DATA DECIPHERED

A VISUAL MIGRATION OF VFX

ROBERT FORDYCE

Victoria University of Wellington 2016

A Thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the Degree of Masters of Design Innovation

ABSTRACT

The visual effects industry is an interconnected network of migratory professionals that is in an on-going state of dynamism. The transient nature of industry contracts and the resultant economic impact of studio ebb and flow is a largely uncharted, yet highly phenomenological subject, within design discourse. In the absence of a reliable metric to quantify employee migration, previous theories in this field have been speculative and conjectural. However, the wealth of data inherent in employment-oriented social-media profiles and online crowd-sourced databases provides a new way in which to identify and analyse collective trends in industry migration.

Data Deciphered: A Visual Migration of VFX reveals the geographical and demographic patterns in the postproduction services industry through the data visualization medium. Furthermore, it investigates the optimal way to comprehend, filter and relate the large volume of information that is the sector's migration patterns.

This thesis first amassed a dataset of 82,711 migratory employment records specific to professionals within the visual effects industry over the previous 35 years. It drew this information from the public-facing pages of both the LinkedIn and Internet Movie Database (IMDB) online Internet platforms. This collection has been subsequently used to drive a 3D visualization tool that was constructed within the Unity5 game engine.

This study has revealed that, despite claims to the contrary, California continues to function as the central hub of the visual effects world and that the majority of industry professionals have been located there at some point throughout their employment histories. Furthermore, environment and matte-painting roles have been identified as the most migratory, while technician and code professions tend to be more static. Finally, skills analysis demonstrates that while proficiency in software packages and coding languages is prevalent within the industry, ultimately, the possession of these abilities has negligible impact upon migration frequency.

ACKNOWLEDGMENTS

Data Deciphered would not have been achievable without the incredible amount of support from the people in my life. While it is not possible to list all of them in this space, I would like to give special mention to those who have had an active hand in this endeavour.

FIRST AND FOREMOST, TO MUM AND DAD

Thanks for the copious amounts of coffee,
Subway and other such amazing meals,
Being my aesthetic and technical sounding board,
For instilling in me an appreciation for hard work,
Motivation to do a quality job,
And for always pushing me to 'Get it done'!

TO KATE AND MISS B

Thanks for the copious amounts of coffee,
Sneaky sessions of Rocket League
When I should have been working,
For application critique and design advice,
And for putting up with my craziness through it all.

TO MY FRIENDS

Those at University and otherwise,
Thanks for the copious amounts of coffee,
Pub quizzes and movie nights,
For reminding me that life still exists outside of research,
And constant encouragement along the way.

TO RHAZES SPELL AND BYRON MALLETT

Original members of the
Digital Workshops of the World team,
Thanks for your hands
In the previous phases of this project,
For your graphical and technical expertise,
And for making this a really fun one to work on!

TO LEON GUREVITCH

Supervisor Supreme,
Thanks for your mentorship and friendship,
Your design and academic critiques
Have been vital to the success of this thesis,
You have continuously encouraged me to
Push to new levels -
A fact of which I am so grateful,
It has truly been a privilege to work on this with you
For the past three years.

FINALLY, TO GOD

I made it Lord!
Thanks for being there with me,
Every minute of every day,
For enabling me and providing for me,
And for being my best friend.

The Lord is my strength and my shield;
My heart trusts in him, and he helps me.
My heart leaps for joy, and with my song I praise him.

PSALM 28:7 NIV

CONTENTS

1 INTRODUCTION

There has been widespread concern that the visual effects (henceforth VFX) industry is experiencing a tumultuous phase that is rife with studio foreclosure and mass migration, however no formal quantitative evidence currently exists that validates and visualises these trends. The limited commentary that surrounds the sector is largely speculative and generally focuses upon specific events, rather than on presenting an objective and comprehensive perspective. In utilising a combination of computer science, media design and statistical methodologies, this thesis endeavours to quantify the industry's migration and present its findings through the medium of data visualization. In doing so, the project will scrutinise the dominant geographic and demographic patterns evident within the global migratory networks of VFX.

INTRODUCTION

The VFX industry of today constitutes a USD$2.5 billion market (Grage, 2015, p.152) and has brought many revolutionary computer graphics milestones to the silver screen during its four decades of evolution (p.62-129). Films famous for their VFX, such as The Matrix (1999), Avatar (2009) and Gravity (2013), are developed through a rigorous postproduction pipeline that consists of various digital processes such as 3D modelling, animation, lighting, rendering and compositing. While these effects are commonly considered in relation to live action footage, it should be noted that in the context of this thesis, studios that deal exclusively in computer generated animation are also to be regarded as VFX companies. It is estimated that the global sector consists of approximately 500 competing firms (Gupta et al, 2013, p.3), although "the precise size of the multimedia and digital visual effects industry is extraordinarily difficult to calculate" (Scott, 1998, p.30). Ever since the VFX industry's emergence within a globalized economy in the early 1990s, there have been two dominant discernible trends. The first of these is the rapid growth of the industry, particularly in regions such as Britain, Canada and Australasia. California meanwhile, the birthplace of VFX, has experienced a professional exodus (Gurevitch & Spell, 2015; VES, 2013, p.5-6). The second observation is a clear rise in the migration of VFX artists (Gurevitch, 2015). This transience and growth is symptomatic of the dynamic status quo, where postproduction work is being increasingly offshored and former industry frontrunners are falling victim to foreign tax subsidies and comparatively cheaper overseas labour (Leberecht, 2014).

In its 2013 white paper on 'The State of the Global Visual Effects Industry' the Visual Effects Society documented the urgent need for quantitative analysis of the sector. It painted a bleak picture of an industry confused by globalization and suggested that a more comprehensive understanding of workforce behaviour is an initial step in rectifying the current issues facing VFX (VES, 2013, p.20). Furthermore, within academic circles, there is a dearth of statistical conclusion that relates to the phenomenon of VFX migration. For computer and social scientists, media designers, and industry recruiters alike, the illustration of how professionals are moving over time is desirable as it provides valid implications of the underlying rationale.

Data Deciphered: A Visual Migration of VFX has intended to anchor anecdotal speculation by presenting the key migratory trends of the VFX industry, based upon irrefutable crowd-sourced data. This thesis assumed its position as the third instalment of the broader Digital Workshops of the World project, headed by Dr Leon Gurevitch at Victoria University of Wellington and funded by the Royal Society of New Zealand. As such, in accordance with the outputs of its predecessors, Data Deciphered has produced a 3D interactive visualization tool that showcases the global migration of the VFX industry over time. The power of this application comes in its ability to allow users to filter and relate the associated data set, which are two of the fundamental tenets of Shneiderman's task taxonomy for information visualization (1996, p.4-5). In providing a platform from which users can observe and query the data, this tool aims to foster the discovery of new population and migratory trends. However, this could not be achieved in the first place without a sampling of the VFX industry. Given the time and resource constraints applied to this study, a full qualitative analysis of the sector that utilises case study, interview and survey methodologies was impossible. Despite this, the collection of thousands of legitimate profiles of VFX professionals from around the world was achieved by using the benefits afforded by the popular social networking platform, LinkedIn. Crowd sourced accumulation and validation methodologies for 'big data' sets were employed to compose a database of 82,711 migrations from 22,554 individual employment histories. Subsequently, statistical analysis was performed on this data to present conclusions that identify the largest VFX regions and their compositions, the most migratory professions and the top coding languages and software packages used by practitioners. This information fulfils the research aims of the original Digital Workshops of the World project and provides valuable insight into the VFX industry.

To summarise this overview, the following quote from Pierre Grage is particularly appropriate due to its appreciation of trends. Facilitating the observation of trends, both known and novel, is the aim of the Data Deciphered project, which is achieved through the medium of visualization. Grage stresses the importance of this "vivid picture" as a tool to make valid predictions regarding the future of the VFX industry (Grage, 2015, p.132):

Trends are like puzzle pieces. They are constantly worldwide in motion. Yet if we are able to correctly analyse these moving trends, …we might get a glimpse of a vivid picture. One that may form right in front of our inner eyes. A forecast in the right direction can tremendously help us making suitable decisions for our best future outcome – no matter if you are making decision for a whole VFX studio or yourself. This is why fortune not only favours the bold but also the attentive one: attentive to trends in motion.

The following documentation initially contextualises the current state of VFX by presenting case studies of studio foreclosures and highlighting the key motivators to migration. Following this, it establishes the case for a crowd-sourced data approach and outlines the methodologies employed to amass the final database. Visualization is then justified as a valid output medium before the final application is discussed with particular emphasis on the design decisions made in its conception. Finally, this thesis presents its data findings before delving into an evaluation of the entire process.

THE DIGITAL WORKSHOPS OF THE WORLD PROJECT

'Digital Workshops of the World' is an on-going research project that is headed by Dr. Leon Gurevitch at Victoria University of Wellington, New Zealand. The primary aim of this initiative is to identify dominant geographic and industrial demographic trends in the collective migration of VFX professionals. The study investigates the industry's evolution over the lifetime of its previous 40 years (Grage, 2015, p.62-129). It specifically aims to present quantitative evidence that validates the widespread speculation and commentary pertinent to this field. In the process of doing this, this project has produced interactive data-visualization tools that enable the discovery of additional phenomena relating to the VFX industry's diaspora.

Digital Workshops of the World has valuable implications in both industrial and academic contexts. For VFX houses, this concise representation of the entire workforce provides a powerful platform from which to gain insight on talent flow, skill acquisition and regional growth. In addition to providing a reflection on the past, the visualization also has potential to function as a forecasting tool through which the extrapolation of current observations and trends can be applied to the future. Despite there being a great deal of blogs and articles online in regards to studio closure and professional transience within the VFX domain (Parish, 2015; Barkan, 2014; Kaufman, 2013), there is a dearth of content that quantifies this industry as a whole, in terms of its migration. From an academic perspective, this formal study holistically illuminates the extent of this workforce's mobility and, in doing so, provides an objective basis to surrounding conjecture.

Preceding this thesis, The Digital Workshops of the World research project had produced a database of migratory employment data and two functional visualization prototypes that exhibited this information. The original data was sourced from IMDB (Internet Movie Database) and consisted of work contracts pertaining to specific professionals. This information had been deduced by cross-referencing names between the credits of various high-profile VFX movies. Subsequently, an interactive prototype was developed within the Google Earth web framework that presented a customised 3D rendering of this data over time. Due to the technical limitations imposed by the Google plugin and adhering to iterative design process, a secondary WebGL realisation was built that offered greater functionality and improved performance.

While there was high acclaim and international recognition for the Digital Workshops of the World research project, there remained room for substantial improvement in both data validation and application performance respects. The original database consisted of approximately 13,000 records of migration from 5,000 employment histories. These numbers were only large enough to offer a tentative reflection of the industry as a whole. Furthermore, the integrity of the database was inherently problematic due to the guesswork involved in linking professionals across movies based upon their associated credits. Additionally, as the database was not built to self-update, records from late 2014 onwards (when it was amassed) were not represented. Moreover, the data visualization was laborious, primarily in regards to its performance and accessibility. Initially coded in JavaScript, the WebGL version was developed as an Internet application and could only be reliably executed in the Google Chrome browser. Due to unoptimised calculations and the simultaneous rendering of thousands of data entries every update, the frame rate for this application suffered

drops in particularly dense sections. An unacceptable performance lag also occurred during the initial loading phase, due to the retrieval and parsing of the required data files. An additional concern was that the only supported projection mode was 3D and therefore it was impossible for users to obtain a complete view of the data at any given time. Collectively, this critique made clear that while the Digital Workshops of the World project had arrived at a respectable milestone, there was still considerable room for improvements in validity, deployment, performance and user-experience.

THE DATA-DECIPHERED PROJECT

'Data Deciphered: A Visual Migration of VFX' was established to advance Digital Workshops of the World to its next milestone, specifically by resolving the key weaknesses of the second iteration, namely data inaccuracies and lagging performance. This thesis utilises a mixed-methodology approach that synthesises computer science and graphic design disciplines. These two facets contribute to the primary output of this project, which is a revised desktop visualization application that exhibits a completely reassembled set of VFX employment data. The chosen development platform was the prevalent game engine Unity5. This was due to its optimised code for graphics-intensive programs, its content-rich asset store, its seamless integration with a variety of computing platforms and its wealth of online documentation and support. The contributing data sources were the public-facing webpages of IMDB and the popular social-networking site LinkedIn. These were selected due to their wealth of embedded information, their profile-centric design and their structured webpage formats.

Fundamentally a data-mining enterprise, Data Deciphered involved a significant number of technical methodologies and techniques. In the collation and extraction of the data from the sources, the Python coding language was heavily utilised to construct scripts that effectively automated the entire process. This was necessary due to the sheer volume of data that needed to be downloaded, pulled, parsed and validated. Text categorization and analysis methodologies were paramount throughout this process as they enabled the encoding of thousands of disparate elements of raw data into a manageable set of restricted values. The development of the visualization also demanded a heavy technical requirement. Spherical mathematics were employed through the C# programming language to create efficient calculations that enabled migrations to update in near real-time. Furthermore, the repeated refinement of polymorphic system architecture allowed for complex filtering, navigation and user-interface behaviours to be achieved through relatively lightweight code. Finally, the tailored use of Unity5 particle systems and line rendering components harnessed the underlying power of the engine to optimise graphics and rendering performance.

The graphic and interaction design of the application was an aspect of the project that underwent continual scrutiny across its development. A balance needed to be achieved between the functionality of the system and the simplicity of the user interface. As the primary responsibility of the visualization is to facilitate the easy discovery of trends within the database, the graphical widgets needed to operate powerfully, yet contain this functionality within a few simple controls. Adhering to the principles of Edward Tufte's information-led design (2001) and Ben Shneiderman's information seeking mantra (1996), a minimalistic aesthetic was prescribed that preserved the data integrity of the presentation and also heightened accessibility.

2 SETTING THE SCENE

The VFX industry provides an ideal backdrop for research into migratory patterns due to the highly transient nature of its workforce. It is important to understand the rationale behind this phenomenon as familiarisation with the cause equips the reader to best comprehend the quantitative results and their wider implications. As such, the following section endeavours to contextualise the VFX industry and explain the primary motivators of its inherent migration. Specifically, it presents the commentaries of Grage, Gurevitch and the Visual Effects Society, which identify a dominant Hollywood exodus in the advent of an industrial awakening to globalisation (2015, p.163-195; 2013, p.5-6). The foreclosures of the Digital Domain and Rhythm and Hues VFX facilities are given as case studies that exemplify the resultant dynamism prevalent in this sector (Leberecht, 2014; Gupta et al, 2013, p.2). Furthermore, this section elaborates on the key inciters to migration. Among these are the international tax rebates imposed to attract movie making deals, the rise of Asia as a VFX hotspot, the advancement and commercialisation of high-end software and the influx of young professionals into the workforce. Finally, in conjunction with the most recent white paper published by the Visual Effects Society, it commends this research as an important aid in providing clarity to a nomadic industry confused by globalisation (2013, p.20).

A HOLLYWOOD DIASPORA

The current business model of the digital VFX industry was birthed in 1970s California following the first instance of computer graphics in George Lucas' Star Wars (1977) (Grage, 2015, p.64-66). The industry has always competed for work from an oligopoly of six film studios, known colloquially as the 'Big Six', whose film units are all based in Hollywood, Los Angeles: Universal Studios, 20th Century Fox, Columbia TriStar, Warner Brothers, Paramount Pictures and The Walt Disney Studios. Before the dawn of the new millennium, there existed no competition to rival American VFX studios and it made sense therefore for these facilities to neighbour their clients. These companies had defined the field of digital VFX. They owned the majority of the people, software, concepts and technologies that were foundational to the industry's inception (Grage, 2015, p.170). However, the 2000s introduced the sector to globalisation – a union that resulted in the phenomenon of 'runaway production' (VES, 2013, p.5-6). Through attractive foreign tax subsides and lower labour costs in offshore markets, film studios were encouraged to auction their postproduction work abroad. In chasing these contracts, VFX artists began to leave California at a fast rate, spurring an industrial exodus from the homeland of film production (Gurevitch, 2015). Many studios were forced to shut up shop due to the intense financial competition from newly formed foreign competitors. This in turn left a further tier of unemployed artists who also had to permanently relocate in order to remain afloat within the industry. In fact, over the decade from 2003 to 2013, 21 major VFX companies either closed or filed for bankruptcy (Leberecht, 2014). Californian companies that have survived the globalisation punch have typically done so by investing in overseas facilities and reducing local staff to creative development, production management and support roles (VES, 2013, p.6). This Hollywood diaspora is indicative of an unstable, dynamic and therefore highly migratory industry.

STUDIO SHUTDOWNS

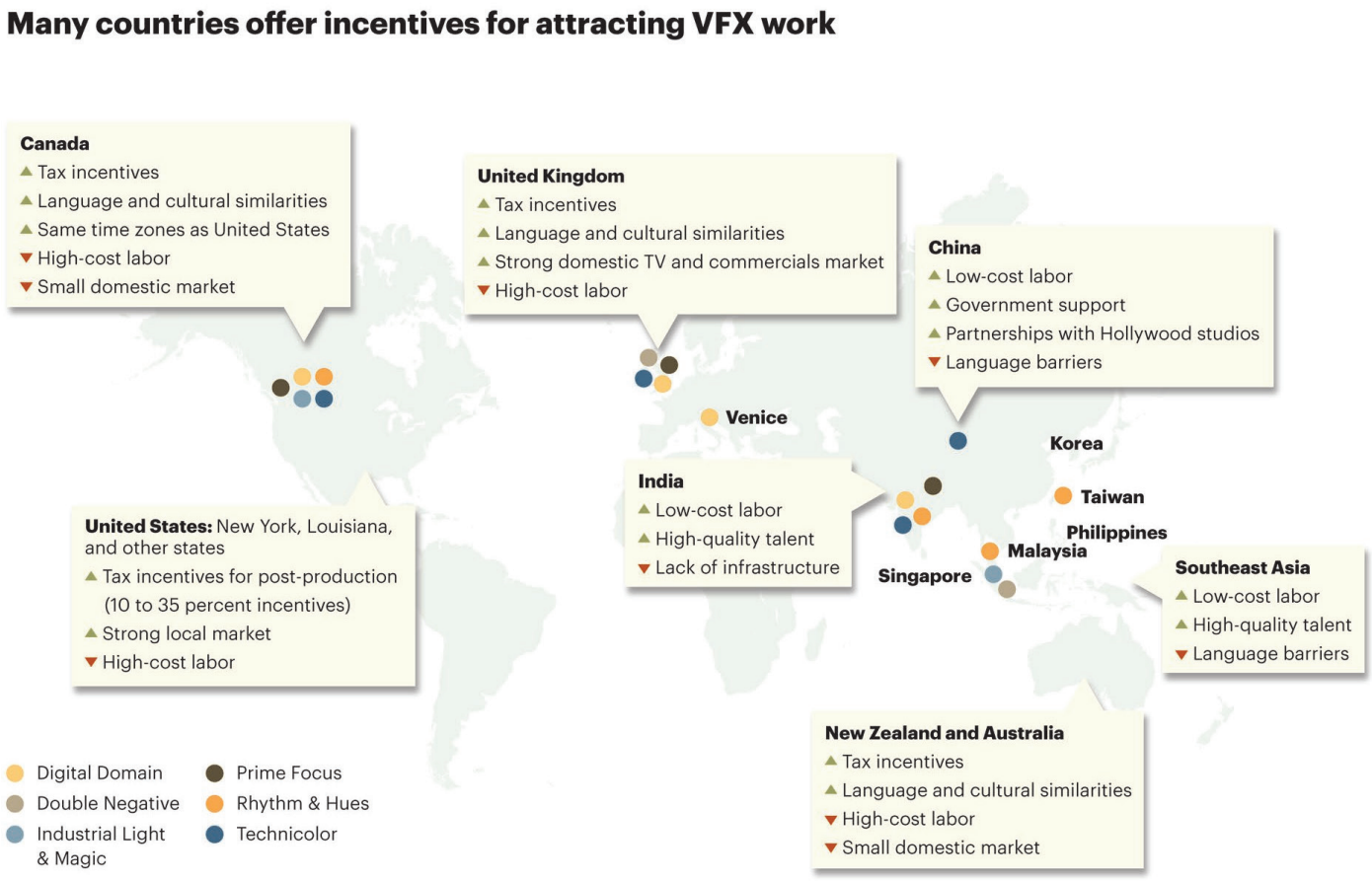
While the frequent formation and foreclosure of studios within the VFX sector has become something of a status quo, the bankruptcies of Digital Domain in 2012 and Rhythm and Hues in 2013 profoundly shocked the industry. In wake of these events there was global confusion at how two of the most prolific and reputable companies within VFX could spontaneously combust. Digital Domain - the VFX company behind Titanic (1997), The Curious Case of Benjamin Button (2008) and 2012 (2009) was regarded as one of the top six VFX studios in the 1990s (Grage, 2015, p.227-232). However, following a change in management in 2006, the facility began its derailment by taking on too many concurrent initiatives. Initially, the Digital Domain Institute was launched in Florida, which functioned as an educational facility. This organisation effectively generated free labour by offloading aspects of Digital Domain's work onto its students. This was seen as a controversial move by the professional industry, as it felt undermined by this program that enlisted students to pay to work for free. Furthermore, the company simultaneously attempted to initiate a computer animation division and the preproduction and production phases of Ender's Game. Additionally, the studio also experimented with 3D hologram technology for the specific purpose of reimagining performers posthumously live on stage. All of these concurrent projects ultimately proved too much for the VFX facility and on September 12th, 2012 it filed for Chapter 11 bankruptcy after defaulting on a $35 million loan (Grage, 2015, p.229). However, the company was subsequently bought out by Chinese and Indian investors and exists today as a subsidiary of Reliance MediaWorks and Sun Innovation.

Rhythm and Hues offers another example of a widely publicised VFX studio foreclosure. The company was renowned for its high-quality effects work in titles such as Babe (1995), The Golden Compass (2007) and Life of Pi (2012). The supreme irony present in this case was that the Oscar that Rhythm and Hues received for Life of Pi was awarded eleven days after the company declared bankruptcy (Gurevitch, 2015; Leberecht, 2014). Many commentators saw the studio's decline as systematic of the 'flawed' fixed-bid business model between the VFX and film industries (Leberecht, 2014; VES, 2013, p.12-13). The unexpected cancellation of Rhythm and Hues' contribution to Snow White and the Huntsman (2012), coupled with tight profit margins as a result of having to compete with foreign agencies, ultimately rendered the studio a victim of a cash crunch. This

occurred despite the company's preventative counter-measures of engagement in co-production deals, expansion to low wage countries and the utilisation of foreign tax subsides (Grage, 2015, p.236-237). Given the studio's textbook means of operation, this tumultuous event served to galvanise the VFX industry in mutual fear and resolve (Leberecht, 2014). The foreclosure of Rhythm and Hues in the immediate wake of Digital Domain exemplifies the recent unpredictability and dynamism of this sector, identifying the industry as a prime candidate for migrational research.

MOTIVATORS TO MIGRATION

Before delving into a thorough analysis of the rationale behind professional movement in the VFX industry, it is important to illustrate the inherent financial and demographic attractions of the various VFX hotspots. Figure [2.1] provides an overview of the incentives to migration for VFX professionals. Published by A.T. Kearney analysis, this chart identifies the primary regions of VFX activity and lists the specific advantages and disadvantages that are associated with production in these locales (Gupta et al, 2013, p.5). While this infographic acts as a visual aide to the following discussion, the particular list of included studios is now out-dated. For instance, both Digital Domain and Rhythm and Hues have both filed for bankruptcy, and Double Negative and Prime Focus have subsequently merged (Bevir, 2014) since this chart's creation in 2013. Grage includes Weta Digital, Sony Pictures Imageworks, The Moving Picture Company and Framestore in his analysis of the 'Big Eight' VFX houses of the 2000s (2015, p.225). Their absence in this figure creates the misleading implication that there is not a VFX presence



Note: VFX is visual effects.
Sources: company websites; A.T. Kearney analysis

[FIGURE 2.1]

Many countries offer incentives for attracting VFX work. From Gupta et al. (2013). And Action - Making Money in the Post-Production Services Industry, A. T. Kearney, p(5).

in regions such as Australia, New Zealand, Montreal and California. In contrast, areas such as Venice, which do warrant a dot, have a comparatively smaller VFX population (Gurevitch & Spell, 2015). However, while imperfect in its portrayal of the current state of the industry, this chart is still useful for reporting on regionally specific motivators to migration. These aspects are further elaborated upon in the following section.

The abundance of commentary surrounding VFX migration collectively identifies four main contributors to the industrial transience. The first of these is the highly controversial issue of regional tax subsides. In order to attract digital technology companies and markets, governments offer significant tax rebates to film studios in return for the outsourcing of VFX contracts. These tax incentives are particularly high in Canada, the United Kingdom, Australia and New Zealand (Fritz, 2013), which explains the dramatic growth of postproduction facilities in those regions. Interestingly, in the filming of The Hobbit trilogy, Warner Bros. threatened the New Zealand government with project relocation unless tax conditions were more favourable. Bowing to this pressure, New Zealand rewrote its labour laws to retain ownership rights and capitalise upon the associated tourism and industrial benefits of blockbuster production. As of 2015, New Zealand taxpayers have contributed more than NZD$150 million to the trilogy (Parish, 2015). This is one of the many examples that indicate the influence of Hollywood in its extensive outsourcing of highly potent movie contracts. In fact, such is the status quo that film studios now expect tax incentives; making them mandatory within film budgets or authorising production in a region based upon qualification for a tax rebate (VES, 2013, p.14-15). In its bankruptcy court filing, Rhythm & Hues estimated that in considering subsides, exchange rates and labour practices, subsidising regions were effectively given financial bidding advantages of between 35% - 60% over Californian companies (Fritz, 2013). Scott Ross, former General Manager of Industrial Light and Magic and Co-Founder of Digital Domain, regards tax credits and subsides as “the most detrimental issue facing the VFX industry by far” (Barkan, 2014). In the Life After Pi documentary, Prashant Buyyala of Rhythm and Hues India explains the crippling effect of subsides upon Californian studios that are trying to stay afloat in an intensely competitive industry (Leberecht, 2014):

> *If a studio decides there’s a $10 million project they want to take to [a region offering tax subsides for visual effects work], the [local VFX company] will bid it at $10 million, but the studio gets, say, $3 million back as a tax rebate. So the only way to even get considered for that film is to bid it at $7 million. So we’re now getting less money for the same amount of work.*

This model has prompted the globalisation of VFX studios. The industry frontrunners have expanded their influence by opening up satellite bases in tax-favourable areas. Many start-up studios have also followed this trend, resulting in an industrial paradigm shift from the traditional Californian-centric picture of VFX (Grage, 2015, p.170-172; VES, 2013, p.5-6). VFX supervisor Scott Squires labels VFX as “an industry based on a house of cards” (Kaufman, 2013), with the implication that the absence of subsides will result in sectorial collapse. This warning is particularly relevant as no governments have guaranteed indefinite incentive support, which implies a vulnerability to external economic, social or political influences that could potentially jeopardise regional stability (VES, 2013, p.6). The volatile nature of VFX could be reiterated if these regions were also to experience runaway production in the future.

On a corporate level, the VFX industry can be regarded as highly dynamic due to the widespread expansion of studios and the significant quantity and frequency of foreclosures and start-ups. In terms of individual professionals however, these events imply routine migration in the pursuit of work. Kristy Barkan of ACM SIGGRAPH describes tax programs as a flawed model. While subsides are intended to attract companies and encourage local job growth, often the emigrating professionals are transplants who set up a temporary residence for the duration of the production, before departing again for the next contract (2014). Squires affirms this by noting that, far from rooting industry, subsides only serve in its rotation. “VFX companies have the expense of setting up satellite places and VFX professionals have to move around the world… and it changes every six months where the subsides are” (Barkan, 2014). This nomadic lifestyle and the typical brevity of VFX work contracts are the primary reasons as to why the industry exhibits such geographic migration among its professionals.

The development of technology and software is another contributing factor to increased professional migration. The typical postproduction house produces gigabytes of new data each day (Dodgson, 2010, p.6-7). It would be impossible for VFX studios to maintain synchronisation across their various satellite facilities without the advantages that modern technologies provide. High-speed communications, efficient data transfer protocols and the decreasing cost of Internet bandwidth have effectively removed the requirement for facilities to operate in close proximity to one another (VES, 2013, p.8-9; Chung, 2011, p.6). In regards to migration, this technological evolution has facilitated studio expansion and therefore the globalisation of the industry's workforce. Additionally, the notion of constant workflow through the strategic deployment of facilities in different time zones is a very attractive aspect to pipeline optimisers. In its heyday, Rhythm and Hues reflected this point with the catch phrase "the sun never sets on Rhythm and Hues" (Chung, 2011, p.2). Working around the clock in this way would not be possible without the technological advancements in software integration, cloud computing and asset management. Furthermore, the large-scale commercialisation of software packages has prompted the establishment of start-ups around the world by empowering the layperson with the ability to produce high quality effects. Where traditionally VFX studios utilised elitist in-house proprietary software, nowadays every major studio employs standardised commercial programs with their own plugin extensions and customisations (Grage, 2015, p142). In terms of pricing, advanced VFX software that once cost several thousand per user is now available for several hundred (Fritz, 2013). Additionally, for students, certain companies provide free licences. This is with the understanding that these students will be more likely to operate in their programs as junior professionals, having previously trained in these environments. This increased accessibility to industry-leading toolsets has effectively removed a once-formidable barrier to entry for prospective competitors. In this way, it inspires start-up initiatives around the world, thus providing greater avenues and potential for migration. Additionally, for VFX recruiters, technological developments and the proliferation of social media networks, such as LinkedIn, have dramatically increased the size of the hiring pool. Nowadays, professionals can be easily identified and interviewed despite geographical separation and this also contributes to an industry-wide increase in mobility.

The rise of VFX in Asia has effectively augmented the predominantly Western industry with a massive workforce, providing additional opportunities for migration into and between Indian, Chinese and Singaporean markets. Asia has always been an appealing option to VFX studios due to its low cost labour (Grage, 2015, p.197-212; VES, 2013, p.9). Working conditions for the majority of professionals within these regions are grossly unfair in comparison to Western standards. This results in a notoriously high rate of workforce turnover, which implies significant internal migration within these regions. Grage reports that of the estimated 10,000 employees that comprise India's overall VFX population, 70% are earning 10,000 rupees per month, which is equivalent to USD$165 (2015, p.200). Despite these humanitarian and political concerns, the outsourcing of work to these regions – particularly less sophisticated effects tasks such as rotoscoping, keying and matchmoving – makes financial sense. Many Western studios have resultantly deployed facilities offshore or chosen instead to partner with prominent Asian effects companies, rather than compete remotely. For example, in 2012, DreamWorks Animation, in association with various Chinese investment companies founded Oriental DreamWorks in Shanghai. That same year Industrial Light and Magic announced that it was to form a strategic alliance with Chinese studio BaseFX (Grage, 2015, p210-211). In 2014, British VFX giant Double Negative merged with India's Prime Focus World in a bid to further expand its global presence (Bevir, 2014). Additionally, in relation to the acquisition of Digital Domain, CEO Daniel Seah is quoted as saying, "Buying a visual effects company will be an easier path for them to potentially link Hollywood and entertainment in China in the short future" (Grage, 2015, p.231). The Chinese market in particular promises significant returns for film and VFX studios alike. The Hollywood Reporter notes that in 2015 the Chinese box office grew an incredible 47.8%, reaching a record of USD$6.78 billion. At its current rate of growth, China is predicted to surpass North America, by the end of 2017, as the largest movie market in the world (Brzeski, 2015). For film studios to specifically gain entry to this market they need to work around the foreign film import regulations imposed by the Chinese government. As of 2014, these quotas allow a mere thirty-four foreign films to be imported annually. However, as Hollywood captures an estimated total market share of 70% (Grage, 2015, p.205), film companies are partnering with Chinese enterprises to produce content internally and thereby take hold of this lucrative market (VES, 2013, p.9). It is therefore in the best interest of VFX

studios to establish a Chinese foothold through which to accommodate this trend. The evident desire to capitalise upon cheaper labour and venture into a highly potent, yet largely untapped Oriental market, implies a greater potential for VFX migration to Asia in the future.

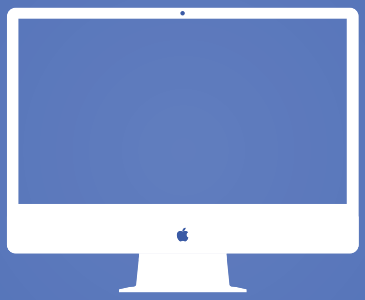
A further contributor to migration is the proliferation of VFX education. In the early days of the industry, the geographical and academic barriers to entry were high. Californian VFX studios originally held an effective monopoly on computer graphics education (Grage, 2015, p.156) and enthusiasts were required to migrate to this region in order to learn the necessary skills. Furthermore, in a 1998 labour survey performed upon the digital VFX industry in Southern California, the majority of participants were reported as being highly qualified, with most having undertaken at least four years of tertiary study and some even possessing Masters degrees (Scott, 1998, p.35). With the popularisation of the Internet in the early 2000s however, these barriers were significantly reduced. Digital training schools, such as the Gnomon Workshop, began to offer specific training DVDs to hobbyists who were in pursuit of industrial internship. Furthermore, web forums developed into an invaluable resource for troubleshooting and thousands of amateur video tutorials became freely available online (Grage, 2015, p.156). This increased accessibility to high quality, low cost education made possible the influx of professionals in the new millennium, as the demand for VFX grew. This saturation of fresh talent is even more evident today. Hundreds of colleges and universities around the world now offer specifically tailored VFX and computer graphics training programs (VES, 2013, p.9). Resultantly, every year a new wave of students graduate and look to be absorbed by the industry. As the sector is composed of predominantly young professionals the transient and global nature of this career is seen as desirable, as it provides a joint opportunity for work and travel (Chung, 2011, p.8; Scott, 1998, p.32). The heightened accessibility to VFX education around the world, coupled with the loosening of qualification requirements, has enabled more people, from a wider variety of locations and demographics, to enter the industry. This influx inherently promotes migration and naturally boosts the dynamism of the VFX sector.

RIPE FOR RESEARCH

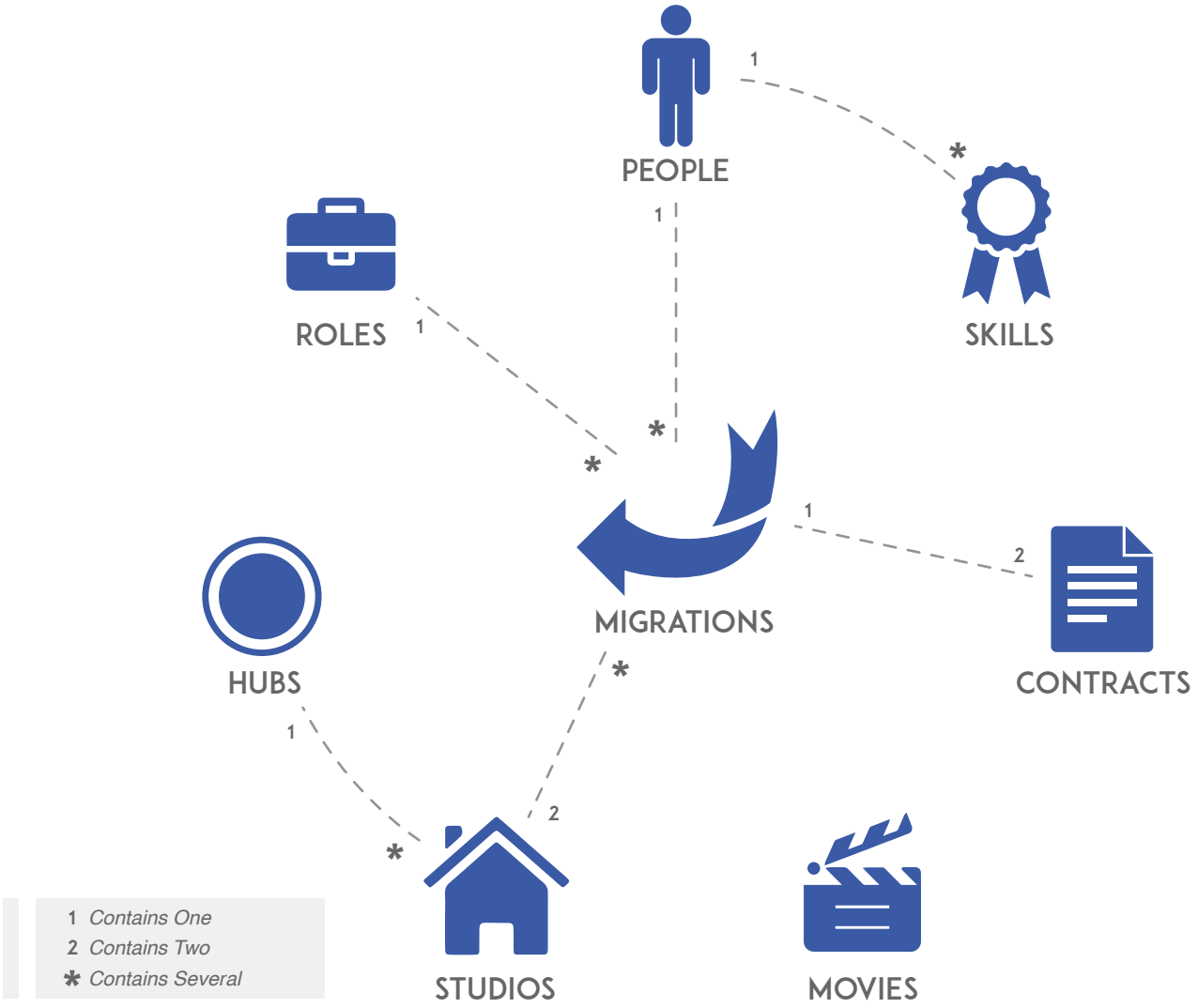
These previous discussions on the state of the VFX industry clearly warrant studies such as Data Deciphered and Digital Workshops of the World. Additionally, in the introduction to his survey on industrial satisfaction, Squires acknowledges that there is a "lack of any real data regarding companies and workers", as no current enterprise or organisation regularly monitors this sector (2013). While the scope of this thesis is too restricted to comprehensively quantify the industry, the research outputs endeavour to clearly portray meaningful trends and insights. To conclude, this quote, extracted from the Visual Effects Society's most recent white paper, summaries the need for quantitative study in an industry ripe for research (VES, 2013, p.20):

> *Part of the reason the demand is so unpredictable is that there is very little hard data about the effects industry. Statistics about the size of the market (in dollars, employees or shots produced) are scant, and any numbers mentioned in articles or interviews tend to be speculation and educated guesses. Without real information, trends are difficult to identify and impossible to forecast, and the impact of changes on policies and standards can't be measured or evaluated. A legitimate set of historical data could be very helpful in understanding the industry better.*

3 THE DATA MINING PIPELINE



Data contextualisation and analysis methodologies comprise the foundation upon which the Digital Workshops of the World initiative is framed. Falling within the discourse of 'Big Data' management, the project aims to categorise, refine, explore and present key features and trends within the database in a way that is intuitive for the layperson to understand. Data Deciphered shares in this endeavour, yet identifies current-standing omissions, both in the process and output of the data-mining pipeline. The following section initially provides a technical breakdown of the database schema and subsequently justifies the need for a new and revised set of data to underpin this project. It engages in the wider conversation on crowd-sourced data gathering, analysis, ethics and management, before finally delving into a thorough documentation of the five-step collation process.



[FIGURE 3.1]

Data Deciphered Database Class Diagram. Shows the relational connectivity between the data entities in the database schema.

DATABASE SCHEMA

Data Deciphered operates upon an input collection of migratory employment data. At a technical level this translates to a set of records, each consisting of a migration component (an origin and a destination, a start and an end date), and an employment component (an association to a particular person and an occupation in the database). The migrations act as the building blocks of the visualization in that they bind the various structures of secondary data together into a primary, meaningful entity that can be displayed. Collectively, the migrations constitute the visualization and their cumulative ebb and flow demonstrate the industrial trends that this thesis aims to scrutinise. In addition, the database also consists of secondary units that are effectively linked through migration structures:

- **Hubs** consist of a name and a geo-location. These units are visualised about the globe in key VFX hotspots and provide points from which a migration can either arrive or depart. The hubs function as a visual aggregate of studios, aiding clarity through solving the predicament of iconising approximately six hundred different studios simultaneously.

- **Studios** describe actual visual effects, television or film-related companies. A studio identifies with a particular hub and is therefore geo-locatable. This structure contains the company's inauguration date and may also contain a closure year if the studio has gone into receivership. Associated filmographies, in addition to industry and population statistics

are included within this entity.

- **Contracts** are implied in the migratory data, but specifically describe a record of employment. Whereas a migration requires two contracts in order to infer the jump between, a contract singularly details a period of time spent working at a place. In this way, a contract is comprised of a studio, a role, a person and start and end dates.

- **Movies** exist somewhat independently of the other secondary units and effectively act as a timestamp in film history. In juxtaposition with migratory movements, movies can aide in the contextualisation of the database by providing rationale for sudden influxes or releases. As a chronological entity, movies have associated release dates, as well as budget information and gross box-office earnings. They can be further grouped by country, language or genre.

- **Roles** form a predefined list of occupations that categorise the various titles given to visual effects workers. This enumeration is important for both data validation in the mining process and data clarity in the visualization. Roles provide an intuitive tag with which to filter the data set.

- **Skills** are additional tags that industry workers associate with their lifetime careers in visual effects. Where a role is used to describe the specific type of work a person performs as part of a contract, a skill can refer to any competency they have attained along the way. Skills can be effectively combined with roles to perform more specific queries of the data. The introduction of skills into the directory was made possible through the information inherent in the LinkedIn data source, and therefore this novelty is unique to the Data Deciphered project.

- **People** are effectively bridge structures that link migrations that are common to a specific person. Furthermore, people have an associated set of skills. A person structure is considered anonymous due to the fact that identifying information, such as name, current residence, age and gender have been deliberately culled from the database for ethical and size-reduction purposes.

THE NEED FOR NEW DATA

A primary concern in regards to the Digital Workshops of the World project was the integrity of its dataset. This was largely due to the singularity of the extraction source and the multiple assumptions made in its process of validation. In cross-referencing the credits of prominent VFX-heavy films on the Internet Movie Database (IMDB), it was possible to assume employment histories for specific individuals, based upon the release dates of the films and the associated production houses. Re-rendering the data for the purpose of establishing visual migrations introduced further suppositions, as artificial dates were injected into the set in order to simulate start dates for the jumps. Unfortunately, this assumption inevitably compromised the validity of the database. While someone who operates at the end of the VFX pipeline may have been roughly represented by the data estimates, pre-production workers have the possibility of being misdated by months or even years, due to the fact that their period of employment is further removed from the known release date of the film. The limitations of the IMDB structure invoked the need for significant assumption and this, in turn, put the project's validity further into question.

In addition to the dubious integrity of the original output, there were also holes present in the database itself that lead to unsavoury Band-Aid solutions in the Digital Workshops of the World visualization. A primary example of this was being unable to map all studios to their respective hubs. Those that were unassigned were, by default, associated with the Californian base, due to the fact that the region is historically the centre point of the industry. However, this was far from an ideal solution, as it again introduced assumptions that were likely a misrepresentation of the actual migratory routes. Furthermore, the original dataset did not allow for studio disambiguation. Many of the larger VFX houses have multiple satellite offices established around the globe and the Digital Workshops of the World convention was to group any branch of a particular VFX studio into the same hub as that studio's headquarters. This ruling was due to the data source's lack of information that could imply a more specific locale. However, these ambiguities commonly resulted in the neglect of hubs that were known to be active. As previously stated, regions such as Singapore, China and India are widely reported to be an attractive financial option for film studios, yet the original visualization displayed an obvious dearth in migrations to these areas. This, in part, was due to the fact that while many Western VFX

houses have at least one offshore Asian facility, their base of operations is typically located in the United States or the United Kingdom. In this instance, the data consolidation on studio structures was too stringent, resulting in a distortion of the predicted picture.

The technical ramifications of an incomplete or incompatible database are significant, especially when trying to re-interpret the data in an external visualizer. Roles that had not been properly mapped and studios that had not been properly geo-located were assigned a 'bad data' tag, which effectively acted as a null value. The introduction of empty values as a possibility implied that more preventative code checks needed to be established in order to mitigate or avoid runtime exceptions. This had the added consequence of fleshing out scripts, making them harder to debug and more prone to error from a software engineering perspective. Furthermore, the migratory output from the database consisted of superfluous information, as well as certain unformatted properties that needed to be reinterpreted by the visualization. Technically, this resulted in longer loading times, as an otherwise preventable massaging phase needed to be performed upon thousands of migrations to adapt them at runtime.

LINKING IN LINKEDIN

The issues detailed above provided rationale for a new and revised data set. The limitations of the IMDB format implied that the mere repetition of the data extraction process was insufficient; it also needed to be pulled from a more comprehensive and valid source. The professional social networking site LinkedIn was an attractive option for several reasons. With 300 million subscribers announced in April 2014, LinkedIn is regarded as the most popular social networking site for professionals (Dai et al, 2015, p.1). In this regard, it provides the capacity to substantiate clear trends in industry. In addition, LinkedIn is primarily a record of people's previous employment contracts. Each contract documents the timespan, the employing company, the occupation and optionally, a note that indicates the region. For the purposes of data acquisition, LinkedIn was an ideal source as the data was essentially in a directly applicable format. Furthermore, this all but removed the issue of assumption, as the interpretation involved in the extraction pipeline was significantly less.

In terms of studio disambiguation, LinkedIn offered solutions through its optional region text-field, (which can be cross-referenced against the studio), or through having several different pages for a particular VFX company (for example, 'Lucasfilm' has two dedicated company pages – both 'Lucasfilm' for its Californian headquarters and 'Lucasfilm Singapore' for its South East Asian office).

As LinkedIn is essentially a vast online spreadsheet, it enforces data validation over user-input information. This is desirable as it means that certain properties are guaranteed to exist, such as the company and the occupation of a contract. Technically, this implies that null values will not be present under these headings in the database, and the flow on effect of this is that the code within the visualizer will not have to be burdened with programming checks that mitigate null-pointer exceptions.

LinkedIn also has the appeal of extra skill information that may be appealing to an end-user. Skills were not originally part of the Digital Workshops of the World database, but they are optionally present on a LinkedIn user's profile. As skills are tied to specific workers over the course of their careers, in conjunction with occupation listings, this information can make for more specific, meaningful and interesting data queries.

The original database was amassed at the end of 2014 and was therefore out of date. The previous data-mining pipeline had never been coded to automatically update, as this would have introduced another level of complication, which was unjustified for the scope of the project. In this regard, a 2016-applicable database was a welcome improvement. For these reasons, and the fact that the construction of a new database and the subsequent cross-confirmation of the sets would further validate the findings, it was decided that LinkedIn would be the data source for the Data Deciphered project.

LEGAL AND ETHICAL CONSIDERATIONS

LinkedIn provided a perfect platform for the data-mining endeavours of Data Deciphered, however, various legal and ethical considerations are raised when one considers extracting this information in significant quantities. The large-scale, automated retrieval of web pages and their inherent data is commonly known as

'scraping', and LinkedIn implements several policies within their User Agreement against this practice.

> *[You agree that you will not] Scrape or copy profiles and information of others through any means (including crawlers, browser plugins and add-ons, and any other technology or manual work)*
>
> *(Section 8.2 - LinkedIn, 2014)*

These protocols are not without teeth. In January 2014, LinkedIn filed a highly publicised lawsuit in California that indicted anonymous 'John Does' with attempting to steal LinkedIn user's personal profile data (Roberts, 2014). Leveraging Amazon's cloud computing service as a virtual smokescreen in their endeavour, these data-miners fabricated thousands of fake user accounts to 'connect' with valid users and thereby siphon their information. The offending company was eventually named as local start-up Robocog, which was trading under the name HiringSolved. In the business of "people aggregation", this recruiting technology agency was capitalising upon LinkedIn data to feed its own candidate database. After months of negotiations, both parties eventually came to a settlement agreement, in which HiringSolved paid $40,000 in damages and agreed to permanently delete all LinkedIn-extracted data. It is important to note that this kind of data collection and aggregation from across various social networking sites is an increasingly common task, as the need for fast, dynamic and accurate talent searching grows. A simple Google query such as 'LinkedIn profile extractor' returns a vast range of diverse results and so it is naïve to assume that this illegal form of scraping and profiteering isn't occurring on a constant basis.

The LinkedIn versus HiringSolved ruling raises fundamental issues pertaining to the wider discourse surrounding the ownership and management of Big Data. The reality of life in the Information Age is that data permeates our very existence; it is recorded as part of our daily activities, measured and analysed by our smart devices, and has become a currency of sorts within social media. As data is so inextricably linked between individuals, companies and media, valid questions are raised when corporations claim exclusive rights or outright ownership. In particular, those in the business of 'information architecture' hold concern. Originally coined by Richard Saul Wurman, former professor of architecture at the University of North Carolina, this vocation is ascribed to those who task themselves with avoiding 'information anxiety', or "the black hole between data and knowledge" (Cairo, 2013, p.15). Information architects advocate for shared information environments, shaping products and experiences to support usability and inference, and bringing the principles of design and architecture to the digital landscape. In essence, these researchers task themselves with bridging "the ever-widening gap between what we understand and what we think we should understand" (Cairo, 2013, p.15). As such, their beliefs regarding the open sourcing of corporately owned data are in direct opposition to business strategies that involve monopolising this data for profit. The unexpected proliferation of data in recent years has left legislation in the lurch and as such, there exists a grey-zone into which this debate between private ownership and public access falls.

In the event that the balance is tipped too far in favour of corporate ownership, the benefits of Big Data innovation are thrown into jeopardy. The Stop Online Piracy Act (SOPA), put before Congress in late 2011 and early 2012, is an example of this that was met with fierce backlash from industry, organisations and individuals (Davis, 2012, p.8). The underlying fear was that the provisions of the law would restrict future innovation in highly potent digital resources such as Big Data. "SOPA represents a classic example of how a lack of transparent and explicit discourse about how a critical piece of our economy and society works had the potential to significantly limit our collective ability to benefit from those tools" (Davis, 2012, p.8). In this regard, it is important to understand the privileged position of Big Data and the potential detriment that withholding this information from the public domain could cause.

Corporate greed is not the only rationale behind limiting data-access to specific companies. The issue of privacy is gaining significance in the area of ethical data management, as more and more people are becoming accustomed to sharing sensitive information across a social networking platform. As the data is frequently about people and their characteristics or behaviour, there is significant potential for abuse through the acts of capturing, aggregation, selling, mining and linking (Davis, 2012, p.9). Information architects must ensure that, in their zeal to assimilate this rapidly expanding body of knowledge, they do not

infringe upon the rights of those they seek to inform.

Data Deciphered is attributed to information architecture in the sense that it aims to shine light on the transient nature of the VFX industry by providing quantitative evidence of frequent employee migrations. The strict terms of use put forward by LinkedIn seek to ensure the integrity and correct use of client proprietary data. However, they also significantly reduce the available data-mining avenues for prospective researchers. The application programming interfaces put forward by LinkedIn to independent developers are severely restrictive and do not accommodate data gathering on any reasonable level. Furthermore, while third party extraction software exists as a potential solution, this inevitably results in a loss of control over the process, raises a legal red flag and also incurs a financial cost. After considerable research, this project decided to embrace a commonly adopted strategy used by several prominent data-wrangling services.

PUBLIC VERSUS PRIVATE WEB

It is important to note that any given LinkedIn profile consists of two different identities and that the choice of identity when scraping data either breaches or circumvents the LinkedIn terms of use. Prospect Visual, a data-mining enterprise, has detailed its procedure, which capitalises upon search-engine functionality. This thesis has replicated this process, in order to adhere to legal and ethical responsibilities.

Social media platforms have an ethical obligation to their members to ensure the integrity and privacy of their data. As such, these companies implement terms that disallow the unauthorised extraction, storage and propagation of this information. Typically, members agree to these regulations when they register for an account. As client proprietary data is privileged information, logging in to export data from these platforms is usually a violation of the terms of service.

Ambiguity is introduced when one considers that search engines, such as Google or Bing, are intentionally granted permission to scrape the platform in order to heighten publicity and accessibility. In this way, profiles can be examined via the search engine without the prior registration of the viewer, and therefore without agreement to the terms of service. The key distinction between an unregistered search engine request, and a registered platform request, is the identity that is returned to the viewer. In the former instance, the requester is only presented with the public-facing identity, while in the latter it is the private, usually more-comprehensive identity that is shown.

Social media platforms commonly allow their members to control which information is made publicly accessible, available to their ‘friends’ or ‘connected’ profiles, and which is to be made private. In this way, it is under the express permission of the owner of the data that information is made available on a public-facing profile. “This choice makes the entire issue wholly transparent. When a user chooses to make their profile public, they are deciding not to restrict their data to the platform they are using, choosing instead to share it with the wider web via search engines” (Prospect Visual, 2014). Furthermore, the question of ethics is rendered irrelevant, as the owner of the information has effectively waived all privacy rights to the data by placing it in the public domain.

It is with this understanding that Data Deciphered maintains a legal and ethical stance in regards to its scraping procedures. The entire data-mining process was undertaken through the window of the Google search engine, and as such all information contained within this project’s database is guaranteed to exist in the public domain.

AN ANONYMOUS DATABASE

One of the single greatest concerns underpinning the digital privacy debate is that online information can be misappropriated to compromise the virtual identities of others. It is ethically responsible therefore to adhere to the highest possible degree of anonymity in a database that deals with personal information. It is important to note that while Data Deciphered collected and analysed personal profiles, it disregarded identifying properties such as name, gender, age, email and address. As this study was concerned with the collective trends of a specific industrial subset, it was counter-intuitive to populate the database with irrelevant information.

The database was primarily constructed from individual contract records, which were conceptually linked to a person through an ID tag. While this initially appeared to preserve anonymity, a study by researchers at the

University of Texas has shown that "even if identifying information such as names, addresses, and Social Security numbers has been removed, the adversary can use contextual and background knowledge, as well as cross-correlation with publicly available databases, to re-identify individual data records" (Narayanan and Shmatikov, 2007, p.1). Specifically, these researchers analysed an academic release of a Netflix movie-rating database. Even though this set did not include names, through cross-checking these ratings against public reviews on the open IMDB, this study was able to re-identify two people out of nearly half a million entries. The significance of this finding is made clearer when one considers that one of the identified had particularly strong views on homosexual-themed and religious films. Suddenly, it was possible for the member's political views to be unintentionally projected onto a public stage. This paper is compelling as it illustrates the dangers of information aggregation; essentially, taking unclassified information sets and combining them into something that is classified.

With consideration to the above, Data Deciphered aimed to disguise the amassed data in order to ensure anonymity to the greatest extent possible. The greatest risk of correlation lay in the combination of individual employment records, and while this link had to exist for technical visualization purposes, it could be purposefully encoded and hidden within the tool. When exported from Unity, the database was effectively encoded into binary, making it incredibly difficult for a human user to find and interpret. Furthermore, within the visualizer itself, the specific person tracking functionality feature from previous iterations was removed. As every migration was represented as a dot, and there was no visual cue as to which dots pertained to the same person, the risk of identification was all but eliminated. In this way, Data Deciphered upheld its ethical responsibility to enforce data anonymity.

THE FIVE-STEP PROCESS

The following section is a detailed documentation of the five-step process employed by this thesis to formulate a JSON (JavaScript Object Notation) file output, consisting of 82,711 encoded migrations. In addition, JSON-encoded hub, studio, movie, role, people and skill files were also generated that contained the secondary data necessary to support the migrations. In comparison to the original Digital Workshops of the World database, which consisted of 13,217 migrations, this output was approximately six times the size. In this sense, the Data Deciphered database intended to paint a more accurate picture of the VFX industry, and also aimed to clarify some of the dominant trends that were present in the original visualization.

STEP ONE: SOURCING

The first step in amassing a collection of LinkedIn profiles was to construct a list of valid Internet URLs that were addressed to the public-facing identities of VFX workers. With this list as an input, a programmatic scraper was employed to automate the download process. With Google as the search engine of preference, specific LinkedIn-targeted queries were formulated to return the profiles of workers with histories at specific VFX studios. The chosen studios corresponded to the 'Big 15' houses identified as part of the Digital Workshops of the World project (Gurevitch & Spell, 2015). Grage also reaffirmed some studios in the selection by including them in his list of the top VFX studios of the 2000s (2015, p.225). These were:

- **Animal Logic**
- **Blue Sky Studios**
- **BUF Compagnie**
- **Digital Domain**
  (Top of the 2000s)
- **Double Negative**
  (Top of the 2000s)
- **Dreamworks Animation**
- **Framestore**
  (Top of the 2000s)
- **Industrial Light & Magic**
  (Top of the 2000s)
- **Moving Picture Company (MPC)**
  (Top of the 2000s)
- **Pixar Animation Studios**
- **Rising Sun Pictures**
- **Sony Pictures ImageWorks**
  (Top of the 2000s)
- **Trixter**

- **Walt Disney Animation Studios**
- **Weta Digital**

(Top of the 2000s)

For each of these studios, a search was submitted, typically returning approximately 600 results in each instance. Google's scraper is unfortunately limited to this figure for technical performance reasons. Naturally, this set consisted of workers with the highest-profile pages and therefore it did not give an adequate representation of the overall demographic and population at the individual studios. In order to gain diversity across the URLs per studio, the input queries were adjusted to specifically target certain combinations of roles. By only requesting 'Animators' and 'Artists' at studios, pages that contain higher frequencies of these terms were more likely to be included in the search results. In this way, a broader collective selection was achieved. In total, ten different role combinations were queried for each studio:

- ***"Animation, Research, Development"***
- ***"Artist, Color, Paint, Shade, Rotoscope"***
- ***"Camera, Layout, Lighting, Editor"***
- ***"Code, Software, Shading, Technical"***
- ***"Compositor, Creature, Environment"***
- ***"Model, Rigging, Texture"***
- ***"Motion Capture, Simulation, Systems, Massive, Stereoscopic"***
- ***"Pipeline, Production, Render, Shading"***
- ***"TD, CTO, Technical Director, Supervisor"***
- ***"Visual Effects, Effects, Visualization"***

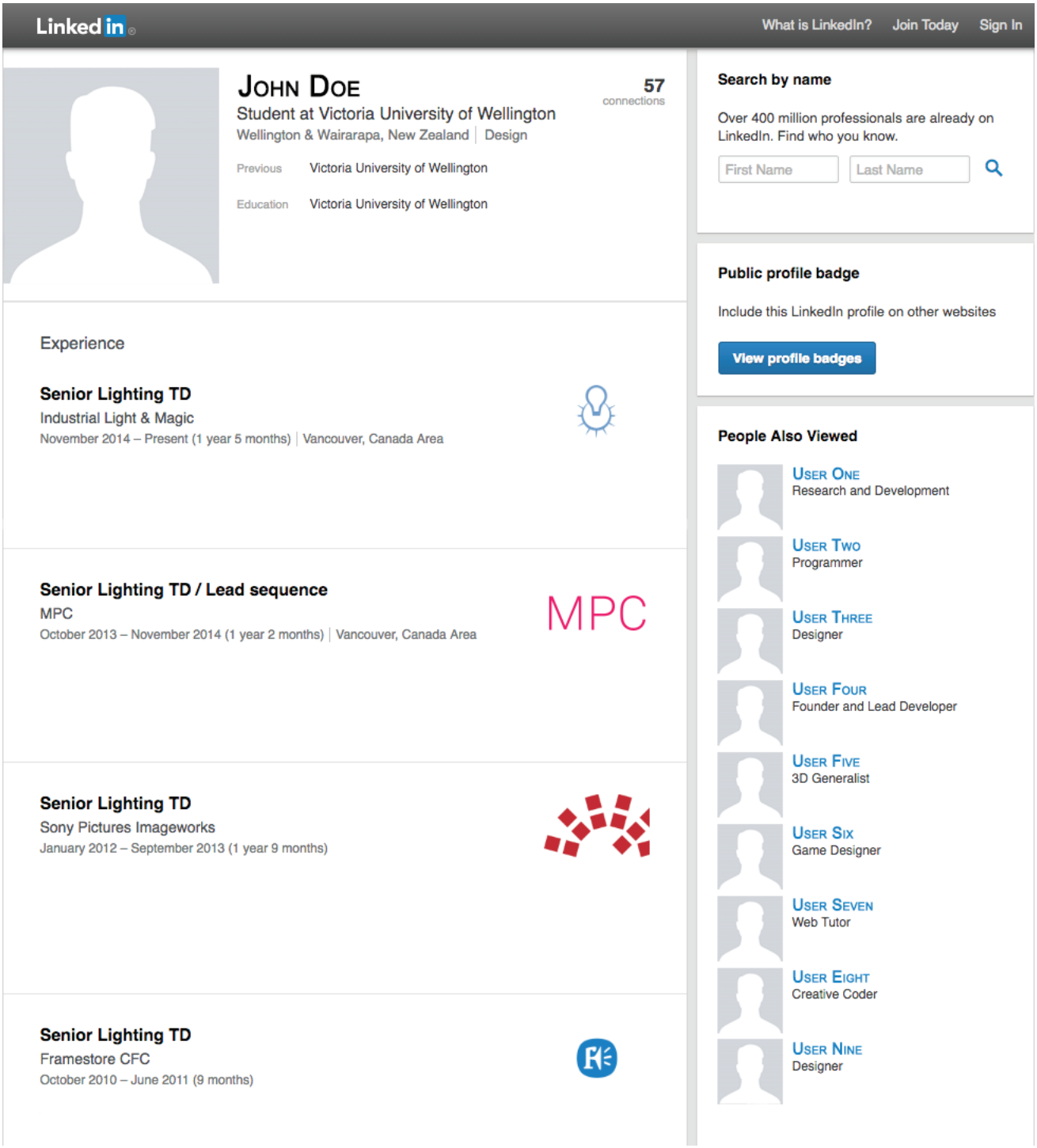
In order to harvest the URLs, a simple JavaScript program was written to export the URLs in the browser to a text file. Once an entire studio module was completed, the eleven respective URL sets were run through a custom-built Python aggregation script. This ensured that duplicate URLs were reduced to only one instance within the current studio context. Subsequently, the resultant set was further refined through reducing duplicates against a global URLs file, which consisted of all the currently known addresses across all studios. The final step in this process per studio was to add the remaining unique URLs to the said global collection.

The eventual output of this first phase was a single text file consisting of 52,068 addresses to various public VFX-related LinkedIn pages. In a pragmatic sense, the final set needed to be completely unique, in order to reduce the download time necessary to amass all profiles. Duplicate entries implied that the same information would have been unnecessarily pulled multiple times.

STEP TWO: DOWNLOADING

The second phase in the pipeline was the actual downloading of the pages from the gathered URLs. Each page could be downloaded as a single .html file and was essentially just structured text, so the file size per profile was negligible. While this entire process could be performed manually through the browser, for tasks of this magnitude it was simply not feasible for a human worker. As such, an automated Python scraper program was written that took a list of URLs as an input and wrote out the corresponding .html files to an output directory. This script was executed repeatedly upon studio-specific groups. For any files that failed to download initially, their corresponding URLs were written out to an error log, which was subsequently reissued to the program until all downloads passed.

Once the first iteration of downloads was complete for all studios, it became apparent that there was further opportunity to mine data. Embedded within each public LinkedIn page was a list of similar public-facing profiles that were associated by either shared company or occupation. Another Python utility script was written and executed upon each file, which added promising URLs, not present in the global URLs collection, to the second iteration of downloads. After the completion of



[FIGURE 3.2]

A Typical LinkedIn Public-Facing Profile. Note this person's 'Contracts' listed under the *Experience* section. Furthermore, associated profiles (used for the second iteration of downloads) can be observed under the *People Also Viewed* section. *LinkedIn* gives their members control over the information displayed on this public-facing page.

this phase, each studio had both a primary collection of files from the initial URL haul and a secondary collection, based upon extraction from the first.

It is important to note the technical challenges that needed to be overcome in order to amass the collection of files. LinkedIn is notorious for being a virtual Fort Knox in regards to web scraping. As there is high demand for its data, the platform employs various policing measures to thwart the attempts of any automated bots that intend to scrape its information en masse. Among these measures is IP monitoring, which is effectively able to determine and destroy the connection of any browser registering unusually high traffic. While it technically may be legal to download public-facing LinkedIn profiles, the question of whether this is achievable without being blocked is another matter entirely.

In order to reduce the chance of IP blocking from the server, it was necessary for the custom Python scraper to hide its identity. TOR (The Onion Router) is an application originally developed in the 1990s by the United States Naval Research Laboratory, as a means to protect American intelligence communications online. Today, it is publicly available and is able to disguise usage and location through a worldwide, volunteer-run relay system. When TOR is used to load a webpage, the request is bounced around the network before finally arriving at a destination router wherefrom the request is issued to the server. This makes it considerably more difficult to identify the IP address of users by anyone conducting traffic analysis or network surveillance. Using bdheath's pyTor library, which is freely available on Github, the Python scraper was effectively able to utilise the TOR system. As an additional layer of security, the Mechanize Python library, which enables programmatic replication of an Internet browser, offers the means to interchange the 'User Agent' string component of a webpage request. This text acts as a virtual passport, identifying to the server the type of the browser that is prompting the page. By rotating IP address and 'User Agent' header information through these two libraries, the custom Python utility was able to bypass LinkedIn's policing measures.

STEP THREE: VALIDATION

The third phase of validation ensured that the 'fluff' – irrelevant or superfluous profiles that were included in the downloaded pages – was removed from the core set. In order to do this, a series of passes were run over the entire directory, filtering out the undesirables based upon an assessed condition. However, certain filters required that secondary-data was collected and verified, and so it was appropriate that the studio, role, skill, people and movie databases were consolidated as a necessary prerequisite to this step.

STUDIO DATABASE

In order to construct a finite list of valid studios to assess the acceptability of contracts, an initial pass of the data occurred to determine studio frequency. A custom Python script was written to iterate through all of the files, extracting the various studio names and ordering them into an output text file, based upon their frequency. From the most common, studios were manually validated and researched, with their associated hub, industry, company size statistics, website and IMDB page entered into a Microsoft Excel spread sheet. This process was undertaken for all studios that appeared across at least twenty or more distinct profiles. It should be noted that not all studios with a high frequency were necessarily VFX studios. Others identified with the animation, film production, broadcast television, design and advertising industries respectively. These questionable candidates were accepted as part of the valid set due to the fact that VFX practices are a part of their production pipeline. To exclude them would imply that a migration to or from the company is invalid. This would have the detrimental effect of restricting the study to a narrow, VFX-rigid scope. In fact, the only significant industry types that were excluded outright were gaming companies and educational institutes, as the introduction of these organisations would have misaligned the VFX-centric quality of the database.

A disambiguation table was also constructed as an important component of the studio database. The dominant studios with multiple offices in different hubs were given distinct entries, in order to accommodate for their geographic diversity within the visualization. In this sense, 'MPC Singapore' was an entirely different entity to 'MPC Vancouver'. Additionally, as each studio entry was processed, mapping tables were manually created to correct common spelling mistakes in the raw data. A custom-built studio utility script was written to incorporate both the disambiguation and mapping tables. This simple program attempted to map an input name to a valid output studio name, or

raised a flag if none existed. This functionality was required by several of the filters in the validation phase.

ROLE DATABASE

The process for extracting and verifying role data was conceptually similar to the studio validation method. The only key point of difference was, due to the fact that there were a vast number of professions pertinent to the VFX industry, an automatic mapping phase was initially conducted to dramatically reduce the number of unassigned roles. Key terms, relating to each of the valid roles, were checked against the unassigned values to see if they occurred within. If so, the script would automatically assign the corresponding valid role to the unmapped value. The remaining values that were left over after the automatic mapping phase, were subsequently manually mapped. Ultimately, this mapping process resulted in a role utility script that was capable of parsing a given string into a valid role.

SKILL AND PEOPLE DATABASES

Constructing databases for both skills and people was a relatively trivial matter. Following an automated extraction of all skills in the profile set, a manual verification ensured that only VFX-related tags were accepted into a shortlist. Of this list, only the 500 most common skills were submitted to the final database, in order to avoid data excess. From this spread sheet, a corresponding people table was formed. This effectively acted as a bridge structure, by linking migrations to their neighbours and skills to migrations through representative ID values.

MOVIE DATABASE

The Movie database was somewhat disjoint from the other data-mining outputs in the sense that movies were not related to migrations in any way. However, they were desirable as they brought context to certain timestamps when viewed in conjunction with the migratory data. The overall collection contained 2,000 films, ranked in terms of the allocated production budget. In order to amass this set, the public-facing IMDB pages of the validated studios were programmatically downloaded with the previously used LinkedIn scraper. These pages contained company filmographies, grouped into production types. In the instance of a 'Special Effects' section, the underlying addresses to movie pages were added to the movie download list. Once captured, these movie profiles could be combed for relevant data. The thumbnail posters that were associated to the films were also gathered for incorporation into the visualizer.

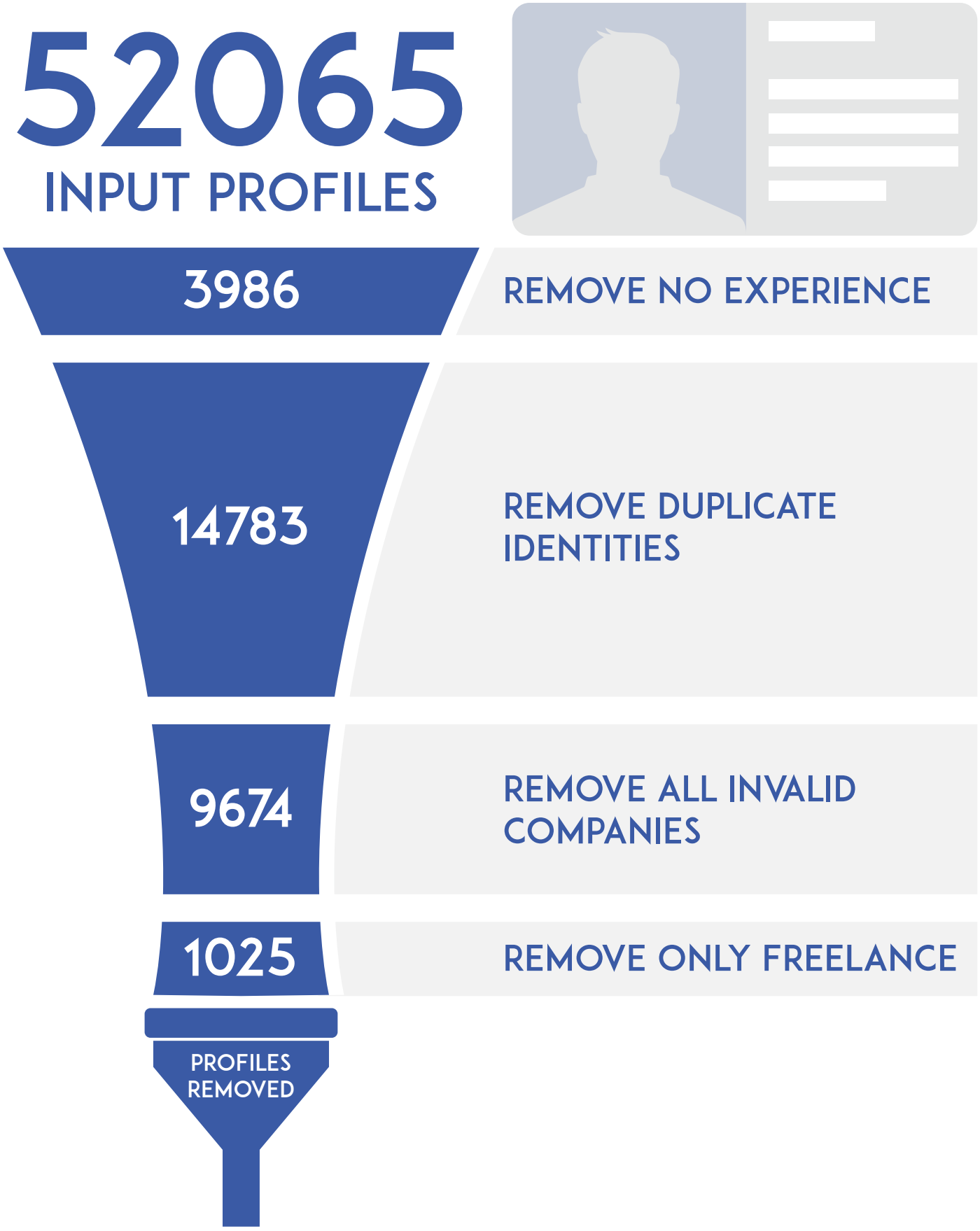
FILTERING THE DATA

Four filters were then run consecutively across the entire dataset in order to remove all unnecessary profiles. Essentially, these small scripts returned a Boolean condition that indicated whether any given profile passed or failed the assessment. They are as follows:

- The first filter excluded all profiles that did not have any record of experience. This may have been due to the fact that the member had not completed the contractual information section of their LinkedIn profile, or that they had not chosen to make their employment history publicly visible. Regardless, migrations could not be extracted from an empty record and so these profiles were deemed redundant.

- Any duplicate profiles were removed as part of the second filter. Despite efforts to eradicate duplicate addresses within the first phase, identical pages were still able to creep into the set of downloads. The primary reason for this was that LinkedIn pages could be common across various subdomains. For example, a person's profile in the United Kingdom (at uk.linkedin.com) differs from the web address of their New Zealand version (at nz.linkedin.com). However, duplicate data had no place in the eventual output, as this would have resulted in misleading, factually incorrect trends and information. In order to find and remove duplicate profiles, each page was turned into an identity, based upon its inherent name and current employer. Any profile that matched an already known identity was excommunicated from the dataset.

- The third filter used the studio mapping utility to determine whether or not a given LinkedIn page had any valid studios. This filter removed profiles that had tentative links to VFX, yet were not themselves directly related.



[FIGURE 3.3]

Illustration of the Data Validation Pipeline. Demonstrates the significance of each filter upon the reducing database.

- The fourth and final filter worked to remove professionals who had a singularly freelance employment history. Freelancing was technically treated as a valid studio, because a vast proportion of the industry regularly engages in independent work. The issue with this consideration was that it is nonsensical to geographically associate a freelance worker with a specific hub, as this type of self-directed work is not tied to any specific location. However, an informed guess can be made when it is known where the worker was employed previously or subsequently. Therefore, only profiles with at least one non-freelance contract were allowed to pass this filter.

STEP FOUR: EXTRACTION

The fourth phase of the data-mining pipeline was concerned with extracting migrations from the downloaded HTML files. The output of this section was the final migrations JSON file that could be directly applied to the visualizer. In utilising the mapper scripts, the raw text could be adapted into references to validated studio and role data objects. The implication of text that failed to be successfully mapped was that the contract to which the information pertained was invalid, and so the record was skipped. A custom Python extraction script was built to automate this process across all remaining files. The comma separated values (.csv) file that was yielded by this process was a direct aggregation of all valid contracts in the downloaded profiles.

This information was then imported into Microsoft Excel to aide in the conversion of contracts to migrations. As migrations were comprised of two neighbouring contracts, this process simply changed the structure of the data objects. The first contract of a person's employment history raised an issue however, as it had no previous contract with which to form a migration. As such, the design decision was made to effectively duplicate the original contract in order to form the migration and preserve the data in the visualization. This resulted in an initial circular, or 'non', migration for each person at his or her inaugural hub, where the migratory role was that of the first contract.

A further complication lay in the chronology of certain employment histories. While chronological order was generally preserved, in some instances this was not the case. The visualization required that there was no overlapping within careers and so an automated refinement of the start and end dates of offending contracts was carried out to ensure this property.

Initial attempts at visualising the data from this point resulted in unsavoury migrations, due to the sheer number of visual dots and the unstandardised duration of each migration. Some migrations would instantaneously conclude, as their contracts shared respective end and start dates, while others would extend for years, indicating a significant duration between VFX-related jobs in that person's career. In order to create a more visually compelling display, the chronological components of the migration data were massaged to fall between a minimum and maximum range. In this way, the duration of a migration was made to be anywhere between two to four months. It is important to note that this editing of dates did not propagate down the migration chain for a particular person. The lengthening or shortening of a migration bore no effect on its neighbours. In this way, the integrity of the data was preserved to the greatest extent possible.

STEP FIVE: ANALYSIS

The final phase of the data-mining pipeline was a pre-computation step to reduce the necessary number of calculations performed by the visualizer at runtime. The original Digital Workshops of the World overlay features, such as the professional composition and studio density graphics, were performance-heavy entities, as they needed their internal statistics to be re-generated upon every time update. As the database was fixed by this stage in the overall process, pre-computation of all possible statistics across time was plausible. In order to do this, a Python utility script was created to iterate across the contract database and formulate hub-specific tallies for each month of the visualization's timespan. In its various passes, this script yielded composition statistics for each hub, based upon specific roles and overall population. This data, now existing in text format, could subsequently be loaded and interpreted by the visualizer, whereby it effectively removed the overlay computation cost.

4

DEVELOPING THE VISUALIZATION

Visualising the amassed database to identify and present key trends and phenomena relevant to VFX migration has been a primary aim of the Data Deciphered project. Yet displaying this quantity of information in an accurate, efficient and comprehensible format has been an on-going challenge throughout the development of the system. The following section endeavours to document this evolution, describing and evaluating the key technical and design decisions made along the way. It provides rationalisation for visualization as the most suitable output medium for this project. Furthermore, it advocates for the display's responsibility to preserve data integrity and subsequently delves into the unique advantages posed by computer-based visualization. Finally, this section concludes with a critique of the final design and an overview of outstanding features scheduled for future development.

ON THE CASE FOR VISUALIZATION

Visualization has been established as convention when it comes to displaying sets of data in the age of social media and the 'internet of things'. In considering relational networks, sensor recordings or historical data, visualization is the de facto way to illustrate this information. However, when designing systems to assist with the presentation and accessibility of large databases, a valid question to ask is why is this the case?

In essence, visualization is the adaptation of raw data into an easily comprehensible graphical format (Wright, 2008, p.78). Edward Tufte, considered by many to be the grandfather of visualization, states that "the special power of graphics comes in the display of large data sets" and that "the most extensive [visualizations] place millions of bits of information on a single page before our eyes. No other method for the display of statistical information is so powerful" (2001, p.26). In this sense, it can be argued that the value of visualization lies in its innate ability to reduce dense amounts of quantitative data into easily comprehensible symbolic representations. Lev Manovich, professor of computer science at The Graduate Centre, City University of New York, suggests that another value of this reduction is that it more efficiently enables the exploration of the data by introducing the means of comparison (2015, p.33). Furthermore, infographics journalist, Alberto Cairo, lists organisation and correlation as further responsibilities of the graphic (2013, p.27). Together, these three cognitive processes work to construe meaning from within the dataset and it is through this ability to express relationships within the data that visualization becomes appropriate to all quantitative inquiry (Tufte, 2001, p.47). In this sense, the abstraction of raw numbers through graphics enables easier realisation of the data's implications. In his publication, 'The Visual Display of Quantitative Information', Tufte asserts that, as "much of the world these days is observed and assessed quantitatively", visual communication is far more effective than words at expressing statistical phenomena. Tufte launches a tirade against the frivolous use of graphics. To simply brandish icons, charts and graphs across an application without respect for their inherent communicative potential is a bastardisation of "graphical excellence". It wastes the tremendous communicative power of graphics to use them merely as decoration (Tufte, 2001, p.111). Therefore, in the context of the Data Deciphered project, visualization is deemed as an appropriate output medium. This is due to the need for the clear expression of the amassed database, despite its significant size, and the identification and presentation of its inherent relationships.

Visualization also boasts psychological advantages, specifically regarding the aiding of cognition and memory retention. Cairo writes that the primary goal of a graphic is "to be a tool for your eyes and brain to perceive what lies beyond their natural reach" (2013, p.10). He elaborates by suggesting that humans, through the act of perception, are constantly imposing hierarchies on their surrounding environments for the purpose of assembling order out of the natural chaos. "The brain always tries to close the distance between observed phenomena and knowledge or wisdom that can help us survive. This is what cognition means" (Cairo, 2013, p.17). Furthermore, Richard Wright notes the power of human vision in its intuitive ability to "compare small and large scale features at the same time" and discern artefacts or irregularities in the data itself (2008, p.79). He rejects the notion that 'pictures do not prove anything' with his reasoning that the suggestion of apparent relationships through visualization can be later confirmed through exact statistical inquiry. Therefore, for the information architect, visualization presents a unique opportunity through which to suggest a semblance of order through visual cues. This guides the viewer's initial comprehension and resultant interpretation of the data. Additionally, visualization proves superior to other display types, such as tables, in its ability to assist with information retention. Regarded as a great inventor of modern graphical designs, William Playfair (1759 – 1823), a Scottish political economist, praised charts for the 'simple' and 'distinct' ideas that they portray. Playfair noted that one of the key issues with employing spreadsheets as presentation media is that, "information, that is imperfectly acquired, is generally as imperfectly retained; and a man who has carefully investigated a printed table, finds, when done, that he has only a very faint and partial idea of what he has read" (Tufte, 2001, p.32). Playfair realised that the advantage of the graphical interface is that it can symbolically portray complex ideas in simple shapes. These shapes leave "distinct impressions", which remain unimpaired due to the simplicity and completeness of the mental picture. In this way, visualization functions as a cognitive usher, by directing the reading of the content, and also as a photograph, with its ability to imprint in the mind of its viewer.

There has been much debate concerning whether

the medium of visualization should be regarded as a purely scientific discipline, or one that lends itself to creativity and art. In exploring this, it is helpful to state the fundamental requirement of visualization. Wright defines this as the "[algorithmic derivation of a] sensory expression from the structures implicit in digital data, even when, and especially when that expression takes us far from the realm of computer code" (2008, p.86). While this statement is exceptionally broad, it does suggest that visualization should utilise technical process to achieve an artistic outcome and therefore functions as a hybrid medium. Wright notes that the fundamental difference between an image and a visualization is that the former is constructed upon known conclusion, whereas the latter is about establishing conclusion based upon connections between the data attributes. "A visualization is not a representation but a means to a representation" (Wright, 2008, p.81). In this way, it is implied that, due to its heightened functionality the medium should be regarded as a scientific method. Writers such as Colin Ware have sought to affirm visualization as such by grounding it in the domains of physiology, human perception and cognitive psychology (2004). These scientific advocates for visualization seek to establish a taxonomy of perceptual techniques that can assist designers in creating the most optimal expression of a dataset. In harnessing the 'automatic processing' stage of human vision through the considerate use of light, pattern, orientation and movement, it is believed that a universal visual language can be established, which is subconsciously intuitive to the reader. The ultimate goal of this is 'computational offloading', which is a description of how innate perception can be capitalised upon to communicate comparison and relationship without the need for explicit calculation on behalf of the user (Wright, 2008, p.82). While this scientific inquiry into visualization has high potential for enhanced readability it is guilty of streamlining the medium into a standardised aesthetic with disregard for artistic liberty. Visualization scientist Chaomei Chen believes that the medium is more appropriately defined as an art than as a science. He notes that there currently exists no template for objectively assessing the quality of a given visualization, due to the fact that viewer interpretation can vary dramatically. Furthermore, in the absence of standardised design conventions, decisions pertaining to the visual encoding of information are highly subjective (Chen, 2006, p.1-2). In addition to this, Wright puts forward the notion of 'non-cognitive visualizations', which "move so far from their source data that the data disappears from relevance entirely" (2008, p.84). In one example of critical design, known as Lung: Slave Labour, a database is effectively given breath as a poignant memento of the foreign labourers that were forced to work under Nazi rule during World War II (Harwood, 2005). Users of the installation are able to inspect the age, gender and height of worker profiles that have been extracted from Nazi records. As they do so, 'Lung' is able to calculate lung capacity and produce a symbolic breath of air for each labourer through its associated speaker system. This example clearly demonstrates a poetic dimension that exists over and above the scientific technicalities of the database. In this way, visualization is a hybrid tool that has relevance to both artistic and scientific disciplines. Depending upon how the designer intends for the viewer to interpret the data, visualization offers both approaches and aesthetics to achieve contrasting effects.

RESPONSIBILITIES OF VISUALIZATION

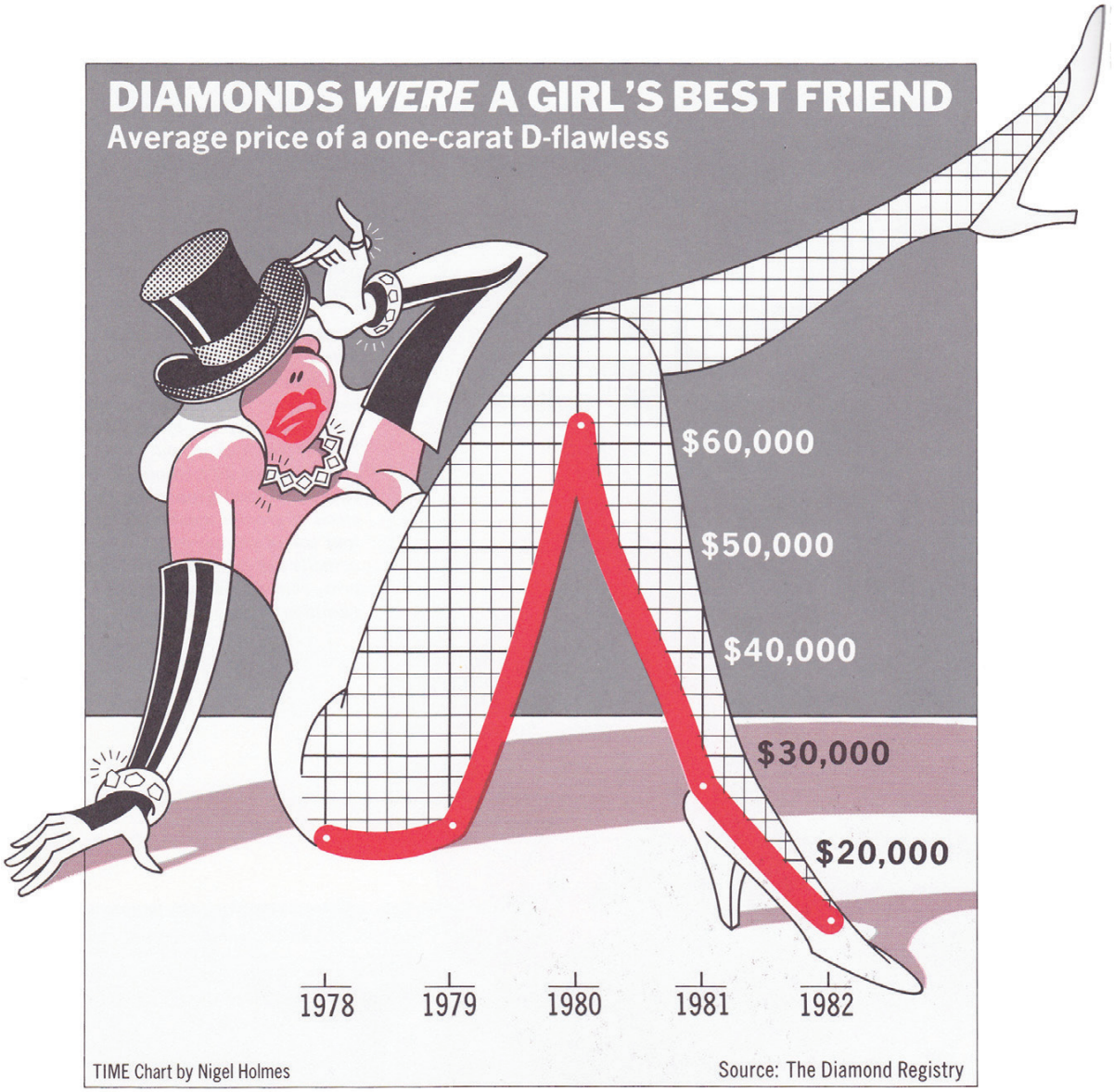
Adherence to 'data integrity' is fundamentally important in conveying an honest representation of the database. Integrity, in regards to visualization, implies that the accurate expression of quantitative information is paramount. It follows therefore that the presentation devices will be truthful and won't attempt to distort or selectively portray the data to serve political ends. Tufte is famously quoted as writing, "Above all else, show the data. This principle is the basis for the theory of data graphics" (2001, p.92). In his vendetta against 'amateurish' design, Tufte is quick to condemn any attempt to misrepresent a graphic, through distortion of the data measures, over-decoration or otherwise. In fact, he proposes various scientific models and equations, based upon the relative size of the graphic's 'ink', to be able to quantify exaggeration and ascertain his point. Tufte challenges designers that "every bit of ink on a graphic requires reason. And nearly always that reason should be that it presents new information"

[FIGURE 4.1]

Opposite. *Lung: Slave Labour.* An example of a non-cognitive vizualisation. From Harwood, G. (2005). Lung: Slave Labour [Installation]. In: "Lung: Slave Labour", ZKM, Karslruhe, Germany, 3/20/2005 - 10/3/2005.


SONY
YAMAHA

(2001, p.96). He concedes that adornment can have value in editorialising a graphic but maintains that it is irresponsible of designers to manipulate the data measures in order to make a comment or fit an aesthetic scheme. This perspective contrasts with the work of Nigel Holmes, a British graphic designer, who has a comparatively looser and more light-hearted approach to chart design. Holmes is renowned for the bold statements made by his heavily illustrated graphical work. He believes that the introduction of humour, through characterisation and colour, can create more attractive, impressionable and therefore memorable designs (Holmes, 1984). In his 1990 publication, 'Envisioning Information', Tufte brought the debate to the forefront of design discourse by labelling one of Holmes' graphics as 'chartjunk'. "Graphics do not become attractive and interesting through the addition of ornamental hatching and false perspective to a few bars. Chartjunk can turn bores into disasters, but it can never rescue a thin data set" (Tufte, 2001, p.121). Tufte's concern, in relation to Holmes' work, is that the beatification of visualization immediately shifts the focus from the subject to the decoration. In doing so, it makes the reading of the data more difficult and introduces the possibility of misinterpretation. Data Deciphered respects these graphical rulings as a means of optimizing data integrity. This influence is evident in the minimalistic and functional visual design of the final output.



[FIGURE 4.2]

Diamonds Were A Girl's Best Friend. An example of 'Chartjunk', where visual decoration detracts from the data itself. From Holmes, N. (1983). Time Magazine.

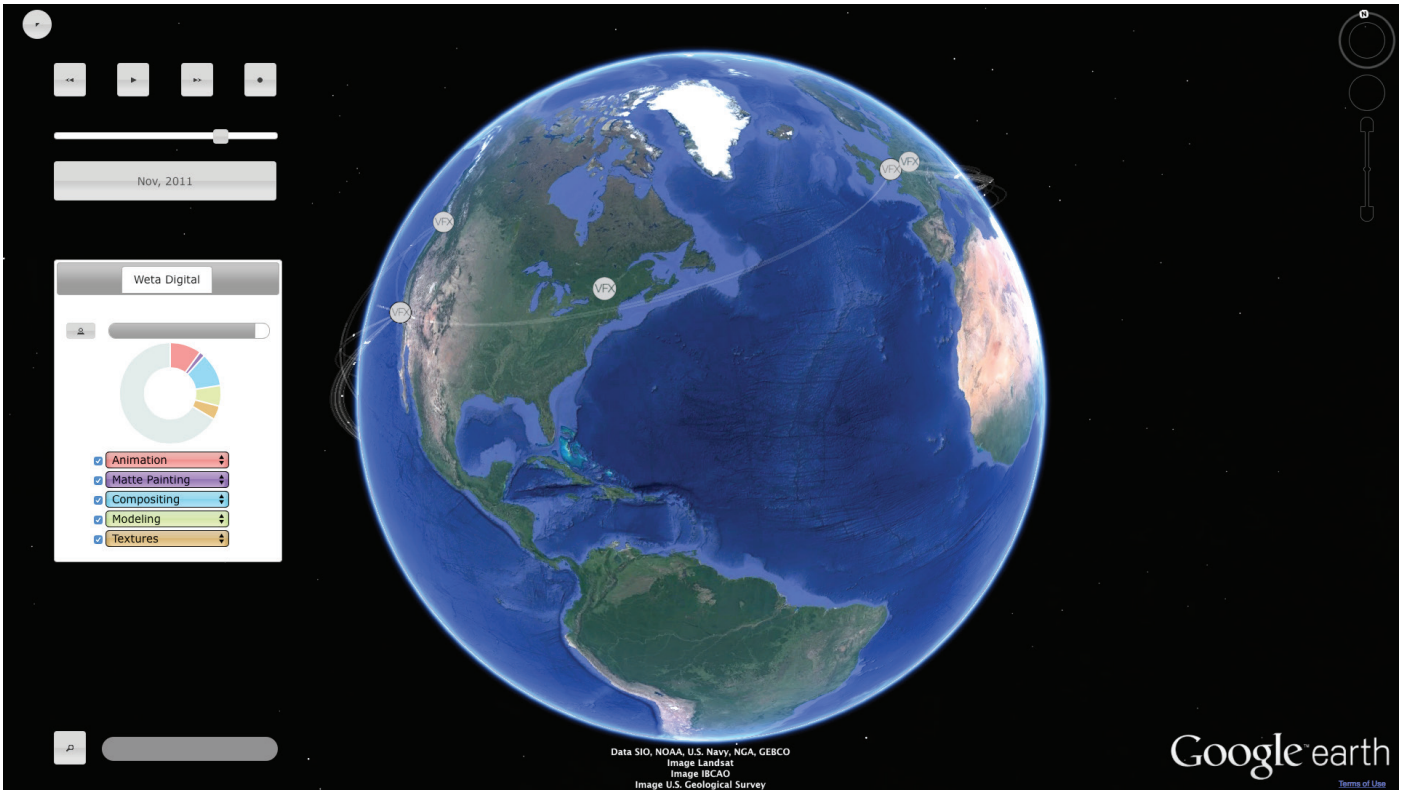
Visualization also has the responsibility of being function-led. There exists a common misconception that visualization is simply an artistic rendering of data. While a strong aesthetic is indeed important for successful visualization, it is important to realise that graphics are primarily a tool and therefore their design should adhere to the function that they are built to perform. Cairo writes, "If you accept that visualization is, above all, a tool, you are implicitly accepting that the discipline it belongs to is not just art, but functional art, something that achieves beauty not through the subjective, freely wandering self-expression of the painter or sculptor, but through the careful and restrained tinkering of the engineer" (2013, p.23). If the goal of visualization is to provide an interface to the database, then it follows that the visual design should accentuate and direct the eye towards the key components or displays that support find, filter and comparison operations. Furthermore, it is important to realise that, in adhering to a function-led manifesto, the data itself should not adopt any arbitrary form, but instead should be constrained by the dimensionality and type of the entries (Cairo, 2013, p.36). For example, migration data is comprised of origin and destination locations and travel duration. It is trivariate and is therefore not suited to a bar graph or scatterplot that only accommodates two variables. Furthermore, since the type of the data is location, function dictates that it is best suited to some form of map. The adage 'form follows function' is particularly suited to data visualization and it is the responsibility of the information architect to ensure that the aesthetic of the work highlights and supports the functional elements.

THE MERITS OF COMPUTER VISUALIZATION

The works of Tufte, Holmes and Cairo predominantly utilised traditional print-based media, however Data Deciphered has been developed in, and is intended for, a virtual environment. Despite their ontological differences, visualizations constructed in three dimensions are still bound to the design principles of their 2D counterparts. The primary difference is the notion of animation. In print, graphics exist as still media and therefore all information, labels and legends need to be allocated space on the page in order to be included. The notion of spatial and temporal animation, synthesised by computers, has led to the introduction of interactive user interface elements that can react to visualization state (Manovich, 2013, p.289-296). Practically, this implies potentially infinite use of layout space, as visualization components can be toggled on and off to met the user's current inquiry. Furthermore, the data itself is no longer required to exist concurrently but instead can be separated across frames, each of which can be recalled instantly upon request. The renowned visualization Google Earth was one of the first commercial applications of this division, with its segmentation of frames based upon chronology. Data Deciphered employs an identical 'time slider' mechanism and practically this enables the 80,000 database migrations to be efficiently rendered and appropriately and clearly displayed. However, for each new benefit that the computer grants visualization, there is a corresponding consideration to be handled. For any computer application, the question of system interaction is inevitable. In 1996, Ben Shneiderman, distinguished computer science professor at the University of Maryland, College Park, published an information seeking mantra that informs user experience design. Shneiderman claims that in order to heighten usability, information visualizations should support the features of: overview, zoom, filter, details-on-demand, relatable data elements, history or 'undo', and extraction of findings (1996). These seven operations were employed as a functional benchmark within Data Deciphered, in order to iteratively develop the system and interface. Even though the computer affords far greater visualization and data exploration potential than print media ever could, this is tempered with the added challenge of designing an intuitive interaction experience.

THE DEVELOPMENT PROCESS

Having justified visualization as an appropriate and necessary output medium, it is now fitting to discuss the development of the Digital Workshops of the World application, across the three prototyping phases. As I was one of the lead designers on both the Google Earth and WebGL variants, I am aware of the technical and aesthetic development challenges and qualified in passing judgement in regards to these prototypes. These three project milestones encapsulate the body of practical work devoted to this thesis. The following section endeavours to provide a summary for each iteration and subsequently, in accordance with Shneiderman's task taxonomy, investigates the



[FIGURE 4.3]

Home Screen of the Google Earth Plugin. Iteration one of the Digital Workshops of the World project was created via the Google Earth API in the Google Chrome browser.

evolution of key visualization features. As comparisons are frequently made between the prototypes, it should be known that each was developed on the same machine – a late 2011, Intel i7 Macbook Pro with 4GBs RAM and an AMD Radeon 6750M graphics card – and therefore the hardware running the programs remains constant. Finally, this section concludes with an achievement analysis and an outline of improvements scheduled for future release.

GOOGLE EARTH

The first realisation of Digital Workshops of the World was conceived within the now-deprecated Google Earth browser plugin. This platform was deemed appropriate due to the target audience's familiarity with this software as a visualization tool. Furthermore, as he plugin natively existed in a 3D scene, a large portion of the mathematics involved with spherical positioning and the translation between latitude and longitude and Cartesian coordinates was provided. Visually, this prototype exhibited Google Earth's default satellite texturing, which supports level of detail redefinition. Initially, this aesthetic was considered desirable, due to the project's clear link to Google Earth in both form and function. However, upon reflection, the high detail offered by the map was detrimental to overlaid elements that required visual emphasis. Furthermore, the mip-map functionality was largely redundant, as it was discovered that a distant overview provided the best perspective from which to observe the data.

Atop the orbit-able 3D view, the application hosted a user-interface that supported search and time control functionality. In addition to this, a non-diegetic pie chart and percentage bar displayed composition and density information respectively, per studio. The currently inspected studio could be changed by clicking on its landmark within the visualization, or by requesting it specifically through the search bar. This notion of studio-specific information was unique to this prototype, as, despite having high potential for dense datasets, the insufficient size of the initial database meant that the majority of studio's had minimal chart activity for the majority of the visualization. Subsequent iterations solved this issue by aggregating studio data into more intuitive hub clusters, which were indicative of the relative size and make-up of VFX regions.

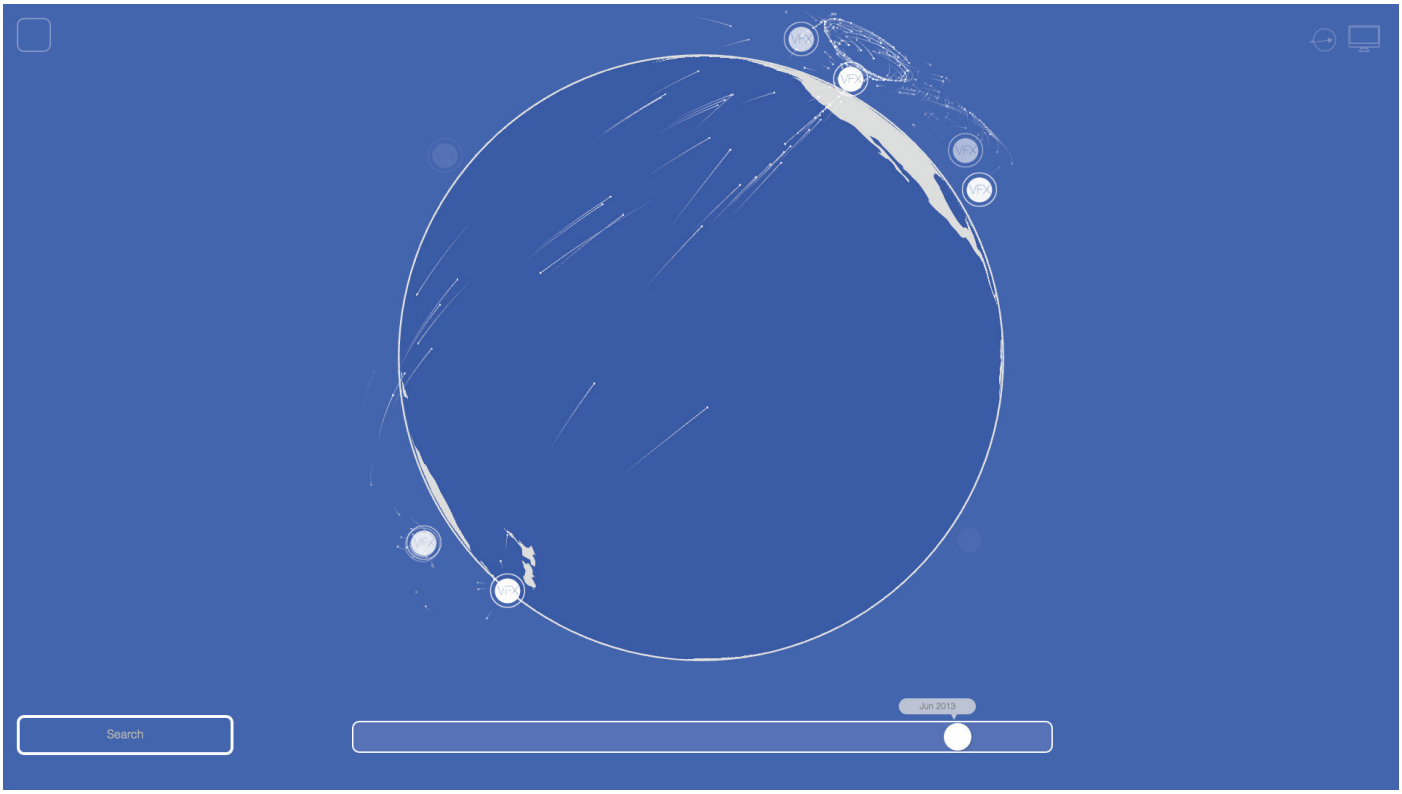
The performance during startup and runtime was

[FIGURE 4.4]

24 Hour European Flight Traffic Vizualisation. Produced by data vizualisation firm 422 South. Designed works such as this were precedent pieces for the Digital Workshops of the World project. From NATS. [opal]. (2014, Mar 9). 24 Hour European Flight Traffic Visualization. [Video File].

subject to undesirable lag, due to the parsing of data and the overhead of the plugin. During the loading process, Google Earth had the unique pitfall of requiring that the input data (in JSON format) be parsed to KML (Keyhole Markup Language) format. KML, Google's custom file specification for geographic data, is the standard across the Maps and Earth applications for the encoding of user-specified landmarks, lines and timestamps. Unfortunately, in order to programmatically apply large amounts of external data, the information first had to be translated into KML in order for it to be palatable by the system. Practically, this meant that an extra step had to be undertaken during the loading phase, which added extra duration to an already labouring process. The runtime performance of the application was also substandard due to the significant quantity of data applied to an already heavy plugin. The level of detail rendering within Google Earth, requires a vast texture library that is sampled based upon the user's current position and proximity to the globe. Furthermore, within city confines, the 3D perspective rendering of buildings and other landmarks becomes apparent. When one considers this functionality and the requirement of having to simultaneously load and unload certain assets upon every camera adjustment, it is conceivable as to why the application may struggle. For the purposes of Digital Workshops of the World, this rendering, while impressive, was unnecessary due to the redundancy of zoom functionality.

The limited range of visual landmarks to signify data also proved to be a disadvantage of the system. The notion of representing migrations as flights that orbit the globe over time was originally inspired by works such as the 24 hour European flight traffic visualization, which was created by data visualization design firm 422 South (NATS, 2014). This design was deemed elegant, intuitive and powerful in its ability to highlight regions and routes of dense activity. Digital Workshops of the World intended to port this aesthetic to the Google Earth plugin; however, the range of drawing tools within the application was severely restricted. Instead of providing primitives that enable designers to combine into more complicated and tailored shapes, the plugin provided its own set of premade visualization components and restricted use to only this set. As such, to illustrate migrations, the default trail renderer was selected. Unfortunately, this had the undesirable effect of rendering the entire path, between origin and destination, in reduced opacity, while also drawing the current position of the migration as a highlighted segment along the line. Compared to the



[FIGURE 4.5]

Home Screen of the WebGL Application. Iteration two of the Digital Workshops of the World project was created via the THREE.JS library in the Google Chrome browser.

[FIGURE 4.6]

Home Screen of the Data Deciphered Unity5 Application. Iteration three of the Digital Workshops of the World project was created via the Unity engine and built for desktop deployment.

precedent piece, this design was resultantly rougher and did not achieve the desired elegancy. Furthermore, this prototype did not adequately display the circular, internal migrations around their corresponding hubs. Rather, due to the naïve default implementation, the migrations rendered as a vertical line above each hub. This did nothing in terms of presenting regional activity and proved to be a major flaw as later prototypes proved that the majority of VFX migration occurs internally. In light of these points, while the Google Earth variant provided a solid proof of concept for the Digital Workshops of the World project, it left much to be desired in technical, informational and visual respects.

WEBGL

The second iteration of the Digital Workshops of the World project was also built for the Chrome web browser but had the advantage of being independent from the Google Earth plugin that had limited performance and design. This variant was built as a traditional webpage, utilising HTML5, CSS3 and JavaScript for content, style and functionality respectively. The prominent Three.JS library was incorporated as third-party software that enabled efficient 3D browser rendering and handled advanced mathematical operations. Stylistically, this prototype boasted a customised minimalistic aesthetic that exhibited thin and crisp elements in navy blue and white. In terms of data, a repetition of the initial gathering process was undertaken with the goal of amassing a larger set that additionally incorporated animation studios. The primary aims of this version were to demonstrate this enhanced database and to provide exploration of this data by supporting filtering operations and statistical analysis. In order to achieve this, the React.JS framework was employed to quickly prototype user interface elements. A filter menu was built to support the display of migrations based upon profession, region or past employment history. Furthermore, this menu also enabled the rendering of diegetic hub overlays, which could report upon regional density, composition and movie production. In an attempt to heighten the cinematic quality of the system, additional toggles were provided to enable automatic planetary rotation and to adjust the brightness and contrast for projector displays.

While this variant of the application contained definite improvements from its predecessor, there were still inherent incomplete aspects that needed to be rectified. In Inside VFX, Pierre Grage notes the rise of Asia as a VFX hotspot in the new millennium (2013, p.197-212). Unfortunately, this observance, which was particularly relevant to the Chinese and Indian markets, was not reflected in the system. Grage further suggests that the Asian workforce is underrepresented in film credits, which provides rationale as to why this data was missing. As the majority of these workers are junior artists that perform menial tasks such as rotoscoping and matchmoving, their contribution is not deemed significant enough to warrant a claim to the limited credit space. In terms of the data accumulation process, which was performed by cross-referencing IMDB film credits, this presented a misrepresentation that led to a significant gap in the database. In order to provide a holistic picture of the global VFX industry in future iterations, oriental activity would need to be represented.

The lengthy loading time of the database into the visualization was a further inadequacy that was amplified due to the online environment. As the application was hosted upon my personal website for a period of time, users were able to access this tool from anywhere in the world. However, hosting this visualization externally incurred a heavy download cost to the client machine. Despite efforts to compress the data into a file size that was more manageable, the webpage required minutes to properly load in some cases. The workaround of rebranding the tool as a downloadable application that could be executed locally was enacted. However, the lack of a graphical user interface to execute the program was a significant usability drawback. In order to run the visualization, users had to first create a Python server, which could subsequently load the required assets within the browser. Practically, this involved opening the command prompt and entering a few lines of shell script. Even though to the experienced user this may seem like a negligible task, for the uninformed this can be overwhelming and discouraging. In order to improve accessibility, a system-integrated executable needed to be created that had the ability to perform all launch processes upon a single click.

UNITY5

The final realisation of the Digital Workshops of the World project was assembled in the increasingly prevalent Unity5 game engine. Unity has the reputation of being the industry leader amongst modern game development platforms and therefore it was a powerful,

well-supported and safe option for the advancement of the iterative application. Data Deciphered utilised the free licence of version 5.1.1 and with this came a variety of advantages. The most appealing of these was Unity's underlying optimised graphics software. Traditionally, Digital Workshops of the World had an intensive rendering requirement and with the higher data load placed upon this version it needed to maintain acceptable performance. Another advantage was the inherent facilitation of deployment to a variety of platforms and operating systems. Previously, the need to repeat code in browser-specific format had proven to be a hassle, so the notion that Unity could automatically apply this translation behind the scenes was appealing. Additionally, a fully redesigned user interface system was incorporated within the platform that enabled the simple creation and placement of responsive elements. For specific development hurdles that could be easily overcome with third-party software, an integrated asset store was on hand with thousands of premade scripts and features. The online documentation was well maintained with ample detail, which allowed for the fast identification and resolution of bugs during development. As shaders had been previously employed to achieve complex graphical effects and accelerated calculations, it was reassuring to note full integration of these scripts with the rest of the platform. Furthermore, all of the pre-packaged software was open source and could therefore be manually inspected for learning purposes. For all of these reasons, Unity5 justified itself as a desirable candidate for the advancement of the Digital Workshops of the World project.

There were several new features employed within the Data Deciphered iteration that aimed to enhance the overall usability of the system. In order to present a holistic and unobstructed picture of the data, migration type toggles that could switch between internal or external migrations and a 3D-to-2D projection shift were added. In order to accommodate for the need of greater exploration within the improved dataset, an advanced filter system was established that enabled the visualization of migrations based upon current location, past employment history, occupation and acquired skillset. Interesting observations that were made through this component could be saved to file and subsequently reloaded at a later time, or within another session, by selecting the record through the newly incorporated radial menu system. Finally, to aid users in mastering these novel elements, a tutorial presentation was included that provided graphics and text to illustrate each concept.

A unique aspect regarding development within the Unity5 engine was that all custom-built software required integration with the underlying system. Whereas previously, I had the added responsibility of code architecture, in this version functionality was added to 'game objects' via scripts that inherited from a base Unity class. In terms of design, this presented a hybrid construction, as, while scripts were still necessary for customised behaviour, the ability to manipulate game objects, in terms of their values and hierarchies, was also needed. In this sense, discovering the best way to overcome specific visualization problems required both a designer's understanding of the 3D workspace and a computer scientist's knowledge of code processes and architecture.

OVERVIEW AND ZOOM

The first two principles of Shneiderman's 'Information Seeking Mantra' relate to the navigability of the system. The default projection of the application throughout the first two prototypes was 3D with orbit controls that allowed the user to rotate, zoom and tilt about the central globe. While this offered an engaging perspective, one of the key criticisms in regards to overview was that users could not simultaneously observe activity on opposite sides of the world, due to the occlusion it posed. To remedy this, the Unity5 variant offered switchable projection modes. For the first time, users could now unwrap the globe to a 2D map and utilise pan controls to infinitely scroll across a looping world texture. In this state, all events are displayed concurrently. This heightened visibility was not without compromise however, as fitting all of this information onto the screen required that the camera zoom out. In this way, the 3D view still retained its value in providing a closer and more targeted view of the data.

In addition to magnification, 'zooming' can also be considered as focusing upon a specific data element. In this sense, the 'fly to' functionality offered in relation to studios and movies can be regarded as a 'zoom'. Across all three prototypes, when the user selected a studio from the search results, the camera would sweep around the globe until it arrived at the associated hub. Furthermore, with the introduction of movies in the second iteration came the chronological 'fly to' functionality to reflect film release. Both of these 'zooms' were desirable as they heightened system feedback and also visually reinforced the information text display.

[FIGURE 4.7]

Data Deciphered 2D Projection Mode. This feature was a new addition to the third iteration and offers a perspective on all data simultaneously.

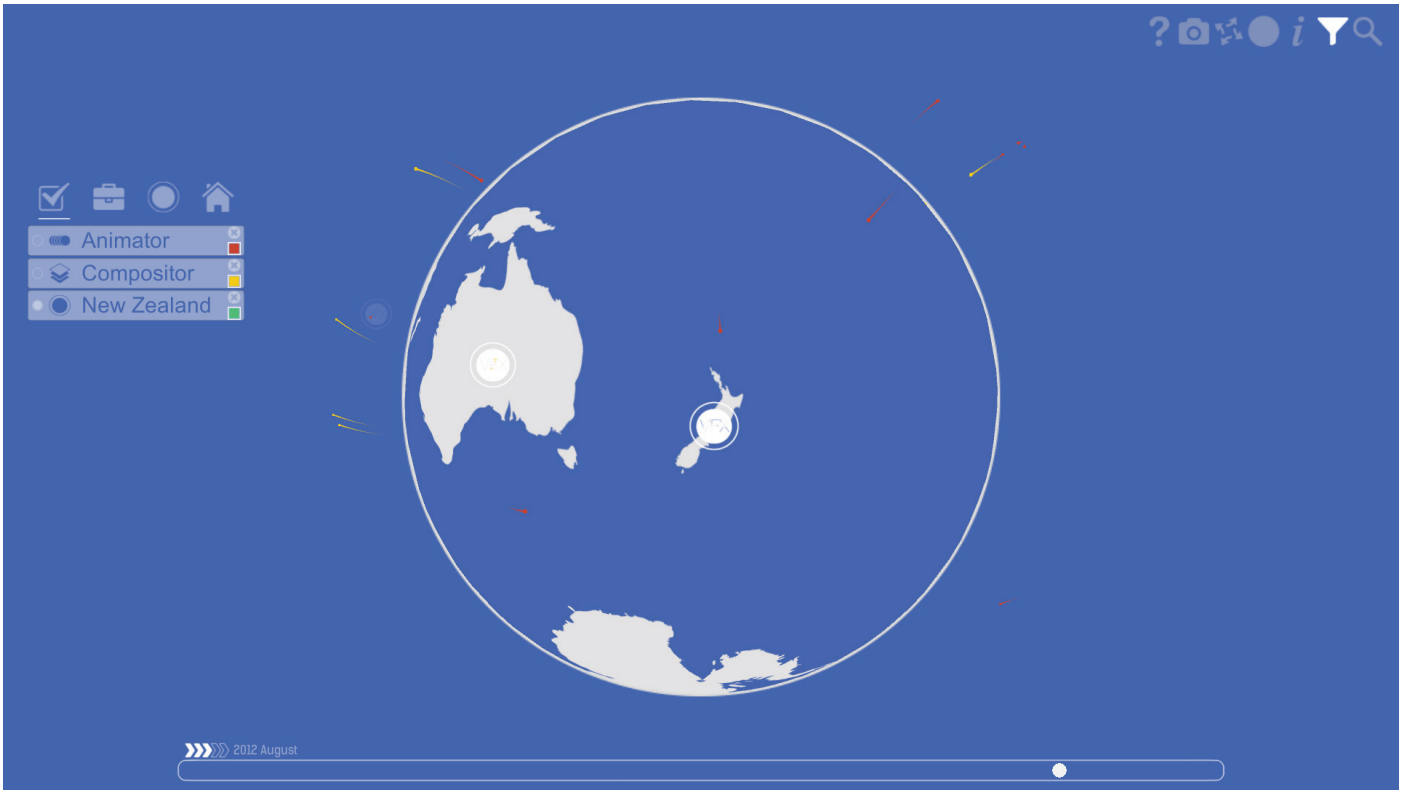
In a technical sense, a well-performing overview will enable instant feedback upon navigational changes and will not suffer from frame rate drops. Unquestionably, the single greatest technical challenge in regards to performance throughout the development of Digital Workshops of the World was that of updating thousands of data entries every frame in real time. As migrations were represented by two elements – a dot and a trail – several positional recalculations were required per instance. The Google Earth iteration had a comparatively smaller database and therefore was able to sidestep optimisation requirements. However, the subsequent iterations needed to employ these measures in order to maintain acceptable usability. Due to the different deployment contexts of the WebGL and Unity5 prototypes – web and desktop respectively – the technical workarounds on offer were distinct, yet shared common concepts. The first optimisation concerned the identification of migrations drawable for the current frame. A naïve implementation would simply iterate through all entries, sorting those that were active from the rest. For small collections, this may be feasible but the database of Digital Workshops of the World was too significant. As such, it became apparent that migrations needed to be sorted into a data structure that enabled efficient retrieval. In the loading phase, the system created a 'draw order' hash table that assigned migrations into reduced groups based upon their chronological period of activity. Having performed this pre-computation, the system could simply request the collection from the table that was identified by the hash of the current time. In terms of data sorting and retrieval, this structure greatly enhanced performance. A second improvement was concerned with the optimisation of visual migration components. Initially, the update algorithm would maintain a pool of all migrations, recalculating for those that were visible during the current frame and hiding the remainders. The issue with this was that, despite being hidden, the superfluous migrations were still occupying system resources and explicit checks needed to occur in order to ascertain their current state. As the total number of necessary migrations was equal to the display load of the densest frame, the improvement was made to limit the number of dots and trails to this figure. Additionally, in divorcing each data entity from its corresponding visual representation, the system could now manage the temporarily assignment of an available visual to a migration data object. This link would exist for the duration of that data's display, after which the visual component would be released back into the 'pool' whereupon it would become available

[FIGURE 4.8]

Data Deciphered Tutorial System. The increased functitonality of the third iteration provided rationale for a tutorial system to explain features to first time users.

for reassignment. This software mechanism proved significantly more efficient, but did not eradicate the frame rate dilemma. Another improvement undertaken was the replacement of the 3D spheres that represented the migration heads with 2D sprites. This dimension reduction heavily reduced the number of points that required rendering on the graphics processing unit (GPU) and afforded a performance increase. Additionally, in researching graphics processes, it became apparent that modern GPUs are significantly more powerful than CPUs (central processing unit), where software is primarily executed. As such, there is a greater efficiency in moving regularly repeated code processes to the graphics card, by way of a shader program. In the WebGL prototype, the reassignment of the positional update code to the GPU yielded the most significant performance increase. This same strategy was therefore attempted in the Unity5 version, however, due to the game engine's inherent graphical optimisation, this tampering seemed to interfere and detriment the overall performance. Resultantly, the code was reverted. The final major frame rate improvement came with the introduction of spherical mathematics in the Unity5 iteration. Previously, the system had interpolated latitude and longitude coordinates, which were effectively converted to spherical representations through third-party libraries. This implied an additional conversion step during each migration update. However, in the final instalment, these geographic coordinates were immediately parsed to their spherical counterparts and all subsequent calculations occurred in this domain. Ultimately, this had the effect of simplifying the dependant algorithms and speeding up the update process. While the current version of Digital Workshops of the World is still subject to frame rate drops in highly active sections, the overall process of optimisation has resulted in acceptable performance for thousands of migrations updating simultaneously.

The tutorial and tooltip elements were incorporated into the visualization in order to provide an instructional overview of the system for first time users. With the development of additional features in the later iterations, it became apparent that with the increased functionality came increased complexity. The Unity5 prototype sought to remedy this with a tutorial system that could be triggered from the main screen. However, despite instructional illustrations, the large amount of explanatory text discouraged many participants in user testing, leaving them uninformed. Furthermore, there was an apparent disconnect between the tutorial, which appeared as an overlay, and the visualization.



[FIGURE 4.9]

Data Deciphered Filter System. The current settings display Animators and Compositors who are arriving in or leaving from New Zealand. Note the option to manually apply color to differentiate a particular type of migration.

The inability to interact with the components as the user was discovering them was a critical flaw in this system. In order to provide a more diegetic interface, the Unity5 iteration reverted to the tooltip model of its predecessor. This means of instruction was advantageous because tooltips only appeared on hover events, making them non-intrusive and specific to the user's current focus. Furthermore, the elements were implicit to the system, which preserved user connection with the application.

FILTER AND RELATE

The clear portrayal of the influence of shared relationships upon the dataset as a whole is the primary aim of this application. Shneiderman asserts that the ability for queries to dynamically filter the database is one of the key tenets of visualization (1996, p.4-6). This has been the functional goal of the filter machine, which is specific to the Unity5 prototype and necessary because of the increased quantity of data. Due to the meaning behind the various filters, this mechanism has greater implications than the mere toggling of layer visibility. To turn on a filter is to select and display all entities that identify with the associated property. In this way, to apply a filter is to visualise a common relationship across the dataset. This should not be understated, especially when one considers that filters can be stacked to display the union of relationships. Furthermore, due to the ambiguity inherent in queries with multiple filters, the option to restrict these conditions was added. With restriction applied, the associated filter is enforced upon the dataset, implying that any migrations that fail to meet its condition will be hidden. For those readers familiar with Boolean logic, restriction sets an 'and' relationship where normally there would be an 'or'. For example, the filters of 'Animator', 'Compositor' and 'New Zealand' will yield all animators, all compositors and everyone arriving in or embarking from New Zealand. However, with restriction applied to 'New Zealand', only animators and compositors moving to or from New Zealand will be shown. The difference is subtle, but the implications of these different forms of queries are profound, as they allow for the inspection of a subset of data that has an already established list of qualities. The filter machine has evolved from a basic implementation in the WebGL version to this more advanced dynamic query creator in the Unity5 prototype. In user testing, it was discovered that while users agreed that the latter was indeed more powerful,

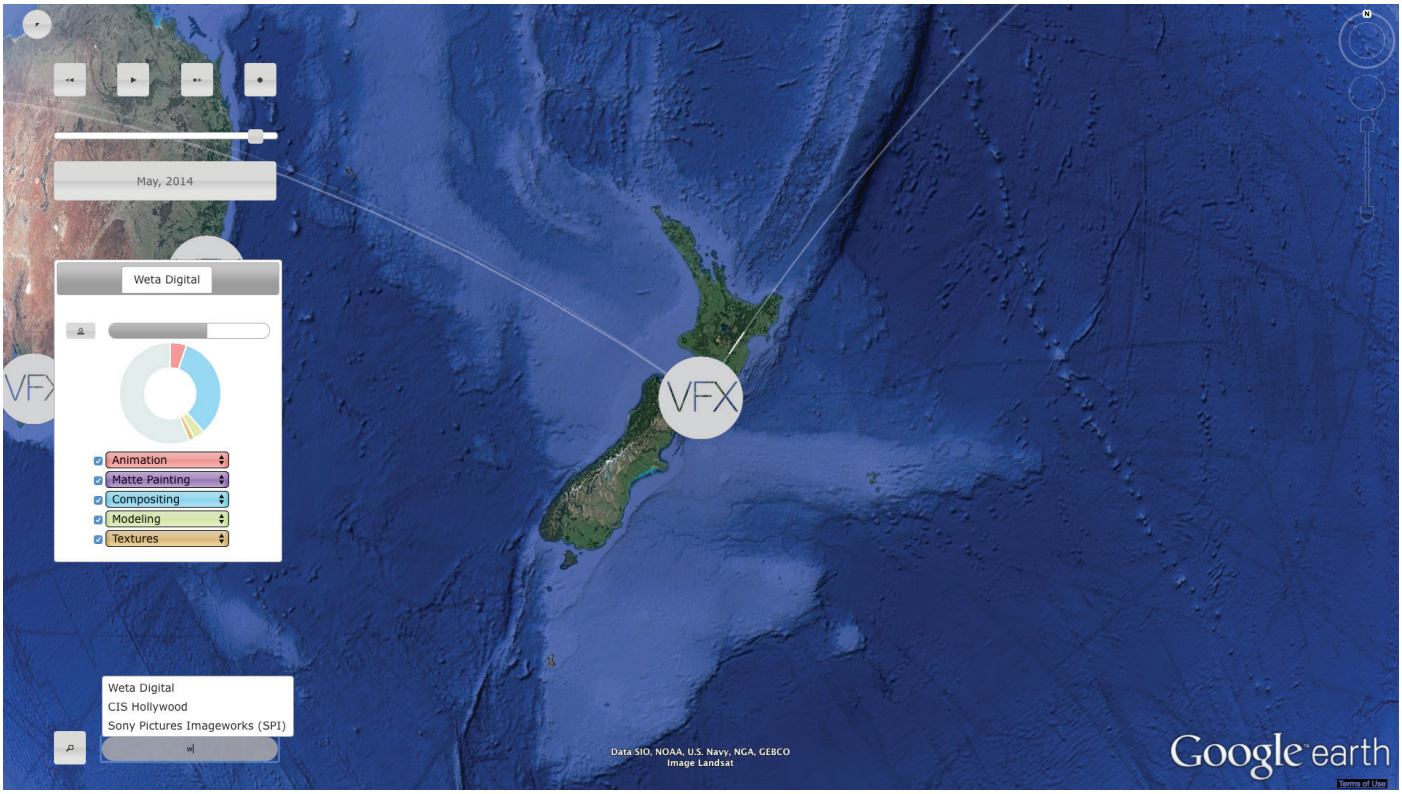
there was also a significant learning curve involved in embracing the feature. As such, it was decided that the Unity5 version would exhibit a duplicate of the WebGL filter widget as the default mechanism and also provide the advanced variant for users who desired more control over the system. To conclude, it should be noted that in this discussion of relationships, we are referring to the shared relationship between individual migrations based upon identical attributes, rather than the observed relationships between migratory subsets. The latter is also interesting however, and brings the data scientist into the realm of information analysis, or the investigation of how one set impacts another. The filtration of the dataset is arguably the most important feature of the visualization as it allows for targeted exploration and the discovery of geographic and demographic relationships within the VFX industry.

DETAILS ON DEMAND

The search bar has always been the primary interface to the database entities. As the data expanded with more categories introduced, there was an increased need for this component to function more efficiently and comprehensively. In the Google Earth prototype, the search bar simply recalled visualization studios. As there were few of these, there was no technical optimisation required and the program could simply iterate through the entire list with no delay. The WebGL iteration saw the addition of movies to the dataset and this implied a greater data load and the complexity of mixed search result types. The React.JS autocomplete module was employed to manage the searching process. This ultimately allowed faster lookup speeds, as the data was efficiently arranged in a trie structure that reduced the number of comparisons required to ascertain the correct search results. While both of these iterations provided a window into the data, the component effectively yielded 'data labels' that had little bearing on the rest of the visualization. It was the Unity5 variant, however, that integrated the search bar most effectively. This version also accommodated professions and skills into the list of acceptable search result types. When clicked, these would automatically append to the active filter list. Also, an information dialog box was established to elaborate upon search results with associated data. Furthermore, the quantity of data was a significant increase from previous

[FIGURE 4.10]

Iteration One Search Bar. This Google Earth version of the database interface only allows users to inspect a restricted list of VFX studios.

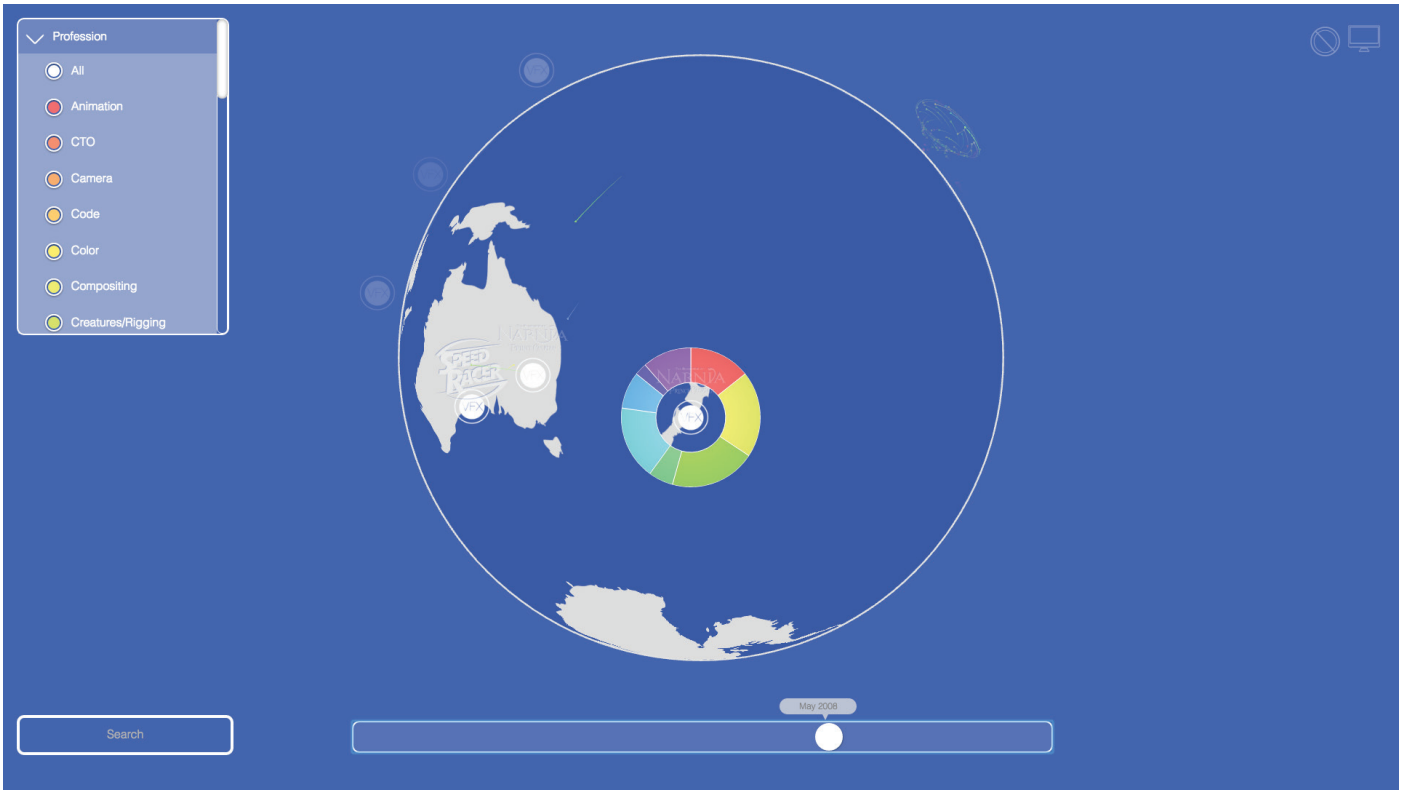


[FIGURE 4.11]

Iteration Two Search Bar. The WebGL version of the search bar allows users to inspect a greater range of VFX studios and also a list of ~600 movies.

[FIGURE 4.12]

Iteration Three Search Bar. The search widget in the Unity5 version accommodates searching by either movie, studio, profession or skill. The database consists of 2000 movies and 600 VFX houses.This variant also allows users to toggle the types of search results returned.



[FIGURE 4.13]

Iteration Two Professional Composition. An overlay element to display the professional percentages of regional populations over time.

[FIGURE 4.14]

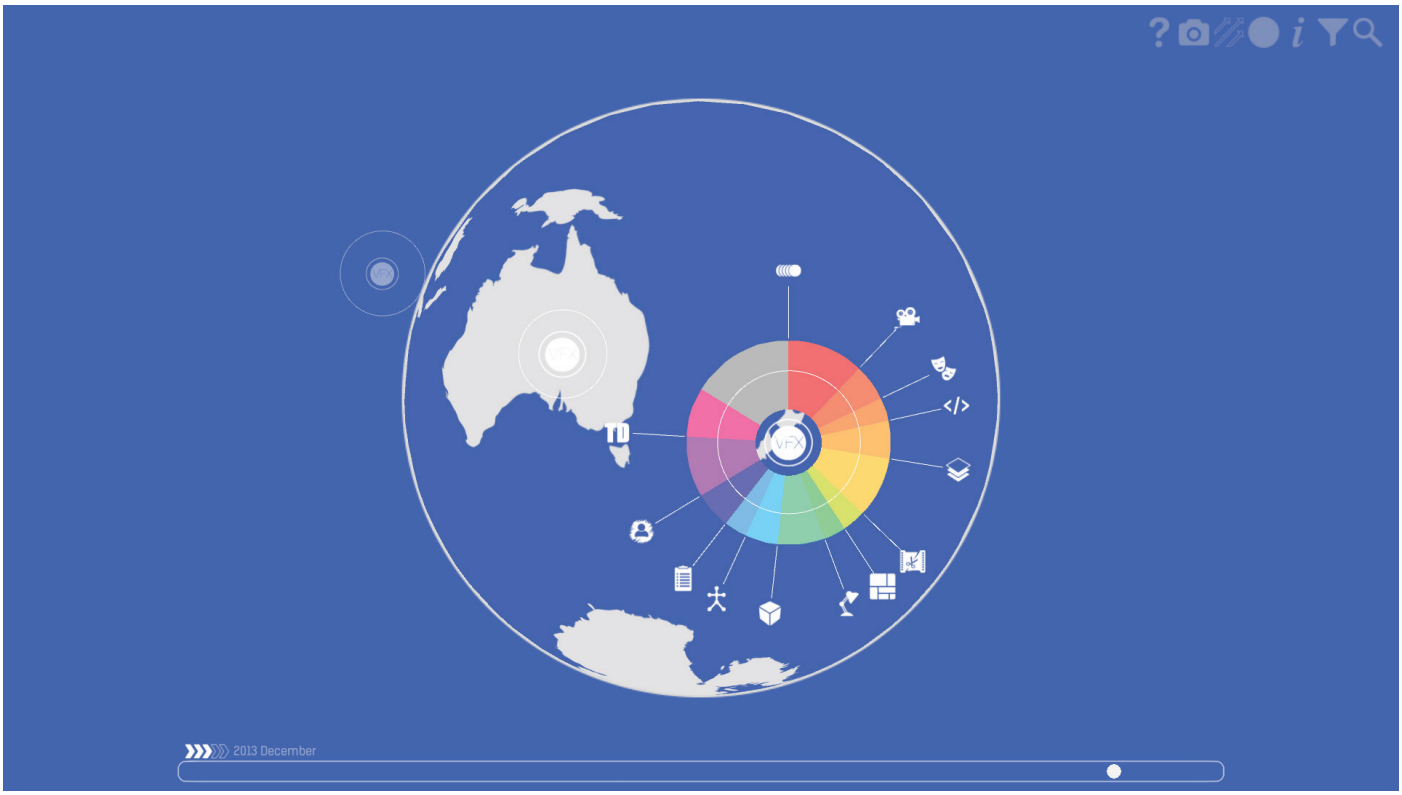
Iteration Two Regional Density. An overlay element that draws circles about hubs with radii indicative of the region's internal population.

versions, which placed a greater importance upon the component's functionality. As the React.JS framework was unavailable within Unity, a custom autocomplete structure was developed that operated efficiently and allowed for certain result types to be omitted from the search, based upon the user's preferences. In conclusion, the final iteration of the search bar was the most successful in providing details on demand due to its speed, size and integration with the surrounding system.

The toggle-able overlay components are also an example of features that provide details on demand. These components were created to render graphical diegetic representations of data associated with the various hubs. Specifically, they show total regional populations, the professional compositions of these populations and the movies being produced within these regions over time. Despite being roughly implemented within the Google Earth prototype, these features have existed as fully functioning overlays since the second iteration. One of the key challenges posed by these graphics is the preservation of data integrity. In 'The Visual Display of Quantitative Information', Tufte describes the 'lie factor' as the size of the effect shown in a graphic over the size of the effect within the data. A lie factor of one implies perfect data integrity. However, lie factors that are more than 0.05 to either side of this mark are considered "substantially distorted" (Tufte, 2001, p.57). The studio density overlay, which renders circles around hubs with radii indicative of internal populations, is an example of data distortion. The reasoning behind this is that across the entire range of hubs the populations vary dramatically in size. Expressing this linearly, thereby achieving perfect data integrity, would compromise visualization aesthetic and readability, as some circles would extend off-screen or some circles would be too small. Therefore, a logarithmic scaling was employed to massage the numbers into an acceptable domain. However, in terms of user perception, this does not adequately suggest the greater significance behind differing radii. While this appears to be a catch-22 situation, it is important to note the criticism against Tufte's rulings and the warnings for designers who take his aphorisms too literally. Wheaton College Professor John Grady states, "each chapter of his books consists of loosely integrated discussions on

[FIGURE 4.15]

Iteration Three Professional Composition and Regional Density. The Unity5 variant allows for both overlays to be displayed simultaneously. In this version, populations are larger and professional breakdowns are more comprehensive, as evidenced by the greater number of roles.



the merit of particular displays" (Cairo, 2013, p.65). He further elaborates by saying that information architects should not treat Tufte's work as guides or analytical texts, but rather as meditations or essays. In some cases, the application of these "abstract principles" to the real world is infeasible (Cairo, 2013, p.65). In this way, illustrations that do not achieve perfect data integrity are not necessarily irresponsible, especially when the scaling applied has been uniform across the entire dataset. The described overlay features portray intuitive and easily comparable views of the data. As they are information graphics, it is important to note the joint responsibilities they have to both data integrity and visualization clarity. The balanced approach discussed here ensures that these details on demand are optimally comprehendible by the viewer.

HISTORY AND EXTRACT

The newly added snapshot feature provides an interface through which to save and load session state. Shneiderman notes that the sheer number of steps involved in adequately exploring large datasets warrants the need for a history of actions and the ability to retrace (1996, p.5). Previous versions of Digital Workshops of the World destroyed all session information upon program termination. However, due to the heightened exploration potential of the Unity5 prototype, it was reasoned that the ability to permanently store interesting observances was desirable, especially when those observations required complex filter setups. As Unity is first and foremost a game engine, there was a lot of support for developing save and load functionality. Essentially, this was achieved through the serialisation of visualization objects into a binary representation and subsequently the writing of these bits to a specified file within the project directory. Upon relaunch, the program would read and parse the file, thereby restoring knowledge of previous session states. In terms of extraction, Shneiderman suggests that the best visualizations allow the means to export and share data in a variety of formats (1996, p.5). Currently, the only output format afforded by the most recent iteration is binary and this is not human-readable. Specialised screenshots, custom data reports or shareable session files could be a useful feature within a future release, especially in regards to promoting collaboration between data scientists.

FUTURE DESIGN IMPROVEMENTS

While Data Deciphered is an improvement to the previous Digital Workshops of the World prototypes, the final output still leaves features to be desired. Firstly, while this end product did a predominantly comprehensive job of exhibiting the database, there were still categories of information that were only visible through the information panel. Specifically, these were studio opening and closing dates and movie budget and box office statistics. Rather than simply being text on screen these numbers would have been better showcased in graphical form. Ideally, they would be embedded within additional overlays that visually indicate this data over time. In the case of studio dates, distinct opening and closing icons could appear around the associated hubs as a timestamp that indicates the internal change in the corporate ecosystem. Additionally, in the case of movie monetary data, the size of movie title graphics could be utilised as a variable representative of budget or box office haul. The strength of visualising these numbers over time is that it immediately introduces comparison and therefore heightens the viewer's understanding of the dataset. Placing a chart around the globe as a diegetic element also increases comprehension by intuitively illustrating location. Essentially, these additional overlays would further promote engagement with the database and therefore strengthen the visualization in terms of its fundamental purpose.

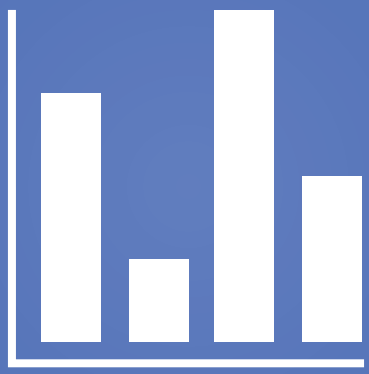
While the filter system provides a powerful utility for exploring the dataset and presenting its key trends, it could be improved to specifically identify extremes. When interviewed in user testing, participants tended to phrase a typical visualization query with superlatives. This notion of only presenting the extents of a specific subset of data is a functionality currently not supported by the filter machine. In future development, it would be desirable to augment the query system with buttons that only show data above or below a user-defined threshold. In this way, questions such as 'what are the highest budget movies of all time?' or 'what types of professionals associate with the largest skill sets?' could be specified and answered. As questions involving extremities are typical in cognitive inquiry, it makes sense that the filter system should support this same process of query formulation.

As it currently stands, the context of the VFX industry has been applied to the visualization. While this particular demographic demonstrates extensive migration, it is

not the only sector to do so. A future improvement of this software is to create a completely generic code architecture that is accommodating of any given input, so long as the database schema adheres to the required format. A customisable user interface would need to be developed alongside this functionality for the purpose of reflecting the altered nature of the data. The creation of this template would allow designers to rapidly prototype customised interfaces and would facilitate the process of data investigation across a wide range of contexts.

The ability to make the visualization and database live online is a further improvement that is scheduled for a future release. Doing so would enable connection with social media platforms, such as Facebook or Twitter. In addition to merely viewing the data, users could be given the option to contribute their own VFX employment history, which would be validated by their profile. Subsequently, this data would be programmatically injected into the database, providing further accuracy and substance to the research. This improvement would change the very nature of the project from a static display into an evolving, communal representation of industry.

5 DATA IMPLICATIONS



The following section presents key data results and the implications of these findings, which are pertinent to the VFX industry. At the genesis of the Digital Workshops of the World project, research questions were suggested with the aim of focussing the processes of data accumulation, visualization design and statistical inquiry. This section will initially discuss these queries before subsequently delving into the key demographic contributions of this thesis. These relate to regional population, professional mobility, coding literacy and software proficiency within the VFX industry. Additionally, this section will deliberate upon the validity of these conclusions, given the crowd-sourced nature of the project's dataset.

DIGITAL WORKSHOPS OF THE WORLD RESEARCH QUESTIONS

The primary research questions posed by the Digital Workshops of the World project to direct exploration through the database were:

- *Which VFX regions demonstrate the most significant growth?*
- *Which VFX regions have historically had the largest population?*
- *What is the average professional composition of VFX regions over time?*
- *Which migratory roles are the most / least mobile?*
- *Which coding languages are the most prevalent within the VFX industry?*
- *What correlation exists between coding ability and mobility?*
- *Which migratory roles demonstrate the highest levels of coding ability?*
- *Which software packages are the most prevalent within the VFX industry?*
- *What correlation exists between software proficiency and mobility?*
- *Which migratory roles demonstrate the highest levels of software proficiency?*

In attempting to answer these questions, the visualization was initially employed to suggest extremities, narratives and general trends. Following this insight, a series of Python scripts were coded to run specific passes across the migration data and automatically extract values based upon the current inquiry. These statistics were subsequently placed into Microsoft Excel, whereupon they were formulated into intuitive conclusions. It should be noted that while the research questions above were used to guide this process of result generation, the database consists of significantly more variables and is therefore capable of producing additional findings outside of this scope. In the absence of time constraints, further inquiry that examined the relationship between, for example, movie production and studio longevity or acquired skills and visited regions would have been desirable.

IMPERFECT DATA ADMISSIONS

While the integrity of process was maintained during the data mining and analysis phases, it is important to note that the conclusions drawn from the database are imperfect and are therefore not 100% reliable. One of the stipulations of crowd-sourced data is that there is no guarantee that the information is 'real'. The greatest appeal of this means of data capture – the fact that anyone can contribute – is also its greatest risk. Disregarding terms of use, anyone can create a misleading profile online and this presents potential for any accumulated data to be flawed. However, many social media platforms, such as LinkedIn, pride themselves on data integrity and are not hesitant in terminating accounts that they believe violate their User Agreement (LinkedIn, 2014, Section 3.4). Furthermore, it is in the best interest of users to provide legitimate information, as failing to do so places their professional reputation and employment candidacy in jeopardy.

Another potential inaccuracy of the data validation process lies in the mapping of studios and roles. This involved translating raw text from an online profile into the identifying tag of a database object. Specifically in regard to roles, the large number of diverse strings required an automated assignment. It is possible that this process may have yielded some incorrect mappings, as it was based upon the identification of particular substrings in the role description. Furthermore, the manual mapping undertaken to consolidate residual studios and roles may also have introduced error. Some text labels were especially vague and this made valid classification dependant upon the researcher's interpretation. The only way to determine the correct

mappings with certainty would be to individually interview each of the 22,554 VFX professionals in the database. As this is logistically impossible, slight data inaccuracies must be accepted as the most precise outcome.

Despite being a comprehensive sample, the database does not fully describe the extent and various parameters of the VFX industry. While the 82,711 entities indicate a significant improvement upon previous versions, it would be wrong to assume that this size implies a complete representation of VFX migration. For example, professionals without a LinkedIn profile are inherently omitted from this research. Furthermore, while every effort has been made to include an extensive set of studios, there will inevitably be some over the course of VFX history that have missed registration. Additionally, in considering the fact that LinkedIn is a predominantly Western organisation and Internet platform, it is conceivable that professionals in Asia or South America may be underrepresented. However, despite not being all-encompassing, the database should still be regarded as a powerful and representative sampling of the VFX industry, due to the significant amount of valid information that it provides.

Skill information was used to determine qualities such as coding literacy and software proficiency in the database analysis, which implies that there is a potential margin of error in these findings. As previously discussed, skill tags were sourced from the 'Skills' section of a LinkedIn profile. However, there is no requirement for members to exhibit this information online. Therefore, it is possible that an adept coder would not be classified as such within the Data Deciphered database, in the instance that they had not listed any programming languages on their profile. With this said, this thesis has advanced under the proviso that it reasonable to assume for the majority of users that important career-related proficiencies will be displayed if attained. In this sense, the conclusions drawn from skills analysis are implicational if not certain.

Even though the findings of the database should not be treated as gospel, they are likely indicative of trends in the industry. Social media offers a data-gathering platform that is imperfect. However, it is important to realise that perfect and ascertainable data accumulation is not achievable given the limitations and formats of current online technologies and systems. Researchers do have control over process however, and in this respect the employed methodologies have rigorously adhered to sound practice and statistical integrity. In this sense, the findings submitted by this thesis are completely valid within the context of the database and they constitute a comprehensive portrayal of the VFX industry as a whole.

REGIONAL DENSITY

Population analysis, both in terms of size and composition, was key in scrutinising the density of VFX regions over time. In utilising the figures generated for the visualization overlays it was possible to calculate the total lifetime population for each hub. From this, an averaged monthly density was interpreted, which allowed for the direct comparison of regional size.

(CONCLUSION 1)

California is, and always has been, the largest region in terms of VFX population.

[FIGURE 5.1]

Opposite. *Averaged Regional VFX Populations.* Infographic comparing the lifetime average populations of VFX hubs across the period Jan 1990 - Jan 2016.

[FIGURE 5.2]

Next. *Regional Growth in the VFX Industry.* Infographic comparing the population growth of VFX hubs across the period Jan 1990 - Jan 2016.

[FIGURE 5.3]

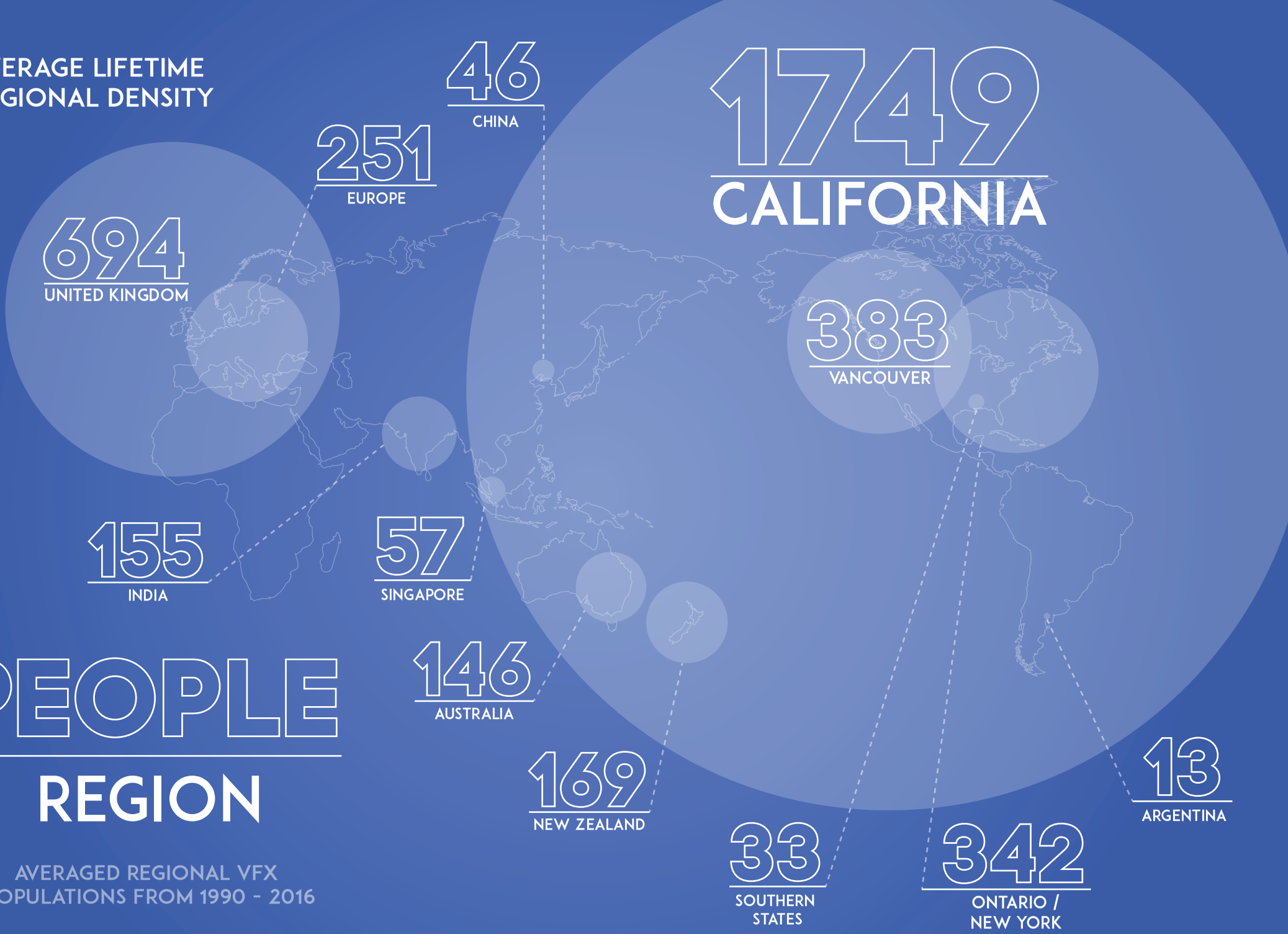
Next Opposite. *Regional Composition in the VFX Industry.* Infographic comparing the lifetime professional compositions of VFX hubs across the period Jan 1990 - Jan 2016.

AVERAGE LIFETIME
REGIONAL DENSITY

PEOPLE

REGION

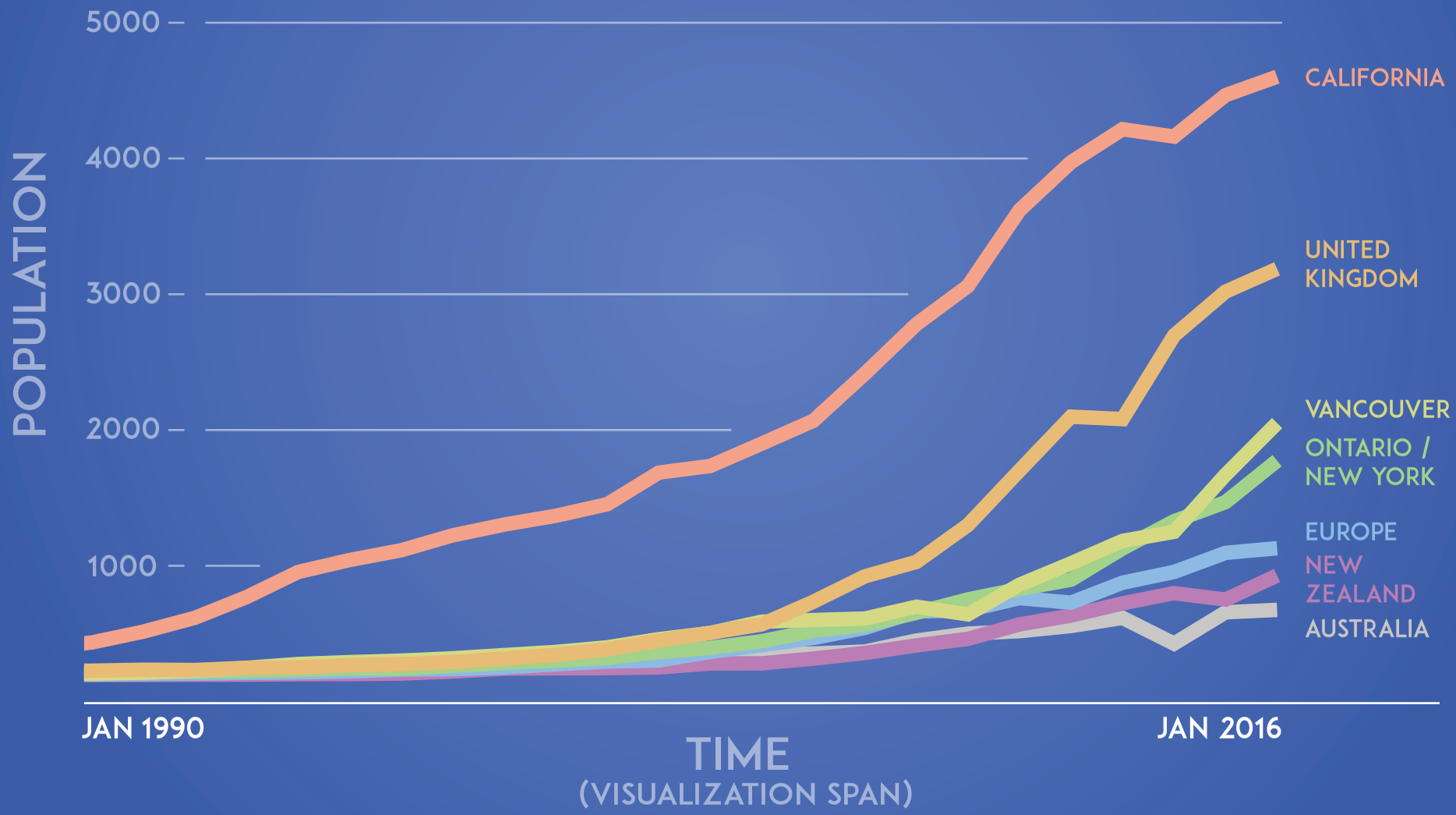
AVERAGED REGIONAL VFX
POPULATIONS FROM 1990 - 2016



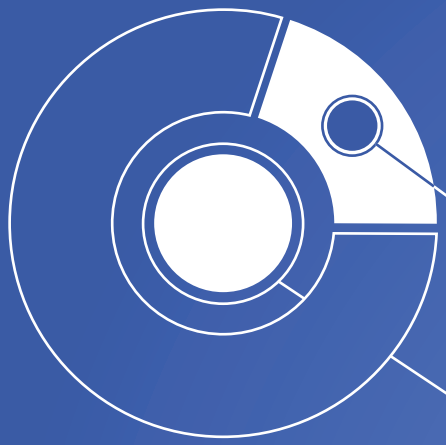
Data Deciphered: A Visual Migration Of VFX

GROWTH OF VISUAL EFFECTS

REGIONAL POPULATION DENSITY BETWEEN 1990 - 2016

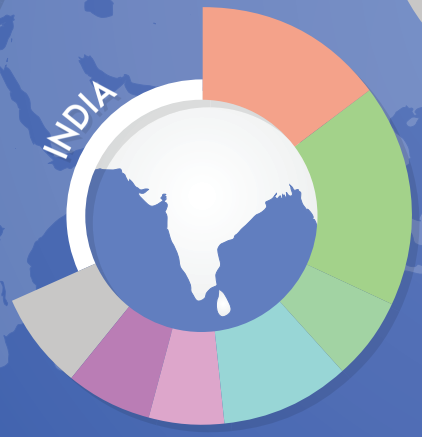


Data Deciphered: A Visual Migration Of VFX



AN ANALYSIS OF ROLE BY REGION

ANIMATION
CHARACTER
CODE
COMPOSITING
EFFECTS
LIGHTING
MODELING
PRODUCTION
ROTOSCOPE
TD



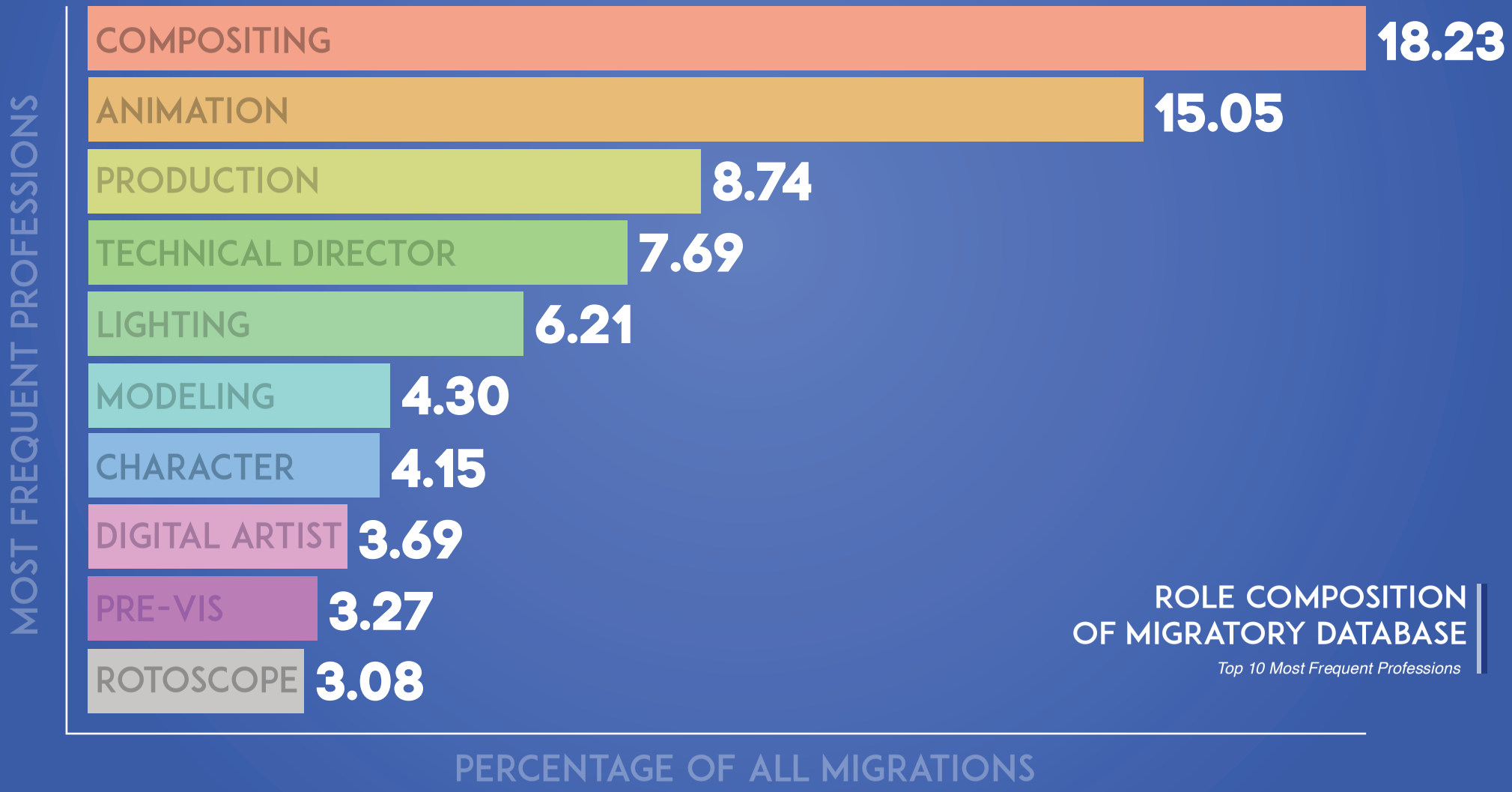
AVERAGED LIFETIME REGIONAL COMPOSITION

Additional Role Categories

Data Deciphered: A Visual Migration Of VFX

MIGRATION BY PROFESSION

82711
MIGRATIONS IN THE DATABASE



ROLE COMPOSITION
OF MIGRATORY DATABASE

Top 10 Most Frequent Professions

Data Deciphered: A Visual Migration Of VFX

The overlay statistics were subsequently charted to produce a line graph that exhibited the relative populations of VFX regions over time (see Figure [5.2]). Through gradient analysis it was possible to identify the fastest regional growth within VFX history.

(CONCLUSION 2)

The United Kingdom exhibits the fastest growth in VFX population across the period of June 2013 – March 2014.

The averaged professional composition displays are indicative of the types of VFX professions that have typically thrived in their respective regions over time. These can be observed in Figure [5.3]. Furthermore, a database illustration of migration in terms of professions can be seen in Figure [5.4].

Commentary that alludes to a Hollywood diaspora (Grage, 2015, p.170-172; VES, 2013, p.5-6) is widespread throughout VFX discourse; however, it is evident from these statistics that the exodus is not as severe as its reputation implies. California has historically functioned as the powerhouse of VFX (Grage, 2015, p.170). Despite concessions to start-ups in regions with desirable tax incentives, it has securely maintained its position at the centre of the industry. Even though substantial growth has been recently observed in locations such as London and Vancouver, it does not appear likely that these regions will surpass California in the immediate future. This is further supported by the current upward trend in Californian VFX population.

ANALYSING MOBILITY

In analysing mobility, the primary aim was to determine a correlation between professions and migration. Specifically, an emphasis was placed on roles with the highest and lowest mobility. However, mobility was not a variable of the database and so no direct comparison could be made without prior inference. In considering the quantification of movement, an initial solution is to use distance as a metric. In this sense, a migration that travels from California to New Zealand, for example, is regarded as highly mobile. In contrast, a migration that moves between VFX houses within California has low mobility and a migration based upon a change in profession within the same studio has no mobility. Technically, the database implements these scenarios as external, internal and non-migrations respectively. However, movement can also be regarded as an expression of activity. Therefore, a professional with an employment history that consists of ten different contracts is considered highly mobile, while one with only two listings has low mobility. In this sense, geography has no bearing on the metric and so a professional can still be deemed mobile even if their entire career is spent at the same company. This ambiguity in regards to mobility implied that two different approaches were necessary in order to thoroughly assess this condition for a particular role.

The problem was further complicated by the fact that a role could also have two interpretations based upon whether it was examined by migration or by person. A role can be regarded as temporary when it is presented in the context of a migration. In this sense, a person is labelled with the role for the duration of their employment, before moving onto another profession. However, a person can also be associated to multiple permanent roles over the course of their career. In this way, a person is regarded as a combination of all the roles that they have ever performed. Therefore, the mobility of a role could be viewed in terms of the averaged geographic movement of all of the migrations associated with the profession. However, it could also be considered as the averaged activity of all people who have ever identified with that particular role.

In order to analyse geographic mobility, a custom Python script was employed to iterate through all migrations and sort them into buckets based upon their associated role. Each bucket was subsequently split into three subcategories depending on whether

[FIGURE 5.4]

Opposite. *Role Composition of Database Migrations.* Infographic illustrating the breakdown of all database migrations by profession.

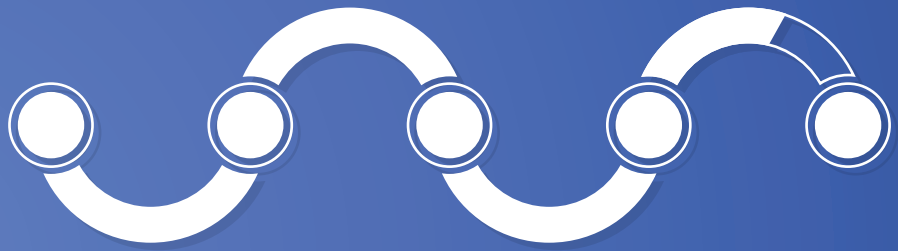
MOBILITY

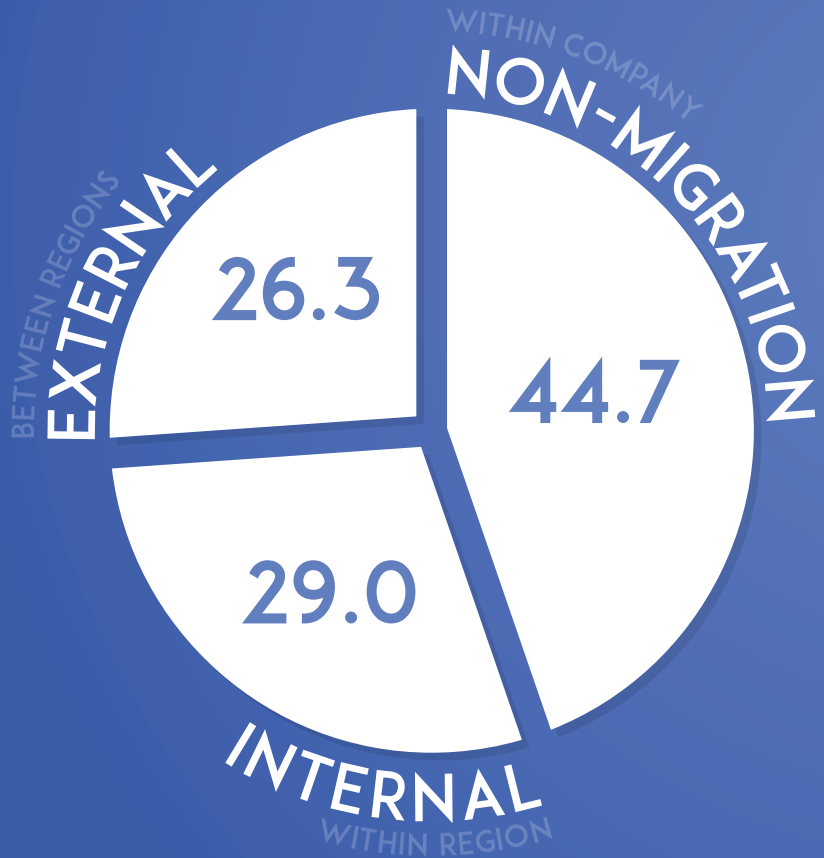
IN VISUAL EFFECTS

THE VFX ARTIST WILL
MOVE ON AVERAGE

3.67 TIMES

OVER THE COURSE OF
THEIR CAREER

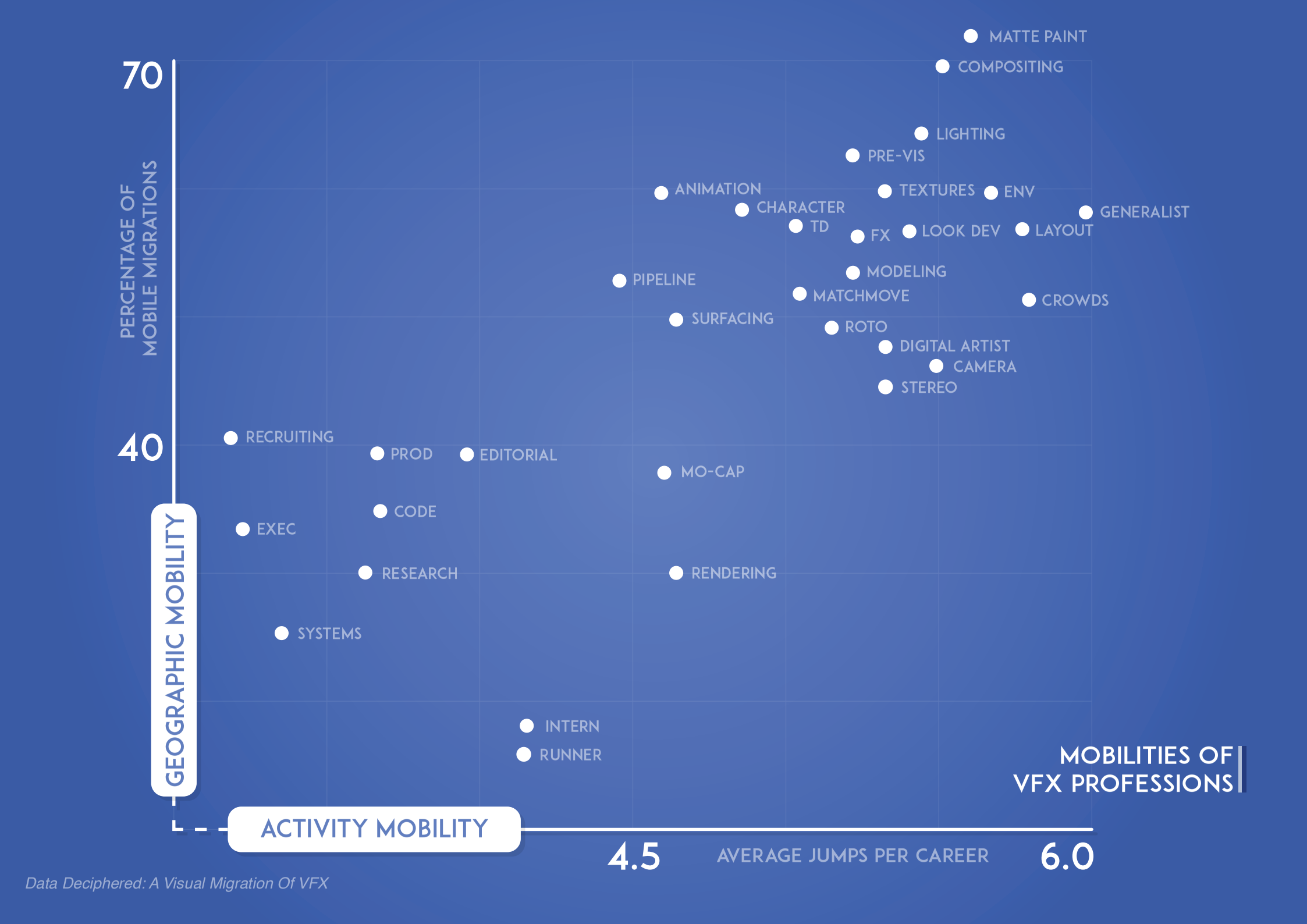




PERCENTAGE OF MOBILITY
TYPE OVER ALL MIGRATIONS

AVERAGED ACTIVITY
AND GEOGRAPHIC MOBILITY

Data Deciphered: A Visual Migration Of VFX



Data Deciphered: A Visual Migration Of VFX

the inspected migration was 'external', 'internal' or 'non'. The percentages of these subgroups were then calculated for each role, relative to the total number of migrations in that role's bucket. In this way, professions that demonstrated a higher percentage of non-migrations were deemed least mobile, while those with greater external figures were regarded as most mobile.

In order to analyse active mobility, a conceptual shift was required from migrations to people as the atomic data entity. Another Python script was written to extract each person's total number of lifetime migrations. Every person was also associated to the professions with which they identified. Using this information, the person's total migration count was subsequently added to the global migration count of the corresponding roles. In dividing this role count by the number of unique individuals associated with that profession, the average number of migrations per person was calculated.

While both variants of the role mobility statistics hold weight, their validity can be criticised based upon the ambiguity inherent in their interpretation. Therefore, in order to provide additional credibility, the stipulation that a particular role needed to lie at the ends of both the geographic and activity spectrums in order to be considered (im)mobile was introduced. This meant that, after the results tables had been calculated, a role could only be regarded as 'most mobile' if it appeared in both the top five geographically mobile and actively mobile professions. Likewise, in order to be seen as 'least mobile', a profession had to rank within the five lowest geographic and activity mobility scores. Based upon this analysis, this thesis presents the following conclusions:

(CONCLUSION 3)

Environments and Matte Painting are the most mobile professions within the VFX industry.

(CONCLUSION 4)

Research and Systems are the least mobile professions within the VFX industry.

In addition to the above conclusions, it should be noted that Lighting, Compositing, Textures, Generalist, Layout and Look Development are also professions with high mobility. Runner, Intern, Executive, Code, Editorial and Production meanwhile are predominantly static roles. Figure [5.6] provides an illustration of these findings. An interesting observation regarding the nature of these roles is that support professions, which are primarily concerned with aiding the studio, are less mobile than those roles that actually work on movies. It appears that due to the contractual nature of VFX film production, compositors, lighters and environment artists will move with the ebb and flow of movie work, while coders, technicians and production staff tend to have an extended employment at studios. An additional observation is the presence of runners and interns within the low mobility set. This could be the result of an implementation artefact and may not accurately depict the migratory nature of these roles. As discussed within the section on data-mining process, in order to include the inaugural contract of a VFX worker within the database, a 'dummy' entry was copied from this element and introduced into the employment history. This allowed for a migration to be inferred that accurately represented the first contract. However, this record will always be classified as a non-migration, due to the fact that its contracts were copied. In this way, the high geographic immobility of these roles may be indicative of the fact that they are typically entry-level titles, rather than a reflection of a predominantly static nature.

[FIGURE 5.5]

Previous. *Mobility in the VFX Industry.* Infographic providing statistical analysis of activity mobility and geographic mobility within the VFX industry.

[FIGURE 5.6]

Previous Opposite. *Mobility Analysis of VFX Industry Professions.* Infographic comparing activity mobility against geographic mobility for all database professions.

CODING LITERACY WITHIN VFX

'Coding literacy' was determined for a person if the skill section on their profile contained at least one designated coding language. The list of coding languages was a subset of the complete skill collection in the database and are as follows:

- **ActionScript**
- **Bash**
- **C**
- **C#**
- **C++**
- **CSS**
- **GLSL**
- **HTML**
- **Java**
- **JavaScript**
- **Lua**
- **MATLAB**
- **MaxScript**
- **MEL**
- **Objective C**
- **Open GL**
- **Perl**
- **PHP**
- **Python**
- **SQL**
- **Unix Shell Scripting**
- **VEX**
- **XML**

A Python script was implemented to iterate through all people in the database and identify those deemed as 'coders' by this list. In comparing the size and frequency of this subset against the total number of people in the database, conclusions five and six were drawn:

(CONCLUSION 5)

One in five VFX professionals are coding literate.

(CONCLUSION 6)

MEL, Python and C++ are the most prevalent coding languages within the VFX industry.

In order to determine a correlation between coding literacy and migration, the methods for assessing mobility were applied to the coding subset. This resulted in a marginal indication of higher geographic and activity mobility for professionals with knowledge of scripting or programming. However, this increase was too small to be statistically significant and therefore the notion that coding literacy increases the chances of employment mobility was not clear enough to warrant publication.

In order to determine a correlation between coding literacy and profession, for a specific role, the total number of migrations pertinent to coders was placed over the total number of migrations.

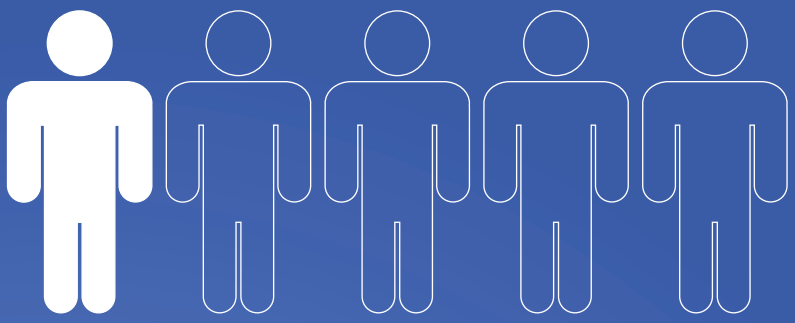
(CONCLUSION 7)

Pipeline, Code and Research professions have the highest percentages of coding literacy.

(CONCLUSION 8)

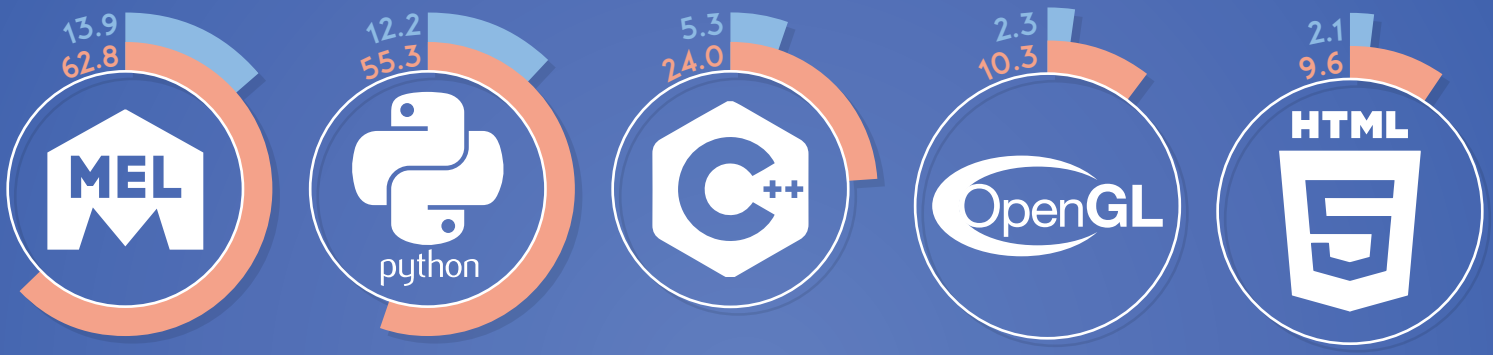
Editorial, Recruiting and Pre-Visualization professions have the lowest percentages of coding literacy.

CODE LITERACY IN VISUAL EFFECTS



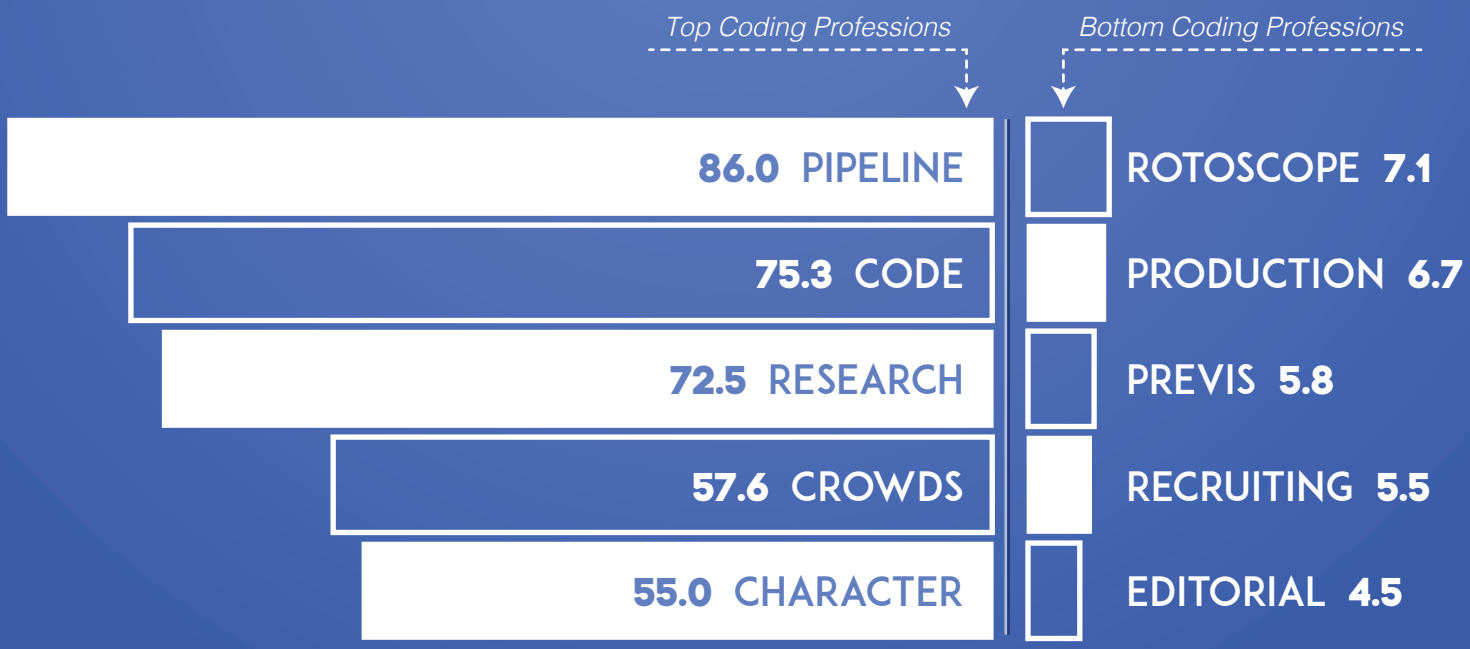
1:5

ONE IN FIVE VFX WORKERS CAN CODE



TOP FIVE MOST POPULAR LANGUAGES

Percentage Among Coders
Percentage Among All Artists



CODING PERCENTAGES IN PROFESSIONS

Data Deciphered: A Visual Migration Of VFX

SOFTWARE PROFICIENCY WITHIN VFX

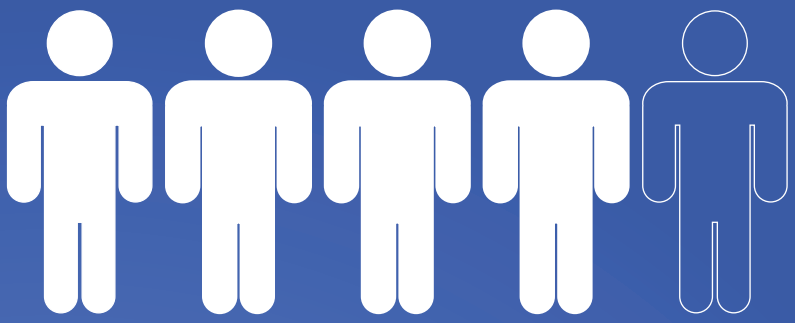
Similar to coding literacy, 'software proficiency' was determined for a professional if their profile's skill section contained at least one tag out of the predefined list of VFX software applications. This list was comprised of the most frequent packages from across the entire database.

- 3D Coast
- 3D Equalizer
- 3D Studio Max
- Adobe After Effects
- Adobe Creative Suite
- Adobe Illustrator
- Adobe InDesign
- Adobe Lightroom
- Adobe Photoshop
- Adobe Premiere Pro
- AutoCAD
- Autodesk MatchMover
- Autodesk Maya
- Autodesk MotionBuilder
- Autodesk Smoke
- Autodesk Software
- Avid
- Avid Media Composer
- Blender
- Bodypaint
- Boujou
- Cinema 4D
- Corel Painter
- DaVinci Resolve
- DVD Studio Pro
- Final Cut Pro
- Final Cut Studio
- Flame
- Flash
- Fume FX
- Fusion
- Houdini
- Katana
- Krakatoa
- Lightwave
- Logic Pro
- Mari
- Massive
- Microsoft Excel
- Microsoft Office
- Microsoft Outlook
- Microsoft Powerpoint
- Microsoft Word
- Mocha
- Modo
- Mudbox
- Naiad
- Nuke
- PFTrack
- Pro Tools
- Rayfire
- Realflow
- RV
- Shake
- Shotgun
- Silhouette
- SketchUp
- Sony Vegas
- Syntheyes
- Thinking Particles
- Toon Boom
- Topogun
- Unity 3D
- Unreal Engine
- Vue

[FIGURE 5.7]

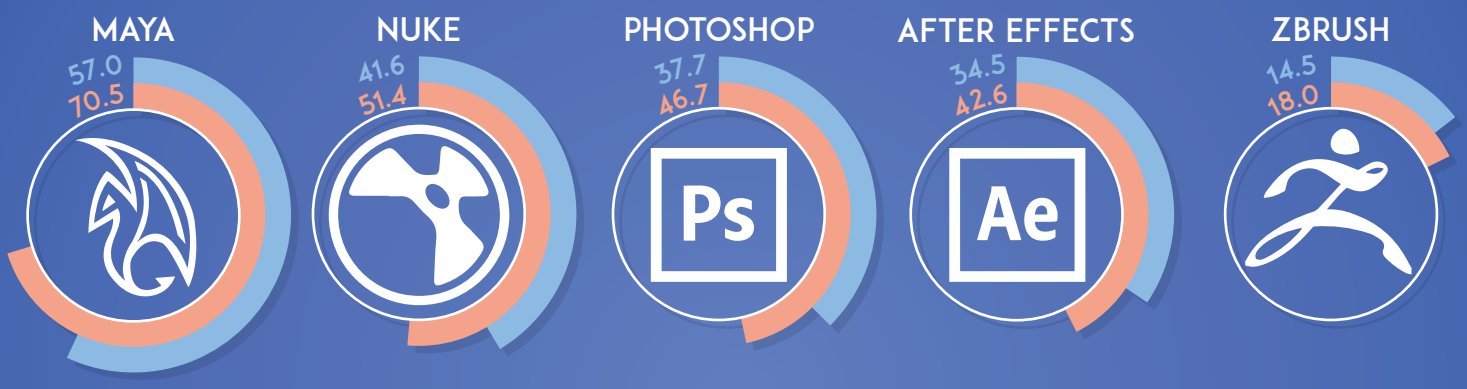
Opposite. *Coding Literacy in the VFX Industry.* Infographic denoting the most popular coding languages and the coding percentages of professions within the VFX industry.

SOFTWARE PROFICIENCY IN VISUAL EFFECTS



4:5

FOUR IN FIVE PROFESSIONALS KNOW VFX SOFTWARE



TOP FIVE MOST POPULAR PROGRAMS

Percentage Among Software Users
Percentage Among All Artists

Top Software Professions

- 93.2 FX
- 92.9 COMPOSITING
- 92.0 GENERALIST
- 91.9 ROTOSCOPE
- 91.8 MODELING

Bottom Software Professions

- PRODUCTION 69.5
- RESEARCH 68.9
- RECRUITING 68.5
- SYSTEMS 58.9
- EXECUTIVE 52.6

SOFTWARE PROFICIENCY PERCENTAGES IN PROFESSIONS

Data Deciphered: A Visual Migration Of VFX

- **Zbrush**

Following the same methodologies as the coding analysis, a Python utility script was created to identify all professionals deemed as 'software users' by this list.

(CONCLUSION 9)

Four in five VFX professionals are proficient in common VFX software packages.

(CONCLUSION 10)

Autodesk Maya, The Foundry Nuke and Adobe Photoshop are the most popular software packages in the VFX industry.

Autodesk Maya holds a 15% margin over its competitors in its position as the most utilised software in the VFX industry. It is interesting to note the correlation between this prevalence and the fact that the three most popular coding languages are those used for interfacing with the Maya API. These figures are testament to Autodesk's dominance in the VFX industry. Grage notes that this standing is partially due to the fact that the company holds the most important patents and is therefore able to constantly develop Maya further (2013, p.140-141).

In correlating software proficiency and migration, a marginal suggestion of higher mobility was observed in both geographic and activity respects. However, as with coding literacy, this increase was too small to be statistically significant and to confirm the notion that software aptitude increases the chances of employment mobility.

In determining the correlation between software proficiency and profession, for a specific role, the total number of migrations associated with software users was placed over the total number of migrations.

(CONCLUSION 11)

The FX profession has the highest percentage of software proficiency.

(CONCLUSION 12)

The Executive profession has the lowest percentage of software proficiency.

The fact that FX professionals are the most versed in VFX software makes sense in considering that nowadays the vast majority of all simulations are constructed digitally. Fire and fog effects, fluid dynamics and cloth and hair simulation are common facets of the FX category, which are all developed within 3D software packages such as Autodesk Maya. Therefore, as it is their digital bread and butter, it follows that these professionals would associate their software proficiencies with their online profiles. In the case of Executives, who appear at the bottom of the results table in regards to software aptitude, it would be wrong to assume technological incompetence. Rather, it is likely an indication that the programs employed by these owners, financiers and managers are not of the VFX persuasion and therefore do not count in this analysis.

[FIGURE 5.8]

Opposite. *Software Proficiency in the VFX Industry.* Infographic denoting the most popular software packages and the software percentages of professions within the VFX industry.

6 CONCLUSION

CONCLUSION

The visualization of the VFX industry makes clear that it is a global interconnected network of professions and skills in constant motion. The primary aims of this thesis were to provide irrefutable quantitative data on industrial migration and to examine the inherent geographic and demographic trends of this dataset through the medium of interactive visualization. In doing so, it both fulfils the project statement of the Digital Workshops of the World initiative and answers the request of the Visual Effects Society for statistical analysis of the sector.

The final database consisted of 82,711 migratory employment records from 22,554 unique VFX professionals. This was assembled through the programmatic downloading of the public-facing pages of LinkedIn profiles. This information was augmented with 2,000 movie box office records from the Internet Movie Database (IMDB). In this way it was able to provide a comprehensive sampling of VFX industry migration and contextualise these jumps with timestamps of specific film release dates. Given further time and resource, a desirable improvement to this data-mining initiative is to dramatically increase the size of the database, in order to be more representative of the sector. As LinkedIn is a predominantly Western Internet platform, it would be advantageous to additionally utilise alternate European and Asian professional networking websites to achieve more comprehensive coverage. Qualitative research that incorporates survey, interview and case study methodologies could be employed in combination with this quantitative assembly to further validate findings and provide inference to the underlying rationale.

The final application is the third in a series of prototypes and has therefore undergone rigorous feature development and continuous refinement over the process of its creation. Therefore, while there remains room for an extension of functionality, as it currently stands the tool comprehensively achieves the aims as outlined in the project's proposition. The visualization provides a platform for the discovery of trends and patterns within the dataset. This is enhanced through an advanced filtering mechanism that allows the user to select subcategories of data based upon profession, skill, region or VFX studio tags. Additionally, the option to specify additive or restrictive relationships between query elements is provided. Ultimately, this has yielded an iteration whose primary advantage over its predecessors is its powerful ability to explore the dataset. It should be noted that this output is much more than just a chart that exhibits migratory data. As Wright writes, "a visualization is not a representation but a means to a representation" (2008, p.81). In this way, the goal of the application is not merely to present conclusive findings but to facilitate the process of perpetual discovery. This is not to say that graphical charts and conclusions are irrelevant, rather that they are a result of visualization.

In utilising statistical analysis and visualization methodologies, this thesis has also generated findings and implications applicable to the VFX industry. Specifically, it has found that despite claims to the contrary, California still continues to function as the primary regional hub of global VFX activity. In its analysis of migratory behaviour, it has determined that environment artists and matte painters are the most mobile professionals in the industry, in both geographic and activity respects. At the opposite end of the spectrum lie the technical, systems-oriented professions. This implies that supportive roles are more likely to stay tied to studios, while professions that work directly with digital film effects have a greater tendency to migrate, due to the temporary nature of VFX contracts. Finally, MEL, Python and C++ have been verified as the most popular coding languages in the industry, while Autodesk Maya is the most prolific software package. Despite common belief that coding literacy and software proficiency increases one's chances of migration, this thesis has denounced this suggestion by discovering that such increases are marginal at best.

Data Deciphered: A Visual Migration of VFX has produced a framework for visualising migratory data over time. Specifically, this has been viewed in the context of the VFX industry to draw valuable conclusion and insight. However, in a general sense, a key contribution of this paper is the documentation of a process for mining, consolidating and verifying raw Internet data and subsequently exhibiting it in an interactive visualization. In this Information Age where big data is being generated at a rate that surpasses humanity's ability to comprehend it all (Cairo, 2013, p.15; Davis, 2012, p.4-5), visualization is becoming more necessary than ever within science and design disciplines. It is fundamentally important that information architects continue in their endeavours to interpret and analyse the vast tsunami of bits on the horizon. There is significant potential in understanding the 'how' and 'why' of social science phenomena and it is a potential that has yet to be fully explored.

BIBLIOGRAPHY

Barkan, K. (2014, Feb 28). *What's Wrong with the Visual Effects Industry?* Retrieved from: http://www.siggraph.org/discover/news/whats-wrong-visual-effects-industry

Bevir, G. (2014, Jul 4). Merger with Prime Focus to Boost Double Negative TV Arm. *Broadcast*. pp.15(1).

Brzeski, P. (2015, Dec 31). China Box Office Grows Astonishing 48.7 Percent in 2015, Hits $6.78 Billion. *The Hollywood Reporter.* Retrieved from: http://www.hollywoodreporter.com/news/china-box-office-grows-astonishing-851629

Cairo, A. (2013). *The Functional Art: An Introduction to Information Graphics and Visualization.* Berkeley, CA: New Riders.

Chaomei, C. (2006). *Information Visualization: Beyond the Horizon (2nd ed.).* London: Springer-Verlag.

Chung, H. J. (2011). Global Visual Effects Pipelines: An Interview with Hannes Ricklefs. *Media Fields Journal,* (2). Retrieved from: http://www.mediafieldsjournal.org/global-visual-effects/

Dai, K. (2015). *Scraping and Clustering Techniques for the Characterization of Linkedin Profiles.* Paper presented at The Fourth International Conference on Information Technology Convergence & Services (ITCS 2015), Zurich, Switzerland. doi: 10.5121/csit.2015.50101.

Davis, K. (2012). *Ethics of Big Data.* Sebastopol, CA: O'Reilly Media Inc.

Dodgson, N. (2010). *What's Up Prof? Current Issues in the Visual Effects and Post-Production Industry*. Leonardo, 43(1), Cambridge.

Fritz, B. (2013, February 22). Visual Effects Industry does a Disappearing Act. *Wall Street Journal*. Retrieved from: http://www.wsj.com/articles

Fry, B. (2008). *Visualising Data*. O'Reilly Media Inc, Sebastopol, CA.

Grage, P. (2015). *Inside VFX: An Insider's View into the Visual Effects and Film Business (2nd ed.).* San Bernardino, CA: CreateSpace Independent Publishing Platform.

Gurevitch, L. (2015). The Innovation Engines: The Convergence of Science and Entertainment in New Zealand's Research Future. *The Journal of the Royal Society of New Zealand*. New Zealand.

Gurevitch, L. (2015). The Straw that Broke the Tiger's Back? Skilled Labour, Social Networks and Protest in the Digital Workshops of the World. *Routledge Companion to Labour and Media*. Routledge.

Gurevitch, L. & Spell, R. (2015). *Digital Workshops of the World Migration Big Data Visualization Tool*. Victoria University of Wellington, New Zealand.

Gupta, G. & Bhasin, R. et al. (2013). *And Action! Making Money in the Post-Production Services Industry.* ATKearney. Retrieved from: http://www.atkearney.com/communications-media-technology/ideas-insights/

Harwood, G. (2005). *Lung: Slave Labour* [Installation]. In: "Lung: Slave Labour", ZKM, Karslruhe, Germany, 3/20/2005 - 10/3/2005.

Hellstrom, S. K. (2013). *More Than Digital Makeup: The Visual Effects Industry as Hollywood Diaspora*. Stockholms universitet. Stockholm, Sweden.

Holmes, N. (1984). *A Designer's Guide to Creating Charts and Diagrams*. New York, NY: Watson-Guptill Publications.

Kaufman, D. (2013). *VFX Crossroads: Causes & Effects of an Industry Crisis, Part 1.* Retrieved from: http://library.creativecow.net/kaufman_debra/VFX_Crossroads-1/1

Leberecht, S. [HollywoodEndingMovie]. (2014, Feb 25). *Life After Pi*. [Video File]. Retrieved from: http://www.youtube.com/watch?v=9lcB9u-9mVE

LinkedIn. (2014, Oct 23). *LinkedIn User Agreement*. LinkedIn. Retrieved from: http://www.linkedin.com/legal/user-agreement

Manovich, L. (2015). Data Science and Digital Art History. *International Journal for Digital Art History*, (1). Retrieved from: http://www.dah-journal.org/issue_01.html

Manovich, L. (2013). *Software Takes Command.* New York, NY: Bloomsbury Academic.

NATS. [opal]. (2014, Mar 9). *24 Hour European Flight Traffic Visualization*. [Video File]. Retrieved from: http://www.youtube.com/watch?v=s2b06qtqpp4

Nayaranan, A. & Shmatikov, V. (2008). *Robust De-anonymization of Large Datasets (How to Break Anonymity of the Netflix Prize Dataset).* Proceedings, IEEE Symposium on Security and Privacy, p111-125. doi: 10.1109/SP.2008.33

Parish, S. (2015). *Practical Magic: The State of the VFX Industry in 2015, Part Two – Hug a VFX Artist.* Retrieved from: http://www.awn.com/blog/practical-magic-state-vfx-industry-2015-part-2-hug-vfx-artist

Prospect Visual. (2014). *Collecting Data from the Web – Is it Legal?* Prospect Visual. Retrieved from: http://prospectvisual.wordpress.com/2014/12/16/collecting-data-from-the-web-is-it-legal-2/

Roberts, J. (2014, Apr 1). *LinkedIn Names Company that used Bots to Steal Profiles for Competing Recruiter Service.* Gigaom. Retrieved from: http://gigaom.com/2014/04/01/linkedin-names-company-that-used-bots-to-steal-profiles-for-competing-recruiter-service/

Scott, A. (1998). Multimedia and Digital Visual Effects: An Emerging Local Labor Market. *Monthly Labor Review*, 121(3), CA.

Shneiderman, B. (1996). *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations.* Proceedings, IEEE Symposium on Visual Languages. doi: 10.1109/VL.1996.545307.

Squires, S. (2013). *Visual Effects Working Conditions Survey*. Retrieved from: http://effectscorner.blogspot.co.nz/2013/05/visual-effects-working-conditions-survey.html

Tufte, E. (2001). *The Visual Display of Quantitative Information (2nd ed.)*. Cheshire, CT: Graphics Press LLC.

VES. (2013). *The State of the Global Visual Effects Industry 2013, An Analysis of Current Business Models and Better Business Practices*. Sherman Oaks, CA: The Visual Effects Society.

Ware, C. (2004). *Information Visualization: Perception for Design (2nd ed.)*. San Francisco, CA: Morgan Kaufmann Publishers.

Wright, R. (2008). *Software Studies: A Lexicon*. Cambridge, MA: The MIT Press. pp 78-86.

Zwerman, S. (2009). *The Visual Effects Producer: Understanding the Art and Business of VFX*. Taylor & Francis, Burlington.

LIST OF FIGURES

[Figure 4.14] *Iteration Two Regional Density.* An overlay element that draws circles about hubs with radii indicative of the region's internal population.

[Figure 4.15] *Iteration Three Professional Composition and Regional Density.* The Unity5 variant allows for both overlays to be displayed simultaneously. In this version, populations are larger and professional breakdowns are more comprehensive, as evidenced by the greater number of roles.

[Figure 5.1] Opposite. *Averaged Regional VFX Populations.* Infographic comparing the lifetime average populations of VFX hubs across the period Jan 1990 - Jan 2016.

[Figure 5.2] Next. *Regional Growth in the VFX Industry.* Infographic comparing the population growth of VFX hubs across the period Jan 1990 - Jan 2016.

[Figure 5.3] Next Opposite. *Regional Composition in the VFX Industry.* Infographic comparing the lifetime professional compositions of VFX hubs across the period Jan 1990 - Jan 2016.

[Figure 5.4] Opposite. *Role Composition of Database Migrations.* Infographic illustrating the breakdown of all database migrations by profession.

[Figure 5.5] Previous. *Mobility in the VFX Industry.* Infographic providing statistical analysis of activity mobility and geographic mobility within the VFX industry.

[Figure 5.6] Previous Opposite. *Mobility Analysis of VFX Industry Professions.* Infographic comparing activity mobility against geographic mobility for all database professions.

[Figure 5.7] *Coding Literacy in the VFX Industry.* Infographic denoting the most popular coding languages and the coding percentages of professions within the VFX industry.

[Figure 5.8] *Software Proficiency in the VFX Industry.* Infographic denoting the most popular software packages and the software percentages of professions within the VFX industry.