Eugene H. Bang. An Analysis of Upwork Profiles: Visualizing Characteristics of Gig Workers using Digital Platforms. A Master's Paper for the M.S. in I.S degree. April, 2019. 35 pages. Advisor: Mohammad H. Jarrahi

In the transitional period of manual and algorithmic hiring, there has been an explosion of new employment opportunities. As a result of the ubiquity of mobile communication technologies, a gig economy has emerged which champions digital platforms as a solution to meet the burgeoning demand for on-demand workers, frequently called independent contractors or freelancers, by hiring organizations. Visualizations and linear regression are used to study information-rich Upwork profiles to determine variables that could predict how users maneuver in the gig economy. A typology of existing gig workers' motivations is combined with visualizations to better understand the situation of the typical gig economy worker.

Headings:

Digital Platforms

Gig Economy

Gig Work

Visualization

AN ANALYSIS OF UPWORK PROFILES: VISUALIZING CHARACTERISTICS OF GIG WORKERS USING DIGITAL PLATFORMS

by
Eugene H. Bang

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2019

Approved by

Mohammad H. Jarrahi

Table of Contents

1 – Introduction

Millennials average the shortest tenures in their job positions, often choosing to jump from one position to the next. In the age of Information, telecommunication technologies and digital communities enable workers to build reputability quicker than ever before from any location and the Internet provides upcomers with the resources necessary to develop technical skills. While college degrees once determined the viability of one's tenure at an organization, it is now becoming less significant. Google recently dropped the educational requirements for many of their most competitive and technical positions such as Senior Data Scientist roles and many other managerial positions – arguably because it is now possible that any user, given time and motivation, can learn the technical skills required to assume those positions by themselves without any physical boundaries (such as having to meet in-person). This has given rise a new trend of employment strategies which has enabled the creation of the "gig economy." A gig refers to a short-term employment opportunity by organizations which contract individuals for tasks based on their expertise in a subject domain. Tasks provided by employers can range from menial data entry and janitor work to providing regional consulting in finance to leading the development of advanced neural networks. This dichotomy in technicality allows gig workers to jump in at any skill level while simultaneously providing a news outlet of employment opportunities for those who require work. There is also the possibility that gig workers can convert to full-time employees for organizations that are particularly thrilled with their work.

Imagine a situation in which you wish to learn a new skill, test your abilities and build reputability simultaneously. The gig economy allows workers, particularly those in the process of looking for new jobs or flexibility, to learn while continuing to maintain a level of agency. Furthermore, gig work can bolster existing revenue streams for families that require more income by equipping these workers with the ability to discover “side gigs,” or supplementary jobs. Over time, experienced gig works can control their price and negotiate their hours/level of responsibility with their clients, thus providing them with choices,. A big question is how well digital platforms such as Upwork enable these transitional workers – the lack of transparency in algorithmic selection may serve to deter this mode of utility. However, this model of the “side gig” may be very valuable as it allows the next generation to become highly interdisciplinary, provide incentive to network and earn portioned revenue without the need of expensive interviews and matches. To reach this point, it is necessary to delve into the algorithmic black box to determine what types of attributes are favorable for such a model to exist and how both new and experienced gig workers can leverage their attributes to be selected for a particular position of their interest.

The data to be analyzed to study digital platforms will be Upwork user profiles. Upwork profiles contain many valuable pieces of information – hours worked, self-proclaimed technical skills and industry knowledge, wages earned, portfolio items, location data, and previous projects. A few visualizations of interest are average wages per year across the USA, wages for projects from inexperienced gig workers, and in particular, the attributes of $0 earned workers. Every gig worker has to start from nothing – how does one become recognized by algorithmic decision-makers and make a living? As of now, over 560,000 freelancers fall under this $0 unearned category of 595,000 total profiles. Digital platforms, and not Upwork in particular, are paving the way to a new standard of work culture in the future as workers transition from traditional education and workplace values to precarious, but diverse opportunities and digital learning. Many of the physical nuances of work is being displaced (ex. having to commute, being at work in-person, having inflexible hours) and it is important to discuss how digital platforms can serve as the transitional period model of employment which can greatly propel productivity and efficiency in connecting prospective workers with employing organizations. Such platforms provide a minimum safety net that helps offset the precarity of traditional freelance work, but in turn, have its own demands and constraints that must be met by gig workers.The research questions of interest are the following:

- What quantitative and qualitative variables correlate to success on digital platforms?
- Can we discover additional geographical variables that may be linked to increased gig work across the USA? For example, high population densities or traffic jams.
- What do digital platforms provide workers that helps them overcome precarious work? For instance, gig workers are not provided the benefit of insurance, retirement pension plans or a signing bonus. At what point is work precarity acceptable for gig workers?

2 – Literature Review

The literature review will cover attributes of the gig economy and gig workers that have already been established. Among these attributes are work precarity (labor rights and protection), potential motivations for gig work and new modalities of communication and contractual labor versus traditional employment. The influx of demand for tailored and mobile employment parallels the developments in machine learning and large-scale analytics as information technologies necessitate swifter action than before the Digital Age. The younger generations in particular have begun transitioning away from traditional long-term employment for more experimental modalities of working to fit a mobile lifestyle.

2.1 – Rise of the Gig Economy

A gig refers to a new modality of employment that undermines traditional lifetime careers by focusing on short-term contractual employment. The gig economy is a freelancing job market which sacrifices corporate loyalty and long-term incentives for a precarious networking environment which emphasizes self-paced growth, mobility, flexibility and uncertainty. According to Stanford (2017), there are five common attributes of work in the gig economy:

1) Employees work on an on-demand schedule where they only perform their services when organizations require them. Prior engagement does not guarantee full-time employment and each job may be the last in a while.

2) Employees must prepare their own equipment to perform their services which include a working environment, appropriate technologies such as a laptop or mobile phone, and any tools or software not directly supplied by the employer (ex. a license to proprietary software would likely be supplied to the employee while a microphone may not).
3) Employees are paid on a piecework basis which means that they will receive payment for their services once they complete a task or reach a certain milestone rather than for their time. This means that the quicker an employee can accomplish a task, the more value they receive. However, the opposite also holds true in that the longer a task takes, the more value they lose. This can result in potential exploitation if not carefully navigated.
4) Gig work is expedited by a digital platform which helps supply prospective employees to an organization at a given time. They can broker work between the parties and serve as an information bridge to keep them in contact
5) Digital platforms can be used to collect final deliverables and facilitate payments for the employee via an escrow system which helps guarantee payments for work that a freelancer

The digitization and subsequent mobilization of information through mobile technologies in the past two decades has greatly increased the demand for knowledge capital and for technical labor. Unlike traditional job markets which were bound by spatial and temporal parameters such as rigid work schedules and proximity to the workplace, the digital space is unique in that work can plausibly be done remotely or virtually (De Stefano, 2015). With the massive improvements in recent information communication technologies such as teleconferencing software like Skype, it is also possible to engage with peers in proxy face-to-face meetings. Wilson et. al (2006) found that while computer-mediated teams generally were more vulnerable to inflammatory verbal attacks, they still fostered a very acceptable level of trust through digital technologies alone. Computer-mediated communication can help reduce awkward encounters at work and help untangle personal and professional sentiments so that peers can have distance to cool off after a heated exchange. It is very possible then, to develop one's social capital virtually while transcending time and space. Consequently, demand for labor has become international – workers in the USA can work for a Chinese firm for example or connect with Californians while taking a trip in Thailand. This shift in mentality has given rise to a necessity for job hubs which can enable this new behavior.

Digital platforms formed as a hub solution for emerging on-demand workers. Digital platforms attempt to bridge the gap between supply and demand for gig workers by securing contracts for freelancers and allowing them to develop reputability over time through crowdsourced evaluations such as user reviews and public work evaluations. The platforms have introduced a new work ecosystem that significantly differentiates new gig work freelancers from traditional freelancers of the past. Digital freelancers seek flexible and remote work where they perceive themselves to exert more control over the pacing and quality of their work (Kuhn & Maleki, 2017). Unlike traditional freelancers, gig workers perform ad hoc piecewise tasks which means that their contractual jobs can become fragmented and at times, incomplete if the contract is terminated early for any reason. However, piecewise tasks can be established such that fixed-price projects can be paid for based on evidence provided by the freelancer that work was done in service for the organization.

Typically, freelance jobs are more supplementary by nature rather than primary, meaning that the tasks do not have a large effect on the overall operations of the organization though this is dependent on the size of the organization. For example, smaller and more obscure organizations are far more likely to provide a wide range of tasks at low wage points which make them fitting entry-level gigs to begin developing reputability, but unsuitable for those who wish to make gig work a full-time commitment immediately. Furthermore, digital gig work typically pays less on average than traditional freelance work of the same kind as time is no longer the rarest commodity, though this is also very dependent on a person's level of domain expertise. The freelance roles are generally more relaxed and freelancers take less accountability for the direction in organizational decision making (Barley et al., 2017), but are also frequently left in the dark for major managerial throughput. However, if that is not a priority, then gig work can be the start of a dynamic work schedule. If a gig worker finds a prospective employer, they can engage in conversation regarding the logistics of the contract and schedule accordingly. This is the point at which gig workers have the most control over their work. Their biggest asset to the workplace is their immense knowledge capital (Gandini, 2016) which allows them produce work that is unmarred by internal organizational culture and provide an outsider's perspective on progressive development in the organization. However, they are bound to digital platforms which broker the work between them and an employer which indicates that the digital platforms must enable an environment that is crucial for this new work ecosystem to succeed (Kalleberg and Dunn 2016). The digital platform imposes unique demands and restrictions on digital freelancers that influence their approach to managing their work lives. As a general online technical freelancing hub, Upwork is an interesting case study as the data will be least likely biased by domain-specific knowledge, allowing findings to be more applicable to a broader set of principles compared to other domain-specific digital platforms. It will instead focus more on the features provided by the platform such as its escrow service, guaranteed labor rights and algorithmic selection of jobs for the gig worker based on previous history.

2.2 – The Evolution of Human Capital in the Workforce

Erik Brynjolfsson and Andrew McAfee (2012) believe that currently, the world is in the midst of great technological advancement. Today, the typical consumer computer machine is incomparably more powerful than any machine of the last century and with each passing year, companies such as Intel, AMD and Nvidia are researching and developing stronger processing units that can revolutionize big data analytics and applications. Brynjolfsson and McAfee believe that technology of today snowballs and accelerates itself; As technology continues to improve, the rate at which technology advances will also increase. This has some very important ramifications for the gig economy. First, the rapid pace of technological advancement also precedes a rapid obsolescence of older technologies (Brynjolfsson & McAfee, 2014). Older programming languages such as C, Fortran and Haskell may be phased out as newer, faster replacements develop over time. Technology is phasing out older skills, forcing older and younger workers alike to learn dynamically and stay updated on current events so as to update their skillsets accordingly. The gig economy can thus serve as an indicator for technological progress and shifting tides in the tech world. Second, communication technologies are breaking down barriers across many different physical barriers such as proximity and as such, talent can cooperate with other talent from around the world given the networking opportunity. Digital platforms in the gig economy greatly expedite this by providing workers with hubs of authority that can reliably connect employers with prospective talent. Finally, the most consequential of all is that knowledge capital in the form of experience is no longer as sacred as it used to be. Seniority and long-term tenures at an organization are no longer guarantees for promotion (Cappelli, 2001) as data analytics has created a strong contrasting dynamic to older, experienced decision-making principles. Work has transitioned from experienced and knowledge-based decision-making to data-driven decision-making (LaValle et al. 2011) where principles and dynamic adaptation to technological changes is preferred to concentrated specializations. In other words, we have reached a point in society where change is normal and those who grasp the chances to be early adopters will be the most likely to succeed (Jones, 2002; De Stefano, 2015).

2.3 – Current State of Labor Protection

Upwork (2018) recently released their annual study, "Freelancing in America" which found that nearly one in seven Americans have freelanced at some point in 2018 and since 2015, more people have chosen to freelance by choice rather than by necessity.

Furthermore, the study found that freelancers valued lifestyle over income which seems to reflect an evolution of typical work-life balance in traditional work environments. If trends remain predictable, digital platforms will be in high demand for most any domain-specific task. Imagine gig handymen for home construction, freelance IT workers who come to your house to fix your computer or maybe even something as intriguing as a gig chef who cooks for you? The realm of possibilities is endless, but with each domain, there are also unique domain-specific nuances that have to be covered which is far too expansive and could each stand alone for comparison endlessly. In that regard, I am not focusing on these nuances, but rather on the most glaring issue in the gig economy at present that is shared among all users of digital platforms: laborer's rights.

Laborer's rights refer to the protections given to precarious workers for their employment which is to offset the loss of entitlements and benefits that could be had through long-term employment (Friedman, 2014; Kalleberg, 2009). Typically, traditional workplaces offer incentivizing benefits such as a retirement pension and insurance and larger organizations may offer amenities such as on-site gyms and cafeterias. However, most gig workers perceive such benefits to be less desirable than having the ability to travel, not commute for work and negotiate a flexible schedule. As more people adopt this lifestyle, it is exceedingly important that legalities regarding vague terminology such as "independent contractors" and their roles in hiring organizations become more lucid.

There is a historical perspective on this matter – precarious on-demand labor is not entirely new. For example, America's entry into World War II in 1941 necessitated significantly more military spending and by proxy, more demand for military products such as planes and rations. This sudden influx in demand for goods meant that jobs suddenly appeared overnight for an on-call workforce who could work in manufacturing plants to produce those goods. (Quinlan, 2012). However, the American government paid factory workers piecewise for goods produced and acted in favor of the exploitation of lower- and middle-class American citizens (Finkin, 2016; Alkhatib, 2017). These new workers were not full-time employees of these factories and were fully expected to return to a normal job at the closing of World War II and therefore not provided with basic protections such as insurance or over-time benefits. As labor rights for contractual employment was still hardly established in those times, the government exploited many of those workers by increasing their workload with meager increases in compensation, if at all. Even in 2019, there are many legal loopholes that have not been addressed which significantly stunts the positive impacts of the gig economy. Gandini (2018) noted that to avoid an exploitative capitalist management environment where the corporations exert significant control over their independent contractors, workers must establish a baseline negotiable power that gives them leeway when discussing the terms of their contracts with their employers to offset this issue. However, gig workers may not necessarily have the tools to do so at this time. Minter (2017) found that some digital platforms are best described as an escrow service where they primarily serve a middleman role of storing advanced payments from the employer and then disseminating payments to the freelancer when the organization approves of their work. However, this situation is still less than ideal as employers ultimately decide what type of work is satisfactory and can sway the flow of money in their favor (for example, reducing payments towards the freelancer due to “shoddy programming” when it may actually be satisfactory labor overall) with little room for the gig worker to retort. Often, the economic burden falls on the gig worker to produce evidence of the contrary and digital platforms rarely insure that such practices are prevented. Screenshots of the work, labeled documentation compilations and having an unbiased third party commenting on the quality of the good or service can all serve as evidence of providing labor for an organization, but nonetheless, the onus remains on the producer to prevent such situations from happening rather than on the digital platform. Digital platforms must rectify this by providing gig workers with some mode of control in the triangular relationship between worker, organization and escrow.

3 – Methods

This research analyzes Upwork profiles of gig workers to study how they are using the available digital platform features to best sell their skills. Qualitative values will be studied through cluster visualizations in Tableau to identify profile variables that have a positive correlation with success variables on digital platforms. Success variables are quantitative profile variables that can indicate whether or not a gig worker is able to make a sustainable foray into the gig economy. It is important to note that I will not be criticizing Upwork's policies nor am I judging their throughput. However, I will attempt to hypothesize why some workers are successful on the platform while many others struggle to make breakthroughs.

3.1 – Data Collection

Upwork is the digital platform I chose to analyze because it does not have domain-specific attributes. Jobs brokered through Upwork range from a variety of different domains which can range from software engineering, web development, image editing accounting, creative writing and more. Unlike other digital platforms such as Uber (rideshare) and Mechanical Turks (human intelligence tasks), Upwork does not necessarily bias itself to any domain and therefore must provide features that can complement the work of all the gig workers in those fields. This can also allow me to drill down to the bare minimum features that are most utilized and determine if there are any domains in particular that greatly benefit from the usage of digital platforms. Using Python and the BeautifulSoup library, I scraped 23 pages of 425 profiles per page for a total of 9,775 Upwork profiles in .csv format.

3.2 – Data Analysis

First, I combined the 23 .csv files together to create an aggregated Upwork profile dataset. During the process, 78 rows were incorrectly parsed so they were thrown out as they had no quantitative values attached to important fields such as income per job. As there were less than 10,000 rows, I opted to use Excel to clean the profile dataset and renamed several column headers to more descriptive monikers. Due to the massive gap between the highest and lowest income values, I normalized values by dividing total revenues by the total number of jobs which was calculated by adding hourly and fixed-price jobs together. Afterwards, the data was loaded into Tableau to create several visualizations to explore the correlations between the profile variables. In particular, I want to focus on variables that have a noticeable dichotomy between $0 income profiles and the money-makers. I wish to highlight geographical differences using a choropleth to outline the distribution of gig work across the United States. Two choropleths will be made with the first using data for income per job while the second will focus primarily on the total number of jobs taken per state. Next, I will create several scatterplots to highlight average income per gig worker on Upwork and cluster groups together based on income and the count of profiles. The next scatterplot will be used for regression analysis on the total number of passed tests, portfolio items and search rank. Previous research hypothesized that as hiring becomes more algorithmic, workers that have many accolades or certifications on display on their profile would be considered more favorably than other profiles without them (Wood et al., 2019). This is important to research as any statistically significant findings would indicate that algorithmic hiring is skewed towards these more embellished profiles and would necessitate that competitive gig workers obtain these accolades for themselves as well.

Rank	Short Name	Title	City of Industry	State	Total Revenue	Income Per Job	Total Jobs
408	Isabella W.	Professional Administrative /	Andover	KS	$343,363.21	$343,363.21	1
751	Ashish C.	B2B2C| InsureTech| .NET| MI	White Plains	NY	$40,318,318.68	$325,147.73	124
538	Travis B.	Fractional CTO and Technolo	Breaux Bridge	LA	$235,369.51	$117,684.76	2
572	Neeraj A.	Php/Wordpress/Magento/.N	San Mateo	CA	$2,417,236.67	$105,097.25	23
365	Stas M.	Expert in Web & Mobile Deve	New York	NY	$1,078,325.77	$89,860.48	12
897	Siddiq M.	Sr. ColdFusion Developer wit	North Brunswick T	NJ	$66,647.72	$66,647.72	1
910	Haley W.	Creative, Intuitive Copywrite	Muscle Shoals	AL	$62,280.61	$62,280.61	1
848	Evan S.	Full Stack Web Developer	Akron	OH	$56,723.09	$56,723.09	1
100	Jon Paul L.	Banner Ad Animation, HTML5	Austin	TX	$163,796.01	$54,598.67	3

Figure 1. Partial Snippet of Scraped Upwork Profile Data

<u>Rank</u>: the order in which they were found by the scraper. Rank N is the Nth search result.

<u>Title</u>: worker's current title, frequently used as a field to highlight industry knowledge.

<u>Income Per Job</u>: Total revenues / Total Jobs

<u>Total Jobs</u>: Hourly Jobs + Fixed-Price Jobs. Both types of jobs are important to study individually as well.

A glaring flaw, but unfortunately unavoidable, is that I cannot obtain temporal data for these profiles as they are snapshot scrapes of their most current profiles and Upwork's site does not provide information about how long a gig worker has been a part of the site. This means that regardless of any findings, I cannot conclusively state that any one variable at this time will be statistically significant in driving success. I fully expect that my correlations and regression analyses for these variables to return incredibly low because success is highly dependent on time and experience. The highest earners may only have one job on record, for example, but may have had multiple points of contact previously with an employer and have had a single contract for multiple piecewise tasks. As such, they heavily skew data towards low job count, high income. This skew may affect the perceptibility of low income, high job count workers who may have middling incomes compared to the highest earners, but have perfectly comfortable wages. Again, since there is no temporal data, I cannot state that higher income necessarily correlates with job counts either. For example, a total revenue of $1,000,000 dollars over 10 jobs turns to an average of $100,000 per job. But the reality is that it was likely more skewed where the first few jobs were low-wage jobs and over time, workers learned skills that increased their ability to perform technical tasks. We also do not know if all 10 of those jobs were taken in one given year. With only a single year's worth of work, $1,000,000 is an incredibly successful foray into digital platforms. However, over 10 years, it tells a slightly different story (though $100,000 a year is still very impressive). Given temporal data, I would have looked into average increases in revenue per year on Upwork and additionally analyzed membership programs such as the "Top Rated" and "Rising

Talent" programs which highlight specific Upwork profiles during searches. For now, I am limited to search rankings, spatial data using the states and total income and jobs data.

3.3 – Results

I began by analyzing my cleaned data and broke down several key statistics of interest. Tests in this case refer to online tests of any sort such as a Windows Competency Test, certifications from Cisco or Red Hat, typing tests and literacy tests. These accolades may also be used as a proxy for experience in a given domain. First, I'll begin with a sample of the passed tests dataset.

[{"name": "Python Test", "id": "15666122", "percentile": 78, "takenOn": "2017-01-03T00:00:00.000Z", "score"
[{"name": "PHP5 Test", "id": "6100644", "percentile": 88, "takenOn": "2012-12-14T00:00:00.000Z", "score": 82
[{"name": "Java Test", "id": "1348618", "percentile": 93, "takenOn": "2015-12-07T00:00:00.000Z", "score": 93
[{"name": "AngularJS Test", "id": "16200175", "percentile": 90, "takenOn": "2017-04-19T00:00:00.000Z", "scor
[{"name": "C# Test", "id": "13783916", "percentile": 98, "takenOn": "2016-01-27T00:00:00.000Z", "score": 98,
[{"name": "Adobe After Effects 7.0 Test", "id": "16685312", "percentile": 74, "takenOn": "2017-08-06T00:00:0
[{"name": "ASP.NET Test", "id": "18998821", "percentile": 93, "takenOn": "2019-03-04T00:00:00.000Z", "score
[{"name": "Swift Test", "id": "18592921", "percentile": 90, "takenOn": "2018-11-27T00:00:00.000Z", "score": 9
[{"name": "Customer Service Test (Old)", "id": "14491930", "percentile": 70, "takenOn": "2016-06-01T00:00:0
[{"name": "Email Etiquette Certification", "id": "17617958", "percentile": 80, "takenOn": "2018-03-26T00:00:
[{"name": "Java Test", "id": "17228305", "percentile": 89, "takenOn": "2017-12-20T00:00:00.000Z", "score": 89
[{"name": "Blog Writing Test", "id": "16906972", "percentile": 68, "takenOn": "2017-09-28T00:00:00.000Z", "s
[{"name": "Customer Service Test", "id": "18223767", "percentile": 75, "takenOn": "2018-08-28T00:00:00.000
[{"name": "Quick Books Pro 2008 Test", "id": "929709", "percentile": 82, "takenOn": "2009-12-11T00:00:00.00
[{"name": "Customer Service Test (Old)", "id": "9062656", "percentile": 52, "takenOn": "2014-01-30T00:00:0
[{"name": "Advanced PHP Test", "id": "1857913", "percentile": 97, "takenOn": "2010-11-06T00:00:00.000Z", "
[{"name": "Customer Service Test (Old)", "id": "10842970", "percentile": 72, "takenOn": "2014-09-19T00:00:0
[{"name": "Ruby on Rails Test", "id": "17271274", "percentile": 94, "takenOn": "2018-01-02T00:00:00.000Z", "

Figure 2. Sample Passed Tests dataset

This dataset shows important data such as the name of the test, the recency of the test, as well as a percentage of the score. Upwork's front-end scripts displays scores not by the numerical scores, but by their percentiles such as “Top 10%.” This negates the slight advantage that marginally higher scores would have had over others who had taken the same test. This is the first dataset that was used to create a set of visualizations to study how it affected associated revenue and number of jobs.

	Total	0 Tests	1 Test	> 1 Tests
# Profiles	9,697	4,523	1,667	3,507
% of Total	100%	47%	17%	36%
Assoc. Revenue	$232,073,145.25	$33,721,405.16	$31,435,178.69	$166,916,561.40
Revenue %	100%	14%	14%	72%
Hourly Jobs	96,821	16,246	18,815	61,760
Fixed-Price Jobs	199,611	32,899	43,948	122,764

Figure 3. Passed Tests v. Success Variables

The first thing to notice is that nearly half of all of the randomly scraped profiles have 0 tests passed on their profiles and have an associated revenue flow of $33.7 million dollars across 4,523 profiles. According to the distribution, 47% of the profiles together only amount to 14% of the total revenue earned. In contrast, the 1-test passed group compose a third of the number of profiles, but also earn significantly more revenue per person. To find just how much more money per capita, I simply divided associated revenue with the number of profiles and found that 0-tests earn $7,455.54 per person while 1-tests earn $18,857.34 per person. However, I was wary that one simple test could increase income per person by more than 150%. What I found was that the 0-test group had a significant number of outliers which I also found in my Tableau visualizations. However, this was also the case for subsequent test groups as well. A previous hypothesis by Wood et al. (2019) was that as hiring transitions from manual sifting by humans to algorithmic selection by machines, there has to be differentiating variables that can help specific individuals in a sea of applicants stand more conspicuously. In this case, personal accolades, user reviews and subjective scores as well as digital platform programs (such as the Rising Talent program: https://support.upwork.com/hc/en-us/articles/211063228-Become-a-Rising-Talent) could be used to alter algorithmic behavior in determining what profiles should be highlighted over others. This can have significant bearing over who is successful and who is not. This highlights a unique power dynamic that digital platforms demand of its users – that is, users must be knowledgeable about the metrics that the platforms use to rank them. The metrics themselves are a black box at this point, however. Despite having more tests passed and more jobs taken, my regression analysis found that there is nearly no correlation at all between passed tests and revenue. I will discuss the importance of this data in my visualizations. The data by totals clearly show that important variables such as hourly and fixed-price jobs provided to 1-test users is increasing. Is this being accounted for in algorithmic selection? It highlights the need for transparency – not necessarily for digital platforms to release their algorithms, but for users to be given a sense of direction. The Rising Talent program, for example, provides key information about how to apply to it and states that it provides users with a distinct badge that helps their profiles become more valuable. This is very similar to why certifications are extremely valued in manual hiring through digital mediums. They often seek validation and verification of quality through external sources. If you were given two project managers, one with his PMP certification and another without it, in a black box situation where no other variables are known, which are we more inclined to choose? The certifications serve as a proxy for previous evaluations and judgment of character (Bozdag, 2013) – obtaining a PMP is a rigorous process and not one that average citizens can afford to do halfheartedly unless they have the skills mastered and it biases our perception of their abilities. Just the same, the rewards from programs offered by digital platforms such as the Top-Rated badge proxies as an evaluation of the profile by the digital platform. This means that employers have to trust the validity of Upwork's principles and metrics and once that trust has been established, it changes the dynamics of how users must interact with the digital platform or they will be left behind. Suddenly, employees want better certified profiles. They want proof through digital proxies such as online examinations, certifications from critically-acclaimed institutions and crowd-sourced reviews. No longer are these badges just slight advantages over other profiles, they are necessities. And once algorithmic selection becomes the norm, it is very important that gig workers are made aware that such changes are being made. For example, Upwork could do this through their annually commissioned study, "Freelancing in America" and distribute descriptive statistics regarding typical accolades of successful gig workers or applicant numbers for their programs. Digital platforms, and really any organization making use of algorithmic selection, should make gig workers aware of the usability dynamics of digital platforms and how to best manage their profiles to obtain success.

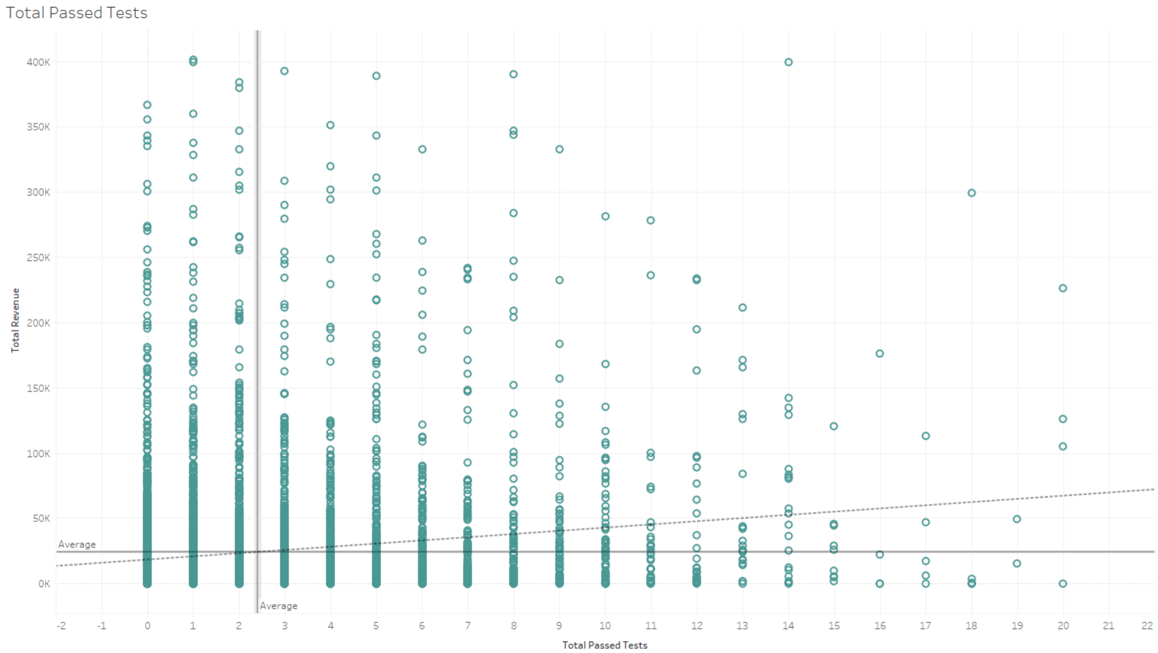

Figure 4. Clustered Total Revenue (y-axis) vs. Total Passed Tests (x-axis)

The visualization shows that of the 9697 profiles, a majority of them had more than one passed test. The distribution of revenue is, however, quite even. This indicates that the problem is not that high paying jobs are unavailable for these workers, but rather that there are significantly more profiles in that cluster to compete with. If we return to Figure 2, we want to highlight the number of profiles again. As the total revenues are very close in value, we can see that of the group of gig workers that were selected for work in the 0-test group, they made as much revenue as the 1-test group. Also, note how the 0-test group had fewer jobs across the board as well. This therefore implies that being selected, and not the amount of money earned per job, is more important. Once a gig worker is selected, we can assume that they will begin to profit, albeit slowly from the beginning. Again, I cannot confirm this without temporal data to show average increases in revenue per year for a typical gig worker on Upwork.

The averages with 95% confidence are total revenue = $24,186 at 2.41 tests passed. The r-squared value is a meager 0.022 and the p-value is < 0.0001. The significantly small r-squared value indicates that the scatter of data around the regression line is all over the place with seemingly no correlation. However, the incredibly small probability value indicates that there is still a real relationship between the predictor variable (passed tests) and the response variable (revenue). I interpret this as saying that the variables do have a correlation, but require additional predictor variables to better explore the patterns.

Following this, I've created a visualization comparing income per job and the total number of jobs per person, colored by total revenue earned.

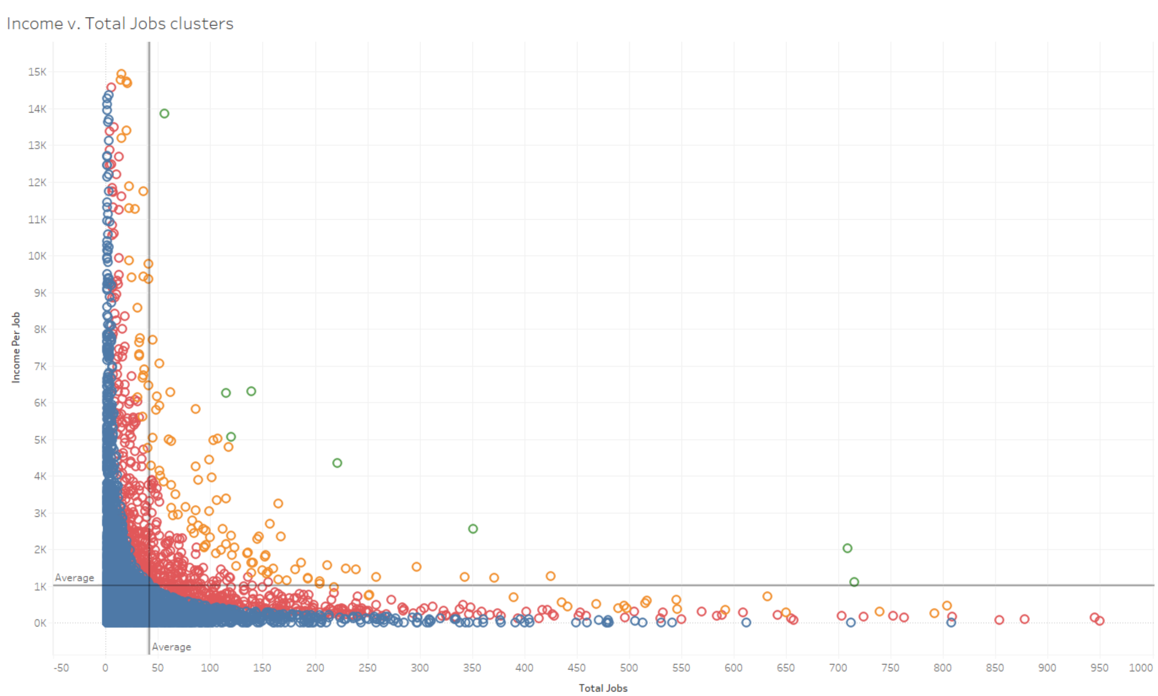

Figure 5. Income Per Job and Total Jobs Cluster

The visualization was aimed to explore how taking on more jobs over time may or may not result in additional income over time. There are four colors total: blue, red, orange and green. Each represent in that order represent a higher magnitude of total revenues earned by that person.

- <u>Blue</u>: $0-$50,000 total revenues; low-barely sustainable revenue, supplementary income
- <u>Red</u>: $50,001-$200,000 total revenues; somewhat sustainable revenue, consistently obtaining decently paying jobs.
- <u>Orange</u>: $200,001-$600,000 total revenues; high total revenues, consistently obtaining well-paying jobs
- <u>Green</u>: $600,001 total revenues; extremely high total revenues, hired at varying (but still expensive) prices

To interpret this, blue circles at the top-left of the scatterplot represent rising stars who have already found some modicum of sustainable income. Extremely lopsided skews (ex. 1 job, $1 million dollars) were removed as they were not representative of the typical gig worker and more focus was placed into encapsulating the majority of data points. Note that there is a large concentration of blue at the bottom left even as the number of jobs increase. This indicates that long-term gig work does not ensure sustainable revenue over time and most users do not make it out of this cluster. The position of the dots provides a temporal measurement for income, though I cannot determine how many jobs are taken per year. The most successful Upwork gig workers are those found as orange or green circles. It indicates sizeable revenues and consistent hiring by external sources which is indicative of a sustainable lifestyle. Next, our average value with 95% confidence can be found at $974 per job at 39.5 jobs taken. Ideally, I would have liked to scrape additional data such as how long a job took on average based on hourly v. fixed-price, but for now, this information states that approximately 47% of Upwork workers have made about $38,473 since they've begun using Upwork. Later analyses should use temporal data to find average annual income based on the clusters as that may tell a vastly different story. To add onto the interpretation, I chose to look into existing typologies of gig workers. Dunn (2016) developed a typology of gig workers that I found extremely useful to grasp the utility of digital platforms and how they can be designed to help reduce work precarity.

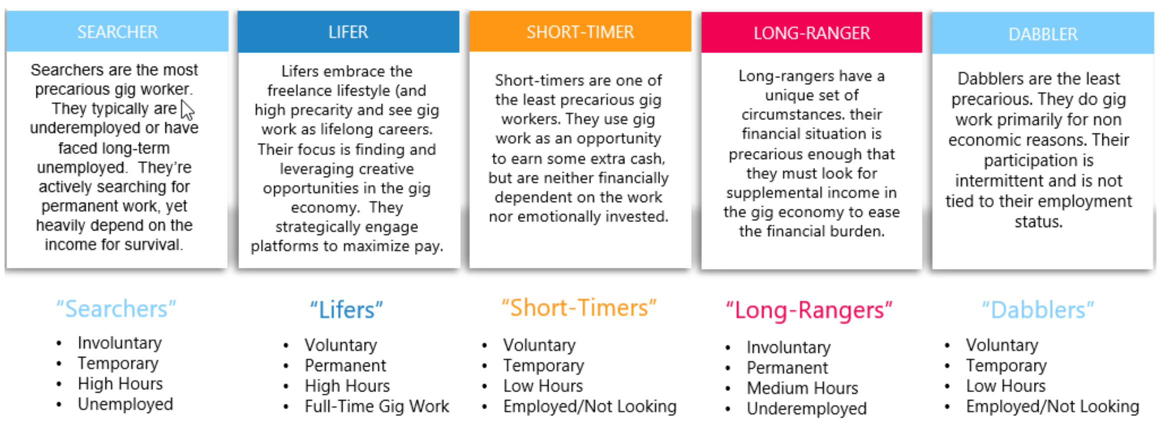

Figure 6. Typology of Gig Workers

This typology denotes motivations. For example, a blue circle found in the bottom-middle of Figure 5 could denote a short-timer or dabbler. We have a lot of information to work with. First, blue colors represent low total revenue and as you go further right, the number of jobs taken increases. If this person was a searcher, he would likely have found a different alternative to Upwork over that long stretch of time. He also is not likely to be a lifer as they are known to adopt strategies to maximize their annual income and is therefore much more likely to increase their revenues significantly over time. Finally, they may be a long-ranger, but as with searchers, they are primarily motivated by economic principles. There are many additional digital platforms available and for many industries so it is unlikely that these groups would stay complacent with their situations. That leaves me with a short-timer or a dabbler. Dabblers may seek to improve their skills in a specific domain and consequently search for jobs on the side to improve those skills. Short-timers are similar to moonlighters, or people who work a second job at night, except they are not financially burdened. Like dabblers, they could be looking to improve their skills in a given area while making some money on the side. The quantitative values such as total revenues, jobs worked and average income per job can be used to tell a story of gig workers. They can be manipulated to derive a potential set of motivations that can be used to attract and design for these types of gig workers. Of course, one major missing piece is the lack of time. Jobs taken is a proxy for time, but it’s a very poor substitute for actual temporal values. For the example above, we could more definitively categorize that worker if we could determine if they had worked for 3 years or 10 years to complement our existing visualization. However, the greatest takeaway here is that it doesn’t require much “insider” information to be able to visualize the types of people flocking to the gig economy and by knowing what motivates them, developers, and by extension, digital platforms, can also learn to aid them in their pursuits.

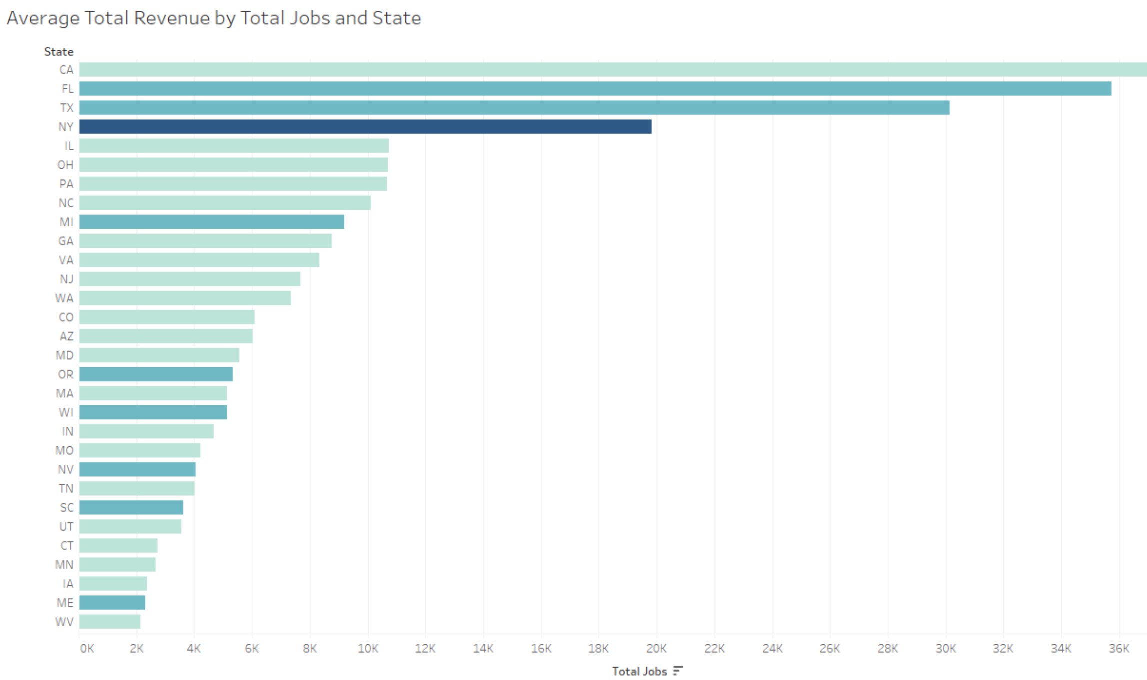

Figure 7. Average Revenue by Jobs and State

The biggest insights come from the top 10 states – they are all the most populous states in America. In fact, they nearly came out in the same order as well. To explain the horizontal bar graph, the height of the bars is determined by the total number of jobs and colored by average total revenue per capita. In Figure 7, that means that while California provided the highest number of jobs available on digital platforms, the average revenue was middling and was more comparable to other states such as North Carolina and Georgia despite its massive population and its industrial anchors such as Silicon Valley and Hollywood. On the other hand, New York has a bit more than half of California's jobs, but has the highest average revenues per capita by far, outstripping even Texas and Florida in a time when New York's population is beginning to decline as people migrate to other states in search of better livelihoods. This is very exciting as it provides several interesting hypotheses as to why this may be. First, it implies that these top states have unique attributes that differentiate them from other, smaller states. Population has a clear correlation to the number of jobs available and while not as linear, revenue does increase accordingly as well as a result of the increasing supply of jobs. This may be a result of increasing population density. For example, Cervero (1989) noted that commuting distances and times were already beginning to increase with minor regional mobility. States are frequently ill-equipped to upscale or downscale their infrastructure and this can be seen in commuting times in California and New York as well as housing prices (http://overflow.solutions/demographic-data/national-data/state-level-analysis/what-states-have-the-longest-commute-times/). With an increasing population, it is expected that commute times will increase year to year and may serve as one of the primary motivations for gig work in these states. It has been frequently stated in interviews of gig workers that time and spatial factors contribute the most to their decision to transition into digital remote work (Kalleberg & Dunn, 2016; De Stefano, 2015). Secondly, the bar graph results imply that industrial specialties do play a significant role in pay as well. New York's Wall Street, California's Silicon Valley and Hollywood, North Carolina's Research Triangle Park and so on are hubs that attract specific types of talent. It is unsurprising that New York would top this list as the financial hub of the United States where many of the most profitable and high-stakes disciplines such as portfolio management and quantitative analysis are grounded. However, it is now possible to work for organizations in these hubs without the need for close proximity or relocation. This reduction in financial burden can also enable behaviors that would otherwise be considered too expensive. For example, New York City's average apartment rent was an incredible $3,667 per month in 2018 (https://www.businessinsider.com/manhattan-rent-by-neighborhood-ranked-from-lowest-to-highest-2018-5). This money, hypothetically, could be used to purchase equipment that enable gig work. This could inject new talent into these industries, particularly those that could not handle the economic burden of relocation but have promising talent. Paired with population information, this can be used to complement current research within the job industry that can denote a transition from analog to digital platforms or other unique employment environments.

4 – Conclusion

I knew going forward that there were going to be many variables that could be used to improve the project. Unfortunately, I was unable to scrape information regarding user reviews and satisfaction ratings, but assumedly, they would have very similar effects as most other crowd-source-driven review systems. For example, Ye et al. (2009) found that online user reviews significantly improved or deterred hotel room sales. Consumers began to judge the business by perceived quality such as having high user ratings and positive and negative user comments – features that are available on individual user profiles to view, but are difficult to scrape from their website without time-consuming manual labor. These are improvements that can easily be made to my visualizations. More importantly, the visualizations show that it is possible to study qualitative and quantitative relationships in variables at a holistic level without requiring insider knowledge about the industry. In fact, I learned along the way that research does not require immense amounts of specificity and dense collections of data, but rather, the ability to connect two points together using logic and possibility. Available details on publicly available profiles are valuable information sources that can be analyzed using a variety of different information tools and create digital stories that tells the current trends and livelihoods of gig workers. Throughout the project, I found significant number of qualitative research studies by acclaimed researchers, but rarely did they invoke quantitative analysis as more than a means to validate facts. There is an increase in population here, more gig work. There is a trend in metropolises where people are preferring to rideshare rather than own their own cars, more gig work. However, I found that using visualizations can help weave stories regarding the lives of these precarious gig workers. Their experiences are not easily explained through semi-structured interviews and surveys and such research often cannot explain the systemic bias of digital platforms. Using quantitative values in visualizations can help ascertain that certain patterns do exist and due to the ubiquity of information, it is possible to continue to build on top of existing visualization models and analytical systems which will only improve as time goes on. My methodology used here is, of course, inherently flawed from the beginning from the lack of temporal data, but would you argue that there is no possible utility of the information that we derived from the visualizations? I would argue not; this transitional period in the gig economy is the ideal time to test any and all of our hypotheses and to see how the world evolves to these developments. In the end, this was a project of serendipity. I knew full well that derivations would be fortuitous and not by my own design, but I believe that the basic quantitative values alone match well with the trends of the real world. Those details alone, no matter how minor, can tell a very interesting story.

Bibliography

Alkhatib, A., Bernstein, M. S., & Levi, M. (2017, May). Examining crowd work and gig work through the historical lens of piecework. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 4599-4616). ACM.

Barley, S. R., Bechky, B. A., & Milliken, F. J. (2017). The changing nature of work: Careers, identities, and work lives in the 21st century. *Academy of Management Discoveries.*

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and information technology*, *15*(3), 209-227.

Brynjolfsson, E., & McAfee, A. (2012). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Lexington, MA: Digital Frontier Press.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Cappelli, P. (2001). Assessing the decline of internal labor markets. In *Sourcebook of labor markets* (pp. 207-245). Springer, Boston, MA.

Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association*, *55*(2), 136-150.

De Stefano, V. (2015). The rise of the just-in-time workforce: On-demand work, crowdwork, and labor protection in the gig-economy. *Comp. Lab. L. & Pol'y J.*, *37*, 471.

Finkin M (2016) Beclouded work in historical perspective. *Comparative Labor Law & Policy Journal* 37(3): 603–618.

Friedman, G. (2014). Workers without employers: shadow corporations and the rise of the gig economy. *Review of Keynesian Economics*, *2*(2), 171-188.

Gandini, A. (2016). Digital work: Self-branding and social capital in the freelance knowledge economy. *Marketing theory*, *16*(1), 123-141.

Gandini, A. (2018). Labour process theory and the gig economy. *Human Relations*, 0018726718790002.

Jones, S. (2002). *Internet goes to college: How students are living in the future with today's technology*. Diane Publishing.

Kalleberg, A. L. (2009). Precarious work, insecure workers: Employment relations in transition. *American sociological review*, *74*(1), 1-22.

Kalleberg, A. L., & Marsden, P. V. (2015). Transformation of the employment relationship. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 1-15.

Kalleberg, Arne L. and Michael Dunn. (2016). "Good Jobs, Bad Jobs in the Gig Economy." *Perspectives on Work* 20(2):10–14.

Kuhn, K. M., & Maleki, A. (2017). Micro-entrepreneurs, Dependent Contractors, and Instaserfs: Understanding Online Labor Platform Workforces. *Academy of Management Perspectives*, 31(3), 183–200.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, *52*(2), 21.

Minter, K. (2017). Negotiating labour standards in the gig economy: Airtasker and Unions New South Wales. *The Economic and Labour Relations Review*, *28*(3), 438-454.

Quinlan M (2012) The 'pre-invention' of precarious employment: the changing world of work in context. *The Economic and Labour Relations Review* 23(4): 3–24

Stewart, A., & Stanford, J. (2017). Regulating work in the gig economy: What are the options? *The Economic and Labour Relations Review*, *28*(3), 420-437.

Stanford, J. (2017). The resurgence of gig work: Historical and theoretical perspectives. *The Economic and Labour Relations Review*, *28*(3), 382-401.

Wilson, J. M., Straus, S. G., & McEvily, B. (2006). All in due time: The development of trust in computer-mediated and face-to-face teams. *Organizational behavior and human decision processes*, *99*(1), 16-33.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: autonomy and algorithmic control in the global gig economy. *Work, Employment and Society*, *33*(1), 56-75.

Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, *28*(1), 180-182.