Projecting Job and Wage Losses Due to Trucking Automation

Anhvinh Doanvo, Jessica Young, Rohan Panikkar

Advised by Dr. Lee Branstetter

CARNEGIE MELLON UNIVERSITY 5000 Forbes Ave, Pittsburgh, PA 15213

Table of Contents

Executive Summary

This report was produced as part of the Methods of Policy Analysis (MOPA) project course at Carnegie Mellon University's Heinz College on behalf of the Keystone Research Group, our client. Our client asked us to estimate job and wage losses over time in the trucking industry due to automation, and to conduct a preliminary exploration of a possible wage insurance scheme. By making these projections, we hoped to inform policymakers of a framework they may use to project job/wage losses over time and to develop literature on the scale of change to come.

We were particularly interested in making projections exploiting our understanding of the heterogeneity of the trucking industry. Although all jobs of the industry share some fundamental characteristics, differences between distinct types of trucking jobs are substantial enough to warrant discrimination between them. For example, we assess that long-haul truckers are far more likely to be embroiled in job losses due to automation because their work is repetitive and occurs over long, strenuous hours. Therefore, we needed a model whose parameters could be adjusted to account for diversity among the trucking industry.

The equation for the sigmoid curve (henceforth, the s-curve) appeared to be best-suited for our analysis, as it models change (i.e., job loss or technological adoption) in three phases: 1) an early phase, where adoption is low and slow, 2) a middle phase, where adoption accelerates rapidly, and 3) where adoption has approached its peak and its pace once again slowed. The s-curve's parameters can be easily modified to reflect the projected change across different categories of truckers. And by summing up the s-curves for each individual category and locality, we can model job loss over time.

We thus identified five attributes to split truckers by, resulting in 32 different job categories:

- Distance traveled: short vs. long haul
- Load type: full-truckload (FTL) vs. less-than-truckload (LTL)
- Specialization: specialized or generalized freight
- Unionization: unionized or not
- Owner-operator status: whether a trucker is an owner-operator (i.e., self-employed) or not

Asking subject-matter-experts to provide s-curve parameters for each of the 32 different job categories is not practical. It is, after all, difficult to conceive of a category that meets all of the following characteristics: truckers that are short haul, FTL, specialized, unionized, and not owner-operators. We therefore created a workflow that streamlined the development of our required s-curve parameters by treating the parameters as additive (or multiplicative, as appropriate) composites of the five attributes' influences. This is explained in more detail in the subchapter, "Using Sigmoid Curve Theory to Model Job Loss", under "Estimating Job and Wage Loss". With these s-curve parameters, we needed only identify the quantity of truckers per locality per subcategory to project job losses over time.

However, national-level population statistics splitting truckers by all five of these attributes do not exist, let alone datasets splitting truckers by these categories and across localities (counties, commuting zones, etc.). We found that two datasets, in particular, might be useful for resolving this gap: the Bureau of Labor Statistics' (BLS) Occupational Employment Statistics (OES) database and the BLS's Quarterly Census on Employment and Wages (QCEW).

Each individual dataset, on its own, is insufficient for this problem set. The OES dataset attempts to estimate the true job count of truckers in each locality, but its statistics are not broken down by any subcategory of truckers. And it may be tempting to solely use the QCEW dataset because it *does* have some of the required subcategory breakdowns in each locality. Indeed, other attempts at estimating job loss due to automation have relied solely on this dataset. But QCEW's statistics are derived solely from employees of establishments that are paying into unemployment insurance programs—a subset of truckers. Using QCEW alone would result in systematic *underestimates* of job loss and fail to reflect the true count of truckers in each locality.

We therefore constructed a process to combine the best of both worlds: we pulled the estimates of the true counts of truckers (adding in owner-operators, which are not counted by OES), in each locality, and merged them with the proportions of truckers belonging to each subcategory of truckers, again, in each locality. This was a nontrivial task, requiring methodological judgments on how these datasets should merge and how we should reconcile differences between them as they appear. But this enabled us to make a considerable refinement relative to other estimates, splitting truckers along the attributes of distance, load type, and specialization. No locality-specific data on unionization and owner-operator status exists, and so we assumed that values for these attributes are evenly distributed throughout the U.S.

At this point, we had had s-curve parameters derived from our consultations with Dr. Steve Viscelli, an expert on the trucking industry's history at the University of Pennsylvania[1]. And with our QCEW-OES data integration process, we derived population statistics per-locality and per-job subcategory across the thousands of counties and the 32 job subcategories. We inputted this data into our s-curve model to calculate wage and job loss statistics. Wage losses were calculated as the difference between the mean wages of truckers in each locality and the mean wages of high school graduates in each locality. All county-level statistics were aggregated at the commuting zone level to better reflect that individuals regularly cross county lines for work.

We then utilized an existing model for a hypothetical wage insurance program to generate preliminary statistics on the cost to insure against wage losses among truckers. Under this model, we explored how much such a scheme might cost under different scenarios of slow, moderate, and fast adoption, different coverage levels (i.e., 50% or full coverage of wage losses), and different timeframes (wage insurance for the first year, five years, or 10 years after losing a job).

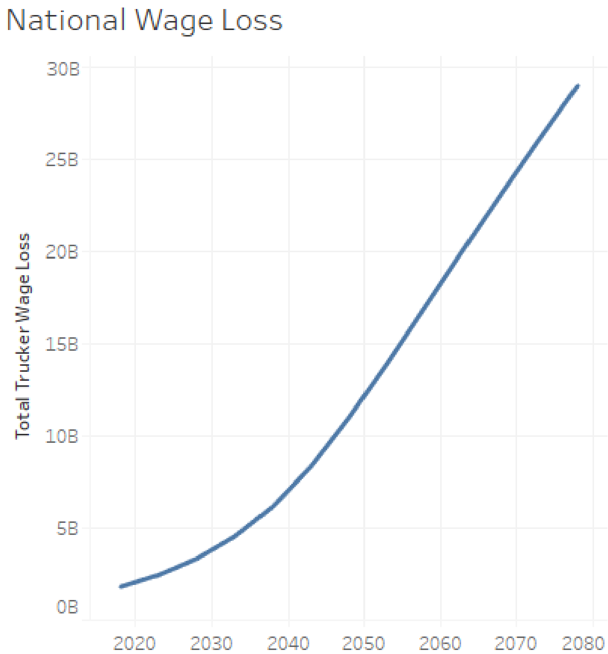
Our final analysis suggests that as many as 300,000 job losses will occur by 2033, 900,000 job losses by 2053, and 1.8 million job losses by 2078. Furthermore, we estimate that annual wage losses on the order of $4.5 billion may occur by 2033, $14 billion by 2053, and $29 billion by 2078. Very preliminary models also suggest that a wage insurance program might be affordable.

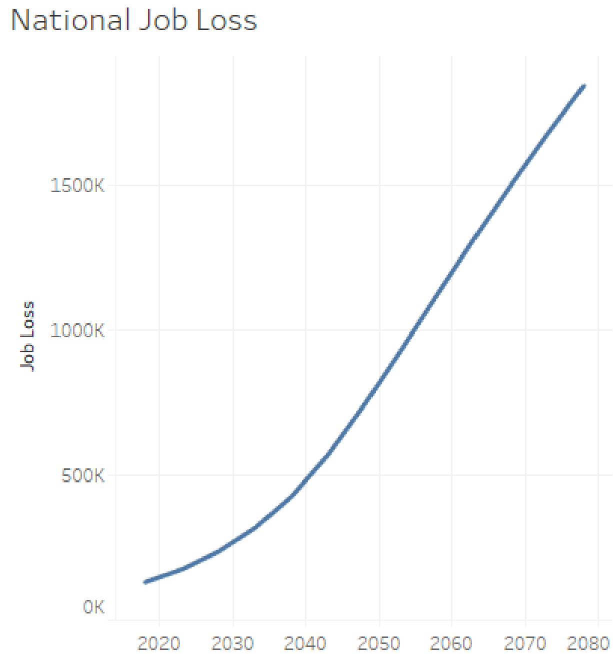
---

[1] Dr. Steve Viscelli is a senior fellow at the University of Pennsylvania's Kleinman Center for Energy Policy. His expertise lies in labor market economics. He is the author of *The Big Rig: Trucking and the Decline of the American Dream*, which chronicles the changes the US trucking industry has undergone over the past several decades.

Estimated National Job and Wage Loss Over Time

Jobs losses are estimated counts. Wage loss is in 2018 US dollars.

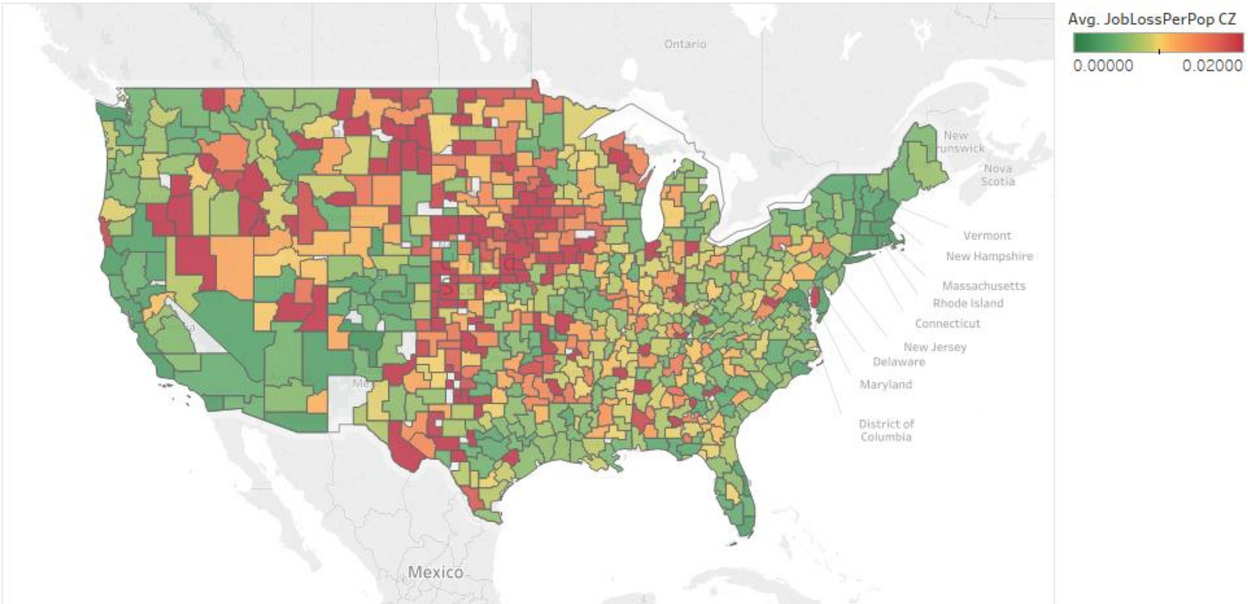




We have also generated dashboards mapping these job and wage losses by commuting zone, and computed metrics on the relative impact of these losses on the local economies. These metrics include the impact of truckers' wage loss relative to total commuting zone wages and the impact of truckers' job loss relative to the working-aged population of each commuting zone. They suggest that changes due to trucking automation will be particularly harsh on the economies of Middle America, including states such as Montana, North Dakota, South Dakota, Iowa, and Nebraska.

Job Loss per Working-Age Population by 2078

Numbers in the legend are in decimal form. For instance, a 2% job loss as a percentage of working-aged individuals, which is the point where commuting zones are completely colored red, is represented by a value of 0.02. Working-age population is defined as the number of individuals in the commuting zone aged 15 to 64.



We would like to stress that our primary innovation here is the framework we created to integrate data on the trucking industry's heterogeneity with experts' judgments on how automation is likely to proceed with each job category. While these judgments served as the foundation for our numbers, our model's parameters can be readily adjusted for more conservative or liberal estimates on job and wage loss.

Our work thus represents an attempt to connect the complexity of the trucking industry to expertise on automation. We strived to put wage insurance policy on firmer ground, exploiting the understanding of the diversity among trucking jobs and how it affects automation. By linking these data sources together and to models translating experts' judgments into numbers, we have developed a workflow to guide further analysis on the subject. We hope that this workflow and the precision it offers serve as a tool enabling policymakers to better respond to the uncertainties posed by the fourth industrial revolution.

Review of Literature

Before diving into the details of automated trucking or estimating its impact, the team undertook a comprehensive survey of existing literature on the subject. The findings of the team are presented in this section.

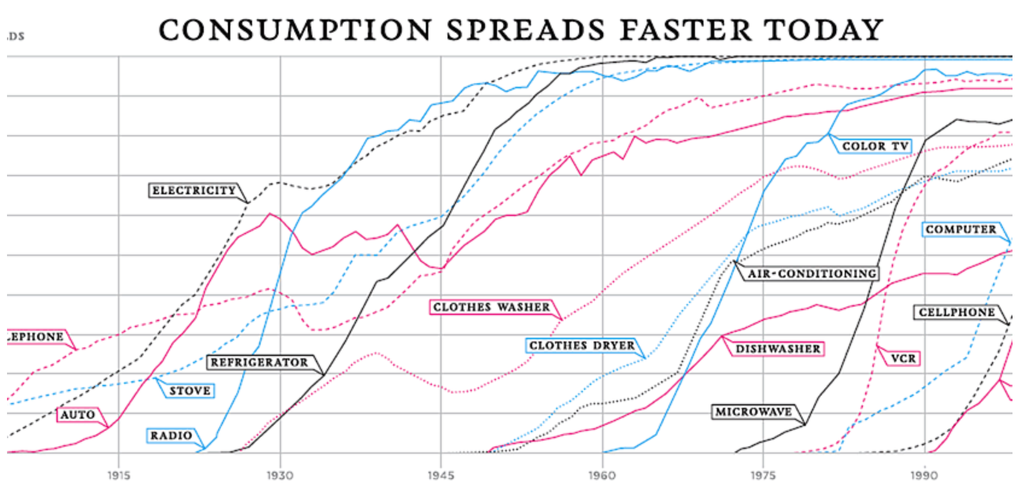
Broadly, the team researched the science behind various technology adoption models to identify the model that would be used to calculate trucker job losses. The Probit Model (S-Curve) was deemed most appropriate. Next, the team researched the trucking industry to determine the basis on which the market is structured. The major segmentation across the industry is by function of the truck and the distance covered by the truck. We also examine other associated attributes of the industry such as unionization and contracting. Finally, we summarize the technological advances in automated trucking, the major players and end with existing research efforts by third-parties to estimate the current and future impact of automated trucking on the US job market.

Technology Adoption Models

Diffusion of Innovations - Everett Rogers' S-Curve

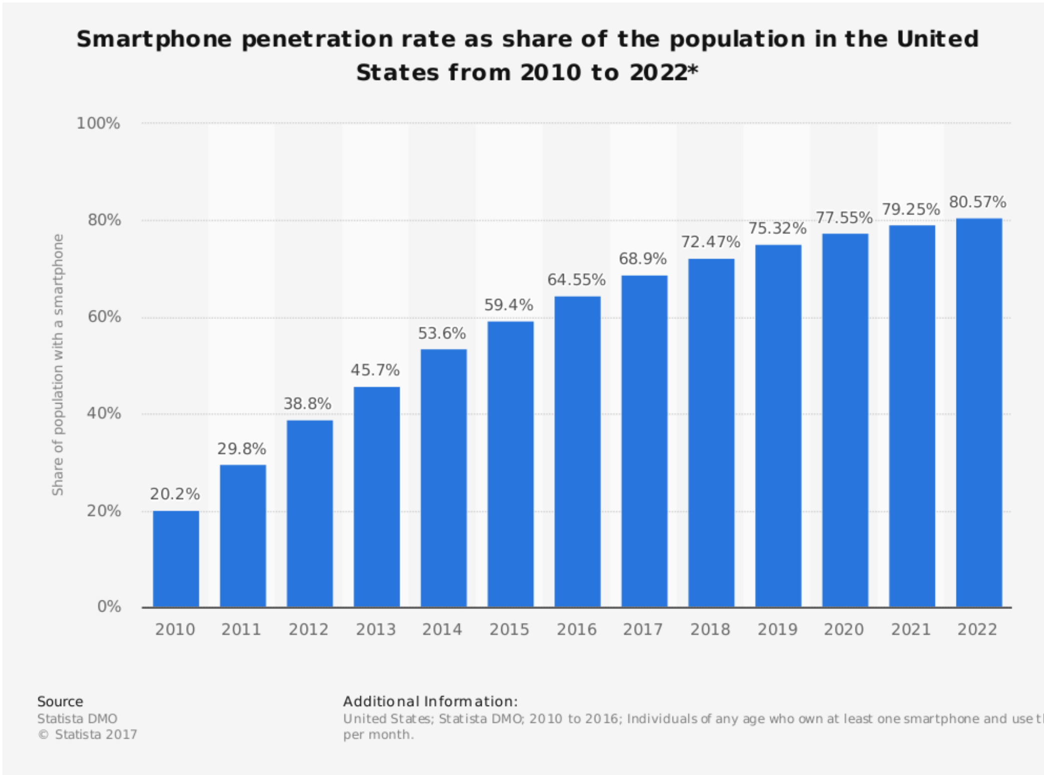
To come to a conclusion on the adoption of Autonomous Trucking technology in the United States, it is useful to first consider technology adoption as a whole. The very first exploration of technology adoption was done in 1962 by Everett Rogers in his book *"Diffusion of Innovations"*. The book was a broad description of how ideas and technologies spread across different cultures. The model has been used as a starting point to drill down on technology adoption and has since led to the creation of many similar models that explain the complications involved in the spread of a new technology.

According to Rogers, adoption follows a standard bell-curve and the various categories of adopters are Innovators, Early Adopters, Early Majority, Late Majority and Laggards.[2] A cumulative view of the same stages turns the bell-curve into an "S-Curve" of adoption.



[2] Technology Adoption Lifecycle - https://en.wikipedia.org/wiki/Technology_adoption_life_cycle

In addition to the technologies shown above, smartphones, smart-speakers and IoT, cars (regular and electric) as well as cryptocurrencies such as Bitcoin follow a S-Curve of technology adoption.[3]



Economic Models of Technology Diffusion - P. A. Geroski

P.A. Geroski's paper *"Economic Models of Technology Diffusion"* surveys literature on the S-curve of adoption and introduces us to a few economic models that could explain the diffusion process. We seek to identify the model which is most relevant to adoption of automated trucking. We discuss two models which come close to explaining it below:[4]

Epidemic Model

In this model, time to adoption is a factor of a lack of information, knowledge and purpose of technology. The relevance to Autonomous Trucking is minimal because adoption of this technology is not limited to knowledge of its existence alone. The fact that autonomous trucking depends on factors other than just the knowledge of the existence of the technology makes this model less applicable to this context. Autonomous Trucking adoption is not a process of spreading news, it has to more to do with persuasion of the interested parties and political climate of the nation. In addition, it depends on the firm that is providing the technology or the market characteristics of the various players. Given the above reasons, the Epidemic Model breaks down.

---

[3] Speculative Bitcoin Adoption/Price Theory - https://medium.com/@mcasey0827/speculative-bitcoin-adoption-price-theory-2eed48ecf7da

[4] Geroski, P. (2000). Models of technology diffusion. *Research Policy*, 29(4-5), pp.603-625.

Probit Model

This model takes other factors into account that are extraneous to the technology in question and the awareness of its existence. Here, time to adoption is a factor of goals, needs, knowledge and abilities of a firm. One of the most important determinants according to this model is firm-type. Immediately we see a relevance to the autonomous trucking landscape. We know for example that smaller trucking companies might not have the resources to adopt newer technology. Larger companies, with their resources, will be quicker to imitate if one or two major players start using automated technology. Each firm could have its own adoption timeline which would be a factor of all the ways in which it is different from other firms in the market.

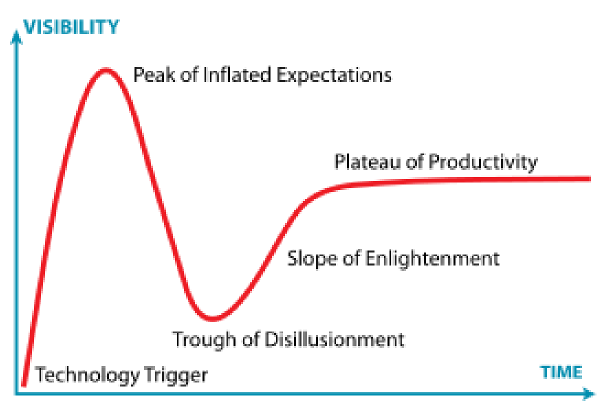
It is also conceivable that supplier pricing and service policies will affect adoption rates. The technology being developed and the supplier that is providing the service will have an impact on the adoption by member firms of the industry. For example, companies like Embark, Uber, Google and Starsky Robotics will develop just the self-driving technology and so it might be cheaper for firms to retrofit their vehicles. On the other hand, vehicle manufacturers like Nissan, Volvo and Tesla will manufacture vehicles with the technology built-in to their trucks which increases purchasing cost for the trucking company. The firm or firms that win the ensuing battle for autonomous trucking and gains market share will affect the adoption curve of the technology on a larger scale.

All of these factors and interdependent scenarios impact adoption and we think the Probit Model is the most relevant in our context. As a next step, we will use this model to develop a most likely scenario of autonomous trucking adoption in the United States

The Gartner Hype Cycle

As an alternative to the S-Curve, Geoffrey Moore suggested that a different flow of technology adoption might be followed by technologies that are more discontinuous or disruptive in nature. This model is called the Gartner Hype Cycle and is shown below. [5]

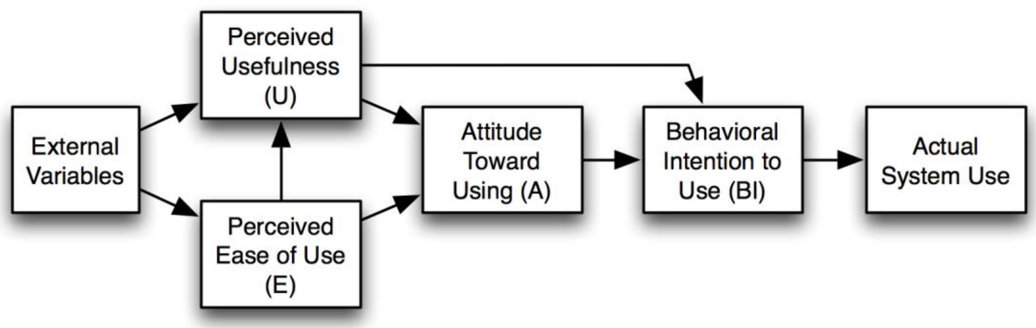
Moore approaches the problem from a marketing standpoint and suggests that the gap between Earl Adopters and Early Majority is the "chasm" which needs to be crossed by a firm to successfully gain traction in the market. An illustration of the Gartner Hype Cycle is shown below.



[5] Moore, G. (2014). *Crossing the chasm*. [New York, NY]: Collins Business Essentials.

Technologies that are new in the market usually accompany a high expectation of usefulness and achieve a "peak of inflated expectations", however reality sets in and its true utility is realized when the demand plateaus. This model does not apply to the automated trucking industry because the societal gains and efficiencies of implementing autonomous trucking are not as varied as suggested here. Most firms are aware of the benefits of automated trucking technology and a sharp decline in this utility is not expected.

Technology Acceptance Model[6]



Another model for technology adoption is shown in the diagram above. Here actual use of the technology is the last step of the process prior to which the attitude of the user as well as the perceived usefulness of the product needs to be assessed.

Again, compared to the Probit Model adapting this model to the Autonomous Trucking industry is not ideal as the TAM is more relevant to cases where a new technology is not sure of improving life for the consumer. In our case, Autonomous Vehicle technology will reduce costs, improve efficiencies and reduce traffic accidents and so acceptance should not be an issue.

To conclude this section, it is our view that Geroski's Probit Model is most suitable to be applied to the case of Autonomous Trucking adoption in the United States.

Trucking Industry Segmentation

Segmentation by Distance and Function

Categorical Definitions

Workers in the trucking industry are most frequently divided into two categories by the length of their trips: **long-haul** and **short-haul**. There are no clear divisions between the two, but generally, long-haul truckers transfer goods between metropolitan areas, while short-haul truckers operate more locally. The 1.7 million large freight truck driving jobs, as counted by the Bureau of Labor

[6] En.wikipedia.org. (2018). *Technology acceptance model*. [online] Available at: https://en.wikipedia.org/wiki/Technology_acceptance_model [Accessed 27 Feb. 2018].

Statistics (BLS)[7] (geographic, breakdowns here[8]), can be split along these terms according to the table below, with an additional 170,000 - 425,000 self-employed and contracting drivers.[9]

Estimates of 2016 Heavy Truck Driver Employment by Sector, Truck Type, and Range of Operations[10]

	50 miles or less	51-100 miles	101-200 miles	201-500 miles	501+ miles	Total
Straight truck private	32.20%	6.60%	1.40%	0.70%	0.40%	41.30%
Tractor trailer private	9.30%	5.40%	2.80%	3.50%	1.70%	22.70%
Straight truck for-hire	2.00%	0.60%	0.20%	0.10%	0.10%	3.00%
Tractor Trailer for-hire	6.10%	4.30%	4.10%	6.10%	12.50%	33.10%
	49.60%	16.90%	8.50%	10.40%	14.70%	

This dichotomy, however, speaks less to the physical constraints of trucking than the division between **less-than-truckload (LTL)** and **full-truckload (FTL)** shipping. As the names of these categories suggest, they represent the extent to which a single shipment fills a truck's trailers. Because of how shipped loads are broken down as they serve more narrow geographies, LTL and FTL shipping oftentimes, but do not always, respectively correspond with short and long-haul shipping.

Furthermore, specialized trucking services—those that involve products that require special care, including refrigerated foods, furniture, and hazardous waste—provide transportation for goods across all distances and LTL/FTL shipping parameters. The immense range of products that these services transport makes analysis of these industries particularly complex, but generally, these industries require specialized pretraining in one form or another.

These breakdowns suggest that to correlate trucking employment with potential job losses, datasets with the following information would be particularly useful for our analysis:

- Geospatial metadata on job counts or flows to break down probable job losses by region

---

[7] "Summary: Heavy and Tractor Trailer Truck Drivers." U.S. Bureau of Labor Statistics. https://www.bls.gov/ooh/transportation-and-material-moving/heavy-and-tractor-trailer-truck-drivers.htm. Note, however, that the "Heavy and Tractor-trailer Truck Drivers" category from BLS is "most(ly)", and not entirely, made of long-haul drivers. Furthermore, one should note that industry estimates frequently overstate employment of drivers by counting administrative/support workers in the industry as drivers.

[8] "53-3032 Heavy and Tractor-Trailer Truck Drivers." U.S. Bureau of Labor Statistics. https://www.bls.gov/oes/current/oes533032.htm#st

[9] Rafter, Michelle. "Feds Say Truckers Over-Counted, Autonomous Tech Threat Overstated." Trucks.com. February 22, 2018. https://www.trucks.com/2018/02/01/truckers-autonomous-tech-threat-overstated/

[10] Monaco paper

- Job counts per short-haul vs. long-haul/LTL vs. FTL driving, or, alternatively, data on shipping flows including:
  - Distances traveled for long-haul vs. short-haul breakdowns
  - Shipment loads as a percentage of truckload
- Job counts or flow quantities for specialized industries

Though this information is rarely provided in a sufficiently disaggregated format to answer all these questions, a combination of databases can provide a more complete picture than a single table. Flow quantities across geographies, industries, and truckload sizes provided by the Community Flow Survey[11] may enable us to get a proportional intuition on these questions even though they are not directly linked with job quantities. Analysis on the CFS should then be appropriately corroborated with estimates from the BLS's geographic breakdowns[12] and the Monaco paper's distance-based classifications.

Vulnerabilities to and Protections Against Automation

Forecasts on automation suggest that non-specialized long-haul and FTL trucking services will be the first to be automated. The punishing schedules and monotonous labor, which are two of the prime factors for employee turnover in the industry, are the very same variables making the industry vulnerable to automation. Eliminating long-haul drivers could save the industry the 26% of revenue it spends on wages.[13]

Short-haul and LTL shipping are likely to face automation later on, primarily because of the capital and labor intensity of their repeated loading/off-loading processes. Unlike in long-haul trucking, each trip in this category may have multiple destinations and "beyond points" requiring special human attention, and repeated trips frequently enable short-haul truckers to build rapport with their customers.

Political Resistance

Industries' paths to automation depend not only on technical capabilities and economic feasibility but also power structures endemic to the profession. It is therefore, worth considering the political influence that "change-makers" hold in trucking.

Market Structure

Long-haul trucking is a highly competitive market, with nearly 90% of "operators" being "non-employers". The top three companies hold produced 5.3% of the industry's revenue in 2017 and the top 50 companies captured less than 30 percent of the market. Individuals can readily start their own trucking business with a loan. Still, though, owner-operator firms generate less than 20% of the market's revenue, and large firms are capable of winning large contracts and setting prices to

---

[11] US Census Bureau, Hang Yu (ESMPD). "US Census Bureau Commodity Flow Survey Main Page." Census.gov. June 17, 2010. https://www.census.gov/econ/cfs/.

[12] "53-3032 Heavy and Tractor-Trailer Truck Drivers." U.S. Bureau of Labor Statistics. https://www.bls.gov/oes/current/oes533032.htm#st

[13] Edward Rivera. "Long-Distance Freight Trucking in the US." IBIS World. November 2017. http://clients1.ibisworld.com/reports/us/industry/competitivelandscape.aspx?entid=1150#MSC
Note: This source also presents other information on other costs, such as maintenance, purchases, barriers to entry, and etc.

some extent.[14] With a limited market share, the adoption of technology among these large firms is likely to lead to slower changes among the remaining operators unless their success portends major changes in the landscape of suppliers.

Unionization and Contracting

Only a small percentage of truck drivers are unionized (less than 5 percent)[15], indicating that as a whole, these organizations are unlikely to be sources of significant resistance to automation. However, they appear to be concentrated in small pockets where the industry bleeds into non-trucking services—for example, 20% of UPS's 1.4 million-strong parcel division are members of the International Brotherhood of Teamsters, the largest player among unionized truckers.

Job Loss Models and Considerations

The Rise and Diffusion of Automated Trucking Technology

Start-up Technology & Affordability

Rapid developments in automated technology can be seen through a survey of leading technology start-ups including Starsky Robotics, Embark and Otto. Starsky Robotics currently operates at Level 3 automation but plans to remove in-vehicle "safety drivers" by end of the year[16]. Its model shows the highest potential for driver displacement, expecting that on highways, trucks will drive independently and for local "last-mile" routes, truck drivers will only be needed for remote monitoring. In contrast, Embark (which just completed its first long-haul journey across the US[17]) and Uber-acquired Otto propose an increase in local truck driving jobs via a transfer hub model whereby drivers will still be needed behind the wheel on local routes which aren't as easily automated.

In any case, these startups offer retrofit options that would make adoption of autonomous trucking technology much more economically viable and cost-effective than fleet replacement. (McKinsey estimates[18] 2M tractor-trailers in the US, typically replaced every 20 years at a cost of $160,000

---

[14] Edward Rivera. "Long-Distance Freight Trucking in the US." IBIS World. November 2017. http://clients1.ibisworld.com/reports/us/industry/competitivelandscape.aspx?entid=1150#MSC
Note: this assessment by IBIS World contains a great deal of information on this industry beyond market competitiveness.

[15] Schulz, John D. "Union-free Carriers Trying Hard to Stay That Way." Recently Filed RSS. December 18, 2013. http://www.logisticsmgmt.com/article/union_free_carriers_trying_hard_to_stay_that_way
The link above cites a figure of 2.5% out of 3.5 million truck drivers. This total count is greater than the estimate of large freight truck drivers in the US, possibly because it includes drivers of smaller trucks. However, the arithmetic suggests that no more than 5% of freight truck drivers are unionized.

[16] Chafkin, Max, and Josh Eidelson. "These Truckers Work Alongside the Coders Trying to Eliminate Their Jobs." Bloomberg.com. June 22, 2017. Accessed February 26, 2018. https://www.bloomberg.com/news/features/2017-06-22/these-truckers-work-alongside-the-coders-trying-to-eliminate-their-jobs.

[17] "Embark Self-Driving Truck Completes Coast-to-Coast Test Run." Transport Topics. February 08, 2018. Accessed February 26, 2018. http://www.ttnews.com/articles/embark-self-driving-truck-completes-coast-coast-test-run.

[18] *A FUTURE THAT WORKS: AUTOMATION, EMPLOYMENT, AND PRODUCTIVITY.* Report. Accessed February 26, 2018.

each – excluding the cost of autonomous driving technology.) Otto for instance, offers its AI system retrofit for around $30,000[19] per truck and the American Transportation Institute estimates that additional costs per truck for hardware and software at each level of automation will be as follows: $13,100/truck for L3; $19,000/truck for L4; and $23,400/truck for L5[20]. With annual earnings of "heavy and tractor trailer drivers" (the segment likeliest to be replaced by highway route automation) estimated at $44,000 on average in 2016[21], the current prices of retrofits bode well for adoption from a cost perspective.

Legislation

While current literature does not speculate on legislation accelerating or slowing the pace of adoption, it does acknowledge that upcoming national legislation can help inform the timeline and constraints of adoption[22]. Currently, five states have enacted legislation permitting autonomous driving pilots with many more considering legislation[23]. The adoption of level 4 autonomation, across long-haul interstate routes for instance, can be impeded if states decided to reject it in their jurisdiction[24].

Forecasted Job Loss by 2050: A Review of Current Models and Hypotheses

ILR Review

In the manuscript, "Truck Driving Jobs: Are They Headed for Rapid Elimination?"[25], submitted to the ILR Review, authors forecast that roughly 300,000 – 400,000 trucking jobs will be displaced due to automated trucking. While they don't attempt to estimated over what time period this will occur, they estimate this will occur as levels 4-5 automation are adopted (given drivers will still be needed at level 3). To arrive at this conclusion, the authors segment drivers based on those likeliest to be impacted by automation and consider only those who can be accurately tracked by national datasets. *As a caveat*, the authors focus on Heavy and Tractor-Trailer Truck Drivers in the Bureau of Labor Statistics (BLS) dataset but note that the Occupational Employment Statistics exclude the self-employed such as Owner-Operators (roughly 10-25% of all heavy truck drivers[26]) and independent contractors (estimated at 800,000 and highly vulnerable to displacement according to experts such as Steve Viscelli). The authors note that at levels 4-5 automation it will be cost-effective to remove drivers for routine routes such as long-haul, interstate routes. Thus, they

---

https://www.mckinsey.com/~/media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx.

[19] Freedman, David H. "If automation is already messing with our economy and our politics, just wait until self-driving trucks arrive." MIT Technology Review. April 06, 2017. Accessed February 26, 2018. https://www.technologyreview.com/s/603493/10-breakthrough-technologies-2017-self-driving-trucks/.

[20] *Identifying Autonomous Vehicle Technology Impacts on the Trucking Industry.* Report. Nov. & dec. 2016. http://atri-online.org/wp-content/uploads/2016/11/ATRI-Autonomous-Vehicle-Impacts-11-2016.pdf.

[21] Truck Driving Jobs: Are They Headed for Rapid Elimination? MS ILR-17-0267, ILR Review. Page 29.

[22] Truck Driving Jobs: Are They Headed for Rapid Elimination? MS ILR-17-0267, ILR Review. Page 25.

[23] "Automated Driving: Legislative and Regulatory Action." Automated Driving: Legislative and Regulatory Action - CyberWiki. Accessed February 26, 2018. http://cyberlaw.stanford.edu/wiki/index.php/Automated_Driving:_Legislative_and_Regulatory_Action.

[24] Truck Driving Jobs: Are They Headed for Rapid Elimination? MS ILR-17-0267, ILR Review. Page 25.

[25] Truck Driving Jobs: Are They Headed for Rapid Elimination? MS ILR-17-0267, ILR Review.

[26] Belman, D., F. Lafontaine, and KA Monaco. "Truck drivers in the age of information: transformation without gain." *Trucking in the Age of Information*, 2005, 183-212.

consider only truck drivers who operate routes of 201 or more miles (419,000 truck drivers) and those most likely to be immediately affected by L4 automation (the 310,000 of the above who operate in the for-hire transportation and warehousing sectors). This 300,000 - 400,000 displacement range is exhibited in Appendix A.

McKinsey

While the authors of the above report do not attempt to forecast the timing of technology deployment and adoption, McKinsey does so via its Center for the Future of Mobility[27]. It predicts Level 4 autonomy will be available between 2020-2022, and that Level 5 will be available by 2030 at the soonest, at which time the industry will start to see widespread adoption of autonomous commercial driving technology. In the meantime, platooning (Level 3 technology by which one driver leads a platoon of trucks) may arrive as soon as 2018, with advanced-driver assistance systems (ADAS) expected to double by 2021. While at these stages drivers are still expected to be utilized behind the wheel rather than displaced, these forecasts show the expected evolution and pacing of autonomous trucking technology.

In its 2017 Automation, Employment & Productivity Report[28], McKinsey models in greater detail its earliest and latest scenarios for technology adoption, based on when Level 4 technology will be available and when it will become on par with the cost of human labor (based on the BLS segment containing Heavy and Tractor-Trailer Truck drivers). It forecast of earliest and latest adoption years is available in Appendix B, and it provides S-curves based on the cost of automation falling under US wage levels 3-10 years after Level 4 autonomy arrives. *As a caveat,* it uses a fleet replacement assumption but acknowledges the opportunity to retrofit at cheaper rates.

Uber

While Uber has a natural interest in promoting an optimistic model for forecasted job impact, it has projected a model with some merit acknowledged by experts such as Steve Viscelli. It predicts .5-1.5M self-driving trucks will be deployed by 2028 and that with a high utilization rate, costs will decrease, demand will increase, and 1M jobs will shift from long-haul to local-haul, with an additional 400,000 new jobs created to meet local hub demand[29]. This contrasts sharply from all other models predicting job loss (the most drastic being those of the White House provided in Appendix C), however both the authors of the ILR Review (Monaco) Report and Steve Viscelli acknowledge the need to explore the "last mile" business model and see potential for increased employment at transfer hubs. [30] Viscelli predicts adoption on interstates first, with exit-to-exit

---

[27] "Autonomous Driving." McKinsey & Company. Accessed February 26, 2018. https://www.mckinsey.com/features/mckinsey-center-for-future-mobility/overview/autonomous-driving.

[28] *A FUTURE THAT WORKS: AUTOMATION, EMPLOYMENT, AND PRODUCTIVITY.* Report. Accessed February 26, 2018. https://www.mckinsey.com/~/media/McKinsey/Global%20Themes/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx.

[29] Uber. "Uber/trucking-labor-analysis." GitHub. January 31, 2018. Accessed February 26, 2018. https://github.com/uber/trucking-labor-analysis.

[30] Madrigal, Alexis C. "Could Self-Driving Trucks Be Good for Truckers?" The Atlantic. February 01, 2018. Accessed February 26, 2018. https://www.theatlantic.com/technology/archive/2018/02/uber-says-its-self-driving-trucks-will-be-good-for-truckers/551879/.

trucking automated and safe in three years, while maintaining an overall estimated job loss forecast of a few hundred thousand jobs.[31]

[31]Chafkin, Max, and Josh Eidelson. "These Truckers Work Alongside the Coders Trying to Eliminate Their Jobs." Bloomberg.com. June 22, 2017. Accessed February 26, 2018. https://www.bloomberg.com/news/features/2017-06-22/these-truckers-work-alongside-the-coders-trying-to-eliminate-their-jobs.

Estimating Job and Wage Loss

To estimate job loss across the nation over time and within each commuting zone, we needed to calculate the quantity of jobs in each of our 32 specified categories and within each of the thousands of locations (initially, counties, then aggregated by commuting zone) across the U.S. Job counts by the 32 categories across the nation do not exist at such a granular level, let alone at a level broken down by both category and location. Therefore, we developed an innovative approach to combine the strengths of two datasets: the Quarterly Census on Employment and Wages (QCEW) and the Occupational Employment Statistics database (OES).

These statistics were fed into an adjustable model (based on the sigmoid curve) for each job category in every locality. By multiplying the quantities of job lost in each location and category by estimated wage loss statistics, we can estimate total wage loss across locations, categories, and the nation. Our model enables us to generate a more granular understanding of job loss by recognizing that the trucking industry is a heterogenous industry with numerous categories, each of which, is likely to experience different effects of automation. We adapted the curve to best reflect the judgments of our adviser, Dr. Steve Viscelli[32]. However, the model's parameters can be readily changed to reflect the judgments of other experts in the field.

Using Sigmoid Curve Theory to Model Job Loss

As discussed in the review of literature ("Trucking Industry Segmentation" section), different segments of the trucking industry are likely to differ in their vulnerability to job loss. For example, long-haul truckers are particularly vulnerable to automation because their work occurs over long, repetitive hours, and because they have fewer loading/unloading cycles than short-haul truckers. We therefore needed to identify key characteristics that might affect job losses over time. Each of these characteristics would need to somehow be inputted into a model that accounts for how these different characteristics affected job losses over time, all the while reflecting our understanding of how technological adoption occurs.

Defining the Job Categories

Key Characteristics

With our review of trucking industry segmentation, we identified several characteristics that might strongly affect job loss:

- Distance type: whether the job is short-haul vs. long-haul
- Load type: whether the job is full-truckload (FTL) or less than truckload (LTL)

[32] Dr. Steve Viscelli is a senior fellow at the University of Pennsylvania's Kleinman Center for Energy Policy. His expertise lies in labor market economics. He is the author of *The Big Rig: Trucking and the Decline of the American Dream*, which chronicles the changes the US trucking industry has undergone over the past several decades.

- Specialization: whether the job is associated with specialized characteristics
- Unionization: whether the job is associated with a union or not
- Owner-operator status: whether the trucker is self-employed

We, in consultation with our adviser (Viscelli, personal communications), reviewed these characteristics for how these characteristics might affect trucking jobs' vulnerability to automation. The table below summarizes our judgements.

Vulnerabilities Associated with Characteristics' Values

Characteristic Type	**Greater Vulnerability**	**Less Vulnerability**
Distance *(strong effect)*	Long-Haul	Short-haul
Load Type *(strong effect)*	Full Truckload (FTL)	Less-Than-Truckload (LTL)
Specialization *(strong effect, but requires clarification)*	Non-specialized	Specialized
Unionization *(exceptionally strong effect)*	Non-Unionized	Unionized
Owner-Operator Status *(not well-divided by more/less vulnerable)*	Employees of Firms; Owner-Operators	

Our judgments on the effect of distance and load types reinforced our literature review. These analyses were, for instance, partly rooted in the frequency of load/unload cycles, as long-haul and FTL shipments are associated with less frequent load/unload cycles, and are thus more repetitive and automatable work. However, our consultations on the impact of unionization were particularly noteworthy. We assessed that unions are likely to strongly oppose both the adoption of automation technology and the replacement of workers. Our conversations with Viscelli suggested that very few of these workers would be replaced over the timeframe, which substantially depresses the probable job loss calculations relative to what they would have been without including unionization.

We also initially judged that specialized jobs would be less vulnerable because they might require unique skill sets among workers that are difficult to automate. But with this interpretation, it is very important to delineate what exactly constitutes a "specialized" job. Refrigerated trucking, for instance, is considered a "specialized trucking job" according to the North American Industry Classification System (NAICS), though our conversations with Viscelli repeatedly pointed to refrigerated trucking as an industry that might be particularly vulnerable to automation.

Jobs that are tallied in datasets as “specialized” or “general” must thus have their definitions verified before we can assign an interpretation of their vulnerability. Upon review of our key datasets (the Bureau of Labor Statistics’ Occupational Employment Statistics database and Quarterly Census of Wages), we judged that specialized jobs might be less vulnerable to automation, though their differences with generalized jobs might be less stark than initially conceived.

Among owner-operators, it is not easy to describe owner-operators as merely more or less vulnerable to automation. Viscelli suggested that while they would not have the capital to automate their own jobs, they would be rapidly pushed out of the market once large firms begin engaging in automation. Their pace of automation would thus be slower than other workers initially, but faster after large firms have automated a substantial proportion of their jobs.

Applying Characteristics to Develop Job Categories

Using these five characteristics, we can, in effect, generate job categories by whether or not they meet specific criteria for each characteristic. Consider, for example, a simple case where we have two characteristics: whether a job is long-haul or short-haul; and whether a job is full-truckload or less-than-truckload. There are thus four possible pairings of these characteristics (i.e., job categories):

- Jobs that are long-haul *and* full-truckload
- Jobs that are long-haul *and* less-than-truckload
- Jobs that are short-haul *and* full-truckload
- Jobs that are short-haul *and* less-than-truckload

As shown in the table below, we can use this same logic to split jobs by all five of the characteristics previously discussed.

Matrix of Job Categories

		General.or.Specialized / Contractor.or.Not / Unionized.or.Not							
		General				Specialized			
		Not Owner-Operator		Owner-Operator		Not Owner-Operator		Owner-Operator	
Long.or.Short	LTL.or.FTL	NotUnionized	Unionized	NotUnionized	Unionized	NotUnionized	Unionized	NotUnionized	Unionized
Long	FTL	■	■	■	■	■	■	■	■
	LTL	■	■	■	■	■	■	■	■
Short	FTL	■	■	■	■	■	■	■	■
	LTL	■	■	■	■	■	■	■	■

Under this framework, the top left cell is a job category that has these characteristics:

- It is a long-haul job
- Its loads are full-truckload
- Its truckers are not unionized
- Its truckers are not owner-operators
- And it focuses on general freight (i.e., non-specialized) loads

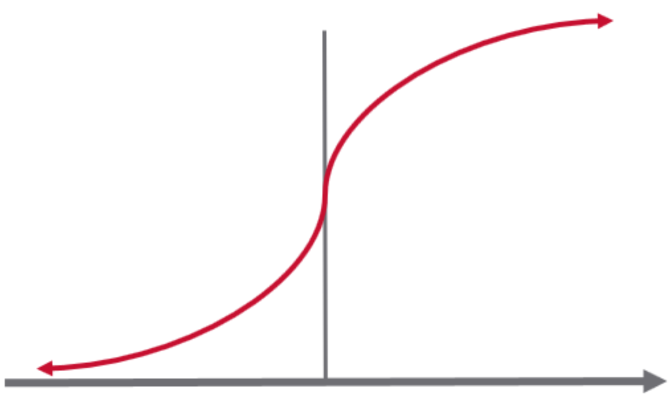
Likewise, the bottom right cell is a job category with these characteristics:

- It is a short-haul job
- Its loads are less-than-truckload
- Its truckers are unionized
- Its truckers are owner-operators
- And it focuses on specialized freight loads

We have, therefore, in effect generated 32 different job categories, each of which has different characteristics. At this point, the challenge is to develop a model that can quickly and transparently reshape job-loss estimates according to experts' specifications.

Applying the Theory of the Sigmoid Curve

We judged the sigmoid curve ("Technology Adoption Models") to be the curve that best-models technology adoption. As shown in the figure below, its slope is near-zero at first, before rapidly increasing, and subsequently approaching near-zero once again.



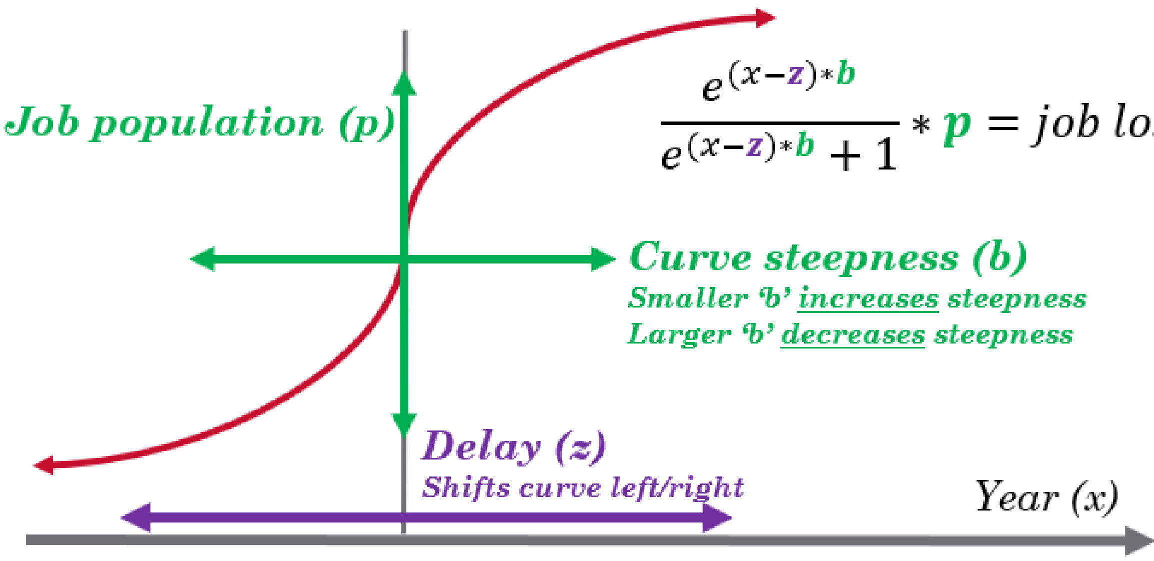
These three phases correspond with a rough understanding on how technological adoption works:

- Change begins slowly due to inertia, possibly due to capital requirements, risk adversity, and scarcity of information
- After an extended period of slow growth, growth rapidly accelerates after approaching some sort of critical mass. This event occurs when large portions of the population recognize the long-term utility of the innovation.
- Growth subsequently slows as the portion of the society that has not absorbed the innovation but can do so in the long-run becomes smaller and smaller. The spread of change becomes slower because there are fewer and fewer people it can spread to.

This curve is best-modeled with the this equation, where y and x represent the vertical and horizontal axes respectively: $y = e^{x}/(1 + e^{x})$

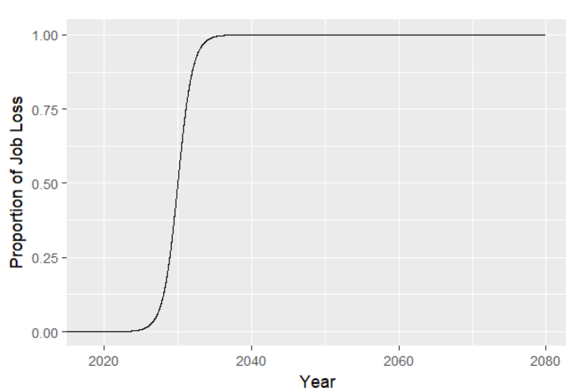
Adjusting S-Curve Parameters

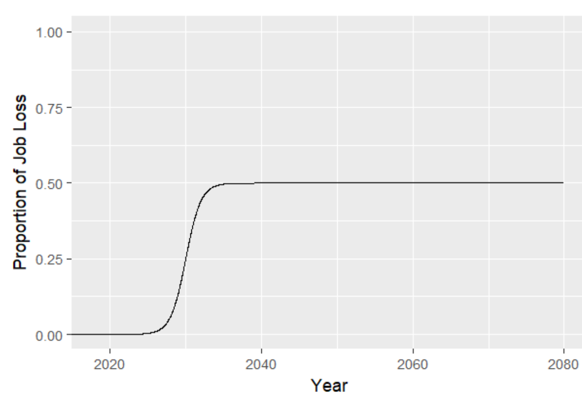
We can therefore imagine a curve where the vertical axis is our job loss at any one point in time, and x, is the year where we're measuring job loss. We can generate specific curves for each specific job category and location. Each of these curves is essentially generates a model of job losses over time for a specific population of truckers (e.g., job loss for one location in one job category). Furthermore, we can change the shape and placement of the sigmoid curve as illustrated by the following figure.



Green parameters ("job population (p)" and "curve steepness (b)") are those that change the shape of the curve, compressing it or stretching it vertically and horizontally. The purple parameter ("delay (z)") shifts the curve to the left and the right. The following text describes these parameters in a little more depth:

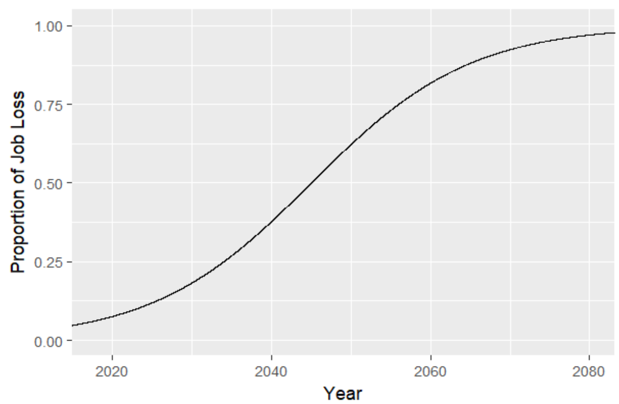
Job population (p): This represents the height of the curve. Increasing this increases the maximum height of the curve (left graph), while decreasing this decreases its maximum height (right graph).

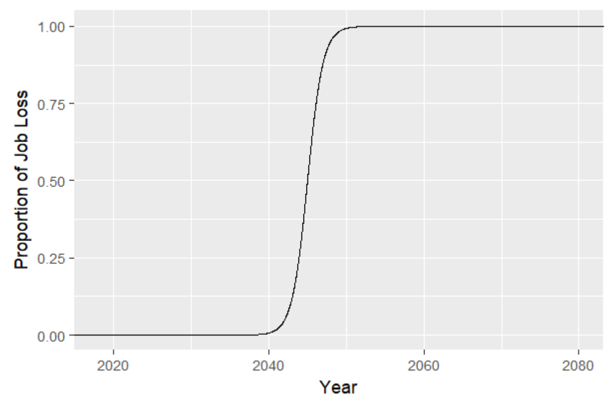




This is useful because different job categories will have different population sizes. There might be more long-haul jobs than short-haul jobs, and so we take these population sizes into account when rescaling to actual job counts by multiplying by *p*.

Curve steepness (b): Increasing *b* decreases the steepness of the curve, while decreasing *b* increases the steepness. We might be inclined to think about this as a measure of vulnerability to the job--after all, job categories that are more vulnerable might experience more sudden job loss. However, this is not so simple. Consider, for example, the two graphs below, where the left has a high *b* value, and the right has a small *b* value.

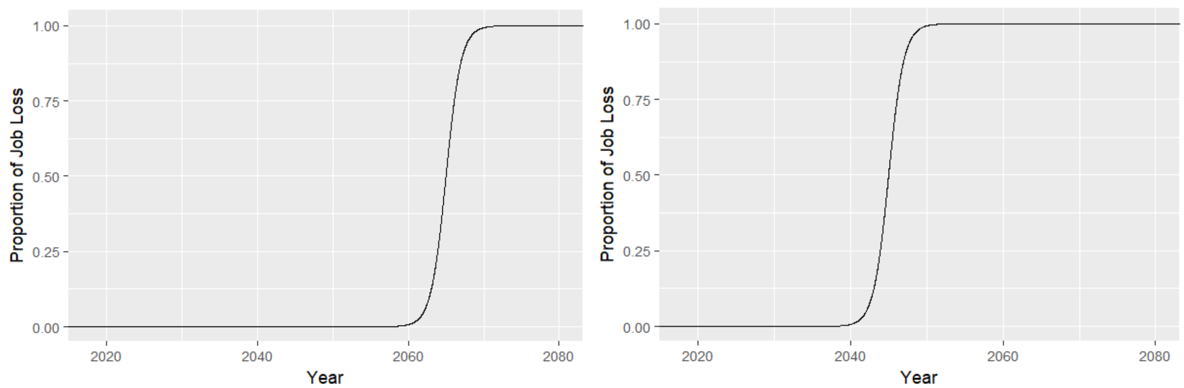




We see that the left graph is less steep than the right. But, with all other parameters held constant, it actually starts off at a higher job loss value. So, the graph on the left, in effect, has job losses occuring over a longer period of time that also *start earlier*. This behavior is a little bit complicated, but can be useful in some instances (e.g. owner-operators, whose job losses might begin later, but occur at a steeper rate once they begin).

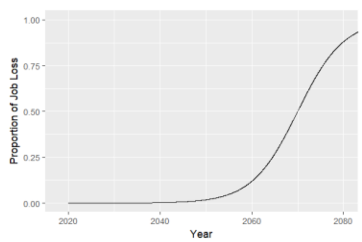
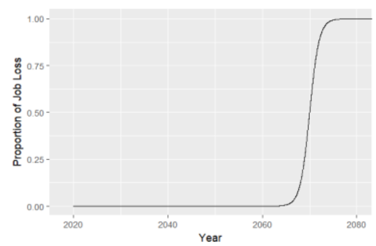
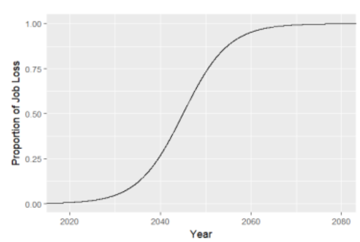
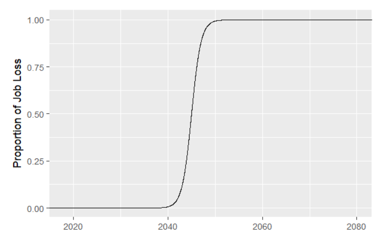
To designate a category as more or less vulnerable, one must manipulate both *b* (steepness) and *z* (delay).

Delay (z): We can move the graph from left to right using this parameter. Its value essentially places the curve's center at z (e.g., if z = 2030, then the center of the curve is at 2030). This is useful because we can adjust this parameter to indicate when job losses occur earlier or later, even if the shape of the curve does not change. Among the figures below, the graph to the left has a high *z* value; the graph to the right has a lower *z* value, which causes this graph to shift leftward.



Given these parameters, we can manipulate variables in combination with one another to generate different curves. Manipulations of b (curve steepness) and z (delay) produce particularly interesting combinations for modeling job loss.

Adjusting S-Curves via *b* and *z*

<table>
<tr><td colspan="2"></td><td colspan="2"><u>Manipulating b (curve steepness)</u></td></tr>
<tr><td colspan="2"></td><td>High b</td><td>Low b</td></tr>
<tr><td rowspan="2"><u>Manipulating z (delay)</u></td><td>High z</td><td>
Job loss is delayed and occurs slowly. Least vulnerable.</td><td>
Job loss is delayed but occurs quickly once it begins.</td></tr>
<tr><td>Low z</td><td>
Job loss is not delayed but occurs slowly.</td><td>
Job loss happens early and occurs quickly. Most vulnerable.</td></tr>
</table>

Applying the S-Curve to Each Job Category

Though we have specified our job characteristics and s-curve parameters of interest, connecting the two does not seem to be a trivial exercise. Let's recall the form of our equation:

$$y = \frac{e^{b(x-z)}}{1+e^{b(x-z)}} * p$$

In theory, with 32 different job categories, we could have 32 different possible values each for *b*, *z*, and *p*. For instance, truckers that meet all of the following characteristics might receive one set of unique values for *b*, *z*, and *p*:

- Short-haul
- Non-unionized
- LTL
- Specialized
- Owner-operator

However, this isn't practical. This category is too specific to conceive of and it is not ideal to request experts to provide 32 different sets of parameters, one for each category. We instead settled on a model that builds each parameter by combining prespecified values associated with each characteristic individually.

For instance, to build *b*, rather than directly creating *b* from scratch with 32 different possible values for the steepness of the curve, we decomposed *b* according to its characteristics. This, in effect, follows the following specification:

$$b = b_1 * b_2 * b_3 * b_4 * b_5 = \prod_{c=1}^{5} b_c$$

Here, each b_1, b_2, and so on is linked to a specific kind of characteristic and the value for that characteristic (index *c* is the index of the characteristic).

A similar logic follows for *z*, though *z* is instead built additively:

$$z = z_1 + z_2 + z_3 + z_4 + z_5 = \sum_{c=1}^{5} z_c$$

By approaching the problem from this framework, we do not need to fit 32 different values (there are 32 different job categories) to *b* and *z* each. Rather, we can create these 32 different values from a base set of 10 values: two possible values for each of the five characteristics.

To make this a little more clear, we can consider the following example using the *b* and *z* values we formulated in consultation with Viscelli. With our consultations with Viscelli, we found it useful to treat the two characteristics, distance and load type, as a single characteristic. Rather than identifying jobs by two characteristics separately, each with two possible different values (e.g., haul type has two possible values: LTL or FTL), we treated distance and load type as a single characteristic with four possible different values:

- Short haul, LTL

- Long haul, LTL
- Short haul, FTL
- Long haul, FTL

There are, in effect, still 32 different job categories, however.[33]

Coefficients *b* and *z* Derived in Consultation with Viscelli

		Curve Steepness: b	Delay: z
Characteristic	Distance and Load Type	Short Haul, LTL: 18 Long Haul, LTL: 6.5 Short Haul, FTL: 13 Long Haul, FTL: 9	Short Haul, LTL: 2080 Long Haul, LTL: 2045 Short Haul, FTL: 2060 Long Haul, FTL: 2045
	Specialization	Specialized: 1.3 General: 1	Specialized: 10 General: 0
	Unionization	Unionized: 1 Non-unionized: 1	Unionized: 60 Non-unionized: 0
	Owner-Operator Status	Owner-Operator: 0.6 Non-owner-operator: 1.4	Owner-Operator: 0 Non-owner-operator: 0

Values for *b* and *z* used for the "distance and load type" characteristic are much larger than the possible values used for other characteristics because we decided that the "distance and load type" characteristic was one that would be useful as a foundational frame of reference. All other characteristics merely modify the parameters set forth by this first characteristic. E.g., given a specific distance and load type, specifying that the trucking job is a "specialized" job type shifts the curve to the right by 10 years, according to *z* the delay parameter.

Let's say we wanted to select a job category that meets the following characteristics:

- Short haul and LTL
- General (non-specialized)
- Non-unionized
- Owner-operator

Recall that the s-curve specification occurs in this form:

$$y = \frac{e^{b(x-z)}}{1 + e^{b(x-z)}} * p$$

Where, now, we specify *b* and *z* as the following:

[33] 4 x 2 x 2 x 2 = 32

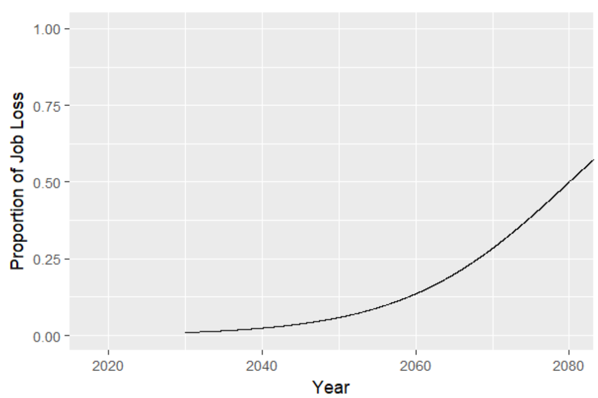
$$b = [Distance\ and\ Load\ Type\ Value] * [Specialization\ Value] * [Unionization\ Value] * [Owner\ Operator\ Value] = 18 * 1 * 1 * 0.6 = 10.8$$

$$z = [Distance\ and\ Load\ Type\ Value] + [Specialization\ Value] + [Unionization\ Value] + [Owner\ Operator\ Value] = 2080 + 0 + 0 + 0 = 2080$$

Our model for job loss for this specific job category would thus be the following, where p is the population of truckers in this category:

$$y = \frac{e^{10.8(x-2080)}}{1 + e^{10.8(x-2080)}} * p$$

Its s-curve would look something like this:



We can apply this modeling workflow to for any of the 32 job categories to develop quantitative models of job loss.

Combining Estimates Across Categories and Locations

As noted earlier, each of these s-curves is meant to represent the job loss over time in a specific category. Now, to generate maps of job loss over time, we would need to understand these s-curves as models that are a little more granular, focusing on job loss over time in a specific category *and in a specific location* (e.g. county or commuting zone). Therefore, we can understand our statistical modeling workflow with some mathematical specification.

We can hold hold the following variable indices to keep track of objects of interest:

i for the job category (one of 32 job categories)

l for the location (one for each county, commuting zone, or some other geographic unit)

A model for job loss in a single category and a single location can be written as the following equation:

$$y_{i,l} = \frac{e^{b_i(x-z_i)}}{1 + e^{b_i(x-z_i)}} * p_{i,l}$$

Note that the *b* (steepness)and *z* (delay) coefficients are the same in all locations if they refer to the same category. Only *p* (size of the population) changes. To put it another way, we assume that the percentage of jobs lost for a specific category over time is the same in all locations if we hold the category constant; it is only the size of the population that changes in each location.

If we wanted identify job loss nationally for one category over time, denoted as Y_i in this case, we would sum up all the s-curves in each location for category *i*:

$$Y_i = \sum_{l=1}^{n} y_{i,l} = \sum_{l=1}^{n} \left(\frac{e^{b_i(x-z_i)}}{1 + e^{b_i(x-z_i)}} * p_{i,l} \right)$$

And if we wanted to identify job loss in a specific location, denoted as Y_l, we would sum all the s-curves in each category for location *l*:

$$Y_l = \sum_{i=1}^{n} y_{i,l} = \sum_{i=1}^{n} \left(\frac{e^{b_i(x-z_i)}}{1 + e^{b_i(x-z_i)}} * p_{i,l} \right)$$

Total national job loss (Y) is simply the sum of all Y_i curves across all categories or the sum of all Y_l curves across all locations:

$$Y = \sum_{i}^{n} Y_i = \sum_{l}^{n} Y_l = \sum_{l}^{n} \left(\sum_{i}^{n} (y_{i,l}) \right)$$

Calculating Populations of Jobs for Each Category: Integrating the Bureau of Labor Statistics' Datasets

Having derived adjustable coefficients *b* (curve steepness) and *z* (delay) for each job category via our conversations with Viscelli, we still needed estimates for the population of truckers in each of the 32 different categories and the thousands of locations (counties, to be aggregated into commuting zones). However, the availability of this data is limited. There are very few datasets that break out the national quantity of truckers into national-level job categories, and no national-level datasets split truckers by all five categories. Furthermore, data on true job counts that applies these splits is typically not available across geographies.

To deal with this, we developed a method to integrate the two most useful datasets that we could find: the [Occupational Employment Statistics] (OES) database[34] and the [Quarterly Census on Employment and Wages][35] (QCEW; used the "annual averages" tables), both from the Bureau of Labor Statistics. We used the 2016 versions of both datasets for internal consistency, as QCEW's latest iteration of its "annual averages" tables are, at the time of writing this report, only current to year 2016, despite OES's last update in 2017.

Why We Integrated OES and QCEW

The OES database contains an estimate of the true level of employment, excluding owner-operators, across numerous professions in the US. It is broken down by "metropolitan statistical areas" (MSA) and "non-metropolitan statistical areas", each of which encompasses multiple counties in the U.S. It also provides other statistics per profession and per MSA including estimated mean and medianwages, hourly wages, and wages at a series of percentiles per profession per MSA. Professions are broken down at a level no more granular than "truckers" (i.e., there are no trucking subcategories, such as FTL or LTL, let alone FTL and long-haul and so on). When excluding owner-operators, it peggs the total national employment level of truckers at approximately 1.7 million.

The QCEW is a *census* for individuals who work at *establishments* that have paid into unemployment insurance. It has a few more granular breakdowns by subcategories of truckers (e.g. general freight trucking, local (i.e., short-haul) and etc.), and each of the statistics are also broken down at the county level. It includes information on the number of establishments, workers, mean wages, taxable wages and contributions to the unemployment insurance program, and last year's percentage changes for these and a number of other statistics.

It might be tempting to use the QCEW dataset, as it appears to have three key strengths (see footnote[36] for critical issues with previous iterations of this project): 1) data is broken down by county rather than MSA, 2) data is disaggregated by some job categories, and 3) QCEW is a census of some kind. However, QCEW, as a census, samples from a specific population—again, those who work at establishments paying into unemployment insurance, which is narrow than the significantly broader population of all truckers. It systematically *underestimates* the population of truckers (or rather, it is not even an attempt to estimate the true population of truckers), pegging the total population at about 25% less than the OES database.

There are also other limitations associated with QCEW. In particular, it classifies occupations using the North American Industry Classification System (NAICS), which classifies job by the nature of the "establishment"—a location-organization entity[37], not the nature of the job itself. For instance,

[34] https://www.bls.gov/oes/special.requests/oesm16all.zip
[35] https://www.bls.gov/cew/datatoc.htm
[36] Both previous years' iterations of this project solely used the QCEW dataset, thereby resulting in significant underestimates of job losses, on the order of ~25%+, regardless of the subjective judgments of their chosen experts. We believe that this is a critical error.
[37] An establishment is a unique combination of an entity/location. For instance, Carnegie Mellon University, a single entity, has multiple locations in Pittsburgh, D.C., and Qatar. These three locations would generally each be identified as a separate "establishment". There are some exceptions to this rule, however.

Carnegie Mellon University—Pittsburgh, being a primarily educational establishment, might have all of its jobs categorized under a NAICS code related to education, even though it has numerous job related to cleaning, food services, and technology only indirectly linked to student instruction.

Therefore, we chose to develop a process that combines the strengths of both worlds, while explicitly recognizing the assumptions we have made. We used the OES's estimate of the true number of truckers (again, excluding owner-operators) across localities of the nation, and combined that with information on the *proportions* of truckers in each industry per locality provided by the QCEW dataset. While we knew that QCEW's counts are not true counts of truckers, we thus assumed for example, that the proportion of all truckers that QCEW identifies being as generalized trucking in Allegheny county represents an estimate of the true proportion of Allegheny trucking jobs that are non-specialized, even though QCEW's statistics are derived solely from a subset of truckers.

Procedure

Pre-Processing the OES Database

The integration process for these datasets was a nontrivial exercise, and so the following section provides a high-level technical overview of the judgments and calculations we made for this procedure. We have omitted minor cleaning procedures from this document.

We first retrieved all rows that corresponded with trucking in OES, using the occupational code[38] of "53-302"[39], and retrieved geographic breakdowns of these job counts by MSA and non-metropolitan statistical areas (n-MSA). Now these MSA and n-MSAs are not particularly useful for our maps as they are areas whose borders are unrelated to commuting zone definitions. Furthermore, they cannot be directly merged or computed in combination with most other datasets (e.g., re-aggregation by commuting zone) because of many-many relationships between MSAs and other geographic breakdowns.

We therefore used the BLS's MSA and n-MSA definitions[40] to link each MSA and n-MSA unit to specific counties and their FIPS county codes. We then assumed that the total population of truckers for each MSA/n-MSA were evenly distributed between each associated county (e.g., if commuting zone 1 had 500 truckers and two counties, both its first and second county would each be associated with 250 truckers). We did this by dividing the total trucking employment in each

[38] This is different from the NAICS occupational code, which, again, classifies job by establishment. The occupational code used here attempts to get at the true classification of a specific job. A occupational code may be linked to multiple NAICS classifications. E.g., where a trucking job is located at an educational vs. a food establishment. Furthermore, a specific NAICS classification may be linked to multiple occupational codes. E.g., where a food establishment has cooks, truck drivers, and managers.

[39] We also filtered by NAICS code of "00000" to ensure that we did not include any redundant, but more granular breakdowns of counts by specific occupation code-NAICS relationships.

[40] MSA and n-MSA definitions are for 2016 are no longer on the BLS website, and so we used the .xlsx provided for 2016 available on the Internet Archive, rather than the BLS website itself. Link: https://web.archive.org/web/20170504163825/https://www.bls.gov/oes/current/area_definitions_m2016.xlsx

MSA by the number of counties in that MSA, and subsequently linking this set of numbers to each county. We did the same here to estimate the total numbers of employed individuals in each county, though we multiplied these figures by 1.101 to account for BLS's estimate of self-employed workers.[41]

Merging OES Statistics with QCEW

Now, the QCEW dataset came in the form of hundreds of spreadsheets for each of the different NAICS codes. We focused on the following NAICS-based occupation identifications because of their direct relationship with the trucking industry: generalized short-haul, generalized long-haul FTL, generalized long-haul LTL, specialized short-haul, specialized long-haul, and movers. We essentially calculated the proportion of workers in each of these categories by dividing the total number of workers counted by QCEW by the number of workers in each of these categories. Again, we focused on county-level breakdowns, though we retrieved state-level breakdowns for later use in some imputation processes. A small percentage of truckers (3.5%) could not be attributed to any county in the census, but because we are using the QCEW to derive *proportions* rather than totals, this does not affect our results much, if at all. We subsequently merged our QCEW statistics with the edited OES dataset.

In the retrieval of this QCEW data, we noticed that there were a few rows where a small number of workers would be placed in a separate row for federal government employment, despite having the same NAICS and FIPS county codes previously seen. This issue was periodically seen throughout our procedure. For each time we saw this occur, we reaggregated statistics to include these workers.

Deriving the count of jobs per job category per county is not as simple as dividing QCEW's trucking category-level estimates by county-level estimates across all trucking jobs identified by QCEW and multiplying this proportion by OES-derived counts. There were some counties where the OES dataset counted a non-zero number of truckers, but where QCEW also identified no truckers. If we were to combine OES's non-zero estimate of the true count of truckers with QCEW's null proportions in these instances, we would create *underestimates* of truckers, as aggregation algorithms would be unable to count the product of a non-zero number and a null value. This was a non-trivial issue, affecting well over 20% of workers as counted by OES.

This is where our state-level statistics were particularly useful. We used state-level statistics aggregated by the BLS itself in its production of QCEW data[42] to impute the proportions in each trucking job category for these rows. If, for instance, county X had 50 truckers estimated by OES, and QCEW estimated 0 truckers in the county, simply multiplying OES's statistic by an incomputable QCEW category-level proportions would result in an undercount of these truckers.

---

[41] https://www.bls.gov/careeroutlook/2014/article/self-employment-what-to-know-to-be-your-own-boss.htm

[42] These statistics are included in the QCEW spreadsheets. They are associated with a different aggregation level code from rows that are aggregated at the county level.

We can instead pull the estimated proportion of truckers in long haul, for example, for the state that county X belongs to, and multiply this proportion by the OES-derived statistics.

This created a problem for some counties, as not every county was linked to every available job category. A quirk in the code therefore caused some category-level proportions to add up to numbers less than 1. We renormalized these proportions to add up to 1.

Dealing with Some Key Categories That QCEW Does Not Break Out

Now, we noticed that the QCEW does not appear to fully break down some categories. The categories provided by the QCEW dataset are listed below. Long-distance is long haul; local is short haul; “tl” refers to full-truckload; general refers to non-specialized trucking:

- NAICS 484121 General freight trucking, long-distance tl
- NAICS 484110 General freight trucking, local
- NAICS 484220 Other specialized trucking, local
- NAICS 484122 General freight trucking, long-distance ltl
- NAICS 484230 Other specialized trucking, long-distance

We can see that there are no LTL/FTL breakdowns for specialized trucking of all types and general short-haul trucking. We cannot assume that all specialized long-haul jobs are FTL, given the substantial verified presence (via QCEW) of LTL jobs among long-haul generalized jobs. And for general short-haul trucking, in particular, our review of literature suggested that the ratio between LTL jobs to FTL jobs might be larger for short-haul trucking than for long-haul trucking. However, we still cannot assume that the number of short-haul FTL jobs is zero, as the literature still suggested that short-haul FTL jobs may occupy a significant place in the economy.

We therefore developed a series of procedures to impute the required proportions (and thus, job counts) across counties. For specialized long-haul trucking, we imputed the proportions of subcategory FTL/LTL jobs by pulling state-level proportions of FTL/LTL jobs among general long-haul truckers. It is less reasonable to assume that state-level FTL/LTL trends hold the same between long-haul truckers and short-haul truckers. However, for the lack of better data, we used these same proportions to impute FTL/LTL subcategory breakdowns for both generalized short-haul truckers and specialized short-haul truckers.

Dealing with Unionization and Owner-Operators

There are few, if any, reliable national statistics on unionization and the presence of owner-operators among truckers, and there are no datasets breaking these numbers down across geographies and trucking job categories. We therefore assumed that the proportion of unionized workers (pegged at 10.4%[43]) was constant across all counties and job categories. Similarly, we assumed that there was a constant proportion of owner-operators (about 25%), which were not counted in OES estimates, across all counties and job categories. We added this 25% of workers to the total, given OES’s underestimation of owner-operators, resulting in a higher total. The

[43] Sourced from http://www.unionstats.com/

assumption of constant proportions across categories and geographics may not be true, as unionized workers and owner-operators may be geographically concentrated or concentrated in specific kinds of trucking (e.g. unionized workers may be concentrated among short-haul LTL jobs). However, we believed that this was the best we could do for the lack of better data.

Recap

The totality of this procedure gives us our best guess at $p_{i,l}$ in the s-curve equation for our job loss model, where job losses per category per location are specified as the following:

$$y_{i,l} = \frac{e^{b_i(x-z_i)}}{1+e^{b_i(x-z_i)}} * p_{i,l}$$

Parameters b and z have previously been specified during our consultations with Steve Viscelli, and so with $p_{i,l}$, we can now calculate job losses over time for each location and category.

Final Data Structures and Calculating Job Loss and Wage Loss After Pre-Processing

Data Structures and Calculating Job Loss

With these proportions, we ultimately structured the data roughly along the lines of the format below:

Job Counts Data Structure

Location	General or Specialized	Long Haul or Short Haul	FTL or LTL	Unionized or Not	Owner-Operator or Not	Employee Calculation
1	General	Long Haul	FTL	Unionized	Owner-Operator	X
1	General	Long Haul	FTL	Unionized	Not owner-operator	X
...	...	...	...	...	...	...

As we can see here, this structure provides multiple rows for each location, with each row being linked being linked to the probable present-day count of workers, derived from the product of proportions and OES statistics integrated and calculated using previously specified procedures, in a specific location and job category. Each category is defined by its association with values in all five attributes: specialization, distance, load type, unionization, and owner-operator status. In theory, we can have up to 32 rows for a single location, for the 32 different categories.

This structure enables us to map coefficient values for curve steepness, *b*, and delay, *z*, according to the specific attributes of each category. See "Applying the S-Curve to Each Job Category" for details of the procedure. And with these coefficients, we can calculate the job loss for each county, category, and year using the s-curve equation, thereby creating a data structure roughly similar to the table below:

Job Loss and Change Over Time Data Structure

Location	General or Specialized	Long Haul or Short Haul	FTL or LTL	Unionized or Not	Owner Operator or Not	Employee Calculation	Year	Job Loss
1	General	Long Haul	FTL	Unionized	Owner Operator	X	2018	J
1	General	Long Haul	FTL	Unionized	Owner Operator	X	2023	J
...	...	...	...	...	...	...		

Now, we essentially have one row per location, per category, per year of the estimate. This data structure also enabled us to scale up population statistics, including, but not limited to, per-location estimates of employee calculations, the working-age population, the total population, and etc. Once again job losses are modeled by the following equation:

$$y_{i,l} = \frac{e^{b_i(x-z_i)}}{1+e^{b_i(x-z_i)}} * p_{i,l}$$

However, to scale the estimates up by a rough estimate of 0.5% population growth over time[44], $p_{i,l}$ is recalculated via the following equation:

$$p_{i,l,x} = \left(1.005^{(x-baseyear)}\right) * \left(p_{i,l,x=baseyear}\right)$$

$p_{i,l,x}$ is thus the population of employees in category *i*, location *l*, and year *x*, and *baseyear* is the year from which we are projecting job loss and population growth (i.e., the starting point of our dataset).

Though this scaling does not affect the *percentage* of trucking jobs lost in each year (percentages remain the same since we assume that trucking jobs grow at the same rate as population growth if

[44] Here, we are, in effect, assuming that the proportion of the U.S. workforce comprised of truckers would have remained constant over time if there were no technological changes related to adoption.

automation were to not occur), this expands the base for job loss to occur over time, resulting in higher totals for job and wage loss in later years. Our research did not yield reliable estimates for the probable growth of the trucking industry over the 2030-2040 period, let alone the 2060-2070 period when we, in consultation with Steve Viscelli, believe that job loss due to automation is likely to take off, and so we did not believe that it would be reasonable to peg the growth of trucking at a specific number that is different from assumed growth of the general population.

Calculating Wage Loss

By multiplying job loss by an expected wage loss per job in each category and location, we can calculate total wage loss aggregatable across a variety of breakdowns. Due to time constraints, we developed a simplified procedure for calculating wage loss across locations in the U.S.

Our review of literature suggested that truckers' demographics most closely resemble the average[45] high school graduate. If a trucker were to lose his job due to automation, his wage loss would thus be the difference between his wages and the average wage of a high school graduate in his locality. Though the QCEW dataset includes statistics on wages broken down by category[46], we simply took the difference between the average wage of a trucker in the county and the average wage of a high school graduate in the county. If this difference were negative (i.e., where truckers earn less than the high school graduate), we computed the per-job wage loss as zero, rather than a negative number. These instances only affected a few counties across the U.S.

We obtained the average wages of high school graduates per county from the 2016 American Community Survey conducted by the Census Bureau.[47]

As a final refinement, we assumed that real wages for all occupations, including trucking and occupations held by non-trucker high school graduates, would increase by an average of 0.2% per year over time. We rescaled all wage statistics over time as appropriate.

Aggregating Statistics by Commuting Zone

All statistics up through this point were computed at the county level. For instances where the original source was not at the county level (e.g., OES data), we had relied upon locality-county relationships and the previously described process to calculate county-level statistics.

We, however, found it useful to reaggregate statistics at the commuting zone level, as people regularly cross county boundaries to reach their jobs. Individual mobility within the commuting zone can be relatively high, even if the commuting zone spans multiple counties. Commuting zones

[45] For wage loss calculations, it is important to use the *average*, rather than median statistics. This is because we can recover a total (e.g., total wage loss) by multiplying averages by counts. On the other hand, the product of median statistics and counts does not necessarily equal the total we are trying to compute.
[46] Despite its limitations, it may be possible to estimate average wages according to some breakdowns, such as generalized vs. specialized truckers in each county.
[47] Source: https://www.census.gov/programs-surveys/acs/

are thus a better representation of communities across the nation. We computed sums and weighted averages as appropriate across counties and commuting zones for each relevant statistic. Though this had some nuanced implications on the structure of our data and how the data might subsequently be displayed in Tableau, we developed internal workarounds to deal with these issues.

Future Work

Despite our attempt to deal with the heterogeneity of the trucking industry, we identified several areas for possible improvement on the process—areas where we were unable to apply substantial refinements to our process in the time given:

- We assess that without major revolutions in the LIDAR foundation for automated driving technology, jobs in areas with substantial snowfall will be difficult to automate. We did not account for the relationship between snowfall metrics (e.g. days of snow per year or snow inches per year) and job loss. Including this in a model would depress job loss estimates.
- We did not come across reliable estimates for the growth of trucking jobs over the decades beyond 2030. However, it may be possible to revise our models and conduct some sensitivity analysis for scenarios where the growth in the trucking population (or, demand for trucking) differs from the growth of the general population.
- Later in the course of this project, we came across a third dataset, the Commodity Flow Survey (CFS), which provides count, mileage, origins and destinations, and other statistics for shipments (not personnel) broken down by numerous variables, including load weight, distance traveled, industry association, and more. This could have been useful in developing a more refined understanding of how the trucking industry behaves. We believe that it would be particularly useful in refining a few points in our analysis:
    - QCEW did not break down categories across all attributes we were interested in exploring. Without an approximation of the ground truth on, for example, how unionized workers might be concentrated in short-haul LTL jobs, we sometimes assumed that there was no internal correlation between some of the attributes; on other occasions, we took on an imputation processes that, as previously specified, may not be ideal. Exploring how categories of workers in one attribute might be particularly concentrated in categories of workers broken down by another attribute could help future analysts refine our estimates.
        - CFS could have enabled us to explore how shipments, and thus job counts, display covariance between these categories.
    - QCEW also did not break down different types of specialization beyond just “specialized” vs. non-specialized. Certain kinds of specialization might be exceptionally more or less vulnerable to automation, and so it may be more productive to reference a more specific breakdown (e.g., the use of hazardous materials, the industry association, etc.)
        - CFS’s records of industry associations with each shipment can yield more nuanced insights on how vulnerable a particular shipment, and thus job,

might be more or less vulnerable to automation. For instance, a shipment with a NAICS code for “grocery and related product merchant wholesalers” might be much more vulnerable than a shipment for “chemical manufacturing” involving hazardous materials.

- Future analysts, however, should note that CFS is a dataset of *shipments*, not *job counts*. There may be a strong correlation between the quantity and mileage of shipments and the number of jobs required to complete them, enabling analysts to indirectly compute job counts that are more refined than using just the OES and QCEW datasets. However, analysts must take the first step in determining 1) whether this correlation can provide reasonable representations of job loss in very specific categories and 2) whether the correlation even exists before they can perform these computations. This is not a trivial exercise, but we believe that integrating CFS would may make future analysis far more rich and nuanced.

Job and Wage Loss Results Analysis

Using the aforementioned procedures, we computed statistics on the scale of wage and job loss over time. They point to annual wage losses of $4.5 billion by 2033, $13.9 billion by 2053, and $29.0 billion by 2078; furthermore, we estimate job losses on the order of 320,000 by 2033, 930,000 by 2053, and 1.8 million by 2078. One of our downloadable dashboards (link[48]) enableis you to visualize national-level statistics over time and break down estimates by single or multple selections of the 32 job categories.

We also computed population-focused and wage-focused statistics for each commuting zone, enabling analysts to understand the impact of job and wage loss relative to the populations and economy of each commuting zone across the U.S. Maps (link[49]) generated from these statistics suggest that the impact of trucking job losses will be felt disproportionately across Middle America.

Key Metrics

The interpretation of national statistics on job loss and wage loss are relatively simple: they represent the totality of job and wage losses occurring over time in present-day dollars, assuming a constant population growth across all jobs and demographics of 0.5% and a constant real wage growth rate of 0.2%.

We also produced several other statistics for each commuting zone, which can be interpreted as the impact relative to the size of the commuting zones from economic and population-focused perspectives:

- Trucking wages loss divided by total wages of the commuting zone: this can be interpreted as a percentage of total wages that are lost due to trucking automation. It presents an understanding of how much change a local economy is going to experience in dollar-wage terms as a result of trucking automation.
- Trucking jobs lost divided by total working-age population of the commuting zone: this represents the number of trucking jobs lost for every individual aged between 15 and 64. We retrieved the quantity of individuals between these ages in each locality from the Census Bureau[50]. It presents a population-focused understanding, measuring how many people are going to be displaced relative to the number of people who may be capable of work.
- Trucking jobs lost as a percentage of trucking jobs in the locality. This focuses on the impact of trucking job loss on the local trucking industries, rather than the communities as a whole.

---

[48] Link: https://public.tableau.com/profile/anhvinh.doanvo#!/vizhome/JobLossProjectionsDisaggregatedbyJobType/NationalTrendsOverTime
[49] Link: https://public.tableau.com/profile/anhvinh.doanvo#!/vizhome/JobLossProjections/Maps
[50] Link: https://www2.census.gov/programs-surveys/popest/datasets/2010-2016/counties/asrh/cc-est2016-alldata.csv

We also developed basic background statistics on the wage loss per job (i.e., the wage difference between a trucker and a high school graduate's wage) for each commuting zone.

National Statistics Over Time and Dashboard Tutorials

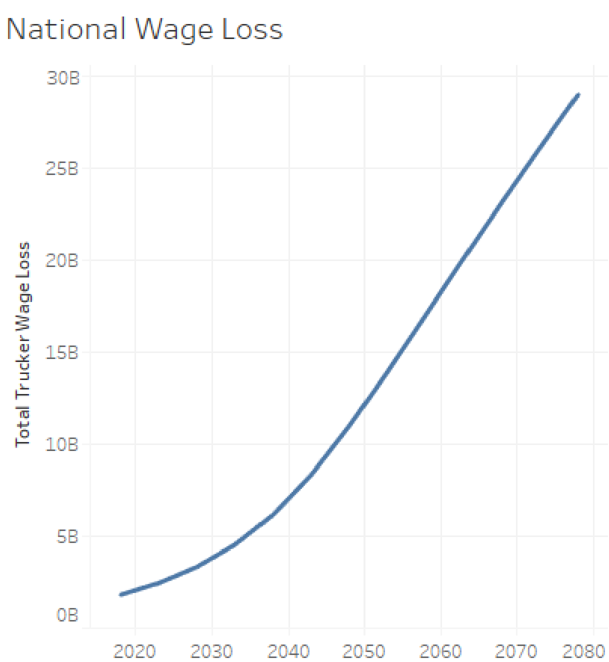
We computed cumulative job loss and resulting annual wage loss every 5 years between from 2018 until 2078, resulting in the following table:

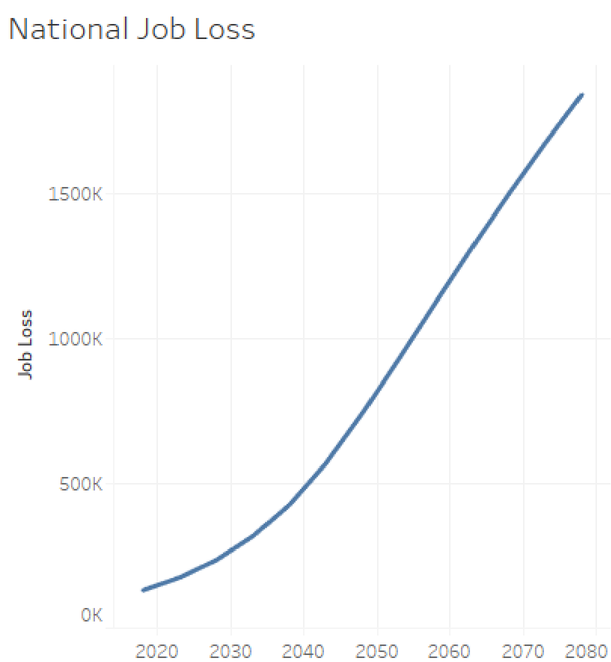
Cumulative Job Loss and Resulting Annual Wage Loss 2018-2078 Due to Trucking Automation

Figures were rounded by significant digits.

Year	Job Loss	Total Annual Trucker Wage Loss
2018	132,000	$ 1,820,000,000
2023	176,000	$ 2,470,000,000
2028	236,000	$ 3,340,000,000
2033	318,000	$ 4,540,000,000
2038	427,000	$ 6,180,000,000
2043	570,000	$ 8,340,000,000
2048	743,000	$ 11,000,000,000
2053	932,000	$ 13,900,000,000

2058	1,130,000	$	17,000,000,000
2063	1,320,000	$	20,100,000,000
2068	1,500,000	$	23,100,000,000
2073	1,680,000	$	26,100,000,000
2078	1,840,000	$	29,000,000,000



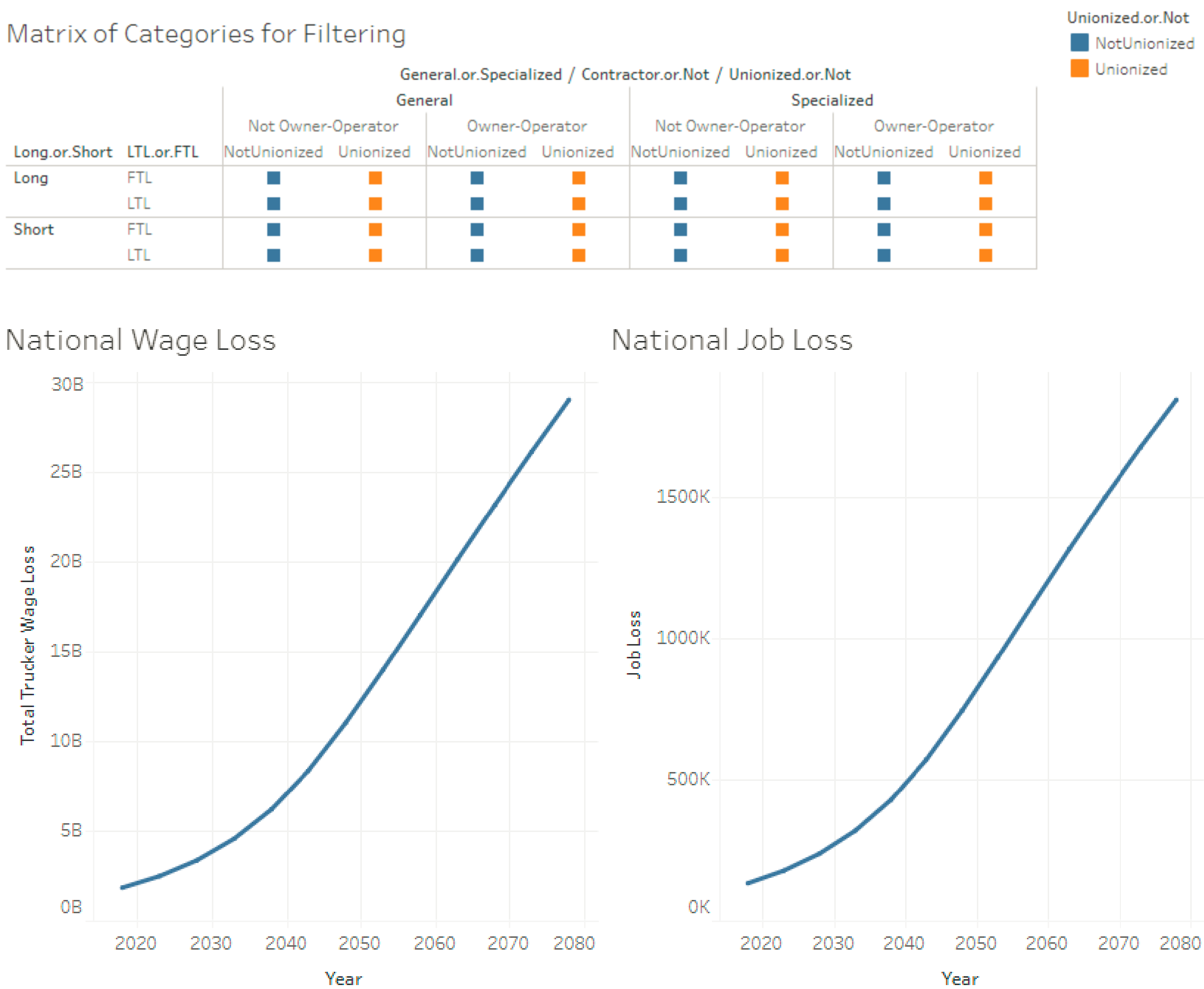


You may notice that our estimates for job loss and wage loss for this year (2018) are nonzero. This is a slight weakness of sigmoid curves: towards their leftward and rightward extremities, they are not as close to 0 or 1 respectively as we would expect. However, estimates in the intervening years (e.g. 2030 and on) may be more well-reflective of our analytical judgments.

On one of our downloadable dashboards (Job Loss Projections Disaggregated by Job Type, link[51], YouTube demonstration[52]), you can drill down on national-level statistics to specific categories. Selecting individual or multiple job categories in the "Matrix of Categories of Filtering" panel will immediately update the plots below.

For example, the current matrix has all categories selected. Therefore, visuals below represent the totality of wage loss and job loss estimates across all categories.

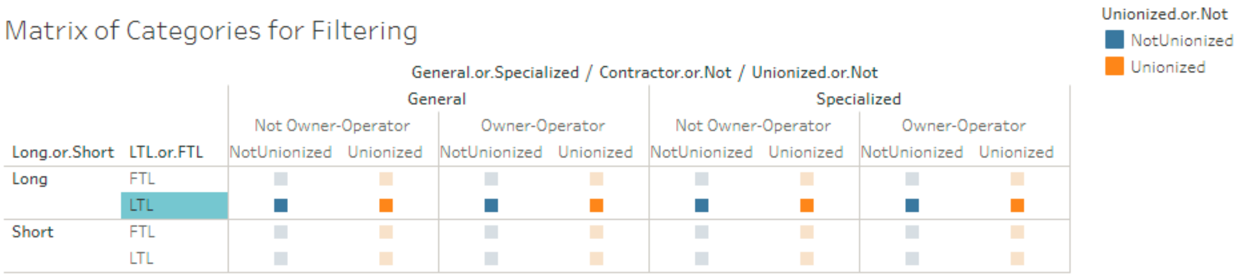
National Statistics Dashboard Demonstration: Across All Categories

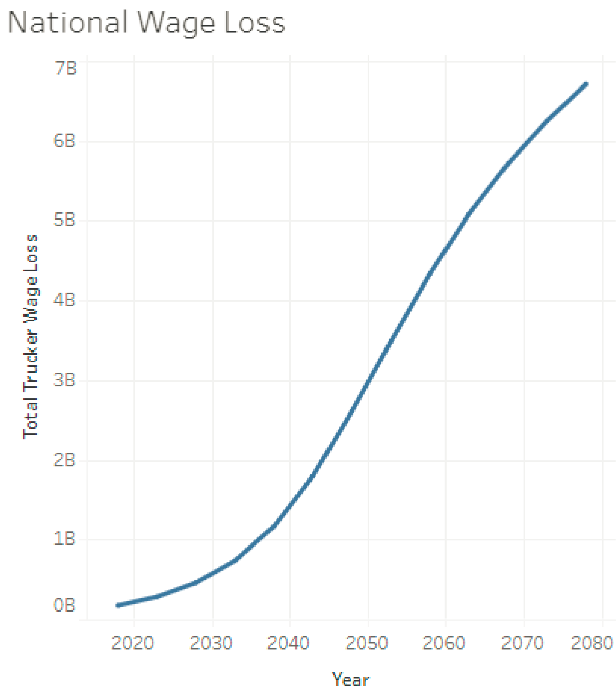


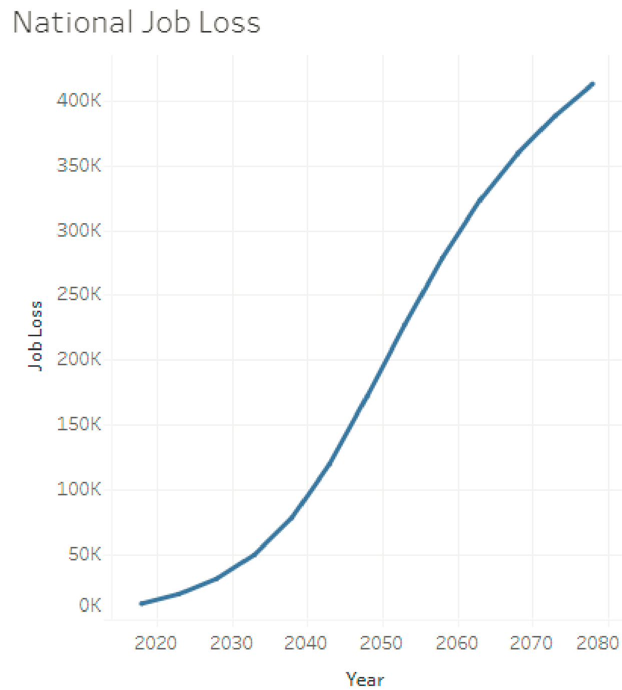
Now, if we selected just long-haul, less-than-truckload (LTL) truckers, we would see the following:

National Statistics Dashboard Demonstration: Long Haul, LTL Truckers

[51] Link: https://public.tableau.com/profile/anhvinh.doanvo#!/vizhome/JobLossProjectionsDisaggregatedbyJobType/NationalTrendsOverTime
[52] Link: https://www.youtube.com/watch?v=uZGHFHfv844







Maps and Dashboard Tutorials

Measuring the Relative Impact on Local Communities

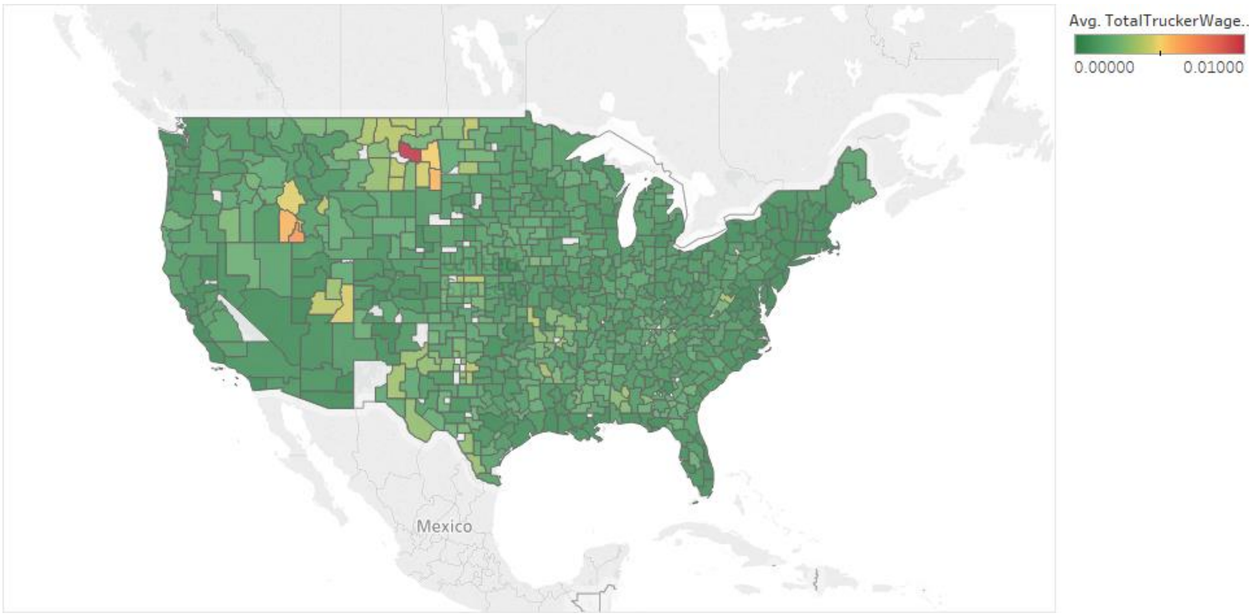
We produced maps (link[53]; sheet 2 in this workbook) of the aforementioned impact statistics in Tableau. When interacting with these maps, you *must* select a single year using the "Year Selection" panel for the statistics to be sensical.

Trucker Wage Loss Due to Automation as a Proportion of Total Wages in the Commuting Zone (YouTube demonstration[54])

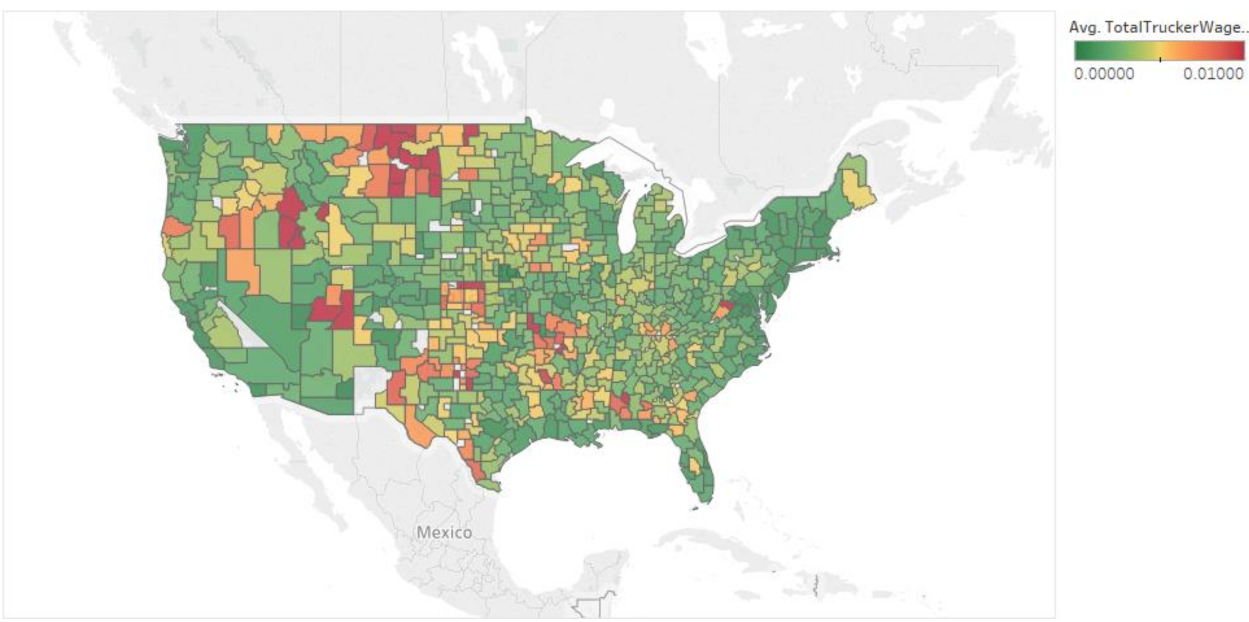
Numbers in the legend are in decimal form. For instance, a 1% wage loss, which is the point where commuting zones are completely colored red, is represented by a value of 0.01.

[53] https://public.tableau.com/profile/anhvinh.doanvo#!/vizhome/JobLossProjections/Maps
[54] Link: https://www.youtube.com/watch?v=6aCu88GO5Jg

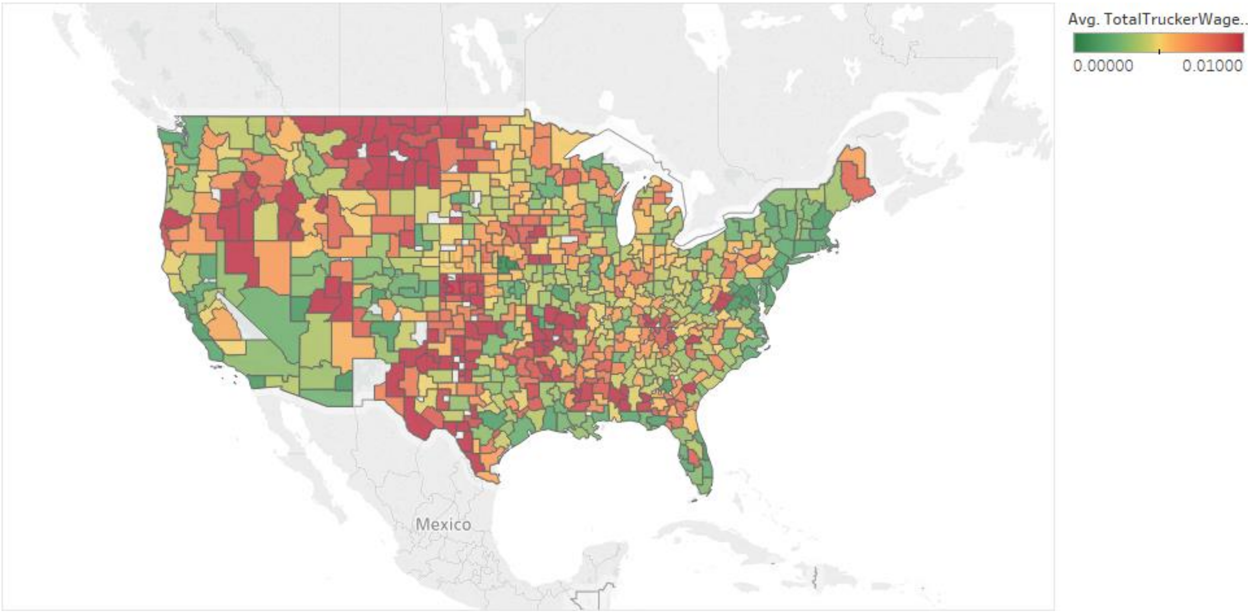
Year 2033



Year 2053



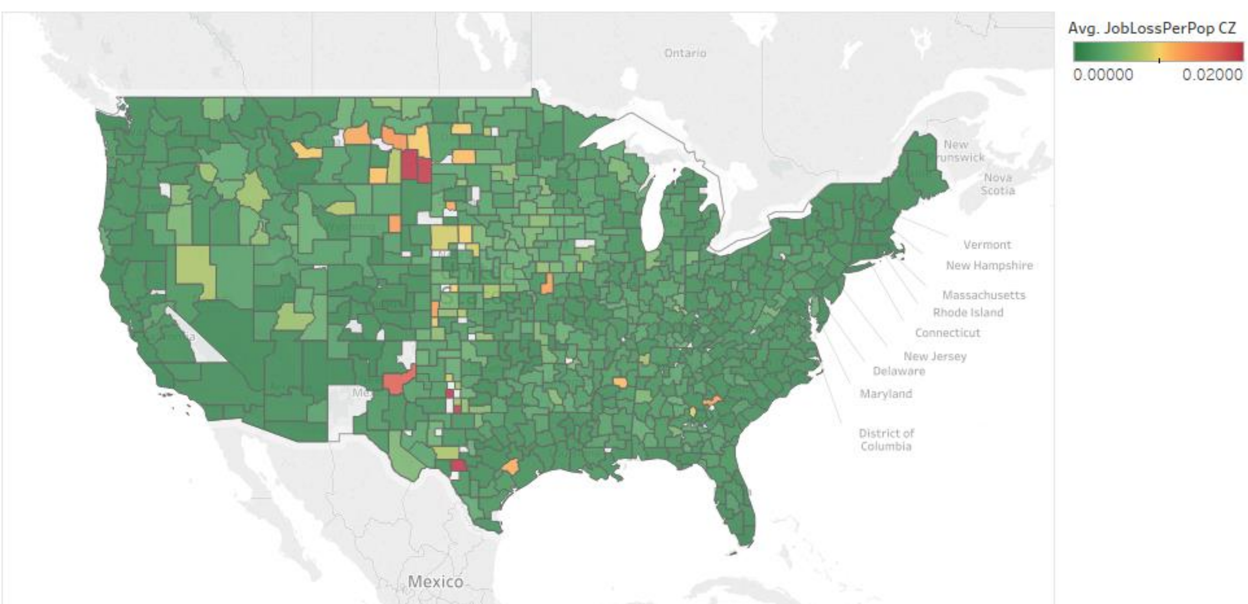
Year 2078



<u>Trucker Job Loss Due to Automation as a Proportion of Working-Age Individuals in the Commuting Zone (YouTube demonstration[55])</u>

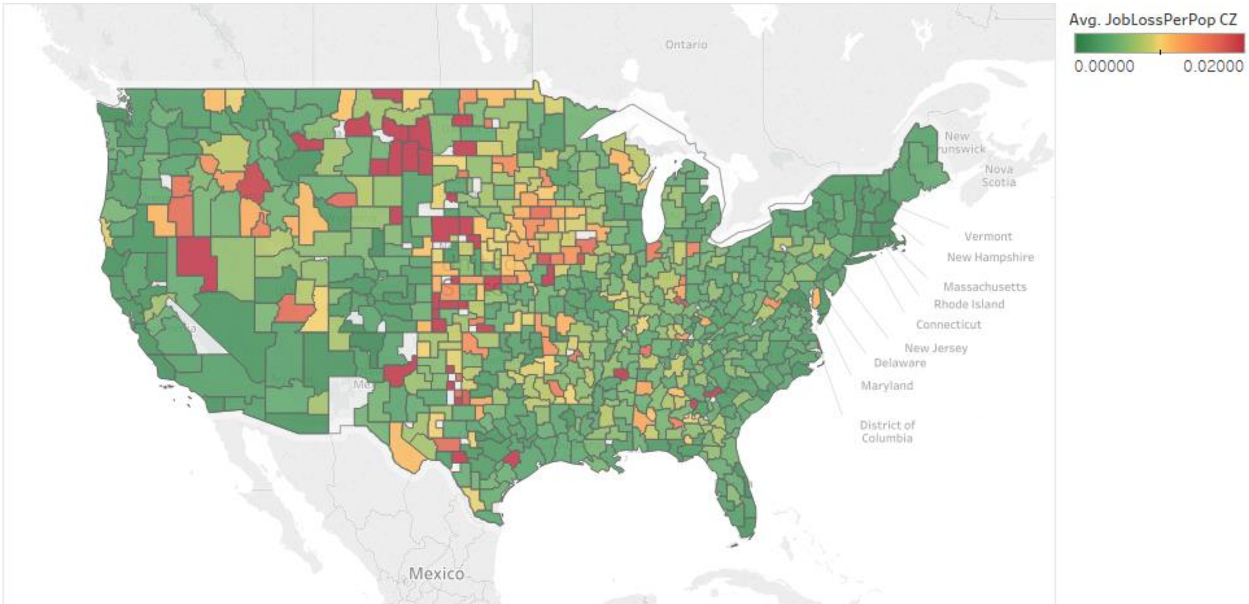
Numbers in the legend are in decimal form. For instance, a 2% job loss as a percentage of working-age individuals, which is the point where commuting zones are completely colored red, is represented by a value of 0.02. Working-age population is defined as the number of individuals in the commuting zone aged 15 to 64.

Year 2033

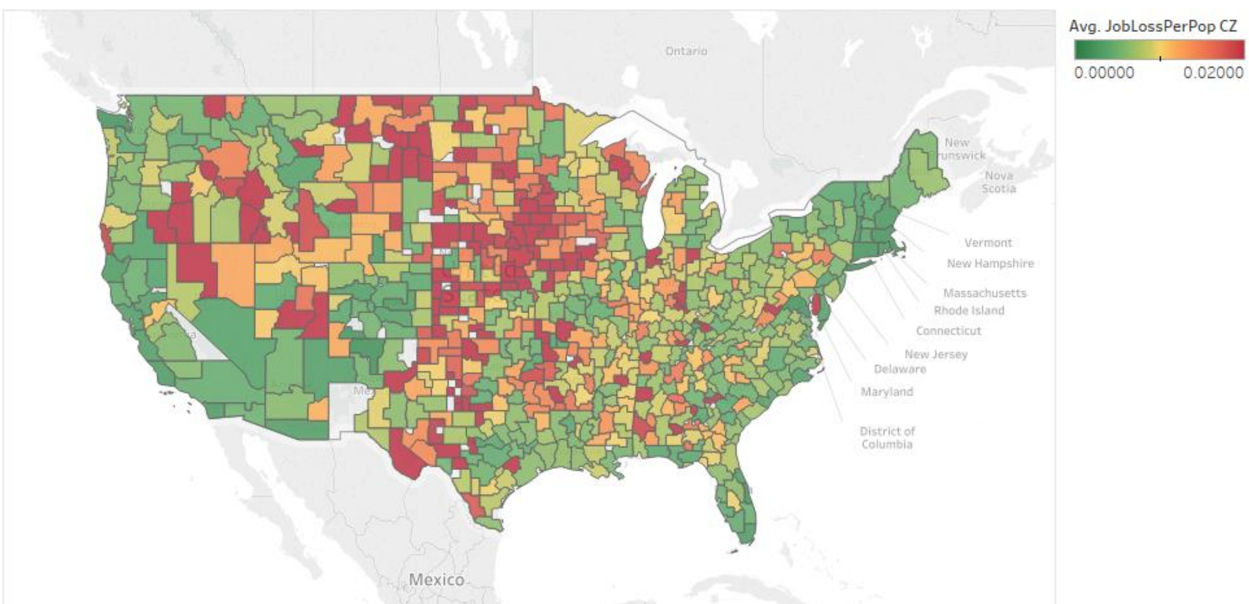


[55] Link: https://www.youtube.com/watch?v=EOxtFA7P-eg

Year 2053



Year 2078



All these maps point to the beginnings of significant job and wage losses occuring by year 2053, with far more widespread losses by 2078. In fact, job and wage losses by 2078 may affect many commuting zones by a factor of 1-2% or more.

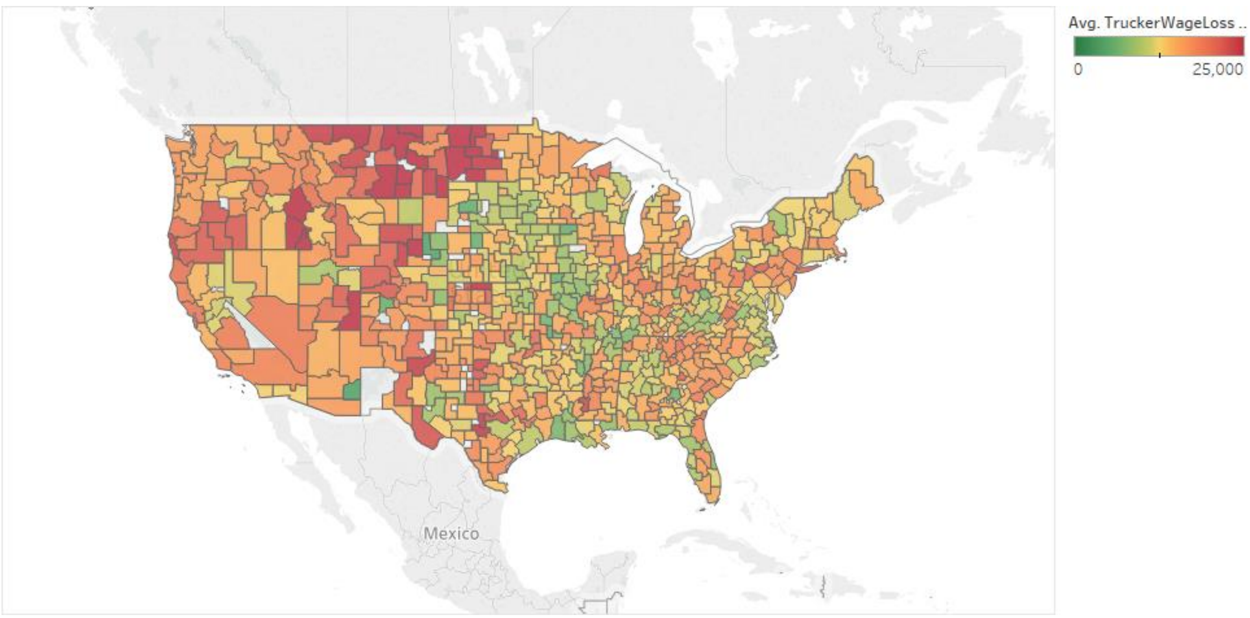
The most dramatic losses relative to local community parameters appear to be concentrated in Middle America. Measures of wage loss relative to community wages appear to increase quite early in the Montana-North Dakota region, while measures of job loss relative to the working population in the community appear to increase quite early around the tri-state region of South Dakota, Iowa, and Nebraska. However, the impact of automation is projected to be more widespread than just these regions, especially by 2078.

Background Statistics

Furthermore, we outputted a few maps to contribute to some background information on why these impacts might be occurring.

Trucking Wage Loss per Trucking Job Loss

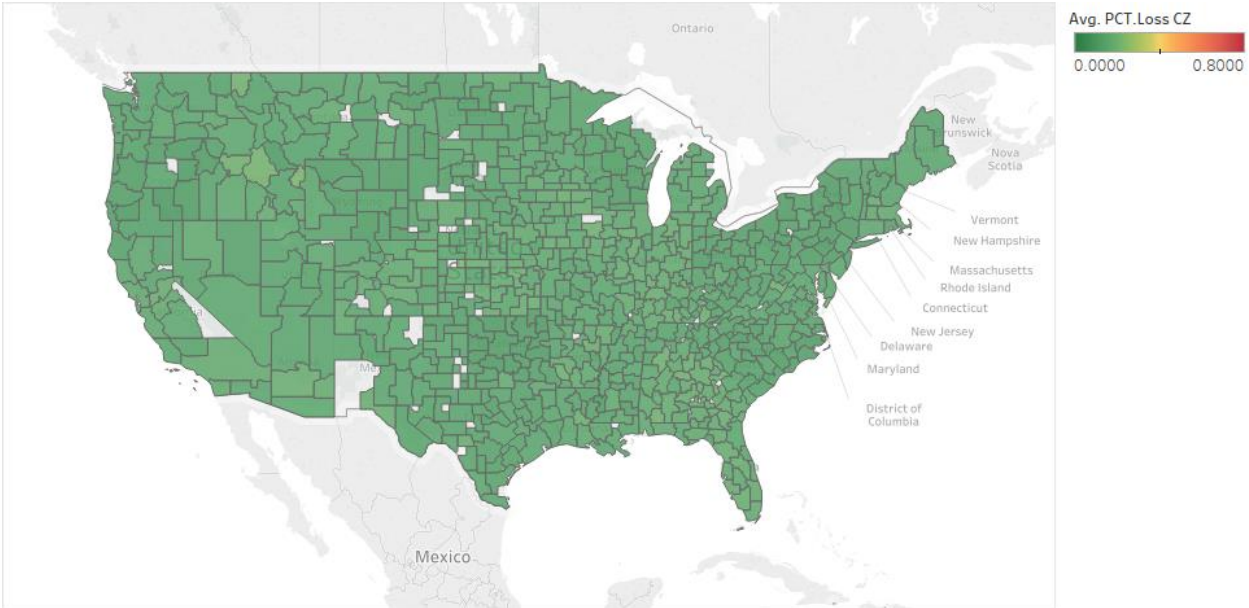
This map represents the quantity of wage loss that would occur for each trucking job that is lost.



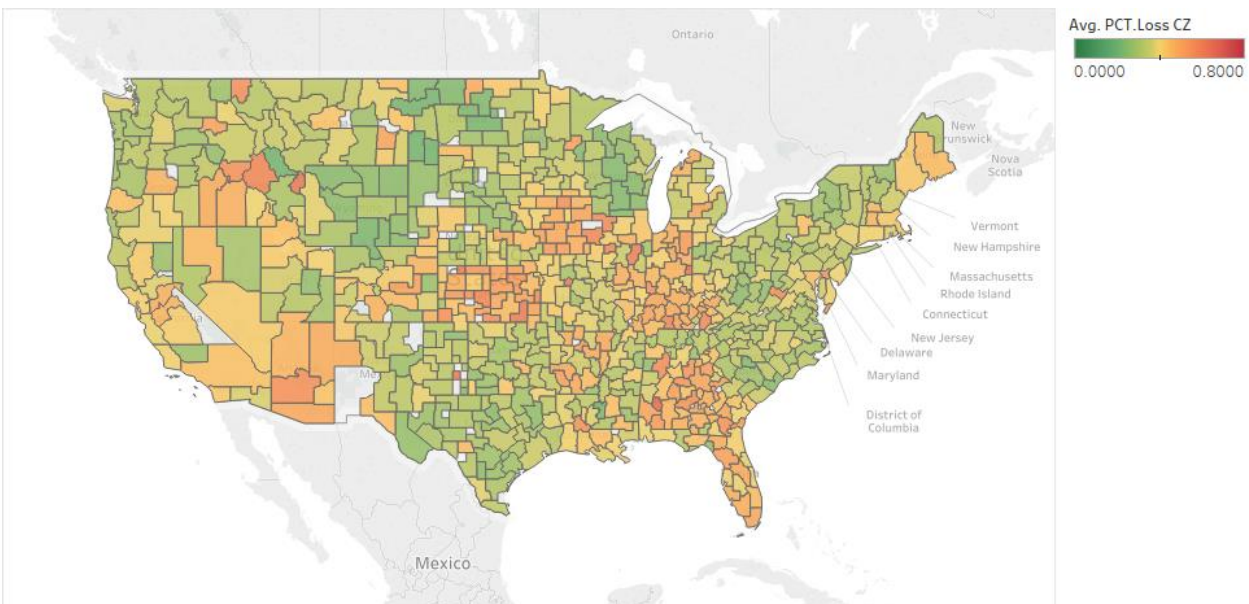
This map essentially represents the premium employers give for truckers and their skillsets, relative to the average high school graduate. Wage losses appear to be especially hurtful per job in the Montana-North Dakota region, and less so in the Iowa-Illinois-Missouri region.

Trucking Jobs Lost per Trucking Population

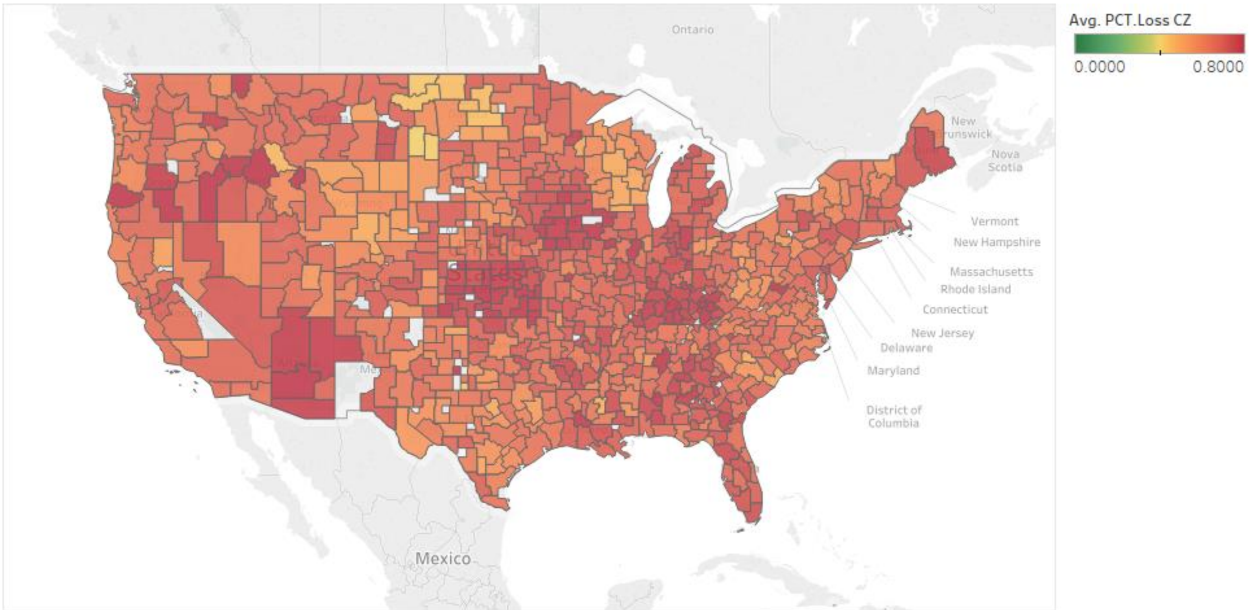
Year 2033



Year 2053



Year 2078



These maps show that as a percentage of the trucking industry, we project that job losses will pick up with substantial activity picking up by 2053 and may be deep and widespread by 2078. A few states are projected to begin this process particularly early, including, but not limited to, Arizona, Kansas, Iowa, Alabama, and Florida.

Scoping a Wage Insurance Program

As a result of our above analysis, we see that wage loss will impact truckers in Middle America most immediately, but that commuting zones across the country will be impacted. Given the nation-wide issue of job displacement due automation, we then explored wage insurance as a policy solution at the federal level. Wage insurance as a policy instrument has received backing from prominent economists such as Robert Lalonde[56] due to the benefits it offers as redistributive tax to protect vulnerable workers, but also in terms of economic efficiency. Policies currently in place such as unemployment insurance, Social Security Disability Income Payments (SSDI), and retraining programs are insufficient to address future wage loss due to automation. Respectively, they are too restrictive; relatively inefficient by disincentivizing people from working; and too poorly funded to be effective at scale.

Wage insurance offers a host of benefits as a potential policy solution. First, it would provide a safety net for workers adversely impacted by automation. We see through this paper how valuable this would be to the trucker population. However, we can extend this premise to the entire population of workers in the U.S. given that a percentage of workers in the economy at large (beyond truckers) will lost their jobs as automation and AI are adopted. Second, wage insurance should result in a reduction in the Social Security Disability Insurance (SSDI) payments that would exist in the absence of a wage insurance program. We hypothesize that 9.9% of trucks drivers who lose their job would fake disability to collect SSDI without wage insurance, based on estimates from Autor, Dorn and Hanson who study labor market shocks and worker responses. Because faking disability is fraudulent and not especially lucrative, we assume that only .9% of those who would otherwise fake disability will continue to do so with wage insurance in place. In this way, a wage insurance scheme would result in a reduction of payments avoided to 9% of displaced truck drivers. By motivating 9% of displaced truck drivers to re-enter the workforce, savings in SSDI could be as large as 1.4B according to prior estimates[57]. Finally, it should be noted that wage insurance would also result in GDP gains due to a larger workforce contributing to national production that would otherwise be the case (e.g. if 9% of displaced truckers left the workforce). The factor by which wage insurance would result in an increase in GDP has been estimated to be between 1.33 and 1.68 US dollars[58].

To scope a wage insurance program, we consider the following criteria. Wage insurance will cover only those workers who lose their jobs and have to move to lower-wage jobs. While our primary focus is on covering trucker job loss due to our precise estimates in this domain, we also include the cost to cover non-truckers in our model, using the best approximations currently available for job and wage loss among highly tenured workers. Our modeling presumes that wage insurance will

---

[56] "The Case for Wage Insurance," Robert LaLonde, CSR No. 30, September 2007, Council on Foreign Relations.

[57] Cristofoletti, Fabio, Jingyi Lu, and Katharine Millard. "Wage Insurance As a Response to Autonomous Truck Driving." 2017.

[58] Cristofoletti, Fabio, Jingyi Lu, and Katharine Millard. "Wage Insurance As a Response to Autonomous Truck Driving." 2017.

- Cover a fraction (f) of lost wages for (n) number of years
- Require mandatory participation from all workers employed in the US (in order to avoid adverse selection), regardless of the probability a given worker will lose his/her job
- Be implemented as a tax on salaries where the premium charged per person is a proportion of one's annual wage.

Further, our modeling uses the following key assumptions and parameters:

1. All truckers who lose their jobs up until a given year (e.g. 2033) receive wage insurance at the same time (2033)
2. Wage insurance only covers those re-employed, and at a lower wage
3. The re-employment rate (for both truckers and non-truckers) will equal the national average re-employment rate plus the SSDI avoidance rate
4. Employment growth per year will occur at the rate of .5% (in line with Bureau of Labor Statistics estimates)
5. Insurance offers 50%[59] or 100% wage loss coverage, for a period of 1, 5 or 10 years
6. A wage insurance program should be actuarially fair (and as such, the program's coffers should balance to 0 at the end of the program period's duration).

Wage Insurance Costs: Methodology

In the modeling the cost of a wage insurance program that will adequately cover a fraction of wages for workers displaced by automation, we use the following equation. It allows us to solve for t, the tax rate that will need to be levied on the entire US employed population given the mean wage in the US, in order to cover the cost of insuring truckers plus the cost of insuring non-truckers.

$$tNW = q_d p_d{}^* f T_d L_d + q_g p_g{}^* f T_g L_g$$

In our model, d represents truck drivers, g represents the general public and variables of interest include the following:

t = tax on the mean wage for wage insurance (endogenous variable)
N = employed population in the US (paying into wage insurance)
W = mean wage in the US
q = percentage that lose their job
p* = percentage that relocate to a lower paid job
f = fraction of loss ensured (endogenous)
T = number of people disrupted by technology

59 In "The Case for Wage Insurance," LaLonde finds, "by limiting benefits to 50 percent of the difference between pre- and post-displacement earnings, most displacement insurance proposals provide incentives for displaced workers to search for more productive jobs at higher wages."

L = average wage loss

After building this model, we then estimate the following figures:

- N, the total number of people employed in the US, equals 160,818,740 according to 2016 US Census data[60]
- W, the mean annual wage in the US, equals $49,630 according to 2016 BLS OES data
- f, the fraction of wage loss covered, will equal 50% or 100% depending on the model scenario below. (While Lalonde proposed that a 50% coverage rate should suffice to incentive workers to continue looking for higher-paid wages, we also forecast the cost of 100% coverage in our single-period program estimates to be conservative. For multi-period calculations, we use an f of 50%.)
- P*, the re-employment rate, is estimated at 82%[61] for both the truckers and non-truckers who would benefit from this program.

Trucker specific figures include the following:

- q, the percentage of truckers displaced, receives three estimates based on three disruption scenarios: low, medium, and high. Using the job loss percent estimates outputted by our S-curve (in the job loss section of this paper), we see that 2033 appear to be the year at which the "shock" of disruption starts to commence. According to our job loss model, 14.42% of truckers will lose their jobs by 2033, so this is used as the percent of job loss for our medium disruption, or "likeliest," scenario. This becomes the baseline year to being disbursements for a wage insurance program. Our low estimate is found by using job loss estimates for 2028 (what our estimate from 2033 would be if the S-curve were shifted to the right by 5 years). Thus, we use 11% for our low-disruption scenario. Finally, our high estimate is the percent of job loss in 2038, as this would correspond to our estimate in 2033 if the S-curve were shifted to the left by five years. This high disruption job loss estimate is thus 18.91%.
- T, the total population of truckers to be considered is 2,043,911 truckers[62]
- L, the average wage loss for truckers, comes to $13,885[63].

Likewise, we estimate these same figures for the general population that would be disrupted by automation:

- q, the percentage of workers displaced amongst the general population of US workers, is estimated at .2% in a low-disruption scenario, .6% for medium disruption, and 1% of the

[60] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk
[61] 82% estimate stems from 73 percent (national average according to BLS: https://www.bls.gov/news.release/disp.nr0.htm) plus 9% (SSDI avoidance estimate from David Autor, MIT)
[62] Consistent with trucker population estimate used to model job loss. This includes the national estimate for owner-operators.
[63] Figure resulting from our job and wage loss modeling

population in our high disruption scenario. Thus, while .6% is our best guess, we explore a high-disruption percentage of 1% to be conservative.[64]

- T, the total base of people who could be disrupted by technology is 160,818,740 (the total employed in US economy as referenced above)
- L, the average wage loss for the general population, comes to $22,009 (as calculated below using 2016 Census data[65]). To get this figure, we use a method similar to the one we employ in calculating average wage loss for the trucking population. We presume those most vulnerable to job displacement and large wage loss are highly tenured workers (in the age 45-64 demographic). To be conservative, we presume the market will not adequately value the skills they have attained, and their best alternative will be a job paying the mean wage received by others their age who have attained a high school education. As a starting point, we consider the mean wage this age group is currently making (noting that the median age is less subject to outliers, so the mean wage will provide us with a more conservative estimate). Subtracting the mean wage for high school attainment for this age group from the mean wage this age group is currently making, we get a wage loss estimate of $22,009.

Age	Total Pop	Weight	Median	Mean
45-54	26,803,000	0.5644	51,854	71,643
55-64	20,685,000	0.4356	51,725	71,401
45-64	47,488,000		51,798	71,538

Age	High School Mean	High School Median	Some College Mean	Some College Median
45-54	48,763	40,704	59,259	47,213
55-64	50,520	41,206	58,752	48,064

Income 45-64			Difference
Mean	High School	49,528	22,009
	Some College	59,038	12,499
	Mean	54,283	17,254
Median	High School	40,923	10,875
	Some College	47,584	4,214
	Mean	44,253	7,545

I. Single-Period Program

We first calculate tax premiums using a single-period scheme. In such a scheme, insurance payments will only cover a fraction of wage loss for one year. Using the parameters identified above, we then calculate the tax premium (%) that will be necessary to cover truckers (identified as

[64] We note a lack of reliable estimates to date for average annual long-term job loss of highly tenured workers due to technological change. Percentages chosen are the best available estimates, provided by Prof. Lee Branstetter, Carnegie Mellon Univ., and used to provide results that are preliminary and illustrative in nature.

[65] https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-03.html#par_textimage_10

"tax premium" as well a pool inclusive of the general population (identified as "tax premium with other beneficiaries"). This lets us then calculate the tax premium that will be assessed per person, as well as the cost that will be incurred per person (calculated given the US mean wage). Illustrated below is the most likely scenario for a single-period program, using the percentages of jobs lost under medium disruption, and outputting the annual cost under both a 50% coverage scheme and a 100% coverage scheme.

Scenario Characteristics		**Known Parameters**		**Our Fractions**		
% non truckers that will be disrupted by technology	0.60%	2016 US Mean Wage **(W) (OESM)**	49630	Fraction of wage loss Wage Insurance will cover **(f)**	0.5	1
% of truckers that will lose their jobs	14.42%	re-employement with wage insurance **(p*)**	82.00%	tax premium that will be assessed per person **(t)**	0.08%	0.15%
		Average wage loss for re-employed truckers **(L)**	13885	Cost per person ($)	38.44	76.88407696
		Number of Truck Drivers **(T)**	2043911			
		Total employed US population in 2016 **(N)**	160,818,740.00			
		Other beneficiaries	791228.2008			
		Tax Premium:	0.00021022			
		Tax Premium with other beneficiaries	0.000774573			

II. Multi-Period Program

We then consider the costs to extend such a program beyond the one-year model. We model the costs of a program covering wage loss for five years beyond the job loss incurred, as well as a period of 10 years. As with the single-period we assume all truckers who lose their jobs up until a given year (2033) receive wage insurance at the same time (2033). As such, for a five-year period program we consider covering workers from 2033-2038, and likewise for a ten-year period program from 2033-2043. Our multi-period program setup assumes taxes are collected from the working population starting in 2019 continuously through the period of disruption (5 or 10 years). From 2019 until the shock year of disruption (2033), the programs net balance benefits from a 5% rate of return on capital per year. Below we see an illustration of cash inflows and outflows in such a five-year program under the medium-disruption scenario[66]. In this example, we see tax revenue is collected every year to cover the cost of the program for both truckers and non-truckers, and that the disruption period when payouts begin commences in 2033. Entering the figures noted in the methodology section, we then program the model to solve for the optimal tax premium rate at which the "year end net amount" will balance to zero at the end of the final year of the program.

[66] Note the columns for years 2021 through 2032 have been hidden to display the outcomes across Disruption Years 1-5.

	Year 1 of Program	Year 2 of Program	Year 3 of Program	Disruption Year 1	Disruption Year 2	Disruption Year 3	Disruption Year 4	Disruption Year 5
Year	2018	2019	2020	2033	2034	2035	2036	2037
Working Population	160,818,740.00	161,622,833	162,430,947	173,311,573	174,178,130	175,049,020	175,924,265	176,803,886
Total Revenue from Tax	$ 26,314,713,420.18	$ 26,446,286,872.74	$ 26,578,518,280.10	$ 28,358,911,255.58	$ 28,500,705,670.32	$ 28,643,209,092.32	$ 28,786,425,121.41	$ 28,930,357,193.84
Previous Year Return	$ -	$ 18,488,024,818.70	$ 32,242,007,371.00	$ 20,884,150,995.66	$ 16,914,188,435.97	$ 12,845,348,770.84	$ 8,673,186,274.83	$ 4,393,035,465.37
Cost of program-Non Trucke	$ 8,707,070,735.70	$ 14,227,638,004.77	$ 19,775,808,104.54	$ 31,456,446,486.06	$ 31,503,363,736.47	$ 31,550,515,584.78	$ 31,597,903,222.10	$ 31,645,527,785.43
Cost of program-Truckers	$ -	$ -	$ -	$ 1,677,864,873.78	$ 1,677,864,873.78	$ 1,677,864,873.78	$ 1,677,864,873.78	$ 1,677,864,873.78
Year End Net Amount	$ 17,607,642,684.47	$ 30,706,673,686.66	$ 39,044,717,546.56	$ 16,108,750,891.40	$ 12,233,665,496.04	$ 8,260,177,404.60	$ 4,183,843,300.35	$ (0.00)

Employment Level	160,818,740.00	
Mean Wage US	$ 49,630.00	
Average Wage Loss Pop	$ 22,009.00	
Average Wage Loss Truckers	$ 13,885.00	Tax premium amount ($):
Number of Truckers	2,043,911.00	$ 163.63
Tax Premium	0.3297%	
Rate of Return	5%	
Non Trucker Job Loss	0.60%	
Wage Insurance Coverage Levels	50%	
Re-employment Rate	82.0%	
Employment Growth per year	0.5%	
Trucker Emp Loss	14.42%	

Wage Insurance Costs: Results

Our Estimates: Single-Period Program (2033)

Given the results calculated in our single-period model[67], we compare the tax premiums and annual cost per person under our three disruption scenarios. We see in our likeliest scenario, such a program would require a tax premium of .08% for 50% wage loss coverage, or .15% for full coverage. In this scenario, the annual cost per person would be $38 for 50% coverage of $77 for 100% coverage. Depending on the degree of disruption, the cost would range from $17 - $60 per person annually for a 50% coverage, single-period program.

Disruption Type:	Low Disruption	Medium Disruption	High Disruption
Job Loss:	.2% of non – truckers 11% of truckers	.6% of non-truckers 14.42% of truckers	1% of non-truckers 18.91% of truckers
Tax Premium: Non-truckers + truckers 50% coverage/100% coverage	.035% / .07%	.08% / .15%	.12% / .24%
Cost per Person: annually	$17 / $35	$38 / $77	$60 / $121

[67] Results available in "Wage Insurance Model one period_2018_FINAL" workbook

Our Estimates: Five-Year Program (2033 - 2038)

As expected, we see tax premiums would rise when extending the model[68] to cover five years, however premiums still remain relatively low below 1%. In our likely scenario of medium disruption, we see an annual cost per person of $164 for a program providing 50% wage loss coverage. Our most conservative forecast for a high disruption scenario sees premiums hit a max of .76% resulting in an annual per person cost of roughly $378.

Disruption Type:	Low Disruption	Medium Disruption	High Disruption
Job Loss:	.2% of non – truckers 11% of truckers	.6% of non-truckers 14.42% of truckers	1% of non-truckers 18.91% of truckers
Tax Premium:	.10%	.33%	.76%
Cost per Person: annually	$49	$164	$378

Our Estimates: 10-Year Program (2033 - 2043)

Finally, when modeling our most extensive version, we see the following results for a 10-year program covering 50% wage loss. In our likeliest disruption scenario, tax premiums reach .78% with an annual per person cost of $388 to fund the program. In our most conservative scenario, these amounts peak with a tax premium of 1.3% and an annual cost per person of $646.

---

[68] Results for the five- and ten-year programs are available in the workbook titled "Wage Insurance MultiPeriod Program Cost_2018_FINAL"

Disruption Type:	Low Disruption	Medium Disruption	High Disruption
Job Loss:	.2% of non – truckers 11% of truckers	.6% of non-truckers 14.42% of truckers	1% of non-truckers 18.91% of truckers
Tax Premium:	.26%	.78%	1.3%
Cost per Person: annually	$130	$388	$646

Key Takeaways & Policy Implications

Our preliminary results indicate that wage insurance as a policy instrument could be reasonably affordable when measuring the costs it would entail relative to the benefits it would extend. Insurance could cost less than $400/person per year to cover 50% wage loss due to automation for both truckers and non-truckers alike over the course of a 10-year period (according to our likeliest scenario). Beyond this, we note that a wage insurance program would protect workers vulnerable to job loss and considerable wage loss, while mitigating the extent to which automation would increase the considerable income disparity that already exists in the United States. We find that a policy solution such as wage insurance will be needed in the near-term by truckers across the country, as exhibited in this paper's section on job and wage loss forecasts. Further, such a solution will be needed in the long-run as automation continues to be deployed across the economy in diverse industries far beyond trucking. We should note, however, that the procedures applied to not account for new entrants to the programs. For example, if wage insurance begins in 2033, we are ensuring everyone who has lost a job up until 2033, but no one who loses a job between 2033 and 2043. If no parameters on job loss estimates for truckers are non-truckers are changed, these numbers thus represent a somewhat conservative estimate on the cost of wage insurance.

Beyond the potential cost-effectiveness of wage insurance and its benefits for social equity, we also note the political promise such a policy solution could have. Job displacement due to automation and artificial intelligence will be non-partisan; this is an issue that will negatively impact workers in both red and blue states alike who, without a policy solution, would be left behind in a changing economy of technological disruption. Indeed, we see in the instance of trucking job loss that states that states that have historically voted for conservative candidates will likely be impacted first and to a greater extent than those in "blue" states. While redistributive policies tend to get more buy-in from liberal candidates, the fact that considerable constituent pools across both political parties will be adversely impacted from automation lends weight to the viability of a federal wage insurance program.

Areas for Future Improvement

The crux of this paper was to advance current research on job and wage loss due to autonomous driving deployment in the trucking industry, and as such, these estimates have been made with considerable techniques to enhance precision. Due to the considerable job and wage loss estimates forecasted, wage insurance was then explored as a potential policy solution to address this upcoming issue in a proactive manner. However, it should be noted that these finding are meant to be preliminary and illustrative in nature. We recommend that future work build upon these wage insurance analyses by enhancing the program's scope to not assume all recipients lose their jobs and enter the system in the same year (in our case, 2033). To make these estimates more robust, future work should allow for new entrants into the program over the period of coverage (e.g. each of the years of the five or ten-year program as opposed to all entering in 2033 at the job loss rate found up until that point). Such a model would consider the increase in the percent of workers who encounter job loss in each year as the program progresses, and in this way consider the cash inflows and outflows necessary to accommodate them. Finally, these estimates could be made even more robust by including both workforce net growth and nominal wage increases over time.