Visualizing Automation Effects on Labor

Lucas Kitzmueller* — Harvard Univerity

Bhavik Nagda† — MIT

ABSTRACT

The last few decades have witnessed rapid advances in automation technology amidst growing anxiety about the labor market. As emerging technologies mature and gain commercial traction, concerns of jobs displacement will only heighten. We present a data article, driven by associated interactive visualizations, that explores the underlying demographic dynamics of automation displacement. In accordance with prior work (cite Webb), we automation-related labor displacement into three categories- AI, robotics, and software, and explore the relationship between these factors and various geographic and educational factors. In particular, we highlight a surprising finding that while AI-related labor displacement disproportionately impacts highly-educated populations.

1 INTRODUCTION

The current public debate on the future of work, and in particular the impact of artificial intelligence (AI), has two major limitations. First, many public articles lack nuance on how the impact of AI will vary across different groups of workers. Second, a large number of stories rely on the same data source, the automation risk estimates by Frey Osborne (2013). Based on the assessment of machine learning researchers on the feasibility of automation of 70 occupations, Frey Osborne predict that 47% of the jobs in the US are at high risk of automation. This figure has been widely cited in both academia (8078 citations on Google Scholar as of May 18, 2021) and the public debate. Frey and Osborne deserve credit for alerting researchers and policymakers to the displacement risk of emerging AI technologies. However, since then new, and arguably better approaches to quantify workers exposure have been developed. While hardly any article on the impact on AI goes without a reference to Frey Osborne, the alternative measures have not yet received a lot of attention.

In our online publication, 'Who's Susceptible To The Automation Boom?', we attempt to address these two shortcomings of the current debate on AI's impact on labor markets in the US. First, we use a new dataset of occupations' automation risk that to our knowledge has not yet been explored in interactive data visualization. The dataset was created by Webb (2020), and measures the overlap between the text of job task descriptions and the text of patents to quantify the exposure of occupations to automation. Second, we illustrate how the exposure to automating technologies varies across educational groups and geographies. We highlight that higher-educated workers have a higher exposure to AI technologies, breaking with historic trends of worker displacement.

2 RELATED WORK

Broadly, our work adds to a growing number of visualizations exploring the impact of technology on labor markets. There already exist many interactive data visualizations of the risk estimates by Frey Osborne (2013). For example, Whitehouse Rojanasakul (2017) from Bloomberg show these occupation-level automation risk estimates their association with wages and education in a scatter plot. They foreground that in the Frey Osborne data, a college degree is associated with a lower automation risk. Sonnad (2013) developed a similar chart earlier for Quartz, also showing the correlation between wage and automation risk. The website https://willrobotstakemyjob.com/ allows users to query the automation probability of different jobs, and provides additional labor market statistics for each job, such as the number of people employed, the projected employment growth, and the median annual wage. To the best of our knowledge, there exist no other interactive data visualizations of the data by Webb (2020). However, there are two static visualizations. First, the original paper itself includes a bar graph showing the exposure to different automation technologies by education. Second, Muro, Whiton, Maxim (2019) in a report for Brookings show similar bar charts and state-level maps of Webb's (2020) exposure scores.

3 METHODS AND DATA

All data is available in the data folder on the class project github. We rely on the following data sources:

- Data from Webb (2020). The data includes the risk percentiles of exposure to AI, robots, and software for 963 occupations. The risk estimates were calculated by measuring the overlap in verb-object pairs in the task description of jobs as listed in the O*Net database by the Department of Labor, and patent texts. For example, one of the tasks of doctors is Interpret tests to diagnose patient's condition. In this case, the verb-object pairs would be (interpret, test) and (diagnose, condition). To then measure this task's exposure to AI, Webb calculates the frequency of these verb-object pairs in the titles of all AI-based patents. By comparison, Frey and Osborne (2013) showed machine learning and robotics researchers at an Oxford University workshop O*NET task and job descriptions and asked them to hand-label occupations where they are confident that the occupation will certainly be fully automated conditional on the availability of big data or will certainly not be fully automated. They then used the 70 hand-labeled occupations to identify engineering "bottlenecks", and model each occupation's probability of automation as a function of the occupation's required bottleneck abilities.

- We merge Webb's (2020) occupation risk scores with data from the IPUMS Current Population Survey (CPS). We primarily use the CPS data to calculate state-level employment shares of different occupations. Unfortunately, these datasets do not have a joint occupation-level identifier. We therefore use a crosswalk file to first convert the occupation codes in the CPS to the occupation codes used in the American Community Survey (ACS). Using these ACS occupation codes, we can then merge CPS data with Webb*s (2020).

- To analyze automation exposure by education, we also merge in the employment projection dataset by the Bureau of Labor Statistics. This dataset includes SOC job codes, which are also included in the Webb (2020) dataset. However, despite using official job codes and crosswalks, some codes were

*e-mail: lucaskitzmueller@hks.harvard.edu

†e-mail:bnagda@mit.edu

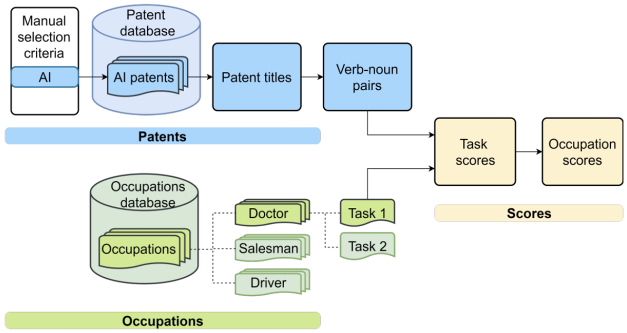

Figure 1: From Webb. et. al. The analytical process that Webb used to arrive at his risk scores.

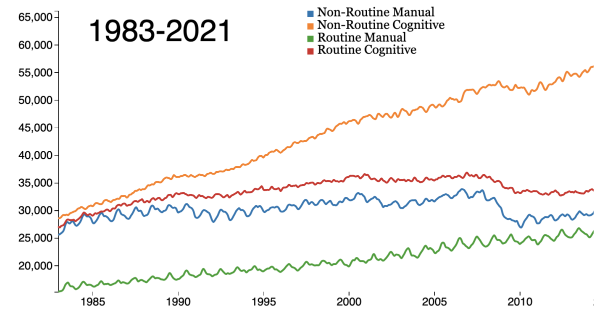

Figure 2: The linechart following the timeline animation. Note that non-routine cognitive tasks have been increasing in volume over time.

inconsistent between data sources, and we had to resolve these errors manually to generate clean data merges.

- To trace historic employment by different occupational groups (cognitive vs. manual, routine vs. non-routine), we use data from the Federal Bank of St. Louis (2016). To generate the data, search in the "Release" view on the FRED home page for the "Employment Situation" report, choose the first link to view the data from the Current Population Survey, and select Table A-13. In the first subsection (Monthly, Employed), select these four series and add them to the graph: Management, professional, and related occupations; Service occupations; Sales and office occupations; and Production, transportation and material moving occupations.

4 Results

4.1 Linechart

We start our data story with a line chart showing the historic trends in employment in different job categories. The linechart is our attempt at depicting net changes in job volumes over time for specific categories: routine manual, non-routine manual, non-routine cognitive, routine cognitive. The largest takeaway that we visually depicted (through the animation and chart formatting) and explicitly highlighted in the accompanying text is that non-routine cognitive tasks were growing in volume in the US- perhaps a good sign for the economy. The linechart depicts, as a timeseries, the labor data, and the date changes alongside the linechart. Inspiration for the chart came from this 2015 Bloomberg chart: **https://www.bloomberg.com/graphics/2015-whats-warming-the-world/**.

4.2 Ridgeline Plot

We then show the distribution of exposure by different educational groups in kernel density plots (also called ridgeline plot or joyplot). The plot shows that robot exposure reinforces historic trends: less-educated workers, typically employed in jobs with a high share of manual routine tasks, are more exposed to disruption from robotics than higher-educated workers. The exposure estimates for AI, however, reveal an opposite pattern. Some lower-skill occupations, such as power plant operators and dispatchers, are still highly exposed. But many highly educated workers will be impacted too, such as clinical laboratory technicians, chemical engineers, and optometrists. In fact, workers with a bachelor's degree appear to be exposed the most.

We arrived at the ridgeline plot after several iterations. Initially, we had designed a treemap chart, where each occupation was scaled by size, arranged by educational group, and colored by technology exposure. Our hope was that this design would give users an immediate sense of what educational groups are more exposed but also allow them to explore the full granularity of our occupation-level data. However, user testing showed that viewers found the treemap chart to be "overwhelming". We therefore decided to visualize a more aggregated (i.e., less granular) version of our data.

The ridgeline plots make it easier for users to directly infer what educational groups are most affected. The interpretation is also aided by more affected groups colored in a darker blue tone – a color scheme that is also repeated in the next graph. In addition, by showing the distribution of exposure, we still show more information than the bar charts previously used to visualize the data. We believe these distributions reveal interesting insights themselves – for example, in the case of AI exposure, we notice that occupations requiring a Master's degree have either very high exposure or very low exposure – only showing the mean exposure could create a wrong impression.

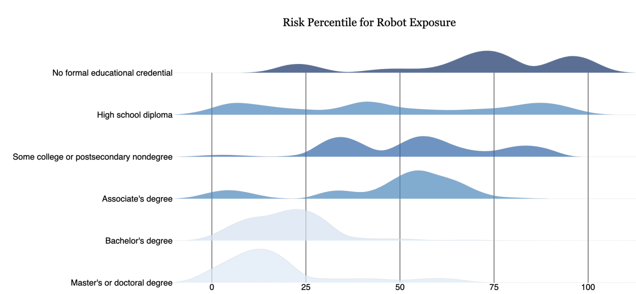

Figure 3: The ridgeline chart depicting robot risk per education level. Note that lower-skilled workers are more exposed than higher-skilled workers.

We were also aware that simply showing the unweighted risk distribution across occupations could be misleading. Some occupations are very small (e.g., according to the Bureau of Labor Statistics, there were only 600 prosthodontists in the US in 2019), while others account for a sizable share of the workforce (e.g., 4.3 million retail salespersons). An analysis of different educational groups' exposure should account for these differences. For doing so, within each educational group, we sampled 1000 occupations with replacement, where the sampling probability was proportional to the size of the

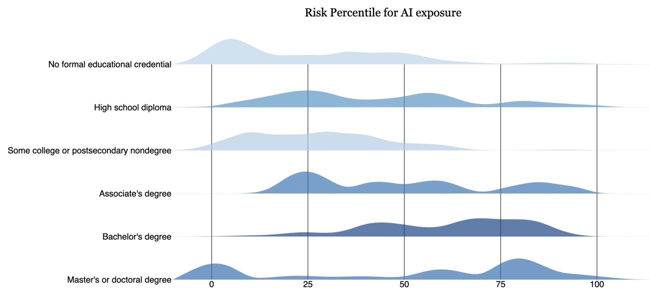

Figure 4: The ridgeline chart depicting AI risk per education level. In contrast to Figure 2, higher-skilled workers are more exposed than lower-skilled workers. This is a key insight we depicted.

occupation. The ridgeline plots show the distribution of Webb's (2020) risk scores across these 1000 sampled occupations.

4.3 Box Map

The tile grid map shows the geographic distribution of exposure to automation technologies. The reason we are interested in the distribution of automation risk across states is that many labor market policies, such as worker retraining programs, are implemented on the state level. The exposure scores encoded to the color channel are weighted averages, where an occupation's weight is the number of workers in 2019 in the state in 2019. The visualization therefore takes into account the occupational composition of the workforce in each state.

The map shows that states involved in manufacturing, such as Wisconsin or Kentucky, are heavily impacted by robotics. However, as AI is increasingly linked to production, the states involved in manufacturing appear to be also disproportionately exposed to AI. At the same time, states with a large high-tech sector and managerial workforce, e.g., Washington DC, Boston, and Washington state have high exposure to AI technologies.

We decided to show these data in a tile grid map to not distract users with the geographic size of the state in their visual perception of automation risk. We also added a tooltip to each state, showing the largest, most at-risk occupations in each state. The color-coding was chosen to be accessible to people with various forms of color-blindness.

4.4 Additional Interactivity

While we do not show statistical confidence intervals in any of our visualizations, we still address the uncertainty of our risk estimates. Unlike most other interactive data visualizations of the Frey Osborne (2013) data, we explain how Webb (2020) derived the occupation-level risk estimates. We believe it is important for users to understand the basic assumption on which this approach rests so that they can assess the validity of the estimates themselves.

To guide users through our visualizations, we implemented scrollytelling. The approach could be described as a repeated "martini glass" technique. As users scroll through the website, key insights from each graph are highlighted and connected to the overall story. However, each graph also has interactive elements itself, so users can stop at each graph and explore the data themselves (for example, by clicking the buttons that display the different risk estimates).

5 Discussion

Citing anecdotal evidence from our peer review, we identify that the audience: (1) finds the broader topic of automation risk

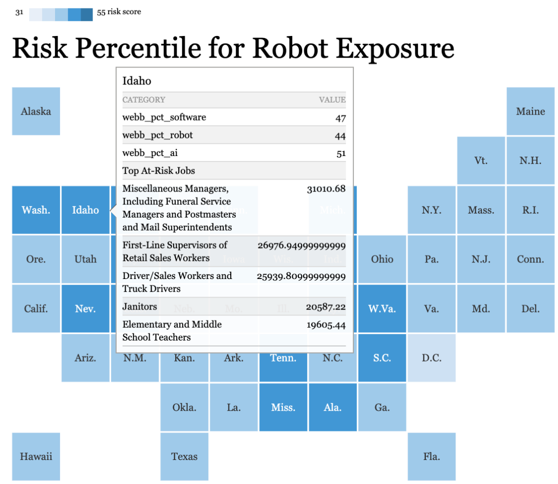

Figure 5: The map depicting robot exposure across geographies, and the associated hover tooltip that provides the top at-risk jobs.

incredibly compelling, (2) appreciates the depiction of labor risk on a per-education or per-geography basis, and (3) is increasingly interested in the methodology and risk-score calculation.

Our audience over the last few weeks- of mostly MIT students- is a deeply analytical, thoughtful, and driven audience. Naturally, our peers are interested in the direct impacts of their work (and for those who are focused on software of any kind, labor displacement tends to be one of them). Many of the insights in the paper- both explicitly mentioned and implicitly gleaned- align with societal expectations: automation tends to disproportionately impact populations that lack formal education (often in lower-skill occupations). Political and economic pundits have indeed been sounding the alarm for the last few years, as job displacement will disproportionately impact lower-skilled, often under served populations. These labor effects, compounded with the outsourcing of manufacturing jobs overseas, have profoundly impacted the American economy and American politics.

Yet the secondary, and perhaps more interesting, conclusion from the data visualization surprised many viewers, as it breaks with historic trends of worker displacement. We find and highlight that while lower-educated workers have a higher exposure to robotics technologies, higher-educated workers have a higher exposure to AI technologies. This spurs many follow-up questions, many of which are actively topics of research in the community: how effective is upskilling in altering economic outcomes for lower-skilled vs. higher-skilled workers? Are there racial disparities (we imagine there are) in the incidence of automation risk? How should municipalities, states, and the federal government prepare for the imminent onslaught of automation?

6 Future Work

As briefly discussed, opportunities for future work abound. For one, we focused our demographic analysis on two main demographics: education and geography. There are opportunities to more granularly explore the data (e.g. county-level analysis as opposed to state-wide). Future work may also focus on other demographic factors, such as age or race.

The data additionally provides us with risk factors for specific jobs. We imagine that workers in certain occupations may choose to transition (or upskill) into other occupations. Future work thus may involve exploring optimal transition paths between jobs based on job similarity. Some work has been done on this topic (see https://royalsocietypublishing.org/doi/10.1098/rsos.182124), but we'd be eager to visually depict and explore patterns in the data.

Finally, the risk values we've depicted are not necessarily time-relevant. That's to say they may change over time, as certain technologies are proven out and others lose steam. One time-dependent proxy for innovation is start-up proliferation and growth. We envision building a dataset, atop on a sampling of startups (perhaps from Crunchbase or LinkedIn), that associates start-ups (based on their industry and mission) with displaced occupations. This may provide a more timely notion of labor risk.

References

[1] Bloomberg - Are you a robot?

[2] Employment Projections.

[3] IPUMS CPS.

[4] Job polarization | FRED Blog.

[5] Michael Webb.

[6] C. B. Frey and M. A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254–280, Jan. 2017.

[7] M. Muro, J. Whiton, and R. Maxim. *WHAT JOBS ARE AFFECTED BY AI?* Brookings.

[8] N. Sonnad. Is your job at risk from robot labor? Check this handy interactive.

[9] M. Webb. The Impact of Artificial Intelligence on the Labor Market. *SSRN Electronic Journal*, 2019.