This document summarizes the replication materials corresponding to the paper “The Evolution of Work in the United States.”

According to the license agreement that we have signed with ProQuest, we are not able to disseminate the raw digitized text that ProQuest has provided to us. Nor are we able to publish a dataset containing ad-level data on jobs’ task content and skill requirements. We instead have posted the data at the most disaggregated level of detail possible: at the job title × year × source combination. Here, source refers to the newspaper posting the ad and whether the ad is a classified or display ad.

Data Sets :

* summary\_by\_year\_bos\_cla.dta : This dataset is constructed using text from Boston Globe classified job ads. For each job title – year combination, the dataset contains information on tasks, skills, and other job characteristics. This dataset contains a subset of the variables described in the Data Library (https://occupationdata.github.io/ ; see in particular the dataset labeled “By job title, source, and year”). The datasets summary\_by\_year\_bos\_dis.dta, summary\_by\_year\_nyt\_cla.dta, summary\_by\_year\_nyt\_dis.dta, summary\_by\_year\_wsj\_cla.dta, and summary\_by\_year\_comp.dta comprise, respectively, of information from Boston Globe display ads, New York Times classified ads, New York Times display ads, Wall Street Journal classified ads, and all of the ads within our dataset (“comp” is short for “composite”).
  + Note: The union of the observations within summary\_by\_year\_bos\_cla.dta, summary\_by\_year\_bos\_dis.dta, summary\_by\_year\_wsj\_cla.dta, summary\_by\_year\_nyt\_cla.dta, and summary\_by\_year\_nyt\_dis.dta equals the observations in the Data Library file labeled “By job title, source, and year”. There are 3,951,950 observations in both sets of files. Also, the number of observations (3,803,219) within summary\_by\_year\_comp.dta equals the number of observations within the Data Library file labeled “By job title and year.”
* boot\_bos\_cla\_title.dta: This dataset is constructed using the text from Boston Globe classified job ads. We resampled from the newspaper job ads from the raw text, clustering ads within each month in a given year. We constructed 40 bootstrapped samples; the final suffix within the variable names give the number of the bootstrap sample. The datasets boot\_bos\_dis.dta, boot\_nyt\_cla.dta, boot\_nyt\_dis.dta, and boot\_comp.dta correspond, respectively, to bootstrapped samples from Boston Globe display ads, New York Times classified ads, New York Times display ads, Wall Street Journal classified ads, and for all ads within our dataset. These datasets are used only to construct the standard errors in Tables 3-5 and 13-20.
* soc\_codes\_corresp.dta: This is a dataset that links job titles to SOC codes. This file is a subset of the file that is in our data library.
* Activities.dta; Knowledge.dta; Skills.dta; Styles.dta: These datasets are taken directly from the O\*NET database (version 22.1). The dataset variables are the SOC code, the O\*NET element code, and the corresponding importance and value scores for the SOC code O\*NET element combination.
* onet\_varname\_crosswalk.csv: This is a correspondence which we have constructed between O\*NET element titles and the variable names according to the datasets we have constructed.
* dot77-8090.dta; dot77-8090-gen.dta; dot91-8090.dta; dot91-8090-gen.dta; oct8497means-gen.dta. These files are taken from the replication materials from Autor, Levy, and Murnane (2003). These files are used to construct our comparisons of our newspaper-based dataset to the Dictionary of Occupational Titles.
* cps\_search.dta, cps\_00029.dta, usa\_00034.dta : These are datasets we have downloaded from IPUMS. In the do files that utilize these datasets, we provide the list of variables, samples, and data quality flags used in downloading these datasets.

Do Files

* read\_ipums\_.do: This do file should be executed first. It takes in data downloaded from IPUMS to construct measures of the number of workers in each occupation, the number of workers in each occupation when restricting to individuals located in the Boston and New York metro areas, and the average education levels by each occupation.
* aej\_calculations.do: This produces the majority of the tables and figures of the paper.
* selection\_appendix.do : This produces the material within Appendix C2:
* alm\_appendix.do: This material is presented within Section 3.3 and Appendix B3.
* centiles-gen-8497-table6.do. This file is taken directly from Autor, Levy, and Murnane’s replication files. We call this do-file within our alm\_appendix do file.