

DOT Project Data Summary

November 6, 2019

This document describes the three data sources available for this project. The goal of this project is to add job complexity information to the 1940 US Census. The available data are all the editions of the Dictionary of Occupational Titles (DOT), the 1940 US Census, and the 1940 Census Alphabetical Index of Jobs.

1 Dictionary of Occupational Titles

The DOT provides job descriptions for thousands of jobs and was meant to be used as a tool for matching job seekers to jobs. The descriptions include information regarding the industry and occupation group the job is in. There are 5 editions of the DOT: 1939, 1949, 1965, 1977, and 1991. Each edition has about 30,000 job titles and 20,000 definitions (one definition can correspond to multiple job titles). Starting with the 1965 edition, the job codes include 3 digits that describe the job complexity with regards to data, people, and things.

Figure 1: Example definition from the 1939 DOT

ACCOUNTING CLERK; accountant clerk (clerical)
1-01.31. Performs the more routine calculating, typing and posting duties necessary in accounting: checks items on various reports, summarizing and posting the data to designated books; performs a variety of other clerical duties, such as making up invoices or monthly statements to be sent to customers, preparing pay rolls, verifying the company bank account, keeping files of various records, making up a periodic report of the company's business activities, and listing and checking detail as instructed. May operate a calculating machine, *ref.* CALCULATING-MACHINE OPERATOR. May operate an adding machine, *ref.* ADDING-MACHINE OPERATOR.

Figure 2: Explanation of Data, People, Things Code

DATA (4th digit)	PEOPLE (5th digit)	THINGS (6th digit)
0 Synthesizing	0 Mentoring	0 Setting-Up
1 Coordinating	1 Negotiating	1 Precision Working
2 Analyzing	2 Instructing	2 Operating-Controlling
3 Compiling	3 Supervising	3 Driving-Operating
4 Computing	4 Diverting	4 Manipulating
5 Copying	5 Persuading	5 Tending
6 Comparing	6 Speaking-Signaling	6 Feeding-Offbearing
7 } No significant relationship	7 Serving	7 Handling
8 } No significant relationship	8 No significant relationship	8 No significant relationship

2 1940 US Census

The United States conducts a census every ten years to collect data about its population. The Census collects a wide range of information about respondents including occupation, income, education, and place of residence. The data is confidential, but becomes publicly available 72 years after Census Day. The data from the 1940 Census became publicly available in April of 2012.

In the 1940 census, every respondent listed their occupation. These occupation names are not standardized and contain many misspellings. In addition to each person's reported occupation, we also have their designated Census occupation code. There are about 400 census occupation codes, and they correspond to coarse categories such as lawyer or manager. The occupation codes and their corresponding categories can be seen here. The census contains around 130 million observations. We instead work with a file that contains the number of observations for every combination of occupation title and census code.

3 1940 Census Alphabetical List of Titles

In every census, the US Census Board creates a list of official job titles and their corresponding census occupation codes and industry. Census employees would use this list to map each respondent's listed title to a census occupation code. This list contains about 12,000 job titles.

4 Current Approach

Our current approach has been to match job titles in the 1940 census to job titles in the 1940 census alphabetical index of titles based on exact matches in the occupation code and exact/inexact matches in the job title. We then plan to match job titles in the DOT to job titles in the alphabetical index. We assume this will be slightly easier since both data sources contain minimal misspellings, as opposed to the 1940 census.