5.7.2 Data Preparation & Documentation

Team 7

December 13th, 2018

INFM 600

5.7.2: Data Preparation & Documentation: 5%

Your data preparation and documentation draft will be part of your final project repository; this is due by Week 10 to help provide early feedback. Take advantage of the opportunity by submitting as complete a document as you can muster! If you are integrating multiple data sources, provide (shorter) background details for each data set.

The first data source is from the Maryland Department of Planning (2018). The dataset from the MDP highlights Maryland’s 24 counties socioeconomic characteristics in 40 categories that include total households, population age, education levels, employment levels, median household incomes, race, and gender. The second data source is from Montgomery County (2018) regarding education and tuition assistance. The dataset, which is updated twice a year, includes a list of classes approved and paid through the County tuition assistance program, the title and description of the course, the school the course will be taken at, the degree this class curriculum is a part of, the Department and title of the requester, the dollar amount reimbursed to the employee.

The third, fourth and fifth data source is from the government site, the Bureau of Labor Statistics(2017/18). For our project, we collected datasets from occupations employment projections, labor force participation, and educational certificates and licenses. These datasets are huge and very encompassing because it was done on a national level by the federal government. They highlight the labor force, by demographic, employment by occupational group, fastest growing occupations, worker characteristics, as well as education and training assignments by detailed occupation. Our final data source comes from the Maryland State Data Center Department of Planning (2017) which dataset highlights occupational categories of the labor force by residence, sex and race/ethnic origin.

The metadata for the data sources ranges from a simple overview of the data described in a sentence to a paragraph summary. Some of the data sources had metadata that explained the frequency the data was being updated, that type of information in the data source and/or limitations of the data. Most of the metadata explains any sources and any specification when reading the dataset. The data sources from the state government had much more metadata than data sources from the federal government which we found to be interesting.

Luckily because our data sources are from government entities, we have not encountered any issues with the data like missing values, unstandardized content, entity matching, etc. From our analysis, the datasets available are cleaned and the data sources address any potential misunderstandings in the metadata. The only remediation that we will be doing on our part is filtering the dataset to highlight the scope of our project. For example, highlighting the data that reflects race, class, education, jurisdiction etc. Because the data we seek might be the only a part and not the sum of the in these very large datasets, we will be doing some small data cleaning to narrow the scope.

Maryland Department of Planning (2018, February 18). *Maryland Counties Socioeconomic Characteristics*. Retrieved from [**https://data.maryland.gov/Demographic/Maryland-Jurisdictions-Socioeconomic-Characteristi/v67y-zu4n**](https://data.maryland.gov/Demographic/Maryland-Jurisdictions-Socioeconomic-Characteristi/v67y-zu4n)

Montgomery County, Maryland (2018, August 15). *Tuition Assistance*. Retrieved from [**https://data.montgomerycountymd.gov/Education/Tuition-Assistance/p7z5-tjrz**](https://data.montgomerycountymd.gov/Education/Tuition-Assistance/p7z5-tjrz)

Bureau of Labor Statistics (2018, April 11). Employment Projections - *Occupations with the most job growth*. Retrieved from [**https://www.bls.gov/emp/tables/occupations-most-job-growth.htm**](https://www.bls.gov/emp/tables/occupations-most-job-growth.htm)

Bureau of Labor Statistics (2017, October 24). *Civilian labor force participation rate, by age, sex, race, and ethnicity*. Retrieved from [**https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm**](https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm)

Bureau of Labor Statistics (2018, February 9). *Labor Force Statistics from the Current Poplucation Survey - Certification and Licensing Status*. Retrieved from [**https://www.bls.gov/cps/tables.htm#certs\_licenses**](https://www.bls.gov/cps/tables.htm#certs_licenses)

Maryland State Data Center Department of Planning (2017, June 8). *American Community Survey 2006-2010 EEO Detailed Report - Persons in Civilian Labor Force by Residence, Sex and Race/Ethnic Origin*. Retrieved from [**https://planning.maryland.gov/MSDC/Pages/american\_community\_survey/2006-2010/eeo/eeo\_acs0610.aspx**](https://planning.maryland.gov/MSDC/Pages/american_community_survey/2006-2010/eeo/eeo_acs0610.aspx)

1. [**https://data.maryland.gov/Demographic/Maryland-Jurisdictions-Socioeconomic-Characteristi/v67y-zu4n**](https://data.maryland.gov/Demographic/Maryland-Jurisdictions-Socioeconomic-Characteristi/v67y-zu4n)
   1. Maryland Department of Planning. (2018). *Maryland Counties Socioeconomic Characteristics* [Data file and code book]. Available from Maryland GovernmentWebsite
2. <https://data.montgomerycountymd.gov/Education/Tuition-Assistance/p7z5-tjrz>
   1. This dataset includes a list of classes approved and paid through the County tuition assistance program, including the title and description of the course, the school the course will be taken at, the degree this class curriculum is a part of, the Department and title of the requester, the dollar amount reimbursed to the employee. Update Frequency: Twice a Year
3. <https://www.bls.gov/emp/tables/occupations-most-job-growth.htm>
4. <https://www.bls.gov/emp/tables/civilian-labor-force-participation-rate.htm>
5. <https://www.bls.gov/cps/tables.htm#certs_licenses>
   1. **CERTIFICATION AND LICENSING STATUS**
      1. 49. Certification and licensing status of the civilian noninstitutional population 16 years and over by employment status ([HTML](https://www.bls.gov/cps/cpsaat49.htm)) ([PDF](https://www.bls.gov/cps/cpsaat49.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat49.xlsx))
      2. 50. Employment status of the civilian noninstitutional population by certification and licensing status and selected characteristics ([HTML](https://www.bls.gov/cps/cpsaat50.htm)) ([PDF](https://www.bls.gov/cps/cpsaat50.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat50.xlsx))
      3. 51. Certification and licensing status of employed persons 16 years and over by selected characteristics ([HTML](https://www.bls.gov/cps/cpsaat51.htm)) ([PDF](https://www.bls.gov/cps/cpsaat51.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat51.xlsx))
      4. 52. Certification and licensing status of the employed by industry and class of worker ([HTML](https://www.bls.gov/cps/cpsaat52.htm)) ([PDF](https://www.bls.gov/cps/cpsaat52.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat52.xlsx))
      5. 53. Certification and licensing status of the employed by occupation ([HTML](https://www.bls.gov/cps/cpsaat53.htm)) ([PDF](https://www.bls.gov/cps/cpsaat53.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat53.xlsx))
      6. 54. Median weekly earnings of full-time wage and salary workers by certification and licensing status and selected characteristics ([HTML](https://www.bls.gov/cps/cpsaat54.htm)) ([PDF](https://www.bls.gov/cps/cpsaat54.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat54.xlsx))
      7. 55. Median weekly earnings of full-time wage and salary workers by certification and licensing status and occupation ([HTML](https://www.bls.gov/cps/cpsaat55.htm)) ([PDF](https://www.bls.gov/cps/cpsaat55.pdf)) ([XLSX](https://www.bls.gov/cps/cpsaat55.xlsx))
6. [**https://planning.maryland.gov/MSDC/Pages/american\_community\_survey/2006-2010/eeo/eeo\_acs0610.aspx**](https://planning.maryland.gov/MSDC/Pages/american_community_survey/2006-2010/eeo/eeo_acs0610.aspx)
   1. **American Community Survey 2006-2010 EEO Detailed Report - Persons in Civilian Labor Force by Residence, Sex and Race/Ethnic Origin:**

**Word Count: 432**