Webscraping Project- Eagles95

Huma Meer

### Data Source:

The data used in this paper is [County Business Patterns Data](https://www.census.gov/programs-surveys/cbp.html) extracted from the Census Bureau website using an API and a census key. CBP is an annual series that provides subnational economic data by industry. This series includes the number of establishments, employment during the week of March 12 2020, first quarter payroll, and annual payroll. Since the data is collected at the county level, all variables used in our analysis are also presented at the county level for the year 2020.

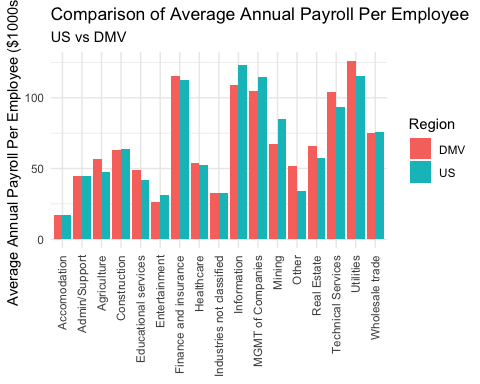
### Data Exploration and Visualization:

For our analysis, we worked on identifying the patterns in the business industries in the US as compared to the DMV metro area. We focused our analysis on the top paying industries by Average Annual Payroll per Employee (AP). We then further analyzed the nature of the business environment in the DMV area that would contribute to the disparities in the AP including which industries had the most establishments and which were the top employers. Additionally, we compared the number of establishments in the DMV area with a few other surrounding states over the years of 2012 to 2020.

## [1] "2af9f600486ef6ad342ed1e8a978c0956a70a52c"

### Plot 1: Average Annual Payroll Per Employee by Industry for the DMV area vs the US

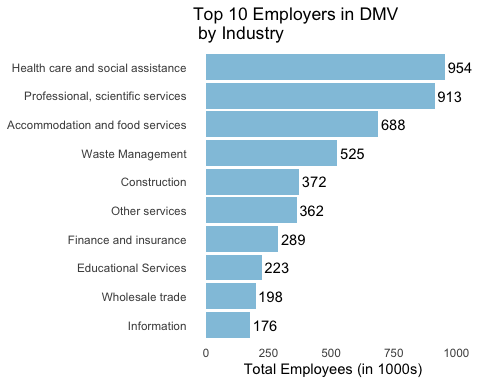
## # A tibble: 6 × 4  
## # Groups: NAICS2017 [3]  
## NAICS2017 NAICS2017\_LABEL Region annual\_payroll  
## <chr> <chr> <chr> <dbl>  
## 1 11 Agriculture DMV 56.9  
## 2 11 Agriculture US 47.9  
## 3 21 Mining DMV 67.5  
## 4 21 Mining US 85.3  
## 5 22 Utilities DMV 126.   
## 6 22 Utilities US 116.



##### The plot above highlights that the Finance, Insurance, Information and Utilities industries are the highest paying industries in the DMV area as well as all over the US. Moreover, Healthcare, Accommodation and Entertainment are the lowest paying industries.

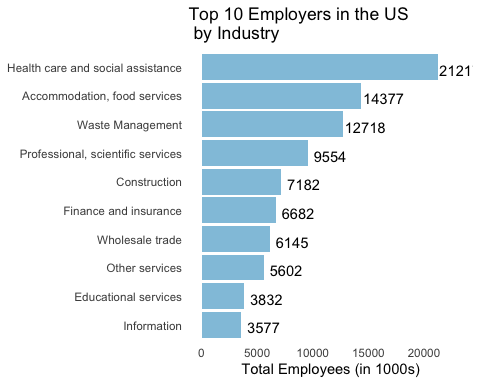
### Plot 2a: Top 10 Industries with the Highest Number of Employees in the DMV

## # A tibble: 6 × 4  
## # Groups: NAICS2017 [6]  
## NAICS2017 NAICS2017\_LABEL sumemp industry  
## <chr> <chr> <dbl> <chr>   
## 1 62 Health care and social assistance 954. Health …  
## 2 54 Professional, scientific, and technical services 913. Profess…  
## 3 72 Accommodation and food services 688. Accommo…  
## 4 56 Administrative and support and waste management and… 525. Waste M…  
## 5 23 Construction 372. Constru…  
## 6 81 Other services (except public administration) 362. Other s…



### Plot 2b: Top 10 Industries with the Highest Number of Employees in the USA

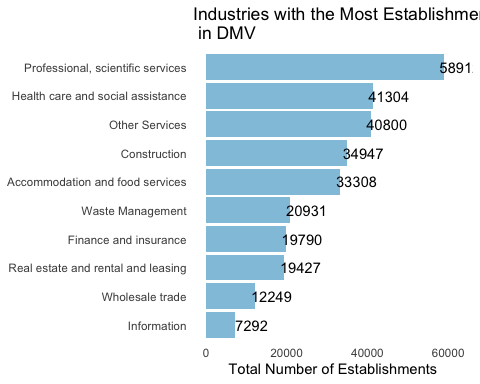
## # A tibble: 6 × 4  
## # Groups: NAICS2017 [6]  
## NAICS2017 NAICS2017\_LABEL sumemp industry  
## <chr> <chr> <dbl> <chr>   
## 1 62 Health care and social assistance 21217. Health …  
## 2 72 Accommodation and food services 14377. Accommo…  
## 3 56 Administrative and support and waste management and… 12718. Waste M…  
## 4 54 Professional, scientific, and technical services 9554. Profess…  
## 5 23 Construction 7182. Constru…  
## 6 52 Finance and insurance 6682. Finance…



##### It is interesting to note that Plot 2a and Plot 2b indicate that Healthcare and Accommodation are industries with the highest number of employees and Plot 1 indicates that these are also the lowest paying industries.

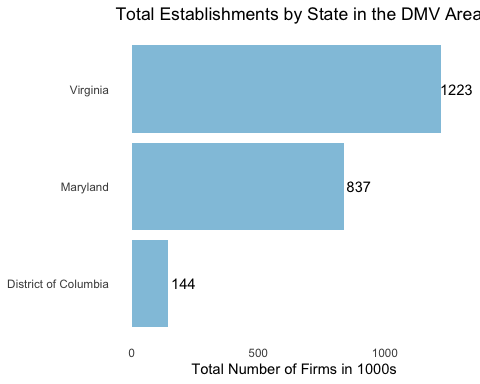
### Plot 3a: Summary Plot for Industries with Highest Number of establishments in DMV

## # A tibble: 6 × 3  
## # Groups: NAICS2017 [6]  
## NAICS2017 NAICS2017\_LABEL sumestab  
## <chr> <chr> <dbl>  
## 1 54 Professional, scientific services 58912  
## 2 62 Health care and social assistance 41304  
## 3 81 Other Services 40800  
## 4 23 Construction 34947  
## 5 72 Accommodation and food services 33308  
## 6 56 Waste Management 20931

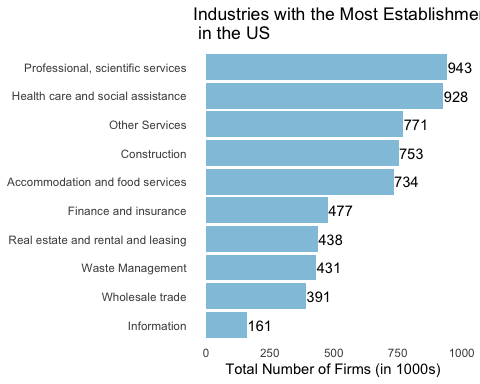


### Plot 3b: Firm Count by State in DMV Area

## # A tibble: 3 × 2  
## # Groups: NAME [3]  
## NAME sumestab  
## <chr> <dbl>  
## 1 Virginia 1223.  
## 2 Maryland 837.  
## 3 District of Columbia 144.



### Plot 3c: Summary Plot for Industries with Highest Number of Establishments in the USA

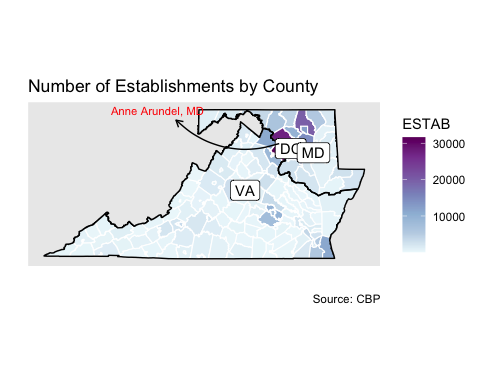


##### Plot 3a and 3c suggest that Professional, technical and Scientific services industry, as well as Healthcare has the highest number of establishments. Additionally, Plot 3b suggests that Virginia has the highest number of establishments as compared to DC and Maryland. This may be due the fact that both Maryland and DC are smaller in terms of geographic area and, therefore, have fewer establishments. This is especially true for DC because DC is a much smaller administrative unit as compared to both VA and MD. To investigate this further we created the following heat map.

### Plot 4: Heat Map of DMV Showing the Number of Establishments by County

## Retrieving data for the year 2020  
## Retrieving data for the year 2020

## Warning in st\_point\_on\_surface.sfc(sf::st\_zm(x)): st\_point\_on\_surface may not  
## give correct results for longitude/latitude data  
  
## Warning in st\_point\_on\_surface.sfc(sf::st\_zm(x)): st\_point\_on\_surface may not  
## give correct results for longitude/latitude data

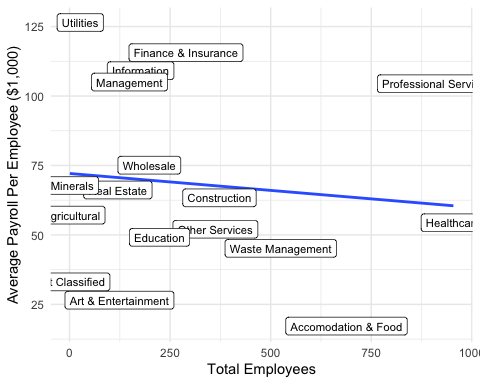


##### Plot 4 indicates that, when analyzing at the county level, there are more establishments per county in Maryland than in Virginia because more counties appear to be darker in shade in MD than in VA. Particularly, the county Anne Arundel in Maryland has the highest number of establishments in the DMV area. This shows that while the total number of establishments in VA are higher than in MD and DC, establishments per county are higher in MD.

### Plot 5: Scatter Plot for Number of Employees and Average Payroll per Employee by Industry in DMV area

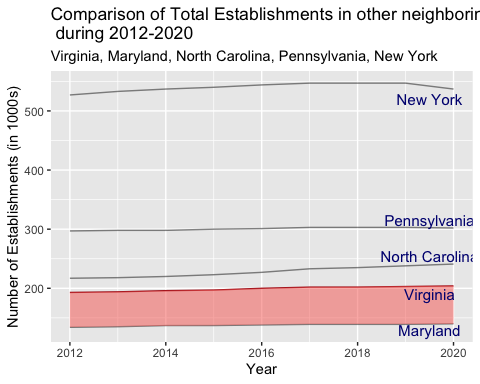
## Warning: Ignoring unknown parameters: check\_overlap

## `geom\_smooth()` using formula 'y ~ x'



##### Plot 5 echoes the conclusion made earlier that Accommodation and Healthcare are the industries with the highest number of employees and the lowest pay.

### Plot 6: Comparison of Virginia and Maryland with Other Neighboring States, in terms of number of Establishments, for the years 2012-2020



##### The plot allows us to compare the standing of the Virginia and Maryland with surrounding states in terms of the number of establishments. It shows that New York, Pennsylvania and North Carolina all perform better as compared to both Maryland and Virginia.

NOTE: Complete codes for this file can be found in the RMD file attached.