Data608 - Final Proposal

Abdelmalek Hajjam

3/21/2020

My Project Proposal

The Goal of the Project:

- Use of visualization to conduct visual analytics on data analysis
- Provide analysis result based on question and answer
- Provide UI to explore the data as well as find insights from it

The Methodology:

- Use shiny app provides efficient ways to manipulate and visualize data
- To explore salary data of 8 professions based on the data about prevailing wage of foreign employers in the United States. The 8 professions include: Data scientist, Software engineers, Data Analyst, Business Analyst, Management Consultant, Assistant Professor, Attorney, Teacher

The Data:

- The dataset is from the United States Department of Labor, Employment & Training Administration (https://www.foreignlaborcert.doleta.gov/performancedata.cfm)
- The final cleaned data for this application contains total 167,278 cases and 17 columns in 2015
- It is about the prevailing wage data of foreign employers seeking to file applications in the Permanent Labor Certification Program
- prevailing wage data of US natives are not included

Cleansing the data

```
library(knitr)
library(dplyr)
library(ggplot2)
library(plotly)
library(sqldf)
library(googleVis)

#load("salary.Rdata")
#head(df)
```

Cleaning, calculating and re-arrange: job, wage and location

- I will be using tidyverse to do arrange and aggregate the datathe data
- I will also be using the ggplot to plot the wage level by state

```
arr <- df[,c(8,5)]
colnames(arr) <- c("region", "WAGE")
arr <- aggregate(arr[, 2], list(arr$region), mean)
colnames(arr) <- c("region", "WAGE")
arr$WAGE<-round(arr$WAGE)
salarylevel <- arr %>% arrange(desc(WAGE))
salarylevel
ordered_states <- rev(salarylevel$region)
salarylevel$region <- factor(salarylevel$region, levels = ordered_states)
ggplot(salarylevel, aes(x = region, y = WAGE)) + geom_bar(stat = "identity", alpha = 0.5, col = "blue")
```

Creating a shiny app to compare the job, wage and location

```
ggplot(df, aes(x=WORK_STATE_ABBREVIATION, y=PAID_WAGE_PER_YEAR, color=JOB_TITLE_SUBGROUP, shape=JOB_TITL
```

I will need the help of the following packages: Reshape2

Ggplot2

Graphics

Ggthemes

GoogleVis

Plotly

Knitr

Dplyr

Plyr

Knitr

and may be more.

The best yet to come - stay tuned for the final project.