

MAT5314 Project 1: Data Visualization

Teng Li(7373086)
Shiya Gao()
Chuhan Yue(300376046)
Yang Lyu(8701121)

Introduction

A data set of the 2016 US election polls was given. In this project we aim to understand the data structure by creating various visualizations.

Method

We use various R packages to present the data set and to plot the graphs.

Result

We first take a look at the raw data set:

```
## state startdate enddate
## 1 U.S. 2016-11-03 2016-11-06
## 2 U.S. 2016-11-01 2016-11-07
## 3 U.S. 2016-11-02 2016-11-06
## 4 U.S. 2016-11-04 2016-11-07
## 5 U.S. 2016-11-03 2016-11-06
## 6 U.S. 2016-11-03 2016-11-06
##
## pollster grade samplesize
## 1 ABC News/Washington Post A+ 2220
## 2 Google Consumer Surveys B 26574
## 3 Ipsos A- 2195
## 4 YouGov B 3677
## 5 Gravis Marketing B- 16639
## 6 Fox News/Anderson Robbins Research/Shaw & Company Research A 1295
## population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1 lv 47.00 43.00 4.00 NA
## 2 lv 38.03 35.69 5.46 NA
## 3 lv 42.00 39.00 6.00 NA
## 4 lv 45.00 41.00 5.00 NA
## 5 rv 47.00 43.00 3.00 NA
## 6 lv 48.00 44.00 3.00 NA
## adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1 45.20163 41.72430 4.626221 NA
## 2 43.34557 41.21439 5.175792 NA
## 3 42.02638 38.81620 6.844734 NA
## 4 45.65676 40.92004 6.069454 NA
## 5 46.84089 42.33184 3.726098 NA
```

```
## 6          49.02208          43.95631          3.057876          NA
```

As we can see, there are a few variables with missing values:

```
##      state      startdate      enddate      pollster
## Length:4208    Length:4208    Length:4208    Length:4208
## Class :character Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character Mode  :character
##
##
##
##      grade      samplesize      population      rawpoll_clinton
## Length:4208    Min.      : 35.0    Length:4208    Min.      :11.04
## Class :character 1st Qu.: 447.5    Class :character 1st Qu.:38.00
## Mode  :character Median : 772.0    Mode  :character Median :43.00
##                      Mean  : 1148.2    Mean  :41.99
##                      3rd Qu.: 1236.5    3rd Qu.:46.20
##                      Max.   :84292.0    Max.   :88.00
##                      NA's    :1
## rawpoll_trump rawpoll_johnson rawpoll_mcmullin adjpoll_clinton
## Min.      : 4.00    Min.      : 0.000    Min.      : 9.0     Min.      :17.06
## 1st Qu.:35.00    1st Qu.: 5.400    1st Qu.:22.5     1st Qu.:40.21
## Median :40.00    Median : 7.000    Median :25.0     Median :44.15
## Mean   :39.83    Mean  : 7.382    Mean  :24.0     Mean  :43.32
## 3rd Qu.:45.00    3rd Qu.: 9.000    3rd Qu.:27.9     3rd Qu.:46.92
## Max.   :68.00    Max.   :25.000    Max.   :31.0     Max.   :86.77
##                      NA's    :1409    NA's    :4178
## adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## Min.      : 4.373    Min.      :-3.668    Min.      :11.03
## 1st Qu.:38.429    1st Qu.: 3.145    1st Qu.:23.11
## Median :42.765    Median : 4.384    Median :25.14
## Mean   :42.674    Mean  : 4.660    Mean  :24.51
## 3rd Qu.:46.290    3rd Qu.: 5.756    3rd Qu.:27.98
## Max.   :72.433    Max.   :20.367    Max.   :31.57
##                      NA's    :1409    NA's    :4178
```

Discussion

Conclusion