

MAT5314 Project 1: Data Visualization

Teng Li(7373086)
Shiya Gao(300381032)
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 create a data variable definition table to give an initial understanding of the data. As we can see, there are a few variables with missing values:

Table 1: Data Variable Definition

Variables	Size	Type	Number.Unique	Number.Missing	Comment
state	4208	character	57	0	The name of the state where the election is held, "U.S." means national polls
startdate	4208	character	352	0	Start date of poll
enddate	4208	character	345	0	End date of poll
pollster	4208	character	196	0	Organization name that conducts or analyzes opinion polls
grade	4208	character	11	429	Grade assigned by Fivethirtyeight (sometimes as 538, an American website that focuses on opinion poll) to pollster
samplesize	4208	integer	1767	1	Sample size of polls for each pollster
population	4208	character	4	0	Type of population being polled
rawpoll_clinton	4208	numeric	1312	0	Percentage for Hillary Clinton
rawpoll_trump	4208	numeric	1385	0	Percentage for Donald Trump
rawpoll_johnson	4208	numeric	585	1409	Percentage for Gary Johnson
rawpoll_mcmullin	4208	numeric	17	4178	Percentage for Evan McMullin
adpoll_clinton	4208	numeric	4200	0	Fivethirtyeight adjusted percentage for Hillary Clinton
adpoll_trump	4208	numeric	4204	0	Fivethirtyeight adjusted percentage for Donald Trump
adpoll_johnson	4208	numeric	2211	1409	Fivethirtyeight adjusted percentage for Gary Johnson
adpoll_mcmullin	4208	numeric	31	4178	Fivethirtyeight adjusted percentage for Evan McMullin

Discussion

Conclusion