

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 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
```

We first explore the column name of raw data.

```
## [1] "state"      "startdate"      "enddate"      "pollster"
## [5] "grade"      "samplesize"      "population"      "rawpoll_clinton"
## [9] "rawpoll_trump"      "rawpoll_johnson"      "rawpoll_mcmullin"      "adjpoll_clinton"
## [13] "adjpoll_trump"      "adjpoll_johnson"      "adjpoll_mcmullin"
```

The following is the values in state:

```
## [1] "U.S."      "New Mexico"      "Virginia"
## [4] "Iowa"      "Wisconsin"      "North Carolina"
## [7] "Georgia"      "Florida"      "Oregon"
```

## [10]	"Ohio"	"South Carolina"	"New York"
## [13]	"Michigan"	"Pennsylvania"	"Missouri"
## [16]	"New Hampshire"	"Arizona"	"Nevada"
## [19]	"Colorado"	"California"	"Washington"
## [22]	"Texas"	"Utah"	"Illinois"
## [25]	"Indiana"	"Tennessee"	"Connecticut"
## [28]	"Massachusetts"	"New Jersey"	"Kansas"
## [31]	"Kentucky"	"Minnesota"	"Oklahoma"
## [34]	"Maryland"	"Alabama"	"Nebraska"
## [37]	"Louisiana"	"Maine"	"Arkansas"
## [40]	"Alaska"	"Vermont"	"Idaho"
## [43]	"Mississippi"	"West Virginia"	"South Dakota"
## [46]	"Montana"	"Hawaii"	"Maine CD-1"
## [49]	"Maine CD-2"	"Rhode Island"	"Nebraska CD-3"
## [52]	"Nebraska CD-1"	"Delaware"	"North Dakota"
## [55]	"District of Columbia"	"Nebraska CD-2"	"Wyoming"

The following is the values in pollster:

##	[1]	"ABC News/Washington Post"
##	[2]	"Google Consumer Surveys"
##	[3]	"Ipsos"
##	[4]	"YouGov"
##	[5]	"Gravis Marketing"
##	[6]	"Fox News/Anderson Robbins Research/Shaw & Company Research"
##	[7]	"CBS News/New York Times"
##	[8]	"NBC News/Wall Street Journal"
##	[9]	"Zia Poll"
##	[10]	"IBD/TIPP"
##	[11]	"Selzer & Company"
##	[12]	"Angus Reid Global"
##	[13]	"Monmouth University"
##	[14]	"Public Policy Polling"
##	[15]	"Marist College"
##	[16]	"The Times-Picayune/Lucid"
##	[17]	"Marquette University"
##	[18]	"Siena College"
##	[19]	"Landmark Communications"
##	[20]	"Quinnipiac University"
##	[21]	"USC Dornsife/LA Times"
##	[22]	"RKM Research and Communications, Inc."
##	[23]	"Remington"
##	[24]	"Christopher Newport University"
##	[25]	"Emerson College"
##	[26]	"CVOTER International"
##	[27]	"Morning Consult"
##	[28]	"SurveyMonkey"
##	[29]	"Clarity Campaign Labs"
##	[30]	"University of New Hampshire"
##	[31]	"Pew Research Center"
##	[32]	"SurveyUSA"
##	[33]	"Rasmussen Reports/Pulse Opinion Research"
##	[34]	"Greenberg Quinlan Rosner/American Viewpoint"
##	[35]	"Saguaro Strategies"

[36] "Hampton University"
 ## [37] "Insights West"
 ## [38] "CNN/Opinion Research Corp."
 ## [39] "Trafalgar Group"
 ## [40] "RAND (American Life Panel)"
 ## [41] "RABA Research"
 ## [42] "GfK Group"
 ## [43] "EPIC-MRA"
 ## [44] "BK Strategies"
 ## [45] "Public Policy Institute of California"
 ## [46] "Columbus Dispatch"
 ## [47] "Keating Research, Inc."
 ## [48] "Opinion Savvy/InsiderAdvantage"
 ## [49] "Research & Polling, Inc."
 ## [50] "University of Massachusetts Lowell"
 ## [51] "TargetSmart/William & Mary"
 ## [52] "Public Opinion Strategies"
 ## [53] "Susquehanna Polling & Research, Inc."
 ## [54] "Stockton College"
 ## [55] "Winthrop University"
 ## [56] "Franklin & Marshall College"
 ## [57] "Dixie Strategies"
 ## [58] "University of North Florida"
 ## [59] "Data Orbital"
 ## [60] "Cygnal Political"
 ## [61] "Suffolk University"
 ## [62] "MassINC Polling Group"
 ## [63] "Muhlenberg College"
 ## [64] "Political Marketing International, Inc./Red Racing Horses"
 ## [65] "Starboard Communications"
 ## [66] "Harper Polling"
 ## [67] "Mitchell Research & Communications"
 ## [68] "Elon University"
 ## [69] "Strategic National"
 ## [70] "Time/Abt SRBI"
 ## [71] "Loras College"
 ## [72] "American Research Group"
 ## [73] "Roanoke College"
 ## [74] "Mason-Dixon Polling & Research, Inc."
 ## [75] "Greenberg Quinlan Rosner (Democracy Corps)"
 ## [76] "University of Denver"
 ## [77] "Y2 Analytics"
 ## [78] "Saint Leo University"
 ## [79] "DFM Research"
 ## [80] "Victory Research"
 ## [81] "Magellan Strategies"
 ## [82] "University of Arkansas"
 ## [83] "JMC Enterprises"
 ## [84] "Bendixen & Amadi International"
 ## [85] "Davis, Hibbitts & Midghall, Inc."
 ## [86] "Associated Industries of Florida"
 ## [87] "Western Kentucky University"
 ## [88] "Crosswind Media & Public Relations"
 ## [89] "Dan Jones & Associates"

[90] "Western New England University"
 ## [91] "Elway Research"
 ## [92] "Princeton Survey Research Associates International"
 ## [93] "Maine People's Resource Center"
 ## [94] "St. Norbert College"
 ## [95] "Braun Research"
 ## [96] "University of Houston"
 ## [97] "Centre College"
 ## [98] "Public Religion Research Institute"
 ## [99] "National Research, Inc."
 ## [100] "SoonerPoll.com"
 ## [101] "Praecones Analytica"
 ## [102] "Nielson Brothers Polling"
 ## [103] "George Washington University (Battleground)"
 ## [104] "North Star Opinion Research/Ayres, McHenry & Associates"
 ## [105] "Southern Media & Opinion Research"
 ## [106] "Craciun Research"
 ## [107] "Fort Hays State University"
 ## [108] "University of Colorado"
 ## [109] "Florida Atlantic University"
 ## [110] "Marketing Resource Group (MRG)"
 ## [111] "Hendrix College"
 ## [112] "Baldwin Wallace University"
 ## [113] "University of New Orleans"
 ## [114] "McLaughlin & Associates"
 ## [115] "Tarrance Group"
 ## [116] "Arizona State University"
 ## [117] "Global Strategy Group"
 ## [118] "Riley Research Associates"
 ## [119] "Ball State University"
 ## [120] "Lake Research Partners"
 ## [121] "Illinois Public Opinion Strategies"
 ## [122] "University of Wyoming"
 ## [123] "Cherry Communications"
 ## [124] "Glengariff Group, Inc."
 ## [125] "Castleton University"
 ## [126] "Southern Illinois University"
 ## [127] "Clout Research/Wenzel Strategies"
 ## [128] "University of Akron"
 ## [129] "icitizen"
 ## [130] "Sacramento State"
 ## [131] "HighGround"
 ## [132] "Edison Research"
 ## [133] "Michigan State University"
 ## [134] "Garin-Hart-Yang Research Group"
 ## [135] "Fairleigh Dickinson University (PublicMind)"
 ## [136] "Vanderbilt University"
 ## [137] "Moore Information"
 ## [138] "High Point University"
 ## [139] "Market Research Insight"
 ## [140] "Echelon Insights"
 ## [141] "Normington, Petts & Associates"
 ## [142] "Hoffman Research Group"
 ## [143] "Strategies 360"

[144] "MBQF"
 ## [145] "Montana State University Billings"
 ## [146] "Anzalone Liszt Grove Research"
 ## [147] "GBA Strategies"
 ## [148] "Middle Tennessee State University"
 ## [149] "University of Delaware"
 ## [150] "Ivan Moore Research"
 ## [151] "Goucher College"
 ## [152] "Meredith College"
 ## [153] "Target Insyght"
 ## [154] "University of Mary Washington"
 ## [155] "Mercyhurst University"
 ## [156] "We Ask America"
 ## [157] "Rutgers University"
 ## [158] "Just Win Strategies"
 ## [159] "Texas Lyceum"
 ## [160] "First Tuesday Strategies"
 ## [161] "OpinionWorks"
 ## [162] "Utah Valley University"
 ## [163] "Tulchin Research"
 ## [164] "Feldman Group"
 ## [165] "R.L. Repass & Partners"
 ## [166] "Zogby Interactive/JZ Analytics"
 ## [167] "Strategy Research"
 ## [168] "Vox Populi Communications"
 ## [169] "Expedition Strategies"
 ## [170] "Penn Schoen Berland"
 ## [171] "Integrated Web Strategy"
 ## [172] "Basswood Research"
 ## [173] "Fairbank, Maslin, Maullin, Metz & Associates"
 ## [174] "Field Research Corporation (Field Poll)"
 ## [175] "OnMessage Inc."
 ## [176] "Hickman Analytics"
 ## [177] "American Strategies"
 ## [178] "National Journal"
 ## [179] "Evolving Strategies"
 ## [180] "Leland Beatty"
 ## [181] "TargetPoint"
 ## [182] "Political Callers, Inc."
 ## [183] "Schoen Consulting"
 ## [184] "Data Targeting"
 ## [185] "Bellwether Research & Consulting"
 ## [186] "Cole Hargrave Snodgrass & Associates"
 ## [187] "Baruch College"
 ## [188] "Behavior Research Center (Rocky Mountain)"
 ## [189] "Harvard University"
 ## [190] "Dartmouth College"
 ## [191] "Merrill Poll"
 ## [192] "Critical Insights"
 ## [193] "Orion Strategies"
 ## [194] "Robert Morris University"
 ## [195] "Florida Southern College"
 ## [196] "Ogden & Fry"

The following is the values in grade:

```
## [1] "A+" "B" "A-" "B-" "A" NA "B+" "C+" "C-" "C" "D"
```

The following is the values in population:

```
## [1] "lv" "rv" "a" "v"
```

```
##
```

```
##      a      lv      rv      v
```

```
##    21 3727  418   42
```

Data dictionary:

* state: The name of the state where the election is held, "U.S." means national polls

* startdate: Start data of poll

* enddate: End data of poll

* pollster: Organization name that conducts or analyzes opinion polls

* grade: Grade assigned by Fivethirtyeight (sometimes as 538, an American website that focuses on opinion poll) to pollster

> "A letter grade from A+ to F that reflects the accuracy of a polling organization's polls and its predictive plus-minus score (a projection of how well we think a pollster's polls will do in the future). Pollsters with a small number of polls get a provisional rating (for example, A/B) rather than a precise letter grade" (Cite from <https://projects.fivethirtyeight.com/pollster-ratings/>).

* samplesize: Sample size of polls from each pollster

* population: Type of population being polled

1.LV (*likely Voters*): population whose most likely to participate in 2016 election.

2.RV (*Registered Voters*): RV includes all eligible and registered voters, whether or not they end up voting.

3.A (*Adults*): Refers to all persons of legal voting age.

4.V (*Voters*): Persons who has voted in an election.

rawpoll_clinton: Percentage for Hillary Clinton

* rawpoll_trump: Percentage for Donald Trump

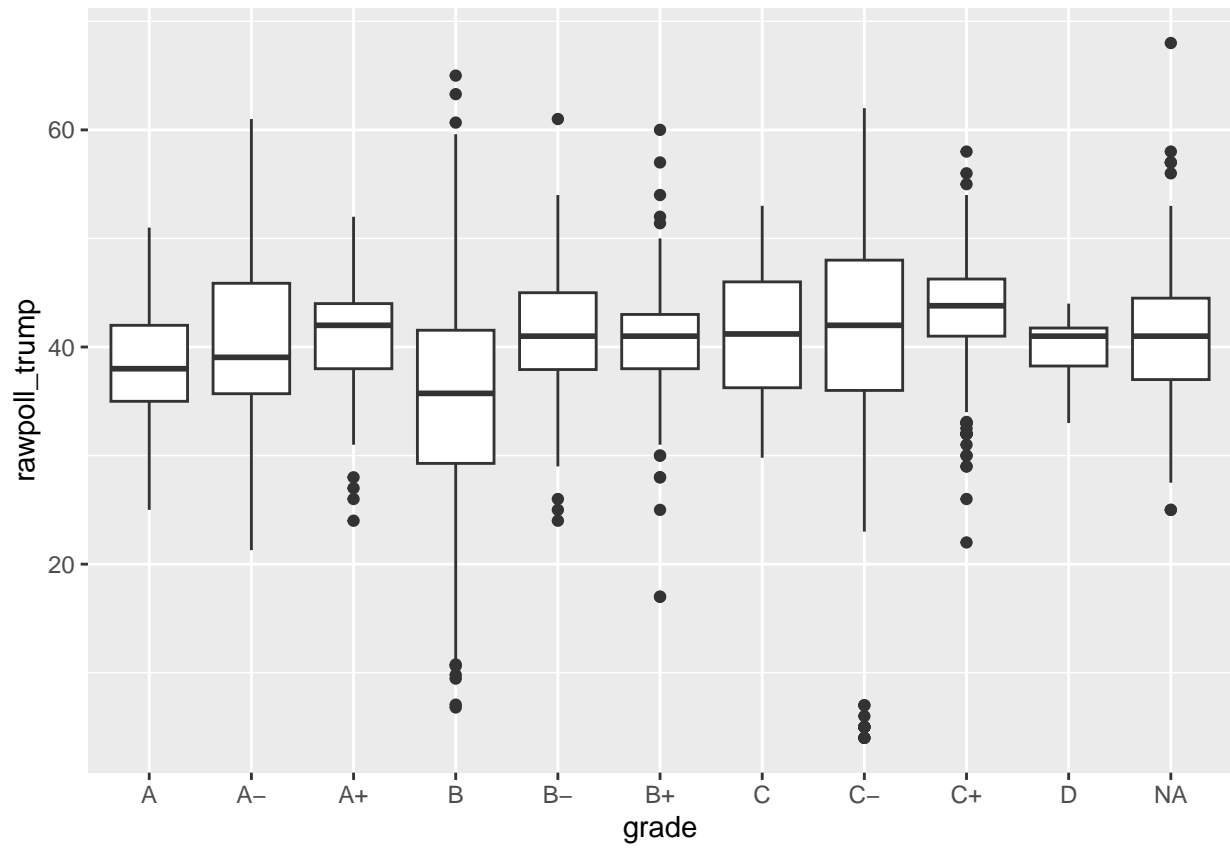
* rawpoll_johnson: Percentage for Gary Johnson * rawpoll_mcmullin: Percentage for Evan McMullin *

adjpoll_clinton: Fivethirtyeight adjusted percentage for Hillary Clinton

* adjpoll_trump: Fivethirtyeight adjusted percentage for Donald Trump

* adjpoll_johnson: Fivethirtyeight adjusted percentage for Gary Johnson

* adjpoll_mcmullin: Fivethirtyeight adjusted percentage for Evan McMullin



Discussion

Conclusion