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


##              state  startdate    enddate                   pollster grade samplesize
## 4203   Washington 2016-10-21 2016-11-02                      Ipsos    A-        538
## 4204      Virginia 2016-09-16 2016-09-22                      Ipsos    A-        452
## 4205     Wisconsin 2016-08-04 2016-08-07    Marquette University     A        683
## 4206          Utah 2016-11-01 2016-11-07 Google Consumer Surveys     B        286
## 4207        Oregon 2016-10-21 2016-11-02                      Ipsos    A-        446
## 4208      Michigan 2016-01-23 2016-01-26                   EPIC-MRA    A-        600
##      population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 4203         lv           46.71         38.33              NA               NA
## 4204         lv           46.54         40.04              NA               NA
## 4205         lv           47.00         34.00            9.00               NA
## 4206         lv           21.33         35.05            9.99               NA
## 4207         lv           46.46         37.41              NA               NA
## 4208         lv           43.00         41.00              NA               NA
##      adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 4203        45.56387      38.22545              NA               NA
## 4204        46.47852      40.48017              NA               NA
## 4205        48.74781      39.07778        4.705020               NA
## 4206        26.65200      40.57738        9.705791               NA
## 4207        45.12949      37.10720              NA               NA
## 4208        42.14966      42.05508              NA               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"
```

Count the missings in state

```
## [1] 0
```

The following is the values in startdate:

```
##    [1] "2015-11-06" "2015-11-07" "2015-11-09" "2015-11-10" "2015-11-11"
##    [6] "2015-11-12" "2015-11-13" "2015-11-15" "2015-11-16" "2015-11-23"
##   [11] "2015-11-27" "2015-11-29" "2015-11-30" "2015-12-02" "2015-12-03"
##   [16] "2015-12-04" "2015-12-05" "2015-12-06" "2015-12-07" "2015-12-08"
##   [21] "2015-12-10" "2015-12-11" "2015-12-13" "2015-12-14" "2015-12-15"
##   [26] "2015-12-16" "2015-12-17" "2015-12-18" "2015-12-19" "2015-12-20"
##   [31] "2015-12-21" "2015-12-22" "2015-12-23" "2015-12-24" "2015-12-25"
##   [36] "2015-12-26" "2015-12-27" "2015-12-28" "2015-12-29" "2015-12-30"
```

```
##  [41] "2015-12-31" "2016-01-01" "2016-01-02" "2016-01-03" "2016-01-04"
##  [46] "2016-01-05" "2016-01-06" "2016-01-07" "2016-01-08" "2016-01-09"
##  [51] "2016-01-10" "2016-01-11" "2016-01-12" "2016-01-13" "2016-01-14"
##  [56] "2016-01-15" "2016-01-16" "2016-01-17" "2016-01-18" "2016-01-19"
##  [61] "2016-01-20" "2016-01-21" "2016-01-22" "2016-01-23" "2016-01-24"
##  [66] "2016-01-25" "2016-01-26" "2016-01-27" "2016-01-28" "2016-01-29"
##  [71] "2016-01-30" "2016-01-31" "2016-02-01" "2016-02-02" "2016-02-03"
##  [76] "2016-02-04" "2016-02-05" "2016-02-06" "2016-02-07" "2016-02-08"
##  [81] "2016-02-09" "2016-02-10" "2016-02-11" "2016-02-12" "2016-02-13"
##  [86] "2016-02-14" "2016-02-15" "2016-02-16" "2016-02-17" "2016-02-18"
##  [91] "2016-02-19" "2016-02-20" "2016-02-21" "2016-02-22" "2016-02-23"
##  [96] "2016-02-24" "2016-02-25" "2016-02-26" "2016-02-27" "2016-02-28"
## [101] "2016-02-29" "2016-03-01" "2016-03-02" "2016-03-03" "2016-03-04"
## [106] "2016-03-05" "2016-03-06" "2016-03-07" "2016-03-08" "2016-03-09"
## [111] "2016-03-10" "2016-03-11" "2016-03-12" "2016-03-13" "2016-03-14"
## [116] "2016-03-15" "2016-03-16" "2016-03-17" "2016-03-18" "2016-03-19"
## [121] "2016-03-20" "2016-03-21" "2016-03-22" "2016-03-23" "2016-03-24"
## [126] "2016-03-25" "2016-03-26" "2016-03-27" "2016-03-28" "2016-03-29"
## [131] "2016-03-30" "2016-03-31" "2016-04-01" "2016-04-02" "2016-04-03"
## [136] "2016-04-04" "2016-04-05" "2016-04-06" "2016-04-07" "2016-04-08"
## [141] "2016-04-09" "2016-04-10" "2016-04-11" "2016-04-12" "2016-04-13"
## [146] "2016-04-14" "2016-04-15" "2016-04-16" "2016-04-17" "2016-04-18"
## [151] "2016-04-19" "2016-04-20" "2016-04-21" "2016-04-22" "2016-04-23"
## [156] "2016-04-24" "2016-04-25" "2016-04-26" "2016-04-27" "2016-04-28"
## [161] "2016-04-29" "2016-04-30" "2016-05-01" "2016-05-02" "2016-05-03"
## [166] "2016-05-04" "2016-05-05" "2016-05-06" "2016-05-07" "2016-05-08"
## [171] "2016-05-09" "2016-05-10" "2016-05-11" "2016-05-12" "2016-05-13"
## [176] "2016-05-14" "2016-05-15" "2016-05-16" "2016-05-17" "2016-05-18"
## [181] "2016-05-19" "2016-05-20" "2016-05-21" "2016-05-22" "2016-05-23"
## [186] "2016-05-24" "2016-05-25" "2016-05-26" "2016-05-27" "2016-05-28"
## [191] "2016-05-29" "2016-05-30" "2016-05-31" "2016-06-01" "2016-06-02"
## [196] "2016-06-03" "2016-06-04" "2016-06-05" "2016-06-06" "2016-06-07"
## [201] "2016-06-08" "2016-06-09" "2016-06-10" "2016-06-11" "2016-06-12"
## [206] "2016-06-13" "2016-06-14" "2016-06-15" "2016-06-16" "2016-06-17"
## [211] "2016-06-18" "2016-06-19" "2016-06-20" "2016-06-21" "2016-06-22"
## [216] "2016-06-23" "2016-06-24" "2016-06-25" "2016-06-26" "2016-06-27"
## [221] "2016-06-28" "2016-06-29" "2016-06-30" "2016-07-01" "2016-07-02"
## [226] "2016-07-03" "2016-07-04" "2016-07-05" "2016-07-06" "2016-07-07"
## [231] "2016-07-08" "2016-07-09" "2016-07-10" "2016-07-11" "2016-07-12"
## [236] "2016-07-13" "2016-07-14" "2016-07-15" "2016-07-16" "2016-07-17"
## [241] "2016-07-18" "2016-07-19" "2016-07-20" "2016-07-21" "2016-07-22"
## [246] "2016-07-23" "2016-07-24" "2016-07-25" "2016-07-26" "2016-07-27"
## [251] "2016-07-28" "2016-07-29" "2016-07-30" "2016-07-31" "2016-08-01"
## [256] "2016-08-02" "2016-08-03" "2016-08-04" "2016-08-05" "2016-08-06"
## [261] "2016-08-07" "2016-08-08" "2016-08-09" "2016-08-10" "2016-08-11"
## [266] "2016-08-12" "2016-08-13" "2016-08-14" "2016-08-15" "2016-08-16"
## [271] "2016-08-17" "2016-08-18" "2016-08-19" "2016-08-20" "2016-08-21"
## [276] "2016-08-22" "2016-08-23" "2016-08-24" "2016-08-25" "2016-08-26"
## [281] "2016-08-27" "2016-08-28" "2016-08-29" "2016-08-30" "2016-08-31"
## [286] "2016-09-01" "2016-09-02" "2016-09-03" "2016-09-04" "2016-09-05"
## [291] "2016-09-06" "2016-09-07" "2016-09-08" "2016-09-09" "2016-09-10"
## [296] "2016-09-11" "2016-09-12" "2016-09-13" "2016-09-14" "2016-09-15"
## [301] "2016-09-16" "2016-09-17" "2016-09-18" "2016-09-19" "2016-09-20"
## [306] "2016-09-21" "2016-09-22" "2016-09-23" "2016-09-24" "2016-09-25"
```

```
## [311] "2016-09-26" "2016-09-27" "2016-09-28" "2016-09-29" "2016-09-30"
## [316] "2016-10-01" "2016-10-02" "2016-10-03" "2016-10-04" "2016-10-05"
## [321] "2016-10-06" "2016-10-07" "2016-10-08" "2016-10-09" "2016-10-10"
## [326] "2016-10-11" "2016-10-12" "2016-10-13" "2016-10-14" "2016-10-15"
## [331] "2016-10-16" "2016-10-17" "2016-10-18" "2016-10-19" "2016-10-20"
## [336] "2016-10-21" "2016-10-22" "2016-10-23" "2016-10-24" "2016-10-25"
## [341] "2016-10-26" "2016-10-27" "2016-10-28" "2016-10-29" "2016-10-30"
## [346] "2016-10-31" "2016-11-01" "2016-11-02" "2016-11-03" "2016-11-04"
## [351] "2016-11-05" "2016-11-06"
```

Count the missings in startdate

```
## [1] 0
```

The following is the values in enddate:

```
##   [1] "2015-11-08" "2015-11-13" "2015-11-15" "2015-11-16" "2015-11-17"
##   [6] "2015-11-19" "2015-11-23" "2015-11-30" "2015-12-01" "2015-12-02"
##  [11] "2015-12-03" "2015-12-06" "2015-12-07" "2015-12-08" "2015-12-09"
##  [16] "2015-12-13" "2015-12-14" "2015-12-15" "2015-12-17" "2015-12-18"
##  [21] "2015-12-19" "2015-12-20" "2015-12-21" "2015-12-22" "2015-12-23"
##  [26] "2015-12-24" "2015-12-25" "2015-12-26" "2015-12-27" "2015-12-28"
##  [31] "2015-12-29" "2015-12-30" "2015-12-31" "2016-01-01" "2016-01-02"
##  [36] "2016-01-03" "2016-01-04" "2016-01-05" "2016-01-06" "2016-01-07"
##  [41] "2016-01-08" "2016-01-09" "2016-01-10" "2016-01-11" "2016-01-12"
##  [46] "2016-01-13" "2016-01-14" "2016-01-15" "2016-01-16" "2016-01-17"
##  [51] "2016-01-18" "2016-01-19" "2016-01-20" "2016-01-21" "2016-01-22"
##  [56] "2016-01-23" "2016-01-24" "2016-01-25" "2016-01-26" "2016-01-27"
##  [61] "2016-01-28" "2016-01-29" "2016-01-30" "2016-01-31" "2016-02-01"
##  [66] "2016-02-02" "2016-02-03" "2016-02-04" "2016-02-05" "2016-02-06"
##  [71] "2016-02-07" "2016-02-08" "2016-02-09" "2016-02-10" "2016-02-11"
##  [76] "2016-02-12" "2016-02-13" "2016-02-14" "2016-02-15" "2016-02-16"
##  [81] "2016-02-17" "2016-02-18" "2016-02-19" "2016-02-20" "2016-02-21"
##  [86] "2016-02-22" "2016-02-23" "2016-02-24" "2016-02-25" "2016-02-26"
##  [91] "2016-02-27" "2016-02-28" "2016-02-29" "2016-03-01" "2016-03-02"
##  [96] "2016-03-03" "2016-03-04" "2016-03-05" "2016-03-06" "2016-03-07"
## [101] "2016-03-08" "2016-03-09" "2016-03-10" "2016-03-11" "2016-03-12"
## [106] "2016-03-13" "2016-03-14" "2016-03-15" "2016-03-16" "2016-03-17"
## [111] "2016-03-18" "2016-03-19" "2016-03-20" "2016-03-21" "2016-03-22"
## [116] "2016-03-23" "2016-03-24" "2016-03-25" "2016-03-26" "2016-03-27"
## [121] "2016-03-28" "2016-03-29" "2016-03-30" "2016-03-31" "2016-04-01"
## [126] "2016-04-02" "2016-04-03" "2016-04-04" "2016-04-05" "2016-04-06"
## [131] "2016-04-07" "2016-04-08" "2016-04-09" "2016-04-10" "2016-04-11"
## [136] "2016-04-12" "2016-04-13" "2016-04-14" "2016-04-15" "2016-04-16"
## [141] "2016-04-17" "2016-04-18" "2016-04-19" "2016-04-20" "2016-04-21"
## [146] "2016-04-22" "2016-04-23" "2016-04-24" "2016-04-25" "2016-04-26"
## [151] "2016-04-27" "2016-04-28" "2016-04-29" "2016-04-30" "2016-05-01"
## [156] "2016-05-02" "2016-05-03" "2016-05-04" "2016-05-05" "2016-05-06"
## [161] "2016-05-07" "2016-05-08" "2016-05-09" "2016-05-10" "2016-05-11"
## [166] "2016-05-12" "2016-05-13" "2016-05-14" "2016-05-15" "2016-05-16"
## [171] "2016-05-17" "2016-05-18" "2016-05-19" "2016-05-20" "2016-05-21"
## [176] "2016-05-22" "2016-05-23" "2016-05-24" "2016-05-25" "2016-05-26"
## [181] "2016-05-27" "2016-05-28" "2016-05-29" "2016-05-30" "2016-05-31"
```

```
## [186] "2016-06-01" "2016-06-02" "2016-06-03" "2016-06-04" "2016-06-05"
## [191] "2016-06-06" "2016-06-07" "2016-06-08" "2016-06-09" "2016-06-10"
## [196] "2016-06-11" "2016-06-12" "2016-06-13" "2016-06-14" "2016-06-15"
## [201] "2016-06-16" "2016-06-17" "2016-06-18" "2016-06-19" "2016-06-20"
## [206] "2016-06-21" "2016-06-22" "2016-06-23" "2016-06-24" "2016-06-25"
## [211] "2016-06-26" "2016-06-27" "2016-06-28" "2016-06-29" "2016-06-30"
## [216] "2016-07-01" "2016-07-02" "2016-07-03" "2016-07-04" "2016-07-05"
## [221] "2016-07-06" "2016-07-07" "2016-07-08" "2016-07-09" "2016-07-10"
## [226] "2016-07-11" "2016-07-12" "2016-07-13" "2016-07-14" "2016-07-15"
## [231] "2016-07-16" "2016-07-17" "2016-07-18" "2016-07-19" "2016-07-20"
## [236] "2016-07-21" "2016-07-22" "2016-07-23" "2016-07-24" "2016-07-25"
## [241] "2016-07-26" "2016-07-27" "2016-07-28" "2016-07-29" "2016-07-30"
## [246] "2016-07-31" "2016-08-01" "2016-08-02" "2016-08-03" "2016-08-04"
## [251] "2016-08-05" "2016-08-06" "2016-08-07" "2016-08-08" "2016-08-09"
## [256] "2016-08-10" "2016-08-11" "2016-08-12" "2016-08-13" "2016-08-14"
## [261] "2016-08-15" "2016-08-16" "2016-08-17" "2016-08-18" "2016-08-19"
## [266] "2016-08-20" "2016-08-21" "2016-08-22" "2016-08-23" "2016-08-24"
## [271] "2016-08-25" "2016-08-26" "2016-08-27" "2016-08-28" "2016-08-29"
## [276] "2016-08-30" "2016-08-31" "2016-09-01" "2016-09-02" "2016-09-03"
## [281] "2016-09-04" "2016-09-05" "2016-09-06" "2016-09-07" "2016-09-08"
## [286] "2016-09-09" "2016-09-10" "2016-09-11" "2016-09-12" "2016-09-13"
## [291] "2016-09-14" "2016-09-15" "2016-09-16" "2016-09-17" "2016-09-18"
## [296] "2016-09-19" "2016-09-20" "2016-09-21" "2016-09-22" "2016-09-23"
## [301] "2016-09-24" "2016-09-25" "2016-09-26" "2016-09-27" "2016-09-28"
## [306] "2016-09-29" "2016-09-30" "2016-10-01" "2016-10-02" "2016-10-03"
## [311] "2016-10-04" "2016-10-05" "2016-10-06" "2016-10-07" "2016-10-08"
## [316] "2016-10-09" "2016-10-10" "2016-10-11" "2016-10-12" "2016-10-13"
## [321] "2016-10-14" "2016-10-15" "2016-10-16" "2016-10-17" "2016-10-18"
## [326] "2016-10-19" "2016-10-20" "2016-10-21" "2016-10-22" "2016-10-23"
## [331] "2016-10-24" "2016-10-25" "2016-10-26" "2016-10-27" "2016-10-28"
## [336] "2016-10-29" "2016-10-30" "2016-10-31" "2016-11-01" "2016-11-02"
## [341] "2016-11-03" "2016-11-04" "2016-11-05" "2016-11-06" "2016-11-07"
```

Count the missings in enddate

```
## [1] 0
```

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 & Amandi 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"
```

Count the missings in pollster

```
## [1] 0
```

The following is the values in grade:

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

Count the missings in grade

```
## [1] 429
```

The following is the values in population:

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

```
##
##    a   lv   rv    v
##   21 3727  418   42
```

Count the missings in population

```
## [1] 0
```

The following is the number of unique and size in rawpoll_clinton:

```
## [1] 1312
```

```
## [1] 4208
```

Count the missings in rawpoll_clinton

[1] 0

The following is the number of unique and size in rawpoll_trump:

[1] 1385

[1] 4208

Count the missings in rawpoll_trump

[1] 0

The following is the number of unique and size in rawpoll_johnson:

[1] 585

[1] 4208

Count the missings in rawpoll_johnson

[1] 1409

The following is the number of unique and size in rawpoll_mcmullin:

[1] 17

[1] 4208

Count the missings in rawpoll_mcmullin

[1] 4178

The following is the number of unique and size in adjpoll_clinton:

[1] 4200

[1] 4208

Count the missings in adjpoll_clinton

[1] 0

The following is the number of unique and size in adjpoll_trump:

[1] 4204

[1] 4208

Count the missings in adjpoll_trump

[1] 0

The following is the number of unique and size in adjpoll_Johnson:

[1] 2211

[1] 4208

Count the missings in adjpoll_Johnson

[1] 1409

The following is the number of unique and size in adjpoll_mcmullin:

[1] 31

[1] 4208

Count the missings in adjpoll_mcmullin

[1] 4178

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 for 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
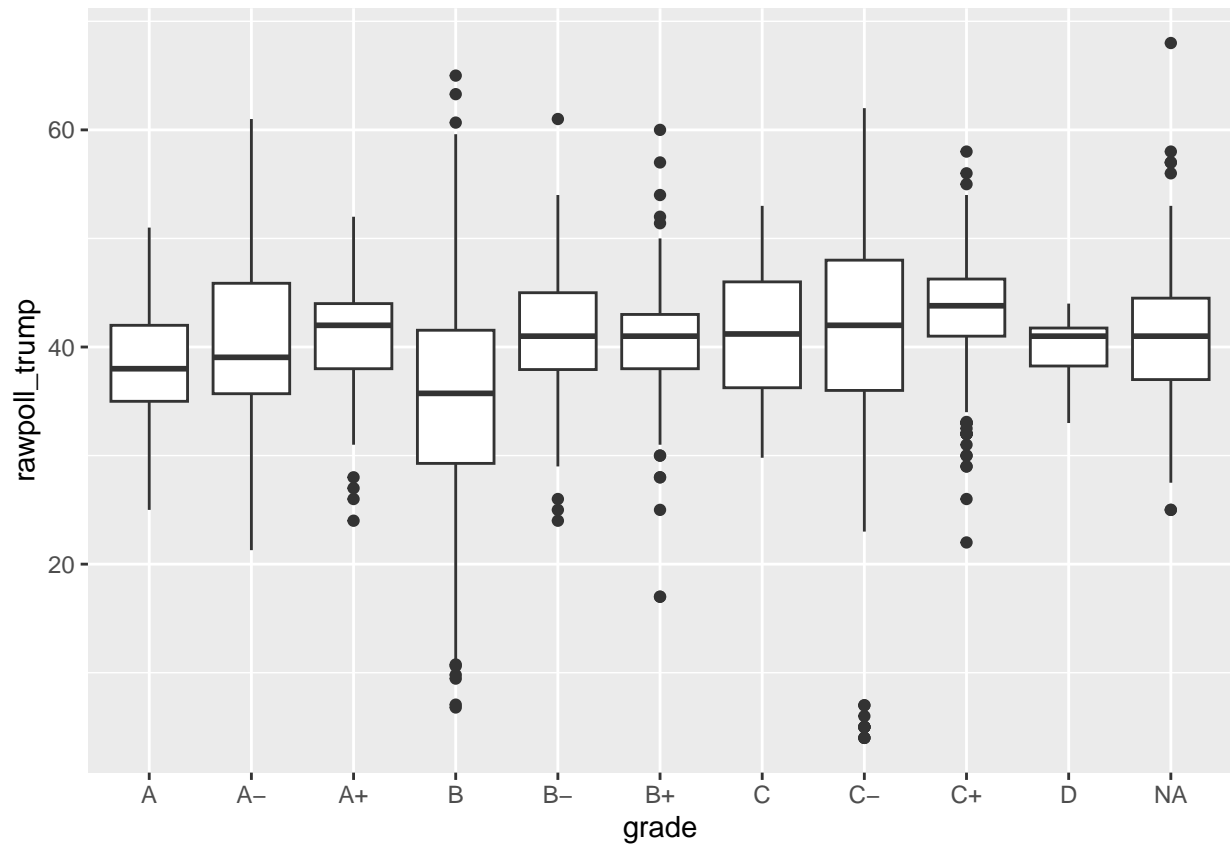* rawpoll_trump: Percentage for Donald Trump
* rawpoll_johnson: Percentage for Gary Johnson * rawpoll_mcmullin: Percentage for Evan Mcmullin * adjpoll_clinton: Fivethirtyeight adjusted percentage for Hillary Clinton
* adjpoll_trump: Fivethirtyeight adjusted percentage for Donald Trump
* adjpoll_johnson: Fivethirtyeight adjusted percentage for Gary Johnson
* adjpoll_mcmullin: Fivethirtyeight adjusted percentage for Evan Mcmullin

Discussion

Conclusion