ProblemSet2Skills

Vera Jónsdóttir

4/12/2021

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
library(dslabs)
```

```
## Warning: package 'dslabs' was built under R version 4.0.5
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## Warning: package 'tibble' was built under R version 4.0.4
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
## Warning: package 'readr' was built under R version 4.0.4
```

```
## Warning: package 'purrr' was built under R version 4.0.4
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
## Warning: package 'stringr' was built under R version 4.0.4
```

```
## Warning: package 'forcats' was built under R version 4.0.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
```

1 Git Concepts

1 List 4 benefits git is a software for distributed version control. List 4 benefits of distributed version control -good way to cooperate with others especially when many are working on code or projects -good way to be able to reinstate old versions -good way to back up work -good way to keep track of big projects and to make sure everybody is using the same version

i. What is the remote repository for this homework? The remote repository for this homework the common class repository which we use to exchange any changes to our R Studio files. The remote repository is stored in GitHub Desktop. It is called datasci-harris/git-github-basics(studentname). Each student has their own private repository.

ii. How do you add a file to staging in github? You need to press Staged and then it will automatically become staged. Then you should press Commit and then push for it to go fully into github.

iii. How do you commit an issue to the local repository? First you need to make sure you start a new project which has version control.Then you make sure that the file/project is connected to your repository before you start editing the file. You find the link for the file at GitHub Desktop. Then you go to File -> New Prokect -> Version Control -> Git and you paste the link. Then each time you need to go to Git -> staged -> commit. Then you also need to remember to select push to push the file to github.

iv. How does github desktop decide what part of your code to show in the main part of the window? Github chooses the most up to date code to show at each time. It chooses the code I upload when I select Commit and Push.

v. What branch are you on right now? Currently I have only one branch for this Problem Set which is the main branch.

vi. If you were to click on "current branch", type a name and click the "New Branch" button, you would create a new branch

a. What would happen to the files in your working directory? They may move between branches and there will be a new version of my code in the data. You must be careful to be in the correct branch at each time.
b. what would happens in the remote repo? There would be a new branch and it may be that the remote repository will now be able to find it.
c. what changes, if anything? There is an extra branch and the data could start floating around between two locations if we're not careful
d. why would you want to work on a different branch? Perhaps it will help organize different projects and problem sets or keep two different versions of the data. it is good to use branches to safely experiment with projects with out destroying the original project

2.1 Debugging mindset 2.1.1

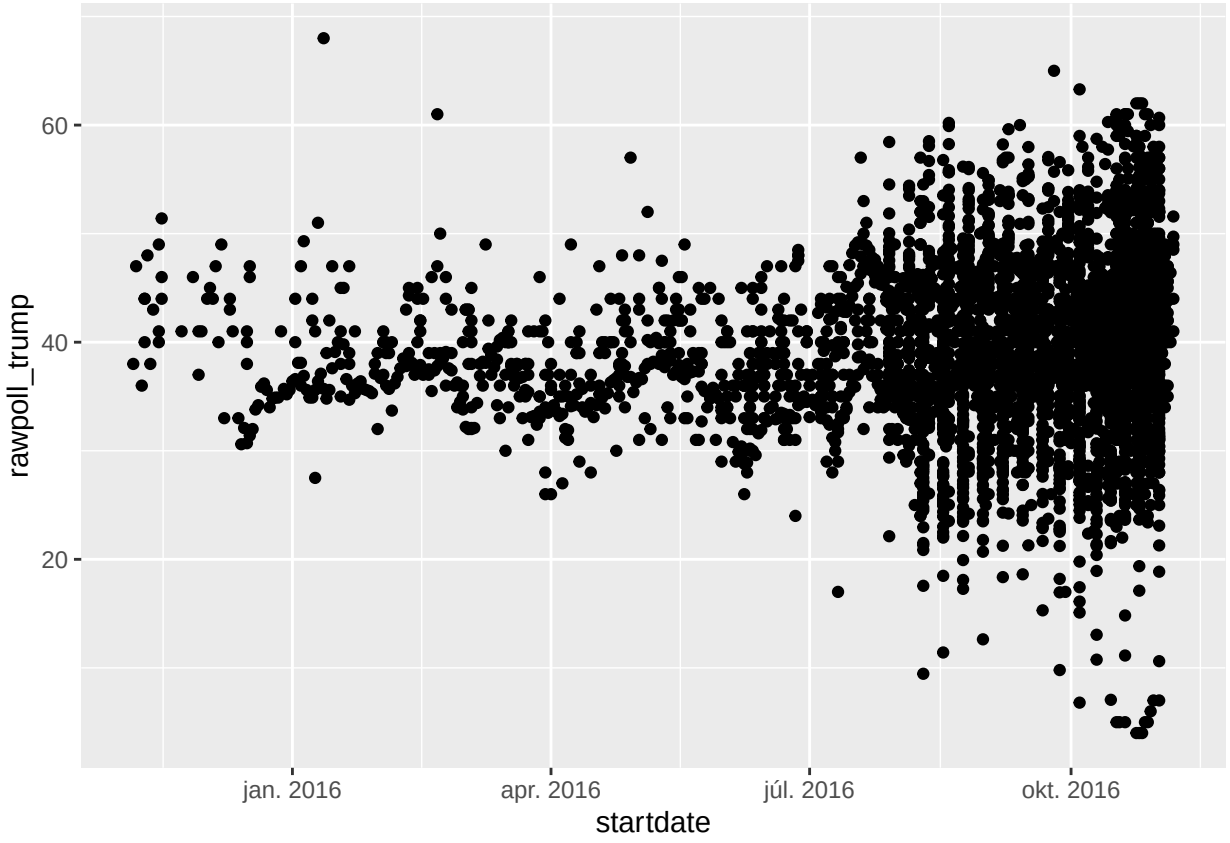
```
my_variable <- 10
my_varıable
```

```
## [1] 10
```

This code runs smoothly in my laptop. However, there may be an issue with running the second variable as it is not written in the same way as the first line of code.

2.1.2 Fix the following code so it works

```
View(polls_us_election_2016)
library(dslabs)
ggplot(data = polls_us_election_2016) +
  geom_point(mapping = aes(x = startdate,
                           y = rawpoll_trump))
```




```
filter(polls_us_election_2016, state == "Florida") %>% head()
```

```
##     state  startdate    enddate                  pollster grade samplesize
## 1 Florida 2016-11-03 2016-11-06      Quinnipiac University    A-        884
## 2 Florida 2016-11-01 2016-11-02                 Remington  <NA>       2352
## 3 Florida 2016-11-02 2016-11-04                    YouGov     B       1188
## 4 Florida 2016-10-20 2016-10-24                  SurveyUSA     A       1251
## 5 Florida 2016-11-01 2016-11-07               SurveyMonkey    C-       4092
## 6 Florida 2016-10-27 2016-11-01 CNN/Opinion Research Corp.    A-        773
##   population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1         lv              46            45               2               NA
## 2         lv              45            48              NA               NA
## 3         rv              45            45              NA               NA
## 4         lv              48            45               2               NA
## 5         lv              47            45               4               NA
## 6         lv              49            47               3               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        46.44315      43.93999        2.098310               NA
## 2        44.85722      46.49677              NA               NA
```

```
## 3        47.07455      46.99468             NA               NA
## 4        46.74555      45.86589        1.520730               NA
## 5        45.59190      44.32744        1.692430               NA
## 6        48.35252      45.23579        2.469063               NA
```

```
filter(polls_us_election_2016, as.numeric(grade) > 3) %>% head()
```

```
##   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
```

2.1.3.Press Alt (Option) + Shift + K. What happens? How can you get to the same place using the menus?

```
#Alt (Option) + Shift + K shows us a preview of Keyboard Shortcut Quick References. In order to do this
```

2.2 Filter 2.2.1. Using the polls_us_election_2016data frame in the ds labs package find the following 2.2.1.1. Polls for the states of Hawaii and Alaska

```
filter(polls_us_election_2016, state == "Alaska" | state == "Hawaii") %>% head(10)
```

```
##     state  startdate    enddate                pollster grade samplesize
## 1  Alaska 2016-11-03 2016-11-06        Gravis Marketing    B-        617
## 2  Hawaii 2016-11-01 2016-11-07            SurveyMonkey    C-        426
## 3  Alaska 2016-11-01 2016-11-07            SurveyMonkey    C-        409
## 4  Alaska 2016-10-25 2016-10-27 Google Consumer Surveys     B        446
## 5  Alaska 2016-10-21 2016-10-26        Craciun Research  <NA>        400
## 6  Alaska 2016-10-11 2016-10-13  Lake Research Partners    B+        500
```

```
## 7  Alaska 2016-10-05 2016-10-06       Moore Information     B        500
## 8  Hawaii 2016-10-30 2016-11-06            SurveyMonkey    C-        426
## 9  Alaska 2016-10-30 2016-11-06            SurveyMonkey    C-        382
## 10 Hawaii 2016-10-04 2016-11-06                  YouGov     B        289
##    population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1          rv            41.0          44.0             3.0               NA
## 2          lv            52.0          28.0             9.0               NA
## 3          lv            31.0          48.0            12.0               NA
## 4          lv            38.0          39.0            11.0               NA
## 5          lv            47.0          43.0             7.0               NA
## 6          lv            36.0          37.0             7.0               NA
## 7          lv            34.0          37.0            10.0               NA
## 8          lv            52.0          29.0             9.0               NA
## 9          lv            31.0          47.0            13.0               NA
## 10         lv            50.3          27.9             4.1               NA
##    adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1         40.84795      43.33498        3.726098               NA
## 2         50.61398      27.33769        6.692430               NA
## 3         29.60705      47.33447        9.692430               NA
## 4         43.43333      46.13872        9.828628               NA
## 5         44.75680      44.77065        6.616653               NA
## 6         36.93134      42.37487        5.751367               NA
## 7         36.22323      41.11790        8.260216               NA
## 8         50.68395      28.39450        6.677361               NA
## 9         29.67449      46.38923       10.677360               NA
## 10        51.16296      30.58176        3.811562               NA
```

2.2.1.2

```
filter(polls_us_election_2016, samplesize > 500)%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                  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
## 7                                  CBS News/New York Times    A-       1426
## 8                               NBC News/Wall Street Journal    A-       1282
## 9                                                 Zia Poll  <NA>       8439
## 10                                                IBD/TIPP    A-       1107
```

```
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786               NA
## 8         43.58576      40.77325        5.365788               NA
## 9         44.82594      41.59978        7.870127               NA
## 10        42.92745      42.23545        6.316175               NA
```

2.2.1.3

```
pdd <- filter(polls_us_election_2016, pollster =="YouGov"|pollster =="Google Consumer Surveys"|pollster
head(pdd)
```

```
##          state  startdate    enddate                pollster grade samplesize
## 1         U.S. 2016-11-01 2016-11-07 Google Consumer Surveys     B      26574
## 2         U.S. 2016-11-04 2016-11-07                  YouGov     B       3677
## 3         Ohio 2016-11-02 2016-11-04                  YouGov     B       1189
## 4      Georgia 2016-11-03 2016-11-05                  YouGov     B        995
## 5 Pennsylvania 2016-11-03 2016-11-05                  YouGov     B        931
## 6      Florida 2016-11-02 2016-11-04                  YouGov     B       1188
##   population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1         lv           38.03         35.69            5.46               NA
## 2         lv           45.00         41.00            5.00               NA
## 3         lv           45.00         46.00              NA               NA
## 4         lv           43.00         49.00            4.00               NA
## 5         lv           45.00         43.00            4.00               NA
## 6         rv           45.00         45.00              NA               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        43.34557      41.21439        5.175792               NA
## 2        45.65676      40.92004        6.069454               NA
## 3        44.93624      45.09646              NA               NA
## 4        43.80799      48.99024        5.169494               NA
## 5        45.82043      42.99602        5.169494               NA
## 6        47.07455      46.99468              NA               NA
```

2.2.1.4

```
filter(polls_us_election_2016, as.numeric(grade) == 10 & rawpoll_trump<30)
```

```
##        state  startdate    enddate                                pollster
## 1   Maryland 2016-09-27 2016-09-30              ABC News/Washington Post
## 2 Washington 2016-08-09 2016-08-13                         Elway Research
## 3 California 2016-06-08 2016-07-02 Field Research Corporation (Field Poll)
## 4   Maryland 2016-03-30 2016-04-03              ABC News/Washington Post
##   grade samplesize population rawpoll_clinton rawpoll_trump rawpoll_johnson
## 1    A+        706         lv              63            27               4
## 2    A+        350         lv              45            24              NA
## 3    A+        495         lv              50            26              10
## 4    A+        752         rv              63            28              NA
##   rawpoll_mcmullin adjpoll_clinton adjpoll_trump adjpoll_johnson
## 1               NA        62.70862      28.73903        1.469119
## 2               NA        44.68377      27.79350              NA
## 3               NA        52.61833      29.79511        5.842896
## 4               NA        60.84333      30.40937              NA
##   adjpoll_mcmullin
## 1               NA
## 2               NA
## 3               NA
## 4               NA
```

2.2.1.5

```
filter(polls_us_election_2016, adjpoll_clinton >=40 & adjpoll_clinton <=60)%>% head()
```

```
##   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
```

2.2.1.6

```
filter(polls_us_election_2016, rawpoll_trump >= rawpoll_clinton + 10)%>% head()
```

```
##      state  startdate    enddate              pollster grade samplesize
## 1 Missouri 2016-10-31 2016-11-01 Public Policy Polling    B+       1083
## 2 Missouri 2016-11-01 2016-11-02 Clarity Campaign Labs     B       1036
## 3 Missouri 2016-10-31 2016-11-01             Remington  <NA>       1722
## 4     Utah 2016-11-03 2016-11-05                YouGov     B        762
## 5     Utah 2016-10-31 2016-11-02       Emerson College     B       1000
## 6     Utah 2016-11-03 2016-11-05       Trafalgar Group     C       1350
##   population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1         lv           37.00         50.00            4.00               NA
## 2         lv           38.00         54.00              NA               NA
## 3         lv           39.00         52.00            4.00               NA
## 4         lv           23.00         40.00            7.00            24.00
## 5         lv           19.60         39.80            3.00            27.60
## 6         lv           29.52         39.95            3.89            24.52
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        36.55124      49.59908        5.028656               NA
## 2        37.65693      52.77196              NA               NA
## 3        39.02343      50.77483        5.124639               NA
## 4        23.83396      40.00230        8.169495         24.00000
## 5        19.75684      38.01440        3.075563         27.70142
## 6        29.32624      37.13456        5.055889         24.52000
```

2.2.1.7

```
filter(polls_us_election_2016, rawpoll_mcmullin > 5)%>% head()
```

```
##   state  startdate    enddate            pollster grade samplesize population
## 1  Utah 2016-11-03 2016-11-05              YouGov     B        762         lv
## 2  Utah 2016-10-31 2016-11-02     Emerson College     B       1000         lv
## 3  Utah 2016-10-30 2016-10-31    Gravis Marketing    B-       1424         rv
## 4  Utah 2016-11-03 2016-11-05     Trafalgar Group     C       1350         lv
## 5  Utah 2016-11-01 2016-11-07        SurveyMonkey    C-       1479         lv
## 6  Utah 2016-10-30 2016-11-02 Monmouth University    A+        402         lv
##   rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1           23.00         40.00            7.00            24.00
## 2           19.60         39.80            3.00            27.60
## 3           29.00         35.00            3.00            24.00
## 4           29.52         39.95            3.89            24.52
## 5           31.00         34.00            7.00            25.00
## 6           31.00         37.00            4.00            24.00
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        23.83396      40.00230        8.169495         24.00000
## 2        19.75684      38.01440        3.075563         27.70142
## 3        29.04086      34.78405        3.422875         24.13522
```

```
## 4        29.32624      37.13456        5.055889         24.52000
## 5        29.59989      33.33115        4.692430         25.00000
## 6        30.06568      36.70382        4.644697         24.10142
```

2.2.2.1

```
filter(polls_us_election_2016, rawpoll_trump > 60 & startdate>="2016-01-01" & startdate<="2016-01-31")
```

```
##     state  startdate    enddate          pollster grade samplesize population
## 1 Alabama 2016-01-12 2016-01-12 Strategy Research  <NA>       2700         rv
##   rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1              32            68              NA               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        30.92318      67.31449              NA               NA
```

This is the poll.

2.2.2.1

```
filter(polls_us_election_2016, is.na(grade))%>% head()
```

```
##          state  startdate    enddate                 pollster grade samplesize
## 1   New Mexico 2016-11-06 2016-11-06                Zia Poll  <NA>       8439
## 2         U.S. 2016-11-05 2016-11-07 The Times-Picayune/Lucid  <NA>       2521
## 3         U.S. 2016-11-01 2016-11-07    USC Dornsife/LA Times  <NA>       2972
## 4     Virginia 2016-11-01 2016-11-02               Remington  <NA>       3076
## 5    Wisconsin 2016-11-01 2016-11-02               Remington  <NA>       2720
## 6 Pennsylvania 2016-11-01 2016-11-02               Remington  <NA>       2683
##   population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1         lv           46.00         44.00               6               NA
## 2         lv           45.00         40.00               5               NA
## 3         lv           43.61         46.84              NA               NA
## 4         lv           46.00         44.00              NA               NA
## 5         lv           49.00         41.00              NA               NA
## 6         lv           46.00         45.00              NA               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        44.82594      41.59978        7.870127               NA
## 2        45.13966      42.26495        3.679914               NA
## 3        45.32156      43.38579              NA               NA
## 4        45.27399      41.91459              NA               NA
## 5        48.22713      38.86464              NA               NA
## 6        45.30896      42.94988              NA               NA
```

you need to change it to is.na. Otherwise R Studio won't understand your function

2.2.2.2

```
filter(polls_us_election_2016, is.na(rawpoll_mcmullin))%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                  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
## 7                                  CBS News/New York Times    A-       1426
## 8                               NBC News/Wall Street Journal    A-       1282
## 9                                                 Zia Poll  <NA>       8439
## 10                                                IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786               NA
## 8         43.58576      40.77325        5.365788               NA
## 9         44.82594      41.59978        7.870127               NA
## 10        42.92745      42.23545        6.316175               NA
```

4,178 rows withh McMullin are n/a. McMullin may not be very popular in the race so they do not keep very accurate polls on his statistics. He is therefore missing from very many polls.

2.2.2.3

```
filter(polls_us_election_2016, as.numeric(grade) == 9)%>% head(10)
```

```
##             state  startdate    enddate
## 1             U.S. 2016-11-03 2016-11-06
## 2             U.S. 2016-11-01 2016-11-03
## 3       Wisconsin 2016-10-26 2016-10-31
## 4  North Carolina 2016-11-04 2016-11-06
```

```
## 5         Florida 2016-10-20 2016-10-24
## 6        New York 2016-11-03 2016-11-04
## 7         Arizona 2016-10-30 2016-11-01
## 8      Washington 2016-10-31 2016-11-02
## 9         Georgia 2016-10-30 2016-11-01
## 10     California 2016-10-28 2016-10-31
##                                                    pollster grade samplesize
## 1  Fox News/Anderson Robbins Research/Shaw & Company Research     A       1295
## 2                                           Marist College     A        940
## 3                                        Marquette University     A       1255
## 4                                          Siena College     A        800
## 5                                              SurveyUSA     A       1251
## 6                                          Siena College     A        617
## 7                                           Marist College     A        719
## 8                                              SurveyUSA     A        681
## 9                                           Marist College     A        707
## 10                                             SurveyUSA     A        747
##    population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1          lv              48            44               3               NA
## 2          lv              44            43               6               NA
## 3          lv              46            40               4               NA
## 4          lv              44            44               3               NA
## 5          lv              48            45               2               NA
## 6          lv              51            34               5               NA
## 7          lv              40            45               9               NA
## 8          lv              50            38               4               NA
## 9          lv              44            45               8               NA
## 10         lv              56            35               4               NA
##    adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1         49.02208      43.95631        3.057876               NA
## 2         42.83406      43.43819        4.780429               NA
## 3         46.10344      40.97982        2.897062               NA
## 4         44.21875      45.08290        2.335250               NA
## 5         46.74555      45.86589        1.520730               NA
## 6         51.30942      35.12664        4.344325               NA
## 7         38.85863      45.72376        7.587354               NA
## 8         48.87802      36.95385        4.691579               NA
## 9         42.81871      45.65459        6.587354               NA
## 10        54.86451      34.27885        4.457467               NA
```

There is no n/a that comes up when we search specifically for a particular letter grade.

2.2.2.4

```
filter(polls_us_election_2016,  is.na(grade) | as.numeric(grade)>0)%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
```

```
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                    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
## 7                                  CBS News/New York Times    A-       1426
## 8                              NBC News/Wall Street Journal    A-       1282
## 9                                                 Zia Poll  <NA>       8439
## 10                                                IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786               NA
## 8         43.58576      40.77325        5.365788               NA
## 9         44.82594      41.59978        7.870127               NA
## 10        42.92745      42.23545        6.316175               NA
```

It leads to printing of all columns with every possible grade and n/a as well.

2.2.2.5

```
filter(polls_us_election_2016,  is.na(grade) | as.numeric(grade) == 0)%>% head(10)
```

```
##            state  startdate    enddate                 pollster grade
## 1     New Mexico 2016-11-06 2016-11-06                Zia Poll  <NA>
## 2           U.S. 2016-11-05 2016-11-07 The Times-Picayune/Lucid  <NA>
## 3           U.S. 2016-11-01 2016-11-07   USC Dornsife/LA Times  <NA>
## 4       Virginia 2016-11-01 2016-11-02               Remington  <NA>
## 5      Wisconsin 2016-11-01 2016-11-02               Remington  <NA>
## 6   Pennsylvania 2016-11-01 2016-11-02               Remington  <NA>
## 7  North Carolina 2016-11-01 2016-11-02               Remington  <NA>
## 8           Ohio 2016-11-01 2016-11-02               Remington  <NA>
## 9        Florida 2016-11-01 2016-11-02               Remington  <NA>
## 10          U.S. 2016-11-04 2016-11-05         Morning Consult  <NA>
```

```
##    samplesize population rawpoll_clinton rawpoll_trump rawpoll_johnson
## 1        8439         lv           46.00         44.00               6
## 2        2521         lv           45.00         40.00               5
## 3        2972         lv           43.61         46.84              NA
## 4        3076         lv           46.00         44.00              NA
## 5        2720         lv           49.00         41.00              NA
## 6        2683         lv           46.00         45.00              NA
## 7        2596         lv           45.00         48.00              NA
## 8        2557         lv           44.00         45.00              NA
## 9        2352         lv           45.00         48.00              NA
## 10       1482         lv           45.00         42.00               8
##    rawpoll_mcmullin adjpoll_clinton adjpoll_trump adjpoll_johnson
## 1                NA        44.82594      41.59978        7.870127
## 2                NA        45.13966      42.26495        3.679914
## 3                NA        45.32156      43.38579              NA
## 4                NA        45.27399      41.91459              NA
## 5                NA        48.22713      38.86464              NA
## 6                NA        45.30896      42.94988              NA
## 7                NA        44.39882      46.04110              NA
## 8                NA        43.03321      42.67245              NA
## 9                NA        44.85722      46.49677              NA
## 10               NA        46.28310      43.27167        6.904304
##    adjpoll_mcmullin
## 1                NA
## 2                NA
## 3                NA
## 4                NA
## 5                NA
## 6                NA
## 7                NA
## 8                NA
## 9                NA
## 10               NA
```

Here we will only get the N/A columns.

2.3.1

```
select(polls_us_election_2016, grade, grade, grade)%>% head(10)
```

```
##    grade
## 1     A+
## 2      B
## 3     A-
## 4      B
## 5     B-
## 6      A
## 7     A-
## 8     A-
## 9   <NA>
## 10    A-
```

Nothing changes if you repeat the same variable multiple times or if you only name the variable once.

2.3.2

```
select(polls_us_election_2016, contains("RAW",ignore.case = TRUE))%>% head(10)
```

```
##    rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1            47.00         43.00            4.00               NA
## 2            38.03         35.69            5.46               NA
## 3            42.00         39.00            6.00               NA
## 4            45.00         41.00            5.00               NA
## 5            47.00         43.00            3.00               NA
## 6            48.00         44.00            3.00               NA
## 7            45.00         41.00            5.00               NA
## 8            44.00         40.00            6.00               NA
## 9            46.00         44.00            6.00               NA
## 10           41.20         42.70            7.10               NA
```

2.3.3

```
select(polls_us_election_2016, contains("clinton") | contains("trump"))%>% head(10)
```

```
##    rawpoll_clinton adjpoll_clinton rawpoll_trump adjpoll_trump
## 1            47.00        45.20163         43.00      41.72430
## 2            38.03        43.34557         35.69      41.21439
## 3            42.00        42.02638         39.00      38.81620
## 4            45.00        45.65676         41.00      40.92004
## 5            47.00        46.84089         43.00      42.33184
## 6            48.00        49.02208         44.00      43.95631
## 7            45.00        45.11649         41.00      40.92722
## 8            44.00        43.58576         40.00      40.77325
## 9            46.00        44.82594         44.00      41.59978
## 10           41.20        42.92745         42.70      42.23545
```

```
select(polls_us_election_2016,rawpoll_clinton:rawpoll_trump | adjpoll_clinton:adjpoll_trump )%>% head(1
```

```
##    rawpoll_clinton rawpoll_trump adjpoll_clinton adjpoll_trump
## 1            47.00         43.00        45.20163      41.72430
## 2            38.03         35.69        43.34557      41.21439
## 3            42.00         39.00        42.02638      38.81620
## 4            45.00         41.00        45.65676      40.92004
## 5            47.00         43.00        46.84089      42.33184
## 6            48.00         44.00        49.02208      43.95631
## 7            45.00         41.00        45.11649      40.92722
## 8            44.00         40.00        43.58576      40.77325
## 9            46.00         44.00        44.82594      41.59978
## 10           41.20         42.70        42.92745      42.23545
```

```
select(polls_us_election_2016,rawpoll_clinton:adjpoll_trump &-rawpoll_johnson & -rawpoll_mcmullin)%>% h
```

```
##    rawpoll_clinton rawpoll_trump adjpoll_clinton adjpoll_trump
## 1            47.00         43.00        45.20163      41.72430
## 2            38.03         35.69        43.34557      41.21439
## 3            42.00         39.00        42.02638      38.81620
```

```
## 4            45.00         41.00        45.65676      40.92004
## 5            47.00         43.00        46.84089      42.33184
## 6            48.00         44.00        49.02208      43.95631
## 7            45.00         41.00        45.11649      40.92722
## 8            44.00         40.00        43.58576      40.77325
## 9            46.00         44.00        44.82594      41.59978
## 10           41.20         42.70        42.92745      42.23545
```

3.1

```
polls_us_election_2016 %>% arrange(desc(rawpoll_clinton))%>%head(1)
```

```
##                  state  startdate    enddate     pollster grade samplesize
## 1 District of Columbia 2016-10-30 2016-11-06 SurveyMonkey    C-        315
##   population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1         lv              88             7               2               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        86.70544      6.406481      -0.3226397               NA
```

3.2

```
polls_us_election_2016 %>% arrange(desc(samplesize))%>%head(5)
```

```
##   state  startdate    enddate     pollster grade samplesize population
## 1  U.S. 2016-10-04 2016-11-06       YouGov     B      84292         lv
## 2  U.S. 2016-10-31 2016-11-06 SurveyMonkey    C-      70194         lv
## 3  U.S. 2016-10-24 2016-10-30 SurveyMonkey    C-      40816         lv
## 4  U.S. 2016-08-29 2016-09-04 SurveyMonkey    C-      32226         rv
## 5  U.S. 2016-10-17 2016-10-23 SurveyMonkey    C-      32225         lv
##   rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1            42.9            39             4.7               NA
## 2            47.0            41             6.0               NA
## 3            47.0            41             6.0               NA
## 4            41.0            37            12.0               NA
## 5            46.0            41             7.0               NA
##   adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1        43.74084      41.38541        4.411561               NA
## 2        45.65592      40.37888        3.677361               NA
## 3        45.69879      41.59350        2.920829               NA
## 4        43.43152      42.04099        4.978575               NA
## 5        44.61705      42.86125        3.103545               NA
```

3.3

```
polls_us_election_2016 %>% arrange(desc(is.na(rawpoll_johnson)))%>% head(10)
```

```
##             state  startdate    enddate              pollster grade samplesize
## 1            U.S. 2016-11-01 2016-11-07 USC Dornsife/LA Times  <NA>       2972
## 2            Ohio 2016-11-02 2016-11-04               YouGov     B       1189
## 3        Virginia 2016-11-01 2016-11-02             Remington  <NA>       3076
## 4  North Carolina 2016-10-31 2016-11-01 Public Policy Polling    B+       1169
```

```
## 5       Wisconsin 2016-11-01 2016-11-02            Remington  <NA>       2720
## 6    Pennsylvania 2016-11-01 2016-11-02            Remington  <NA>       2683
## 7  North Carolina 2016-11-01 2016-11-02            Remington  <NA>       2596
## 8            Ohio 2016-11-01 2016-11-02            Remington  <NA>       2557
## 9            U.S. 2016-10-31 2016-11-06  CVOTER International    C+       1625
## 10        Florida 2016-11-01 2016-11-02            Remington  <NA>       2352
##    population rawpoll_clinton rawpoll_trump rawpoll_johnson rawpoll_mcmullin
## 1          lv           43.61         46.84              NA               NA
## 2          lv           45.00         46.00              NA               NA
## 3          lv           46.00         44.00              NA               NA
## 4          lv           49.00         47.00              NA               NA
## 5          lv           49.00         41.00              NA               NA
## 6          lv           46.00         45.00              NA               NA
## 7          lv           45.00         48.00              NA               NA
## 8          lv           44.00         45.00              NA               NA
## 9          lv           48.91         46.13              NA               NA
## 10         lv           45.00         48.00              NA               NA
##    adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin
## 1         45.32156      43.38579              NA               NA
## 2         44.93624      45.09646              NA               NA
## 3         45.27399      41.91459              NA               NA
## 4         47.90348      45.94813              NA               NA
## 5         48.22713      38.86464              NA               NA
## 6         45.30896      42.94988              NA               NA
## 7         44.39882      46.04110              NA               NA
## 8         43.03321      42.67245              NA               NA
## 9         47.01806      42.04561              NA               NA
## 10        44.85722      46.49677              NA               NA
```

```
polls_us_election_2016 %>% arrange(desc(is.na(rawpoll_clinton)))%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                                 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
## 7                                              CBS News/New York Times    A-       1426
## 8                                         NBC News/Wall Street Journal    A-       1282
## 9                                                           Zia Poll  <NA>       8439
## 10                                                          IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786               NA
## 8         43.58576      40.77325        5.365788               NA
## 9         44.82594      41.59978        7.870127               NA
## 10        42.92745      42.23545        6.316175               NA
```

```
polls_us_election_2016 %>% arrange(desc(is.na(rawpoll_trump)))%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                                 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
## 7                                          CBS News/New York Times    A-       1426
## 8                                      NBC News/Wall Street Journal    A-       1282
## 9                                                         Zia Poll  <NA>       8439
## 10                                                        IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
```

```
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786               NA
## 8         43.58576      40.77325        5.365788               NA
## 9         44.82594      41.59978        7.870127               NA
## 10        42.92745      42.23545        6.316175               NA
```

```
polls_us_election_2016 %>% arrange(desc(is.na(rawpoll_mcmullin)))%>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                 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
## 7                              CBS News/New York Times    A-       1426
## 8                          NBC News/Wall Street Journal    A-       1282
## 9                                                 Zia Poll  <NA>       8439
## 10                                                IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               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
## 7         45.11649      40.92722        4.341786              NA
## 8         43.58576      40.77325        5.365788              NA
## 9         44.82594      41.59978        7.870127              NA
## 10        42.92745      42.23545        6.316175              NA
```

You put the poll you want to observe within the brackets of is.na()

4.1

```
rawcandidate <- polls_us_election_2016 %>%
  mutate(raw_trump=rawpoll_trump/100*samplesize) %>%
  mutate(raw_clinton=rawpoll_clinton/100*samplesize) %>%
   mutate(raw_johnson=rawpoll_johnson/100*samplesize) %>%
  mutate(raw_mcmullin=rawpoll_mcmullin/100*samplesize)

View(rawcandidate)%>% head(10)
```

```
## NULL
```

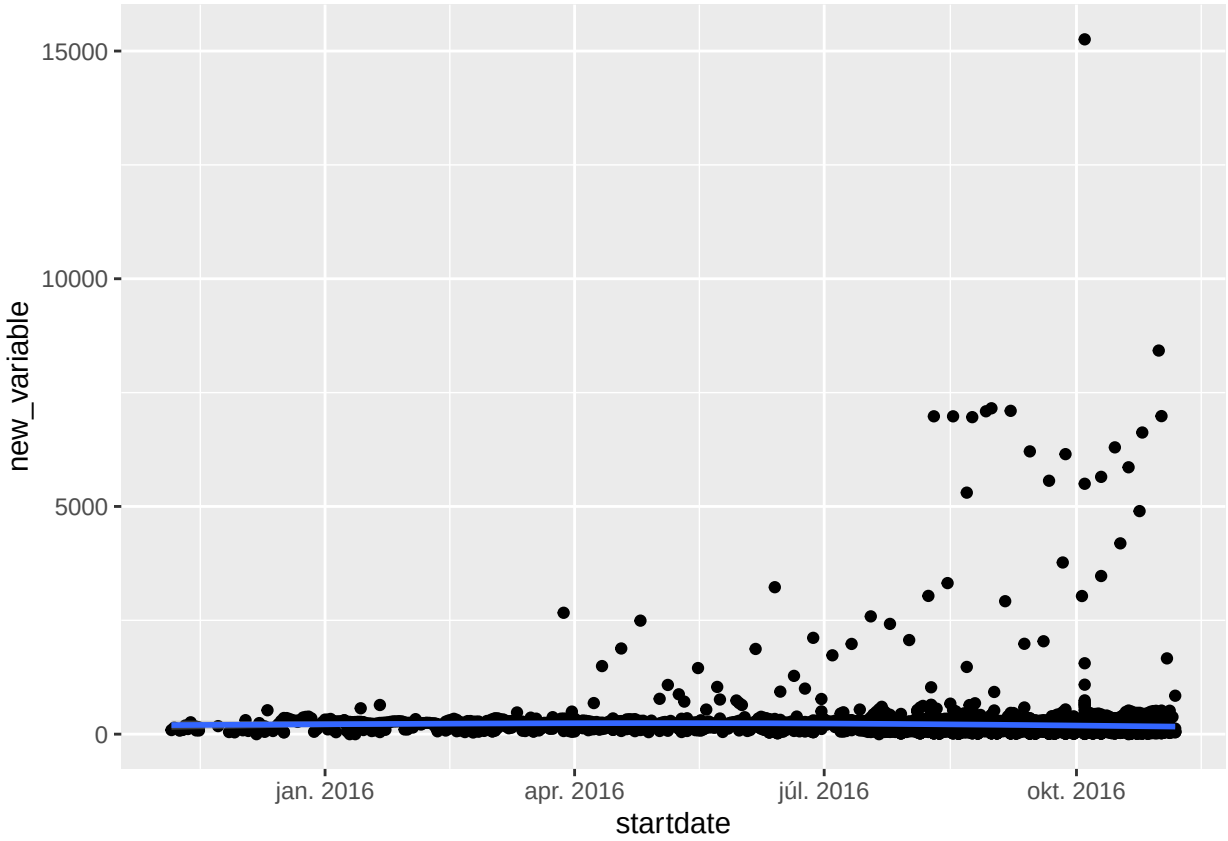
4.2

```
df1 <- rawcandidate %>%
  mutate(new_variable=samplesize-(raw_trump+raw_clinton))
ggplot(data = df1,
       mapping = aes(x = startdate,
                     y = new_variable)) +
  geom_point() +
  geom_smooth(se=TRUE)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 1 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 1 rows containing missing values (geom_point).
```



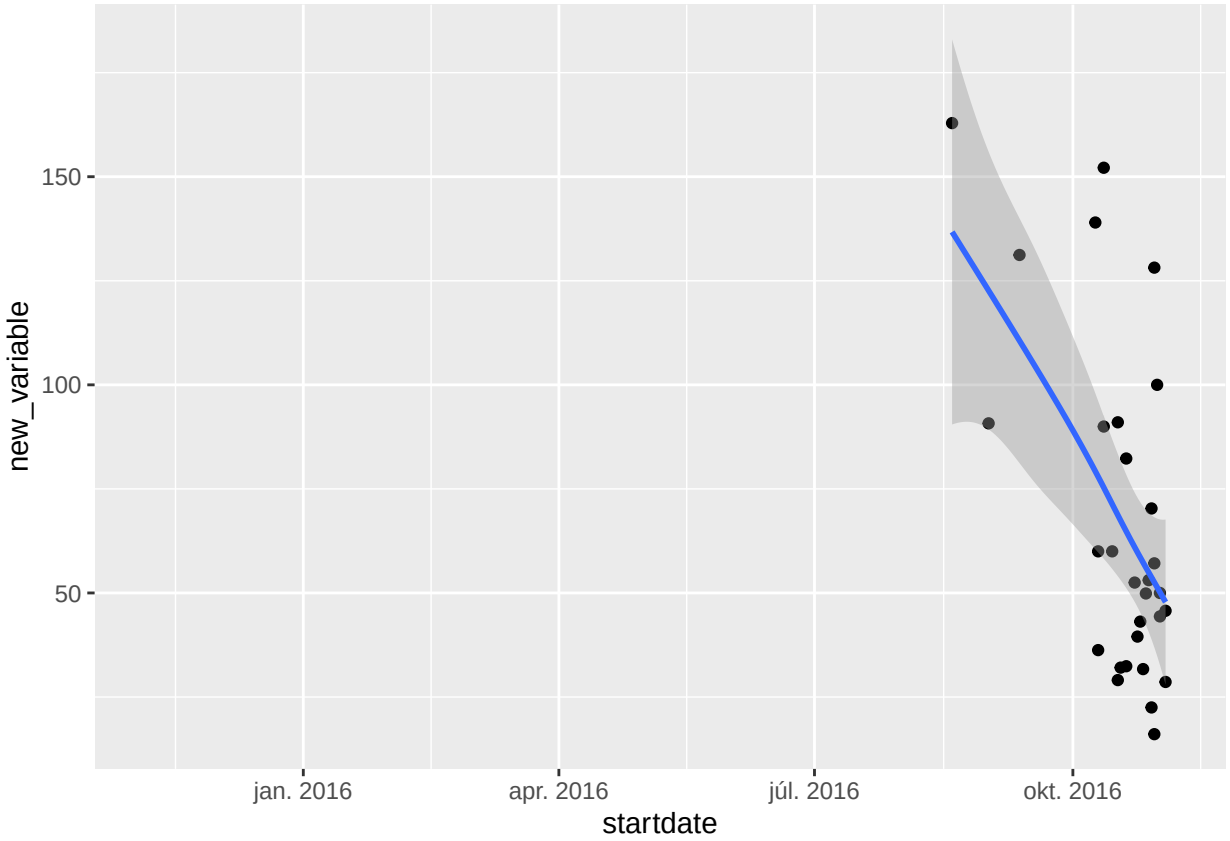
4.3

```
df2 <- df1%>%
  mutate(new_variable = samplesize-(raw_trump+raw_clinton+raw_johnson+raw_mcmullin))
ggplot(data = df2,
       mapping = aes(x = startdate,
                     y = new_variable)) +
  geom_point() +
  geom_smooth(se=TRUE)
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

```
## Warning: Removed 4178 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 4178 rows containing missing values (geom_point).
```



4.4 The reason why is because Johnson is a big candidate and he polled very well close to the election. We also need to take McMuller out of the picture although he was less significant. He polled very well in Utah for example with 21.5 percent of the votes and third in Idaho.

4.5 Ths mostly fixes the problem since Johnson was such a big candidate and when he is out of the picture, there is very little disturbing the data.

4.6

```
polls_us_election_2016 %>%
 select(startdate, pollster, rawpoll_johnson) %>%
 group_by(startdate) %>%
 mutate(rank = min_rank(desc(rawpoll_johnson)),ties.method = 'first')%>%
  head(10)
```

```
## # A tibble: 10 x 5
## # Groups:   startdate [5]
##    startdate  pollster                          rawpoll_johnson  rank ties.method
##    <date>     <fct>                                       <dbl> <int> <chr>
##  1 2016-11-03 ABC News/Washington Post                     4       12 first
##  2 2016-11-01 Google Consumer Surveys                      5.46    85 first
##  3 2016-11-02 Ipsos                                        6        2 first
##  4 2016-11-04 YouGov                                       5        5 first
##  5 2016-11-03 Gravis Marketing                             3       22 first
##  6 2016-11-03 Fox News/Anderson Robbins Resea~             3       22 first
##  7 2016-11-02 CBS News/New York Times                      5        6 first
##  8 2016-11-03 NBC News/Wall Street Journal                 6        5 first
```

```
##  9 2016-11-06 Zia Poll                                    6        2 first
## 10 2016-11-04 IBD/TIPP                                    7.1      2 first
```

We break ties with the first row that comes up.

5.1

```
polls_us_election_2016 %>%
count(pollster)%>%
head(10)
```

```
##                            pollster  n
## 1          ABC News/Washington Post 28
## 2           American Research Group  9
## 3               American Strategies  1
## 4                 Angus Reid Global  1
## 5     Anzalone Liszt Grove Research  2
## 6          Arizona State University  2
## 7  Associated Industries of Florida  3
## 8        Baldwin Wallace University  2
## 9            Ball State University  1
## 10                   Baruch College  1
```

```
polls_us_election_2016 %>%
  group_by(pollster) %>%
  summarise(pollster_n=n())%>%
  head(10)
```

```
## # A tibble: 10 x 2
##    pollster                          pollster_n
##    <fct>                                  <int>
##  1 ABC News/Washington Post                  28
##  2 American Research Group                    9
##  3 American Strategies                        1
##  4 Angus Reid Global                          1
##  5 Anzalone Liszt Grove Research              2
##  6 Arizona State University                   2
##  7 Associated Industries of Florida           3
##  8 Baldwin Wallace University                 2
##  9 Ball State University                      1
## 10 Baruch College                             1
```

5.2

```
polls_us_election_2016 %>%
  group_by(grade)%>%
  filter(state=="Florida")%>%
  summarise(mean=mean(samplesize))
```

```
## # A tibble: 10 x 2
##    grade  mean
##    <fct> <dbl>
```

```
##  1 C-    2290.
##  2 C      780.
##  3 C+     714.
##  4 B-    1322.
##  5 B     1285.
##  6 B+     754.
##  7 A-     825.
##  8 A     1221.
##  9 A+     536.
## 10 <NA>  1000.
```

5.3

```
polls_us_election_2016 %>%
  group_by(pollster)%>%
  count(grade)
```

```
## # A tibble: 196 x 3
## # Groups:   pollster [196]
##    pollster                          grade     n
##    <fct>                             <fct> <int>
##  1 ABC News/Washington Post          A+       28
##  2 American Research Group           C+        9
##  3 American Strategies               <NA>      1
##  4 Angus Reid Global                 A-        1
##  5 Anzalone Liszt Grove Research     C         2
##  6 Arizona State University          C+        2
##  7 Associated Industries of Florida  <NA>      3
##  8 Baldwin Wallace University        <NA>      2
##  9 Ball State University             <NA>      1
## 10 Baruch College                    B-        1
## # ... with 186 more rows
```

It seems so?

5.4

```
polls_us_election_2016 %>%
  group_by(state)%>%
  arrange(rawpoll_mcmullin)
```

```
## # A tibble: 4,208 x 15
## # Groups:   state [57]
##    state startdate  enddate    pollster              grade samplesize population
##    <fct> <date>     <date>     <fct>                 <fct>      <int> <chr>
##  1 Utah  2016-08-19 2016-08-21 Public Policy Polling B+          1018 lv
##  2 Utah  2016-09-01 2016-09-09 Dan Jones & Associat~ C+           605 lv
##  3 Utah  2016-09-12 2016-09-19 Dan Jones & Associat~ C+           820 lv
##  4 Utah  2016-10-09 2016-10-09 Google Consumer Surv~ B            500 a
##  5 Utah  2016-10-10 2016-10-12 Monmouth University   A+           403 lv
##  6 Utah  2016-10-12 2016-10-14 YouGov                B            951 lv
##  7 Utah  2016-10-29 2016-10-31 Rasmussen Reports/Pu~ C+           750 lv
##  8 Utah  2016-10-10 2016-10-11 Y2 Analytics          C+           500 lv
```

```
##  9 Utah  2016-11-03 2016-11-05 YouGov                B            762 lv
## 10 Utah  2016-10-30 2016-10-31 Gravis Marketing      B-          1424 rv
## # ... with 4,198 more rows, and 8 more variables: rawpoll_clinton <dbl>,
## #   rawpoll_trump <dbl>, rawpoll_johnson <dbl>, rawpoll_mcmullin <dbl>,
## #   adjpoll_clinton <dbl>, adjpoll_trump <dbl>, adjpoll_johnson <dbl>,
## #   adjpoll_mcmullin <dbl>
```

5.5

```
polls_us_election_2016 %>%
  group_by(state) %>%
    summarize(pollster_n=n()) %>%
    filter(pollster_n>10) %>%
  arrange(desc(pollster_n))
```

```
## # A tibble: 54 x 2
##    state          pollster_n
##    <fct>               <int>
##  1 U.S.                 1106
##  2 Florida               148
##  3 North Carolina        125
##  4 Pennsylvania          125
##  5 Ohio                  115
##  6 New Hampshire         112
##  7 Nevada                 93
##  8 Virginia               91
##  9 Michigan               86
## 10 Colorado               80
## # ... with 44 more rows
```

5.6

```
  polls_us_election_2016 %>%
    group_by(grade,population) %>%
    summarize(max_n=max(rawpoll_trump),min_n=min(rawpoll_trump)) %>%
  arrange(desc(max_n), desc(min_n))
```

```
## `summarise()` has grouped output by 'grade'. You can override using the `.groups` argument.
```

```
## # A tibble: 29 x 4
## # Groups:   grade [11]
##    grade population max_n min_n
##    <fct> <chr>      <dbl> <dbl>
##  1 <NA>  rv            68  28
##  2 B     lv            65   6.8
##  3 C-    lv            62   4
##  4 B-    lv            61  25
##  5 A-    lv            61  21.3
##  6 B+    lv            60  17
##  7 <NA>  lv            58  25
##  8 C+    lv            58  22
##  9 B+    v             57  28
## 10 C-    rv            57  24
## # ... with 19 more rows
```

5.7

```
polls_us_election_2016 %>%
  group_by(state) %>%
    summarize(pollster_n=n()) %>%
  filter(state =="Alabama" | state == "Arkansas") %>%
  arrange(desc(pollster_n))
```

```
## # A tibble: 2 x 2
##   state    pollster_n
##   <fct>         <int>
## 1 Arkansas         45
## 2 Alabama          43
```

```
polls_us_election_2016 %>%
  group_by(state) %>%
  filter(state !="Alabama" & state != "Arkansas") %>%
  summarize(pollster_n=n()) %>%
  arrange(desc(pollster_n))
```

```
## # A tibble: 55 x 2
##    state          pollster_n
##    <fct>               <int>
##  1 U.S.                 1106
##  2 Florida               148
##  3 North Carolina        125
##  4 Pennsylvania          125
##  5 Ohio                  115
##  6 New Hampshire         112
##  7 Nevada                 93
##  8 Virginia               91
##  9 Michigan               86
## 10 Colorado               80
## # ... with 45 more rows
```

They are running a total of 1106 polls in the United States excluding the polls in Alabama and Arkansas

6.1.a

```
df3 <- polls_us_election_2016%>%
  mutate(spread=(rawpoll_trump - rawpoll_clinton)/100)
df3 %>% head(10)
```

```
##         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
## 7        U.S. 2016-11-02 2016-11-06
## 8        U.S. 2016-11-03 2016-11-05
```

```
## 9  New Mexico 2016-11-06 2016-11-06
## 10       U.S. 2016-11-04 2016-11-07
##                                                    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
## 7                                  CBS News/New York Times    A-       1426
## 8                              NBC News/Wall Street Journal    A-       1282
## 9                                                 Zia Poll  <NA>       8439
## 10                                                IBD/TIPP    A-       1107
##    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
## 7          lv           45.00         41.00            5.00               NA
## 8          lv           44.00         40.00            6.00               NA
## 9          lv           46.00         44.00            6.00               NA
## 10         lv           41.20         42.70            7.10               NA
##    adjpoll_clinton adjpoll_trump adjpoll_johnson adjpoll_mcmullin  spread
## 1         45.20163      41.72430        4.626221               NA -0.0400
## 2         43.34557      41.21439        5.175792               NA -0.0234
## 3         42.02638      38.81620        6.844734               NA -0.0300
## 4         45.65676      40.92004        6.069454               NA -0.0400
## 5         46.84089      42.33184        3.726098               NA -0.0400
## 6         49.02208      43.95631        3.057876               NA -0.0400
## 7         45.11649      40.92722        4.341786               NA -0.0400
## 8         43.58576      40.77325        5.365788               NA -0.0400
## 9         44.82594      41.59978        7.870127               NA -0.0200
## 10        42.92745      42.23545        6.316175               NA  0.0150
```

```
df4 <- polls_us_election_2016%>%
  mutate(spread=(rawpoll_trump - rawpoll_clinton)/100) %>%
  group_by(pollster) %>%
    summarize(spread, mean_spread_n=mean(spread))
```

```
## `summarise()` has grouped output by 'pollster'. You can override using the `.groups` argument.
```

```
df4 %>% head(10)
```

```
## # A tibble: 10 x 3
## # Groups:   pollster [1]
##    pollster                 spread mean_spread_n
##    <fct>                     <dbl>         <dbl>
##  1 ABC News/Washington Post  -0.04       -0.0668
##  2 ABC News/Washington Post  -0.06       -0.0668
##  3 ABC News/Washington Post  -0.04       -0.0668
##  4 ABC News/Washington Post  -0.05       -0.0668
```

```
##  5 ABC News/Washington Post  -0.36       -0.0668
##  6 ABC News/Washington Post  -0.04       -0.0668
##  7 ABC News/Washington Post  -0.03       -0.0668
##  8 ABC News/Washington Post  -0.02       -0.0668
##  9 ABC News/Washington Post   0          -0.0668
## 10 ABC News/Washington Post   0.01       -0.0668
```

6.1.b.a

```
df5 <- df3%>%
  mutate(p=(spread+1)/2)

View(df5 %>% head(10))
```

6.1.b.b

```
df6 <- df5%>%
  mutate(sd=2*sqrt((p*(1-p)/samplesize)))
View(df6)
```

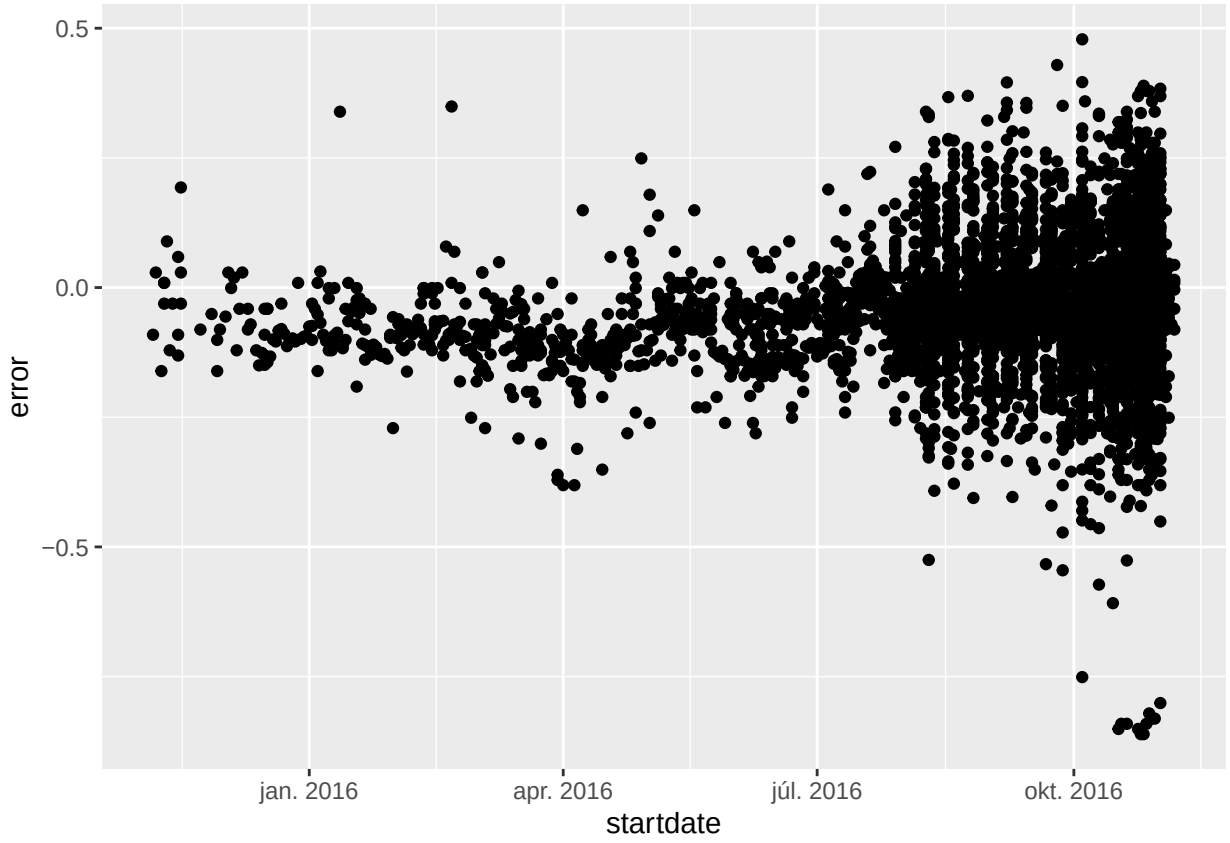
It's true because this is a typical proportional sample and the two candidates dominate.

6.1.b.c

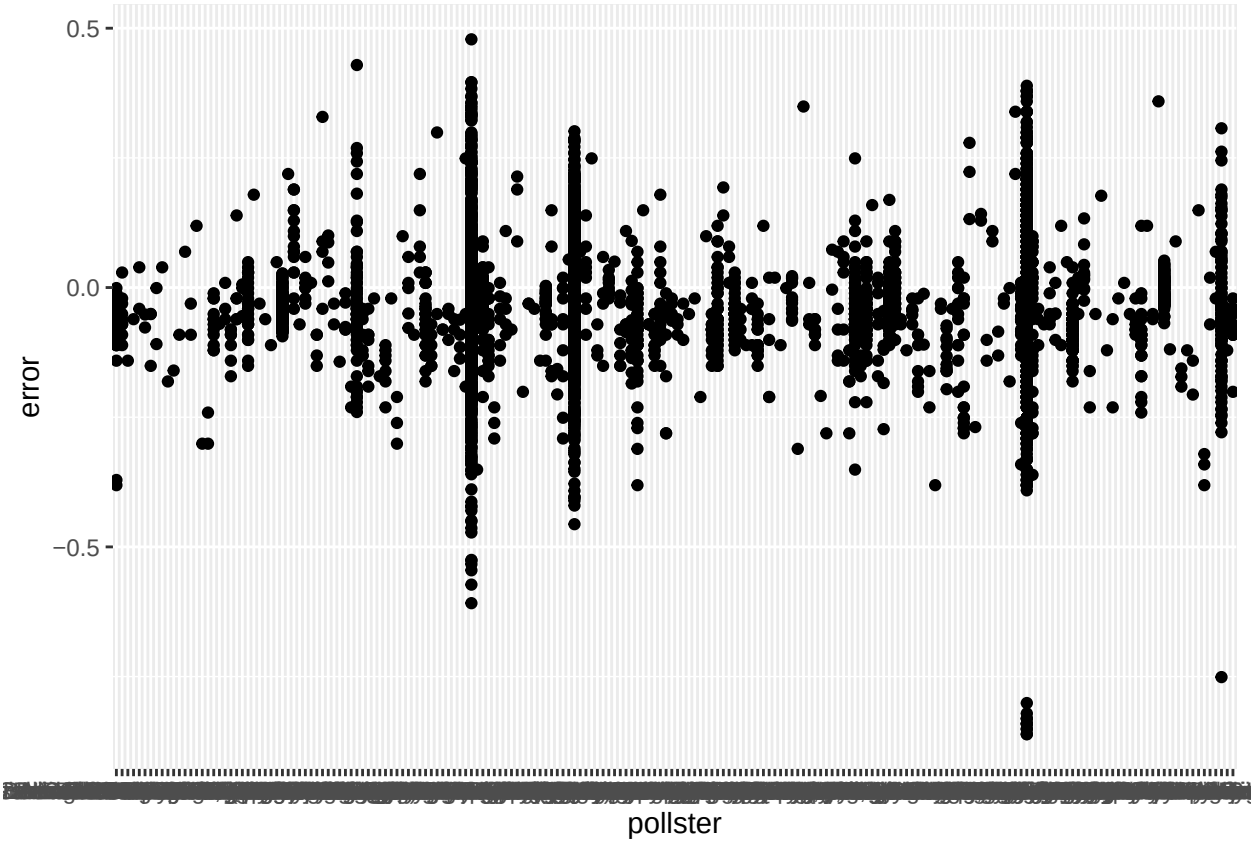
```
df7 <-df6%>%
  mutate(lci=spread-qnorm(0.975)*sd) %>%
  mutate(uci=spread+qnorm(0.975)*sd)
View(df7)
```

6.1.c

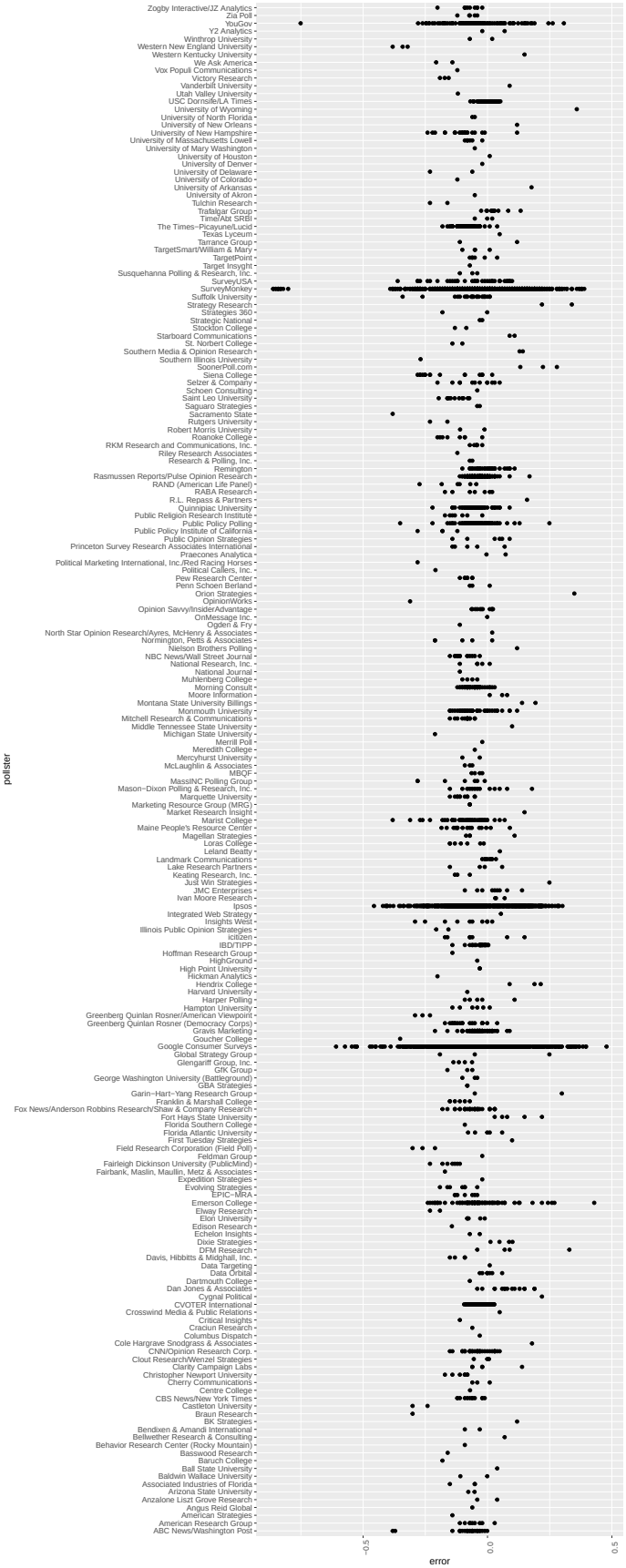
```
df8<- df7%>%
  mutate(error=spread-0.021)
View(df8 %>% head(10))

ggplot(data = df8, aes(x = startdate, y = error,
                       group = pollster))+
  geom_point()
```



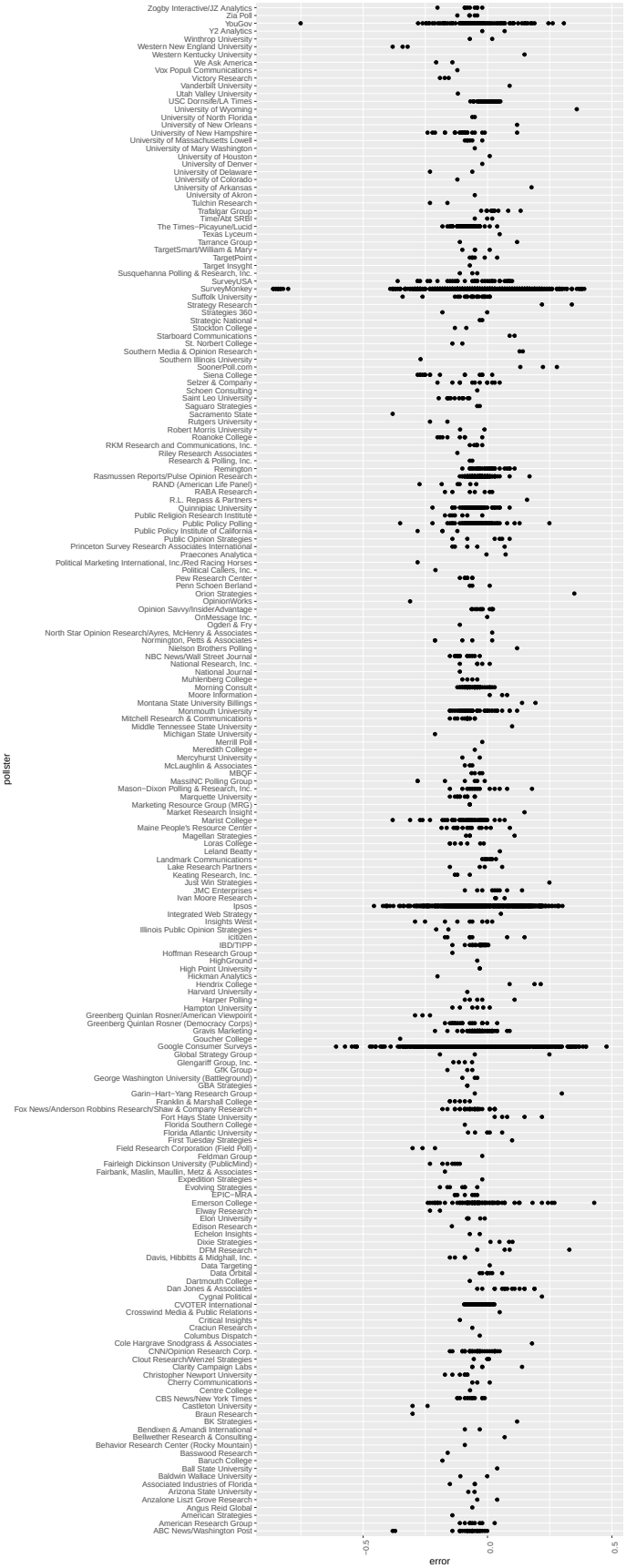

```
ggplot(data = df8, aes(x = pollster, y = error,
                       group = pollster))+
  geom_point()
```




```
ggplot(data = df8, aes(x = error, y = pollster,
                       group = pollster))+
  theme(axis.text.x = element_text(angle = 90))+
  geom_point()
```


pollster
Zogby Interactive/JZ Analytics
Zia Poll
YouGov
Y2 Analytics
Winthrop University
Western New England University
Western Kentucky University
We Ask America
Vox Populi Communications
Victory Research
Vanderbilt University
Utah Valley University
USC Dornsife/LA Times
University of Wyoming
University of North Florida
University of New Orleans
University of New Hampshire
University of Massachusetts Lowell
University of Mary Washington
University of Houston
University of Denver
University of Delaware
University of Colorado
University of Arkansas
University of Akron
Tulchin Research
Trafalgar Group
Time/Abt SRBI
The Times-Picayune/Lucid
Texas Lyceum
Tarrance Group
TargetSmart/William & Mary
TargetPoint
Target Insyght
Susquehanna Polling & Research, Inc.
SurveyUSA
SurveyMonkey
Suffolk University
Strategy Research
Strategies 360
Strategic National
Stockton College
Starboard Communications
St. Norbert College
Southern Media & Opinion Research
Southern Illinois University
SoonerPoll.com
Siena College
Selzer & Company
Schoen Consulting
Saint Leo University
Saguaro Strategies
Sacramento State
Rutgers University
Robert Morris University
Roanoke College
RKM Research and Communications, Inc.
Riley Research Associates
Research & Polling, Inc.
Remington
Rasmussen Reports/Pulse Opinion Research
RAND (American Life Panel)
RABA Research
R.L. Repass & Partners
Quinnipiac University
Public Religion Research Institute
Public Policy Polling
Public Policy Institute of California
Public Opinion Strategies
Princeton Survey Research Associates International
Praecones Analytica
Political Marketing International, Inc./Red Racing Horses
Political Callers, Inc.
Pew Research Center
Penn Schoen Berland
Orion Strategies
OpinionWorks
Opinion Savvy/InsiderAdvantage
OnMessage Inc.
Ogden & Fry
North Star Opinion Research/Ayres, McHenry & Associates
Normington, Petts & Associates
Nielson Brothers Polling
NBC News/Wall Street Journal
National Research, Inc.
National Journal
Muhlenberg College
Morning Consult
Moore Information
Montana State University Billings
Monmouth University
Mitchell Research & Communications
Middle Tennessee State University
Michigan State University
Merrill Poll
Meredith College
Mercyhurst University
McLaughlin & Associates
MBQF
MassINC Polling Group
Mason-Dixon Polling & Research, Inc.
Marquette University
Marketing Resource Group (MRG)
Market Research Insight
Marist College
Maine People's Resource Center
Magellan Strategies
Loras College
Leland Beatty
Landmark Communications
Lake Research Partners
Keating Research, Inc.
Just Win Strategies
JMC Enterprises
Ivan Moore Research
Ipsos
Integrated Web Strategy
Insights West
Illinois Public Opinion Strategies
icitizen
IBD/TIPP
Hoffman Research Group
HighGround
High Point University
Hickman Analytics
Hendrix College
Harvard University
Harper Polling
Hampton University
Greenberg Quinlan Rosner/American Viewpoint
Greenberg Quinlan Rosner (Democracy Corps)
Gravis Marketing
Goucher College
Google Consumer Surveys
Global Strategy Group
Glengariff Group, Inc.
GfK Group
George Washington University (Battleground)
GBA Strategies
Garin-Hart-Yang Research Group
Franklin & Marshall College
Fox News/Anderson Robbins Research/Shaw & Company Research
Fort Hays State University
Florida Southern College
Florida Atlantic University
First Tuesday Strategies
Field Research Corporation (Field Poll)
Feldman Group
Fairleigh Dickinson University (PublicMind)
Fairbank, Maslin, Maullin, Metz & Associates
Expedition Strategies
Evolving Strategies
EPIC-MRA
Emerson College
Elway Research
Elon University
Edison Research
Echelon Insights
Dixie Strategies
DFM Research
Davis, Hibbitts & Midghall, Inc.
Data Targeting
Data Orbital
Dartmouth College
Dan Jones & Associates
Cygnal Political
CVOTER International
Crosswind Media & Public Relations
Critical Insights
Craciun Research
Columbus Dispatch
Cole Hargrave Snodgrass & Associates
CNN/Opinion Research Corp.
Clout Research/Wenzel Strategies
Clarity Campaign Labs
Christopher Newport University
Cherry Communications
Centre College
CBS News/New York Times
Castleton University
Braun Research
BK Strategies
Bendixen & Amandi International
Bellwether Research & Consulting
Behavior Research Center (Rocky Mountain)
Basswood Research
Baruch College
Ball State University
Baldwin Wallace University
Associated Industries of Florida
Arizona State University
Anzalone Liszt Grove Research
Angus Reid Global
American Strategies
American Research Group
ABC News/Washington Post
−0.5
0.0
0.5
error

```r
ggplot(data = df8, aes(x = error, y = pollster,
                       group = pollster))+
  theme(axis.text.x = element_text(angle = 90))+
  geom_point()
```


pollster
Zogby Interactive/JZ Analytics
Zia Poll
YouGov
Y2 Analytics
Winthrop University
Western New England University
Western Kentucky University
We Ask America
Vox Populi Communications
Victory Research
Vanderbilt University
Utah Valley University
USC Dornsife/LA Times
University of Wyoming
University of North Florida
University of New Orleans
University of New Hampshire
University of Massachusetts Lowell
University of Mary Washington
University of Houston
University of Denver
University of Delaware
University of Colorado
University of Arkansas
University of Akron
Tulchin Research
Trafalgar Group
Time/Abt SRBI
The Times-Picayune/Lucid
Texas Lyceum
Tarrance Group
TargetSmart/William & Mary
TargetPoint
Target Insyght
Susquehanna Polling & Research, Inc.
SurveyUSA
SurveyMonkey
Suffolk University
Strategy Research
Strategies 360
Strategic National
Stockton College
Starboard Communications
St. Norbert College
Southern Media & Opinion Research
Southern Illinois University
SoonerPoll.com
Siena College
Selzer & Company
Schoen Consulting
Saint Leo University
Saguaro Strategies
Sacramento State
Rutgers University
Robert Morris University
Roanoke College
RKM Research and Communications, Inc.
Riley Research Associates
Research & Polling, Inc.
Remington
Rasmussen Reports/Pulse Opinion Research
RAND (American Life Panel)
RABA Research
R.L. Repass & Partners
Quinnipiac University
Public Religion Research Institute
Public Policy Polling
Public Policy Institute of California
Public Opinion Strategies
Princeton Survey Research Associates International
Praecones Analytica
Political Marketing International, Inc./Red Racing Horses
Political Callers, Inc.
Pew Research Center
Penn Schoen Berland
Orion Strategies
OpinionWorks
Opinion Savvy/InsiderAdvantage
OnMessage Inc.
Ogden & Fry
North Star Opinion Research/Ayres, McHenry & Associates
Normington, Petts & Associates
Nielson Brothers Polling
NBC News/Wall Street Journal
National Research, Inc.
National Journal
Muhlenberg College
Morning Consult
Moore Information
Montana State University Billings
Monmouth University
Mitchell Research & Communications
Middle Tennessee State University
Michigan State University
Merrill Poll
Meredith College
Mercyhurst University
McLaughlin & Associates
MBQF
MassINC Polling Group
Mason-Dixon Polling & Research, Inc.
Marquette University
Marketing Resource Group (MRG)
Market Research Insight
Marist College
Maine People's Resource Center
Magellan Strategies
Loras College
Leland Beatty
Landmark Communications
Lake Research Partners
Keating Research, Inc.
Just Win Strategies
JMC Enterprises
Ivan Moore Research
Ipsos
Integrated Web Strategy
Insights West
Illinois Public Opinion Strategies
icitizen
IBD/TIPP
Hoffman Research Group
HighGround
High Point University
Hickman Analytics
Hendrix College
Harvard University
Harper Polling
Hampton University
Greenberg Quinlan Rosner/American Viewpoint
Greenberg Quinlan Rosner (Democracy Corps)
Gravis Marketing
Goucher College
Google Consumer Surveys
Global Strategy Group
Glengariff Group, Inc.
GfK Group
George Washington University (Battleground)
GBA Strategies
Garin-Hart-Yang Research Group
Franklin & Marshall College
Fox News/Anderson Robbins Research/Shaw & Company Research
Fort Hays State University
Florida Southern College
Florida Atlantic University
First Tuesday Strategies
Field Research Corporation (Field Poll)
Feldman Group
Fairleigh Dickinson University (PublicMind)
Fairbank, Maslin, Maullin, Metz & Associates
Expedition Strategies
Evolving Strategies
EPIC-MRA
Emerson College
Elway Research
Elon University
Edison Research
Echelon Insights
Dixie Strategies
DFM Research
Davis, Hibbitts & Midghall, Inc.
Data Targeting
Data Orbital
Dartmouth College
Dan Jones & Associates
Cygnal Political
CVOTER International
Crosswind Media & Public Relations
Critical Insights
Craciun Research
Columbus Dispatch
Cole Hargrave Snodgrass & Associates
CNN/Opinion Research Corp.
Clout Research/Wenzel Strategies
Clarity Campaign Labs
Christopher Newport University
Cherry Communications
Centre College
CBS News/New York Times
Castleton University
Braun Research
BK Strategies
Bendixen & Amandi International
Bellwether Research & Consulting
Behavior Research Center (Rocky Mountain)
Basswood Research
Baruch College
Ball State University
Baldwin Wallace University
Associated Industries of Florida
Arizona State University
Anzalone Liszt Grove Research
Angus Reid Global
American Strategies
American Research Group
ABC News/Washington Post
−0.5
0.0
0.5
error