Jessenia Garcia

library(readxl)

table02\_6 <- read\_excel("R/table02\_6.xls")

View(table02\_6)

str (table02\_6)

names(table02\_6)

#mean

mean(table02\_6$`Total Citizen Population`)

mean (table02\_6$`Reported registered`)

mean (table02\_6$`Reported not registered`)

mean (table02\_6$`Reported voted`)

mean (table02\_6$`Reported did not vote`)

#median

median(table02\_6$`Total Citizen Population`)

median (table02\_6$`Reported registered`)

median (table02\_6$`Reported not registered`)

median (table02\_6$`Reported voted`)

median (table02\_6$`Reported did not vote`)

#mode function

temp1<-table(as.vector(table02\_6$'Total Citizen Population'))

names(temp1[temp1 == max(temp1)])

View(temp1)

temp2<-table(as.vector(table02\_6$`Reported registered`))

names(temp2[temp2 == max(temp2)])

View(temp2)

temp3<-table(as.vector(table02\_6$`Reported not registered`))

names(temp3[temp3 == max(temp3)])

View(temp3)

temp4<-table(as.vector(table02\_6$`Reported voted`))

names(temp4[temp4 == max(temp4)])

View(temp4)

temp5<-table(as.vector(table02\_6$`Reported did not vote`))

names(temp5[temp5 == max(temp5)])

View(temp5)

#SD

sd(table02\_6$`Reported voted`)

#variance

var(table02\_6$`Reported voted`)

#sd

sd(vote) = sqrt(var(vote))

var (vote)

#range

max(table02\_6$`Reported voted`) - min (table02\_6$`Reported voted`)

boxplot(table02\_6$`Reported voted`)

hist(table02\_6$`Reported voted`)

Complete the following tasks:

* Find a dataset that you find interesting. You can look at a variety of sources for data including:
  + Religion, health, etc.: [www.Thearda.com](http://www.Thearda.com)
  + Census, Education, Wealth, Time-Use, etc: [www.ipums.org](http://www.ipums.org)
  + Education: [www.nces.ed.gov](http://www.nces.ed.gov)
  + Mental Health: [www.cdc.gov/mentalhealth/data\_publications/index.htm](http://www.cdc.gov/mentalhealth/data_publications/index.htm)
  + Many Other Sources
* Once you have identified a dataset, explain how you accessed it. What links did you click? Was there any registration required? Did you download directly or was there an online system you navigated?

I googled Hispanic voting records / US Census borough. I came to a link <https://www.census.gov/data/tables/time-series/demo/voting-and-registration/p20-580.html> . I then selected Hispanic. I was able to directly download it as an excel sheet.

* In R:
  + Import the data
  + Provide a list of variables in the dataset
    - Copy and paste this list in your word document
* > names(table02\_6)
* [1] "Age" "Total Citizen Population"
* [3] "Reported registered" "Reported not registered"
* [5] "No response to registration 1" "Reported voted"
* [7] "Reported did not vote" "No response to voting 2"
  + What is the structure of the data? Which variables are character and which are numeric?
    - Copy and paste this list in your word document
* Classes ‘tbl\_df’, ‘tbl’ and 'data.frame': 5 obs. of 8 variables:
* $ Age : chr "18 to 24 years" "25 to 44 years" "45 to 64 years" "65 to 74 years" ...
* $ Total Citizen Population : num 5084 10729 7614 1993 1242
* $ Reported registered : num 2340 6059 4776 1315 777
* $ Reported not registered : num 1580 2811 1408 356 240
* $ No response to registration 1: num 1165 1859 1430 322 225
* $ Reported voted : num 1743 4951 4153 1176 659
* $ Reported did not vote : num 2213 3966 2094 487 356
* $ No response to voting 2 : num 1128 1812 1367 330 226

They are all numeric except for Age.

Describe the data through Central Tendencies:

* + - Give me the mean, median, and mode of **five variables**.

Mean:

Total population -5332.4

Reported Registered – 3053.4

Reported not registered 1279

Reported Voted – 2536.4

Reported Not vote - 1823.2

Median

Total population -5083

Reported Registered – 2340

Reported not registered- 1408

Reported Voted – 1743

Reported Not vote - 2093

Mode

Total population -None

Reported Registered – None

Reported not registered- none

Reported Voted – None

Reported Not vote - None

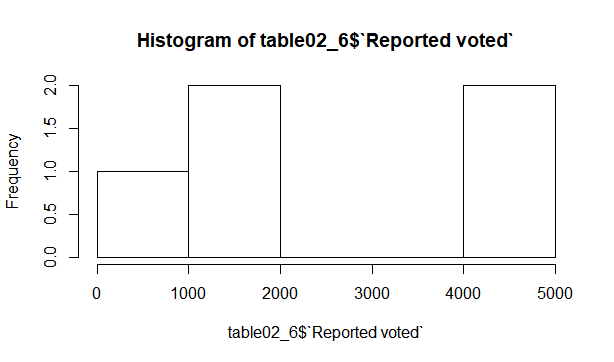
* + Describe the data through variation:
    - Choose one variable and give me: Reported Voted
    - The variance, range, and standard deviation.

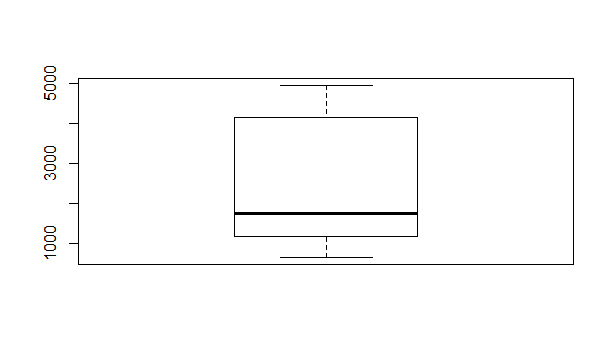
Variance = 3612123

Range 4292

Standard Deviation - 1900.559

* + - A histogram and either box-plot or stem-leaf plot





**What to turn in**

* Publish your homework to GitHub under your user account, the appropriate format (check syllabus for instructions). This means turn in a separate document from your code that answers the questions above.
* Upload the code you used to GitHub.
* Provide me the links of these two files via e-mail no later than Thursday, September 20th at 6:00pm.