nasa\_lab

Ryan Green

2023-02-15

#1  
astronauts <- read\_csv("Astronauts.csv")

## Rows: 357 Columns: 19  
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (13): Name, Status, Birth Date, Birth Place, Gender, Alma Mater, Undergr...  
## dbl (6): Year, Group, Space Flights, Space Flight (hr), Space Walks, Space ...  
##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

astronauts

## # A tibble: 357 x 19  
## Name Year Group Status Birth~1 Birth~2 Gender Alma ~3 Under~4 Gradu~5  
## <chr> <dbl> <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Joseph M. ~ 2004 19 Active 5/17/1~ Inglew~ Male Univer~ Geology Geology  
## 2 Loren W. A~ NA NA Retir~ 3/7/19~ Lewist~ Male Montan~ Engine~ Solar ~  
## 3 James C. A~ 1984 10 Retir~ 3/3/19~ Warsaw~ Male US Mil~ Engine~ Aerosp~  
## 4 Thomas D. ~ 1987 12 Retir~ 5/20/1~ St. Lo~ Male Univer~ Applie~ Applie~  
## 5 Buzz Aldrin 1963 3 Retir~ 1/20/1~ Montcl~ Male US Mil~ Mechan~ Astron~  
## 6 Andrew M. ~ 1987 12 Retir~ 8/4/19~ Philad~ Male Villan~ Mechan~ Busine~  
## 7 Joseph P. ~ 1967 6 Retir~ 6/27/1~ Crawsf~ Male DePauw~ Mathem~ Physics  
## 8 Scott D. A~ 1995 15 Retir~ 8/15/1~ Lincol~ Male Univer~ Aerona~ Aerona~  
## 9 William A.~ 1963 3 Retir~ 10/17/~ Hong K~ Male US Nav~ Nuclea~ Nuclea~  
## 10 Clayton C.~ 1998 17 Retir~ 2/23/1~ Omaha,~ Male Hastin~ Physics Aerosp~  
## # ... with 347 more rows, 9 more variables: `Military Rank` <chr>,  
## # `Military Branch` <chr>, `Space Flights` <dbl>, `Space Flight (hr)` <dbl>,  
## # `Space Walks` <dbl>, `Space Walks (hr)` <dbl>, Missions <chr>,  
## # `Death Date` <chr>, `Death Mission` <chr>, and abbreviated variable names  
## # 1: `Birth Date`, 2: `Birth Place`, 3: `Alma Mater`,  
## # 4: `Undergraduate Major`, 5: `Graduate Major`

#2  
nrow(astronauts)

## [1] 357

ncol(astronauts)

## [1] 19

#3  
women\_astronauts <- astronauts %>%   
 filter(Gender == "Female")  
  
women\_astronauts

## # A tibble: 50 x 19  
## Name Year Group Status Birth~1 Birth~2 Gender Alma ~3 Under~4 Gradu~5  
## <chr> <dbl> <dbl> <chr> <chr> <chr> <chr> <chr> <chr> <chr>   
## 1 Serena M. ~ 2009 20 Active 4/9/19~ Indian~ Female George~ Electr~ Medici~  
## 2 Ellen S. B~ 1984 10 Retir~ 4/27/1~ Fayett~ Female State ~ Geology Medici~  
## 3 Yvonne D. ~ 1996 16 Manag~ 4/24/1~ West P~ Female San Fr~ Bioche~ <NA>   
## 4 Tracy E. C~ 1998 17 Active 8/14/1~ Arcadi~ Female Califo~ Chemis~ Physic~  
## 5 Kalpana Ch~ 1995 15 Decea~ 6/1/19~ Karnal~ Female Punjab~ Aerona~ Aerosp~  
## 6 Laurel B. ~ 1996 16 Decea~ 3/10/1~ Ames, ~ Female Univer~ Zoology Medici~  
## 7 Mary L. Cl~ 1980 9 Retir~ 2/5/19~ Southa~ Female Colora~ Biolog~ Microb~  
## 8 Catherine ~ 1992 14 Active 12/14/~ Charle~ Female MIT; U~ Chemis~ Polyme~  
## 9 Eileen M. ~ 1990 13 Retir~ 11/19/~ Elmira~ Female Syracu~ Mathem~ Operat~  
## 10 Nancy J. C~ 1990 13 Manag~ 12/29/~ Wilmin~ Female Ohio S~ Biolog~ Safety~  
## # ... with 40 more rows, 9 more variables: `Military Rank` <chr>,  
## # `Military Branch` <chr>, `Space Flights` <dbl>, `Space Flight (hr)` <dbl>,  
## # `Space Walks` <dbl>, `Space Walks (hr)` <dbl>, Missions <chr>,  
## # `Death Date` <chr>, `Death Mission` <chr>, and abbreviated variable names  
## # 1: `Birth Date`, 2: `Birth Place`, 3: `Alma Mater`,  
## # 4: `Undergraduate Major`, 5: `Graduate Major`

nrow(women\_astronauts)

## [1] 50

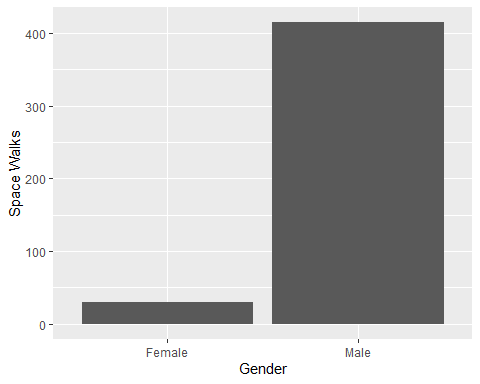
#4  
mil\_ranK\_and\_branch <- women\_astronauts %>%   
 select("Name", "Military Rank", "Military Branch")  
  
mil\_ranK\_and\_branch

## # A tibble: 50 x 3  
## Name `Military Rank` `Military Branch`   
## <chr> <chr> <chr>   
## 1 Serena M. Aunon <NA> <NA>   
## 2 Ellen S. Baker <NA> <NA>   
## 3 Yvonne D. Cagle Colonel US Air Force   
## 4 Tracy E. Caldwell (Dyson) <NA> <NA>   
## 5 Kalpana Chawla <NA> <NA>   
## 6 Laurel B. Clark Captain US Navy   
## 7 Mary L. Cleave <NA> <NA>   
## 8 Catherine G. Coleman Colonel US Air Force (Retired)  
## 9 Eileen M. Collins Colonel US Air Force (Retired)  
## 10 Nancy J. Currie Colonel US Army (Retired)   
## # ... with 40 more rows

#5  
act\_vs\_nonact <- women\_astronauts %>%   
 mutate(act\_dummy = if\_else(Status == "Active", 1, 0))  
  
sum(act\_vs\_nonact$act\_dummy)

## [1] 13

#6  
ggplot(data = astronauts)+  
 geom\_bar(mapping = aes(x=Gender, y= `Space Walks`), stat="identity")



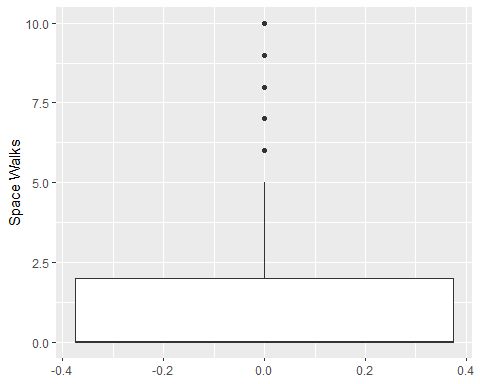
#7  
spc\_walk <- women\_astronauts %>%   
 select("Name", "Gender", "Space Walks") %>%   
 #count(Name, Gender, Space Walks) %>%   
 filter(`Space Walks`>=1)   
   
  
spc\_walk

## # A tibble: 11 x 3  
## Name Gender `Space Walks`  
## <chr> <chr> <dbl>  
## 1 Tracy E. Caldwell (Dyson) Female 3  
## 2 Linda M. Godwin Female 2  
## 3 Susan J. Helms Female 1  
## 4 Tamara E. Jernigan Female 1  
## 5 Kathleen Rubins Female 2  
## 6 Heidemarie M. Stefanyshyn-Piper Female 2  
## 7 Nicole P. Stott Female 1  
## 8 Kathryn D. Sullivan Female 1  
## 9 Kathryn C. Thornton Female 3  
## 10 Peggy A. Whitson Female 7  
## 11 Sunita L. Williams Female 7

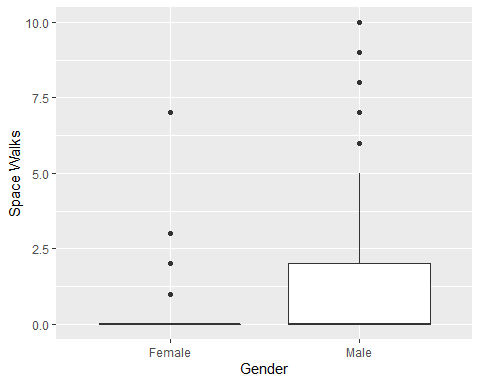
#8  
box\_plot <- ggplot(astronauts)+  
 geom\_boxplot(mapping = aes(y=`Space Walks`))  
  
IQR(astronauts$`Space Walks`)

## [1] 2

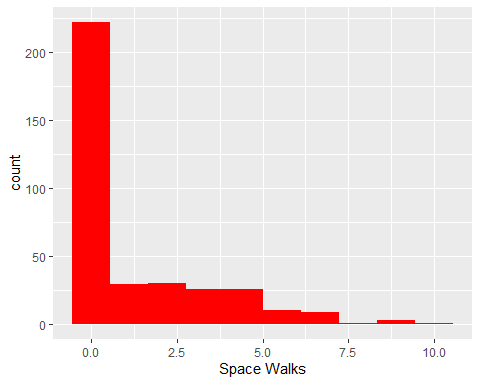
box\_plot

 There are some outliers. Based on the output I would estimate the IQR to be around 2

#9  
box\_plot2 <- ggplot(astronauts)+  
 geom\_boxplot(mapping=aes(x=`Gender`, y=`Space Walks`))  
  
box\_plot2



#10  
ggplot(data = astronauts)+  
 geom\_histogram(mapping = aes(x = `Space Walks`), bins = 10,   
 fill = "red")

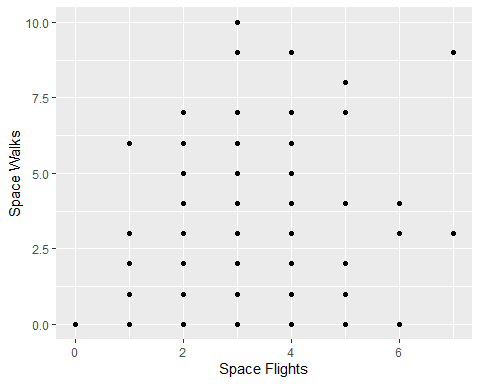


#11  
astronauts %>%   
 group\_by(Gender) %>%   
 summarize(Mean = mean(`Space Walks`))

## # A tibble: 2 x 2  
## Gender Mean  
## <chr> <dbl>  
## 1 Female 0.6   
## 2 Male 1.35

#THIS IS THE QUESTION I EMAILED YOU AND WENT OVER IN CLASS. NOT SURE WHY IT WASN'T GROUPING

#12  
ggplot(data=astronauts) +  
 geom\_point(mapping=aes(x = `Space Flights`,  
 y = `Space Walks`))

 There doesn’t seem to be a strong linear relationship

#13  
mae\_table <- astronauts %>%   
 filter(Name=="Mae C. Jemison") %>%   
 select("Name", "Undergraduate Major", "Graduate Major", "Alma Mater")  
  
mae\_table

## # A tibble: 1 x 4  
## Name `Undergraduate Major` `Graduate Major` `Alma Mater`   
## <chr> <chr> <chr> <chr>   
## 1 Mae C. Jemison Chemical Engineering Medicine Stanford University; Co~

#14  
tbl\_14 <- women\_astronauts %>%   
 filter(`Military Branch`=="US Air Force"|`Military Branch`=="US Air Force (Retired)") %>%   
 select("Name", "Gender", "Military Branch")  
  
number\_of\_female\_astron\_us\_airforce <- nrow(tbl\_14)  
  
number\_of\_female\_astron\_us\_airforce

## [1] 5

tbl\_14

## # A tibble: 5 x 3  
## Name Gender `Military Branch`   
## <chr> <chr> <chr>   
## 1 Yvonne D. Cagle Female US Air Force   
## 2 Catherine G. Coleman Female US Air Force (Retired)  
## 3 Eileen M. Collins Female US Air Force (Retired)  
## 4 Susan J. Helms Female US Air Force   
## 5 Pamela A. Melroy Female US Air Force (Retired)