Stat123 Assignment #3

Evan O’Toole

library(tidyverse) #observe which packages are loaded any potential conflicts

Warning: package 'lubridate' was built under R version 4.3.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(nycflights13)  
library(gapminder)  
library(ggplot2)  
set.seed(42)  
age = sample(18:65, 50, replace = TRUE)  
height = runif(50, 150, 190)  
weight = runif(50,45,100)  
gender = sample(c("Male", "Female"), 50, replace = TRUE)  
data = data.frame(Age = age, Height = height, Weight = weight, Gender = gender)

Question 1

#a)  
AH <- data |>   
 select(Age, Height)  
head(AH)

Age Height  
1 54 170.5765  
2 18 177.0243  
3 42 189.3127  
4 27 180.3818  
5 53 172.6595  
6 35 183.9876

#b)  
A\_30 <- data |>  
 filter(Age > 30)  
head(A\_30)

Age Height Weight Gender  
1 54 170.5765 45.12501 Female  
2 42 189.3127 91.02409 Female  
3 53 172.6595 69.90024 Male  
4 35 183.9876 74.46845 Female  
5 64 157.5790 74.55572 Female  
6 41 160.8515 45.07595 Male

#c)  
Height\_d <- data |>   
 arrange(desc(Height))  
head(Height\_d)

Age Height Weight Gender  
1 42 189.3127 91.02409 Female  
2 51 188.5043 87.67038 Male  
3 21 187.6982 82.32261 Male  
4 60 187.3214 87.86525 Female  
5 32 187.0258 85.34404 Female  
6 35 183.9876 74.46845 Female

#d)  
sumar <- data |>   
 summarize(mean\_age = mean(Age), median\_age = median(Age), sd\_age = sd(Age), var\_age = var(Age))  
head(sumar)

mean\_age median\_age sd\_age var\_age  
1 43.24 43.5 13.46888 181.4106

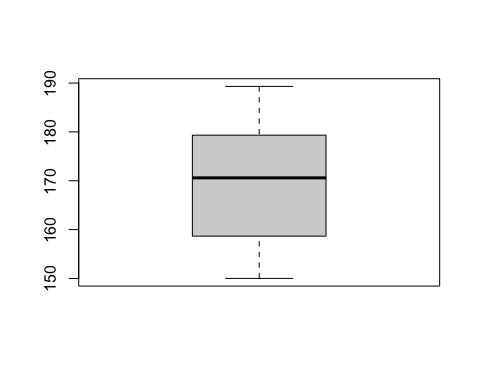
#e)  
quarts <- data |>   
 summarize(Q1 = quantile(Height, 0.25), Q3 = quantile(Height, 0.75))  
head(quarts)

Q1 Q3  
1 158.6686 179.1909

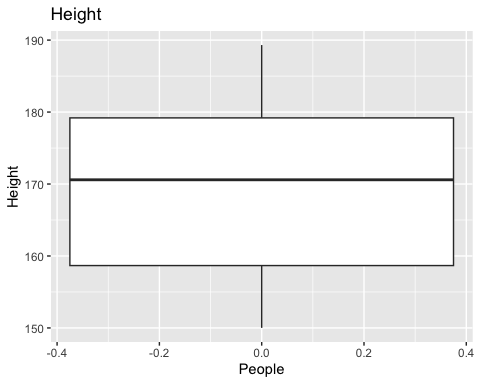
# or just quantile(Height, c(0.25, 0.75))  
#f)  
quantile(data$Weight)

0% 25% 50% 75% 100%   
45.07595 63.25895 77.81169 87.51476 97.94137

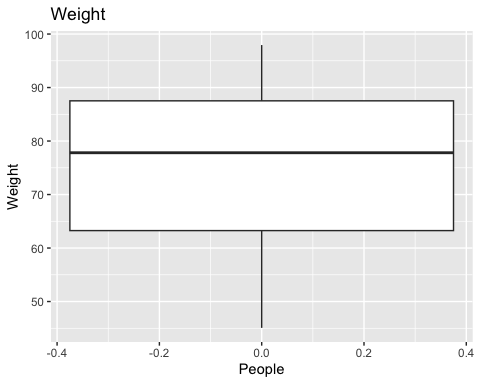
#g)  
boxplot(data$Height)



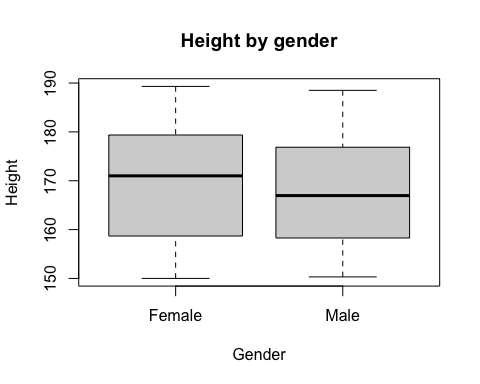
#h)  
ggplot(data = data, aes(, y = Height)) + geom\_boxplot() + ggtitle("Height") + labs(x = "People")



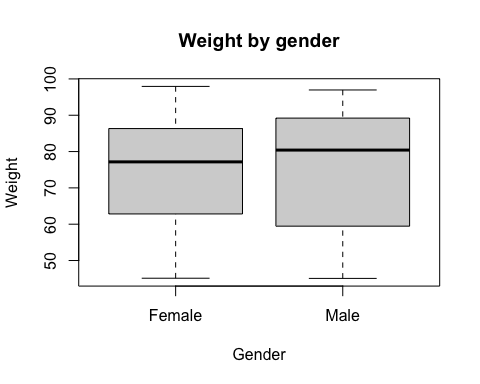
ggplot(data = data, aes(, y = Weight)) + geom\_boxplot() + ggtitle("Weight") + labs(x = "People")



#i)  
boxplot(Height ~ Gender, data = data, main = "Height by gender")

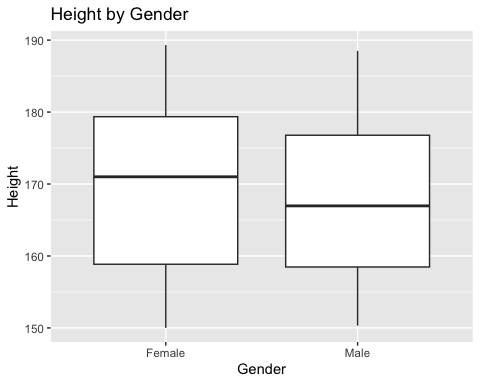


boxplot(Weight ~ Gender, data = data, main = "Weight by gender")



#j)  
ggplot(data = data, aes(x = as.factor(data$Gender), y = Height)) + geom\_boxplot() + ggtitle("Height by Gender") + labs(x = "Gender")

Warning: Use of `data$Gender` is discouraged.  
ℹ Use `Gender` instead.



Question 2

#a  
data <- iris  
sel <- data |>  
 select(iris.Species = Species)  
head(sel)

iris.Species  
1 setosa  
2 setosa  
3 setosa  
4 setosa  
5 setosa  
6 setosa

#b  
data <- data |>  
 arrange(Sepal.Length)  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
1 4.3 3.0 1.1 0.1 setosa  
2 4.4 2.9 1.4 0.2 setosa  
3 4.4 3.0 1.3 0.2 setosa  
4 4.4 3.2 1.3 0.2 setosa  
5 4.5 2.3 1.3 0.3 setosa  
6 4.6 3.1 1.5 0.2 setosa

#c  
data <- data |>   
 mutate(Ratio = Sepal.Width/Petal.Width)  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Species Ratio  
1 4.3 3.0 1.1 0.1 setosa 30.000000  
2 4.4 2.9 1.4 0.2 setosa 14.500000  
3 4.4 3.0 1.3 0.2 setosa 15.000000  
4 4.4 3.2 1.3 0.2 setosa 16.000000  
5 4.5 2.3 1.3 0.3 setosa 7.666667  
6 4.6 3.1 1.5 0.2 setosa 15.500000

#d  
data <- data |>   
 mutate(Ratio = Sepal.Width/Petal.Width) |>  
 relocate(Ratio, .before = Sepal.Length)  
head(data)

Ratio Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
1 30.000000 4.3 3.0 1.1 0.1 setosa  
2 14.500000 4.4 2.9 1.4 0.2 setosa  
3 15.000000 4.4 3.0 1.3 0.2 setosa  
4 16.000000 4.4 3.2 1.3 0.2 setosa  
5 7.666667 4.5 2.3 1.3 0.3 setosa  
6 15.500000 4.6 3.1 1.5 0.2 setosa

#e  
data <- data |>   
 mutate(Ratio = Sepal.Width/Petal.Width) |>  
 relocate(Ratio, .after = Petal.Width)  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Ratio Species  
1 4.3 3.0 1.1 0.1 30.000000 setosa  
2 4.4 2.9 1.4 0.2 14.500000 setosa  
3 4.4 3.0 1.3 0.2 15.000000 setosa  
4 4.4 3.2 1.3 0.2 16.000000 setosa  
5 4.5 2.3 1.3 0.3 7.666667 setosa  
6 4.6 3.1 1.5 0.2 15.500000 setosa

#f  
spec <- data |>  
 group\_by(Species) |>  
 summarize(mean\_p\_len = mean(Petal.Length))  
head(spec)

# A tibble: 3 × 2  
 Species mean\_p\_len  
 <fct> <dbl>  
1 setosa 1.46  
2 versicolor 4.26  
3 virginica 5.55

#g  
data <- data |>  
 mutate(Greater.half = data$Sepal.Width > (data$Sepal.Length/2))  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Ratio Species  
1 4.3 3.0 1.1 0.1 30.000000 setosa  
2 4.4 2.9 1.4 0.2 14.500000 setosa  
3 4.4 3.0 1.3 0.2 15.000000 setosa  
4 4.4 3.2 1.3 0.2 16.000000 setosa  
5 4.5 2.3 1.3 0.3 7.666667 setosa  
6 4.6 3.1 1.5 0.2 15.500000 setosa  
 Greater.half  
1 TRUE  
2 TRUE  
3 TRUE  
4 TRUE  
5 TRUE  
6 TRUE

#h  
data <- data |>   
 filter(Species == "setosa")  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Ratio Species  
1 4.3 3.0 1.1 0.1 30.000000 setosa  
2 4.4 2.9 1.4 0.2 14.500000 setosa  
3 4.4 3.0 1.3 0.2 15.000000 setosa  
4 4.4 3.2 1.3 0.2 16.000000 setosa  
5 4.5 2.3 1.3 0.3 7.666667 setosa  
6 4.6 3.1 1.5 0.2 15.500000 setosa  
 Greater.half  
1 TRUE  
2 TRUE  
3 TRUE  
4 TRUE  
5 TRUE  
6 TRUE

#i  
data <- data |>   
 rename(iris.Species = Species)  
head(data)

Sepal.Length Sepal.Width Petal.Length Petal.Width Ratio iris.Species  
1 4.3 3.0 1.1 0.1 30.000000 setosa  
2 4.4 2.9 1.4 0.2 14.500000 setosa  
3 4.4 3.0 1.3 0.2 15.000000 setosa  
4 4.4 3.2 1.3 0.2 16.000000 setosa  
5 4.5 2.3 1.3 0.3 7.666667 setosa  
6 4.6 3.1 1.5 0.2 15.500000 setosa  
 Greater.half  
1 TRUE  
2 TRUE  
3 TRUE  
4 TRUE  
5 TRUE  
6 TRUE