annand\_DSE5001\_assignment2

2023-05-19

# Import libraries

library(readr)  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ purrr 1.0.1  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ── 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(ggplot2)  
library(dplyr)  
library(psych)

##   
## Attaching package: 'psych'  
##   
## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

library(moments)

## CHAPTER 3

weight\_data <- read\_csv("C:\\Users\\janna\\OneDrive\\Documents\\DSE5001\\Week 2\\weight.csv")

## Rows: 6068 Columns: 8  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): gender  
## dbl (7): subjectid, height, height\_selfreport, weight, weight\_selfreport, ag...  
##   
## 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.

w1 <- select(weight\_data, subjectid, height, weight, age) # returns specific columns  
w1

## # A tibble: 6,068 × 4  
## subjectid height weight age  
## <dbl> <dbl> <dbl> <dbl>  
## 1 10027 178. 81.5 41  
## 2 10032 170. 72.6 35  
## 3 10033 174. 92.9 42  
## 4 10092 166. 79.4 31  
## 5 10093 191. 94.6 21  
## 6 10115 172 80.2 39  
## 7 10117 181 116. 32  
## 8 10237 185 95.4 23  
## 9 10242 178. 99.5 36  
## 10 10244 181. 70.2 23  
## # ℹ 6,058 more rows

w2 <- select(weight\_data, subjectid, height:age) # returns subjectid col and all cols between height and age (inclusive)  
w2

## # A tibble: 6,068 × 6  
## subjectid height height\_selfreport weight weight\_selfreport age  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 178. 180. 81.5 81.7 41  
## 2 10032 170. 173. 72.6 72.6 35  
## 3 10033 174. 173. 92.9 93.0 42  
## 4 10092 166. 168. 79.4 79.4 31  
## 5 10093 191. 196. 94.6 96.6 21  
## 6 10115 172 175. 80.2 79.4 39  
## 7 10117 181 183. 116. 113. 32  
## 8 10237 185 188. 95.4 95.7 23  
## 9 10242 178. 178. 99.5 99.8 36  
## 10 10244 181. 183. 70.2 72.6 23  
## # ℹ 6,058 more rows

w3 <- select(weight\_data, -weight\_selfreport) # return data frame without weight\_selfreport col  
w3

## # A tibble: 6,068 × 7  
## subjectid gender height height\_selfreport weight age race  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 41 1  
## 2 10032 Male 170. 173. 72.6 35 1  
## 3 10033 Male 174. 173. 92.9 42 2  
## 4 10092 Male 166. 168. 79.4 31 1  
## 5 10093 Male 191. 196. 94.6 21 2  
## 6 10115 Male 172 175. 80.2 39 1  
## 7 10117 Male 181 183. 116. 32 2  
## 8 10237 Male 185 188. 95.4 23 1  
## 9 10242 Male 178. 178. 99.5 36 1  
## 10 10244 Male 181. 183. 70.2 23 1  
## # ℹ 6,058 more rows

w4 <- select(weight\_data, starts\_with("w")) # returns all cols that start with w  
w4

## # A tibble: 6,068 × 2  
## weight weight\_selfreport  
## <dbl> <dbl>  
## 1 81.5 81.7  
## 2 72.6 72.6  
## 3 92.9 93.0  
## 4 79.4 79.4  
## 5 94.6 96.6  
## 6 80.2 79.4  
## 7 116. 113.   
## 8 95.4 95.7  
## 9 99.5 99.8  
## 10 70.2 72.6  
## # ℹ 6,058 more rows

w5 <- select(weight\_data, ends\_with("t")) # returns all cols that end in t  
w5

## # A tibble: 6,068 × 4  
## height height\_selfreport weight weight\_selfreport  
## <dbl> <dbl> <dbl> <dbl>  
## 1 178. 180. 81.5 81.7  
## 2 170. 173. 72.6 72.6  
## 3 174. 173. 92.9 93.0  
## 4 166. 168. 79.4 79.4  
## 5 191. 196. 94.6 96.6  
## 6 172 175. 80.2 79.4  
## 7 181 183. 116. 113.   
## 8 185 188. 95.4 95.7  
## 9 178. 178. 99.5 99.8  
## 10 181. 183. 70.2 72.6  
## # ℹ 6,058 more rows

w6 <- select(weight\_data, matches('^h|t$')) # returns cols that start with h or end in t  
w6

## # A tibble: 6,068 × 4  
## height height\_selfreport weight weight\_selfreport  
## <dbl> <dbl> <dbl> <dbl>  
## 1 178. 180. 81.5 81.7  
## 2 170. 173. 72.6 72.6  
## 3 174. 173. 92.9 93.0  
## 4 166. 168. 79.4 79.4  
## 5 191. 196. 94.6 96.6  
## 6 172 175. 80.2 79.4  
## 7 181 183. 116. 113.   
## 8 185 188. 95.4 95.7  
## 9 178. 178. 99.5 99.8  
## 10 181. 183. 70.2 72.6  
## # ℹ 6,058 more rows

w7 <- rename(weight\_data, gender\_identity=gender, ethnicity=race) # renames cols  
w7

## # A tibble: 6,068 × 8  
## subjectid gender\_identity height height\_selfreport weight weight\_selfreport  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 81.7  
## 2 10032 Male 170. 173. 72.6 72.6  
## 3 10033 Male 174. 173. 92.9 93.0  
## 4 10092 Male 166. 168. 79.4 79.4  
## 5 10093 Male 191. 196. 94.6 96.6  
## 6 10115 Male 172 175. 80.2 79.4  
## 7 10117 Male 181 183. 116. 113.   
## 8 10237 Male 185 188. 95.4 95.7  
## 9 10242 Male 178. 178. 99.5 99.8  
## 10 10244 Male 181. 183. 70.2 72.6  
## # ℹ 6,058 more rows  
## # ℹ 2 more variables: age <dbl>, ethnicity <dbl>

w8 <- slice(weight\_data, c(10,20,30,40,100)) # returns rows 10,20,30,40,and 100  
w8

## # A tibble: 5 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age race  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10244 Male 181. 183. 70.2 72.6 23 1  
## 2 10310 Male 179. 178. 90.8 91.7 32 1  
## 3 10361 Male 182. 185. 88.2 90.7 33 1  
## 4 10423 Male 174. 175. 88.1 87.6 26 1  
## 5 10701 Male 176. 180. 96.8 97.5 27 1

w9 <- slice(weight\_data, 3:140) # returns rows 3 through 140 inclusive  
w9

## # A tibble: 138 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10033 Male 174. 173. 92.9 93.0 42  
## 2 10092 Male 166. 168. 79.4 79.4 31  
## 3 10093 Male 191. 196. 94.6 96.6 21  
## 4 10115 Male 172 175. 80.2 79.4 39  
## 5 10117 Male 181 183. 116. 113. 32  
## 6 10237 Male 185 188. 95.4 95.7 23  
## 7 10242 Male 178. 178. 99.5 99.8 36  
## 8 10244 Male 181. 183. 70.2 72.6 23  
## 9 10246 Male 178 178. 88.2 88.5 32  
## 10 10260 Male 173. 175. 70.1 70.3 28  
## # ℹ 128 more rows  
## # ℹ 1 more variable: race <dbl>

w10 <- slice(weight\_data, -(2:40)) # returns all but rows 2 through 40 inclusive  
w10

## # A tibble: 6,029 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 81.7 41  
## 2 10427 Male 174. 178. 57.5 61.3 25  
## 3 10447 Male 168. 170. 87.6 85.3 44  
## 4 10451 Male 184. 185. 85 79.4 19  
## 5 10460 Male 173 173. 86.5 86.2 24  
## 6 10467 Male 176. 178. 95.8 97.5 33  
## 7 10471 Male 167. 168. 66.7 65.8 21  
## 8 10473 Male 174 175. 70.6 70.3 29  
## 9 10482 Male 171. 175. 84.7 83.9 31  
## 10 10495 Male 188. 190. 101. 101. 39  
## # ℹ 6,019 more rows  
## # ℹ 1 more variable: race <dbl>

w11 <- filter(weight\_data, gender == 'female', age < 40) # include observations of women under 40 years old  
w11

## # A tibble: 0 × 8  
## # ℹ 8 variables: subjectid <dbl>, gender <chr>, height <dbl>,  
## # height\_selfreport <dbl>, weight <dbl>, weight\_selfreport <dbl>, age <dbl>,  
## # race <dbl>

w12 <- filter(weight\_data, weight > 100 | weight < 75) # include observations that are over 100 kg or under 75 kg  
w12

## # A tibble: 3,092 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10032 Male 170. 173. 72.6 72.6 35  
## 2 10117 Male 181 183. 116. 113. 32  
## 3 10244 Male 181. 183. 70.2 72.6 23  
## 4 10260 Male 173. 175. 70.1 70.3 28  
## 5 10265 Male 181. 183. 104. 104. 36  
## 6 10272 Male 186. 190. 112. 113. 26  
## 7 10312 Male 178 180. 71.4 74.9 28  
## 8 10319 Male 168 168. 74.6 74.0 24  
## 9 10341 Male 163 163. 67.5 61.3 32  
## 10 10349 Male 182. 185. 107. 104. 30  
## # ℹ 3,082 more rows  
## # ℹ 1 more variable: race <dbl>

calculate\_bmi <- function(w,h){ w/((h/100)^2) }  
mutate(weight\_data, bmi = calculate\_bmi(weight,height)) # adds bmi col using cutsom function

## # A tibble: 6,068 × 9  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 81.7 41  
## 2 10032 Male 170. 173. 72.6 72.6 35  
## 3 10033 Male 174. 173. 92.9 93.0 42  
## 4 10092 Male 166. 168. 79.4 79.4 31  
## 5 10093 Male 191. 196. 94.6 96.6 21  
## 6 10115 Male 172 175. 80.2 79.4 39  
## 7 10117 Male 181 183. 116. 113. 32  
## 8 10237 Male 185 188. 95.4 95.7 23  
## 9 10242 Male 178. 178. 99.5 99.8 36  
## 10 10244 Male 181. 183. 70.2 72.6 23  
## # ℹ 6,058 more rows  
## # ℹ 2 more variables: race <dbl>, bmi <dbl>

weight\_data

## # A tibble: 6,068 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 81.7 41  
## 2 10032 Male 170. 173. 72.6 72.6 35  
## 3 10033 Male 174. 173. 92.9 93.0 42  
## 4 10092 Male 166. 168. 79.4 79.4 31  
## 5 10093 Male 191. 196. 94.6 96.6 21  
## 6 10115 Male 172 175. 80.2 79.4 39  
## 7 10117 Male 181 183. 116. 113. 32  
## 8 10237 Male 185 188. 95.4 95.7 23  
## 9 10242 Male 178. 178. 99.5 99.8 36  
## 10 10244 Male 181. 183. 70.2 72.6 23  
## # ℹ 6,058 more rows  
## # ℹ 1 more variable: race <dbl>

w13 <- arrange(weight\_data, height, weight) # arranges data by height and then weight in ascending order  
w13

## # A tibble: 6,068 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 20429 Female 141. 142. 42.6 45.4 27  
## 2 18454 Female 144. 145. 47.8 47.6 20  
## 3 20093 Female 144. 142. 58.5 59.0 29  
## 4 27070 Female 144. 145. 46.7 45.4 35  
## 5 10222 Female 145. 142. 48.8 48.5 23  
## 6 18333 Female 145. 145. 68.3 67.2 36  
## 7 15160 Female 145. 150. 73.6 68.1 22  
## 8 25764 Female 146. 147. 44.5 46.7 21  
## 9 18126 Female 146. 145. 51.9 53.5 31  
## 10 29284 Female 146. 147. 62.6 61.7 20  
## # ℹ 6,058 more rows  
## # ℹ 1 more variable: race <dbl>

w14 <- arrange(weight\_data, desc(height), desc(weight)) # arrranges data by height and then weight both in descending order  
w14

## # A tibble: 6,068 × 8  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 27557 Male 199. 198. 94.7 99.8 27  
## 2 27992 Male 198 201. 102. 102. 25  
## 3 23270 Male 198. 203. 92.2 95.3 25  
## 4 27316 Male 197. 198. 109. 107. 31  
## 5 20601 Male 197 198. 120. 123. 33  
## 6 11453 Male 197 198. 90.2 91.2 33  
## 7 14094 Male 197. 201. 89.7 93.0 28  
## 8 22118 Male 197. 203. 87.3 90.7 20  
## 9 29207 Male 196 198. 98.1 98.9 38  
## 10 10339 Male 196. 201. 91.5 97.5 26  
## # ℹ 6,058 more rows  
## # ℹ 1 more variable: race <dbl>

summarize(weight\_data,  
 mean\_height = mean(height, na.rm = T),  
 median\_height = mean(height, na.rm = T),  
 sd\_height = sd(height, na.rm = T))

## # A tibble: 1 × 3  
## mean\_height median\_height sd\_height  
## <dbl> <dbl> <dbl>  
## 1 171. 171. 9.00

weight\_by\_race <- group\_by(weight\_data, race)  
weight\_by\_race

## # A tibble: 6,068 × 8  
## # Groups: race [7]  
## subjectid gender height height\_selfreport weight weight\_selfreport age  
## <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 10027 Male 178. 180. 81.5 81.7 41  
## 2 10032 Male 170. 173. 72.6 72.6 35  
## 3 10033 Male 174. 173. 92.9 93.0 42  
## 4 10092 Male 166. 168. 79.4 79.4 31  
## 5 10093 Male 191. 196. 94.6 96.6 21  
## 6 10115 Male 172 175. 80.2 79.4 39  
## 7 10117 Male 181 183. 116. 113. 32  
## 8 10237 Male 185 188. 95.4 95.7 23  
## 9 10242 Male 178. 178. 99.5 99.8 36  
## 10 10244 Male 181. 183. 70.2 72.6 23  
## # ℹ 6,058 more rows  
## # ℹ 1 more variable: race <dbl>

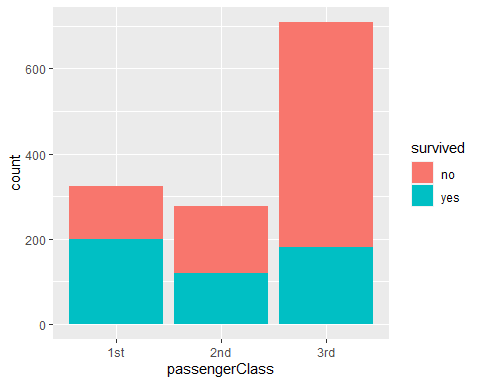
weight\_pivot <- pivot\_longer(weight\_data, cols = height:weight\_selfreport, names\_to = 'measure\_type', values\_to = 'cm')

## CHAPTER 4

titantic\_df <- read\_csv("C:\\Users\\janna\\OneDrive\\Documents\\DSE5001\\Week 2\\TitanicSurvival.csv")

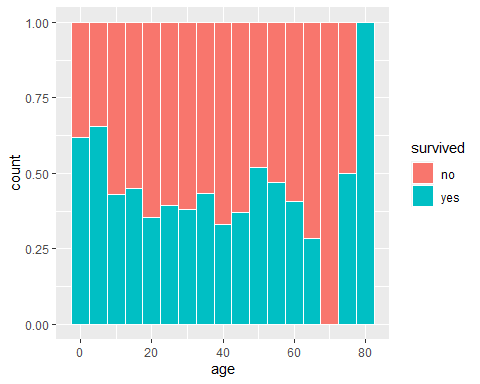
## New names:  
## Rows: 1309 Columns: 5  
## ── Column specification  
## ──────────────────────────────────────────────────────── Delimiter: "," chr  
## (4): ...1, survived, sex, passengerClass dbl (1): age  
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ  
## Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## • `` -> `...1`

# Create bar chart showing survival by passeneger class  
ggplot(titantic\_df,  
 mapping = aes(x = passengerClass, fill = survived)  
) + geom\_bar()



# Create histogram of distribution of ages in Titantic passengers  
ggplot(titantic\_df,  
 mapping = aes(x = age, fill = survived)  
) + geom\_histogram(binwidth = 5, colour = 'white',   
 position = 'fill')

## Warning: Removed 263 rows containing non-finite values (`stat\_bin()`).



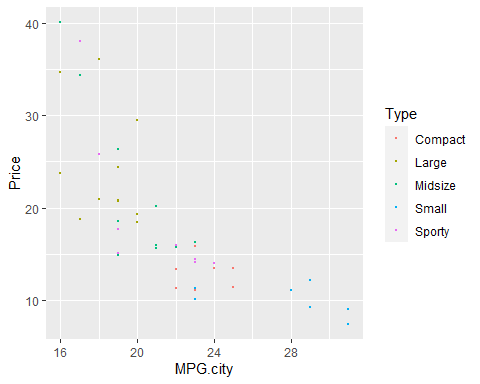
# Box plot of heights by gender  
ggplot(weight\_data,  
 mapping = aes(x = gender, y = height)  
) + geom\_boxplot(width = 0.25, outlier.shape = NA) + coord\_flip()



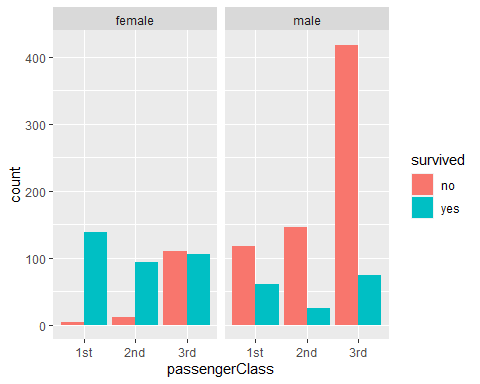
carprice\_df <- read.csv("C:\\Users\\janna\\OneDrive\\Documents\\DSE5001\\Week 2\\carprice.csv")  
  
# Scatter plot showing relationship between price and city mpg for cars that are not vans  
carprice\_filtered <- filter(carprice\_df, Type != 'Van')  
carprice\_filtered

## X Type Min.Price Price Max.Price Range.Price RoughRange gpm100 MPG.city  
## 1 6 Midsize 14.2 15.7 17.3 3.1 3.09 3.8 22  
## 2 7 Large 19.9 20.8 21.7 1.8 1.79 4.2 19  
## 3 8 Large 22.6 23.7 24.9 2.3 2.31 4.9 16  
## 4 9 Midsize 26.3 26.3 26.3 0.0 -0.01 4.3 19  
## 5 10 Large 33.0 34.7 36.3 3.3 3.30 4.9 16  
## 6 11 Midsize 37.5 40.1 42.7 5.2 5.18 4.9 16  
## 7 12 Compact 8.5 13.4 18.3 9.8 9.80 3.3 25  
## 8 13 Compact 11.4 11.4 11.4 0.0 -0.01 3.4 25  
## 9 14 Sporty 13.4 15.1 16.8 3.4 3.38 4.2 19  
## 10 15 Midsize 13.4 15.9 18.4 5.0 5.01 4.0 21  
## 11 18 Large 18.0 18.8 19.6 1.6 1.60 4.7 17  
## 12 19 Sporty 34.6 38.0 41.5 6.9 6.88 4.8 17  
## 13 20 Large 18.4 18.4 18.4 0.0 -0.01 4.2 20  
## 14 21 Compact 14.5 15.8 17.1 2.6 2.59 3.9 23  
## 15 22 Large 29.5 29.5 29.5 0.0 0.02 4.3 20  
## 16 23 Small 7.9 9.2 10.6 2.7 2.68 3.2 29  
## 17 24 Small 8.4 11.3 14.2 5.8 5.80 3.8 23  
## 18 25 Compact 11.9 13.3 14.7 2.8 2.81 4.1 22  
## 19 27 Midsize 14.8 15.6 16.4 1.6 1.60 4.2 21  
## 20 28 Sporty 18.5 25.8 33.1 14.6 14.60 4.8 18  
## 21 29 Small 7.9 12.2 16.5 8.6 8.60 3.2 29  
## 22 30 Large 17.5 19.3 21.2 3.7 3.69 4.2 20  
## 23 31 Small 6.9 7.4 7.9 1.0 1.00 3.1 31  
## 24 32 Small 8.4 10.1 11.9 3.5 3.49 3.8 23  
## 25 33 Compact 10.4 11.3 12.2 1.8 1.82 4.1 22  
## 26 34 Sporty 10.8 15.9 21.0 10.2 10.21 3.9 22  
## 27 35 Sporty 12.8 14.0 15.2 2.4 2.40 3.7 24  
## 28 37 Midsize 15.6 20.2 24.8 9.2 9.21 3.9 21  
## 29 38 Large 20.1 20.9 21.7 1.6 1.59 4.5 18  
## 30 51 Midsize 33.3 34.3 35.3 2.0 1.99 4.7 17  
## 31 52 Large 34.4 36.1 37.8 3.4 3.42 4.5 18  
## 32 60 Sporty 13.3 14.1 15.0 1.7 1.71 4.1 23  
## 33 61 Midsize 14.9 14.9 14.9 0.0 -0.02 4.4 19  
## 34 68 Compact 13.0 13.5 14.0 1.0 0.99 3.6 24  
## 35 69 Midsize 14.2 16.3 18.4 4.2 4.19 3.7 23  
## 36 71 Large 19.5 20.7 21.9 2.4 2.41 4.2 19  
## 37 72 Sporty 11.4 14.4 17.4 6.0 6.01 3.8 23  
## 38 73 Small 8.2 9.0 9.9 1.7 1.69 2.8 31  
## 39 74 Compact 9.4 11.1 12.8 3.4 3.39 3.7 23  
## 40 75 Sporty 14.0 17.7 21.4 7.4 7.40 4.2 19  
## 41 76 Midsize 15.4 18.5 21.6 6.2 6.19 4.3 19  
## 42 77 Large 19.4 24.4 29.4 10.0 10.00 4.2 19  
## 43 79 Small 9.2 11.1 12.9 3.7 3.70 3.0 28  
## MPG.highway  
## 1 31  
## 2 28  
## 3 25  
## 4 27  
## 5 25  
## 6 25  
## 7 36  
## 8 34  
## 9 28  
## 10 29  
## 11 26  
## 12 25  
## 13 28  
## 14 28  
## 15 26  
## 16 33  
## 17 29  
## 18 27  
## 19 27  
## 20 24  
## 21 33  
## 22 28  
## 23 33  
## 24 30  
## 25 27  
## 26 29  
## 27 30  
## 28 30  
## 29 26  
## 30 26  
## 31 26  
## 32 26  
## 33 26  
## 34 31  
## 35 31  
## 36 28  
## 37 30  
## 38 41  
## 39 31  
## 40 28  
## 41 27  
## 42 28  
## 43 38

ggplot(carprice\_filtered,  
 mapping = aes(x=MPG.city, y=Price, colour = Type)  
) + geom\_point(size = 0.5)

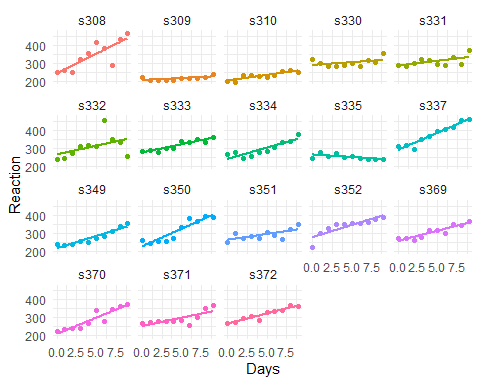


# Facet plot by gender of bar graphs for survival by passenger class  
ggplot(titantic\_df,  
 mapping = aes(x = passengerClass, fill = survived)  
) + geom\_bar(position = 'dodge') +  
 facet\_wrap(~sex)



# Create facet plot with subplots of relationship between Reaction and time in Days for each subject  
sleepstudy\_df <- read.csv("C:\\Users\\janna\\OneDrive\\Documents\\DSE5001\\Week 2\\sleepstudy.csv")  
  
ggplot(sleepstudy\_df,  
 mapping = aes(x = Days, y = Reaction, colour = Subject)  
) + geom\_point() +  
 geom\_smooth(method = 'lm', se = F) +  
 facet\_wrap(~Subject) +  
 theme\_minimal() +  
 theme(legend.position = 'none')

## `geom\_smooth()` using formula = 'y ~ x'



## CHAPTER 5

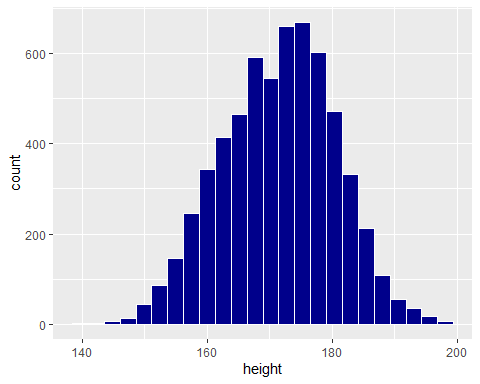
check\_excess\_kurtosis <- function(x) {  
 if(x - 3.0 > 0.2) {  
 'leptokurtic'  
} else if (x - 3.0 < -0.2) {  
 'platykurtic'  
} else {  
 'mesokurtic'  
}  
}  
  
unfiltered\_height\_summ <- summarize(weight\_data, mean = mean(height, na.rm=T), median = median(height, na.rm=T), sd = sd(height, na.rm=T), mad = mad(height, na.rm=T), iqr = IQR(height), skewness = psych::skew(height), kurtosis = moments::kurtosis(height)) %>%  
 mutate(excessKurtosis = check\_excess\_kurtosis(kurtosis))  
  
unfiltered\_height\_summ

## # A tibble: 1 × 8  
## mean median sd mad iqr skewness kurtosis excessKurtosis  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 171. 172. 9.00 9.34 12.7 -0.107 2.69 platykurtic

ggplot(weight\_data,  
 mapping = aes(x = '', y = height)  
) + geom\_boxplot() + coord\_flip()



ggplot(weight\_data,  
 mapping = aes(x = height),  
) + geom\_histogram(binwidth = 2.54, colour = 'white',  
 fill = 'darkblue')



heights\_by\_gender <- group\_by(weight\_data, gender)  
gender\_heights\_summ <- summarize(heights\_by\_gender, mean = mean(height, na.rm=T), median = median(height, na.rm=T), sd = sd(height, na.rm=T), mad = mad(height, na.rm=T), iqr = IQR(height), skewness = psych::skew(height), kurtosis = moments::kurtosis(height)) %>%  
 mutate(excessKurtosis = case\_when(  
 (kurtosis - 3.0) > 0.2 ~ 'leptokurtic',  
 (kurtosis - 3.0) < -0.2 ~ 'platykurtic',  
 (kurtosis - 3.0) < 0.2 & (kurtosis - 3.0) > -0.2 ~ 'mesokurtic'  
 ))  
  
gender\_heights\_summ

## # A tibble: 2 × 9  
## gender mean median sd mad iqr skewness kurtosis excessKurtosis  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>   
## 1 Female 163. 163. 6.42 6.23 8.57 0.0875 3.00 mesokurtic   
## 2 Male 176. 176. 6.86 6.82 9.20 0.111 3.07 mesokurtic

ggplot(weight\_data,  
 mapping = aes(x = gender, y = height)  
) + geom\_boxplot() + coord\_flip()



ggplot(weight\_data,  
 mapping = aes(x=height)  
) + geom\_histogram(binwidth = 2.54, colour = 'white') + facet\_wrap(~gender)

