

NAME : Hariprasad K K

REG NO : 19BCE7079

COURSE : Data_analytics

DATE : 11-11-2021

```
library(data.table)
```

```
url <- http://jse.amstat.org/datasets/fat.dat.txt
```

```
bodyfat <- fread(url, col.names = c("case", "brozek", "siri",  
"density", "age",  
"weight_lbs",  
"height_in", "bmi",  
"fat_free_weight", "neck_cm",  
"chest_cm", "abdomen_cm",  
"hip_cm", "thigh_cm",  
"knee_cm", "ankle_cm",  
"biceps_cm", "forearm_cm",  
"wrist_cm"))
```

```
> library(data.table)  
> url <- "http://jse.amstat.org/datasets/fat.dat.txt"  
>  
> bodyfat <- fread(url, col.names = c("case", "brozek", "siri",  
+                                     "density", "age",  
+                                     "weight_lbs",  
+                                     "height_in", "bmi",  
+                                     "fat_free_weight", "neck_cm",  
+                                     "chest_cm", "abdomen_cm",  
+                                     "hip_cm", "thigh_cm",  
+                                     "knee_cm", "ankle_cm",  
+                                     "biceps_cm", "forearm_cm",  
+                                     "wrist_cm"))  
trying URL 'http://jse.amstat.org/datasets/fat.dat.txt'  
Content type 'text/plain' length 37758 bytes (36 KB)  
downloaded 36 KB
```

```
str(bodyfat)
```

```
> str(bodyfat)
Classes 'data.table' and 'data.frame': 252 obs. of 19 variables:
 $ case      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ brozek    : num  12.6 6.9 24.6 10.9 27.8 20.6 19 12.8 5.1 12 ...
 $ siri      : num  12.3 6.1 25.3 10.4 28.7 20.9 19.2 12.4 4.1 11.7 ...
 $ density   : num  1.07 1.09 1.04 1.08 1.03 ...
 $ age       : int  23 22 22 26 24 24 26 25 25 23 ...
 $ weight_lbs : num  154 173 154 185 184 ...
 $ height_in  : num  67.8 72.2 66.2 72.2 71.2 ...
 $ bmi       : num  23.7 23.4 24.7 24.9 25.6 26.5 26.2 23.6 24.6 25.8 ...
 $ fat_free_weight: num  135 161 116 165 133 ...
 $ neck_cm    : num  36.2 38.5 34 37.4 34.4 39 36.4 37.8 38.1 42.1 ...
 $ chest_cm   : num  93.1 93.6 95.8 101.8 97.3 ...
 $ abdomen_cm : num  85.2 83 87.9 86.4 100 94.4 90.7 88.5 82.5 88.6 ...
 $ hip_cm     : num  94.5 98.7 99.2 101.2 101.9 ...
 $ thigh_cm   : num  59 58.7 59.6 60.1 63.2 66 58.4 60 62.9 63.1 ...
 $ knee_cm    : num  37.3 37.3 38.9 37.3 42.2 42 38.3 39.4 38.3 41.7 ...
 $ ankle_cm   : num  21.9 23.4 24 22.8 24 25.6 22.9 23.2 23.8 25 ...
 $ biceps_cm  : num  32 30.5 28.8 32.4 32.2 35.7 31.9 30.5 35.9 35.6 ...
 $ forearm_cm : num  27.4 28.9 25.2 29.4 27.7 30.6 27.8 29 31.1 30 ...
 $ wrist_cm   : num  17.1 18.2 16.6 18.2 17.7 18.8 17.7 18.8 18.2 19.2 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
cor(bodyfat$age, bodyfat$brozek, method="pearson")
```

```
cor.test(bodyfat$age, bodyfat$weight_lbs, method="pearson")
```

```
> cor(bodyfat$age, bodyfat$brozek, method="pearson")
[1] 0.2891735
>
> cor.test(bodyfat$age, bodyfat$weight_lbs, method="pearson")

Pearson's product-moment correlation

data: bodyfat$age and bodyfat$weight_lbs
t = -0.20155, df = 250, p-value = 0.8404
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.1361046  0.1110016
sample estimates:
cor
-0.01274609
```

```
cor(bodyfat$age, bodyfat$brozek, method="spearman")
```

```
cor.test(bodyfat$age, bodyfat$brozek, method="spearman")
```

```
> cor(bodyfat$age, bodyfat$brozek, method="spearman")
[1] 0.273383
>
> cor.test(bodyfat$age, bodyfat$brozek, method="spearman")

Spearman's rank correlation rho

data: bodyfat$age and bodyfat$brozek
S = 1937979, p-value = 1.071e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.273383
```

```
install.packages("ggplot2")
```

```
library(ggplot2)
```

```
install.packages("plotly")
```

```
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/HP/OneDrive/Documents/R/win-library/4.1'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ggplot2_3.3.5.zip'
Content type 'application/zip' length 4130482 bytes (3.9 MB)
downloaded 3.9 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\HP\AppData\Local\Temp\RtmpEZJaJ5\downloaded_packages
> library(ggplot2)
>
> install.packages("plotly")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/HP/OneDrive/Documents/R/win-library/4.1'
(as 'lib' is unspecified)
also installing the dependencies 'later', 'htmlwidgets', 'lazyeval', 'crosstalk', 'promise:
```

```
library(plotly)
```

```
p <- ggplot(data = bodyfat, aes(x = density, y = brozek,
color = case)) +
```

```
geom_point() +
```

```
theme_bw()
```

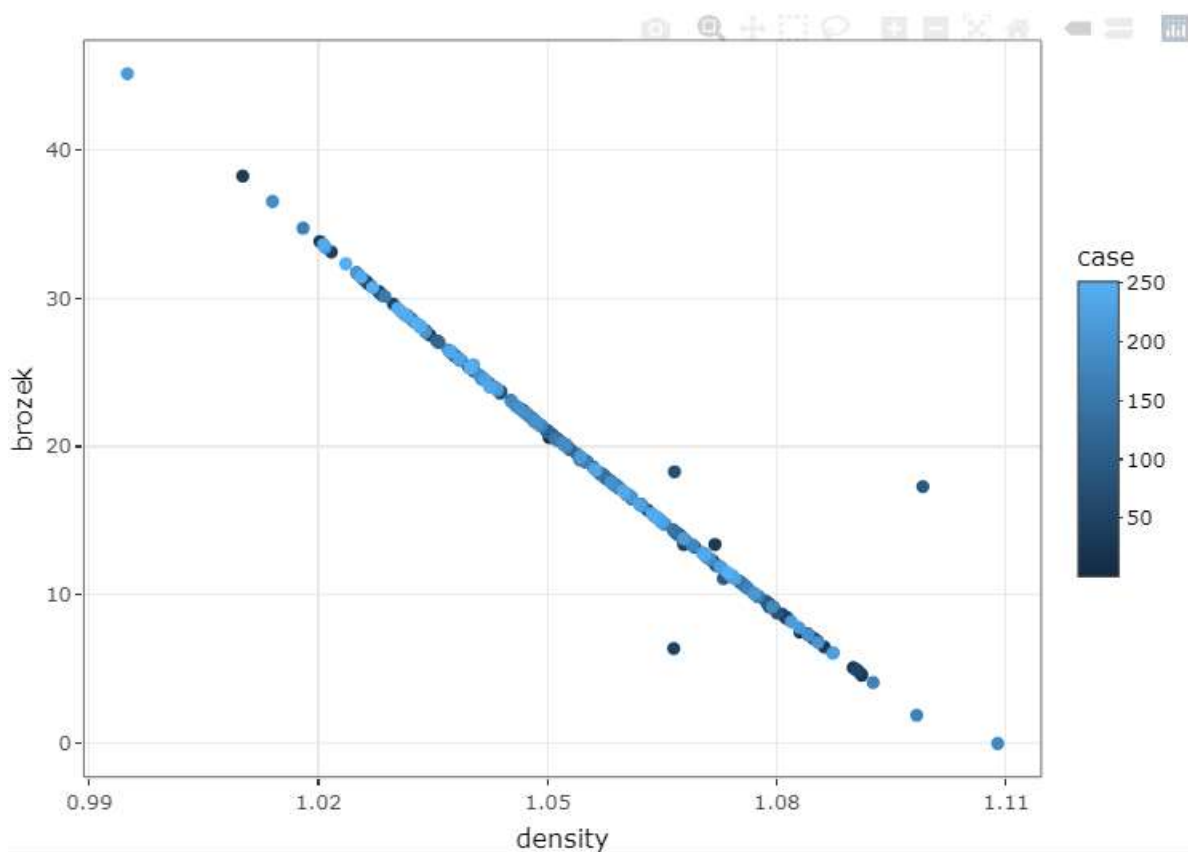
```
g <- ggplotly(p)
```

```
g
```

```

> library(plotly)
Attaching package: 'plotly'
The following object is masked from 'package:ggplot2':
  last_plot
The following object is masked from 'package:stats':
  filter
The following object is masked from 'package:graphics':
  layout
>
> p <- ggplot(data = bodyfat, aes(x = density, y = brozek,
+                               color = case)) +
+   geom_point() +
+   theme_bw()
>
> g <- ggplotly(p)
>
> g

```



```
library(dplyr)
```

```
bodyfat[c(48, 76, 96, 42, 182),
c("density", "brozek", "siri", "height_in")]
```

```

> library(dplyr)

Attaching package: 'dplyr'

The following objects are masked from 'package:data.table':
  between, first, last

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

> bodyfat[c(48, 76, 96, 42, 182),
+         c("density", "brozek", "siri", "height_in")]
  density brozek siri height_in
1:  1.0665   6.4  5.6    71.25
2:  1.0666  18.3 18.5     67.50
3:  1.0991  17.3 17.4     77.75
4:  1.0250  31.7 32.9     29.50
5:  1.1089   0.0  0.0     68.00
> |

```

```
bodyfat$density[c(48, 76, 96)] <- c(1.0865, 1.0566, 1.0591)
```

```
bodyfat <- bodyfat %>%
```

```
mutate(siri_C = round(495/density - 450, 1),
```

```
brozek_C = round(457/density - 414.2, 1),
```

```
bmi_C = round((weight_lbs*0.453592) /
(height_in*2.54/100)^2, 1))
```

```
bodyfat[c(48, 76, 96, 42, 182), c("density", "brozek",
"brozek_C", "siri_C",
"siri")]
```

```

> bodyfat$density[c(48, 76, 96)] <- c(1.0865, 1.0566, 1.0591)
>
> bodyfat <- bodyfat %>%
+ mutate(siri_C = round(495/density - 450, 1),
+        brozek_C = round(457/density - 414.2, 1),
+        bmi_C = round((weight_lbs*0.453592) /
+                        (height_in*2.54/100)^2, 1))
>
>
> bodyfat[c(48, 76, 96, 42, 182), c("density", "brozek",
+                                   "brozek_C", "siri_C",
+                                   "siri")]
  density brozek brozek_C siri_C siri
1:  1.0865   6.4   6.4    5.6   5.6
2:  1.0566  18.3  18.3   18.5  18.5
3:  1.0591  17.3  17.3   17.4  17.4
4:  1.0250  31.7  31.7   32.9  32.9
5:  1.1089   0.0  -2.1  -3.6   0.0
> |

```

```
bodyfat[c(48, 76, 96, 42, 182), c("density", "bmi_C", "bmi",
"height_in", "weight_lbs")]
```

```
> bodyfat[c(48, 76, 96, 42, 182), c("density", "bmi_C", "bmi",
+ "height_in", "weight_lbs")]
  density bmi_C  bmi height_in weight_lbs
1:  1.0865  20.6 20.6    71.25    148.50
2:  1.0566  22.9 22.9    67.50    148.25
3:  1.0591  26.1 26.1    77.75    224.50
4:  1.0250 165.6 29.9    29.50    205.00
5:  1.1089  18.0 18.1    68.00    118.50
> |
```

```
weight_k <- 205 * 0.453592
```

```
height_m <- sqrt(weight_k / 29.9)
```

```
height_m height_in <- height_m*100 / 2.54
```

```
height_in
```

```
bodyfat$
```

```
> weight_k <- 205 * 0.453592
> height_m <- sqrt(weight_k / 29.9)
> height_m
[1] 1.763494
>
> height_in <- height_m*100 / 2.54
> height_in
[1] 69.4289
> |
```

```
bodyfat$density[c(48, 76, 96)] <- c(1.0865, 1.0566, 1.0591)
```

```
bodyfat$height_in[42] <- 69.5
```

```
bodyfat <- bodyfat %>%
```

```
mutate(siri_C = round(495/density - 450, 1),
```

```
brozek_C = round(457/density - 414.2, 1),
```

```
bmi_C = round( (weight_lbs*0.453592) /
```

```
(height_in*2.54/100)^2, 1) )
```

```
bodyfatCH <- subset(bodyfat, abs(siri_C - siri) > 0.11 |
```

```
abs(brozek - brozek_C) > 0.11)
```

```
bodyfatCH[, c("siri", "siri_C", "brozek", "brozek_C",
"density")]
```

```

> bodyfat$density[c(48, 76, 96)] <- c(1.0865, 1.0566, 1.0591)
> bodyfat$height_in[42] <- 69.5
>
> bodyfat <- bodyfat %>%
+   mutate(siri_C = round(495/density - 450, 1),
+          brozek_C = round(457/density - 414.2, 1),
+          bmi_C = round( (weight_lbs*0.453592) /
+                        (height_in*2.54/100)^2, 1) )
> bodyfatCH <- subset(bodyfat, abs(siri_C - siri) > 0.11 |
+                    abs(brozek - brozek_C) > 0.11)
> bodyfatCH[, c("siri", "siri_C", "brozek", "brozek_C", "density")]
  siri siri_C brozek brozek_C density
1: 20.9  21.3  20.6    21.0  1.0502
2:  7.1   7.1   7.5     7.8  1.0830
3: 11.8  11.8  13.4    12.1  1.0719
4: 13.6  13.6  13.4    13.8  1.0678
5: 11.3  11.3  11.1    11.7  1.0730
6: 19.6  19.6  19.1    19.3  1.0542
7: 34.3  36.2  34.7    34.7  1.0180
8:  0.0  -3.6   0.0    -2.1  1.1089
9: 23.6  23.1  22.6    22.6  1.0462
10: 25.8  25.8  25.5    25.1  1.0403
11: 24.8  24.9  24.0    24.2  1.0424
> |

```

```
bodyfat$brozek[11] <- 7.8
```

```
bodyfat$brozek[33] <- 12.1
```

```
bodyfat$brozek[49] <- 13.8
```

```
bodyfat$brozek[98] <- 11.7
```

```
bodyfat$brozek[152] <- 19.3
```

```
bodyfat$brozek[235] <- 25.1
```

```
bodyfat$siri[169] <- 36.2
```

```
bodyfat$siri[200] <- 23.1
```

```
bodyfat$density[6] <- 1.0512
```

```
bodyfat <- bodyfat %>%
```

```
mutate(siri_C = round(495/density - 450, 1),
```

```
brozek_C = round(457/density - 414.2, 1),
```

```
bmi_C = round( (weight_lbs*0.453592) /
```

```
(height_in*2.54/100)^2, 1),
```

```
fat_free_weight_B = round((1 - brozek_C/100) *
```

```
weight_lbs ,1)
```

```
)
```

```
bodyfatCH2 <- subset(bodyfat, abs(siri_C - siri) > 0.11 |
```

```
abs(brozek - brozek_C) > 0.11)
```

```
bodyfatCH2[,c("siri", "siri_C", "brozek", "brozek_C",
"density")]
```

```
> bodyfat$brozek[11] <- 7.8
> bodyfat$brozek[33] <- 12.1
> bodyfat$brozek[49] <- 13.8
> bodyfat$brozek[98] <- 11.7
> bodyfat$brozek[152] <- 19.3
> bodyfat$brozek[235] <- 25.1
> bodyfat$siri[169] <- 36.2
> bodyfat$siri[200] <- 23.1
> bodyfat$density[6] <- 1.0512
> bodyfat <- bodyfat %>%
+   mutate(siri_C = round(495/density - 450, 1),
+           brozek_C = round(457/density - 414.2, 1),
+           bmi_C = round( (weight_lbs*0.453592) /
+                           (height_in*2.54/100)^2, 1),
+           fat_free_weight_B = round((1 - brozek_C/100) *
+                                     weight_lbs, 1)
+   )
> bodyfatCH2 <- subset(bodyfat, abs(siri_C - siri) > 0.11 |
+   abs(brozek - brozek_C) > 0.11)
> bodyfatCH2[,c("siri", "siri_C", "brozek", "brozek_C", "density")]
  siri siri_C brozek brozek_C density
1:  0.0   -3.6     0    -2.1  1.1089
2: 24.8   24.9    24    24.2  1.0424
> |
```

```
bodyfat[which(abs(bodyfat$fat_free_weight -
bodyfat$fat_free_weight_B) > 0.101),
c("fat_free_weight", "fat_free_weight_B")]
```

```
> bodyfat[which(abs(bodyfat$fat_free_weight -
+   bodyfat$fat_free_weight_B) > 0.101),
+   c("fat_free_weight", "fat_free_weight_B")]
  fat_free_weight fat_free_weight_B
1:         172.3         171.7
2:         142.5         147.7
3:         117.6         117.0
4:         127.8         128.0
5:         125.9         125.7
6:         151.2         151.7
7:         168.4         167.8
8:         159.3         159.0
9:         149.3         149.0
10:        141.7         141.4
11:        118.5         121.0
12:        151.3         133.8
13:        117.5         118.2
> |
```

```
sum(abs(bodyfat$fat_free_weight -
bodyfat$fat_free_weight_B) > 0.101)
```

```
> sum(abs(bodyfat$fat_free_weight -
+   bodyfat$fat_free_weight_B) > 0.101)
[1] 13
> |
```

```
sum(bodyfat$fat_free_weight != bodyfat$fat_free_weight_B)
```



```
> sum(bodyfat$fat_free_weight != bodyfat$fat_free_weight_B)
[1] 130
> |
```

```
bodyfat <- bodyfat[-182, ]
sum(bodyfat$fat_free_weight != bodyfat$fat_free_weight_B)
```

```
> bodyfat <- bodyfat[-182, ]
> sum(bodyfat$fat_free_weight != bodyfat$fat_free_weight_B)
[1] 129
>
>
```

```
sum(bodyfat$siri != bodyfat$siri_C)
```

```
> sum(bodyfat$siri != bodyfat$siri_C)
[1] 29
>
```

```
sum(bodyfat$bmi != bodyfat$bmi_C)
```

```
> sum(bodyfat$bmi != bodyfat$bmi_C)
[1] 98
```

```
sum(bodyfat$brozek != bodyfat$brozek_C)
```

```
> sum(bodyfat$brozek != bodyfat$brozek_C)
[1] 39
```

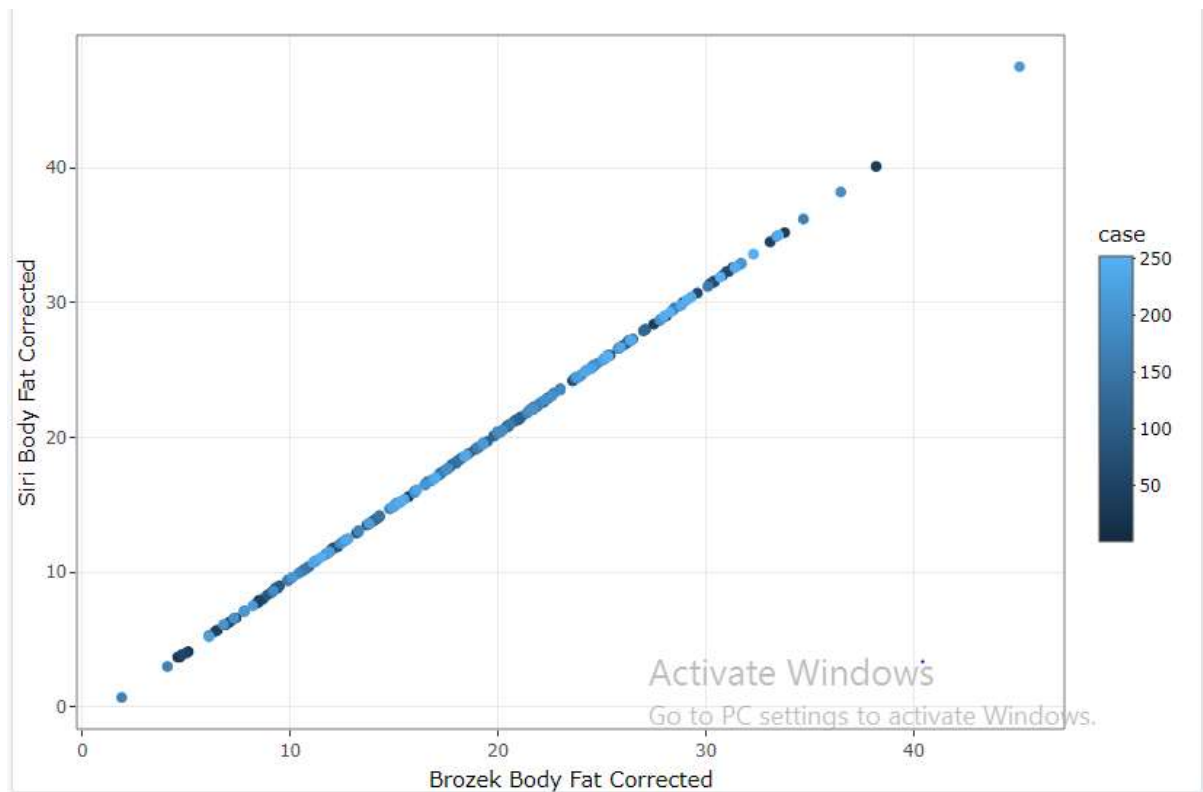
```
gp <- ggplot(data = bodyfat, aes(x = brozek_C, y = siri_C, color =
case)) +
```

```
geom_point() +
```

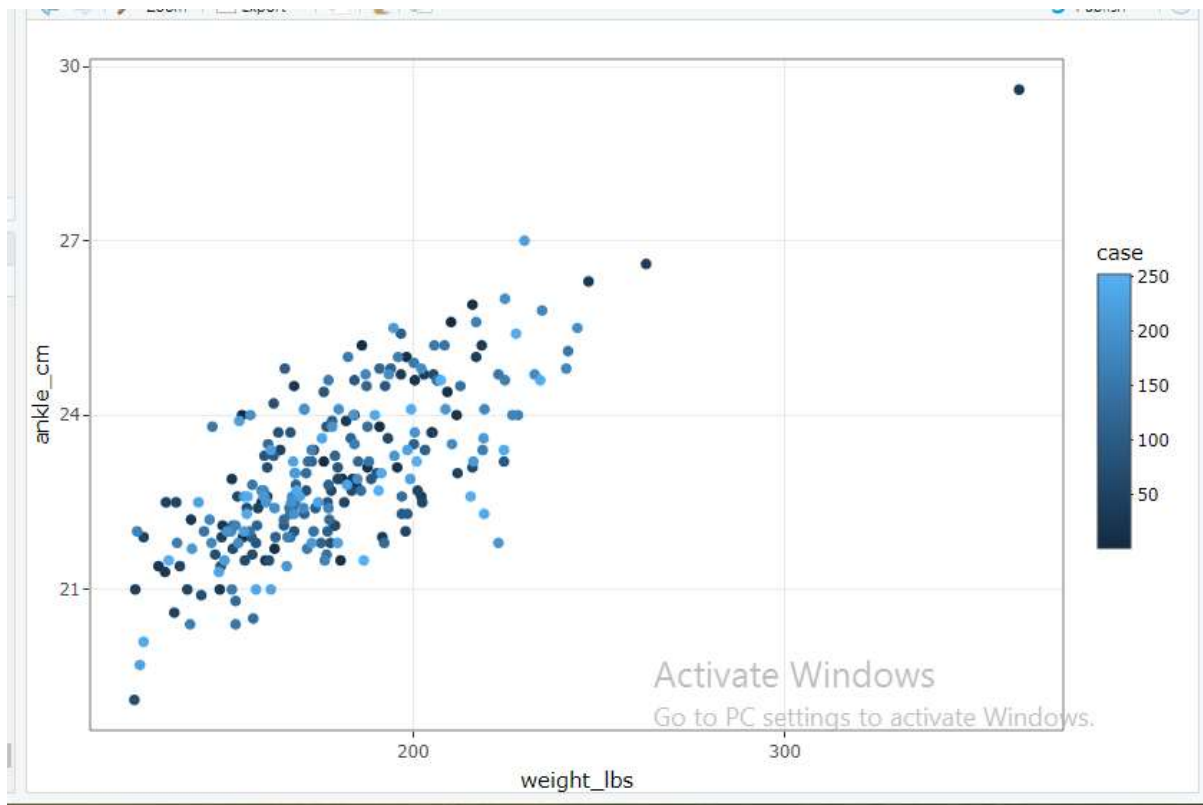
```
theme_bw() +
```

```
labs(x = "Brozek Body Fat Corrected", y = "Siri Body Fat Corrected")
```

```
ggplotly(gp)
```



```
p <- ggplot(data = bodyfat,  
aes(y = ankle_cm, x = weight_lbs, color = case)) +  
geom_point() +  
theme_bw()  
ggplotly(p)
```



```
bodyfat$ankle_cm[31] <- 23.9
```

```
bodyfat$ankle_cm[86] <- 23.7
```

```
bodyfat$forearm_cm[159] <- 24.9
```

```
bodyfatClean <- bodyfat[, -c(1, 2, 3, 4, 8, 9, 20, 23)]
```

```
> # Fixing likely typos / data entry errors
> bodyfat$ankle_cm[31] <- 23.9
> bodyfat$ankle_cm[86] <- 23.7
> bodyfat$forearm_cm[159] <- 24.9
>
> bodyfatClean <- bodyfat[, -c(1, 2, 3, 4, 8, 9, 20, 23)]
> names(bodyfatClean)
[1] "age"      "weight_lbs" "height_in"  "neck_cm"   "chest_cm"  "abdomen_cm" "hip_cm"    "thigh_cm"  "knee_cm"
[10] "ankle_cm" "biceps_cm"  "forearm_cm" "wrist_cm"  "brozek_c"  "bmi_c"      "hip_cm"    "thigh_cm"  "knee_cm"
```

```
bodyfatClean <- bodyfatClean %>%
```

```
mutate(age_sq = age^2, abdomen_wrist = abdomen_cm - wrist_cm,
```

```
am = (weight_lbs/0.453592)^1.2/(height_in*2.54/100)^3.3)
```

```
write.csv(bodyfatClean, "./bodyfatClean.csv",
```

```
row.names = FALSE)
```

```
>
> bodyfatClean <- bodyfatClean %>%
+   mutate(age_sq = age^2, abdomen_wrist = abdomen_cm - wrist_cm,
+          am = (weight_lbs/0.453592)^1.2/(height_in*2.54/100)^3.3)
> write.csv(bodyfatClean, "./bodyfatClean.csv",
+           row.names = FALSE)
> |
```