wi25\_lab4.R

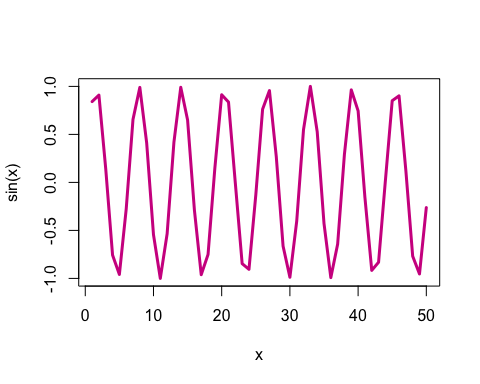
user

2025-01-16

x <- 1:50  
plot(x,sin(x))



# make plot look nicer  
plot(x, sin(x), col="violetred", type="l", lwd=3)



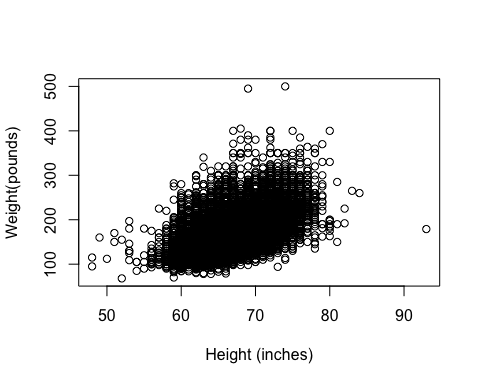
# load in cdc dataset  
source("http://thegrantlab.org/misc/cdc.R")  
  
#check first and last rows  
head(cdc$height)

## [1] 70 64 60 66 61 64

tail(cdc$weight, n=20)

## [1] 195 210 171 190 180 120 140 200 230 230 195 210 180 165 224 215 200 216 165  
## [20] 170

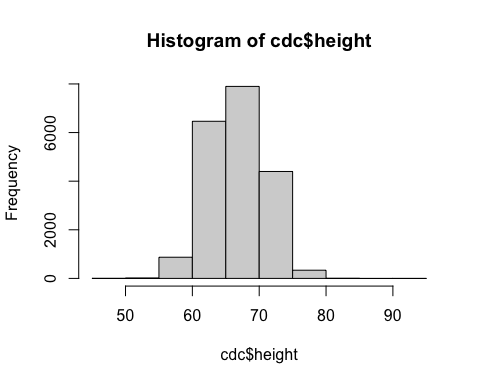
# creates plot of height vs weight  
plot(cdc$height, cdc$weight, type="p", xlab="Height (inches)",   
 ylab="Weight(pounds)")



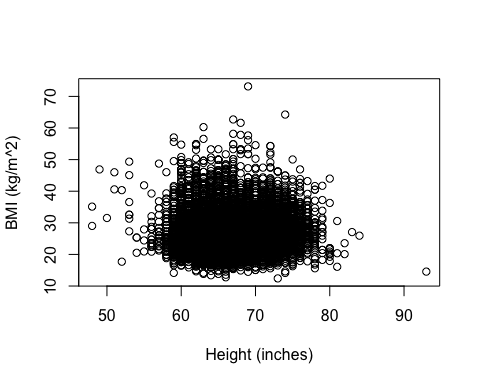
# check correlation of the height vs weight datapoints  
cor(cdc$height, cdc$weight)

## [1] 0.5553222

# generate histogram of the heights  
hist(cdc$height)



height\_m <- cdc$height \* 0.0254  
weight\_kg <- cdc$weight \* 0.454  
bmi <- (weight\_kg)/(height\_m^2)  
plot(cdc$height, bmi, xlab="Height (inches)", ylab="BMI (kg/m^2)")



cor(cdc$height, bmi)

## [1] 0.03251694

head(bmi >= 30, 100)

## [1] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [25] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE  
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE FALSE  
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE TRUE FALSE FALSE  
## [61] FALSE FALSE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [73] FALSE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  
## [85] FALSE FALSE FALSE FALSE TRUE FALSE TRUE FALSE FALSE FALSE FALSE FALSE  
## [97] TRUE FALSE FALSE FALSE

eg <- c(TRUE, TRUE, FALSE, FALSE)  
sum(eg)

## [1] 2

sum(bmi >= 30)

## [1] 3897

sum(bmi >= 30)/length(bmi)

## [1] 0.19485

(sum(bmi >= 30)/length(bmi)) \* 100

## [1] 19.485

round( (sum(bmi >= 30)/length(bmi)) \* 100, 1)

## [1] 19.5

cdc[567,6]

## [1] 160

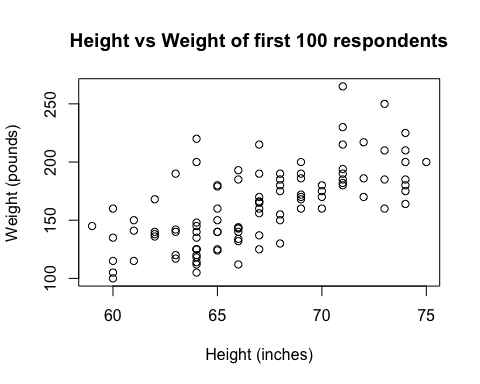
cdc[1:10, 6]

## [1] 175 125 105 132 150 114 194 170 150 180

cdc[1:10, ]

## genhlth exerany hlthplan smoke100 height weight wtdesire age gender  
## 1 good 0 1 0 70 175 175 77 m  
## 2 good 0 1 1 64 125 115 33 f  
## 3 good 1 1 1 60 105 105 49 f  
## 4 good 1 1 0 66 132 124 42 f  
## 5 very good 0 1 0 61 150 130 55 f  
## 6 very good 1 1 0 64 114 114 55 f  
## 7 very good 1 1 0 71 194 185 31 m  
## 8 very good 0 1 0 67 170 160 45 m  
## 9 good 0 1 1 65 150 130 27 f  
## 10 good 1 1 0 70 180 170 44 m

plot(cdc[1:100, "height"], cdc[1:100, "weight"], xlab="Height (inches)",   
 ylab="Weight (pounds)", main="Height vs Weight of first 100 respondents")



head(cdc, n=20)

## genhlth exerany hlthplan smoke100 height weight wtdesire age gender  
## 1 good 0 1 0 70 175 175 77 m  
## 2 good 0 1 1 64 125 115 33 f  
## 3 good 1 1 1 60 105 105 49 f  
## 4 good 1 1 0 66 132 124 42 f  
## 5 very good 0 1 0 61 150 130 55 f  
## 6 very good 1 1 0 64 114 114 55 f  
## 7 very good 1 1 0 71 194 185 31 m  
## 8 very good 0 1 0 67 170 160 45 m  
## 9 good 0 1 1 65 150 130 27 f  
## 10 good 1 1 0 70 180 170 44 m  
## 11 excellent 1 1 1 69 186 175 46 m  
## 12 fair 1 1 1 69 168 148 62 m  
## 13 excellent 1 0 1 66 185 220 21 m  
## 14 excellent 1 1 1 70 170 170 69 m  
## 15 fair 1 0 0 69 170 170 23 m  
## 16 good 1 1 1 73 185 175 79 m  
## 17 good 0 0 1 67 156 150 47 m  
## 18 fair 0 1 1 71 185 185 76 m  
## 19 good 1 1 1 75 200 190 43 m  
## 20 very good 1 1 0 67 125 120 33 f

# test approach using datasets...  
add\_bmi\_df <- cdc  
add\_bmi\_df$bmi <- bmi  
obese\_only\_df <- add\_bmi\_df[add\_bmi\_df$bmi >= 30,]  
nrow(obese\_only\_df[obese\_only\_df$gender=="m",])

## [1] 1961

# test Professor's tip  
gender\_vector <- cdc$gender  
obese <- subset(gender\_vector, bmi>=30)  
table(obese)

## obese  
## m f   
## 1961 1936