HW 1 087 F21

Lucía Carrera

9/10/2021

```
knitr::opts_chunk$set(echo = TRUE)
```

Part 1

Enter vectors:

```
#   Enter the vector (variable) height with the values
#    72, 70, 64, 68, 63, 60, 73, 62, 66
height <- c(72, 70, 64, 68, 63, 60, 73, 62, 66)

#   Enter the vector weight with the values
#    195, 200, 120, 165, missing, 100, 210, 140, 145
weight <- c(195, 200, 120, 165, NA, 100, 210, 140, 145)

#   Enter the vector instate with the values
#   In, In, Out, Out, Out, Out, In, Out, In
instate <- c('In', 'In', 'Out', 'Out', 'Out', 'Out', 'In', 'Out', 'In')

#   Use a shortcut to create a vector called ID
#   with the integers from 1 to 9.
ID <-1:9

#  Make a new vector, called bmi, that contains
#  body mass index, using the height and
#  weight vectors that you entered.
#  Here is the equation:
#      703.07 x W / (H squared)
#  You'll need to make sure it is written in
#  a form that R understands.
bmi <- 703.07 * weight / (height**2)
```

```
# print each vector below
height
```

```
## [1] 72 70 64 68 63 60 73 62 66
```

```
weight
```

```
## [1] 195 200 120 165  NA 100 210 140 145
```

```
instate
```

```
## [1] "In"  "In"  "Out" "Out" "Out" "Out" "In"  "Out" "In"
```

```
ID
```

```
## [1] 1 2 3 4 5 6 7 8 9
```

```
bmi
```

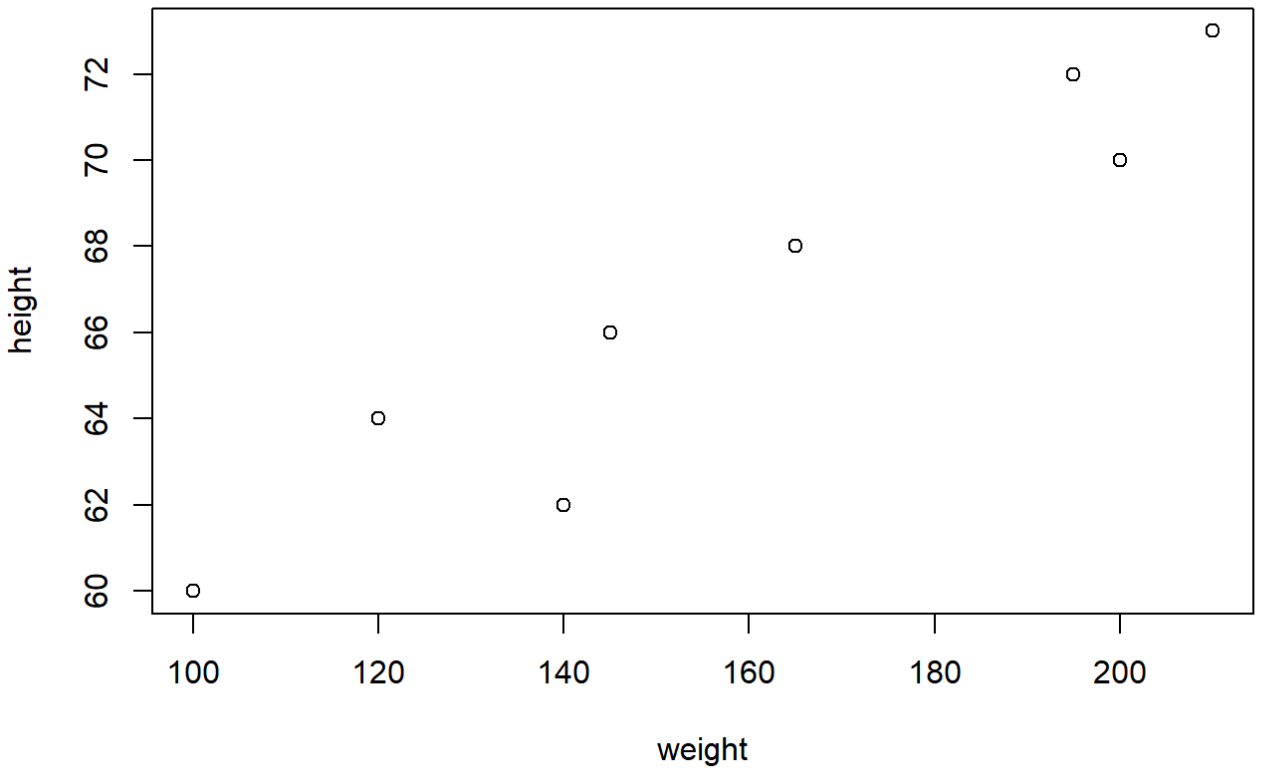
```
## [1] 26.44650 28.69673 20.59775 25.08792       NA 19.52972 27.70589 25.60609
## [9] 23.40339
```

Make a plot

Describe the relationship here: strong/weak, positive/negative. What does this mean about the relationship between height and weight?

I would describe the relationship between the points in this scatterplot as moderate or weak, with a positive and linear association

```
#   Create a scatterplot of weight by height
#   (i.e., height on the x axis)
plot(x = weight, y = height)
```



Make a data frame

```
#  Create a data frame, called students, that includes all of the vectors you creat
ed above.  Print it below.
students <- data.frame(height,weight,instate,ID,bmi)
```

Part 2

Read in a data file

```
#   Download the data file surveyA_F21.csv, then put it in your class folder,  and
read it using the function read.csv. Be sure you are in your project, and you have
saved your homework Rmd and data in the directory that you see in the lower right p
ane.  If there are all true, you can just put the file name of the data in the quot
es:
survey <- read.csv("surveyA_F21.csv")

#   Run the following:
head(survey) #first 6 entries of the dataframe
```

```
##   Response_id Height Weight Earnings Relationship GPA
## 1     1166955     65    160    15000          Yes 3.5
## 2     1167010     65    120    16000          Yes 3.0
## 3     1167023     69    130     6000           No 3.7
## 4     1167071     62    125     3000          Yes  NA
## 5     1167545     65    120     3000           No  NA
## 6     1168285     71    165     3500           No  NA
```

```
dim(survey)
```

```
## [1] 188   6
```

```
names(survey)
```

```
## [1] "Response_id"  "Height"       "Weight"       "Earnings"     "Relationship"
## [6] "GPA"
```

```
summary(survey)
```

```
##   Response_id          Height          Weight         Earnings
##  Min.   :1166955   Min.   :59.00   Min.   : 99.2   Min.   :     0
##  1st Qu.:1170956   1st Qu.:66.00   1st Qu.:135.0   1st Qu.:  2375
##  Median :1171741   Median :68.50   Median :150.0   Median :  5000
##  Mean   :1171346   Mean   :68.44   Mean   :154.8   Mean   :  8266
##  3rd Qu.:1172212   3rd Qu.:71.00   3rd Qu.:170.0   3rd Qu.:  9000
##  Max.   :1173323   Max.   :79.00   Max.   :260.0   Max.   :120000
##                                    NA's   :16
##  Relationship            GPA
##  Length:188         Min.   :1.900
##  Class :character   1st Qu.:3.220
##  Mode  :character   Median :3.585
##                     Mean   :3.456
##                     3rd Qu.:3.800
##                     Max.   :4.000
##                     NA's   :42
```

Stats from Survey data file

```
#   Find the mean and median of the variables,
#   earnings and GPA. Use the functions
#    mean() and median(), and make sure
#   the values match those in the summary above.
mean(survey$Earnings)
```

```
## [1] 8265.729
```

```
mean(survey$GPA, na.rm = TRUE)
```

```
## [1] 3.456164
```

```
median(survey$Earnings)
```

```
## [1] 5000
```

```
median(survey$GPA, na.rm = TRUE)
```

```
## [1] 3.585
```

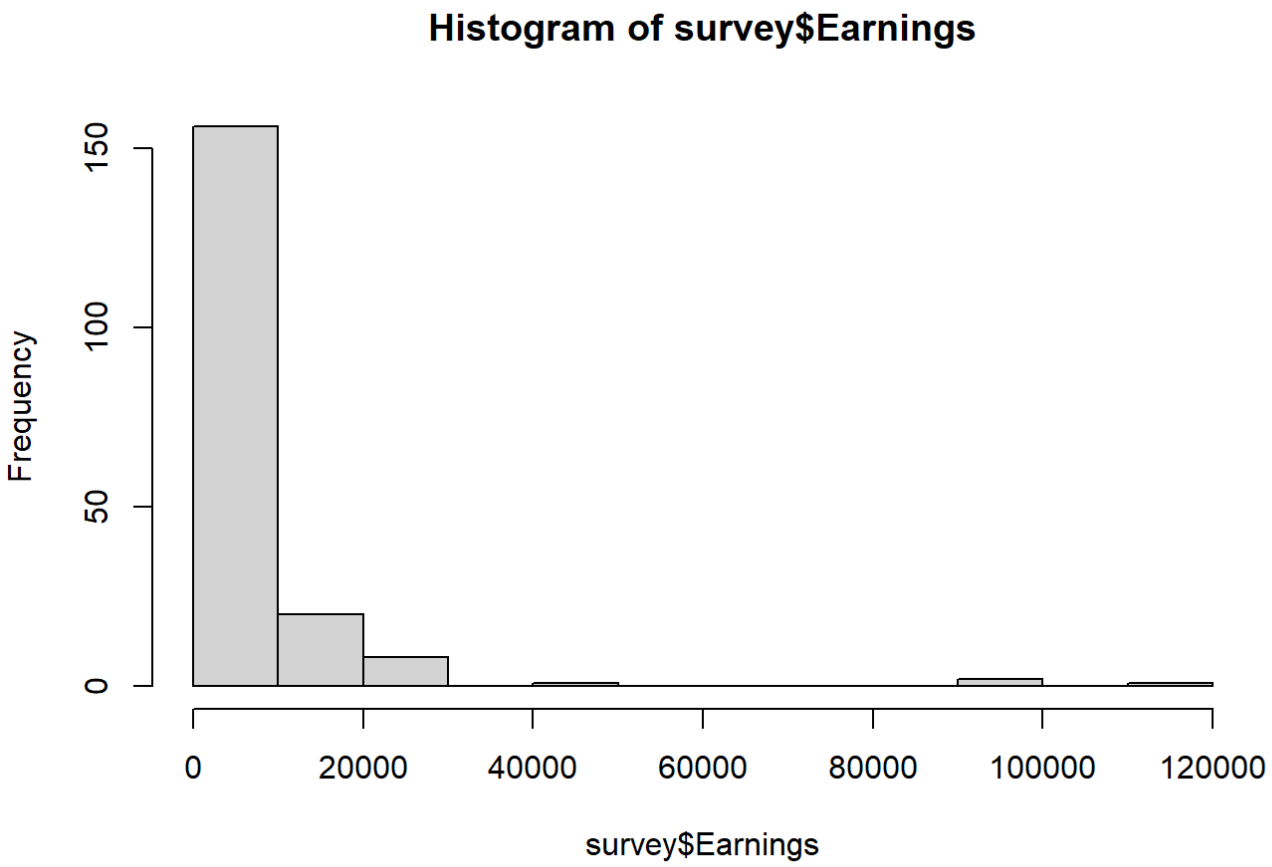
Graphs from Survey data file

Describe the Weight by Height plot made with the survey data. Is it similar to the one above? How does it differ and why?

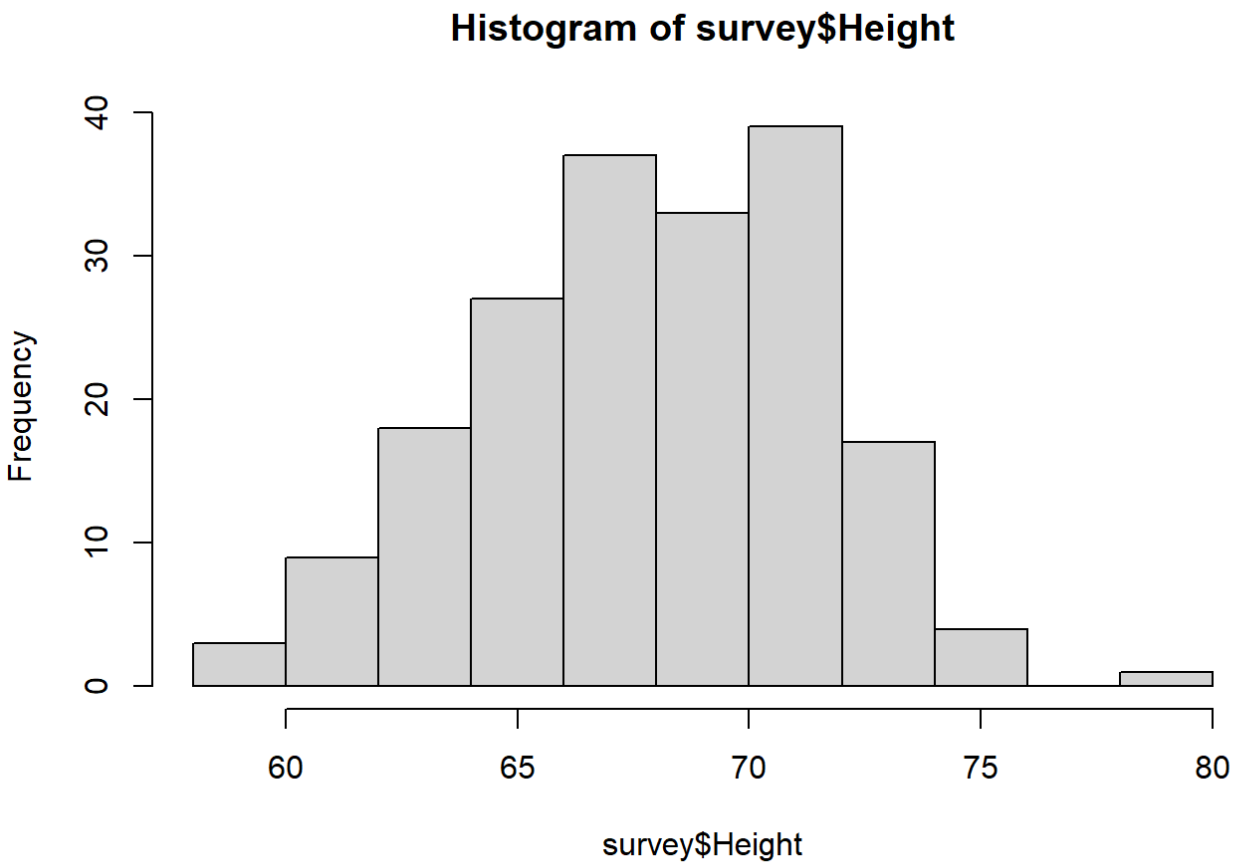
This scatterplot is similar to the first one in the sense that they are both are positive and linear.

It differs because it has a very distinct strong relationship and because we can clearly identify outliers.

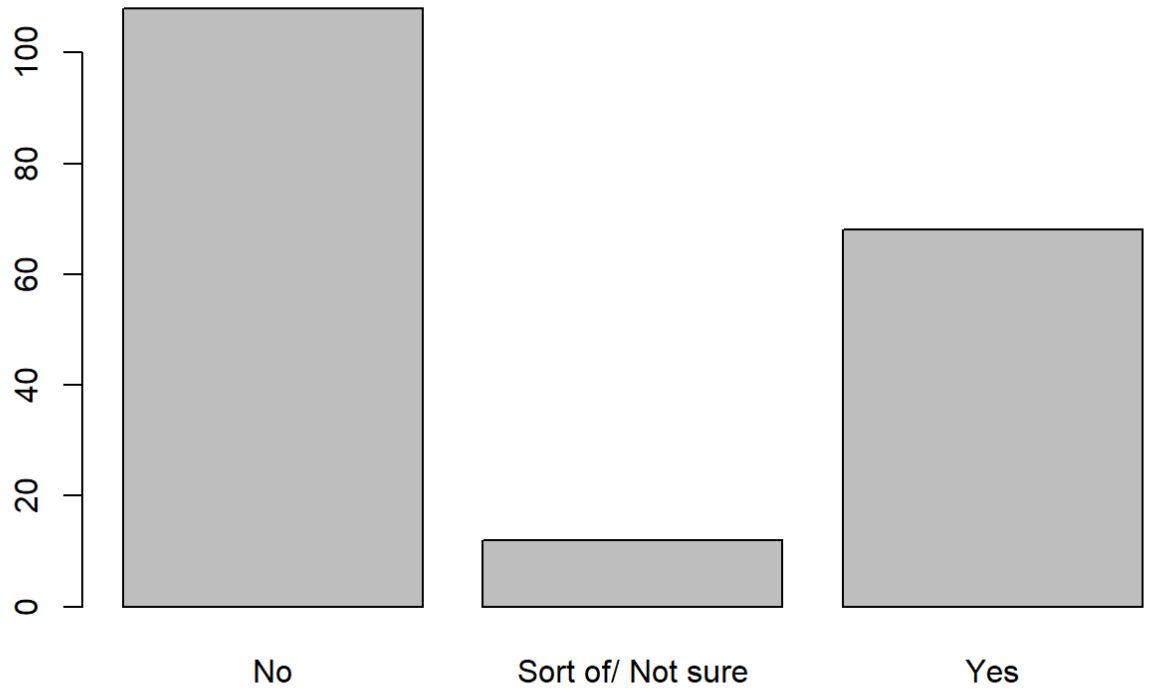
```
#  Make these graphs using the survey data that you read in above:

#   Have R make a histogram of earnings
hist(survey$Earnings)
```



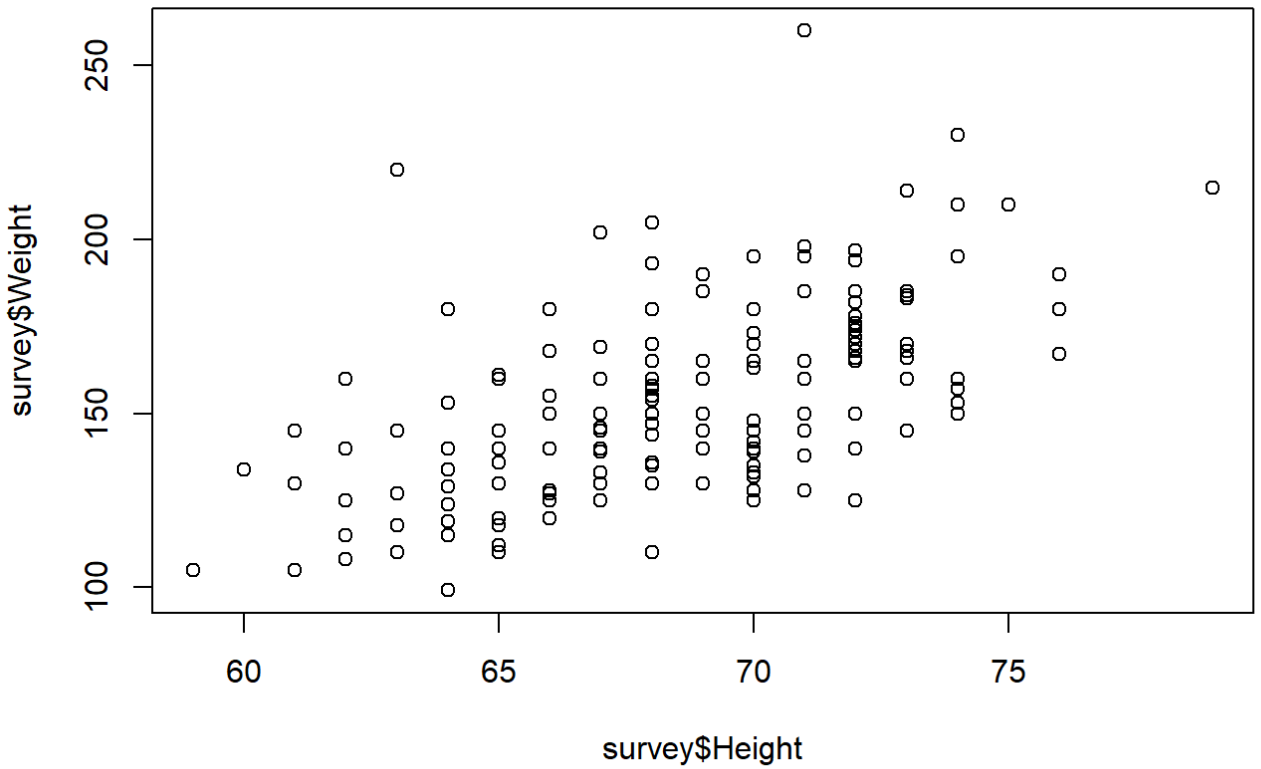

```
#   Have R make a histogram of height
hist(survey$Height)
```




```
#   Have R do a barplot of relationship
barplot(table(survey$Relationship))
```




```
#   Finally, plot weight by height
plot(x=survey$Height, y=survey$Weight)
```


survey$Weight
100
150
200
250
60
65
70
75
survey$Height