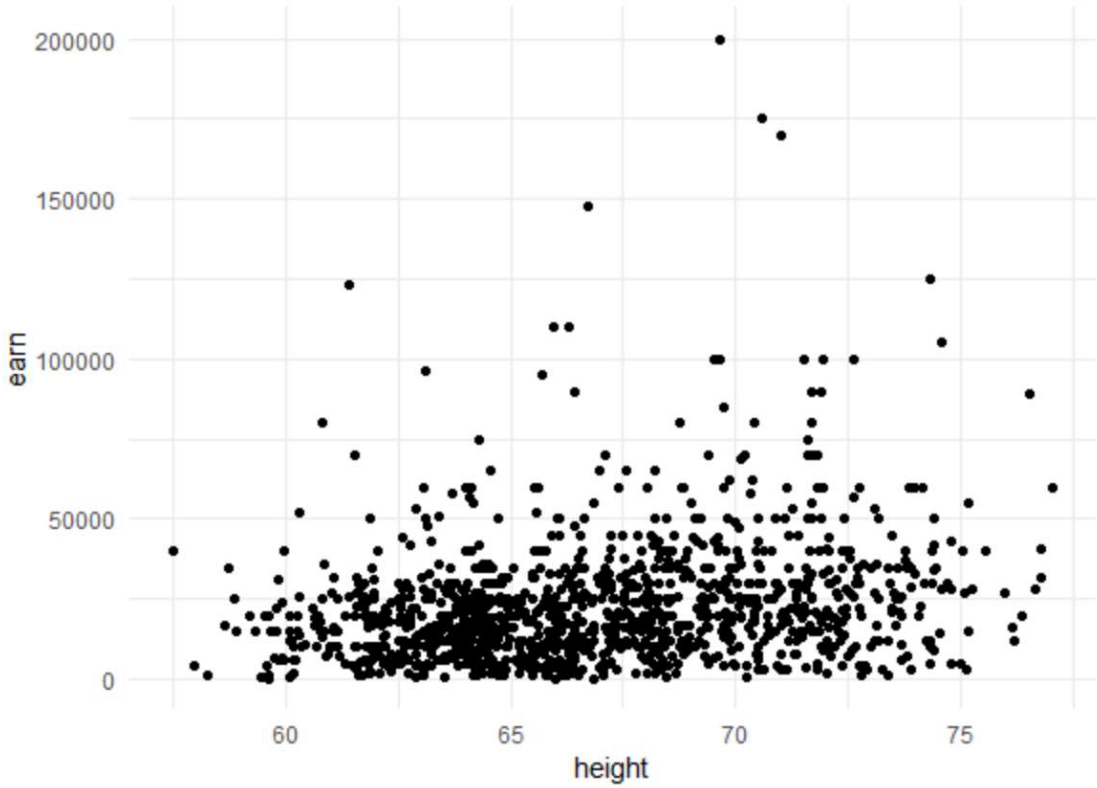
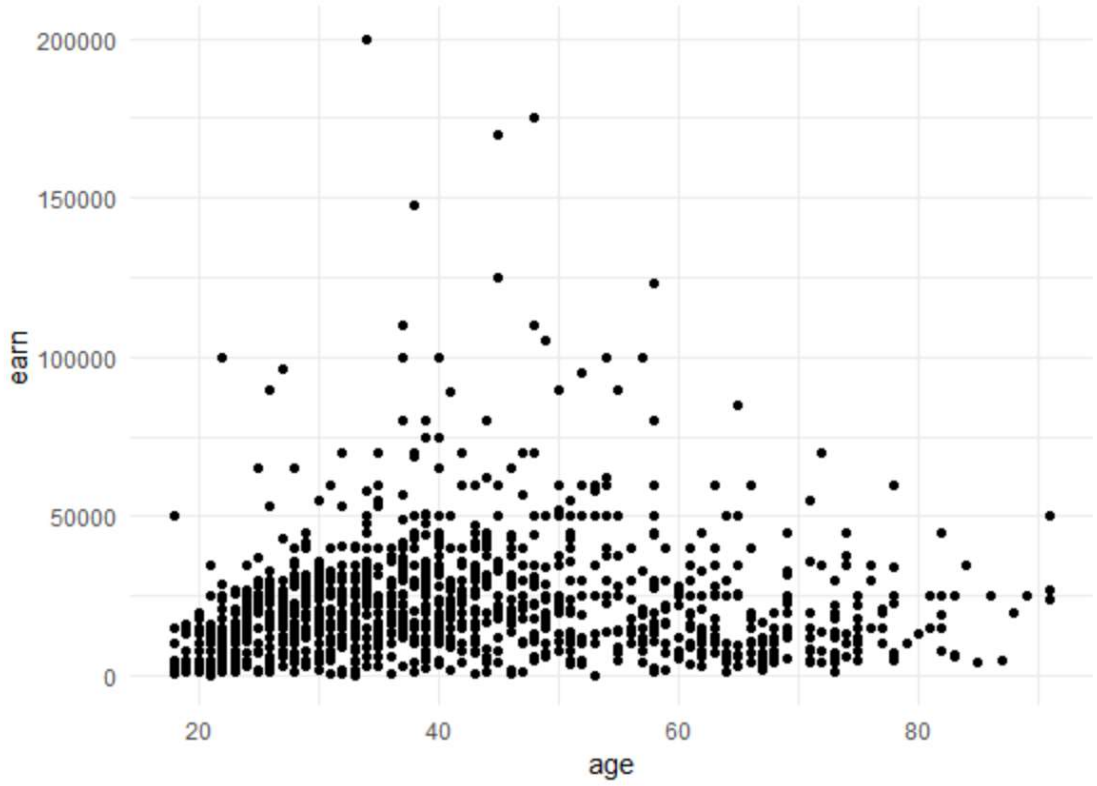
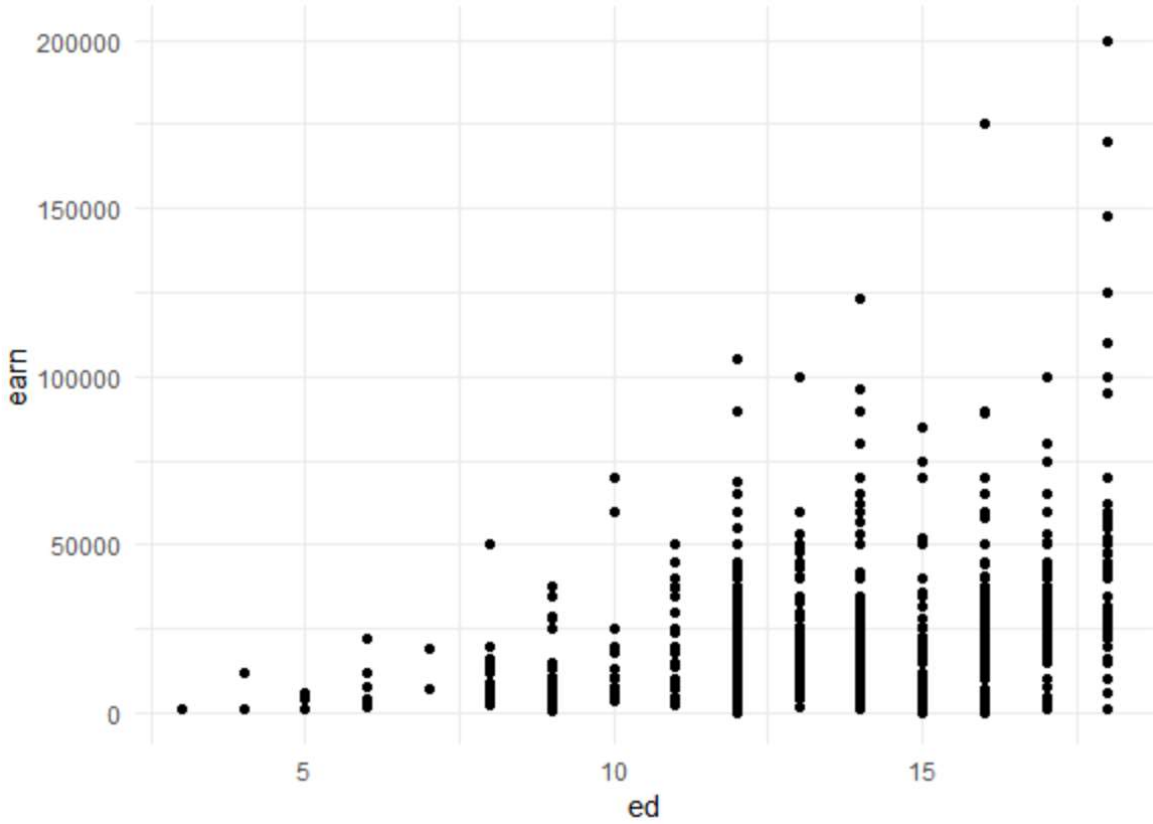
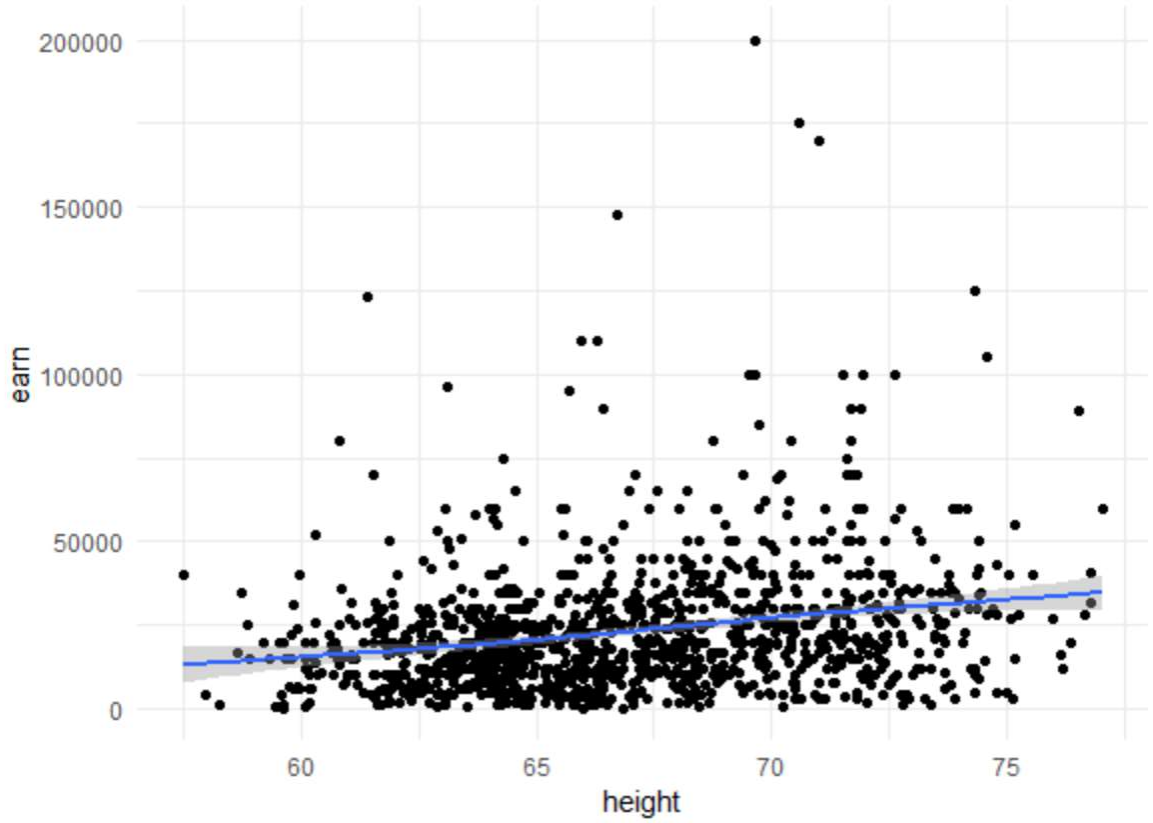
```
# Assignment: ASSIGNMENT 3
# Name: Smitshoek, Stephen
# Date: 2022-03-29

## Load the ggplot2 package
library(ggplot2)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory
setwd("C:\\Users\\sksmi\\PeytoAccess\\Personal\\Bellevue\\DSC520\\dsc520")

## Load the `data/r4ds/heights.csv` to
heights_df <- read.csv("data\\r4ds\\heights.csv")

# https://ggplot2.tidyverse.org/reference/geom_point.html
## Using `geom_point()` create three scatterplots for
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y=earn)) + geom_point()
## `age` vs. `earn`
ggplot(heights_df, aes(x=age, y=earn)) + geom_point()
## `ed` vs. `earn`
ggplot(heights_df, aes(x=ed, y=earn)) + geom_point()

## Re-create the three scatterplots and add a regression trend line using
## the `geom_smooth()` function
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y=earn)) + geom_point() + geom_smooth()
## `age` vs. `earn`
ggplot(heights_df, aes(x=age, y=earn)) + geom_point() + geom_smooth()
## `ed` vs. `earn`
ggplot(heights_df, aes(x=ed, y=earn)) + geom_point() + geom_smooth()

## Create a scatterplot of `height`` vs. `earn`.  Use `sex` as the `col`
(color) attribute
ggplot(heights_df, aes(x=height, y=earn, col=sex)) + geom_point()

## Using `ggtitle()`, `xlab()`, and `ylab()` to add a title, x label, and y
label to the previous plot
## Title: Height vs. Earnings
## X label: Height (Inches)
## Y Label: Earnings (Dollars)
ggplot(heights_df, aes(x=height, y=earn, col=sex)) + geom_point() +
ggtitle("Height vs. Earnings") + xlab("Height (Inches)") + ylab("Earnings
(Dollars)")

# https://ggplot2.tidyverse.org/reference/geom_histogram.html
## Create a histogram of the `earn` variable using `geom_histogram()`
ggplot(heights_df, aes(earn)) + geom_histogram()

## Create a histogram of the `earn` variable using `geom_histogram()`
## Use 10 bins
ggplot(heights_df, aes(earn)) + geom_histogram(bins=10)

# https://ggplot2.tidyverse.org/reference/geom_density.html
## Create a kernel density plot of `earn` using `geom_density()`
```

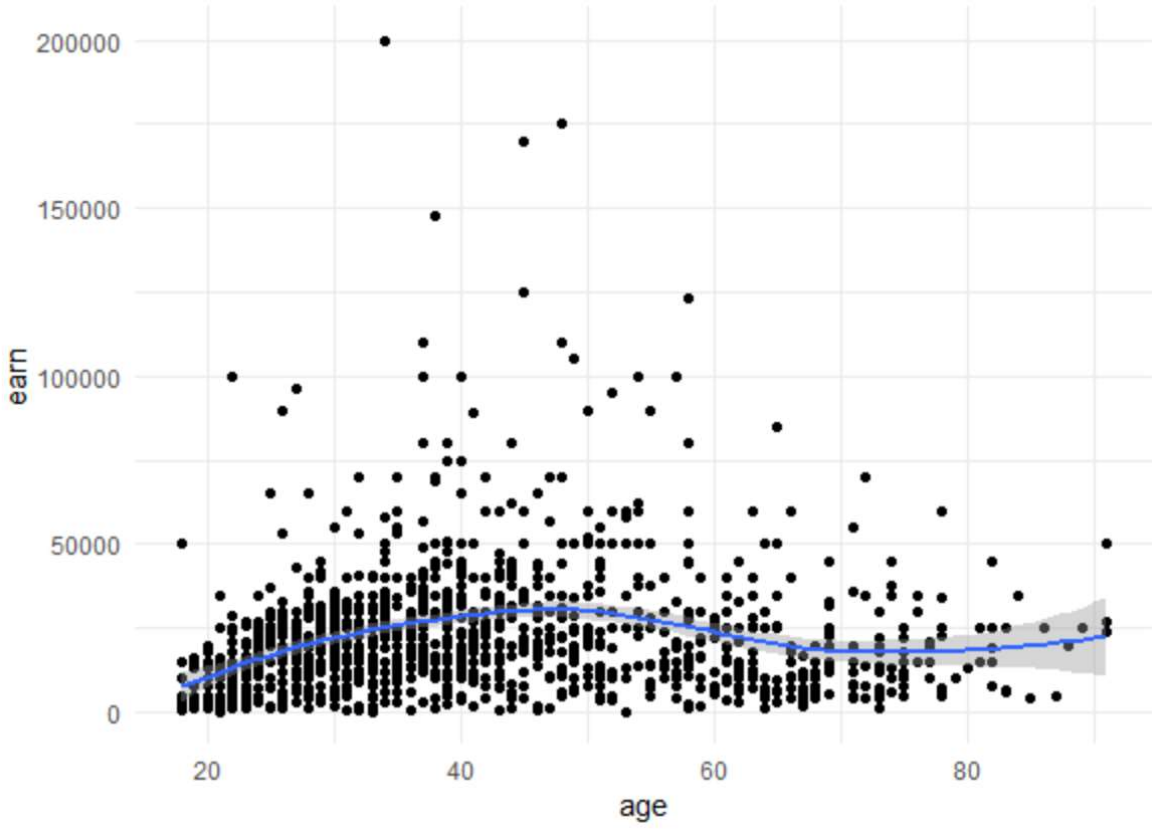
```
ggplot(heights_df, aes(earn)) +  geom_density()
```

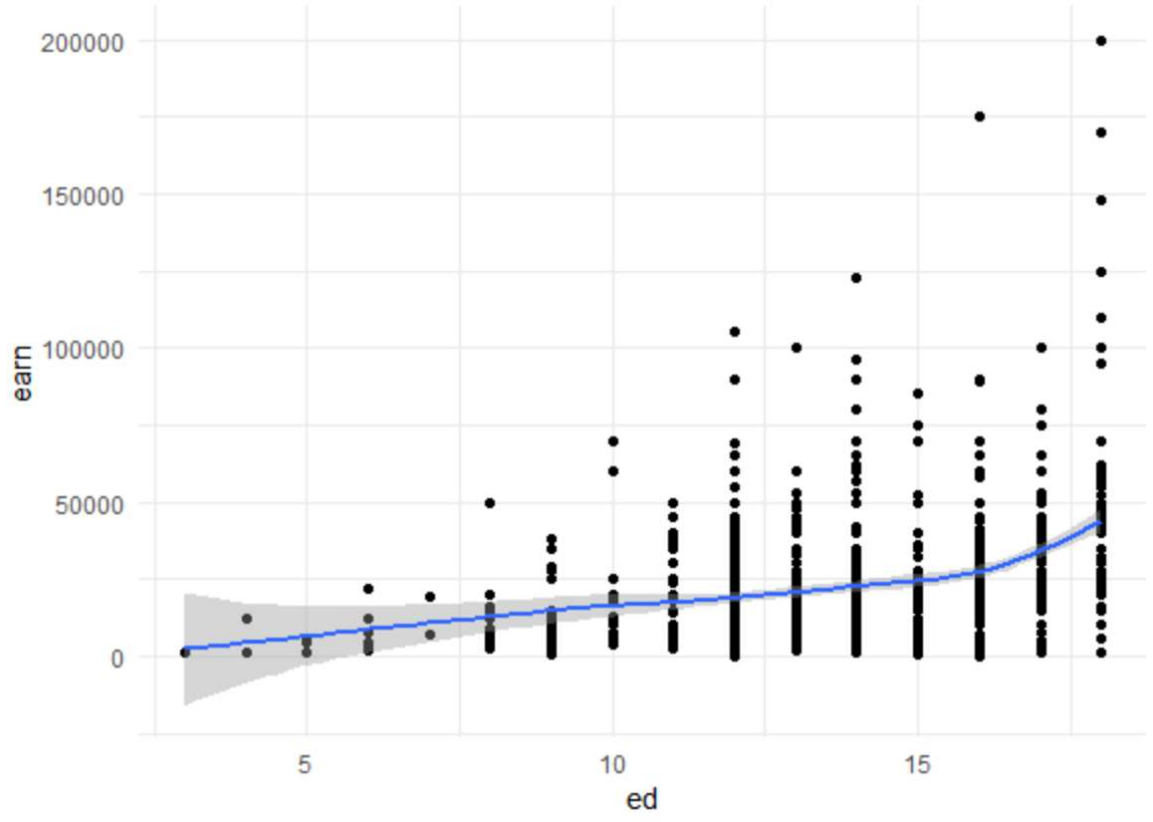

200000
150000
100000
50000
0
earn
60
65
70
75
height

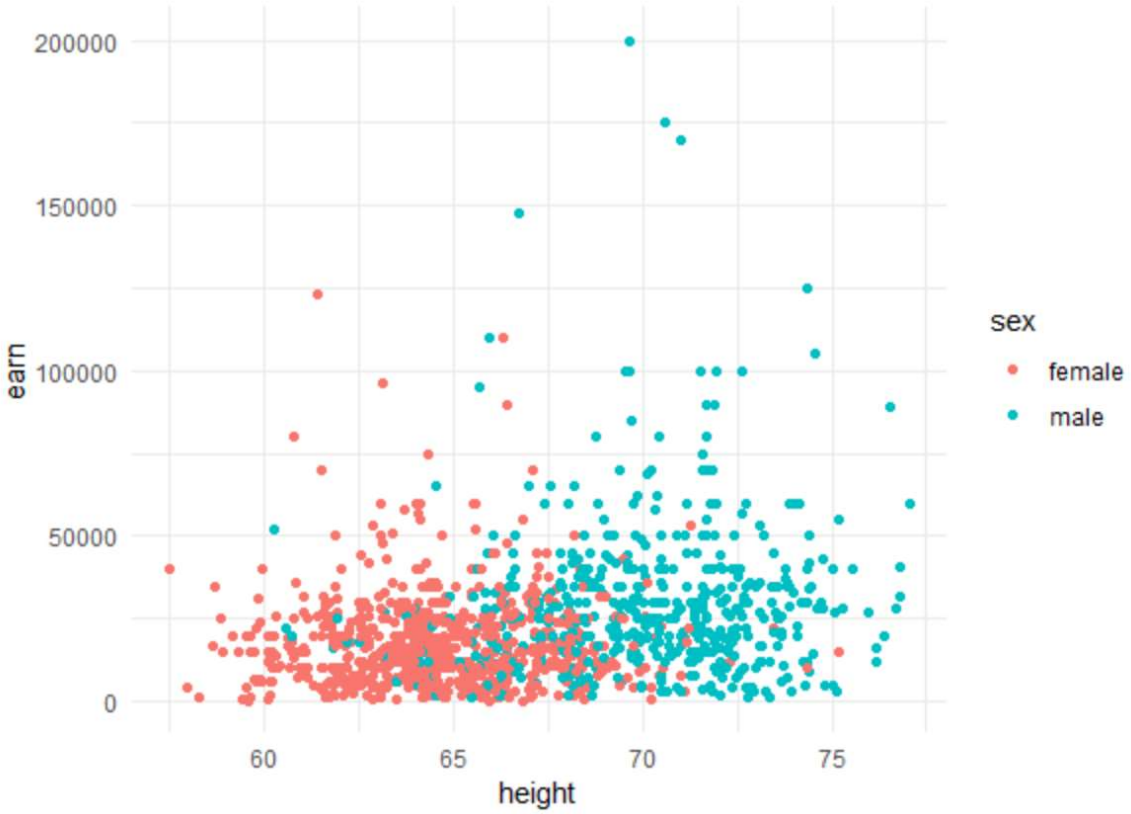

200000
150000
100000
50000
0
earn
20
40
60
80
age


200000
150000
100000
50000
0
earn
5
10
15
ed

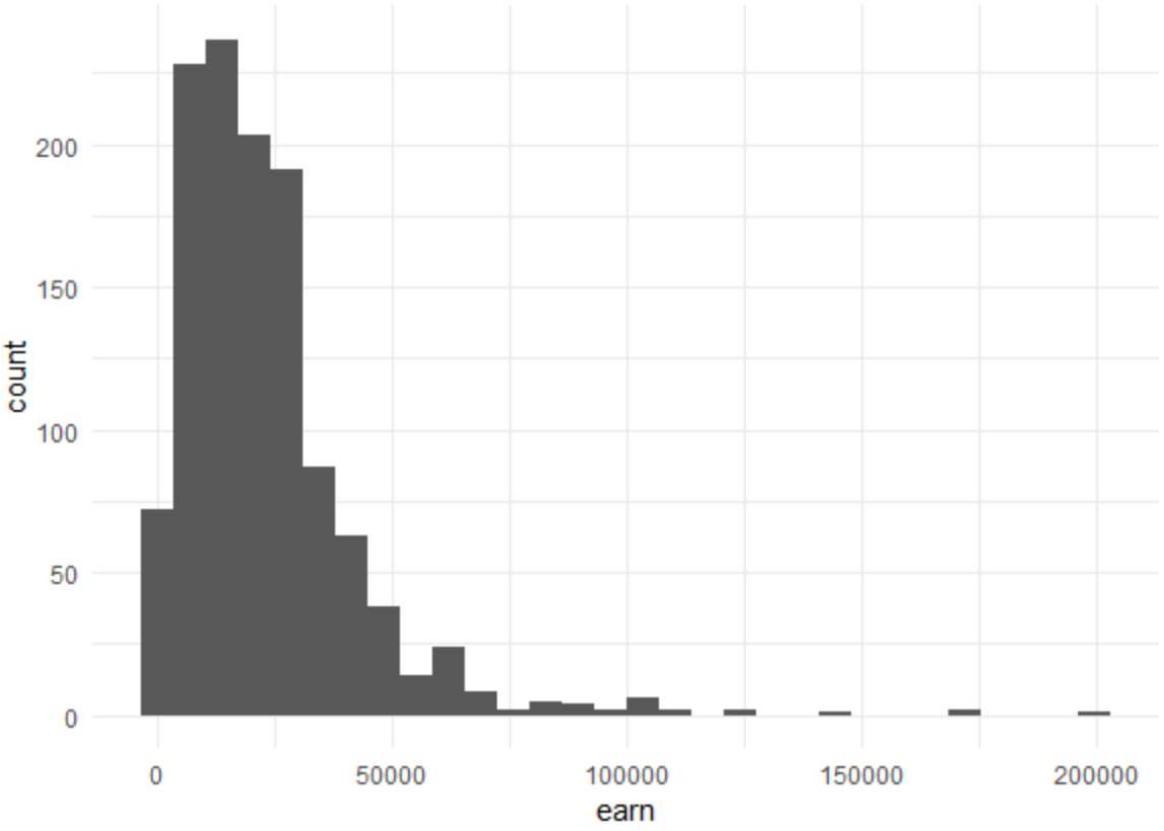

200000
150000
100000
50000
0
earn
60
65
70
75
height

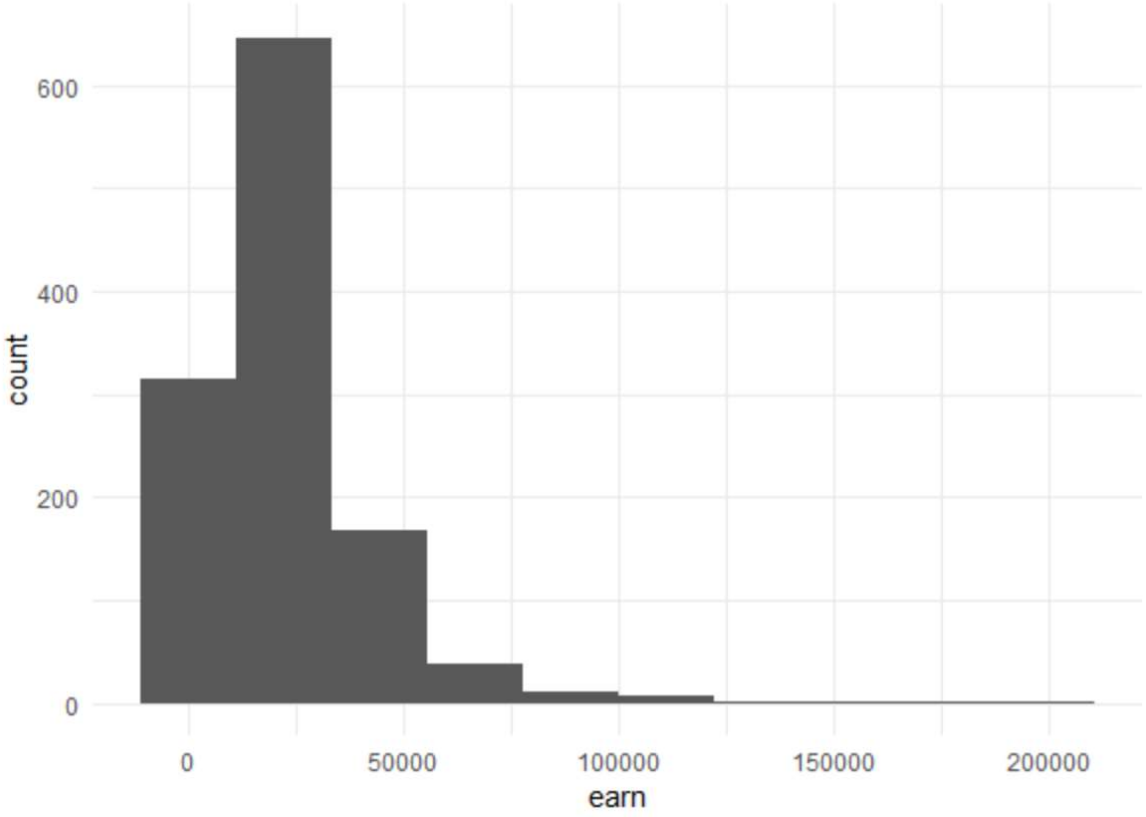

200000
150000
100000
50000
0
earn
20
40
60
80
age

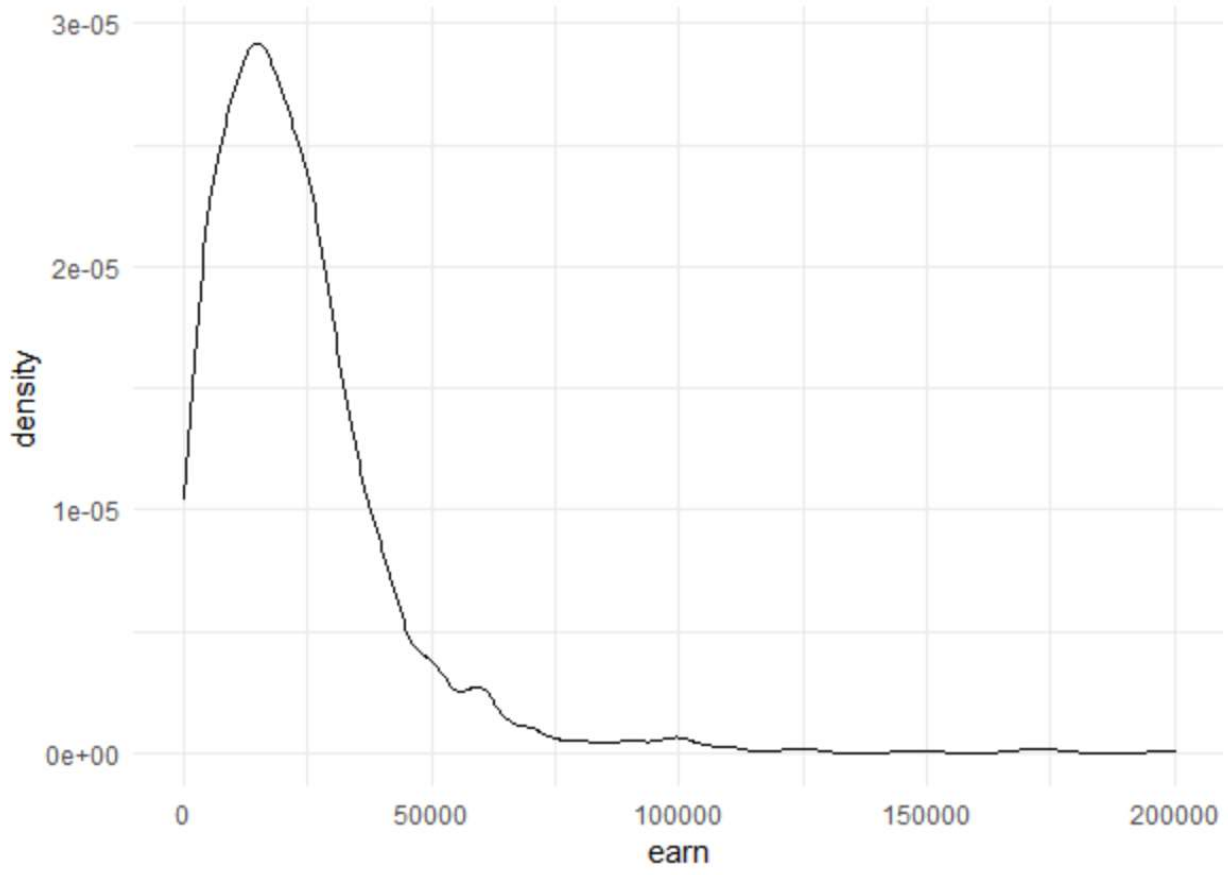

200000
150000
100000
50000
0
earn
5
10
15
ed


200000
150000
100000
50000
0
earn
60
65
70
75
height
sex
female
male


Height vs. Earnings
200000
150000
100000
50000
0
Earnings (Dollars)
60
65
70
75
Height (Inches)
sex
female
male


count
200
150
100
50
0
0
50000
100000
150000
200000
earn


count
600
400
200
0
0
50000
100000
150000
200000
earn


3e-05
2e-05
1e-05
0e+00
density
0
50000
100000
150000
200000
earn