assignment_03_KummarikuntaVidyasagar.R

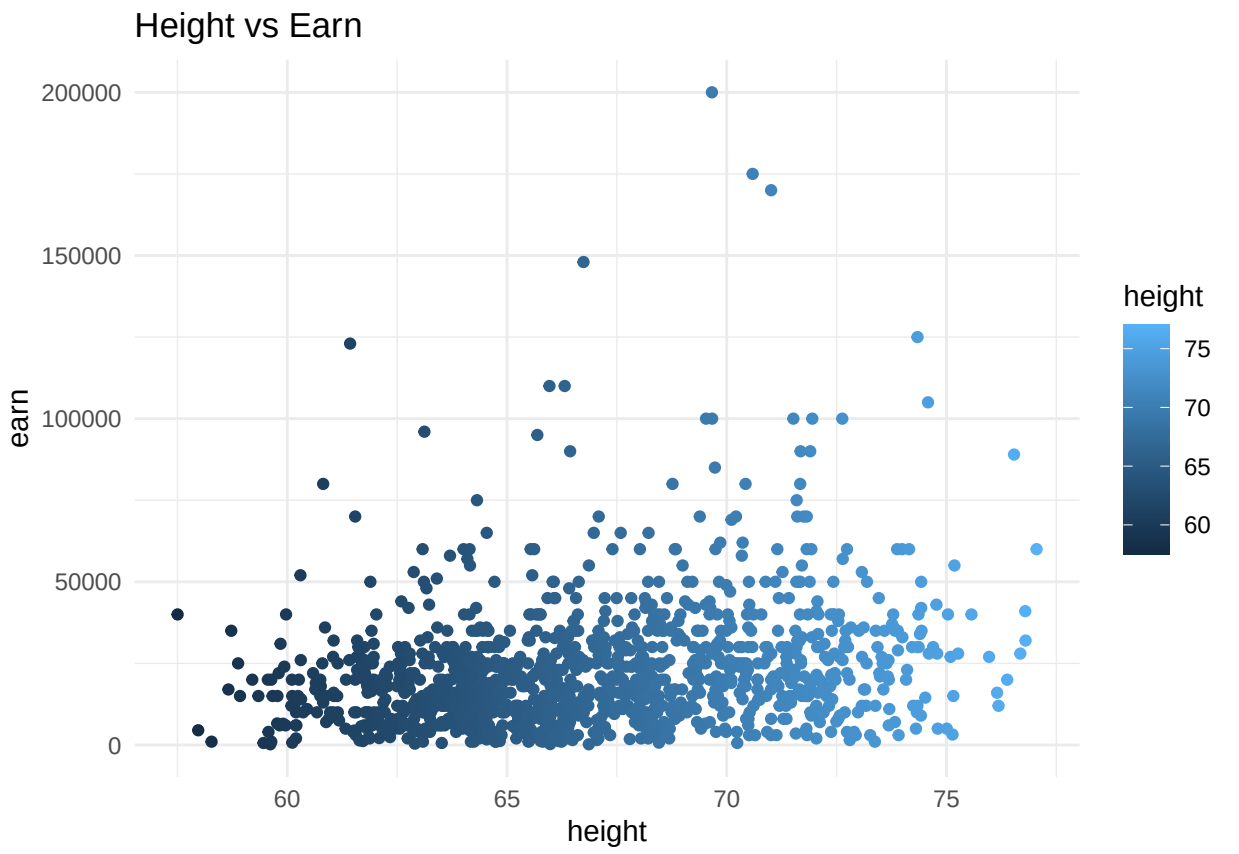
12702

2020-09-18

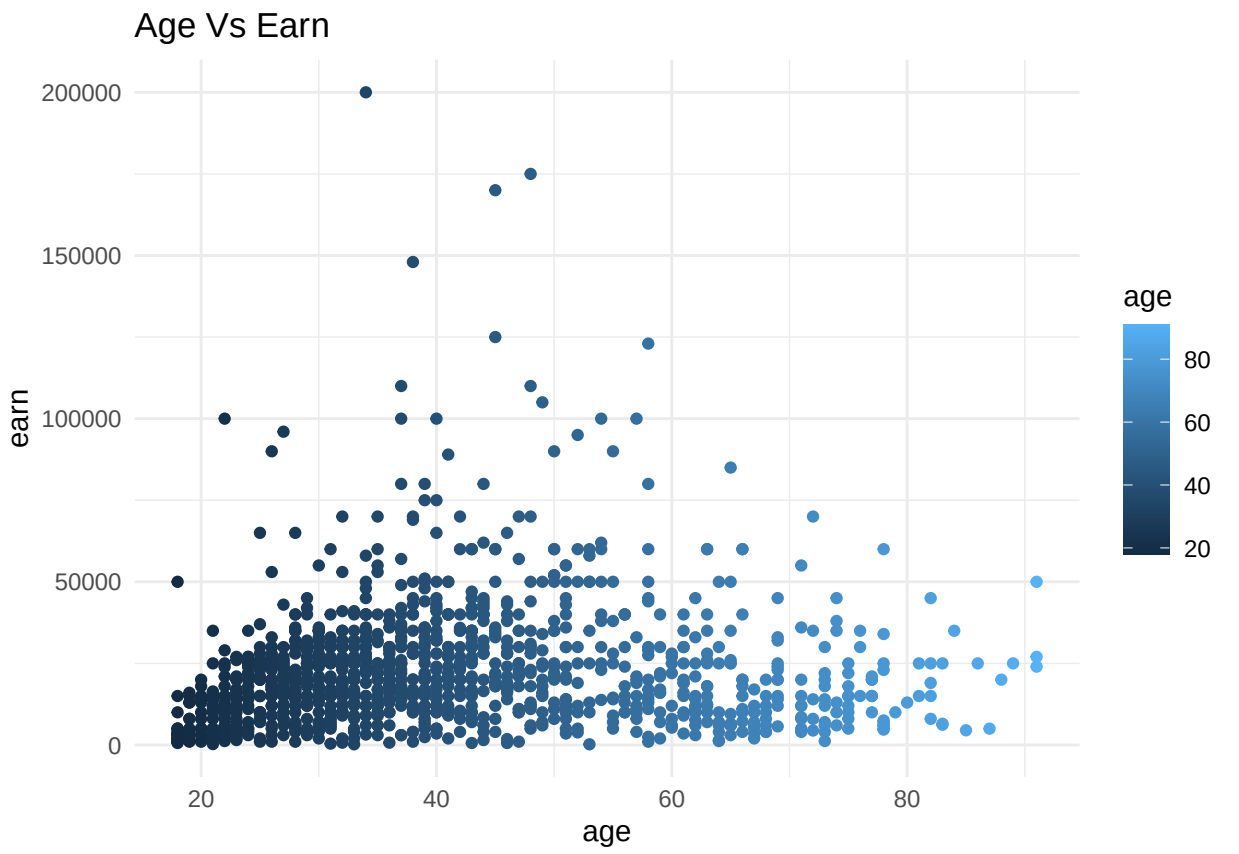
```
# Assignment: ASSIGNMENT 3
# Name: Kummarikunta, Vidyasagar
# Date: 2010-02-14

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

## Set the working directory to the root of your DSC 520 directory
setwd("/Users/12702/Desktop/MODatascience/DSC-520")

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

head(heights_df)
```

```
##    earn   height    sex ed age  race
## 1 50000 74.42444   male 16  45 white
## 2 60000 65.53754 female 16  58 white
## 3 30000 63.62920 female 16  29 white
## 4 50000 63.10856 female 16  91 other
## 5 51000 63.40248 female 17  39 white
## 6  9000 64.39951 female 15  26 white
```

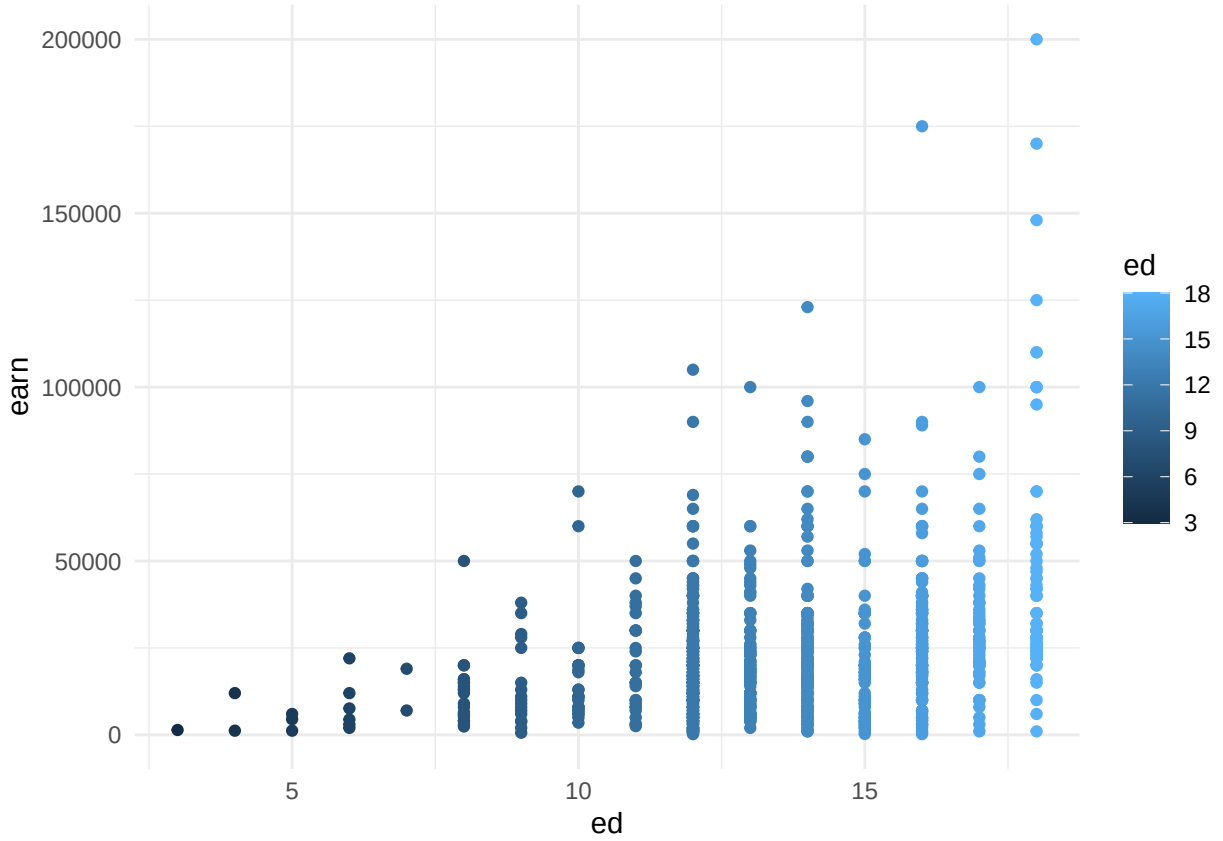
```
# https://ggplot2.tidyverse.org/reference/geom_point.html
## Using `geom_point()` create three scatterplots for
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y=earn)) + geom_point(aes(color = height)) +
         labs(
           title = "Height vs Earn"
         )
```




```
## `age` vs. `earn`
ggplot(heights_df, aes(x= age, y= earn)) + geom_point(aes(color = age)) +
      labs(
          title = "Age Vs Earn"
      )
```

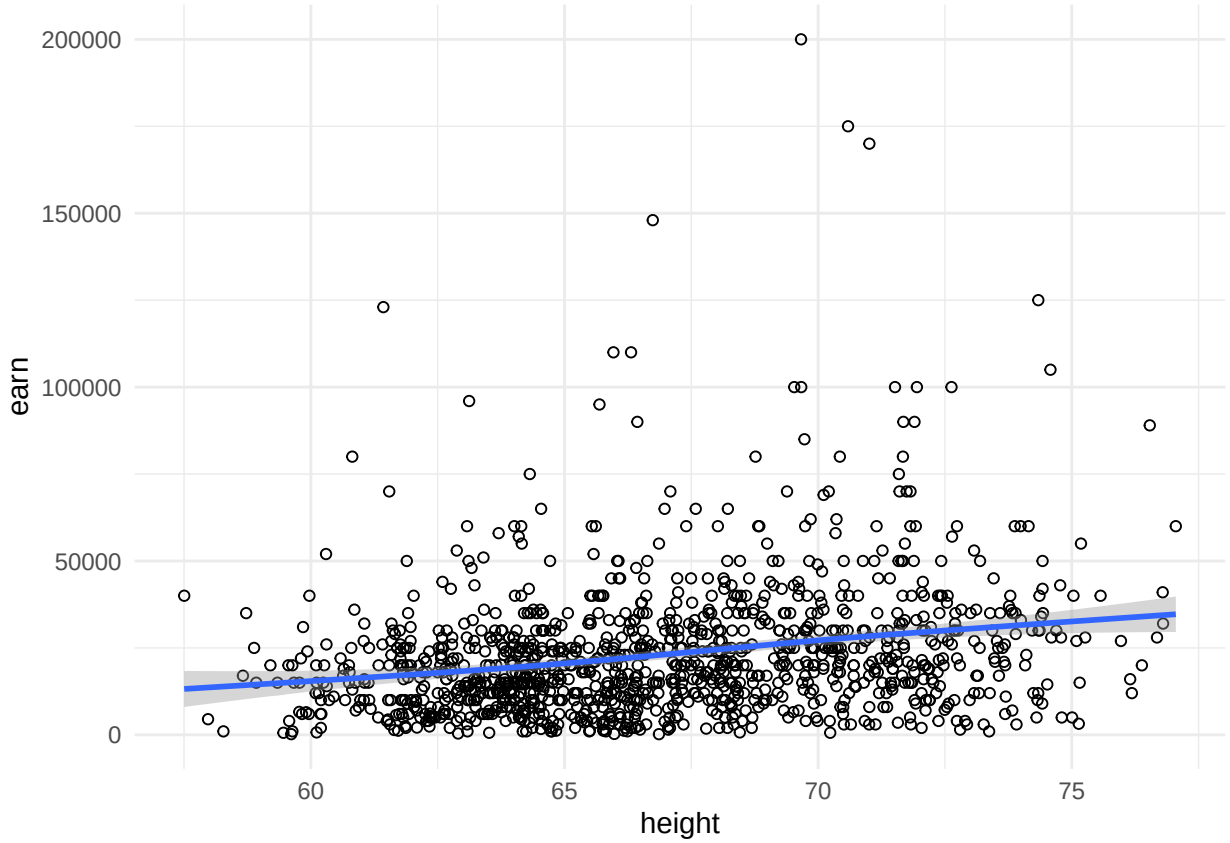



```
## `ed` vs. `earn`
ggplot(heights_df, aes(x=ed, y=earn)) + geom_point(aes(color = ed))
```



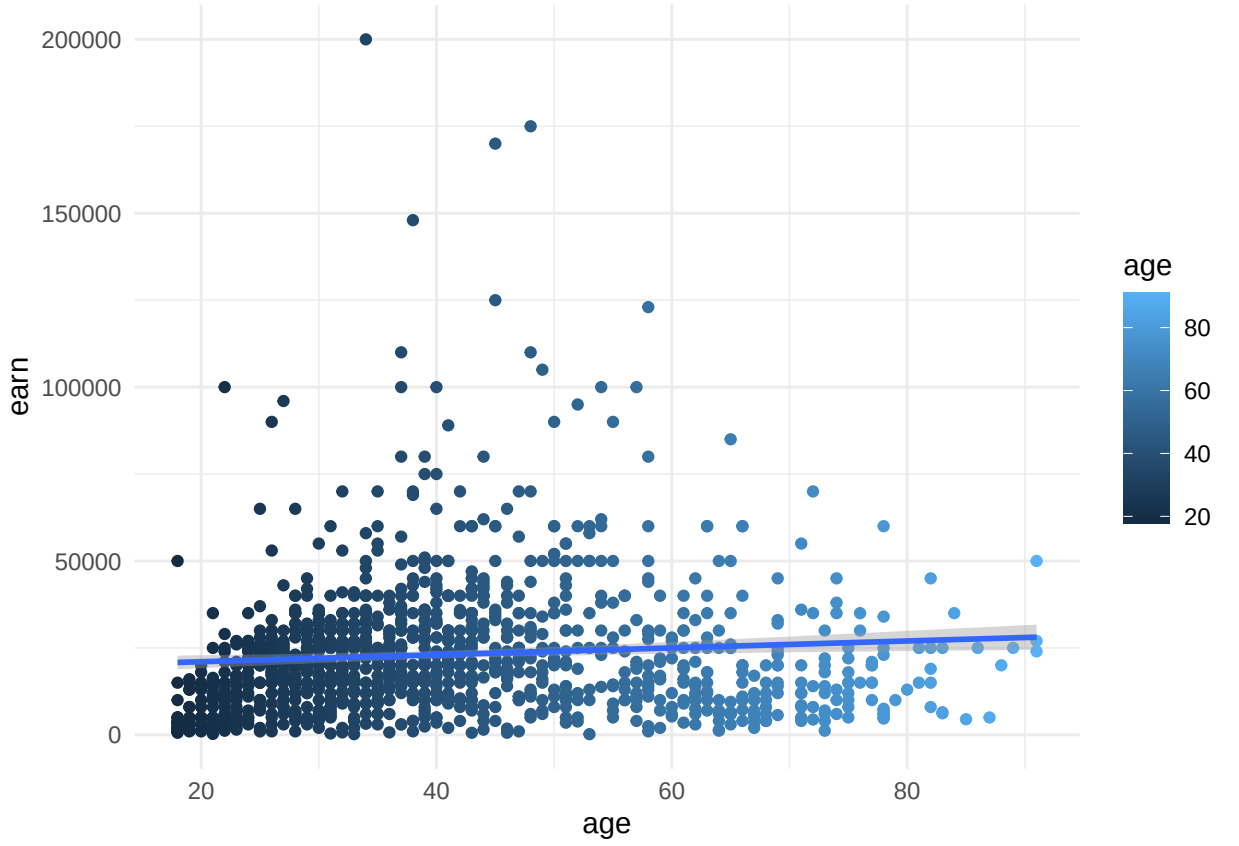

```
## Re-create the three scatterplots and add a regression trend line using
## the `geom_smooth()` function
## `height` vs. `earn`
ggplot(heights_df, aes(x=height, y= earn)) + geom_point(shape = 1) + geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



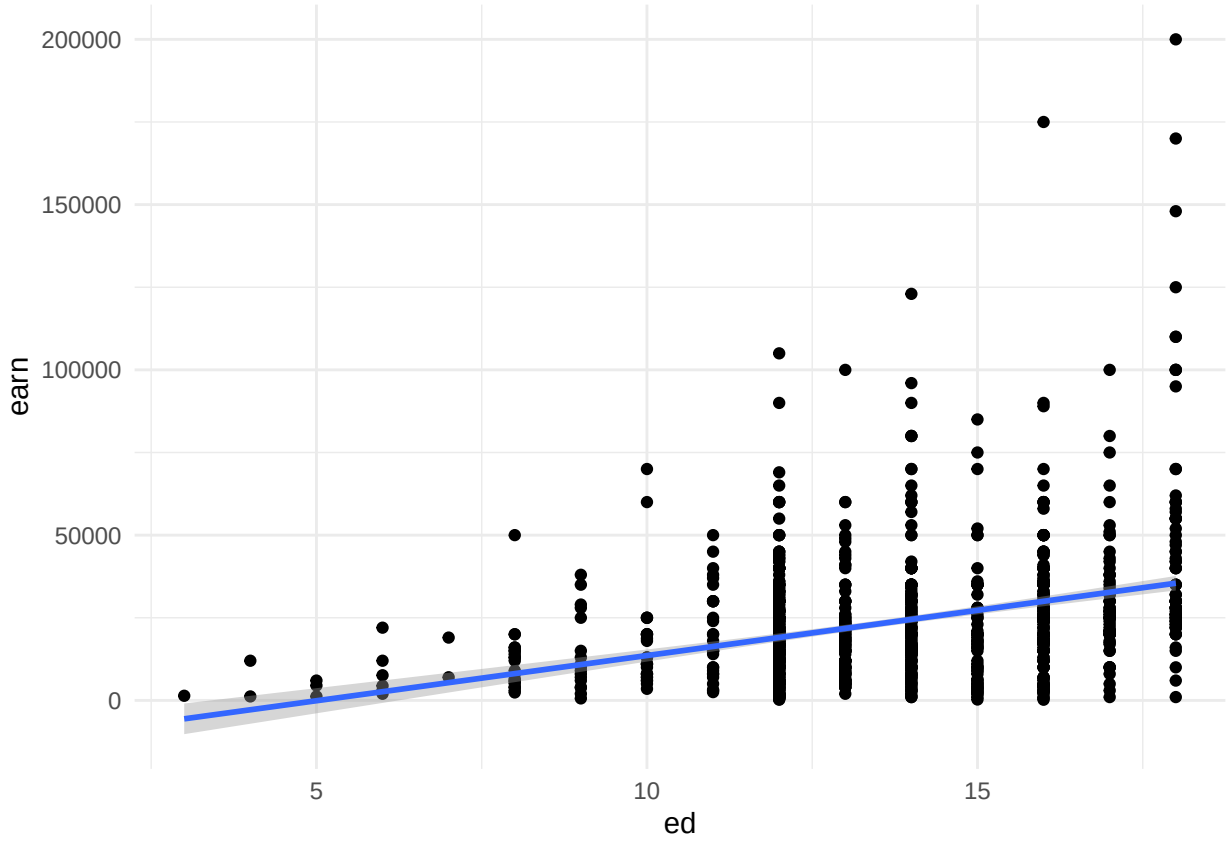

```
## `age` vs. `earn`
ggplot(heights_df, aes(x=age, y= earn)) + geom_point(aes(color = age)) +
  geom_smooth(method = lm)
```

```
## `geom_smooth()` using formula 'y ~ x'
```

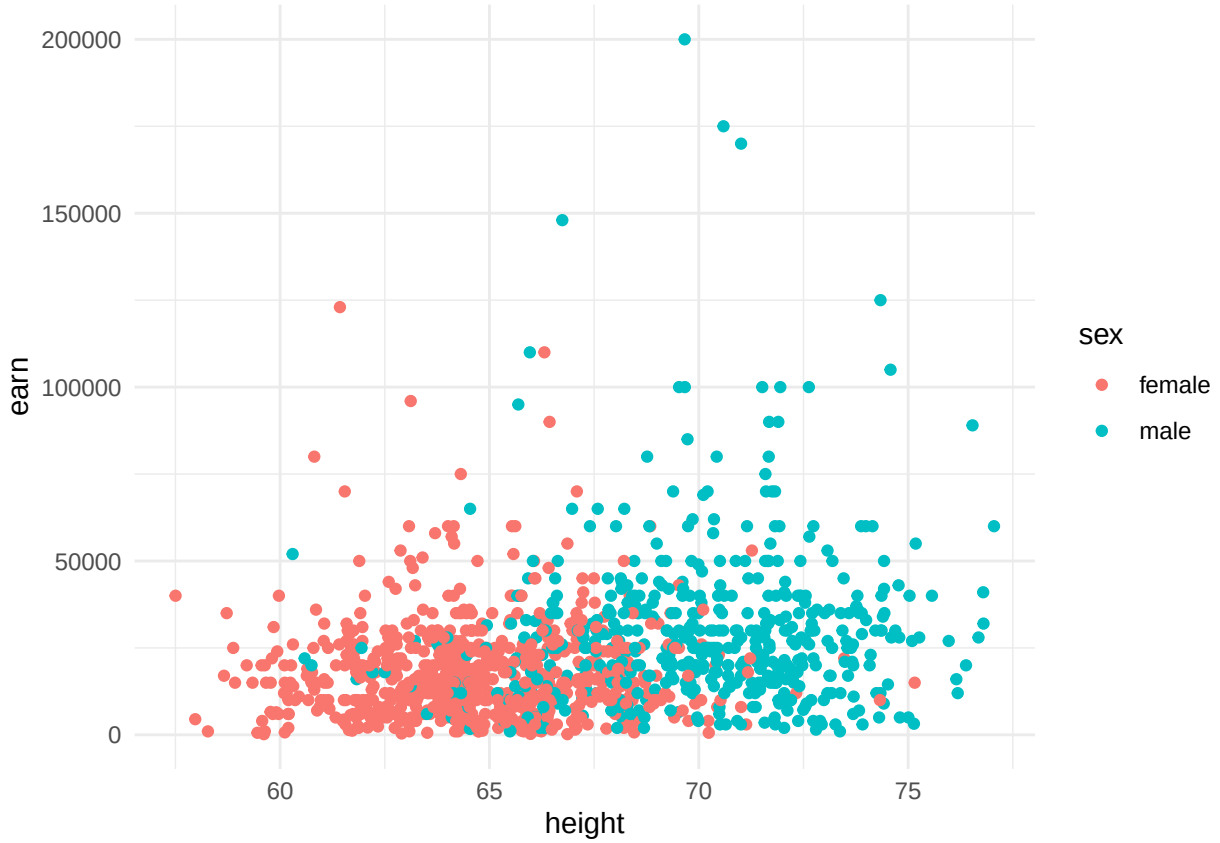



```
## 'ed' vs. 'earn'
ggplot(heights_df, aes(x=ed, y= earn)) + geom_point() + geom_smooth(method =lm)
```

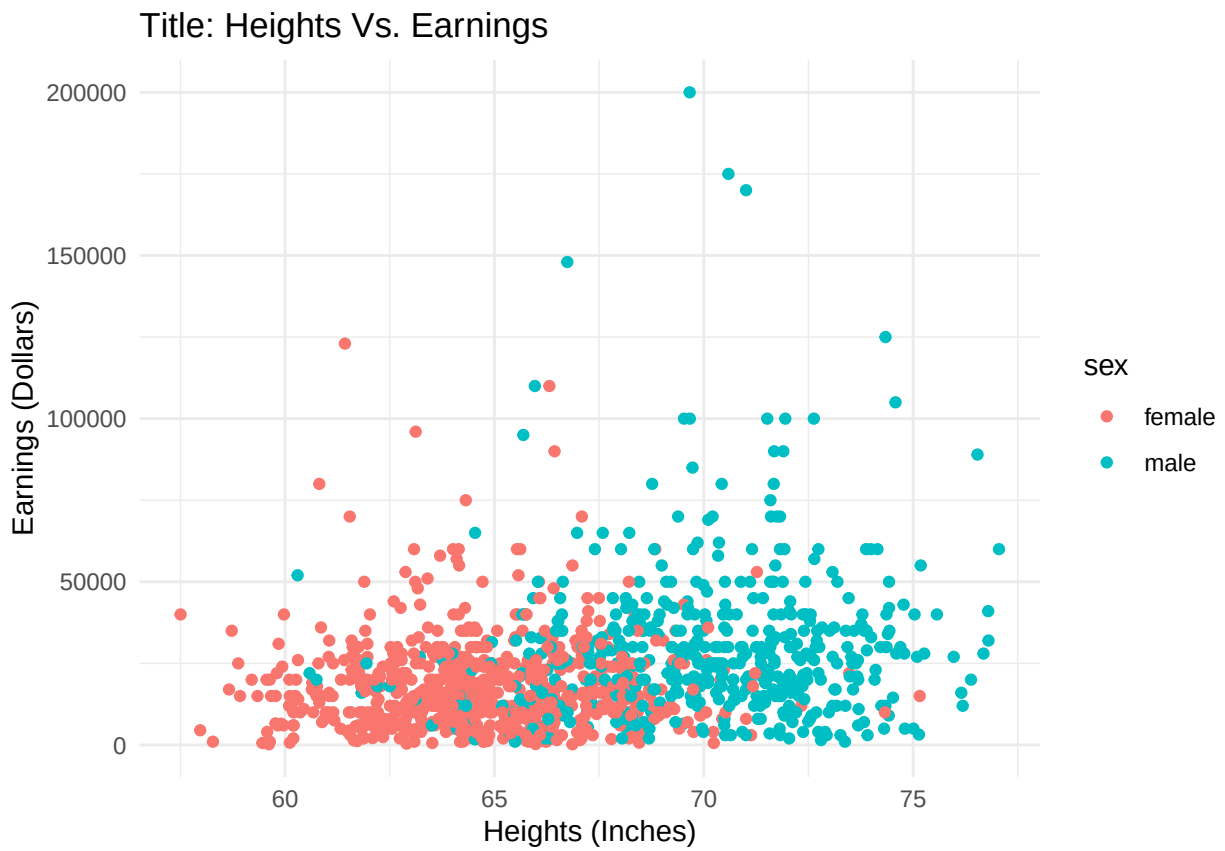
```
## 'geom_smooth()' using formula 'y ~ x'
```




```
## 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()
```

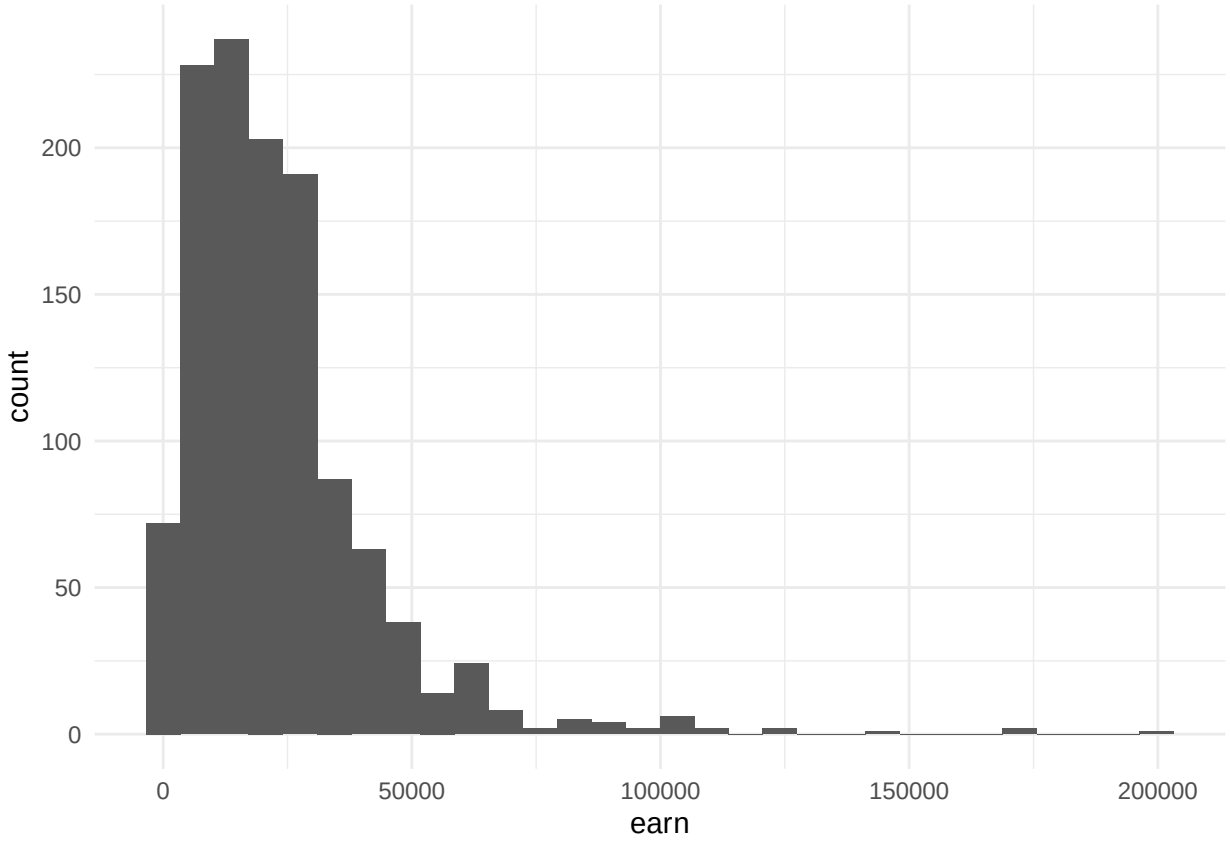



```
## Using 'ggtitle()', 'xlab()', and 'ylab()' to add a title, x label, and y label to the previous plot
## Title: Height vs. Earnings
## X label: Height (Inches)
## Y Label: Earnings (Dollars)
ggplot(heights_df, aes(x=height, y= earn, col=sex)) + geom_point() +
      ggtitle("Title: Heights Vs. Earnings")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

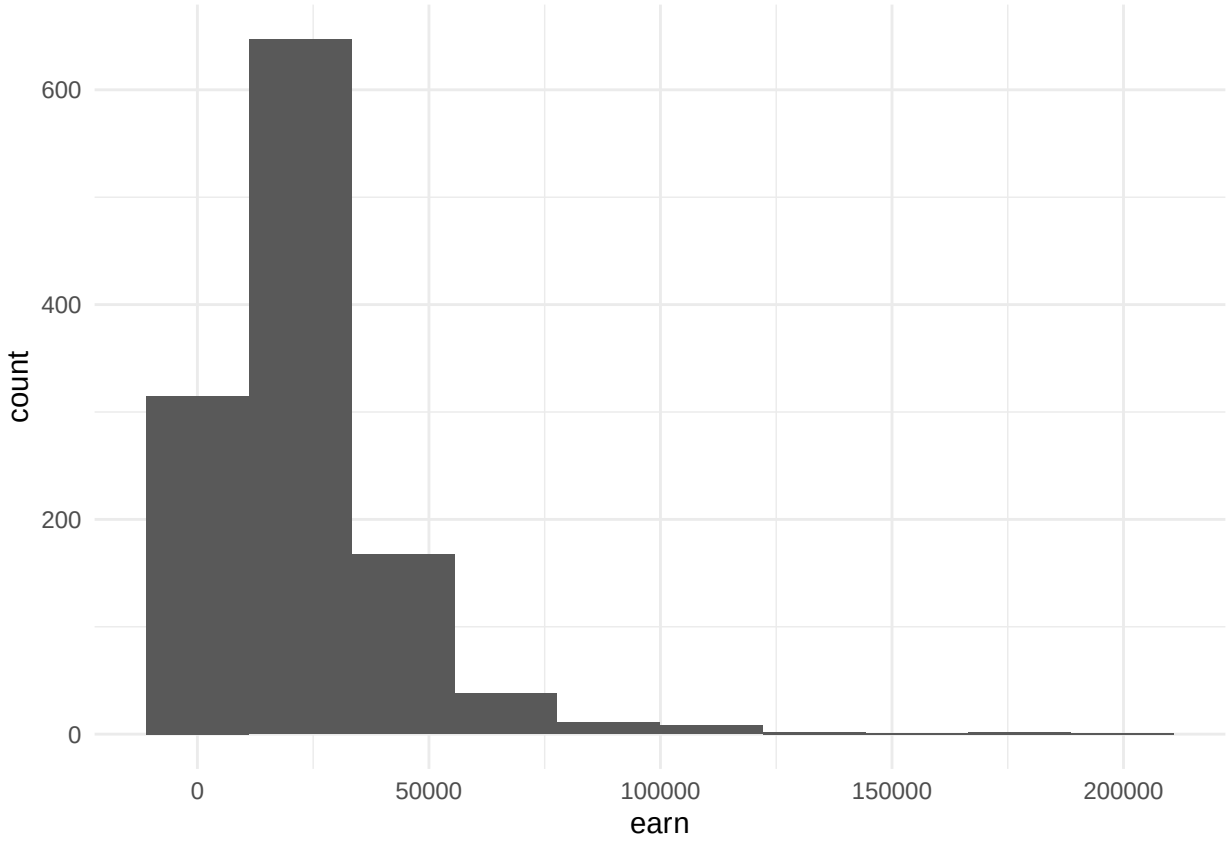



```
# https://ggplot2.tidyverse.org/reference/geom_histogram.html
## Create a histogram of the 'earn' variable using 'geom_histogram()'
ggplot(heights_df, aes(x=earn)) + geom_histogram()
```

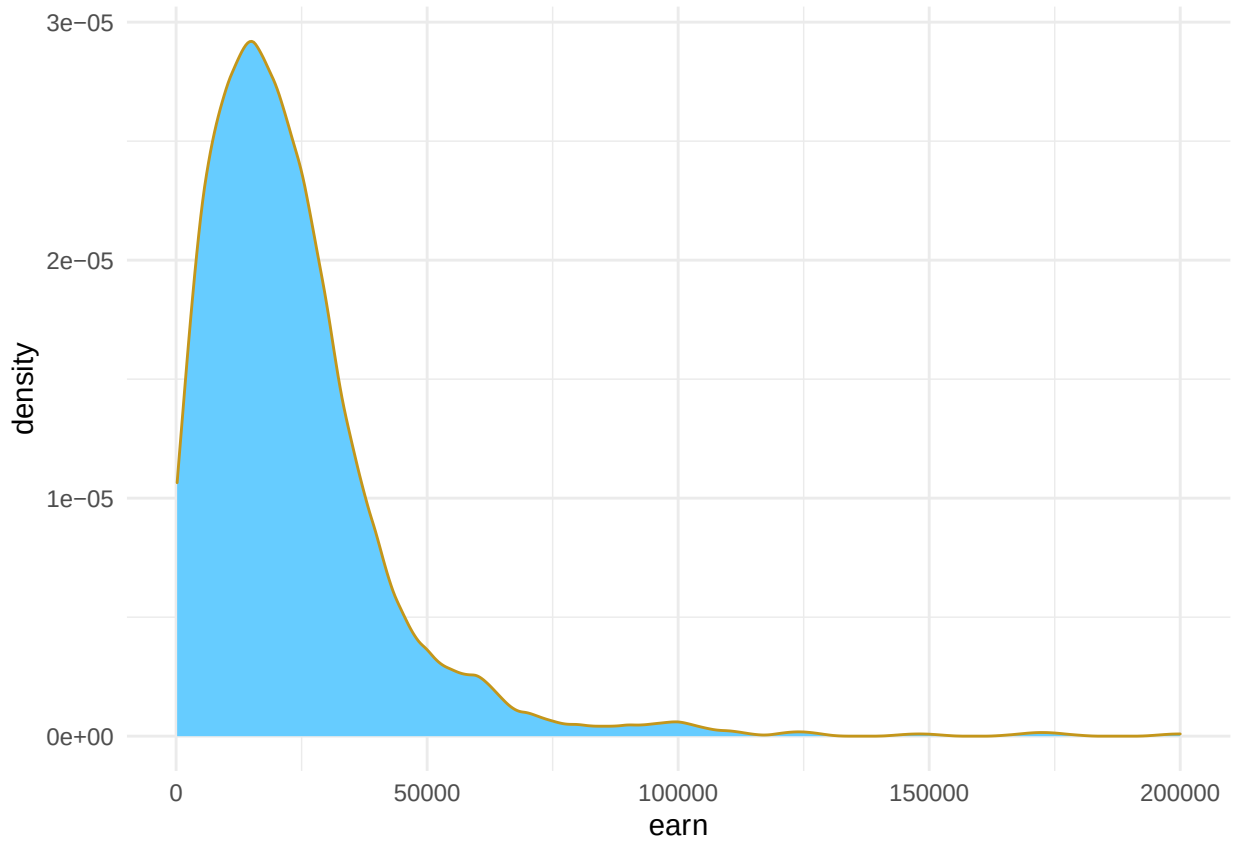
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```




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




```
# https://ggplot2.tidyverse.org/reference/geom_density.html
## Create akernel density plot of ‘earn‘ using ‘geom_density()‘
ggplot(heights_df, aes(x=earn)) +  geom_density(aes(x= earn), fill = "#66CCFF",
                                                color="#C4961A" , kernel ="cosine")
```


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