Assignment: ASSIGNMENT 3

Name: Shekhar, Manish

Date: 2021-03-28

Load the ggplot2 package

```
library(ggplot2)
theme_set(theme_minimal())
```

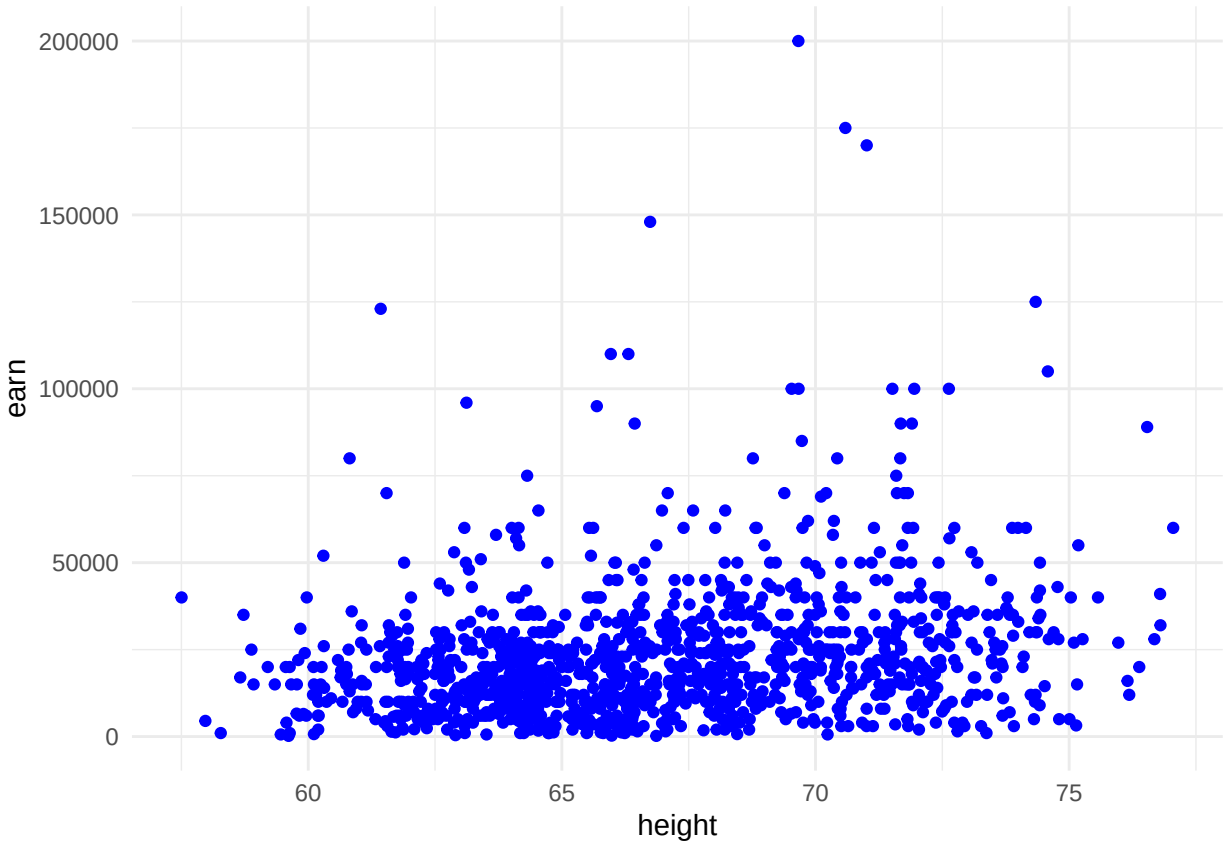
Set the working directory to the root of your DSC 520 directory setwd("/home/jdoe/Workspaces/dsc520") This step was not needed as copied heights.csv to my working directory Load the `data/r4ds/heights.csv` to dataframe

```
heights_df <- read.csv("heights.csv")
head(heights_df)
```

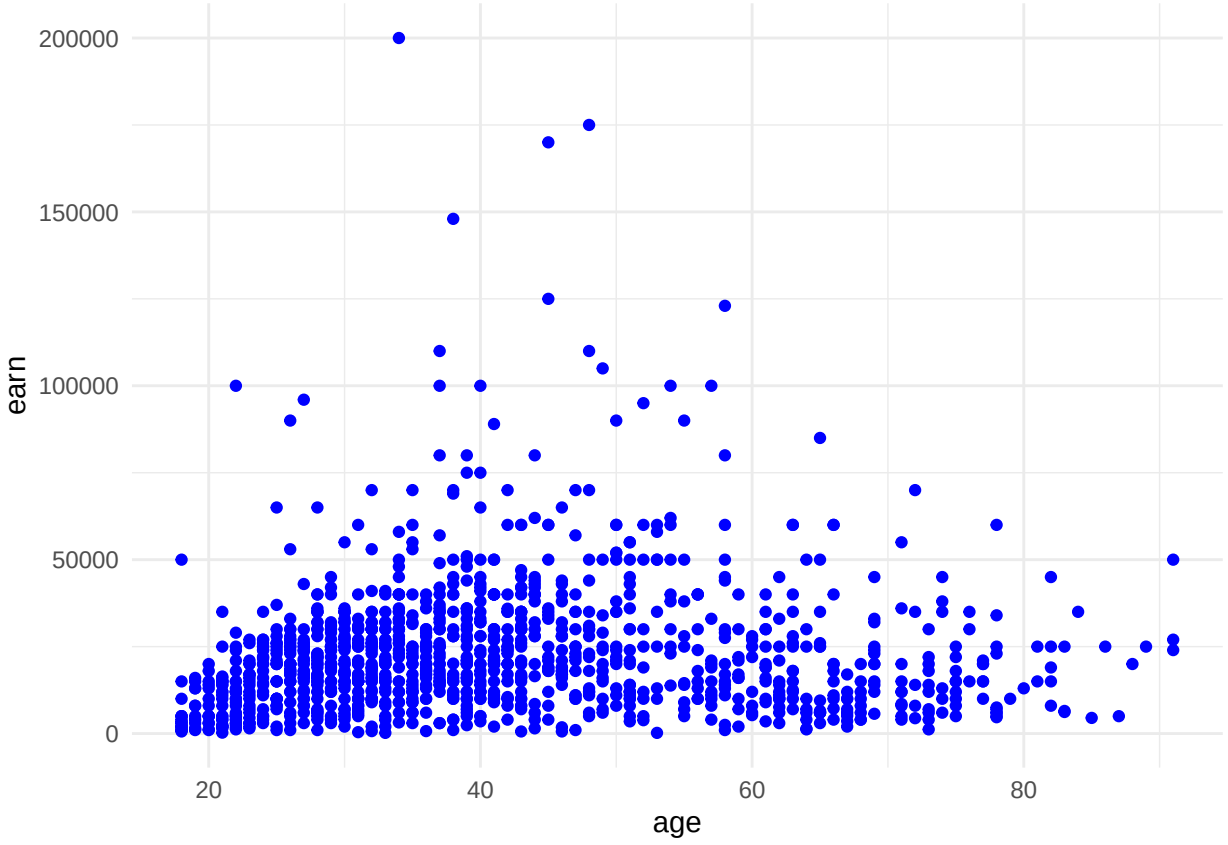
```
##    earn   height    sex ed age  race
## 1 50000 74.42444   male 16  45 white
## 2 60000 65.53754 female 16  58 white
## 3 30000 63.62920 female 16  29 white
## 4 50000 63.10856 female 16  91 other
## 5 51000 63.40248 female 17  39 white
## 6  9000 64.39951 female 15  26 white
```

https://ggplot2.tidyverse.org/reference/geom_point.html Using `geom_point()` create three scatterplots for `height` vs. `earn` `age` vs. `earn` `ed` vs. `earn` creating a data object for reusability

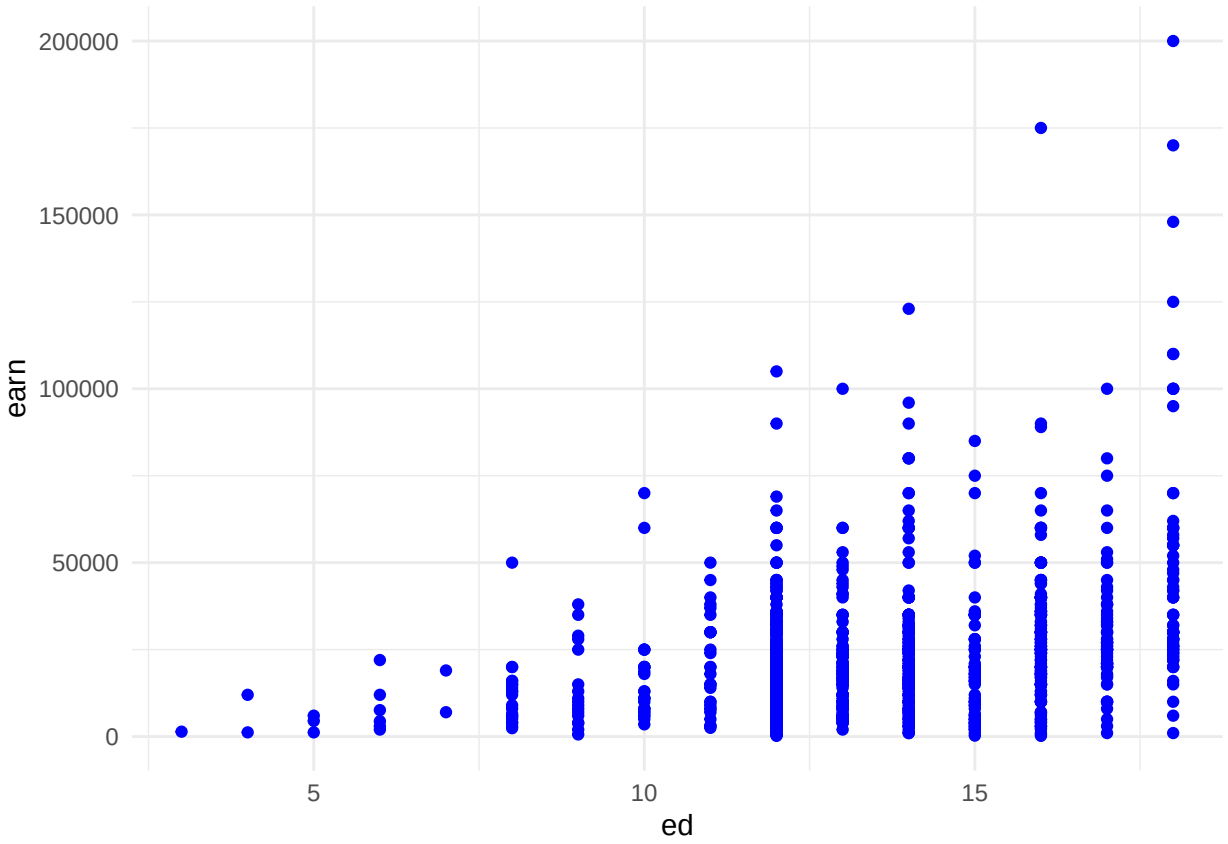
```
data <- ggplot(data = heights_df)
data + geom_point(aes(x=height, y=earn), color = "Blue")
```




```
data + geom_point(aes(x=age, y=earn), color = "Blue")
```



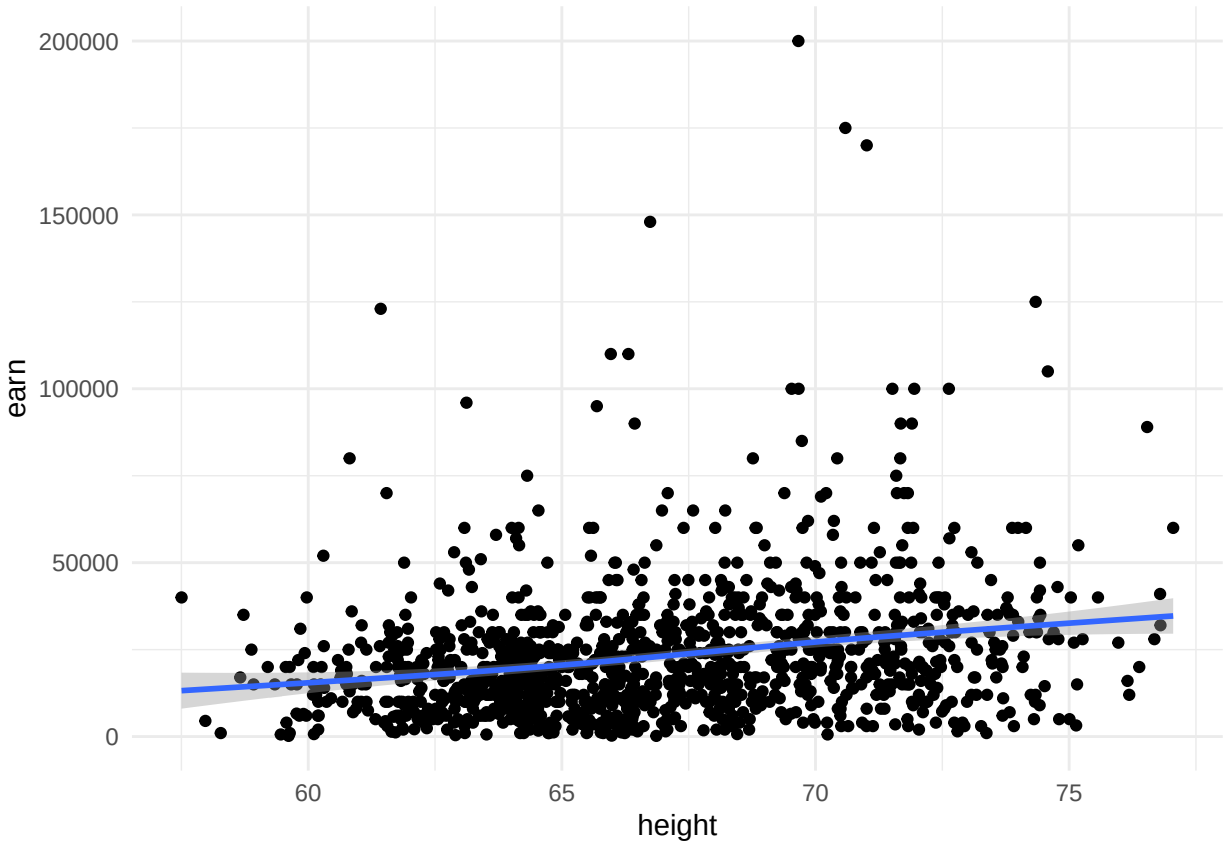

```
data + geom_point(aes(x=ed, y=earn), color = "Blue")
```



Re-create the three scatterplots and add a regression trend line using the `geom_smooth()` function `height` vs. `earn` one way using layering, not sure why geom_smooth() was crying for missing aesthetics x and y. It seems it is not able to inherit from aesthetics in geom_point(), thus I had to pass aethetics separately

```
data + geom_point(aes(x=height, y=earn)) + geom_smooth(aes(x=height, y=earn))
```

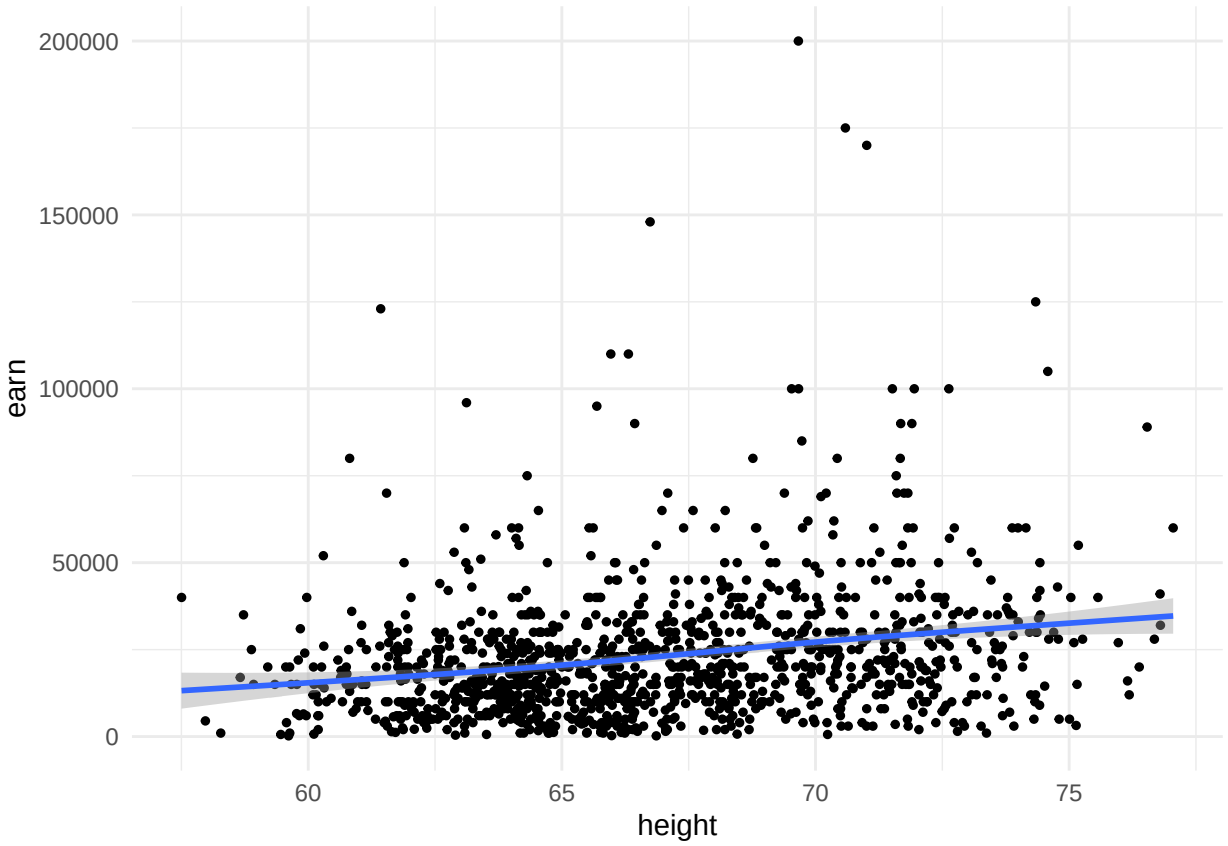
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



another way, geom_smooth() interits aesthetics from ggplot() just fine

```
ggplot(data = heights_df, aes(x=height, y=earn)) + geom_point(size = 1) + geom_smooth()
```

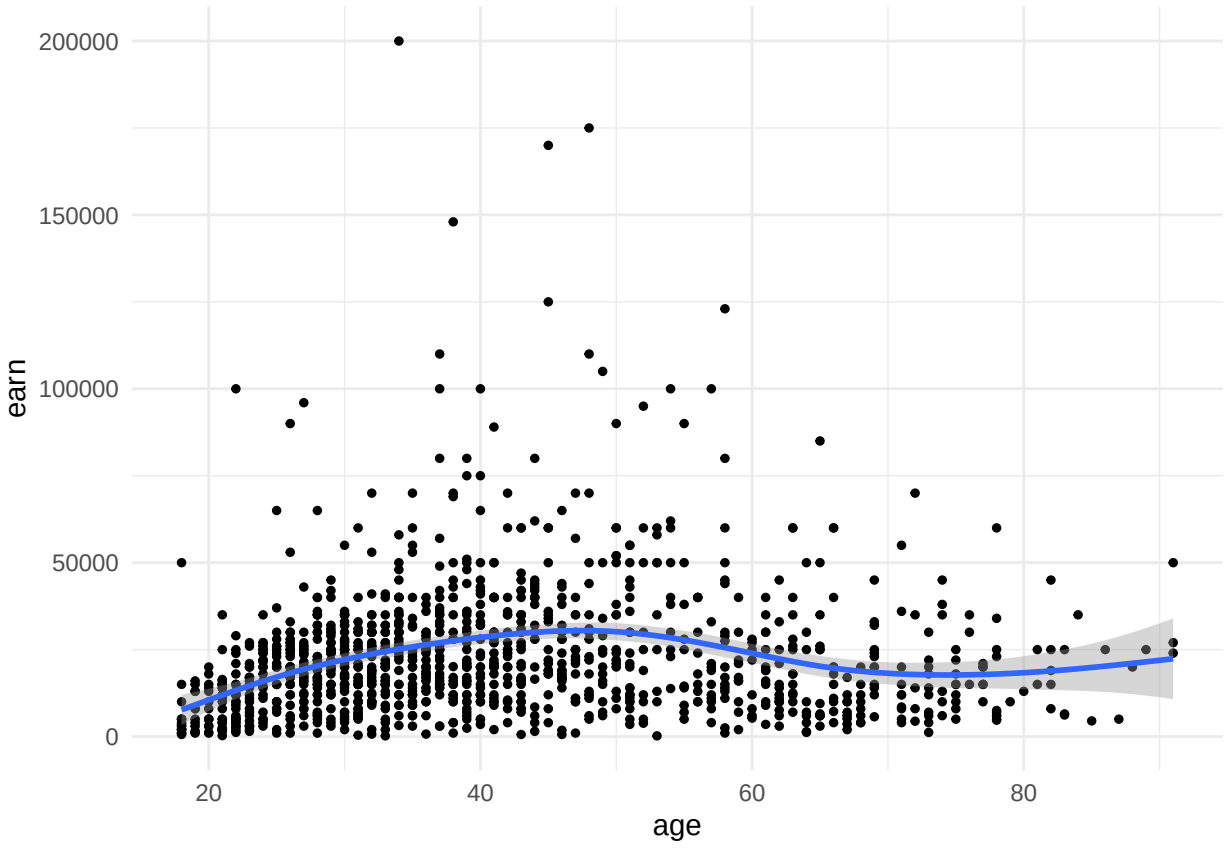
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



age vs. earn

```
ggplot(data = heights_df, aes(x=age, y=earn)) + geom_point(size = 1) + geom_smooth()
```

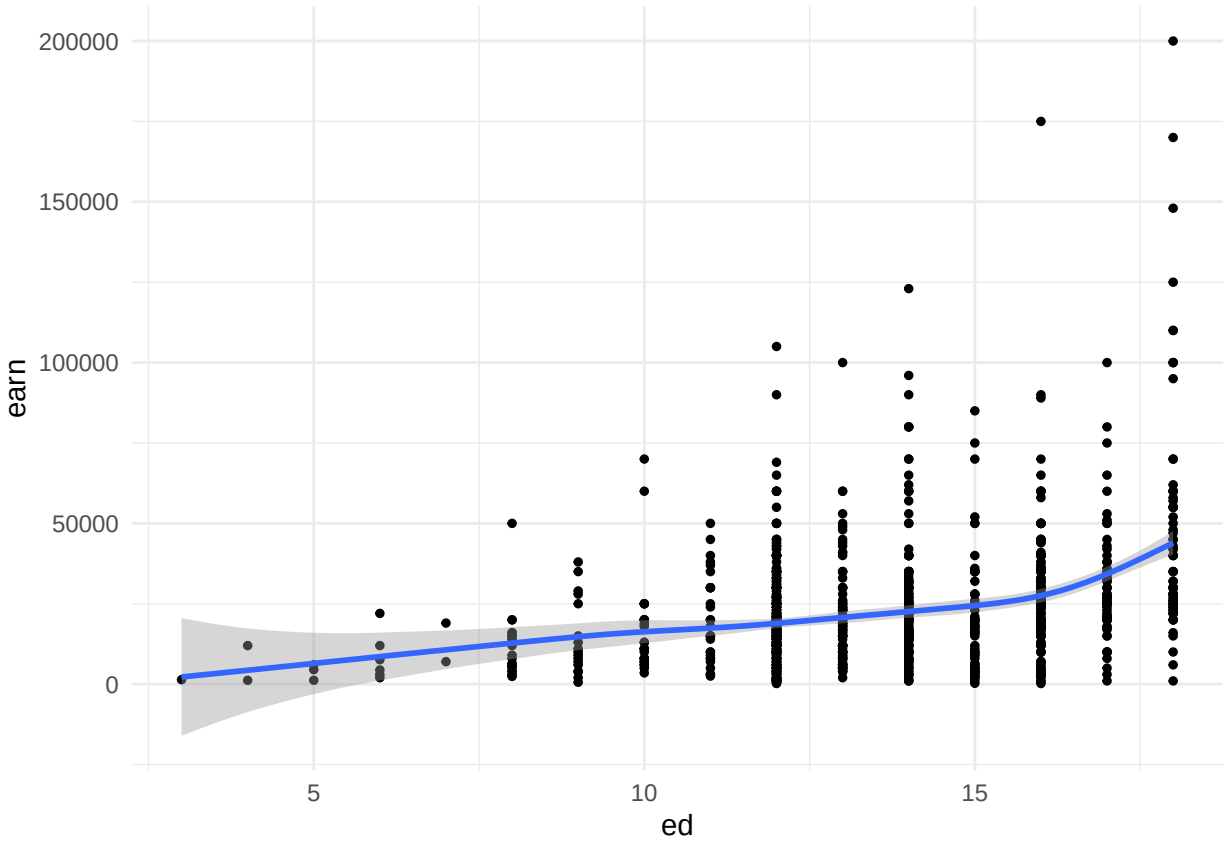
```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



ed vs. earn

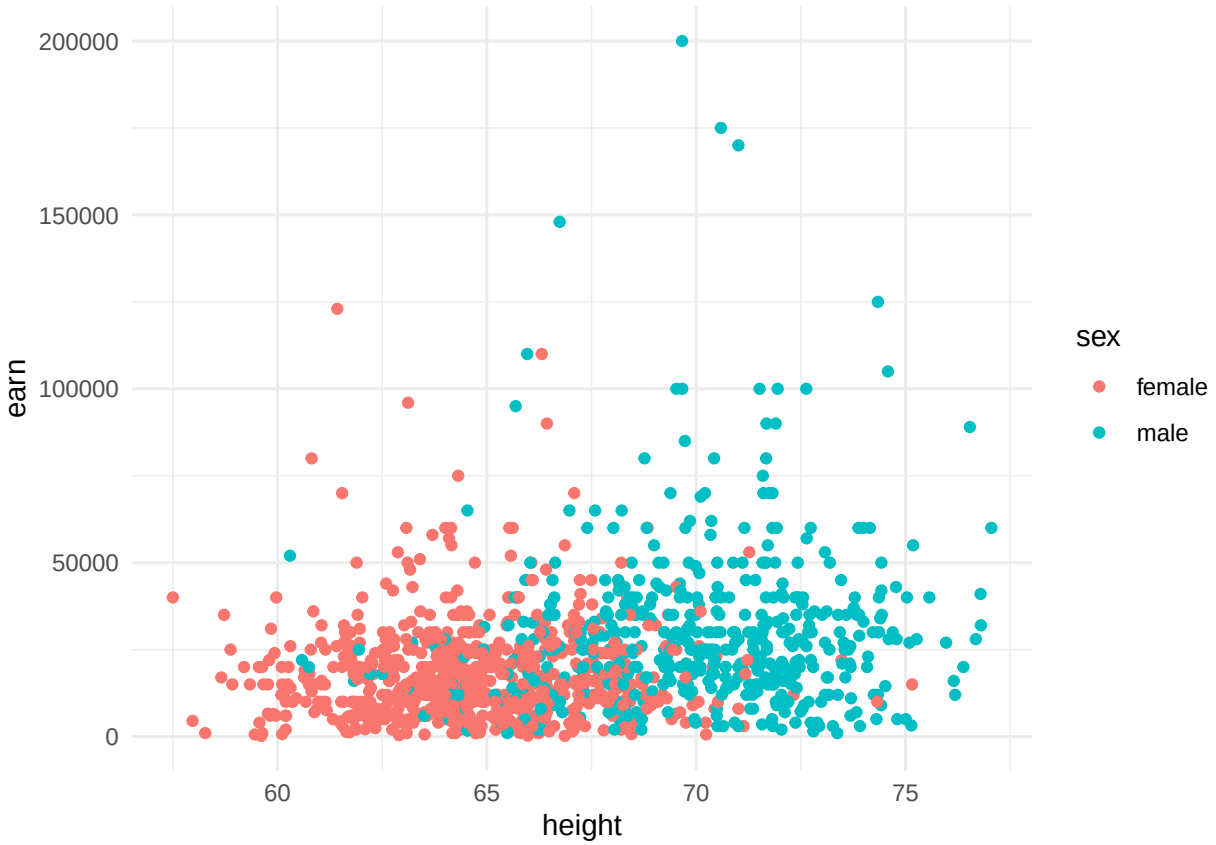
```
ggplot(data = heights_df, aes(x=ed, y=earn)) + geom_point(size = 1) + geom_smooth()
```

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



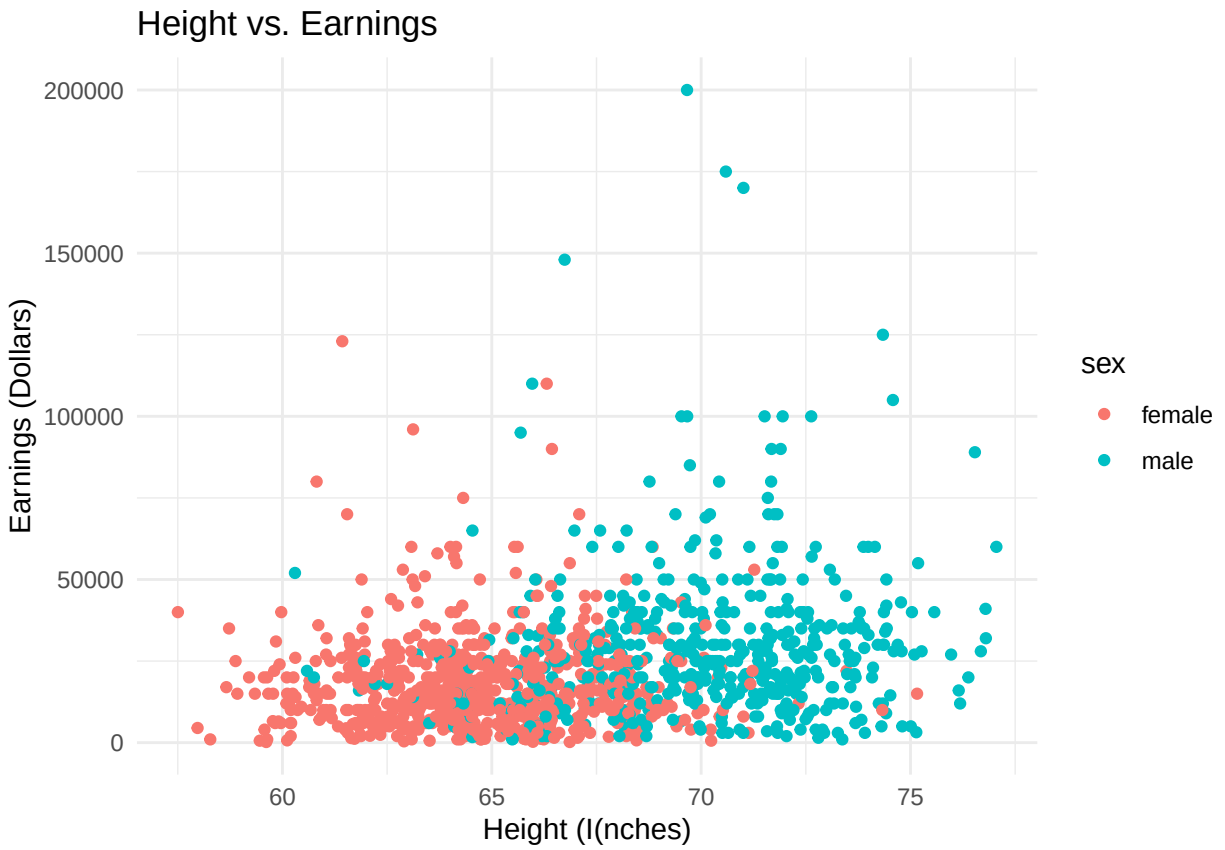
Create a scatterplot of `height`` vs.earn. Usesexas thecol‘ (color) attribute

```
ggplot(data = 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)

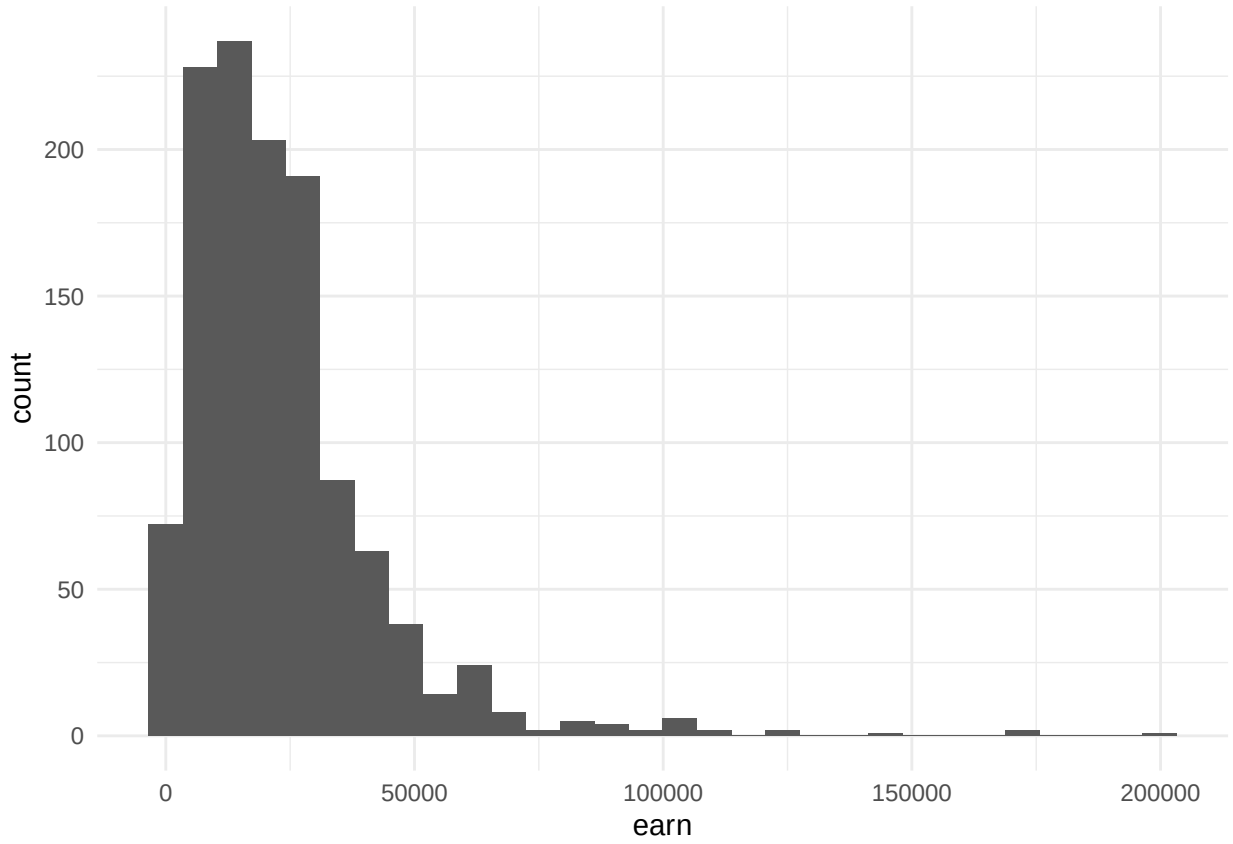
```
ggplot(data = heights_df, aes(x=height, y=earn, col=sex)) +
  geom_point() +
  ggtitle(label = "Height vs. Earnings") +
  xlab(label = "Height (I(nches)") +
  ylab(label = "Earnings (Dollars)")
```



https://ggplot2.tidyverse.org/reference/geom_histogram.html Create a histogram of the `earn` variable using `geom_histogram()`

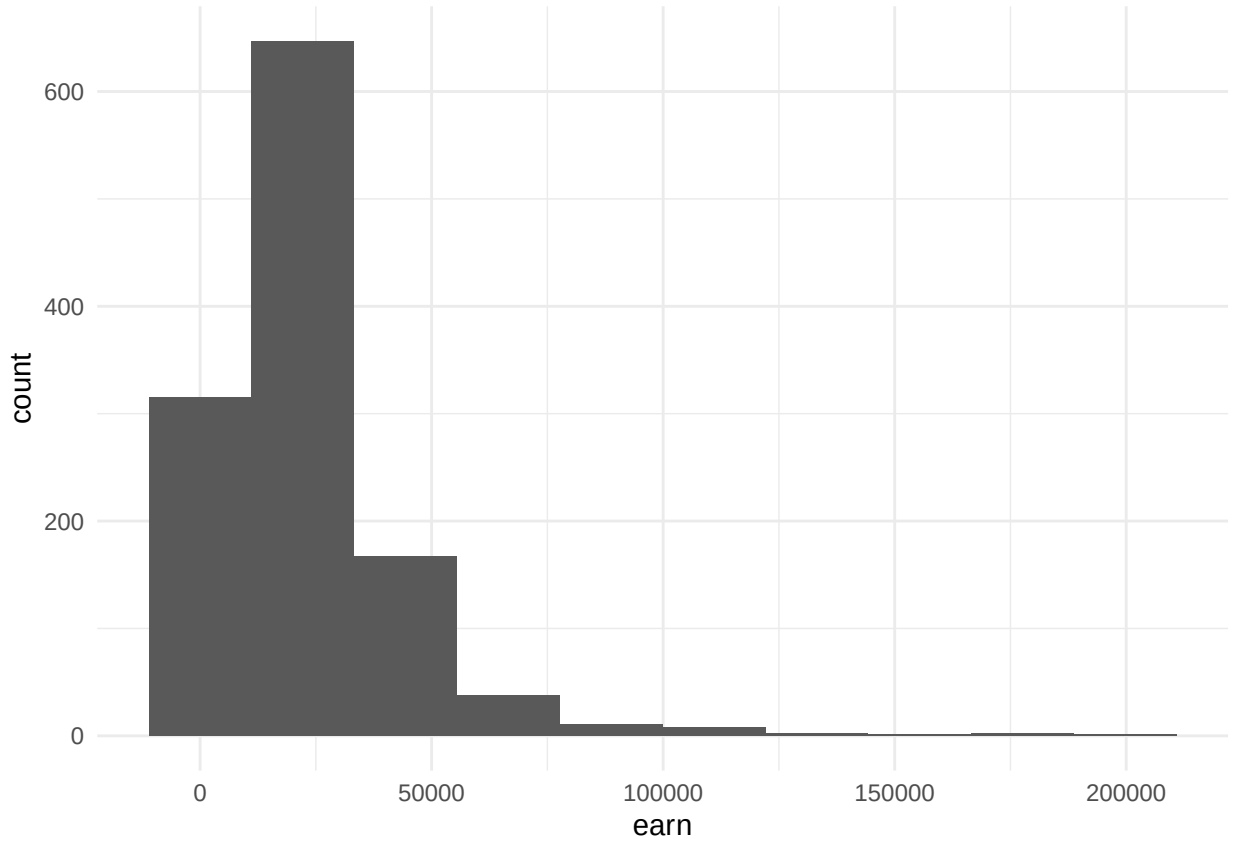
```
ggplot(data = heights_df, aes(x=earn)) + geom_histogram()
```

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



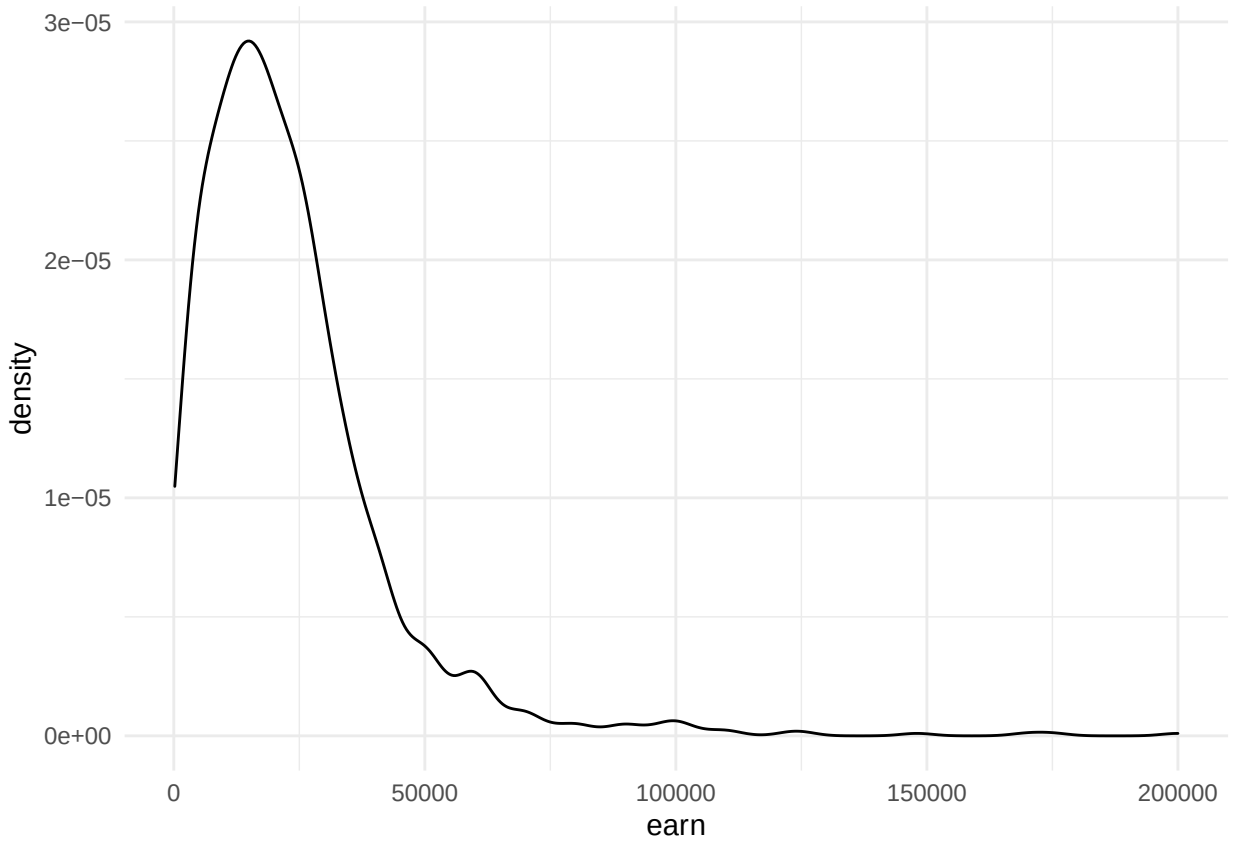
Create a histogram of the `earn` variable using `geom_histogram()` Use 10 bins

```
ggplot(data = heights_df, aes(x=earn)) + geom_histogram(bins=10)
```



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

```
ggplot(data = heights_df, aes(x=earn)) + geom_density()
```


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