

Week3 Assignment 03

Henry Pham

2023-04-02

```
## Load the ggplot2 package

library(ggplot2)
theme_set(theme_minimal())

## Set the working directory to the root of your DSC 520 directory

getwd()

## [1] "C:/R/DSC520"

dir("C://R//DSC520//data")

## [1] "acs-14-1yr-s0201.csv"          "G04ResultsDetail2004-11-02.xls"
## [3] "r4ds"                          "scores.csv"
## [5] "tidynomicon"

if (!file.exists("r4ds"))
{
  # set data working directory
  setwd("C://R//DSC520//data//r4ds")
}

## Load the `data/r4ds/heights.csv`

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

## Using `geom_point()` create three scatterplots for

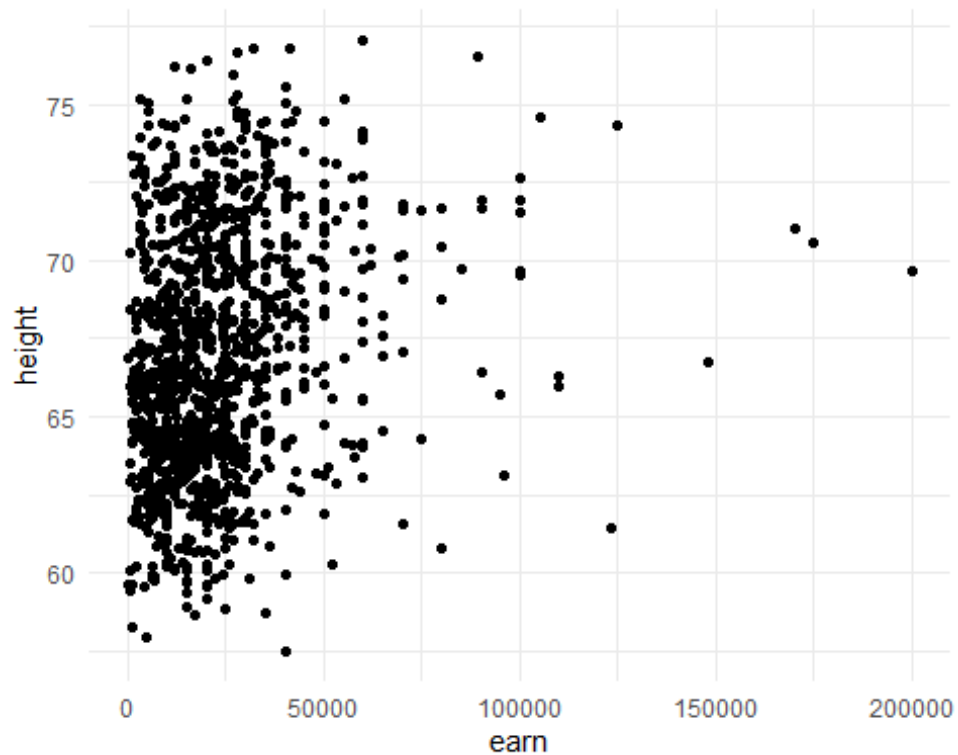
library(ggplot2)

# Load the data
```

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

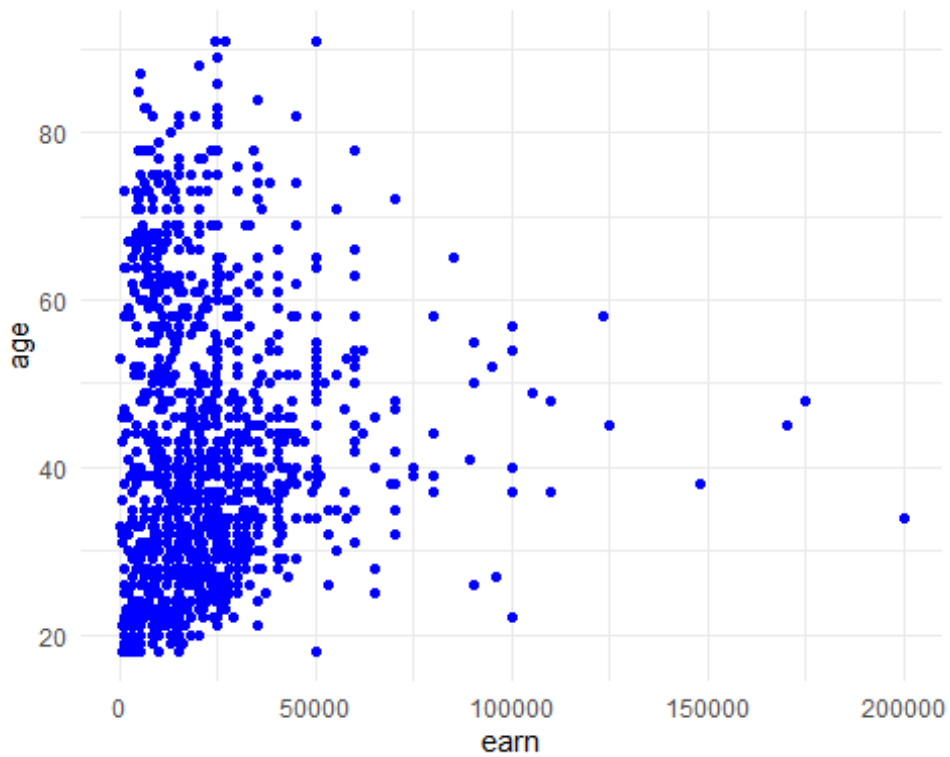
```
# Create the first scatterplot  
## `height` vs. `earn`
```

```
ggplot(heights_df, aes(x = earn, y = height)) + geom_point()
```



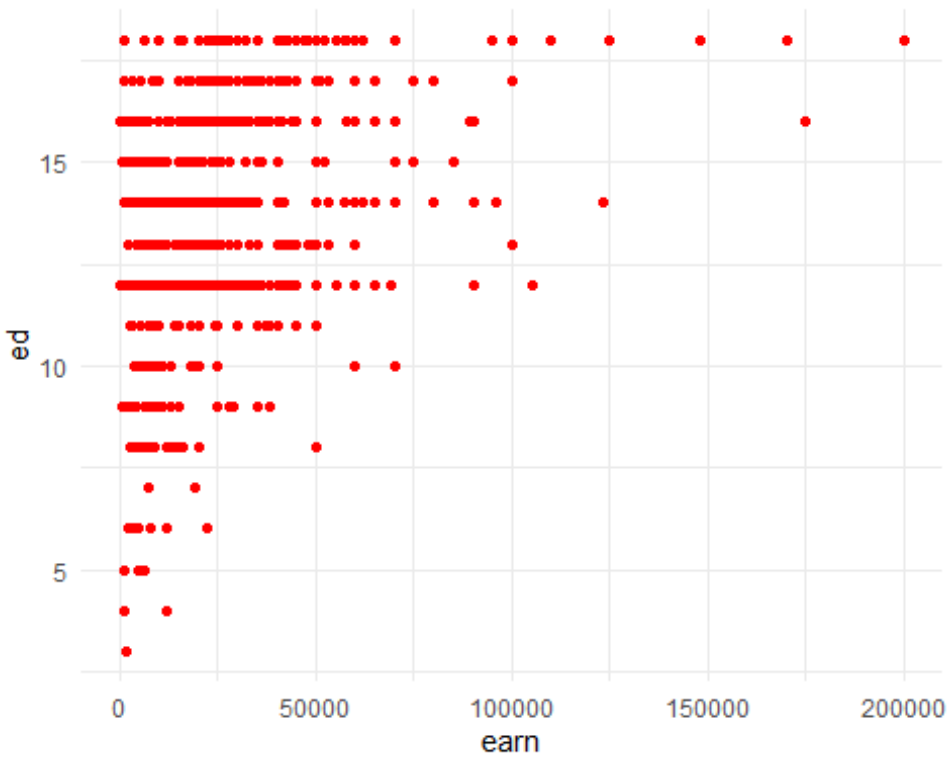
```
## `age` vs. `earn`
```

```
ggplot(heights_df, aes(x = earn, y = age)) + geom_point(color="blue")
```



`ed` vs. `earn`

```
ggplot(heights_df, aes(x = earn, y = ed)) + geom_point(color = "red")
```



```
## Re-create the three scatterplots and add a regression trend line using  
## the `geom_smooth()` function  
## `height` vs. `earn`
```

```
library(ggplot2)
```

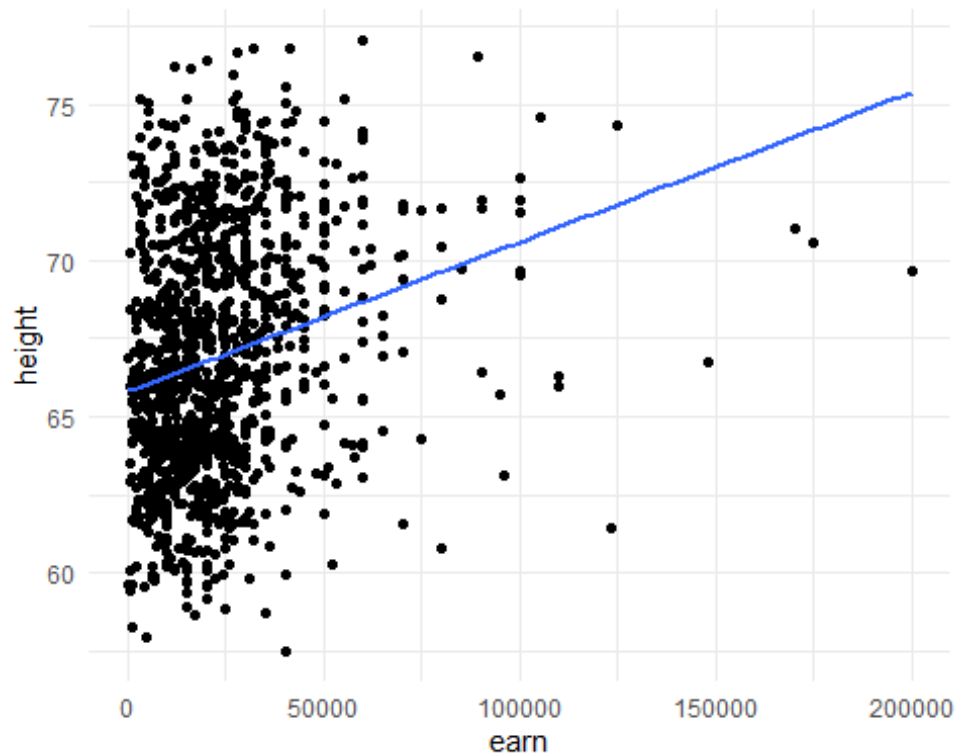
```
# Load the data
```

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

```
# Create the first scatterplot with a regression line
```

```
ggplot(heights_df, aes(x = earn, y = height)) + geom_point() +  
geom_smooth(method = "lm", se = FALSE)
```

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

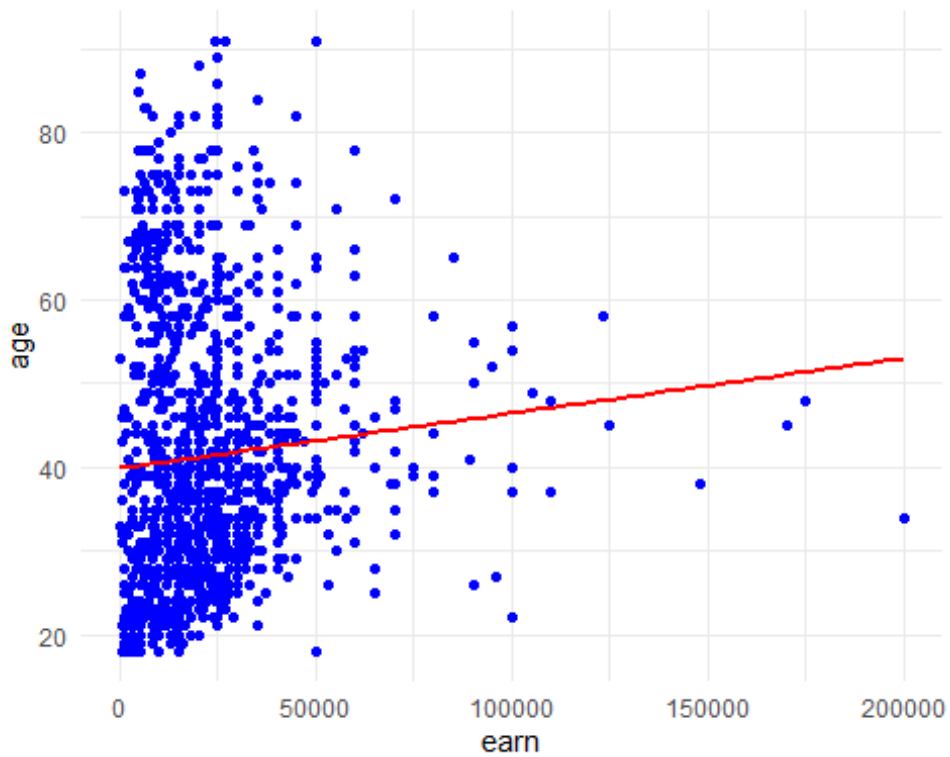


```
# Create the second scatterplot with a regression line
```

```
## `age` vs. `earn`
```

```
ggplot(heights_df, aes(x = earn, y = age)) + geom_point(color = "blue") +  
geom_smooth(method = "lm", se = FALSE, color = "red")
```

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

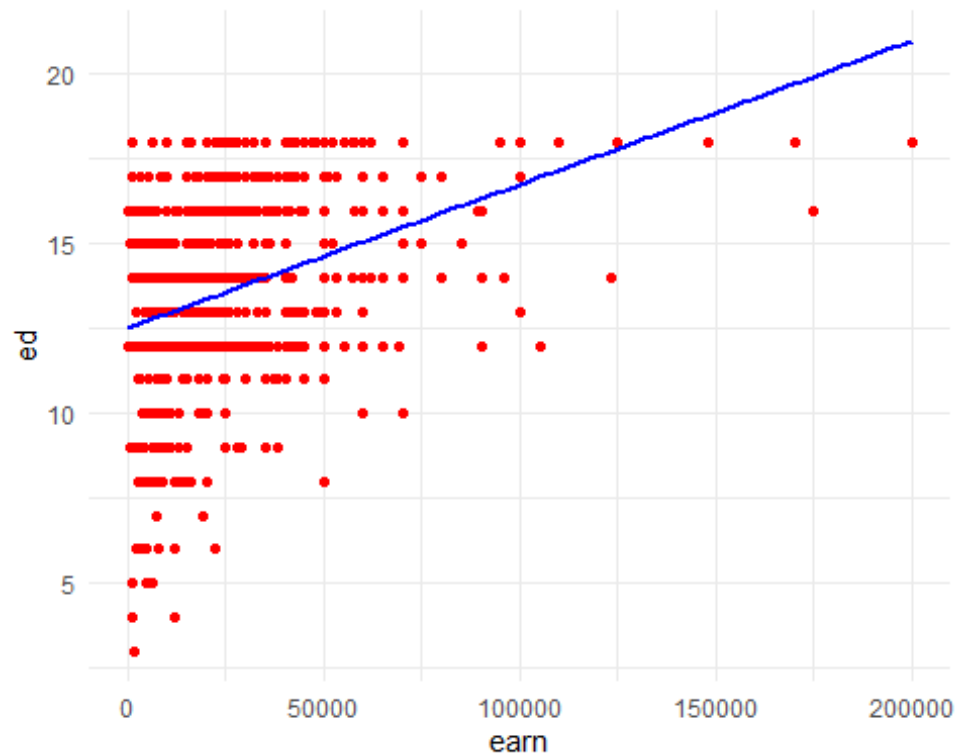


```
# Create the third scatterplot with a regression line
```

```
## `ed` vs. `earn`
```

```
ggplot(heights_df, aes(x = earn, y = ed)) + geom_point(color = "red") +  
  geom_smooth(method = "lm", se = FALSE, color = "blue")
```

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



```
## Create a scatterplot of `height` vs. `earn`.  
## Use `sex` as the `col` (color) attribute  
library(ggplot2)  
  
# Load the data  
heights_df <- read.csv("heights.csv")  
  
# Create the scatterplot with sex as the color attribute  
ggplot(heights_df, aes(x = earn, y = height, 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)
```

```
library(ggplot2)
```

```
# Load the data
```

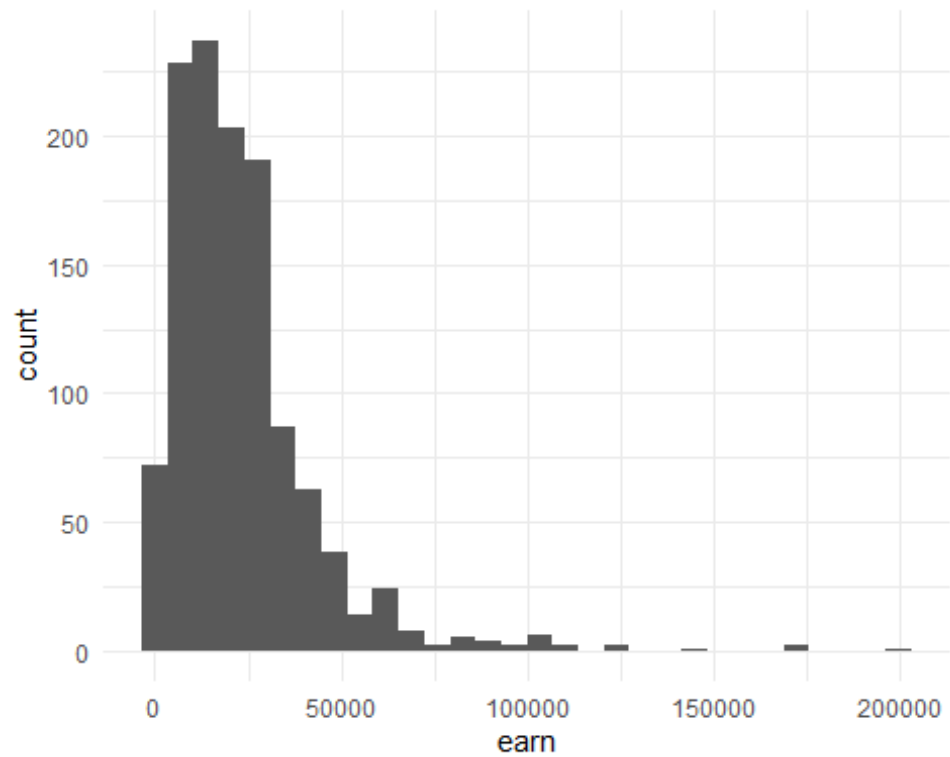
```
heights_df <- read.csv("heights.csv")
```

```
# Create the scatterplot with sex as the color attribute, and add title and  
labels
```

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



```
## Create a histogram of the `earn` variable using `geom_histogram()`  
  
library(ggplot2)  
  
# Load the data  
heights_df <- read.csv("heights.csv")  
  
# Create the histogram of the earn variable  
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
```

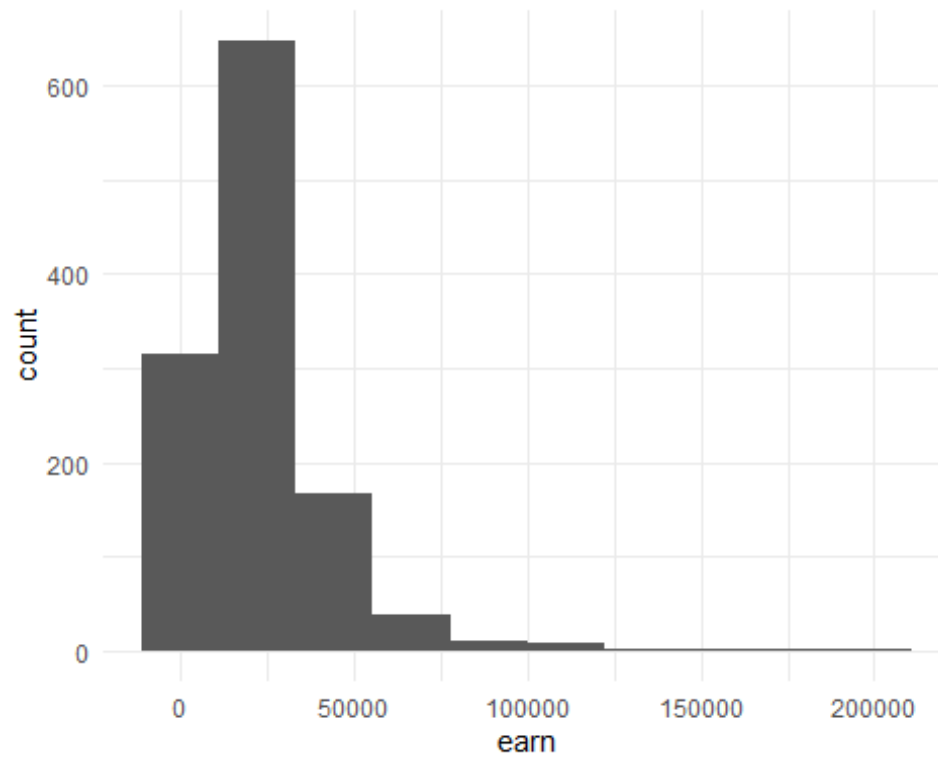
```
library(ggplot2)
```

```
# Load the data
```

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

```
# Create the histogram of the earn variable with 10 bins
```

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



```
## Create a kernel density plot of `earn` using `geom_density()`  
library(ggplot2)  
# Load the data  
heights_df <- read.csv("heights.csv")  
# Create the kernel density plot of the earn variable  
ggplot(heights_df, aes(x = earn)) + geom_density()
```

