Homework #01: Data Visualization

due January 27, 2021 11:59 PM

Ayush Jain

01/24

Load Packages

```
library(tidyverse)
library(viridis)
library(ggridges)
```

Load Data

```
anes <- read_csv("anes2020_subset.csv")
```

Exercise 1

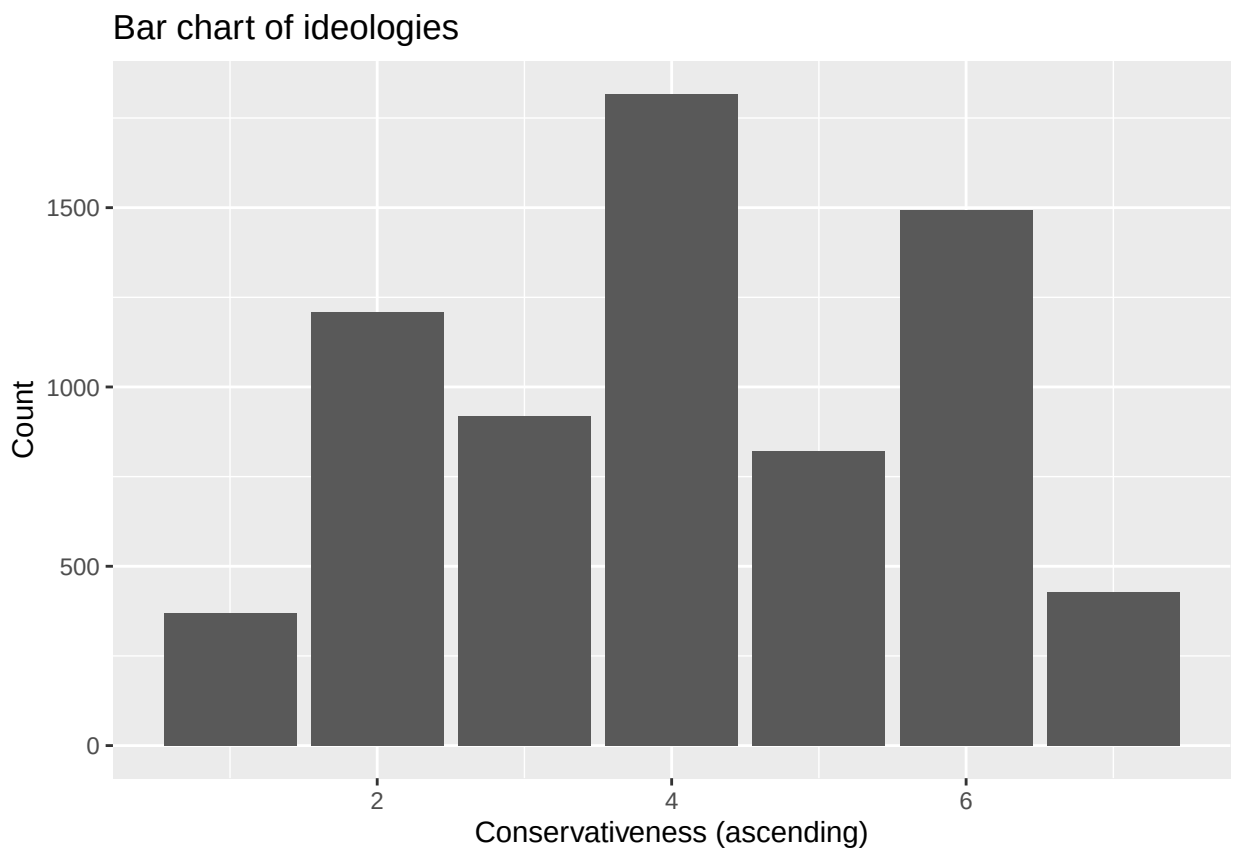
```
glimpse(anes)
```

```
## Rows: 8,280
## Columns: 6
## $ CASEID     <dbl> 200015, 200022, 200039, 200046, 200053, 200060, 200084, 200~
## $ hunt_fish  <dbl> 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0,~
## $ scientists <dbl> 100, 70, 100, 85, 60, 85, 85, NA, 60, 50, 85, 100, 70, 70, ~
## $ education  <dbl> 6, 3, 2, 4, 8, 3, 4, 2, 2, 4, 2, 2, 2, 7, 3, 3, 6, 6, 6, 2,~
## $ ideology   <dbl> 6, 4, 2, 3, 5, 4, 4, NA, NA, NA, NA, 4, 6, 3, 4, 6, 3, 3, 2~
## $ urbanrural <chr> "city", "suburb", "suburb", "small town", "city", "suburb",~
```

Answer: The dataset has 8280 rows and 6 columns

Exercise 2

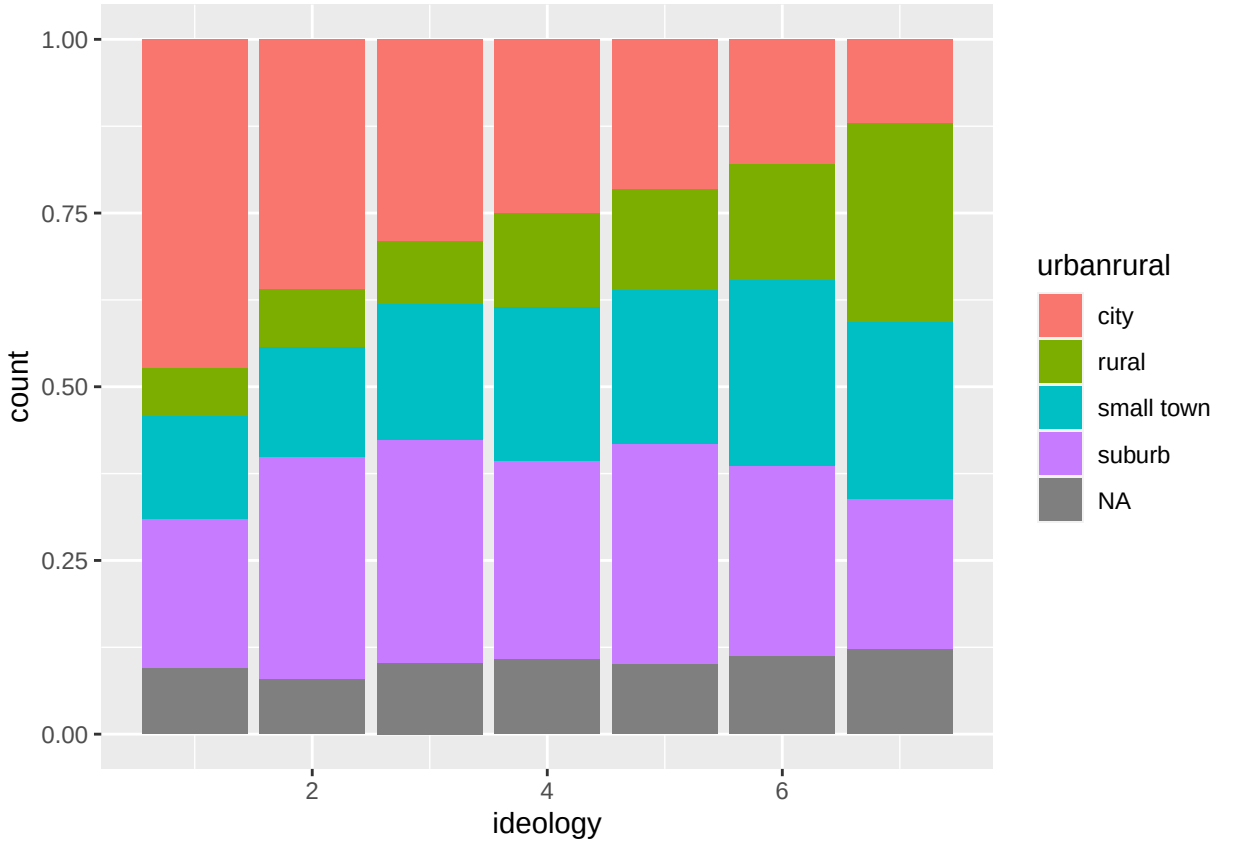
```
ggplot(data = anes,aes(x = ideology)) +
 geom_bar() +
  labs(title = "Bar chart of ideologies",
       x = "Conservativeness (ascending)",
       y = "Count")
```



Answer: Most respondents tend to be ideologically moderate. The most common rating is 4. The lowest ratings are 1 and 7, which indicates that the spread is concentrated around the moderate ratings

Exercise 3

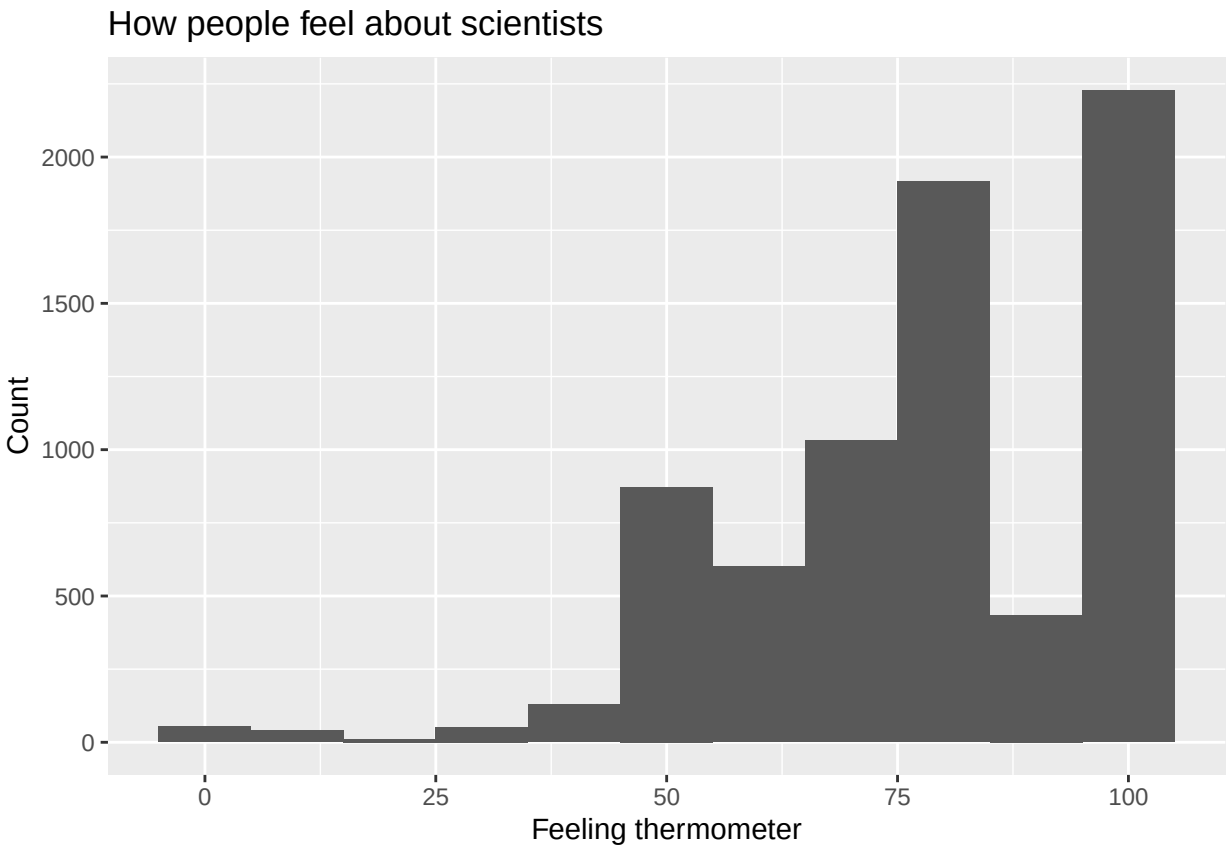
```
ggplot(data = anes,
       aes(x = ideology, fill = urbanrural)) +
 geom_bar(position = "fill")
```



Answer: People with the most liberal ideologies tend to live in cities while people with the most conservative ideologies tend to live in small towns or rural. The people who said NA for where they stay tend to stay the same for all ideologies

Exercise 4

```
ggplot(data = anes,aes(x = scientists)) +
 geom_histogram(binwidth = 10) +
  labs(title = "How people feel about scientists",
       x = "Feeling thermometer",
       y = "Count")
```

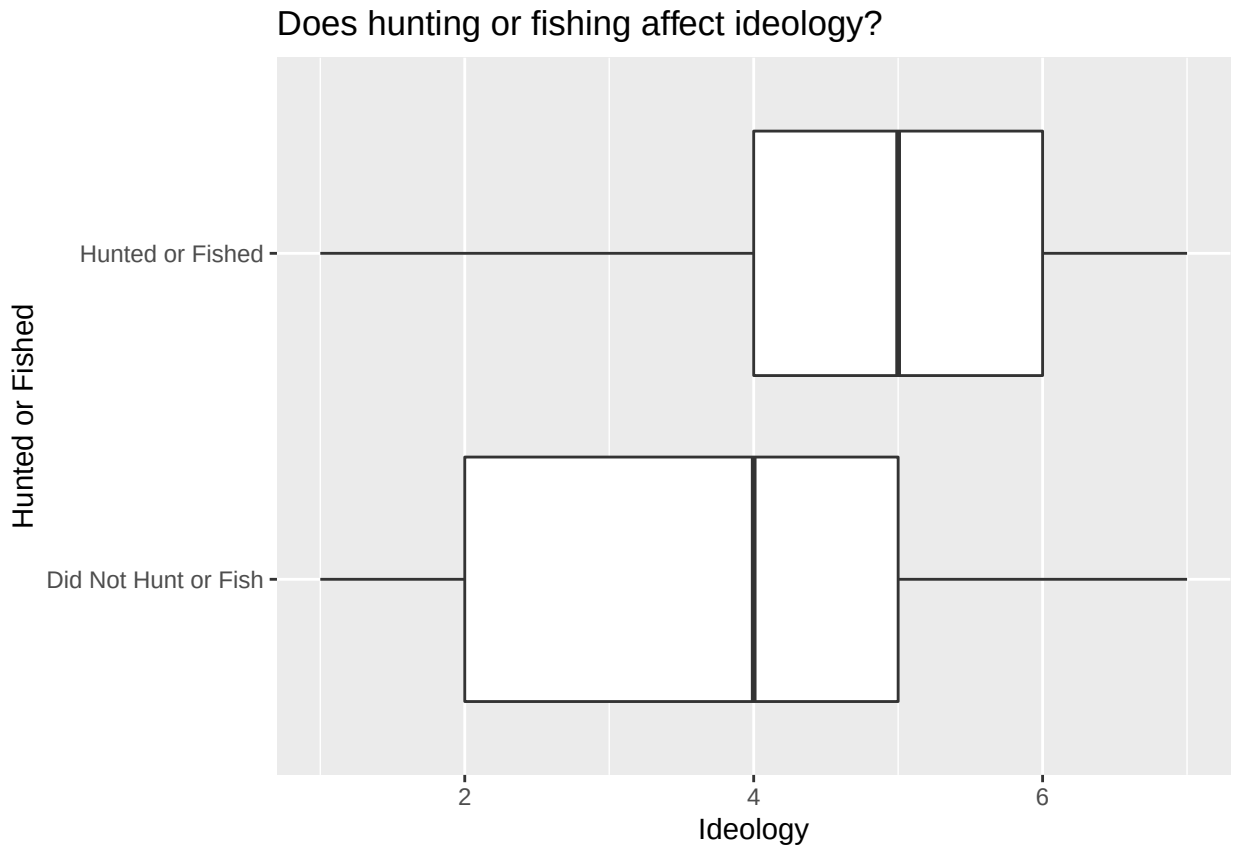


Answer: The graph is generally left skewed because most people feel positively about scientists. There are peaks around the 95-100 and 80-85 marks. Most people feel positively about scientists.

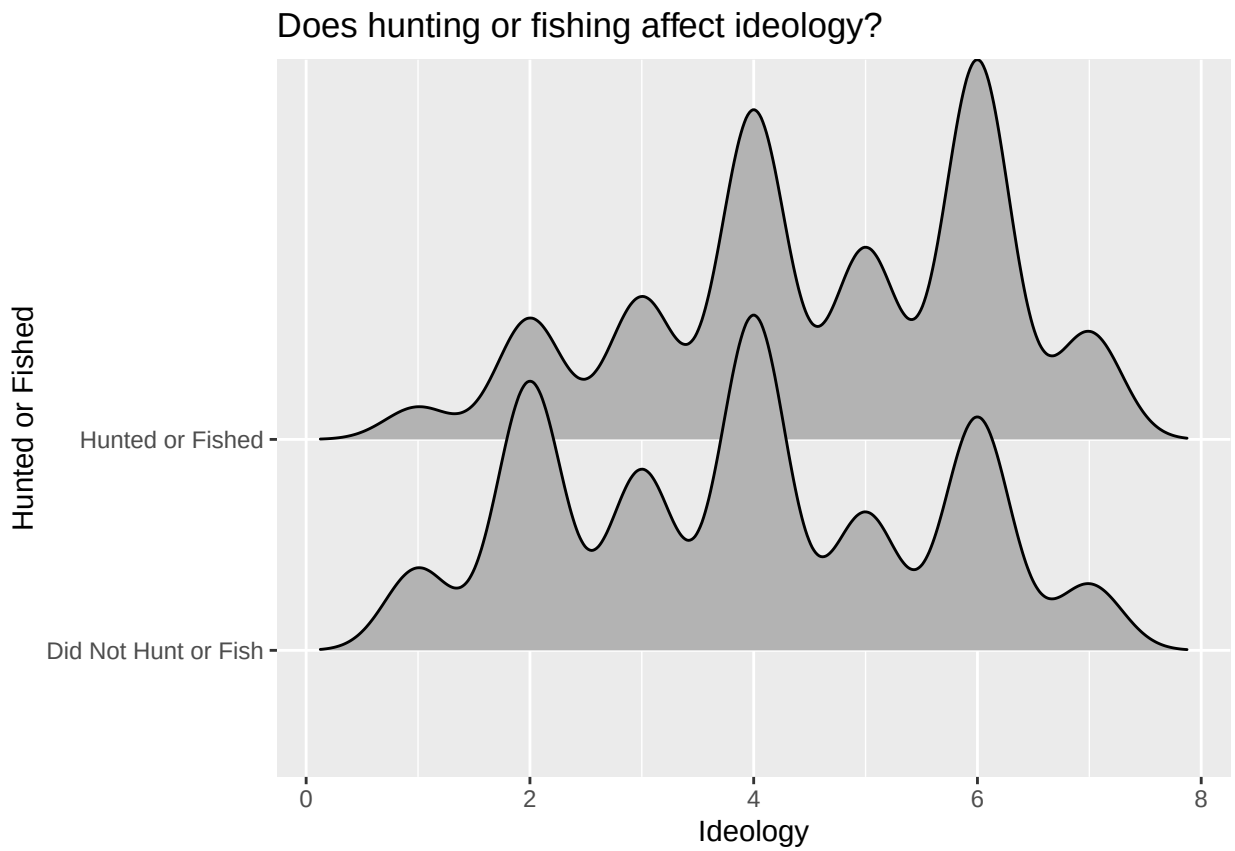
Exercise 5

```
data2 <- anes %>%
  drop_na(hunt_fish) %>%
  mutate(hunted_fished = ifelse(hunt_fish == 0,
                                "Did Not Hunt or Fish",
                                "Hunted or Fished"))

ggplot(data = data2, mapping = aes(x = ideology,
                                   y = hunted_fished)) +
 geom_boxplot() +
  labs(title = "Does hunting or fishing affect ideology?",
       x = "Ideology", y = "Hunted or Fished")
```



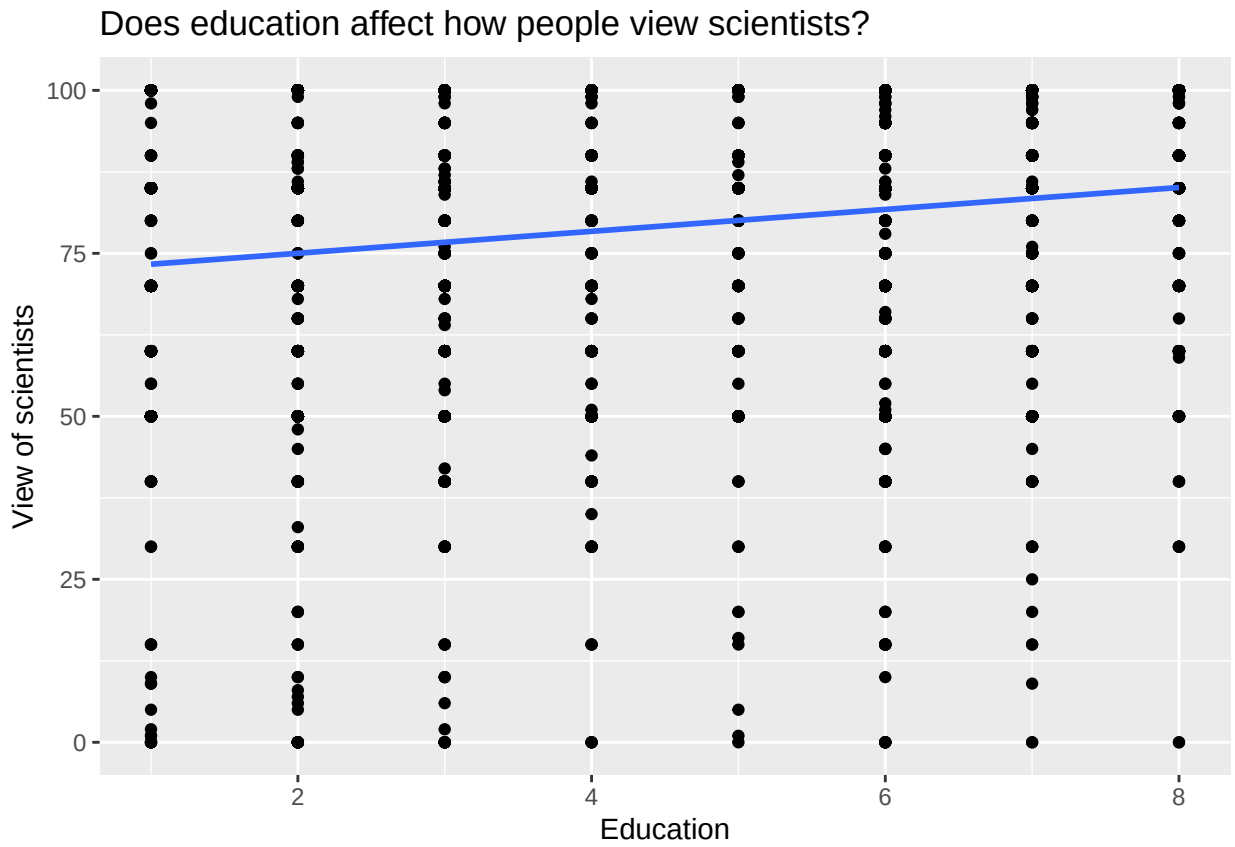

```
ggplot(data = data2,
       mapping = aes(x = ideology, y = hunted_fished)) +
 geom_density_ridges() +
  labs(title = "Does hunting or fishing affect ideology?",
       x = "Ideology", y = "Hunted or Fished")
```



Answer: It seems like people who have not gone hunting or fishing not only encompass a larger range of ideologies, but are also more likely to be liberal on average. The density ridge graph also shows peaks around 4, consistent with the previous observation that most people tend to be moderate. While the boxplot gives a better idea about the range of the data, the density ridges give a better dynamic view of how the data moves and where it peaks or dips

Exercise 6

```
ggplot(anes, aes(x = education, y = scientists))+
  geom_point()+
  geom_smooth(method=lm, se=FALSE)+
  labs(title = "Does education affect how people view scientists?",
       x = "Education",
       y = "View of scientists")
```

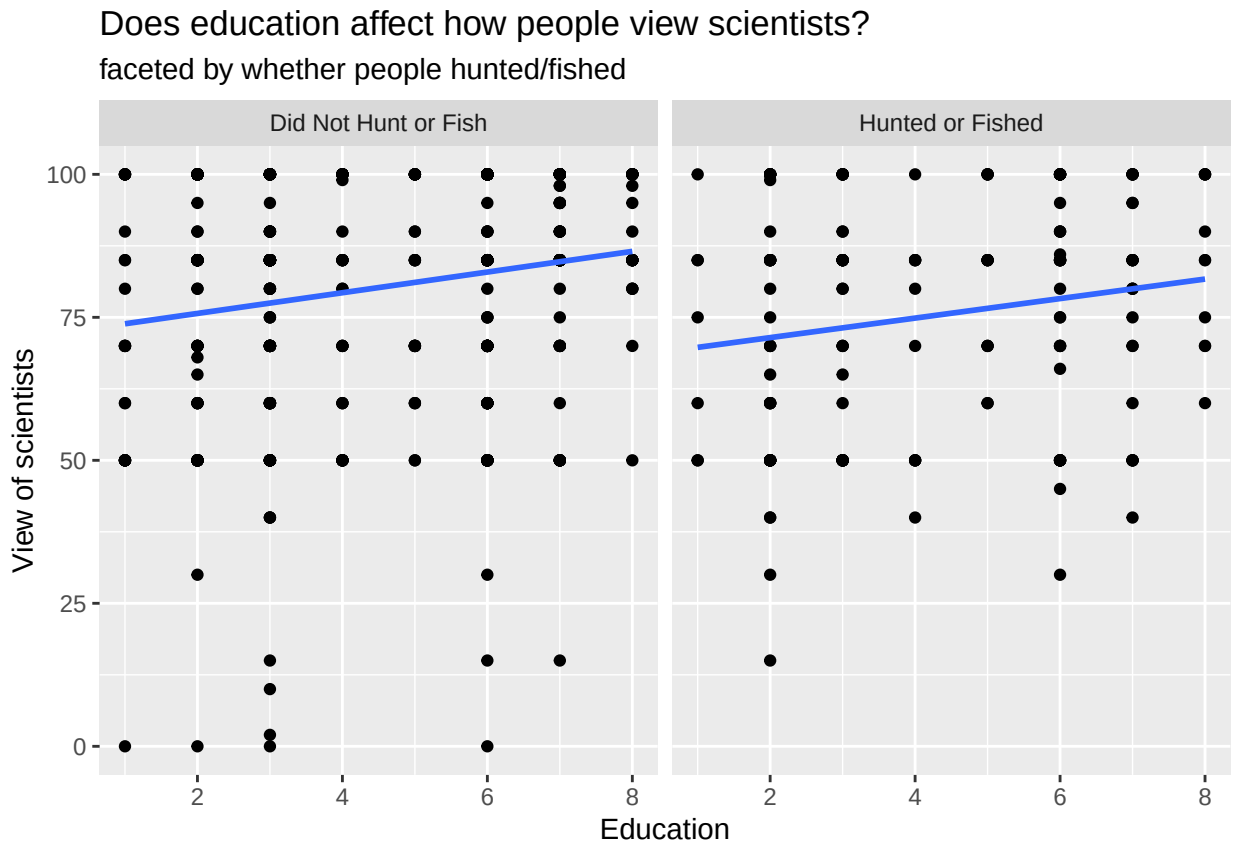


Answer: An increase in education correlates to an increase in view of scientists according to the graph. I find this visualization useful because it shows a clear relation between two variables and allows us to make inferences

Exercise 7

```
set.seed(18)
anes2 <- anes %>%
  sample_frac(.10) %>%
  drop_na(hunt_fish) %>%
  mutate(hunted_fished = ifelse(hunt_fish == 0,
                                "Did Not Hunt or Fish",
                                "Hunted or Fished"))

ggplot(anes2, aes(x = education, y = scientists))+
  geom_point()+
  geom_smooth(method=lm, se=FALSE)+
  labs(title = "Does education affect how people view scientists?",
       subtitle = "faceted by whether people hunted/fished",
       x = "Education",
       y = "View of scientists") +
  facet_wrap(~hunted_fished)
```



Answer: We notice that the line of best fit is lower for people who hunted or fished in the past year. Although their view of scientists still increases with education, it is still lower on average