

DSC520_Week6_Assignment00

Reenie Christudass

2022-07-15

```
library(ggplot2)
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.1
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
## Load the ggplot2 package
```

```
theme_set(theme_minimal())
```

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

```
setwd("C:/Users/chris/dsc520")
```

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

```
heights_df <- read.csv("C:/Users/chris/dsc520/data/r4ds/heights.csv")
```

```
head(heights_df)
```

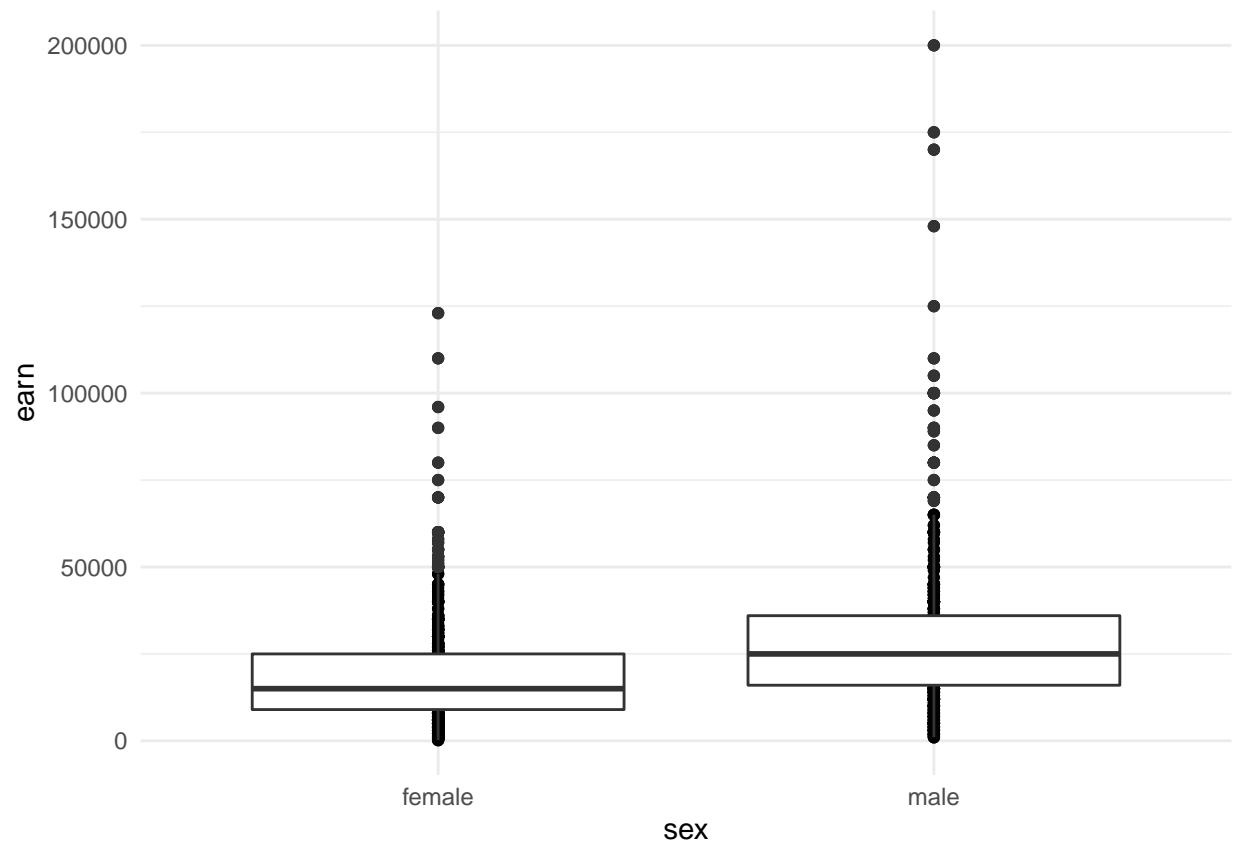
```
##   earn  height  sex ed age race
## 1 50000 74.42444 male 16 45 white
## 2 60000 65.53754 female 16 58 white
## 3 30000 63.62920 female 16 29 white
## 4 50000 63.10856 female 16 91 other
## 5 51000 63.40248 female 17 39 white
## 6  9000 64.39951 female 15 26 white
```

```
# https://ggplot2.tidyverse.org/reference/geom_boxplot.html
```

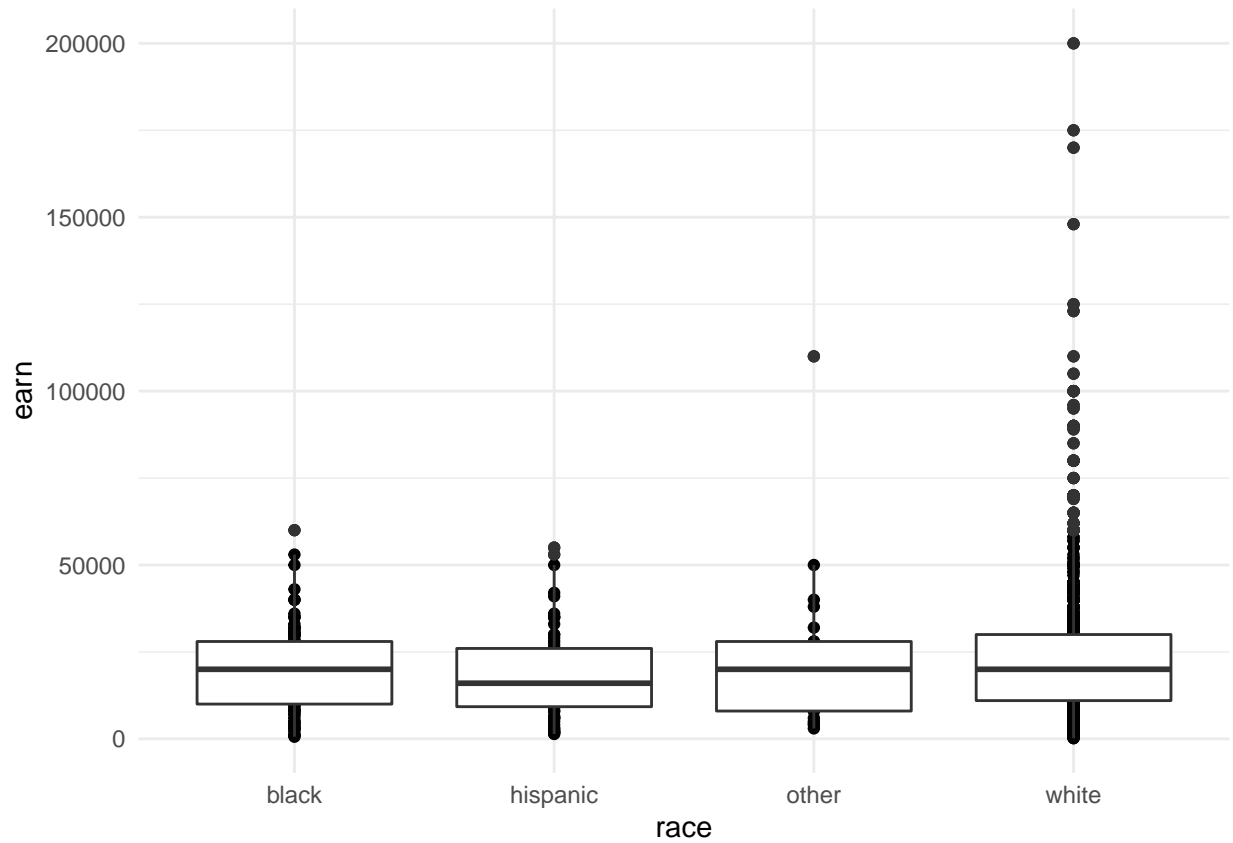
```
## Create boxplots of sex vs. earn and race vs. earn using `geom_point()` and `geom_boxplot()`
```

```
## sex vs. earn
```

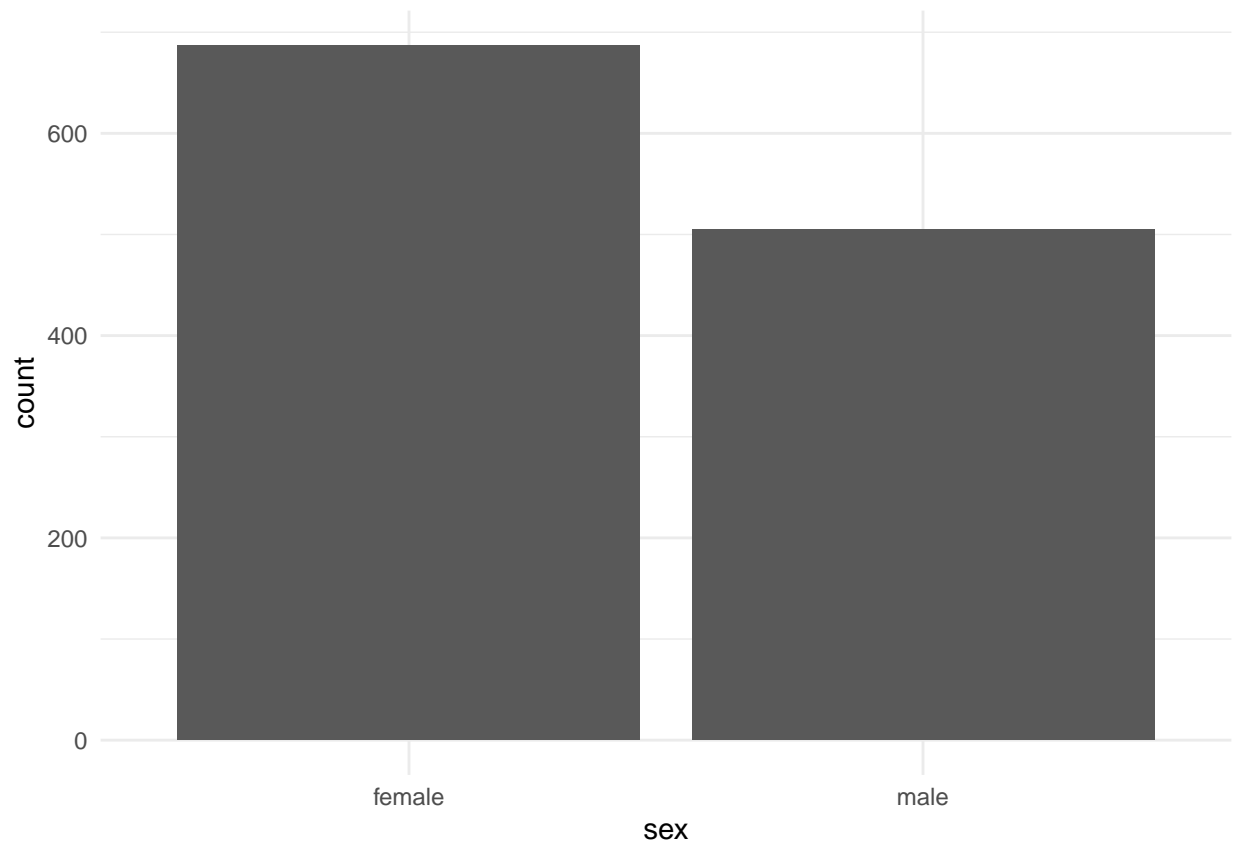
```
ggplot(heights_df, aes(x=sex, y=earn)) + geom_point() + geom_boxplot()
```



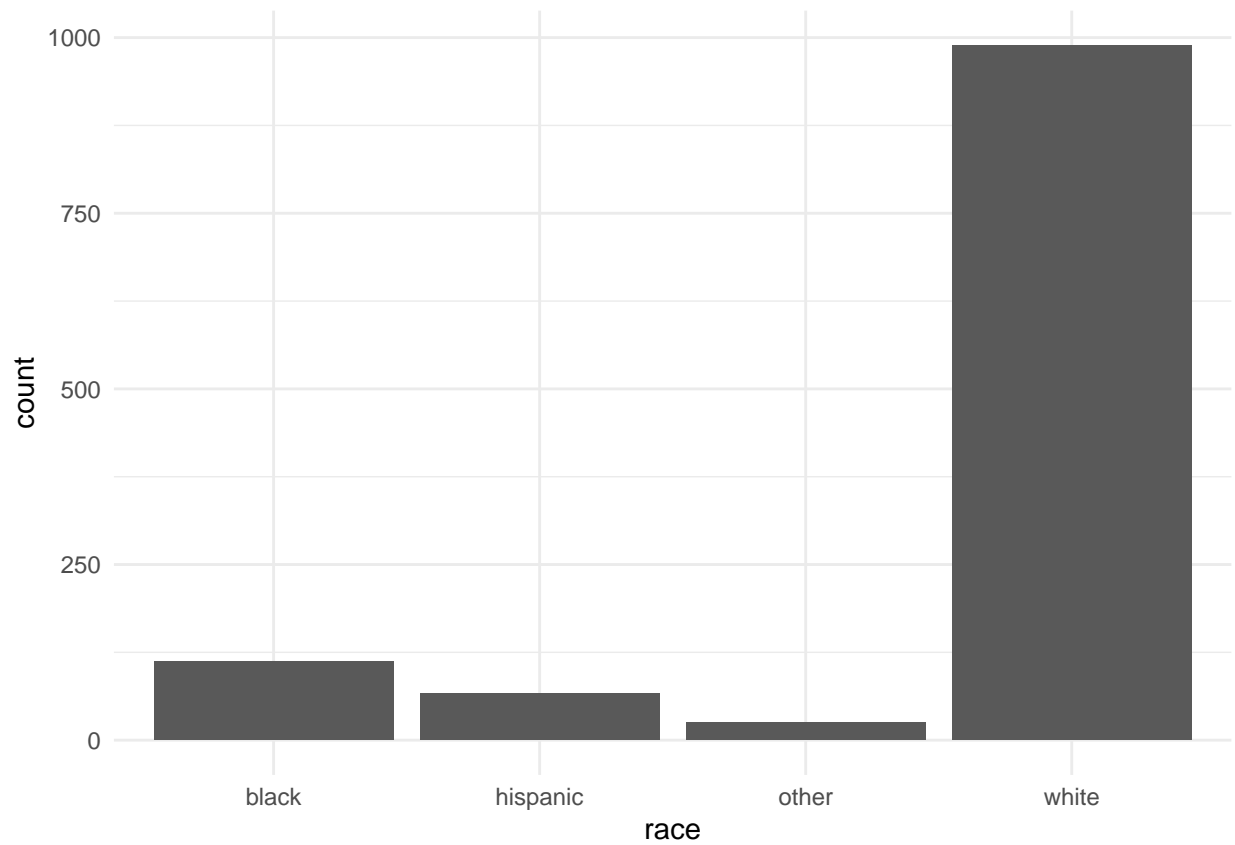
```
## race vs. earn  
ggplot(heights_df, aes(x=race, y=earn)) + geom_point()+ geom_boxplot()
```



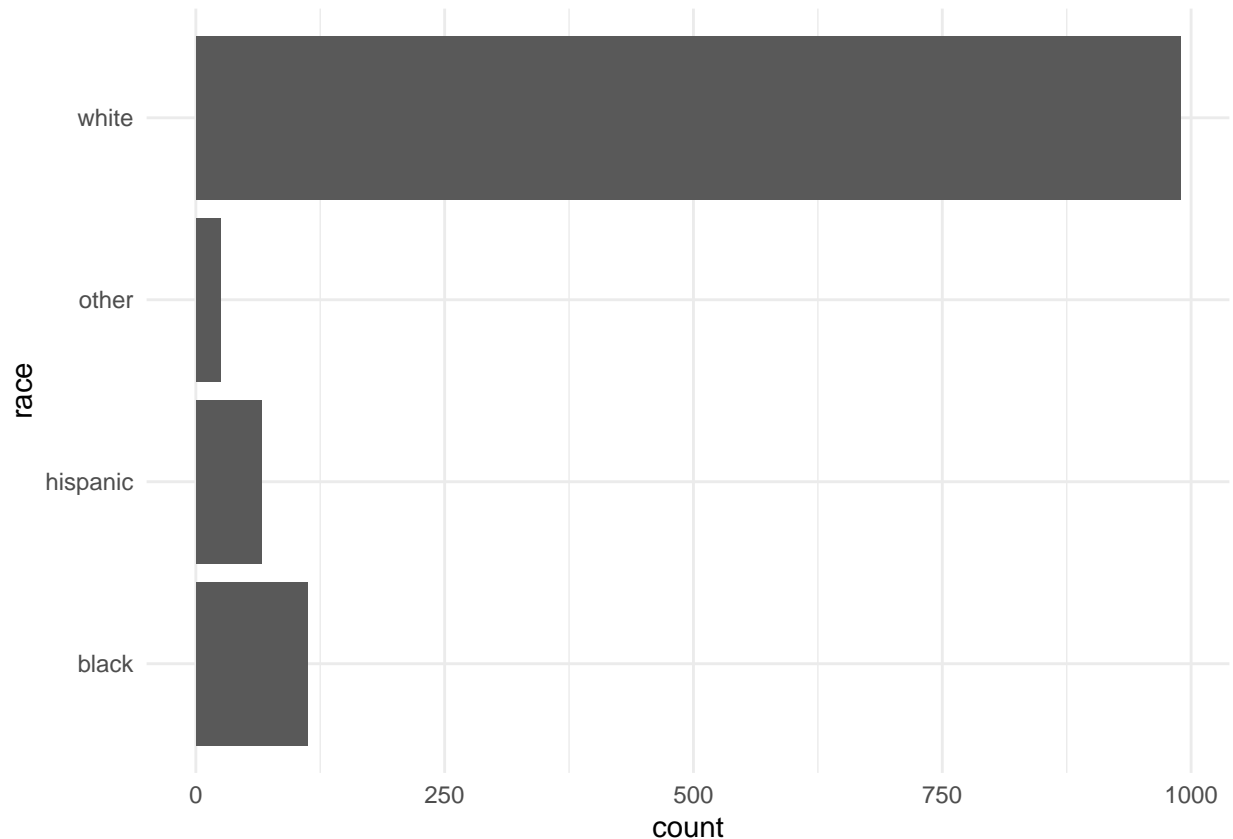
```
# https://ggplot2.tidyverse.org/reference/geom\_bar.html  
## Using `geom_bar()` plot a bar chart of the number of records for each `sex`  
ggplot(heights_df, aes(sex)) + geom_bar()
```



```
## Using `geom_bar()` plot a bar chart of the number of records for each race  
ggplot(heights_df, aes(race)) + geom_bar()
```



```
## Create a horizontal bar chart by adding `coord_flip()` to the previous plot  
ggplot(heights_df, aes(race)) + geom_bar() + coord_flip()
```



```
# https://www.rdocumentation.org/packages/ggplot2/versions/3.3.0/topics/geom\_path
## Load the file `data/nytimes/covid-19-data/us-states.csv` and
## assign it to the `covid_df` dataframe
covid_df <- read.csv("C:/Users/chris/dsc520/data/nytimes/covid-19-data/us-states.csv")
head(covid_df)
```

```
##      date      state fips cases deaths
## 1 2020-01-21 Washington   53     1      0
## 2 2020-01-22 Washington   53     1      0
## 3 2020-01-23 Washington   53     1      0
## 4 2020-01-24  Illinois   17     1      0
## 5 2020-01-24 Washington   53     1      0
## 6 2020-01-25 California    6     1      0
```

```
## Parse the date column using `as.Date()`
covid_df$date <- as.Date(covid_df$date)
```

```
## Create three dataframes named `california_df`, `ny_df`, and `florida_df`
## containing the data from California, New York, and Florida
california_df <- covid_df[ which( covid_df$state == "California"), ]
ny_df <- covid_df[ which( covid_df$state == "New York"), ]
florida_df <- covid_df[ which( covid_df$state == "Florida"), ]
head(california_df)
```

```
##      date      state fips cases deaths
```

```
## 6 2020-01-25 California 6 1 0
## 10 2020-01-26 California 6 2 0
## 14 2020-01-27 California 6 2 0
## 18 2020-01-28 California 6 2 0
## 22 2020-01-29 California 6 2 0
## 26 2020-01-30 California 6 2 0
```

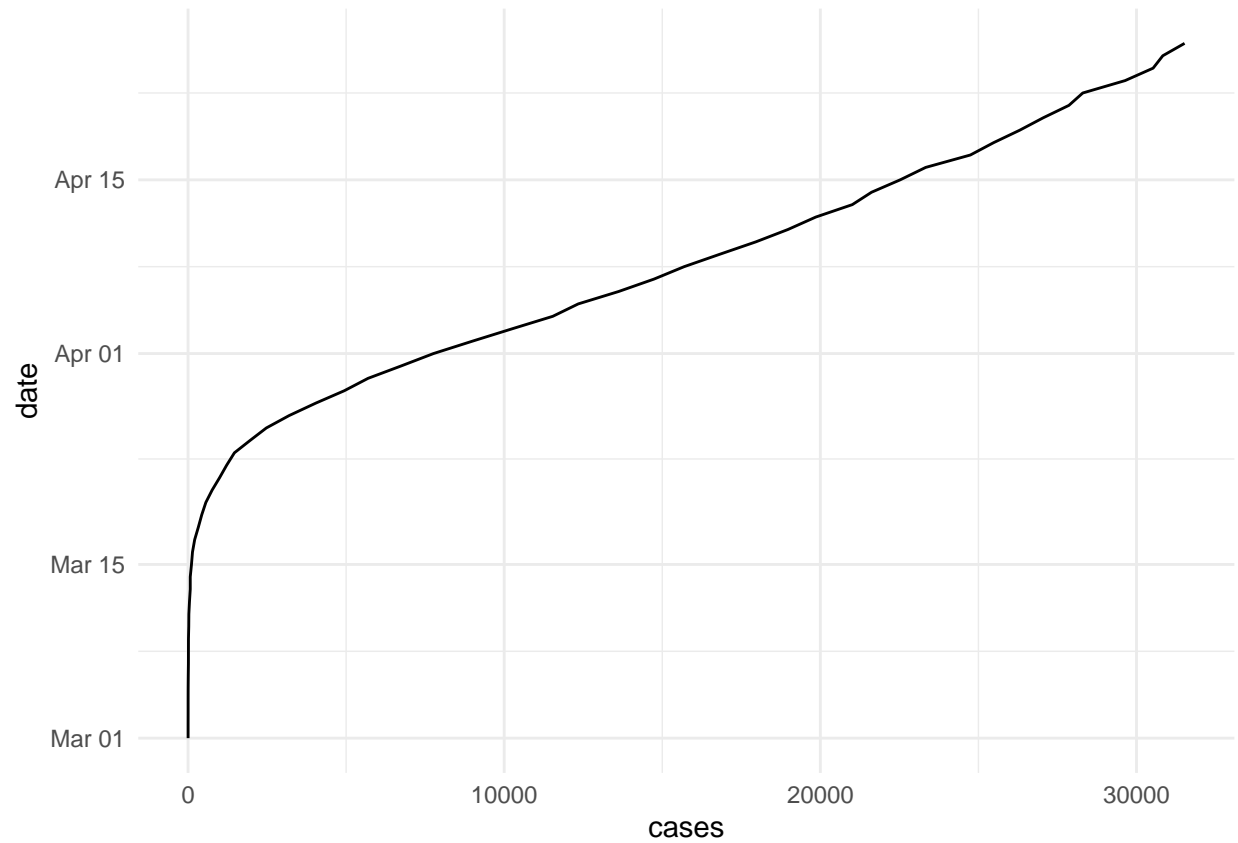
```
head(ny_df)
```

```
##      date      state fips cases deaths
## 247 2020-03-01 New York 36 1 0
## 262 2020-03-02 New York 36 1 0
## 277 2020-03-03 New York 36 2 0
## 294 2020-03-04 New York 36 11 0
## 314 2020-03-05 New York 36 22 0
## 339 2020-03-06 New York 36 44 0
```

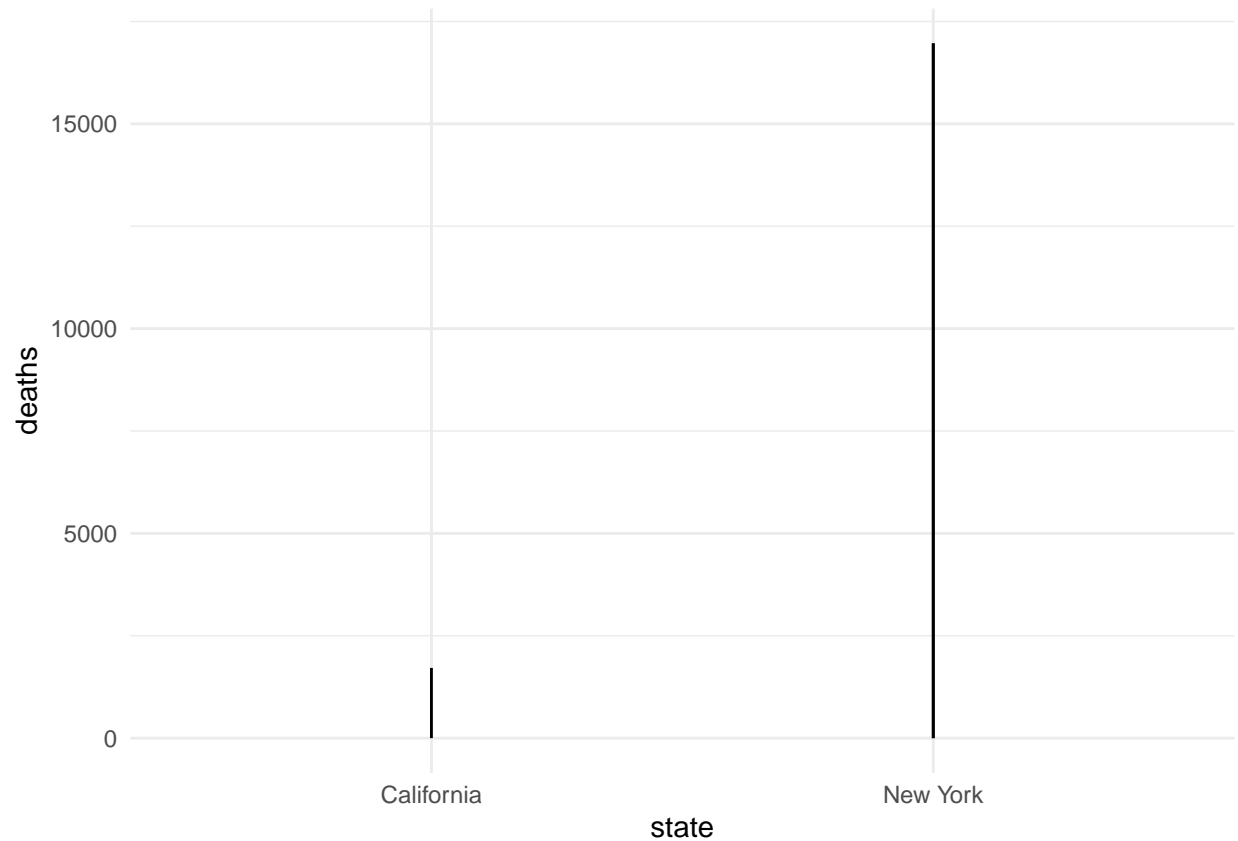
```
head(florida_df)
```

```
##      date      state fips cases deaths
## 243 2020-03-01 Florida 12 2 0
## 256 2020-03-02 Florida 12 2 0
## 271 2020-03-03 Florida 12 3 0
## 287 2020-03-04 Florida 12 3 0
## 305 2020-03-05 Florida 12 4 0
## 326 2020-03-06 Florida 12 7 2
```

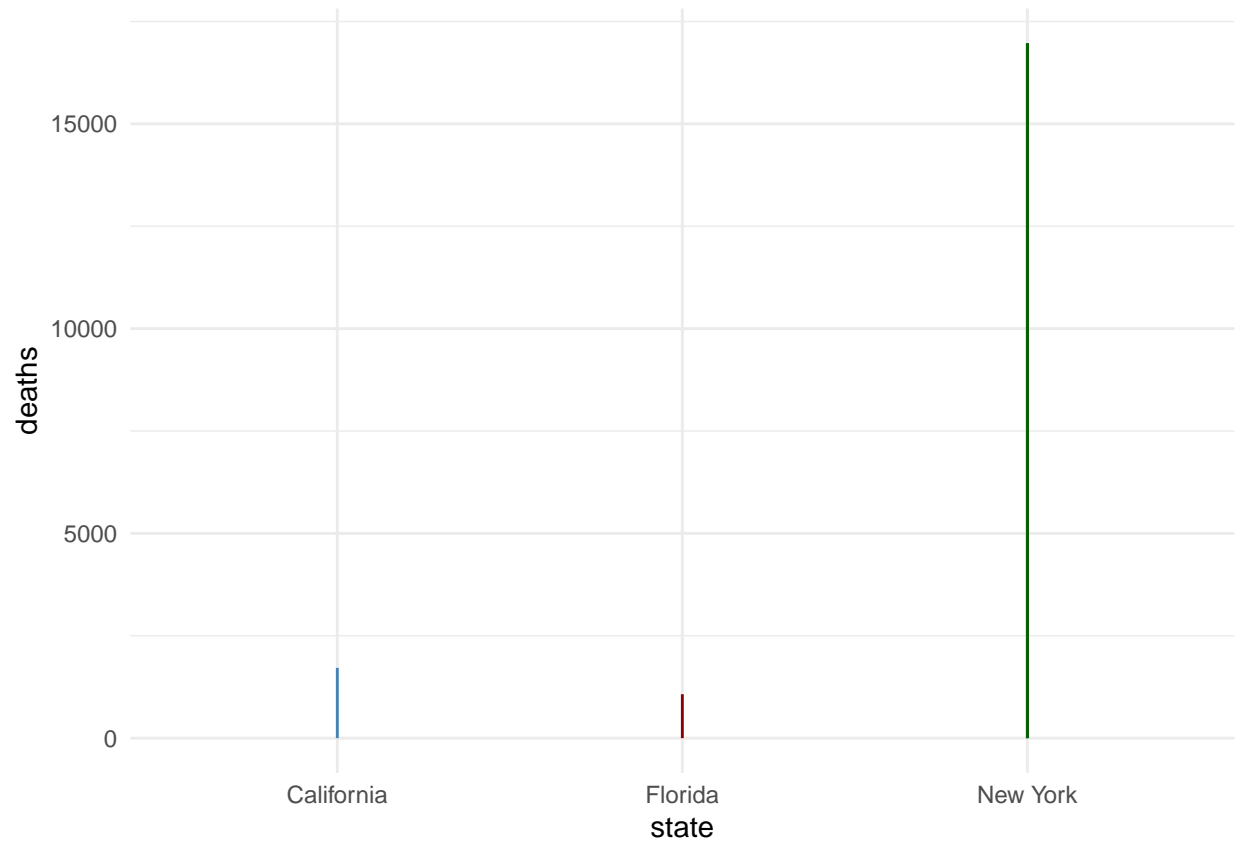
```
## Plot the number of cases in Florida using `geom_line()`
ggplot(data=florida_df, aes(x=cases, y=date, group=1)) + geom_line()
```



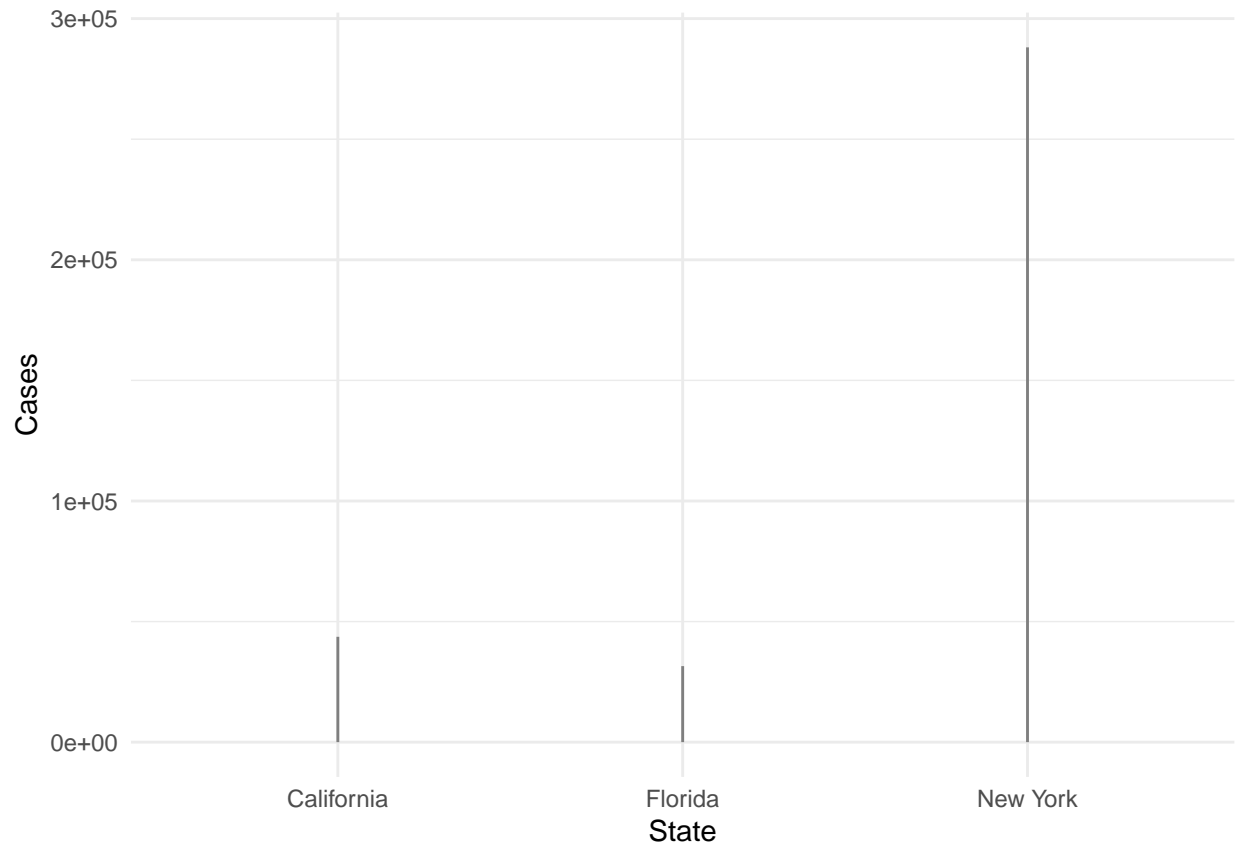
```
## Add lines for New York and California to the plot  
ggplot(data=ny_df, aes(x=state, group=1)) +  
  geom_line(aes(y = deaths)) +  
  geom_line(data=ny_df, aes(y = deaths)) +  
  geom_line(data=california_df, aes(y = deaths))
```

```
## Use the colors "darkred", "darkgreen", and "steelblue" for Florida, New York, and California
ggplot(data=ny_df, aes(x=state, group=1)) +
  geom_line(aes(y = deaths)) +
  geom_line(data=ny_df, aes(y = deaths) ,color="darkgreen") +
  geom_line(data=california_df, aes(y = deaths) ,color="steelblue")+
  geom_line(data=florida_df, aes(y = deaths),color="darkred")
```



```
## Add a legend to the plot using `scale_colour_manual`  
## Add a blank (" ") label to the x-axis and the label "Cases" to the y axis  
  
ggplot(data=ny_df, aes(x=state, group=1)) +  
  geom_line(aes(y = cases, colour = "New York")) +  
  geom_line(data=california_df, aes(y = cases, colour="California")) +  
  geom_line(data=florida_df, aes(y = cases, colour="Florida")) +  
  scale_colour_manual(values = c("darkgreen", "steelblue", "darkred") , breaks = c(9000, 1000000, 10000000))  
  xlab("State") + ylab("Cases")
```



```
## Scale the y axis using `scale_y_log10()`  
ggplot(data=ny_df, aes(x=state, group=1)) +  
  geom_line(aes(y = cases, colour = "New York")) +  
  geom_line(data=california_df, aes(y = cases, colour="California")) +  
  geom_line(data=florida_df, aes(y = cases, colour="Florida")) +  
  scale_colour_manual(values = c("darkgreen", "steelblue", "darkred") , breaks = c(9000, 1000000, 10000000))  
  xlab("State") + ylab("cases") + scale_y_log10()
```

