final project 1

Elizabeth Chin

10/25/2021

```
library(fivethirtyeight)
```

```
## Some larger datasets need to be installed separately, like senators and
## house_district_forecast. To install these, we recommend you install the
## fivethirtyeightdata package by running:
## install.packages('fivethirtyeightdata', repos =
## 'https://fivethirtyeightdata.github.io/drat/', type = 'source')
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(mdsr)
library(Hmisc)
```

```
## Loading required package: lattice
```

```
## Loading required package: survival
```

```
## Loading required package: Formula
```

```
##
## Attaching package: 'Hmisc'
```

```
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

Looking at bad_drivers data set

```
?bad_drivers
head(bad_drivers)
```

```
## # A tibble: 6 x 8
##   state      num_drivers perc_speeding perc_alcohol perc_not_distra~ perc_no_previous
##   <chr>            <dbl>         <int>        <int>            <int>             <int>
## 1 Alabama           18.8            39           30               96                80
## 2 Alaska            18.1            41           25               90                94
## 3 Arizona           18.6            35           28               84                96
## 4 Arkansas          22.4            18           26               94                95
## 5 California        12              35           28               91                89
## 6 Colorado          13.6            37           28               79                95
## # ... with 2 more variables: insurance_premiums <dbl>, losses <dbl>
```

Turning bad_drivers$state into an as.factor

```
bad_drivers$state <- as.factor(bad_drivers$state)
typeof(bad_drivers$state)
```

```
## [1] "integer"
```

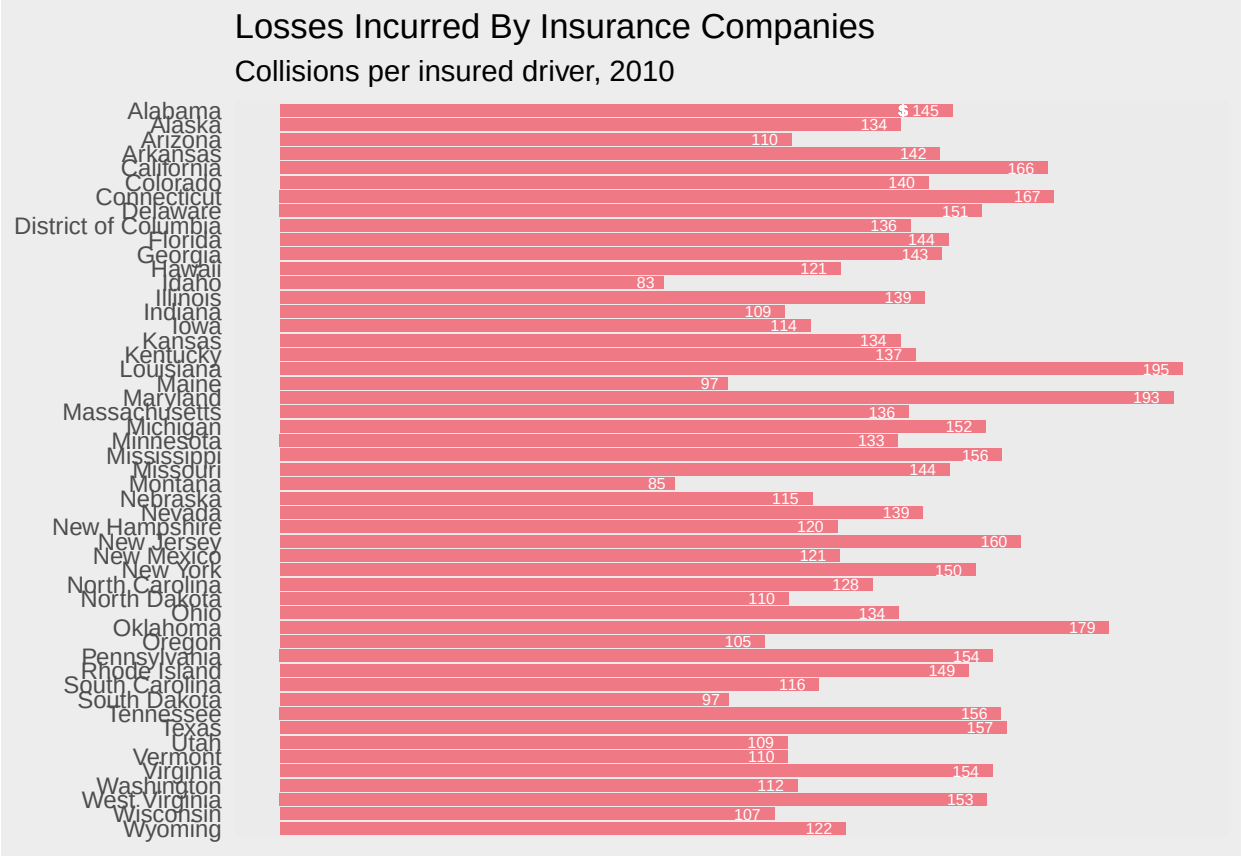
```
levels(bad_drivers$state)
```

```
##  [1] "Alabama"              "Alaska"               "Arizona"
##  [4] "Arkansas"             "California"           "Colorado"
##  [7] "Connecticut"          "Delaware"             "District of Columbia"
## [10] "Florida"              "Georgia"              "Hawaii"
## [13] "Idaho"                "Illinois"             "Indiana"
## [16] "Iowa"                 "Kansas"               "Kentucky"
## [19] "Louisiana"            "Maine"                "Maryland"
## [22] "Massachusetts"        "Michigan"             "Minnesota"
## [25] "Mississippi"          "Missouri"             "Montana"
## [28] "Nebraska"             "Nevada"               "New Hampshire"
## [31] "New Jersey"           "New Mexico"           "New York"
## [34] "North Carolina"       "North Dakota"         "Ohio"
## [37] "Oklahoma"             "Oregon"               "Pennsylvania"
## [40] "Rhode Island"         "South Carolina"       "South Dakota"
## [43] "Tennessee"            "Texas"                "Utah"
## [46] "Vermont"              "Virginia"             "Washington"
## [49] "West Virginia"        "Wisconsin"            "Wyoming"
```

```
g6 <- bad_drivers %>%
  ggplot(aes(x = state, y = losses)) +
  scale_x_discrete(limits = rev(levels(bad_drivers$state))) +
    geom_bar(position = "stack", stat="identity", fill = "#EF7A85") +
    geom_text(aes(label = round(losses, 0)), hjust = 1.5, size = 2, vjust = 0.5, colour = "white") +
    geom_text(aes(x = state[1], y = 145, label = "$"), hjust = 6, size = 2, vjust = 0.5, colour = "whit
```

```
  xlab(NULL) +
  ylab (NULL) +
  ggtitle("Losses Incurred By Insurance Companies") +
  labs(subtitle = 'Collisions per insured driver, 2010') +
  coord_flip() +
  theme(axis.ticks = element_blank())+
  theme(panel.grid.major = element_blank()) +
  theme(panel.grid.minor = element_blank()) +
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5)) +
  theme(axis.text.x = element_blank()) +
  theme(plot.background = element_rect(fill = "grey93"))

g6
```



Link to the dataset

https://fivethirtyeight.com/features/which-state-has-the-worst-drivers/

Information about the data visualization

- Since the sum of car collisions is very large and unequally distributed among states and among the number of insured drivers, the data set divided the insurance companies' losses in each state by the number of insured registered drivers located there. Based on the graph, we can see that Idahoans is

America's best drivers because it costs insurers on average $83 for each collisions in 2010. The most expensive state is Louisiana, where it costs $195 for each collisions in 2010.

Data wrangling-visualization statements

- First I had to convert bad_drivers(state) into an as.factor so that I can arrange it in alphabetical order when plotting it on the x-axis of the graph. Next, I found the fill color of the graph by using colorilla in order to get as close as possible to the original graph's color. Afterwards, I learned how to place a "$" onto one of the states on the graph by using the geom_text function. Lastly, I used the ggtile function to add the graph's title and I used the theme function to make the background of the graph completely grey and without any axis lines.