Assignment 7

Nathan Underwood

2023-11-21

```
library(rvest)
```

```
## Warning: package 'rvest' was built under R version 4.3.2
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter()         masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()            masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Question 1

At the Insurance Institute for Highway Safety, they have data about human fatalities in vehicle crashes. From this web page, import the data from the Fatal Crash Totals data table and produce a bar graph gives the number of deaths per 100,000 individuals. Be sure to sort the states by highest to lowest mortality. *Hint: If you have a problem with the graph being too squished vertically, you can set the chunk options* ***`fig.height`*** *or* ***`fig.width`*** *to make the graph larger, but keeping the font sizes the same. The result is that the text is more spread apart. The chunk options* ***`out.height`*** *and* ***`out.width`*** *shrink or expand everything in the plot. By making the* ***`fix.XXX`*** *options large and* ***`out.XXX`*** *options small, you are effectively decreasing the font size of all the elements in the graph. The other trick is to reset the font size using a theme* ***`element_text`*** *option:* ***`theme(text = element_text(size = 9))`***.

```
url <- "https://www.iihs.org/topics/fatality-statistics/detail/state-by-state"
page <- read_html(url)

fatalCrashTotal <- page %>%
  html_nodes("table") %>%
  .[[1]] %>% # grab the first available table
  html_table(header = FALSE, fill = TRUE) %>% # convert the HTML table into a data frame
  slice(-1 * 1:2 )

# rename the data
fatalCrashTotal <- fatalCrashTotal %>%
  magrittr::set_colnames(c("State", "Population", "Vehicle Miles Traveled (Milions)", "Fatal Crashes",

# clean the data
```

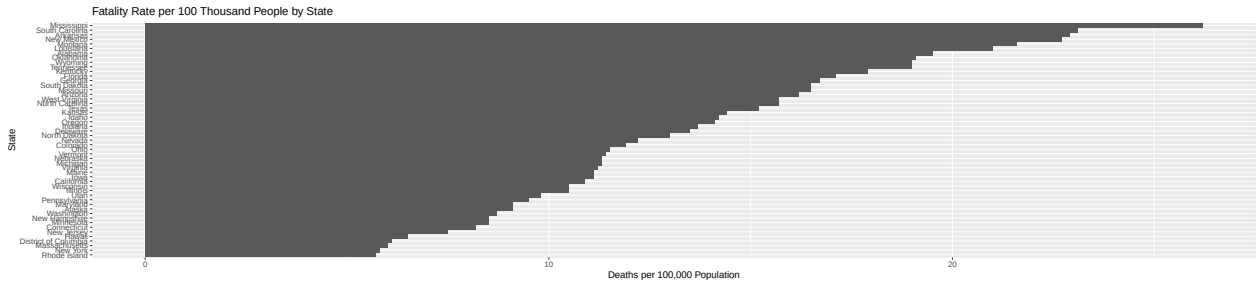
```
fatalCrashTotal <- fatalCrashTotal %>%
  mutate(across(2:7, str_remove_all, ',')) %>%
  mutate(across(2:7, as.numeric)) %>% # correct the data types
  filter(!(State %in% "U.S. total")) %>%
  mutate(State = fct_reorder(`State`, `Deaths per 100,000 Population`))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `across(2:7, str_remove_all, ",")`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

```
head(fatalCrashTotal)
```

```
## # A tibble: 6 x 7
##   State      Population Vehicle Miles Traveled (Milions~1 `Fatal Crashes` Deaths
##   <fct>           <dbl>                             <dbl>           <dbl>  <dbl>
## 1 Alabama       5049846                             79569             885    983
## 2 Alaska         734182                              5752              59     67
## 3 Arizona       7264877                             73760            1063   1180
## 4 Arkansas      3028122                             38427             631    693
## 5 California   39142991                            310823            3983   4285
## 6 Colorado      5811297                             53840             638    691
## # i abbreviated name: 1: `Vehicle Miles Traveled (Milions)`
## # i 2 more variables: `Deaths per 100,000 Population` <dbl>,
## #   `Deaths per 100 Million Vehicles Miles Traveled` <dbl>
```

```
# plot the data in a bar graph
ggplot(fatalCrashTotal, aes(x = `State`, y = `Deaths per 100,000 Population`)) +
  geom_col() +
  coord_flip() +
  labs(title='Fatality Rate per 100 Thousand People by State')
```



Question 2

From the same IIHS website, import the data about seat belt use. Join the Fatality data with the seat belt use and make a scatter plot of percent seat belt use vs number of fatalities per 100,000 people.