Homework 5 of STAT 540
Section 001, Fall 2024
Due: Wednesday Oct 9 (before class)
Total Points: 60

Please hand in a hard copy of your homework (compiled pdf file from R markdown) in class and email your R code to Kaniz Fatema (KFATEMA@email.sc.edu). Please use the R markdown Homework template (HWtemplate.Rmd) to write your homework solutions. Work on the homework independently.

The exercises questions can be found in Chapter 9 in the book "Introduction to Data Science: Data Wrangling and Visualization with R "

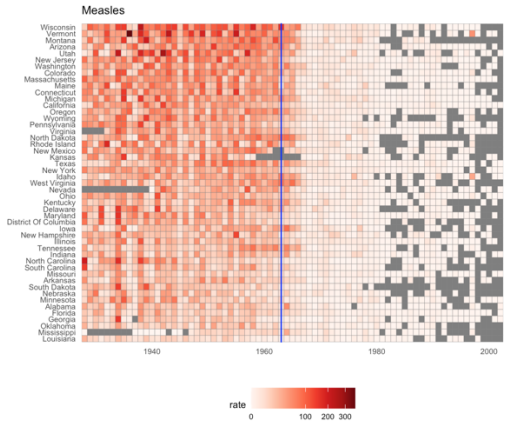
https://rafalab.dfci.harvard.edu/dsbook-part-1/dataviz/dataviz-principles.html

Problem 1. Question 2 in Chapter 9 in the book by Dr. Irizarry. (3 points)
Problem 2. Question 3 in Chapter 9 in the book by Dr. Irizarry. (2 points)
Problem 3. Question 5 in Chapter 9 in the book by Dr. Irizarry and make the plot better. (10 points)

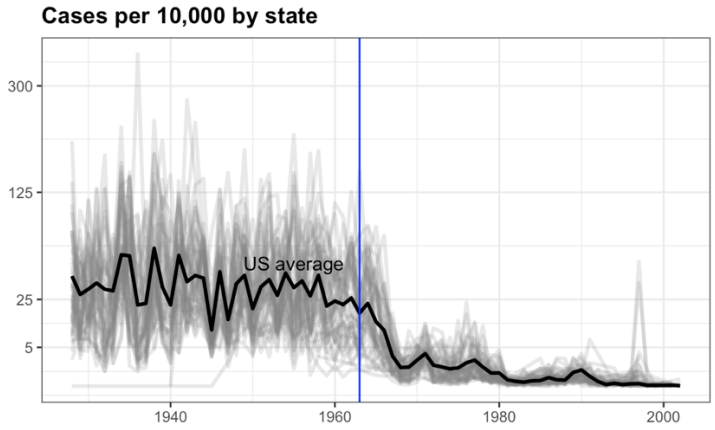
The following questions are related to Chapter 10 in the book by Dr. Irizarry.

Problem 4. Use the Vaccine data (us_contagious_diseases) in the dslabs R package and make this plot for smallpox (20 points)

```
library(tidyverse)
library(RColorBrewer)
library(dslabs)
names(us_contagious_diseases)
```



Problem 5. Use the Vaccine data (us_contagious_diseases)in the dslabs R package and make this time series plot. For the state of California, make a time series plot showing rates for all diseases. Include only years with 10 or more weeks reporting. Use a different color for each disease. (25 points)



Problem 6. Start by loading the **dplyr** and **ggplot2** library as well as the `murders` and `heights` data. (4 points each)

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

(i) With **ggplot2**, plots can be saved as objects. For example we can associate a dataset with a plot object like this

```
p <- ggplot(data = murders)
```

Because `data` is the first argument we don't need to spell it out

```
p <- ggplot(murders)
```

and we can also use the pipe:

```
p <- murders |> ggplot()
```

What is class of the object `p`?

(ii) Now we are going to add a layer and the corresponding aesthetic mappings to show the relationship between population (x) and total gum murder (y). Use the code below to create the plot.

```
murders |> ggplot(aes(x = , y = )) +
  geom_point()
```

(iii) If instead of points we want to add text, we can use the `geom_text()` or `geom_label()` geometries. The following code

```
murders |> ggplot(aes(population, total)) +
geom_label()
```

will give us the error message: `Error: geom_label requires the following missing aesthetics: label`

Why is this?

a. We need to map a character to each point through the label argument in aes.
b. We need to let `geom_label` know what character to use in the plot.
c. The `geom_label` geometry does not require x-axis and y-axis values.
d. `geom_label` is not a ggplot2 command.

(iv) Now we are going to change the x-axis to a log scale to account for the fact the distribution of population is skewed. Let's start by defining an object `p` holding the plot we have made up to now

```
p <- murders |>
  ggplot(aes(population, total, label = abb, color =
region)) +
```

```
  geom_label()
```
To change the x-axis to a log scale we learned about the scale_x_log10() function. Add this layer to the object p to change the scale and render the plot.

Problem 7. We can also assign groups through the fill argument. This has the added benefit that it uses colors to distinguish the groups, like this:

```
heights |>
  ggplot(aes(height, fill = sex)) +
  geom_density()
```
However, here the second density is drawn over the other. We can make the curves more visible by using alpha blending to add transparency. Set the alpha parameter to 0.2 in the geom_density function to make this change. (4 points)