

# Mini-Project

```
library(viridis)
```

Loading required package: viridisLite

```
library(maps)
```

Attaching package: 'maps'

The following object is masked from 'package:viridis':

unemp

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
x purrr::map()     masks maps::map()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(leaflet)
```

The data for this project will be using the leading causes of death in the United States dataset, linked here: [[https://data.cdc.gov/NCHS/NCHS-Leading-Causes-of-Death-United-States/bi63-dtpu/about\\_data](https://data.cdc.gov/NCHS/NCHS-Leading-Causes-of-Death-United-States/bi63-dtpu/about_data)]

For this (mini) project, we will be analyzing the different death statistics across the US for each individual state, specifically for the year 2016. The numeric data we will be analyzing is the age-adjusted rate for the highest cause of death, while the categorical data used is the leading cause of death for each state.

Let's load in the data

```
main_data <- read_csv("C:/Users/charl/Documents/SDS_264/NCHS_-_Leading_Causes_of_Death_Unit
```

```
Rows: 10868 Columns: 6
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
chr (3): 113 Cause Name, Cause Name, State
```

```
dbl (3): Year, Deaths, Age-adjusted Death Rate
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Here we are only getting the data from 2016, and renaming the states for consistency reasons later.

```
main_data <- main_data |>
  filter(Year == "2016") |> #filter only 2016
  select(-c(`113 Cause Name`)) |>
  mutate(State = str_to_lower(State)) #uniformity for state names later
```

```
library(maps) #load in maps
us_states <- map_data("state")
head(us_states)
```

	long	lat	group	order	region	subregion
1	-87.46201	30.38968	1	1	alabama	<NA>
2	-87.48493	30.37249	1	2	alabama	<NA>
3	-87.52503	30.37249	1	3	alabama	<NA>
4	-87.53076	30.33239	1	4	alabama	<NA>
5	-87.57087	30.32665	1	5	alabama	<NA>
6	-87.58806	30.32665	1	6	alabama	<NA>

Next, let's find out which is the highest cause of death in each state, we will plot these results later.

```
main_data <- main_data |>
  group_by(State) |>
  filter(`Cause Name` != "All causes") |>
  mutate(highest_death = max(Deaths),
         highest_cause = `Cause Name`[which.max(Deaths)]) |>
  ungroup()
```

Then we create a new table of just the highest cause of death.

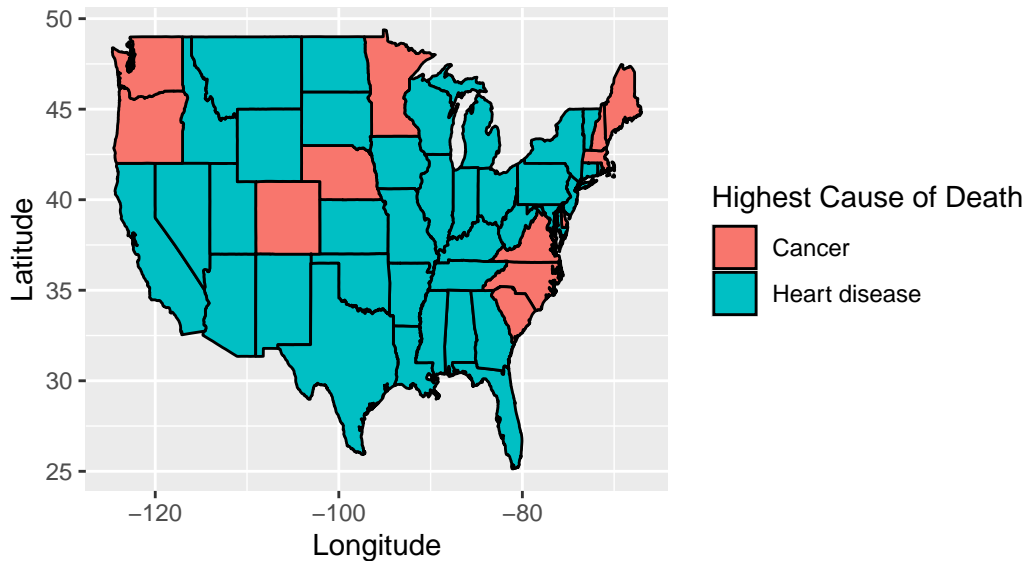
```
highest_cause_data <- main_data |>
  filter(`Cause Name` == highest_cause)
```

Let's plot the categorical data, what exactly is causing the greatest amount of death for every state?

```
highest_cause_data |>
  right_join(us_states, by = c("State" = "region")) |>
  rename(region = State) |>
  ggplot(mapping = aes(x = long, y = lat, group = group)) +
  geom_polygon(aes(fill = highest_cause), color = "black") +
  labs(
    title = "Leading Cause of Death by State in the United States",
    subtitle = "Comparing Heart Disease and Cancer as the Primary Causes of Mortality",
    fill = "Highest Cause of Death",
    x = "Longitude",
    y = "Latitude"
  )
```

## Leading Cause of Death by State in the United States

### Comparing Heart Disease and Cancer as the Primary Causes of Mortality



```
str(main_data$highest_cause)
```

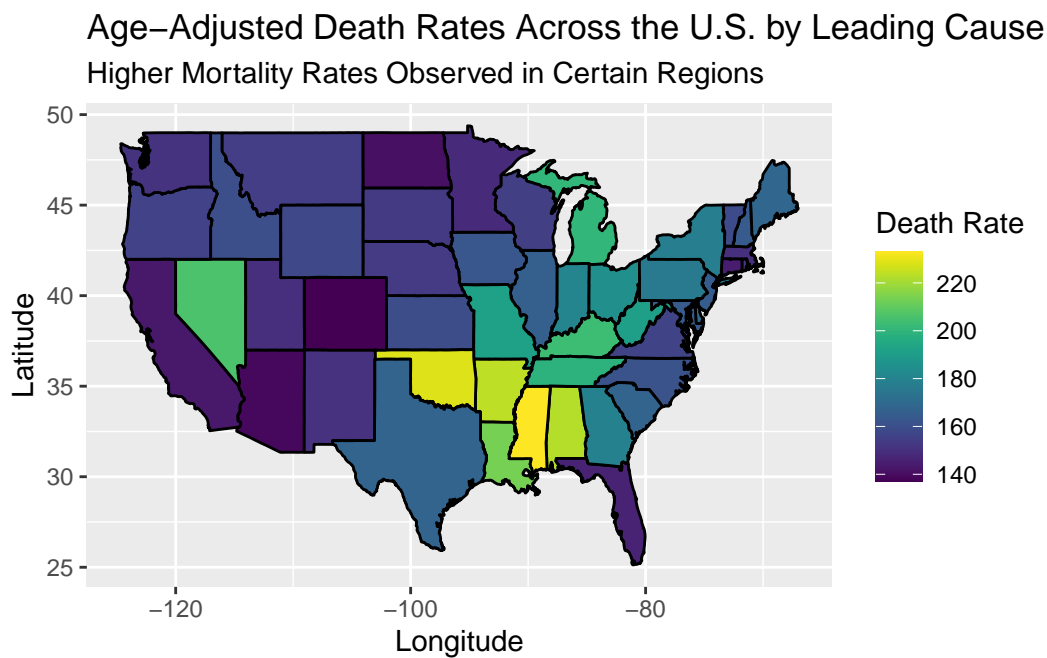
```
chr [1:520] "Heart disease" "Cancer" "Heart disease" "Heart disease" ...
```

ALT TEXT: Here we have a map of the United States, divided by the different states in the US. On the y axis we have the latitude of the map, with values ranging from 25 to 50, on the x axis we have the longitude of the map, with values ranging from -125 to - 65. From this graph, we can see that the majority of the US has their main cause of death coming from heart disease, with the rare exception of cancer. Cancer is the leading cause of death in Washington, Oregon, Colorado, Nebraska, Minnesota, South Carolina, North Carolina, Virginia, DC, Massachusetts, New Hampshire, and Maine, where everywhere else has heart disease as a higher cause of death.

We will also see just how deadly these are through Age-adjusted Death Rates in the data. An excerpt from <https://www.health.ny.gov/diseases/chronic/ageadj.htm> defines Age-adjusted Death Rate as “usually expressed in terms of the number of cases/deaths per 100,000 people per year”.

```
highest_cause_data |>
  right_join(us_states, by = c("State" = "region")) |>
  rename(region = State) |>
  ggplot(mapping = aes(x = long, y = lat, group = group)) +
```

```
geom_polygon(aes(fill = `Age-adjusted Death Rate`), color = "black") +
labs(
  title = "Age-Adjusted Death Rates Across the U.S. by Leading Cause of Death",
  subtitle = "Higher Mortality Rates Observed in Certain Regions",
  fill = "Death Rate",
  x = "Longitude",
  y = "Latitude"
) +
scale_fill_viridis()
```



From this graph we can see that areas such as the south are more impacted by their death rates than others areas. Specifically, their highest cause is heart disease, so we see that that is impacting them at a rate higher than other areas.

Next we will load in our interactive plots using the sf library:

```
library(sf)
```

①

Linking to GEOS 3.13.0, GDAL 3.10.1, PROJ 9.5.1; sf\_use\_s2() is TRUE

```
states <- read_sf("https://rstudio.github.io/leaflet/json/us-states.geojson") ②
```

Here we get the sf version of states, so we can then add our data as follows:

```
library(htmltools)
library(glue)

states <- states |>
  mutate(name = str_to_lower(name)) |> #align state names and get data from highest cause
  right_join(highest_cause_data, by = c("name" = "State"))

states <- states |> #create labels for graphs
  mutate(labels = str_c(name, ", Age-adjusted death: ", `Age-adjusted Death Rate`, " from: ")

bins <- c(100, 125, 150, 175, 200, 225, 250, 275, Inf)
pal <- colorBin("YlOrRd", domain = states$density, bins = bins)
# If want more HTML formatting, use these lines instead of those above:
#states <- states |>
# mutate(labels = glue("<strong>{name}</strong><br/>{density} people / #mi<sup>2</sup>"))

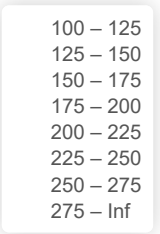
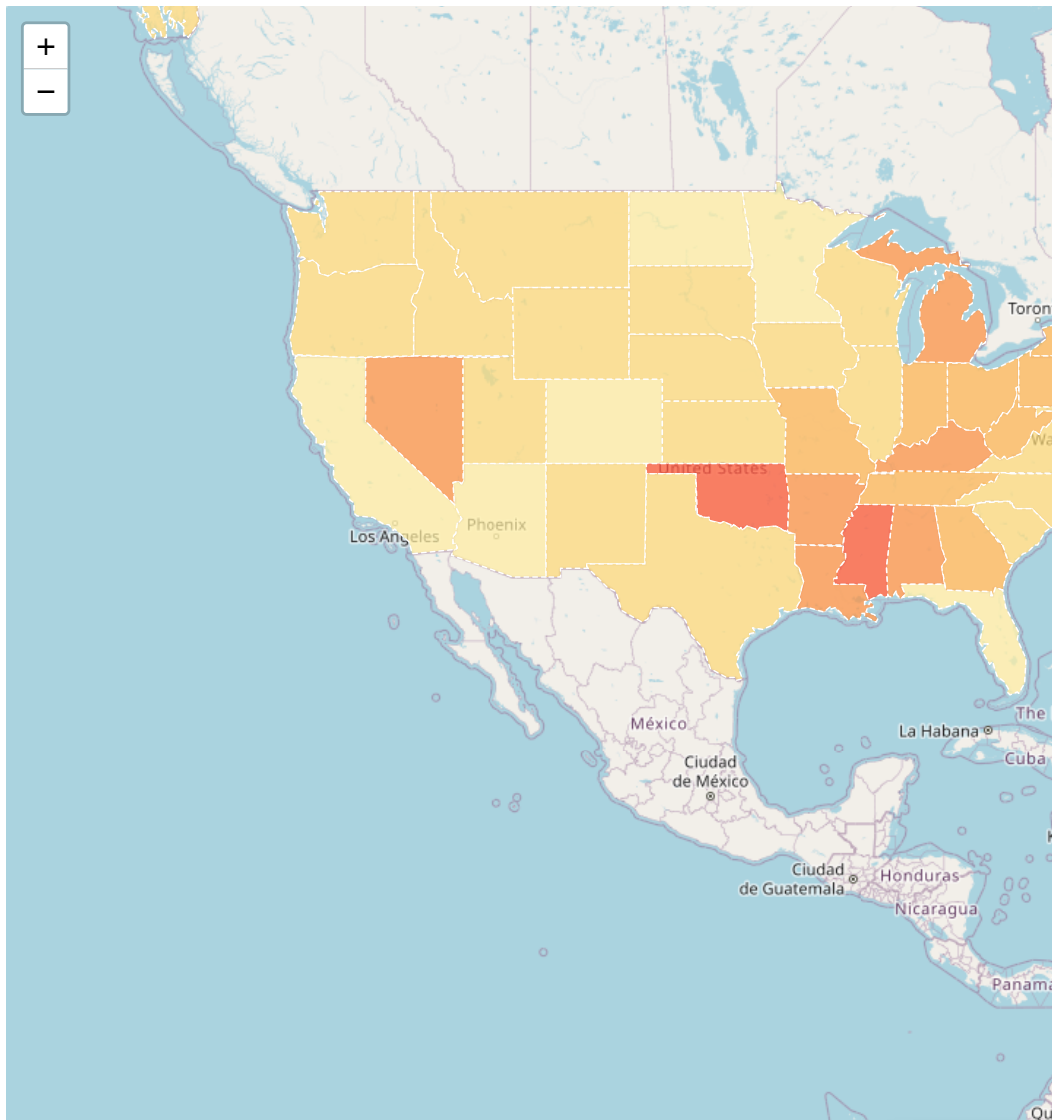
labels <- lapply(states$labels, HTML)

leaflet(states) |>
  setView(-96, 37.8, 4) |>
  addTiles() |>
  addPolygons(
    fillColor = ~pal(`Age-adjusted Death Rate`),
    weight = 1,
    opacity = 1,
    color = "white",
    dashArray = "3",
    fillOpacity = 0.7,
    highlightOptions = highlightOptions(
      weight = 5,
      color = "#666",
      dashArray = "",
      fillOpacity = 0.7,
      bringToFront = TRUE),
    label = labels,
    labelOptions = labelOptions(
      style = list("font-weight" = "normal", padding = "3px 8px"),
```

```
    textsize = "15px",  
    direction = "auto")) |>  
addLegend(pal = pal, values = ~`Age-adjusted Death Rate`, opacity = 0.7, title = NULL,  
          position = "bottomright")
```

Warning in sf::st\_is\_longlat(x): bounding box has potentially an invalid value  
range for longlat data

file:///C:/Users/charl\AppData/Local/Temp/Rtmp6r4ToD/file105506585b46/widget10550628523df.htm



Leaflet | © OpenStreetMap, ODbL



```

states <- states |> #create labels for graphs
  mutate(labels = str_c(name, ", Leading cause of death: ", highest_cause))

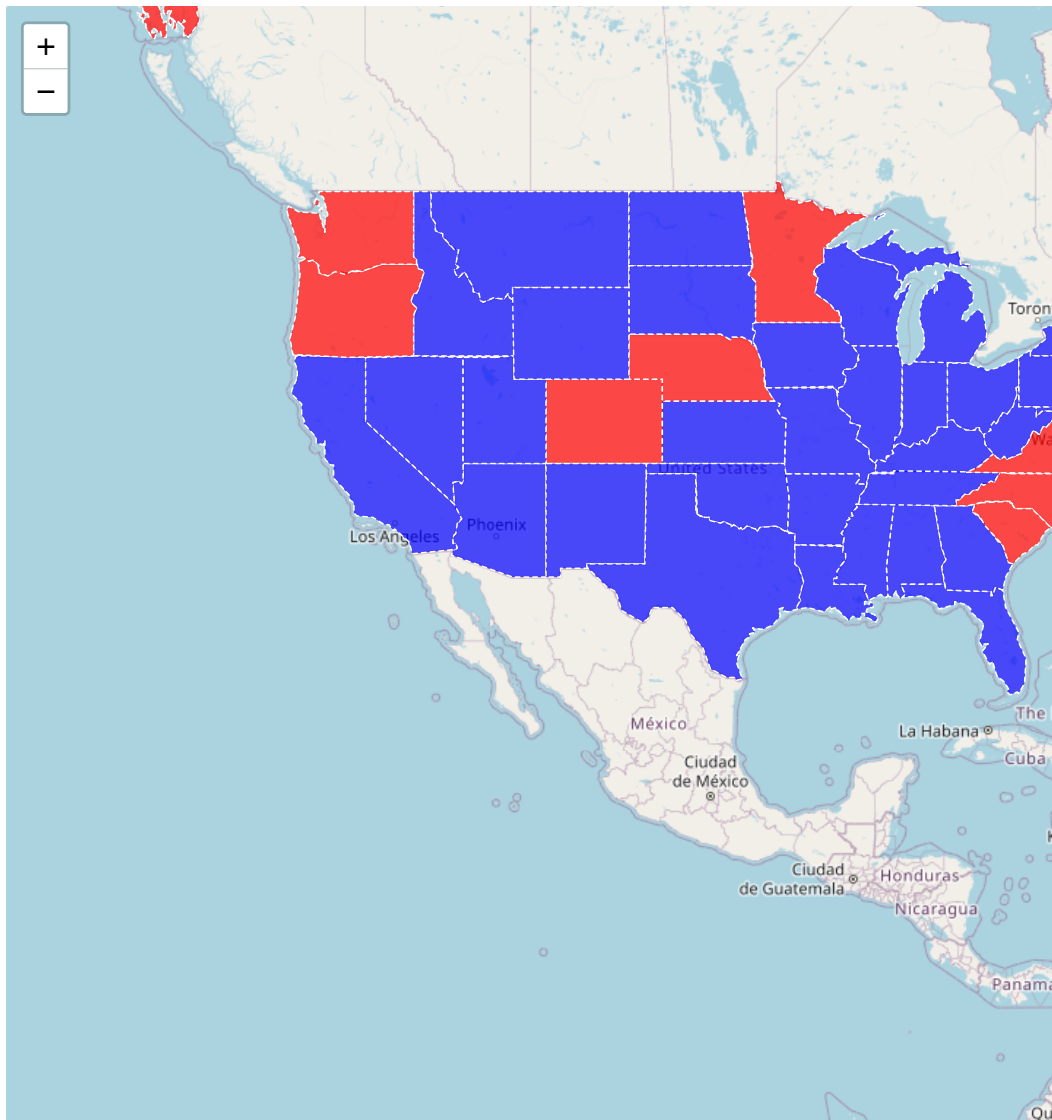
pal <- colorFactor(c("red", "blue"), domain = c("Cancer", "Heart disease"))

leaflet(states) |>
  setView(-96, 37.8, 4) |>
  addTiles() |>
  addPolygons(
    fillColor = ~pal(highest_cause), # Use the fixed color mapping
    weight = 1,
    opacity = 1,
    color = "white",
    dashArray = "3",
    fillOpacity = 0.7,
    highlightOptions = highlightOptions(
      weight = 5,
      color = "#666",
      dashArray = "",
      fillOpacity = 0.7,
      bringToFront = TRUE),
    label = labels,
    labelOptions = labelOptions(
      style = list("font-weight" = "normal", padding = "3px 8px"),
      textSize = "15px",
      direction = "auto")) |>
  addLegend(pal = pal, values = ~highest_cause, opacity = 0.7, title = "Leading Cause",
    position = "bottomright")

```

Warning in sf::st\_is\_longlat(x): bounding box has potentially an invalid value range for longlat data

file:///C:/Users/charl/AppData/Local/Temp/Rtmp6r4ToD/file10550222b5688/widget10550621438b6.h



**Leading Cause**

Cancer

Heart disease

Leaflet | © OpenStreetMap, ODbL