Mini Project 1

https://data.hrsa.gov/maps/mchb

```
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 1.9.3     v tidyr     1.3.1
v purrr     1.0.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becor
```

```
library(dplyr)
```

```
cesarean_delivery <- read_csv("~/Desktop/Cesarean Delivery - All (%) 2020-2022 - Output fo
```

```
Rows: 3272 Columns: 7
-- Column specification --------------------------------------------------------
Delimiter: ","
chr (4): County FIPS Code, State, County, NCHS Urban-Rural Classification
dbl (3): HRSA Region, Cesarean Delivery - All (%) 2020-2022, NCHS Urban-Rura...

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.
```

```
cesarean_delivery
```

```
# A tibble: 3,272 x 7
   `County FIPS Code` `HRSA Region` State   County        Cesarean Delivery - ~1
   <chr>                      <dbl> <chr>   <chr>                          <dbl>
 1 01001                          4 Alabama Autauga Coun~                   40.1
 2 01003                          4 Alabama Baldwin Coun~                   31.7
 3 01005                          4 Alabama Barbour Coun~                   35.7
 4 01007                          4 Alabama Bibb County                     34
 5 01009                          4 Alabama Blount County                   31.7
 6 01011                          4 Alabama Bullock Coun~                   38
 7 01013                          4 Alabama Butler County                   39.3
 8 01015                          4 Alabama Calhoun Coun~                   33.8
 9 01017                          4 Alabama Chambers Cou~                   38.1
10 01019                          4 Alabama Cherokee Cou~                   31.1
# i 3,262 more rows
# i abbreviated name: 1: `Cesarean Delivery - All (%) 2020-2022`
# i 2 more variables: `NCHS Urban-Rural Classification` <chr>,
#   `NCHS Urban-Rural Code` <dbl>
```

```
#loading in info to draw US states for geom_polygon
library(maps)
```

```
Attaching package: 'maps'

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

    map
```

```
states_polygon <- as_tibble(map_data("state")) |>
  select(region, group, order, lat, long)

# See what the state (region) levels look like in states_polygon
unique(states_polygon$region)
```

```
 [1] "alabama"              "arizona"              "arkansas"
 [4] "california"           "colorado"             "connecticut"
```

```
 [7] "delaware"            "district of columbia" "florida"
[10] "georgia"             "idaho"                "illinois"
[13] "indiana"             "iowa"                 "kansas"
[16] "kentucky"            "louisiana"            "maine"
[19] "maryland"            "massachusetts"        "michigan"
[22] "minnesota"           "mississippi"          "missouri"
[25] "montana"             "nebraska"             "nevada"
[28] "new hampshire"       "new jersey"           "new mexico"
[31] "new york"            "north carolina"       "north dakota"
[34] "ohio"                "oklahoma"             "oregon"
[37] "pennsylvania"        "rhode island"         "south carolina"
[40] "south dakota"        "tennessee"            "texas"
[43] "utah"                "vermont"              "virginia"
[46] "washington"          "west virginia"        "wisconsin"
[49] "wyoming"
```

```r
#loading in info to draw US states for geom_sf and leaflet
library(sf)
```

```
Linking to GEOS 3.11.0, GDAL 3.5.3, PROJ 9.1.0; sf_use_s2() is TRUE
```

```r
states_sf <- read_sf("https://rstudio.github.io/leaflet/json/us-states.geojson") |>
  select(name, geometry)

# See what the state (name) levels look like in states_sf
unique(states_sf$name)
```

```
 [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"
[52] "Puerto Rico"
```

```
c_delivery <- as_tibble(cesarean_delivery) |>
  filter()

# See what the state (State) levels look like in c_delivery
unique(c_delivery$State)
```

```
 [1] "Alabama"                        "Alaska"
 [3] "Arizona"                        "Arkansas"
 [5] "California"                     "Colorado"
 [7] "Connecticut"                    "Delaware"
 [9] "District of Columbia"           "Florida"
[11] "Georgia"                        "Hawaii"
[13] "Idaho"                          "Illinois"
[15] "Indiana"                        "Iowa"
[17] "Kansas"                         "Kentucky"
[19] "Louisiana"                      "Maine"
[21] "Maryland"                       "Massachusetts"
[23] "Michigan"                       "Minnesota"
[25] "Mississippi"                    "Missouri"
[27] "Montana"                        "Nebraska"
[29] "Nevada"                         "New Hampshire"
[31] "New Jersey"                     "New Mexico"
[33] "New York"                       "North Carolina"
[35] "North Dakota"                   "Ohio"
[37] "Oklahoma"                       "Oregon"
[39] "Pennsylvania"                   "Rhode Island"
[41] "South Carolina"                 "South Dakota"
[43] "Tennessee"                      "Texas"
[45] "Utah"                           "Vermont"
[47] "Virginia"                       "Washington"
[49] "West Virginia"                  "Wisconsin"
[51] "Wyoming"                        "American Samoa"
[53] "Federated States of Micronesia" "Guam"
[55] "Marshall Islands"               "Northern Mariana Islands"
[57] "Republic of Palau"              "Puerto Rico"
[59] "U.S. Minor Islands"             "U.S. Virgin Islands"
```

```r
# Matching the key values
library(lubridate)

states_polygon <- states_polygon |>
  mutate(region = str_to_title(region))
states_polygon
```

```
# A tibble: 15,537 x 5
   region  group order   lat  long
   <chr>   <dbl> <int> <dbl> <dbl>
 1 Alabama     1     1  30.4 -87.5
 2 Alabama     1     2  30.4 -87.5
 3 Alabama     1     3  30.4 -87.5
 4 Alabama     1     4  30.3 -87.5
 5 Alabama     1     5  30.3 -87.6
 6 Alabama     1     6  30.3 -87.6
 7 Alabama     1     7  30.3 -87.6
 8 Alabama     1     8  30.3 -87.6
 9 Alabama     1     9  30.3 -87.7
10 Alabama     1    10  30.3 -87.8
# i 15,527 more rows
```

```r
# Checking what states/regions need to be filtered out
c_delivery |>
  anti_join(states_polygon, by = c("State" = "region")) |>
  count(State)
```

```
# A tibble: 12 x 2
   State                              n
   <chr>                          <int>
 1 Alaska                            31
 2 American Samoa                     5
 3 District of Columbia               1
 4 Federated States of Micronesia     4
 5 Guam                               1
 6 Hawaii                             5
 7 Marshall Islands                  11
 8 Northern Mariana Islands           4
 9 Puerto Rico                       78
10 Republic of Palau                 13
```

```
11 U.S. Minor Islands                 9
12 U.S. Virgin Islands                3
```

```
c_delivery |>
  anti_join(states_sf, by = c("State" = "name")) |>
  count(State)
```

```
# A tibble: 8 x 2
  State                            n
  <chr>                        <int>
1 American Samoa                   5
2 Federated States of Micronesia    4
3 Guam                              1
4 Marshall Islands                 11
5 Northern Mariana Islands          4
6 Republic of Palau                13
7 U.S. Minor Islands                9
8 U.S. Virgin Islands               3
```

```
# Filtering
c_delivery <- c_delivery |>
  filter(!(State %in% c("alaska",
                        "american samoa",
                        "federated states of micronesia",
                        "guam",
                        "hawaii",
                        "marshall islands",
                        "northern mariana islands",
                        "puerto rico",
                        "republic of palau",
                        "u.s minor islands",
                        "u.s. virgin islands")))
c_delivery
```

```
# A tibble: 3,272 x 7
   `County FIPS Code` `HRSA Region` State   County        Cesarean Delivery - ~1
   <chr>                      <dbl> <chr>   <chr>                          <dbl>
 1 01001                          4 Alabama Autauga Coun~                   40.1
 2 01003                          4 Alabama Baldwin Coun~                   31.7
 3 01005                          4 Alabama Barbour Coun~                   35.7
```

```
 4 01007                          4 Alabama Bibb County                      34
 5 01009                          4 Alabama Blount County                    31.7
 6 01011                          4 Alabama Bullock Coun~                    38
 7 01013                          4 Alabama Butler County                    39.3
 8 01015                          4 Alabama Calhoun Coun~                    33.8
 9 01017                          4 Alabama Chambers Cou~                    38.1
10 01019                          4 Alabama Cherokee Cou~                    31.1
# i 3,262 more rows
# i abbreviated name: 1: `Cesarean Delivery - All (%) 2020-2022`
# i 2 more variables: `NCHS Urban-Rural Classification` <chr>,
#   `NCHS Urban-Rural Code` <dbl>
```

```r
# Merging for static
c_del_polygon <- c_delivery |>
   group_by(State) |>
  summarise(mean_c_section_delivery = round(mean(`Cesarean Delivery - All (%) 2020-2022`),
            mean_urban_rural = round(mean(`NCHS Urban-Rural Code`, na.rm = TRUE), digits =
  right_join(states_polygon, by = c("State" = "region"))
c_del_polygon
```

```
# A tibble: 15,537 x 7
   State   mean_c_section_delivery mean_urban_rural group order   lat  long
   <chr>                     <dbl>            <dbl> <dbl> <int> <dbl> <dbl>
 1 Alabama                      36                5     1     1  30.4 -87.5
 2 Alabama                      36                5     1     2  30.4 -87.5
 3 Alabama                      36                5     1     3  30.4 -87.5
 4 Alabama                      36                5     1     4  30.3 -87.5
 5 Alabama                      36                5     1     5  30.3 -87.6
 6 Alabama                      36                5     1     6  30.3 -87.6
 7 Alabama                      36                5     1     7  30.3 -87.6
 8 Alabama                      36                5     1     8  30.3 -87.6
 9 Alabama                      36                5     1     9  30.3 -87.7
10 Alabama                      36                5     1    10  30.3 -87.8
# i 15,527 more rows
```

```r
#Merging for interactive
temp <- c_delivery |>
   group_by(State) |>
  summarise(mean_c_section_delivery = round(mean(`Cesarean Delivery - All (%) 2020-2022`),
            mean_urban_rural = round(mean(`NCHS Urban-Rural Code`, na.rm = TRUE), digits =
```

```
c_del_sf_1 <- states_sf |>
  left_join(temp, by = c("name" = "State"))
```

Average Percent Cesarian Deliveries By State 2020-2022 (static plot)

```
library(viridis)
```

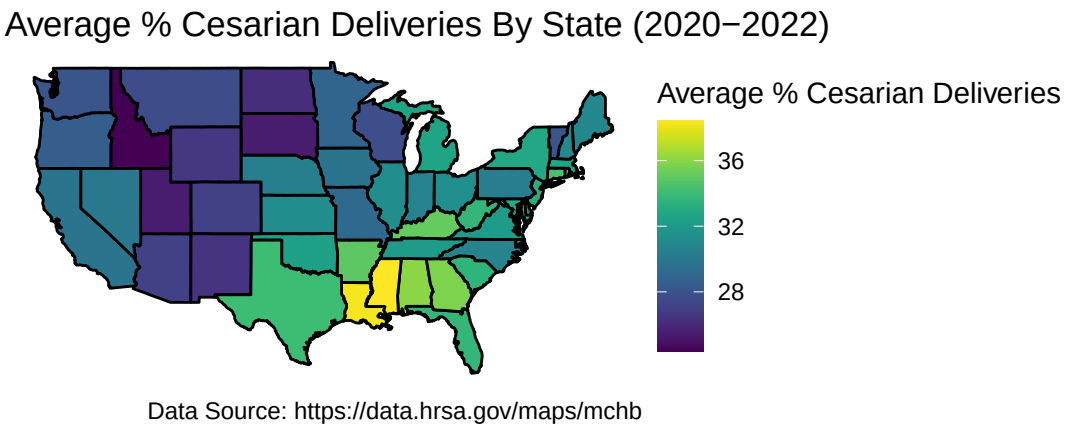
```
Loading required package: viridisLite
```

```
Attaching package: 'viridis'
```

```
The following object is masked from 'package:maps':

    unemp
```

```
c_del_polygon |>
  ggplot(mapping = aes(x = long, y = lat, group = group)) +
    geom_polygon(aes(fill = mean_c_section_delivery), color = "black") +
    labs(fill = "Average % Cesarian Deliveries", caption = "Data Source: https://data.hrsa
    coord_map() +
   theme_void() +
   scale_fill_viridis()
```



- This graph is a map of the continental United States color coded with state-level data of the average percent of deliveries which were cesarean (c-sections) in the years 2020-2022, aggregated from county-level data. The average percent of deliveries which were cesarean varies from approximately 24% to 38%. The graph highlights that the South tends to have higher rates of cesarian deliveries, especially Louisiana and Mississippi, while the West and upper-western Midwest tend to have lower rates of cesarian sections.

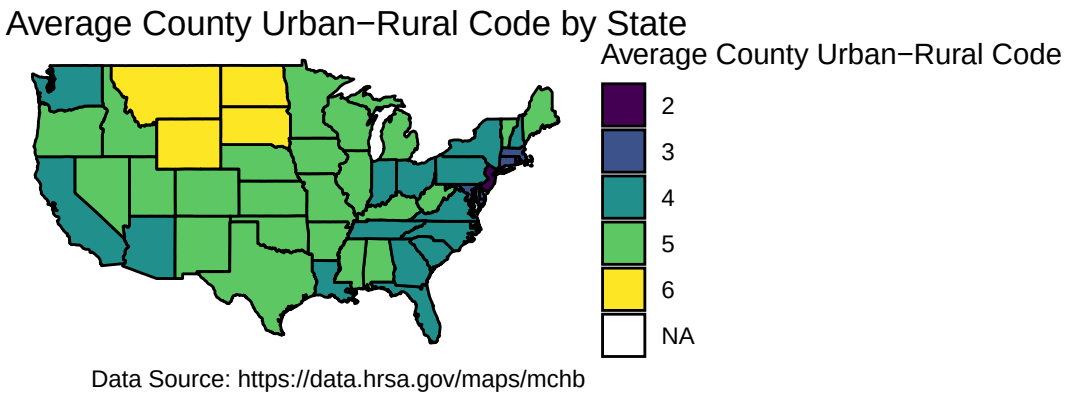
Average County Urban-Rural Code by State (static plot)

```
table(c_del_polygon$mean_urban_rural)
```

```
  2    3    4    5    6
271 1037 6509 7133  577
```

```
c_del_polygon |>
    ggplot(mapping = aes(x = long, y = lat, group = group)) +
    geom_polygon(aes(fill = as.factor(mean_urban_rural)), color = "black") +
    labs(fill = "Average County Urban-Rural Code", caption = "Data Source: https://data.hr
    coord_map() +
```

```
  theme_void() +
scale_fill_viridis(discrete = TRUE)
```



- This graph is a map of the continental US color-coded at the state-level by the mean county urban-rural code. The mean county urban-rural code ranges from 2 (most urban) to 6 (most rural). States on the east coast have the lowest urban-rural code, meaning that the average of state's counties' urban-rural codes is the most urban. States in the South and on the west coast are next-most urban. The states that are the most rural are North Dakota, South Dakota, Montana, and Wyoming.