mini_project_1

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

```
library(mdsr)
library(maps)
```

```
Attaching package: 'maps'

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

    map
```

```
library(viridis)
```

```
Loading required package: viridisLite

Attaching package: 'viridis'

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

```
    unemp
```

```
library(leaflet)
library(sf)
```

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

##Dataset: CigarettesSW: This dataset is about cigarette consumption for the 48 continental US States from 1985–1995.

##Source: Stock, J.H. and Watson, M.W. (2007). Introduction to Econometrics, 2nd ed. Boston: Addison Wesley.

##Website: https://vincentarelbundock.github.io/Rdatasets/datasets.html

```
# Load US map data
library(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>
```

```
# Load CigarettesSW dataset
CigarettesSW <- read.csv("https://raw.githubusercontent.com/Ela-Kanade/CigarettesMaps/main/C:

# Filter for 1985 data
cigarettes_data <- CigarettesSW %>%
  filter(year == 1985)

# Create a short version of the US map data (only one entry per region)
us_states_short <- us_states %>%
  group_by(region) %>%
  slice(1)
```

#Set up the CigarettesSW dataset!

```r
# Map state abbreviations to full names
state_abbr_to_fullname <- data.frame(
  abbr = state.abb,
  name = state.name)

# Convert the 'region' column in us_states_short to title case for consistency
us_states_short$region <- str_to_title(us_states_short$region)

# Join the state abbreviation data to cigarettes_data and rename column for consistency
cigarettes_data <- cigarettes_data %>%
  left_join(state_abbr_to_fullname, by = c("state" = "abbr")) %>%
  rename(states_names = name)

# Merge the cleaned US states data with cigarettes data
combined_data <- us_states_short %>%
  left_join(cigarettes_data, by = c("region" = "states_names"))

us_map <- map_data("state")
```

#Static plot of Cigarette Packs Sold by State!

```r
# Create a static map of cigarette packs sold by state
library(ggplot2)

# Prepare the map data and add cigarette packs sold per state
state_packs <- combined_data %>%
  select(region, packs) %>%
  mutate(region = tolower(region))  # Match lowercase region names in map_data

# Merge the map data with cigarette packs data
map_with_packs <- left_join(us_map, state_packs, by = "region")

# Create the plot
ggplot(data = map_with_packs, aes(x = long, y = lat, group = group, fill = packs)) +
  geom_polygon(color = "black", linewidth = 0.2) +
  scale_fill_viridis(option = "D", na.value = "gray90") +
  theme_void() +
  labs(title = "Cigarette Packs Sold by State", fill = "Packs Sold")
```
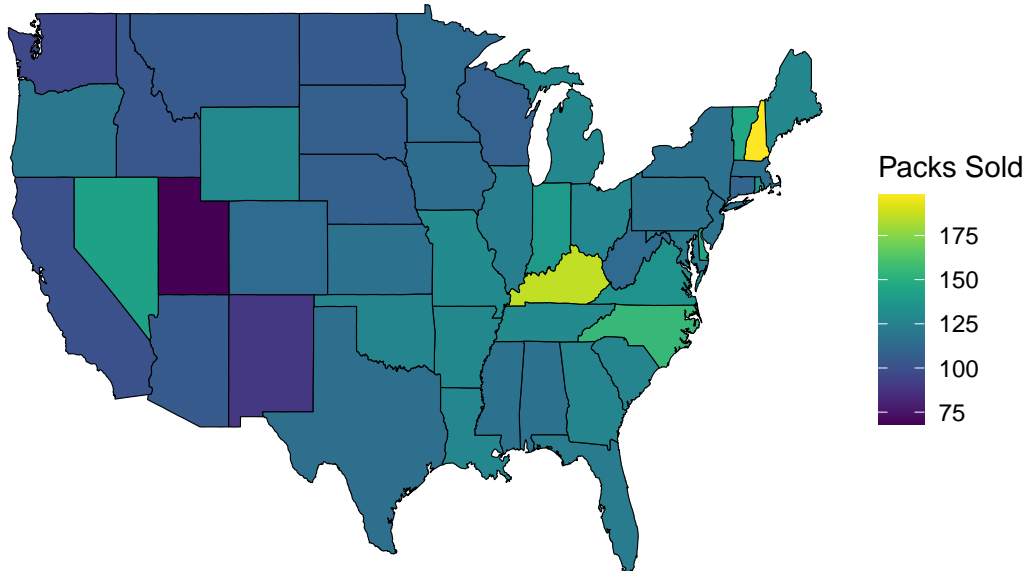
Cigarette Packs Sold by State

#More set up for interactive plots

```
# Convert long and lat to numeric and ensure packs are numeric as well
combined_data$long <- as.numeric(combined_data$long)
combined_data$lat <- as.numeric(combined_data$lat)
combined_data$packs <- as.numeric(combined_data$packs)

# Convert to an sf object for spatial manipulation
us_states_sf <- st_as_sf(combined_data, coords = c("long", "lat"), crs = 4326, agr = "constar
```

First interactive plot - numerical

```
# Load spatial data for US states
states_sf <- read_sf("https://rstudio.github.io/leaflet/json/us-states.geojson") %>%
  select(name, geometry)

# Merge spatial data with cigarette data
cigarettes_sf <- states_sf %>%
  left_join(cigarettes_data, by = c("name" = "states_names")) %>%
  filter(!(name %in% c("alaska", "hawaii"))) %>%
  mutate(across(where(is.numeric), ~round(.x, 1)))
```

```r
# Create color bins for the numerical data (packs sold per capita)
bins <- c(0, 30, 60, 90, 120, 150, 200, 250, Inf)
pal <- colorBin("YlOrRd", domain = cigarettes_sf$packs, bins = bins)


# Add labels to each state and create the interactive leaflet map
library(htmltools)
library(glue)

# Add labels for each state with the number of packs sold
cigarettes_sf <- cigarettes_sf %>%
  mutate(labels = str_c(name, ": ", packs, " packs of cigarettes sold per capita in 1985"))

# Convert labels to HTML format for leaflet
labels <- lapply(cigarettes_sf$labels, HTML)

# Create the interactive map with leaflet
leaflet(cigarettes_sf) %>%
  setView(-96, 37.8, 4) %>%
  addTiles() %>%
  addPolygons(
    fillColor = ~pal(packs),
    weight = 2,
    opacity = 1,
    color = "black",
    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 = ~packs, opacity = 0.7, title = NULL, position = "bottomright"
```
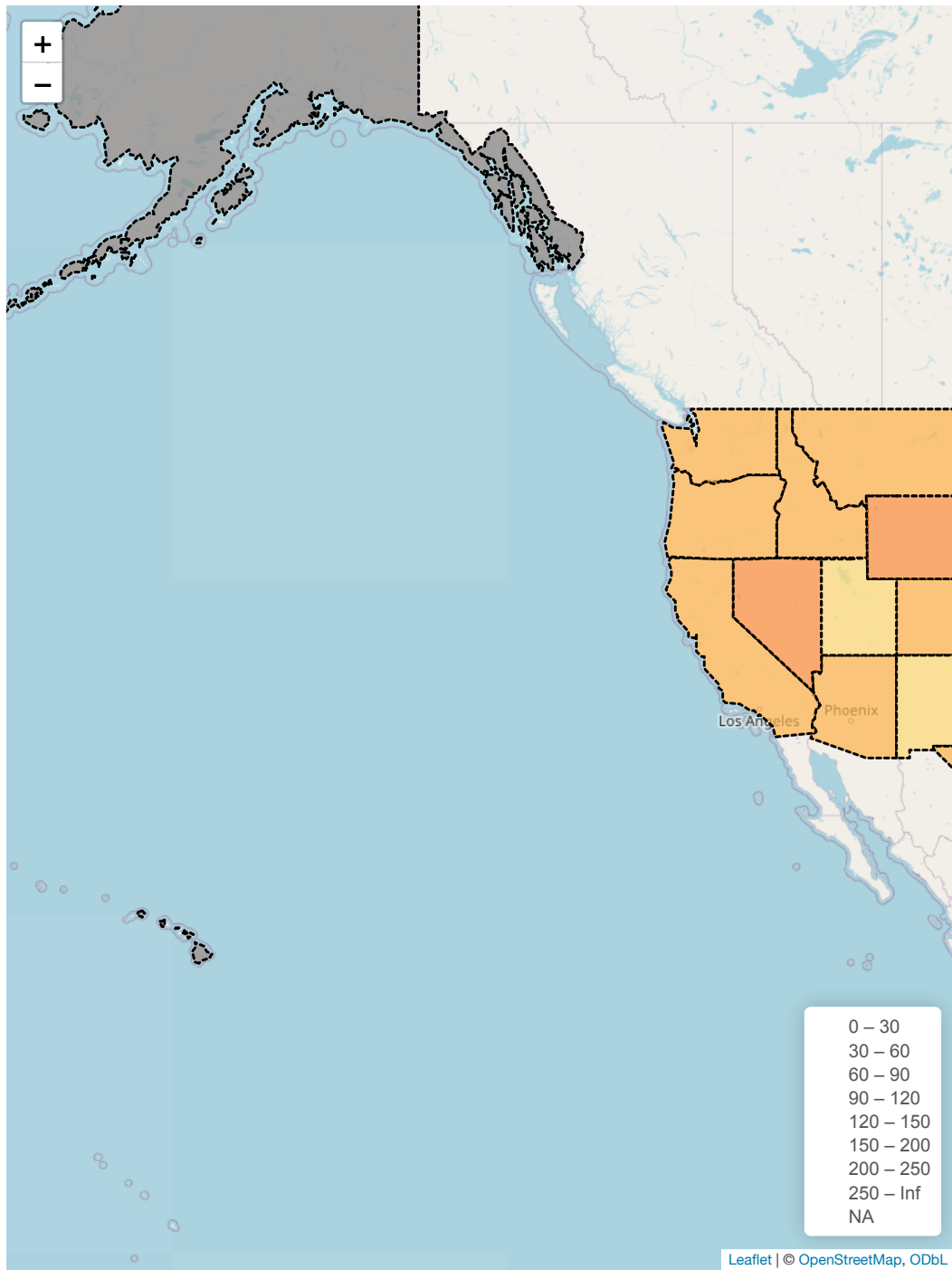
	0 − 30
	30 − 60
	60 − 90
	90 − 120
	120 − 150
	150 − 200
	200 − 250
	250 − Inf
	NA

Leaflet | © OpenStreetMap, ODbL

Second interactive plot - categorical

```r
# Categorize cigarette sales by packs sold
cigarettes_sf <- states_sf %>%
  left_join(cigarettes_data, by = c("name" = "states_names")) %>%
  filter(!(name %in% c("alaska", "hawaii"))) %>%
  mutate(across(where(is.numeric), ~round(.x, 1))) %>%
  mutate(cigarette_category = case_when(
    packs < 90 ~ "Low",
    packs >= 91 & packs < 110 ~ "Medium",
    packs >= 111 & packs < 130 ~ "High",
    packs >= 130 ~ "Very High"
  ))

# Add labels for each state with the number of packs sold
cigarettes_sf <- cigarettes_sf %>%
  mutate(labels = str_c(name, ": ", packs, " packs of cigarettes sold per capita in 1985"))

# Convert labels to HTML format for leaflet
labels <- lapply(cigarettes_sf$labels, HTML)

# Create color palette for categorical data
factpal <- colorFactor(c("lightblue", "cyan2", "mediumturquoise", "darkblue"),
                        levels = c("Low", "Medium", "High", "Very High"))

# Create the interactive map with leaflet
leaflet(cigarettes_sf) %>%
  setView(-96, 37.8, 4) %>%
  addTiles() %>%
  addPolygons(
    weight = 2,
    opacity = 1,
    color = "black",
    fillColor = ~factpal(cigarette_category),  # Color by the categorical variable
    dashArray = "3",
    fillOpacity = 0.7,
    highlightOptions = highlightOptions(
      weight = 5,
      color = "maroon",
      dashArray = "",
      fillOpacity = 0.7,
```

```
    bringToFront = TRUE),
  label = labels,
  labelOptions = labelOptions(
    style = list("font-weight" = "normal", padding = "3px 8px"),
    textsize = "15px",
    direction = "auto")) %>%
addLegend(
  pal = factpal,
  values = ~cigarette_category,   # Use the categorical values for the legend
  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:////private/var/folders/kg/90zf5swd5j9f16nv0swcbdf00000gn/T/RtmpeRnoU3/file8aec5371df14/

High
Low
Medium
Very High
NA