Assignment 3

Jiawei Zhao

2024-06-30

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

# 1.input data  
data <- read\_csv("/Users/jiawei/McDaniel Study/ANA 515/Week 6/Assignment 3/StormEvents\_details-ftp\_v1.0\_d1998\_c20220425.csv")

## Warning: One or more parsing issues, call `problems()` on your data frame for details,  
## e.g.:  
## dat <- vroom(...)  
## problems(dat)

## Rows: 50973 Columns: 51  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (20): STATE, MONTH\_NAME, EVENT\_TYPE, CZ\_TYPE, CZ\_NAME, WFO, BEGIN\_DATE\_T...  
## dbl (24): BEGIN\_YEARMONTH, BEGIN\_DAY, BEGIN\_TIME, END\_YEARMONTH, END\_DAY, EN...  
## lgl (7): MAGNITUDE\_TYPE, FLOOD\_CAUSE, CATEGORY, TOR\_OTHER\_WFO, TOR\_OTHER\_CZ...  
##   
## 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.

library(dplyr)  
# 2.Select the required columns  
data <- data %>% select(BEGIN\_YEARMONTH, EPISODE\_ID, STATE, STATE\_FIPS, CZ\_NAME, CZ\_TYPE, CZ\_FIPS, EVENT\_TYPE)

# 3.Arrange by state name  
data <- data %>% arrange(STATE)

library(stringr)  
# 4.Convert state and county names to title case  
data <- data %>%   
 mutate(STATE = str\_to\_title(STATE),  
 CZ\_NAME = str\_to\_title(CZ\_NAME))

# 5.Filter for county events and remove CZ\_TYPE column  
data <- data %>%   
 filter(CZ\_TYPE == "C") %>%   
 select(-CZ\_TYPE)

# 6.Pad state and county FIPS and unite into one column  
data <- data %>%   
 mutate(STATE\_FIPS = str\_pad(STATE\_FIPS, 2, pad = "0"),  
 CZ\_FIPS = str\_pad(CZ\_FIPS, 3, pad = "0"),  
 FIPS = paste0(STATE\_FIPS, CZ\_FIPS)) %>%   
 select(-STATE\_FIPS, -CZ\_FIPS)

# 7.Change all column names to lower case  
data <- data %>% rename\_all(tolower)

# Load base R state data  
data("state")  
# 8.Create dataframe with state name, area, and region  
state\_info <- data.frame(state = state.name, area = state.area, region = state.region)

# 9.Count number of events per state  
events\_per\_state <- data %>%   
 group\_by(state) %>%   
 summarize(num\_events = n())  
# Merge with state information  
merged\_data <- merge(events\_per\_state, state\_info, by.x = "state", by.y = "state")  
# Remove states not in state information dataframe  
merged\_data <- merged\_data %>%   
 filter(state %in% state\_info$state)

# 10.Plot number of events per state  
library(ggplot2)  
  
ggplot(merged\_data, aes(x = area, y = num\_events, color = region)) +  
 geom\_point(size = 1, alpha = 0.7) +  
 labs(title = "Number of Storm Events per State in 1998", x = "State Area", y = "Number of Events") +  
 theme\_minimal() +  
 scale\_color\_brewer(palette = "Set1") +  
 theme(legend.position = "top")

