ANA 515 Assigment 3

Daniel Appadurai

2023-07-01

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── 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

library(dplyr)  
library(knitr)  
library(rmarkdown)  
library(officer)

# 1. Reading the csv file and checking the columns  
data<-read.csv("StormEvents\_details-ftp\_v1.0\_d2013\_c20230118.csv")  
colnames(data)

[1] "BEGIN\_YEARMONTH" "BEGIN\_DAY" "BEGIN\_TIME"   
 [4] "END\_YEARMONTH" "END\_DAY" "END\_TIME"   
 [7] "EPISODE\_ID" "EVENT\_ID" "STATE"   
[10] "STATE\_FIPS" "YEAR" "MONTH\_NAME"   
[13] "EVENT\_TYPE" "CZ\_TYPE" "CZ\_FIPS"   
[16] "CZ\_NAME" "WFO" "BEGIN\_DATE\_TIME"   
[19] "CZ\_TIMEZONE" "END\_DATE\_TIME" "INJURIES\_DIRECT"   
[22] "INJURIES\_INDIRECT" "DEATHS\_DIRECT" "DEATHS\_INDIRECT"   
[25] "DAMAGE\_PROPERTY" "DAMAGE\_CROPS" "SOURCE"   
[28] "MAGNITUDE" "MAGNITUDE\_TYPE" "FLOOD\_CAUSE"   
[31] "CATEGORY" "TOR\_F\_SCALE" "TOR\_LENGTH"   
[34] "TOR\_WIDTH" "TOR\_OTHER\_WFO" "TOR\_OTHER\_CZ\_STATE"  
[37] "TOR\_OTHER\_CZ\_FIPS" "TOR\_OTHER\_CZ\_NAME" "BEGIN\_RANGE"   
[40] "BEGIN\_AZIMUTH" "BEGIN\_LOCATION" "END\_RANGE"   
[43] "END\_AZIMUTH" "END\_LOCATION" "BEGIN\_LAT"   
[46] "BEGIN\_LON" "END\_LAT" "END\_LON"   
[49] "EPISODE\_NARRATIVE" "EVENT\_NARRATIVE" "DATA\_SOURCE"

# 2. Limiting the columns to the ones listed below  
selected\_col<-c("BEGIN\_YEARMONTH","EPISODE\_ID","STATE","STATE\_FIPS",  
"CZ\_NAME","CZ\_TYPE","CZ\_FIPS","EVENT\_TYPE")  
  
data2<-select(data, selected\_col)

Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.  
ℹ Please use `all\_of()` or `any\_of()` instead.  
 # Was:  
 data %>% select(selected\_col)  
  
 # Now:  
 data %>% select(all\_of(selected\_col))  
  
See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.  
This warning is displayed once every 8 hours.  
Call `lifecycle::last\_lifecycle\_warnings()` to see where this warning was  
generated.

colnames(data2)

[1] "BEGIN\_YEARMONTH" "EPISODE\_ID" "STATE" "STATE\_FIPS"   
[5] "CZ\_NAME" "CZ\_TYPE" "CZ\_FIPS" "EVENT\_TYPE"

#3. Arranging the data by State Name  
data3<-data2[order(data2$STATE),]  
sub\_data3<-data3[1:3]  
kable(head(sub\_data3))

|  | BEGIN\_YEARMONTH | EPISODE\_ID | STATE |
| --- | --- | --- | --- |
| 287 | 201301 | 71249 | ALABAMA |
| 288 | 201301 | 71249 | ALABAMA |
| 289 | 201301 | 71249 | ALABAMA |
| 2302 | 201301 | 71249 | ALABAMA |
| 2303 | 201301 | 71249 | ALABAMA |
| 2304 | 201301 | 71249 | ALABAMA |
|  |  |  |  |

#4.Change state and county names to title case   
library(stringr)  
data4 <- data3 %>% mutate(STATE = str\_to\_title(STATE)) %>% mutate(CZ\_NAME=str\_to\_title(CZ\_NAME))  
sub\_data4<-data4[1:5]  
kable(head(sub\_data4))

|  | BEGIN\_YEARMONTH | EPISODE\_ID | STATE | STATE\_FIPS | CZ\_NAME |
| --- | --- | --- | --- | --- | --- |
| 287 | 201301 | 71249 | Alabama | 1 | Elmore |
| 288 | 201301 | 71249 | Alabama | 1 | Jefferson |
| 289 | 201301 | 71249 | Alabama | 1 | Calhoun |
| 2302 | 201301 | 71249 | Alabama | 1 | Calhoun |
| 2303 | 201301 | 71249 | Alabama | 1 | Jefferson |
| 2304 | 201301 | 71249 | Alabama | 1 | Randolph |

unique(data4$CZ\_TYPE)

## [1] "C" "Z"

#5. Limit to the events listed by county FIPS (CZ\_TYPE of “C”)   
data5<-subset(data4, CZ\_TYPE =="C")  
unique(data5$CZ\_TYPE)

## [1] "C"

#5. removing the CZ\_TYPE column  
data5<-subset(data5, select = -c(CZ\_TYPE))  
colnames(data5)

## [1] "BEGIN\_YEARMONTH" "EPISODE\_ID" "STATE" "STATE\_FIPS"   
## [5] "CZ\_NAME" "CZ\_FIPS" "EVENT\_TYPE"

#6.Pad the state FIPS with a “0” at the beginning (hint: there’s a function in stringr to do this) and then unite the two columns to make one FIPS column with the new state-county FIPS code  
data5$STATE\_FIPS<-formatC(data5$STATE\_FIPS, width = 2, format = "d", flag="0")  
  
  
head(data5$STATE\_FIPS)

[1] "01" "01" "01" "01" "01" "01"

#6. Pad the county FIPS with a “0”   
data5$CZ\_FIPS<-formatC(data5$CZ\_FIPS, width = 4, format = "d", flag="0")  
head(data5$CZ\_FIPS)

## [1] "0051" "0073" "0015" "0015" "0073" "0111"

#6. Unite the two columns to make one FIPS column with the new state-county FIPS code  
data5$STATE\_COUNTY\_FIPS<-paste0(data5$STATE\_FIPS,data5$CZ\_FIPS)  
head(data5$STATE\_COUNTY\_FIPS)

## [1] "010051" "010073" "010015" "010015" "010073" "010111"

#7.Change all the column names to lower case (you may want to try the rename\_all function for this)  
colnames(data5) <- tolower(colnames(data5))  
  
colnames(data5)

[1] "begin\_yearmonth" "episode\_id" "state"   
[4] "state\_fips" "cz\_name" "cz\_fips"   
[7] "event\_type" "state\_county\_fips"

unique(data5$event\_type)

## [1] "Thunderstorm Wind" "Flash Flood" "Hail"   
## [4] "Tornado" "Flood" "Lightning"   
## [7] "Heavy Rain" "Funnel Cloud" "Dust Devil"   
## [10] "Debris Flow"

#8.There is data that comes with base R on U.S. states (data("state")). Use that to create a dataframe with these three columns: state name, area, and region  
data("state")  
df<-data.frame(state.name,state.area,state.region)  
head(df)

state.name state.area state.region  
1 Alabama 51609 South  
2 Alaska 589757 West  
3 Arizona 113909 West  
4 Arkansas 53104 South  
5 California 158693 West  
6 Colorado 104247 West

#9. gathering storm events count for each state and creating a dataframe "events"  
storm\_events<- table(data5$state)  
events<-data.frame(storm\_events)  
head(events)

Var1 Freq  
1 Alabama 784  
2 Alaska 23  
3 American Samoa 1  
4 Arizona 356  
5 Arkansas 768  
6 California 207

#9. Renaming columns to create matching columns "state.name" to faciliate merging storm events dataframe with the state dataframe.  
colnames(events)<- c("state.name","Number of storm events")  
head(events)

state.name Number of storm events  
1 Alabama 784  
2 Alaska 23  
3 American Samoa 1  
4 Arizona 356  
5 Arkansas 768  
6 California 207

#9. Merging storm events with state dataframe  
new\_df<-merge(df, events, by="state.name")  
head(new\_df)

state.name state.area state.region Number of storm events  
1 Alabama 51609 South 784  
2 Alaska 589757 West 23  
3 Arizona 113909 West 356  
4 Arkansas 53104 South 768  
5 California 158693 West 207  
6 Colorado 104247 West 617

#9. Embellishing the columns with title case  
colnames(new\_df)<-c("State\_Name","State\_Area", "State\_Region", "Number\_of\_Storm\_Events" )  
head(new\_df)

State\_Name State\_Area State\_Region Number\_of\_Storm\_Events  
1 Alabama 51609 South 784  
2 Alaska 589757 West 23  
3 Arizona 113909 West 356  
4 Arkansas 53104 South 768  
5 California 158693 West 207  
6 Colorado 104247 West 617

#10. Creating a plot with land area on the X axis and Number of Storm events on the Y axis.   
library(ggplot2)  
plt<-ggplot(new\_df, aes(x = State\_Area, y = Number\_of\_Storm\_Events , fill = State\_Region)) +  
 geom\_point(shape = 21, size = 4) +  
 labs(x = "Land Area", y = "# of storm events", fill = "Region") +  
 ggtitle("Number of Storm Events by Land Area in 2013")  
  
plt

