ANA515_Assignmnet3_Jiaxuan_Ren

Jiaxuan Ren

2023-07-01

1. Download the bulk storm details data for 1987 and save it into my working directory. Read and save the data into the dataframe.

```
all_data <- read_csv("StormEvents_details-ftp_v1.0_d1987_c20220425.csv")
```

2. Limit the dataframe to the required columns.

```
mydata <- subset(all_data, select = c(BEGIN_YEARMONTH, EPISODE_ID,
    STATE, STATE_FIPS, CZ_NAME, CZ_TYPE, CZ_FIPS, EVENT_TYPE))
glimpse(mydata)
```

```
Rows: 7,363
Columns: 8
$ BEGIN_YEARMONTH <dbl> 198705, 198705, 198705, 198705, 198705, 198710, 198709~
$ EPISODE_ID      <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ STATE           <chr> "COLORADO", "COLORADO", "COLORADO", "COLORADO", "COLOR~
$ STATE_FIPS      <dbl> 8, 8, 8, 8, 8, 4, 10, 27, 28, 30, 20, 20, 20, 20, 20, ~
$ CZ_NAME         <chr> "PHILLIPS", "DENVER", "DOUGLAS", "WASHINGTON", "WASHIN~
$ CZ_TYPE         <chr> "C", "C", "C", "C", "C", "C", "C", "C", "C", "C", "C",~
$ CZ_FIPS         <dbl> 95, 31, 35, 121, 121, 13, 3, 111, 129, 95, 59, 3, 3, 1~
$ EVENT_TYPE      <chr> "Thunderstorm Wind", "Tornado", "Thunderstorm Wind", "~
```

3. Arrange the data by the state name.

```
mydata <- mydata %>%
    arrange(STATE)
head(mydata)
```

```
# A tibble: 6 x 8
  BEGIN_YEARMONTH EPISODE_ID STATE STATE_FIPS CZ_NAME CZ_TYPE CZ_FIPS EVENT_TYPE
            <dbl> <lgl>      <chr>      <dbl> <chr>   <chr>     <dbl> <chr>
1          198708 NA         ALAB~          1 LAUDER~ C            77 Thunderst~
2          198708 NA         ALAB~          1 COVING~ C            39 Hail
3          198708 NA         ALAB~          1 WALKER  C           127 Thunderst~
4          198708 NA         ALAB~          1 COLBERT C            33 Thunderst~
5          198708 NA         ALAB~          1 CLAY    C            27 Thunderst~
6          198708 NA         ALAB~          1 JEFFER~ C            73 Hail
```

4. Change state and county names to title case.

```
mydata <- mydata %>%
    mutate(STATE = str_to_title(STATE)) %>%
    mutate(CZ_NAME = str_to_title(CZ_NAME))
head(mydata)
```

```
# A tibble: 6 x 8
  BEGIN_YEARMONTH EPISODE_ID STATE STATE_FIPS CZ_NAME CZ_TYPE CZ_FIPS EVENT_TYPE
            <dbl> <lgl>      <chr>      <dbl> <chr>   <chr>     <dbl> <chr>
1          198708 NA         Alab~          1 Lauder~ C            77 Thunderst~
2          198708 NA         Alab~          1 Coving~ C            39 Hail
3          198708 NA         Alab~          1 Walker  C           127 Thunderst~
4          198708 NA         Alab~          1 Colbert C            33 Thunderst~
5          198708 NA         Alab~          1 Clay    C            27 Thunderst~
6          198708 NA         Alab~          1 Jeffer~ C            73 Hail
```

5. Limit to the events listed by county FIPS (CZ_TYPE of "C") and then remove the CZ_TYPE column.

```
filtered_data <- mydata %>%
    filter(CZ_TYPE == "C") %>%
    select(-c(CZ_TYPE))
head(filtered_data)
```

```
# A tibble: 6 x 7
  BEGIN_YEARMONTH EPISODE_ID STATE   STATE_FIPS CZ_NAME     CZ_FIPS EVENT_TYPE
            <dbl> <lgl>      <chr>        <dbl> <chr>         <dbl> <chr>
1          198708 NA         Alabama          1 Lauderdale      77 Thunderstorm~
2          198708 NA         Alabama          1 Covington       39 Hail
3          198708 NA         Alabama          1 Walker         127 Thunderstorm~
4          198708 NA         Alabama          1 Colbert         33 Thunderstorm~
5          198708 NA         Alabama          1 Clay            27 Thunderstorm~
6          198708 NA         Alabama          1 Jefferson       73 Hail
```

6. Pad the state and county FIPS with a "0" at the beginning and then unite the two columns to make one FIPS column with the new state-county FIPS code.

```
padded_data <- mutate(filtered_data, STATE_FIPS = str_pad(filtered_data$STATE_FIPS,
    width = 3, side = "left", pad = "0"))
padded_data <- mutate(filtered_data, CZ_FIPS = str_pad(filtered_data$CZ_FIPS,
    width = 4, side = "left", pad = "0"))
padded_data <- padded_data %>%
    unite("fips", STATE_FIPS, CZ_FIPS)
head(padded_data)
```

```
# A tibble: 6 x 6
  BEGIN_YEARMONTH EPISODE_ID STATE   fips   CZ_NAME    EVENT_TYPE
            <dbl> <lgl>      <chr>   <chr>  <chr>      <chr>
1          198708 NA         Alabama 1_0077 Lauderdale Thunderstorm Wind
2          198708 NA         Alabama 1_0039 Covington  Hail
3          198708 NA         Alabama 1_0127 Walker     Thunderstorm Wind
4          198708 NA         Alabama 1_0033 Colbert    Thunderstorm Wind
5          198708 NA         Alabama 1_0027 Clay       Thunderstorm Wind
6          198708 NA         Alabama 1_0073 Jefferson  Hail
```

7. Change all the column names to lowercase.

```
padded_data <- padded_data %>%
    rename_all(tolower)
head(padded_data)
```

```
# A tibble: 6 x 6
  begin_yearmonth episode_id state   fips   cz_name    event_type
            <dbl> <lgl>      <chr>   <chr>  <chr>       <chr>
1          198708 NA         Alabama 1_0077 Lauderdale Thunderstorm Wind
2          198708 NA         Alabama 1_0039 Covington  Hail
3          198708 NA         Alabama 1_0127 Walker     Thunderstorm Wind
4          198708 NA         Alabama 1_0033 Colbert    Thunderstorm Wind
5          198708 NA         Alabama 1_0027 Clay       Thunderstorm Wind
6          198708 NA         Alabama 1_0073 Jefferson  Hail
```

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")
us_state_info <- data.frame(state = state.name, region = state.region,
    area = state.area)
head(us_state_info)
```

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

9. Create a dataframe with the number of events per state in the year of your birth. Merge in the state information dataframe you just created in step 8. Remove any states that are not in the state information dataframe.

```
# Count the event frequency of each state
events_count <- data.frame(table(padded_data$state))

# Merge in the state information dataframe
events_count <- events_count %>%
    rename(c(state = "Var1")) %>%
    rename_all(tolower)
merged_state_info <- merge(x = events_count, y = us_state_info,
    by.x = "state", by.y = "state")
head(merged_state_info)
```

```
        state freq    region    area
1     Alabama  155     South   51609
2     Arizona   33      West  113909
3    Arkansas  162     South   53104
4  California   13      West  158693
5    Colorado  186      West  104247
6 Connecticut    4 Northeast    5009
```

10. Create the plot

```
library(ggplot2)
storm_plot <- ggplot(merged_state_info, aes(x = area, y = freq)) +
    geom_point(aes(color = region)) + labs(x = "Land area (square miles)",
    y = "# of storm events in 1987")
storm_plot
```
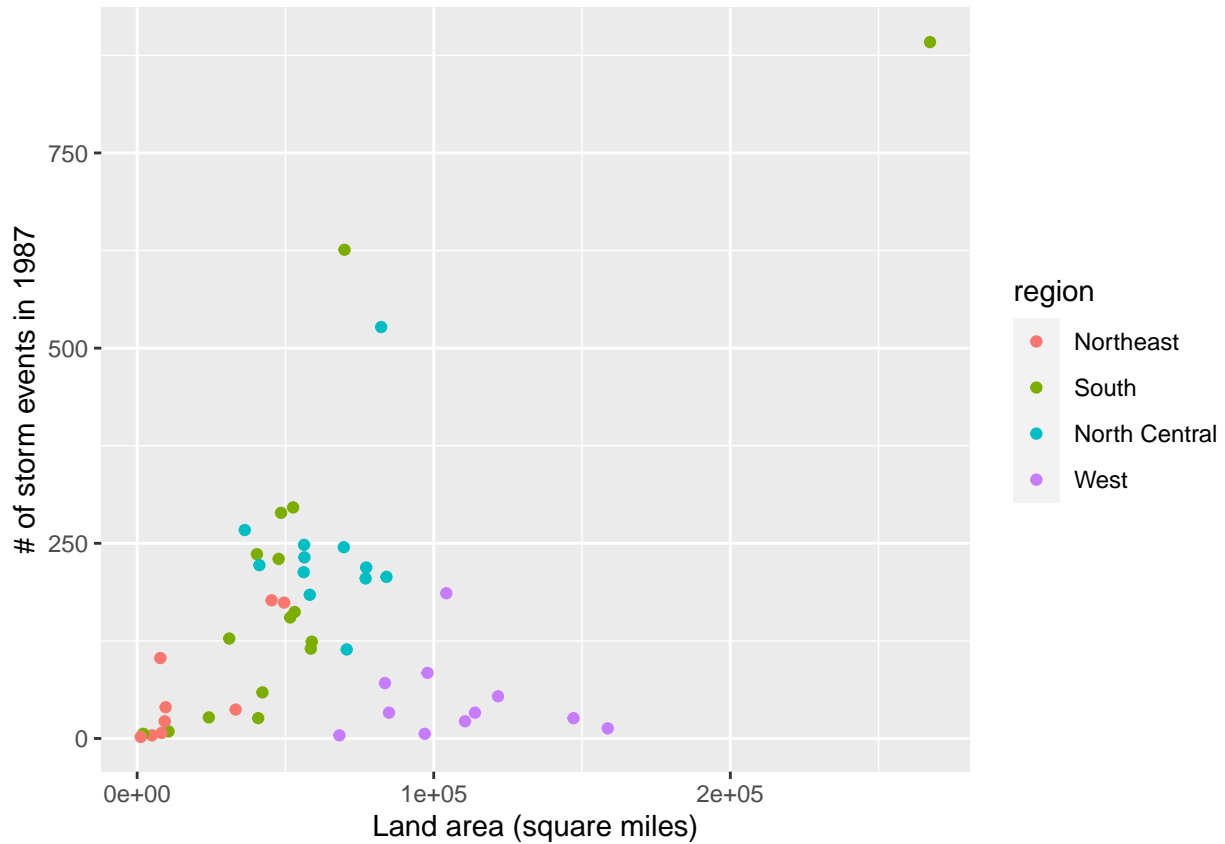


```
# Save the plot into a png file
ggsave("storm_plot.png")
```