ANA 515 Assignment 3

Siliang Gong

Nov 28, 2021

Read the data

```
## read data
 analysis.data <- read.csv("StormEvents_details-ftp_v1.0_d1991_c20210803.csv")
```

Select a subset of features

```
## select a subset of features

event.data <- analysis.data%>%select(BEGIN_DATE_TIME, END_DATE_TIME, EPISODE_ID, EVENT_ID,
STATE, STATE_FIPS, CZ_NAME, CZ_TYPE, CZ_FIPS, EVENT_TYPE, SOURCE, BEGIN_LAT,
BEGIN_LON, END_LAT, END_LON)
```

Convert the beginning and ending dates

```
event.data <- event.data%>%mutate(BEGIN_DATE_TIME = dmy_hms(BEGIN_DATE_TIME),
                                  END_DATE_TIME = dmy_hms(END_DATE_TIME))
```

Change state and county

```
event.data <- event.data%>%mutate(STATE = str_to_title(STATE))
head(event.data$STATE, 5)
```

```
## [1] "Colorado" "Colorado" "Colorado" "Colorado" "Colorado"
```

Filter by CZ_TYPE and then remove CZ_TYPE

```
event.data <- event.data%>%filter(CZ_TYPE == "C")%>%select(-CZ_TYPE)
```

Pad state and county FIPS

```
event.data <- event.data%>%mutate(STATE_FIPS = str_pad(STATE_FIPS, width = 3, side = "left", pad = "0")


event.data <- unite_(event.data, "FIPS", c("STATE_FIPS", "CZ_FIPS"), sep = "")

head(event.data$FIPS, 5)
```

```
## [1] "008013" "008013" "008059" "008001" "008073"
```

Rename all column names to lower case

```
event.data = rename_all(event.data, tolower)

colnames(event.data)
```

```
##  [1] "begin_date_time" "end_date_time"   "episode_id"
##  [4] "event_id"        "state"           "fips"
##  [7] "cz_name"         "event_type"      "source"
## [10] "begin_lat"       "begin_lon"       "end_lat"
## [13] "end_lon"
```

load dataset state and create a dataframe

```
data("state")
us.state.info <- data.frame(state = state.name, region =
                              state.region, area = state.area)
head(us.state.info, 5)
```

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

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

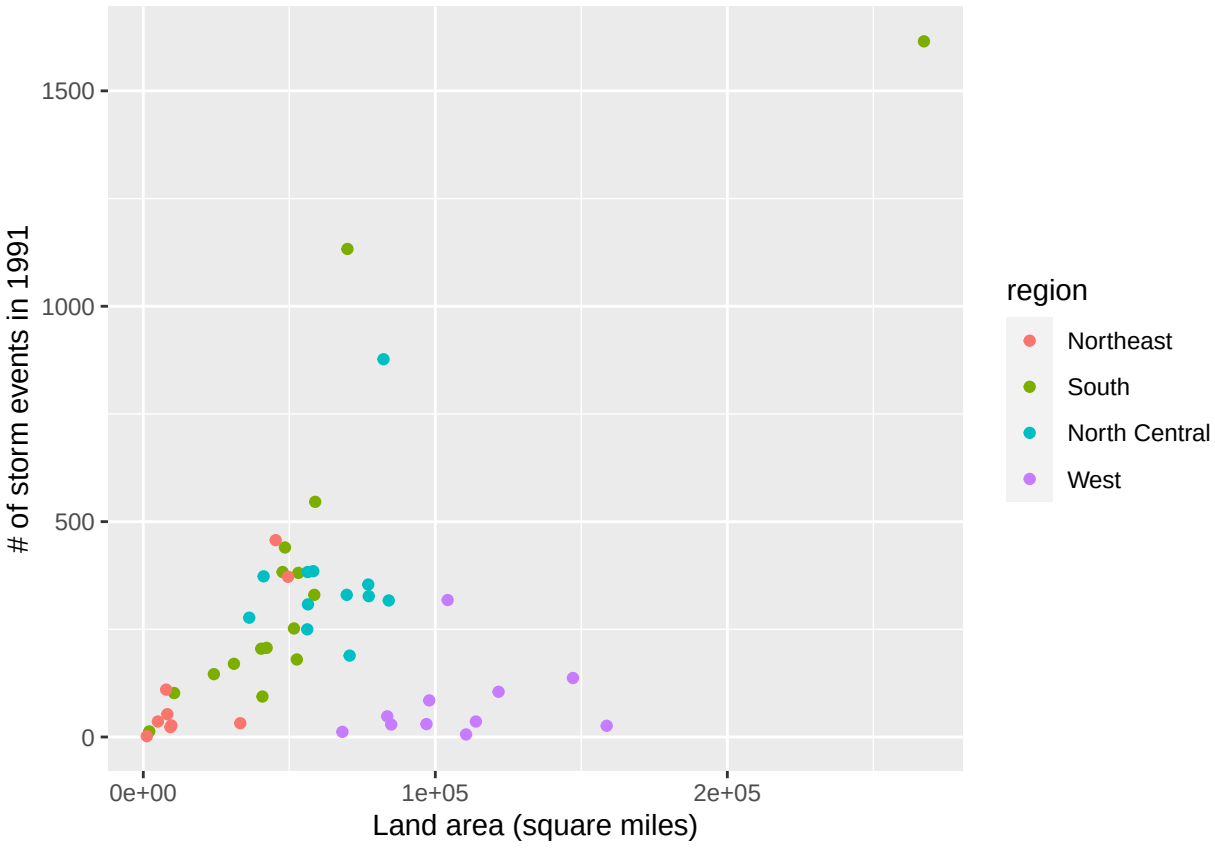
```
newset = data.frame(table(event.data$state))
newset1 <-  rename(newset, c("state" = "Var1"))

merged <- merge(newset1, us.state.info, by = "state")
head(merged, 5)
```

```
##        state Freq region   area
## 1    Alabama  252  South  51609
## 2    Arizona   36   West 113909
## 3   Arkansas  381  South  53104
## 4 California   26   West 158693
## 5   Colorado  318   West 104247
```

Create a plot

```
storm_plot <- ggplot(merged, aes(x = area, y = Freq))+
  geom_point(aes(color = region)) +
  labs(x = "Land area (square miles)",
       y = "# of storm events in 1991")
storm_plot
```


1500
1000
500
0
# of storm events in 1991
0e+00
1e+05
2e+05
Land area (square miles)
region
Northeast
South
North Central
West