Anita Blege - Assignment 3 (Getting & Cleaning Data)

Anita Blege

2022-10-01

The data has about 13534 rows and 51 columns.

fil\_dropped <- select(filtered, -CZ\_TYPE)  
head(fil\_dropped)

## BEGIN\_YEARMONTH BEGIN\_DATE\_TIME END\_DATE\_TIME BEGIN\_DAY BEGIN\_TIME  
## 1 199201 1/14/1992 10:30 1/14/1992 10:30 14 1030  
## 2 199201 1/14/1992 10:31 1/14/1992 10:31 14 1031  
## 3 199201 1/14/1992 10:34 1/14/1992 10:34 14 1034  
## 4 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 5 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 6 199201 1/14/1992 9:50 1/14/1992 9:50 14 950  
## END\_YEARMONTH END\_DAY END\_TIME EPISODE\_ID EVENT\_ID STATE STATE\_FIPS  
## 1 199201 14 1030 NA 10115132 PENNSYLVANIA 42  
## 2 199201 14 1031 NA 10115133 PENNSYLVANIA 42  
## 3 199201 14 1034 NA 10115134 PENNSYLVANIA 42  
## 4 199201 14 1130 NA 10079696 NEW JERSEY 34  
## 5 199201 14 1130 NA 10079697 NEW JERSEY 34  
## 6 199201 14 950 NA 10114276 PENNSYLVANIA 42  
## MONTH\_NAME CZ\_NAME CZ\_FIPS EVENT\_TYPE SOURCE BEGIN\_LAT BEGIN\_LON  
## 1 January SUSQUEHANNA 115 Thunderstorm Wind NA 41.78 -75.70  
## 2 January PHILADELPHIA 101 Thunderstorm Wind NA 39.97 -75.17  
## 3 January MONTGOMERY 91 Thunderstorm Wind NA 40.28 -75.38  
## 4 January UNION 39 Thunderstorm Wind NA 40.63 -74.27  
## 5 January MIDDLESEX 23 Thunderstorm Wind NA 40.58 -74.30  
## 6 January LACKAWANNA 69 Thunderstorm Wind NA 41.58 -75.50  
## END\_LAT END\_LON  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA

united <- unite(fil\_dropped,col='fips'  
, c('STATE\_FIPS', 'CZ\_FIPS') , sep = " ", remove = TRUE)  
  
head(united)

## BEGIN\_YEARMONTH BEGIN\_DATE\_TIME END\_DATE\_TIME BEGIN\_DAY BEGIN\_TIME  
## 1 199201 1/14/1992 10:30 1/14/1992 10:30 14 1030  
## 2 199201 1/14/1992 10:31 1/14/1992 10:31 14 1031  
## 3 199201 1/14/1992 10:34 1/14/1992 10:34 14 1034  
## 4 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 5 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 6 199201 1/14/1992 9:50 1/14/1992 9:50 14 950  
## END\_YEARMONTH END\_DAY END\_TIME EPISODE\_ID EVENT\_ID STATE fips  
## 1 199201 14 1030 NA 10115132 PENNSYLVANIA 42 115  
## 2 199201 14 1031 NA 10115133 PENNSYLVANIA 42 101  
## 3 199201 14 1034 NA 10115134 PENNSYLVANIA 42 91  
## 4 199201 14 1130 NA 10079696 NEW JERSEY 34 39  
## 5 199201 14 1130 NA 10079697 NEW JERSEY 34 23  
## 6 199201 14 950 NA 10114276 PENNSYLVANIA 42 69  
## MONTH\_NAME CZ\_NAME EVENT\_TYPE SOURCE BEGIN\_LAT BEGIN\_LON END\_LAT  
## 1 January SUSQUEHANNA Thunderstorm Wind NA 41.78 -75.70 NA  
## 2 January PHILADELPHIA Thunderstorm Wind NA 39.97 -75.17 NA  
## 3 January MONTGOMERY Thunderstorm Wind NA 40.28 -75.38 NA  
## 4 January UNION Thunderstorm Wind NA 40.63 -74.27 NA  
## 5 January MIDDLESEX Thunderstorm Wind NA 40.58 -74.30 NA  
## 6 January LACKAWANNA Thunderstorm Wind NA 41.58 -75.50 NA  
## END\_LON  
## 1 NA  
## 2 NA  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA

to\_lower <- rename\_all(united, tolower)  
head(to\_lower)

## begin\_yearmonth begin\_date\_time end\_date\_time begin\_day begin\_time  
## 1 199201 1/14/1992 10:30 1/14/1992 10:30 14 1030  
## 2 199201 1/14/1992 10:31 1/14/1992 10:31 14 1031  
## 3 199201 1/14/1992 10:34 1/14/1992 10:34 14 1034  
## 4 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 5 199201 1/14/1992 11:30 1/14/1992 11:30 14 1130  
## 6 199201 1/14/1992 9:50 1/14/1992 9:50 14 950  
## end\_yearmonth end\_day end\_time episode\_id event\_id state fips  
## 1 199201 14 1030 NA 10115132 PENNSYLVANIA 42 115  
## 2 199201 14 1031 NA 10115133 PENNSYLVANIA 42 101  
## 3 199201 14 1034 NA 10115134 PENNSYLVANIA 42 91  
## 4 199201 14 1130 NA 10079696 NEW JERSEY 34 39  
## 5 199201 14 1130 NA 10079697 NEW JERSEY 34 23  
## 6 199201 14 950 NA 10114276 PENNSYLVANIA 42 69  
## month\_name cz\_name event\_type source begin\_lat begin\_lon end\_lat  
## 1 January SUSQUEHANNA Thunderstorm Wind NA 41.78 -75.70 NA  
## 2 January PHILADELPHIA Thunderstorm Wind NA 39.97 -75.17 NA  
## 3 January MONTGOMERY Thunderstorm Wind NA 40.28 -75.38 NA  
## 4 January UNION Thunderstorm Wind NA 40.63 -74.27 NA  
## 5 January MIDDLESEX Thunderstorm Wind NA 40.58 -74.30 NA  
## 6 January LACKAWANNA Thunderstorm Wind NA 41.58 -75.50 NA  
## end\_lon  
## 1 NA  
## 2 NA  
## 3 NA  
## 4 NA  
## 5 NA  
## 6 NA

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

events\_freq\_data <- data.frame(table(united$STATE))  
head(events\_freq\_data)

## Var1 Freq  
## 1 ALABAMA 347  
## 2 ARIZONA 82  
## 3 ARKANSAS 567  
## 4 CALIFORNIA 33  
## 5 COLORADO 266  
## 6 CONNECTICUT 46

events\_freq\_data1 <- rename(events\_freq\_data, c("state" = "Var1"))  
head(events\_freq\_data1)

## state Freq  
## 1 ALABAMA 347  
## 2 ARIZONA 82  
## 3 ARKANSAS 567  
## 4 CALIFORNIA 33  
## 5 COLORADO 266  
## 6 CONNECTICUT 46

upper\_us\_state\_info <- mutate\_all(us\_state\_info, toupper)  
  
head(upper\_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

merged\_data <- merge (x=events\_freq\_data1, y=upper\_us\_state\_info, by.x="state", by.y = "state")  
  
head(merged\_data)

## state Freq region area  
## 1 ALABAMA 347 SOUTH 51609  
## 2 ARIZONA 82 WEST 113909  
## 3 ARKANSAS 567 SOUTH 53104  
## 4 CALIFORNIA 33 WEST 158693  
## 5 COLORADO 266 WEST 104247  
## 6 CONNECTICUT 46 NORTHEAST 5009

final\_plot <- ggplot(merged\_data, aes(x=area, y=Freq)) + geom\_point(aes(color=region)) + labs(x= "Land Area (Square Miles)", y= "Number of Storm Events in 1992")  
  
final\_plot

