Mapping

Maggie Sha

2020/10/27

Introduction

This document works on the data from FEMA, and shows every step I have taken to acquire, clean, organize and map the data.

Data Source

I am using a publically available dataset from FEMA.

- The data is from https://www.fema.gov/openfema-data-page/public-assistance-funded-projects-details-v1
- In this report, I only focus on Hurricane data from year 2009 to 2018.

Data Cleaning and Organization

```
# import data and filter hurricane data from 2009 to 2018
data <- read.csv("PublicAssistanceFundedProjectsDetails.csv",header=TRUE)
data <- data %>% filter(incidentType == "Hurricane")
data$declarationDate <- ymd_hms(data$declarationDate)
data <- data %>% filter(2009 <= year(data$declarationDate))
data <- data %>% filter(year(data$declarationDate) <= 2018)

#unique(data$state)
 # state <- c("Alabama","Texas","Virgin Islands of the U.S.","North Carolina","Massachusetts","Puerto R
 #           "New York","Virginia","New Hampshire","Maryland","Delaware","West Virginia","Louisiana","
 #           "New Jersey","Vermont","Connecticut","Pennsylvania","Rhode Island","Maine","District of C
 #           "Mississippi","Ohio","Georgia","South Carolina","American Samoa","Hawaii")

myState <- tolower(unique(data$state))
county <- tolower(unique(data$county))

MainStates <- map_data("state",myState)
AllCounty <- map_data("county",myState)

# length(unique(MainStates$region))
# [1] 23

# length(unique(AllCounty$subregion))
# [1] 914
```

There are 23 unique States and 914 counties in my hurricane data.

This is how the hurricane data looks like:

```
head(data,n=10)
```

```
##    disasterNumber     declarationDate incidentType pwNumber applicationTitle
## 1            1866 2009-12-22 05:00:00    Hurricane        1     DIW-097-02F
## 2            1866 2009-12-22 05:00:00    Hurricane        2     DIW-097-01F
## 3            1866 2009-12-22 05:00:00    Hurricane        3     DIW-097-03F
## 4            1866 2009-12-22 05:00:00    Hurricane        4     DIW-097-04F
## 5            1866 2009-12-22 05:00:00    Hurricane        5     DIW-097-01B
## 6            1866 2009-12-22 05:00:00    Hurricane        6         FOL-01B
## 7            1866 2009-12-22 05:00:00    Hurricane        7         BAL-01B
## 8            1866 2009-12-22 05:00:00    Hurricane        8        D102ADLR
## 9            1866 2009-12-22 05:00:00    Hurricane        9         D102CJM
## 10           1866 2009-12-22 05:00:00    Hurricane       10         DI02CDR
##       applicantId      damageCategoryCode dcc     damageCategory projectSize
## 1   097-U15P3-00      F - Public Utilities  F     Public Utilities       Large
## 2   097-U15P3-00      F - Public Utilities  F     Public Utilities       Large
## 3   097-U15P3-00      F - Public Utilities  F     Public Utilities       Large
## 4   097-U15P3-00      F - Public Utilities  F     Public Utilities       Small
## 5   097-U15P3-00 B - Protective Measures    B Protective Measures       Small
## 6   003-26992-00 B - Protective Measures    B Protective Measures       Small
## 7   003-99003-00 B - Protective Measures    B Protective Measures       Small
## 8   097-19744-00      A - Debris Removal    A      Debris Removal       Small
## 9   097-19744-00   C - Roads and Bridges    C   Roads and Bridges       Small
## 10  097-19744-00   C - Roads and Bridges    C   Roads and Bridges       Small
##      county countyCode   state stateCode stateNumberCode projectAmount
## 1    Mobile         97 Alabama        AL               1          0.00
## 2    Mobile         97 Alabama        AL               1      58425.34
## 3    Mobile         97 Alabama        AL               1          0.00
## 4    Mobile         97 Alabama        AL               1      12778.47
## 5    Mobile         97 Alabama        AL               1      15290.26
## 6   Baldwin          3 Alabama        AL               1       9820.02
## 7   Baldwin          3 Alabama        AL               1      22003.57
## 8    Mobile         97 Alabama        AL               1      32126.92
## 9    Mobile         97 Alabama        AL               1       8876.92
## 10   Mobile         97 Alabama        AL               1       2998.96
##    federalShareObligated totalObligated              obligatedDate
## 1                   0.00           0.00 2010-01-28T03:33:40.000Z
## 2               43819.01       43819.01 2010-01-28T03:33:40.000Z
## 3                   0.00           0.00 2010-01-28T03:33:40.000Z
## 4                9583.85        9583.85 2010-01-28T03:33:40.000Z
## 5               11467.70       11467.70 2010-01-28T03:33:40.000Z
## 6                7365.02        7365.02 2010-01-28T03:33:40.000Z
## 7               16502.68       16502.68 2010-01-28T03:33:40.000Z
## 8               24095.19       24095.19 2010-01-28T03:33:40.000Z
## 9                6657.69        6657.69 2010-01-28T03:33:40.000Z
## 10               2249.22        2249.22 2010-08-20T22:27:11.000Z
##                                  hash              lastRefresh
## 1  31ce5206bca27f03ed8f35a8a42ec896 2020-06-15T11:33:48.982Z
## 2  2208b99d98f3ad0858e5bfe615c63ccc 2020-06-15T11:33:48.983Z
## 3  a60cf74f0c30e90dc5eadceb0512307e 2020-06-15T11:33:48.984Z
```

```
## 4   81e4c0dcd9873684e927069be0274394 2020-06-15T11:33:48.985Z
## 5   2296b207e4b3118ec6a8ded1f11dcfd3 2020-06-15T11:33:48.986Z
## 6   ab0cd713b8357092a33c2564ebb4c083 2020-06-15T11:33:48.987Z
## 7   bc596df1be80e806d4496ccef7b568a1 2020-06-15T11:33:48.988Z
## 8   cf748f9cdef691f0168f5a6ef86c5706 2020-06-15T11:33:48.990Z
## 9   c046f8d00903d9f12c64f37bc2b4f662 2020-06-15T11:33:48.991Z
## 10 739266eb91cf0137338f3bb35dce1818 2020-06-15T11:33:48.992Z
##                               id
## 1   5ee75c9c556129600ddb60c3
## 2   5ee75c9c556129600ddb60c4
## 3   5ee75c9c556129600ddb60c5
## 4   5ee75c9c556129600ddb60c6
## 5   5ee75c9c556129600ddb60c7
## 6   5ee75c9c556129600ddb60c8
## 7   5ee75c9c556129600ddb60c9
## 8   5ee75c9c556129600ddb60ca
## 9   5ee75c9c556129600ddb60cb
## 10 5ee75c9c556129600ddb60cc
```

There are many attributes that I am not going to use or compare. I will only keep several columns and do future data cleaning from here.

```r
# select columns I need
cTable <- data%>%select("declarationDate","county","countyCode","state","projectSize","projectAmount","

# get a new table contains projectTotal
sumProject <- cTable %>% group_by(countyCode) %>% dplyr::summarise(projectTotal = sum(projectAmount))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```r
# merge cTable and projectTotal
cTable <- merge(sumProject,cTable,by = "countyCode")

cTable %<>% dplyr::rename(region=state, subregion=county)
cTable$region %<>% tolower()
cTable$subregion %<>% tolower()

# correct those counties' names which contain " (city)"
cTable$subregion <- gsub("\\ \\(city\\)","",cTable$subregion)


# join the table with Allcounty to match latitude and longitude
# filter out miss-matched rows
cTable <- left_join(cTable,AllCounty,by = "subregion")
cTable %<>% filter(long != "NA")

# find which counties have project amount greater than 100 million.
cTable %<>% mutate(`projectAmount > 100000000` = ifelse(cTable$projectTotal > 100000000, "True", "False


# length(unique(cTable$region.x))
# [1] 22
```

```r
# length(unique(cTable$subregion))
# [1] 525
```

Then I am done with the data cleaning and organization part. Now the cTable has 22 distinct States and 525 counties. This is how cTable looks like:

```r
head(cTable,n=10)
```

```
##    countyCode projectTotal     declarationDate subregion    region.x
## 1           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 2           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 3           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 4           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 5           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 6           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 7           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 8           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 9           1    173086414 2011-09-02 04:00:00 fairfield connecticut
## 10          1    173086414 2011-09-02 04:00:00 fairfield connecticut
##    projectSize projectAmount federalShareObligated totalObligated      long
## 1        Small       2366.33               1774.75        1774.75 -73.53341
## 2        Small       2366.33               1774.75        1774.75 -73.49902
## 3        Small       2366.33               1774.75        1774.75 -73.49902
## 4        Small       2366.33               1774.75        1774.75 -73.50475
## 5        Small       2366.33               1774.75        1774.75 -73.49329
## 6        Small       2366.33               1774.75        1774.75 -73.46465
## 7        Small       2366.33               1774.75        1774.75 -73.44746
## 8        Small       2366.33               1774.75        1774.75 -73.39017
## 9        Small       2366.33               1774.75        1774.75 -73.37297
## 10       Small       2366.33               1774.75        1774.75 -73.32714
##         lat group order   region.y projectAmount > 100000000
## 1  41.67695    68  3160 connecticut                      True
## 2  41.66549    68  3161 connecticut                      True
## 3  41.65403    68  3162 connecticut                      True
## 4  41.63684    68  3163 connecticut                      True
## 5  41.60819    68  3164 connecticut                      True
## 6  41.55090    68  3165 connecticut                      True
## 7  41.49934    68  3166 connecticut                      True
## 8  41.51079    68  3167 connecticut                      True
## 9  41.49360    68  3168 connecticut                      True
## 10 41.48214    68  3169 connecticut                      True
```

```r
write.csv(cTable,"D:\\MA615\\Mapping-pro\\cTable.csv", row.names = FALSE)
```

Mapping

First I want to see the distribution of project size (either "large" or "small")

```r
ggplot() +
  geom_polygon(data=AllCounty, aes(x=long, y=lat, group=group),
               color="gray", fill="white", size = .1 ) +
```

```
geom_polygon(data = cTable, aes(x = long, y = lat, group = group,
                                fill = `projectSize`),
             color = "grey", size = 0.2, alpha = 1.6) +
geom_polygon(data = MainStates, aes(x = long, y = lat, group = group),
             color="black", fill="white", size = 0.2, alpha = 0.3) +
scale_fill_manual(values = c("gold", "blue"))+
ggtitle("Project Size")
```
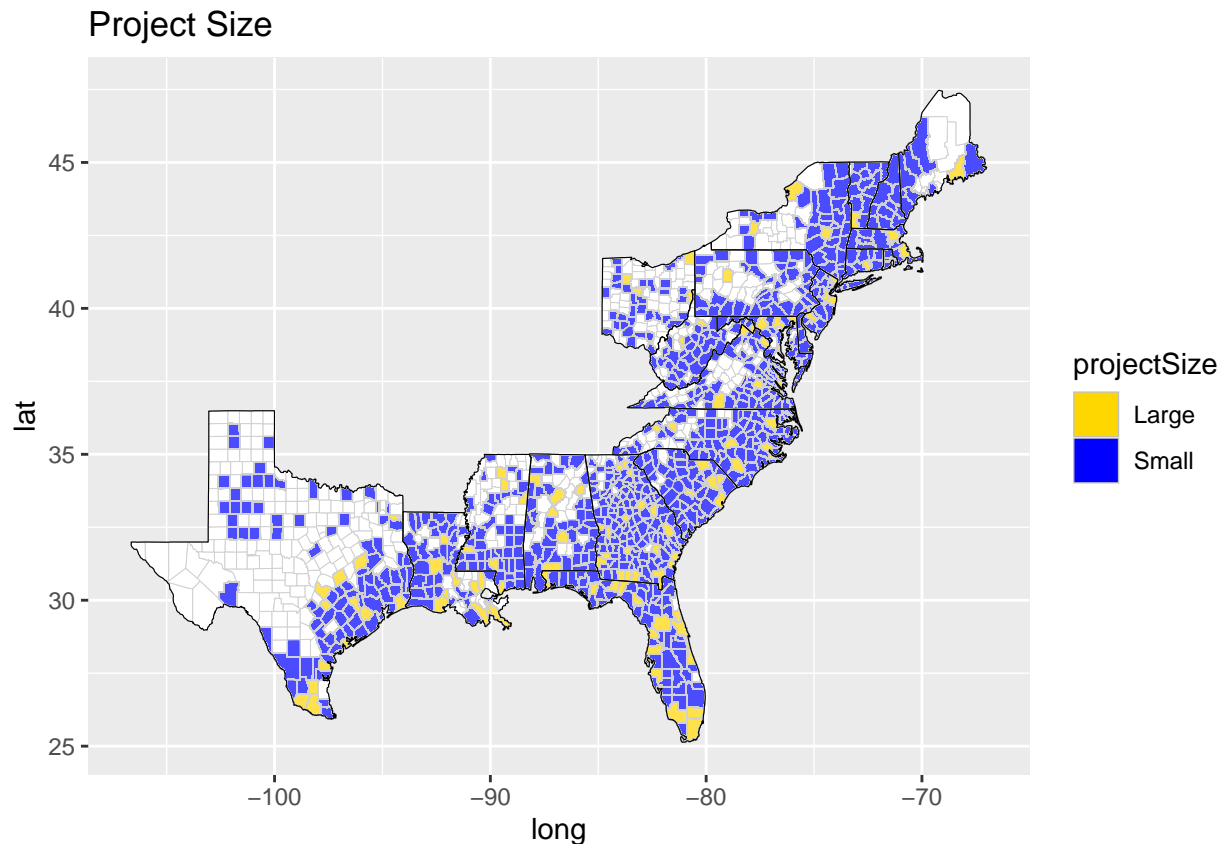

Then I map the project total amount grouped by county too see which exceed 100 million.

```
# use ggplot
ggplot() +
geom_polygon(data=AllCounty, aes(x=long, y=lat, group=group),
             color="gray", fill="white", size = .1 ) +
geom_polygon(data = cTable, aes(x = long, y = lat, group = group,
                                fill = `projectAmount > 100000000`),
             color = "grey", size = 0.2, alpha = 1.6) +
geom_polygon(data = MainStates, aes(x = long, y = lat, group = group),
             color="black", fill="white", size = 0.2, alpha = 0.3) +
scale_fill_manual(values = c("darkblue", "red"))+
ggtitle("County Based TOtal Project Amount") +

# Center the title
theme(plot.title = element_text(hjust = 0.5))
```

County Based TOtal Project Amount

projectAmount > 100000000

- False
- True