# Extended Mapping Assignment

Yifan Liu, Zhaosheng Xie

11/9/2020

## Importing Data

The dataset used in this project is from FEMA.

```
d1 = read.csv("PublicAssistanceFundedProjectsDetails.csv")
```

## Data Cleaning

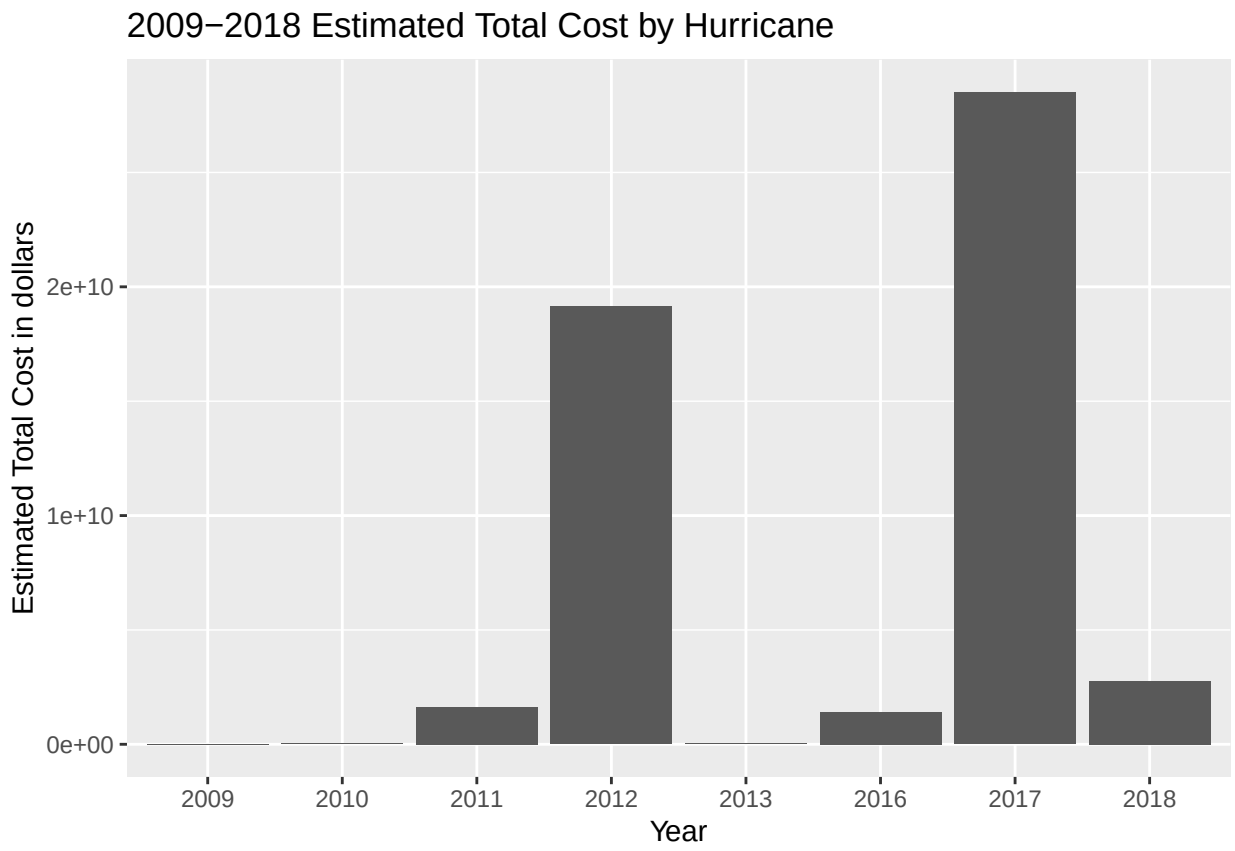
```
# filter the data from 2009 to 2018 and select related columns
d1 = separate(col = declarationDate, into = c("declarationYear","others"), sep = "-", remove = FALSE, da
d2 = select(d1, c(3,5,8,9,13,15,17,18,19,20))
d3 = d2 %>% filter(d2$declarationYear %in% c(2009:2018))
d4 = filter(d3, incidentType=="Hurricane")

#write.table(d4,"D:/MSSP/Rdata/615/615 Midterm project/d4.csv",col.names = TRUE,row.name = FALSE,sep =
```

```
#checking for NA values
length(which(!is.na(d4)=="FALSE"))
```

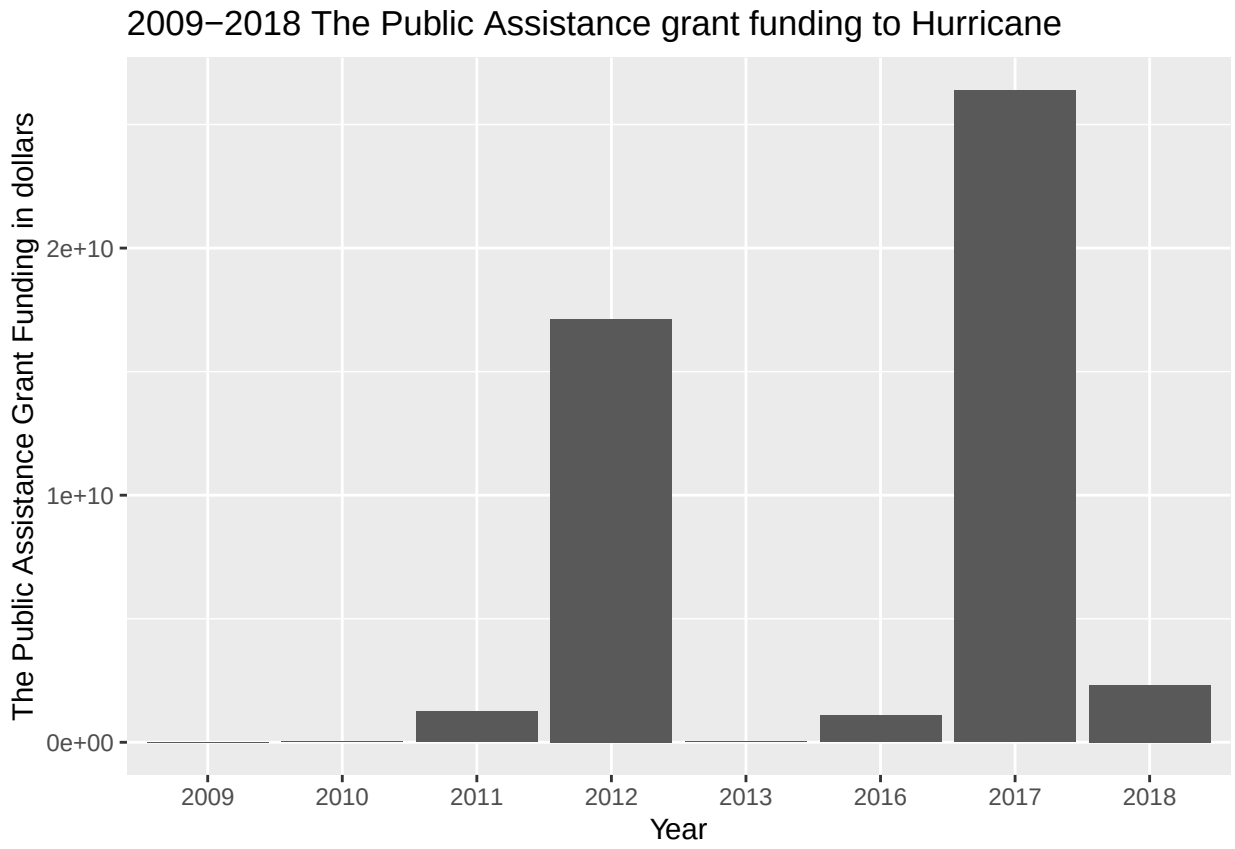
```
## [1] 0
```

## EDA

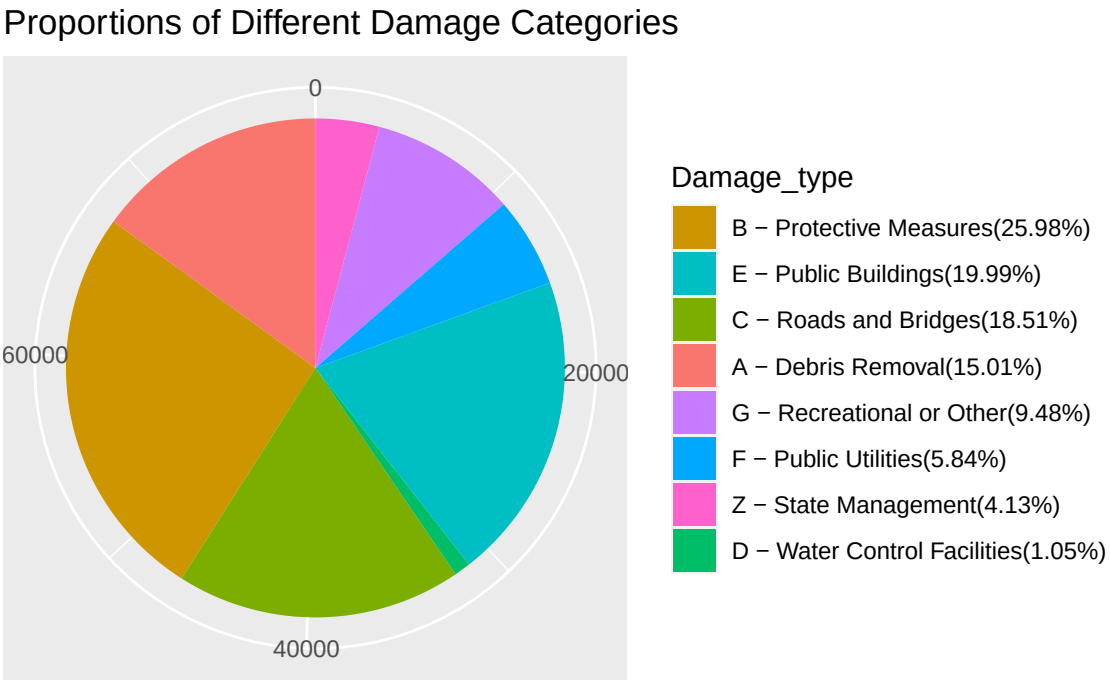
```
# estimate total cost due to hurricane
p1 = aggregate(projectAmount~declarationYear, data = d4, FUN = sum)
p1$declarationYear = as.numeric(p1$declarationYear)

ggplot(data=p1, aes(x = factor(declarationYear), y = projectAmount)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "Estimated Total Cost in dollars", title = "2009-2018 Estimated Total Cost by Hu
```




```
# The Public Assistance grant funding
p2 = aggregate(federalShareObligated~declarationYear, data = d4, FUN = sum)
p2$declarationYear = as.numeric(p2$declarationYear)

ggplot(data=p2, aes(x = factor(declarationYear), y = federalShareObligated)) +
  geom_bar(stat = "identity") +
  labs(x = "Year", y = "The Public Assistance Grant Funding in dollars", title = "2009-2018 The Public
```



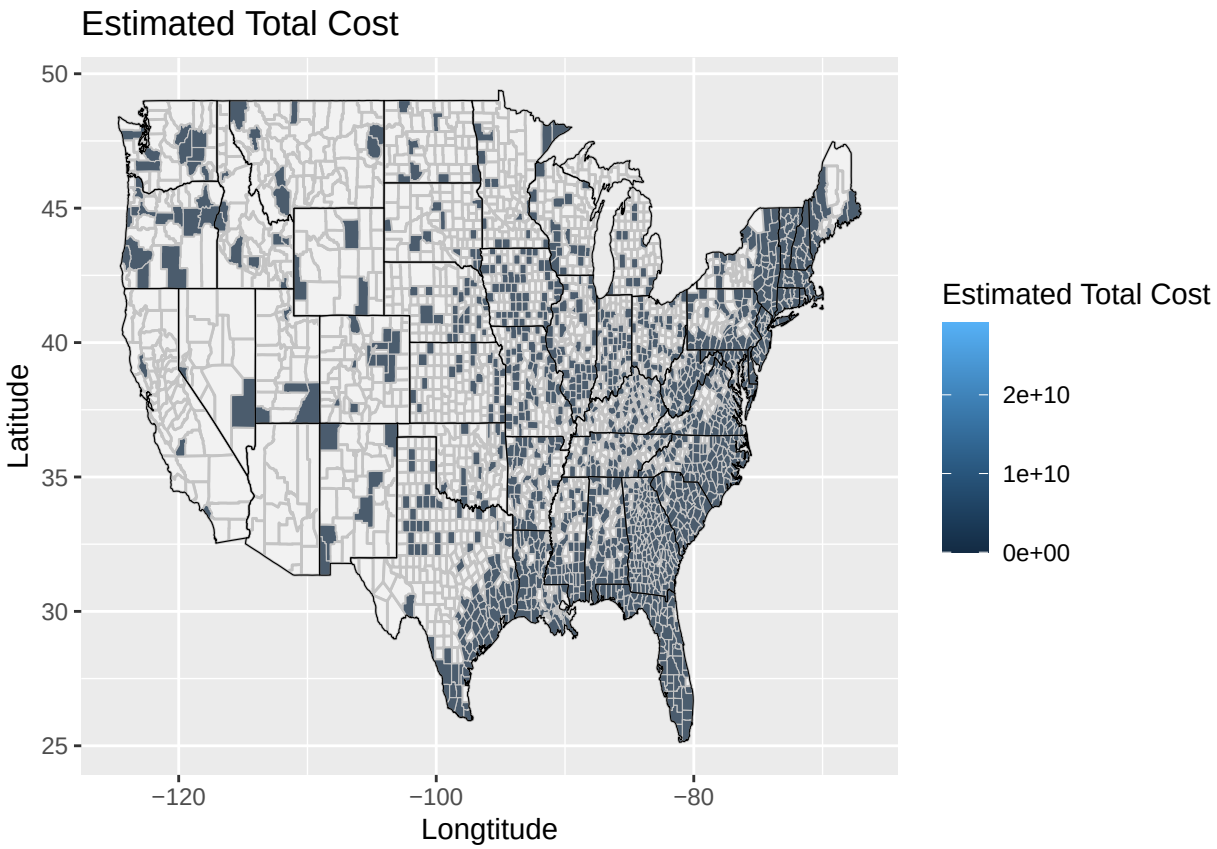

```
# Proportions of Different Damage Categories
t1 = data.frame(table(d4$damageCategoryCode))
t1 = rename(t1, "Damage_type" = "Var1")
t1 = t1[order(t1$Freq, decreasing = TRUE),]
label = as.vector(t1$Damage_type)
label = paste(label, "(", round(t1$Freq / sum(t1$Freq) * 100, 2), "%)        ", sep = "")

ggplot(t1, aes(x = "", y = Freq, fill = Damage_type)) +
  geom_bar(width = 1, stat = "identity") +
  coord_polar("y", start = 0) +
  theme(axis.ticks = element_blank()) +
  labs(x = "", y = "", title = "Proportions of Different Damage Categories") +
  scale_fill_discrete(breaks = t1$Damage_type, labels = label)
```




## Mapping

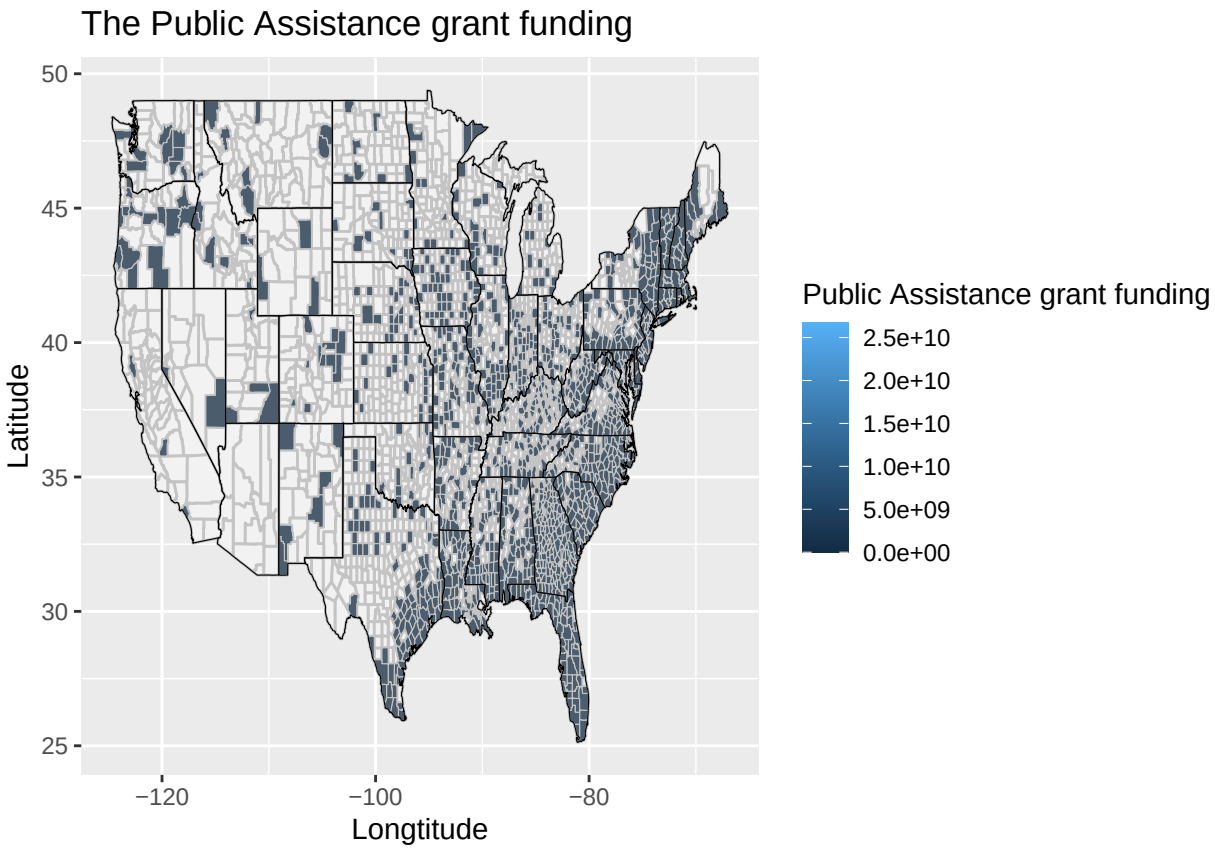
```
#Estimate Total Cost
a1 = aggregate(projectAmount~county, data = d4, FUN = sum)
a2 = aggregate(applicantId~county, data = d4, FUN = unique)
a2$number = lengths(a2$applicantId)
m1 = merge(a1, a2)
for (i in 1:nrow(m1)){
  m1$subregion[i] = tolower(m1$county[i])
}
m1 = select(m1, -c(1,3))
geo = map_data("county")
state = map_data("state")
geo1 = right_join(geo, m1, by = c('subregion'='subregion'))
```

```
ggplot() +
  ggtitle("Estimated Total Cost") +
  geom_polygon(data = geo, aes(x = long, y = lat, group = group), color = "grey", fill = "white") +
  geom_polygon(data = geo1, aes(x = long, y = lat, group = group, fill = projectAmount), color = "grey"
  labs(x = "Longtitude", y = "Latitude", fill = "Estimated Total Cost")+
  geom_polygon(data = state, aes(x = long, y = lat, group = group), color="black", fill = "lightgray",
```




```
# The Public Assistance grant funding
a3 = aggregate(federalShareObligated~county, data = d4, FUN = sum)
a2$number = lengths(a2$applicantId)
m2 = merge(a3, a2)
for (i in 1:nrow(m2)){
  m2$subregion[i] = tolower(m2$county[i])
}
m2 = select(m2, -c(1,3))
geo = map_data("county")
state = map_data("state")
geo2 = right_join(geo, m2, by = c('subregion'='subregion'))
```

```
ggplot() +
  ggtitle("The Public Assistance grant funding") +
  geom_polygon(data = geo, aes(x = long, y = lat, group = group), color = "grey", fill = "white") +
  geom_polygon(data = geo2, aes(x = long, y = lat, group = group, fill = federalShareObligated), color =
  labs(x = "Longtitude", y = "Latitude", fill = "Public Assistance grant funding")+
  geom_polygon(data = state, aes(x = long, y = lat, group = group), color="black", fill = "lightgray",
```


The Public Assistance grant funding
50
45
40
35
30
25
Latitude
−120
−100
−80
Longtitude
Public Assistance grant funding
2.5e+10
2.0e+10
1.5e+10
1.0e+10
5.0e+09
0.0e+00