Extended Mapping Assignment

Yuxin Zeng & Haoqi Wang

2020/11/12

Contents

Introduction

In this assignment, we use the data of public assistance funded projects information from FEMA to produce maps with more features. The purpose of this assignment is to make sure that we know the basics of using R to produce documents, presentations, and shiny apps.

Data Collection

1.Read the data

The source of data is from https://www.fema.gov/openfema-data-page/public-assistance-funded-projects-details-v1

```
dt <- read.csv(file="PublicAssistanceFundedProjectsDetails.csv",header=T)
```

2.Data Cleaning

We not only choose the data from 2009 to 2018 and remove the missing data,but also filter the hurricane data and tornado data from the data we got.

```
#Filter the information from 2009 to 2018 and remove meaningless columns
dt %>%
  filter(2009<=year(declarationDate)&year(declarationDate)<=2018) %>%
  select(c(1,2,3,5,6,9,11,13,16)) -> dt1

#Focus on hurricane and tornado
unique(dt1$incidentType)
```

```
##  [1] "Severe Ice Storm" "Flood"            "Severe Storm(s)"  "Tornado"
##  [5] "Earthquake"       "Hurricane"        "Snow"             "Tsunami"
##  [9] "Fire"             "Other"            "Terrorist"        "Chemical"
## [13] "Coastal Storm"    "Mud/Landslide"    "Freezing"         "Volcano"
## [17] "Typhoon"
```

```
hurricane <- subset(dt1,incidentType=="Hurricane")
tornado <- subset(dt1,incidentType=="Tornado")

#Check for missing values
length(which(!is.na(hurricane)=="FALSE"))
```

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

```
length(which(!is.na(tornado)=="FALSE"))
```

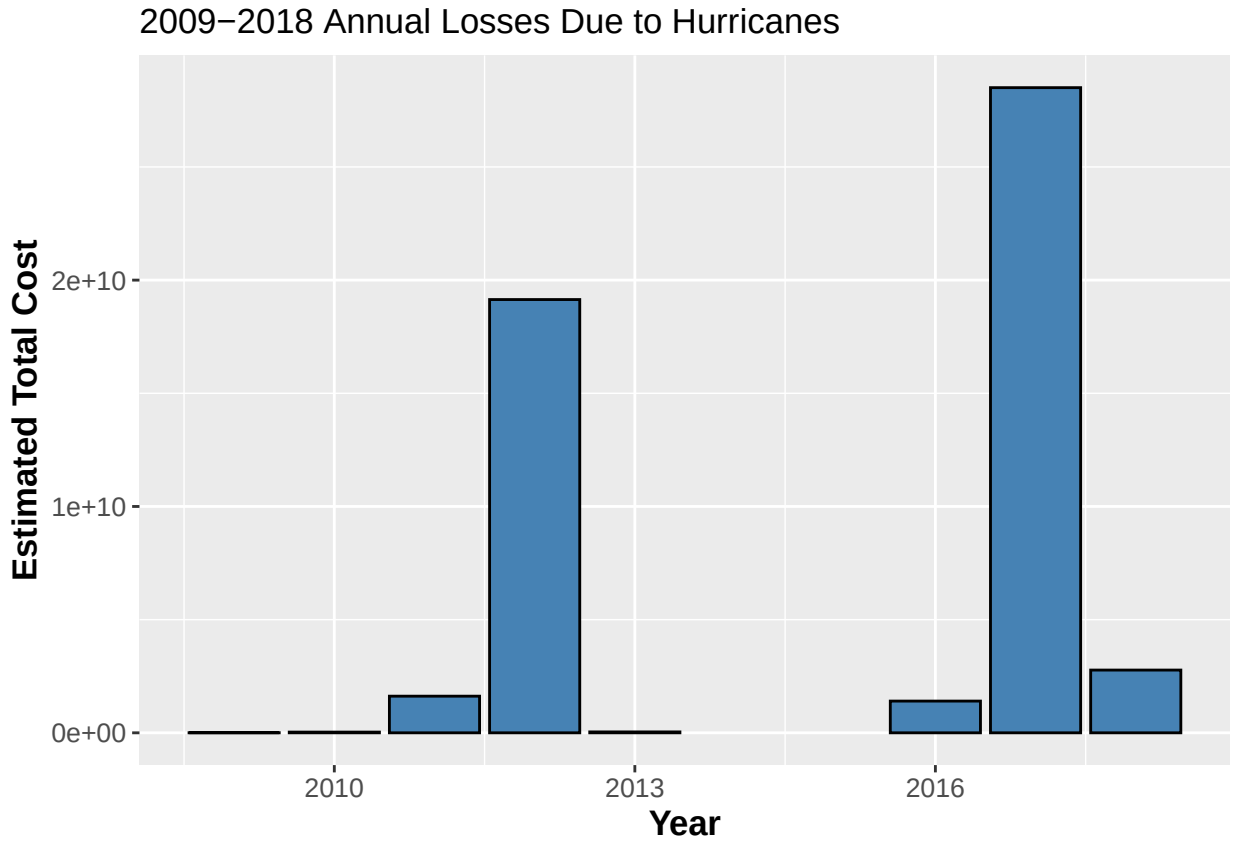
```
## [1] 0
```

EDA

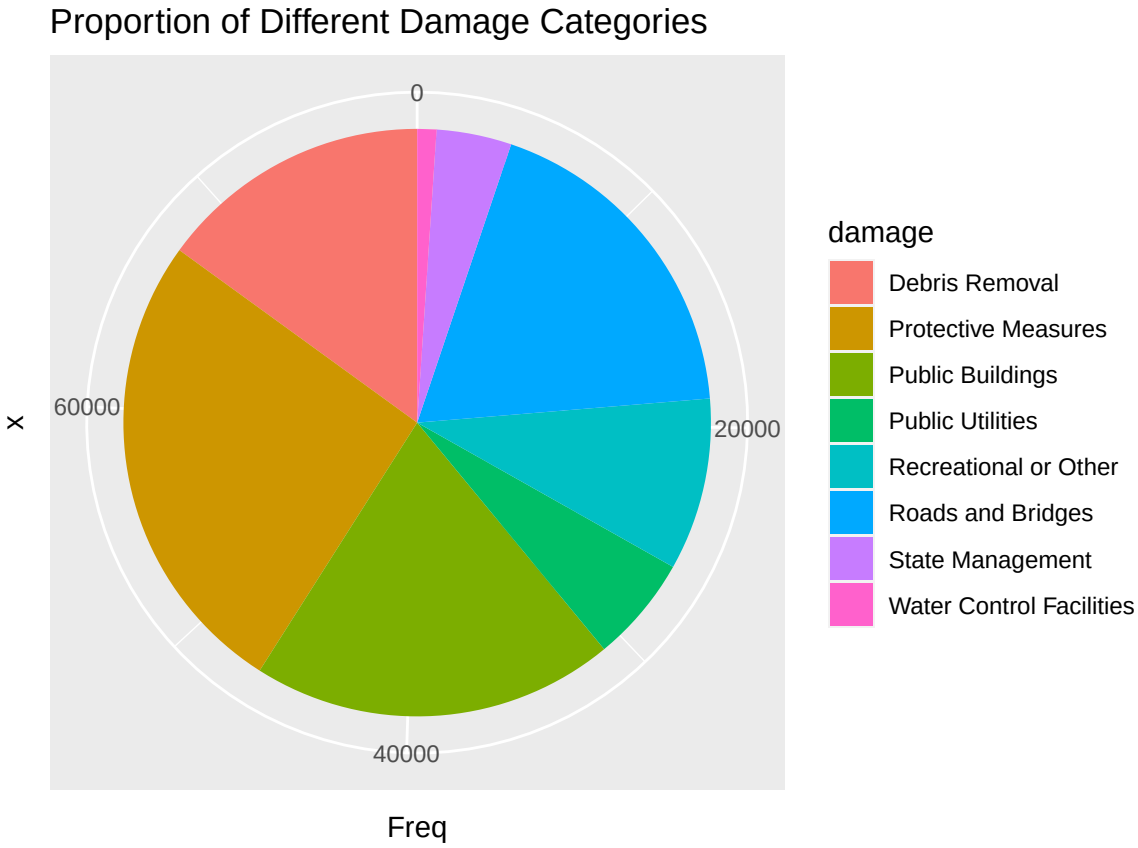
In this part, we focus on column `projectAmount` – the estimated total cost of the Public Assistance grant project in dollars, without administrative costs. This amount is based on the damage survey. We calculate the total project amount for each state, each year, different project sizes and different damage categories and display it using barplot.

1.EDA for Hurricane

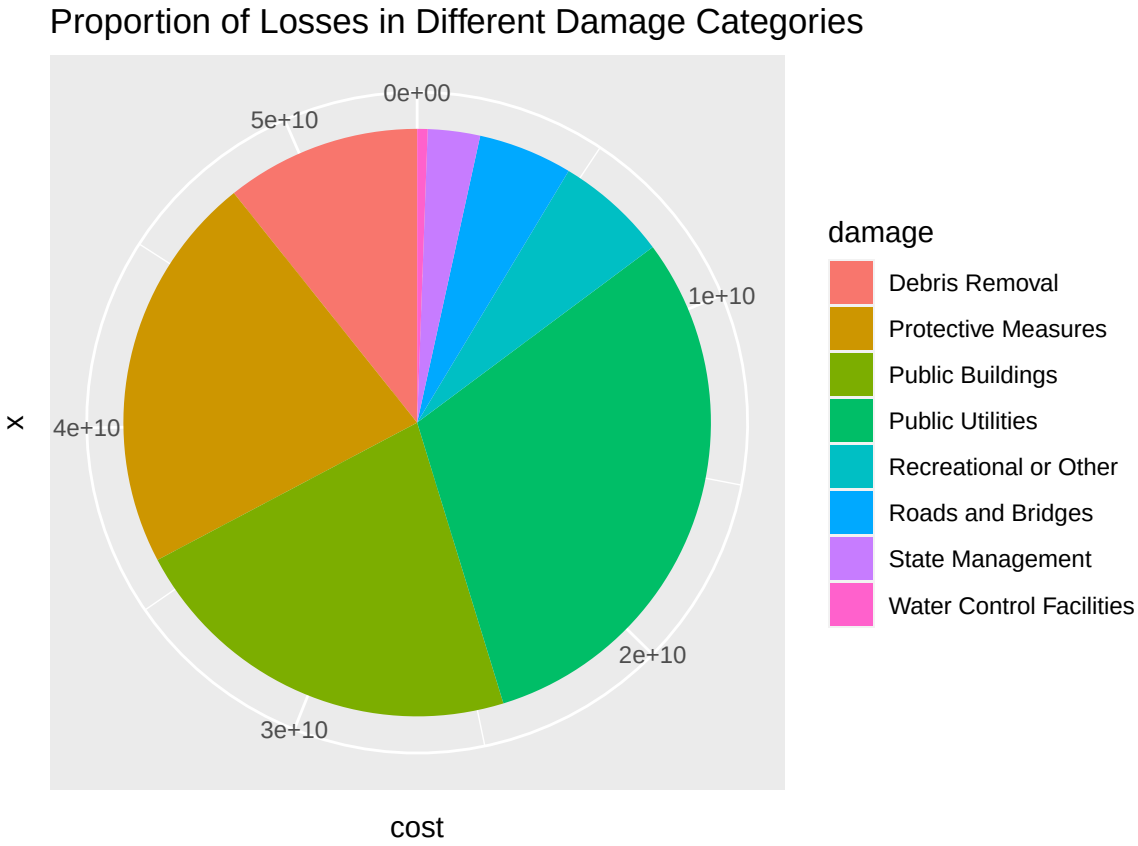
```
#2009-2018 Annual Losses Due to Hurricane
aggregate(projectAmount~year(declarationDate),data=hurricane,FUN=sum) %>%
  rename("year"="year(declarationDate)","cost"="projectAmount") ->
  a1

#Figure
ggplot(data=a1,aes(x=year,y=cost)) +
  geom_bar(stat="identity",fill="steelblue",color="black") +
  labs(x="Year",y="Estimated Total Cost",title="2009-2018 Annual Losses Due to Hurricanes") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 13, face = "bold"))   ->
  h_bar
h_bar
```




```
#Frequency of each damage type
t1=table(hurricane$damageCategory)

#Cost of each damage type
a2=aggregate(projectAmount~damageCategory,data=hurricane,FUN=sum)

#Pie chart data
cbind(a2,t1) %>%
  rename("damage"="damageCategory","cost"="projectAmount") %>%
  select(-3) ->
  h_piedata

#Proportion of Different Damage Categories
ggplot(data=h_piedata, mapping=aes(x="",y=Freq,fill=damage)) +
  geom_bar(stat="identity",width=1) +
  coord_polar(theta="y") +
  labs(title="Proportion of Different Damage Categories") +
  theme(axis.ticks=element_blank()) -> h_pie1
h_pie1
```

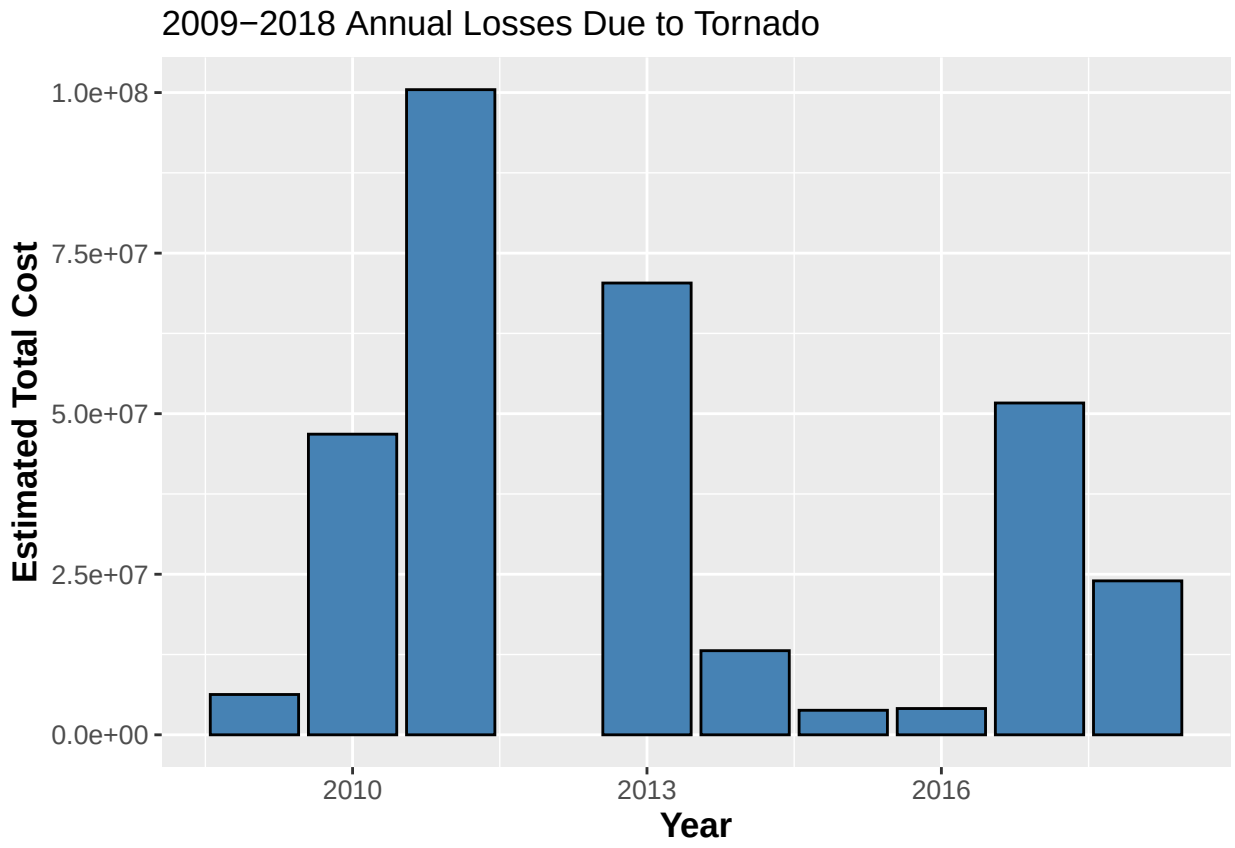



```
#Proportion of Losses in Different Damage Categories
ggplot(data=h_piedata, mapping=aes(x="",y=cost,fill=damage)) +
  geom_bar(stat="identity",width=1) +
  coord_polar(theta="y") +
  labs(title="Proportion of Losses in Different Damage Categories") +
  theme(axis.ticks=element_blank()) -> h_pie2
h_pie2
```

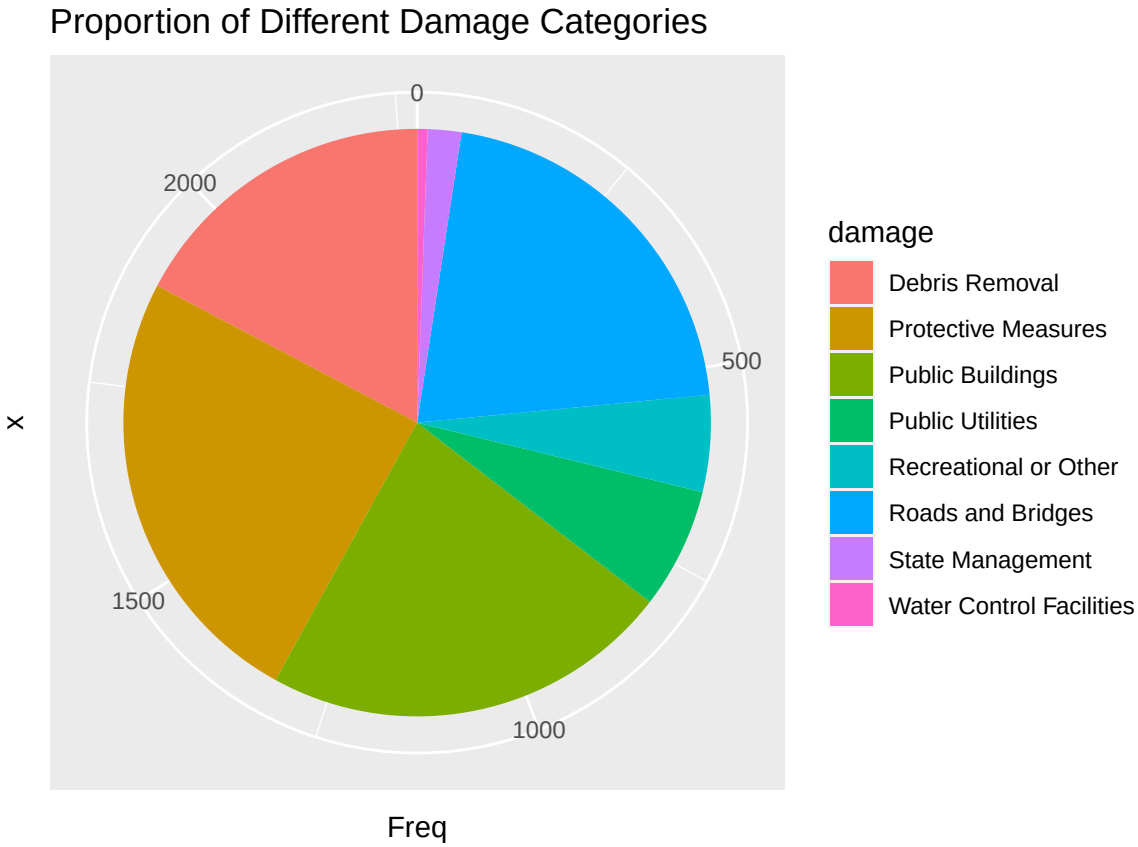


2.EDA for Tornado

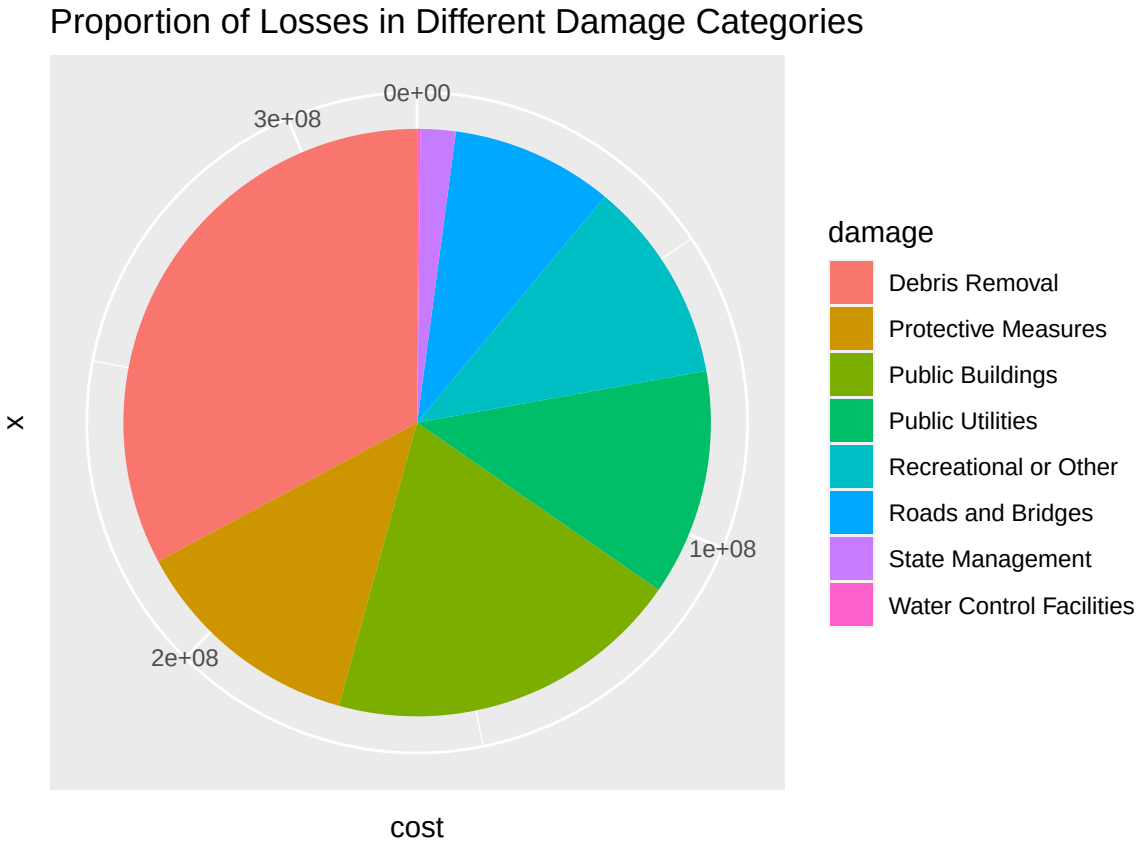
```
#2009-2018 Annual Losses Due to Tornado
aggregate(projectAmount~year(declarationDate),data=tornado,FUN=sum) %>%
  rename("year"="year(declarationDate)","cost"="projectAmount") ->
  a3

#Figure
ggplot(data=a3,aes(x=year,y=cost)) +
  geom_bar(stat="identity",fill="steelblue",color="black") +
  labs(title="2009-2018 Annual Losses Due to Tornado") +
  xlab("Year") +
  ylab("Estimated Total Cost") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 13, face = "bold"))   ->
  t_bar
t_bar
```




```
#Frequency of each damage type
t2=table(tornado$damageCategory)

#Cost of each damage type
a4=aggregate(projectAmount~damageCategory,data=tornado,FUN=sum)

#Pie chart data
cbind(a4,t2) %>%
  rename("damage"="damageCategory","cost"="projectAmount") %>%
  select(-3) ->
  t_piedata

#Proportion of Different Damage Categories
ggplot(data=t_piedata, mapping=aes(x="",y=Freq,fill=damage)) +
  geom_bar(stat="identity",width=1) +
  coord_polar(theta="y") +
  labs(title="Proportion of Different Damage Categories") +
  theme(axis.ticks=element_blank()) -> t_pie1
t_pie1
```




```
#Proportion of Losses in Different Damage Categories
ggplot(data=t_piedata, mapping=aes(x="",y=cost,fill=damage)) +
  geom_bar(stat="identity",width=1) +
  coord_polar(theta="y") +
  labs(title="Proportion of Losses in Different Damage Categories") +
  theme(axis.ticks=element_blank()) -> t_pie2
t_pie2
```



We can see that there is no estimated cost in 2014 and 2015 for Hurricane, but the estimated cost in 2011 was drastically large, over 20 billion. And he estimated cost in 2011 was the largest (100 million) for Tornado, and there is no estimated cost in 2012.

Mapping

In this part, we draw different maps to show the total cost and average cost in counties, which were caused by hurricane and tornado.

1.Hurricane Mapping

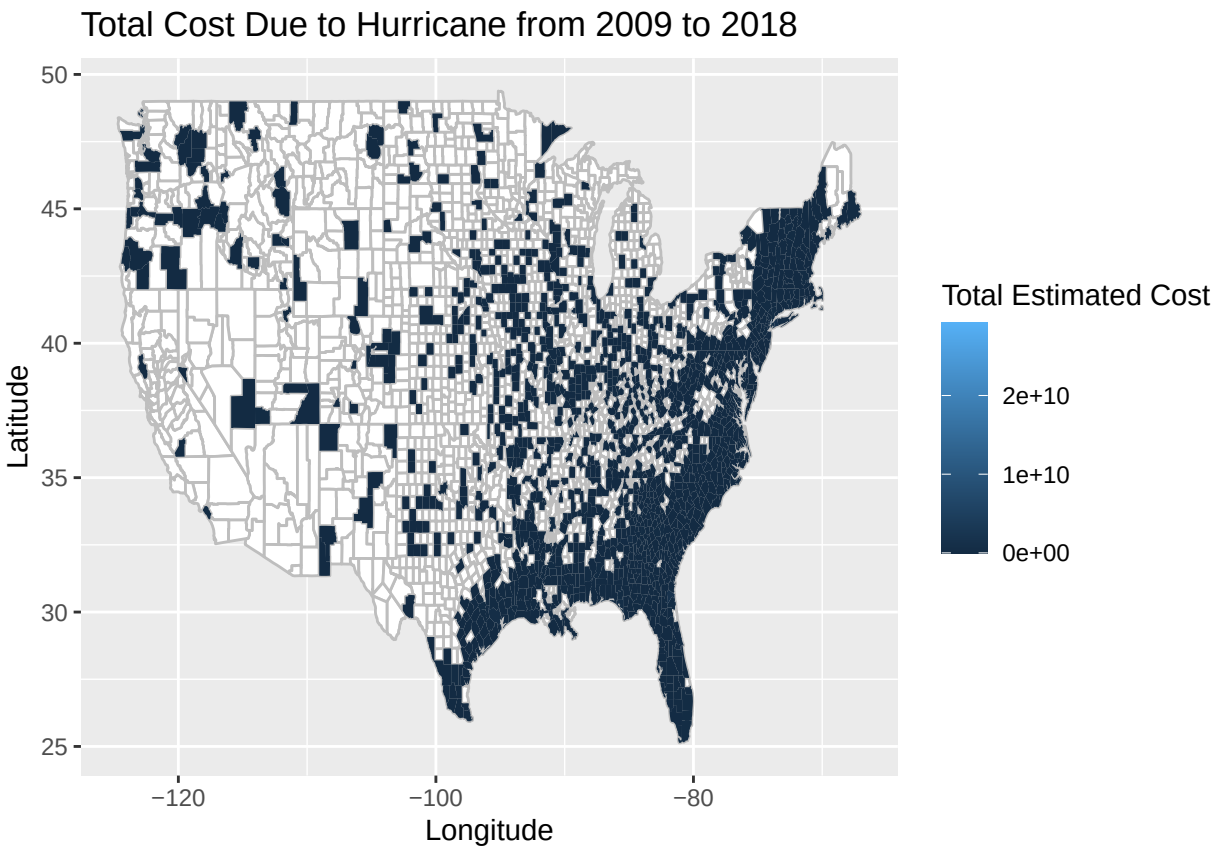
Firstly, we calculate the average cost per applicant of each country and reorganise the data so that we can use them to draw maps.

```
#Total cost of each county
a5=aggregate(projectAmount~county,data=hurricane,FUN=sum)
#Unique applicant ID of each county
a6=aggregate(applicantId~county,data=hurricane,FUN=unique)
#Number of unique applicants of each county
a6$number=lengths(a6$applicantId)
#Join the tables (a5 & a6)
h_data=merge(a5,a6)
#Average cost per applicant of each county
h_data$average=h_data$projectAmount/h_data$number

#Organize the data
for (i in 1:nrow(h_data)){
```

```
  tolower(h_data$county[i]) -> h_data$subregion[i]
}

h_data %>%
  select(-c(1,3)) %>%
  rename("total"="projectAmount") ->
  h_data

#Add long&lat information
map_data("county") -> geo
geo %>% right_join(h_data,by=c('subregion'='subregion')) ->
  h_mapdata
```

This is the map that shows the total cost from 2009 to 2018, which is caused by Hurricane.

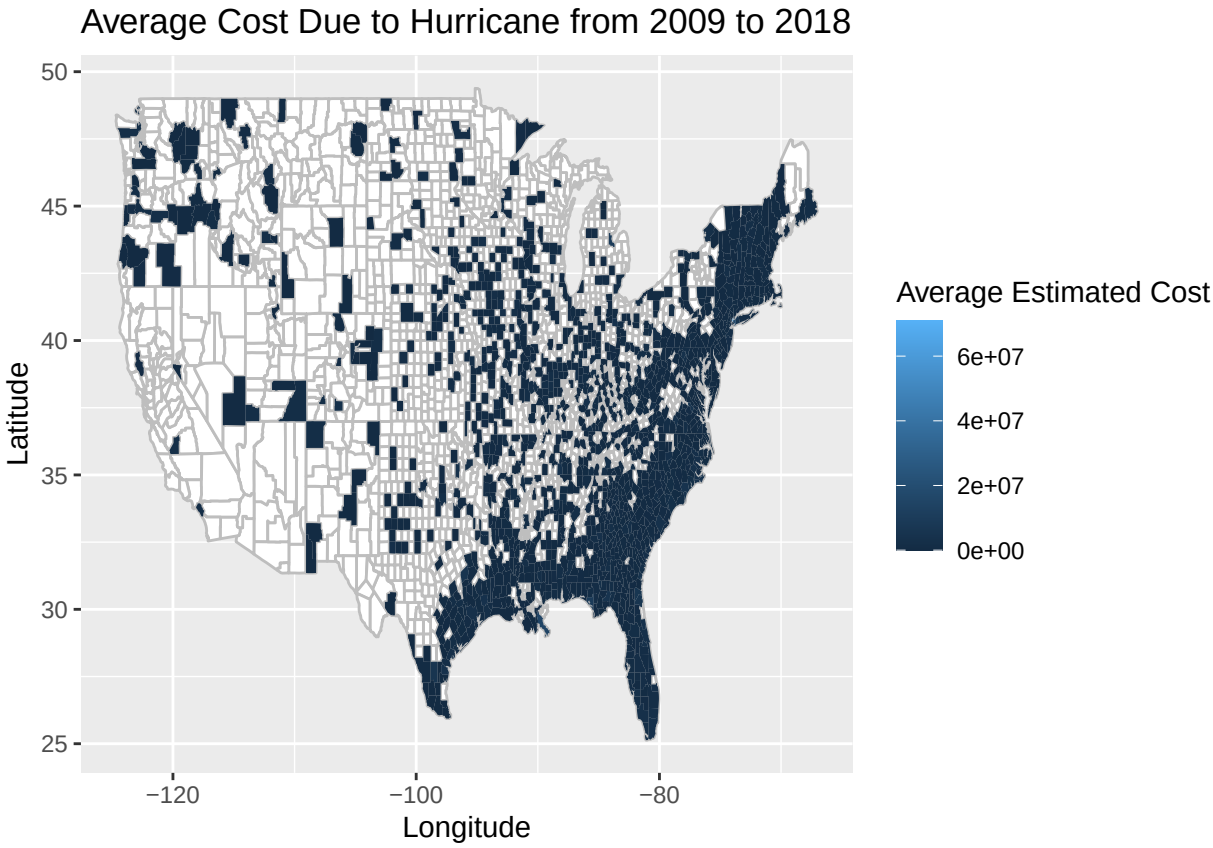
```
#Total Cost Due to Hurricane from 2009 to 2018
ggplot() +
  geom_polygon(data=geo,aes(x=long,y=lat,group=group),colour="grey",fill="white") +
  geom_polygon(data=h_mapdata,aes(x=long,y=lat,group=group,fill=total),colour="transparent") +
  labs(title="Total Cost Due to Hurricane from 2009 to 2018 ",x="Longitude",y="Latitude",fill="Total Es
  h_map1
h_map1
```



This is the map that shows the average cost from 2009 to 2018, which is caused by Hurricane.

```
#Average Cost Due to Hurricane from 2009 to 2018
ggplot() +
  geom_polygon(data=geo,aes(x=long,y=lat,group=group),colour="grey",fill="white") +
  geom_polygon(data=h_mapdata,aes(x=long,y=lat,group=group,fill=average),colour="transparent") +
```

```
  labs(title="Average Cost Due to Hurricane from 2009 to 2018 ",x="Longitude",y="Latitude",fill="Averag
  h_map2
h_map2
```



2.Tornado Mapping

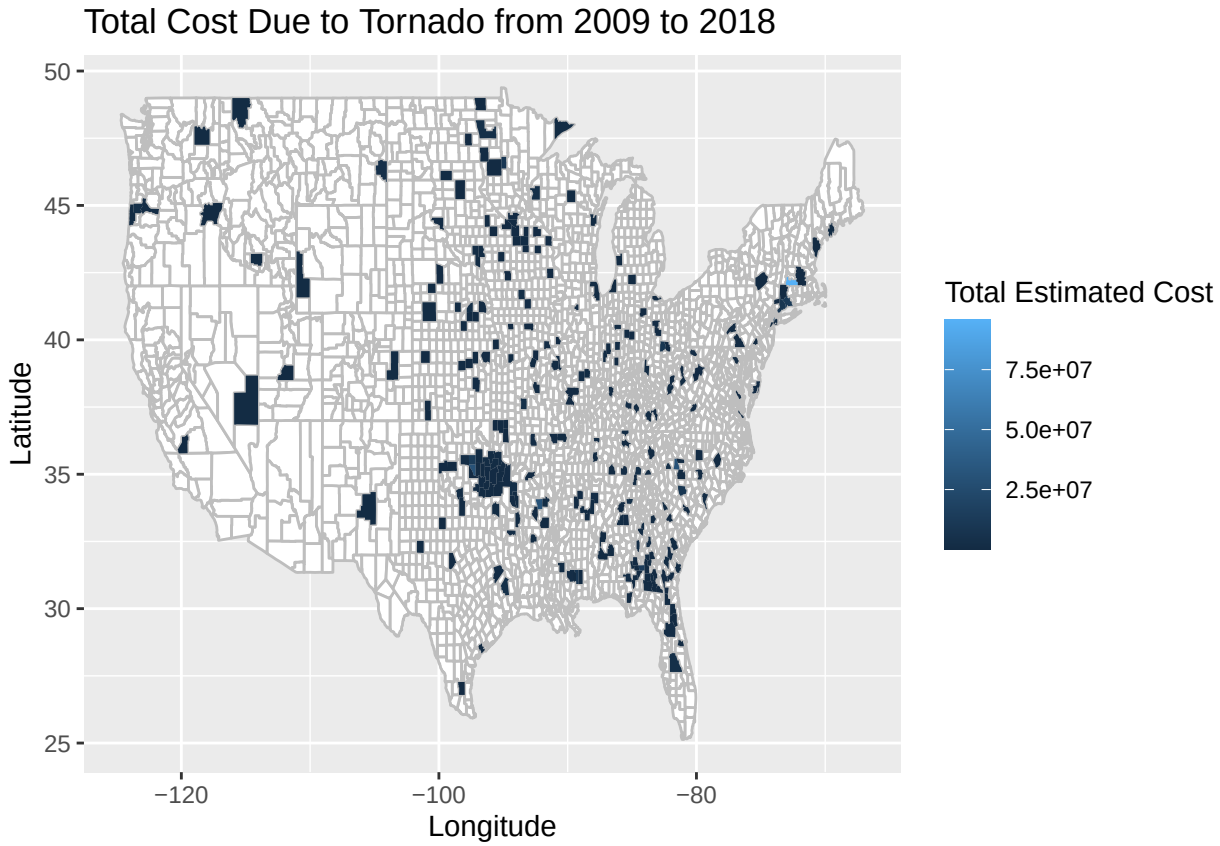
Firstly, we calculate the average cost per applicant of each country and reorganise the data so that we can use them to draw maps.

```
#Total cost of each county
a7=aggregate(projectAmount~county,data=tornado,FUN=sum)
#Unique applicant ID of each county
a8=aggregate(applicantId~county,data=tornado,FUN=unique)
#Number of unique applicants of each county
a8$number=lengths(a8$applicantId)
#Join the tables (a7 & a8)
t_data=merge(a7,a8)
#Average cost per applicant of each county
t_data$average=t_data$projectAmount/t_data$number

#Organize the data
for (i in 1:nrow(t_data)){
  tolower(t_data$county[i]) -> t_data$subregion[i]
}

t_data %>%
  select(-c(1,3)) %>%
```

```
  rename("total"="projectAmount") ->
  t_data

#Add long&lat information
map_data("county") -> geo
geo %>% right_join(t_data,by=c('subregion'='subregion')) ->
  t_mapdata
```

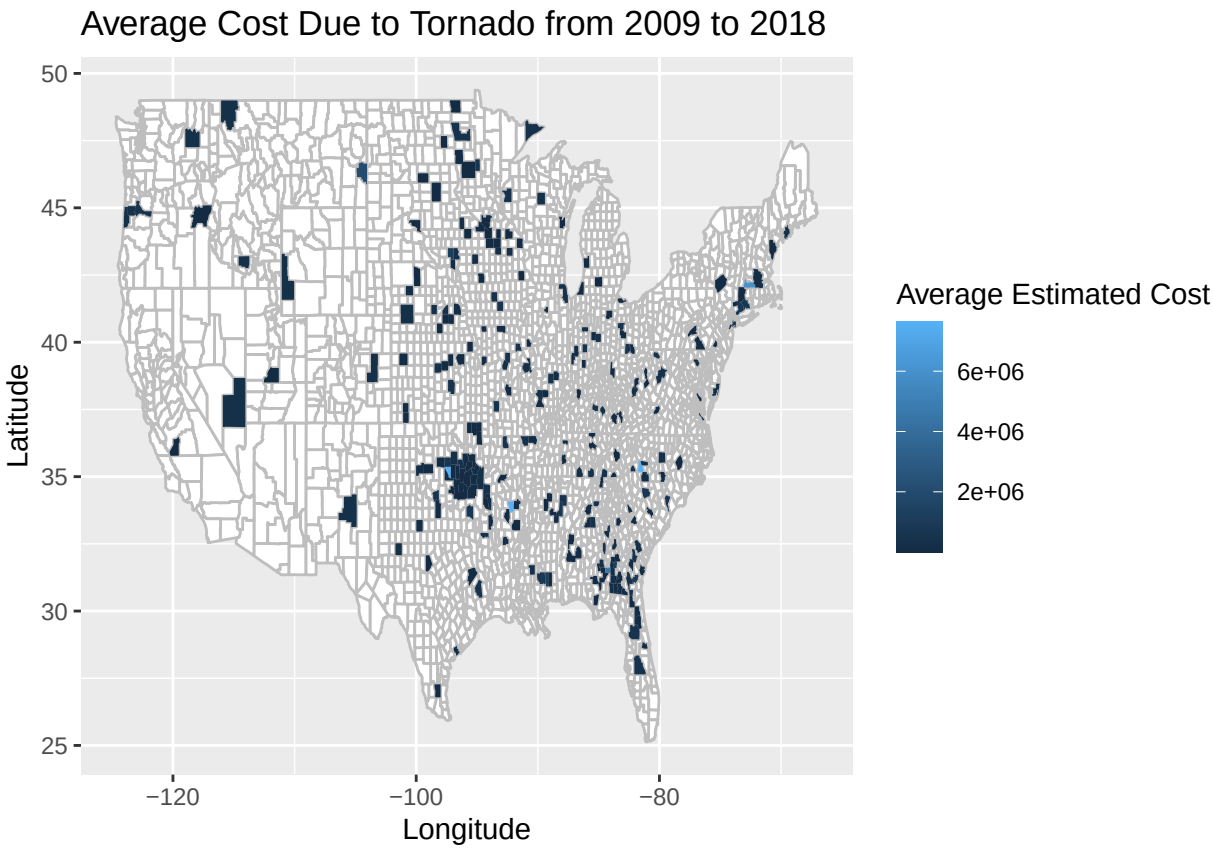
This is the map that shows the total cost from 2009 to 2018, which is caused by Tornado.

```
#Total Cost Due to Tornado from 2009 to 2018
ggplot() +
  geom_polygon(data=geo,aes(x=long,y=lat,group=group),colour="grey",fill="white") +
  geom_polygon(data=t_mapdata,aes(x=long,y=lat,group=group,fill=total),colour="transparent") +
  labs(title="Total Cost Due to Tornado from 2009 to 2018 ",x="Longitude",y="Latitude",fill="Total Esti
  t_map1
t_map1
```



This is the map that shows the average cost from 2009 to 2018, which is caused by Tornado.

```
#Average Cost Due to Tornado from 2009 to 2018
ggplot() +
  geom_polygon(data=geo,aes(x=long,y=lat,group=group),colour="grey",fill="white") +
  geom_polygon(data=t_mapdata,aes(x=long,y=lat,group=group,fill=average),colour="transparent") +
  labs(title="Average Cost Due to Tornado from 2009 to 2018 ",x="Longitude",y="Latitude",fill="Average
  t_map2
t_map2
```


Average Cost Due to Tornado from 2009 to 2018
50
45
40
35
30
25
Latitude
−120
−100
−80
Longitude
Average Estimated Cost
6e+06
4e+06
2e+06