Extended Mapping Assignment

Yuxin Zeng

2020/11/9

目录

Introduction

Various natural disasters bring different types and degrees of economic losses to people every year. Using FEMA data, we took Hurricane and Tornado from 2009 to 2018 as the research objects, conducted a simple EDA analysis, and drew some loss maps.

Our data source is website: https://www.fema.gov/openfema-data-page/public-assistance-funded-projects-details-v1.

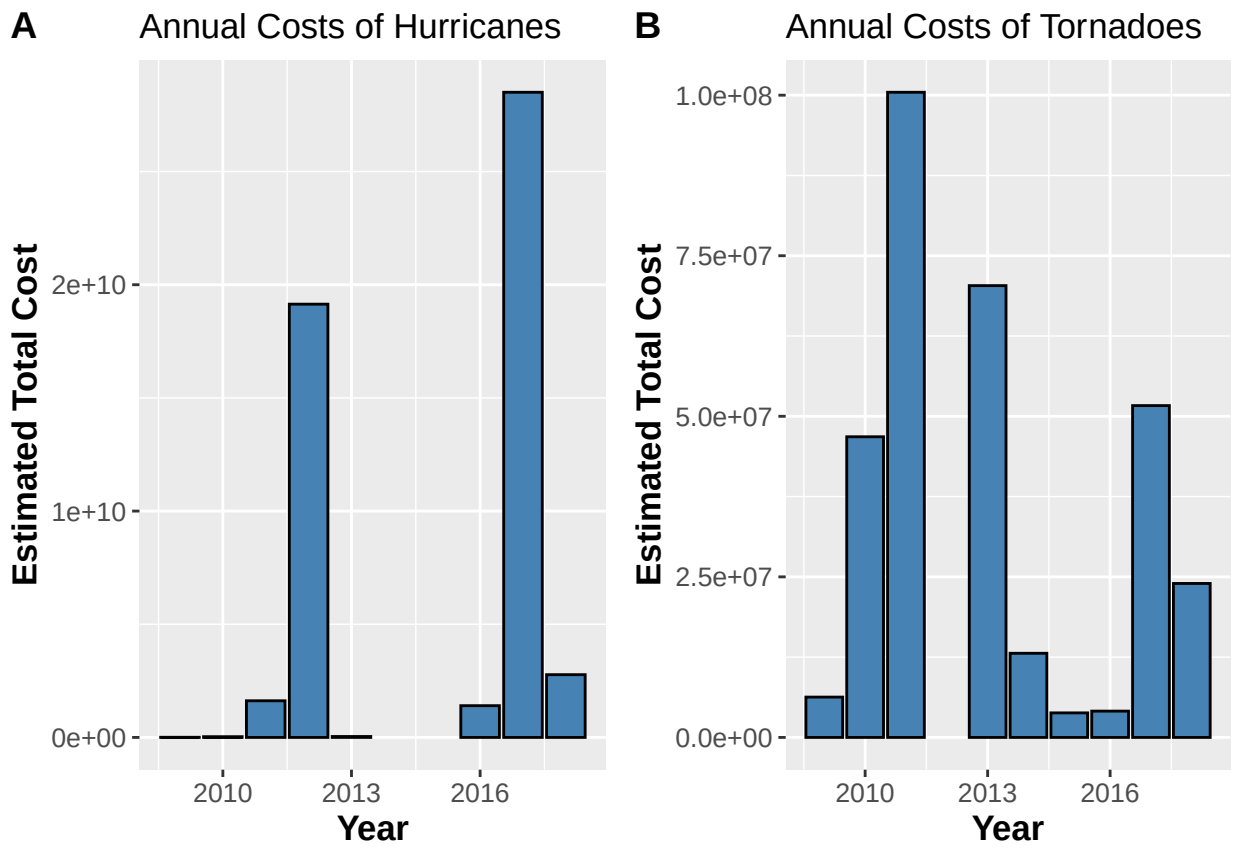
Hurricane v.s. Tornado

Hurricanes and tornadoes appear to be similar in their general structure. Both are characterized by extremely strong horizontal winds swirling around the center, strong upward motion dominating the circulation with some downward motion in the center.

EDA: 2009-2018 Annual Costs

```
#2009-2018 Annual Costs of Hurricanes
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="Annual Costs of Hurricanes") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 13, face = "bold"))   ->
  h_bar

#2009-2018 Annual Costs of Tornadoes
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="Annual Costs of Tornadoes") +
  xlab("Year") +
  ylab("Estimated Total Cost") +
  theme(axis.text = element_text(size = 10),
        axis.title = element_text(size = 13, face = "bold"))    ->
  t_bar

cowplot::plot_grid(h_bar,t_bar,labels="AUTO")
```

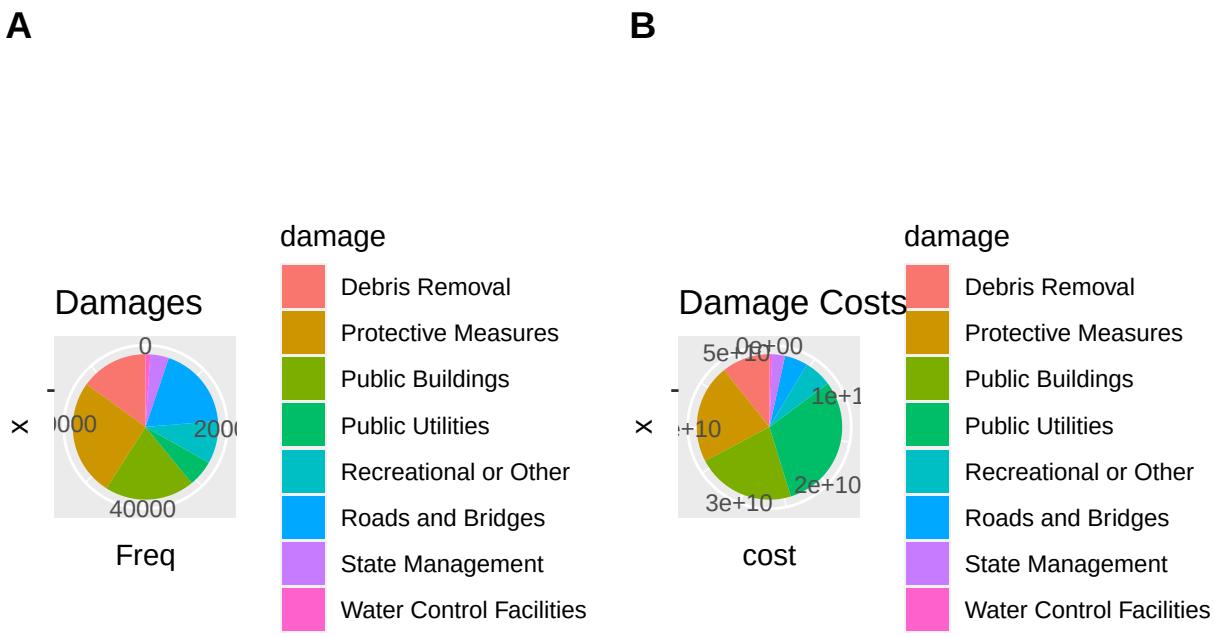


Hurricane: There is no estimated cost in 2014 and 2015, but the estimated cost in 2011 was drastically large, over 20 billion. Tornado: The estimated cost in 2011 was the largest (100 million), and there is no estimated cost in 2012.

EDA: Different Damage Categories of Hurricanes

```
#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="Damages") -> 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="Damage Costs") -> h_pie2

cowplot::plot_grid(h_pie1,h_pie2,labels ="AUTO")
```

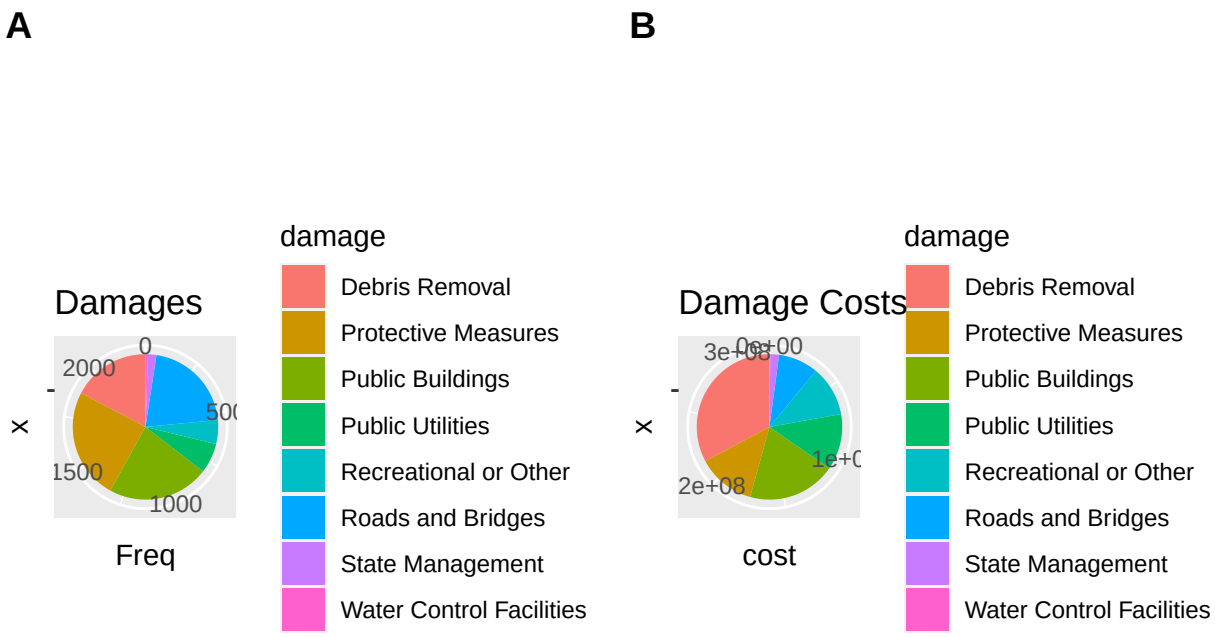


Protective Measures accounts for the largest proportion among all the damage categories caused by hurricanes, but its cost is not the largest. The cost of Public Utilities is the largest.

EDA: Different Damage Categories of Tornadoes

```
#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="Damages") -> 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="Damage Costs") -> t_pie2

cowplot::plot_grid(t_pie1,t_pie2,labels="AUTO")
```

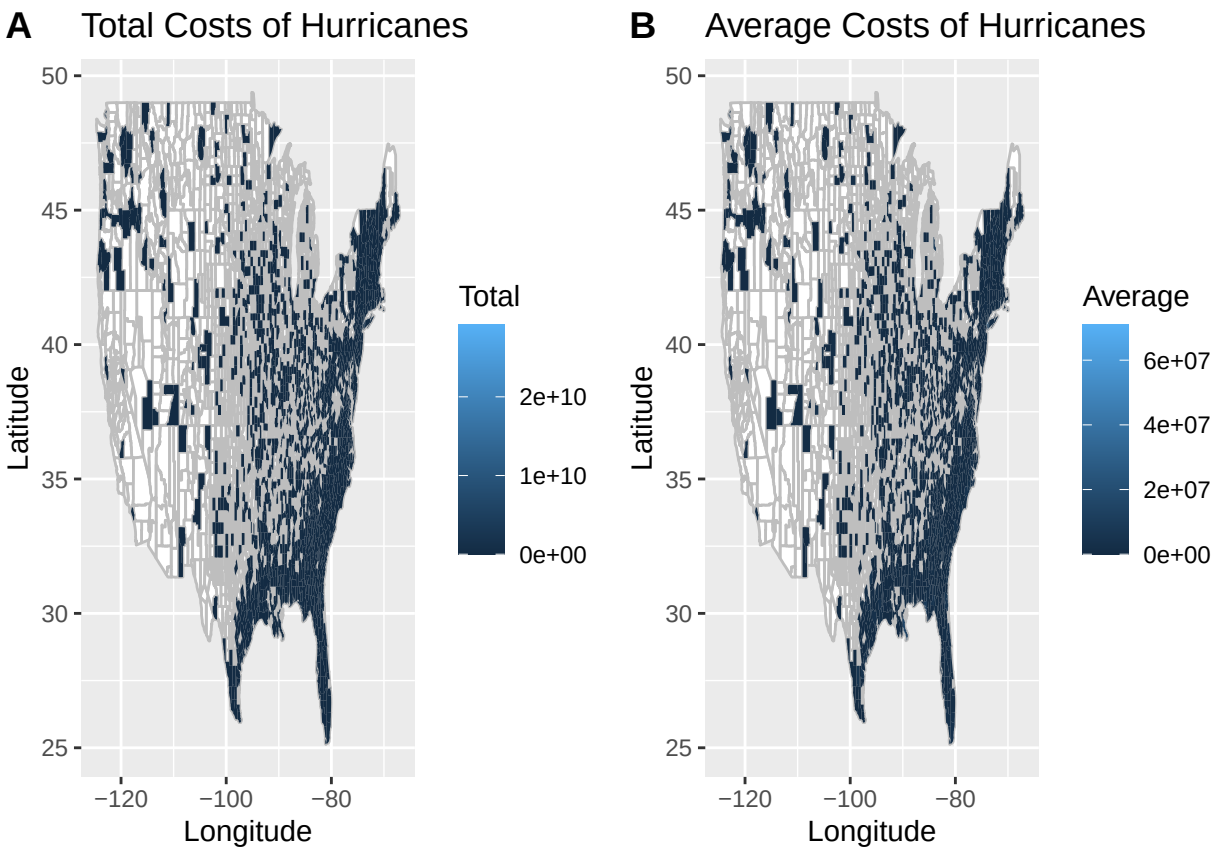


Debris Removal is not the most frequent appliaction among all the damage categories caused by tornadoes, but its cost is the largest.

Hurricane Maps

```
#Total cost of each county
a5=aggregate(projectAmount~county,data=hurricane,FUN=sum)
#Unique applicant ID of each county
```

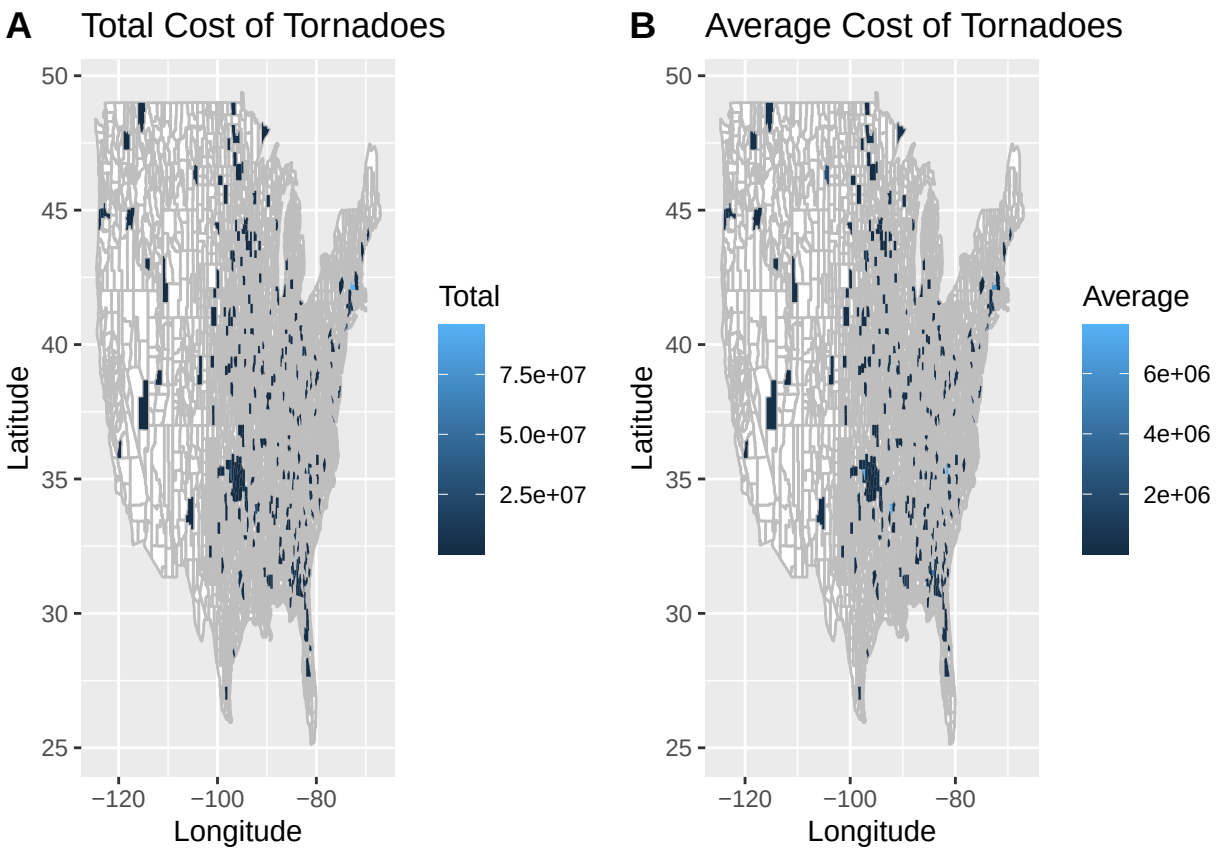
```
a6=aggregate(applicantId~county,data=hurricane,FUN=unique)
#Number of unique applicants of each county
a6$number=lengths(a6$applicantId)
#Join the tables (a5 & a6)
h_data=merge(a5,a6)
#Average cost per applicant of each county
h_data$average=h_data$projectAmount/h_data$number

#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

#Total Costs of Hurricanes 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 Costs of Hurricanes",x="Longitude",y="Latitude",fill="Total") ->
  h_map1

#Average Costs of Hurricanes 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 Costs of Hurricanes",x="Longitude",y="Latitude",fill="Average") ->
  h_map2

cowplot::plot_grid(h_map1,h_map2,labels="AUTO")
```



Tornado Maps

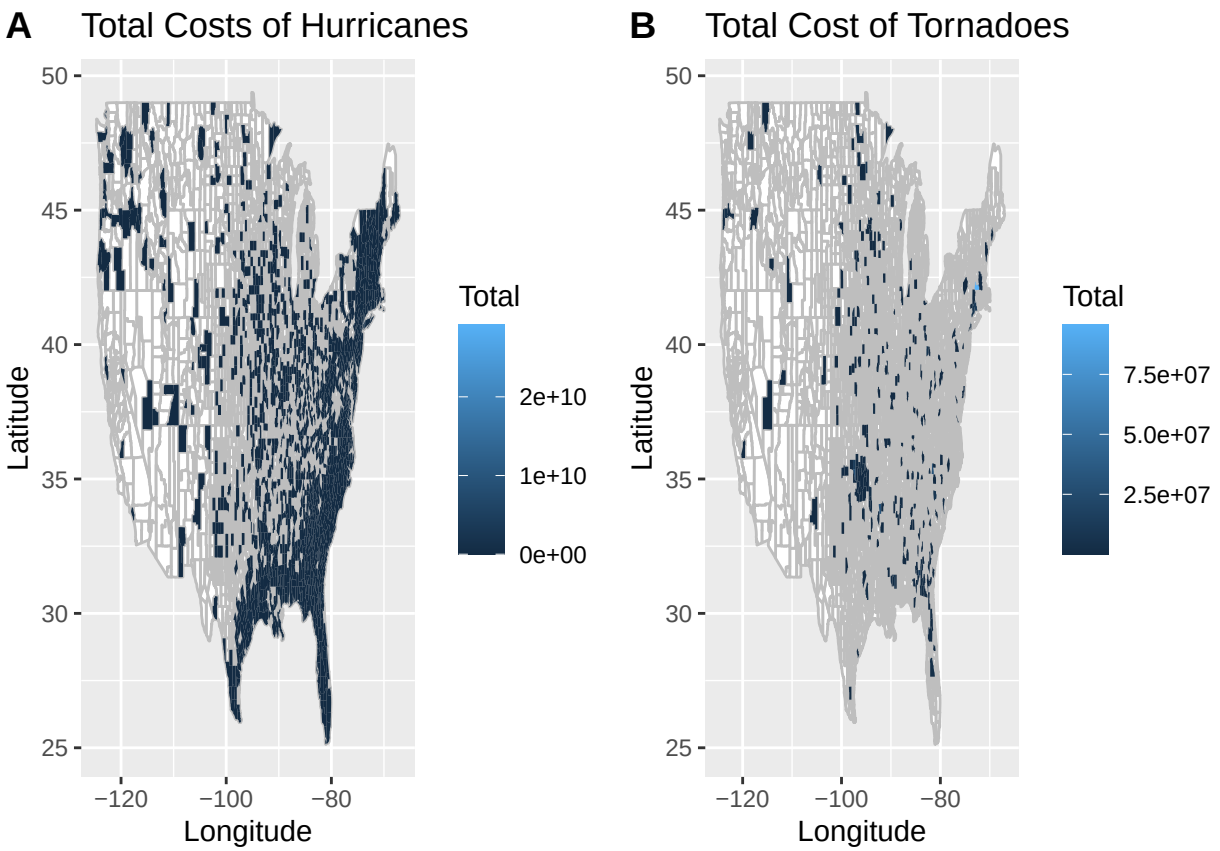
```
#Total cost of each county
a7=aggregate(projectAmount~county,data=tornado,FUN=sum)
#Unique applicant ID of each county
a8=aggregate(applicantId~county,data=tornado,FUN=unique)
#Number of unique applicants of each county
a8$number=lengths(a8$applicantId)
#Join the tables (a7 & a8)
t_data=merge(a7,a8)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

#Total Cost of Tornadoes 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 of Tornadoes",x="Longitude",y="Latitude",fill="Total") ->
  t_map1

#Average Costs of Tornadoes 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 of Tornadoes",x="Longitude",y="Latitude",fill="Average") ->
  t_map2

cowplot::plot_grid(t_map1,t_map2,labels="AUTO")
```

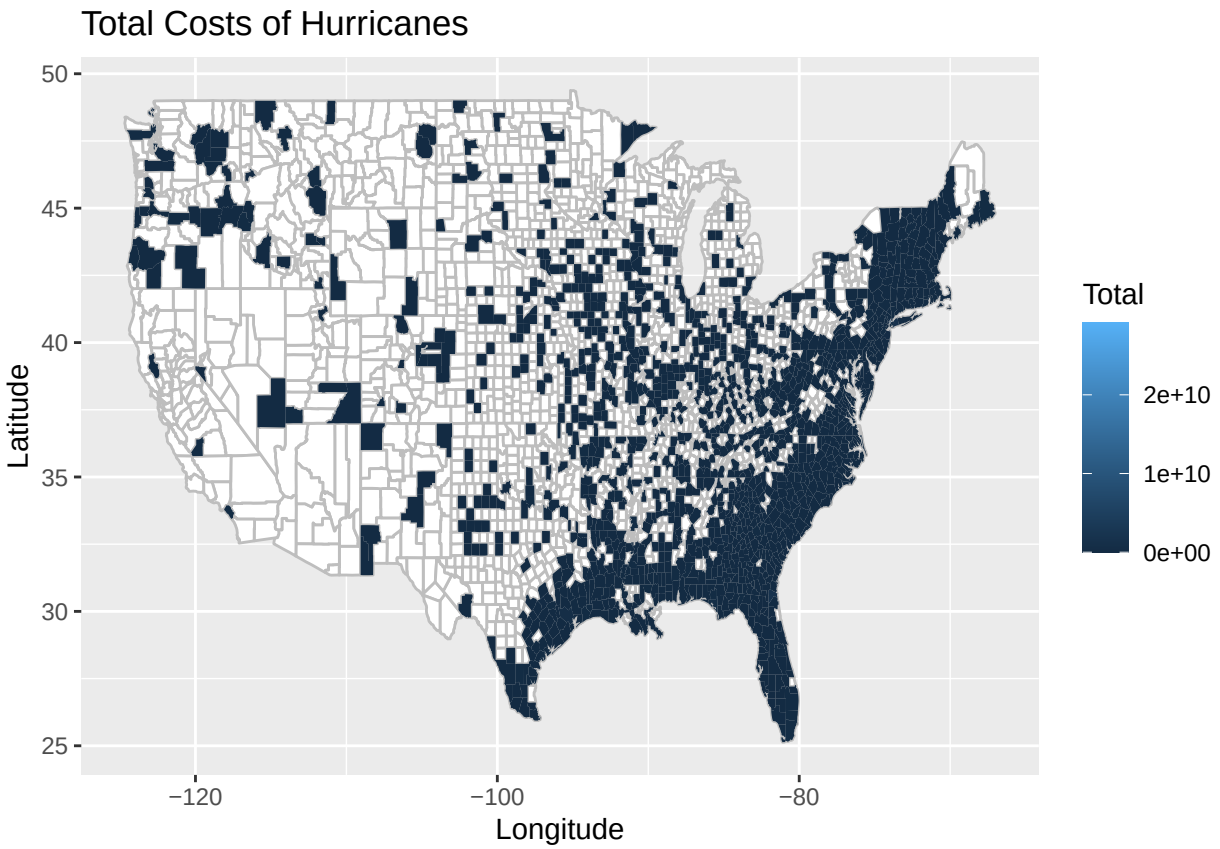


Difference?

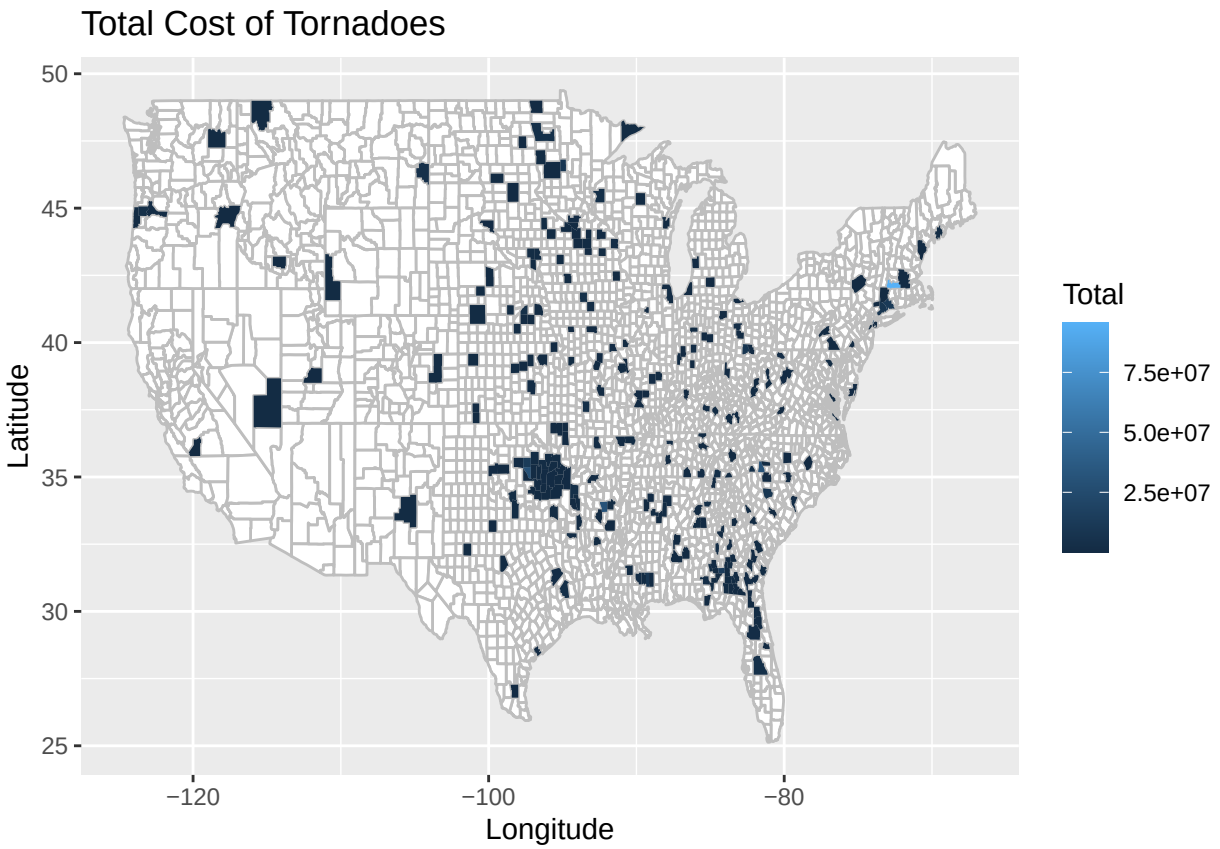
```
cowplot::plot_grid(h_map1,t_map1,labels="AUTO")
```


A
Total Costs of Hurricanes
B
Total Cost of Tornadoes
Latitude
Longitude
50
45
40
35
30
25
−120
−100
−80
Total
2e+10
1e+10
0e+00
7.5e+07
5.0e+07
2.5e+07

```
h_map1
```


Total Costs of Hurricanes
50
45
40
35
30
25
Latitude
−120
−100
−80
Longitude
Total
2e+10
1e+10
0e+00

```
t_map1
```



The estimated cost of hurricanes is relatively large near the east coastline, while the estimated cost of tornadoes does not show this trend. This is because a hurricane almost always forms over sea while a tornado usually forms over land.

Another obvious difference between tornadoes and hurricanes is that they have significantly different scales.