In Search of the Most Impactful Type of Weather Events Across the U.S.

Mark Rhine

Synopsis

In this report, I will analyze a listing of almost every weather event that occurred in the United States since 1996 and the health and economic consequences. The data comes from the National Oceanic & Atmospheric Administration. In terms of human health, I will show that tornadoes are the most harmful weather event. Also, in terms of economic damages, I will show that floods are the most harmful type of weather event.

Data Processing

First, I loaded the libraries needed in R to do the analysis.

```
library(dplyr)
library(lubridate)
library(quantmod)
library(ggplot2)
library(knitr)
library(grid)
library(gridExtra)
```

Then I downloaded the data file and loaded into R. Below is the code:

```
url <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url, destfile = "./weatherDataSet.csv.bz2")
#Load the raw data into a data frame called "mydata"
mydata <- read.csv(bzfile("./weatherDataSet.csv.bz2"))
```

Then I created a new data frame, that kept only the variables I deemed necessary for the analysis. It makes the data frame easier to work with. I kept variables related to date, event type, human fatalities, human injuries, property damage, crop damage, any remarks, and the unique reference number of the specific event. I also renamed some of the variables to make them easier to read.

```
#create a new data frame with only necessary variables & renamed a few.
dataf <- mydata[, c(2, 8, 23, 24, 25, 26, 27, 28, 36, 37)]
names(dataf)[1] <- "Date"
names(dataf)[2] <- "EventType"
names(dataf)[5] <- "PropertyDamage"
names(dataf)[6] <- "PropertyDamageScale"
names(dataf)[7] <- "CropDamage"
names(dataf)[8] <- "CropDamageScale"
```

The “Date”" variable is not yet a date data type, so I converted it to a date type. Now, I created a new variable, “Year”, that is just the year each event occurred. I do not think the day and month are necessary for the analysis, it is too much detail.

```
dataf$Date <- strptime(as.character(dataf$Date), format = "%m/%d/%Y %H:%M:%S" )
dataf$Year <- year(dataf$Date)
```

The reason the year each event occurred is so important is because it affects the analysis. First off, there is inflation, so dollar costs of damages cannot be compared as is. They need to be converted to the same year’s base prices.

To adjust event damages for inflation, a Consumer Price Index is needed to know the conversion rate. So I created a data frame that consists of the monthly date and CPI level.

```
#Create a data frame with Consumer Price Index by Year
getSymbols("CPIAUCSL", src='FRED')
```

```
## [1] "CPIAUCSL"
```

```
rawCPI <- CPIAUCSL
rawCPIdf <- data.frame(rawCPI)
rnames <- row.names(rawCPIdf)
rawCPIdf$date <- as.Date(rnames)
rawCPIdf$cpi <- rawCPIdf$CPIAUCSL
rawCPIdf$year <- year(rawCPIdf$date)
```

To make the analysis easier, I averaged the CPI level by each year instead of each month and year. “CPIdf” is the data frame that has the year and average CPI level.

```
#now get mean CPI per year
grouped <- group_by(rawCPIdf, year)
summed <- summarise(grouped, avgCPI = mean(cpi))
CPIdf <- data.frame(summed)
```

Next, I merged the “dataf” and “CPIdf” data frames together so that each weather event observation will have a variable specifying what the CPI index level was for that year in which it occurred. So we can adjust for inflation.

```
#Merge the two tables into one. Year variable is the primary-foreign key
finaldf <- merge(dataf, CPIdf, by.x = "Year", by.y = "year", all.x = T)
```

The Property Damage amount is currently a combination of 2 variables. So I had to create just 1 variable for the Property Damage amount. The same exact thing occurred with the Crop Damage amount.

```
#convert the variables to numeric and character
finaldf$PropertyDamageScale <- as.character(finaldf$PropertyDamageScale)
finaldf$CropDamageScale <- as.character(finaldf$CropDamageScale)
finaldf$PropertyDamage <- as.numeric(finaldf$PropertyDamage)
finaldf$CropDamage <- as.numeric(finaldf$CropDamage)
finaldf2 <- finaldf
```

I had to convert the Crop Damage scale variable from a letter representing thousands (k), millions (m), or billions (b) to their scale in nominal amounts. Uppercase and lowercase were used interchangeably in the raw data. I did not have the time to check each case, so I decided to give the variable a 0 value, meaning $0 in damages. I figured that with a data set so large any errors of non-inclusion would balance itself out. (the same proportion of weather event types will be non-included because it did not have the proper character in this variable). I did the same thing for the Crop Damage Scale.

```
finaldf2$CropDamageScale[!(finaldf2$CropDamageScale %in% c("K", "k", "M", "m", "B",
"b"))] <- 0
finaldf2$CropDamageScale[finaldf2$CropDamageScale == ""] <- 0
finaldf2$CropDamageScale[finaldf2$CropDamageScale %in% c("K", "k")] <- 1000
finaldf2$CropDamageScale[finaldf2$CropDamageScale %in% c("M", "m")] <- 1000000
finaldf2$CropDamageScale[finaldf2$CropDamageScale %in% c("B", "b")] <- 1000000000
finaldf2$CropDamageScale[finaldf2$CropDamageScale %in% c("?", "2")] <- 0

finaldf2$PropertyDamageScale[!(finaldf2$PropertyDamageScale %in% c("K", "k", "m", "M",
"b", "B"))] <- 0
finaldf2$PropertyDamageScale[finaldf2$PropertyDamageScale == ""] <- 0
finaldf2$PropertyDamageScale[finaldf2$PropertyDamageScale %in% c("K", "k")] <- 1000
finaldf2$PropertyDamageScale[finaldf2$PropertyDamageScale %in% c("M", "m")] <- 1000000
finaldf2$PropertyDamageScale[finaldf2$PropertyDamageScale %in% c("B", "b")] <- 1000000000

finaldf2$PropertyDamageScale <- as.numeric(finaldf2$PropertyDamageScale)
finaldf2$CropDamageScale <- as.numeric(finaldf2$CropDamageScale)
```

Then I simply multiplied the Damage factor variable by the damage scale amount to create a new single variable that was the dollar amount of Property damages. Also, a new variable was created for the dollar amount of crop damages. The amounts were still in historic dollars.

```
finaldf2$PropertyDamageHistoricDollars <- finaldf2$PropertyDamageScale * finaldf2$Propert
yDamage
finaldf2$CropDamageHistoricDollars <- finaldf2$CropDamageScale * finaldf2$CropDamage
```

Again these damages variables are still in historic cost. Which due to inflation makes them hard to compare. So I converted the historic cost to 2015 dollars using the average consumer Price Index (avgCPI) variable and the 2015 CPI value.

```
#Convert Historic Dollars of Damages to 2015 Dollars for Comparison
#2015 CPI index to bring all damages to current dollar
CPI2015 <- CPIdf[CPIdf$year == 2015, 2]
finaldf2$PropertyDamageCurrentDollar <- finaldf2$PropertyDamageHistoricDollars * (CPI2015
/ finaldf2$avgCPI)
finaldf2$CropDamageCurrentDollar <- finaldf2$CropDamageHistoricDollars * (CPI2015 / final
df2$avgCPI)
```

So now the Property and Crop damage amounts per event are all on the same scale of 2015 dollars. So now they can be compared.

Since my analysis deals with finding which type of weather event causes the most economic consequences in terms of damages, I don’t see a significant difference between property damages and crop damages. So I added them together into 1 variable, called “TotalDamage”.

```
finaldf2$TotalDamage <- finaldf2$PropertyDamageCurrentDollar + finaldf2$CropDamageCurrent
Dollar
```

On the same token, my analysis also deals with finding which types of weather event are the most harmful to human health. While fatalities are more serious, In terms of being harmful to health, I cannot discern between fatalities and injuries. So I combined the 2 variables together into 1 variable called “TotalInjuries”.

```
finaldf2$TotalInjuries <- finaldf2$FATALITIES + finaldf2$INJURIES
```

I checked to see if there are any obvious outliers, just by getting a visual of a scatterplot of specific events and their total damage. Below is the scatterplot (Figure 1).

```
plot(finaldf2$Year, finaldf2$TotalDamage, main = "Specific Storm Events and their Total D
amage", xlab="Year", ylab = "Total Damages (Dollars)")
```

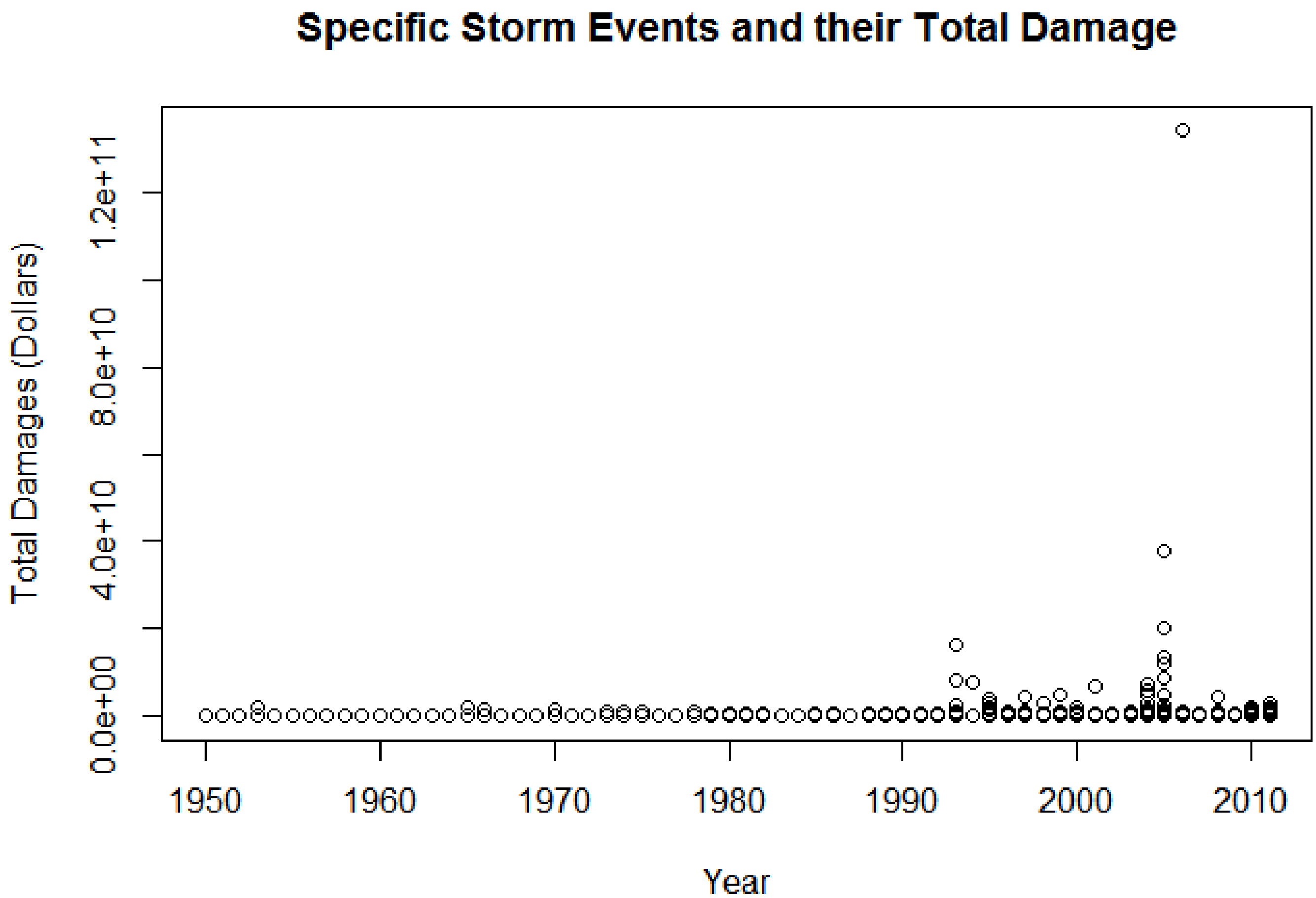

Figure 1: “Specific Storm Events and their Total Damage”. A scatterplot that plots each specific event by year and total damages. There seems to be one serious outlier.

There was one specific storm event that stood way above the others at over $120 billion. That seemed unreasonable.

```
temp <- which.max(finaldf2$TotalDamage)
#find out the RefNum of the event
h <- finaldf2[temp, 11]
h
```

```
## [1] 605943
```

Weather Event with a reference number of 605943 was the extreme outlier. Upon further review of the remarks of the event, it is clear that the amount of damages are supposed to be in millions instead of billions. So I corrected the Total Damages amount for this event.

```
#convert the outlier from billions to millions
finaldf2$TotalDamage[finaldf2$REFNUM == h] <- finaldf2$TotalDamage[finaldf2$REFNUM == h]
/ 1000
```

I had to remove the “Date” variable because the date data type is not supported by the ‘dplyr’ package I will later use to summarize the data.

```
#remove DATE column because it is an unsupported data type for dplyr library.
finaldf3 <- finaldf2[, -2]
```

There are many more events in the years 1996 - 2012 as opposed to 1950 - 1995. This is because there were more types of weather events being recorded starting January of 1996. This is per the NOAA website. So in order to compare the types of events on a fair basis, I excluded any events that occurred pre-1996.

```
finaldf4 <- finaldf3[finaldf3$Year > 1995,]
```

So now I am ready to summarize the Total Injuries and Total Damages by weather event.

```
grouping <- group_by(finaldf4, EventType)
sumfinal <- summarise(grouping, TotalDamagesCurrentDollars = sum(TotalDamage), TotalInjur
y = sum(TotalInjuries))
sumfinaldf <- data.frame(sumfinal)
```

This summarized data frame is still not ready to be analyzed and needs to be cleaned up. First, I removed any weather event type that had a total injuries = 0 AND total damages = 0. So if it had no consequences, it is irrelevant to my analysis.

```
#Clean out summarized data frame. Remove any Event Type that has 0 Injuries AND 0 Monetar
y Damages
sumfinaldf2 <- sumfinaldf
sumfinaldf3 <- sumfinaldf2[sumfinaldf2$TotalDamagesCurrentDollars > 0 & sumfinaldf2$Total
Injury >0,]
```

Some weather type events listed were the same exact event types as others, only they had different case characters. So I converted the EventType variable to lowercase, so when summarized again, they will be added together into 1 event type.

```
#now need to combine Event types that are the same but just character different
sumfinaldf4 <- sumfinaldf3
sumfinaldf4$EventType <- tolower(sumfinaldf4$EventType)
```

Many of the weather event types listed deviated slightly from those allowed in the dataset created by National Oceanic & Atmospheric Administration. For example, “Fog” should have been listed as “Dense Fog”. So I used best judgement to take those weather event types that were not listed exactly in the raw data documentation, and convert them to allowed weather type events. Furthermore, I took it upon myself to aggregate some weather type events into a single event type. I combined “flood” and “flashflood” together as just “flood”. In substance they are the same type of event. Any event type dealing with snow I combined into one type called “winter storm”. Also, any event type that was “wind” I combined into 1 event type called “high wind”. I don’t think the speed of the wind makes them a different type of event. I excluded “thunderstorm wind”" from “high wind” because they were caused by thunderstorms.

```
sumfinaldf4$EventType[grep("hurric", sumfinaldf4$EventType)] <- "hurricane"
sumfinaldf4$EventType[grep("coast", sumfinaldf4$EventType)] <- "coastal flood"
sumfinaldf4$EventType[grep("^cold", sumfinaldf4$EventType)] <- "cold"
sumfinaldf4$EventType[grep("^drought", sumfinaldf4$EventType)] <- "drought"
sumfinaldf4$EventType[grep(".cold", sumfinaldf4$EventType)] <- "extreme cold"
sumfinaldf4$EventType[grep("^high s", sumfinaldf4$EventType)] <- "high surf"
sumfinaldf4$EventType[grep("^heavy surf", sumfinaldf4$EventType)] <- "high surf"
sumfinaldf4$EventType[grep("extreme windchill", sumfinaldf4$EventType)] <- "extreme cold"
#converged flood and flash flood
sumfinaldf4$EventType[grep("^flood", sumfinaldf4$EventType)] <- "flood"
sumfinaldf4$EventType[grep("^flash", sumfinaldf4$EventType)] <- "flood"
sumfinaldf4$EventType[grep("^minor", sumfinaldf4$EventType)] <- "flood"
sumfinaldf4$EventType[grep("^river", sumfinaldf4$EventType)] <- "flood"

sumfinaldf4$EventType[grep("^water", sumfinaldf4$EventType)] <- "waterspout"
sumfinaldf4$EventType[grep("^wild", sumfinaldf4$EventType)] <- "wildfire"
sumfinaldf4$EventType[grep("strong wind", sumfinaldf4$EventType)] <- "high wind"
sumfinaldf4$EventType[grep("high wind", sumfinaldf4$EventType)] <- "high wind"
sumfinaldf4$EventType[grep("^wind", sumfinaldf4$EventType)] <- "high wind"
sumfinaldf4$EventType[grep("surf", sumfinaldf4$EventType)] <- "high surf"
sumfinaldf4$EventType[grep("^marine", sumfinaldf4$EventType)] <- "marine thunderstorm win
d"
sumfinaldf4$EventType[grep("thundersnow", sumfinaldf4$EventType)] <- "winter storm"
sumfinaldf4$EventType[grep("^thunder", sumfinaldf4$EventType)] <- "thunderstorm wind"
sumfinaldf4$EventType[grep("^tst", sumfinaldf4$EventType)] <- "thunderstorm wind"
sumfinaldf4$EventType[grep("^ic", sumfinaldf4$EventType)] <- "ice storm"

#converge anything with snow
sumfinaldf4$EventType[grep("snow", sumfinaldf4$EventType)] <- "winter storm"
sumfinaldf4$EventType[grep("blizzard", sumfinaldf4$EventType)] <- "winter storm"
sumfinaldf4$EventType[grep("winter storm", sumfinaldf4$EventType)] <- "winter storm"
sumfinaldf4$EventType[grep("wint", sumfinaldf4$EventType)] <- "winter storm"


sumfinaldf4$EventType[grep("^gust", sumfinaldf4$EventType)] <- "high wind"
sumfinaldf4$EventType[grep("heat", sumfinaldf4$EventType)] <- "heat"
sumfinaldf4$EventType[grep("warm", sumfinaldf4$EventType)] <- "heat"
sumfinaldf4$EventType[grep("microburst", sumfinaldf4$EventType)] <- "thunderstorm wind"
sumfinaldf4$EventType[grep("tornado", sumfinaldf4$EventType)] <- "tornado"
sumfinaldf4$EventType[grep("funnel", sumfinaldf4$EventType)] <- "tornado"
sumfinaldf4$EventType[grep("tropical", sumfinaldf4$EventType)] <- "tropical storm"
sumfinaldf4$EventType[grep("typhoon", sumfinaldf4$EventType)] <- "hurricane"
sumfinaldf4$EventType[grep("fire", sumfinaldf4$EventType)] <- "wildfire"
sumfinaldf4$EventType[grep("fog", sumfinaldf4$EventType)] <- "dense fog"
sumfinaldf4$EventType[grep("^freezing", sumfinaldf4$EventType)] <- "sleet"
sumfinaldf4$EventType[grep("glaze", sumfinaldf4$EventType)] <- "freeze"
```

```
sumfinaldf4$EventType[grep("heavy rain", sumfinaldf4$EventType)] <- "heavy rain"
sumfinaldf4$EventType[grep("wind damage", sumfinaldf4$EventType)] <- "high wind"
sumfinaldf4$EventType[grep("landslide", sumfinaldf4$EventType)] <- "landslide"
sumfinaldf4$EventType[grep("high water", sumfinaldf4$EventType)] <- "flood"
sumfinaldf4$EventType[grep("rip current", sumfinaldf4$EventType)] <- "rip current"
sumfinaldf4$EventType[grep("hail", sumfinaldf4$EventType)] <- "hail"
sumfinaldf4$EventType[grep("storm surge", sumfinaldf4$EventType)] <- "storm surge"
sumfinaldf4$EventType[grep("tidal flooding", sumfinaldf4$EventType)] <- "coastal flood"
sumfinaldf4$EventType[grep("urban", sumfinaldf4$EventType)] <- "flood"
sumfinaldf4$EventType[grep("whirl", sumfinaldf4$EventType)] <- "high wind"
```

Now I re-summarized my dataframe that combines the event types as seen in the above code

```
groupfin <- group_by(sumfinaldf4, EventType)
sumgroupfin <- summarise(groupfin, TotalInjuries = sum(TotalInjury), TotalDamages = sum(T
otalDamagesCurrentDollars))
sumgroupfin2 <- arrange(sumgroupfin, desc(TotalInjuries))
summarizedData <- data.frame(sumgroupfin2)
summarizedData$EventType <- as.factor(summarizedData$EventType)
summarizedData$EventType <- factor(summarizedData$EventType, as.character(summarizedDat
a$EventType))
```

So now there is a summarized data frame (“summarizedData”). It has the event type and total sum of human injuries (including fatalities) and total sum of damages (adjusted for inflation) for each event type since 1996. There are 28 weather event types in total.

Results

Most Harmful to Population Health

So which weather event type causes the most injuries (including fatalities) to humans? Tornadoes, with 22,179 total injuries since 1996. You can see our summarized data frame of event types sorted in descending order by total injuries (including fatalities).

```
summarizedData
```

```
##                  EventType TotalInjuries TotalDamages
## 1                  tornado         22179 2.941692e+10
## 2                    flood          9842 6.235106e+10
## 3                     heat          9664 5.849565e+08
## 4        thunderstorm wind          5533 1.110613e+10
## 5                lightning          4792 9.221458e+08
## 6             winter storm          3426 3.585226e+09
## 7                high wind          1889 7.519479e+09
## 8                 wildfire          1543 1.027501e+10
## 9                hurricane          1451 1.080079e+11
## 10             rip current          1045 2.139168e+05
## 11               dense fog           924 2.682290e+07
## 12                    hail           730 2.102234e+10
## 13               ice storm           427 4.758478e+09
## 14          tropical storm           395 1.086385e+10
## 15               high surf           392 1.058007e+08
## 16              dust storm           387 1.023726e+07
## 17               avalanche           379 4.371188e+06
## 18            extreme cold           366 1.936511e+09
## 19              heavy rain           324 1.679612e+09
## 20                  freeze           213 8.637976e+04
## 21                 tsunami           162 1.555326e+08
## 22                    cold           137 3.616560e+06
## 23               landslide            91 3.988694e+08
## 24 marine thunderstorm wind           56 7.323506e+06
## 25             storm surge            55 5.698342e+10
## 26              dust devil            41 7.785667e+05
## 27                   sleet            17 9.025569e+05
## 28           coastal flood            15 4.379881e+08
## 29                 drought             4 1.896648e+10
## 30                   frost             4 2.192901e+04
## 31                   other             4 1.519770e+06
## 32              waterspout             4 6.175713e+06
```

Here is a bar chart (Figure 2) showing the top 5 weather events in terms of total injuries.

```
sumTop5 <- summarizedData[1:5,]
g <- ggplot(sumTop5, aes(EventType, TotalInjuries))
g2 <- g + geom_bar(stat = "identity") + labs(title = "Top 5 Weather Event Types by Human Injuries", y = "Total Injuries", x = "Event Type")
g2
```

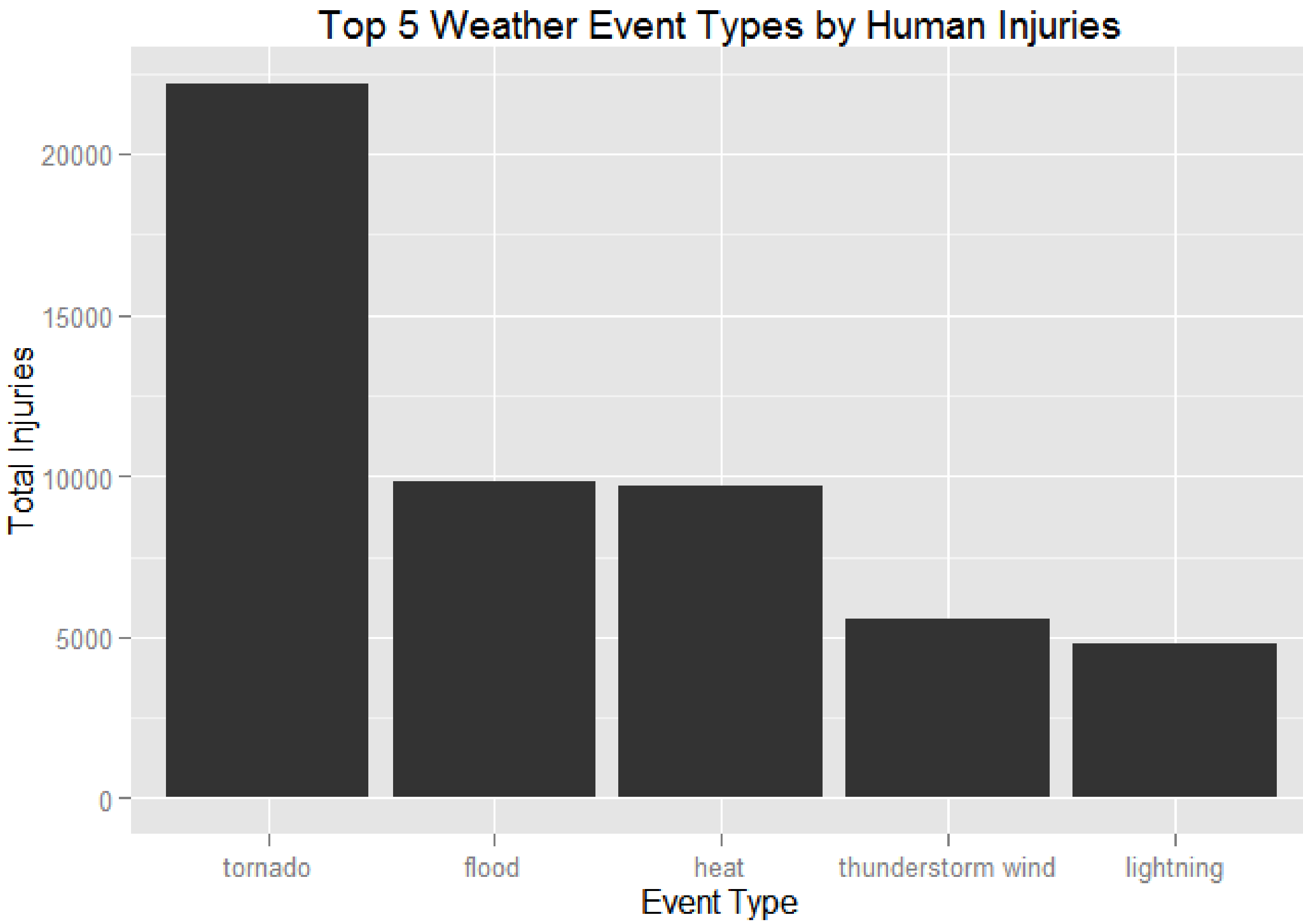

Figure 2: ‘Top 5 Weather Event Types by Human Injuries’. This figure shows the top 5 weather event types in terms of human injuries sustained since 1996. This includes body injuries and fatalities. Tornadoes are the number 1 most harmful weather type.

So tornadoes are clear and away the most harmful weather event. Followed distantly by floods with 9,842 and heat waves with 9,664. One could make the conclusion that thunderstorm wind and lightning should be combined together as just a “thunderstorm” event. Even so, it would not come close to the amount of total injuries and fatalities caused by tornadoes.

Greatest Economic Consequences

Which weather event type caused the most monetary damages from 1996-2011? Hurricanes, with over $108 Billion total dollars in damage. All costs are converted to 2015 dollars.

```
sumgroupfin3 <- arrange(sumgroupfin, desc(TotalDamages))
summarizedData9 <- data.frame(sumgroupfin3)
summarizedData9$EventType <- as.factor(summarizedData9$EventType)
summarizedData9$EventType <- factor(summarizedData9$EventType, as.character(summarizedDat
a9$EventType))
```

I created a bar chart showing the top 5 most costly type of events. It will be shown later as the left plot in Figure 3.

```
sumTop5Dam <- summarizedData9[1:5,]
sumTop5Dam$TotalDamages <- sumTop5Dam$TotalDamages / 1000000000
k <- ggplot(sumTop5Dam, aes(EventType, TotalDamages))
k2 <- k + geom_bar(stat = "identity") + labs(title = "1996-2011", y = "Total Cost of Dama
ges (in billions)", x = "Event Type")
k2 <- k2 + coord_cartesian(ylim=c(0, 120))
```

Below is the data frame showing the top 5 weather events in terms of total damages.

```
sumTop5Dam
```

```
##     EventType TotalInjuries TotalDamages
## 1   hurricane          1451    108.00786
## 2       flood          9842     62.35106
## 3 storm surge            55     56.98342
## 4     tornado         22179     29.41692
## 5        hail           730     21.02234
```

So we can see that hurricanes have caused the most economic harm across the United States with over $108 billion in damages since 1996. Followed by floods that have cost 62.4 billion in damages since 1996. In third place are storm surges with 57 billion in damages since 1996.

Hurricane Katrina

When looking through some of the highest costing specific storm events, I noticed several were attributed to Hurricane Katrina. So I decided to see how much Total Damage this one event caused. Since it was attributed to many different types of events in addition to Hurricane (ex: storm surge or flood), I searched the Remarks column for ‘Katrina’.

```
CostKatrina <- finaldf3$TotalDamage[grep("[Kk]atrina", finaldf3$REMARKS)]
TotCostKat <- sum(CostKatrina)
TotCostKat
```

```
## [1] 41433568699
```

This may not have captured all of the costs of Hurricane Katrina, but it means that it caused at least $41 Billion in total damages. This was a unique situation as the levees of New Orleans broke and so the area flooded causing an abnormally high amount of damages. So what if we ignored this storm? Would that change the outcome? In order to check, I decided to ignore the entire year of 2005 so that it would exclude all events related to Katrina but also so my exclusion would affect all weather event types as fairly as possible. So I went back to the data frame that was used right before I subsetted by year originally. Now I also excluded the year 2005 and ran through all of the processing code once again.

```
#finaldf3 was the data frame of pre-summarized data, right before t was subsetted by yea
r.
#now need to subset by year, ecluding 2005.
#proccess code in same way as before.
#Copied code from earlier, put 99 or 100 at end of variable names.
finaldf99 <- finaldf3[finaldf3$Year > 1995 & finaldf3$Year != 2005,]
grouping99 <- group_by(finaldf99, EventType)
sumfinal99 <- summarise(grouping99, TotalDamagesCurrentDollars = sum(TotalDamage), TotalI
njury = sum(TotalInjuries))
sumfinaldf99 <- data.frame(sumfinal99)
```

I again, cleaned out the first draft of the summarized data frame. I Removed any Event Type that has 0 Injuries AND 0 Monetary Damages

```
sumfinaldf99 <- sumfinaldf99[sumfinaldf99$TotalDamagesCurrentDollars > 0 & sumfinaldf99$T
otalInjury >0,]
```

I again combined events in the exact same manner as I did before.

```
#now need to combine Event types that are the same but just character different
sumfinaldf100 <- sumfinaldf99
sumfinaldf100$EventType <- tolower(sumfinaldf100$EventType)

sumfinaldf100$EventType[grep("hurric", sumfinaldf100$EventType)] <- "hurricane"
sumfinaldf100$EventType[grep("coast", sumfinaldf100$EventType)] <- "coastal flood"
sumfinaldf100$EventType[grep("^cold", sumfinaldf100$EventType)] <- "cold"
sumfinaldf100$EventType[grep("^drought", sumfinaldf100$EventType)] <- "drought"
sumfinaldf100$EventType[grep(".cold", sumfinaldf100$EventType)] <- "extreme cold"
sumfinaldf100$EventType[grep("^high s", sumfinaldf100$EventType)] <- "high surf"
sumfinaldf100$EventType[grep("^heavy surf", sumfinaldf100$EventType)] <- "high surf"
sumfinaldf100$EventType[grep("extreme windchill", sumfinaldf100$EventType)] <- "extreme c
old"
#converged flood and flash flood
sumfinaldf100$EventType[grep("^flood", sumfinaldf100$EventType)] <- "flood"
sumfinaldf100$EventType[grep("^flash", sumfinaldf100$EventType)] <- "flood"
sumfinaldf100$EventType[grep("^minor", sumfinaldf100$EventType)] <- "flood"
sumfinaldf100$EventType[grep("^river", sumfinaldf100$EventType)] <- "flood"

sumfinaldf100$EventType[grep("^water", sumfinaldf100$EventType)] <- "waterspout"
sumfinaldf100$EventType[grep("^wild", sumfinaldf100$EventType)] <- "wildfire"
sumfinaldf100$EventType[grep("strong wind", sumfinaldf100$EventType)] <- "high wind"
sumfinaldf100$EventType[grep("high wind", sumfinaldf100$EventType)] <- "high wind"
sumfinaldf100$EventType[grep("^wind", sumfinaldf100$EventType)] <- "high wind"
sumfinaldf100$EventType[grep("surf", sumfinaldf100$EventType)] <- "high surf"
sumfinaldf100$EventType[grep("^marine", sumfinaldf100$EventType)] <- "marine thunderstorm
wind"
sumfinaldf100$EventType[grep("thundersnow", sumfinaldf100$EventType)] <- "winter storm"
sumfinaldf100$EventType[grep("^thunder", sumfinaldf100$EventType)] <- "thunderstorm wind"
sumfinaldf100$EventType[grep("^tst", sumfinaldf100$EventType)] <- "thunderstorm wind"
```

```
sumfinaldf100$EventType[grep("^ic", sumfinaldf100$EventType)] <- "ice storm"

#converge anything with snow
sumfinaldf100$EventType[grep("snow", sumfinaldf100$EventType)] <- "winter storm"
sumfinaldf100$EventType[grep("blizzard", sumfinaldf100$EventType)] <- "winter storm"
sumfinaldf100$EventType[grep("winter storm", sumfinaldf100$EventType)] <- "winter storm"
sumfinaldf100$EventType[grep("wint", sumfinaldf100$EventType)] <- "winter storm"


sumfinaldf100$EventType[grep("^gust", sumfinaldf100$EventType)] <- "high wind"
sumfinaldf100$EventType[grep("heat", sumfinaldf100$EventType)] <- "heat"
sumfinaldf100$EventType[grep("warm", sumfinaldf100$EventType)] <- "heat"
sumfinaldf100$EventType[grep("microburst", sumfinaldf100$EventType)] <- "thunderstorm win
d"
sumfinaldf100$EventType[grep("tornado", sumfinaldf100$EventType)] <- "tornado"
sumfinaldf100$EventType[grep("funnel", sumfinaldf100$EventType)] <- "tornado"
sumfinaldf100$EventType[grep("tropical", sumfinaldf100$EventType)] <- "tropical storm"
sumfinaldf100$EventType[grep("typhoon", sumfinaldf100$EventType)] <- "hurricane"
sumfinaldf100$EventType[grep("fire", sumfinaldf100$EventType)] <- "wildfire"
sumfinaldf100$EventType[grep("fog", sumfinaldf100$EventType)] <- "dense fog"
sumfinaldf100$EventType[grep("^freezing", sumfinaldf100$EventType)] <- "sleet"
sumfinaldf100$EventType[grep("glaze", sumfinaldf100$EventType)] <- "freeze"
sumfinaldf100$EventType[grep("heavy rain", sumfinaldf100$EventType)] <- "heavy rain"
sumfinaldf100$EventType[grep("wind damage", sumfinaldf100$EventType)] <- "high wind"
sumfinaldf100$EventType[grep("landslide", sumfinaldf100$EventType)] <- "landslide"
sumfinaldf100$EventType[grep("high water", sumfinaldf100$EventType)] <- "flood"
sumfinaldf100$EventType[grep("rip current", sumfinaldf100$EventType)] <- "rip current"
sumfinaldf100$EventType[grep("hail", sumfinaldf100$EventType)] <- "hail"
sumfinaldf100$EventType[grep("storm surge", sumfinaldf100$EventType)] <- "storm surge"
sumfinaldf100$EventType[grep("tidal flooding", sumfinaldf100$EventType)] <- "coastal floo
d"
sumfinaldf100$EventType[grep("urban", sumfinaldf100$EventType)] <- "flood"
sumfinaldf100$EventType[grep("whirl", sumfinaldf100$EventType)] <- "high wind"

#re-summarize the data frame by sum.
groupfin100 <- group_by(sumfinaldf100, EventType)
sumgroupfin100 <- summarise(groupfin100, TotalInjuries = sum(TotalInjury), TotalDamages =
sum(TotalDamagesCurrentDollars))
```

I sorted the data frame by Total Damages.

```
#sort the summarized data frame by Total Damages, decreasing order
sumgroupfin200 <- arrange(sumgroupfin100, desc(TotalDamages))
summarizedData100 <- data.frame(sumgroupfin200)
summarizedData100$EventType <- as.factor(summarizedData100$EventType)
summarizedData100$EventType <- factor(summarizedData100$EventType, as.character(summarize
dData100$EventType))

#Only exract the Top 5, to be used in a bar chart
sumTop5Dam100 <- summarizedData100[1:5,]
sumTop5Dam100$TotalDamages <- sumTop5Dam100$TotalDamages / 1000000000
```

Below is a plot (Figure 3) that shows the different results between including 2005 and excluding 2005.

```
u <- ggplot(sumTop5Dam100, aes(EventType, TotalDamages))
u2 <- u + geom_bar(stat = "identity") + labs(title = "1996-2011, Excluding 2005", y = "To
tal Cost of Damages (in billions)", x = "Event Type")
u2 <- u2 + coord_cartesian(ylim=c(0, 120))
grid.arrange(k2, u2, ncol = 2, main = "Top 5 Weather Event Types by Cost of Damages")
```

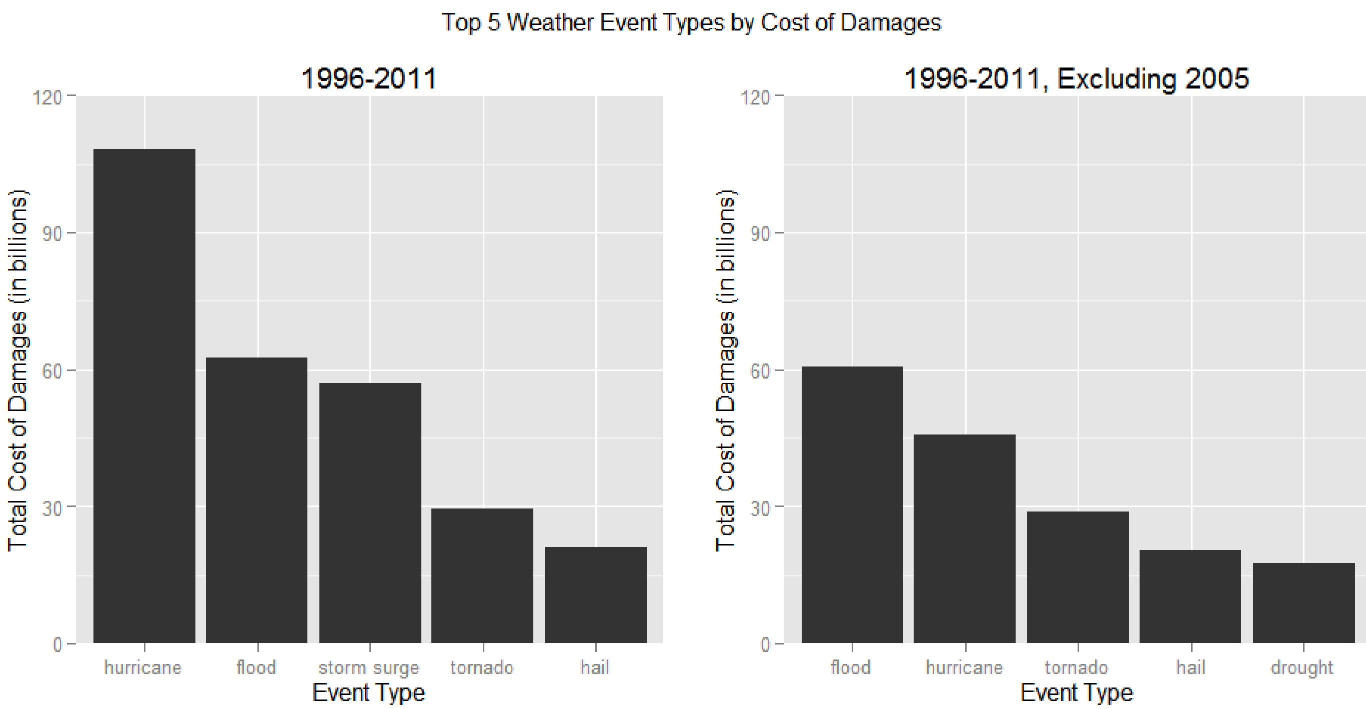

Figure 3: ‘Top 5 Weather Event Types by Cost of Damages’. This figure shows the top 5 weather event types in terms of their total cost in damages (in billions) since 1996. All costs were adjusted for inflation and are in 2015 dollars. The left plot includes 2005 and thus Hurricane Katrina. The right plot excludes 2005. Before excluding 2005, Hurricanes caused the most damage. After exclusion, it was Floods.

Interestingly, now floods have the greatest economic consequences. Hurricanes fall to second place. Storm surges are no longer even in the top 5.

In my opinion, this is a more faithful representation of the true economic consequences of each weather type. Hurricane Katrina was very unique in that the town of New Orleans is below sea level and with the levees breaking it caused much more damage than it would have. So, I would argue floods have the greatest economic consequences.

*Excluding 2005 had no effect on the relationship between Total Injuries caused by event type.