Most Harmfull Events Across United States From 1950 To 2011

Synopsis

In this report we explore U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. From these data, we found out, that Tornado is the most harmful type of event with respect to population health. We also saw that while Excessive Heat was the third most harmful event in 1991-2000, it became the second most harmful event type in 2001-2011. In terms of economic consequences, Flood is the most harmful type of event, followed by Hurricane/Typhoon and Tornado.

Data Processing

Here is a description of the steps taken to download data, read data into R and preprocess it.
1. Create a directory which will hold the NOAA Storm Database file
2. Download data from the web
3. Read data into R using read.csv. This may take several minutes (no need to unzip)

```
# 1
if(!dir.exists("./NOAA_Storm_Database")){dir.create("./NOAA_Storm_Database")}

# 2
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
if(!file.exists("./NOAA_STORM_DATABASE/storm_data.csv.bz2")) {download.file(fileUrl,
destfile = "./NOAA_Storm_Database/storm_data.csv.bz2", method = "curl")}

# 3
storm_data <- read.csv("./NOAA_Storm_Database/storm_data.csv.bz2")
```

We can now look at the first few rows of the dataset (there are 902297 rows in the dataset). We also look at the column names.

```
dim(storm_data)
```

```
## [1] 902297     37
```

```
head(storm_data[,1:8])
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE
## 1 TORNADO
## 2 TORNADO
## 3 TORNADO
## 4 TORNADO
## 5 TORNADO
## 6 TORNADO
```

```
names(storm_data)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Results

We want to answer two questions. Here is the first one.

Across the United States, which types of events (as indicated in the `EVTYPE` variable) are most harmful with respect to population health?

Step_by_step description of analysis which is followed by the corresponding code.

1. For this part we only need 4 columns from the dataset (**BGN_DATE**, **EVTYPE**, **FATALITIES**, **INJURIES**). Let's sumbset the data frame to have only these 4 columns (using **dplyr** package).
2. Using **lubridate** package convert the **BGN_DATE** column to **POSIXct**.
3. Add a new column to the data frame called **YEAR** to have only the years from the **BGN_DATE** column.
4. We know that there are fewer records in earlyer years so we will create a new column which we will call **INTERVAL_YEAR**. This new column divides **YEAR** column into several intervals to allow us observe the

more recent years separately from the earlyer years.

5. Make two plots.

- **PANEL PLOT** of top 5 types of events per interval.
    a. Add two columns to our data_frame which count total number of event types and total fatalities and injuries in each interval. From these columns create a data frame with only 4 rows (one row for each interval totals) which we will later use to annotate our plot. We are going to compare every type of event in each interval against these interval totals.
    b. Create a new column called **PERCENT**. For each interval this column will have the percentages of the number of fatalities and injuries. From **PERCENT** we will create another column **PERCENT_CHAR** which is the character representation of **PERCENT** (with "%" added at the end of each number).
    c. Create a data frame called **health_df_top_5** which will have only the top 5 events for each interval.
- **SINGLE PLOT** of top 10 types of events that caused fatalities and injuries during 1950-2011.
    a. Create a data frame (**health_df_total**) which will have total number of fatalities and injuries per event type.
    b. Calculate total number of event types (1 number), as well as total number of fatalities and injuries (1 number). We will use these numbers to annotate our plot. And we will compare fatalities and injuries of each of the 10 event types to this total number of fatalities and injuries.
    c. In the same way as above, create **PERCENT** and **PERCENT_CHAR** columns to make our plot more informative.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
# 1
health_df <- storm_data %>% select(BGN_DATE, EVTYPE, FATALITIES, INJURIES)
dim(health_df)
```

```
## [1] 902297      4
```

```
head(health_df, 3)
```

```
##            BGN_DATE  EVTYPE FATALITIES INJURIES
## 1 4/18/1950 0:00:00 TORNADO          0       15
## 2 4/18/1950 0:00:00 TORNADO          0        0
## 3 2/20/1951 0:00:00 TORNADO          0        2
```

```
# 2
health_df$BGN_DATE <- mdy_hms(as.character(health_df$BGN_DATE))
dim(health_df)
```

```
## [1] 902297      4
```

```
head(health_df, 3)
```

```
##     BGN_DATE  EVTYPE FATALITIES INJURIES
## 1 1950-04-18 TORNADO          0       15
## 2 1950-04-18 TORNADO          0        0
## 3 1951-02-20 TORNADO          0        2
```

```
# 3
health_df <- health_df %>% mutate(YEAR = year(health_df$BGN_DATE))
dim(health_df)
```

```
## [1] 902297      5
```

```
head(health_df, 3)
```

```
##     BGN_DATE  EVTYPE FATALITIES INJURIES YEAR
## 1 1950-04-18 TORNADO          0       15 1950
## 2 1950-04-18 TORNADO          0        0 1950
## 3 1951-02-20 TORNADO          0        2 1951
```

```
# 4
health_df <- health_df %>%
        mutate(INTERVAL_YEAR = cut(YEAR,
                                   breaks = c(1950, 1970, 1990, 2000, 2011),
                                   labels = c("1950-1970", "1971-1990", "1991-2000",
"2001-2011"),
                                   include.lowest = TRUE))
dim(health_df)
```

```
## [1] 902297      6
```

```
head(health_df, 3)
```

```
##     BGN_DATE  EVTYPE FATALITIES INJURIES YEAR INTERVAL_YEAR
## 1 1950-04-18 TORNADO          0       15 1950     1950-1970
## 2 1950-04-18 TORNADO          0        0 1950     1950-1970
## 3 1951-02-20 TORNADO          0        2 1951     1950-1970
```

```
# 5

# Panel Plot

# a
health_df_interval_event <- health_df %>%
        group_by(INTERVAL_YEAR, EVTYPE) %>%
        summarize(FATALITY_INJURY = sum(FATALITIES, INJURIES))

health_df_interval_event_count <- health_df_interval_event %>%
        group_by(INTERVAL_YEAR) %>%
        mutate(count_event_type = length(EVTYPE),
               count_fatal_inj = sum(FATALITY_INJURY))
dim(health_df_interval_event_count)
```

```
## [1] 1105    5
```

```
head(health_df_interval_event_count, 3)
```

```
## Source: local data frame [3 x 5]
## Groups: INTERVAL_YEAR [1]
##
##   INTERVAL_YEAR    EVTYPE FATALITY_INJURY count_event_type count_fatal_inj
##           <fctr>    <fctr>           <dbl>            <int>           <dbl>
## 1     1950-1970      HAIL               0                3           35524
## 2     1950-1970   TORNADO           35524                3           35524
## 3     1950-1970 TSTM WIND               0                3           35524
```

```
counts_df <- health_df_interval_event_count %>%
        select(INTERVAL_YEAR, count_event_type, count_fatal_inj) %>%
        group_by(INTERVAL_YEAR) %>%
        filter(row_number() <= 1)
dim(counts_df)
```

```
## [1] 4 3
```

```
counts_df
```

```
## Source: local data frame [4 x 3]
## Groups: INTERVAL_YEAR [4]
##
##   INTERVAL_YEAR count_event_type count_fatal_inj
##           <fctr>            <int>           <dbl>
## 1     1950-1970                3           35524
## 2     1971-1990                3           37283
## 3     1991-2000              930           45019
## 4     2001-2011              169           37847
```

```
# b
health_df_interval_event_count_percentage <- health_df_interval_event_count %>%
        group_by(INTERVAL_YEAR, EVTYPE) %>%
        mutate(PERCENT = FATALITY_INJURY * 100 / count_fatal_inj) %>%
        mutate(PERCENT_CHAR = paste(as.character(round(PERCENT, digits = 2)), "%"))
dim(health_df_interval_event_count_percentage)
```

```
## [1] 1105    7
```

```
head(health_df_interval_event_count_percentage, 3)
```

```
## Source: local data frame [3 x 7]
## Groups: INTERVAL_YEAR, EVTYPE [3]
##
##   INTERVAL_YEAR    EVTYPE FATALITY_INJURY count_event_type count_fatal_inj
##          <fctr>    <fctr>           <dbl>            <int>           <dbl>
## 1     1950-1970      HAIL               0                3           35524
## 2     1950-1970   TORNADO           35524                3           35524
## 3     1950-1970 TSTM WIND               0                3           35524
## # ... with 2 more variables: PERCENT <dbl>, PERCENT_CHAR <chr>
```

```
# c
health_df_top_5 <- health_df_interval_event_count_percentage %>%
        group_by(INTERVAL_YEAR) %>%
        arrange(INTERVAL_YEAR, desc(FATALITY_INJURY)) %>%
        top_n(n = 5, wt = FATALITY_INJURY)
dim(health_df_top_5)
```

```
## [1] 16  7
```

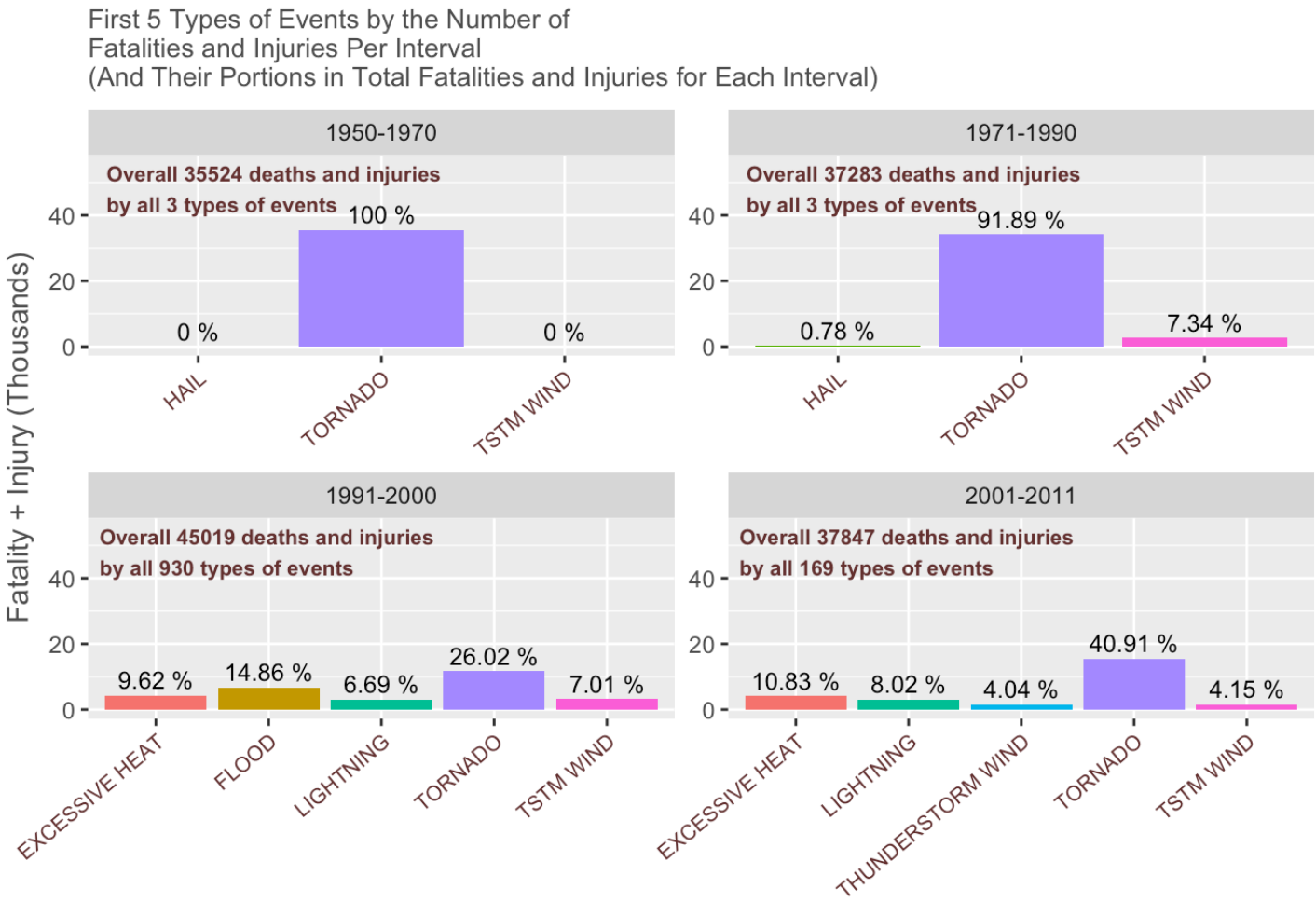
```
head(health_df_top_5)
```

```
## Source: local data frame [6 x 7]
## Groups: INTERVAL_YEAR [2]
##
##   INTERVAL_YEAR    EVTYPE FATALITY_INJURY count_event_type count_fatal_inj
##          <fctr>    <fctr>           <dbl>            <int>           <dbl>
## 1     1950-1970   TORNADO           35524                3           35524
## 2     1950-1970      HAIL               0                3           35524
## 3     1950-1970 TSTM WIND               0                3           35524
## 4     1971-1990   TORNADO           34259                3           37283
## 5     1971-1990 TSTM WIND            2735                3           37283
## 6     1971-1990      HAIL             289                3           37283
## # ... with 2 more variables: PERCENT <dbl>, PERCENT_CHAR <chr>
```

We have everything we need to make the panel plot.

```
# The plot
library(ggplot2)

div = 1000
ggplot(health_df_top_5, aes(x = EVTYPE, y = FATALITY_INJURY/div, fill = EVTYPE))+
        facet_wrap(~INTERVAL_YEAR, nrow = 2, scales = "free") +
        geom_bar(stat = "Identity") +
        geom_text(aes(label = PERCENT_CHAR), vjust = -0.4, size = 3.2) +
        labs(title = "First 5 Types of Events by the Number of \nFatalities and Injur
ies Per Interval\n(And Their Portions in Total Fatalities and Injuries for Each Inter
val)",
             x = "", y = "Fatality + Injury (Thousands)") +
        theme(legend.position = "none") +
        theme(axis.text.x = element_text(angle = 40, hjust = 1, vjust = 1, size = 8,
color = rgb(0.4,0.2,0.2))) +
        theme(axis.title.y = element_text(color = rgb(0.3,0.3,0.3))) +
        theme(plot.title = element_text(color = rgb(0.3,0.3,0.3), size = 10)) +
        scale_y_continuous(limits = c(min(health_df_top_5$FATALITY_INJURY),
                                      max(health_df_top_5$FATALITY_INJURY)+20000)/div
) +
        annotate("text", x = 0.5, y = 55, size = 2.8, hjust = "inward", vjust = "inwa
rd",
                 color = rgb(0.4,0.2,0.2), fontface = 2,
                 label = paste("Overall", as.character(counts_df$count_fatal_inj),
                               "deaths and injuries\nby all",
                               as.character(counts_df$count_event_type),
                               "types of events"))
```



From the plot above we clearly see that in all 4 interval the most harmful event type is Tornado. The second harmful event is different from interval to interval. First of all, we notice that in 1950-1970 and 1971-1990 only three types of events were recorded (and that is why the sum of percentages is 100%).
1950-1970 - Tornado is the only winner in terms of human health harms.
1971-1990 - Tornado is the first harmful event type. Second harmful event type is Tstm Wind (which is Marine Thunderstorm Wind, by the way, as we can see in Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)). And Hail is the third harmful event type with respect to population health.

In 1991-2000 and 2001-2011 there are many types of events recorded (that is why the sum of percentages is not equal to 100%, as only 5 event types are shown on the plot).

1991-2000 - Tornado is the first harmful event type, Flood is the second. Then come Excessive Heat, Tstm Wind and Lightning.
2001-2011 - Tornado is again the first harmful event type, but this time Excessive Heat goes up one level from the previous decade and is now the second harmful event (global warming?). And then come Lightning, Tstm Wind, and Thunderstorm Wind.

We can now move on to see the picture of these 4 intervals taken togetheer.

```
# Single Plot
library(dplyr)
library(ggplot2)

# a
health_df_total <- health_df %>%
        group_by(EVTYPE) %>%
        summarize(FATALITY_INJURY = sum(FATALITIES, INJURIES)) %>%
        arrange(desc(FATALITY_INJURY))

# b
count_event_types <- nrow(health_df_total)
count_fatal_injur <- sum(health_df_total$FATALITY_INJURY)

# c
health_df_total <- health_df_total %>%
        mutate(PERCENT = FATALITY_INJURY * 100/count_fatal_injur) %>%
        mutate(PERCENT_CHAR = paste(as.character(round(PERCENT, digits = 2)),"%"))
dim(health_df_total)
```

```
## [1] 985   4
```

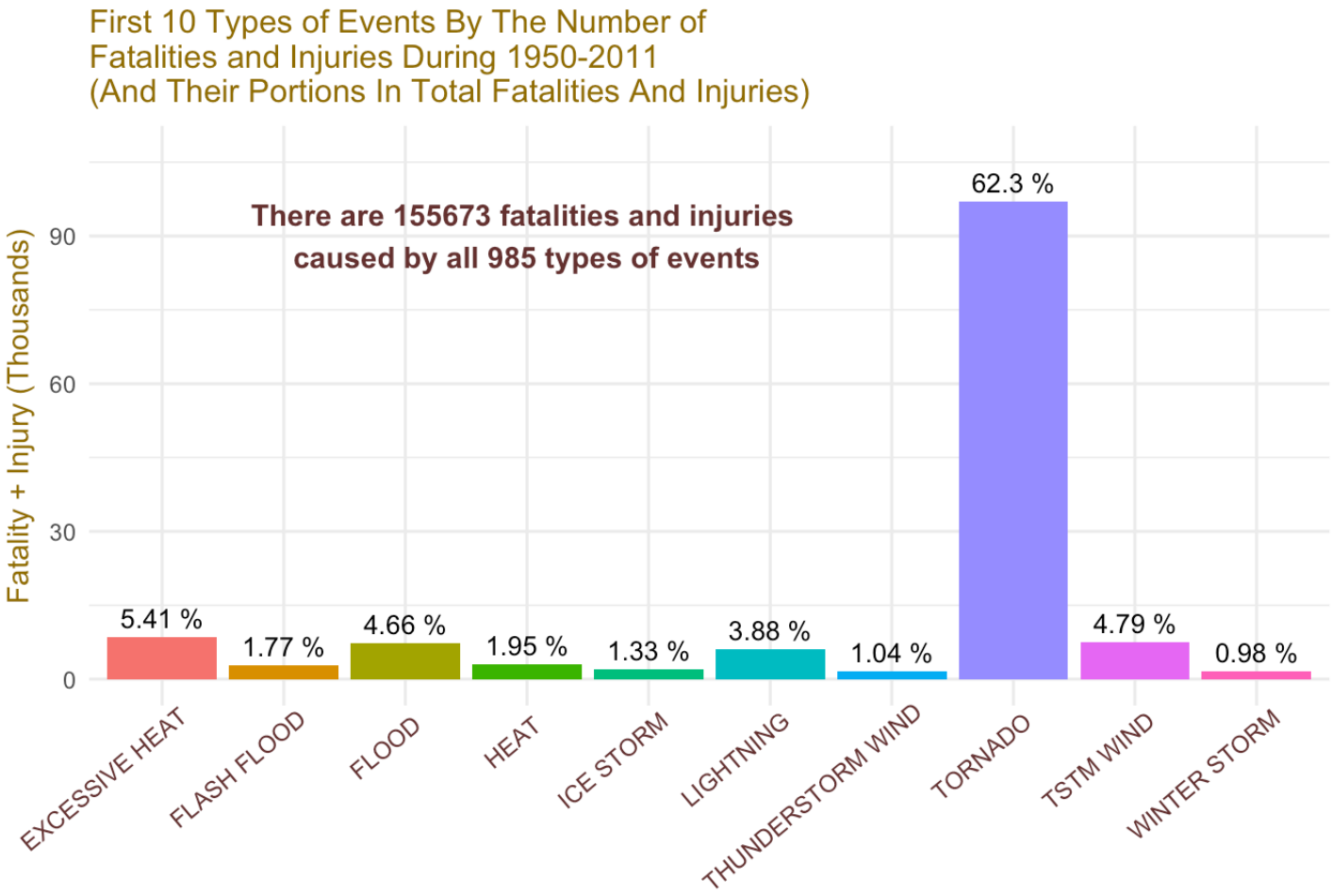
```
head(health_df_total, 3)
```

```
## # A tibble: 3 × 4
##           EVTYPE FATALITY_INJURY   PERCENT PERCENT_CHAR
##           <fctr>           <dbl>     <dbl>        <chr>
## 1        TORNADO           96979 62.296609       62.3 %
## 2 EXCESSIVE HEAT            8428  5.413912       5.41 %
## 3      TSTM WIND            7461  4.792739       4.79 %
```

And we have everything needed to see the most harmful events with respect to population health in one single plot.

```
# The plot
library(ggplot2)

div = 1000
ggplot(health_df_total[1:10,], aes(x = EVTYPE, y = FATALITY_INJURY/div, fill = EVTYPE
)) +
        geom_bar(stat = "identity") +
        geom_text(aes(label = PERCENT_CHAR), vjust = -0.5, size = 3.5) +
        labs(title = "First 10 Types of Events By The Number of \nFatalities and Inju
ries During 1950-2011\n(And Their Portions In Total Fatalities And Injuries)") +
        labs(x = "") +
        labs(y = "Fatality + Injury (Thousands)") +
        theme_minimal() +
        theme(legend.position = "none") +
        theme(axis.text.x = element_text(angle = 40, hjust = 0.9, vjust = 1, size = 9
, color = rgb(0.4,0.2,0.2))) +
        theme(axis.title.y = element_text(color = colorRampPalette(c("green", "red"))
(15)[9])) +
        theme(plot.title = element_text(color = colorRampPalette(c("green", "red"))(1
5)[9], size = 12)) +
        scale_y_continuous(limits = c(min(health_df_total$FATALITY_INJURY),
                                      max(health_df_total$FATALITY_INJURY)+10000)/div
) +
        annotate("text", x = 4, y = 90, color = rgb(0.4,0.2,0.2), fontface = 2,
                 label = paste("There are", as.character(count_fatal_injur),
                               "fatalities and injuries \ncaused by all",
                                as.character(count_event_types), "types of events"))
```



As we would guess Tornado is the first harmful event, Excessive Heat is scaringly the second harmful event, then come Tstm Wind, Flood, Lightning and the rest. We also see that there are many more types of events that are not shown on the plot and among all these 985 types of events Tornado causes a gigentic harm to population health.

Now let's move on to the second question.

Across the United States, which types of events have the greatest economic consequences?

To answer this question we will make a single plot with top 10 harmful event types that have the greatest economic consequences. We do it in three steps.

- Creating the initial data frame called **economic_df**
- Processing **economic_df** data frame
- Making the plot

Creating the initial data frame called economic_df

We need 6 columns (EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP, REMARKS).

```
library(dplyr)

economic_df <- storm_data %>%
        select(EVTYPE, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP, REMARKS)
dim(economic_df)
```

```
## [1] 902297      6
```

```
head(economic_df, 3)
```

```
##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP REMARKS
## 1 TORNADO    25.0          K       0
## 2 TORNADO     2.5          K       0
## 3 TORNADO    25.0          K       0
```

```
str(economic_df)
```

```
## 'data.frame':    902297 obs. of  6 variables:
##  $ EVTYPE    : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 834 834 834 834 83
4 834 834 834 834 834 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 17 17 17 17 17 17 17 17 17 1
7 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REMARKS   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1 ...
```

Processing economic_df data frame

The two columns **PROPDMGEXP** and **CROPDMGEXP** that we have selected are the alphabetical characters signifying the magnitudes of the numbers in **PROPDMG** and **CROPDMG**. “Alphabetical characters used to signify magnitude include **“K”** for thousands, **“M”** for millions, and **“B”** for billions” as explained in section 2.7 of Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf). But we notice that in these columns there are other values as well. We need to do something about these other values. Let’s deeg a little bit.

1. **Dealing with 0s**
    a. Make **PROPDMGEXP** and **CROPDMGEXP** from factor to character (also change REMARKS column to character as we will need it later).
    b. We can start simplifying **PROPDMGEXP** by first looking at the values of **PROPDMGEXP** when **PROPDMG** is equal to zero. We create a new data frame **economic_1** that has only rows where

PROPDMG == 0. **PROPDMGEXP** column of **economic_1** shows what values correspond to the 0s of **PROPDMG**. We definitely need to change these values to be equal to 0.

c. Wherever **PROPDMG** is 0 in **economic_df** data frame, set the value of **PROPDMGEXP** to 0. Do the same with **CROPDMGEXP** wherever **CROPDMG** is equal to 0.

d. We now hope that we will have fewer rows to worry about. And we are right. Let's check how many rows of **PROPDMGEXP** and **CROPDMGEXP** are left that have values other then "K", "M", "B", and "0".

```
library(dplyr)

# 1
# 1_a
economic_df$PROPDMGEXP <- as.character(economic_df$PROPDMGEXP)
economic_df$CROPDMGEXP <- as.character(economic_df$CROPDMGEXP)
economic_df$REMARKS <- as.character(economic_df$REMARKS)

# 1_b
economic_1 <- economic_df[economic_df$PROPDMG == 0,]
economic_1 %>% group_by(PROPDMGEXP) %>% summarize(count = length(PROPDMGEXP))
```

```
## # A tibble: 12 × 2
##    PROPDMGEXP  count
##         <chr>  <int>
## 1             465858
## 2           ?      8
## 3           0      7
## 4           1     25
## 5           2     12
## 6           3      3
## 7           5     10
## 8           6      1
## 9           7      3
## 10          8      1
## 11          K 197184
## 12          M     11
```

```
# 1_c
economic_df <- within(economic_df, PROPDMGEXP[PROPDMG == 0] <- "0")
economic_df <- within(economic_df, CROPDMGEXP[CROPDMG == 0] <- "0")

# 1_d
economic_2 <- economic_df[grep("[^0KMB]", economic_df$PROPDMGEXP),]
economic_3 <- economic_df[grep("[^0KMB]", economic_df$CROPDMGEXP),]
dim(economic_2)
```

```
## [1] 49  6
```

```
dim(economic_3)
```

```
## [1] 22  6
```

2. **Dealing with non-0s**
    a. First, let's look at the unique values in **PROPDMGEXP** and **CROPDMGEXP** that are not "K", "M", "B" or "0", and thus need to be changed. We see **11** values of **PROPDMGEXP** and **2** values of **CROPDMGEXP** that need our attention.
    b. **PROPDMGEXP** - Let's explore "m" as it may indicate million entered in lowercas instead of an uppercase "M". Then we will look at the rest of the values one by one. We will find **REMARKS** column handy for this task. Also, we can read the **Appendix B** of Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) to get the idea of property damage estimates.
        - **"m"** - By reading the remarks column and Appendix B, we can definitely say that "m" stands for millions. Change "m" of **PROPDMGEXP** with "M" in **economic_df**.
        - **"H" and "h"** - Maybe they stand for hundreds? Again, let's read remarks. They really seem to stand for hundreds. Change "H" and "h" of **PROPDMGEXP** to "100" in economic_df.
        - To change **"7"**, **"6"**, **"5"**, **"4"**, **"3"**, **"2"**, **"+"**, **"-"** values, we need to do a lot of guessing. So we will dare to ignore these values (hopefully, we will not loose a lot of information as there are only few rows with these values). In other words, we will replace these values with 0s. To do all replacements with one function call we will create a function called **mgsub** ("multiple gsub""") as done in the post on Stackoveflow (link to the post on Stackoverflow (https://stackoverflow.com/questions/15253954/replace-multiple-arguments-with-gsub)).
    c. **CROPDMGEXP** - We now look at the values of **CROPDMGEXP** and do the necessary replacements. Again, we use **REMARKS** column. Looks like "m" stands for million and "k" stands for thousand. In case of "k", **REMARKS** column has many empty entries but from the property damage columns we can guess that it is a damage to the crop in thousands. So we will replace "k" and "m" of **CROPDMGEXP** with "K" and "M" in **economic_df** data frame.
    d. We are almost finished with processing columns **PROPDMGEXP** and **CROPDMGEXP**. Let's look at the values that each column has and their counts. We see emtpy entries. These empty entries are the last item we are going to worry about. We'll look at columns **PROPDMG** and **REMARKS** when **PROPDMGEXP** is empty, as well as **CROPDMG** and **REMARKD** when **CROPDMGEXP** is empty. We see that **REMARKS** column has descriptions of damages. So the least we can do is to change these empty entries in **PROPDMGEXP** and **CROPDMGEXP** with "K" to make these damages appear in our analysis. And that may be the exact amount of the damage but we will be careful, anyway, and ignore the rows where no remarks are made. In short, we change empty entries with "K"s when there are remarks and with "0"s when remarks are missing (missing remarks are noted by two spaces: " ").
    e. We will have a final look at the values in **PROPDMGEXP** and **CROPDMGEXP** to make sure we

are only left with “K”, “M”, “B”, “0” and “100”. We can now change “K” to “1000”, “M” to “1000000”, “B” to “1000000000” and convert these columns from character to numeric.

```
library(dplyr)

# 2
# 2_a
unique(economic_2$PROPDMGEXP)
```

```
##  [1] "m" "+" "5" "6" "4" "h" "2" "7" "3" "H" "-"
```

```
unique(economic_3$CROPDMGEXP)
```

```
## [1] "m" "k"
```

```
# 2_b
economic_2 %>% group_by(PROPDMGEXP) %>% summarise(count = length(PROPDMGEXP))
```

```
## # A tibble: 11 × 2
##    PROPDMGEXP count
##         <chr> <int>
## 1           -     1
## 2           +     5
## 3           2     1
## 4           3     1
## 5           4     4
## 6           5    18
## 7           6     3
## 8           7     2
## 9           h     1
## 10          H     6
## 11          m     7
```

```
economic_2[economic_2$PROPDMGEXP == "m",]
```

```
##                    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 187584     HURRICANE OPAL    20.0          m      10          m
## 195369            TORNADO     0.5          m       0          0
## 196712            TORNADO    10.0          m       0          0
## 202891               HAIL     4.9          m       0          0
## 203087 THUNDERSTORM WINDS     2.0          m       0          0
## 207157 THUNDERSTORM WINDS     0.5          m       0          0
```

```
## 211464             TORNADO      1.0           m        0           0
##
REMARKS
## 187584 Hurricane Opal moved northward across south Alabama during the evening and
early morning hours. Widespread minor damage occurred to homes and other structures w
ith sporadic occurrences of more serious damage. Trees, many large, were downed causi
ng damage to vehicles, buildings, as well as power and telephone lines.  Over 90 perc
ent of the region's electric customers lost power.  Serious damage to the area's agri
culture was reported including partial losses of cotton and peanut crops.  Five thous
and chickens were lost in one instance of damage to chicken houses.  While not direct
ly attributable to the storms effects, three people lost their lives in traffic accid
ents and one was killed in a fire caused by candle usage during the power outage. (Da
mage figures are estimated.)
## 195369
A severe thunderstorm caused sportatic damage from downtown Tallahassee  across the e
ast and north sides of the city and spawned a tornado that caused considerable damage
.  Hundreds of mature pines and hardwoods were damaged or blown completely down causi
ng damage to many residences and apartment complexes.  The Tallahassee Civic Center s
ustained considerable roof damage.  No serious injuries were reported but several peo
ple reportedly received cuts and bruises.
## 196712
A tornado first touched down on the north side of the Albany Dougherty County Airport
then moved east north-east for five miles to just past U.S. Highway 19.  In two spots
along the path the tornado lifted briefly.  One of these breaks in the damage path oc
curred at an elementary school that was in session.  The worst damage was to a Winn D
ixie grocery store where many of the 36 injuries occurred.  One side of the store was
pulled out and collapsed to the side parking area.  In the parking lot, numerous cars
were moved about and a few turned over.  Over 40 homes were destroyed with over 50 mo
re severely damaged.  Five businesses were severely damaged.
## 202891
Numerous houses and automobiles damaged or destroyed from baseball to softball size h
ail.
## 203087
Rear flank downdraft damage over five miles wide and 12 miles long.  Wind speeds esti
mated to be over 100 mph.  Over 150 power poles were snapped off and 40 Pivot Sprinkl
ers were damaged or destroyed, along with trees and fences.
## 207157
Part of gynasium roof was blown off Northside High School. Several windows were blown
out and one wall collapsed. Several homes in the area also received minor damage such
as shingles torn off roofs, and trees blown down on sheds.
## 211464
The tornado touched down 5 miles northwest of Jackson and moved east-northeast, and e
xited the county, 7 miles northeast of Jackson at 2055CST. The tornado then moved acr
oss Rankin County before lifting 1 mile northeast of Fannin. Numerous homes were dama
ged, many trees and power lines were blown down, and several small buildings were des
troyed.
```

```
economic_2[economic_2$PROPDMGEXP == "H",]
```

```
##                    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 216476 THUNDERSTORM WINDS       5          H       0          0
## 232397 THUNDERSTORM WINDS       5          H       0          0
## 232398 THUNDERSTORM WINDS       2          H       0          0
## 232735 THUNDERSTORM WINDS       3          H       0          0
## 248019 THUNDERSTORM WINDS       5          H       0          0
## 248038               HAIL       5          H       0          0
##
REMARKS
## 216476
Thunderstorm winds downed a tree in Loomis and blew down large tree limbs at Hazard.
Just west of Loomis, thunderstorm winds damaged a grain bin, a barn and another outbu
ilding and also flipped over a hay trailer.  Heavy rain and small hail combined with
the wind to damage the corn crop.
## 232397
A line of severe thunderstorms with a "bow echo" appearance moved swiftly through the
Upstate during the morning hours.  Highest winds with the bow were around the Georgia
border in Oconee County and one small tornado was generated west of Fair Play.  Large
trees were uprooted and pines were snapped by winds about 85 knots along a path 30 ya
rds wide and one mile long.  Gust front winds in the same area accounted for some dam
age.  A roof of an addition to a mobile home was removed in Walhalla and trees fell i
n many locations.  The wet ground contributed to some of the tree damage.  In Clemso
n a tree fell through a residence.  Other trees brought down power lines across the a
rea and power was out to several thousand customers well into the evening.  In Simpso
nville tree limbs were punched through the roof of a residence, the limbs were carrie
d for some distance since there was no trees near the house.  Power outages were quit
e widespread across western parts of Greenville County although Donaldson Center Air
Park only recorded gusts to 41 knots.
## 232398 A strong cold front moving through the area generated high winds in the mou
ntainous section of Greenville County downing a number of trees and power lines.  Pow
er remained out until the next morning in some areas.  A line of thunderstorms develo
ped along the front.  A very small tornado (possibly a "gustnado") was spotted by air
traffic controllers six miles southeast of the Greenville Downtown Airport.  A weak t
ornado touched down in a wooded area with only tree damage reported.  Some siding was
blown off a house in Taylors and significant thunderstorm winds occurred with the sam
e part of the line at Lake Robinson north of Greer.  A spotter reported wind gusts to
80 mph with a number of trees downed.  Weaker wind damage was reported around Landrum
where a few trees were downed and a railroad sign was broken off. A stronger tornado
(strong F0/weak F1) was generated southwest of Greenwood from a weakly rotating thund
erstorm.  The tornado touched down along Highway 10 destroying one mobile home and da
maging several others - along with some homes.  The tornado lifted after a one-mile t
rack otherwise the storm could have moved near downtown Greenwood.  The tornado track
was almost due north while the parent thunderstorm moved northeast.  The thunderstorm
rapidly collapsed producing strong damaging winds in an area northeast of town - from
```

```
near Coronaca over to Lake Greenwood.  A couple of carports were torn from houses, a
number of trees were downed, and power lines were taken down.  A home under construct
ion was severely damaged at the lake.  Across the lake in Laurens County, the only ev
idence of damage was large tree limbs that were blown out.
## 232735
## 248019
Strong thunderstorm winds knocked down a tree, many tree limbs and a flagpole in Grey
bull.  One and three quarter inch diameter hail fell in Greybull.
## 248038
One inch diameter hail fell in Recluse and in Gillette.  The hail damaged trees and g
ardens in Gillette.  One outdoor booth at a rodeo in Gillette was demolished and some
trailers were damaged.  At the same rodeo, 2 people sustained minor injuries due to t
he hail.  The hail was in drifts 6 to 8 inches deep in Gillette.
```

```
economic_2[economic_2$PROPDMGEXP == "H",]$REMARKS
```

```
## [1] "Thunderstorm winds downed a tree in Loomis and blew down large tree limbs at
Hazard.  Just west of Loomis, thunderstorm winds damaged a grain bin, a barn and anot
her outbuilding and also flipped over a hay trailer.  Heavy rain and small hail combi
ned with the wind to damage the corn crop. "
## [2] "A line of severe thunderstorms with a \"bow echo\" appearance moved swiftly t
hrough the Upstate during the morning hours.  Highest winds with the bow were around
the Georgia border in Oconee County and one small tornado was generated west of Fair
Play.  Large trees were uprooted and pines were snapped by winds about 85 knots along
a path 30 yards wide and one mile long.  Gust front winds in the same area accounted
for some damage.  A roof of an addition to a mobile home was removed in Walhalla and
trees fell in many locations.   The wet ground contributed to some of the tree damage
.  In Clemson a tree fell through a residence.  Other trees brought down power lines
across the area and power was out to several thousand customers well into the evening
.  In Simpsonville tree limbs were punched through the roof of a residence, the limbs
were carried for some distance since there was no trees near the house.  Power outage
s were quite widespread across western parts of Greenville County although Donaldson
Center Air Park only recorded gusts to 41 knots. "
## [3] "A strong cold front moving through the area generated high winds in the mount
ainous section of Greenville County downing a number of trees and power lines.  Power
remained out until the next morning in some areas.  A line of thunderstorms developed
along the front.  A very small tornado (possibly a \"gustnado\") was spotted by air t
raffic controllers six miles southeast of the Greenville Downtown Airport.  A weak to
rnado touched down in a wooded area with only tree damage reported.  Some siding was
blown off a house in Taylors and significant thunderstorm winds occurred with the sam
e part of the line at Lake Robinson north of Greer.  A spotter reported wind gusts to
80 mph with a number of trees downed.  Weaker wind damage was reported around Landrum
where a few trees were downed and a railroad sign was broken off. A stronger tornado
(strong F0/weak F1) was generated southwest of Greenwood from a weakly rotating thund
erstorm.  The tornado touched down along Highway 10 destroying one mobile home and da
maging several others - along with some homes.  The tornado lifted after a one-mile t
rack otherwise the storm could have moved near downtown Greenwood.  The tornado track
was almost due north while the parent thunderstorm moved northeast.  The thunderstorm
rapidly collapsed producing strong damaging winds in an area northeast of town - from
near Coronaca over to Lake Greenwood.  A couple of carports were torn from houses, a
number of trees were downed, and power lines were taken down.  A home under construct
ion was severely damaged at the lake.  Across the lake in Laurens County, the only ev
idence of damage was large tree limbs that were blown out. "
## [4] "  "
## [5] "Strong thunderstorm winds knocked down a tree, many tree limbs and a flagpole
in Greybull.  One and three quarter inch diameter hail fell in Greybull. "
## [6] "One inch diameter hail fell in Recluse and in Gillette.  The hail damaged tre
es and gardens in Gillette.  One outdoor booth at a rodeo in Gillette was demolished
and some trailers were damaged.  At the same rodeo, 2 people sustained minor injuries
due to the hail.  The hail was in drifts 6 to 8 inches deep in Gillette. "
```

```
economic_2[economic_2$PROPDMGEXP == "h",]
```

```
##                       EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 209285 THUNDERSTORM WIND G50       2          h       0          0
##
REMARKS
## 209285 Wind damage reported at Parks Rd. and U.S. Highway 27.  Trees downed and wi
ndow of a home blown out.
```

```
economic_df$PROPDMGEXP <- gsub("m", "M", economic_df$PROPDMGEXP)
economic_df$PROPDMGEXP <- gsub("H", "100", economic_df$PROPDMGEXP)
economic_df$PROPDMGEXP <- gsub("h", "100", economic_df$PROPDMGEXP)

mgsub <- function(pattern, replacement, x, ...) {
  if (length(pattern)!=length(replacement)) {
    stop("pattern and replacement do not have the same length.")
  }
  result <- x
  for (i in 1:length(pattern)) {
    result <- gsub(pattern[i], replacement[i], result, ...)
  }
  result
}

economic_df$PROPDMGEXP <- mgsub(c("2", "3", "4", "5", "6", "7", "\\+", "\\-"),
                                c("0", "0", "0", "0", "0", "0", "0", "0"),
                                economic_df$PROPDMGEXP)

# 2_c
economic_3[economic_3$CROPDMGEXP == "m",]
```

```
##                EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 187584 HURRICANE OPAL      20          m      10          m
##
REMARKS
## 187584 Hurricane Opal moved northward across south Alabama during the evening and
early morning hours. Widespread minor damage occurred to homes and other structures w
ith sporadic occurrences of more serious damage. Trees, many large, were downed causi
ng damage to vehicles, buildings, as well as power and telephone lines.  Over 90 perc
ent of the region's electric customers lost power.  Serious damage to the area's agri
culture was reported including partial losses of cotton and peanut crops.  Five thous
and chickens were lost in one instance of damage to chicken houses.  While not direct
ly attributable to the storms effects, three people lost their lives in traffic accid
ents and one was killed in a fire caused by candle usage during the power outage. (Da
mage figures are estimated.)
```

```
economic_3[economic_3$CROPDMGEXP == "k",]
```

```
##                   EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 195667     FLASH FLOODING       2          K      10          k
## 200542 THUNDERSTORM WINDS      50          K       1          k
## 200543 THUNDERSTORM WINDS      50          K       1          k
## 200618               HAIL       2          K       1          k
## 200632               HAIL       1          K       3          k
## 200880               HAIL       3          K       3          k
## 200881               HAIL       3          K       1          k
## 200882               HAIL      20          K      15          k
## 200883               HAIL      10          K      15          k
## 200884               HAIL       3          K       1          k
## 201144               HAIL       3          K       1          k
## 201145 THUNDERSTORM WINDS       5          K       1          k
## 201674               HAIL       3          K       1          k
## 201675               HAIL      50          K      10          k
## 201676               HAIL       3          K       1          k
## 201677               HAIL      20          K       5          k
## 201977               HAIL      15          K       5          k
## 201978               HAIL      10          K       5          k
## 202101 THUNDERSTORM WINDS      25          K       1          k
## 202257               HAIL      50          K     350          k
## 202262        WIND DAMAGE      10          K       5          k
##
REMARKS
## 195667 Fifteen roads in the north part of the county, generally north of Interstat
e 10, were closed due to high water.  Highway 85, north of Interstate 10, had a foot
of water in it and had to be closed for a few hours.
## 200542
```

```
## 200543
## 200618
## 200632
## 200880
## 200881
## 200882
## 200883
## 200884
## 201144
## 201145
## 201674
## 201675
## 201676
## 201677
## 201977
## 201978
## 202101
## 202257
## 202262
Trees and power lines down, some damage to homes.
```

```
economic_df$CROPDMGEXP <- gsub("m", "M", economic_df$CROPDMGEXP)
economic_df$CROPDMGEXP <- gsub("k", "K", economic_df$CROPDMGEXP)

# 2_d
economic_df %>% group_by(PROPDMGEXP) %>% summarize(count = length(PROPDMGEXP))
```

```
## # A tibble: 6 × 2
##   PROPDMGEXP  count
##        <chr>  <int>
## 1                76
## 2          0 663367
## 3        100      7
## 4          B     40
## 5          K 227481
## 6          M  11326
```

```
economic_df %>% group_by(CROPDMGEXP) %>% summarize(count = length(CROPDMGEXP))
```

```
## # A tibble: 5 × 2
##   CROPDMGEXP  count
##        <chr>  <int>
## 1                 3
## 2          0 880210
## 3          B      7
## 4          K  20158
## 5          M   1919
```

```
cbind(economic_df[economic_df$PROPDMGEXP == "",]$PROPDMG,
      economic_df[economic_df$PROPDMGEXP == "",]$REMARKS)
```

```
##       [,1]
##  [1,] "0.41"
##  [2,] "3"
##  [3,] "2"
##  [4,] "4"
##  [5,] "4"
##  [6,] "10"
##  [7,] "10"
##  [8,] "10"
##  [9,] "4"
## [10,] "5"
## [11,] "10"
## [12,] "35"
## [13,] "75"
## [14,] "3"
## [15,] "10"
## [16,] "1"
## [17,] "3"
## [18,] "20"
## [19,] "2"
## [20,] "20"
## [21,] "10"
## [22,] "1"
## [23,] "20"
## [24,] "5"
## [25,] "4"
## [26,] "5"
## [27,] "4"
## [28,] "6"
## [29,] "7"
## [30,] "7"
## [31,] "10"
## [32,] "9"
```

```
## [33,] "3"
## [34,] "2"
## [35,] "8"
## [36,] "8"
## [37,] "6"
## [38,] "3"
## [39,] "2"
## [40,] "7"
## [41,] "4"
## [42,] "1"
## [43,] "3"
## [44,] "5"
## [45,] "6"
## [46,] "5"
## [47,] "3"
## [48,] "1"
## [49,] "10"
## [50,] "3"
## [51,] "5"
## [52,] "3"
## [53,] "5"
## [54,] "2"
## [55,] "3"
## [56,] "9"
## [57,] "4"
## [58,] "3"
## [59,] "5"
## [60,] "6"
## [61,] "3"
## [62,] "3"
## [63,] "6"
## [64,] "3"
## [65,] "9"
## [66,] "2"
## [67,] "4"
## [68,] "5"
## [69,] "5"
## [70,] "2"
## [71,] "4"
## [72,] "5"
## [73,] "3"
## [74,] "6"
## [75,] "20"
## [76,] "3"
##       [,2]
##  [1,] "Wind gusts to 96 mph Mt Tamalpias and 89 mph at the Golden Gate Bridge Peta
luma river at Petuluma went 1.6 feet over flood stage. "
```

```
##  [2,] "A small tornado touched down at the North Florida Prison Reception Center d
amaging the building before dissipating. "
##  [3,] "The sheriff's office reported numerous power lines and trees were down.  "
##  [4,] "Thunderstorm winds produced minor damage to five homes and ripped off a tin
roof from a barn in Hartwell.      ~ Hart County   1 E Hartwell,31,2032EST,0.5,30,0,0
,3,?,Tornado (F0) \nA short lived tornado touched down and toppled two large trees an
d blew over a farm shed.  Two large columns that supported the front porch of a house
were blown away. "
##  [5,] "Several large limbs were downed on U.S. Route 24 west of Logansport. Four v
ehicles were damaged, and the highway was blocked. "
##  [6,] "  "
##  [7,] "  "
##  [8,] "  "
##  [9,] "Trees downed, some across Route 525 West. "
## [10,] "High winds accompanied by large hail knocked down power lines. "
## [11,] "A severe thunderstorm produced golf ball- to baseball-size hail from Highla
nd to just north of Denton.  An apparant tornado formed two miles north of Denton and
tracked east for 12 miles to near Wathena.  Damage was intermittant along the narrow
path, which averaged about 100-yards in width.  Roofs were blown off of several barns
and outbuildings. (F0) "
## [12,] "A severe thunderstorm produced golf ball- to baseball-size hail from Highla
nd to just north of Denton.  An apparant tornado formed two miles north of Denton and
tracked east for 12 miles to near Wathena.  Damage was intermittant along the narrow
path, which averaged about 100-yards in width.  Roofs were blown off of several barns
and outbuildings. (F0) "
## [13,] "Heavy rains in a short period of time produced widespread street and low ly
ing flooding in and around the city of Topeka. Several city streets were under water
and closed for an hour or two during the morning rush hour. "
## [14,] "Thunderstorm winds blew siding off a house in the vicinity of Hopkinsville.
"
## [15,] "  "
## [16,] "Power lines downed. "
## [17,] "A spotter reported several large trees uprooted due to thunderstorm winds.
"
## [18,] "A maintenance building collapsed and only had one wall left standing. "
## [19,] "Several trees were blown down. "
## [20,] "  "
## [21,] "  "
## [22,] "Trees were downed, one on an automobile. \nWarren County   Springboro,14,00
33EST,,,0,0,0,0,Hail (0.75) "
## [23,] "Thunderstorm rains produced flooding of streets, poor drainage areas and ba
sements. A rainfall of 3.5 inches was measured in less than two hours. "
## [24,] "Two inches of rain fell in 90 minutes on top of saturated ground causing fl
ooding of several roads, especially in Middletown.  Low lying areas and poor drainage
areas also flooded. "
## [25,] "  "
## [26,] "Nearly three inches of rain fell in 90 minutes resulting in flooding of str
```

```
eets, streams, and poor drainage areas.  In Wayne Township, a bridge was water covere
d for a short time.  High water was reported across State Route 4, south of Middletow
n. "
## [27,] "Numerous trees downed, some taking down power lines. "
## [28,] "Trees were downed in several locations.  Three miles south of Springfield,
winds were measured at 65 mph.  60 mph winds were also measured at Charleston.  Golf
ball-size hail damaged crops. "
## [29,] "Thunderstorms brought one to two inches of rain, causing Stonelick creek to
run out of its banks.  Some county road and poor drainage areas also flooded. "
## [30,] "Heavy rains on top of saturated ground caused flooding of streets and poor
drainage areas.  Several roads were reported to have been water covered. "
## [31,] "A series of thunderstorms dropped two to five inches of rain in nearly six
hours on saturated ground.  This produced some flooding of roads, streams, and baseme
nts.  Five roads in Versailles and Ansonia were closed due to high water.  Standing w
ater remained over a few roads into the evening hours. "
## [32,] "Numerous trees were downed.  In Shawnee Hills, a large tent was toppled. "
## [33,] "Limbs and utility pole downed. "
## [34,] "Trees were downed. "
## [35,] "Trees were downed at several locations. "
## [36,] "Thunderstorm rains brought up to two inches of rain in one hour causing flo
oding of streams and streets.  Water as high as four feet was reported across some ro
ads.  Black Lick Creek overflowed its banks onto nearby roads. "
## [37,] "Thunderstorm winds downed large trees and power lines. "
## [38,] "Additional heavy rains of three inches in six hours caused more flooding of
roads, streams and poor drainage areas.  Blacklick Creek near Reynoldsburg overflowed
its banks. "
## [39,] "Large limbs were downed. "
## [40,] "Several large trees were blown down. "
## [41,] "Trees were downed in southern Cincinnati, while three-quarters inch hail wa
s also reported. "
## [42,] "Several large trees were downed, some across roads. "
## [43,] "Several large trees were downed, some across roads. "
## [44,] "Trees and limbs were downed in several locations including Wakeman. "
## [45,] "About a half dozen trees were downed. "
## [46,] "Trees and limbs were downed, some on power lines. "
## [47,] "Thunderstorms dropped around two inches of rain causing flooding of streets
, streams, and poor drainage areas.  Water was reported over U.S. Highway 40 near Etn
a for a short time. "
## [48,] "Thunderstorm winds toppled two large trees. "
## [49,] "Large hail was reported in Boardman and trees were downed in Youngstown. "
## [50,] "Trees were downed. "
## [51,] "Trees were downed in several locations.  Seven miles southwest of Celina, t
wo mobile homes were destroyed by fallen trees, and another mobile home was seriously
damaged.  Fallen trees also crushed a car at the same location. "
## [52,] "Heavy rains brought nearly two inches of rain in less than three hours.  Th
is caused minor street flooding and flooding of low lying areas. High water was repor
ted across some low lying roads in Piqua.    "
```

```
## [53,] "Another round of thunderstorms brought an additional one to three inches of
rain in less than two hours causing additional flooding of roads and streams.  Six mo
re roads had to be closed due to high water, including State Routes 202 and 571. "
## [54,] "A few large trees were downed, some into power lines.  "
## [55,] "Strong winds behind a cold front downed trees, some into power lines.  Some
trees also fell across roadways. "
## [56,] "Trees were downed in numerous locations. "
## [57,] "The Scioto river at Piketon exceeded flood stage of 16 feet and crested at
19 feet at 1800 EST.  Flooding was confined to low lying areas. "
## [58,] "Several trees were reported downed by winds estimated to be 60 mph. "
## [59,] "Heavy rains produced one and one-half to two inches of rain in less than tw
o hours.  Water was reported running across State Routes 109 and 613 near Leipsic and
West Leipsic. "
## [60,] "Numerous trees were downed, some knocked down power lines. "
## [61,] "The Paint Creek at Bourneville exceeded flood stage of 10 feet and crested
at 12.75 feet at 1030 EST.  Some farmlands experienced minor flooding. "
## [62,] "Several trees were downed.   \nClermont County  Countywide,25,0945EST- *,25
,1200EST,,,0,0,6K,0,Flash Flood \nHeavy rains of one and one-half to two inches on to
p of saturated ground produced flooding of streets, small streams, and basements.  15
roads were closed due to high water, including State Route 222. "
## [63,] "Power lines and trees were downed. "
## [64,] "Trees were downed, some across State Route 139. "
## [65,] "Numerous trees and power lines were blown down, some across roads.  Near Si
dney, a steel shed was blown off its base and thrown into a nearby corn field. "
## [66,] "Thunderstorm rains of two to three inches in less than two hours caused flo
oding of some streets and low lying areas.  A few roads in Sidney had high water acro
ss them for a short period of time. "
## [67,] "A wind gust to 58 mph was measured at Dixon, and three-quarters inch hail w
as reported at Ohio City. "
## [68,] "Thunderstorm rains produced flooding of streets and poor drainage areas. "
## [69,] "  "
## [70,] "  "
## [71,] "  "
## [72,] "Several trees and power lines were knocked down. "
## [73,] "  "
## [74,] "  "
## [75,] "  "
## [76,] "  "
```

```
cbind(economic_df[economic_df$CROPDMGEXP == "",]$CROPDMG,
      economic_df[economic_df$CROPDMGEXP == "",]$REMARKS)
```

```
##      [,1]
## [1,] "3"
## [2,] "4"
## [3,] "4"
##      [,2]
## [1,] "  "
## [2,] "Thunderstorms produced widespread large hail, accompanied by damaging winds
that knocked down tree limbs, stripped leaves from trees and knocked out power and te
lephone communications to San Marcos for several hours.  The hailstones broke windows
in homes and school as well as Southwest State University. "
## [3,] "Thunderstorms moving eastward through Medina County produced widespread wind
damage.  The Red Cross reported that 32 homes were destroyed, 44 had major damage, an
d 194 homes had minor damage.  Numerous mobile homes suffered roof and wall damage.
The city of Castroville was without power from 1639 CST until 0530 CST the next morni
ng.  Very heavy rain accompanied the storms, reducing visibility to near zero.  Some
of the residents reported a dark green color to the clouds just before the storm stuc
k.  Although no large hail was reported, the hail piled into drifts along the side of
the road just west of D'Hanis.  Very little damage was reported at the Castroville Ai
rport as most aircraft were tied down or put away at the time of the storms. "
```

```
economic_df <- within(economic_df,
                      PROPDMGEXP[PROPDMGEXP == ""][REMARKS[PROPDMGEXP == ""] == "  "]
<- "0")
economic_df <- within(economic_df,
                      CROPDMGEXP[CROPDMGEXP == ""][REMARKS[CROPDMGEXP == ""] == "  "]
<- "0")

economic_df <- within(economic_df,
                      PROPDMGEXP[PROPDMGEXP == ""][REMARKS[PROPDMGEXP == ""] != "  "]
<- "K")
economic_df <- within(economic_df,
                      CROPDMGEXP[CROPDMGEXP == ""][REMARKS[CROPDMGEXP == ""] != "  "]
<- "K")

# 2_e
economic_df %>% group_by(PROPDMGEXP) %>% summarize(count = length(PROPDMGEXP))
```

```
## # A tibble: 5 × 2
##   PROPDMGEXP  count
##        <chr>  <int>
## 1          0 663381
## 2        100      7
## 3          B     40
## 4          K 227543
## 5          M  11326
```

```
economic_df %>% group_by(CROPDMGEXP) %>% summarize(count = length(CROPDMGEXP))
```

```
## # A tibble: 4 × 2
##   CROPDMGEXP  count
##        <chr>  <int>
## 1          0 880211
## 2          B      7
## 3          K  20160
## 4          M   1919
```

```
economic_df$PROPDMGEXP <- as.numeric(mgsub(c("K", "M", "B"),
                                c("1000", "1000000", "1000000000"),
                                economic_df$PROPDMGEXP))
economic_df$CROPDMGEXP <- as.numeric(mgsub(c("K", "M", "B"),
                                c("1000", "1000000", "1000000000"),
                                economic_df$CROPDMGEXP))
```

Making the plot

1. Create a new data frame called **economic_plot_df** which will have combined **PROPDMG** and **CROPDMG** columns (first we will multiply **PROPDMG** with **PROPDMGEXP** and **CROPDMG** with **CROPDMGEXP**.
2. Count total number of event types and total number of damage to property and crop.
3. Create **PERCENT** and **PERCENT_CHAR** columns to make our plot more informative.
4. Make the plot

```
library(dplyr)

# 1
economic_plot_df <- economic_df %>%
        group_by(EVTYPE) %>%
        summarise(PROP = sum(PROPDMG * PROPDMGEXP), CROP = sum(CROPDMG * CROPDMGEXP))
%>%
        mutate(PROP_CROP_DMG = PROP + CROP) %>%
        arrange(desc(PROP_CROP_DMG))
dim(economic_plot_df)
```

```
## [1] 985   4
```

```
head(economic_plot_df, 3)
```

```
## # A tibble: 3 × 4
##               EVTYPE         PROP       CROP PROP_CROP_DMG
##               <fctr>        <dbl>      <dbl>         <dbl>
## 1              FLOOD 144657716800 5661968450  150319685250
## 2 HURRICANE/TYPHOON  69305840000 2607872800   71913712800
## 3            TORNADO  56937163480  414953110   57352116590
```

```
# 2
count_event_types <- nrow(economic_plot_df)
sum_prop_crop <- sum(economic_plot_df$PROP_CROP_DMG)

# 3
economic_plot_df <- economic_plot_df %>%
        mutate(PERCENT = PROP_CROP_DMG * 100/sum_prop_crop) %>%
        mutate(PERCENT_CHAR = paste(as.character(round(PERCENT, digits = 2)),"%"))
dim(economic_plot_df)
```

```
## [1] 985   6
```

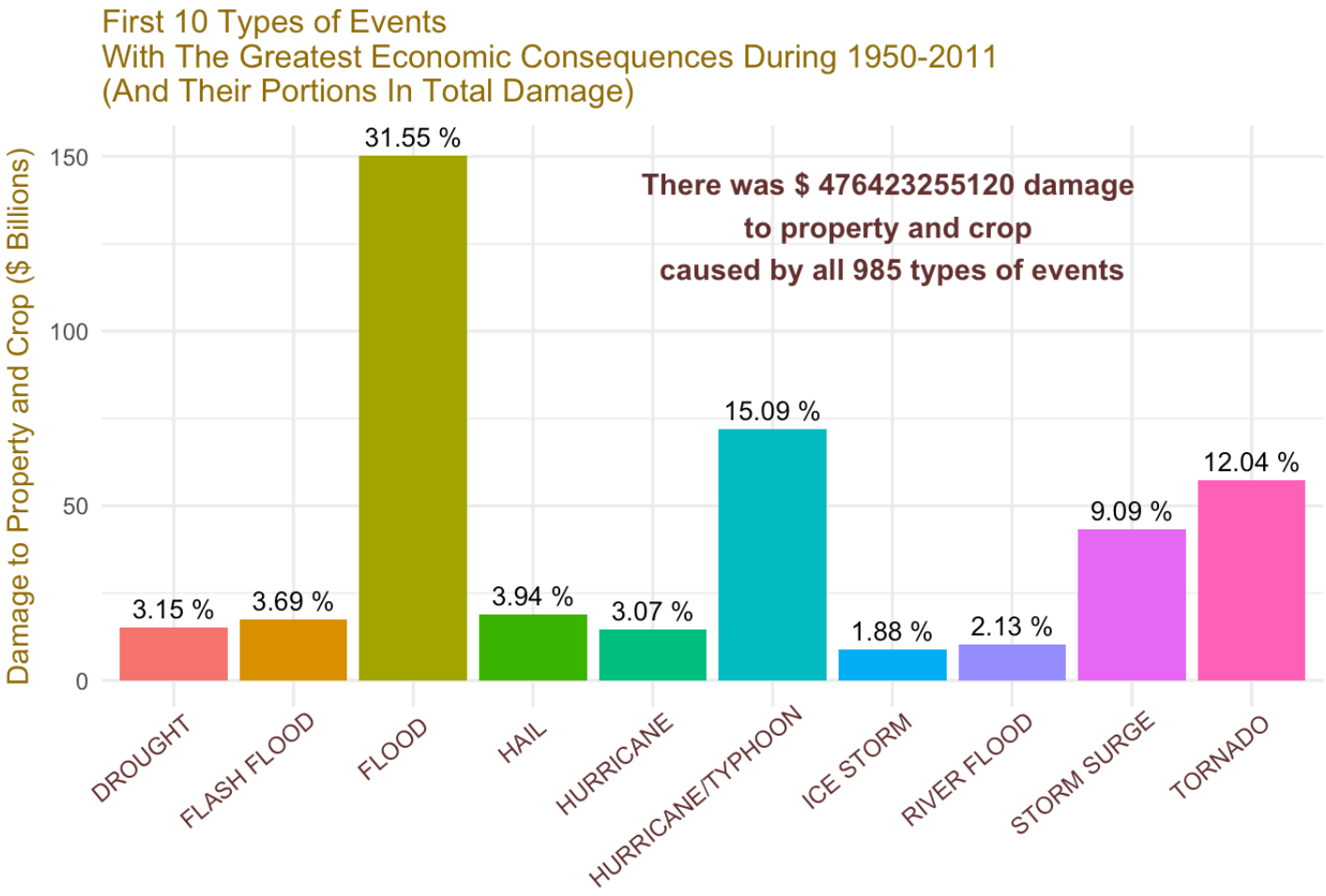
```
head(economic_plot_df, 3)
```

```
## # A tibble: 3 × 6
##              EVTYPE         PROP       CROP PROP_CROP_DMG  PERCENT
##              <fctr>        <dbl>      <dbl>         <dbl>    <dbl>
## 1             FLOOD 144657716800 5661968450  150319685250 31.55171
## 2 HURRICANE/TYPHOON  69305840000 2607872800   71913712800 15.09450
## 3           TORNADO  56937163480  414953110   57352116590 12.03806
## # ... with 1 more variables: PERCENT_CHAR <chr>
```

And we can make our plot.

```
library(ggplot2)

# 4 plot
div = 1000000000
ggplot(economic_plot_df[1:10,], aes(x = EVTYPE, y = PROP_CROP_DMG/div, fill = EVTYPE)
) +
        geom_bar(stat = "identity") +
        geom_text(aes(label = PERCENT_CHAR), vjust = -0.5, size = 3.5) +
        labs(title = "First 10 Types of Events\nWith The Greatest Economic Consequenc
es During 1950-2011\n(And Their Portions In Total Damage)") +
        labs(x = "") +
        labs(y = "Damage to Property and Crop ($ Billions)") +
        theme_minimal() +
        theme(legend.position = "none") +
        theme(axis.text.x = element_text(angle = 40, hjust = 0.9, vjust = 1, size = 9
, color = rgb(0.4,0.2,0.2))) +
        theme(axis.title.y = element_text(color = colorRampPalette(c("green", "red"))
(15)[9])) +
        theme(plot.title = element_text(color = colorRampPalette(c("green", "red"))(1
5)[9], size = 12)) +
        scale_y_continuous(limits = c(min(economic_plot_df$PROP_CROP_DMG),
                                      max(economic_plot_df$PROP_CROP_DMG)+1000000000)
/div) +
        annotate("text", x = 7, y = 130, color = rgb(0.4,0.2,0.2), fontface = 2,
                 label = paste("There was $", as.character(sum_prop_crop),
                               "damage \nto property and crop \ncaused by all",
                                as.character(count_event_types), "types of events"))
```



From the plot above we see that Flood had the greates economic consequences across the United States from 1950 to 2011. Then come Hurricane/Typhoon, Tornado, Storm Surge, and the rest.