Storm data analysis

Yaroslav Tsaregorodtsev

3/31/2020

Librarys loading

```
library(dplyr)
library(data.table)
library(readr)
library(ggplot2)
library(stringr)
library(stringi)
```

Storm Data description

Storm Data is an official publication of the National Oceanic and Atmospheric Administration (NOAA) which documents:

a. The occurrence of storms and other significant weather phenomena having sufficient intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce;

b. Rare, unusual, weather phenomena that generate media attention, such as snow flurries in South Florida or the San Diego coastal area; and

c. Other significant meteorological events, such as record maximum or minimum temperatures or precipitation that occur in connection with another event.

Data Processing

```
#Loadind Storm data
data_path <- list.files(pattern = "repdata_data_StormData.csv",
                        recursive = TRUE, ignore.case = TRUE)
data <- read_csv(data_path,
                 col_types = cols(BGN_DATE = col_date(format = "%m/%d/%Y %H:%M:%S"),
                                  WFO = col_character(),
                                  END_TIME = col_character(),
                                  BGN_AZI = col_character(),
                                  BGN_LOCATI = col_character(),
                                  END_DATE = col_character(),
                                  CROPDMGEXP = col_character(),
                                  ZONENAMES = col_character(),
                                  REMARKS = col_character(),
```

```
                                  END_LOCATI = col_character(),
                                  END_AZI = col_character(),
                                  STATEOFFIC = col_character())) %>%
  data.table()

#Loadind event table
event_table_path <- list.files(pattern = "event_table",
                               recursive = TRUE, ignore.case = TRUE)
event_table <- read_delim(event_table_path, delim = ";") %>%
  data.table()

#Loadind State list (50 States and District of Columbia)
state_list_path <- list.files(pattern = "State_list",
                              recursive = TRUE, ignore.case = TRUE)
state_list <- read_delim(state_list_path, delim = ";") %>%
  data.table()

#Grouping data by main event types and classification for the 3 designators: County/Parish; Zone; and M
storm_data <- data %>%
  mutate(Event_Name = case_when(str_detect(EVTYPE %>% toupper(), "ASTRONOMICAL LOW TIDE") ~ "ASTRONOMIC
                                str_detect(EVTYPE %>% toupper(), "AVALANCHE") ~ "AVALANCHE",
                                str_detect(EVTYPE %>% toupper(), "BLIZZARD") ~ "BLIZZARD",
                                str_detect(EVTYPE %>% toupper(), "COASTAL FLOOD") ~ "COASTAL FLOOD",
                                str_detect(EVTYPE %>% toupper(), "COLD/WIND CHILL") ~ "COLD/WIND CHILL"
                                str_detect(EVTYPE %>% toupper(), "DEBRIS FLOW") ~ "DEBRIS FLOW",
                                str_detect(EVTYPE %>% toupper(), "DENSE FOG") ~ "DENSE FOG",
                                str_detect(EVTYPE %>% toupper(), "DENSE SMOKE") ~ "DENSE SMOKE",
                                str_detect(EVTYPE %>% toupper(), "DROUGHT") ~ "DROUGHT",
                                str_detect(EVTYPE %>% toupper(), "DUST DEVIL") ~ "DUST DEVIL",
                                str_detect(EVTYPE %>% toupper(), "DUST STORM") ~ "DUST STORM",
                                str_detect(EVTYPE %>% toupper(), "EXCESSIVE HEAT") ~ "EXCESSIVE HEAT",
                                str_detect(EVTYPE %>% toupper(), "EXTREME COLD/WIND CHILL") ~ "EXTREME C
                                str_detect(EVTYPE %>% toupper(), "FLASH FLOOD") ~ "FLASH FLOOD",
                                str_detect(EVTYPE %>% toupper(), "FLOOD") ~ "FLOOD",
                                str_detect(EVTYPE %>% toupper(), "FROST/FREEZE") ~ "FROST/FREEZE",
                                str_detect(EVTYPE %>% toupper(), "FUNNEL CLOUD") ~ "FUNNEL CLOUD",
                                str_detect(EVTYPE %>% toupper(), "FREEZING FOG") ~ "FREEZING FOG",
                                str_detect(EVTYPE %>% toupper(), "HAIL") ~ "HAIL",
                                str_detect(EVTYPE %>% toupper(), "HEAT") ~ "HEAT",
                                str_detect(EVTYPE %>% toupper(), "HEAVY RAIN") ~ "HEAVY RAIN",
                                str_detect(EVTYPE %>% toupper(), "HEAVY SNOW") ~ "HEAVY SNOW",
                                str_detect(EVTYPE %>% toupper(), "HIGH SURF") ~ "HIGH SURF",
                                str_detect(EVTYPE %>% toupper(), "HIGH WIND") ~ "HIGH WIND",
                                str_detect(EVTYPE %>% toupper(), "HURRICANE|TYPHOON") ~ "HURRICANE (TYPH
                                str_detect(EVTYPE %>% toupper(), "ICE STORM") ~ "ICE STORM",
                                str_detect(EVTYPE %>% toupper(), "LAKE-EFFECT SNOW") ~ "LAKE-EFFECT SNOW
                                str_detect(EVTYPE %>% toupper(), "LAKESHORE FLOOD") ~ "LAKESHORE FLOOD"
                                str_detect(EVTYPE %>% toupper(), "LIGHTNING") ~ "LIGHTNING",
                                str_detect(EVTYPE %>% toupper(), "MARINE HAIL") ~ "MARINE HAIL",
                                str_detect(EVTYPE %>% toupper(), "MARINE HIGH WIND") ~ "MARINE HIGH WIN
                                str_detect(EVTYPE %>% toupper(), "MARINE STRONG WIND") ~ "MARINE STRONG
                                str_detect(EVTYPE %>% toupper(), "MARINE THUNDERSTORM WIND") ~ "MARINE T
                                str_detect(EVTYPE %>% toupper(), "RIP CURRENT") ~ "RIP CURRENT",
```

```
                                str_detect(EVTYPE %>% toupper(), "SEICHE") ~ "SEICHE",
                                str_detect(EVTYPE %>% toupper(), "SLEET") ~ "SLEET",
                                str_detect(EVTYPE %>% toupper(), "SNOW") ~ "SNOW",
                                str_detect(EVTYPE %>% toupper(), "STORM SURGE/TIDE") ~ "STORM SURGE/TIDE",
                                str_detect(EVTYPE %>% toupper(), "STRONG WIND") ~ "STRONG WIND",
                                str_detect(EVTYPE %>% toupper(), "THUNDERSTORM WIND|TSTM WIND|THUNDERSTORM",
                                str_detect(EVTYPE %>% toupper(), "TORNADO") ~ "TORNADO",
                                str_detect(EVTYPE %>% toupper(), "TROPICAL DEPRESSION") ~ "TROPICAL DEPRESSION",
                                str_detect(EVTYPE %>% toupper(), "TROPICAL STORM") ~ "TROPICAL STORM",
                                str_detect(EVTYPE %>% toupper(), "TSUNAMI") ~ "TSUNAMI",
                                str_detect(EVTYPE %>% toupper(), "VOLCANIC ASH") ~ "VOLCANIC ASH",
                                str_detect(EVTYPE %>% toupper(), "WATERSPOUT") ~ "WATERSPOUT",
                                str_detect(EVTYPE %>% toupper(), "WILDFIRE") ~ "WILDFIRE",
                                str_detect(EVTYPE %>% toupper(), "WINTER STORM") ~ "WINTER STORM",
                                str_detect(EVTYPE %>% toupper(), "WINTER WEATHER") ~ "WINTER WEATHER",
                                TRUE ~ "Other")) %>%
  merge(event_table, by = "Event_Name", all.x = TRUE) %>%
  merge(state_list, by.x = "STATE", by.y = "Code", all.x = TRUE)
```

Analysis

Storm Data ware considered. All state codes converted to state names (states that have a different code from code 50 states + District of Columbia receive empty name). Also, all weather events ware grouped into general types from Storm Data Event Table (see https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf, section 2.1.1) and there are three designators: County/Parish; Zone; and Marine.

Analysis consists of 3 parts:

1. Crop damage and fatalities for all time of observed data and all states.
2. Total level of Crop damage and fatalities for all states.
3. Crop damage and fatalities for all events and designators.

Calculating

```
##calculating total Crop Damage and Fatality for the all States


 damage_state_year <- storm_data %>%
   group_by(BGN_DATE, State, Event_Name) %>%
   summarise(Crop_damage = sum(CROPDMG),
             Fatality = sum(FATALITIES)) %>%
  arrange(desc(Crop_damage))
```

```
## `summarise()` regrouping output by 'BGN_DATE', 'State' (override with `.groups` argument)
```

```
 summary(damage_state_year)
```

```
##     BGN_DATE             State            Event_Name         Crop_damage
##  Min.   :1950-01-03   Length:240766      Length:240766      Min.   :   0.000
##  1st Qu.:1990-08-17   Class :character   Class :character   1st Qu.:   0.000
##  Median :2000-02-14   Mode  :character   Mode  :character   Median :   0.000
##  Mean   :1996-01-04                                         Mean   :   5.723
##  3rd Qu.:2006-07-05                                         3rd Qu.:   0.000
##  Max.   :2011-11-30                                         Max.   :9630.000
##     Fatality
##  Min.   :  0.0000
##  1st Qu.:  0.0000
##  Median :  0.0000
##  Mean   :  0.0629
##  3rd Qu.:  0.0000
##  Max.   :583.0000
```

```
head(damage_state_year, 10)
```

```
## # A tibble: 10 x 5
## # Groups:   BGN_DATE, State [10]
##    BGN_DATE   State       Event_Name        Crop_damage Fatality
##    <date>     <chr>       <chr>                   <dbl>    <dbl>
##  1 2011-04-27 Mississippi TORNADO                 9630        31
##  2 2011-05-01 Arkansas    FLOOD                   7848         5
##  3 2008-05-29 Nebraska    FLASH FLOOD             7750         0
##  4 1998-05-21 Nebraska    HAIL                    7160.        0
##  5 2008-06-16 New York    HAIL                    6778         0
##  6 2003-06-03 Texas       HAIL                    6505         0
##  7 2008-06-04 Nebraska    HAIL                    6025         0
##  8 2008-06-08 Wisconsin   FLOOD                   5962         0
##  9 2008-08-02 Mississippi THUNDERSTORM WIND       5812         0
## 10 1996-09-06 Virginia    FLASH FLOOD             5721.        4
```

```
damage_state <- storm_data %>%
  group_by(State, Event_Name) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Crop_damage))
```

```
## `summarise()` regrouping output by 'State' (override with `.groups` argument)
```

```
summary(damage_state)
```

```
##     State            Event_Name         Crop_damage          Fatality
##  Length:1534        Length:1534        Min.   :     0.0   Min.   :  0.000
##  Class :character   Class :character   1st Qu.:     0.0   1st Qu.:  0.000
##  Mode  :character   Mode  :character   Median :     0.0   Median :  0.000
##                                        Mean   :   898.2   Mean   :  9.873
##                                        3rd Qu.:    50.0   3rd Qu.:  4.000
##                                        Max.   :201203.1   Max.   :653.000
```

```
head(damage_state, 10)
```

```
## # A tibble: 10 x 4
## # Groups:   State [6]
##    State        Event_Name        Crop_damage Fatality
##    <chr>        <chr>                   <dbl>    <dbl>
##  1 Nebraska     HAIL                  201203.        0
##  2 Texas        HAIL                  104050.        4
##  3 Kansas       HAIL                   80734.        0
##  4 Nebraska     THUNDERSTORM WIND      49874.        2
##  5 Iowa         HAIL                   47901.        4
##  6 Iowa         FLOOD                  46148.        1
##  7 Iowa         FLASH FLOOD            29658         6
##  8 North Dakota HAIL                   28820.        0
##  9 Wisconsin    FLASH FLOOD            25645.        7
## 10 Nebraska     FLASH FLOOD            25018.        3
```

```
fat_state_year <- storm_data %>%
   group_by(BGN_DATE, State, Event_Name) %>%
   summarise(Crop_damage = sum(CROPDMG),
             Fatality = sum(FATALITIES)) %>%
   arrange(desc(Fatality))
```

```
## `summarise()` regrouping output by 'BGN_DATE', 'State' (override with `.groups` argument)
```

```
summary(fat_state_year)
```

```
##     BGN_DATE             State            Event_Name         Crop_damage
##  Min.   :1950-01-03   Length:240766      Length:240766      Min.   :   0.000
##  1st Qu.:1990-08-17   Class :character   Class :character   1st Qu.:   0.000
##  Median :2000-02-14   Mode  :character   Mode  :character   Median :   0.000
##  Mean   :1996-01-04                                         Mean   :   5.723
##  3rd Qu.:2006-07-05                                         3rd Qu.:   0.000
##  Max.   :2011-11-30                                         Max.   :9630.000
##     Fatality
##  Min.   :  0.0000
##  1st Qu.:  0.0000
##  Median :  0.0000
##  Mean   :  0.0629
##  3rd Qu.:  0.0000
##  Max.   :583.0000
```

```
head(fat_state_year, 10)
```

```
## # A tibble: 10 x 5
## # Groups:   BGN_DATE, State [10]
##    BGN_DATE   State         Event_Name     Crop_damage Fatality
##    <date>     <chr>         <chr>                <dbl>    <dbl>
##  1 1995-07-12 Illinois      HEAT                     0      583
##  2 2011-04-27 Alabama       TORNADO                  0      235
##  3 2011-05-22 Missouri      TORNADO                  0      158
```

```
##  4 1965-04-11 Indiana       TORNADO                0      137
##  5 1953-05-11 Texas         TORNADO                0      127
##  6 1953-06-08 Michigan      TORNADO                0      125
##  7 1952-03-21 Arkansas      TORNADO                0      112
##  8 1971-02-21 Mississippi   TORNADO                0      110
##  9 1999-07-28 Illinois      EXCESSIVE HEAT         0      100
## 10 1953-06-09 Massachusetts TORNADO                0       90
```

```
fat_state <- storm_data %>%
  group_by(State, Event_Name) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Fatality))
```

```
## `summarise()` regrouping output by 'State' (override with `.groups` argument)
```

```
summary(fat_state)
```

```
##     State            Event_Name         Crop_damage          Fatality
##  Length:1534        Length:1534        Min.   :     0.0   Min.   :  0.000
##  Class :character   Class :character   1st Qu.:     0.0   1st Qu.:  0.000
##  Mode  :character   Mode  :character   Median :     0.0   Median :  0.000
##                                        Mean   :   898.2   Mean   :  9.873
##                                        3rd Qu.:    50.0   3rd Qu.:  4.000
##                                        Max.   :201203.1   Max.   :653.000
```

```
head(fat_state, 10)
```

```
## # A tibble: 10 x 4
## # Groups:   State [9]
##    State        Event_Name     Crop_damage Fatality
##    <chr>        <chr>                <dbl>    <dbl>
##  1 Illinois     HEAT                  60.4      653
##  2 Alabama      TORNADO             1653.       617
##  3 Texas        TORNADO             4866.       538
##  4 Mississippi  TORNADO            24964.       450
##  5 Missouri     TORNADO             2286        388
##  6 Arkansas     TORNADO              388.       379
##  7 Pennsylvania EXCESSIVE HEAT         0        376
##  8 Tennessee    TORNADO              681        368
##  9 Illinois     EXCESSIVE HEAT         0        330
## 10 Oklahoma     TORNADO              607.       296
```

```
damage_fat_event_year <- storm_data %>%
  group_by(BGN_DATE, Event_Name) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  filter(Crop_damage > 0, Fatality > 0)
```

```
## `summarise()` regrouping output by 'BGN_DATE' (override with `.groups` argument)
```

```
damage_event <- storm_data %>%
  group_by(Event_Name) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Crop_damage))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(damage_event, 10)
```

```
## # A tibble: 10 x 3
##    Event_Name          Crop_damage Fatality
##    <chr>                     <dbl>    <dbl>
##  1 HAIL                    585957.       45
##  2 THUNDERSTORM WIND       194820.      714
##  3 FLASH FLOOD             186484.     1035
##  4 FLOOD                   177967.      484
##  5 TORNADO                 100027.     5636
##  6 DROUGHT                  33954.        6
##  7 HIGH WIND                21063.      297
##  8 Other                    18888.      575
##  9 HEAVY RAIN               11694.       98
## 10 HURRICANE (TYPHOON)      11628.      133
```

```
fat_event <- storm_data %>%
  group_by(Event_Name) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Fatality))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(fat_event, 10)
```

```
## # A tibble: 10 x 3
##    Event_Name        Crop_damage Fatality
##    <chr>                   <dbl>    <dbl>
##  1 TORNADO                100027.     5636
##  2 EXCESSIVE HEAT           494.      1920
##  3 HEAT                     923       1212
##  4 FLASH FLOOD            186484.     1035
##  5 LIGHTNING                3586.      817
##  6 THUNDERSTORM WIND      194820.      714
##  7 RIP CURRENT                 0       577
##  8 Other                   18888.      575
##  9 FLOOD                  177967.      484
## 10 HIGH WIND               21063.      297
```

```
damage_designator_year <- storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(Year, Designator) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Crop_damage))
```

```
## `summarise()` regrouping output by 'Year' (override with `.groups` argument)
```

```
head(damage_designator_year, 10)
```

```
## # A tibble: 10 x 4
## # Groups:   Year [10]
##     Year Designator    Crop_damage Fatality
##    <int> <chr>               <dbl>    <dbl>
##  1  2008 County/Parish     175255       261
##  2  2011 County/Parish      93096       794
##  3  1998 County/Parish      91846.      359
##  4  1994 County/Parish      83575.      217
##  5  2000 County/Parish      73925.      157
##  6  2007 County/Parish      68273       231
##  7  2004 County/Parish      67006.      176
##  8  1997 County/Parish      66891.      291
##  9  1996 County/Parish      60804.      236
## 10  2009 County/Parish      57227       137
```

```
fat_designator_year <- storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(Year, Designator) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES)) %>%
  arrange(desc(Fatality))
```

```
## `summarise()` regrouping output by 'Year' (override with `.groups` argument)
```

```
head(fat_designator_year, 10)
```

```
## # A tibble: 10 x 4
## # Groups:   Year [9]
##     Year Designator    Crop_damage Fatality
##    <int> <chr>               <dbl>    <dbl>
##  1  1995 Zone                1896.     1148
##  2  2011 County/Parish      93096       794
##  3  1999 Zone                4489.      639
##  4  1953 County/Parish          0       519
##  5  2006 Zone               12547.      389
##  6  1974 County/Parish          0       366
##  7  1998 County/Parish      91846.      359
##  8  2005 Zone                5769.      314
##  9  1965 County/Parish          0       301
## 10  1998 Zone                5130.      297
```

```
damage_fat_designator <- storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(Year, Designator) %>%
  summarise(Crop_damage = sum(CROPDMG),
            Fatality = sum(FATALITIES))
```

```
## `summarise()` regrouping output by 'Year' (override with `.groups` argument)
```

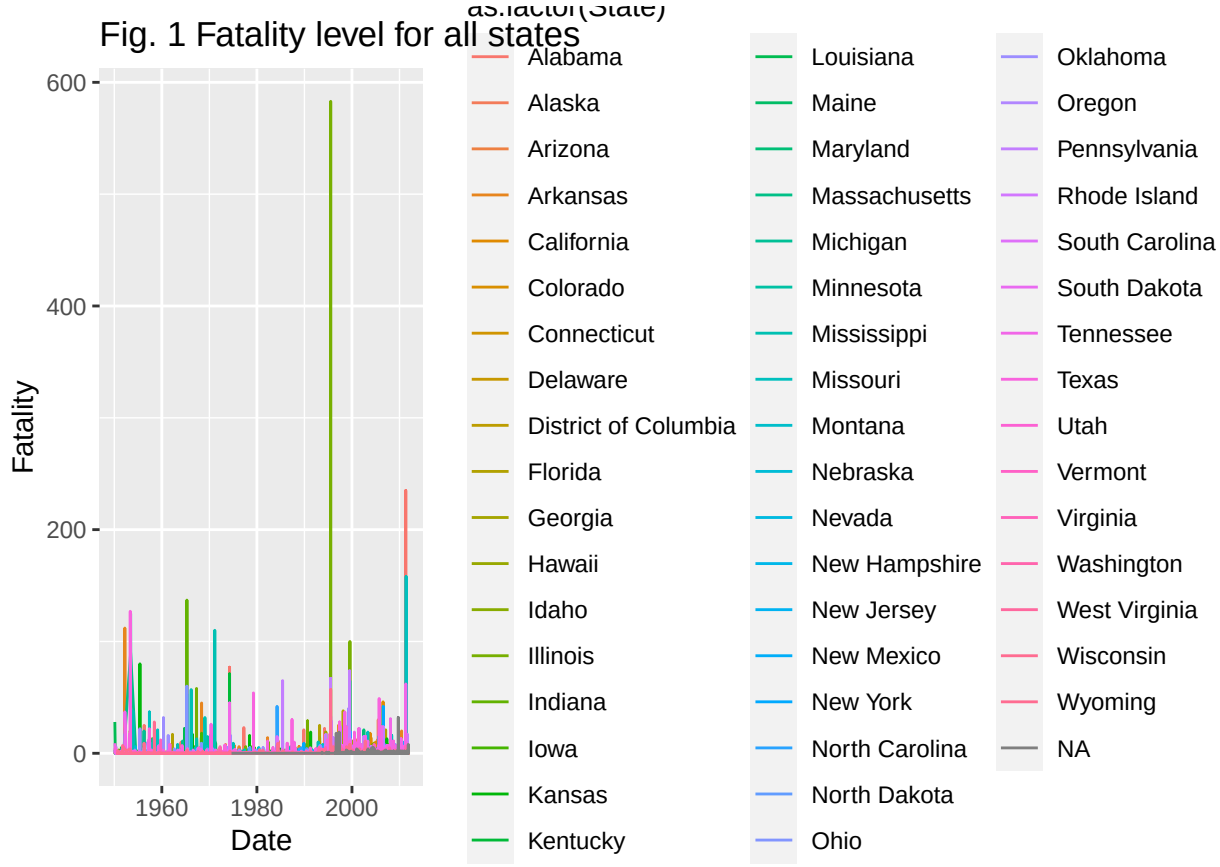
```
head(damage_fat_designator)
```

```
## # A tibble: 6 x 4
## # Groups:   Year [6]
##    Year Designator    Crop_damage Fatality
##   <int> <chr>               <dbl>    <dbl>
## 1  1950 County/Parish           0       70
## 2  1951 County/Parish           0       34
## 3  1952 County/Parish           0      230
## 4  1953 County/Parish           0      519
## 5  1954 County/Parish           0       36
## 6  1955 County/Parish           0      129
```

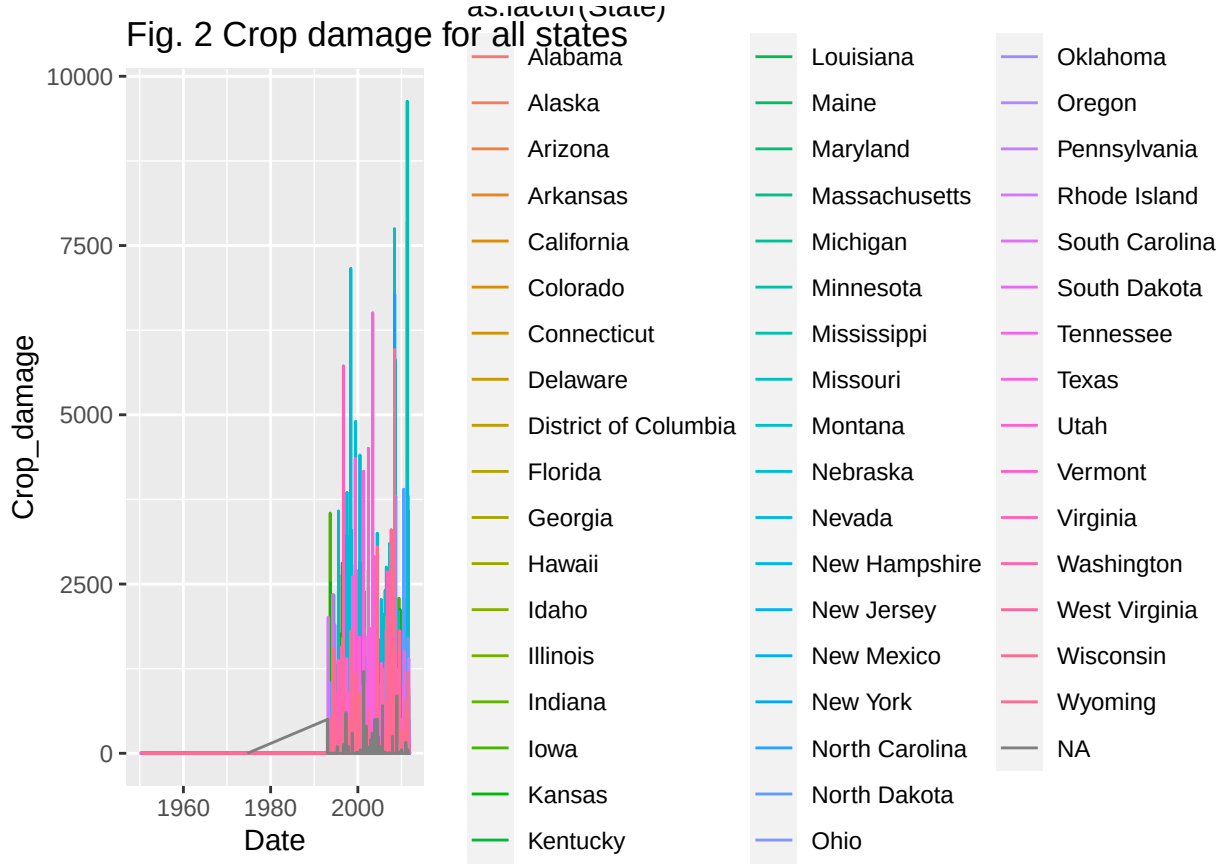
Plots

You can also embed plots, for example:

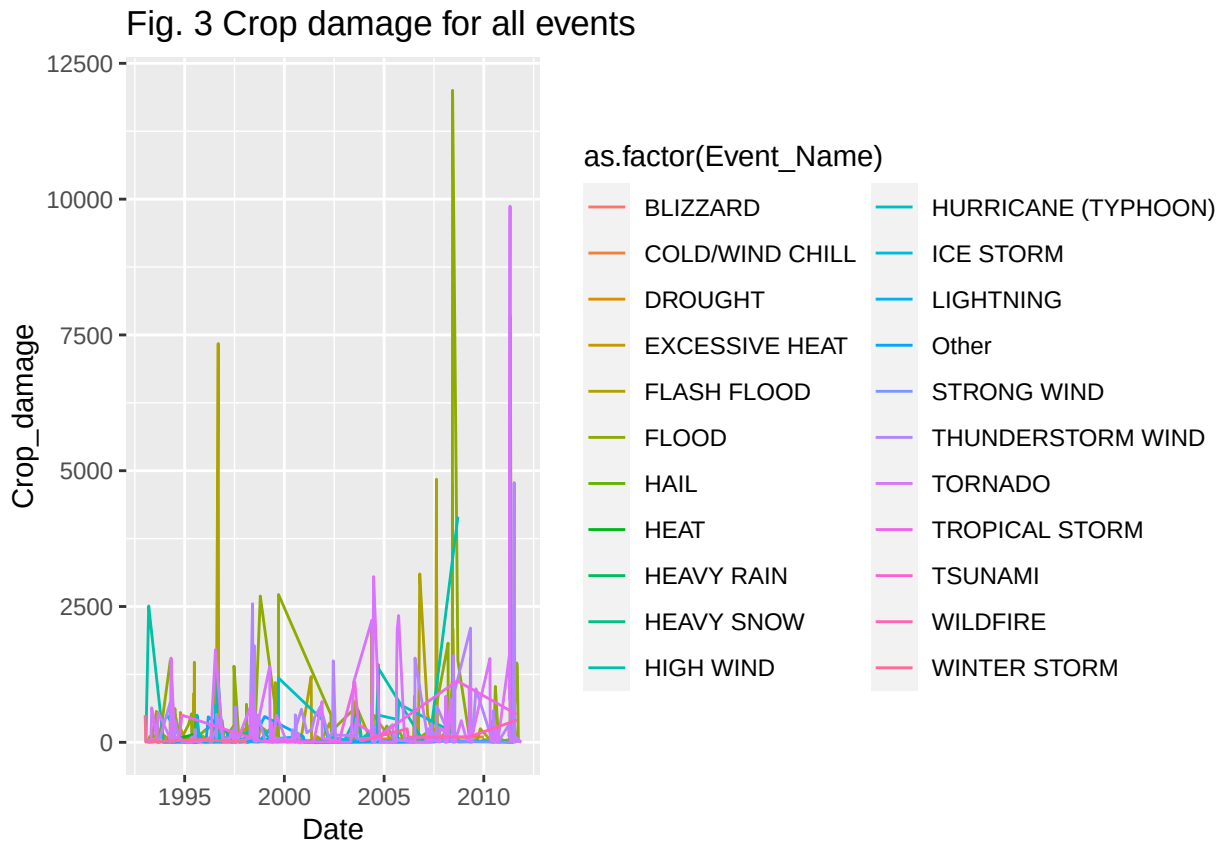
```
g <- ggplot(damage_state_year, aes(BGN_DATE, Fatality, group = as.factor(State),  color = as.factor(Sta
g + geom_line() + xlab("Date") + ggtitle("Fig. 1 Fatality level for all states")
```




```
d <- ggplot(fat_state_year, aes(BGN_DATE, Crop_damage, group = as.factor(State),  color = as.factor(Sta
d + geom_line() + xlab("Date") + ggtitle("Fig. 2 Crop damage for all states")
```

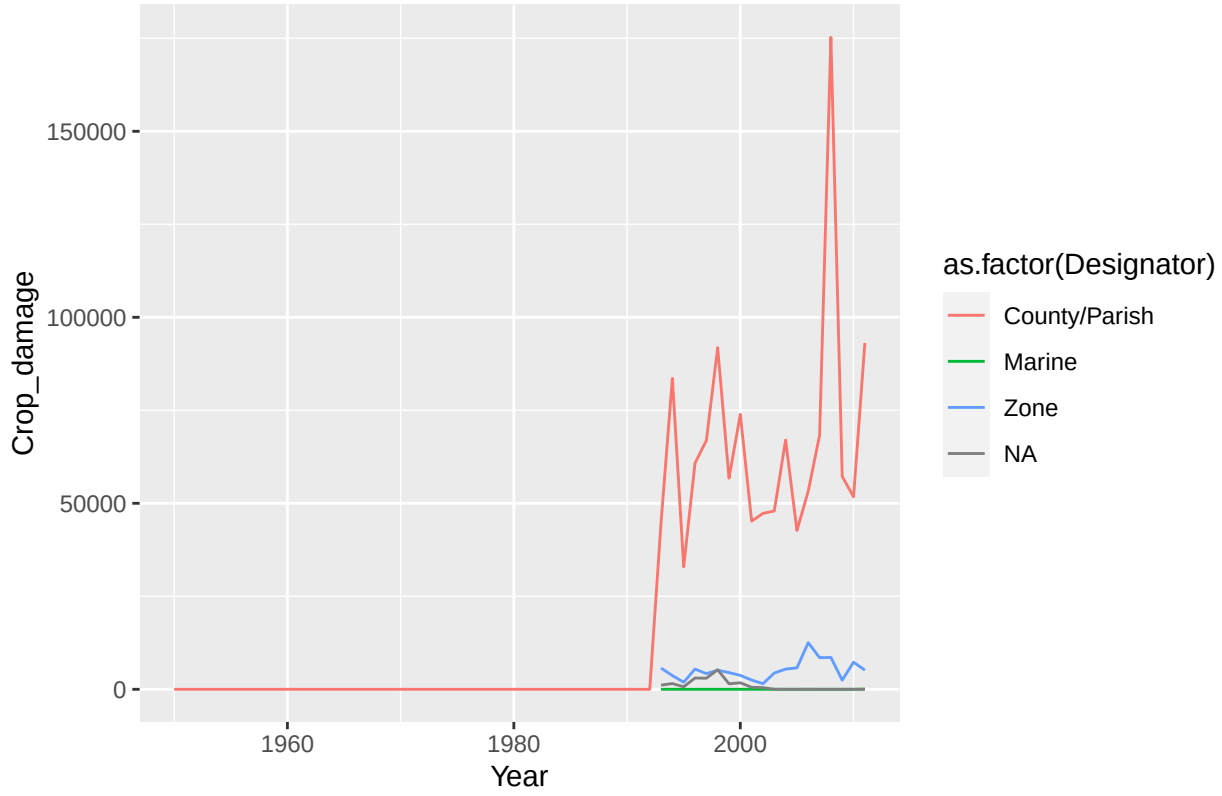



```
e <- ggplot(damage_fat_event_year, aes(BGN_DATE, Crop_damage, group = as.factor(Event_Name),  color = as
e + geom_line() + xlab("Date") + ggtitle("Fig. 3 Crop damage for all events")
```



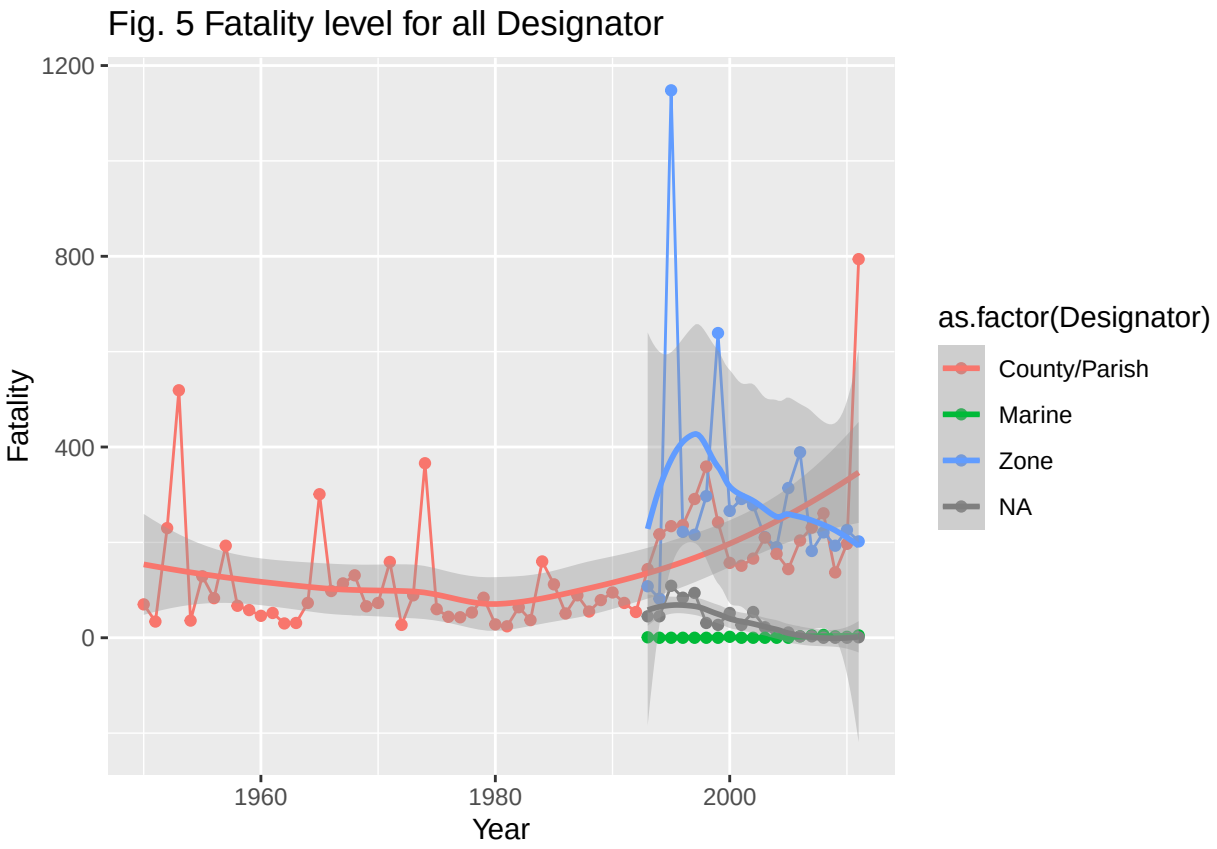

```
f <- ggplot(damage_designator_year, aes(Year, Crop_damage, group = as.factor(Designator),  color = as.fa
f + geom_line() + xlab("Year") + ggtitle("Fig. 4 Crop damage for all Designator")
```

Fig. 4 Crop damage for all Designator




```
h <- ggplot(fat_designator_year, aes(Year, Fatality, group = as.factor(Designator),  color = as.factor(
h + geom_line() + geom_point() + geom_smooth() + xlab("Year") + ggtitle("Fig. 5 Fatality level for all
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Results

1. Crop damage and fatalities for all time of observed data and all states. From the table 1 we can see that the event with a biggest Crop damage is TORNADO in Smithville, Mississippi April 27, 2011 (fig. 2). Nebraska state has a total biggest Crop damage from HAIL. The most mortatality event (HEAT) was happened in Chicago, Illinois in The 1995 (fig. 1)

2. Total level of Crop damage and fatalities for all states. The state with a most mortality level is Illinois from HEAT and Nebraska has a maximum Crop damage from HAIL.

3. Crop damage and fatalities for all events and designators. In the total level event with a maximum Crop damage is a HAIL and an event with a maximum mortalities is TORNADO. How we can see from fig. 4 and fig. 5 County/Parish have the biggest Crop damage, but an absolute maximum for Fatality is achieved for a Zone.