

Reproducible Research - Project 2

Hayden Roche

December 19, 2017

Data Processing

Using a `download.file` function I was able to pull the zip file right into R Studio and read the csv in through a local storage file. The next step was implementing `gsub` to clean up the `PROMDMG` column because there were a lot of inconsistencies with the naming conventions.

Synopsis

This analysis is looking into fatalities, injuries and property damages from various natural disasters. These natural disasters have been recorded over from 1950-2011. The overall findings are quite interesting: torandoes, excessive heat and flash floods are the deadliest of natural disasters. Over the 60 years of data collection, tornadoes accounted for 5.6K fatalities.

The story is similar when we look at injuries, but the magnitude increases significantly. Tornadoes led the injury count by miles, with a total 91.3K and then excessive heat followed with 8.6K injuries. An interesting next step would be creating a weighted damage measure that factored in both deaths and injuries. This would then take a nice consistent measure for damages.

```
str(Clean_Storm_data)
```

```
## 'data.frame':    902297 obs. of  9 variables:
##  $ BGN_DATE      : Date, format: "1950-04-18" "1950-04-18" ...
##  $ EVTYPE        : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ FATALITIES    : num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES      : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ 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 17 ...
##  $ 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 ...
##  $ TotalDamage   : chr  "25 - K" "2.5 - K" "25 - K" "2.5 - K" ...
```

```
head(Clean_Storm_data)
```

##	BGN_DATE	EVTYPE	FATALITIES	INJURIES	PROPDMG	PROPDMGEXP	CROPDMG
## 1	1950-04-18	TORNADO	0	15	25.0	K	0
## 2	1950-04-18	TORNADO	0	0	2.5	K	0
## 3	1951-02-20	TORNADO	0	2	25.0	K	0
## 4	1951-06-08	TORNADO	0	2	2.5	K	0
## 5	1951-11-15	TORNADO	0	2	2.5	K	0
## 6	1951-11-15	TORNADO	0	6	2.5	K	0
##	CROPDMGEXP	TotalDamage					
## 1		25 - K					
## 2		2.5 - K					
## 3		25 - K					
## 4		2.5 - K					
## 5		2.5 - K					
## 6		2.5 - K					

Results

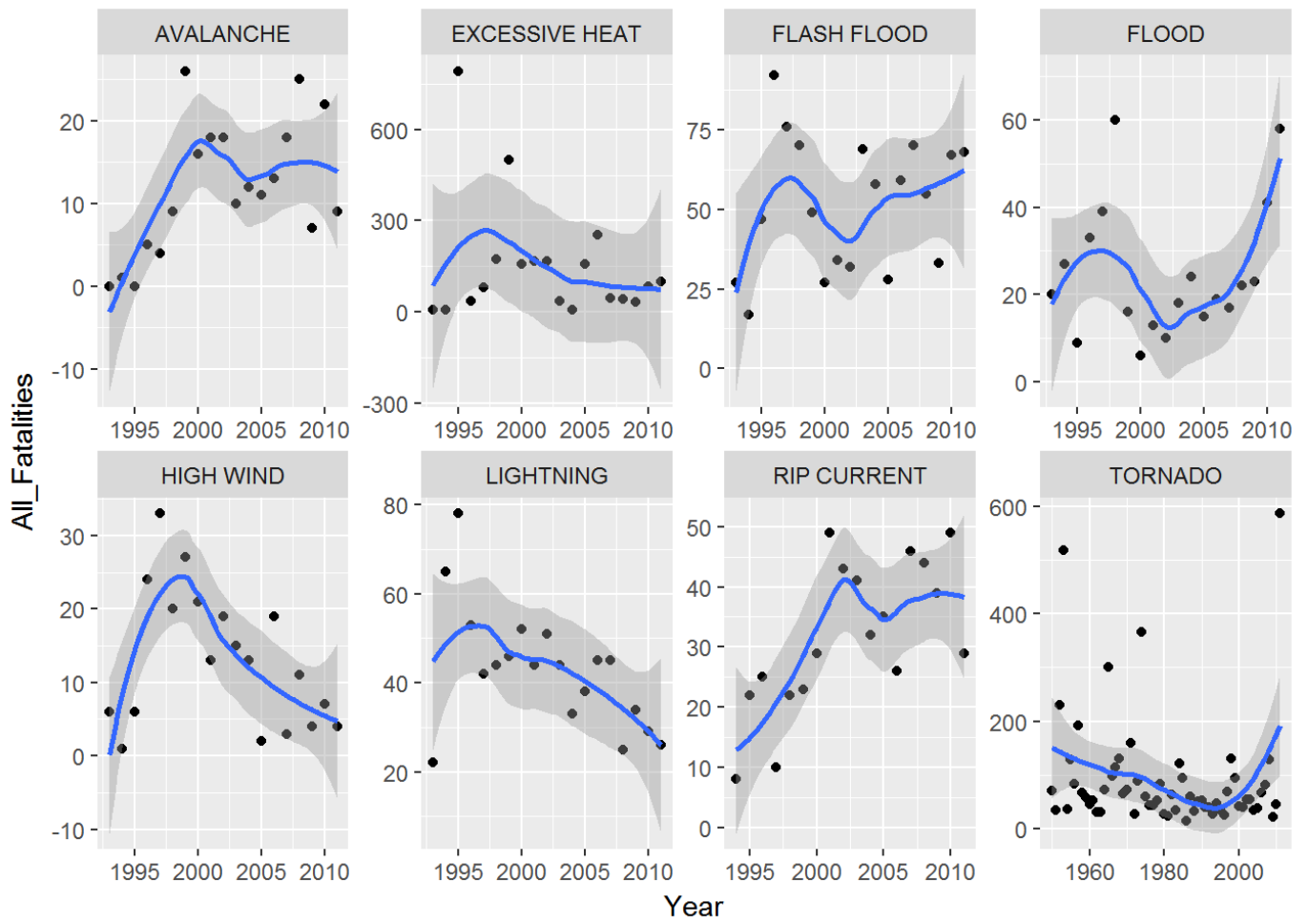
Fatalities: Manipulating the data to the top 5 most dangerous natural events type

```
Fatalities_DF <- Clean_Storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(Year, EVTYPE) %>%
  summarise(All_Fatalities = sum(FATALITIES)) %>%
  filter(EVTYPE == 'EXCESSIVE HEAT' | EVTYPE == 'TORNADO' | EVTYPE == 'LIGHTNING'
| EVTYPE == 'FLASH FLOOD' | EVTYPE == 'RIP CURRENT' | EVTYPE == 'FLOOD' | EVTYPE =
= 'HIGH WIND' | EVTYPE == 'AVALANCHE')
```

```
## Warning: package 'bindrcpp' was built under R version 3.4.1
```

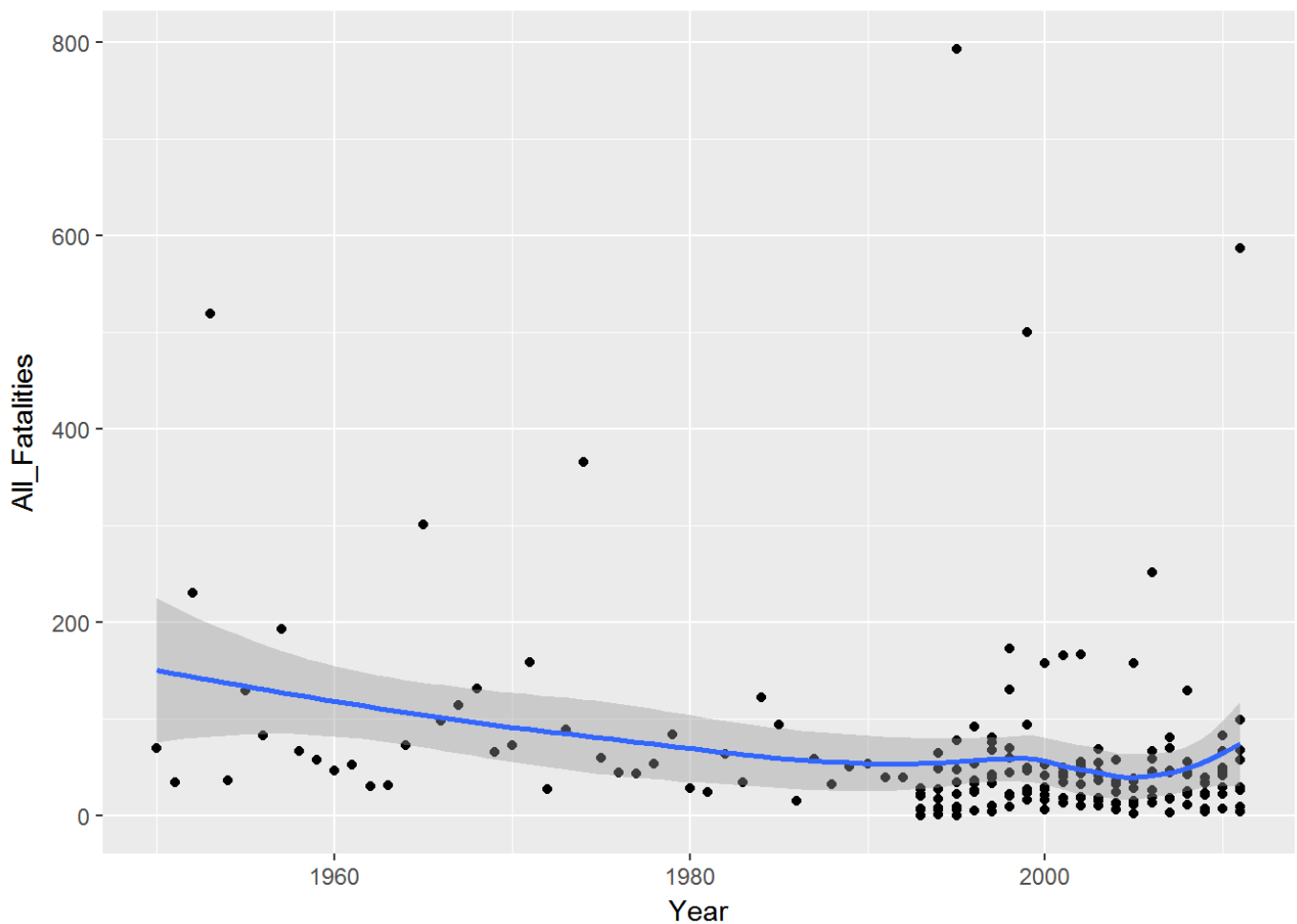
```
m <- ggplot(data = Fatalities_DF, aes(x = Year, y = All_Fatalities)) +
  geom_point() +
  stat_smooth()
m + facet_wrap(~ EVTYPE, ncol=4, scales="free")
```

```
## `geom_smooth()` using method = 'loess'
```



```
print(m)
```

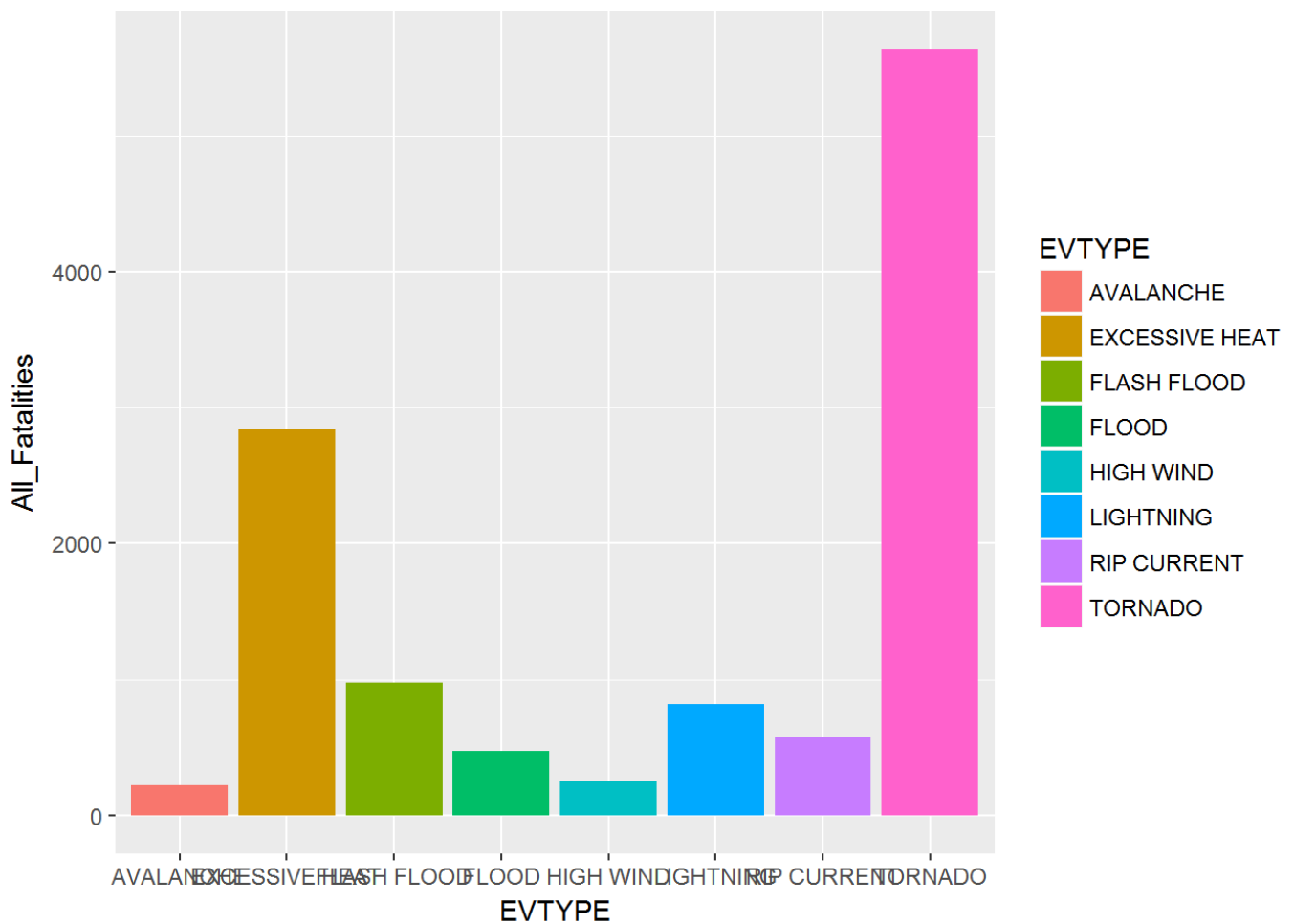
```
## `geom_smooth()` using method = 'loess'
```



Fatalities: Manipulating the data to the top 5 most dangerous natural events type

```
Fatalities_Total <- Clean_Storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(EVTYPE) %>%
  summarise(All_Fatalities = sum(FATALITIES)) %>%
  as.data.frame() %>%
  arrange(desc(All_Fatalities)) %>%
  filter(EVTYPE == 'EXCESSIVE HEAT' | EVTYPE == 'TORNADO' | EVTYPE == 'LIGHTNING'
| EVTYPE == 'FLASH FLOOD' | EVTYPE == 'RIP CURRENT' | EVTYPE == 'FLOOD' | EVTYPE =
= 'HIGH WIND' | EVTYPE == 'AVALANCHE')
View(Fatalities_Total)

b <- ggplot(Fatalities_Total, aes(x = EVTYPE, y = All_Fatalities, fill = EVTYPE))
b + geom_bar(stat = "identity")
```

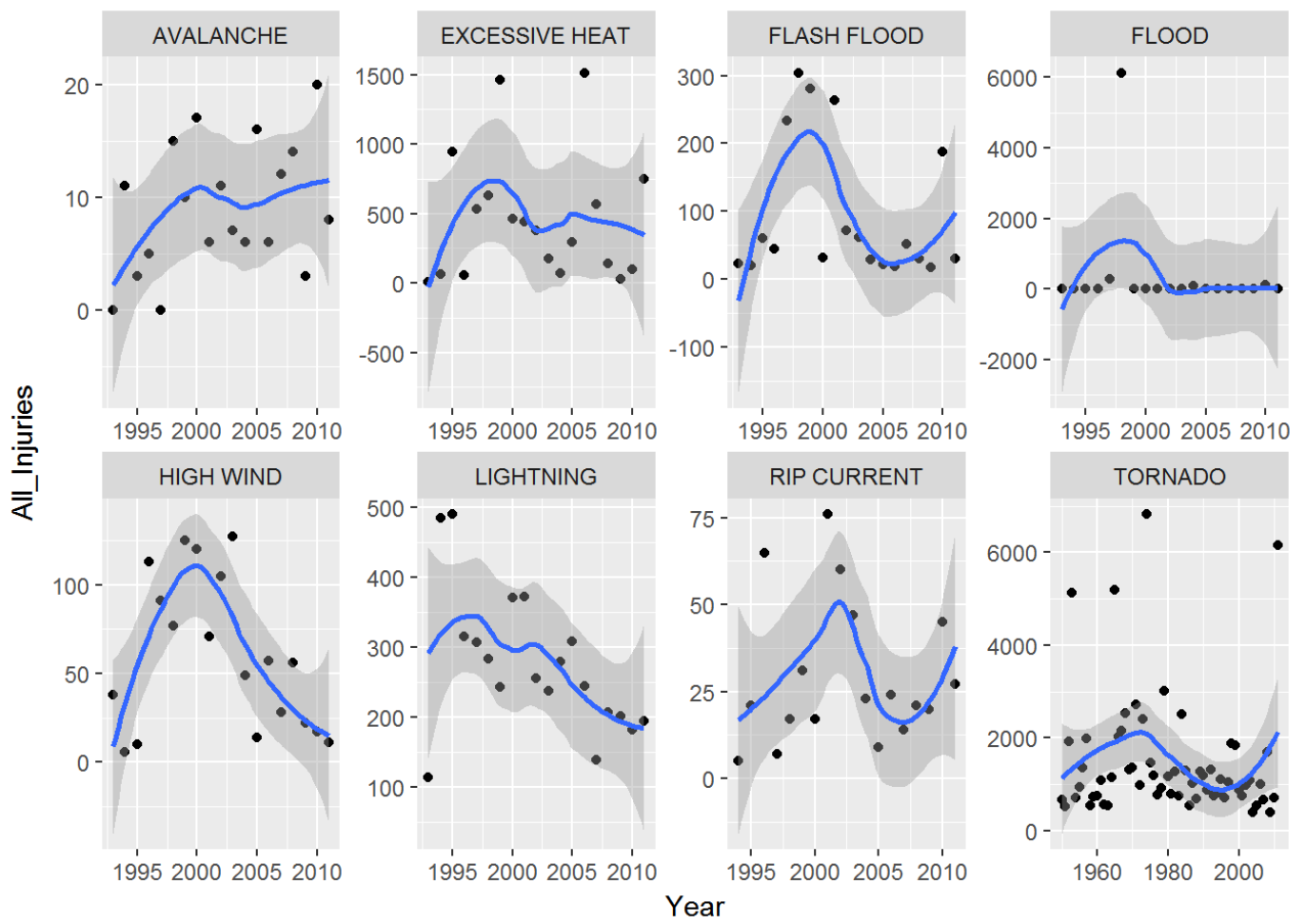


Injuries: Manipulating the data to the top 8 most dangerous natural events type

```
Injuries_DF <- Clean_Storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(Year, EVTYPE) %>%
  summarise(All_Injuries = sum(INJURIES)) %>%
  filter(EVTYPE == 'EXCESSIVE HEAT' | EVTYPE == 'TORNADO' | EVTYPE == 'LIGHTNING'
| EVTYPE == 'FLASH FLOOD' | EVTYPE == 'RIP CURRENT' | EVTYPE == 'FLOOD' | EVTYPE =
= 'HIGH WIND' | EVTYPE == 'AVALANCHE')

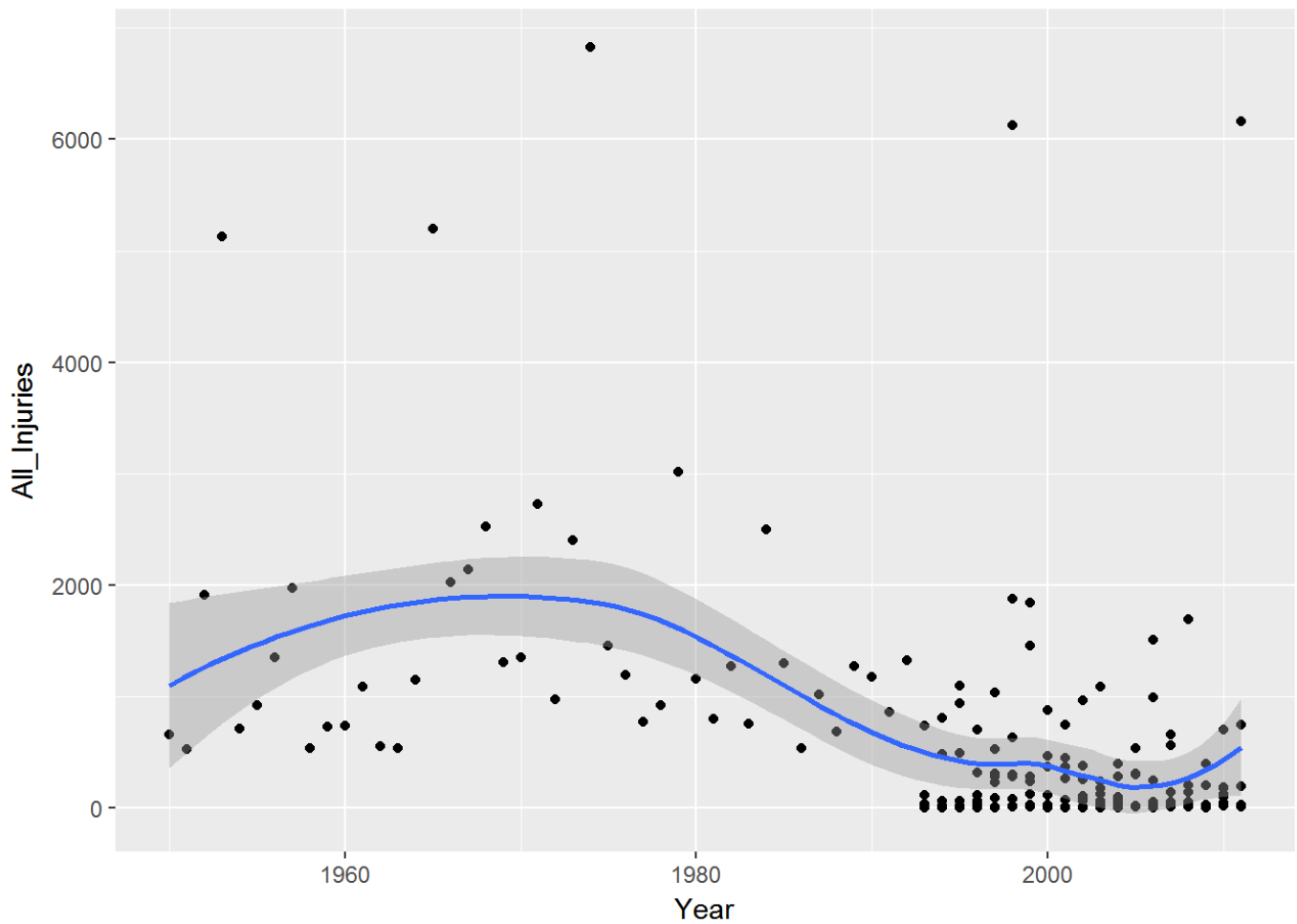
m <- ggplot(data = Injuries_DF, aes(x = Year, y = All_Injuries)) +
  geom_point() +
  stat_smooth()
m + facet_wrap(~ EVTYPE, ncol=4, scales="free")
```

```
## `geom_smooth()` using method = 'loess'
```



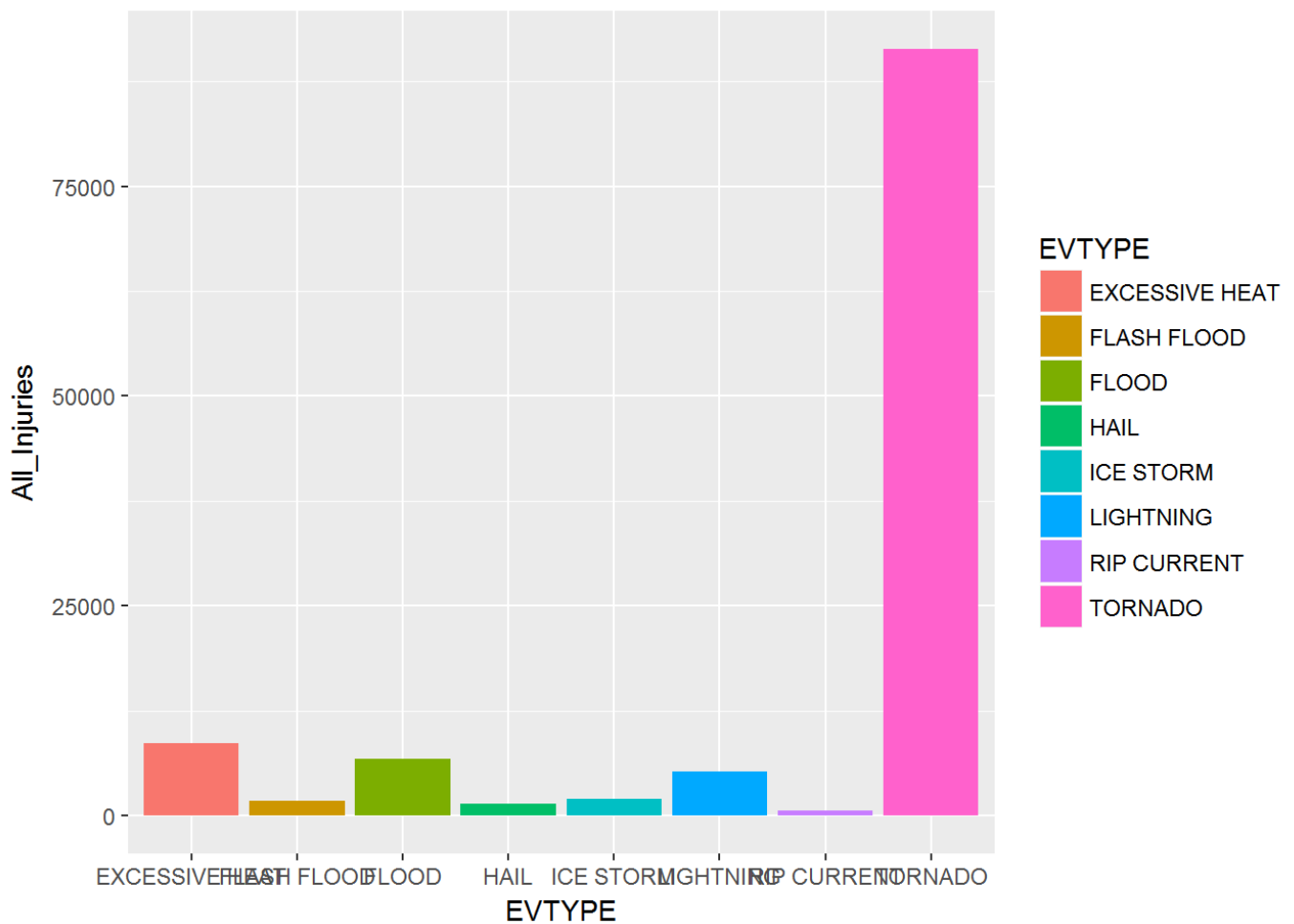
```
print(m)
```

```
## `geom_smooth()` using method = 'loess'
```



```
Injuries_Total <- Clean_Storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(EVTYPE) %>%
  summarise(All_Injuries = sum(INJURIES)) %>%
  as.data.frame() %>%
  arrange(desc(All_Injuries)) %>%
  filter(EVTYPE == 'EXCESSIVE HEAT' | EVTYPE == 'TORNADO' | EVTYPE == 'LIGHTNING'
| EVTYPE == 'FLASH FLOOD' | EVTYPE == 'RIP CURRENT' | EVTYPE == 'FLOOD' | EVTYPE =
= 'ICE STORM' | EVTYPE == 'HAIL')
View(Injuries_Total)

b <- ggplot(Injuries_Total, aes(x = EVTYPE, y = All_Injuries, fill = EVTYPE))
b + geom_bar(stat = "identity")
```



Property Damage

```
Property <- Clean_Storm_data %>%
  mutate(Year = year(BGN_DATE)) %>%
  group_by(EVTYPE) %>%
  summarise(All_Damage = sum(PROPDMG)) %>%
  as.data.frame() %>%
  arrange(desc(All_Damage)) %>%
  filter(EVTYPE == 'WINTER STORM' | EVTYPE == 'TORNADO' | EVTYPE == 'LIGHTNING' |
EVTYPE == 'FLASH FLOOD' | EVTYPE == 'THUNDERSTORM WINDS' | EVTYPE == 'FLOOD' | EVT
YPE == 'HIGH WIND' | EVTYPE == 'HAIL')
View(Property)

d <- ggplot(Property, aes(x = EVTYPE, y = All_Damage, fill = EVTYPE))
d + geom_bar(stat = "identity")
```