**# Commands and Results #**

echo = TRUE # Always make code visible

options(scipen = 1) # Turn off scientific notations for numbers

library(R.utils)

library(ggplot2)

library(plyr)

require(gridExtra)

getwd()

setwd("C:/Users/shaya/Desktop/Reprocible\_Research\_Course\_Project\_2")

if (!"stormData.csv.bz2" %in% dir("./data/")) {

print("hhhh")

download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile = "data/stormData.csv.bz2")

bunzip2("data/stormData.csv.bz2", overwrite=T, remove=F)

}

if (!"stormData" %in% ls()) {

stormData <- read.csv("data/stormData.csv", sep = ",")

}

dim(stormData)

head(stormData, n = 2)

[1] 902297 37

R Console

|  |
| --- |
|  |

|  | **STATE\_\_**  <dbl> | **BGN\_DATE**  <fctr> | **BGN\_TIME**  <fctr> | **TIME\_ZONE**  <fctr> |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 4/18/1950 0:00:00 | 0130 | CST |  |
| 2 | 1 | 4/18/1950 0:00:00 | 0145 | CST |  |

2 rows | 1-5 of 37 columns

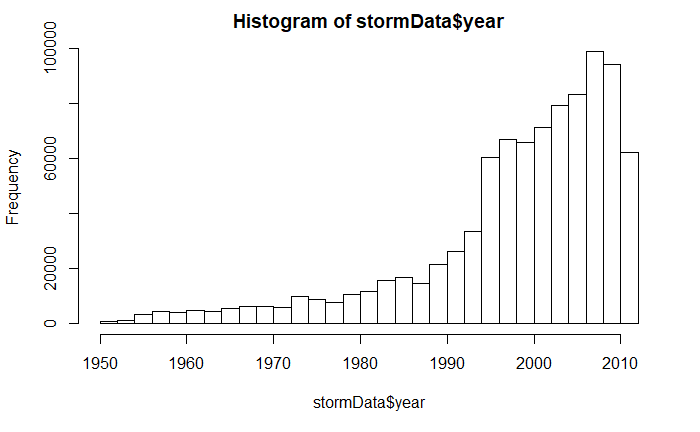
data.frame

2 x 37

[1] 902297 37

|  | **STATE\_\_**  <dbl> | **BGN\_DATE**  <fctr> | **BGN\_TIME**  <fctr> | **TIME\_ZONE**  <fctr> | **COUNTY**  <dbl> | **COUNTYNAME**  <fctr> | **STATE**  <fctr> | **EVTYPE**  <fctr> | **BGN\_RANGE**  <dbl> |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 4/18/1950 0:00:00 | 0130 | CST | 97 | MOBILE | AL | TORNADO | 0 |  |
| 2 | 1 | 4/18/1950 0:00:00 | 0145 | CST | 3 | BALDWIN | AL | TORNADO | 0 |  |

|  | **STATE\_\_**  <dbl> | **BGN\_DATE**  <fctr> | **BGN\_TIME**  <fctr> | **TIME\_ZONE**  <fctr> | **COUNTY**  <dbl> | **COUNTYNAME**  <fctr> | **STATE**  <fctr> | **EVTYPE**  <fctr> | **BGN\_RANGE**  <dbl> |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 1 | 4/18/1950 0:00:00 | 0130 | CST | 97 | MOBILE | AL | TORNADO | 0 |  |
| 2 | 1 | 4/18/1950 0:00:00 | 0145 | CST | 3 | BALDWIN | AL | TORNADO | 0 |  |



storm <- stormData[stormData$year >= 1995, ]

dim(storm)

[1] 681500 38

sortHelper <- function(fieldName, top = 15, dataset = stormData) {

index <- which(colnames(dataset) == fieldName)

field <- aggregate(dataset[, index], by = list(dataset$EVTYPE), FUN = "sum")

names(field) <- c("EVTYPE", fieldName)

field <- arrange(field, field[, 2], decreasing = T)

field <- head(field, n = top)

field <- within(field, EVTYPE <- factor(x = EVTYPE, levels = field$EVTYPE))

return(field)

}

fatalities <- sortHelper("FATALITIES", dataset = storm)

injuries <- sortHelper("INJURIES", dataset = storm)

convertHelper <- function(dataset = storm, fieldName, newFieldName) {

totalLen <- dim(dataset)[2]

index <- which(colnames(dataset) == fieldName)

dataset[, index] <- as.character(dataset[, index])

logic <- !is.na(toupper(dataset[, index]))

dataset[logic & toupper(dataset[, index]) == "B", index] <- "9"

dataset[logic & toupper(dataset[, index]) == "M", index] <- "6"

dataset[logic & toupper(dataset[, index]) == "K", index] <- "3"

dataset[logic & toupper(dataset[, index]) == "H", index] <- "2"

dataset[logic & toupper(dataset[, index]) == "", index] <- "0"

dataset[, index] <- as.numeric(dataset[, index])

dataset[is.na(dataset[, index]), index] <- 0

dataset <- cbind(dataset, dataset[, index - 1] \* 10^dataset[, index])

names(dataset)[totalLen + 1] <- newFieldName

return(dataset)

}

storm <- convertHelper(storm, "PROPDMGEXP", "propertyDamage")

storm <- convertHelper(storm, "CROPDMGEXP", "cropDamage")

names(storm)

options(scipen=999)

property <- sortHelper("propertyDamage", dataset = storm)

crop <- sortHelper("cropDamage", dataset = storm)

NAs introduced by coercionNAs introduced by coercion [1] "STATE\_\_" "BGN\_DATE" "BGN\_TIME" "TIME\_ZONE" "COUNTY" "COUNTYNAME"

[7] "STATE" "EVTYPE" "BGN\_RANGE" "BGN\_AZI" "BGN\_LOCATI" "END\_DATE"

[13] "END\_TIME" "COUNTY\_END" "COUNTYENDN" "END\_RANGE" "END\_AZI" "END\_LOCATI"

[19] "LENGTH" "WIDTH" "F" "MAG" "FATALITIES" "INJURIES"

[25] "PROPDMG" "PROPDMGEXP" "CROPDMG" "CROPDMGEXP" "WFO" "STATEOFFIC"

[31] "ZONENAMES" "LATITUDE" "LONGITUDE" "LATITUDE\_E" "LONGITUDE\_" "REMARKS"

[37] "REFNUM" "year" "propertyDamage" "cropDamage"

fatalities

|  |
| --- |
|  |
|  | **EVTYPE**  <fctr> | **FATALITIES**  <dbl> |  |  |
| 1 | EXCESSIVE HEAT | 1903 |  |  |
| 2 | TORNADO | 1545 |  |  |
| 3 | FLASH FLOOD | 934 |  |  |
| 4 | HEAT | 924 |  |  |
| 5 | LIGHTNING | 729 |  |  |
| 6 | FLOOD | 423 |  |  |
| 7 | RIP CURRENT | 360 |  |  |
| 8 | HIGH WIND | 241 |  |  |
| 9 | TSTM WIND | 241 |  |  |
| 10 | AVALANCHE | 223 |  |  |

injuries

|  |
| --- |
|  |
|  | **EVTYPE**  <fctr> | **INJURIES**  <dbl> |  |  |
| 1 | TORNADO | 21765 |  |  |
| 2 | FLOOD | 6769 |  |  |
| 3 | EXCESSIVE HEAT | 6525 |  |  |
| 4 | LIGHTNING | 4631 |  |  |
| 5 | TSTM WIND | 3630 |  |  |
| 6 | HEAT | 2030 |  |  |
| 7 | FLASH FLOOD | 1734 |  |  |
| 8 | THUNDERSTORM WIND | 1426 |  |  |
| 9 | WINTER STORM | 1298 |  |  |
| 10 | HURRICANE/TYPHOON | 1275 |  |  |

fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", binwidth = NULL) +

scale\_y\_continuous("Number of Fatalities") +

theme(axis.text.x = element\_text(angle = 45,

hjust = 1)) + xlab("Severe Weather Type") +

ggtitle("Total Fatalities by Severe Weather\n Events in the U.S.\n from 1995 - 2011")

injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, geom = "bar", binwidth = NULL) +

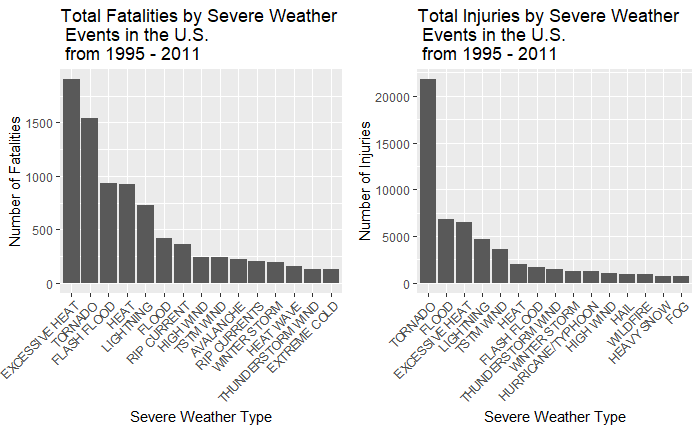
scale\_y\_continuous("Number of Injuries") +

theme(axis.text.x = element\_text(angle = 45,

hjust = 1)) + xlab("Severe Weather Type") +

ggtitle("Total Injuries by Severe Weather\n Events in the U.S.\n from 1995 - 2011")

grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)



Property

|  |
| --- |
|  |
|  | **EVTYPE**  <fctr> | **propertyDamage**  <dbl> |  |  |
| 1 | FLOOD | 144022037057 |  |  |
| 2 | HURRICANE/TYPHOON | 69305840000 |  |  |
| 3 | STORM SURGE | 43193536000 |  |  |
| 4 | TORNADO | 24935939545 |  |  |
| 5 | FLASH FLOOD | 16047794571 |  |  |
| 6 | HAIL | 15048722103 |  |  |
| 7 | HURRICANE | 11812819010 |  |  |
| 8 | TROPICAL STORM | 7653335550 |  |  |
| 9 | HIGH WIND | 5259785375 |  |  |
| 10 | WILDFIRE | 4759064000 |  |  |

crop

|  |
| --- |
|  |
|  | **EVTYPE**  <fctr> | **cropDamage**  <dbl> |  |  |
| 1 | DROUGHT | 13922066000 |  |  |
| 2 | FLOOD | 5422810400 |  |  |
| 3 | HURRICANE | 2741410000 |  |  |
| 4 | HAIL | 2614127070 |  |  |
| 5 | HURRICANE/TYPHOON | 2607872800 |  |  |
| 6 | FLASH FLOOD | 1343915000 |  |  |
| 7 | EXTREME COLD | 1292473000 |  |  |
| 8 | FROST/FREEZE | 1094086000 |  |  |
| 9 | HEAVY RAIN | 728399800 |  |  |
| 10 | TROPICAL STORM | 677836000 |  |  |

propertyPlot <- qplot(EVTYPE, data = property, weight = propertyDamage, geom = "bar", binwidth = NULL) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + scale\_y\_continuous("Property Damage in US dollars")+

xlab("Severe Weather Type") + ggtitle("Total Property Damage by\n Severe Weather Events in\n the U.S. from 1995 - 2011")

cropPlot<- qplot(EVTYPE, data = crop, weight = cropDamage, geom = "bar", binwidth = NULL) +

theme(axis.text.x = element\_text(angle = 45, hjust = 1)) + scale\_y\_continuous("Crop Damage in US dollars") +

xlab("Severe Weather Type") + ggtitle("Total Crop Damage by \nSevere Weather Events in\n the U.S. from 1995 - 2011")

grid.arrange(propertyPlot, cropPlot, ncol = 2)

