Thong Pham
John Hopkins Data Science
March 6th, 2024

Weather Data Analysis

Synopsis:

Utilizing R Markdown to document the process of addressing project questions. Due to the large size of the data frame repdata_data_StormData.csv, the initial step involves splitting the data and extracting only the required information. The code first loads the data, then creates a new data frame containing the relevant columns "FATALITIES", "INJURIES", and "EVTYPE". The EVTYPE column contains numerous instances of weather event types with varying recorded injuries and fatalities. Using this information, I identify the unique weather types and calculate the total fatalities and injuries for each type, storing the results in a new data frame for further analysis. The final step involves extracting the necessary information to address the project questions.

Loading the data:

Loading the data using read.csv with the document file_path as the input

```{r, echo = TRUE, cache = TRUE}
#Load the data
file_path <- "Downloads/repdata_data_StormData.csv"
data <- read.csv(file_path)
```

Process data for question 1:

```{r, echo = TRUE, cache = TRUE}
#Create a new data frame with only relevant columns (FATALITIES, INJURIES, EVTYPE)
new_df <- data[, c("FATALITIES", "INJURIES", "EVTYPE")]
head(new_df)

#Sum up all the recorded injuries and fatalities with the unique event type
unique_events <- unique(new_df$EVTYPE)

# Initialize an empty data frame to store the results
event_totals <- data.frame(EVTYPE = character(), FATALITIES = numeric(), INJURIES =
numeric(), stringsAsFactors = FALSE)

# Loop through each unique event type
for (event in unique_events) {
  # Subset the data for the current event type
  subset_data <- new_df[new_df$EVTYPE == event, ]
```

```
  # Calculate total fatalities and injuries
  total_fatalities <- sum(subset_data$FATALITIES, na.rm = TRUE)
  total_injuries <- sum(subset_data$INJURIES, na.rm = TRUE)

  # Add the totals to the results data frame
  event_totals <- rbind(event_totals, data.frame(EVTYPE = event, FATALITIES =
total_fatalities, INJURIES = total_injuries))
}

# Print the result
head(event_totals)
```

Question 1: Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```{r, echo = TRUE, cache = TRUE}
event_fatalities_order <- event_totals[order(-event_totals$FATALITIES), ]
event_injuries_order <- event_totals[order(-event_totals$INJURIES), ]

head(event_fatalities_order)
head(event_injuries_order)
```

Results Question 1:

- event_fatalities_order <- event_totals[order(-event_totals$FATALITIES),]
- Using the command to order the weather event with the most fatalities, results can be interpreted from the table.

	EVTYPE	FATALITIES	INJURIES
1	TORNADO	5633	91346
99	EXCESSIVE HEAT	1903	6525
20	FLASH FLOOD	978	1777
27	HEAT	937	2100
15	LIGHTNING	816	5230
2	TSTM WIND	504	6957

- event_injuries_order <- event_totals[order(-event_totals$INJURIES),]
- Using the command to order the weather event with the most injuries, results can be interpreted from the table.

	EVTYPE	FATALITIES	INJURIES
1	TORNADO	5633	91346
2	TSTM WIND	504	6957
36	FLOOD	470	6789
99	EXCESSIVE HEAT	1903	6525
15	LIGHTNING	816	5230
27	HEAT	937	2100

Process data for question 2:

```{r, echo = TRUE, cache = TRUE}
# Create a new data frame with only relevant columns (PROPDMG, PROPDMGEXP, CROPDMG, EVTYPE)
new_df1 <- data[, c("PROPDMG", "PROPDMGEXP", "CROPDMG", "EVTYPE")]
```

```
# Function to convert PROPDMGEXP to numeric multiplier, using PROPDMGEXP for
CROPDMG
exp_to_multiplier <- function(exp) {
  multiplier <- switch(exp,
                       "K" = 1e3,
                       "M" = 1e6,
                       1)
  return(multiplier)
}

# Convert the letters in PROPDMGEXP to numeric multipliers
multipliers <- sapply(new_df1$PROPDMGEXP, exp_to_multiplier)

# Multiply PROPDMG,  by the multipliers to get total property damage
new_df1$total_prop_damage <- new_df1$PROPDMG * multipliers
new_df1$total_crop_damage <- new_df1$CROPDMG * multipliers

# Determine unique event types in EVTYPE
unique_events <- unique(new_df1$EVTYPE)

# Initialize an empty data frame to store the results
event_totals1 <- data.frame(EVTYPE = character(),
                            total_prop_damage = numeric(),
                            total_crop_damage = numeric(),
                            stringsAsFactors = FALSE)

# Loop through each unique event type
for (event in unique_events) {
  # Subset the data for the current event type
  subset_data <- new_df1[new_df1$EVTYPE == event, ]

  # Calculate total property damage and crop damage
  total_prop_damage <- sum(subset_data$total_prop_damage, na.rm = TRUE)
  total_crop_damage <- sum(subset_data$total_crop_damage, na.rm = TRUE)

  # Add the totals to the results data frame
  event_totals1 <- rbind(event_totals1, data.frame(EVTYPE = event,
                                                   total_prop_damage = total_prop_damage,
                                                   total_crop_damage = total_crop_damage))
}
```

```
# Add a new column for total damage
event_totals1$total_damage <- event_totals1$total_prop_damage +
event_totals1$total_crop_damage

head(event_totals1)
```

Question 2: Across the United States, which types of events have the greatest economic consequences?

```{r, echo = TRUE, cache = TRUE}
# Order Total_Damage from highest to lowest
event_total_damage <- event_totals1[order(-event_totals1$total_damage), ]

head(event_total_damage)
```

Results Question 2:

- event_total_damage <- event_totals1[order(-event_totals1$total_damage),]
- Using the command to order the weather event with the most fatalities, results can be interpreted from the table.

EVTYPE	total_prop_damage	total_crop_damage	total_damage
TORNADO	51625660796	28269878512	79895539307
FLOOD	22157709930	53751978402	75909688331
FLASH FLOOD	15140812068	38822140319	53962952387
HAIL	13927367054	15314323923	29241690977
TSTM WIND	4484928495	7684657862	12169586357
HIGH WIND	3970046296	7174066263	11144112559

Conclusion:

After analyzing the data, it is evident that tornadoes caused the highest number of fatalities and injuries, as well as the most damage to property and crops.