***Storm and Climate Data Record (SCDR)***

Implementation Plan

Date of record:

Jan 1, 2017 to Dec 1, 2019

Principal Investigator: David France

## I. Introduction

### 1.1 Overview and purpose

The Miami police department has requested a report on the rising costs of crimes when storms are occurring. This report will provide information on crimes committed during storms and crimes committed when no storms are present. It will further detail the cumulative monetary loss to victims of each category over time. It will conclude with suggestions for further analysis. The police department will use this report to attempt to combat future crimes and lower the overall costs to the community.

### 1.2 Define why you need data analysis

Data analysis is a fundamental tool in making educated decisions for both public and private institutions. It includes both data mining and predictive analytics. Data mining allows us to perform several tasks on the data, including description, estimation, prediction, classification, clustering, and association. Predictive analytics then takes that information to make an educated guess at future outcomes. One example of this in use is the Commonwealth of Massachusetts using predictive analytics to cut down on future Medicaid fraud (Larose & Larose, 2015).

In this scenario the Miami police department has a similar problem. They are looking to cut down on costs to crime victims by using predictive analysis. They have a hypothesis about the connection between storms and crimes. They think the costs to victims are higher during storm activity than when no storms are present and are continuing to rise. We can use data analysis to confirm or deny that hypothesis, thus helping the Miami PD efficiently use their resources to combat criminal activity. We could also use the analysis to suggest further areas of inquiry as we gain insights during our exploration of the data.

## II. Data Preparations

### 2.1 Name data sources

The data comes from the Miami police department. They have provided us with data that details crimes, dates, storm activity, and loss during the time frame of January 2017 – January 2019.

### 2.2 Filter through unnecessary data

The data we need is crime activity when storms occur and when they don’t occur. Two files were created - one from a query where crime data was selected with a Storm\_Id present, and one from a query where no Storm\_Id was present. The relevant fields at this point were Date and Loss. Summing the Loss by month for each data set gave us the filtered data in crimenostormQ.csv and crimestormQ.csv. We will be able to use those files when we run our R scripts.

### 2.3 Define your parameters

Parameters are the variables we will use in our analysis. In this case, we are using Crime\_Id, Storm\_Id, Date, and Loss. Crime\_Id will be used to count the number of crimes committed both when there is a Storm\_Id associated, and when there is not. If there was also a Storm\_Id, that means the crime was committed during storm activity. No Storm\_Id means no storm activity was present. Loss will be used to track the cumulative expense of the crimes committed – summed by month, and Date will be used to create a time series for crimes committed during storms and when there are no storms.

### 2.4 Identify measurement priorities

The Miami PD wants a report on the rising costs of crimes that occur during storms. The measurement priorities are the cumulative victim loss in thousands of dollars from January 2017 – January 2019 when there was storm activity present, as well as the cumulative victim loss over the same time frame when no storm activity was present.

### 2.5 Ensure collected data fits the need

A visual inspection of the data shows it has the information we need. After analyzing the data, both in loading the files and creating a visual, we should be able to confirm that this is correct.

## 

## III. Data Analysis

### 3.1 Identify scripts used

Using R-Studio to analyze the data, the following scripts were used:

**// Install and load necessary packages to create time series and plots**

install.packages("tframe");

install.packages("tfplot");

library("tframe");

library("tfplot");

**// Set working drive**

setwd("C:/Users/Public/Desktop/DAT-375")

**// Load data for crimes with storms and print data to screen to confirm it is correct**

crimestormdataQ <- read.csv("crimeStormQ.csv")

print(crimestormdataQ)

**// Load data for crimes without storms and print data to screen to confirm it is correct**

crimenostormdataQ <- read.csv("crimenostormQ.csv")

print(crimenostormdataQ)

**// Create time series for both data sets, summing total loss in thousands starting January 2017 // and plotting the data by month**

z<-ts(cumsum(crimestormdataQ$Loss)/1000,start=c(2017,1), frequency=12)

x<-ts(cumsum(crimenostormdataQ$Loss)/1000,start=c(2017,1), frequency=12)

**// Plot both time series sets in a line graph**

tfplot(z,x,

ylab="Victim Monetary Loss in $K",

xlab="By Month and Year",

title="Victim Loss From Crimes for Jan 2017 - Dec 2019",

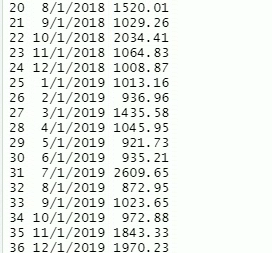
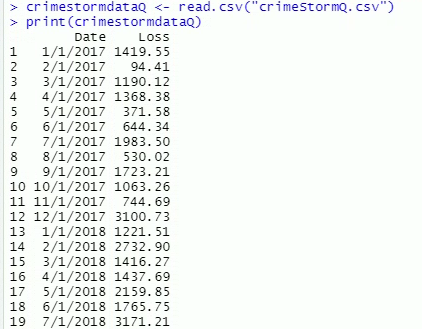
subtitle = "Cumulative Loss in Thousands of Dollars",

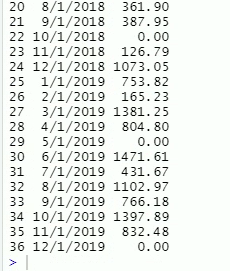
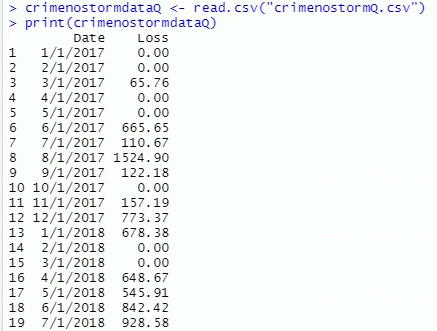
legend=c("Crimes During Storms (black)", "Crimes When No Storms (red)"),

source="Source: DAT Data")

### 3.2 Run the scripts to analyze the data and validate the output

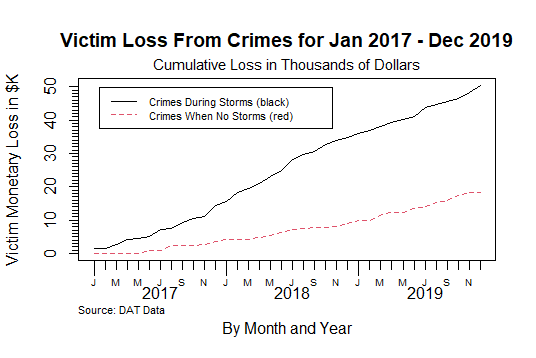
The scripts were run in R-Studio. Looking at the data for crimestormdataQ and crimenostormdataQ, we can validate that the data was loaded correctly. Each data set has 36 rows with the first of each month representing the data and the total loss for that month in the Loss colomn:





## IV. Drawing Conclusions

### 4.1 Present the results of the analysis to stakeholders



Looking at the graph of the data, it is clear to see the cost of crime during storms (in black) is much higher than with no storms (in red), and the difference is increasing. The slope of the Crimes During Storms line appears to increase during the period of December 2017 to July 2018. A look at the raw data confirms this range. The slope of the Crimes When No Storms line, by comparison, appears to remain relatively constant. This is the non-hurricane season. It is possible the criminals aren’t able or are unwilling to operate during hurricanes, as opposed to less dangerous storm events. A further look into the type of crime and storm activity could yield further valuable information for the department.

### 4.2 Determine whether the problem was addressed, including any challenges and limitations

Looking at the graph, we can see that the hypothesis of the Miami PD is confirmed. The costs to victims are higher during storm activity and continue to rise. We can also see the real costs of crimes during storms v. crimes with no storms.

The data from the police department was clean and easy to work with, so there were not any challenges to speak of. One limitation of this study is that there is a lot of storm activity in Florida overall. It is reasonable to say that the cost of crimes during such activity will be higher simply because there is storm activity present nearly every day. This can be seen when looking at the crimenostormQ data – there are 10 months out of 36 that have 0.0 loss due to criminal activity. The most likely explanation for this is there was storm activity every day during those months. Controlling for this factor in the future would be wise.

Another limitation is we didn’t investigate crime type or storm type in this study. It would be beneficial for the department to know if certain crimes increased or decreased during the different types of storm activity. Knowing this information would allow the department to better focus their efforts.

### 4.3 Report potential new findings

It looks like not only does crime increase during storms, resulting in a large economic cost, but that increase widens during the non-hurricane season. The department might want to consider increasing its presence during non-evacuation storm activity.

For future studies, the department has a few more things they could investigate. They should look closer into what types of crime increase/decrease during different types of storm activity. They should also examine the months that have no monetary loss at all in conjunction with controlling for the prevalence of storm activity in the area. I would recommend looking at the overall economic loss due to crime, then seeing how the percentage of loss during storm activity compares to the percentage of time there is storm activity. Having a more detailed picture of what is happening will allow the Miami PD to spend limited resources in the most effective way.

# References

Larose, D. T., & Larose, C. D. (2015). *Data Mining and Predictive Analytics.* Hoboken, NJ: John WIley & Sons, Inc.