Assignment 1. Examining data using SPSS

General: This is a group assignment (please use the same groups you have for your final project and time-series). All group members are expected to participate. (Zero tolerance free ridders).

Format: submitted report should include a detailed discussion of the techniques you used to answer the suggested questions (See below). Include only relevant graphs and tables (of course with a discussion of these). Any additional insights into your data is positively rewarded.

Objective: A detailed examination of data to identify and treat missing values as well as outliers. Test assumptions in data namely linearity, normality as well as homoscedasticity.

Dataset: US Mass Shooting

Source: https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years

Context: From 1996 to 2016, The US has witnessed 398 mass shootings resulting in 1,996 deaths and 2,488 injured. The latest and the worst mass shooting of October 2, 2017 killed 58 and injured 515 so far. The number of people injured in this attack is more than the number of people injured in all mass shootings of 2015 and 2016 combined. The average number of mass shootings per year is 7 for the last 50 years that would claim 39 lives and 48 injured per year.

Content:

Geography: United States of America

Time period: 1966-2017

Unit of analysis: Mass Shooting Attack

Dataset: The dataset contains detailed information of 398 mass shootings in the United States of America that killed 1996 and injured 2488 people.

Variables: The dataset contains Serial No, Title, Location, Date, Summary, Fatalities, Injured, Total Victims, Mental Health Issue, Race, Gender, and Latitude -Longitude information.

There are five versions of this dataset; you will be using the latest one, provided as a CSV file on Campus Online / Additional Documentation / Assignment 1.

Questions to consider:

- How many people got killed and injured per year?

- Is there any correlation between shooter and his/her race, gender?
- Any correlation with calendar dates? Do we have more deadly days, weeks or months on average?
- What cities and states are more prone to such attacks?
- Do you have any missing values in your data? Determine extent as well as pattern of randomness. Discuss techniques used to impute data and rational behind it.
- How many shooters have some kind of mental health problem?
- Can you model the number of killed (and/or injured people) based on the data you collect?

Submission date: October 25th, 2019 by 5:00pm. NO LATE SUBMISSION WILL BE ALLOWED.

A turnitin assignment is already created for that purpose (Campus Online/Assignment).

One submission per group.