Projectplan

Jonas Kouwenhoven
Martijn Maiwald
Folkert Stijnman
Casper Wortmann

June 2018

Log

05/06/2018

After the first meeting with Nick van de Wolff (Later on revered to as 'Nick'), we started the project of with getting git hub to work. Casper created the public repository: 'DAV_23' and invited the rest of us. After Github was working we discussed the changes we wanted to make to the dataset and called it a day.

06/06/2018

The day started off with the classification of all datatypes. On the Github page of *jamesgo*[1] there is a table that helped us in this progress. Later on we discussed the differences between JSON and CSV, so we could make a better decision between the two of them. We called it a day after a unanimous decision for JSON.

07/06/2018

The group met up at the coffee corner on Science Park to discuss the questions we wanted to ask Nick. He then helped us to convert the file from CSV format to a JSON format with the help of 'Nested Keys'. After the meeting we went on to clean up the dataset a bit more, and got rid of all the unnecessary enters and punctuation. We also got rid of all columns containing URLs e.g. *incident_url* because they are unnecessary for our research.

08/06/2018

We spent the day to clean up the dataset as much as possible. Now that the dataset is in order we can start programming right away next week.

---

[1]https://github.com/jamesqo/gun-violence-data

Research idea's

1. Correlation between surname's and incidents.

2. Incidents correlated to universities/campus/students with the help of keywords like: 'campus','university','school','student','professor','teacher'.

3. Participant relationships; family,friends,co-workers,gang vs gang, acquaintance, significant others.

4. Potential forecasting of incidents, the course of gun violence in the future.

5. Woman/man distribution in family incidents.

Steps taken

1. **From CSV to JSON**: The provided dataset was in a CSV Format, but the discovery was made that Panda[2] would work better with a JSON format.

2. **Deleting columns**: Columns containing url's were deleted, because it looked messy and they where not needed.

3. **Tidy up**: By removing unnecessary enters and punctuation the data set got a better appearance.

4. **Empty cells**: Because empty cells could cause some problems during plotting and programming, we replaced them all with 'NaN' or 'Unknown'.

5.

---

[2]https://pandas.pydata.org/