

<b>LOG</b>	<b><i>DA-DV</i></b>
TA	<u><i>Willemijn Becks</i></u>
GROUP	<i>15</i>
Participants	<i>Sarah Yeguez</i> <i>River Vaudrin</i> <i>Jelle Bosscher</i> <i>Serge van Haag</i>

*This is a logbook which includes the decisions that are made, solutions to problems and each step of the process is described per day.*

*According to canvas, only the logbook was required to submit and nothing was mentioned about the process book.*

*However, after we had contact with our TA, we decided to process quite a few things of the process book in our logbook (in example screenshots).*

Date	Topic of Interest	Indication of Timespan	Planning
------	-------------------	------------------------	----------

<p>DAY ONE</p> <p>06/04/18</p>	<p>The four of us thought of questions to ask.</p> <p>We met with our TA</p> <ul style="list-style-type: none"> <li>- Asked how we need to start.</li> <li>- Questions about the project in general.</li> </ul> <p>Created a repository on GitHub.</p> <p>Made a journal on Google Docs.</p>	<p>2 HOURS</p>	<p>Think about questions to ask our TA</p> <p>Meeting with our TA</p> <p>Create a repository</p> <p>Create a journal</p> <p>Think about which dataset to choose</p>
<p>DAY TWO</p> <p>06/05/18</p>	<p>Jelle taught us how to work with Github.</p> <p>Everyone made branches on Github.</p> <p>We choose for the dataset: Gun Violence in the USA.</p> <p>River found out that Panda fully supports our dataset (CSV).</p> <p>Updated the journal.</p>	<p>2 HOURS</p>	<p>Serge, Sarah and River need to understand github.</p> <p>Choose a dataset</p> <p>Check if Panda supports the format</p> <p>Think about questions to ask Willemijn</p>
<p>DAY TREE</p> <p>06/06/18</p>	<p>We met with our TA.</p> <p>We did a few things with the dataframe:</p> <ol style="list-style-type: none"> <li>1.Counted the missing classes.</li> <li>2.We are changing address to street.</li> <li>3. There is a problem with age of participants because it can contain more values. We are thinking about it how to handle this as a team.</li> <li>4.Thought about what to do with missing values: <ul style="list-style-type: none"> <li>• If long/lat is not missing, but street is missing; then we can fill in the street value by making use of the long/lat</li> </ul> </li> </ol>	<p>3 HOURS</p>	<p>Cleaning data</p> <ul style="list-style-type: none"> <li>- Missing values</li> <li>- Correcting inconsistent data</li> </ul> <p>Meet with Willemijn</p>

	<ul style="list-style-type: none"> <li>Other missing data will be changed to 'Unknown'.</li> </ul> <p>Updated the journal.</p>		
<p>DAY FOUR</p> <p>06/07/18</p>	<p>Sarah found Health Care data. Sarah found Population data.</p> <p>Jelle cleaned the data for 80%.</p> <p>River and Serge started to work in a Jupyter notebook.</p>	2 HOURS	<p>Clean the data</p> <p>Search for datasets regarding:</p> <ul style="list-style-type: none"> <li>Health Care per state</li> <li>Population per state</li> <li>Gun licenses distributed per state</li> </ul> <p>Understand Pandas and create a working environment for pandas.</p>
<p>DAY FIVE</p> <p>06/08/18</p>	<p>Visualize an overview of the importance and completeness of the dataset.</p> <p>Sarah scraped Health Care data.</p> <p>Jelle started on a script that should clean the rest.</p> <p>River and Serge got a few charts from pandas in Matplotlib.</p> <p>Updated the journal.</p>	4 HOURS	<p>Clean the final part of the data.</p> <p>Reduce the data if possible</p> <p>Keep searching for datasets regarding:</p> <ul style="list-style-type: none"> <li>Health Care per state</li> <li>Population per state</li> <li>Gun licenses distributed per state</li> </ul> <p>Create a few graphs using pandas and matplotlib.</p>
<p>DAY SIX</p> <p>06/09/18</p>	<p>Jelle cleaned the data.</p> <p>Sarah cleaned the Health Care data.</p> <p>Updated the journal.</p>	1.5 HOURS	<p>Clean the final part of the data since it took more work than expected.</p>
<p>DAY SEVEN</p>	<p>Sarah scraped and cleaned Population data.</p>	1 HOUR	<p>Make sure all data is consistent and clean.</p>

06/10/18	Serge checked if no data was missing.  Updated the journal.		
----------	---	--	--

<p>DAY EIGHT</p> <p>06/11/18</p>	<p>Meeting Willemijn</p> <ul style="list-style-type: none"> <li>- Willemijn said that mental health care might be interesting as well.</li> </ul> <p>Serge and River read the EDA.</p> <p>Made an extra question about mental health.</p>	<p>3 HOURS</p>	<p>Clean the data further.</p> <p>Read the EDA</p> <p>Think of extra questions.</p>
<p>DAY NINE</p> <p>06/12/18</p>	<p>Every member of the team read the EDA.</p> <p>Serge and River created some explorative questions.</p> <p>Updated the journal.</p>	<p>3 HOURS</p>	<p>Read the EDA</p> <p>Creating explorative questions for the EDA and categorize them in one of the four categories. (univariate, multivariate, graphical, non-graphical)</p>
<p>DAY TEN</p> <p>06/13/18</p>	<p>Applied techniques discussed in the EDA to the three main questions.</p> <p>Updated the journal.</p>	<p>2 HOURS</p>	<p>Apply knowledge from the EDA to the three main questions of our dataset.</p> <p>Make sure we can show our TA more in-depth graphs.</p>
<p>DAY ELEVEN</p> <p>06/14/18</p>	<p>Met with Willemijn.</p> <ul style="list-style-type: none"> <li>- Displayed a few graphs</li> <li>- We need to make some values more clear</li> </ul> <p>Sarah read about clustering.</p>	<p>3 HOURS</p>	<p>Meet with our TA.</p> <p>Read about various machine learning techniques.</p> <p>Sarah started doing research on clustering.</p>

DAY TWELVE 06/15/18	Updated the journal.  Sarah clustered a couple of columns using k-means techniques.		Sarah is going to apply clustering techniques.
DAY THIRTEEN 06/16/18	Q1 is answered for states and cities, Serge still need to answer the years-part of the question.  River found out that there is a peak in violence on the 1th of January and 4th of July. Also found some other notable data needed to answer Q2.  Sarah tried to implement clustering techniques in Python.	4 HOURS	Serge answers Q1.  River answers Q2.  Jelle and Sarah work further on clustering.
DAY FOURTEEN 06/17/18	Sarah reduced the clustering implementation code.  Updated the journal.  Jelle is working on getting geo-locations from the dataframe.	2 HOURS	Jelle will cluster geo-locations in order to display an interactive map of the USA.

DAY FIFTEEN 06/18/18	Plotted graphs for question 4  Jelle created a map that is working by using the Google Maps api.	2 HOURS	Work on plotting gun violence incidents per capita for question 4.  Work on clustering.
DAY SIXTEEN 06/19/18	Meeting Willemijn  Serge and River read more about: - Linear Regression	6 HOURS	Meet with Willemijn.  Ask our TA questions about regression and clustering.

	<ul style="list-style-type: none"> <li>- Polynomial Regression</li> </ul> <p>Jelle en Sarah read in-depth about:</p> <ul style="list-style-type: none"> <li>- Machine learning</li> <li>- Clustering</li> </ul> <p>Updated the journal.</p>		Read on regression and clustering as well.
<p>DAY SEVENTEEN</p> <p>06/20/18</p>	<p>Jelle, River and Serge met because Sarah had physiotherapy.</p> <p>We wrote down a few todo's which we shared with Sarah as well.</p>	3 HOURS	Meet with the team to discuss the progress and todo's of this week.
<p>DAY EIGHTEEN</p> <p>06/21/18</p>	<p>Meeting Willemijn</p> <p>Everyone met at Sarah's house.</p> <p>Applied techniques:</p> <ul style="list-style-type: none"> <li>- Polynomial Regression</li> <li>- Linear Regression</li> <li>- K-means (clustering)</li> </ul> <p>Updated the journal.</p>	5 HOURS	<p>Meet with our TA</p> <p>Go to Sarah's house</p> <p>Work on making a regression lines.</p> <p>Clustering map is done.</p>
<p>DAY NINETEEN</p> <p>06/22/18</p>	<p>River made a report in ShareLatex.</p> <p>River and Serge added labels to graphs.</p> <p>Everyone searched for interactive alternatives for Matplotlib.</p> <ul style="list-style-type: none"> <li>- Bokeh</li> <li>- Plotly</li> <li>- ggPlot</li> </ul>	3 HOURS	<p>Add labels to graphs.</p> <p>Find out how we are going to make graphs interactive in HTML.</p> <p>Read about data visualisation for the website.</p> <p>Added labels to graphs.</p>

<p>DAY TWENTY</p> <p>06/23/18</p>	<p>We choose to use plotly instead of Matplotlib.</p> <p>Tried to implement Plotly in jupyter.</p> <p>Updated the journal.</p>	<p>1.5 HOURS</p>	<p>Choose an alternative to matplotlib for interactive plotting.</p>
<p>DAY TWENTY ONE</p> <p>06/24/18</p>	<p>River and Jelle wrote a concept for the introduction and method.</p> <p>Serge plotted charts and dates in plotly.</p> <p>Sarah finished clustering characteristics of each incident.</p>	<p>4 HOURS</p>	<p>Create a basis for the report and start with introduction and method.</p> <p>Make charts that support the answers to questions.</p> <p>Work on clustering.</p>

<p>DAY TWENTY TWO</p> <p>06/25/18</p>	<p>Met at Jelle's house.</p> <p>Report layout is completely done by River.</p> <p>Serge created graphs using Plotly.</p> <p>River plotted a regression line in plotly.</p> <p>Jelle is working on question three.</p> <p>Website is now accessible via Github.</p> <p>Updated the journal.</p>	<p>4 HOURS</p>	<p>Meeting at Jelle's house.</p> <p>Work on the Report.</p> <p>Work on answering the questions.</p> <p>Host the website on github.</p>
---------------------------------------	--	----------------	--

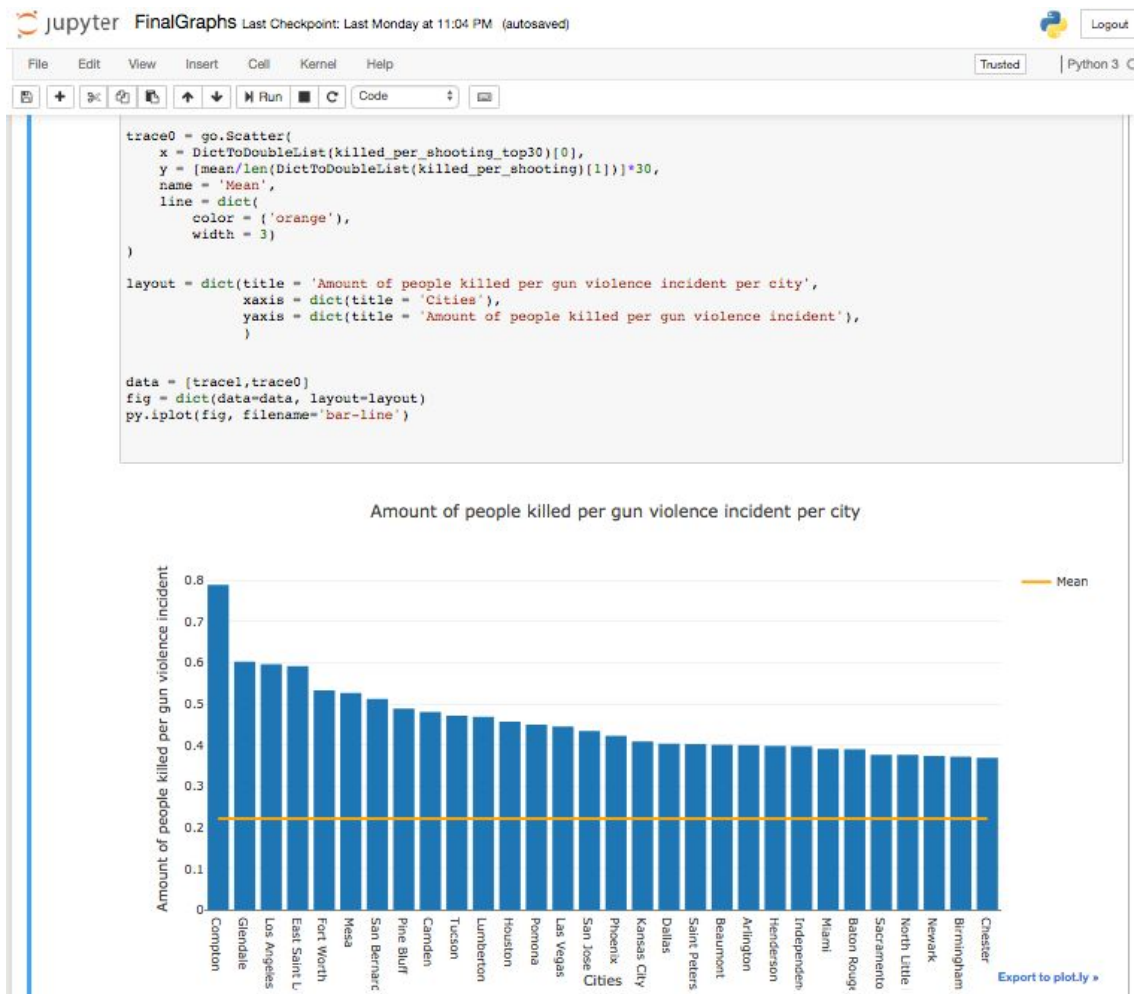


<p>DAY TWENTY THREE</p> <p>06/26/18</p>	<p>Meeting Willemijn</p> <ul style="list-style-type: none"> <li>- She told us that we should write down mathematical methods in the method part of the report. Also that results are purely factual and that the discussion is needed for interpretation.</li> </ul> <p>Jelle did not feel well so we met at Sarah's place.</p> <p>Worked on graphs.</p> <p>River made a website</p>	<p>5 HOURS</p>	<p>Meet willemijn</p> <ul style="list-style-type: none"> <li>- Ask about report</li> </ul> <p>Meet at jelle</p> <p>Work on graphs.</p> <p>Make a website.</p>
<p>DAY TWENTY FOUR</p> <p>06/27/18</p>	<p>Uploaded graphs to our website.</p> <p>Added descriptive text to the site.</p> <p>Updated the journal.</p>	<p>3 HOURS</p>	<p>Make sure that we got all the graphs for the website.</p> <p>Put text on website</p>
<p>DAY TWENTY FIVE</p> <p>06/28/18</p>	<p>Meeting Willemijn</p> <ul style="list-style-type: none"> <li>- Asked about our report and our website</li> <li>- We got advice on what to do this day.</li> </ul> <p>Updated the journal one last time.</p> <p>Serge is working on the website.</p> <p>Everyone is writing in the report.</p> <p>Sarah puts her data in the report and on the website.</p>	<p>10 HOURS</p>	<p>Finish the website</p> <p>Finish the report</p>

DAY TWENTY SIX 06/29/18			Presenting our study.
----------------------------	--	--	-----------------------

# Screenshots:

Creating graphs with plotly in Jupyter.



We wrote the report in LaTeX and used ShareLatex to share updates with each other.

Technisch Rapport Dream T...

J S R Review History Ch

Recompile

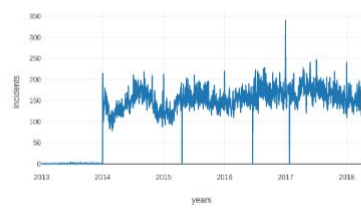
```
273 \includegraphics[width=\textwidth]{images/Plot361.png}
274 \caption{...}
275 \end{figure}
276 \subsection{Most dangerous states taking population into consideration}
277 One could say that the number of incidents in a state defines how
threatening or dangerous a state is. On the other hand, the number of
deaths per shooting might define the real danger of a state. Considering
the population of a state, every graph is plotted per capita. Graph one
shows the number of gun violence incidents per capita for every state,
while graph two shows the number of people killed per incident for every
state.
278 \begin{figure}[H]
279 \centering
280 \includegraphics[width=\textwidth]{images/plot371.png}
281 \caption{...}
282 \end{figure}
283 \begin{figure}[H]
284 \centering
285 \includegraphics[width=\textwidth]{images/Plot372.png}
286 \caption{...}
287 \end{figure}
288 \subsection{The relation between gun-laws and gun-violence}
289 Trump stated this on October 19, 2016. Graph one makes it obvious that
Chicago has, by far, more gun violence incidents than any other city from
2014 up to October 19, 2016.
290 \begin{figure}[H]
291 \centering
292 \includegraphics[width=\textwidth]{images/plot381.png}
293 \caption{...}
294 \end{figure}
295 According to scraped data per year from State firearm Laws we found out
that Illinois, the state where Chicago is located, has a mean amount of
gun laws of 65 over the years 2014-2016; see table 4. Taking into
consideration all states, Illinois is on the seventh spot. However, the
law center wrote that Illinois ranks eight if you look at the toughest gun
laws per state.
296 \begin{center}
297 \includegraphics[width=\textwidth]{images/table382.png}
298 \captionof{table}{Your caption here}
299 \end{center}
300 \section{Discussion}
301 % Question 1 -
302
303 % Question 2 - Interesting patterns
304 So it seems that victims in overall get mostly injured and suspects get
mostly killed. Moreover if an incidents does not have any killings, one
can assume that no females were involved but if females mainly take part
in incidents where there are more injuries than killings.
305
306 Furthermore, gun violence report during the holiday seasons in December
and January are far lower than other months. Holidays around this time
tend to make people more caring and compassionate towards each other.
What's more, people tend to have more free time during the weekend rather
than during the week. Which explains why incidents are more active on
```

The data about mental health in the U.S per state was acquired from the 'Mental Health America' website[3]. The data shows a table with states ordered by ranking, from better to worse, alas the data consisted only of year 2018. The data showing the gun laws per state were found on the website of the State Firearms Laws[2], it depicts a map of the number of gun laws per state. The data taken covered the same time span as the data about gun violence, 2013 to 2018.

The final dataset was taken off the US Census Bureau website[4]. The data shows a table of the population per state per year from 2010 to 2017. The data was then converted to estimate the population per state per month from January 2013 to and including December 2017. This data was used to calculate per capita, such as incidents per capita.

## 2.2 Data processing

To properly use the data it first had to be cleaned. As can be seen in figure 1, the data before was inconsistent. The data sheet only showed a maximum of five incidents for any day in 2013 even though there were more. Therefore, the analysis of the data has only been applied to 2014 and onward.



We used the documentation of plot.ly to understand plotly

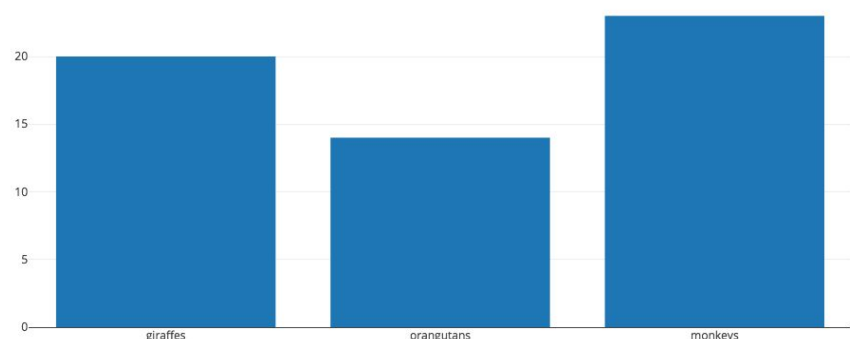
Beveiligd | <https://plot.ly/python/bar-charts/>

plotly Developer Support Workshops Consulting SIGN IN SIGN UP REQUEST DEMO

API Libraries Python Bar Charts Fork on Gith

```
py.iplot(data, filename='basic-bar')
```

Out[2]:



Species	Incidents
giraffes	20
orangutans	14
monkeys	22

EDIT CHART