Portfolio Final Version

**Crime and the Media**

Research question

1 Which regional news source is the most accurate in representing crime abuse and theft, compared to the registered crime in the region Amsterdam in the years 2012 and 2013?

This research will be based on the crime statistics on abuse and theft in Amsterdam in the years 2012 and 2013. The three (digital) regional news sources for comparison are Het Parool newspaper, information from the Amsterdam police and AT5. These sources were chosen because they all broadcast news about and for Amsterdam citizens. Thus the news sources were qualified as regional news. This question is researched by Sarah Hazm.

1.1 To what extent do the news sources align with the reported feelings of insecurity of Amsterdam citizens in the years 2012 and 2013?

The regional news sources will be compared with reported feelings of insecurity regarding theft and abuse in government census held between August and November. This period has been selected based on the availability of the census data on feelings of insecurity. This question was answered by Hugo Maat.

The date of the first research question on registered crime in Amsterdam will by obtained through the Dutch statistical institute: Central Bureau of Statistics. The website of the institute supports selection of specific data which can be downloaded. Available extensions are .csv or .xlsx. So far the ordinary excel extension proved to be complicated for visualization, so we have relied on csv. Data has been selected for headings of registered theft and abuse in the years 2012 and 2013 of Amsterdam.

Media news reports on crime are used from Het Parool, Police(website) and AT5. These sources have been selected for its user-friendly interface including Amsterdam specific search options, enabling exclusion of news from other regions.

None of the regional news sources are accessible through API. I have attempted to obtain an alternative, as I wanted to read the search results represented on the website in a (downloadable) format that could be read through the Terminal. This has proved unsuccessful. We eventually generated useful articles by selecting it by ‘hand’ and counting the relevant articles about abuse and theft. The document of the counted articles is available in GitHub ‘News sources data’.

The following is a brief description per media source on the method of retrieving the required information.

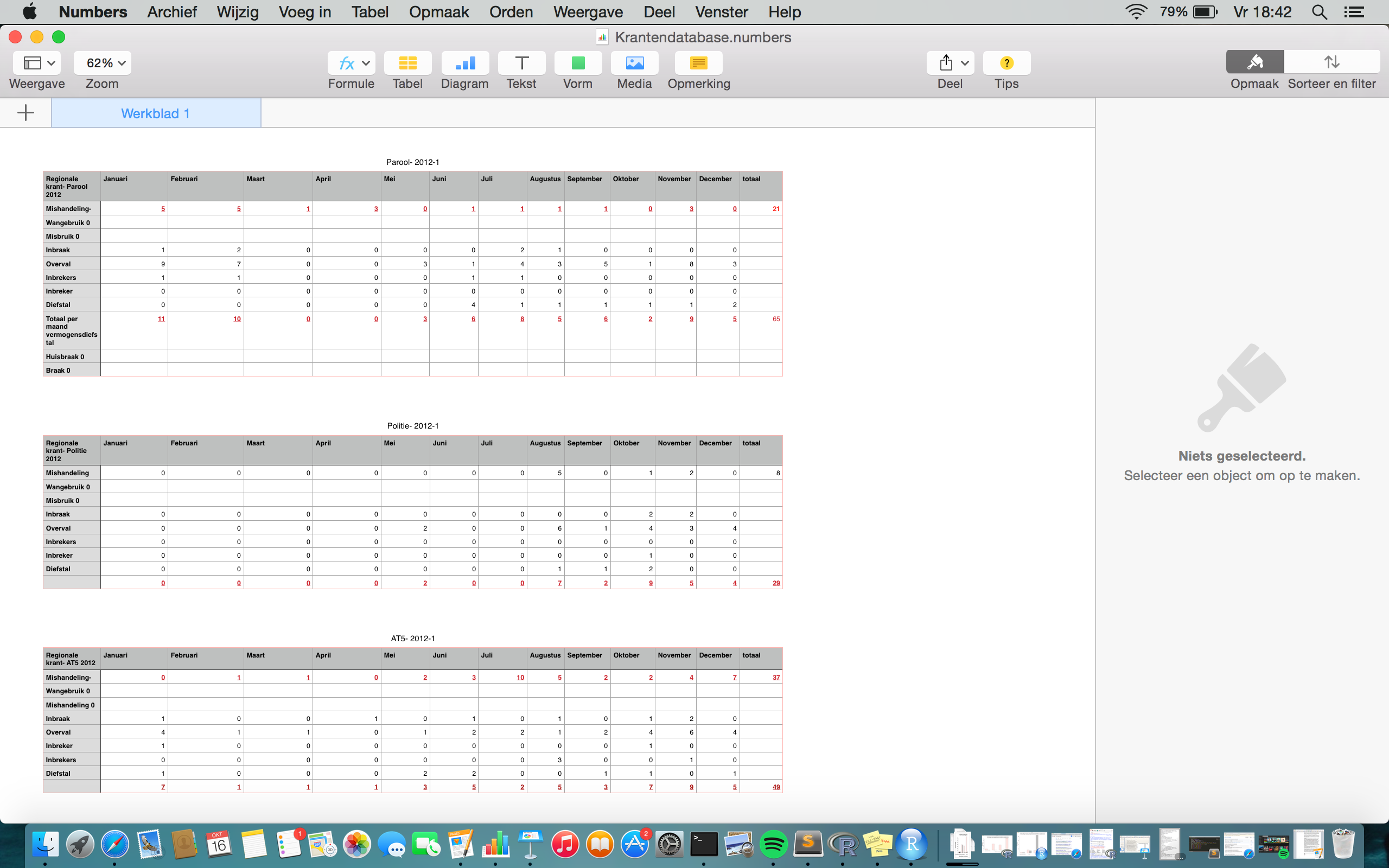
Het Parool: I have used Google search engine for the following keywords: “overval, inbraak, inbrekers, mishandeling” (robbery, burglary, burglars, abuse) within the consignment of Amsterdam crime. The keyword search for abuse, for instance through Google, is as follows: *site:parool.nl/parool/nl/7/MISDAAD/ “Amsterdam” “mishandeling”*. It became apparent that some of the results were irrelevant as several of the results only contained references to other search results which in themselves showed up as part of the search. Because of this double result the articles had to be screened to prevent pollution of the database. As this proved very time-consuming, less time has been available thus far to work on the visualization part for the first portfolio.

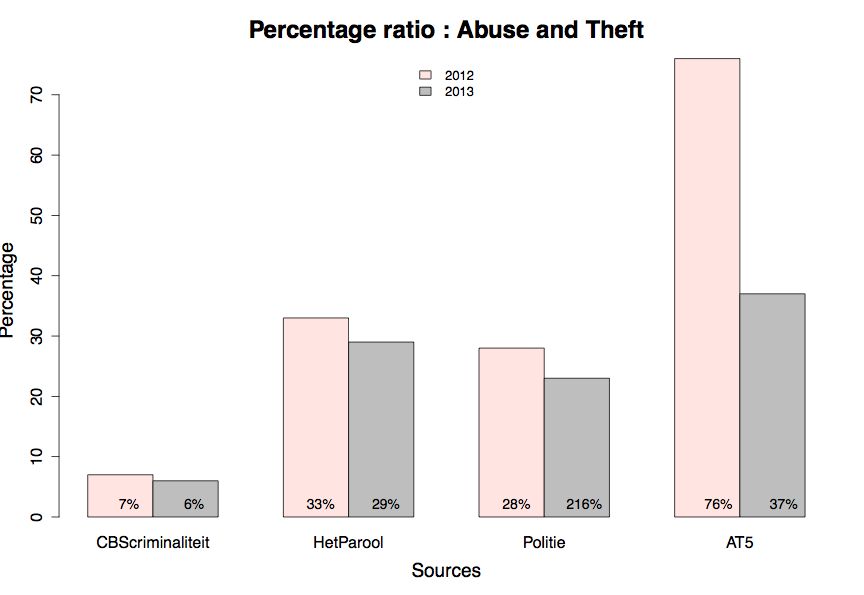
Police: It is uncomplicated to specify the searches on the official website. *The interface was user-friendly for searching news items.* A number of results proved irrelevant on this site as well. *This source is not the same as the data from the CBS, but the police news source is, like the other two news sources, also a website that publishes several articles about the committed crime.*

AT5: Distilling relevant research results from the searches on this site proved time-consuming as well. *We have also generated these articles through Google with the following query: site:at5.nl "Amsterdam" "inbreker".*

All the relevant results on crime and theft for 2012 and 2013 have been counted and manually entered into a document as shown aside.

Retrieving Data





Visualization

For research question 1 I began entering the figures for registered crime with a .csv extension in the terminal. The results were not represented in a clear manner. The next step was to feed the file into R (Crime <- read.csv(Crimi.csv)). Several attempts to visualize this .csv file in R failed as well, prompting me to enter the data in R manually. This has resulted in the script as uploaded in Github: ‘Script1.R’. The approximate result I wished to create in R was as follows:

I had some problems to recreate the first scripts version into a hist bar. There was a recurring error message with every attempt ‘there was no “x” value detected’. I needed to experiment with other codes.

The formula for ‘x’ needed to be simply written down in an other code format (Script1.R). The file ‘ReadFirstScript1’ in GitHub describes the steps I took. In the final script the values for news sources were also added. Eventually through many efforts this is the final result via ‘R’ also shown in the file ‘PDFresearch1’ via GitHub:

Visualization progress Hugo

The second research question was divided for Hugo Maat. The purpose of this section was to portrait the feeling of insecurity of Amsterdam citizens and the numbers of reported crime in news sources. The visualization should illustrate a possible relation between this two subjects. The illustration shown aside is an example on how to possibly give answer to the second research question.

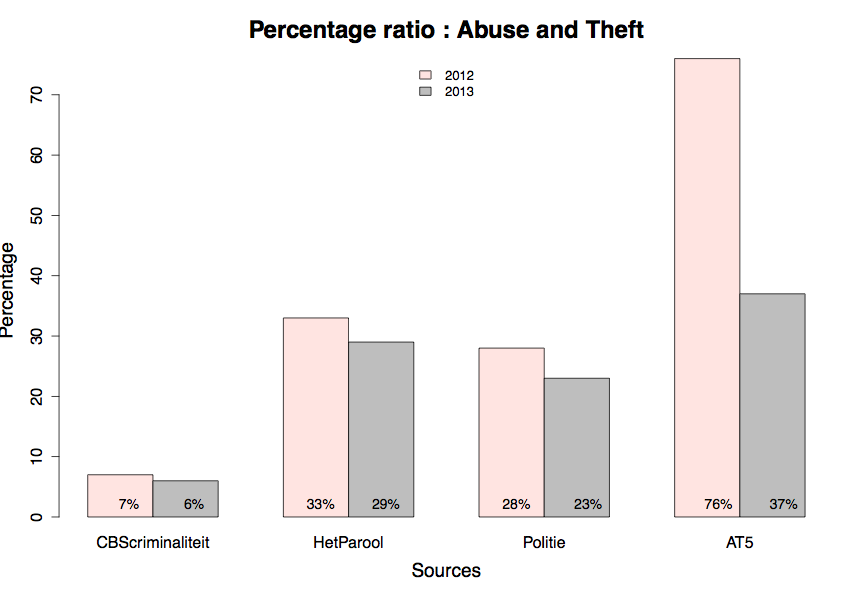
|  |  |  |  |
| --- | --- | --- | --- |
| Step 1:  Formulating research question | Stap 2: Acquiring data | Step 3: Structuring data | Step 4: Visualizing data |
| The research question has been formulated | So far only the data on 2012 has been collected. | Data is yet to be properly structured, as it has not been possible to visualize it in R. | Visualize the data in bar or line graphs using R. |
|  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Step 1:  Formulating research question | Stap 2: Acquiring data | Step 3: Structuring data | Step 4: Visualizing data | Step 5:  Presentation of the results in class | Step 6:  Creating the final report |
| The research question has been formulated | All the data of the regional news source and CBS statistics is been retrieved. | Data is structured in a way that ‘R’ recognizes it to visualize the data. | Visualize the data in bar or line graphs using R. | Showing our project in class. | Adding new progress information into the portfolio and adding the feedback and unclarity from the class after the presentation. |
|  |  |  |  |  |  |

Steps

Steps undertaken for the first portfolio

Steps undertaken for the final portfolio



The table below illustrates the answer to the first research question: *Which regional news source is the most accurate in representing crime abuse and theft, compared to the registered crime in the region Amsterdam in the years 2012 and 2013?*

The percentages in this table represent the ratio between the abuse and theft by dividing abuse by theft per each source. The main focus relays on the representation between registered crime and news sources. The percentage is a relative number instead of an absolute number, because now with the relative numbers the relationship and thus the research question is better illustrated. If the numbers in the table were absolute the results were shown out of proportion and not relevant to be shown to support the question.

The percentages only give an illustration about the articles that were chosen to be relative towards the search query. There could only be conclusions made based on those regional sources chosen: The Parool, Politie, and AT5 within the years 2012 and 2013.

You could interpret the table as following. The percentage of the news sources that are the same as to the registered crime shows an equal news report about abuse and theft in Amsterdam. In the year 2012 only 7 percent were crimes about abuse according to CBS crime statistics. Wherein the actual news reports of the news sources are way more then 7 percent about abuse. The news source that comes the closest to reporting representational crime news is in the year 2012 the ‘Politie’ with a percent of 28. In the following year there was a lower number of registered crime on abuse. The same shift also occurred in the news sources. Still the percentage remained not close to equal. In the year 2013 the news source ‘Het Parool’ came the nearest one to report representational news on abuse and theft in comparison to CBS crime statistics.

In general you could conclude that not a single regional news source (Het Parool, Politie, AT5) reported in the same ratio as the actual crime events of crime and abuse (2012,2013). The news reports on abuse are way much more highlighted then the amount of actual theft crimes. In the same breath, you could also state that the theft crimes were not important as much as abuse crimes to report.

Results