Portfolio

Research question

1 *Which news source is the most accurate in representing crime levels compared to 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) news sources for comparison are the Parool newspaper, information from the Amsterdam police and AT5.

1.1 *To what extent does the most accurate news source align with the reported feelings of insecurity of Amsterdam citizens in the years 2012 and 2013?*

The most accurate news source will be selected using the results of question 1, and its results 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.

The figures on registered crime in Amsterdam will by obtained through the Dutch statistical institute CBS. The website of the institute supports selection and download of specific data. Available extensions are .csv or .xlsx. So far the ordinary excel extension proved complicated for visualization, so we have relied on csv. Data has been selected for headings of theft and abuse in the years 2012 and 2013.

Media news reports on crime used are from the Parool, police and AT5. These sources have been selected for user-friendly interface including Amsterdam-specific search options, enabling exclusion of news from other regions clouding the material.

None of the 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 format that could be read through the Terminal. This has proved unsuccessful.

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

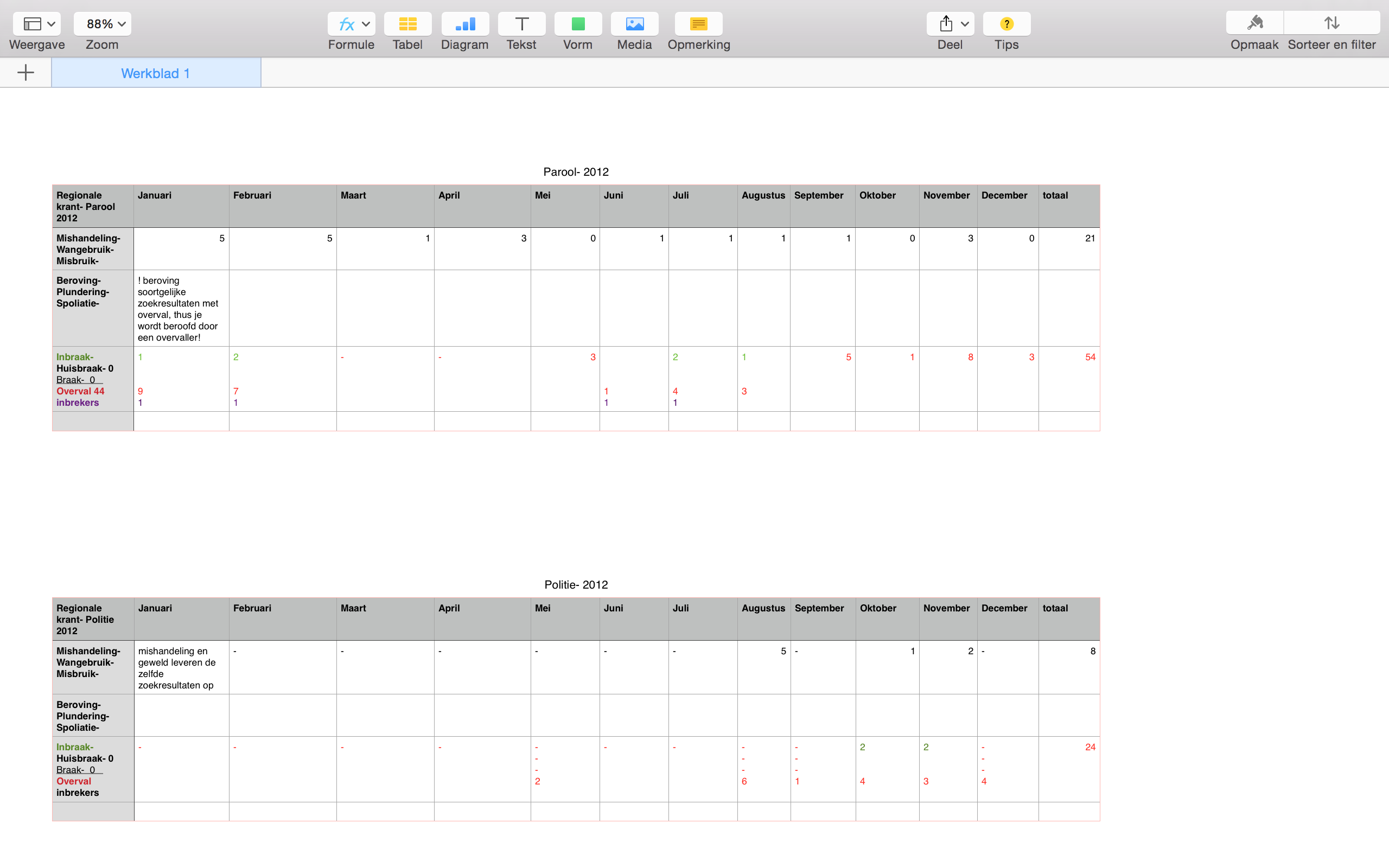
Parool: I have used Google searches for the following keywords: “overval, inbraak, inbrekers, mishandeling” (robbery, burglary, burglars, abuse) within the consignment of Amsterdam crime. The keyword search for abuse, for instance, is as follows: site:parool.nl/parool/nl/7/MISDAAD/ “Amsterdam” “mishandeling”. It became apparent that many 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 of the portfolio.

Police news source: It is uncomplicated to specify the searches on the official website. A number of results proved irrelevant on this site as well.

AT5: Distilling relevant research results from the searches on this site proved time-consuming as well.

The relevant results on crime and theft for 2012 have been counted and manually entered into a document. Results for 2013 are yet to be obtained, but considering the time-restraints I began visualizing this data in R first. Once the visualization and method of accomplishing this visualization is completed will we continue obtaining data for 2013 and the feelings of insecurity noted in the secondary research question.

Retrieving Data



Visualization

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 in R failed as well, prompting me to enter the data in R manually. This has resulted in the script as uploaded in Github: script.txt. The result I wish to create in R is as follows:

I have not been able to recreate this result using the R ‘hist’ function. There is a recurring error message with every attempt. I will experiment with amongst others the following code:

'write.table(x,file='',row.names=TRUE,col.names=TRUE, sep=' ')' - '

Once I have found a way to add the ‘x’ value to the file I will visualize the data on feelings of insecurity and the representations of theft and abuse in the media. The results will likely resemble the following. We will look into alternatives if necessary for the powerpoint presentation in the class of week 6.

|  |  |  |  |
| --- | --- | --- | --- |
| 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. |
|  |  |  |  |

Steps