

assignment-02

September 23, 2020

1 EPA-1316 Introduction to *Urban* Data Science

1.1 Assignment 2: Geographic Visualisation

TU Delft Q1 2020 Instructor: Trivik Verma **TAs:** Aarthi Meenakshi Sundaram, Jelle Egbers, Tess Kim, Lotte Lourens, Amir Ebrahimi Fard, Giulia Reggiani, Bramka Jafino, Talia Kaufmann
Computational Urban Science & Policy Lab

1.2 1. Introduction

Note: If you have not gone through **labs and homeworks 04-05**, kindly do so before starting this assignment, as those will help you with all the necessary knowledge for this assignment. This assignment will be useful for you when exploring and visualising your data, and discussing observations, outliers or relationships situated geographically, as part of the final project for this course or in the future.

1.1 Submission Please submit the results by Brightspace under **Assignment 02**, using a single file as example,

firstname_secondname_thirdname_lastname_02.html

If your file is not named in lowercase letters as mentioned above, your assignment will not be read by the script that works to compile 200 assignments and you will miss out on the grades. I don't want that, so be exceptionally careful that you name it properly. Don't worry if you spelled your name incorrectly. I want to avoid a situation where I have 200 assignments all called assignment_02.html

Please **do not** submit any data or files other than the `html` file.

1.2 How do you convert to HTML? There are 2 ways,

1. from a running notebook, you can convert it into html by clicking on the file tab on the main menu of Jupyter Lab
 - File → Export Notebooks as... → Export Notebook to HTML
2. go to terminal or command line and type
 - `jupyter nbconvert --to html <notebook_name>.ipynb`

1.3 Learning Objectives This assignment is designed to support three different learning objectives. After completing this laboratory you will be able to:

- Combine different datasets
- Work with Geographic data
- Plot important information about the data (graphs, scatter plots, choropleth, etc..) using the principles of graphical excellence and guidelines of exploratory data analysis.

1.4 Tasks Problem Structure: First formulate a hypothesis/RQ, identify two/three key variables, search for appropriate datasets (csv, excel, etc.), merge the csv and the given shape file, do a bit of EDA/spatial analysis and then report your result in a list of **5 neighbourhoods** that discuss the RQ.

What we expect: Provide a list of **five neighbourhoods** in The Hague characterised as “safe” or ”” (insert your measure based on what data you collected). Give at least two measurements that may be linked to each other (for example: your hypothesis is that neighbourhoods with no green space are more prone to populations with mental health issues). Be explicit about what definition you use for each measurement using markdown cells. I am not looking for mathematical equations as justification, but you are welcome to also form simple relations and show them in markdown. I shouldn’t have to call you to find out what you mean by measurements of “Accessibility” or “Safety” or “Popularity” etc. Do not assume that there is one normative definition and skip your reasoning. * Note that the measurement can also just be number of people living in a neighbourhood, their voting choice or availability of jobs. But it will not hurt to plot more than just a column from a csv.

This assignment requires you to go through the following tasks in eda and visualisation, and networks and spatial weights.

1. Use at least two datasets: merge at least 1 shapefile (already provided) and a csv file (you find and obtain).
2. Justify your data collection process using markdown cells as you go through the notebook
3. Clean your data and make it tabular for your own good! (think about weeks 1-3 and assignment 1)
4. Carry out an exploratory data analysis
5. Justify your choice of the list on the basis of your analysis.
 - Use at least **3 figures** to support your analysis. Think about exploratory data analysis (build data, clean data, explore global, explore group properties).
 - These figures should have followed the principles of graphical excellence. Using markdown, write explicitly under each figure at least **3 principles of excellence** you have used to make it.
 - Create **choropleths** to display region-specific information (ex. population, voting choice or jobs availability)
 - Be careful with the use of color (and try to be inclusive of color blind people)
 - Use **one method** from the lectures to discuss what you observe for your variable(s). Examples below, * local or global spatial autocorrelation * network measures * spatial weights / spatial lag * binning * feature scaling, normalisation or standardisation
6. **[Optional]** Plot and Overlay Shapefiles to show Den Haag with some other elements like the sea, canals, centroids, amenities (try Open Street Maps data - using `osmnx`?), climate or

land-use patterns, etc.

Remember to always document your code!

1.3 2. Download the Data

For this assignment I am providing you with the shapefiles of The Hague. Mikhail Sirenko has prepared these files with love and care so that you can connect it with either [Den Haag Cijfers](#) or [CBS](#) datasets without having to clean badly collected data.

Note: For data from CBS, data is only complete upto 2017. You will have to subset the data on municipality using the variable name `gm_naam = 's-Gravenhage` and then subset on neighbourhood resolution using variable name `recs = Buurt` to get the data that can match the shapefiles we have provided.

So after you unzip, we'll work with the file `neighborhoods.xxx`, which is in one of many geographic formats. Put the data in a convenient location on your computer or laptop, ideally in a folder called **data** which is next to this **jupyter notebook**. I recommend taking a look at the file with format `.json` in a text editor like *atom* for any system or notepad++ for windows. These will also make your life easy for everything else on your computer. Make sure you've set your working directory in the correct manner – okay?

It's a big file and it may take a while to load onto your laptop and into Python (running on the jupyter labs environment).

So, to summarise, you will use at least two datasets.

1. Download Shapefiles provided with the assignment
2. Get a second dataset of your choice from The Hague city region using the links above (curate them as you like)

More Data Sources You can find more data sources on Cities and Population, Climate indicators and Land-use in the following links in case you are attempting the **[Optional]** exercise.

- <http://citypopulation.de/>
- <https://www.census.gov/programs-surveys/geography.html>
- <https://www.eea.europa.eu/data-and-maps>
- <http://download.geofabrik.de/>

In case you get more data as shapefiles, and want to play with projections, a nice guide for it is [here](#)

1.4 3. Start your analysis

```
[1]: # your code here
     # use many cells if you like to structure your code well
```