Introduction to Data Visualization

Winter Term 2021/2022 – Assignment Sheet # 1

Hasso-Plattner-Institut
Fachgebiet Computergrafische Systeme
Prof. Dr. Jürgen Döllner

Objectives

With these assignments you will learn how to

- learn theoretical aspects about the assessment of visualizations;
- apply data filtering and preprocessing to prepare data for further analysis and visualization;
- apply D3.js and matplotlib for visualization using basic charts.

For your programming tasks, you will use both Python and JavaScript for preprocessing and visualization, respectively. You are expected to comment and document your code in a focused and clear manner.

Your solutions must be uploaded via Moodle by **November 25, 2021, 9am (UTC+1)** as one ZIP file that contains all answers and source files. The naming convention for this ZIP file is **sheet1_<group_name>.zip**.

Instructions

Implementation As visualization framework we use *D3.js* and *matplotlib* and their respective languages JavaScript and Python; for details we refer to their documentation[1].

The implementations with D3.js should be implemented as a web page based using the framework. Make sure that relative paths are used. Supplementary exercise-specific data should be organised in a sub-folder called "data". Colour schemes, in particular, can be explored and loaded via colorbrewer2.org[2]. Local provisioning for debugging and testing can be done using a local web server such as the http-server by node[3]. The server can be installed with the following command:

- npm install http-server -g

In order to load e.g. CSV files you need to start a local server by

- http-server -a 127.0.0.1 -o

in your working directory containing both the sources and the data.

The implementations with Python and matplotlib should be implemented as a ready-to-use python script. The result of a script execution is expected to be a single image file. For details refer to the documentation. If you use any external libraries additionally, hand in the requirements file as exported via the *pip freeze* command[4]. The requirements about the data folder and relative paths from above also apply here.

Pair Programming On these assignments, you are encouraged (not required) to work with a partner provided you practice pair programming. Pair programming "is a practice in which two programmers work side-by-side at one computer, continuously collaborating on the same design, algorithm, code, or test." One partner is driving (designing and typing the code) while the other is navigating (reviewing the work, identifying bugs, and asking questions). The two partners switch roles every 30–40 minutes, and on demand, brainstorm.

Violation of Rules A violation of rules results in grading the affected assignments with 0 points.

- Writing code with a partner without following the pair programming instructions listed above (e.g., if one partner does not participate in the process) is a serious violation of the course collaboration policy.
- Plagiarism represents a serious violation of the course policy.

[1] D3 documentation can be found at `https://github.com/d3/d3/blob/master/API.md` and matplotlib documentation can be found at `https://matplotlib.org/`

[2] `https://colorbrewer2.org`

[3] `https://nodejs.org/en`

[4] `https://pip.pypa.io/en/stable/cli/pip_freeze/`

Exercise 1.1: Reading "Views on Visualization" (5 Points)

The paper "Views on Visualization"[5] by van Wijk introduces views on visualization as art, design, and a scientific field. He introduces models that support the formal understanding of visualization and to assess the value of visualizations. Read the paper and answer the following questions:

- The author argues that visualization is not good by definition. Why? (1 Point)
- Describe in a few sentences what positive and negative *Knowledge* mean in the model presented by the author. (1 Point)
- Based on the economic model introduced, when is the visualization considered to be successful? (1 Point)
- Why, according to the author's model, is *insight* not considered to be a feasible factor of the assessment of visualizations? (1 Point)
- What contributes to making visualization a stronger scientific discipline? (1 Point)

The answers should be brief and handed in as a PDF file.

Exercise 1.2: Preprocessing, Filtering, and Plotting Data (6 Points)

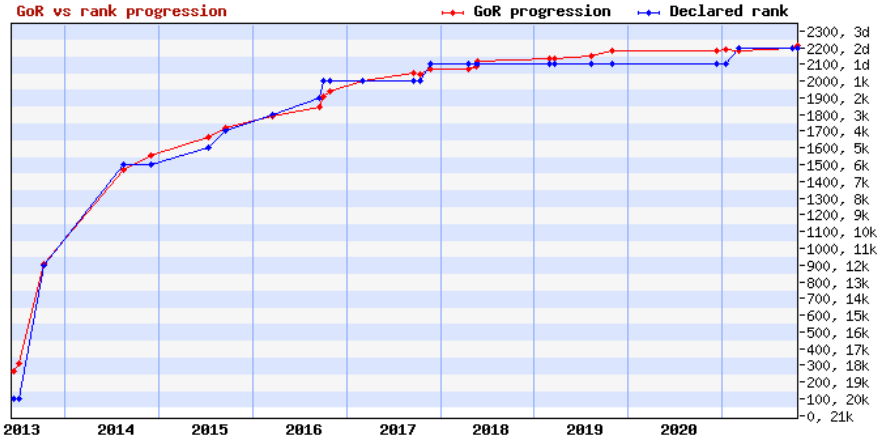

Figure 1: An example line plot for the GoR data.

Go is an ancient board game originated in China. The goal of this exercise is to visualize the playing strength of a selected Go player over time, called *Go Rating* (GoR). An example visualization is shown in Fig. 1 in which the red line shows the GoR progression (ignore the blue line). The *all.hst* file contains data for the GoR evaluation from all European Go players over time. Use Python and matplotlib for the implementation.

a) Parse the given *all.hst* file in a structured format such as csv. (Additionally hand in the result of this step; 1 Point)

b) Filter the data for data points of the player *"Tim Cech"*. (1 point)

c) Parse and extract the dates of the tournaments played from the tournament column (`yymmdd`-format). (1 Point)

d) Extract the appropriate GoR values. The IGor column contains the GoR value before the tournament, the FGor column contains the GoR value after the tournament. (1 Point)

e) Generate a plot layout with time as x axis and GoR values as y axis including proper labels. Plot a simple line diagram with the data from part c) and d). (1.5 Points)

f) Briefly justify your choice of an adequate and correct interpolation technique. (0.5 Points)

[5] Jarke J. Jarke J. van Wijk. "Views on Visualization." IEEE Transactions on Visualization and Computer Graphics (2006): 20–28. doi: 10.1109/TVCG.2006.80

Exercise 1.3: Scatter Plot for Displaying Data Correlation (6 Points)

Given the data set *UnRegionsGdp.csv*, which contains in each row name, the Human Development Index (HDI), and the Gross Domestic Product (GDP) among other data for the year 2017. Visualize the correlation of HDI and GDP by a scatter plot as shown in the examples in figure 2. Use D3.js for visualization.

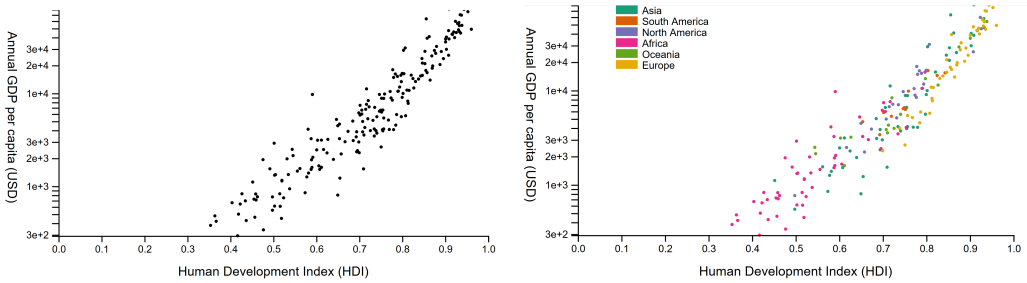

Figure 2: Two example scatter plots, uniform (left) and categorized (right).

a) Each country should be represented by a point in the two-dimensional Cartesian coordinate system. HDI is mapped onto the x axis, GDP on the y axis. Read the data using *D3.js* functionality and select only valid data points (HDI > 0 and GDP > 0). (2 Points)

b) Expand the implementation by coordinate axes including labels for both x and y. For equalization of the y axis a logarithmic scale should be used. (2 Points)

c) Use a qualitative color scheme to map the continent for each data point. The legend of this representation should be added to the visualization. (2 Points)

Exercise 1.4: Return Triangle Plot for Share Price (6 Points)

A return triangle plot visualizes the stock performance over the years. Performance numbers, derived from the price development, are arranged in the form of a triangle, with the horizontal direction representing the year of purchase and the vertical direction representing the year of sale. Figure 3 shows an example for a return triangle for the German financial service company *Wirecard AG* from 2010 to 2019[6].

Wirecard Aktie Renditedreieck

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
2011	22,6									
2012	34,9	48,4								
2013	41,2	51,6	54,9							
2014	37,6	43,0	40,3	27,2						
2015	35,4	38,8	35,7	27,0	26,8					
2016	26,1	26,8	21,9	12,5	5,8	-11,7				
2017	37,2	39,7	38,1	34,2	36,6	41,7	127,6			
2018	37,8	40,2	38,8	35,8	38,1	42,1	80,3	42,8		
2019	30,0	30,9	28,6	24,7	24,2	23,5	38,1	7,6	-18,9	
2020	-29,4	-33,6	-40,0	-47,6	-54,8	-63,2	-70,4	-85,0	-95,2	-99,7
Ø	27,3	31,7	27,3	16,3	12,8	6,5	43,9	-11,6	-57,0	-99,7

Figure 3: Return triangle for the Wirecard AG

Create a return triangle for the American credit card company *American Express Company* from January 1, 2001 to December 31, 2021. Use *D3.js* for the implementation. Compute the annual returns from the dataset *AXP.csv*[7] and display them by a return triangle. Add a straight-forward color schema.

[6] `https://www.boerse.de/renditedreieck/Wirecard-Aktie/DE0007472060`
[7] `https://finance.yahoo.com/quote/AXP/history?p=AXP`