CS-E4840 Information Visualization

Assignment 1

Specific instructions for Assignment 1:

- Deadline is on 15 March 2020 at 23:55, local time.
- Maximum number of points from this assignment is 15.
- Assignment 1 has four exercises that must all be completed to obtain full points.

General instructions:

- The assignment should be completed by one person and discussions with others are encouraged. However, your final solution must be your own. Please read the Aalto University Code of Academic Integrity and Handling Violations Thereof for further details.
- The language of the assignments is English.
- The deadline has a late submission policy: each day being late automatically reduces 3 points of the assignment. However, you cannot get negative points for each assignment.
- If you have a pressing and verifiable (e.g., by a doctor's certificate) reason that causes you to miss the deadline, you can send an email to the lecturers to request an extension (without penalty). The extension must be requested before the deadline. Otherwise, the extension will be refused.
- The submitted report should be in Portable Document Format (pdf). If you are using software such as Word, then export the final document as pdf.
- Do not attach any source code.
- State clearly your name and your student id in the report.
- Number your answers to correspond the questions in each assignment, and do it in order corresponding to the questions.

The data sets

The datasets given for Exercises 3 and 4 are in the CSV format, where the lines present rows of a table and the numbers for each column in a row are separated by commas.

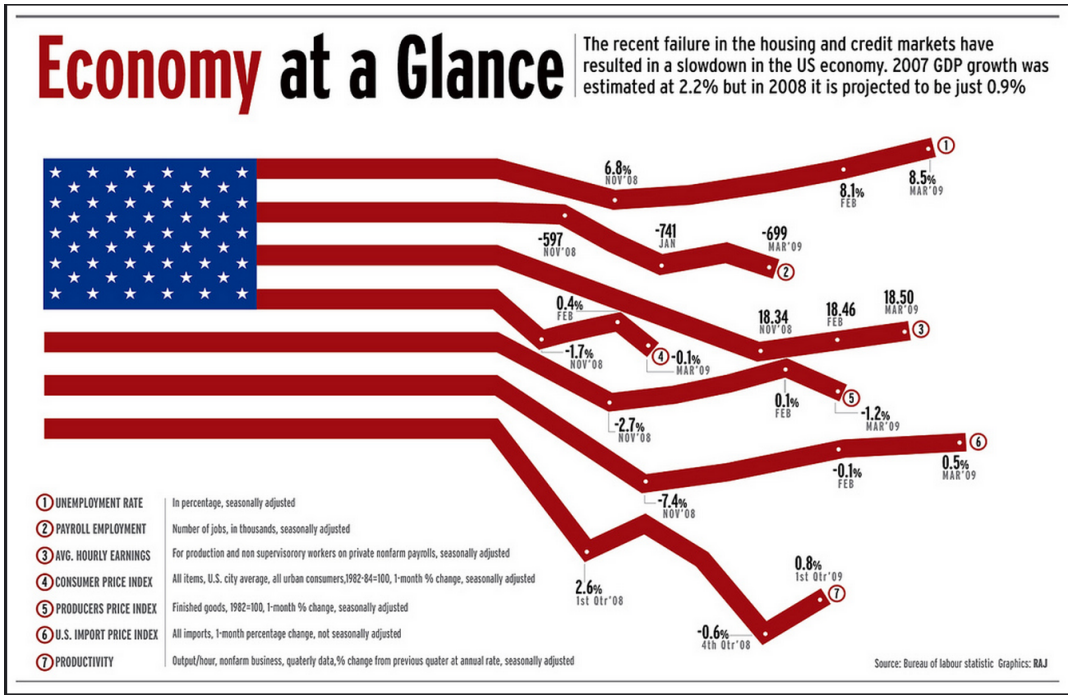

Figure 1: Some indicators of U.S. economy during the crisis in 2008.

Exercise 1 (3 points)

Figure 1 shows a published visualisation (origin unknown) of various measures indicating economic development in the USA at the beginning of the financial crisis a decade ago.

Your task is to:

(a) Analyse the visualisation in Figure 1, starting from Tufte's principles. List at least 4 items that contradict these good design principles.

(b) Give a suggestion for a better visualisation for the same task, using the data shown in Figure 1, and explain your design choices. For a full mark you should provide an image (e.g. drawing, even by hand) and an argumentation, why the proposal is better than the original.

Exercise 2 (2 points)

Look for an example of visualisation that you find particularly beautiful or disturbingly bad in a recent issue (recent = published on or after June 2019) of a high profile journal (Nature, Science, etc.) or mainstream media (CNN / Helsingin Sanomat / Tilastokeskus.fi). Try to explain what makes it appealing or useful or horrible. The journals are accessible from within Aalto.

Specify exactly the visualization you have selected. If a link is provided, it must be functional and unambiguous (otherwise you get zero points). It is safer to inset the picture in your report.

Exercise 3 (7 points)

(a) Donald is the president of a large country whose economy is highly dependent on fossil energy. While there are concerned voices about global warming, his conservative stand is to keep the economy

running. Your goal is to help Donald, based on the data set (climate.data) [1], to convince the public that the news about warming are fake and there is no crisis ahead. You can use every trick given in your book and lectures: chartjunk, optical illusions, "creative" layout, use only part of the data. You can use any plotting software available (R, Matlab, Python, Excel, OpenOffice, gnuplot etc.).

(b) Bernie is a senator of the major opposition party. Naturally, he disagrees with Donald, and wants to change the environmental politics. Use the same data to make the opposite case. Again you can use every creative trick you can think of.

(c) Use the notion of Lie factor (see slides of Lecture 2 or Tufte's book, page 57–58) to measure whether the above plots are underestimating or overestimating the change. Provide a clear explanation of how you measure the lie factor in this particular case.

(d) Maija is a student at Aalto university, who is tired of manipulative plots. She decides to start a blog of graphical designs of important topical datasets. Help Maija and create an objective plot for the case of global warming, following the principles of Tufte as closely as possible. Justify your choices, and describe how/whether you can improve your plot even more.

Exercise 4 (3 points)

Visualize the Wine dataset (wine.data). This dataset contains 13 measurements of 178 different wines, fermented from 3 different grape varieties. The first variable indicates the grape variety. The remaining 13 variables (table columns) are as follows:

1. Alcohol
2. Malic acid
3. Ash
4. Alcalinity of ash
5. Magnesium
6. Total phenols
7. Flavanoids
8. Nonflavanoid phenols
9. Proanthocyanins
10. Color intensity
11. Hue
12. OD280/OD315 of diluted wines
13. Proline

Select at least 4 of these features, and create small multiples (=trellis), a visualization with scatterplots of each pair of features, arranged as a matrix; see an example of such arrangement for an Iris dataset at `https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Iris_dataset_scatterplot.svg`. Indicate with different colors the three grape varieties. Try to maximize the data-ink ratio, within reason.

[1] from `http://www.drroyspencer.com/latest-global-temperatures/`