UCLouvain - EPL - LDATA2010 - Exercise session

Graphic critique and design

Exercises credit: CS-E4840 Aalto University, CPSC 547 UBC.

Exercise 1. List 4 graphical mistakes in the plot in Fig. 1.

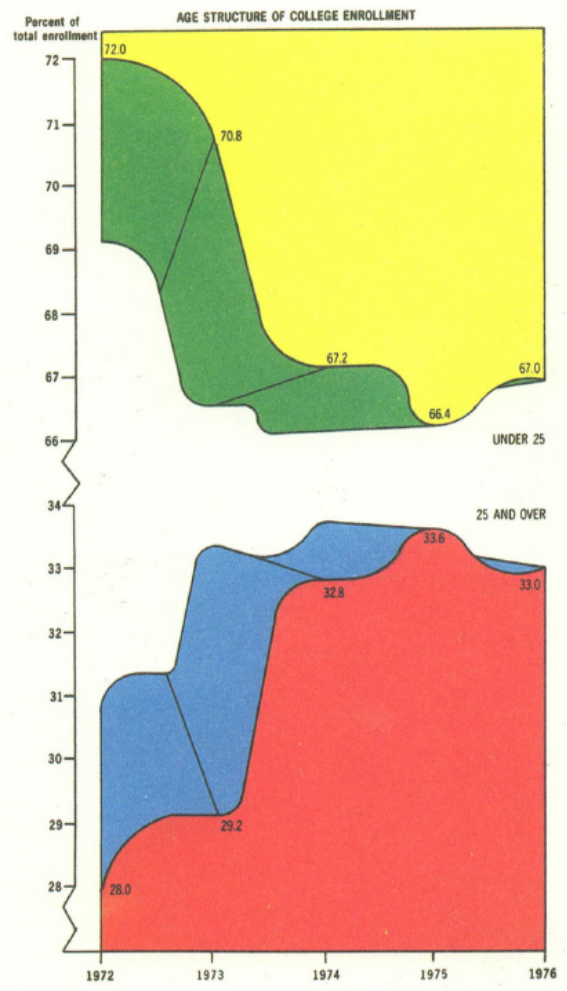

Figure 1: American Education, 1970s

Exercise 2. Consider the suitability of the chart in Fig. 2 for each of these five tasks by ordering them from best supported to least supported:

- Comparisons;
- Organizing/Ordering;
- Presenting information;
- Relationships/Correlations/Trends;
- Showing exceptions.

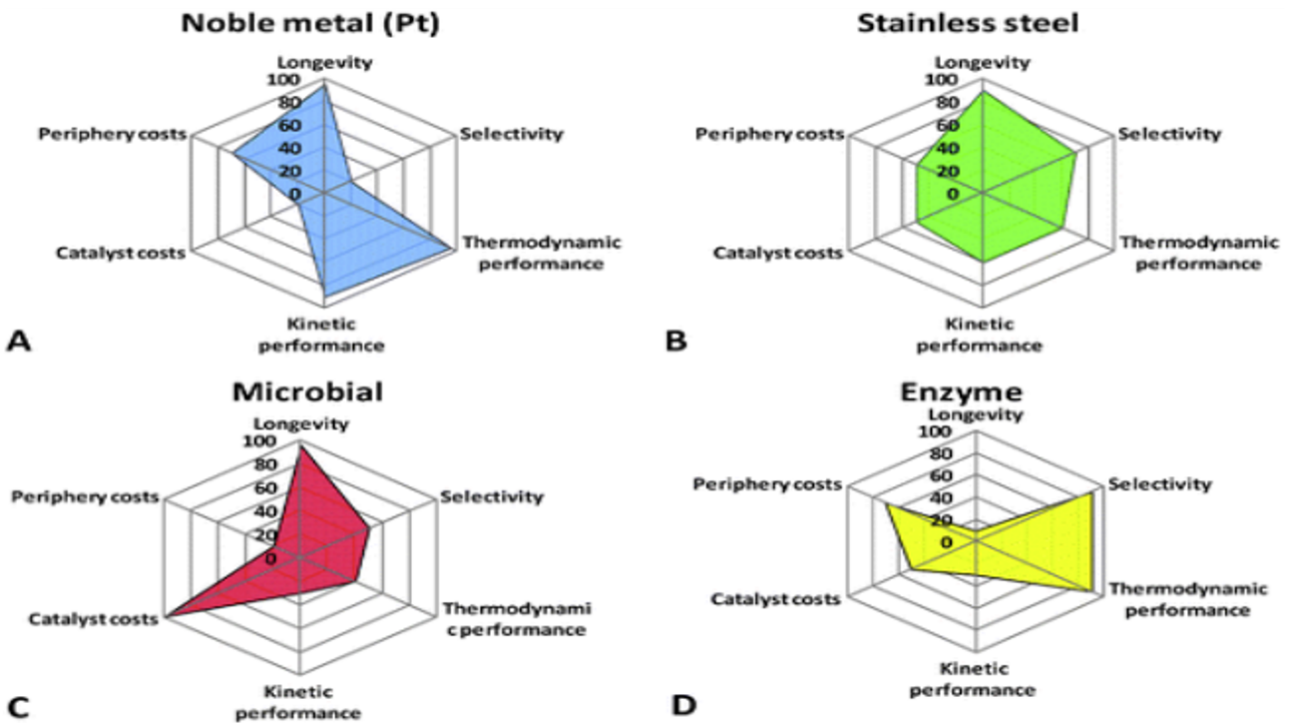

Figure 2

Exercise 3. For each task below, which chart between 3a and 3b is more suitable?

- Comparisons;
- Organizing/Ordering;
- Presenting information;
- Relationships/Correlations/Trends;
- Showing exceptions.

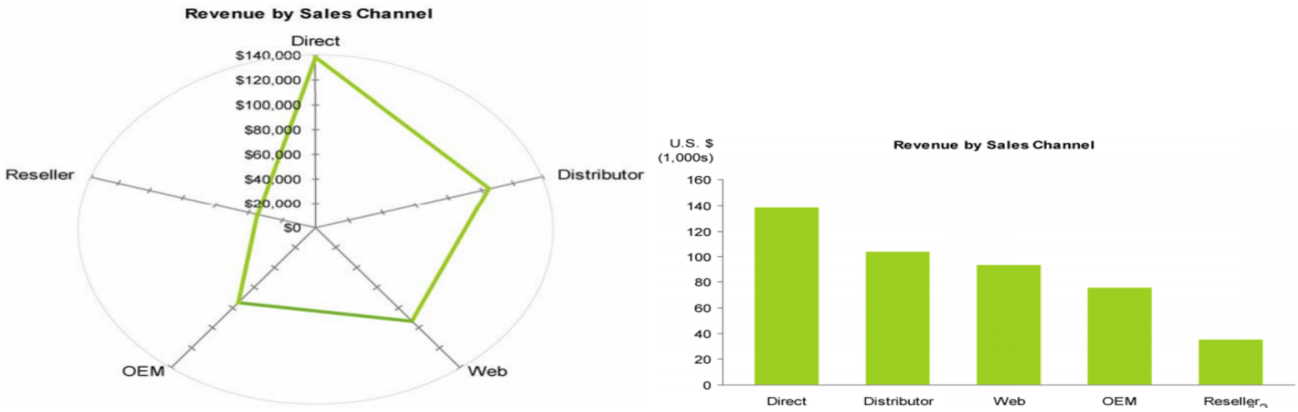

Figure 3

Exercise 4. Juha is the Prime Minister of a small, sparsely populated Nordic country whose economy is highly dependent on exports. Due to several years of world-wide economic downturn as well as plunging oil prices, the debt of your country is on the rise.

1. Help Juha to convince the public that the country is headed for an unprecedented crisis. Use the provided data (debt.txt) to make your case. You can use every trick of your choice: chartjunk, optical illusions, creative layout, use only part of the data. You can employ the plotting software of your preference (Matlab, Python, R, etc.).

2. Antti is the speaker of the major opposition party. Naturally, he disagrees with everything that Juha

proposes. Use the same data to make the opposite case. Again you can you use every creative trick you can think of.

3. Use the notion of Lie factor to measure whether the above plots are underestimating or overestimating the change. Since Lie factor is defined between two values, use the ones for the years 1940 and 2015.

4. Robert is a student at Université catholique de Louvain, and he is tired of manipulative plots. He decides to start a blog of graphical designs of important topical datasets. Help Robert and create a plot for the national debt by following the principles of Tufte as closely as possible. Justify your choices, and describe how/whether you can improve your plot even more.

Exercise 5. Visualize the Wine dataset (wine.txt). This dataset contains 13 measurements of 178 different wines, fermented from 3 different grape varieties. Each row of wine.txt corresponds to a wine sample, with the first variable indicating the grape variety and the remaining 13 ones 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.

The variables as separated by commas.

Select at least 4 features, and create small multiples (= trellis), a visualization with scatterplots of each pair of features, arranged as a matrix[a]. Indicate with different colors the three grape varieties. Try to maximize the data-ink ratio, within reason.

[a] See `https://en.wikipedia.org/wiki/Iris_flower_data_set#/media/File:Iris_dataset_scatterplot.svg` for an example of such arrangement for the Iris data set.