

Assignment 1

Specific instructions for Assignment 1:

- › The deadline is on March 19, 2023 at 16:00, Finnish time.
- › The maximum number of points from this assignment is 15.
- › This assignment has four exercises that must all be completed to obtain full points.
- › 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.

General instructions:

- › The general grading criteria are available on the MyCourses web page.
- › Students should complete the assignment individually, without the aid of AI tools, but discussions with other students 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 reports should be English.
- › Policy for late submissions: The score is reduced by three points per day after the deadline. (However, you cannot get negative points for an assignment.)
- › If you have a pressing reason that causes you to miss the deadline, you can send an email to the lecturers (cs-e4840@aalto.fi) to request an extension without the late submission penalty. The reason must be such that it would entitle you to be absent from work (e.g., illness) and verifiable (e.g., doctor's certificate). The extension must be requested before the deadline.
- › The submitted report should be in a single Portable Document Format (PDF) file. If you are using software such as Word, then export the final document as PDF. If you have several PDFs, then please merge them into one before submitting the assignment.
- › Do not attach any source code.
- › State your name and your student ID clearly in the report.
- › Number your answers by the number of the questions, and keep the order.

Exercise 1 (3 points)

- Analyze the visualization in Figure 1, starting from Tufte's principles. List at least four items that contradict these good-design principles.
- Suggest an improved visualization and explain your design choices. For a full mark, you should provide an image (e.g., drawing, even by hand) and explain why your proposal is better than the original.

Exercise 2 (7 points)

- Satoshi is running a crypto business. It is very turbulent with fake media is spreading rumors of bubbles and pyramid schemes. Your goal is to help Satoshi convince the public that Bitcoin has performed better than the the Nasdaq. Use the provided data (Bitcoin.csv and Nasdaq.csv), which contains the daily closing prices in US dollars for the Bitcoin and the Nasdaq-100 index, respectively, to make your case. You can every trick in your book: 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.).
- Warren is a passive investor irritated by the whole Bitcoin fuzz. Use the same data to make the opposite case. Again, you can use every creative trick imaginable.
- 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 relative performance of the two financial instruments.
- Jorma is a student at Aalto University. He is impartial because he has no cash, Bitcoins, or Nasdaq ETFs. He decides to start a blog about graphic design and data visualization. Help Jorma and follow the principles of Tufte as closely as possible, and create a plot for the relative performances of Bitcoin and the Nasdaq-100 index. Justify your choices, and describe how/whether you can improve your visualization further.

Exercise 3 (2 points)

Look for an example of a visualization that you find particularly beautiful or disturbingly bad in a recent issue (published on or after June 2021) of a high-profile scientific journal (*Nature*, *Science*, etc.) or mainstream media (CNN / Helsingin Sanomat / Tilastokeskus.). Try to explain what makes it appealing, purposeful, horrible, etc. The journals are accessible from within Aalto. Insert the picture (a screenshot or photo) in the report.

Exercise 4 (3 points)

Visualize the penguin dataset (Penguins.csv), available at the MyCourses page. This dataset contains beak measurements of 344 penguins sampled in Antarctica. Create a small-multiples (trellis) visualization with scatterplots of each pair of features, arranged as a matrix; see an example of such arrangement for the Iris dataset (see Wikipedia's "Iris Flower data set"). Indicate with different colors the three species (Chinstrap, Gentoo, Adelie). Try to show the difference between the regions, and maximize the data-ink ratio, within reason. More info about this data set (like what those measurements actually refer to) can be found here: <https://allisonhorst.github.io/palmerpenguins/>

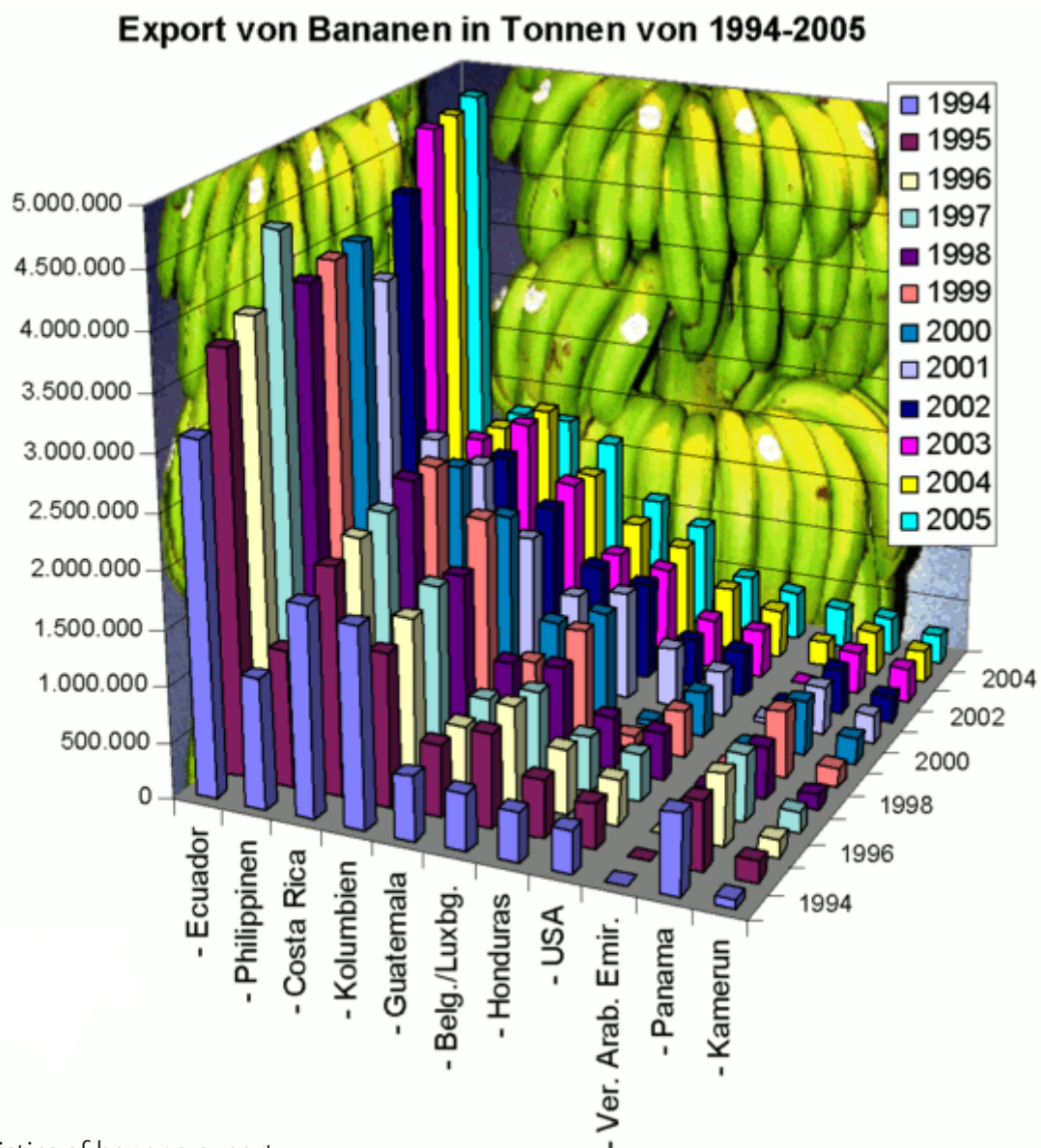


Figure 1. Statistics of banana export.