CS-E4840 Information Visualization

Assignment 1

Specific instructions for Assignment 1:

- Deadline is on 11 March 2018 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.

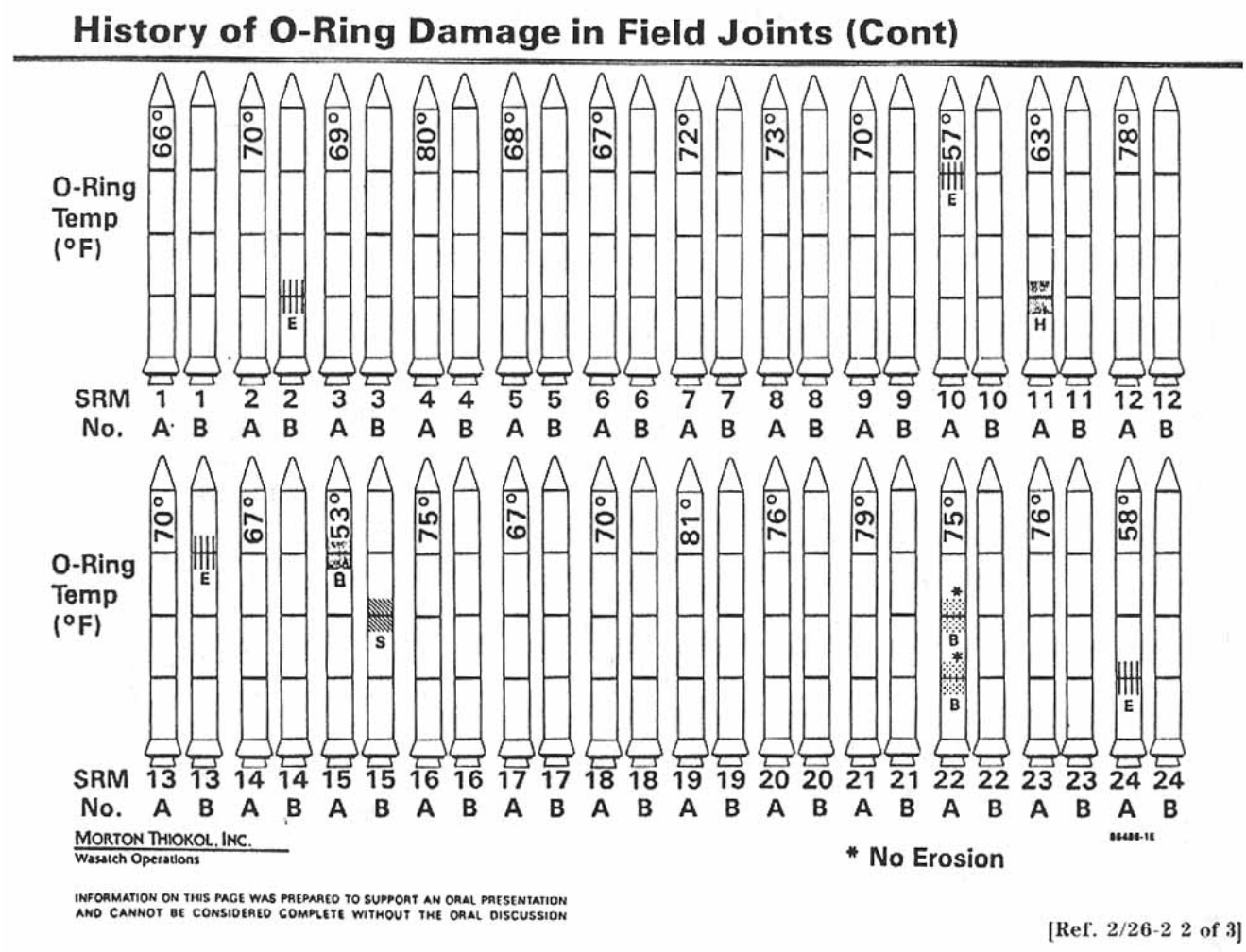

Figure 1: History of o-ring damage.

Exercise 1 (3 points)

The space shuttle Challenger was destroyed in an accident shortly after launch on 28 January 1986. The disaster was caused by hot gases leaking from one of the two solid rocket boosters, which resulted to catastrophic damage to the vehicle. The cold weather—29 degrees Fahrenheit—was a contributing factor: the hot gases in the solid rocket boosters were sealed in the booster by a structure that was made tight by rubber o-rings. In the cold weather the rubber hardened and lost some of its elastic properties, making it easier for the hot gases to escape from the boosters with disastrous consequences. The Challenger disaster is commonly used example of the role of visualisations in decision-making.[1]

Figure 1 shows one of the visualisations that were used when it was being decided whether Challenger should be launched or not. In Figure 1, you can see the images of the solid rocket boosters with launch number (SRM) and the two boosters (A and B) for each of the historic launches. Figure also shows the o-ring temperature in degrees Fahrenheit and damage to the o-rings. Each of the boosters have 3 pairs of o-rings, totalling 6 pairs of o-rings for a launch. The o-rings come in pairs for redundancy: if there is blowby of hot gases past the primary o-ring the secondary o-ring becomes primary. If the secondary o-ring

[1] Although some of the criticism may have been unfair for the engineers, see `https://doi.org/10.1007/978-3-319-45193-0_12` and `https://people.rit.edu/wlrgsh/FINRobison.pdf`

fails as well then the hot gases can leak out, which can be catastrophic. The letter codes are as follows: "S" is "heating of secondary o-ring" (because of the blowby past the primary o-ring), "B" is "primary o-ring blowby", "E" is "primary o-ring erosion", and "H" "means heating of the primary o-ring". No letter means no damage to the o-ring. The darker the color, the greater the amount of blowby. The "badness scale" is therefore, from the best to the worst: no damage, H, E, B (primary o-ring fails here), and S (secondary o-ring takes the punishment, this should not happen), with greater amount of blowby being worse. For example, in launch 1 at 66 degrees Fahrenheit there was no o-ring damage. In launch 2 at 70 degrees Fahrenheit there was erosion to the lowest primary o-ring of the B booster etc. On mission 15 one of the the secondary o-rings heated when the hot gases blew by the primary o-ring which did not seal quickly enough. Mission 22 experienced a small amount of primary o-ring blowby.

In other words, the objective here is to study the relation of o-ring temperature to the probability and magnitude of blowby, in order to decide whether it is safe to launch the space shuttle at 29 degrees Fahrenheit.

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.

What would your recommendation be, to launch or not to launch?

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 2017) 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.

Exercise 3 (7 points)

(a) Satoshi is running a dot-com business of buying and selling bitcoins. It is a very turbulent business however and fake media is spreading rumours of bubbles and pyramid schemes. You goal is to help Satoshi to convince the public that bitcoin has performed better than the S&P 500. Use the provided data (BTCvsSP500.csv), which contains the daily closing prices in US dollars for the bitcoin and S&P 500 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.).

(b) Warren is a passive investor who is irritated by the whole bitcoin fuzz. Use the same data to make the opposite case. Again, you can you 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 relative performance of the two financial instruments.

(d) Jorma is a student at Aalto University. He is impartial, because he has no money, bitcoins, or S&P 500 ETFs. He decides to start a blog of graphical designs of important topical datasets. Help Jorma and follow the principles of Tufte as closely as possible, and create a plot for the relative

performances of the bitcoin and S&P 500. Justify your choices, and describe how/whether you can improve your plot even more.

Exercise 4 (3 points)

Visualize the *Olive* dataset, available at `https://raw.githubusercontent.com/ggobi/ggobi/master/data/olive.csv`

This dataset contains 572 olive oil samples from three different regions of Italy. For each sample, the normalized concentrations of eight fatty acids are given. The first variable indicates the row name, the second region, the third the area, and the remaining columns give the fatty acid concentrations. All numbers are separated by a comma and the first row gives the column labels.

Select at least 4 features, and create small multiples (=trellis), a visualisation with scatterplots of each pair of features, arranged as a matrix; see an example of such arrangement for an Iris dataset. Indicate with different colors the three regions. Try to show the difference between the regions, and maximize the data-ink ratio, within reason.