T-61.5010 INFORMATION VISUALIZATION

April 5, 2016.

To pass the course you must pass this written examination as well as the course assignment. Results of this examination will be valid for one year after the examination date.

This examination has five questions, 6 points each. The exam consists of two pages of questions, and two pages of figures, two figures in total. Please write clearly and leave a wide left or right margin. You can have a calculator, with memory erased. No other extra material is allowed.

The results will be announced in mycourses. The questions must be returned.

Each question should be answered on a fresh new page. Questions 1, 4, 5 should be answered on one booklet (sheet of paper), and questions 2 and 3 on another, as different lecturers will mark these questions.

Question 1:

(a) Briefly explain data-ink maximization. Give an example (describe briefly in words and/or draw), where redundant data-ink is useful. [1.5 points]

(b) Briefly explain chartjunk. [1 point]

(c) Consider a visualization given in Figure 1.

(i) Compute the Lie factor between the largest and the smallest quantity. Use the *square area* as a measure of quantity, ignore the rounded corners. What does this Lie factor indicate in this particular figure? The widths of the squares are as follows: 14 12.5 5 2 units. [0.75 points]

(ii) Compute the Lie factor between the largest and the smallest quantity. Use the *square width* as a measure of quantity. What does this Lie factor indicate in this particular figure? [0.75 points]

(iii) Explain in words and/or redraw a better visualization, and provide brief justification why your visualization is better. [2 points]

Question 2:

(a) What are visual acuities? Briefly explain two acuities. [1 point]

(b) Compare and contrast the two receptors cells in the retina. What is the fovea and why is it important? [2 points]

(c) Give one preattentive feature and one non-preattentive feature. Briefly explain why these features are preattentive or non-preattentive. [1 point]

(d) Briefly explain the Gestalt laws of (*i*) proximity, (*ii*) connectedness, and (*iii*) similarity, and the design principles each one conveys when visualizing information. Which of these grouping principles is most powerful? [2 points]

Question 3:

(a) Explain the data abstraction level of Munzner's model of visualization design. Mention three major dataset types. [1 point]

(b) Explain the task abstraction level of Munzner's model of visualization design. What does the task {action, target} pair refers to? Choose a classic visualization, such as Nightingale's diagram on the causes/mortality in the army or Snow's map of cholera deaths (given in Figure 2), and briefly explain a task {action, target} pair that this visualization tried to achieve. [2 points]

(c) Which visual variable can accurately communicate quantitative, ordinal and nominal data? Give an example of a visual variable that should be avoided when graphically encoding quantitative or ordinal data. [1 point]

(d) Choose a visual idiom and indicate: (*i*) the marks and visual variables it uses; (*ii*) the data type and tasks it is most appropriate for. [2 points]

Question 4:

(a) List four aesthetic criteria that make a graph layout easier to understand. [1 point]

(b) List two positive aspects and two negative aspects of force-directed layout. [1 point]

(c) Describe briefly the layered layout. [1.5 points]

(d) Describe three scenarios where linear layout, layered layout, or adjacency matrix is more suitable choice than force-directed layout. (one scenario per layout = 3 scenarios in total) [1.5 points]

(e) Describe briefly focus+context and Fisheye distortion in the context of graph navigation. [1 point]

Question 5:

(a) Give the *mathematical definition* of a principal component (PCA). Describe a technique on how to obtain the components (you don't need to prove that it is correct). [3 points]

(b) Describe briefly a scenario where PCA is not a suitable choice. [1 point]

(c) Describe briefly the connection between Torgerson scaling (classic MDS) and PCA. [1 point]

(d) Metric MDS and Sammon mapping produce (typically slightly) different results. How these results are different? Describe briefly the main difference between non-metric MDS and metric MDS. (Stating the exact optimization criteria is not needed but can be used if helpful) [1 point]

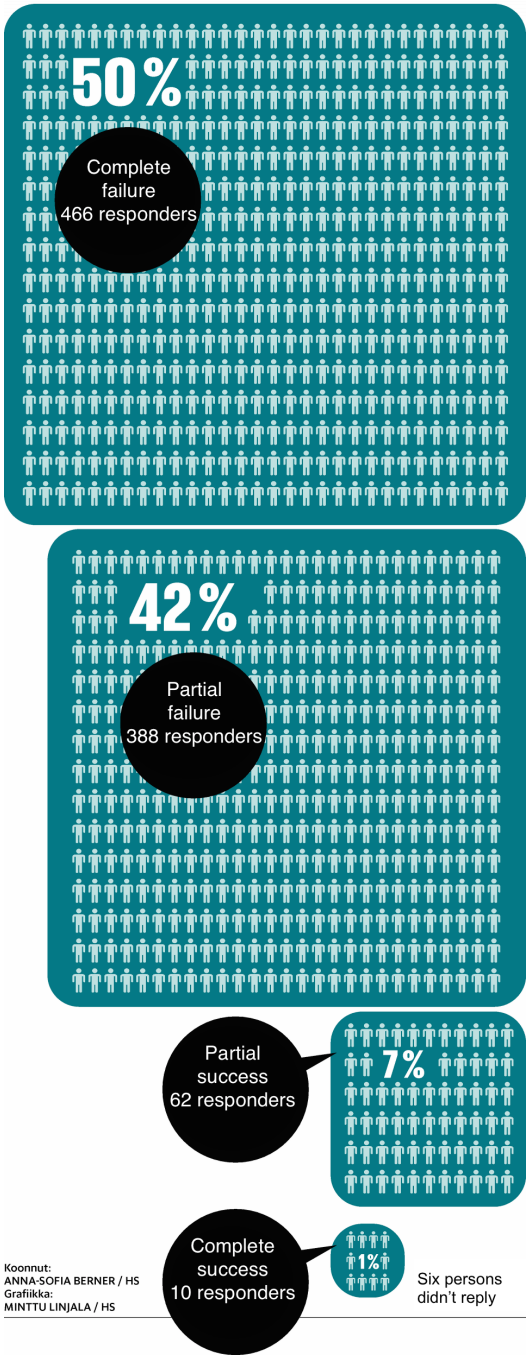

Figure 1: Visualization of a survey among professors. A major newspaper, Helsingin sanomat, surveyed Finnish professors on government's university policy. Professors were asked to rate the policy on scale (1 = complete failure, 2 = partial failure, 3 = partial success, 4 = complete success). Translated and cropped, original in Finnish language was published in Helsingin Sanomat 13.3.2016.

Figure 2: Snow's map on Cholera's deaths