Data Mining Homework 5

Madis Nõmme

March 18th, 2014

Task 1

Listen to the presentation by Tamara Munzner: Keynote on Visualization Principles - http://vizbi.org/Videos/26205288.
Summarize the key take-home messages from her presentation.

Messages I got:

- Visualization is meant for humans to understand the underlying data better. If your graph does not make data clearer, you need better graph.
- Choose visual channels (what/where, grouping, how much) to fit the data. Otherwise the visualization can become mis-communicating.
- Controlling your voice is important to avoid strange sounds on recording (25:31)
- If the goal is to signify one thing, the popout principle can be used - but mostly it works on only one channel at a time (red & blue, squares & circles).
- Eyes beat memory. Carefully pick when to use animation (requires remembering what happened in the past).

1 Task 2

Choose one of them and answer following questions:

Mobile Makeover

- What is the key message for the chart?

Key message was to show that mobile devices and services are on the rise and will continue rising.

- Is it easy to grasp?

Two bottom ones were relatively easy. It would have been easier if they were completely horisontal or vertica. The artistic approach distracted the message. The largest graph comparing number of users of different devices since their launches was least obvious. The scale for quarters after launch could have been years.

- Is it possible to improve it according to Tamara's suggestions? How?

Yes. Using just a bar chart for the users/time from launch graph. It was mixing too many different channels (hue, spacial, vertical-horizontal axis, 3d)

2 Task 3

Read data from the file, data.txt, calculate mean and variance for every feature (column). Compute correlation between pairs of features x1 vs y1, x2 vs y2 etc. Compare results. Plot each of these pairs separately (such that x feature is on the x-axis and y feature on the y-axis). Interpret the results.

	x1	x2	x3	x4	y1	y2	y3	y4
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.1	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.1	5.39	12.5
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89
Mean	9	9	9	9	7.500909091	7.500909091	7.5	7.500909091
Variance	11	11	11	11	4.127269091	4.127629091	4.12262	4.123249091
Correlation	0.8164205163	0.816236506	0.8162867395	0.8165214369				
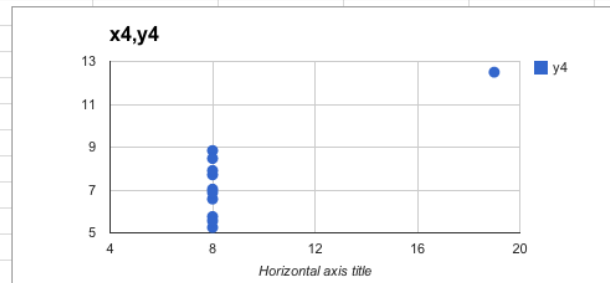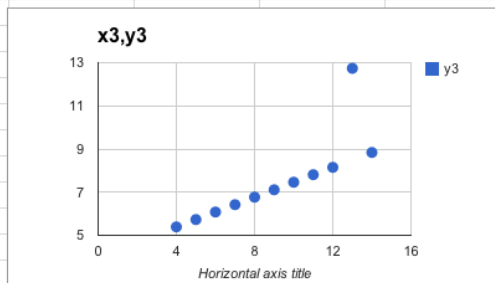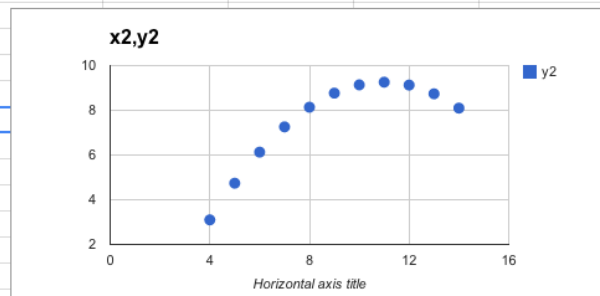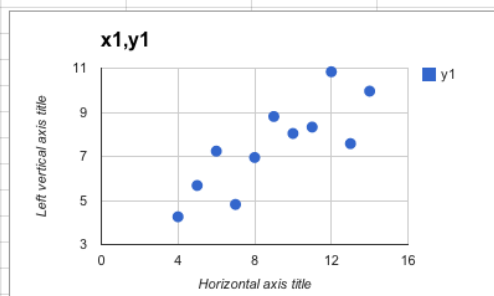
3 Task 5

Think of university study information system - all data about every student, grades, and dates, teachers, feedback, etc. Propose some visual report/summary that might help programme managers, head of the institutes, deans, and rector to decide about student progress in a curriculum (or comparisons in between curricula and institutes, for example). What information would need to be fitted on such a visualisation? Make an illustration of your idea (can be done on paper). Be as creative as possible.

Reports, summaries:

1. Report of students' exam results changing over multiple tries. Could show whether doing exam multiple times would make the result better.

2. Exam results in morning classes vs. afternoon classes. To that could be added results of a year where the morning classes were helt in afternoon and the other way around. If there was correlation with exam results, the harder class (with average result being lower) could be moved to a time slot where people get better grades (hopefully due to better comprehension).