Data Mining
Homework Assignment #4

Dmytro Fishman, Anna Leontjeva and Jaak Vilo

March 13, 2014

You are free to use any programming language you are comfortable with. Hints in R are optional.

Task 1

Listen to the presentation by Tamara Munzner: Keynote on Visualization Principles - `http://vizbi.org/Videos/26205288`

Task 2

Summarize the key take-home messages from her presentation. Take a look at three following charts:

- `http://www.billboard.com/biz/articles/news/digital-and-mobile/5827354/the-download-hits-middle-age-and-it-shows`
- `http://junkcharts.typepad.com/.a/6a00d8341e992c53ef019b021a098f970b-pi`, source
- `http://www.technologyreview.com/graphiti/520491/mobile-makeover/`

Choose one of them and answer following questions:

- What is the key message for the chart?
- Is it easy to grasp?
- Is it possible to improve it according to Tamara's suggestions? How?
- (Optional) Redraw it (either on paper or using some visualization tool) according to your suggestions.

Task 3

Read data from the file, data.txt, calculate mean and variance for every feature (column). Compute pairwise correlation (correlation(x1,y1),correlation(x2,y2),etc). Compare results. Make 4 plots: x on x-axis and y on y-axis. Make a conclusion.

Task 4

Perform a "Single Link" clustering of 2-D data from slide 28. Use Euclidean distance as a distance measure. Draw a dendrogram/tree with node height at the distance at where the clusters were merged. Hint: Draw the points first on 2D and then perform manual simulation. (Solutions on paper are ok).

	X	Y
A	2	4
B	7	3
C	3	5
D	5	3
E	7	4
F	6	8
G	6	5
H	8	4
I	2	5
J	3	7

Task 5

Task 6 (2pt)