Universität Stuttgart

Institut für Visualisierung und Interaktive Systeme

Assignment 7

Information Visualization & Visual Analytics (WS 2019/20)

Due: Monday, 9.12.2019, 12:00 **Discussion:** Wednesday, 11.12.2019

Point of contact: Max Franke <Max.Franke@vis.uni-stuttgart.de>

Please solve the assignment in **groups of up to three (3) students**. Choose one student who uploads your solution on the assignments page in ILIAS as PDF (for theoretical submissions) or ZIP (for practical submissions Impl). The submitted files should follow the naming scheme `yourlastname1_yourlastname2_yourlastname3` with respective file-ending, of course. Make sure that you create your team before uploading the solution.

Please post general questions regarding the exercise, but *no solutions,* to the forum. In case of more specific problems, especially such that cannot be posed outside of the context of your own solution, please send an email to, or make an appointment with, the tutor responsible for the exercise.

Task 1 The Information Visualization Reference Model *[Points: 8]*

Consider the Information Visualization reference model as discussed in the lecture. Describe at which stages in the model the interaction takes place for different user interactions listed below. In some cases, more than one stage might be affected. Justify your answers briefly.

(a) Our source data[1] is a list of citizens of the Greater London area. Each data item has a number of data attributes, one of them being their address. Initially, the dataset is visualized as a scatterplot map, where each point represents the place of residence for one citizen in the dataset (see Fig. 1a).

i. *(1 points)* The user now interacts with the visualization to visualize instead the average income per capita for each borough of London. The data is encoded by color into the areas in a choropleth map[2] (see Fig. 1b).

ii. *(1 points)* Continuing with the visualization from item i, the user interacts with the visualization to encode voter turnout per borough into the color of the areas instead (see Fig. 1c).

iii. *(1 points)* The user wants to see in which boroughs the voter turnout is larger than 50 %, and selects a diverging colorscheme (see Fig. 1d).

iv. *(1 points)* The user is now only interested in voter turnout in the southeast of the Greater London area, and interacts with the visualization to show only that part (see Fig. 1e).

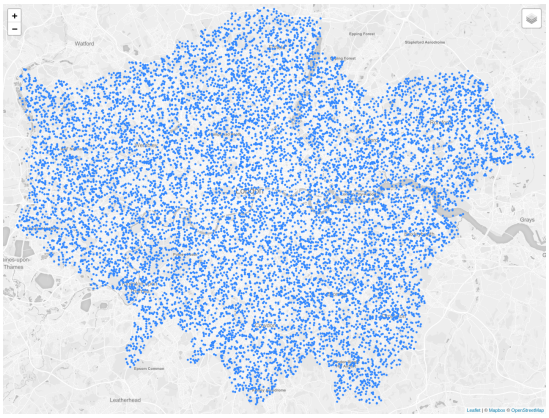
(b) *(1 points)* A barchart visualization shows the amount of cars with a certain number of cylinders in the `cars` dataset (see Fig. 2a). The user interacts with the visualization so that only the number of cylinders for cars manufactured 1970–1980 are shown (see Fig. 2b).

(c) *(1 points)* The user reorders the axes in a parallel coordinates plot (PCP) by drag-and-drop, because they want to see the relation between different axes (see Fig. 3).

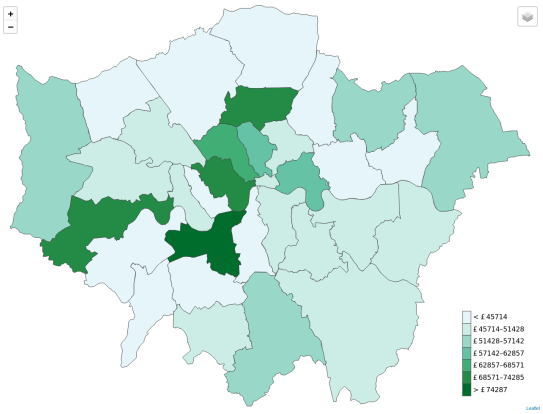
(d) *(2 points)* Consider the difference between *geometric zooming* and *semantic zooming*. How do they differ? With which stage or stages in the reference model do we interact in the two cases?

[1] The data was randomly generated and is not intended to convey any political or socioeconomical opinion.

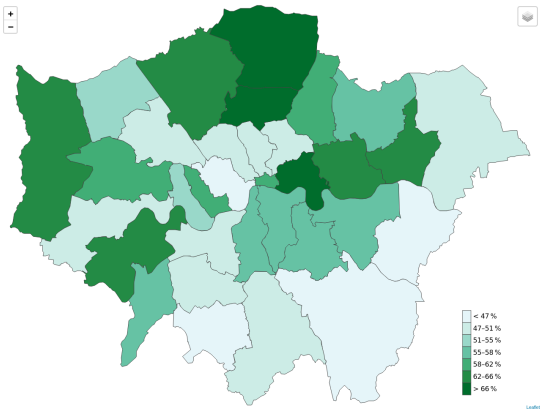
[2] `https://en.wikipedia.org/wiki/Choropleth_map`



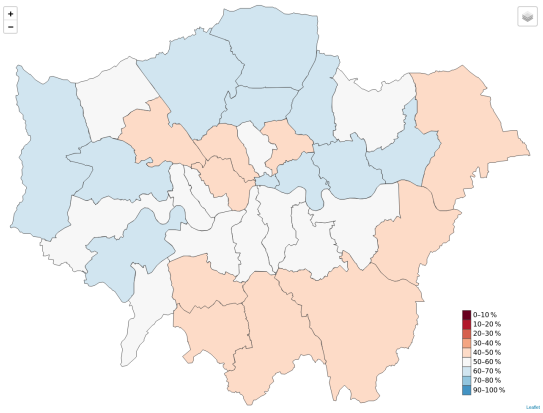
(a) Scatterplot map.



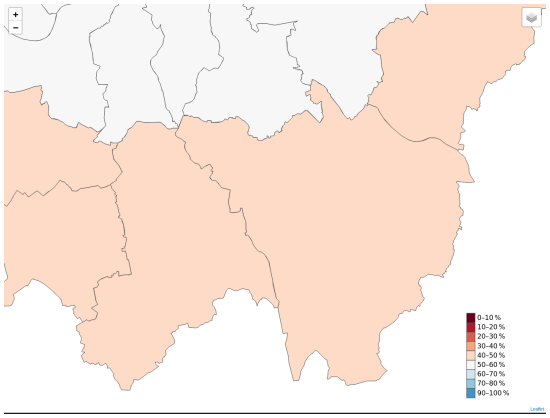
(b) Average income per capita.



(c) Voter turnout.



(d) Voter turnout, diverging colorscheme.



(e) Voter turnout, diverging colorscheme, southeast London.

Figure 1: Visualizations for the different stages of subtask 1a. Map data © OpenStreetMaps.

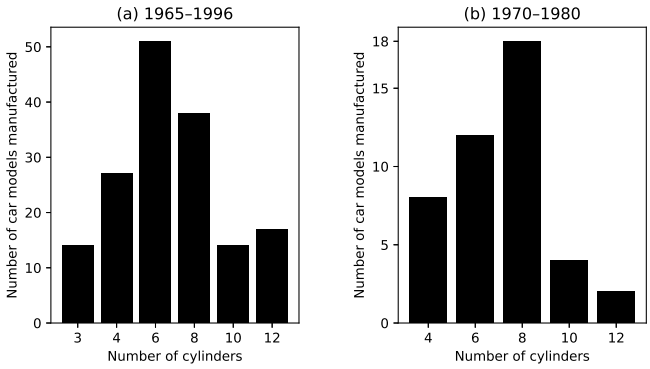
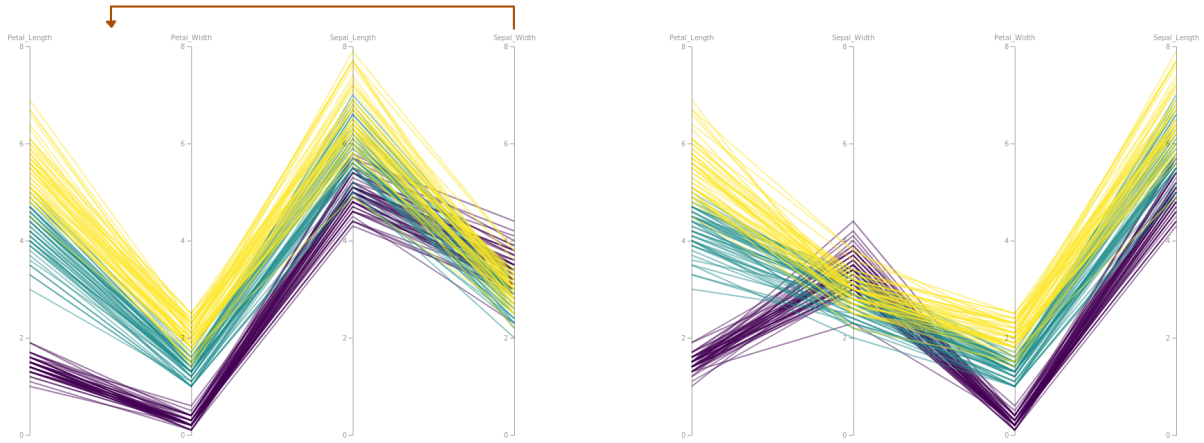


Figure 2: Number of cars manufactured with a certain amount of cylinders, for the entire dataset (a) and for the interval 1970–1980 (b).



(a) Initial order.

(b) After reordering.

Figure 3: Parallel coordinates plots of the Iris dataset.

Task 2 Interaction with Interactive Visualizations *[Points: 6]*

In this task, you will come up with interaction techniques to solve a particular task in a visualization. For the described situations, describe in detail how you would solve the problem. Add a sketch or picture that clearly shows your proposed solution, and discuss the scalability and limitations of your approach.

(a) *(2 points)* Consider a visualization of a graph structure as a node-link diagram. The edges of the graph are associated with a number of different data attributes which cannot all be visually encoded. Interactively selecting a link in the node-link diagram would therefore show this information in a separate view.

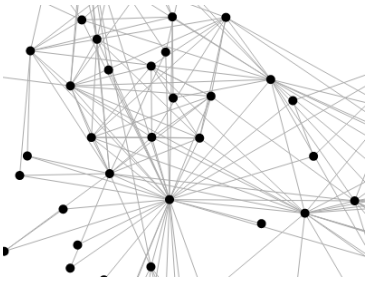
How would you design interaction to select one edge? Be aware that the node-link diagram might be very cluttered (see Fig. 4a).

(b) *(2 points)* Consider a scatterplot with very many markers. The markers are very dense in some areas, and may overlap (see Fig. 4b).

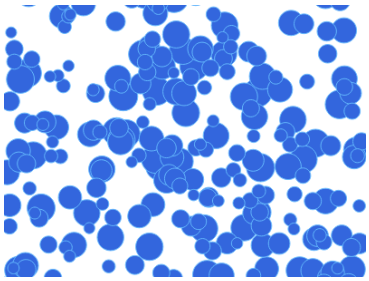
How would you design interaction to select one marker?

(c) *(2 points)* Now consider a scatterplot where markers usually do not overlap. However, the markers are very small, and hitting them with the mouse can be very difficult (see Fig. 4c).

How would you design interaction to select single markers without much hassle?



(a) Links in a node-link diagram.



(b) Overlapping, large markers.

(c) Sparse, small markers.

Figure 4: Figures for Task 2.

Task 3 k-d Trees *[Points: 7]*

Spatial data structures such as the k-d tree are required to efficiently lookup data points close to a specific position, without having to iterate over the entire dataset. This becomes necessary when the size of a visualized dataset grows.

(a) *(4 points)* Using the algorithm from the lecture, add the following datapoints to a 2-d tree in the given order:

$$(4, 2.5), (0.5, -0.5), (-2, 0.5), (1, -2.5), (2, 1), (3, 3.5), (0, 3)$$

Visualize the resulting k-d tree *both* as areas and as a node-link diagram (see Fig. 5).

(b) *(3 points)* The resulting tree from subtask a is not ideal. What is the problem with this tree, and how can it be fixed? Apply the fix and plot the resulting tree again as both areas and as a node-link diagram.

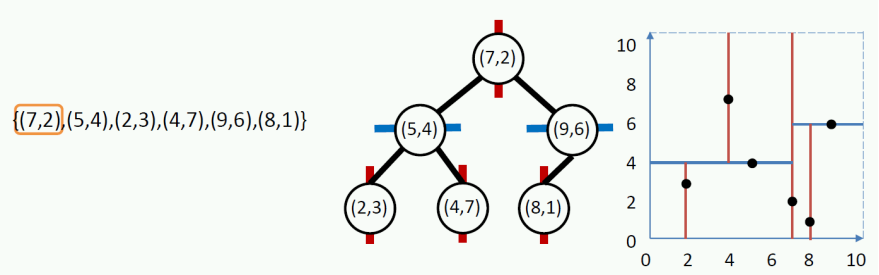


Figure 5: A k-d tree for a set of data points, visualized as areas (center) and as a node-link diagram (right).