**Review sheet**

**● Why is it important to learn data visualization?**

**○ The example of the historical cholera outbreak in London and its historical importance**

* Jon Snow made a map with dots and it changed history because you could see a pattern

**○ Voronoi diagram: a simple way to improve the cholera map even further. What is it and how does it work?**

* A partitioning of a plane into regions based on distance to points in a specific subset of the plane
* Drawing a Voronoi diagram: draw the midpoint lines between all the points
* Cholera map can be redrawn into a Voronoi diagram

**● History of visualizations**

* Started with data maps

**● Visualization integrity**

**○ Various types of misleading visualizations**

* + Simply wrong – the data is wrong
  + Correlation – correlation does not mean causation but if you try hard enough and long enough you can cook up as many crazy correlations as you want

**● Perception and design**

**○ Psychophysics**

* The branch of psychology that deals with the relationships between physical stimuli and mental phenomena

**○ Weber's law**

**○ Visual encodings**

**■ What are the visual encodings?**

**○ Stevens' power law**

**■ What is the Stevens' power law?**

**■ Why is Stevens' power law important?**

**○ Colors**

**■ Why are they important?**

**■ How to encode colors?**

**■ What are the basic concepts in colors?**

**■ Main usage types**

**■ How should we pick colors for categorical data?**

**■ How should we pick colors for quantitative data?**

**■ Accessibility of visualizations when using colors**

**○ Design principles**

**■ What are the gestalt principles?**

**■ Why are they important?**

**■ Data-ink ratio**

**■ Chart junk**

**■ Pre-attentive processing**

**● Data**

**○ Data types**

**■ Major data types and classification**

**○ Tidy data**

**■ What is tidy data?**

**■ Why is it important?**

**● 1D (univariate) visualization**

**○ 1D scatter plot (strip-chart)**

**○ Boxplot**

**■ Key ideas**

**○ Histogram**

**■ Key ideas**

**■ Choosing bins**

**■ Cumulative histogram**

**○ CDF & CCDF**

**■ Key ideas**

**○ Density Estimation**

**■ Why and how**

**■ Parametric vs. non-parametric approaches**

**■ Kernel density estimation**

**● Choices in KDE**

**○ Characteristics and comparison of univariate density visualization methods**

**○ Interpolation**

**○ Extrapolation**

**○ Moving average**

**○ Regression**

**● Log-scale**

**○ Log-scale basics**

**○ Why is it important?**

**○ Visualizing ratios**

**○ Drawing a histogram in a log-scale**

**○ CCDF**

**● 2D (bivariate) visualizations**

**● High dimensional data**

**○ Scatterplot matrix**

**○ Parallel coordinates**

**○ Radar charts**

**○ Dimensionality reduction**

**■ Data manifold. Why can dimensionality reduction work?**

**■ Common techniques: how do they work?**

**● MDS**

**● PCA**

**● t-SNE**