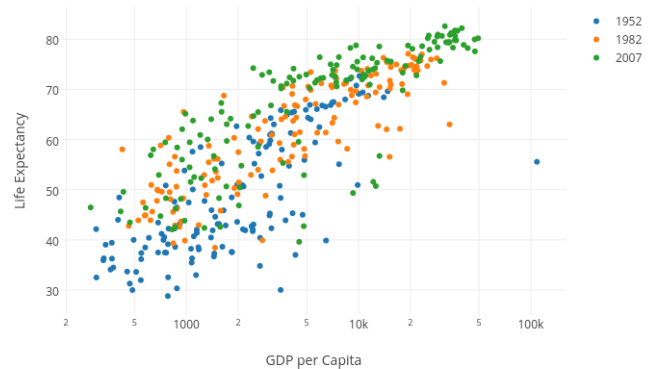
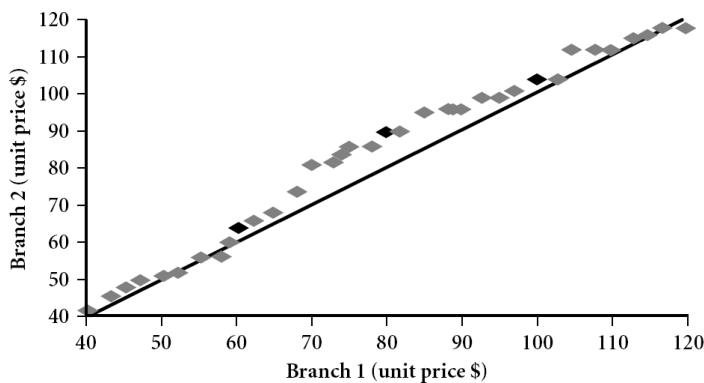
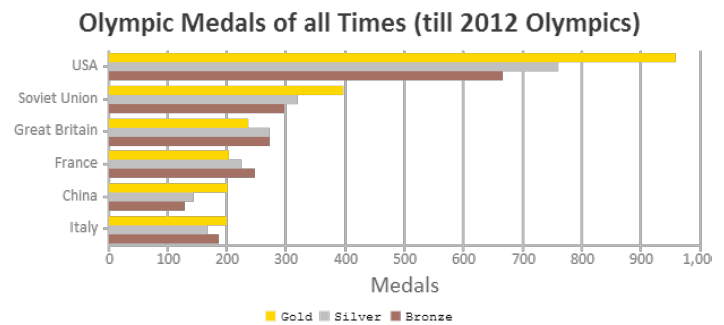
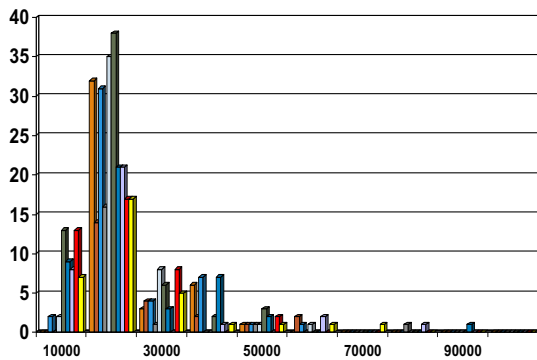


CSE 40647/60647 Data Science (Spring 2018)

Lecture 3: Data Preprocessing: Data Visualization

Goals:

- Understand dimensionality, sparsity, and resolution
- Describe and generate plots
 - Boxplot, histogram, bar chart
 - Quantile plot, Q-Q plot
 - Scatter plot
- Understand graphical integrity
- Describe data proximity
 - Describe data dissimilarity: Calculate three Minkowski distance measures
 - Describe data similarity: Calculate Jaccard similarity, cosine similarity, and KL divergence



Can you tell (what kind of question we want to answer) when we use scatter plot, and when we use Q-Q plot?

Name:

NetID:

Please write down whatever question you have about this course: