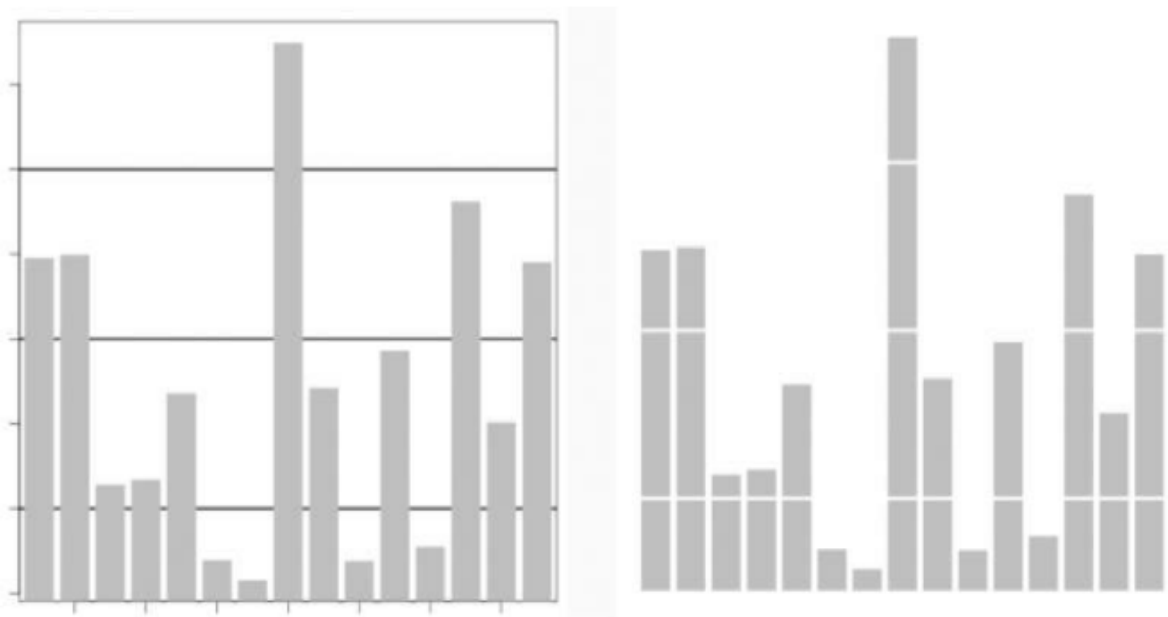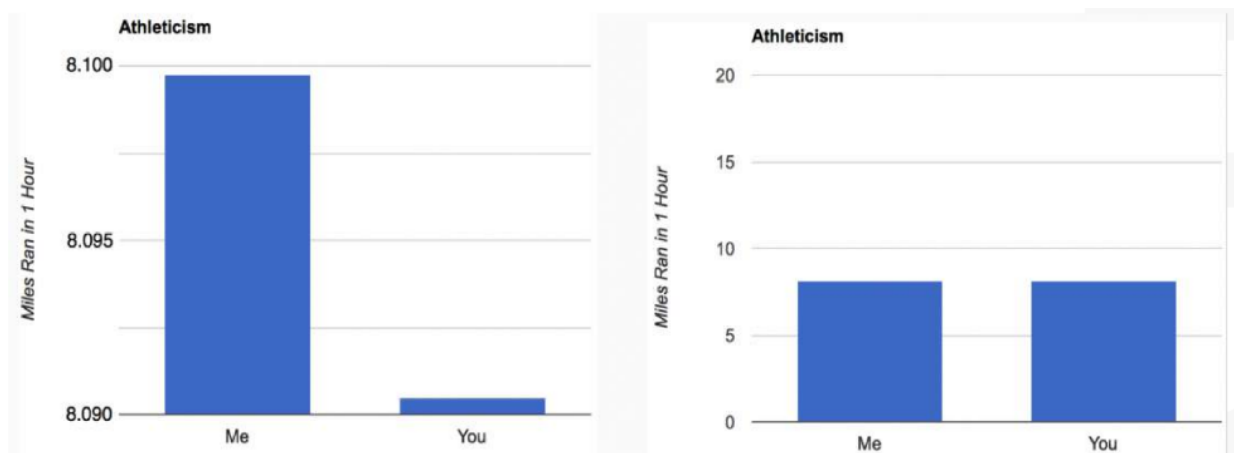# INT303 W5

PRINCIPLES OF VISUALIZATIONS

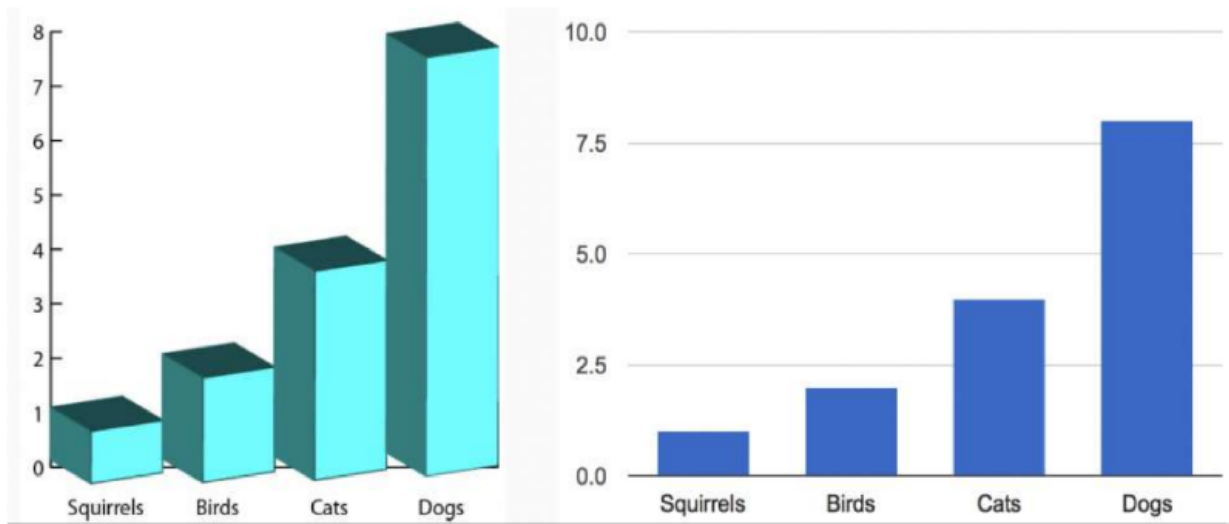Some basic data visualization guidelines from Edward Tufte:

1. Maximize data to ink ratio: show the data
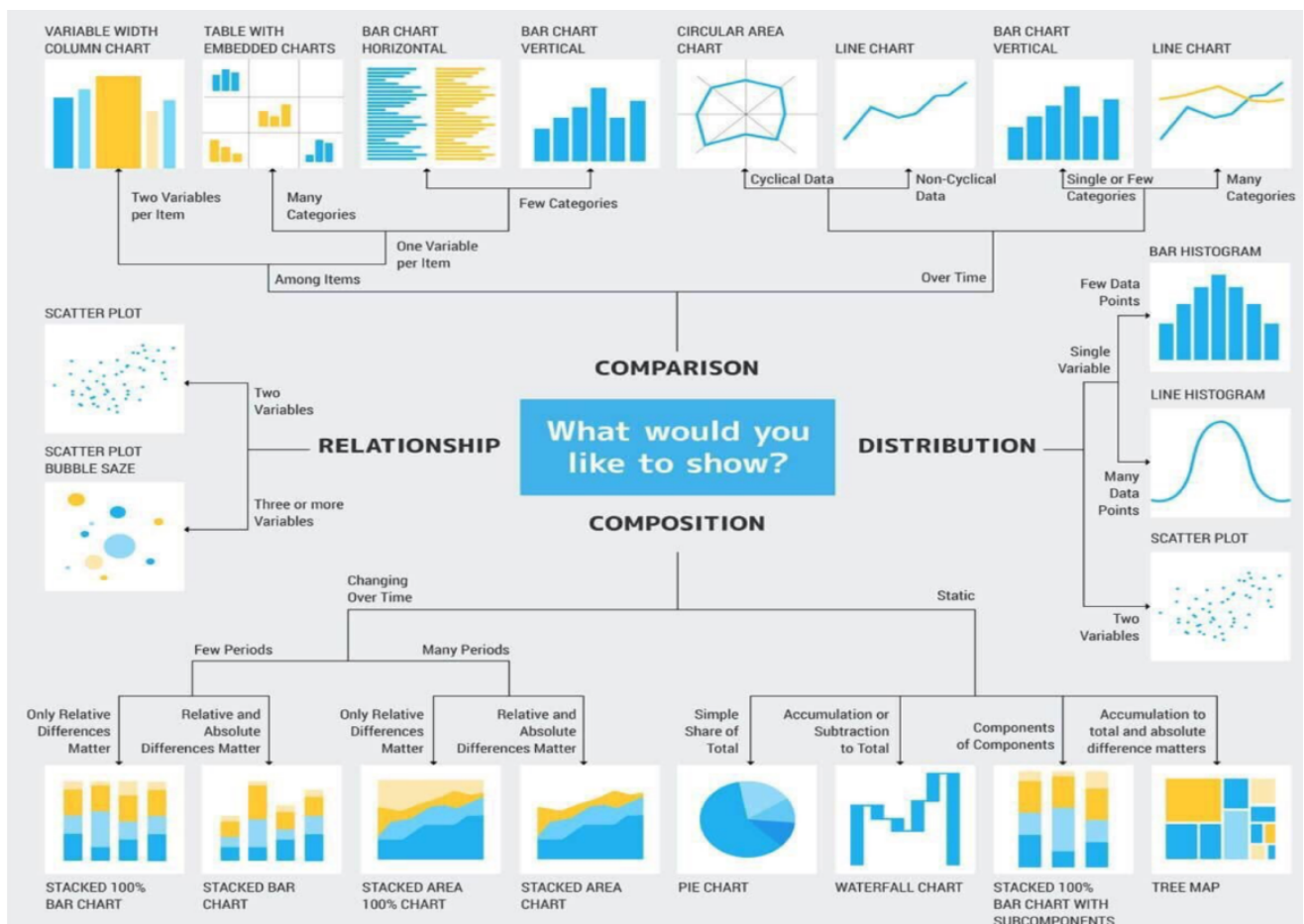

2. Don't lie with scale: minimize

3. Minimize chart-junk: show data variation, not design variation

4. Clear, detailed and thorough labeling
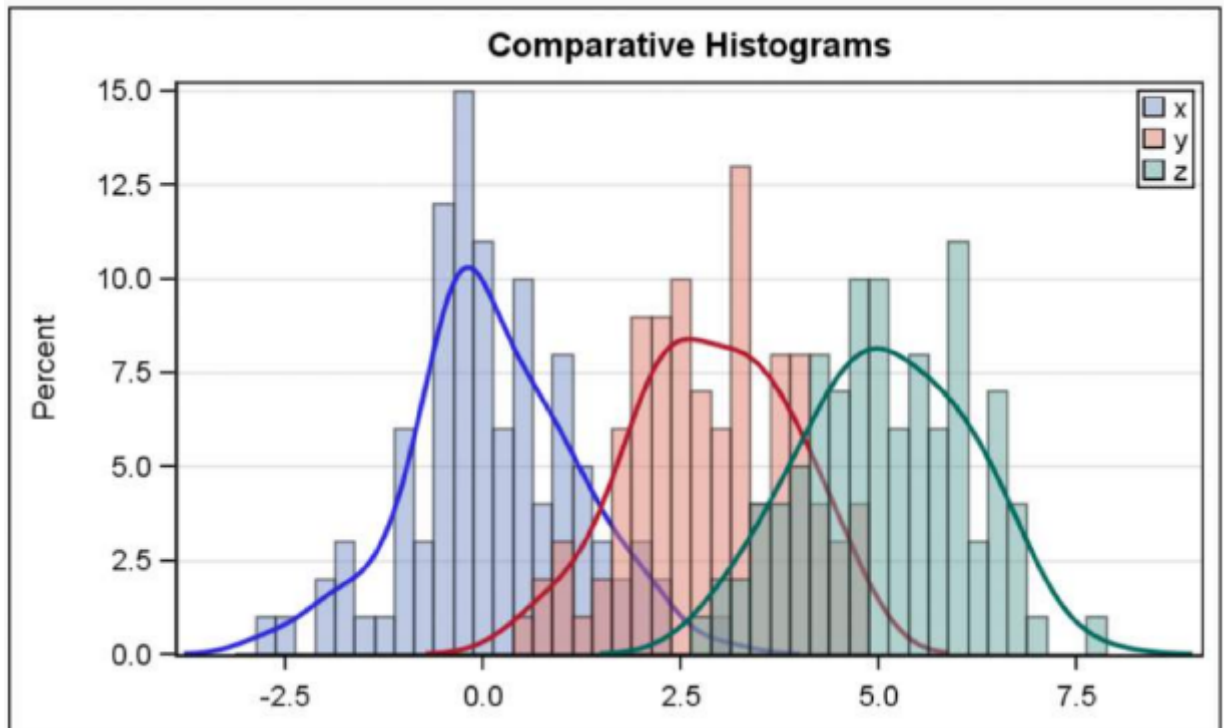
TYPES OF VISUALIZATIONS

- **Distribution**: how a variable or variables in the dataset distribute over a range of possible values.
- **Relationship**: how the values of multiple variables in the dataset relate
- **Composition**: how a part of your data compares to the whole.
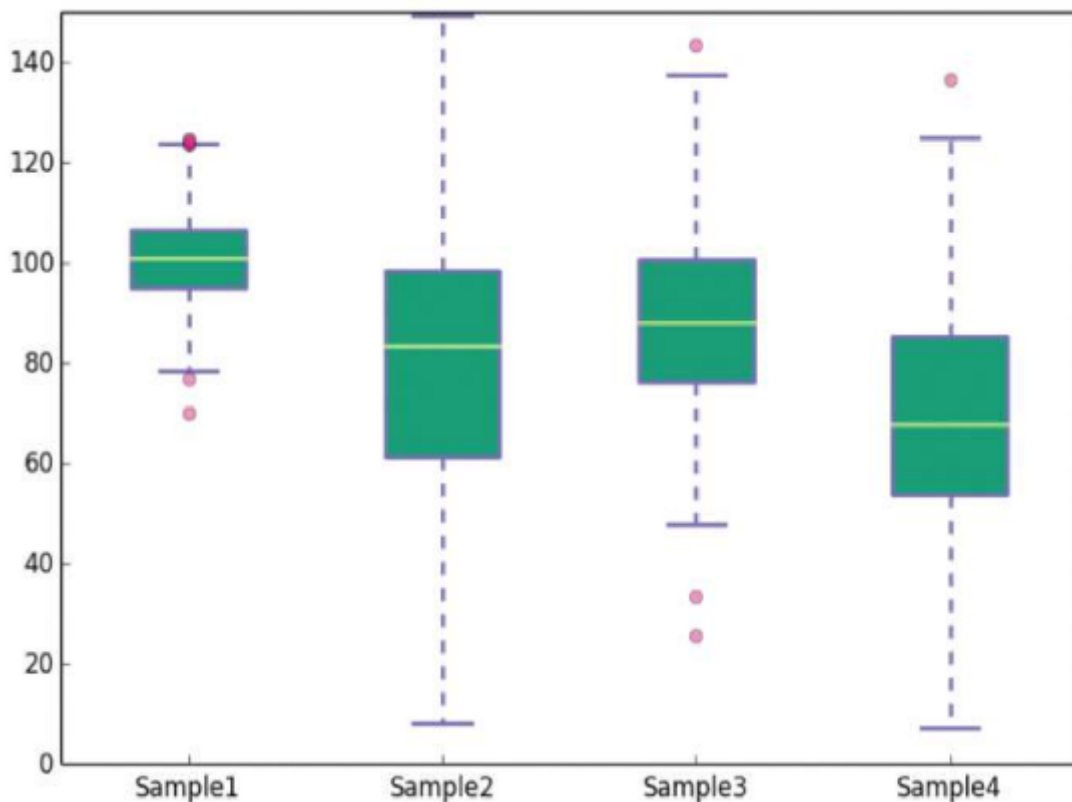- **Comparison**: how trends in multiple variable or datasets compare

补充：

Comparison

- multiple histograms

- boxplot: highlights the range, quartiles, median and any outliers present in a data set

QA

Q1: WHAT ARE SOME IMPORTANT FEATURES OF A GOOD DATA VISUALIZATION?

The data visualization should be light and must highlight essential aspects of the data; looking at important variables, what is relatively important, what are the trends and changes. Besides, data visualization must be visually appealing but should not have unnecessary information in it.

One can answer this question in multiple ways: from technical points to mentioning key aspects, but be sure to remember saying these points:

1. Maximize data to ink ratio: show the data
2. Don't lie with scale: minimize
3. Minimize chart-junk: show data variation, not design variation
4. Clear, detailed and thorough labeling

QUESTION2: WHAT IS A SCATTER PLOT? FOR WHAT TYPE OF DATA IS SCATTER PLOT USUALLY USED FOR?

A scatter plot is a chart used to plot a correlation between two or more variables at the same time. It's usually used for numeric data.

QUESTION3: WHAT TYPE OF PLOT WOULD YOU USE IF YOU NEED TO DEMONSTRATE "RELATIONSHIP" BETWEEN VARIABLES/PARAMETERS?

When we are trying to show the relationship between 2 variables, scatter plots or charts are used. When we are trying to show "relationship" between three variables, bubble charts are used.