

SCI1020: Introduction to Statistical Reasoning

EXPLORING DATA

CHAPTER 1: Picturing distributions with graphs

Individuals and variables

- **Individuals:** objects described by a set of data.
- **Variable:** characteristic of a variable.
 - o **Categorical variable:** places an individual into one of several groups or categories.
 - **Nominal:** no natural ranking or ordering.
 - **Ordinal:** ordering exists.
 - All categorical variables have discrete number of sub-groups and a count in each.
 - o **Quantitative variable:** takes numerical values for which arithmetic operations such as adding, and averaging make sense.
 - **Continuous:** can take on any value in an interval. Arithmetic operations such as adding, and averaging make sense.
 - **Discrete:** can only take particular values (e.g. number of students in a course).

Categorical variables: Pie charts and bar graphs

- Exploring data:
 - o Begin by examining each variable by itself. Then move on to study the relationships among the variables.
 - o Begin with a graph or graphs. Then add numerical summaries of specific aspects of the data.
- **Distribution:** tells us what values it takes and how often it takes these values.
 - o Distribution of a categorical variable lists the categories and gives either the count or the percent of individuals who fall in each category.
 - o Use a pie chart or bar graph to display this.
 - o Bar chart easy to read and can compare set of quantities measured in the same units.

Quantitative variables: Histograms

- Histogram displays distribution of a quantitative variable.

Interpreting histograms

- Displays cross-sectional data.
- Examining a histogram:
 - o Look for overall pattern and for striking deviations from that pattern.
 - Shape
 - Center
 - Variability (spread).
 - o **Outlier:** individual value that falls outside the overall pattern.
- Symmetric and skewed distributions