

Organizing and Displaying Categorical Data

Categorical data can be organized into a **frequency table** which counts the number of cases that fall into each category, or a **relative frequency table** which measures the percentage of the data set that falls into each category.

Categorical data can be visualized in a **bar graph**. Separate bars labelled by category have heights determined by the frequency or proportion (percentage of the total) of data in that category. A **pie chart** does the same: categories are represented by labeled sectors of a circle, and the proportion of data in that category equals the percentage of the circle assigned to that sector.

All statistical displays obey **the Area Principle**: *each data value should be represented by the same amount of area.*

Paired Categorical Data

- **contingency table**
table of counts of units that fall into each of the different pairs of categories when considering data representing two categorical variables, the **row variable** and **column variable**
- **marginal distribution**
information provided by the **row totals** (frequencies for the row variable) and **column totals** (frequencies for the column variable) in a contingency table
- **conditional distribution**
distribution of data across one categorical variable which fall into a specific category of the other variable
- **independence**
variables are independent if the conditional distribution of one of the variables is the same across each category of the other
- **Simpson's paradox**
the reversal in an association between categorical variables that occurs when data are averaged across groups