

## Chapter 3 Summary

Three Rules of Data Analysis- Make a picture-a display of the data will reveal things that can't be seen in a table of numbers and will help to think clearly about the patterns and relationships that may be hiding in the data. Make a picture- A well designed display will show the important features and patterns of the data. A picture will also show us the things we did not expect to see. Make a picture- the best way to tell others about the data us with a picture.

Category	Count	Percentage
A	10	25%
B	15	37.5%
C	5	12.5%
D	10	25%
Total	40	100%

In order to make a picture of data, the first thing we have to do is make piles. For categorical data we just count the number of cases corresponding to each category and pile them up. We can organize. Sometimes we want to know the fraction or proportion of the data in each category so we divide the counts by the total number of cases. Usually we multiply by 100 to express these **proportions** as percentages. A **relative frequency table** displays the **percentages**, rather than the counts, of the values in each category. Both types of tables show how the cases are distributed across the categories. In this way, they describe the **distribution** of a categorical variable because they name the possible categories and tell how frequently each occurs.

**The Area Principle** says that the area occupied by a part of the graph should correspond to the magnitude of the value it represents.

A **bar chart** displays the distribution of a categorical variable, showing the counts for each category next to each other for easy comparison.

If we want to draw attention to the relative proportions then we could replace the counts with percentages and use a **relative frequency bar chart**.

**Pie charts** show the whole group as a circle. They slice the circle into pieces whose size is proportional to the fraction of the whole in each category.

Before you make a bar chart or a pie chart always check the **Categorical Data Condition**: The data are counts or percentages of individuals in categories.

In a **contingency table** the table shows how the individuals are distributed along each variable.

In a contingency table, the distribution of either variable along is called the **marginal distribution**. The counts or percentages are the totals found in the margins of the table.

The distribution of a variable restricting the Who to consider only a smaller group of individuals is called a **conditional distribution**.

In a contingency table when the distribution of one variable is the same for all categories of another it is **independent**.

**Simpson's paradox** is when the averages are taken across different groups and they seem to contradict the overall averages.

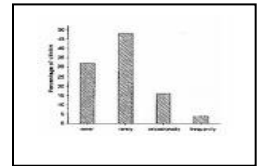


Figure 17

	High Income	Not High Income
Deception	10	5
No Deception	5	10