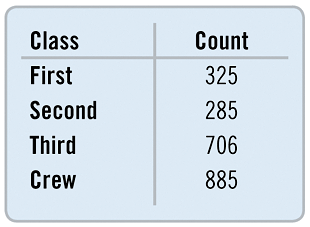
**Chapter 2 Displaying and Describing Categorical Data**

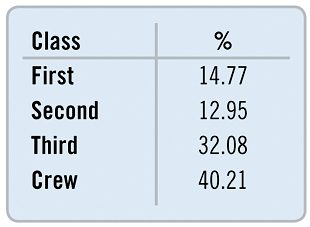
Three Rules of Data Analysis:

1. **Make a Picture** - to “Think” clearly; to see patterns and relationships you might not be able to see in a table of numbers.
2. **Make a Picture** – to “Show” the important features and patterns of your data.
3. **Make a Picture** – to “Tell” others about your data.

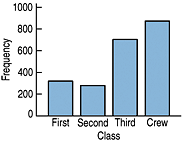
Frequency Table: A table with counts (frequencies) in each category.

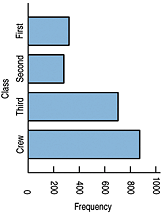


Relative Frequency Table: A table with decimals or percents in each category. We find the decimals by taking the frequency of each category / total number of data values. We can leave that as a decimal or express it as a percentage.



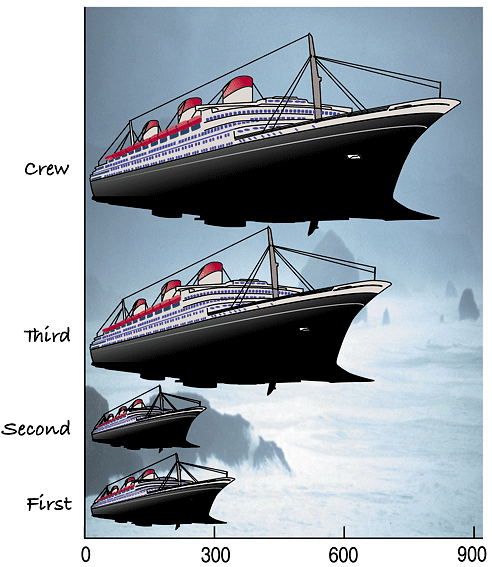
Bar Chart – visual display of frequencies/relative frequencies for categorical data.

or

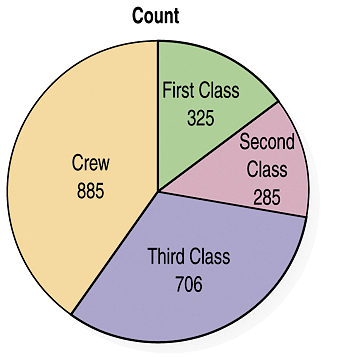


Area Principle: the area occupied by a part of the graph should correspond to the magnitude of the value it represents.

Example: Titanic data, p.16. Since there were about 3 times as many crew as second-class passengers, the ship depicting crew is about 3 times longer as the picture depicting 2nd class passengers, but it occupies about 9 times the area.

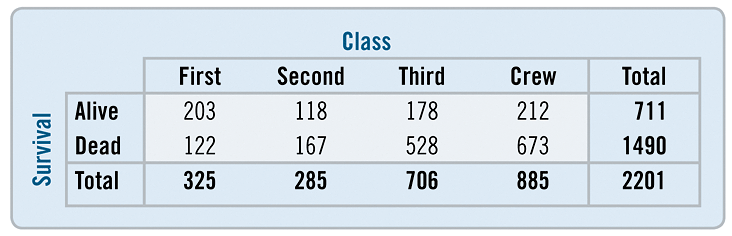


Pie Chart – visual display of categorical data; each “slice” is found by taking the relative frequency of each category and applying that same percentage to the 360circle.



Contingency Tables: a table that allows us to look at two variables of the data at the same time.

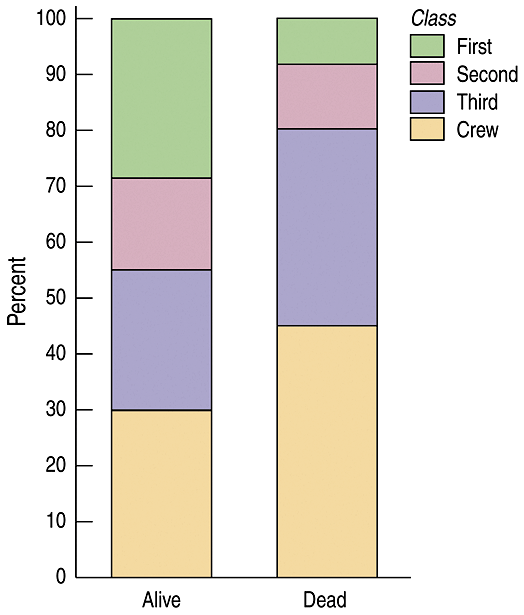
The variables are **independent** when the distribution of one variable is the same for all categories of the other.



**Marginal distributions**: row/column total percentages (margins of the table); a frequency distribution for one of the variables.

**Conditional distributions**: a specific row/column (NOT totals); show the distribution of one variable for just those cases that satisfy a condition on another variable.

Segmented Bar Charts: the bars represent the “whole” of one variable and is divided proportionally into segments corresponding to the percentage in each group of the other variable.

****

Simpson’s Paradox: unfair averaging over different groups; comparing individual averages against overall averages can yield different (opposite) results.

Example, p.29

