Categorical variables represent types of data which may be divided into groups. Examples of categorical variables are race, sex, age group, and educational level. While the latter two variables may also be considered in a numerical manner by using exact values for age and highest grade completed, it is often more informative to categorize such variables into a relatively small number of groups.
Analysis of categorical data generally involves the use of data tables. A ***two-way table*** presents categorical data by counting the number of observations that fall into each group for two variables, one divided into rows and the other divided into columns. For example, suppose a survey was conducted of a group of 20 individuals, who were asked to identify their hair and eye color. A two-way table presenting the results might appear as follows:

 Eye Color

Hair Color Blue Green Brown Black Total

-----------------------------------------------------

Blonde 2 1 2 1 6

Red 1 1 2 0 4

Brown 1 0 4 2 7

Black 1 0 2 0 3

-----------------------------------------------------

Total 5 2 10 3 20

The totals for each category, also known as ***marginal distributions***, provide the number of individuals in each row or column without accounting for the effect of the other variable (in the example above, the total number of individuals with blue eyes, regardless of hair color, is 5).

Since simple counts are often difficult to analyze, two-way tables are often converted into percentages. In the above example, there are 4 individuals with red hair. Since there were a total of 20 observations, this means that 20% of the individuals survered are redheads. One also might want to investigate the percentages within a given category -- of the 4 redheads, 2 (50%) have brown eyes, 1 (25%) has blue eyes, and 1 (25%) has green eyes.

For a more detailed example, consider the following dataset, "Weights of 1996 US Olympic Rowing Team." The first column gives the name of the rower, the second gives his event, and the third gives his weight. There are 8 different event categories, with weight given as numeric data.

Auth LW\_double\_sculls 154 Klepacki four 205

Beasley single\_sculls 224 Koven eight 200

Brown eight 214 Mueller quad 215

Burden eight 195 Murphy eight 220

Carlucci LW\_four 160 Murray four 205

Collins,D LW\_four 155 Peterson,M pair 210

Collins,P eight 195 Peterson,S LW\_double\_sculls 160

Gailes quad 205 Pfaendtner LW\_four 160

Hall four 195 Schnieder LW\_four 158

Holland pair 195 Scott four 208

Honebein eight 200 Segaloff coxswain 121

Jamieson quad 210 Smith eight 207

Kaehler eight 210 Young quad 207