

2.1 Relationships Between Two Categorical Variables Part 2

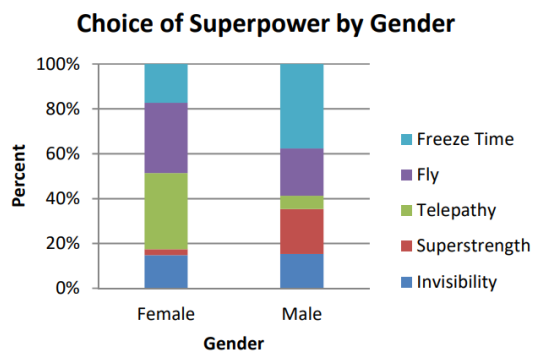
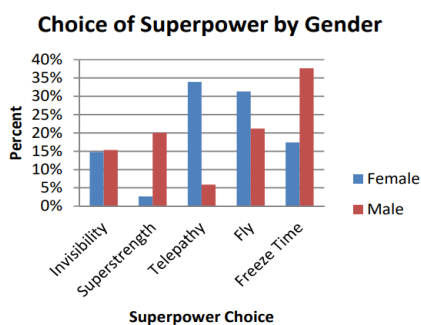
2.1 Relationships Between Two Categorical Variables (Part 2)

After identifying the explanatory variable, the next step is to display the distribution of the response variable for each category of the explanatory variable. This can be done with the side-by-side bar charts from Lesson 1.2 or with **segmented bar charts**.

A **segmented bar chart** displays the possible values of a categorical variable as segments of a rectangle, with the area of each segment proportional to the percent of individuals in the corresponding category.

Segmented bar charts are sometimes called "stacked" bar charts because they are the result of stacking the bars from a side-by-side bar chart.

In both types of graphs, we calculate the relative frequencies separately *for each gender* because gender is the explanatory variable. That is, we calculate the percent of *the females* who prefer each superpower. Then, we calculate the percent of *the males* who prefer each superpower.



2.1 Relationships Between Two Categorical Variables Part 2

How to Make a Segmented Bar Chart

1. **Identify the variables.** Determine which variable is the explanatory variable and which is the response variable.
2. **Draw and label the axes.** Put the name of the explanatory variable under the horizontal axis. To the left of the vertical axis, indicate if the graph shows the percent (or proportion) of individuals in each category of the response variable.
3. **Scale the axes.** Write the names of the categories of the explanatory variable at equally spaced intervals under the horizontal axis. On the vertical axis, start at 0% (or 0) and place tick marks at equal intervals until you reach 100% (or 1).
4. **Draw "100%" bars** above each of the category names for the explanatory variable on the horizontal axis so that each bar ends at the top of the graph. Make the bars equal in width and leave gaps between them.
5. **Segment each of the bars.** For each category of the explanatory variable, calculate the relative frequency for each category of the response variable. Then divide the corresponding bar so that the area of each segment corresponds to the proportion of individuals in each category of the response variable.
6. **Include a key** that identifies the different categories of the response variable.

Example: In 1912, the luxury liner *Titanic*, on its first voyage across the Atlantic, struck an iceberg and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table gives information about adult passengers who survived and who died, by class of travel. Make a segmented bar chart to display the relationship between survival status and class of travel for passengers on the *Titanic*.

		Class of travel			
		First	Second	Third	Total
Survival status	Survived	197	94	151	442
	Died	122	167	476	765
Total		319	261	627	1207

- ① go to <http://highschool.bfwpub.com/spa3e>
- ② click on "student site" on the left
- ③ scroll till you find "Two Categorical Variables"
- ④ enter the two-way table.