Are you more likely to survive a boat crash if you have a first-class ticket? Do boys and girls like different colors? These are questions that can be answered by examining ***distributions*** of data.

Individuals: The objects described by a set of data. Individuals may be people, animals, or objects.

Categorical Variable: A characteristic of an individual that places the individual into one of several groups or categories. Examples: Eye color, favorite ice cream flavor, gender, whether or not the person did their homework…

Relative Frequency: The *fraction or percent* of a group who fall into a category.

Distribution of a Categorical Variable: Lists the different categories that the individuals in the data set fall into and states how many (or what percent) of individuals fall into each category.

Two-way Table: A table that is broken down into rows and columns. The values of one categorical variable go along the rows, and the values of another categorical variable go down the columns. A two-way table helps us see if there's any relationship between the two variables.

Marginal Distribution: Ignores the inside of the table, and just gives the *percent of all the individuals in the whole sample* who fall into each category. **Look at the totals in the <u>margins</u>!**

Conditional Distribution: Gives the *percent of individuals in just one sub-group* who fall into each category. (For example, the percent of <u>just the boys</u> who like each color instead of the percent of everyone in the entire sample who likes each color.)

Example: In 1912, the *Titanic* hit an iceberg on its first voyage across the Atlantic and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table gives information about adult passengers who lived and who died, based on what type of ticket they had.

	Survival Status		
Type of Ticket	**Survived**	**Died**	**Total**
First Class	203	122	
Second Class	118	167	
Third Class	178	528	
Crew	212	673	
Total			

a) Give the marginal distribution of survival status. (For everyone in the entire sample, what percent survived and what percent died?)

b) Give the marginal distribution of type of ticket. What does this distribution tell you about?

c) Give the conditional distribution of survival status for people with 1st class tickets.

d) Give the conditional distribution of survival status for people with 2nd class tickets.

e) Give the conditional distribution of survival status for people with 3rd class tickets.

f) Give the conditional distribution of survival status for the crew.

g) Draw side-by-side bar graphs to compare the distributions in parts c-f. Then write a few sentences comparing and contrasting the conditional distributions.

h) One of your friends tries to argue that 3rd class tickets were actually better than 2nd class tickets by saying, "A higher number of 3rd class ticket holders survived the Titanic disaster than 2nd class ticket holders." Explain what is misleading about this statement.

i) What percent of the passengers in 1st class survived?

j) What percent of the survivors were in 1st class?

Take a survey in your class of favorite colors and fill in the following table:

	Favorite Color								
	Red	Orange	Yellow	Green	Blue	Purple	Pink	Other	Total
Boys									
Girls									
Total									

a) Give the marginal distribution of favorite color for your class.

b) Give the marginal distribution of gender for your class.

c) Give the conditional distributions of favorite color for boys and girls.
     Boys: Girls:

d) Write a few sentences comparing and contrasting the conditional distributions of favorite color for boys and girls.

e) What percent of the girls in the class chose blue?

f) What percent of the people who chose blue are girls?

g) Was your answer to part f) part of a marginal distribution or a conditional distribution?

h) What percent of the people in the class chose red, orange, or yellow as a favorite color?

i) Was your answer to part h) part of a marginal distribution or a conditional distribution?