

Graphing Categorical Data

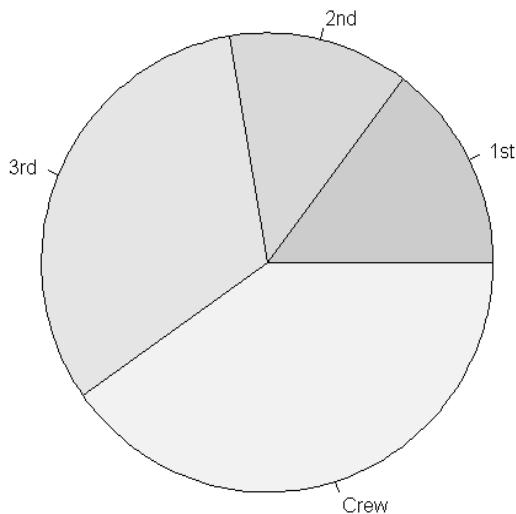
- A. one-way frequency tables
  - 1. pie charts
  - 2. bar graphs
- B. two-way frequency tables (contingency tables)
  - 1. stacked bar graph
  - 2. side-by-side bar graph
  - 3. bar graph of proportions

A. one-way frequency tables

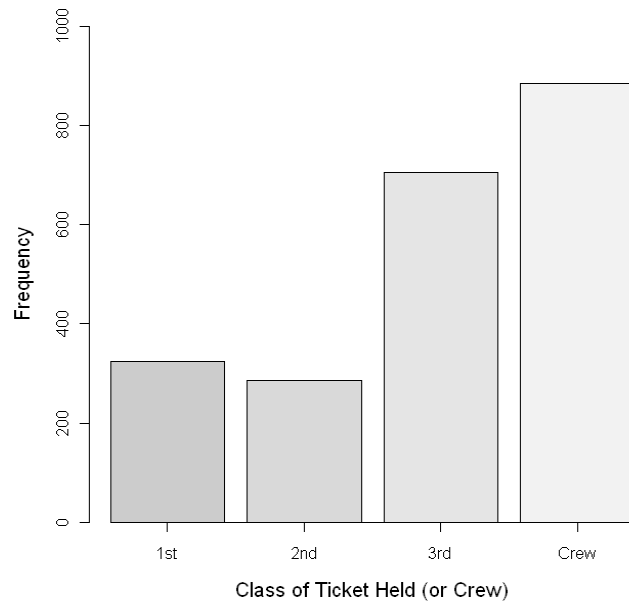
Class

1st	2nd	3rd	Crew
325	285	706	885

**Passengers by Class Aboard Titanic (N=2201)**



**Passengers by Class Aboard Titanic (N=2201)**

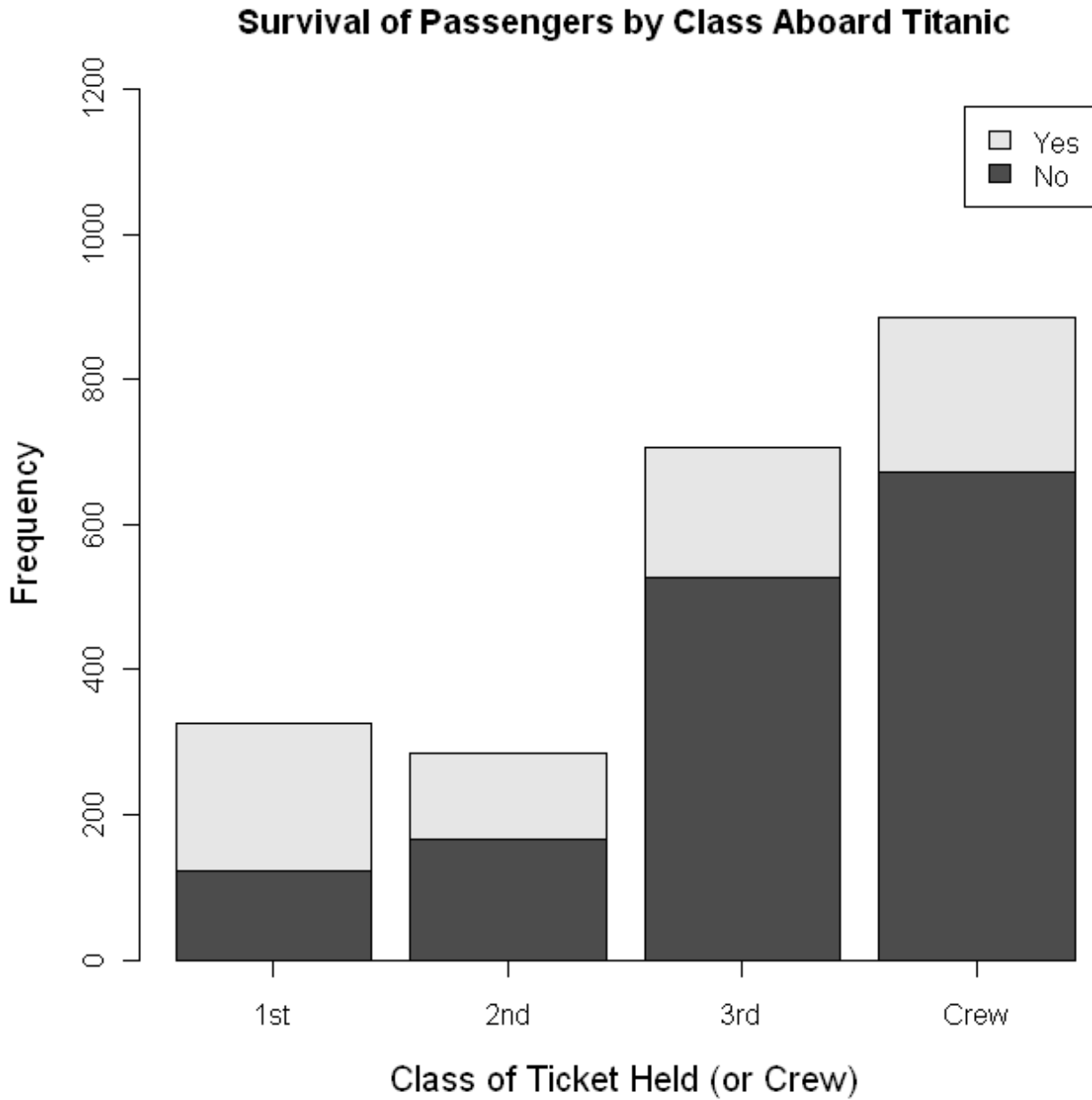


Most statisticians object strenuously to pie charts, and even more strenuously to fancy exploded pie charts or three dimensional ones. Such flashy but uninformative features of statistical graphs are sometimes called "chart junk." It distracts from the information in the graph rather than adding to it.

The reason for this objection to pie charts is illustrated well in this case. In the pie chart, it's simply hard to tell how much of a difference there is between 1st and 2nd class. This is much more obvious in the bar graph. In addition, the pie chart doesn't actually contain the frequency information (although it can be added). In short, use a bar graph unless you have a really good reason for doing otherwise.

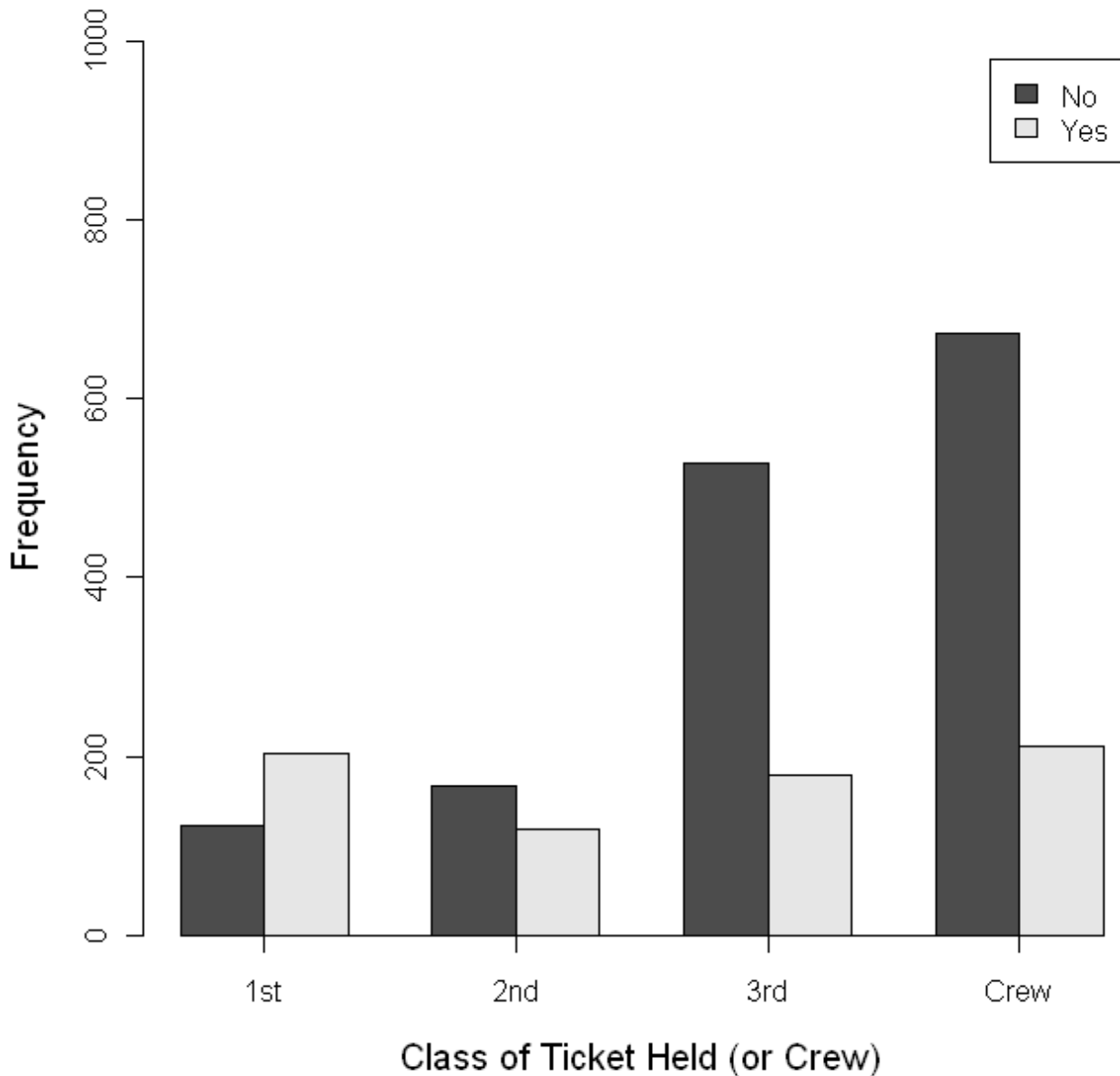
B. two-way frequency tables (contingency tables)

Class	Survived	
	No	Yes
1st	122	203
2nd	167	118
3rd	528	178
Crew	673	212



Stacked bar graphs seem to be popular with magazines and newspapers. I can't imagine why.

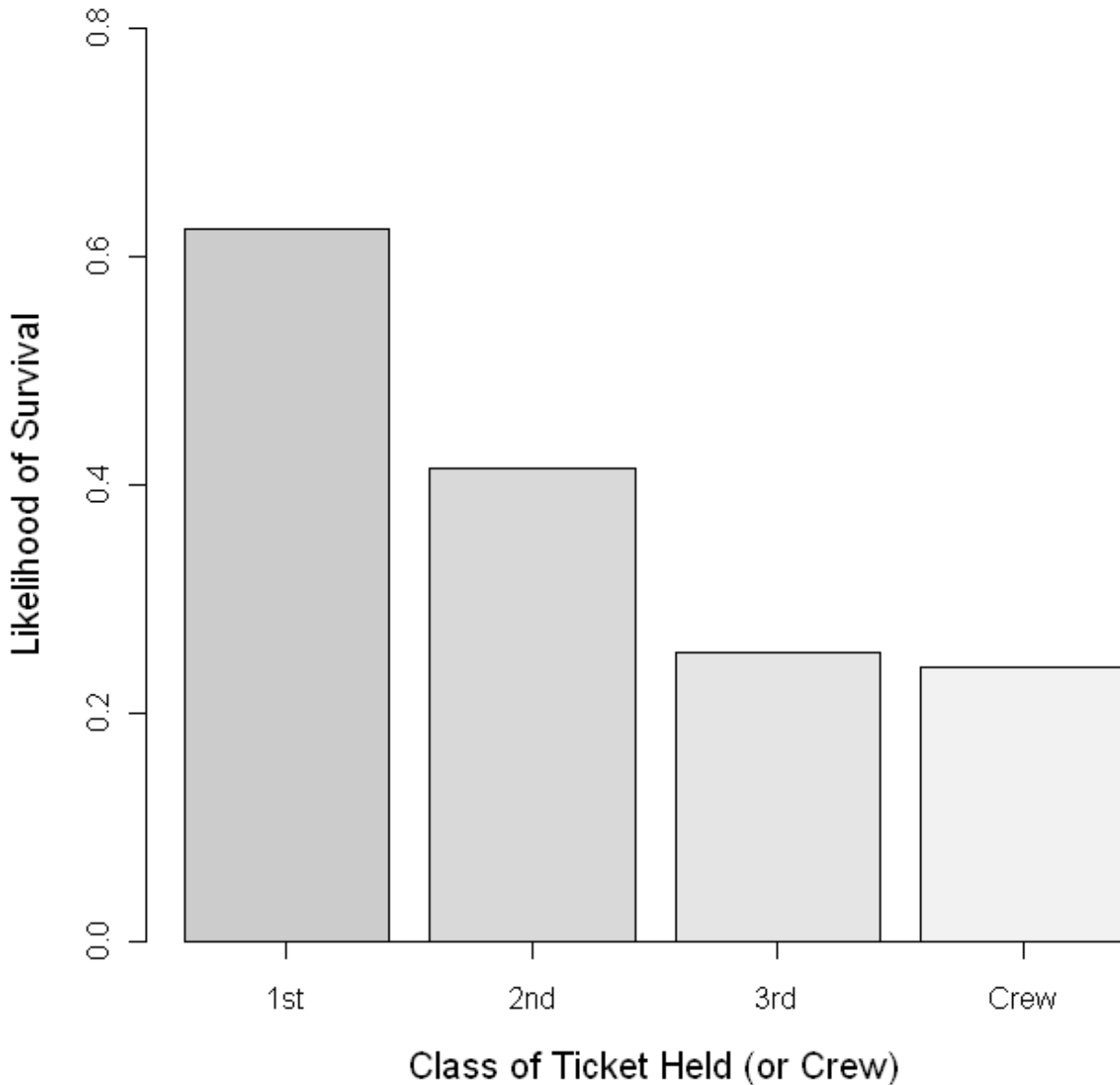
## Survival of Passengers by Class Aboard Titanic



A side-by-side bar graph is much more informative. Notice that all the elements of a good graph are present: clear axis labels, clear axis value labels, and a legend so that we can tell which bars are which. My criticism here would be of the legend. No and Yes what? A graph should stand on its own. The reader should not have to hunt around through the text of your research paper to find out what the graph is telling her. In the paper, graphs (called figures) will be accompanied by a figure caption that will explain it further. Notice that this particular graph does not directly compare the class groups because of the widely different group sizes. If that is your

objective, you might consider plotting proportions (also called relative frequencies or likelihoods), which adjust for group sizes. On the other hand, the graph above does display more information than a graph of proportions would. The one you should choose depends upon the story you want your data to tell.

### Survival of Titanic Passengers by Class of Ticket Held



How does this graph tell a different story from the previous one? Why was likelihood of survival chosen rather than likelihood of dying? And why not plot them both in a side-by-side bar graph? (In such a graph, the vertical axis would simply be labeled "Likelihood," and each "Class" would have two bars above it, one for survival and one for dying or nonsurvival.)