

Chapter 03: Displaying and Describing Categorical Data

Frequency Tables

The first step in creating a graph for qualitative data is to create a frequency table. This is a list of the values that were observed, and a count of how many times each was observed (the *frequency*).

In reality, you would have to do this yourself...in AP Statistics, it is likely that the frequency table would be given to you.

Example

[1.] A game uses a six sided die that does not have numbers on the sides—rather, each side has a color. The colors of the sides are white, black, red yellow, green and blue. The die is rolled twenty times and the color noted for each roll:

Table 1 - Die Roll Results for Example 1

black	black	black	yellow	green	green	black	red	blue	yellow
blue	white	blue	red	green	blue	blue	black	green	red

Construct a frequency table of these data.

Just start counting! As you encounter a new value, add a row (or column) to your table—the order of the values doesn't really matter at this point.

Table 2 - Die Roll Frequency Table for Example 1

Color	Black	White	Red	Yellow	Green	Blue
Count	5	1	3	2	4	5

Graph Types

Bar Charts

I would find it hard to believe that you've never seen a bar chart before. There are, though, several varieties! One word of caution—**never ever, under any circumstances, construct a 3D bar chart.**

For all bar charts, be sure to scale and label each axis. *Scale* means to write out the values of the variable along the axis; *label* means to tell the name of the variable being measured along that axis.

Standard

The *Standard Bar Chart* lists values of the variable on the horizontal axis and frequency on the vertical axis (there are some people who reverse those axes...). A bar is drawn for each variable value, and the bars do not touch one another. The order of the values on the horizontal axis does not matter.

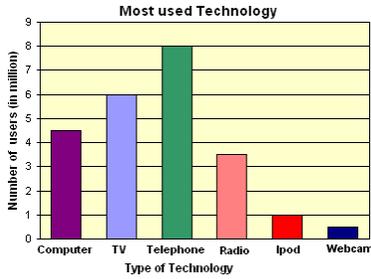


Figure 1 - Standard Bar Chart

Pareto Charts

A *Pareto Chart* is a bar chart where the values of the variable are arranged so that the bars decrease in height from left to right.

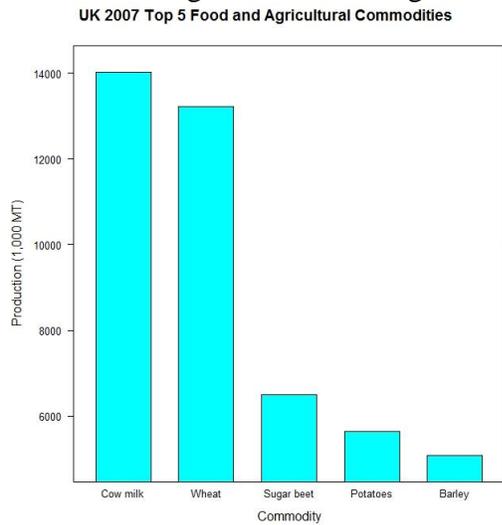


Figure 2 - Pareto Chart

Side-by-side

Sometimes we have two groups of measurements using the same variable—for example, the colors of cars and the colors of trucks. Side-by-side bar charts have two bars where the standard bar chart has only one—one bar for each group from which measurements were taken.

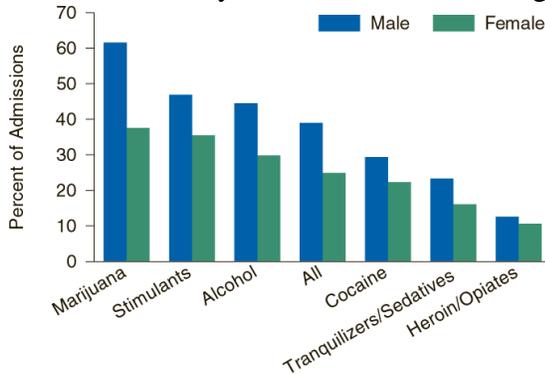


Figure 3 - Side-by-side Bar Chart

Segmented

A segmented bar chart (also known as a *stacked bar chart*) is used when we have two different qualitative variables in a single frequency table. One of the variables is listed along the horizontal axis, and the values of the other variable make up segments of each bar vertically.

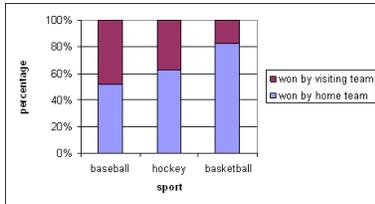


Figure 4 - Segmented Bar Chart

Pie Charts

This is another type that I believe you've seen before. To construct one, extend your frequency table so that you can measure relative frequency—the percent of the total. Convert those percents to angles, and then measure out the pie slices accordingly. When you're doing this by hand, just make the angle measures close—there's no need to go and use a protractor. If you need something that exact, use software.

Actually, you should probably try to avoid using pie charts in the first place. They are almost always a poor choice when trying to display data. If you do use one, make sure to label each pie slice in some manner.

...and never ever, under any circumstances, construct a 3D pie chart!

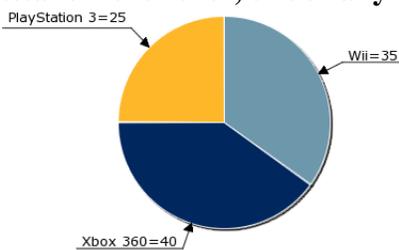


Figure 5 - Pie Chart

Pictograms

This type uses pictures of objects to represent the frequencies—sometimes it is the size of the objects, sometimes it is the number of objects shown. They are most often seen in publications like USA Today, and are rarely used in statistics.

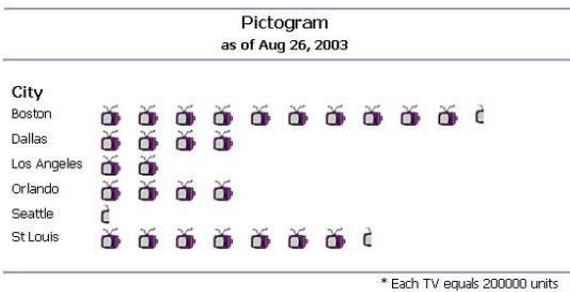


Figure 6 - Pictogram

Examples

[2.] A group of cars were classified according to type. The results are shown below:

Table 3 - Vehicle Type Data for Example 2

Compact	Large	Midsize	Small	Sporty	Van
16	11	22	21	14	9

Construct a Pareto Chart of these data.

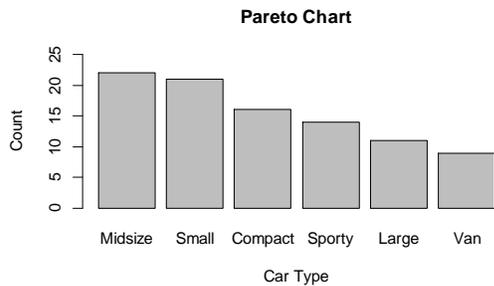


Figure 7 - Pareto Chart for Example 2

[3.] As it turns out, there were more measurements on those cars...in particular, each car was also classified by its drive type (front wheel drive, rear wheel drive, all wheel drive). Here are the updated data:

Table 4 - Cross-Tabular Data for Example 3

Drive Type	Compact	Large	Midsize	Small	Sporty	Van
Front	13	7	17	19	7	4
Rear	2	4	5	0	5	0
All-Wheel	1	0	0	2	2	5

Construct a segmented bar chart of Drive Type by Vehicle Type.

Since you are told to make a bar chart “of Drive Type,” the drive type variable should make up the segments and vehicle type should be the variable for the bars—meaning that there should be six bars, each with up to three segments.

The first step is to convert each column into percentages. For example, the total in the “compact” column is 16, so divide each of the entries in that column by 16. Similarly, since the “large” column total is 11, divide each of those values by 11. Here’s the result of that:

Table 5 - Relative Frequencies of Drive Type by Vehicle Type for Example 3

Drive Type	Compact	Large	Midsize	Small	Sporty	Van
Front	0.8125	0.6364	0.7727	0.9048	0.5000	0.4444
Rear	0.1250	0.3636	0.2273	0.0000	0.3571	0.0000
All-Wheel	0.0625	0.0000	0.0000	0.0952	0.1429	0.5556

Now, the first bar will be for the “compact” vehicle type, and the first segment of that bar will have height 0.8125. The second segment will have height 0.125, and the third segment will have height 0.0625—for a total bar height of 1.0.

Continue this process for all of the other bars to get the following:

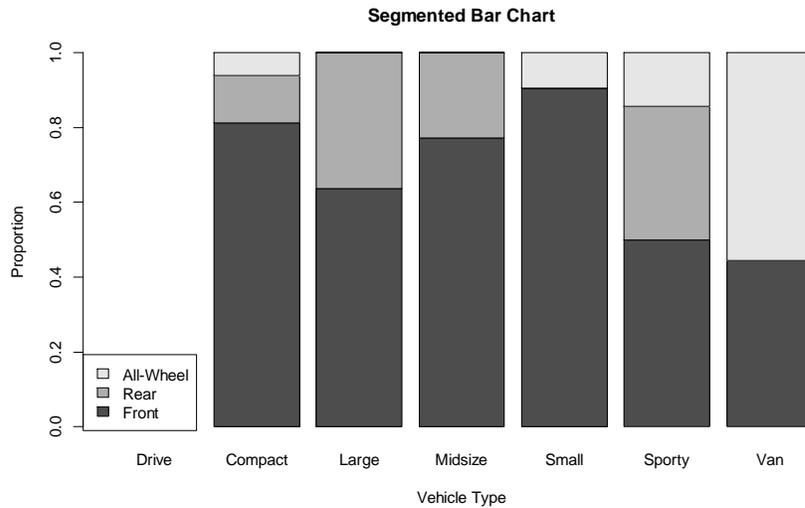


Figure 8 - Segmented Bar Chart for Example 3

[4.] A teacher keeps track of the number and types of errors that her students make on their research papers. The results are shown below:

Table 6 - Error Data for Example 4

Error	Count
Punctuation	22
Grammar	15
Spelling	10
Typing/Formatting	3

Construct a pie chart of these data.

We need to convert those counts (*frequencies*) into relative frequencies...the total number of errors is 50, so divide each count by 50.

Table 7 - Example 4 Data with Proportions

Error	Count	Proportion
Punctuation	22	0.44
Grammar	15	0.30
Spelling	10	0.20
Typing/Formatting	3	0.06

Now multiply each of those relative frequencies by 360° to get the angles...then we can make the pie chart.

Table 8 - Example 4 Data with Angles

Error	Count	Proportion	Angle
Punctuation	22	0.44	158.4°
Grammar	15	0.30	108°
Spelling	10	0.20	72°
Typing/Formatting	3	0.06	21.6°

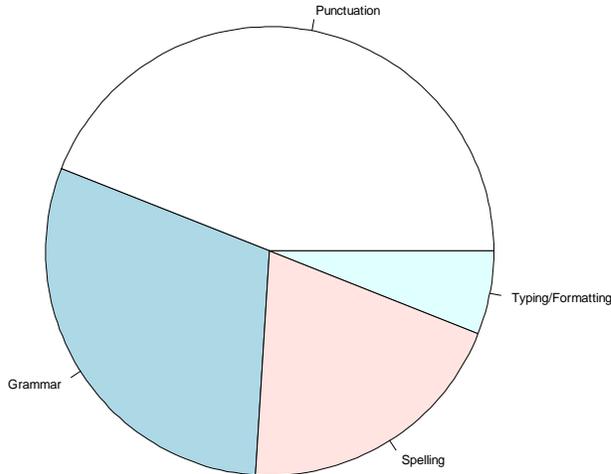


Figure 9 - Pie Chart for Example 4

Describing Qualitative Variables

The best thing that you can do to describe qualitative variables is to tell what occurred most (or least) frequently—or, at least change all of the frequencies to relative frequencies (percents) so that comparisons are more easily made. In the case of cross-tabular data (like example 3), creating percents within each column (or row) can be helpful (these are called *conditional distributions*—more on those in another chapter).

Example

[5.] 48 patients in a diet study were randomly assigned to eat pre-meal crackers with different types of added fiber (bran fiber, gum fiber, a combination of both fibers, or crackers with no additional fiber added). As part of the experiment, the patients reported on the amount of bloating they experienced after eating the crackers. The results are shown below.

Table 9 – Cross-Tabular Data for Example 5

Bloat	Bran	Gum	Combo	Control
High	0	5	2	0
Medium	1	3	3	2
Low	4	2	5	4
None	7	2	2	6

What can you say about the relationship (if any) between fiber type and amount of bloating?

Since there are equal number of patients in each fiber category, there is no need to convert to percentages—making fair comparisons is possible right away. Speaking of right away, let's make a graph:

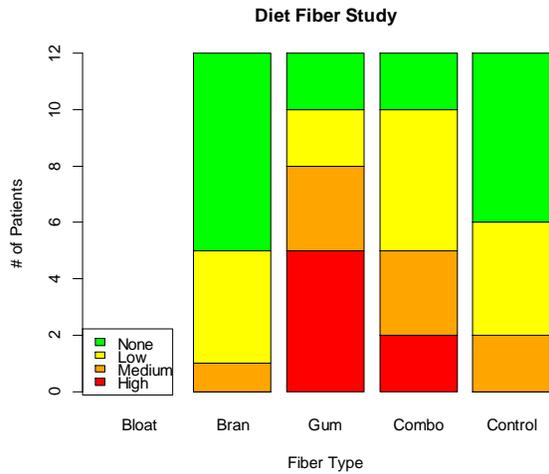


Figure 10 - Segmented Bar Chart for Example 5

What I notice is that gum fiber seems to be the worst for bloating—it has the most people reporting high bloating and the fewest reporting no bloating. I also notice that bran fiber has no people reporting high bloating, and the highest number of people reporting no bloating—in fact, in terms of bloating, the bran fiber appears even better than the control crackers! The combo fiber crackers seem to be in the middle—which makes sense if the gum fiber is the worst and the bran fiber is the best.

Compared to the control, adding bran fiber appears to decrease bloated feelings, and adding gum fiber appears to greatly increase bloated feelings.