

We will explore tables, bar graphs, and pie charts.

General Education Statistics

Class Notes

Organizing Qualitative Data (Section 2.1)

If we have raw data but cannot display it accurately and succinctly, it does us no good. We will explore these common ways for displaying qualitative data. Quantitative data will be dealt with later.

Organizing Data in Tables:

Consider this data obtained (deliciously) from a bag of M&Ms. We are interested in the distribution of colors of the M&Ms in the bag.

EXAMPLE Organizing Qualitative Data into a Frequency Distribution

brown, brown, yellow, red, red, red, brown, orange, blue, green, blue, brown, yellow, yellow, brown, red, red, brown, brown, brown, green, blue, green, orange, orange, yellow, yellow, yellow, red, brown, red, brown, orange, green, red, brown, yellow, orange, red, green, yellow, yellow, brown, yellow, orange

expl 1: First, we will make a Tally chart. This is the most efficient way to count repeated data. Go through the list one-by-one and make a tally next to each color as it occurs. Then count the tally marks at the end and place those numbers in the right column.

Color	Tally	Count (Frequency)
brown		
yellow		
red		
orange		
blue		
green		
		Total =

Good to check total against the raw data.

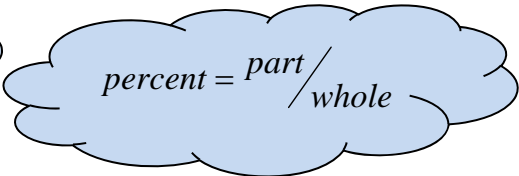
This is called a frequency distribution. Remember that the data is still qualitative, even though we now have counted each category.

Definition: A **frequency distribution** lists each category of data and the number of occurrences for each category.

Well, now that's well and good. But what if we wanted to show what proportion or percentage of the bag was red or blue? We could construct a **relative frequency distribution**.

Definition: The **relative frequency** is the proportion (or percent) of observations within a category and is found using the following formula.

$$\text{relative frequency} = \frac{\text{frequency}}{\text{sum of all frequencies}}$$

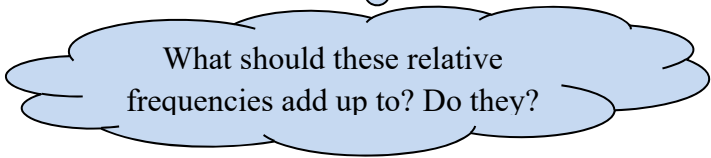


$$\text{percent} = \frac{\text{part}}{\text{whole}}$$

Definition: A **relative frequency distribution** lists each category of data alongside their relative frequencies.

expl 2: Let's create a relative frequency table. Divide each category's frequency by the total number of M&Ms to get each category's relative frequency. Then record those numbers in the new table below. Round to three decimal places.

Color	Frequency	Relative Frequency	Percentage
brown	12		
yellow	10		
red	9		
orange	6		
blue	3		
green	5		
	Total = 45		



What should these relative frequencies add up to? Do they?

Recall: Percentages:

Do you remember how to find percents? How do you convert these numbers to percent form? When you do, remember to include the percent symbol (%).

Organizing Data in Bar Graphs:

Definition: A **bar graph** is constructed by labeling each category of data on either the horizontal or vertical axis, with the frequency (or relative frequency) of the category on the other axis. The height of each rectangle represents the category's frequency (or relative frequency).

Let's return to the M&Ms example.

expl 3: Use the counts for each color of M&M to draw a bar graph for this bag of M&Ms.

Color	Frequency
brown	12
yellow	10
red	9
orange	6
blue	3
green	5
	Total = 45

Label the horizontal axis "Color" and space the options along it. Label the vertical axis "Count" and draw equally spaced tick marks.

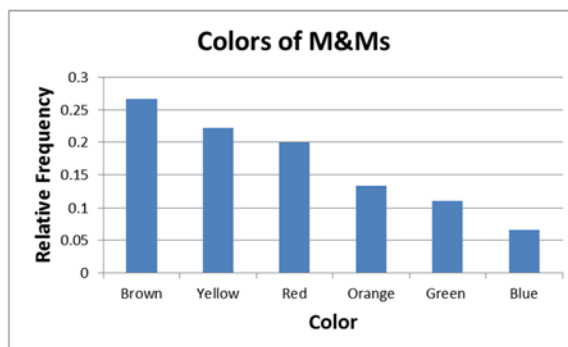
Rectangles of equal width are drawn for each category.

Provide a title too.

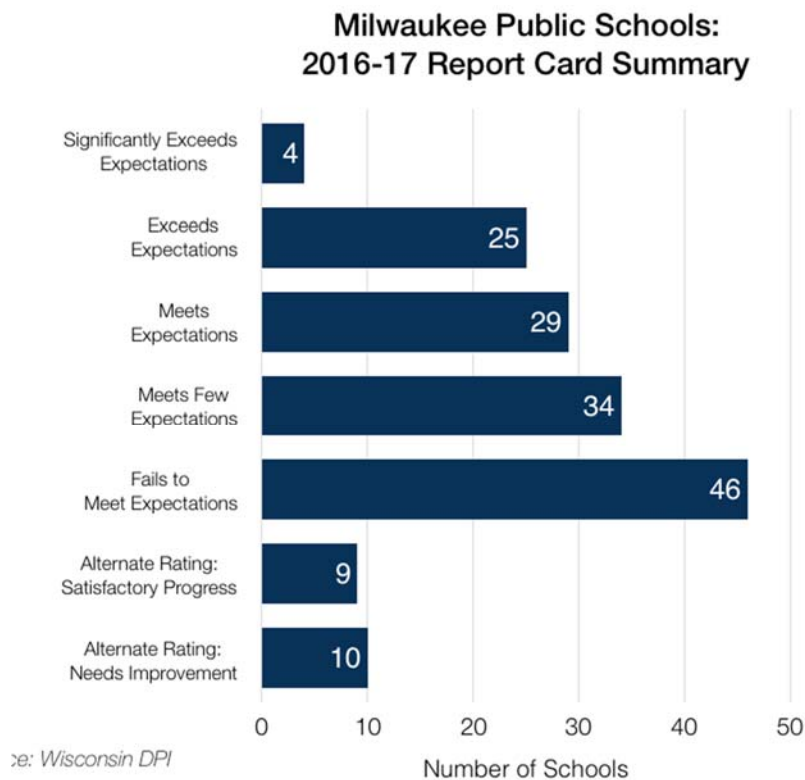
Rectangles in bar graphs should *not* touch each other.

A bar graph can show frequencies like ours does above. You may also choose to use relative frequencies (percentages). In either case, the bars will be the same relative to each other. Comparing the areas of the bars is an easy way to compare categories.

Definition: A **Pareto chart** is a bar graph where the bars are drawn in decreasing order of frequency or relative frequency.



expl 4: Consider the bar graph below to answer the following questions.



er @MacIverWisc

(Source: <http://www.goethes-farbenlehre.com/bar-graph-vs-column-graph.html>)

a.) What does this graph display?

b.) How many schools “meet expectations”?

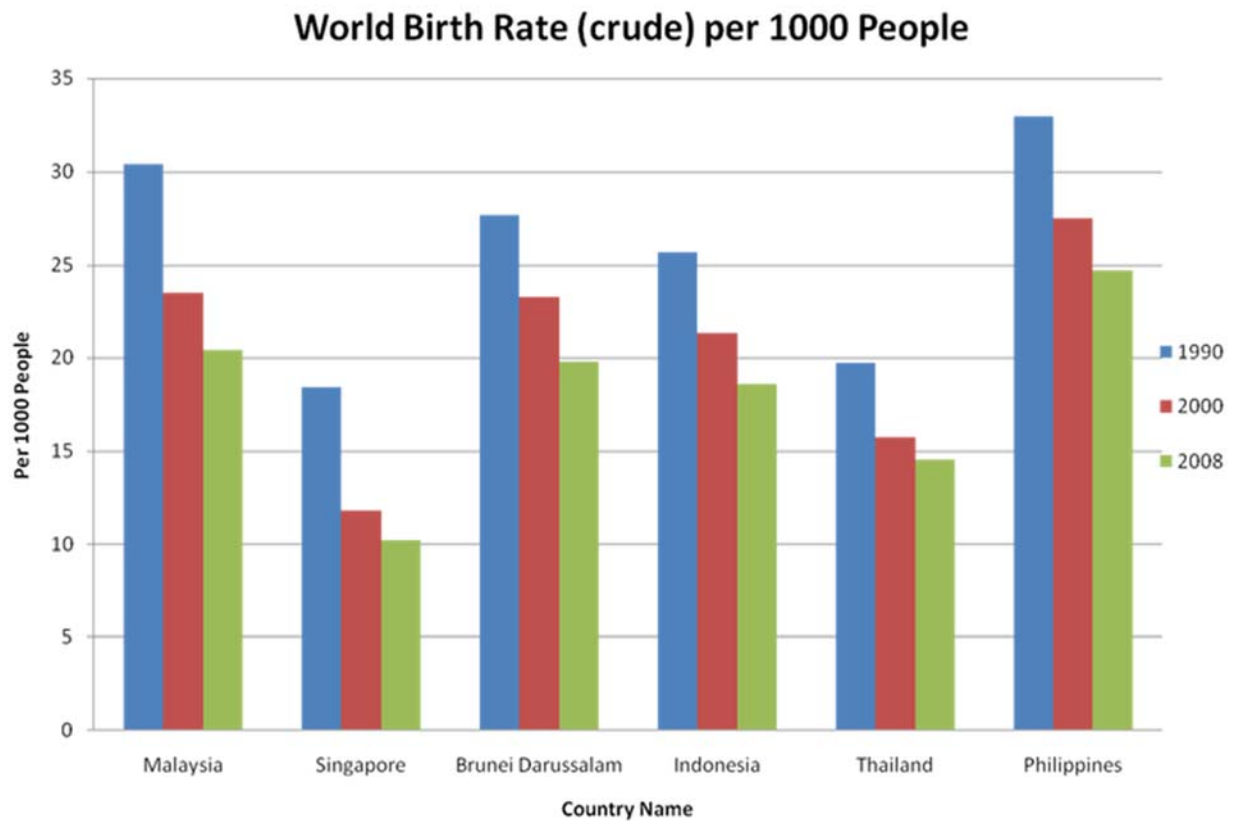
c.) How many schools score at “meets expectations” or above?

d.) Of the 157 schools represented, what percentage “fails to meet expectations”? Round to the nearest tenth of a percent.

Side-by-Side Bar Graphs:

Consider the information this graph provides. Here we are able to compare each country with one another, as well as how each country changed over time.

These data sets should be compared by using relative frequencies, because different sample or population sizes make comparisons using frequencies difficult or misleading.



(Source: <https://cloudfour808.wordpress.com/2011/03/29/bar-graph-aisyah-zahidah-107168/>)

expl 5: Use the above graph to answer these questions.

- What country and year do we see the highest birth rate given on the graph? What is that birth rate?
- Describe the trend over time for the country of Malaysia.
- Who has the higher birth rate in 2008, Indonesia or Malaysia?

expl 6: (refers to graph on previous page) Let's delve deeper into the question asked in example 5c. The populations and total counts of births for Indonesia and Malaysia in 2008 are given below in the table. Fill in the birth rates, estimated from the graph, in the table below. Explain why we use birth rates (per 1000 people) and *not* the number of births to compare these countries.

Country	2008 Birth Rates (per 1000 people, read from graph)	Approximate number of births	Population (2008)
Indonesia		4,400,512	234,693,997
Malaysia		506,354	24,821,286

Organizing Data in Pie Charts:

Definition: A **pie chart** is a circle divided into sectors. Each sector represents a category of data. The area of each sector is proportional to the frequency of the category.

expl 7: Consider the frequency table for the M&Ms example. Let's draw a pie chart for this data using StatCrunch.

Color	Frequency
brown	12
yellow	10
red	9
orange	6
blue	3
green	5
	Total = 45

Instructions for how to make pie charts in StatCrunch are given on the next page.

Instructions: Select **StatCrunch** from the menu in **MyStatsLab**. From there, select **Visit StatCrunch website**. Once there, select **Open StatCrunch** from the tabs at the top.

Label the column marked “var1” by clicking on “var1” and renaming it “Color”. Do the same for “var2”, renaming it “Count”. Enter the data into the columns.

Select **Graph > Pie Chart > With Summary**. You will need to tell it which columns have the categories and counts. Also, title your graph under **Graph Properties > Title**. There are other options but you can leave them on default. Then click “Compute”. It produces the pie chart, with counts and percentages in place.

The textbook has good instructions on this and other displays as well as different software.

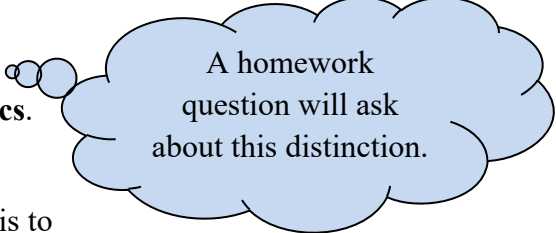
Recall: Inferential versus Descriptive Statistics:

Recall: Definitions: Descriptive Statistics versus Inferential Statistics:

A **statistic** is a numerical summary of a *sample*. **Descriptive statistics** consist of organizing and summarizing data. It uses numerical summaries, tables, and graphs.

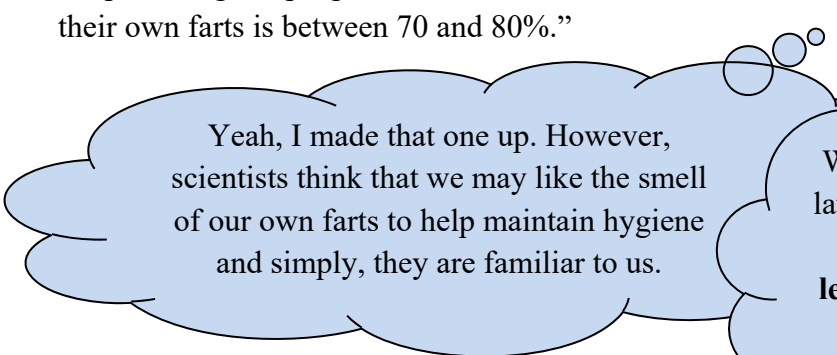
Inferential statistics uses methods that take the result from a sample and extend it to the population. It also provides a measure of the reliability of the result. A **parameter** is a numerical summary of the *population*. It is this information that statistics can help us estimate by looking at the sample.

If you are told about the result of a sample in word or graph form, consider this to be **descriptive statistics**.

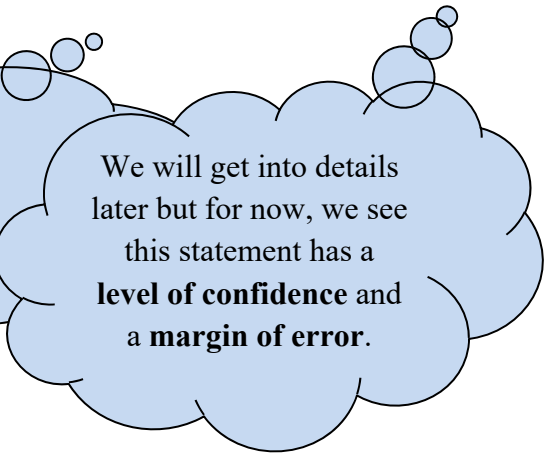


A homework question will ask about this distinction.

If you are told information about a whole population (i.e. for all adults, all cars, all states, etc.), consider this to be **inferential statistics**. The statisticians took a sample and then *inferred* from it something about the larger population. They are said to be making a **prediction** using the limited information from a sample. Often, such things will be worded like, “We are 95% confident that the percentage of people who like the smell of their own farts is between 70 and 80%.”



Yeah, I made that one up. However, scientists think that we may like the smell of our own farts to help maintain hygiene and simply, they are familiar to us.



We will get into details later but for now, we see this statement has a **level of confidence** and a **margin of error**.