LESSON 5: GRAPHS and CHARTS
How Do We Visualize Data?

The terms "graph" and "chart" are used interchangeably in statistics. (See Footnote 1.)

PART A: FREQUENCY HISTOGRAMS

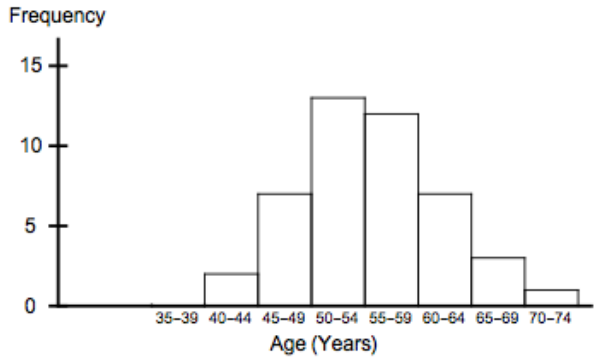
We can use a frequency histogram to graphically represent **quantitative** data.

Example 1 (Frequency Histogram: Presidents' Ages)

Here is the **frequency table** for presidents' ages from Lesson 4, Example 1.

Age class (years)	Frequency
35-39	0
40-44	2
45-49	7
50-54	13
55-59	12
60-64	7
65-69	3
70-74	1
	Sum = N = 45

The **frequency histogram** below is a graphical representation of the **frequency distribution** of presidents' ages. The **height** of each bar gives the **frequency** of the corresponding class. (Bars for consecutive, non-empty classes should touch.)



• The histogram is roughly symmetric (the left half and the right half are roughly mirror images), and it resembles a bell curve  .
This is common for data on natural measures such as ages and heights.

• The histogram allows us to **estimate frequencies**. For example, between 10 and 15 presidents were 50 to 54 years old on becoming president.

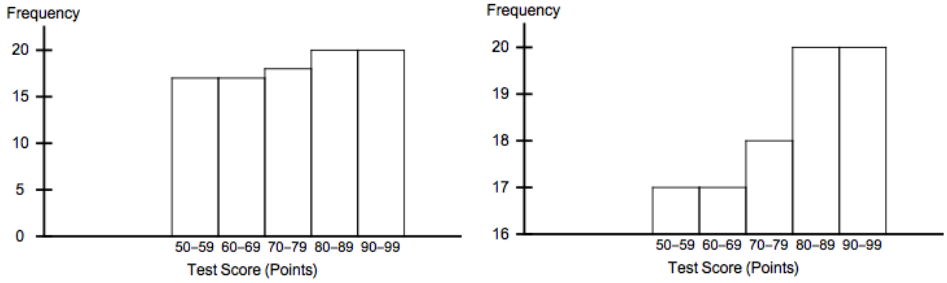
Notes on Scaling:

• Highest tick mark. The highest tick mark (15) on the Frequency axis must be **at least as high as** the highest frequency (13) among the classes. (The class with this highest frequency is called the modal class; here, it is the age class from 50 to 54 years.) We don't want our bars to extend past the top of our figure!

• Consistent spacing along each axis. The **tick marks** and the **numerical values** on the Frequency axis are both **evenly (equally) spaced**. Also, the bars must have the **same width**; remember that the age classes must also have the same width.

• The issues of different starting points and stretching (or squeezing) the Frequency axis. The Frequency axis should **start at 0**, although this is sometimes impractical.

Different starting points tend to encourage creators to **stretch or squeeze** the Frequency axis. We then obtain **different (inconsistent) spacings** between, say, 17 and 19 along the Frequency axis in the two histograms below.



All these differences will lead to **different pictures**. Differences among the bar heights (and the heights themselves) could be **emphasized** or **de-emphasized**, depending on the creator's intent. This can lead to **statistical abuse**!

Compare the two histograms above. They actually correspond to the **same** frequency distribution. The Frequency axis in the second histogram **starts at 16** and is more **stretched** out.

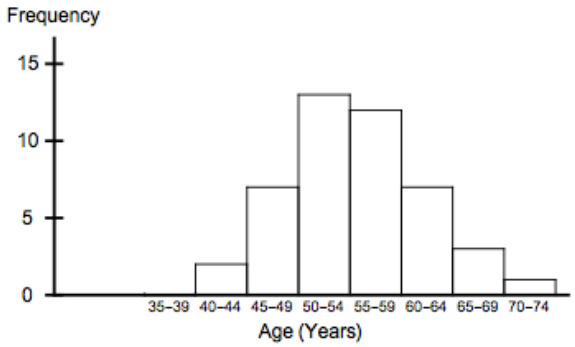
(The **horizontal axis** tends to be less controversial. A different starting point does lead to a horizontal shift, but creators may be less likely to stretch or squeeze than in the case of the Frequency axis.)

§

PART B: RELATIVE FREQUENCY HISTOGRAMS

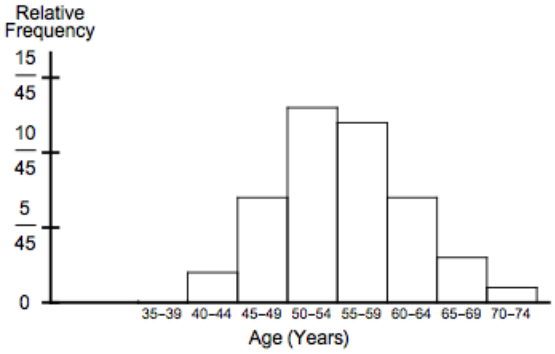
Example 2 (Relative Frequency Histogram: Presidents' Ages)

In Example 1, we constructed a **frequency histogram** for presidents' ages.



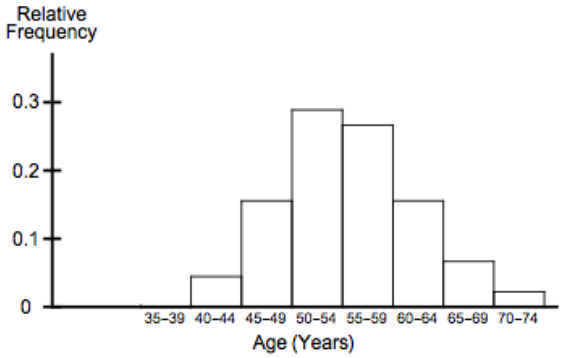
To construct a relative frequency histogram, we can **stretch** the vertical axis by a factor of N (45 here) and keep the same picture.

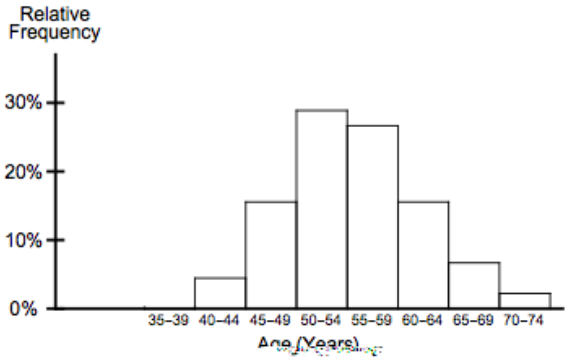
We could simply divide the numbers on the vertical tick marks by N (45 here) and put "Relative Frequency" on the vertical axis.



Note: For **sample data**, you would divide by n, the sample size.

We may prefer to use "nice" decimals or percents along the vertical axis.





• The **relative frequency table** from Lesson 4, Example 2 can help us.

Age class (years)	**Frequency**	**Relative Frequency**	**Relative Frequency (as a percent)**
35-39	0	0	0%
40-44	2	0.044	4.4%
45-49	7	0.156	15.6%
50-54	13	0.289	28.9%
55-59	12	0.267	26.7%
60-64	7	0.156	15.6%
65-69	3	0.067	6.7%
70-74	1	0.022	2.2%
	Sum = N = 45	Sum = 1	Sum = 100%

• This histogram allows us to **estimate relative frequencies**. For example, a little under 30% of the presidents were 50 to 54 years old on becoming president.

• The relative frequency histogram graphically represents the relative frequency distribution of the data values. Here, we have a roughly bell-shaped distribution.

§

Think About It

If a standard six-sided die is rolled 100 times, would you expect a relative frequency histogram for the die results (numbered 1-6) to have a similar shape?

PART C: DOTPLOTS

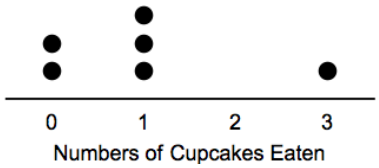
Dotplots are like histograms with "thin" classes, except that stacked dots replace the bars.

Example 3 (Dotplot)

An office has six employees. Cupcakes are brought to the office. The numbers of cupcakes eaten by the employees are:

1, 0, 3, 1, 0, 1

Here is a dotplot:



We can quickly see that two employees ate 0 cupcakes, three ate 1 cupcake, and one ate 3 cupcakes.

§

PART D: STEMPLOTS, OR STEM-AND-LEAF PLOTS

Stemplots, or stem-and-leaf plots, resemble sideways histograms, but you can use them to:

- recover the original data values
- sort the data values (place them in numerical order)

Example 4 (Stemplot)

Nine students take a test. Their scores (in points) are as follows:

77 93 73 51 74 85 82 73 100

Stage 1: Assign leaves to stems.

- The leaf of a quantitative data value is the value's last digit.

  (If the data values are not integers, they must be written out to the same number of decimal places.)

- The stem consists of the other digits.

Write the stems in increasing order.

		Comments
5	1	← Corresponds to 51
6		← No scores in the 60s
7	7343	← Correspond to 77, 73, 74, and 73
8	52	
9	3	
10	0	

- Do **not** skip the "6"; include it as a stem. Otherwise, the shape of the distribution will be distorted.

- **Include repetitions**. There were two "73"s.

Stage 2: Sort the leaves for each stem (in increasing order).

5	1
6	
7	3347
8	25
9	3
10	0

Stages 1 and 2 can be done simultaneously.

We can quickly read off the scores (in points) in **increasing** order:

51 73 73 74 77 82 85 93 100

§

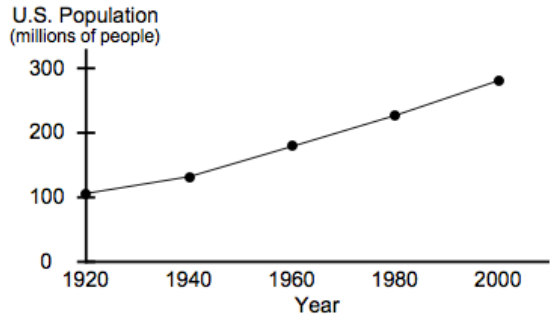
PART E: TIME-SERIES GRAPHS

Time-series graphs are used for data that changes over time. Examples include population, stock prices, CO_2 concentration in the atmosphere, etc.

Example 4 (Time-Series Graph: U.S. Population)

Year	U.S. Population (in millions)
1920	106
1940	132
1960	179
1980	227
2000	281

The table above can be represented graphically by this time-series graph:



§

PART F: BAR GRAPHS

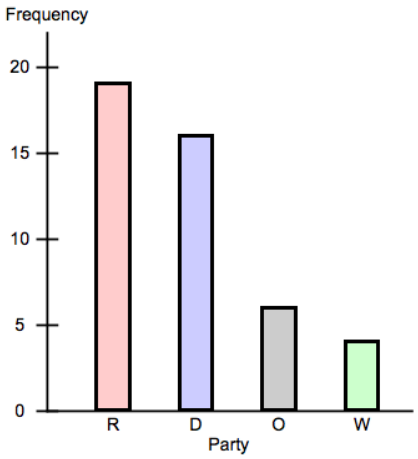
Bar graphs are like histograms for **qualitative data** consisting of categories, such as political parties in Example 5 below. Bars for qualitative data should not touch.

In a Pareto chart, categories are sorted by descending order of their frequencies.

Example 5 (Bar Graph / Pareto Chart: Presidents' Parties)

We can use the **frequency table** from Lesson 4, Example 3 to construct a bar graph (Pareto chart) for the presidents' ages.

Party	Frequency
Democratic (**D**)	16
Republican (**R**)	19
Whig (**W**)	4
Other (**O**)	6



§

PART G: PIE CHARTS

Pie charts are circular disks that can represent both quantitative and qualitative data. Possible values correspond to wedges of the pie.

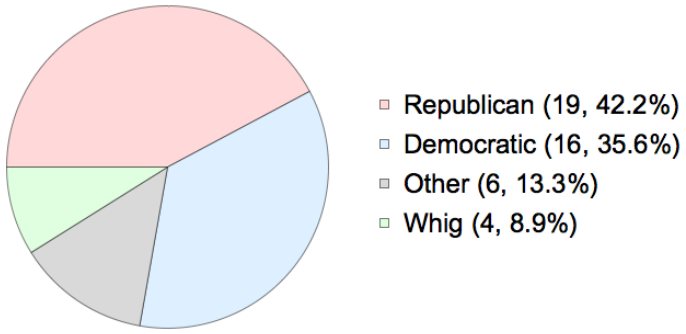
The wedges are circular sectors whose sizes (and central angles) are **proportional** to their frequencies and relative frequencies.

Example 6 (Pie Chart: Presidents' Parties)

We can use the **frequency and relative frequency tables** from Lesson 4, Example 3 to construct a pie chart for the presidents' parties.

Party	**Frequency**	**Relative Frequency**	**Relative Frequency (as a percent)**
Democratic (**D**)	16	$\frac{16}{45} \approx 0.356$	35.6%
Republican (**R**)	19	$\frac{19}{45} \approx 0.422$	42.2%
Whig (**W**)	4	$\frac{4}{45} \approx 0.089$	8.9%
Other (**O**)	6	$\frac{6}{45} \approx 0.133$	13.3%
	Sum = N = 45	Sum = 1	Sum = 100%

Here is a pie chart with an explanatory legend:



§

FOOTNOTES (OPTIONAL)

#1) Graphs and charts. Some sources see these two terms as interchangeable and synonymous. Charts may do more to summarize data in presentations. One refers to "pie charts" much more so than "pie graphs." See:
https://www.reference.com/math/difference-between-chart-graph-8775a512630cf520

#2) Pie charts – one advantage. Although pie charts can vary in size and color, the central angles must be the same if the same data is being used. In this sense, pie charts are not as vulnerable to abuse as histograms or bar graphs (with respect to the rescaling issue mentioned in Part A).

#3) Moiré effect. Try not to stripe bars or wedges. Stripes tend to lead to illusions of vibration, referred to as the Moiré effect.