

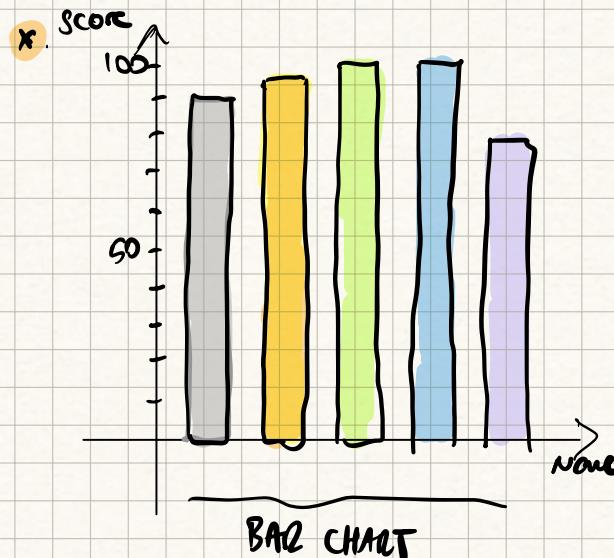
Displaying Quantitative Data with Graphs

16.06.2025

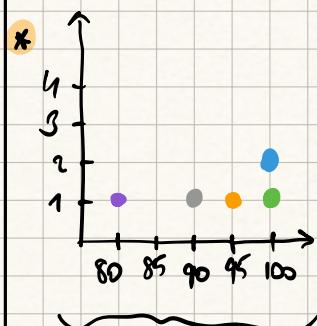
Representing Data

* Name	Column
Amy	90
Bill	95
Cam	100
Efra	100
Farah	80

TABLE



BAR CHART



FREQUENCY DOT PLOT

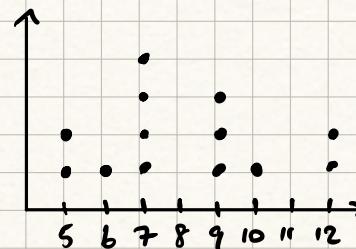
* [90, 95, 100, 100, 80] } list

* Many ways to represent data! Representing data in these ways allows us to interpret them easily. For ex, imagine we had thousands of data points represented in a frequency table. Answering the question, "What's the range of our data?" would be very easy.

Frequency tables and dot plots

* Ages of Students in class = [5, 7, 5, 9, 7, 7, 6, 9, 9, 9, 10, 12, 12, 7]

Age	Frequency
5	2
6	1
7	4
8	0
9	3
10	1
11	0
12	2



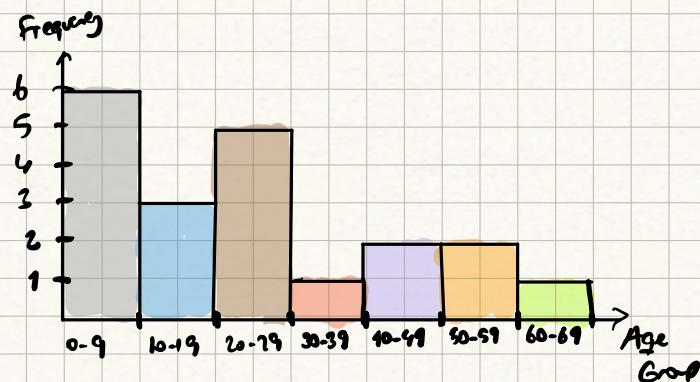
* Data can be represented in many ways, and each method highlights different aspects, making patterns and trends easier to understand.

* A frequency table organizes data by listing each category or value along with its corresponding count (frequency). A frequency dot plot visually represents this data using dots stacked above a number line, where each dot represents one occurrence of a value. Both tools help summarize and display the distribution of a dataset.

Creating a histogram

* Ages = [1, 3, 27, 32, 5, 63, 26, 25, 18, 16, 4, 45, 29, 19, 22, 51, 58, 9, 42, 6]

Bin	Frequency
0-9	6
10-19	3
20-29	5
30-39	1
40-49	2
50-59	2
60-69	1



Interpreting a histogram

- * By creating a histogram, we've realized that the highest frequencies are (0-9) and then (20-29). Maybe this is a restaurant that gives away toys with food? A lot of children (0-9) and their young parents (20-29)?
- * By looking at the histogram we can also tell how many people there are in the restaurant ($6+3+5+1+2+2+1 = 20$ people)
- * A histogram displays numerical data by grouping data into "bins" of equal width. Each bin is plotted as a bar whose weight corresponds to how many data points are in that bin.

* A histogram is a graphical representation of data that uses bars to display the frequency of values within specified intervals (bins). Unlike bar graphs, the bars in a histogram touch each other, emphasizing the continuous nature of quantitative data. It helps visualize the distribution, shape, and spread of a data set.

Stem-and-leaf plots



* A statistician for a basketball team tracked the number of points that each of the 12 players on the team had in one game, and then made a stem-and-leaf plot to show the data. How many points did the team score?

Stem	Leaf
0	0 0 2 4 7 7 9
1	1 1 3 8
2	0

means that players scored:

$$= [0, 0, 2, 4, 7, 7, 9, 11, 11, 13, 18, 20]$$

sum = 102 points

* A stem-and-leaf plot is a method of organizing numerical data by splitting each value into a stem (leading digit(s)) and a leaf (trailing digit). The stems are listed vertically, and the leaves extend horizontally, showing the distribution while preserving the original data points. It's useful for displaying small datasets while maintaining exact values.

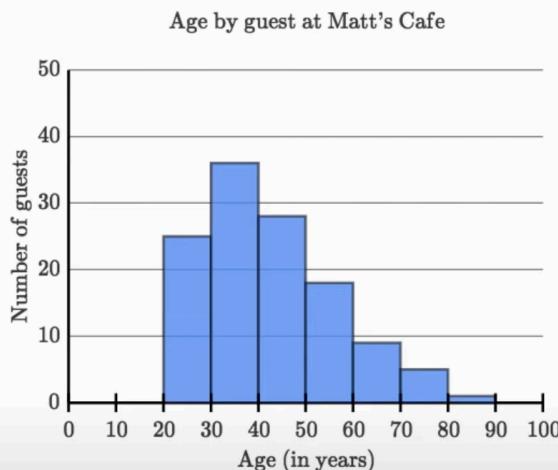


* A stem-and-leaf plot is a method of organizing numerical data by splitting each value into a stem (leading digit(s)) and a leaf (the trailing digit). The stems are listed vertically, and the leaves extend horizontally, showing the distribution while preserving the original data points. It's useful for displaying small datasets while maintaining exact values.

Describing and Comparing Distributions

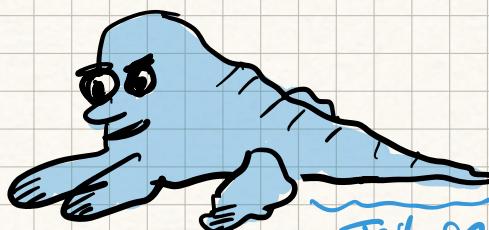
16.06.2025

Shapes of distributions



What's the shape of the distribution?

- (A) Left-tailed
- (B) Approximately symmetrical
- (C) Right-tailed



Tail on the right!

- * Left-tailed = Skewed to the left = Negatively skewed
Mean < Median < mode
- * Right-tailed = Skewed to the right = Positively skewed
Mean > Median > Mode
- * Perfectly symmetrical \Rightarrow Mean = Median = Mode

Clusters, gaps, peaks, outliers

- * Clusters : Groups of data points concentrated closely together in a distribution.
- * Gaps : Unusually empty ranges where no data points appear between values.
- * Peaks : The highest frequencies in distribution, representing modes or local maxima.
- * Outliers : Extreme values that fall far outside the overall pattern of the data.

* Left-tailed = Skewed-to-the-left \Rightarrow Mean < median < mode

Right-tailed = Skewed-to-the-right \Rightarrow Mean > median > mode

Symmetrical \Rightarrow Mean = median = mode

* Clusters : Groups of data points concentrated closely together in a distribution.

* Gaps : Unusually empty ranges where no data points appear between values.

* Peaks : The highest frequencies in distribution, representing modes or local maxima.

* Outliers : Extreme values that fall far outside the overall pattern of the data.

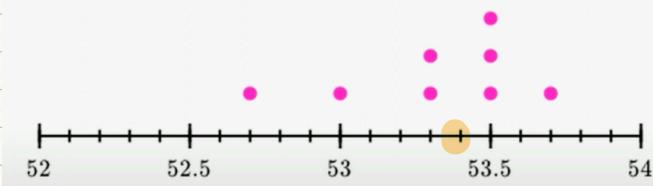
Comparing distributions with dot plots



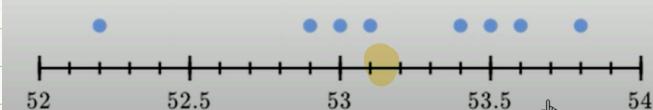
At the Olympic games, many events have several rounds of competition. One of these is the men's 100 meter backstroke. The upper dot plot show the times (in seconds) of the top 8 finishers in the semifinal round at the 2012 Olympics. The lower dot plot shows the times of the same 8 swimmers, but in the final round.

Which pieces of information can be gathered from these dot plots?

Semifinal round:



Final round:



Answer



Select all that apply.

- The swimmers had faster times on average in the finals.
- One of the swimmers was disqualified from the finals.
- The times in the finals vary noticeably more than the times in the semifinals.
- Individually, the swimmers all swam faster in the finals than they did in the semifinals.

[Check Answer](#)

[Show me how](#)

Comparing dot plots, histograms, and box-plots



A statistician recorded the length of each of Pixar's first 14 films. The statistician made a dot plot (each dot is a film), a histogram, and a box plot to display the running time data. [\[Show me the displays\]](#)

Which display could be used to find the median?

Select all that apply.

- The dot plot → b/c it gives all the data points
- The histogram → b/c it doesn't give all the data points, just the bins
- The box plot → b/c it tells the median explicitly



Nam owns a used car lot. He checked the odometers of the cars and recorded how far they had driven. He then created both a histogram and a box plot to display this same data (both diagrams are shown below).

Which display can be used to find how many vehicles had driven more than 200,000 km (kilometers)?

- The histogram (as long as 200,000 km is where a bin starts)
- The box plot

Which display can be used to find that the median distance was approximately 140,000 km?

- The histogram
- The box plot → b/c it tells the median explicitly

! You can skip the quizzes of this section and comeback after unit 3 because many concepts asked in these quizzes won't be covered until Unit 3.

Reading Line Graphs



An investment firm creates a graph showing the performance of a specific stock over 12 months. Over the course of the year, is the price of the stock rising, falling, or staying the same?



- * Line graphs display data points connected by straight lines, showing trends or changes over time. It's ideal for visualizing continuous data and comparing multiple datasets (using different colored lines).

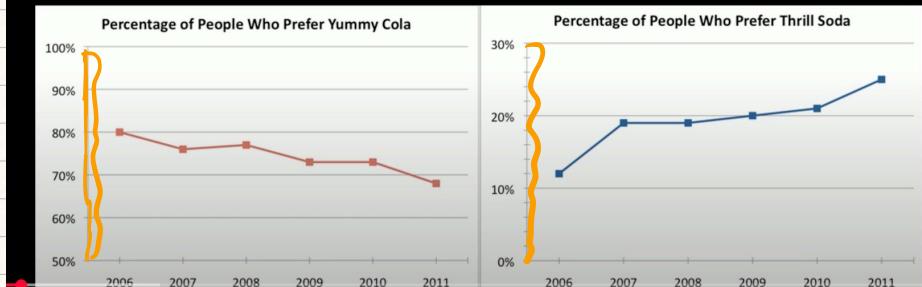
Misleading line graphs



Thrill Soda hired a marketing company to help them promote their brand against Yummy Cola. The company gathered the following data about consumers' preference of soda:

Year	% of respondents who prefer Yummy Cola	% of respondents who prefer Thrill Soda	% of respondents who have no preference
2006	80%	12%	8%
2007	76%	19%	5%
2008	77%	19%	4%
2009	73%	20%	7%
2010	73%	21%	6%
2011	68%	25%	7%

The advertising company created the following two graphs to promote Thrill Soda:



Data is accurate but the y-axes are scaled in a way that as if the percentage of people who prefer thrill soda is higher!

- * Line graphs display data points connected by straight lines, showing trends or changes over time. It's ideal for visualizing continuous data, and comparing multiple datasets (using different colored lines).
- * Line graphs can be misleading if...
 - ① The y-axis is truncated (doesn't start at 0), making small changes look dramatic
 - ② The x- or y-axis is stretched/compressed, distorting the rate of change
 - ③ Uneven intervals are used, creating false patterns.