



M2-04: Histograms

Part of the "Exploratory Data Analysis" Learning Badge

Video Walkthrough: <https://discovery.cs.illinois.edu/m2-04/>

Bar Graphs vs. Histograms

A _____ is a familiar way of visualizing qualitative (categorical) variables/distributions. It displays a bar for each category. The bars are equally spaced and equally wide. You can create bar graphs with horizontal bars or vertical bars.

Bar graphs use height to depict frequencies or counts. The widths of the blocks in bar graphs don't mean anything. They are all the same regardless of the size of the category.

Puzzle #1: Draw a bar graph using the following data from the GPA dataset. First, find the number of sections in each category, then label your x and y-axis, and lastly, draw the bar graph.

Category:	STAT	CS	IS	MATH
# of Sections:				

A _____ is a common way of visualizing quantitative variables/distributions. Unlike bar graphs, the x-axis of a histogram is always drawn to scale. Histograms can help us visualize the shape of a distribution.

There are two types of histograms that are commonly used:

- 1) _____ **Histogram:** Frequency histograms show how often different values in a dataset occur. The width of each block indicates the size of the interval and the height of each bar indicates the number of observations in each interval (or bin). Frequency histograms are very *simple* to interpret and read.
- 2) _____ **Histogram:** Density histograms use areas to depict percentages. The width of each block indicates the size of the interval and the area of each block indicates what percentage of the data belongs to that category. The height indicates density or how crowded the block is. Density histograms are used a lot in *statistics*.

IMPORTANT NOTE: In a density histogram, the total area of all the blocks is 1 or 100%.