# Glossary

**Action Research Model** – a model of teaching that focuses on collecting and analyzing data in a real life context. Practitioners conduct research on the effectiveness of various activities and lesson plans and then use that research to make informed decisions about their next course of action.

**Average** – a term for the median, mean, or mode of a data set. Used to describe the most common value. Synonym for *central tendency*.

**Bar chart/bar graph** – a visualization that displays values through the use of narrow rectangles (bars).

**Benchmark statistics** – a reliable statistic used to determine the validity or importance of another statistic in a related field.

**Bins** – consecutive numeric ranges or intervals used in histograms, such as 20-30, 0-100, or 0-1.

**Box plot** – a visual icon placed on a graph that displays a wide range of numerical values, including the mean, median, first and third quartiles, and maximum and minimum values.

**Bubble charts** – visualizations that represent the relationship between three separate, numeric variables, one on the y-axis, another on the x-axis, and a third mapped via the the size and/or color of the bubble around the data point itself.

**Causation** – a relationship of cause and effect in which change in one variable always provokes a change in the other.

**Categorical information** – information that is sorted into groups according to the presence of named characteristics, such as age or gender.

**Central tendency** – synonym for *average*.

**Chart** – a way of organizing information in tabular form, much like a table. Also, a synonym for *graph*.

**Cherry-picking** – the practice of selecting information that confirms your argument while ignoring information that contradicts your argument.

**Choropleth map** – a map that conveys information through color shading, where a range of hues demonstrates the density of a variable. Population density is the most common variable mapped on a choropleth map.

**Close reading** – the process of reading material actively for information while considering the context in which the information is presented and the language which is used.

**Computational data** – data which results from a simulation or computer model, usually displayed in numeric format.

**Confirmation bias** – the inherent tendency to favor information that agrees with our pre-existing hypothesis or ideas.

**Correlation** – term used when one observes a relationship or connection between two or more elements. A correlation does not indicate that one variable causes another to happen. Positive correlation means that variables increase or decrease together, while negative correlation means that one variable decreases as the other variable increases. Much more common than causation.

**Data** – facts and statistics collected together for reference or analysis. Note that while data is technically a plural term, it is often used either as a singular or plural term in practice.

**Data collection** – the process of gathering qualitative or quantitative information through experiments, interviews, focus groups, test scores, surveys, field notes, observations, etc.

**Data literacy** – the ability to comprehend, evaluate, and synthesize data and numeric information in all of its different forms.

**Data point** – a single piece of data or information.

**Data projection** – the act of predicting data trends through statistical analysis.

**Datasets** – a collection of data points.

**Data visualization** – a display of information in visual form, which may take the form of a table, chart, graphic, infographic, etc., for the purpose of discovering trends, patterns, or anomalies that we would otherwise miss.

**Density** – the compactness or volume of information associated with a geographic location, variable, or element.

**Distribution** – the range of values or intervals in a dataset.

**Dot plot** – a graph that represents data as dots on an x- and y-axis. May be used as an alternative to bar graphs to avoid visual clutter.

**Evidence-based practice** – decision-making based on research.

**Filter function** – a function in Excel that allows the user to limit the data they are looking at.

**GAP** – an acronym used by Reading Apprenticeship program that stands for genre, audience, and purpose.

**Graphic visualization** – See *Data visualization*.

**Frequency** – the number of times a variable, instance, or number appears in a dataset.

**Goldilocks principle** – term for a right-sized dataset. Also, a rule of thumb that says that while some data points will be extremes, most will fall somewhere in the middle.

**Graph** – a visual diagram showing the relationship between two or more variables.

**Histogram** – a visualization that shows the distribution of a numeric set of variables, like height or income, in a bar graph-like format, with each bar representing a *bin*, or range of values.

**Infographic** – a collection of visual, numerical, and text-based bits that together create a compelling argument or synthesis.

**Key** – See *Legend*.

**Labels** – words or phrases assigned to the axes or headings of graphs to provide information on what each value represents.

**Legend** – a text box usually located in the lower right corner of a graph that provides contextual information needed to interpret the graph, such as what specific colors, lines, or shapes represent. Also known as a *key*.

**Line chart/line graph** – a visualization that uses lines to connect plotted data points.

**Logarithmic graphs** – visualizations that may use non-standardized distances between intervals to consolidate a wide range of information so it fits into a smaller graph.

**Margin of error** – the statistical allowance for differences between a sample and the actual population from which the sample is derived.

**Maximum** – the largest numeric value in a dataset.

**Mean** – the result after one adds a series of numbers and divides that sum by the total quantity of numbers. Also known as the *arithmetic mean*.

**Median** – a number that is the midpoint in a range of numbers, so that there is equal probability that a number will fall above or below this number's value.

**Metadata** – information about a dataset, such as labels, headers, scale, and other information that is required to understand a dataset.

**Metrics** – a mathematical function that takes a set of numbers and compares the difference in values. Specifically, metrics create standards by which figures and statistics can be judged.

**Minimum** – the smallest numeric value in a data set.

**Mode** – the value that occurs most often in a data set.

**Online portal** – a website that brings together information or links to information from a diverse number of sources.

**Outliers** – a value that falls outside the expected range of a dataset and is numerically exceptional in comparison to the other values.

**Parameters** – a measurable factor or number that helps define an operation or sets the conditions by which a study is conducted.

**Percentile** – the percentage of scores that were lower than that of a given person or instance. On a standardized test, a student scoring in the 93rd percentile means 93% of the scores were lower than the student's.

**Pie chart** – showcases the parts of a whole or percentages of a total through the use of "slices" within a circle.

**Pivot table** – function in Excel allowing users to create a temporary table of a specific column (or category) and check for duplicates in the data.

**P-value** – a statistical term that refers to evaluating the significance of a hypothesis. While widely used in scholarly work, the importance of calculating p-value has been debated of late.

**Quartiles** – the three values that divide a numeric dataset into four equal parts.

**Qualitative data** – information that is best captured in words or "qualities." For example, answers in an interview are qualitative data.

**Quantitative data** – data that is measured by counting and/or in numbers.

**Range** – the difference in values from the maximum to the minimum.

**Research question** – the question or hypothesis that is used as the grounds for experimentation or research.

**Regression analysis** – a statistical process for determining the relationships among variables, specifically one independent variable and a number of dependent variables (causation instead of correlation).

**Sample** – a part or fraction that represents a larger group. Using a sample in lieu of studying a larger group is cost-effective and efficient, but to be accurate, samples should be representative of that whole group and large enough to discern patterns.

**Scatterplot/Scatter graph** – a graph in which each data point is represented by a dot. Lines do not connect the dots.

**Segment** – a portion of a whole, such as a slice in a pie chart.

**Self-reported** – an adjective that describes data donated by and about participants. When people report data about themselves, there is a tendency to provide answers skewed toward one's "best self."

**Skew** – A distribution is considered skewed if, when graphed, one of its "tails" is longer than the other.

**Sort function** – an Excel feature that that allows the user to sort and sequence the information in a spreadsheet's column.

**Statistics** –  a branch of mathematics focused on the collection, analysis, interpretation, presentation, and organization of quantitative or numerical data.

**Storyframing** – a process tool designed and named by Connie Williams  in which students use sticky notes to experiment with the content and format of an infographic.

**Table** – a set of data arranged in rows and columns.

**Terms of art** – specialized terminology, or jargon, associated with a particular field of study.

**Variables** – A characteristic or object that can be counted.

**Vetted** – describes information which has been checked by a professional for accuracy.

**Waffle chart** – a square graph used to compare percentages between different categories.

**Web polling** – a survey conducted over the Internet.

**Web portal** – see *Online portal*.

**Wireframing** – a rapid model showing the placement of images, graphic design and content of a web page or document.