STATISTICS AND PROBABILITY

01 ANALYZING CATEGORICAL DATA

* Individual: A single member or unit of a population or sample being studied.
* Variable: A characteristic or attribute that can take on different values among the individuals.
* Categorical variable: A variable that represents distinct groups or categories (e.g. gender, color, class...)
* Bar chart: A graphical representation of categorical data using rectangular bars where the length or height of each bar corresponds to the value it represents.
* Venn Diagrams: A visual representation of mathematical or logical relationships between different sets, using overlapping circles (or other shapes) to show common and distinct elements.
* Two-way Table: A table that displays the frequency distribution of two categorical variables, showing how the categories of one variable relates to the categories of another.
* When analyzing two-way relative tables, pay attention to whether it is column-relative or row-relative or both. For example, if it's only column-relative, sum of the values on the same row does not equal to 1.00 (100%).
* Distribution in Two-way Tables.
  * Joint Distribution: Two (or more) random variables occurring together (e.g. 20 out of 200 students studied 21 to 40 minutes per day and answered 60-79% of all questions correctly.)
  * Marginal Distribution: Ignoring (marginalizing out) other variables and giving the distribution of remaining variables (e.g. 70 out of 200 students answered 40-59% of all questions correctly.)

* Conditional Distribution: The distribution of one random variable given that another random variable takes a specific value (e.g. 14.28% of the students who studied more than 60 minutes per day answered 60-79% questions correctly)