| **Statistics** | **Probability** |
| --- | --- |
| * We use information from a sample to make an assumption about a population | * We assume certain characteristics of a population and apple those to a sample |
| **Descriptive** - Graphical and numerical summary of data  **Inferential** - techniques used to analyze data and draw conclusions | |

| **Population**  Definition: The entire collection of the individuals we are studying  **Parameter** - A value that describes a population  E.g. 650,000 acres of land in this public landscape | **Sample**  Definition: The subset of a population that we can observe and measure  **Statistics** - A value that can describes a sample.  E.g. The mean circumference of our sample of trees is 16.5. |
| --- | --- |

# Types of Data

| **Qualitative** (Categorical) - variables are classified into categories | **Quantitative** (Numerical) - Characteristics that are measurable |
| --- | --- |
| **Nominal** - Variables that don’t have order   * **Identifier** - A number that isn’t used to | **Discrete(ratio)** - Variables represented as a whole number |
| **Ordinal** - Variables that do have order | **Continuous** - Variables that have a range of possible values |

# 

# Sampling methods

**Simple random sample** - Like

a lottery, every individual in the population has an equal chance of being chosen. Usually we achieve this with technology, a random number generator. Like a lottery, every individual in the population has an equal chance of being chosen. Usually we achieve this with technology, a random number generator.

**Sample of convenience** - No method of randomization, but rather individuals are chosen by proximity or ease of access.

**Stratified sample -** The population is divided in groups(stratas) and the same number of random people are chosen from each group.

**Cluster sample -** A random selection of clusters or groups are chosen and every member from those clusters are selected

**Systematic sample -** A starting place is randomly chosen then a fixed number of individuals are selected.

**Voluntary response sample -** Members who participate volunteered to do so

# 

# 

# Experiment types

| **Experimental unit(People are subjects) -** Wherever we are studying or observing  **Outcome(response) -** What we measure on each unit  **Treatment -** The procedures applied to each uni**t** |
| --- |

* **Randomized experiment -** The investigator assigns treatments to the units at random
  + **Double-blind -** Neither the investigator or the subjects know who get assigned what
  + **Randomized block -** We account for major differences(such as age) by assigning the same number of subjects in each group to each treatment
* **Observational study -**
* The assignments to the treatment groups is not chose by the investigator
* This is necessary when applying treatments to people(subjects) would be unethical.
  + **Confounder**
  + **Cohort study**
    - **Prospective**
    - **Cross sectional**
    - **Retrospecive**
      * **Case control studies**

# Sources of bias

* **Voluntary bias** - People with strong opinions( negative or positive) are more likely to volunteer their responses.
* **Self-interest bias** - When people have some interest in the outcome.
* **Social acceptability bias** - When people feel judged, embarrassed, etc by their responses
* **Leading questions** - The way questions are worded can influence our responses
* **Nonresponse** - A person refuses to respond
* -

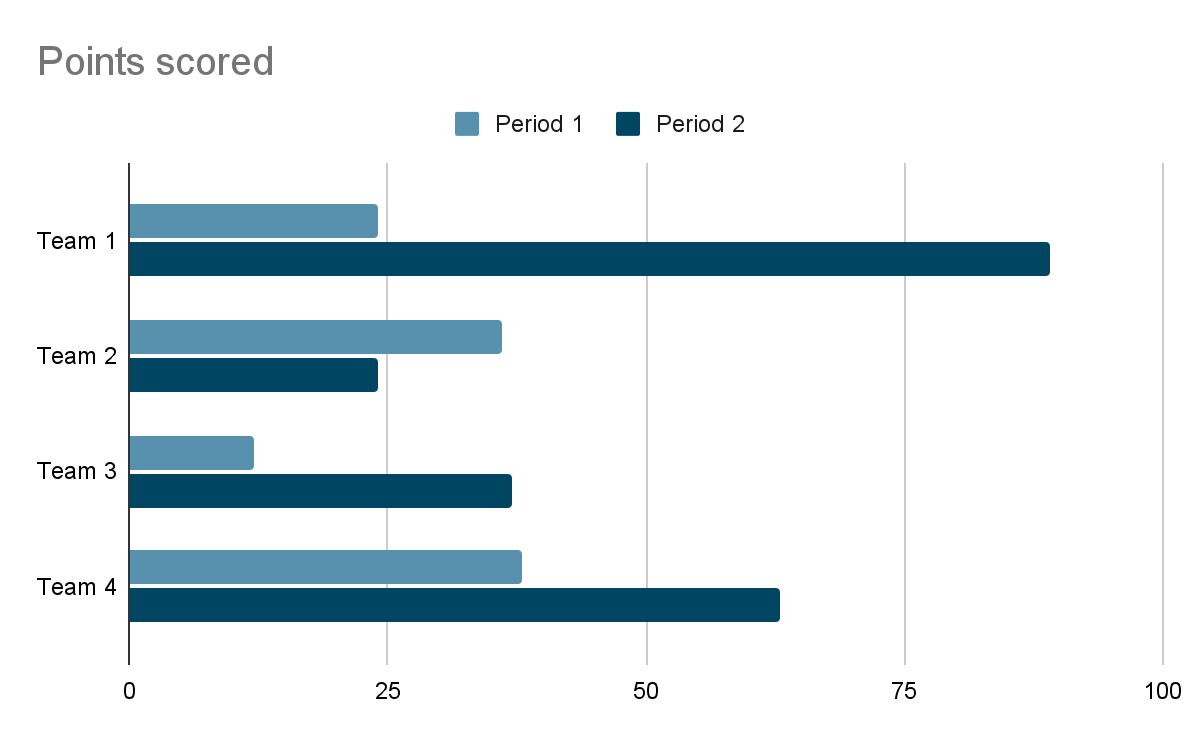
# Graphs of Qualitative data

**Frequency** - the number of items in each category.

**Relative frequency** - the proportion of something relative to the whole; the count divided by the total number in the entire sample or population.

Relative frequency = Frequency/sum of all frequencies

## Bar graph



## Histogram

Skewed - When the distribution is clustered in one direction

* Left skewed - the tail is in the left direction
* Right skewed - the tail is in the right direction

Symmetrical

Modes of histograms - the most common occurrence. In histograms these are peaks

* Unimodal - One distinct mode
* Bimodal - Two or more distinct modes
* Multimodal - More then two mores
* Uniformal - There is no distinct mode. All values are similar

## Stem leak

| stem | Leaf |
| --- | --- |
| 1  2  3  4 | 2 4 8  2 2 3 5 5 8  0  5 7 8 |
|
|

The stem represents how each data starts. The “leaf” is all possible combinations with the stem. Ex 1 - 12, 14, 18

## 