Lecture 1

- Data is a set of measurements taken on a set of individual units, usually called cases
- **Categorical** variables divide the cases into groups, e.g. Movie genres
- **Quantitative** variables measure the numerical quantity for each case e.g. Age
- **Explanatory/Response variables** e.g. Does meditation (**explanatory)** help reduce stress **(response**)?
- **Population** includes all individuals or objects of interest
- **Sample** is all the cases that we have collected data on (a subset of the population)
- **Statistical inference** is the process of using data from a sample to gain information about the population
- **Sampling bias** occurs when the method of selecting a sample causes the sample to differ from the population in some relevant way - if sampling bias exists, we cannot trust generalisations from the sample to the population
- SAMPLING GOAL - Select a sample that is similar to the population, only smaller
- Avoid sampling bias by taking a **random sample**, only random samples can be trusted to make generalisations
- Forms of bias to look out for:
- Question wording
- Context
- Inaccurate responses