# What is statistics?

A collection of principles and parameters for gaining information in order to make decisions when faces with uncertainty.

It is a fact or piece of data from a study of a large quantity of numerical data

Statistics is the art of connection numbers to these questions soi that the answers evolve,

To make quantitative connections to qualitative questions is the heart of statistics

To make decisions based on data and not only on instinct

# How are we going to do this?

Statistics are not about pretty neat data: it’s about carefully selected data

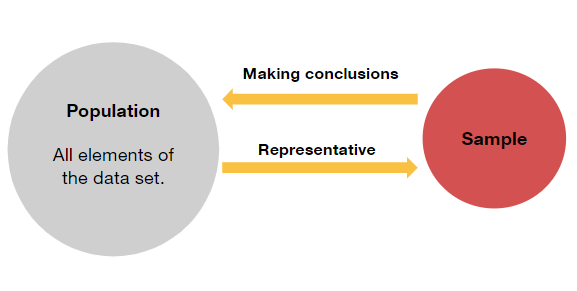
It is about asking skeptical questions like:

* ” Are we sure that this cool result isn’t just white noise?”
* “How was this data gathered?”
* “Is the sample Biased?”
* “If it is, what can be done to remedy that?”
* “How many metrics did we look at before arriving at that amazing finding? A lot?”
* “What can we do about that?”
* “Could something other than causation explain why X and Y move together?”
* “How can we control for other factors Z that might be involved?”

# Types od statistical analysis

* Descriptive:
  + Describes the characteristics of a data set in the form of coefficients, tables and graphs.
  + Ex: Average temperature, average height and weight, average sales per month.
* Inferential
  + Asks a hypothetical question about the **populations** and finds the answer by testing a smaller **sample** of data using descriptive statistics.
  + Ex: Opinion polls, predicted sales, volume for a new product, customer surveys.

# Population vs. sample



The population is defined based on your problem

* The population could be the entire population of a country, or it could be all of your current customers.

Rather than analyzing the whole population, most analyses are based on samples.

**The results from the analysis of the sample are used to infer the behavior of the population**

If your population consists of 10000 people, your sample might consist of just a few hundred

* i.e. a survey of a sample of you customers.

Equally, if you produce 10000 products a day you will take a sample to analyze performance, identify rates of defects etc.

When performing inferential statistics, it is important to make sure your sample size is large enough to be statistically significant

* ex. A sample of 10 from 10000 would not be enough to provide reliable estimates.