University of BRISTOL

WEEK 1

The role of experimentation

Felipe Campelo

In this lecture…

In this lecture, we'll discuss different ***classes of experiments*** and their main characteristics. In summary, we'll cover:

- The role of experimentation in (data) science.
- Important experimental principles.
- Classes of experiments – retrospective vs. observational vs. designed.
- Some basic experimental strategies and guidelines for designed experiments.

The scientific process

- Normally shown as a simple “recipe”.
- Oversimplification of a complex and iterative process
- Actually includes:
  - Several activities, performed at different stages.
  - Interaction with the community.
  - Creative, *out-of-the-box* thinking.
  - Preliminary conclusions subject to revision when new / better data becomes available.
  - Learning from failure as much as from success.

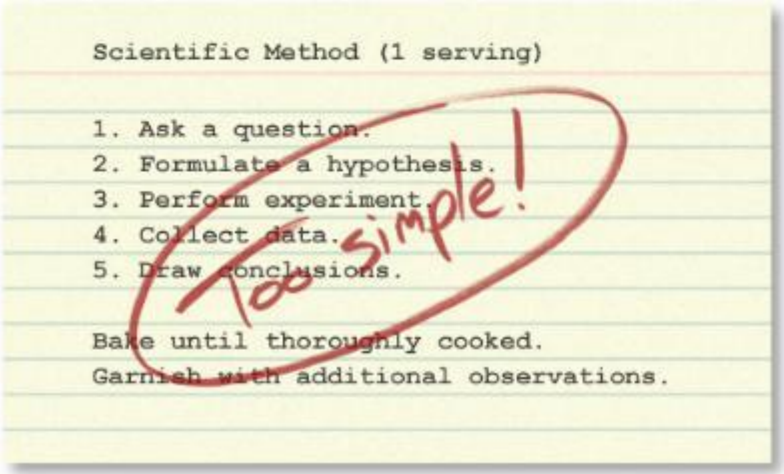


Image: Understanding Science, 2015 (http://goo.gl/7cCGaz).

The scientific process

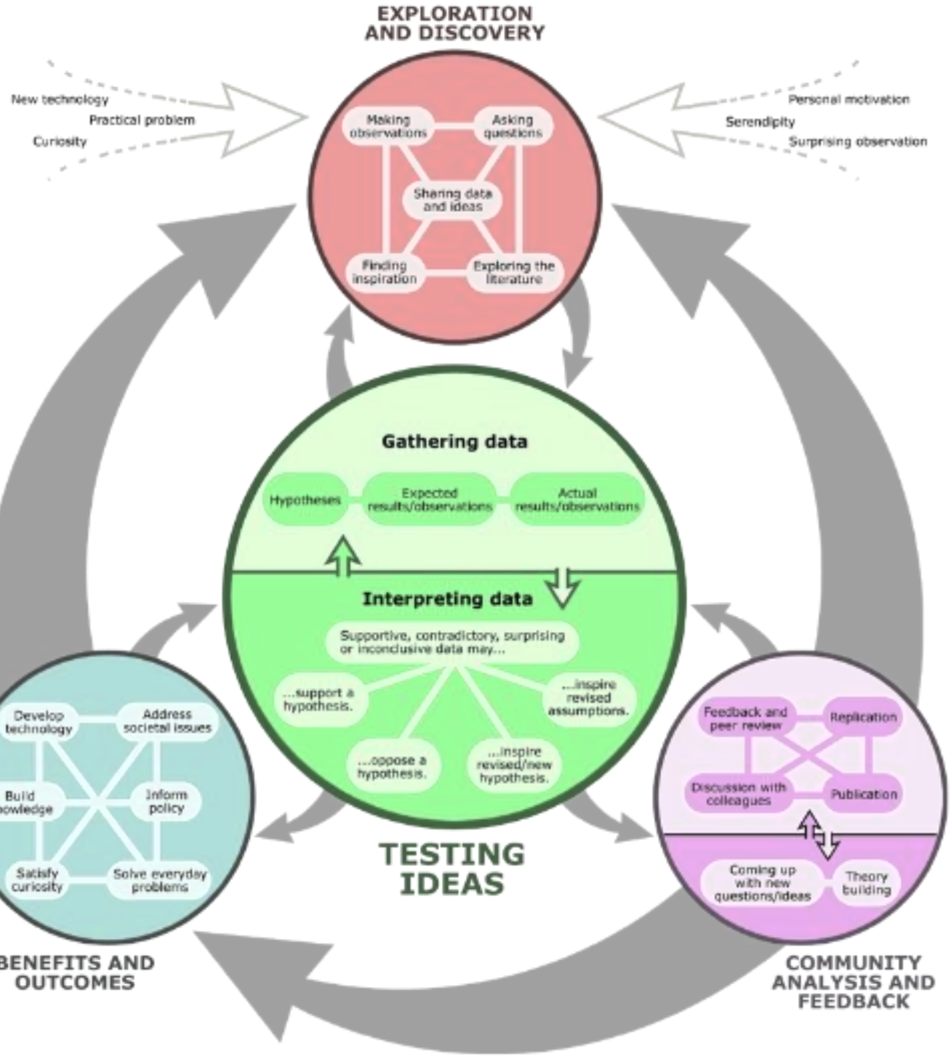
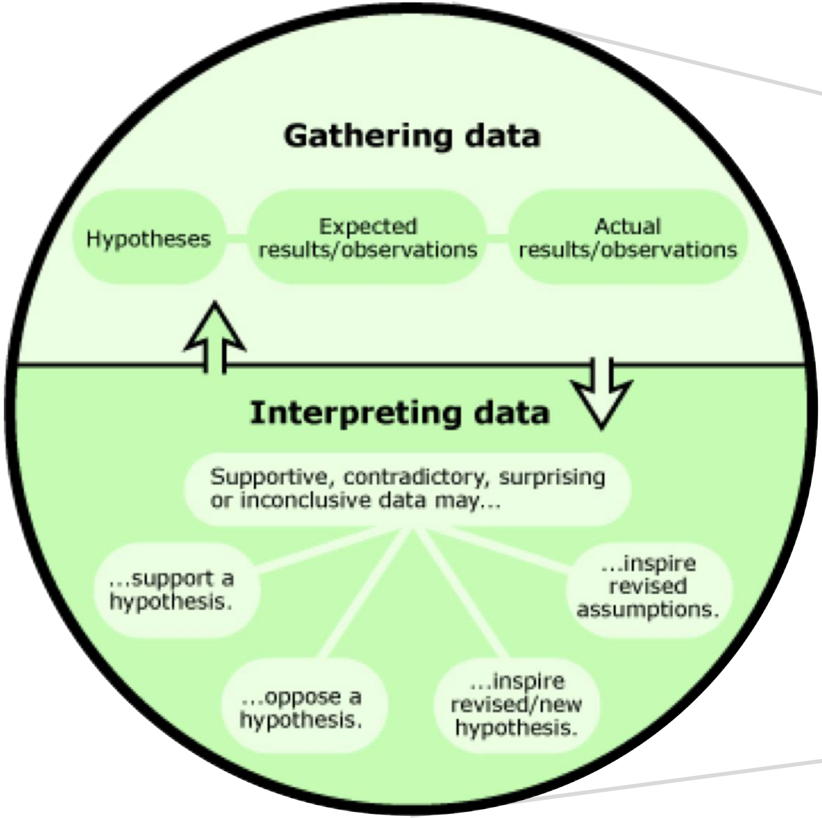


Image: Understanding Science, 2015 (http://goo.gl/7cCGaz).

The scientific process



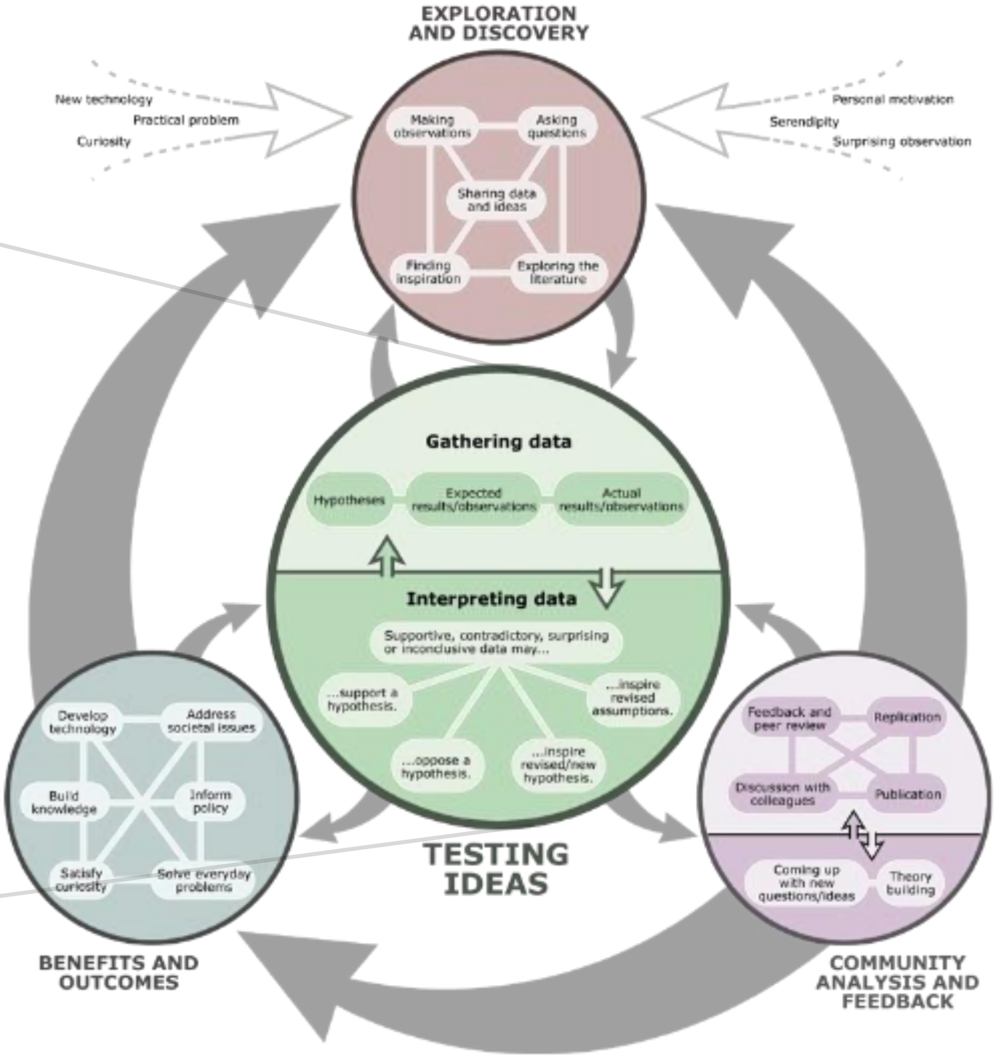
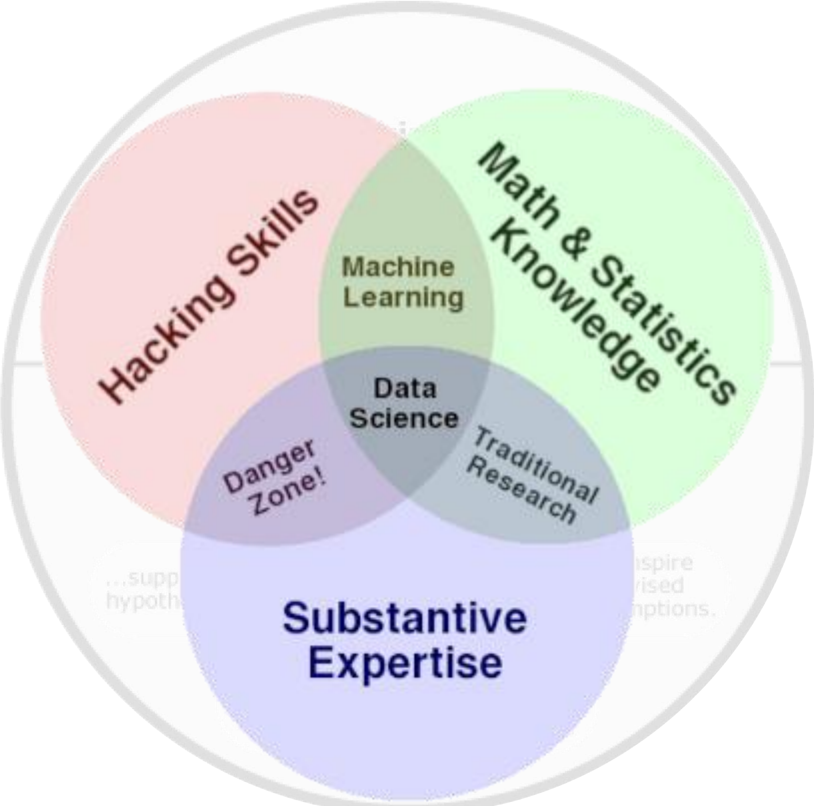


Image: Understanding Science, 2015 (http://goo.gl/7cCGaz).

Why is it important?



Even if you don’t work as a researcher, having a solid grasp of the methods of science – how to test and compare things, how to quantify and account for uncertainty, how to model and predict, and how to think statistically – is a critical skill to effectively use data science or ML/AI in any domain.

Image: http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

What is an experiment?

An experiment can be defined as a test (or a series of tests) wherein changes are introduced in the state of a system or process, enabling the observation and characterisation of effects that can occur because of these changes.

Performed with some *objective* in mind, e.g.:

- Uncovering influential variables in a system.
- Determining desired values for certain parameters.
- Characterize the behaviour of the system under study.

Pre-experimental design

Before planning an experiment (computational or physical), it is important to check some things, such as:

- Do you (or does your team) have all the required **background knowledge**?
  - It is important to understand the *problem domain*.

- Is the experiment **necessary**?
  - Maybe someone else has already answered your question of interest?
  - Checking the (technical or scientific) literature is important.
  - Knowing how to select your information sources too!

- Is the experiment **feasible**?
  - Can it be done with the available resources?
  - Is it sufficient to help you answer your specific question?

Main types of experiment

Retrospective studies

Use of historical data / existing datasets of previous observations.

Cheap and convenient, but no control over how or which data was collected, experimental confounders etc.

Example: using data from blood donations serving ~20M people to investigate hidden variables (such as proportion of asymptomatic infected individuals) in Covid-19 epidemiological models.

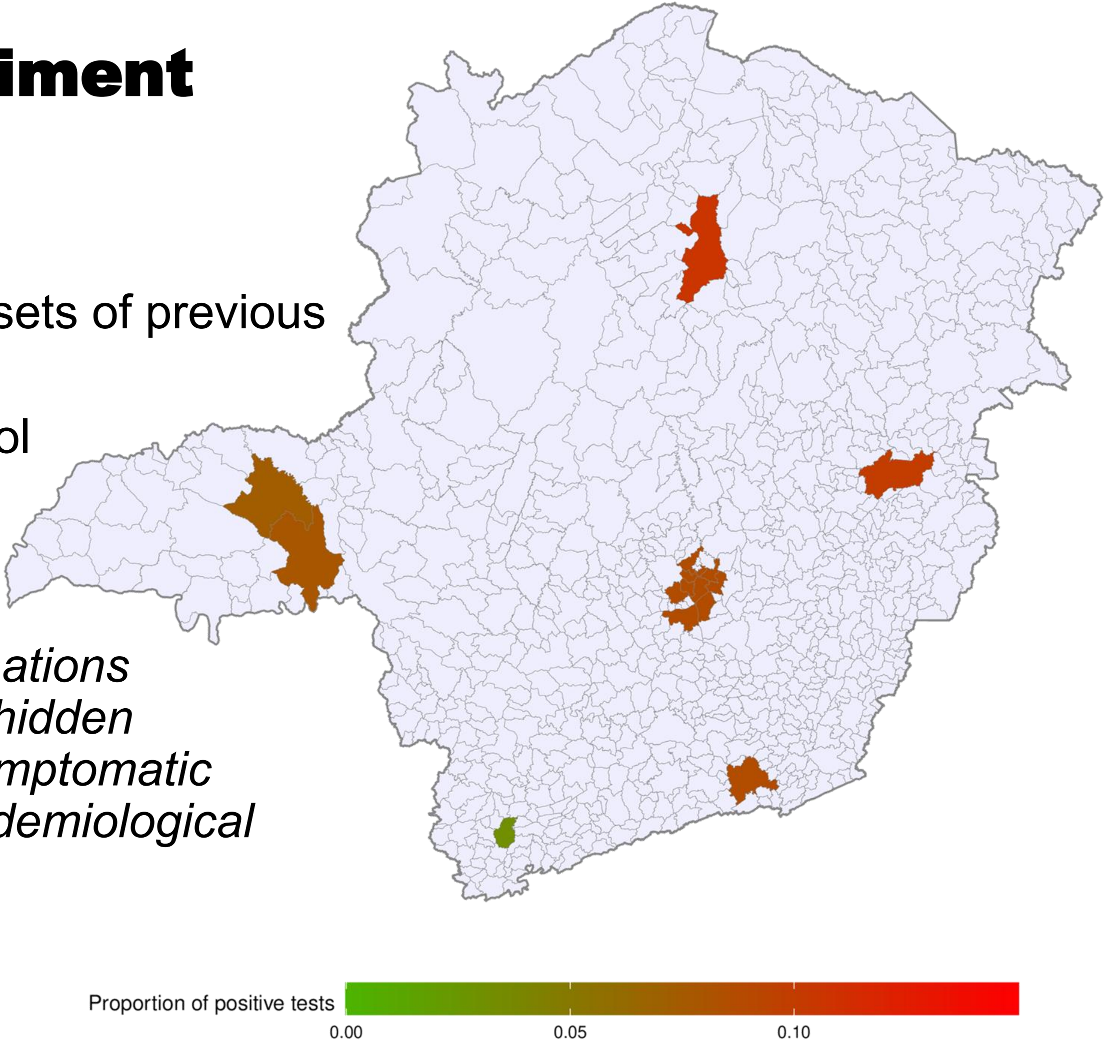


Image: https://doi.org/10.3201/eid2804.211961

Main types of experiment

<u>Observational (a.k.a. natural) experiments</u>

Observation of a system or phenomenon under “usual conditions”, with minimal disturbance.

Some control over which data is collected, but potential issues with data representativeness, especially of uncommon cases.

Example: looking at the spatial distribution of cholera cases to infer the likely location of an infected water source.

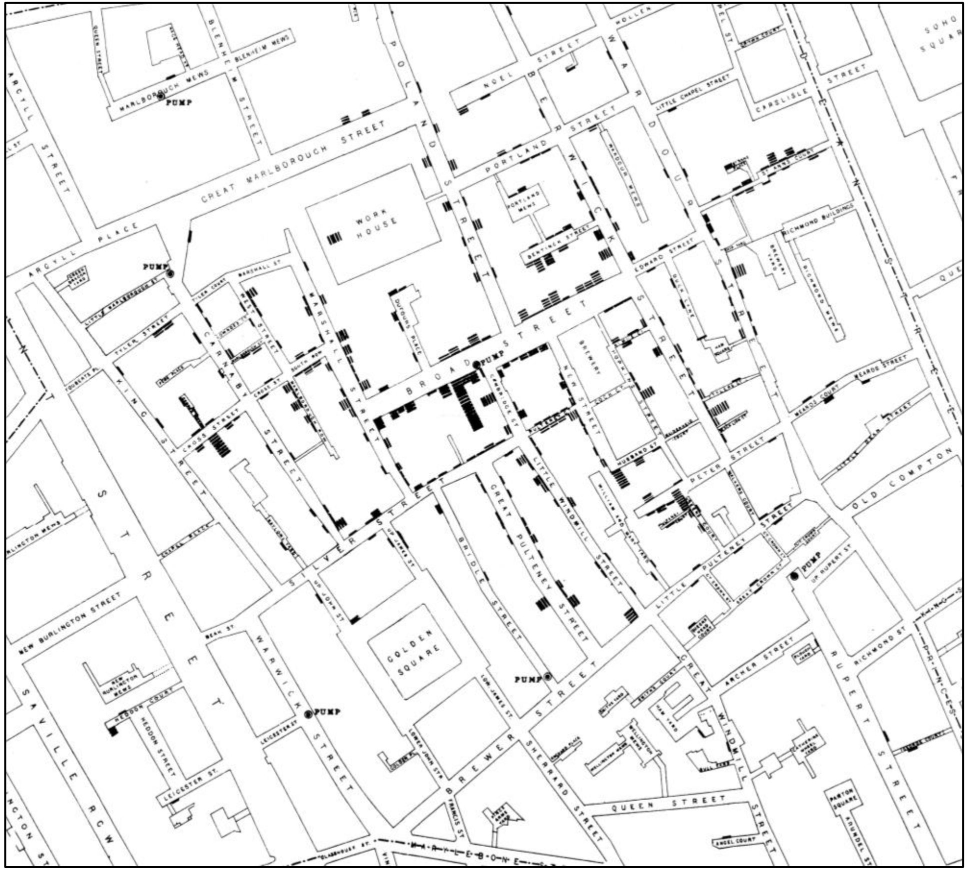


Image: https://commons.wikimedia.org/wiki/File:Snow-cholera-map-1.jpg

Main types of experiment

Designed experiments

Introduction of deliberate / planned changes in predefined *experimental factors,* while controlling or accounting for other potential confounding variables.

When possible, it is considered the gold standard of experimentation.

Example: designing biological datasets with different data volumes and proportions of observations from a certain organism to investigate the effect of those factors on the performance of predictive models.

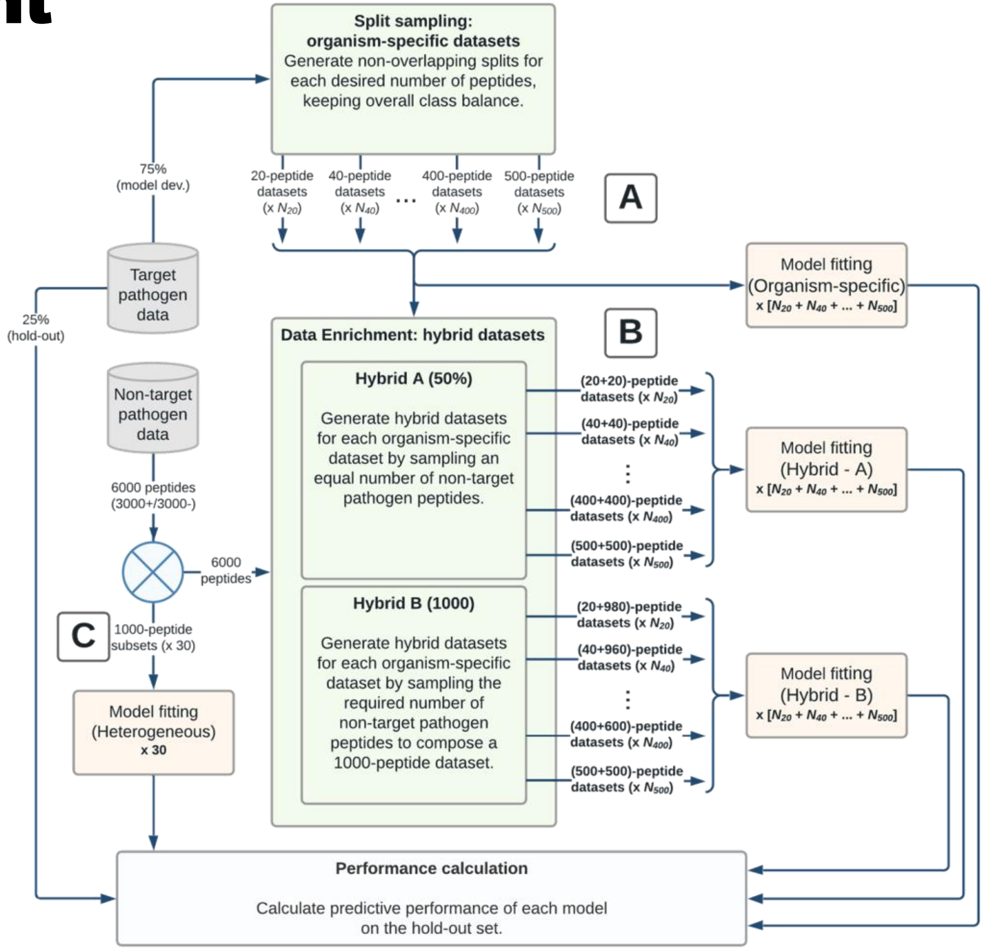


Image: https://doi.org/10.1109/BIBM58861.2023.10385381

Experimental strategies

<u>**Heuristic (“educated guess”)**</u>

- Select a combination of values for the experimental factors/variables (e.g., based on experience or plausibility);
- Test and record/observe the result;
- Change one or two factors at a time based on observed results, then re-test.

Can occasionally achieve reasonable results, but has lots of limitations

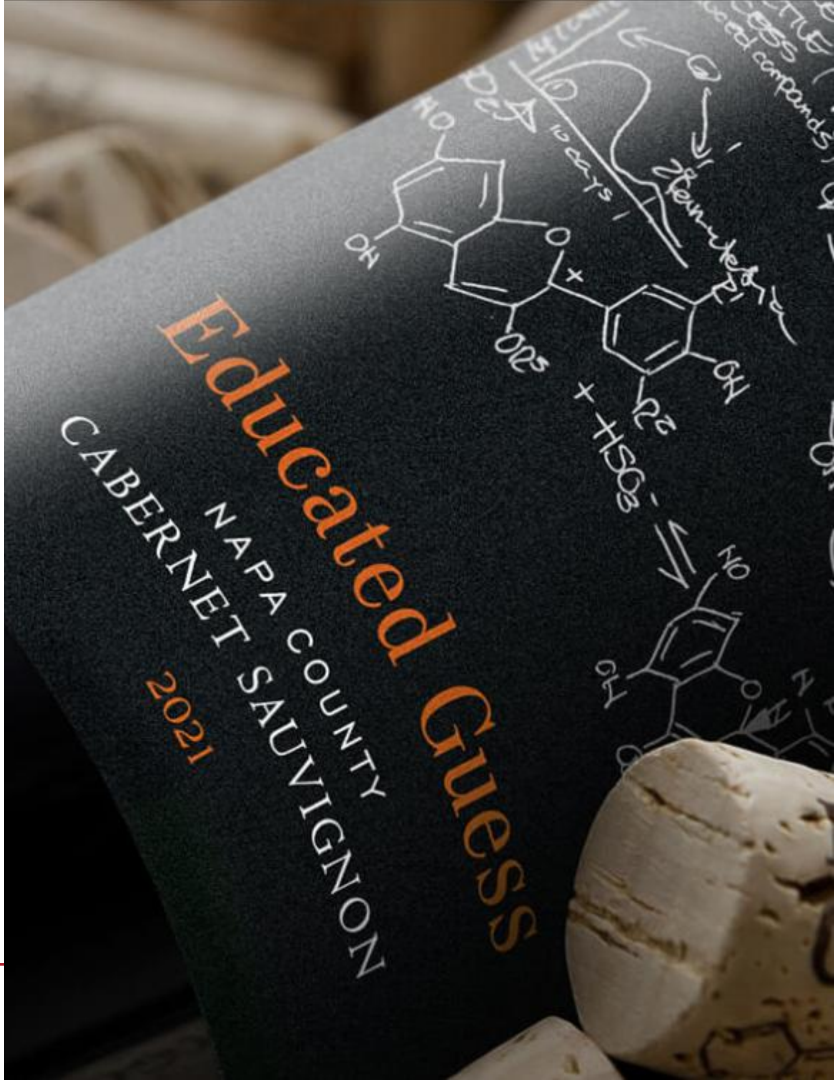
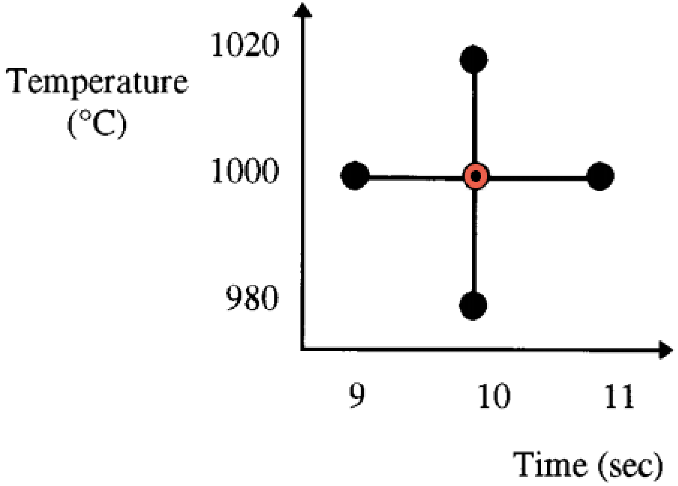


Image: https://rootsrundeep.com

Experimental strategies

OFAT (one factor at a time)

- Select a combination of values for the experimental factors/variables (e.g., based on experience or plausibility) and set it as the reference point
- Change each factor individually in a systematic manner, keeping all others constant.



Can achieve good results *as long as there are no interaction effects*

Image: D.C. Montgomery, “Design and Analysis of Experiments”.

Experimental strategies

<u>Factorial experiments</u>

- Select different *levels* of each factor/variable
- Change the factors simultaneously in a planned, systematic manner.

Can estimate both main (marginal) effects and interactions.

Provides greater precision of model coefficients.

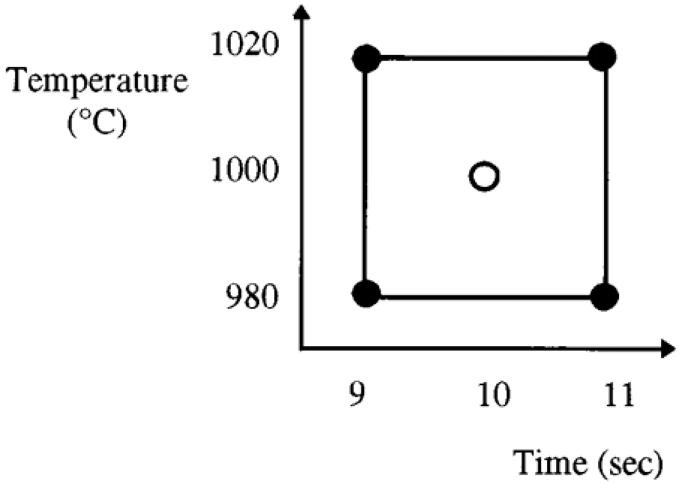


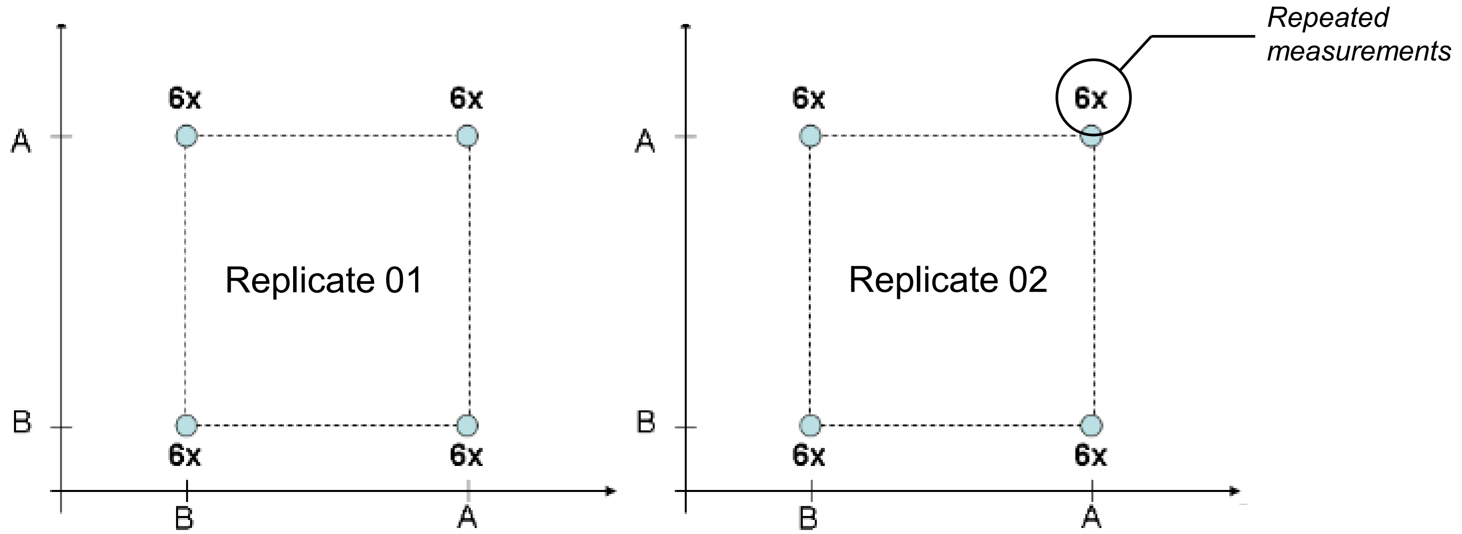
Image: D.C. Montgomery, "Design and Analysis of Experiments".

Important concepts in the design of experiments

Repetition and replication:

Replication helps estimate the *experimental error.*

Repeated measurements allow estimation of *within-group variability, and* allow estimation of *statistical uncertainty,* which is key to adequately test hypotheses with a quantifiable level of confidence.



Important concepts in the design of experiments

Randomisation:
Needed when the ordering or arrangement of samples may induce unwanted biases in the data (e.g., allocation of users in an A/B test or patients in a clinical trial).

Helps reduce bias due to the influence from uncontrolled covariates;

Stratified randomisation is often used to ensure balance across groups (e.g., stratify users by some demographic such as age band or sex, then randomise allocation within each group).

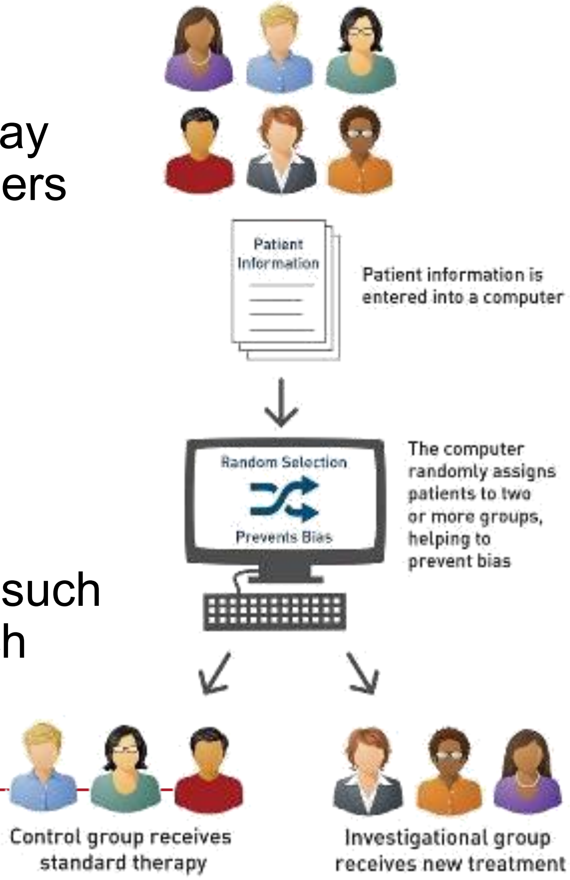


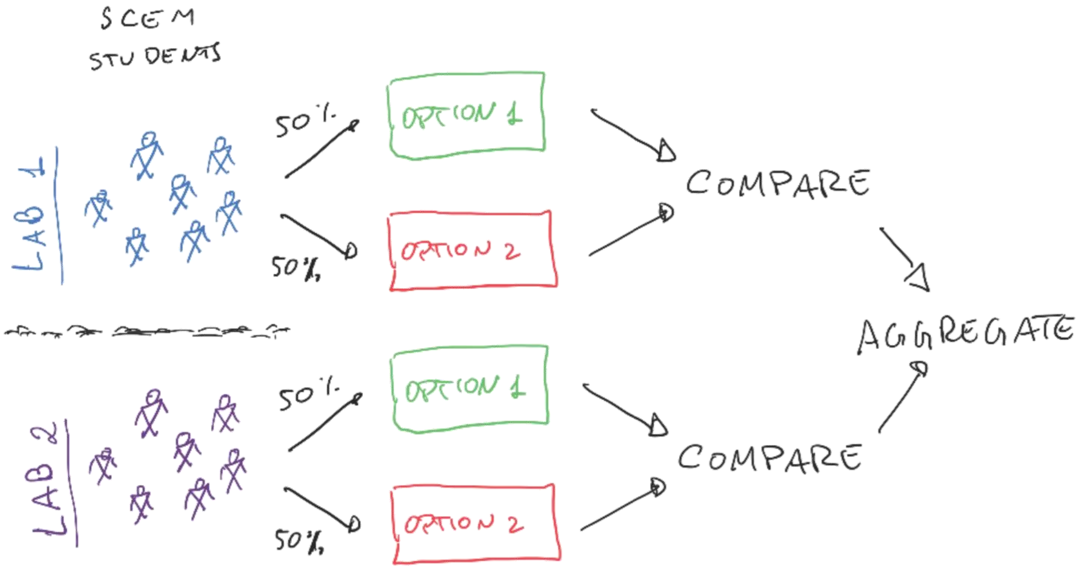
Image: https://www.cancer.gov/clinical-trial-randomization-infographic

Important concepts in the design of experiments

Blocking:

To reduce the effect of ***nuisance variables*** (sources of variability you're not primarily interested in), it is common to group experimental units into relatively homogeneous "blocks".

Randomisation is then done within each block, ensuring that comparisons between "treatments" are made under similar conditions and increasing the precision of estimates.



Important concepts in the design of experiments

Blinding:

In many cases, knowledge of group assignment or expected outcomes can **bias** behaviour or analysis.

(*Example: in clinical trials, it is common for patients, doctors and data analysists are not informed if they are in the "new treatment" or "old treatment" groups, to prevent this knowledge from influencing their behaviour, observation, or analysis).*

Whenever that's the case, it is important to "blind" the people involved with the execution and analysis of the experiment, to prevent biases from contaminating the results.

Sometimes these biases can creep in during data collection and measurement[1], so it is important to control or to know (as much as possible) how the data was collected.

---

[1] For a famous example of biases arising due to non-blinding, check
Maddox, Randi and Stewart (1988), *"High dilution" experiments a delusion*, Nature 334.

Guidelines for a good experimental design

- Pre-experimental planning
  - Acquisition of domain knowledge and problem definition
  - Selection of experimental factors and of nuisance variables to control or measure.

- Experimental design
  - Translation of the technical/scientific question into a statistical one.
  - Determination of sampling regime (randomisation, stratification, blocking, blinding etc.)
  - Selection of adequate type of inferential test
  - Determination of desired statistical properties and sample size estimation

- Running the experiment according to the plan.

Why design experiments?

A good experimental design (or, alternatively, a solid and domain-informed analysis plan for retrospective experiments) is essential to prevent nuisance factors or personal biases from creeping into the results. It allows the experimenter to select which experimental factors to test, how to test them, at which levels/values and in which combinations.

A good experiment is not done to "*show that your idea is right*" or that "*your algorithm is better than another*".

Instead, it is ***a systematic way to carefully and thoroughly observe some data-generating phenomenon and use consistent logic to evaluate the results***. Results must always reflect the observed reality, not expectation or convenience.

"The great tragedy of Science – the slaying of a beautiful hypothesis by an ugly fact"
T.H. Huxley