Lecture notes for this lesson:

Introductory video: https://docs.google.com/presentation/d/16CTB8spdjaz3xy4PZ7cvB4f28XKNduvANuesITypvbU/edit?usp=sharing

Causality: https://docs.google.com/presentation/d/1DqxwVDvV5JRdzlGdEB9FKcK2UnZ9j1YANlgxwVbCbV8/edit?usp=sharing

This is the longest module of the class. If you can make it through this one, you're home free. In this module we discuss the ideal (the experiment is carefully designed) with what is often reality (the data is observational) and cover some solutions.

A key topic of the lecture is **confounding.** Confounding occurs when you want to compare two things and a third gets in the way. As an example, you want to look at ad performance and purchases. However, the ads ran on different sites, so were thus seen by a different audience. The different audiences may have different purchasing patterns, so any difference seen may not be due to the ad campaign but instead may be due to the audience. Read more about confounding here https://en.wikipedia.org/wiki/Confounding

An interesting component of randomization in experiments is the ability to estimate **causal** effects. We define causal effects as the difference between the outcome for a subject observed at a particular treatment minus the outcome observed as a a control. However, a subject can only receive one of the treatment levels, so only one of these two gets observed. The other, is called a counterfactual. We can estimate counterfactual effects because of this. However, we can, under assumptions estimated averaged counterfactual effects if we have randomization. The study of how to estimate causal effects using data is called causal inference. I find the most useful aspect of causal inference and counterfactuals is the way thinking about them helps me think about different experimental designs. You can read more about causal inference here https://en.wikipedia.org/wiki/Rubin_causal_model