

Causal Inference

MIXTAPE SESSION

**MIXTAPE
SESSIONS**



Roadmap

Hidden curriculum

- Background

- Empirical workflow

- Hierarchical folder structure

- Naming conventions

- Version control

- Soft skills

Name stuff well

1. hierarchical folder structure
2. automation
3. **naming conventions**
4. version control

Different elements

- Three different things you must name:
 1. variables,
 2. datasets
 3. programs
- You do these three things repeatedly therefore you need a system otherwise errors will occur due to the repetition
- Anything that can happen will happen with enough trials. – Murphy's law

Naming conventions for variables

- Variables should be readable to a stranger
 - Say that you want to create the product of two variables. Name it the two variables with an underscore
 - `gen price_mpg = price * mpg`
- Otherwise name the variable exactly what it is
 - `gen bmi = weight / (height^2 * 703)`
- Avoid meaningless words (e.g., `1mb2`), dating (e.g., `temp05012020`) and numbering (e.g., `outcome25`) as your future self will be confused

Naming datasets and do files

- The overarching goal is always to name things so that a stranger seeing them can know what they are
- One day you will be the stranger on your own project! Make it easy on your future self!
- Choose some combination of simplicity and clarity but whatever you do, be consistent
- Avoid numbering datasets unless the numbers correspond to some meaningful thing, like randomization inference where each file is a set of coefficients and numbered according to FIPS index