Causal Inference

Mixtape Session

Roadmap

- Hidden curriculum
  - Background
  - Empirical workflow
  - Hierarchical folder structure
  - Naming conventions
  - Version control
  - Soft skills

Automation

- A second problem I had was that I was copying and pasting information
    1. hierarchical folder structure
    2. **automation**
    3. naming conventions
    4. version control
- Automation does not allow errors to creep in, whereas hand copying does

Code not Command Line

- Your future self doesn't remember making any of these tables or figures
- All tables and figures must be replicable using a scripting file, not the user interface or command line
  - → It's fine to use the command line
  - → But it must eventually go into the program
- Final output must be produced by running the entire code, not just chunks

Beautiful code

- Your ideal goal is to make beautiful code which is easier for some than others
- At minimum, don't make it ugly – if your future self can't read it, it's ugly or it's confusing but both are bad
- Consider a new text editor like Visual Studio Code which allows for colored syntax, indentation, column editing and bundles
- Stata and Rstudio also come with built-in text editors which are fine

Headers

```
*************************************************************************
* name: texas.do
* author: scott cunningham (baylor university)
* description: estimates the causal effect of prison capacity
*              expansion on incarceration rates using synth
* date: march 19, 2018
*************************************************************************
```

Speak slowly in your programs

“Be conservative in what you do; be liberal in what you accept from others.” - Jon Postel

- Smart sounding quote about both programming and relationships
- Your future self is time constrained, so explain *everything* to her as well as write clear code
- Optimally document your programs
- But speak your future self’s love language so she understands

Automating tables

- Your goal is to make “beautiful tables” that are never edited post-production as well as readable on their own
- Large fixed costs learning commands like -estout- or -outreg2-: incur them bc marginal costs are zero
- I use -estout- because Jann has written an excellent help file at `http://repec.org/bocode/e/estout/hlp_esttab.html` but many like -outreg2-
- I’ve uploaded code in Stata that will automate some simple tables

Automate figures

- Again goal is replication, accuracy and efficiency
- If you're doing something a few times, learn to automate it
- Learn how to make customizable graphs using automation, not by tinkering in post-production
- Examples include Stata's -twoway-, R's -ggplot2-, Python's -matplotlib- and its various wrappers

Data visualization resources

Study other people's pictures and get help from experts

1. Kieran Healy's 2018 Visualization: A Practical Introduction (Princeton University Press); free version is http://socviz.co/index.html#preface.
2. Ed Tufte's book Visual display of quantitative information is classic, but more a coffee table book plus no programming assistance.
3. Ben Jann deck for making beautiful graphs in Stata https://www.stata.com/meeting/uk18/slides/uk18_Jann.pdf

```
. set scheme s2color
. two (lpolyci pfor length, clstyle(p1line))
>     (lpolyci pdom length, clstyle(p2line))
>     (scatter price? length, pstyle(p3 p4 p5))
>     , ytitle(Price) legend(order(2 "Domestic"
>         4 "Foreign" 5 6 7))
```