

Applied Econometrics and Data Analysis

ECON 121

People

▶ Me: Tom Vogl

- ▶ User, not developer, of econometrics
- ▶ Researcher of health and population in developing countries
- ▶ Have been teaching variants of this course for over a decade
- ▶ Teaching is very important to me

▶ TA: Regina Calles-Martínez

- ▶ Will hold biweekly problem set labs, super useful



Roadmap for the Quarter

1. Estimation

- ▶ “Review” of OLS
- ▶ Departures from i.i.d.
- ▶ Maximum likelihood
- ▶ Limited dependent variables
- ▶ Panel data

2. Causality

- ▶ Difference-in-differences designs
- ▶ Potential outcomes
- ▶ Randomized experiments
- ▶ Instrumental variables
- ▶ Regression discontinuity designs



Prerequisites

- ▶ **Option 1: ECON 120C**
 - ▶ Targets Econ, Math/Econ, ManSci majors after full 120 series
- ▶ **Option 2: ECON 5 & ECON 120B**
 - ▶ Targets BusEcon majors, who must take 5 but not 120C
- ▶ **Either is solid**
 - ▶ Students w/ 120C will have seen many 121 topics before
 - ▶ Students w/ 5 will have more experience with statistical computing



Course Structure

▶ Text

- ▶ No textbook, will rely on course notes

▶ Participation

- ▶ Participation matters for grade but can take many forms

▶ Deliverables

- ▶ Problem sets (5): R-based, can code in groups (max 4 people) but must write own answers, lowest score dropped
- ▶ Academic articles (4):
 - ▶ Group presentation for one article
 - ▶ Multiple choice quizzes for all articles, lowest score dropped

▶ Final Exam

- ▶ Open book, R-based, basically an extra problem set
-



Lectures and Assignments

▶ Lectures:

- ▶ If people behind you might see non-course-related material on your laptop, please sit toward the back of the class
- ▶ I will post written notes by the night before each lecture, so you can print them or put them on your tablet for notetaking
- ▶ I will post my whiteboard afterward

▶ Problem sets:

- ▶ Work in groups; let us know if you need help finding partners
- ▶ Late problem sets not accepted
 - ▶ Lowest score dropped
 - ▶ If you are late, you should do it anyway



Grading

- ▶ I care about your learning, not about your grades
- ▶ I would love to teach you without giving you grades
- ▶ But both students and UCSD expect me to give grades, so I try to do it fairly and generously
- ▶ I assign letter grades based on your final course score:
 - ▶ 40% final exam
 - ▶ 40% problem sets
 - ▶ 20% quizzes, presentation, and participation
- ▶ Grades are curved to typical upper division ECON distribution, but the curve **helps** you, never hurts you
- ▶ If you fall behind, e-mail me – I am happy to offer a path to get back on track, but you will have to work for it



Statistical Computing

- ▶ Course used to use Stata

- ▶ Easier implementation of methods we study
- ▶ Used in ECON 120 series
- ▶ Less common outside academia (\$\$\$)

- ▶ Switched to R last year

- ▶ Free, more common in industry
- ▶ Avoided in the past because of the package zoo, but I figured out we can do most topics using `tidyverse` and `fixest`
- ▶ I hope you find it useful, but feedback is very useful

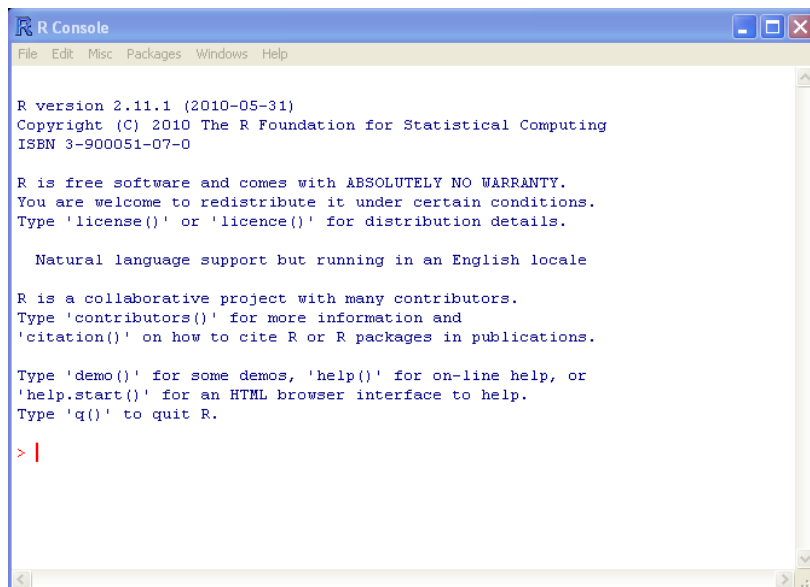
- ▶ Stuck on R?

- ▶ Refer to my examples from class (on GitHub)
 - ▶ Use help files / Stack Exchange / Google / ChatGPT
-



Background on R and RStudio

- ▶ R: programming language for statistical computing
 - ▶ Like my first kitchen: only basic tools, every task requires work
- ▶ RStudio: integrated development environment for R
 - ▶ Like Chancellor Khosla's kitchen: lots of appliances, easy cooking
- ▶ Download both: <https://posit.co/download/rstudio-desktop/>



```
R Console
File Edit Misc Packages Windows Help

R version 2.11.1 (2010-05-31)
Copyright (C) 2010 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

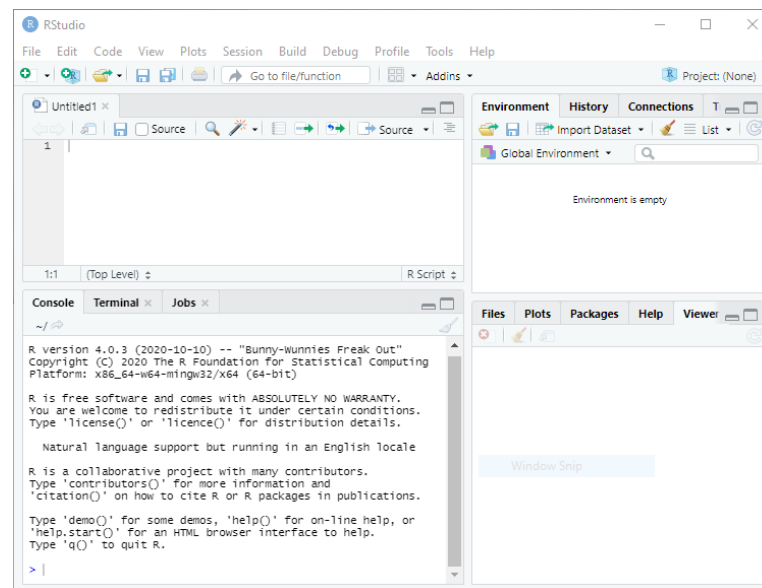
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```



Packages

- ▶ Of **many** user-written packages that extend base R's capabilities, we will use three:
 - ▶ `tidyverse` (a suite of packages) for basic tasks and graphing
 - ▶ `fixest` for regression estimation
 - ▶ `marginaleffects` for post-estimation tools
- ▶ To download a package, type:
 - ▶ `install.packages('packageName')`
- ▶ To load the package, type:
 - ▶ `library(packageName)`
- ▶ You only need to install each package once, but you must load it every time you open R or RStudio



Operators

- ▶ Operators perform operations on variables and values
- ▶ Arithmetic operators: `+` `-` `*` `/` `^`
- ▶ Comparison operators: `==` `!=` `>` `<` `>=` `<=`
- ▶ Logical operators: `&` `|`
- ▶ Assignment operators: `<-` `=`
 - ▶ Assign values to objects
 - ▶ They are the same, don't be confused when used interchangeably
- ▶ Pipe operators: `%>%` `|>`
 - ▶ String together sequences of operations: 'and then'
 - ▶ Original is `%>%` from `tidyverse`, `|>` is new addition to base R
 - ▶ Very similar, but we will mostly use `%>%`



Data Frames and Variables

- ▶ Unlike Stata, R has always been able to store multiple datasets (“data frames”) in memory simultaneously
- ▶ You need to specify the data frame when you ask R to perform calculations on a variable (or set of variables)
- ▶ The **\$** operator is the standard way
 - ▶ `mean(census$educ)` estimates the mean of the variable `educ` from the data frame `census`
- ▶ For tidyverse functions, pipes or the 1st argument do it
 - ▶ `census %>% summarize(mean(educ))` or `summarize(census, mean(educ))`
- ▶ Missing values? Change to `mean(educ, na.rm=TRUE)`



Tidyverse Functions

▶ We will use `tidyverse` functions to ‘wrangle’ data

▶ A few that we will use often:

▶ **Modifying data:**

`arrange()` orders observations by the variable(s) inside the parentheses

`filter()` subsets the data to observations that satisfy the statement inside the parentheses

`mutate()` creates, modifies, or deletes variables

`select()` keeps the variables inside the parentheses

▶ **Grouping data:**

`group_by()` groups the data by the variable(s) inside the parentheses

`summarize()` summarizes the data in a group → useful w/ `mean()`, `sd()`, `sum()`

`n()` gives the number of observations in a group

▶ **Evaluating conditional statements:**

`if_else()` evaluates truth of statement in parentheses → useful to create binary variables

`case_when()` is like `if_else()` but with multiple categories



Guidelines for Programming in R

- ▶ I will do most programming instruction by writing/running R scripts in class, but here are a few basic principles:
 1. Write code in a script (*.R) or Markdown (*.Rmd) file
 - ▶ R scripts are just code, do not automatically save output
 - ▶ R Markdown files save code, prose, and output to html or PDF
 2. Keep track of your working directory
 - ▶ If you are using or saving files locally, set a working directory
 - ▶ `getwd()` tells the current directory, `setwd()` sets a new one
 3. Annotate, annotate, annotate
 - ▶ Write comments to explain each step of your code
 - ▶ The `#` symbol starts a comment

