# Getting Started with the Capstone Project

ECON 490

**Email:** tmackay@fullerton.edu

# Overview

In these slides, we'll talk about:

- Requirements for the capstone project
- Basic components of your projects

# Key Components of Your Capstone Papers

Core components of your capstone projects:

1. Clear research question and motivation
2. Literature review
3. Data analysis

We'll complete several intermediate assignments throughout the semester

- First step is the Research and Data Proposal
- Capstone Outline will include a literature review and draft data analysis

# Final Deliverables

Project overall is worth 40% of your final grade

- Presentations during Week 15 are worth 10% of final grade
- Final paper (due at the end of Finals Week) is worth 15% of final grade

Final papers should be roughly 12-14 pages

- ***Including*** data analysis output and works cited
- Presentations should be approximately 15 minutes long

We'll talk more about writing and presenting later in the semester

- Data analysis is the most important component, so we'll focus on that first

# Capstone Research and Data Proposal

First capstone assignment is due Week 8 (10/15)

Steps to complete this assignment:

1) Decide if you want to work alone or with a partner
2) Identify a clear research question
3) Identify the data you'll use to answer your research question
4) Complete the proposal worksheet on Canvas

***NOTE:*** *Before moving forward with your project, you will need approval on the topic and data source(s) – I may ask you to submit a revised proposal.*

# Working Backwards

Trying to find a research question can be daunting

To make things easier, try ***working backwards***

- Start thinking about the general topics you're interested in – this could be labor markets, housing, education, poverty, etc.
- Then ask, "What kind of data analysis can I do?" related to that topic

Your data analysis requires (1) data and (2) an analysis strategy

- Start by identifying and exploring the data set that best fits your interest
- Then, think about (2) possible analysis strategies
- Finally, return to research question and orient around your analysis

# A Bit of Background

First semester of 490, everyone had to find their own data (some folks struggled)

- Second semester, I had everyone pick from 4-5 data sets as a starting point
- This felt overly restrictive

Starting last semester, I'm trying to find a happy medium

- You are welcome to use Canvas data sets as the starting point for your project
- However, you are ***not*** required to use any of these data sets

**KEY REQUIREMENT:** If you want to use your own data, you ***need*** to have a clean version of that data available by Week 8 and ***submit it with your proposal***

# Canvas Data Sets for Your Capstone Project

Data sets on Canvas have been selected as easy & flexible project starting points

First step in looking for data – review files listed under “Primary Data Sets”

- “Supplemental Data Sets” offers additional data to expand the range of possible projects and topics you can explore
- The supplemental files can be easily merged with the primary data sets

Several suggestions (more on Canvas data page):

- Individual-level CPS data is the most flexible = easiest to do analysis with
- State- and county-level data are best suited for merging with other data sets
- Interested in macro-style topics? Use the longer-run state level data set

# Two Choices for Your Capstone Data Analysis

As part of your data analysis, you will need to either:

1. Find and analyze data you've collected yourself (i.e., not something I provided)
2. Use an advanced analysis strategy (more on this later)

There is a list of possible secondary data sources on Canvas

- Using them requires downloading the data from the source and cleaning it
- Using any of these data sets would satisfy option (1) above

# Analysis Strategies

In general, I'll refer to analysis strategies as ways of exploring your outcome and explanatory variable relationship

What should I do for my analysis?

- Depends on specific research question
- We'll talk about specific suggestions and examples next week

Suggestion – wait until ***after*** next lecture to think about specific analysis strategies

# Advanced Analysis Strategies

This option entails more focus on analysis and relatively less on data processing:

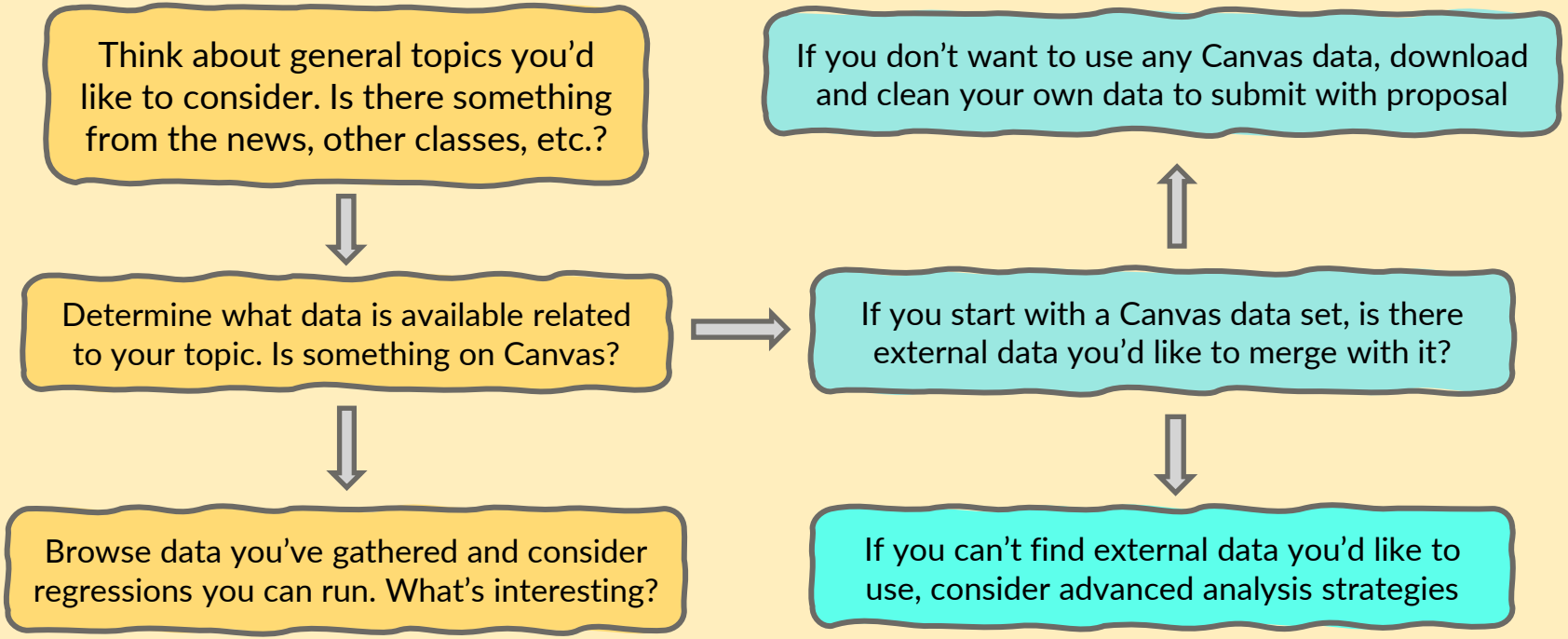
- Requires producing an additional stage of results (i.e., another table, etc.)
- Requirements for execution and accuracy in analysis will be higher

This output should go beyond basic descriptive statistics + regression output

- Possible examples include using GIS tools to generate maps, using non-linear regression methods, conducting causal inference, etc.
- If you have some background in stats / ML, can use prediction models, etc.

Note that the exact requirements will depend on your project

# Proposal Flowchart



# Capstone Data Analysis Glossary

Doing your capstone data analysis requires:

1. Finding some data and cleaning it up to create your ***working data set***
2. Using your working data to generate your ***data analysis output*** (graphs, regressions, etc.)

You'll use two broad types of data to create your working data set:

- ***Canvas data sets*** are data sets directly available from Canvas that I have cleaned
    - This includes both the primary and supplemental data sets
- ***External data sets*** are any data that you find that I did not give to you already cleaned up
    - This includes both data you found yourself and data you found using the links posted under "External Data Sets" on the Capstone Data Resources page

# General Suggestions on Research Questions

***Don't*** start thinking about research questions before looking at data

- See the "Working Backwards" slide from earlier
- Instead, think about a ***general topic*** of interest – labor markets, crime, etc.

Your first draft of a research question will likely be too superficial

- Remember, the goal is to write a 12+ page paper
- As part of this, you'll spend 2-3 pages talking about your data analysis output

*If your research question is something like, "What's the correlation between GDP and unemployment?" you're not going to have much to say!*

# Selection of Research Questions from Last Semester

*How do recessions impact household food expenditures?*

*How do the returns to graduate education vary across demographic groups?*

*How did the Covid-19 pandemic impact household transportation choices?*

*What factors drive variation in home prices across counties in California?*

*How much of NBA player compensation can be explained by future expected performance (compared to past performance or prior accolades)?*

# Thinking Descriptively

How to go beyond, “What’s the correlation between X and Y?”

- We’ll talk about specific analysis strategies next week
- Here, some conceptual suggestions and a simple example

In general, think about ***exploring variation*** in either:

1. The level of one economic variable as a function of other variables
2. A relationship between economic variables

In both cases, you can explore variation across time, places, economic factors, etc.

# Canvas Proposal Example

Suppose you're interested in migration – why do people live where they live?

From the data sets on Canvas, CPS data is best-suited for this question

- Individual-level data on migration choices – *Did you move last year? From where?*
- Explore variation across demographic characteristics – *Who's most likely to move?*

Possible analysis strategies

- Baseline regression – relationship between migration and education + income
- Exploring variation – how does this relationship differ across regions of the US? How did the Pandemic affect migration?

# An Example Using Data Sets on Canvas

Suppose we're interested in the relationship between unemployment and crime

- We can estimate this general relationship using our state-level data…
- How do we explore things further?

One option – see how this relationship differs across regions of the country

- Is crime more responsive to economic conditions in certain places?

***NOTE:*** *I picked this example to set up some practice with R – for your capstone analysis, you'll want to be able to "tell a story" about your results!*