Intro to Data Science

What R we doing?

Prof. Weldzius

Villanova University

Slides Updated: 2024-08-26

Agenda

1. Meet the instructor
2. Course Motivation
    - What is data science (DS) & why should we care?
3. Course Objectives
    - **Content:** Critical thinking, analysis, presentation
    - **Skills:** Computing and analysis in R
4. ChatGPT and data science
5. Course Expectations & Syllabus review

Meet the instructor

- Education
  - PhD from UCLA Political Science in 2018
  - Postdocs at WashU and Princeton Niehaus
- Published some things
  - Substantive: global value chains & currency manipulation; European politics; economics of land use; book review on currency statecraft
  - Policy: 2014 World Trade Report; polarization & lobbying
- Work
  - World Trade Organization
  - Commercial Real Estate (2007/8 global financial crisis)

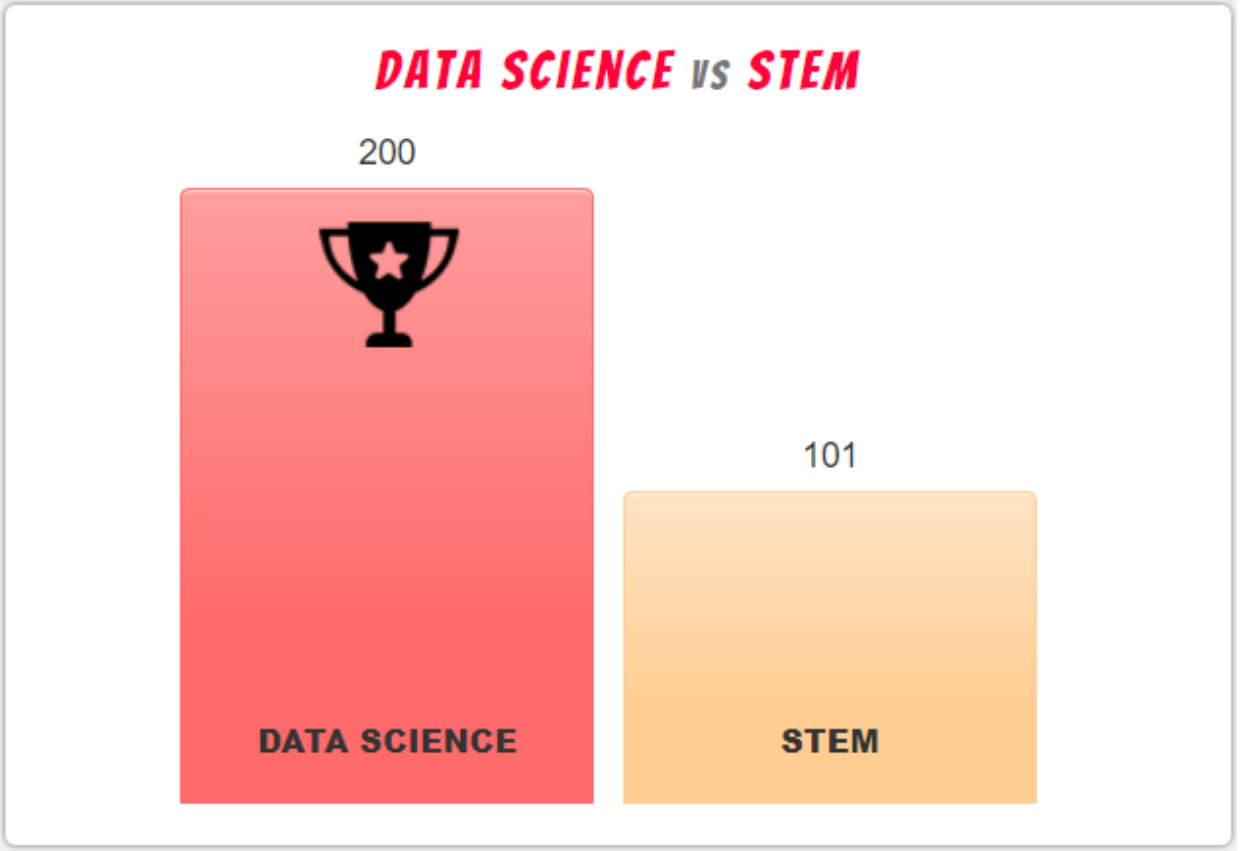
Meet the instructor

- Current research
  - Global value chains + currency manipulation
  - Financial crises + polarization
  - Trade agreements + monopolies
  - Inequality + economic growth

Why are you here?

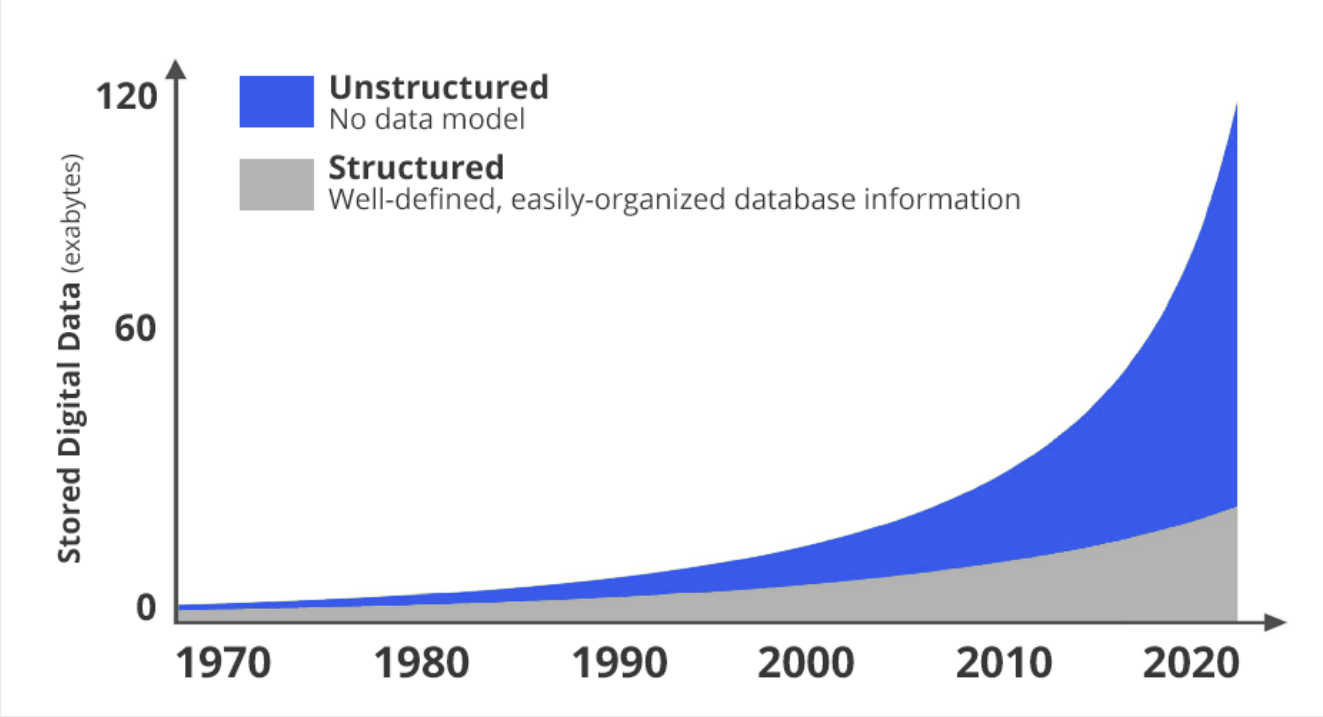
Suggested fights

20 last fights



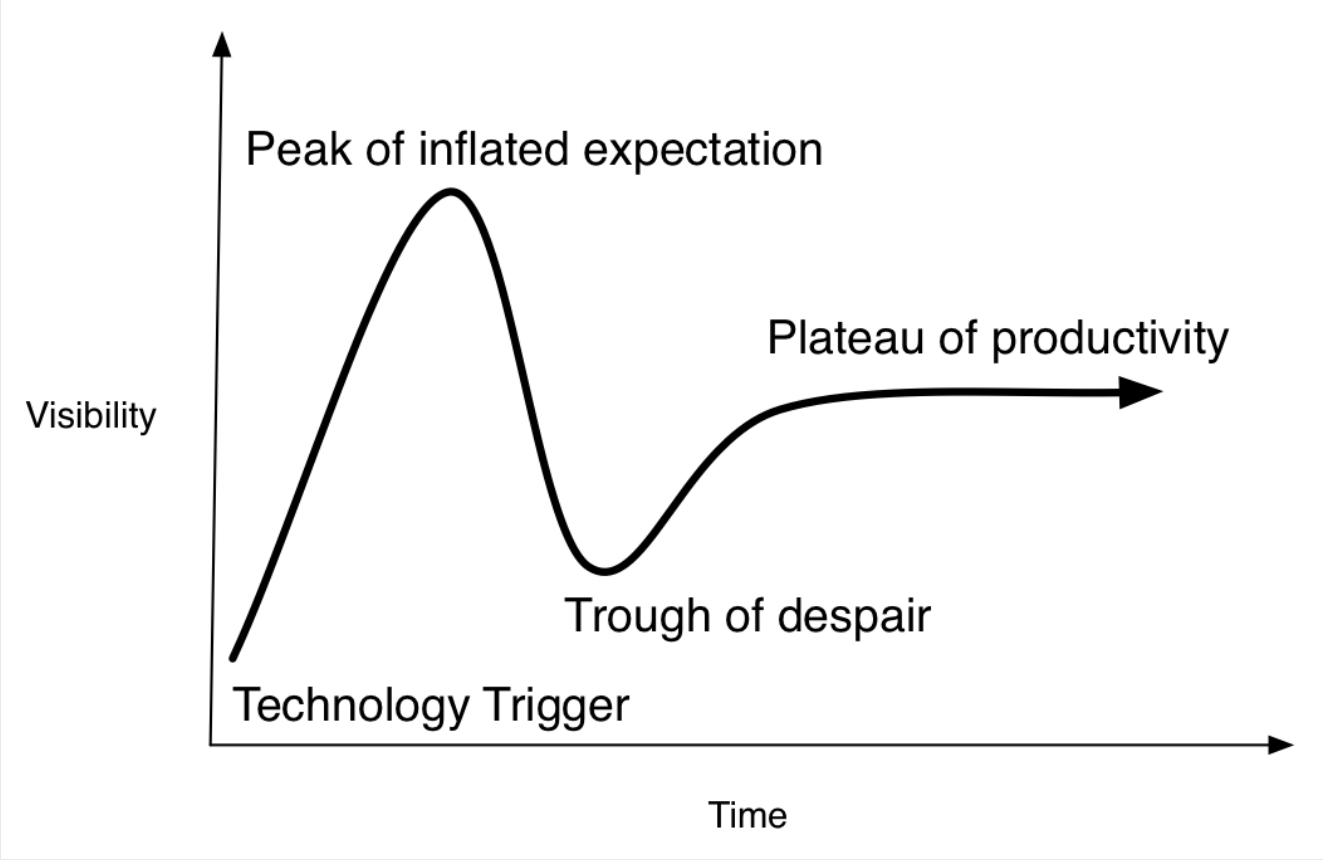
Is this all just a fad?

- No



Is this all just a fad?

- But there are faddish qualities



So what IS data science?

- Split into two camps

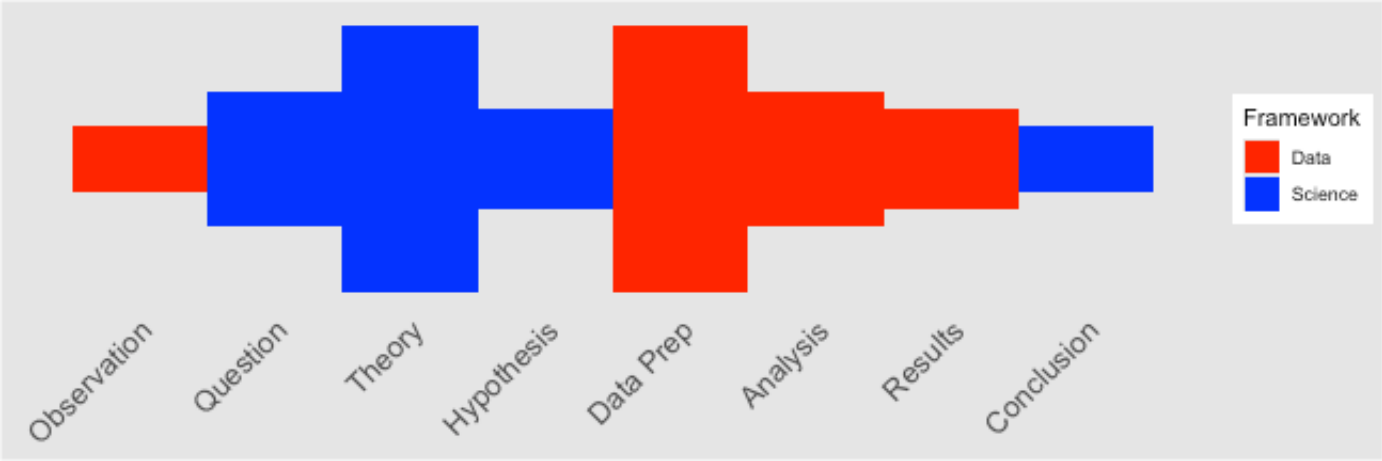
1. Research camp
    - Focused on **answering a research question**
    - Follows the "scientific method"
    - Goal: contribute to knowledge
    - Domain: academia
2. Prediction camp
    - Focused on **making a prediction**
    - Typically unconcerned with theory or *why* a model works
    - Goal: inform a decision / policy
    - Domain: private sector

The Two Camps

The Two Camps

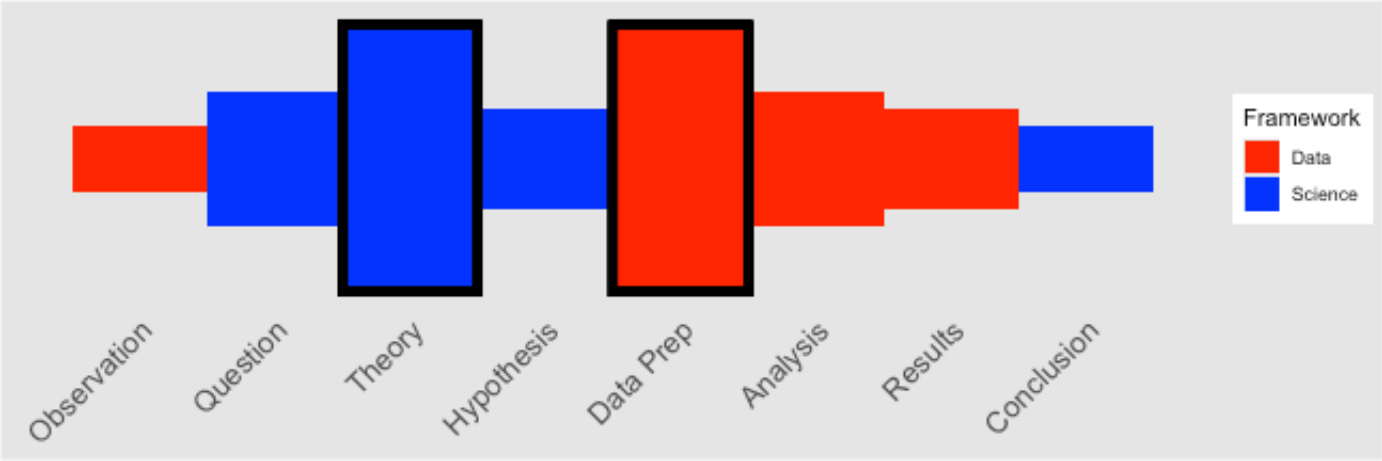
Research Camp

- The scientific method
  1. Observation → Question
  2. Theory → Hypothesis
  3. Data Collection / Wrangling → Analysis
  4. Results → Conclusion



Research Camp

- The scientific method
  1. Observation → Question
  2. Theory → Hypothesis
  3. Data Collection / Wrangling → Analysis
  4. Results → Conclusion



Research Camp

Received: 5 September 2019 | Revised: 18 October 2020 | Accepted: 1 December 2020

DOI: 10.1111/ecpo.12184

ORIGINAL ARTICLE

ECONOMICS & POLITICS WILEY

Global Value Chains

The end of currency manipulation? Global production networks and exchange rate outcomes

Ryan M. Weldzius

Department of Political Science, Villanova University, Villanova, PA, USA

Correspondence
Ryan M. Weldzius, Department of Political Science, Villanova University, Villanova, PA, USA.
ryan.weldzius@villanova.edu

Funding information
National University of Singapore; Washington University in St. Louis; Niehaus Center for Globalization and Governance, Princeton University; Villanova University

Abstract

Between 2000 and 2017, eight major exporting countries engaged in currency manipulation in order to increase their trade surpluses with the rest of the world. As of 2018, however, no currency manipulators remained. This change in policy is puzzling given the past successes of this export-led growth model. I argue that the state-level decision to stop depreciating its exchange rate stems from the reduced benefits and increased costs of currency manipulation. As production becomes more global, an increase in tradable inputs decreases the traditional benefits of a depreci-

Research Camp

1. **Observation** → **Question**
    - Observation is facilitated by data (Descriptive analysis)

Research Camp

1. **Observation** → **Question**
    - Observation is facilitated by data (Descriptive analysis)

Research Camp

1. **Observation** → **Question**
    - Observation is facilitated by data (Descriptive analysis)
    - Number of currency manipulators per year according to Bergsten-Gagnon criteria

Research Camp

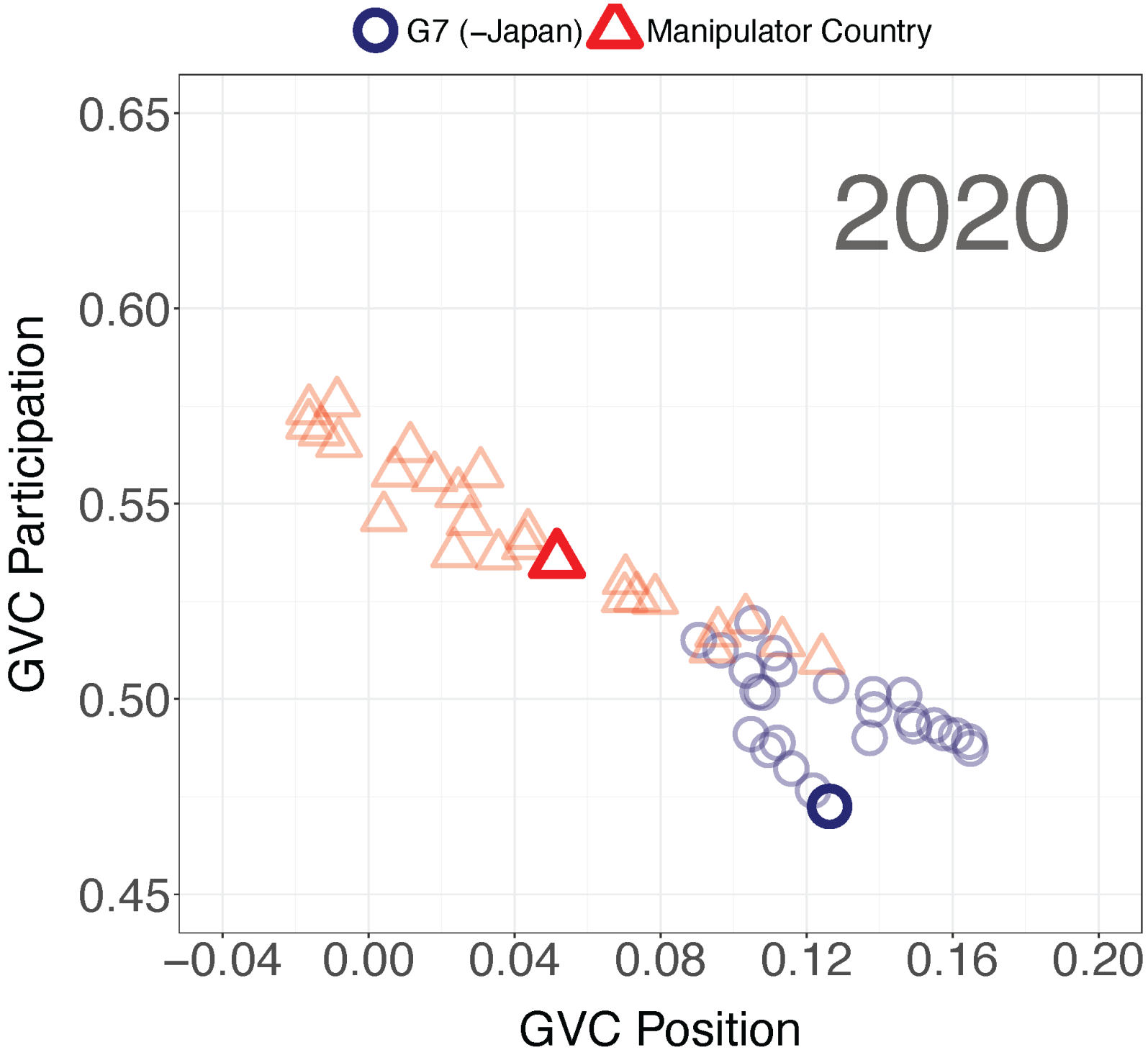
1. **Observation** → **Question**
    - The question pertains to science
    - I.e., why have countries stopped (or attenuated) currency manipulation?

Research Camp

1. **Theory** → **Hypothesis**
    - Theorizing requires abstraction & simplification
    - I.e., trade has changed and therefore exchange rate preferences have changed
    - Global value chains now account for 70% of global trade (~50% in 1990s)
    - Hypotheses fall out naturally from well-done theory
    - **H1**: *global value chain trade should moderate the relationship between export dependence and currency manipulation*

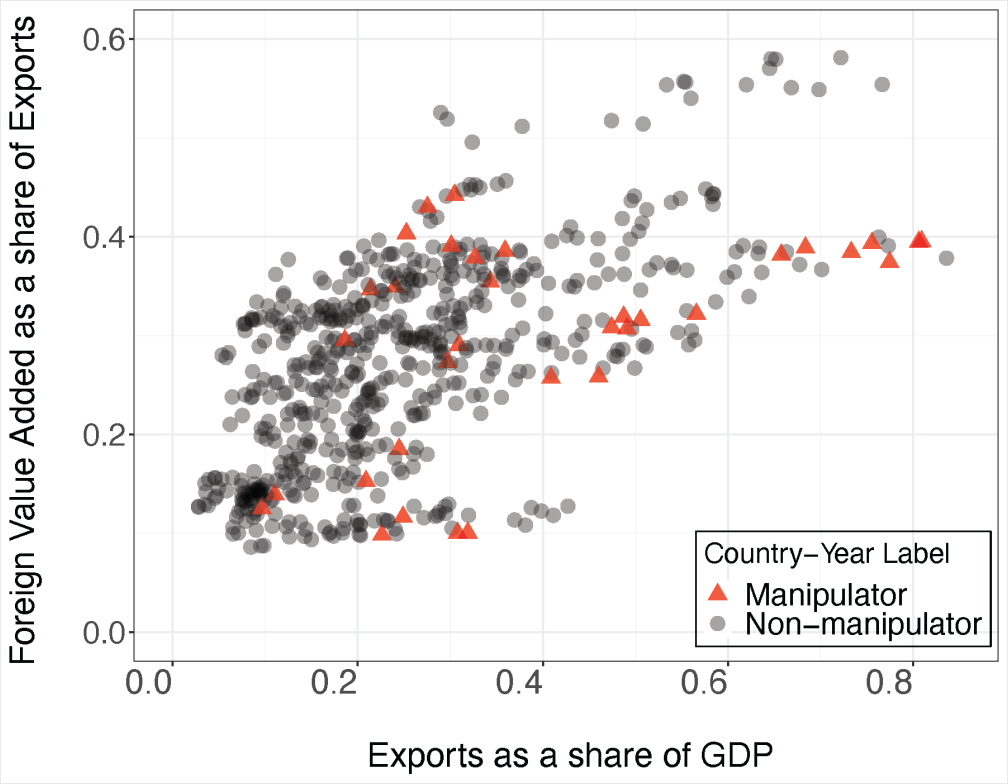
Research Camp

3. **Data Collection / Wrangling** → **Analysis**
    - Data collection separates "Data Science"...
    - ...from "Science, with data"
    - collect national input-output tables (matrices) & estimate global value chain linkages


G7 (–Japan)
Manipulator Country
2020
GVC Participation
0.65
0.60
0.55
0.50
0.45
–0.04
0.00
0.04
0.08
0.12
0.16
0.20
GVC Position

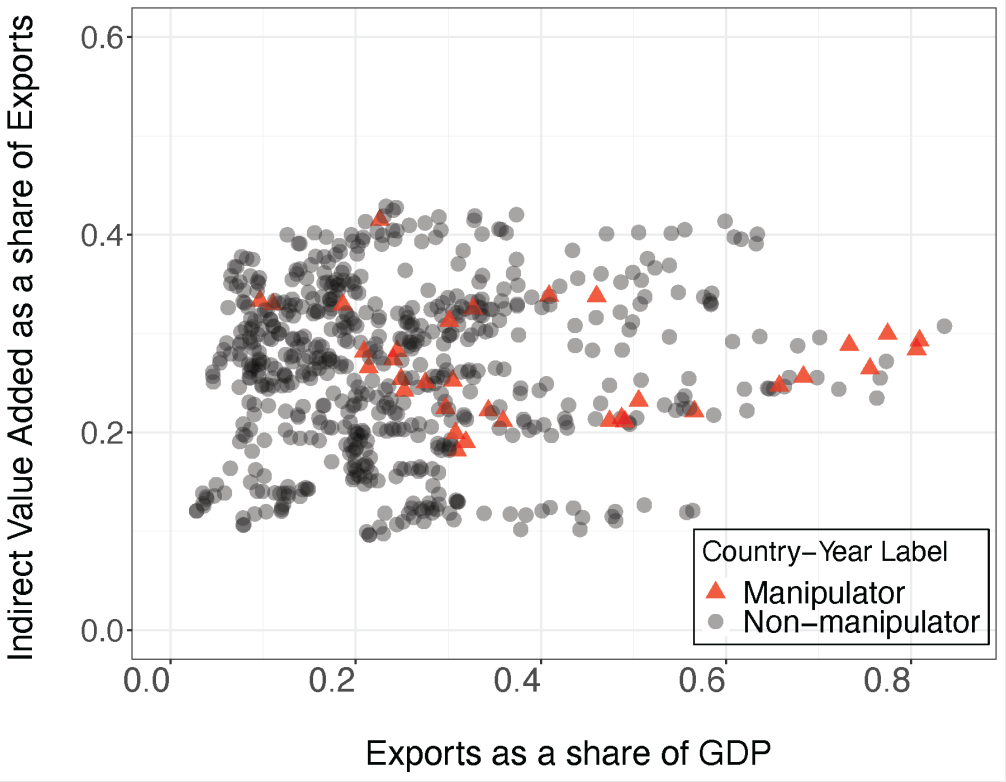
Research Camp

3. **Data Collection / Wrangling** → **Analysis**
    - Analysis is informed by the data you have collected...
    - ...and the hypotheses you have generated



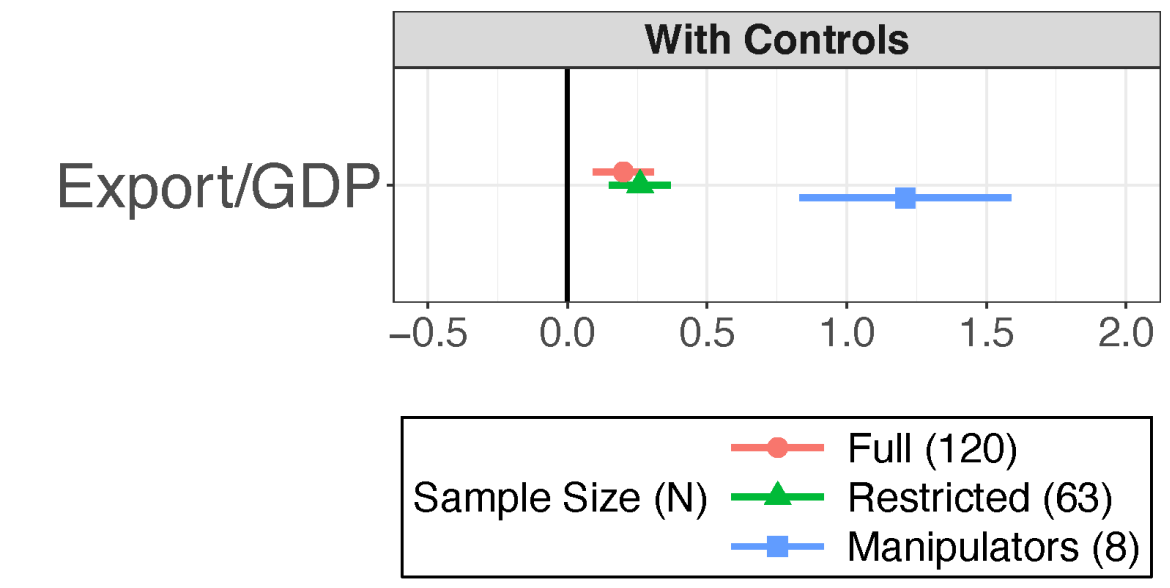
Research Camp

3. **Data Collection / Wrangling** → **Analysis**
    - Analysis is informed by the data you have collected...
    - ...and the hypotheses you have generated



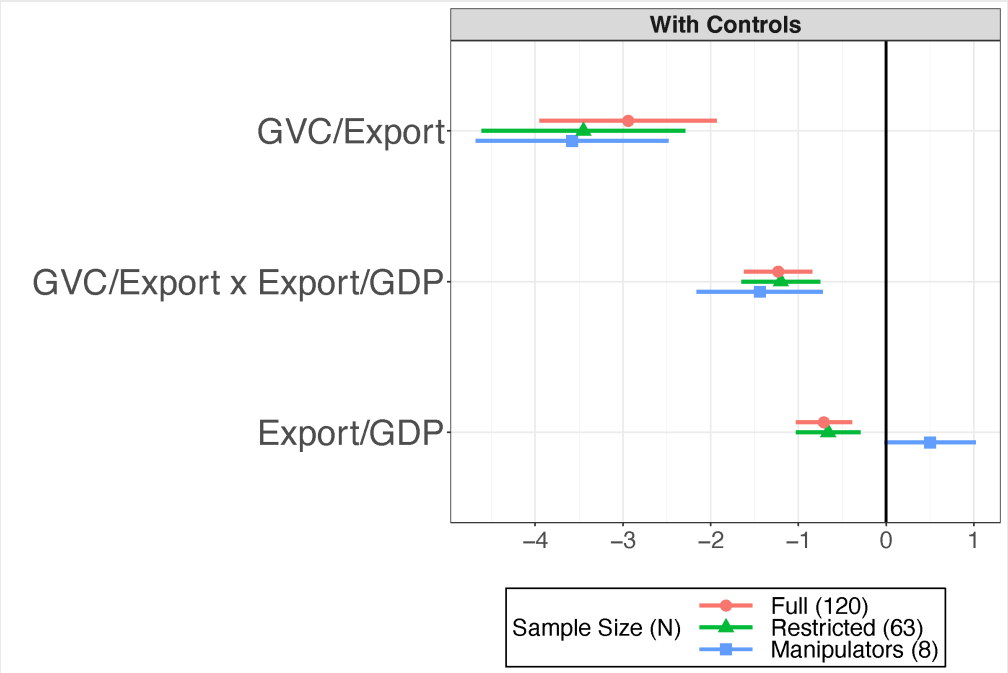
Research Camp

4. **Results** → **Conclusion**
   - Results fall out naturally from the analysis...
   - ...and must be interpreted in terms of the theory and hypotheses...
   - ...to draw conclusions



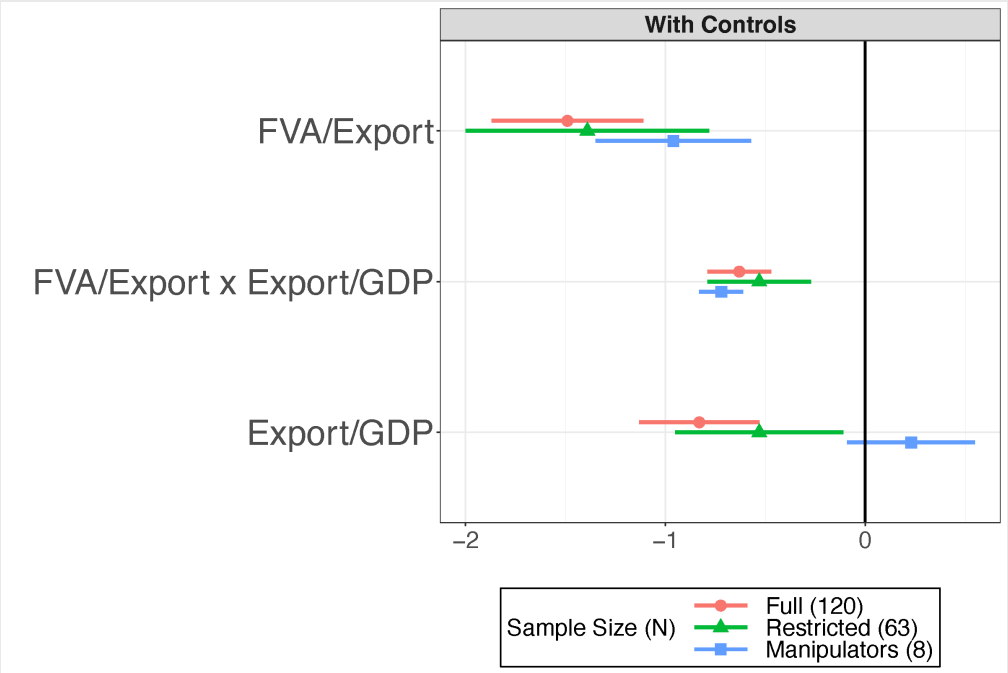
Research Camp

4. **Results** → **Conclusion**
   - Results fall out naturally from the analysis…
   - …and must be interpreted in terms of the theory and hypotheses…
   - …to draw conclusions



Research Camp

4. **Results** → **Conclusion**
   - Results fall out naturally from the analysis…
   - …and must be interpreted in terms of the theory and hypotheses…
   - …to draw conclusions



The Two Camps

Course Objectives

- This course is the menu, not the food
  - Look over many different fields, methods, and tools
  - You pick those you like, and take more advanced classes to dig into them
- But we are very **hands on**
  - You must download `R` and `RStudio` prior to next class (Problem Set 0)
  - You must work through first HW using an `.Rmd` file

Learning goals

1. Generate a sophisticated research question based on clearly described assumptions and a narrowly defined hypothesis.
2. Describe the data used to investigate this research question, including univariate and multivariate visualizations and summary statistics.
3. Apply the appropriate methods to answer the research question and evaluate the hypothesis.
4. Acknowledge limitations of method and results, and describe a superior empirical setting that would overcome these limitations.

ChatGPT in the classroom

- Are we at the precipice of a new era in human-computer relations?
  - ChatGPT can achieve these learning goals!
  - But it needs to be used wisely...it is still a tool
- It can make coding (the hardest part of this class) easier
- But it can also prevent you from learning

AI in the labor market

- McKinsey told AT&T in 1980 that, by 2000, cell phones would be a niche market of 900,000 subscribers
- Is AI-assisted work is the future?
  - Profound gains in productivity already
- Will this be like automation and globalization for US manufacturing?
  - What skills will be valuable in 5 years? 10 years?

AI in the labor market

- My answer: prepare you for both possibilities
  - If AI is a "fad", make sure you can do this work unassisted
  - If AI is the new normal, make sure you can work with it productively
- The one thing you **shouldn't** do
  - Take shortcuts / cheat
- You will still have an interview in which you are asked something like the following: "How is overfitting different from underfitting, and why should we care?"
  - **You** need to know this answer

Grades

Item	Percent	Points	EC	Max
pset1	5%	5	1	6
pset2	5%	5	1	6
pset3	5%	5	1	6
pset4	5%	5	1	6
pset5	5%	5	1	6
pset6	5%	5	1	6
pset7	5%	5	1	6
pset8	0%	0	6	6
Quizzes	15%	15	4	19
Midterm exam	20%	20	2	22
Final project	10%	10	2	12
Final exam	20%	20	2	22
Totals	**100%**	**100**	**25**	**125**

Grades: PSets

- 9 in total, only 8 are graded (one freebie)
  - Pset 0 doesn't count
  - 5 points each + 1 EC
- Posted to **GitHub** on Mondays at noon
- Due **Friday by midnight**
  - Submit/Upload to **Blackboard**
  - Each day late is -1 point
  - After 3 days, scored zero
- Restrictions:
  - Open book / open note / open Campuswire
  - Can collaborate but submissions must be your own

Grades: Final Project

- Final project is 10 points
    - 8 points for final submission
    - 2 points for presentation in final week (5-8 minutes)
    - 2 EC points: 1 point for each classmate you give feedback to after the presentations, max. 2 points (submit via Campuswire)
- Task: create a report that takes a dataset of your choice and uses the skills you learn in this class to tell us something about it. Use at least two skills from first half, two skills from second half. Minimum 8 pages, double-spaced, 12pt font.

Grades: Exams

- 1 in total: midterm on October 9th
- Midterm is 20 points + 2 EC
- Final is 20 points + 2 EC (may opt out of exam!)

Grades: Quizzes

- Taken at beginning of each lecture
- Each quiz is worth one point
- 19 total quizzes, only 15 count toward final grade (four freebies)
- Password protected
  - Only students in class can take them
  - 50% of quiz grade is just taking it (effectively a way to track attendance)
  - 50% of quiz grade are questions related to **homework**

Not Graded: HW

- You should work through the homeworks prior to each lecture
- Open the `.Rmd` file and Knit it
- Read the output and try and answer the prompts
- **Not graded**, but enormously helpful in preparing you to keep up with lectures

The Syllabus

- (Go to Github)

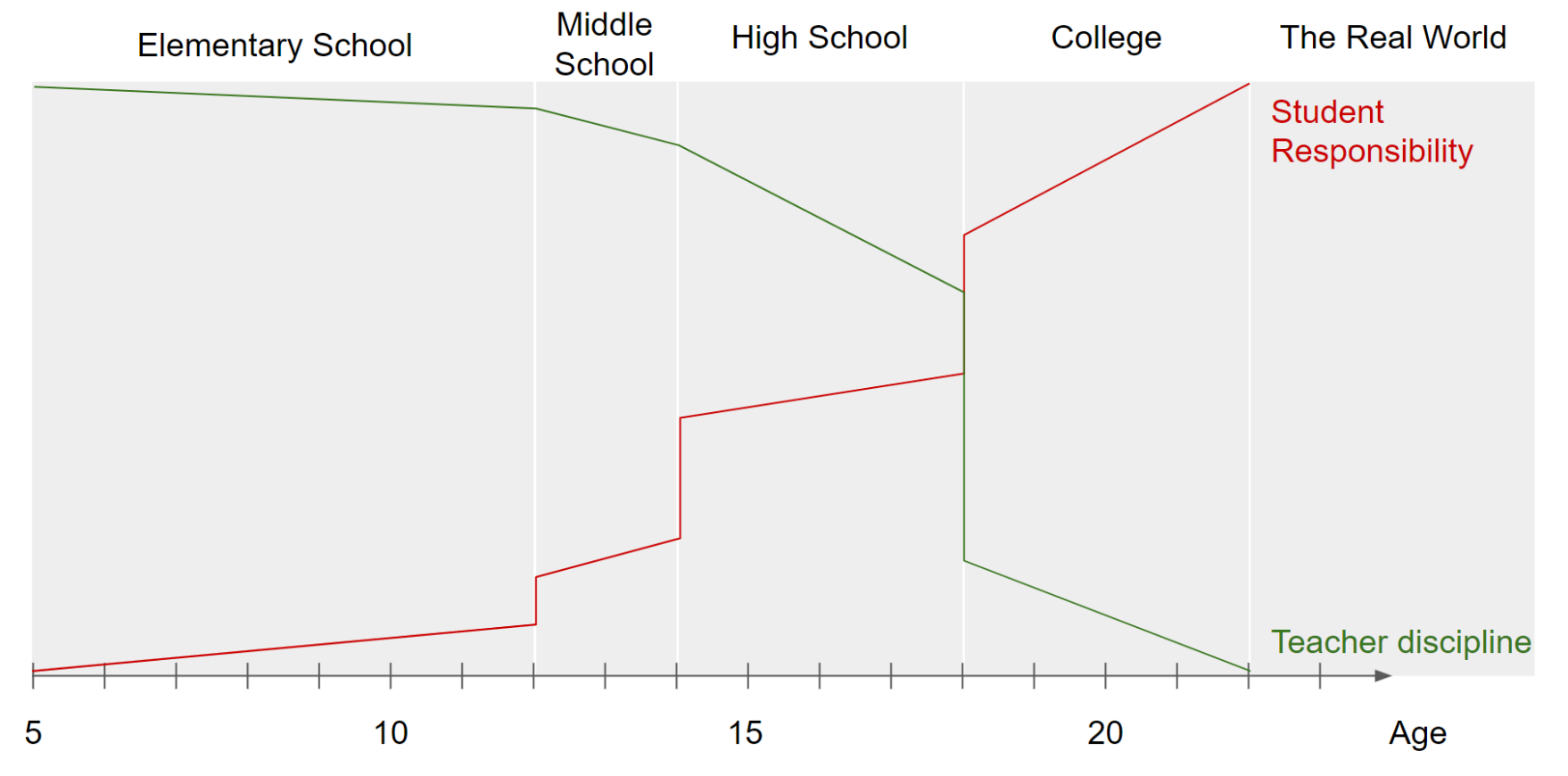
Honor Code

- Students are assumed to have read and agreed with the Villanova University Academic Integrity policy
- Violations of this policy may result in:
  - An F for the semester (at minimum)
  - Suspension for a semester
  - Expulsion
- However, except where **explicitly noted**, this course is collaborative
  - Open book, open note, open internet
  - Can rely on Campuswire for help
  - Can work together on problem sets (but must submit own work)
- **Can't collaborate on exams**

Resources

- Campuswire (place for **questions**)
  - Post questions on the class feed
- Blackboard (place for **submissions**)
  - Submit problem sets, quizzes, and exams
- GitHub (place for **materials**)
  - Find all in-class materials
- Office hours (place for **hands-on help**)

Teaching Philosophy



Teaching Philosophy

- This course is **inherently** hard
  - Learning R is challenging
- But the goal is to **encourage** you to pursue data science
- As such, the **nature** of the material is at odds with the **goal** of the class
- My solution: grade leniently
  - I.e., lots of extra credit

Conclusion

- Let's have a great semester!
- Homework:
    1. Work through psc4175_hw_1.Rmd
    2. Complete Problem Set 0 (on Blackboard)
    3. Create a Copilot account (https://platform.openai.com/signup)