

Intro to Data Science

What **R** we doing?

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/01/09

Slides Updated: 2023-01-07

Agenda

1. Meet the instructor

- Prof. Bisbee: *james.h.bisbee@vanderbilt.edu*

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. Course Expectations & Syllabus review

Meet the instructor

- PhD from NYU Politics in 2019
- Postdocs at Princeton Niehaus & NYU CSMaP
- Published some things
 - Methods-ey: external validity [1](#), [2](#); measurement [3](#), [4](#)
 - Substantive: economics & populism [1](#); Covid-19 & U.S. politics [2](#), [3](#); IPE [4](#); academic naval-gazing [5](#)
- Popular press
 - Monkey Cage articles [1](#), [2](#)
 - [Podcast](#) / Radio interviews

Meet the instructor

- Current research
 - YouTube + polarization
 - Twitter + misinformation
 - Telegram + white supremacists
- Is my current research agenda data science?

Why are you here?



Suggested fights

20 last fights



DATA SCIENCE vs STEM

200



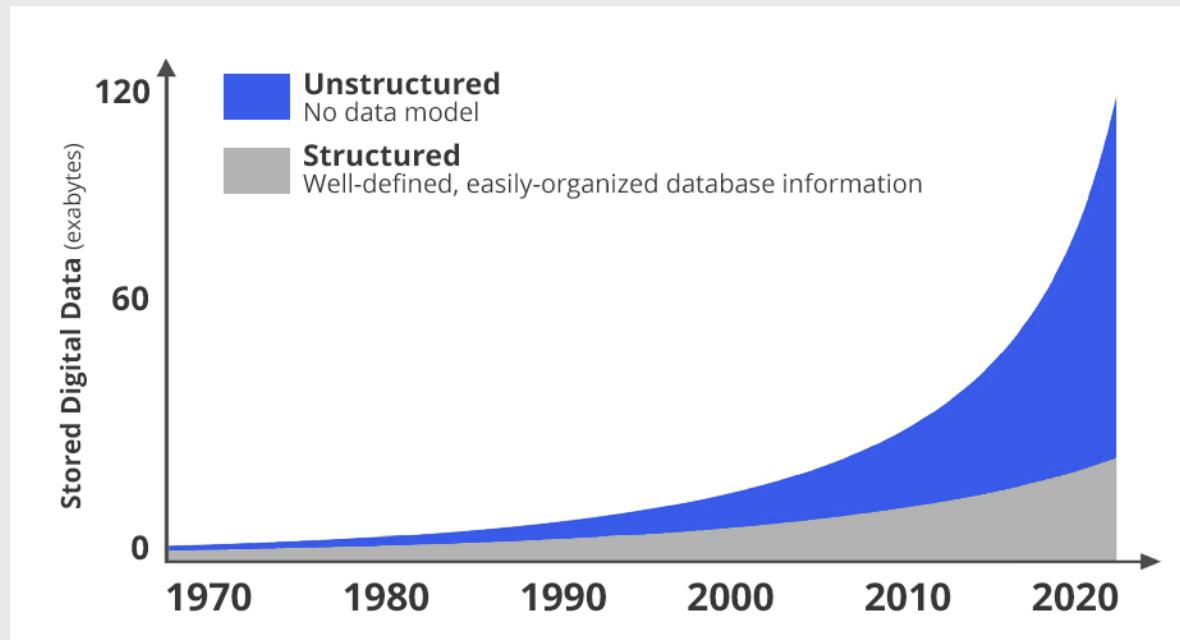
DATA SCIENCE

101

STEM

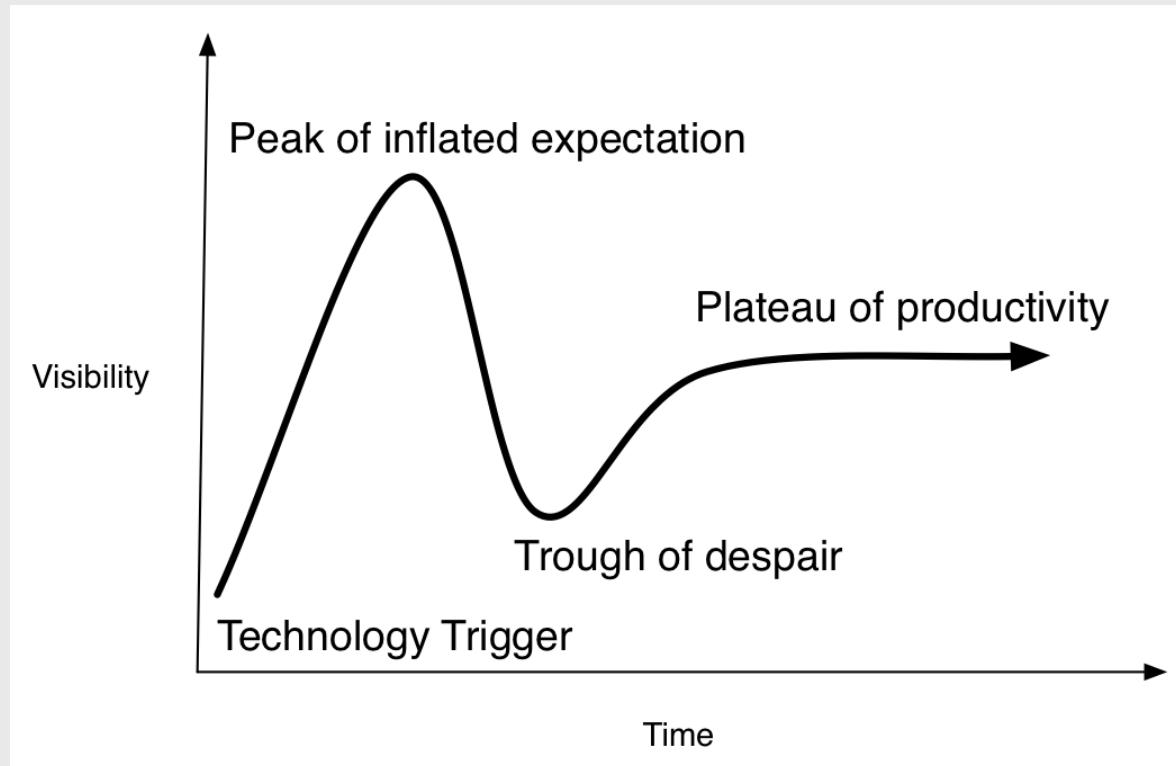
Is this all just a fad?

- 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

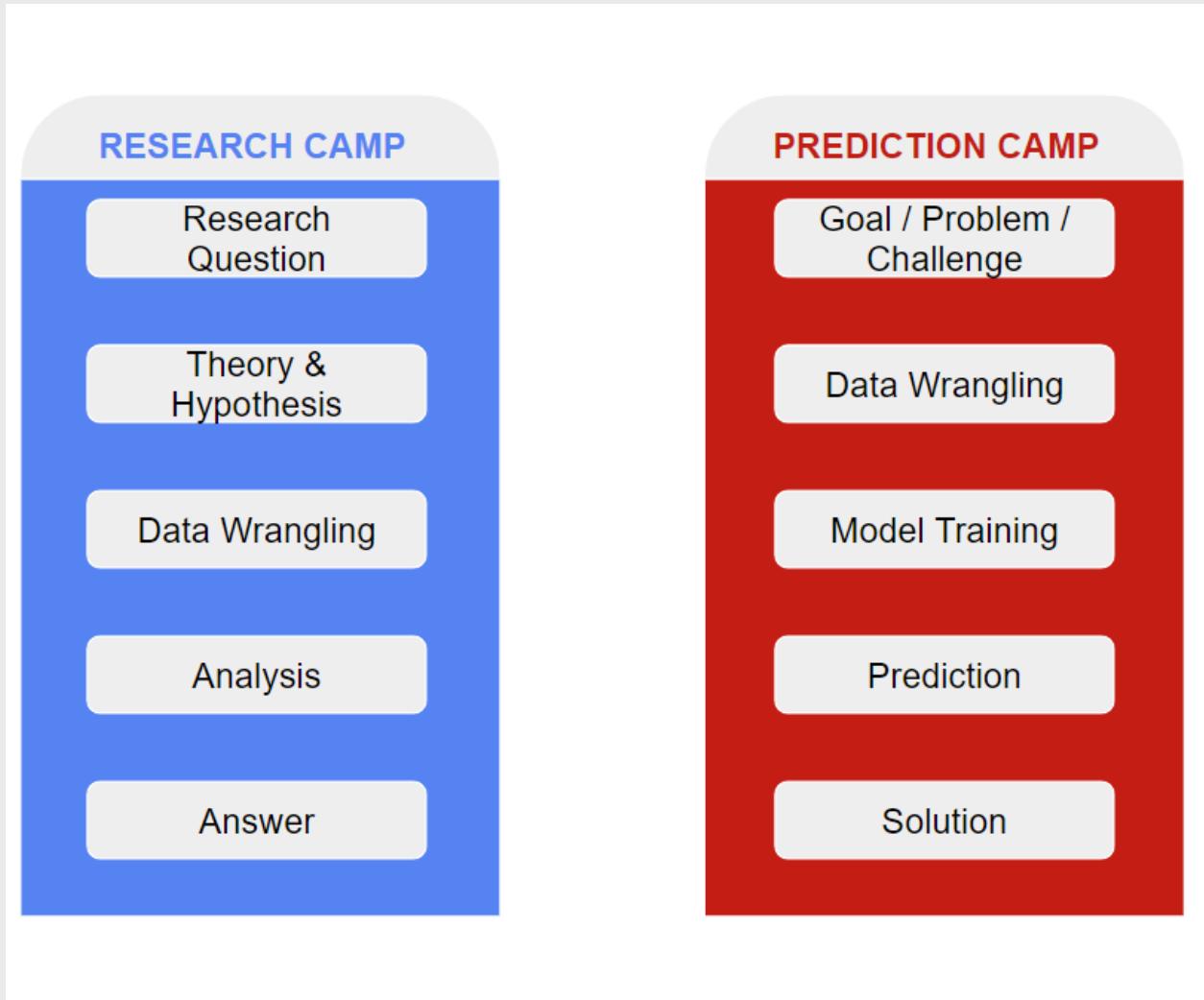
2. Prediction camp

- Focused on **making a prediction**
- Typically unconcerned with theory or *why* a model works
- Goal: inform a decision / policy
- Domain: private sector

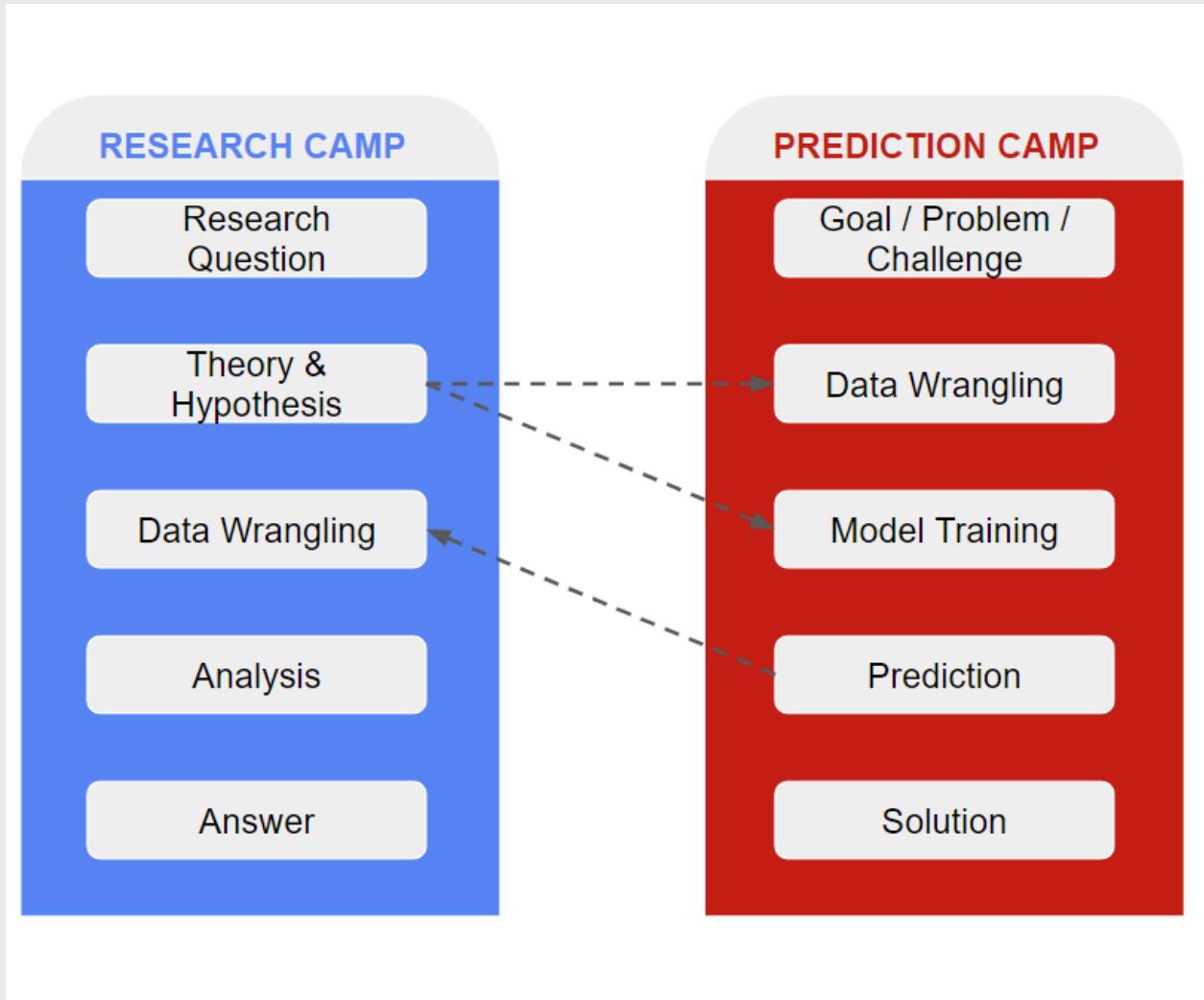
The Two Camps



The Two Camps

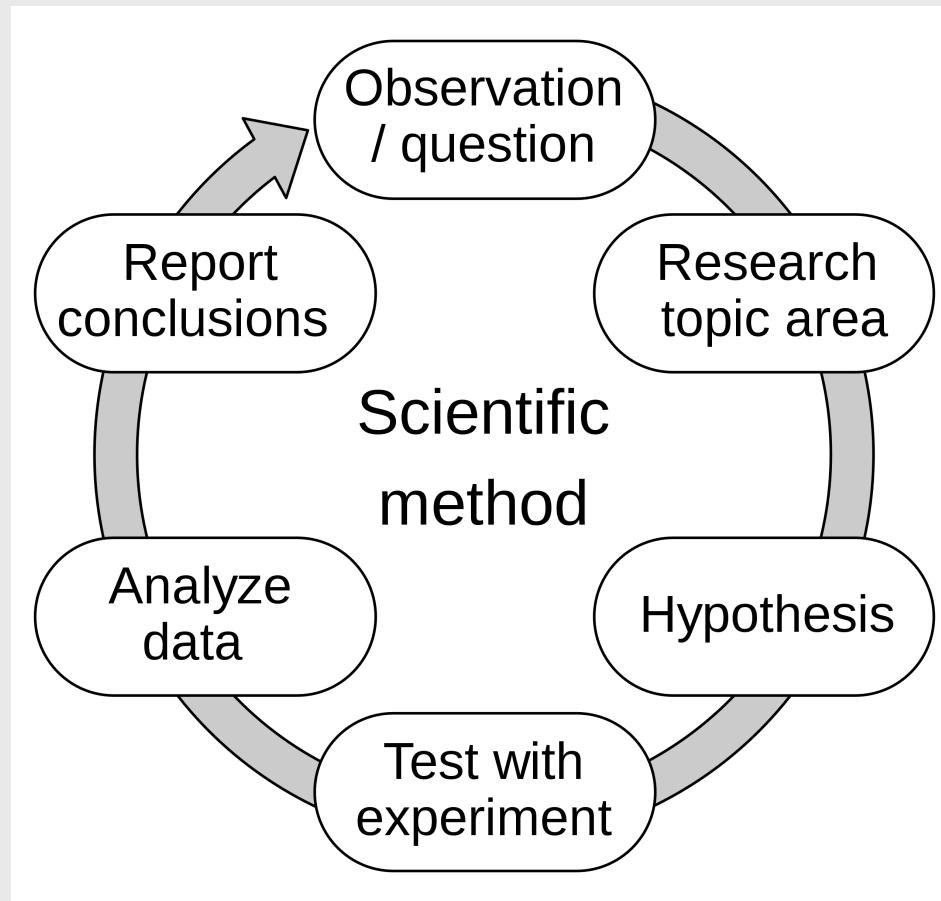


The Two Camps



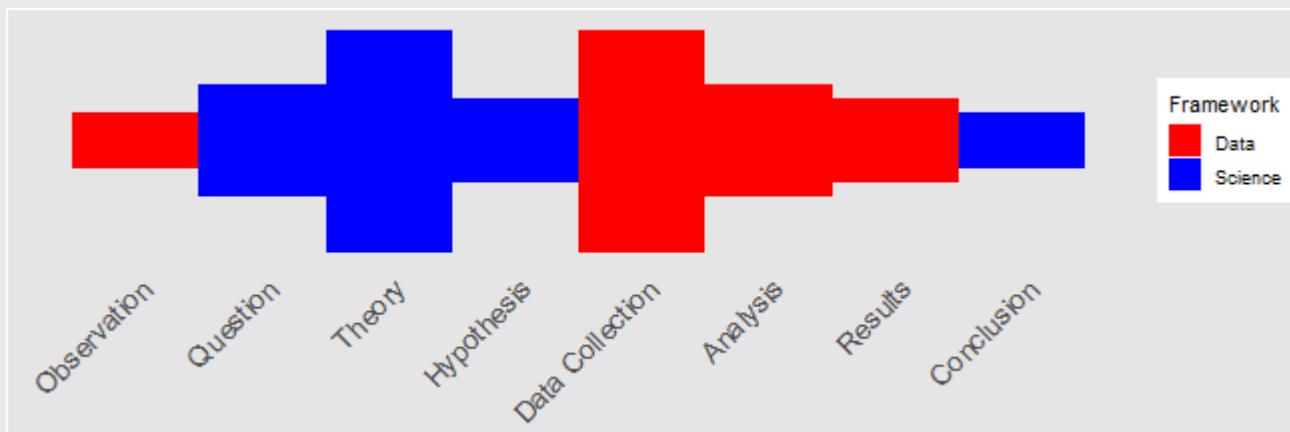
Research Camp

- The scientific method



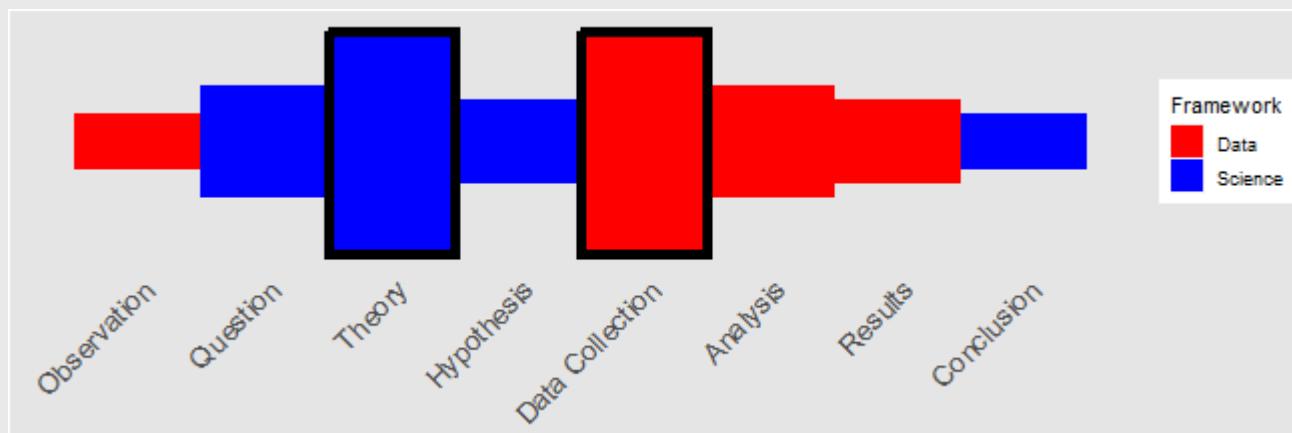
Research Camp

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



Research Camp

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



Research Camp

Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users

Megan A. Brown,^{1‡} James Bisbee,¹ Angela Lai,^{1,4}
Richard Bonneau,^{1,3,4} Jonathan Nagler,^{1,2,4} Joshua A. Tucker^{1,2,4}

¹Center for Social Media and Politics, New York University

²Politics Department, New York University

³Biology Department, New York University

⁴Center for Data Science, New York University

[‡]To whom correspondence should be addressed: meganbrown@nyu.edu

August 24, 2022

Abstract

To what extent does the YouTube recommendation algorithm push users into echo chambers, ideologically biased content, or rabbit holes? Despite growing popular concern, recent work suggests that the recommendation algorithm is not pushing users into these echo chambers. However, existing research relies heavily on the use of anonymous data collection that does not account for the personalized nature of the recommendation algorithm. We asked a sample of real users to install a browser extension that downloaded the list of videos they were recommended. We instructed these users to start on an assigned video and then click through

Research Camp

1. Observation → Question

- Observation is facilitated by **data** (Descriptive analysis)



Research Camp

1. Observation → Question

- Observation is facilitated by **data** (Descriptive analysis)

The image shows a screenshot of a CBS News video player. At the top, there are two video frames: one of a man in a suit and tie, and another of a woman in a pink top. Below them is a news ticker with the text "SENATE HEARING ON RUSSIAN INTERFERENCE IN 2016 ELECTION cbsnews.com/hearing". A red banner at the bottom of the screen reads "PRES. USES SOCIAL MEDIA TO DENOUNCE 'RIGGED' ELECTION". The CBSN logo is visible in the bottom right corner. The video player interface includes a play button, a progress bar showing 0:06 / 11:40, and a "SHOW ME" button. Below the video, there is a section titled "U.S. elections" with the note "The AP has called the Presidential race for Joe Biden. See more on Google." and a link "Robust safeguards help ensure the integrity of elections and results. Learn more". A caption at the bottom left says "Trump continues to push false claims of election fraud in Facebook video" with "12,798 views • Dec 3, 2020". On the right side of the video player, there is a sidebar titled "Up next" showing several news items from other networks like FOX, NBC, and CNN.

Research Camp

1. Observation → Question

- Observation is facilitated by **data** (Descriptive analysis)

The image shows a screenshot of a CBS News video player. The main video frame is split horizontally, with the left side showing a Black male anchor in a white shirt and tie, and the right side showing a female anchor in a pink top. Below the video, a red banner displays the headline: "PRES. USES SOCIAL MEDIA TO DENOUNCE 'RIGGED' ELECTION". The CBSN logo is visible in the bottom right corner of the video frame. At the bottom of the screen, there is a navigation bar with icons for play, volume, and time (0:06 / 11:40). A small text overlay at the bottom left says "U.S. elections". To the right of the video player is a sidebar titled "Up next" which lists several other news items from various sources like ABC News, NBC News, and CNN, each with a thumbnail, title, and view count.

PRES. USES SOCIAL MEDIA TO DENOUNCE "RIGGED" ELECTION

SENATE HEARING ON RUSSIAN INTERFERENCE IN 2016 ELECTION
cbsnews.com/hearing

LIVE CBSN

U.S. elections

Trump continues to push false claims of election fraud in Facebook video

12,798 views • Dec 3, 2020

CBS News 3.29M subscribers

President Trump posted a long Facebook video where he repeatedly denounced the November election as "rigged," even though Attorney General William Barr said the Justice Department has seen no evidence of election fraud. CBS News White House correspondent Paula Reid joins CBSN's

942 182 SHARE SAVE

SUBSCRIBE

Up next

HOW IT STARTED: Senate Hearing On FBI Investigation I...
ABC News 47K views • 3 hours ago

Attorney General William Barr's job in jeopardy
ABC News 57K views • 5 hours ago

The Full Story of Trump and COVID-19 NowThis
NowThis News 1.8M views • 1 month ago

Live: New York Gov. Andrew Cuomo Holds Briefing On Cov...
NBC News 9.2K watching

See Bernie Sanders' reaction to Trump floating 2024...
CNN 963K views • 18 hours ago

Mary Trump Says Trump's Legal Battles Could Prevent a 2024...
The View 5.5K views • 1 hour ago

Trump releases Facebook video full of false claims about...
CBS News 14K views • 4 hours ago

Election Lawsuits Meltdown... With Prejudice!
LegalEagle 995K views • 4 days ago

Second Georgia Senate election hearing
11Alive 9K watching

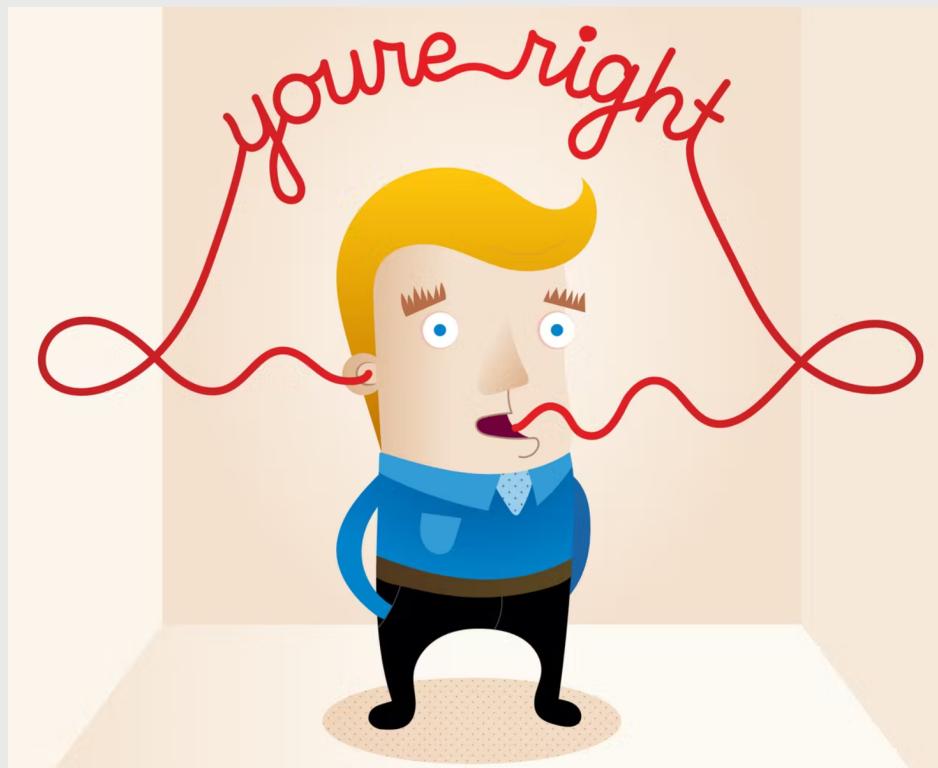
'A Fool': MAGA Fans Turn On Bar After Debunking Trump's...
MSNBC 875K views • 19 hours ago

19 / 70

Research Camp

1. **Observation** → **Question**

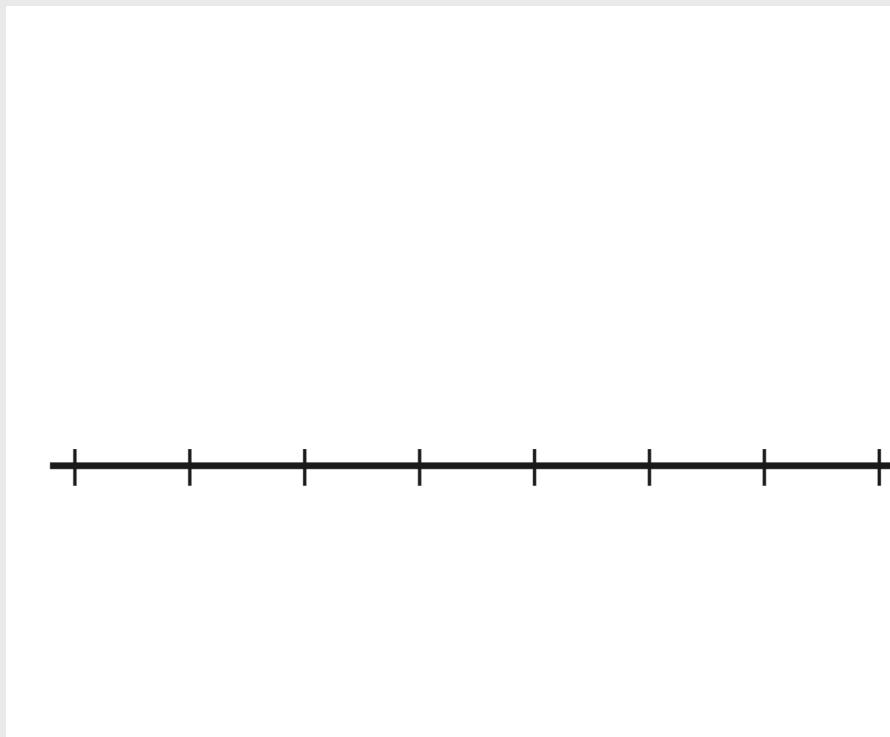
- The question pertains to science
- I.e., does YouTube's algorithm put users into "echo chambers"?



Research Camp

2. Theory → Hypothesis

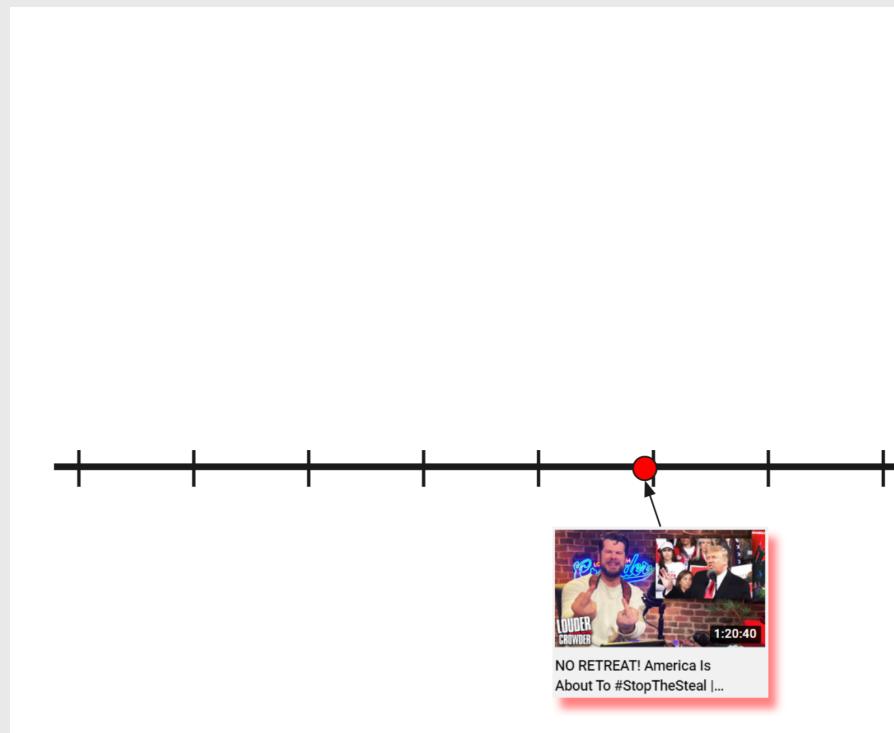
- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict



Research Camp

2. Theory → Hypothesis

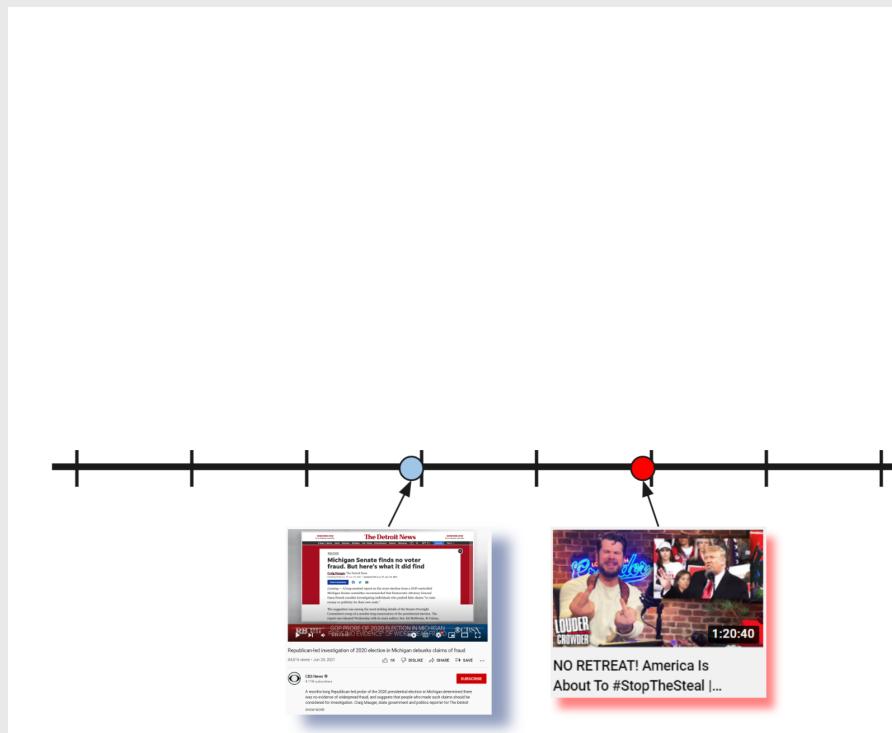
- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict



Research Camp

2. Theory → Hypothesis

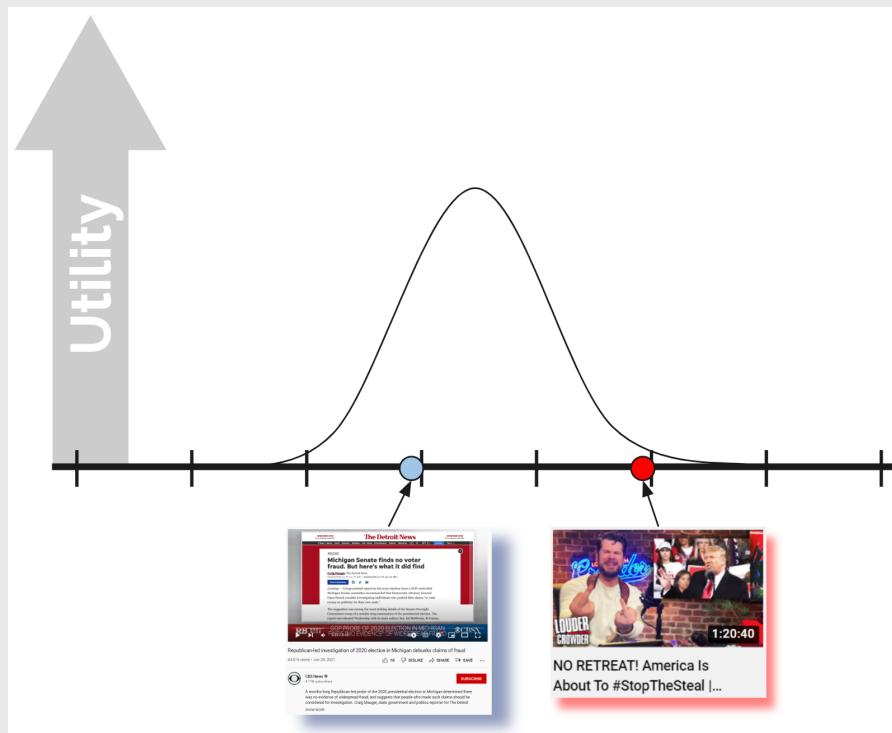
- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict



Research Camp

2. Theory → Hypothesis

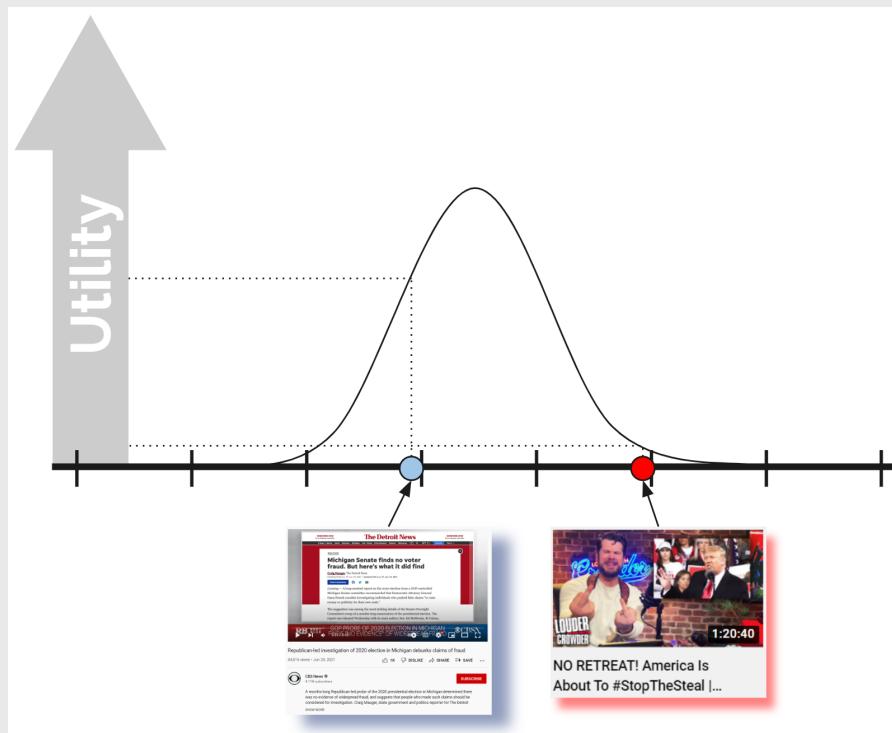
- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict



Research Camp

2. Theory → Hypothesis

- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict



Research Camp

2. Theory → Hypothesis

- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict
- YouTube wants users to watch more videos

Deep Neural Networks for YouTube Recommendations

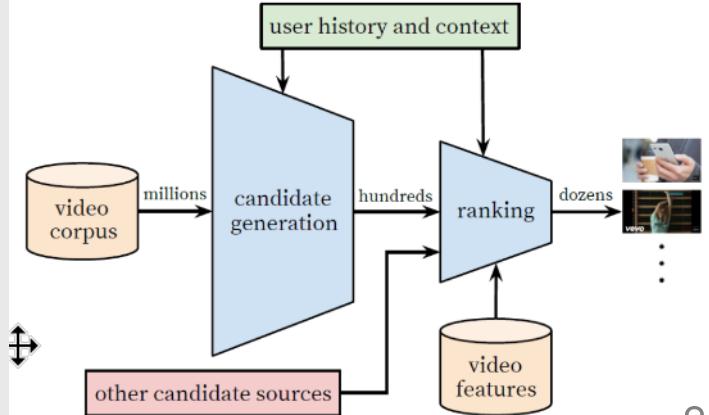
Paul Covington, Jay Adams, Emre Sargin
Google
Mountain View, CA
{pcovington,jka,msargin}@google.com

ABSTRACT
YouTube represents one of the largest scale and most sophisticated industrial recommendation systems in existence. In this paper, we describe the system at a high level and focus on the dramatic performance improvements brought by deep learning. The paper is split according to the classic two-stage information retrieval dichotomy: first, we detail a deep candidate generation model and then describe a separate deep ranking model. We also provide practical lessons and insights derived from designing, iterating and maintaining a massive recommendation system with enormous user-facing impact.

Keywords
recommender system; deep learning; scalability

1. INTRODUCTION
YouTube is the world's largest platform for creating, sharing and discovering video content. YouTube recommendations are responsible for helping more than a billion users



$$P(w_t = i | U, C) = \frac{e^{v_i, u}}{\sum_{j \in V} e^{v_j, u}}$$


The diagram illustrates the YouTube recommendation system architecture. It starts with a large "video corpus" (millions of videos) which feeds into a "candidate generation" module. This module outputs "hundreds" of candidates to a "ranking" module. The "ranking" module outputs "dozens" of recommended videos. The "ranking" module also receives input from "user history and context" and "video features". Additionally, there is a feedback loop from the "ranking" module back to the "candidate generation" module. Other candidate sources are also integrated into the system.

Research Camp

2. Theory → Hypothesis

- Theorizing requires abstraction & simplification
- I.e., people (in general) avoid conflict
- YouTube wants users to watch more videos
- Hypotheses fall out naturally from well-done theory
- **H1:** *YouTube's recommendation algorithm should suggest liberal content to liberals and conservative content to conservatives.*

Research Camp

3. Data Collection → Analysis

- Data collection separates "Data Science"...
- ...from "Science, with data"

- Recruit YouTube users to install [extension](#)



YouTube Recommendation Downloader

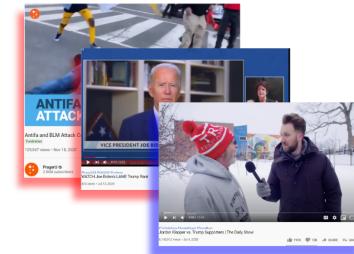
Offered by: csmappplugin

Research Camp

3. Data Collection → Analysis

- Data collection separates "Data Science"...
- ...from "Science, with data"

- Recruit YouTube users to install **extension**
- Start on randomly assigned **seed video**



Research Camp

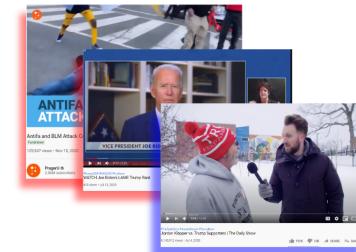
3. Data Collection → Analysis

- Data collection separates "Data Science"...
- ...from "Science, with data"

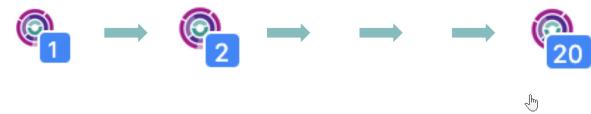
- Recruit YouTube users to install **extension**



- Start on randomly assigned **seed video**



- Follow **traversal rule** to select recommended video



Research Camp

3. Data Collection → Analysis

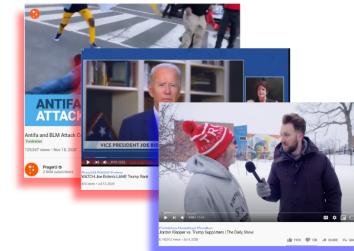
- Data collection separates "Data Science"...
- ...from "Science, with data"

- Recruit YouTube users to install **extension**



YouTube Recommendation Downloader

Offered by: csmapp plugin



- Start on randomly assigned **seed video**

- Follow **traversal rule** to select recommended video

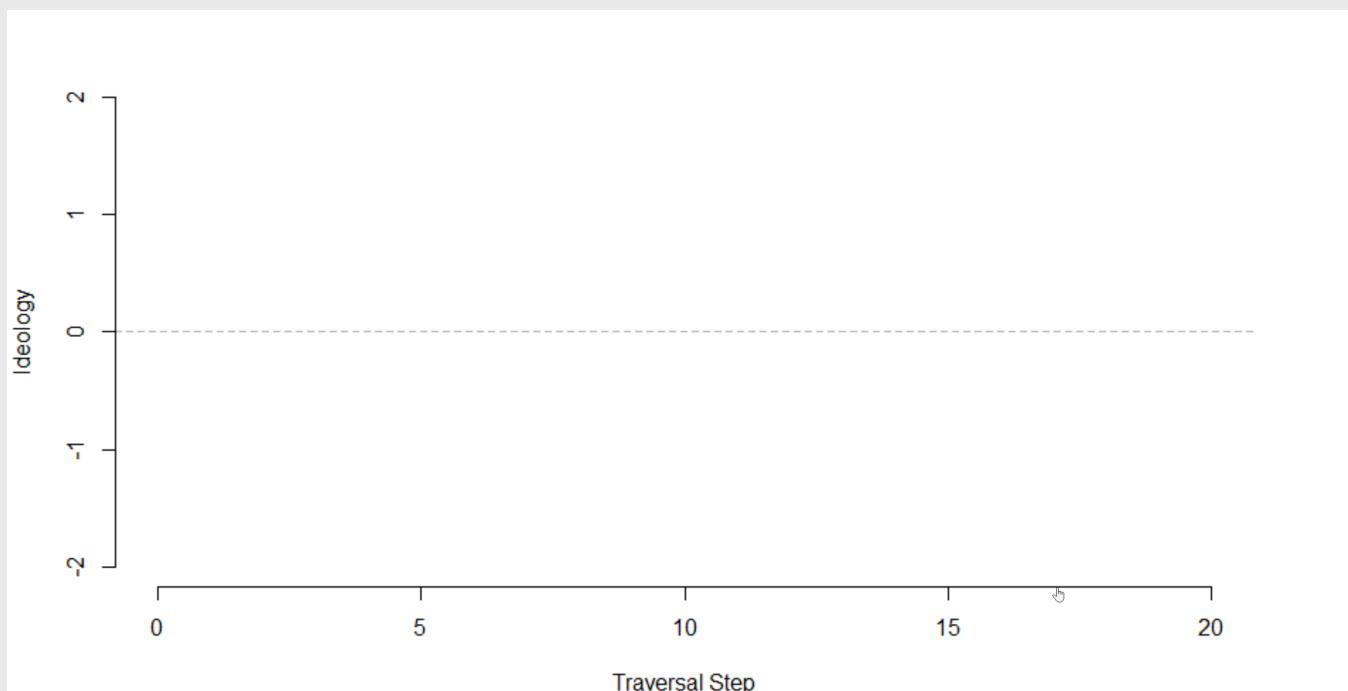


- Short **survey** on demographics, politics, and **BELIEFS ABOUT THE 2020 ELECTION**

Research Camp

3. Data Collection → Analysis

- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

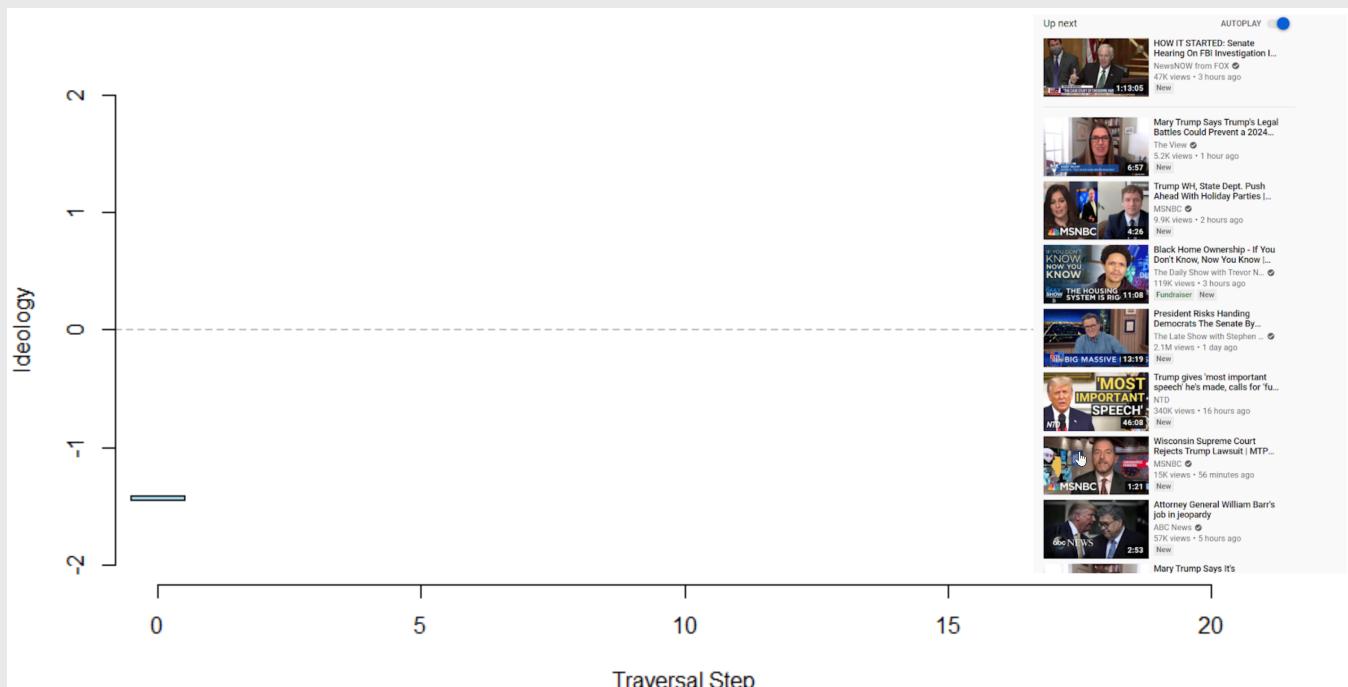
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

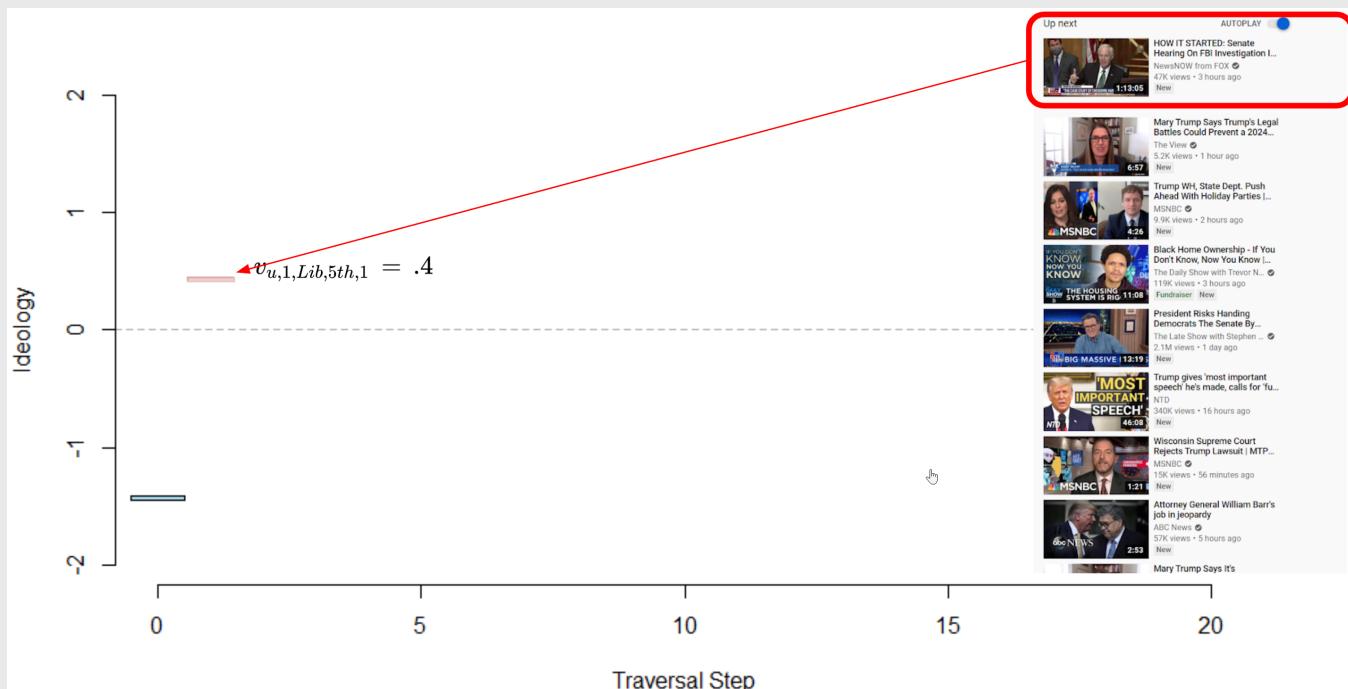
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

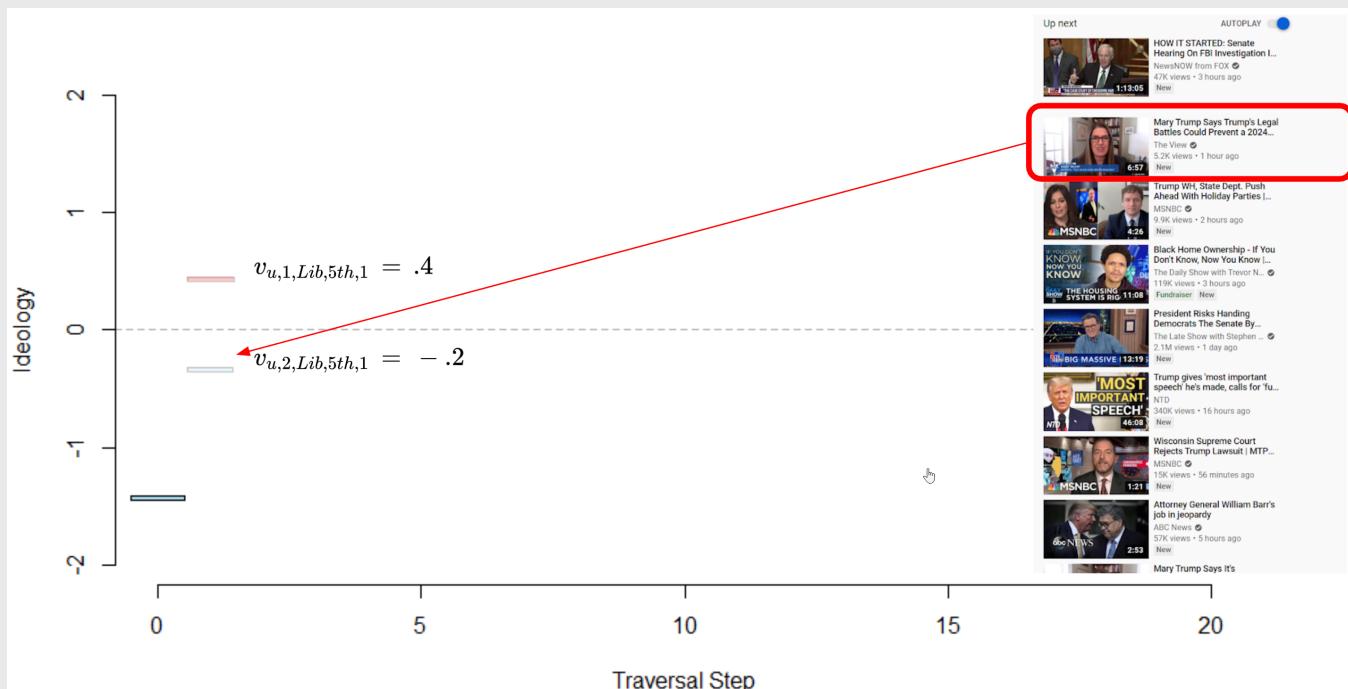
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

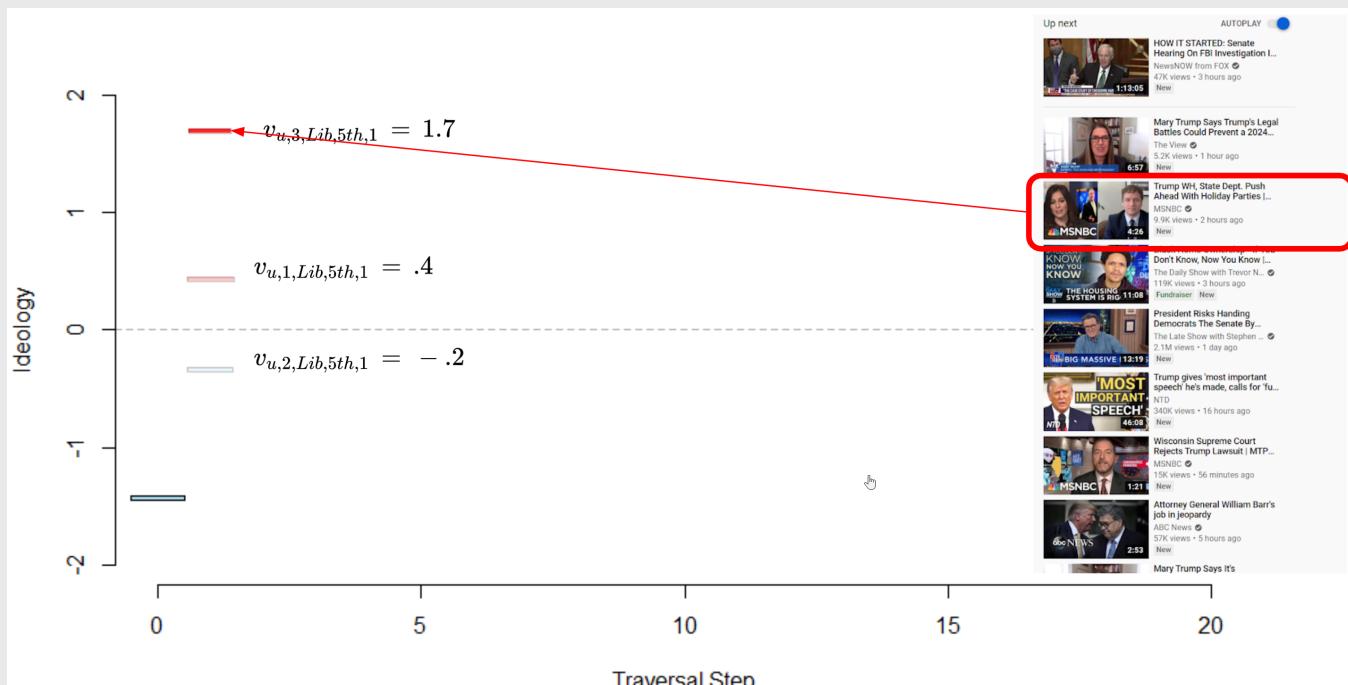
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

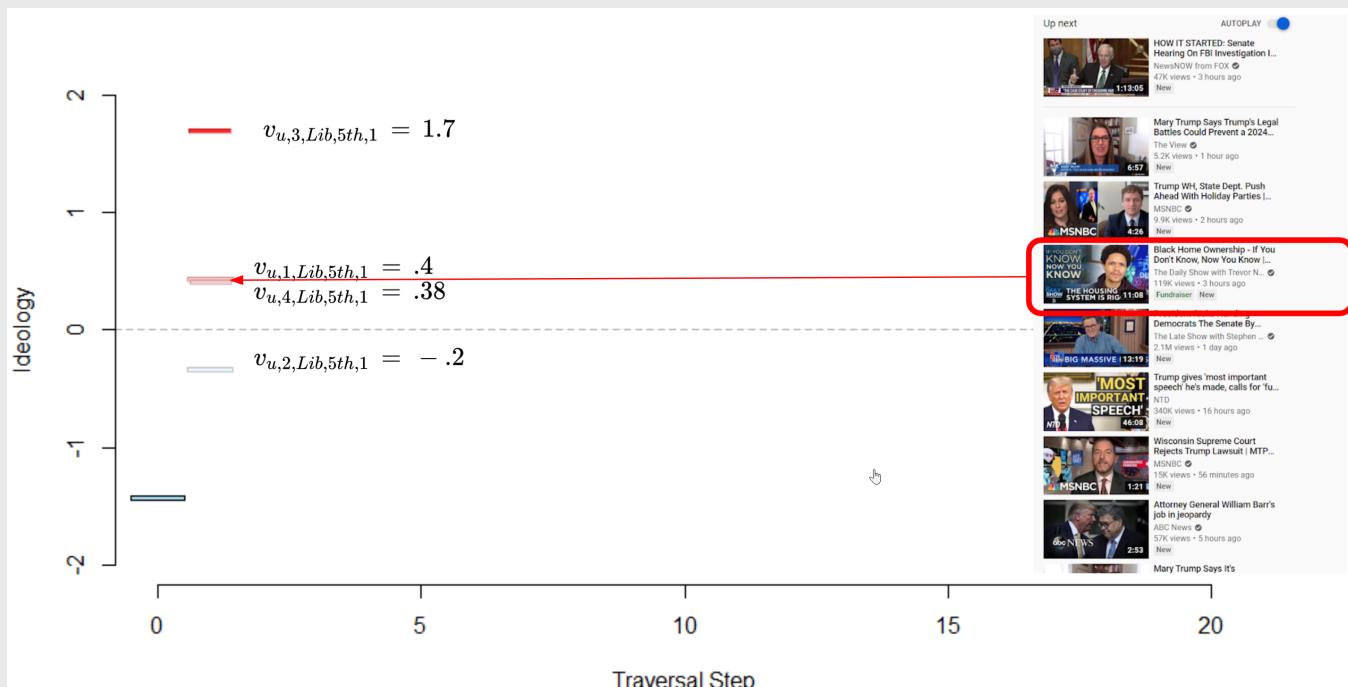
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

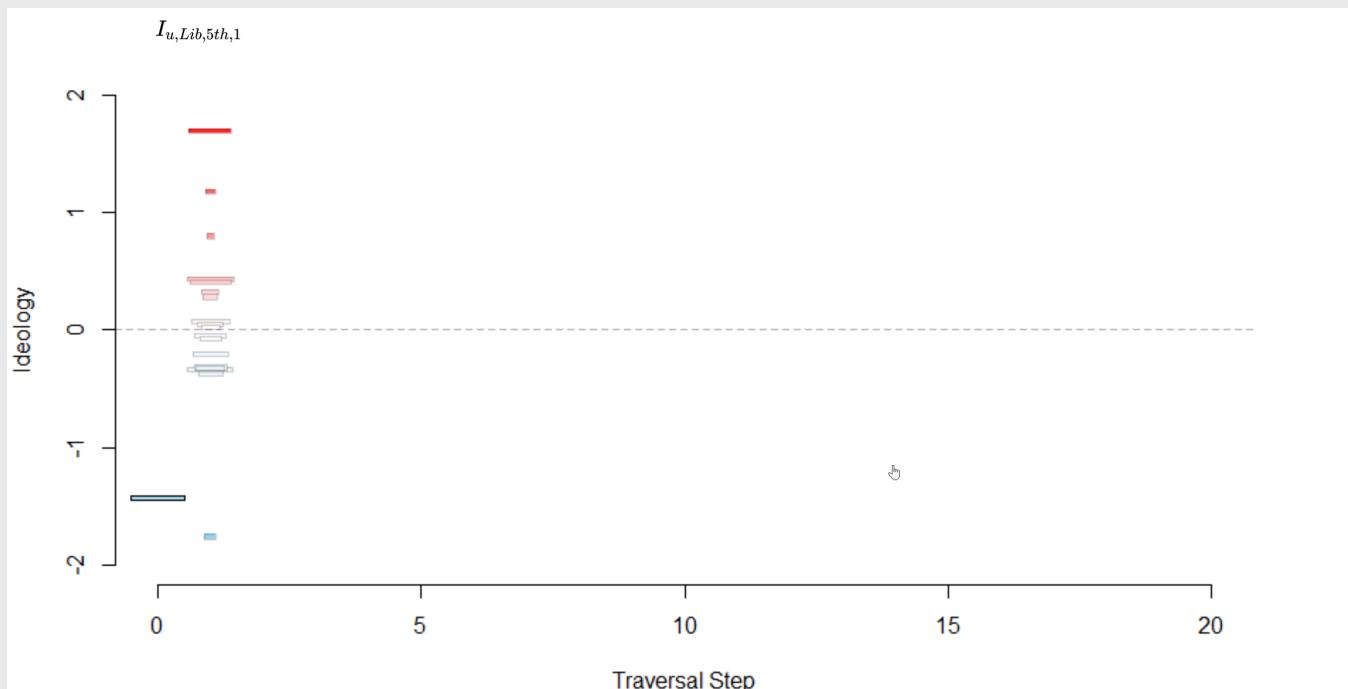
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

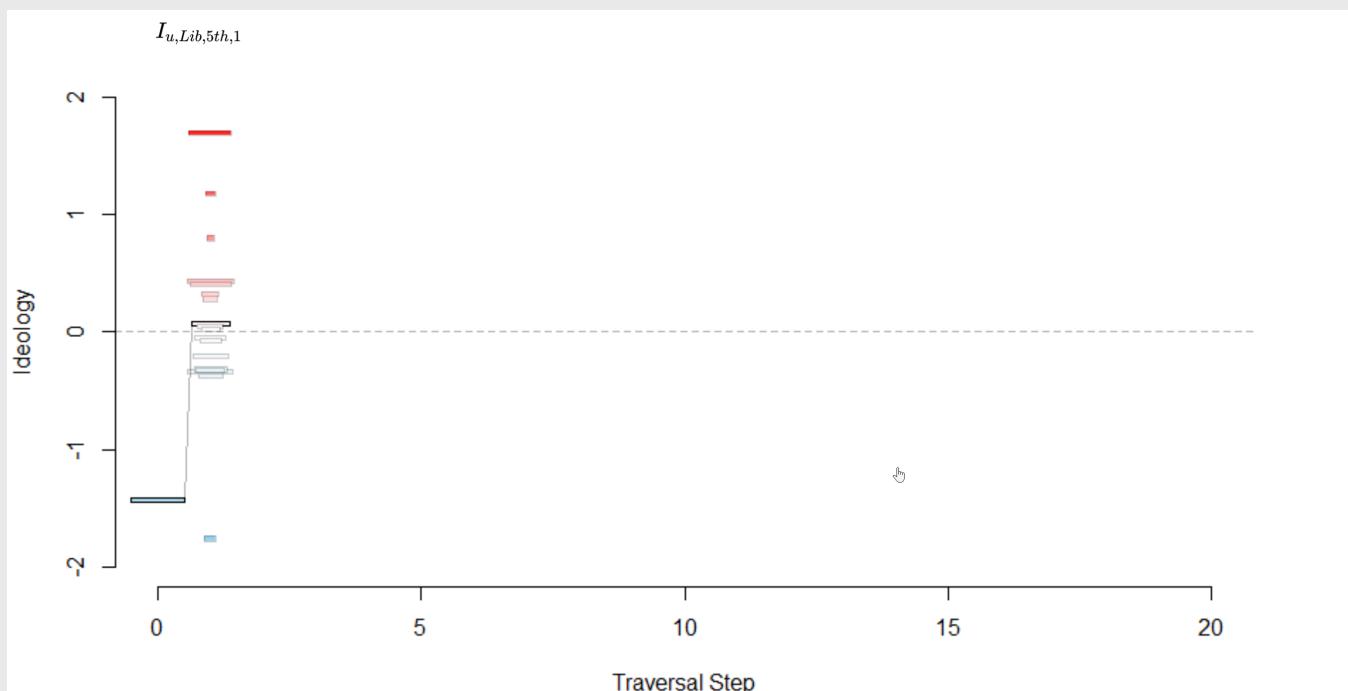
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

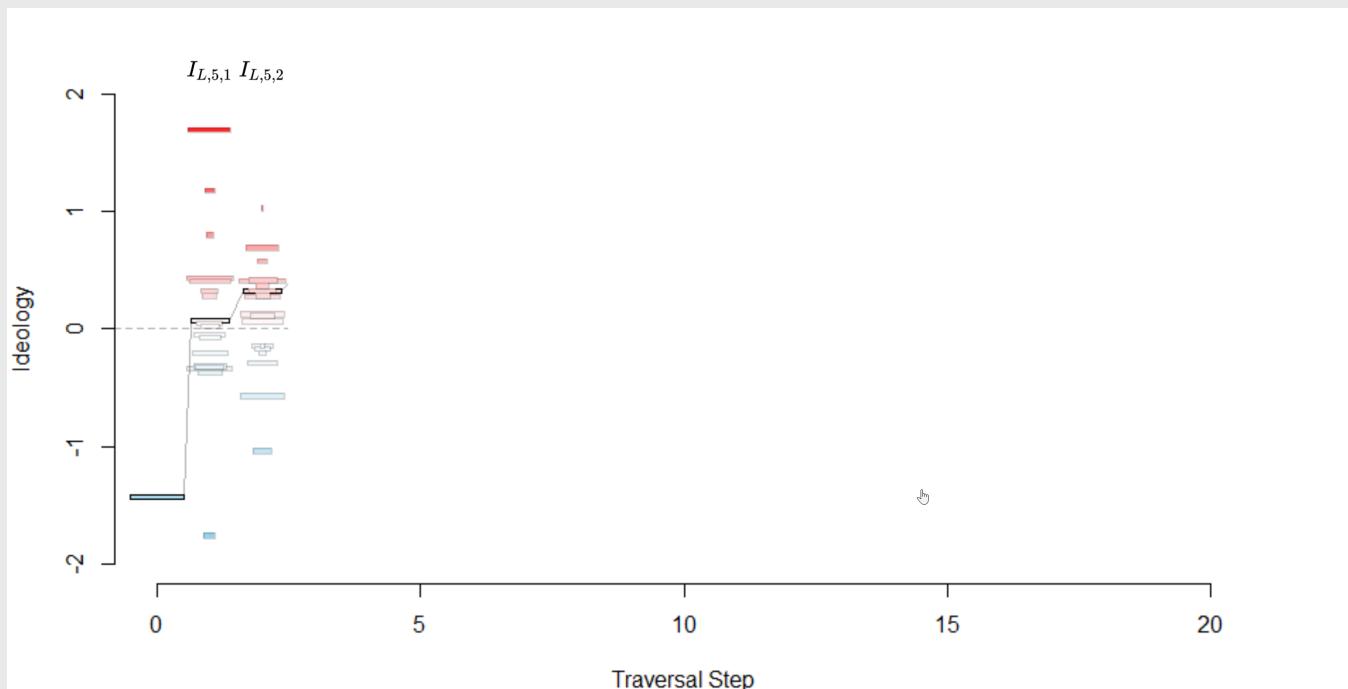
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

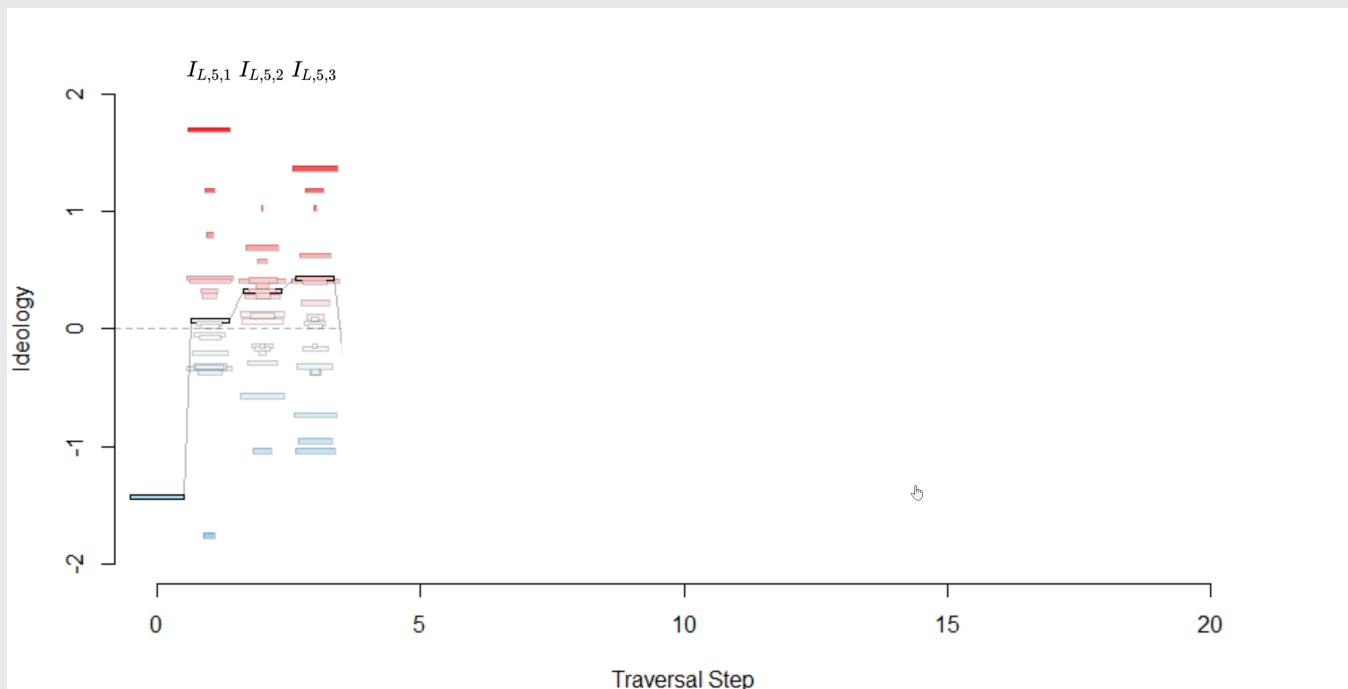
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

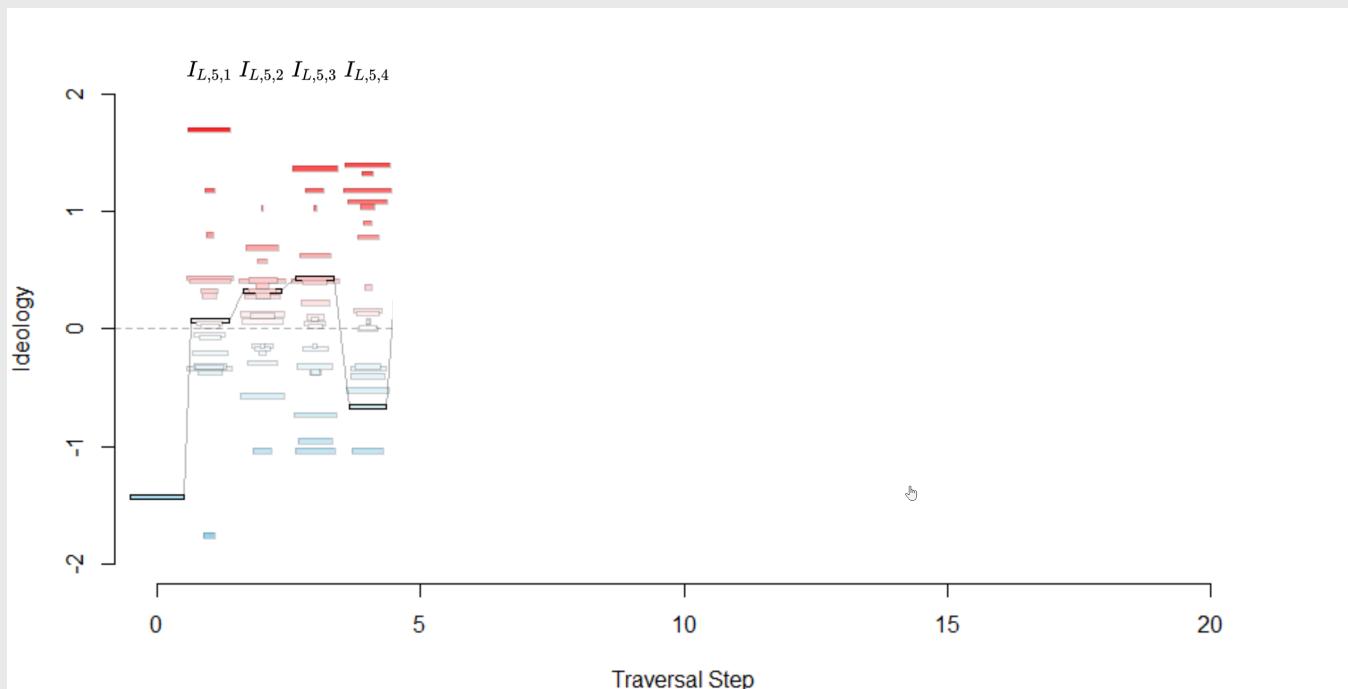
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

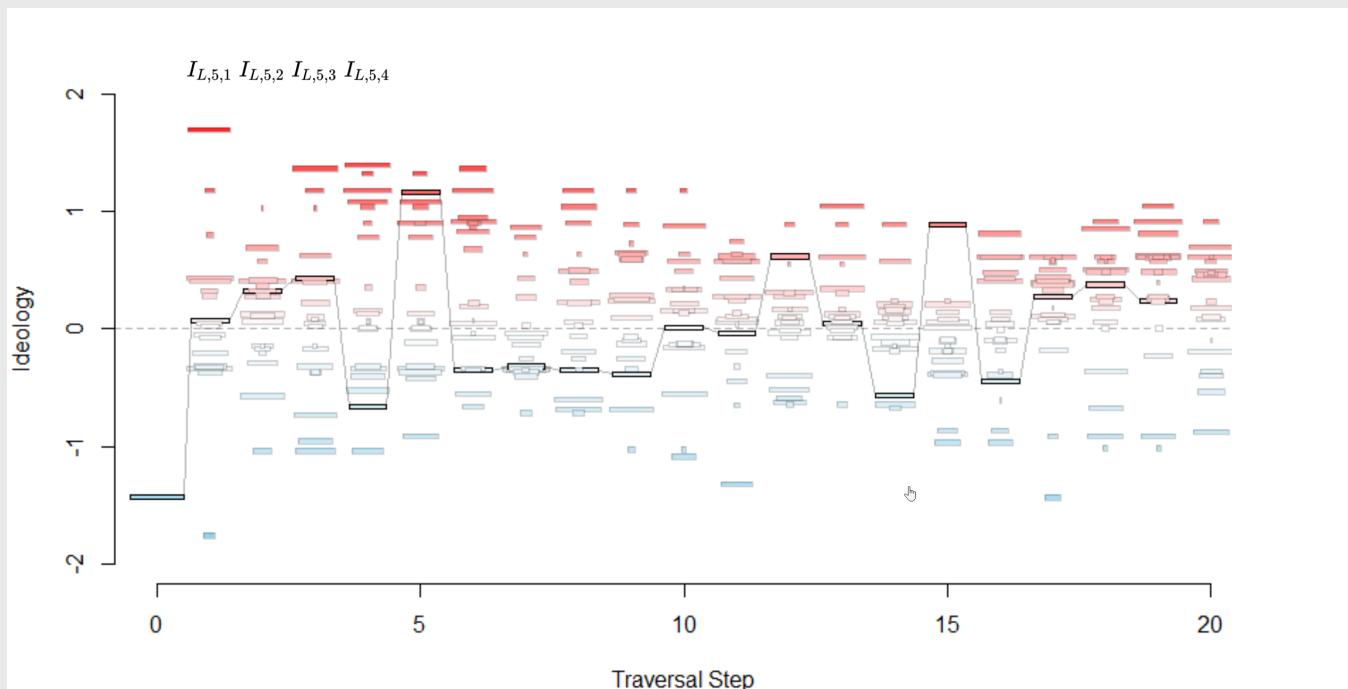
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

3. Data Collection → Analysis

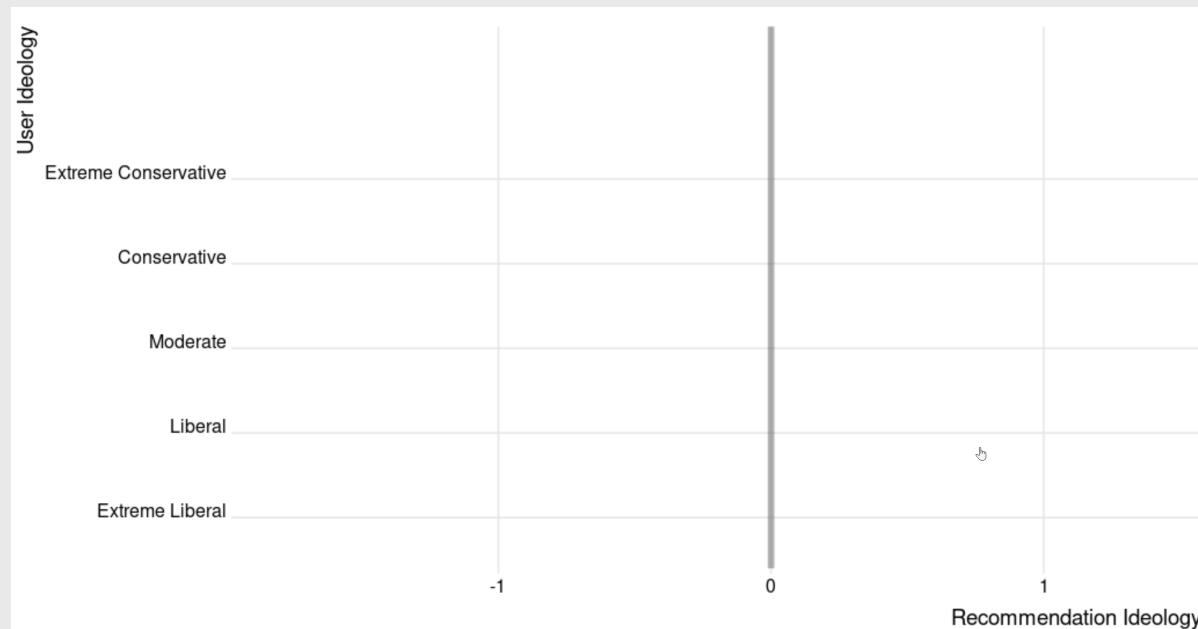
- Analysis is informed by the **data** you have collected...
- ...and the **hypotheses** you have generated



Research Camp

4. Results → Conclusion

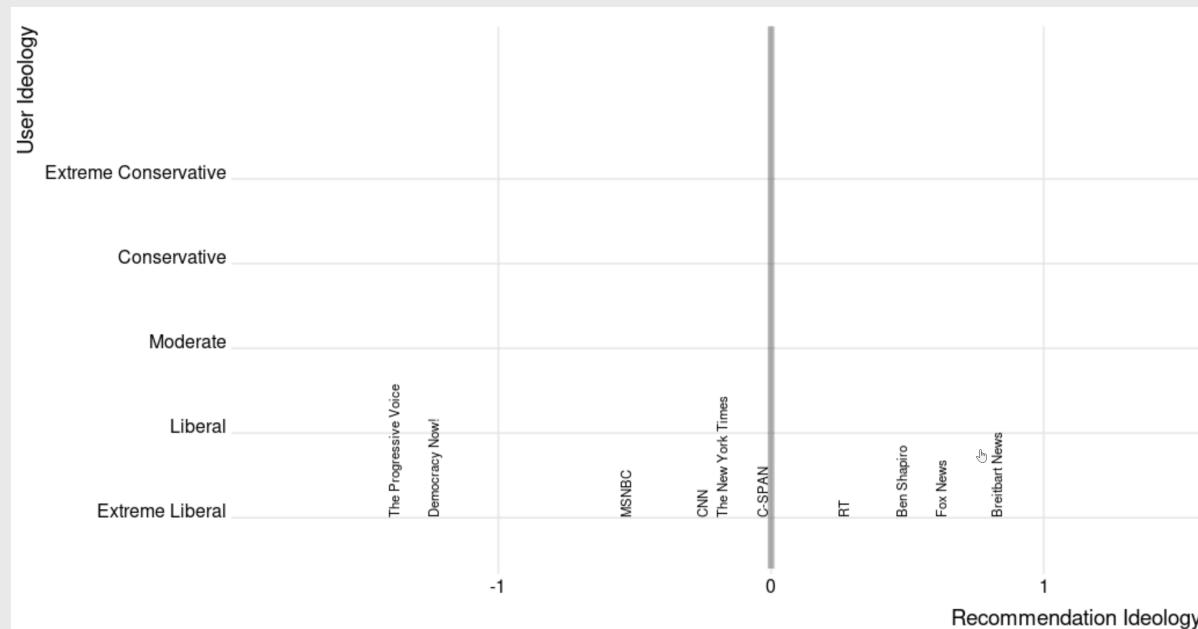
- Results fall out naturally from the analysis...
- ...and must be interpreted in terms of the theory and hypotheses...
- ...to draw conclusions



Research Camp

4. Results → Conclusion

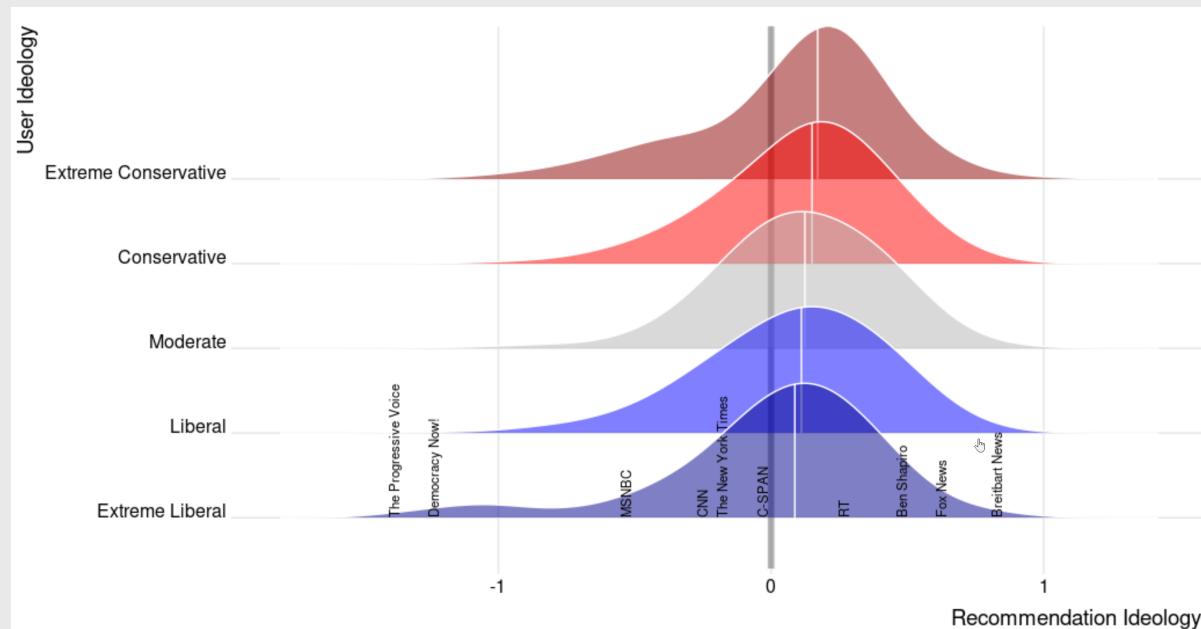
- Results fall out naturally from the analysis...
- ...and must be interpreted in terms of the theory and hypotheses...
- ...to draw conclusions



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



Prediction Camp

- **Goal/Problem/Challenge:** Measure the ideology of a YouTube

Prediction Camp

- **Data Wrangling:** Get matrix of links shared on political subreddits

The Ideology of a Video in 3 Steps: Step 1

Prediction Camp

- **Data Wrangling:** Get matrix of links shared on political subreddits

The Ideology of a Video in 3 Steps: Step 1

Behavior:
Sharing URLs

Posted by u/santanzchild Constitutional Conservative
6 hours ago 2

AOC, a Sitting Member of Congress,
Weaponized Her Followers in an
Attempt to Silence a Free Press
redstate.com/jenav... 2

1.2k 323 Comments Share ...

Posted by u/oz4ut Conservative 3 hours ago

Joe Biden's Abortion Policies Are
Grounds For Excommunication
thefederalist.com/2021/0... 2

308 131 Comments Share ...

Posted by u/guanaco55 Conservative 3 hours ago

The QAnon Takeover Of The GOP Is
A Fantasy Of Dems And The Media --
The GOP Civil War Is Between
Populists and the Establishment
thefederalist.com/2021/0... 2

303 43 Comments Share ...

Prediction Camp

- **Data Wrangling:** Get matrix of links shared on political subreddits

The Ideology of a Video in 3 Steps: Step 1

Behavior:
Sharing URLs

+

Domain:
Subreddits

Posted by u/santanzchild Constitutional Conservative
6 hours ago 2

AOC, a Sitting Member of Congress,
Weaponized Her Followers in an
Attempt to Silence a Free Press
redstate.com/jenav...



r/Conservative

1.2k 323 Comments Share ...

Posted by u/oz4ut Conservative 3 hours ago

Joe Biden's Abortion Policies Are
Grounds For Excommunication
thefederalist.com/2021/0...



r/neutralnews

308 131 Comments Share ...

Posted by u/guanaco55 Conservative 3 hours ago

The QAnon Takeover Of The GOP Is
A Fantasy Of Dems And The Media --
The GOP Civil War Is Between
Populists and the Establishment
thefederalist.com/2021/0...



r/SandersForPresident

303 43 Comments Share ...

Prediction Camp

- **Data Wrangling:** Get matrix of links shared on political subreddits

The Ideology of a Video in 3 Steps: Step 1



Posted by u/santanzchild Constitutional Conservative
6 hours ago 2

AOC, a Sitting Member of Congress,
Weaponized Her Followers in an
Attempt to Silence a Free Press
redstate.com/jenav... 2

↑ 1.2k ↓ 323 Comments Share ...

Posted by u/oz4ut Conservative 3 hours ago

Joe Biden's Abortion Policies Are
Grounds For Excommunication
thefederalist.com/2021/0... 2

↑ 308 ↓ 131 Comments Share ...

Posted by u/guanaco55 Conservative 3 hours ago

The QAnon Takeover Of The GOP Is
A Fantasy Of Dems And The Media --
The GOP Civil War Is Between
Populists and the Establishment
thefederalist.com/2021/0... 2

↑ 303 ↓ 43 Comments Share ...



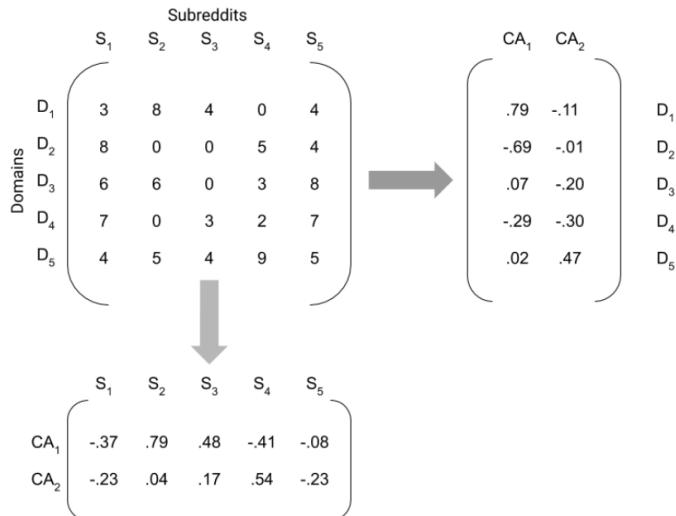
r/Conservative



r/neutralnews



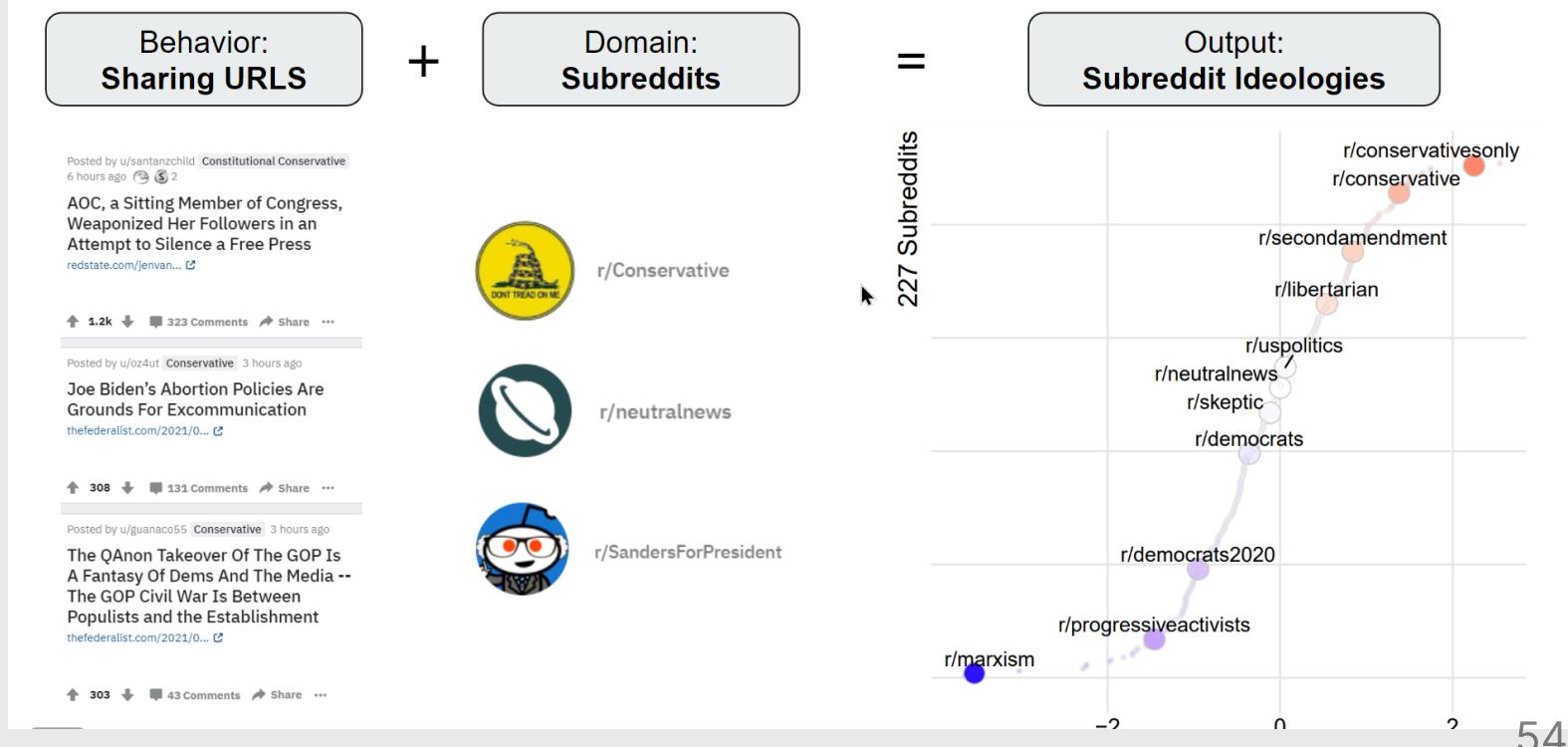
r/SandersForPresident



Prediction Camp

- **Data Wrangling:** Correspondence Analysis to estimate ideology scores for subreddits

The Ideology of a Video in 3 Steps: Step 1



Prediction Camp

- **Data Wrangling:** Get matrix of YouTube videos shared on scored subreddits

The Ideology of a Video in 3 Steps: Step 2

Prediction Camp

- **Data Wrangling:** Get matrix of YouTube videos shared on scored subreddits

The Ideology of a Video in 3 Steps: Step 2

Behavior:
Sharing Videos

Interview with Thomas Biryani by a reporter from an abc local texas affiliate's live feed:
<https://www.youtube.com/watch?v=X3WYY0fsF-I>
r/PublicFreakout Posted by u/elseman 20 days ago

52 37 Comments Share ...

A wand with a twist! I posted a "how to" on YouTube.
<https://m.youtube.com/watch?v=7QnhkNLUAvw> Credit:tpowen!_
r/Wandsmith Posted by u/torunay3 16 days ago

44 15 Comments Share ...

Made a video about the G14 and my setup! Check it out if you're interested! It would be greatly appreciated! <https://www.youtube.com/watch?v=crcTp9vYE&feature=youtu.be> Credit:unwqm...
r/Zephyrus14 Posted by u/alexszurkus 1 month ago

11 30 Comments Share ...

why is jimin like this full video:
<https://www.youtube.com/watch?v=iIhaZl1436M&t=173s> Meme
r/heungtan Posted by u/yangtiglighthere 14 days ago

74 6 Comments Share ...

Prediction Camp

- **Data Wrangling:** Get matrix of YouTube videos shared on scored subreddits

The Ideology of a Video in 3 Steps: Step 2

Behavior:
Sharing Videos

+

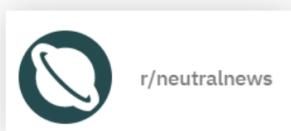
Domain:
Ideological Reddit

Interview with Thomas Biryani by a reporter from an abc local texas affiliate's live feed:
<https://www.youtube.com/watch?v=X3WYY0fsF-I>
r/PublicFreakout Posted by u/elseman 20 days ago
52 37 Comments Share ...

A wand with a twist. I posted a "how to" on YouTube.
<https://m.youtube.com/watch?v=7QikhNUlAew> Credit:tpowen!_ r/Wandsmith Posted by u/tprunify3 16 days ago
44 15 Comments Share ...

Made a video about the G14 and my setup! Check it out if you're interested! It would be greatly appreciated! <https://www.youtube.com/watch?v=crcTp9vYEY&feature=youtu.be> r/Zephyrus14 Posted by u/alexszurkus 1 month ago
11 30 Comments Share ...

why is jimin like this full video:
<https://www.youtube.com/watch?v=vIhaZtI436M&t=173s> r/heungtan Posted by u/yangtiglighthere 14 days ago
74 6 Comments Share ...



Prediction Camp

- **Data Wrangling:** Get matrix of 60k YouTube videos shared on scored subreddits

The Ideology of a Video in 3 Steps: Step 2

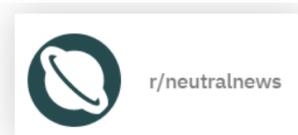
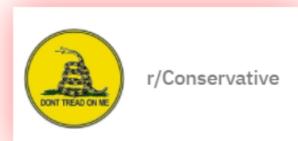
Behavior:
Sharing Videos + Domain:
Ideological Reddit

Interview with Thomas Biryani by a reporter from an abc local Texas affiliate's live feed:
<https://www.youtube.com/watch?v=X3WYY0fsF-I>
r/PublicFreakout Posted by u/eliseann 20 days ago
52 comments Share ...

A wond with a twist. I posted a "how to" on YouTube.
<https://m.youtube.com/watch?v=7QnhkNUlAew> Credit:tpowen!_
r/Wandsmith Posted by u/timurh3 36 days ago
44 comments Share ...

Made a video about the G14 and my setup! Check it out if you're interested! It would be greatly appreciated! <https://www.youtube.com/watch?v=crcTp9vYEY&feature=youtu.be>
r/Zephyrus14 Posted by valeruszurkus 1 month ago
11 comments Share ...

why is jimin like this? full video:
<https://www.youtube.com/watch?v=1haZtI436M&t=173s> Home
r/heungtan Posted by u/yangtighere 14 days ago
74 comments Share ...

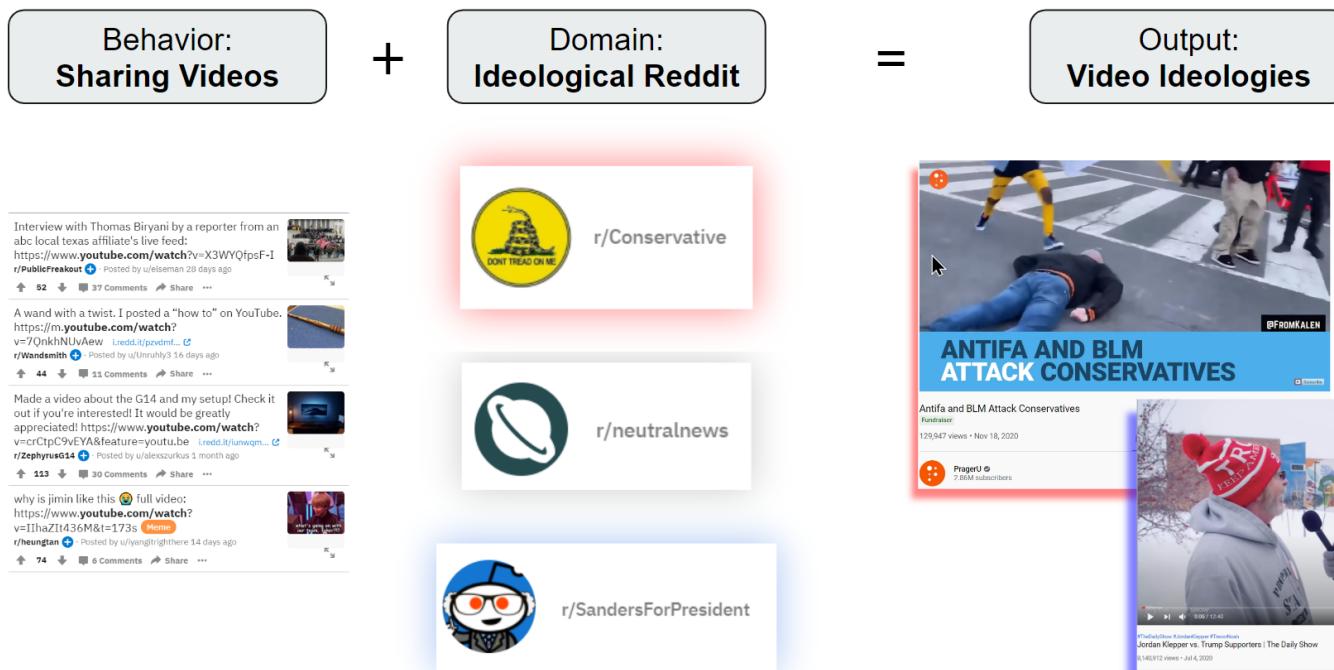


YouTube Videos	Subreddits						
	S ₁	S ₂	S ₃	S ₄	S ₅	CA ₁	CA ₂
V ₁	3	8	4	0	4	.79	-.11
V ₂	8	0	0	5	4	-.69	-.01
V ₃	6	6	0	3	8	.07	-.20
V ₄	7	0	3	2	7	-.29	-.30
V ₅	4	5	4	9	5	.02	.47

Prediction Camp

- **Data Wrangling:** Calculate video ideology as weighted mean of subreddits

The Ideology of a Video in 3 Steps: Step 2



Prediction Camp

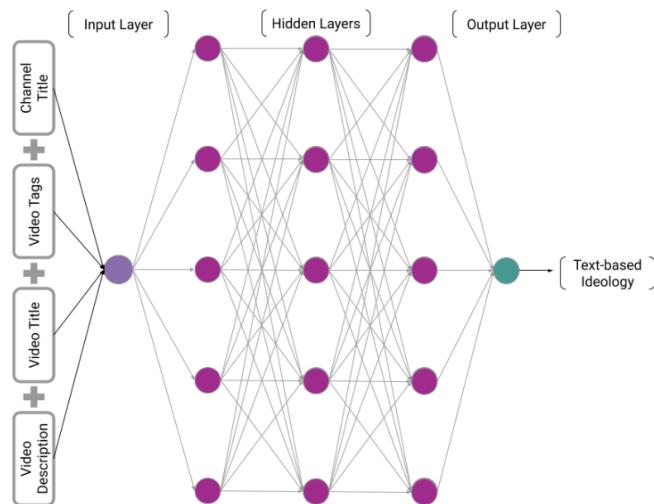
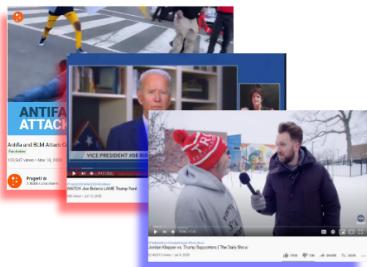
- **Model Training:** BERT transformer trained on 60k videos

The Ideology of a Video in 3 Steps: Step 3

Training Data:
67k Coded Videos

+

Classifier:
BERT Transformer



Prediction Camp

- **Prediction:** Measure the ideology of a YouTube video

The Ideology of a Video in 3 Steps: Step 3

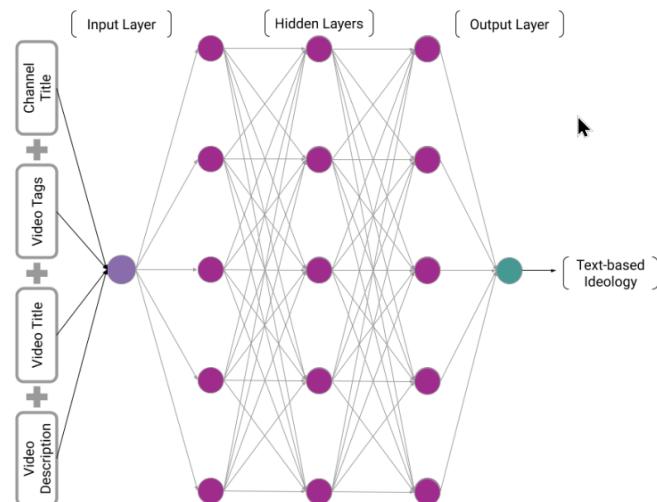
Training Data:
67k Coded Videos

+

Classifier:
BERT Transformer

=

Output:
Any Video's Ideology



Preview of Semester

- 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
 - You must work through the first homework assignment using an **.Rmd** file

Grades

Item	Percent	Points	
pset 1	5%	10	
pset 2	5%	10	
pset 3	5%	10	
pset 4	5%	10	
pset 5	5%	10	
pset 6	5%	10	
pset 7	5%	10	
pset 8	5%	10	
Midterm	20%	40	
Final Exam	20%	40	
Quizzes	20%	40	
Totals	100%	200	

The Syllabus

Date	Lecture	DOW	Goal	Assignments	Quizzes
9-Jan-23	Intro to Data Science	M	The scientific method, the camps of analysis	Pset 0 assigned	Quiz 1
11-Jan-23	Intro to R Part 1	W	Install and open R, packages, tidyverse		Quiz 2
16-Jan-23	BREAK	M			
18-Jan-23	Intro to R Part 2	W	Objects, functions, %>%, and <-		Quiz 3
23-Jan-23	Intro to R Part 3	M	Visualization in R	Pset 1 assigned	Quiz 4
25-Jan-23	Intro to R Review	W			
30-Jan-23	Data Wrangling	M	Replicability, R, and tabular data	Pset 2 assigned	Quiz 5
1-Feb-23	Univariate Analysis	W	Summaries of a single variable		Quiz 6
6-Feb-23	Multivariate Analysis Part 1	M	Summaries of multiple variables	Pset 3 assigned	Quiz 7
8-Feb-23	Multivariate Analysis Part 2	W	Visualizations of multiple variables		Quiz 8
13-Feb-23	Multivariate Analysis Part 3	M	Uncertainty and bootstrapping	Pset 4 assigned	Quiz 9
15-Feb-23	Multivariate Review	W			
20-Feb-23	Regression Part 1	M	The concept of a linear regression	Pset 5 assigned	Quiz 10
22-Feb-23	Regression Part 2	W	Interpreting a linear regression output and evaluating model performance		Quiz 11
27-Feb-23	Regression Part 3	M	Multiple regression and categorical predictors		Quiz 12
1-Mar-23	Regression Review	W			
6-Mar-23	Midterm Review	M			
8-Mar-23	Midterm Exam	W			
13-Mar-23	BREAK	M			
15-Mar-23	BREAK	W			
20-Mar-23	Classification Part 1	M	The concept of a logistic regression	Pset 6 assigned	Quiz 13
22-Mar-23	Classification Part 2	W	Interpreting a logistic regression output and evaluating model performance		Quiz 14
27-Mar-23	Classification Part 3	M	Using models for prediction	Pset 7 assigned	Quiz 15
29-Mar-23	Classification Review	W			
3-Apr-23	Clustering Part 1	M	k-means clustering	Pset 8 assigned	Quiz 16
5-Apr-23	NLP Part 2	W	k-means clustering on text		Quiz 17
10-Apr-23	NLP Part 3	M	Sentiment analysis	Pset 9 assigned	Quiz 18
12-Apr-23	NLP Review	W			
17-Apr-23	Advanced Topics in DS	M	Random forests, neural networks, image as data		Quiz 19
19-Apr-23	Ethics	W	The risks of rapid technological change		Quiz 20
24-Apr-23	Final Review	M			
26-Apr-23	Final Exam	W			

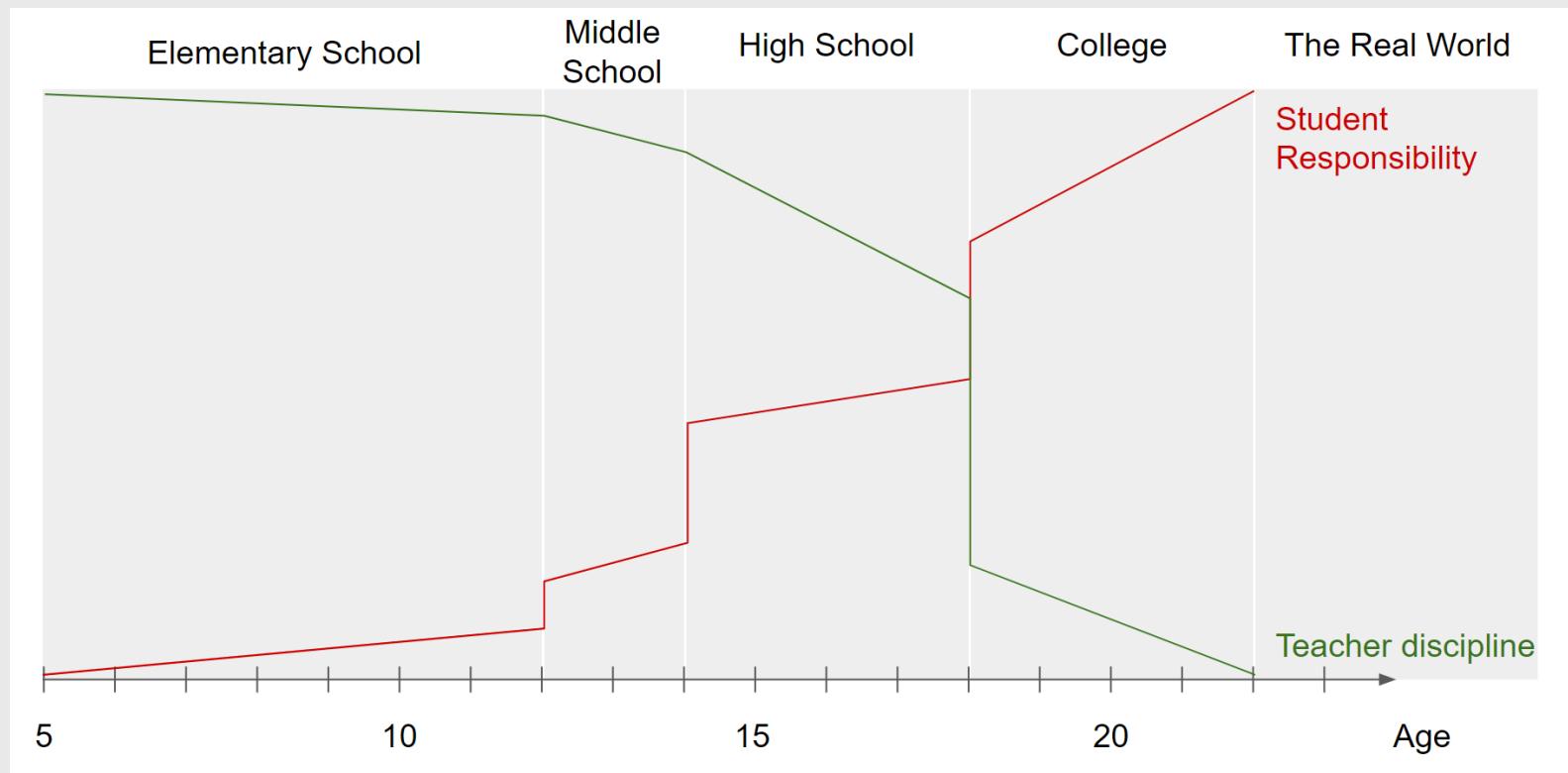
Honor Code

- Students are assumed to have read and agreed with the [Vanderbilt University Academic Honesty 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
- Brightspace (place for **submissions**)
 - Submit problem sets, quizzes, and exams
- GitHub (place for **materials**)
 - Find all in-class materials
- TA recitations / labs (place for **hands-on help**)
- Office hours (place for **hands-on help**)

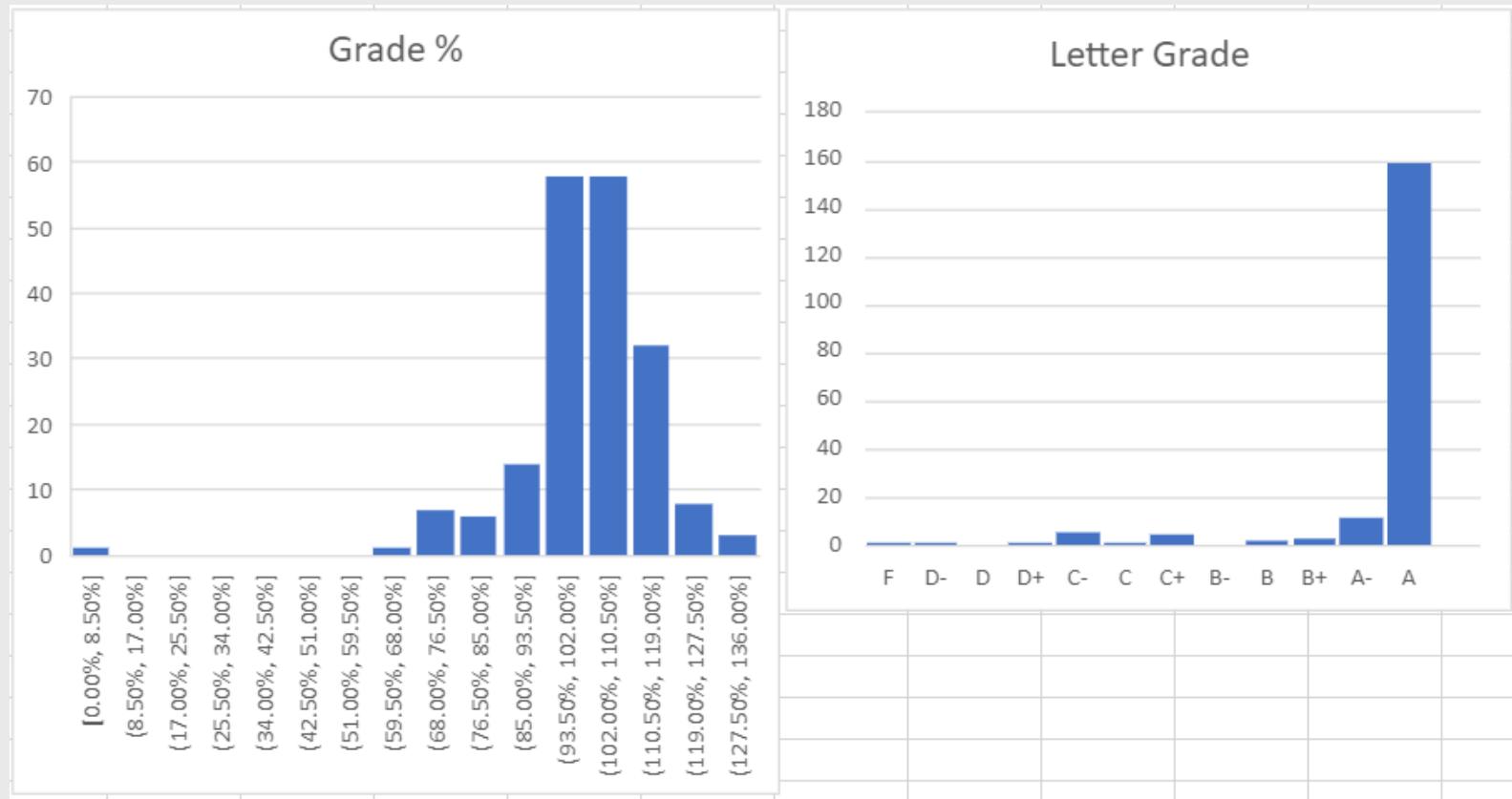
Teaching Philosophy



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
 - + lots of extra credit

Previous Semester



Conclusion

- Go to Brightspace and take the **1st** quiz
 - The password to take the quiz is 3326
- Homework:
 1. Work through Intro_Data_Science_hw.Rmd
 2. Complete Problem Set 0 (on Brightspace)