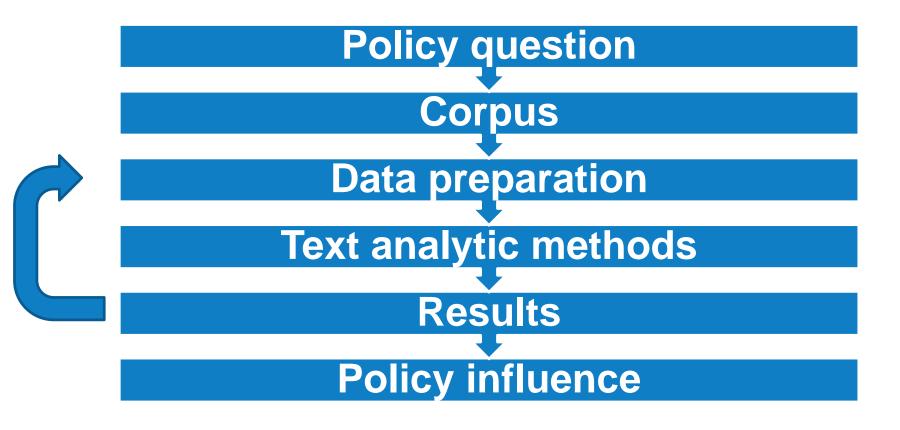


Agenda

- Introduction
- Framework
- **Policy Question**
- Text Analysis Strategies
 - Word use frequency
 - Sentiment analysis
 - Topic models
- Policy influence

Analytical framework



NOTE: Sample R code will be made available for all methods described in this presentation

Policy Question: Strategic Planning for UC 2030









TOPICS

COMMENTARIES

PROJECTS

AUDIO & VIDEO

DATA PUBLICATIONS

S ABOUT

ABOUT EDSOURCE

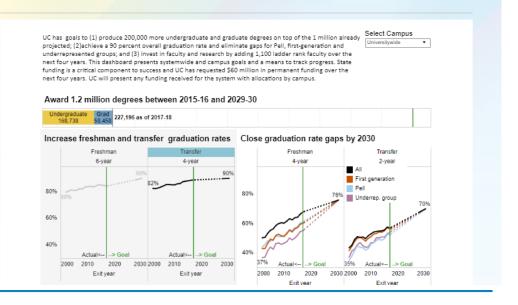
University of California aims to improve graduation rates by 2030

Focus is on low-income, African-American, Latino and first generation students.





UC 2030 Dashboard



Define your corpus

What is one of the most meaningful learning experiences you have had at "your UC campus"?

Taking sociology with [professor]. Hearing where he came from and the experiences he went through to be able to make it to this level of academia was very inspiring.

Joining the rowing team. Learning how to study and be successful in my classes.

Projects!

College is expensive.

Learning you won't be reached out to, you have to be the one that reaches out.

Data Preparation:

- Merge survey responses to student characteristics
- Generate list of all popular words in responses
 - Omit words that appear fewer than 50 times
 - Omit "stop words" like prepositions
 - "De-stem" words (e.g. remove plural 's' and verb conjugations)
 - Remove "invalid" words that will impair results (e.g. campus names)
- Add common two- or three-word phrases

Three text analysis tools:

1. Word use frequency by group

2. Sentiment Analysis

3. Topic Models

Methodology:

- Produce a data frame with:
 - Student characteristics, usually defined as binary variables
 - 0/1 indicators for every identified word
- Estimate linear regressions of each student characteristic on all of the word indicators
- Produce word clouds of the words with the highest estimated t-statistics (from the null hypothesis)
 - t-statistics are higher if (1) the word is more common (bypassing) outliers) and (2) the word disproportionately appears in the responses of students with the given characteristic

Word frequency examples: simple word cloud

```
meaningful learning experiences
           learning experiences
thing also ableresearch
world time professor
different
way people student community
made work life learn real
opportunit
                                         opportunity
year experience major
   lot learning friend study meaningful one many
   meaningful learning like
              course campus
    learning experience
```

Word frequency examples: UC campuses

UC Davis

connecting experience freshmen understood childadvisorleading man working connection conference community degree native CINIC hand bad feel CINIC run access accesswork christian intern theory first year specific expectation spanish quarter system belong

UC Santa Cruz

withinextreme leadership skills progress university psych travel natural model housing computer incredibly college game without forever senior theater stem field teacher moment section org active important spacetutoring creating color intro faculty member

UC Merced

meaningful experiences whether writing stay grant one another talent type add attending uc generation friend bridgepublic health journey club core strive stuck semester study habits matter expect engineering leadership bit staff status campus opportunity everyone service assist one meaningful amazing people meaningful experience develop

Word frequency examples: Ethnicity

Asian

student organization interacting meet met met met broaden handson computer science whether joining diverse driven utilize grew biothus pressure japanese learn asian growth improve meeting dancerigor assistant comfort feel like mature definitely business college mentorship graduate student perspective interpersonal

Hispanic

leadership skills knowing injustice united states advisor meaningful experience something home navigate program latino spanish biological easierlowincome scholar support _ overall manner knew able assignment help us hispanic meant meaningful experiences english undocumented amazing open minded

White

```
best learning
beneficial natural
core two surrounding
class taught
spent field fantastic far
seminar senior white physics town
sort
man coursework explain
graduate thesis animal
global research source
roommate humanity incredible acceptance
incredibly literature
independent research
getting faculty requirement
```

Three text analysis tools:

1. Word use frequency by group

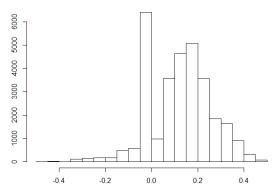
2. Sentiment Analysis

3. Topic Models

Methodology:

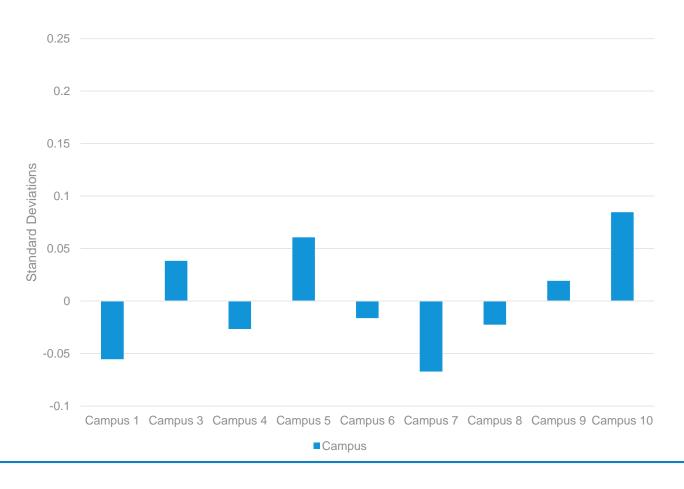
- Estimate 'Positive' and 'Negative' sentiment of each response
 - Uses the "SentimentAnalysis" R package
- Estimate linear regressions of response sentiment against student characteristics
- Produce word clouds of the words most-associated to positive and negative responses

Histogram of response sentiment measures



Note: sentiment ranges from -1 to 1

Regression estimates: effect size of estimates



Regression estimates: effect size of estimates



Three text analysis tools:

1. Word use frequency by group

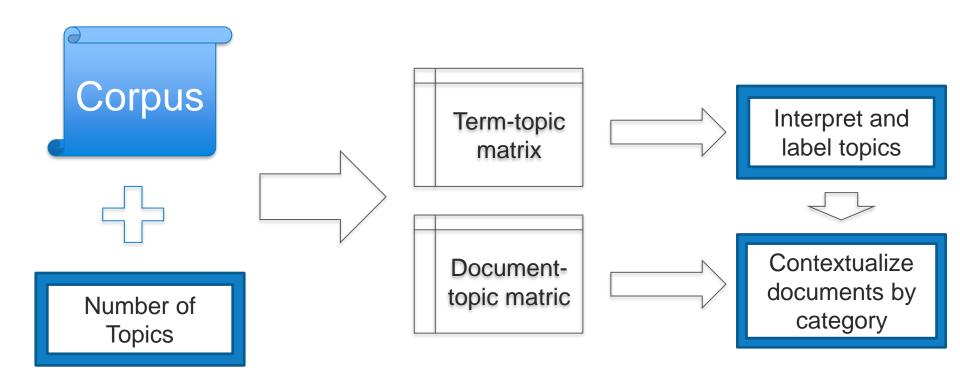
2. Sentiment Analysis

3. Topic Models

Methodology:

- Computationally identify "topics" in responses
 - Uses the "topicmodels" R package
 - You choose the number of topics (we chose 30). The computer finds them, characterized by often-adjacent words
- Each topic is characterized by the words that appear more-frequently in that topic; each response is characterized by its topics
- Regress student characteristics on topic composition to identify which topic is mostassociated

Methodology:



UC Berkeley is most closely associated with topic 17, which looks like this:

UC Merced is most closely associated with topic 10, which looks like this:

majority exist remember quarter system assist project setting term academically engagement motivate speakeraw world application meeting complete recognize aspect

accomplishextreme academically benefitinternational family piece pretty music history sitting difficulty believe overall hardest follow word midterm ever word incredible exist facing today shape final play whatever appreciate advisor essential abroad ethnicity government studied roommate understand lecturer

Example: Pell Students are most closely associated with topic 25, which looks like this:

Topic 25:

```
mostly found
     design closermindset occurred improving survey
   blackshapetouch treat belong
reasontaken serve create negative attitude find
standsaw walkseek felt please imagine home cool will now
       voice applicable knew
       communication advocate invaluable
                positive campus
```

Using results to influence policy

Using the data to inform 2030 strategic goal setting:

- Identify experiences that matter
- Identify programs and strategies that have impact
- Find artifacts that tell a representative story

Using results to influence policy

Survey responses that highlight program impact

"For me the most meaningful learning experience was being disqualified from the university because of low grades that were caused by difficulty with anxiety that I developed at college.

However the support of my CAPS counselors, Marshall academic advisors and some faculty helped me return and succeed. These faculty that helped me...really helped me overcome my condition and regain confidence in myself."

Survey responses that support sense of belonging

"Being able to make new friends knowing you're not the only one that may experience certain struggles. It made my first year college experience somewhat more easy to handle."

"The most meaningful learning experience that I have had so far was discovering the importance of self care... I have had to deal with the most stress and anxiety I have ever had...but there are campus resources to help me deal with it."

Fixed-effect word regression, sentiment analysis, and topic modeling can be implemented "out-of-the-box."

Computational text analysis is:

- Feasible
- Illuminating
- Actionable

