

# Mirror, Flashlight, and Roadmap: How Institutions of Higher Education Use Data to Advance Student Outcomes

And Why We Need Your Help

Brad Weiner

University of Colorado Boulder

2022-04-22



# About Me

-  **Interim Chief Data Officer, University of Colorado Boulder**
-  **20 years experience in higher education**
-  **15 years on campus (Kansas, Vanderbilt, Minnesota, Colorado)**
-  **5 years in Ed-Tech/Consultancy**
-  **13 years Higher Ed Analytics/Data Science**

# Contact

 [brad.weiner@colorado.edu](mailto:brad.weiner@colorado.edu)

 [brad\\_weiner](https://twitter.com/brad_weiner)

 [bradweiner.info](http://bradweiner.info)

A photograph of a grand, classical-style university building made of light-colored stone. The building features a central entrance with a triangular pediment, flanked by two wings with arched windows. A prominent clock tower rises from the center. In the foreground, a paved walkway leads towards the building, lined with green lawns and small, rounded green shrubs.

What Does a Research University Do?

- ✓ Create new knowledge in the form of academic research, scholarly output, and creative works
- ✓ Teach undergraduate and graduate students so they can be engaged members of society and informed, participatory citizens in a democracy
- ✓ Generate credentials to signal to society and the labor market that graduates have achieved certain skills
- ✓ Engage with the public in order to be broadly useful to society and so we may create economic and social value

✓ Play football

- ✓ Preserve knowledge through libraries, archives, curation, and art
- ✓ Heal patients and create life-improving treatments
- ✓ Act as an economic engine for local and global businesses
- ✓ Other stuff that we haven't even listed here

A photograph taken from behind a graduate wearing a black cap and gown with a purple stole. The graduate has long, straight, reddish-brown hair. A purple tassel hangs from the cap. In the background, a large crowd of people in similar graduation attire is visible, though they are out of focus.

# Who Does a University Serve?

 **Students**

 **Faculty**

 **Staff**

 **The Public**

 **Alumni, Friends, Fans, and Donors**

 **The Economy**

 **Its Local Community**

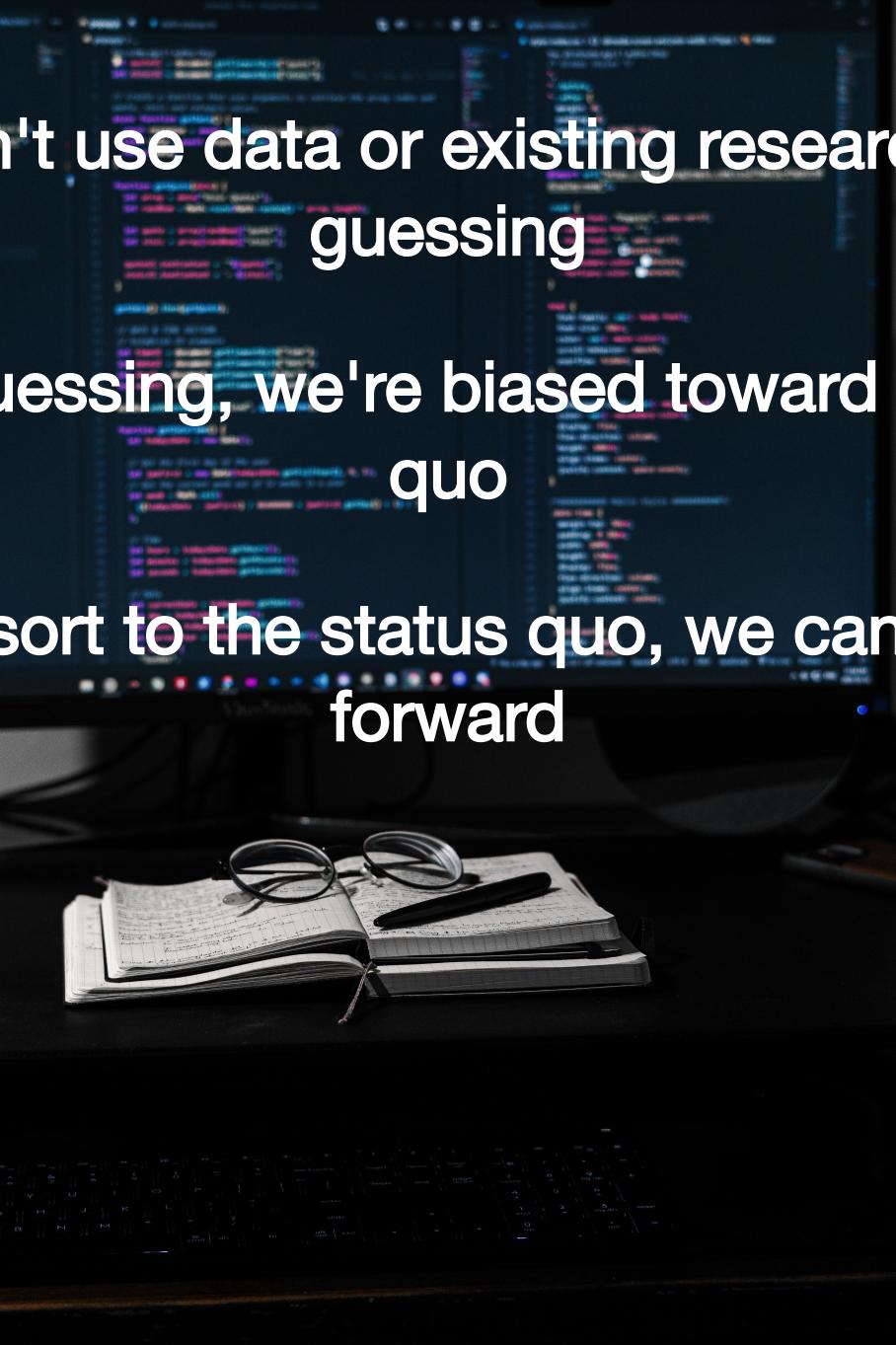
A wide-angle photograph of a green soccer field. In the upper left foreground, a white soccer goal is positioned on the grass. A player in a blue shirt and black shorts is seen from behind, kicking a black soccer ball towards the goal. The field has white boundary lines and a small number '12' is visible near the top left corner of the goal area. The background shows a blurred stadium seating area.

**Universities have diffuse and sometimes competing goals**

**Universities have myriad stakeholders**

**Universities have to demonstrate our value to taxpayers and to society**

**Universities are resource constrained**



If we don't use data or existing research, we're guessing

If we're guessing, we're biased toward the status quo

If we resort to the status quo, we can't move forward

**Universities use data as a "mirror" to see how we're doing**



**Universities use data as a "flashlight" to find areas for improvement**



# Descriptive Statistics and Reporting



University  
of Colorado  
Boulder

Reset  
Filters



## CU Boulder Retention & Graduation Rates

First-time, full-time undergraduate students entering in the summer/fall

School/College: All Major: All

Entry Year  
2010 to 2021

Entry College  
All

Entry Major  
All

Gender  
All

Race/Ethnicity  
All

Residency  
All

First Generation Status  
All

Financial Aid Status

Disability Status  
All

RAP Participant  
All

ACO Status  
All

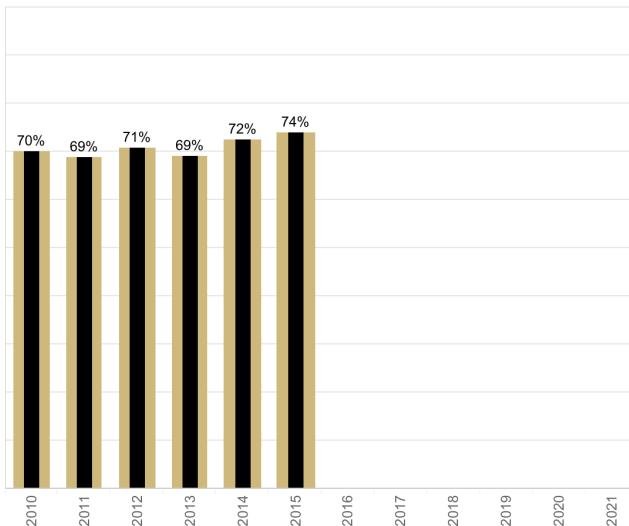
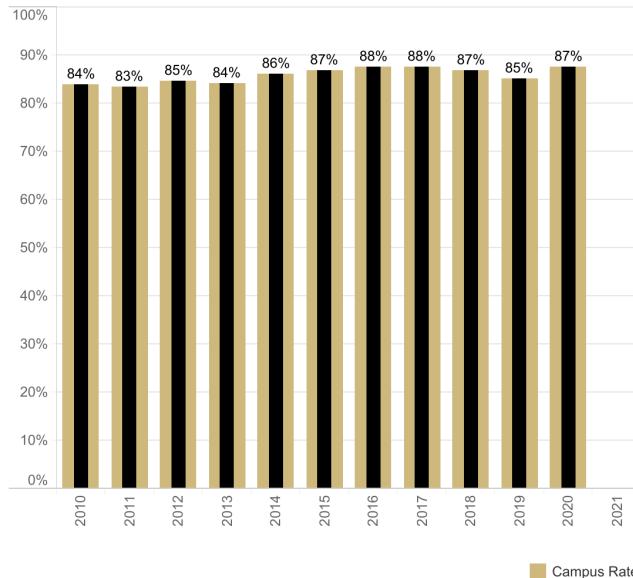
Greek Status  
All

Select Retention Rate  
Enrolled 2nd fall, any college

Select Graduation Rate  
Graduated by 6th summer, any college

**Retention Rate:** Enrolled 2nd fall, any college

**Graduation Rate:** Graduated by 6th summer, any college



These are historical retention and graduation rates for CU Boulder for first-time, full-time undergraduates



Reset  
Filters



Entry Year  
2010 to 2021

Entry College  
All

Entry Major  
All

Gender  
All

Race/Ethnicity  
All

Residency  
All

First Generation Status  
First Gen

Financial Aid Status i  
All

Disability Status  
All

RAP Participant  
All

ACO Status  
All

Greek Status  
All

## CU Boulder Retention & Graduation Rates

First-time, full-time undergraduate students entering in the summer/fall

School/College: All Major: All

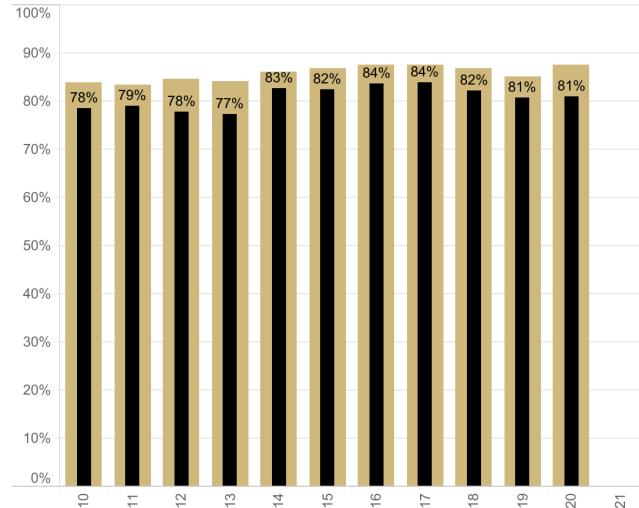
### Select Retention Rate

Enrolled 2nd fall, any college

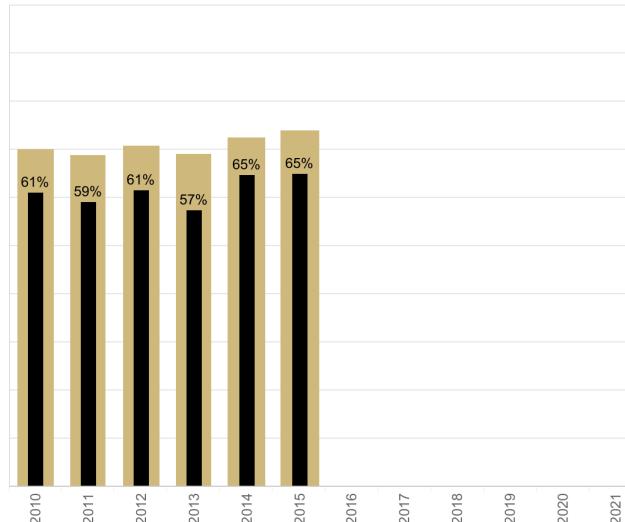
### Select Graduation Rate

Graduated by 6th summer, any college

#### Retention Rate: Enrolled 2nd fall, any college



#### Graduation Rate: Graduated by 6th summer, any college



These are historical retention and graduation rates for CU Boulder for first-time, full-time undergraduates who are ALSO first-generational

It appears as though, on average, first-generational students are retained and graduate at a lower rate than peers whose parents attended college.

A dark, vertical photograph of a person from behind, standing in a forest at night. The person is holding a flashlight that beams a bright white light towards the upper right corner of the frame. The sky is filled with numerous small stars. The foreground is dark, showing the silhouettes of trees.

# Data Science and Predictive Analytics



**Research Question: Can we hypothesize why first-generational students graduate at lower rates?**

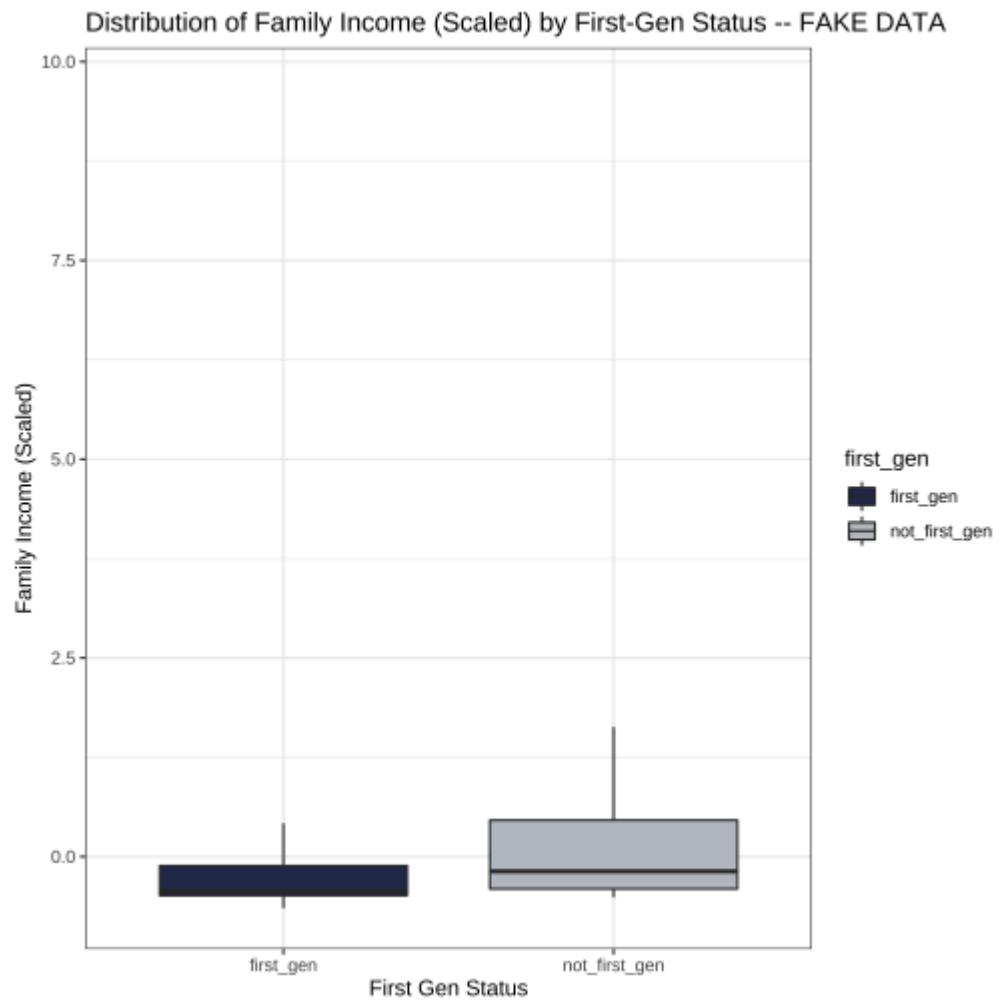
# Compile the Raw Data (Not Real Data)

```
##  
## Attaching package: 'kableExtra'  
  
## The following object is masked from 'package:dplyr':  
##  
##     group_rows  
  
## Warning: package 'tidyverse' was built under R version 4.0.2  
  
## — Attaching packages ————— tidyverse 1.3  
  
## ✓ readr  2.0.1      ✓forcats 0.5.1  
## ✓ stringr 1.4.0  
  
## Warning: package 'forcats' was built under R version 4.0.2  
  
## — Conflicts ————— tidyverse_conflicts  
## x readr::col_factor()    masks scales::col_factor()  
## x purrr::discard()      masks scales::discard()  
## x dplyr::filter()       masks stats::filter()  
## x stringr::fixed()      masks recipes::fixed()  
## x kableExtra::group_rows() masks dplyr::group_rows()  
## x dplyr::lag()          masks stats::lag()  
## x readr::spec()         masks yardstick::spec()  
  
## Warning: package 'magrittr' was built under R version 4.0.2
```

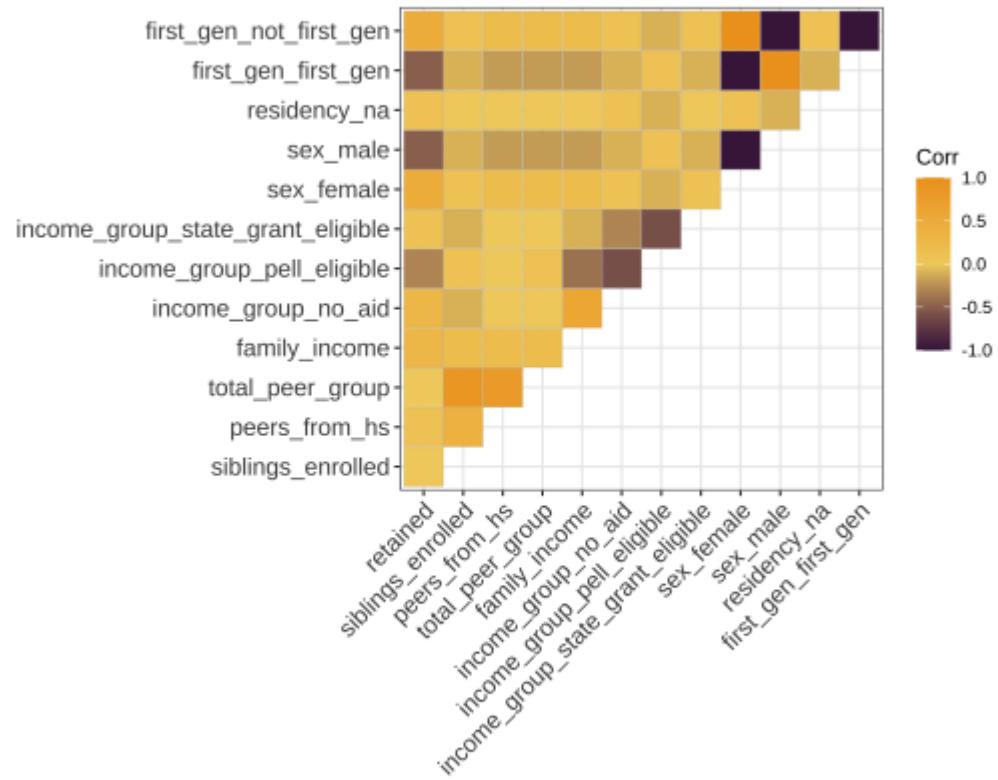
# Pre-Process the Data (Again Not Real Data)

student_id	1	2	3	4	5	6
retained	0	1	1	1	0	0
income_group	Pell Eligible	No Aid	Pell Eligible	No Aid	Pell Eligible	Pell Eligible
sex	male	female	female	female	male	male
age	-0.5300051	0.5714304	-0.2546462	0.3649113	0.3649113	NA
siblings_enrolled	0.4325504	0.4325504	-0.4742788	0.4325504	-0.4742788	-0.4742788
peers_from_hs	-0.4734077	-0.4734077	-0.4734077	-0.4734077	-0.4734077	-0.4734077
residency	Resident	Non-Resident	Resident	Resident	Resident	International
total_peer_group	1	1	0	1	0	0
first_gen	first_gen	not_first_gen	not_first_gen	not_first_gen	first_gen	first_gen
family_income	-0.5021568	0.7865640	-0.4887541	0.4205673	-0.4861766	-0.4779288
income_group_no_aid	0	1	0	1	0	0
income_group_pell_eligible	1	0	1	0	1	1
income_group_state_grant_eligible	0	0	0	0	0	0
sex_female	0	1	1	1	0	0
sex_male	1	0	0	0	1	1
residency_international	0	0	0	0	0	1
residency_non_resident	0	1	0	0	0	0
residency_resident	1	0	1	1	1	0

# Explore the Data (Again Not Real Data)



# Explore the Data (Again Not Real Data)



# Model the Outcome: Split Into Test and Training Sets

retn_train\$retained	n	percent
0	412	0.6158445
1	257	0.3841555

retn_test\$retained	n	percent
0	137	0.6171171
1	85	0.3828829

# Model the Outcome: Build Basic Regression Model (This is Hacky and Poorly Specified)

```
mod.1 <- glm(retained ~  
              total_peer_group +  
              family_income +  
              first_gen_first_gen,  
              data = retn_train,  
              family = "binomial")
```

# Review and Interpret the Results

term	estimate	std.error	statistic	p.value
(Intercept)	4.59998	0.19685	7.75218	0.00000
<b>total_peer_group</b>	<b>0.72952</b>	<b>0.07324</b>	<b>-4.30613</b>	<b>0.00002</b>
<b>family_income</b>	<b>2.11703</b>	<b>0.15248</b>	<b>4.91882</b>	<b>0.00000</b>
<b>first_gen_first_gen</b>	<b>0.06219</b>	<b>0.21840</b>	<b>-12.71770</b>	<b>0.00000</b>

## Interpretation (Fake Data)

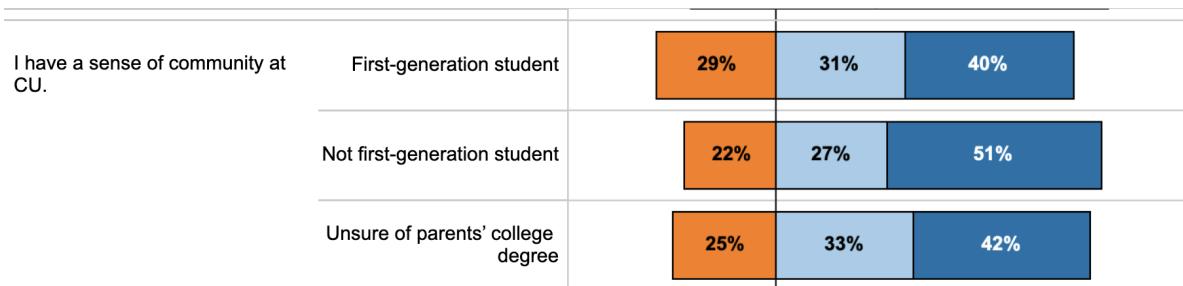
Being classified as first-gen\_first\_gen is correlated with being 0.6 times less likely to be retained in the second year when controlling for other factors in the model.

For each additional 'unit' of family\_income, a student is 2.11 times more likely to be retained in the second year when controlling for other factors in the model.

# Surveys and Data Collection

Go Ask Them!

This set of questions addresses your experiences with CU Boulder overall. Indicate how strongly you disagree or agree with each of the following statements:



- Agree/ Strongly agree
- Somewhat agree
- Somewhat disagree/ Disagree

## 4 Chart Types

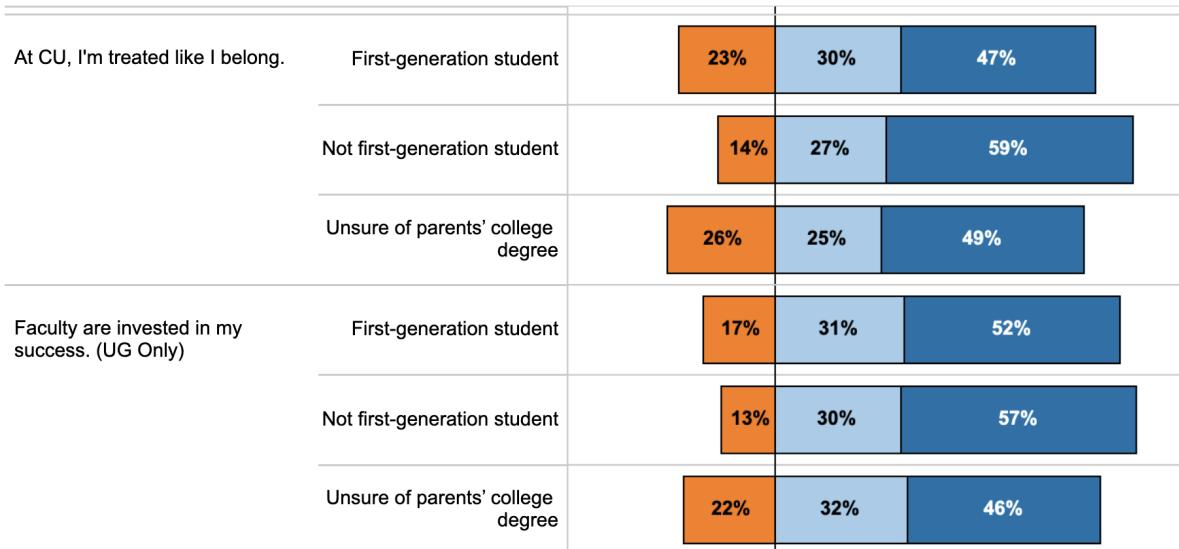
**Change Chart Type:**

- Display Distribution
- Display Averages

**First-generational students disagree that they have a sense of community at CU Boulder 7% more than non-first-generational students.**

## Belonging at CU Boulder

This set of questions addresses your experiences with CU Boulder overall. Indicate how strongly you disagree or agree with each of the following statements:



### Legend:

click to highlight on chart

- Agree/ Strongly agree
- Somewhat agree
- Somewhat disagree/ Disagree

## 4 Chart Types

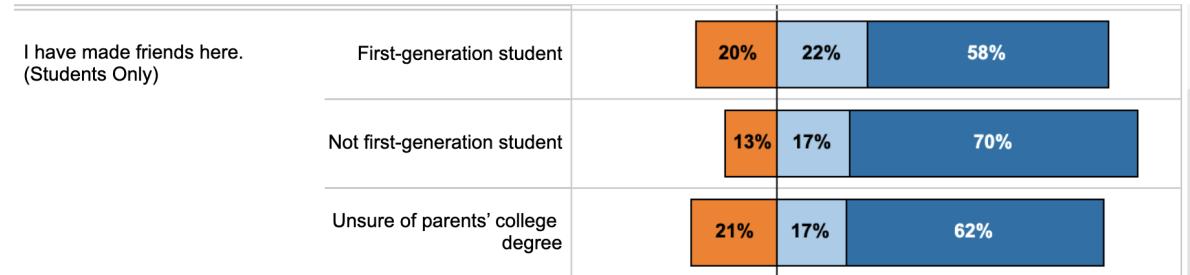
### Change Chart Type:

- Display Distribution
- Display Averages

**First-generational students disagree that are treated like they belong 9% more than non-first-generational students.**

## Belonging at CU Boulder

This set of questions addresses your experiences with CU Boulder overall. Indicate how strongly you disagree or agree with each of the following statements:



### Legend:

click to highlight on chart

  Agree/ Strongly agree

  Somewhat agree

  Somewhat disagree/ Disagree

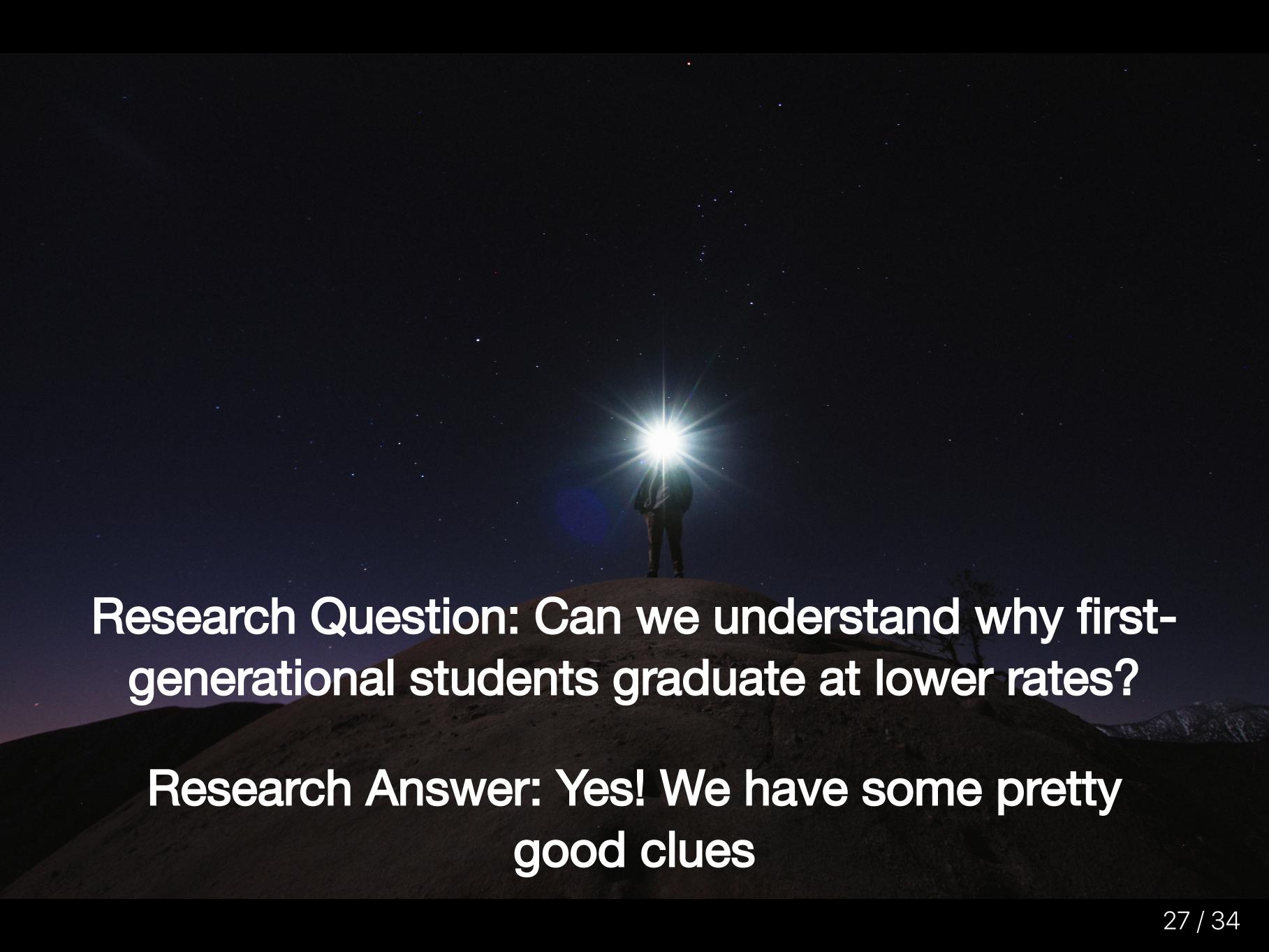
## 4 Chart Types

### Change Chart Type:

Display Distribution

Display Averages

**First-generational students disagree that they have made friends here 7% more than non-first-generational students.**



**Research Question: Can we understand why first-generational students graduate at lower rates?**

**Research Answer: Yes! We have some pretty good clues**

- ✓ Retention and graduation are key indicators of student and institutional success
  - ✓ Differential rates in retention and graduation are inequitable and demand attention
- ✓ Theoretical evidence posits many theories as to why first-generational students are less likely to persist until graduation.
  - ✓ Our fake data and clunky model specification indicated two things:
    - ✓ That family income is also highly correlated and something institutions can impact through financial aid. And...
    - ✓ That students from first-generational backgrounds feel less connected, less engaged and less likely to have friends or support.
  - ✓ So what can we do about it?

A wide-angle photograph of a green soccer field. In the upper left foreground, a white soccer goal is positioned on the grass. A player in a blue shirt and black shorts is seen from behind, kicking a black soccer ball towards the goal. The field has white boundary lines and a small number '12' is visible near the top left corner of the goal area. The background shows a paved path or road curving away from the field.

**Universities have diffuse and sometimes competing goals**

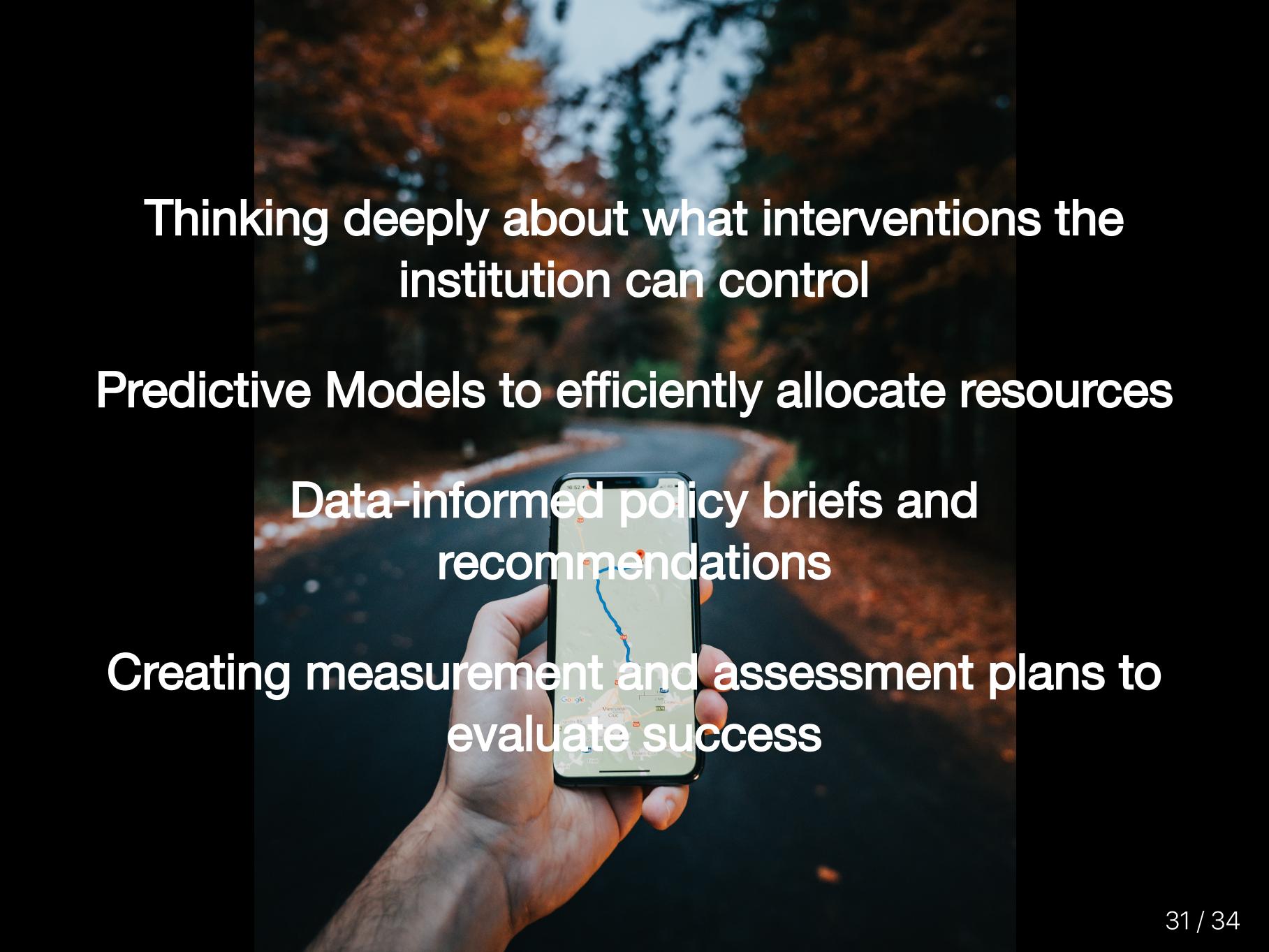
**Universities have myriad stakeholders**

**Universities have to demonstrate our value to taxpayers and to society**

**Universities are resource constrained**



# How Do We Use Data To Improve the University?



Thinking deeply about what interventions the institution can control

Predictive Models to efficiently allocate resources

Data-informed policy briefs and recommendations

Creating measurement and assessment plans to evaluate success

# We Need Your Help

**The public and non-profit sectors need descriptive, predictive, and prescriptive analytics to advance prosocial outcomes**

**These sectors aren't as far along as the private sector. This provides lots of room for innovation!**

**Find a business problem you care about. For me, that's helping colleges and universities, but there are endless ways to make a difference.**

# Some Blue Sky Projects in ODA and Higher Ed

**NLP Model to Detect Incidents of Self-Harm in Survey Responses**

**IoT Experiment for Rapid, Random Surveys via Smart Phone**

**Deep Learning exploration of optimal curricular pathways**

**Recommender Systems for Courses, Instructors, Majors**

**Prescriptive Models for Guiding "Next Likely Action" for students and advisors**

**[Your Great Idea Here]**

**Thank you: Marcos, Marie, Jordyn, Poom, Professor Larsen and the MSBA Team for the Invitation**

**This slide deck was created using R, Rmarkdown and the Xaringan Package**

**All photos are freely available from Unsplash.com**

**Errors, Typos, and Oopsies Are Mine. Please let me know if you see something wacky**

**Code and Slides available at**

**[bradweiner.info/talk](http://bradweiner.info/talk)**

