

# IR 101 DATA VISUALIZATION

CAIR 2020

Scott Heil  
UC Riverside  
scott.heil@ucr.edu

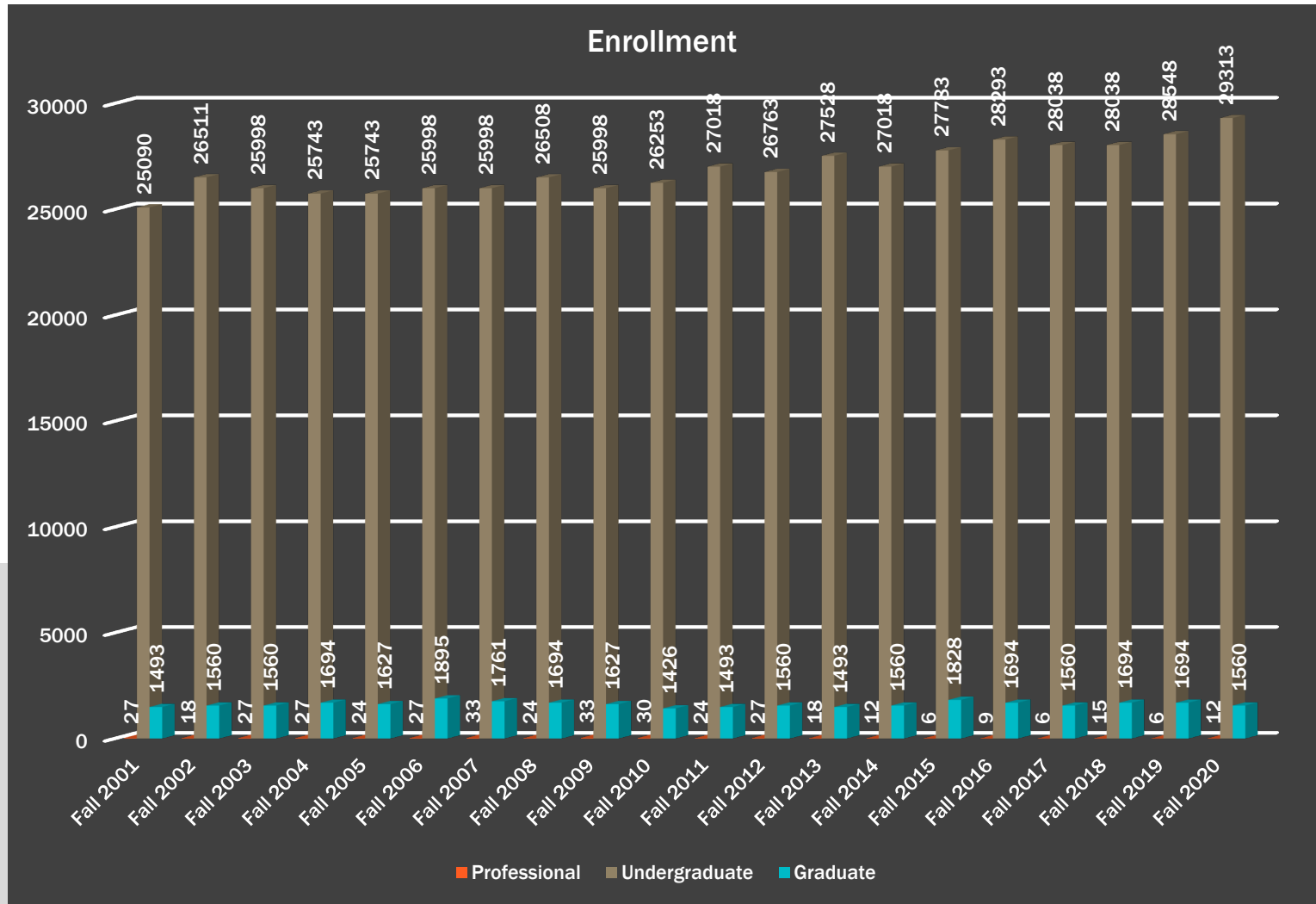


# **OVERVIEW OF THIS SESSION**

- **Platform agnostic**
- **Focus on concepts, not technical how-to**
- **Practical examples, including some that work well and some that need help**

- **Examine how data types relate to visual conventions**
- **Top 10 tips for upping your visualization game**
- **A few suggested resources**

**DO THESE WORK?**



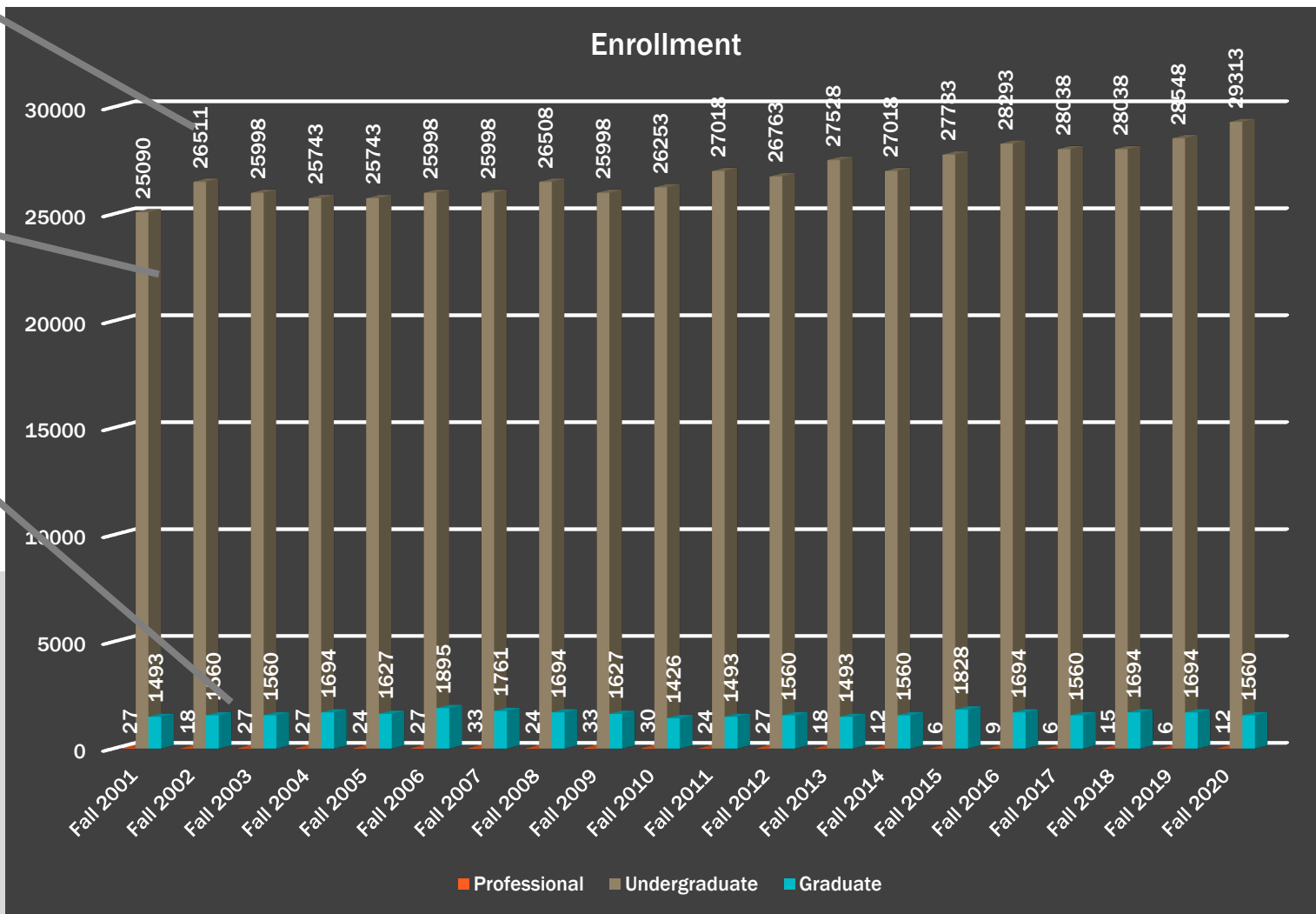
*Simulated data*

Both rotated labels and small text are hard to read

3-D adds no meaning

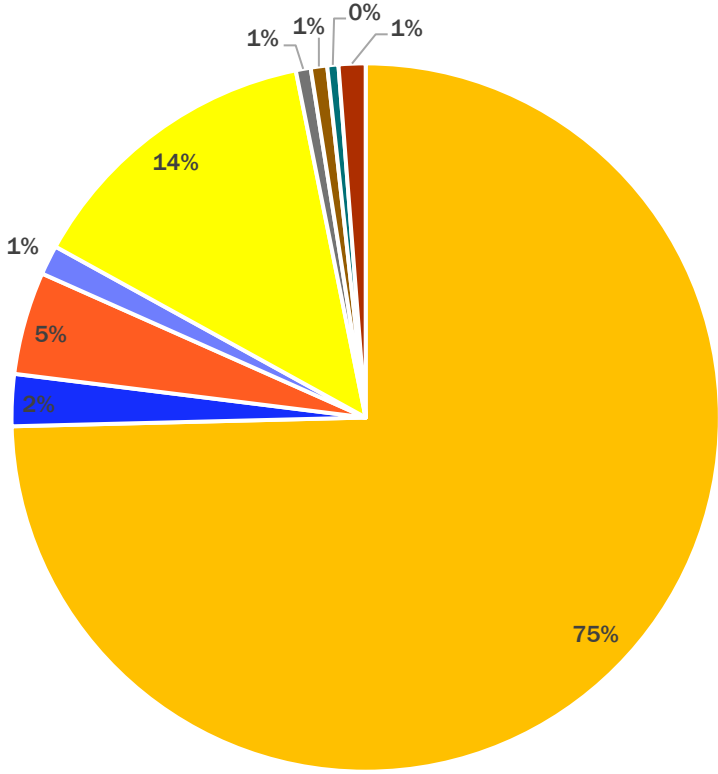
Hard to see any patterns in the smaller categories

Overall effect is low readability and high clutter



Simulated data

Faculty by College



- Literature Arts and Sciences
- Law School
- School of Medicine
- Research Foundation
- College of Engineering
- Architecture
- Graduate School of Urban Policy
- Journalism and Professional School
- College of Social Work

*Simulated data*

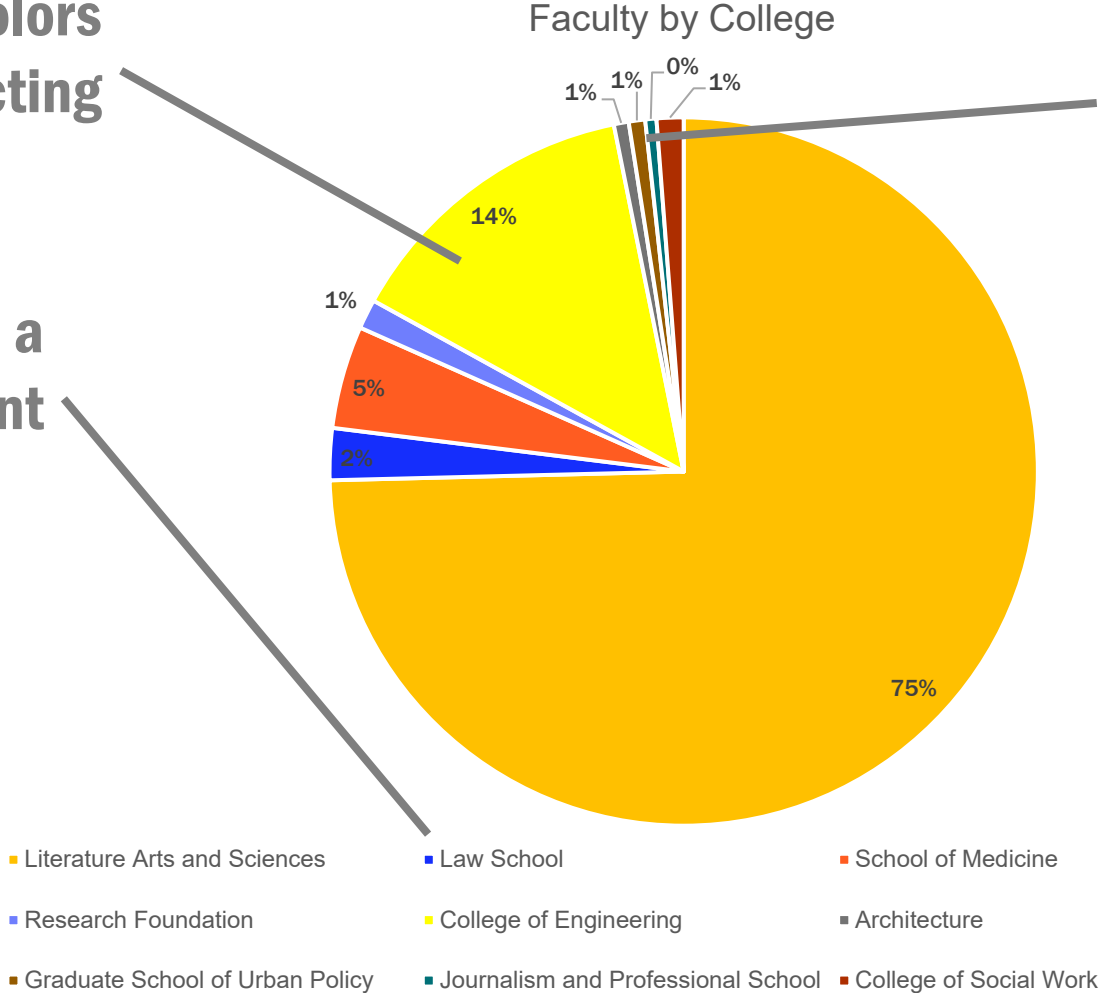


# This uses a lot of space to convey very little information

Rainbow of strong colors can be distracting

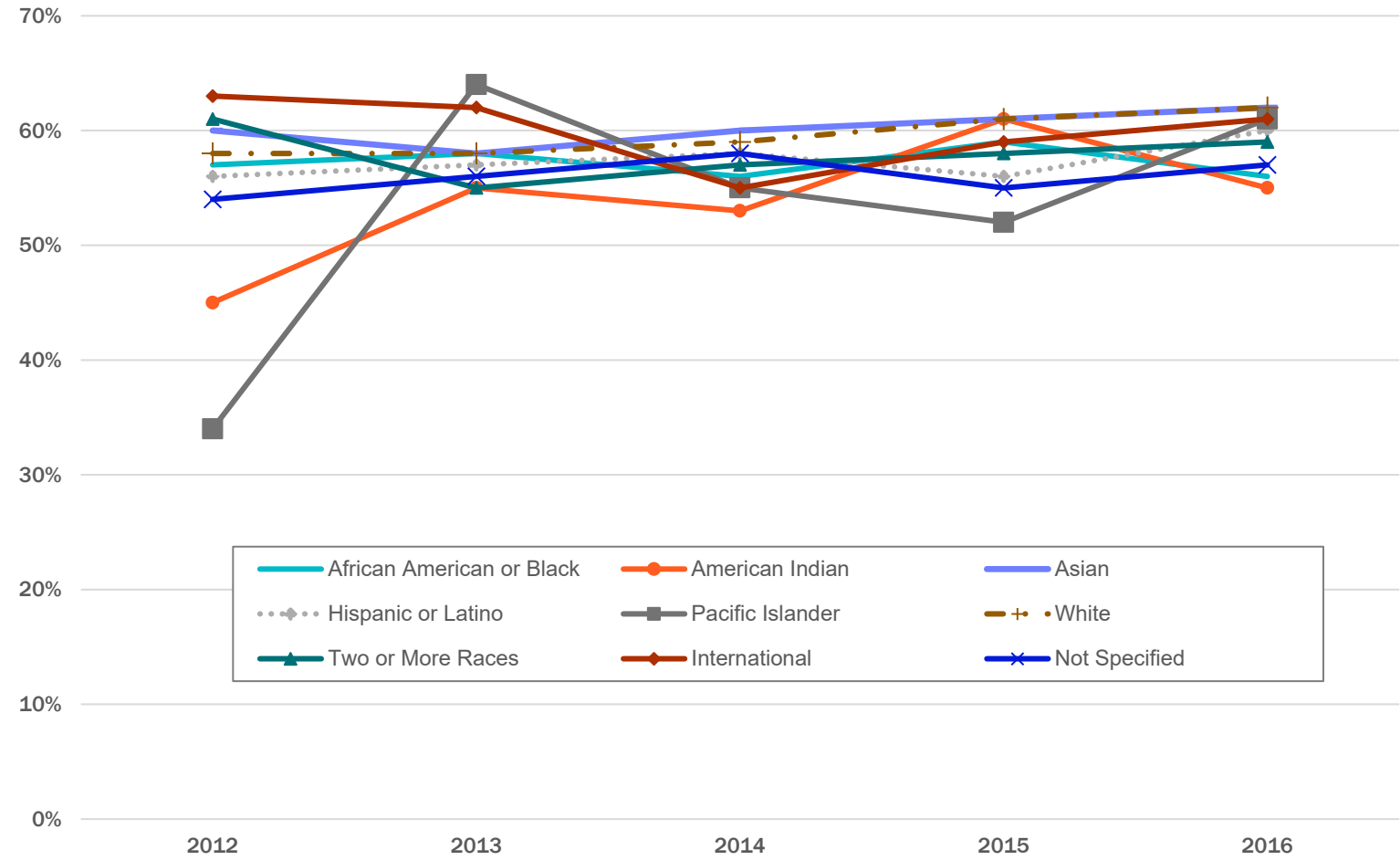
Small categories are hard to make out

Legend sends the reader on a tedious color-matching hunt



*Simulated data*

4-Year Graduation Rate by Ethnicity



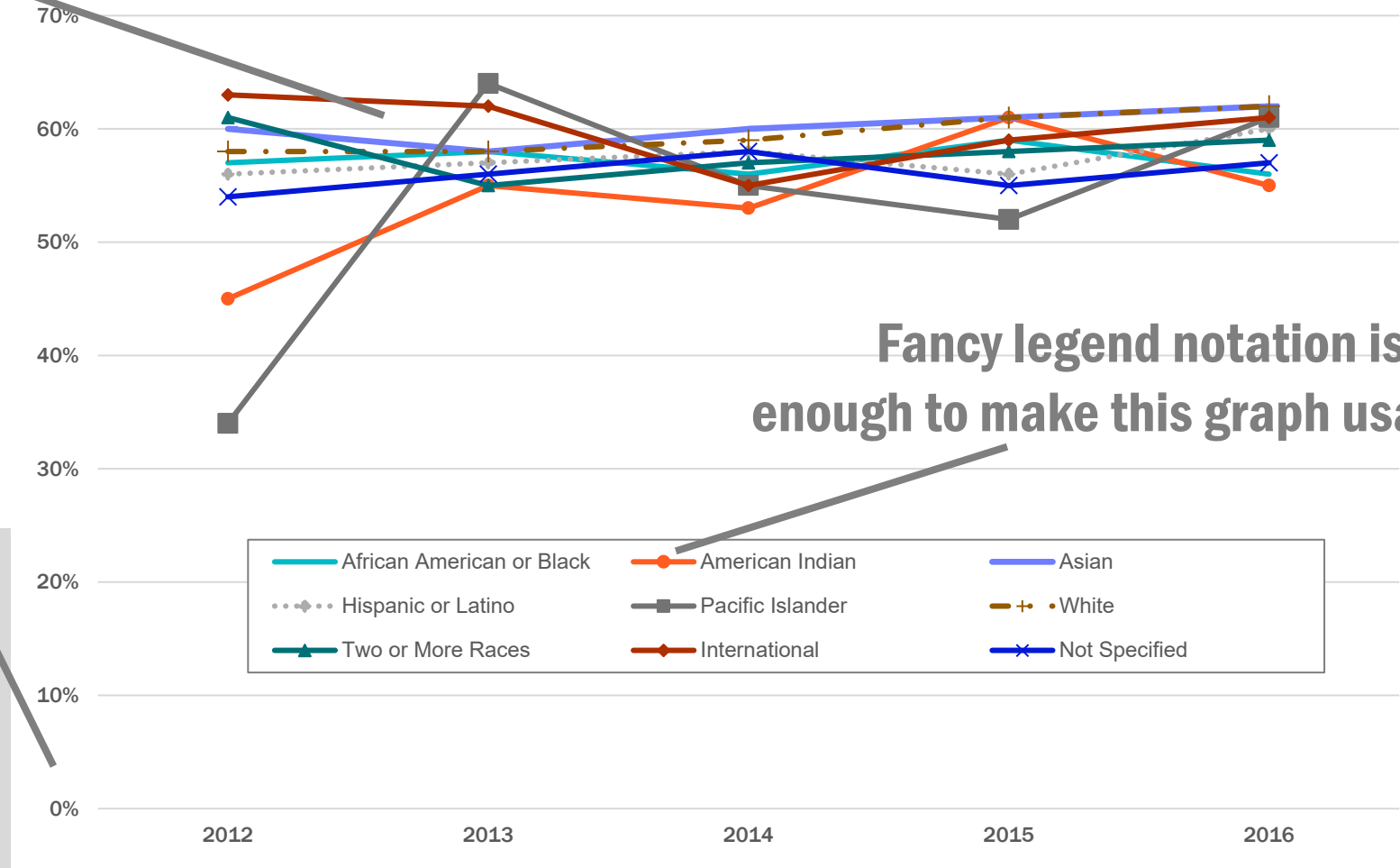
*Simulated data*

# Most of this graph is wasted space

Extremely difficult to follow a tangle of lines like this

Starting from zero is not necessary for trend lines and crowds them closer together

4-Year Graduation Rate by Ethnicity



Fancy legend notation is not enough to make this graph usable

*Simulated data*

**WERE THOSE HARDER  
THAN THEY NEEDED  
TO BE?**

**A FEW SAMPLES THAT  
WORK WELL**

## Responses by Race/Ethnicity and Year at UC

Category	Year	#	At least somewhat dissatisfied	Percentage of respondents by response category	At least somewhat satisfied	Select Year(s) Multiple values
Afr Amr	2006	1,328	21%		79%	Characteristics Race/Ethnicity
	2018	2,190	25%		75%	
Amr Ind	2006	341	18%		82%	
	2018	369	21%		79%	
Asian	2006	22,590	22%		78%	
	2018	20,993	24%		76%	
His/Lat	2006	7,103	15%		85%	
	2018	14,558	21%		79%	
Interntnl	2006	1,199	20%		80%	
	2018	5,021	19%		81%	
White	2006	19,700	13%		87%	
	2018	13,916	18%		82%	



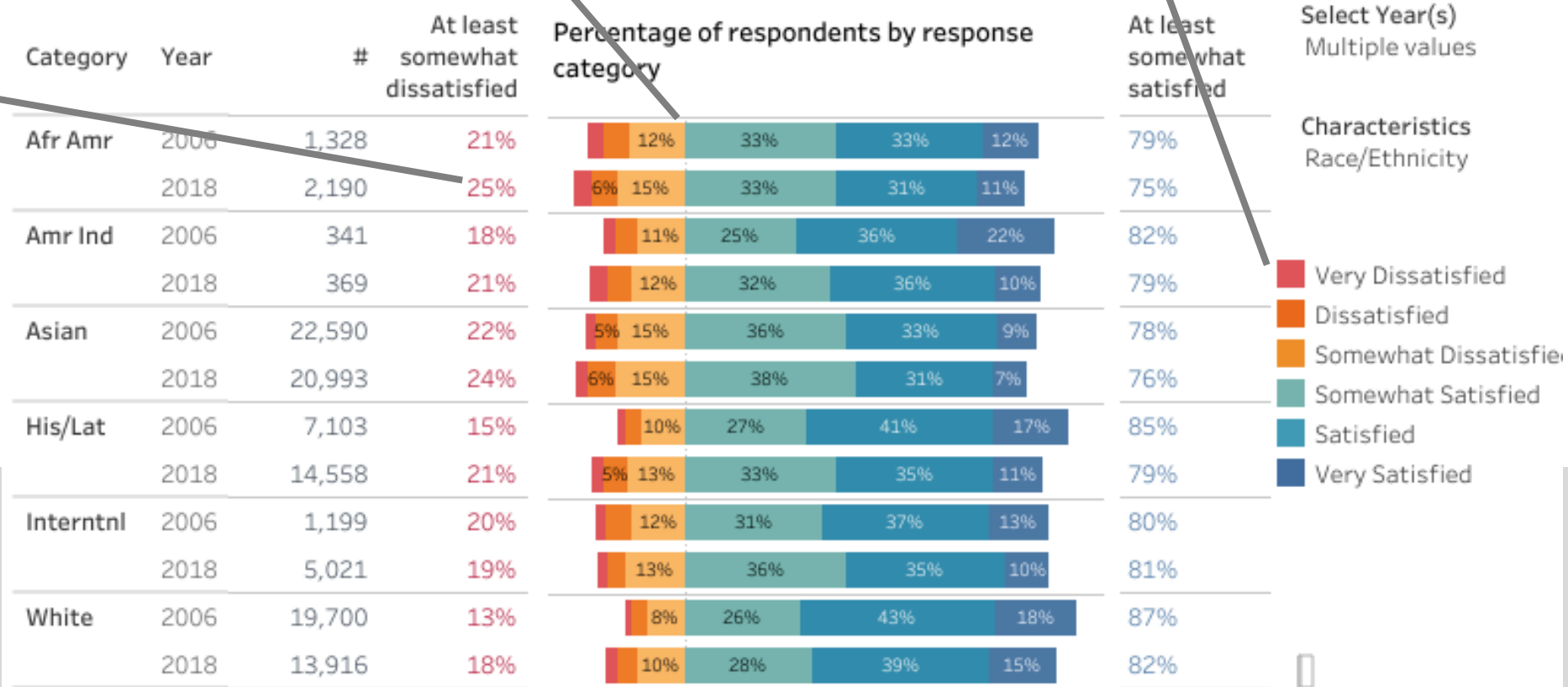
Source: University of California, Office of the President, 2018

Scale centered at the midpoint

Colors are meaningful and match visual expectations

Provides multiple levels of summary and detail along with change over time

Responses by Race/Ethnicity and Year at UC

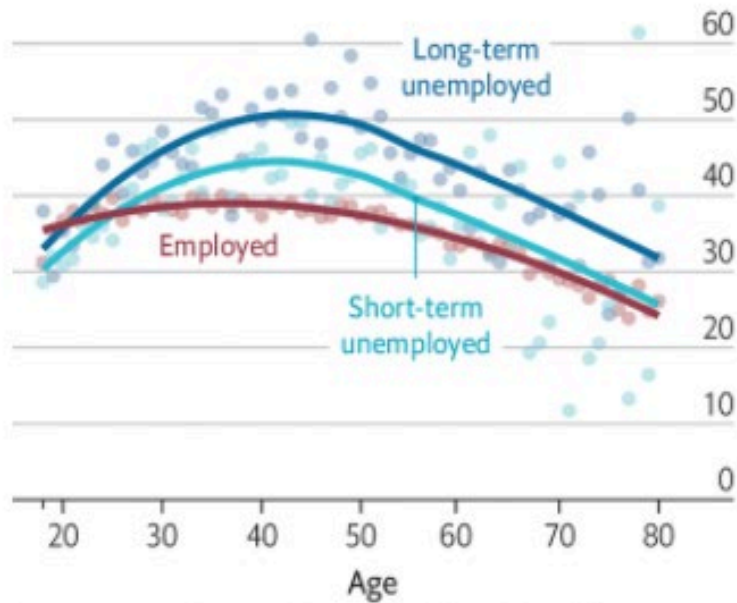


Source: University of California, Office of the President, 2018

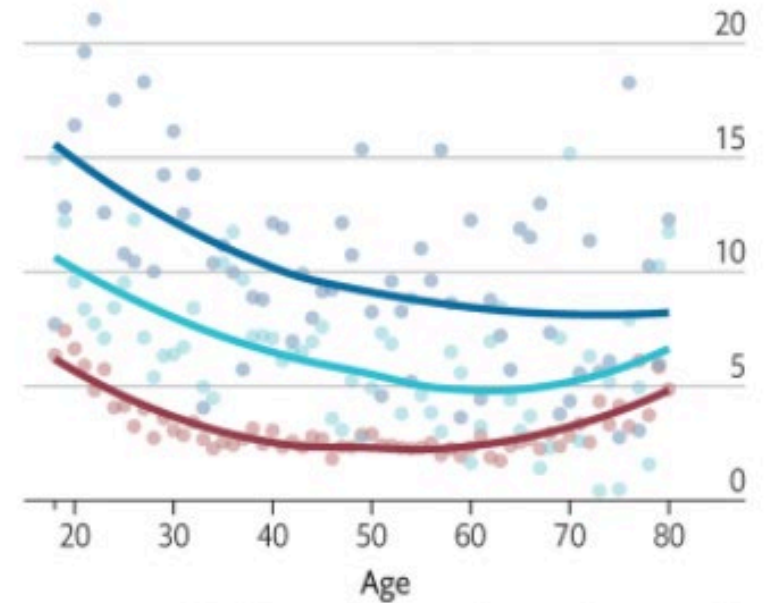
## Not worth losing sleep over

Sleep length per night by age and employment status

Less than seven hours, % of total\*



Ten hours or more, % of total†



Source: "Unemployment disrupts sleep" by D. Blanchflower and A. Bryson, NBER working papers, 2020

\*Excludes people getting at least ten hours per night  
†Excludes people getting less than seven hours per night

The Economist

Source: The Economist, September 2020



# Reduces complex pattern to illustrate the message

Signals the main message in the title

Definitions and context in the subtitle

Coordinated use of color

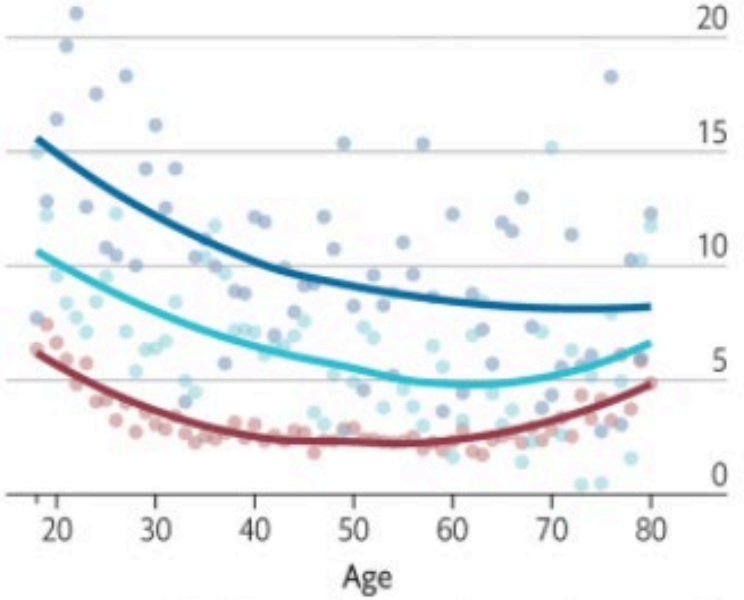
## Not worth losing sleep over

Sleep length per night by age and employment status

Less than seven hours, % of total\*



Ten hours or more, % of total†



Source: "Unemployment disrupts sleep" by D. Blanchflower and A. Bryson, NBER working papers, 2020

\*Excludes people getting at least ten hours per night  
†Excludes people getting less than seven hours per night

The Economist

Source: The Economist, September 2020

## REVENUE AND MARGIN GROWTH SCENARIOS Assuming 9% membership growth.

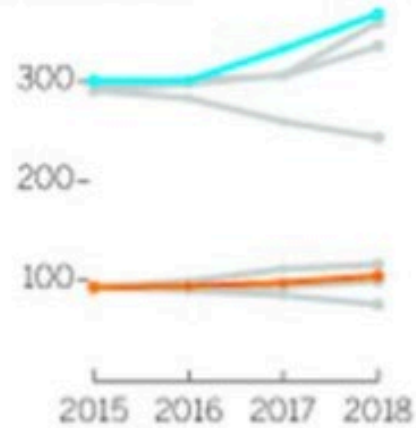
### STATUS QUO

\$400 million



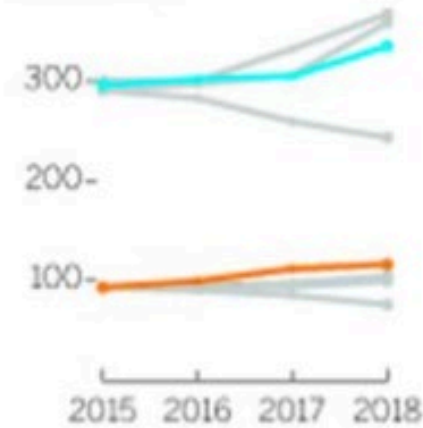
### SCENARIO 1

\$400 million



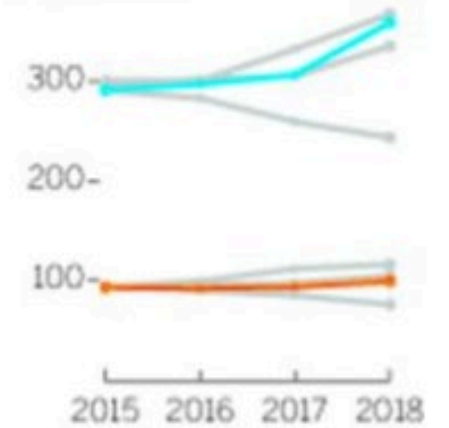
### SCENARIO 2

\$400 million



### SCENARIO 3

\$400 million

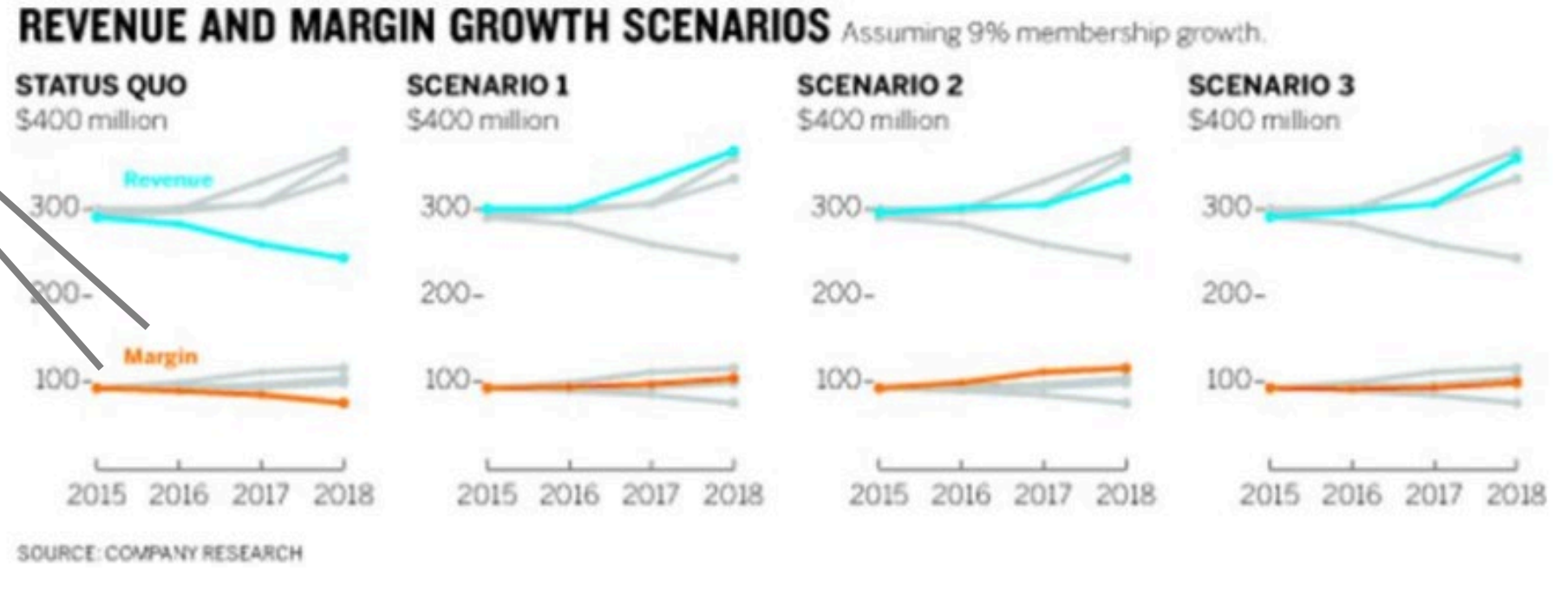


SOURCE: COMPANY RESEARCH

Source: Scott Berinato, Harvard Business Review, August 2016

Isolates each scenario with the others in the background

Coordinated use of color, no need for separate legend



Source: Scott Berinato, Harvard Business Review, August 2016

## DEMOGRAPHICS



Traditionally  
Underrepresented



Pell Grant  
Recipient



1st in Family  
to Earn Degree



1st Generation to  
Earn Degree in Family

## ETHNICITY

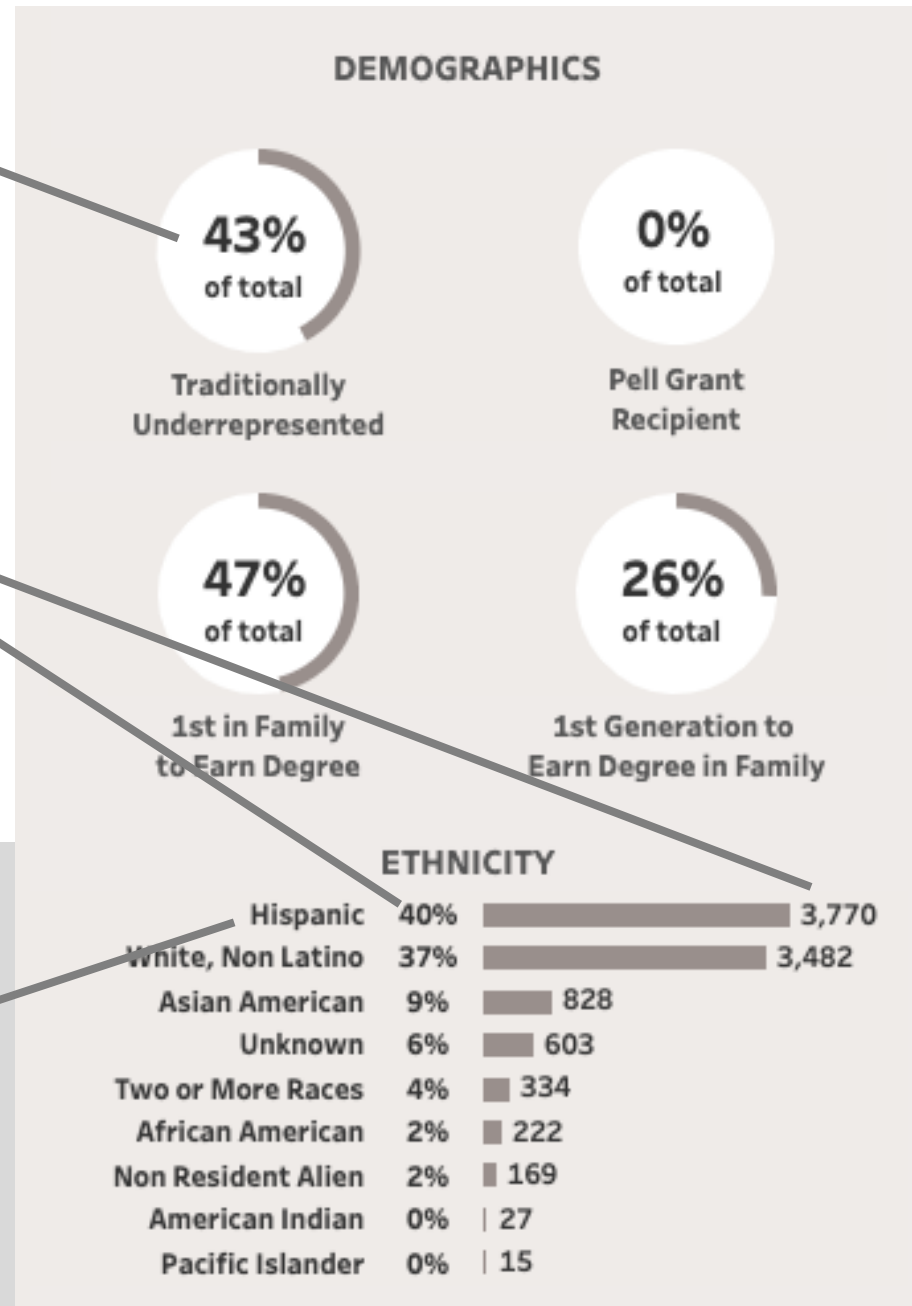
Hispanic	40%	3,770
White, Non Latino	37%	3,482
Asian American	9%	828
Unknown	6%	603
Two or More Races	4%	334
African American	2%	222
Non Resident Alien	2%	169
American Indian	0%	27
Pacific Islander	0%	15

Source: California State University, Chancellor's Office, 2020

Simple, highly readable  
chart types and text

Focus on illustrating  
essential information  
without excess formatting

Ranked listing  
quickly shows  
which are the  
largest and  
smallest groups



**WHAT MAKES A VISUAL  
SUCCESSFUL?**

- **Sound quantitative reasoning**
- **Clear message or focus**
- **Intuitive and easy to read**
- **Aesthetically pleasing**

# **DATA TYPES AND VISUAL REPRESENTATIONS**



# Nominal comparisons

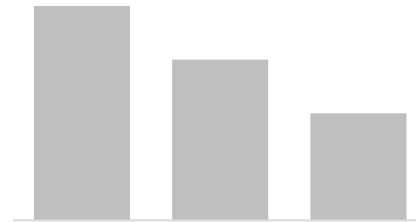
- **Categorical**
- **No natural order**
- **Typical examples: demographic groups, student types, academic units**

# Nominal examples

Horizontal bar



Column



Scatterplot/Bubble

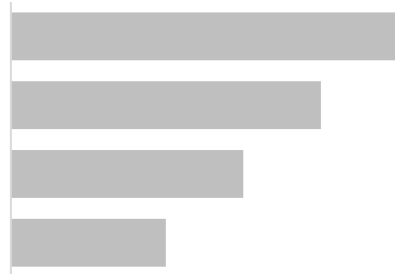


# **Ordinal comparisons**

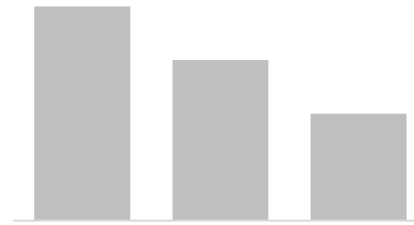
- **Clear sequence of levels**
- **Typical examples: Likert scales, age brackets, score ranges, letter grades**

# Ordinal examples

Horizontal bar



Column



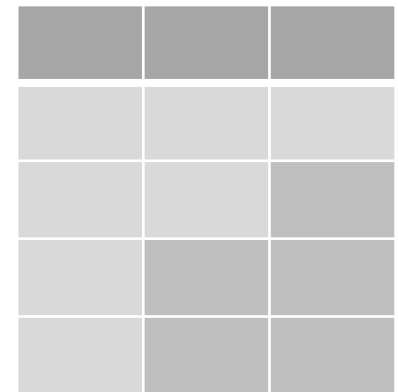
Scatterplot



Stacked



Heatmap



# Proportions of the whole

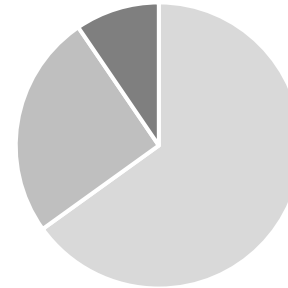
- **Measures that are expressed as percentages**
- **Typical examples: Demographic distributions, graduation rates, retention rates, admission rates**

# Proportion examples

Horizontal bar



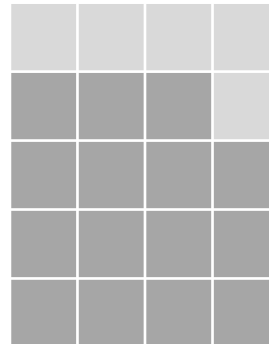
Pie



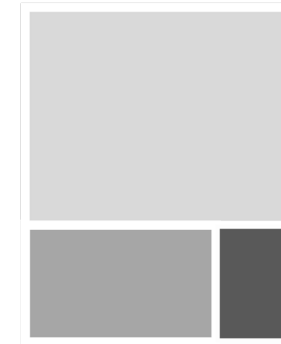
Donut



Waffle/Gridplot



Treemap



# Time series

- **Multiple measurements over some time interval**
- **Typical examples: Annual graduation rates, enrollment by term, demographic change over time**

# Time series examples

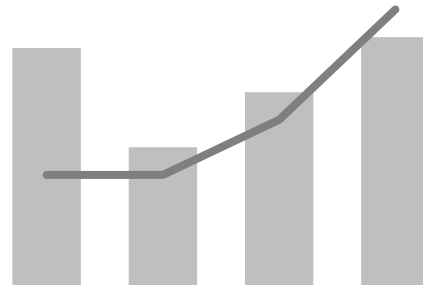
Line



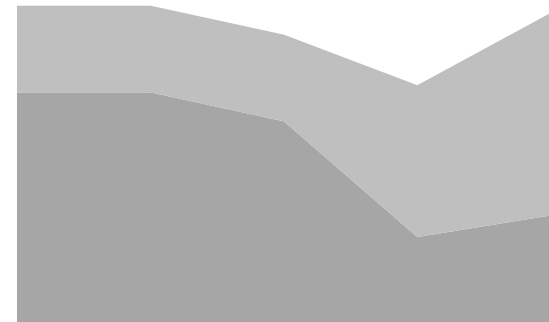
Scatterplot/Bubble



Combination



Area





# Correlations

- **One or more quantitative measures**
- **Typical examples: Subgroup comparisons, associations across multiple measures**

# Correlation examples

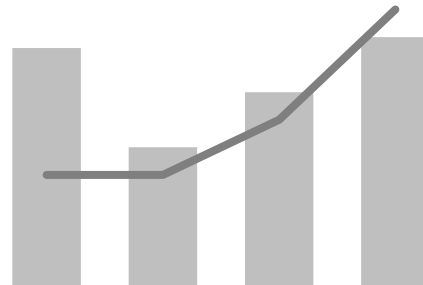
Multiple Line



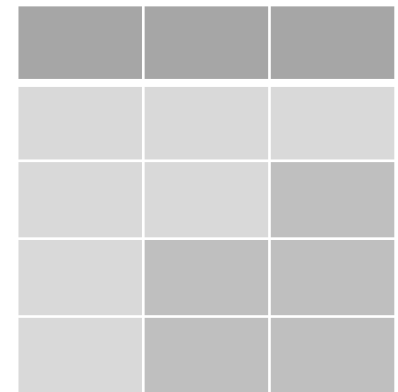
Scatterplot/Bubble



Combination



Heatmap



# TIPS FOR BETTER DATA GRAPHICS



# Use ranked lists

Easier to use



Requires more effort



\*Avoid what Edward Tufte called the “tyranny of the alphabet”

(But this may not be appropriate for ordinal categories)

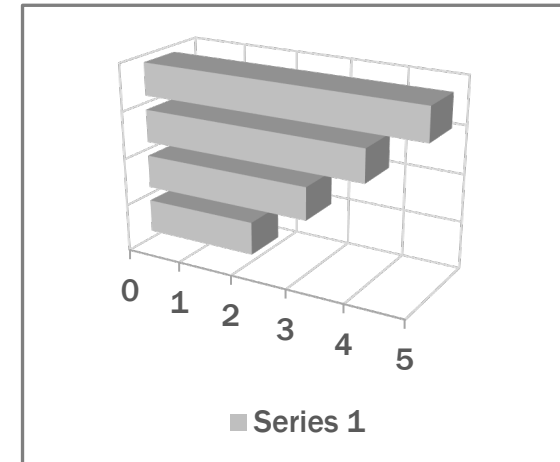


# Remove clutter

Just the essentials



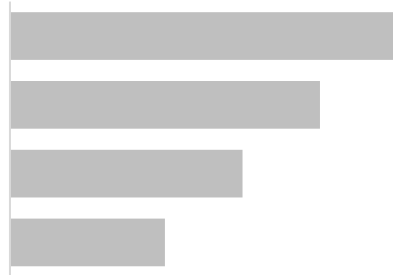
More formatting, but not necessarily more meaning





# Small graphs can be effective

Reasonable size for four  
data points



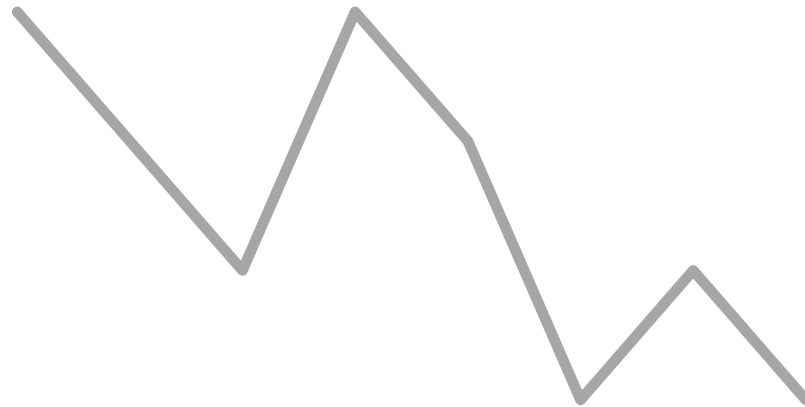
Bigger does not always  
improve readability





# Ensure the visual matches the message

Here the visual conclusion is lots of change



But in this format the message is stability over time

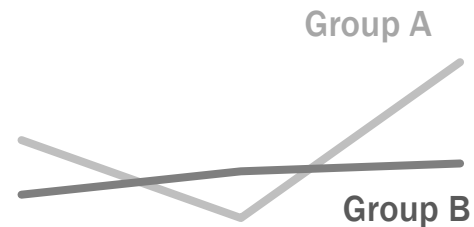


(Data values are identical for the two graphs)



# Make the legend part of the graph

Here the meaning is instantly clear and the reader does not have to hunt for anything in a separate legend





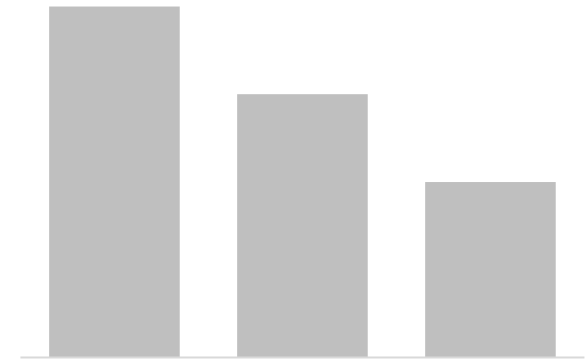


# Zero is not needed in trend line graphs

Nothing is lost here if the y-axis does not begin at zero



But for bars and columns zero is necessary





# Try small multiples

In this format the reader can quickly scan across multiple series and spot similarities and differences

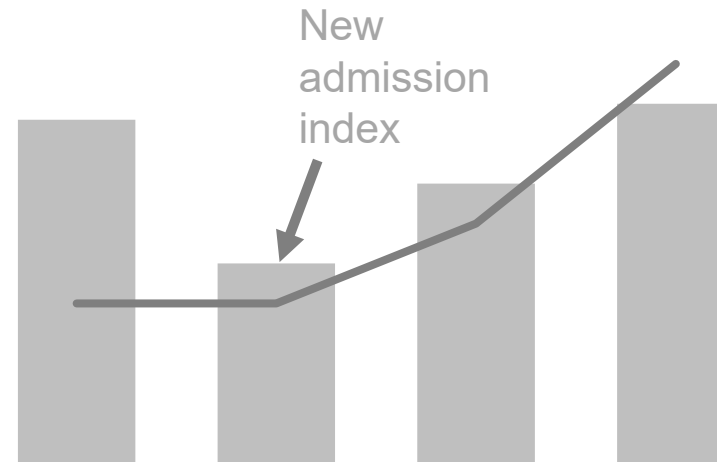




# Use descriptive titles and annotations

## Incoming GPA vs. Retention Rate

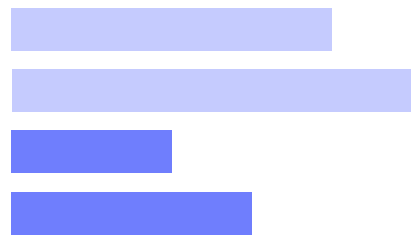
Cohort grade point average compared with one-year fall-to-fall retention



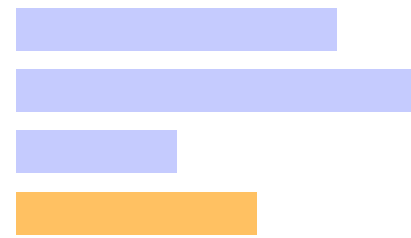


# Be deliberate in the use of color

Gradients of the same color often convey the meaning well



Higher contrast colors can be reserved for highlights and special information

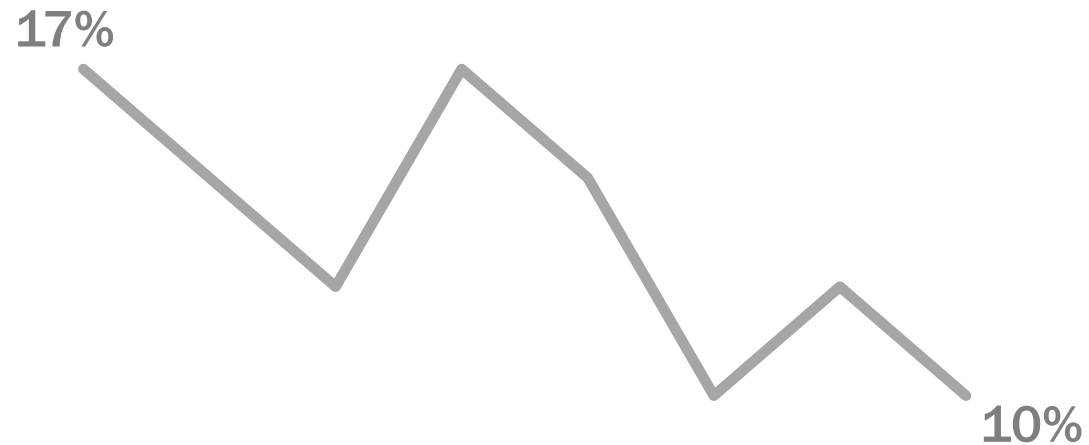


The rainbow effect is often a distraction and a lost opportunity to add meaning





# Not every data point needs to have a label



Labeling key points with easily readable text is almost always better than many points with tiny labels

**WRAP-UP**

# Closing tips

- **Let the meaning of the data help drive your visualizations**
- **Simple is often better than complicated**
- **Favor readability and intuitive understanding over complexity**

# RESOURCES



# Cheat sheets

- *Core Principles of Data Visualization* at [policyviz.com](https://policyviz.com)
- *Visual Vocabulary* at [ft.com](https://ft.com)
- *Data Visualization Cheatsheet* at [kaggle.com](https://kaggle.com)

# Books

- *Good Charts*, Scott Berinato, 2016
- *Effective Data Visualization*, Stephanie Evergreen, 2019

**THANK YOU**