

EXAMPLES OF MISLEADING VISUALIZATIONS AND HOW IR CAN SUPPORT PREVENTION

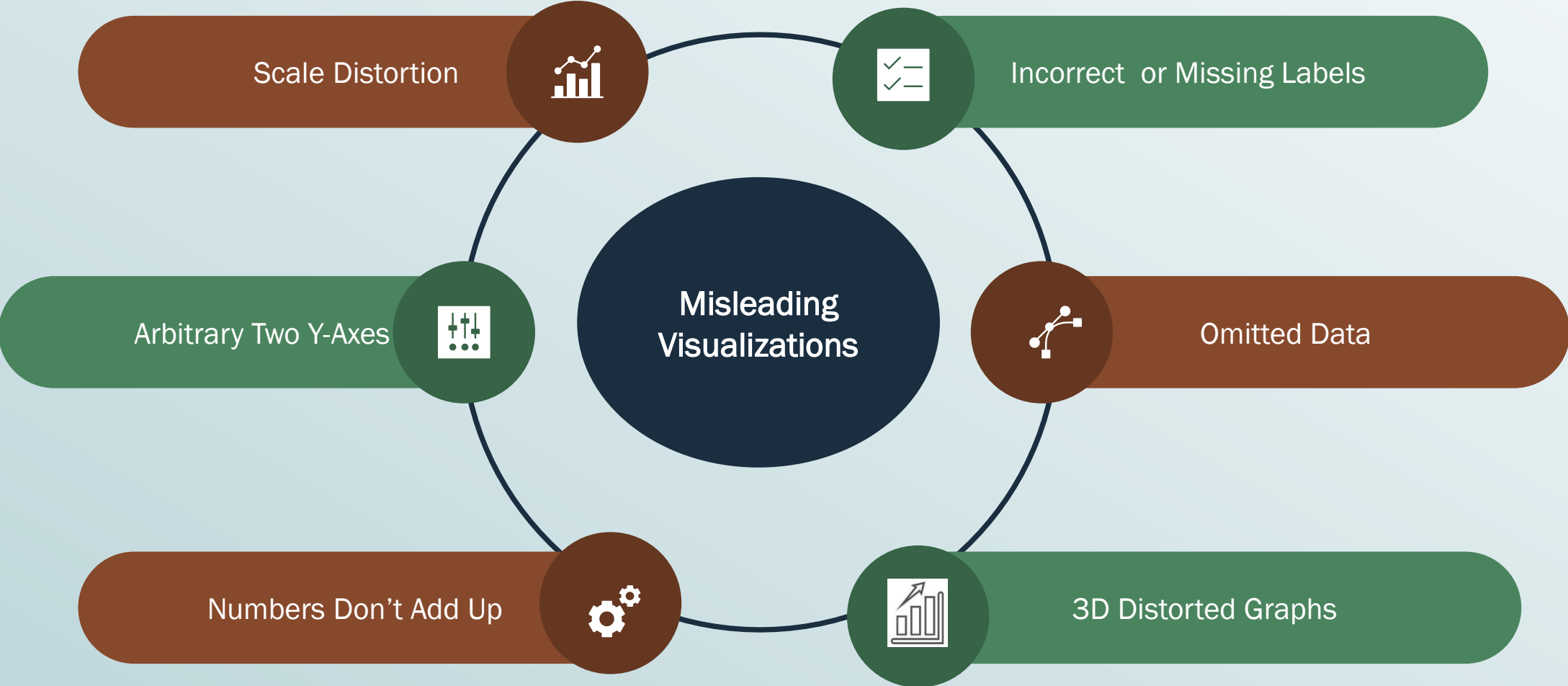
Heidi M. Carty, Ph.D., Research Analyst IV, IR, UCSD

Galina Belokurova, Ph.D., Assistant Director of Operations, IR, UCSD

Outline

1. Introduction of Presenters
2. Importance of Accurate Visualizations
3. Most Common Types of Misleading Visuals
4. Are the Statistics Correct?
5. Vertical Scale Distortion
6. Data Not Adding Up
7. Arbitrary Dual Y-Axes
8. Missing or Incorrect Labels
9. 3D Distortion
10. Concealed or Omitted Data
11. How varying definitions can change data visualizations

Misleading Visuals

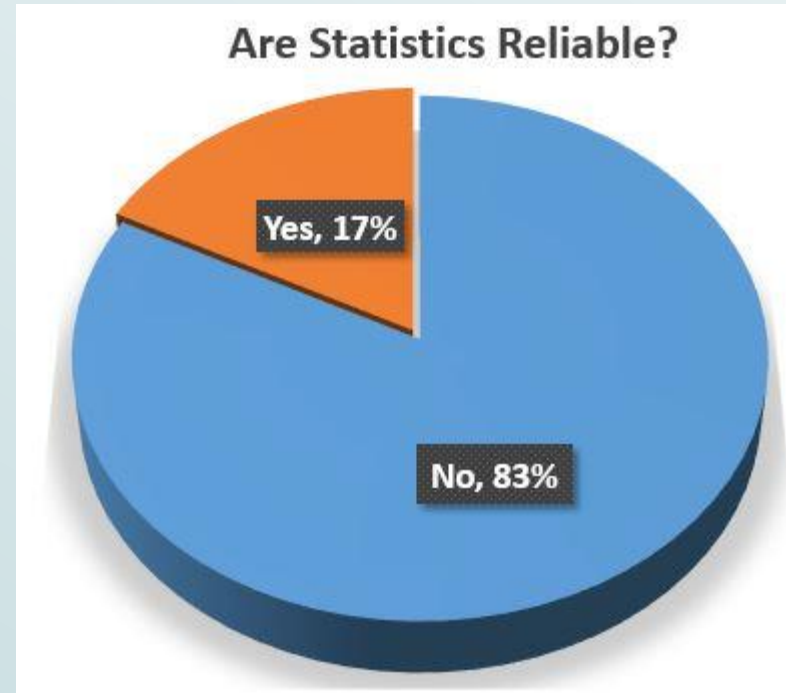


Institutional Research utilizes data to inform leadership, administrators, faculty, students, parents and the public about who we are, where we have been and where we are going as an institution in a variety of ways. The accuracy of our data, and statistics and their visualizations are critical to portray an accurate picture.

- Are the data accurate?

83% of published statistics are incorrect. Really? Can we trust the data?

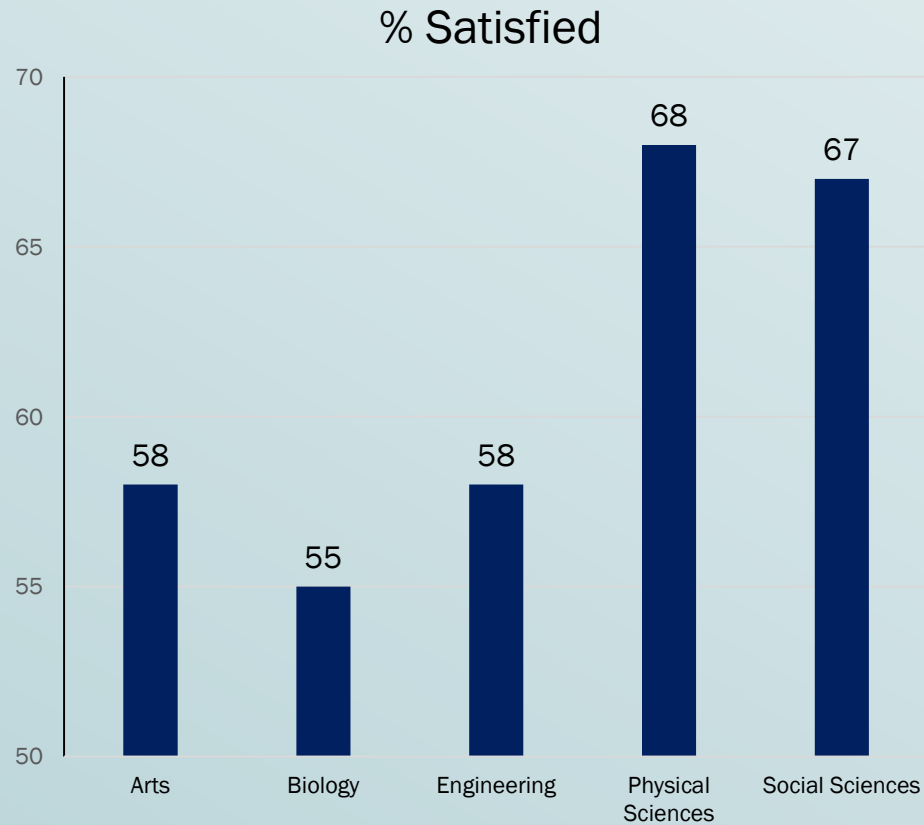
- Check your source. Look for consistency.



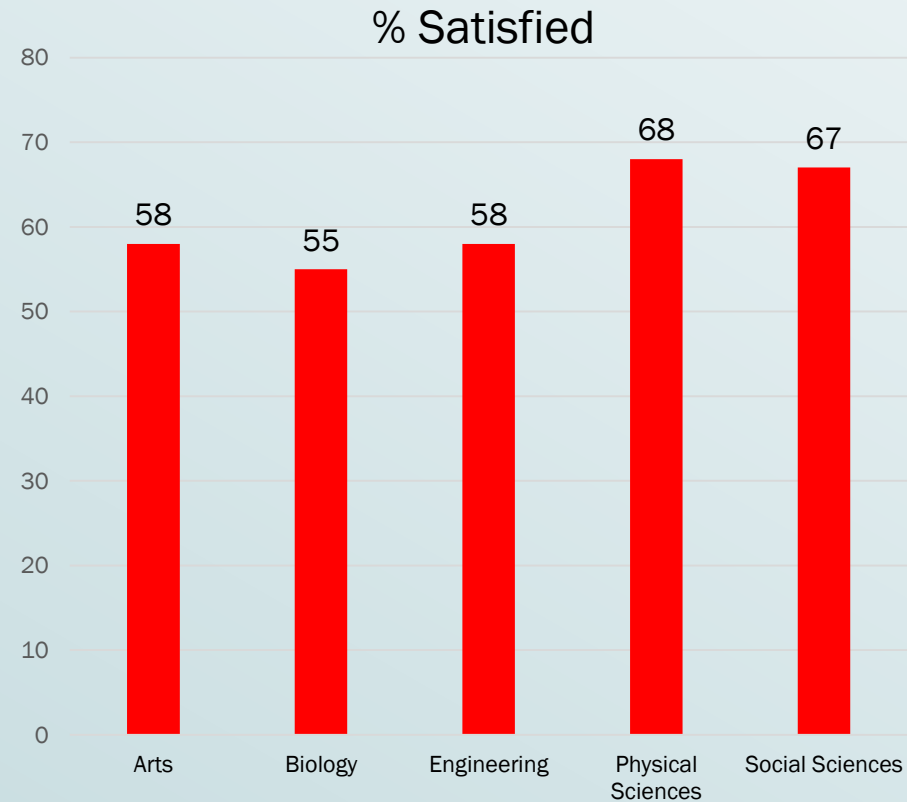
Scale Distortion

Vertical axis not starting at zero or skipping numbers – truncated y-axis leads to distorted patterns and over-exaggerated differences, suggesting differences that do not exist. In most cases, the baseline should start at zero.

Bad



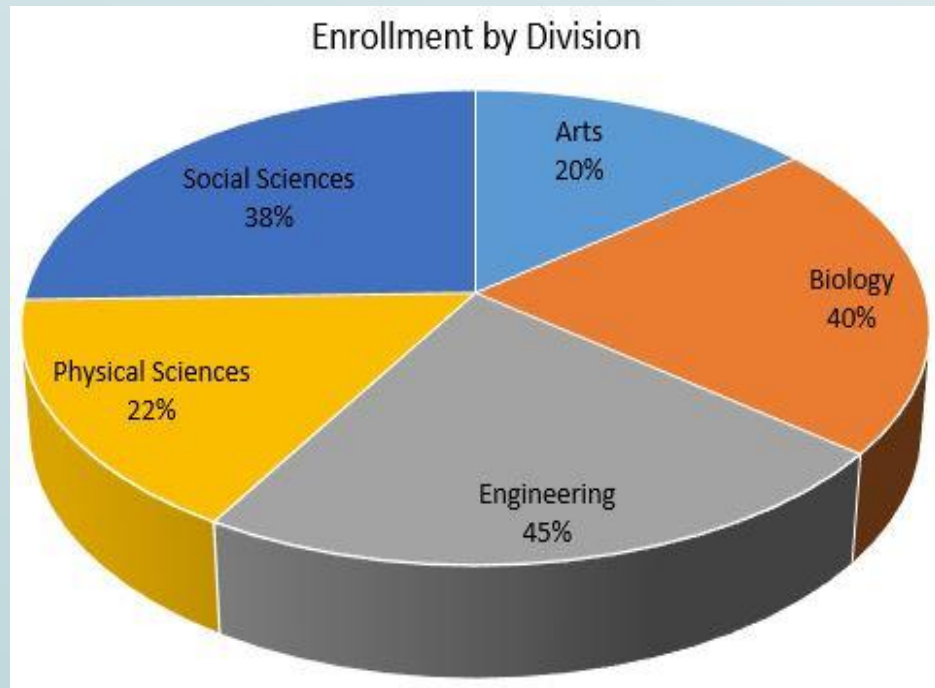
Good



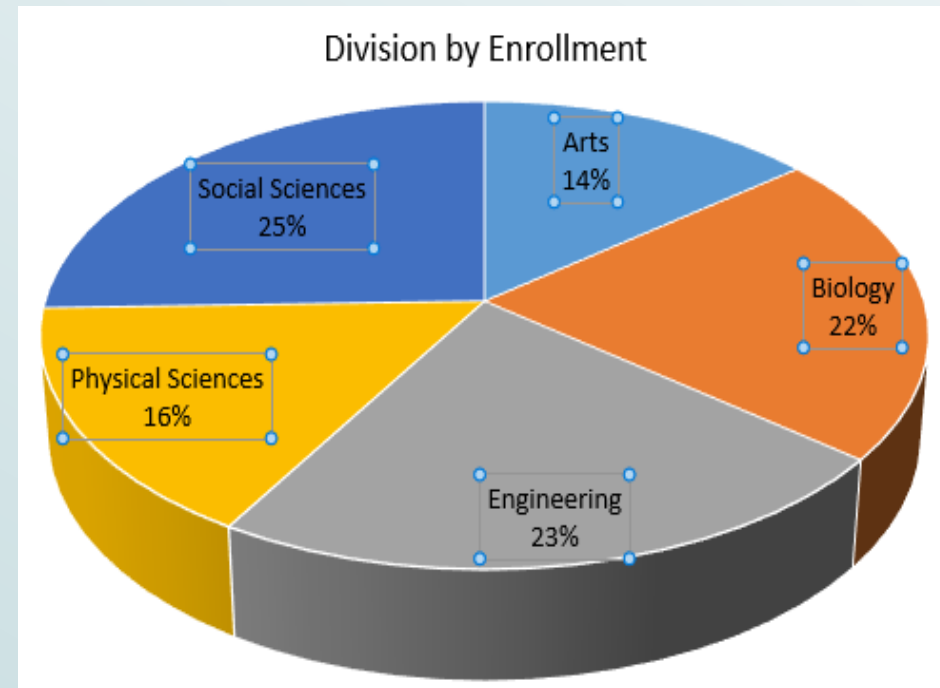
Data Not Adding Up

Graphs that don't add up! A pie chart are used to show the parts of a whole, not the difference between groups. Pie charts should always add up to 100%.

Bad

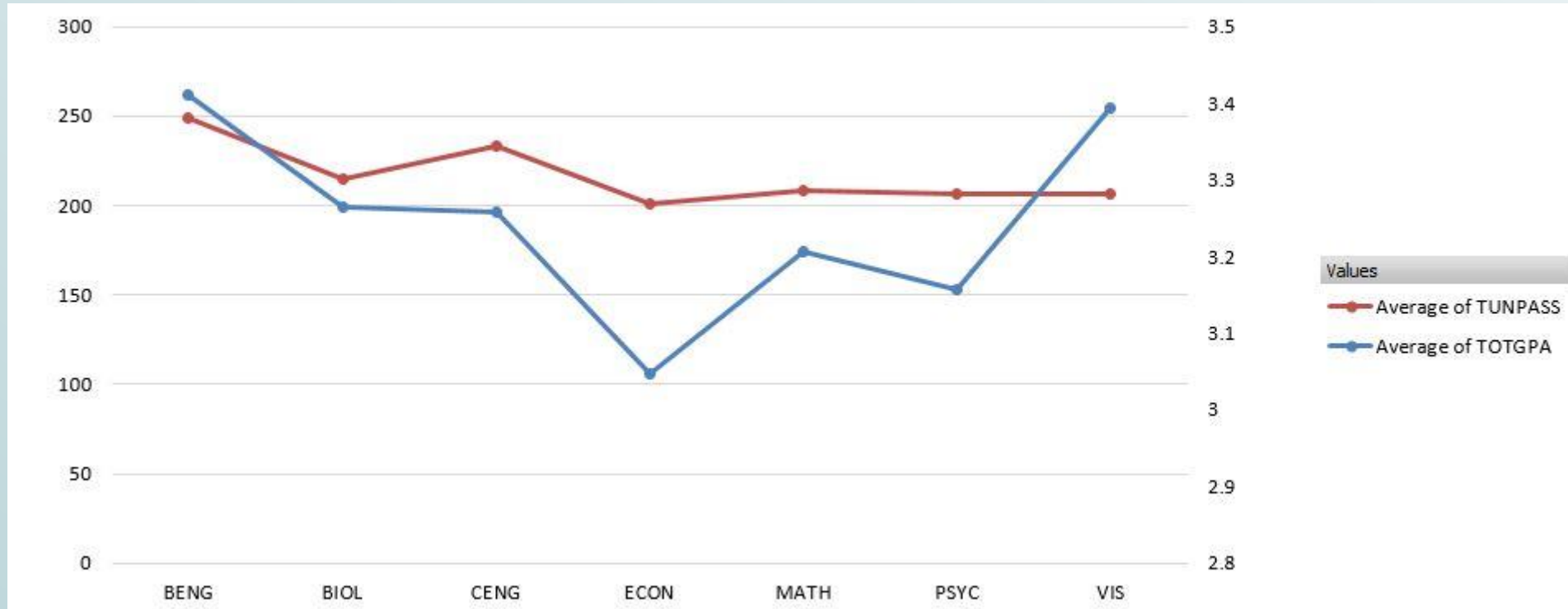


Good



• Arbitrary Dual Y-Axes •

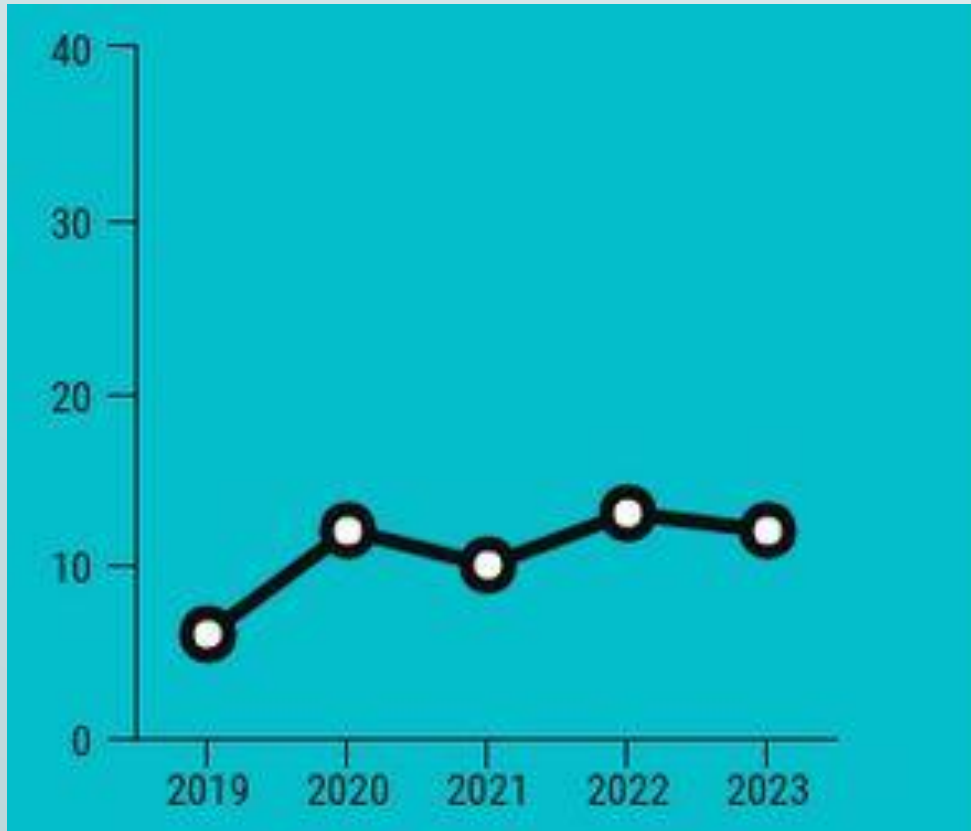
Dual y-axes – the scales of dual axes can be set arbitrarily and deliberately mislead readers about the relationship between the two data series. Small difference in the blue GPA field (GPA) appear to be large while large differences red line (units passed) appear small.



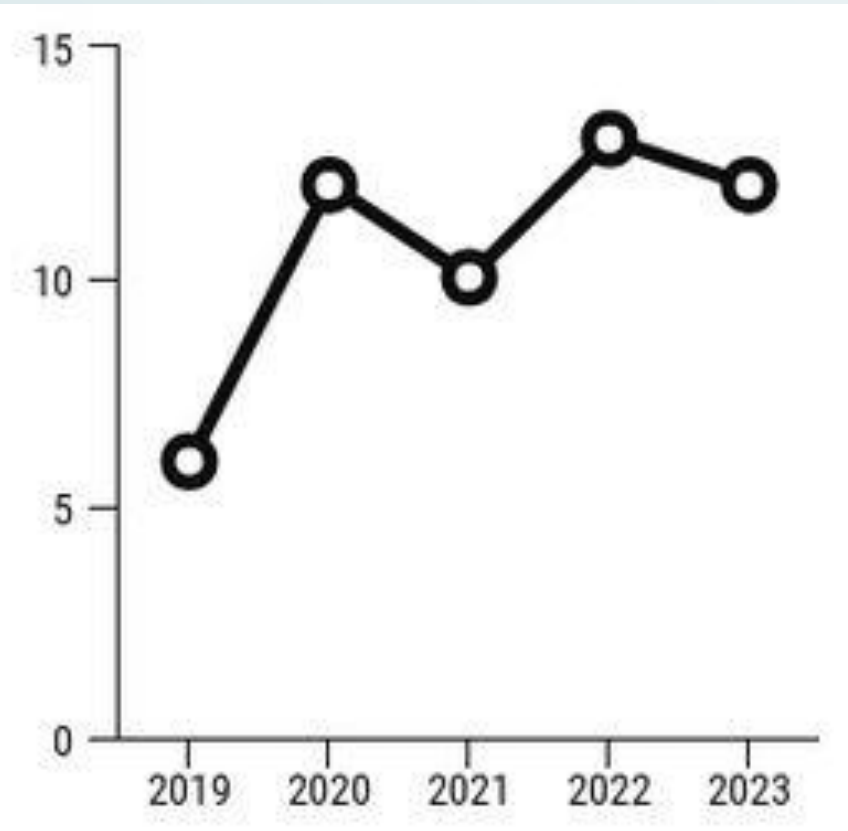
Manipulated Y-Axis

Compressing or expanding the scale skews how the data are perceived by presenting a false impression of the data. The scale is disproportionate to the data making change over time seem smaller. It would also be helpful to have labels showing the actual values by year.

Bad

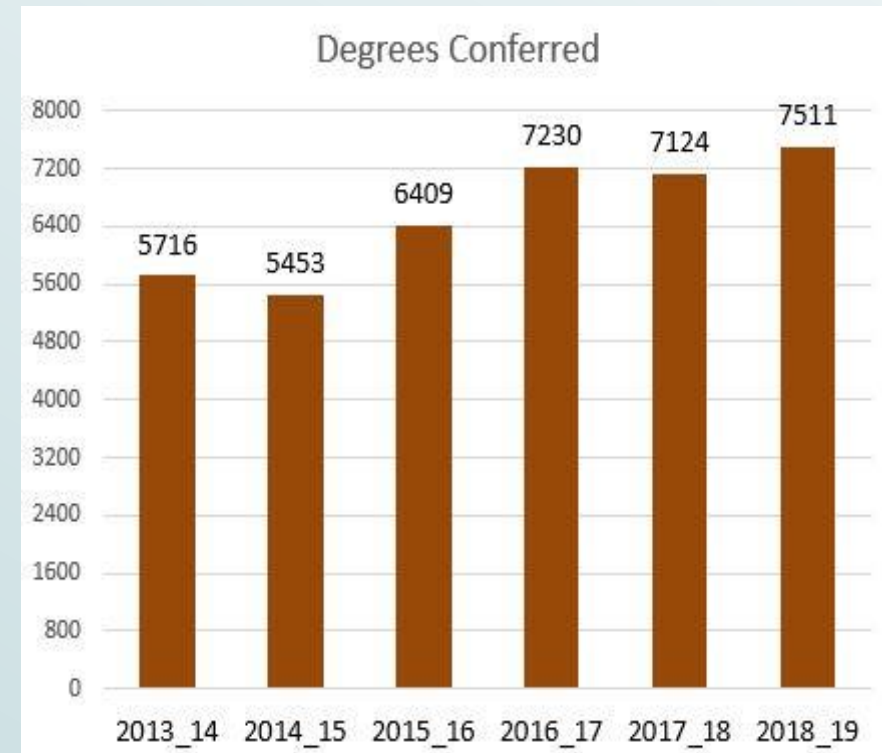
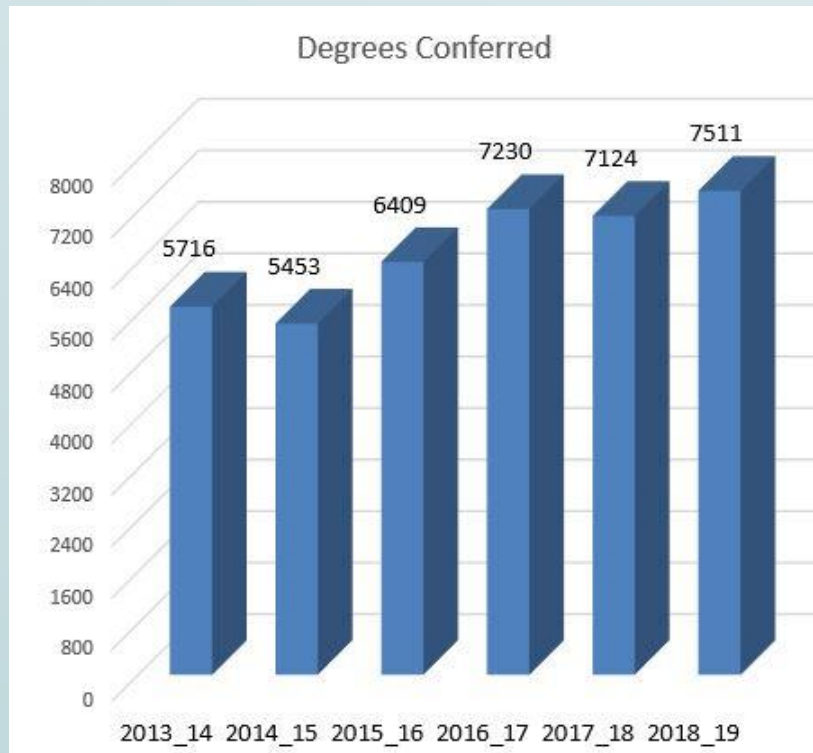


Good



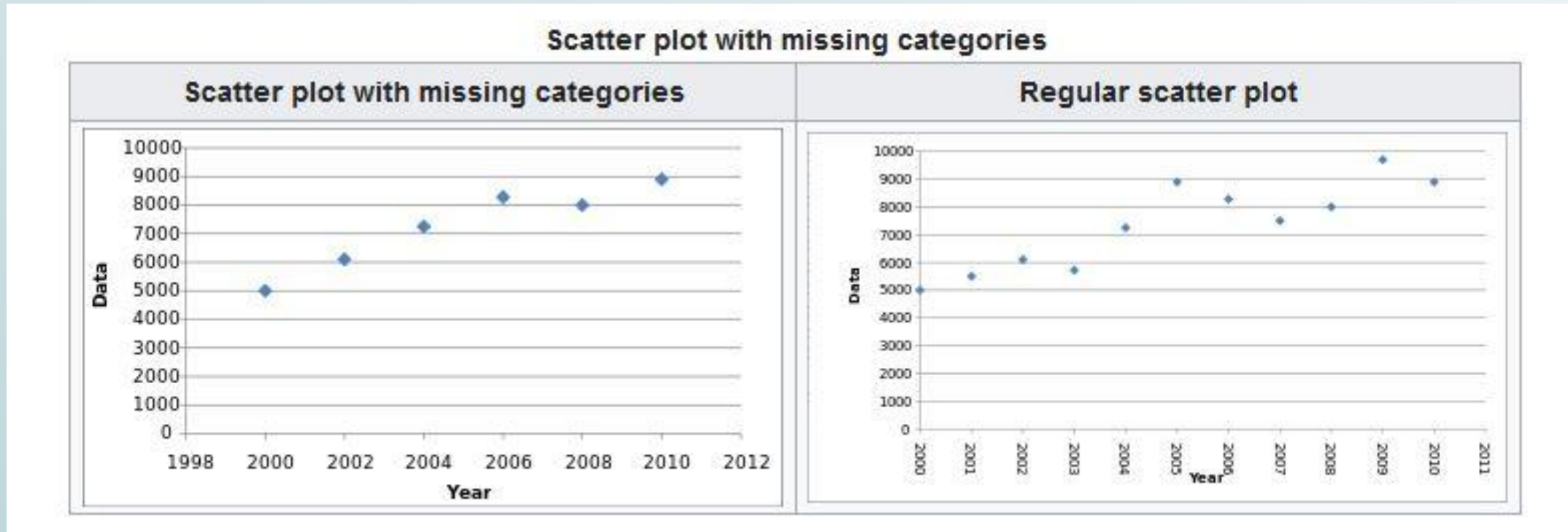
3D Distortion

3D Graphs look cool and add dimension to an otherwise boring bar chart; however, they can distort the bars making differences appear larger compared to a 2D graphic.



• Omitted or Concealed Data •

Omitted or concealed data can be either intentional (cherry picking) or unintentional. Researchers may only include data which reinforces their narrative. This can provide a false representation of the data.



Varying Definitions

Experimenting with varying the definition of our data and the change in statistics and visualizations.

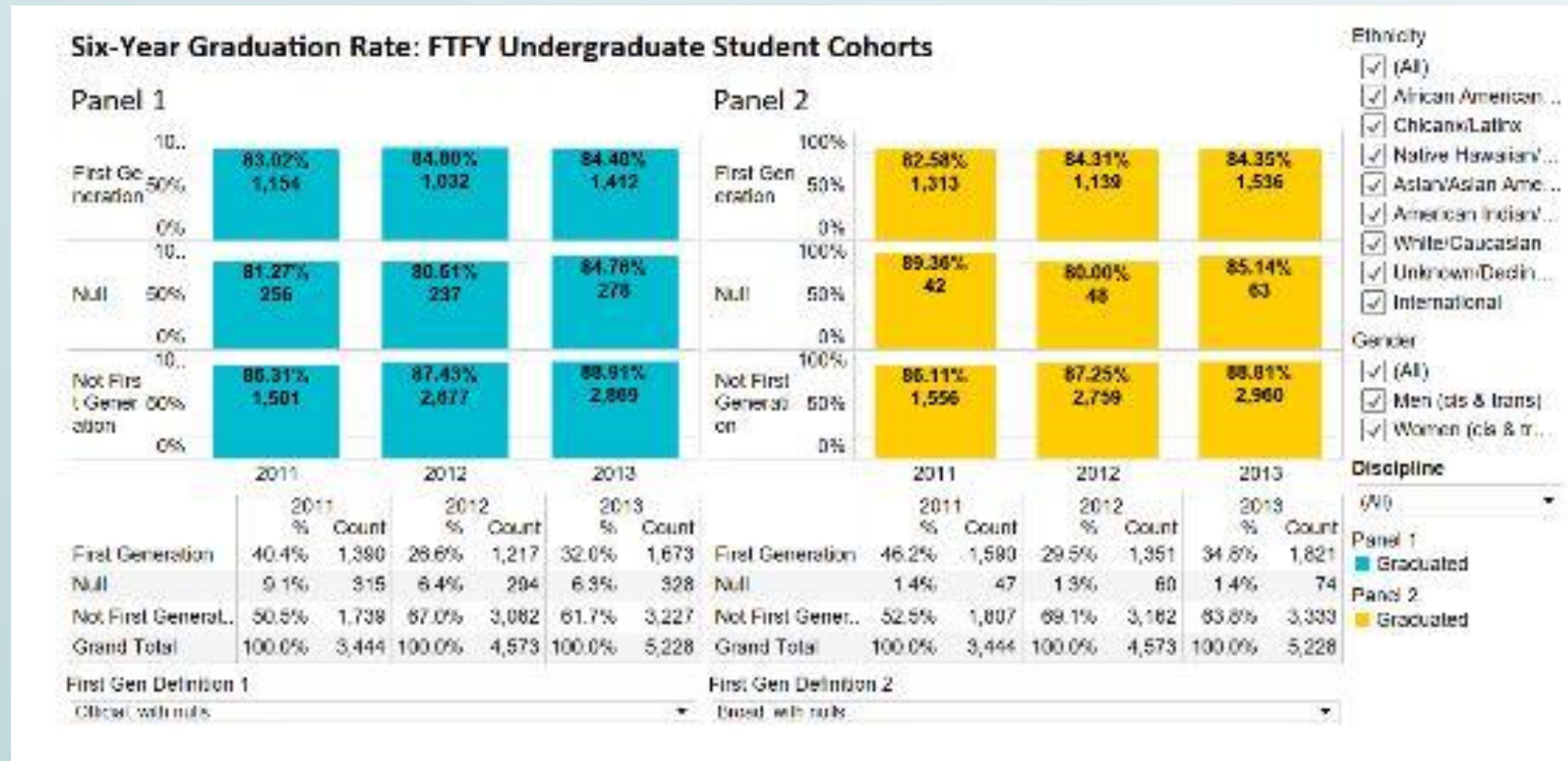
Six-Year Graduation Rate: FTFY Undergraduate Student Cohorts												ETHNICNEWINT		
	2011		2012		2013			2011		2012		2013		
	%	Count	%	Count	%	Count		%	Count	%	Count	%	Count	
First Generation	40.4%	1,390	26.6%	1,217	32.0%	1,673	First Generation	46.2%	1,590	29.5%	1,351	34.8%	1,821	<input checked="" type="checkbox"/> (All)
Null	9.1%	315	6.4%	294	6.3%	328	Null	1.4%	47	1.3%	60	1.4%	74	<input checked="" type="checkbox"/> African American...
Not First Generation	50.5%	1,739	67.0%	3,062	61.7%	3,227	Not First Generation	52.5%	1,807	69.1%	3,162	63.8%	3,333	<input checked="" type="checkbox"/> Chicano/Latinx
Grand Total	100.0%	3,444	100.0%	4,573	100.0%	5,228	Grand Total	100.0%	3,444	100.0%	4,573	100.0%	5,228	<input checked="" type="checkbox"/> Native Hawaiian/...

First Gen Definition 1 <input type="text" value="Official with nulls"/>	First Gen Definition 2 <input type="text" value="Broad with nulls"/>
---	--

Gender <input checked="" type="checkbox"/> (All) <input checked="" type="checkbox"/> Men (cis & trans) <input checked="" type="checkbox"/> Women (cis & tr...	Discipline <input type="text" value="(All)"/>
---	---

Varying Definitions Bar Charts

Comparison of bar charts, frequency distribution and outcome statistics between field definitions.



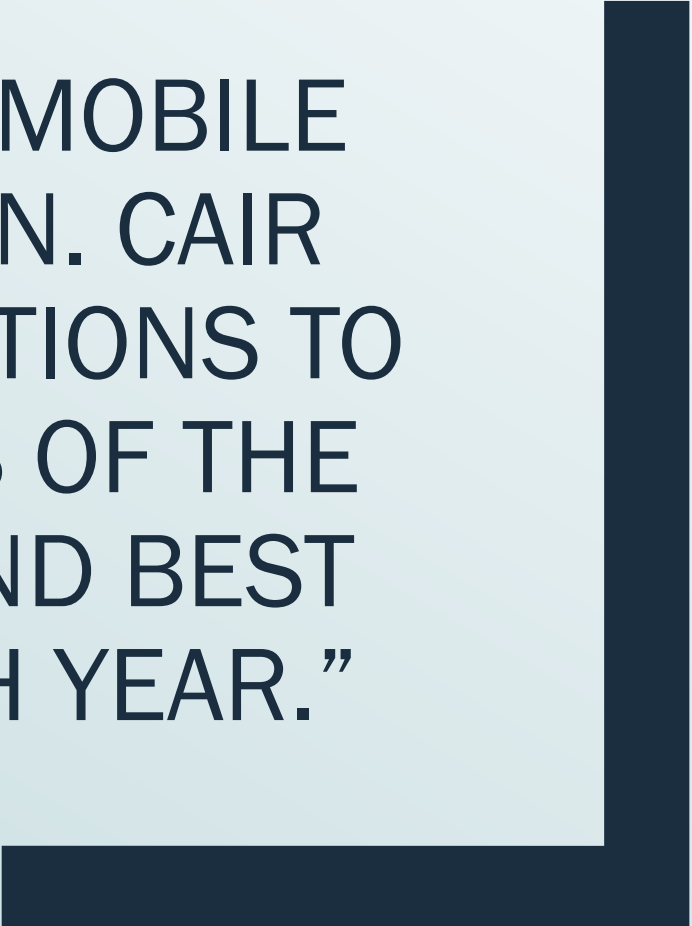

References

<https://blog.datawrapper.de/dualaxis/>

https://en.wikipedia.org/wiki/Misleading_graph

<https://www.datapine.com/blog/how-to-choose-the-right-data-visualization-types/>

<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.725.7355&rep=rep1&type=pdf>



“PLEASE USE THE WHOVA MOBILE APP TO RATE THIS SESSION. CAIR USES THE SESSION EVALUATIONS TO DETERMINE THE WINNERS OF THE BEST NEW PRESENTER, AND BEST PRESENTER AWARDS EACH YEAR.”

Questions?



CONTACT US:



- Heidi M. Carty – hcarty@ucsd.edu
- Galina Belokurova – gbelokurova@ucsd.edu