

Data Viz

April 2, 2019

Data Science CSCI 1951A

Brown University

Instructor: Ellie Pavlick

HTAs: Wennie Zhang, Maulik Dang, Gurnaaz Kaur

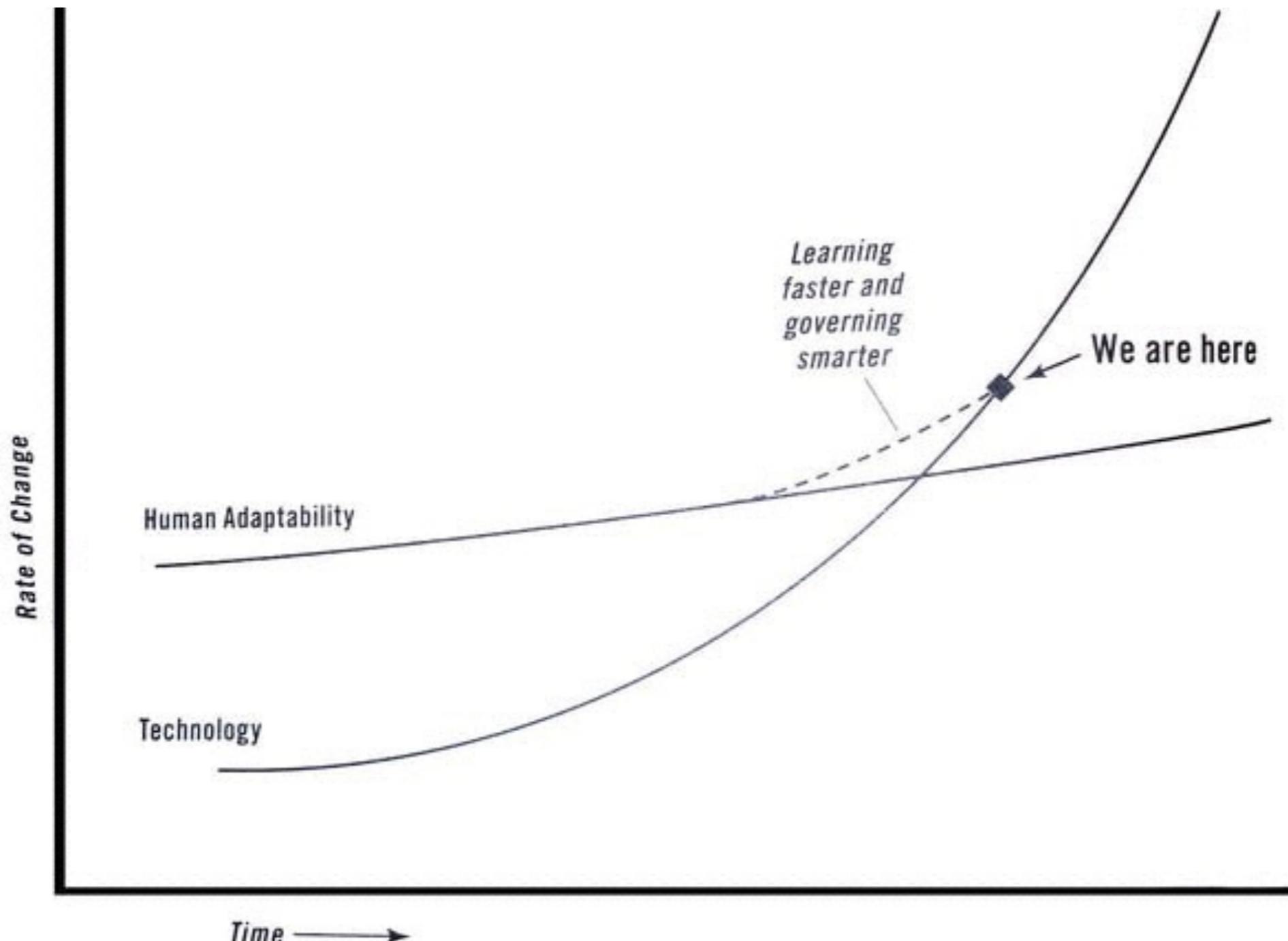
Announcements

- Welcome back!
- Data Viz assignment released (using D3)
- Summer Project with RI Hospital (email Adeel;
`adeel_abbas@brown.edu`)

Today

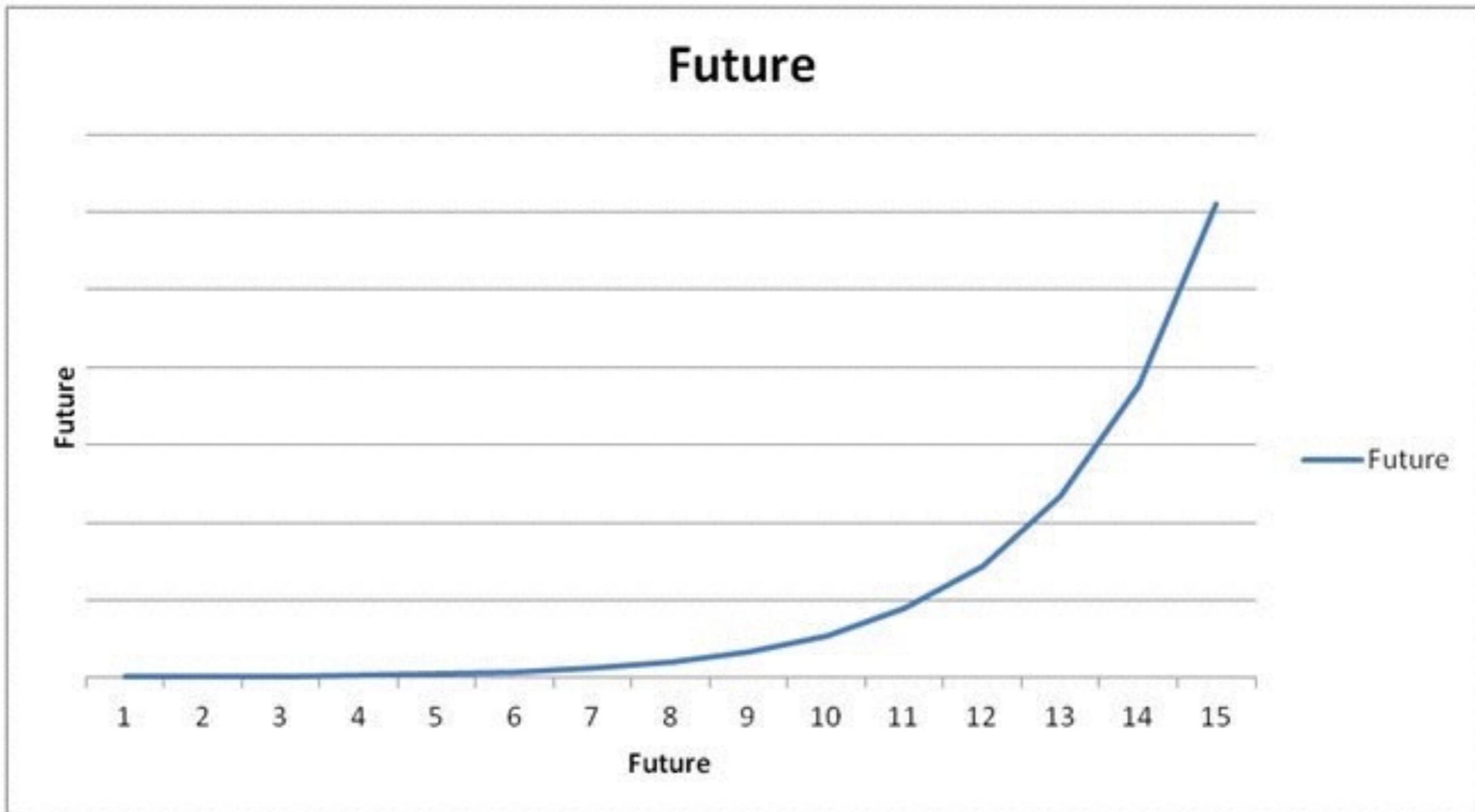
- Data Viz best practices
- D3 demo

So many bad figures...



<https://www.rollingstone.com/politics/politics-news/who-won-the-make-the-most-meaningless-thomas-friedman-graph-contest-111259/>

So many bad figures...



<https://www.rollingstone.com/politics/politics-news/who-won-the-make-the-most-meaningless-thomas-friedman-graph-contest-111259/>
<https://twitter.com/OttawaFencer/status/802210174450757632>

My “three pillars” of Data Viz

My* “three pillars” of Data Viz

*I feel like I don't have the gravitas to be a person who has a “three pillar” but nonetheless, I decided to have them...

My “three pillars” of Data Viz

clarity — Your figures should speak for themselves. The analysis should be understandable and your conclusions should be obviously supported, without too much effort

My “three pillars” of Data Viz

Clarity — Your figures should speak for themselves. The analysis should be understandable and your conclusions should be obviously supported, without too much effort

Don't obfuscate the data or **H**ide the pr**O**cess you used to come to your co**N**clusions. Giv**E** people enough data **S**o that **T**hey can disagree with **Y**ou if they want to.

My “three pillars” of Data Viz

Clarity — Your figures should speak for themselves. The analysis should be understandable and your conclusions should be obviously supported, without too much effort

Don't obfuscate the data or **H**ide the pr**O**cess you used to come to your co**N**clusions. Giv**E** people enough data **S**o that **T**hey can disagree with **Y**ou if they want to.

Minimalism = **Substance over style. Make your point concisely, without redundant or distracting information or ornamentation.**

Ellie rants about culture for
2 seconds. Indulge me....

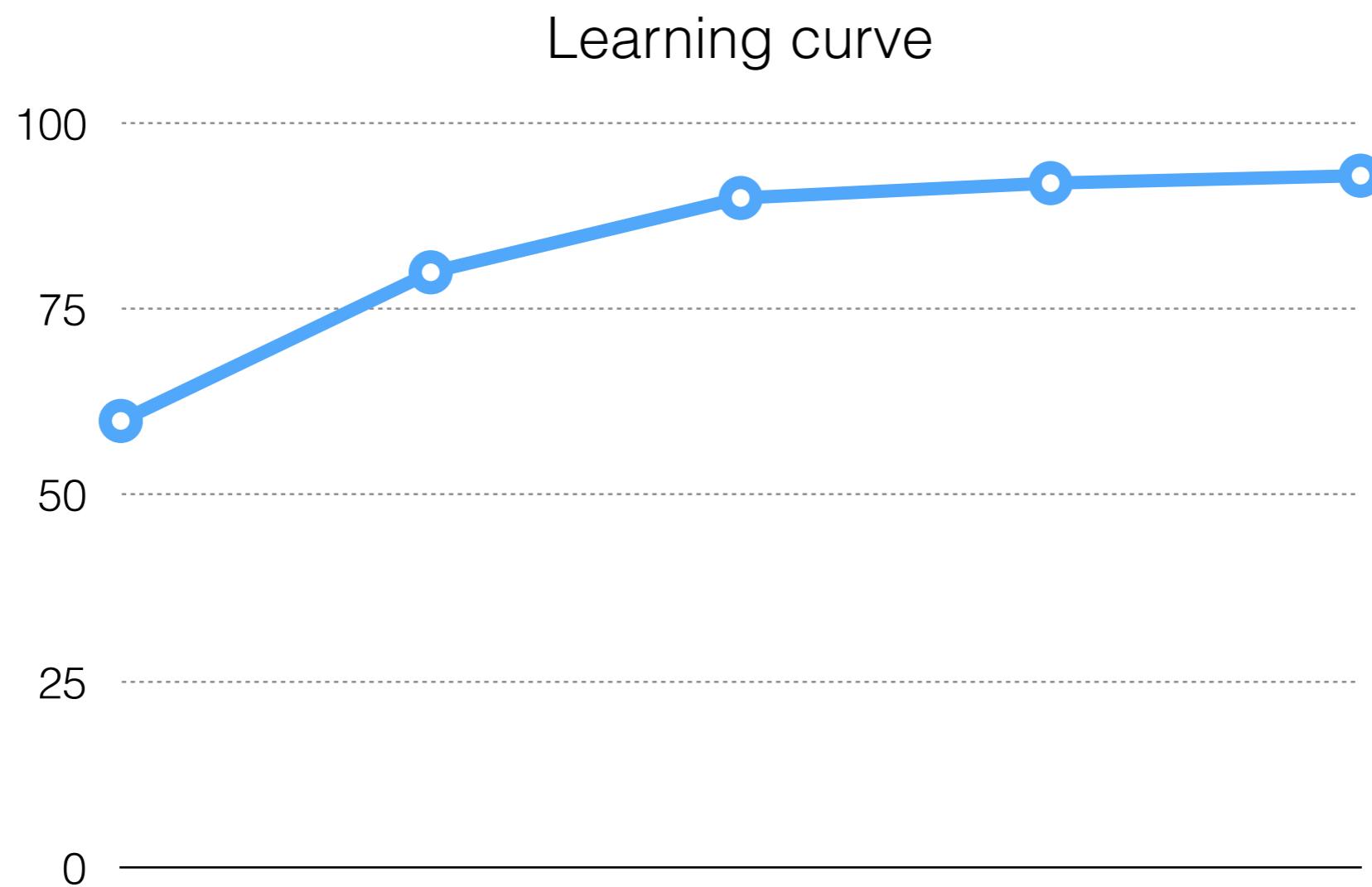
Ellie rants about culture for
2 seconds. Indulge me....

“form follows function”

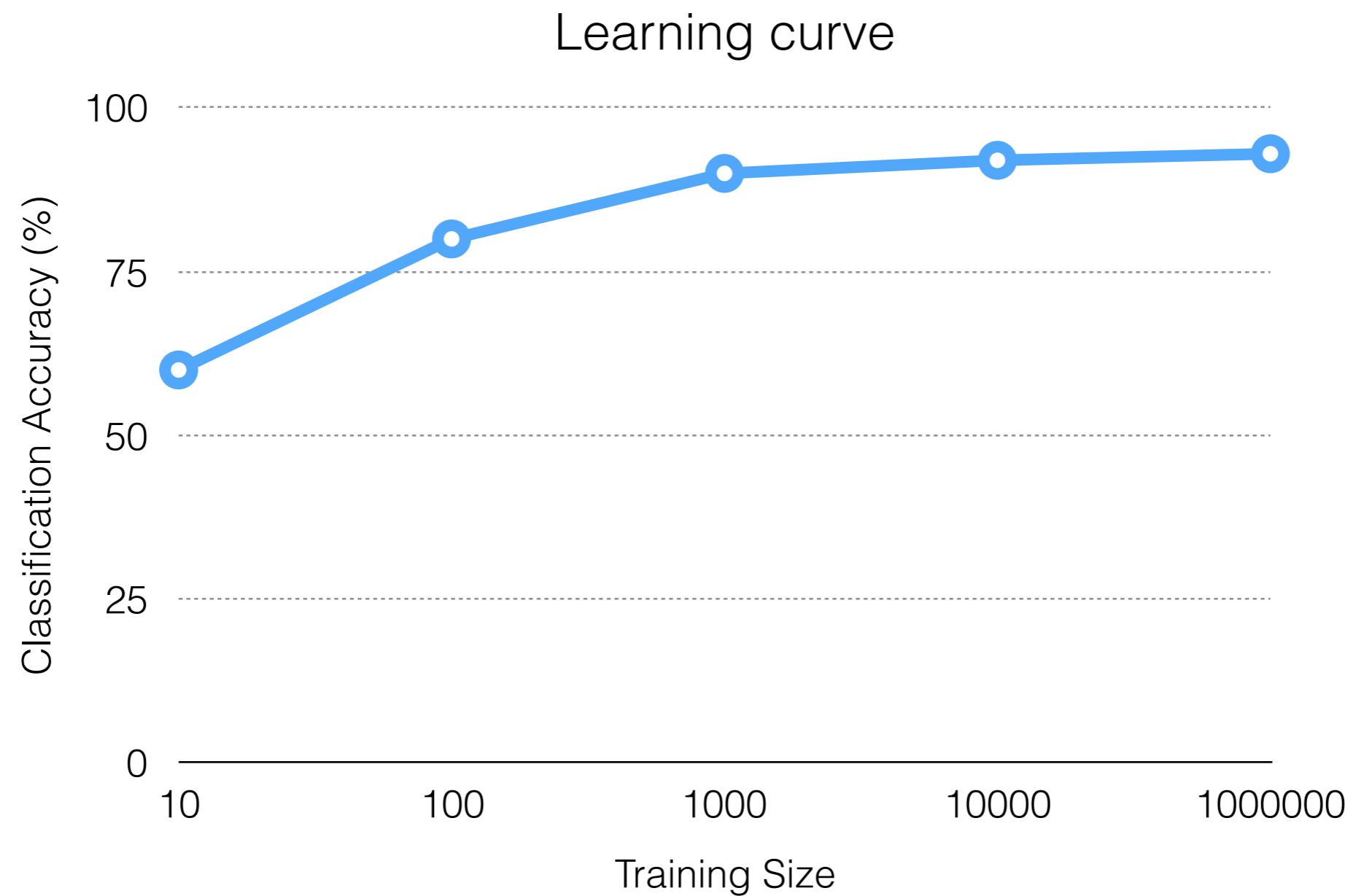
My “three pillars” of Data Viz

clarity — Your figures should speak for themselves. The analysis should be understandable and your conclusions should be obviously supported, without too much effort

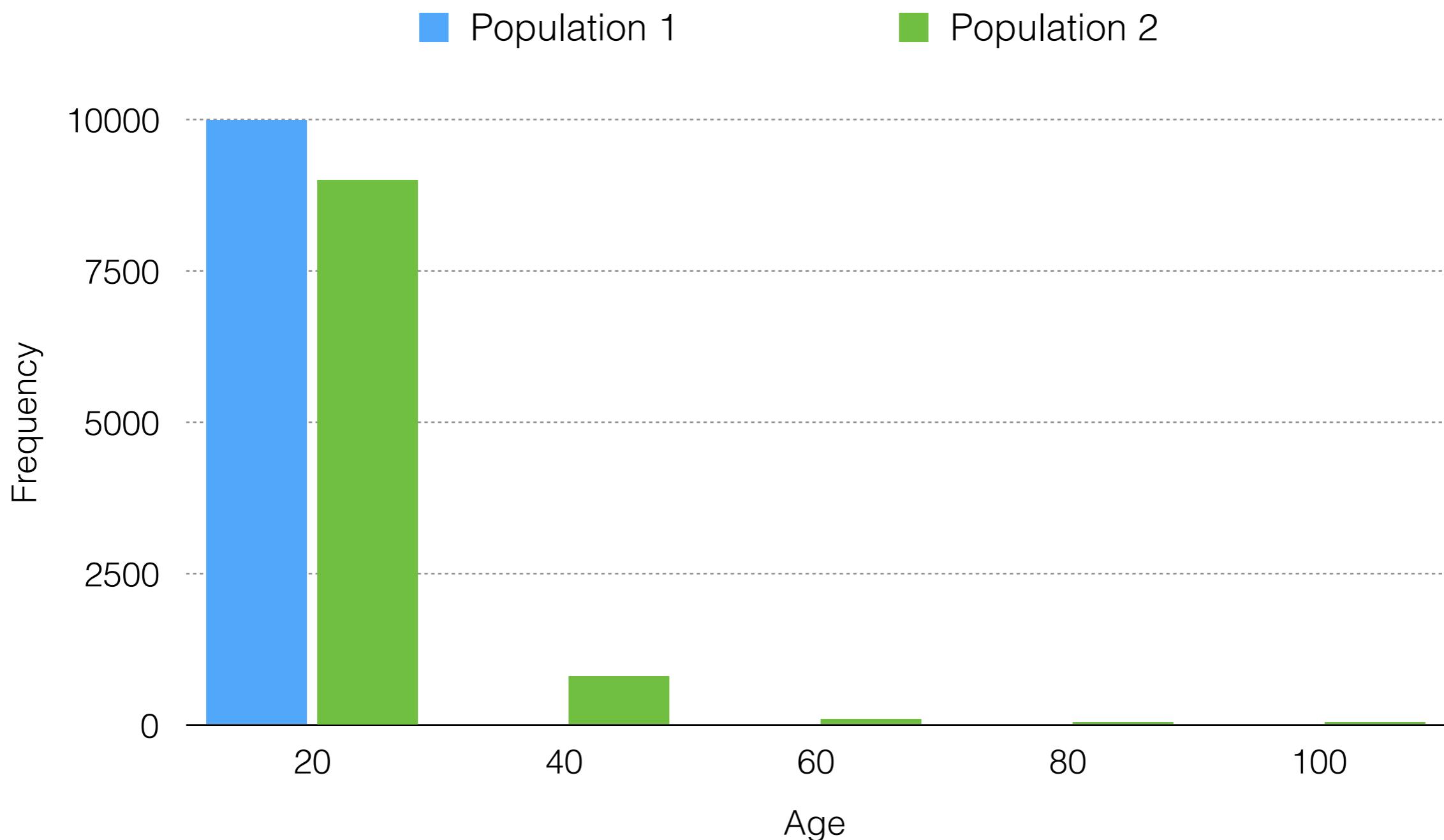
Missing or Cryptic Labels



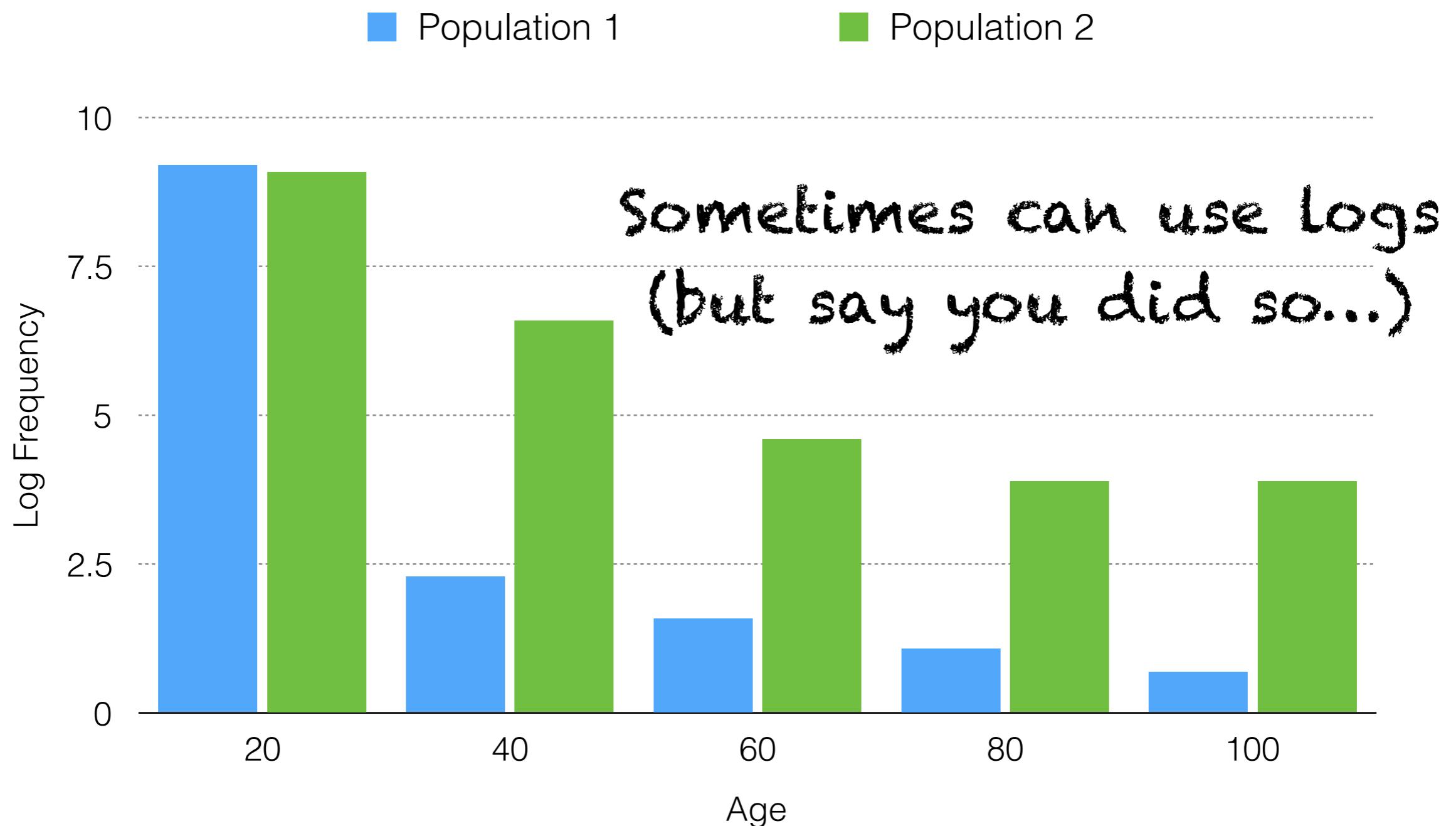
Missing or Cryptic Labels



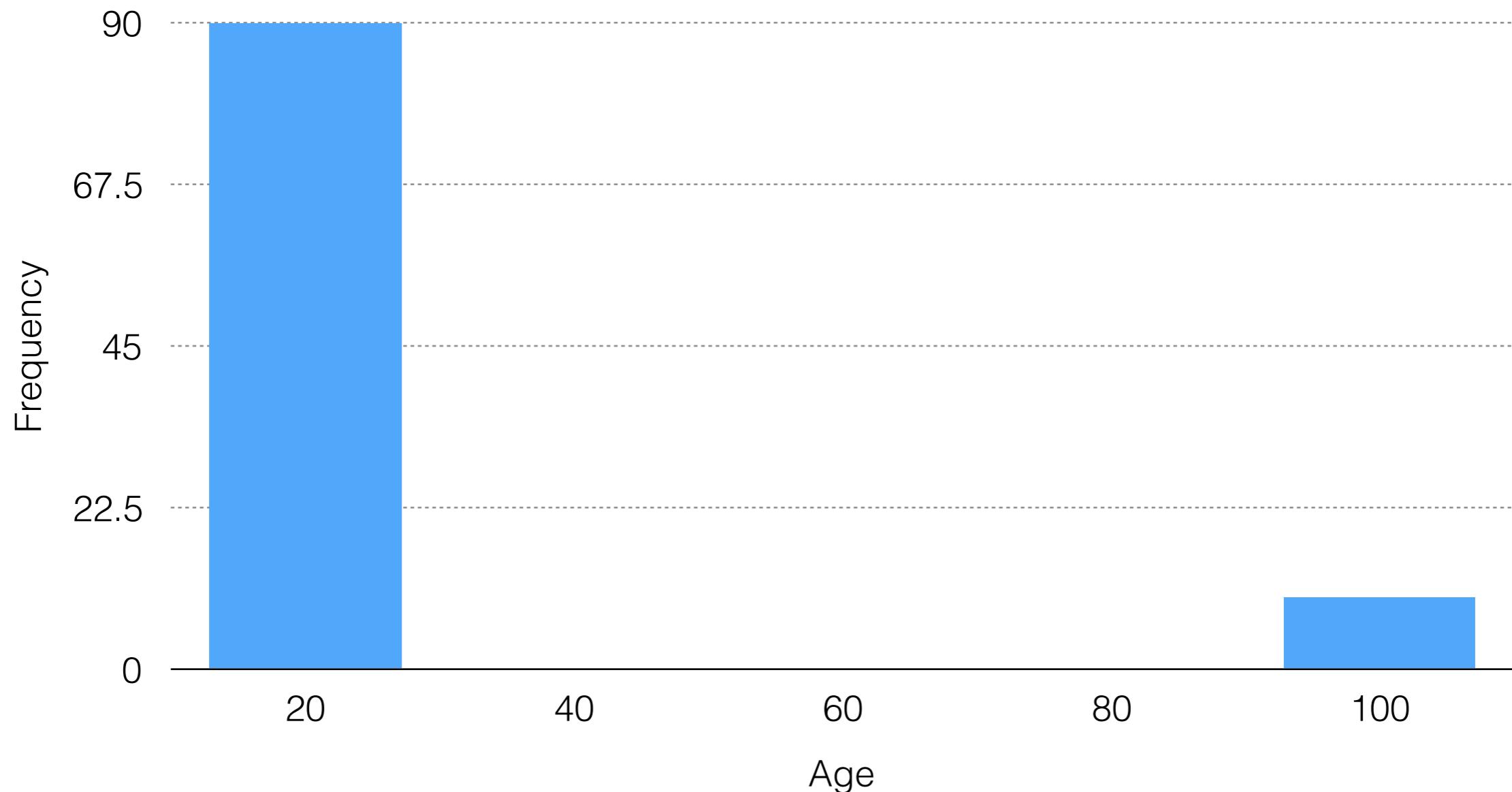
Skewed or Crunched Data



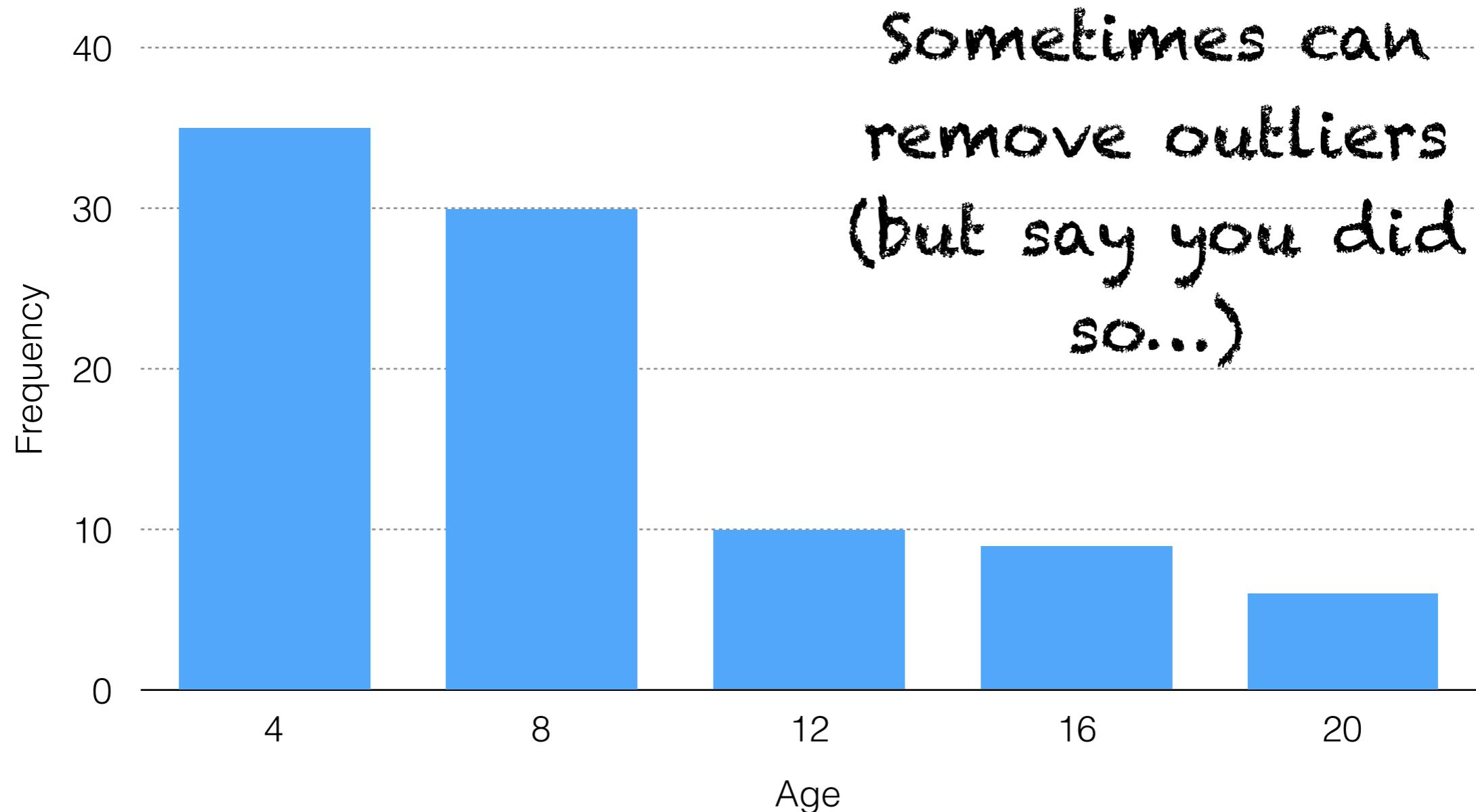
Skewed or Crunched Data



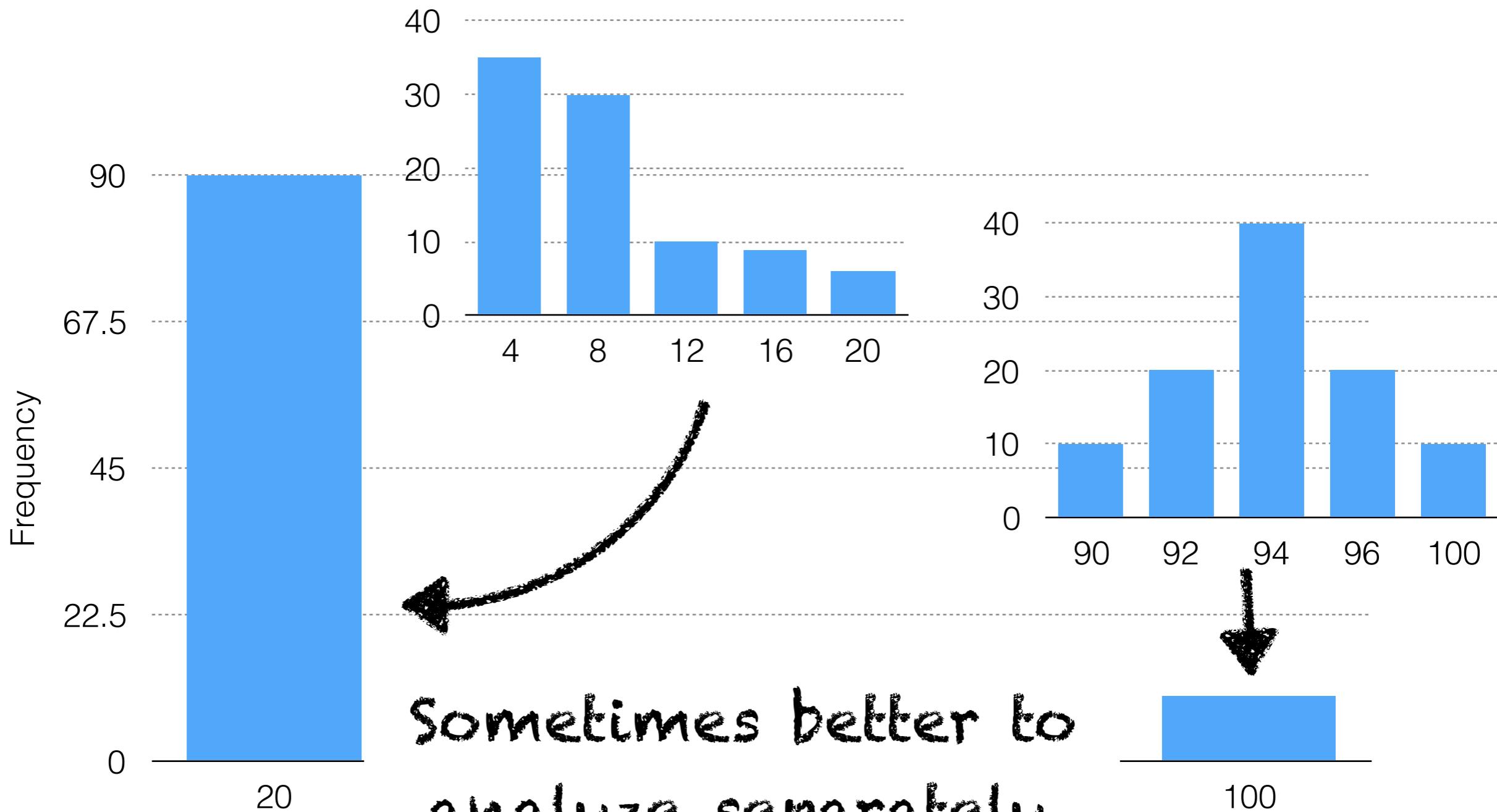
Skewed or Crunched Data



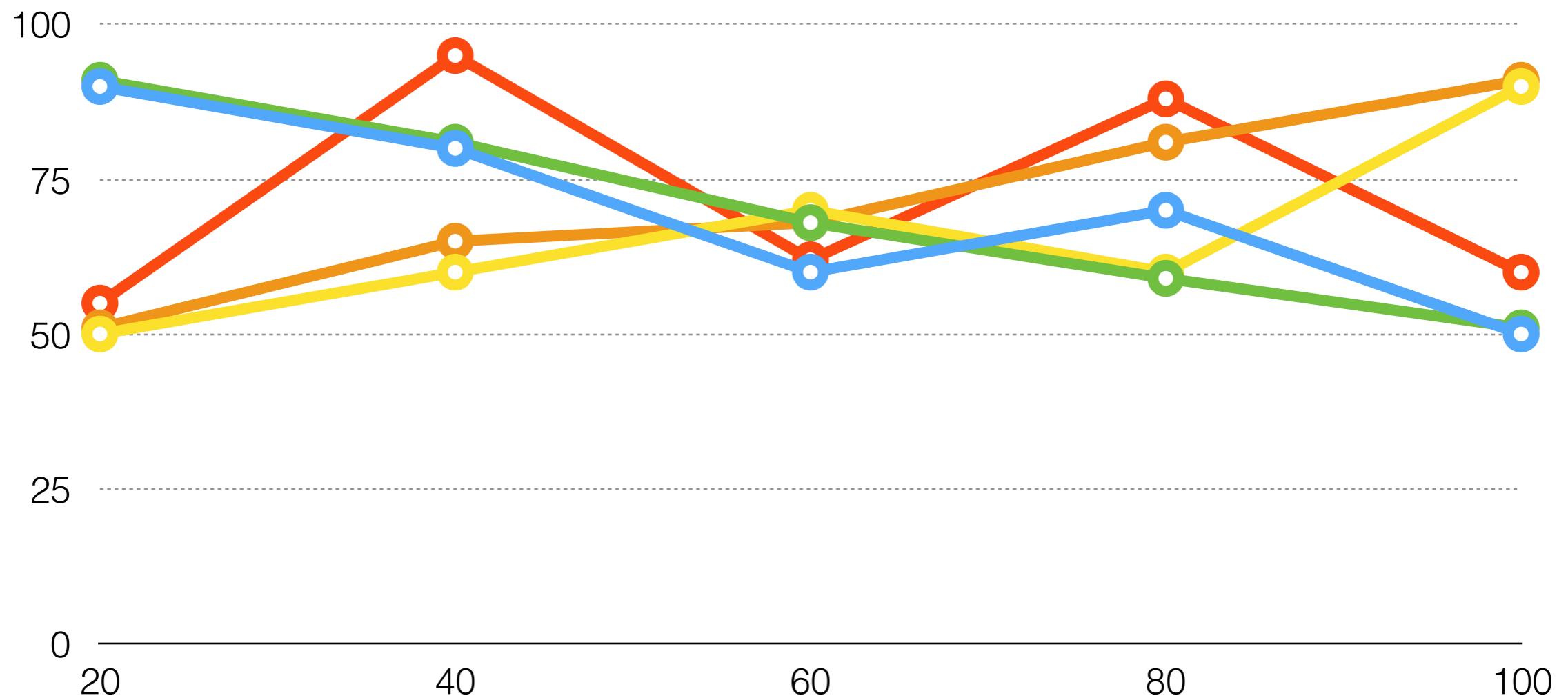
Skewed or Crunched Data



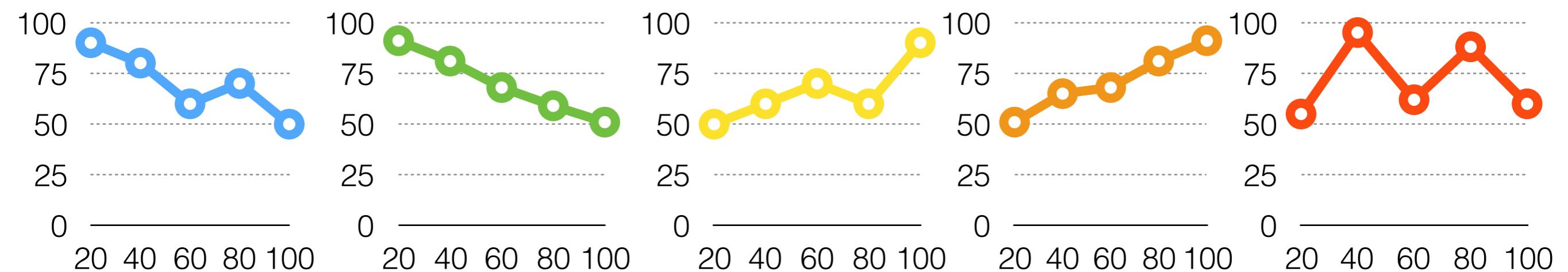
Skewed or Crunched Data



Skewed or Crunched Data

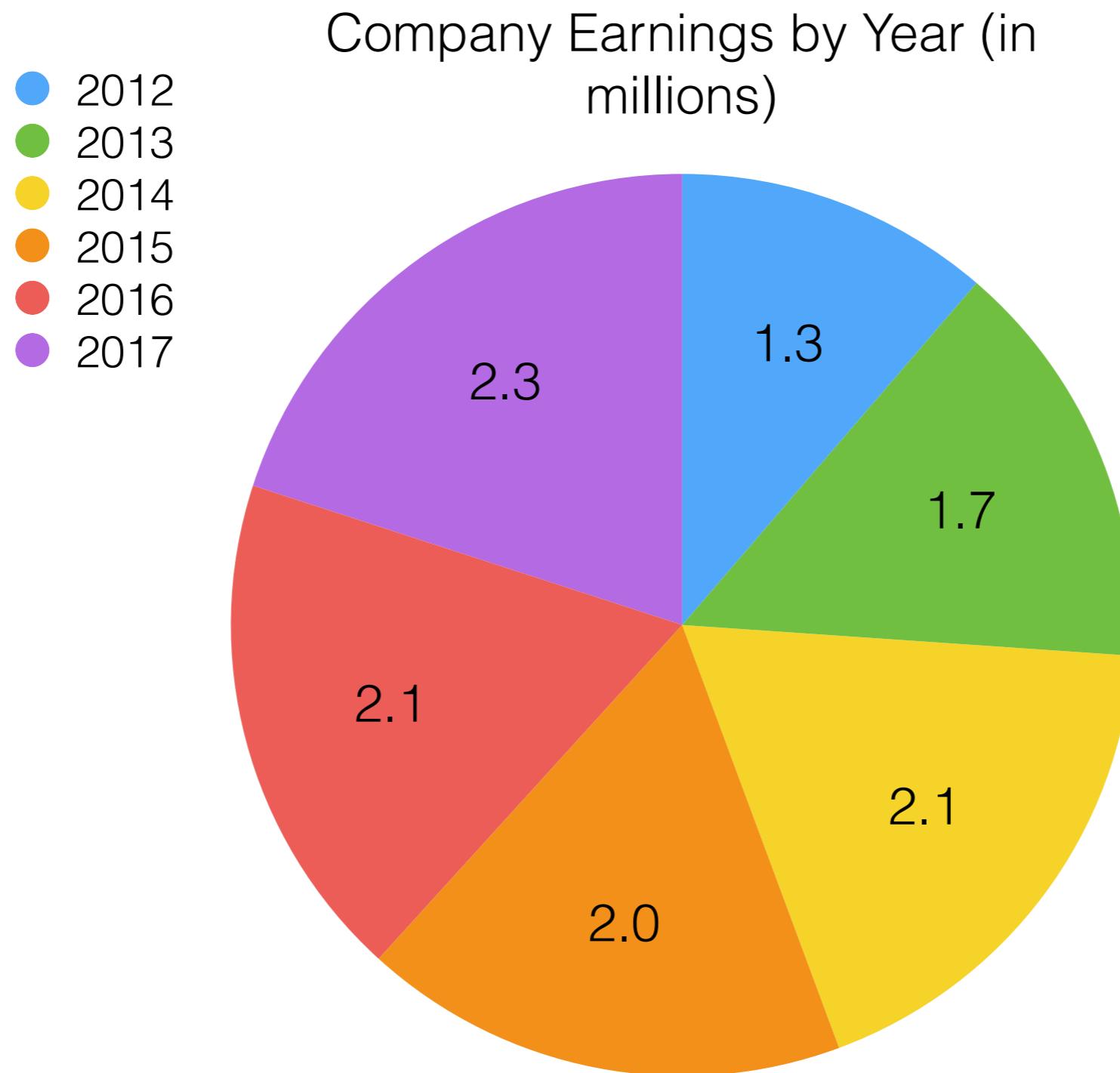


Skewed or Crunched Data

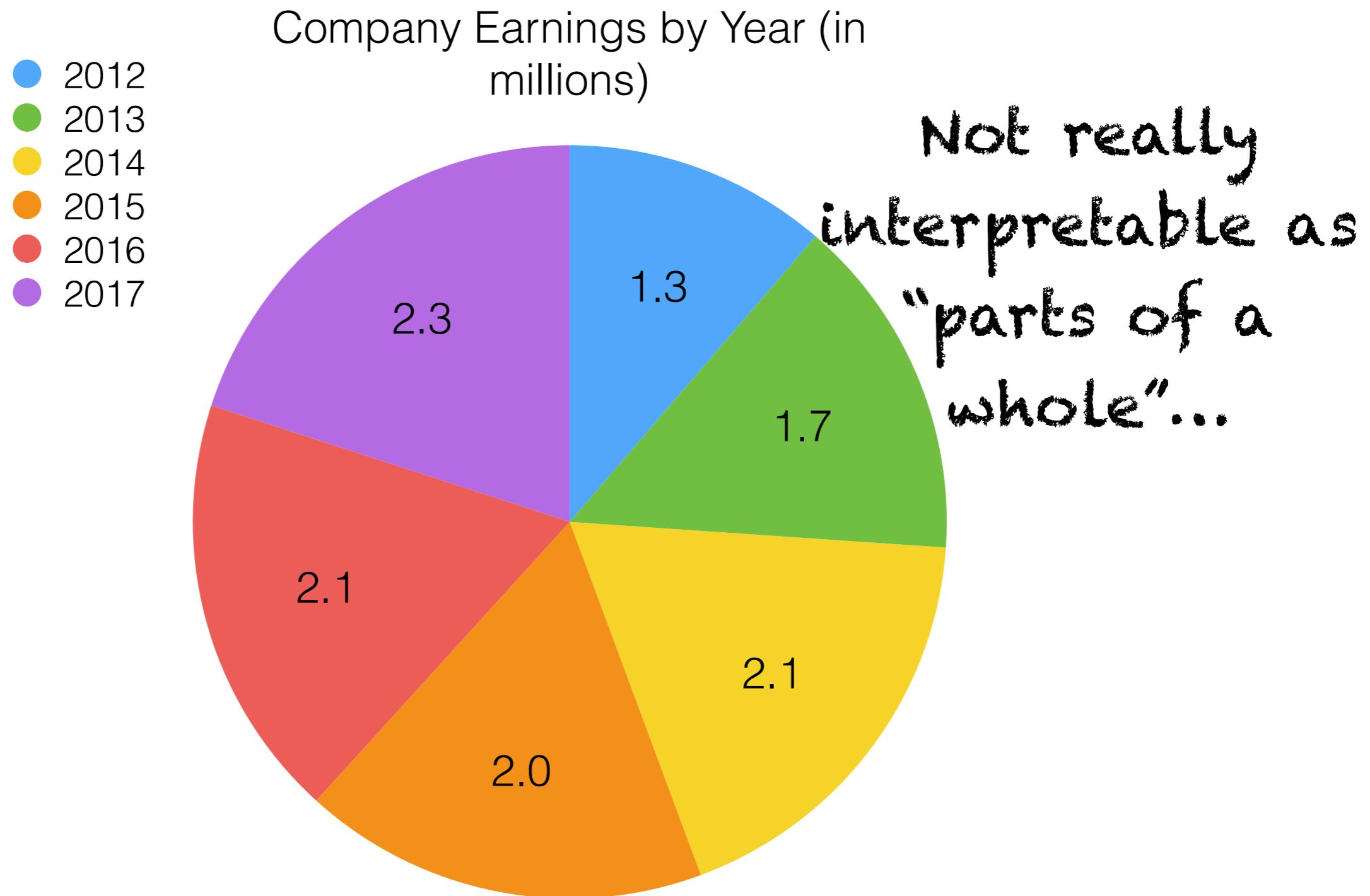


Sometimes better to split into
multiple charts...

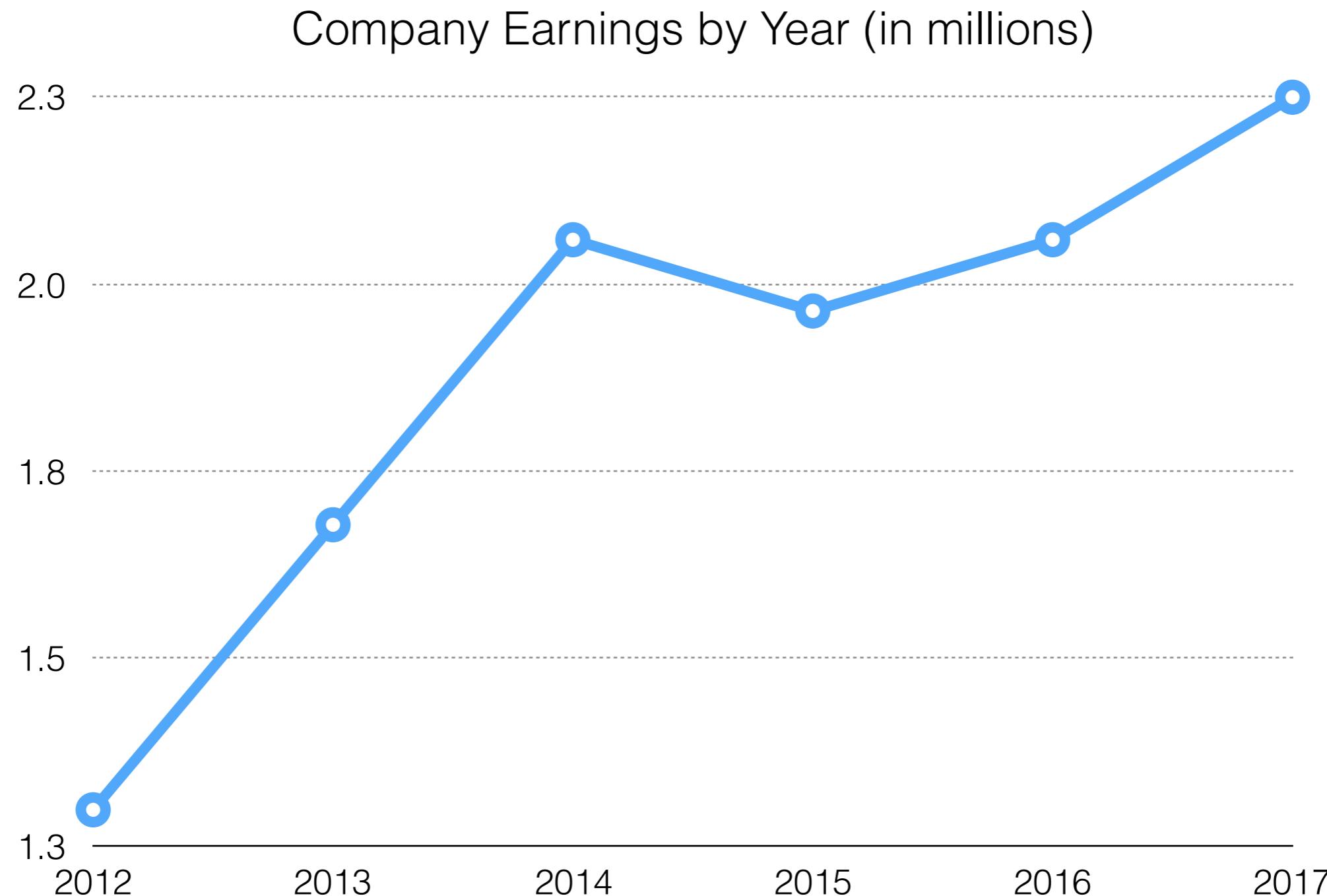
Chart/Data Type Mismatch



Chart/Data Type Mismatch



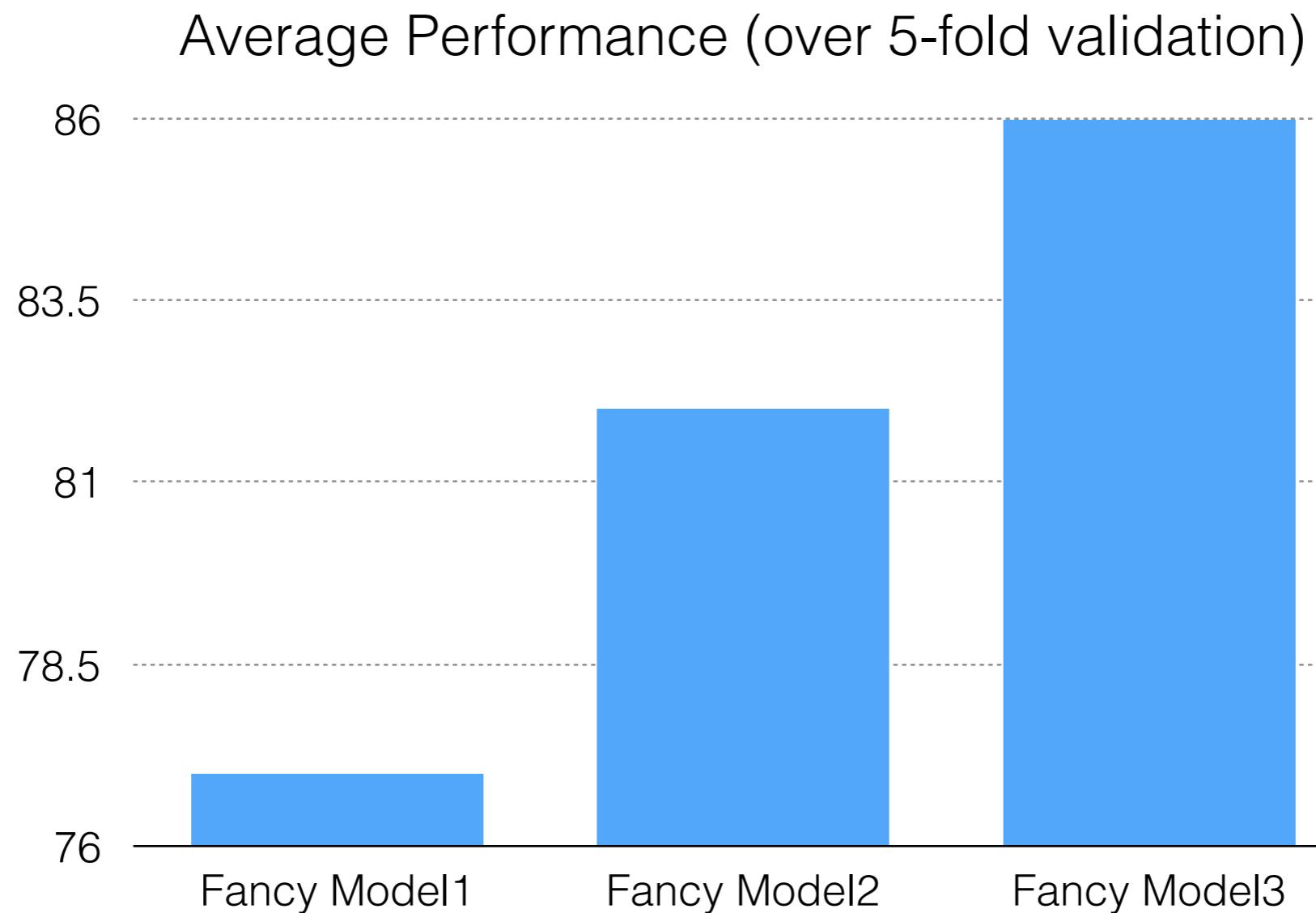
Chart/Data Type Mismatch



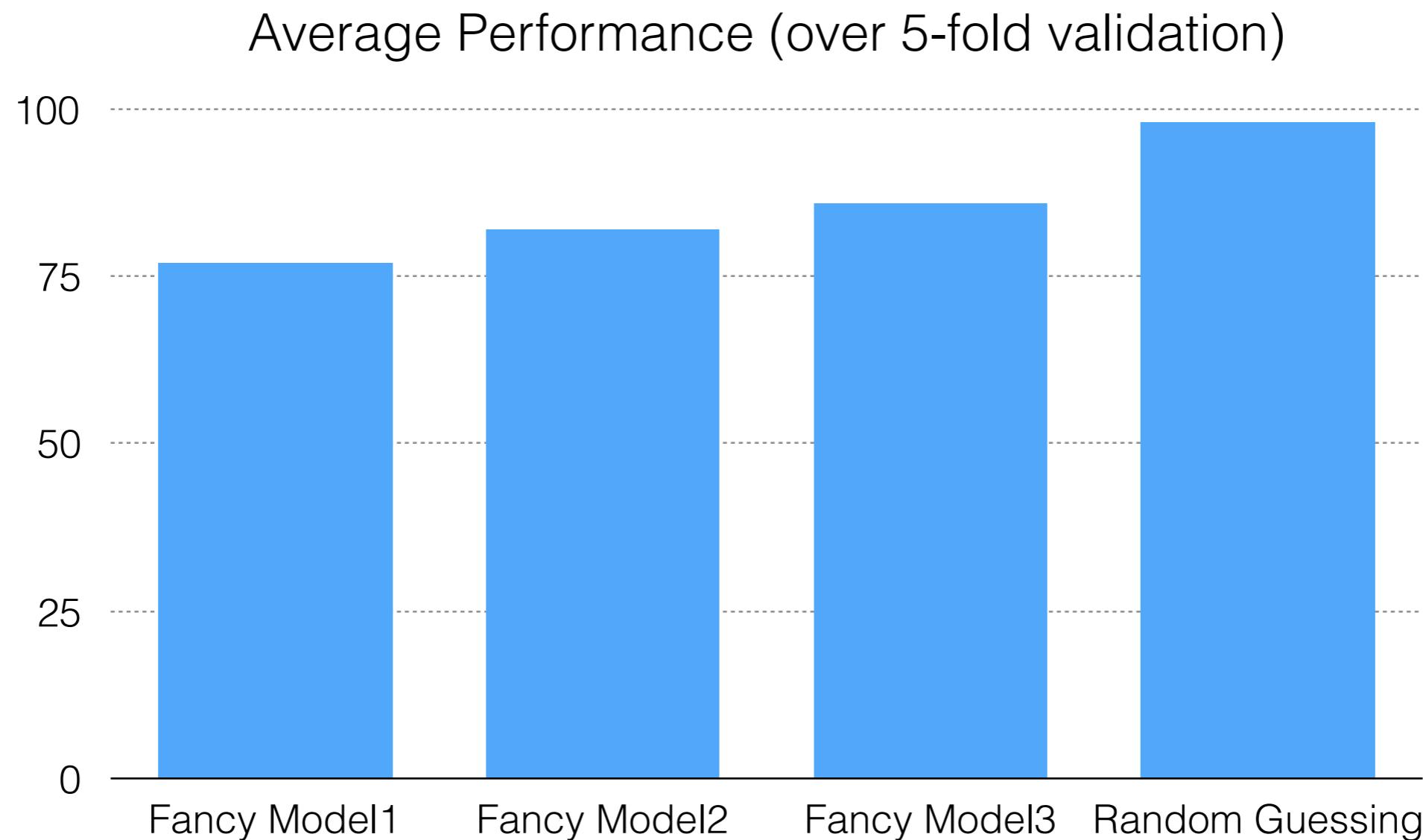
My “three pillars” of Data Viz

Don't obfuscate the data or **H**ide the pr**O**cess you used to come to your co**N**clusions. Giv**E** people enough data **S**o that **T**hey can disagree with **Y**ou if they want to.

No Point of Comparison

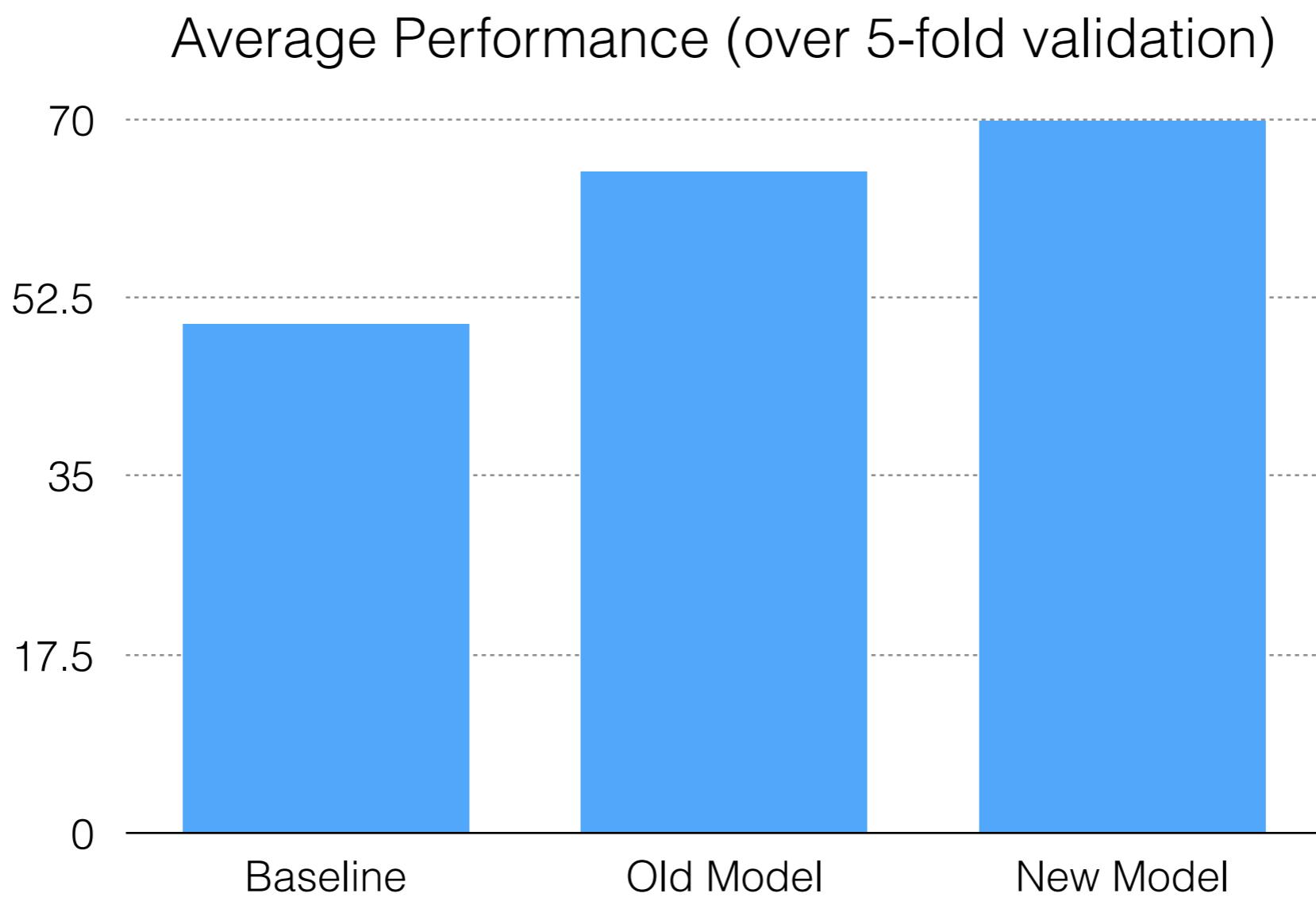


No Point of Comparison

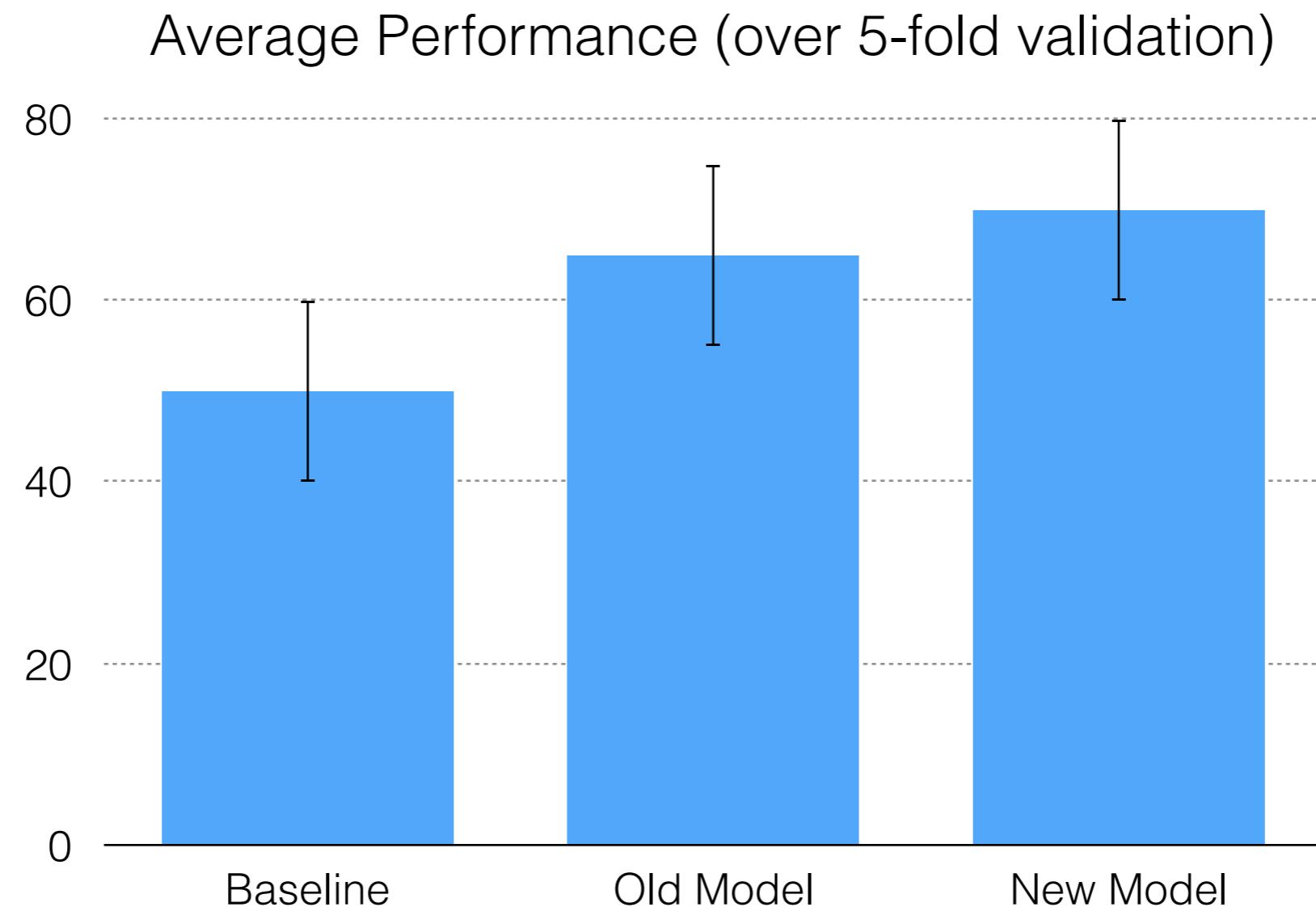


Help calibrate how easy/hard the problem is,
what types of numbers to expect a priori...

Summary Stats Only

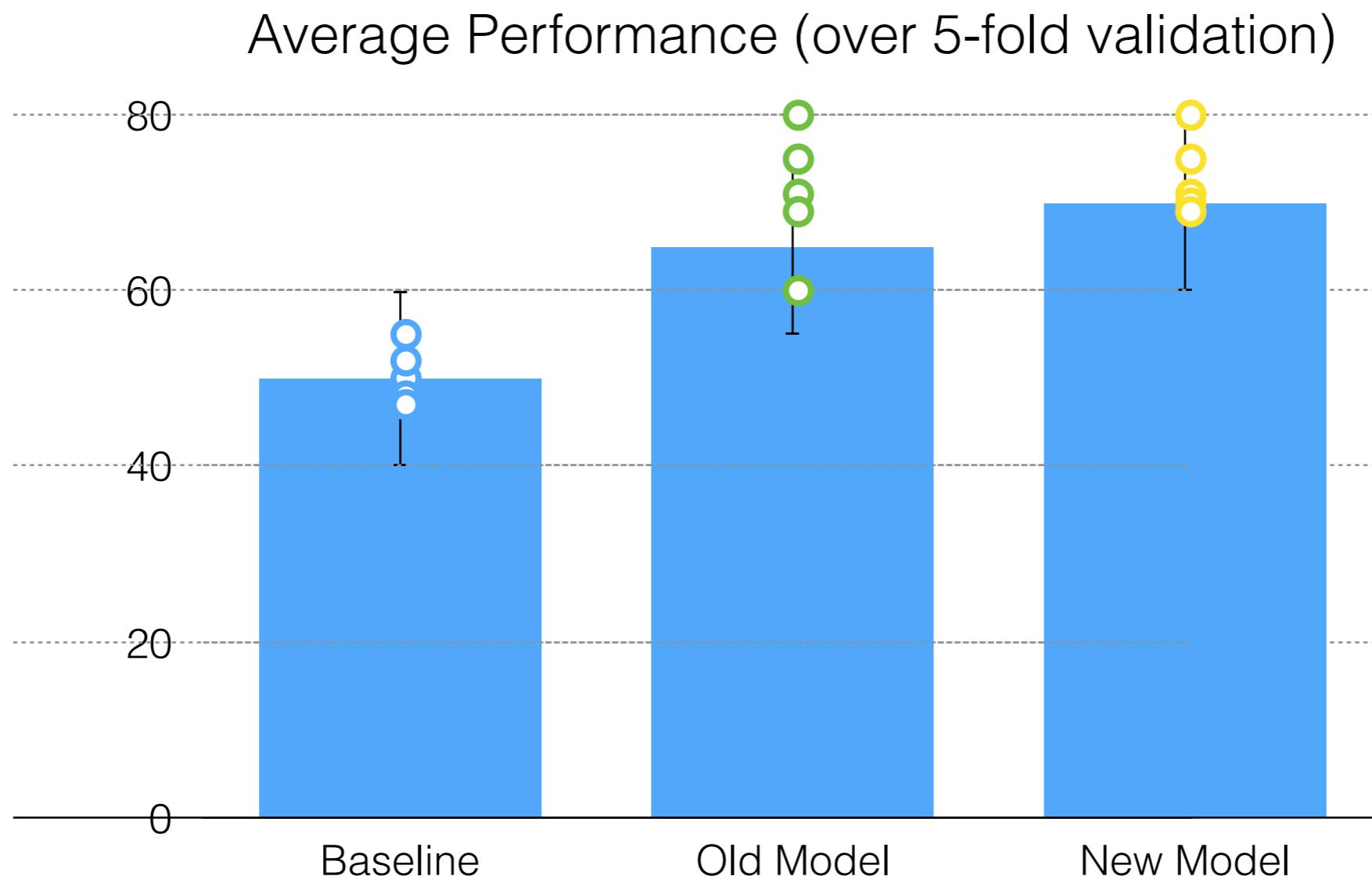


Summary Stats Only



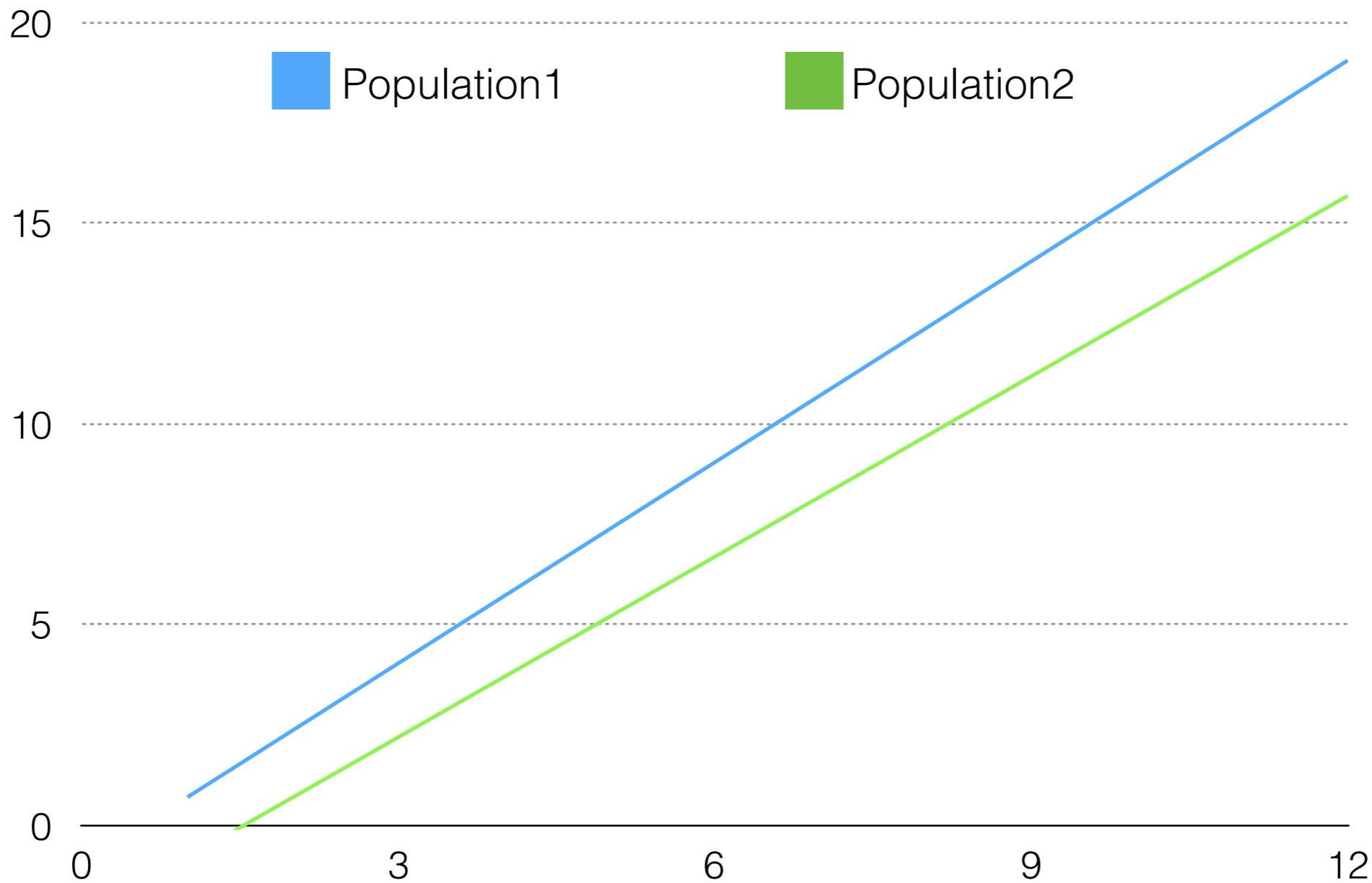
Sometimes can include error bars/
confidence intervals...

Summary Stats Only

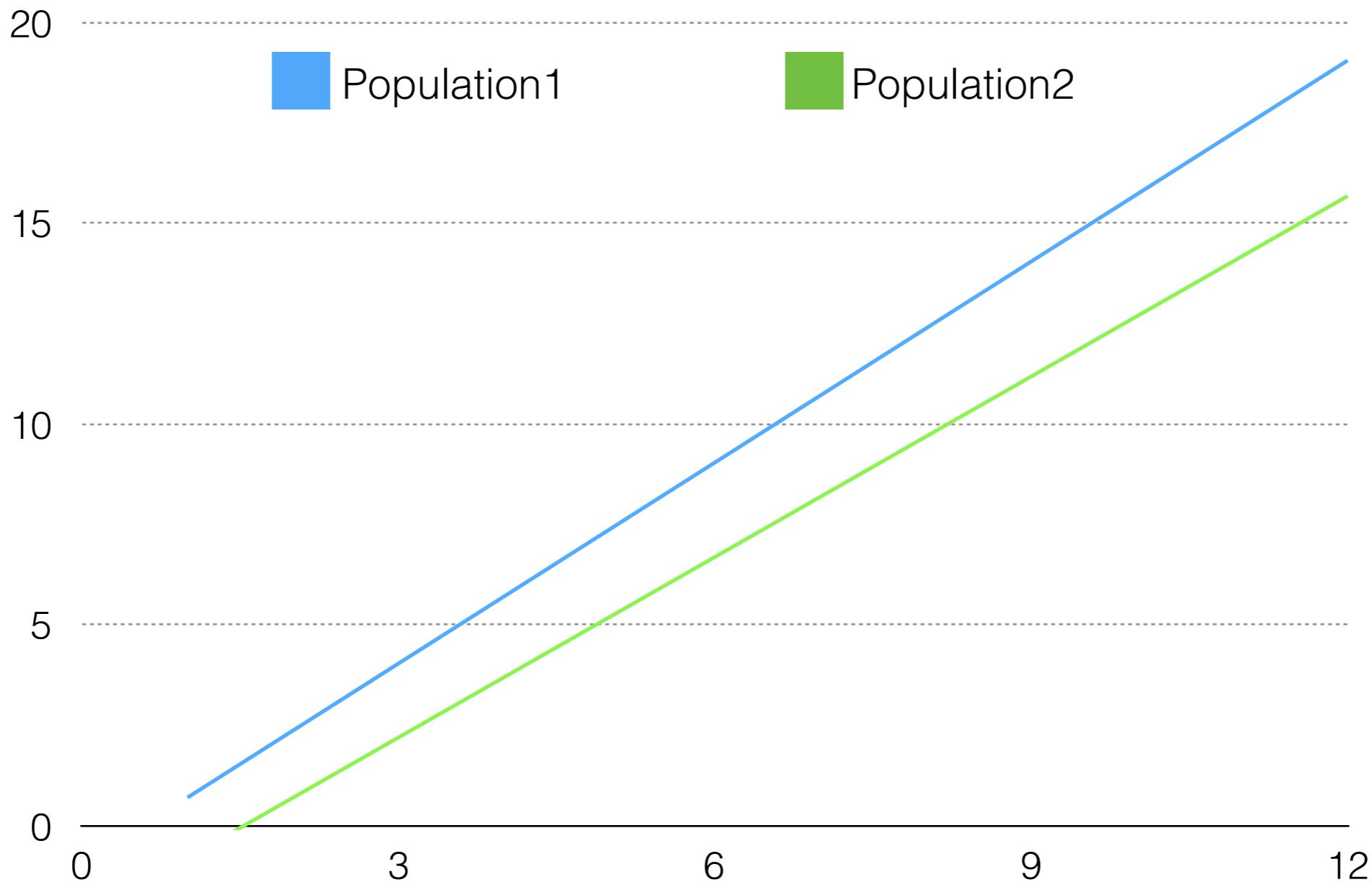


Even better, just show all the data alongside
the summary stats.

Summary Stats Only

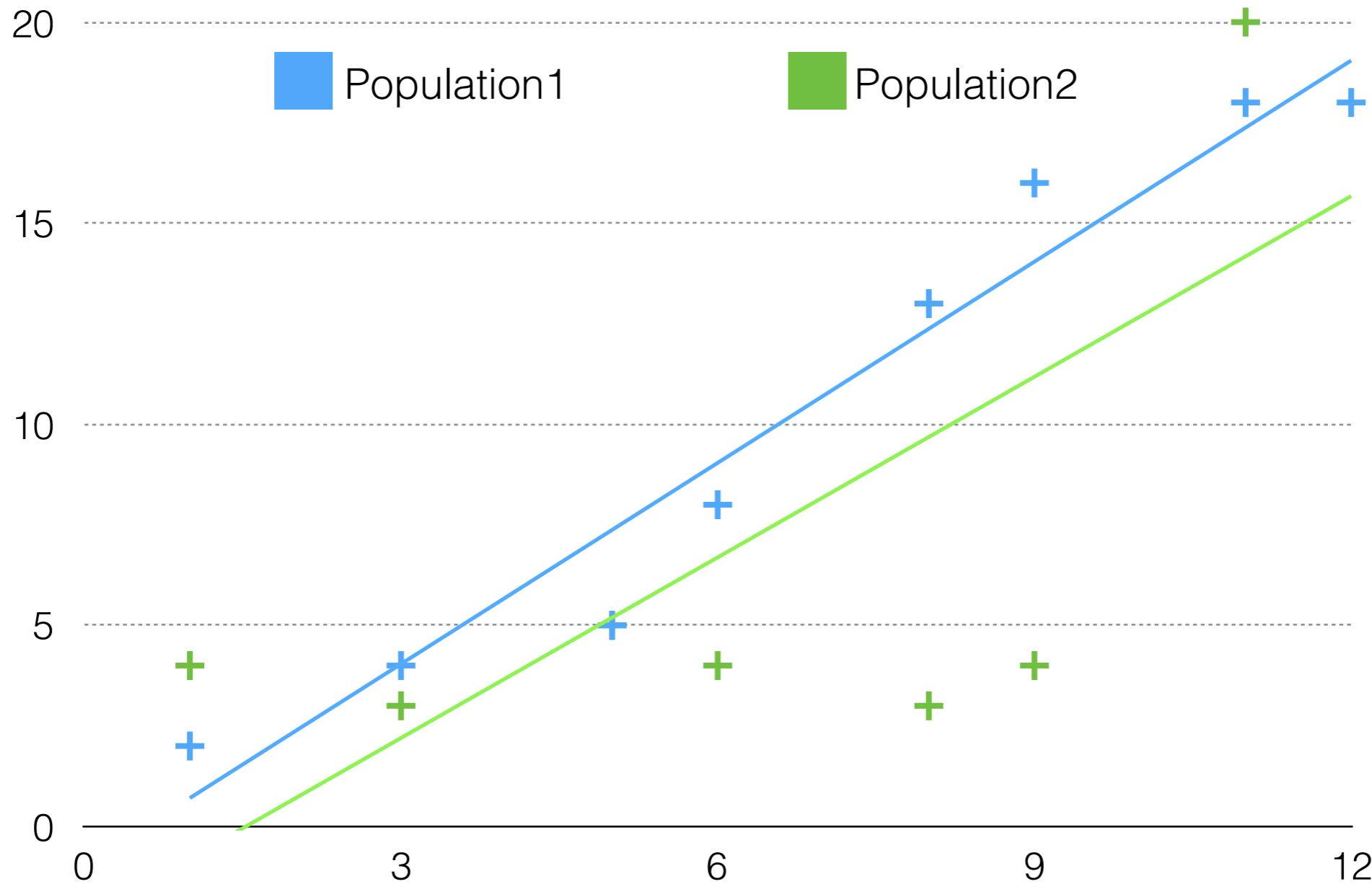


Summary Stats Only



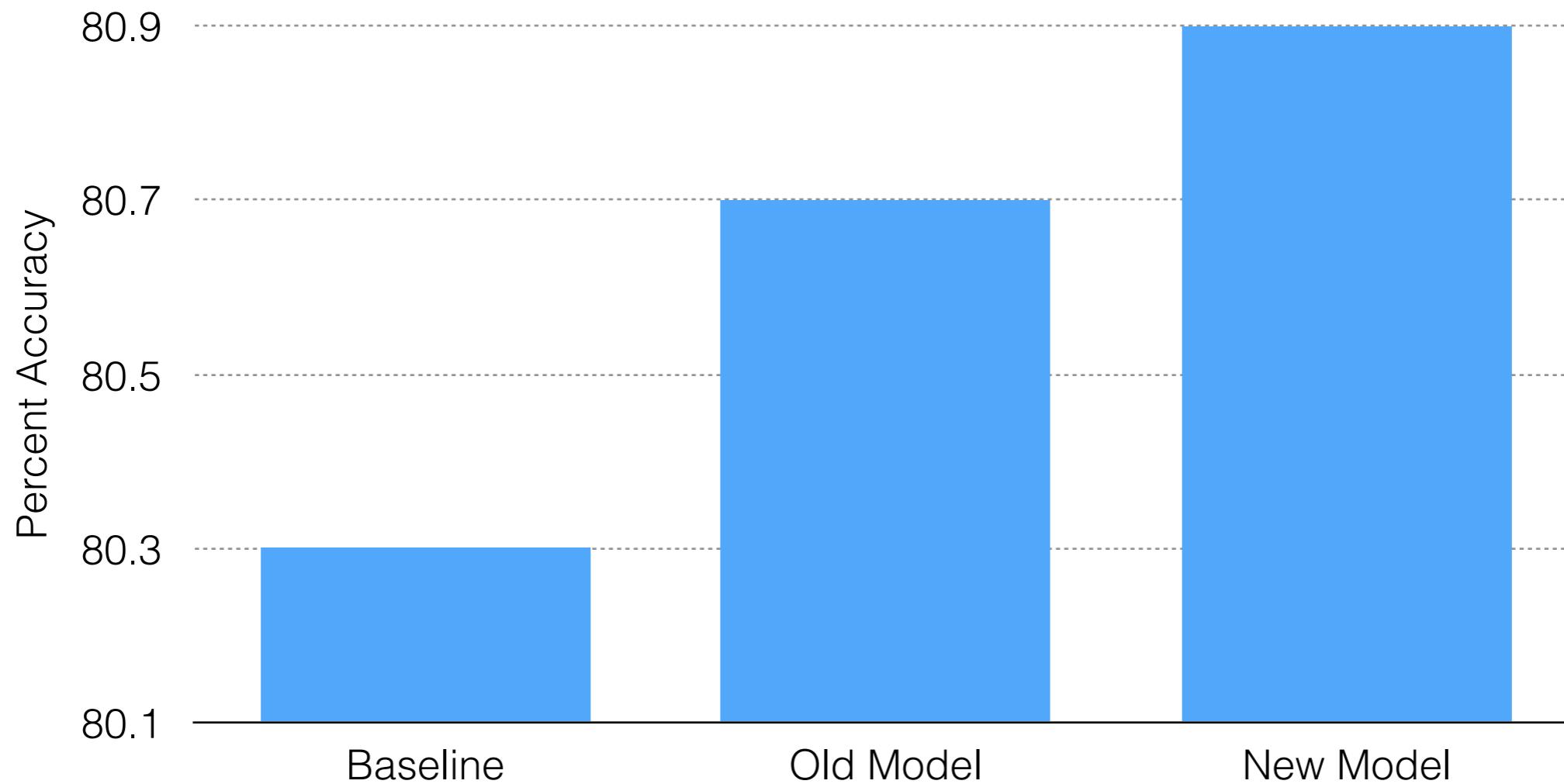
Be careful about showing smoothed/
estimated/aggregated trends only

Summary Stats Only

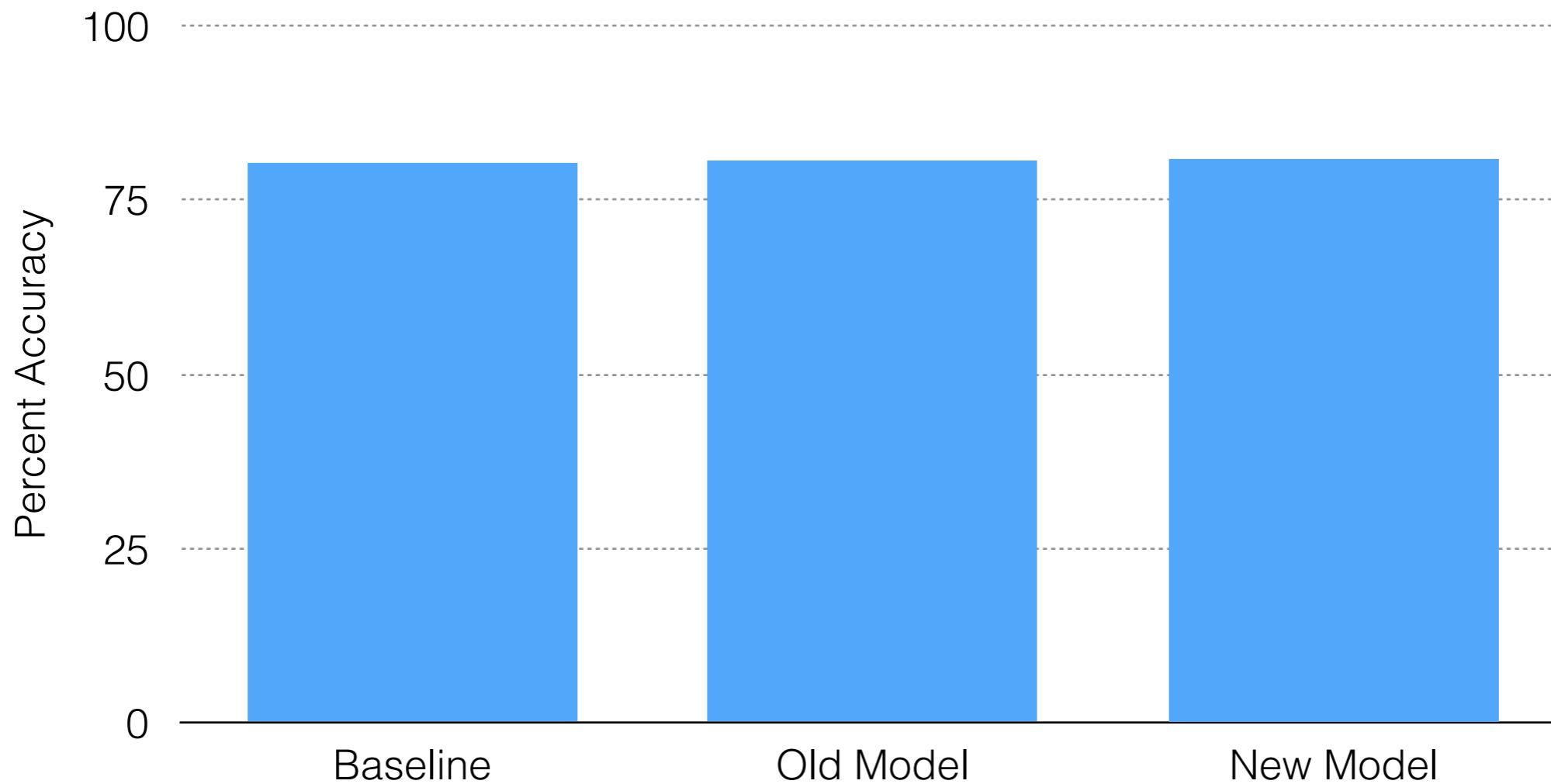


Whenever possible, show underlying data

Misleading/Badly Scaled Axes

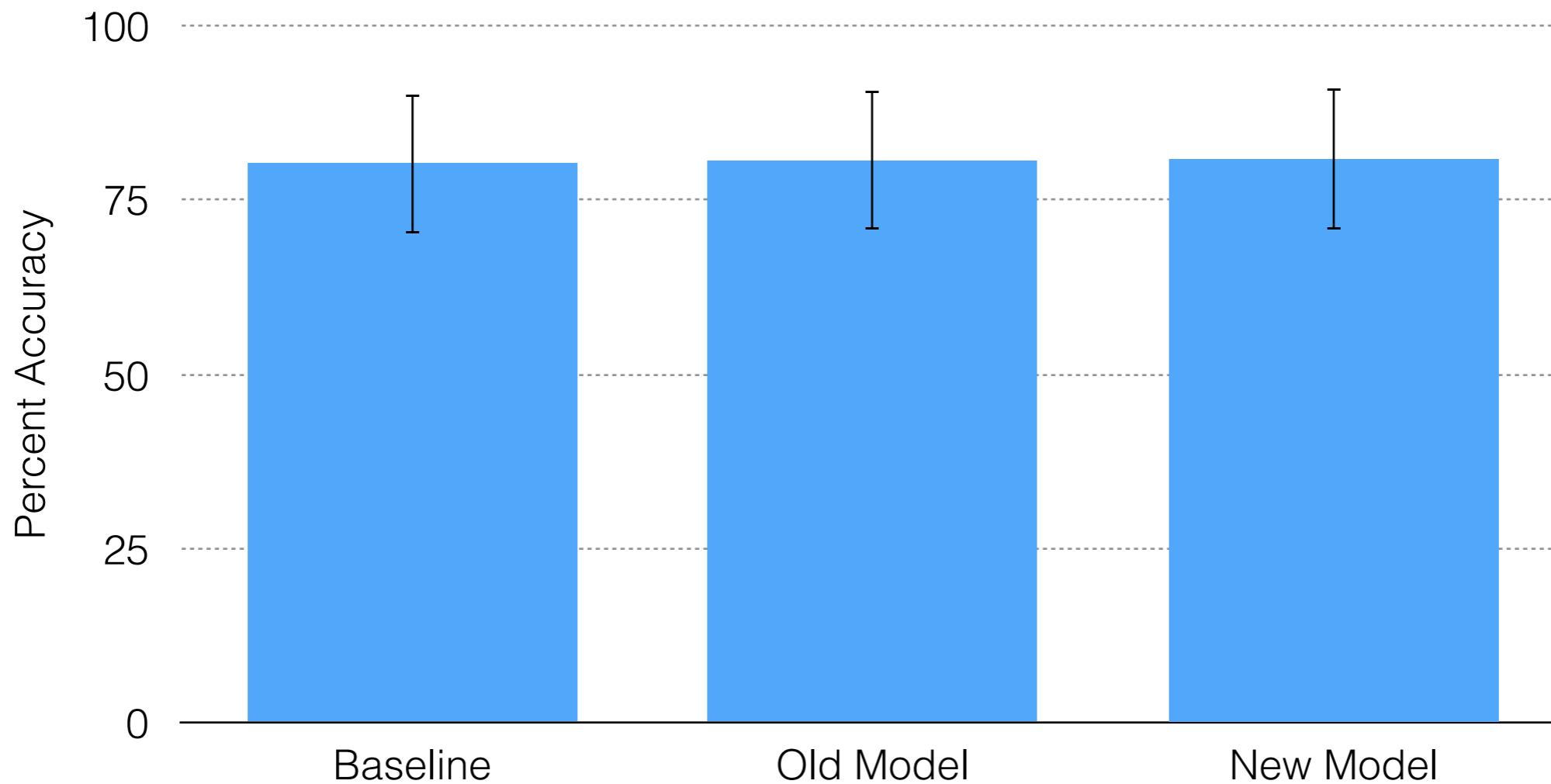


Misleading/Badly Scaled Axes



Rescale to a meaningful range (use full range of values that are interesting/expected)

Misleading/Badly Scaled Axes

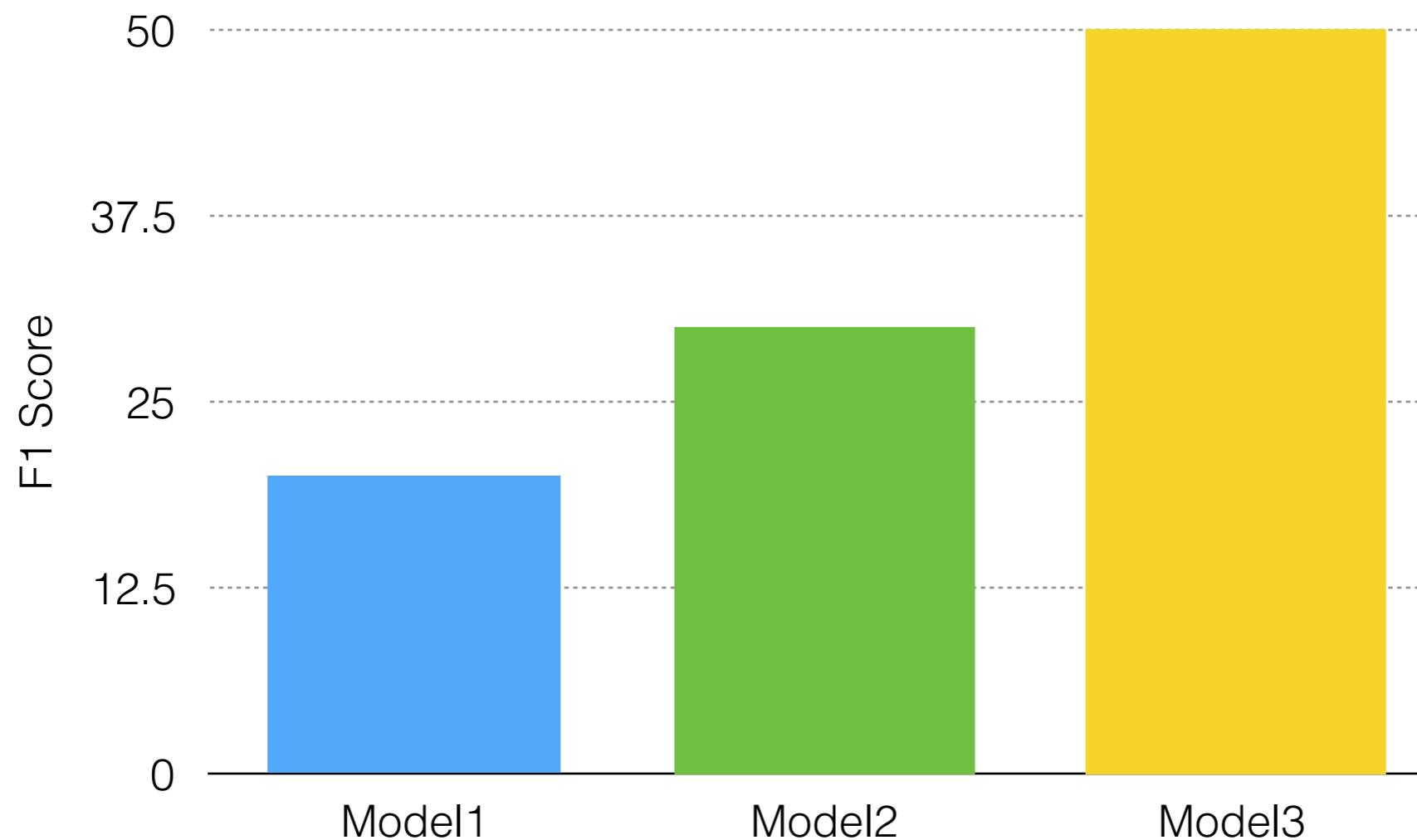


And/or include error bars

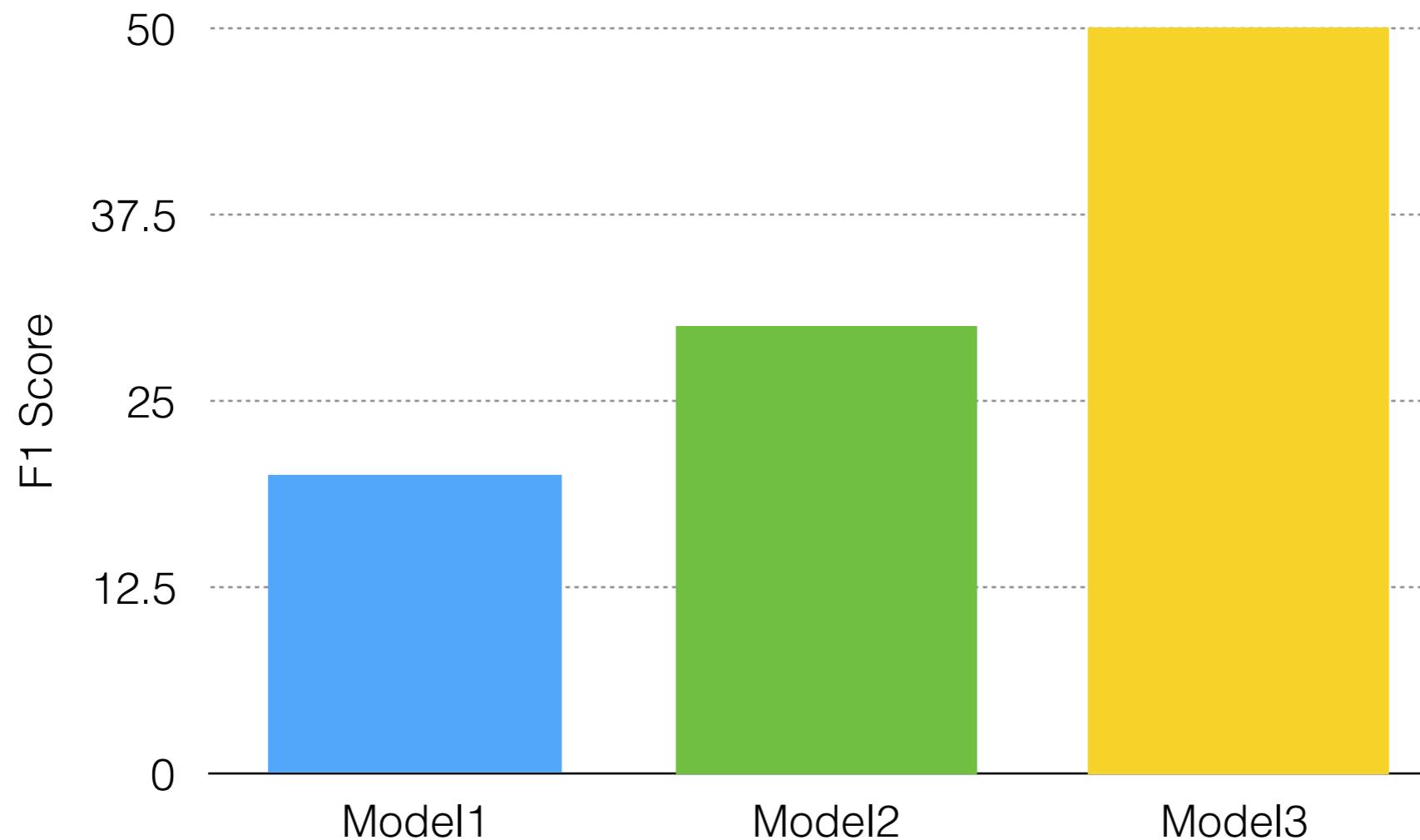
My “three pillars” of Data Viz

Minimalism = Substance over style. Make your point concisely, without redundant or distracting information or ornamentation.

Redundancy

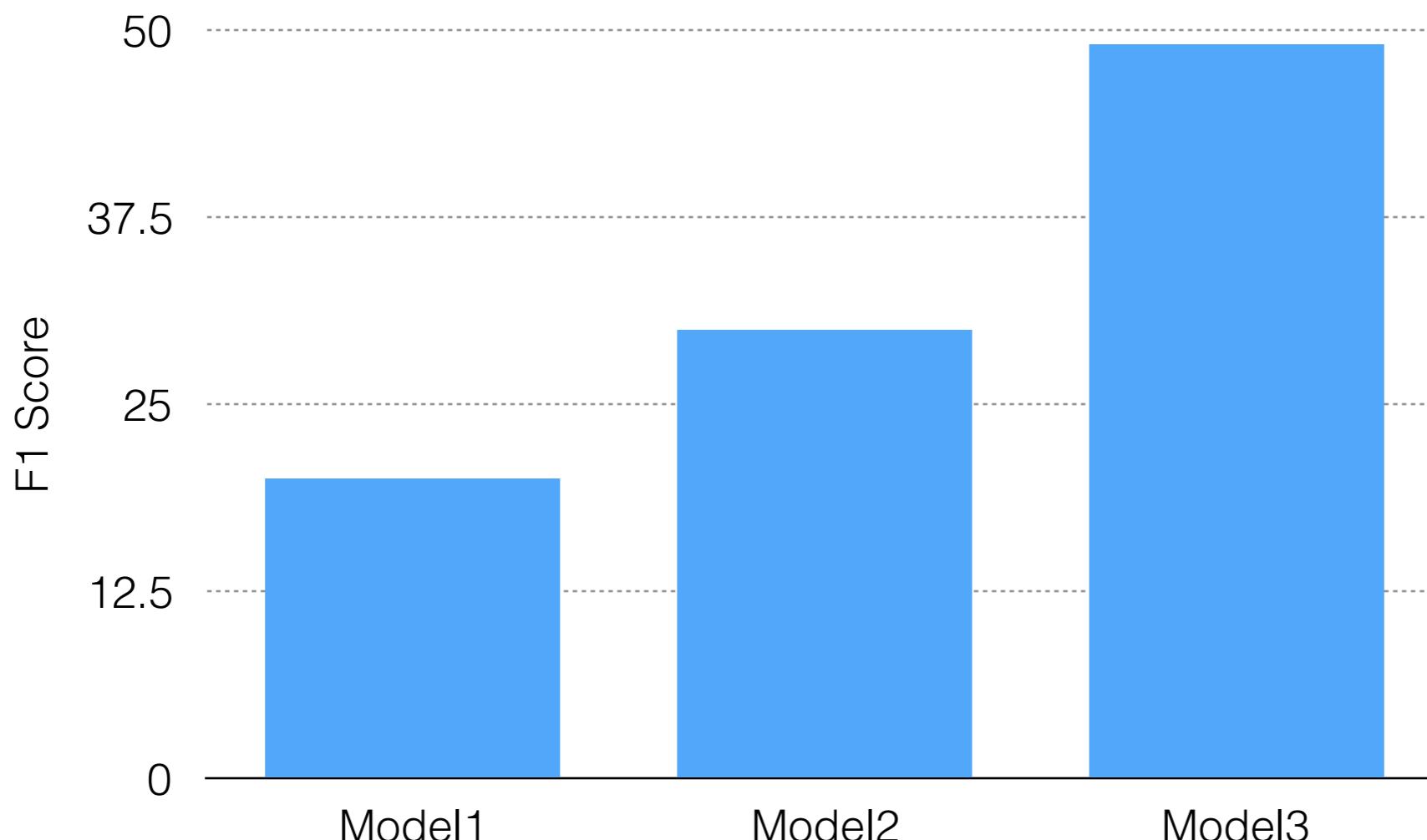


Redundancy



Dont use colors/decorations unless they add new information

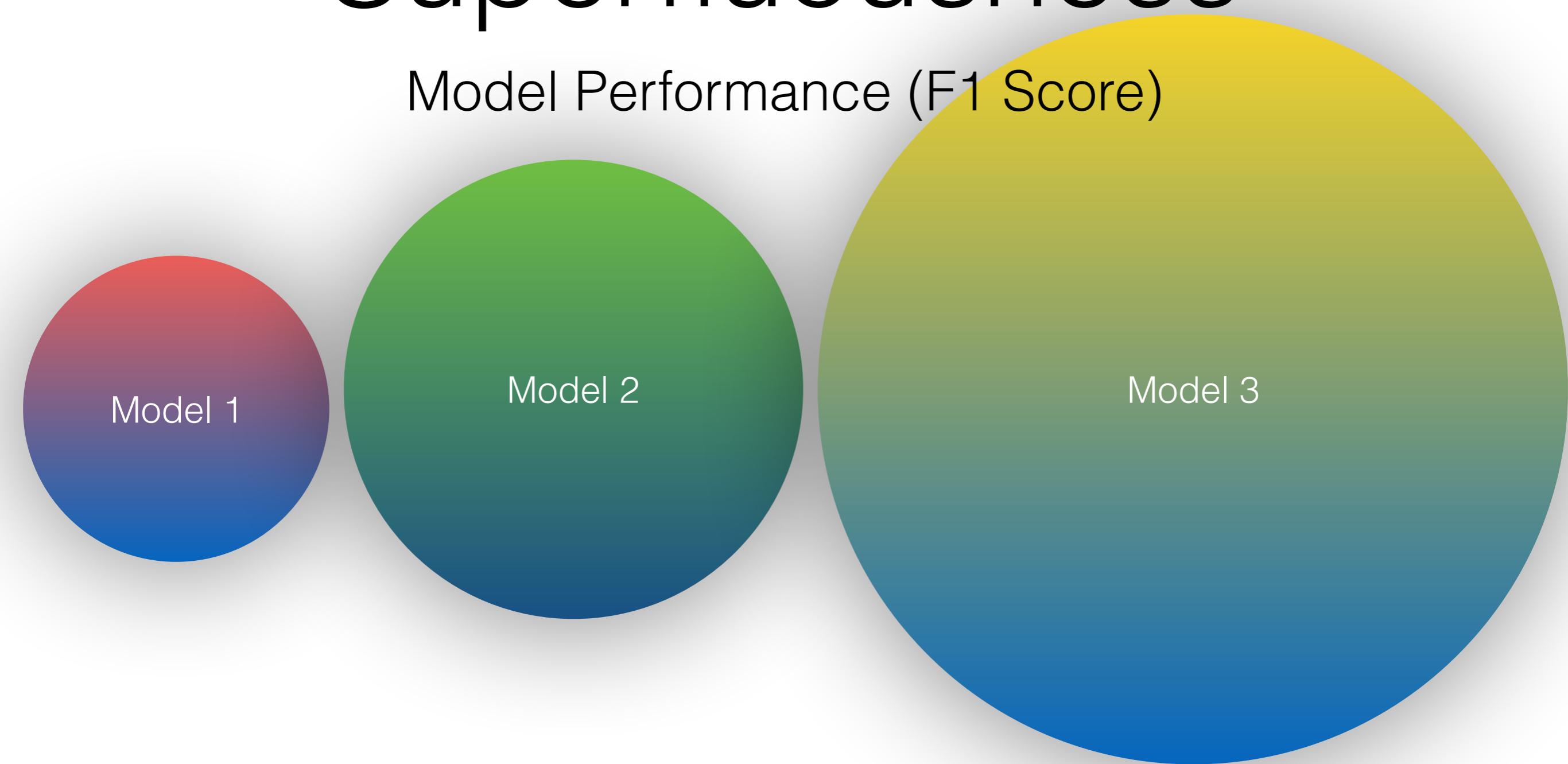
Redundancy



Just look how pretty that is

Superfluously

Model Performance (F1 Score)



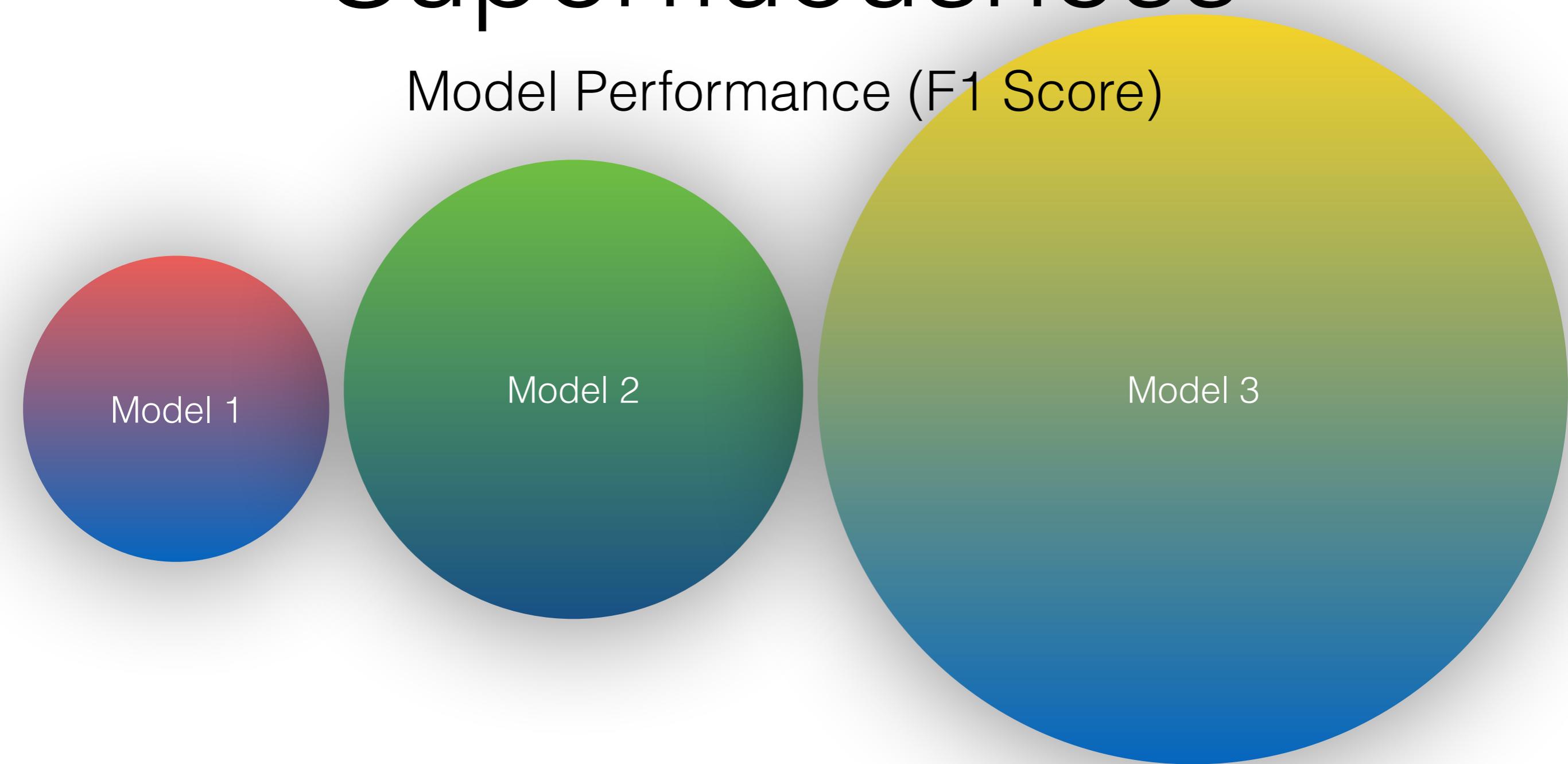
Model 1

Model 2

Model 3

Superfluouslyness

Model Performance (F1 Score)



Model 1

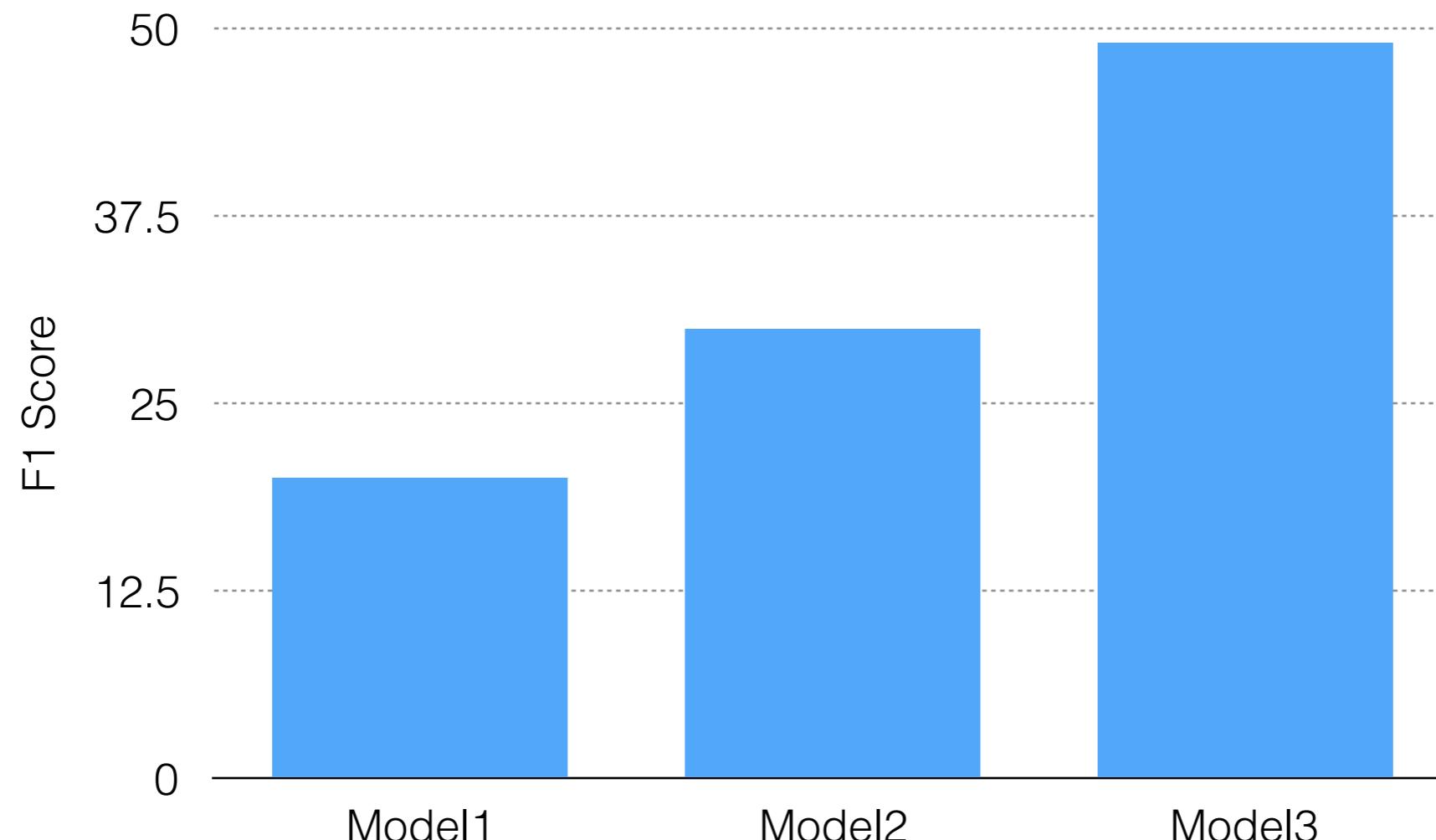
Model 2

Model 3

Dont use colors/decorations unless they add new information

Superfluousness

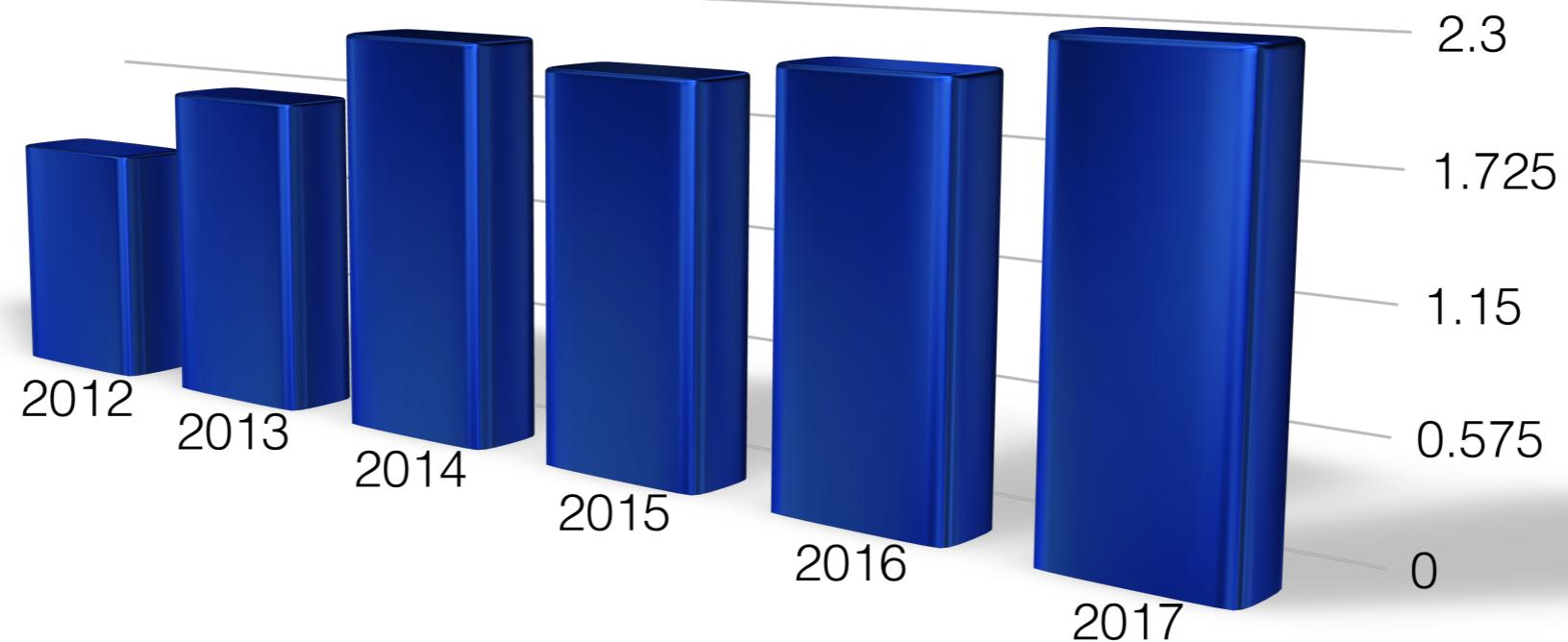
Model Performance (F1 Score)



Just look how pretty that is

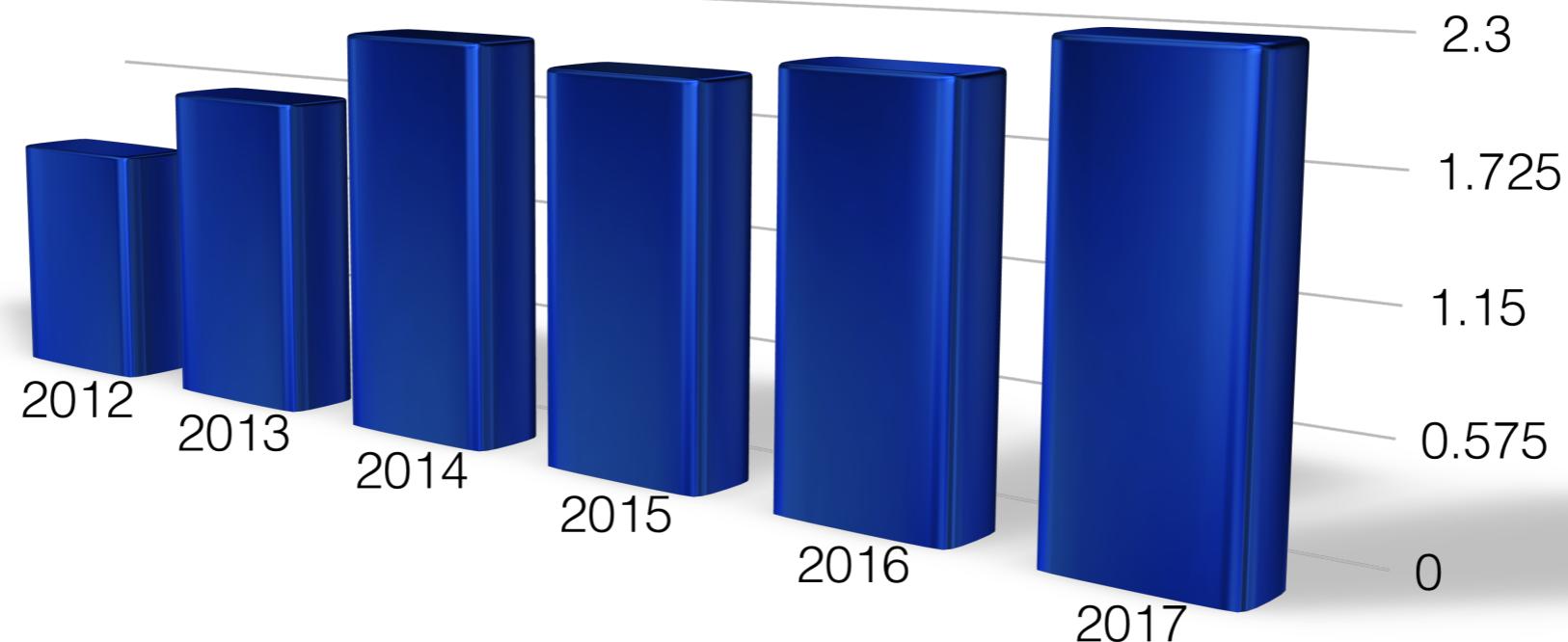
Why is that chart 3D?

Company Earnings by Year (in millions)



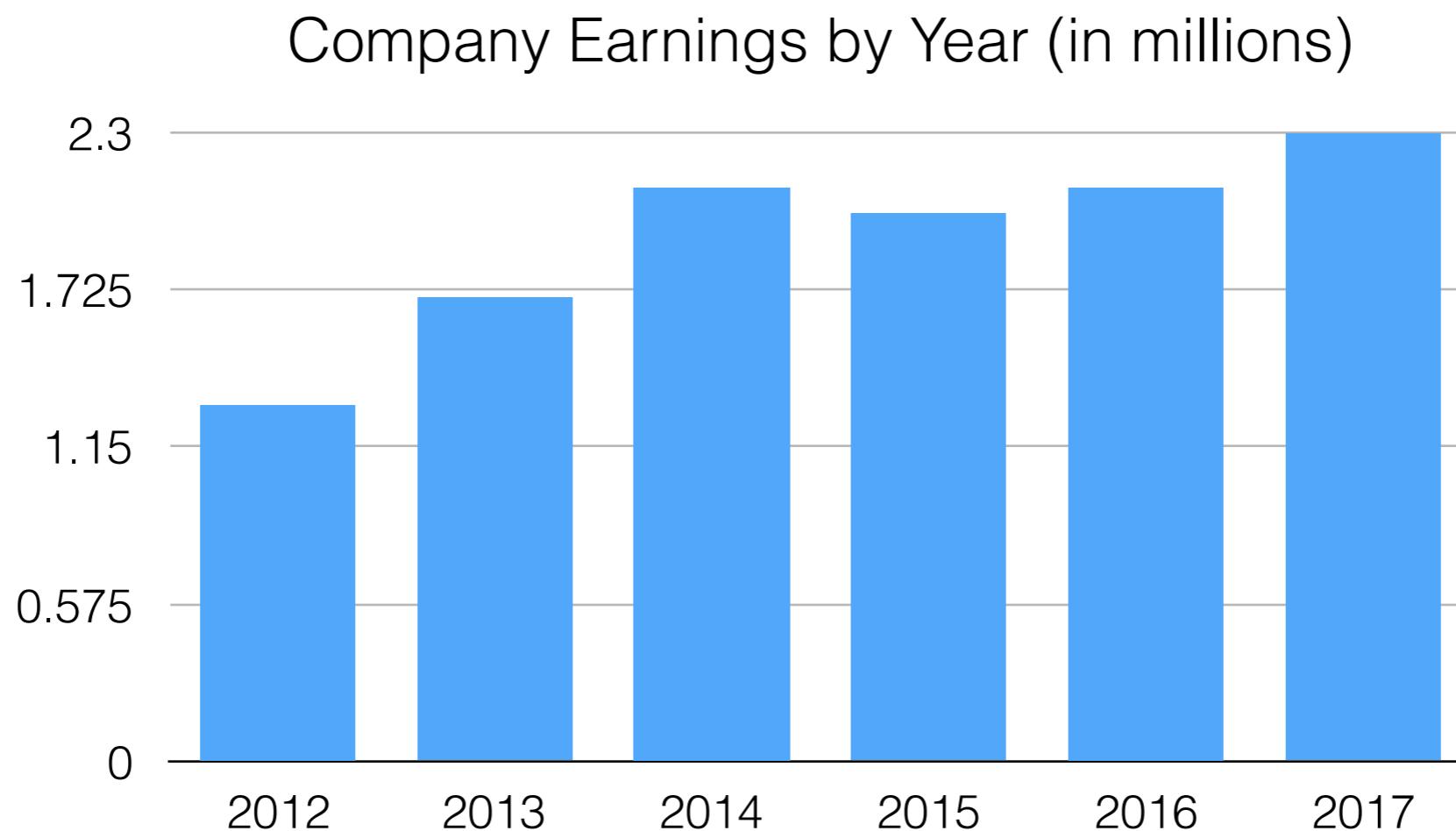
Why is that chart 3D?

Company Earnings by Year (in millions)

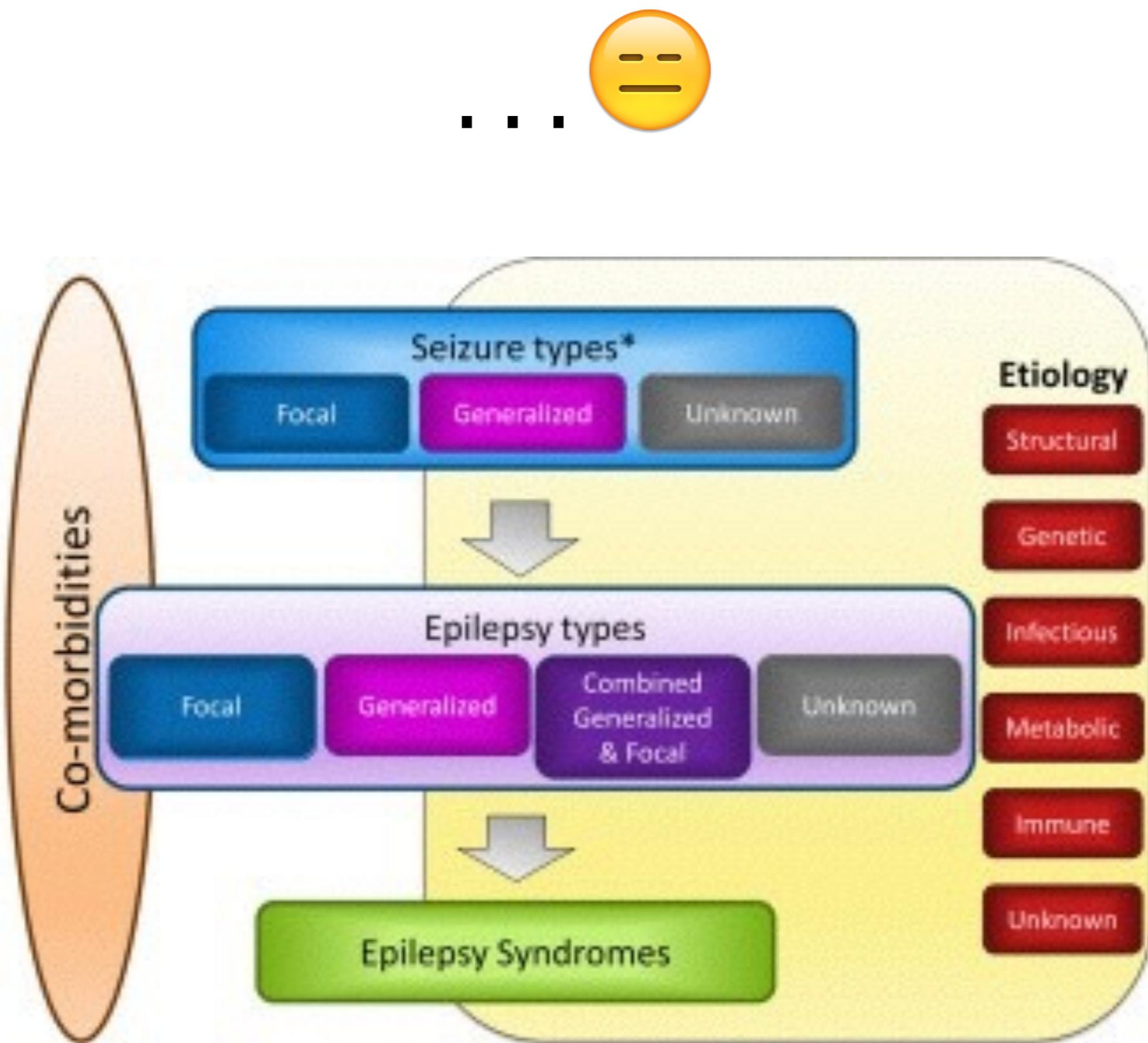


Just...don't

Why is that chart 3D?



Just look how pretty that is





- No where to go from there, so we'll just do a D3 demo...