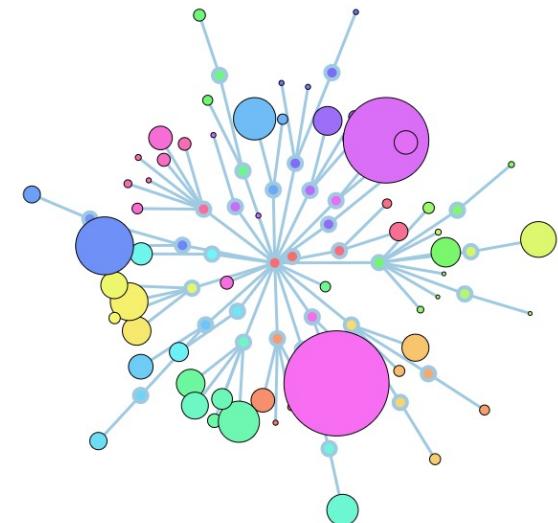


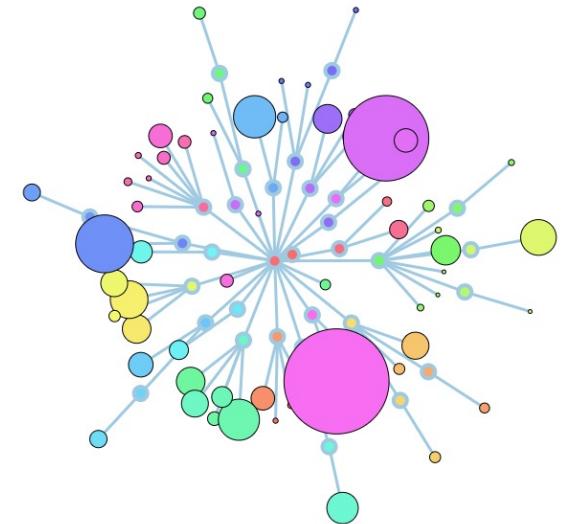
# This is a talk about **data viz**

Dr. Jennifer Piscionere, PhD

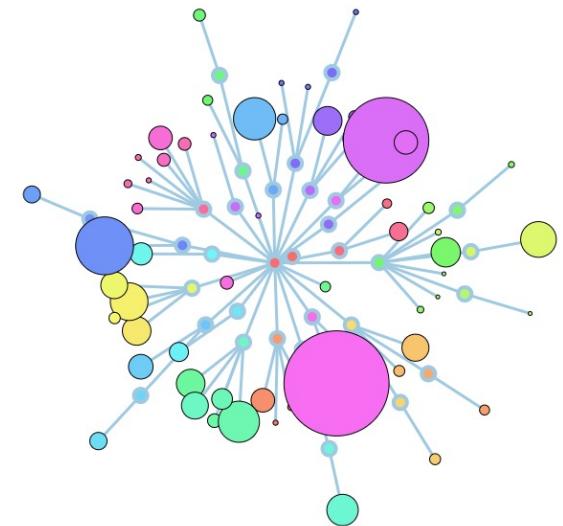
Data Scientist @ GenV



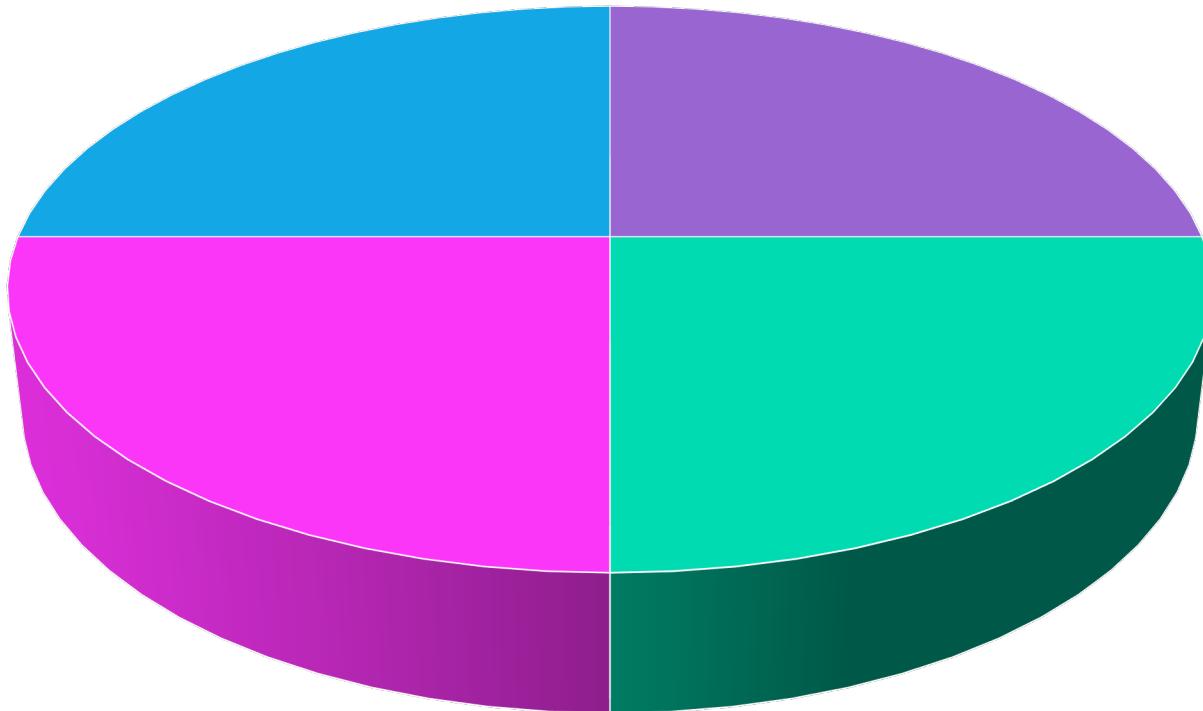
# This is a talk about communication



This is a talk about **science**

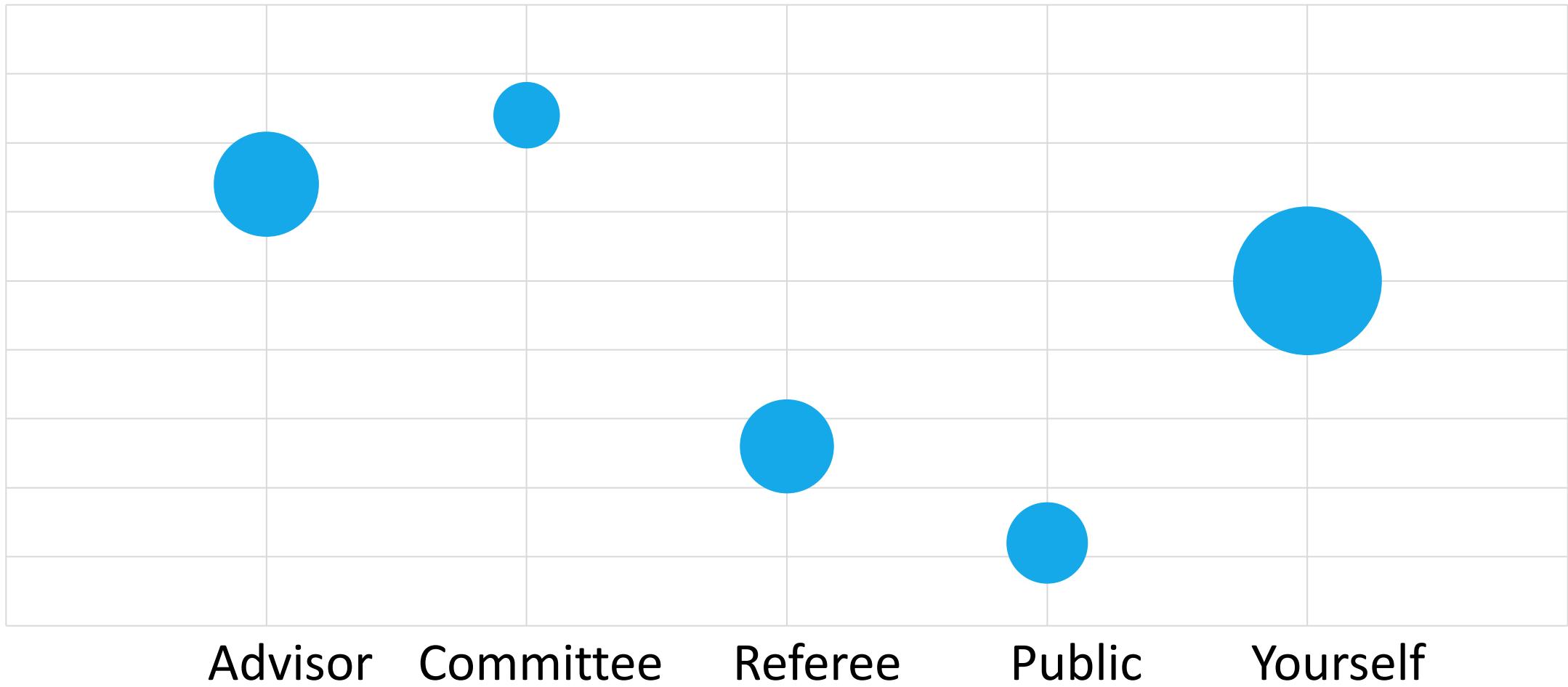


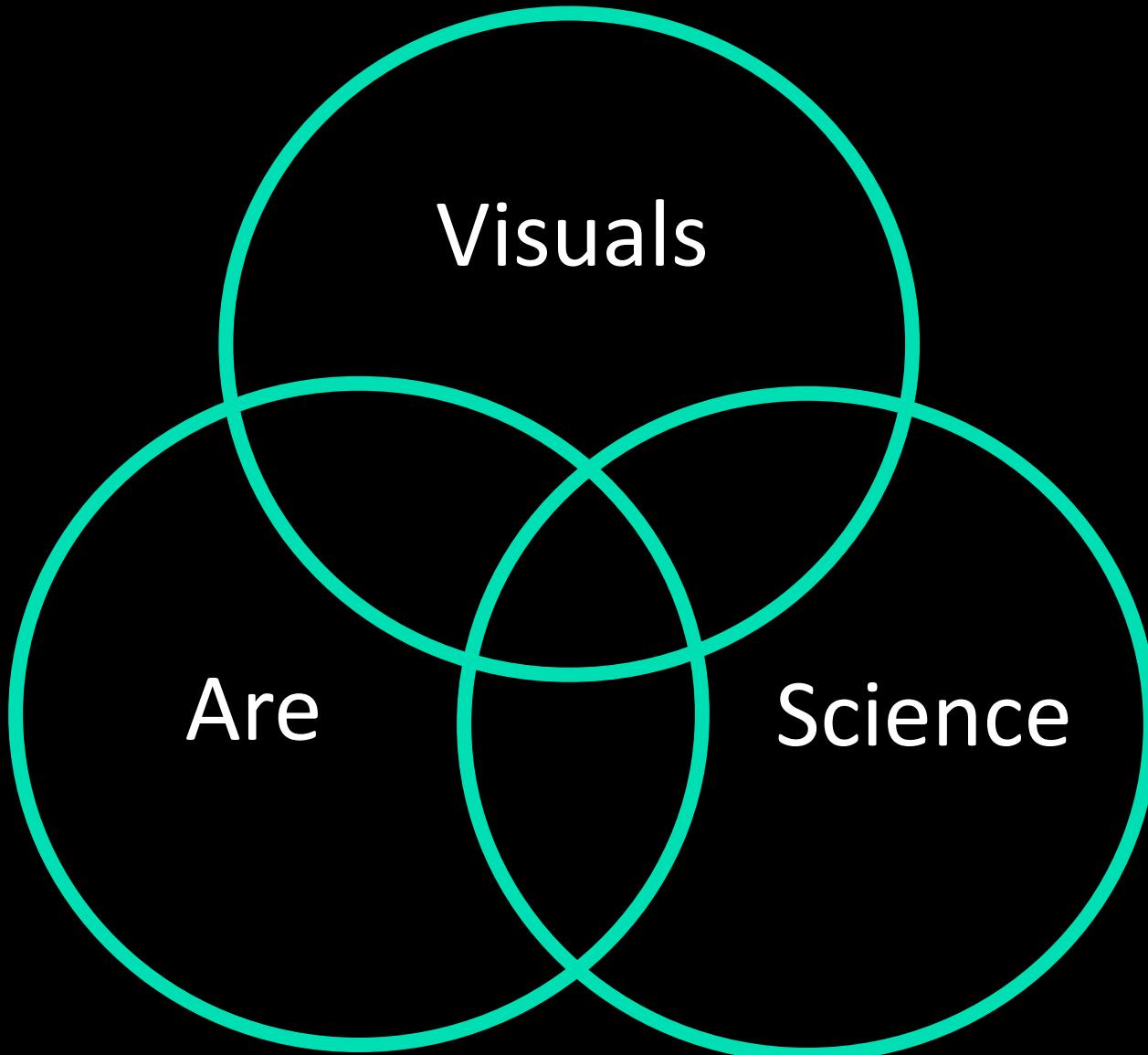
# We



■ Learn ■ Through ■ Visual ■ Aids

Your job is to convince {\ref} of your science





■ Good ■ Communication ■ Requires

Quality



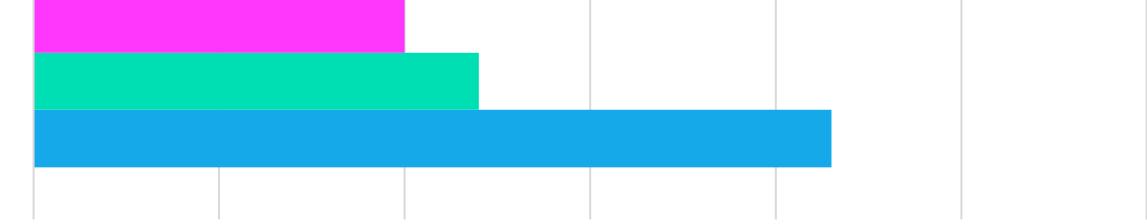
Quantity



Relevance



Manner



Say everything you need to say, and nothing you don't.

Say everything you need to say, and nothing you don't.

Everything you say should be true and backed by evidence.

Say everything you need to say, and nothing you don't.

Everything you say should be true and backed by evidence.

Everything you say should be appropriate for the audience you are speaking to.

Say everything you need to say, and nothing you don't.

Everything you say should be true and backed by evidence.

Applicable to plots as well as talks

Everything you say should be appropriate for the audience you are speaking to.

Say everything you need to say, and nothing you don't.

Everything you say should be true and backed by evidence.

**No viz is created in a vacuum**

Everything you say should be appropriate for the audience you are speaking to.

Papers

Presentations

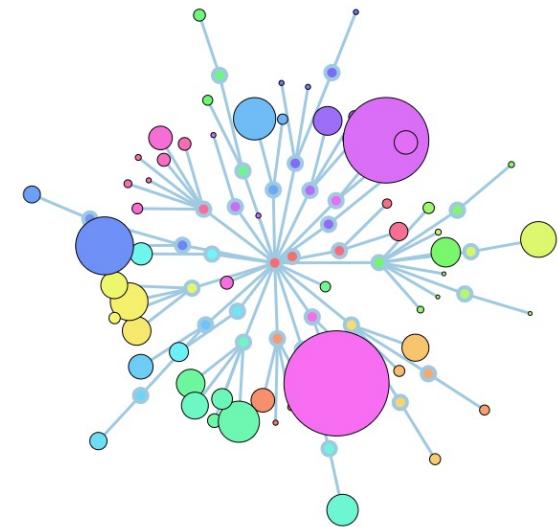
Special  
Projects

Papers

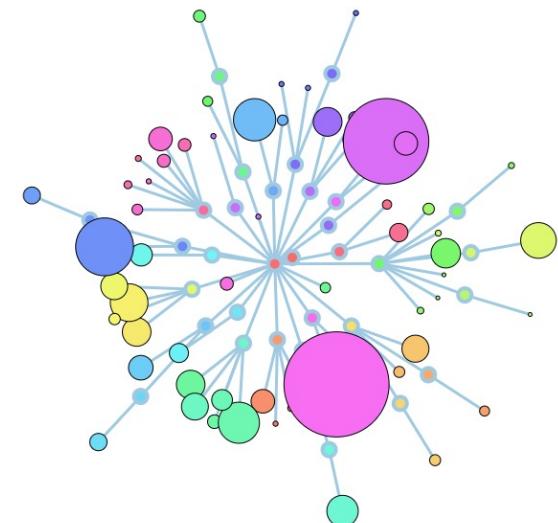
Presentations

Special  
Projects

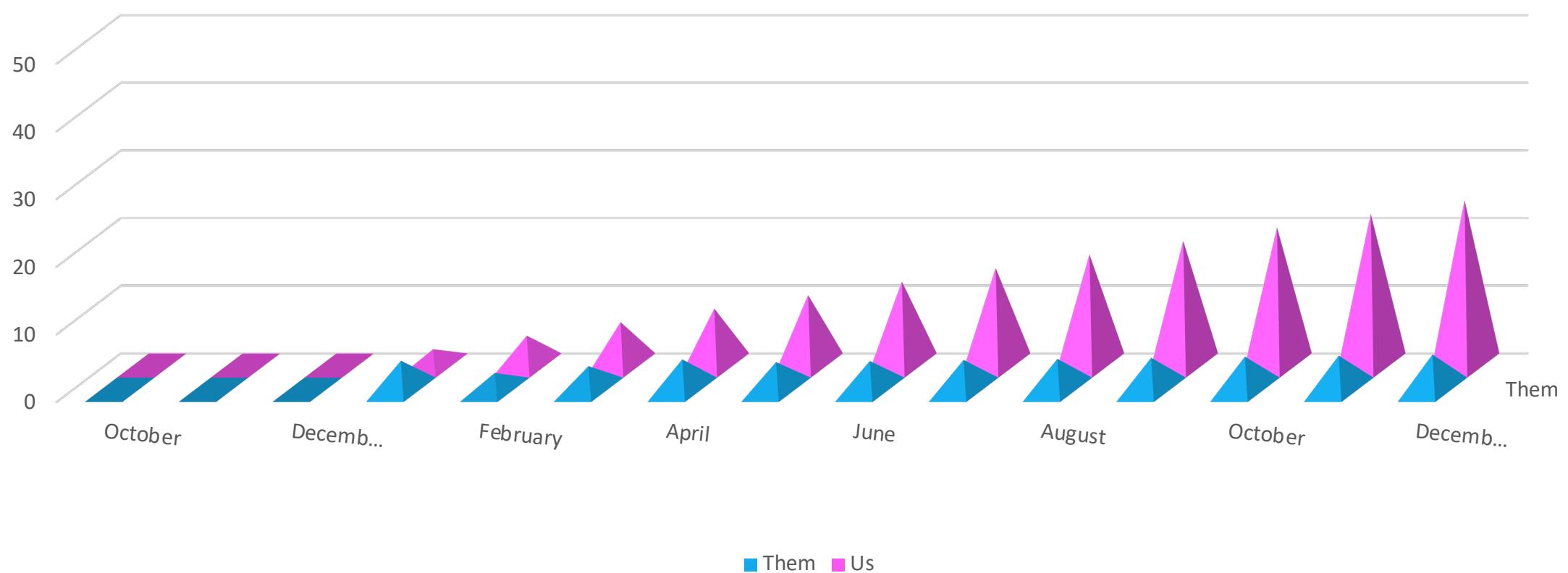
# The **key** to a good plot



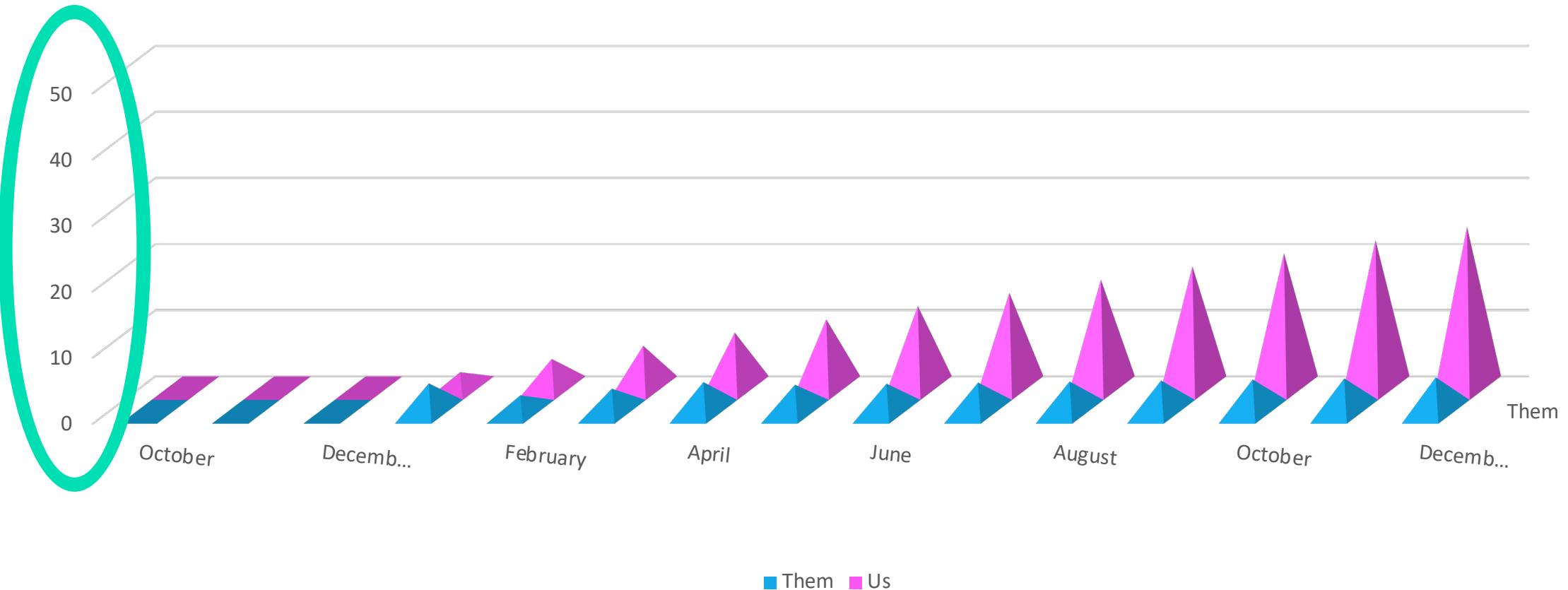
Remove unnecessary data ink  
and **emphasize** the most  
important data ink left.



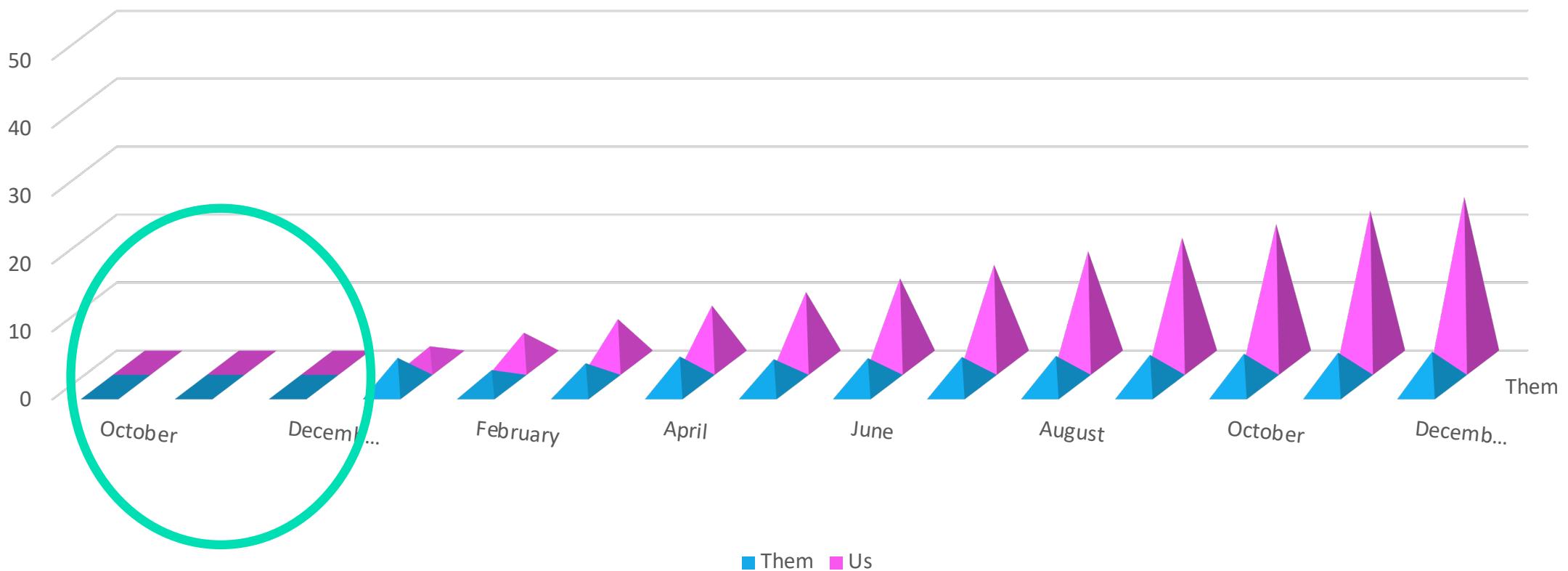
# WHY is this a terrible plot?



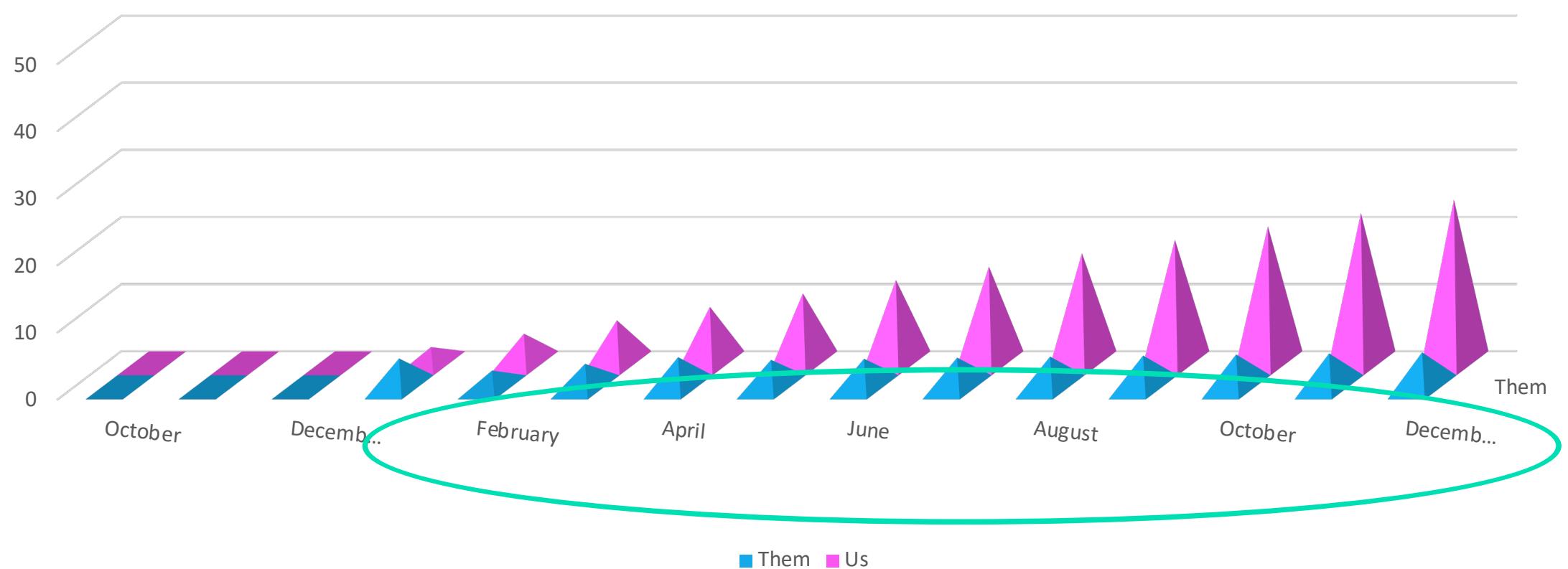
# WHY is this a terrible plot?



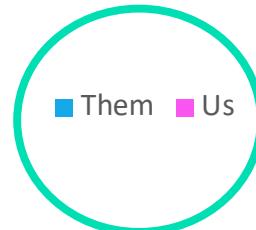
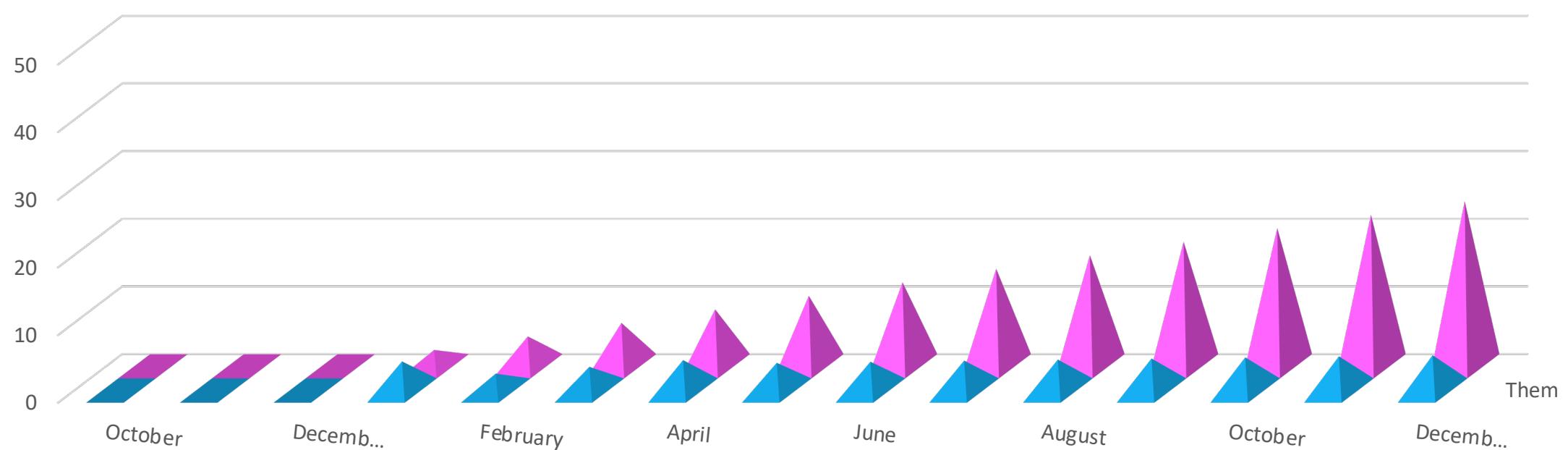
# WHY is this a terrible plot?



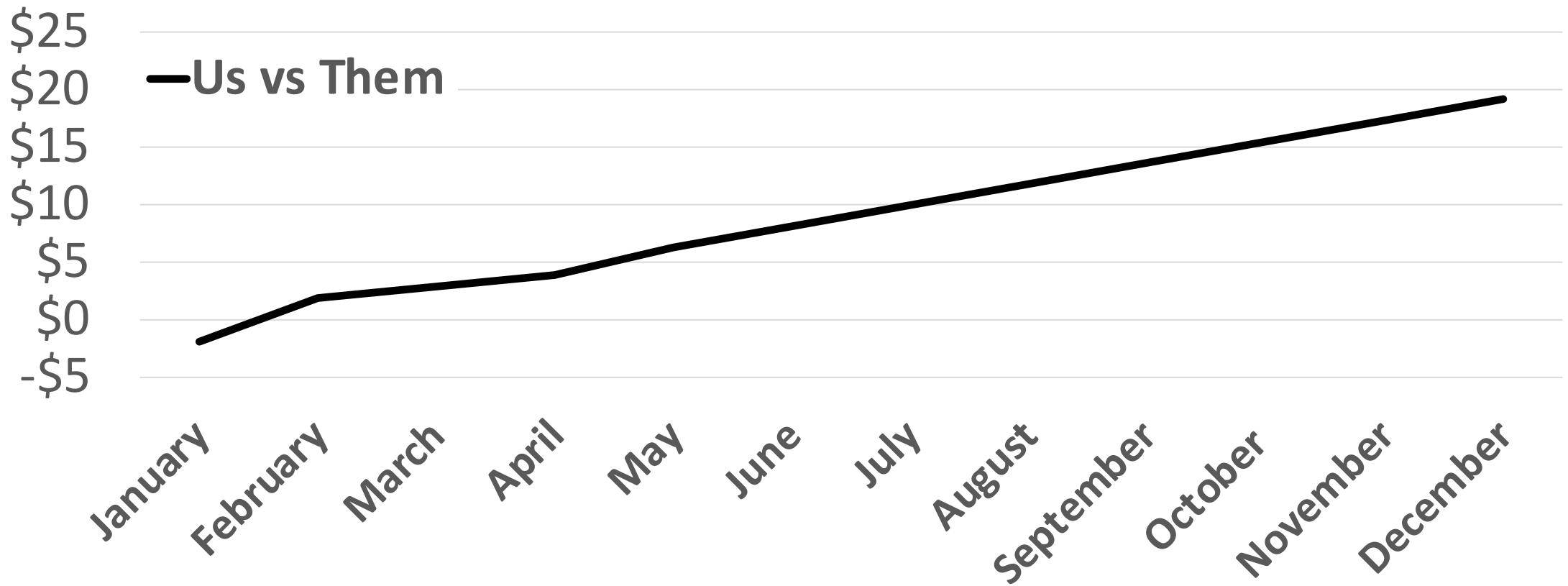
# WHY is this a terrible plot?



# WHY is this a terrible plot?



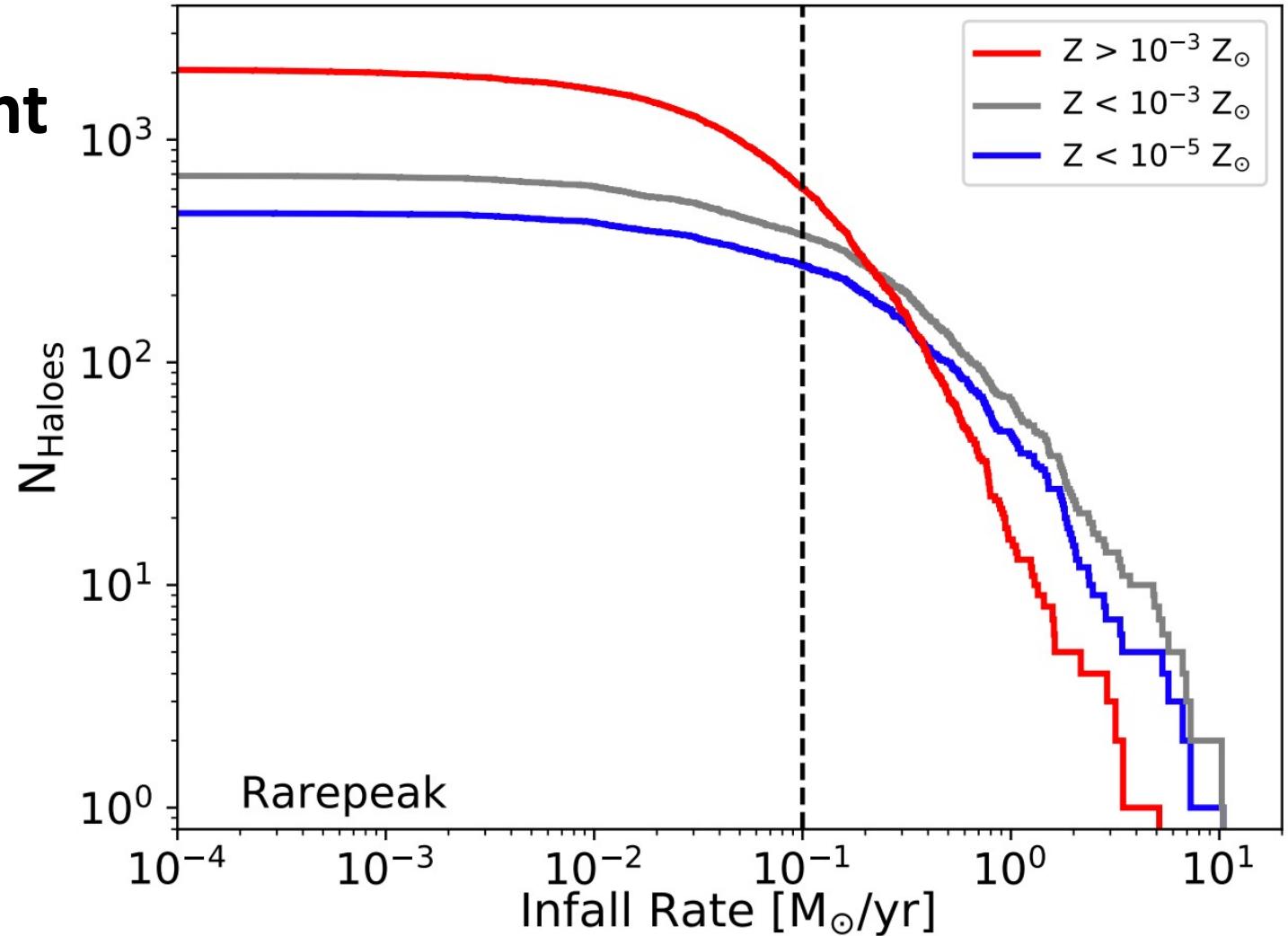
# We Are Selling More Than Them



# Anatomy of a Plot

1

## Graphical Element



# When we are examining data, what can we look for?

- Does this data describe a **geometric** object?
- Are the data points **connected** to each other?
- Can we describe data points with a fixed set of **categories**?
- Is there a **quantity** associated with the data?
- Are the datapoints **continuous** along one or more dimensions?

# When we are examining data, what can we look for?

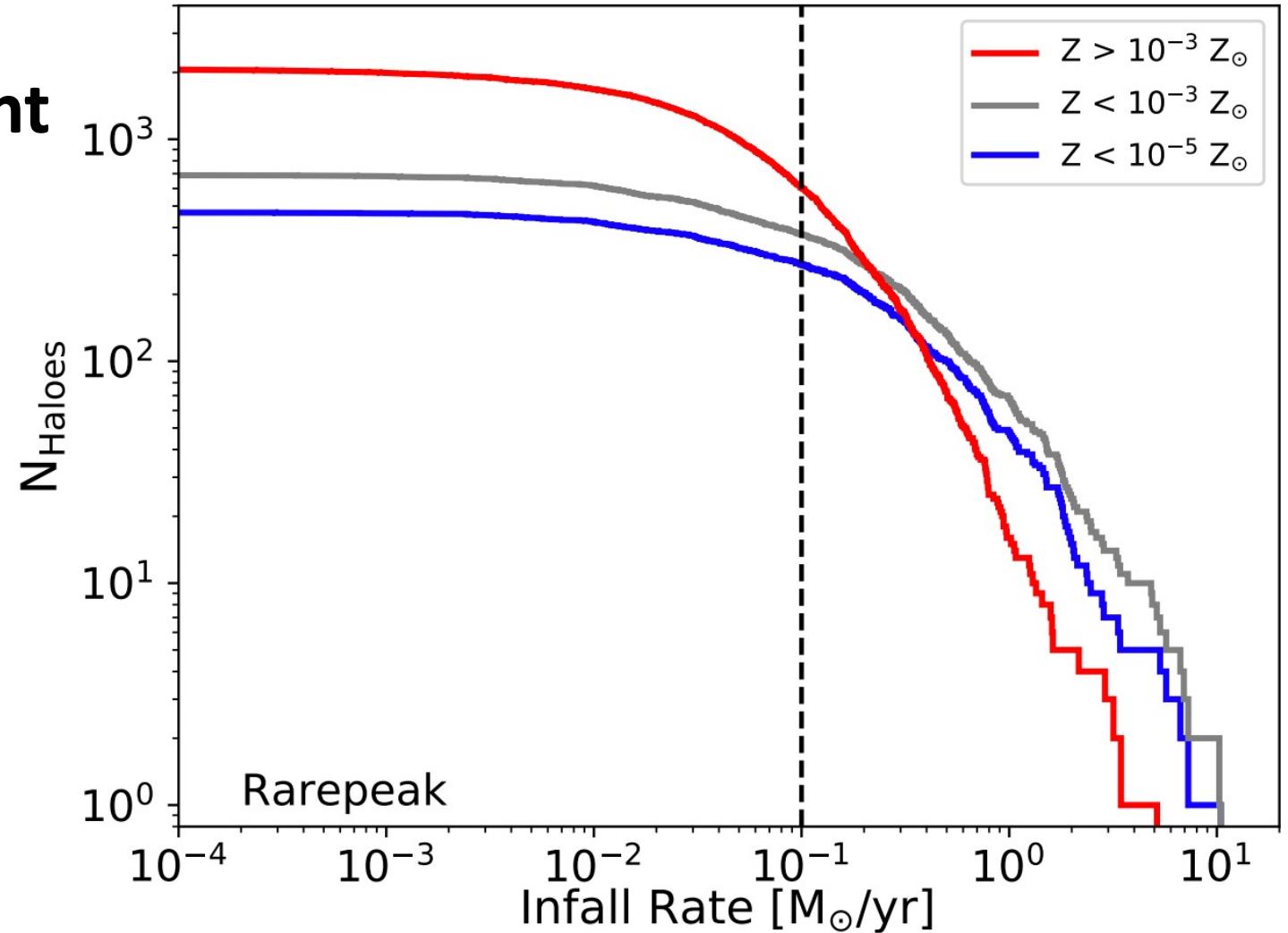
- Does this data describe a **geometric** object?
- Are the data points **connected** to each other?
- Can we describe data points with a fixed set of **categories**?
- Is there a **quantity** associated with the data?
- Are the datapoints **continuous** along one or more dimensions?

## Is your data **categorical** or **continuous**?

# Anatomy of a Plot

1

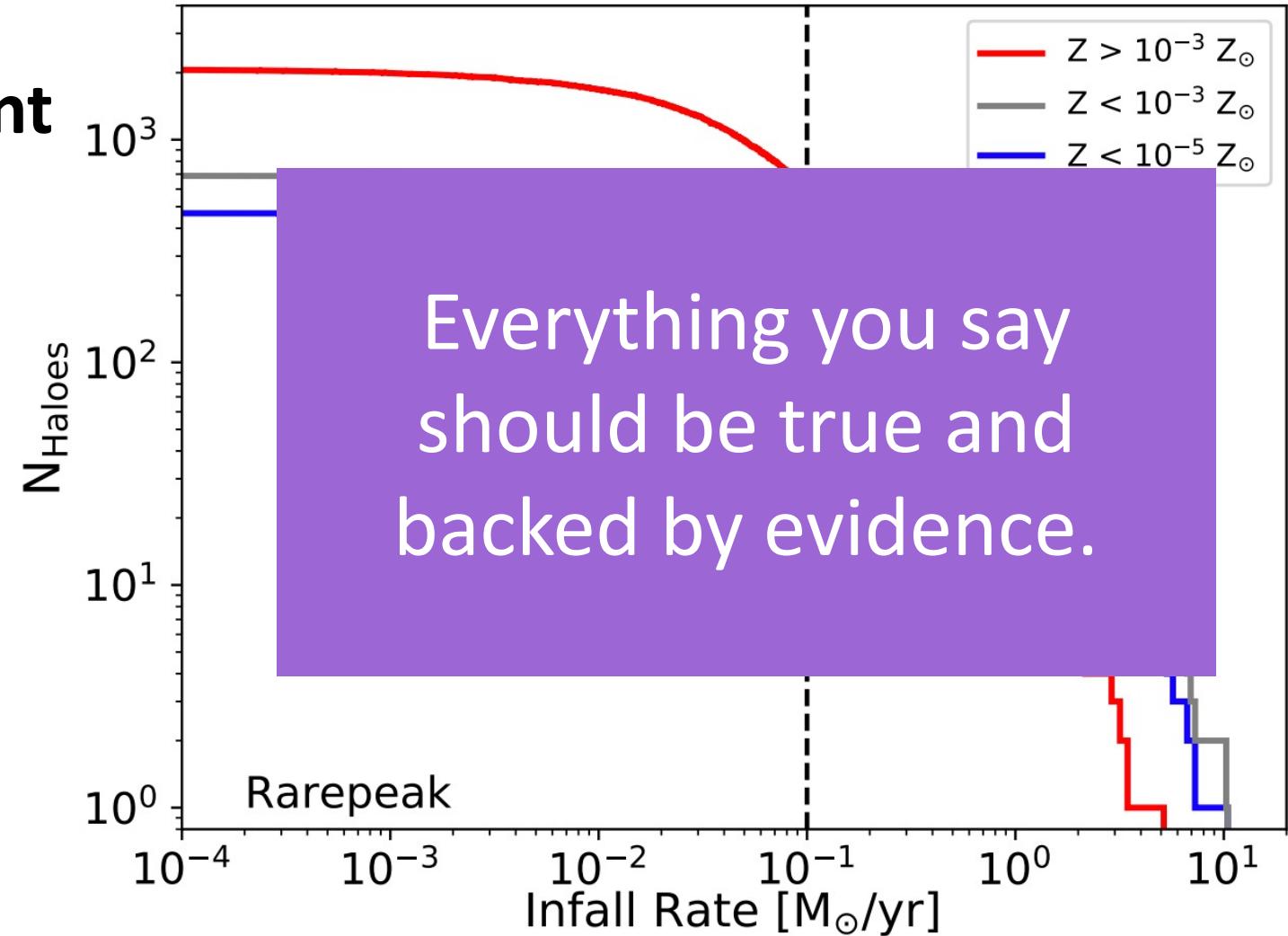
## Graphical Element



# Anatomy of a Plot

1

## Graphical Element



A few  
tips

Be Mindful of  
Binning

Special  
Projects

A few  
tips

Avoid automatic line  
fitters and  
smoothers

Special  
Projects

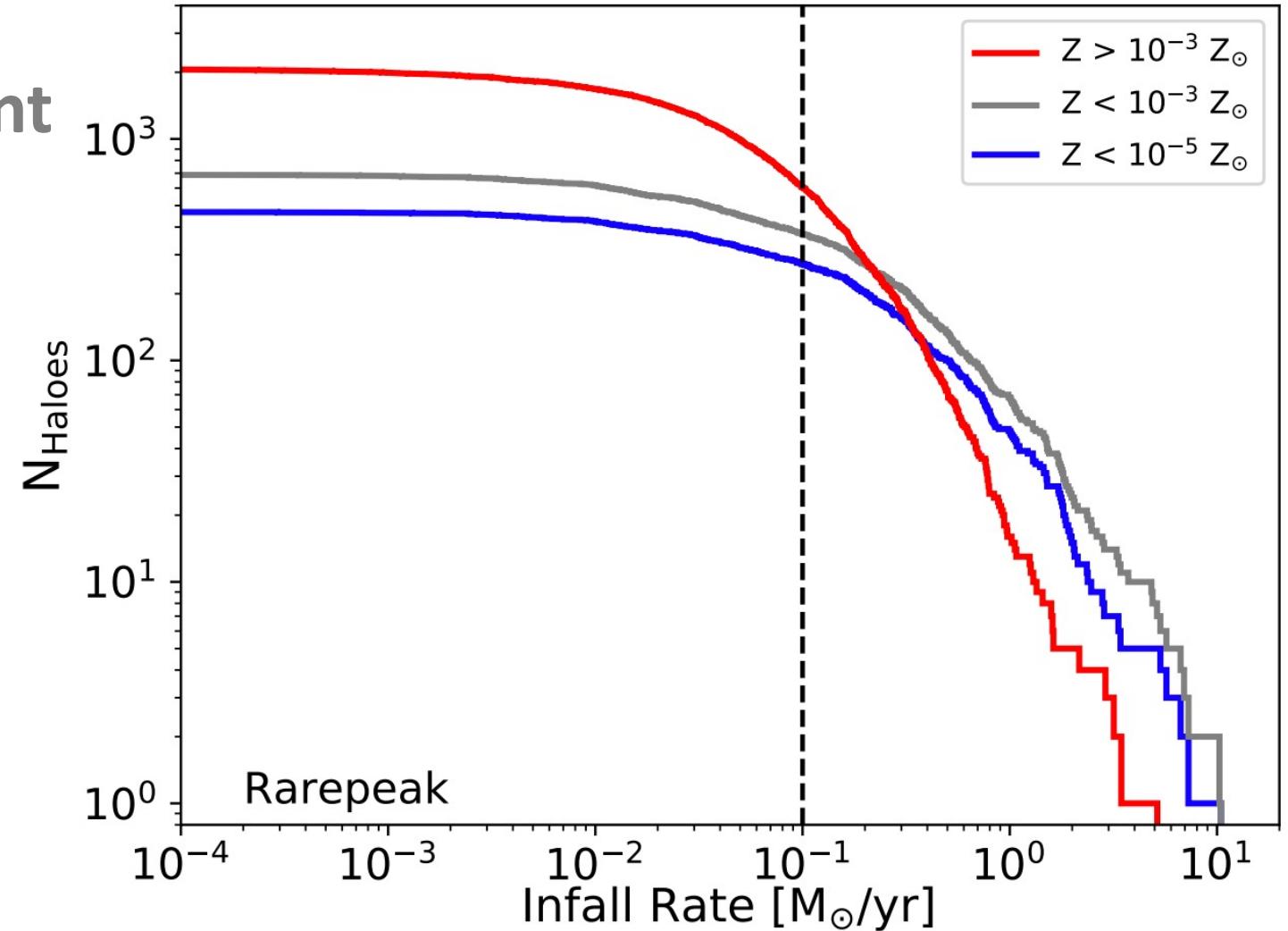
# Anatomy of a Plot

1

Graphical Element

2

Colors



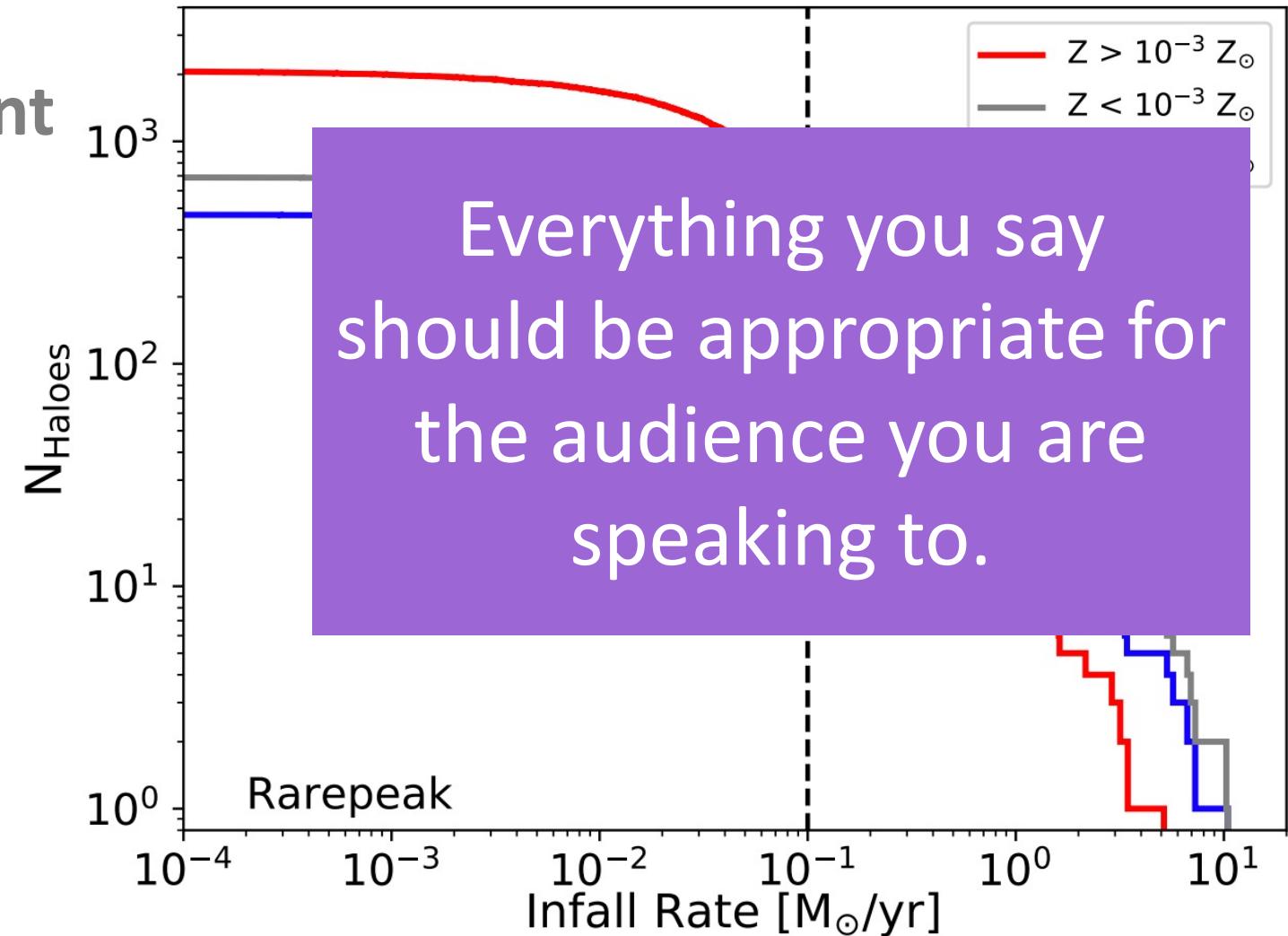
# Anatomy of a Plot

1

Graphical Element

2

Colors



A few  
tips

Make it black and  
white friendly

Special  
Projects

A few  
tips

Make it colorblind  
friendly

Special  
Projects

A few  
tips

Avoid Green

Special  
Projects

# Anatomy of a Plot

1

**Graphical Element**

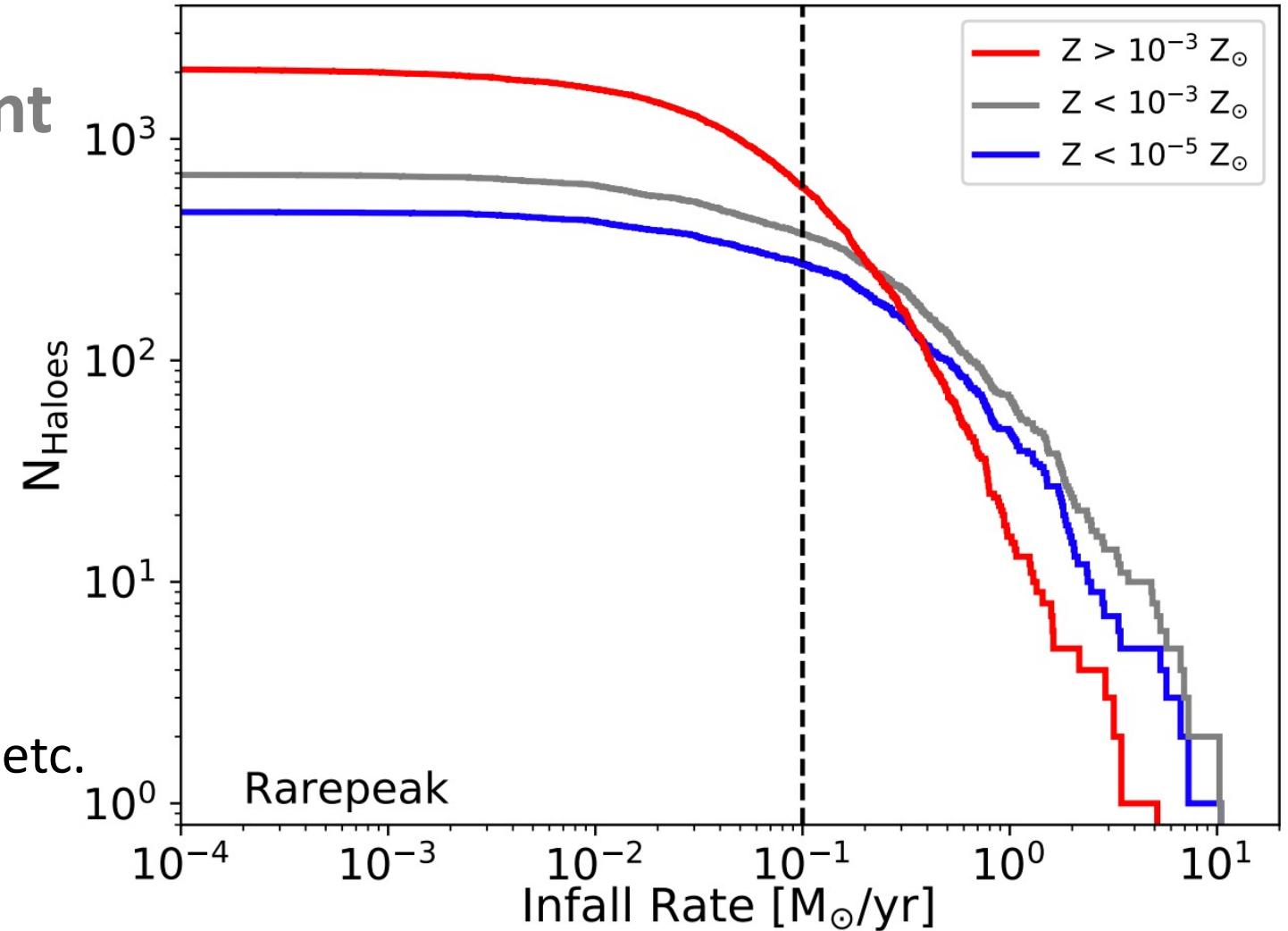
2

**Colors**

3

**Plot theme**

Axis labels, tick marks, etc.



# Anatomy of a Plot

1

**Graphical Element**

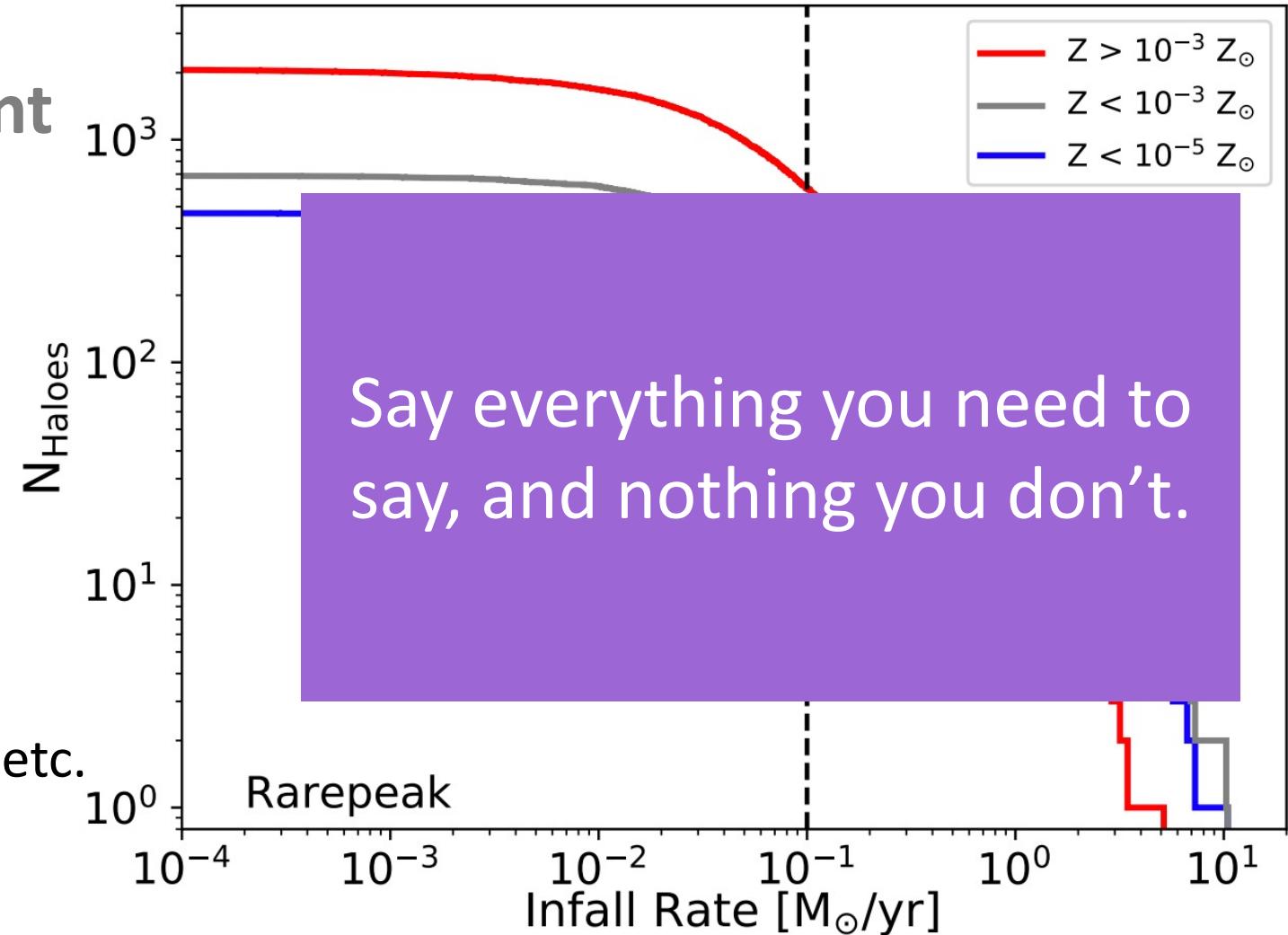
2

**Colors**

3

**Plot theme**

Axis labels, tick marks, etc.



A few  
tips

Set up your plot  
theme first

Special  
Projects

## A few tips

This includes colors,  
tick mark size, axis  
placement, fonts,  
etc.

## Special Projects

A few  
tips

Be confident about  
your(log-log) axes

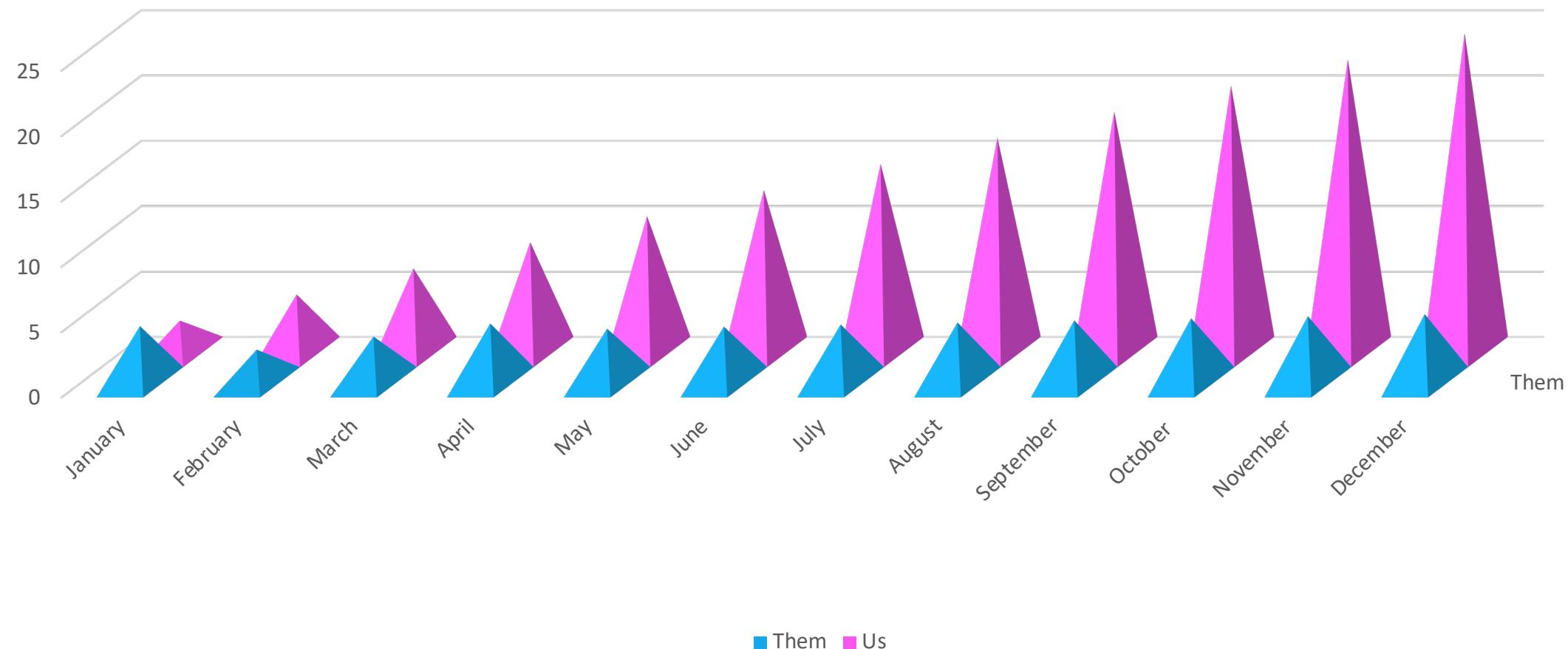
Special  
Projects

A few  
tips

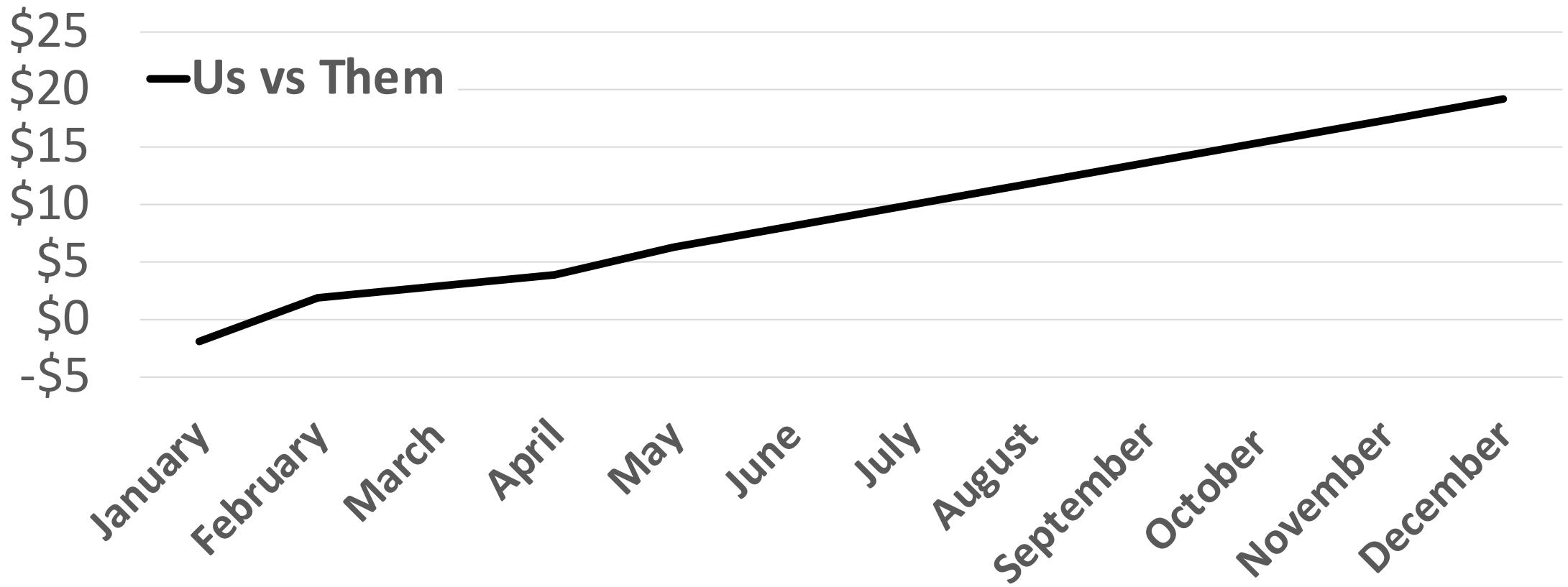
Don't Make People  
Figure Out the Point

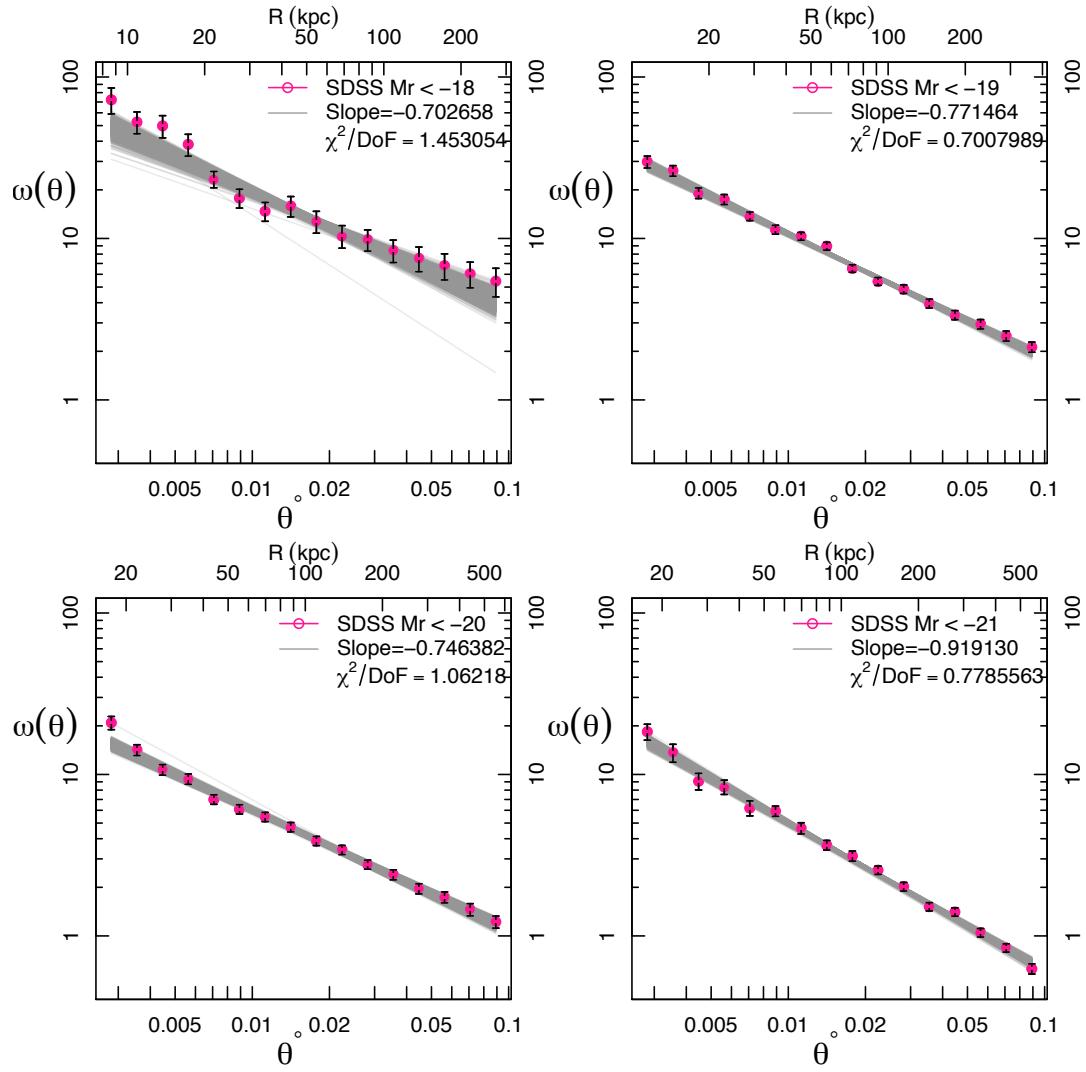
Special  
Projects

This is a terrible plot. You would NEVER make this kind of plot in excel right?



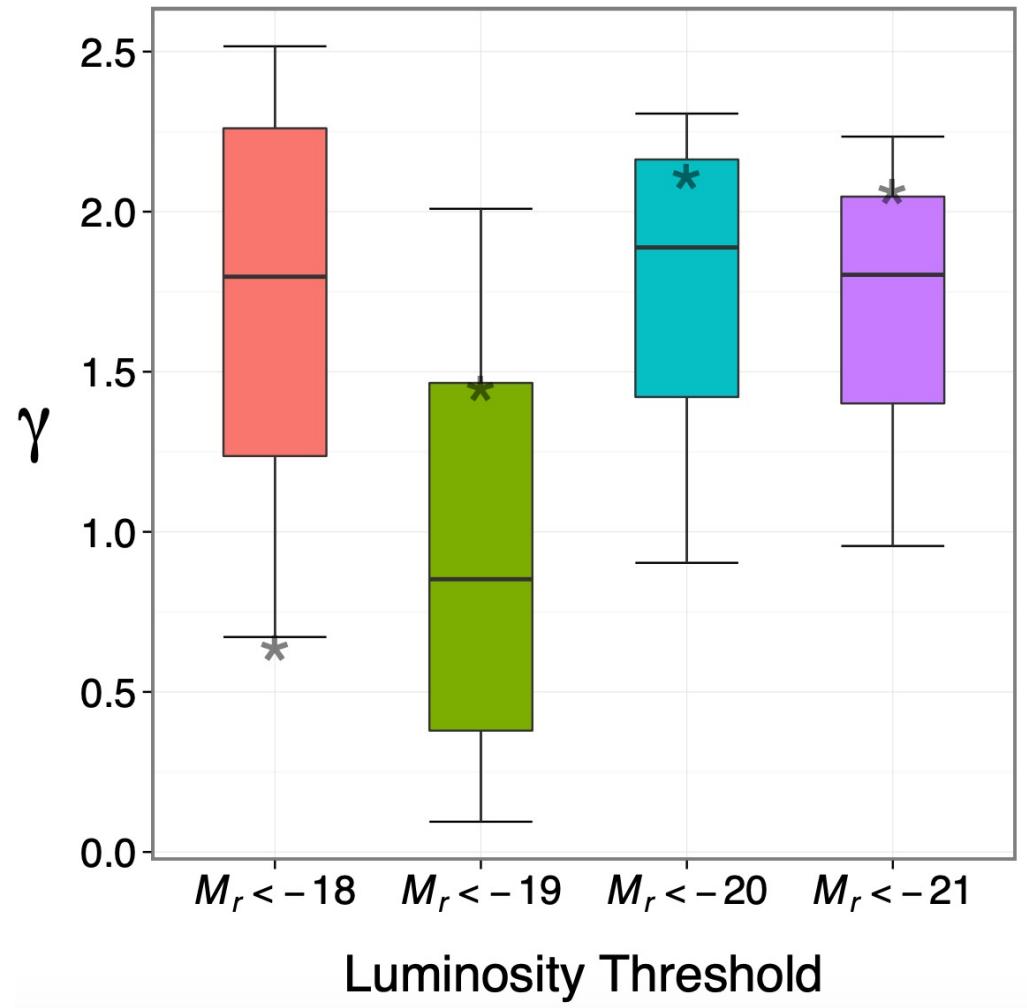
# We Are Selling More Than Them





VS

## Best Fit Slope $\gamma$ of Each Sample



Papers

Presentations

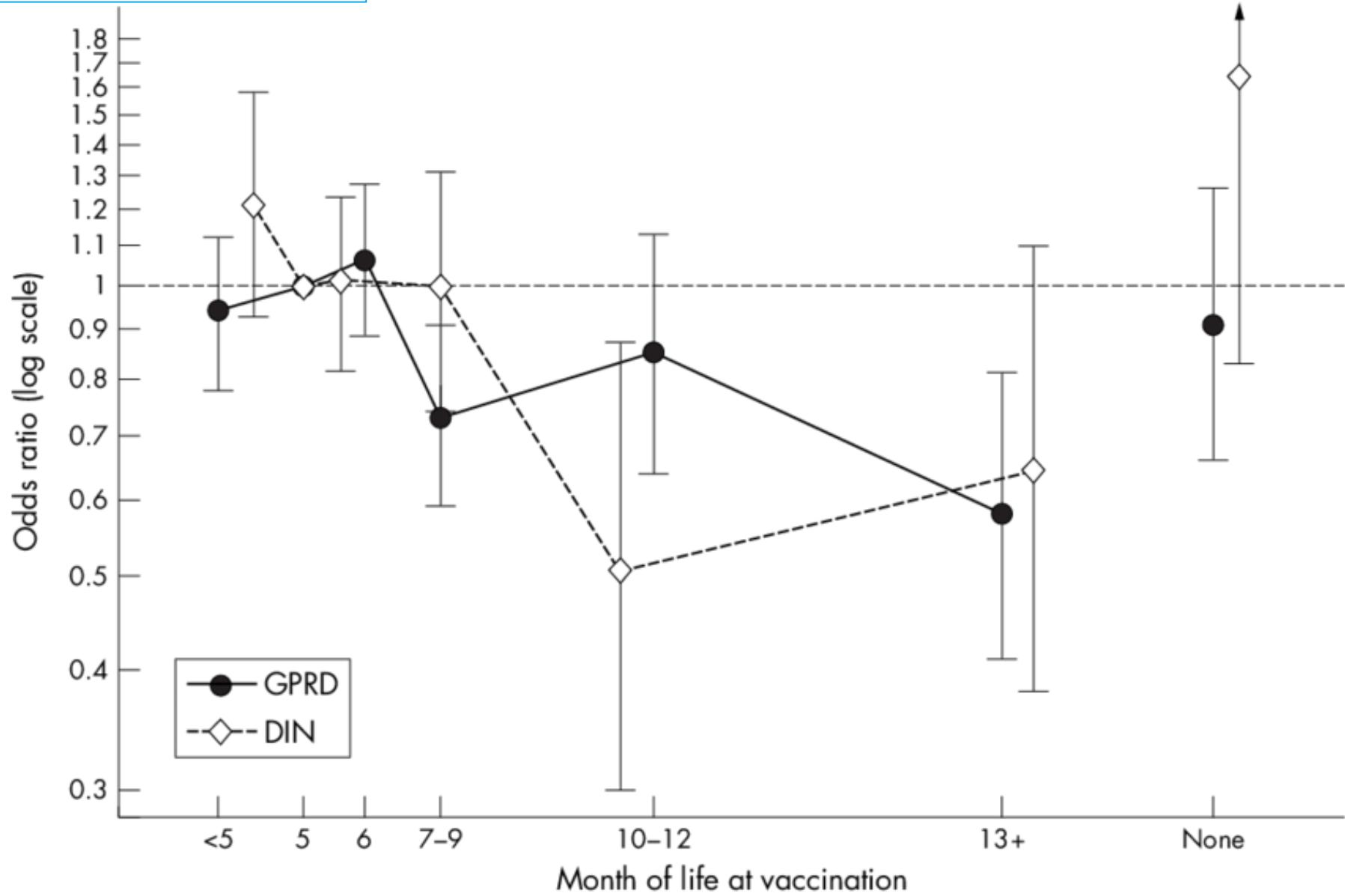
Special  
Projects

Papers

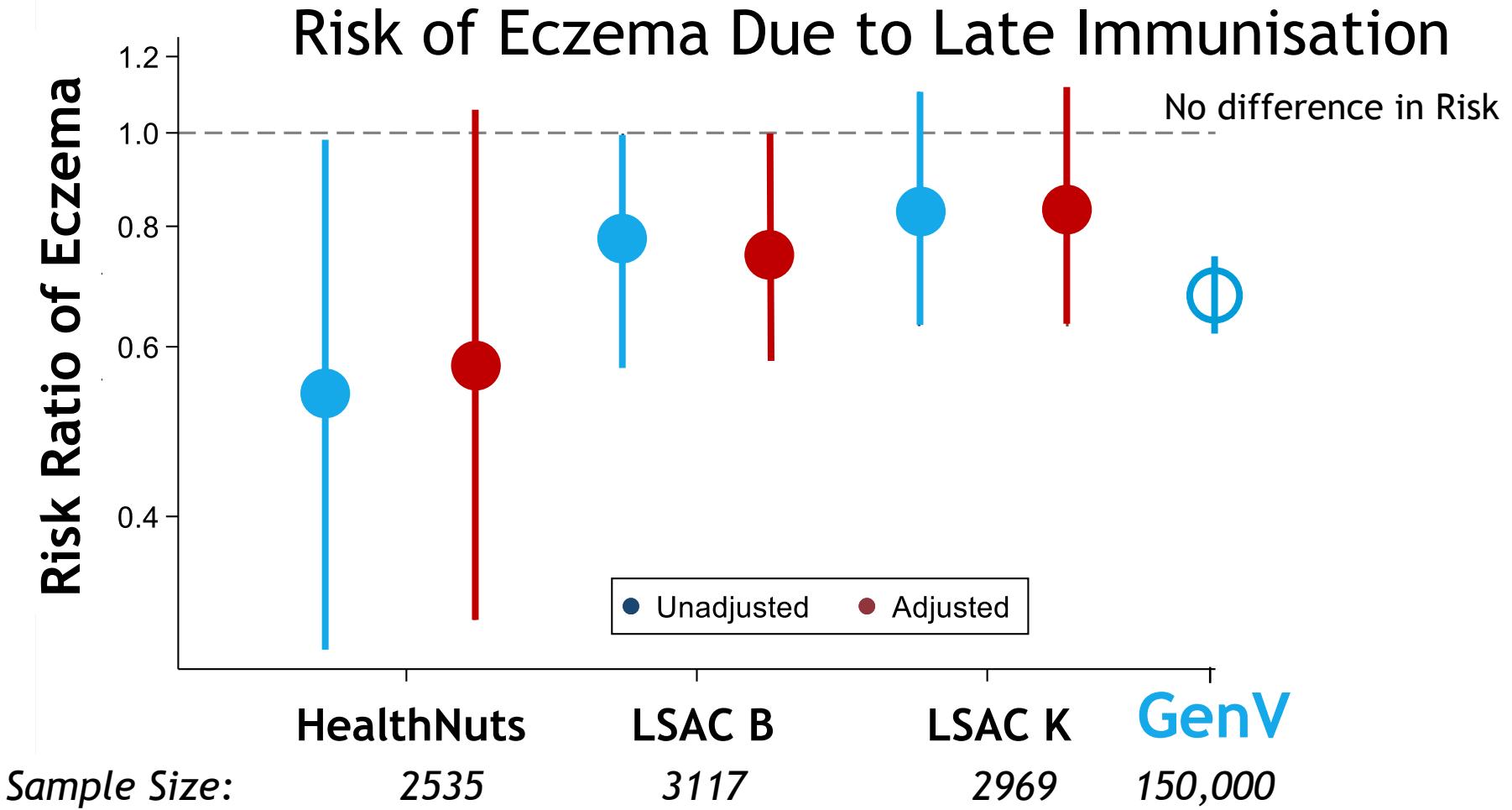
Turning your paper  
plots into  
presentation plots

Special  
Projects

# This is a paper plot

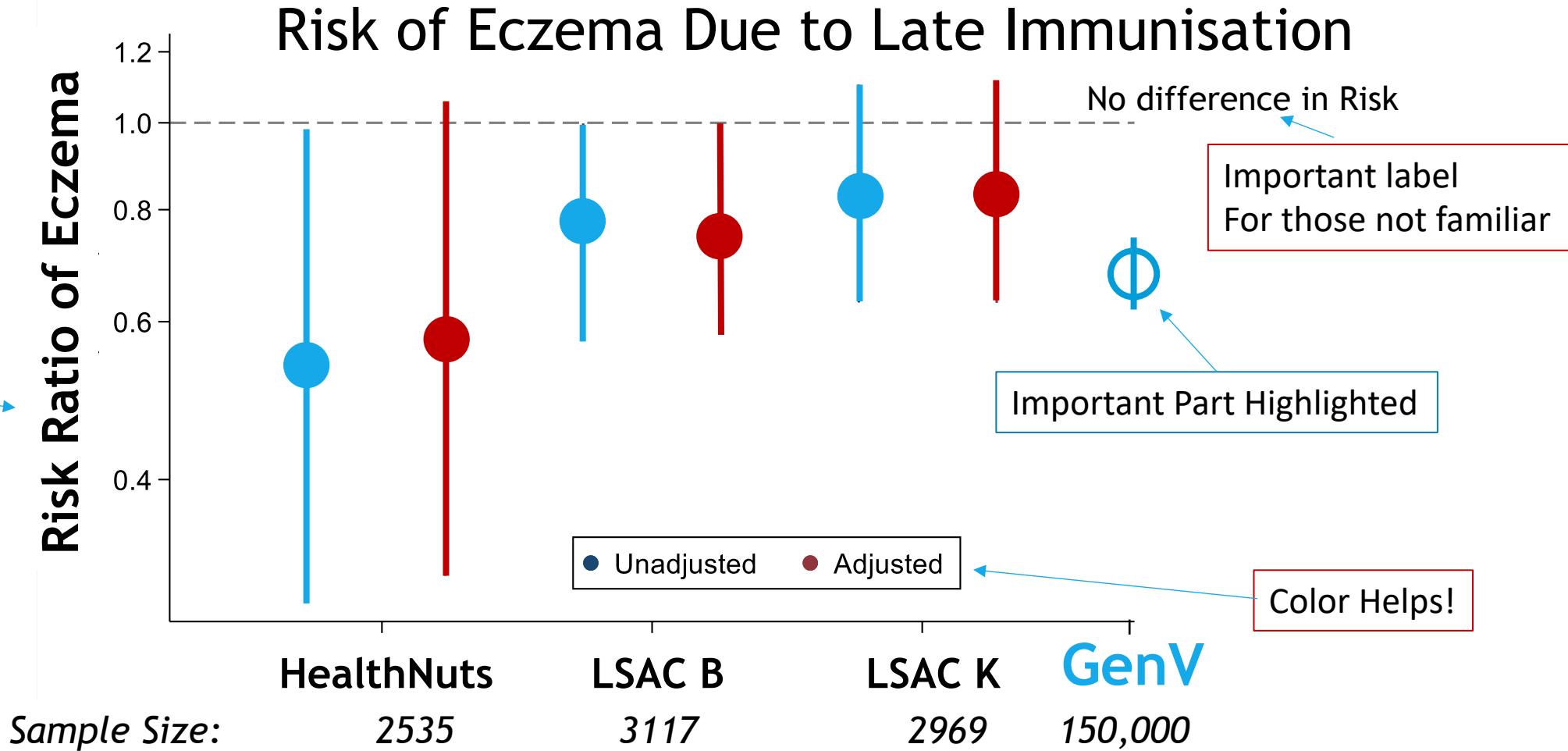


# This is a presentation plot

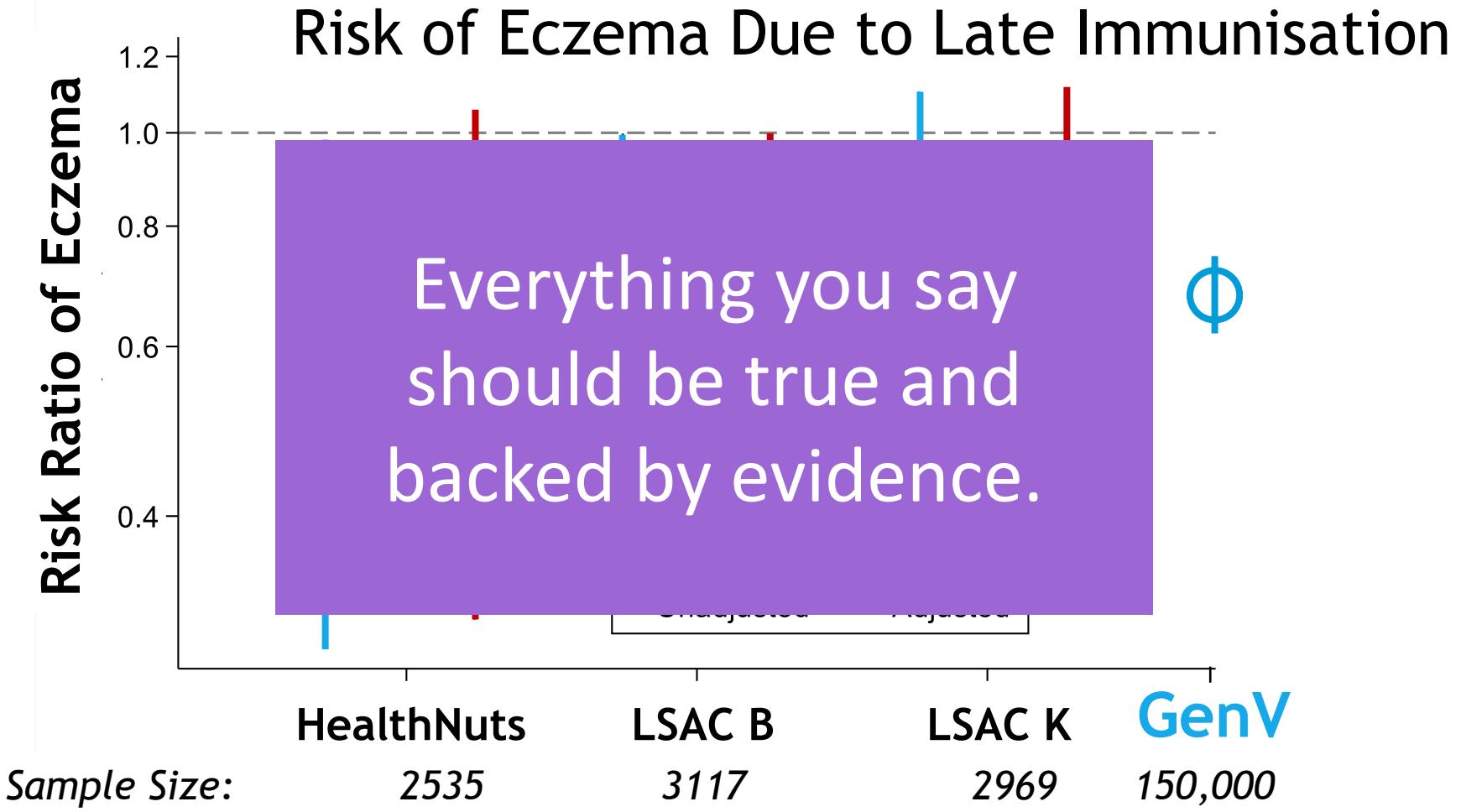


# This is a presentation plot

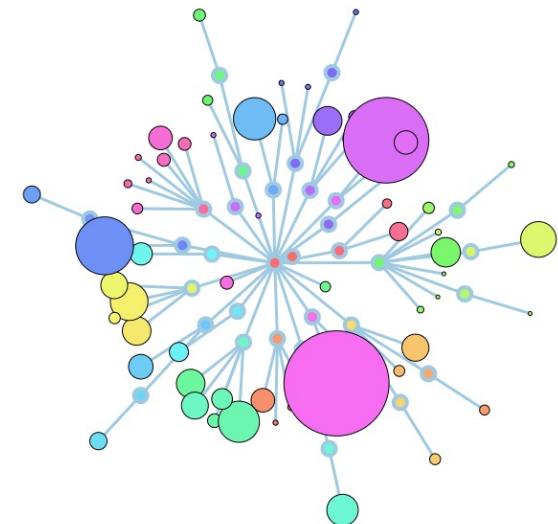
Big Title For Sleeping Person to Catch up



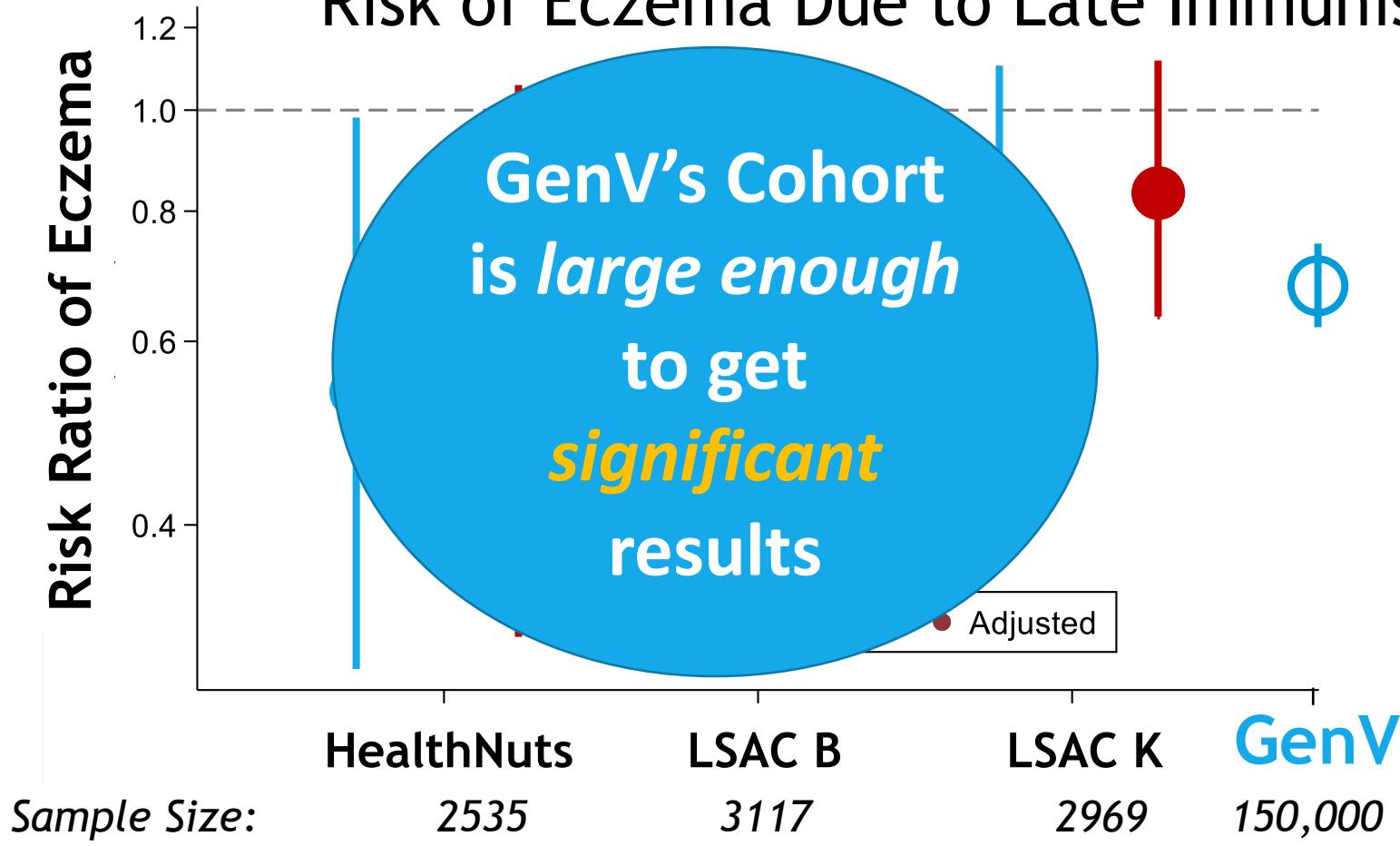
# This is a presentation plot



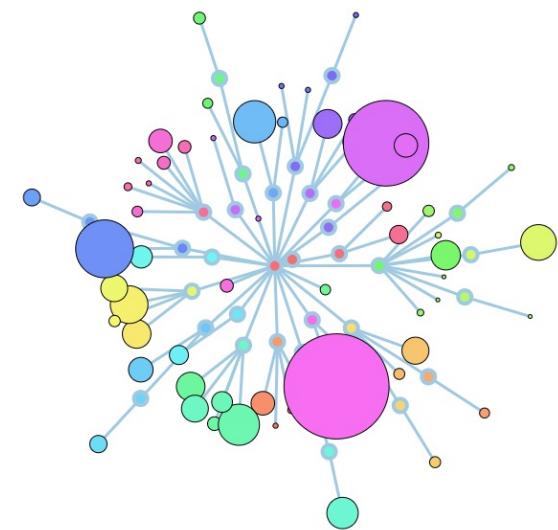
Have a **simple** message you  
**repeat** often



## Risk of Eczema Due to Late Immunisation



An aside on **powerpoint**



# I Obviously Prefer Powerpoint

- It's on every computer and does not require an internet connection
- Powerpoint wizardry is appreciated everywhere else besides academia
- Other options like keynote have less functionality and will not transfer well between computers
- Keep it simple and clean- avoid themes and weird fonts.
- No fancy transitions
- Find the compress pictures option and use liberally

A few  
tips

For general talks,  
pitch your  
presentation to a  
first year grad  
student

Special  
Projects

A few  
tips

Use Words Sparingly

Special  
Projects

# If words are here

- They are not listening to you
  - This is a VERY important paragraph about how if you put everything on your slide, people won't pay attention to you and instead read all the words in this very long run on sentence in too small font and maybe I should've used Helvetica?
- Definitely should've used Helvetica
- Is Helvetica even available on macs?
- I should watch the Helvetica documentary again
- That was wild.
- Update: Helvetica IS available on macs

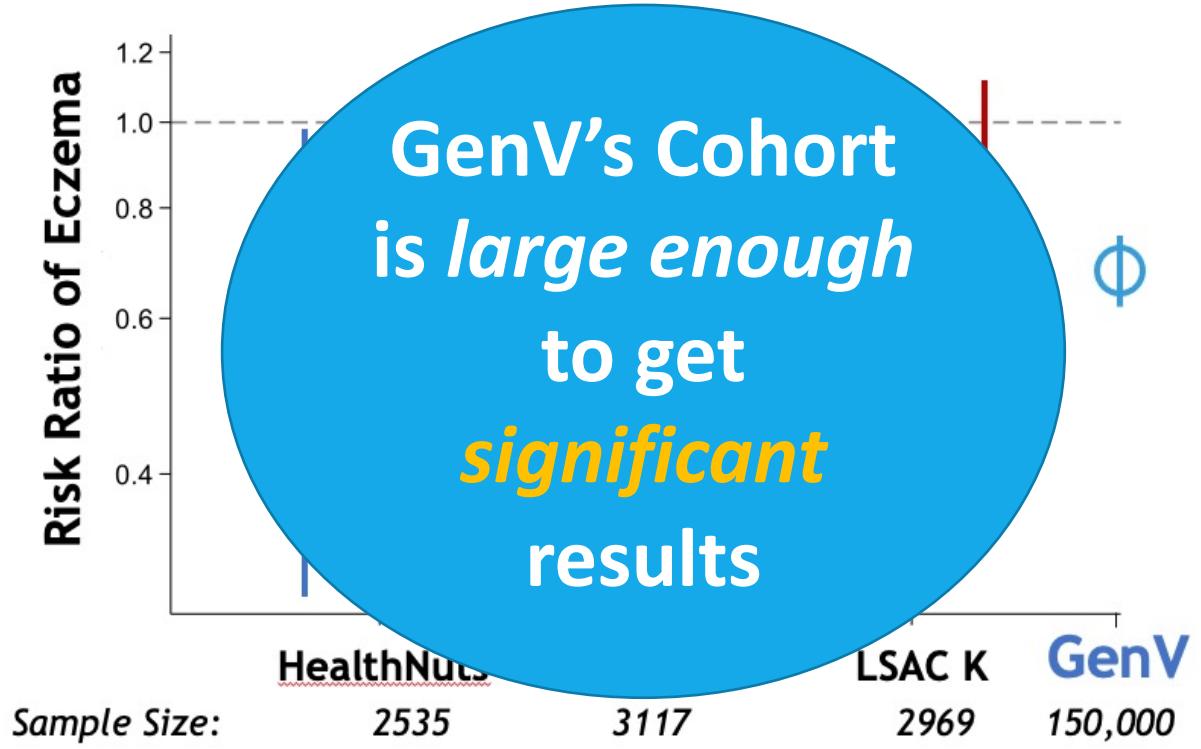
A few  
tips

Include reminder or  
catch-up slides

Special  
Projects

# Did you fall asleep?

- Repeat your simple message often
  - Say everything you need and nothing you don't
  - Everything you say should be true and backed by evidence
  - Everything you say should be appropriate for your audience



A few  
tips

Do not put tables in  
your talk

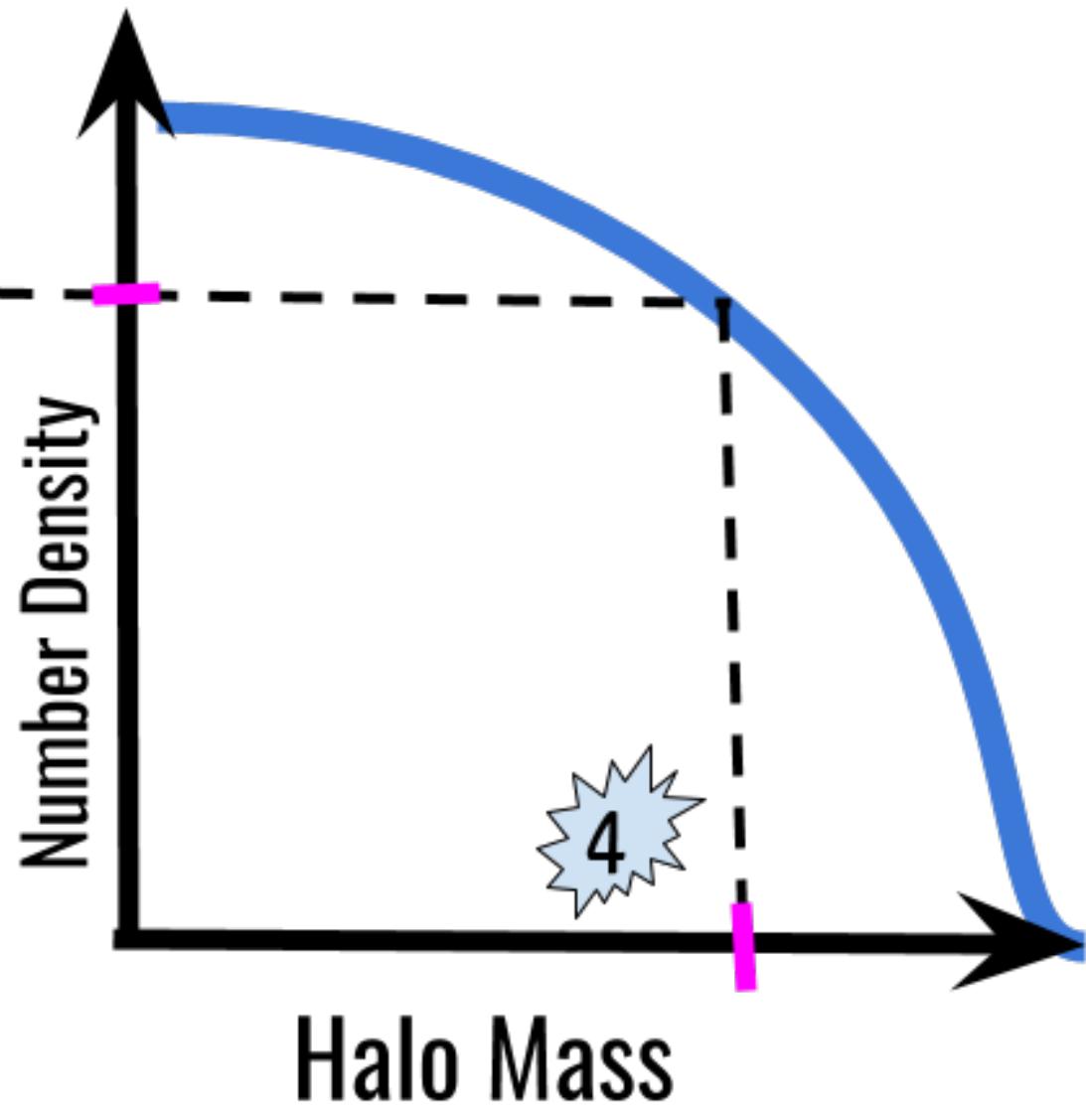
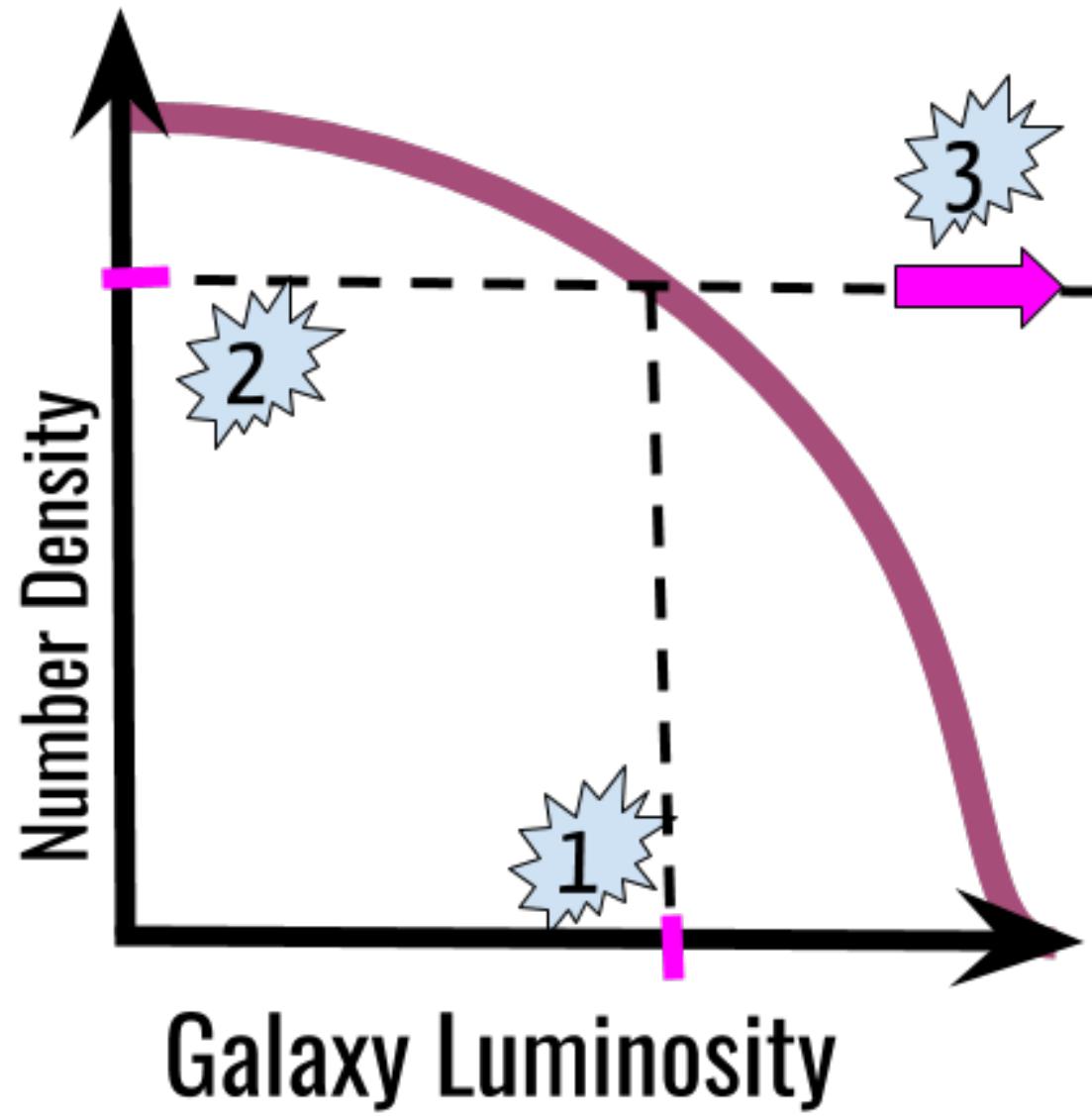
Special  
Projects

Please	Do not	Put	Tables in	Your	Presentations
$0.2 < z < 0.5$	$(0.40_{-0.05}^{+0.05}, 0.22_{-0.06}^{+0.06})$	1.5	$(1.09_{-0.10}^{+0.28}, 10.33_{-0.27}^{+0.86}, -0.77_{-0.31}^{+0.22})$	0.7	7.2 0.2
$0.5 < z < 0.8$	$(0.46_{-0.06}^{+0.07}, 0.45_{-0.08}^{+0.09})$	8.0	$(1.42_{-0.06}^{+0.13}, 10.3_{-0.17}^{+0.39}, -0.83_{-0.17}^{+0.18})$	0.57	22.8 16.9
$0.8 < z < 1.1$	$(0.46_{-0.3}^{+0.08}, 0.68_{-0.13}^{+0.45})$	4.1	$(1.83_{-0.20}^{+7.1}, 10.7_{-0.44}^{+33}, -0.69_{-0.35}^{+0.54})$	25.0	14.9 21.9
$1.1 < z < 1.5$	$(0.63_{-0.06}^{+0.05}, 0.59_{-0.06}^{+0.09})$	5.8	$(1.93_{-0.19}^{+0.18}, 10.62_{-0.3}^{+0.56}, -0.93_{-0.16}^{+0.23})$	1.3	9.1 2.1
$1.5 < z < 2.0$	$(0.58_{-0.06}^{+0.05}, 0.86_{-0.07}^{+0.11})$	2.8	$(2.22_{-0.27}^{+6.3}, 10.95_{-0.5}^{+13.7}, -0.82_{-0.35}^{+0.30})$	10.77	-8.4 -36.3

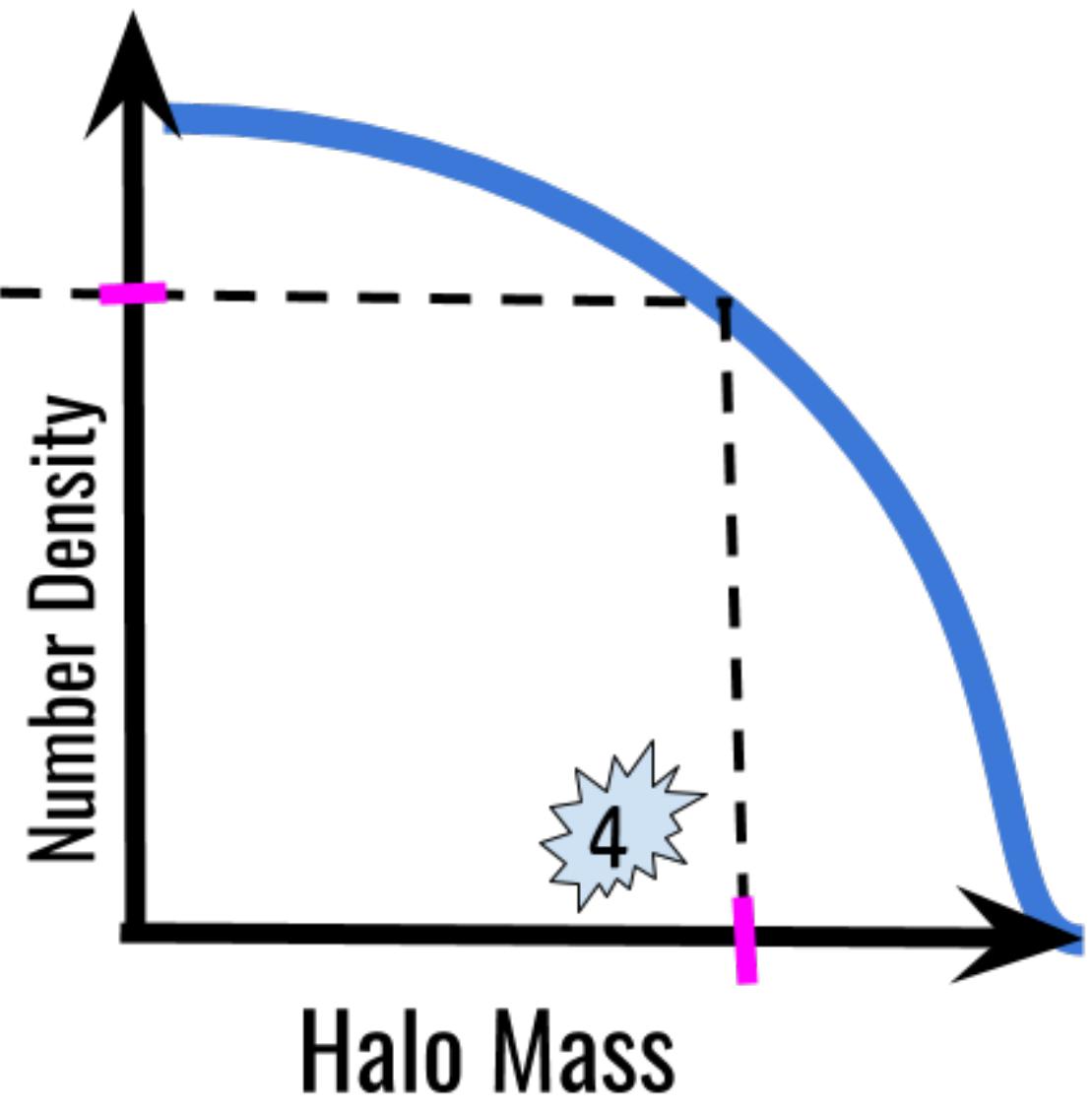
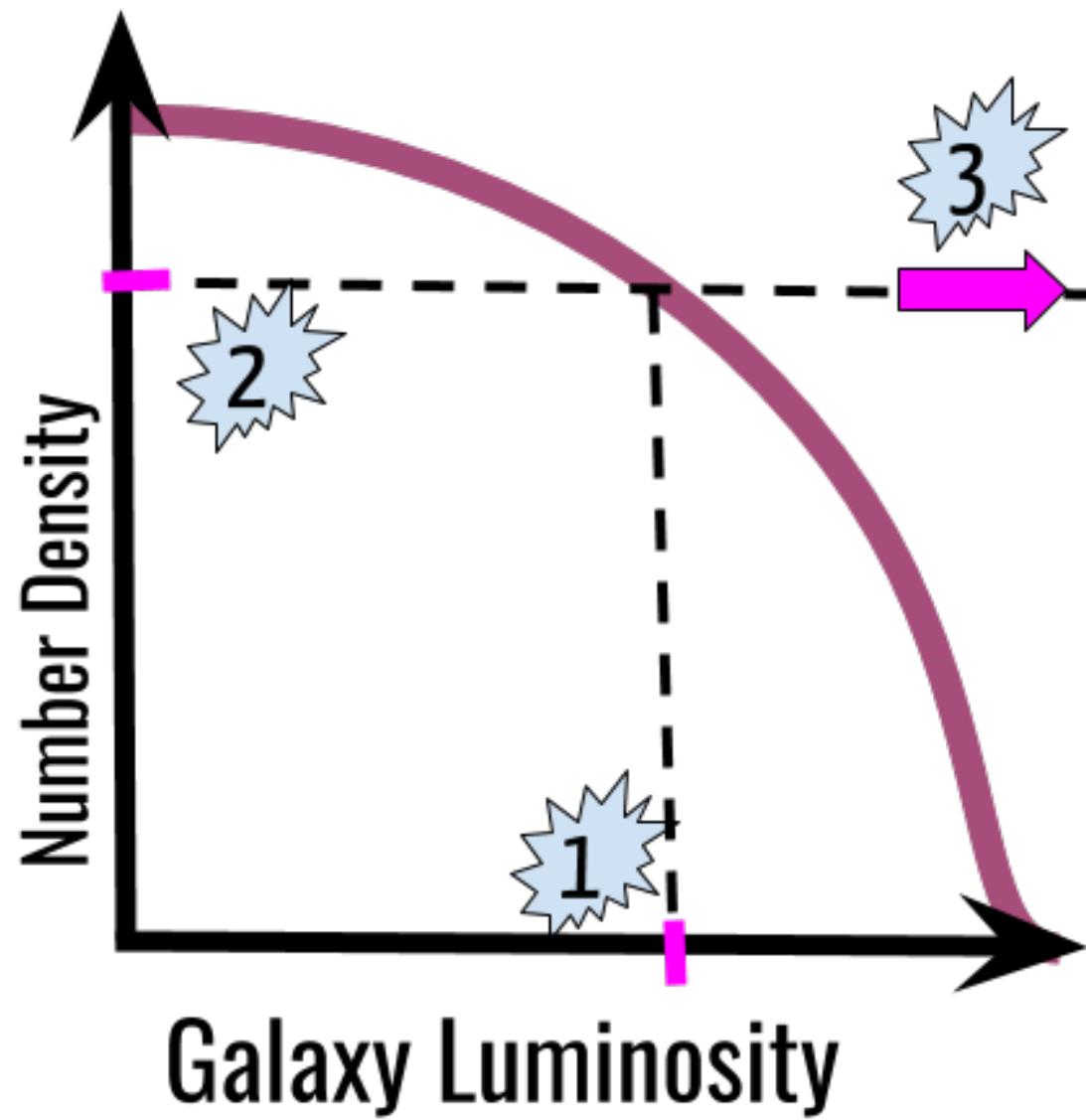
A few  
tips

Use Cartoons

Special  
Projects



\*Made With Google Drawings



A few  
tips

Avoid Green

Special  
Projects

A few  
tips

Don't mess with  
fancy fonts

Special  
Projects

Final tip

Do not make a .ppt  
over 50mb

Special  
Projects

Papers

Presentations

Special  
Projects

Papers

Presentations

Posters

Papers

Presentations

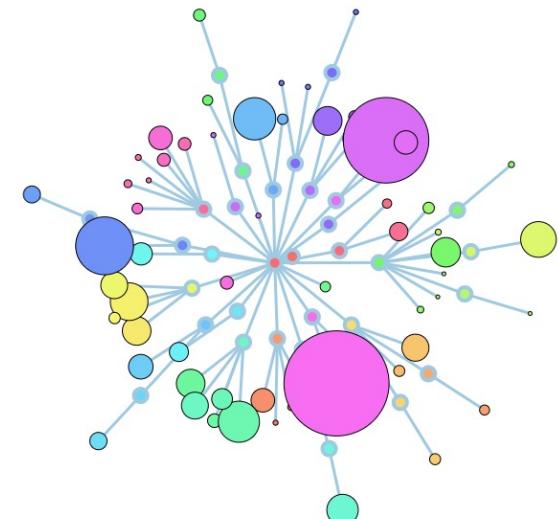
Websites

Papers

Presentations

Interactive  
objects

A ‘special project’ is how I made  
the transition from **astronomy** to  
**data science**





## Artist Tags



## User Searches

Papers

Presentations

Posters

# General Tips

- First time- use a .ppt template
- Simple Background
- Edit on a simple printed page
- Don't print out full glossy – too heavy
  - Fabric is even better!
- Have printed handouts



## The Very Small Scale Clustering of SDSS-II and SDSS-III Galaxies

Jennifer A. Piscionere, Andreas A. Berlind, Vanderbilt

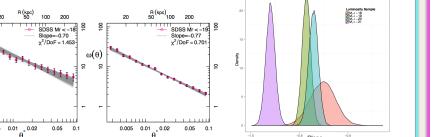


### Abstract

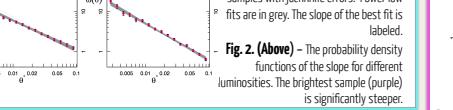
The very small scale clustering of galaxies can tell us about their spatial distribution within dark matter halos. To study the local universe, we measure the very small-scale angular clustering of galaxies in volume-limited luminosity samples drawn from the SDSS DR7. These angular scales correspond to 20 to 500 kpc at the median redshift of the  $M_r < -20$  galaxy sample. We model this clustering using mock galaxy catalogues produced from the LasDamas simulations and the Halo Occupation Distribution (HOD) framework, assuming a flexible density profile of satellite galaxies within halos. We find that luminous galaxies have a steeper correlation function, and are thus more centrally concentrated in halos than the underlying dark matter. Lower luminosity galaxies, however, have a density profile that is consistent with that of dark matter. In order to see if this trend continues to higher redshift, we also measure the projected correlation function of SDSS-III BOSS CMASS galaxies on similar scales.

### Measuring $\omega(\theta)$ on Sloan Digital Sky Survey Galaxies

We measure the angular correlation in four volume-limited galaxy samples with absolute  $r$ -band magnitude thresholds of  $M_r = -18, -19, -20, -21$ , from the SDSS. Using the angular correlation function frees us from the SDSS fiber collision limit, a technical aspect of the survey that causes incompleteness in the galaxy sample for angular separations less than  $55''$ . A power law fit was done for  $0.002'' < \theta < 0.1$  using jackknife errors.



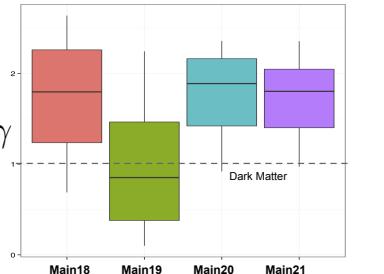
**Fig. 1. (Left)** - The angular correlation function for the 4 volume limited SDSS samples with jackknife errors. Power law fits are in grey. The slope of the best fit is labeled.



**Fig. 2. (Above)** - The probability density functions of the slope for different luminosities. The brightest sample (purple) is significantly steeper.

### Fitting For The Halo-Galaxy Connection

The very small clustering of galaxies depends strongly on the number of satellites in a single halo. To quantify the relationship between host halos and satellite galaxies, we produce mock catalogues of galaxies using the Halo Occupation Distribution (HOD; Berlind & Weinberg 2002). Using the Zheng et al. 2007 parameterization, we vary the probability that halo of a given mass will host a satellite galaxy, as well as the spatial distribution of these galaxies within the halo. By comparing the parameter estimation of different galaxy classes, we can investigate how the spatial distribution of galaxies depends on luminosity.



**Fig. 5 (Above)**: How the GFW profile density parameter gamma depends on luminosity. A higher value of  $\gamma$  corresponds to a steeper density profile. The two brighter samples differ from NFW with more than 99.7% confidence. The extent of the whiskers is the 99.7% CI.

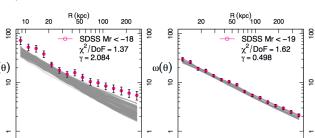
**Fig. 6 (Left)**: The joint probability distribution of HOD parameters for four luminosity samples. Each parameter controls how many satellite galaxies are added to a halo.

### The Spatial Distribution of Galaxies:

How well do galaxies trace the dark matter distribution?

**Not well.** Figure 3 (below) shows the angular correlation function for the SDSS  $M_r < -20$  sample compared to mock galaxy catalogues using a standard Navarro, Frenk and White density profile. The grey lines are  $\omega(\theta)$  for each NFW mock. The upper axis describes the physical separation of the galaxies as if they were all located at the median redshift. The SDSS fiber collision limit of  $55''$  is shown in the magenta line.

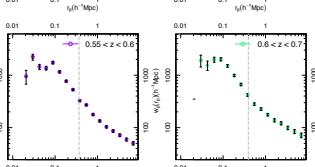
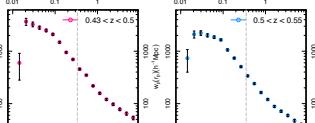
**What Next?** From this, we are motivated to introduce a Generalized NFW (GNFW) profile, with two additional parameters. The  $c_{gal}$  parameter (Eq. 1) allows galaxies to have a different concentration than the dark matter. The parameter  $\gamma$  (Eq. 2) allows the inner slope of the density profile to be adjusted. A value of  $\gamma = 1$  recovers the NFW profile of dark matter.



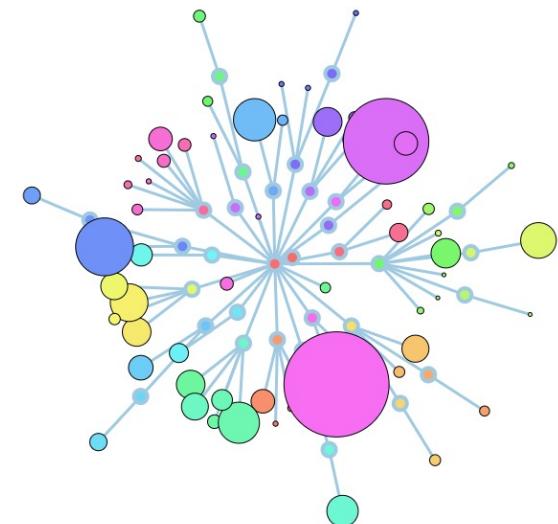
**Fig 4 (Right):** By using a GNFW profile, we were successful at recovering the clustering of galaxies on small scales. Each panel shows model fits allowing both the GNFW to vary as well as other HOD parameters. The  $\chi^2/\text{DoF}$  values are noted. The two brighter samples,  $M_r < -20$  and  $M_r < -21$ , strongly prefer a density profile steeper than NFW. The other two samples are consistent with NFW. See Fig. 5 for more detail. Each line represents a different model from the MCMC chain.

### Measuring Clustering on BOSS CMASS Galaxies: Is there Evolution in the Very Small Scale Clustering?

Next, we investigate whether there is an evolution in the density profile of one class of galaxy. We measure the projected correlation function,  $w_p(r_p)$  for four different redshift bins. Figure 7 hints at a redshift evolution in CMASS galaxy clustering. We see a steepening in  $w_p(r_p)$  towards lower redshift bins. It is possible that satellites which have been accreted at the highest z bin have not undergone enough dynamical times to be destroyed by the lowest bin.



Have **fun** and **care** about your  
(whole) audience



# Resources

- ‘Show me the numbers’ by Stephen Few
  - [https://nces.ed.gov/programs/slds/pdf/08\\_F\\_06.pdf](https://nces.ed.gov/programs/slds/pdf/08_F_06.pdf)
- Data Visualization by Jill P. Naiman
  - <https://uiuc-ischool-dataviz.github.io/spring2019online/>
- My slides on how to rip off a cool visualisation and put it on your website
  - <http://jpiscionere.github.io/d3.pdf>