

Everything your advisor

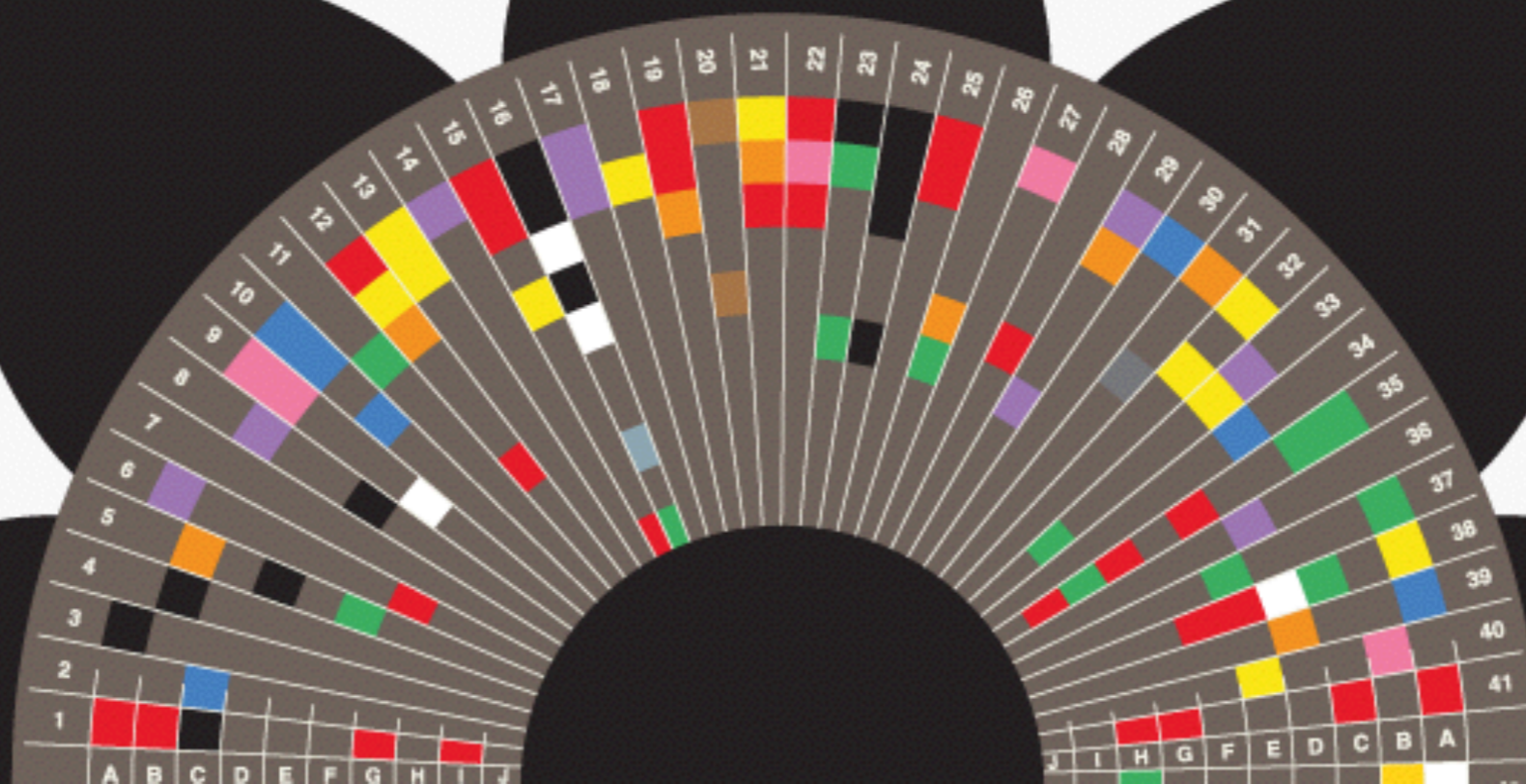
**didn't** teach you about giving talks

## Visuals II

Renée Hlozek

[hlozek@dunlap.utoronto.ca](mailto:hlozek@dunlap.utoronto.ca)

AB 222



Data

Relationships

Dimensions

Structure

Task:

In pairs, scrape a data set off the internet that you think has a great deal of complexity (decide together why you chose the data).

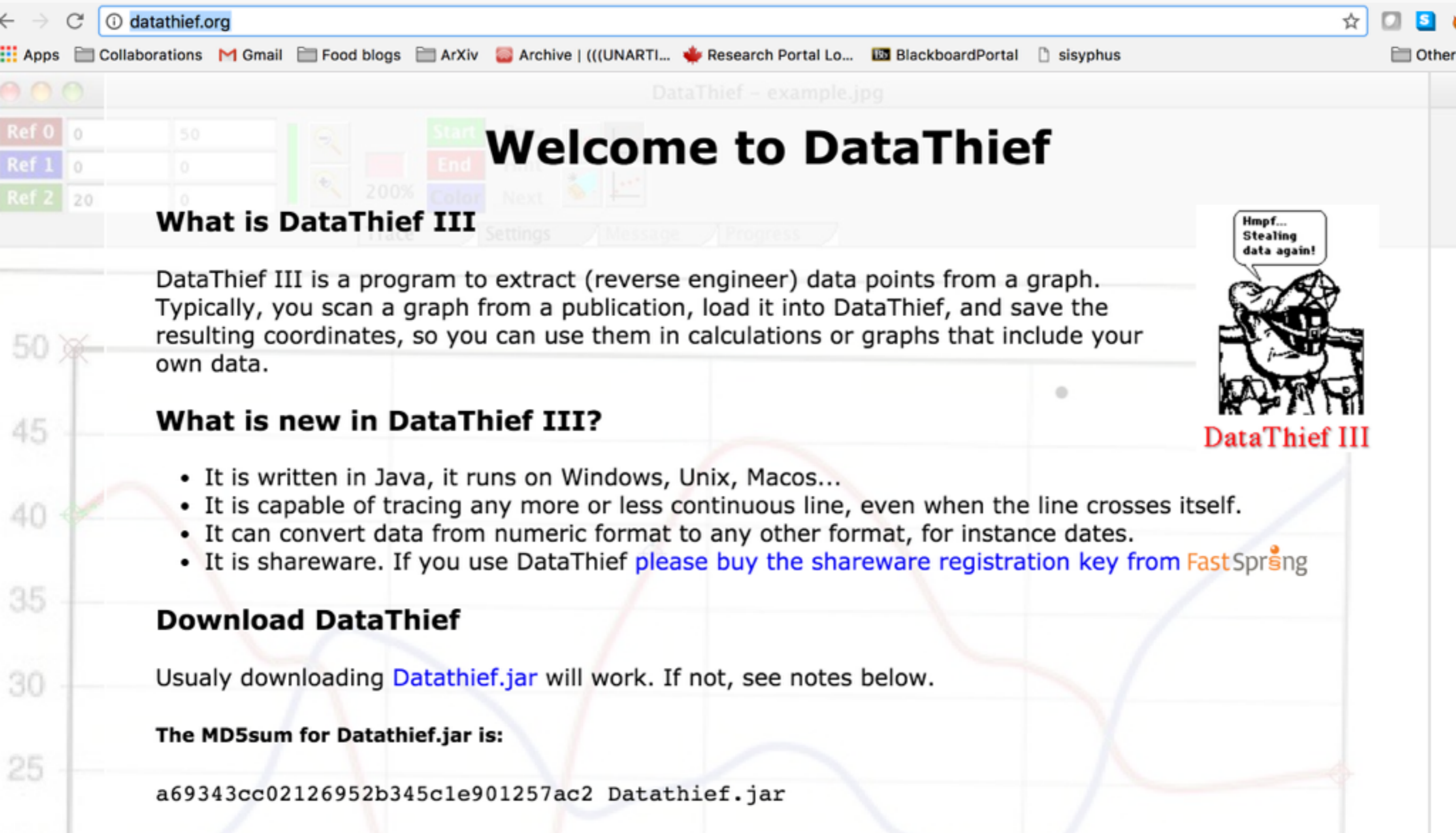
Split off from each other and make independent plots of the dataset

# Tufte's design principles

## **1) Representation (aka tell the truth):**

“The representation of numbers, as physically measured on the *surface of the graphic itself*, should be directly proportional to the numerical quantities measured.”

# An aside:



The screenshot shows a web browser window with the URL [datathief.org](http://datathief.org). The browser's address bar and tabs are visible at the top. The main content area features a large heading "Welcome to DataThief" and a sub-heading "What is DataThief III". Below this is a paragraph explaining the program's purpose. A section titled "What is new in DataThief III?" contains a bulleted list of features. Further down, there is a "Download DataThief" section with instructions on how to download the software and its MD5sum. On the right side, there is a cartoon illustration of a thief with a speech bubble that says "Hmpf... Stealing data again!" and the text "DataThief III" below it. The background of the page is a large, faint graph with two lines, one red and one blue, and a grid. The graph has a y-axis ranging from 25 to 50. There are also some UI elements like "Start", "End", "Color", "Next", "Settings", "Message", and "Progress" buttons visible in the background.

## Welcome to DataThief

### What is DataThief III

DataThief III is a program to extract (reverse engineer) data points from a graph. Typically, you scan a graph from a publication, load it into DataThief, and save the resulting coordinates, so you can use them in calculations or graphs that include your own data.

### What is new in DataThief III?


- It is written in Java, it runs on Windows, Unix, MacOS...
- It is capable of tracing any more or less continuous line, even when the line crosses itself.
- It can convert data from numeric format to any other format, for instance dates.
- It is shareware. If you use DataThief [please buy the shareware registration key from FastSpring](#)

### Download DataThief

Usually downloading [Datathief.jar](#) will work. If not, see notes below.

**The MD5sum for Datathief.jar is:**

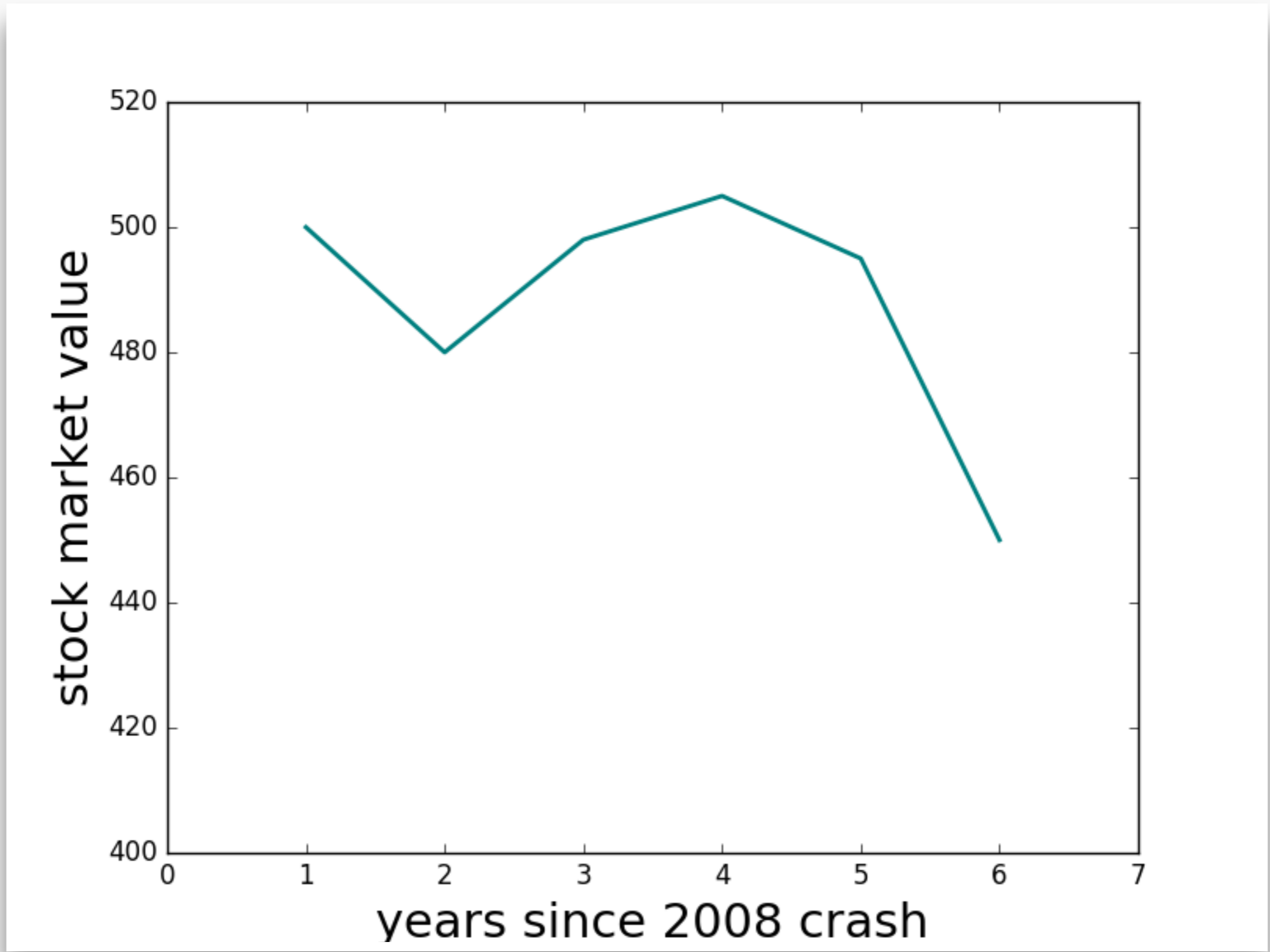
```
a69343cc02126952b345c1e901257ac2 Datathief.jar
```



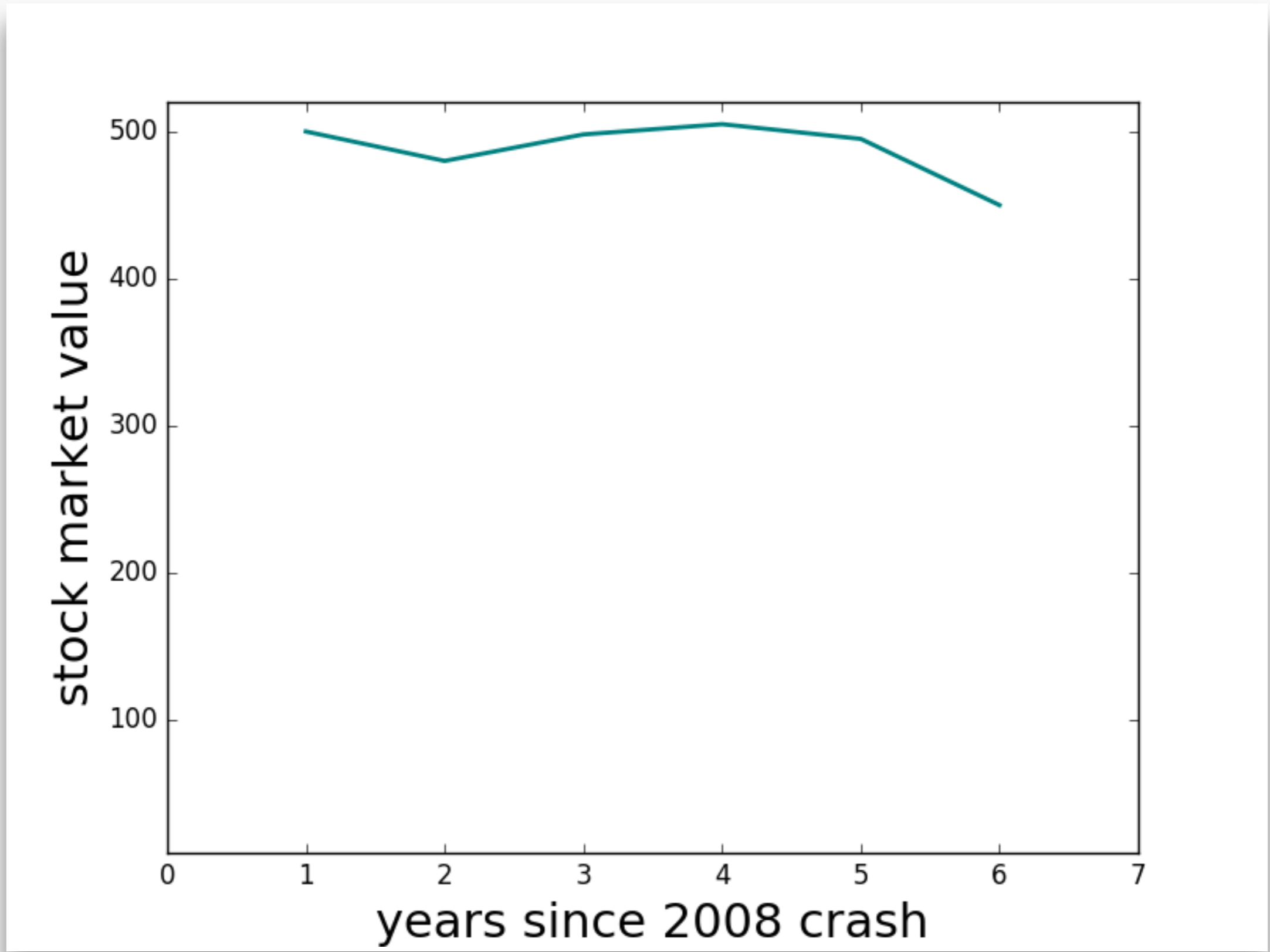
Hmpf... Stealing data again!

DataThief III

# Tufte's design principles



# Tufte's design principles



# Tufte's design principles

## **1) Representation (aka tell the truth):**

“The representation of numbers, as physically measured on the *surface of the graphic itself*, should be directly proportional to the numerical quantities measured.”

## **2) Unambiguous (label it right):**

“Clear, detailed, and thorough labels should be used to defeat graphical distortions. Labels should be out explanations of the data. Labels should be out important events.

Does my plot actually need this additional detail?

# Tufte's design principles

$$\text{The Lie factor} = \frac{\text{size of an effect as portrayed in the plot}}{\text{size of an effect in the real data}}$$

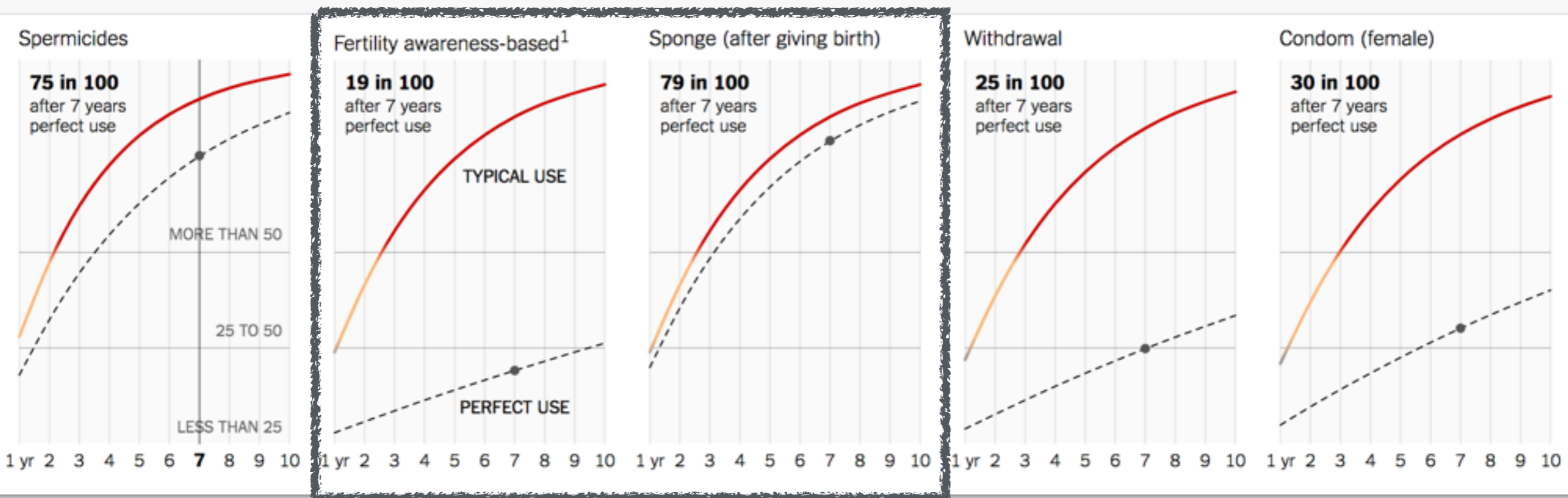
This is often  $\gg 1$ ; we magnify the effect to illustrate

# Tufte's design principles

**130mm-15mm difference**

**The Lie factor =**

**19-79 cases per hundred**

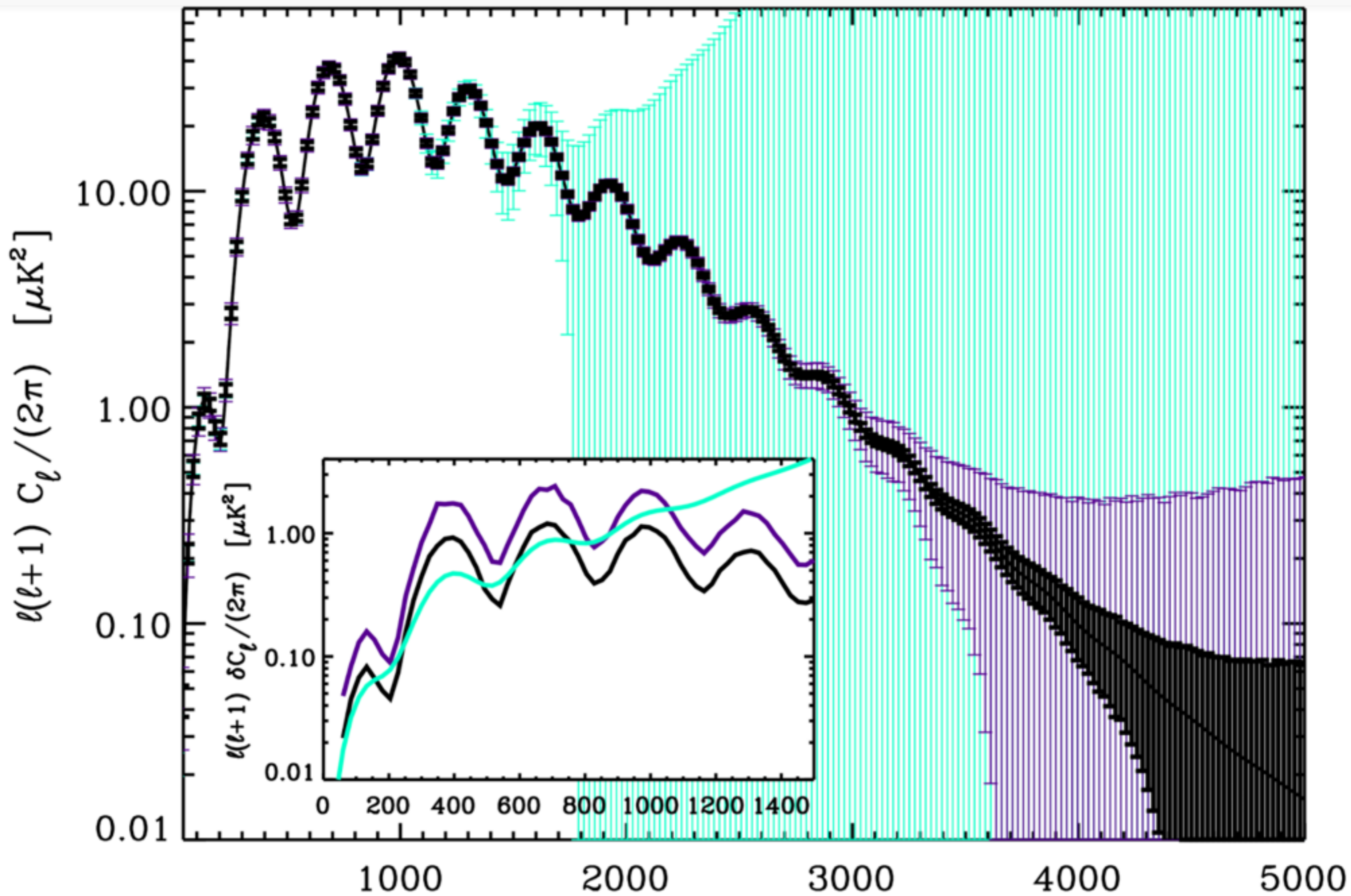


How Likely Is It That Birth Control Could Let You Down? - NYTime article

# Tufte's design principles

## **3) Data variation, not design variation**

Don't unnecessarily change the style of your plot unless the data needs it



# Tufte's design principles

## **3) Data variation, not design variation**

Don't unnecessarily change the style of your plot unless the data needs it

## **4) Your plots shouldn't 'age':**

When making plots that contain money etc. standardised units are better than absolute units. Less useful for science?  
[Also percentage plots?]

# Tufte's design principles

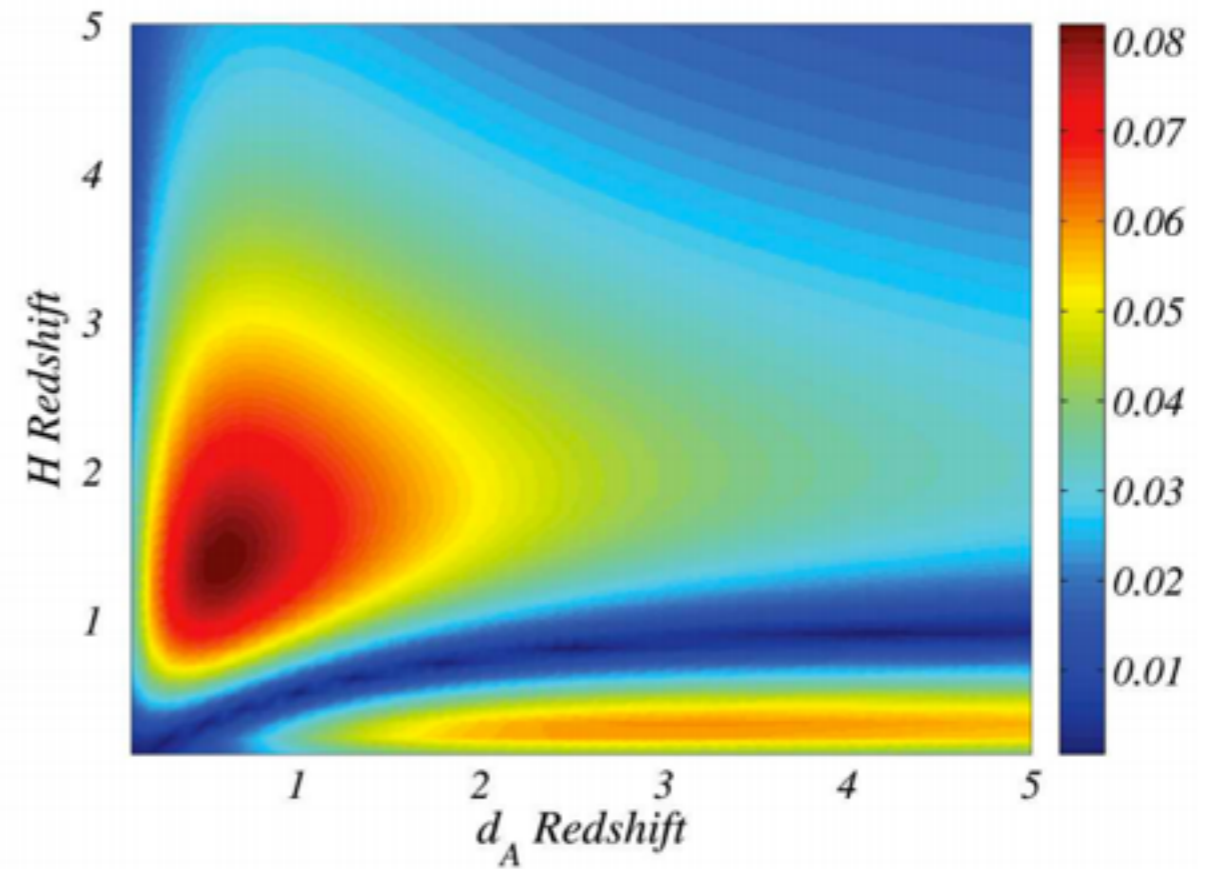
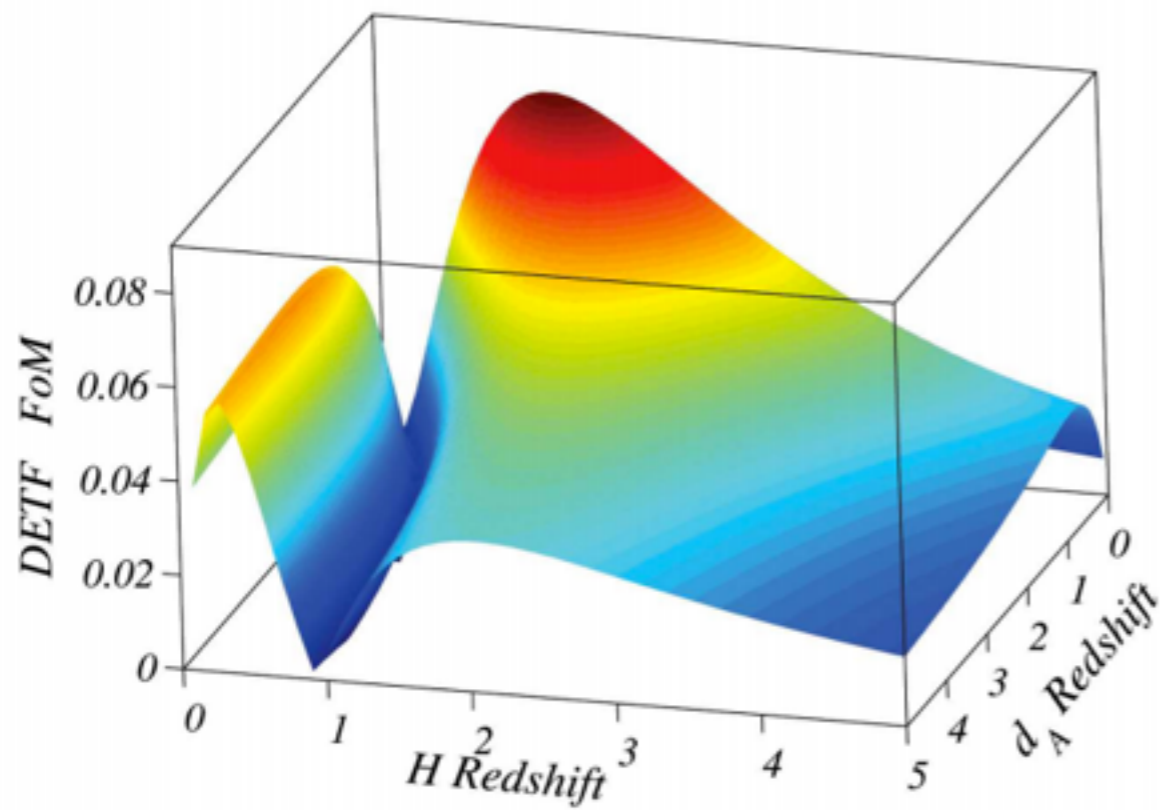
**5) Number of information-carrying dimensions should not exceed the number of data dimensions.**

Beware the 3D plot!

# Tufte's design principles

DEFT FoM=1  
H,  $d_A=2$

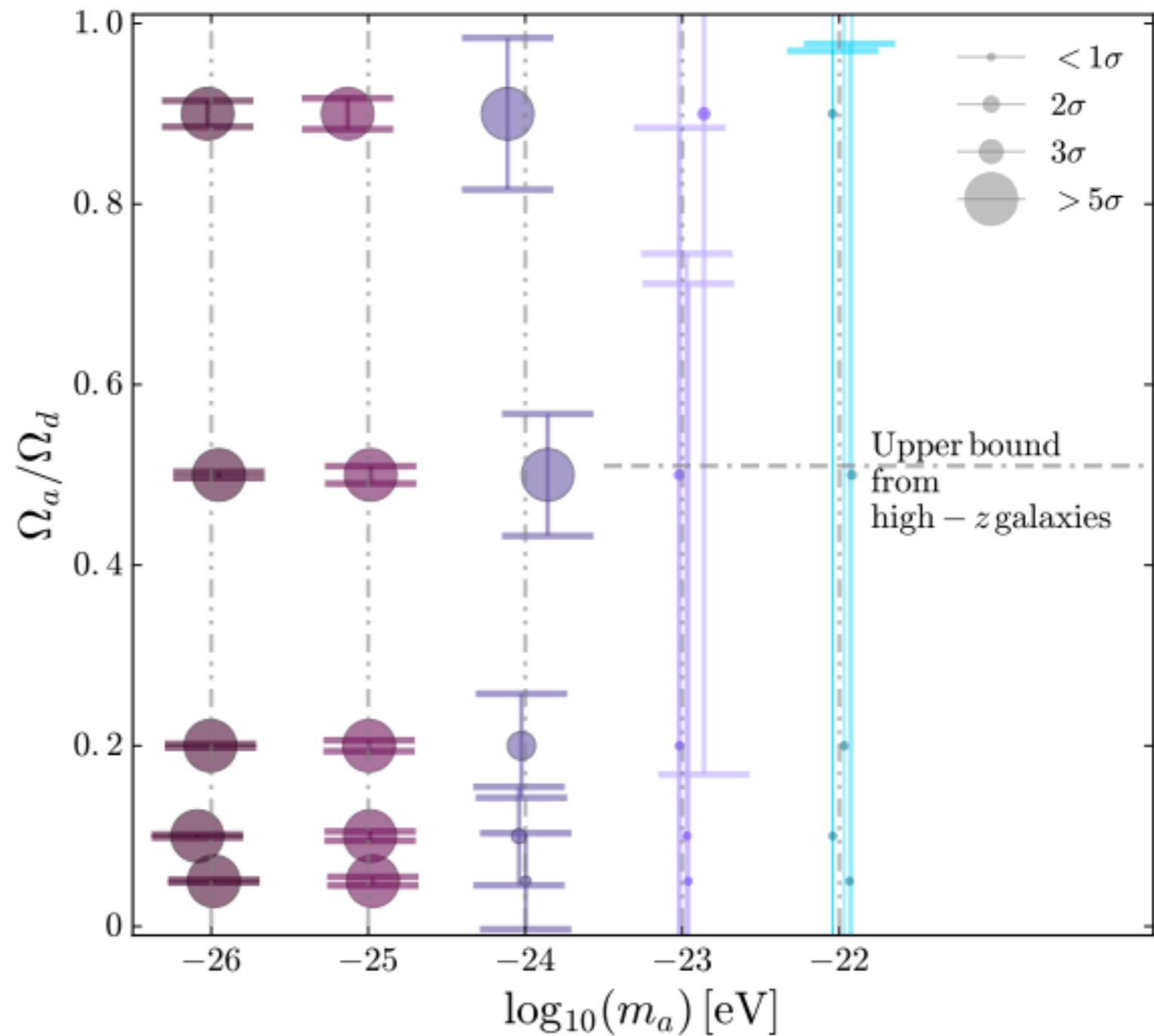
color=1  
x,y,z=3



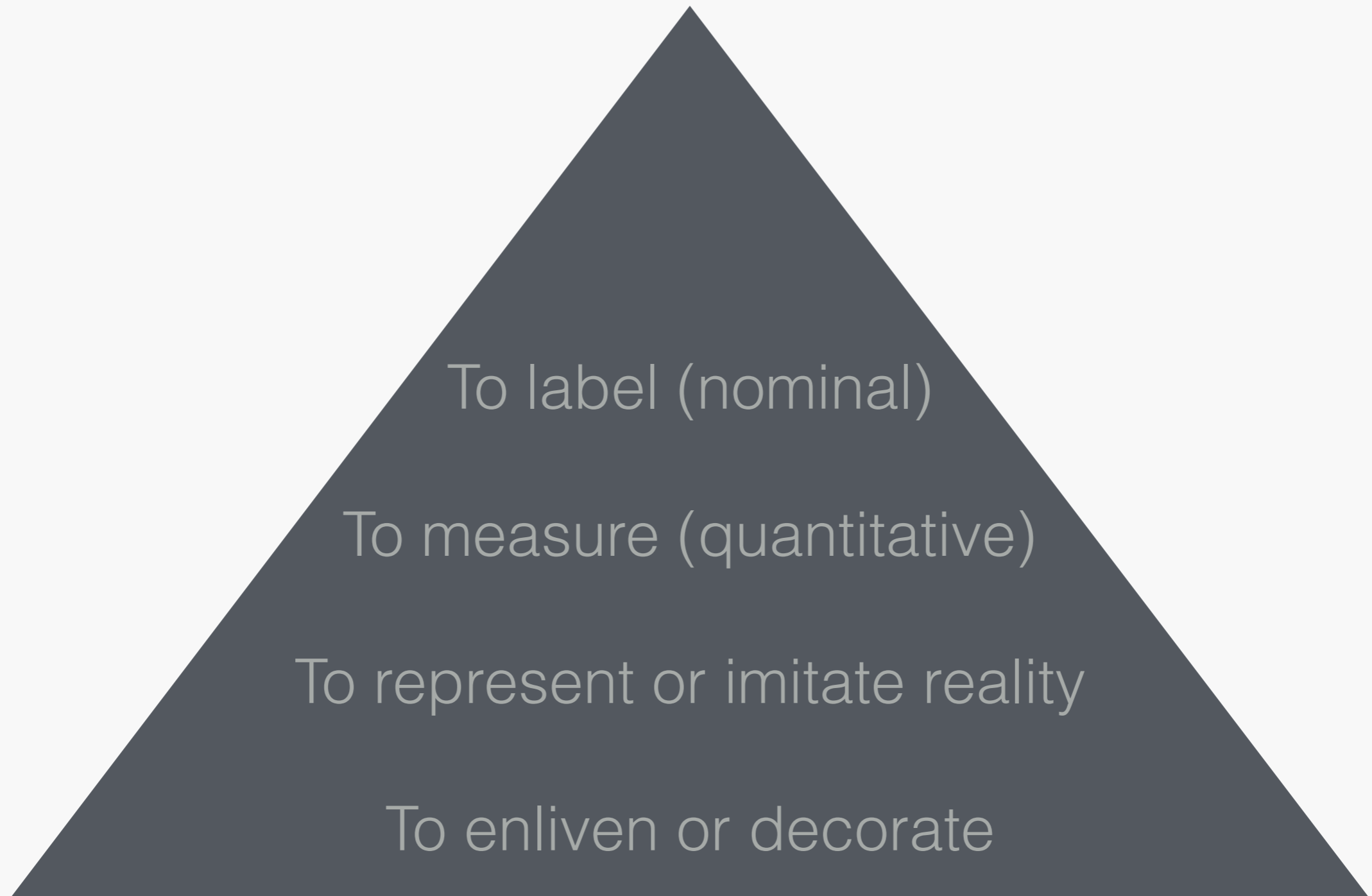
# Tufte's design principles

mass = 1  
axfrac=1  
errorbar=1  
significance=1  
previous bound = 1

area = 2  
color=1  
position = 1  
y,x=2  
central line=1  
high-z galaxy line=1



# Using colour correctly



# Using colour correctly

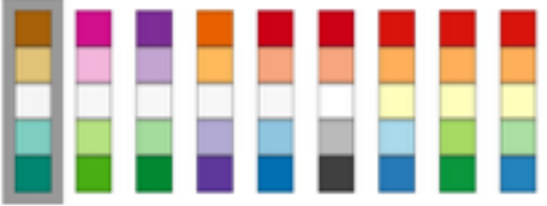
colorbrewer2.org/#type=diverging&scheme=BrBG&n=6

Collaborations Gmail Food blogs ArXiv Archive | ((UNARTI... Research Portal Lo... BlackboardPortal sisyphus

Number of data classes: 6

Nature of your data:  
 sequential  diverging  qualitative

Pick a color scheme:



Only show:  
 colorblind safe  
 print friendly  
 photocopy safe

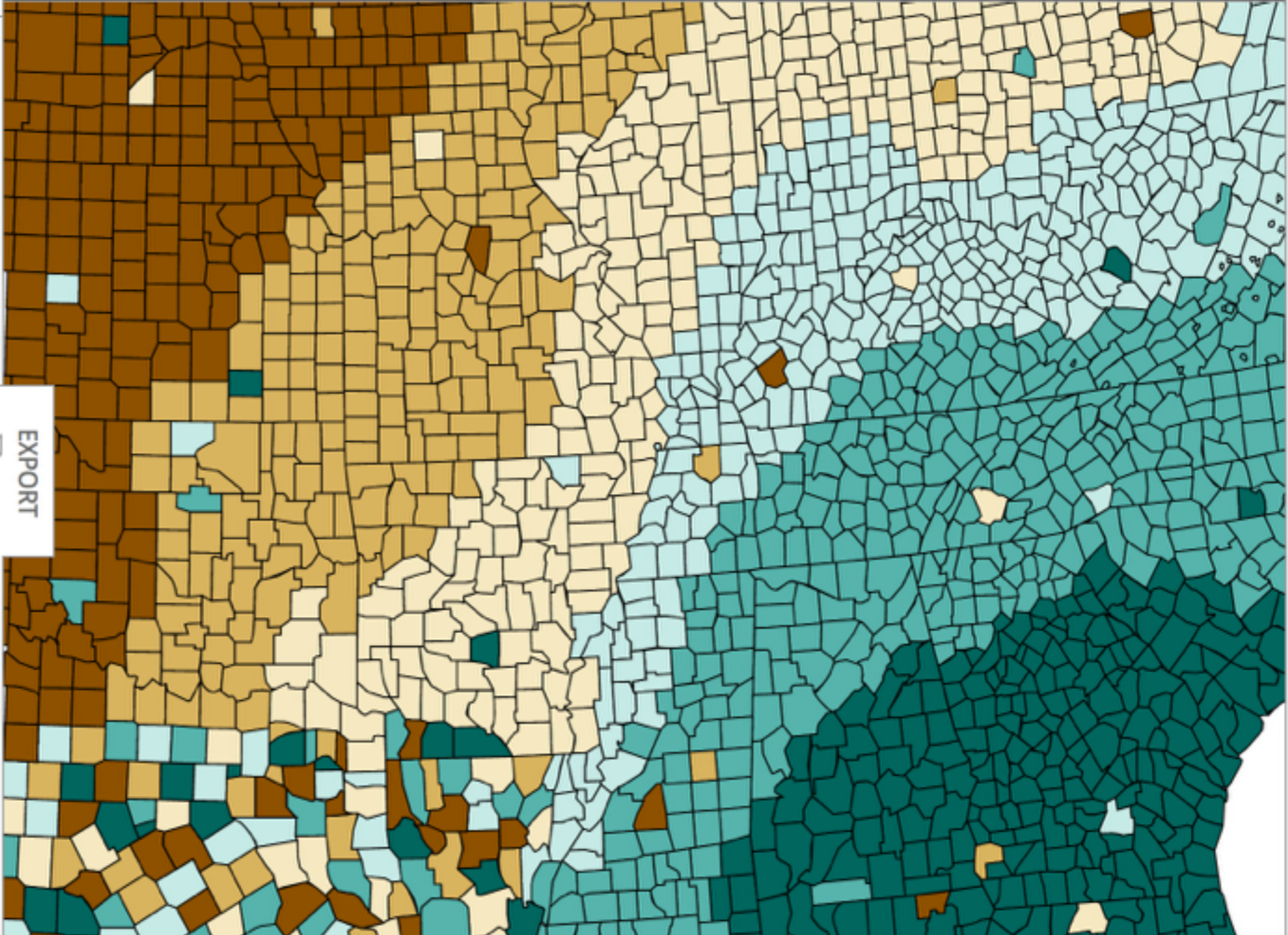
Context:  
 roads  
 cities  
 borders

Background:  
 solid color  
 terrain

color transparency

how to use | updates | downloads | credits

**COLORBREWER 2.0**  
color advice for cartography



6-class BrBG

EXPORT

HEX

- #8c510a
- #d8b365
- #f6e8c3
- #c7eae5
- #5ab4ac
- #01665e

# Using colour correctly

The image shows a browser window with the URL <https://coolors.co>. The browser's address bar and tabs are visible at the top. The website header includes the logo "COOLORS + SKILLSHARE 3 months free" and navigation links for "Generate", "Explore", "iOS App", "Add-on", "Chrome Extension", "More", and "Login".

## The super fast color schemes generator!

Create, save and share perfect palettes in seconds!

[Start the Generator, it's free!](#)

[iOS App](#) [Adobe Add-on](#)

[Chrome Extension](#) [New](#)

The main content area features a large grid of various color palettes. A prominent palette in the foreground consists of five vertical bars: red, cyan, yellow, light grey, and white. Another palette in the background shows a grid of horizontal bars with various color combinations.

# Tufte's design principles

$$\text{data ink ratio} = \frac{\text{amount of ink used on data}}{\text{total amount of ink used in figure}}$$

Also can be thought of as the the proportion of a graphic's ink devoted to the **non-redundant** display of data information - can you get away with less ink overall?

**Maximize the data ink ratio, within reason**

Do you need those background grid lines?

First erase non-data ink, then erase non-redundant data ink.

# Tufte's design principles

**5) Number of information-carrying dimensions should not exceed the number of data dimensions.**

Beware the 3D plot!

**6) Graphics should quote data out of context:**

Labels matter! Plots should survive their captions.

# Tufte's design principles

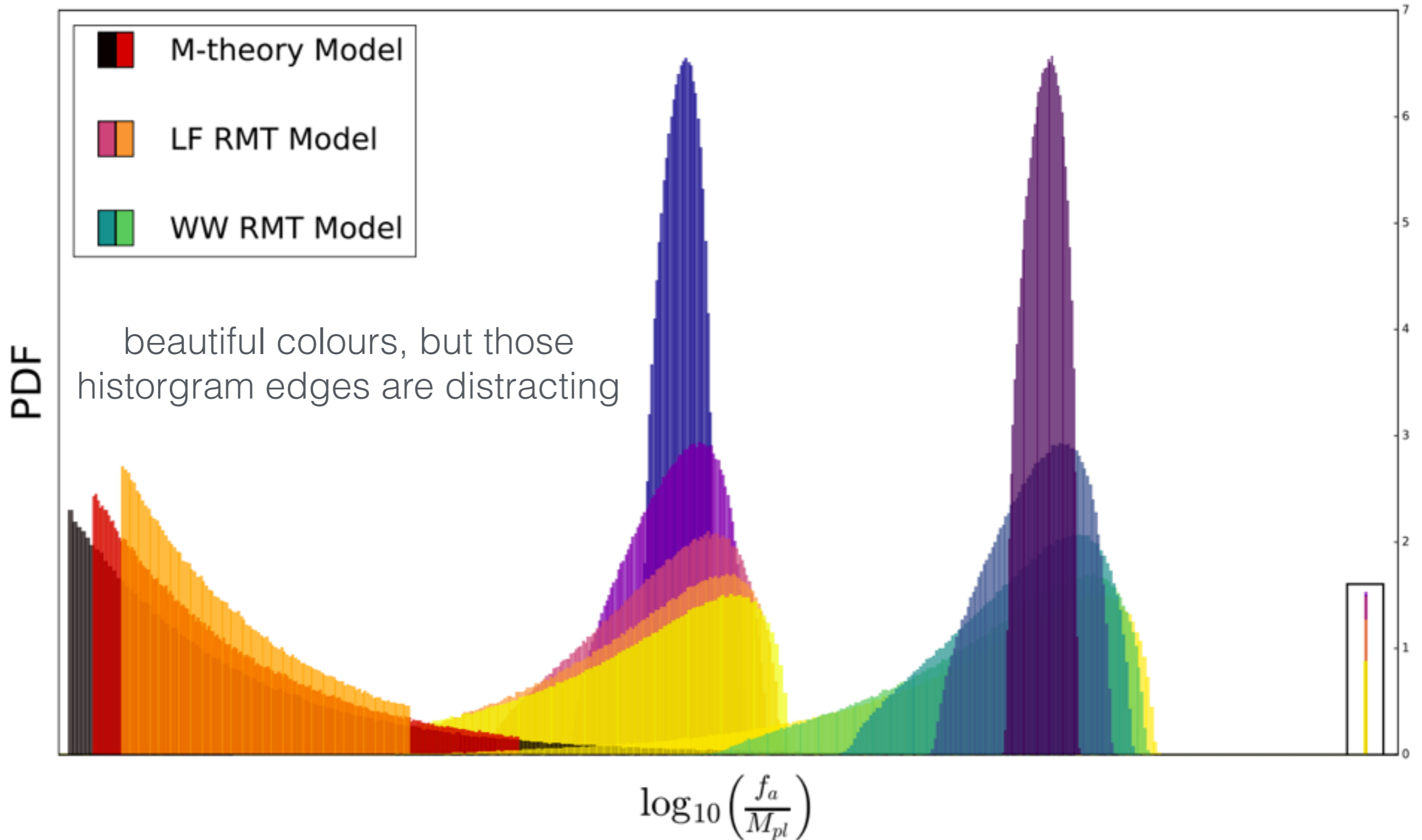
**Chartjunk:** “visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information”

# Types of chartjunk

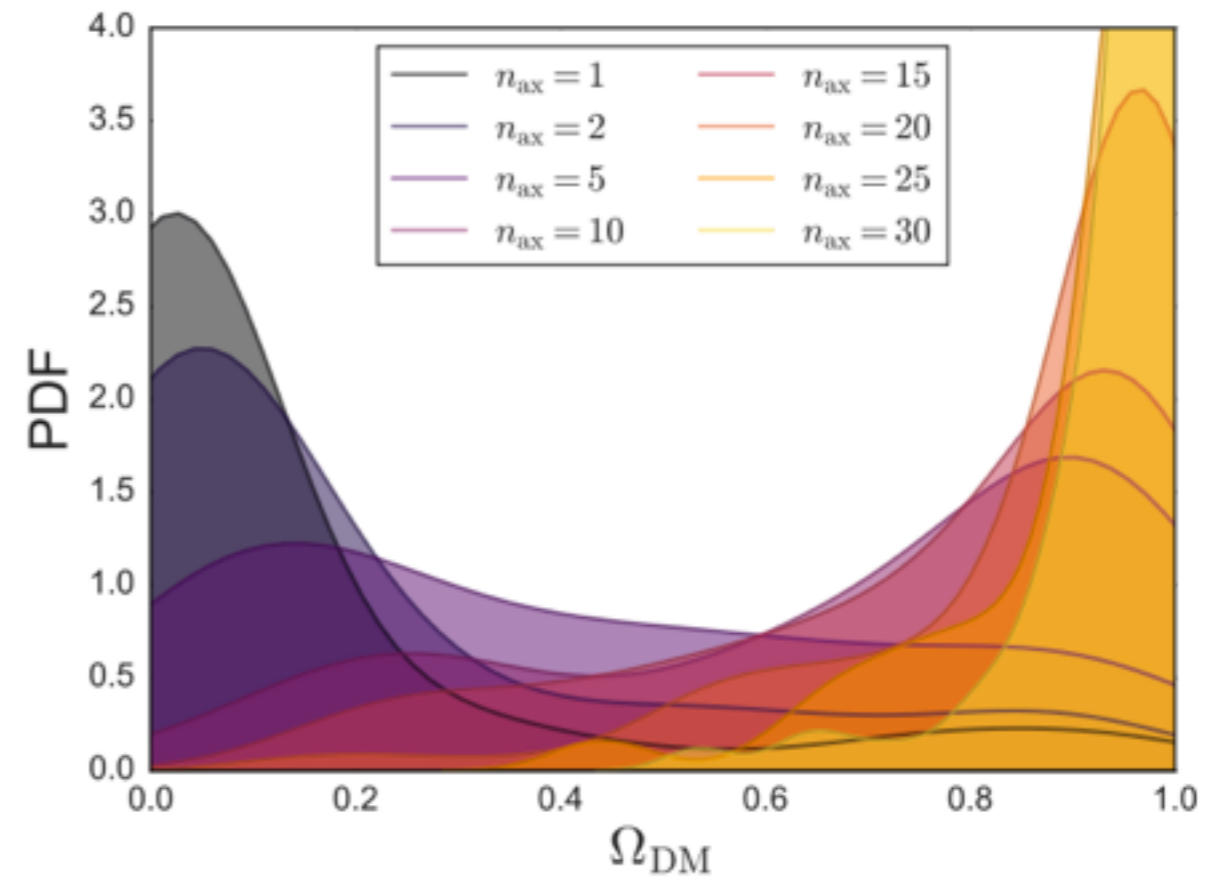
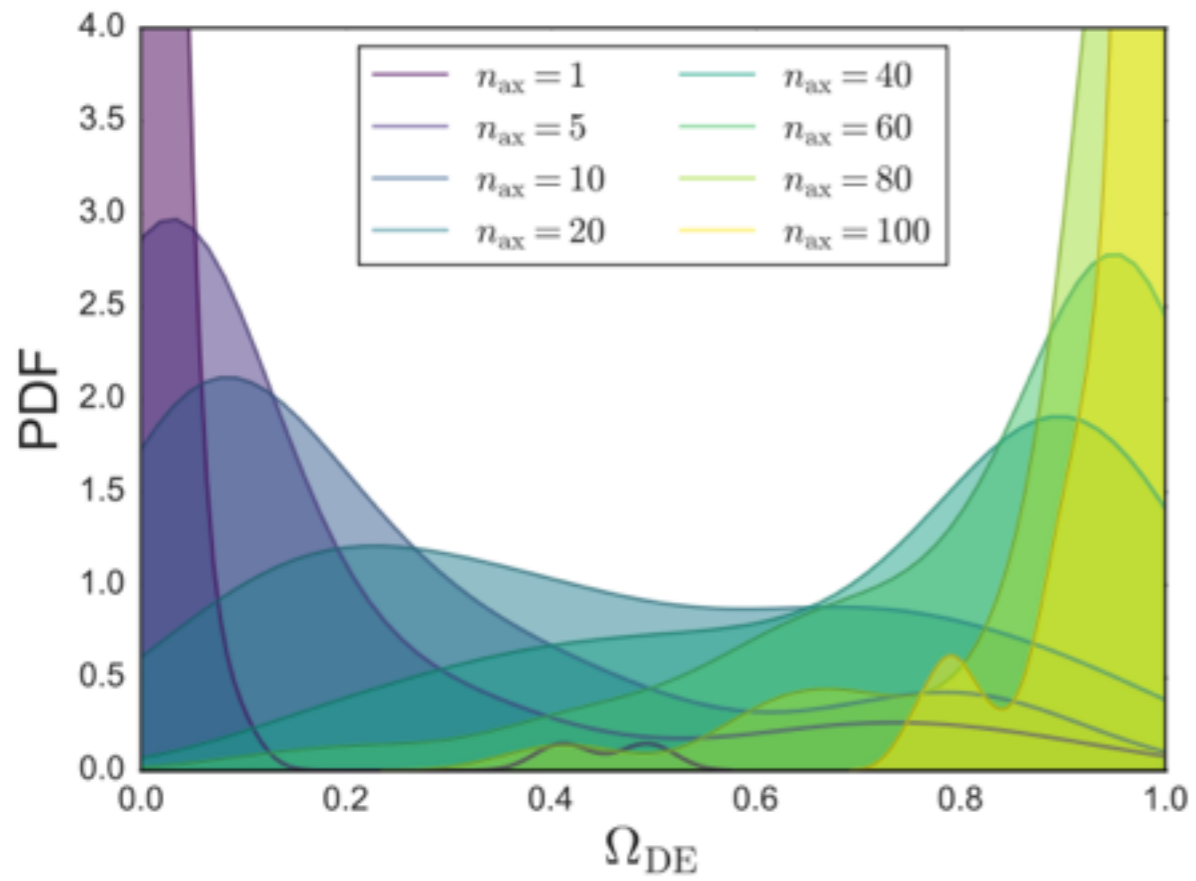
## **1. Vibrating chartjunk:**

cross-hatching or other patterns that distract the mind from the information being presented

# Types of chartjunk



# Types of chartjunk



same paper, same colours, smooth histograms!

# Types of chartjunk

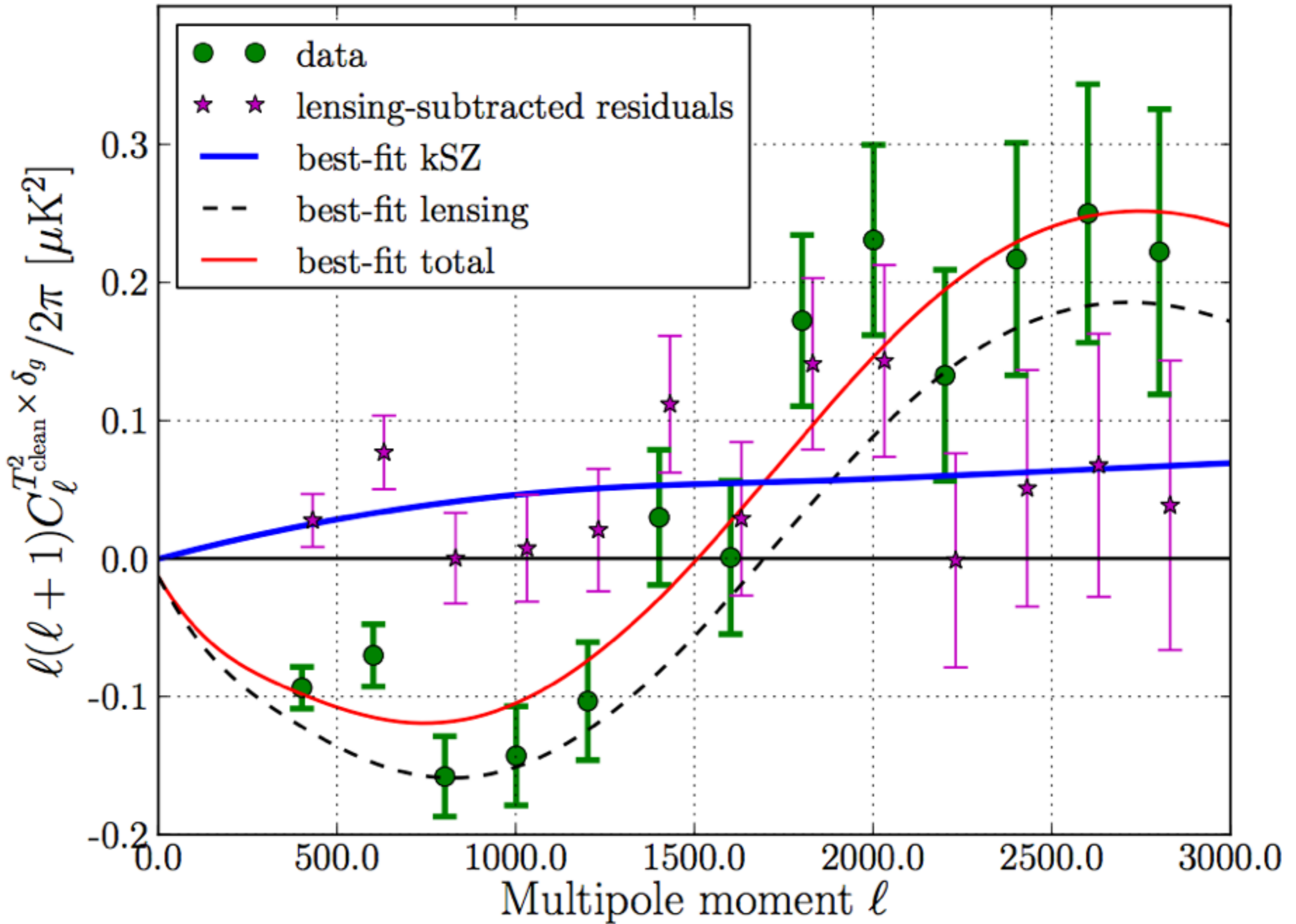
## **1. Vibrating chartjunk:**

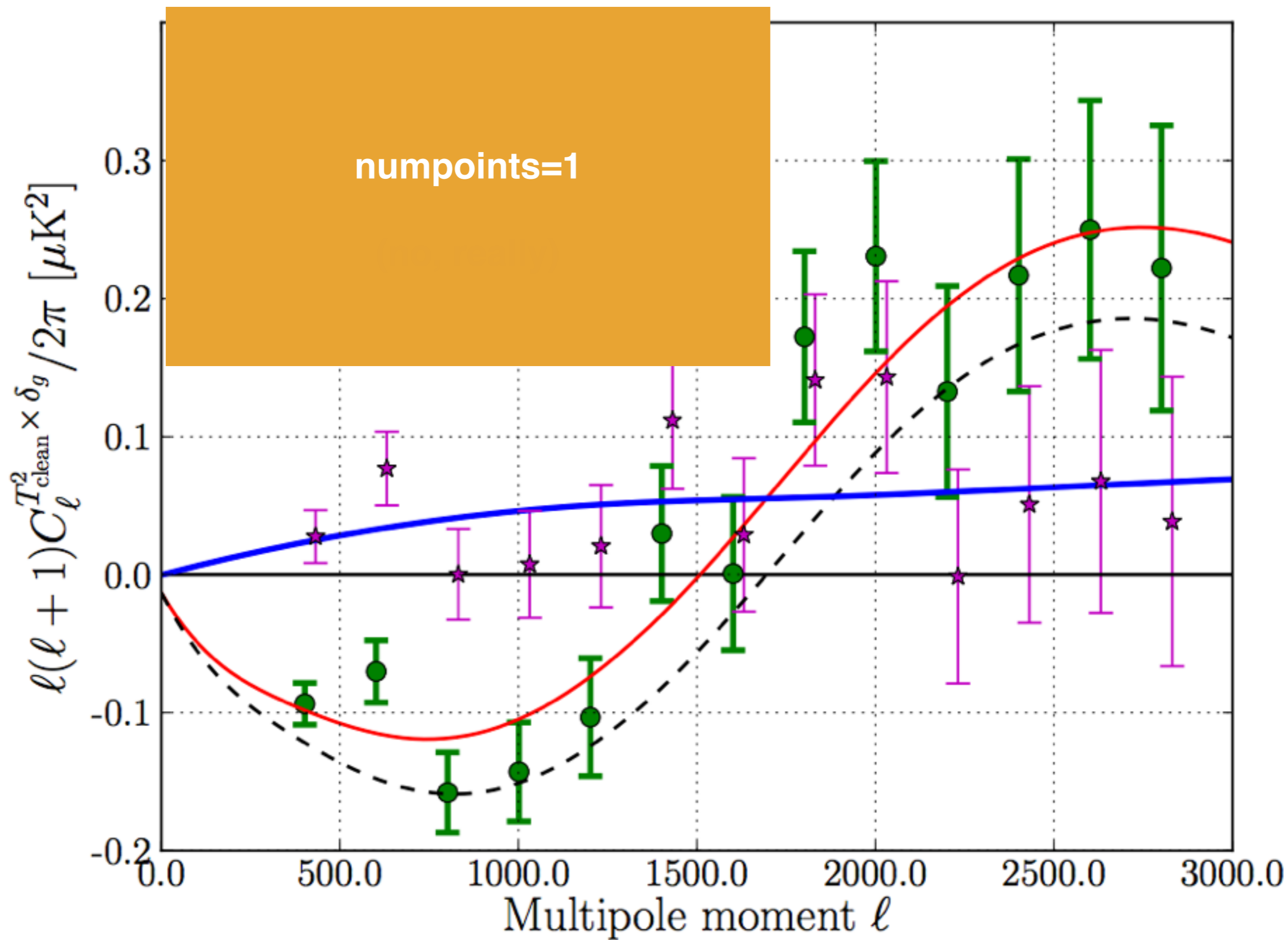
cross-hatching or other patterns that distract the mind from the information being presented

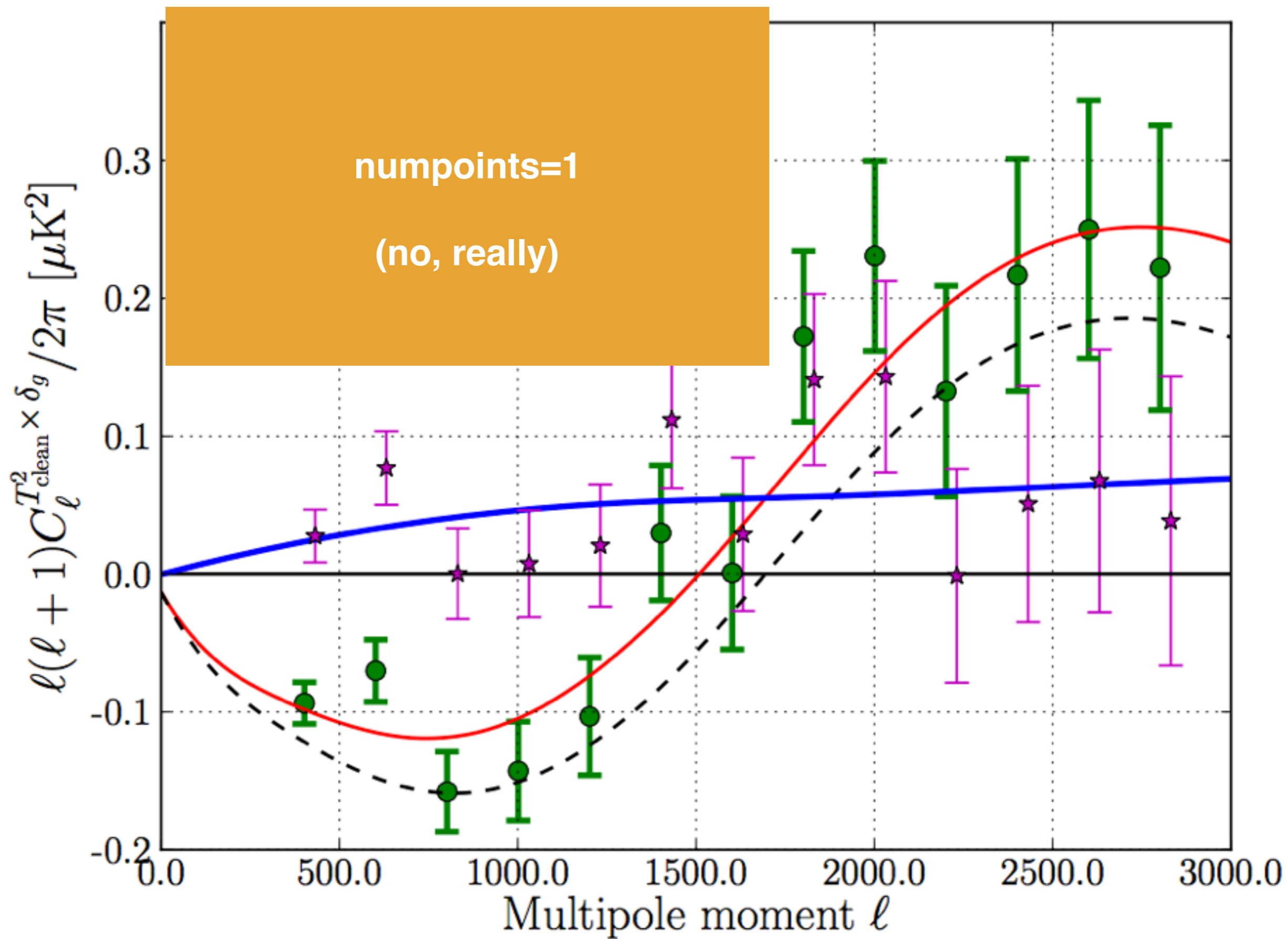
## **2. Grids:**

make them a lighter gray, not black, if you are going to use them

check when grids are necessary







# Types of chartjunk

## **1. Vibrating chartjunk:**

cross-hatching or other patterns that distract the mind from the information being presented

## **2. Grids:**

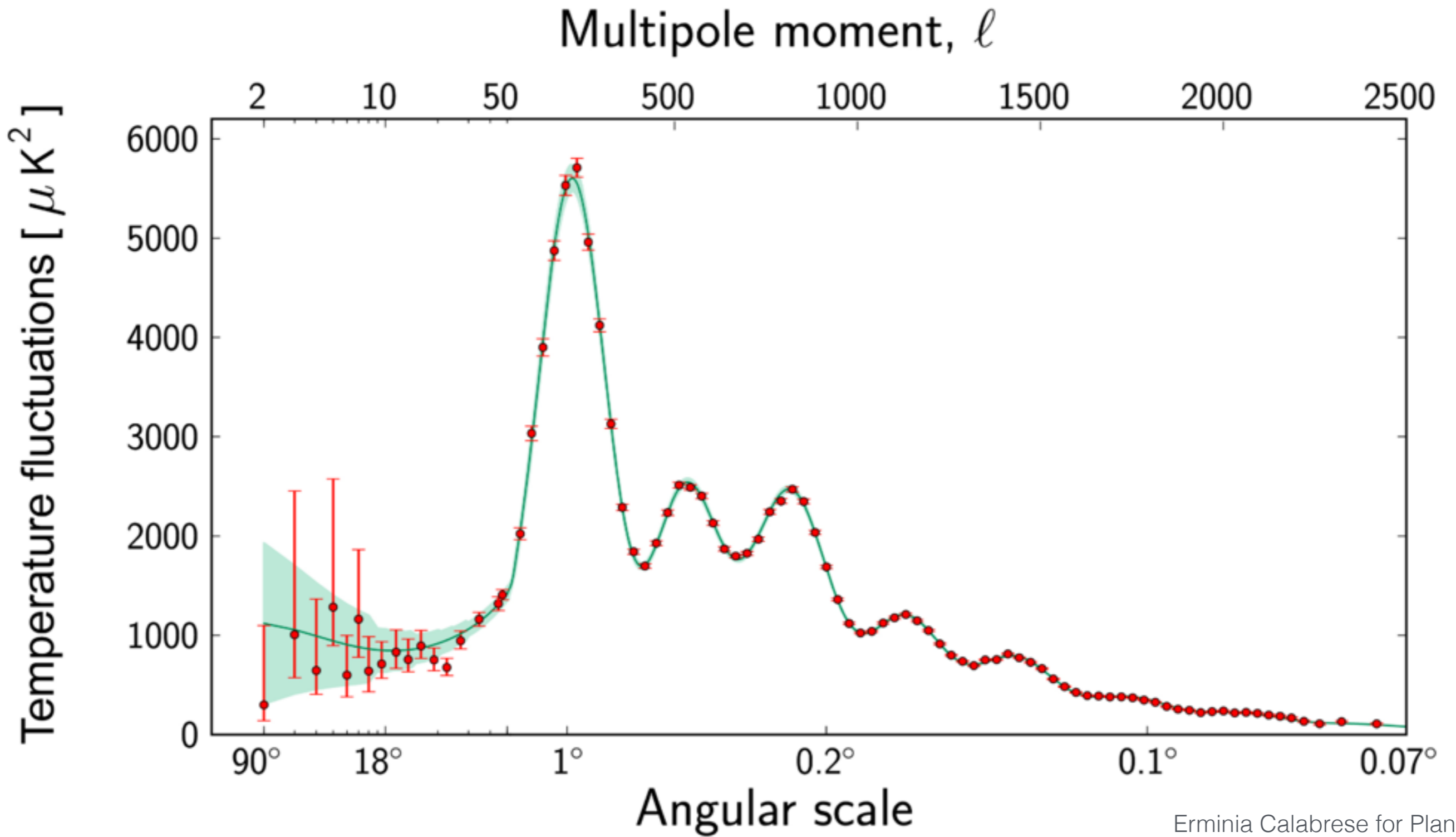
make them a lighter gray, not black, if you are going to use them

## **3. Self-promoting graphics (“The Duck”):**

color schemes and patterns are introduced for artistic appeal rather than information content.

# Scale/axes

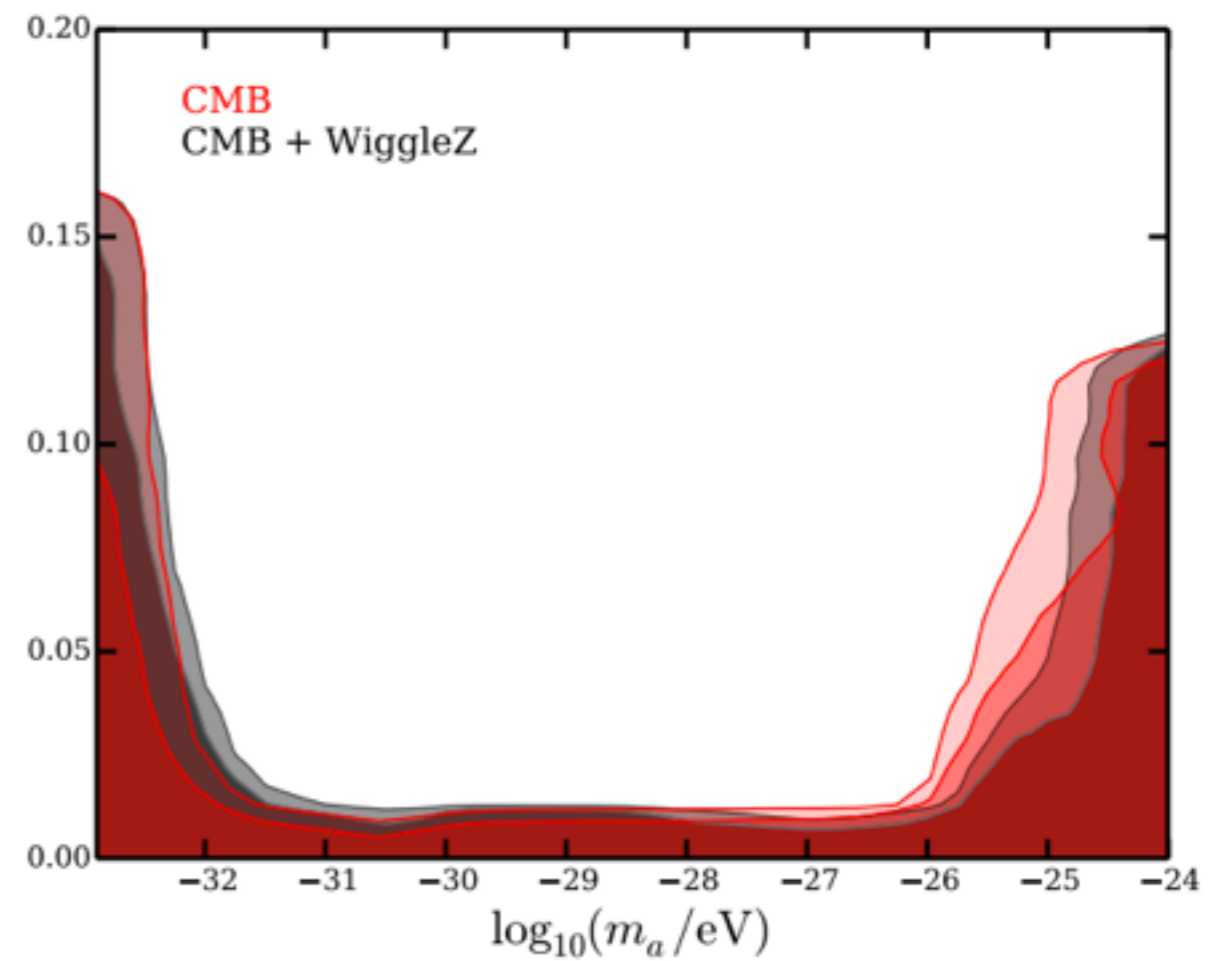
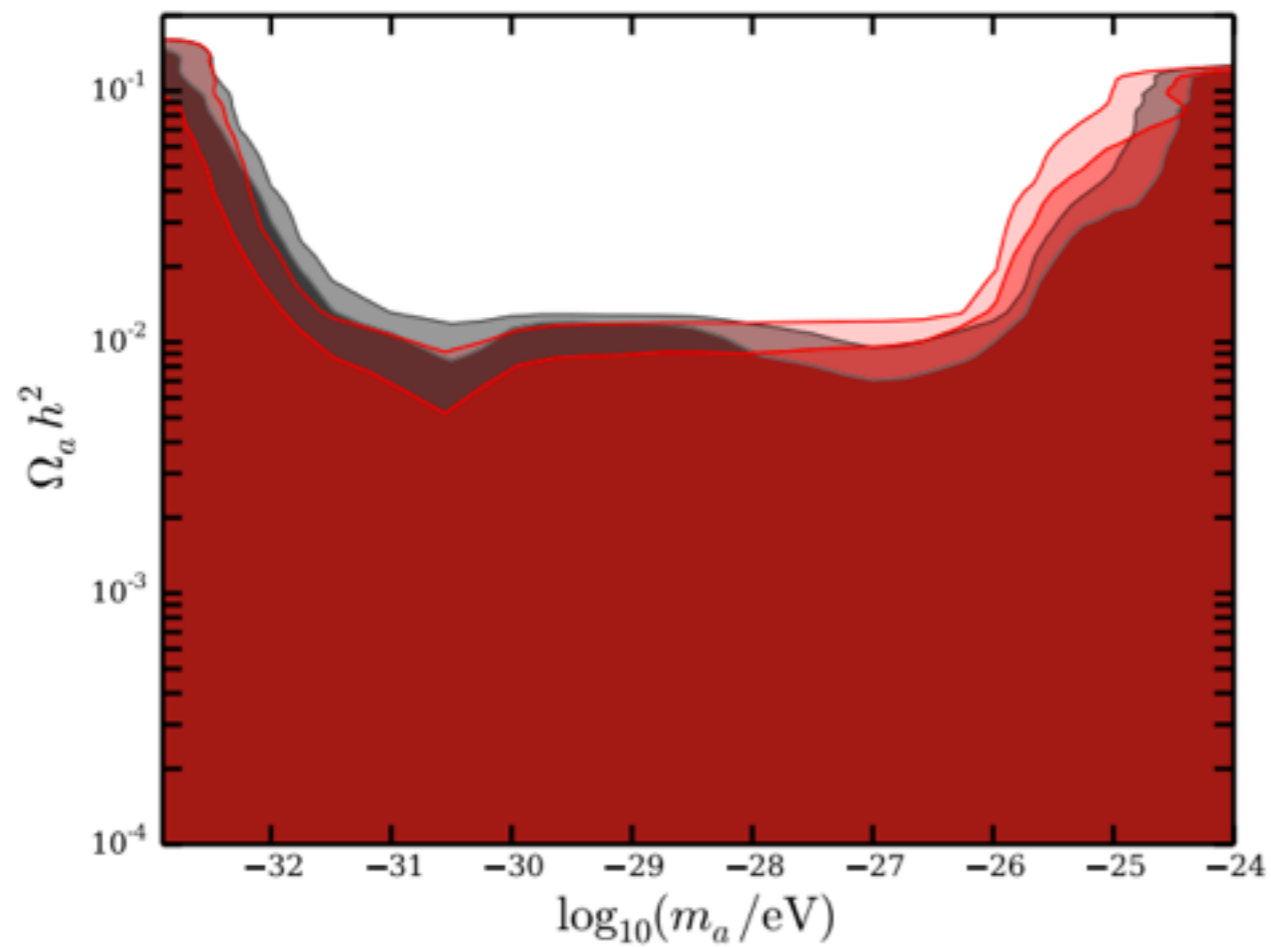
1. Choose the range of the tick marks to include or nearly include the range of data. (to white space or not..)
2. Subject to the constraints that scales have, choose the scales so that the data fill **up as much of the region as possible**.
3. It is sometimes helpful to use the pair of scale lines for a variable to show two different scales.



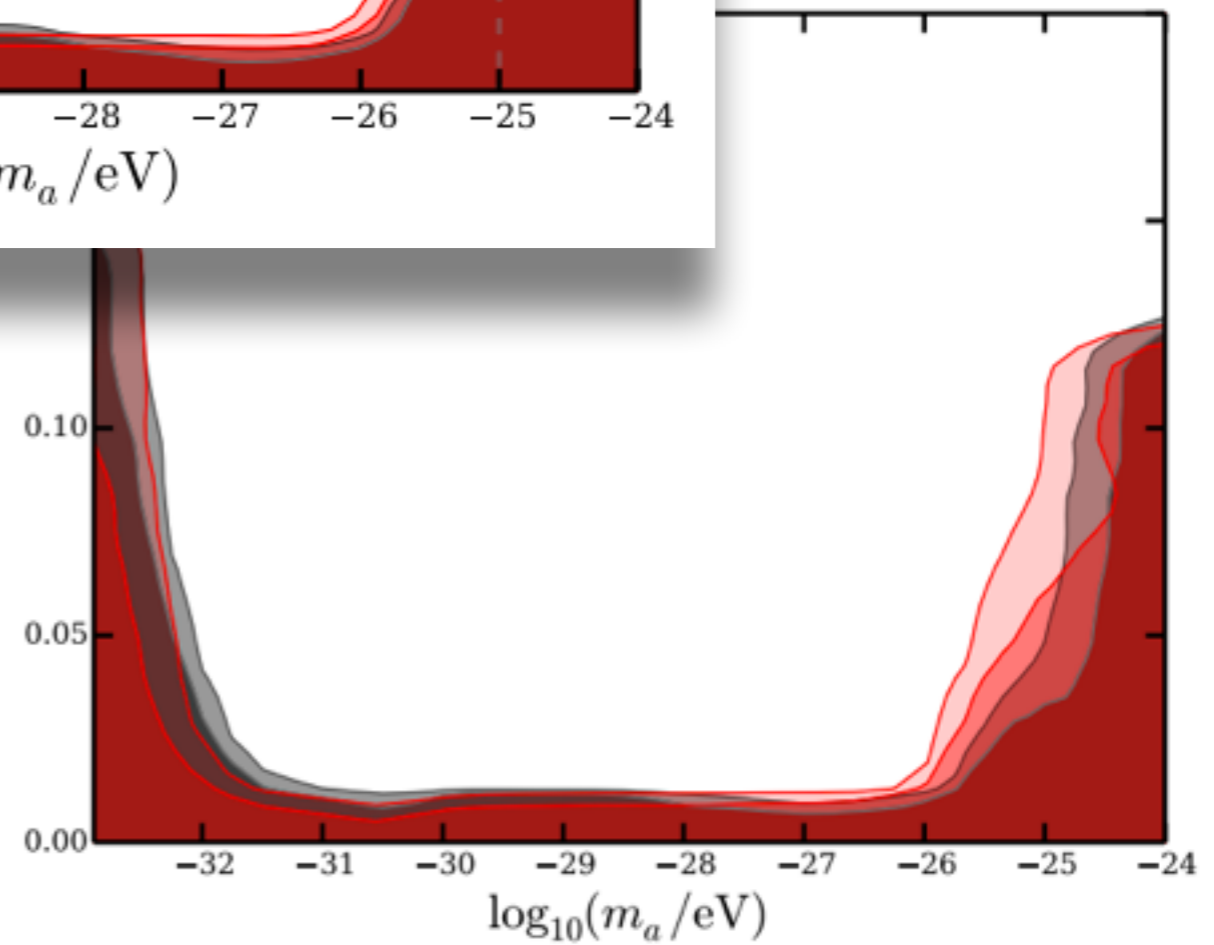
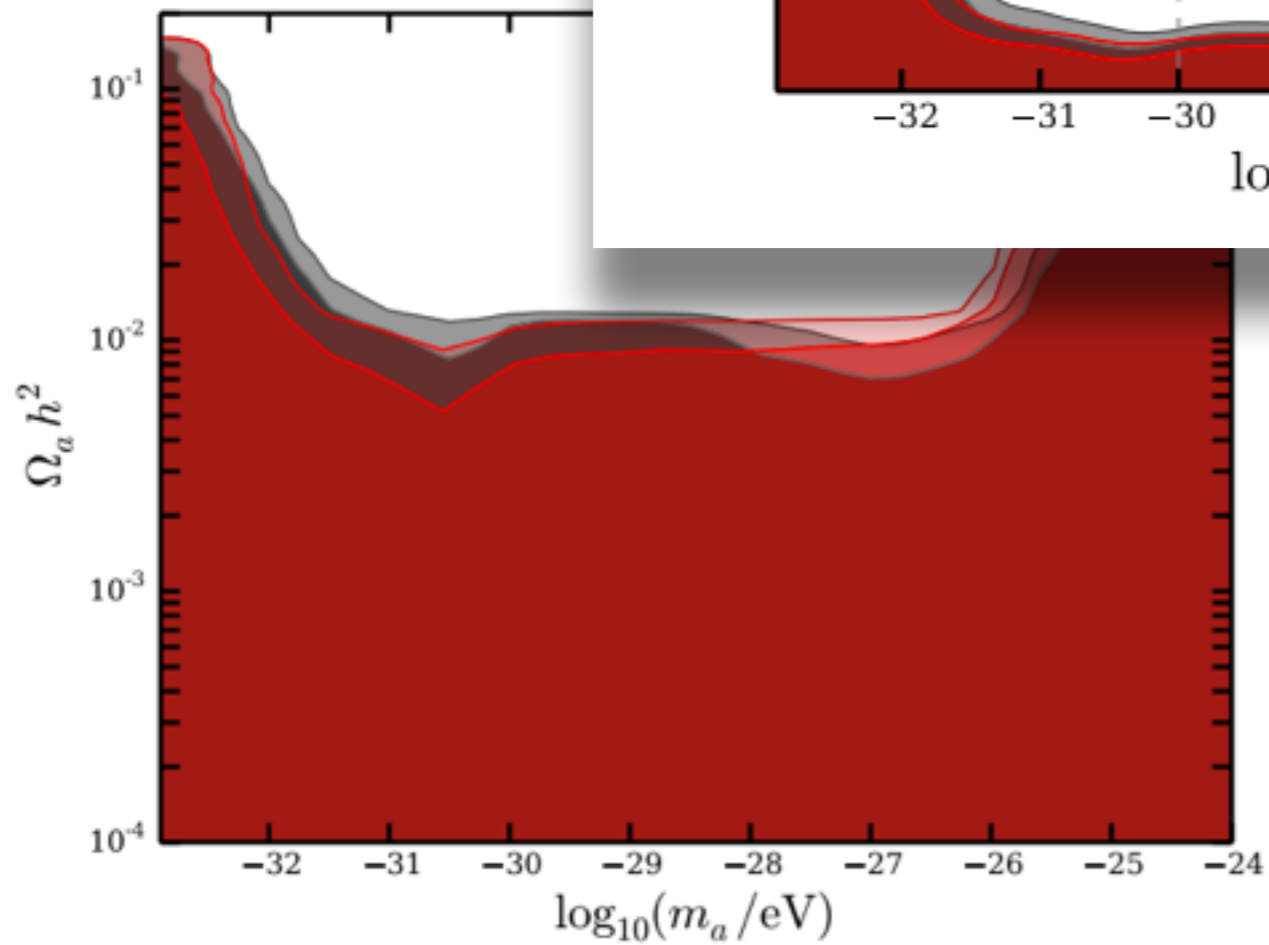
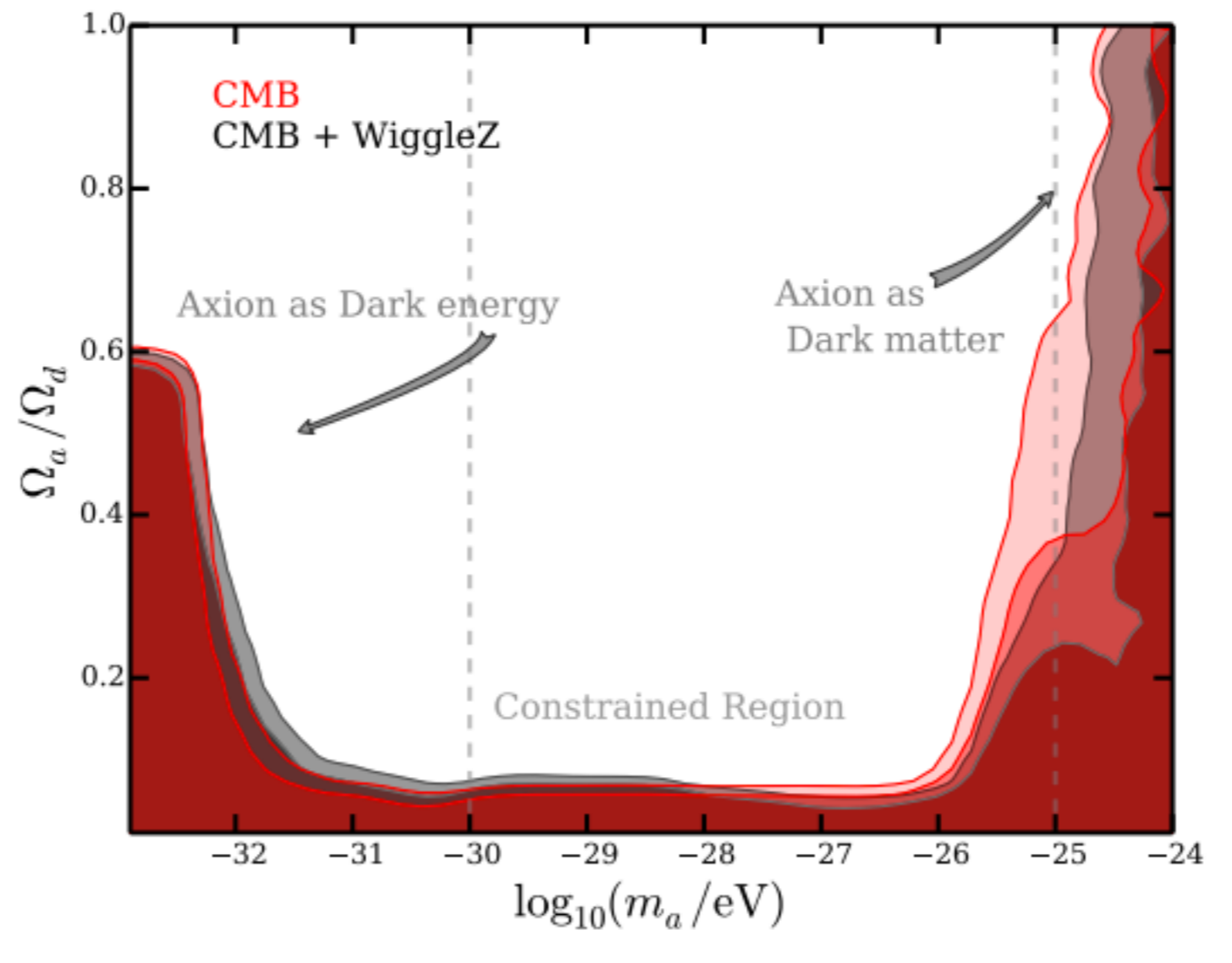
# Scale/axes

1. Choose the range of the tick marks to include or nearly include the range of data. (to white space or not..)
2. Subject to the constraints that scales have, choose the scales so that the data fill **up as much of the region as possible**.
3. It is sometimes helpful to use the pair of scale lines for a variable to show two different scales.
4. Choose appropriate scales when graphs are compared.
5. Use a logarithmic scale when it is important to understand percent change or multiplicative factors.
6. Showing data on a logarithmic scale can improve resolution - but you should check first (**intentionality**).

# Scale/axes



# Scale/axes

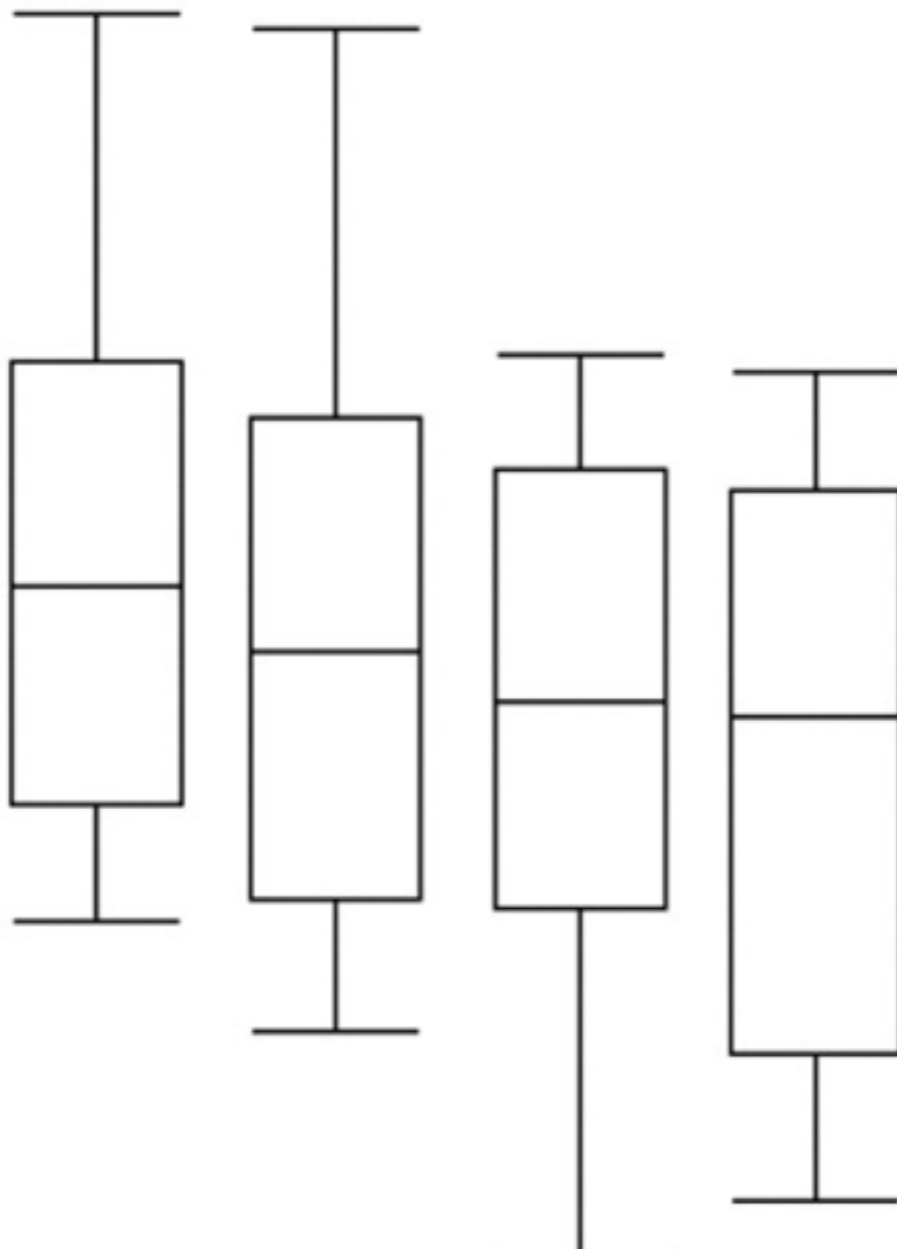


# Redesigning the box plot

Typical series of box plots

Tufte's "box" plot

Maximum



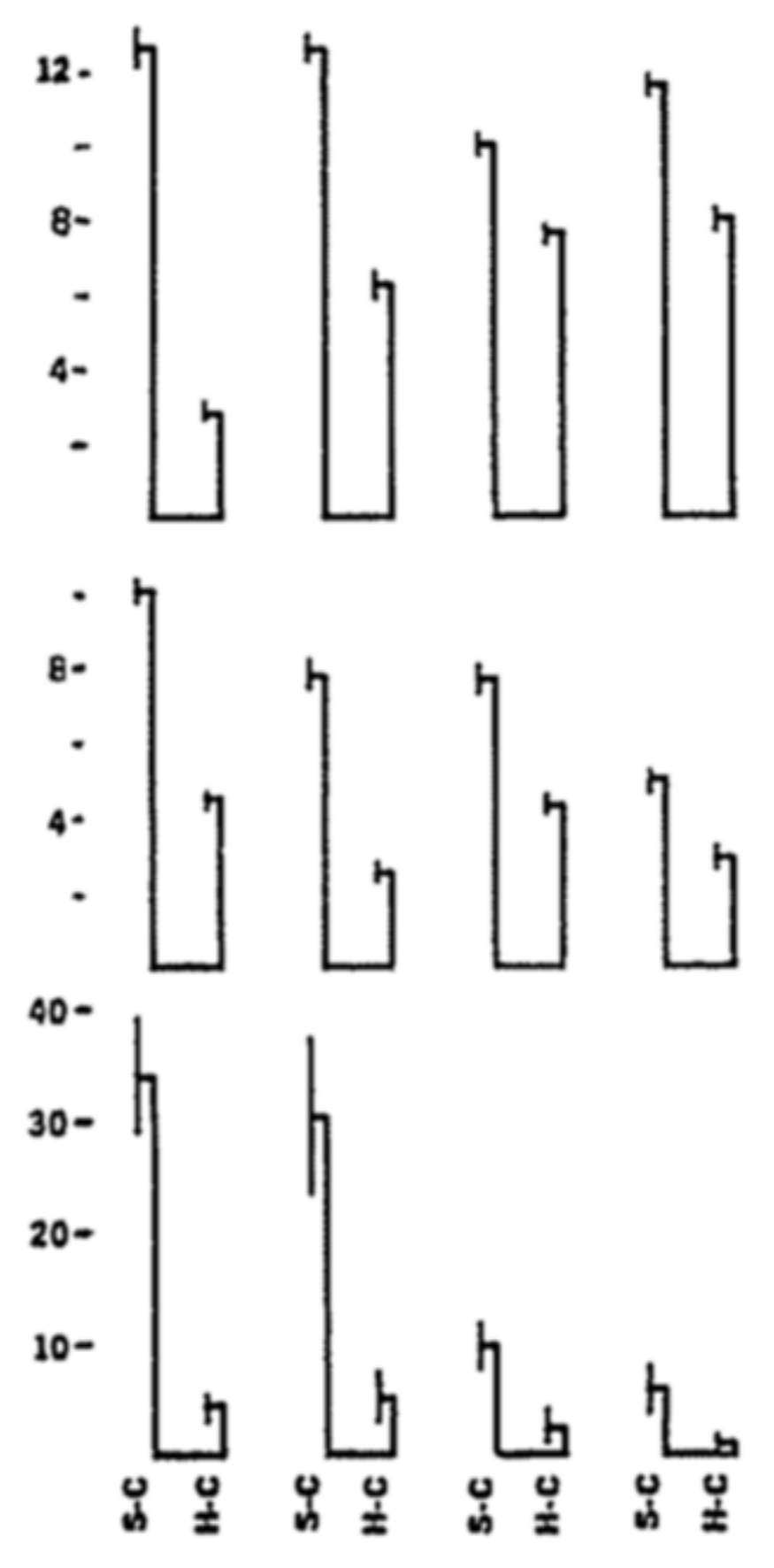
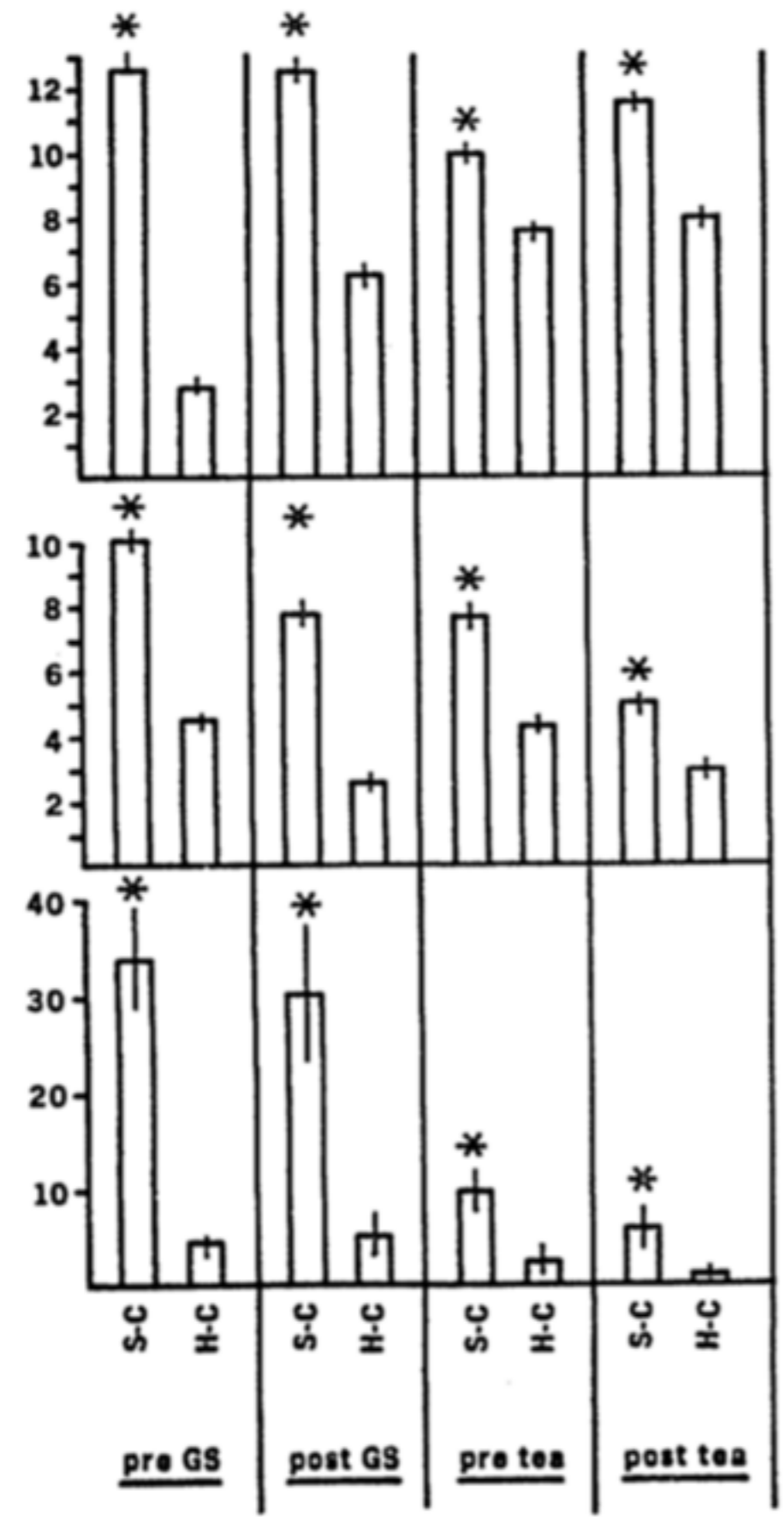
Quartile

Median

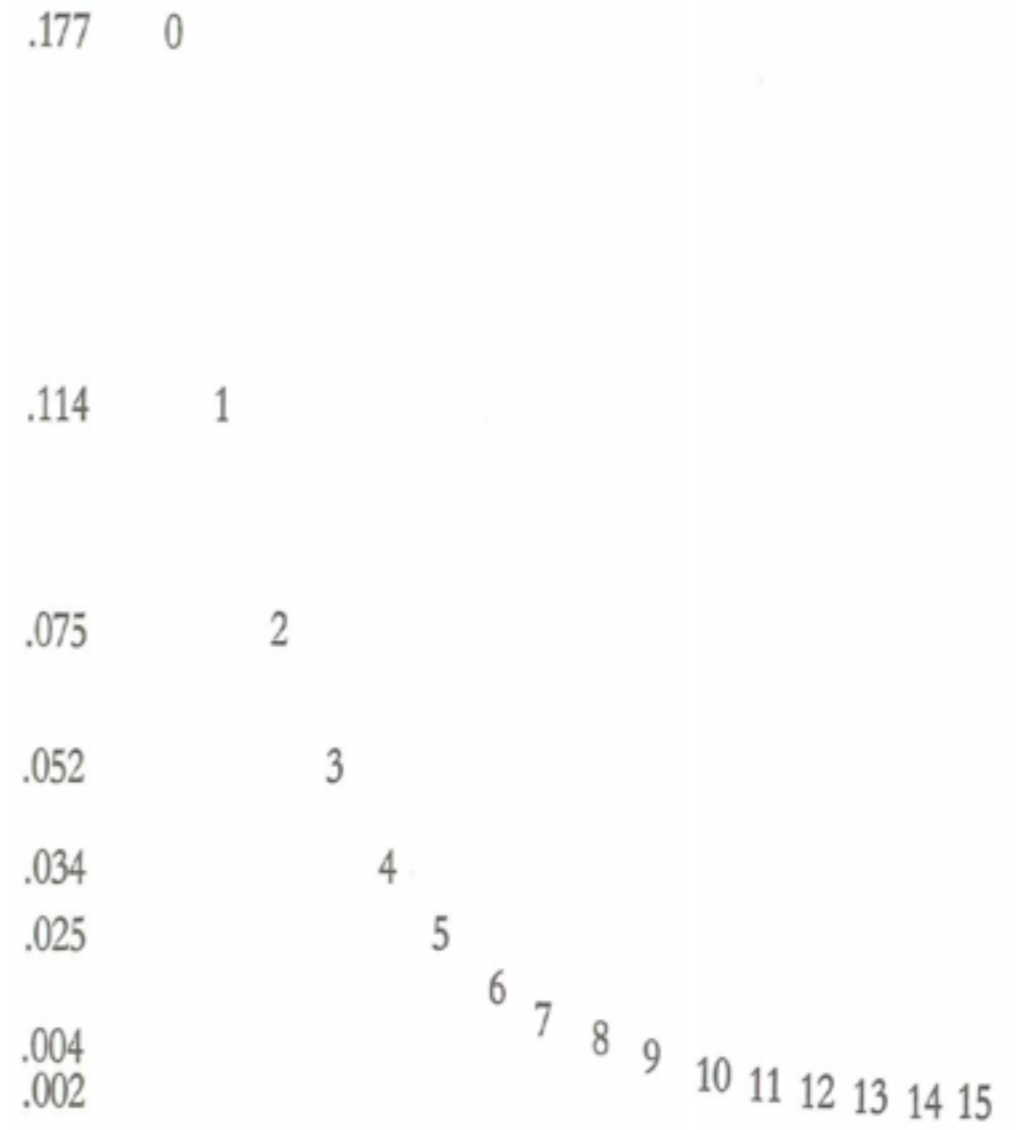
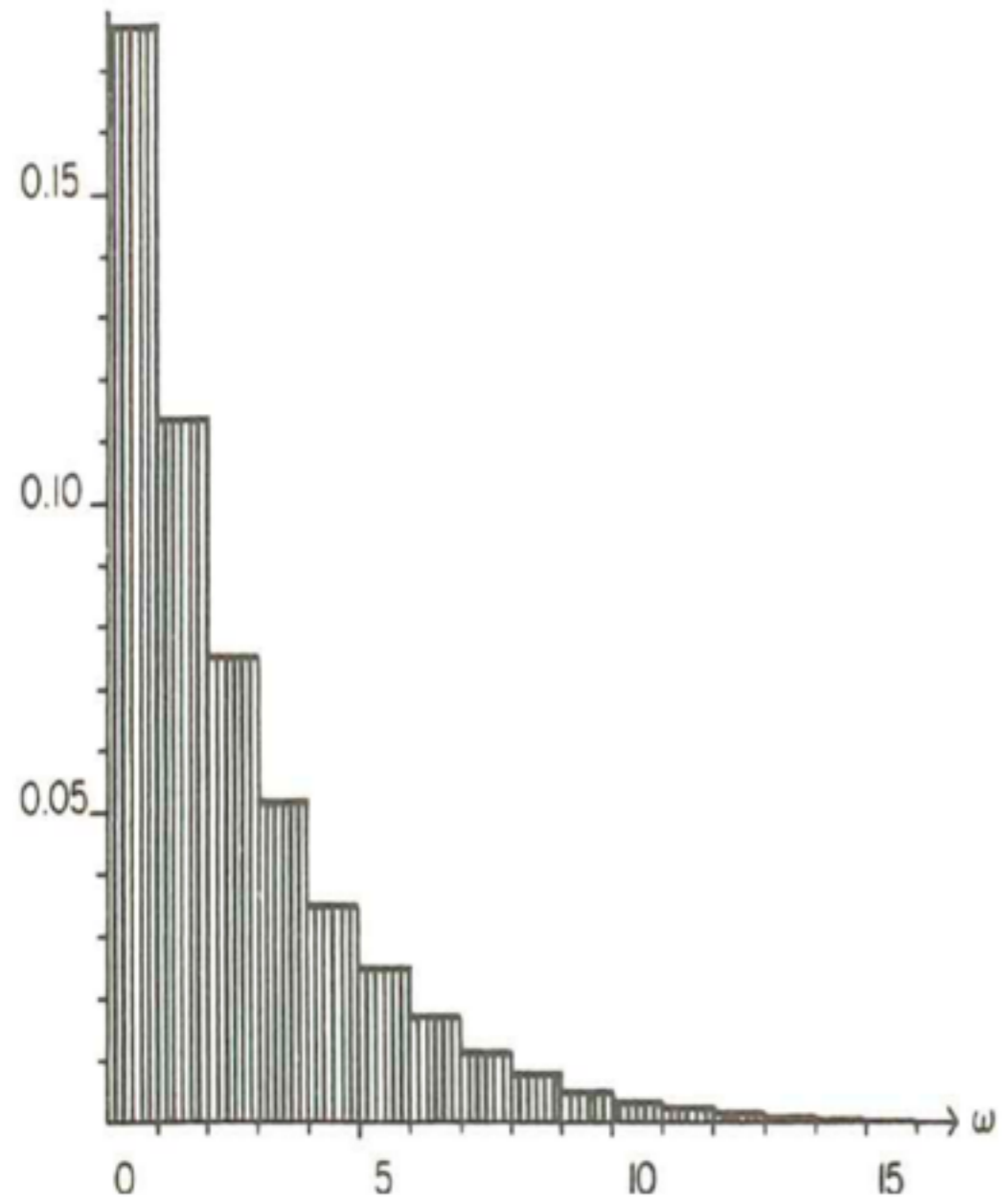
Quartile

Minimum

# Redesigning the bar chart



# labels to show structure



# Tufte's design principles

$$\text{data density} = \frac{\text{number of entries in data matrix}}{\text{total area of the figure}}$$

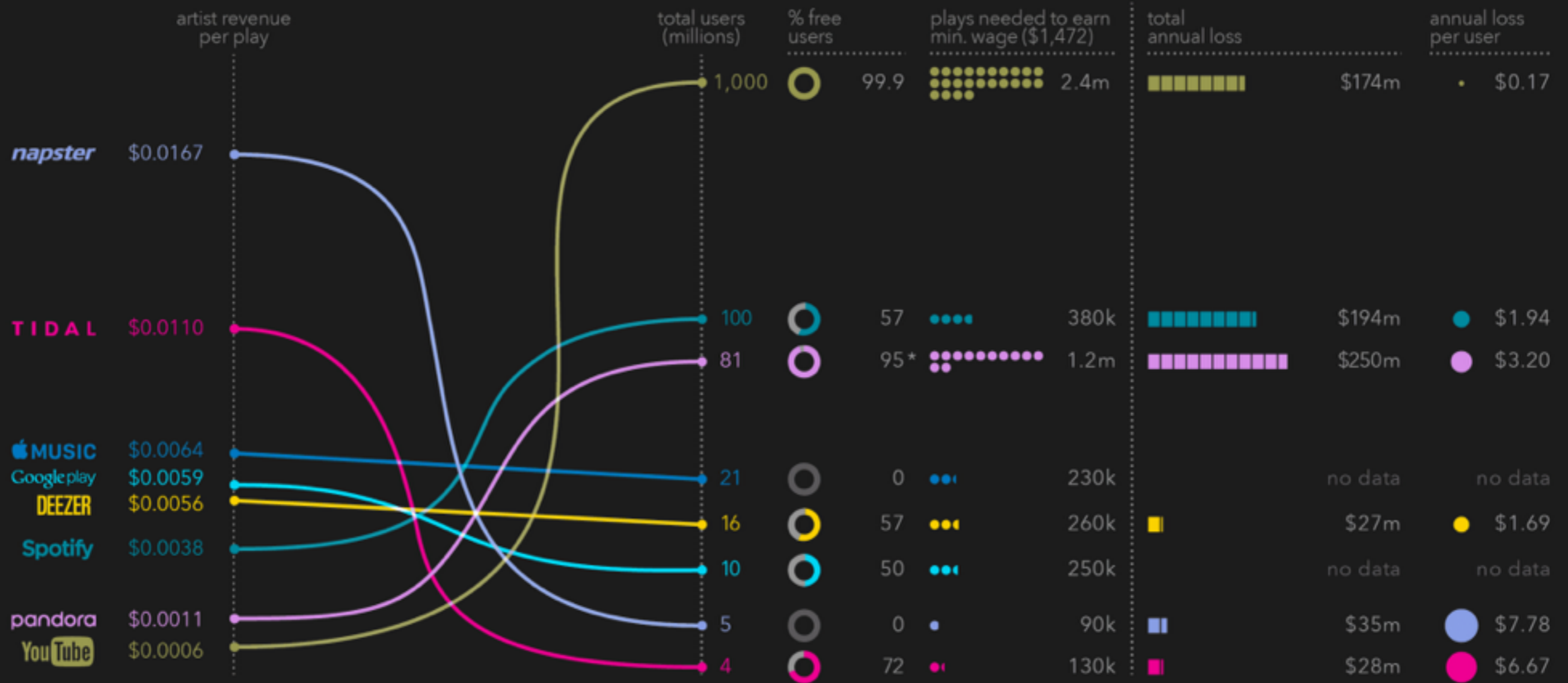
This is not the same as the data ink ratio - this is about making sure that your eyes are used for the important stuff!

**Minimise white space**

# Tufte is not always beautiful!

## Money Too Tight to Mention? Unsigned Artists

Major music streaming services compared



Last Update: 21 Apr 2017

\* based on 5% of standard Pandora users taking up their new on-demand service

data: bit.ly/KIB\_stream

informationisbeautiful.net

<http://www.informationisbeautiful.net/visualizations/spotify-apple-music-tidal-music-streaming-services-royalty-rates-compared/>

Using graphs to guide learning not to show 'work'



[https://www.ted.com/talks/hans\\_rosling\\_the\\_truth\\_about\\_hiv](https://www.ted.com/talks/hans_rosling_the_truth_about_hiv)

# Final thoughts/principles

1. **Think first**: what is the take home message?
2. Make the data stand out.
3. Use **visually prominent** graphical elements to show the data - be intentional.
4. Don't **clutter the data area** (but also don't waste space)
5. Use a **reference line only when it matters**, but do not let the line interfere with the data.
6. Do not allow data labels in the data region to interfere with the quantitative data or to clutter the graph. **Check your legend!**

## Final thoughts/principles

5. Overlapping plotting symbols must be **visually distinguishable**.
6. Superposed data sets must be **readily visually discriminated**.
7. **Visual clarity** must be preserved under reduction and reproduction.
8. Check your figures in **B&W and colour** when printing.
9. Make sure you **respect colourblindness** wherever possible
10. Proofread your graphs!
11. Get **a friend to explain the graph** without you speaking.
12. (come to astroph coffee)