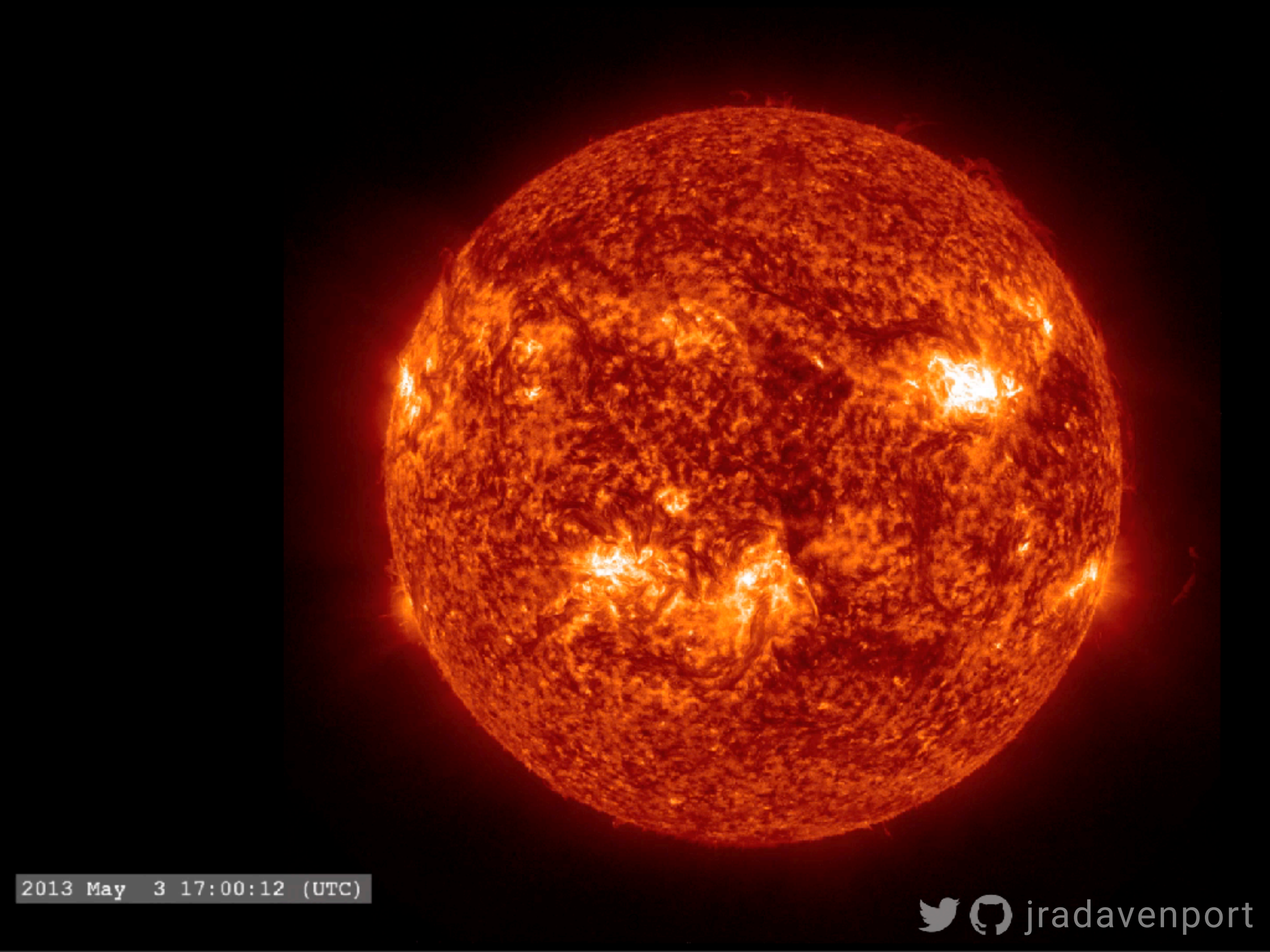
Data Visualization for Academic Writing

James R. A. Davenport
NSF Astronomy & Astrophysics Postdoctoral Fellow, Western Washington University
DIRAC Fellow, University of Washington

jradavenport

about me

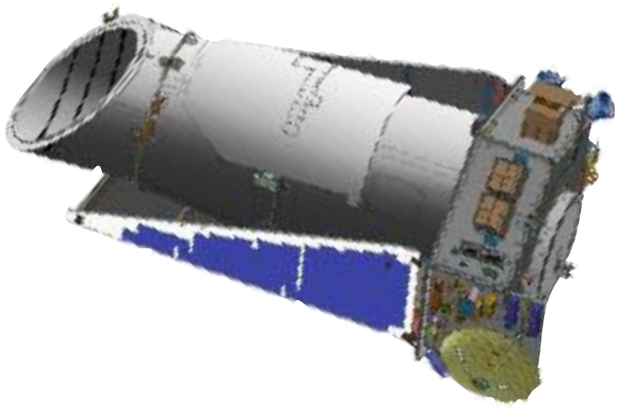
jradavenport

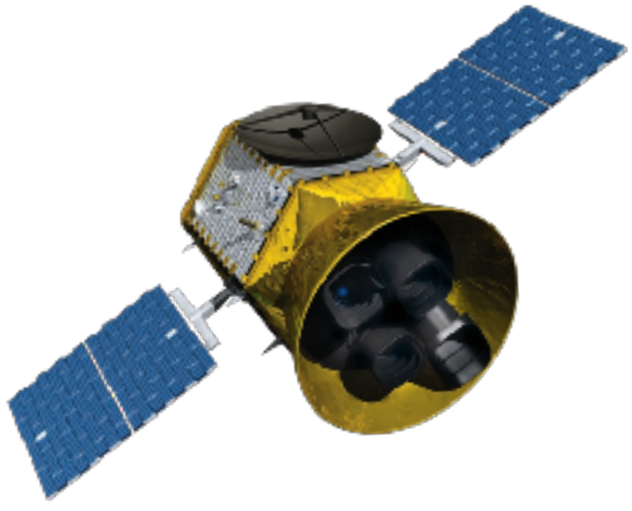

2013 May 3 17:00:12 (UTC)
jradavenport

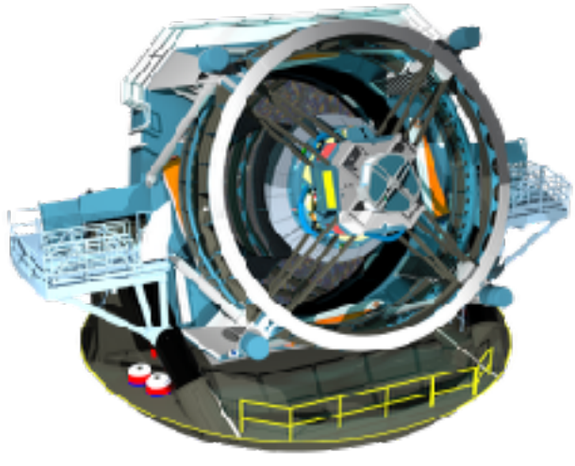
Aurora Borealis - Frederic Edwin Church (1865)

jradavenport

DiRAC







jradavenport

If We Assume

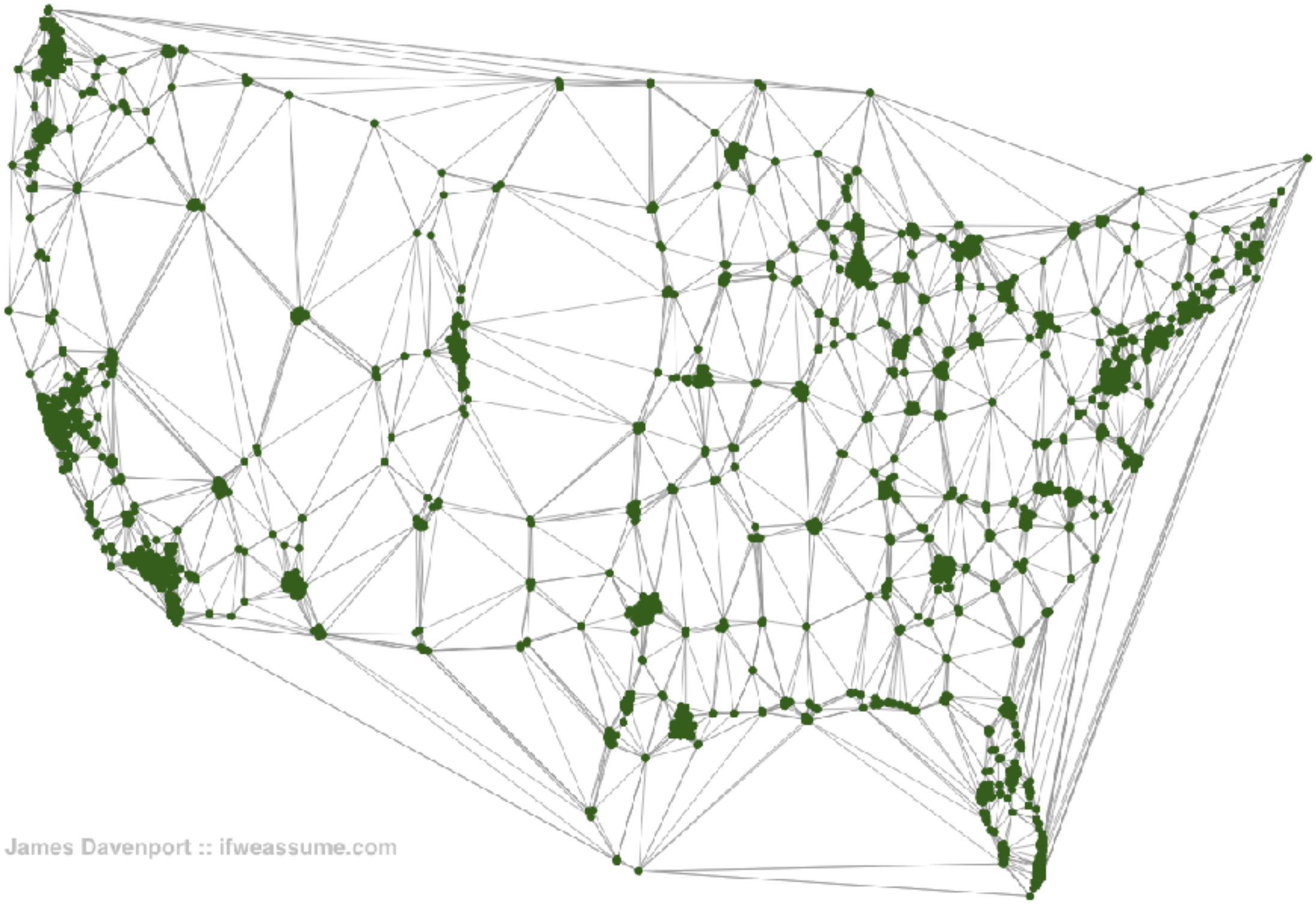
Better Living Through Data

Topics: Apple, Everyday Data, happiness, visualization

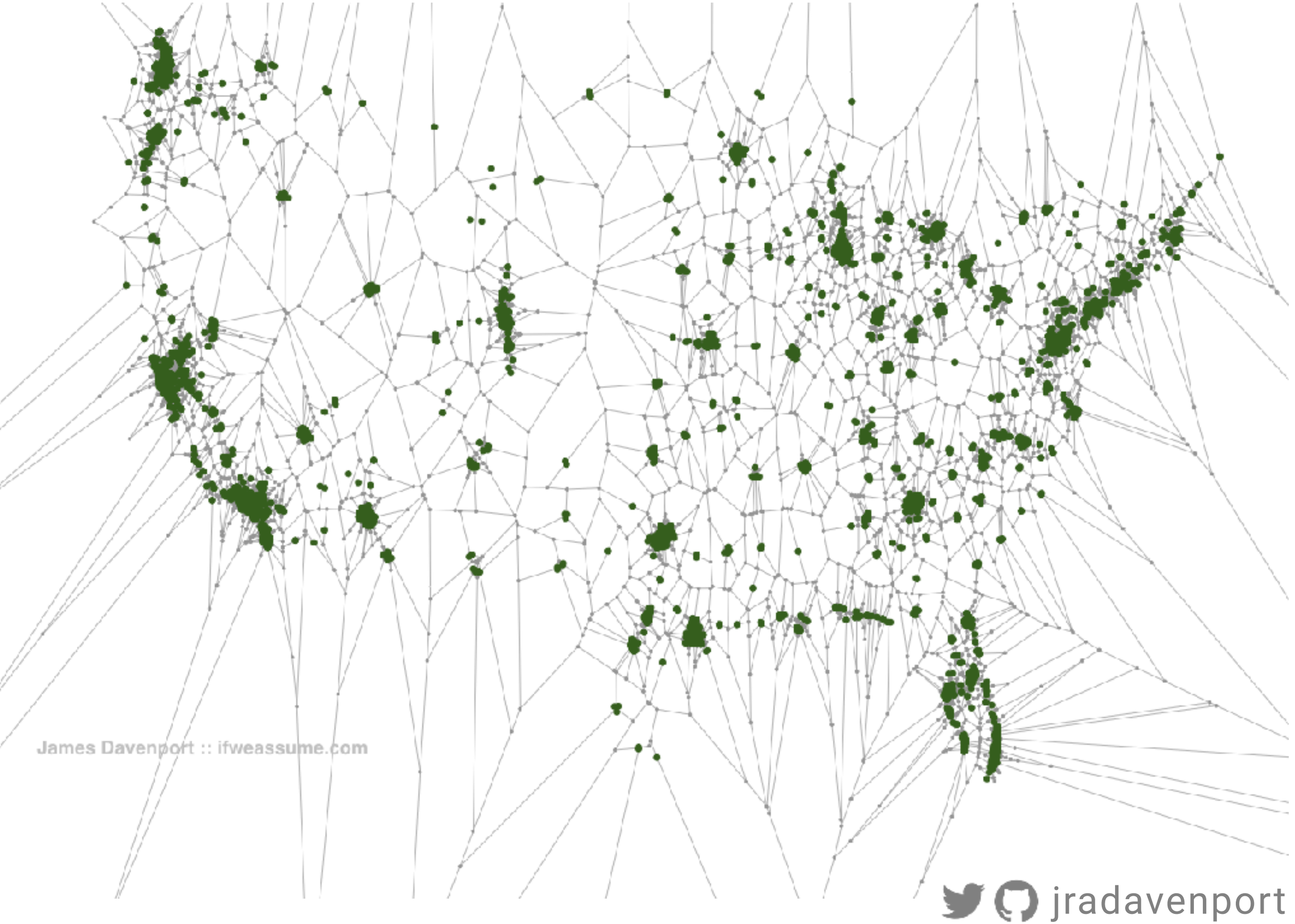
One running theme on this blog has been that of data-driven self study. A favorite source for data about myself is my laptop battery logs. Last summer I shared what an entire year of laptop battery usage looks like, in remarkable detail. Today I'm excited to show the follow up data!

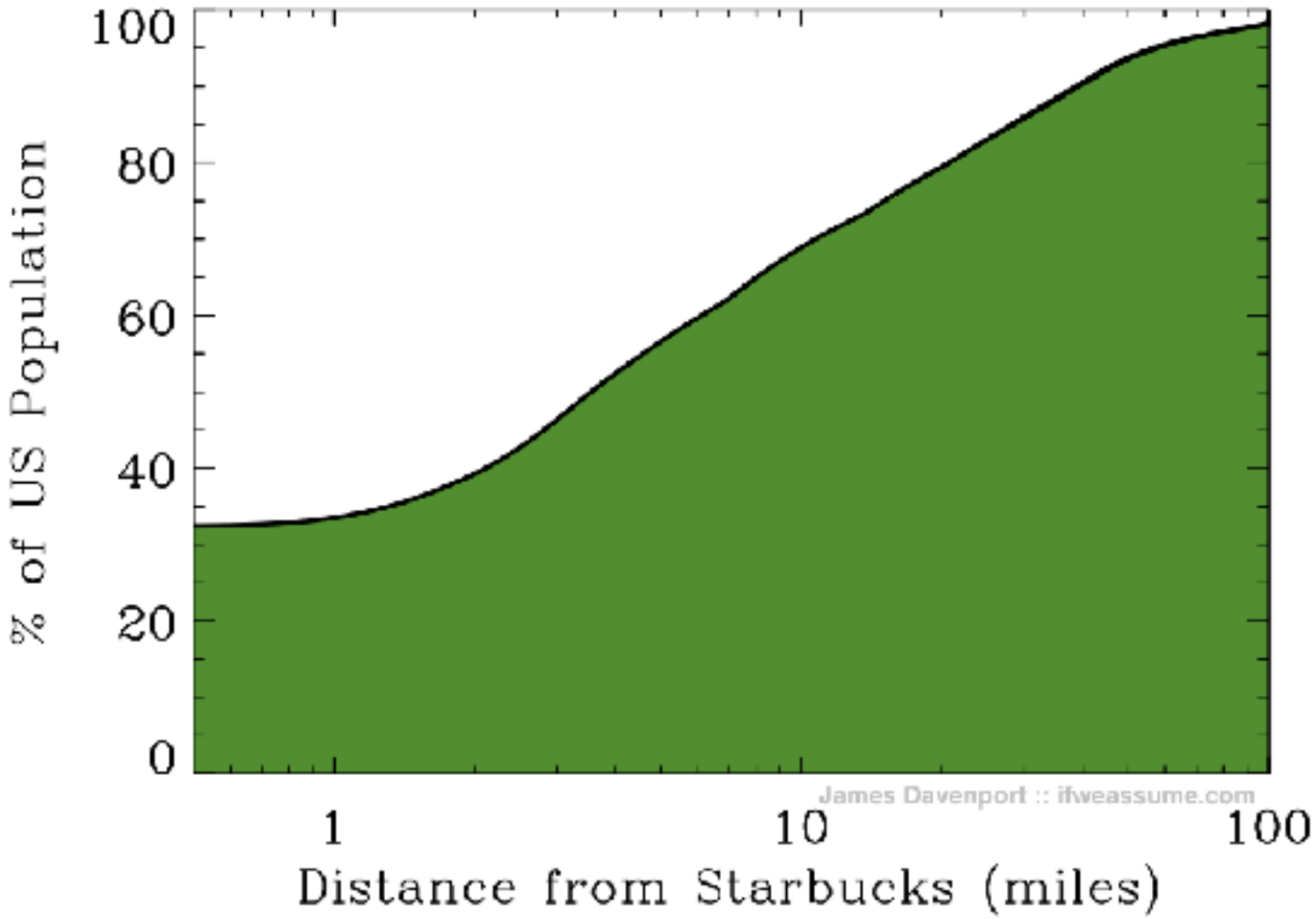
Here is what **two years of laptop battery use looks like**, sampled every minute I've used my computer(s). This includes 293,952 data points, at time of writing. Since the "batlog" script runs *every minute*, that translates to over 204 days of computer use in the last ~2 years! Yowza

Update: Per several requests, I have added a more detailed install guide in the README file on github.



jradavenport


James Davenport :: ifweassume.com
jradavenport



jradavenport

24 hours of King County Metro

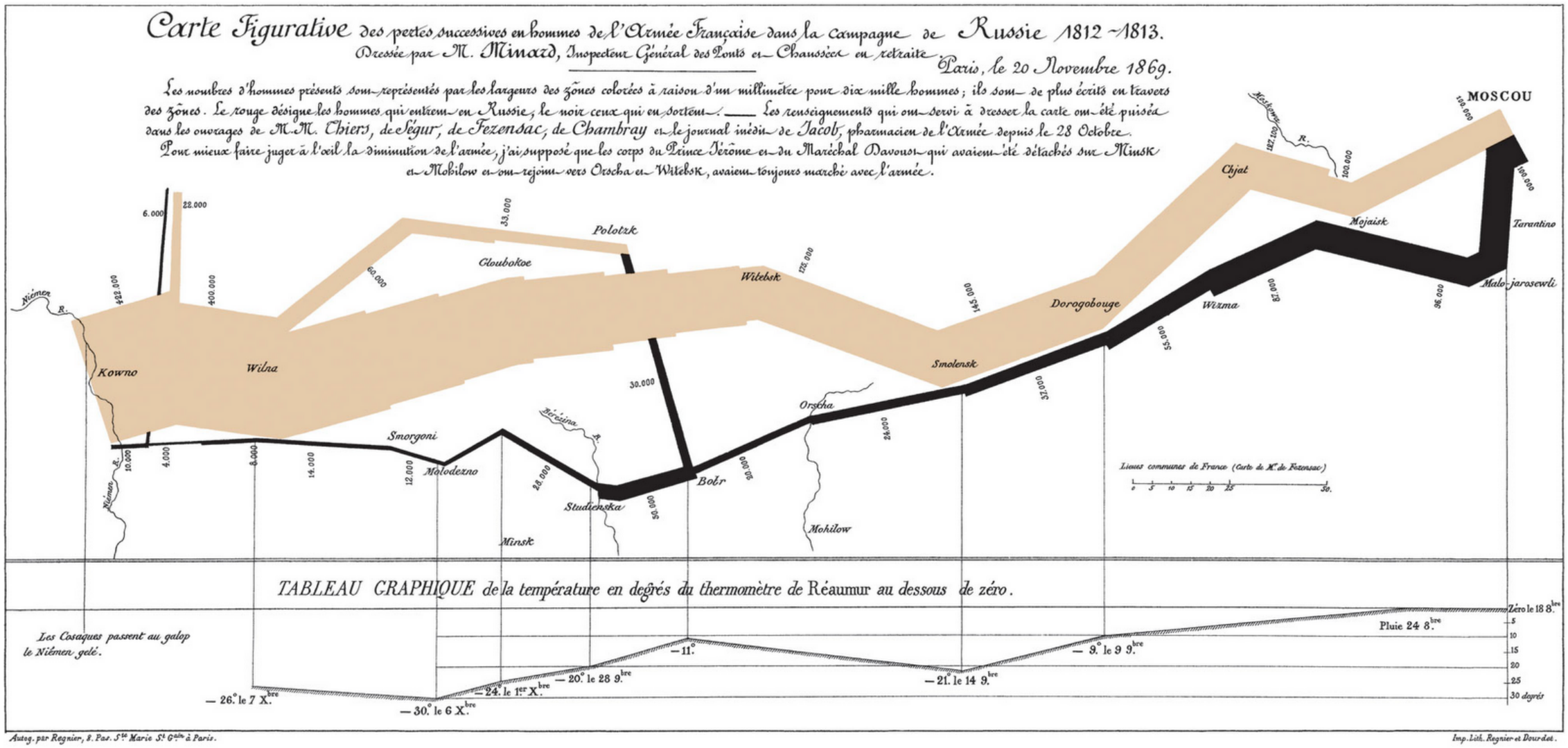
jradavenport

Data Visualization

or

Visual Storytelling

jradavenport



"The best statistical graph ever drawn…" -Edward Tufte

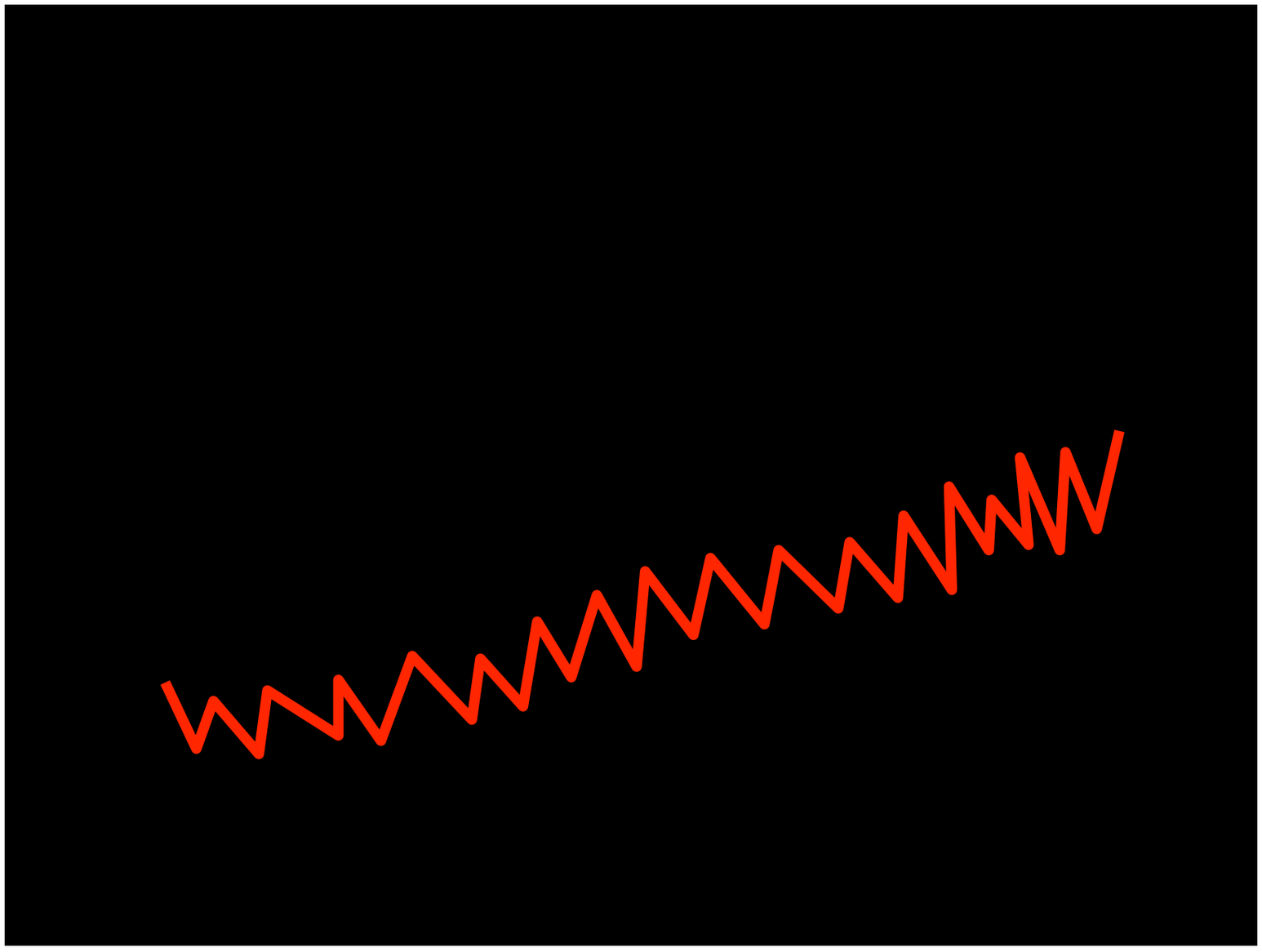
(except it wasn't drawn for PowerPoint or computers)

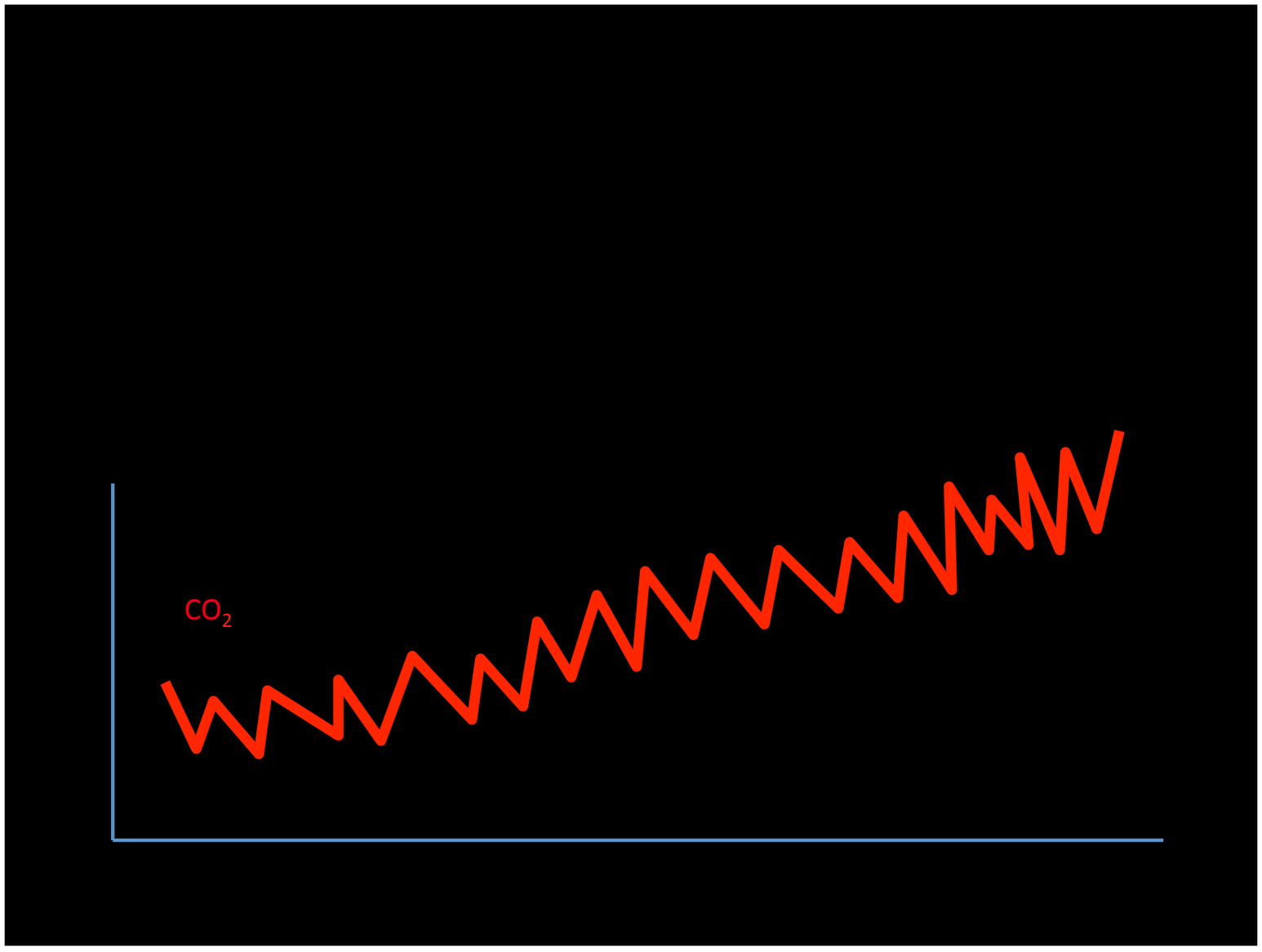
jradavenport

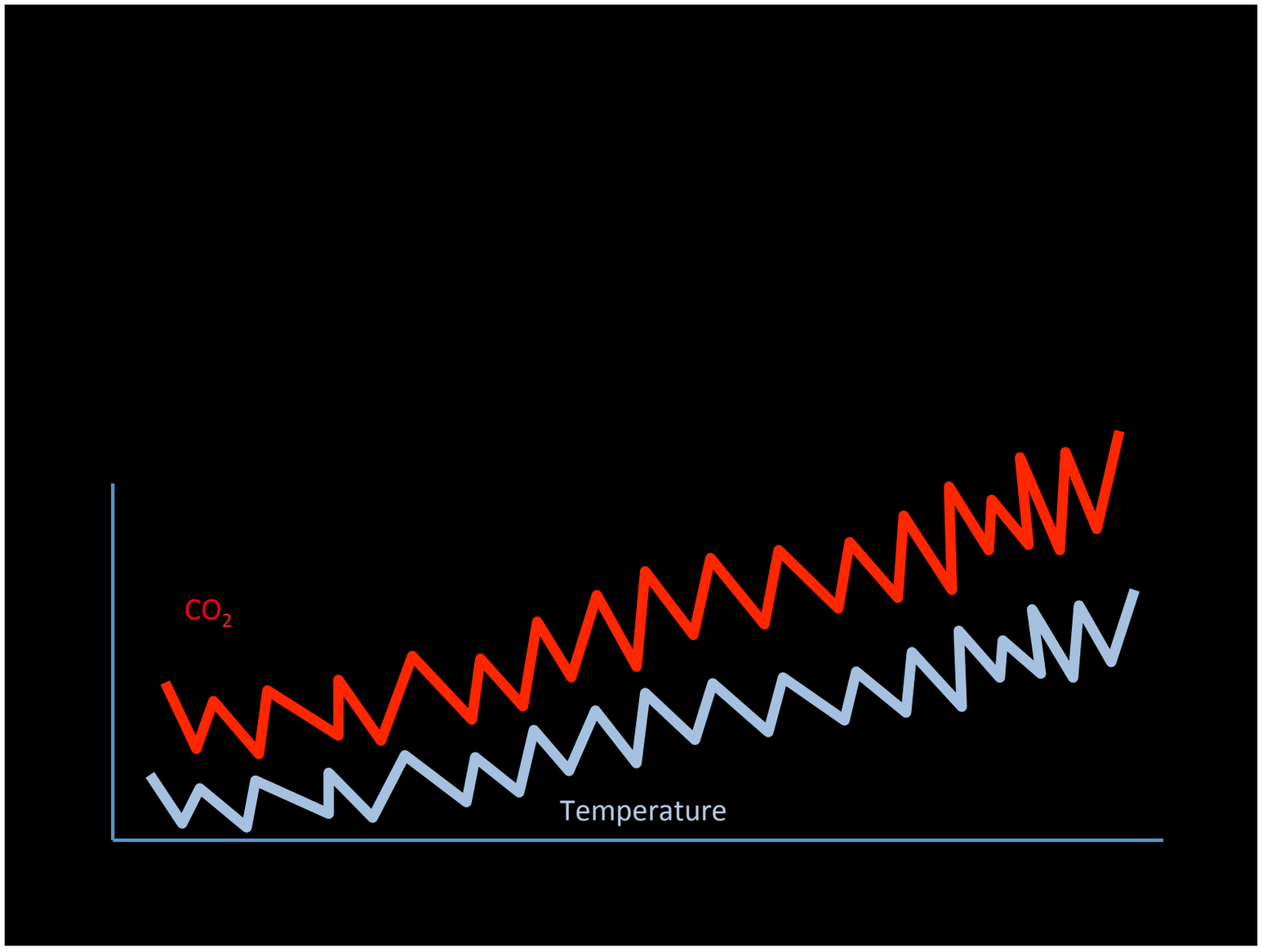
"The thing I've spent more time on… is trying to identify all those things in peoples minds that serve as obstacles to them understanding"

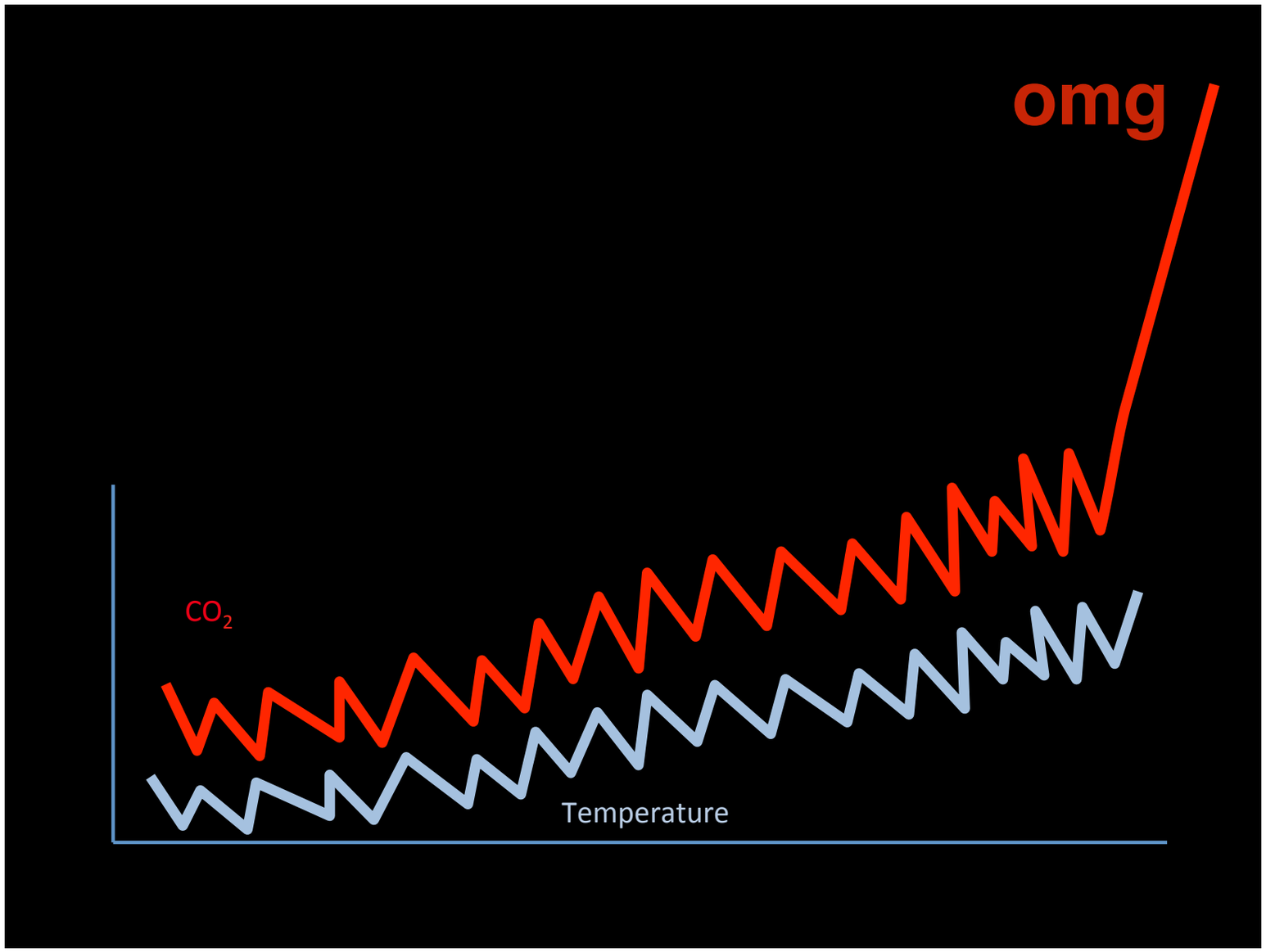
- Al Gore

jradavenport

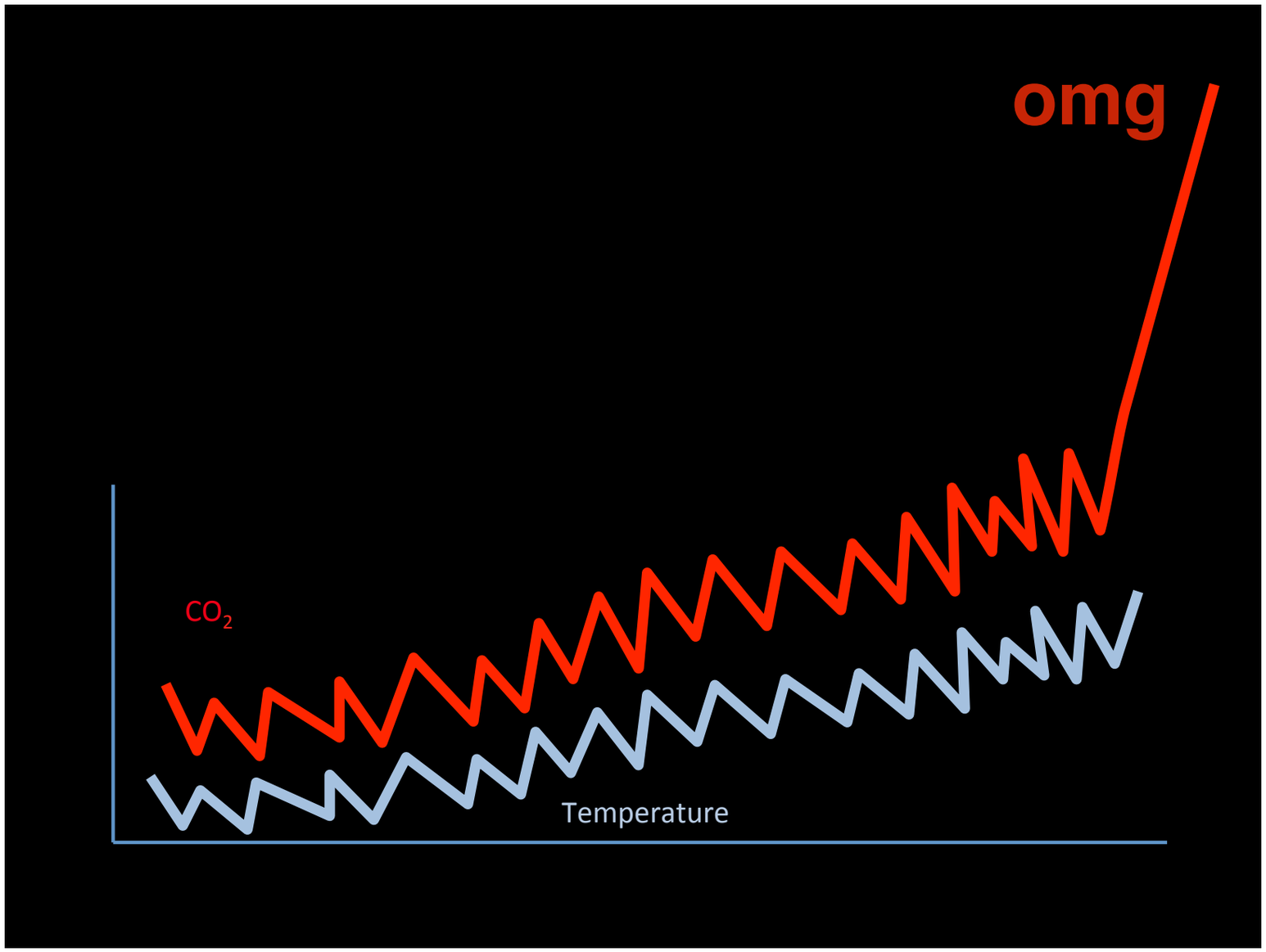



CO_2


CO_2
Temperature


omg
CO_2
Temperature

A great visualization: sells the message, worth 1000 words (or $1.2 Trillion annually)



What I *won't* cover:

- how to make graphs
- types of graphs/plots
- graphing languages

What I *want to* cover:

- why we make plots
- why we publish plots
- 5 points to consider

Why do we make plots?

a.k.a. charts, graphs, visualizations, figures, diagrams, maps…

jradavenport

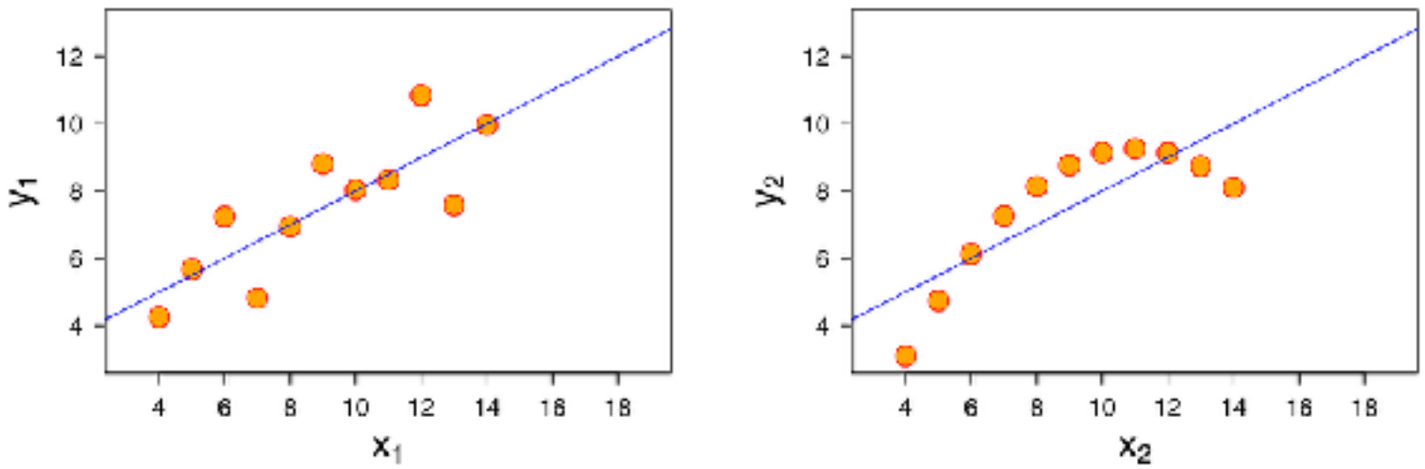
Why do <u>we</u> make plots?

- **Development**
  - data exploration
  - idea generation
  - debug code
- **Presentation**
  - talks
  - posters
  - papers

Each use/format has unique concerns!

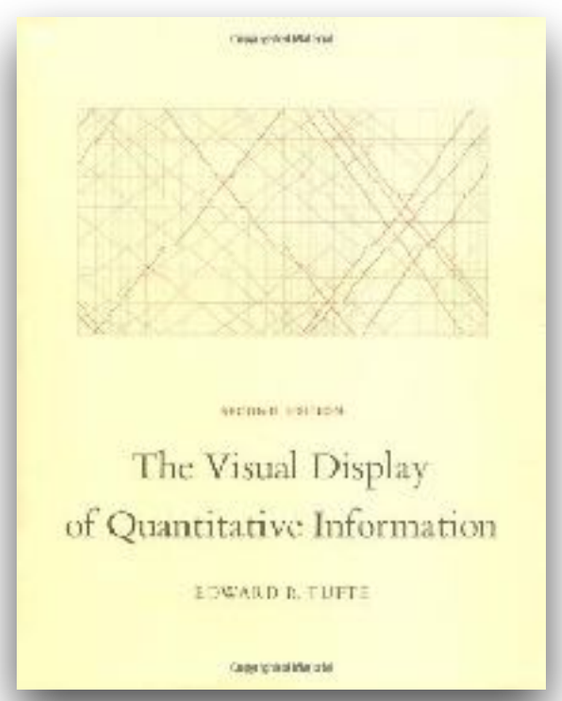
Development: Exploratory Data Analysis

Anscombe's Quartet



All panels have the same

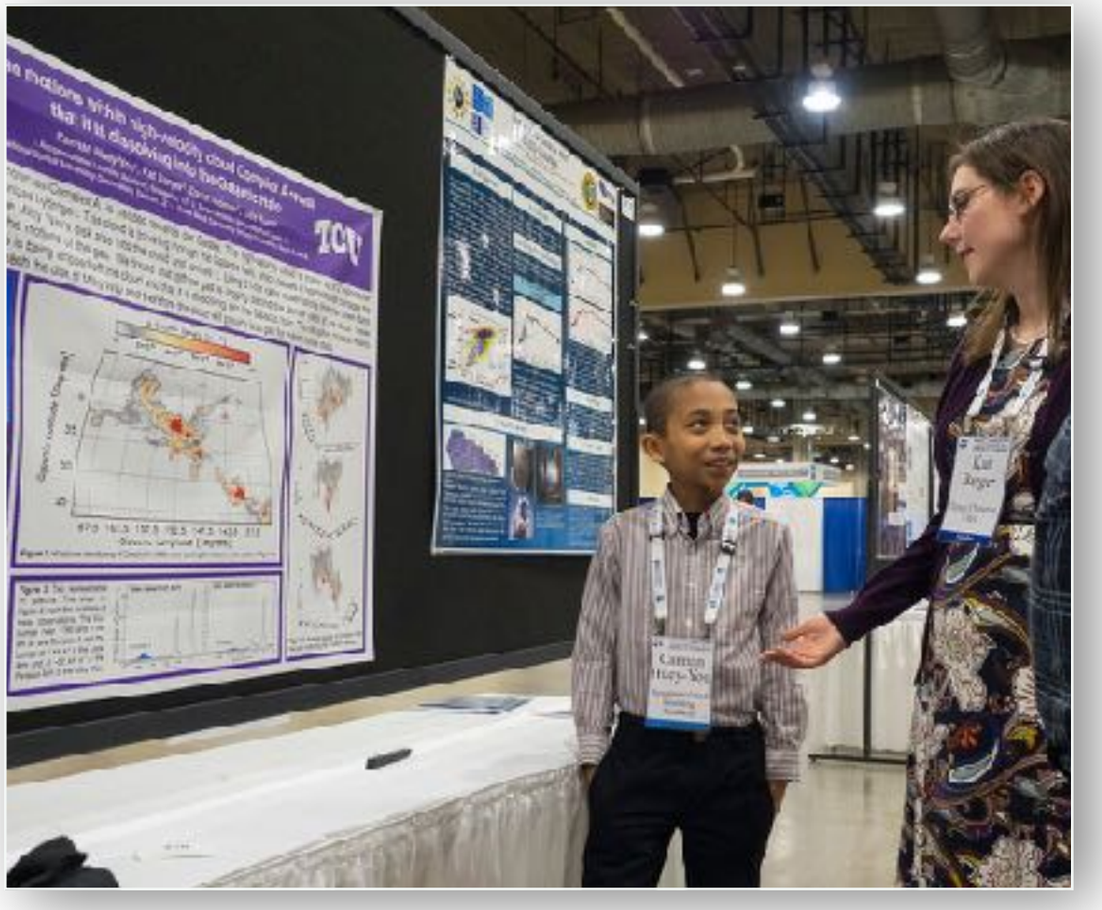
- mean (in X and Y)
- variance (in X and Y)
- linear regression
- "r" coefficient



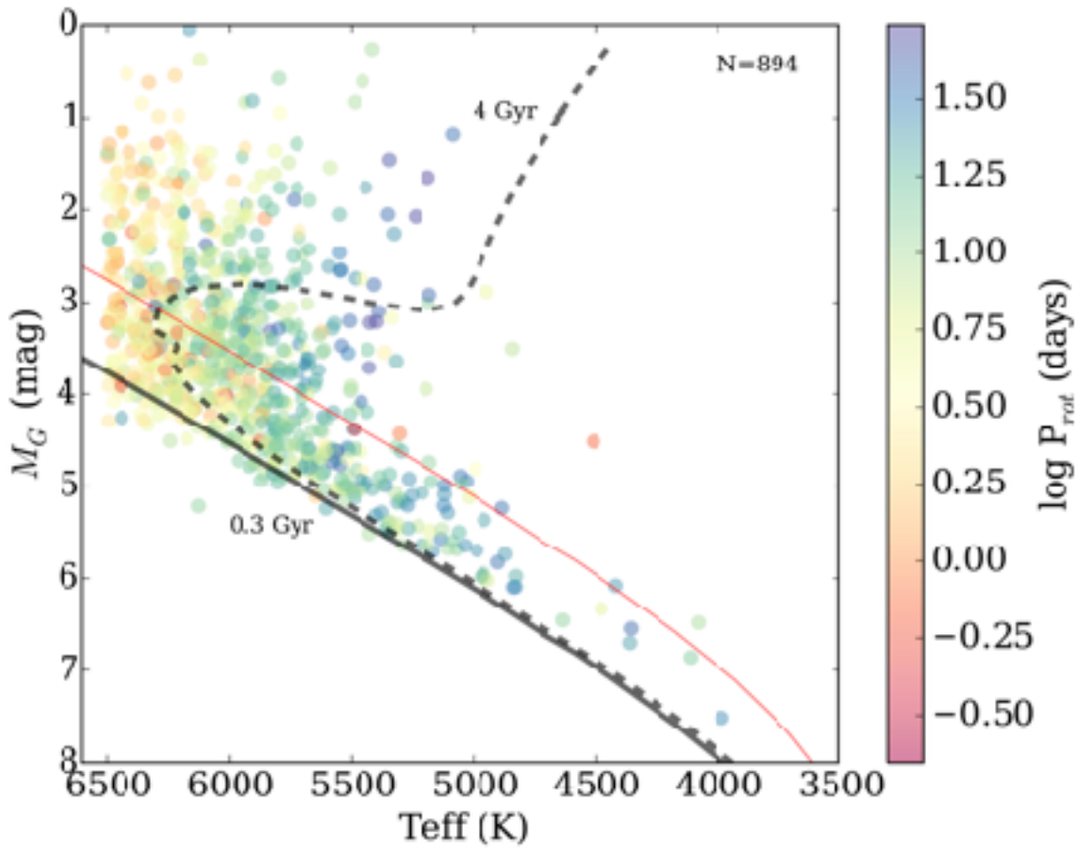
made popular by Tufte

jradavenport

Presentation: Talks & Posters



A Tale of One Plot



jradavenport

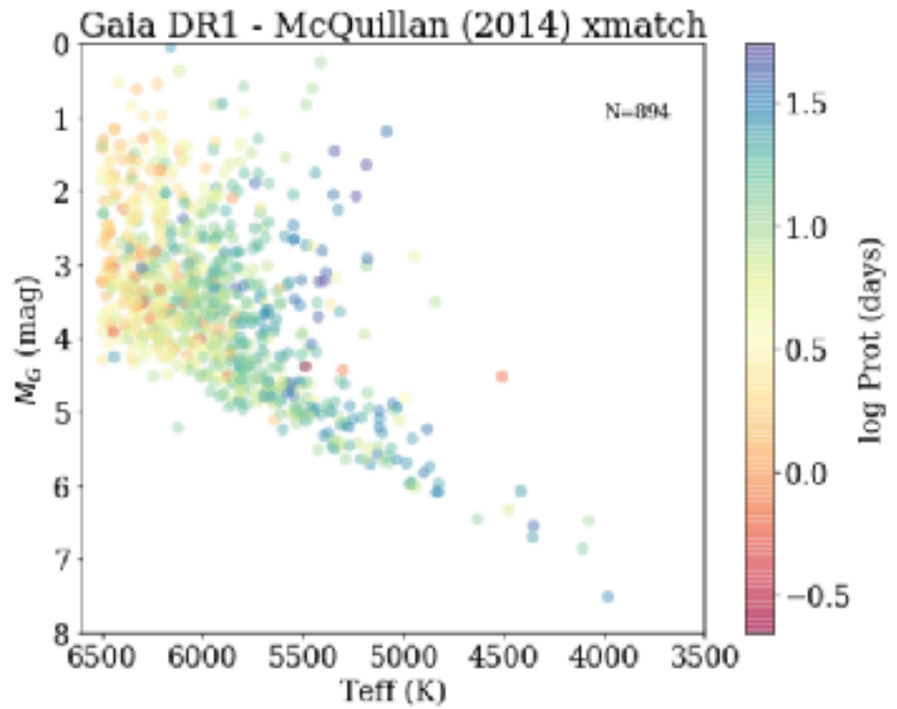
Exploration

In [15]:
```
plt.figure(figsize=(9,7))
cm = plt.cm.get_cmap('Spectral')

plt.scatter(gaia[u'Teff'].values[ok], MG.values[ok],
            c=np.log10(gaia[u'Prot'].values[ok]),
            lw=0, alpha=0.45, cmap=cm, s=60)

plt.xlim(6600, 3500)
plt.ylim(8, 0)
plt.xlabel('Teff (K)')
plt.ylabel('$M_G$ (mag)')
plt.title('Gaia DR1 - McQuillan (2014) xmatch')
plt.text(4000, 1, 'N=894', fontsize=12)

cb = plt.colorbar()
cb.set_label('log Prot (days)')
```



In [16]:
```
# let's put an isochrone down on that figure...
iso_file1 = 'output661464910329.dat' # Padova isochrone, 0.3gyr
iso_file2 = 'output223087321109.dat' # 1.0Gyr
iso_file3 = 'output816155020438.dat' # 4Gyr
```

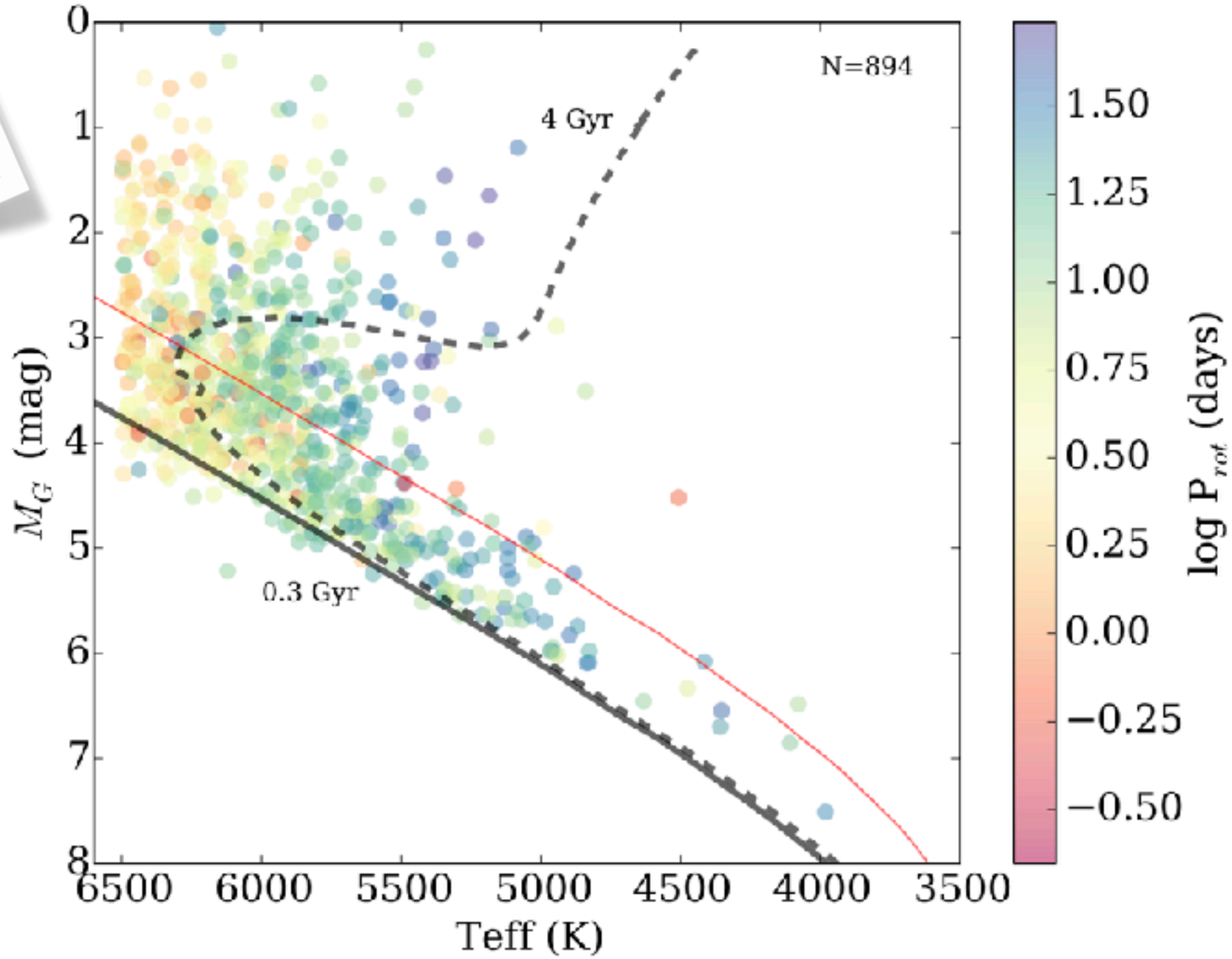
In [17]:
```
names=('Z','log(age)','M_ini', 'M_act','logL/Lo','logTe','logG','mbol','G','G_BP','G_RP','int_IMF'
,'stage')
iso1 = pd.read_table(iso_file1, delim_whitespace=True, header=None, comment='#', names=names)
iso2 = pd.read_table(iso_file2, delim_whitespace=True, header=None, comment='#', names=names)
iso3 = pd.read_table(iso_file3, delim_whitespace=True, header=None, comment='#', names=names)
```

In [18]:
```
plt.figure(figsize=(9,7))
cm = plt.cm.get_cmap('Spectral')

plt.scatter(gaia[u'Teff'].values[ok], MG.values[ok],
            c=np.log10(gaia[u'Prot'].values[ok]),
            lw=0, alpha=0.5, cmap=cm, s=60, rasterized=False)
```

jradavenport

Finished Product



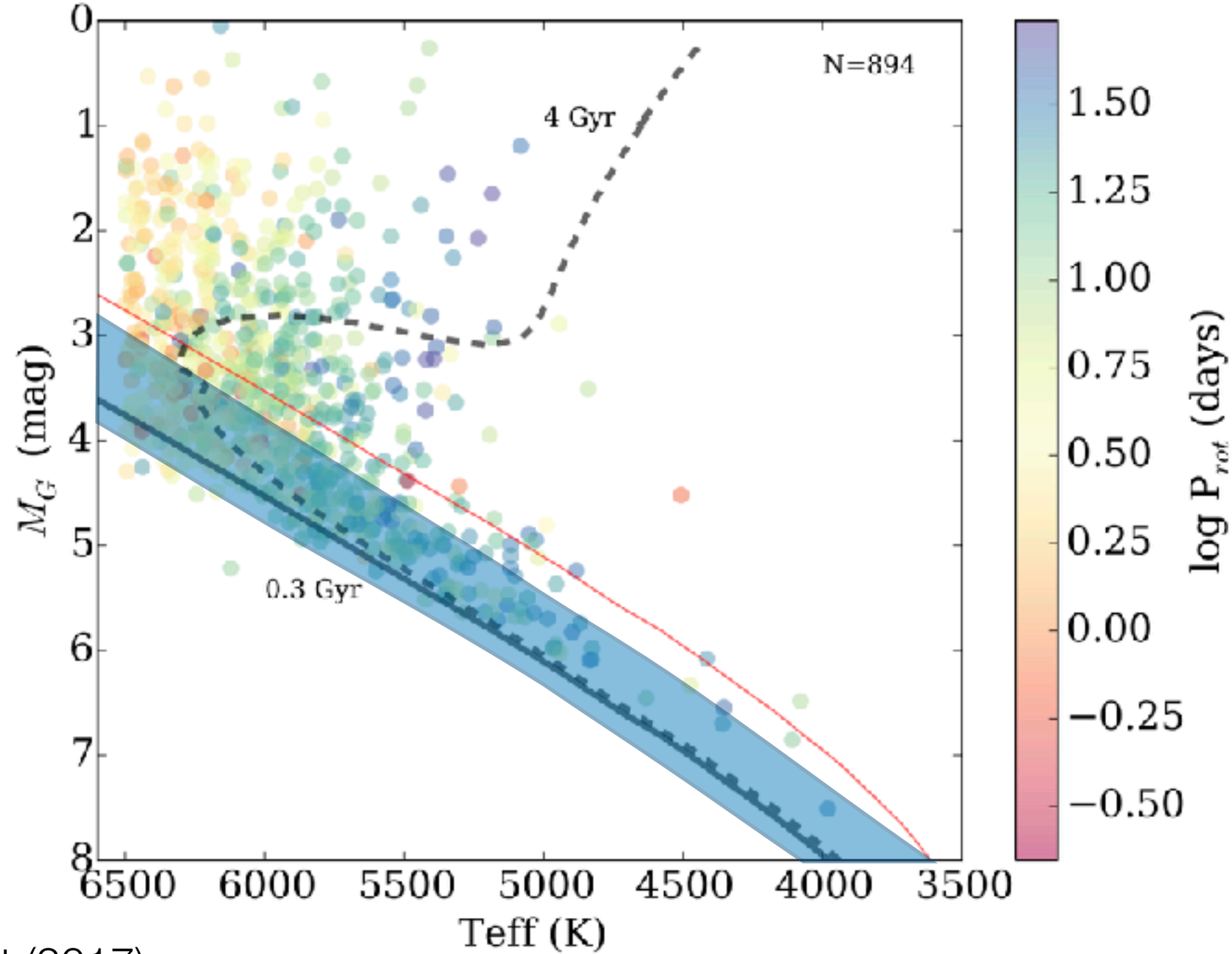
jradavenport

Match Kepler to Gaia (DR1/TGAS)

Select Main Sequence

Filter out “junk”

Lecture Slide



Davenport (2017)

jradavenport

Poster

Rotating stars from *Kepler* observed with *Gaia* DR1

James R. A. Davenport
NSF Postdoctoral Fellow
Western Washington University

Executive Summary – A bimodal rotation period distribution, discovered in *Kepler* by McQuillan et al. (2013) for M dwarfs, is shown to extend through G and K dwarfs. The bimodality was previously undetected for hotter stars due to subgiant contamination, which is eliminated using distances from *Gaia* DR1 (Lindegren et al. 2016). Two explanations for this bimodality have been suggested: variation in the local star formation history, and a new rapid dynamo phase transition. The former scenario is preferred by this data.

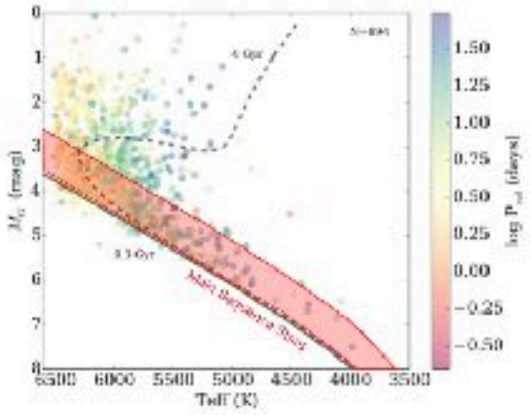


Figure 1 – HR Diagram for 894 *Kepler* rotating stars from McQuillan et al. (2014) passing photometric & parallax quality cuts from Gaia, colored by the *Kepler* rotation period. Two Bressan et al. (2012) isochrones are shown for reference (black lines). 440 Main Sequence stars are selected within the shaded region. This data shows significant contamination of subgiants was present in the *Kepler* rotation period sample.

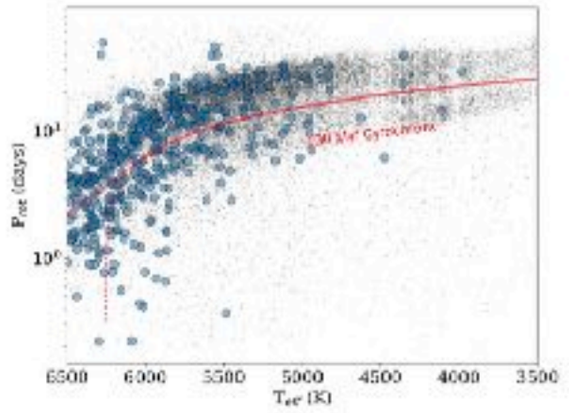


Figure 2 – Rotation period distribution for 33,855 *Kepler* stars from McQuillan et al. (2014) with detections in *Gaia* DR1 (black dots), and 440 "main sequence" stars from **Figure 1** (blue circles). The period bimodality, seen clearly for *Kepler* stars with T_{eff} < 4000 K, extends to T_{eff} ~ 6000 K in the subgiant-free sample. A Meibom et al. (2011) 600 Myr gyrochrone roughly traces the bimodality.

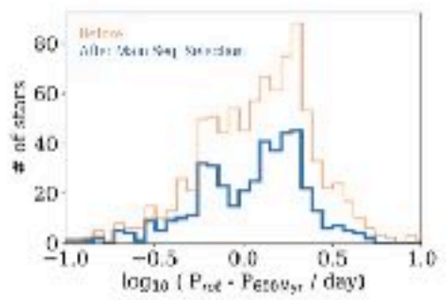


Figure 3 – log rotation period distribution for stars from **Figure 1** centered around the 600 Myr gyrochrone in **Figure 2**, before (orange) and after (blue) the main sequence selection. A bimodal period distribution is revealed for these hotter stars, suggesting the feature is due to variations in local star formation history

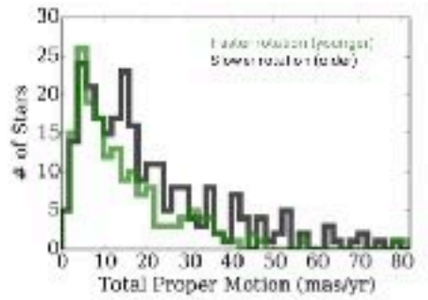


Figure 4 – Total Gaia proper motions for the faster (green) and slower (black) rotating populations. The two rotation period samples appear to be drawn from different kinematic populations, further indicating a stellar age origin for the bimodality, rather than a rapid dynamo transition at intermediate ages.

References
Bressan, A., et al. 2012, MNRAS, 427, 127
Lindegren, L., et al. 2016, A&A, 595, A
McQuillan, A., et al. 2013, MNRAS, 432, 1203
McQuillan, A., et al. 2014, ApJS, 211, 24
Meibom, S., et al. 2011, ApJL, 733, L9

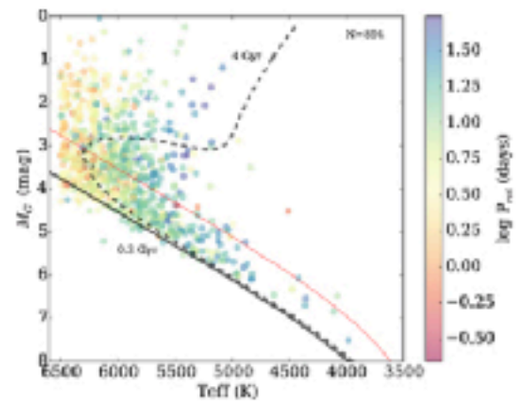
jradavenport

arxiv.org/abs/1610.08563

jradavenport

ApJ Webpage

distances, and therefore luminosities, parallaxes were required to have errors <0.4 mas. These cuts left a total of 894 stars from the *Kepler*–TGAS-matched sample (68%). The Hertzsprung–Russell (H-R) diagram for these stars is shown in Figure 2, with each point colored by its *Kepler*-measured rotation period. Example isochrones from the grid of Bressan et al. (2012) are shown for two ages. A systematic offset of ~0.5 mag is found between the measured absolute G-band and the isochrone's main sequence. This offset is likely due to calibration differences between the nominal and actual G-band (A. Brown 2016, private communication).



Zoom In Zoom Out Reset image size

Figure 2. Hertzsprung–Russell (H-R) diagram using temperatures from McQuillan et al. (2014) and *Gaia* DR1 G-band absolute magnitudes for the 894 stars that pass cuts on photometric and parallax quality described in the text. Points are colored by their measured *Kepler* rotation periods. Two isochrones from Bressan et al. (2012) are shown, with ages of 300 Myr and 4 Gyr (solid and dashed lines). Main-sequence stars are selected as those between the 300 Myr isochrone and the isochrone shifted up by 1 mag (red line).

Download figure:

Standard Image High-resolution Image Export PowerPoint slide

The H-R–period diagram in Figure 2 shows stars with a range of evolutionary states, and could help to test models of the evolution of post-main-sequence angular momentum (e.g., do Nascimento et al. 2012). Outliers in this diagram are either due to erroneous cross-matching in the *Kepler* and *Gaia* catalogs or represent interesting systems such as rare binary star configurations or stars that have undergone mergers or ingested giant planets (Massarotti 2008; Tayar et al. 2015). For example, examination of the 2MASS (Skrutskie et al. 2006) image on SIMBAD for the rapidly rotating star at $T_{\rm eff} = 4500$ K, $M_G = 4.5$ mag, $P_{\rm rot} = 0.652$ day, in Figure 2 (KIC 07957709) shows that this source appears to be highly contaminated with three point sources clustered within ~10 arcsec, likely leading to an erroneous position in the H-R diagram, and possibly incorrect variability measurements with *Kepler*. Investigating all such outliers in Figure 2 is beyond the scope of this work, but these targets are worth further study because they may reveal new physics.

Back to top

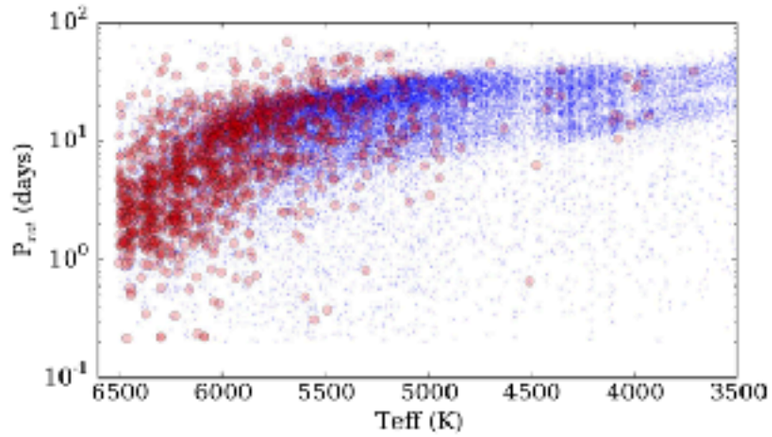


Figure 1. Rotation period distribution for 33,855 *Kepler* stars from McQuillan et al. (2014) with detections in *Gaia* DR1 (blue dots). The period bimodality can be seen most clearly for stars with $T_{\rm eff} < 4000$ K as a dearth of sources with periods of ~25 days, but extends to at least $T_{\rm eff} \sim 5500$ K according to McQuillan et al. (2014). The subsample of 1299 nearby objects found in TGAS are highlighted (red circles), and are mostly hotter stars due to the faint limit of the TGAS sample.

sources by extending the astrometric solutions from Tycho and *Hipparcos*. While the TGAS data are not a complete astrometric survey, and have possible systematics in the reported parallaxes (Lindegren et al. 2016; Stassun & Torres 2016), they represent a significant improvement in the astrometry and kinematics available for stars in the *Kepler* field.

Using the CDS X-Match service, I cross-matched the available catalogs from these two surveys. A default cross-match radius of 5 arcsec was used. A total of 33,855 stars were found in the cross match between these catalogs, which is 99.5% of the sample from McQuillan et al. (2014). The small number of stars not recovered from McQuillan et al. (2014) may be missed because of source confusion within the matching radius. A subset of 1299 objects were recovered in the TGAS sample. Very few K and M dwarfs were recovered in that sample because of its brightness limits. Future releases of *Gaia* data will provide full astrometric solutions for nearly all *Kepler* stars. The rotation periods versus stellar effective temperatures for the *Kepler–Gaia* matched stars are shown in Figure 1.

3. SELECTING MAIN-SEQUENCE STARS

Though McQuillan et al. (2014) attempted to measure periods only for dwarf stars, the sample of *Kepler–Gaia* matched stars contains both main-sequence dwarfs and evolved stars (giants and subgiants). Previous studies have shown that significant contamination by giants or subgiants can affect the implied variability properties of dwarf stars (Ciardi et al. 2011; Mann et al. 2012). Therefore to properly understand the nature of the period distribution and its implications for age-dating field stars, a robust sample of main-sequence stars must be selected.

Faint stars were removed by requiring sources to have *G*-band flux errors <1%. To ensure accurate distances, and therefore luminosities, parallaxes were required to have errors <0.4 mas. These cuts left a total of 894 stars from the *Kepler*–TGAS-matched sample (68%). The Hertzsprung–Russell (H-R) diagram for these stars is shown in Figure 2, with each point colored by its *Kepler*-measured rotation period. Example isochrones from the grid of Bressan et al. (2012) are shown

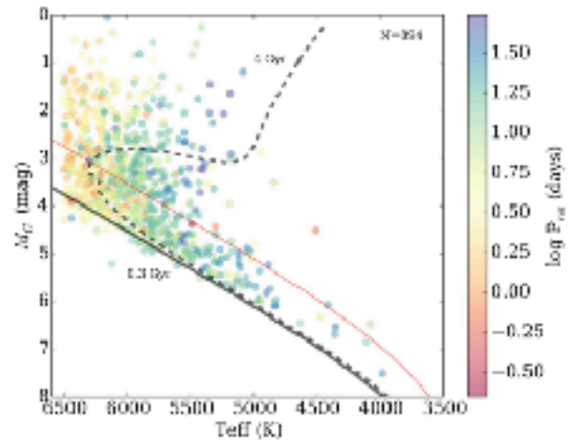


Figure 2. Hertzsprung–Russell (H-R) diagram using temperatures from McQuillan et al. (2014) and *Gaia* DR1 *G*-band absolute magnitudes for the 894 stars that pass cuts on photometric and parallax quality described in the text. Points are colored by their measured *Kepler* rotation periods. Two isochrones from Bressan et al. (2012) are shown, with ages of 300 Myr and 4 Gyr (solid and dashed lines). Main-sequence stars are selected as those between the 300 Myr isochrone and the isochrone shifted up by 1 mag (red line).

for two ages. A systematic offset of ~0.5 mag is found between the measured absolute *G*-band and the isochrone's main sequence. This offset is likely due to calibration differences between the nominal and actual *G*-band (A. Brown 2016, private communication).

The H-R–period diagram in Figure 2 shows stars with a range of evolutionary states, and could help to test models of the evolution of post-main-sequence angular momentum (e.g., do Nascimento et al. 2012). Outliers in this diagram are either due to erroneous cross-matching in the *Kepler* and *Gaia* catalogs or represent interesting systems such as rare binary star configurations or stars that have undergone mergers or ingested giant planets (Massarotti 2008; Tayar et al. 2015). For example, examination of the 2MASS (Skrutskie et al. 2006) image on SIMBAD for the rapidly rotating star at $T_{\rm eff} = 4500$ K, $M_G = 4.5$ mag, $P_{\rm rot} = 0.652$ day, in Figure 2 (KIC 07957709) shows

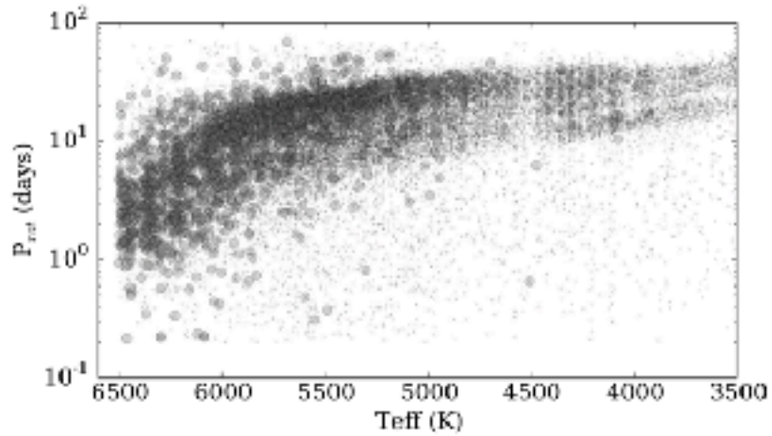


Figure 1. Rotation period distribution for 33,855 *Kepler* stars from McQuillan et al. (2014) with detections in *Gaia* DR1 (blue dots). The period bimodality can be seen most clearly for stars with $T_{\rm eff} < 4000$ K as a dearth of sources with periods of ~25 days, but extends to at least $T_{\rm eff} \sim 5500$ K according to McQuillan et al. (2014). The subsample of 1299 nearby objects found in TGAS are highlighted (red circles), and are mostly hotter stars due to the faint limit of the TGAS sample.

sources by extending the astrometric solutions from Tycho and *Hipparcos*. While the TGAS data are not a complete astrometric survey, and have possible systematics in the reported parallaxes (Lindegren et al. 2016; Stassun & Torres 2016), they represent a significant improvement in the astrometry and kinematics available for stars in the *Kepler* field.

Using the CDS X-Match service, I cross-matched the available catalogs from these two surveys. A default cross-match radius of 5 arcsec was used. A total of 33,855 stars were found in the cross match between these catalogs, which is 99.5% of the sample from McQuillan et al. (2014). The small number of stars not recovered from McQuillan et al. (2014) may be missed because of source confusion within the matching radius. A subset of 1299 objects were recovered in the TGAS sample. Very few K and M dwarfs were recovered in that sample because of its brightness limits. Future releases of *Gaia* data will provide full astrometric solutions for nearly all *Kepler* stars. The rotation periods versus stellar effective temperatures for the *Kepler–Gaia* matched stars are shown in Figure 1.

3. SELECTING MAIN-SEQUENCE STARS

Though McQuillan et al. (2014) attempted to measure periods only for dwarf stars, the sample of *Kepler–Gaia* matched stars contains both main-sequence dwarfs and evolved stars (giants and subgiants). Previous studies have shown that significant contamination by giants or subgiants can affect the implied variability properties of dwarf stars (Ciardi et al. 2011; Mann et al. 2012). Therefore to properly understand the nature of the period distribution and its implications for age-dating field stars, a robust sample of main-sequence stars must be selected.

Faint stars were removed by requiring sources to have G-band flux errors <1%. To ensure accurate distances, and therefore luminosities, parallaxes were required to have errors <0.4 mas. These cuts left a total of 894 stars from the *Kepler–*TGAS-matched sample (68%). The Hertzsprung–Russell (H-R) diagram for these stars is shown in Figure 2, with each point colored by its *Kepler*-measured rotation period. Example isochrones from the grid of Bressan et al. (2012) are shown

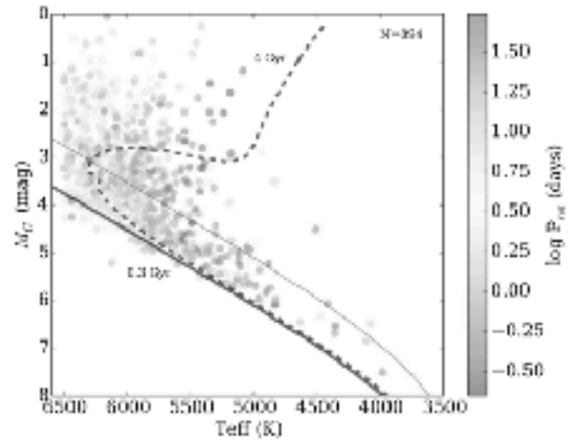


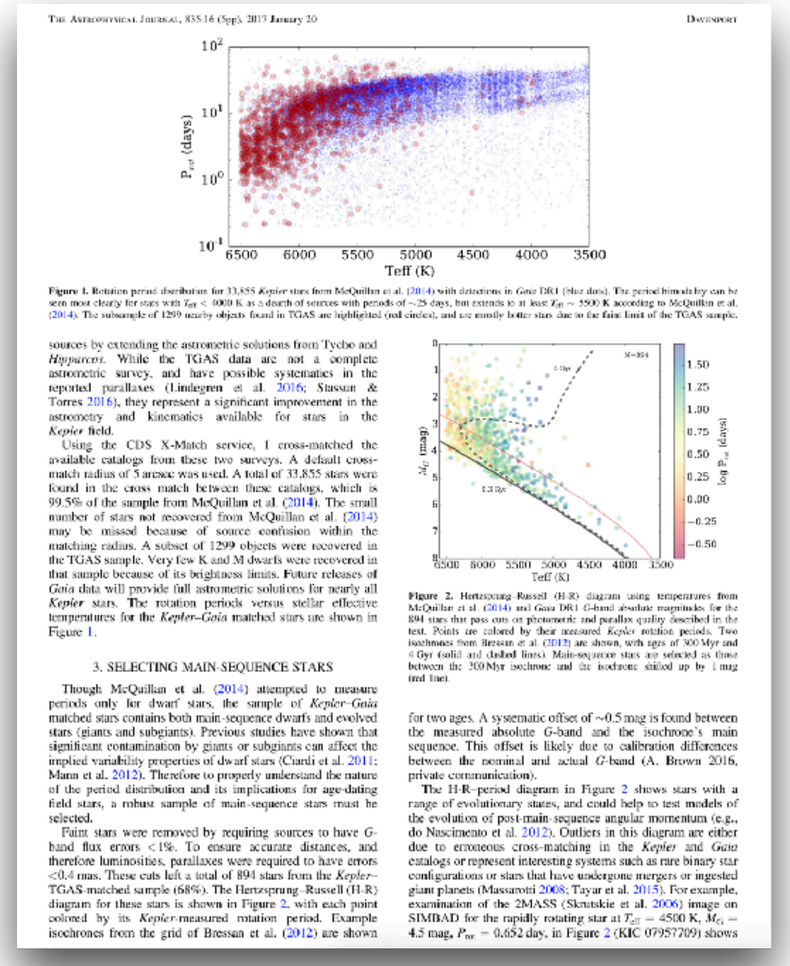
Figure 2. Hertzsprung–Russell (H-R) diagram using temperatures from McQuillan et al. (2014) and *Gaia* DR1 G-band absolute magnitudes for the 894 stars that pass cuts on photometric and parallax quality described in the text. Points are colored by their measured *Kepler* rotation periods. Two isochrones from Bressan et al. (2012) are shown, with ages of 300 Myr and 4 Gyr (solid and dashed lines). Main-sequence stars are selected as those between the 300 Myr isochrone and the isochrone shifted up by 1 mag (red line).

for two ages. A systematic offset of ~0.5 mag is found between the measured absolute G-band and the isochrone's main sequence. This offset is likely due to calibration differences between the nominal and actual G-band (A. Brown 2016, private communication).

The H-R–period diagram in Figure 2 shows stars with a range of evolutionary states, and could help to test models of the evolution of post-main-sequence angular momentum (e.g., do Nascimento et al. 2012). Outliers in this diagram are either due to erroneous cross-matching in the *Kepler* and *Gaia* catalogs or represent interesting systems such as rare binary star configurations or stars that have undergone mergers or ingested giant planets (Massarotti 2008; Tayar et al. 2015). For example, examination of the 2MASS (Skrutskie et al. 2006) image on SIMBAD for the rapidly rotating star at $T_{\rm eff} = 4500$ K, $M_G = 4.5$ mag, $P_{\rm rot} = 0.652$ day, in Figure 2 (KIC 07957709) shows

Point #1

Your plots (need to) tell the story at a glance!



The Astrophysical Journal, 835:16 (5pp), 2017 January 20 — Davenport

Figure 1. Rotation period distribution for 33,855 *Kepler* stars from McQuillan et al. (2014) with detections in *Gaia* DR1 (blue dots). The period bimodality can be seen most clearly for stars with $T_{\rm eff} < 4000$ K, as a dearth of sources with periods of ~25 days, but extends to at least $T_{\rm eff} \sim 5500$ K according to McQuillan et al. (2014). The subsample of 1299 nearby objects found in TGAS are highlighted (red circles), and are mostly hotter stars due to the faint limit of the TGAS sample.

sources by extending the astrometric solutions from Tycho and *Hipparcos*. While the TGAS data are not a complete astrometric survey, and have possible systematics in the reported parallaxes (Lindegren et al. 2016; Stassun & Torres 2016), they represent a significant improvement in the astrometry and kinematics available for stars in the *Kepler* field.

Using the CDS X-Match service, I cross-matched the available catalogs from these two surveys. A default cross-match radius of 5 arcsec was used. A total of 33,855 stars were found in the cross match between these catalogs, which is 99.5% of the sample from McQuillan et al. (2014). The small number of stars not recovered from McQuillan et al. (2014) may be missed because of source confusion within the matching radius. A subset of 1299 objects were recovered in the TGAS sample. Very few K and M dwarfs were recovered in that sample because of its brightness limits. Future releases of *Gaia* data will provide full astrometric solutions for nearly all *Kepler* stars. The rotation periods versus stellar effective temperatures for the *Kepler–Gaia* matched stars are shown in Figure 1.

Figure 2. Hertzsprung–Russell (H-R) diagram using temperatures from McQuillan et al. (2014) and *Gaia* DR1 *G*-band absolute magnitudes for the 894 stars that pass cuts on photometric and parallax quality described in the text. Points are colored by their measured *Kepler* rotation periods. Two isochrones from Bressan et al. (2012) are shown, with ages of 300 Myr and 6 Gyr (solid and dashed lines). Main-sequence stars are selected as those between the 300 Myr isochrone and the isochrone shifted up by 1 mag (red line).

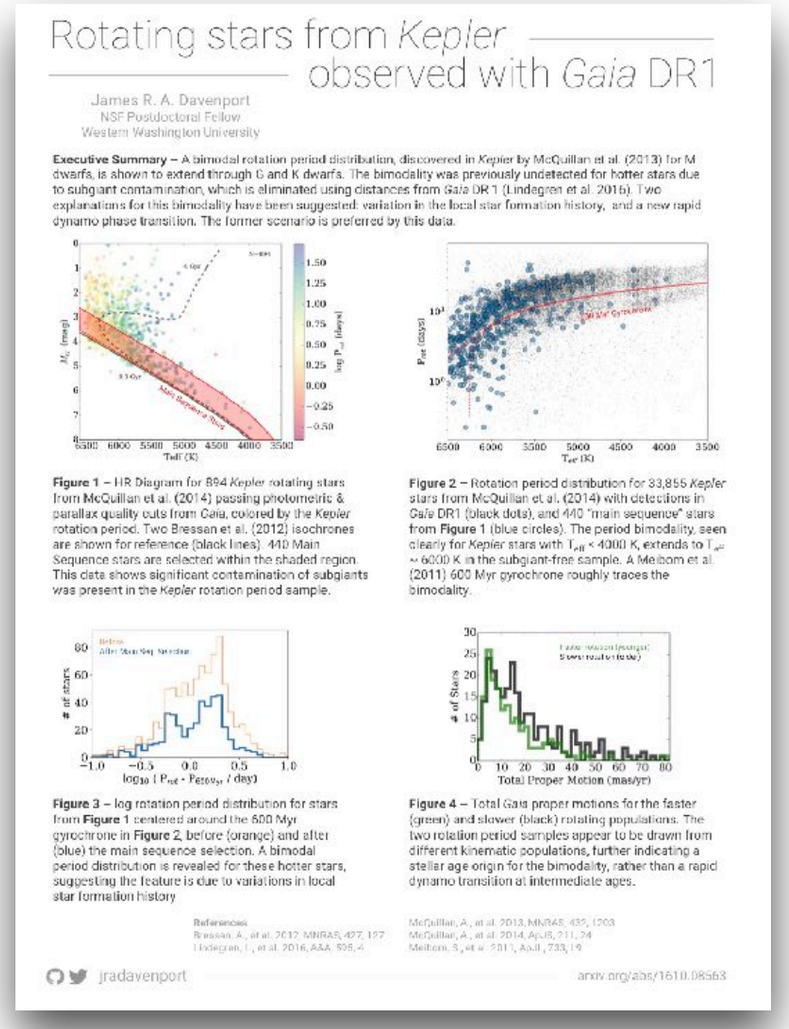
3. SELECTING MAIN-SEQUENCE STARS

Though McQuillan et al. (2014) attempted to measure periods only for dwarf stars, the sample of *Kepler–Gaia* matched stars contains both main-sequence dwarfs and evolved stars (giants and subgiants). Previous studies have shown that significant contamination by giants or subgiants can affect the implied variability properties of dwarf stars (Ciardi et al. 2011; Mann et al. 2012). Therefore to properly understand the nature of the period distribution and its implications for age-dating field stars, a robust sample of main-sequence stars must be selected.

Faint stars were removed by requiring sources to have *G*-band flux errors <1%. To ensure accurate distances, and therefore luminosities, parallaxes were required to have errors <0.4 mas. These cuts left a total of 894 stars from the *Kepler*–TGAS-matched sample (68%). The Hertzsprung–Russell (H-R) diagram for these stars is shown in Figure 2, with each point colored by its *Kepler*-measured rotation period. Example isochrones from the grid of Bressan et al. (2012) are shown for two ages. A systematic offset of ~0.5 mag is found between the measured absolute *G*-band and the isochrone's main sequence. This offset is likely due to calibration differences between the nominal and actual *G*-band (A. Brown 2016, private communication).

The H-R–period diagram in Figure 2 shows stars with a range of evolutionary states, and could help to test models of the evolution of post-main-sequence angular momentum (e.g., do Nascimento et al. 2012). Outliers in this diagram are either due to erroneous cross-matching in the *Kepler* and *Gaia* catalogs or represent interesting systems such as rare binary star configurations or stars that have undergone mergers or ingested giant planets (Massarotti 2008; Tayar et al. 2015). For example, examination of the 2MASS (Skrutskie et al. 2006) image on SIMBAD for the rapidly rotating star at $T_{\rm eff} = 4500$ K, $M_G = 4.5$ mag, $P_{\rm rot} = 0.652$ day, in Figure 2 (KIC 07957709) shows

paper



Rotating stars from *Kepler* observed with *Gaia* DR1

James R. A. Davenport
NSF Postdoctoral Fellow
Western Washington University

Executive Summary – A bimodal rotation period distribution, discovered in *Kepler* by McQuillan et al. (2013) for M dwarfs, is shown to extend through G and K dwarfs. The bimodality was previously undetected for hotter stars due to subgiant contamination, which is eliminated using distances from *Gaia* DR1 (Lindegren et al. 2016b). Two explanations for this bimodality have been suggested: variation in the local star formation history, and a new rapid dynamo phase transition. The former scenario is preferred by this data.

Figure 1 – HR Diagram for 894 *Kepler* rotating stars from McQuillan et al. (2014) passing photometric & parallax quality cuts from Gaia, colored by the *Kepler* rotation period. Two Bressan et al. (2012) isochrones are shown for reference (black lines). 440 Main Sequence stars are selected within the shaded region. This data shows significant contamination of subgiants was present in the *Kepler* rotation period sample.

Figure 2 – Rotation period distribution for 33,855 *Kepler* stars from McQuillan et al. (2014) with detections in *Gaia* DR1 (black dots), and 440 "main sequence" stars from **Figure 1** (blue circles). The period bimodality, seen clearly for *Kepler* stars with $T_{\rm eff} < 4000$ K, extends to $T_{\rm eff} \sim 6000$ K in the subgiant-free sample. A Meibom et al. (2011) 600 Myr gyrochrone roughly traces the bimodality.

Figure 3 – log rotation period distribution for stars from **Figure 1** centered around the 600 Myr gyrochrone in **Figure 2**, before (orange) and after (blue) the main sequence selection. A bimodal period distribution is revealed for these hotter stars, suggesting the feature is due to variations in local star formation history

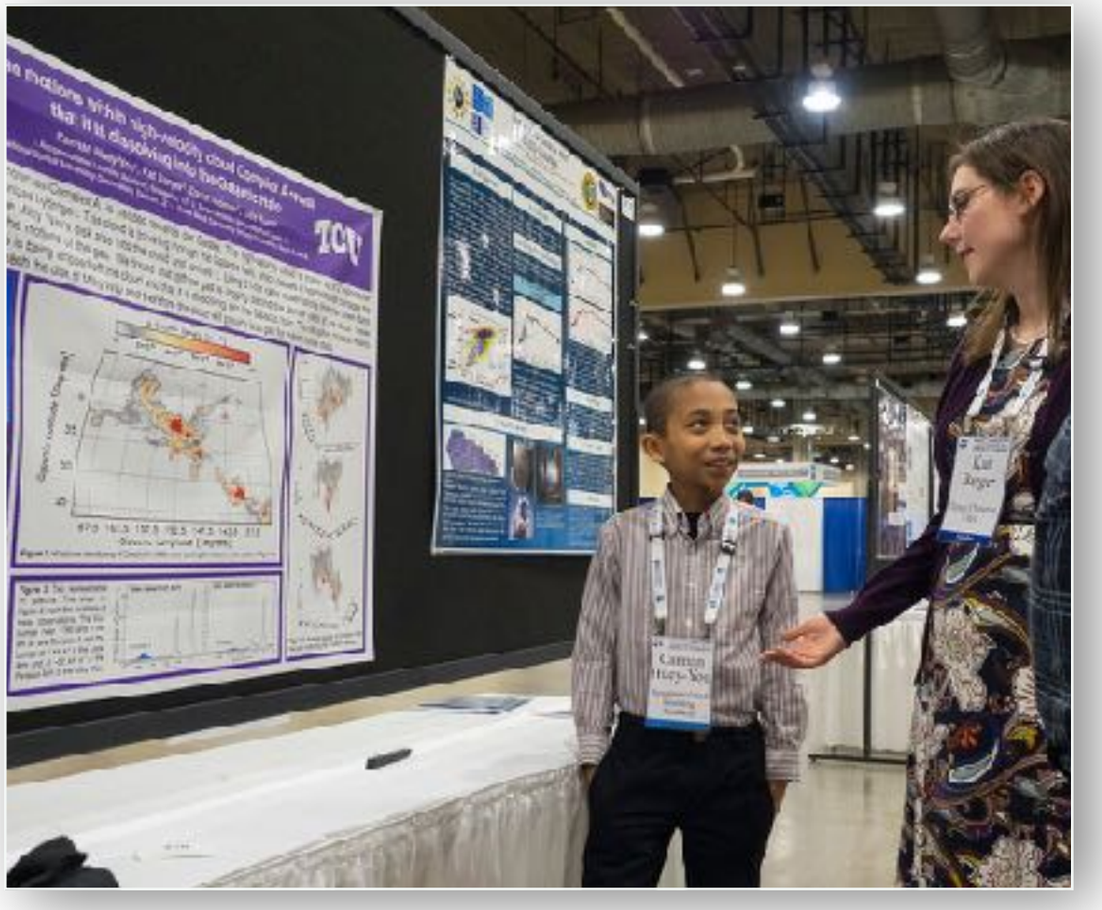
Figure 4 – Total Gaia proper motions for the faster (green) and slower (black) rotating populations. The two rotation period samples appear to be drawn from different kinematic populations, further indicating a stellar age origin for the bimodality, rather than a rapid dynamo transition at intermediate ages.

References
Bressan, A., et al. 2012, MNRAS, 427, 127
Lindegren, L., et al. 2016, A&A, 595, A
McQuillan, A., et al. 2013, MNRAS, 435, 1203
McQuillan, A., et al. 2014, ApJS, 211, 24
Meibom, S., et al. 2011, ApJ, 733, 115

poster

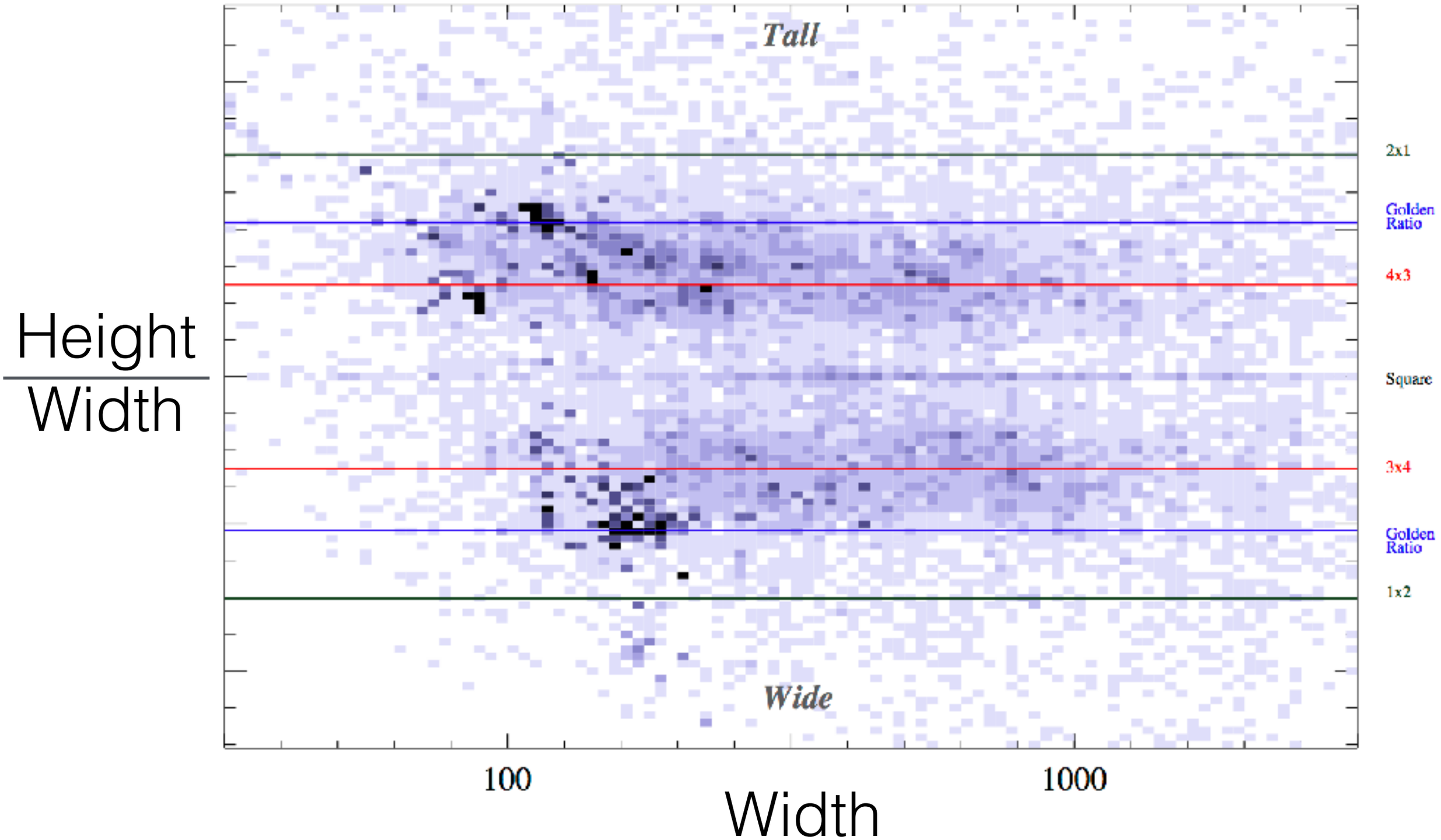
jradavenport

One plot does not work for all audiences/needs



The Dimensions of Art

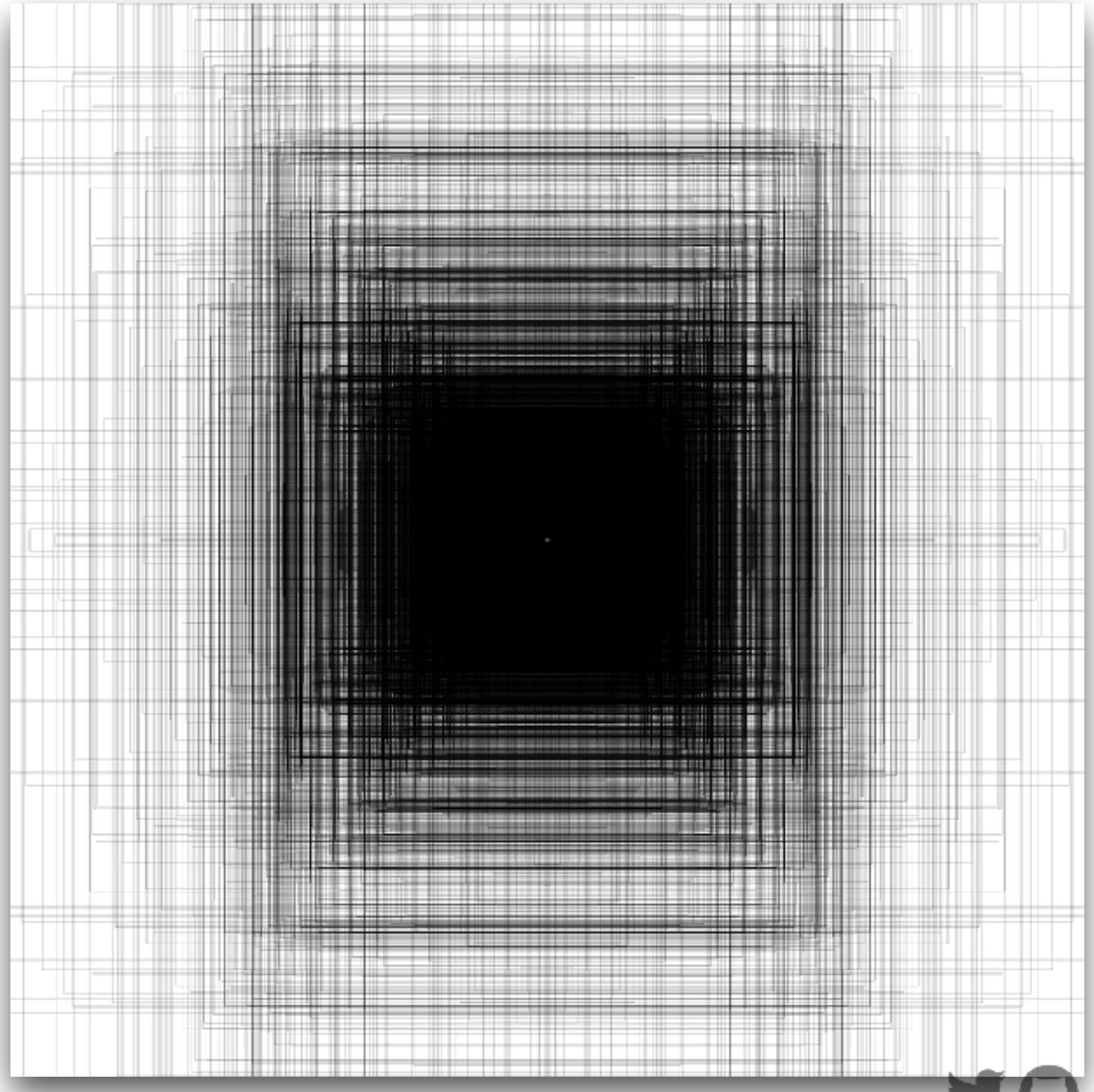
65,000 pieces of art from the Tate Modern



jradavenport

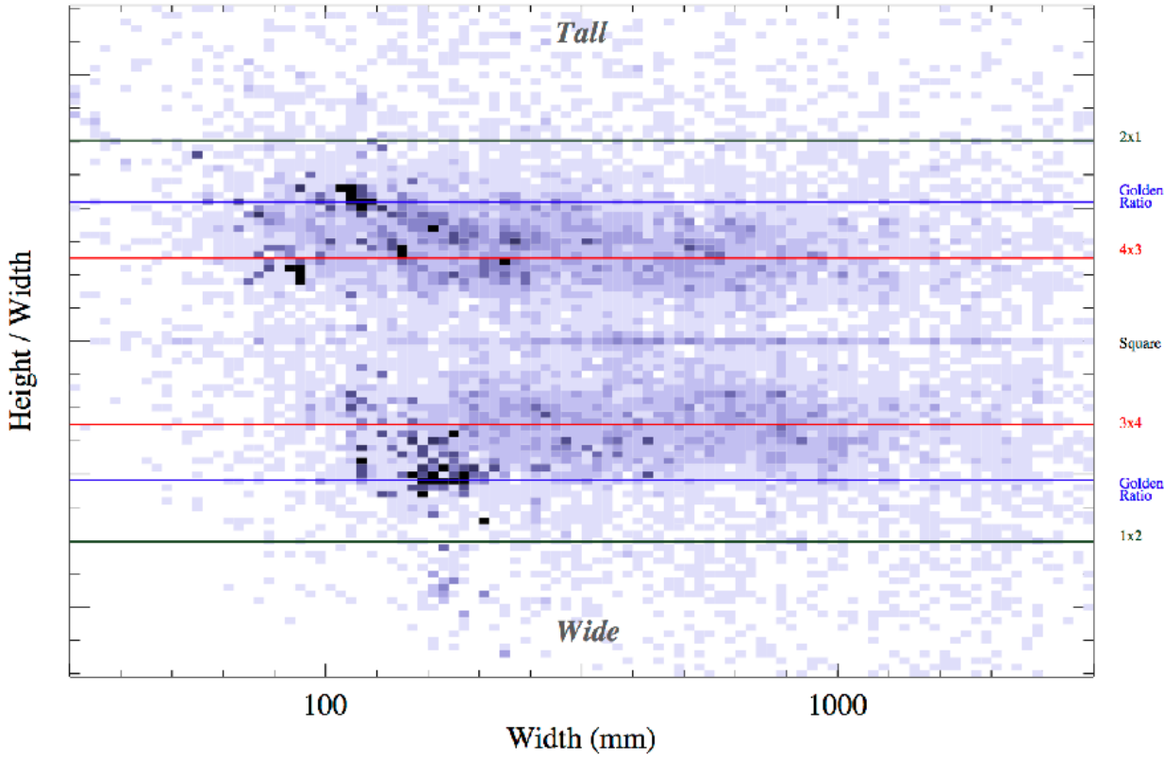
The Dimensions of Art

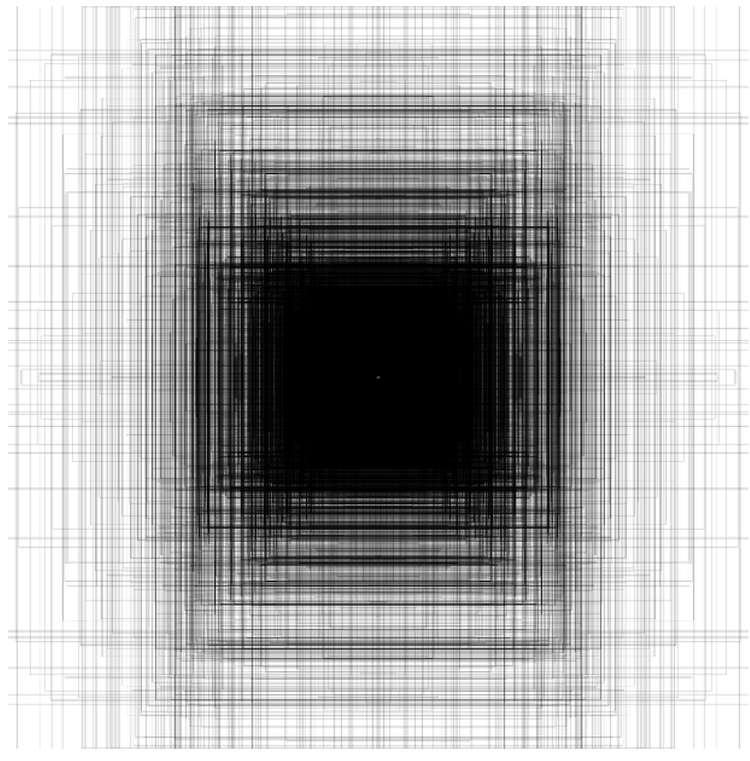
65,000 pieces of art from the Tate Modern



jradavenport

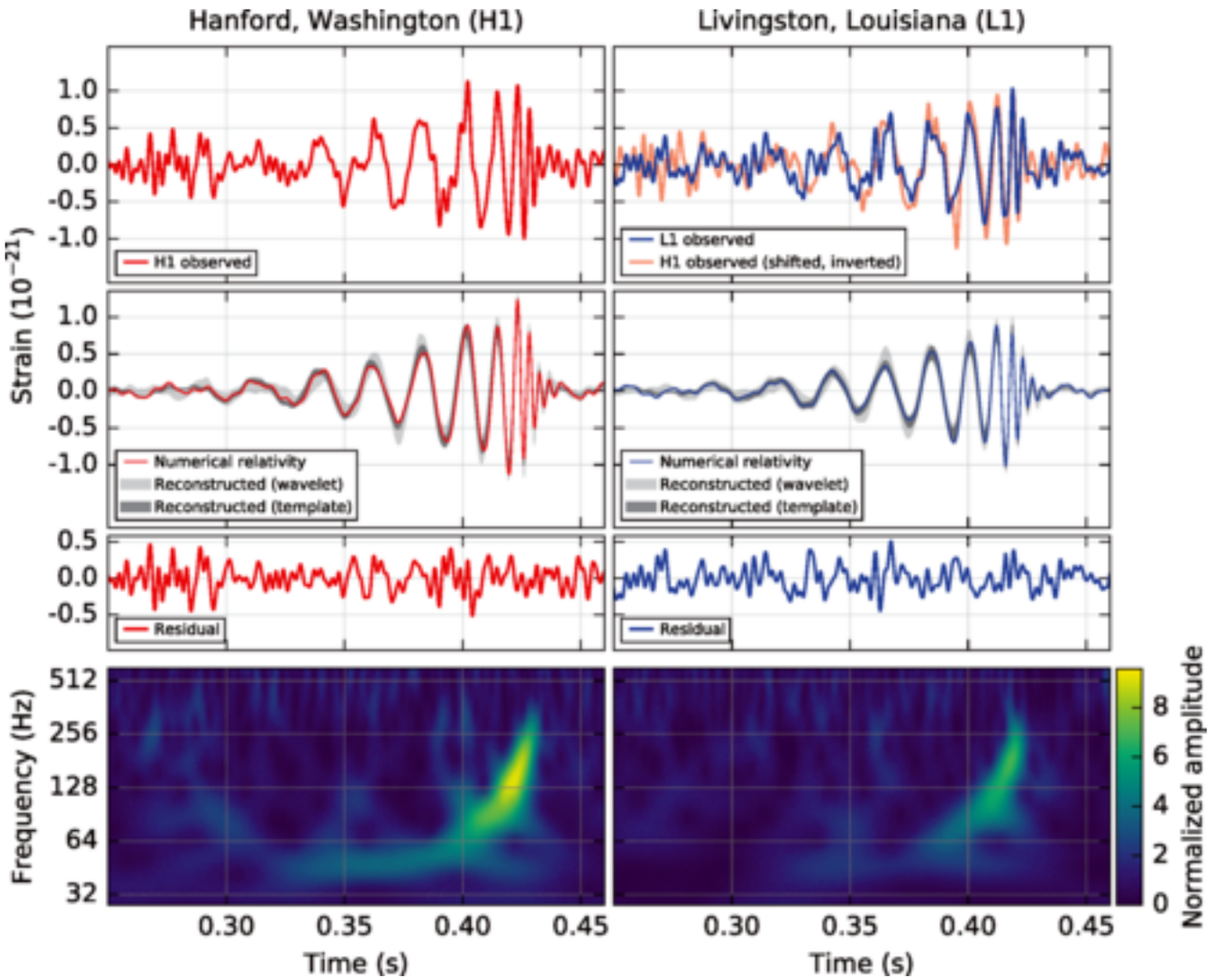
One plot does not work for all audiences/needs



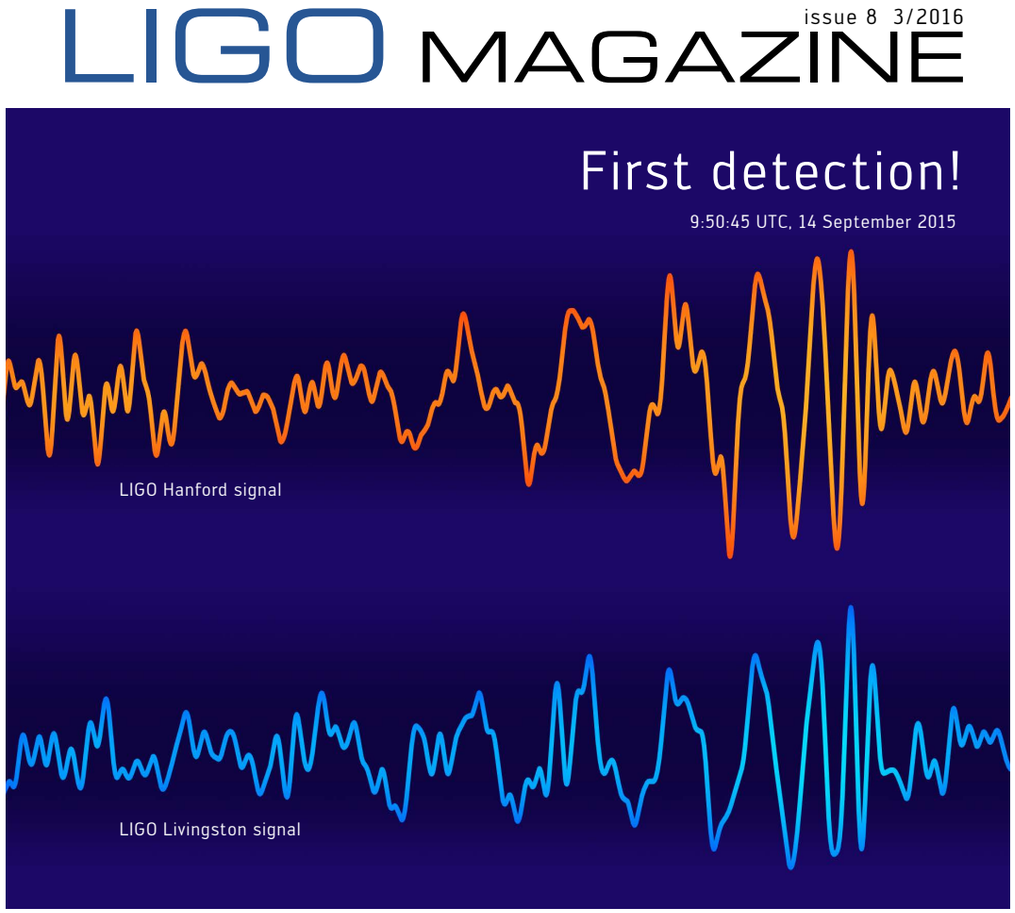


jradavenport

One plot does not work for all audiences/needs



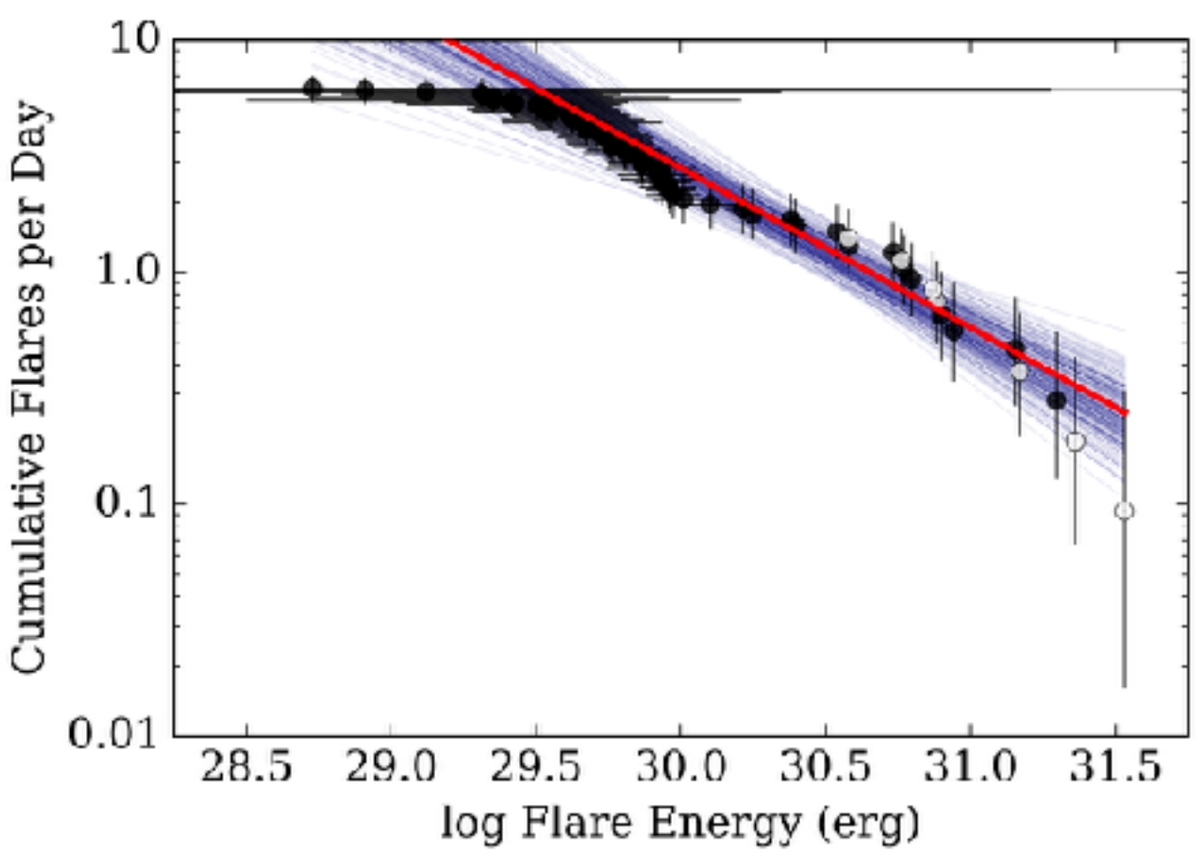
1st LIGO detection, Abbott et al. (2016)



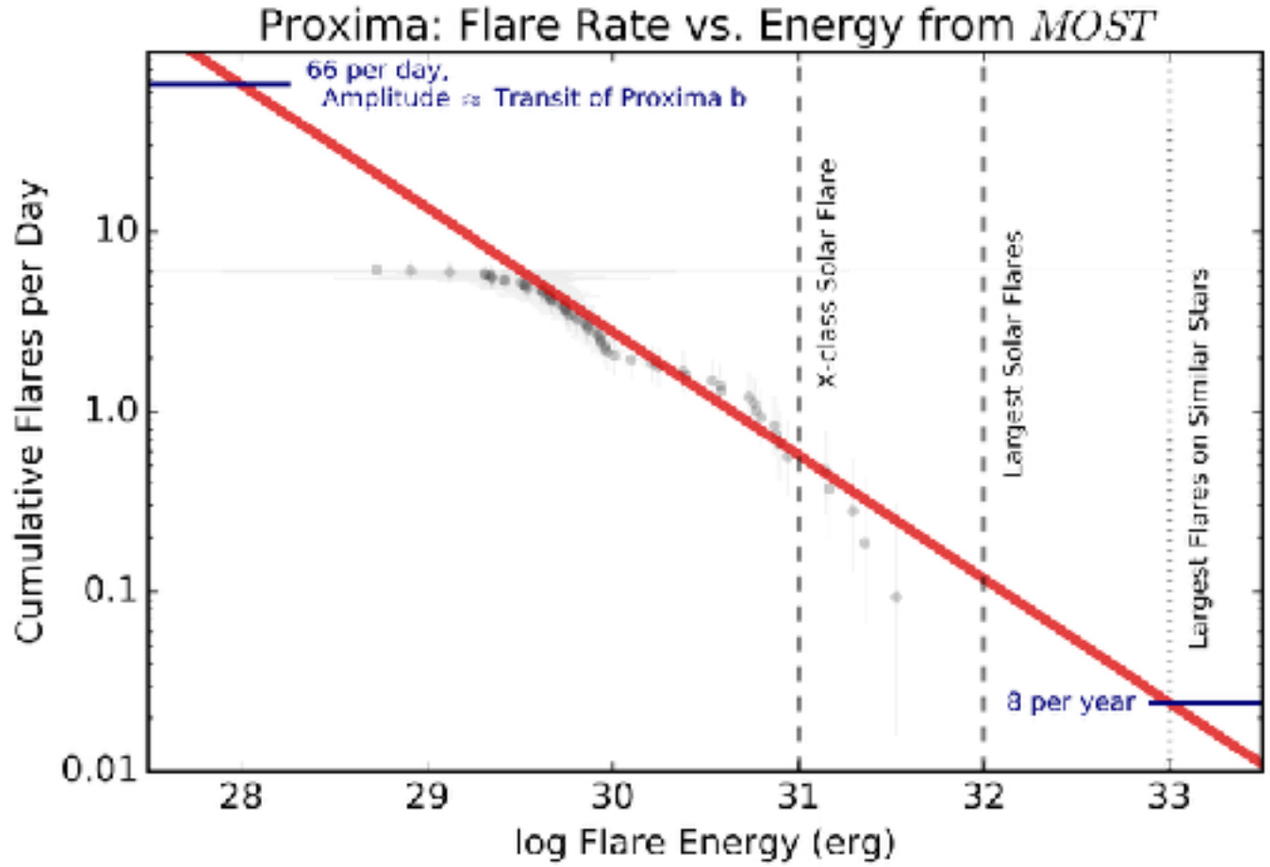
LIGO Magazine, 2016

jradavenport

One plot does not work for all audiences/needs



Davenport et al. (2016)

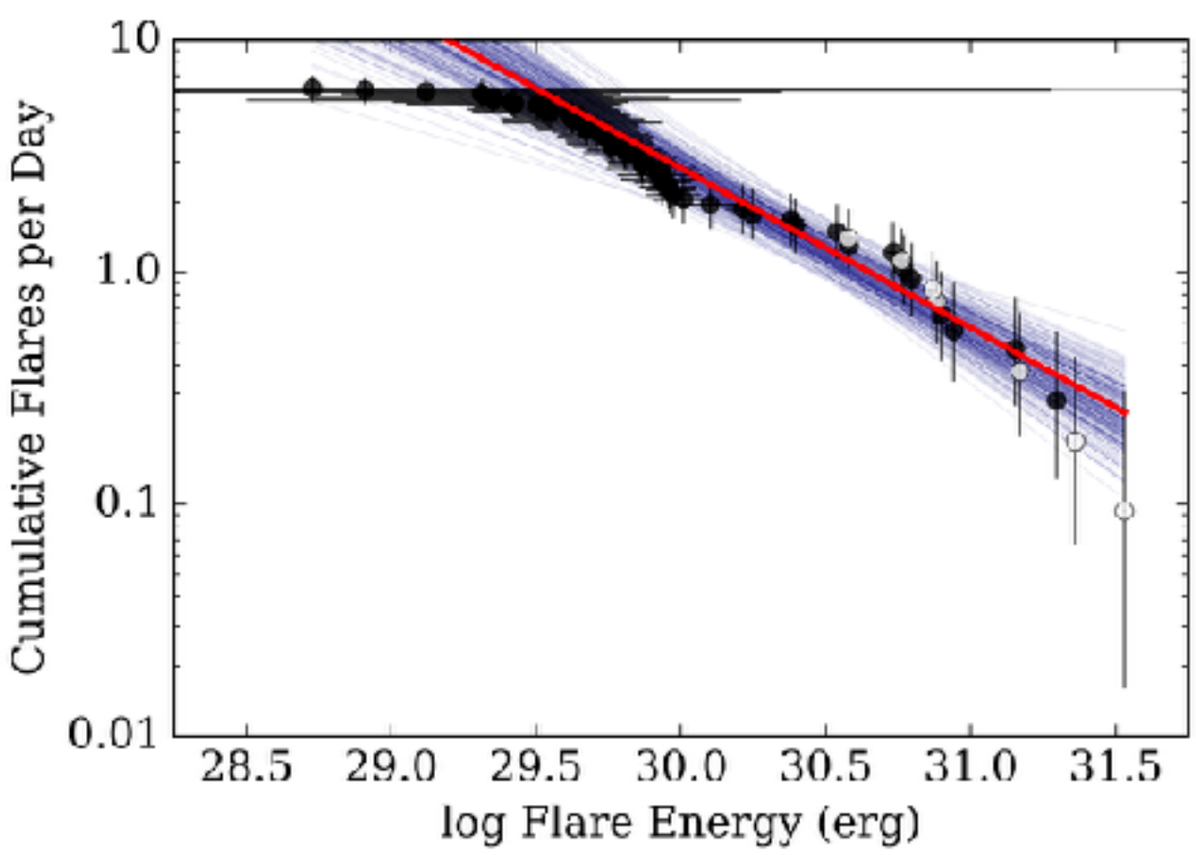


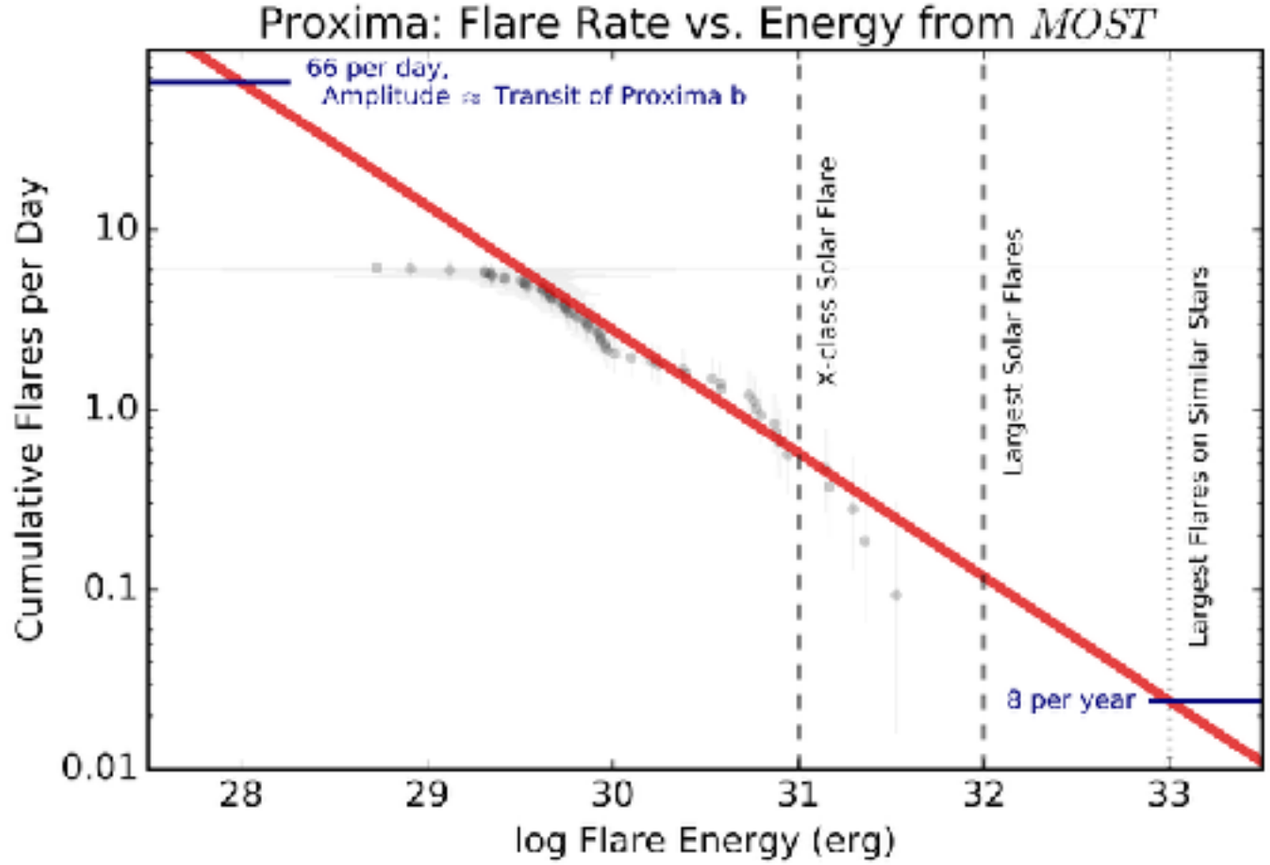
blog post

jradavenport

Point #2

Make multiple versions of a plot!





& make them available for people to use!

jradavenport

A brief aside….

Your camera doesn't matter

– Ken Rockwell (Photographer)

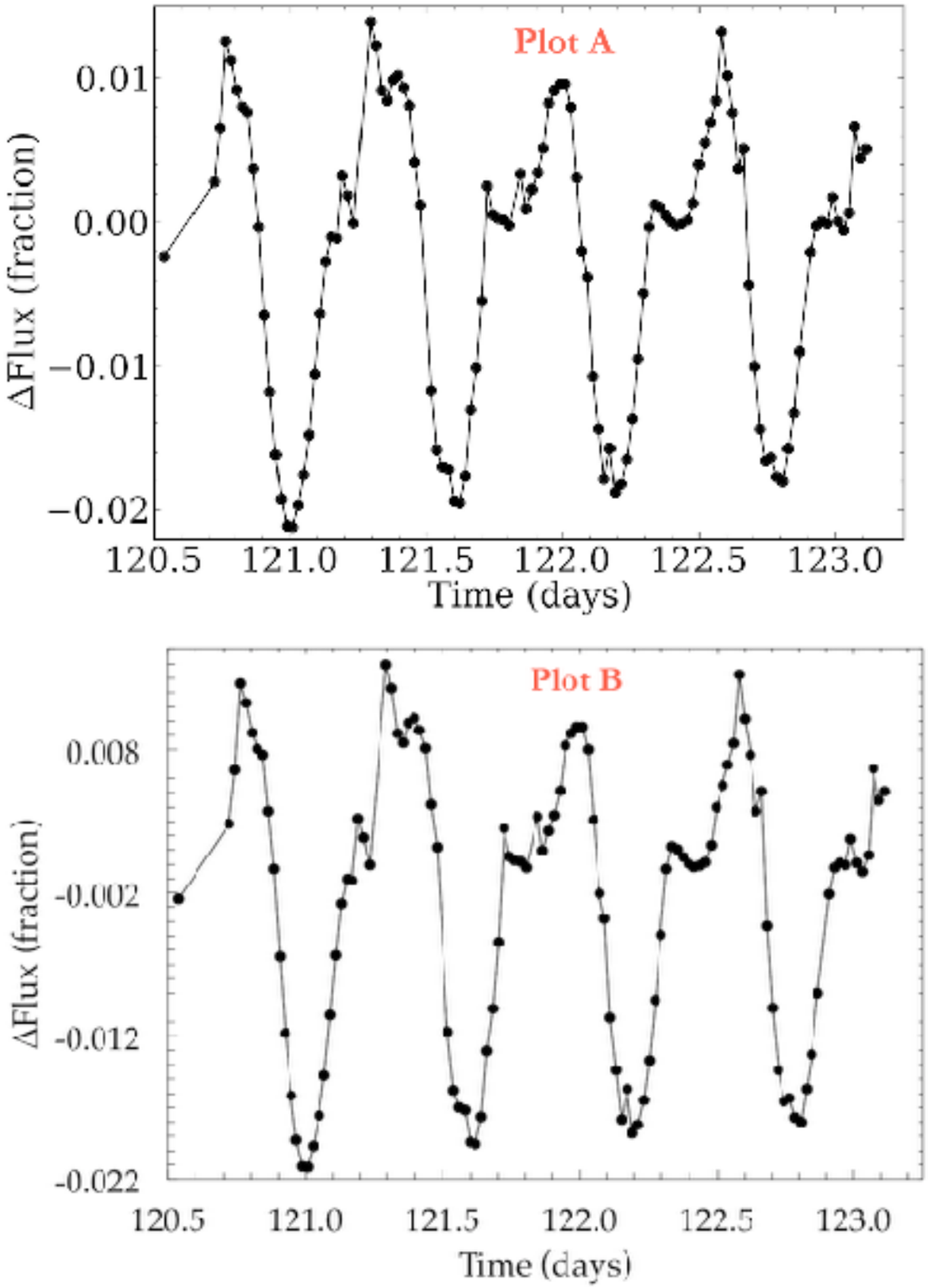
A brief aside….

Your graphing language doesn't matter

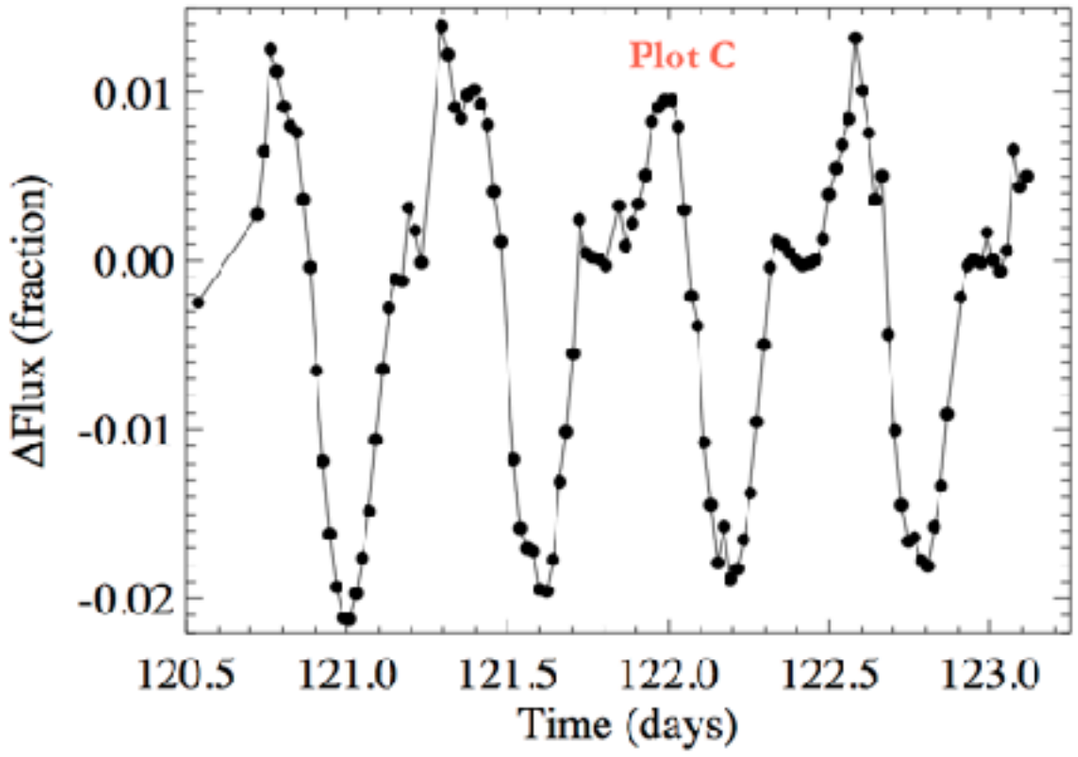
– me, just now

A short example:

A brief aside….



Can you tell which is IDL, Python, Excel ?



Kepler data from Davenport et al. (2014)

jradavenport

Point #3

Don’t tool shame

jradavenport

DESIGN

jradavenport

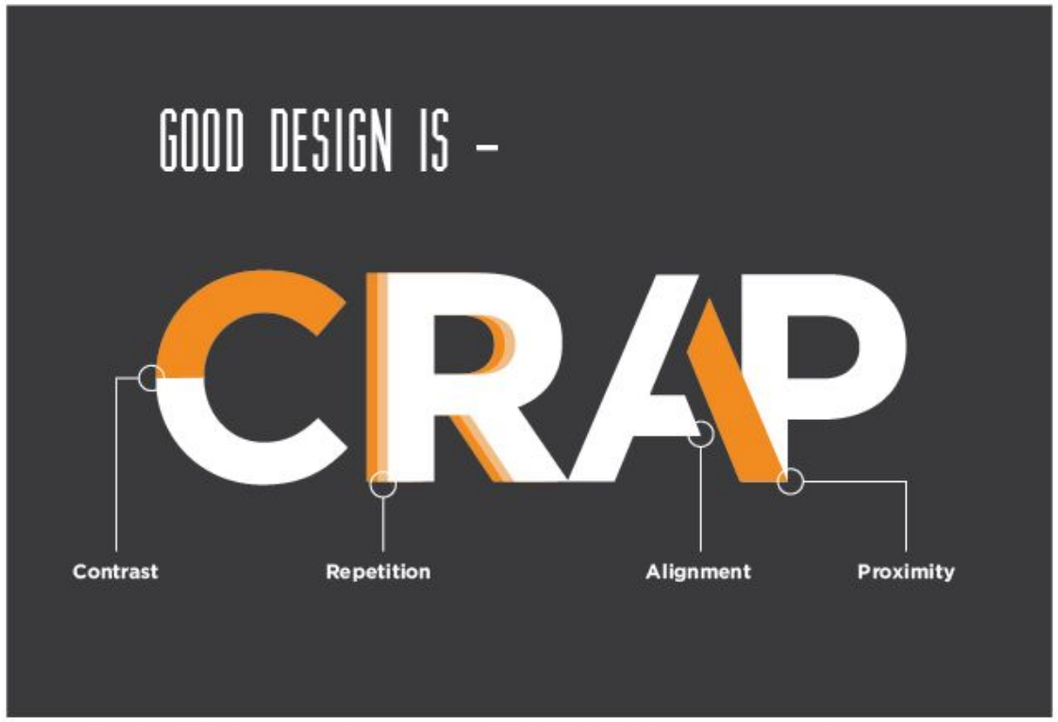
A handy design acronym: PARC

Proximity
Alignment
Repetition
Contrast

Gwen Eadie, DIRAC

A handy design acronym: PARC

Proximity
Alignment
Repetition
Contrast

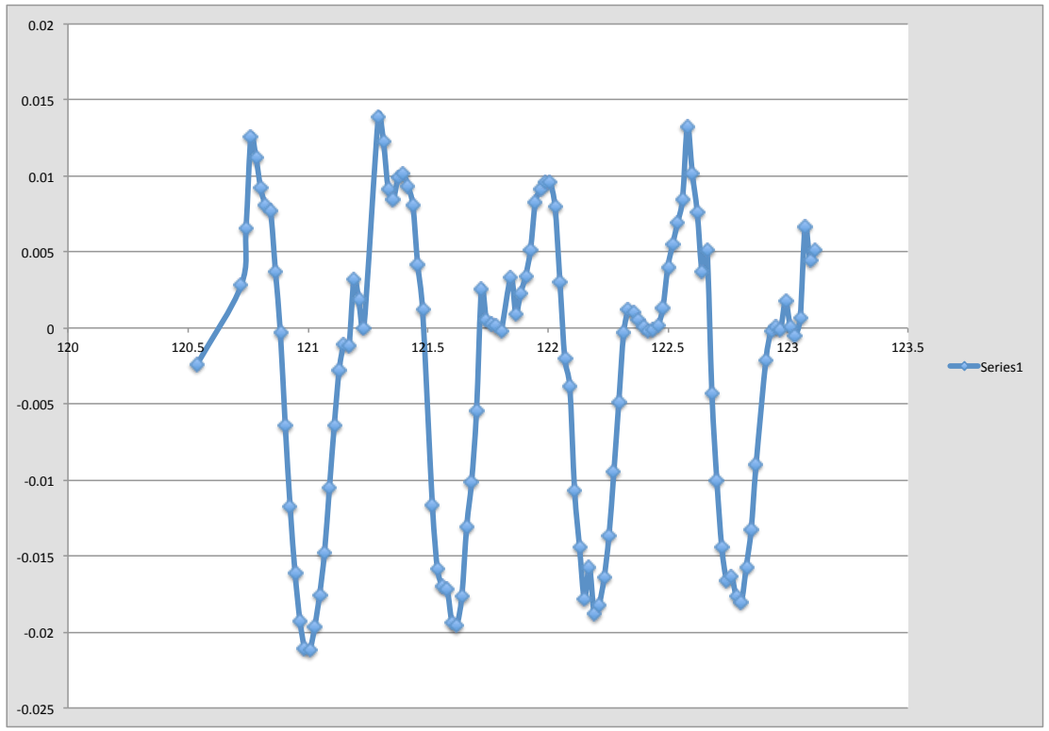


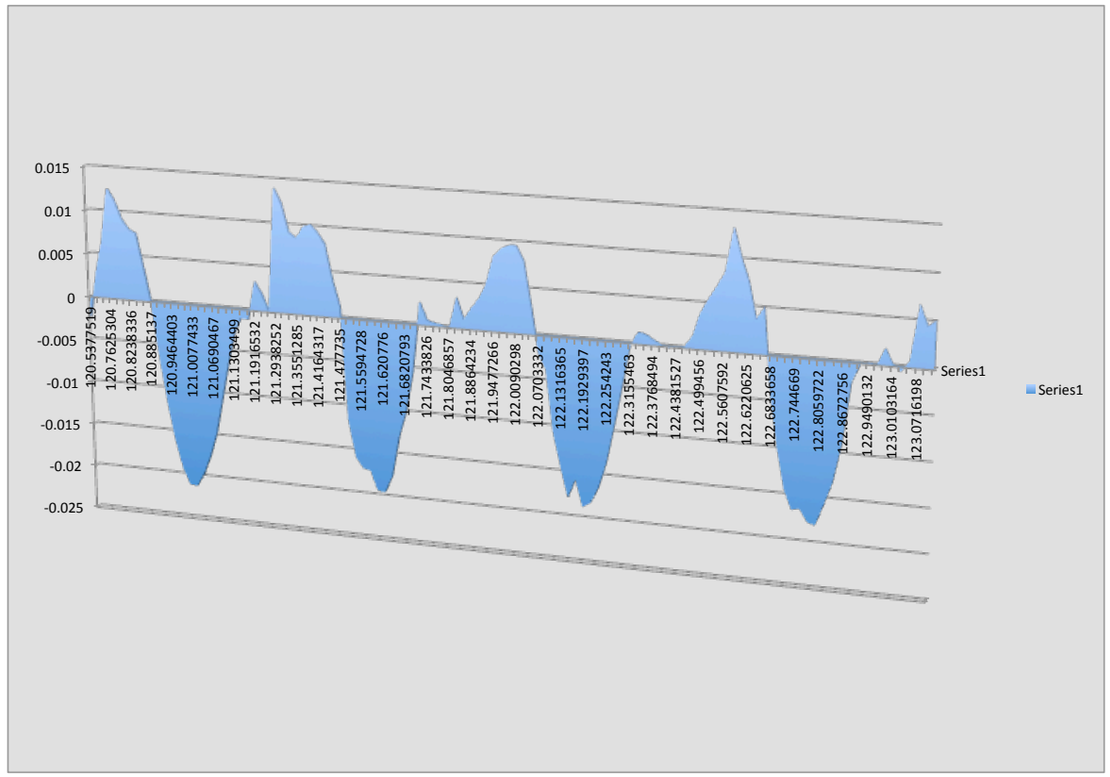
https://mdst485class.wordpress.com/2016/06/16/this-is-c-r-a-p/

Gwen Eadie, DIRAC

Keep it Clean

Default plots from Excel

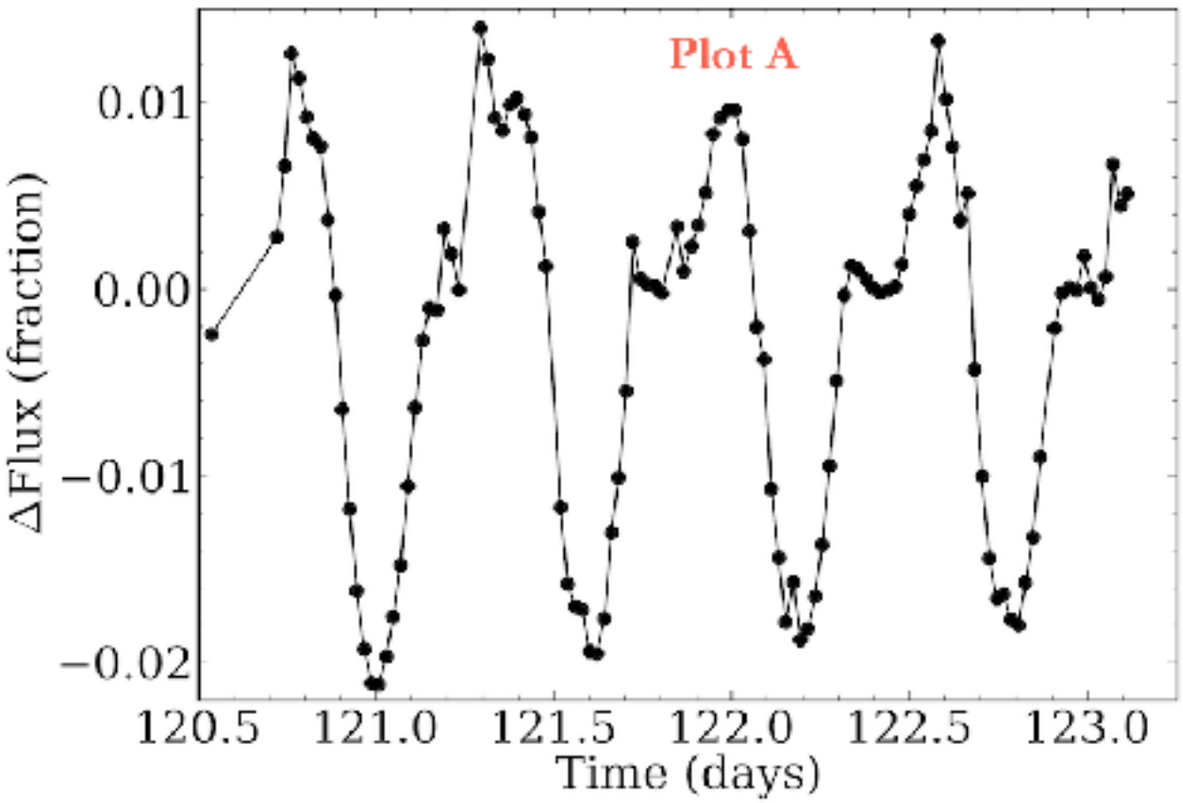


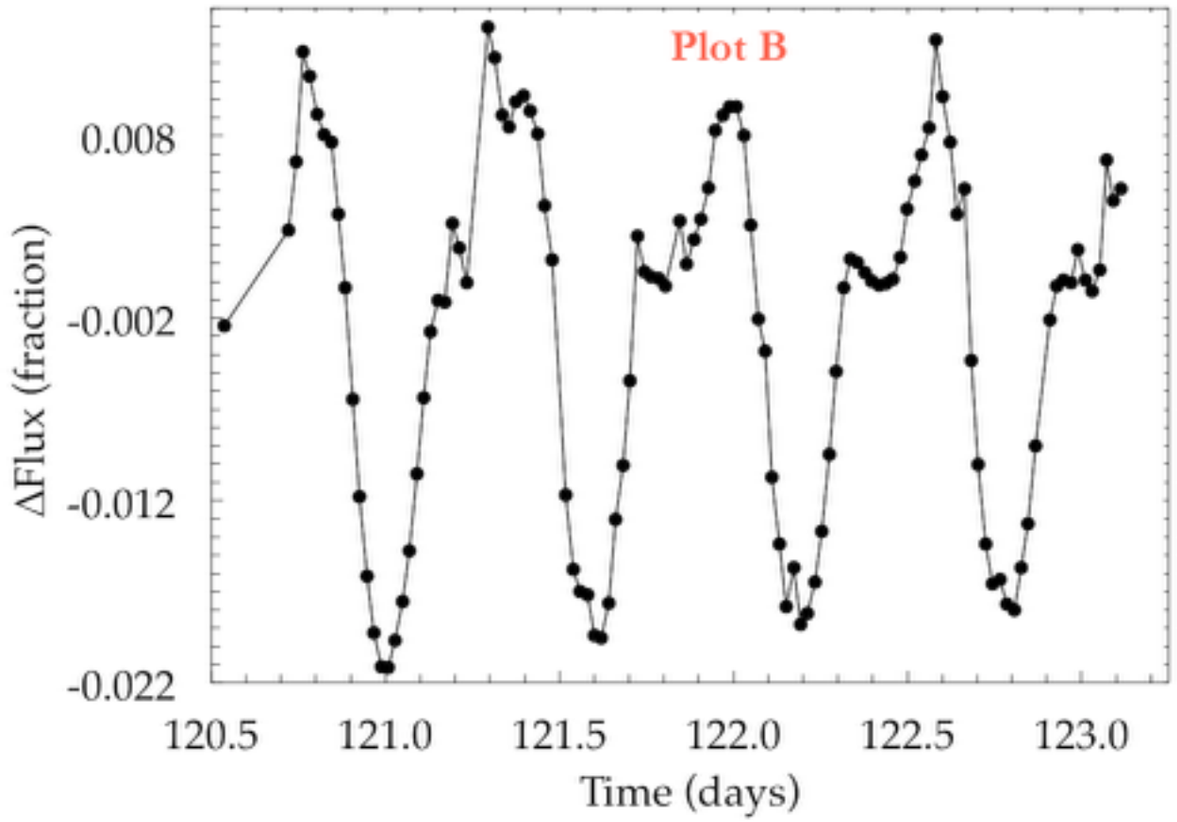


jradavenport

Keep it Clean

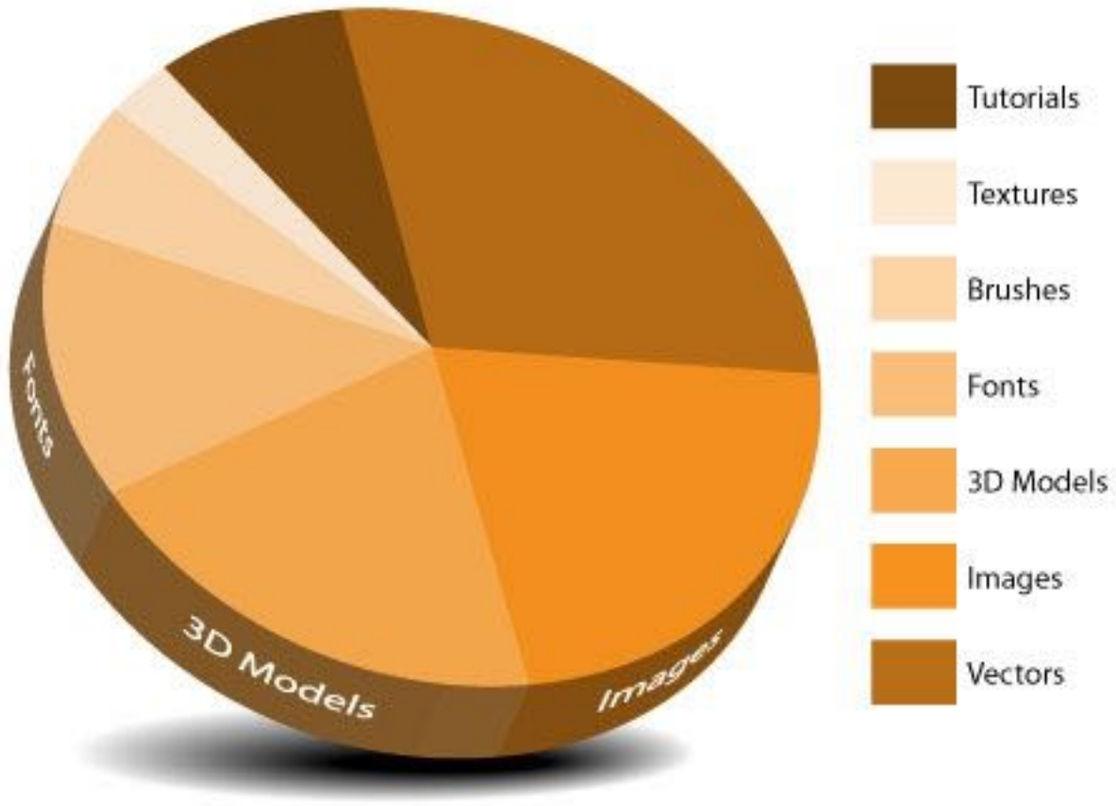
Clean plots from Excel and Matplotlib (Python)





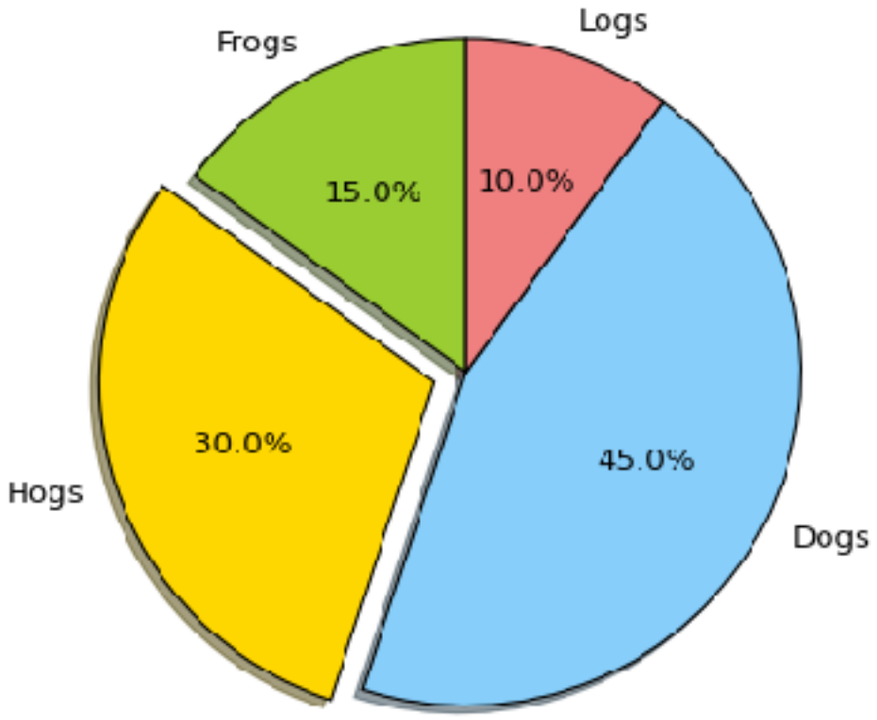
jradavenport

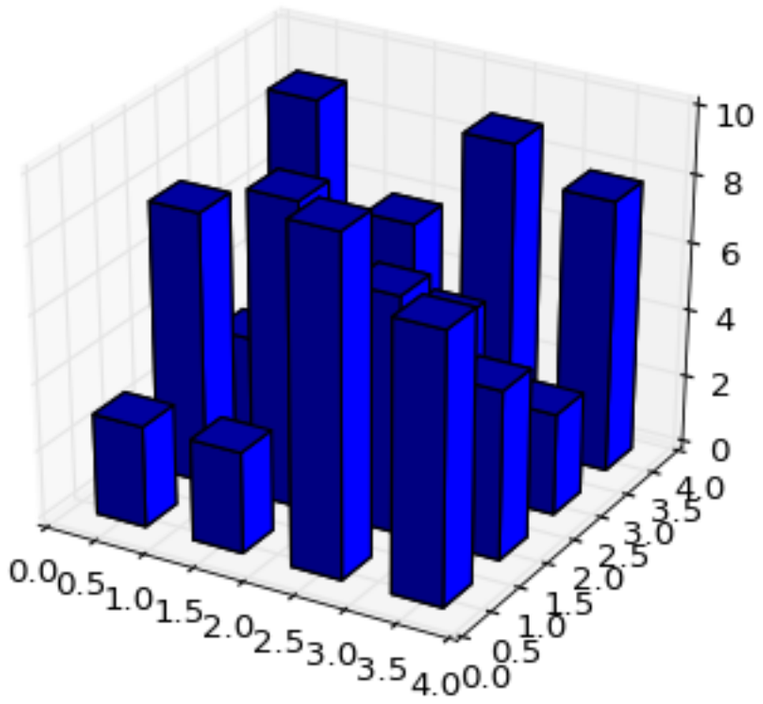
Avoid Gimmicks



jradavenport

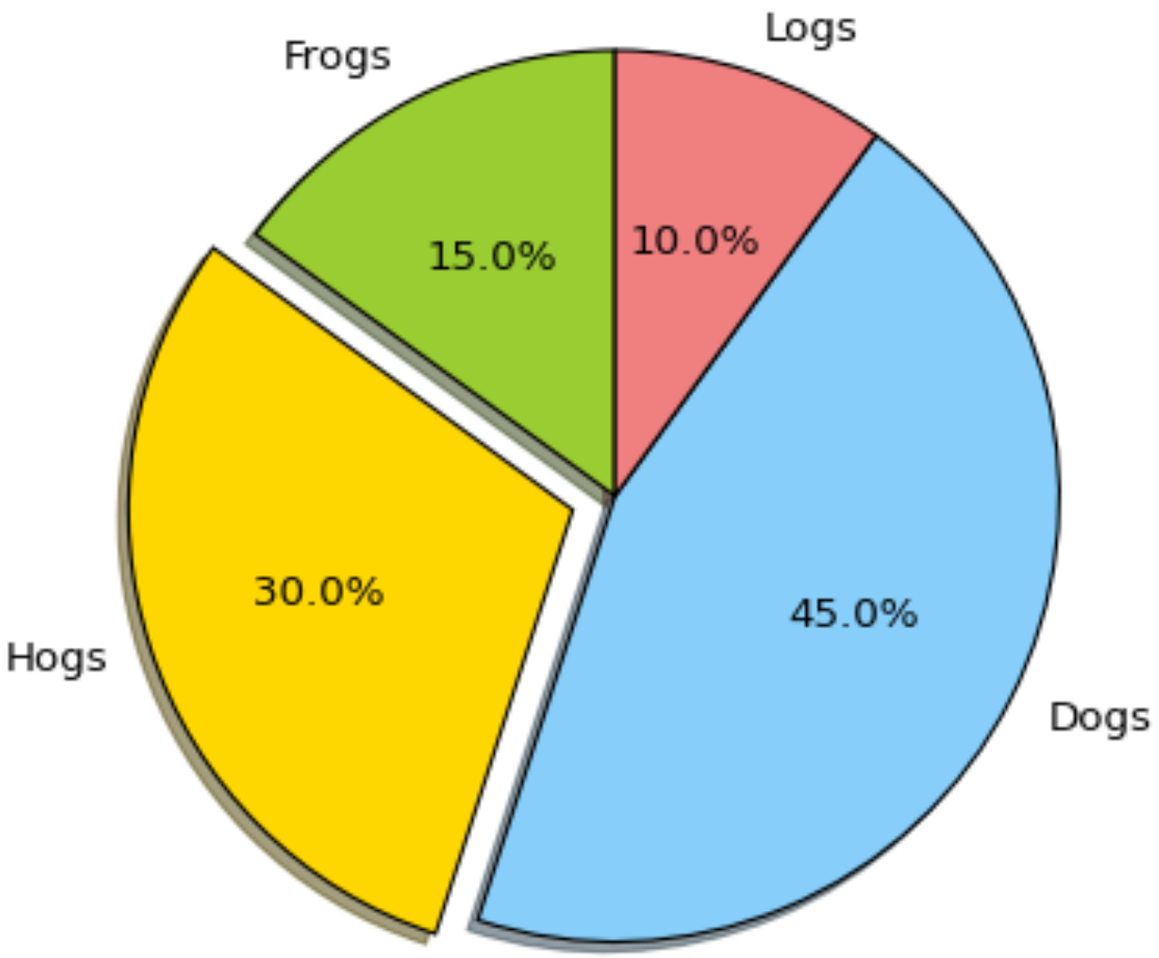
Avoid Gimmicks

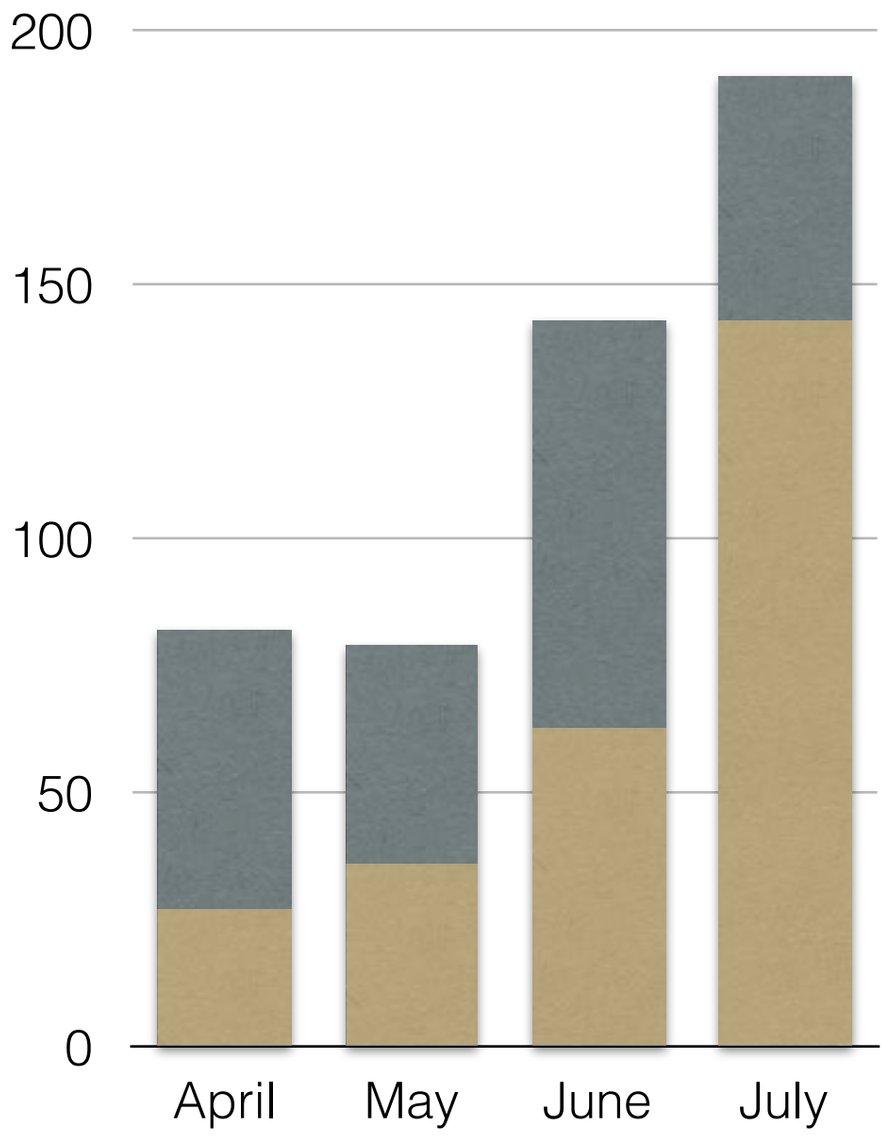




examples given for matplotlib…

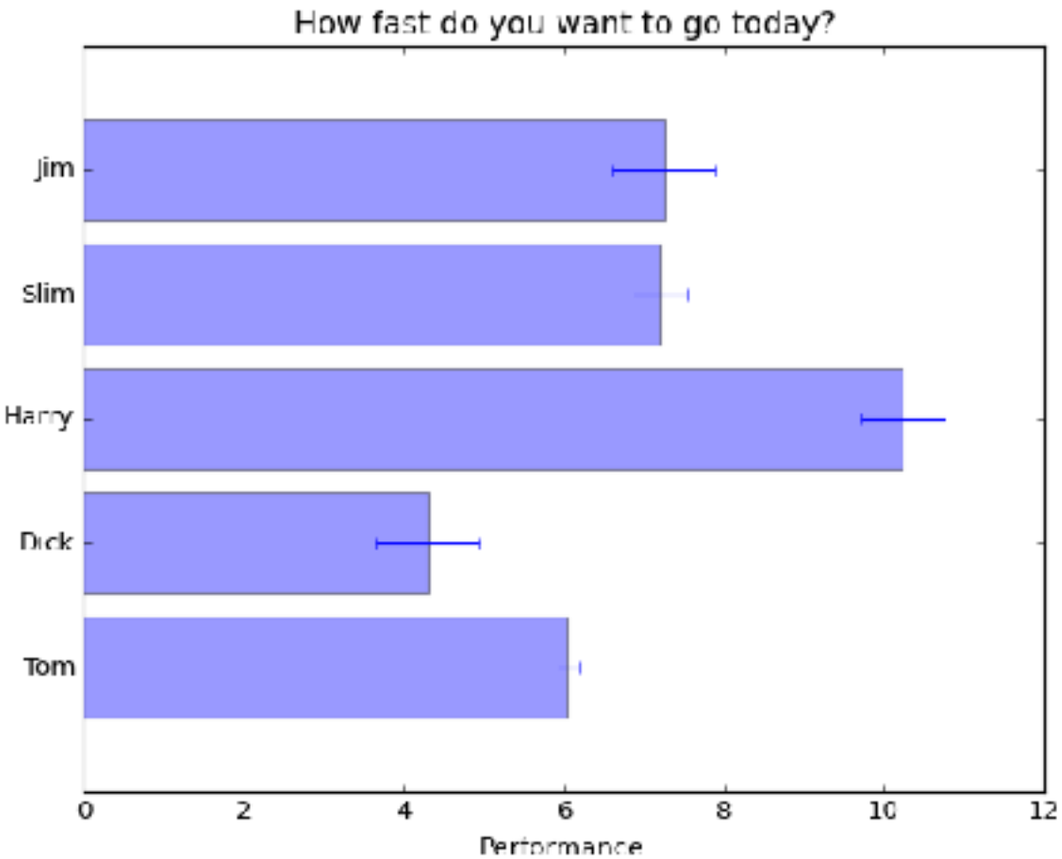
Make things comparable

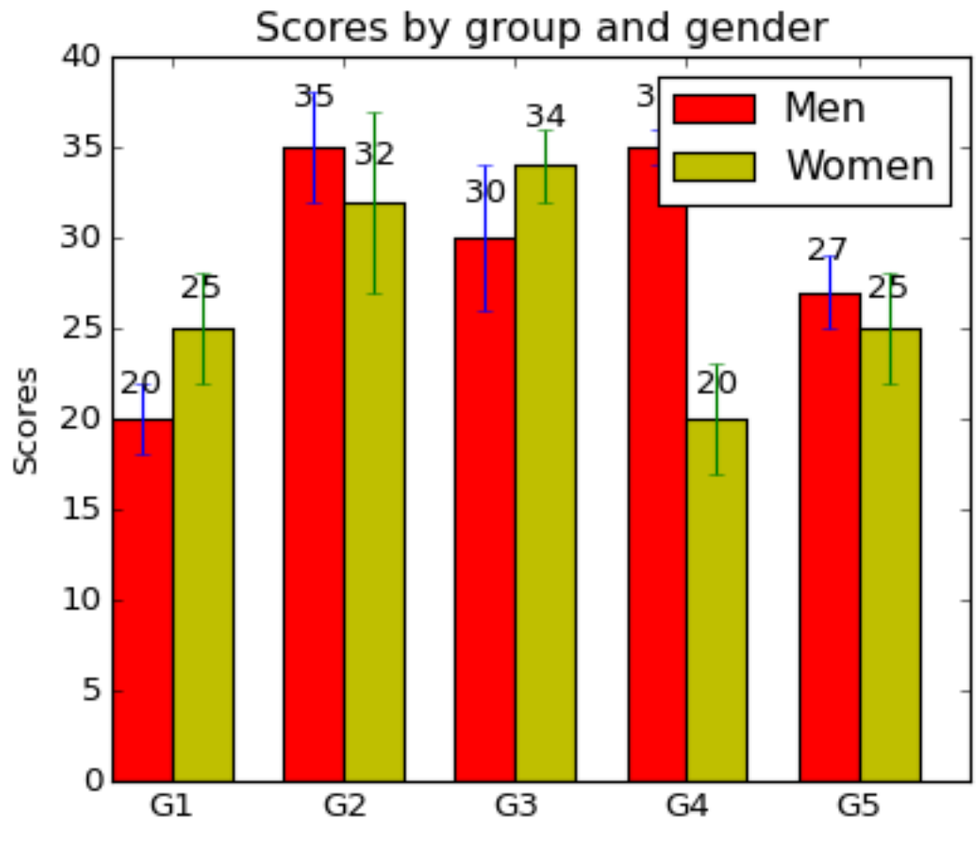




jradavenport

Be Consistent





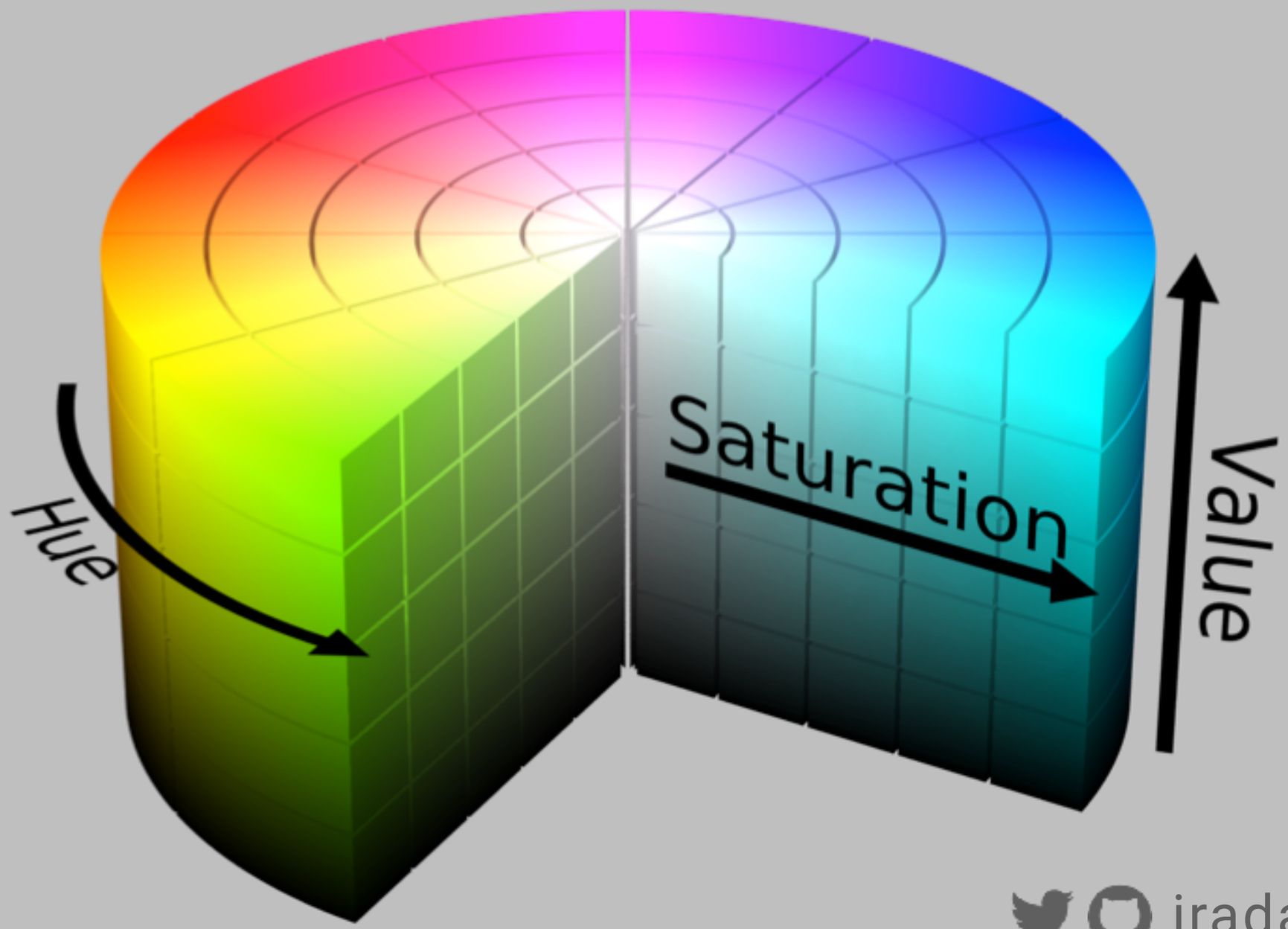
Color in visualizations: best friend, worst enemy

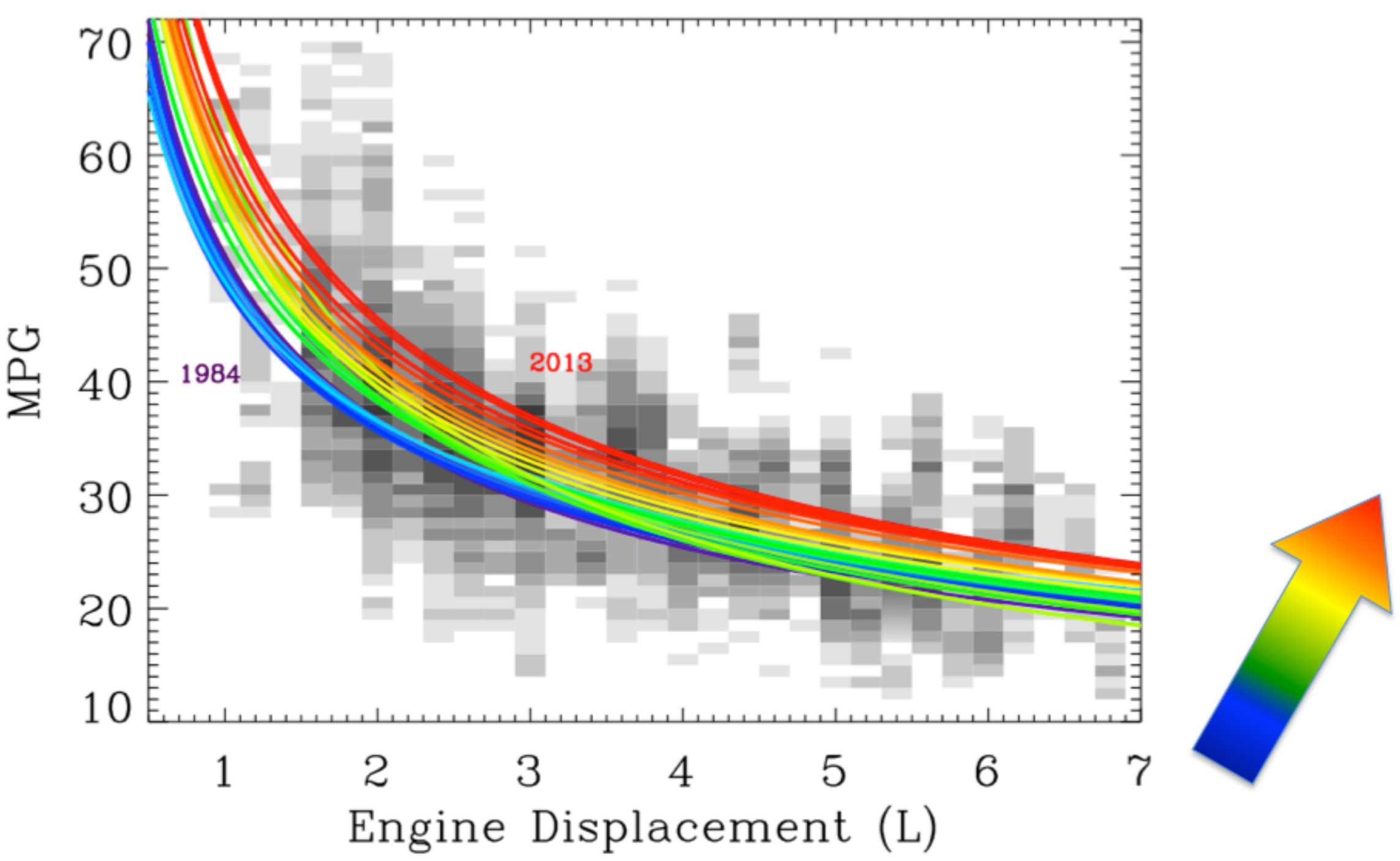
jradavenport

Rainbow: worst/best

Pro: (almost) everybody understands it

Con: it's inaccurate,
doesn't print well,
bad for colorblind people





jradavenport

Spare a moment for colorblindness

- 5-10% of population!
- Again, projector is evil

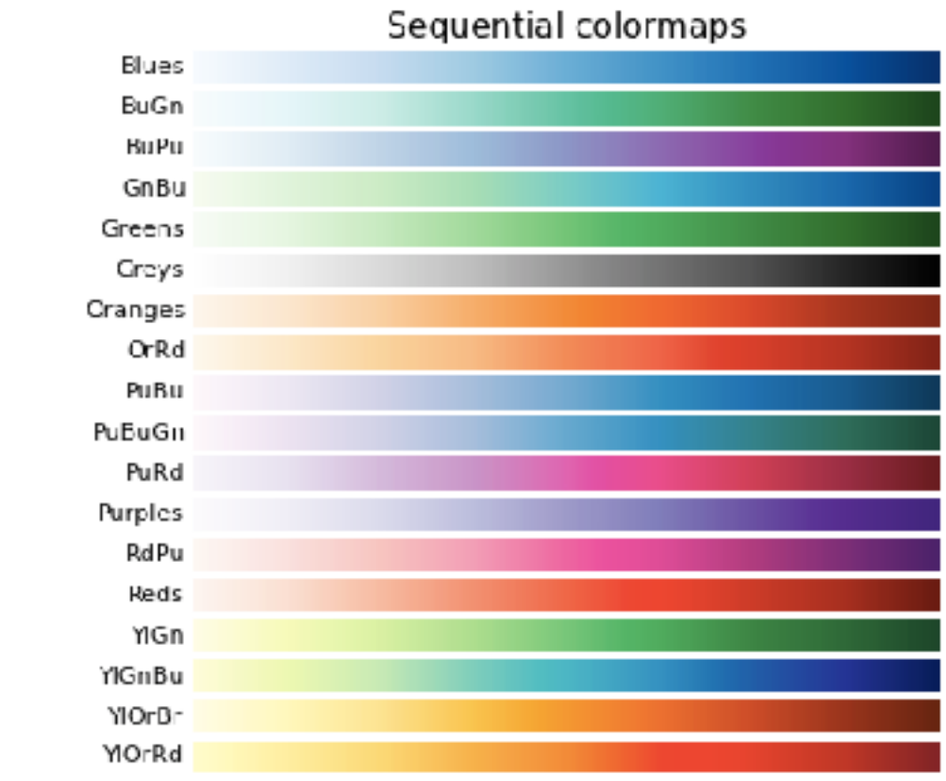
Some colorful hats.

As seen by a person with deuteranopia.

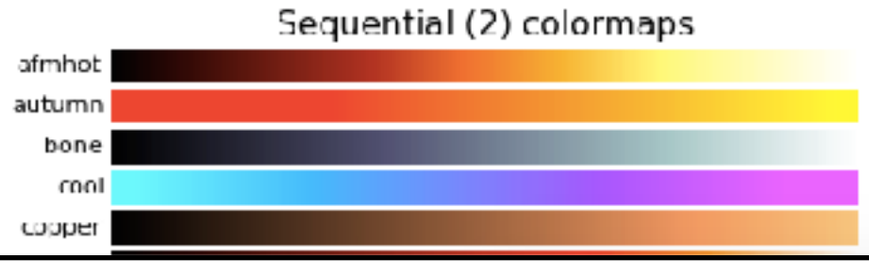
As seen by a person with protanopia, another form of red/green deficit.

color example code: colormaps_reference.py

(Source code)



(png, hires.png, pdf)



Previous topic

color example code: color_cycle_demo.py

Next topic

color example code: named_colors.py

This Page

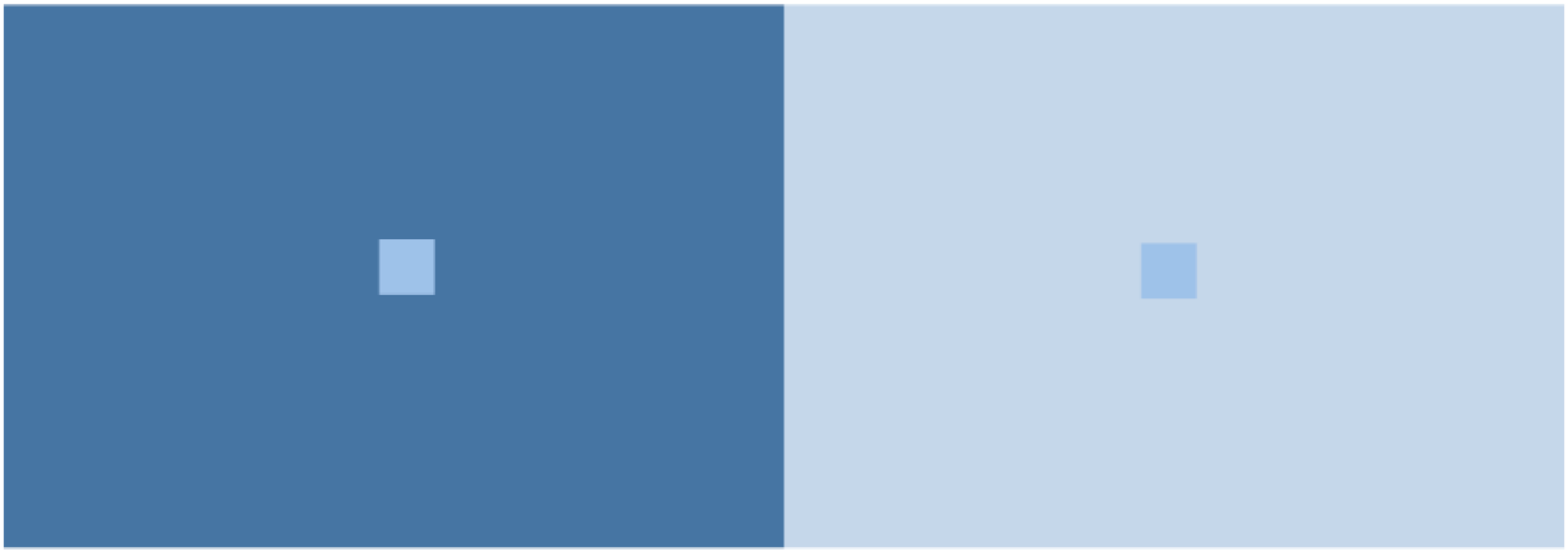
Show Source

Quick search

Go

Enter search terms or a module, class or function name.

Contrast is King



Concept from Edward Tufte

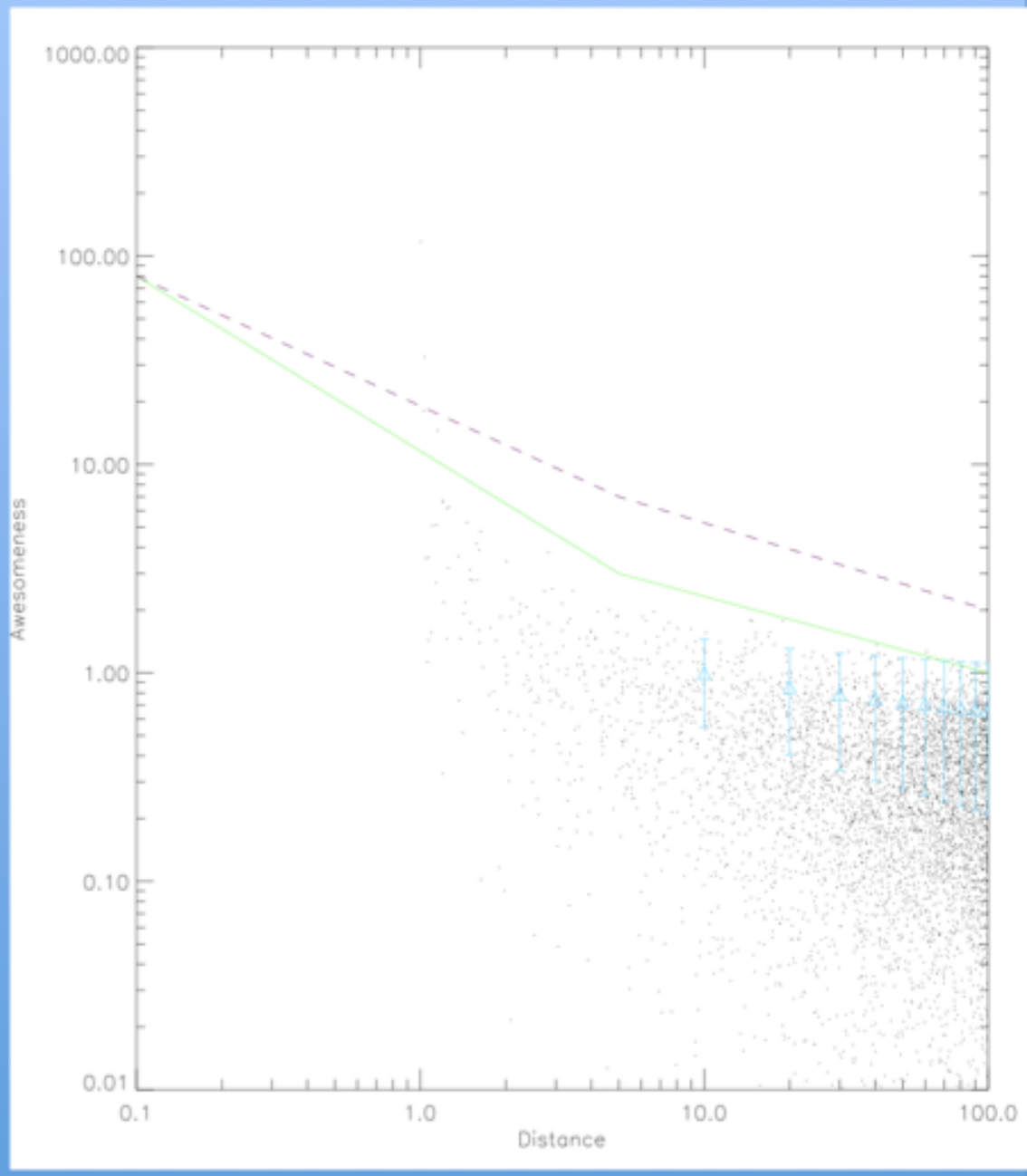
jradavenport

Contrast is King

Concept from Edward Tufte

Contrast is King

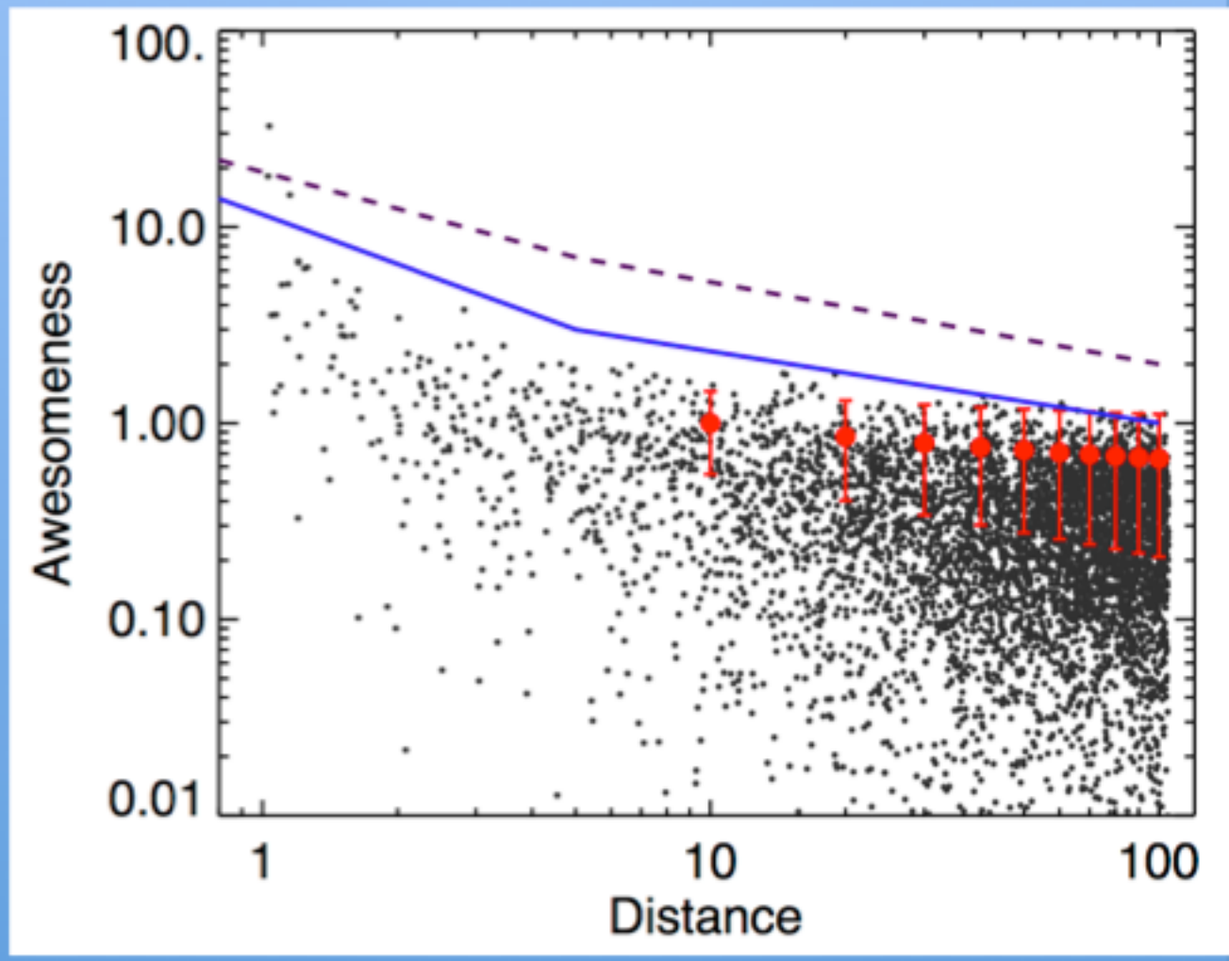
- The data matches the model remarkably well
- This changes everything



- Oh, this looks better on *my* screen…

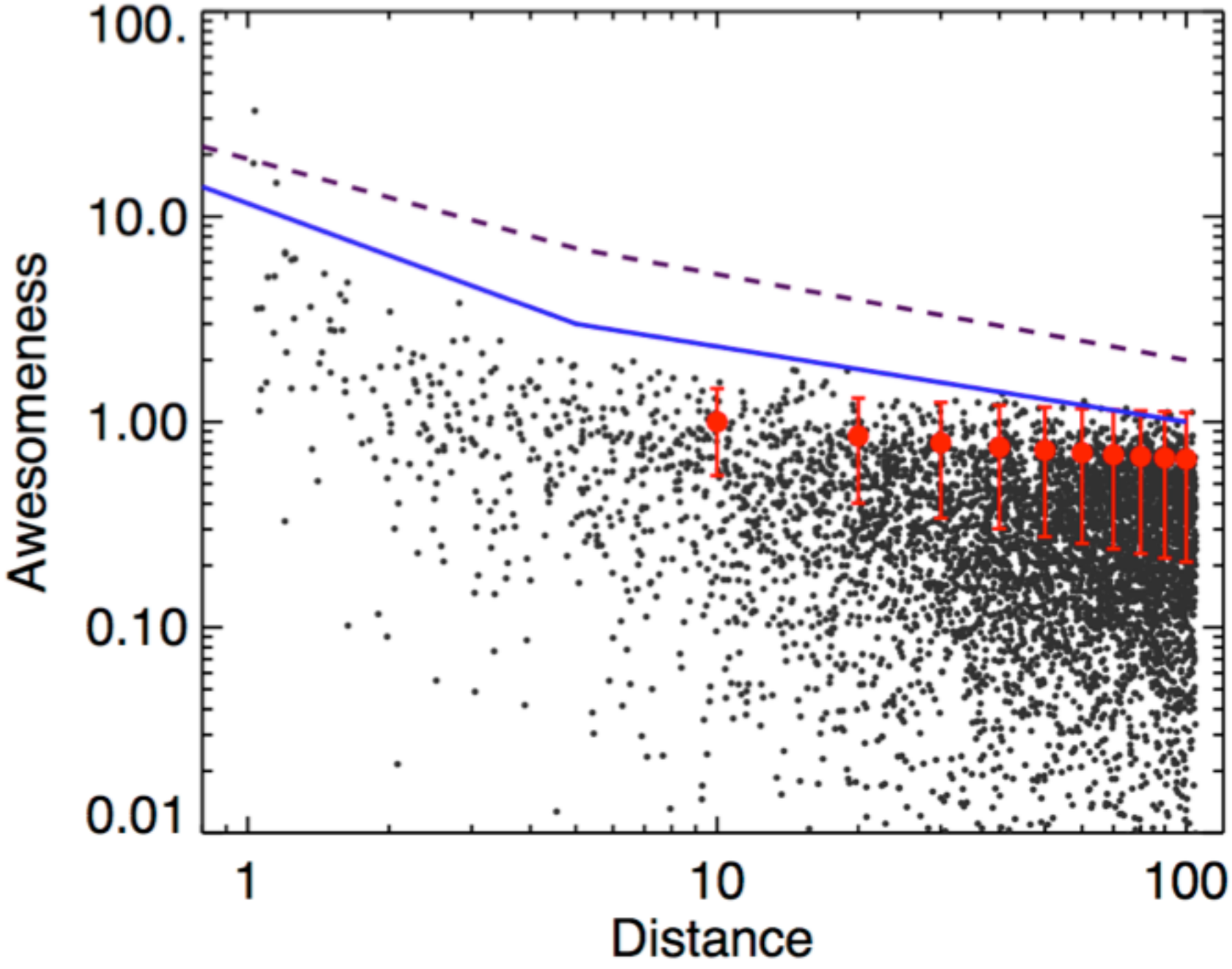
Contrast is King

- The data matches the model remarkably well
- This changes everything



- Now *you* can see it
- **Seeing is believing**

jradavenport



Ingredients:

1. X-range
2. Y-range
3. Empty space
4. Log-log space
5. Y-tick labels
6. Axis labels
7. Point colors
8. Point sizes
9. Filled/empty
10. Line styles
11. Line colors
12. Line thickness
13. Plotting orders
14. Font size
15. Font thickness

jradavenport

Point #4

Try to use good design principles, especially **contrast**.

Attention to Details

jradavenport

Visualization works …

"… when data understanding is supported by perceptual rather than cognitive processes"

– Enrico Bertini

Daniela Huppenkothen, DIRAC

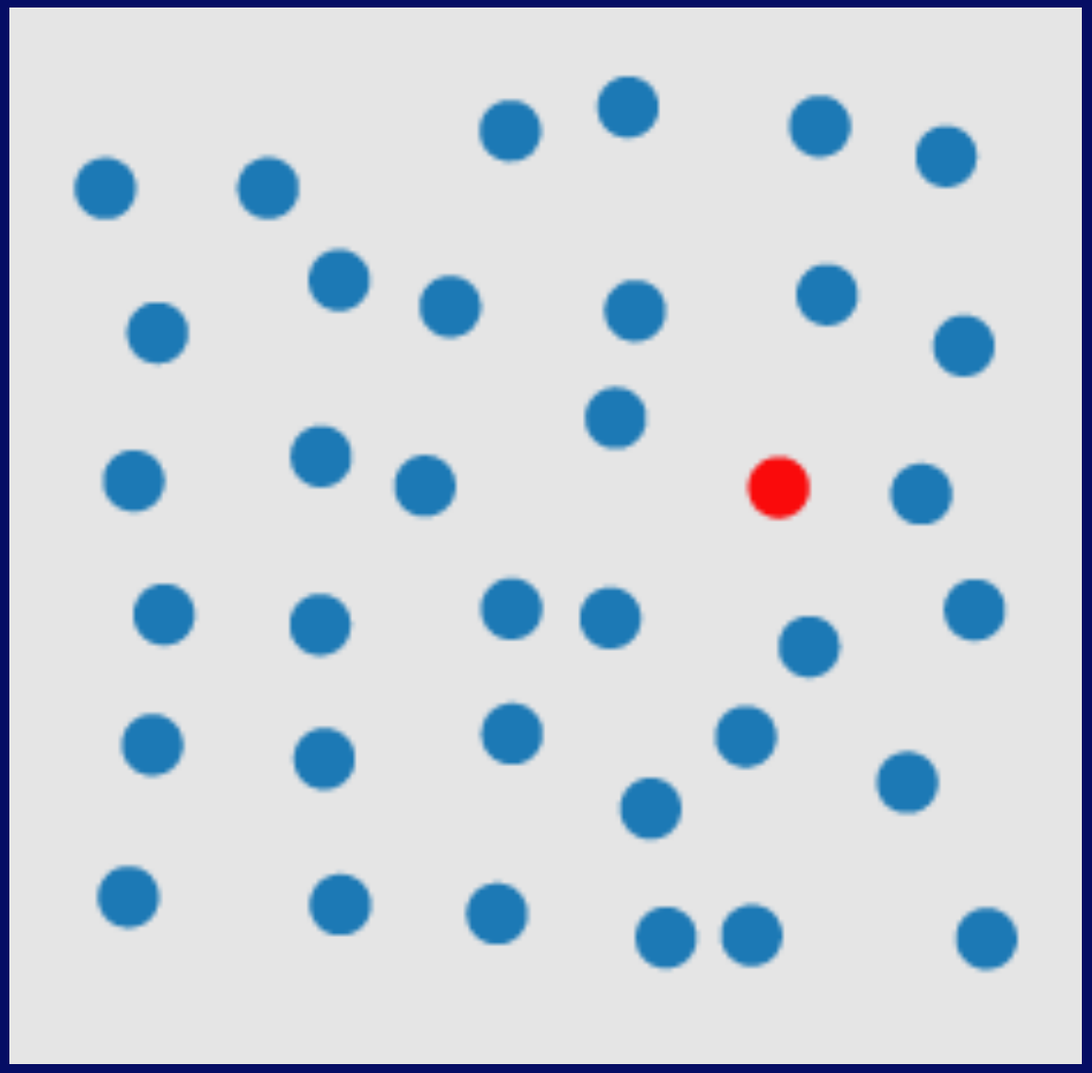
Visualization is about managing the viewer's attention!

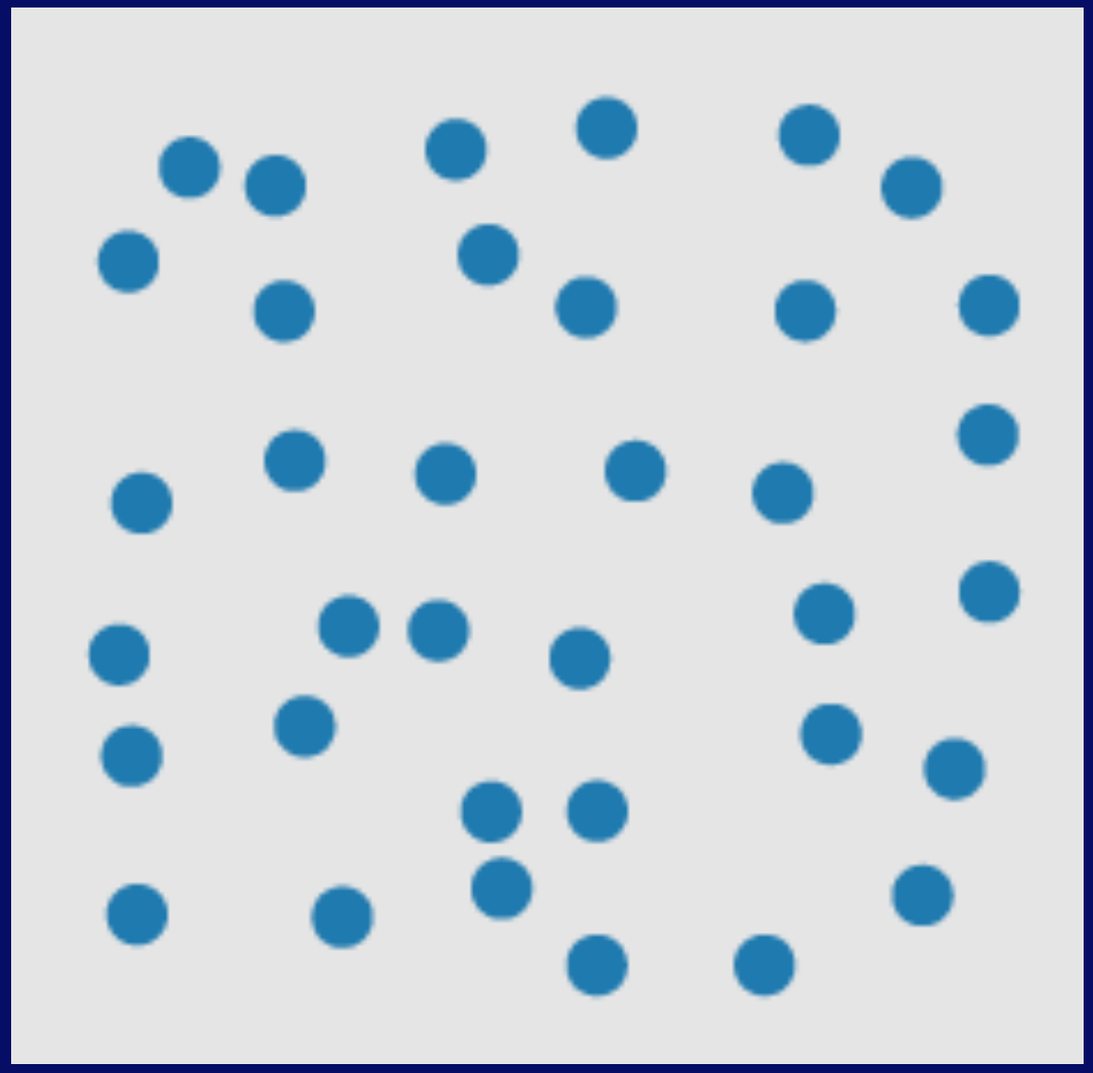
Daniela Huppenkothen, DIRAC

Daniela Huppenkothen, DIRAC

human vision is not like photography!

Daniela Huppenkothen, DIRAC

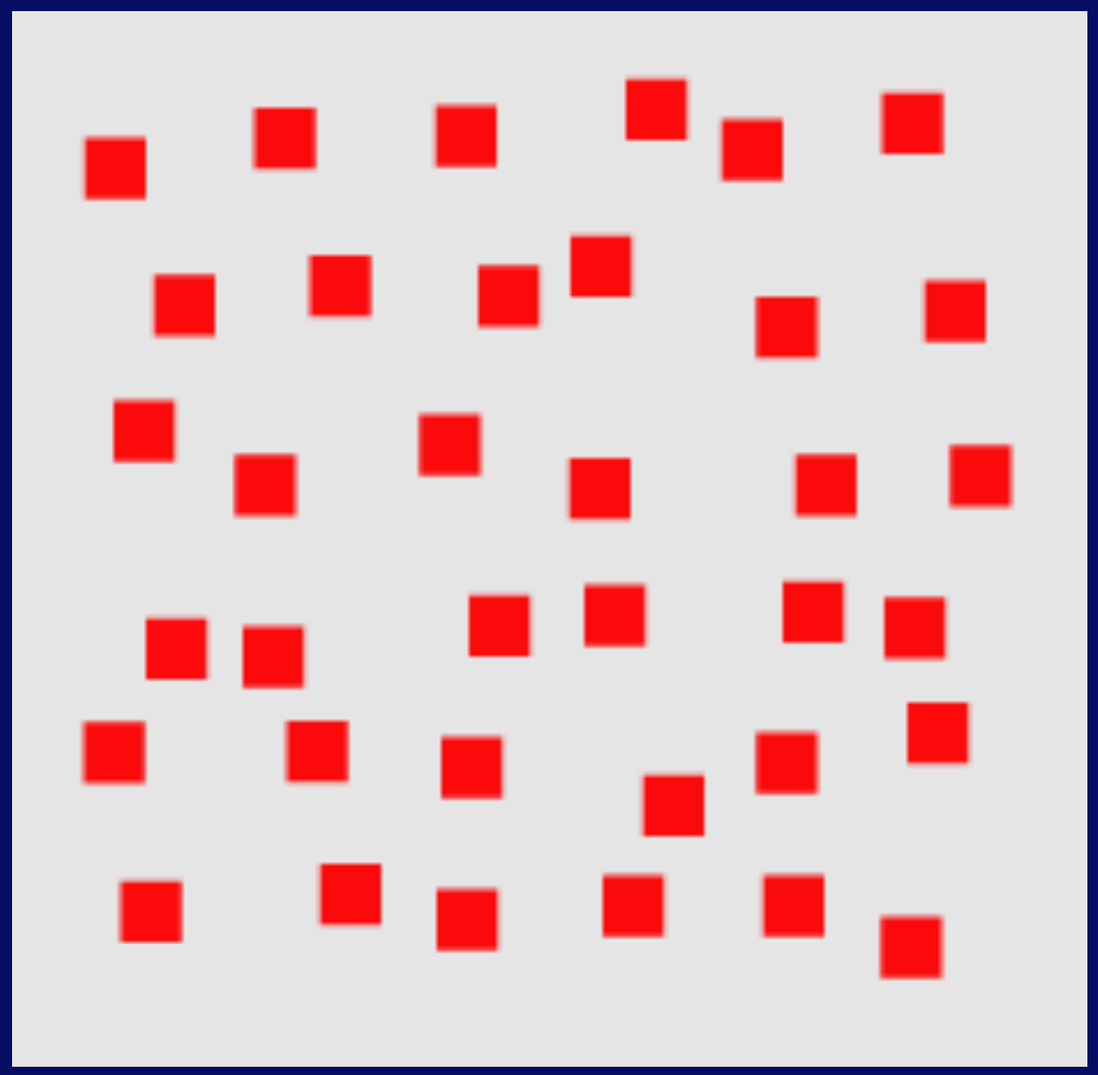


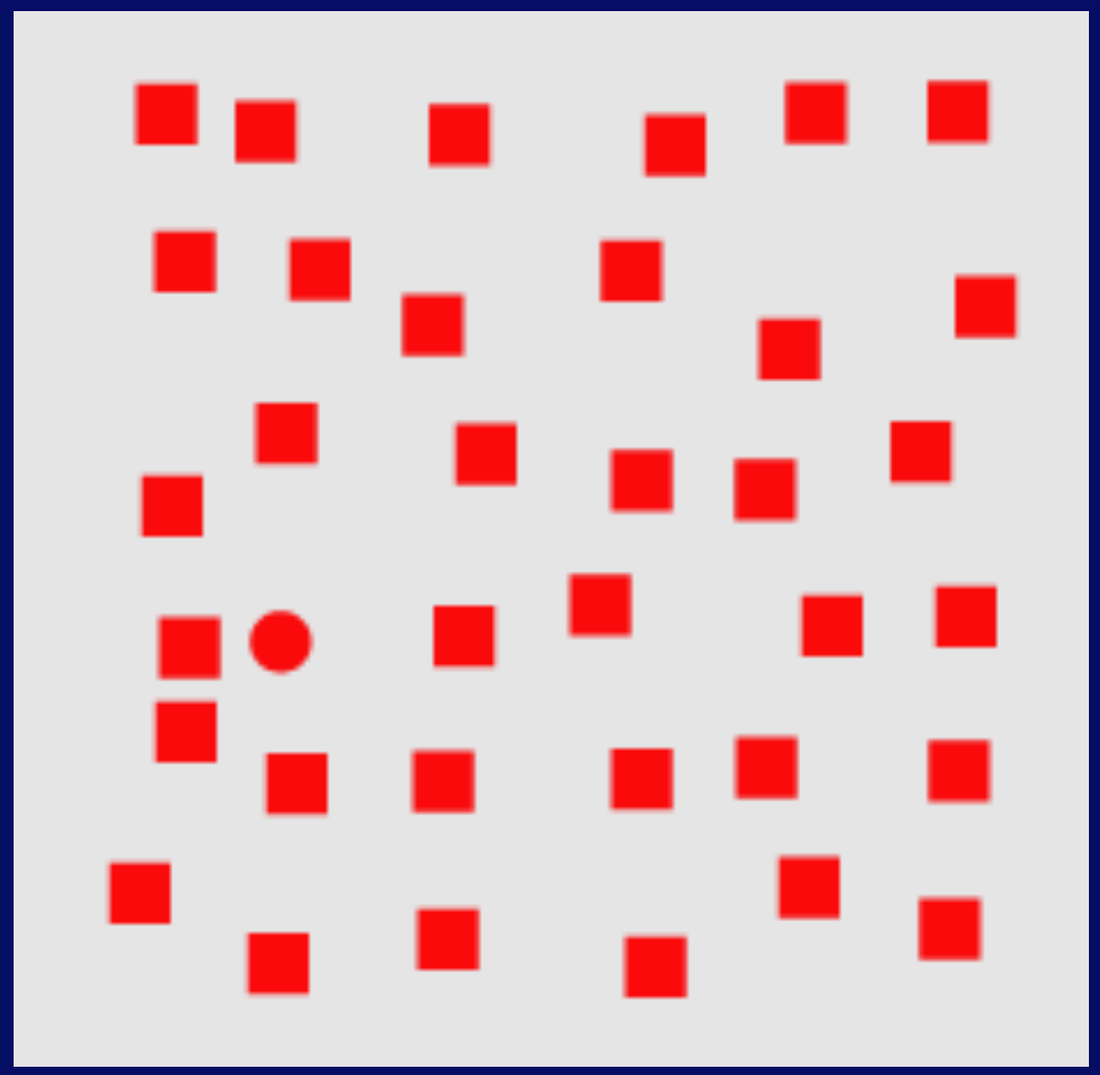


pre-attentive task

https://www.csc2.ncsu.edu/faculty/healey/PP/

Daniela Huppenkothen, DIRAC

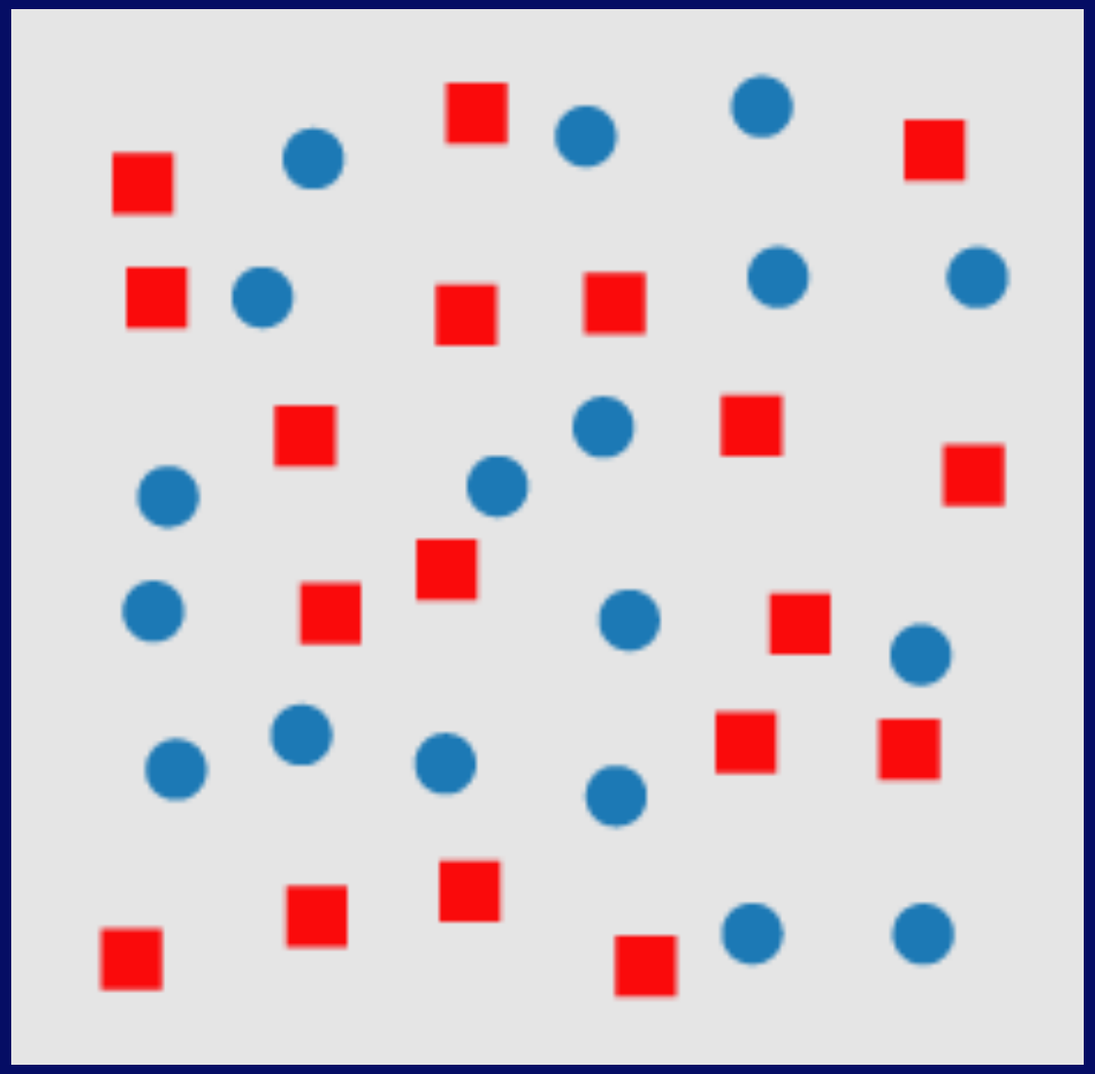


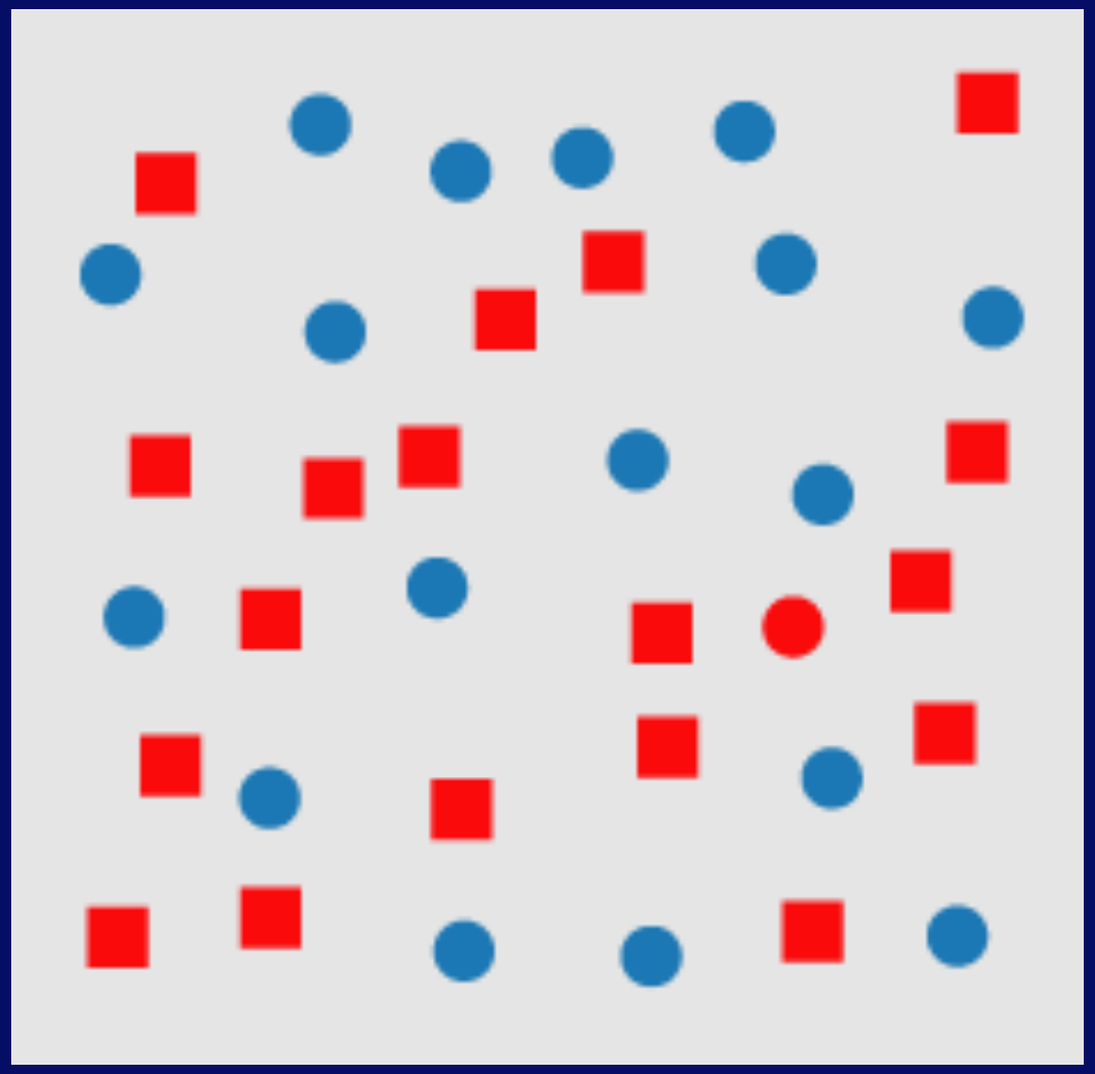


pre-attentive task

https://www.csc2.ncsu.edu/faculty/healey/PP/

Daniela Huppenkothen, DIRAC



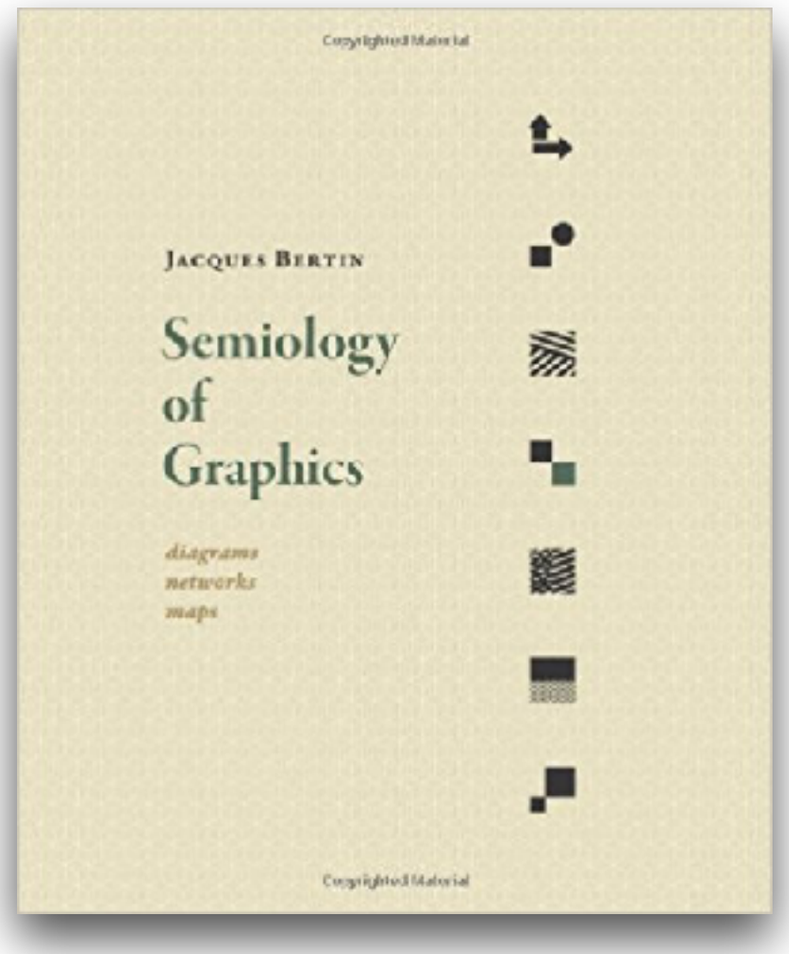


serial search

https://www.csc2.ncsu.edu/faculty/healey/PP/

Daniela Huppenkothen, DIRAC

Everything must have **meaning**

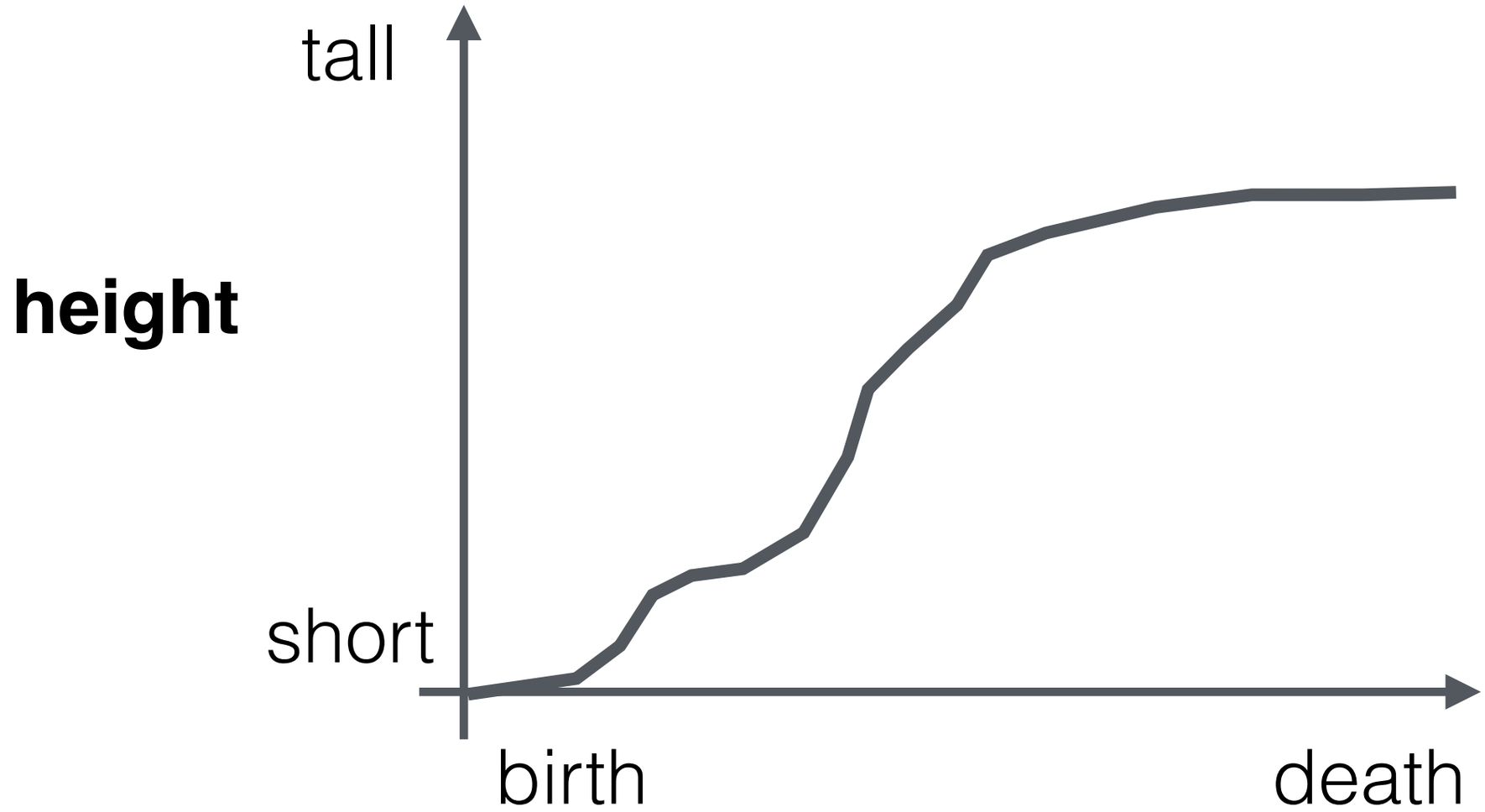


brief vocabulary lesson

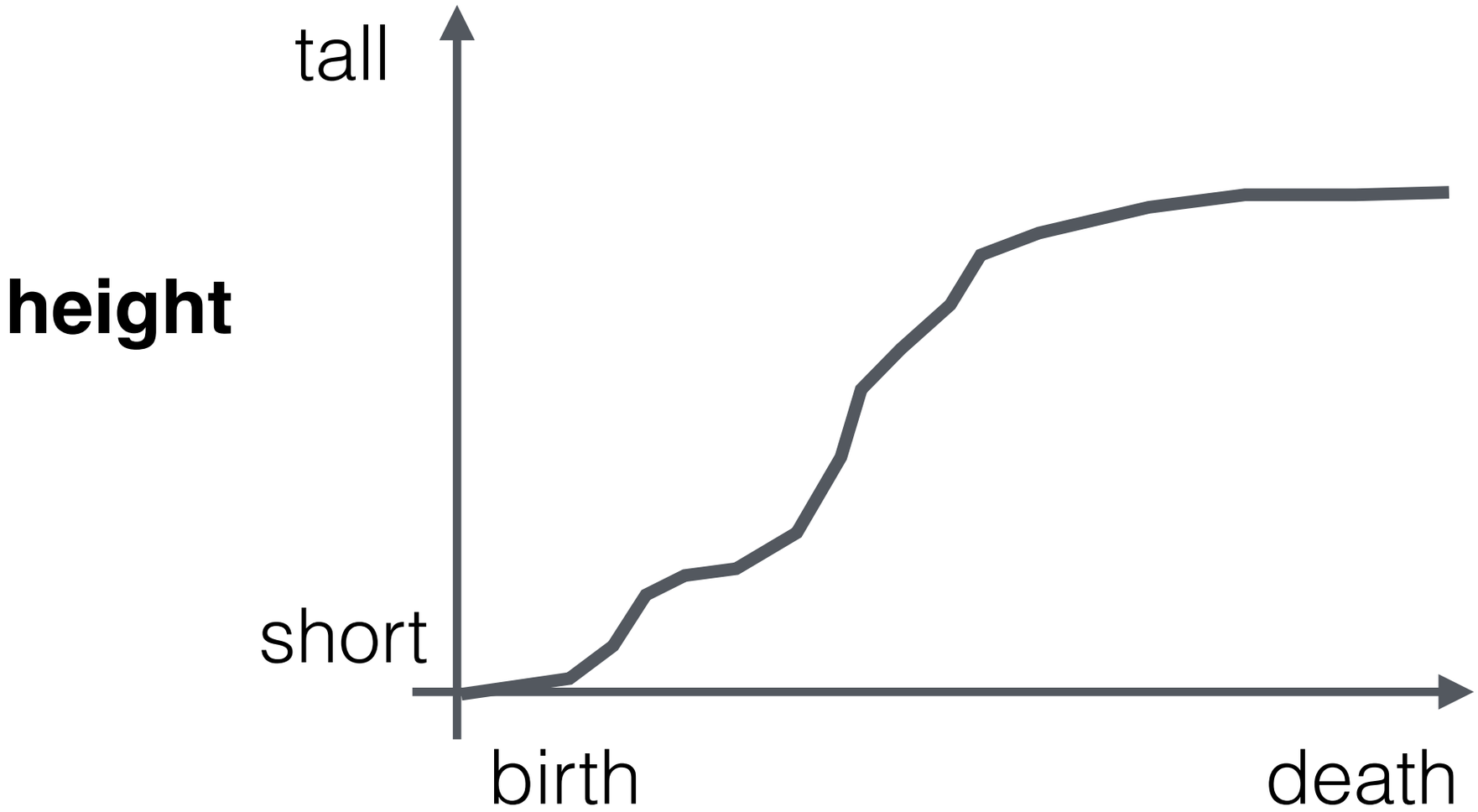
Semiotics: the study of signs and symbols in communication

jradavenport

Everything must have **meaning**



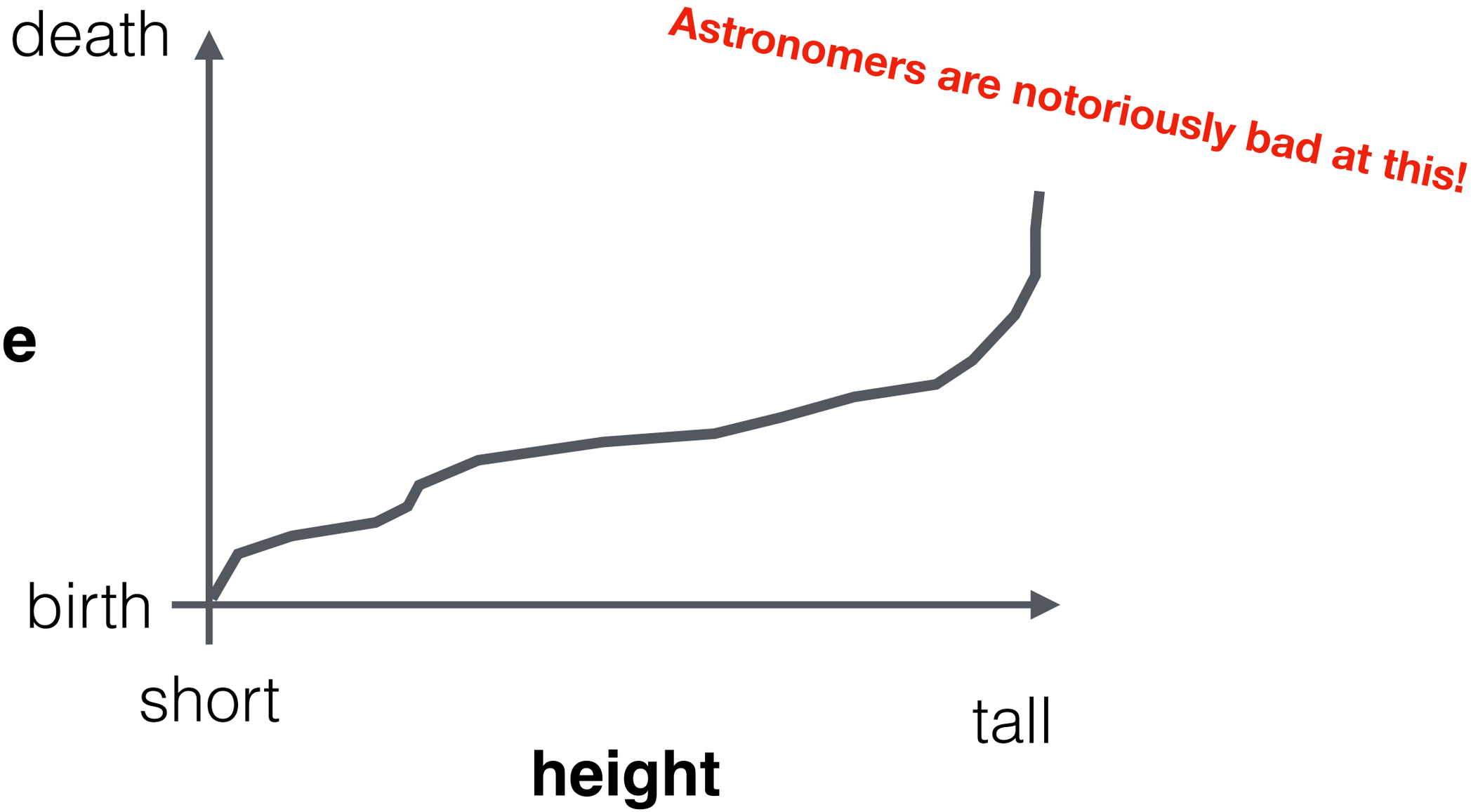
Everything must have **meaning**



possible meanings: growth over time, growth-spurts,

Everything must have **meaning**

Everything must have **meaning**



possible meanings: …tall people die?

Point #5

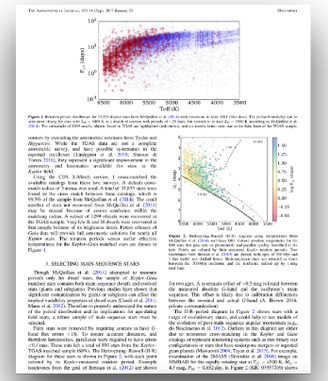
Use value-added meaning when possible!

- Repeat layouts/designs
- give colors/shapes meanings
- re-use colors consistently

Point #5 *restated*

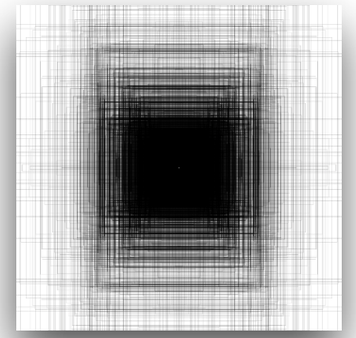
If they’re thinking about your plot,
they’re not thinking about your science

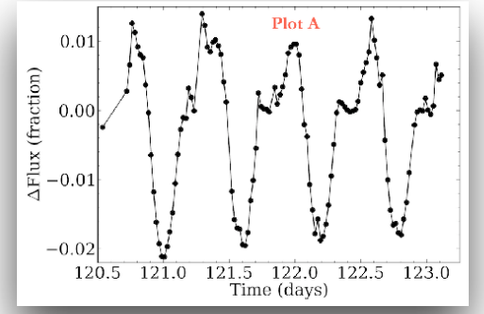
CONCLUSIONS



Your plots (need to) tell the story at a glance!

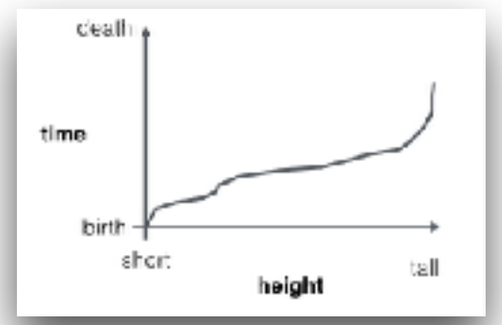
Make multiple versions of a plot!





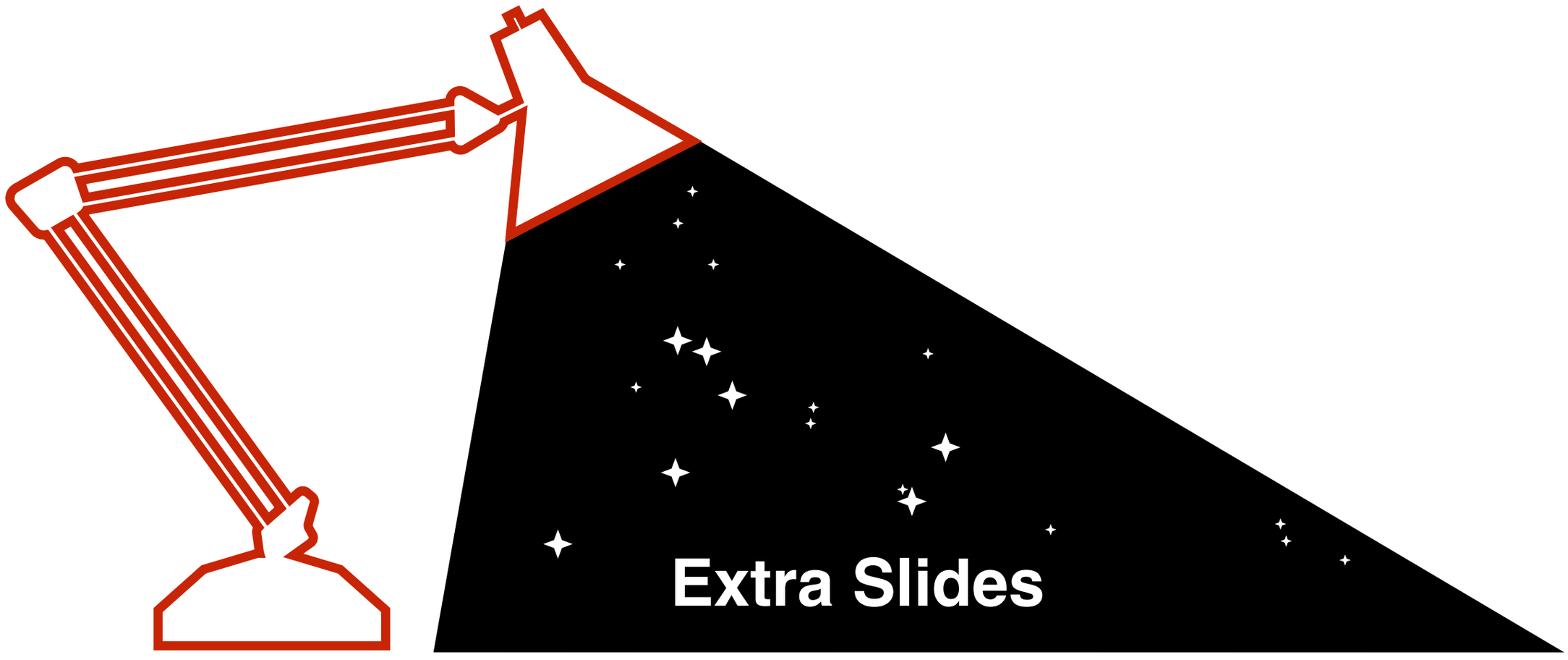
Don’t tool shame

Try to use good design principles, especially **contrast**.



If they’re thinking about your plot, they’re not thinking about your science

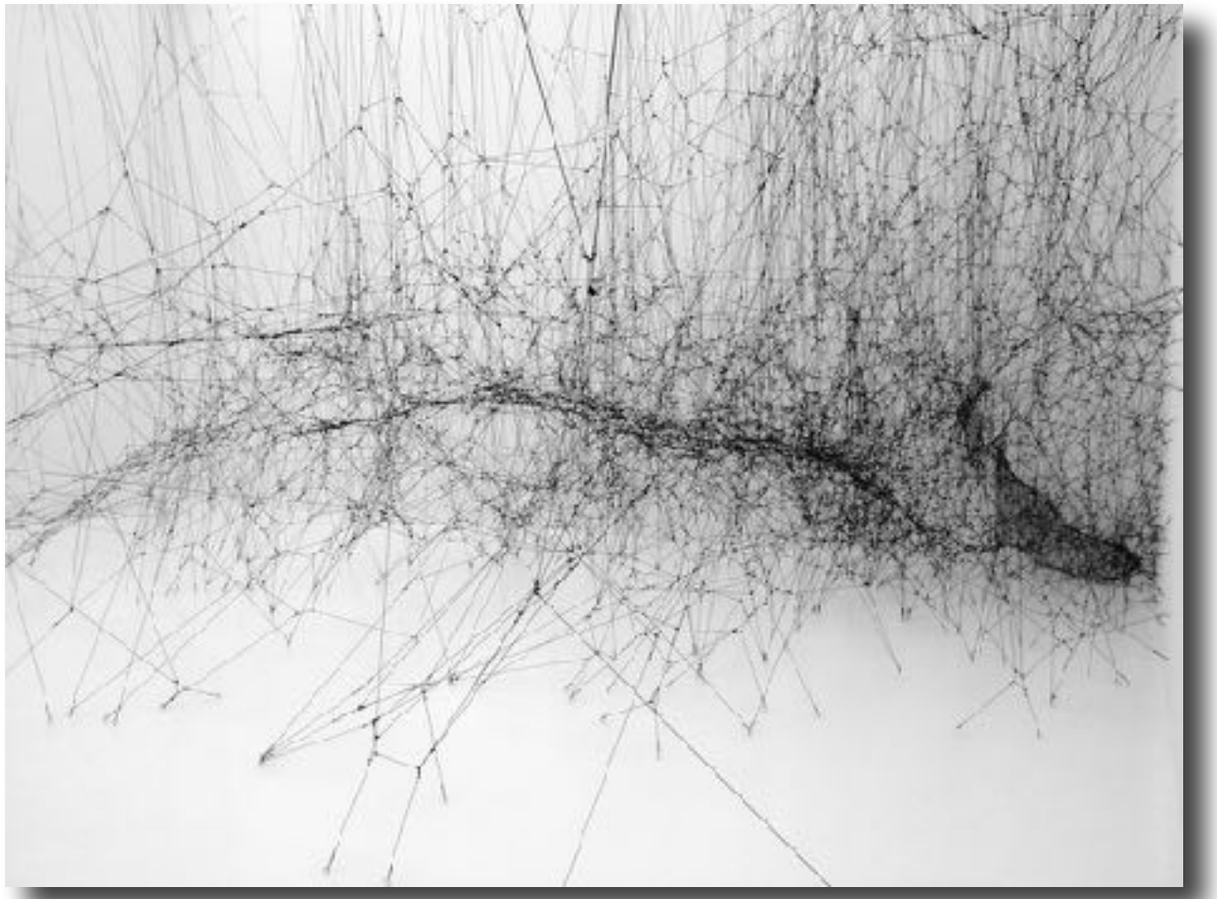
jradavenport

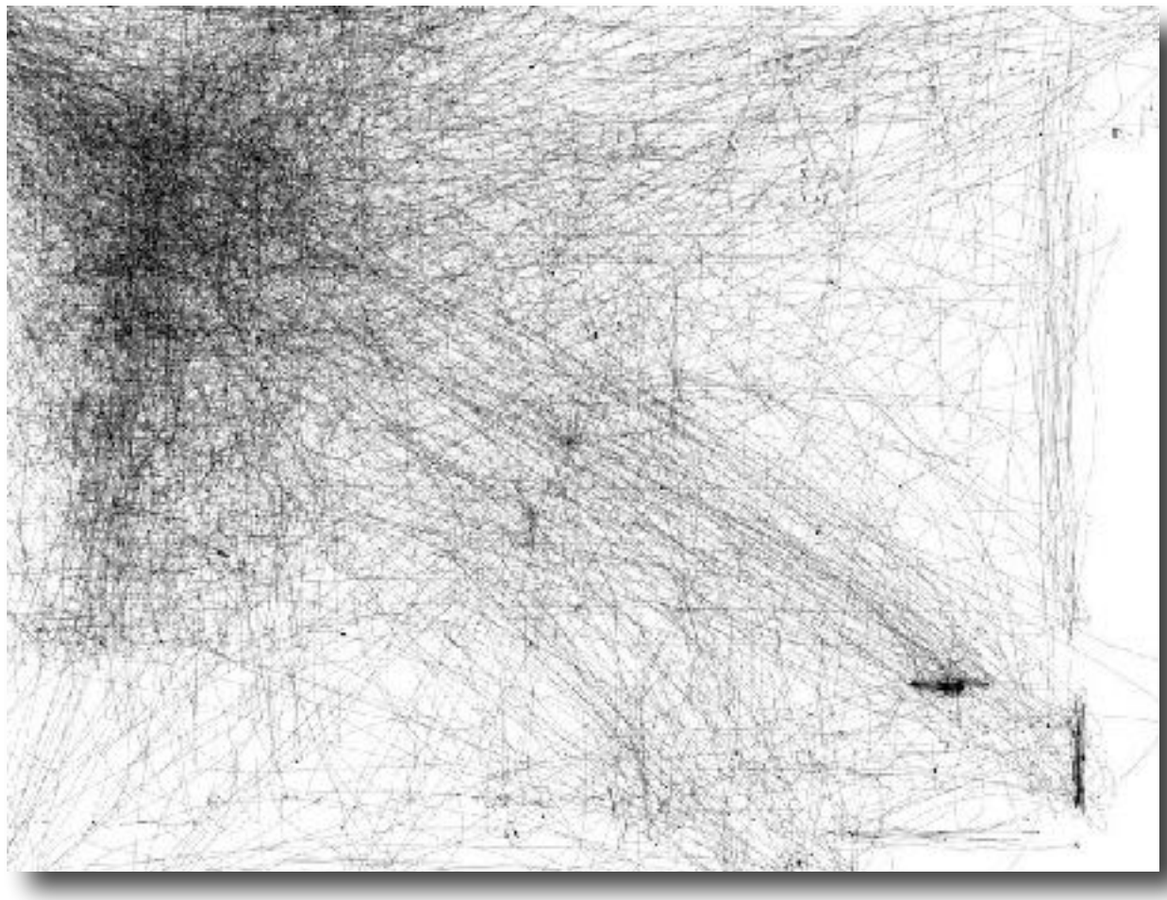


jradavenport

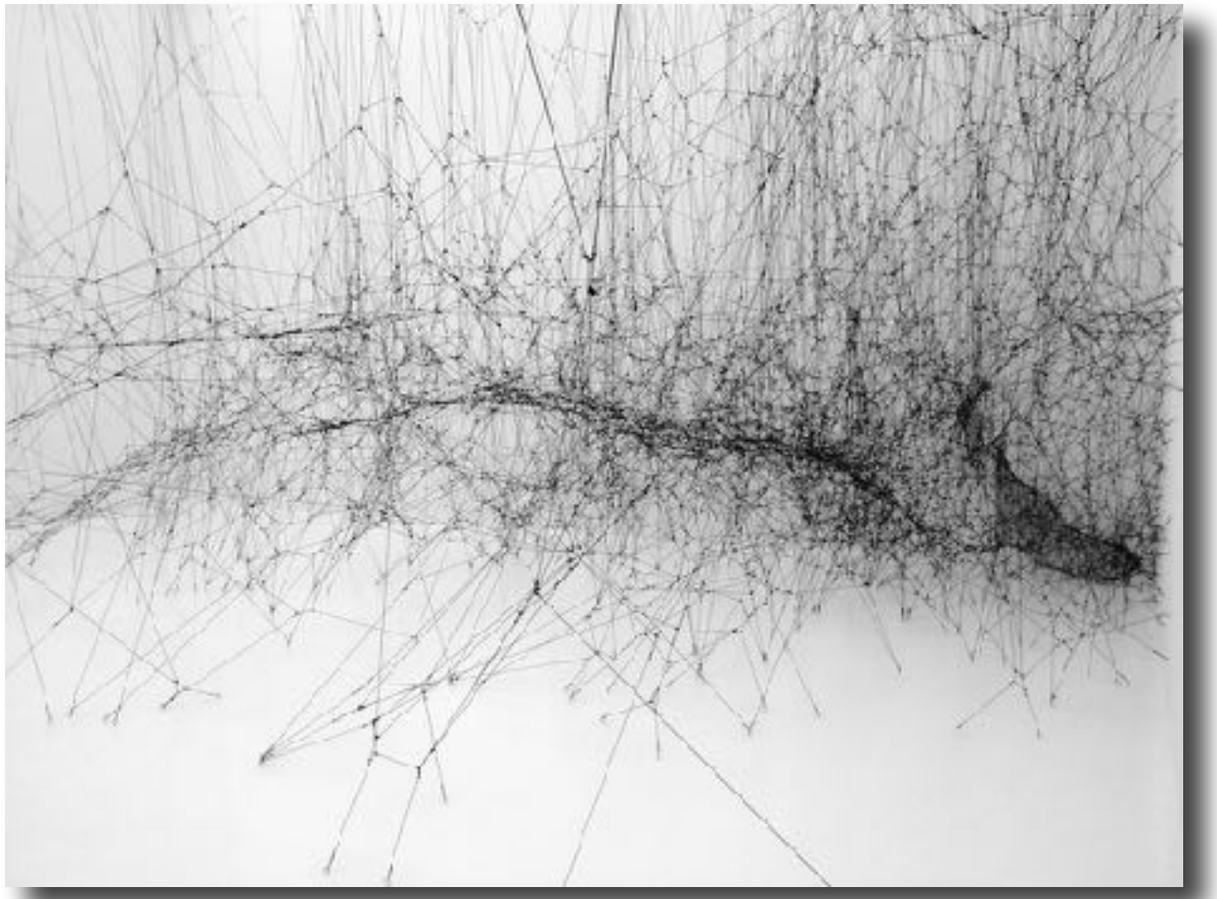
Let’s play a game...

Art or Data?



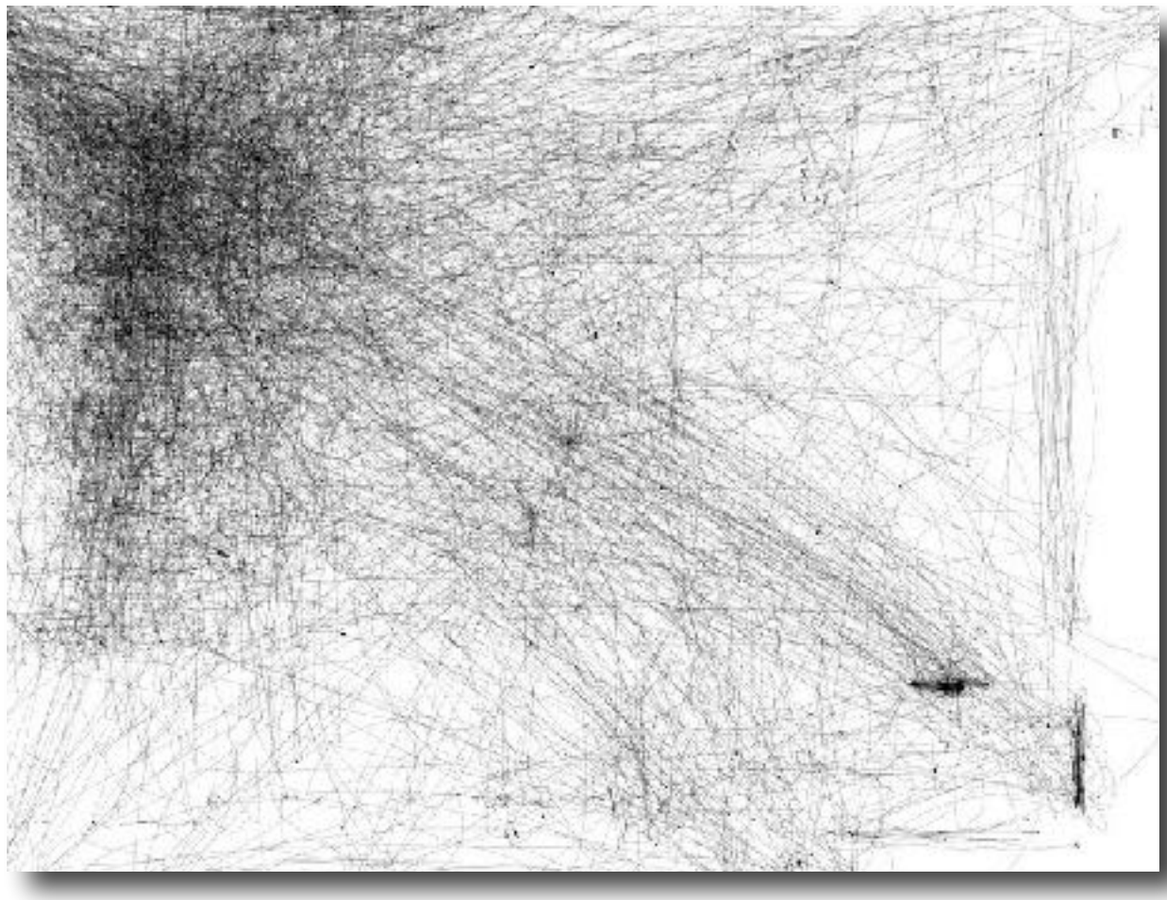


Art



"14 Billions" (2010)
by Tomás Saraceno

Data

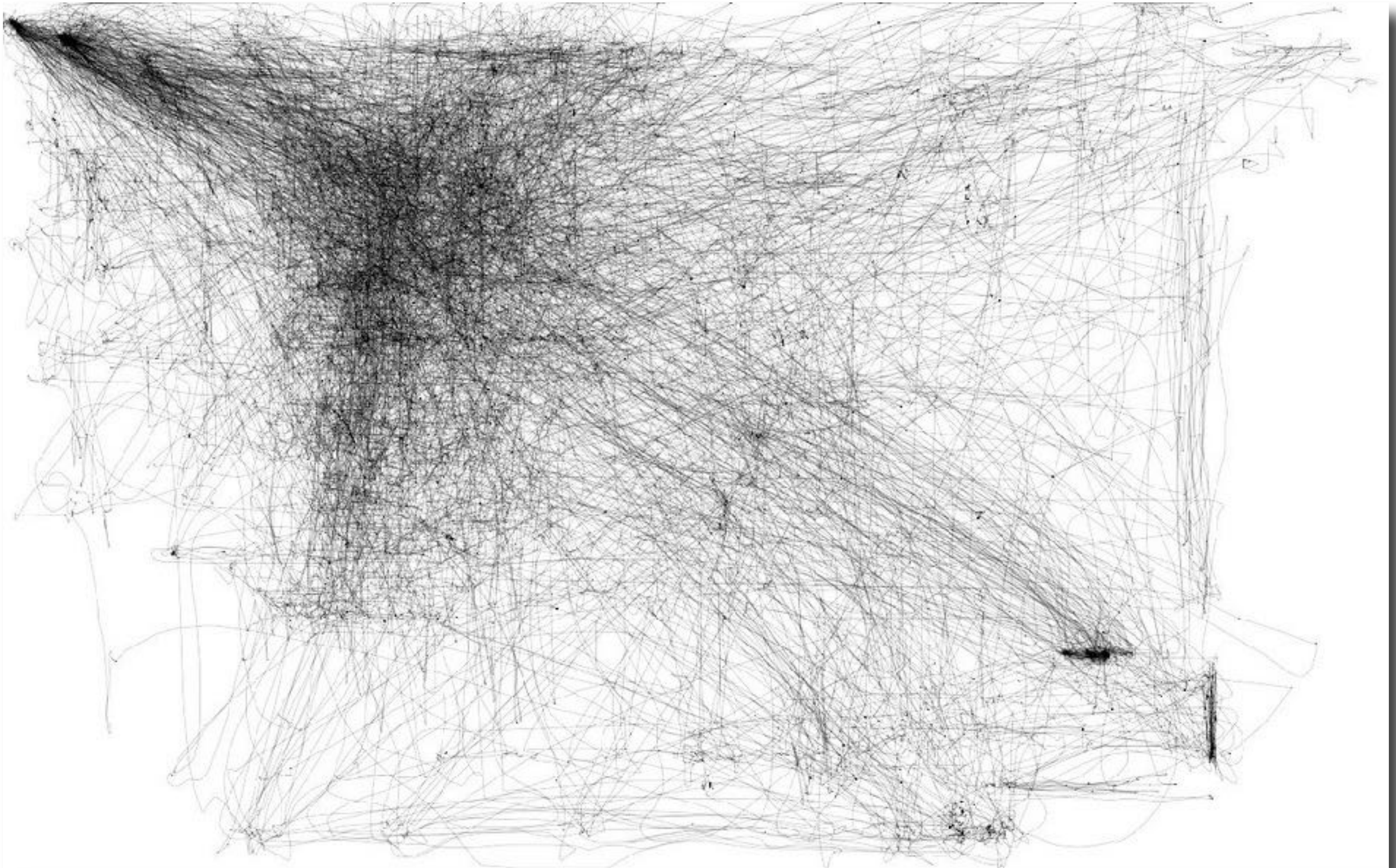


4 hours using Eclipse,
traced with IOGraph

Art

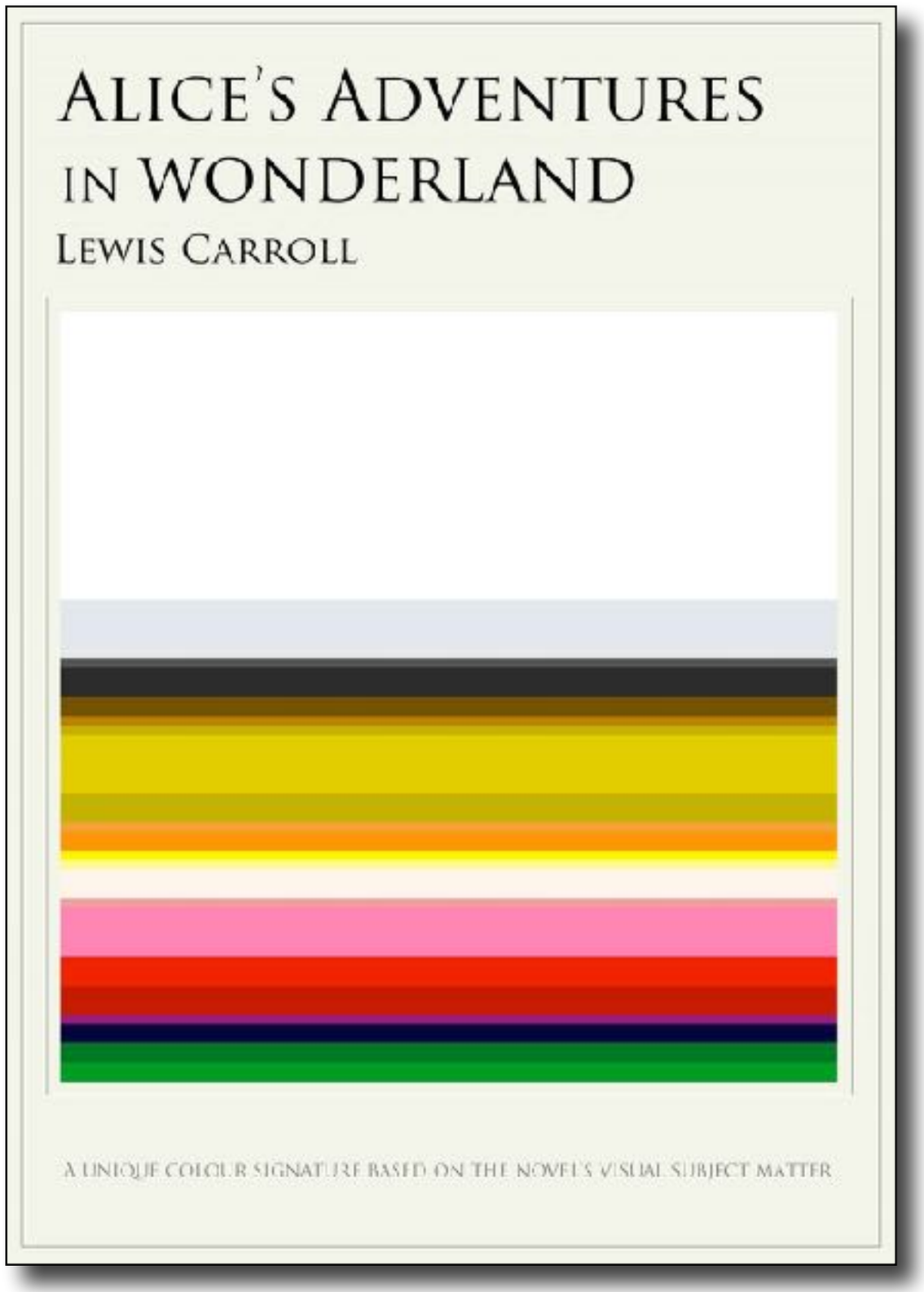
“14 Billions” (2010)
by Tomás Saraceno

Data



4 hours using Eclipse,
traced with IOGraph

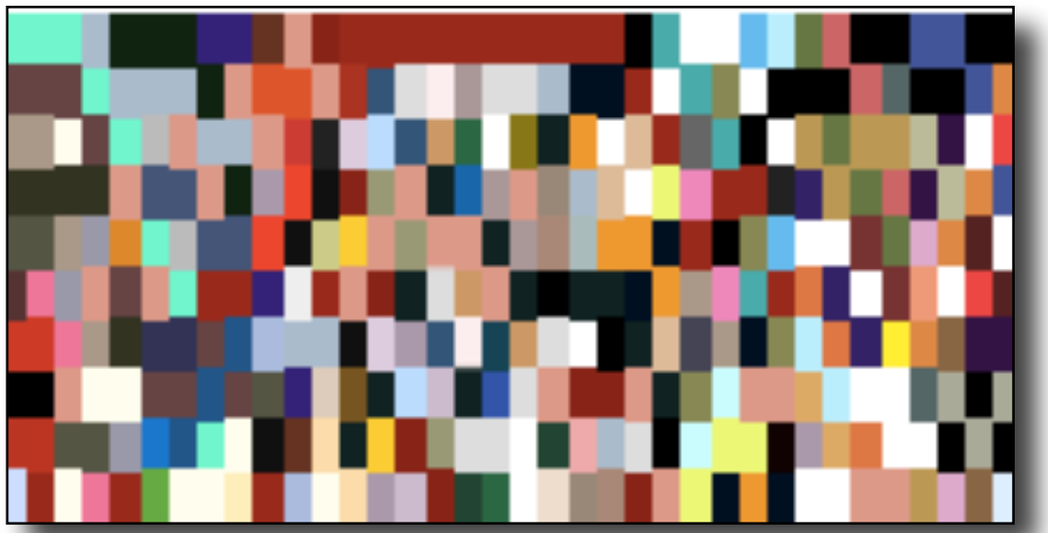
Data

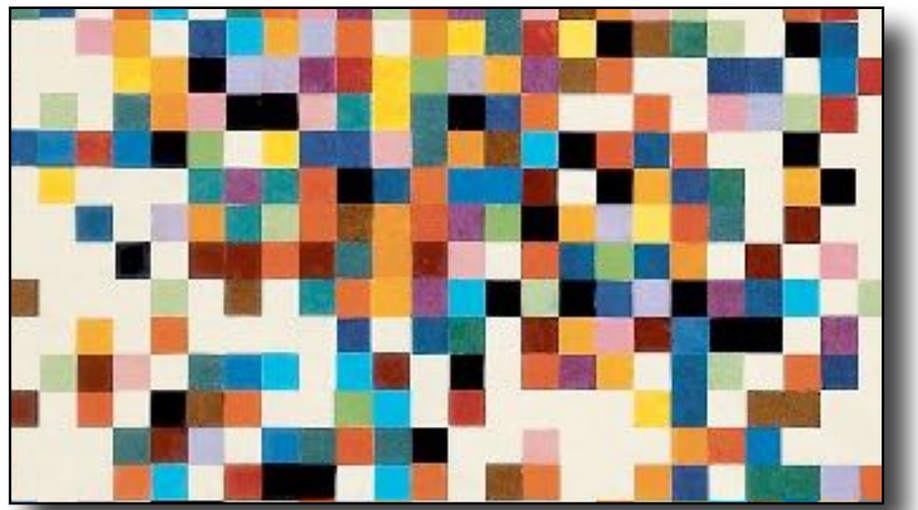


Jaz Parkinson

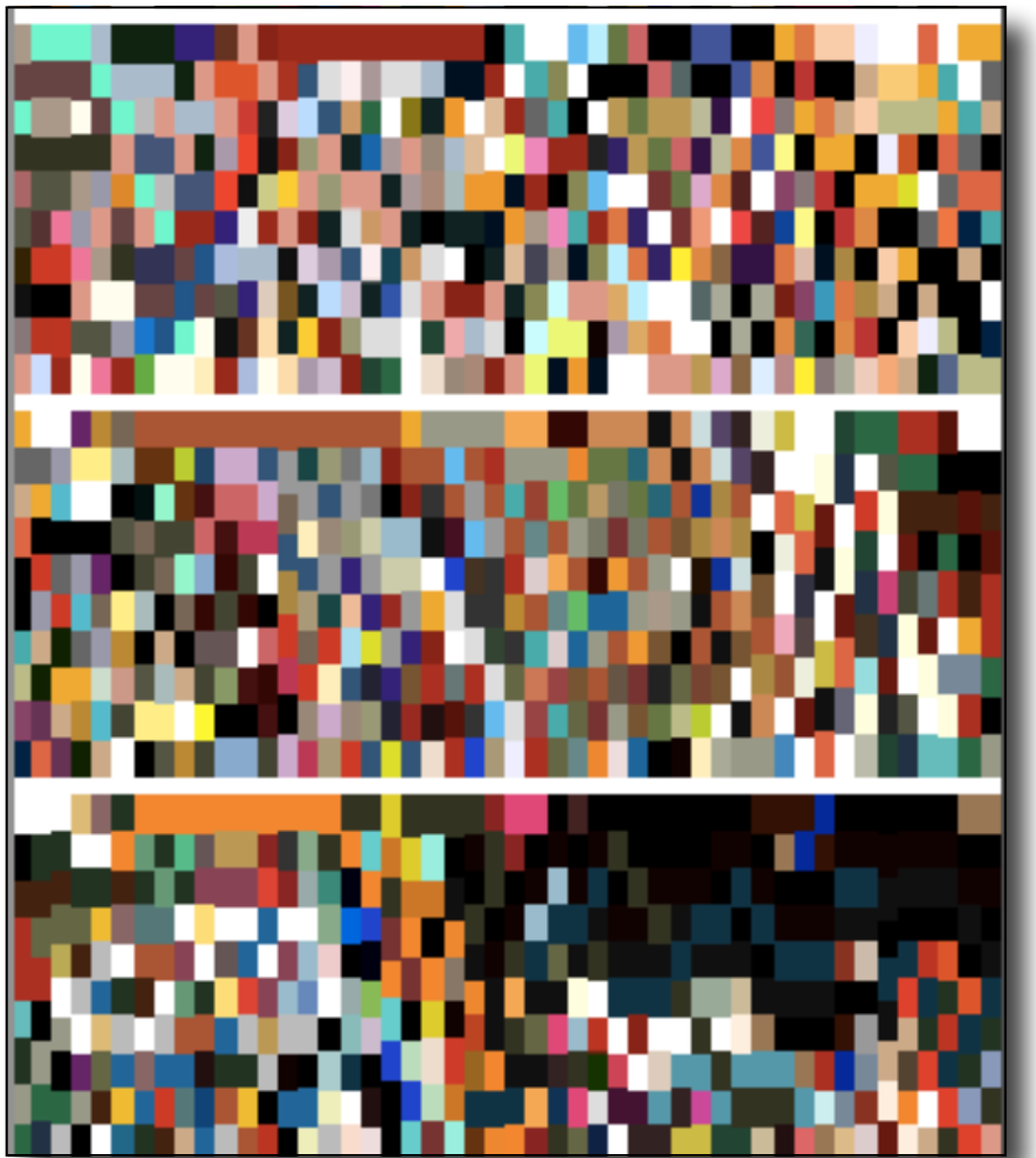
Art

Ellsworth Kelly



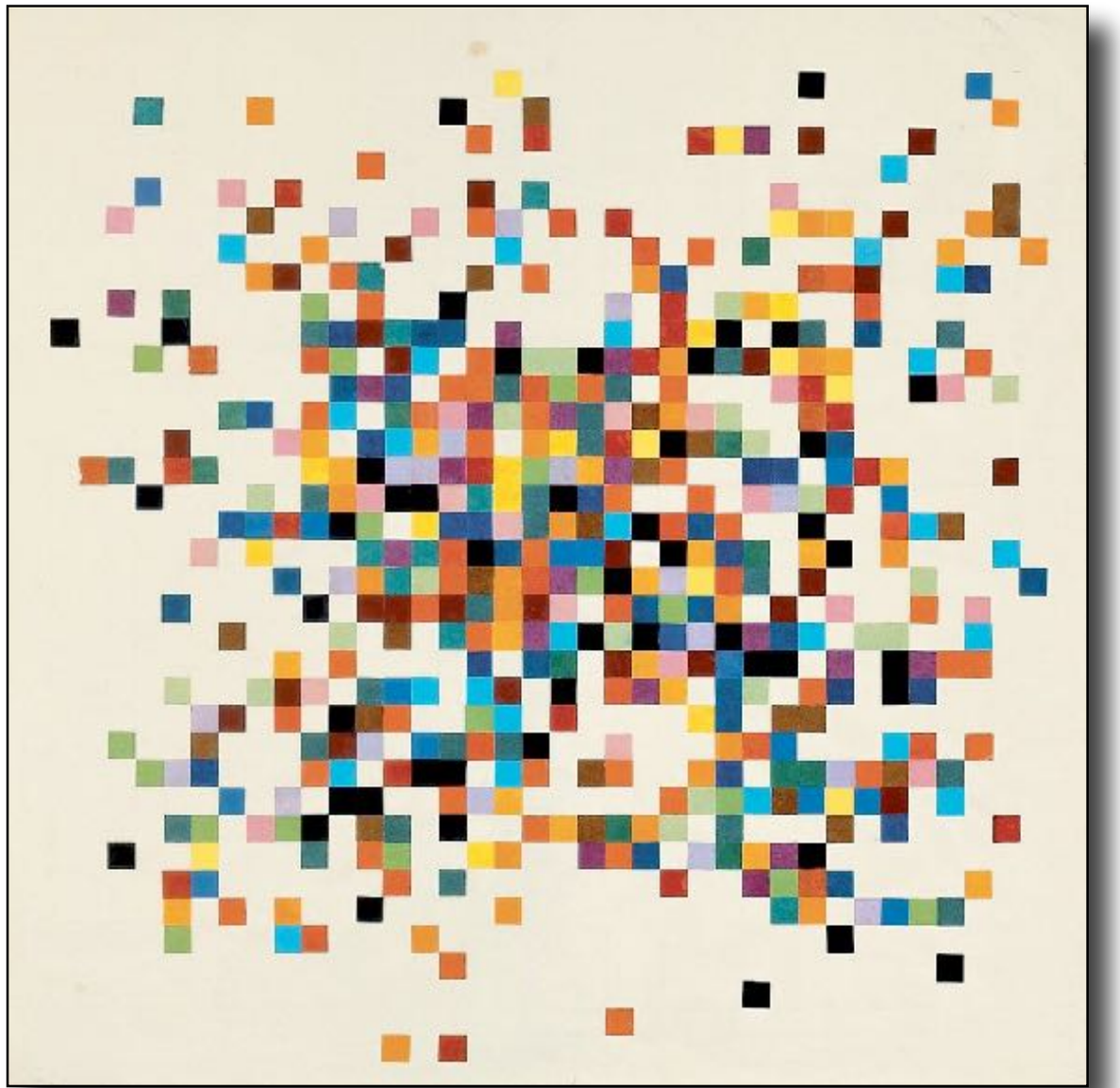


Data



James Davenport

Art



Ellsworth Kelly