



Please scan this code to register your presence on Ilias  
for contact-tracing purposes. Your Ilias identity will be used.  
– Only do this if you are *physically present!* –

# DATA LITERACY

## LECTURE 12

### HOW TO MAKE A FIGURE

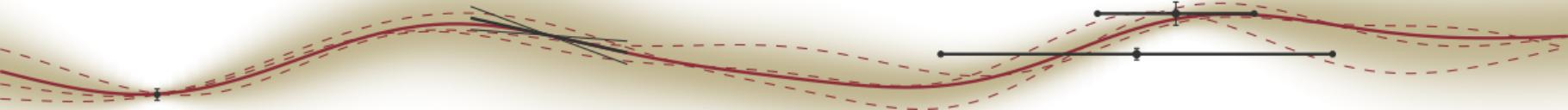
Philipp Hennig

24 January 2022

EBERHARD KARLS  
**UNIVERSITÄT**  
TÜBINGEN



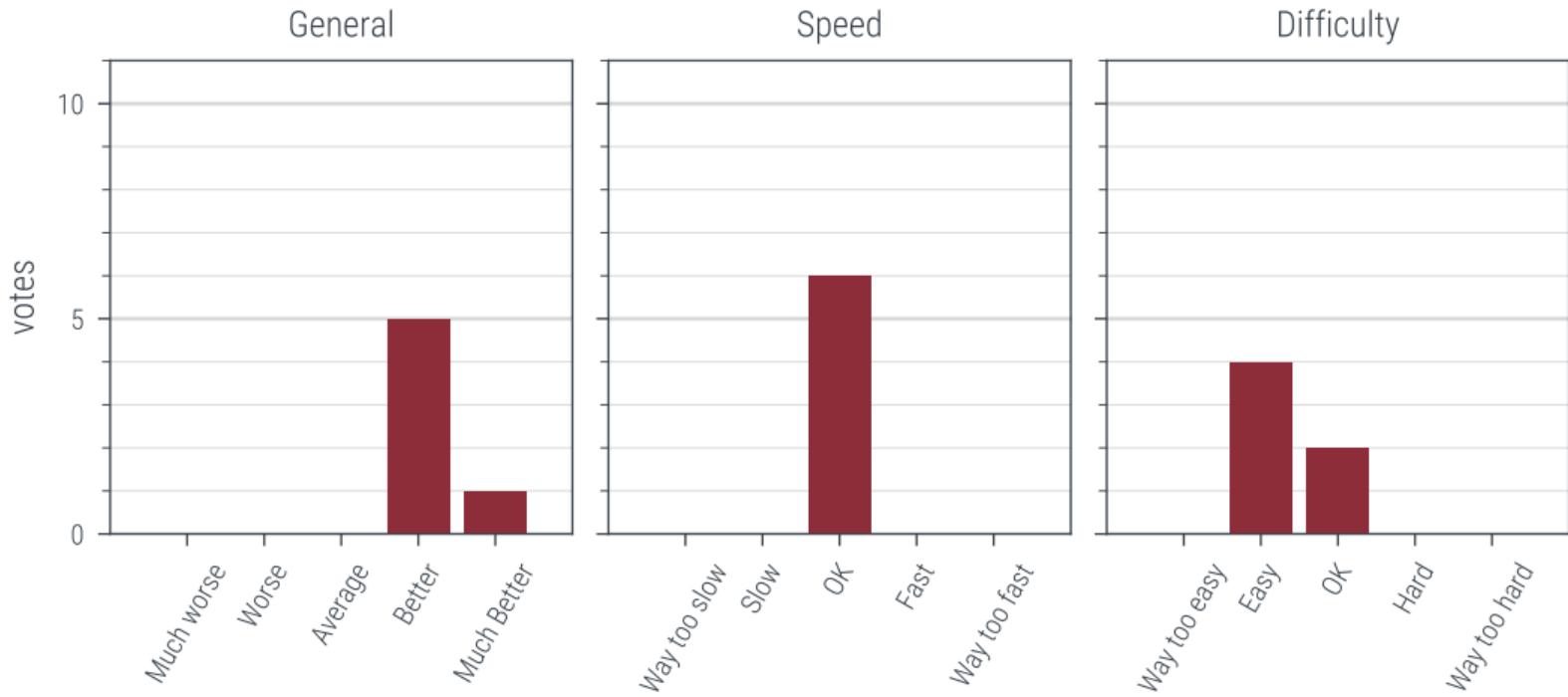
FACULTY OF SCIENCE  
DEPARTMENT OF COMPUTER SCIENCE  
CHAIR FOR THE METHODS OF MACHINE LEARNING





# Feedback

quantitative





# Detailed Feedback

your answers

## Things you *didn't* like



## Things you liked

- ▶ industry part
- ▶ personal examples
- ▶ using examples from own practice, explaining the whole data management process

## Things you *didn't* understand





Making figures for an academic paper or a technical presentation is a **craft**.  
It's neither a hard **science** nor a liberal **art**.

- ▶ find what works for you
- ▶ feel free to disagree with my aesthetic preferences

Today:

**Why** should you care about figures?

**What** should a plot look like?

**How** do you get results into a paper?



# Why should you care about figures?



# Brains Vary

Cater to those in your audience with a visual mind

Which part of today's lecture did you enjoy **most**?

The algebraic explanations

Which part of today's lecture did you **least** enjoy?

The examples (bread and butter??)

Which part of today's lecture did you enjoy **most**?

- The example with butter and bread
- The explanation why Gaussians are so important

Which part of today's lecture did you **least** enjoy?

Transformation to Gaussian

Which part of today's lecture did you enjoy **most**?

graph examples

Which part of today's lecture did you enjoy **most**?

- fanciest graphs ever seen
- great way of explaining stuff

Which part of today's lecture did you **least** enjoy?

long eq. with  $\phi$

Communication modes in scientific writing / teaching:

- ▶ mathematical formulae
- ▶ graphs and figures
- ▶ examples and experiments
- ▶ code and experimental protocols



# Graphical People are not Stupid

don't feel bad about your visual way of thinking

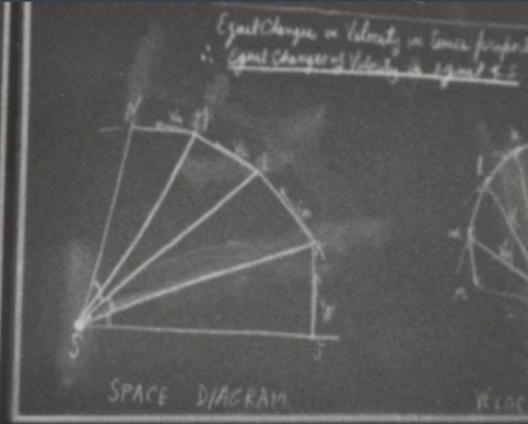
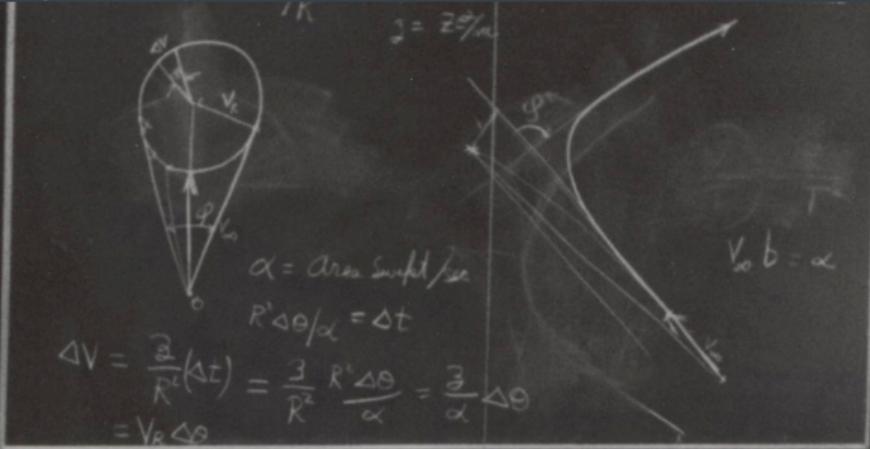
$\rightarrow + FP$ , then it is tangent to ellipse

$G' \text{ (image of } F)$

$$FP + F'P = FP + GP = FG'$$

$$FQ + F'Q = FQ + G'Q > FG'$$

$\therefore Q$  lies outside (on, at) the ellipse



center  
(a) + (b), then (c)

ce toward Sun

$\therefore$  Force as  $1/r^2$

1 times are

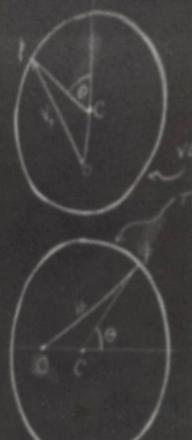
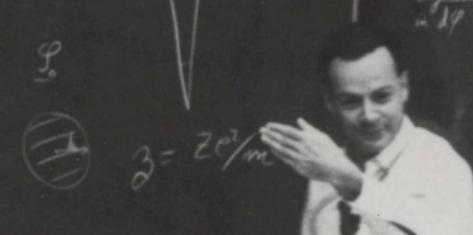
$$V_R = \frac{r}{\alpha}$$

$$v_b = c \sqrt{\frac{3}{2}} \tan \frac{\phi}{2}$$

$$\tan \frac{\phi}{2} = \frac{V_R}{V_\infty} = \frac{3}{\alpha V_\infty} = \frac{3}{b V_\infty}$$

X-section area for deflection  $> \phi$

$$\omega \pi b^2 \text{ or } \pi \beta^2$$





# Graphical People are not Stupid

don't feel bad about your visual way of thinking

[images: public, getty, mfo]

**Figure 1.** John Wheeler lecturing at a conference in Cambridge, UK, in 1971. Wheeler's style was to cover the blackboard with inspirational colored-chalk diagrams and phrases before the lecture, then work his way through them, one by one.



# Graphical People are not Stupid

don't feel bad about your visual way of thinking

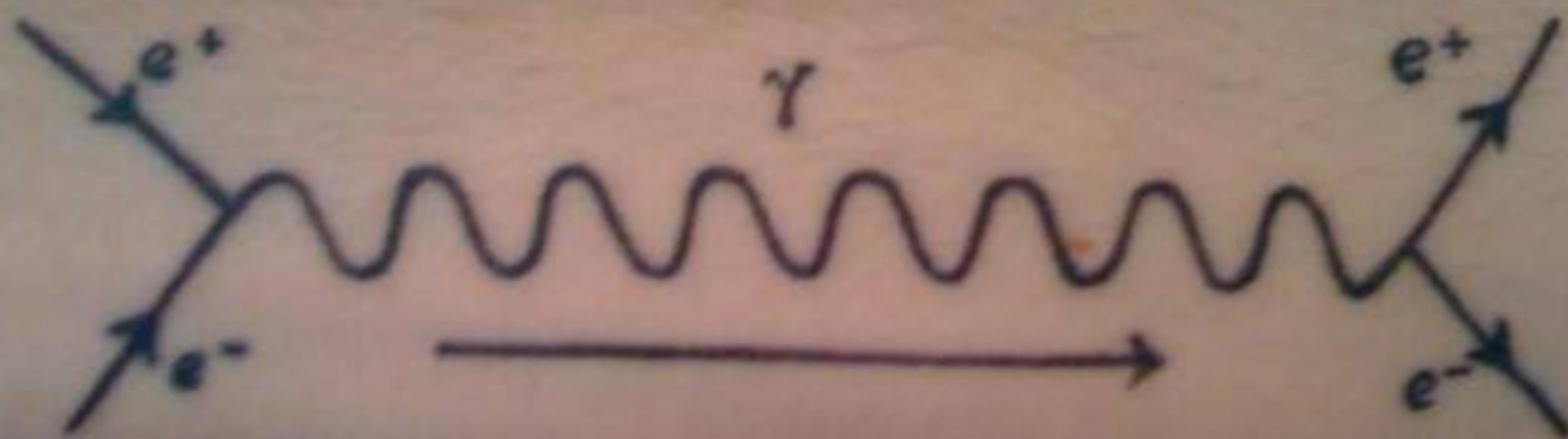


[images: public, getty, mfo]



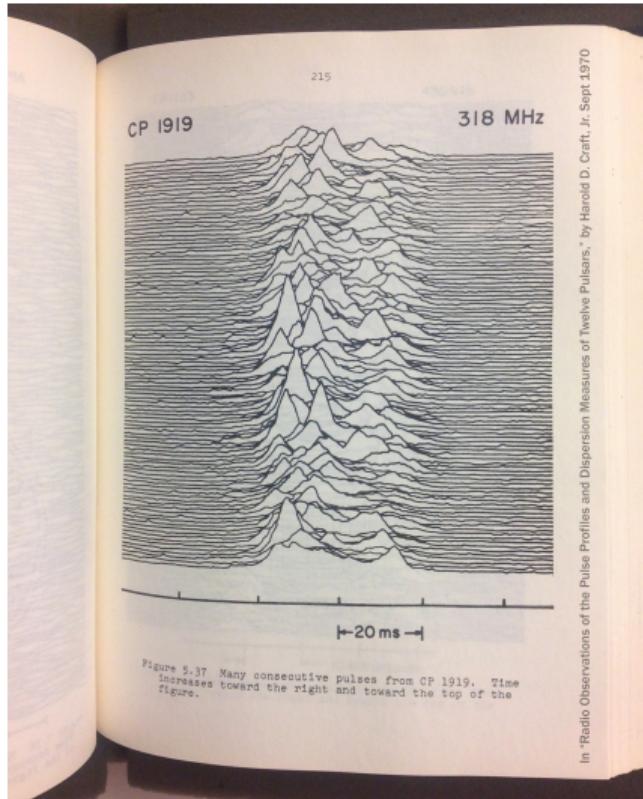
# A Good Visualization Can Transcend the Science

Presenting your idea is half the trick



# A Good Visualization Can Transcend the Science

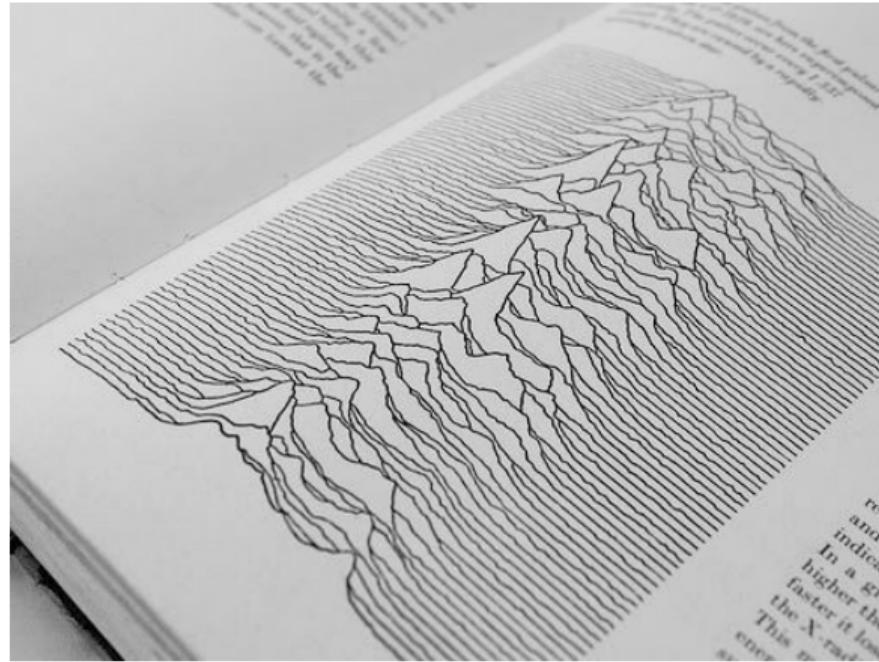
Presenting your idea is half the trick



*Radio Observations of the Pulse Profiles and Dispersion Measures  
of Twelve Pulsars*  
Harold D. Craft, Jr.  
PhD Thesis, Cornell U  
Sept. 1970

# A Good Visualization Can Transcend the Science

Presenting your idea is half the trick



The Cambridge Encyclopaedia of Astronomy  
1977 (no image credit!)

# A Good Visualization Can Transcend the Science

Presenting your idea is half the trick



1979

# A Good Visualization Can Transcend the Science

Presenting your idea is half the trick



# A Good Visualization Can Transcend the Science

Presenting your idea is half the trick

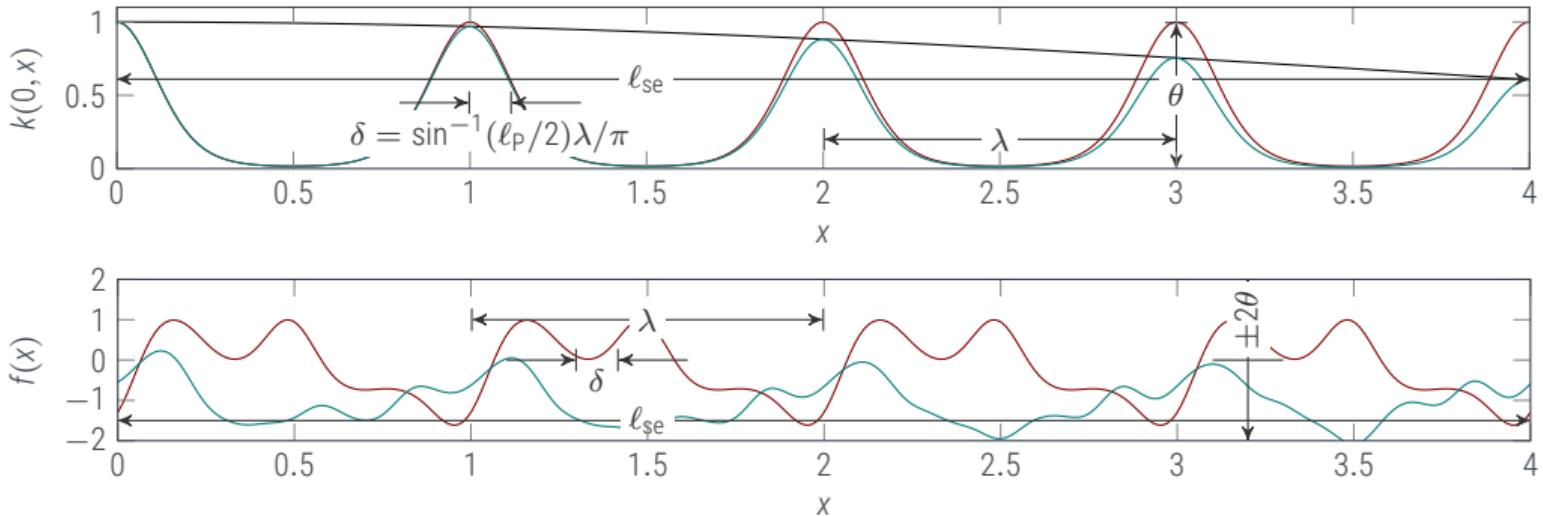




# How should your figure *look*? scientific and graphical considerations

# What are you trying to achieve?

A good paper tells a story, with math, **figures**, and experiments

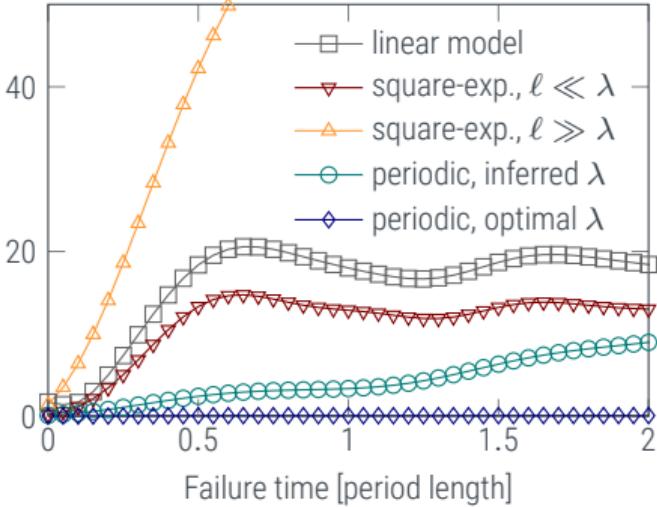
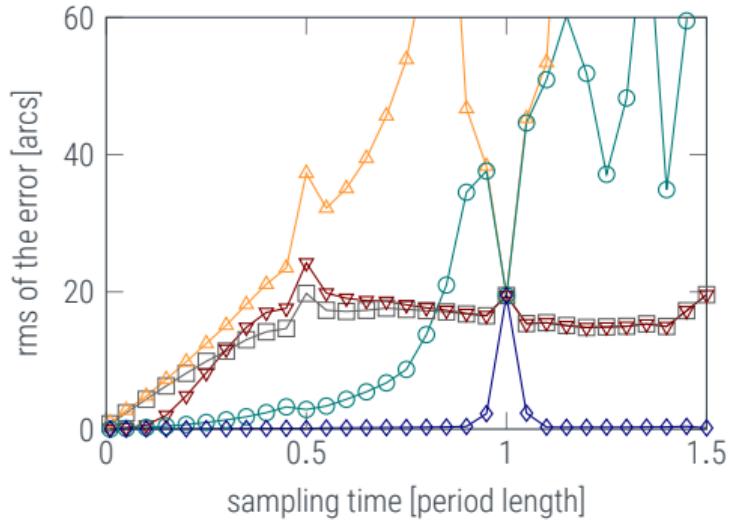


## Conceptual Sketches

- ▶ can be low-dim, simplified
- ▶ must be particularly clear
- ▶ are not just eye-candy!

# What are you trying to achieve?

A good paper tells a story, with math, **figures**, and experiments



## Empirical Results

- ▶ must not hide problems
- ▶ should be legible
- ▶ are a core part of the paper



# The Basic Rules

this should all go without saying

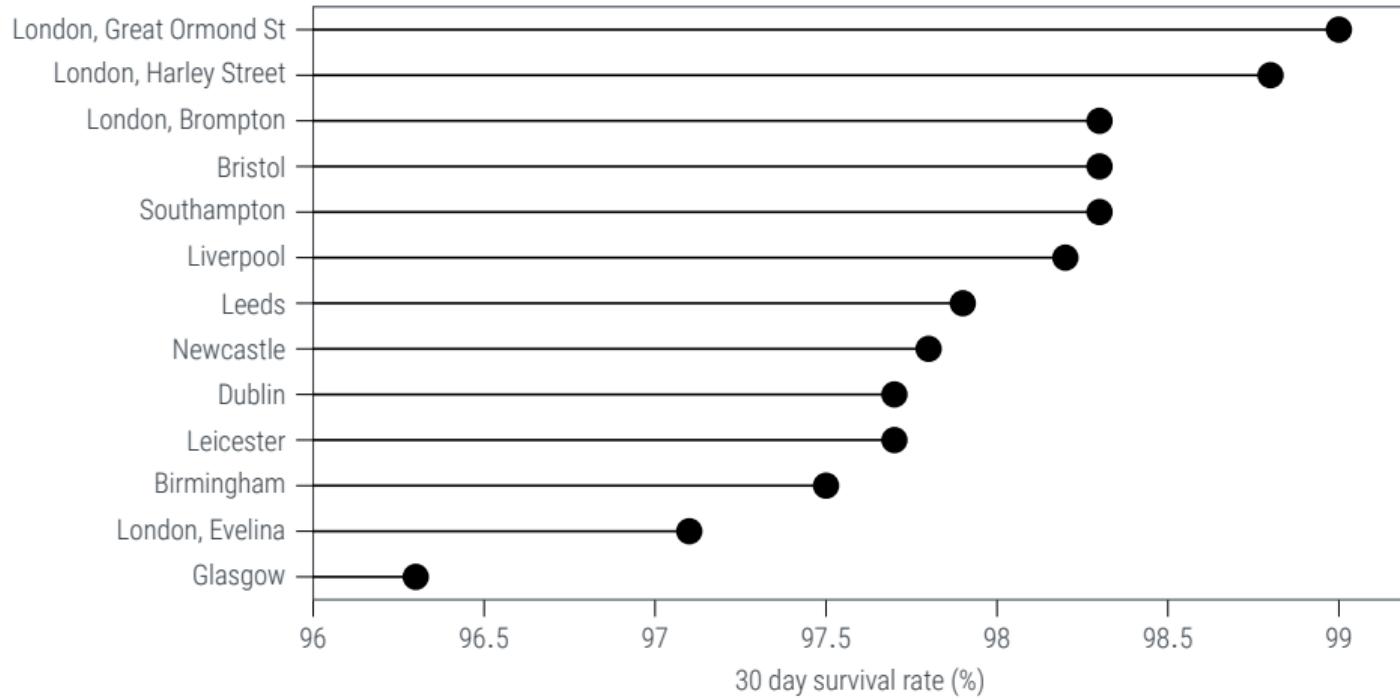
- ▶ proper **axis labels**
  - ▶ units!
  - ▶ right label for all lines?
- ▶ proper **axis limits**
  - ▶ if relevant, do not cut off 0
  - ▶ sometimes 1 is just as important!
  - ▶ sometimes, it can be better to cut off a bit of irrelevant stuff to zoom in on the crucial bits
- ▶ **legends** are important,  
unless they aren't
- ▶ **font size!**

# Some obvious things

or are they?



Survival rates of pediatric heart surgery in British hospitals

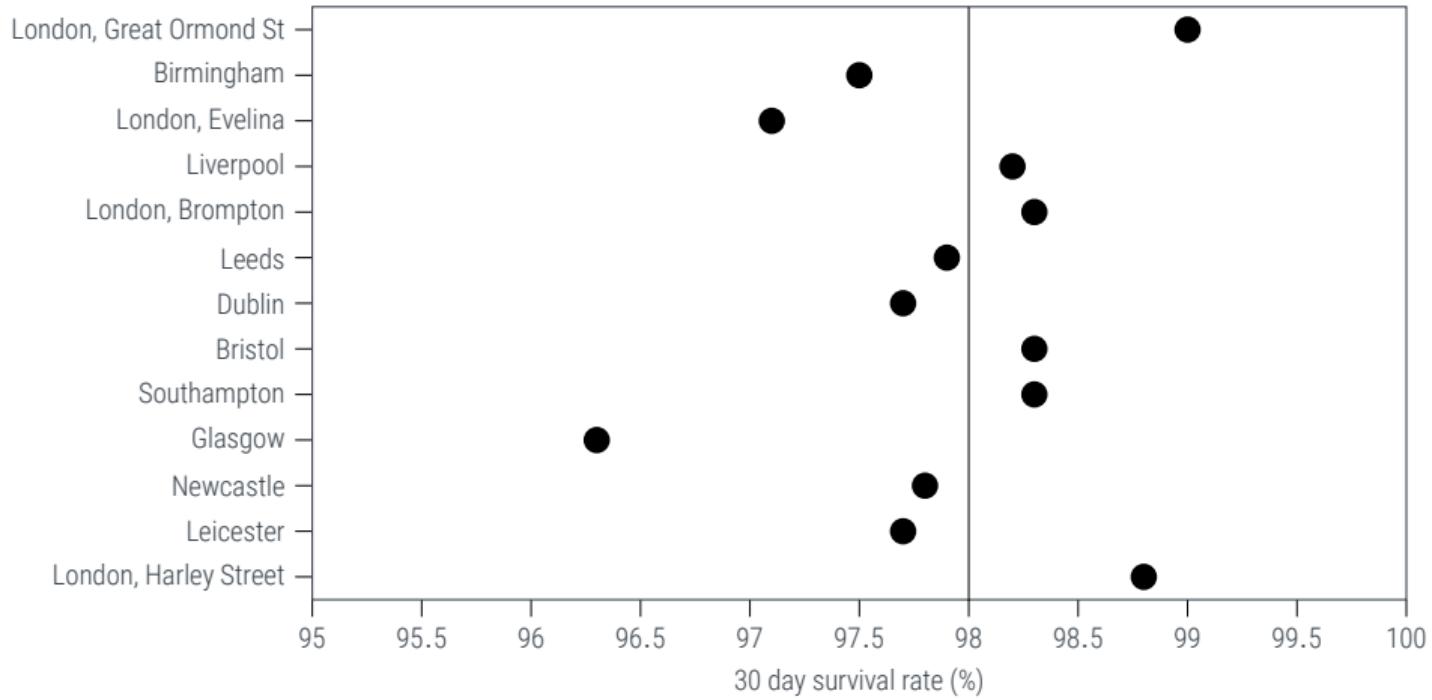


# Some obvious things

or are they?



Survival rates of pediatric heart surgery in British hospitals

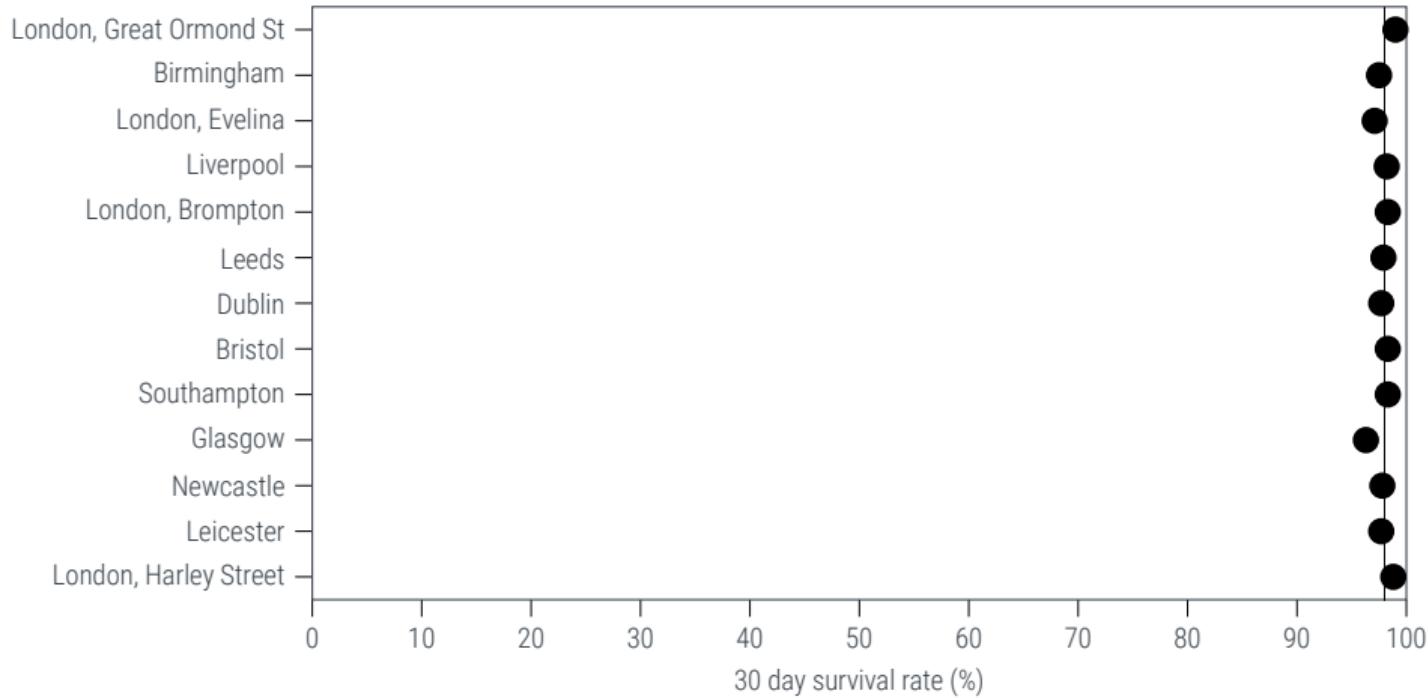


# Some obvious things

or are they?



Survival rates of pediatric heart surgery in British hospitals

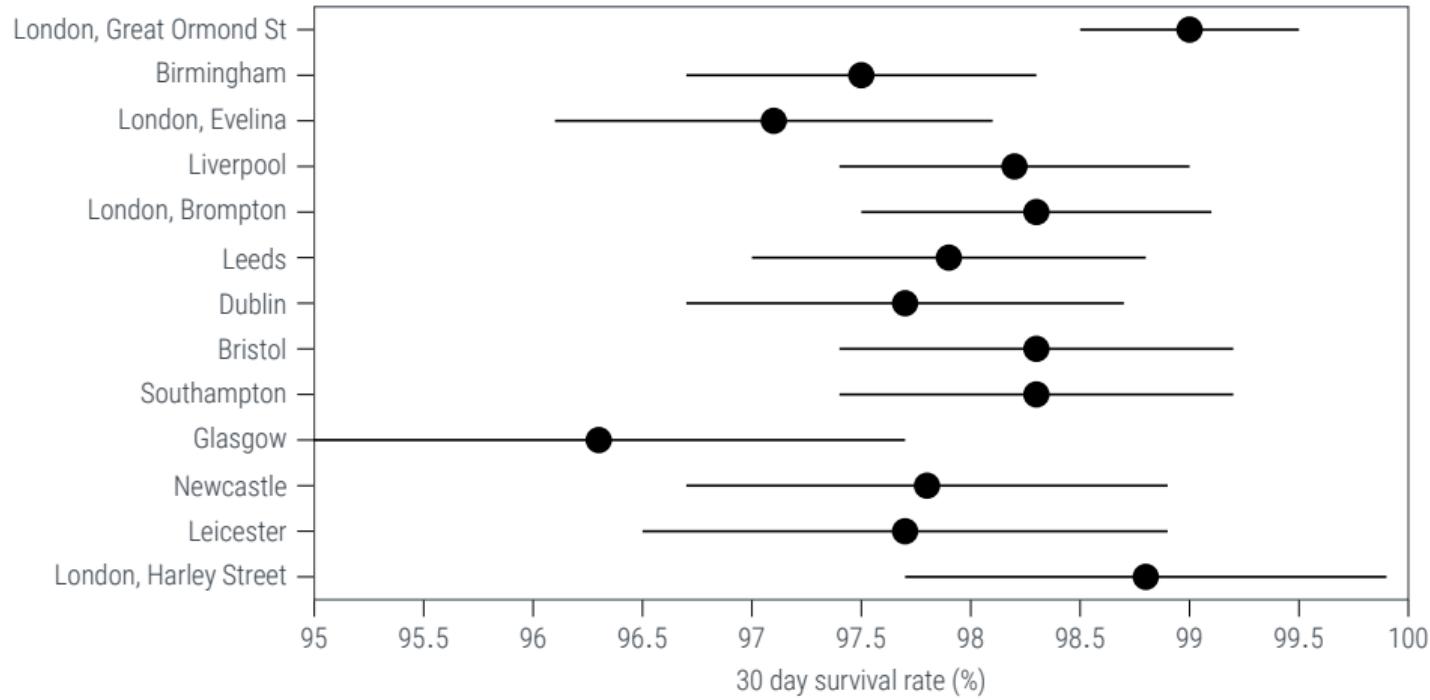


# Some obvious things

or are they?



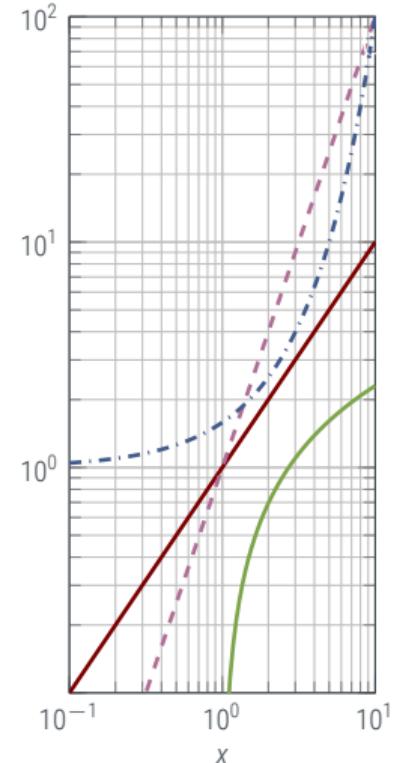
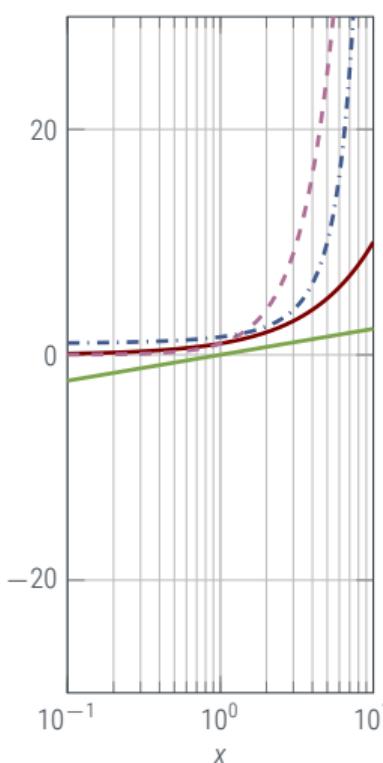
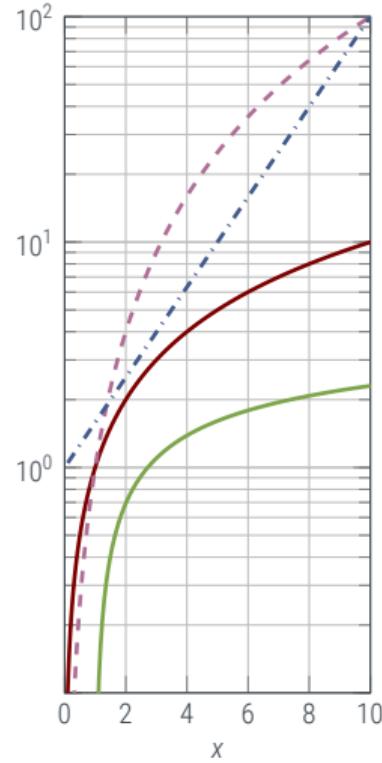
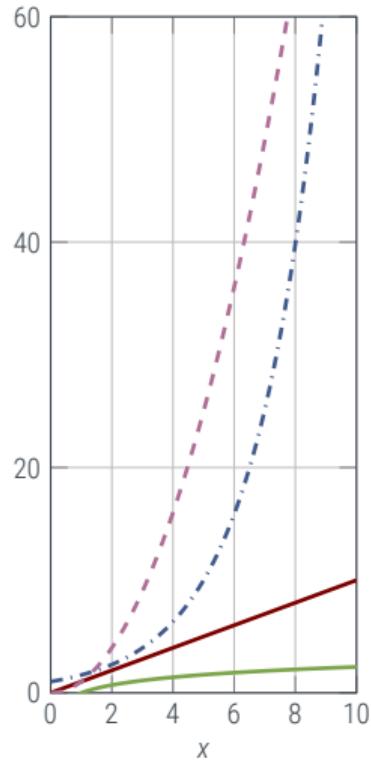
Survival rates of pediatric heart surgery in British hospitals





# Think about Scaling

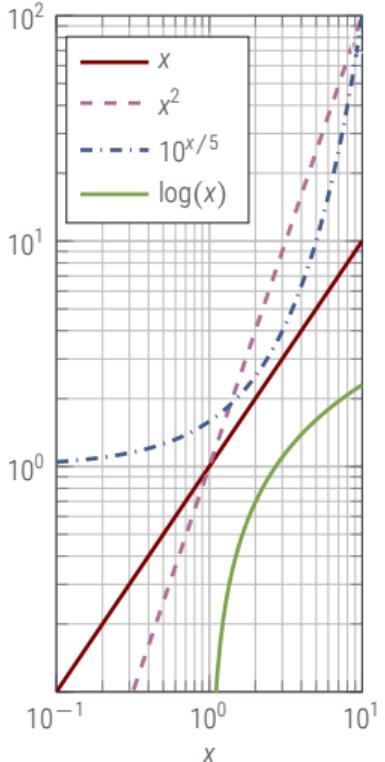
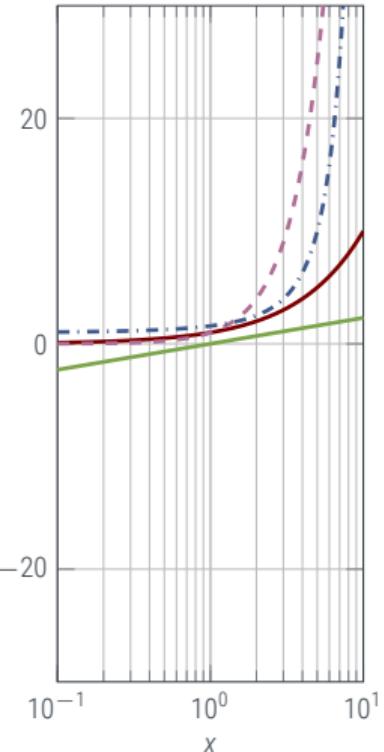
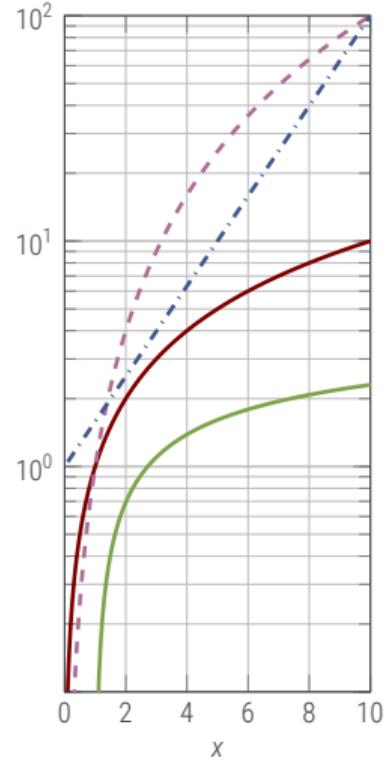
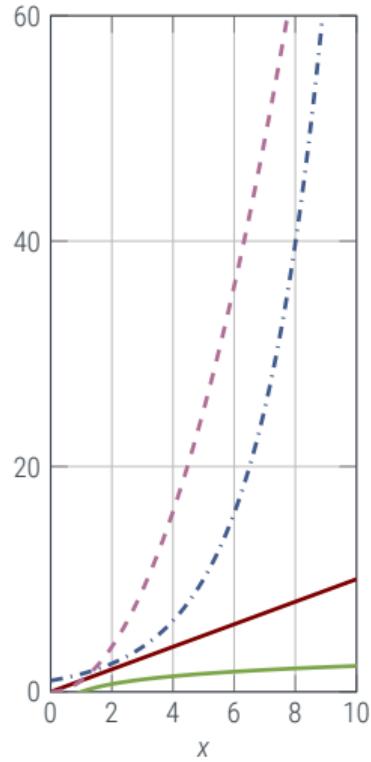
plotting can reveal structure





# Think about Scaling

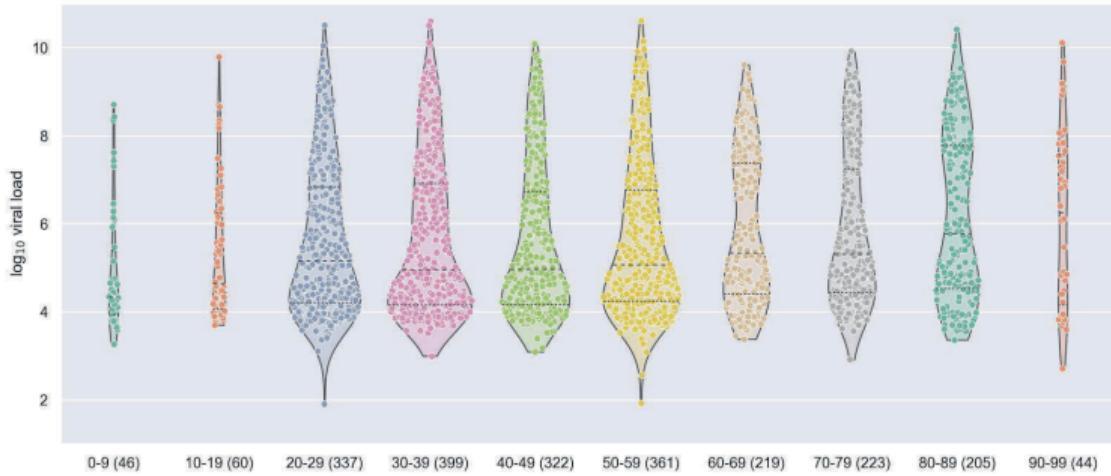
plotting can reveal structure





# Let the Data Speak ...

a figure is both a rhetorical tool and a documentation of research



An analysis of SARS-CoV-2 viral load by patient age

Jones, Mühlemann, Veith, Biele, Zuchowski, Hofmann, Stein, Edelmann, Corman, Drosten, 2020  
(preprint)



Christian H.M. Drosten

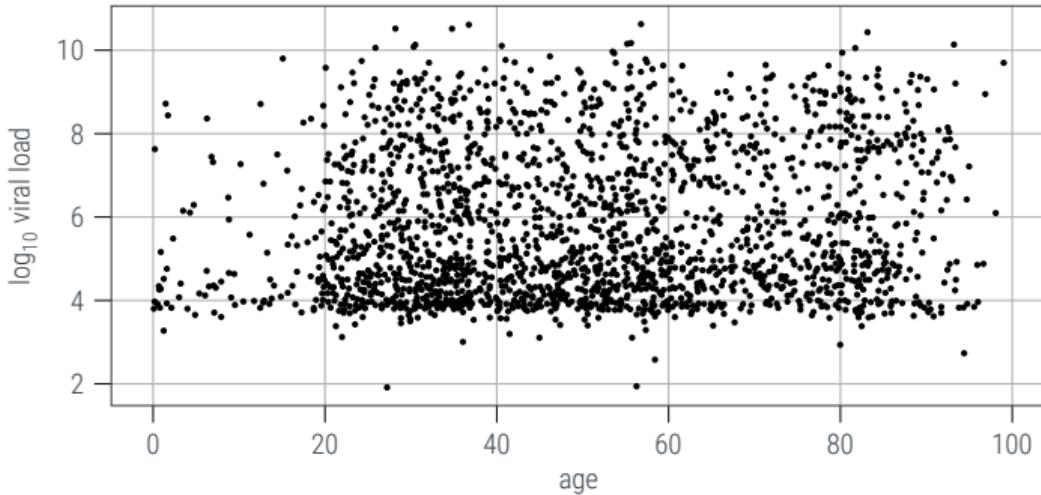
1972-

[https://virologie-ccm.charite.de/fileadmin/user\\_upload/microsites/m\\_cc05/virologie-ccm/dateien\\_upload/Weitere\\_Dateien/Charite\\_SARS-CoV-2\\_viral\\_load\\_2020-06-02.pdf](https://virologie-ccm.charite.de/fileadmin/user_upload/microsites/m_cc05/virologie-ccm/dateien_upload/Weitere_Dateien/Charite_SARS-CoV-2_viral_load_2020-06-02.pdf)



# Let the Data Speak ...

a figure is both a rhetorical tool and a documentation of research



## An analysis of SARS-CoV-2 viral load by patient age

Jones, Mühlemann, Veith, Biele, Zuchowski, Hofmann, Stein, Edelmann, Corman, Drosten, 2020  
(preprint)



Christian H.M. Drosten

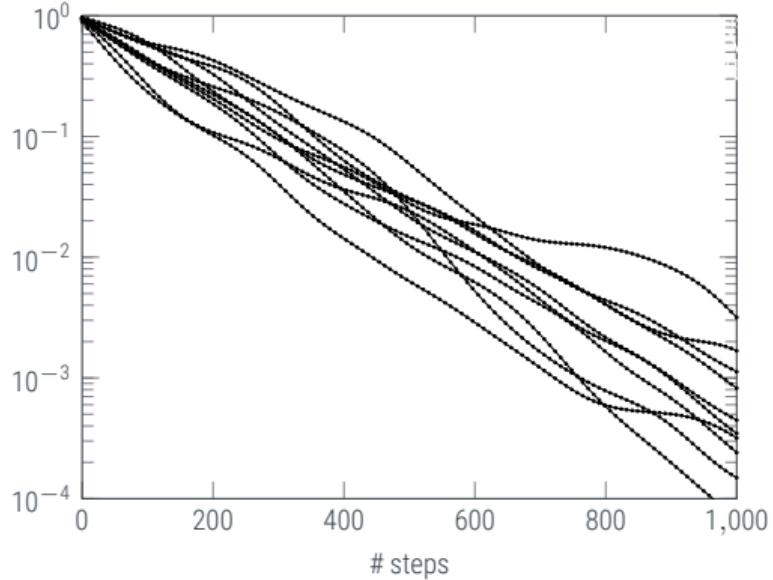
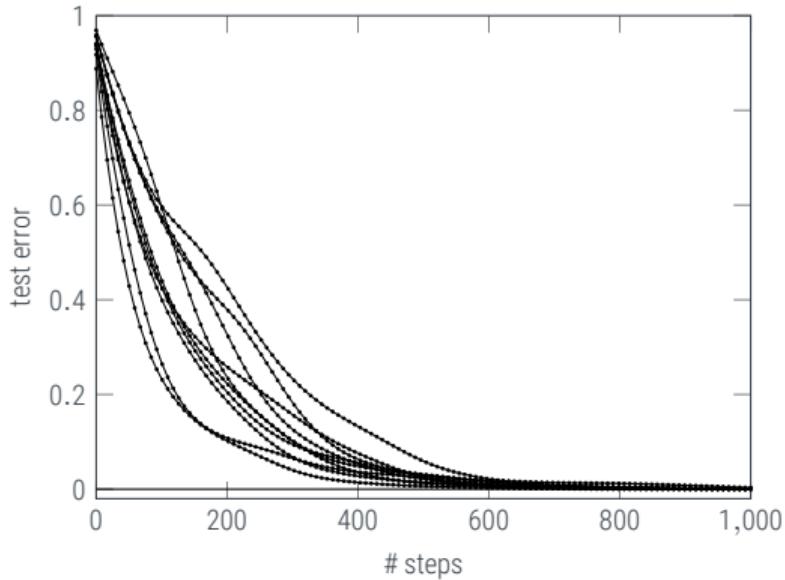
1972–



# Careful with Averages

where do you want to average?

$$\log \sum_i y_i \neq \sum_i \log y_i$$

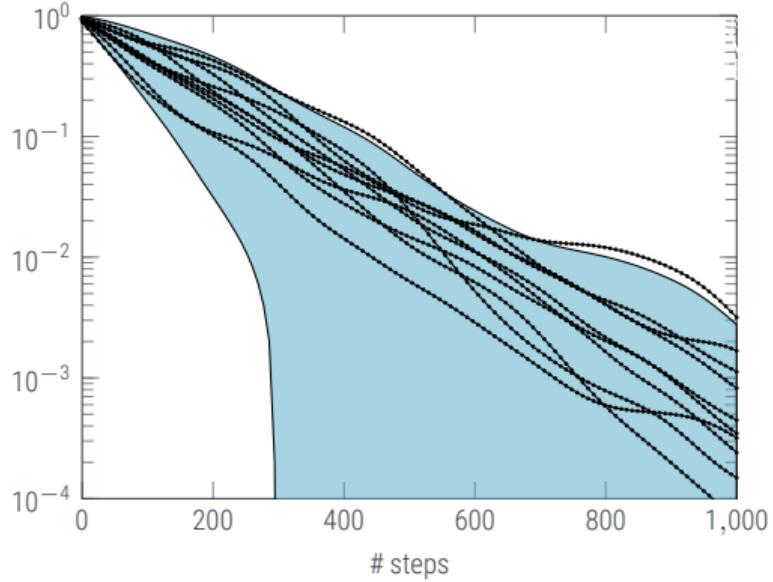
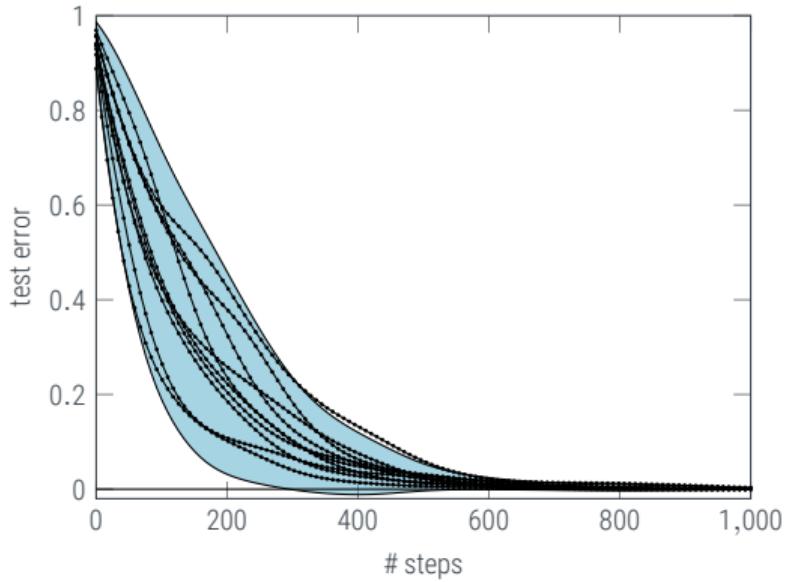




# Careful with Averages

where do you want to average?

$$\log \sum_i y_i \neq \sum_i \log y_i$$

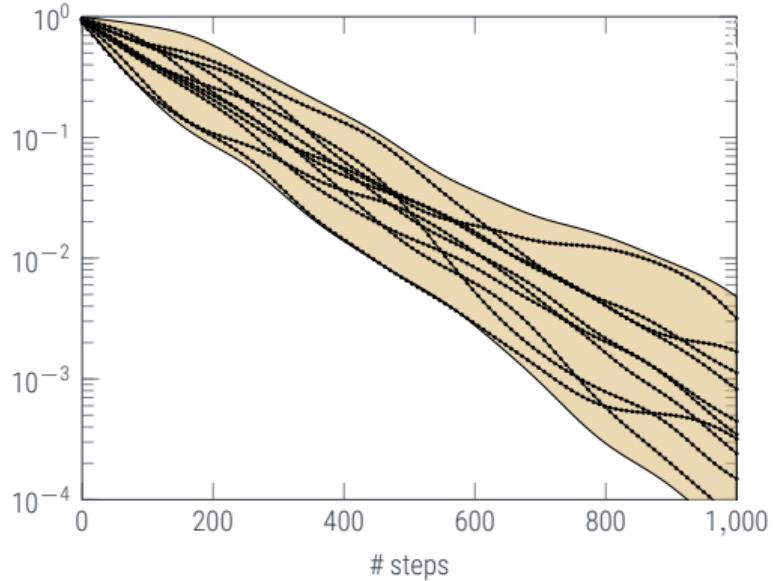
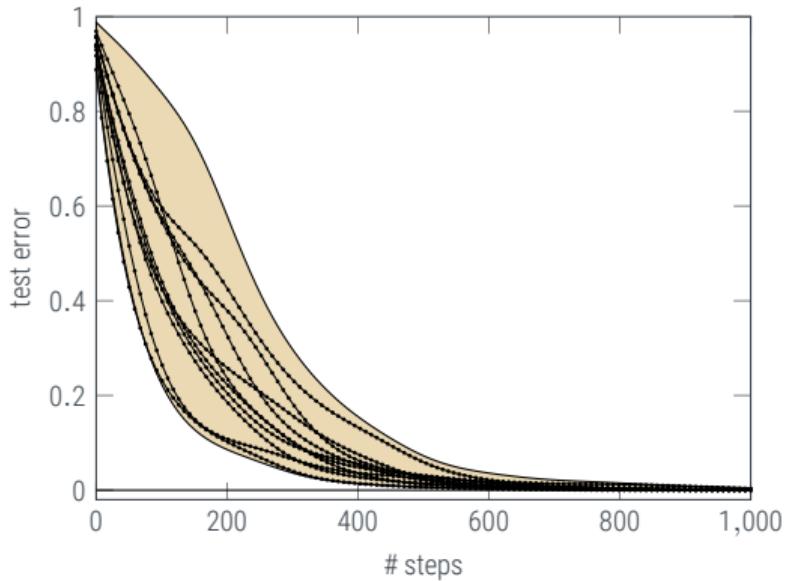




# Careful with Averages

where do you want to average?

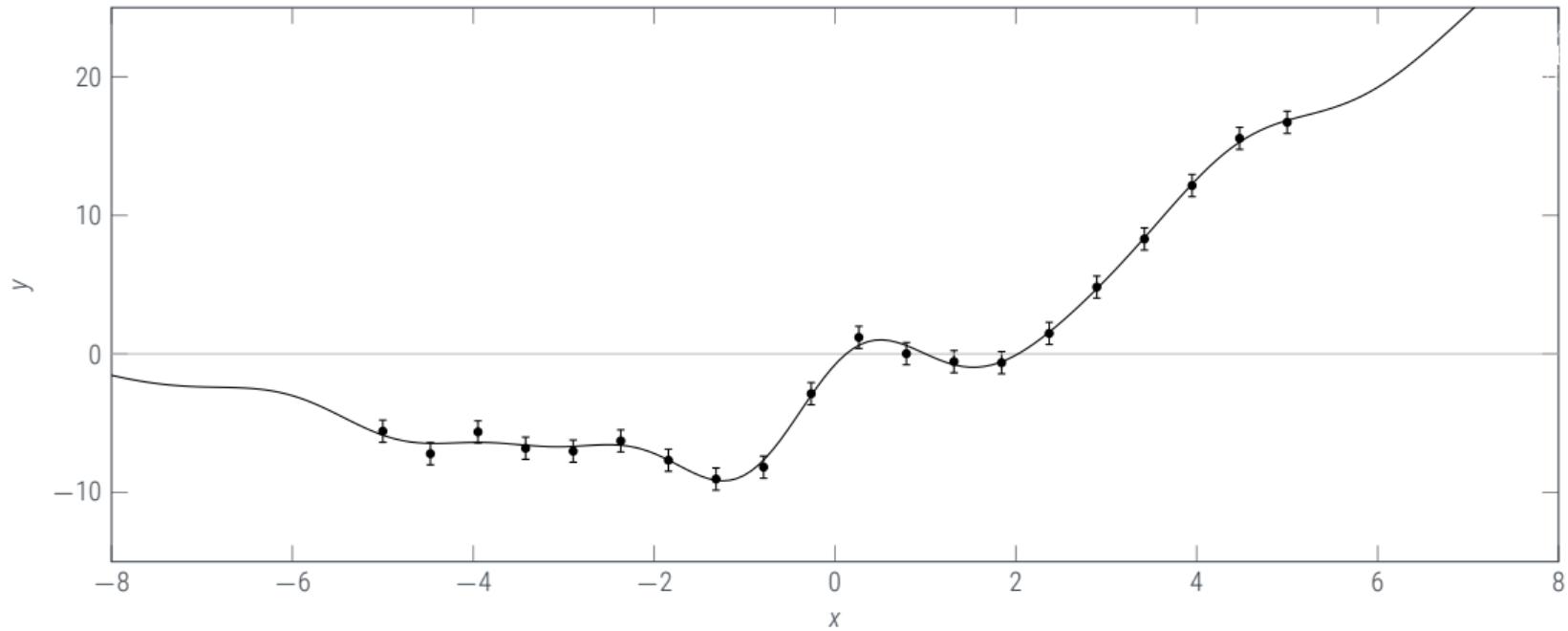
$$\log \sum_i y_i \neq \sum_i \log y_i$$





# Sometimes, a black line is fine

if it's just one





# But colors do help

when things get crowded



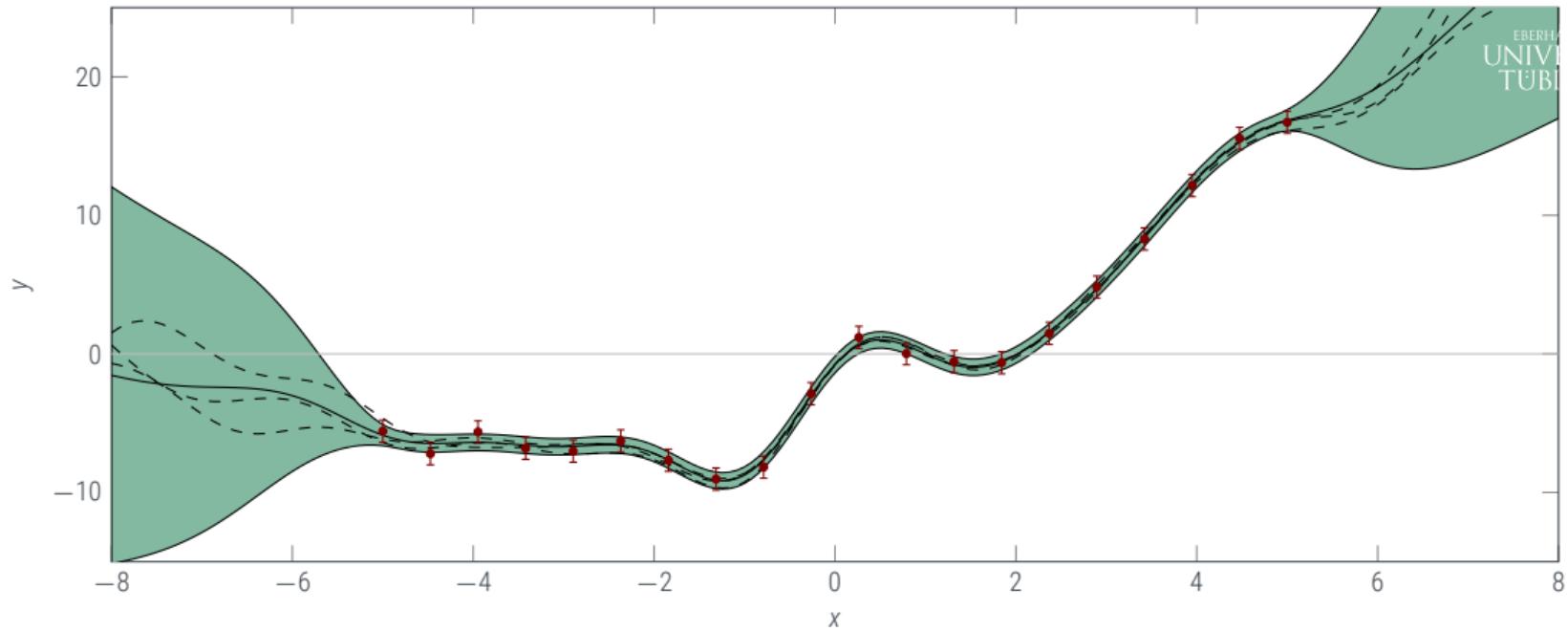
# Shading Matters, too

different normalizations move the focus



# Flat fills are ok as well

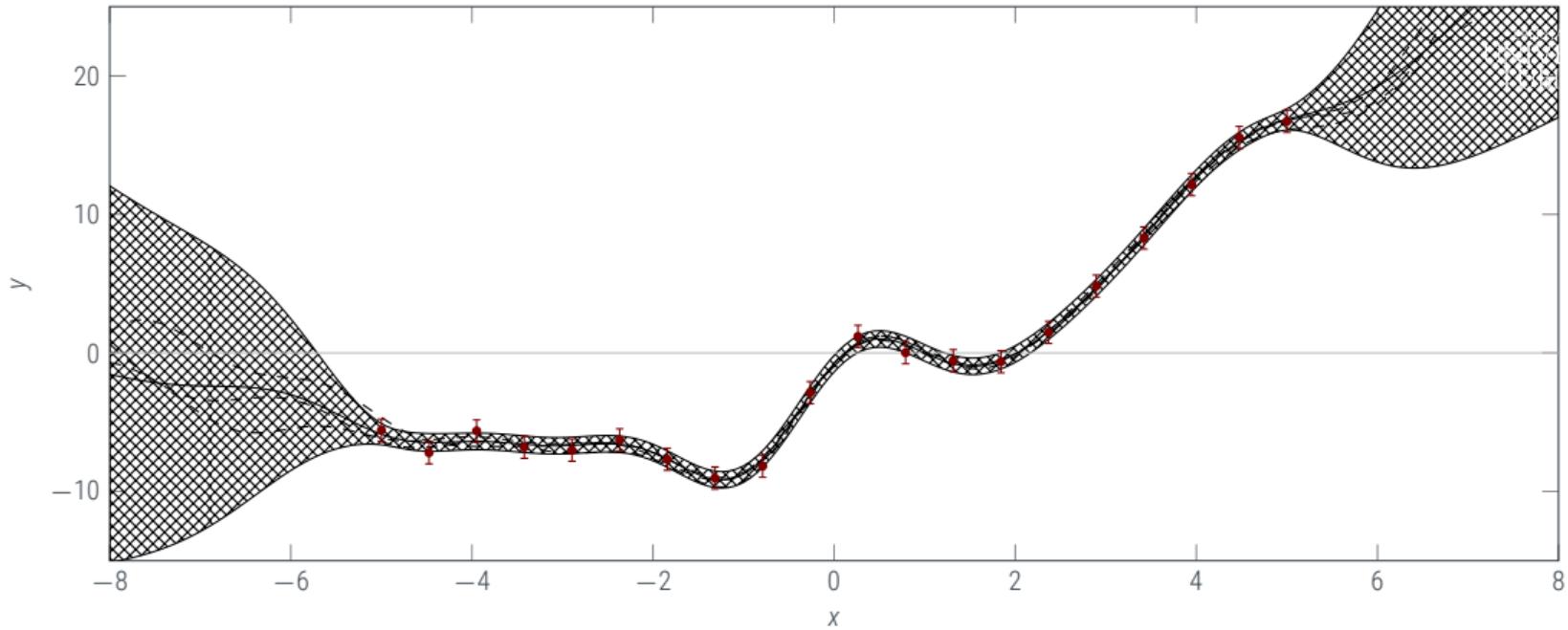
and can be overlayed better





# Patterns rarely make sense

potential exception: stacked plots, in b/w print



# Colors

think about them!

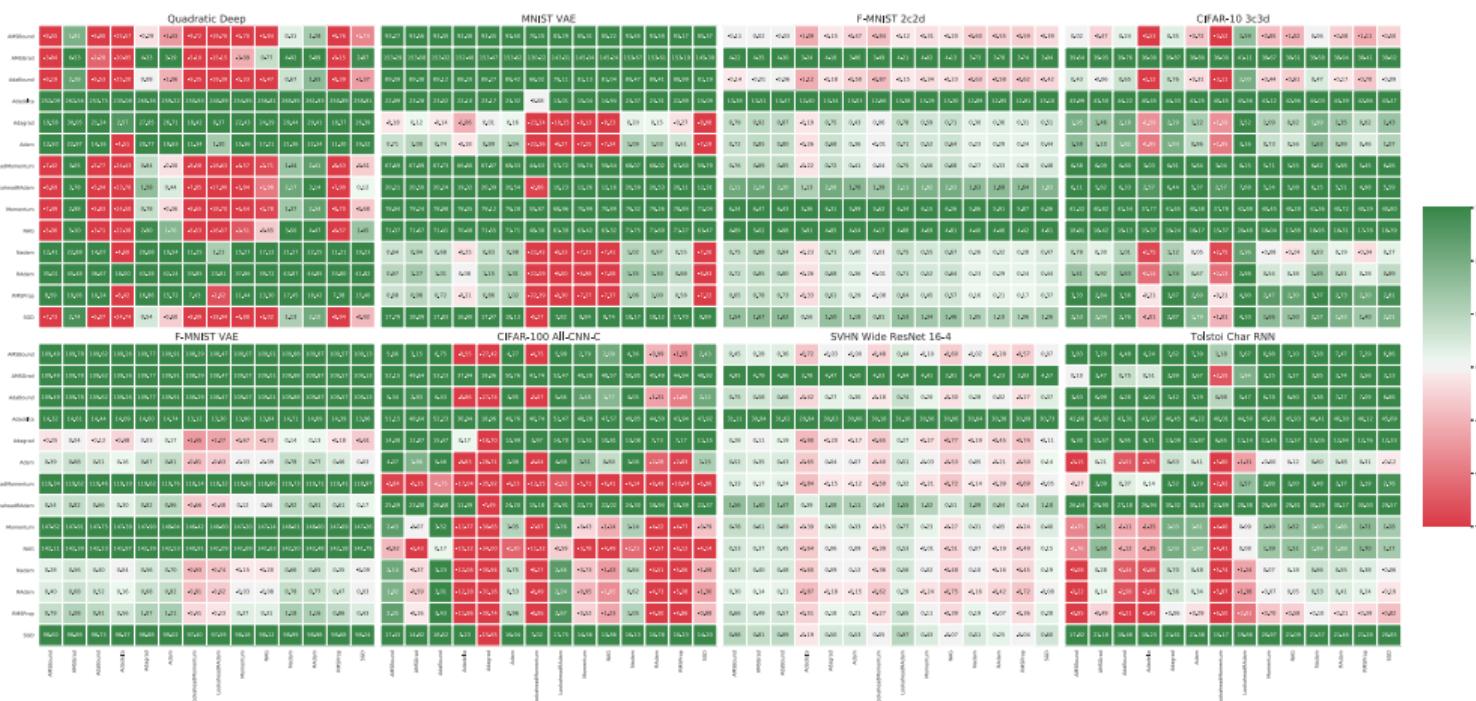


<https://matplotlib.org/stable/tutorials/colors/colormaps.html>

## Accessibility

graphics: arxiv 2007.01547

### Mean Improvement when tuning



# Accessibility

8% of the population (mostly male) is affected by color blindness



F-MNIST VAE

	109.5	109.8	109.8	109.6	109.5	109.8	109.9	107.7	108.5	109.1	109.1	109.9	109.9	109.7	109.2
AMSGrad	109.5	109.8	109.8	109.6	109.5	109.8	109.9	107.7	108.5	109.1	109.1	109.9	109.9	109.7	109.2
AdaBelief	0.3	0.6	0.7	0.4	0.3	0.7	0.7	-1.4	-0.7	-0.1	-0.0	0.7	0.7	0.6	0.1
AdaBound	109.5	109.8	109.8	109.6	109.5	109.8	109.9	107.7	108.5	109.1	109.1	109.9	109.9	109.7	109.2
Adadelta	14.3	14.6	14.6	14.4	14.3	14.6	14.7	12.5	13.3	13.9	14.0	14.7	14.7	14.6	14.1
Adagrad	-0.2	0.0	0.1	-0.1	-0.2	0.1	0.1	-2.0	-1.3	-0.7	-0.6	0.1	0.1	-0.0	-0.5
Adam	0.4	0.7	0.7	0.5	0.4	0.7	0.8	-1.4	-0.6	-0.0	0.0	0.8	0.8	0.6	0.1
LA(Mom.)	119.3	119.6	119.7	119.4	119.3	119.6	119.7	117.6	118.3	118.9	119.0	119.7	119.7	119.6	119.1
LA(RAdam)	0.5	0.8	0.9	0.6	0.5	0.8	0.9	-1.2	-0.5	0.1	0.2	0.9	0.9	0.8	0.3
Mom.	147.6	147.9	147.9	147.7	147.6	147.9	148.0	145.8	146.6	147.2	147.3	148.0	148.0	147.9	147.4
NAG	142.1	142.4	142.4	142.2	142.1	142.4	142.5	140.3	141.1	141.7	141.7	142.5	142.5	142.3	141.8
Nadam	0.3	0.6	0.6	0.4	0.3	0.6	0.7	-1.5	-0.8	-0.2	-0.1	0.6	0.7	0.5	0.0
RAdam	0.4	0.7	0.7	0.5	0.4	0.7	0.8	-1.4	-0.6	-0.0	0.0	0.8	0.8	0.6	0.1
RMSProp	0.8	1.1	1.1	0.9	0.8	1.1	1.2	-1.0	-0.2	0.4	0.4	1.2	1.2	1.0	0.5
SGD	98.6	98.9	98.9	98.7	98.6	98.9	99.0	96.8	97.6	98.2	98.2	99.0	99.0	98.8	98.3

One-shot

SVHN Wide ResNet 16-4

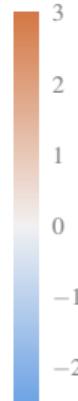
AMSGrad	0.3	0.2	-0.1	0.3	0.1	-0.1	-0.2	0.7	0.5	0.3	0.4	-0.1	-0.3	-1.3	-0.2
AdaBelief	4.8	4.7	4.4	4.8	4.6	4.4	4.3	5.2	5.0	4.8	4.9	4.4	4.2	3.3	4.3
AdaBound	0.3	0.2	-0.1	0.3	0.2	-0.0	-0.2	0.8	0.6	0.3	0.5	-0.1	-0.3	-1.2	-0.2
Data Literacy – P. Hennig, WS 2021/22 – Lecture 12: How to Make a Figure – © Philipp Hennig, 2022 CC BY-NC-SA 3.0	0.6	0.5	0.2	0.6	0.4	0.2	0.1	1.0	0.8	0.6	0.7	0.2	0.0	-1.0	0.1

CIFAR-100 All-CNN-C

5.0	2.8	3.3	0.2	0.3	-12.0	3.1	7.8	5.2	6.1	7.9	3.4	3.6	1.3	1.7
51.5	49.3	49.8	46.7	46.8	34.5	49.6	54.3	51.7	52.6	54.4	49.9	50.1	47.8	48.2
1.6	-0.5	-0.1	-3.2	-3.1	-15.4	-0.3	4.4	1.8	2.7	4.5	0.0	0.2	-2.1	-1.7
4.7	2.5	3.0	-0.1	-0.0	-12.3	2.8	7.5	4.9	5.8	7.6	3.1	3.3	1.0	1.3
50.5	48.3	48.8	45.7	45.8	33.5	48.6	53.3	50.7	51.6	53.4	48.9	49.1	46.8	47.2
13.8	11.6	12.0	9.0	9.0	-3.2	11.8	16.5	13.9	14.8	16.6	12.2	12.3	10.1	10.4
3.8	1.6	2.0	-1.0	-1.0	-13.2	1.8	6.5	3.9	4.8	6.6	2.1	2.3	0.1	0.4
-3.5	-5.7	-5.2	-8.3	-8.2	-20.5	-5.4	-0.7	-3.3	-2.4	-0.6	-5.1	-4.9	-7.2	-6.8
25.0	22.8	23.2	20.2	20.2	8.0	23.0	27.7	25.2	26.0	27.8	23.4	23.5	21.3	21.6
1.8	-0.4	0.1	-3.0	-2.9	-15.2	-0.1	4.6	2.0	2.9	4.7	0.2	0.4	-1.9	-1.6
-1.5	-3.7	-3.3	-6.3	-6.3	-18.5	-3.5	1.2	-1.4	-0.5	1.3	-3.1	-3.0	-5.2	-4.9
1.5	-0.7	-0.2	-3.3	-3.2	-15.5	-0.4	4.3	1.7	2.6	4.4	-0.1	0.1	-2.2	-1.9
1.3	-0.9	-0.4	-3.5	-3.5	-15.7	-0.6	4.1	1.5	2.4	4.1	-0.3	-0.1	-2.4	-2.1
1.7	-0.5	-0.0	-3.1	-3.0	-15.3	-0.2	4.5	1.9	2.8	4.6	0.1	0.3	-2.0	-1.6
16.8	14.6	15.1	12.0	12.1	-0.2	14.9	19.6	17.0	17.9	19.7	15.2	15.4	13.1	13.4

Tolstoi Char RNN

3.8	7.2	6.7	4.5	5.0	7.6	7.4	5.8	5.6	7.5	7.7	7.6	7.5	7.8	6.8
0.1	3.5	3.0	0.7	1.3	3.9	3.7	2.1	1.9	3.8	3.9	3.8	3.8	4.1	3.1
-3.0	0.4	-0.1	-2.4	-1.8	0.8	0.6	-1.0	-1.2	0.7	0.8	0.7	0.7	1.0	0.0
3.6	7.0	6.5	4.3	4.8	7.4	7.2	5.6	5.4	7.3	7.5	7.4	7.3	7.6	6.6



# Colors and Shade

why is this so complicated?



Hue



Saturation

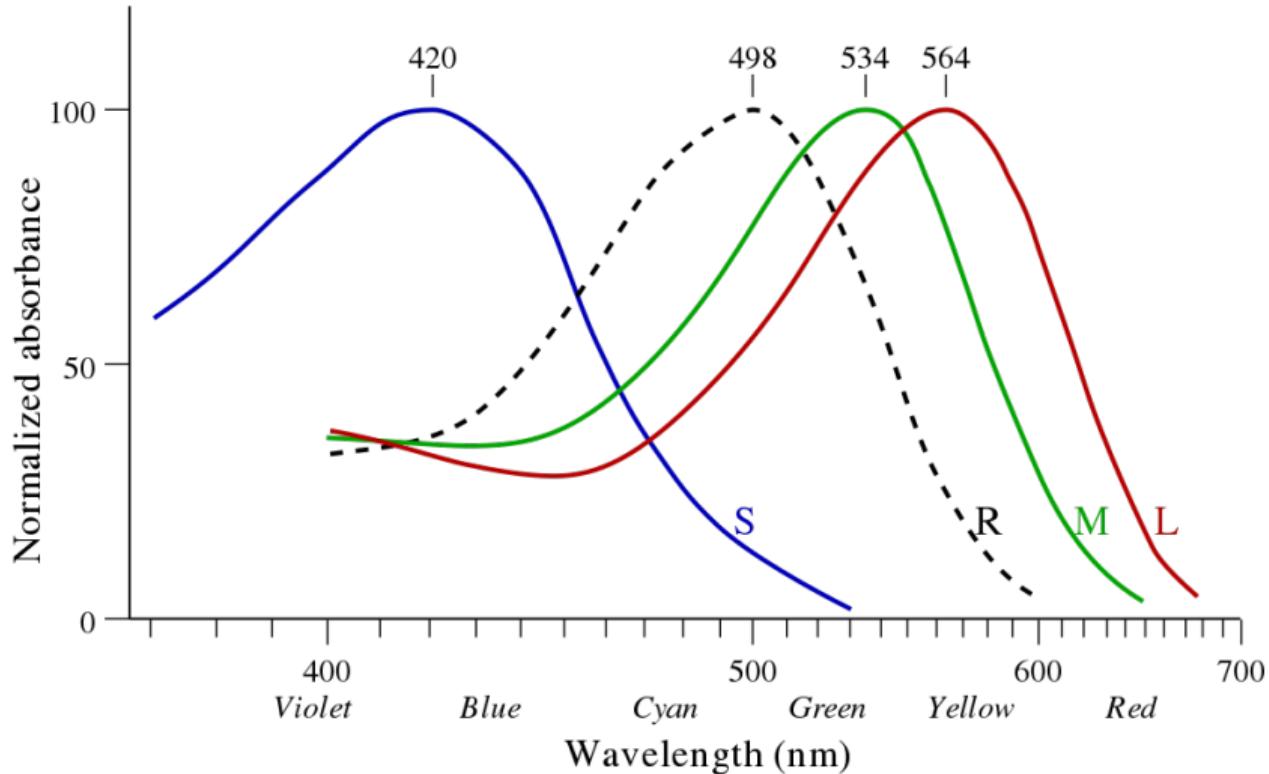


Luminance



# Colors and Shade

why is this so complicated?





# The Physics of Light and Dark

gray scale isn't linear

gamma corrected: linear intensity on screen (looks lighter in print)



gamma uncorrected: linear density of black (looks darker on screen)



linear density of white (looks whiter on screen)



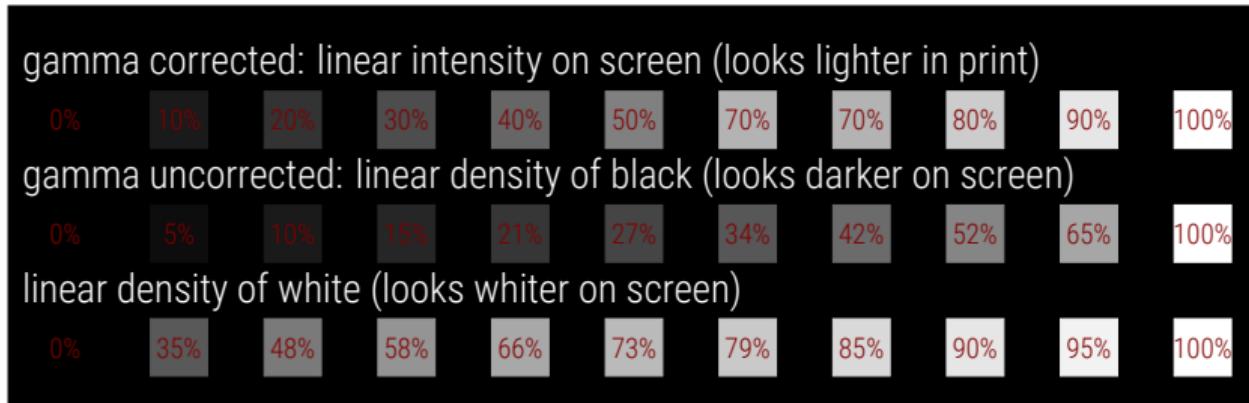
- ▶ there's a difference between covering paper in black ink and increasing the current in an LED
- ▶ when shading an area, consider how it looks in print *and* on screen
- ▶ for more, google *gamma correction*.

Don't treat grayscale as a calibrated scale!



# The Physics of Light and Dark

gray scale isn't linear



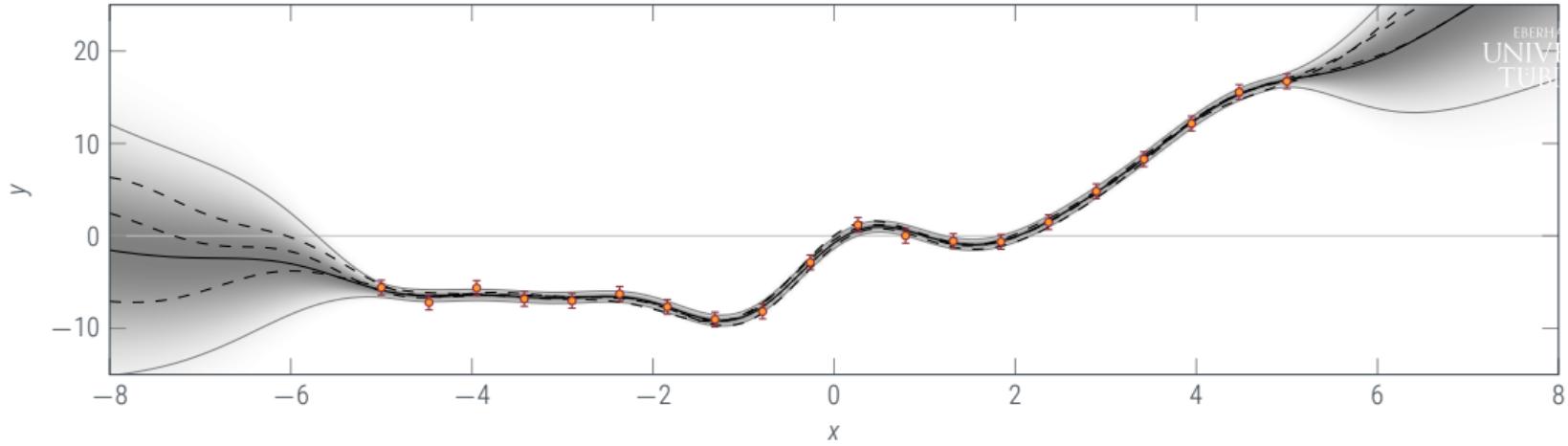
- ▶ there's a difference between covering paper in black ink and increasing the current in an LED
- ▶ when shading an area, consider how it looks in print *and* on screen
- ▶ for more, google *gamma correction*.

Don't treat grayscale as a calibrated scale!



# Captions are important

make them compact and informative

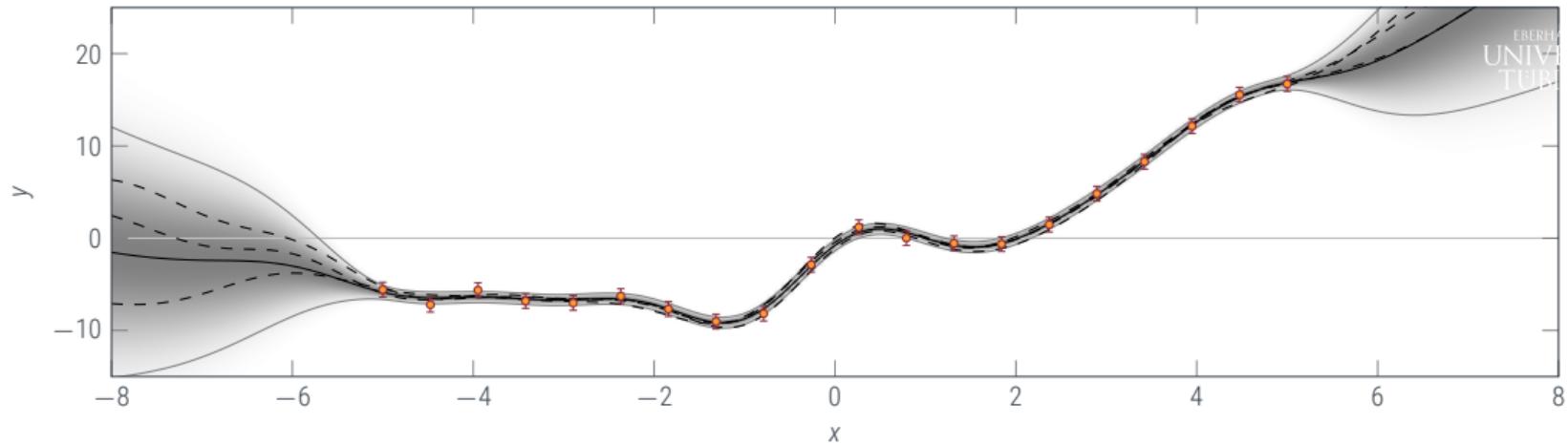


**Figure 1.** This figure shows us the posterior distribution under a Gaussian process prior. It was created using the dataset explained in the text, and using the kernel described in Eq. 1. The data is shown as yellow dots with red circles and error bars. The dashed lines are samples, the black line is the posterior mean function. This figure shows us that the algorithm can find a smooth function interpolating all the data points.



# Captions are important

make them compact and informative



**Figure 1.** Gaussian process posterior distribution. Data (●) shown with errorbars at 2 standard deviations. Mean function (—); three sample paths (---); marginal density as gray shading (— at 2 standard deviations). Note smooth interpolation. Details in text.



# How to *make* your figure

building an efficient toolchain

# Not like this!

Klein, Falkner, Bartels, Hennig, Hutter, *Fast Bayesian Optimization of Machine Learning Hyperparameters on Large Datasets*, AISTATS 2017

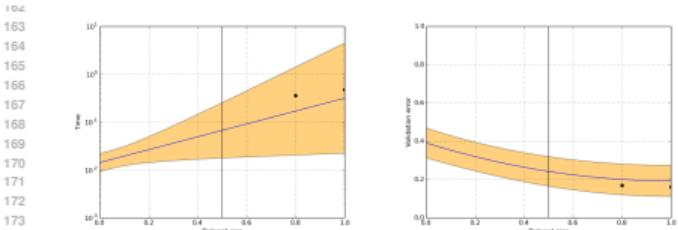


Figure 2: Extrapolation of the models for a random test point for the objective functions and the cost functions after trained on the initial design. Both models have just seen points up to  $\frac{d}{2}$  indicated by the vertical line.

to learn a distribution instead of collecting low function values it does not necessarily evaluate the point where it assumes to be the minimum. Due to noise inside the model this can prevent to hit the global optimia. Note that this is known shortcoming of entropy search when the model is not perfect, and improving this is subject of future work.

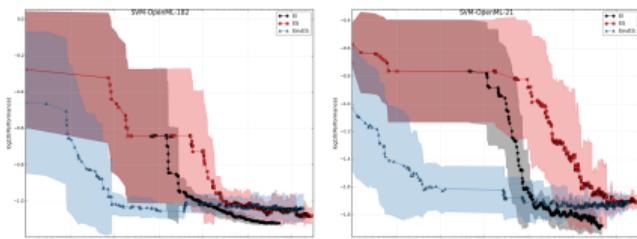


Figure 3: Comparison of our entropy search method (EnvES) against expected improvement and standard entropy search on the support vector machine task on two different datasets. Note the logarithmic x-axis.

# But like this!

tueplots, by Nico Krämer

► pip install tueplots

---

```
1 from tueplots import bundles
2
3 # global change:
4 plt.rcParams.update(bundles.neurips2021(usetex=False))
5
6 # or contextual:
7 with plt.rc_context(bundles.neurips2021(usetex=False,family='serif')):
8     fig, ax = plt.subplots()
9     ax.plot([1.0, 2.0], [3.0, 4.0])
10    ax.set_title("Title")
11    ax.set_xlabel("xlabel")
12    ax.set_ylabel("ylabel")
13    plt.show()
14
15    # this file can now be used as \includegraphics{filename}
16    plt.savefig('figname_neurips.pdf',bbox_inches='tight')
```

---





## Summary: Making Figures

- ▶ Figures are more than decoration. They are a core means of scientific communication, on par with mathematics and experiments
- ▶ Put effort into your figures to convey ideas, convince the reader and clarify details
- ▶ Use an automated toolchain to ensure reproducibility, increase the quality of your figures, and avoid bugs

Please provide feedback:

