Graphs, Tables and Captions

My biased take on how to view your data

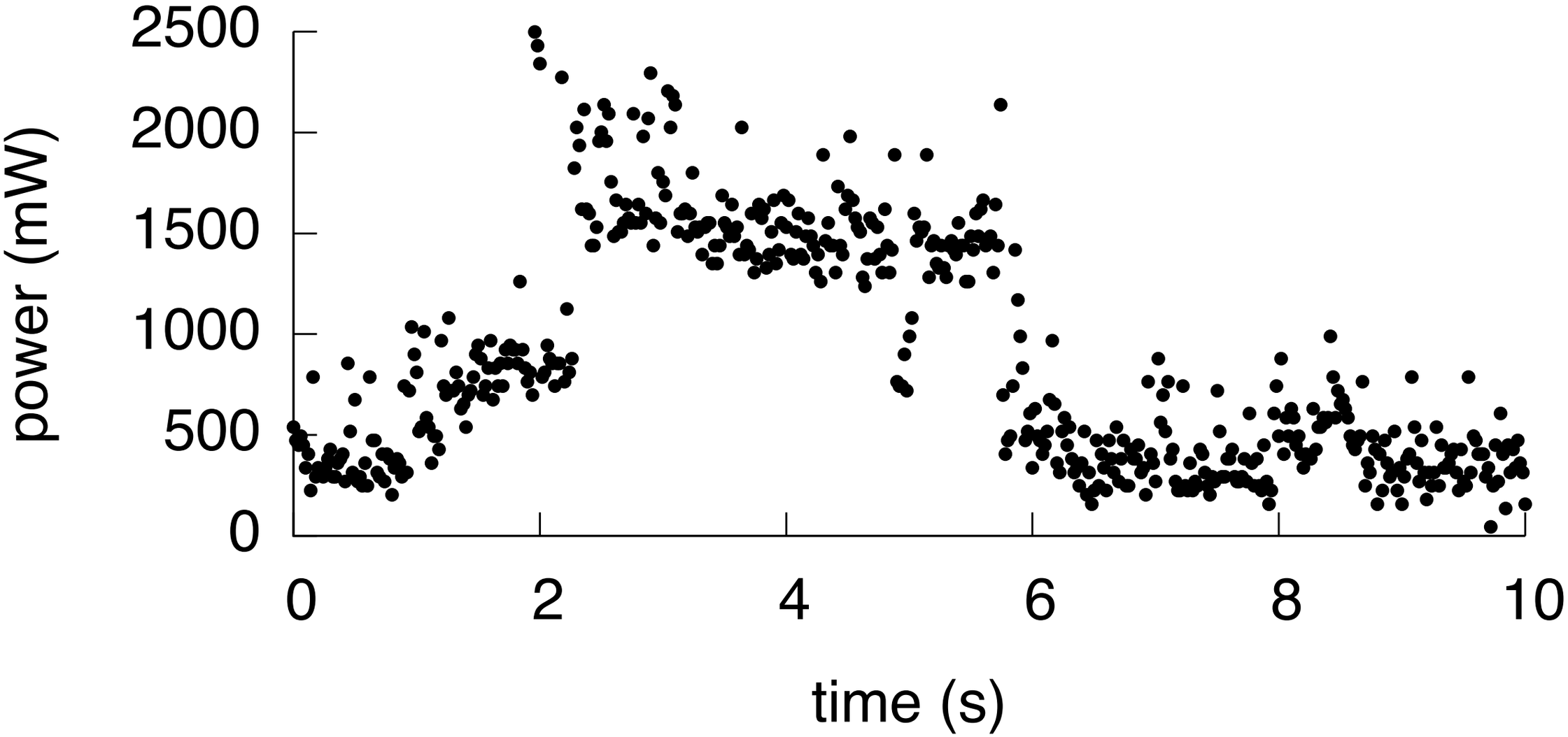
Aaron Schulman
Featuring graphs from many talented artists

Why do we graph?

- To grasp lots of observations
- To find patterns in observations
- To see the distribution of observations
- To determine the relationship between observations (be careful with correlation)
- To find bugs in the experiment
- **To show others our observations**

A graph

X - independent
Y - dependent axis labels (and units) data

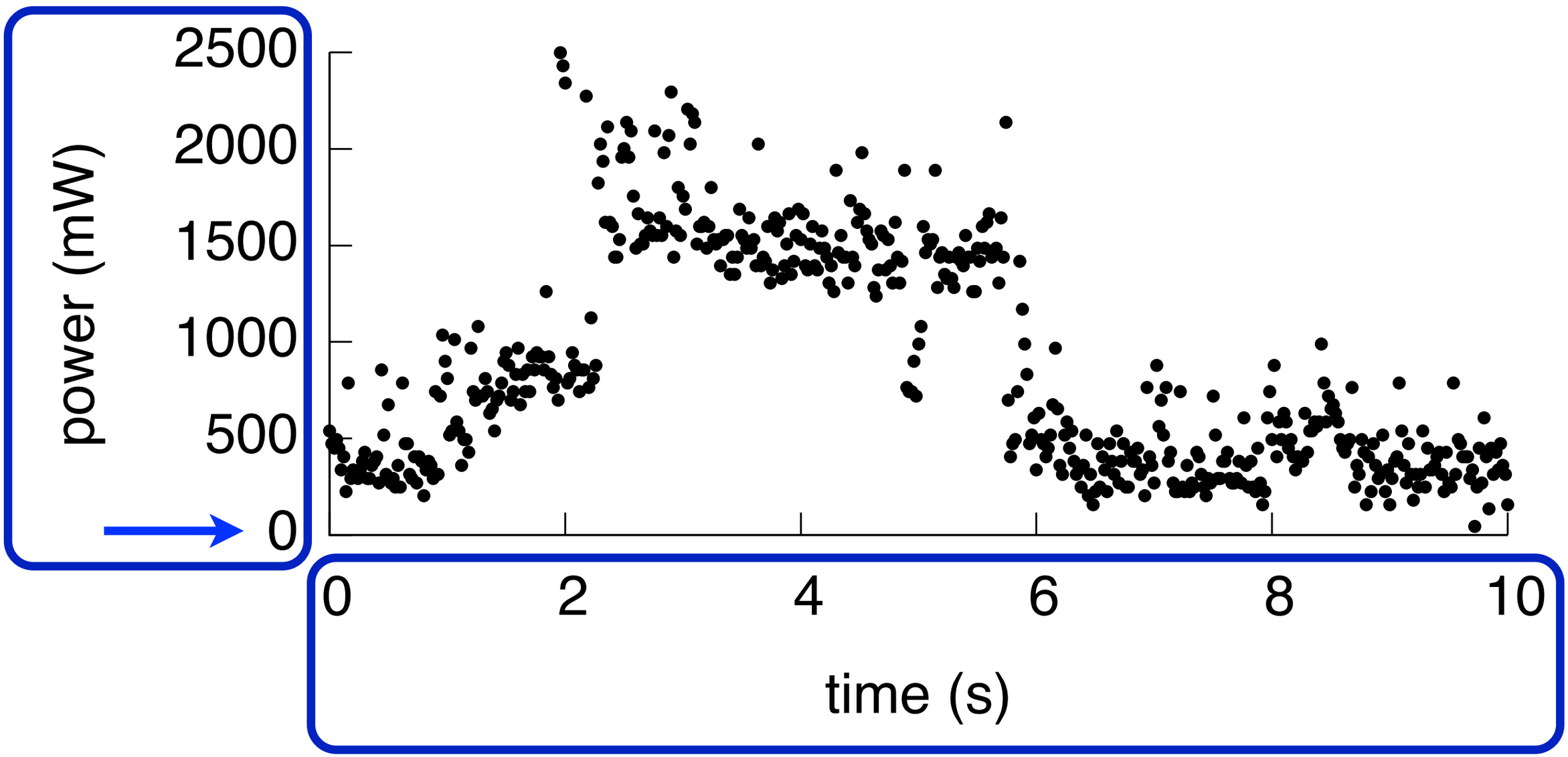


A graph

X - independent
Y - dependent

axis labels (and units)

data

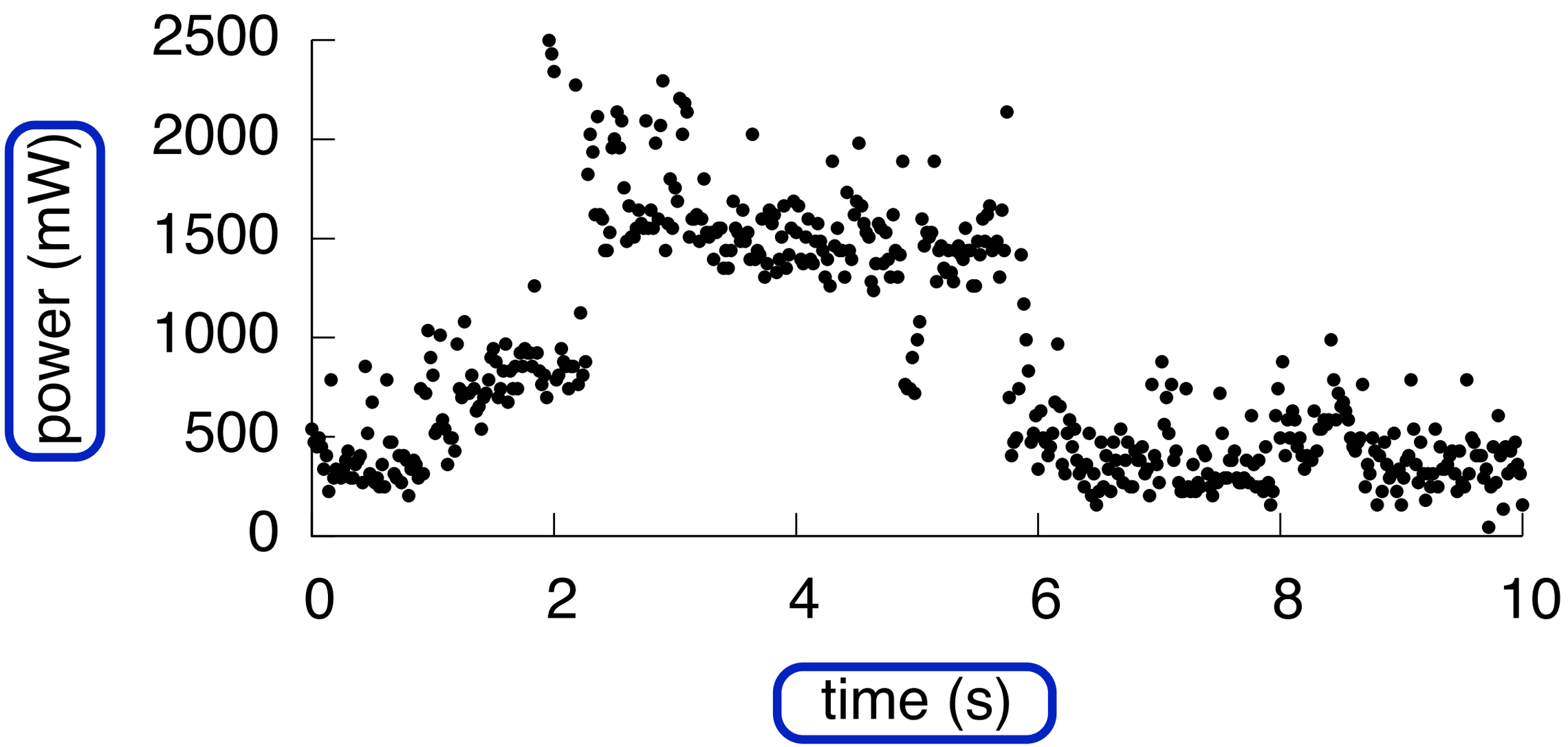


A graph

X - independent
Y - dependent

axis labels (and units)

data

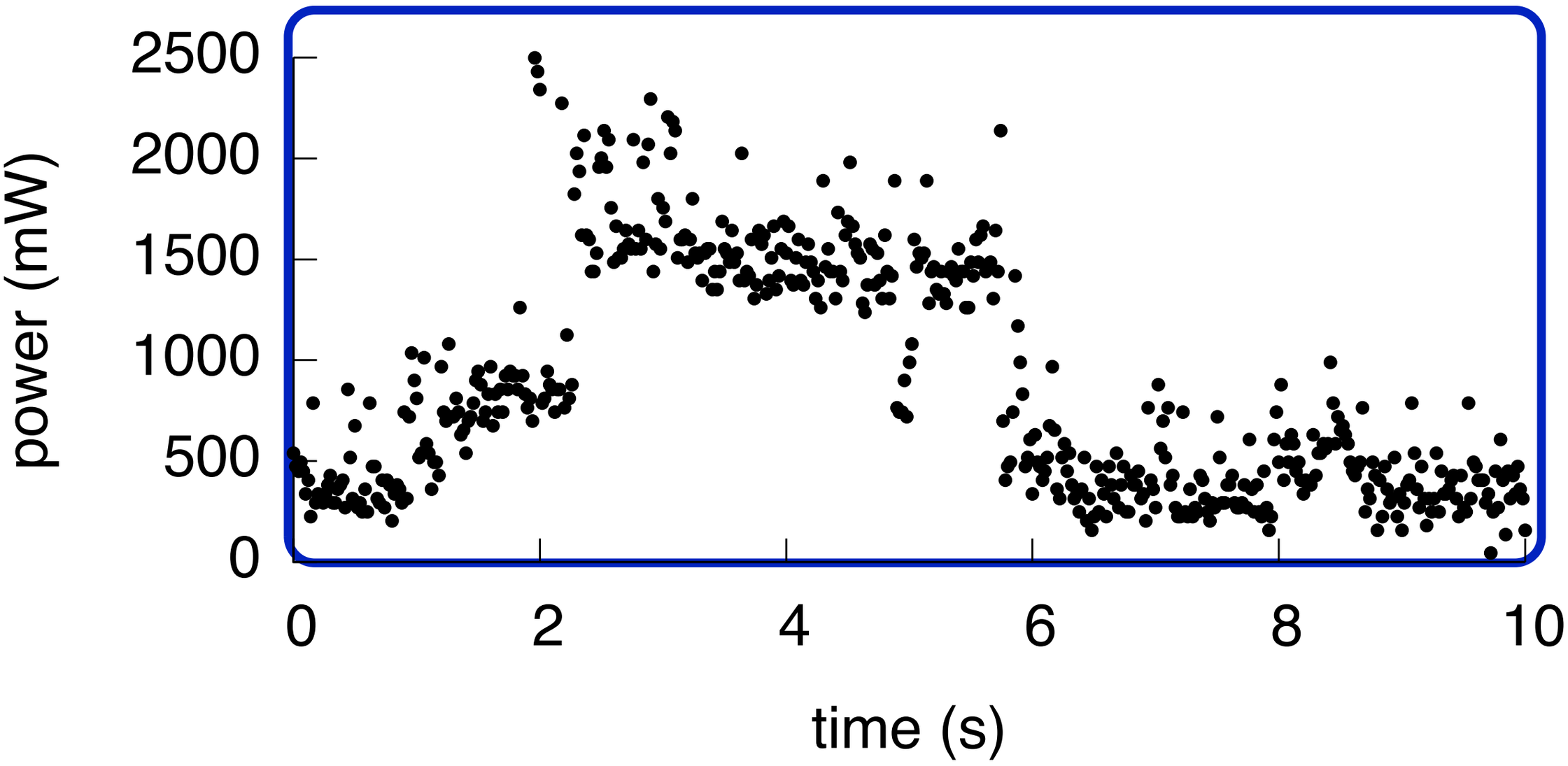


A graph

X - independent
Y - dependent

axis labels (and units)

data



Scatter plot

Always view the raw data first

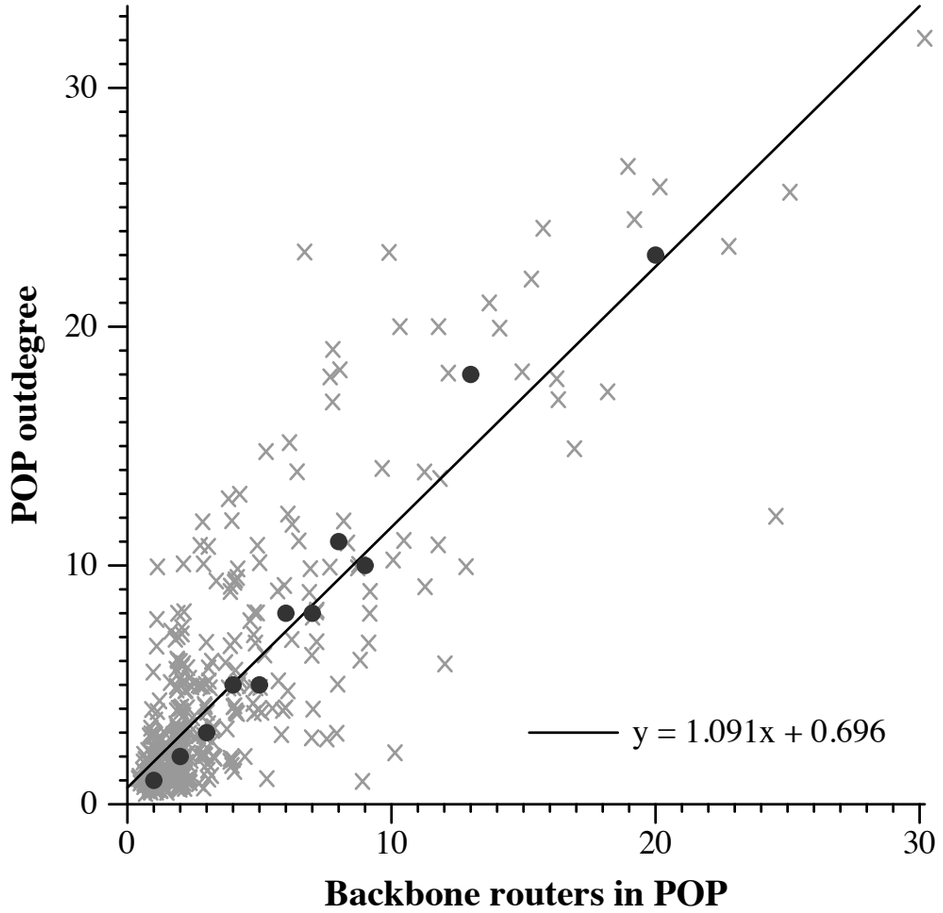


Fig. 11. POP outdegree vs backbone routers in the POP. A small random jitter was added to the data points to expose their density. Circles represent the median of at least ten nearby values: fewer medians are present for the few large POPs. The solid line traces a linear regression fit, with $R^2 = 0.70$. This is an aggregate graph over nine ISPs, excluding Level3 due to its logical mesh topology that gives POPs very high outdegree.

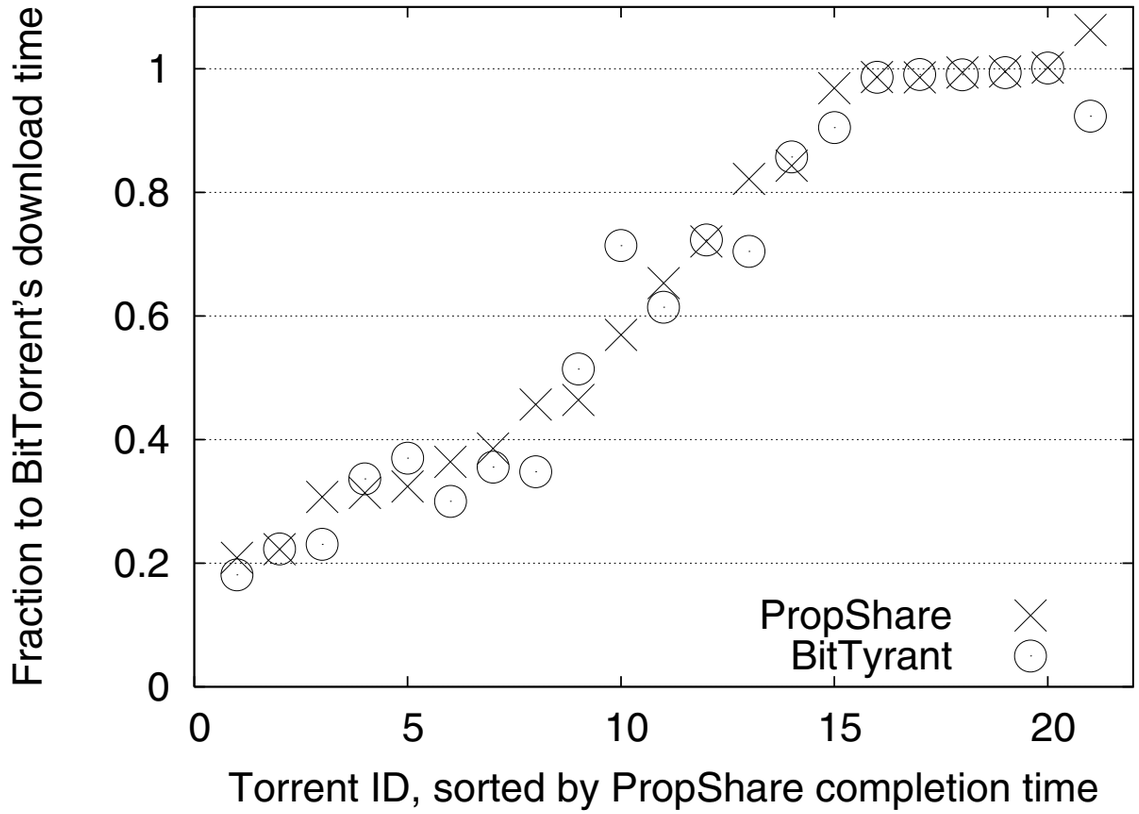


Figure 6: Runs on live swarms

Credit: Neil Spring (left) Dave Levin (right)

Bar graph

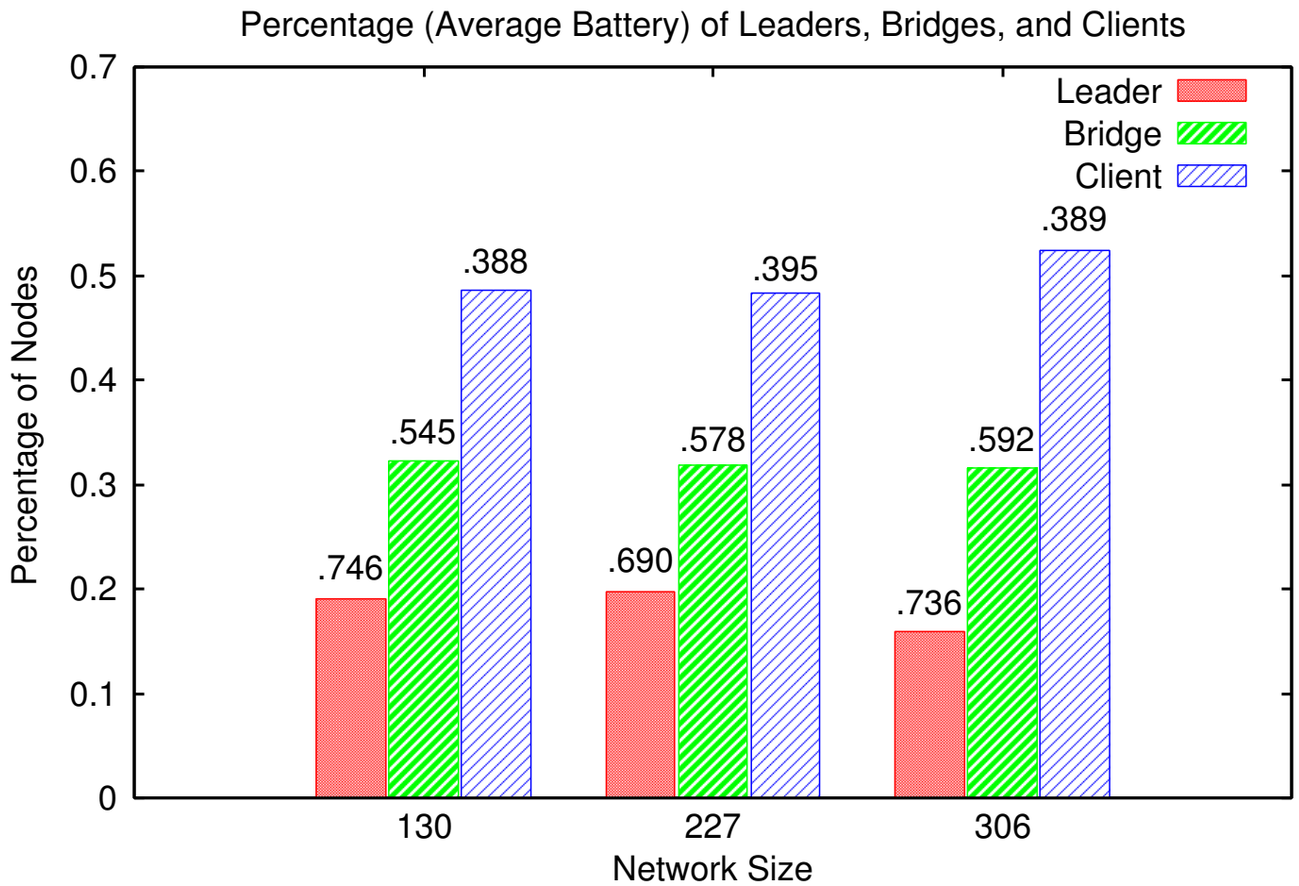
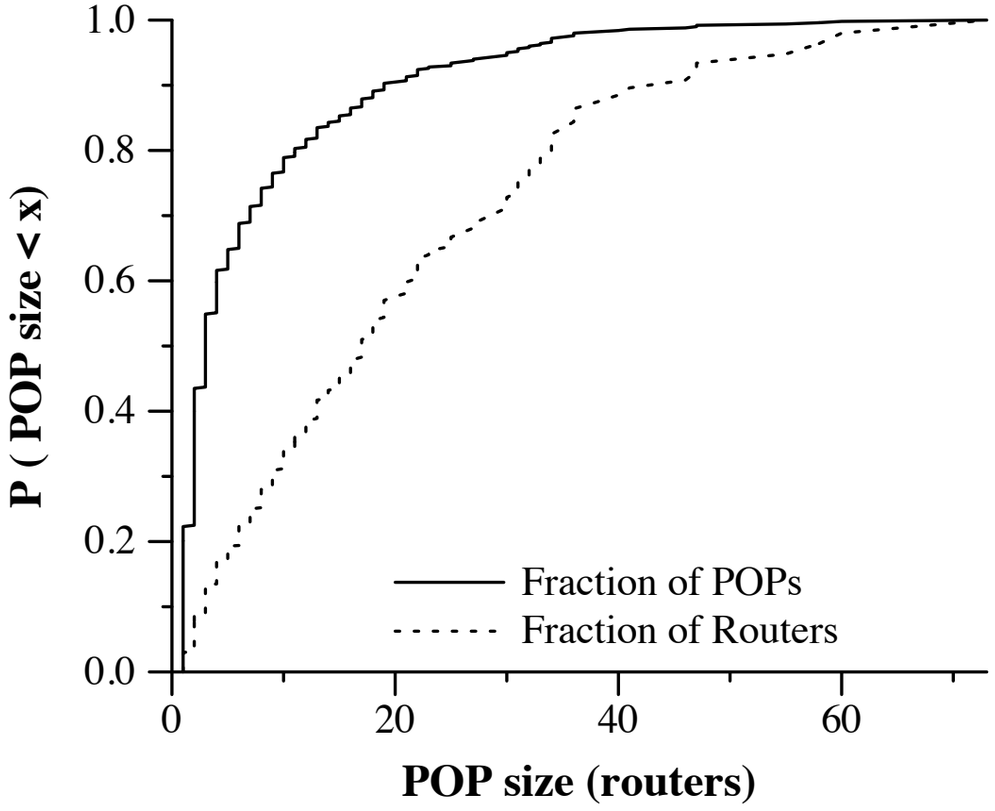


Figure 5: Percentage of leaders, bridges, and clients in different topologies. The value on top of each bar denotes the average remaining battery for each case.

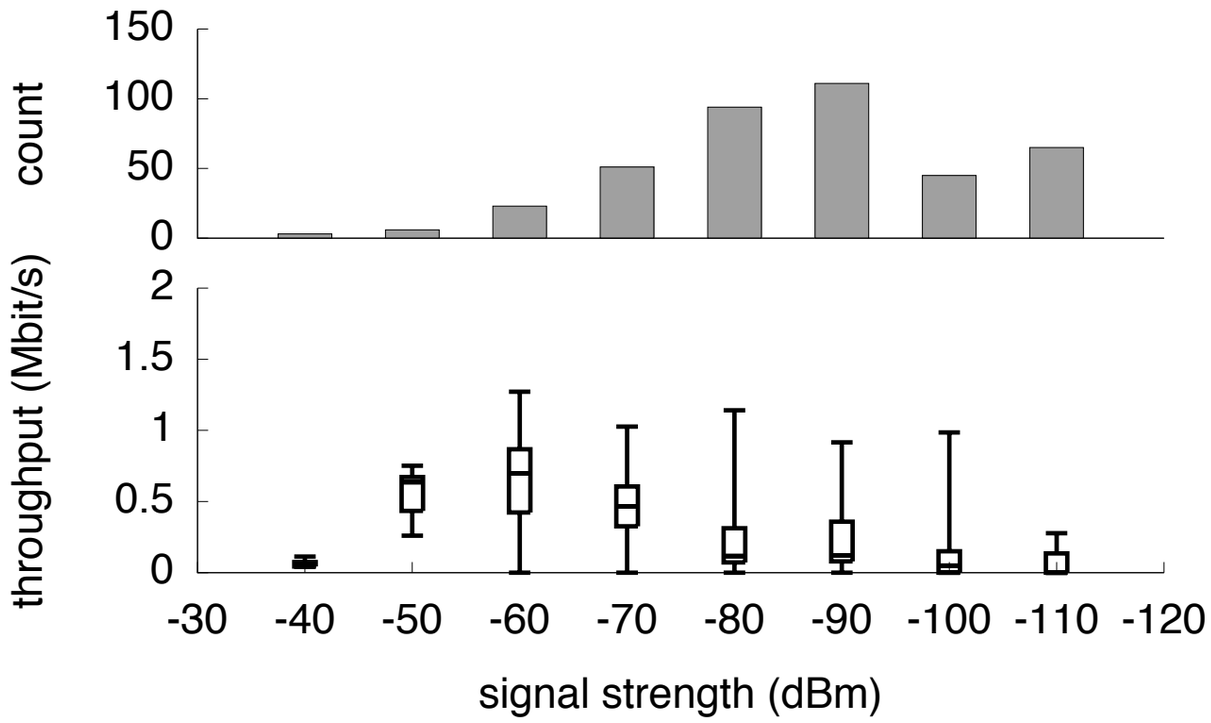
Credit: Seungjoon Lee

Distributions

CDF

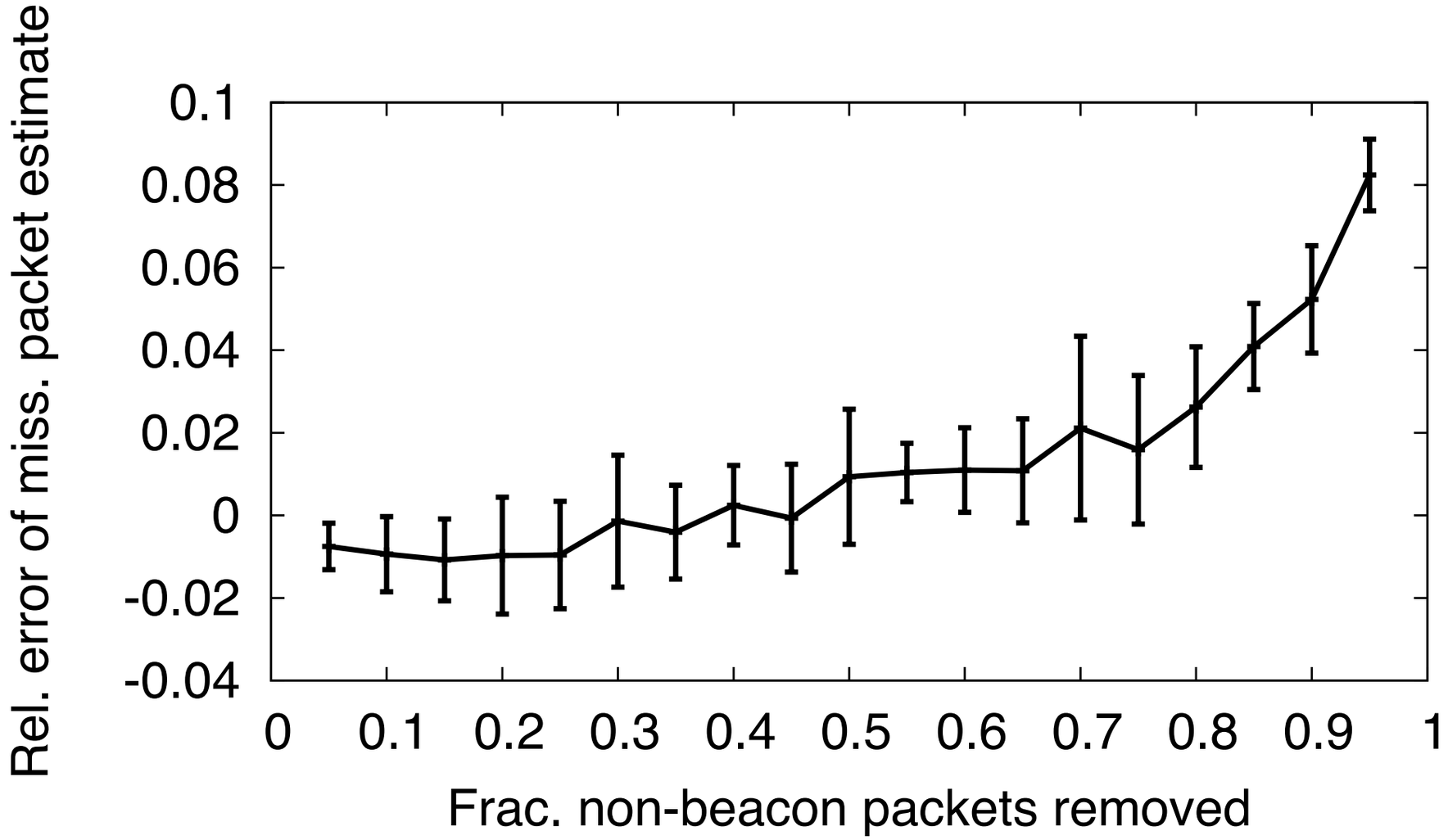


Box and whisker



Credit: Neil Spring

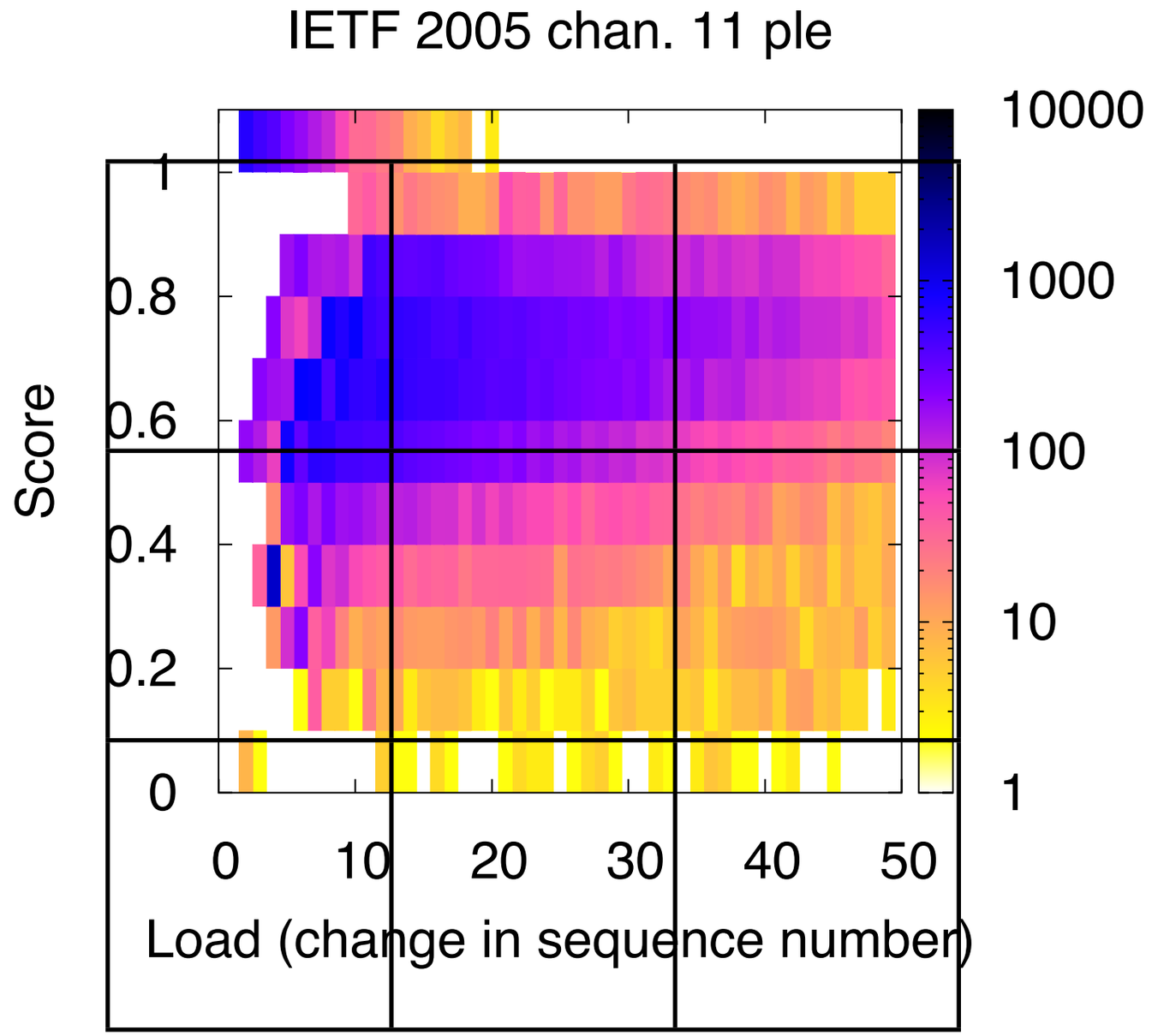
Standard deviation and confidence interval



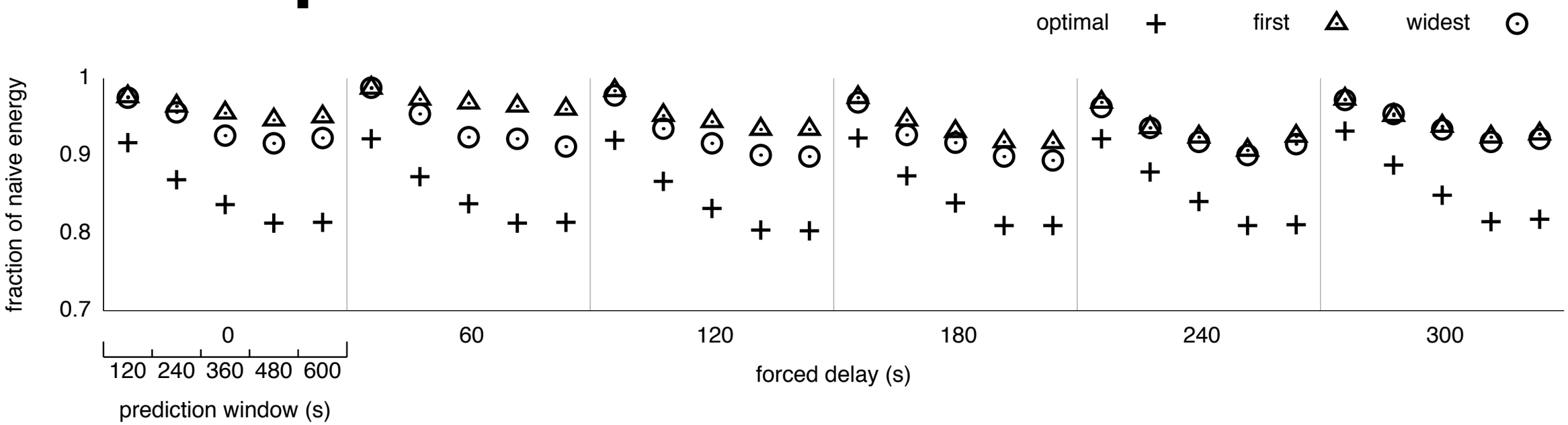
Know your data

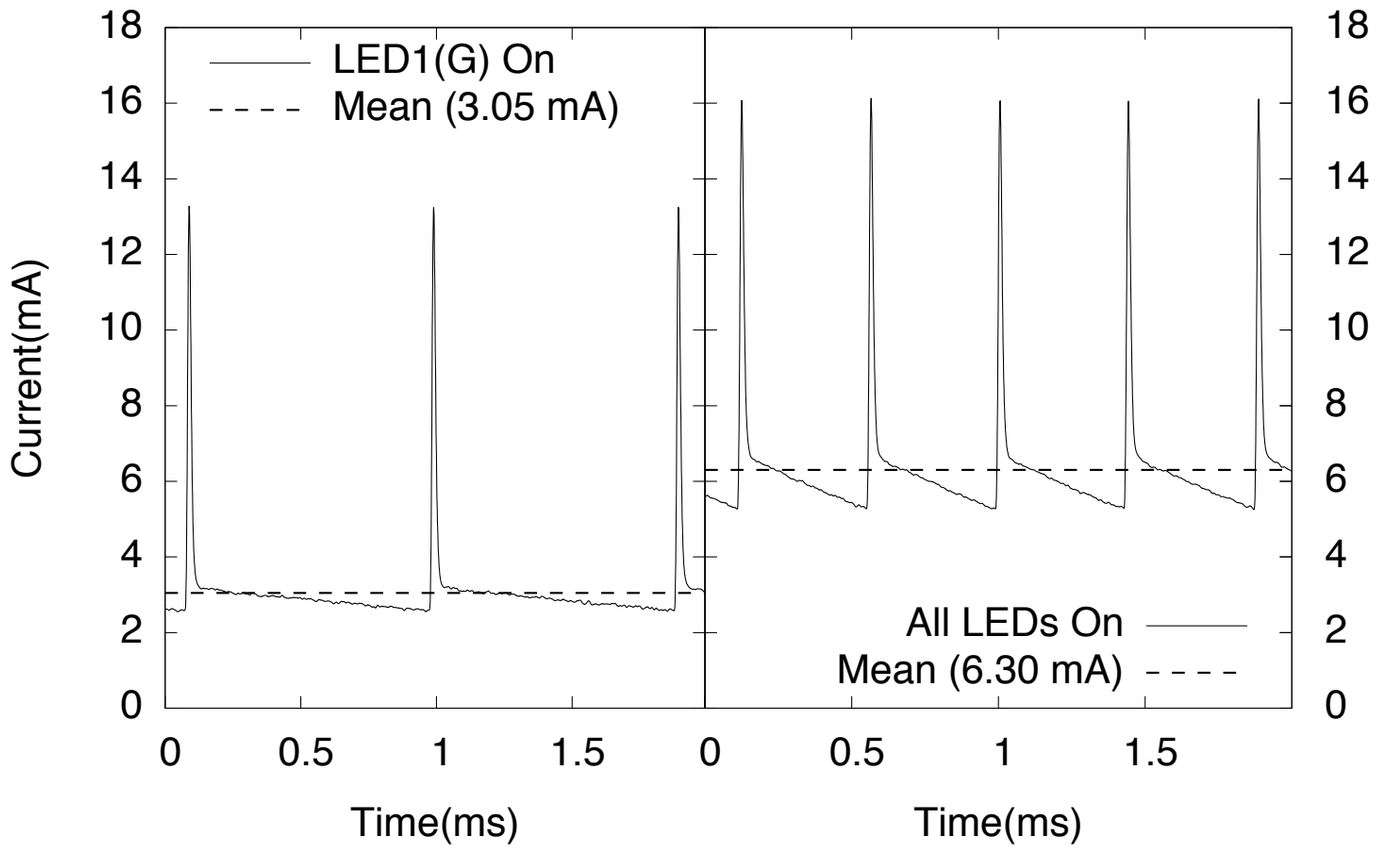
3D and heatmap

- **3D is not your friend**
- Use a heatmap



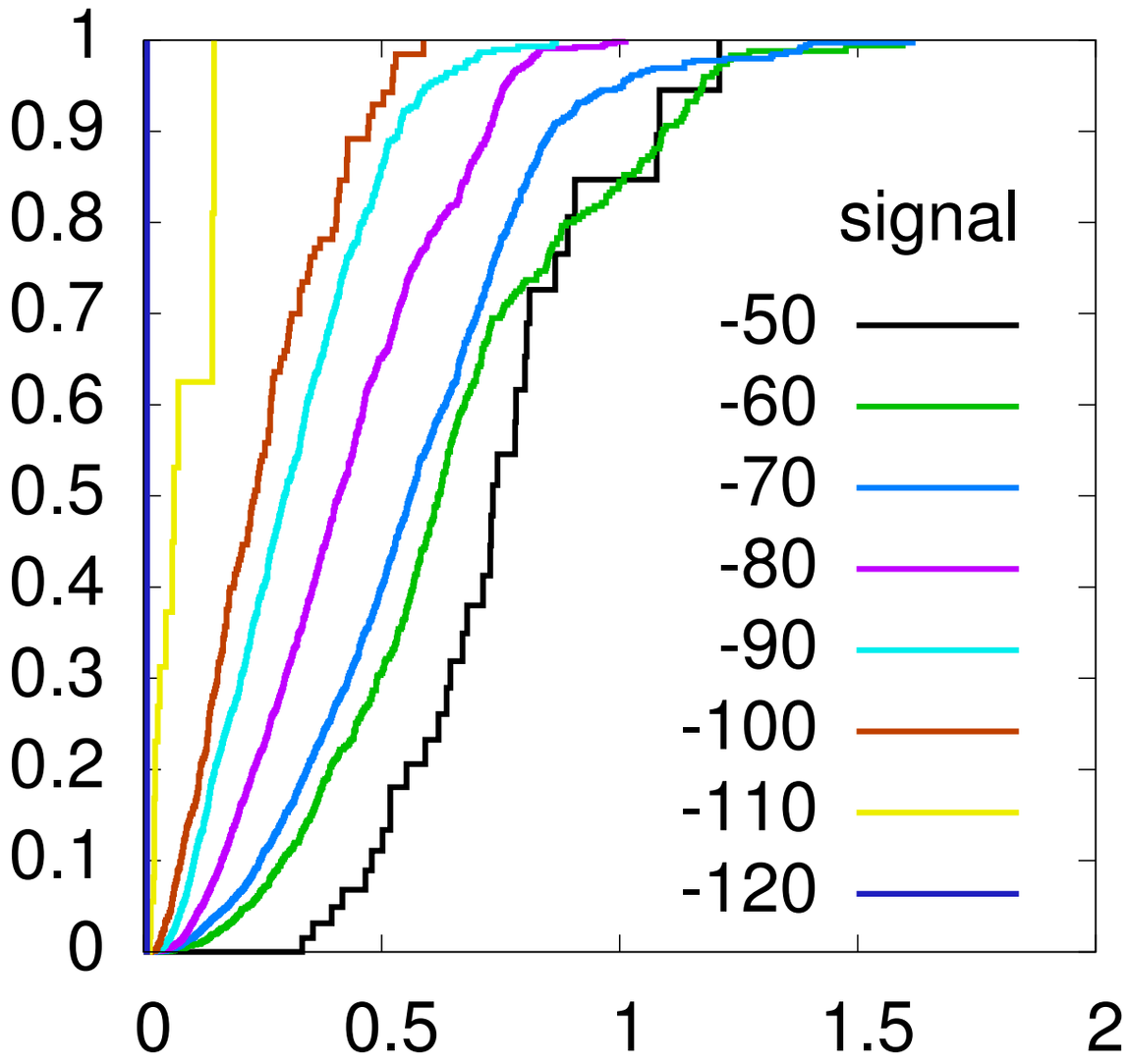
Multiple axis and shared axis

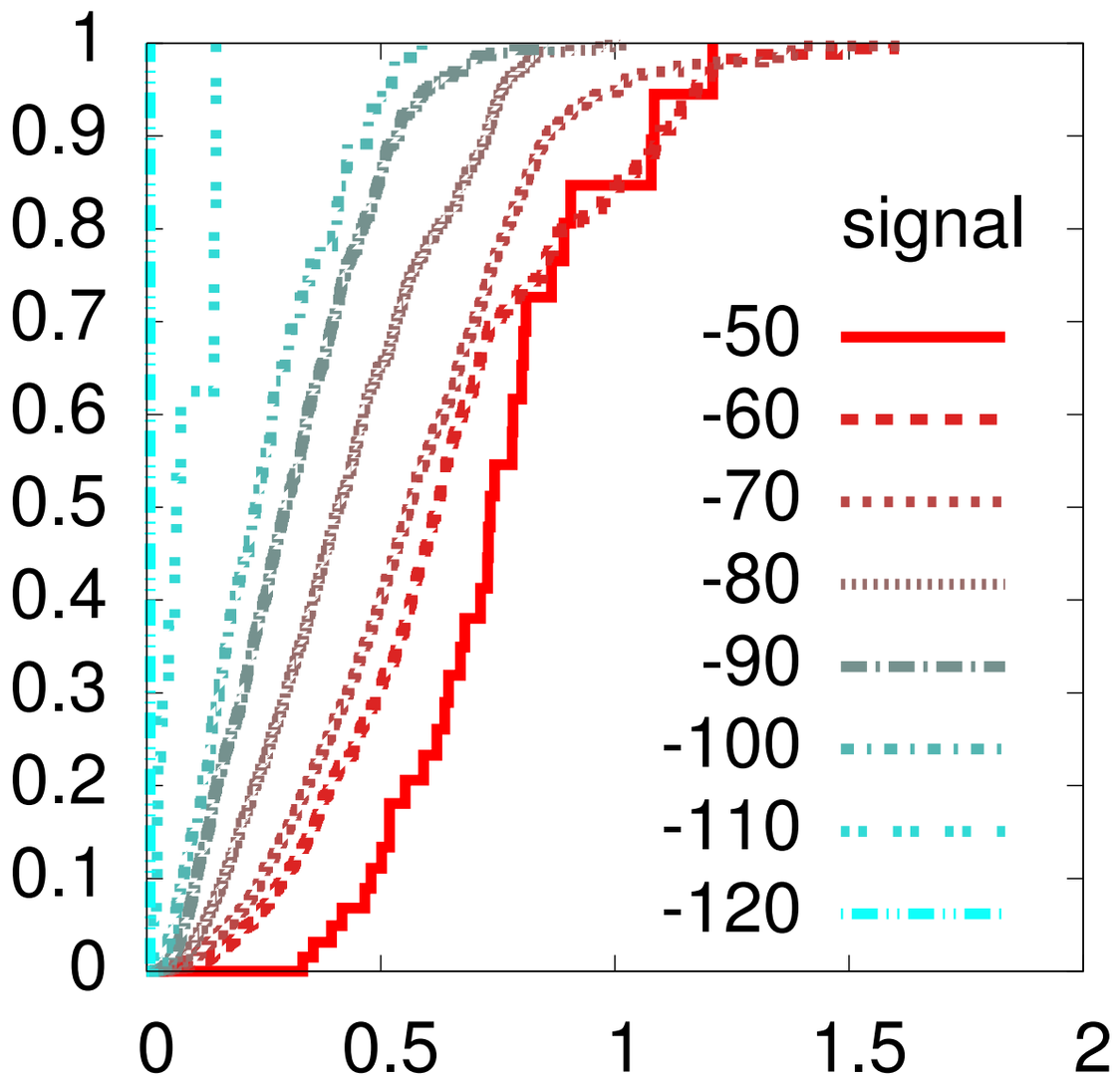




Credit: Prabal Dutta (bottom)

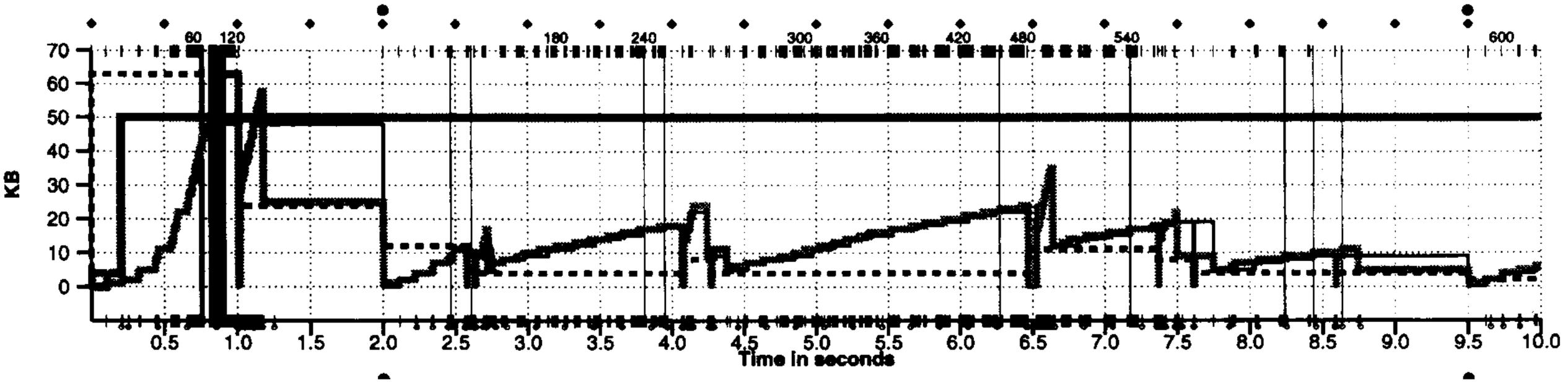
Watch out for grayscale





Captions

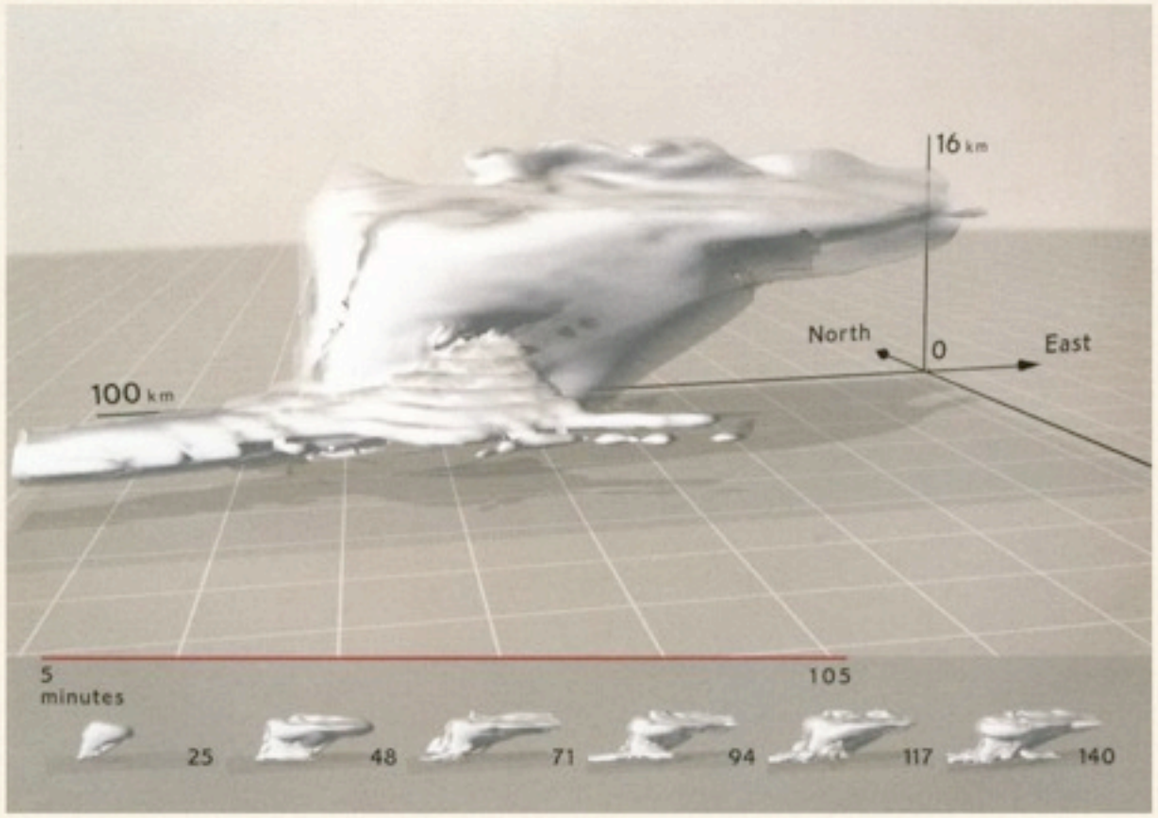
- Hypothesis or conclusion from the figure
- Description the experiment
- Description of the data points
  - Point out interesting ones
  - Give statistics
  - Explain outliers



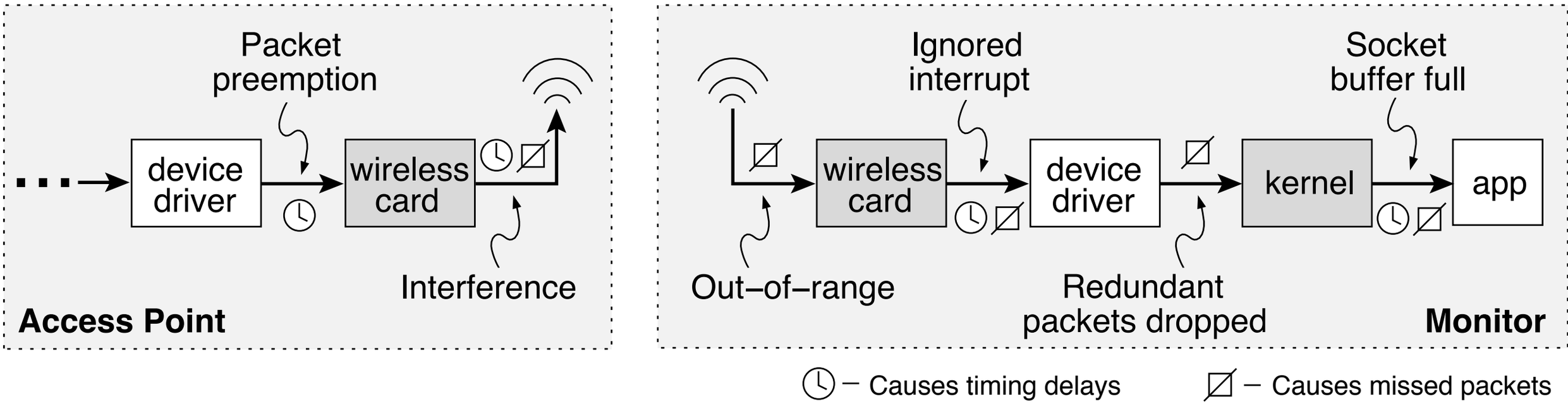
Credit: Lawrence Brakmo

EDWARD R. TUFTE

VISUAL EXPLANATIONS



IMAGES AND QUANTITIES, EVIDENCE AND NARRATIVE


Packet
preemption
device
driver
wireless
card
Interference
Access Point
Ignored
interrupt
Socket
buffer full
wireless
card
device
driver
kernel
app
Out–of–range
Redundant
packets dropped
Monitor
– Causes timing delays
– Causes missed packets

Demo

Graphing tools

- Gnuplot - **does everything,** horrible defaults
- matplotlib - **python**
- Matlab, R - **statistics**, not flexible?
- Jgraph - **nice defaults**, obscure
- Excel - **simple,** horrible defaults, plot and data are one, not easy to script

Why make a table

- If you only have a few data points
- If the interesting data is obvious
- If you want to present a visual comparison

Tables

Technique	Protocol	No extra bits for correct packets	Maintain link latency	Compatible with 802.11	Incremental deployment	Partial Packet Recovery
Checksum	**Maranello**	✓	✓	✓	✓	✓
	Seda [6]			N/A	✓	✓
	FRJ [11]				✓	✓
FEC	ZipTx [14]				✓	✓
PHY layer hints	PPR [12]	✓	✓	N/A		✓
	SOFT [27]	✓	✓	N/A		
Diversity	MRD [20]	✓	✓		✓	
	SPaC [4]	✓	N/A	N/A	✓	
	PRO [16]	✓	✓	✓	✓	

Table 1: Desired behavior and functionality of wireless error recovery protocols

Parameter	**Value**
Hardware	iPhone 3GS [1]
Software	SignalScope Pro [2]
Function	Signal Generator
Output	Headphones
Type	Tone
Frequency	20 Hz to 24 kHz (5 kHz nom)
Amplitude	0.00 dB
Pan	0.000%
Volume	Maximum

Table 1: Experiment parameters for determining the available power from the iPhone 3GS headset port.

Deadlines & camera ready

- Easiest to edit➔smallest to distribute
- Script + data files➔ EPS,PS ➔PDF
- You will have to make changes after submission
- PDF Embedded fonts