Data Science and Making Quality Plots

Rory Hartong-Redden
roryhr@gmail.com

Agenda

- Philosophizing
- Jupyter Notebook
- Examples in the wild (if time)

Discovery of the Higgs Boson

http://home.cern/sites/home.web.cern.ch/files/file/scientists/CCJulAug17_HIGSAT5.pdf

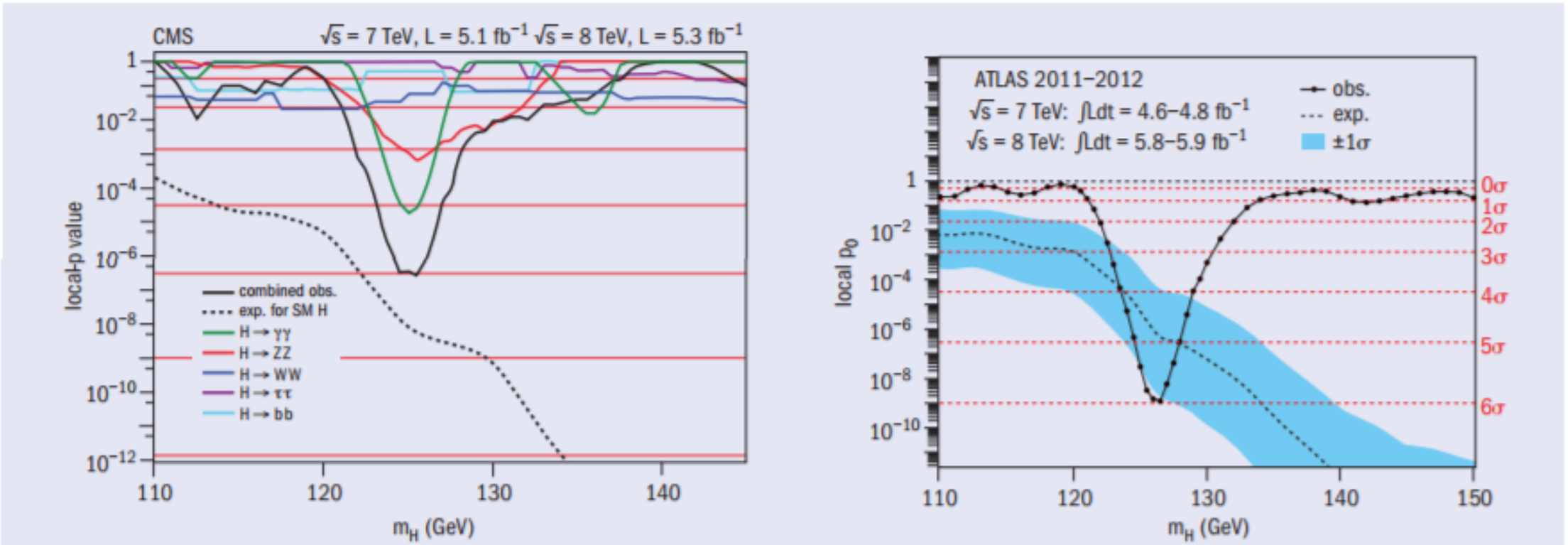


Fig. 3. The discovery of the Higgs boson at ATLAS and CMS, as reported in two papers (arXiv:1207.7214 and arXiv:1207.7235) published after the 4 July announcement. Black lines show the local "p-value", which is the probability that the observation is a statistical fluctuation and not the Higgs boson. This p-value is less than one part in a million, similar to the probability of flipping a coin 21 times and it coming up heads on every occasion, and the significance is peaked at the same mass for both experiments.

Gravitational Waves

https://physics.aps.org/featured-article-pdf/10.1103/PhysRevLett.116.061102

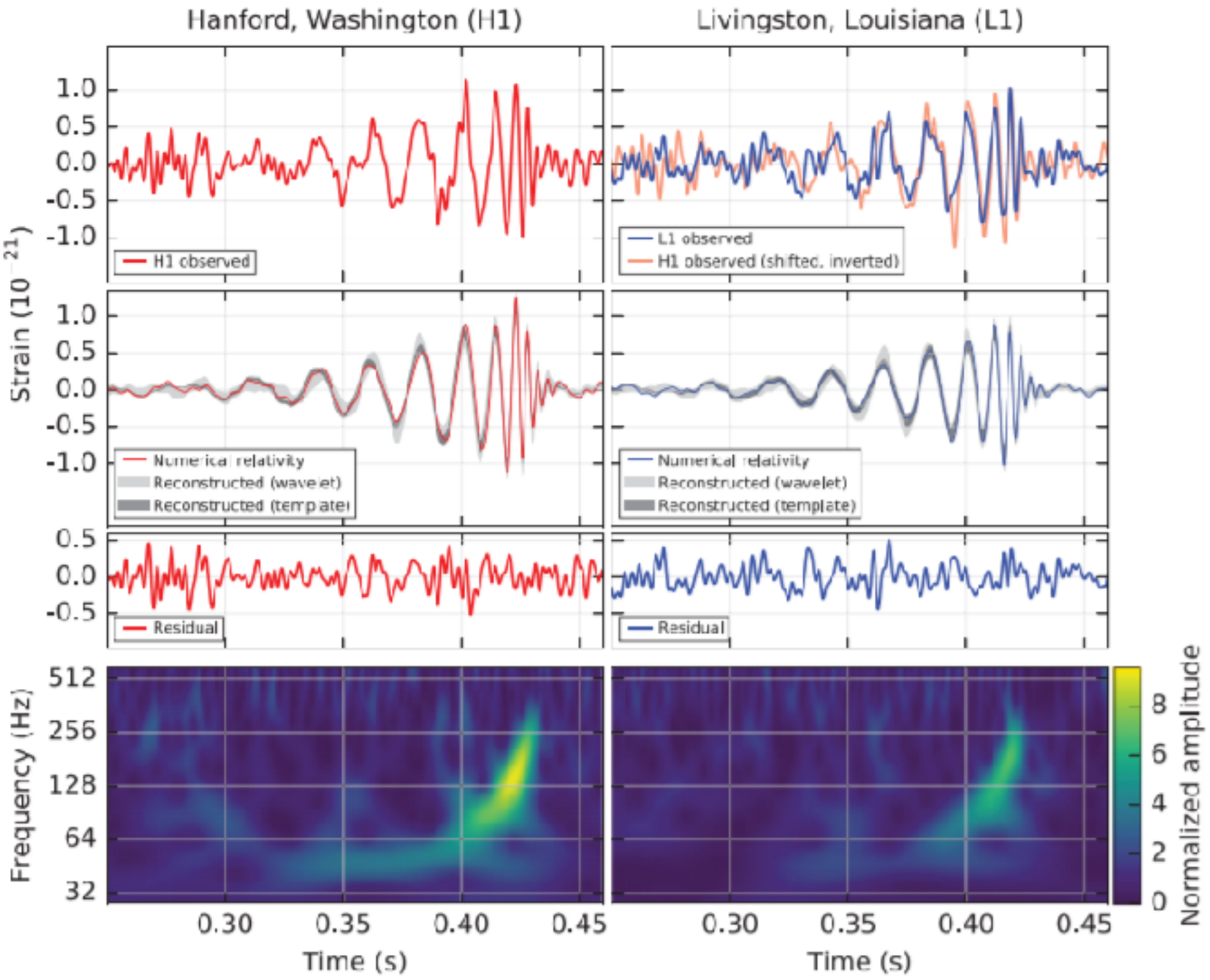
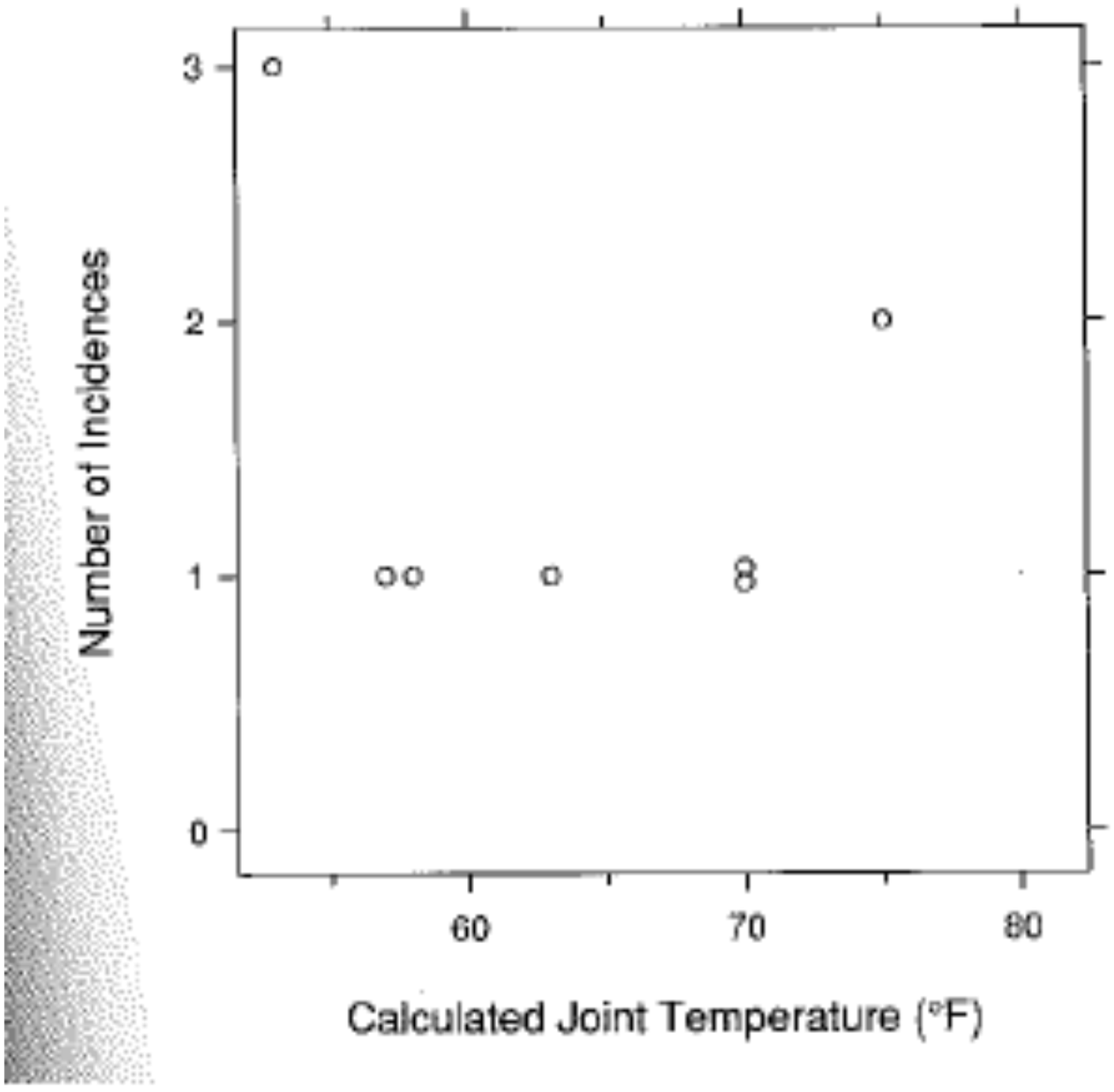


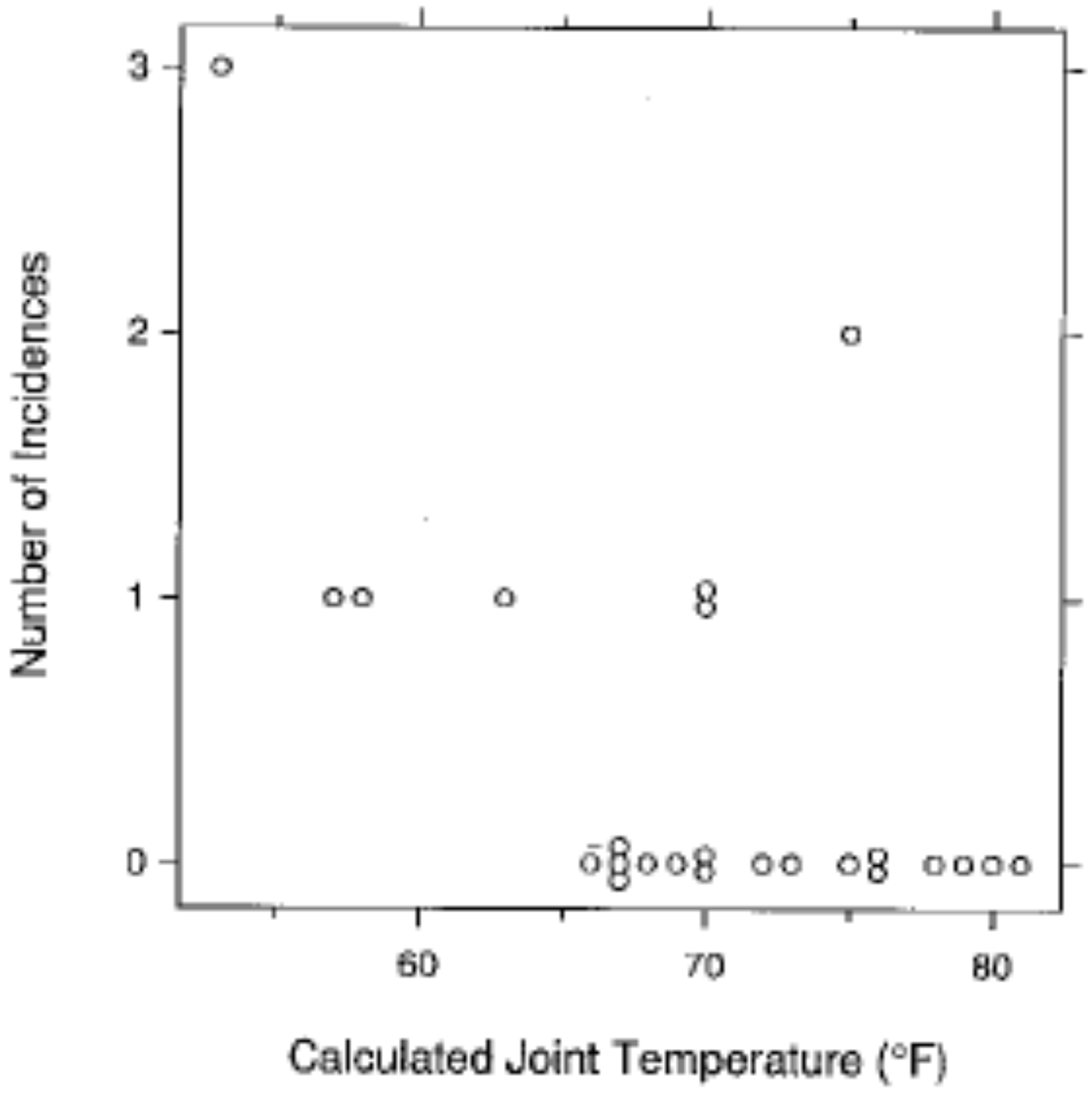
FIG. 1. The gravitational-wave event GW150914 observed by the LIGO Hanford (H1, left column panels) and Livingston (L1, right column panels) detectors. Times are shown relative to September 14, 2015 at 09:50:45 UTC. For visualization, all time series are filtered with a 35–350 Hz bandpass filter to suppress large fluctuations outside the detectors' most sensitive frequency band, and band-reject filters to remove the strong instrumental spectral lines seen in the Fig. 3 spectra. *Top row, left:* H1 strain. *Top row, right:* L1 strain. GW150914 arrived first at L1 and $6.9^{+0.5}_{-0.4}$ ms later at H1; for a visual comparison, the H1 data are also shown, shifted in time by this amount and inverted (to account for the detectors' relative orientations). *Second row:* Gravitational-wave strain projected onto each detector in the 35–350 Hz band. Solid lines show a numerical relativity waveform for a system with parameters consistent with those

Plots tell Stories — Part I



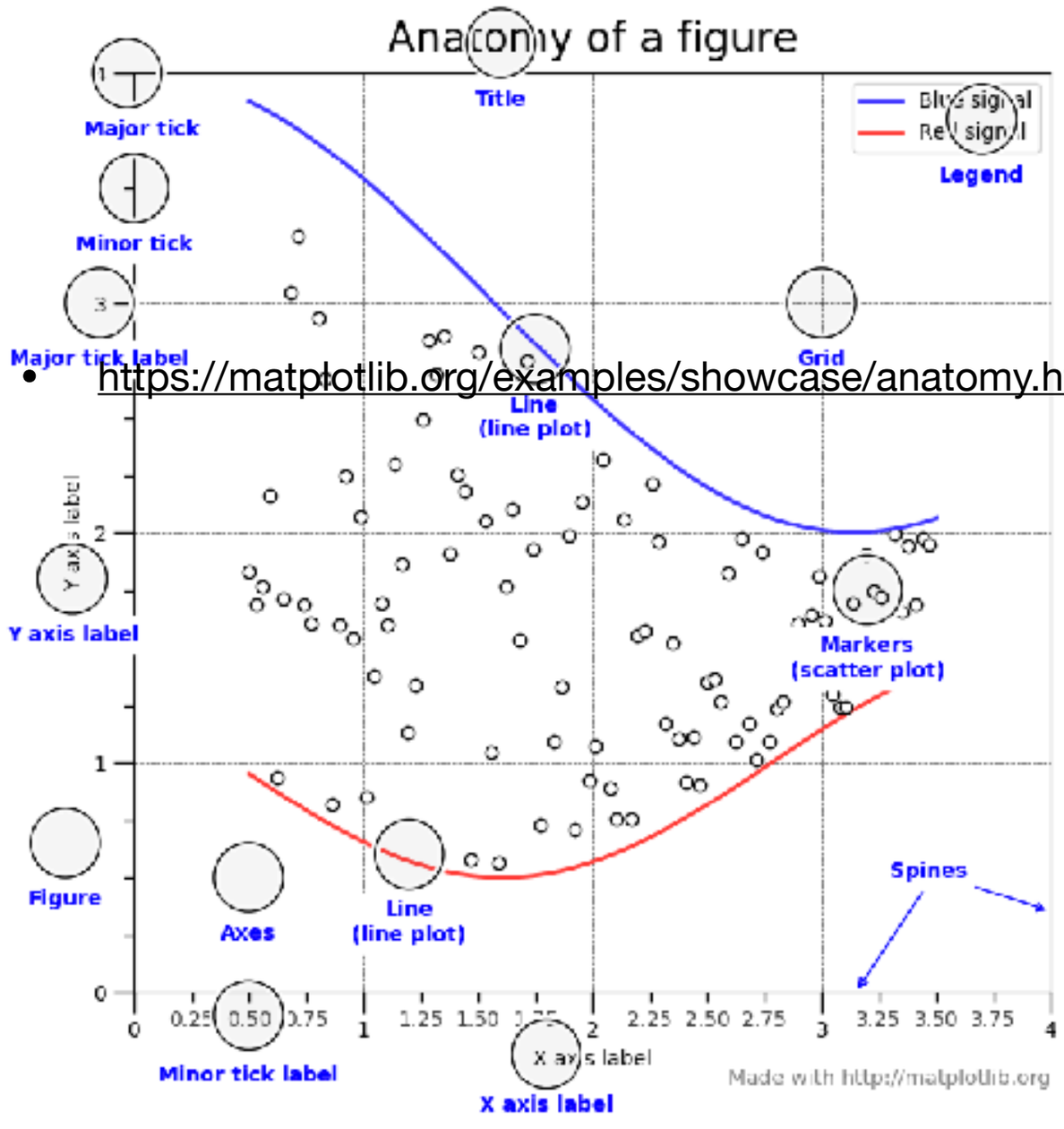
Plots Tell Stories — Part II

- O-ring data for the Challenger disaster
- Forecast temperature was 31°
- Source: The Elements of Graphing Data



Components of a Good Plot

- Title
- Descriptive axes labels (with units)
  - Force (kN)
  - Stress (psi)
- Legend for markers and lines
- Appropriate scale
  - loglog
  - semilogy, semilogx
- Caption which fully describes the plot



- https://matplotlib.org/examples/showcase/anatomy.html

https://matplotlib.org/examples/showcase/anatomy.html

Matplotlib

Pros	Cons
Self-contained	No GUI
Stable	
Well-known API	Weird API
Integrates with Pandas	
Plots look good	Plots don’t look great
Customizable	Too custimizable