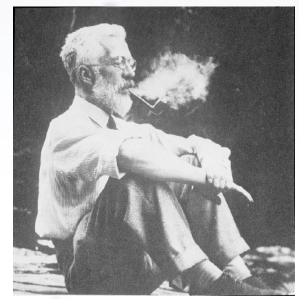
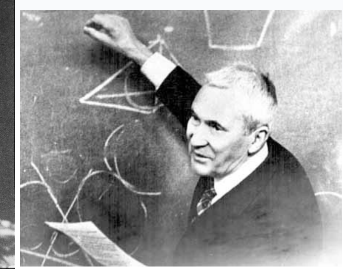
Experimental Physics

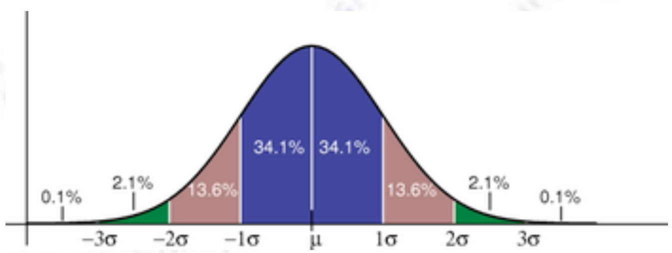
Good Experimental and Statistical practices





Troels C. Petersen (NBI)



"Statistics is merely a quantisation of common sense"

A little bit of statistics…

Types of data

To first order, data comes in two general types and then "the rest":

- **Discreet** (typically counting data, i.e. positive integers)
- **Continuous** (at least more or less)
- **The rest**, i.e. text, images - but typically convertible into two first.

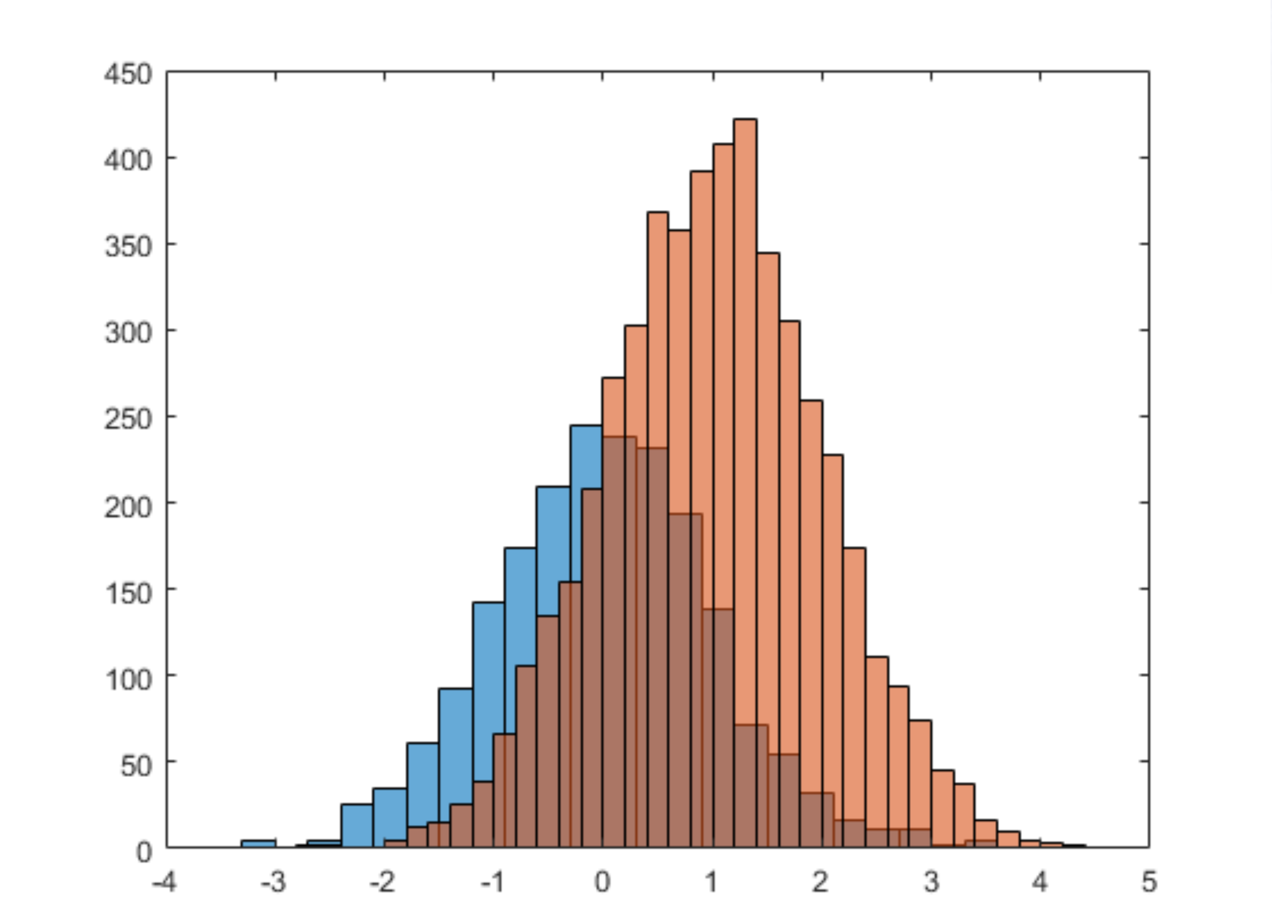
A pitfall is that continuous data is not always continuous, but may seem so!!! The problem arise, if plotting in a histogram with binning comparable (and possibly prime) to steps.

Most basically, one has repeated measurements of things (i.e. 1D distributions). However, often there are several dimensions in the data (possibly 1000s), leading to near-infinite complexity.

Data can be paired in different ways, and/or divisible into groups, experiments, periods, etc. This "meta-data" is important to keep track of.

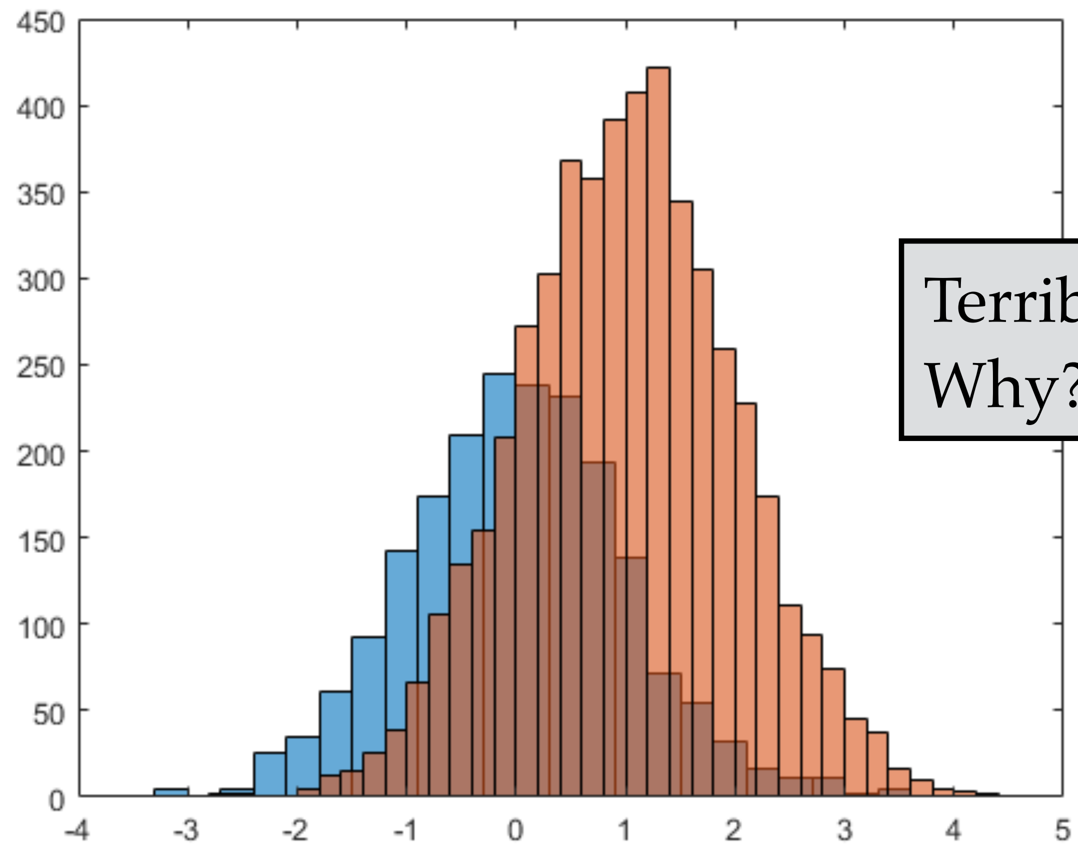
Ways of displaying data

Given repeated measurement of a quantity, the most common way of displaying it, is with a 1D histogram. It is simple and easy to understand, but of course doesn't include more complex information.



Ways of displaying data

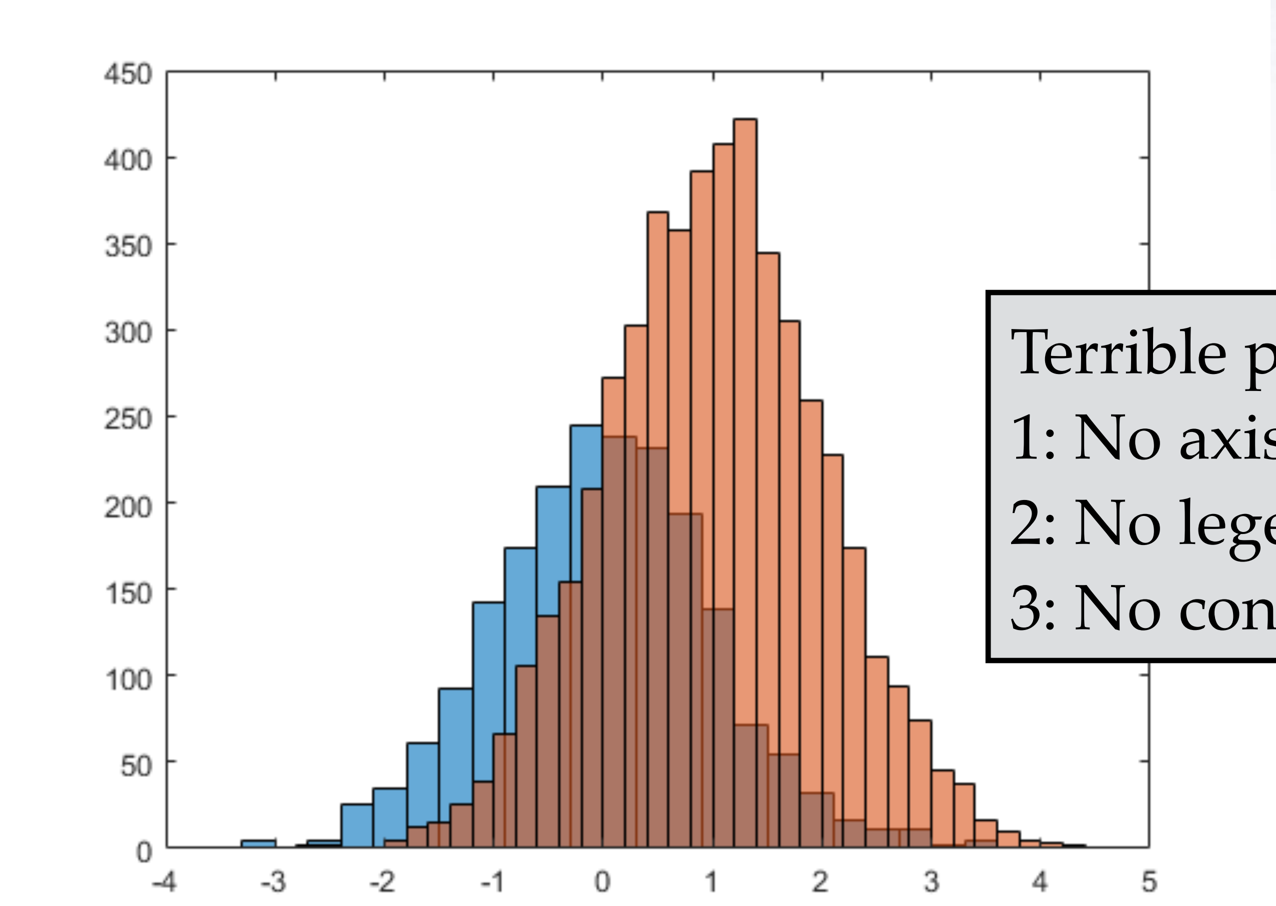
Given repeated measurement of a quantity, the most common way of displaying it, is with a 1D histogram. It is simple and easy to understand, but of course doesn't include more complex information.



Terrible plot!!!
Why?

Ways of displaying data

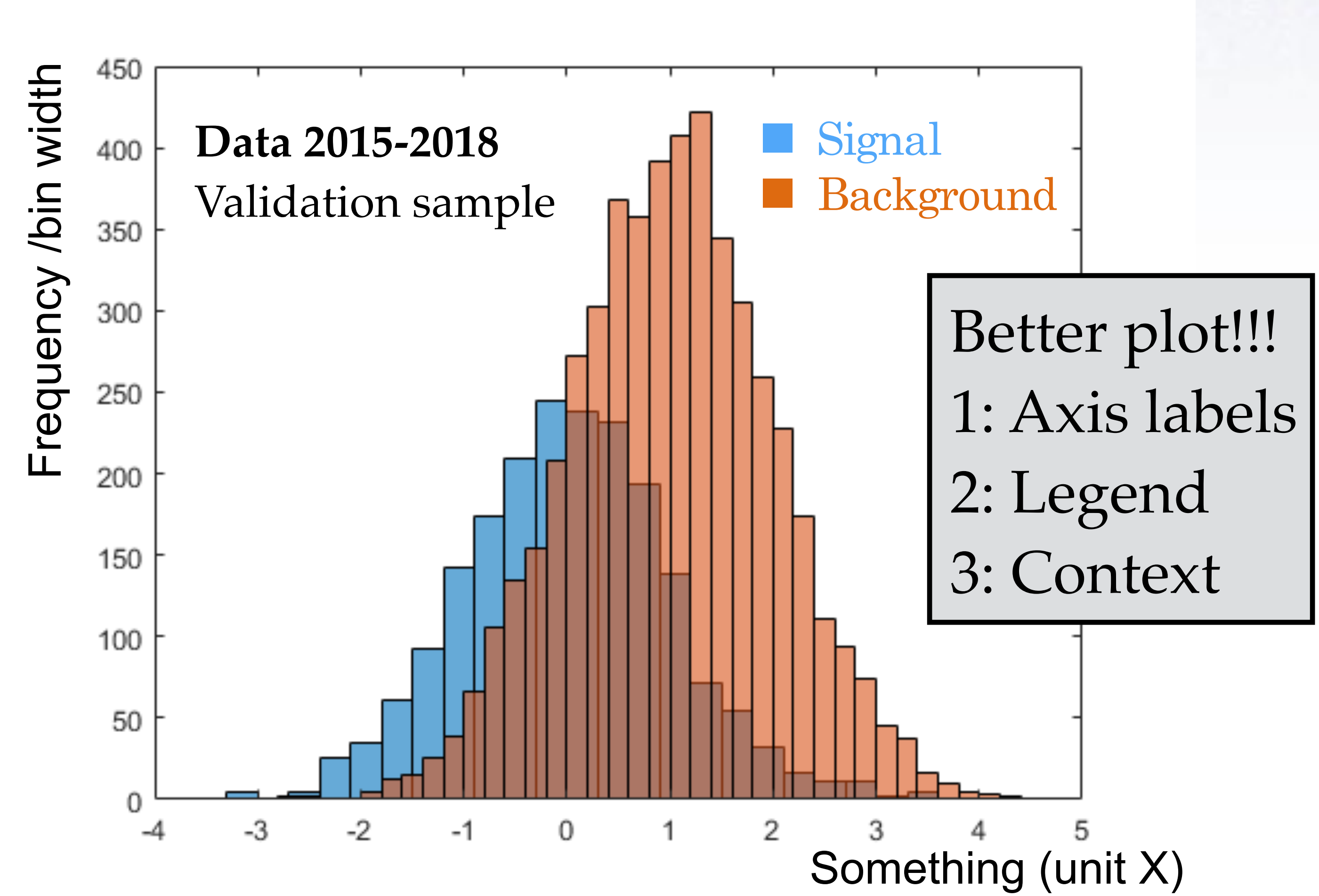
Given repeated measurement of a quantity, the most common way of displaying it, is with a 1D histogram. It is simple and easy to understand, but of course doesn't include more complex information.



Terrible plot!!!
1: No axis labels
2: No legend
3: No context

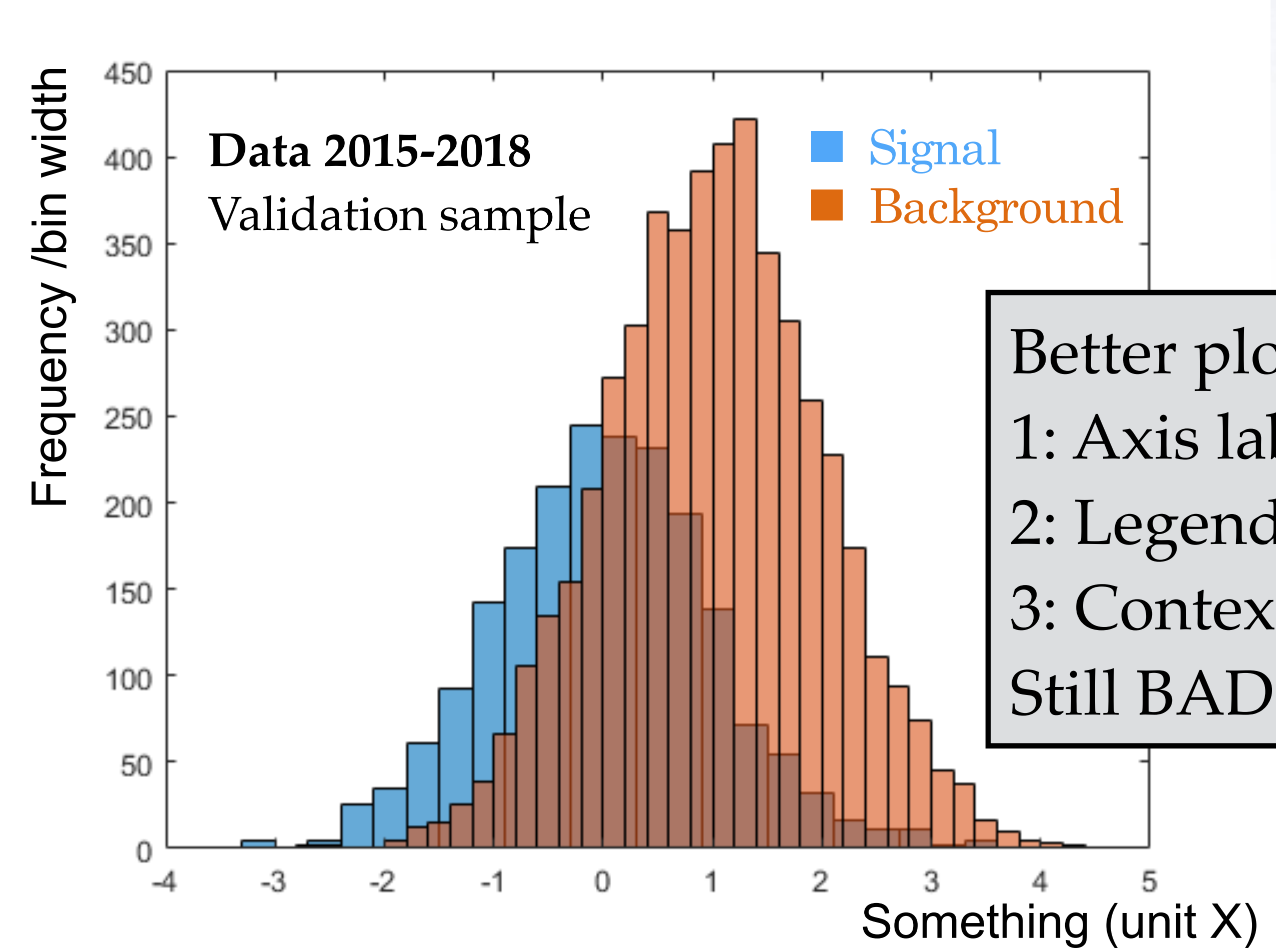
Ways of displaying data

Given repeated measurement of a quantity, the most common way of displaying it, is with a 1D histogram. It is simple and easy to understand, but of course doesn't include more complex information.



Ways of displaying data

Given repeated measurement of a quantity, the most common way of displaying it, is with a 1D histogram. It is simple and easy to understand, but of course doesn't include more complex information.

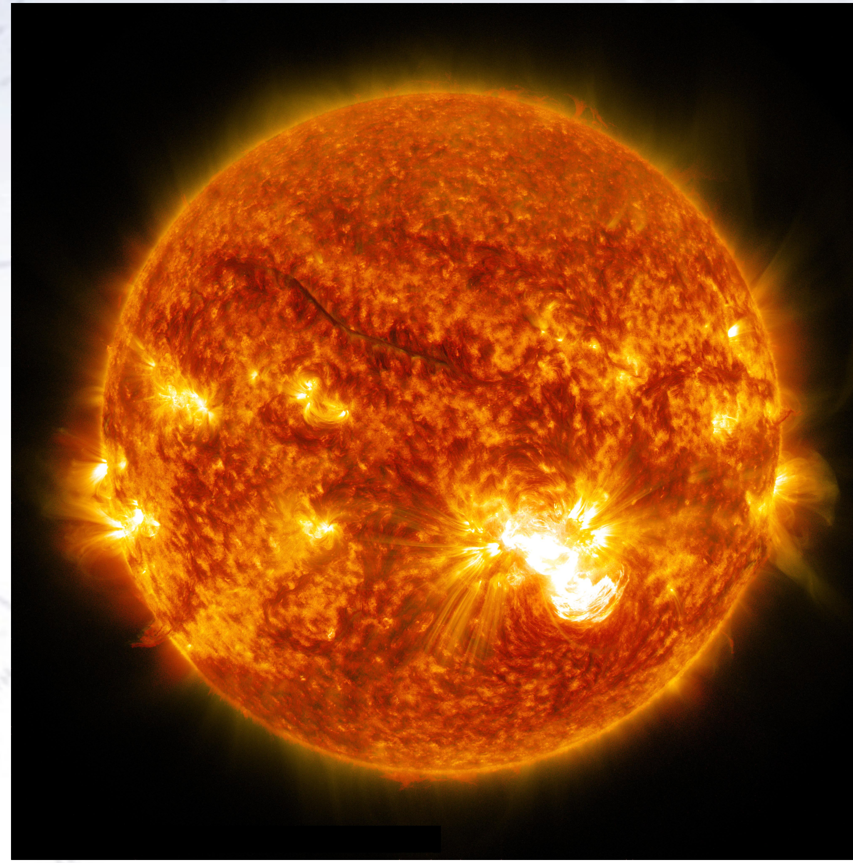


Better plot!!!
1: Axis labels
2: Legend
3: Context
Still BAD! Why?

Coincidence?

Seen from the Earth, the angular sizes of the Moon and Sun are uncorrelated, but:

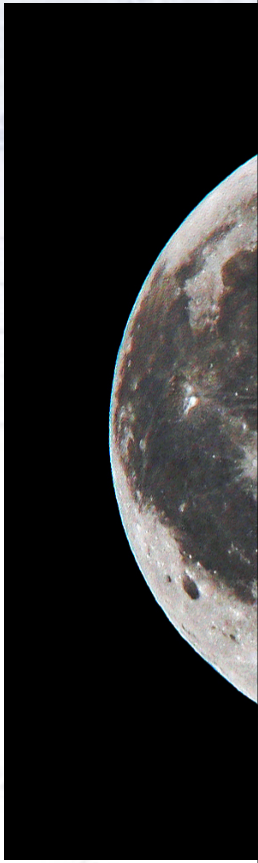
$\theta_{moon} = 0.558^{\circ}$

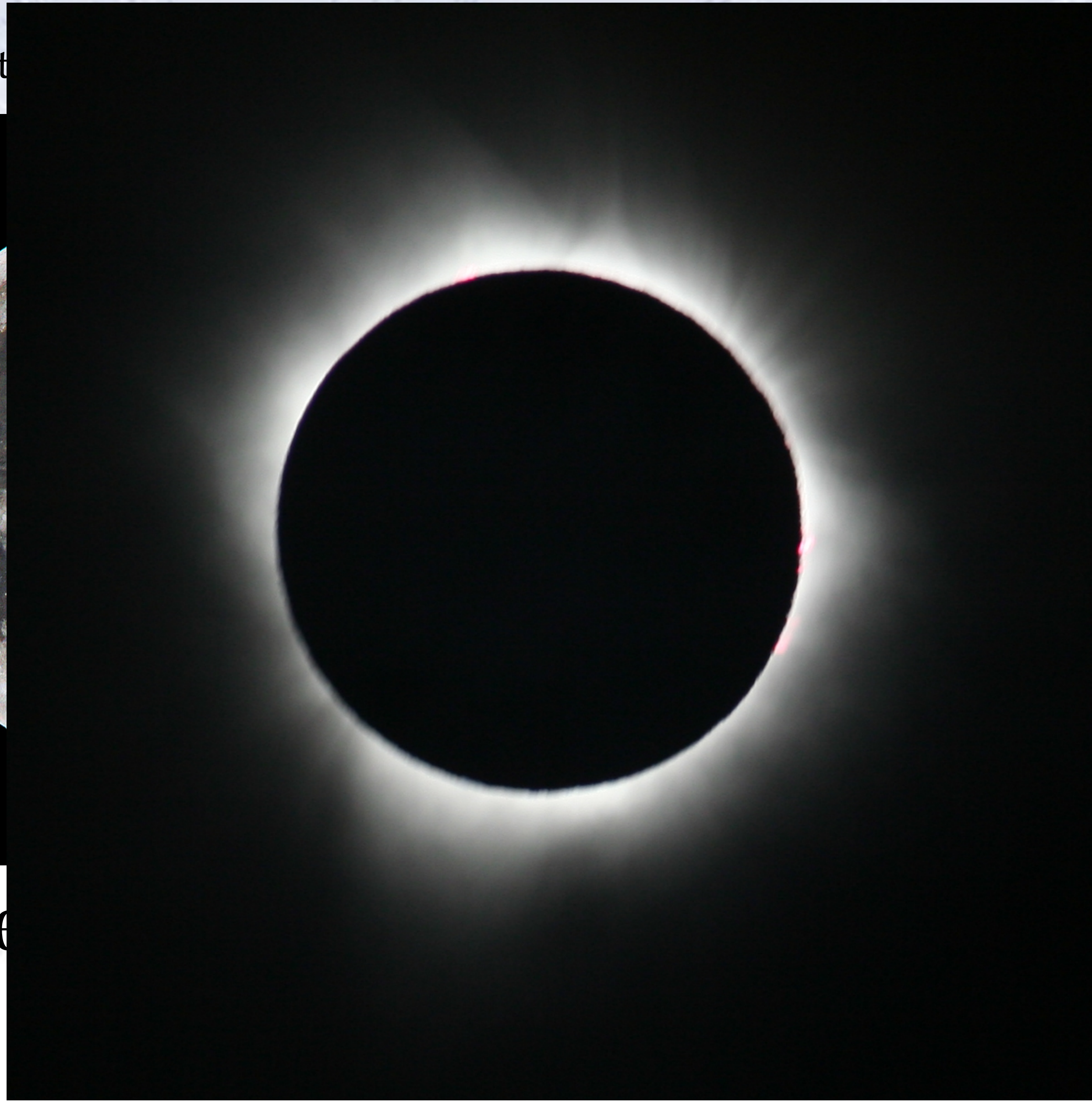


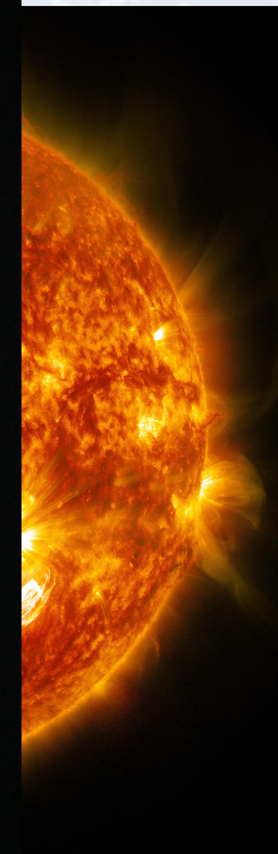
$\theta_{sun} = 0.545^{\circ}$

Coincidence?

Seen from t … elated, but:







6 … °

Coincidence?

“Coincidences, in general, are great stumbling-blocks in the way of that class of thinkers who have been educated to know nothing of the theory of probabilities [and statistics (red.)] - that theory to which the most glorious objects of human research are indebted for the most glorious of illustration.”

[Edgar Allan Poe, The Murders in the Rue Morgue]

Coincidence?

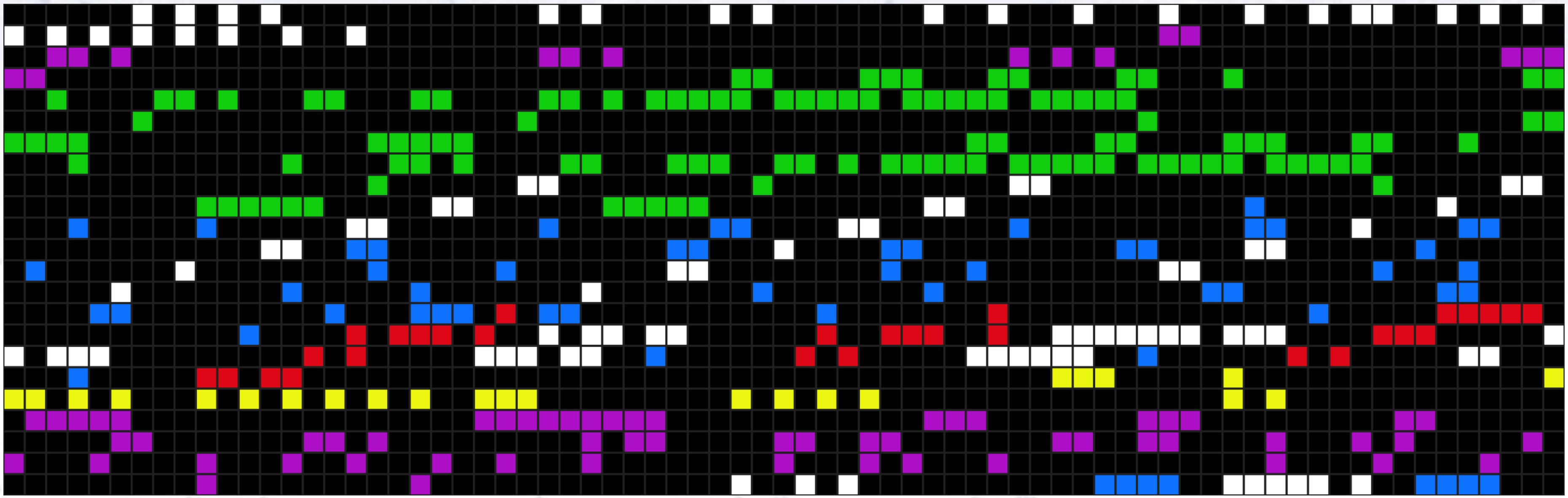
"Coincidences, in general, are great stumbling-blocks in the way of that class of thinkers who have been educated to know nothing of the theory of probabilities [and statistics (red.)] - that theory to which the most glorious objects of human research are indebted for the most glorious of illustration."

[Edgar Allan Poe, The Murders in the Rue Morgue]

The aim of science in general and statistics in particular is
NOT to be misled by coincidences,
but rather to calculate probabilities and judge their significance.

Random???

You sit and play around with a large radio telescope (as one does!), and you receive 1679 data points that you (wisely) decide to plot in a (random?) format (the colors are artificial!):



Is this random noise, or is there some pattern to it? Well, that is hypothesis testing, as we have two hypothesis:

H_0 - Null: It is random.

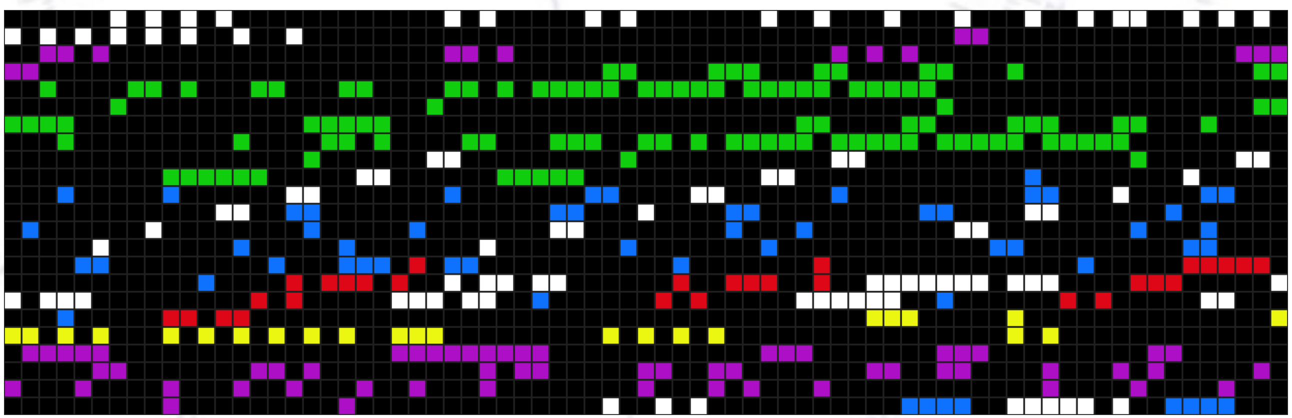
H_1 - Alternative: It is not random.

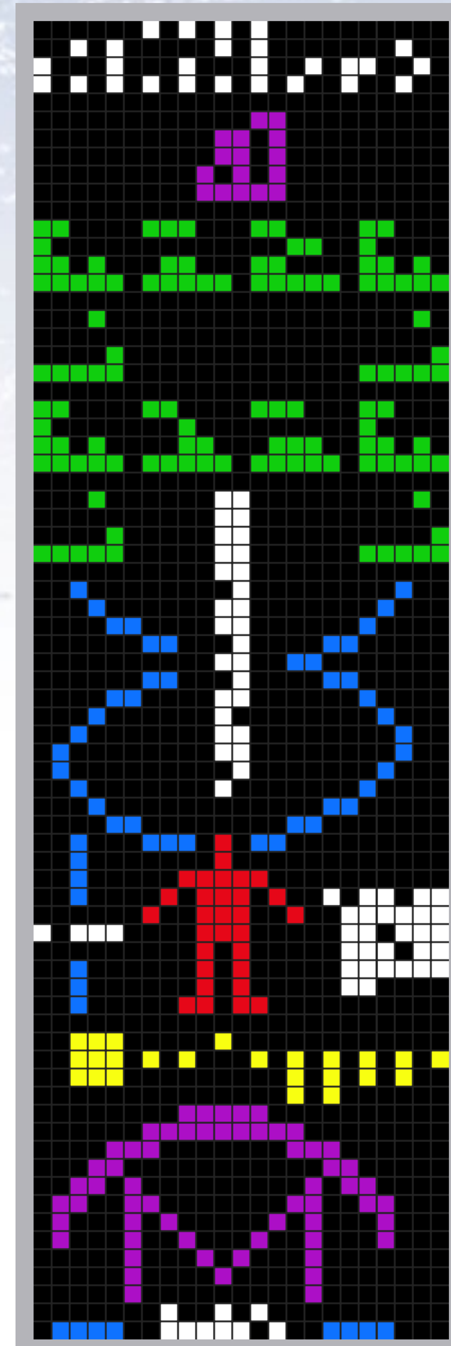
How would you test this?

Random???

It becomes very clear, that this is NOT random. That does not prove that it is anything interesting, but simply that you didn't pick up pure noise in your apparatus.

You realise the "right" format (23 rows x 73 columns), and suddenly a message appears… The Arecibo message (sent into space in 1974).

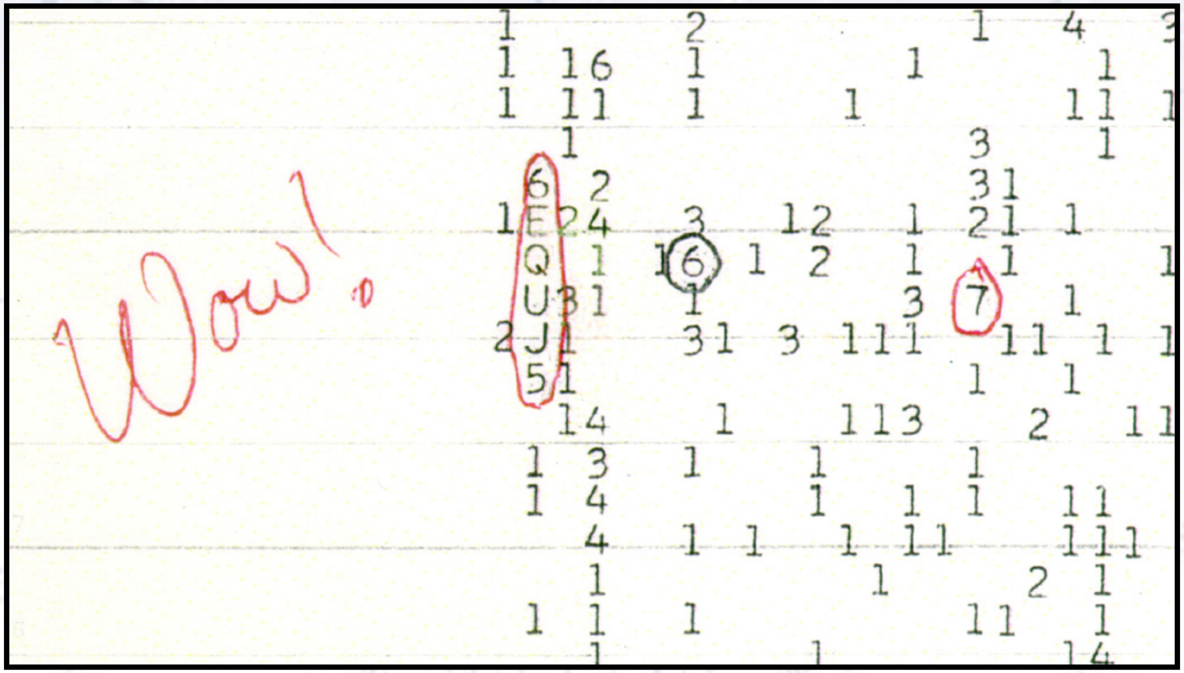


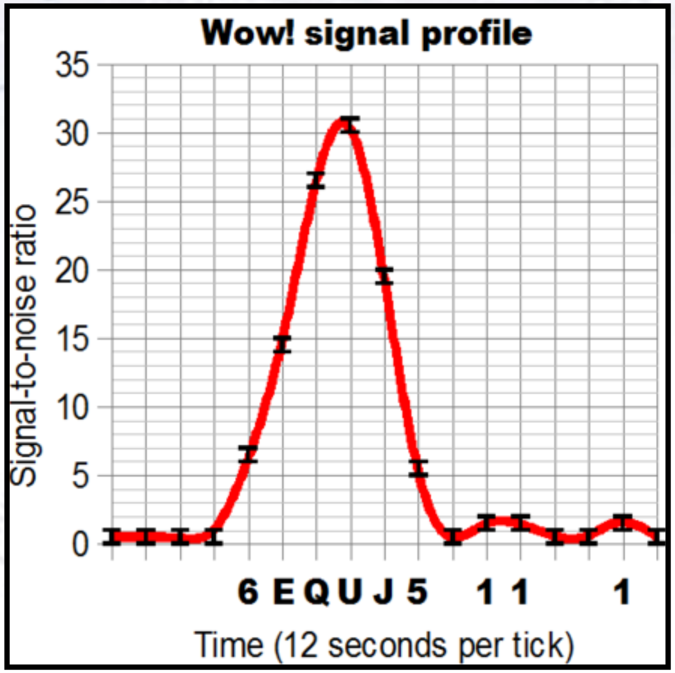


https://en.wikipedia.org/wiki/Arecibo_message

Signal???

However, one should always be careful of not being fooled by spurious signal, effects of apparatus or unknown backgrounds.





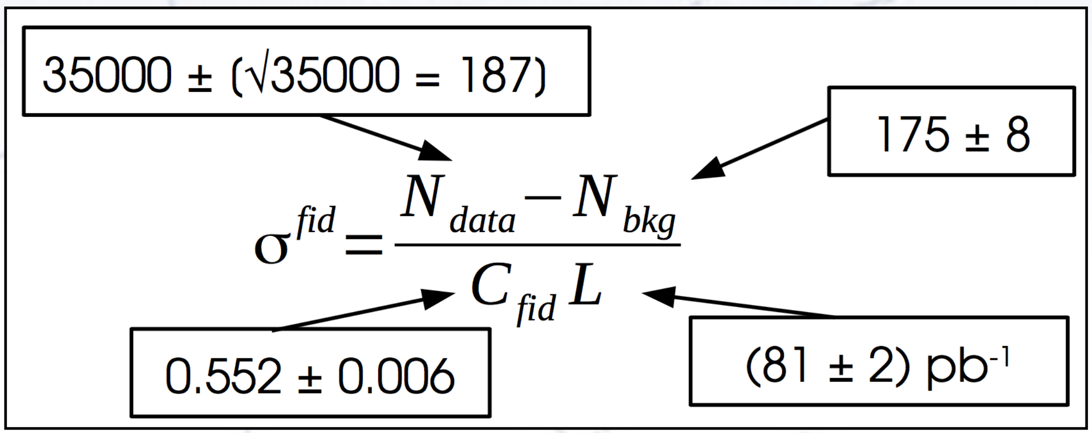
Given Earth's rotation, an extraterrestrial signal was expected to be 72s long and only observed in one (of 50) 10kHz bands - exactly like observed!

The signal could have been produced by an unknown (human) background.

Error propagation

The simplest experiments and subsequent extraction of a measurement are usually the best!

When you count something, then there is (mostly) not doubt about the result:



$$\sigma^{fid} = \frac{N_{data} - N_{bkg}}{C_{fid} L}$$

Error propagation

The simplest experiments and subsequent extraction of a measurement are usually the best!

When you count something, then there is (mostly) not doubt about the result:

$$\sigma^{fid} = \frac{N_{data} - N_{bkg}}{C_{fid} L}$$

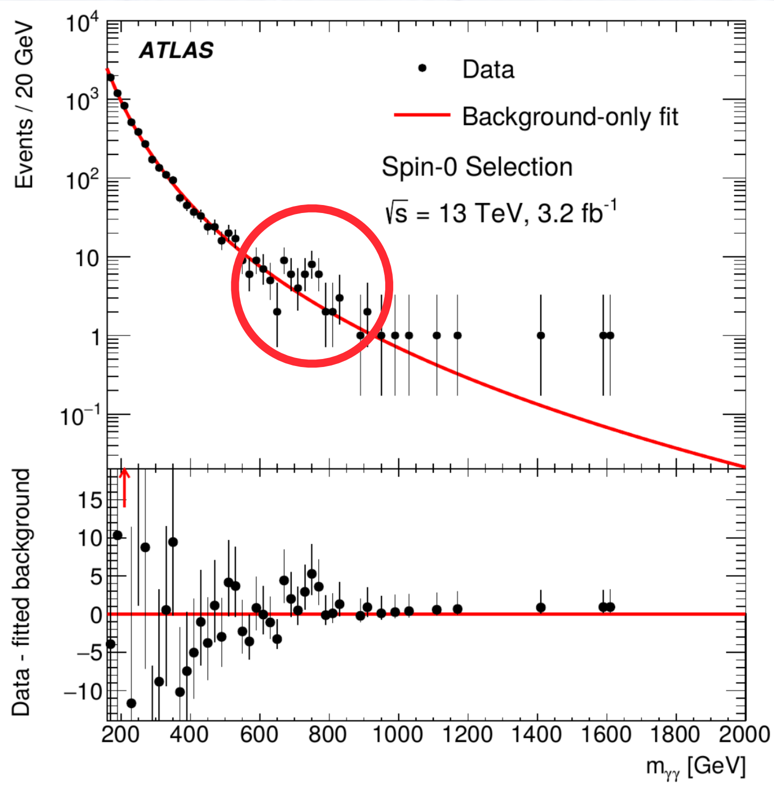
- N_{data}: 35000 ± (√35000 = 187)
- N_{bkg}: 175 ± 8
- C_{fid}: 0.552 ± 0.006
- L: (81 ± 2) pb^{-1}

Quantifying your uncertainties is hard work, that requires forethought! Did you carefully think about your experimental setup and data taking **before** you actually did the experiment?

σ^{fid} = 0.781 ± 0.004 (stat) ± 0.008 (syst) ± 0.016 (lumi)

Significance of signal

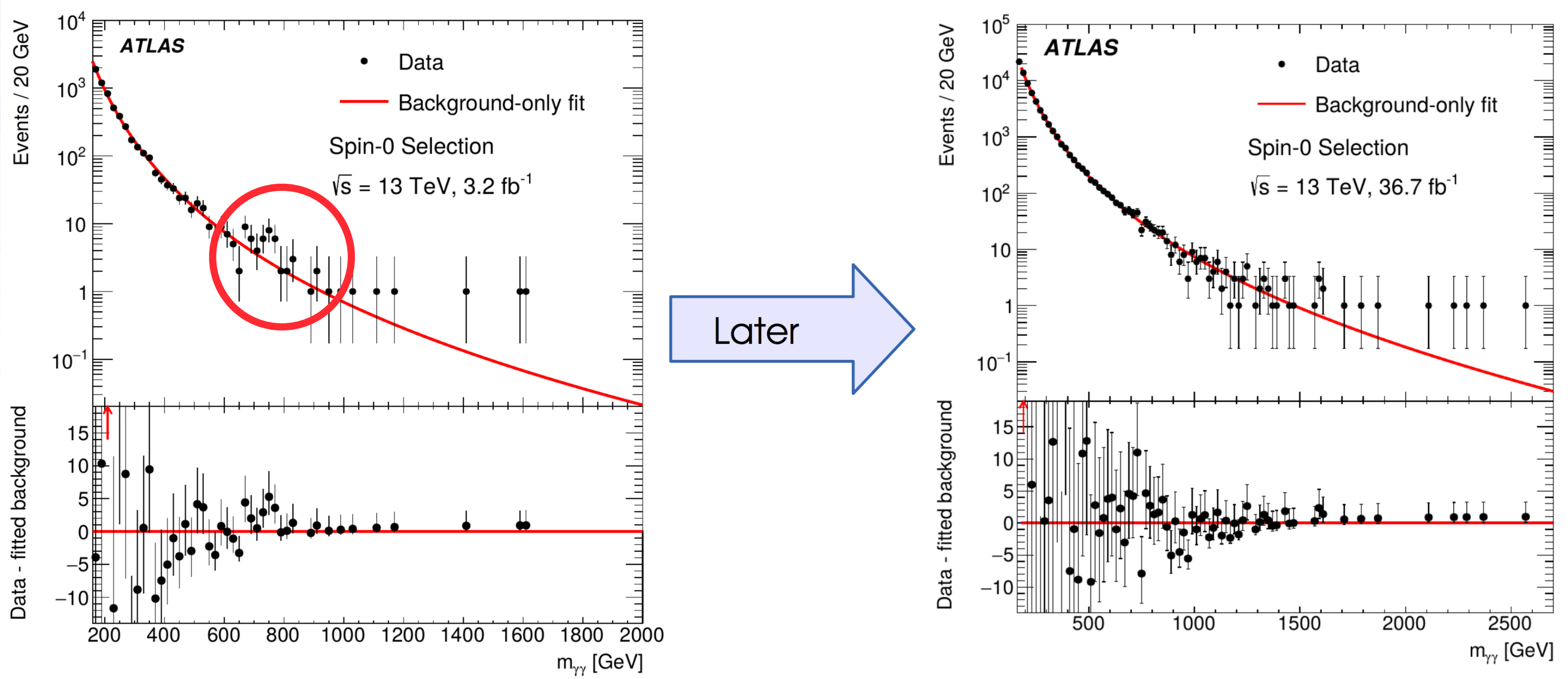
You do an experiment, consider data, and see something that could be a signal:



How do you determine, if it is actually real, or if it is just a statistical fluctuation of noise?

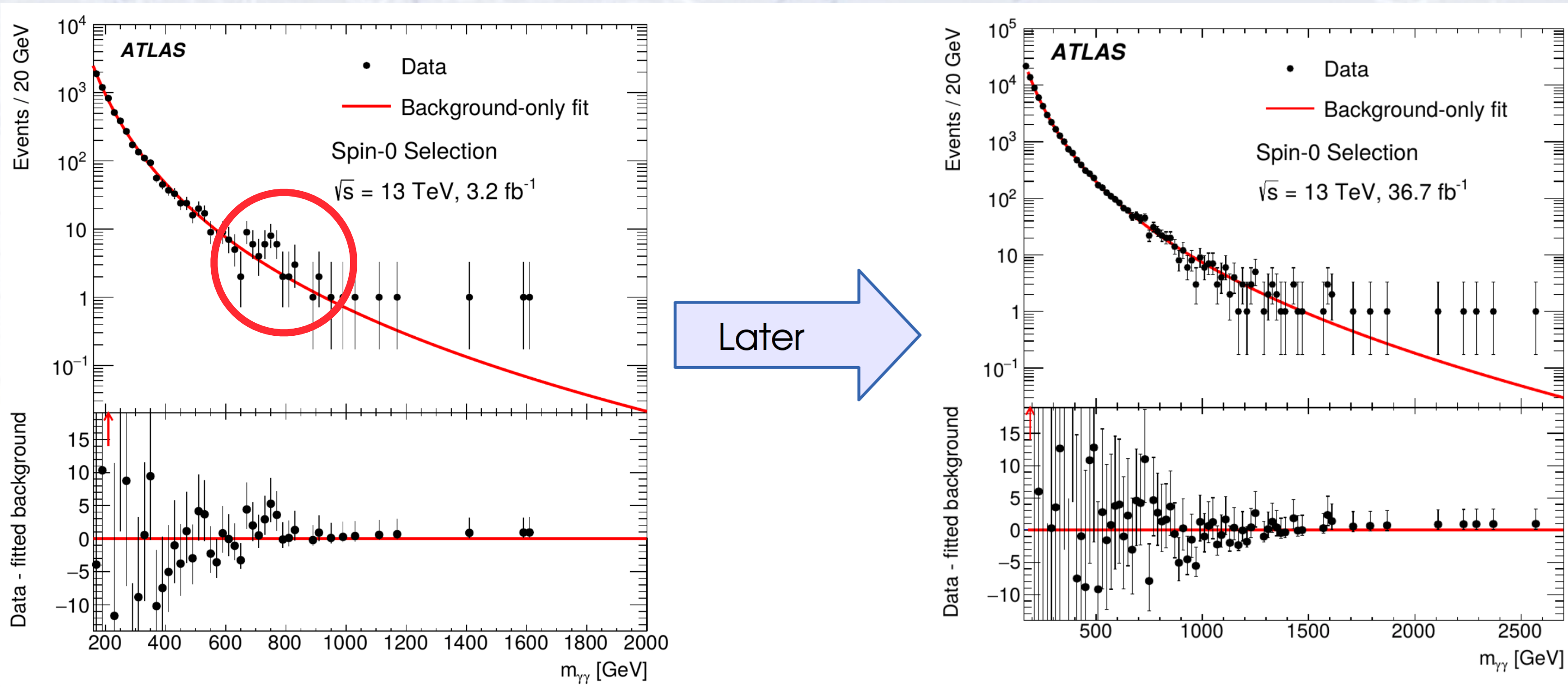
Significance of signal

You do an experiment, consider data, and see something that could be a signal:



Significance of signal

You do an experiment, consider data, and see something that could be a signal:



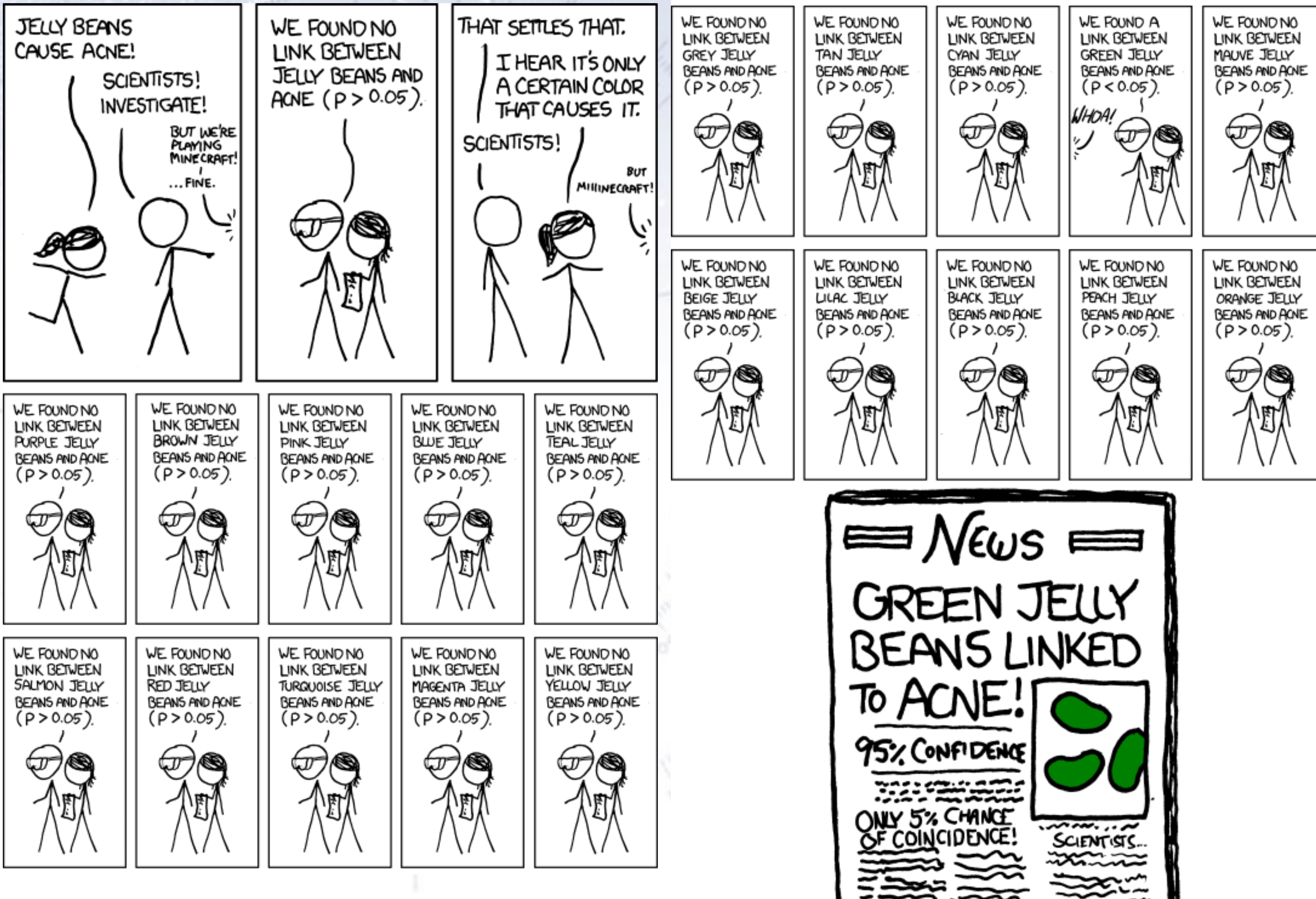
Global p-value →

$$p_{global} = 1 - \left(1 - p_{local}\right)^N \approx N \; p_{local}$$

← Local p-value

↑ trials factor, number of independent experiments

XKCD on statistics



XKCD on statistics

XKCD on statistics

Trial factor: 20

Defining the Chi-Square

Problem Statement: Given N data points (x,y), adjust the parameter(s) θ of a model, such that it fits data best.

The best way to do this, given uncertainties σ_i on y_i is by minimising:

$$\chi^2(\theta) = \sum_{i}^{N} \frac{(y_i - f(x_i, \theta))^2}{\sigma_i^2}$$

The power of this method is hard to overstate!

Not only does it provide a simple, elegant and unique way of fitting data, but more importantly it provides a **goodness-of-fit measure**.

This is the Chi-Square test!

Optimal data analysis

As an observational astronomer, you have survived a long airline flight, a terrifying taxi ride to the summit of a volcano, days of dodgey weather, hours coaxing flakey equipment back into an orderly lifestyle, exhaustion. At last, you attain that exalted state of resonance with machine and sky. Your equipment is working in miraculous defiance of Murphy's Law.

Everything that could go wrong did, but now you have emerged to savour a long clear night plucking data from the sky. Thus you succeed in acquiring an astronomical dataset. **After such an ordeal, giving birth to the data, it seems shameful, even criminal, to analyse your data with anything less than optimal methods.**

[Keith Horne (Univ. of St. Andrews), "The Ways of Our Errors" (book in preparation)]

A little bit on experiments…

Measurement situation

There are four possible situations in experimental measurements of a quantity:

One measurement, no error:

X = 3.14

Situation: You are f***ed!

You have no clue about uncertainty, and you can not obtain it!

Several measurements, no errors:

X1 = 3.14
X2 = 3.21
X3 = …

Situation: You are OK

You can combine the measurements, and from RMS get error on mean.

One measurement, with error:

X = 3.14 ± 0.13

Situation: You are OK

You have a number with error, which you can continue with.

Several measurements, with errors:

X1 = 3.14 ± 0.13
X2 = 3.21 ± 0.09
X3 = …

Situation: You are on top of things!

You can both combine to a weighted, average and check with a chi-square.

Pendulum - comments

Time measurement:

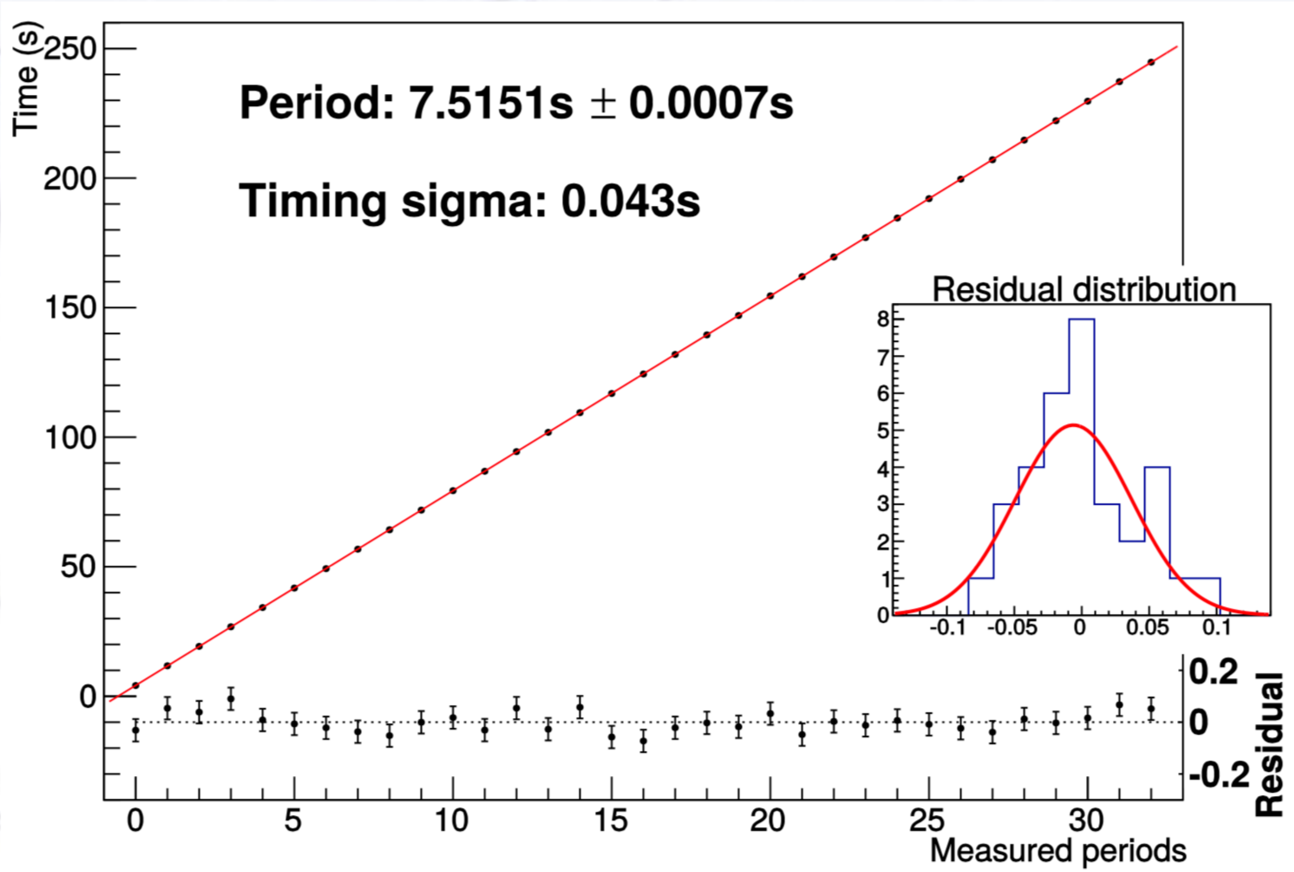
Many independent measurements, little systematic $\Rightarrow$ Good error estimate

Length measurement:

Some independent measurements but also some systematics $\Rightarrow$ check difference between instruments.

You can not reduce the uncertainty by multiple measurements, if the main limitation is some inherent systematic!

Several groups managed to get uncertainties below 0.1%.

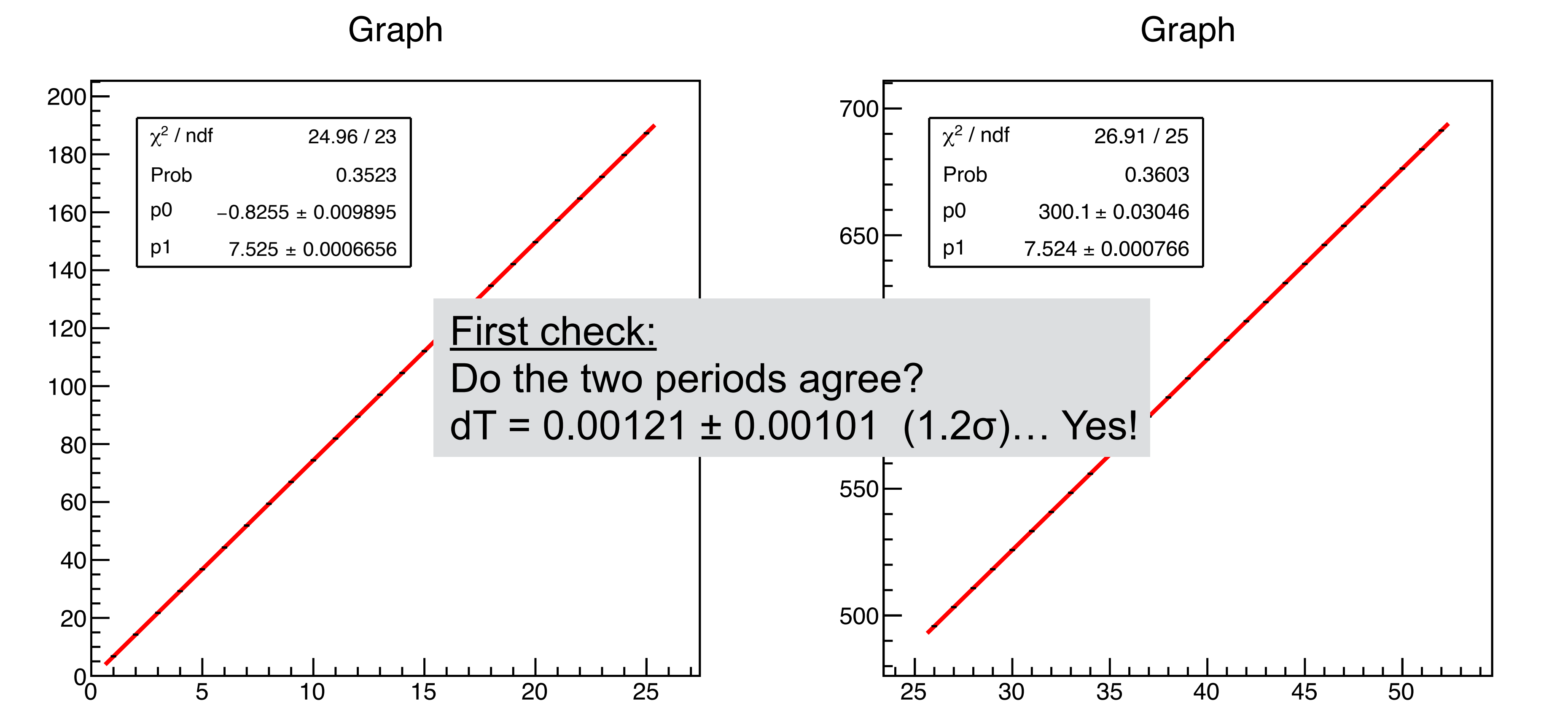


Additional ideas

It is possible to “leave” the pendulum swinging between to sets of measurements. This maximises the period over which you measure, without requiring your activity all the time…

Additional ideas

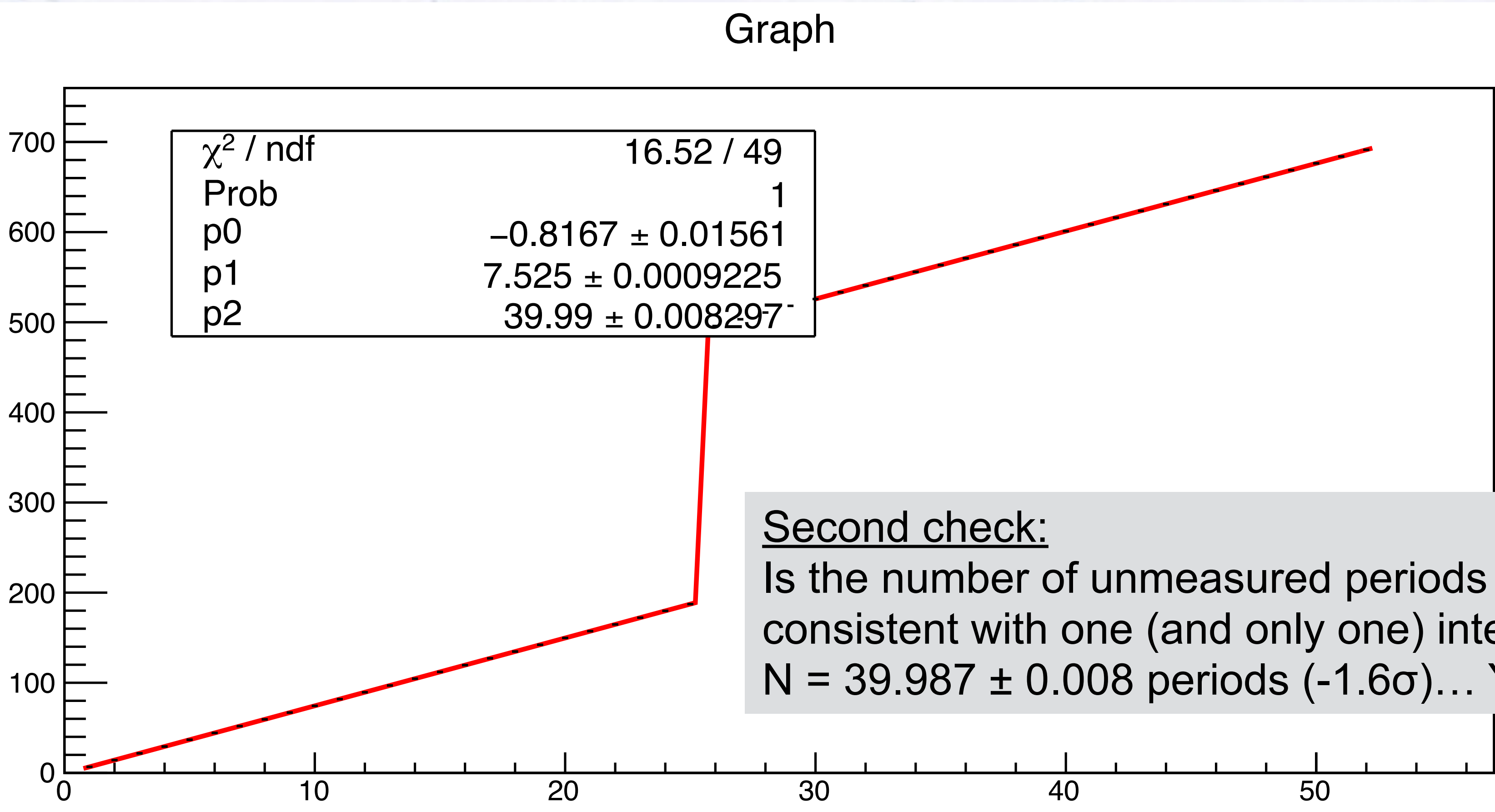
It is possible to “leave” the pendulum swinging between to sets of measurements. This maximises the period over which you measure, without requiring your activity all the time…



First check:
Do the two periods agree?
dT = 0.00121 ± 0.00101 (1.2σ)… Yes!

Additional ideas

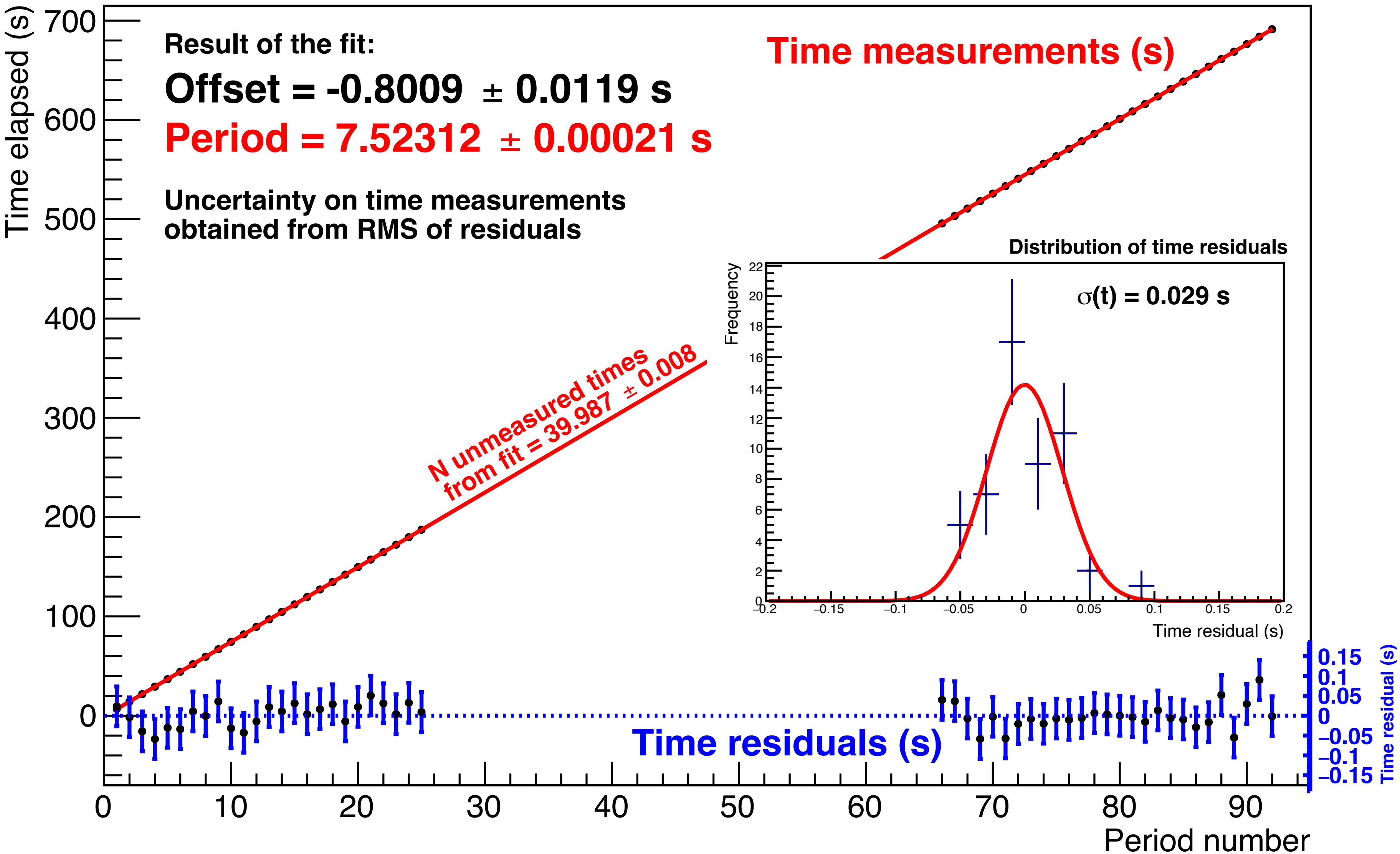
It is possible to "leave" the pendulum swinging between to sets of measurements. This maximises the period over which you measure, without requiring your activity all the time…



<u>Second check:</u>
Is the number of unmeasured periods consistent with one (and only one) integer?
N = 39.987 ± 0.008 periods (-1.6σ)… Yes!

Additional ideas

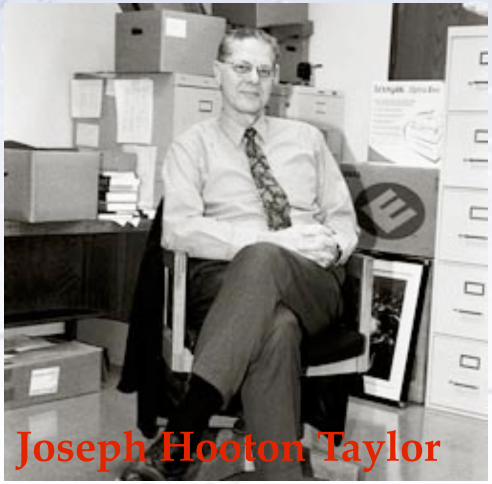
After checks, fit the entire time span to get "insanely great" precision.

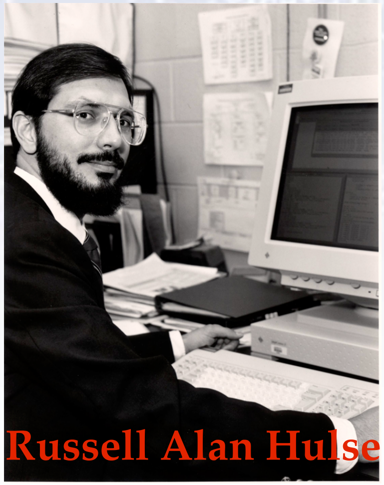


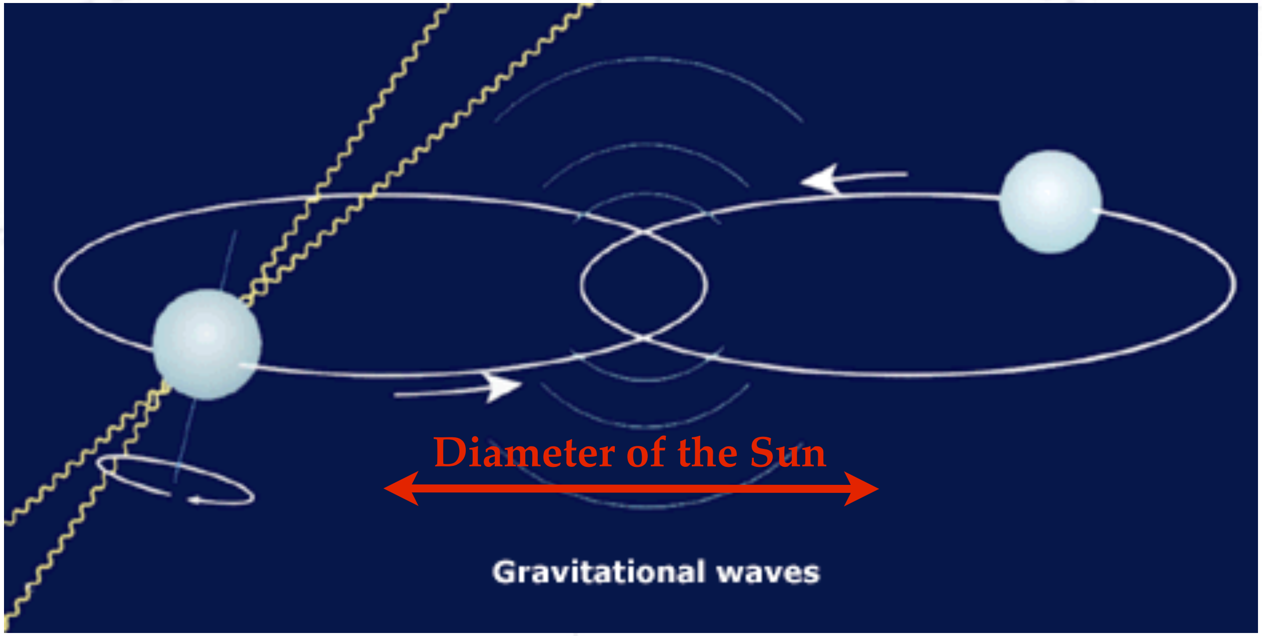

A small intermezzo…
…the Hulse-Taylor-pulsar

Hulse-Taylor-pulsaren

In 1973 Hulse & Taylor discovered a very special pulsar… the period for its signal was NOT constant, but had a variation with a period of 8 hours! As it turns out, this was to become a "jewel" in the test of Einstein's theory of relativity.







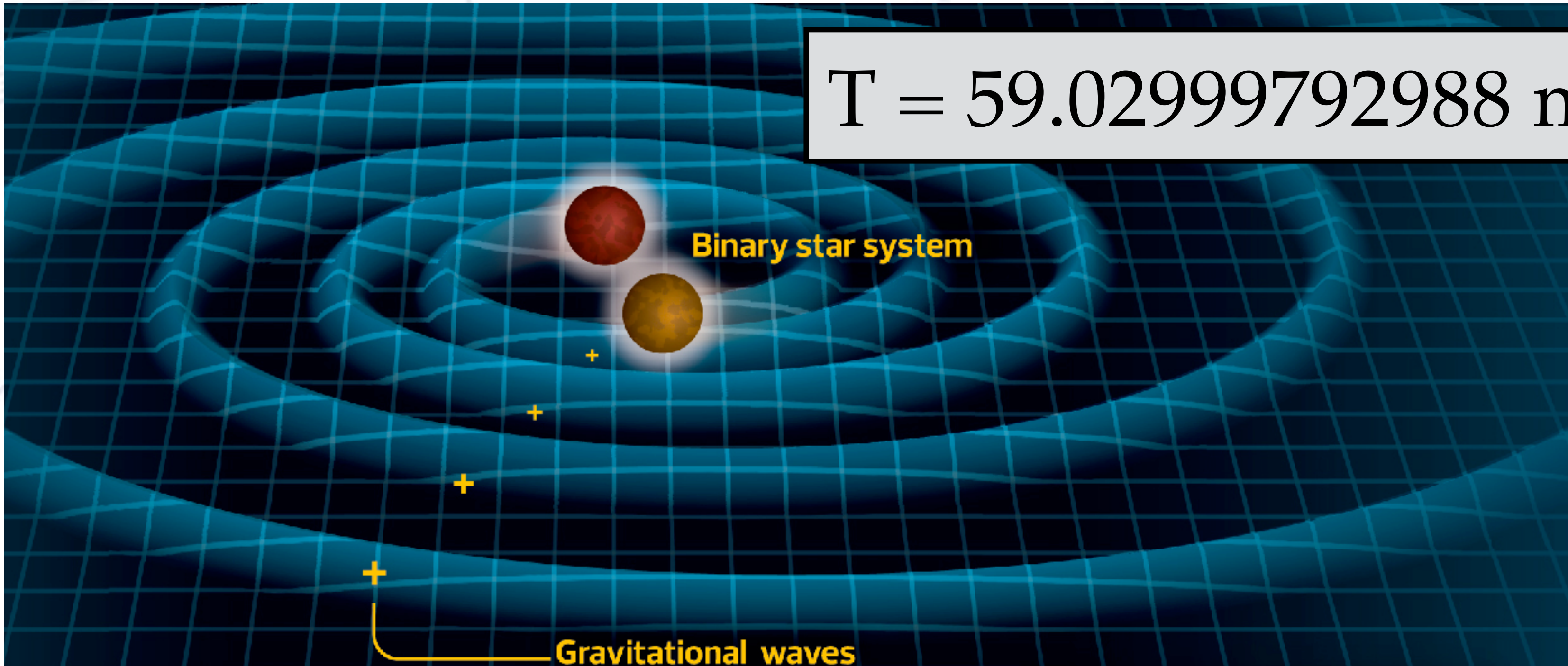
Præcisionsmålinger

In the following years, they measured the pulsar parameters with great precision:

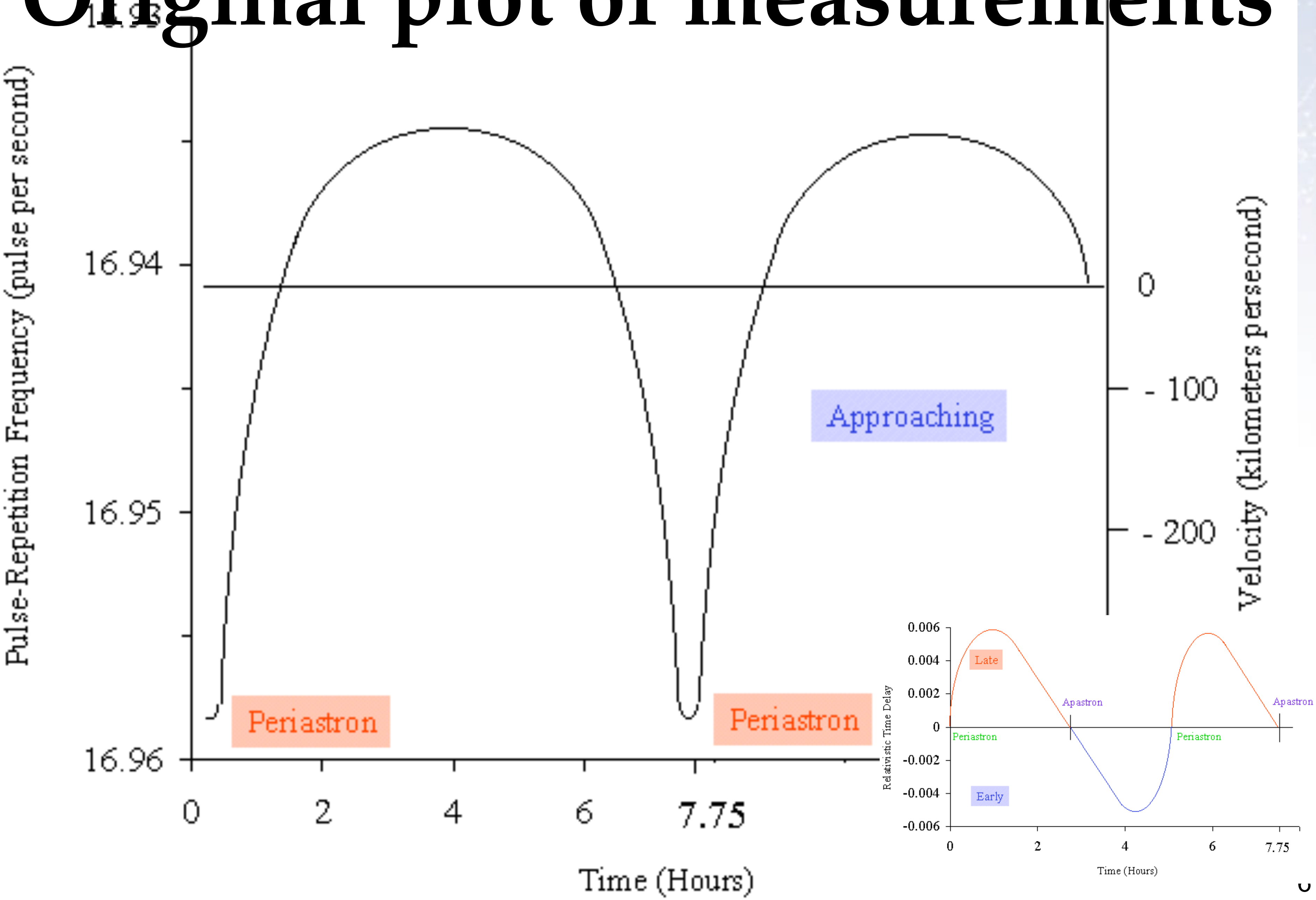
- Mass of companion: 1.387 MSun
- Total mass of the system: 2.828378(7) MSun
- Orbital period: 7.751938773864 hr
- Eccentricity: 0.6171334
- Semimajor axis: 1,950,100 km
- Periastron separation: 746,600 km
- Apastron separation: 3,153,600 km
- Orbital velocity of stars at periastron (relative to center of mass): 450 km/s
- Orbital velocity of stars at apastron (relative to center of mass): 110 km/s

The measurements were possible, partly because of the large relativistic effects. What takes **a century** for Mercury, takes **a day** for the Hulse-Taylor-pulsar!

T = 59.02999792988 ms



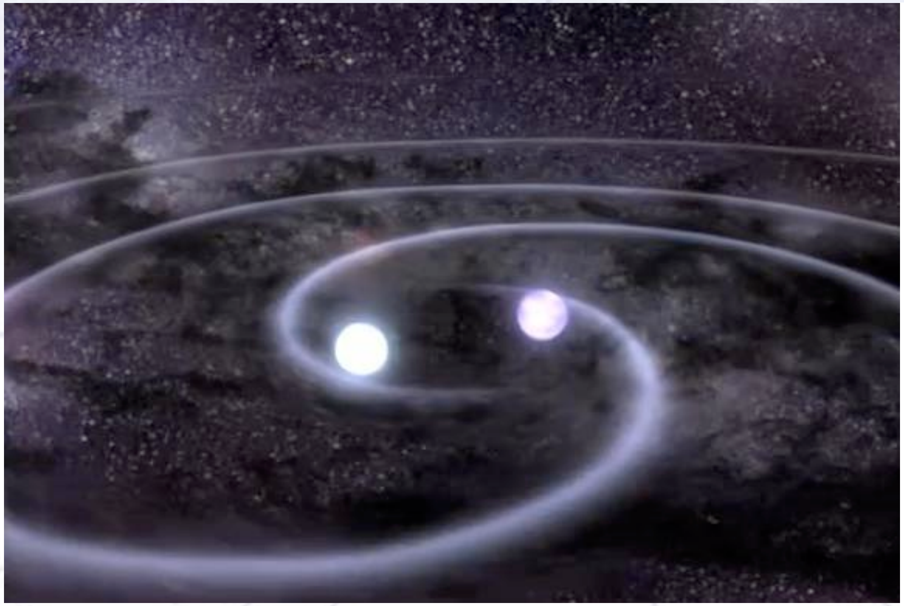
Original plot of measurements

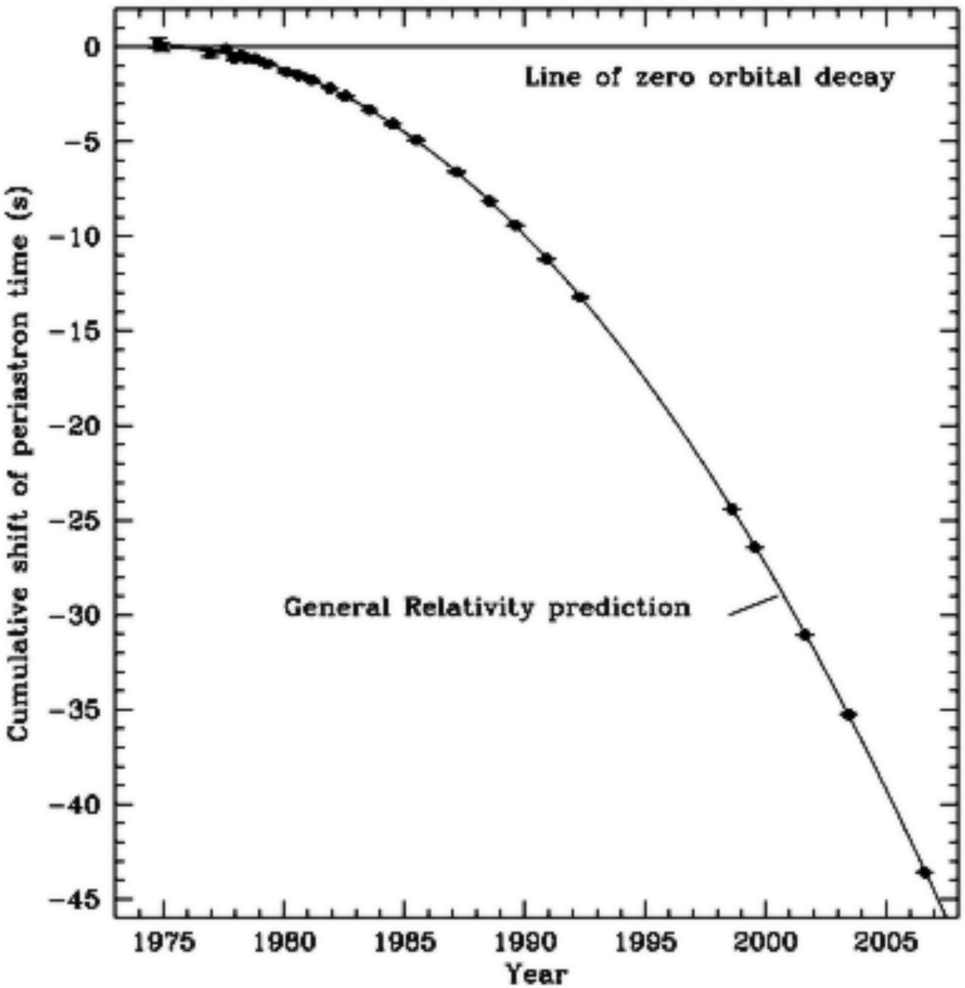


The discovery

After years of observation it became clear, that the pulsars spiral towards each other.

Conclusion: They loose energy (fast). Immediate question: How?!?





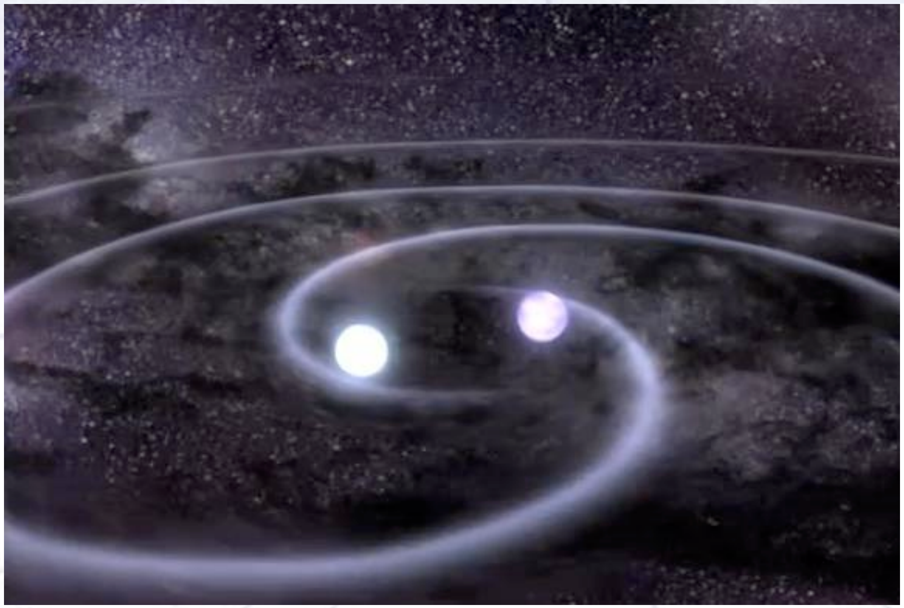
The answer is gravitational waves. What is normally a tiny effect is here so large (100x suns output!), that we (with a natural high precision clock) can see it.

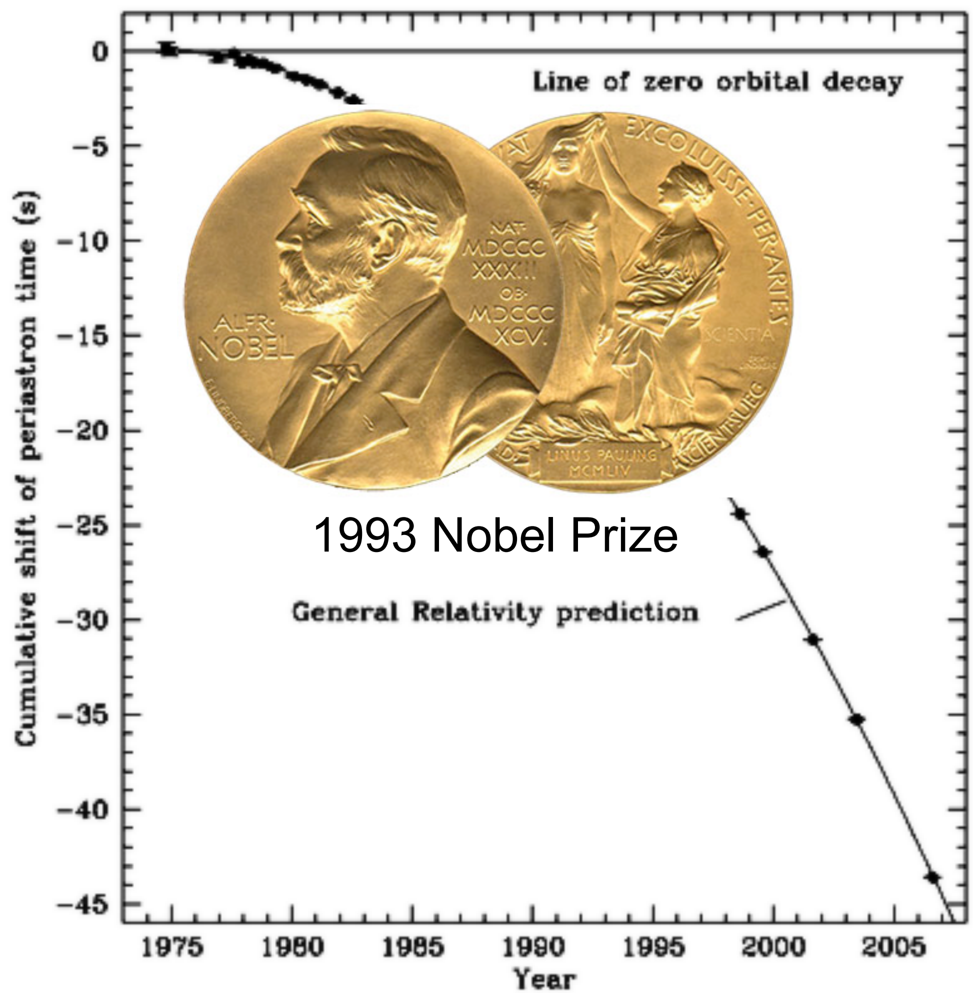
(Sun+Earth: 200 W, 10^{-15} m radius "lost" pr. day)

The discovery

After years of observation it became clear, that the pulsars spiral towards each other.

Conclusion: They loose energy (fast). Immediate question: How?!?





The answer is gravitational waves. What is normally a tiny effect is here so large (100x suns output!), that we (with a natural high precision clock) can see it.

(Sun+Earth: 200 W, 10^{-15} m radius "lost" pr. day)

The good experimenter

The good experimenter

<u>The good experimenter will always:</u>

- inspect data visually.
- test assumptions.
- keep an accurate record.
- perform cross checks.
- do a ChiSquare test (also).
- plan the experiments carefully.
- try to "blind" results until final.

<u>The good experimenter will never:</u>

- rely on untested assumptions.
- "just let someones program do it".
- make changes in data.
- look for only some effects.
- not look at the raw data.