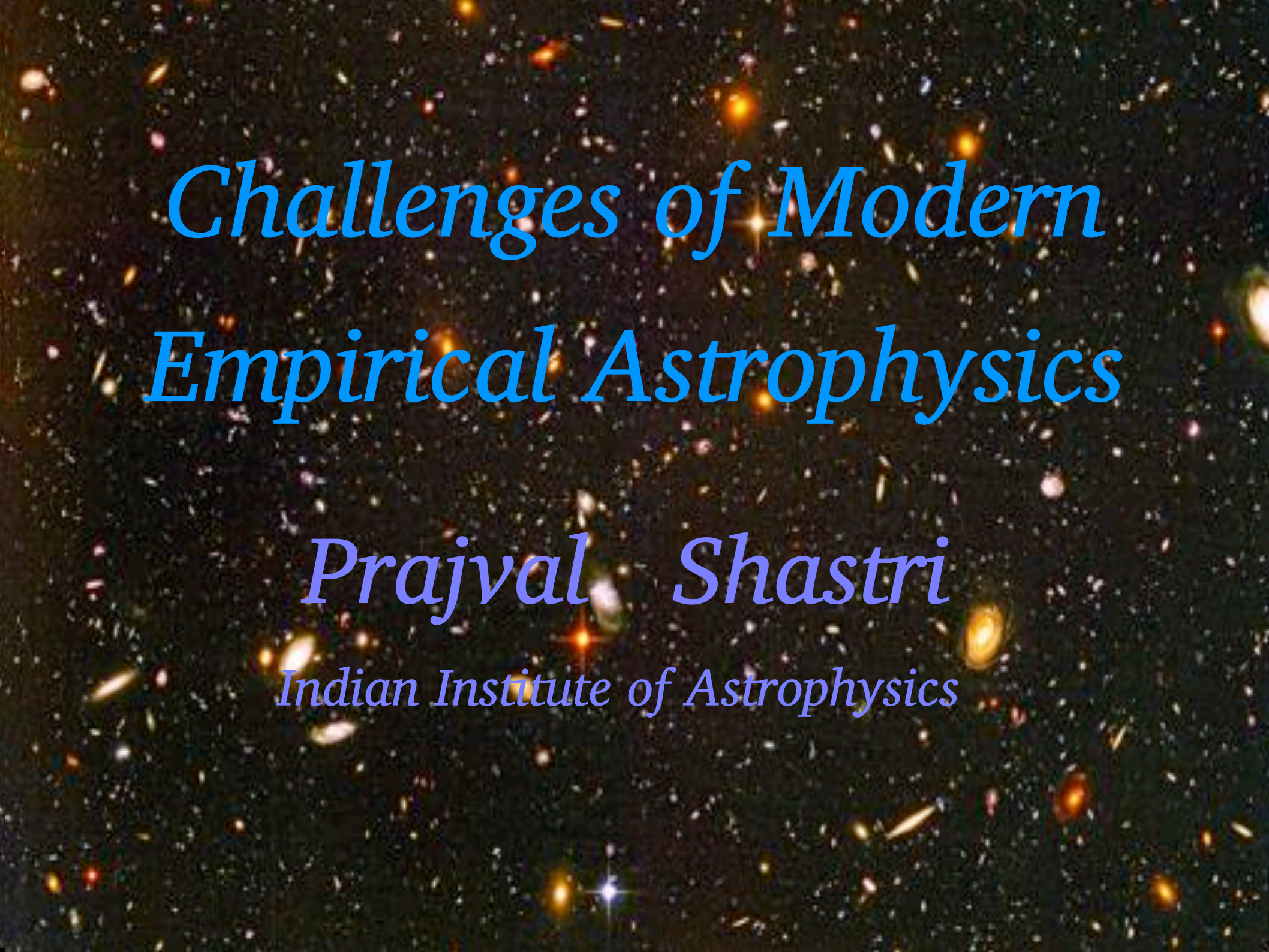

Challenges of Modern Empirical Astrophysics
Prajval Shastri
Indian Institute of Astrophysics

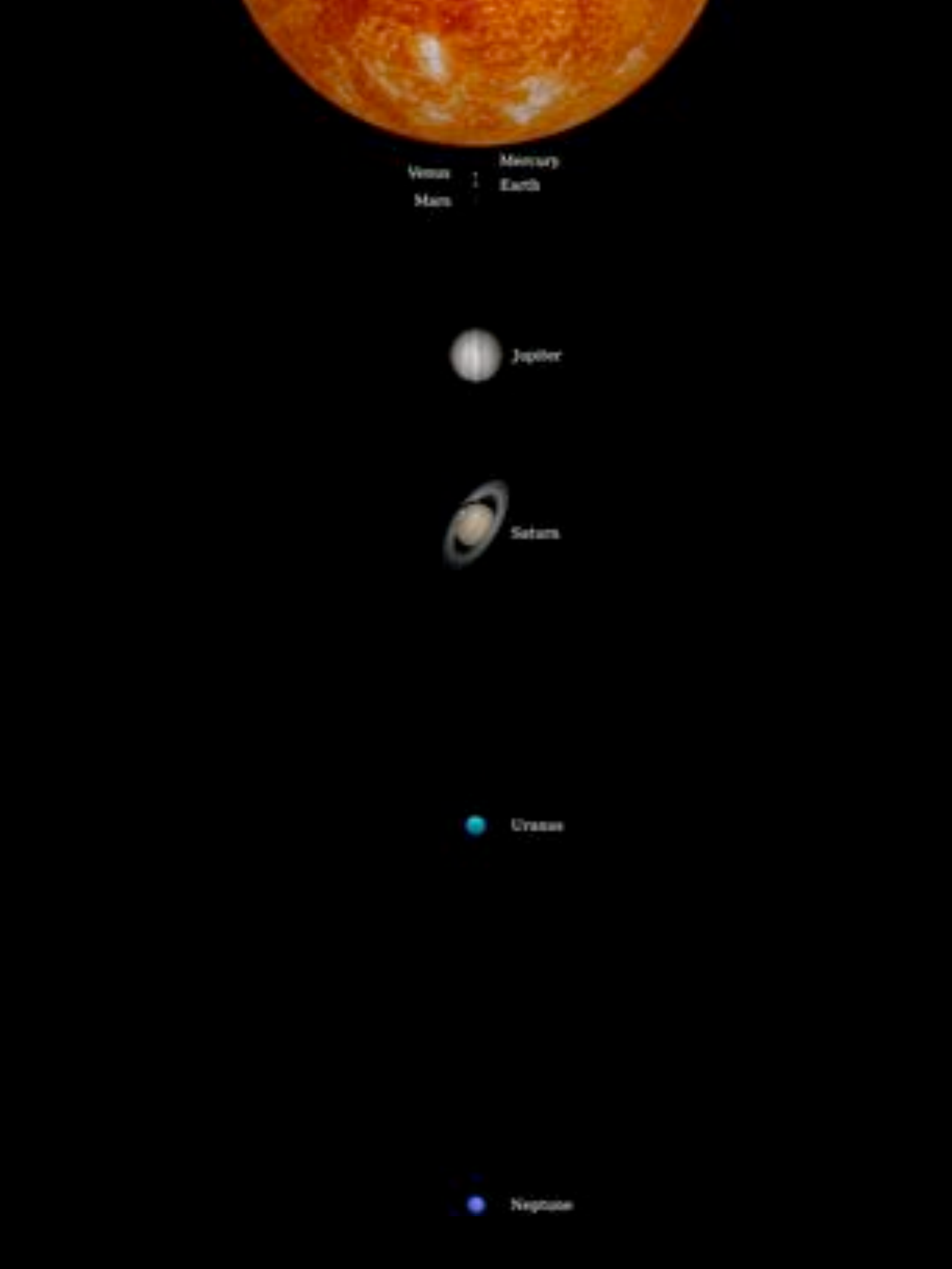
Astrophysics *is the study of the universe...*

and therefore also the study of ourselves, because we are evolutes of the physical processes that occur on cosmological scales..

Astrophysics *is the*
"last bastion of the generalist"

- Ter Haar

ಚಂದ್ರ


Mercury
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune


Wade Clark

Spiral Galaxy NGC 4414

Spiral Galaxy NGC 4622

Hubble Heritage

NASA and The Hubble Heritage Team (STScI/AURA) • Hubble Space Telescope WFPC2 • STScI-PRC02-03

Starburst Galaxy NGC 3310

Hubble Heritage

NASA and The Hubble Heritage Team (STScI/AURA)
Hubble Space Telescope WFPC2 • STScI-PRC01-26

Spiral Galaxy NGC 4414

Hubble Heritage

PRC99-25 • Hubble Space Telescope WFPC2 • Hubble Heritage Team(AURA/STScI/NASA)

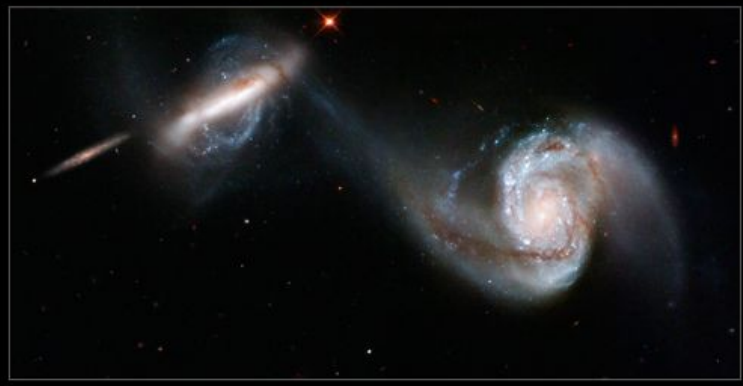
Sombrero Galaxy • M104

Hubble Heritage

Hickson Compact Group 87

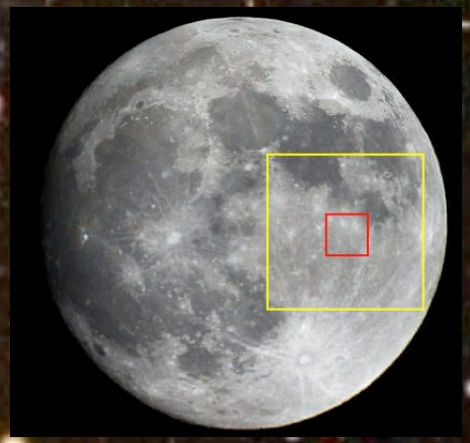
Hubble Heritage

Interacting Galaxies • Arp 87

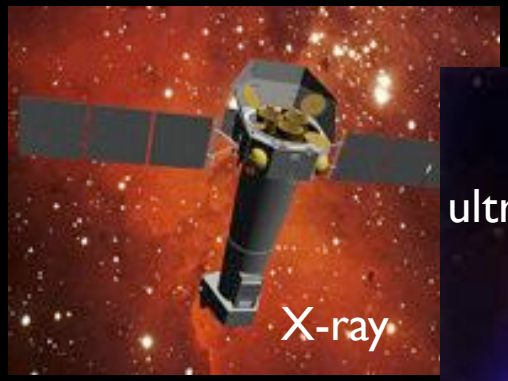


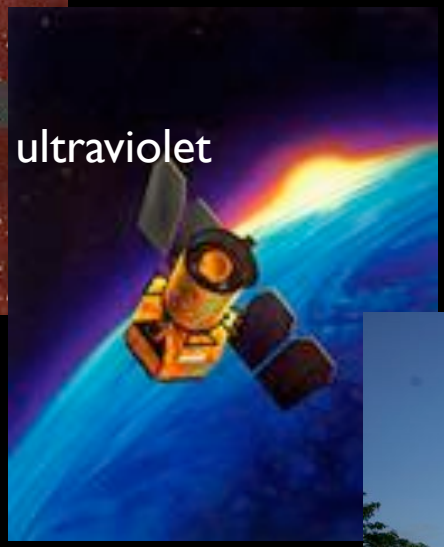
Hubble Heritage

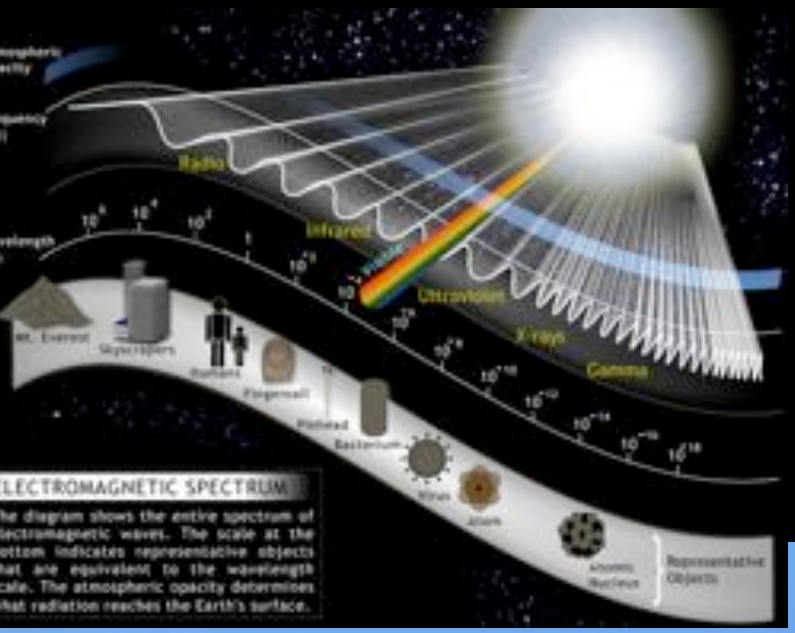
NASA, ESA, and the Hubble Heritage Team (STScI/AURA) • Hubble Space Telescope WFPC2 • STScI-PRC07-36




gamma-ray

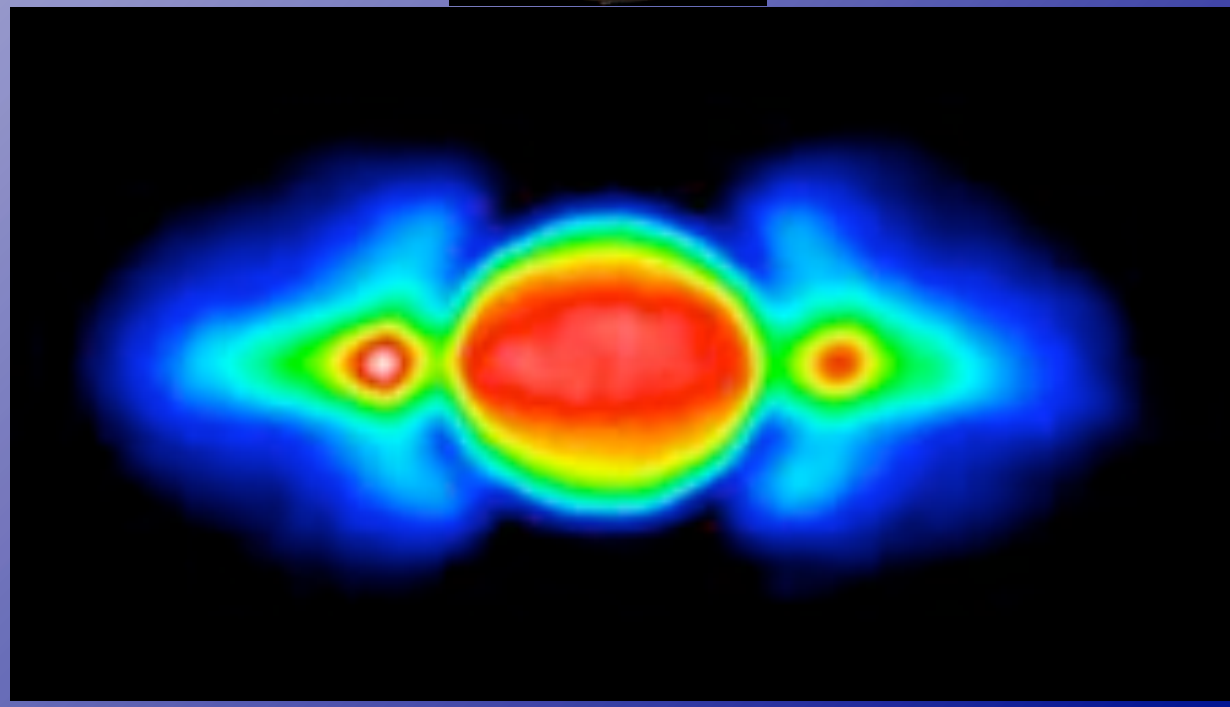

X-ray

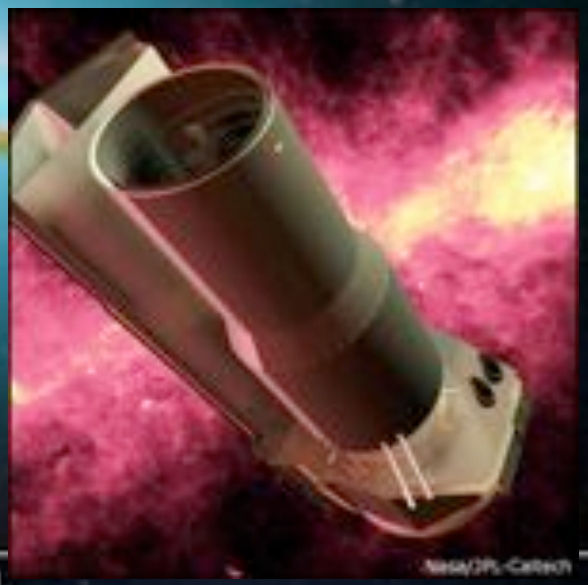

ultraviolet


Radio
Infrared
Ultraviolet
X-rays
Gamma
ELECTROMAGNETIC SPECTRUM
Representative Objects


Infrared


d Grafton
Saturn
CM1 250
CM2 336
CM3 053
C14 @ f/27 taken with a ST5c CCD from Houston Texas on December 11th 2002 at 5:40 UT





Visible

Infrared

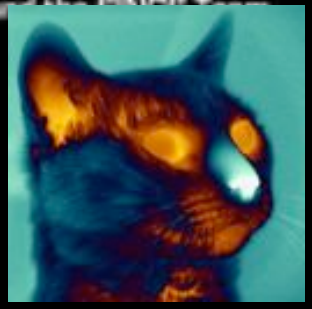
Sombrero Galaxy/Messier 104

Spitzer Space Telescope • IRAC

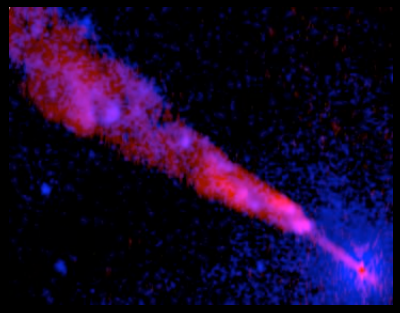
Visible: Hubble Space Telescope/Hubble Heritage Team

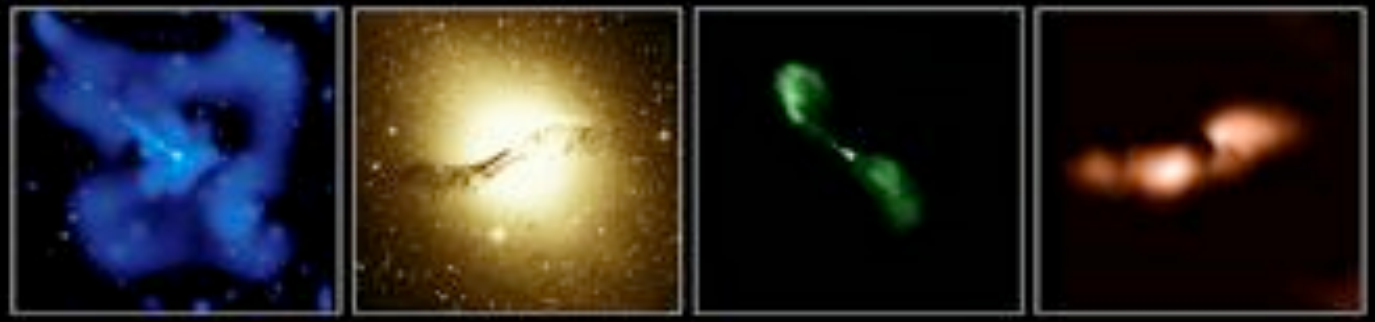
NASA / JPL-Caltech / R. Kennicutt (University of Arizona), and the SINGS Team

ssc2005-11a

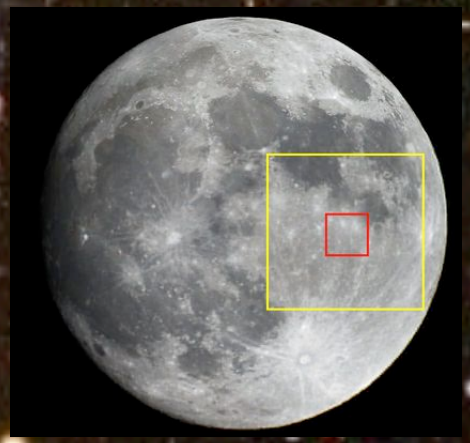


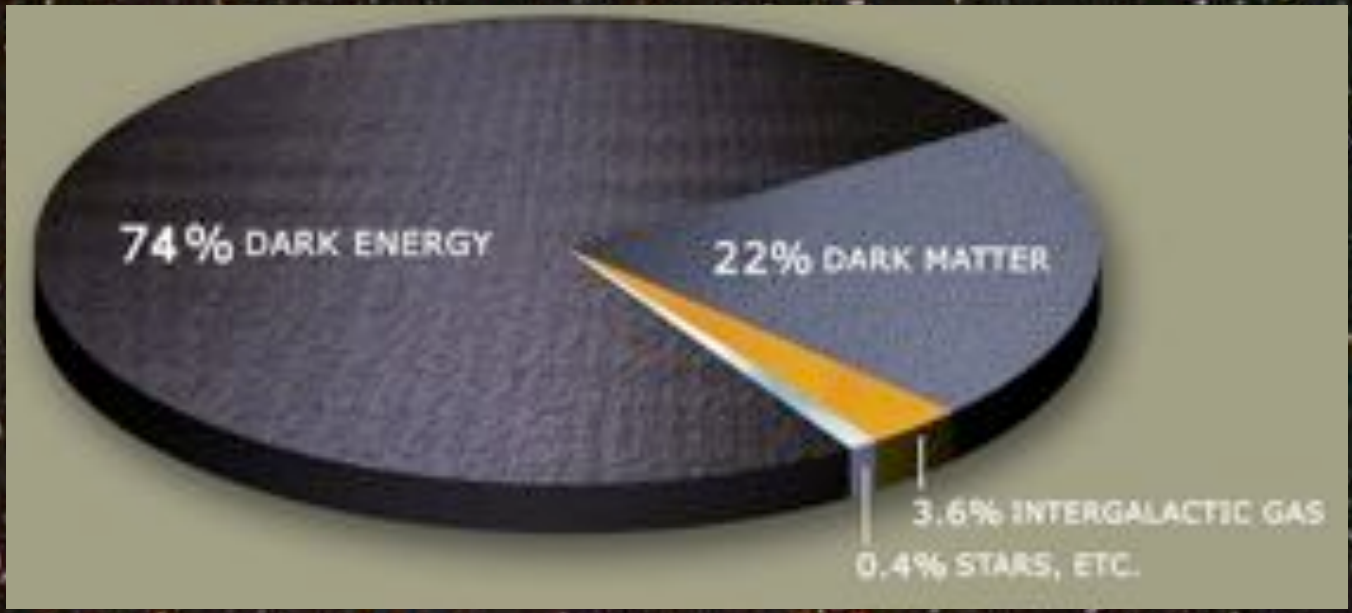
Centaurus A



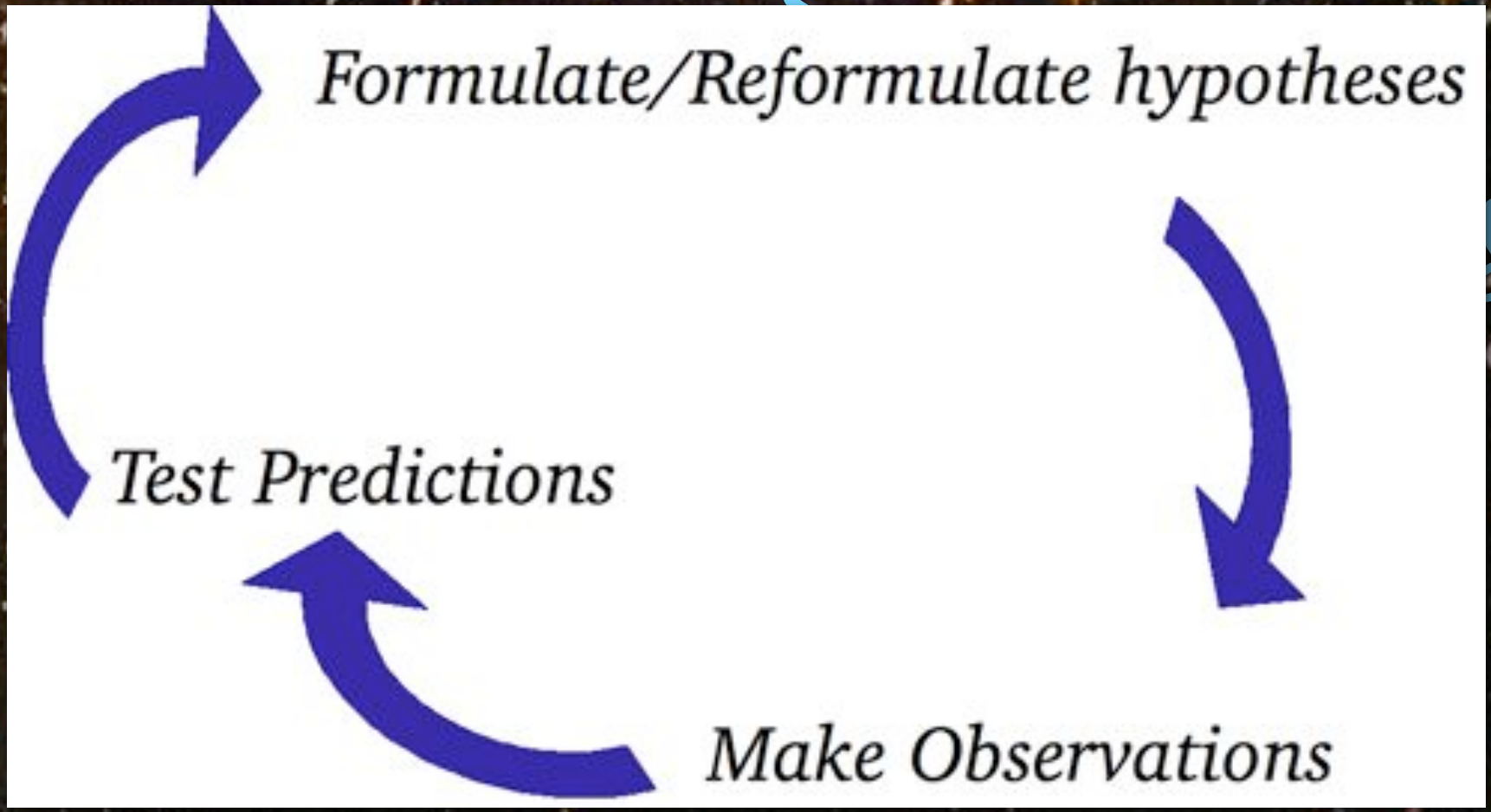


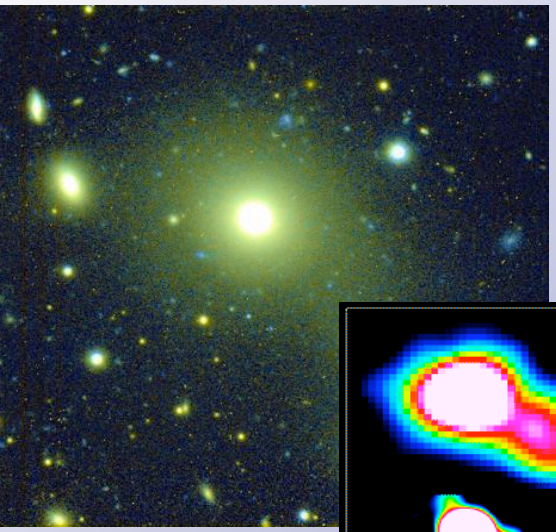
CHANDRA X-RAY | DSS OPTICAL | NRAO RADIO CONTINUUM | NRAO RADIO (21-CM)

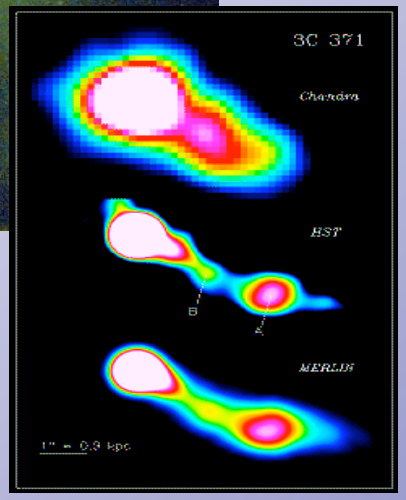


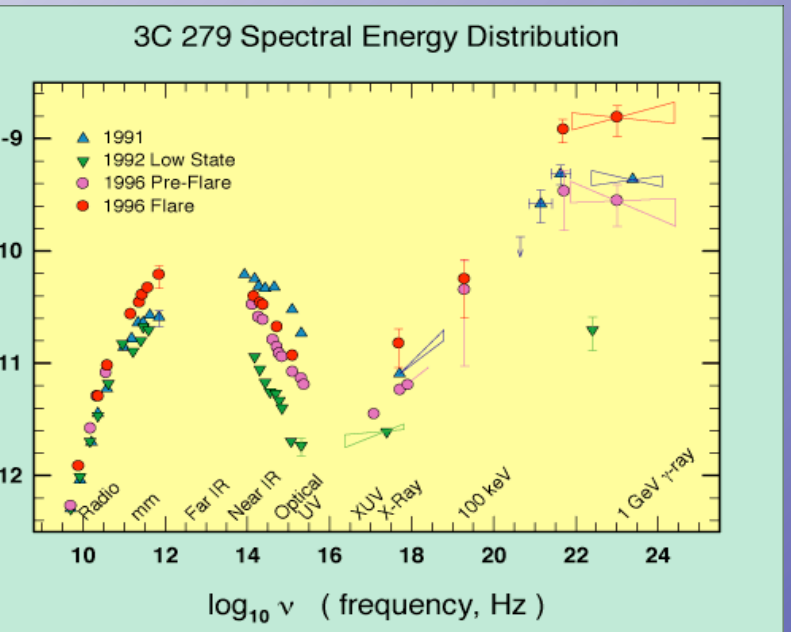

74% DARK ENERGY
22% DARK MATTER
3.6% INTERGALACTIC GAS
0.4% STARS, ETC.

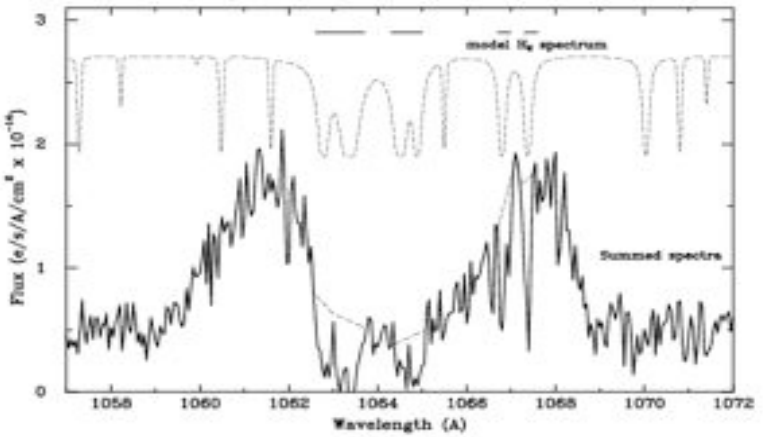
Patterns

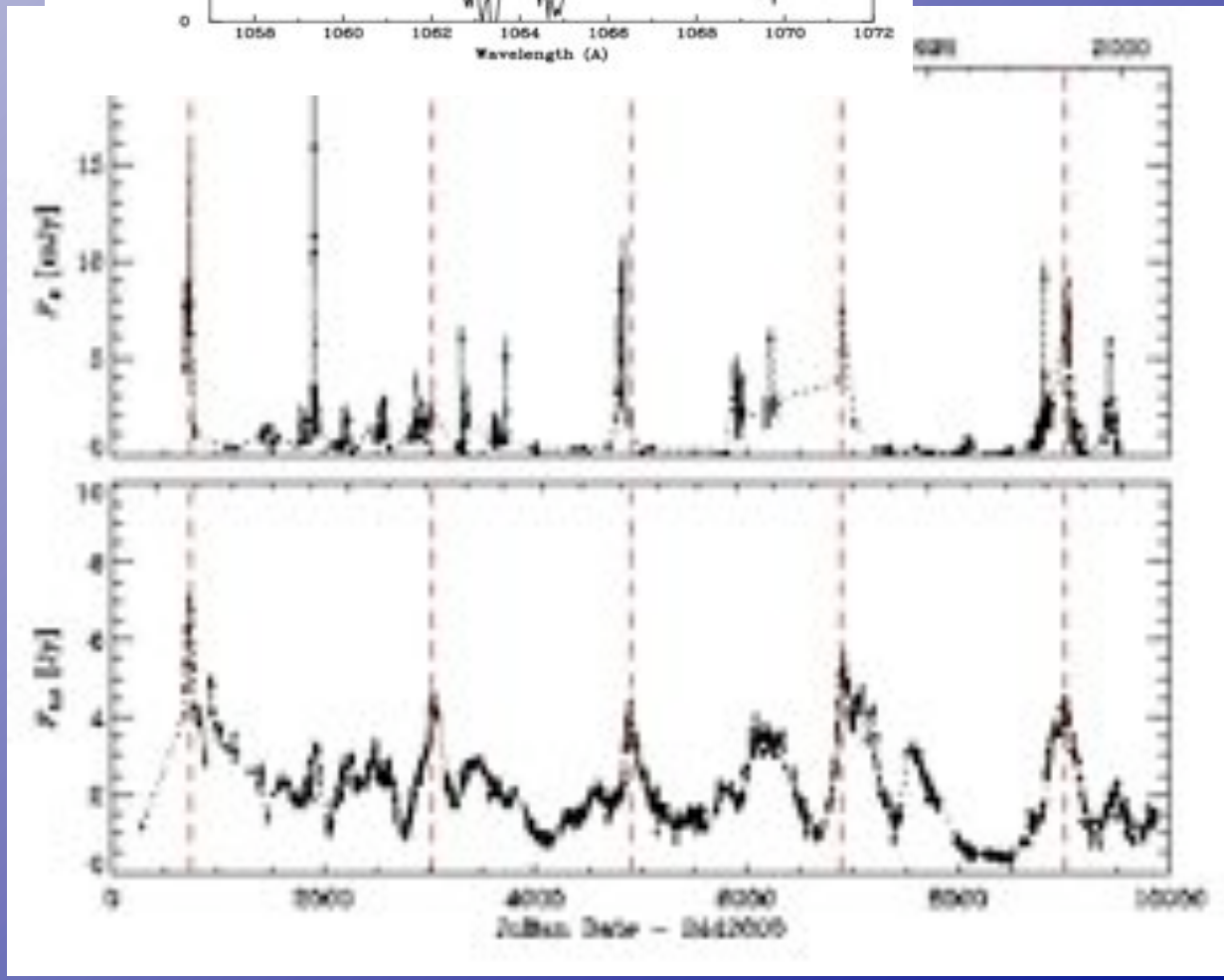

Formulate/Reformulate hypotheses
Test Predictions
Make Observations

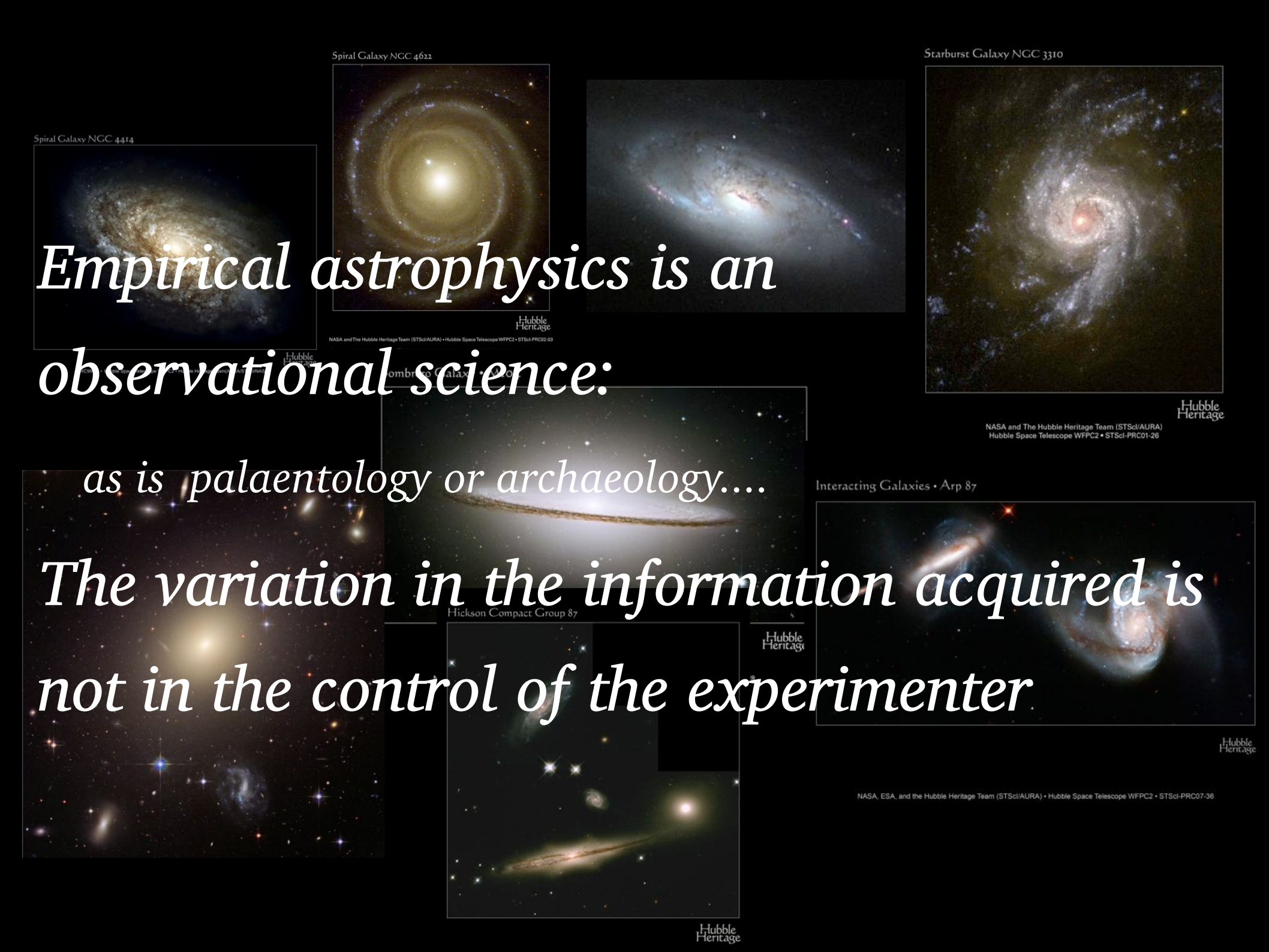



3C 371
Chandra
HST
MERLIN
1" = 0.9 kpc


3C 279 Spectral Energy Distribution
1991
1992 Low State
1996 Pre-Flare
1996 Flare
Radio
mm
Far IR
Near IR
Optical
UV
XUV
X-Ray
100 keV
1 GeV γ-ray
10
12
14
16
18
20
22
24
log10 ν (frequency, Hz)


model H2 spectrum
Summed spectra
Wavelength (A)





Empirical astrophysics is an observational science:

as is palaentology or archaeology....

The variation in the information acquired is not in the control of the experimenter

- *Rigour in data reductions & analysis*
- *Multiwavelength measurements (increase in the dimensions)*
- *Increase in amount of data makes automation inevitable*

The need to intelligently use state-of-the-art statistics based on a reasonable conceptual understanding cannot be overstated

Rigour in data reductions & analysis

- *When is a blip in a spectrum, image or data stream a real signal?*
- *Are these stars/galaxies/sources an unbiased sample of the vast underlying population?*
- *When should these objects be divided into 2/3/… classes?*
- *What is the intrinsic relationship between two properties of a class (especially with confounding variables)?*
- *Can we answer such questions when our data have measurement errors & flux limits?*
- *How is the very common variability in stars/galactic nuclei etc. to be modelled?*

Rigour in data reductions & analysis

- *When is a blip in a spectrum, image or data stream a real signal?* ***Statistical Inference***
- *Are these stars/galaxies/sources an unbiased sample of the vast underlying population?* ***Sampling***
- *When should these objects be divided into 2/3/… classes?* ***Multivariate Classification***
- *What is the intrinsic relationship between two properties of a class (especially with confounding variables)?* ***Multivariate Regression, Principal Component Analysis***
- *How is the variability in stars or galactic nuclei to be modelled?* ***Time Series Analysis***
- *Can we answer such questions when our data have measurement errors & flux limits?* ***Censoring, Truncation & Measurement Errors***

- ***Maximum Entropy Method in imaging***

  Gull & Skilling 1984

seeks to extract as much information from a measurement as is justified by the data's signal-to-noise ratio

- ***Two-point correlation function for galaxies***

  Bhavsar 1990

  The data points are pairs of galaxies (ie galaxy co-ordiates in the sky), and to take into account the fact that the /N error bars are not independent, the bootstrap methodology is applied

- ***Concept similar to Mahalanobis Distance in object detection***
  Babu, Mahabal, Djorgovski, Williams 2008

gives a very robust object detection technique that is capable of detecting faint sources especially in multi-epoch frames, i.e., even those objects that are not visible at all epochs (which would normally be smoothed out by traditional methods)

- ***Oscillation Analysis of Solar Corona***
  Gissot & Hochedez 2008

ability of a motion estimation algorithm to explore and analyse the oscillating motions of coronal loops present in extreme Ultraviolet image sequences, using Morlet wave analysis.

▸ ***Nonparametric Inference for the Cosmic Microwave Background***

Genovese, Miller, Nichol, Arjunwadkar & Wasserman 2004

- construction of non-parametric confidence set for the unknown Cosmic Microwave Background Spectrum, to give an estimated spectrum based on minimal assumptions, leading to a wide range of additional inferences in addition to those similar to the cosmologists' model-based estimates.

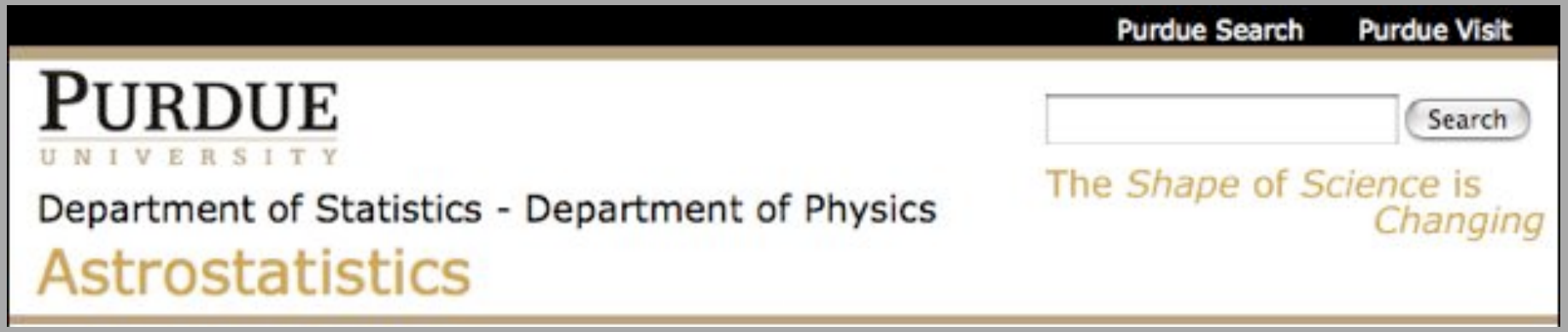
▸ ***Image reconstruction with error estimates***

van Dyk, Connors, Esch, et al 2007

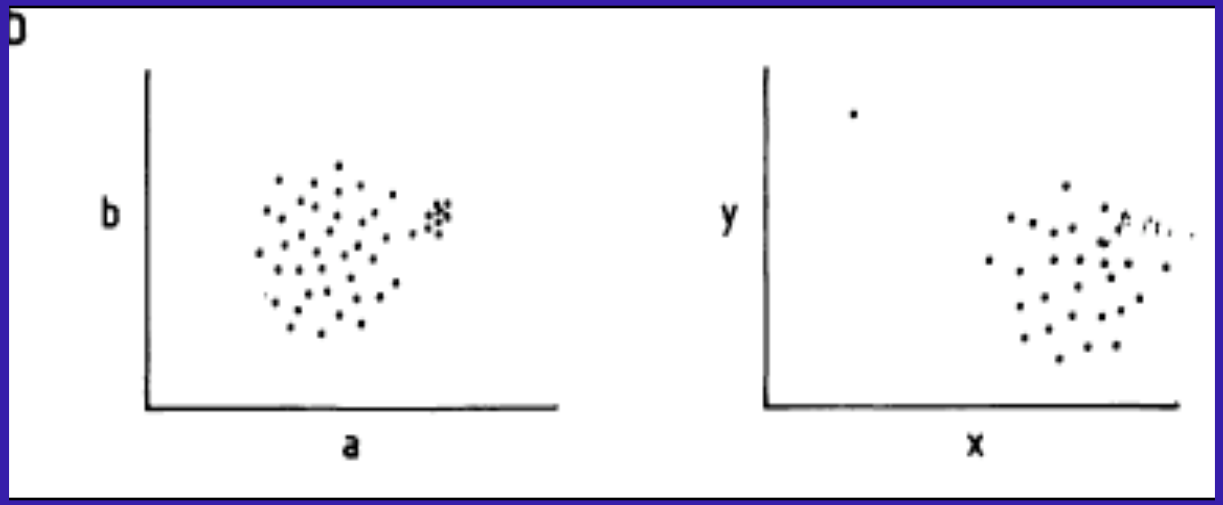
explicitly model the complexities of both astronomical sources and the data generation mechanisms inherent in new high-tech instruments, i.e., non-uniform stochastic censoring, heteroscedastic errors in measurement, and background contamination.

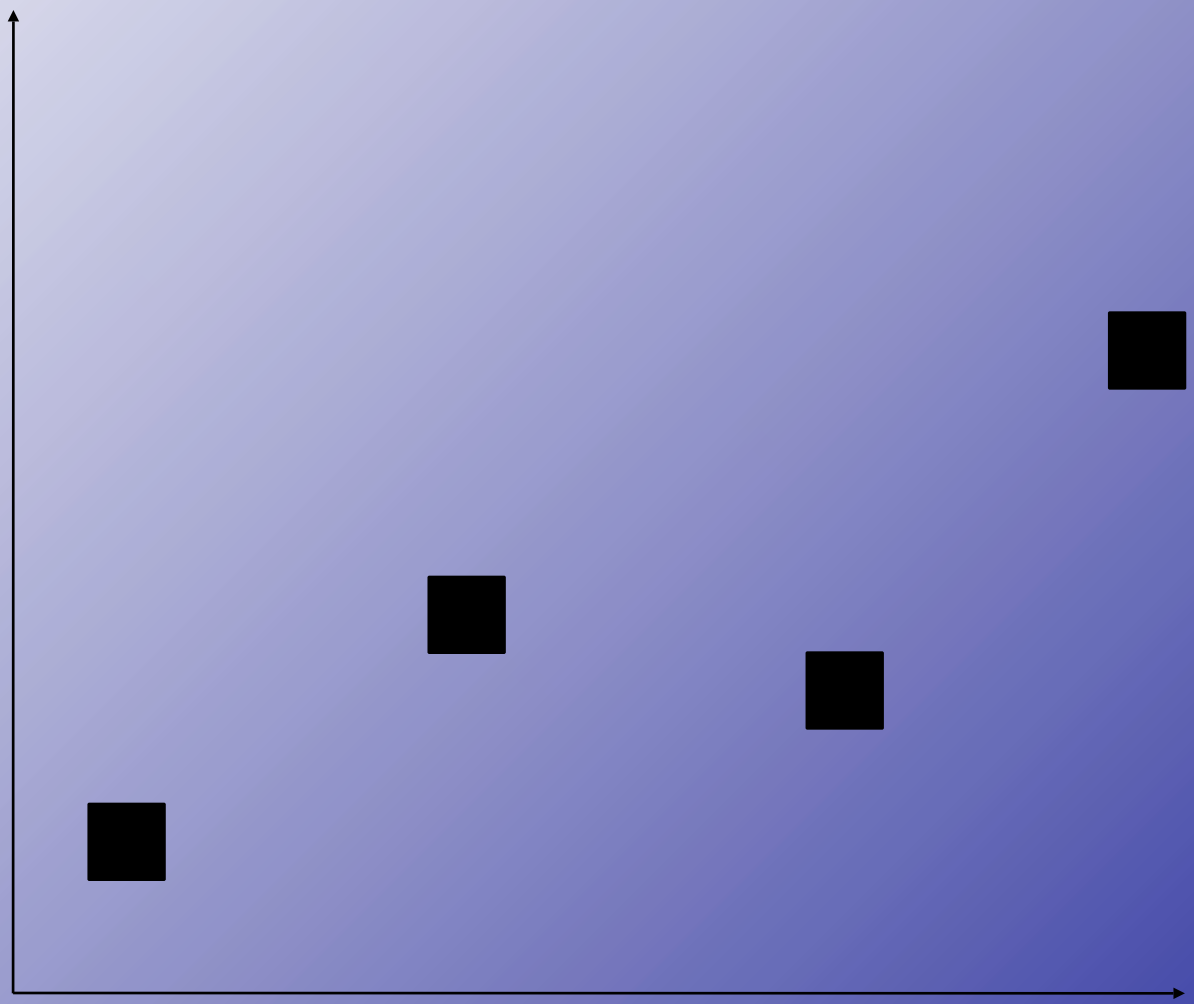
CHASC Astro-Statistics Collaboration

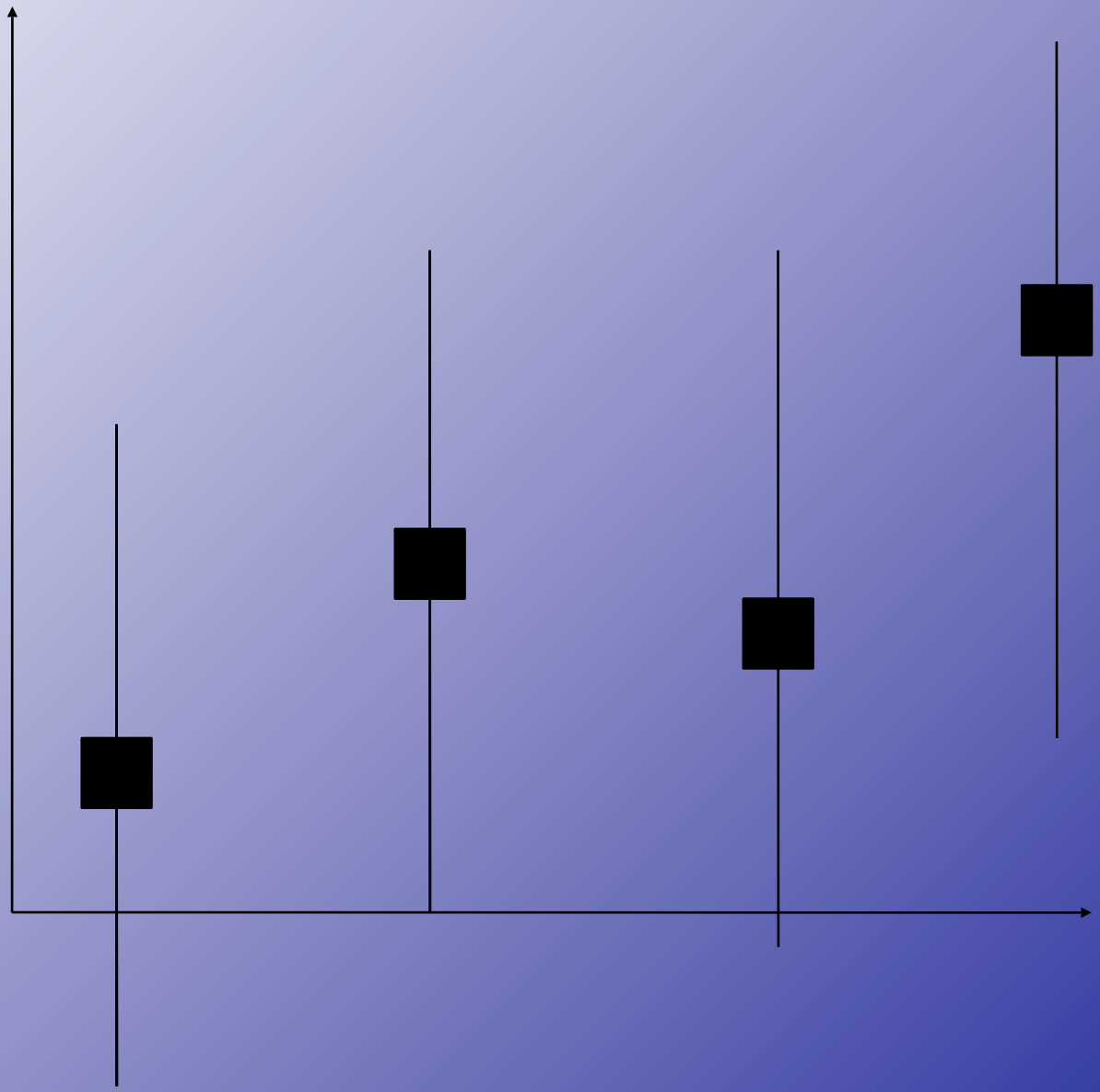
(California/Harvard/ASC AstroStatistics Collaboration)

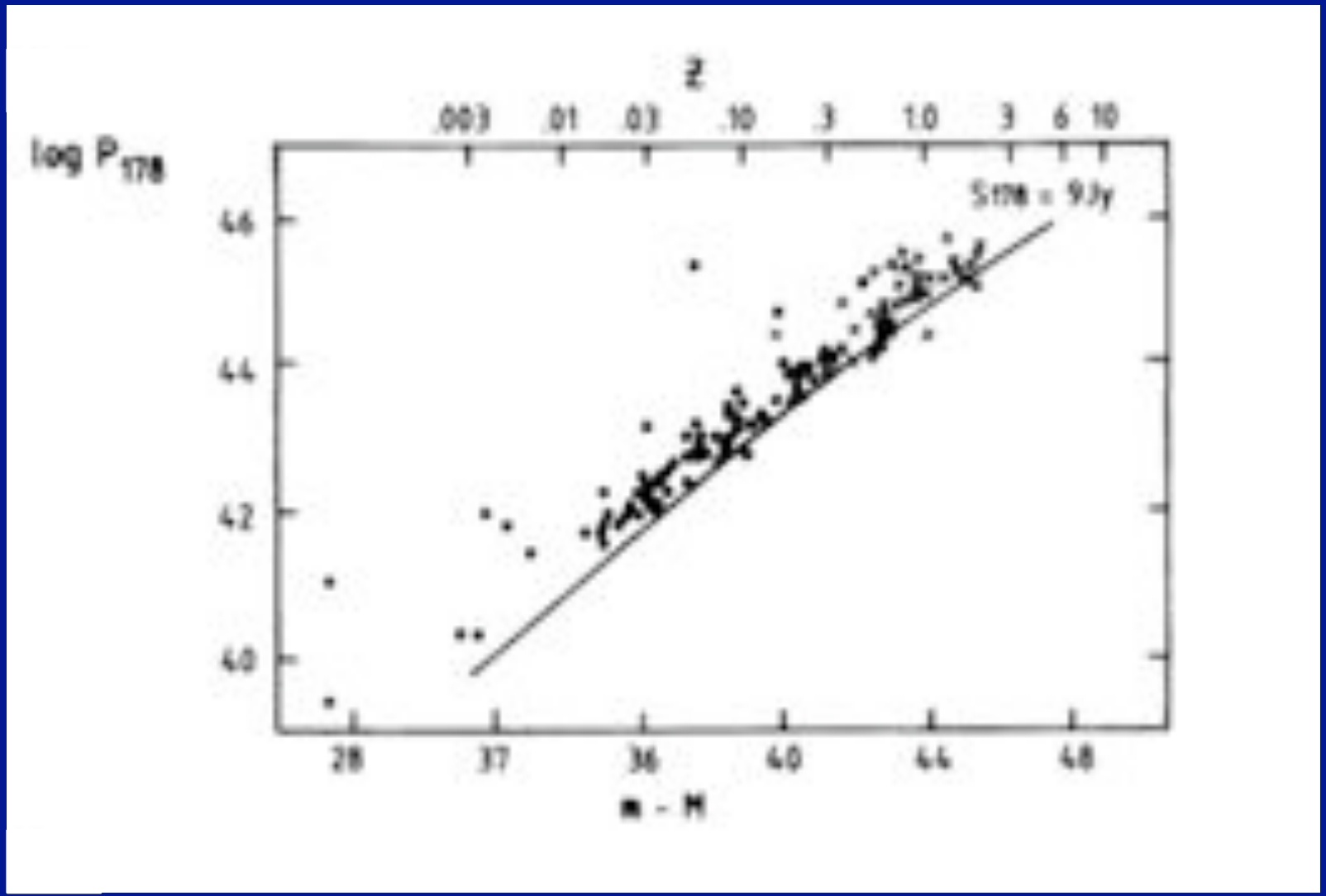


Carnegie-Mellon and University of Pittsburgh

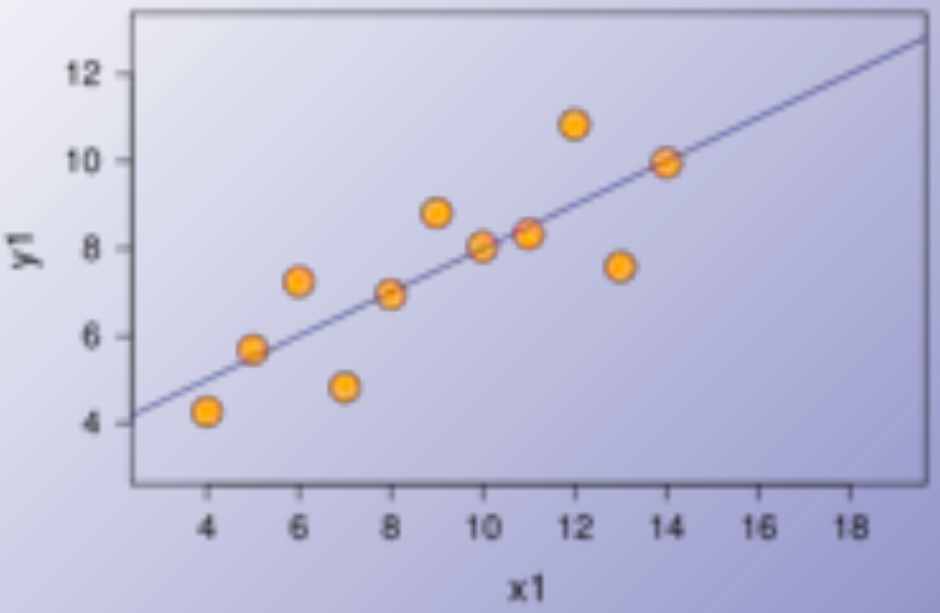

b
a
y
x

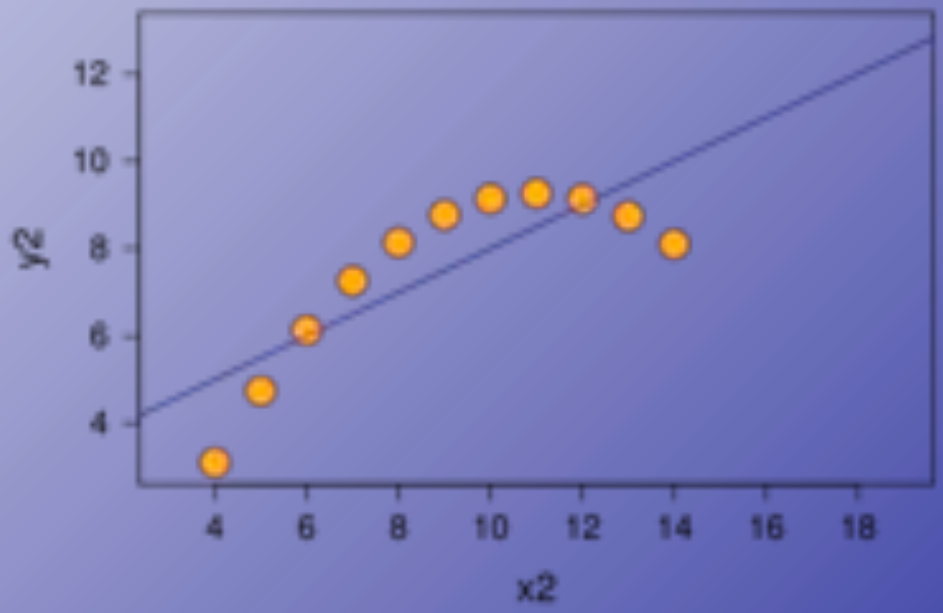


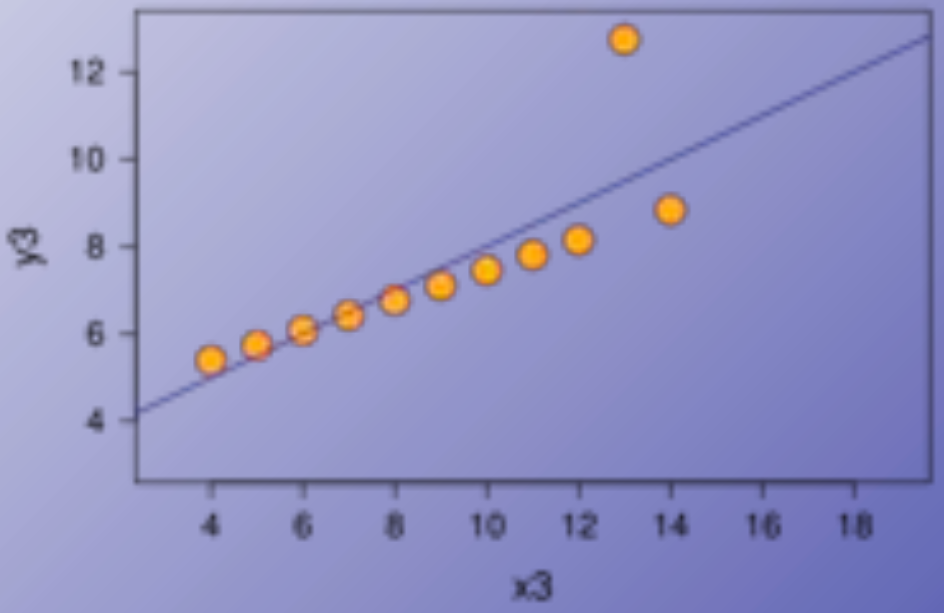


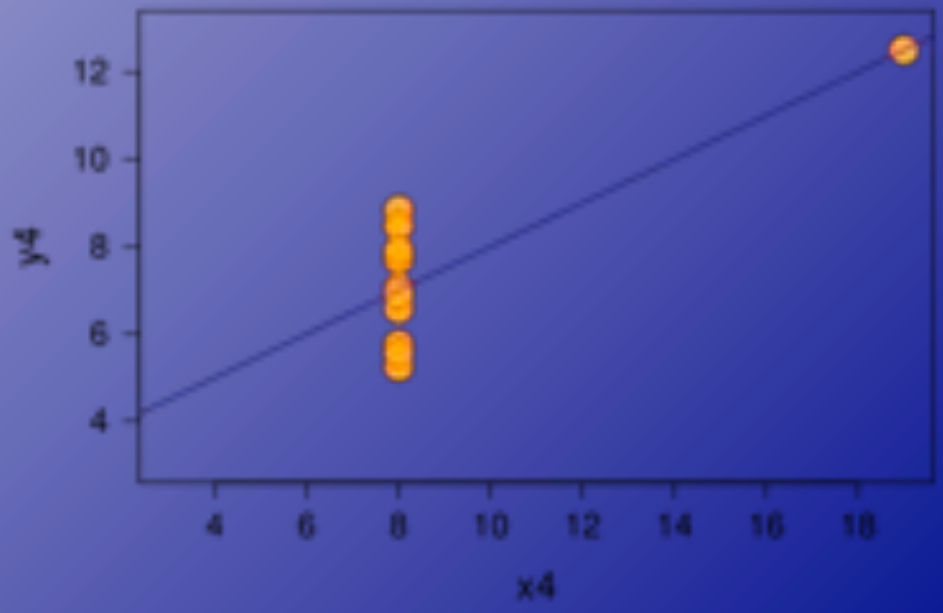

log P_{178}
Z
.003
.01
.03
.10
.3
1.0
3
6
10
S178 = 9.3y
46
44
42
40
28
32
36
40
44
48
m - M

Anscombe's Quartet









Were taught that:

"we must use non-parametric tests"

But we tend to work in "recipe book style"

The Prevailing doctrine:

"Does the eye see much correlation? If not, calculation of a formal correlation statistic is probably a waste of time."

Thus, in general, the awareness and exposure is poor, and the ignorance is profound…and worse….quite unabashed!

unabashed astrophysicist: *Is the difference between these magnitudes significant?*

eminent statistician: *Don't ask me, go look at the data!*

Smithsonian/NASA ADS Astronomy Abstract Service

· Find Similar Abstracts (with default settings below)
· Full Refereed Journal Article (PDF/Postscript)
· Full Refereed Scanned Article (GIF)
· **Citations to the Article (6)** (Citation History)
· Refereed Citations to the Article

Title: **Practical statistics for astronomers. I -**
Definitions, the normal distribution, detection of signal
Authors: *Wall, J. V.*
Affiliation:AA(Mullard Radio Astronomy Observatory, Cambridge, England)
Publication: Royal Astronomical Society, Quarterly Journal, vol. 20, June 1979, p. 138-152.
Publication Date: 06/1979
Category: Astronomy
NASA/STI Keywords: ASTRONOMY, NORMAL DENSITY FUNCTIONS, PROBABILITY DENSITY FUNCTIONS, SIGNAL DETECTION, DEFINITION, HISTOGRAMS, STATISTICAL ANALYSIS
Bibliographic Code: 1979QJRAS..20..138W

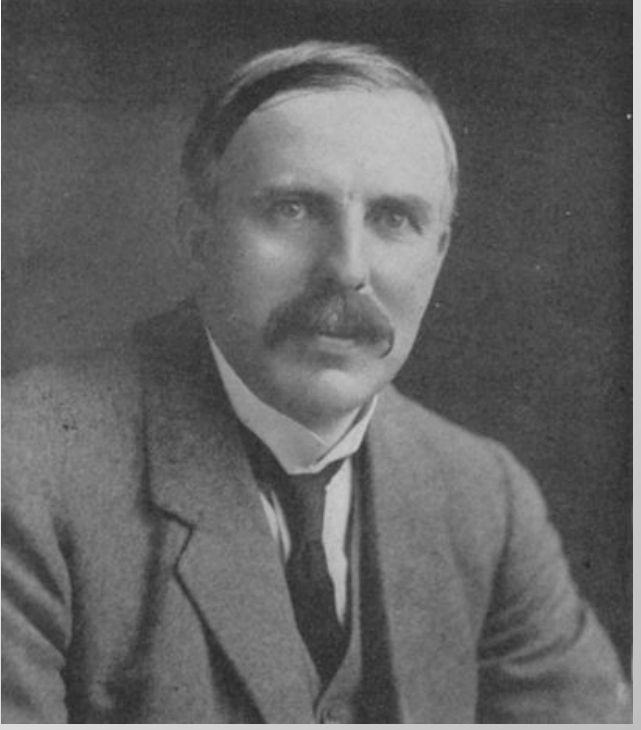
Smithsonian/NASA ADS Astronomy Abstract Service

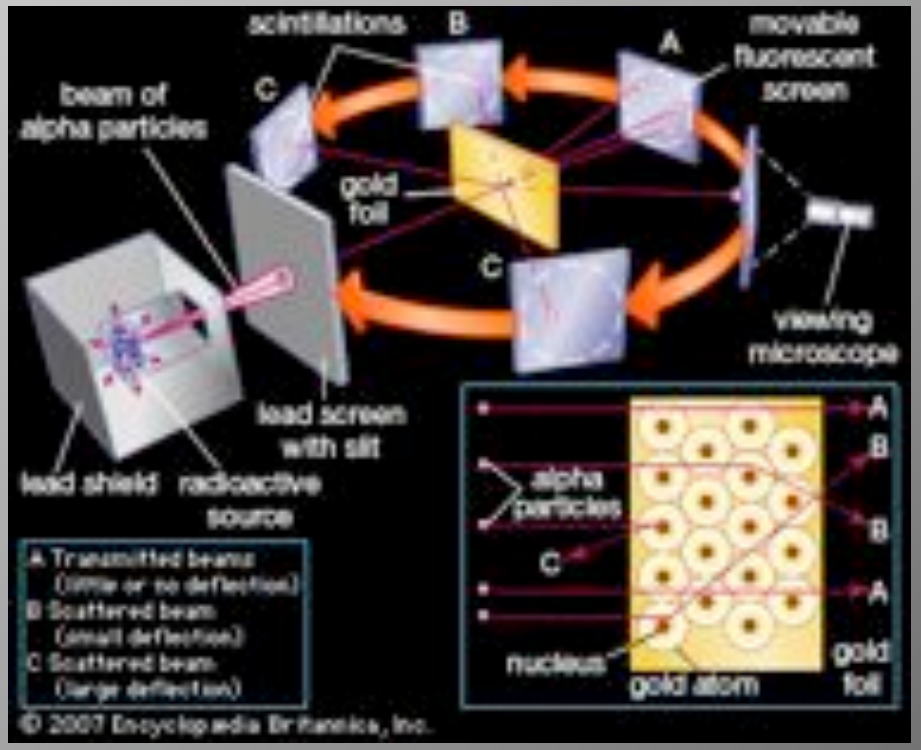
· Find Similar Abstracts (with default settings below)
· Full Refereed Journal Article (PDF/Postscript)
· Full Refereed Scanned Article (GIF)
· References in the article
· **Citations to the Article** (9) (Citation History)
· Refereed Citations to the Article

Title:Practical Statistics for Astronomers - II. Correlation, Data-modelling and Sample Comparison
Authors: Wall, J. V.
Publication:Quarterly Journal of the Royal Astronomical Society, Vol. 37, p.519
Publication Date:12/1996
Origin: QJRAS
Bibliographic Code:
1996QJRAS..37..519W

Institutional barriers

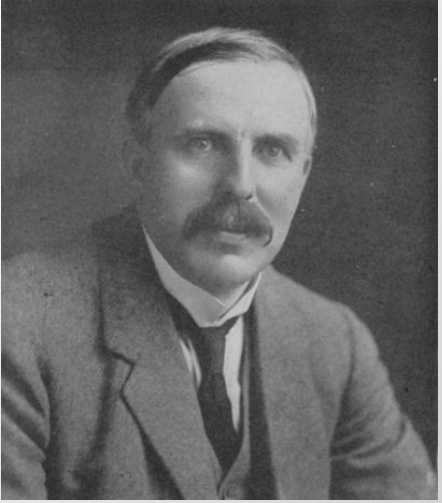
- *Statistics is not part of an astrophysicist's formal training*
- *Astrophysicists tend to be housed in research institutes rather than in universities*
- *Astrophysicists come with their "physicist" baggage:*





"If your experiment needs statistics, you ought to have done a better experiment."

- *Ernest Rutherford*



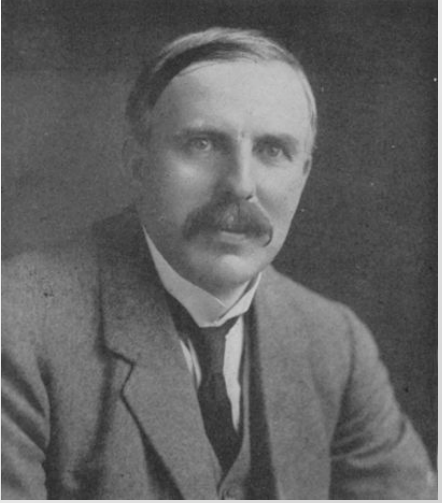
"If your experiment needs statistics, you ought to have done a better experiment."

- Ernest Rutherford

*Acts are not only of **omission**: not using state-of-the-art statistical methodology, but*

Even traditional methods are often misused:

- *Unweighted bivariate least-squares fits are used interchangeably in Hubble constant studies with wrong confidence intervals*
  Feigelson & Babu ApJ 1992
- *Likelihood-ratio test (F test) usage typically inconsistent with asymptotic statistical theory*
  Protassov et al. ApJ 2002
- *Kolmogorov-Smirnov goodness-of-fit probabilities are inapplicable when the model is derived from the data*
  Babu & Feigelson ADASS 2006



"If your experiment needs statistics, you ought to have done a better experiment."

- Ernest Rutherford

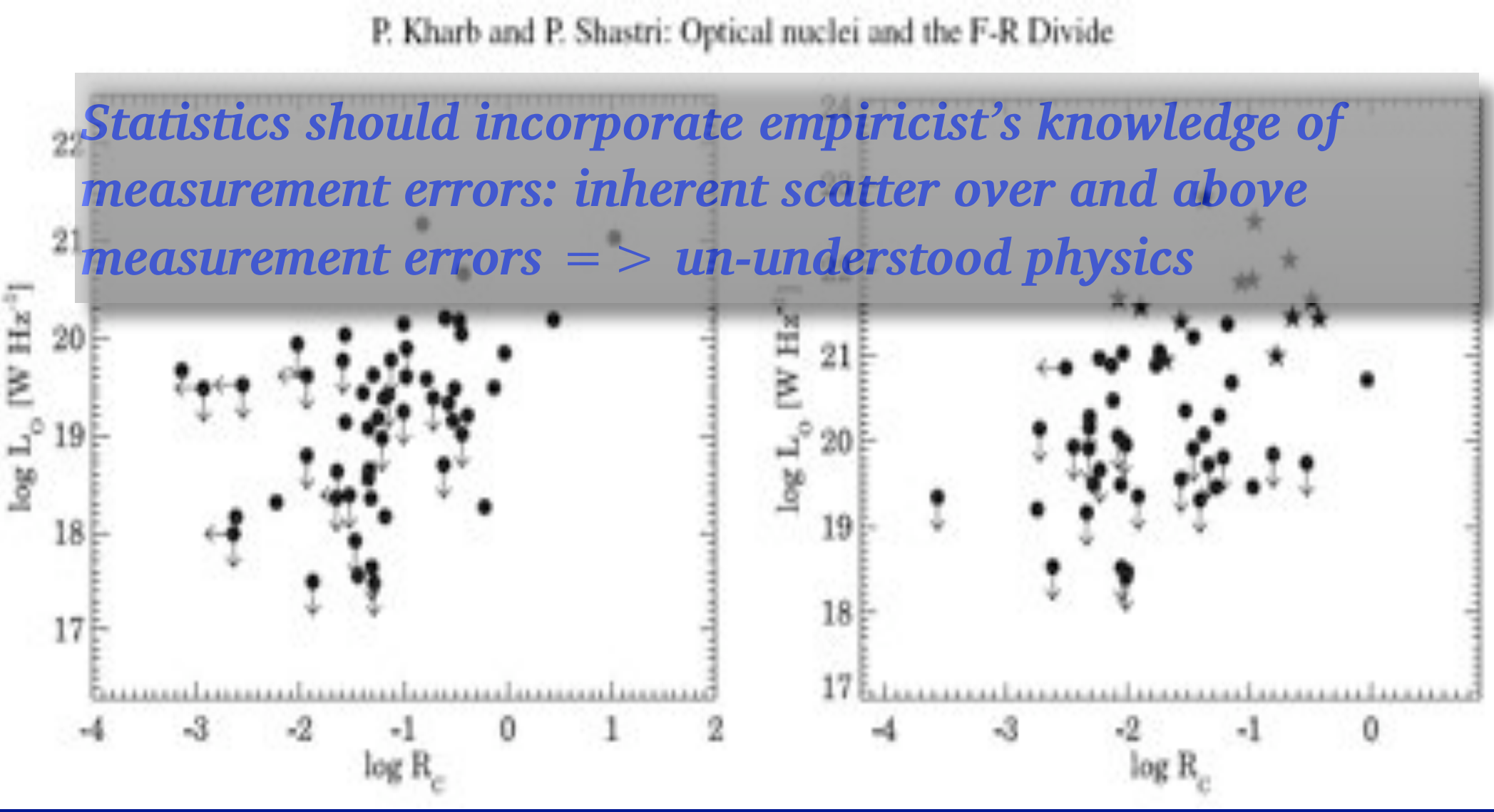
Empirical astrophysics differs from empirical physics:

- *it is an observational science, i.e., The variation in the information acquired is not in the control of the experimenter*
- *data are constantly gathered at the limit of the instrument capability*

In other observational disciplines, experimental design and inference, and hypothesis testing develop together, but this has only very rarely happened in astrophysics.

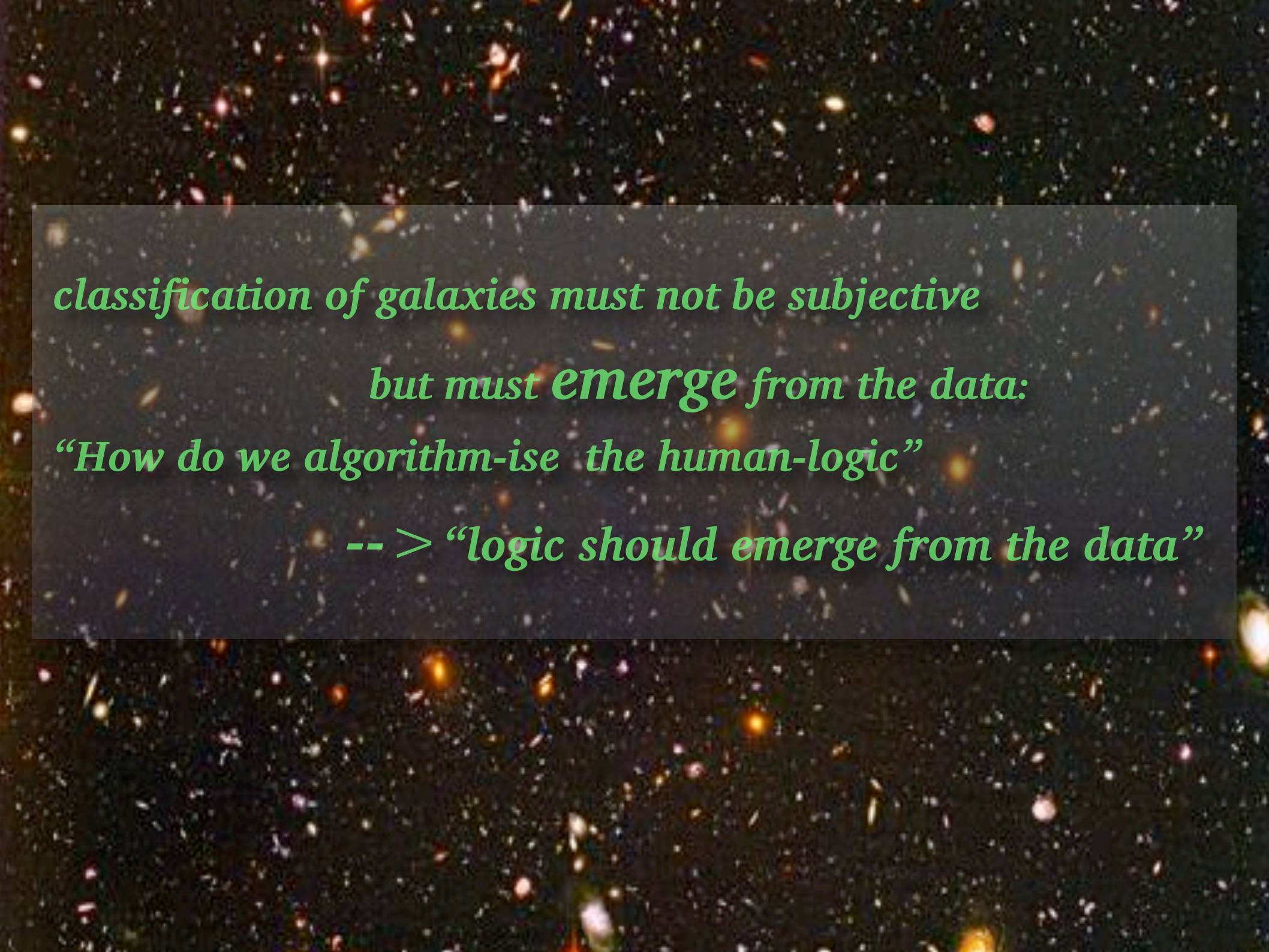
There is a GAP in the pedagogy!

Kharb & Shastri 2004



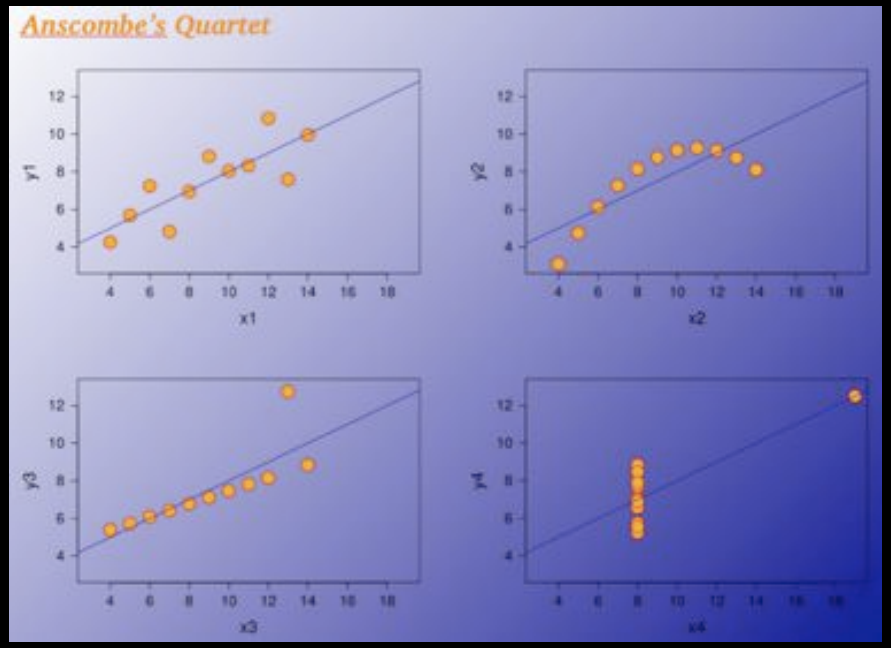
Statistics should incorporate empiricist's knowledge of measurement errors: inherent scatter over and above measurement errors => un-understood physics

The scope for collaboration between astrophysicsts and statisticians is immense


classification of galaxies must not be subjective
but must **emerge** *from the data:*
"How do we algorithm-ise the human-logic"
-- > "logic should emerge from the data"

Probability:

- *Coin flips, conditional probabilities*
- *density of a continuous random variable*
- *normal and chi-square distributions*
- *The Central Limit Theorem*



Exploratory Data Analysis:

- *uncover the underlying structure*
- *detect outliers and anomalies*
- *extract important variables*
- *formulate hypotheses for testing*

Uses the R software environment

Statistical Inference:

- *Going beyond the immediate data*
- *Is the observed difference between groups dependable or could it have happened by chance?*

Bayesian Inference:

- *Taking prior knowledge into account*

Likelihood Estimation:

- *Difference between likelihood and probability!*
- *probability -> occurrence of future events*
- *Likelihood -> past events with known outcomes*

Fitting mathematical models to the data

Tuning the free parameters to obtain a good fit

Non-parametric statistics:

- *which make no assumptions about the probability distributions of a population*
- *therefore applicability is wider*

Concepts of Regression:

- *Applications in the astronomy literature*

Model Selection:

Goodness of Fit: Bootstrap

Cluster Analysis: Grouping, data mining

Multivariate Analysis: of data with two or more dependent variables

Monte-Carlo Markov Chain techniques that use pseudo-random (simulated) values to estimate mathematical solutions

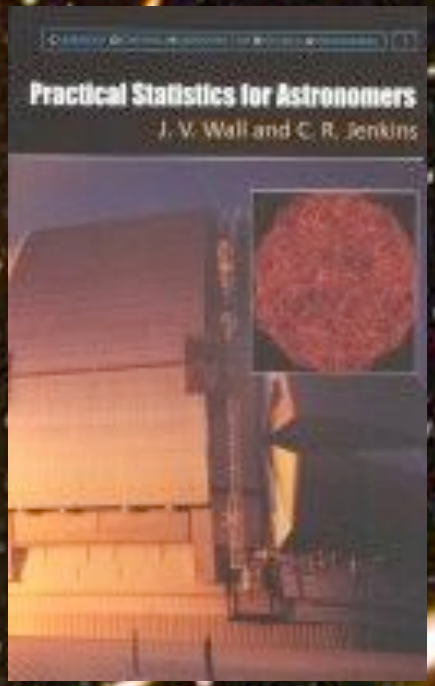
Time Series Analysis

Discussion of some statistical applications in the astronomy literature

Discussion sessions: opportunity to discuss your individual research problems with the statisticians

Dont do the tutorials blind!:

No laptops to be open during the lectures!!


Practical Statistics for Astronomers
J. V. Wall and C. R. Jenkins

Thank you!


ASTROSTATISTICS
Interdisciplinary Statistics
G. J. Babu and E. D. Feigelson