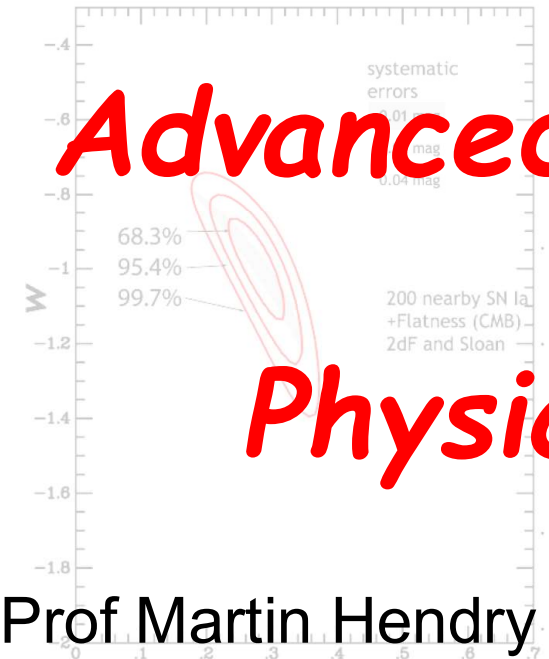
Scottish Universities Physics Alliance



Advanced Data Analysis for the Physical Sciences

Prof Martin Hendry

School of Physics and Astronomy
University of Glasgow

Martin.Hendry@glasgow.ac.uk

10 lecture course which aims to:

- Review the theoretical foundations and practical methods of modern data analysis in the physical sciences
- Summarise the basic tools of parameter estimation and testing goodness of fit
- Explore and contrast Bayesian and classical (frequentist) methods
- Highlight modern computational advances in analysis of (very large) datasets.

Graduate School: Investing in Education

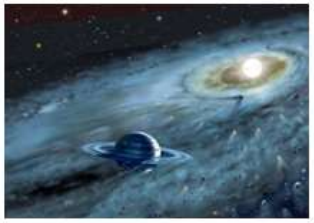
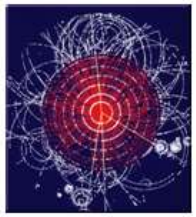
Scottish Universities Physics Alliance

The SUPA Graduate School is a cornerstone of the SUPA Alliance.

Shared Delivery of Graduate Education:

- New specialist courses are made viable by pooling expertise.
- Delivery by a variety of modes, lectures, distance learning, short courses.
- SUPA-funded Access Grid rooms in each Institution for two-way video delivery.
- Scotland-wide relay of research seminars and colloquia.





Courtesy: Horiba Jobin Yvon IBH Ltd; Coherent Scotland Ltd; CRLO Displays Ltd

Prize Studentship Competition:

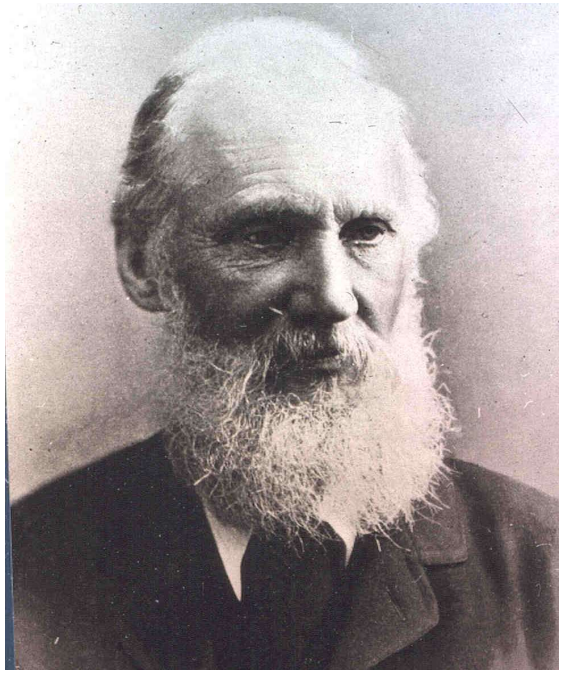
- High profile and prestigious awards.
- Attracting the brightest and best, worldwide.
- 8 awards per year, starting October 2005.
- To be held at any SUPA University and in any theme.

International Summer Schools:

- Building on a 40-year track-record of SUSSP.
- Attracting world-renowned lecturers.

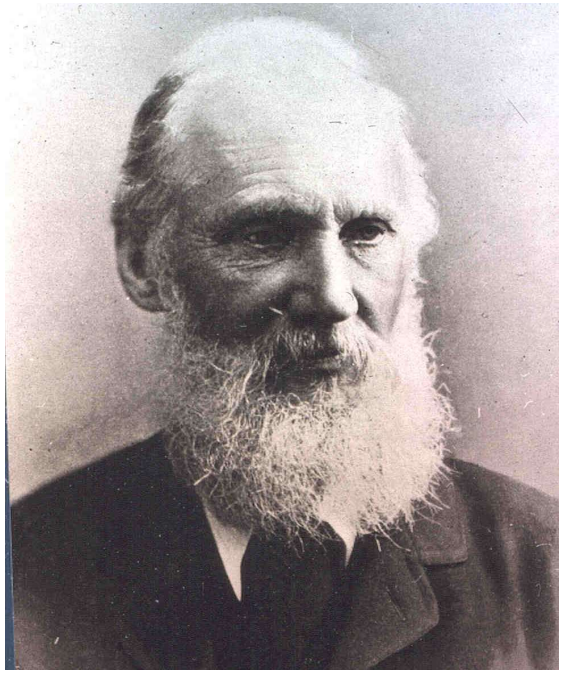
Distinguished Visitor Programme:

- The best scientists in the world spending time in Scotland.
- Adding variety and weight to the local teaching base.
- Stimulating research in new areas.
- Enhancing the profile of Scottish Physics on the world stage.



William Thompson
(Lord Kelvin)
1824 - 1907

William Thompson
(Lord Kelvin)
1824 - 1907



Kelvin in 1900

"There is nothing new to be discovered in physics now. All that remains is more and more precise measurement"

Course Programme

Lectures 1 to 5

1. **Introduction and theoretical foundations – part 1**
2. **Introduction and theoretical foundations – part 2**
3. **Parameter estimation and goodness-of-fit – part 1**
4. **Parameter estimation and goodness-of-fit – part 2**
5. **Parameter estimation and goodness-of-fit – part 3**

Course Programme

Lectures 6 to 10

6. **An Advanced Toolbox for Bayesian Inference**
7. **An Advanced Toolbox for Bayesian Inference**
8. **Bayesian Model Selection**
9. **Monte Carlo Simulation Methods**
10. **Fourier Methods**

Course Assessment

- Series of 'Pop Quiz' questions, spread throughout the lectures (**compulsory** for SUPA students only)
- Series of numerical problems, to be posted on my.SUPA (**optional** for SUPA students)
- Mock data challenge to be posted on my.SUPA. (optional for SUPA students).

SUPAADA is also an Glasgow MSc course: "Advanced Data Analysis for Physics and Astronomy". There are separate assessment arrangements for these MSc students.

[Note: this slide is not featured in the video and audio files]

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

Not logged in Talk Contributions Create account Log in

Article Talk | Read Edit View history | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Big data

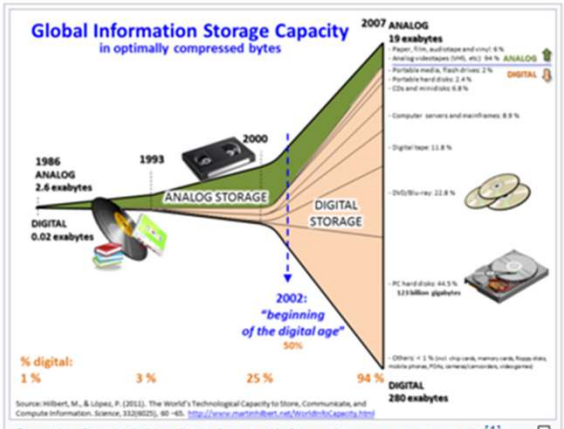
From Wikipedia, the free encyclopedia

This article is about large collections of data. For the band, see Big Data (band).

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.[2] "There is little doubt that the quantities of data now available are indeed large, but that's not the most relevant characteristic of this new data ecosystem."[3]

Analysis of data sets can find new correlations to "spot business trends, prevent diseases, combat crime and so on".[4] Scientists, business executives, practitioners of medicine, advertising and governments alike regularly meet difficulties with large data-sets in areas including Internet search, finance, urban informatics, and business informatics. Scientists encounter limitations in e-Science work, including meteorology, genomics,[5] connectomics, complex physics simulations, biology and environmental research.[6]

Data sets grow rapidly - in part because they are increasingly gathered by cheap and numerous information-sensing mobile devices, aerial (remote sensing), software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks.[7][8] The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s;[9] as of 2012, every day 2.5 exabytes (2.5×10^{18}) of data is generated.[10] One question for large enterprises is determining who should own big-data initiatives that affect the entire organization.[11]



Growth of and digitization of global information-storage capacity[1]

Scottish Universities Physics Alliance

[Note: this slide is not featured in the video and audio files]

IBM Products Services Industries Developers Support Careers Marketplace Search

A Smarter Planet >

Big Data

Data savvy. Insight driven.

What is big data? | Big data technology | Big data conversations

Contact IBM

What is Big Data?

Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills.

Get started with big data and analytics

Three additional slides covered in the video and audio files, relevant to the local arrangements for when the lectures were previously filmed in Glasgow

Three additional slides covered in the video and audio files, relevant to the local arrangements for when the lectures were previously filmed in Glasgow

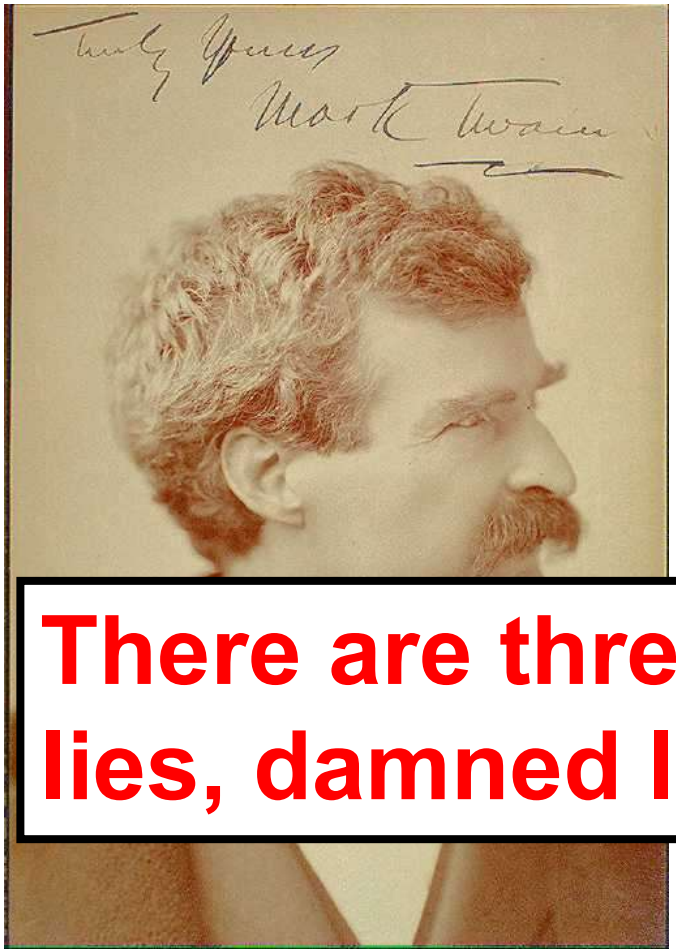
Three additional slides covered in the video and audio files, relevant to the local arrangements for when the lectures were previously filmed in Glasgow

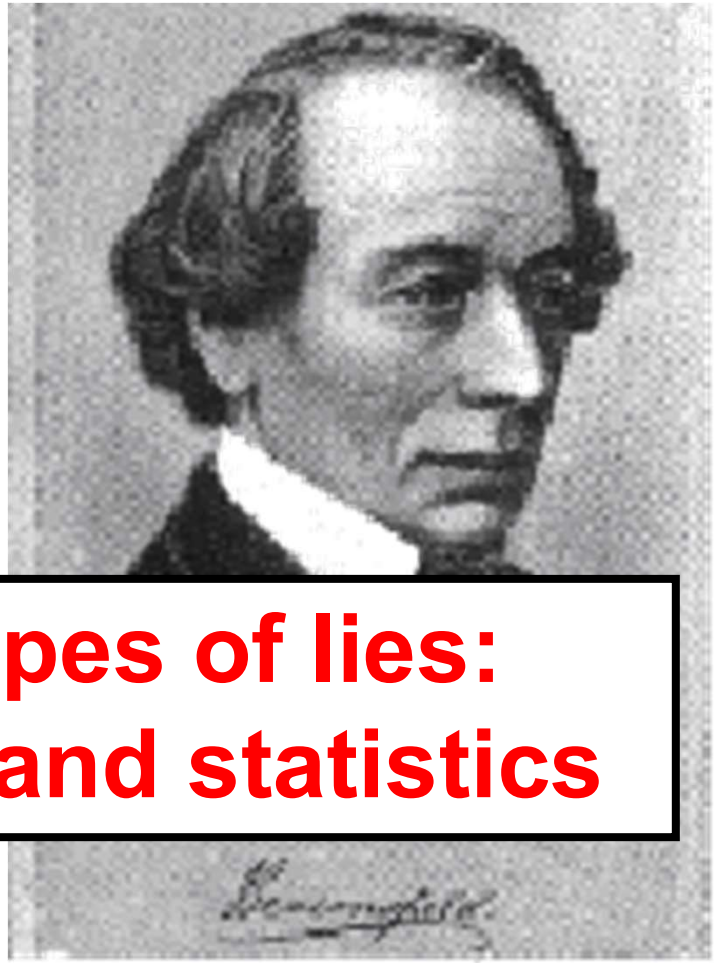
Why a course on data analysis?…

Mark Twain

Benjamin Disraeli

Why a course on data analysis?…





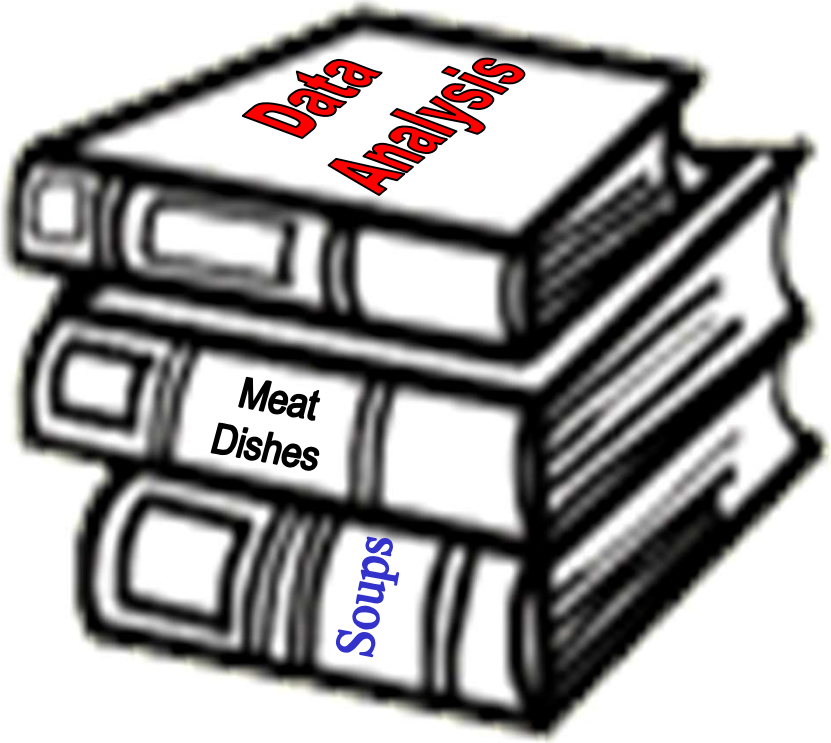
There are three types of lies: lies, damned lies and statistics

Mark Twain

Benjamin Disraeli

Why a course on data analysis?…

Data analysis methods are often regarded as simple recipes…



…but in physics, sometimes the recipes don't work!!!

Why a course on data analysis?…

Data analysis methods are often regarded as simple recipes…

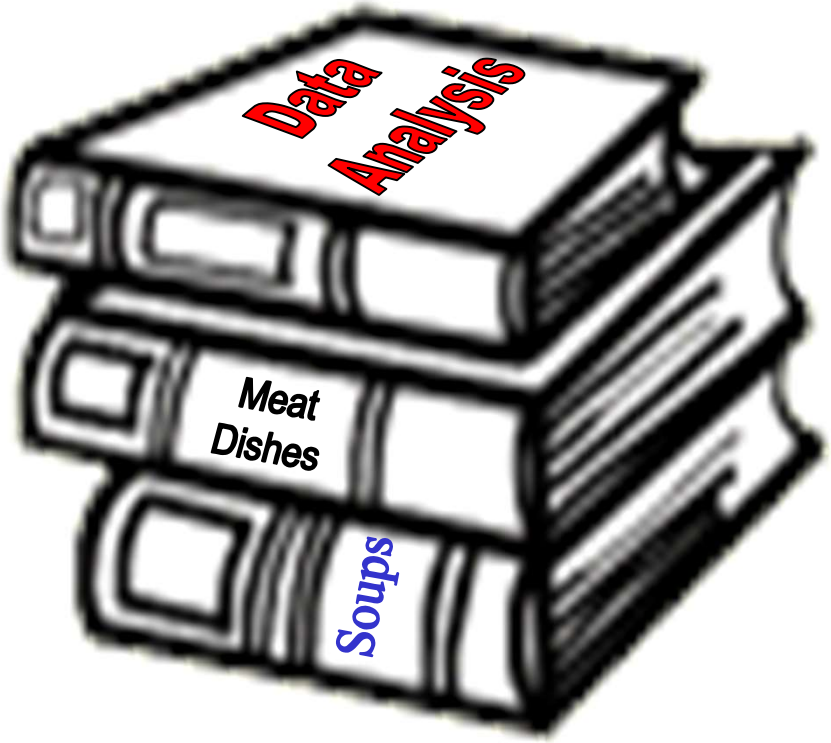
…but in physics, sometimes the recipes don't work!!!

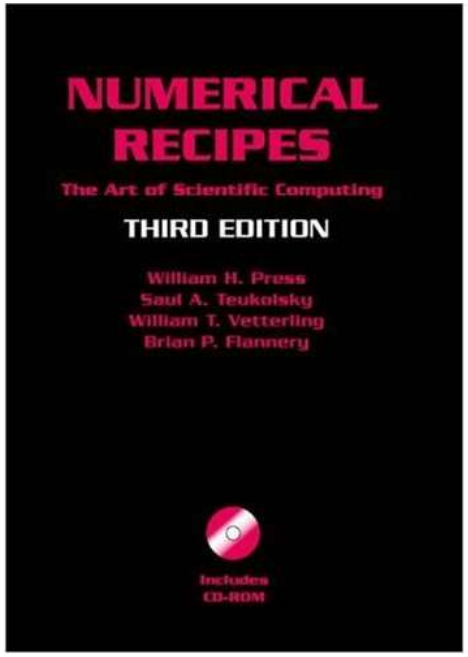
- Very weak signals
- Correlated 'residuals'
- Incorrect assumptions

SYSTEMATIC ERRORS

Why a course on data analysis?…

Data analysis methods are often regarded as simple recipes…





http://www.numerical-recipes/

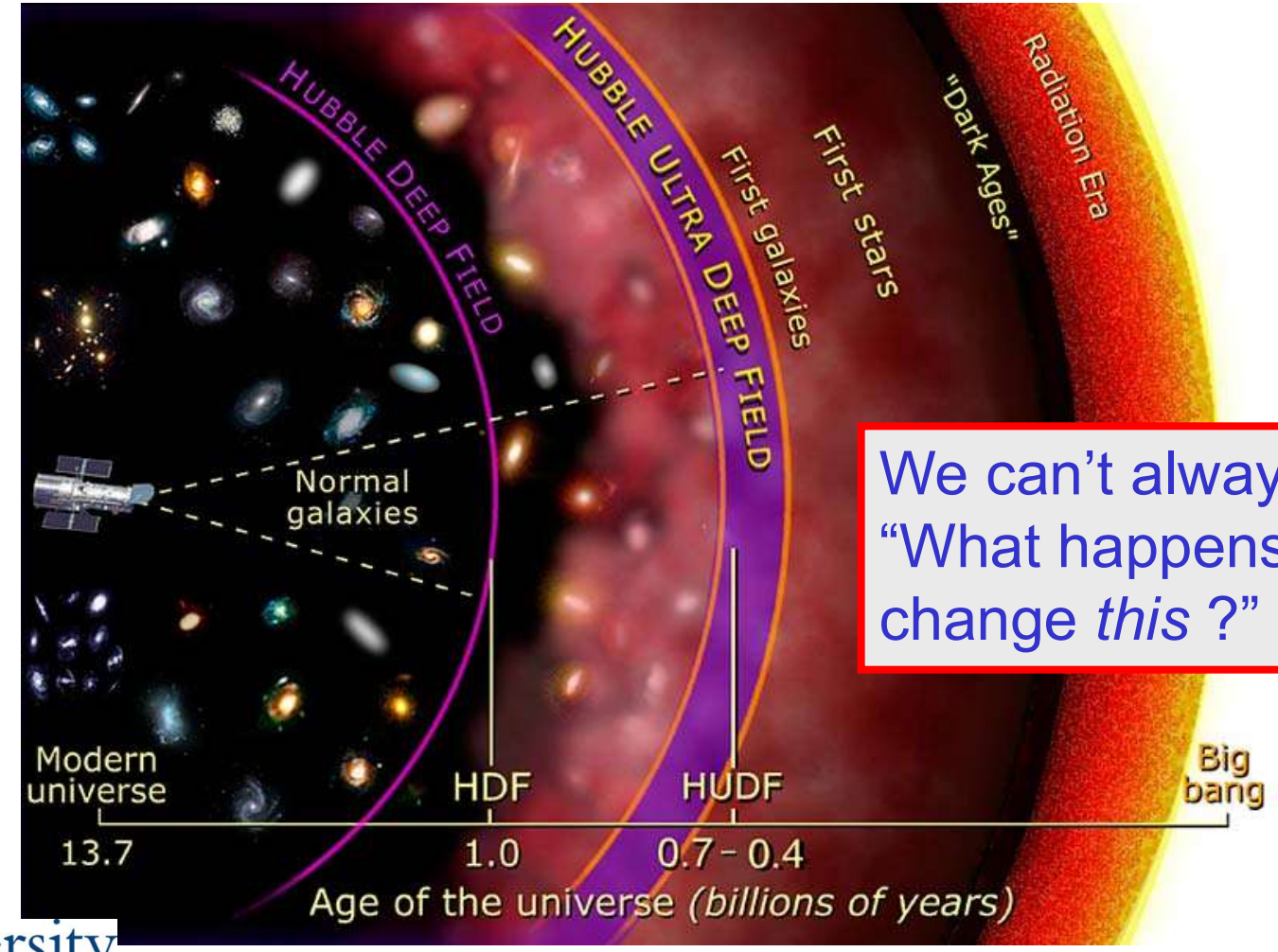
http://www.nr.com/olderswitcher.html

Why a course on data analysis?…

"You must unlearn what you have learned"

Why a course on data analysis?…

Many areas of physics are *Remote sensing*



We can't always ask "What happens if we change *this* ?"

Why a course on data analysis?…

Even if you (will) do little data analysis yourself, you will need to assess critically results in the literature of your field.

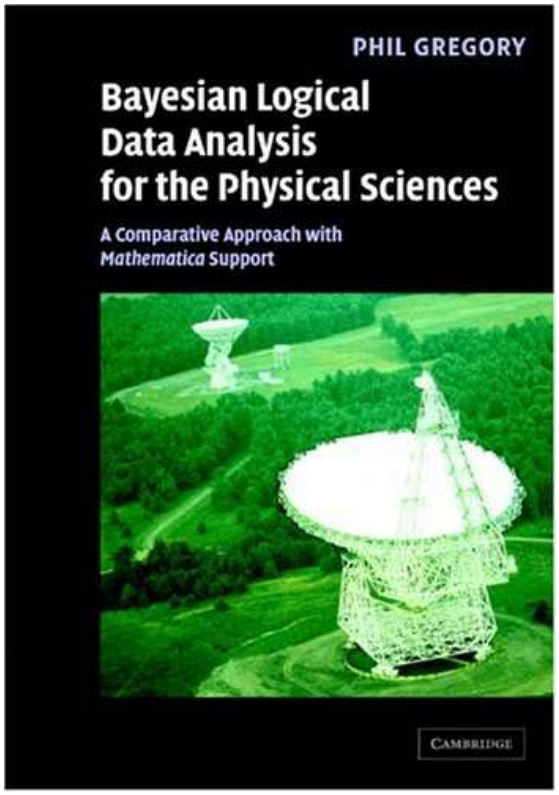
→ Determine **significance** of old and new theories and models.

Many (all?) current analyses are carried out within a particular framework – the 'frequentist' approach to probability:

- 'classical' data analysis methodology
- rapidly losing ground to **Bayesian inference**.

Important to understand the differences between these two approaches, and the strengths of the new Bayesian paradigm.

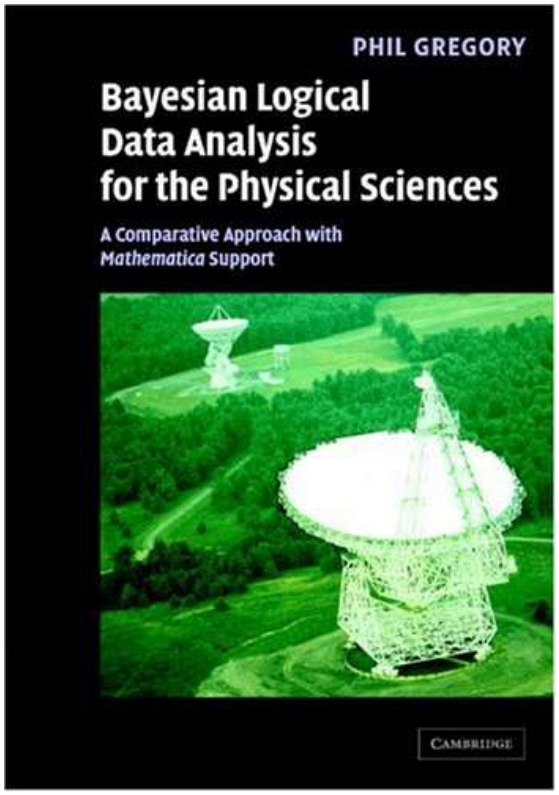
Why a course on data analysis?…



Cambridge Univ. Press

ISBN: 052184150X

Why a course on data analysis?…



Cambridge Univ. Press

ISBN: 052184150X

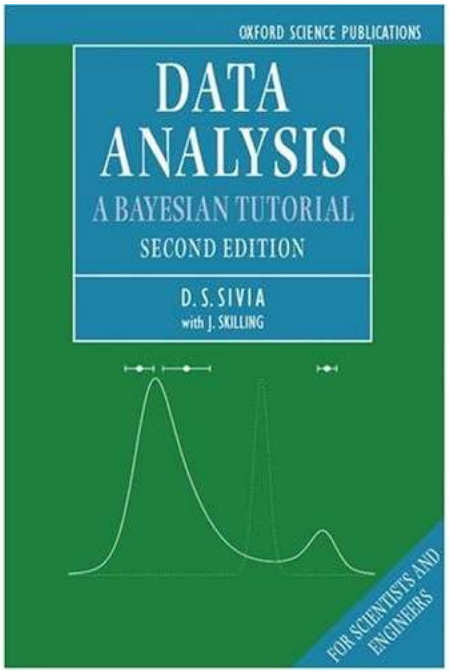
PREFACE

The goal of science is to unlock nature's secrets…Our understanding comes through the development of theoretical models capable of explaining the existing observations as well as making testable predictions…Statistical inference provides a means for assessing the plausibility of one or more competing models, and estimating the model parameters and their uncertainities. These topics are commonly referred to as "data analysis".

We are currently in the throes of a major paradigm shift in our understanding of statistical inference based on…Bayesian Probability Theory…The Bayesian paradigm is becoming very visible at international meetings of physicists and astronomers. However, the majority of scientists are still not at home with the topic and much of the current scientific literature still employs the conventional "frequentist" statistical paradigm.

This book is an attempt to help new students to make the transition while at the same time exposing them to some of the essential ideas of the frequentist paradigm that will allow them to comprehend much of the current and earlier literature and interface with his or her research supervisor.

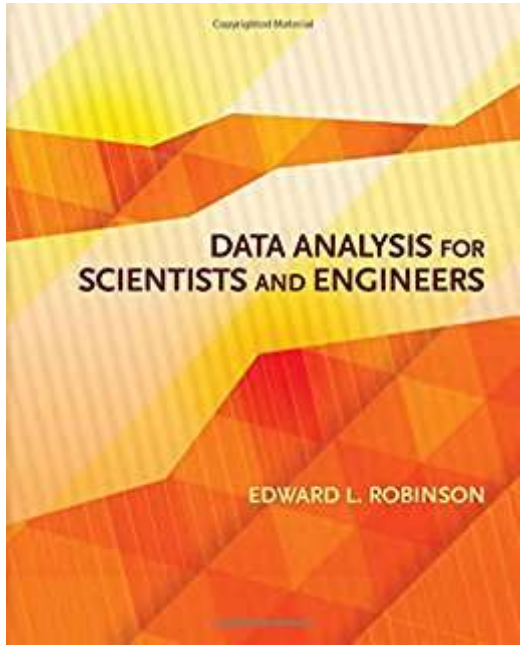
Why a course on data analysis?…



Data Analysis:
A Bayesian Tutorial

(Oxford Univ Press)
D.S. Sivia & J. SKilling

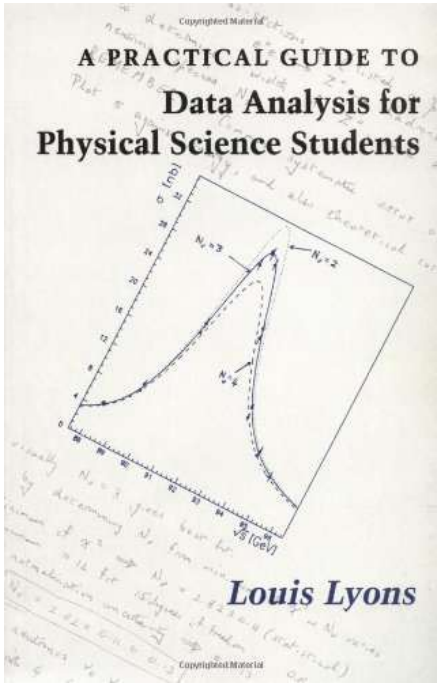
ISBN: 0198568312



Data Analysis for Scientists and Engineers

(Princeton Univ Press)
E.L. Robinson

ISBN: 9781400883066



A Practical Guide to Data Analysis for the Physical Sciences

(Oxford Univ Press)
Louis Lyons

ISBN: 0521424631

Mark Twain

Mark Twain

“It is better to keep your mouth closed and let people think you are a fool than to open it and remove all doubt.”