

# Business Forecasting using R: Welcome

## CMAF Practitioner Workshop

London, 29 April 2019



# Marketing Analytics & Forecasting



Lancaster University  
Management School

*Europe's leading research centre in applied forecasting*

## Services

- Short courses (open & bespoke)
- Consultancy
- MSc summer projects
- Software development
- Knowledge-transfer partnerships
- PhD research projects

## Expertise

- Marketing analytics
- Supply chain forecasting
- Forecasting & planning processes
- Machine learning & AI

*Applied in a wide variety of  
sectors (eg FMCG, govt, pharma)*



Prof John  
Boylan



Dr Sven  
Crone



Prof Robert  
Fildes



Prof Nikolaos  
Kourntzes



Dr Nicos  
Pavlidis



Dr Ivan  
Svetunkov

# Keeping Updated with CMAF

- Website ([www.lancaster.ac.uk/lums/cmaf](http://www.lancaster.ac.uk/lums/cmaf))
  - Open-source software packages
  - Blog posts – summarising recent research
  - Newsletter – courses and events
- Twitter (@LancasterCMAF)
- LinkedIn (Centre for Marketing Analytics and Forecasting)
- Contact us directly ([r.fildes@lancaster.ac.uk](mailto:r.fildes@lancaster.ac.uk))



# Upcoming CMAF Activities

Training Course - Business Forecasting with R **9<sup>th</sup>-10<sup>th</sup> May, London**

## Day 1

- **Forecasting principles:** what makes forecasting different from conventional statistical modelling?
- **Time series exploration and decomposition:** what can be captured in a time series? How to identify what is in our data?
- **Forecasts evaluation:** what is a good forecast? What are useful metrics of forecast performance and how to evaluate our forecasting process?
- **Exponential smoothing family of models:** what it can and cannot model, as well as understand how to best implement the models in the data at hand.

## Day 2

- **Exponential smoothing with events and promotions:** we introduce ways to model special events, bank holidays, promotions, etc.
- **The basics of regression:** causal modelling demonstrating the ease to build regression models that capture effects of exogenous variables and events with our forecasts.
- **Advanced forecasting methods:** introduction to advanced approaches such as ARIMA, MAPA and simple Neural Networks.



# Business Forecasting using R: R we there yet?

## CMAF Practitioner Workshop

Nikolaos Kourentzes

Professor, Department of Management Science, LUMS  
Centre for Marketing Analytics and Forecasting  
Visiting Professor, Stockholm School of Economics

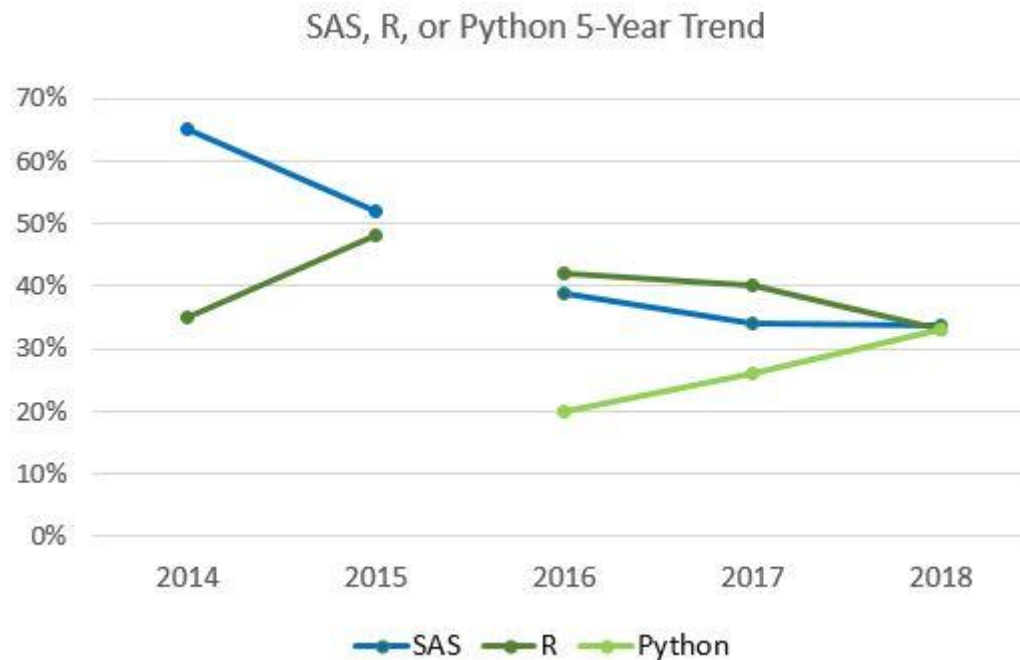
London, 29 April 2019

Marketing Analytics  
and Forecasting



Lancaster University  
Management School

# We do analytics, let's get the data

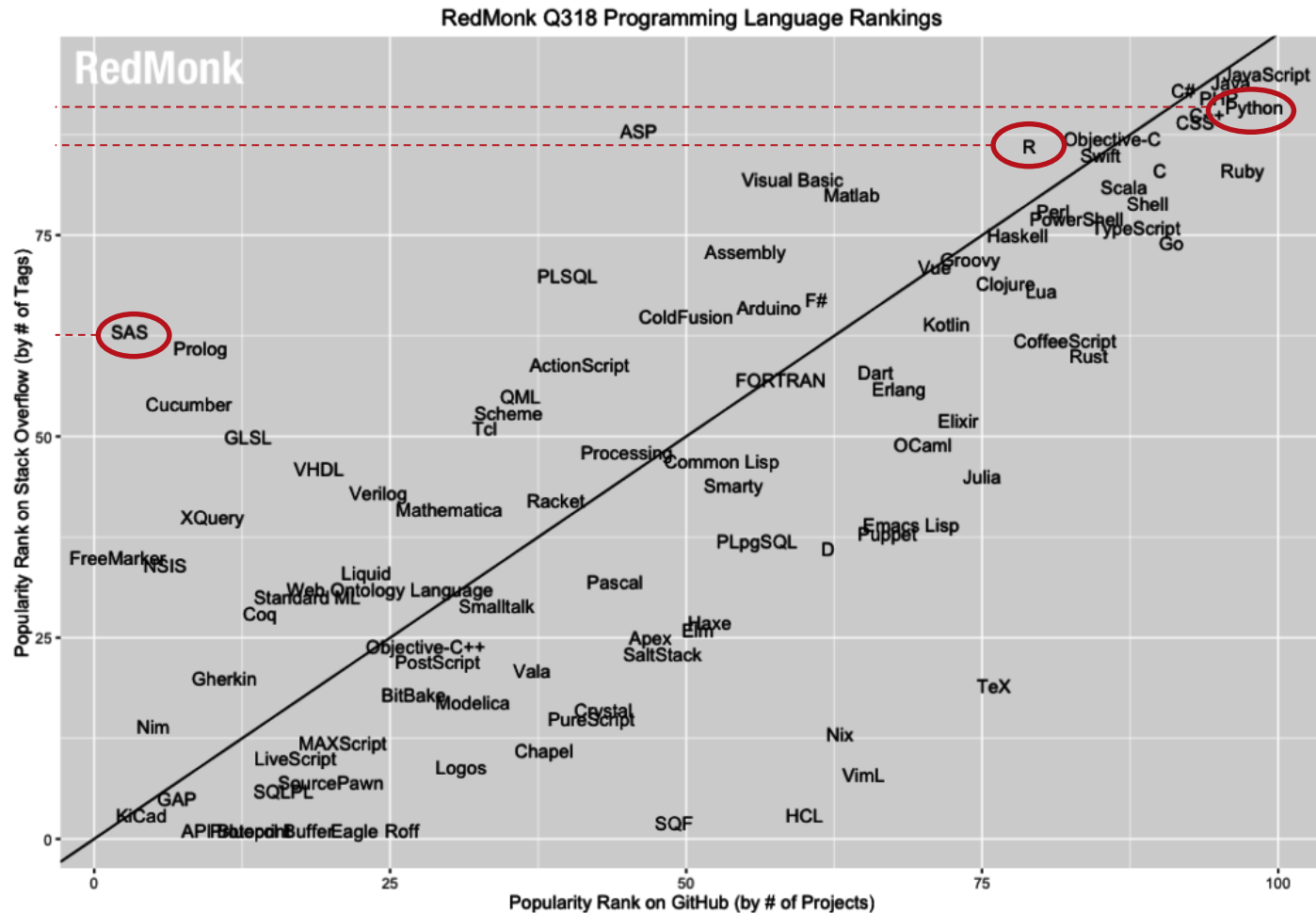


Note: Python option added in 2016

Data ©2018 Burtch Works LLC

<https://www.burtchworks.com/2018/07/16/2018-sas-r-or-python-survey-results-which-do-data-scientists-analytics-pros-prefer/>

# We do analytics, let's get the data

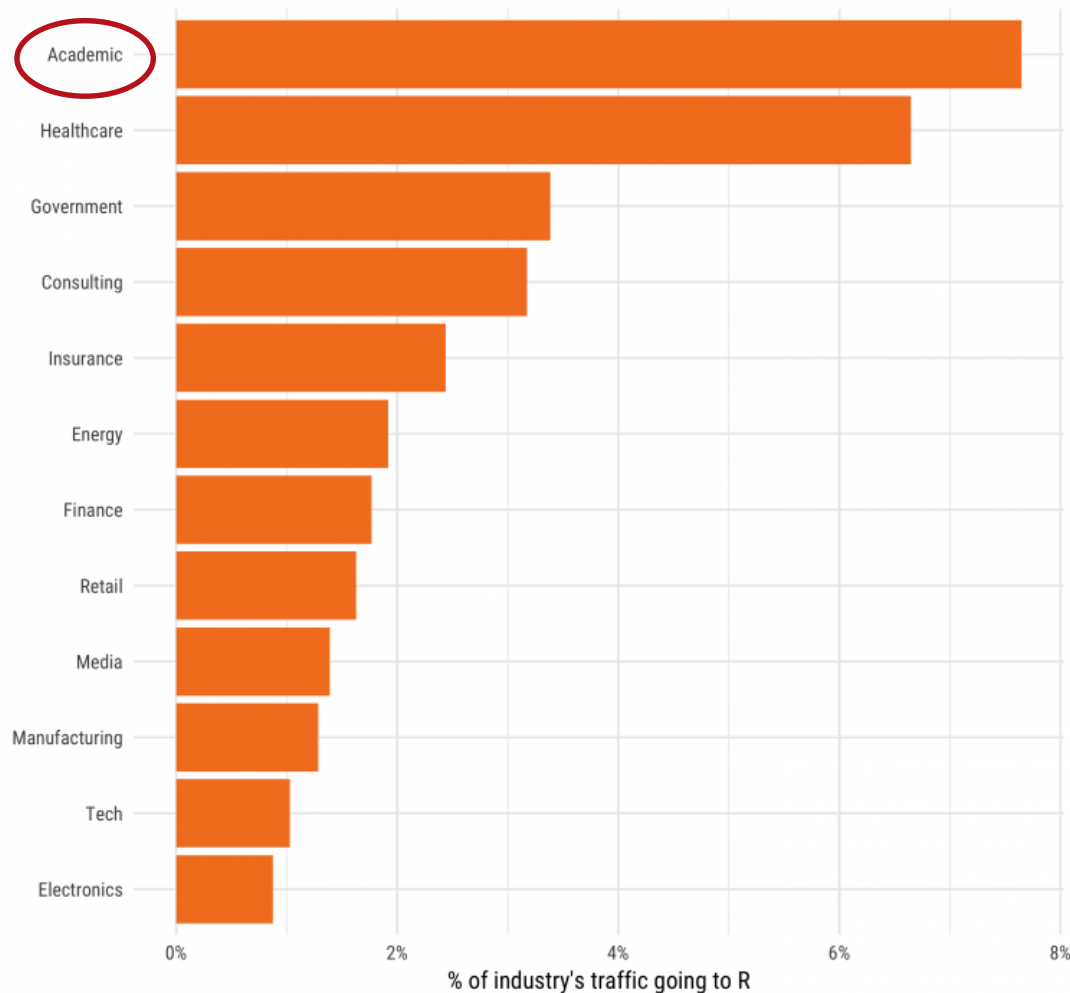


<https://redmonk.com/sograzy/2018/08/10/language-rankings-6-18/>

# We do analytics, let's get the data

## Visits to R by industry

Based on visits to Stack Overflow questions from the US/UK in January-August 2017.  
The denominator in each is the total traffic from that industry.



<https://stackoverflow.blog/2017/10/10/impressive-growth-r/>



# What the data tells us?

- Open source is no longer fringe, industry is using it, as evident by the large community of corporate users.
- **BUT** Academia is still leading the usage.

Academia has different motivations to industry: **(i)** need to innovate; **(ii)** need to make research transparent; **(iii)** need to make research open to the public (if you want public funds!); **(iv)** academics often have unorthodox IT practices; **(v)** ridiculously high skill level. **Only open source ticks all boxes.**

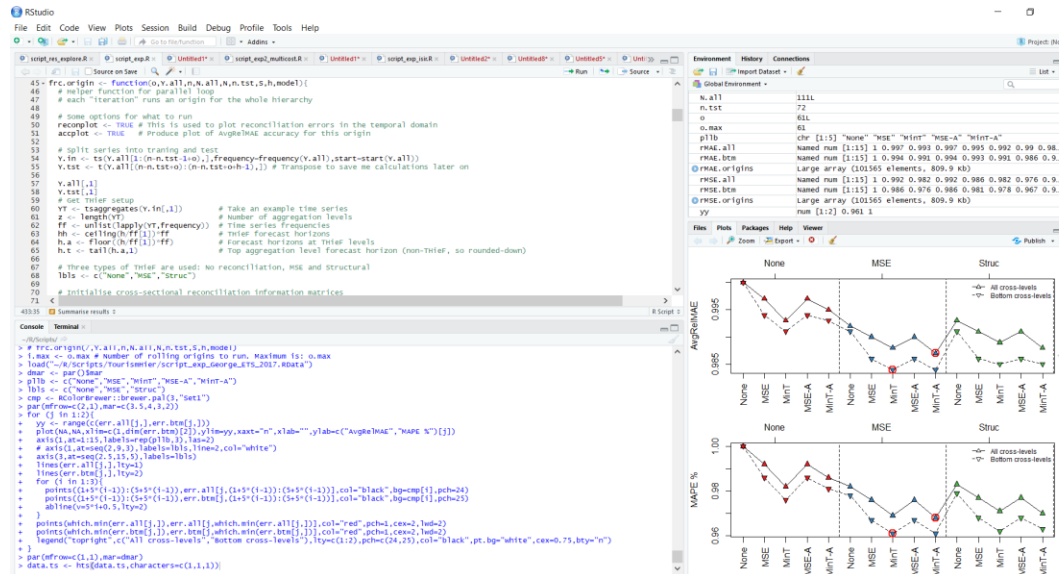
- Industry has different requirements:
  - No pressure for transparency to externals (review process);
  - Solutions are products and competitive advantages and (often) not to be given for free;
  - Solutions are **typically not the code! Solving a business problem**;
  - Companies cannot afford ad-hoc IT practices → otherwise: jumbled Excel mess;
  - BUT still need to **innovate**.

# Where did big software houses lost the bet?

- Could never hire the top talent → academia is more fun + you don't have to show up in the office!
  - Typically software houses follow existing research with limited quality control.
- Are in the job of solving customer problems, often that becomes not forward looking.
  - Biggest failing: Why algorithm X is not implemented? Our customers do not ask for it.
    - Customers (should!) pay for expertise, and get the solutions they truly need, not what they think they need!
- Have to wrangle through a complete intellectual property legal mess (companies do not help themselves with being extra lawyery: scare away academics - also US/EU differences in laws doesn't help either).
  - Developing (plagiarising) everything in house, without the experts, leads to "interesting" results.
- Even if you put lots of capital, eventually companies have X number of employees.
  - The open source community/academia is massive: we have grad students!

# Where did big software houses won the bet?

- The open source community develops tools, packages, libraries, code snippets **but not solutions!**



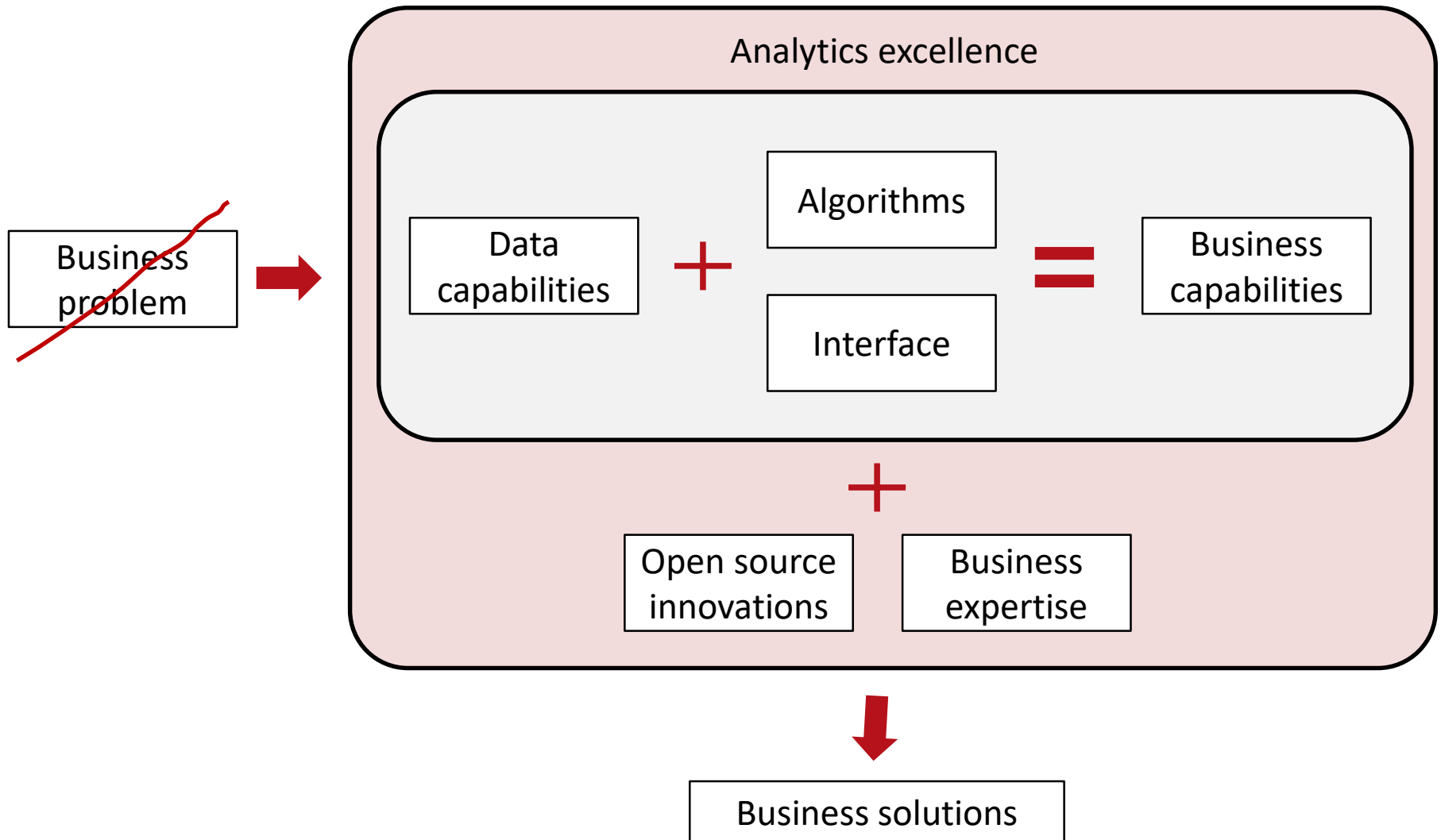
Me solving a case study, interface and data considerations not for the faint hearted.

- Not all data “scientists” are made equal: some are programmers!
- Analytics is not a bunch of code snippets put together

- Data considerations
- Business insights
- Access to non-PhD users

**Open source is missing the “business glue”**  
Serious software houses offer exactly that.

# The way forward



# Not all algorithms are made equal

- Even basic forecasting techniques (e.g. exponential smoothing) can be implemented in many ways
  - Many hidden (often hard-coded) parameters.
  - Improved solutions (e.g. better model selection/combination).
  - Additional capabilities
    - Exponential smoothing is often implemented with a small set of options
    - **forecast** package for R has become the standard of a “full” implementation
    - **smooth** package for R provides much needed capabilities, such as explanatory variables, promotions, etc.
    - Other packages go further to enhance the modelling methodology, e.g. **MAPA** and **thief**
    - Yet many proprietary software do not even provide reliable automatic modelling.
    - Some have even coding errors.

# Not all algorithms are made equal

- A case to the point: **AI for forecasting**
- Needs two elements:
  1. Good AI models and maths, especially in terms of implementation.
  2. Good modelling methodology to take the barebone maths to forecasting applications
    - Which variables?
    - How to train?
    - How big networks, etc.
- There are multiple packages for R that implement neural networks, or facilitate using external libraries, like **Tensorflow**, but these are forecasting specific or take care of business aspects of forecasting → check our **nnfor** package.

# Not all algorithms are made equal

- Open source offers some unique capabilities that many proprietary software do not offer yet.
  - New product forecasting: **diffusion** package
  - Temporal hierarchies: **thief** package
  - Cross-sectional hierarchies: **hts** package
  - Intermittent demand: **tsintermittent** and **smooth** packages
  - Google trends for forecasting: **GTT** package
- Some proprietary software call these packages to provide these capabilities, however not all options/parameters are open to the users.
  - Not necessarily a bad thing!

# A matter of interface

- Who are the users? What are their needs? What are the business needs?
- Interface is much more important than the thought users and software houses put into it.
- The interface encapsulates the analytical logic behind the capability/solution → is it fit for purpose?
- Standard omissions/bad design elements in software:
  - Data exploration
  - Model diagnostics
  - Incorporation of soft information and judgement
  - Feedback – you got accuracy X%, **why?**

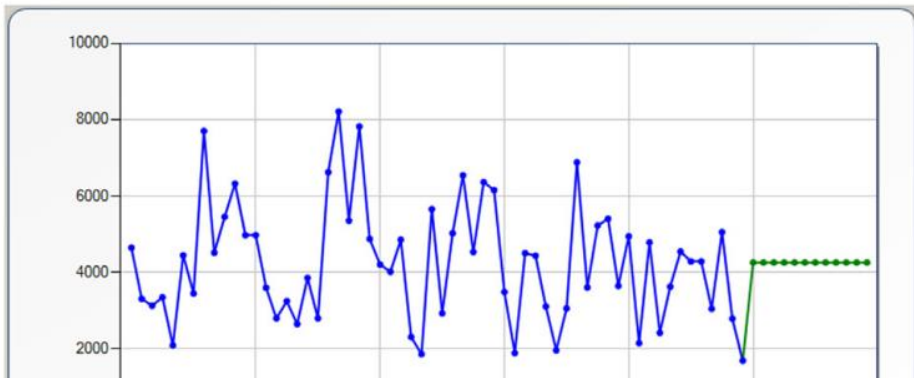


# A matter of interface



Experiment on interface and selection of forecasting method

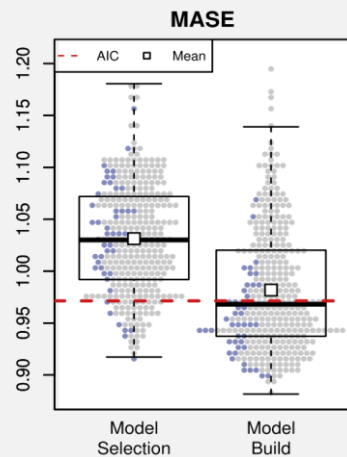
Model Selection



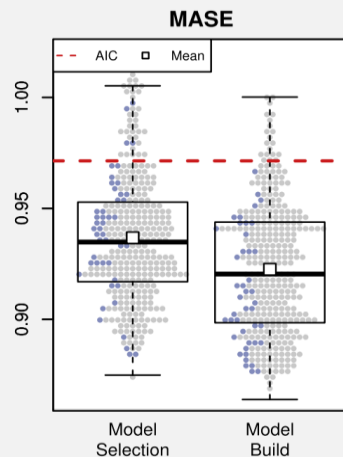
Model Build



All participants



Experts



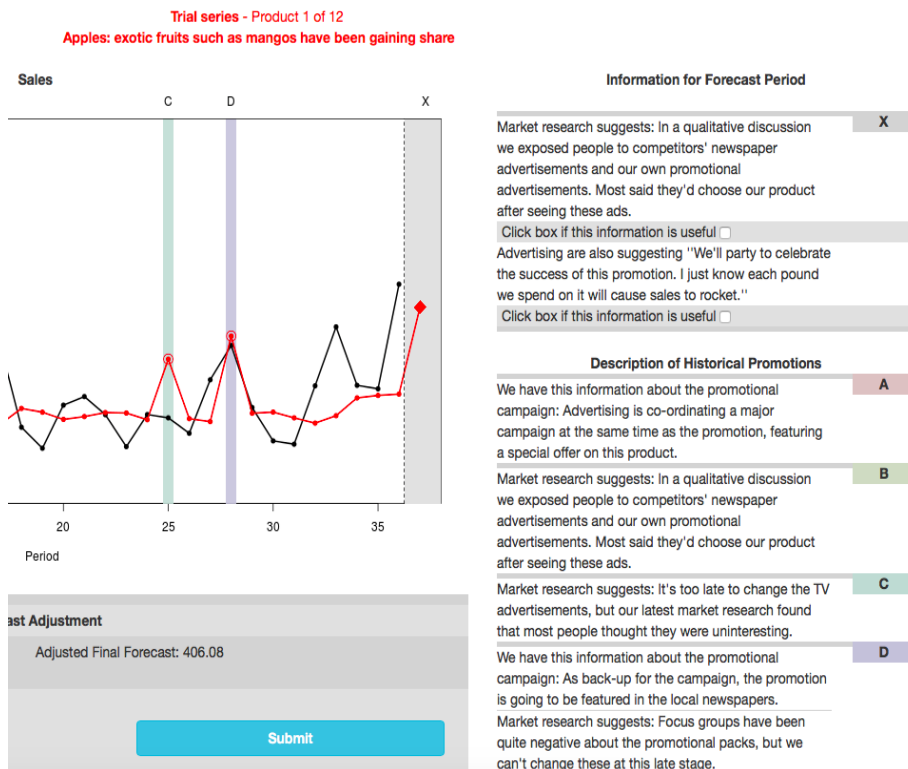
- Interface had a dramatic effect on performance of users, for exactly the same data and forecasting methods.
- This is only one part of the forecasting process. How many choices are arbitrary in the design process?
- How flexible is your forecasting interface? Fit for purpose?

# A matter of interface

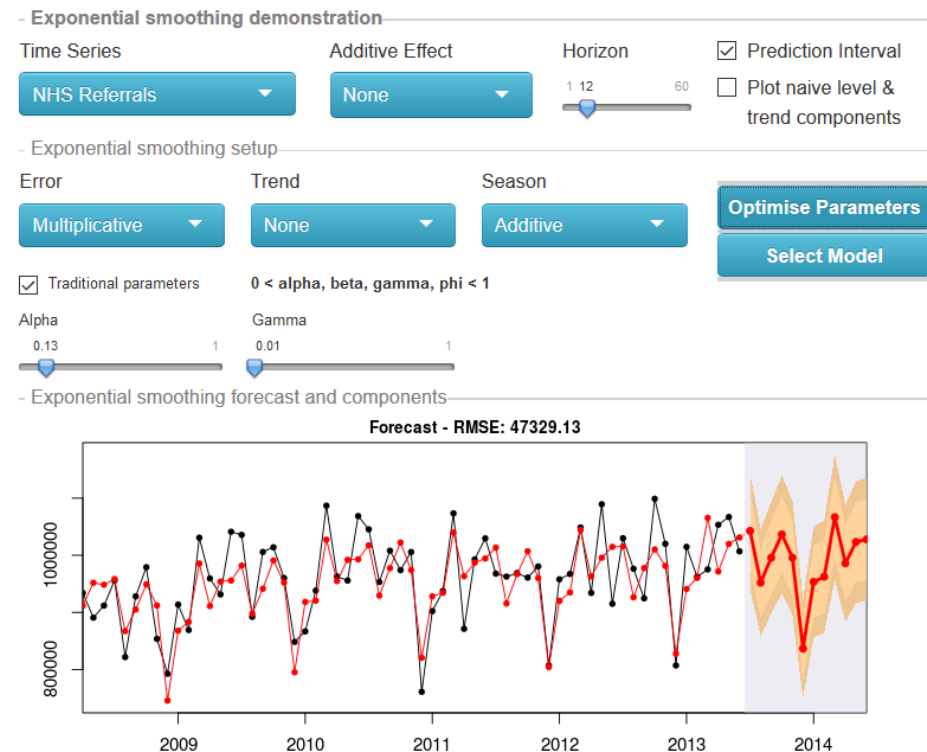
Dashboards are very effective ways to summarise and present information

- Specialised open source widgets and tools. Box plots → Violin plots → Bean plots
- Horses for courses: different tasks require different interface.

## Experiment for incorporating soft information



## Training interface for exponential smoothing



# So, can we get software cheap?

- Open source: **it is not zero cost!**
  - Support?
  - Licenses?
  - Open source ethics and eco-system!
- Quality of implementation and scalability
  - You still get what you pay!
- State-of-the-art techniques?
  - Surprisingly, you don't get what you pay. Actually you typically pay too much!
  - A call for action: innovation happens with your help!
- Develop in-house?
  - What expertise do you have in-house?
  - Economies of scale?

# Conclusion – (no plural, it is only one!)

- **It is not either or!**
    - Proprietary software cannot compete with open source in innovation.
    - Open source cannot compete with proprietary software in:
      - Implementation;
      - Interface;
      - Data considerations.
    - They are two different complimentary worlds – use them as such!
  - Demand from your software to access open source solutions.
    - Support open source, so that it can keep innovating for you.
    - Big software companies do (Google, Facebook, Microsoft, SAS, etc. actively do!)
- ... and analytics/AI is not just code and interfaces. **It is the business expertise of the users.**

# The most important slide: A call to action

- If only I had a PhD for every time a company asks me for a solution to their **specific** problem...
  - Academia has the incentive to do the research to find a good quality solution, rather than just a solution.
  - Academia provides resources for free.
  - But for academia to do this work it needs data and resources from industry, **with publication rights**
  - Otherwise, we cannot solve your problems. **The IP is not in the maths or the modelling – an open culture in business means better solutions in business.**
    - Look at examples of leading companies, they contribute to the open source ecosystem and even provide their own solutions for others to use.
    - Why? Innovation happens faster from a community, rather than a small team.
- **Help us solve your problems by sharing your challenges, your data and understanding that PhD students do require food and water at least once per week (i.e. pay for expertise and not flashy colours)**
- **CMAF needs to demonstrate use of its work and research, so it can continue offering it for free → help us demonstrate that.**

# Resources

Centre for Marketing Analytics and Forecasting ([www.lancaster.ac.uk/lums/cmaf](http://www.lancaster.ac.uk/lums/cmaf))

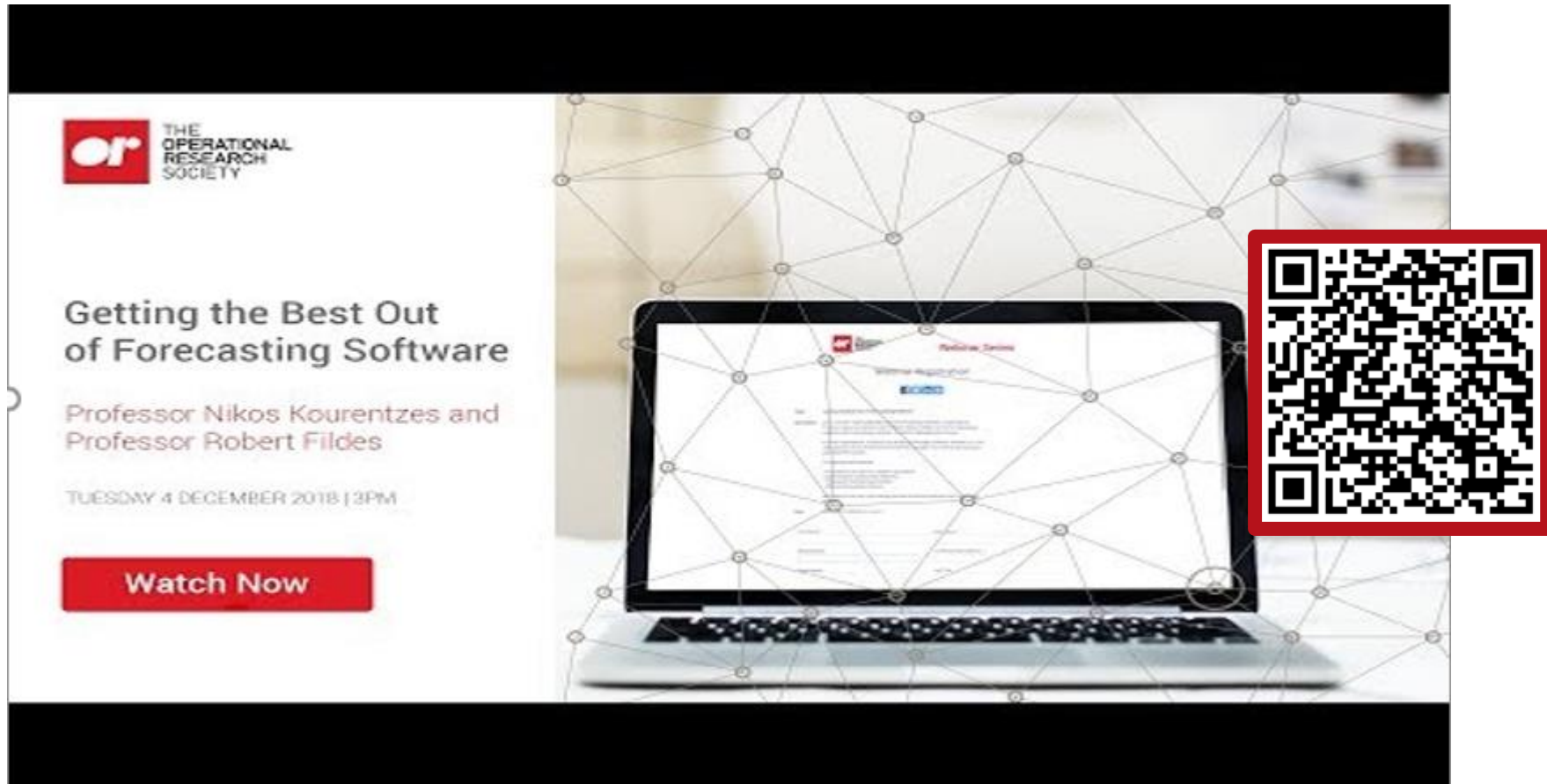
CMAF R forecasting packages:

- **MAPA**: forecasting with multiple temporal aggregation;
- **thief**: forecasting with temporal hierarchies
- **nnfor**: (shallow) neural networks for time series forecasting
- **smooth**: ETS and ARIMA (incl. high frequency data and explanatory variables)
- **tsintermittent**: forecasting intermittent series
- **diffusion**: life-cycle modelling
- **TStools**: supporting functions for time series modelling (available on Github)
- **greybox**: supporting functions for time series modelling
- **GTT**: handle Google Trends data

Research blog ([nikolaos.kourentzes.com](http://nikolaos.kourentzes.com))

- Examples and tutorials for forecasting related libraries and packages
- Research (hopefully!) explained in non-academic terms

# Resources: Webinar on Forecasting Software



The banner features a white background on the left with the Operational Research Society logo (a red square with 'or' in white) and the text 'THE OPERATIONAL RESEARCH SOCIETY'. Below this is the title 'Getting the Best Out of Forecasting Software' in bold, followed by the speakers 'Professor Nikos Kourentzes and Professor Robert Fildes' and the date 'TUESDAY 4 DECEMBER 2018 | 3PM'. A red button with 'Watch Now' is at the bottom left. The right side of the banner shows a laptop displaying a network diagram, with a large QR code overlaid on the right side of the laptop screen.

**or** THE OPERATIONAL RESEARCH SOCIETY

**Getting the Best Out of Forecasting Software**

Professor Nikos Kourentzes and Professor Robert Fildes

TUESDAY 4 DECEMBER 2018 | 3PM

**Watch Now**

<http://kourentzes.com/forecasting/2019/01/14/webinar-getting-the-best-out-of-forecasting-software/>

**Thank you for your attention!**  
**Questions?**

Nikolaos Kourentzes

email: [nikolaos@kourentzes.com](mailto:nikolaos@kourentzes.com)

twitter [@nkourentz](https://twitter.com/nkourentz)

Blog: <http://nikolaos.kourentzes.com>

Full or partial reproduction of the slides is not permitted without authors' consent.

Marketing Analytics  
& Forecasting



Lancaster University  
Management School