

SatRdays London 2024

Schedule and Abstracts



27 April 2024

Invalid Date

Schedule

Time	Title	Room
09:00 - 09:30	Registration	
09:30 - 09:45	Welcome	
09:45 - 10:15	Andrie de Vries	
10:15 - 10:45	Nicola Rennie	
10:45 - 11:15	Break	
11:15 - 11:45	Myles Mitchell	
11:45 - 12:15	Matt Thomas	
12:15 - 13:15	Lunch	
13:15 - 13:30	Welcome from CUSP London	
13:30 - 14:00	Charlie Gao	
14:00 - 14:30	Michael Hogers	
14:30 - 15:00	Break	
15:00 - 15:30	Matthew Lam & Matthew Law	
15:30 - 16:00	Hannah Frick	
16:00 - 16:15	Close	



Abstracts

Andrie de Vries

Director of Product Strategy at Posit



Lessons learnt from Product Management, applied to Data Science

As a Data Scientist you build data products all the time. You may even have worked with a Product Manager to create analyses and dashboards for decision making.

But are you applying the skills of product management in your data science role?

In this talk Andrie provides an overview of Product Management (PM), and what he's learnt over two decades of managing products, ranging from hardware (Psion PDAs) to software (Microsoft R Open, Posit Workbench) and hosted services (MRAN).

Every product manager must consider the new product adoption life cycle, managing the stages from finding the first innovators, managing growth and ultimately the end-of-life process.

During this process you must manage your product so that it's usable (customers want it), feasible (you can build it) and valuable (you can do this sustainably). Many frameworks exist to think about discovering what customers want, the jobs they must get done, forming a value proposition, managing a product roadmap, working with dev teams to build it, and working with marketing and sales to create a compelling sales pitch.

As a data scientist, you can benefit from product management knowledge by thinking of your app as a product. You must convince your users (internal customers) to use this app (at the cost of changing their workflow).

I will leave you with a map to get started with classic resources, including Geoffrey Moore, Marty Cagan, Teresa Torres, April Dunford and Lenny's Podcast.

Hannah Frick

Software Engineer at Posit



Survival analysis is coming to tidymodels!

If you have time-to-event data, such as data on customer churn, data on the lifetime of machines, or similar, survival analysis with its censored regression models gives you the ability to include all your observations in the model appropriately, including those where you may not have observed the event yet. The tidymodels framework is a collection of packages for safe, performant, and expressive supervised predictive modeling on tabular data. The framework's consistency makes switching between models easy, its guardrails against common pitfalls such as overfitting due to data leakage make it safe. It covers the entire modeling workflow: preprocessing and feature engineering, models, resamples, performance metrics, and tuning. We are now extending support for survival analysis across the entire tidymodels framework with dedicated models and metrics, allowing the same ease and expressiveness as for classification and regression, across all steps of the modeling process.

Charlie Gao

Director & Research Scientist at Hibiki AI Limited



'mirai' for Shiny and Plumber Applications

'mirai' is Japanese for 'future'. Some of the existing solutions for parallelization in R have not fundamentally changed in 20 years. The technologies behind 'mirai' are, in contrast, modern and minimalist, and provide a level of performance that will be noticeable for demanding, client-facing workloads typical of Shiny and Plumber applications. As a scheduler for distributed tasks, 'mirai' currently powers the high performance computing needs for the 'targets' reproducible-workflow ecosystem, whether locally, on traditional HPC clusters or the cloud. It has undergone the validation required to reliably handle demanding scientific workloads such as clinical trials simulations. At R Project Sprint 2023, it was integrated as a backend for the base R 'parallel' package at the request of R-Core. The same industrial-strength, yet incredibly lightweight solution is now available to power large-scale Shiny and Plumber applications. This presentation demonstrates how 'mirai' works in typical example situations which benefit from parallelization of computations, and the different ways they may be distributed to background processes on the same machine or across a network of servers. A particular highlight will be the zero-configuration TLS option. This 'just works' to protect remote connections using single-use certificates generated on-the-fly. This was developed under an R Consortium infrastructure grant that aims to make such technologies available to the wider R community.

Michael Hogers

Tech Lead at NPL Markets Ltd



Modular Shiny(Proxy) - a SaaS setup

I aim to provide a talk that displays how one can use R, Shiny and ShinyProxy (or other deployment methods) to create a modular SaaS platform that later allows to swap out modules of the platform with new languages or frameworks. The key ingredients are: use a database back-end across Shiny modules, deploy modules as relatively small apps to dedicated URL endpoints, use a shared UI library across Shiny modules and package your Shiny apps (+ use CI/CD) while keeping business logic separated to later on export business logic functions.

Matthew Lam

Graduate GIS Consultant at Mott MacDonald



Matthew Law

GIS Consultant at Mott MacDonald



How Mott MacDonald unlocks the power of geospatial data with R

Mott MacDonald is a global engineering, management, and development consultancy with a broad portfolio of projects across various engineering disciplines. Geospatial data plays an instrumental role in supporting projects in these sectors, enabling us to understand the world around us so that we can make better informed decisions, improve efficiencies, and drive digital innovation. In this proposed presentation, we will illustrate how we use R at Mott MacDonald to harness the power of geospatial data with two examples – Risk Modelling for Ash Dieback and Creative Geospatial Visualisation for Impactful Communication. The Ash Dieback Pipeline is a computer vision project which attempts to identify trees with the Ash Dieback disease from video footage of roadways around the UK. We intend to showcase how we use R to process a variety of geospatial datasets and attempt to model the risk to road users associated with a diseased tree remaining untreated. Our work at Mott MacDonald often involves wrangling complex datasets to answer multifaceted questions. R provides excellent toolkits for integrating, analysing, and visualising geospatial datasets. We intend to demonstrate how R can be used for creative visualisation of geospatial data to extract and communicate actionable insights. Through these examples, we hope to outline our team's maturity journey towards building multilingual spatial data science capabilities alongside traditional GIS platforms.

Myles Mitchell

Data Scientist at Jumping Rivers



Using R to teach R

At Jumping Rivers, we teach over forty courses covering data science topics, including programming, data visualisation and machine learning, in R as well as Python, Tableau, Git, Docker and Stan. Most courses follow the same template: static notes, live coding scripts and presentation slides. For every taught course we also have to spin up a bespoke virtual environment, collect feedback and generate certificates.

In this talk, I will explain how we have used R to streamline the course writing process, automate the course build and deployment to Posit Workbench, and conduct post-course administrative tasks. With over 100 courses taught every year, each step in this pipeline must be rigorously tested so that, on the day, the trainer can focus on the attendees without having to worry about technical issues.

I will draw on our process's successes (and shortcomings) and share some take-home lessons applicable to any big coding project, including packaging of source code, automated testing and scheduled builds.

Nicola Rennie

Lecturer in Health Data Science at Lancaster University



Typst or LaTeX? Styling PDF documents with Quarto extensions

Quarto is an open-source scientific and technical publishing system that allows you to combine text with code to create fully reproducible documents in a variety of formats. The addition of custom styling to documents can make them look more professional and recognisable. In this talk, I'll give an overview of ways to create customised PDF documents using Quarto. Until recently, this meant getting to grips with LaTeX. Now, there's a new kid on the block: Typst. Typst is an open-source typesetting system that is designed to be as powerful as LaTeX while being much easier to learn and use.

Extensions are a powerful way to modify and extend the behaviour of Quarto, including adding styling to your documents with LaTeX or Typst. To demonstrate the differences between LaTeX and Typst, I'll walk through the process of converting a LaTeX-based style extension to Typst, allowing users to easily switch between them. We'll compare the two – discussing error messages (we all get them!), render time, and customisability along the way.

Matt Thomas

Head of Strategic Insight & Foresight at British Red Cross



Where data meets disaster: A journey through the British Red Cross's 'humaniverse'

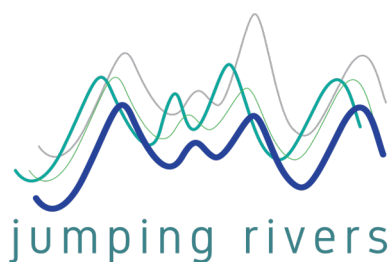
The 'Humaniverse' is a suite of R packages produced by the British Red Cross's data scientists for sharing humanitarian data and tools. Open data and analyses are vital for 21st Century humanitarianism and these packages have transformed the speed and scale at which we can provide answers about emerging and ongoing humanitarian crises in the UK. In this talk, I will offer an overview of the Humaniverse and will share some of the ways we have used this infrastructure to inform how the British Red Cross supports people affected by disasters, displacement, and health crises. I will cover our core R packages, discuss how and why we work in the open, demonstrate some of the analyses and apps we've built using this infrastructure, and share our ambitions for the future of the Humaniverse.

Sponsors



CUSP London is the Centre for Urban Science and Progress based at King's College London, UK. Their mission is to support interdisciplinary research and innovation using Data Science in and for London.

Find them on Twitter @CuspLondon



Jumping Rivers is an analytics company specialising in creating bespoke solutions for modern business problems. Their team of data science and engineering experts come from many different backgrounds, and their wealth of knowledge and experience allows them to think outside the box and solve problems in new and innovative ways.

Find them on Twitter @jumping_uk



The central mission of the R Consortium is to work with and provide support to the R Foundation and to the key organizations developing, maintaining, distributing and using R software through the identification, development and implementation of infrastructure projects.

Find them on Twitter @RConsortium