



# From SPSS to R

Designing a scalable R infrastructure to support health analytics in Scotland

**Terry McLaughlin**  
Principal Information Analyst, Consultancy Services

Information Services Division

# The Organisation



**The Scottish  
Government**  
Riaghaltas na h-Alba

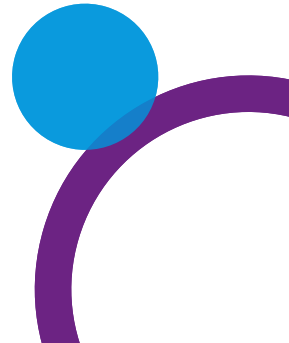


**NHS**  
National  
Services  
Scotland

- Provides information & intelligence services to support the delivery of Health & Social care across Scotland.



- Employs 300+ analysts, mostly based in our offices in Edinburgh and Glasgow, with some embedded in local Health Boards, Councils and within Scottish Government.



# Public Health Reform in Scotland



# SPSS

- IBM SPSS Statistics has been the analytical tool of choice in ISD Scotland for decades.
- Until very recently, analysts have relied on SPSS for the production of National Statistics publications and to deliver business-critical analytical functions.
- 101 concurrent SPSS licenses, with desktop clients connected to a dedicated SPSS server.
- Most analysts use SPSS for extracting data from databases, data wrangling, basic statistics, and outputting flat files e.g. CSV.
- These files are formatted in Microsoft Excel for publication and Tableau for internal NHS use.



## **SPSS *cont.***

- Analysts have felt for some time that our publications and routine analytical processes could be streamlined, automated and made more reproducible.
- There's general acknowledgement that advancements with this have been hampered by the tools we are using:
  - SPSS macro language is just a string parser;
  - Control SPSS from Python and call R scripts from SPSS, but why not just use the far more capable Python or R independently of SPSS instead?
- Analysts are frustrated by the limited features of the tools we are using in comparison with modern data science tools such as R.
- There's also a huge financial outlay for the organisation of renewing SPSS licenses every few years.



# Why has ISD Scotland chosen R?

- Skills of new graduate recruits – most already have experience of working with R
- Open Source – fits with our requirements under the UK Statistics Code of Practice to produce statistics using transparent and reproducible methods.
- It is a full interpreted programming language, extensible with community-developed packages – significantly more flexible and powerful than SPSS.
- Large, world-wide, established user base and community for technical and methodological support.
- Reduced cost – not free, but substantially lower cost to the organisation than existing SPSS infrastructure.



# How did ISD Scotland decide on R?

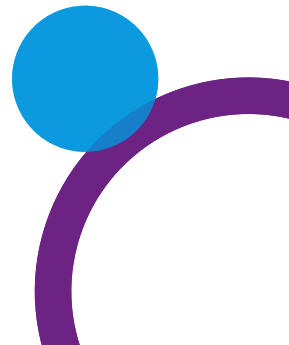
The use of R amongst analysts developed organically:

- Started with a few analysts trying it out on their laptops.
- One analyst secured agreement from our IT department for RStudio Server to be installed on a Virtual Machine as a pilot.
- The pilot was successful and the server's use grew exponentially.



# How did ISD Scotland decide on R?

- We reached a point where the RStudio Server was being used as a production analytical environment, for publications and business-critical functions, despite never being set up for this.
- It quickly became obvious that the infrastructure was not capable of meeting the demand.
- The server became slow and analysts lost faith in its capabilities.
- As a result, the organisation took a few steps backwards in its adoption of R.
  - What was needed was a proper assessment of the infrastructure required to support the analytical demand now and in the future.

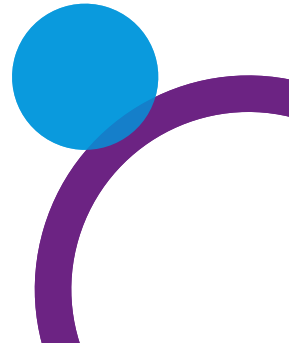





# R User Group

Senior Management recognised the desire from analysts to adopt R as their principle analytical tool:

- An R User Group was therefore established to represent users' strategic, technical and operational interests and aspirations in respect of R.
- The group meets every 3 months to share good practice amongst the user community and acts as a support mechanism for those users requiring assistance.
- Sub-groups are established to focus on specific areas of interest and / or issues.



# R Infrastructure Requirements

- There was recognition by the R User Group and Senior Management that our R infrastructure required formalisation and upgrading.
  - I was tasked by the R User Group to lead a small team (2 analytical colleagues and I) to research, specify and make recommendations on the required infrastructure for a production R analytical environment.
  - It was essential that our recommendations would deliver an analytical environment that would exceed the capacity and capability of our existing SPSS infrastructure, and be able to scale as the organisation grows.
- 

# R Infrastructure Recommendations

## Software for Analysis

### Desktop



- v3.5.1
- Annual x.x.1 upgrades
- Desktop and Server R versions kept in sync

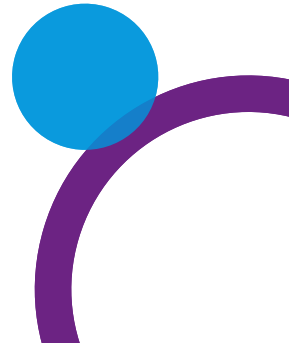
 **Studio<sup>®</sup>**  
Open Source

- Desktop and Server versions of RStudio kept in sync

### Server



 **Studio<sup>®</sup>**  
Server Pro



# R Infrastructure Recommendations

## Benefits of RStudio Server Setup

- Connect from anywhere through web browser and VPN.
- Close to data, faster and more secure.
- More powerful systems, with faster processors and more RAM than can be delivered with standard laptops or desktop PCs.
- Management of packages for all users providing the same analytical environment for all users.



# R Infrastructure Recommendations

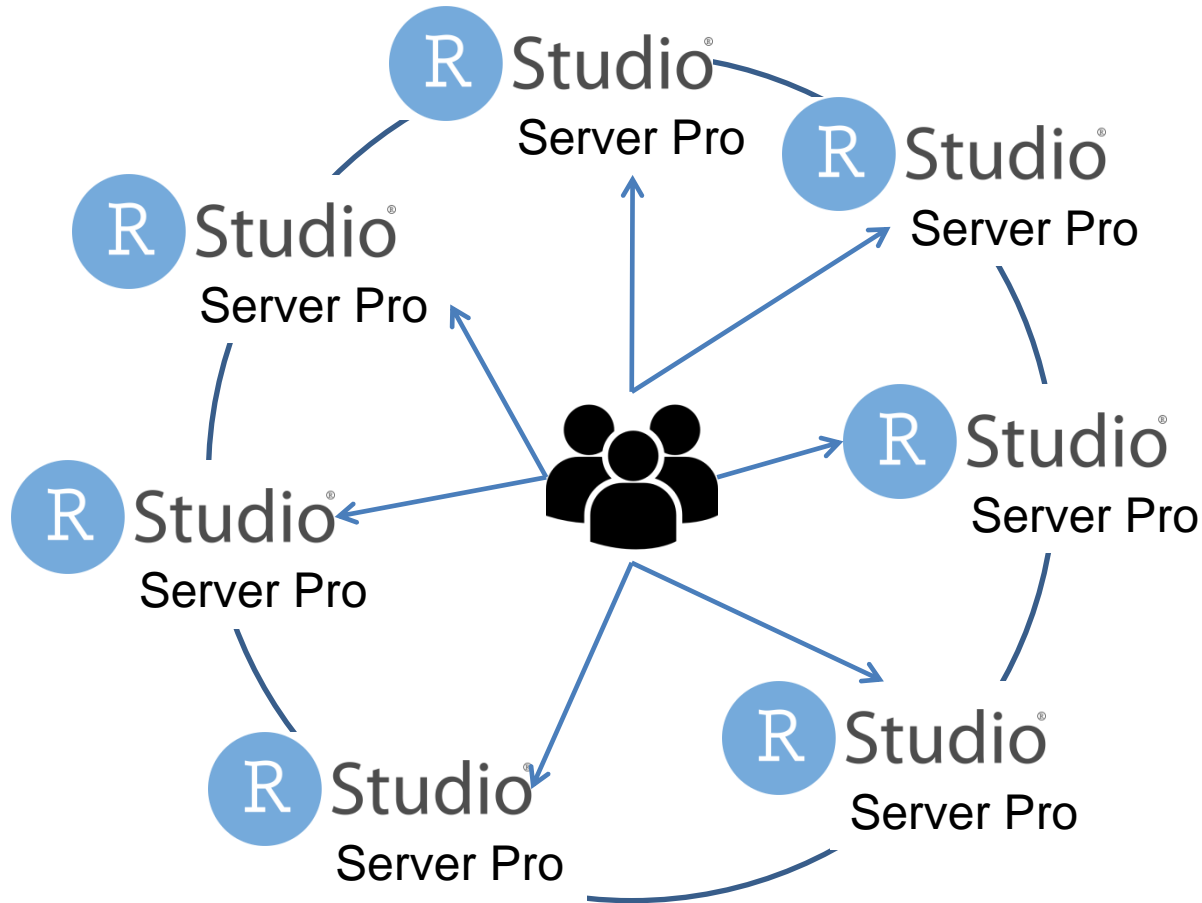
## Why RStudio Server Pro?

- Scalability
  - Load balancer spreads compute demand across cluster of servers
  - New servers can be added to the cluster as demand increases
- Multiple R Versions
  - Installed side-by-side on the same server
  - Accessible on a per session basis
  - Allows older versions of R to be retained to support legacy projects
- Multiple R Sessions
  - Users can have multiple R sessions running concurrently.



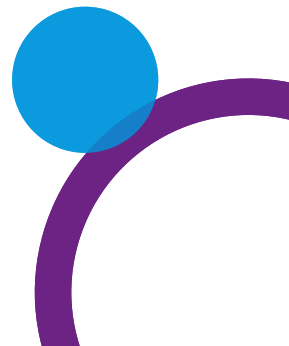
# R Infrastructure Recommendations

## Hardware for Analysis



- 16-core CPU
- 256GB RAM

- Would support 105 concurrent R sessions of mixed size.



# R Infrastructure Recommendations

- We presented these recommendations to the R User Group.
- I was then tasked with leading a large team of 20 analysts to evaluate RStudio Server Pro.
- We carried out structured testing over a 45 day period, focussing on the features unique to the “Pro” version of RStudio Server (excluding load-balancing).
- The tests were all successful, and the analysts were impressed with and excited by the prospect of being able to use RStudio Server Pro for their day-to-day work.
- With testing complete, the green light was given for:
  - 7x RStudio Server Pro licenses
  - Hardware to install those licenses on
  - 1x RStudio Package Manager license



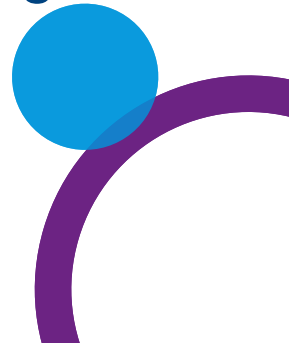
# Planned Implementation

## Hardware for Analysis



- 2x AMD EPYC 7000 series server CPU
  - Each CPU is 32-core and supports 64 threads
- As R is single-threaded, the server would support up to 128x R sessions before blocking would become an issue.
- Configured with 4TB of RAM, dividing equally would give each session 32GB of RAM.

*Ignoring OS, RStudio Server and other overheads!*





# Planned Implementation

## Hardware for Analysis



R Studio<sup>®</sup>  
Server Pro



R Studio<sup>®</sup>  
Server Pro



R Studio<sup>®</sup>  
Server Pro



R Studio<sup>®</sup>  
Server Pro



R Studio<sup>®</sup>  
Server Pro



# Planned Implementation

## Additional Infrastructure



Hypervisor



# Next Steps and Future Plans

- We currently do not have a group with responsibility for the delivery of the new, production R environment, and this has been one of the reasons for the delay in implementation.
- It is recognised that the establishment of this group is key to the successful delivery and adoption of R across the organisation.
- The group will also need to develop and put in place business processes for, and assign resource to testing new versions of R and approving new versions of packages.
- Lead on the organisation-wide adoption of Version Control using Git and Gitea.
- Migrate data visualisations and dashboards from Tableau to R Shiny.



Contact Details:

Email: [terry.mclaughlin@nhs.net](mailto:terry.mclaughlin@nhs.net)

Web: [www.isdscotland.org](http://www.isdscotland.org)



# Questions & Comments?

**Terry McLaughlin**  
Principal Information Analyst, Consultancy Services

[Information Services Division](#)