



Driving R adoption in an NHS information service: barriers and solutions

Chris Mainey

Intelligence Analyst
University Hospitals Birmingham NHS Foundation Trust

chris.maine@uhb.nhs.uk



Introduction





University Hospitals Birmingham NHS FT

- Second largest Acute-Hospital Trust
- Formed by merger of former UHB and Heart of England trusts
- Four hospitals seeing 440,000 in-patients per year

- Large informatics team(s):
 - Core Information
 - SQL / Dashboard developers
 - Analytics
 - Research
 - Commercial



Benchmarking Platforms

Healthcare Evaluation Data (HED)

www.hed.nhs.uk

- Online hospital benchmarking system
- Statistical models and analysis tools
- Training and support
- Used by ~60 NHS and other organisations





Data analysis tools in the NHS

- Excel and Access were baseline 10-years ago
- 'Complicated' Database work conducted by IT department
- Some Trusts sent regular extracts from system suppliers
- Increasing use of Qlik, PowerBI and Tableau recently

UHB:

- Recruited SQL developers and slowly moved department to SQL Server
- Excel, SQL Server Reporting Services and TIBCO Spotfire used for outputs



Keenies...

- R was used only for rare 'statistical' calculations
- A few keen medical student or researchers
- Conserved pattern at other Trusts
- The case in our team...

One keen user...





Allowed space to develop skills:

- My own PhD project, funded by UHB, on statistical modelling in NHS.
- Chose to use R from the outset, rather than SAS or Stata.
- Python is good candidate too, but harder to find training when I started
- Developed statistical, visualisation and data manipulation skills
- Moved some HED models to R
- Supported other team members to use it
- This generated a demand for basic training



Followed established patterns:

- Set up a local user group
- Trained senior team
- Encouraged new starters to use it
 - Plan work that requires it
- External relationships:
 - Twitter has been key
 - NHS-R Community: <https://nhsrcommunity.com/>
 - External R user groups
 - Regional relationships: Other Trusts, PHE, City Council
 - Advice from Mango, RStudio and others





Barriers

- **Open Source fears:** IT teams sometimes reluctant to install
- **Wrestling with default settings:**
 - Mapped user folder that aren't writable
 - Can't write to package library, or temp files
- **Poor hardware**
 - Mostly desktop users at hospitals
 - Defaults can be low on disk space with minimal RAM
- **Lack of understanding of fundamentals:**
 - Networking/file locations
 - Suspensions around security
 - R coding or application to problems



Solutions

- **Permissions:** *Explicitly identify default settings*
 - Set a specific library, using `R_libs_user` variable
 - Set temporary directories using `.Renv` file
- **Hardware:**
 - Make the case for more RAM! It doesn't cost much
 - Consider server approaches: VMs or 'Desktop' on a 'Server'
 - RStudio Server / RStudio Connect
- **Skill/Awareness:**
 - Allow space to develop knowledge
 - Windows users often don't know file paths, SSH etc.
 - Central guidance on R/Python implementation





How do we use it?

1. Regression, and related models (`mgcv`, `lme4`, `glmmTMB`, `glmnet`)
2. Machine Learning technique (`caret`, `randomForest`, `gbm`)
3. Optimisation problems (solving equations using `nloptr`)
4. Summary stats and Exploration
5. RMarkdown Reports and Presentation
6. API interaction (NHS ODS service, CQC register organisations)



Initial adoption:

```

74 base_table_1718[,PROCOD3]=as.factor(PROCOD3))
75 base_table_1718[,donl_dflag_chap]=as.factor(donl_dflag_chap))
76 base_table_1718[,LOS_PQ]=as.factor(ifelse(LOS==4, 'LOS_5_plus', paste0("LOS_",LOS)))
77
78 #use this to claim back memory
79 gc()
80
81
82
83 # 4. This code sets up the parallel structures.
84 ## If doing this on a laptop, don't run section 5, and comment out the
85 ## set up cluster and register with 'doparallel', an implicit parallel function
86 ## If doing this in Microsoft R open (MRO), run all 'mc' commands. MRO parallelises silently and gets confused
87 ## explicitly parallelised like we are here. i.e. on four cores, you want 4 single threads, not 4 x 4 threaded
88 ## If doing this in 'regular' R, skip over or comment out the 'mc' commands.
89
90 getMCthreads() #do not use in 'regular' R, use if running Microsoft R open (faster version)
91 setMCthreads(1) #do not use in 'regular' R, use if running Microsoft R open (faster version)
92
93 #make an register clusters
94 cl=makeCluster(1)
95 registerDoParallel(cl)
96
97 #set clusters to single thread each if required
98 clusterEvalQ(cl, setMCthreads(1)) #do not use in 'regular' R, use if running Microsoft R open (faster version).
99 clusterEvalQ(cl, getMCthreads()) #do not use in 'regular' R, use if running Microsoft R open (faster version)
100
101 #Package collect on cluster
102 clusterEvalQ(cl, gc())
103
104
105
106 # 5. work out HRO4 chapters for iteration
107 chapters=distinct(base_table_1718,HRO4C_LETTER) %>% as.data.frame()
108 chapters=t(chapters) %>% as.vector()
109
110
111 # 6. Set up blank data frame for AUC data
112 AUC_gms<-data.frame(HRO4C_LETTER= NULL, AUC=NULL)
113
114
115
116 #7. set start time and run loop iterating through
117 start<-Sys.time()
118
119 #start of loop
120-
121-
122-
  
```

- Collections of R scripts saved to directory
 - Required lead analyst to write
 - Required team skills to run and/or debug
- Constant debugging:
 - Data changing
 - Littered with comment, hard to decipher





Creating a local package

Wrapped functions in to an R package structure

- Built binary and save to file structure
- Use roxygen2 to write your support material as you define functions

The screenshot shows the RStudio interface with the 'HED-Functions' package. The left pane displays the source code for the 'funnel_plot' function, which includes comments and R code for handling predictions, observed values, and model parameters. The right pane shows the package documentation, including the title 'HED-Functions', version '0.2.3', and a list of help pages such as 'buildFormulaTable', 'buildFormula', 'buildModelFile', 'bulkSQLInsert', 'exportModelOutputs', 'funnelPlot', 'getFormulaParameters', 'getRRLOSBaseTable', 'minimizeModelReadmission', 'runRRLOS', 'runRRReadmission', and 'writeCSV'.



Source control

- Moved to source control
 - Install git for windows
 - Git learning curve
 - Happy Git with R!
 - Git repo on TFS
 - Figure out https connections
 - Stef Locke's tfsr
- <https://happygitwithr.com/>
- <https://github.com/lockedata/tfsR>

The screenshot shows the Visual Studio Team Foundation Server interface for the 'HEDAnalytics' repository. The left pane shows the file explorer with folders like 'man', 'R', and files like '.gitignore', '.Rbuildignore', 'DESCRIPTION', 'NAMESPACE', 'RunningRRLOS.R', and 'RunningRRReadmission.R'. The right pane shows the commit history for the 'master' branch, listing commits with their dates and descriptions, such as 'PR 4: Merge ImproveRRReadmission to master - Chris Mainey' and 'sorted funnel plot, bumped version number to force update - Chris'.





NHS-R community

- Funded by grant from the Health Foundation
- Aimed to support and promote R use
- Training provision
- Blog posts
- Discussion and dissemination of best practice
- Development of tools
- National Conference



What's next?

- Continuing to develop analyst skills in R
- Contributing to, or building Open Source tools with NHS-R community
 - **FunnelPlotR**: <https://github.com/chrismainey/FunnelPlotR>
 - **NHSRdatasets**: <https://github.com/nhs-r-community/NHSRdatasets>
- Wider adoption of Git/Git Hub
- Pushing further into ML techniques for prediction
 - Improving current models: Ensemble techniques etc.
 - Unsupervised techniques for cluster analysis





Thanks for your time!

✉ chris.mainey@uhb.nhs.uk

🌐 <http://www.hed.nhs.uk>

🐦 [@chrismainey](https://twitter.com/chrismainey)

👤 [chrismainey](https://github.com/chrismainey)

🌐 <http://www.mainard.co.uk>

🆔 <https://orcid.org/0000-0002-3018-6171>



Health Informatics
Information. Analysis. Statistics. Reporting.