



PD01: My Journey from SAS to R

Lucy Watson

PHUSE EU Connect 2022, Belfast

Agenda

1
My coding history

2
SAS vs R

3
Challenges

4
Solutions

5
Summary & what next

1. My coding history

University

Applied Mathematics
and Physics

MATLAB, C++

2nd Job

Assistant Statistician,
NISRA

Moved to SPSS

Now

Principal Programmer,
GSK

SAS & R

1st Job

Research Officer, Office
for National Statistics
Introduced to SAS

3rd Job

Snr Programmer, Exploristics
Back to SAS (yay!)
First introduction to R

2. SAS vs R

SAS

- ☐ Licensed commercial software with a cost
- ☐ Dedicated customer support
- ☐ Less frequently updated

R

- ☐ Open source and free
- ☐ Large online community
- ☐ Continuously updated

3. Challenges

- ❑ Questioning necessity of new language
- ❑ Lack of opportunity / time
- ❑ Comparisons with SAS
- ❑ Assumptions made from SAS knowledge

Challenges - Examples

Midpoint rounding issues

Numvar	SAS	R
0	0	0
0.5	1	0
1.5	2	2
7.5	8	8
8.5	9	8

Let's round Numvar to an integer...

SAS => `round(numvar, 1)`

R => `round(numvar, 0)`

SAS rounds the midpoint **away from 0**

R rounds to the **nearest even number**

Solution: New function created to match SAS

Challenges - Examples

Sorting differences – numeric values

```
SAS => proc sort;  
        by numvar;  
run;
```

```
R1 => arrange (Numvar)  
R1 => order (df$Numvar)
```

Numvar	SAS	R1	R2
0	.	0	NA
10	0	1	0
1	1	2	1
2	2	9	2
.	9	10	9
9	10	NA	10

Potential 'Solution': R2 => order (df\$Numvar, **na.last = FALSE**)

Challenges - Examples

Missing data

SAS

- ❑ Missing represented as . (num) & '' (char) [.a-z, . _ for special]
- ❑ Treated as smaller than all numbers / printable character values
- ❑ Handled differently in various procedures

R

- ❑ Missing represented as NA
- ❑ NA is an indicator of missingness (not string or numeric)
- ❑ NA options: na.actions, na.omit, is.na, na.last, na.rm, useNA

Numvar	SAS		R		
	$x < 0$	$x < 0$	$x < 0$	$x == \text{NA}$	is.na(x)
0	N	N	FALSE	NA	FALSE
1	N	N	FALSE	NA	FALSE
.	Y	Y	NA	NA	TRUE
5	N	N	FALSE	NA	FALSE

4. Solutions / Ways of working

1. R Adoption at GSK
2. R Code Along
3. R Code Corners
4. R Book Club



➤ R Adoption at GSK

- Project designed to integrate R into clinical reporting
- Guidance documents and training courses
- R code along, R code corners, R Book Club
- External collaborations (e.g. pharmaverse)
- Dedicated support team



R Code Along

- 1-2 hr hands on class with support
- Common displays used as examples
- Good basis for starting project work

R Code Corners



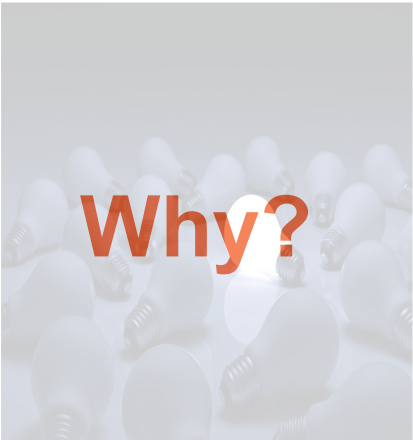
What?

R challenges on the algorithms we use in our daily programming



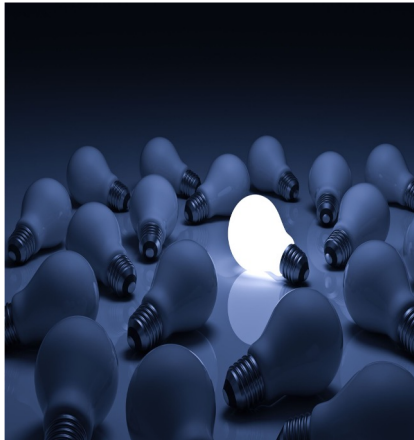
When?

Every 1-2 months with 1-2 weeks to submit



Why?

To build expertise and bridge gap between training and application



Examples
&
Benefits

R Code Corners - Example

Calculating Big N counts for each treatment is a common derivation. While SAS makes use of *Proc FREQ/SQL* to do so, how can we do this in R?

The challenge:

- ❑ Read in ADSL dataset provided
- ❑ Count the frequency of treatment (TRT01A) based on Safety population
- ❑ Create a new column “LABEL”: Concatenate TRT01A values and N with frequency counts calculated in above step
- ❑ Display the data frame

Sample output as follows:

TRT01AN	TRT01A	LABEL
1	Active	Active (N=XX)
2	Placebo	Placebo (N=XX)

Key Learnings:

- Loading libraries
- Reading data
- Sub-setting data
- Using summary functions
- Using string functions
- Codes reviewed for GPP: commenting, indentation, headers

Now co-leading this initiative!

➤ R Book Club

- Books selected at different levels / interests
- Weekly/bi-weekly meetings
- Groups kept small (<15)
- Take turns leading sessions

Example books:

- R for Data Science
(Hadley Wickham, Garrett Grolemund)
- R Packages (2e)
(Hadley Wickham, Jenny Bryan)



5. Summary & what next...

- ❑ Differences in SAS vs R
- ❑ Challenges of moving to a new language (+ examples)
- ❑ Support I received from GSK
- ❑ Solutions I adopted on my R journey

And next ... Python??



thank you!

go raibh maith agaibh

Contact information

Name:	Lucy Watson
Job title:	Principal Programmer
Email:	lucy.x.watson@gsk.com