The lazy and easily distracted report writer

Mike K Smith (Pfizer R&D UK Ltd.)
@MikeKSmith

RStudio::conf(2019)

TL;DR - Disclaimer

- I used rmarkdown notebooks / documents to write up an exploratory analysis to share with a drug development team.
  - Statistician
  - Clinical Pharmacologist (including my manager)
  - Clinician
- The analysis presented here is ***NOT*** that analysis (for confidentiality) but it has similar attributes.

Your (my) brain is lazy, shallow, and easily distracted.

https://www.slideshare.net/CJAtherton/chris-atherton-at-presentation-camp-london

Cutlery drawers & what they say about YOU

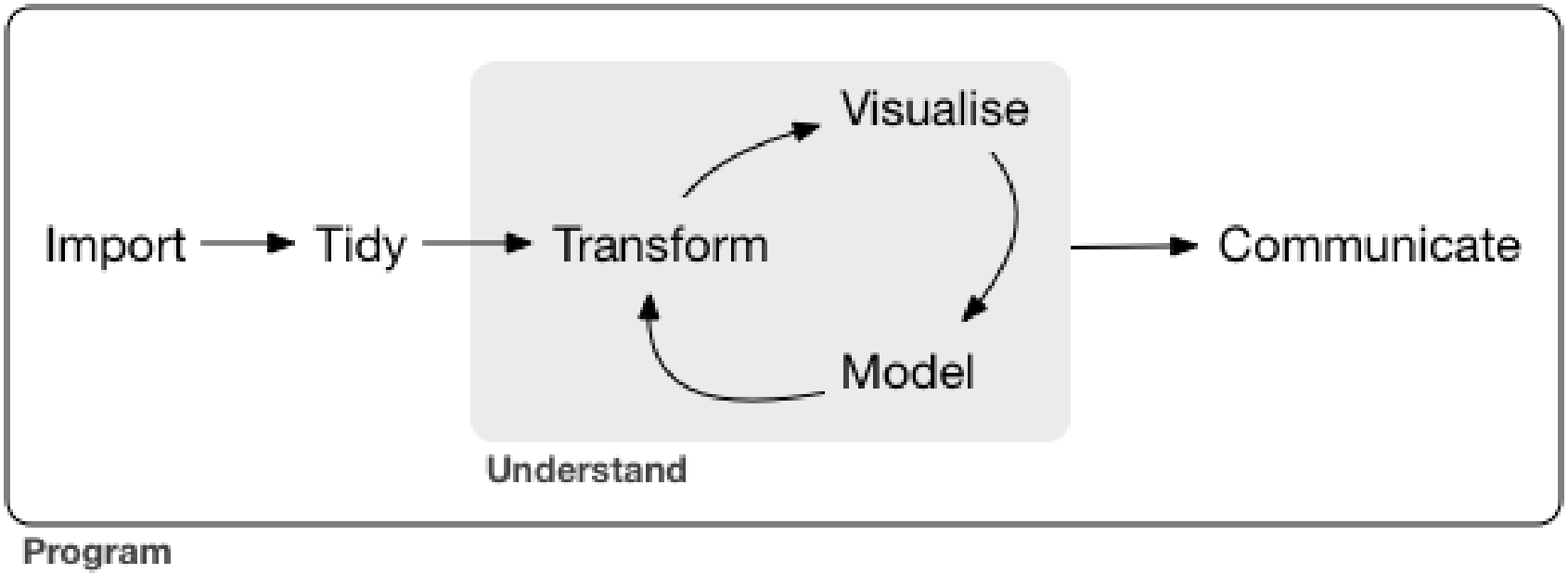
HT: @HadleyWickham, @jimhester_, @dataandme

Mine… (sorry / not sorry)

HOME ORGANIZATION TIP:
JUST GIVE UP.

https://xkcd.com/1077/

Data analysis - THEORY



https://r4ds.had.co.nz/

Data Analysis – In practice….

(***DISCLAIMER***: I'm ***sure*** the experiences recounted here are ***unique to me alone.***)

Go to email with link to data source…

read and respond to 3 other emails…

Download and read data into R…

stop and answer colleague's question(s) about the tidyverse…

Wrangle data and plot it…

LUNCH

go to an (unrelated) meeting / teleconference call.

Make better plots.

follow an interesting link that Mara Averick (@dataandme) just posted on Twitter

Fit preliminary model to data.

<Next day>

Team find problem with data,
share new version of data.

Change input data and redo analysis.

Check new version against
previous version.

Discuss findings with my boss.

file expenses.

Circulate report.

DONE!!!

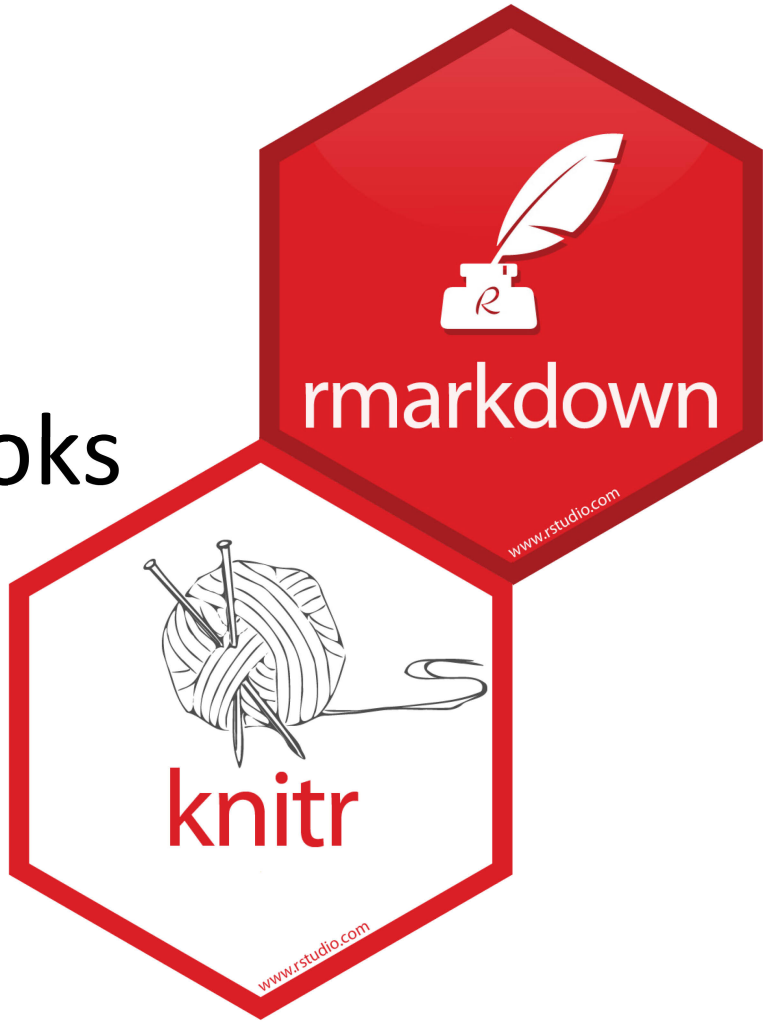
< 6 months pass >

Review comments come back…

Wait… Erm… ***WHAT*** was I thinking?

To the rescue…

rmarkdown & notebooks



Who is your audience?

- Present (distracted) ***me***
- Future (6 months later) ***me***
- Quantitative colleagues / reviewers
- Decision makers (may not be quantitative)

Notebooks / markdown ***vs*** scripts ***(for analysis)***

BUT, see also: https://yihui.name/en/2018/09/notebook-war/

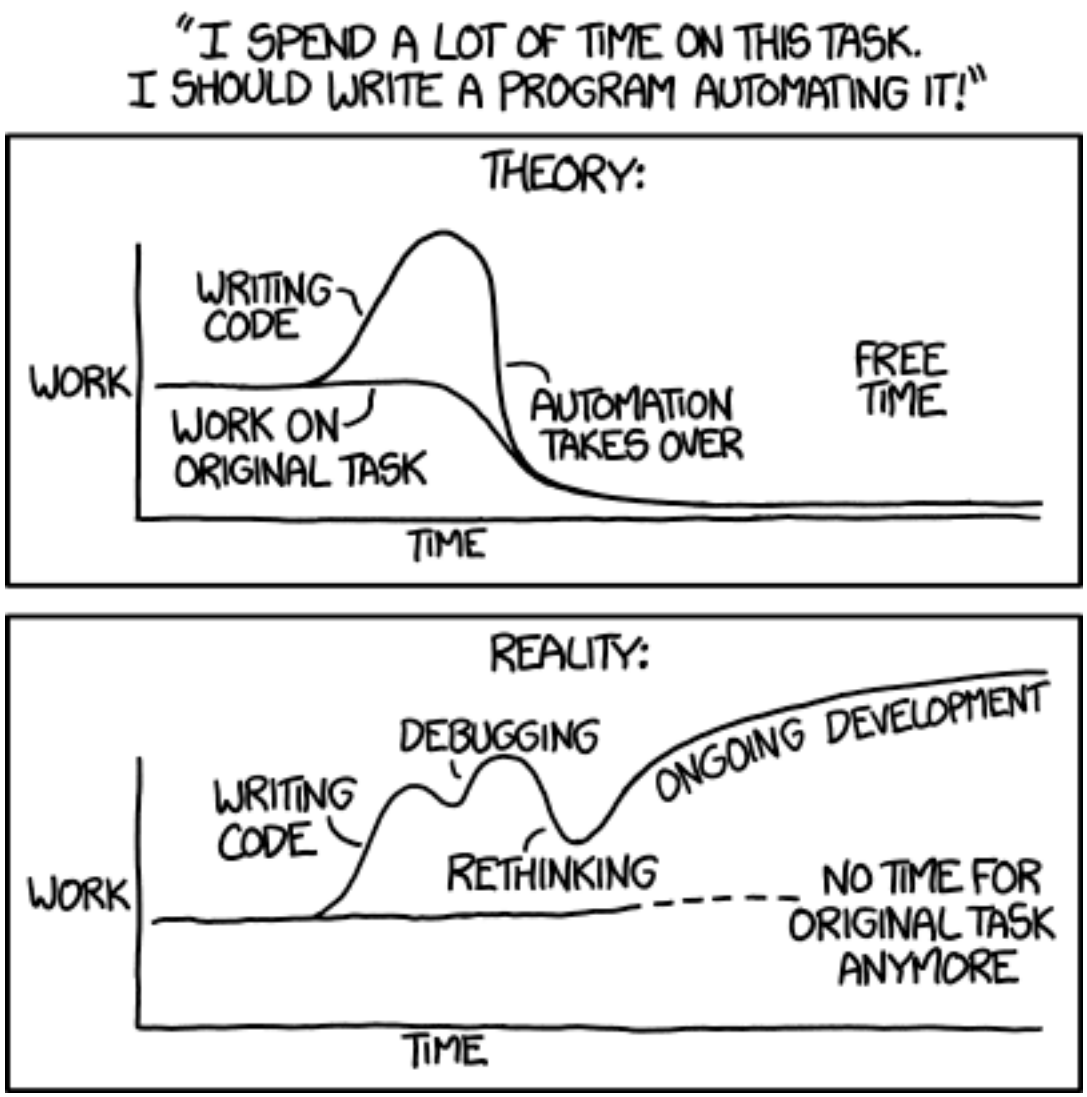
Also…

I ***knew*** my manager / other reviewers would ask for reports on the ***THREE*** different endpoints.

Rule of three

- Copy & paste code ≥3 times?
  - Write and use a function

- Perform analysis across ≥ 3 endpoints?
  - Multiple markdown reports?
  - ***NOPE***. ***Parameterised*** reports.



https://xkcd.com/1319/

YAML header parameters

My Parameterised report with child doc... | ModelDiagnostics_text.Rmd | DataManipulation_text.Rmd

Knit | Insert | Run

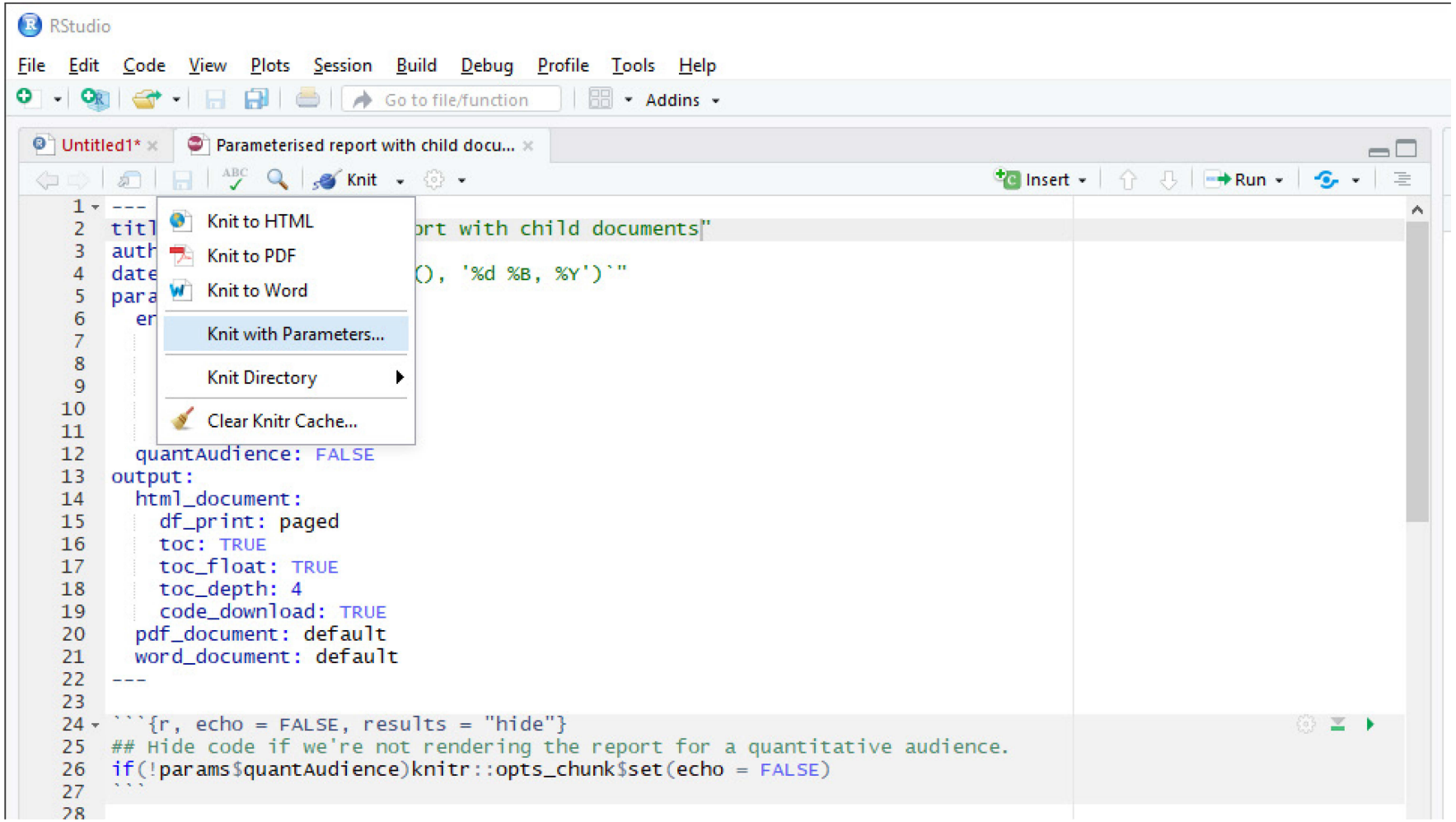
```
1  ---
2  title: "Parameterised Report with child documents"
3  author: "Mike K Smith"
4  date: "`r format(Sys.time(), '%d %B, %Y')`"
5  params:
6    endpoint:
7      value: HAMDTL17
8      choices:
9        - HAMDTL17
10       - HAMATOTL
11       - PGIIMP
12   quantAudience: FALSE
13 output:
14   html_document:
15     df_print: paged
16     toc: TRUE
17     toc_float: TRUE
18     toc_depth: 4
19     code_download: TRUE
20   pdf_document: default
21   word_document: default
22 ---
```

Parameters that can be used in the code / knitr options.

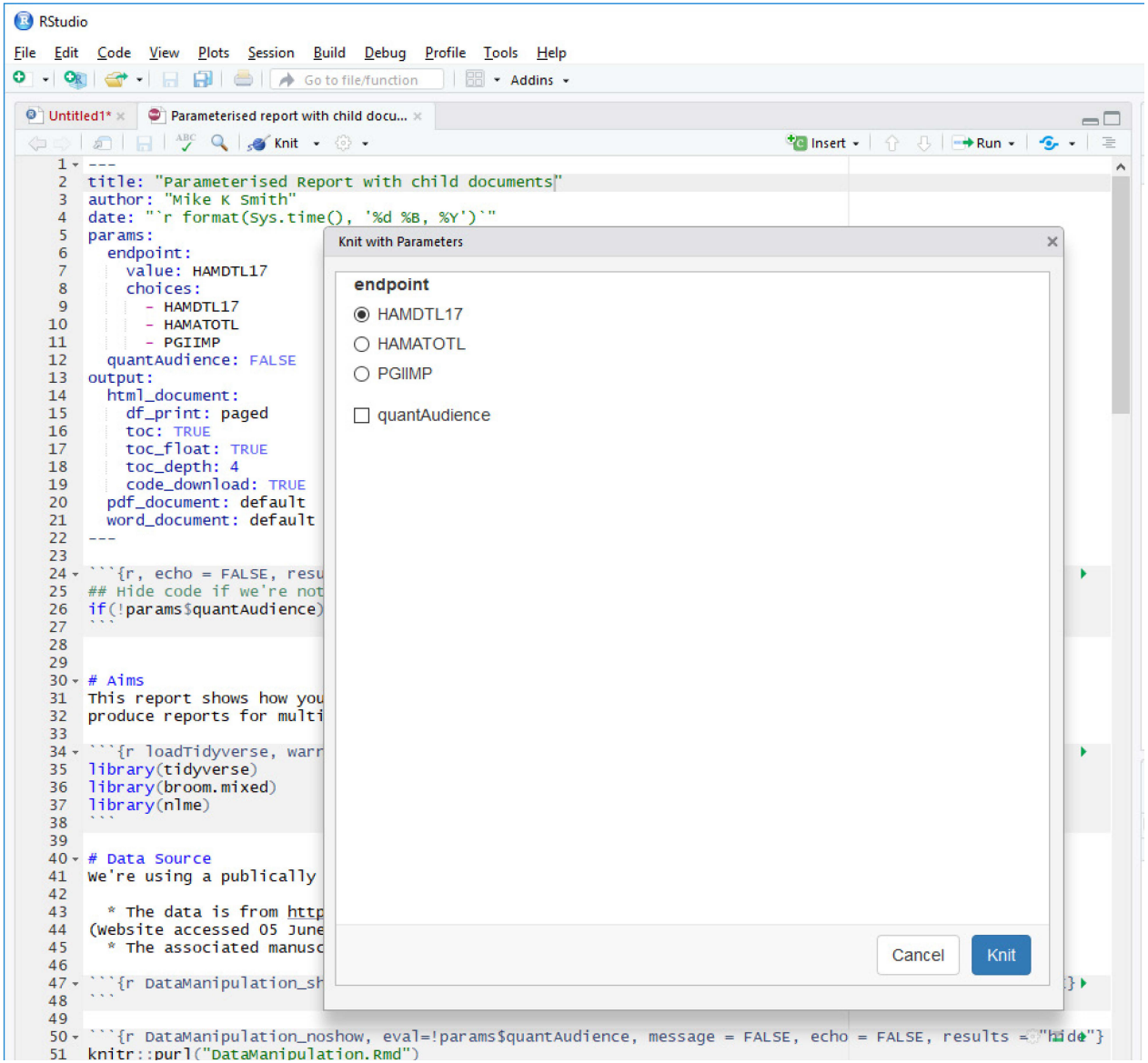
Note that endpoint has only three choices, and a default value (HAMDTL17)

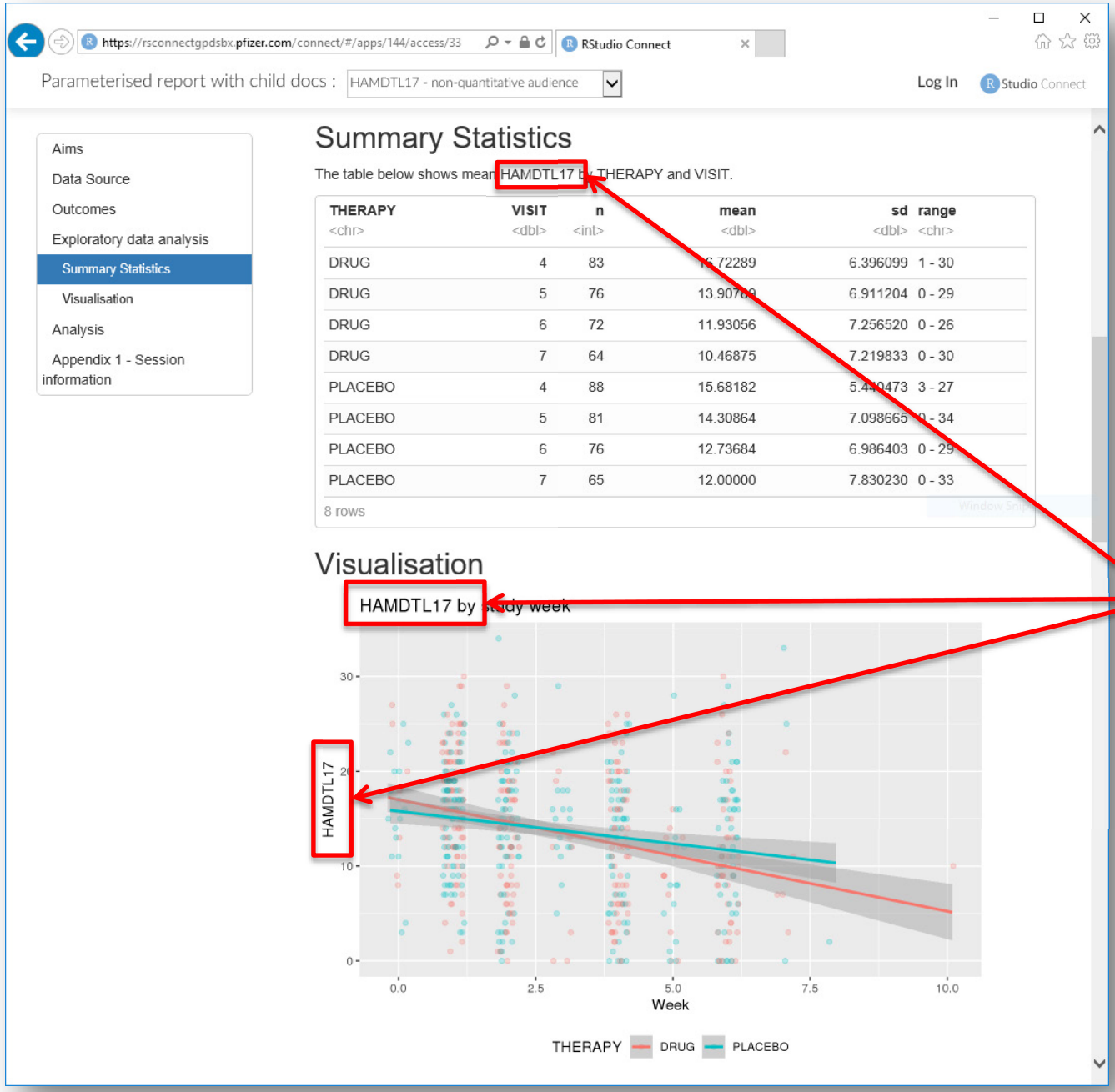
NB. A bit like Shiny inputs

Render with parameters



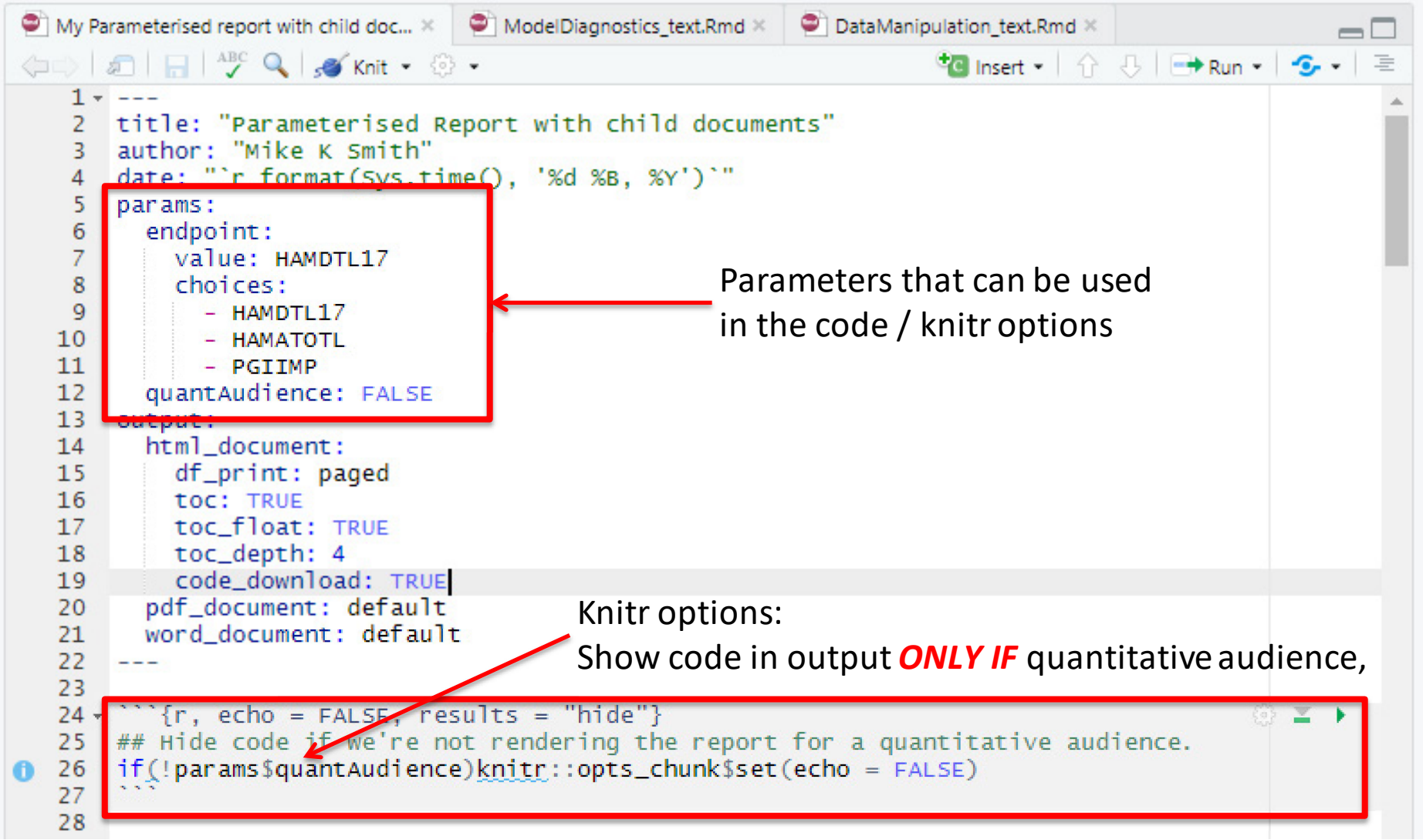
Render with parameters






```
params:
  endpoint: HAMDTL17
  quantAudience: FALSE
```

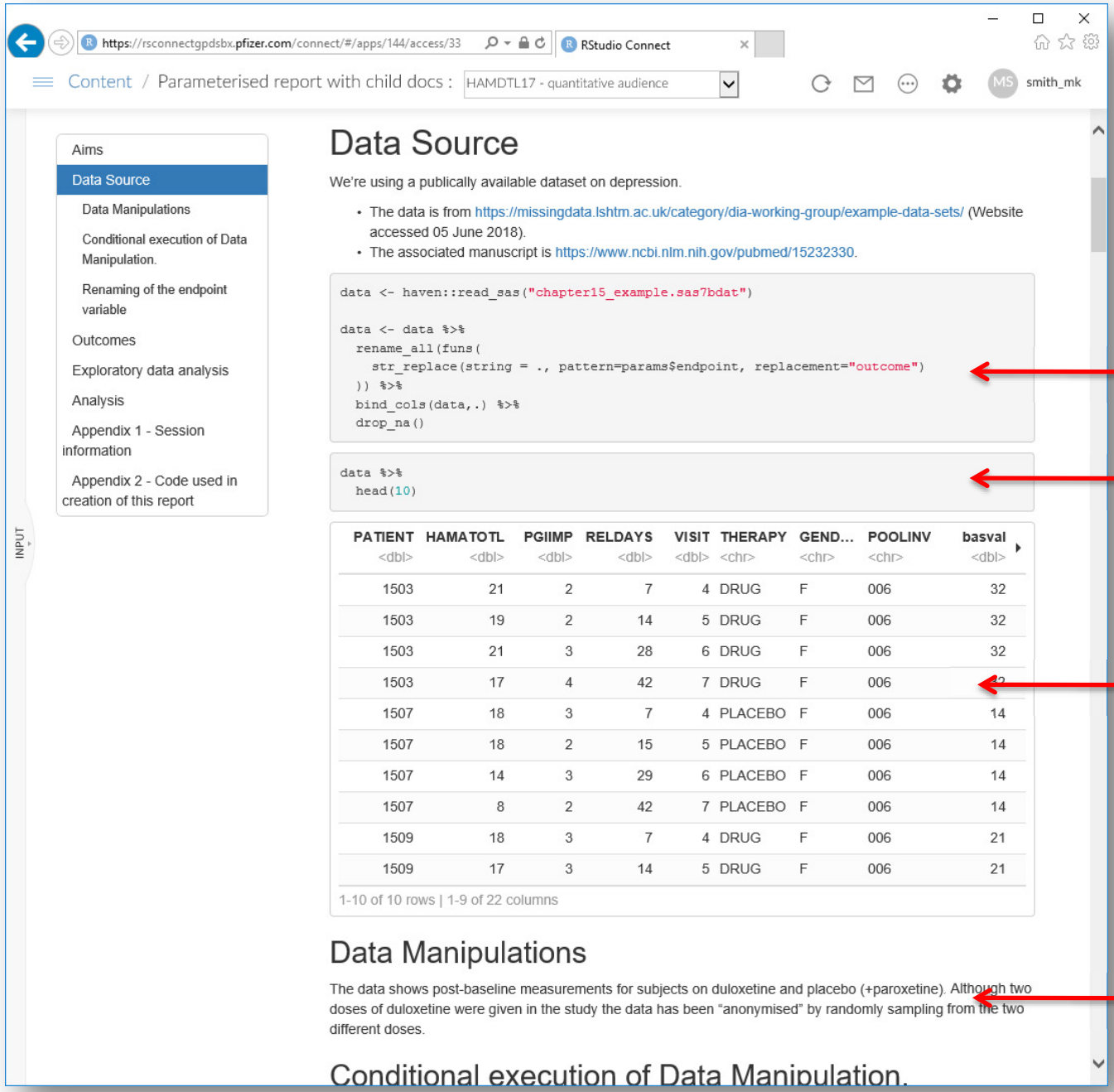
Using **`params$endpoint`** in markdown text, plot code


My Parameterised report with child doc...
ModelDiagnostics_text.Rmd
DataManipulation_text.Rmd
Knit
Insert
Run
---
title: "Parameterised Report with child documents"
author: "Mike K Smith"
date: "`r format(Sys.time(), '%d %B, %Y')`"
params:
endpoint:
value: HAMDTL17
choices:
- HAMDTL17
- HAMATOTL
- PGIIMP
quantAudience: FALSE
output:
html_document:
df_print: paged
toc: TRUE
toc_float: TRUE
toc_depth: 4
code_download: TRUE
pdf_document: default
word_document: default

```{r, echo = FALSE, results = "hide"}
## Hide code if we're not rendering the report for a quantitative audience.
if(!params$quantAudience)knitr::opts_chunk$set(echo = FALSE)
```
Parameters that can be used in the code / knitr options
Knitr options:
Show code in output ***ONLY IF*** quantitative audience,

Rename endpoint variable(s) to “outcome” (simplifies later code)

```
```{r DataManipulation, results="hide", message=FALSE, warnings=FALSE}
data <- haven::read_sas("chapter15_example.sas7bdat")

data <- data %>%
  rename_all(funs(
    str_replace(string = ., pattern=params$endpoint, replacement="outcome")
  )) %>%
  bind_cols(data,.) %>%
  drop_na()
```
```

Run this code ***ONLY IF*** **`params$quantAudience = TRUE`**

```
```{r Show_data, eval = params$quantAudience}
data %>%
  head(10)
```
```

Run ***ONLY IF*** **`params$quantAudience = TRUE`**, to pull in text ***from child document***

```
```{r DataManipulationChildDoc, eval=params$quantAudience,
child="DataManipulation_text.Rmd"}
```
```



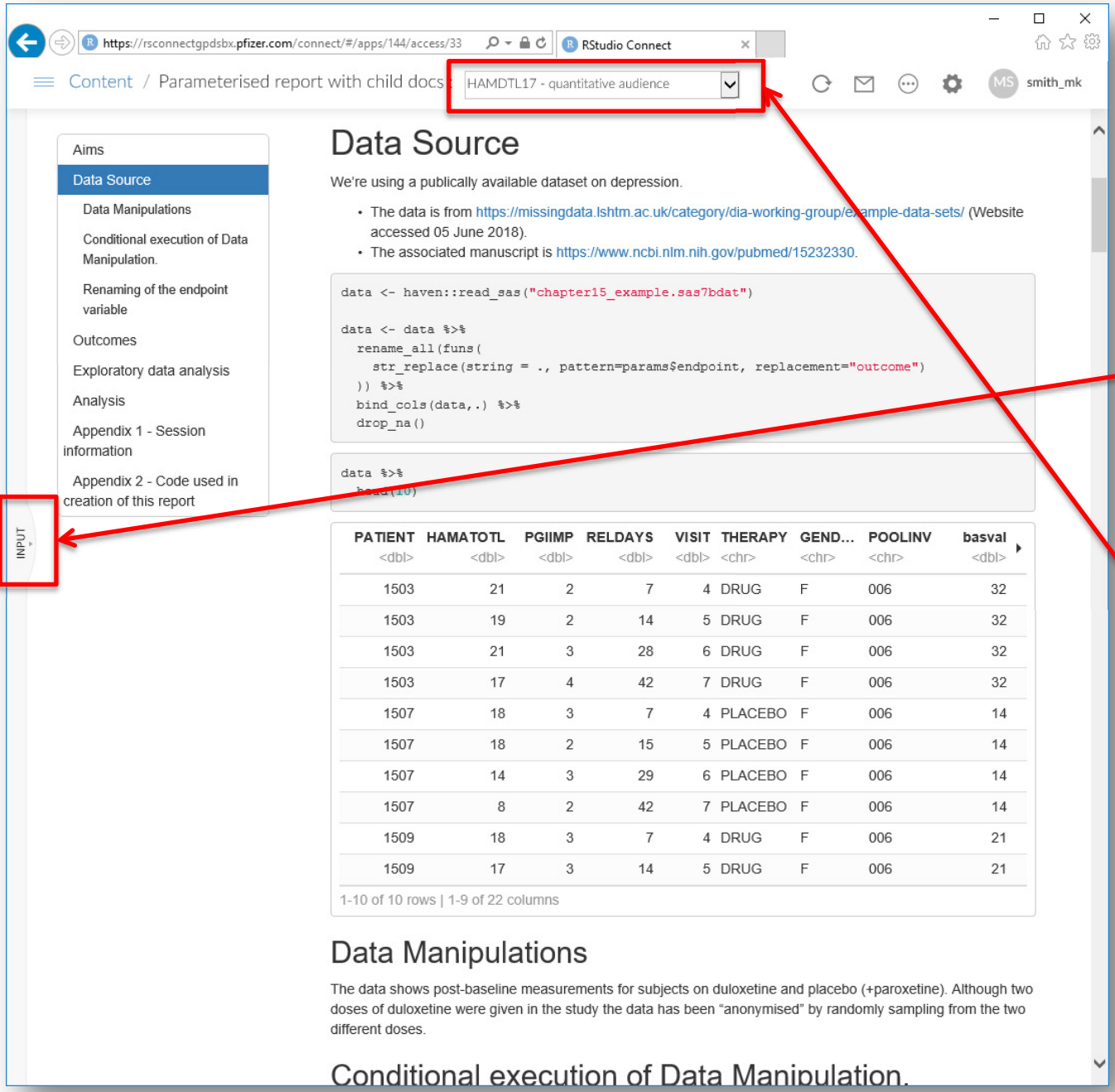

```
params:
  endpoint: HAMDTL17
  quantAudience: TRUE
```

Code shown

Chunk run

Data shown

Child doc text




```
params:
  endpoint: HAMDTL17
  quantAudience: TRUE
```

RStudio Connect allows you (or visitor to your page) to specify parameters and render a parameterised report and then to save that report as a named item.

You can then have pre-rendered reports for various audiences ready to go…

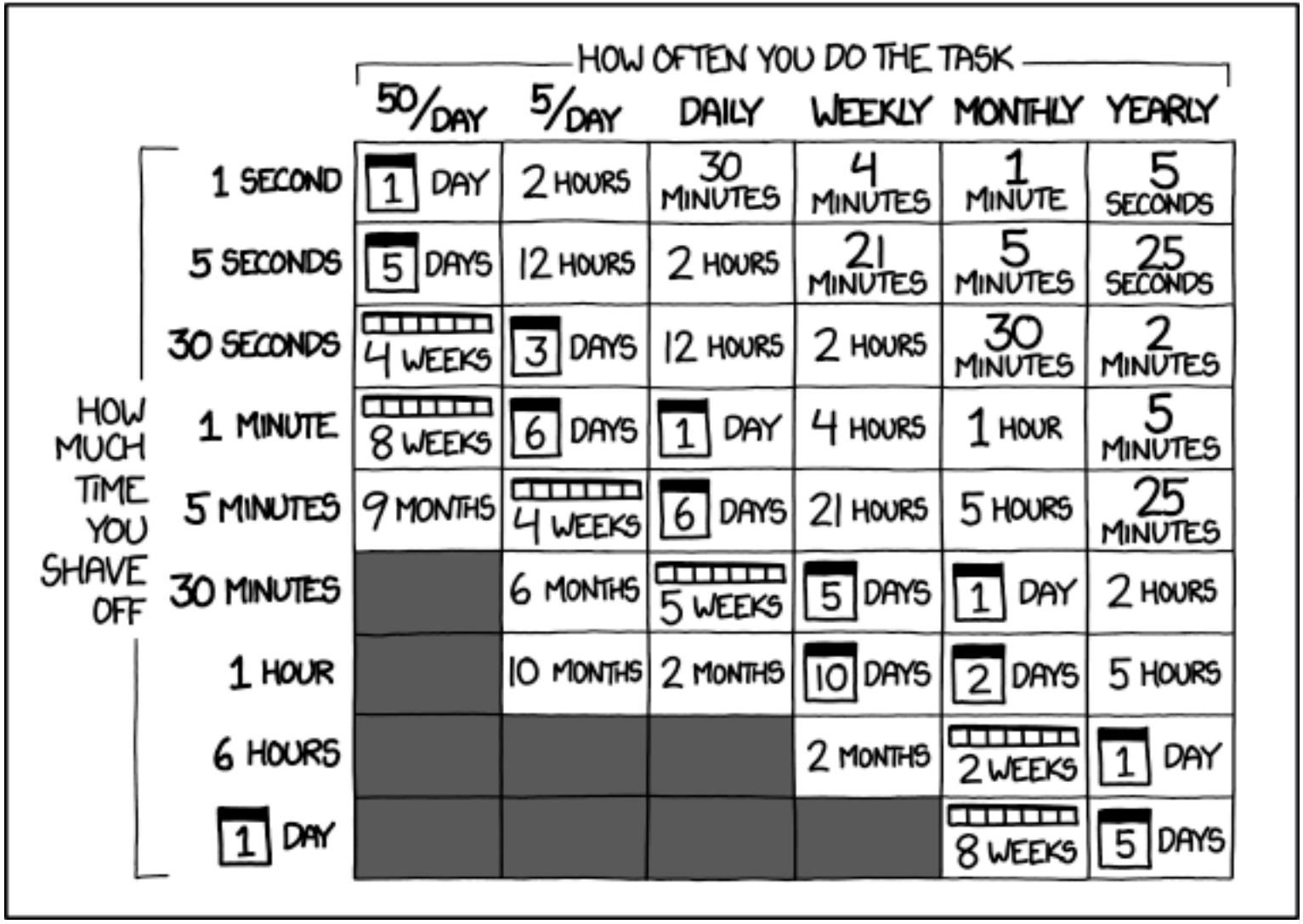
More parameterisation

- Question: how to show correct analysis for non-continuous endpoint?
  - Change analysis type in code depending on **`params$endpoint`**.

- Question: what to do if something goes wrong in the analysis?
  - Check for errors and handle appropriately using **`tryCatch(…)`**
  - Insert child document text: "***EMERGENCY!*** Something has gone wrong… Contact your data scientist!"

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?

(ACROSS FIVE YEARS)



https://xkcd.com/1205/

Feel free to ask me questions, but remember….