



RStudio Projects: Making project QA less of a project

Rachel Naylor
Modelling and Analysis Officer



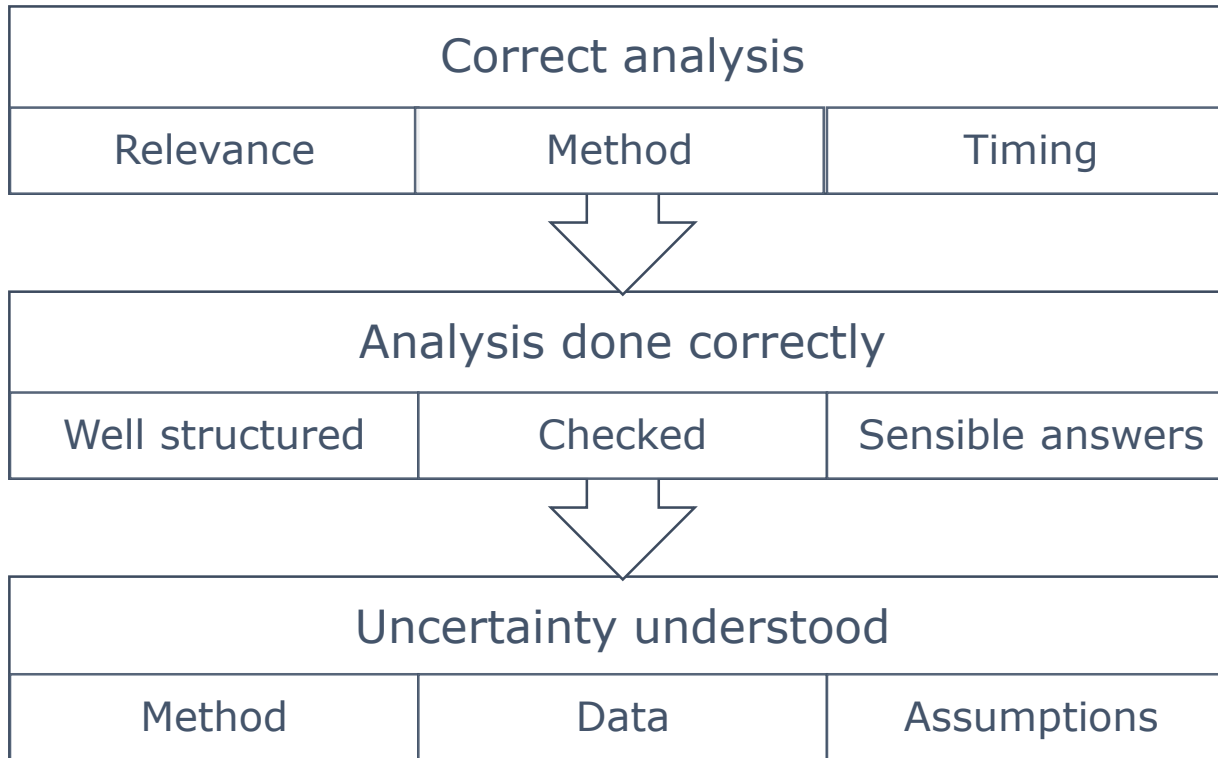


Quality Assurance

“If analysis and any supporting models, data and assumptions are not fit-for-purpose then the consequences can be severe ranging from financial loss through to reputational damage and legal challenge. In the most severe of consequences, lives and livelihoods can be affected.” – The Aqua Book



What is quality?





RStudio Projects

- Help you keep all your files together
- Give you a single working directory for the project
- Give you a separate R environment for the project
- Allow you to restore your RStudio set-up





Using RStudio Projects

The screenshot displays the RStudio interface for a project named "Example Project". The main editor shows an R script with the following content:

```
1 # Header -----
2 # Title: Example RStudio Project script
3 # Author: Rachel Naylor
4 # Date: 09/05/2018
5
6 # Packages -----
7 library(tidyverse)
8
9 # Data -----
10 iris_data <- read_csv("iris_data copy.csv")
11
12 # Plots -----
13
14 iris_data %>%
15   ggplot(aes(sepal.Length, sepal.width)) +
16   geom_point(aes(colour = factor(species))) +
17   scale_colour_manual(values=c("#433E91", "#188700", "#E05499"), name = "Species") +
18   labs(x = "Sepal length", y = "Sepal width", title = "Iris Sepal length-width relationship by species")
19
20 ggsave("iris sepal width length plot.pdf")
21
```

The Environment pane on the right shows the "Global Environment" and a "Data" section containing the variable "iris_data" with 150 observations and 5 variables. The Files pane at the bottom shows the project structure:

| Name | Size | Modified |
|----------------------------------|--------|-----------------------|
| .. | | |
| .RData | 3.8 KB | May 11, 2018, 9:57 AM |
| .Rhistory | 1.6 KB | May 11, 2018, 9:57 AM |
| Example Project.Rproj | 218 B | May 11, 2018, 9:57 AM |
| Example script.R | 813 B | May 11, 2018, 9:57 AM |
| iris_data copy.csv | 3.6 KB | May 11, 2018, 9:20 AM |
| iris sepal width length plot.pdf | 5.5 KB | May 11, 2018, 9:58 AM |

The Console pane at the bottom shows the execution output:

```
v readr 1.1.1 v forcats 0.3.0
> conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
>
> # Data -----
> iris_data <- read_csv("iris_data copy.csv")
Parsed with column specification:
cols(
  sepal.Length = col_double(),
  sepal.width = col_double(),
  petal.Length = col_double(),
  petal.width = col_double(),
  species = col_character()
)
>
> # Plots -----
>
> iris_data %>%
+ ggplot(aes(sepal.Length, sepal.width)) +
+ geom_point(aes(colour = factor(species))) +
+ scale_colour_manual(values=c("#433E91", "#188700", "#E05499"), name = "Species") +
+ labs(x = "Sepal length", y = "Sepal width", title = "Iris Sepal length-width relationship by species")
>
> ggsave("iris sepal width length plot.pdf")
Saving 8.89 x 5.73 in image
> |
```





Summary

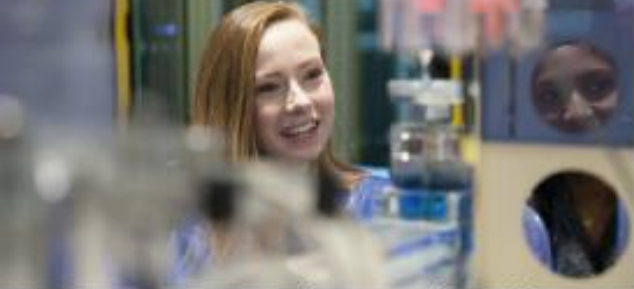
- Quality assurance matters
- RStudio Projects help you do your analysis correctly
- RStudio Projects help other people check your analysis
- Setting up RStudio Projects is an easy habit to get into



References

- 1) Rstudio Projects web overview:
<https://support.rstudio.com/hc/en-us/articles/200526207-Using-Projects>
- 2) R for Data Science, chapter 8 “Workflow: projects”: <http://r4ds.had.co.nz/workflow-projects.html#rstudio-projects>
- 3) The Aqua Book:
<https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government>





0161 244 0888



info@transportforthenorth.com



Transport for the North
2nd Floor, 4 Piccadilly Place
Manchester, M1 3BN

Ground Floor
West Gate, Grace Street
Leeds, LS1 2RP



transportforthenorth.com



TRANSPORT FOR THE
NORTH

