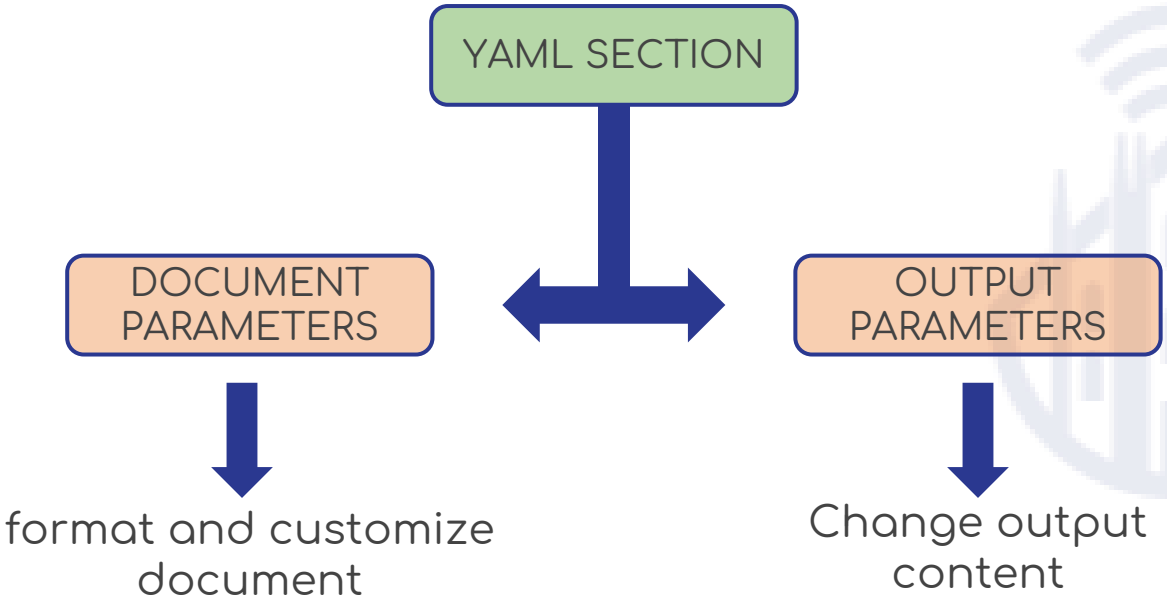
A UNIFIED APPROACH FOR WRITING AUTOMATIC REPORTS

Parameterization and Generalization of R-Markdown

Cristina Muschitiello
Niccolò Stamboglis

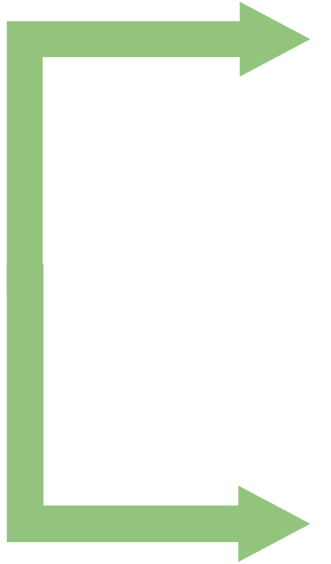
03_Parametrization

Parametrization: 2 tipologies



1.DOCUMENT PARAMETERS (and sub-parameters)

Indentation matters:



Not working

```yaml
---
title: "My First Markdown"
output:
 html_document:
 number_sections: true
 toc: true
 toc_depth: 3
 toc_float: false
author: "Nick Stamboglis"
date: "`r format(Sys.time(), '%e %B %Y')`"
link-citations: yes
---
```

Working

```yaml
---
title: "My First Markdown"
output:
 html_document:
  number_sections: true
  toc: true
  toc_depth: 3
  toc_float: false
author: "Nick Stamboglis"
date: "`r format(Sys.time(), '%e %B %Y')`"
link-citations: yes
---
```

1. DOCUMENT PARAMETERS

A Stylish Rmarkdown

```
---
title: "My First Markdown"
output:
 html_document:
  theme: cerulean
  number_sections: true
  toc: true
  toc_depth: 3
  toc_float:
   collapse: true
   smooth_scroll: true
author: "Nick Stamboglis"
date: "`r format(Sys.time(), '%e %B %Y')`"
link-citations: yes
---
```

Themes: https://www.datadreaming.org/post/r-markdown-theme-gallery/

Section numbering

Table of contents

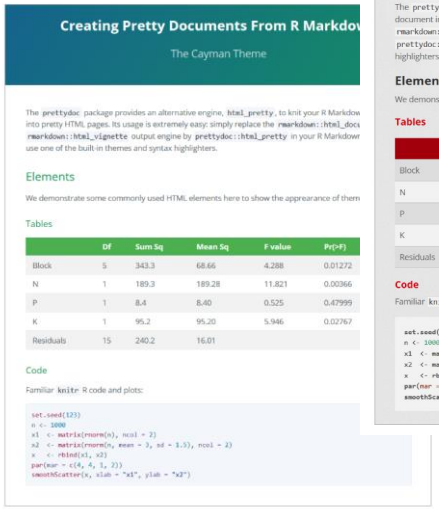
R Code

Link to bibliography

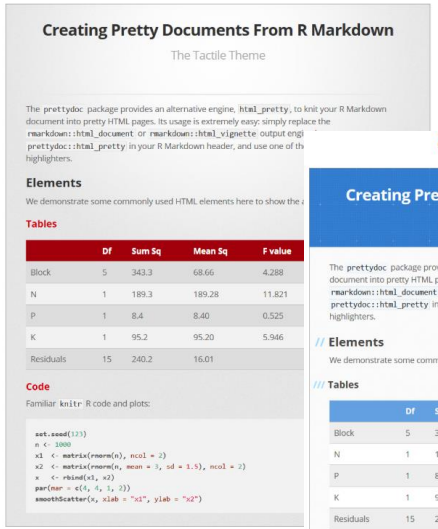
1.DOCUMENT PARAMETERS

Prettydoc styles

Cayman (Demo Page)

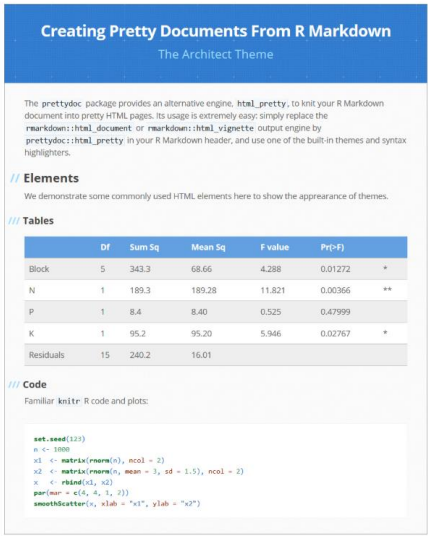


Tactile (Demo Page)

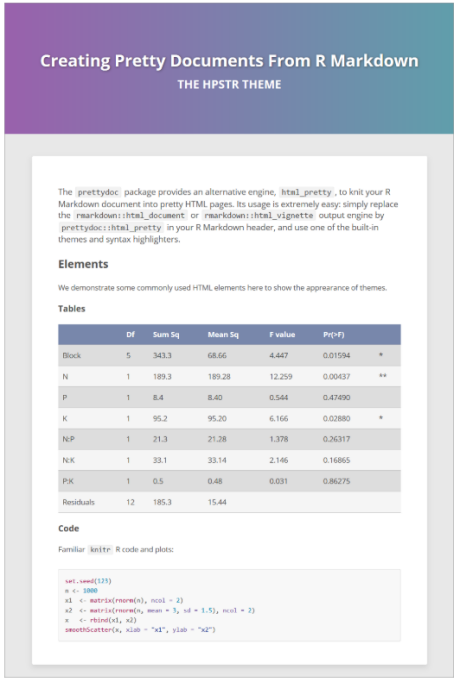


Leonids (Demo Page)

Architect (Demo Page)



HPSTR (Demo Page)



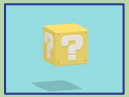
Sources:
https://prettydoc.statr.me/themes.html

Your turn: Exercise 4

E04_document_parameters.Rmd

1. Add a floating toc
2. Add numbered sections
3. Change theme

```yaml
---
title: "Working with document parameters"
output:
 html_document:
  theme: cerulean
  number_sections: true
  toc: true
author: "Nick Stamboglis"
date: "`r format(Sys.time(), '%e %B %Y')`"
---
```



Time: 5 min

2.OUTPUT PARAMETERS

- ➔ Make the document parameter-dependent
- ➔ Replicate the analysis for a specific value of the parameters

define:

```yaml
---
title: "Working with document parameters"
output: html_document
author: "Cristina Muschitiello"
date: "`r format(Sys.time(), '%e %B %Y')`"
params:
  month: "June"
  year: 2020
  name: "Cristina"
  sequence: 100
---
```

Use:

```
`r params$name`
`r params$month`
`r params$year`
```

Your turn: Exercise 5

E05_output_parameters.Rmd

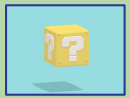
Change your welcome message

```
18 ## Welcome message
19
20 Hi everybody!
21
22 I'm **`r params$name`** and this report was created in ***`r params$month` `r params$year`***
23
```

Welcome Message

Hi everybody!

I'm **Cristina** and this report was created in ***May 2020***



Time: 2 min

2.OUTPUT PARAMETERS: Why?

Useful if you need to run

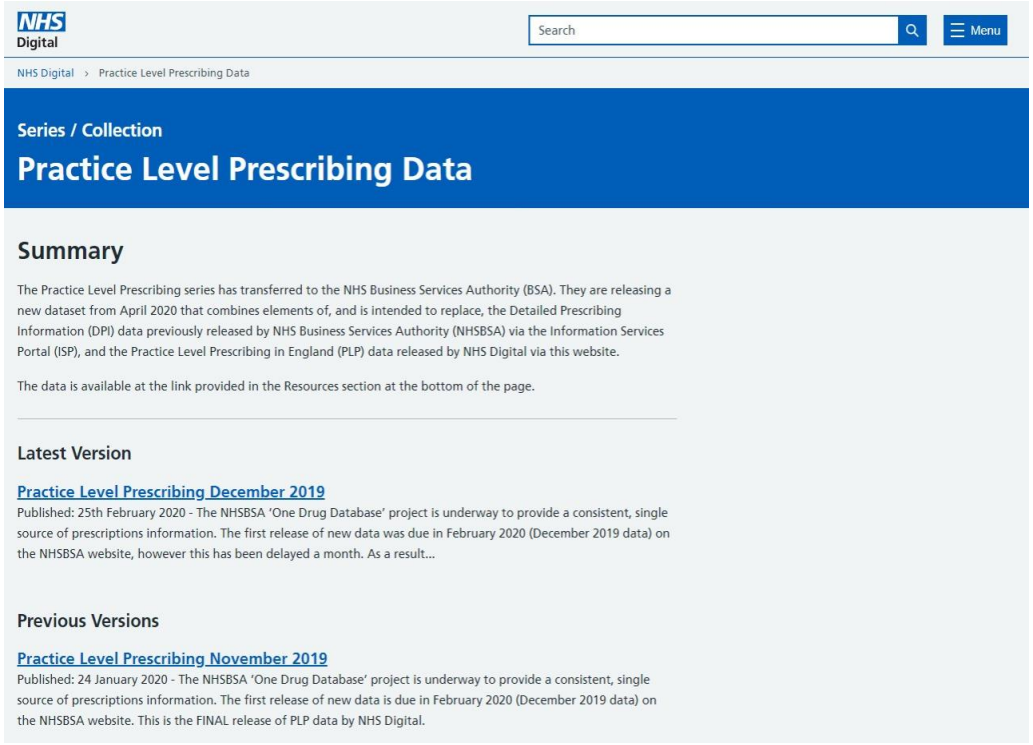
➔ Report for a specific customer

➔ Report for a specific section of your data

In general useful for ***reproducibility***

2.OUTPUT PARAMETERS: Reproducibility

First-up though: a real-world example



- ➔ English NHS Open Data
- ➔ Single GP Prescriptions
- ➔ Rich Monthly Data (>1.5Gb)
- ➔ From Sept. 2011

In practice:

	A	B	C	D	E	F	G	H	I	J
1	SHA	PCT	PRACTICE	BNF CODE	BNF NAME	ITEMS	NIC	ACT COST	QUANTITY	PERIOD
2	Q44	RTV	Y04937	0304010W0BBABAL	Phenergan_Tab 25mg	3	8,15	7,89	98	201801
3	Q44	RTV	Y04937	0401010Z0AAAAAA	Zopiclone_Tab 7,5mg	7	2,88	3,35	98	201801
4	Q44	RTV	Y04937	0401020K0AAAHAH	Diazepam_Tab 2mg	5	3,76	3,94	191	201801
5	Q44	RTV	Y04937	0402010ABAAAVAV	Quetiapine_Tab 50mg M/R	1	31,6	29,37	28	201801
6	Q44	RTV	Y04937	0402010ADAAAAAA	Aripiprazole_Tab 10mg	3	13,6	12,88	63	201801
7	Q44	RTV	Y04937	0402010ADAAADAD	Aripiprazole_Tab 5mg	1	1,36	1,37	7	201801
8	Q44	RTV	Y04937	0402010S0AAADAD	Promazine HCl_Oral Soln 25mg/5ml	2	26	24,14	300	201801
9	Q44	RTV	Y04937	040201030AAABAB	Risperidone_Tab 2mg	1	0,15	0,25	7	201801
10	Q44	RTV	Y04937	040201030AAACAC	Risperidone_Tab 3mg	1	0,36	0,45	14	201801
11	Q44	RTV	Y04937	040201060AAALAL	Olanzapine_Tab 15mg	1	13,2	12,33	14	201801
12	Q44	RTV	Y04937	0403010R0AAAAAA	Lofepramine HCl_Tab 70mg	1	4,69	4,46	28	201801
13	Q44	RTV	Y04937	0403010X0AAABAB	Trazodone HCl_Cap 100mg	1	3,75	3,59	28	201801
14	Q44	RTV	Y04937	0403010X0AAADAD	Trazodone HCl_Tab 150mg	1	7,54	7	28	201801
15	Q44	RTV	Y04937	0403030E0AAAAAA	Fluoxetine HCl_Cap 20mg	1	0,35	0,44	14	201801

Presentation level -> data is provided at the level of each individual drug name

Granularity -> GP practice (no individual prescription).

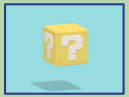
Identifiable data -> No.

Your turn: Exercise 6

E06_England_Report.Rmd

- Knit the Rmd to see its content
- Then change *year*, *month* and *executor* parameters to see what changes

```yaml
---
title: "***England report***"
output:
  html_document:
    theme: cerulean
params:
  month: 4
  year: 2019
  executor: "Cristina Muschitiello"
---
```



Time: 5 min

Changing parameters

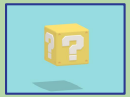
Parameters could be useful to run your analysis on a specific subset of your dataset (in our case a practice).

Your turn: Exercise 7

E07_GP_report.Rmd

- Knit the Rmd to see its content
- Then change parameters and see how different GPs behave in different months

```yaml
---
title: "**Practice reports**"
output:
  html_document:
    theme: cerulean
params:
  month: "05"
  year: 2019
  gp: "A81005"
---
```



Time: 5 min

Changing parameters

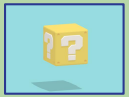
Parameters can also be useful to parametrize reports on specific quantitative thresholds (in our case a threshold on prescription rates).

Your turn: Exercise 8

E08_High_prescribing_report.Rmd

- Knit the Rmd to see its content
- Then change the threshold to see how High prescribing GPs behave

```yaml
---
title: "***High Prescribing GPs report***"
output:
  html_document:
    theme: cerulean
params:
  month: "07"
  year: 2019
  threshold: 10
---
```



Time: 5 min

Questions

Next up: Generalizability

Source: giphy.com