NSC-R Workshop

Literate programming with RMarkdown

Asier Moneva

06 Jul 2021

The Basics

Markdown syntax

> Help > RMarkdown Quick Reference

```
How can I format my text?
```

1. This is an
2. ordered list

- This is an unordered list.
  - And this is its sublist.
- For emphasis:
  - *This* is italic and *this* too.
  - **This** is bold and **this** too.
- Special formats:
  - $^{\text{This}}$ is superscript.
  - $_{\text{This}}$ is subscript.
  - ~~This~~ is strikethrough.
  - `This` is inline code.
- This is hyperlinked text.
- And this.[1]

You can also add math expressions, such as:

$$y = X\beta + \epsilon$$

Want to run a spell-check in the document? Press `F7`.

[1] is a footnote

Compile your document

Most used output formats:

- `html_document`
- `pdf_document`
- `word_document`

How to customize the output format? See, for example, `?html_document`. For Word, you can use a template document. I named mine `word_template_v0_1.docx`. You can then specify you are using a template in the YAML with the parameter `reference_docx`. Check out this blog post by Layton (2015) for a step-by-step tutorial.

Advanced formats with the `bookdown` package (Xie 2017):

- `html_document2`
- `pdf_document2`
- `word_document2`

Why? Cross-referencing.

Chunks and their parts

Use this first chunk to configure your global options and chunk options. Then keep a *lineal*, *organized* workflow. Remember to name your chunks for organization and quick access!

```
# Global options
options(scipen = 999)

# Chunk options: eval, echo, results, collapse, warning, message, error, include, cache, fig.width, fig
knitr::opts_chunk$set(echo = TRUE)
```

Here you can *load* your `packages`. To import crime data, I will use the `crimedata` R package by Ashby (2018) that connects to the Crime Open Database.

```
# Check if you have already installed this package. If not, install it
if (!requireNamespace("crimedata"))
  install.packages("crimedata")
# Then load it
library(crimedata)

if (!requireNamespace("here"))
  install.packages("here")
library(here)

if (!requireNamespace("kableExtra"))
  install.packages("kableExtra")
library(kableExtra)

if (!requireNamespace("knitr"))
  install.packages("knitr")
library(knitr)
```

Table 1: Top three property crime groups in Austin in 2019

Offense group	Count
larceny/theft offenses	304
destruction/damage/vandalism of property (except arson)	62
fraud offenses (except counterfeiting/forgery and bad checks)	56

```
if (!requireNamespace("tidyverse"))
  install.packages("tidyverse")
library(tidyverse)
```

And here you can *import* your data.

```
crime_data <- get_crime_data()
```

Then, you can *transform* your data to, for example, identify the top three property crime groups in Austin in 2019.

```
crime_table <- crime_data %>%
  filter(city_name == "Austin" & offense_against == "property") %>%
  group_by(offense_group) %>%
  summarise(count = n()) %>%
  ungroup() %>%
  arrange(desc(count)) %>%
  slice_head(n = 3)
```

In addition, you can display this information in a nice table using the `kableExtra` package. Table 1 shows that larceny/theft offenses is the most frequent offense group against property, with 304 records.

Now you may want to *visualize* the data to provide additional context and facilitate interpretation. Figure 1 ranks property offense groups by number of registries.

```
crime_figure <- crime_data %>%
  filter(city_name == "Austin" & offense_against == "property") %>%
  group_by(offense_group) %>%
  summarise(count = n()) %>%
  ungroup() %>%
  arrange(desc(count)) %>%
  ggplot(mapping = aes(
    x = reorder(
      x = offense_group,
      X = count
    ),
    y = count
  )) +
  geom_col() +
  scale_x_discrete(labels = function(x) str_wrap(
    string = x,
    width = 30
  )) +
```

```
  coord_flip() +
  labs(
    x = "Offense group",
    y = "Registries"
  ) +
  theme_classic()
print(crime_figure)
```

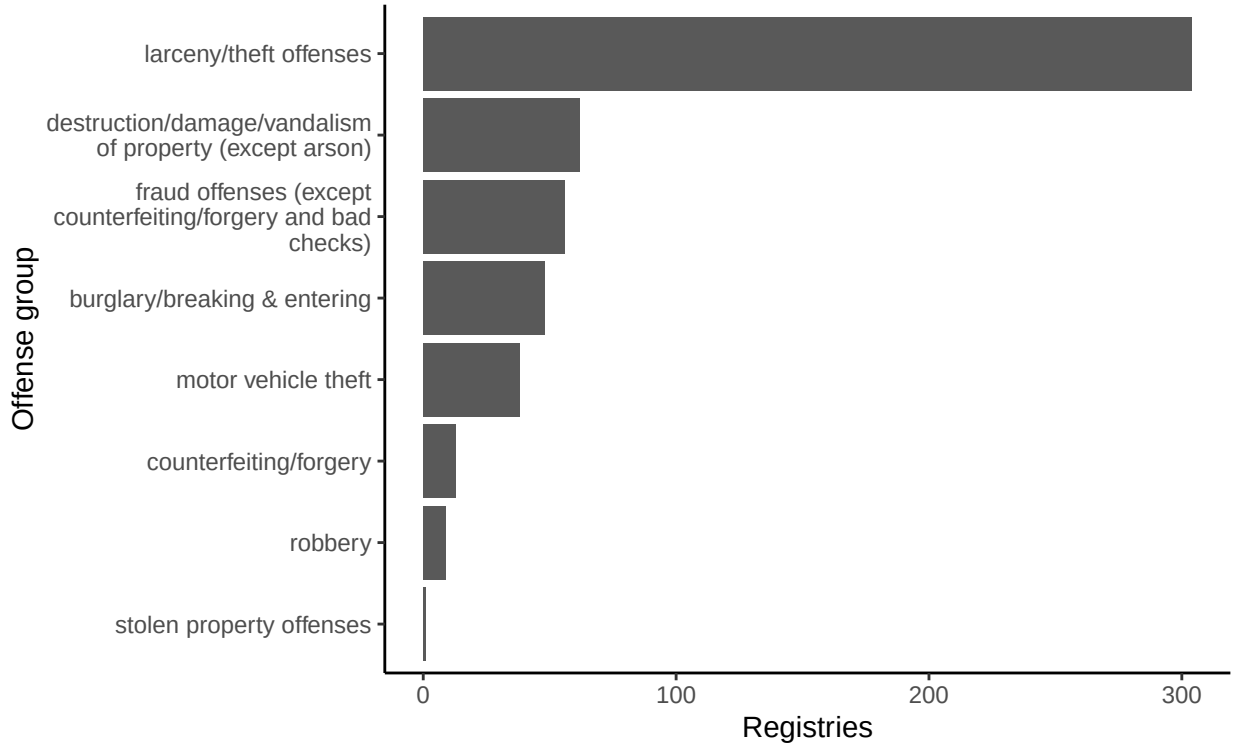

Figure 1: Property crime registries by offense group

And now imagine you want to insert a picture in your report. Let's include the NSCR logo as Figure 2.

To indicate the `path` on which the picture is stored, we used the `here` package by Müller (2020). To learn more, check out this blog post on *Why should I use the here package when I'm already using projects?*

```
include_graphics(path = here("scripts", "nscr_logo.jpg"))
```

Figure 2: Netherlands Institute for the Study of Crime and Law Enforcement logo

References with Zotero

This blog post by Joseph J. Allaire (2020), creator of RStudio, details how to integrate citations from Zotero

using RMarkdown.

In addition to citing papers, you can also cite R packages like `rmarkdown` (Allaire, Xie, McPherson, et al. 2021)! To retrieve the citation information for a specific package, you can call `citation()` and specify the name of the R package using the `package` parameter.

```
citation(package = "rmarkdown")
```

```
##
## To cite the 'rmarkdown' package in publications, please use:
##
##   JJ Allaire and Yihui Xie and Jonathan McPherson and Javier Luraschi
##   and Kevin Ushey and Aron Atkins and Hadley Wickham and Joe Cheng and
##   Winston Chang and Richard Iannone (2021). rmarkdown: Dynamic
##   Documents for R. R package version 2.9. URL
##   https://rmarkdown.rstudio.com.
##
##   Yihui Xie and J.J. Allaire and Garrett Grolemund (2018). R Markdown:
##   The Definitive Guide. Chapman and Hall/CRC. ISBN 9781138359338. URL
##   https://bookdown.org/yihui/rmarkdown.
##
##   Yihui Xie and Christophe Dervieux and Emily Riederer (2020). R
##   Markdown Cookbook. Chapman and Hall/CRC. ISBN 9780367563837. URL
##   https://bookdown.org/yihui/rmarkdown-cookbook.
##
## To see these entries in BibTeX format, use 'print(<citation>,
## bibtex=TRUE)', 'toBibtex(.)', or set
## 'options(citation.bibtex.max=999)'.
```

You can also format your references:

- Regular citation: (Allaire, Xie, McPherson, et al. 2021);
- In-text citation: Allaire, Xie, McPherson, et al. (2021) or just (2021);
- Citation with prefixes or suffixes: (i.e., Allaire, Xie, McPherson, et al. 2021, 10)

The reference list will automatically be compiled at the end of your document. But you can relocate it for your convenience using the following syntax.

```
:::{#refs}
:::
```

Templates

> New file > R Markdown > From Template

Learn more about the `rticles` package by Allaire, Xie, Wickham, et al. (2021) here!

References

Allaire, Joseph J. 2020. "RStudio 1.4 Preview: Citations." https://blog.rstudio.com/2020/11/09/rstudio-1-4-preview-citations/.

Allaire, Joseph J., Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2021. *Rmarkdown: Dynamic Documents for r*. https://rmarkdown.rstudio.com.

Allaire, Joseph J., Yihui Xie, Hadley Wickham, Ramnath Vaidyanathan, Carl Boettiger, Karl Broman, Kirill Müller, et al. 2021. *Rticles: Article Formats for r Markdown*. https://CRAN.R-project.org/package=rticles.

Ashby, Matthew P. J. 2018. "Studying Crime and Place with the Crime Open Database." https://doi.org/10.31235/osf.io/9y7qz.

Layton, Richard. 2015. "Happy Collaboration with Rmd to Docx." https://rmarkdown.rstudio.com/articles_docx.html.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Xie, Yihui. 2017. *Bookdown: Authoring Books and Technical Publications with r Markdown*. Boca Raton, FL: CRC Press.

You can wrap up your script with `sessionInfo()`. This function retrieves information about your OS and helps you to keep track the software versions you used, which is useful to write *reproducible examples*.

```
sessionInfo()
```

```
## R version 4.1.0 (2021-05-18)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] forcats_0.5.1    stringr_1.4.0    dplyr_1.0.7      purrr_0.3.4
##  [5] readr_1.4.0      tidyr_1.1.3      tibble_3.1.2     ggplot2_3.3.5
##  [9] tidyverse_1.3.1  knitr_1.33       kableExtra_1.3.4 here_1.0.1
## [13] crimedata_0.2.0
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.6         svglite_2.0.0      lubridate_1.7.10   class_7.3-19
##  [5] assertthat_0.2.1   rprojroot_2.0.2    digest_0.6.27      utf8_1.2.1
##  [9] R6_2.5.0           cellranger_1.1.0   backports_1.2.1    reprex_2.0.0
```

```
## [13] evaluate_0.14      e1071_1.7-7        highr_0.9          httr_1.4.2
## [17] pillar_1.6.1       rlang_0.4.11       readxl_1.3.1       rstudioapi_0.13
## [21] rmarkdown_2.9      labeling_0.4.2     webshot_0.5.2      munsell_0.5.0
## [25] proxy_0.4-26       broom_0.7.8        compiler_4.1.0     modelr_0.1.8
## [29] xfun_0.24          pkgconfig_2.0.3    systemfonts_1.0.2  htmltools_0.5.1.1
## [33] tidyselect_1.1.1   bookdown_0.22      fansi_0.5.0        viridisLite_0.4.0
## [37] withr_2.4.2        crayon_1.4.1       dbplyr_2.1.1       sf_1.0-1
## [41] grid_4.1.0         jsonlite_1.7.2     gtable_0.3.0       lifecycle_1.0.0
## [45] DBI_1.1.1          magrittr_2.0.1     units_0.7-2        scales_1.1.1
## [49] KernSmooth_2.23-20 cli_3.0.0          stringi_1.6.2      farver_2.1.0
## [53] fs_1.5.0           xml2_1.3.2         ellipsis_0.3.2     generics_0.1.0
## [57] vctrs_0.3.8        tools_4.1.0        glue_1.4.2         hms_1.1.0
## [61] yaml_2.2.1         colorspace_2.0-2   classInt_0.4-3     rvest_1.0.0
## [65] haven_2.4.1
```