Writing reproducible documents

Gbadamassi G.O. Dossa

last updated: 2022-11-04

Acknowledgements

The content of this module are based on materials from:

olivier gimenez's materials

Context

We will use the awesome `palmerpenguins` dataset 🐧, an alternative to Fisher's `iris` dataset, to explore and visualize data.

These data have been collected and shared by Dr. Kristen Gorman and Palmer Station, Antarctica LTER.

The package was built by Drs Allison Horst and Alison Hill, check out the official website.

The package `palmerpenguins` has two datasets.

```
library(palmerpenguins)
data(package = 'palmerpenguins')
```

The dataset `penguins` is a simplified version of the raw data; see `?penguins` for more info:

```
head(penguins)
```

```
## # A tibble: 6 × 8
##   species island    bill_length_mm bill_depth_mm flipper_l…¹ body_…² sex    year
##   <fct>   <fct>              <dbl>         <dbl>       <int>   <int> <fct> <int>
## 1 Adelie  Torgersen           39.1          18.7         181    3750 male   2007
## 2 Adelie  Torgersen           39.5          17.4         186    3800 fema…  2007
## 3 Adelie  Torgersen           40.3          18           195    3250 fema…  2007
## 4 Adelie  Torgersen           NA            NA            NA      NA <NA>   2007
## 5 Adelie  Torgersen           36.7          19.3         193    3450 fema…  2007
## 6 Adelie  Torgersen           39.3          20.6         190    3650 male   2007
## # … with abbreviated variable names ¹flipper_length_mm, ²body_mass_g
```

The other dataset `penguins_raw` has the raw data; see `?penguins_raw` for more info:

```
head(penguins_raw)
```

```
## # A tibble: 6 × 17
##   study…¹ Sampl…² Species Region Island Stage Indiv…³ Clutc…⁴ `Date Egg` Culme…⁵
##   <chr>     <dbl> <chr>   <chr>  <chr>  <chr> <chr>   <chr>   <date>       <dbl>
## 1 PAL0708       1 Adelie… Anvers Torge… Adul… N1A1    Yes     2007-11-11    39.1
## 2 PAL0708       2 Adelie… Anvers Torge… Adul… N1A2    Yes     2007-11-11    39.5
## 3 PAL0708       3 Adelie… Anvers Torge… Adul… N2A1    Yes     2007-11-16    40.3
## 4 PAL0708       4 Adelie… Anvers Torge… Adul… N2A2    Yes     2007-11-16    NA
## 5 PAL0708       5 Adelie… Anvers Torge… Adul… N3A1    Yes     2007-11-16    36.7
## 6 PAL0708       6 Adelie… Anvers Torge… Adul… N3A2    Yes     2007-11-16    39.3
## # … with 7 more variables: `Culmen Depth (mm)` <dbl>,
## #   `Flipper Length (mm)` <dbl>, `Body Mass (g)` <dbl>, Sex <chr>,
## #   `Delta 15 N (o/oo)` <dbl>, `Delta 13 C (o/oo)` <dbl>, Comments <chr>, and
## #   abbreviated variable names ¹studyName, ²`Sample Number`, ³`Individual ID`,
## #   ⁴`Clutch Completion`, ⁵`Culmen Length (mm)`
## # ℹ Use `colnames()` to see all variable names
```

For this exercise, we're gonna use the `penguins` dataset.

Question 1

- Create a new R Markdown document, name it and save it.
- Delete everything after line 12.
- Add a new section title, simple text and text in bold font.
- Compile ("Knit").

Question 2

- Add a chunk in which you load the `palmerpenguins`. The corresponding line of code should be hidden in the output.
- Load also the `tidyverse` suite of packages.
- Modify the defaults to suppress all messages.

Question 3

- Add another chunk in which you build a table with the 10 first rows of the dataset.

Question 4

- In a new section, display how many individuals, penguins species and islands we have in the dataset. This info should appear directly in the text, you might want to use inline code 😄.
- Calculate the mean of the (numeric) traits measured on the penguins.

Question 5

- In another section, entitled 'Graphical exploration', build a figure with 3 superimposed histograms, each one corresponding to the body mass of a species.

Question 6

- Install package `citr` to manage citations following the guidelines here. If everything goes well, you should see it in the pulldown menu `Addins` 💪.
- Pick a recent publication from the researcher who shared the data, Dr Kristen Gorman. Import this publication in your favorite references manager (we use Zotero, no hard feeling), and create a bibtex reference that you will add to to the file `mabiblio.bib`.
- Add `bibliography: mabiblio.bib` at the beginning of your R Markdown document (YAML).
- Cite the reference iin the text using `Insert citations` in the pull-down menu `Addins`.
- Compile.

Question 7

- Change the default citation format (Chicago style) into the The American Naturalist format. It can be found here https://www.zotero.org/styles. To do so, add `csl: the-american-naturalist.csl` in the YAML.

Question 8

- Build your report in html, pdf and docx format. 🎉