Reports in R Markdown

July 28 2022

Getting started

Download this application exercise by pasting the code below into your **console**

```
download.file("https://sta101.github.io/static/appex/ae15.Rmd",
destfile = "ae15.rmd")
```

Next download the references file by pasting the code below into your **console**

```
download.file("https://sta101.github.io/static/appex/references.bib",
    destfile = "references.bib")
```

Cleaning up the mess below

```
knitr::opts_chunk$set(message = TRUE,
                      warning = TRUE,
                      echo = TRUE,
                      fig.width = 6, #width of figure
                      fig.asp = .618, #set figure height based on aspect ratio
                      out.width = "75%", #width relative to text
                      fig.align = "center" #alignment
                      )
```

```
library(tidyverse)
```

```
## -- Attaching packages ----------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.6      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.1.1      v forcats 0.5.1
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(palmerpenguins) #use the penguins data frame
library(knitr)
```

Today

- Code chunk options
- Citations & links
- Customizing plots
- Neatly display tables and output

Introduction

For this analysis, we will use the `penguins` data set in the **palmerpenguins** R package (Horst, Hill, and Gorman 2020). This data set contains measurements and other characteristics for over 300 penguins observed near Palmer Station in Antarctica. The data were originally collected by Dr. Kristen Gorman.

Click here to learn more on the palmerpenguins website.

Code chunk options

Code chunk options are used to customize how the code and output is displayed in the knitted R Markdown document. There are two ways to set code chunk options:

- In the header of an individual code chunk
- As a global setting to apply to all code chunks

A few options to change what we show/hide in the knitted document:

- `message = FALSE`: hide messages (default == TRUE)
- `warning = FALSE`: hide warnings (default == TRUE)
- `echo = FALSE`: hide code (default = TRUE)
- `include = FALSE`: runs code but hides all code and output (default = FALSE). (Avoid using this option as a global setting.)

For the project, you will set the option `echo = FALSE` to hide all code in your final report.

- Change the global code chunk settings so the document is more suitable for a general audience. Let's take a look at the updated PDF.

Citations

Your report will include citations, e.g. the data source, previous research, and other sources as needed. At a minimum, you should have a citation for the data source.

All of your bibliography entries will be stored in a `.bib` file. The entries of the bibliography are stored using BibTex, i.e., a format to store citations in LaTeX. Let's take a look at `references.bib`.

In addition to the .bib file:

- Include `bibliography: references.bib` in the YAML.
- At the end of the report, include `## References`. This will list all of the references at the end of the document.
- If you want to include an Appendix, include the additional code shown at the end of this document.

Citation examples

1. In Gorman and LTER (2014), the authors focus on the analysis of Adelie penguins.

2. Studies have shown whether environmental variability in the form of winter sea ice is associated with differences in male and female pre-breeding foraging niche (Gorman and LTER 2014).

Practice

- Add a citation for *R markdown: The definitive guide* to this document.

Links

Add URLs to your document using the following syntax:

DISPLAYED TEXT

- Add `links-as-notes: true` to the YAML to display URLs as a footnote in the knitted PDF. (This is useful if your document will be printed)
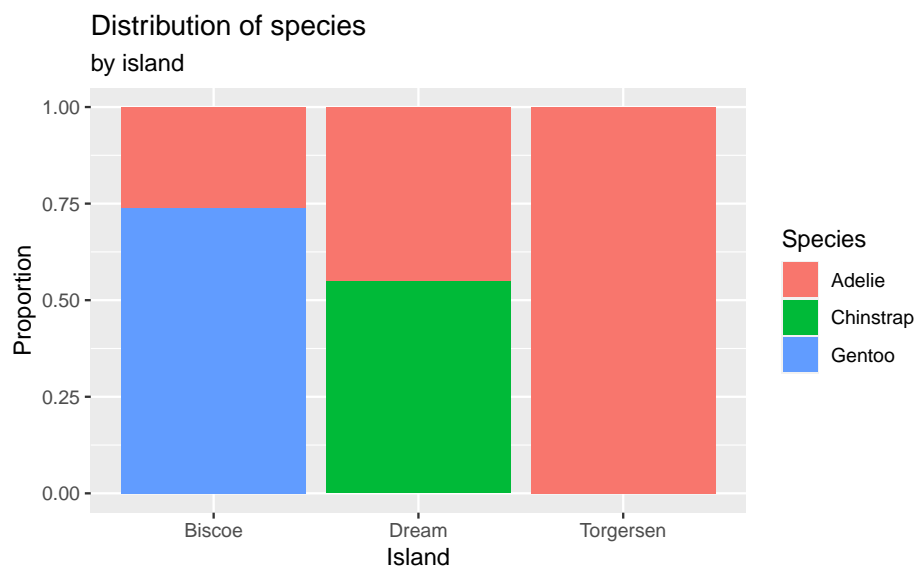
Link example

You can learn more about Markdown on the R Markdown website.

Customizing plots

Let's start with a plot looking at the species vs. the island.

```r
ggplot(data = penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill") +
  labs(x = "Island",
       y = "Proportion",
       fill = "Species",
       title = "Distribution of species",
       subtitle = "by island")
```

Distribution of species
by island

Standard color palette + theme

You can set a standard color palette and theme at the top of the document to make the plots look coordinated throughout the document. Navigate to the code chunk labeled `ggplot2-options` and let's take a look.

Choose 3 colors from the color palette, then use the code below to apply the colors to the segmented bar plot. Remove `eval = FALSE` from the code chunk header.

```r
#fill in the code and remove #eval = FALSE from the code chunk header
ggplot(data = penguins, aes(x = island, fill = species)) +
  geom_bar(position = "fill") +
  labs(x = "Island",
       y = "Proportion",
       fill = "Species",
       title = "Distribution of species",
```

```
        subtitle = "by island") +
  scale_fill_manual(values = c(color_palette$____,
                               color_palette$_____,
                               color_palette$_____))
```

- Make a histogram of `bill_depth_mm` and fill in the histogram using the one of the colors from the color palette. Notice we have also added a caption that will display in the knitted document.

```
# add code here
ggplot(data = penguins, aes(x = flipper_length_mm)) +
  geom_histogram()
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

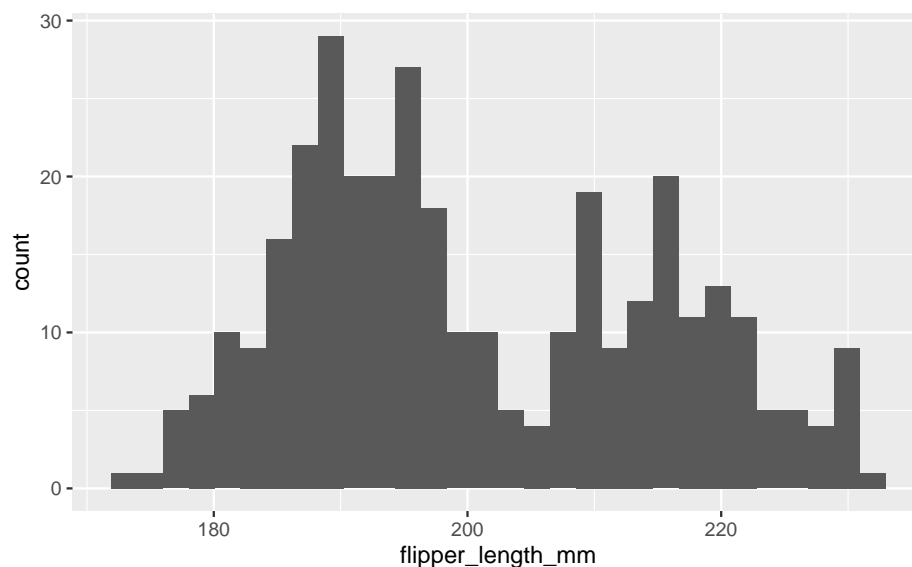## Warning: Removed 2 rows containing non-finite values (stat_bin).

Figure 1: Distribution of penguin bill depth

Neatly display table and output

- Complete the code below to calculate the mean, median, and standard deviation of `bill_depth_mm`. Display the results. Remove `eval = FALSE` from the code chunk header.

```
# Complete the code and remove eval = FALSE from the code chunk header
penguins %>%
  filter(!is.na(bill_depth_mm)) %>%
```

- Let's neatly display the results using the `kable` function from the **knitr** package. We will

  - Display results to 3 digits
  - Customize column names
  - Add a caption

```
## add code
```

Acknowledgements

These notes were adapted from the following:

- Communicating data analysis by Dr. Lucy McGowan
- Customizing R Markdown by Dr. Mine Çetinkaya-Rundel
- Take a sad plot and make it better by Dr. Alison Hill

References

Gorman, Kristen, and Palmer Station Antarctica LTER. 2014. "Structural Size Measurements and Isotopic Signatures of Foraging Among Adult Male and Female Gentoo Penguins (Pygoscelis Papua) Nesting Along the Palmer Archipelago Near Palmer Station, 2007-2009."

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data.* https://allisonhorst.github.io/palmerpenguins/.