# Higher, further, faster with Marvelous R Markdown

athomas\_mock



# The Heroine's Journey (sans spoilers)

Act 1: Story is set up (Here's the tool)

Act 2: Complication arises (Here's the problem)

Act 3: Heroine finds resolution (Here's a solution)



Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>

# But coding is not a linear journey

Less about:

**Beginner < Intermediate < Expert** 

More about:

Expanding your toolset for the right job at the right time



Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>

# RMarkdown Taxonomy

**Literate Programming** 

**Data Products** 

**Control Documents** 

**Templating** 



# Literate Programming



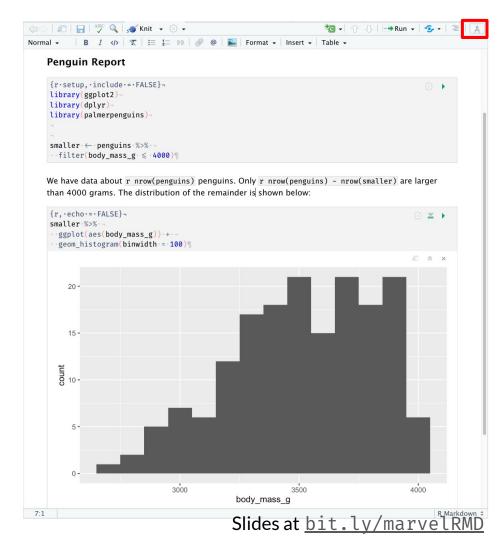
# Literate Programming

A programming paradigm introduced by Donald Knuth in which a computer program is given an explanation of its logic in a natural language, such as English, interspersed with snippets of macros and traditional source code, from which compilable source code can be generated.

```
title: "Penguins"
                            YAML
date: 2020-08-11
output: html document
```{r setup, include = FALSE}
library(ggplot2)
library(dplyr)
                                 Code
library(palmerpenguins)
smaller <- penguins %>%
  filter(body mass g <= 4000)
We have data about `r nrow(penguins)` penguins. Only
r nrow(penguins) - nrow(smaller) are larger than
4000 grams. The distribution of the remainder is
shown below:
```{r. echo = FALSE}
smaller %>%
                                   Code
  ggplot(aes(body_mass_g)) +
  geom_histogram(binwidth = 100)
```

# Literate Programming

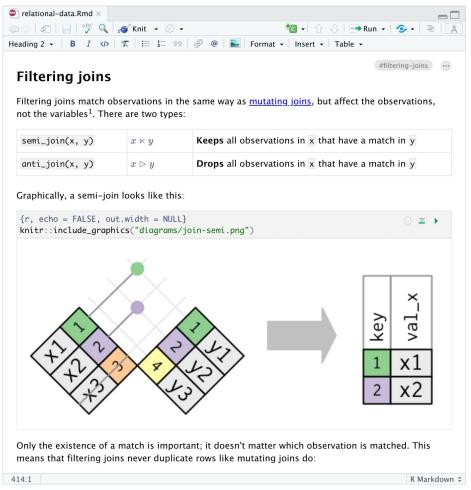
```
(□□) 2□ | → Run - 5 - 2 | A
 9 ### Penguin Report
 10
 11 * ```{r setup, include = FALSE}-
 12 library(ggplot2)
 13 library(dplyr)
    library(palmerpenguins)
 16
 17 smaller ← penguins %>% -
      filter(body_mass_g ≤ 4000)
 19 -
 20
21 We have data about `r nrow(penguins)` penguins. Only `r
     nrow(penguins) -- nrow(smaller) are larger than 4000 grams. The
     distribution of the remainder is shown below:
 22
 23 * ``{r, echo = FALSE}
 24 smaller - %>% - -
      -ggplot(aes(body_mass_g)) + --
      geom_histogram(binwidth = 100)
 27 - ***-
         20 -
         15-
                                                            4000
                                 body mass g
      ## Penguin Report $
                                                           R Markdown $
```



# Visual RMarkdown

Coming in the RStudio 1.4 release!

- Visual editing for all of Pandoc markdown
- Extensive support for citations
- Scientific and technical writing features, including LaTeX
- Writing productivity w/ real time spell-checking
- Tight integration with source editing
- Rich keyboard support and can use the \( \mathre{\*} \)
   / shortcut to insert anything

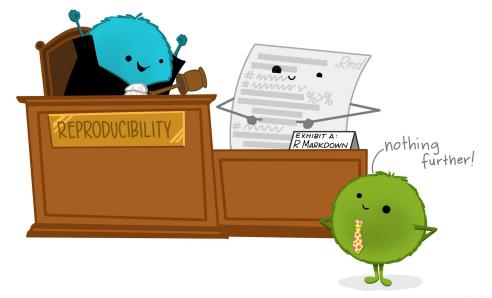


Guide to Visual RMarkdown

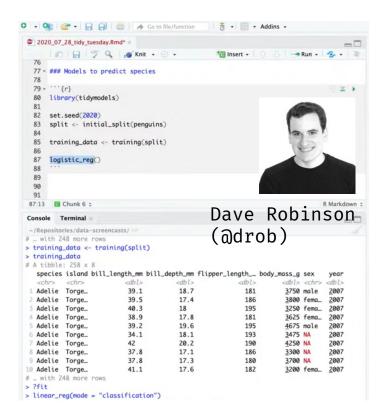
# M.V.P of Reproducibility

### R Markdown

- Output is self-documenting (code is embedded)
- Has to run successfully to save/knit output
- Is diffable AND human readable in version control



# TidyTuesday screencasts exemplars



```
🔊 palmer_penguins.R<u>md* 🔀</u>
                  Knit - 🌣 -
                                                    Onsert → A J Run → O →
        geom_point(alpha = 0.7) +
        facet_wrap(~species)
    penguins_df <- penguins %>%
        filter(!is.na(sex)) %>%
        select(-year, -island)
 44 - ## Build a model
                                                Julia Silge :
 47 library(tidymodels)
                                                (@juliasilge)
 49 set.seed(123)
 50 penguin_split <- initial_split(penguins_df, strata = sex)</pre>
 51 penguin_train <- training(penguin_split)
    penguin_test <- testing(penguin_split)</pre>
 54 set.seed(234)
    penguin_boot < bootstraps(penguin_train)
55:16 Chunk 4
```

https://www.youtube.com/watch?v=z57i2GVcdww

https://www.youtube.com/watch?v=ImpXawPNCfM

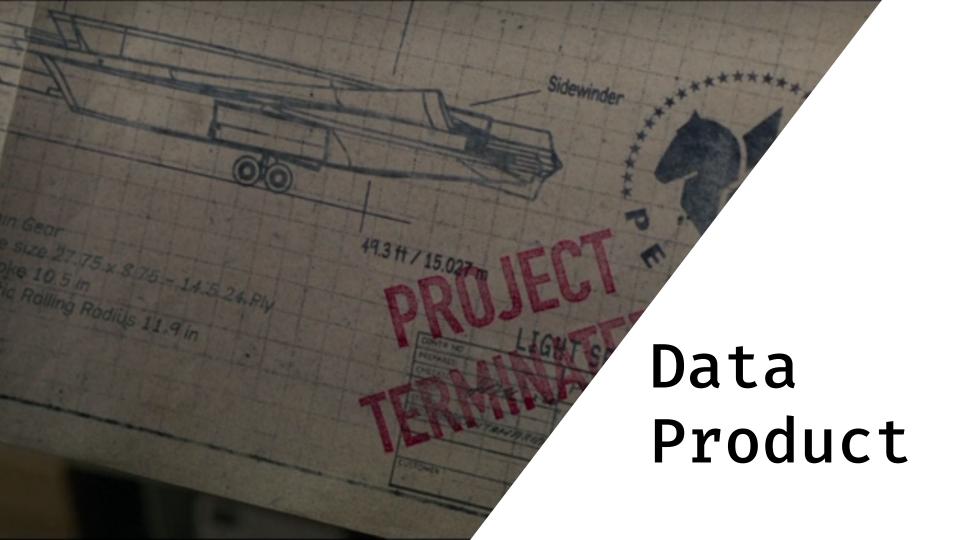
Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>

# Not just for R!

- Python natively or via reticulate
- SQL
- CSS or <u>Javascript</u>
- Bash, Rcpp, Stan and others

All together a total of **52** possible language engines from knitr







## Data Product(s)

Reports - HTML, PDF, Word, Markdown, RTF, Tufte handout, GitHub document

Presentations - Xaringan, Powerpoint, ioslides, reveal.js, beamer

<u>Dashboards</u> - flexdashboard as standalone or with Shiny

Entire websites - Blogdown, Hugodown, distill, R Markdown site

Books - bookdown

HTMLWidgets - DT, reactable, plotly, crosstalk and more

### Crosstalk

```
output: html document
```{r setup. include=FALSE}
library(tidyverse)
library(reactable)
library(plotly)
library(crosstalk)
librarv(htmltools)
```{r, echo = FALSE}
small df <- palmerpenguins::penguins %>%
 select(species, sex, flipper length mm, body mass g) %>%
 na.omit()
pen df <- SharedData$new(small df)</pre>
                                    Crosstalk data
plotly graphic <- plot ly(pen df,</pre>
                         x = \sim flipper length mm.
                         y = ~body mass g,
 plotly graph
                         color = ~species,
                         text = ~species) %>%
 add markers() %>%
 layout(xaxis = x, yaxis = y) %>%
 highlight(on = "plotly selected",
           off = "plotly doubleclick".
           persistent = FALSE))
reactable table
reactable table <- reactable(pen df,
           columns = list(
             species = colDef("Species"),
             sex = colDef("Sex", cell = function(value){
               if (value == "male") "Male" else "Female"}),
             flipper length mm = colDef("Flipper Length"),
             body mass g = colDef("Body Mass"))))
```

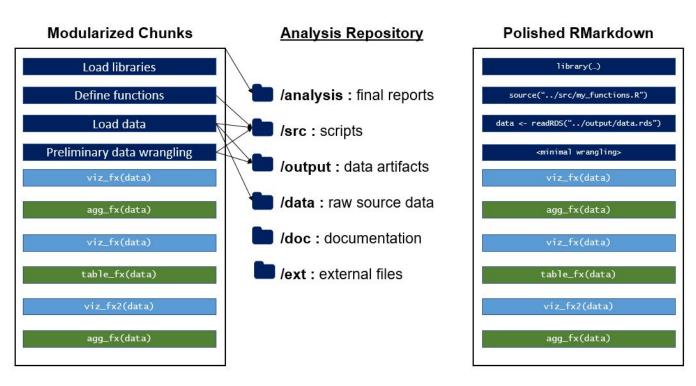
```
```{r, echo = FALSE, warning=FALSE, message=FALSE}
x <- list(title = "Flipper Length (mm)", range = c(170, 240))
v \leftarrow list(title = "Body Mass (g)", range = c(2500, 6500))
div(
  h1("Penguin Data"),
  h3("Filter by: Sex, Species, or Body Mass"),
  br().
  bscols(
    widths = c(2. 10).
    list(
      filter checkbox("sex", "Sex", pen df, ~sex),
      filter checkbox("species", "Species", pen df, ~species),
      filter slider("mass", "Body Mass (g)", pen df, ~body mass g)
    plotly graphic,
  br().
  reactable table
```







# Meta RMarkdown Driven Development



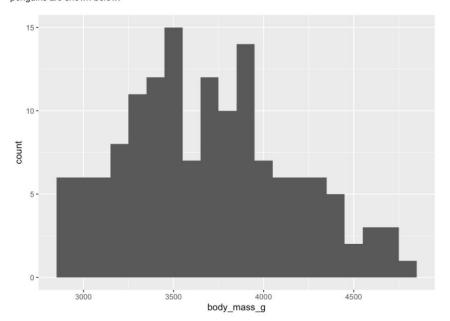
### **Knit with Parameters**

```
title: "Penguins"
date: 2020-08-11
output: html document
params:
                  Create in YAML
 species: Adelie
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
                       Reference in Code
smaller <- penguins %>%
 filter(species == params$species,
        !is.na(body mass g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified
as r params$species. The distribution of the
r params$species penguins are shown below:
                      Reference in Line
smaller %>%
 ggplot(aes(body_mass_g)) +
geom_histogram(binwidth = 100)
```

#### **Penguins**

2020-08-11

We have data about 344 penguins. Only 198 are classified as Adelie. The distribution of the Adelie penguins are shown below:



Resource Links: Ex. 1, Ex. 2, Ex. 3
Slides at bit.ly/marvelRMD

### Reference .R files

```
title: "Penguins"
date: 2020-08-11
output: html document
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
knitr::read_chunk('penguins.R') Read.R code
      Reference by name
```{r, smaller-penguins, echo = FALSE}
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified as
Adelie. The distribution of the
Adelie penguins are shown below:
```{r, plot-penguins, echo = FALSE}
```

### penguins.R

```
## ---- smaller-penguins Label block of code
smaller <- penguins %>%
  filter(species == "Adelie",
            !is.na(body mass g))
## ---- plot-penguins
smaller %>%
  ggplot(aes(body mass g)) +
  geom histogram(binwidth = 100)
Penguins
 We have data about 344 penguins. Only 198 are classified as Adelie. The distribution of the Adelie
```

### Child documents

#### adelie-report.Rmd

```
output: html document
                            Uses all code
smaller <- penguins %>%
 filter(species == "Adelie",
        !is.na(body mass g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified as
Adelie. The distribution of theAdelie penguins are
shown below:
```{r, echo = FALSE}
smaller %>%
 ggplot(aes(body_mass_g)) +
 geom_histogram(binwidth = 100)
```

```
title: "Penguins"
date: 2020-08-11
output: html_document
---

```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)

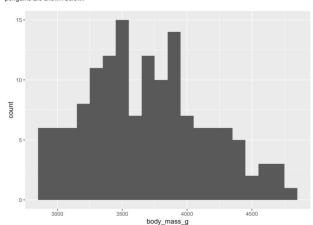
Uses report code

```{r, child=c("adelie-report.Rmd")}
```

#### **Penguins**

#### 2020-08-11

We have data about 344 penguins. Only 198 are classified as Adelie. The distribution of the Adelie penguins are shown below:



## Logical Child documents

### adelie-report.Rmd

```
output: html document
                            Uses all code
smaller <- penguins %>%
 filter(species == "Adelie",
        !is.na(body mass g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified as
Adelie. The distribution of the Adelie penguins are
shown below:
```{r, echo = FALSE}
smaller %>%
 ggplot(aes(body_mass_g)) +
 geom_histogram(binwidth = 100)
```

```
title: "Penguins"
date: 2020-08-11
output: html_document
---
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)

species_sel <- "Chinstrap"
Logical
```{r, child=if (species_sel == "Adelie") "a-report.Rmd"}</pre>
```

Uses report code conditionally

#### **Penguins**

2020-08-11

### Blastula emails

```
title: "Penguins"
date: 2020-08-11
output: html document
  `{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
library(blastula)
                     Referenced code
  {r penguin-plot, echo = FALSE}
penguins %>%
 filter(species == "Adelie",
         !is.na(body mass g)) %>%
 ggplot(aes(body_mass_g)) +
 geom_histogram(binwidth = 100)
```{r connect email, echo = FALSE}
render_connect_email(input = "penguin-email.Rmd") %>%
 attach connect email(
   subject = "Penguin Report"
                                 Generates email
```

title: "Penguins Report" penguin-email.Rmd

date: 2020-08-11

output: blastula::blastula\_email

---

```{r packages, include = FALSE}
library(tidyverse)

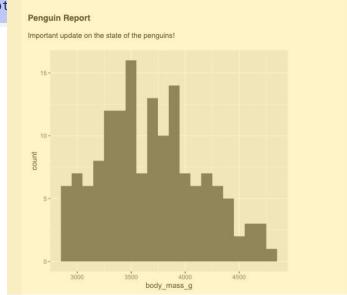
library(palmerpenguins)

This is an email preview for RStudio Connect
Use attach connect email(preview = FALSE) to attach without this preview.

Important update c

EMAIL SUBJECT: Penguin Report





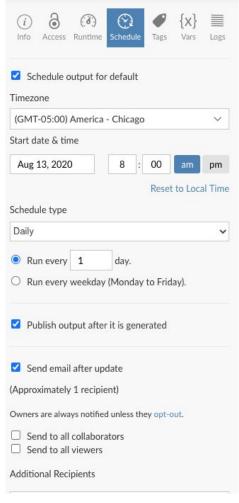
Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>

### RMarkdown + RStudio Connect

RStudio Connect is a hosting and execution platform for Shiny, RMarkdown, Plumber (also Jupyter, Flask, Dash, & Streamlit)

### Can execute/schedule RMarkdown for all sorts of things like:

- <u>Self-service</u> parameterized RMarkdown for non-tech users
- Extract Transform Load from SQL, APIs or Spark for example
- Automated reports with logging/history
- Long model <u>training steps</u> and save model upon completion
- Send blastula emails conditionally or on a schedule





Templating



### Render with Params

```
title: "Penguins"
date: 2020-08-11
output: html document
params:
                  Create in YAML
 species: Adelie
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
                        Reference in Code
smaller <- penguins %>%
 filter(species == params$species,
        !is.na(body mass g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified
as r params$species. The distribution of the
r params$species penguins are shown below:
                      Reference in Line
smaller %>%
 ggplot(aes(body_mass_g)) +
 geom_histogram(binwidth = 100)
```

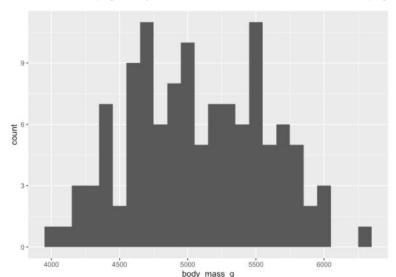
```
rmarkdown::render(
    "penguin.rmd",
    params = list(species = "Gentoo")
)
```

#### Uses report code w/ new parameter

#### **Penguins**

2020-08-11

We have data about 344 penguins. Only 225 are classified as Gentoo. The distribution of the Gentoo penguins are shown below



Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>

### Render with Params

```
title: "`r params$species` Penguins"
date: 2020-08-11
output: html document
params:
 species: Adelie Create in YAML
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
                       Reference in Code
smaller <- penguins %>%
 filter(species == params$species,
        !is.na(body mass g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified
as r params$species. The distribution of the
r params$species penguins are shown below:
                      Reference in Line
smaller %>%
 ggplot(aes(body_mass_g)) +
geom_histogram(binwidth = 100)
```

#### Uses report code w/ new parameter

```
for (penguin in unique(penguins$species)) {
  rmarkdown::render(
    input = "penguins-report.rmd",
     params = list(species = penguin),
     output file = glue::glue("{penguin}-report.html")
render fun <- function(penguin){</pre>
  rmarkdown::render(
    input = "penguins-report.rmd",
    params = list(species = penguin),
    output file = glue::glue("{penguin}-report.html")
penguins %>%
  distinct(species) %>%
  pull() %>%
  as.character() %>%
  purrr::walk(render fun)
```

```
      ■ Adelie-report.html
      584 KB
      Aug 13, 2020, 11:45 AM

      ■ Chinstrap-report.html
      672 KB
      Aug 13, 2020, 11:45 AM

      ■ Gentoo-report.html
      672.3 KB
      Aug 13, 2020, 11:45 AM
```

# Loop within a doc

```
title: "Penguin Report"
output: html document
```{r penguin function, include=FALSE}
library(tidyverse)
library(palmerpenguins)
library(glue)
knitr::opts_chunk$set(echo=FALSE, fig.width=6,
 message=FALSE)
                      Source function
source("multiplot.R")
penguins <- palmerpenguins::penguins %>%
 filter(!is.na(bill length mm),
   !is.na(flipper_length_mm))
                       Generate raw output
```{r loop-output, results="asis"}
penguins %>%
 distinct(species) %>%
 pull(species) %>%
 as.character() %>%
 walk(multiplot)
        Loop across input w/ function
```

#### Penguin Report

#### Adelie

multiplot.R

flipper len <- df

pull(mean) %>%

glue::glue("There

plot out <- df per

geom point() +

print(plot\_out)

cat(" \n \n")

ggplot(aes(x = t))

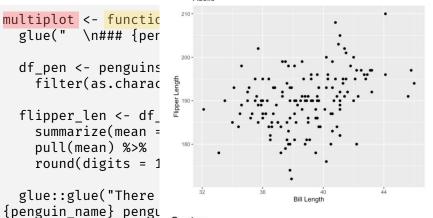
labs(x = "Bill l

y = "Flippε

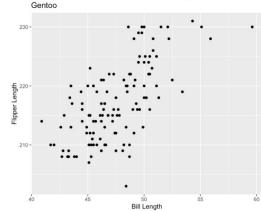
title = per

{flipper len}.

There are 151 observations of Adelie penguins. The average flipper length is 190



23 observations of Gentoo penguins. The average flipper length is 217.2.



# whisker vs glue ...

... is one vs two

```
glue::glue("There are {nrow(mtcars)} rows in the mtcars dataset")

# There are 32 rows in the mtcars dataset

rows_in <- nrows(mtcars)

whisker::whisker.render('There are {{rows_in}} rows in the mtcars dataset')

# [1] "There are 32 rows in the mtcars dataset"</pre>
```

# Whisker for #TidyTuesday

```
use tidytuesday readme <- function(title subject, dataset name){</pre>
  raw readme <- readLines(here::here("static/readme template.md"))</pre>
  title subject <- title subject
  dataset name <- dataset name
  core url <- "https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/"</pre>
  year <- lubridate::year(Sys.Date())</pre>
  tues date <- tidytuesdaymeta::next tuesday()</pre>
  week num <- tidytuesdaymeta::next week num()</pre>
  filled readme <- whisker::whisker.render(raw readme)</pre>
  writeLines(
                                  Fill the template
    filled readme,
    here::here(
      lubridate::year(lubridate::today()),
      tidytuesdaymeta::next tuesday(),
      "readme.md"
```

# Whisker output

```
### Get the data here
```{r}
# Get the Data
# Read in with tidytuesdayR package
# Install from CRAN via: install.packages("tidytuesdayR")
# This loads the readme and all the datasets for the week of interest
# Either ISO-8601 date or year/week works!
tuesdata <- tidytuesdayR::tt load('2020-08-11')
tuesdata <- tidytuesdayR::tt_load(2020, week = 33)</pre>
avatar <- tuesdata$avatar
                                            Filled readme.md
# Or read in the data manually
avatar <- readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2020/2020-08-11/avatar.csv')</pre>
### Data Dictionary
# `avatar.csv`
```

# Whisker inputs

```
title: "{{species}} Penguin"
output: html document
```{r setup, include = FALSE}
library(tidvverse)
                              Pass short text
library(palmerpenguins)
smaller <- penguins %>%
  filter(species == "{{species}}",
        !is.na(body_mass_g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified as {{species}}.
The distribution of the {{species}} penguins are shown below:
smaller %>%
  ggplot(aes(body_mass_g)) +
  geom_histogram(binwidth = 100)
{{long_prose}}
```

### Pass long text

```
title: "Adelie Penguin"
output: html_document
```{r setup, include = FALSE}
library(tidyverse)
library(palmerpenguins)
smaller <- penguins %>%
 filter(species == Adelie, !is.na(body_mass_g))
We have data about `r nrow(penguins)` penguins. Only
`r nrow(penguins) - nrow(smaller)` are classified as Adelie. The
distribution of the Adelie penguins are shown below:
smaller %>%
 ggplot(aes(body mass g)) +
 geom histogram(binwidth = 100)
```

### Mating

waterproof feathers.

The mating season begins with the Antarctic spring in October. The penguins create nests by piling little stones in circles. Once the egg is laid in December the parents take turns incubating the egg and going to hunt. The parent that stays behind does not eat during their turn with the egg. Once the hatched chick is about 3 weeks old both parents will abandon it, returning to the sea to hunt. The downy chicks gather into a group called a crèche to keep each other warm. They will start to hunt at about 9 weeks old once their down has been replaced with

Slides at bit.ly/marvelRMD

### whisker function

```
Read in
Whisker Fill
```

Write out

```
Input
```

```
use_penguin_template <- function(species, long_prose){
  raw_rmd <- readLines("penguin-whisker.Rmd")

filled_rmd <- whisker::whisker.render(raw_rmd)

writeLines(
  text = filled_rmd,
  con = glue::glue("{species}-report.rmd")
)
}</pre>
```

```
long_prose <- "### Mating\n\nThe mating season begins with the
Antarctic spring in October. The penguins create nests by piling
little stones in circles. Once the egg is laid in December the
parents take turns incubating the egg and going to hunt. The
parent that stays behind does not eat during their turn with the
egg. Once the hatched chick is about 3 weeks old both parents will
abandon it, returning to the sea to hunt. The downy chicks gather
into a group called a crèche to keep each other warm. They will
start to hunt at about 9 weeks old once their down has been</pre>
```

### Use template

use\_penguin\_template(species, long\_prose)

replaced with waterproof feathers."

species <- "Adelie"</pre>

#### usethis::use\_template

Creates a file from data and a template found in a package. Provides control over file name, the addition to .Rbuildignore, and opening the file for inspection.

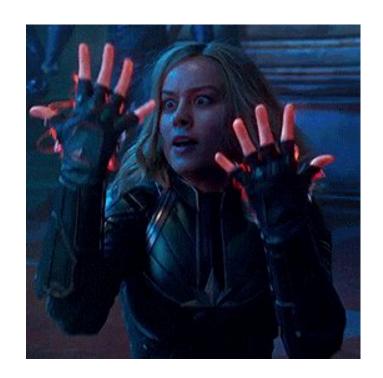
```
use_template(
  template,
  save_as = template,
  data = list(),
  ignore = FALSE,
  open = FALSE,
  package = "usethis"
)
```

# You have the power, now use it!

### Follow up reading

- Emily Riederer's <u>RMD Driven Development</u>
- Sharla Gelfand's RMD templating
- My <u>Meta RMarkdown blogpost</u>
- RMarkdown Cookbook
- RMarkdown <u>Definitive Guide</u>

Slides at <a href="mailto:bit.ly/marvelRMD">bit.ly/marvelRMD</a>



OVES MATERIALION

BOMUS HOMARD SLILDE