my first notebook

‽

today’s date

Table of Contents

This is an [R Notebook](https://rmarkdown.rstudio.com/r_notebooks.html) illustrating [literate programming](https://en.wikipedia.org/wiki/Literate_programming) techniques with R code **inside** *code chunks* and incorporating explanations (enhanced via [R Markdown](http://rmarkdown.rstudio.com/)) **outside** the *code chunks*. You can execute code within the R Notebook. Results appear beneath the code will also be integrated into the output documents identified in the [YAML](https://en.wikipedia.org/wiki/YAML) header .

Try executing the code chunk, below, by clicking the green-arrow *Run* button (within the chunk), or by placing your cursor inside the chunk and pressing *Ctrl+Shift+Enter.*

plot(cars)



Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.

*Ctrl+Shift+M*: Generates a pipe – %>% – between functions (When you use the tidyerse package.)

###### <https://rmarkdown.rstudio.com/r_notebooks.html>

# Packages and the Tidyverse

###### <https://tidyverse.org>

library(tidyverse)

# Hello World

This is my very first R Notebook.

My[[1]](#footnote-31) *favorite* ***search*** *engines* are

1. [Google](http://google.com)
2. [Yahoo](http://yahoo.com)
3. [Bing](http://bing.com)

###### <https://rmarkdown.rstudio.com/>

# Load Data

###### <https://support.rstudio.com/hc/en-us/articles/218611977-Importing-Data-with-RStudio>

###### <https://www.rstudio.com/resources/webinars/importing-data-into-r/>

### Hurricane data from github

## # A tibble: 249 x 16  
## order storm intensity casualties `damage (mn)` `peak wind`  
## <int> <chr> <chr> <dbl> <chr> <int>  
## 1 1 ONE td 0 0 30  
## 2 2 TWO td 0 0 35  
## 3 3 ALBERTO 3 0 0 125  
## 4 4 FOUR td 0 0 35  
## 5 5 BERYL ts 1.00 0.027 50  
## 6 6 CHRIS ts 0 0 40  
## 7 7 DEBBY 1 1.00 735 85  
## 8 8 ERNESTO ts 0 0 40  
## 9 9 NINE td 0 0 35  
## 10 10 FLORENCE 1 3.00 0 80  
## # ... with 239 more rows, and 10 more variables: `minimum pressure` <int>,  
## # start\_date <chr>, `end date` <chr>, `us affected` <chr>, US\_LAT <dbl>,  
## # US\_LON <dbl>, US\_STATE <chr>, COUNTRY\_LAT <dbl>, COUNTRY\_LON <dbl>,  
## # COUNTRY\_SPECIFIC <chr>

have a look at the data structures and data types

###### <http://r4ds.had.co.nz/tibbles.html>

class(canes)

## [1] "tbl\_df" "tbl" "data.frame"

###### <http://r4ds.had.co.nz/vectors.html>

glimpse(canes)

## Observations: 249  
## Variables: 16  
## $ order <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, ...  
## $ storm <chr> "ONE", "TWO", "ALBERTO", "FOUR", "BERYL", "...  
## $ intensity <chr> "td", "td", "3", "td", "ts", "ts", "1", "ts...  
## $ casualties <dbl> 0, 0, 0, 0, 1, 0, 1, 0, 0, 3, 26, 2, 1, 0, ...  
## $ `damage (mn)` <chr> "0", "0", "0", "0", "0.027", "0", "735", "0...  
## $ `peak wind` <int> 30, 35, 125, 35, 50, 40, 85, 40, 35, 80, 80...  
## $ `minimum pressure` <int> 1008, 1008, 950, 1009, 1007, 1008, 991, 100...  
## $ start\_date <chr> "6/7/2000", "6/23/2000", "8/3/2000", "8/8/2...  
## $ `end date` <chr> "6/8/2000", "6/25/2000", "8/23/2000", "8/11...  
## $ `us affected` <chr> "y", "n", "y", "n", "y", "n", "y", "n", "y"...  
## $ US\_LAT <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 30.0, NA, 2...  
## $ US\_LON <dbl> NA, NA, NA, NA, NA, NA, NA, NA, -94.0, NA, ...  
## $ US\_STATE <chr> NA, NA, NA, NA, NA, NA, NA, NA, "Texas", NA...  
## $ COUNTRY\_LAT <dbl> NA, NA, NA, NA, 24.9, NA, NA, NA, 30.0, NA,...  
## $ COUNTRY\_LON <dbl> NA, NA, NA, NA, -98.6, NA, NA, NA, -94.0, N...  
## $ COUNTRY\_SPECIFIC <chr> NA, NA, NA, NA, "Mexico", NA, NA, NA, "Unit...

# dplyr

* filter (rows)
* arrange (sort rows by variables)
* select (columns)
* mutate (change cell values)
* Count
* Summarize

###### <http://r4ds.had.co.nz/transform.html#introduction-2>

###### <http://dplyr.tidyverse.org/>

starwars

## # A tibble: 87 x 13  
## name height mass hair\_color skin\_color eye\_color birth\_year gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr>   
## 1 Luke Sk~ 172 77.0 blond fair blue 19.0 male   
## 2 C-3PO 167 75.0 <NA> gold yellow 112 <NA>   
## 3 R2-D2 96 32.0 <NA> white, bl~ red 33.0 <NA>   
## 4 Darth V~ 202 136 none white yellow 41.9 male   
## 5 Leia Or~ 150 49.0 brown light brown 19.0 female  
## 6 Owen La~ 178 120 brown, gr~ light blue 52.0 male   
## 7 Beru Wh~ 165 75.0 brown light blue 47.0 female  
## 8 R5-D4 97 32.0 <NA> white, red red NA <NA>   
## 9 Biggs D~ 183 84.0 black light brown 24.0 male   
## 10 Obi-Wan~ 182 77.0 auburn, w~ fair blue-gray 57.0 male   
## # ... with 77 more rows, and 5 more variables: homeworld <chr>,  
## # species <chr>, films <list>, vehicles <list>, starships <list>

## Filter

starwars\_small <- starwars %>%   
 filter(!is.na(mass), !is.na(height))

## Arrange

starwars\_small %>%   
 arrange(desc(height), desc(name)) %>%   
 head()

## # A tibble: 6 x 13  
## name height mass hair\_color skin\_color eye\_color birth\_year gender  
## <chr> <int> <dbl> <chr> <chr> <chr> <dbl> <chr>   
## 1 Tarfful 234 136 brown brown blue NA male   
## 2 Lama Su 229 88.0 none grey black NA male   
## 3 Chewbac~ 228 112 brown unknown blue 200 male   
## 4 Roos Ta~ 224 82.0 none grey orange NA male   
## 5 Grievous 216 159 none brown, wh~ green, ye~ NA male   
## 6 Tion Me~ 206 80.0 none grey black NA male   
## # ... with 5 more variables: homeworld <chr>, species <chr>, films <list>,  
## # vehicles <list>, starships <list>

## Select

starwars\_small %>%   
 arrange(species, height) %>%   
 select(name, species, height, mass, birth\_year)

## # A tibble: 59 x 5  
## name species height mass birth\_year  
## <chr> <chr> <int> <dbl> <dbl>  
## 1 Ratts Tyerell Aleena 79 15.0 NA   
## 2 Dexter Jettster Besalisk 198 102 NA   
## 3 Ki-Adi-Mundi Cerean 198 82.0 92.0   
## 4 Zam Wesell Clawdite 168 55.0 NA   
## 5 R2-D2 Droid 96 32.0 33.0   
## 6 R5-D4 Droid 97 32.0 NA   
## 7 C-3PO Droid 167 75.0 112   
## 8 IG-88 Droid 200 140 15.0   
## 9 Sebulba Dug 112 40.0 NA   
## 10 Wicket Systri Warrick Ewok 88 20.0 8.00  
## # ... with 49 more rows

starwars\_small %>%   
 select(name, species, height, mass, birth\_year) %>%   
 arrange(species, desc(height), mass, name) %>%   
 filter(species == "Human") %>%   
 arrange(birth\_year)

## # A tibble: 22 x 5  
## name species height mass birth\_year  
## <chr> <chr> <int> <dbl> <dbl>  
## 1 Luke Skywalker Human 172 77.0 19.0  
## 2 Leia Organa Human 150 49.0 19.0  
## 3 Wedge Antilles Human 170 77.0 21.0  
## 4 Biggs Darklighter Human 183 84.0 24.0  
## 5 Han Solo Human 180 80.0 29.0  
## 6 Lando Calrissian Human 177 79.0 31.0  
## 7 Boba Fett Human 183 78.2 31.5  
## 8 Lobot Human 175 79.0 37.0  
## 9 Darth Vader Human 202 136 41.9  
## 10 Anakin Skywalker Human 188 84.0 41.9  
## # ... with 12 more rows

## Mutate

starwars\_small %>%   
 select(name, species, height, mass, birth\_year) %>%   
 arrange(species, desc(height), mass, name) %>%   
 filter(species == "Human") %>%   
 mutate(BMI = round(height / mass, 2))

## # A tibble: 22 x 6  
## name species height mass birth\_year BMI  
## <chr> <chr> <int> <dbl> <dbl> <dbl>  
## 1 Darth Vader Human 202 136 41.9 1.49  
## 2 Dooku Human 193 80.0 102 2.41  
## 3 Qui-Gon Jinn Human 193 89.0 92.0 2.17  
## 4 Raymus Antilles Human 188 79.0 NA 2.38  
## 5 Anakin Skywalker Human 188 84.0 41.9 2.24  
## 6 Mace Windu Human 188 84.0 72.0 2.24  
## 7 Gregar Typho Human 185 85.0 NA 2.18  
## 8 Boba Fett Human 183 78.2 31.5 2.34  
## 9 Jango Fett Human 183 79.0 66.0 2.32  
## 10 Biggs Darklighter Human 183 84.0 24.0 2.18  
## # ... with 12 more rows

## Count

starwars %>%   
 count(mass) %>%   
 arrange(desc(n))

## # A tibble: 39 x 2  
## mass n  
## <dbl> <int>  
## 1 NA 28  
## 2 80.0 6  
## 3 79.0 4  
## 4 75.0 3  
## 5 77.0 3  
## 6 84.0 3  
## 7 32.0 2  
## 8 45.0 2  
## 9 48.0 2  
## 10 50.0 2  
## # ... with 29 more rows

## Summarize

starwars %>%   
 filter(!is.na(height)) %>%   
 group\_by(species) %>%   
 summarise(Count = n(), mean\_ht = mean(height), min\_ht = min(height),  
 max\_ht = max(height)) %>%   
 arrange(desc(Count))

## # A tibble: 38 x 5  
## species Count mean\_ht min\_ht max\_ht  
## <chr> <int> <dbl> <dbl> <dbl>  
## 1 Human 31 177 150 202   
## 2 Droid 4 140 96.0 200   
## 3 <NA> 4 160 96.0 183   
## 4 Gungan 3 209 196 224   
## 5 Kaminoan 2 221 213 229   
## 6 Mirialan 2 168 166 170   
## 7 Twi'lek 2 179 178 180   
## 8 Wookiee 2 231 228 234   
## 9 Zabrak 2 173 171 175   
## 10 Aleena 1 79.0 79.0 79.0  
## # ... with 28 more rows

# More information

<https://rfun.library.duke.edu/intro2r>

1. My Full Name Here ;- [↑](#footnote-ref-31)