Teaching data science with puzzles

useR! 2019

Irene Steves

i_steves isteves

bit.ly/ds-puzzles


Advent of Code

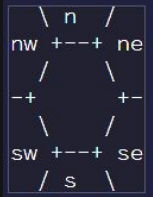
--- Day 11: Hex Ed ---

Crossing the bridge, you've barely reached the other side of the stream when a program comes up to you, clearly in distress. "It's my child process," she says, "he's gotten lost in an infinite grid!"

Fortunately for her, you have plenty of experience with infinite grids.

Unfortunately for you, it's a hex grid.

The hexagons ("hexes") in this grid are aligned such that adjacent hexes can be found to the north, northeast, southeast, south, southwest, and northwest:



You have the path the child process took. Starting where he started, you need to determine the fewest number of steps required to reach him. (A "step" means to move from the hex you are in to any adjacent hex.)

For example:

- `ne,ne,ne` is `3` steps away.
- `ne,ne,sw,sw` is `0` steps away (back where you started).
- `ne,ne,s,s` is `2` steps away (`se,se`).
- `se,sw,se,sw,sw` is `3` steps away (`s,s,sw`).

To begin, get your puzzle input.

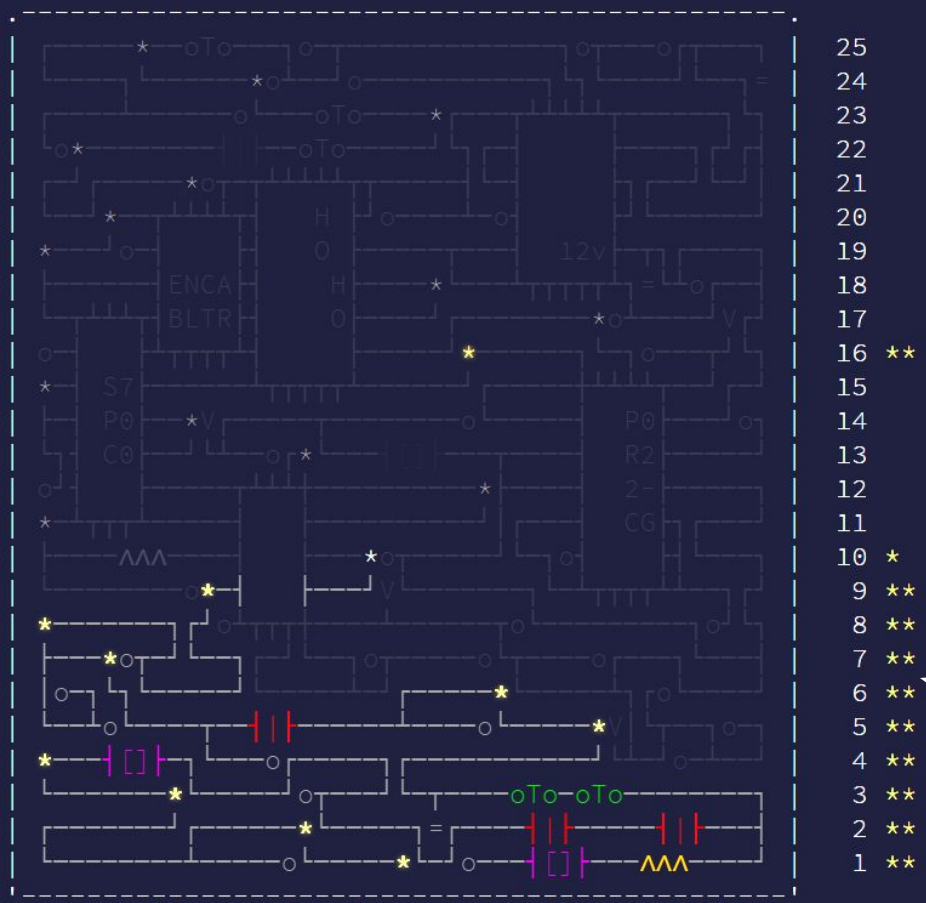
Answer: [Submit]

Puzzle text

Unique puzzle input

Answer submission

Advent of Code [About] [AoC++] [Events] [Settings] [Log Out] (anonymous user #288538) 21*
/^2017$/ [Calendar] [Leaderboard] [Stats] [Sponsors]



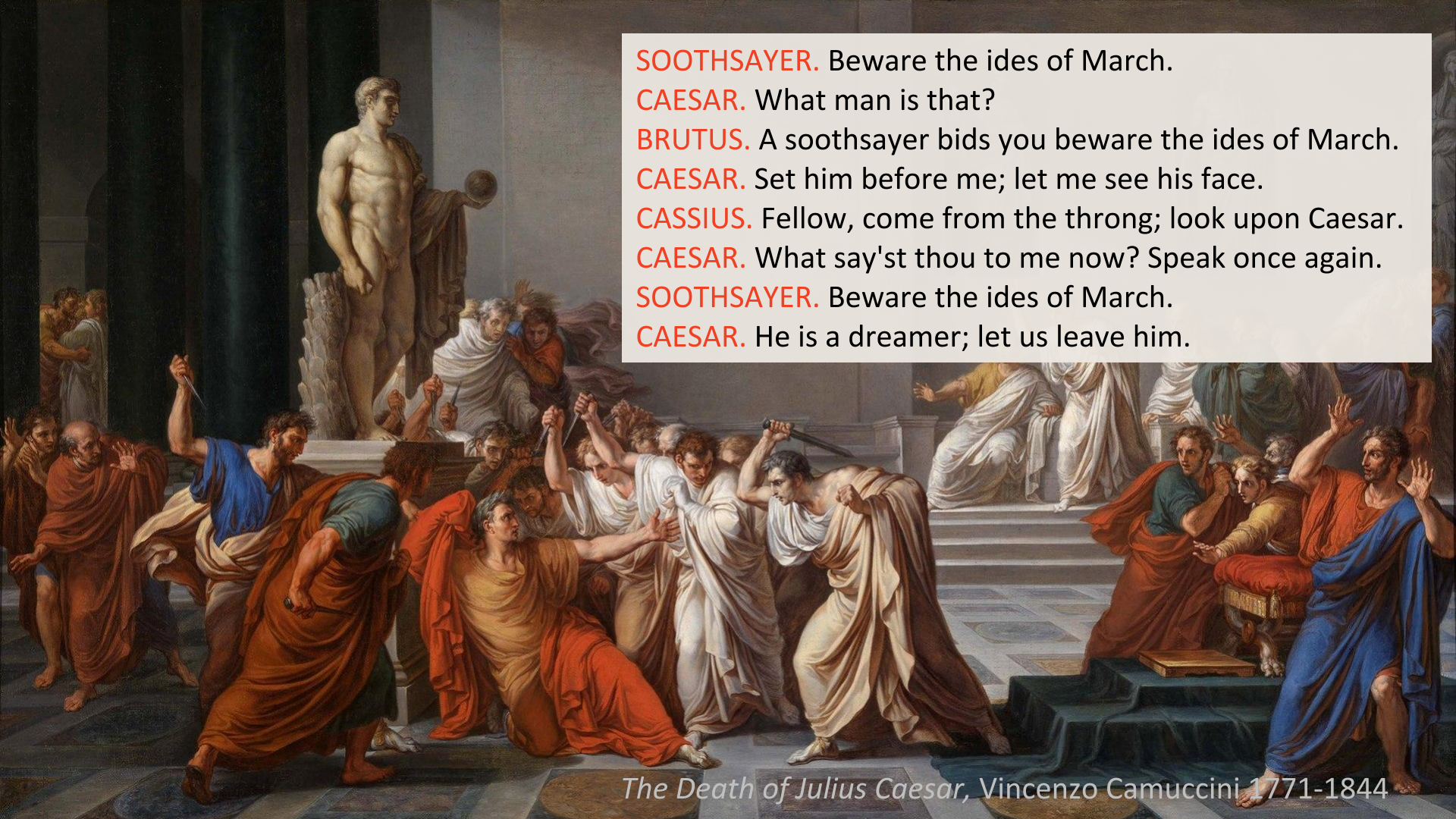
25
24
23
22
21
20
19
18
17
16 **
15
14
13
12
11
10 *
9 **
8 **
7 **
6 **
5 **
4 **
3 **
2 **
1 **

I solved these with R, but boy was it clunky!

Let’s make puzzles that highlight what R/the tidyverse are good at!

Bite-sized puzzles that focus on ***core data science skills*** as championed by the tidyverse set of packages

[still unreleased!]



SOOTHSAYER. Beware the ides of March.
CAESAR. What man is that?
BRUTUS. A soothsayer bids you beware the ides of March.
CAESAR. Set him before me; let me see his face.
CASSIUS. Fellow, come from the throng; look upon Caesar.
CAESAR. What say'st thou to me now? Speak once again.
SOOTHSAYER. Beware the ides of March.
CAESAR. He is a dreamer; let us leave him.

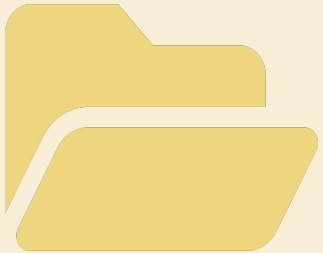
The Death of Julius Caesar, Vincenzo Camuccini 1771-1844

Photo: flickr clement127

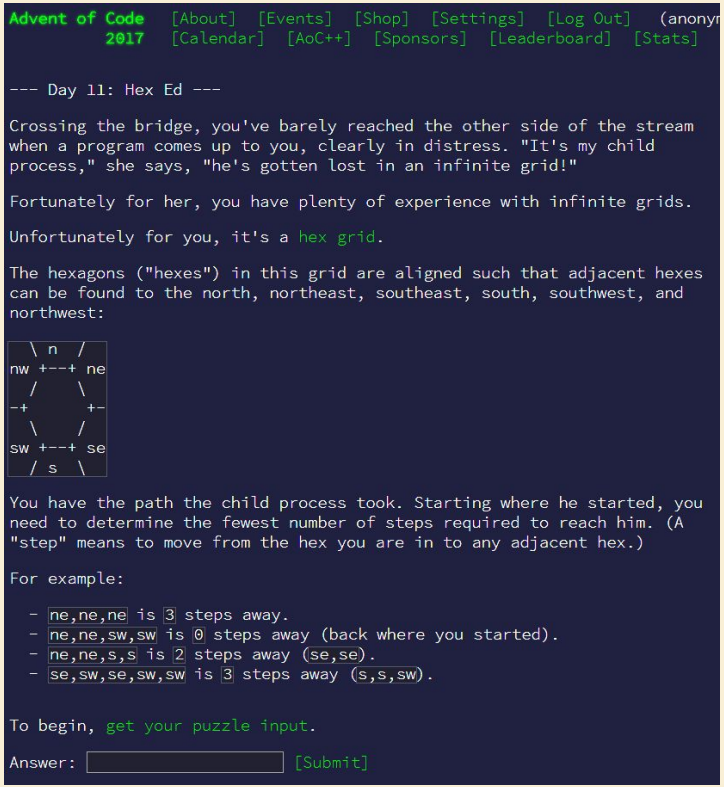
Wrangling

Wrangling

Workflow



Web-based experience



Advent of Code 2017 [About] [Events] [Shop] [Settings] [Log Out] (anonyr
[Calendar] [AoC++] [Sponsors] [Leaderboard] [Stats]

--- Day 11: Hex Ed ---

Crossing the bridge, you've barely reached the other side of the stream when a program comes up to you, clearly in distress. "It's my child process," she says, "he's gotten lost in an infinite grid!"

Fortunately for her, you have plenty of experience with infinite grids.

Unfortunately for you, it's a hex grid.

The hexagons ("hexes") in this grid are aligned such that adjacent hexes can be found to the north, northeast, southeast, south, southwest, and northwest:

```
  \ n  /
nw +--+ ne
  /    \
-+      +-
  \    /
sw +--+ se
  / s  \
```

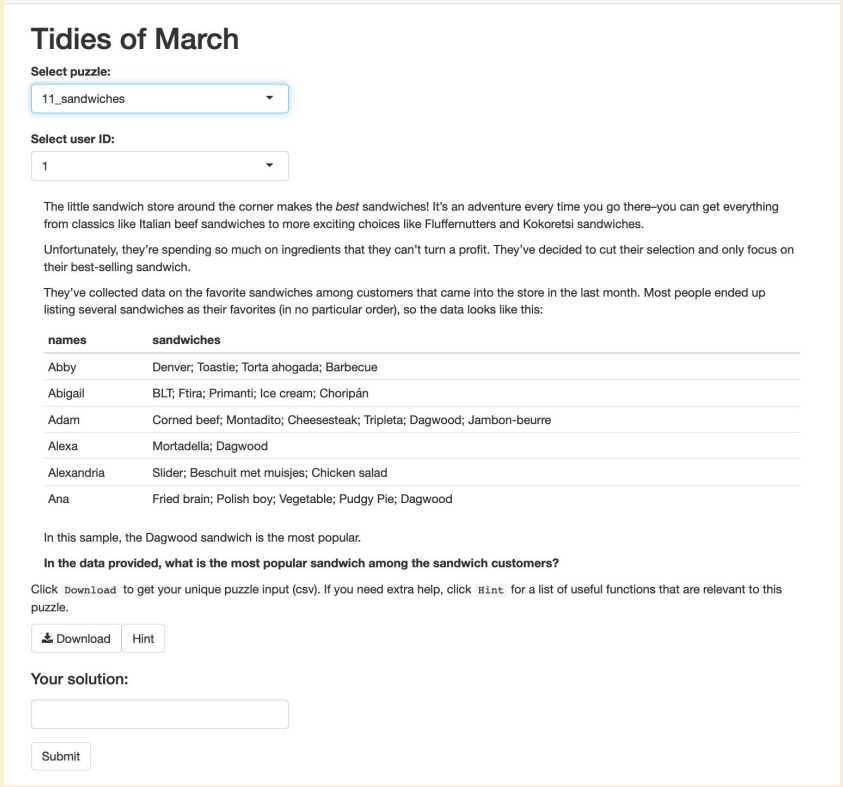
You have the path the child process took. Starting where he started, you need to determine the fewest number of steps required to reach him. (A "step" means to move from the hex you are in to any adjacent hex.)

For example:

- ne,ne,ne is 3 steps away.
- ne,ne,sw,sw is 0 steps away (back where you started).
- ne,ne,s,s is 2 steps away (se,se).
- se,sw,se,sw,sw is 3 steps away (s,s,sw).

To begin, get your puzzle input.

Answer: [Submit]



Tidies of March

Select puzzle:

11_sandwiches

Select user ID:

1

The little sandwich store around the corner makes the *best* sandwiches! It's an adventure every time you go there–you can get everything from classics like Italian beef sandwiches to more exciting choices like Fluffernutters and Kokoretsi sandwiches.

Unfortunately, they're spending so much on ingredients that they can't turn a profit. They've decided to cut their selection and only focus on their best-selling sandwich.

They've collected data on the favorite sandwiches among customers that came into the store in the last month. Most people ended up listing several sandwiches as their favorites (in no particular order), so the data looks like this:

names	sandwiches
Abby	Denver; Toastie; Torta ahogada; Barbecue
Abigail	BLT; Ftira; Primanti; Ice cream; Choripán
Adam	Corned beef; Montadito; Cheesesteak; Tripleta; Dagwood; Jambon-beurre
Alexa	Mortadella; Dagwood
Alexandria	Slider; Beschuit met muisjes; Chicken salad
Ana	Fried brain; Polish boy; Vegetable; Pudgy Pie; Dagwood

In this sample, the Dagwood sandwich is the most popular.

In the data provided, what is the most popular sandwich among the sandwich customers?

Click `Download` to get your unique puzzle input (csv). If you need extra help, click `Hint` for a list of useful functions that are relevant to this puzzle.

Download Hint

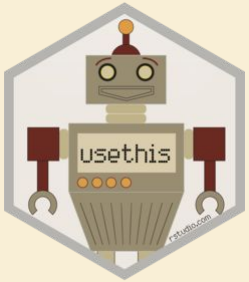
Your solution:

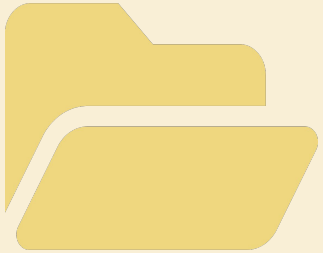
Submit

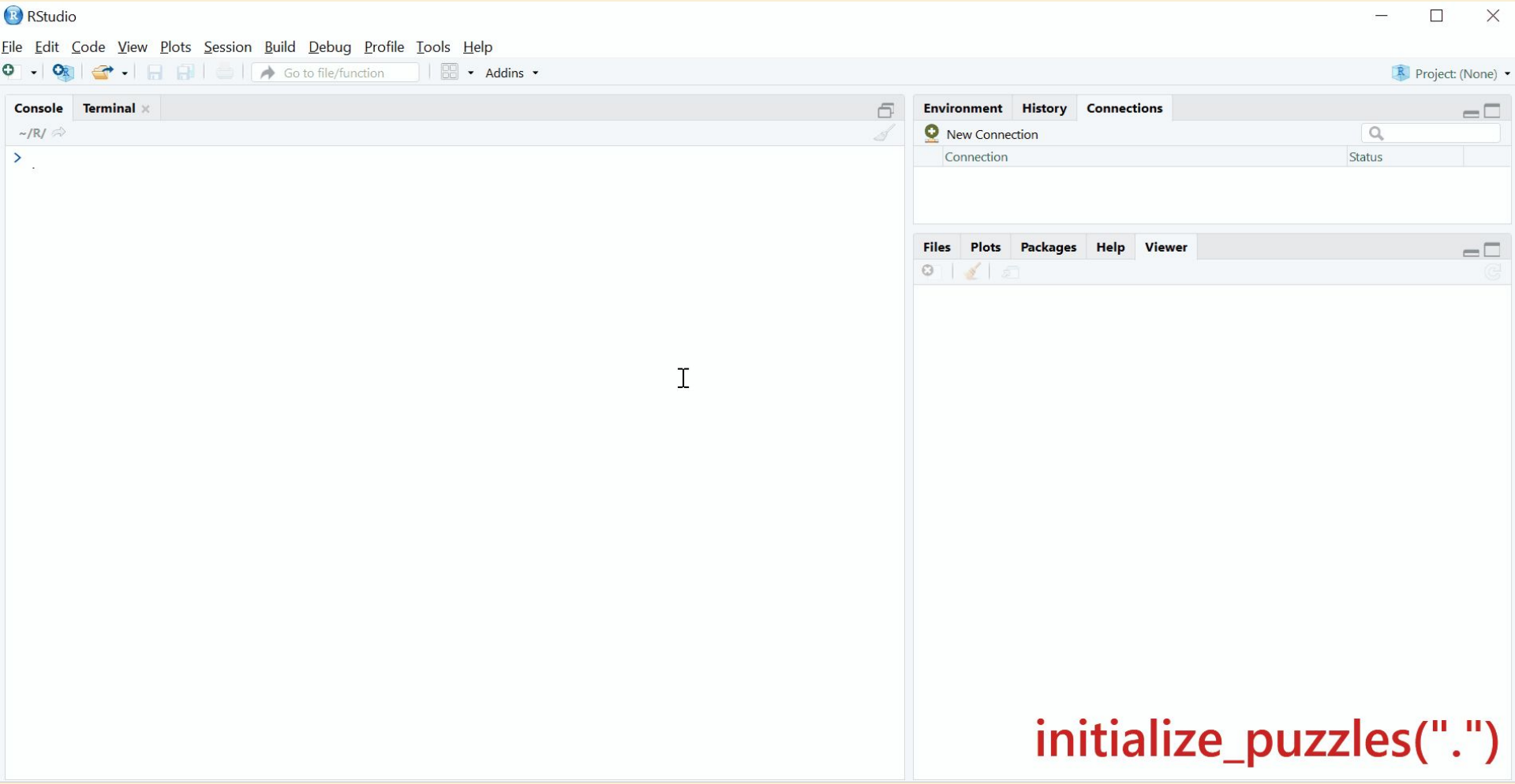
Language & platform agnostic

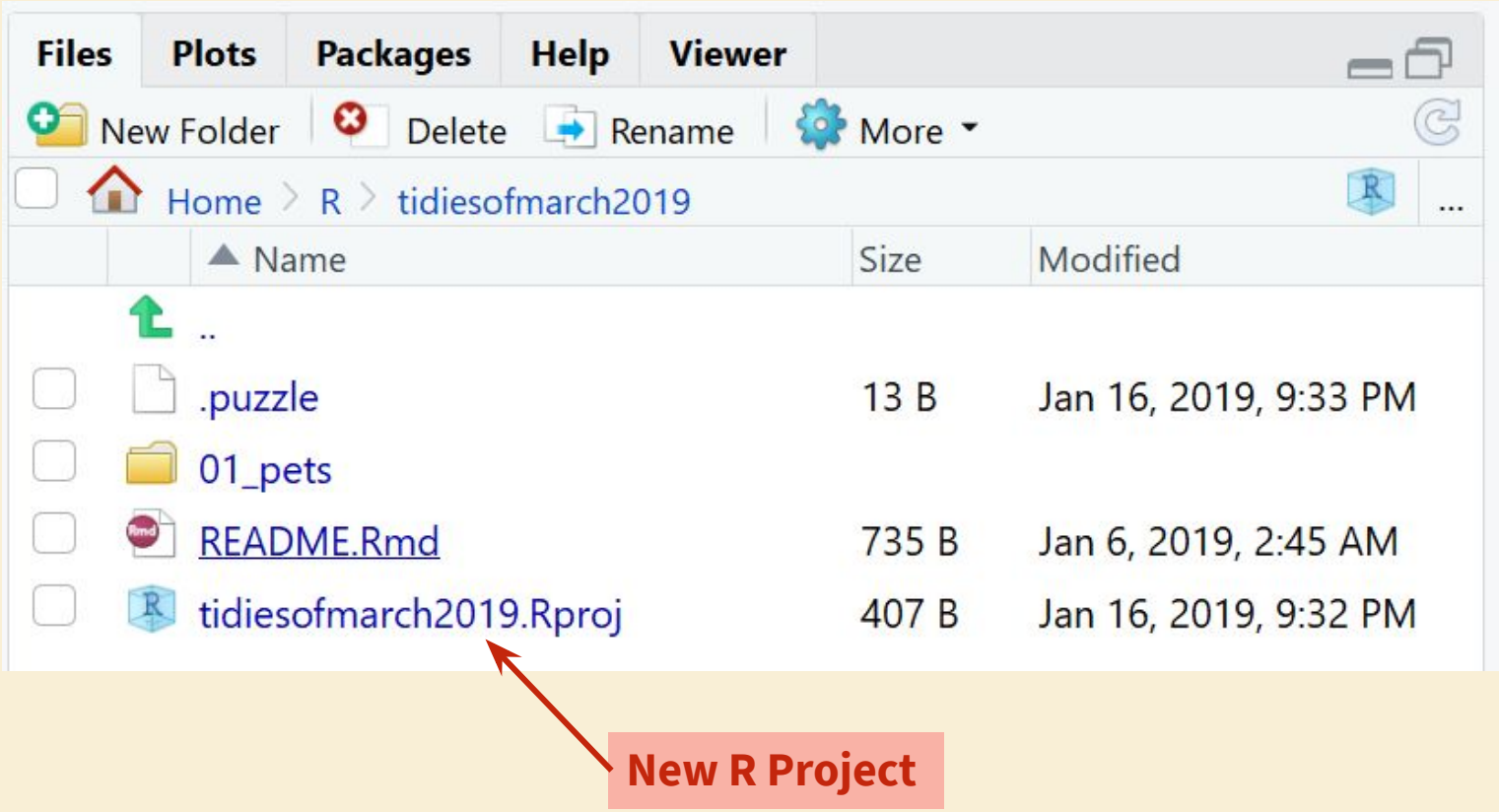
Workflow

R-mediated experience






RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Project: (None)
Console
Terminal
~/R/
Environment
History
Connections
New Connection
Connection
Status
Files
Plots
Packages
Help
Viewer
initialize_puzzles(".")


Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
R
tidiesofmarch2019
Name
Size
Modified
..
.puzzle
13 B
Jan 16, 2019, 9:33 PM
01_pets
README.Rmd
735 B
Jan 6, 2019, 2:45 AM
tidiesofmarch2019.Rproj
407 B
Jan 16, 2019, 9:32 PM
New R Project

README.Rmd

Knit

```
---
title: "Tidies of March 2019"
output: html_document
---

```{r setup, include=FALSE}
library(tidyverse)
library(fs)
library(here)
library(glue)
```

Table of Contents

```{r echo = FALSE, results = "asis"}
puzzle_dirs <- dir_ls(".", type = "directory", regexp = "^[0-9]{2}_")

puzzle_no <- str_extract(puzzle_dirs, "^[0-9]{2}(?=_)")
puzzle_name <- str_extract(puzzle_dirs, "(?<=_)[a-z-]+")

glue("{puzzle_no}.
[{puzzle_name}]({puzzle_dirs}/{puzzle_no}_solution.R)")
```
```

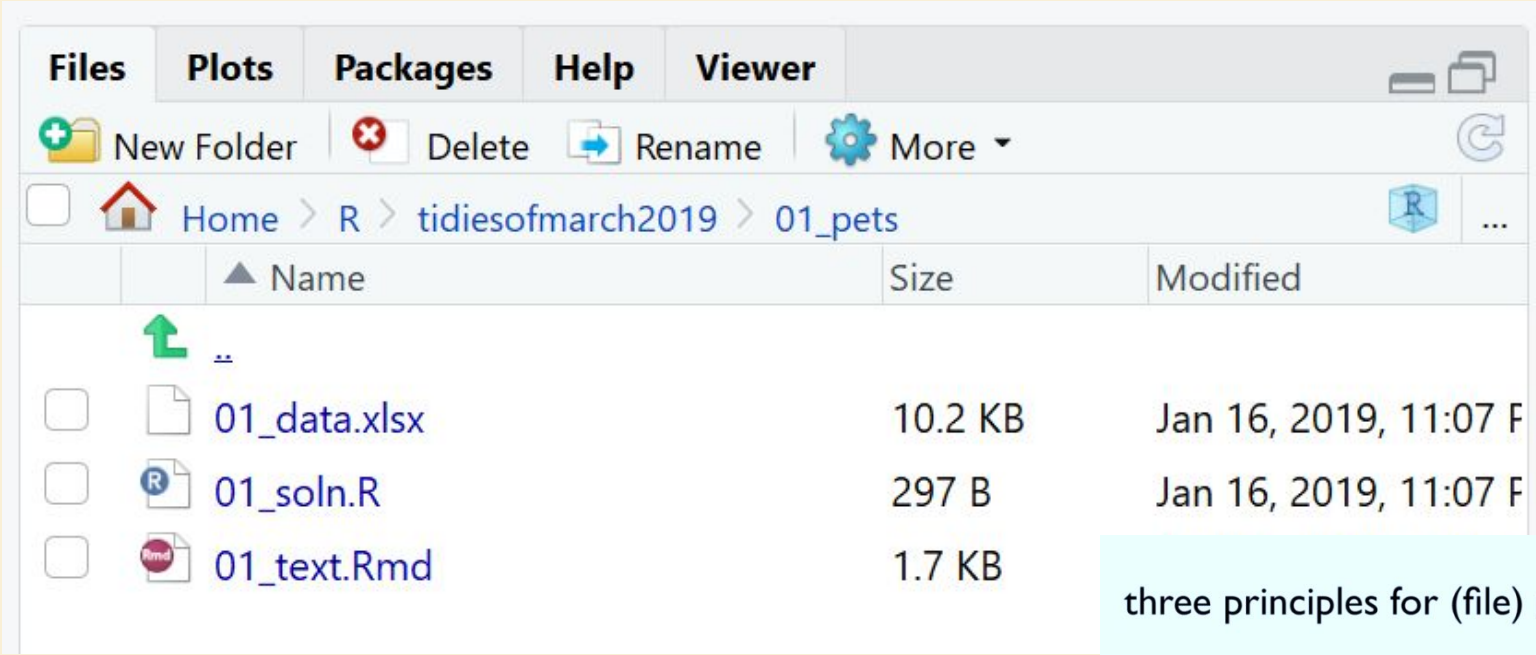
Files | Plots | Packages | Help | Viewer

Tidies of March 2019

Table of Contents

1. pets

Auto-generated table of contents



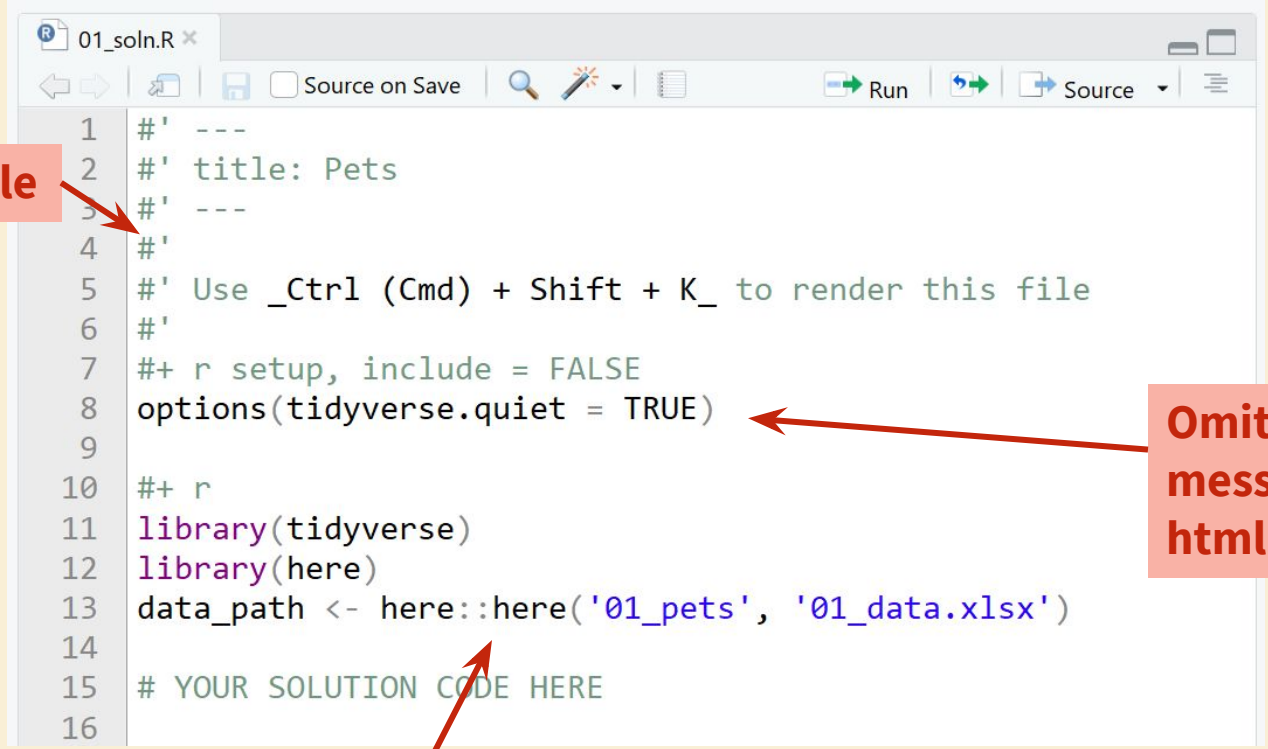
three principles for (file) names

machine readable

human readable

plays well with default ordering

@JennyBryan


01_soln.R
Source on Save
Run
Source
#' ---
#' title: Pets
#' ---
#'
#' Use _Ctrl (Cmd) + Shift + K_ to render this file
#'
#+ r setup, include = FALSE
options(tidyverse.quiet = TRUE)
#+ r
library(tidyverse)
library(here)
data_path <- here::here('01_pets', '01_data.xlsx')
# YOUR SOLUTION CODE HERE
Knittable .R file
Omit tidyverse messages from html output
Paths that work in the console & when rendered

The neighborhood deli makes amazing sandwiches--from classics like BLTs to dessert sandwiches like Fluffernutters. Since many of their specialty ingredients keep going bad, they've decided to cut their selection and focus on their best-selling sandwich.

Photo: flickr skywhisperer

To help with the decision, the storeowners collected data on their customers' favorites. Most people listed several varieties (in no particular order). Here's a sample of the data:

names	sandwiches
Abby	Denver; BLT; Torta ahogada; Barbecue
Abigail	BLT; Ftira; Primanti; Ice cream; Choripán
Adam	Corned beef; Montadito; Cheesesteak; Tripleta; Dagwood; Jambon-beurre
Alexa	Dagwood; Mortadella
Alexandria	Slider; Beschuit met muisjes; Chicken salad
Ana	Fried brain; Polish boy; Vegetable; Pudgy Pie; Dagwood

In this sample, the Dagwood sandwich is the most popular.

In the full dataset, **what is the most popular sandwich among the customers?**

names	sandwiches
Abby	Denver; BLT; Torta ahogada; Barbecue
Abigail	BLT; Ftira; Primanti; Ice cream; Choripán
Adam	Corned beef; Montadito; Cheesesteak; Tripleta; Dagwood; Jambon-beurre
Alexa	Dagwood; Mortadella
Alexandria	Slider; Beschuit met muisjes; Chicken salad
Ana	Fried brain; Polish boy; Vegetable; Pudgy Pie; Dagwood

test case

In this sample, the Dagwood sandwich is the most popular.

```
sw
```

```
## # A tibble: 6 x 2
##   names      sandwiches
##   <chr>      <chr>
## 1 Abby       Denver; BLT; Torta ahogada; Barbecue
## 2 Abigail    BLT; Ftira; Primanti; Ice cream; Choripán
## 3 Adam       Corned beef; Montadito; Cheesesteak; Tripleta; Dagwood; Jamb
## 4 Alexa      Dagwood; Mortadella
## 5 Alexandria Slider; Beschuit met muisjes; Chicken salad
## 6 Ana        Fried brain; Polish boy; Vegetable; Pudgy Pie; Dagwood
```

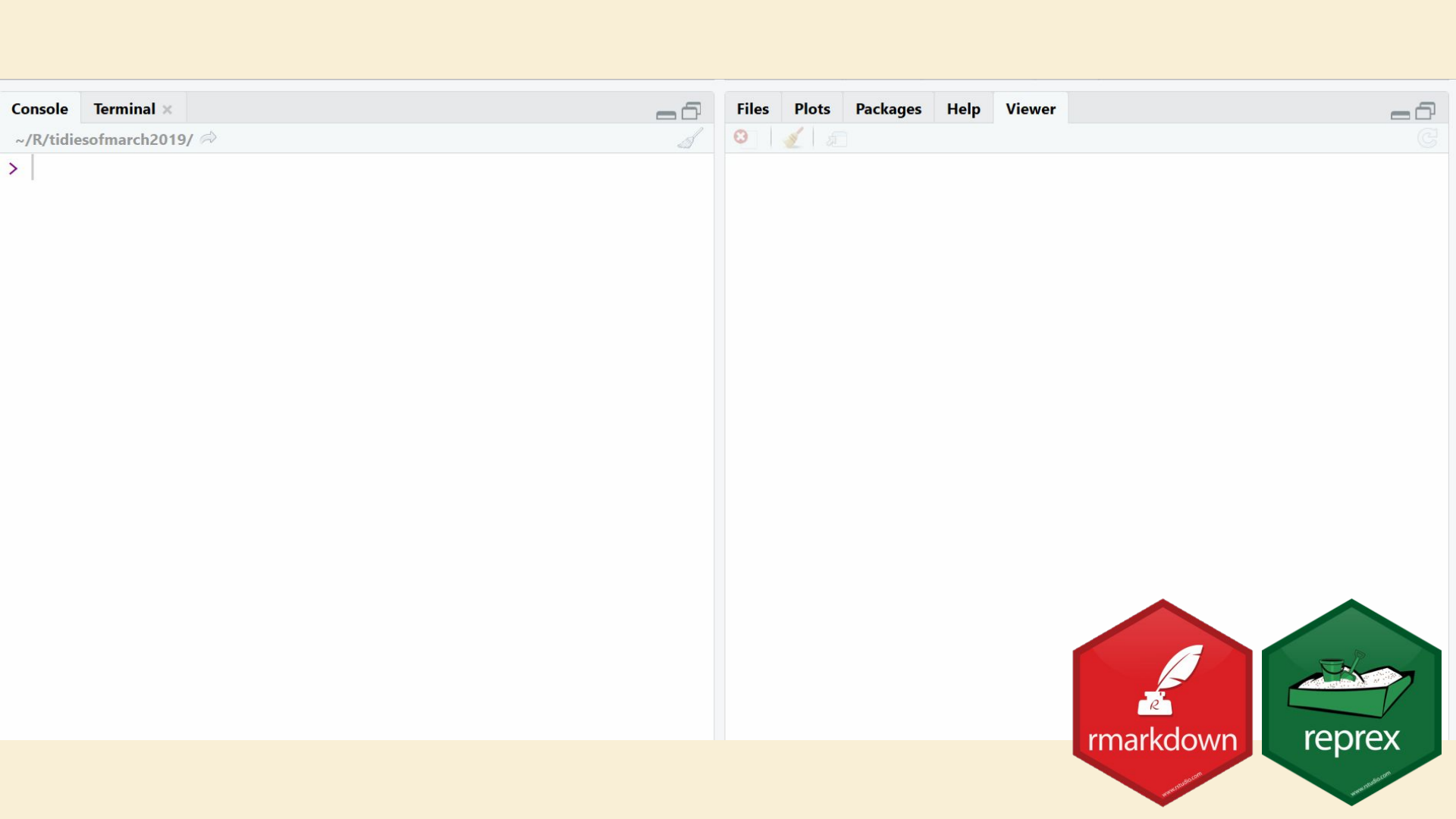
```
sw %>%
  separate_rows(sandwiches, sep = "; ")
```

```
## # A tibble: 25 x 2
##    names   sandwiches
##    <chr>   <chr>
##  1 Abby    Denver
##  2 Abby    BLT
##  3 Abby    Torta ahogada
##  4 Abby    Barbecue
##  5 Abigail BLT
##  6 Abigail Ftira
##  7 Abigail Primanti
##  8 Abigail Ice cream
##  9 Abigail Choripán
## 10 Adam    Corned beef
## # ... with 15 more rows
```

```
sw %>%
  separate_rows(sandwiches, sep = "; ") %>%
  count(sandwiches, sort = TRUE)
```

```
## # A tibble: 22 x 2
##    sandwiches                n
##    <chr>                 <int>
##  1 Dagwood                   3
##  2 BLT                       2
##  3 Barbecue                  1
##  4 Beschuit met muisjes      1
##  5 Cheesesteak               1
##  6 Chicken salad             1
##  7 Choripán                  1
##  8 Corned beef               1
##  9 Denver                    1
## 10 Fried brain               1
## # ... with 12 more rows
```

Photo: Wikipedia


Console
Terminal
~/R/tidiesofmarch2019/
Files
Plots
Packages
Help
Viewer
rmarkdown
reprex


tidyverse

Beyond the tidyverse

Projects & version control

Consistent and parseable names

Test cases

Self-contained code

Thank you!

Irene Steves

 i_steves  isteves

bit.ly/ds-puzzles