Chapter 3

Wrangling data

We've gone through different methods of bringing in various types of data sets into R.

Now we're going to learn the proper methods of transforming and analyzing the data.

We'll be using the **dplyr** and **tidyr** packages.

With verb-like names for functions, **dplyr** and **tidyr** turn data analysis into a sort-of grammar exercise and avoids the mess of nested functions that can tangle users in Base R.

This lends itself to better documentation because others will be able to read your code later on and follow what you were trying to accomplish.

Goals

- Transforming and analyzing data
    - Learn and implement data wrangling verbs
- Tidying and joining data
    - Learn how to tidy up and join data
- Murders case study
    - Looking for evidence of serial killers
- Handling strings
    - Strategies on dealing with strings in data
- Dealing with dates
    - Introduction on how to deal with dates

Files and folders

The repo for this class is on Github, but can be easily downloaded to your desktop with the following commands:

```
install.packages("usethis")
```

```
usethis::use_course("https://github.com/r-journalism/learn-chapter-3/archive/master.zip")
```

Test yourself

There are links to exercise what you've learned spread through out this section.

It's possible to run these files locally to test yourself if you've downloaded the files for the chapter as instructed above.

Make sure your project directory is correct and then run these lines in the console:

```
install.packages("learnr")
install.packages("rmarkdown")
install.packages("tidyverse")
```

and then

```
rmarkdown::run("chapter-3/index.Rmd")
```

Readings

- Text analysis of Trump tweets confirms he only writes anger ones -Variance Explained
- Tidy Data - Hadley Wickham
- Serial Killers Should Fear This Algorithm - Bloomberg
- When algorithms decide what you pay - ProPublica
- Doctors and Sex Abuse - Atlanta Journal-Constitution