

The Lesser Known ⭐S of the Tidyverse

Emily Robinson

@robinson_es



About Me

- Data Analyst at Etsy
- R User for ~6 years
- Enjoy talking about:
 - A/B Testing
 - Building and finding Data Science community
 - R

Disclaimers

**This talk represents my
own views, not those
of Etsy**

It's not Base R vs. Tidyverse



Hadley Wickham @hadleywickham · 30 Aug 2017

Please use as much or as little of the [#tidyverse](#) as you feel useful. I want to be as effective in [#rstats](#) as possible

15

50

306



Hadley Wickham

@hadleywickham

Following

You can not use [#tidyverse](#) without base R. It's not a dichotomy. Pick the tools that make you most effective.

8:52 PM - 30 Aug 2017

Talk Goals

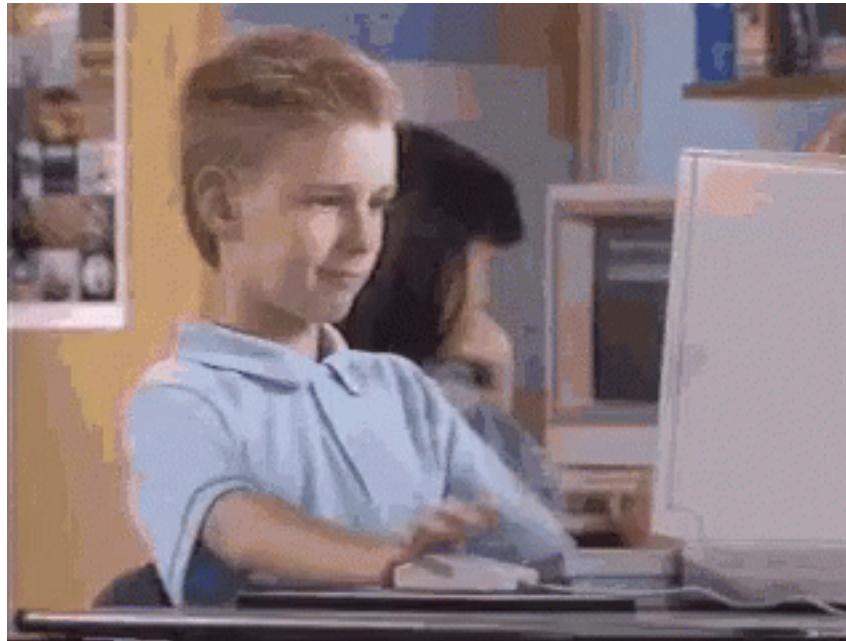
1. Keep you hip to the lingo



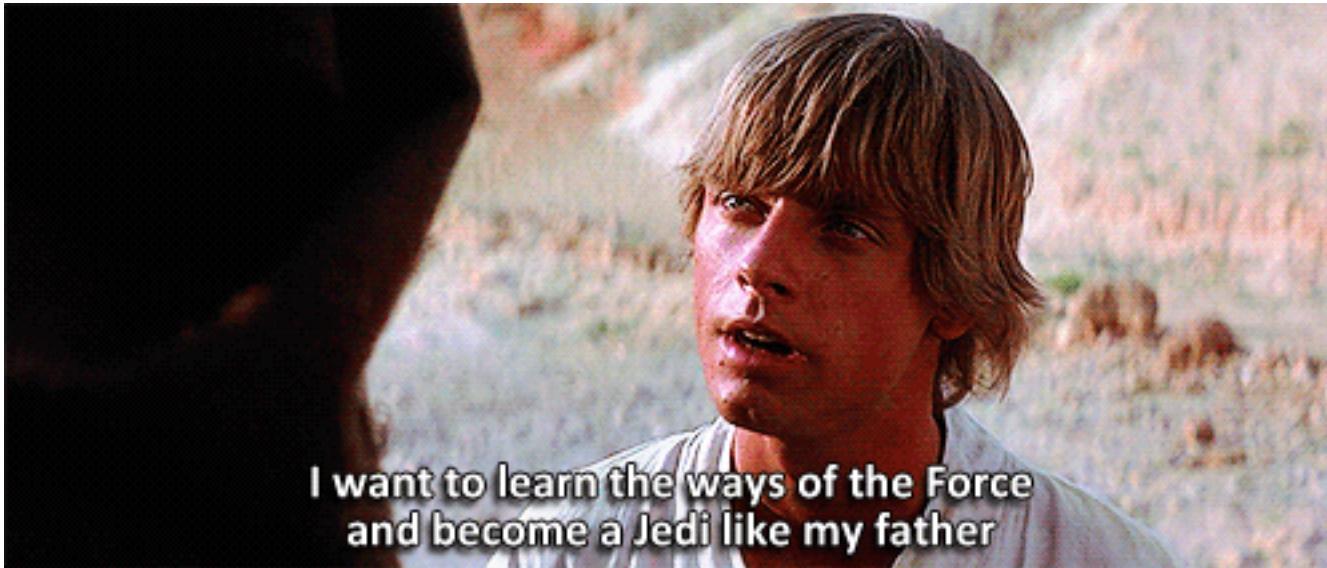
2. Stop you from doing this ...



... by sharing useful functions



3. Point you to resources



I want to learn the ways of the Force
and become a Jedi like my father

The Tidyverse

Coherent system of packages for data manipulation, exploration and visualization that share a common design philosophy.

Import

readr
readxl
haven
xml2

Tidy → Transform

tibble
tidyr

dplyr
forcats
hms

lubridate
stringr

purrr
magrittr

Program

Visualise

ggplot2

Model

broom
modelr

Communicate

shiny
rmarkdown

Hadley Wickham, the Man Who Revolutionized R

· 52,351 views · More stats

 Share

 Like 191

 Tweet



If you don't spend much of your time coding in the open-source statistical programming language **R**, his name is likely not familiar to you -- but the statistician Hadley Wickham is, in his own words, "nerd famous." The kind of famous where people at statistics conferences line up for selfies, ask him for autographs, and are generally in awe of him. "It's utterly utterly bizarre," he admits. "To be famous for writing R programs? It's just crazy."

CODE

Most of my work is in the form of open source R code, which you can find on [my github](#). You can roughly divide my work into three categories: tools for data science, tools for data import, and software engineering tools.

DATA SCIENCE

- `ggplot2` for visualising data.
- `dplyr` for manipulating data.
- `tidyverse` for tidying data.
- `stringr` for working with strings.
- `lubridate` for working with date/times.

DATA IMPORT

- `readr` for reading .csv and fwf files.
- `readxl` for reading .xls and .xlsx files.
- `haven` for SAS, SPSS, and Stata files.
- `httr` for talking to web APIs.
- `rvest` for scraping websites.
- `xml2` for importing XML files.

SOFTWARE ENGINEERING

- `devtools` for general package development.
- `roxygen2` for in-line documentation.
- `testthat` for unit testing

Tidyverse

=



?

Tidyverse

=



!



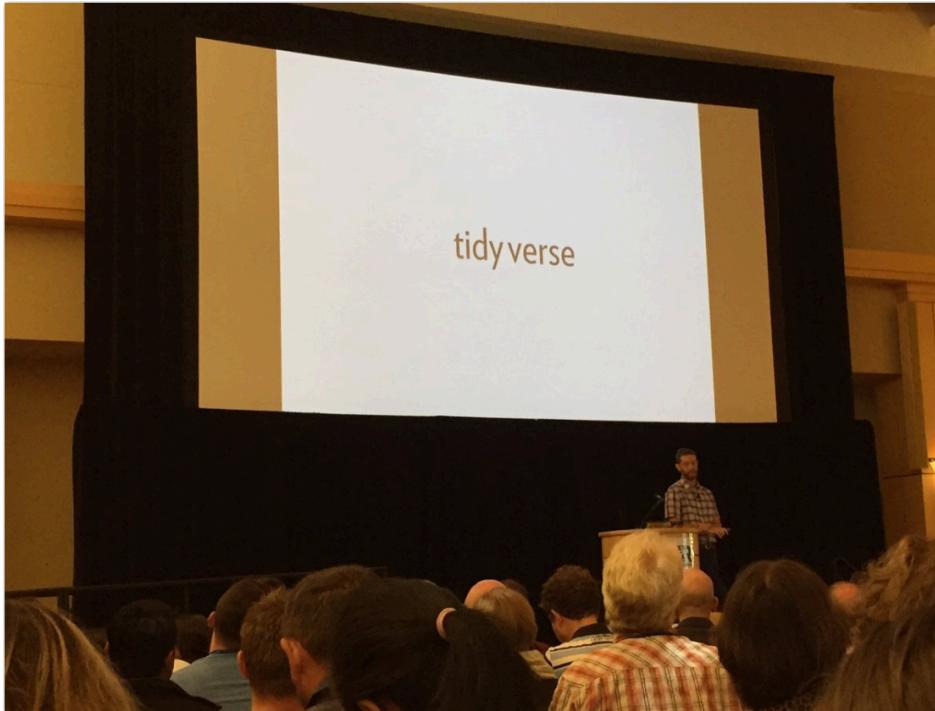
David Robinson

@drob

Following



.@hadleywickham proposes we stop saying
"Hadleyverse", start saying "tidyverse"
#useR2016 #rstats





R4DS
Book



Tidyverse != Hadleyverse



Tidyverse != Hadleyverse

Many other contributors





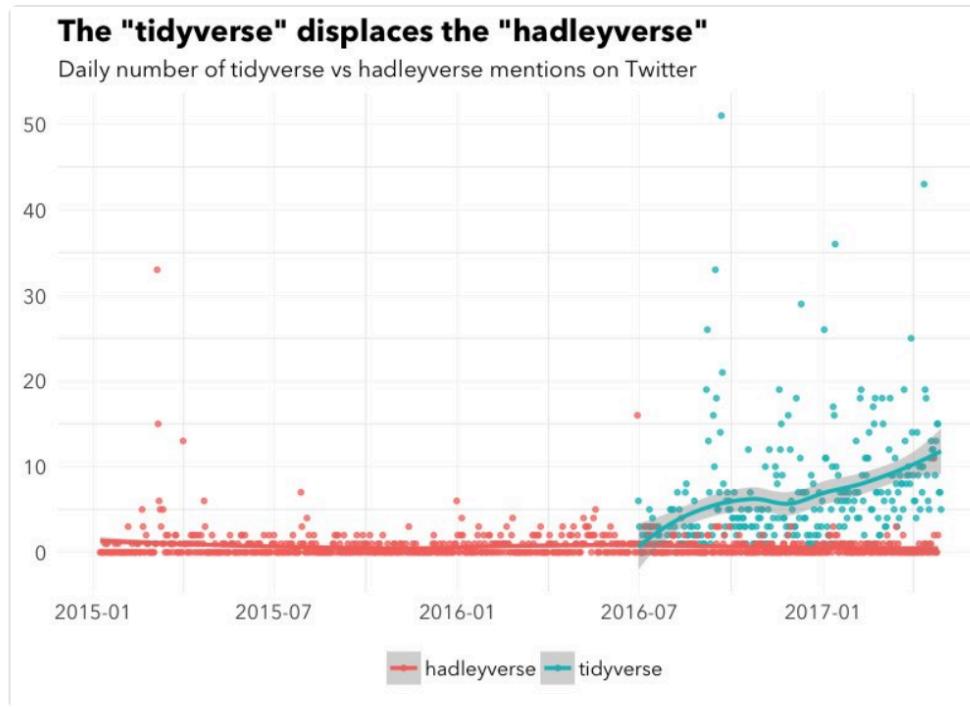
Mike Kearney 

@kearneymw

Following



"Hadleyverse" vs "tidyverse" mentions on Twitter. #rstats



6:35 AM - 19 Sep 2017

Demonstration

2017 The State of Data Science & Machine Learning

This year, for the first time, we conducted an industry-wide survey to establish a comprehensive view of the state of data science and machine learning. We received over **16,000 responses** and learned a ton about who is working with data, what's happening at the cutting edge of machine learning across industries, and how new data scientists can best break into the field. The below report shares some of our key findings and includes interactive visualizations so you can easily cut the data to find out exactly what you want to know. Here are some sample takeaways:

Review

Dplyr

na_if
count
select_if
n_distinct

Other

readr::read_csv
skimr::skim
%<>%
purrr::map_df
tidyr::gather
tidyr::unnest
stringr::str_split

Forcats

fct_infreq
fct_relevel
fct_reorder

Resources

Welcome



R for Data Science

1 Introduction

I Explore

2 Introduction

3 Data visualisation

4 Workflow: basics

5 Data transformation

6 Workflow: scripts

7 Exploratory Data Analysis

8 Workflow: projects

II Wrangle

9 Introduction

10 Tibbles

11 Data import

12 Tidy data

13 Relational data

R for Data Science

Garrett Grolemund

Hadley Wickham

Welcome

This is the website for “**R for Data Science**”. This book will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, visualise it and model it. In this book, you will find a practicum of skills for data science. Just as a chemist learns how to clean test tubes and stock a lab, you’ll learn how to clean data and draw plots—and many other things besides. These are the skills that allow data science to happen, and here you will find the best practices for doing each of these things with R. You’ll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. You’ll also learn how to manage cognitive resources to facilitate discoveries when wrangling, visualising, and exploring data.

#rstats Twitter



Emily Robinson

@robinson_es



#rstats twitter: is there a way to globally set the color scale for ggplot2? theme_set()
seems to work only for background.

7:15 AM - 20 Dec 2017

#rstats Twitter



Ildi Czeller

@czeildi

Following



Replying to @robinson_es

Not native ggplot2, but I use {ggthemr}, it provides define_palette,
ggthemr(your_palette) to set palette for all plots and ggthemr_reset to go back to defaults. Concrete example:

ildiczeller.com/2017/10/15/cus...

community.rstudio.com/c/tidyverse



This category is for anything and everything about the tidyverse.

■ tidyverse ▶ all tags ▶ **Latest** New (10) Unread Top + New Topic O

Topic	Users	Replies	Views	Activity
<input checked="" type="checkbox"/> <input type="checkbox"/> About the tidyverse category This category is for anything and everything about the tidyverse. If you're asking a question about code, please make sure to include a reprex!		0	305	Aug '17
<input type="checkbox"/> All.equal on tibbles ignores attributes? • new	V	4	29	1h
<input checked="" type="checkbox"/> How do you do time-series forecasting? Best practices, tidy ways etc • new forecasting, ets		8	26	2h
<input checked="" type="checkbox"/> Plotting a number of plots by looping using ggplot2 • new ggplot2	J	4	27	2h

All Data Science Courses



Introduction to the Tidyverse

Get started on the path to exploring and visualizing your own data with the tidyverse, a powerful and popular collect...

🕒 4 hours

▶ Play preview



DAVID ROBINSON

Data Scientist, Stack Overflow



Communicating with Data in the Tidyverse

Leverage the power of tidyverse tools to create publication-quality graphics and custom-styled reports that communica...

🕒 4 hours



TIMO GROSSENBACHER

Data Journalist at SRF Data



Working with Dates and Times in R

Learn the essentials of parsing, manipulating and computing with dates and times in R.

🕒 4 hours

▶ Play preview



CHARLOTTE WICKHAM

Assistant Professor at Oregon State University

And more!

- Tidyverse.org
- <https://www.rstudio.com/resources/cheatsheets/>
- <https://medium.com/@kierisi/r4ds-the-next-iteration-d51e0a1b0b82>
- <http://tiny.cc/hadleyvid>

Conclusion



Christine Zhang @christinezhang · Jun 5

TIL that there are packages called "forcats" & "purrr" in @hadleywickham's tidyverse 🐈🐈🐈 #rstats #hadleyverse tidyverse.org

The tidyverse

Come for the stickers
and package names ...

Stay for the friendly
community and happy
workflow.

Components



#rstats Twitter



Emily Robinson @robinson_es · 24 Aug 2017

Yay! This is for presentation on the lesser known stars of the tidyverse. What do you think of topic? Internal for now, considering external

4



3



Hadley Wickham ✅

@hadleywickham

Following



Replying to @robinson_es @jimhester_

You should submit it to rstudio::conf!

2:14 PM - 24 Aug 2017

#rstats Twitter



Jim Hester @jimhester_ · 24 Aug 2017

Maybe this calls for a Tidyverse face collage that forms a picture of
@hadleywickham

3



12



Maëlle Salmon 🐟

@ma_salmon

Following

Replies to @jimhester_ @robinson_es and 6 others

Lucy now I expect you to download the
Github avatars of all tidyverse contributors, I
know you can make pretty R collages 😊

10:38 PM - 24 Aug 2017



Jim Hester

@jimhester_

Following



Replying to @ma_salmon @robinson_es and 6 others

I'll just leave this here...

[gist.github.com/jimhester/c408 ...](https://gist.github.com/jimhester/c408...)



7:58 AM - 25 Aug 2017

Thank You!



tiny.cc/datadaytalk



robinsones.github.io



@robinson_es

