TEACHING THE TIDYVERSE TO *COWORKERS*

Erin Grand

David Robinson

Chief Data Scientist at DataCamp, works in R and Python.

Email
Twitter
Github

Teach the tidyverse to beginners

A few years ago, I wrote a post Don't teach built-in plotting to beginners (teach ggplot2). I argued that ggplot2 was not an advanced approach meant for experts, bu rather a suitable introduction to data visualization.

> *Many teachers suggest I'm overestimating their students: "No, see, my students are beginners...". If I push the point, they might insist I'm not understanding just how much of a beginner these students are, and emphasize they're looking to keep it simple and teach the basics, and that that students can get to the advanced methods later....*

> ***My claim is that this is precisely backwards.*** *ggplot2 is easier to teach beginners, not harder, and makes constructing plots simpler, not more complicated.*

TEACH THE TIDYVERSE TO ~~BEGINNERS~~ COWORKERS

Preparing for Good PD

1. Tools to measure current skill level 🛠️
2. Structures to produce PD sessions that align with skill level growth 📖
3. Create PD Materials 💡
4. Plan for practice / homework 📚
5. Feedback 💬

Tools to measure current skill level

	New		Beginner				Proficient
	1	2	3	4	5	6	7
Package List		readr stringr (very basics) datapasta (addin)	dplyr	dplyr (continued) janitor databases: DBI, ODBC, RODC (connection basics) WriteXL (requires preparing of computer)	dplyr tidyr	dplyr tidyr purrr (map_dfr) assertr here	stringr purrr
Functionand Skills List	- Syntax (i.e. use of <- vs =) - Data Types - Operators & Comparative operators - Loading R packages - Calling functions - Naming variables - Basic indexing - Rstudio workflow - Commenting	- Reading in files as data frames - Factors vs strings - Dates and Times (basics) - Changing between data types - Creation of lookup table - Check for missing values	- select - filter - mutate - group by & summarize - *_join - pipe	- get_dupes - clean_names - arrange - group by & mutate - Connect to a SQL database and query the database in R - case_when - WriteXL	- spread & gather - unite & separate - mutate_* statements - Start writing functions	- map_dfr - assert & verify - Rnotebook parameters	- string functions - basic regular expressions - Work with lists and lists of lists
Workflow and Style	Introduction to the Rstudio IDE. Install and update packages inside Rstudio and be able to identify where new objects are shown and where to search for help. Inside a Rnotebook, write both R code and real text in the same file, knowing how and when to separate R chunks.	- Basic understanding of difference between Rmarkdown and .R script. - Continue to include comments in text and code to explain business rules or decisions made along the way. - Best practices for Project structure. (Input, Output, Scripts...etc) - Writing cleaned data files to an output directory - Computer literacy around file paths and basic coding principles.	Write R script to produce a full analysis without errors. Style of code should follow the Tidyverse style guide. http://style.tidyverse.org/	Script should be using pipes all the way through, and obey the general rules of pipes and tidyverse styling.	Understand tidy data principles, and be able to visualize the form the data needs to take to proceed (long or wide). Don't repeat code more than twice, use functions to replace this repetition. Organize functions and code with comments explaining choices made.	Use best practices around SQL queries in Rstudio. Use R projects to house data work	Write unit tests into script.

Rubric measure skills in 3 categories (packages, functions and skills, and workflow)
and
covers Beginner to Proficient levels

Structures to produce PD sessions that align with skill level growth 📖

Use a scope and sequence to allow for detailed planning of a topic.

- Level
- Topic
- Skill
- Outcome goals
- Practice actives
- Required materials
- Time

Organize skills into topics, and topics into individual PD sessions.

Topic: Cleaning Data

Skill: janitor::get_dupes()

Goal: Students will be able to find and fix duplicates.

Practice: Students will clean a student roster with duplicates caused by different grade, dates and schools.

Required Materials: Student roster set up with duplicates

Create PD Materials 💡

1. Slides

2. R notebook examples and activities

3. Cheat Sheet

4. Homework

Create PD Materials 💡

1. Slides

2. R notebook examples and activities

3. Cheat Sheet

4. Homework

Note: All of the RStudio Tidyverse workshop materials are available on Github - go forth and use them!

Feedback 💬

1. Check in with your coworkers to assess their skills and progress

2. Create new goals *together*

3. Have action steps for how to reach those goals