Intermediate R Topics

Data Wrangling, Unit Testing, and Engineering Principles with R

Donald Sawyer | Lead BI Engineer | Item COE Scrum

Github Location of Today's Training:

https://github.com/donaldsawyer/Demos

Discussion Goals

- Demo R Data Wrangling to replace Excel

- Improve the quality of your code
  - Engineering Principles
  - Test-driven development (TDD)
  - Unit Testing

- All code is on my git repo
  - https://github.com/donaldsawyer/Demos/R

R Examples Covered

- Wrangling
  - Vector Operations
  - Functions
  - Read data from csv, webpage, database
  - Write data to csv
  - Exploring the structure of your data
  - Cleansing data
  - Dplyr data manipulations (forget about sql)
    - Mutate, select, filter, join, group, summarise

- Unit Testing
  - Setting up a test suite
  - TDD w/ R
  - Running tests from command line

- If there's time
  - My completed Datacamp exercises
  - Quick data visualizations

What Are The SOLID Principles?

- **Single Responsibility**
- Open/Closed Principle
- Liskov Substitution
- Interface Segregation
- Dependency Inversion

- http://blog.gauffin.org/2012/05/solid-principles-with-real-world-examples/

Why Does This Matter?

- Create software that is:
  - Maintainable
  - **Testable**
  - Readable
  - Reusable
- Un-clutter your code

How to Become SOLID (IMHO)

- You're the engineer, you decide
- But, if you need a jump:
  - Review your team's code
    - Look for ways to be more SOLID
    - If you need to, focus on one or two principles
  - Practice refactoring
    - LOTS of practicing
  - Lots of unit testing
    - TDD, if that's what it takes
    - Writing automated tests will help you find clunky code
- TRY IT!

Check out More

- R References
  - Free Book: [An Introduction to Statistical Learning with Applications in R](#)
  - Books that I like
    - R for Everyone
    - R Cookbook
    - R Graphics Cookbook
  - [Runit Basics](#)
  - Google! (tip: [R] *your question here*)
- R Training
  - Free Beginner R Hands-On Practice: http://tryr.codeschool.com/
  - Datacamp (my favorite): https://www.datacamp.com/
  - Coursera
  - Others?
- Engineering
  - Refactoring.com
  - [Joel vs. Uncle Bob: Relevance of TDD and SOLID Principles](#)

I'm Done.

Questions?

What else can I show you?