

Foundations of Data Science I

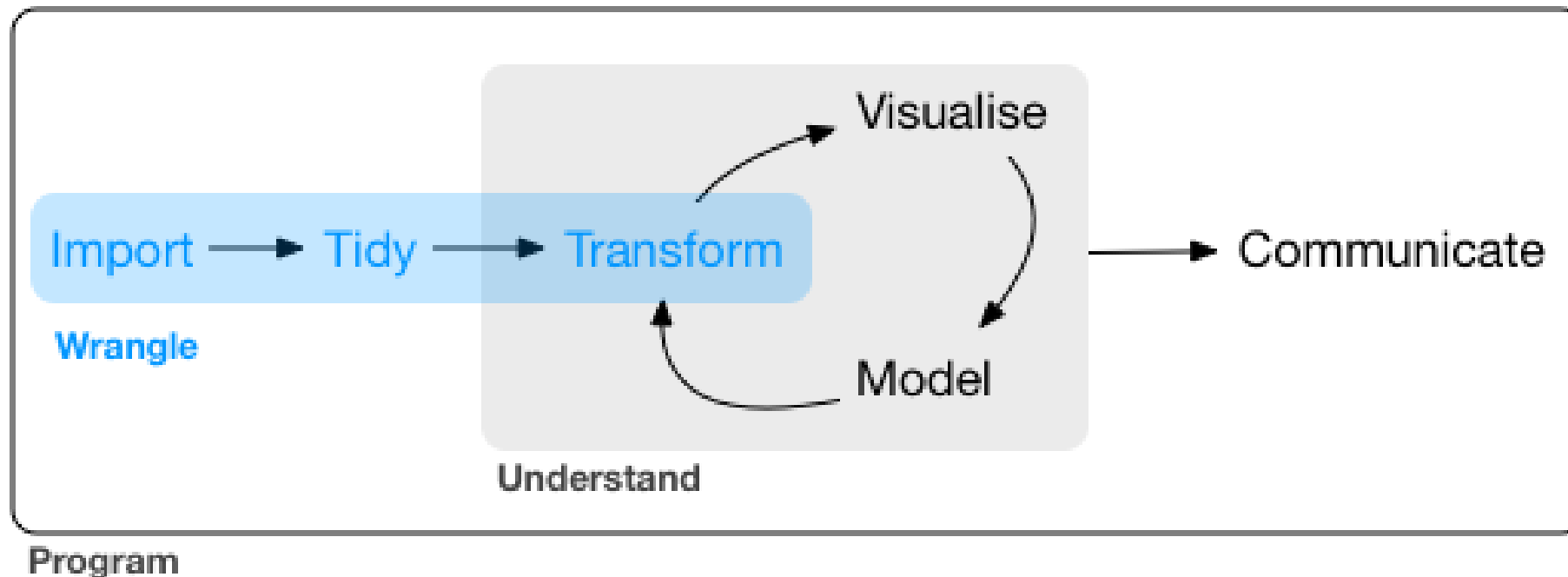
September 12, 2018

Course Materials from Professor Jesse Lecy, Arizona State University

Cristian Nuno

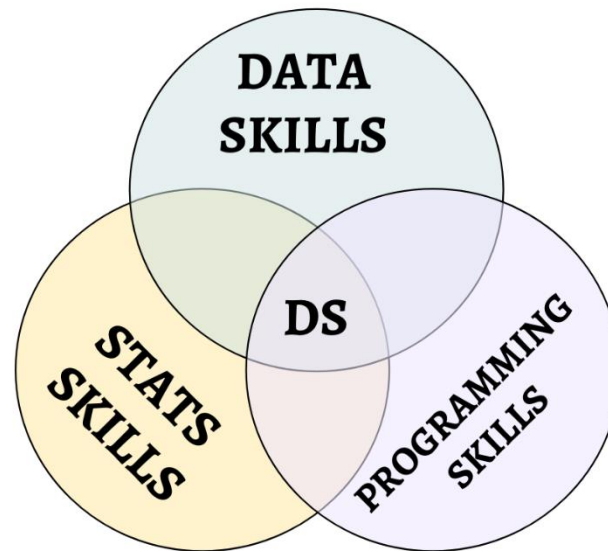
Welcome

- This course introduces students to the field of data science and its applications in the public sector.



Course Goal

- “...[A] data scientist is someone who knows how to extract meaning from and interpret data, which requires both tools and methods from statistics and machine learning, as well as being human.”



Course Outline

- Intro to R, RStudio, and the Tidyverse
- Data types and structures
- Basic Graphs
- Subsets and merges
- Maps
- Control structures (if-then, loops)
- Regular expressions and strings

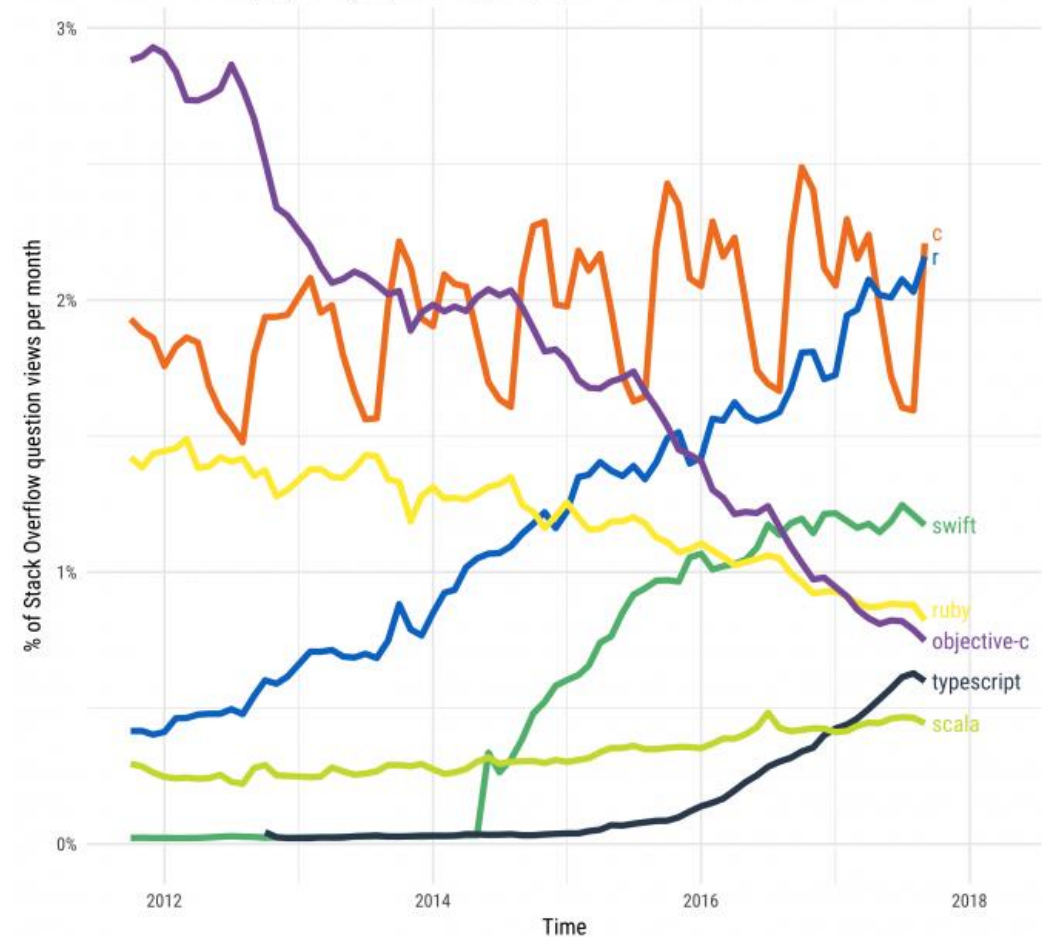
This Course Uses



- R is a language and environment for statistical computing and graphics.
- It's free and entirely open source
- Easier to learn Python or other programming languages after learning R
- Install Tutorials:
 - Your own machine:
 - Macs [\[link\]](#)
 - PCs [\[link\]](#)
 - Work machine:
 - Submit a ticket to ask for it [\[link\]](#)
- For more on Python v. R, please see the following:
 - <https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis>
 - <http://r4ds.had.co.nz/introduction.html>
 - <https://rweekly.org/>

Stack Overflow Traffic to Programming Languages

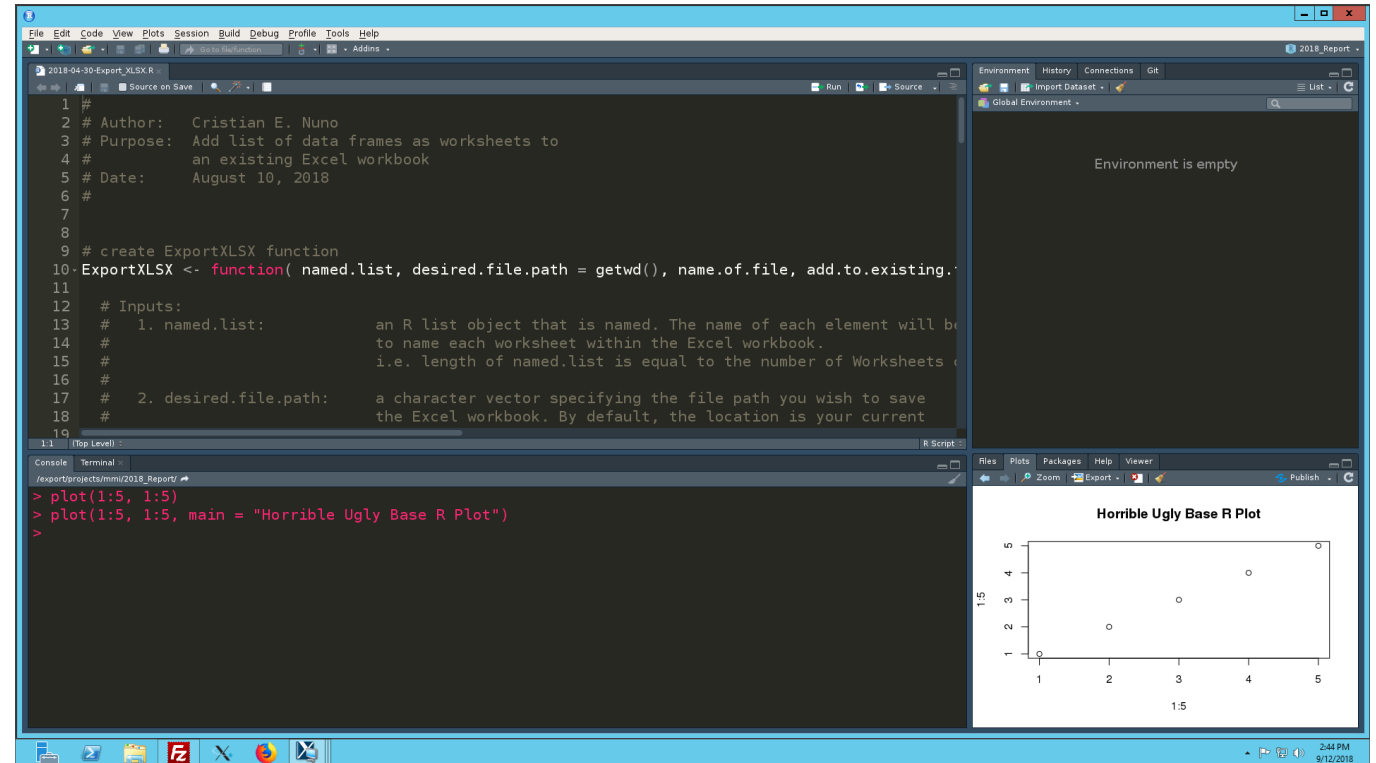
Based on visits to Stack Overflow questions from World Bank high-income countries.
The more-visited languages of Python, JavaScript, Java, C#, and PHP were omitted.



Source: Stack Overflow

To use R, we'll use RStudio

- An interactive development environment (IDE) composed of 4 tabs:
 - Scripts
 - Console
 - Environment*
 - Plot Viewer*
- Click here for a helpful tutorial on navigating RStudio [\[link\]](#)
- * These tabs also contain others as well



To program in R, we'll use the Tidyverse

- The tidyverse is an opinionated [collection of R packages](#) designed for data science. All packages share an underlying design philosophy, grammar, and data structures.
- Learn to use verbs to filter, produce statistics, and visualize your results all in one flow

LITTLE MISS TIDYverse

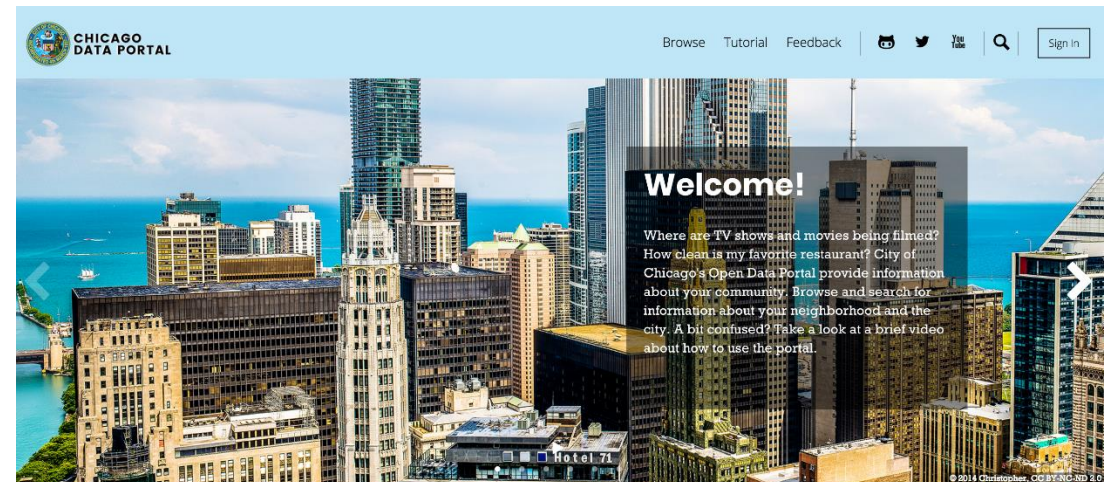
by Roger Hargreaves



Source: Danielle Navarro

To learn, we'll use City of Chicago Open Data Portal

- School profile information for all schools in the Chicago Public School district for the school year 2016-2017 [[link](#)]



What we won't get to in this course

- GitHub/GitLab for version control
- Rmarkdown
- Shiny applications