Section 1: About this course and R

STA 141A – Fundamentals of Statistical Data Science

Instructor: Akira Horiguchi

Fall Quarter 2025 (Sep 24 – Dec 12)
MWF, 9:00 AM – 9:50 AM, TLC 1215
University of California, Davis

Overview

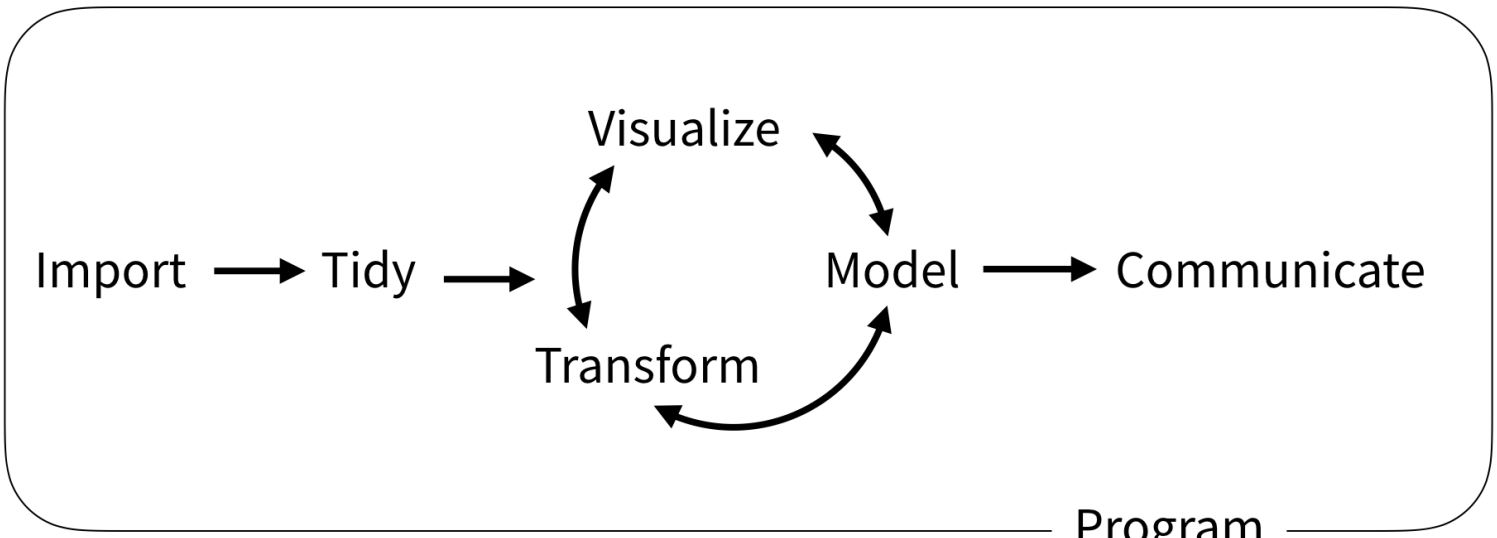
Section 1: About this course and R

- What is the course about? – The model of data science
- Programming
- About R

About this course and R

What is the course about? – The model of data science

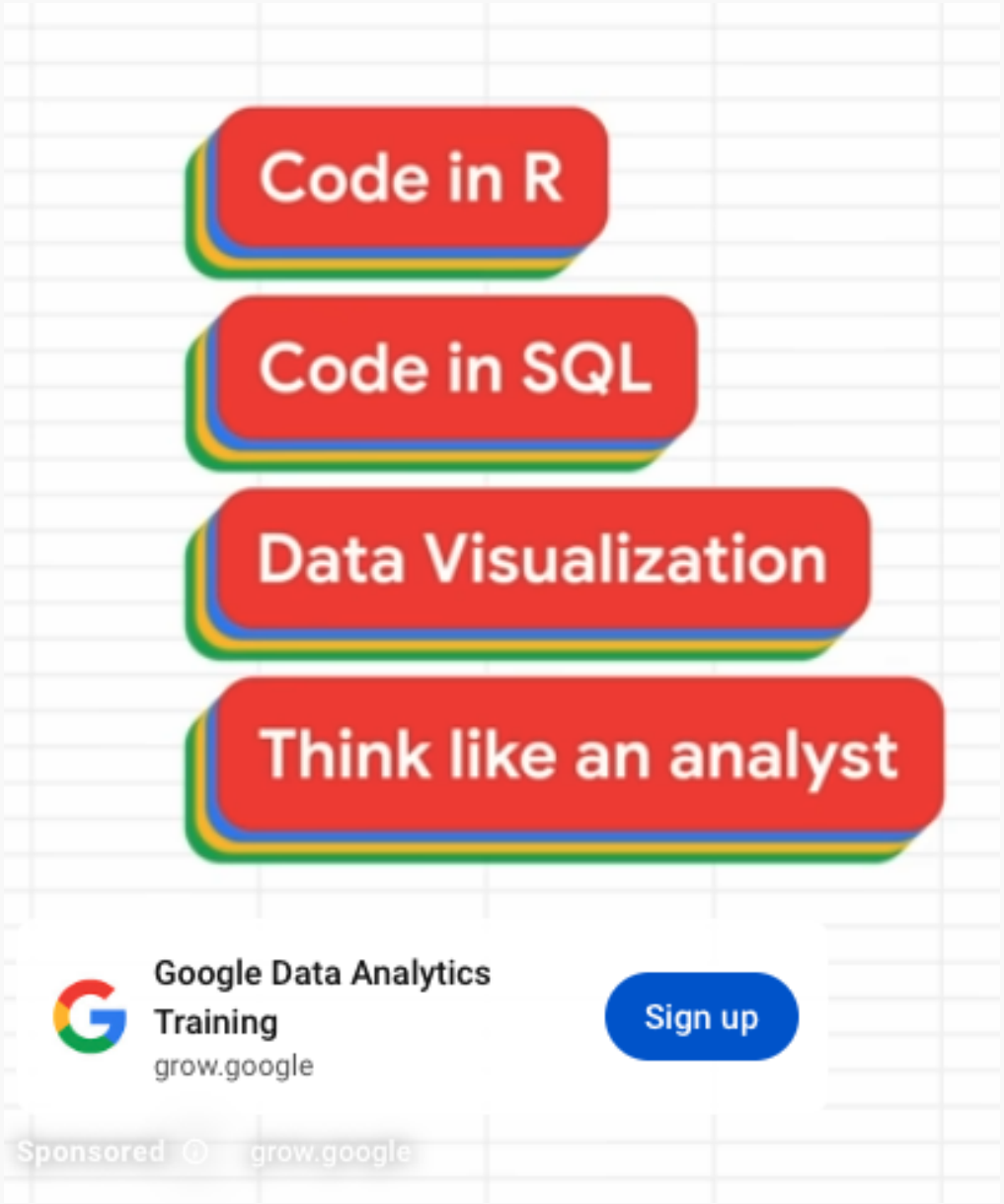
What is the course about? – The model of data science



Wickham and Grolemund (2017)

⇒ We want to do this with R !!!

Ad I got when watching YouTube



About this course and R

Programming

General comments

- Surrounding all the mentioned tools is programming.
- Being an expert programmer or data scientist is not needed. However, learning more about programming pays off since it allows to automate and simplify common tasks

In this class we are going to learn...

- ... general programming concepts;
- ... visualize our results;
- ... statistical programming, computation techniques for data analysis/statistics purposes.

Key high-level programming concepts

- Data Objects (vectors, arrays, matrices, lists, data frames, etc.)
- Operations (vector arithmetic, selecting and modifying, element-wise operations, matrix multiplication, matrix decompositions, etc.)
- Control statements (conditional execution, repetitive execution, etc.)
- Functions (built-in functions, writing own functions)
- Data manipulation (how to manipulate/transform data frame objects)
- Data visualization

About this course and R

About R

Why R?

R is a programming language and software for statistical computing and graphics. It provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques.

- Easy to learn and to use.
- R can be used to generate graphics based on complex data sets very quickly.
- Very popular and one of the standard languages for statistics, data science, computational biology, finance, industry, etc.
- New technology and ideas often appear first in R.
- Supported by a vast community that maintains and updates R.
- A lot of high quality packages.
- Free and open-source.
- Runs on basically any platform.

What is the effect of learning R?

Learn general concepts of high-level programming and languages.

- Since R is a complete programming language, learning it allows you to transfer the concepts to other languages.
- Syntax and available libraries may differ between languages, but how you approach a computational task and reason about the computations is similar.
- It enables you to learn another programming language much easier.

R libraries/packages

R is divided into:

1. The *base R system*
   - This contains, among other things, the base package which is required to run R, and the most fundamental functions.
   - The 'base' system contains also some other packages.
2. In about *20,000 libraries* (or packages) that you can install and use:
   - CRAN [1] 'contributed' packages (or sometimes in BioConductor project or in Github repositories).
   - These already do pretty much anything you have in mind (data manipulation, advanced visualizations, machine learning models, etc.).

[1] The Comprehensive R Archive Network

Integrated Development Environment (IDE)

R console has basic functionality. Typically want to use an IDE.

- An IDE is an application that enables programmers to consolidate the different aspects of writing a computer program by combining common activities of writing software into a single application: editing source code by syntax highlighting and autocomplete, debugging.
- An IDE makes it easier to interface yourself with R, and comes with extra functionalities.

The company *Posit* offers some of its products for free, including the IDEs:

- *RStudio* `https://posit.co/download/rstudio-desktop/`
- *Positron* (first stable release July 2025) `https://positron.posit.co/`
- A quick tour of Positron: `https://www.youtube.com/watch?v=4Ir_HX4riHw`

Please make sure you can run R and an IDE from a machine you have access to.

Some more announcements

- This quarter, the other STA 141A instructor will teach using Python
- *Homework due in seven days* (W Oct 1)