Lecture 1: Intro to R, RStudio, and the Data Science Workflow

FNCE 5352 — Financial Programming & Modeling

Matt McDonald

2026-01-20

Intro

Matt McDonald - CV

Work:

KBRA KROLL BOND RATING AGENCY

GE Capital

IBM

edgelab FRESH IDEAS. BREWED DAILY.

priceline.com®

pwc

Education:

Columbia University
MS Statistics

University of Connecticut
MBA Finance

Colgate University
BA Mathematics

Career Path


Actuarial
IT/Software Development
Credit Modeling / Model Risk Management

Course Framing

The Core Idea

Most financial questions aren’t about the “right” answer.

They’re about making the **best decision**
given **noise, uncertainty, and tradeoffs**.

What This Course Is — And Isn’t

This course is not - learning every algorithm - optimizing models for their own sake

This course is - applied statistical learning for finance - regression and classification - interpretation and model assessment - understanding tradeoffs (accuracy vs robustness, simplicity vs flexibility)

Tools And Transfer

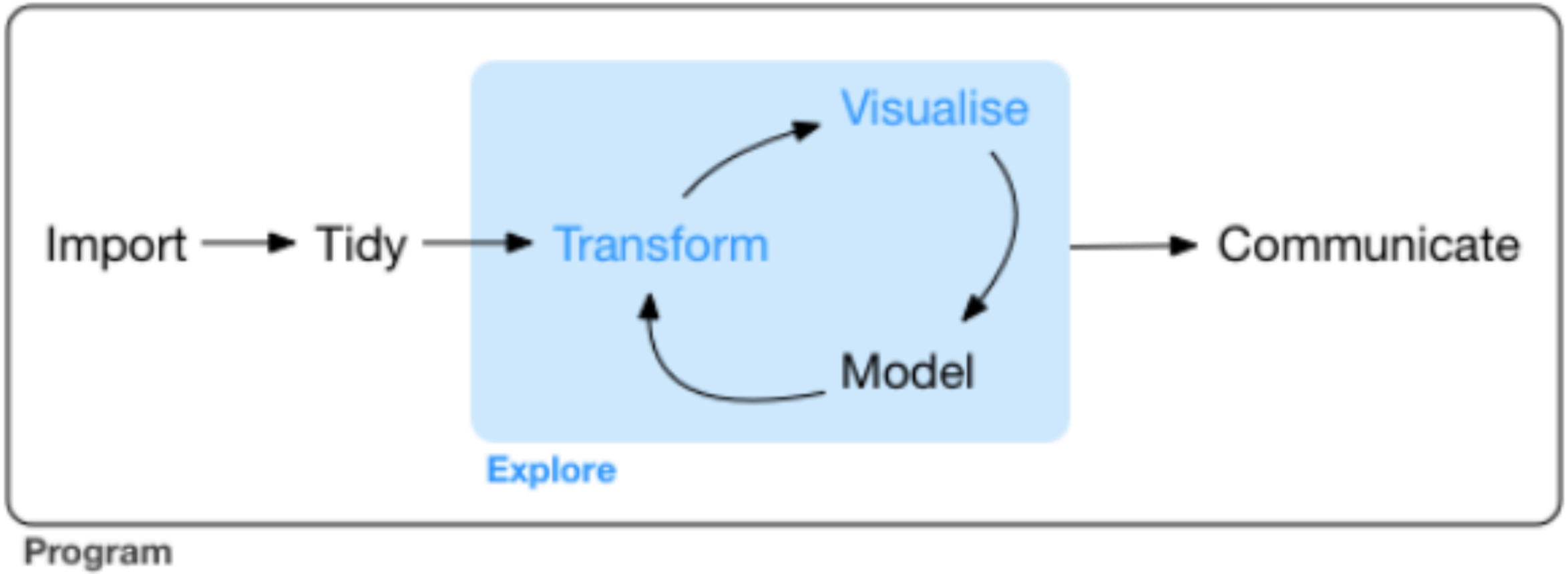
The **ideas** matter more than the language.

- Part 1 uses **R** (lecture language)
- Part 2 uses **Python**
- Assignments may be completed in either language
- The workflow and concepts transfer directly across both

The Data Science And Modeling Workflow

(Finance-First)

The Workflow We Will Use All Semester



This is not a one-time sequence.
It's a **loop** you will revisit constantly.

Why This Is A Loop (Not A Checklist)

- New data arrives
- Assumptions change
- Results raise new questions
- Decisions require iteration

In finance, the workflow rarely ends.

What Each Step Looks Like In Finance

- **Import**
  Prices, returns, fundamentals, loan tapes, macro series
- **Tidy**
  Dates, joins, missing data, panel structure
- **Explore**
  Distributions, correlations, regime changes
- **Model**
  Regression and classification to support decisions
- **Communicate**
  Plots, narrative, and a clear “so what?”

R And Python: Portability

The Key Claim

> **If you understand what we're doing in R, you'll recognize it instantly in Python.**

What Actually Transfers

- Data frames / tables
- Functions, objects, and libraries
- Wrangle → visualize → model → evaluate
- Train / test thinking (coming soon)

RStudio Orientation

(Minimum Viable Comfort)

R vs RStudio

Packages

R: New phone

R Packages:
Apps you can download

GET IT ON Google Play

Download on the App Store

Projects (Strongly Recommended)

- One folder per project
- Relative paths
- Easier reproducibility

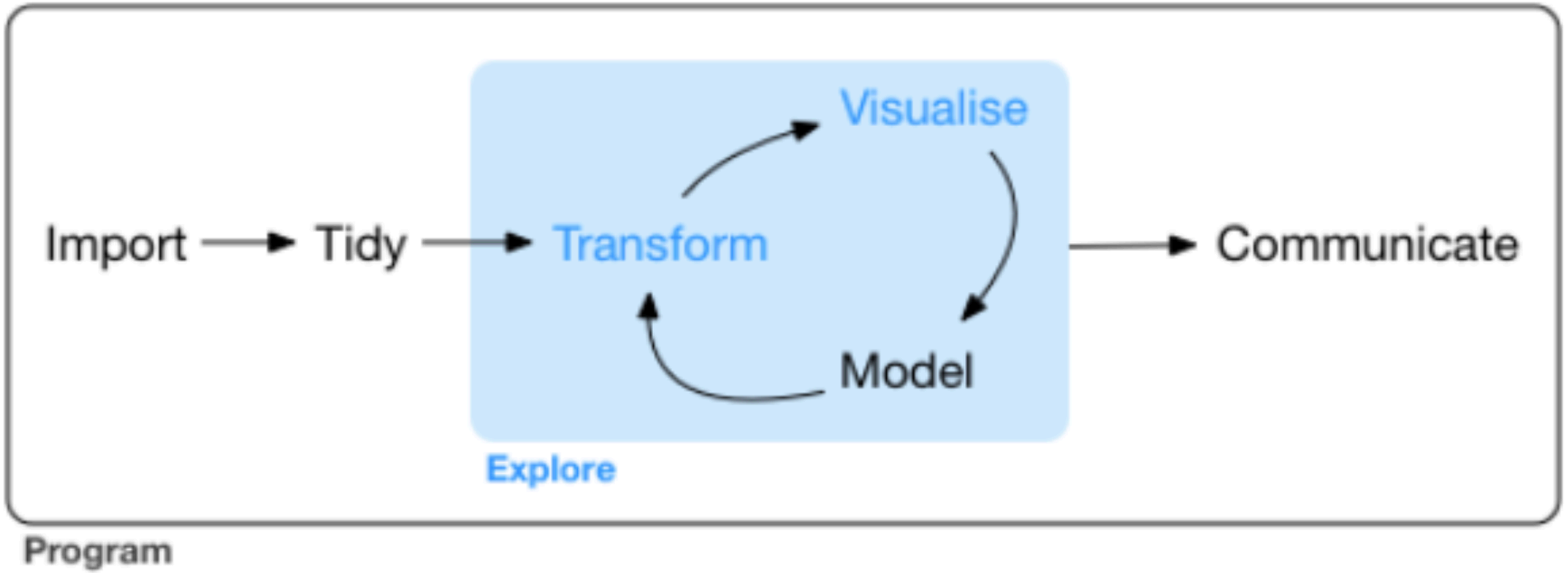
RStudio Walkthrough (Live)

Hands-On Workflow Demo

(AAPL + SPY)

Live Coding In RStudio

What We Just Did



ISL Chapter 1 (conceptual)

> This chapter is about how we talk about models, not how we compute them.

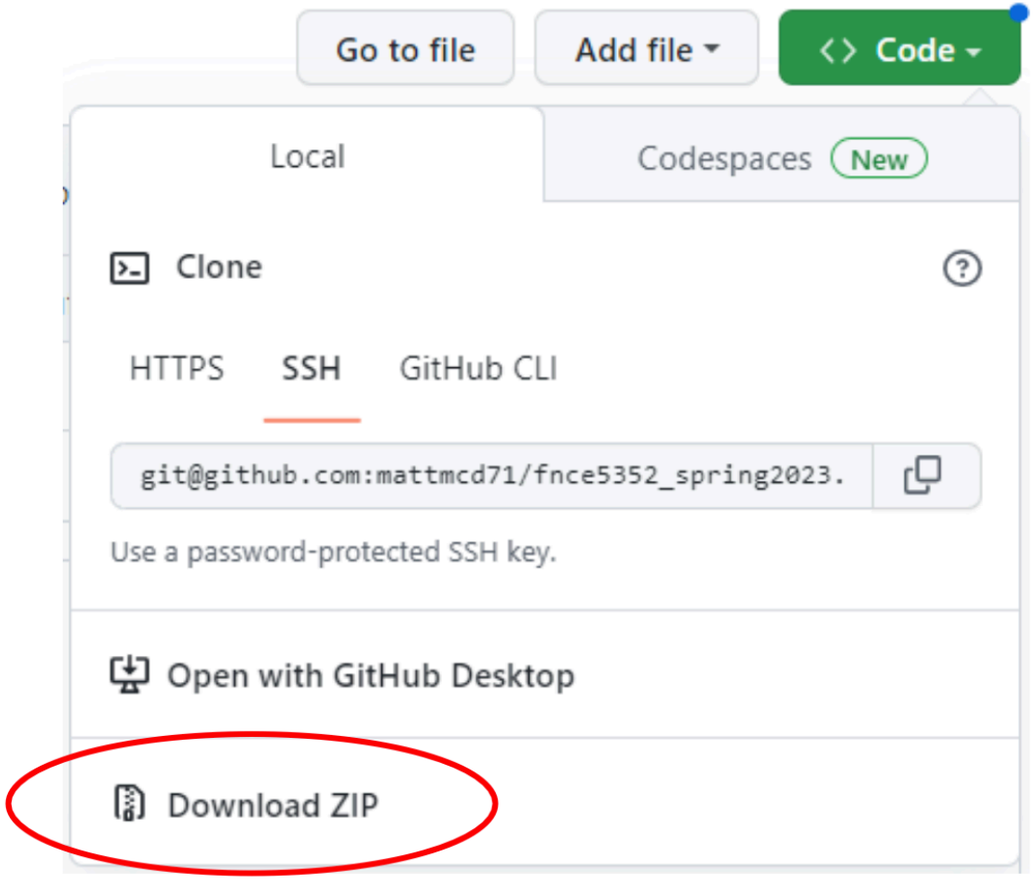
Vocabulary we’ll use all semester

- statistical learning
- prediction vs inference
- training vs test performance (intuition today)
- overfitting (why “looks great in-sample” can be a trap)
- tradeoffs (accuracy vs interpretability; complexity vs robustness)

Wrap Up

How to Access the Class GIT Repo

- Download ZIP



- Clone
  - HTTPS or SSH
  - SSH will require a setup of your SSH keys
    - See happygitwithr page for complete instructions on that

Next Steps

- Read: R4DS Sections 1–8
- Read: ISL Chapter 1 (conceptual)
- Make sure you can run code and re-render slides