Welcome!

Introduction to Data Science

Today we are going to get all of the administrative details dealt with. Here is a quick outline:

– give a brief overview of the course material
– go through the syllabus
– tell you a bit about myself
– install software
– answer any additional questions

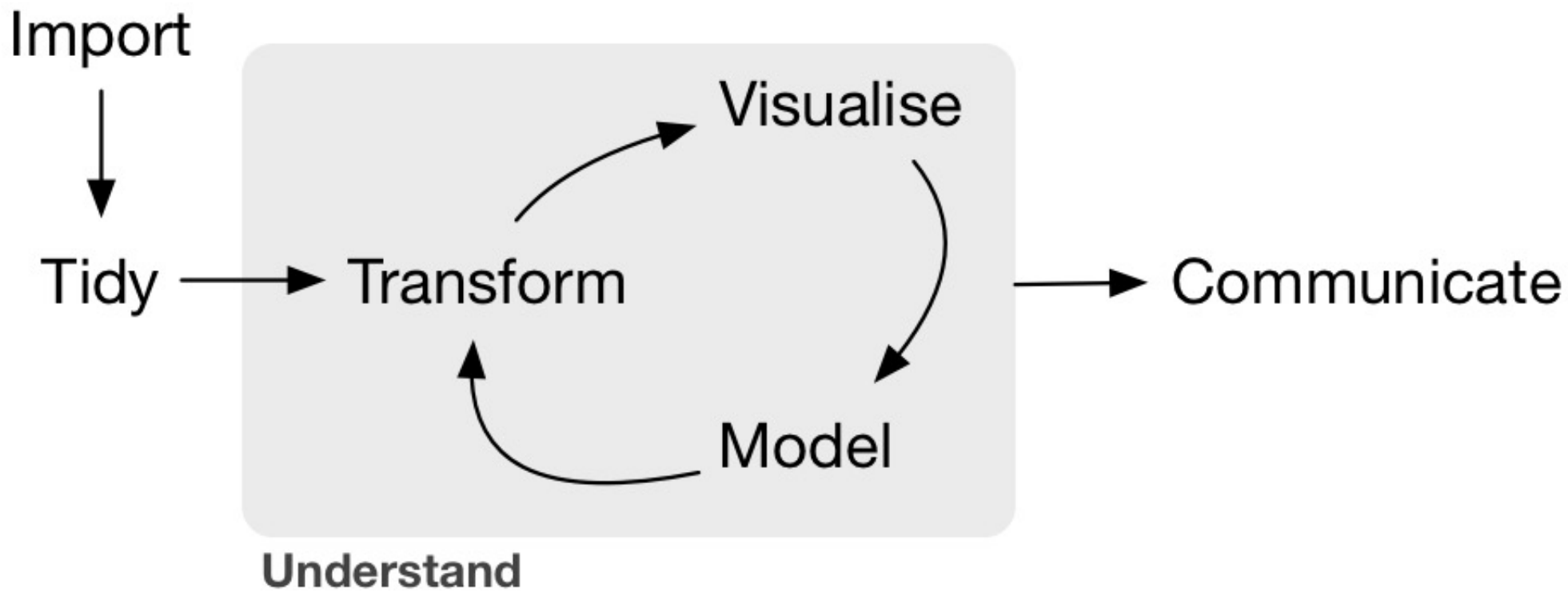
Data Science?

Data science is an interdisciplinary field concerned with drawing knowledge from data and communicating those results to various audiences.

Unlike many other fields that you may be familiar with, data science focused on the creation and application of **methods**, rather than theoretical or foundational questions.

This semester we will learn and practice a series of methods for organizing, collecting, visualizing, manipulating, and exploring different kinds of data.

Data Science Pipeline



Grolemund & Wickham R4DS Illustration

Perspectives

Most students really enjoy this course. We learn skills that are important to doing technical work in almost any field while playing around with a variety of interesting datasets in a relatively low-stress setting.

The most important thing is to have an open mind about what we will be covering this semester. Note that:

- this is NOT a mathematics class, despite the name of the course
- the topics covered are very different from those in an introductory statistics course
- it often resembles a computer science course, but keep in mind that different techniques are needed for data science scripting than you may have learned in other CS courses

Programming

There are several different programming languages for data science. By far the two most popular are R and Python.

We will be using R this semester but will learn a version that is easily adapted to other languages such as Python.

In the last week I will demo the use of Python and JavaScript based on the the class material.

Grading

There are three elements that you will get graded on this semester:

- **Exams**: four take-home, home-book exams
- **Homework**: readings posted on website; graded on self-reported completion using course form
- **Engagement**: short essay due on last day of class reflecting on effort/participation/engagement

Your final grade is determined by averaging together these six grades. More details and the specific grading scales used to convert between number and letter grades are given in the syllabus.

Class Structure

Most course meetings are organized as follows:

— **homework I:** carefully read any posted notes on the website and formulate questions for the next class (30-90 minutes)
— **course form**: fill out at the start of class (< 1 minute)
— **discussion / slides:** review readings or previous notebooks, discuss any questions, perhaps start classwork together (15-30 minutes)
— **classwork**: work individually or in small groups to answer questions in the form of programming notebooks (45-60 minutes)
— **homework II**: finish classwork or review posted solutions (0-60 minutes)

All materials can be found on the course website.

Attendance

It is important to attend class as much as possible, though it is also important to stay home if you are feeling ill. The class form will ask, when you are absent, to give an explanation. Just a few words will suffice (i.e., "feeling ill", "job interview", "needed a personal day").

If absences are relatively rare and/or justified,† they will not affect your grade. When I feel that they are excessive, I will reach out to individuals before taking any punitive action.

† A specific cut-off is hard to define. As a guideline, 2-3 absences for any reason is fine. A few more in the case of illness are also okay. Beyond that we will likely have to talk.

Questions & Office Hours

We will usually have a lot of time in class to answer any questions you have about the course material. I am usually around before and after class for additional questions.

Please also feel free to send questions by email. I typically respond within 24 hours.

Finally, I am of course happy to schedule an office hours meeting for extended questions or personal concerns. Just send me an email with your availability at least 1 day before you'd like to meet.

Coffee chat?

I get a lot of requests for longer discussions about graduate school, careers in data science, or research projects.

I love having these conversations with students. Just send me an email or ask after class and we can find a time to grab coffee/tea/whatever and answer any questions you have.

Masks and Such

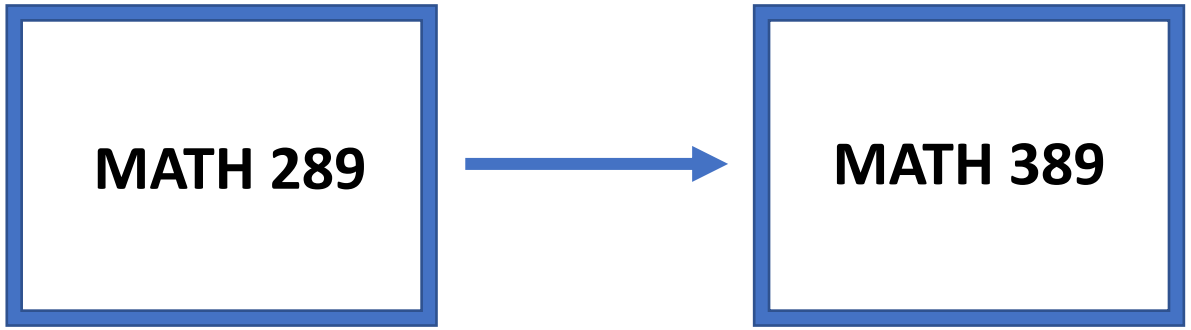
As I mentioned in my first email to the class, policies regarding COVID-19 are constantly changing. In fact, they already have!

Following the UR guidelines, for the first three weeks we will wear masks while in class. We can re-assess the situation after that point.

Faculty are also asked to create a seating chart (kindergarten-style, see image). Think about where and with whom you would like to sit. I will use whatever configuration we have in class on Thursday.

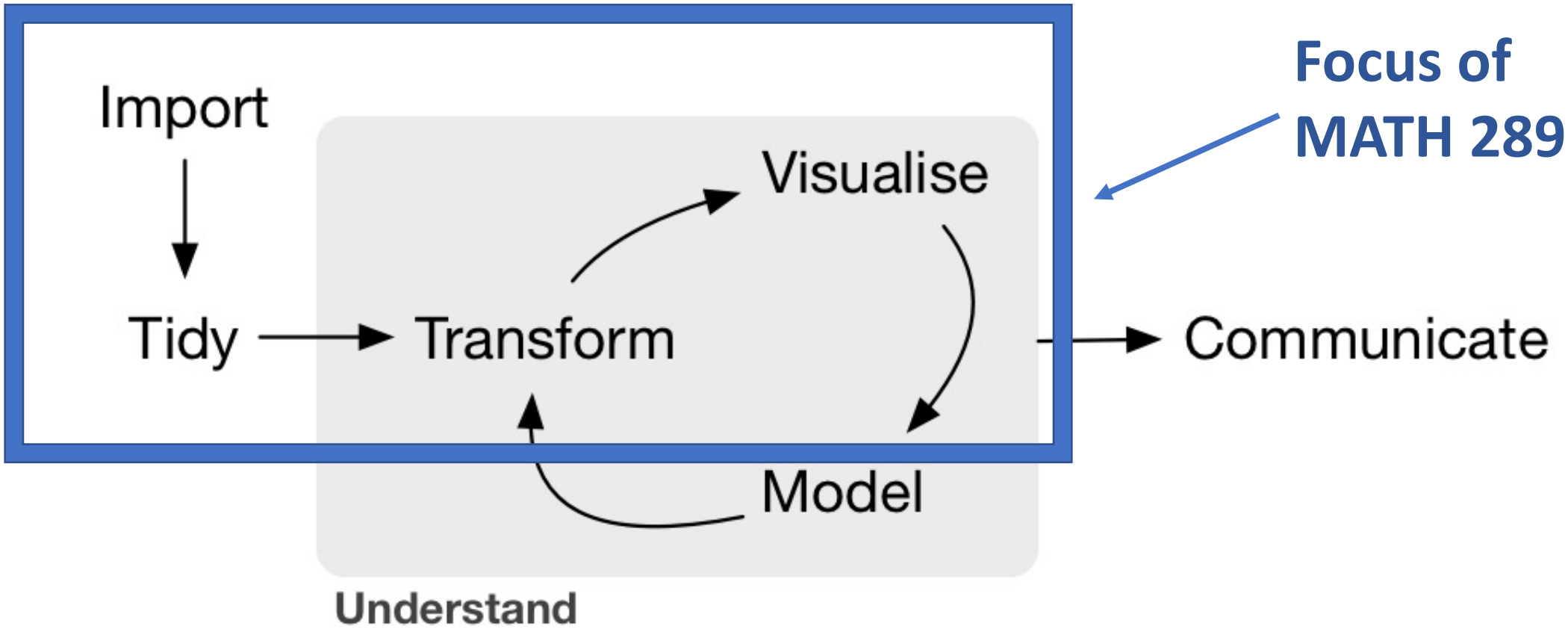
Course Sequence

This course assumes you have some experience using code to manipulate data. It makes no assumptions about your knowledge of any specific programing language or knowledge of statistical inference.



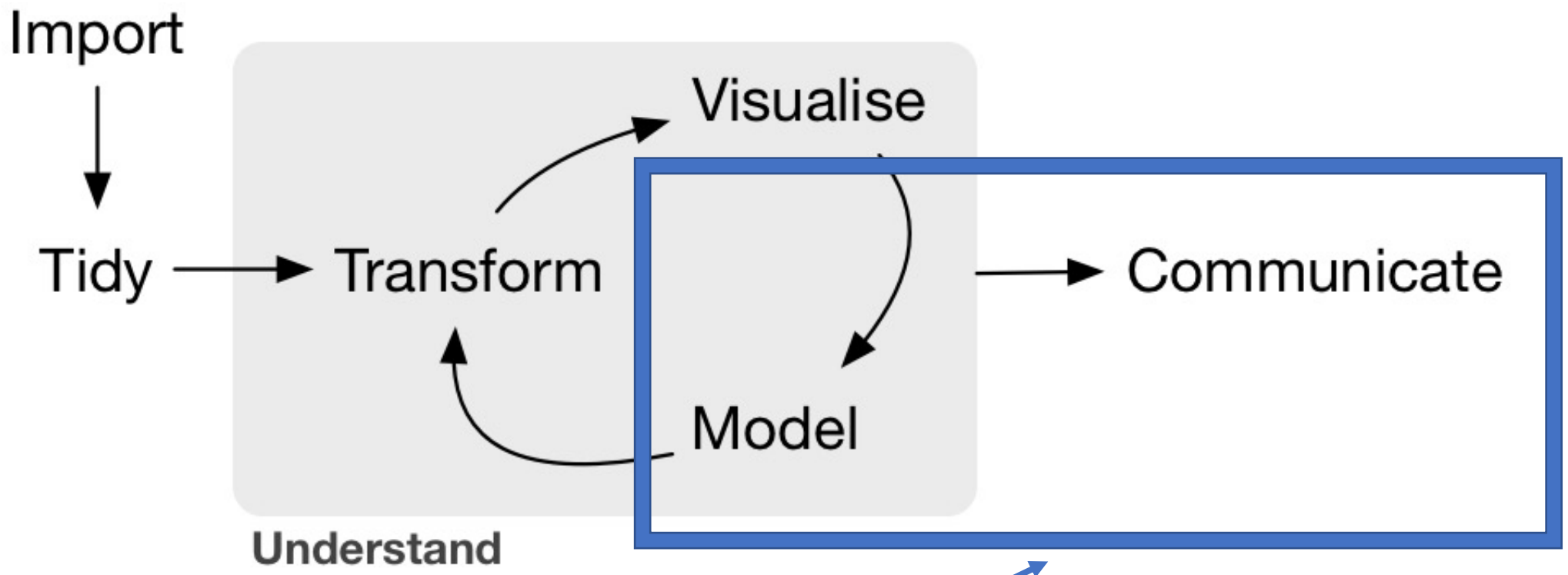
The class is designed as a year-long sequence paired with MATH 389; I strongly suggest taking both during the same academic year if possible.

Data Science Pipeline



Grolemund & Wickham R4DS Illustration

Data Science Pipeline



Grolemund & Wickham R4DS Illustration

Focus of MATH 389

Course Topics

See tentative topics posted on the course website.

About Me

- From New England: born in Maine, school in MA, ME, CT
- Moved to Richmond in 2016
- Research on large text and image datasets in linguistics and cultural studies

Photogrammar

R Journal

ACLS

Distant Viewing Lab

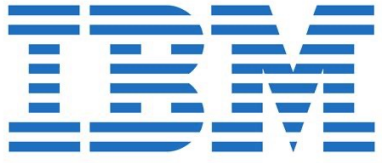
THE ANDREW W. MELLON FOUNDATION

NATIONAL ENDOWMENT FOR THE HUMANITIES

COLLEGIUM DE LYON
Université de Lyon

About Me

- Lots of industry experience in DS:
  - IBM (Healthcare)
  - Travelers (Insurance)
  - DARPA (social media)
  - AT&T (location analytics)
  - Telperian (pharmaceuticals)



About Me

- I have two Shih-Tzus: Roux and Sargent
- Roux is often in my office; please come say hello