DSST 330:
Mathematical Statistics

Welcome!

T. ARNOLD

1. Syllabus

Course Website

As I mentioned in my email, these notes and all others for this semester will be posted on the course website:

https://statsmaths.github.io/dsst330-s24/

2. Course Content

Terminology

There has always been some uncertainty about what is covered by the term ***statistics***. At stake are the several closely related questions:

Q1. Does it cover all aspects of collecting, organizing, analyzing, interpreting, and presenting data? Or is it just the mathematical modeling of randomness and/or applied probability?

Q2. Is it a branch of mathematics? A domain within the larger mathematical sciences? Or an interdisciplinary field that cuts across many others?

Q3. How is it related to other domains such as data science, information science, and machine learning?

Terminology

These kinds of debates can be important in some contexts, but too many people waste too much time thinking of them when they could just state whatever working definition they are using in some context and then move. So, let's do that!

Terminology

For us, we will use the following definitions:

Data Science: Large umbrella term for the academic domain concerned with the collecting, organizing, analyzing, interpreting, and presenting data. It is its own interdisciplinary field that naturally overlaps with many other domains.

Statistics: An academic domain focused on the modeling of randomness in data. Often focused on making inferences and predictions about new data.

So, for us, nearly all of statistics falls under the umbrella of data science, though the inverse is certainly not true. Also, because it is focused on modelling, almost all of statistics is part of the larger mathematical sciences. Very little, however, would be considered a sub-domain of mathematics itself.

330

The implications for this class are that:

Mathematical Statistics We will be using mathematical/probabilistic methods to modelling uncertainty and randomness in data. The first term is a bit redundant but helps to signal our definition of statistics.

DSST Starting this semester, this course has a DSST prefix (it still counts as a mathematics elective though!). We are taking a broad data science-based view of the material. We focus on theoretical results that help us do a better job of working with data rather than introducing theory for its own sake. We will use and create data throughout the semester and will be doing a fair amount of statistical computing.

Fair Warning

While I am very well qualified to teach this class—I have a Ph.D. in statistics, five years of industry experience, and a decade of teaching experience—I have never actually taught a mathematical statistics course. So, expect a few bumps and changes along the way.

You can see a tentative list of topics on the course website. We will roughly follow this set of topics, but the specific pacing might change.

3. Introductions

You

- How does your name show up in BannerWeb? Do you have a different preferred name? Is there anything else I should know about the pronunciation of your name or your preferred pronouns?
- Please give a brief 2-3 sentence introduction.
- What other statistics or data sciences courses have you taken at UR?
- In one sentence, why are you taking this class?
- Anything else I should know? At a minimum, please let me know if you have a DAN or are involved in any UR sports teams or other official organizations that may cause you to miss class

4. Course Setup

Installing Course Software

We need to install three different components for this semester:

1. The R Programming Language
2. The RStudio IDE
3. A set of R Packages and data

All of these components are open-source and available for all modern operating systems. You may have trouble, however, if you have an older OS and have not updated it recently.

Even if you already have R installed, I suggest doing a fresh update for the semester.

1. Installing the R Language

To install R, go to https://cran.r-project.org/ and select your operating system. Then:

macOS => click on R-4.3.2.pkg and follow instructions
Windows => click on **base** followed by "Download R"

For Linux, either install from source or use your favorite package manager.

You need to install R before anything else, but we will never actually open it directly. So feel free to remove and shortcuts or links that are created during installation.

2. Installing RStudio

To install RStudio, follow the following link and download either the dmg (macOS) or exe (Windows):

https://posit.co/download/rstudio-desktop/

Note: On macOS, you need to drag the RStudio icon into your Applications directory after downloading.

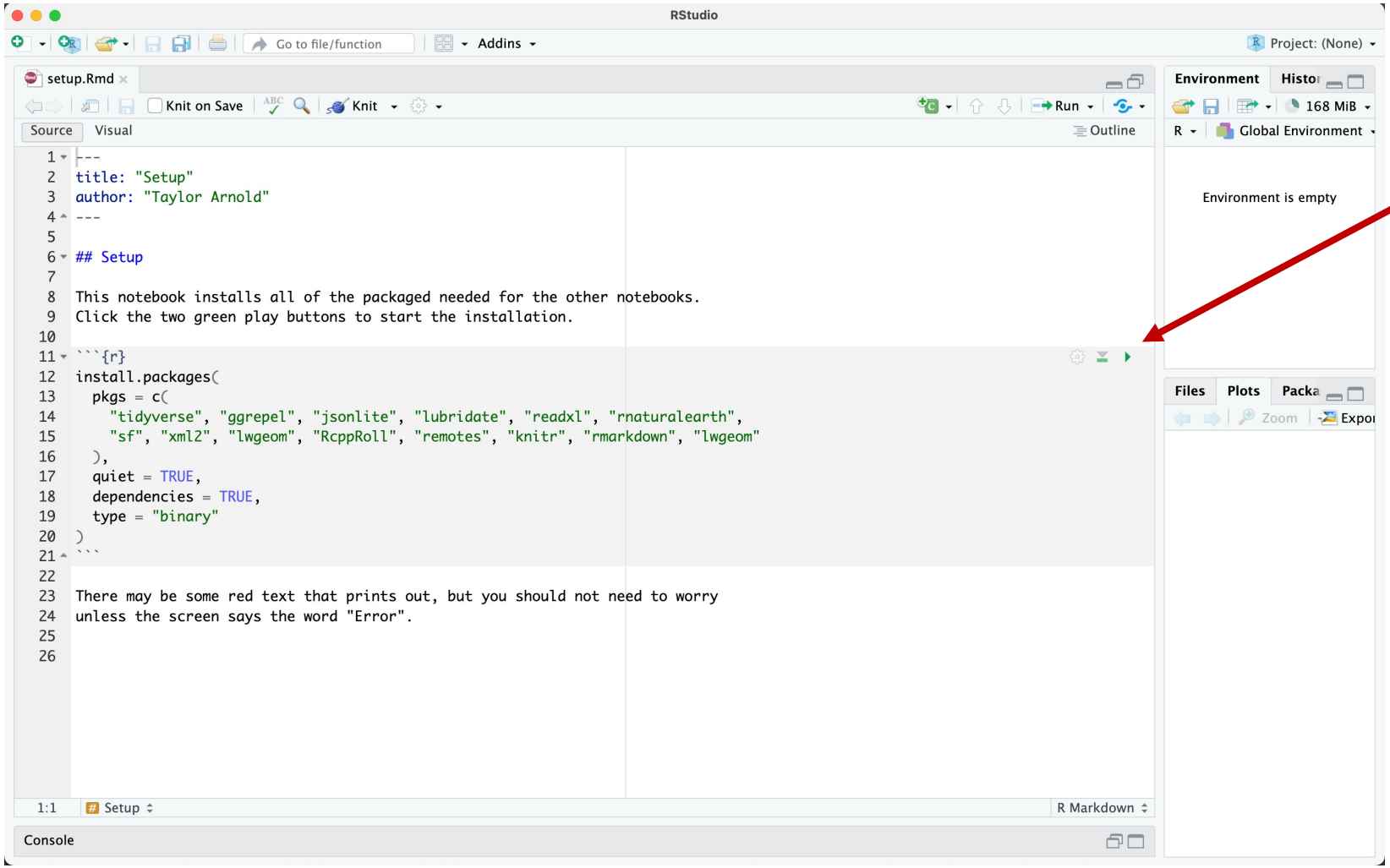
3. Install R Packages

Finally, download and unzip the "materials.zip" file from the class website. We will use this directory throughout the semester, so put it somewhere you will remember it.

Then, open the setup.Rmd file in RStudio, and click the green play buttons (see next slide).

Make sure to put this folder somewhere for the semester. You will need it!

3. Install R Packages



Click here

For the Semester

You should plan on bringing a laptop with a working version of R, RStudio, and all of the installed packages to each class meeting. If that is or becomes an issue, just let me know and we will find a solution.

Note that if you are having computer issues, particularly during an exam, it is always possible to use the lab computers in Jepson as a back-up. They have R and RStudio installed, but not all of the class R packages. Simply start from Step 3 in these notes before getting started.