

# **431 Fall 2023 Syllabus**

Thomas E. Love

2023-08-23

# Table of contents

<b>1</b>	<b>Welcome to 431!</b>	<b>5</b>
1.1	Working With This Syllabus . . . . .	5
1.2	Who, When and Where? . . . . .	5
<b>2</b>	<b>Getting Started</b>	<b>6</b>
2.1	Welcome to 431 Survey . . . . .	6
<b>3</b>	<b>Course Web Site</b>	<b>7</b>
3.1	Contact Information . . . . .	7
<b>4</b>	<b>Course Description</b>	<b>8</b>
4.1	Course Objectives . . . . .	8
4.2	Key Topics in 431-432 . . . . .	9
4.3	431 Course Outline & Format . . . . .	9
4.3.1	Part A: R and Exploring Data . . . . .	9
4.3.2	Part B: Making Comparisons . . . . .	10
4.3.3	Part C: Linear Models . . . . .	10
4.4	Prerequisites and Intended Student Population . . . . .	11
4.5	Motivations for our Approach . . . . .	11
4.6	Is 432 Required? . . . . .	12
<b>5</b>	<b>Required Texts</b>	<b>13</b>
5.1	Course Notes . . . . .	13
5.2	Buy This Book! . . . . .	13
5.3	Other Books to Download . . . . .	13
5.4	Key Articles and Posts . . . . .	14
<b>6</b>	<b>Required Software</b>	<b>16</b>
6.1	R and RStudio . . . . .	16
6.2	Getting the Software . . . . .	17
6.2.1	System Requirements . . . . .	17
6.2.2	What will I need for 431? . . . . .	18
6.3	Installing R and R Studio . . . . .	18
6.4	Installing R Packages and Data/Code for 431 . . . . .	18
6.5	Need Installation Help? . . . . .	18
6.6	Updating Your R Packages . . . . .	19

6.7 Why do we teach R in 431-432? . . . . .	19
6.8 Additional Resources . . . . .	20
<b>7 Professor Love</b>	<b>21</b>
7.1 Email . . . . .	22
7.2 Offices . . . . .	22
7.3 Name, Pronouns, Social Media . . . . .	23
<b>8 Teaching Assistants</b>	<b>24</b>
8.1 TA office hours . . . . .	24
8.2 Getting To Know The TAs . . . . .	24
8.2.1 Ali Elsharkawi . . . . .	25
8.2.2 Chenyu Liu . . . . .	26
8.2.3 Alex Olejko . . . . .	26
8.2.4 Lindsay Petrenchik . . . . .	27
8.2.5 Miza Salim Hammoud . . . . .	27
8.2.6 Monika Strah . . . . .	29
8.2.7 Ria Tilve . . . . .	29
<b>9 Getting Help</b>	<b>31</b>
9.1 Campuswire is the place to ask questions and discuss the course . . . . .	31
9.2 TA Office Hours is the place to get one-on-one help for 431 . . . . .	32
9.3 How We'll Use Canvas . . . . .	32
9.4 How We'll Use Google Drive . . . . .	32
9.5 Meeting with Professor Love . . . . .	33
9.6 Other Sources of Help . . . . .	33
<b>10 Assignments</b>	<b>34</b>
10.1 Deadlines for Assignments . . . . .	34
10.2 Using Artificial Intelligence Tools . . . . .	34
10.3 Minute Papers . . . . .	34
10.4 Class Participation . . . . .	35
10.5 Labs . . . . .	36
10.5.1 Lab X . . . . .	36
10.5.2 A few tips for Early Lab Work . . . . .	36
10.6 Quizzes . . . . .	37
10.6.1 Campuswire and the Quizzes . . . . .	37
10.6.2 About the Quizzes . . . . .	37
10.7 Projects . . . . .	38
10.8 Why Two Projects? . . . . .	38
<b>11 Grading</b>	<b>39</b>
11.1 Grading Breakdown . . . . .	39

11.2 Grading Class Participation . . . . .	40
11.3 Grading Labs . . . . .	41
11.4 Lab Appeal Policy - Request a Review via Google Form . . . . .	41
11.5 Grading the Quizzes . . . . .	43
11.6 Grading the Projects . . . . .	43
11.7 Late Assignments . . . . .	43
11.8 Incomplete Grades . . . . .	44
<b>12 General Policies</b>	<b>45</b>
12.1 Attendance . . . . .	45
12.2 Auditing . . . . .	45
12.3 Recordings . . . . .	46
12.4 Diversity and Inclusion . . . . .	46
12.5 Disability Resources . . . . .	46
12.6 Laptop and cell phone use . . . . .	47
12.7 Academic Integrity . . . . .	47
12.8 Plagiarism . . . . .	48
12.9 Use of AI is Not Permitted without Express Written Consent from Dr. Love . . .	48
12.10 University Resources / Student Support . . . . .	49
<b>13 Movies</b>	<b>50</b>
<b>14 General Advice</b>	<b>51</b>
14.1 On Graduate School . . . . .	51
14.2 On Seeking a Job . . . . .	52
<b>15 On Writing</b>	<b>53</b>
15.1 Writing, Presenting & Communicating . . . . .	53
15.2 Campus Resources . . . . .	53
15.3 Artificial Intelligence (ChatGPT) . . . . .	53
15.4 Advice from Other People . . . . .	54
15.5 A Few Tips from Professor Love . . . . .	54

# 1 Welcome to 431!

This is the Fall 2023 syllabus for PQHS / CRSP / MPH 431: Statistical Methods in Biological & Medical Sciences at Case Western Reserve University with Professor Thomas E. Love. While some things will change here as the semester gets started, this is almost a complete document at this point.

## 1.1 Working With This Syllabus

1. This syllabus includes **fifteen** chapters. Use the sidebar on the left side of the screen to navigate between chapters. Make sure you look at each chapter.
2. Click the download icon under the title at the top left to download a PDF or ePub version of the syllabus.
3. Click the switch next to the download icon to switch from light to dark mode.
4. The sidebar also lets you search the entire syllabus.
5. The current chapter's table of contents is shown on the right of your screen.
6. The syllabus will be updated occasionally. The date of the current version is listed above under **PUBLISHED**.

## 1.2 Who, When and Where?

- The course instructor is Professor Thomas E. Love. Learn more about him in Chapter [7](#).
  - Email him at **Thomas dot Love at case dot edu** if you have any questions before class.
- There are several teaching assistants helping with the course this semester. See Chapter [8](#) for more on them.
- The course is given on Tuesdays and Thursdays from 1:00 to 2:15 PM Eastern Time, in Room E321-323 in Robbins Hall, within CWRU's School of Medicine, starting on 2023-08-29.

## 2 Getting Started

Your first few steps in this course are to:

1. Register and enroll via the [CWRU Student Information System \(SIS\)](#).
  - Section 100 of PQHS 431, CRSP 431 and MPHP 431 are identical.
2. Review the [main course website](#), and be sure to visit the [Course Calendar](#).
3. Skim through each chapter of the Course Syllabus (this document.)
4. Obtain the texts required for the course. The book you need to purchase is David Spiegelhalter's *The Art of Statistics: How to Learn from Data*, originally published in 2020 and available at all major book retailers for around \$20. All other required texts are free. See Chapter [5](#) for details on how to get the other required texts.
5. Complete the [Welcome to 431 survey](#) - see below for details.
6. Get started installing the software you'll need for the course. See Chapter [6](#) and [the Software link on our main website](#) for more details.
7. Sign up for the (free) [Campuswire discussion forum](#).

### 2.1 Welcome to 431 Survey

Please fill out the survey at <https://bit.ly/431-2023-welcome-survey> to help us get to know you a little better.

- Note that you must **log into Google via your CWRU account** in order to access the survey.
- The survey should take 15-20 minutes to complete, and we encourage you to complete it as soon as you can.

# 3 Course Web Site

The main course web site is found at [this link](#).

By the first class, this page will house links to everything we will provide, including:

- The [431 Syllabus](#) (this document)
- Professor Love's [Course Notes](#), the main textbook for the course
- The [Course Calendar](#), which is the final word for all deadlines and assignments
  - The [Calendar](#) also links to materials for each class, including presentation slides and announcements
- [Data \(and Code\)](#) for the course's assignments, presentations, and notes
- Our [Campuswire discussion forum](#) for questions and answers about R and the course
- Our [Canvas page](#) for turning in assignments and downloading recordings.

## 3.1 Contact Information

Our main website at <https://thomaselove.github.io/431-2023/> includes a [Contact Us](#) page which provides detailed information on the best ways to get help from Professor Love and the Teaching Assistants during the semester.

- The course Teaching Assistants will hold regular office hours, starting in early September and running through early December, over Zoom, and the schedule for these sessions will be found on the [Contact Us page](#) and the [Course Calendar](#).
  - Once the semester begins, questions about course materials are best asked through [Campuswire](#) or in TA Office Hours, over Zoom.
- Professor Love will arrive 10 minutes before each class, and stay afterwards, as well, to discuss any student concerns appropriate to that setting.
  - Professor Love's office at CWRU is on the ground floor of the Wood building (in the WG-82 Suite), specifically WG-82J. Email him if you need an appointment, either in person or via Zoom.
- If you have any unique concerns about the course, need special accommodations or have any other issues you want to discuss with Professor Love specifically, please email him directly (**Thomas dot Love at case dot edu**) and he'll respond as quickly as possible.

# 4 Course Description

PQHS 431 (cross-listed as CRSP 431 and MPHP 431) is the first half of a two-semester sequence (with 432) focused on modern data analysis and advanced statistical modeling, with a practical bent and as little theory as possible. We emphasize the key roles of thinking hard, and well, about design and analysis in research.

The course is formally titled *Statistical Methods in Biological & Medical Sciences, Part 1*. A more accurate title is **Data Science for Biological, Medical or Health Research**.

We'll learn about managing and visualizing data, building models and making predictions, and other data science activities. This highly applied course focuses on modern tools for learning from data. We'll learn a lot of R, and we'll use RStudio and Quarto as tools to help make R work better, and help perform our research in rigorous and replicable ways.

## 4.1 Course Objectives

During the 431-432 sequence, students will:

1. Use modern data science tools to import, tidy/manage, explore (through transformation, visualization and modeling) and communicate about data.
2. Think hard and well about rigorous design and analysis in scientific research.
3. Gain sufficient background in the practical issues regarding linear and generalized linear models to give you a starting place for meaningful applied work, particularly in terms of making comparisons to address general types of statistical and analytic questions (exploratory, predictive, inferential, and causal, in particular.)
4. Learn about the importance of replicable research, and develop facility and practice in open source tools for doing it.
5. Complete a series of assignments (labs, projects and quizzes) designed to help you demonstrate what you've learned.
6. Using R and RStudio to make all of the things above happen; with particular emphasis on doing replicable research and using [Quarto](#) to document the work in a replicable way.

This is **NOT** a course in mathematical statistics or statistical inference. It's far more applied than that.

## 4.2 Key Topics in 431-432

1. Exploratory Data Analysis: “All graphs are comparisons” including data exploration, statistical graphics and more general visualization of information.
2. Placing biological, medical and health research questions into a statistical framework.
3. Study Development - making choices in designing and executing the collection and aggregation of data.
4. Data Handling - including important issues in importing, tidying and transforming data, as well as methods for dealing with missing data, including imputation.
5. Statistical Comparisons: “All of statistics are comparisons” - including methods for discrete and continuous variables: intervals, assumptions, some thoughts on statistical power, and the bootstrap, design of visualizations and models for rates, proportions and contingency tables.
6. The proper and rigorous use of multi-predictor models for continuous and discrete data, including...
  - Fitting, evaluating, and interpreting linear and generalized linear models.
  - Prediction and validation.
  - Critical role of graphics, including diagnostics and residual analysis.
  - Model choice, including variable selection, shrinkage and model uncertainty.
  - Dealing with categorical predictors and interactions meaningfully.
  - Causal inference using regression: controlling for covariates meaningfully.
7. Using R and RStudio to make all of the things above happen; with particular emphasis on doing replicable research and using [Quarto](#) to document the work in a replicable way.

## 4.3 431 Course Outline & Format

The main group sessions for the 431 course will include 24 in-person lecture sessions led by Professor Love, to be held on Tuesdays and Thursdays from 1:00 to 2:15 PM in E321-323 in the Robbins Building at the CWRU School of Medicine.

- The [Course Calendar](#) provides additional detail on specific sessions, and links to materials used in those sessions, including slides.

### 4.3.1 Part A: R and Exploring Data

Classes 1-10 (roughly) focus on this material.

- Exploratory Data Analysis
  - Descriptive Numerical and Graphical Summaries

- Histograms and their cousins
- Scatterplots and related tools from correlation and linear regression
- Dealing with Missing Data
- The Importance of the Normal Distribution
- Exploring Data with the Tidyverse, Getting Up To Speed with R
  - Visualizing Data with `ggplot2`
  - Data Transformation and `dplyr`
- Using Quarto
  - Building Quarto documents within R Projects
  - Building R code using Quarto, and publishing/sharing results

#### **4.3.2 Part B: Making Comparisons**

Classes 11-16 (roughly) focus on this material.

- Estimation and Inference for Means and Proportions (especially)
  - Confidence Intervals
  - Design Implications: Matched vs. Independent Samples
  - Hypothesis Testing Strategies and why significance isn't so helpful
  - Cross-Tabulations
  - Randomized Trials vs. Non-Randomized Studies

#### **4.3.3 Part C: Linear Models**

Classes 17-24 (roughly) focus on this material.

- Estimation and Inference using Ordinary Least Squares
  - Simple and Multivariate Linear Regression Models
  - Building Prediction Models, and Validating Them
  - Categorical Variables, Analysis of Variance
  - Analysis of Covariance
  - Residual and Influence Analyses
  - Foundations of Model / Feature / Variable Selection
  - What you've learned in the past and how it wasn't so helpful

## 4.4 Prerequisites and Intended Student Population

What do we expect you to know already before you start the course? **Not much.**

Useful prior experience includes training/experience in statistics, coding/programming and biology/biomedical science. We expect most people will have some experience in one or two of these areas, but very few will have all three.

- Some students have lots of prior training in statistics. But there are many students in the class with no statistical training at all that they use regularly. We assume only that everyone knows what an average is, and has some sense of why statistics might be useful to them in their chosen field.
- Some students have lots of prior coding and programming experience, including experience with R and/or Quarto. Some have never written a line of code in their life. We assume only that everyone is willing to learn how to do modern work with data, and that means writing computer code, but that some people will be starting from nothing.
- Some students have lots of prior experience with biological and biomedical science, and know a lot of useful things in those areas which relate directly to our work. Others have zero experience in this area, and will learn a lot from their colleagues. We assume only that everyone is willing to learn, and to put in some effort to do so.

People succeed in this course with a wide range of backgrounds and a common interest in using data effectively in research related to biology, health or medicine. There will be multiple people in the class who are years away from their last statistics class. We expect the majority of students will have no prior experience using R, or any meaningful recollection of using statistical software.

The pace can be brisk at times, but all CWRU students who feel up to it are welcome, regardless of their field of study or prior experience. Section 100 (Professor Love's section) is specifically geared towards students in programs under the auspices of the Department of Population & Quantitative Health Sciences, as well as students who intend to continue on and take 432 this Spring. Section 101 (with Professor Zhang) is more appropriate for most other students.

## 4.5 Motivations for our Approach

Professor Love has a lot of thoughts on this issue and you'll hear about them through the semester, but you may prefer to hear from other people on the subject. So here are a few references that guided our thinking in developing the course.

- [A Guide to Teaching Data Science](#) by Stephanie C. Hicks, Rafael A. Irizarry (pdf)

- ... our (case-study) approach (in a graduate-level, introductory data science course) teaches students three key skills needed to succeed in data science, which we refer to as creating, connecting, and computing.
- [Data Visualization on Day One: Bringing Big Ideas into Intro Stats Early and Often](#) by Xiaofei Wang, Cynthia Rush, Nicholas Jon Horton (pdf)
- [50 Years of Data Science](#) by David Donoho in the *Journal of Computational and Graphical Statistics*, 2017.
- [Why You Should Master R \(Even if it might eventually become obsolete\)](#) blog post from Sharp Sight, 2016-12-27
- [Teaching R to New Users - From tapply to the Tidyverse](#), A YouTube video by Roger D. Peng
- [Teach the Tidyverse to Beginners](#) and a related post on teaching `ggplot2`, specifically from David Robinson. There is also a related video from `rstudio::conf` 2018.
- Video from Hadley Wickham, [You can't do data science in a GUI](#), 2018 in Chicago.

## 4.6 Is 432 Required?

If I take 431 this semester, do I have to take 432 in the Spring?

It is the natural thing to do, and I assume that almost all of you will do so. The 431 course is part 1 of a two-semester sequence. Frankly, 432 contains some of the most interesting material and is generally regarded by students who take both as the more entertaining course. Every year, some students take only 431, though. The decision is up to you. The 432 course assumes you have completed 431, whether with me or another instructor.

# 5 Required Texts

## 5.1 Course Notes

Professor Love maintains a set of [Course Notes](#). Professor Love revises the Notes every year, and so they appear in pieces as the semester progresses.

Although these Notes share some of the features of a textbook, they are neither comprehensive nor completely original. The main purpose is to give 431 students a set of common materials on which to draw during the course, providing a series of examples using R to work through issues that are likely to come up during the semester, and in later work.

In addition, **slides** and video recordings from each of Professor Love's lectures, plus other in-class materials will be posted for your use throughout the semester.

Once class begins, you'll be able to access all materials (including the Course Notes) through the main course website at <https://thomaselove.github.io/431-2023/>.

## 5.2 Buy This Book!

During the course, we will read David Spiegelhalter's **The Art of Statistics**, which was first published by Penguin in March 2019 (and February 2020) in the UK and then by Basic Books in the US in September 2019. You can purchase any of the available versions (hard-cover, paperback or e-reader) online or in your local bookstore for about \$20.

- [Dr. Spiegelhalter's website](#) has lots of useful information.
- The [book's website](#) contains R code, corrections and other materials.
- You are welcome to read this book before class starts, if you'd like to get a jump on things, but that's not necessary: we'll link readings to the [Course Calendar](#).

## 5.3 Other Books to Download

There are three additional **free** books that you will definitely need to obtain during the semester and may be interested in looking at before class begins. Simply visit the links below.

1. [R for Data Science](#) (2nd edition) by Hadley Wickham, Mine Cetinkaya-Rundel and Garrett Grolemund.
  - Solutions to the exercises in R4DS can be found [here](#) and may be very helpful to you.
2. [R Graphics Cookbook](#) (2nd edition) by Winston Chang.
3. [Biostatistics for Biomedical Research](#) by Frank E. Harrell Jr.
  - The related set of YouTube videos can be found [here](#).

There are many other free R books available online, which may be helpful to you. Here are a few that some students in past versions of this course have liked.

- [OpenIntro Statistics](#) by David Diez, Mine Cetinkaya-Rundel, Christopher Barr and OpenIntro.
- [Introduction to Data Science: Data Analysis and Prediction Algorithms with R](#) by Rafael A. Irizarry
- [Fundamentals of Data Visualization](#) by Claus O. Wilke
- [Data Visualization: A practical introduction](#) by Kieran Healy

## 5.4 Key Articles and Posts

While I will recommend dozens, perhaps hundreds of articles, blog posts and the like to you over the course of the year, these are especially important in 431.

1. Several of the guides prepared by Jeff Leek and his group, including:
  - [Finally, a Formula for Decoding Health News](#), from fivethirtyeight.com
  - [Reading academic \(scientific\) papers](#),
  - [Writing your first academic paper](#)
  - [Write papers like a modern scientist](#)
  - [How to Share Data for Collaboration](#) by Shannon E. Ellis and Jeffrey T. Leek in *The American Statistician*, 2018 Special Issue on Data Science, or you can [read the PeerJ preprint version here](#).
2. [Data Organization in Spreadsheets](#) by Karl W. Broman and Kara H. Woo in *The American Statistician*, 2018 Special Issue on Data Science, or you can [read the PeerJ preprint version](#).
  - The Ellis/Leek and Broman/Woo papers are part of the [Practical Data Science for Stats](#) collection, which may be of interest.
3. [Project-oriented workflow](#) at tidyverse.org from Jenny Bryan.
4. From the Ten Simple Rules series at PLOS Computational Biology:

- [Ten Simple Rules for Effective Statistical Practice](#) by Kass RE et al. 2016
  - [Ten Simple Rules for Graduate Students](#) by Gu J Bourne PE 2007
  - [Ten Simple Rules for Better Figures](#) by Rougier NP Droettboom M Bourne PE 2014
  - [Ten Simple Rules for Creating a Good Data Management Plan](#) by Michener WK 2015
5. [Statistical Inference in the 21st Century: A World Beyond  \$p < 0.05\$](#)  from 2019 in *The American Statistician*
  6. The American Statistical Association's 2016 [Statement on p-Values: Context, Process and Purpose](#).

Professor Love's class-specific READMEs will provide links to these articles and other recommended readings as the semester goes on.

# 6 Required Software

The course makes heavy use of the R statistical programming language, and several related tools, most especially the RStudio development environment. Every bit of this software is free to use, and open-source.

You will need access to a computer to do your work for this class, not just an iPad or other tablet, but an actual computer. You do not need a state of the art machine, nor should you need any special hardware to run things for this course.

- There will be many people in the course for whom R is a new experience. I assume no prior R work in the course. You will know a fair amount of R (and some other things, too) after taking the course, though.
- We'll be using [Quarto](#) for most of our work, as well, which can be used to generate reproducible reports that appear as .html websites, PDF files or Word documents, among other things.
- For some people, working with R and Quarto is the best part of the class, and the part that they're most excited about.
- For others, it's a real source of anxiety. We understand and encourage patience. There will definitely be some pain, but our experience is that things are much smoother for most people by early October than they appear to be in August.

## 6.1 R and RStudio

You will do all of your analysis with the open source (and free!) [programming language R](#). You will use [RStudio](#) as the main program to access R. Think of R as an engine and RStudio as a car dashboard. R handles all the calculations and the actual statistics, while RStudio provides a nice interface for running R code.

R is free, but it can sometimes be a pain to install and configure. Information about getting R and RStudio on your computer is provided below, and [also on our main website](#).

Learning R can be difficult at first - it's like learning a new language, just like Spanish, French, or Chinese. Hadley Wickham-the chief scientist at Posit (makers of Rstudio) and the author of some amazing R packages you'll be using like `ggplot2` [made this wise observation](#):

It's easy when you start out programming to get really frustrated and think, "Oh it's me, I'm really stupid," or, "I'm not made out to program." But, that is absolutely not the case. Everyone gets frustrated. I still get frustrated occasionally when writing R code. It's just a natural part of programming. So, it happens to everyone and gets less and less over time. Don't blame yourself. Just take a break, do something fun, and then come back and try again later.

If you're finding yourself taking way too long hitting your head against a wall and not understanding, take a break, talk to the teaching assistants, talk to classmates, ask questions, e-mail Professor Love, etc.

I promise you can do this.

Some of this material is also borrowed from Andrew Heiss, especially [here](#) and [here](#).

## 6.2 Getting the Software

Everything is free, but it does require some patience to get control over your computer.

### 6.2.1 System Requirements

You will need access to a computer to do your work for this class, not just an iPad or other tablet, but an actual computer. Whether or not you want to bring that computer to class is up to you. All of the software we will use in this class is either free and open source, or available to you for free through your affiliation with CWRU, so there is nothing to buy in terms of software.

- We've made some effort in terms of course requirements to set the bar low. You do not need a state of the art machine, nor should you need any special hardware to run things for this course.
- You will need a computer, either PC (running Windows 10 or 11, ideally) or Macintosh (running a reasonably recent OS), but your choice should be determined by your personal preferences and how you believe you will use the machine in your research life. RStudio and R will look and work the same on either a PC or a Macintosh.
- We **do not** recommend the use of a Chromebook for 431 or 432.
- R and RStudio Desktop also run on Linux systems but Professor Love knows essentially nothing about that. Consult the documentation [at CRAN for R](#) and [at the download page for RStudio](#).

### 6.2.2 What will I need for 431?

These instructions are also available on [the Software page](#) at our main course website: <https://thomaselove.github.io/431-2023/>.

## 6.3 Installing R and R Studio

The steps you need to complete are:

1. Download and install **the latest version of R** (version **4.3.1 or later**) from <https://cran.case.edu/> or, if you prefer, from <https://cloud.r-project.org/>, which automatically chooses a fast, nearby mirror for you.
  - If you have a pre-existing installation of R and/or RStudio, you will need to re-install both to get current.
2. Download and install **RStudio Desktop** (Open Source Edition - the free version **2023.06.1+524 or later**) by visiting <https://posit.co/download/rstudio-desktop/>.

## 6.4 Installing R Packages and Data/Code for 431

After you've completed steps 1 and 2 above, move on to these tasks:

3. Install some **R packages** - an R “package” is a collection of functions, data, and documentation that extends the capabilities of R, and is the critical way to get R doing interesting work.
  - Visit <https://github.com/THOMASELOVE/431-packages> for details and the list of packages we'll be using in 431.
4. Download **data and code** (functions) we've developed specifically for 431 at <https://github.com/THOMASELOVE/431-data>.
  - Follow the instructions you'll find there.

## 6.5 Need Installation Help?

If you need more help, you might look at [this terrific resource for Installing R and RStudio](#) from Jenny Bryan and the STAT 545 project. These are the people responsible for the great [Happy Git with R](#) project, which will also be worth your time when we are using Git and GitHub.

- If you’re having trouble with installation before our first class, don’t worry too much. The TAs and Professor Love will be available to help once the class gets going.
- Once the class starts, if you’re having installation problems or problems getting started in R, please consider visiting TA office hours or perhaps asking a question on Campuswire. We want to hear from you!

## 6.6 Updating Your R Packages

Every once in a while, it’s a good idea to update your R packages. To do so,

1. Go to the **Packages** tab on the right side of your RStudio screen, and click on **Update**.
2. This will bring up a dialog box. I usually click **Select All**, then click **Install Updates**.
  - A popup box may appear, asking “Do you want to install from sources the packages which need compilation?” to which I usually answer **No**. A **Yes** response leads to a slower installation, but can solve problems if you still have them after updating.
  - This may take a few minutes. As long as you’re seeing activity in the Console window, things are progressing.
  - Eventually, you’ll get a message that “The downloaded source packages are in ...” with a directory name. That’s the sign that the updating is done.
  - Updating packages is something you’ll do occasionally throughout the semester, mostly when a problem happens.
3. Finally, choose **File ... Quit Session** from the top menu, and accept or deny (I usually deny) RStudio’s request to save your workspace. Then close RStudio, and re-open it if you want to do some work.

## 6.7 Why do we teach R in 431-432?

Why do we teach using R, rather than SAS or SPSS or Python or whatever?

1. Because it is by far the better choice for what we’re trying to do, which is to help you become effective data scientists. And effective scientists, period.
2. Because being a data scientist means writing code and actually doing (not just talking about) replicable research, which R facilitates in an immense variety of ways.
3. Because R is free to you, me and everyone, and its community is a daily delight.

To read comments from other people on the subject, I suggest reading [Why R?](#) from Chester Ismay and Patrick Kennedy.

Also, the question of “Why R and not SPSS?” was nicely addressed by Greg Snow in [this 2010 post at StackOverflow...](#)

When talking about user friendliness of computer software I like the analogy of cars vs. busses: Busses are very easy to use, you just need to know which bus to get on, where to get on, and where to get off (and you need to pay your fare). Cars on the other hand require much more work, you need to have some type of map or directions (even if the map is in your head), you need to put gas in every now and then, you need to know the rules of the road (have some type of drivers licence). The big advantage of the car is that it can take you a bunch of places that the bus does not go and it is quicker for some trips that would require transferring between busses. Using this analogy programs like SPSS are busses, easy to use for the standard things, but very frustrating if you want to do something that is not already preprogrammed. R is a 4-wheel drive SUV (though environmentally friendly) with a bike on the back, a kayak on top, good walking and running shoes in the passenger seat, and mountain climbing and spelunking gear in the back. R can take you anywhere you want to go if you take time to learn how to use the equipment, but that is going to take longer than learning where the bus stops are in SPSS.

## 6.8 Additional Resources

If you're interested in getting started with the tools you'll be using in 431 before the class gets rolling, the great folks at [RStudio Education](#) provide [these 6 ways to begin learning R](#). Pick the one that appeals to you, and give it a shot.

- Note that if you're having trouble installing things, you can still learn a lot about R, RStudio and Data Science basics with the interactive tutorials at <https://rstudio.cloud/learn/primers>.
- If you're already a strong coder, and have some R experience, there are also [learning pathways for intermediates at RStudio Education](#) which might be appealing to you.

Our goal is to get everyone well into the intermediate level by December. Some people will get there in September, for others it will take longer. But you can do this, and we'll be there to help you.

In addition, there are many, many online resources to help you with working in R and in [Quarto](#), and we'll point you to some of the best of them during the semester.

## 7 Professor Love



Thomas E. Love, Ph.D.

Hi. I am Thomas E. Love, Ph.D. and I have at least three different jobs.

- I am a Professor in the Departments of Medicine and Population & Quantitative Health Sciences at Case Western Reserve University. I teach three courses per year there (PQHS 431, 432 and 500) and also lead the Health Care Analytics track of the MS program in Biostatistics.
- I direct the [Biostatistics and Data Science Division](#) at the [Population Health Research Institute](#) and at the Center for Health Care Research & Policy, which are jointly supported by The MetroHealth System and by CWRU.
- For ten years, I was the (founding) Data Director for [Better Health Partnership](#), an alliance of people who provide, pay for and receive care in Northeast Ohio. I now serve as Chief Data Scientist there.
- I am a Fellow of the American Statistical Association, and was recently elected 2024 Chair for its Council of Sections Governing Board.
- I have won numerous awards for my teaching and my research, including the 2018 [John S. Diekhoff Award for Graduate Teaching](#) from CWRU.
- I have been teaching at CWRU since 1994, and have taught every manner of CWRU student over the years, especially students in biostatistics, medicine, and management.

In research, I use statistical methods to look at questions in health policy and in particular the provision of health services. I mostly work with observational data, rather than data that emerge from randomized clinical trials, and I have a special interest in working with data from electronic health records.

- For example, you may be interested in a [study in Health Affairs](#) showing the impact of a [Medicaid-like expansion plan on care and outcomes of poor patients in Cleveland](#).
- Or you might be interested in our [New England Journal of Medicine study](#) of the effect of electronic health records on the care and outcomes of people with diabetes.
- In 2011, [James O'Malley](#) and I chaired the [Ninth International Conference on Health Policy Statistics](#), here in Cleveland. Here's a [recap](#).
- I've also worked on many projects involving the use of propensity scores to make causal inferences from observational studies, particularly in heart failure.

If you want to see a [pretty complete list of my publications](#), knock yourself out.

I hold degrees from Columbia University in the City of New York and from the University of Pennsylvania. My dissertation adviser was Paul Rosenbaum. I am married to a brilliant woman who is an attorney at Savant Systems, and we have two terrific sons, one a recent graduate of the University of Pittsburgh, and the younger a junior at Columbia University. I live in Shaker Heights. I also sing and act occasionally in [community theater](#). I'll let you know if I'm in any shows this Fall.

## 7.1 Email

- **Thomas dot Love at case dot edu** (for matters related to grades or individual concerns)
- Professor Love is best reached via email in an emergency.

## 7.2 Offices

- Wood WG-82J on the ground floor of the Wood building (Tuesday and Thursday afternoons)
- Rammelkamp R-229A at MetroHealth Medical Center (Friday mornings)

At other times, Professor Love works from his home in Shaker Heights.

Professor Love is available for the 10 minutes before and approximately 20 minutes after each class for drop-in discussions and otherwise by appointment (send email to schedule).

### **7.3 Name, Pronouns, Social Media**

- Professor Love uses he/him/his pronouns, and identifies as male.
  - Most students refer to him either as Professor Love or Doctor Love, and he prefers this in class.
  - He prefers his given name to be written “Thomas” as opposed to “Tom”.
  - Most of his friends and colleagues call him “Tom”. You are welcome to do so, outside of class time, if it makes you comfortable.
- At Github, he is [THOMASELOVE](#).
  - The Github repository [for this syllabus is here](#).

# 8 Teaching Assistants

Almost all of this year's stellar group of teaching assistants have been in your shoes - they've taken the course in the past, and they enjoyed it enough to come back for more. They are volunteering their precious time and energy to help make the course happen, and we couldn't be more delighted to welcome you to the course. All office hours and all teaching assistant work will be done remotely.

There is a lot to learn in this class. Don't suffer in silence - talk to us! As mentioned earlier, **you shouldn't spend more than 15 minutes struggling with anything in this class without asking for help.**

Visit Getting Help (Chapter 9) of this Syllabus for more details.

## 8.1 TA office hours

TA office hours are held via Zoom, will begin on Tuesday 2023-09-05 and continue through Monday 2023-12-11. The weekly schedule will appear on our [Contact Us page](#) as well as our [Course Calendar](#).

- Our Shared Google Drive (431 Fall 2023 Dr Love and Students) will eventually contain Zoom links for each of these sessions in a document called **TA office hours schedule and Zoom links**.
  - No appointment is necessary. Please just drop in.
  - Be sure to log into Google via CWRU to access the shared Drive.
- TA office hours are cancelled on 2023-10-23 and 2023-10-24 for Fall Break and from 2023-11-19 through 2023-11-25 for Thanksgiving break.

Professor Love will also hold "office hours" before and after each class to the extent possible.

## 8.2 Getting To Know The TAs

The Fall 2023 Teaching Assistants for 431 are:

TA Name	For more, visit	Current Roles and CWRU Experiences
Ali Elsharkawi	Section <a href="#">8.2.1</a>	third year MS student in Clinical Research
Chenyu Liu	Section <a href="#">8.2.2</a>	first year PhD student in Epidemiology & Biostatistics
Alex Olejko	Section <a href="#">8.2.3</a>	fifth year PhD candidate in the Developmental, Cognitive, and Affective Sciences program in the Department of Psychological Sciences
Lindsay Petrenchik	Section <a href="#">8.2.4</a>	PGY1 Pharmacy Resident at University Hospital Samaritan Medical Center and Clinical Instructor of Pharmacy Practice at Northeast Ohio Medical University and MS graduate
Miza Salim Hammoud	Section <a href="#">8.2.5</a>	MS student in Clinical Research and Postdoc Fellow at Cleveland Clinic
Monika Strah	Section <a href="#">8.2.6</a>	PhD student in Clinical Translational Science and MS graduate
Ria Tilve	Section <a href="#">8.2.7</a>	first year MD student and MPH graduate

### 8.2.1 Ali Elsharkawi



Figure 8.1: Ali Elsharkawi

Ali Elsharkawi is in his third year in the M.S program in Clinical Research. He completed his bachelor's in medicine, and fellowship in Electrophysiology at Metrohealth in Cleveland. For his research interest, he studied the outcome of atrial fibrillation ablation procedures. Ali took 431 and 432 two years ago and he is applying the skills he learned in his research projects. Ali enjoys spending time with his family, playing music, watching soccer, and traveling.

### **8.2.2 Chenyu Liu**



Figure 8.2: Chenyu Liu

Chenyu Liu is a first-year Ph.D. student in Epidemiology and Biostatistics within the Department of Population & Quantitative Health Sciences. Chenyu completed her undergraduate studies in statistics at Wuhan University of Technology in China and her master's program in statistics at Case. Her current research interest is Bayesian Statistics, multi-omics data integration, and proteomics analysis. Outside of her Ph.D. life, Chenyu spends her time traveling, cooking, and watching movies.

### **8.2.3 Alex Olejko**

Alex is a 5th year PhD candidate in the Developmental, Cognitive, and Affective Sciences program in the Department of Psychological Sciences. He completed my undergraduate education at Ohio State, and earned my Master's degree from CWRU in 2022. His research mainly resides in the field of behavior genetics, where he focuses on polygenic influences on behavioral outcomes. These outcomes mostly encompass educational and health phenotypes in the early to mid-twenties. Alex grew up outside of Cleveland and am a fervent Cleveland sports fan. He loves going to Guardians games and have already been to 20 this year (and



Figure 8.3: Alex Olejko

counting). Some of his other favorite past-times include building Lego, mixology, exercise, and (mainly) single-player video games.

#### **8.2.4 Lindsay Petrenchik**

Lindsay Petrenchik is a PGY1 Pharmacy Resident at University Hospital Samaritan Medical Center and a Clinical Instructor of Pharmacy Practice at Northeast Ohio Medical University. Prior to residency, Lindsay worked at Novartis for two years where she evaluated Patient Reported Outcomes (PROs) in oncology clinical trials. She is excited to share with you how she has used 431/432 super-power skills daily in both pharmaceutical industry and residency research projects. Come to her TA sessions and you will have a blast learning R and statistical methods!

#### **8.2.5 Miza Salim Hammoud**

Miza Salim Hammoud is a postdoctoral research fellow at Cleveland Clinic's Pediatric and Congenital Heart Surgery Department and is in her second year of the Masters of Science in Clinical Research (CRSP) program at CWRU. Her research goal is to construct predictive models that illuminate the path ahead for little lives touched by congenital heart defects, predicting both clinical outcomes and the quality of life perceived by patients and their families. Miza took 431 and 432 last year and is applying the amazing research skills gained from these two classes in her research projects. Beyond the realm of academia and research, Miza enjoys playing the piano's melody, painting colors on canvas, and embarks on adventures that span the globe embracing the wonders of new countries and cultures.



Figure 8.4: Lindsay Petrenchik



Figure 8.5: Miza Salim Hammoud

### **8.2.6 Monika Strah**



Figure 8.6: Monika Strah

Monika Strah is a Ph.D. student in Clinical Translational Science at CWRU and recently completed the Biostatistics M.S. program. She is looking forward to returning as a TA for PQHS 431/432 and also enjoyed TAing for PQHS 453 this summer. Monika's past research includes mathematical approaches to questions in population genetics and human evolution and is presently studying dietary factors associated with lung cancer risk and multiple sclerosis related fatigue management strategies. Monika enjoys reading, hiking and is a member of the National Ski Patrol.

### **8.2.7 Ria Tilve**

Ria Tilve is a first year medical student at CWRU. After completing her bachelor's in Biochemistry, she obtained her MPH in Population Health Research at CWRU last spring. Her current research focuses on musculoskeletal health and biomedical imaging. Ria took 431 and 432 during the 2021-2022 school year, and she is applying those skills in her research projects. Outside of studying, Ria enjoys spending time with her family, running, watching sports, and exploring new places.



Figure 8.7: Ria Tilve

# 9 Getting Help

Data science and statistical programming can be difficult. Computers are stupid and little errors in your code can cause hours of frustration (even if you've been doing this stuff for years!). **You shouldn't spend more than 15 minutes struggling with anything in this class without asking for help.**

We are fully committed to making sure that you learn everything you were hoping to learn from this class. If you tell us you're having trouble, I will not judge you or think less of you. I hope you'll extend me the same grace. You never owe us personal information about your health (mental or physical). If you need extra help, or if you need more time with something, or if you feel like you're behind or not understanding everything, do not suffer in silence! Talk to us! We will work with you.

Once the course begins, you have several ways to get help from us:

## 9.1 Campuswire is the place to ask questions and discuss the course

- We will use [Campuswire](#) this semester to facilitate questions and answers. We'll be conducting all class-related discussion there, instead of via email.
- You can sign up for this class on Campuswire once Dr. Love makes it available.
  - See the [Campuswire page](#) on our website for sign-up details, including the Class Join Code.
  - Be sure to sign in with your full name and an email address that you will check almost every day.
  - If you're successfully enrolled in the class, you should be able to sign up immediately and gain access to Professor Love's 431 course.
    - \* The quicker you begin asking questions on Campuswire (rather than via emails), the quicker you'll benefit from the collective knowledge of your classmates and instructors.
    - \* We encourage you to ask questions when you're struggling to understand anything related to the class, and you can even do so anonymously.
    - \* You'll also be able to help your colleagues by answering their questions (and you'll get some class participation credit for doing so), as well.

In general, we don't have a way to diagnose your problem with R, RStudio or Quarto if you don't show us what you're typing that causes an error, or a lack of results. If you wrote a Quarto file, we need to see it, along with a specific question (or series of them) about specific error messages or strange results you are getting. We need to replicate the problem in order to know how to fix your problem, and it also helps if we know what error message you're seeing, or what strange result you are getting.

## 9.2 TA Office Hours is the place to get one-on-one help for 431

You are strongly encouraged to visit TAs during office hours that are convenient for you, whenever you need help with the course. This is the best way to get one-on-one help, or try out a new idea.

- Teaching Assistant (TA) office hours will be held via Zoom, beginning in early September and ending in early December
- The schedule for TA office hours will appear on the [Course Calendar](#), and on our [Contact Us](#) page.
- Office Hours are held for your benefit. Please visit the TAs, especially if you have questions about how to do things in R, or about the content of the class slides or course notes. They're volunteering their time to help you, and most of them have been in your shoes in previous years. They want to help you!
- More information on the TAs will be found in Chapter [8](#).

## 9.3 How We'll Use Canvas

The course's primary listing on [Canvas](#) is PQHS 431, but students in CRSP 431 and MPHP 431 should find the same information. We'll use the Canvas platform for (at most) two things:

- (definitely) turning in lab and project assignments
- (maybe) providing information about (and maintaining recordings of) our class sessions

Visit <https://canvas.case.edu/> to log in and see what's available there.

## 9.4 How We'll Use Google Drive

Professor Love will share some data and some other materials (answer sketches for lab assignments, for example) through a Shared Google Drive. You will need to log into Google through your CWRU account in order to view and use these materials.

## 9.5 Meeting with Professor Love

Professor Love does not hold regular office hours. He will arrive 10 minutes before each class, and stay afterwards, as well, to address any drop-in student concerns that can be addressed in that setting.

- If you have any unique concerns about the course, need special accommodations or have any other issues you want to discuss with Professor Love specifically, please email him ([Thomas dot Love at case dot edu](mailto:Thomas.dot.Love@case.edu)) and he'll respond as quickly as possible.
- Remember that most questions about course materials are best asked through [Campuswire](#) or at TA Office Hours so that we can share the answers across all students in the course.

## 9.6 Other Sources of Help

In addition, there are many, many online resources to help you learn R beyond those we provide.

- Building internet searches (“googling”) effectively is an important skill that requires practice. Don’t be afraid to use search engines to figure out how to code things in R, for example.
- The [RStudio Community](#) at Posit is a forum specifically designed for people using RStudio and the tidyverse (and that will be you.)

# 10 Assignments

Course assignments include (in increasing order of importance):

- minute papers (and class participation)
- labs
- quizzes, and
- projects

The Assignments page linked on the [main course website](#) provides links to additional information on these assignments.

## 10.1 Deadlines for Assignments

All deadlines are posted as part of the [Course Calendar](#), and the Calendar is the final word for all deadlines.

- On every assignment, Lab, Quiz, project-related task, whatever, we will be delighted to respond to questions **up to 12 hours before** the assignment is due. After that time, you may wind up on your own. The reason for this is that Professor Love and the teaching assistants will regularly post responses to frequently asked questions about assignments, and we need sufficient time to accomplish this task.

## 10.2 Using Artificial Intelligence Tools

See Section [12.9](#) for details on this course's policies on using AI composition software (such as ChatGPT) for assignments in this course.

## 10.3 Minute Papers

During the semester, there are **10** Minute Papers scheduled. In this class, a Minute Paper is a short survey (completed using a Google Form) where you will answer a few questions about how the course is going for you. To complete the Minute Paper, you'll need to log into Google

with your CWRU account. The [Welcome to 431 Survey](#) uses a similar format, but is more extensive than a typical Minute Paper.

Each Minute Paper ideally takes less than five minutes to complete, and is best done right at the end of Tuesday's class. The Minute Papers are actually due at noon Wednesday to give you some time to reflect, if you'd like, or if you're in a hurry at the end of Tuesday's class. The Minute Papers are linked on the [431 Minute Paper page](#).

You **should** complete the Minute Paper even if you weren't able to attend the most recent class.

Only Professor Love sees the results of the Minute Papers with identifying information. He then summarizes findings without identifying people and addresses those as part of the README discussion in the Thursday class whenever possible.

Professor Love reserves the right to change the number of minute papers depending on how the class is going during the semester.

## 10.4 Class Participation

We cannot emphasize enough how much we want to hear from you about things that are relevant to this course.

1. If you're not shy, ask questions during class. Come to the (virtual) TA office hours if you need help.
2. Contribute to Campuswire with questions, answers and helpful comments. That'll help us recognize you as someone trying to improve their understanding.
  - If you find **typos** in the materials (code, slides, the Notes, this syllabus) send them to us on Campuswire.
  - See a cool visualization online? A nice use of statistical methods or design in a paper? Share them with us on Campuswire.
3. Email Professor Love if you want to discuss something that doesn't fit on Campuswire or in a conversation with a TA. Visit with him before or after class. Make an appointment to talk with him if you have something to discuss that doesn't work well in email.
  - It's always helpful if Professor Love knows who you are, and this is an especially difficult thing to achieve with a large class and masks.

## 10.5 Labs

During the semester, there are **seven** main lab assignments scheduled. Professor Love reserves the right to adjust the number of labs depending on how the class is going during the semester.

The [Course Calendar](#) provides deadlines for all Labs, and the [431-Labs page](#) links to additional information on all Labs.

- Most labs require straightforward demonstrations of mastery for core principles and fundamental skills. Some require deeper dives into more technically sophisticated material.
- Most labs ask you to clean/visualize/analyze some data that we provide to you, and also to react to something we're reading at the same time in the Spiegelhalter book.
- You should anticipate that some coding in R **and** writing at least one essay will be required in most labs.

Lab assignments are submitted through [Canvas](#). The link to post your responses for each Lab will appear in time for you to submit the work, usually just after the deadline for the preceding Lab has passed.

Each lab is associated with a deadline. Late work is permitted, but penalized (see the Grading section below, and the [431-Labs page](#) for more details.)

### 10.5.1 Lab X

There will be an (optional) bonus assignment called Lab X, which involves building a website for yourself. See the [Lab X instructions](#) for more details.

### 10.5.2 A few tips for Early Lab Work

1. When writing anything for this course in English, use complete sentences, not just bullet points.
2. Clearly mark each Question in each Lab. There is no need to repeat the question before answering it, although you are welcome to do so.
3. You are welcome to discuss each Lab with anyone, including Professor Love, the teaching assistants, or your colleagues, but your answer must be prepared by you alone. We especially encourage you to take advantage of TA office hours and Campuswire.
4. Don't leave anything until the last day. Allow time for computer problems.

## 10.6 Quizzes

During the semester, you will have two quizzes to complete. Each quiz is taken online via a Google Form (like the Minute Papers or the [Welcome to 431 survey](#).) Each quiz includes both short-answer and multiple-choice items. The quizzes focus on whether you can apply the tools developed in class to address small problems.

The [Course Calendar](#) provides deadlines for each of the Quizzes, and the [431-Quiz](#) page provides links to instructions for Quizzes.

You will receive access to the quiz on the Thursday before it is due (the next Tuesday at noon), and each quiz is designed to take about six hours to complete, so the time pressure should be minimal. **If you need to make alternate arrangements for a Quiz, please contact Professor Love via email as soon as possible**, at least a week before a Quiz is released.

### 10.6.1 Campuswire and the Quizzes

As we've discussed, for most of the course, Campuswire will be our tool for asking and answering questions.

**However, for questions about the Quizzes**, we may provide an alternate method for you to ask questions of Professor Love and the Teaching Assistants, and if so, we will share that approach at the top of the Quiz instructions when they become available.

### 10.6.2 About the Quizzes

1. Quizzes typically involve 20-30 questions.
2. The questions are not arranged in any particular order, and you should answer all questions.
3. All questions involve relatively short responses, sometimes after working through a detailed analysis.
4. You will have the opportunity to edit your responses after completing the Quiz, but this must be completed by the deadline.
5. You are welcome (even encouraged) to consult the materials provided on the course website, but you are **not** allowed to discuss the questions on the Quizzes with anyone other than Professor Love or the teaching assistants.
6. We do not guarantee to answer questions we receive about the Quiz less than 3 hours prior to the Quiz submission deadline. As noted above, we will **not** use Campuswire for Quiz questions, but instead an alternate approach we will provide to you with each quiz.
7. Quizzes that are more than an hour late will **not** be accepted, except in truly remarkable circumstances.
8. An answer sketch for each Quiz will be made available within 48 hours of the deadline.

9. Grades for the Quizzes are usually available within 48 hours of the deadline.
10. If you feel Professor Love has made an error in grading your Quiz, please let him know directly, by email, as soon as possible.

Dr. Love prepared a document containing five old quiz questions, as well as an answer sketch for those five questions, and some information on results. You'll find it all here, if you are interested.

## 10.7 Projects

During the semester, you will have two projects to complete.

- [Project A](#) anticipates you using materials from the first 15 classes, working with a data set Dr. Love provides to you.
- [Project B](#) anticipates you using materials from the entire term, working with a data set you identify yourself.

The [Course Calendar](#) provides deadlines for all Project elements, and each project will have a special website of its own which will provide full directions.

## 10.8 Why Two Projects?

The main reason is that I can't figure out a way to get you to think about all of the things I hope you'll learn from this course in a single Project. Another important reason is that I want you to be able to make mistakes during the semester without worrying about it too much, and having two projects spreads out this learning a bit.

1. I set different tasks for Project A and for Project B, allowing us to touch on a wider fraction of the things I hope you'll learn in 431.
2. I give much more guidance in Project A than in Project B.
3. We have to evaluate each of your projects quickly, and there are many students in the class. Knowing the data set you'll be working with helps us manage this.
4. Having a broad range of activities to evaluate helps reduce the cost of a mistake on any one of them, so that we can build on what you do well.
5. All of Project A can be done using materials discussed in Classes 1-15.

# 11 Grading

Over the past five years, all students taking the course have eventually earned either an “A”, a “B” or an “Incomplete” in 431. I don’t see a reason why this won’t also be the case this year. The course cannot be taken “pass/fail.”

## 11.1 Grading Breakdown

Grading standards apply in the same way for all students, regardless of whether they are enrolled in PQHS 431, CRSP 431 or MPHP 431. Students in this class are not in competition with each other for grades. I have no set percentage of students who will receive any particular grade. We aim for substantial and timely feedback from the teaching assistants and myself on all elements of the course.

The course grade is based on four key areas of demonstrated accomplishment.

Course grades are based primarily on performance on labs, projects and quizzes. Your grade on each of these three elements will be determined separately. Then we will combine these elements as follows:

Element	Weight 1	Weight 2	Weight 3	Weight 4
Labs	35%	30%	25%	-
Projects	45%	40%	35%	-
Quizzes	25%	20%	15%	-
Participation	-	-	-	10%

Among Labs, Projects and Quizzes, your element with the strongest score will receive Weight 1, your element with the second strongest score will receive Weight 2 and your element with the weakest score will receive weight 3. The Class Participation score (which includes the Minute Papers) is worth 10% for all students.

Professor Love will then determine your final course grade using the following scale:

- course averages of 88.0% to 100% will receive a A
- course averages of 85.0% to 87.9% will be reviewed more closely to determine whether the grade will be A or B

- course averages of 70.0% to 84.9% will receive a B

For students in the 85.0-87.9% range, students who have improved from the first half of the course to the second will be more likely to receive an A than those whose performance has declined.

Students whose course average is near or below 70% usually have failed to complete some element of the course, and need to take an Incomplete to complete their work before I assign a final grade. We follow the University's Incomplete Policy, as described below.

## 11.2 Grading Class Participation

Student Class Participation grades are determined by Professor Love out of a maximum of 100 points. This determination is based on:

- 30 points for completing the [Welcome to 431 survey](#) that most of you did prior to class, and if you haven't done it, please do so now.
- plus up to 45 additional points for Minute Paper responses (5 points for each on-time submission, 3 points for late submissions within 24 hours of the deadline.)
- plus up to 35 bonus points for actions that help the class, as awarded during the semester

These bonus points have a small impact, generally, but include things like:

- finding typographical or coding errors in Professor Love's slides, notes and other published materials. Professor Love makes occasional typographic and grammatical errors, which irritate him enormously. Please post to Campuswire the details of any typos or errors you find in this syllabus or any other course materials. If you are the first to let us know, and we make the change, you will receive some credit in your class participation grade.
- asking especially helpful questions in the Minute Papers or on Campuswire
- providing helpful responses to questions and engaging in the discussion on Campuswire
- identifying interesting visualizations or uses of ideas we're discussing in class in the popular or scientific media (this is also done via Campuswire.)
- completing (in a timely fashion) special "partially hidden" tasks that Professor Love buries within longer documents to ascertain whether or not you're reading these documents effectively. For an example, see section 14 of this Syllabus.
- plus a few other things that come under the general heading of "being helpful."

Thanks to the bonus, scores up to 110 / 100 in Class Participation are technically possible. Bonus credit is capped so that no one's class participation score will exceed 105 points.

## 11.3 Grading Labs

The seven Main Labs are graded by the Teaching Assistants, according to a rubric and answer sketch. In grading labs, everything counts, including the quality of the presentation, as well as the coding and statistical work.

Grading for Lab X is discussed as part of the [Lab X instructions](#), but completion of Lab X can only help your grade in the course.

For each of the main 7 Labs, we publish a detailed answer sketch 48 hours after the Lab's submission deadline. Later, when the grades become available (roughly one week after the deadline) you will also receive the grading rubric used by the teaching assistants. Along with the grade, on-time labs will receive modest feedback about the work.

All Labs should be turned in by the deadline specified on the [Course Calendar](#). In this class, it is far better to turn in timely, but only partially complete work than nothing at all. We don't give extensions on Labs. Instead, we use the following grading policy.

- To receive full credit on a Lab, it must be received on Canvas no later than 59 minutes after the posted deadline. (This allows for small issues with uploading to Canvas to occur without penalty.)
- Labs that are turned in 1-12 hours after the deadline will lose 10% of available points.
- Labs turned in more than 12 but less than 48 hours after the deadline will lose 30% of available points.
- No extensions to Lab deadlines will be permitted this semester. Labs turned in more than 48 hours after the deadline will receive no credit, since by then the Lab Sketch will be posted.

Your lowest lab score (out of the first seven) over the course of the semester will be dropped before we calculate your overall lab grade.

- If you decide to skip a lab, please submit a note to Canvas by the deadline saying that you are skipping the lab.
- Should you run the risk of needing to miss more than one lab during the semester, contact Professor Love via email.

## 11.4 Lab Appeal Policy - Request a Review via Google Form

Clarification of concerns related to potential typographical or other errors in our Lab answer sketches is welcome at any time, but haggling over points on assignments can be a real time sink in a large class. To that end, students are **requested not to dispute** any grading on Labs during the semester, but instead to request a review using the [Lab Regrade Request Form](#).

Should you receive a lower grade on a Lab than you feel you deserve based on the feedback the TA provides, the answer sketch and the grading rubric, you can submit a request for a regrade by Professor Love as follows...

- On the course's [Labs page](#), we include a section about Lab Regrade Requests. There, you'll find a link to the [Lab Regrade Request Form](#) (you must log into Google via CWRU to see it) listing each of the Labs. Any student who wishes to dispute points can specify the number of points in question for each item, and the details of the issues that concern them.
  - All forms must be submitted by the deadline in December specified on the Course Calendar.
  - If you wish to dispute a grade, just fill out the form at any time before the deadline in December. You can edit the form after submitting it so that you can add additional requests up until the end of the semester.
  - Disputing a grade on a homework with a teaching assistant is pointless.
  - Professor Love is responsible for all final grading decisions, and if you have a concern, you should submit your request through the form.

Students are welcome to ask questions of Professor Love about grading during the term. The TAs and Professor Love are happy to discuss in a general sense any questions about an assignment, but no grades will be changed until the end of the term. - The one exception is if there is a mistake in adding up points, or some similar clerical error. If you find such an issue, please bring it to Professor Love's attention via email, and such problems will be corrected immediately.

In mid-December, **after** Professor Love has worked out what letter grade to give each student, he will go through the requests on [the form](#) and determine for each whether the student's letter grade would change if all of the points in dispute were granted. If the answer is no, then he won't even look at the disputed grade(s). If the answer is yes, then he will look very carefully to see if enough extra points are merited to change a grade. (It will not help your case to submit frivolous requests.)

The main advantage of this system is that it saves all of us the hassle of haggling over points that are never going to mean anything anyway. It also provides "equal access" to students who are too timid to express their concerns. Finally, if there is an issue with grading a particular problem or assignment that needs to be reconsidered, Professor Love will have access to all papers and can make a universal decision<sup>1</sup>.

---

<sup>1</sup>I got this idea from Jessica Utts.

## **11.5 Grading the Quizzes**

Each quiz will be graded solely by Professor Love. A detailed answer sketch and grading rubric will be provided to you when grades are available, within 48 hours of the quiz deadline.

In calculating your quiz average for your course grade, Professor Love weights your stronger Quiz performance at 60% and your other Quiz at 40%.

## **11.6 Grading the Projects**

Some elements of each project will be graded by Professor Love, and others by the teaching assistants.

In calculating your project average for your course grade, I will weight Project A at 40% and Project B at 60%.

Students who cannot complete a project (either A or B) in a timely fashion will receive an incomplete grade until both Projects are completed. A penalty will be applied to late work on projects.

Once the complete project instructions are available, please contact Professor Love directly if you are in any way concerned about your ability to complete a project on time.

## **11.7 Late Assignments**

If an emergency arises that will keep you from timely completion of work that has a deadline, like a lab, minute paper, or the like, send an email directly to Professor Love saying the following:

I have an emergency that will keep me from completing (list things you're worried about completing) in a timely fashion. I will contact you again when I am able to make plans.

Professor Love will respond with “OK” and temporarily excuse you from things until you send a “re-connection” email to make plans for making up the work you missed, at which time he’ll evaluate the situation and together you can make that plan. At no time do you owe us any information about your health or the nature of the emergency.

## **11.8 Incomplete Grades**

Here is the official University policy:

The grade of Incomplete is assigned at the discretion of an instructor provided that:

1. There are extenuating circumstances, explained to the instructor before the assignment of the grade, which clearly justify an extension of time beyond the requirements established for other students in the class. It is the student's responsibility to notify the instructor of the circumstances preventing completion.
2. The student has been passing the course and only a small segment of the course remains to be completed, such as a term paper, for which the extenuating circumstances justify a special exception.

An Incomplete grade may not be assigned if a student is absent from a final examination, unless the dean has authorized the absence. Unauthorized absence from a final examination will result in a failing grade. When the student completes the work, the Incomplete is changed to an A, B, C, D, P, F, or NP.

All work for the incomplete grade must be made up and the change of grade recorded in the Office of the University Registrar by the date specified by the instructor, but no later than the last day of class in the semester following the one in which the Incomplete was received. A student who has a permanent Incomplete for a required course must retake the course in a later term. If the student cannot complete the work by the end of the following semester, he or she must petition for an extension which must be endorsed by the instructor, explain the reasons why the work has not been completed, and include a new date for completion. Students will be allowed only one extension of no more than one additional semester to complete the work for an Incomplete grade.

# 12 General Policies

All Case Western Reserve University policies apply to this class, including those summarized below. To the extent that those policies are unclear, I would regard an appropriate code of conduct as mirroring [this one from the R Consortium](#).

Any concerns or questions regarding these general policies, the teaching assistants or the course itself should be directed to Professor Love, if at all possible.

## 12.1 Attendance

Things happen, and people occasionally cannot make it to class in person. For 431, we expect all students to attend **at least 19** of the 24 sessions in person.

- If you need to miss a class session, or even two, just review the materials and move on, and please feel encouraged to ask for help.
- If you must miss **more than two classes** in a row, then it is your responsibility to email Professor Love as soon as possible to let him know.
  - An appropriate email would read “I am unable to attend class on [specific dates] and I will complete all necessary work while I am away.” No additional explanation is required.

## 12.2 Auditing

Enrolled students and teaching assistants are the only people permitted to attend the course, without special permission from Professor Love.

- I do not permit anyone to audit the course who has not previously taken it, without exception.

## **12.3 Recordings**

We aim to have video recordings available the day after each class session.

- “Live” Zoom attendance will not (generally) be possible.
- Technical problems may occur during the semester with the recordings.
- You should assume that anything you say in class will be recorded.

## **12.4 Diversity and Inclusion**

It is our intent that all students regardless of their background and perspective be well-served by this class. Further, we intend to present material whose content is respectful of diversity (gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture) and deliver it in a way that respects these differences as well. We expect that all students, instructors and guests will help foster an atmosphere of respect, trust and safety in the classroom.

If you have suggestions for how to make the class content or environment more inclusive, or have specific incidents to report, please reach out to Professor Love directly. If you are not comfortable reaching out to him, please feel encouraged to reach out to someone else, such as the School of Medicine Graduate Education Office ([som-geo@case.edu](mailto:som-geo@case.edu)) or the Office of Inclusion, Diversity and Equal Opportunity (OIDEO) ([oideo@case.edu](mailto:oideo@case.edu)). More information on University policy and resources are available on OIDEO’s website at <https://case.edu/diversity/>.

## **12.5 Disability Resources**

In accordance with federal law, if you have a documented disability, you may be eligible to request accommodations from Disability Resources. In order to be considered for accommodations you must first register with the Disability Resources office. Please contact their office to register at 216.368.5230 or visit their website at <https://case.edu/studentlife/disability/> to get more information on how to begin the process.

Once you have registered with Disability Resources, they will contact Professor Love to discuss any necessary arrangements. Please feel encouraged to get in touch with Professor Love directly to discuss any specific concerns you have.

## **12.6 Laptop and cell phone use**

Here is the official University policy:

“Although having a laptop in class opens up new learning possibilities for students, it can be used in ways that are inappropriate. It is easy for your laptop to become a distraction to you and to those around you. Laptops are to be used only when essential to the task at hand. Please turn off or silence all cell/smart phones, tablets, and other electronic devices for the duration of the course. Inappropriate uses will be noted and may affect the final grade.”

Professor Love adds that, in this class, the use of a laptop or other device is welcomed, and even encouraged, for taking notes on our slides, or following along with the presentation, or trying out ideas in R, but it is inappropriate for you to use such a machine for activities unrelated to the class during our time together. Silence your phone during class. The temptation to look at your phone or Facebook or email during class is nearly irresistible. Resist anyway, if only to avoid distracting your instructor and your fellow students.

Thank you.

## **12.7 Academic Integrity**

Here is the official policy of the University on Academic Integrity.

Any violation of the University’s Code of Ethics will not be tolerated. All forms of academic dishonesty including cheating, plagiarism, misrepresentation, and obstruction are violations of academic integrity standards and will result in a minimum penalty of receiving a zero for the assignment, the potential for failing the entire course. Cheating includes copying from another’s work, falsifying problem solutions or laboratory reports, or using unauthorized sources, notes or computer programs. Plagiarism includes the presentation, without proper attribution, of another’s words or ideas from printed or electronic sources. It is also plagiarism to submit, without the instructor’s consent, an assignment in one class previously submitted in another. Misrepresentation includes forgery of official academic documents, the presentation of altered or falsified documents or testimony to a university office or official, taking an exam for another student, or lying about personal circumstances to postpone tests or assignments. Obstruction occurs when a student engages in unreasonable conduct that interferes with another’s ability to conduct scholarly activity. Destroying a student’s computer file, stealing a student’s notebook, and stealing a book on reserve in the library are examples of obstruction. In addition, the incident will be reported to the Dean of Undergraduate Studies and

Academic Review Board for undergraduates or Senior Associate Dean of Graduate Studies, for Graduate Students.

The CWRU Statement of Ethics for graduate students can be found here: <http://case.edu/gradstudies/about-the-school/policies-procedures/>.

The University's [Academic Integrity Policy](#) is also worth reading.

## **12.8 Plagiarism**

Professor Love, and the faculty of the Department of Population & Quantitative Health Sciences expect and require original writing for all assignments given. Submitting plagiarized work for an academic requirement is a violation of the academic integrity standards set forth by the University. Plagiarism is the representation of another's work or ideas as one's own; it includes the unacknowledged, word-for-word use and/or paraphrasing of another person's work, and/or the inappropriate unacknowledged use of another person's ideas. Submitting substantially the same work to satisfy requirements for one course that has been submitted in satisfaction of requirements for another course, without permission of the instructor of the course for which the work is being submitted, is also prohibited.

If you have any questions about this policy as it applies to 431, please direct them to Professor Love.

## **12.9 Use of AI is Not Permitted without Express Written Consent from Dr. Love**

You must obtain written permission from Dr. Love before using AI composition software (such as ChatGPT) for assignments in this course. Using these tools without my permission puts your academic integrity at risk. Intellectual honesty is central to the academic experience at CWRU. Our Academic Integrity Policy prohibits submitting AI-generated text as one's own, just as it prohibits claiming authorship for text that was originally produced by another person.

More on AI and Academic Writing can be found at <https://case.edu/artscli/writing/resources/ai-academic-writing>

## **12.10 University Resources / Student Support**

Case Western Reserve University has a wide range of student support resources on campus. These resources can assist students both in the classroom and outside of it. CWRU offers these resources to enrich and enhance each student's academic experience and to create an environment that encourages personal growth and development. The Department of Population and Quantitative Health Sciences provides a detailed list [on its website in the Resources section](#). Please take advantage of these resources.

In particular, visit <https://case.edu/studentsuccess/academic-resources> for links to various programs and opportunities to help you learn more effectively. In particular, students needing some help with English, or writing skills will find multiple resources that have proven very helpful in the past.

## 13 Movies

Once you have read through this Syllabus, please e-mail Professor Love the poster for one of your favorite movies. Use the subject line: **431 Favorite Movie**. If he receives your poster by 9 AM Eastern time on Tuesday 2023-09-05, he will add some bonus credit to your class participation grade.

Thank you.

# 14 General Advice

My most important piece of general advice to people is to be kind. That's not always the thing I do as well as I'd like.



**The Meanest TA**  
@MeanestTA

Hell hath no fury like a full professor mildly inconvenienced.

11:32 AM · Jun 12, 2019 · [Twitter for iPhone](#)

## 14.1 On Graduate School

For graduate students, I urge you to take as much advantage of this learning experience as you can. While I'll refer to some of the pieces below during the course, I've gathered a few favorites here.

1. From [matt.might.net...](#)
  - [12 Resolutions for Grad Students](#)
  - [How to get a great letter of recommendation](#)
  - [How to send and reply to email](#). You might also want to look at [Email Etiquette: Guidelines for Writing to Your Professors](#)
2. [Four Things You Should Do When You're Bored](#), on YouTube (the four things are Exercise / Read / Meditate / Find and Engage a Hobby with Passion). Also, [Get Up and Move. It May Make You Happier](#) by Gretchen Reynolds, in the *New York Times*, 2017-01-25.
3. [Why academics need to focus on structuring their time](#) from *University Affairs*.

4. Most people are really bad at meetings, including Professor Love. In a series of tweets, [Greg Wilson argues that](#) “the single most useful training you can give an adult is how to run a meeting and how to participate in someone else’s.” He then provides some nice tips in subsequent tweets, although the link provided there to a blog post is no longer working.
5. Some people need help taking notes. You might be interested in [Cornell Note Taking](#) or the [5 other methods described here](#).
6. You may be interested in the American Statistical Association, and its [This is Statistics](#) program.

## 14.2 On Seeking a Job

If I have a job or internship to offer, I'll be noisy about it in class. In the meantime, I'd consider joining the [American Statistical Association as a student member](#) and perhaps joining the [Greater Cleveland R Meetup Group](#).

Here are some gathered thoughts from other folks that you might enjoy:

1. [General Advice on an Academic Career Path](#) (which is filled with useful advice, especially for those studying biostatistics.)
2. [Advice for Applying to Data Science Jobs](#) from Emily Robinson
3. [Academic job search advice](#) from [matt.might.net](#)
4. [Career Advice for Data Scientists](#) was a panel at rstudio::conf 2020 focused on building a career around R.

# 15 On Writing

## 15.1 Writing, Presenting & Communicating

I write all the time. It's my job. It's yours, too. You'll do more of it here than you may be expecting in this class, and in life. So you'll need to take advantage of every opportunity you have to write more effectively tomorrow than you do today.

Here's what I find to be a compelling argument from George Cobb ...

If you want your work with data to make a difference, devote time and effort to choosing the words and pictures you use to present your evidence and conclusions. If you teach or supervise, seek to reward those - they who learn from you, and they who report to you – when they spend time crafting their message.

Think back to the last “report” you were expected to read. Is it easy to recall the main points? The answer, yes or no, depends not so much on the quality of the data, the effort, and the thinking that went into the report, nor on your own dutiful diligence in reading the report, but rather, and mainly, on whether the people who wrote the report had learned and practiced the skills of how to use words and pictures, first, to claim attention, and second, to claim retention: to deliver a message that sticks in the mind.

## 15.2 Campus Resources

The Department of Population and Quantitative Health Sciences provides a detailed list [on its website in the Resources section](#) of links to several important resources for improving your English and your communication skills. Also helpful is <https://case.edu/studentsuccess/academic-resources>. We strongly encourage you to take advantage of these resources.

## 15.3 Artificial Intelligence (ChatGPT)

We will discuss some of the flaws and opportunities of AI composition software as part of the Communication goals for this course. Using these tools without my permission puts your

academic integrity at risk. Our Academic Integrity Policy prohibits submitting AI-generated text as one's own, just as it prohibits claiming authorship for text that was originally produced by another person.

More on this course's policies regarding AI and Academic Writing can be found in Section Section 12.9.

## 15.4 Advice from Other People

1. [Preparing Manuscripts for Submission to Medical Journals: The Paper Trail](#) by H. Gilbert Welch, from *Effective Clinical Practice* in 1999.
  - Start early, focus on high-visibility components, develop a systematic approach to the body of the paper, finish strong.
  - Use [thinkchecksubmit](#) and [see this tweet](#) for some related suggestions when selecting a journal or publisher for your work.
2. [Writing a Scientific Paper in Four Easy Steps](#) from Claus Wilke at [The Serial Mentor blog](#).
3. [Rules to write a good research paper](#) from Daniel Lemire.
4. [Hey-here are some tips on communicating data and statistics!](#) from Andrew Gelman 2017-06-02.
5. [Writing Pet Peeves: Correctness, References, and Style](#) from Tamara Munzner.
6. Frank Harrell's [Checklist for Authors of Statistical Problems to Document and to Avoid](#)

## 15.5 A Few Tips from Professor Love

1. Statistics is a “getting the details right” business - we care deeply about details, and this applies to writing code or complete English sentences. RStudio has a spell-checker. To use it, click F7.
2. Nothing impresses us as much as a clear and concise argument, presented using well-written English sentences, effective and well-labeled figures and tables.
3. Don’t parrot back material that Professor Love wrote or said. State ideas in your own words. Stating them in my words is, technically, plagiarism.
4. Edit your more adventurous output; don’t present everything you know how to do in R, and don’t forget that someone is trying to read both your code and your results.
5. Make your work easy to evaluate. In responding to an assignment, be sure to answer the question that was asked, restating it as necessary.
6. Clearly label everything: graphs, tables, your answer to a specific question. Everything. Again, make your work easy to evaluate.
7. Simplify. Emphasize ideas in plain language. Avoid jargon. Use English well.

8. Data are plural. Use “the data **are** ...” rather than “the data *is* ...”
9. A paragraph must contain more than one sentence.
10. Don’t switch tenses. If you want to write in the present tense, stick to it throughout.
11. Don’t write or say random sample unless you used a random number generator. If you used haphazard sampling or convenience sampling, call it what it is, and indicate whether any problems could have cropped up as a result.
12. Similarly, don’t defend a method of data collection because it is random. Most of the time we want to represent some population, and a random sample is just one way to ensure that certain types of biases have a low probability of creeping in.
13. If you want to write that you used  $\alpha = 0.05$  as your significance level, then state that your results were obtained using a 95% confidence level, not a 95% confidence interval, unless you are actually interpreting a confidence interval.
14. If you find yourself in the appalling situation of writing about a  $p$ -value, then you should state something like:
  - [1] We’re using a 95% confidence level.
  - [2] We’re using a 5% significance level.
  - [3] We’re using  $\alpha = 0.05$ .
  - Don’t use more than one of these expressions.
15. Again, don’t use  $p$ -values in most settings, but if you must, refer to all  $p$ -values that are less than 0.001 or perhaps less than 0.0001 as  $p < 0.001$ , rather than, for instance,  $p = 0.00000001$  or, worse yet,  $p = 0$ . In a similar vein, write all  $p$ -values that exceed 0.99 as  $p > 0.99$  instead of, for instance,  $p = 1$ .
16. To the extent possible, don’t use **computer-ese** to label variables, plots or tables. R and Quarto allow you to change the labels on graphs and tables to meaningful things – do so. Use meaningful abbreviations, as necessary, explaining what they mean on the first usage.
17. When in doubt, err on the side of clarity. Clear thinking leads to clear writing.