PQHS 431 Syllabus - Fall 2019

Thomas E. Love, Ph.D.

Version: 2019-09-18 06:49:43

Contents

Critical Information

This is the Fall 2019 syllabus for PQHS / CRSP / MPHP 431: Statistical Methods in Biological & Medical Sciences, Section 1 (with Professor Thomas Love).

Course Home Page

The course home page, with everything you'll need, is at https://github.com/THOMASELOVE/2019-431.

Working with This Document

1. This document is broken down into multiple sections. Use the table of contents at left to navigate.
2. At the top of the document, you'll see icons which you can click to
   - search the document,
   - change the size, font or color scheme of the page, and
   - download a PDF version of the entire document.
3. The document is updated occasionally through the semester. Check the Version information above to verify the last update time.

Who, When and Where?

- Section 1 is taught by Professor Thomas Love. He uses he/him pronouns. More on him here. Most of my students call me Professor Love, or Dr. Love.
- In this effort, he is assisted by five veterans of the class. Details on the Teaching Assistants here.
- Once the semester begins, reach us at **431-help at case dot edu**.
- The course is given on Tuesdays and Thursdays from 1:00 to 2:15 PM, in Room E321-323 in the Robbins building at the School of Medicine.

- The first class is Tuesday 2019-08-27.

Getting Help

To get help for anything related to the course, email **431-help at case dot edu**.

- Professor Love is available on Tuesdays and Thursdays at CWRU, by appointment. To make an appointment, email him at `thomas.love@case.edu`.
- For drop-in conversations, Professor Love is always available for 15 minutes before and 30 minutes after class. If he's not in the classroom then, stop by his office on the ground floor of the Wood building, WG-82J.
  - If you have any special concerns about the course, need special accommodations or have any other issues for Professor Love, please email or speak with him before or after class.
- TA office hours are held on the ground floor of the Wood building, in the computing lab (WG-56) or the student lounge (WG-67), beginning 2019-09-03. See the Course Calendar for the schedule.

Data science and statistical programming can be difficult. Computers are stupid and little errors in your code can cause hours of headache (even if you've been doing this stuff for years!). There are many, many online resources to help you with this, in addition to emailing us at `431-help at case dot edu`. An especially useful one is the RStudio Community (a forum specifically designed for people using RStudio and the tidyverse (i.e. you)).

Before the First Class

The first class meeting is from 1:00 to 2:15 PM on Tuesday 2019-08-27, in Room E321-323 of the Robbins building at the School of Medicine.

"Welcome to 431" Survey

At your earliest convenience, please fill out the survey at http://bit.ly/431-2019-welcome-survey so we can get to know you a little bit better.

What Do I Need To Buy?

We'll read two books that you'll need to purchase (the combined price is about $25.):

1. Nate Silver's *The Signal and The Noise*. ISBN-13: 978-1594204111 Amazon Link, and
2. Jeff Leek's *The Elements of Data Analytic Style*, available at https://leanpub.com/datastyle.

Everything else that you will need is free, and is described in the remainder of this syllabus or on the course website at https://github.com/THOMASELOVE/2019-431.

Chapter 1

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 role 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 R Markdown as tools to help make R work better, and help perform our research in replicable ways.

1.1 What Students Should Expect

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 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, including homeworks and quizzes on data provided for you, and course projects using data you select/develop.
6. Program ("Code") in R sufficiently to accomplish the tasks above, with enough self-sufficiency afterwards to be able to debug and use new R tools without substantial troubleshooting help. What separates "doing data science" from "doing data analysis" is programming.

1.2 431 Class Outline

The 431 course **calendar** is linked at https://github.com/THOMASELOVE/2019-431. Go there for details on each class throughout the semester.

The course is split in three parts.

Part A (Classes 2-10) is mostly about R and Visualizing Data.

- Exploratory Data Analysis
  - Descriptive Numerical and Graphical Summaries
  - Distributions, specifically the Normal
  - Histograms and their cousins
  - Scatterplots and related tools from correlation and linear regression
- Exploring Data with the Tidyverse, Getting Up To Speed with R
  - Visualizing Data with `ggplot2`
  - Data Transformation and `dplyr`
  - Using scripts and projects, Building Code

Part B (which starts around Class 11) is about Making Comparisons.

- Estimation and Inference for Means and Proportions
  - Confidence Intervals
  - Design Implications: Matched vs. Independent Samples
  - Hypothesis Testing Strategies
  - Cross-Tabulations
  - Dealing with Missing Data
  - Randomized Trials vs. Non-Randomized Studies

Part C (which starts around Class 16) is about Building Regression Models.

- Estimation and Inference using Ordinary Least Squares
  - Categorical Variables, Analysis of Variance
  - Simple and Multivariate Linear Regression Models
  - Building Prediction Models, and Validating Them
  - Analysis of Covariance
  - Residual and Influence Analyses
  - Foundations of Model / Feature / Variable Selection

1.3 Key Topics in 431 and 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 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 Markdown to document the work.

1.4 What We Expect You To Know Already

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. 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.

1.5 R and RStudio

(borrowed from Andrew Heiss)

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. To make life easier, you can (and should!) use the free RStudio.cloud service, which lets you run a full instance of RStudio in your web browser. This means you won't have to install anything on your computer to get started with R! We will have a shared class workspace in RStudio.cloud that will let you quickly copy templates for some of your work, too.

RStudio.cloud is convenient, but it can be slow and it is not designed to be able to handle larger datasets or more complicated analysis. Over the course of the semester, you'll probably want to get around to installing R, RStudio, and other R packages on your computer and wean yourself off of RStudio.cloud. This isn't absolutely necessary early on, but you'll probably feel the need to have done it by the end of September. You'll find complete instructions for installing, R, RStudio, the R packages we'll use and the data and code we've built for 431 here.

1.6 Pep Talk! (Thanks, Andrew Heiss)

Learning R can be difficult at first - it's like learning a new lanugage, just like Spanish, French, or Chinese. Hadley Wickham-the chief data scientist at 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 classmates, ask questions at `431-help at case dot edu`, e-mail Dr. Love, etc.

I promise you can do this.

1.7 Why I Teach 431 Like This

I have a lot of thoughts on this issue, but you may prefer to hear from other people on the subject. So here are a few references that have guided my recent thinking.

- 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 by Roger D. Peng, which is also available as a YouTube Video
- Teach the Tidyverse to Beginners from David Robinson at `rstudio::conf 2018` (video), and here is the blog post and a related post on teaching `ggplot2`, specifically from David.
- Video from Hadley Wickham, You can't do data science in a GUI, 2018 in Chicago.

Chapter 2

Professor Love

Thomas E. Love, Ph.D.

- Professor of Medicine, Population and Quantitative Health Sciences, CWRU
- Director of Biostatistics and Evaluation, Center for Health Care Research & Policy, MetroHealth Medical Center
- Chief Data Scientist, Better Health Partnership
- Track Lead for Health Care Analytics, MS in Biostatistics, Department of Population and Quantitative Health Sciences, CWRU
- Fellow, American Statistical Association

2.1 A More Complete Biography

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 Biostatistics and Evaluation at the Center for Health Care Research & Policy, which is a joint venture of CWRU and MetroHealth Medical Center.
- 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 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.

- You may be interested in a recent 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 GE Lighting, and we are raising two terrific sons, one in college (University of Pittsburgh), and one finishing high school this year. I live in Shaker Heights. I also sing and act occasionally in community theater.

2.2 Email

- Email to get help with the course: **431-help at case dot edu** (seen by Professor Love and the TAs)

- `Thomas dot Love at case dot edu` (for matters related to grades or individual concerns)
- Professor Love is hard to reach by phone. Email is always the best way to reach him.

2.3 Offices

- Wood WG-82J on the ground floor of the Wood building (Tuesdays and Thursdays)
- Rammelkamp R-229A at MetroHealth Medical Center (Wednesdays and Fridays)

Professor Love is available for the 15 minutes before and the 30 minutes after each class, and otherwise by appointment on Tuesdays and Thursdays (send email to schedule).

2.4 Name and Pronouns

- Professor Love uses he/him/his pronouns.
- Most students refer to him either as Professor Love or Dr. Love.
- 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, as well, if that makes you more comfortable.

2.5 Web

- Professor Love's GitHub pages website.
  - His GitHub name is THOMASELOVE.
- His Twitter handle is ThomasELove.

Chapter 3

Teaching Assistants

Each of this year's stellar group of teaching assistants has 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.

The TAs this year are:

- [Benjamin Booker, BS]
- Julijana Conic, MD
- Joseph Hnath, BA
- Noah Lorincz-Comi, MSc
- Amr Mahran, MD MS
- Harry Persaud, BS
- Amin Saad, MD

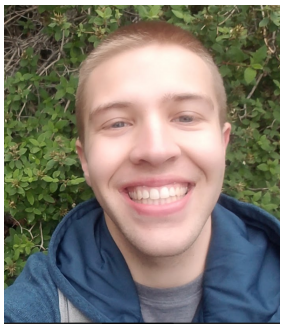
There are two sections of this course (Section 1 with Professor Love, and Section 2 with Professor Li) but all TAs work with both sections. To contact Professor Love, Professor Li and the TAs, email **431-help at case dot edu**. This is a difficult class for many people. Don't suffer in silence - talk to us!

3.1 Office Hours for TAs

Teaching Assistant Office Hours are held in WG-56 (Computing Lab) or WG-67 (Student Lounge) on the ground floor of the Wood building, so be sure to look in both places if you need help.

TA Office Hours begin on 2019-09-03. The schedule is posted at the bottom of the Course Calendar.

3.2 Benjamin (Ben) Booker, BS



Benjamin (Ben) Booker is a first year PhD student in the Epidemiology & Biostatistics program in the Department of Population & Quantitative Health Sciences. Ben holds a BS in Molecular Biology from the University of Cincinnati, and then completed two years of additional training in Biostatistics there. He has worked at Cincinnati Children's Hospital performing DNA methylation analysis, and as a data scientist consultant for Givaudan Flavors. Outside of work and school I enjoy rock climbing/bouldering (novice level), playing soccer and watching the European football leagues.

3.3 Julijana Conic, MD

Julijana Conic was born in Serbia and received her MD from the University of Belgrade Faculty of Medicine last year. Since enrolling in the MS in Clinical Research program the same year she has been conducting research focusing on ischemic mitral regurgitation in the Department of Cardiovascular Imagining at the Cleveland Clinic. Currently, she is working on a project incorporating machine learning to improve existing algorithms for automatic quantification of cardiac volumes on MRI images and to aid in risk stratification of ischemic mitral regurgitation patients. She hopes to start internal medicine residency next year and ultimately establish herself as a physician investigator. During her free time Julijana enjoys hiking, watching movies and volunteering in the community.

3.4 Joseph Hnath, BA

Joseph Hnath is in his second year of the Master of Public Health program on the Intensive Research Pathway with concentrations in Population Health Research and Health Policy & Management. He finished his undergraduate studies at CWRU this May where he majored in Chemical Biology, Cognitive Science, and Economics. Having taken 431 & 432 last year, the skills he learned have been invaluable in his research projects, such as his capstone on the health economics of abortion policy and helping with the NEO-CASE cancer disparities resource. Joseph enjoys playing basketball, watching Master Chef, and reading *The Complete Works of F. Scott Fitzgerald.*

3.5 Noah Lorincz-Comi, MSc

Noah Lorincz-Comi is a first-year student in the Epidemiology & Biostatistics PhD program at Case. Before beginning his PhD, Noah earned a Master's degree in Psychiatric Research from King's College London in the UK, and before that he earned a Bachelor's degree in Psychology from Ohio University. He is very interested in better understanding psychotic illnesses like bipolar and schizophrenia, and hope that good training in statistical methods can help him and colleagues better understand these diseases. Outside of class, Noah likes

soccer, drawing, and painting.

3.6 Amr Mahran, MD MS

Amr Mahran is a urologist who is working as a senior research associate in the department of urology, CWRU School of Medicine. He received his MD degree from Assiut University School of Medicine in Upper Egypt. He also finished a residency in urology along with earning a Master of Science degree. Before joining CWRU, Amr was a practicing urologist and was appointed as a faculty at the department of urology, Assiut University Hospitals. Amr took 432 in the spring of 2019 and learned many skills that helped him in his clinical research. Amr's research focus on prostate cancer, pelvic pain, and voiding dysfunction. He does outcome research on large databases as NSQIP, National Trauma Database (NTDB), and NIS databases. Amr enjoys playing soccer, table tennis, and reading.

3.7 Harry Persaud, BS

Harry Persaud is a second-year student in the M.S. Biostatistics program. He obtained his Bachelor's degree in Health Science from DePaul University in Chicago. Before enrolling at Case, he worked at Boston Children's Hospital

and Harvard Medical School. Harry took 431 and 432 last semester, which helped prepare him for a research internship at Better Health Partnership where he currently provides research support on observational studies and projects dealing with social determinants of health. In his free time, he enjoys trail running, backpacking, and finding any excuse to spend time in the Sierra Nevada mountains.

3.8 Amin Saad, MD

Amin Saad is an international medical graduate from Syria with two years of General Surgery residency training experience in the United States. Amin is currently enrolled in the CRSP Master's program and is seeking a Ph.D. degree in Clinical and Translational Research with a focus toward lowering surgical site infection rates. Amin took 431 and 432 two cycles ago and has appreciated how the skills he learned in those classes have helped him with his clinical outcomes research at the Department of Colorectal Surgery at University Hospitals. Amin enjoys playing soccer, swimming, and spending time with his family.

Chapter 4

Required Texts

4.1 Professor Love's Materials

The main text is a set of Notes for the course, maintained by Professor Love. The title is "Data Science for Biological, Medical and Health Research: Notes for PQHS 431".

Professor Love is revising the Notes extensively this year, so they will appear as the semester progresses. The 2018 version of the Notes is available, too.

Although the 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. Again, this is work in progress, and updates will occur irregularly as the semester progresses.

Slides from each of Professor Love's lectures, plus other in-class materials from each session of the class will be posted before and after each class as part of daily README files discussed at the start of each session.

Visit https://github.com/THOMASELOVE/2019-431 for links to all materials.

4.2 Two Books To Purchase

In addition, we'll read two books that you'll need to purchase (the combined price is about $25.):

1. Nate Silver's *The Signal and The Noise* ISBN-13: 978-1594204111 Amazon Link, and

2. Jeff Leek's *The Elements of Data Analytic Style*, available at https://leanpub.com/datastyle.

With regard to *The Signal and the Noise*, you can watch Nate discuss the book's ideas in many places, for instance, at this YouTube link, or this one on the Art and Science of Prediction, or this one at Google. We'll also spend considerable time visiting the FiveThirtyEight website, where Nate is editor-in-chief.

4.3 Four Books to Download

There are four books that you will definitely need to obtain during the semester. All are freely available to you, at the links below.

1. R for Data Science by Garrett Grolemund and Hadley Wickham
2. Biostatistics for Biomedical Research (pdf) by Frank E. Harrell Jr and James C Slaughter
3. OpenIntro Statistics by David Diez and Mine Cetinkaya-Rundel (4th Edition) and supplementary material is here
4. Modern Dive: Statistical Inference via Data Science (A Modern Dive into R and the Tidyverse) by Chester Ismay and Albert Y. Kim.

4.4 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.

Links to all required and recommended materials appear on the Course Website's READINGS page.

Chapter 5

Worthy (and Free) Resources

Many of these resources will come up again in class, but no one can keep up with all of this material. Pick things that interest you to follow up with. And I'm always eager to receive additional suggestions from students in the class. If you find a helpful resource, please send it along to **431-help at case dot edu**.

5.1 Data Visualization Books

1. R Graphics Cookbook, 2nd Edition by Winston Chang.
2. Data Visualization: A Practical Introduction by Kieran Healy.
3. Fundamentals of Data Visualization by Claus O. Wilke
4. Data Visualization with R by Rob Kabacoff
5. Interactive web-based data visualization with R, plotly, and shiny by Carson Sievert

5.2 Statistics/Data Analysis/Data Science Books

1. The Art of Data Science by Roger D. Peng and Elizabeth Matsui (book is also available with lecture videos). An earlier edition is available at bookdown.
2. Exploratory Data Analysis with R by Roger D. Peng. An earlier edition is available at bookdown

3. Data Analysis for the Life Sciences by Rafael A. Irizarry and Michael I. Love
4. Modern Statistics for Modern Biology by Susan Holmes and Wolfgang Huber
5. Regression Models for Data Science in R by Brian Caffo
6. Introduction to Data Science: Data Analysis and Prediction Algorithms with R by Rafael A. Irizarry
7. Practical Regression and ANOVA using R by Julian J. Faraway (pdf)
8. A First Course in Design and Analysis of Experiments by Gary W. Oehlert (pdf)

5.3 R and R Markdown Books

1. Cookbook for R by Winston Chang
2. Learning Statistics with R and its bookdown repository by Danielle Navarro
3. R Programming for Data Science by Roger D. Peng. An earlier edition is available at bookdown
4. R Markdown: The Definitive Guide by Yihui Xie, J. J. Allaire, and Garrett Grolemund
5. R Markdown for Scientists by Nicholas Tierney
6. R Packages by Hadley Wickham and Jenny Bryan
7. What They Forgot to Teach You About R by Jenny Bryan and Jim Hester
8. Advanced R by Hadley Wickham (2nd edition)
9. Hands-On Programming with R by Garrett Grolemund

5.4 Blogs and Internet Columns

1. Andrew Gelman and friends at Statistical Modeling, Causal Inference, and Social Science
2. Simply Statistics by Jeff Leek, Brian Caffo, Roger Peng, Rafael Irizarry and others
3. Frank Harrell's Statistical Thinking blog
4. FlowingData by Nathan Yau
5. JunkCharts by Kaiser Fung
6. New York Times What's Going On in this Graph?
7. Edward Tufte on the Web
8. Tidy Tuesdays: A weekly data project in R from the R for Data Science online learning community
9. FiveThirtyEight on Politics, Sports, Science & Health, Economics and Culture. Nate Silver is Editor-in-Chief.

5.5 Resources for Learning R

1. I recommend the Community-Sourced Data Science Guide of resources for learning data science.
    - I no longer support DataCamp in any way, and suggest that you don't, either. See R-Ladies Global on the matter.
2. RStudio Cheat Sheets are definitely worth your time. In 431, you'll especially like:

- Data Transformation with `dplyr`
- Data Visualization with `ggplot2`
- Data Import
- R Markdown

3. The `swirl` package in R can be a great help for people learning R programming and data science. Find out more about it at http://swirlstats.com/students.html
4. UCLA's Institute for Digital Research and Education has some great Data Analysis Examples using R (and other software.)

5.6 Videos about R and Data Science

1. Resources from RStudio is a great source of all kinds of useful stuff. For example:

- Getting Started with R Markdown
- Getting Your Data into R
- Data Wrangling with R and RStudio
- Six part series on RStudio Essentials (Parts 1 and 2 of each section are likely to be of greater interest in 431.)
- Data Science Essentials
- RStudio Webinars on a variety of subjects, most of which are gathered in this YouTube playlist as well.

2. Data Wrangling with R and the Tidyverse YouTube Playlist from Garrett Grolemund
3. Hadley Wickham's Whole Game
4. Tidy Tuesday Screencasts from David Robinson on YouTube
5. Hans Rosling: The Best Stats You've Ever Seen TED Talk from 2006.
6. This is Statistics: Roger Peng explains in less than two minutes why statistics is an amazing field.
7. Mona Chalabi's TED Talk on 3 ways to spot a bad statistic, 2017.
8. The beauty of data visualization from David McCandless at TEDGlobal 2010.
9. Six Types of Questions You Can Ask in a Data Analysis from Roger Peng.
10. Videos from Coursera's 4 week course "Computing for Data Analysis" in R

11. Learn R by Intensive Practice list of tutorials on YouTube.

5.7 Podcasts

1. Not So Standard Deviations by Hilary Parker and Roger Peng talking about the latest in data science and data analysis in academia and industry.
2. The Effort Report by Elizabeth Matsui and Roger Peng talking about life in the academic trenches, telling it "like it is". Every graduate student in this course looking at a career in academia would benefit from listening.
3. More or Less: Behind the Stats from Tim Harford and BBC Radio 4
4. Stats + Stories from the American Statistical Association and Miami University
5. The R-Podcast

5.8 Books on related topics

1. Broadening Your Statistical Horizons: Generalized Linear Models and Multilevel Models by Julie Legler and Paul Roback
2. Computer Age Statistical Inference: Algorithms, Evidence and Data Science by Bradley Efron and Trevor Hastie (pdf)
3. An Introduction to Statistical Learning with Applications in R by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani
4. *Statistical Rethinking* with `brms`, `ggplot2` and the tidyverse by A. Solomon Kurz, building off of Richard McElreath's *Statistical Rethinking* text and freely-available lectures.
5. Think Bayes: Bayesian Statistics Made Simple by Allen B. Downey (pdf)

Additional references and links will appear over the course of the semester in the README files associated with each class.

Chapter 6

On Software, and R

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.

- 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 also be using the R Markdown tool within RStudio. R Markdown will be taught in our class, and can be used to generate reproducible reports that appear as .html files, PDF files or Word documents, among other things.
- For some people, working with R 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 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 or Mac, but the style should be determined by your personal preferences and how you believe you will use the machine in your research life.
  - In this class, you'll be using RStudio and R, which look and work the same on either a PC or a Mac.
  - Any reasonably recent PC or Macintosh machine will work well.
  - We **do not** recommend the use of a Chromebook for this class.
  - R and RStudio also run on Linux systems. If you use one, you know more than Professor Love does about how to accomplish that.

6.2 RStudio Cloud

(borrowed in part from Andrew Heiss)

R is free, but it can sometimes be a pain to install and configure. To make life easier, you can (and should!) use the free RStudio.cloud service, which lets you run a full instance of RStudio in your web browser. This means you won't have to install anything on your computer to get started with R! We will have a shared class workspace in RStudio.cloud that will let you quickly copy templates for some of your work, too. RStudio.cloud is convenient, but it can be slow and it is not designed to be able to handle larger datasets or more complicated analysis. Over the course of the semester, you'll probably want to get around to installing R, RStudio, and other R packages on your computer and wean yourself off of RStudio.cloud. This isn't absolutely necessary (especially early on), but you'll probably feel the need to have done it by the end of September.

6.3 How Do I Install The Software?

We've built a Software page on the course website to help with installation and getting started with R.

There, you will find specific instructions to install everything you need, specifically:

- [R] The latest version of the R statistical software.
- [RStudio] The latest version of the RStudio development environment.
- [Packages] Some R "packages" of functions, data and documentation.
- [431 Data] Some data and functions specific to the 431 class.

In brief, the steps you need to take for 431 are:

1. Download and install the latest version of R (version `3.6.1` at this writing) at http://cran.case.edu/ or from https://cloud.r-project.org) which automatically chooses a fast, nearby mirror for you.

2. Download and install RStudio (version `1.2.1335` or later at this writing) at https://www.rstudio.com/products/rstudio/download/#download. If you prefer, you can run the Preview Version of RStudio to get the very latest features, but that requires you to update your setup more frequently, and, very occasionally, deal with some additional troubleshooting.
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. To install the packages for our course, follow the instructions in [the Packages description at our Software page (to be posted before class begins.)
4. Download the data and code (functions) we've developed specifically for this course from our Data and Code page (also to be posted before class begins.)

You'll find complete instructions, with a step-by-step walk through for PC or Mac machines on the Software page.

6.4 Need More 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 is worth your time, too, if you intend to use Git and GitHub.

If you're having installation problems or problems getting started in R, please consider asking a question of us at **431-help at case dot edu**, although a visit to office hours is often more helpful, as it's difficult for us to diagnose your problem without seeing your computer.

There are many, many online resources to help you with working in R, in addition to emailing us at `431-help at case dot edu`. An especially useful one is the RStudio Community (a forum specifically designed for people using RStudio and the tidyverse (i.e. you)).

6.5 Why do we teach R, instead of SPSS or SAS or whatever, in 431-432?

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, there's always Google, but 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 friendlyness 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 transfering 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 pasenger seat, and mountain climbing and spelunking gear in the back. R can take you anywhere you want to go if you take time to leard how to use the equipment, but that is going to take longer than learning where the bus stops are in SPSS.

Chapter 7

Deliverables, Expectations and Assessment

All students are expected to attend all sessions, participate vigorously in class discussions and in group work, complete all individual work in a timely fashion, demonstrate improvement of skills over the term, and perform well on the Quizzes and in the final portfolio presentation. Such a performance is the minimum standard required to receive a grade of B.

To receive an A, students are expected to complete all the requirements described above, demonstrate excellent work in both the final portfolio presentation, and outstanding work in at least one of the following: [a] in class participation, [b] assignments, [c] quizzes.

7.1 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. The courses are identical.

The course grade is based on four key areas of demonstrated accomplishment. The planned breakdown is as follows, but Professor Love may make adjustments as the semester progresses, especially once we're sure about the numbers of Homeworks and Quizzes through the term.

Weight	Task
15%	In-Class and Outside-of-Class Participation
25-35%	Completion and Quality of Homework
20-30%	Performance on Quizzes

Weight	Task
30%	Performance on Project Work

Any questions regarding how you are doing in the course should be directed to Professor Love alone. In particular, the TAs do not have full access to the final grades.

Regarding the Homework, Dr. Love will drop your lowest score before calculating your grade in the course. So you can miss one Homework over the course of the semester with no explanation. Note that this doesn't apply to the Quizzes or to any element of the Project.

7.2 Participation

I 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 in class. The TAs help me assess participation, so they are paying attention, too. Come to the TA office hours if you need help. Make an appointment to talk to us if you have something to discuss that doesn't work well in email.
2. Email `431-help at case dot edu`. with your questions and comments. That'll lead to faster answers, typically, and help us recognize you as someone trying to improve their understanding.
    - Find **typos** in the materials (code, slides, the Notes, this syllabus)? Send them to us at `431-help at case dot edu`.
    - See a cool visualization online? A nice use of statistical methods or design in a paper? Share them with us, at `431-help at case dot edu`.
3. If we ask you to do something after class (responding to a minute paper survey, for example, which will happen 8-10 times over the course of the term), getting that done in a timely fashion will help your grade in this area.
4. Visit the Teaching Assistants and ask them questions about the course, or things you're having trouble with.
5. Talk to Professor Love about your questions/comments/concerns. Make sure he knows who you are - which can be challenging with such a large group.

7.3 There are Quizzes.

Each Quiz will be taken online, exclusively. Details on the Quizzes are found at the Course website's Quizzes page.

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.

7.3.1 About the Quizzes

1. Quizzes typically involve 25-40 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 via email less than 3 hours prior to the Quiz submission deadline.
7. Quizzes that are more than ten minutes 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.

7.4 Homework assignments

There are several homeworks scheduled, and described in detail on the Homework page of our web site.

Most require straightforward demonstrations of mastery for core principles and fundamental skills. Some require deeper dives into more technically sophisticated material. Some also require reflection, particularly based on materials we'll be reading throughout the semester, especially from Nate Silver's book.

7.4.1 How do the Homeworks work?

1. Almost every Homework will require you to analyze some data, and prepare a report using R Markdown. You will submit both your Markdown file, and an HTML document built using RStudio from that Markdown file.
2. Several Homeworks will require you to write an essay. After Homework A, essays must be composed as part of your Markdown file, and thus included in your HTML document. Do not edit the result of your R Markdown conversion into Word.
3. When writing in English, use complete sentences, rather than bullet points.
4. Clearly mark each Question in each Homework. There is no need to repeat the question before answering it, although you are welcome to do so.
5. Read and heed the advice of Jeff Leek in *The Elements of Data Analytic Style*. Chapters 5, 9, 10 and 13 of that book are especially relevant to our early assignments.
6. You are welcome to discuss each Homework 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 email **431-help at case dot edu**.
7. In general, we do not guarantee to provide answers to questions that we receive in the last 18 hours before a Homework is due, especially once we've gotten into mid-September. So don't leave anything until the last day. Allow time for computer problems.
8. Failure to turn in a Homework within one hour of the deadline will result in a very poor grade on the Homework when it is (eventually) turned in, and a zero (from which it can be difficult to recover) if it is not turned in. **Submission of timely, partial work is usually better than no submission at all.**
9. **Things happen.** Dr. Love drops your worst Homeworks grade of the semester before calculating your grade. So if some disaster occurs and you have to miss a deadline, just submit a note saying that this Homework is the one you'll skip. No need for any explanation. If you need to miss more than one, though, you will need to discuss that with Dr. Love, in advance.
10. Grades on Homeworks are usually available one week after the submission deadline.
11. If you have a complaint about your grade on a Homework, please first review the Grade Appeal Policy at the end of this syllabus.

7.4.2 Where do I turn in the Homeworks?

For the most part, you'll do this using the Canvas system at https://canvas.case.edu. The course's primary listing is PQHS 431, but students in CRSP 431 and MPHP 431 should find the same information. The link to post your responses for each Homework will appear in time for you to submit the work, usually just after the deadline for the preceding Homework has passed.

7.5 Course Project

The course project is a major part of the course. Materials related to the project will be posted to the Course Website's Project page which will update frequently through the semester.

The project includes tasks you'll complete throughout the semester, culminating in a final presentation to Professor Love of your work in mid-December. For deadlines, see the Course Calendar.

Chapter 8

Some Advice for Graduate Students

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.

8.1 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, especially including me. Here are some extremely useful suggestions from Greg Wilson. In a tweet, 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."
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.

8.2 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. 10 Hints for Conferences from Creative Maths, 2016-12-07

Chapter 9

On 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.

9.1 On Campus Resource

The CWRU Writing Resource Center is definitely a good place to get some help.

9.2 Advice from Other People I Respect

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.3. [thinkchecksubmit.org] to choose the right journal for your research, or at least ones you can trust, and see this tweet for some related suggestions.
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

9.3 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 other 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. or
    - [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 Markdown 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, clear writing.

Chapter 10

General Course Policies

1. 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.
2. All student work is subject to the University's policies and procedures.
3. **Registration is required**. I do not permit anyone to audit the course who has not previously taken it, without exception. If you've taken the course before and want to sit in again, you will need to volunteer or be hired as a teaching assistant.
4. **Grading**. 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.
5. **Attendance** is expected, but it is expected that sometimes life will intervene.

- If you need to miss two or more classes in a row, inform Professor Love via email beforehand, or as soon as possible thereafter. I will assume you have a good reason - details are **not** necessary.
- You are responsible for all missed work, regardless of the reason for your absence.

6. You get to miss one Homework over the course of the semester, with no explanation, as Dr. Love will drop your lowest score across all Homeworks at the end of the semester. With that exception, **late work is unacceptable** under anything but the most harrowing of circumstances. Professor Love (via email) is the person to discuss this with, prior to the deadline, if you feel your circumstances are sufficiently dire to warrant an exception. In this class, it is far better to turn in timely, but only partially complete work than nothing at all.
7. **Feedback on assignments - deadline.** On every assignment, Quiz, project-related task, whatever, we will be delighted to respond to email

questions **up to 18 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.

8. **On Getting Help Quickly and Effectively**: In general, we don't have a way to diagnose your problem with R, RStudio or Markdown if you don't show us what you're typing that causes an error, or a lack of results. If you wrote a Markdown 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. **Using a Laptop** Using a laptop to follow along, take notes, or try things out during class, can be very helpful. Feel encouraged to do so.
10. **Computer** You will need access to a computer (PC or Mac - a ChromeBook won't do) outside of class to do every assignment. You will need to be able to install software on this computer, and update it frequently, although the use of RStudio Cloud may help alleviate this problem a bit.
11. **Distractions**. 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. **Professor Love has no shame about embarrassing people on this issue. If it's critical, just step out of the room**.
12. **Research Usage**. Any and all results of in-class and out-of-class assignments and activities are data sources for research and may be used in published research. All such use will always be anonymous.
13. **Audio-Recording.** It is our intention to provide audio recordings of each class after they are complete. Anything you say during a class session *may* be audio-recorded.
14. **Typos.** Professor Love makes occasional typographic and grammatical errors, which irritate him enormously. Please email him if you find any 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.

Again, all Case Western Reserve University policies apply to this class. To the extent that those policies are unclear, I would regard an appropriate code of conduct as mirroring this one from the R Consortium.

10.1 Grade Appeal Policy - Request a Review in December!

For each Homework, we publish a detailed answer sketch and grading rubric. You will also learn your scores on each individual item on each Homework.

Clarification of concerns related to potential typographical or other errors in these 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 Homeworks during the semester, but instead to request a review on a Google Form that will be reviewed by Professor Love in December.

- On the course's Homework page, we include a section about Grading Errors and Regrade Requests. There, you'll find a link to a Google form (you must log into Google via CWRU to see it) listing all of the Homeworks. The form is found at http://bit.ly/431-2019-regrade-requests. 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.
  - If you wish to dispute a grade, just fill out the form, and Professor Love will review it in December.
  - You can fill out the form at any time, and you can edit the form after submitting it once so that you can add additional requests up until the end of the semester.
  - All forms must be submitted by 2019-12-12 at 12 Noon.
  - 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, of course.

In mid-December, **after** Professor Love has worked out what letter grade to give each student, he will go through the requests 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 approach us in person with

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[1]

[1] I got this idea from Jessica Utts at http://www.amstat.org/publications/jse/v22n2/rossmanint.pdf.