

POLS 3316: STATISTICS FOR POLITICAL SCIENCE

Summer 2022

Instructor:	Tom Hanna	Time:	2 PM - 4 PM, Mon - Fri
Email:	tlhanna@central.uh.edu	Place:	PGH 200

Office Hours: Tuesday and Thursday 12 to 1 PM, after class, and by appointment

Welcome to Statistics for Political Science:

Today the use of data and statistics is not limited to scientists, engineers and academics. Business, government, non-profits, and the media rely on data for decision making. Unfortunately, much of the way statistics is used brings to mind Mark Twain's famous remark that "There are three kinds of lies: lies, damned lies, and statistics." As voters, we are faced with competing information from candidates often presenting the same statistics to push competing visions, proving Twain's point! There are many examples of activists on all parts of the ideological spectrum relying on their supposed expert credentials and misuse of statistics to sway the opinion of voters. At the least by the end of this course you should be able to appropriately question whether a statistical argument in the public arena is valid or deceptive. You will also gain skills useful in political science and many professions.

This course is intended to introduce the use of statistical methods and software for the analysis of quantitative data to advanced undergraduate students with little to no previous background in statistics, mathematics, or programming. You will not need to use calculus or linear algebra. Course examples will be drawn heavily from political science particularly with some examples from the life sciences and other social sciences. But the use of these methods will serve students in many careers, and you will have the opportunity to explore the techniques with data of your own choosing from other fields. This is primarily a political science course, but it is also a core math/reasoning course and appropriate to other disciplines.

Data science is a rapidly growing profession in its own right and the tools of data science are being widely used in political science, other academic disciplines, government, non-profits, and business. As part of this course, you will learn the basics of one of the two most widely used data science programming languages, the R programming language. We will cover the fundamentals in a step by step fashion to better equip you for learning more later in formal settings or on your own. You will also learn to share your results using a public Github repository and develop a simple first project, giving you a widely recognized way to present yourself as a professional user of all three tools.

Specific course objectives include:

- Understanding basic descriptive/summary statistics
- Understanding basic probability theory to prepare for learning more advanced topics
- Applying critical thinking skills to quantitative (numerical) data including the relationship between statistics and causal inference
- Learning the language of causal inference including how to properly present statistical results in the narrative
- Learning the basic concepts behind linear regression (OLS or Ordinary Least Squares)
- Learning the basics of the R statistical programming language, R Studio development environment, and version control with Github

- Developing a Github profile and your first public repositories to demonstrate your statistics and coding proficiency to future employers or graduate schools

Prerequisites:

MATH 1310, POLS 1336, and 1337 or equivalent, or consent of instructor. No higher math is required, though we may occasionally discuss some for those who are interested. The ability to clearly present and explain results in written sentences and paragraphs is expected.

Email:

I expect emails to be courteous and respectful of my time. I will be courteous and respectful of your outside time as well, so when I do email, I expect you to read and act on them as needed. When emailing me, please include the course and section number in the email subject, along with the reason for the email. I am not particular about format, but a well written, brief email is easier to respond to quickly. I am a Ph.D. candidate not a full Ph.D., so I am Professor Hanna during this course. I will respond to emails within 24 hours barring emergencies during normal business days and sooner if possible. You should bear in mind that a degree of professionalism in your emails is an important career skill you should start learning now and that professionalism is a graded component of this course. Again, as long as you are courteous and respectful of my time, I am not concerned about precise format, but here is a great reference you should consider reading before emailing any professor: <https://marktomforde.com/academic/undergraduates/Email-Etiquette.html>

Course Policies: My basic expectation is that as a student preparing to enter the workforce, school is your profession while you are here. Many of you, of course, work at other jobs as well. The point is that this is a professional environment, not recreation. Professional courtesy toward your fellow students and the instructor are a bare minimum to succeed in this class.

- Respect for other people is non-negotiable. This includes other students, faculty, staff, and guests. This also includes people not present, such as the subjects of survey or observational data we may use, and even authors whose ideas we may disagree with.
- **Academic honesty is non-negotiable. If you are uncertain whether something is not appropriate, ask me.**
- **Online collaboration may be done using the class Microsoft Teams group Study Group channel. If you set up a GroupMe or similar chat for this class, the instructor must be notified and provided access on request. If you use a GroupMe without doing this, I will assume it is for purposes of academic dishonesty.**
- Quizzes will be regular and unannounced with a short quiz during many class periods. The time of quizzes may vary from beginning to middle to end of class.
- It is impossible to earn full professionalism credit without being here, being here on time, and remaining for the full class.
- If arriving to class early or leaving late, please take a seat near the door and do so with a minimum of noise.
- If you need to take an extra break, please keep disruption to a minimum.
- Up to two quizzes may be exempted without excuse.
- Up to two absences, late arrivals, and early departures (total) may be exempted from professionalism grade without excuse.

- The grade structure includes generous points beyond what is needed to max out the grading scale. Other opportunities including correction of errors on assignments, quizzes, and tests may be offered. Take advantage of this when it is offered.
- Generous extra credit will be made available throughout the semester to the entire class. Take advantage of this when it is offered. This extra credit may only be announced in class.
- To get an A, regardless of extra credit, you must earn at least an A- on your regular work.
- **Regrading:** If you are not satisfied with any grade, you may request an official regrade after a 24-hour cooling off period and within one week following the grade. Regrade requests must be in writing, by email, and provide a reasoned argument why a better grade is justified. Any regrade will be a full re-examination of all aspects of the work and may result in an increase, decrease, or no change in the grade.
- Use of electronic devices is required for the class, but must be nondisruptive, limited to class material, and with sound turned off.

Grading Policy: Grades will be computed from the following: Professionalism and participation (10 points), Quizzes (10 points), Problem Sets (40 points), Midterm (20 points), Project (20 points), Final (20 points).

Grading Scale: Earning the following will guarantee the associated grade. Note that these are points, not percent, and that over 100 points is possible. *For technical reasons, the points in Blackboard will be 10x the points here. Simply divide your Blackboard total points by 10 to see how many points you have accumulated to that point.* Depending on class performance overall, I reserve the right to adjust the scale in the student's favor:

- A: 94.000 Points and above
- A- : 90.000 to 93.999 Points
- B+ : 87.000 to 89.999 Points
- B: 84.000 to 86.999 Points
- B - : 80.000 to 83.999 Points
- C+ : 77.000 to 79.999 Points
- C: 74.000 to 76.999 Points
- C- : 70.000 to 73.999 Points
- D+ : 67.000 to 69.999 Points
- D: 64.000 to 66.999 Points
- D- : 60.000 to 63.999 Points
- 59.999 and below: F Points

Professionalism and participation (0-10 Points): As a student, you may consider school as merely preparation for a future profession. A better approach is to think of school as your current profession. A portion of your grade will be determined by the degree of professionalism you show in this course. Specific aspects of your approach to the course that may impact this grade include punctuality,

courtesy to others, preparation for class, participation in discussions, good written communication including email, among others. A portion of each class period will be devoted to answering questions and resolving questions from the previous problem sets and readings. A portion of each lab period will be spent covering errors from the previous session. The issue you ask about today may be encountered by someone else next week, so asking appropriate questions is encouraged as a major portion of participation. Because some errors take research to resolve, your participation grade is not limited to in class participation, but includes participation in the class Microsoft Teams channel, sending questions to the instructor ahead of class time, and participation in virtual or in person office hours. That said, participation through Office hours, Teams, and email contact is a bonus that works with attendance, not a substitute for it.

Quizzes (0-10 Points): Quizzes will cover prior lecture material and topics of discussion from previous problem sets. Quizzes will be very short answers or calculations during class time, or R exercises during lab time. Expect 15-20 quizzes total. It is possible that there could be a quiz during lecture and a quiz during lab on the same day. They will typically last 5 minutes or less. Up to two quizzes may be exempted without excuse. Answers will be posted, so make-ups will not be allowed.

Problem/reading sets (0-40 Points): There will be a total of 8 problem/reading sets each worth 5 points. These assignments will include a combination of written responses, statistical calculations, and exercises in R. For statistical calculations and R programming exercises, it is expected that students will run into errors. This is a critical part of learning advanced skills. Written responses detailing the reasoning used, attempts to correct errors, and showing the work done will count for a substantial portion of the grade, up to full credit for high quality work and writing. You may work together on problem/reading sets but must do so through the official Teams channel or provide notice to the instructor of who you are working with and how. If you use a GroupMe for work in this course, the instructor must be given access or academic dishonesty will be assumed. Late work will be accepted with 50% deducted unless a valid reason for an extension is provided at least 24-hours before the assignment is due. No late assignments will be accepted after answers are posted.

Midterm exam/Final Exam (20 Points each - 40 Points total): Both exams will be allowed 90 minutes. The Final Exam will be cumulative. Students will be allowed to bring a one-page, single sided handwritten reference sheet containing any material they find useful. The exams will be a combination of short answers, statistical calculations, and interpretation of output from R. Calculations should be simple enough to finish by hand, but handheld calculators will be allowed. No other electronic devices will be allowed – **phones may not be used as calculators.** There will be a 30-minute exam review session at the end of the previous class session, with time allowed after class for those with additional questions. The Midterm is tentatively scheduled for Tuesday, July 26 at 2:00 PM. The day following the mid-term will be a half class period from 3 to 4 PM. The Final exam is tentatively scheduled for Wednesday, August 10, at 2 PM. Makeup exams will be given only for documented emergencies.

Project (20 Points:) At the beginning of the course, you will learn to setup a Github account, install R and R Studio, and integrate them with Github. As the course progresses, you will complete a short statistical analysis on a subject of your choice applying the methods from the course. You will compile the results and write a short paper explaining the results of the analysis in narrative form and including nicely formatted tables of results. This paper should be 1 to 2 pages of text with some tables and graphs. Though the final product will be due at the end of the semester, it will actually consist of compiling your lab work in your chosen data set from throughout the semester. If you keep up with the lab work and ask questions as we go about problems, completing the project should be fairly easy. If you don't keep up with the lab work, you will probably do poorly on both labs and the project. You will share this project and your **corrected** problem sets through a Github repository which you may choose to make public to showcase your work to

future employers.

Main References:

Required readings will be provided through Teams and/or Github.

Additional texts:

While all topics will be covered in lectures, lecture slides, and provided readings, two of the works I am using in preparing this course and which you may find useful are:

- Introductory Statistics: A Conceptual Approach Using R by Miller, Ferron, and Ware. ISBN-13: 9780415996006 DIGITAL ISBN-13: 9781136870101 ISBN-10: 0415996007
- Even You Can Learn Statistics & Analytics by Levine and Stephan ISBN-13: 9780133382662 DIGITAL ISBN-13: 9780133382679 ISBN-10: 133382664

Both of these books are available through the university bookstore in print and digital editions, through the Kindle store on Amazon, and available used at lower prices.

In class labs and required Software

This course will use the R statistical programming language, the R Studio development environment for R, and Github version control. All of these are available for free. The second class period will be dedicated to the basics of Github, R, and R Studio. **Those who have created a Github account, and pre-installed these programs on their own laptop configured to work with the Github account will get more out of the first lab session (July 12) including an opportunity for extra credit, including partial credit for each step completed.**

A portion of each class period will be dedicated to learning the R statistical programming language. We will cover the topic in small bite sized pieces, allowing time between classes to deal with software errors, questions, and concerns. At the end of the semester, those who invest the time will have the beginnings of their first public Github repository to demonstrate these skills (plus up to 120 points possible out of 100 required).

Errors in code are one of the most frustrating parts of learning R or any other programming language. Learning to deal with them is also one of the most important parts of the process. Your grade will not depend on error free R output, as I would rather work with you to resolve errors than have you not turn in work and learn nothing. You will be expected to be able to interpret R results on both the mid-term and final exams.

Installing R, R Studio, and Git: The main point to remember is that installing things in order will reduce the potential for errors. Following is an excellent resource for installing R and R Studio, setting up a Github account, and making them all work together. If you encounter problems, don't hesitate to contact me by email or Teams so we can devote class time to substance:

<https://happygitwithr.com/index.html> (Ch. 4, 5, 6, 7 and 12 are most important.)

Another option I highly recommend is the following Coursera course. It can easily be completed during their two week free trial and includes a certificate which can be linked to your LinkedIn profile. (I completed it in a few hours. Those with less computer experience may take longer, but may benefit even more.):

<https://www.coursera.org/learn/data-scientists-tools?specialization=jhu-data-science>

Microsoft Teams, Github, and R Studio Cloud

This class has a Microsoft Teams group and Github repository. Participation in the Teams discussion channels is encouraged and will add to your class participation grade. All required readings, lecture slides, written materials, and additional recorded video examples in R will be posted to Teams. Lab materials and R exercises for problem sets will be posted to Github for easy integration to R.

To assure that everyone will be able to follow along with minimal errors, everything will also be reproduced in an R Studio Cloud project. R Studio Cloud is a tool I think many of you will find helpful if too much frustration sets in. I have created this project using the "Instructor Pays" option, so you can access it with a free account. While you can theoretically do most of your work in R Studio Cloud, I still highly encourage you to install the software on your own device and attempt everything there first. First, you'll need to create your own R Studio Cloud project to keep your problem sets separate and this isn't really any harder than using R Studio. Second, you'll have a working copy of the examples which you can keep and refer to when you use R in the future, share through Github, etc. Third, you'll learn more, be better prepared for the exams, final project, and problem sets, and in some cases may be able to earn extra credit points. Finally, you'll actually have the experience of successfully setting everything up on your own which is valuable.

Tentative Course Outline: Important dates:

- Midterm: 2 PM, July 26
- Half day: 3-4 PM, July 27 (possibly a Teams meeting)
- Final Exam: 2 PM, August 10
- July 14: Last day to drop a course or withdraw without receiving a grade. Last day to drop a course without hours counting towards the Enrollment Cap for Texas Residents.
- August 1: Last day to drop a course or withdraw with a "W".
- August 8: Last day of class

Lectures, labs, and problem set due dates:

Generally, each class will begin with a lecture, discussion, and practice with the material followed by a lab period using R. *With the exception of Problem Set 1 and the Final Exam Review Problem Set 8, problem sets will be due on Wednesdays and Fridays so your weekends are free, but I will post assignments ahead of the weekend for those who prefer that. Problem Set 8 is due on Monday, August 8 so I can post answers and corrections before the Final Exam on the 10th.*

- July 11: Course introduction - Introductions, Course policies, brief overview of the topic
- July 12: Introductory Concepts in Probability and Statistics (No lab)
- July 13: Introduction to R and R Studio (Lab only – bring laptop with R and R Studio installed, Git installed, first lab scripts pulled from Github). Finish installation and getting started with R working directories, projects, scripts
- July 14:
 - Lecture: Descriptive and summary statistics
 - Discussion: Choosing your project data set

- Lab R: Math functions
- Lab: Compute basic descriptive stats with R math functions
- July 15:
 - Lecture: Descriptive and summary statistics
 - Discussion: Project data set questions
 - Problem Set 1 Due
 - Lab: R Scalars, vectors with a first look at matrices and arrays
 - Lab: Descriptive and summary statistics R functions
- July 18
 - Lecture: Percentiles, Quartiles, Standard Deviations
 -
 - Lab: Lab: R matrices and arrays
 - Lab: R percentiles, quartiles, standard deviations
- July 19:
 - Lecture: Probability and sets
 - Lab: R Lists
 - Lab: R Probability functions
- July 20:
 - Lecture: Frequency distributions
 - Lab: R Data frames and tibbles - loading your data
 - Lab: Frequency distributions in R
 - Problem Set 2 Due
- July 21:
 - Frequency distributions and z-scores
 - Lab R: Factors and tables - loading your data
 - Lab: R Frequency distributions with your data
- July 22:
 - Correlation
 - Lab R: Loops
 - Lab R: Correlations in R - calculating
 - Problem Set 3 Due
- July 25:
 - Hypothesis testing: Basic introduction
 - Midterm review
- July 26:
 - Problem Set 4 (Midterm Review Problems) Due

- Midterm 2 PM
- July 27:
 - Half day - recovery
 - 3 PM: Hypothesis testing in R
- July 28:
 - Statistics and causal inference
 - Lab R: Plotting in R - correlations, scatterplots, histograms
 - Lab R: Plotting with your data
- July 29:
 - Lecture: Hypothesis testing: Chi square
 - Problem Set 5 Due
 - Lab R: *if* and *ifelse*
 - Lab R: Chi square
 - Lab R:
- August 1:
 - Lecture: Hypothesis testing: Z-scores and t-tests
 - Lab R: Simple simulations
 - Lab R: Z-scores and t-tests
- August 2:
 - Lecture: Hypothesis testing: Anova
 - Lab R: Data organization and preparation
 - Lab R: Hypothesis testing with your data
- August 3:
 - Lecture: Confidence intervals
 - Problem Set 6 Due
 - Lab R: Data organization and preparation
 - Lab R: Confidence intervals
- August 4:
 - Lecture: Introduction to OLS Regression
 - Lab R: A simple regression
 - Lab R: Reading regression results
- August 5:
 - Lecture: Analyzing regression results
 - Problem Set 7 Due
 - Lab R: Two variable regression
 - Lab R: Writing regression results

- August 8:
 - Lecture: Final Exam Review
 - Problem Set 8 (Final Exam Review) Due
 - Lab R: Regression with your data
 - Lab R: Writing your results
- August 9: Optional Work and Review Period Online!
 - Lecture: Final Exam Review 2 PM
 - Lab R: Review and project work period 3 PM
- August 10. 2 PM:
 - 30 minute project work session
 - 2:30 PM Final exam
- August 11. Project Due.

Academic Integrity

Cheating and plagiarism will not be tolerated and will be penalized. Each student in this course is expected to abide by the University of Houston's policies against cheating and plagiarism. The University's statement on academic honesty is available here: [urlhttps://www.uh.edu/provost/policies-resources/honesty/](https://www.uh.edu/provost/policies-resources/honesty/)

You are encouraged to work, study, and discuss information and concepts covered in class and readings with other students. Cooperation between students should never result in one student possessing a copy of work done by someone else, either in electronic form or hard copy. **Cooperation in studying for tests ends when the test begins.** Should copying occur, both the student who copied the work and the student who provided the copied material will automatically receive no points for that test. Repeated and/or egregious violations will be dealt with more severely, and may include failing the course and facing further disciplinary action.

Honor Code Statement

Students may be asked to sign an honor code statement as part of their submission of any graded work including but not limited to projects, quizzes, and exams: "I understand and agree to abide by the provisions in the University of Houston Academic Honesty Policy. I understand that academic honesty is taken very seriously and, in the cases of violations, penalties may include suspension or expulsion from the University of Houston."

Accommodations for Students with Disabilities

The Americans with Disabilities Act is a federal law providing comprehensive protections, including reasonable academic accommodations, for persons with disabilities. The University of Houston is committed to providing reasonable accommodations for students who have learning disabilities, psychiatric disabilities, and health impairments. Any student in this course who has a disability that may prevent him or her from fully demonstrating his or her abilities should inform me personally. If you believe you have a disability which is eligible and requires accommodation, please contact the Center for Students with Disabilities (CSD). For detailed information, including documentation requirements, listings of available academic support services, and test administration policies, please visit <http://www.uh.edu/csd/>

Statement on Counseling and Psychological Services Counseling and Psychological Services (CAPS) can help students who are having difficulties managing stress, adjusting to college, or feeling sad and hopeless. To reach CAPS (www.uh.edu/caps), call 713-743-5454 during and after business hours for routine appointments or if you or someone you know is in crisis. No appointment is necessary for the

“Let’s Talk” program, a drop-in consultation service at convenient locations and hours around campus. http://www.uh.edu/caps/outreach/lets_talk.html

Required Daily Health Assessment

Your presence in class each session means that you have completed a daily self-assessment of your health/exposure and you:

- Are NOT exhibiting any Coronavirus Symptoms
- Have NOT tested positive for COVID-19
- Have NOT knowingly been exposed to someone with COVID-19 or suspected/presumed COVID-19

If you are experiencing any COVID-19 symptoms that are not clearly related to a pre-existing medical condition, do not come to class. Please see COVID-19 Diagnosis/Symptoms Protocols for what to do if you experience symptoms and Potential Exposure to Coronavirus for what to do if you have potentially been exposed to COVID-19. Consult the (select: Undergraduate Excused Absence Policy Absence Policy) for information regarding excused absences due to medical reasons.

Excused Absence Policy Regular class attendance, participation, and engagement in coursework are important contributors to student success. Absences may be excused as provided in the University of Houston Undergraduate Excused Absence Policy (<https://uh.edu/provost/policies-resources/student/excused-absence-policy/>) for reasons including: medical illness of student or close relative, death of a close family member, legal or government proceeding that a student is obligated to attend, recognized professional and educational activities where the student is presenting, and University-sponsored activity or athletic competition. Additional policies address absences related to military service, religious holy days, pregnancy and related conditions, and disability.

Recording of Class

Students may not record all or part of class, livestream all or part of class, or make/distribute screen captures, without advance consent of the instructor. If you have or think you may have a disability such that you need to record class-related activities, please contact the Center for Students with DisABILITIES.

If you have an accommodation to record class-related activities, those recordings may not be shared with any other student, whether in this course or not, with any other person, or on any other platform. Classes may be recorded by the instructor. Students may use instructor’s recordings for their own studying and notetaking. Instructor’s recordings are not authorized to be shared with anyone without the prior approval of the instructor. Failure to comply with requirements regarding recordings will result in a disciplinary referral to the Dean of Students Office and may result in disciplinary action.

Syllabus Changes

Due to the changing nature of the COVID-19 pandemic, please note that the instructor may need to make modifications to the course syllabus and may do so at any time. Notice of such changes will be announced as quickly as possible on Blackboard.

UH Email

Email communications related to this course will be sent to your Exchange email account which each University of Houston student receives. The Exchange mail server can be accessed via Outlook, which provides a single location for organizing and managing day-to-day information, from email and calendars to contacts and task lists. Exchange email accounts can be accessed by logging into Office 365 with your CougarNet credentials or through Access UH. They can also be configured on IOS and Android mobile devices. Additional assistance can be found at the Get Help page.

Webcams

Access to a webcam may be required for students participating remotely in this course. Webcams must be turned on for office hours.

Delivery Format and Final Exams

Lectures being offered in the Face-to-Face format with a safe number of students in a socially distanced classroom. Lectures and participation occur in-person and in real time. Final exams will be conducted in face-to-face format on the day assigned by the University schedule.

Helpful Information COVID-19 Updates: <https://uh.edu/covid-19/> Coogs Care: <https://www.uh.edu/dsaes/coogs-care/>
Laptop Checkout Requests: <https://www.uh.edu/infotech/about/planning/off-campus/index.php#do-you-need-a-laptop> Health FAQs: <https://uh.edu/covid-19/faq/health-wellness-prevention-faqs/> Student Health Center: <https://uh.edu/class/english/lcc/current-students/student-health-center/index.php>

Course Resources

- Teams Link: https://teams.microsoft.com/l/team/19%3aoDhb2KU9AuTFmPqmsdjT12J2rU0cpj_Q-mV0Gng-S40thread.tacv2/conversations?groupId=67ef7322-91d1-4c63-847f-2b13cd6e889f&tenantId=170bbabd
- Github Repository: <https://github.com/tomhanna-uh/pols3316-summer2022>
- R Studio Project Link: https://rstudio.cloud/spaces/260761/join?access_code=mv0pyFqMtTixD7jrgLYFXH

Useful Resources

- **Getting started!**
 - <https://happygitwithr.com/index.html> (Ch. 4, 5, 6, 7 and 12 are most important.)
- Official R and R Studio
 - The R Project homepage - <https://www.r-project.org/>
 - R Studio - <https://www.rstudio.com/>
 - CRAN: The Comprehensive R Archive Network - <https://cran.r-project.org/>
- Cheatsheets
 - base R cheatsheet <https://github.com/rstudio/cheatsheets/raw/main/base-r.pdf>
 - advanced R cheatsheet <https://www.rstudio.com/wp-content/uploads/2016/02/advancedR.pdf>
 - R Studio Cheatsheets (directory of cheatsheets) <https://www.rstudio.com/resources/cheatsheets/>
- Bloggers, developers, advanced users
 - R bloggers - <https://www.r-bloggers.com/>
 - Jenny Bryan - R Studio Developer with lots of useful stuff <https://jennybryan.org/>
 - Hadley Wickham - Rice Statistics Professor, R developer and author <https://hadley.nz/>
- R related subreddits:
 - <https://www.reddit.com/r/rstats/>
 - <https://www.reddit.com/r/RStudio/>
 - <https://www.reddit.com/r/rprogramming/>
 - <https://www.reddit.com/r/Rlanguage/>
- Free books, courses, etc.
 - R For Data Science - complete book in a website <https://r4ds.had.co.nz/>
 - ggplot2: elegant graphics for data analysis - <https://ggplot2-book.org/>
 - An Introduction to Statistical Learning with Applications in R - complete PDF of 612 page book available for download. The best introduction to machine learning available: <https://www.statlearning.com/>
 - Advanced R - <https://adv-r.hadley.nz/>
 - Hewlett Packard Enterprise Data Science Institute at UH - <https://hpedsi.uh.edu/education/training>