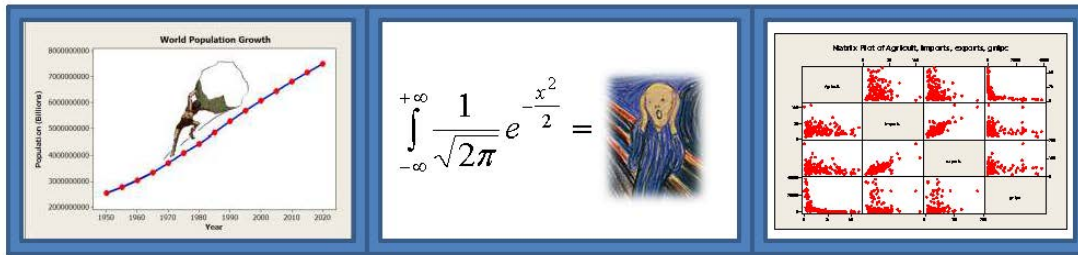


# Course Syllabus

[Jump to Today](#)



## S&DS 230e: Data Exploration and Analysis

Summer 2024

Instructor: [Jonathan Reuning-Scherer](http://environment.yale.edu/profile/ruening-scherer) (<http://environment.yale.edu/profile/ruening-scherer>)

## Course Syllabus

[The Goals](#)

[The Decision: Is this course for you?](#)

[The Software and Materials](#)

[The Lectures](#)

[The Modules](#)

[The Requirements](#)

[The Project](#)

[The Grades](#)

[The Ways to Get Help](#)

[The Schedule](#)

### The Goals

- Explore a variety of examples of real world data problems
- Learn the R programming language
- Learn how to scrape data off websites, deal with missing data
- Extend knowledge of multiple regression, non-parametric tests, and logistic regression
- Learn about new statistical graphics
- **Learn how to write about data analysis results**

### The Decision: Is this course for you?

This course is for you if:

- You've had at least one intro stat class (such as S&DS 10x, S&DS 100, AP Stats, ENV 728, etc)
- You want to learn R
- You don't want to see too many proofs and you don't need calculus with your statistics
- You want hands-on, applied experience with data!

This course is probably not for you if:

- You've never had an intro stat class
- You like proofs and calculus
- You don't intend on keeping up - we move **quickly!**

## The Software and Materials

You'll need a computer that can run R (pretty much every computer can run R).

There **are no required textbooks** for this class. However, you may wish to check out the following :

R is available **for FREE by downloading online**. [Click here](#). [\(http://cran.r-project.org/\)](http://cran.r-project.org/). Available for Windows, Macintosh, Linux. I suggest installing R 4.3.1. A great program, but pretty challenging to learn. Some online intros are available : <http://www.r-project.org/> and search for Contributed Documentation. See resources folder on this site for some R intros.



**RStudio** is a [free](#) 'front end' way to use R. Some people prefer to use just straight up R, but I'll be using this all semester long. Certainly a nice way to organize R code, R graphics, R data, and R Markdown.



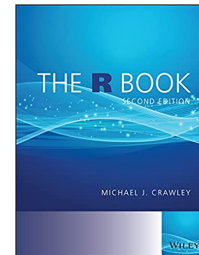
**DataCamp**. Free (and paid) online interactive intro to R. A great supplemental way to get started in R.



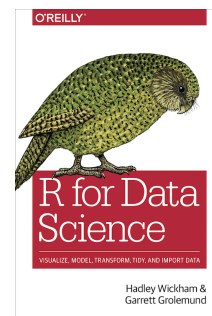
**Code School**. Free, quick, very basic intro to R



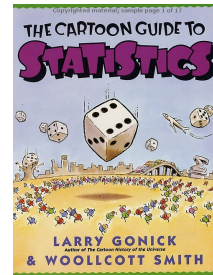
**Crawley : The R Book (Second Edition - 2012)**. Lots of positive reviews, good companion to learn about most of the topics we'll cover. **FREE as an electronic resource on ORBIS**. (if off campus, just connect with Yale VPN). Also available on **AMAZON**.



**Wickham and Golemund : R for Data Science**. Lots of positive reviews, however strikes me as a good 'second book'. Seems to me to start in the middle of an idea. **FREE as an electronic resource on ORBIS**. (if off campus, just connect with Yale VPN). Also available on **AMAZON**.



**Gonick and Smith : *The Cartoon Guide to Statistics*.** A humorous, non-threatening, and almost entirely accurate introduction to probability and statistics.



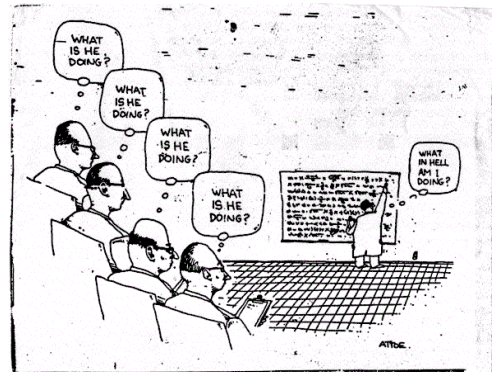
### Additional R Help

[R Reference Card](ftp://cran.r-project.org/pub/R/doc/contrib/Short-refcard.pdf)  [Reference Card](https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf)  [R Markdown](#)

### The Lectures

 [\(http://www.stat.yale.edu/Courses/2017-18/\)](http://www.stat.yale.edu/Courses/2017-18/)

All PDF notes and relevant R Markdown / R code is provided for each of the 24 modules.



### The Modules

This class consists of 24 modules. Each module includes video recordings (about one hour per class), R Markdown files, PDF notes, etc. Click on [MODULES \(https://yale.instructure.com/courses/87932/modules\)](https://yale.instructure.com/courses/87932/modules) at left.

This class is offered entirely asynchronously - there is no time you have to be present.

### The Requirements

**Homework** : Assigned and due about every 2.5 days (9 assignments total). Homework is VERY important (for you, not me) - you just have to get your hands dirty to understand statistics. Working together is encouraged. **Copying is not.** You must turn in your own solutions to each problem set. I'm generally a pretty easy going guy : however, **homework is due on the date specified. Homework to be uploaded to Gradescope via CANVAS as a PDF file. You can either knit directly to PDF and then upload, OR you can to knit to Word, open the file in Word (not in preview as on a Mac), save as a PDF, upload the PDF.**

As a part of submission, you need to indicate the pages in your PDF file on which particular questions can be found. A penalty of 5% is assessed if you fail to do this for homeworks 2 through 10.

Here is a [VIDEO TUTORIAL](#)   [\(http://Here%20is%20a%20video%20TUTORIAL%20on%20how%20to%20upload%20to%20Gradescope%20and%20assign%20pages.\)](http://Here%20is%20a%20video%20TUTORIAL%20on%20how%20to%20upload%20to%20Gradescope%20and%20assign%20pages.) on how to upload to Gradescope and assign pages.

A penalty of 10% is assessed on homework up to 24 hours late, 20% if up to 48 hours late. No homework accepted after 24 hours without a really good excuse . . .




Regrade request may be submitted within 3 days of grades being released by submitting a request to GradeScope. Be aware that your entire assignment will be regraded and your score may go down.

Homework grades will be released to you approximately 3 days after the due date, and solutions are available on CANVAS approximately 2 days after the due date.

**Exams :** Two exams: each is an untimed take home, open note exam. See assignments for due dates. Also submitted to Gradescope. Late exams are not accepted without a dean's excuse.

**Project :** See below

## The Project

- You can choose a data set that you are working on in another class or any data set you find interesting. [Kaggle](https://www.kaggle.com/datasets)  (<https://www.kaggle.com/datasets>) is a great option : for social science data, see [ICPSR](https://www.icpsr.umich.edu/web/pages/)  (<https://www.icpsr.umich.edu/web/pages/>) (must be accessed through Yale portal).
- You can also collect your own survey - I can make it a requirement for class folks to fill it out - see [www.surveymonkey.com](http://www.surveymonkey.com/)  (<http://www.surveymonkey.com/>)
- Work in groups of 2-3 people if you like, but not a requirement
- Projects are due by Sunday, August 6, 2023, midnight, uploaded to CANVAS
- Complete instructions can be found under Files --> Final Project Information (or just click [HERE](https://yale.instructure.com/courses/87932/files/8029435/download?wrap=1) (<https://yale.instructure.com/courses/87932/files/8029435/download?wrap=1>)).

**As you work on this project, I expect you will regularly pester myself and the TAs.**

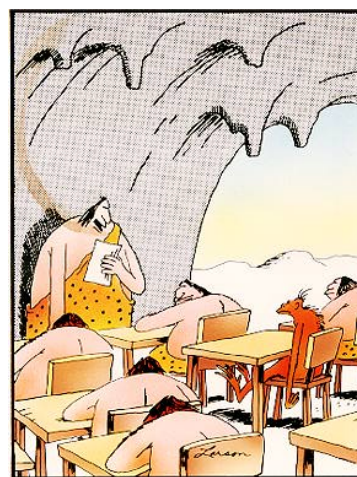
## The Grades

Your grade will be calculated as follows :

Homework	50%
Exams	30%
Project	20%

**All Grades are posted online so you can check our record keeping - click on GRADES at left.**


**The Bottom Line : Do the work, learn statistics, forget the grade.**



"Well, I've got your final grades ready, although I'm afraid not everyone here will be moving up."

## The Ways to Get Help – Office Hours, Help Sessions, Ed Discussion

**OFFICE HOURS** - There are over 25 hours per week of office hours between JDRS and TA's. Use Them! Click [HERE](https://yale.instructure.com/courses/87932/pages/office-hours) (<https://yale.instructure.com/courses/87932/pages/office-hours>) (<https://yale.instructure.com/courses/65145/pages/office-hours>) for weekly times and Zoom Links. Click [HERE](https://yale.instructure.com/courses/87932/pages/instructor-and-tas) (<https://yale.instructure.com/courses/87932/pages/instructor-and-tas>) to learn about TA's and ULA's.

**Ed Discussion-** a great online chat site- post questions, TA's and Instructors will answer them, and student can answer each others questions. Click on [Ed Discussion](#)   
(<https://edstem.org/us/courses/6360/discussion/>) at left.

**The Experience of past students:**

About 5% of students find my humor corny and my voice and personality 'grating'.

**The Schedule: See Home Page**