Stat 201: Stat Analysis Via Computers

Spring 2016 – University of Vermont

Instructor:	Samuel V. Scarpino	**Time:**	T/R 13:15 – 14:30
Email:	svscarpi@uvm.edu	**Location:**	Votey 205
Office:	Farrell Hall 200B	**Office Hours:**	W 9:30 – 12:00

Course Pages: http://scarpino.github.io/teaching/ and Blackboard.

Teaching assistant: Our teaching assistant is Lucy Greenberg. Her email address is lgreenb1@uvm.edu. Lucy will hold office hours on Tuesdays from 3-4pm on the top floor of Pearl House.

Main References:

- Nolan, D. and Lang, D.T., 2015. *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving.* CRC Press[1].

- Cody, R., 2007. *Learning SAS By Example: A Programmer's Guide.* SAS Institute[2].

Objectives: This course offers an introduction to computational methods for analyzing, visualizing, and performing statistical investigations of data. It is designed to teach you a variety of methods and give you a chance to use them to study real-world data. There will be no traditional in-class exams; instead, the major assignments–an open-book take-home midterm and a small group data challenge–will ask you to demonstrate your mastery of the methods we will cover in class. Learning to program and work with data is hard, but I know you're up to the challenge and am sure you'll be glad to have these skills.

Assignments & Grading Breakdown:
Take-home midterm (30%) – The take-home midterm must be done individually, but it is open-note, open-book, and open-internet. It will challenge you to apply the skills learned during the first half of the semester. If you plan on applying for a job as a data scientist, you'll probably have to do something very similar to this assignment during your interview.

Data challenge (30%) – The data challenge–an event where someone releases a data set and awards prizes to those who find something interesting in the data and/or produce a striking visualization–lets us experience an application of data science methods that has become very popular. This will be a group project, but will have an individual written component worth 1/2 of your grade and a group presentation portion worth 1/2 of your grade. Two of my collaborators from Silicon Valley have agreed to attend the presentations in order to give you feedback on how your results might be received at technology companies.

Homework (30%) – There will be 5 homework assignments, which are designed to evaluate your mastery of the skills we learn in class. You will have the option to replace one homework grade with your grade on the midterm exam. This does not include missed homeworks. All homeworks will be due by 5pm on Thursday of the "due by" week.

Participation (10%) – Your participation grade is based on your attendance and respect for the technology policy. I will not take attendance, but will assign non-graded in-class assignments. Each missed in-class assignment will reduce your participation grade by 0.5 percentage points. Everyone will receive one free missed assignment. For example, after you miss two in-class assignments your maximum participation grade will be 9%, meaning you cannot score higher than a 99% in the class.

---

[1]Available as an eBook through the UVM library – http://proquest.safaribooksonline.com/9781482234817
[2]Available as an eBook through the UVM library – http://proquest.safaribooksonline.com/book/databases/sas/9781599941653

Prerequisites: An introductory-level understanding of probability and statistics is expected (STAT 111 with Instructor permission, or STAT 141. Co-requisite: STAT 211.). Previous experience using R for statistics and/or data analysis is beneficial, but not required.

Important Dates:

Take-home midterm assigned .... March 17th
No class, work on midterm .... March 22nd & 24th
Take-home midterm due .... March 24th by 17:00
Data challenge begins .... April 14th
No class, work on data challenge .... April 19th
Data challenge written due .... May 3rd by 17:00
Data challenge presentations ...... Scheduled final time

Course Schedule:

Date(s)	Material	To Do
Jan. 19th & 21st	Introduction to data science & R	Install R and Rstudio[3]
Jan. 26th & 28th	Working with data & creating reports	Attend class
Feb. 2nd & 4th	Plotting and custom functions	HW1 due[4]
Feb. 9th & 11th	Modeling race times	Read DSIR Chpt. 2
Feb. 16th & 18th	Airline data analysis	HW2 due & Read DSIR Chpt. 5
Feb. 23rd & 25th	Simulations in R	Read DSIR Chpt. 7
Mar. 1st & 3rd	Zika virus	HW3 due & Read DSIR Chpt. 6
Mar. 15th & 17th	Trading stocks	HW4 due & midterm assigned
Mar. 22nd & 24th	Work on midterm	Midterm due by 5pm on the 24th
Mar. 29th & 31st	Midterm review and SAS intro	Read LSBE Chpts. 1 & 2
Apr. 5th & 7th	Data analysis in SAS	Read LSBE Chpts. 3 & 4
Apr. 12th & 14th	Adv. SAS and data challenge	HW5 due & Read LSBE Chpt. 25
Apr. 19th & 21st	Data challenge & how to give talks	Work on data challenge
Apr. 26th & 28th	Skills for the data challenge	Work on data challenge
May 3rd.	Exploring data science jobs	Data challenge report due
Scheduled final period	Data challenge presentations	Present and attend

---

[3]R: https://cran.r-project.org/ RStudio: https://www.rstudio.com/
[4]All HWs are due by 5pm on Thursday and should be turned in on Blackboard.

Academic assistance: Anyone needing accommodation, e.g., per the ACCESS program, please contact me as soon as possible.

Religious holidays: As per University policy, you have the right to practice the religion of your choice and can make-up missed work due to your religious holidays. For those requesting an accommodation due to a religious holiday, please submit a schedule of your holidays to me by the end of the second full week of classes. You must submit documentation verifying your participation in these holidays.

Course policies:

I. Grades – 100–98% (A+), 97–93% (A), 92–90% (A-), 89–87% (B+), 86–83% (B), 82–80% (B-), 79–77% (C+), 76–73% (C), 72–70% (C-), 69–60% (D), <60% (F).

II. Technology – Please silence and put away all electronics before coming to class–there should be zero texting in class. Computers should be used only for course-related work and only when someone isn't addressing the class. Violation of these policies will negatively affect your participation grade (and your understanding of course material). I will provide a short email break each class.
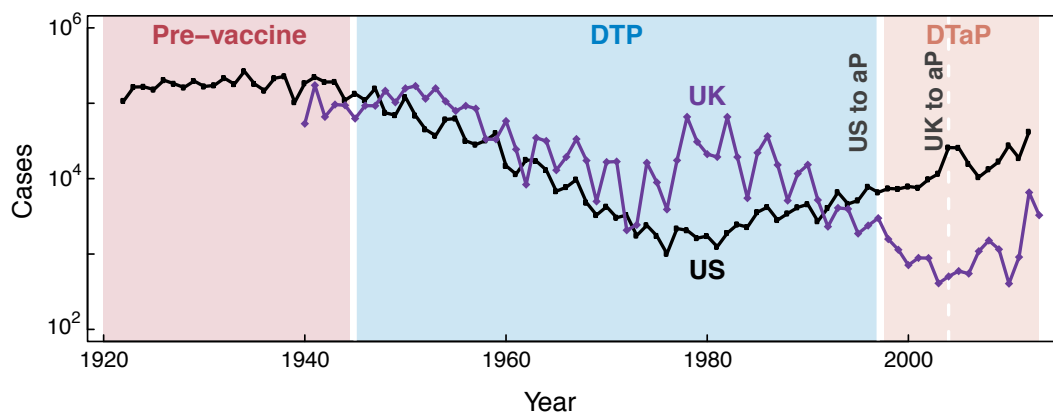
III. Turning in assignments – All assignments must be turned in on Blackboard.

IV. Late assignments – Late or missed assignments will be given a score of 0%. Please contact me if you have a documented emergency.

V. Email – I am happy to answer questions via email, but cannot promise to respond same-day. Please remember that email is a professional, mostly-permanent record, so please communicate in a respectful manner.

VI. Academic honesty – As in all UVM classes, academic honesty will be expected and departures will be dealt with appropriately. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation, see http://www.uvm.edu/cses/ for guidelines.

Whooping cough in the U.S.A. and U.K.

Recreated from Althouse, B.M. & Scarpino, S.V. (2015) Asymptomatic transmission and the resurgence of Bordetella pertussis. *BMC medicine* 13(1): 146.