

# Stat 201: STAT ANALYSIS VIA COMPUTERS

Spring 2017 – 14936 – University of Vermont

---

<b>Instructor:</b>	Samuel V. Scarpino	<b>Time:</b>	T/R 10:05 – 11:20
<b>Email:</b>	<a href="mailto:svscarpi@uvm.edu">svscarpi@uvm.edu</a>	<b>Location:</b>	Votey 205
<b>Office:</b>	Farrell Hall 206	<b>Office Hours:</b>	W 9:30 – 11 & R 12 – 13:30

---

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

**Teaching assistant:** Our teaching assistant is Deven Gokhale. His email address is [deven-vishwas.gokhale@uvm.edu](mailto:deven-vishwas.gokhale@uvm.edu). Deven will hold office hours on Fridays from 3-4pm on the second floor of Farrell Hall.

## Main References:

- Kabacoff, 2015. *R in Action: Data analysis and graphics with R, 2nd Ed.*. Manning Press<sup>1</sup>.
- Delwiche and Slaughter, 2003. *The Little SAS Book*. SAS Institute<sup>2</sup>.

**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:

*R Quiz (10%)* – After the first two weeks of class, on Tuesday, Jan. 31st, there will be an in-class quiz to evaluate how much your basic skills in the R programming language have advanced. The quiz will be done individually, but will be open-note, open-book, and open-Internet. Nevertheless, it will be challenging, so please prepare by completing the first homework assignment and practicing with the swirl package.

*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.

*Data challenge (20%)* – 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 and will have a written component worth 1/2 of your grade and a presentation worth 1/2 of your grade.

*Homeworks (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 midterm exam grade. This does not include missed homeworks, i.e. homeworks you did not turn in or were turned in late. All homeworks will be due by 5pm on Thursday of the “due by” week.

*Participation (10%)* – Your participation grade is based on your in-class work, 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 1 percentage point. 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, and after three missed in-class assignments your max. participation grade will be 8%.

---

<sup>1</sup> Available as an eBook through the UVM library – <http://proquest.safaribooksonline.com/book/programming/r/9781617291388>

<sup>2</sup> Available as an eBook through the UVM library – <http://proquest.safaribooksonline.com/book/databases/sas/9781590473337>

**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:

In-class R Quiz .....	Jan. 31st
Take-home midterm assigned .....	March 9th
Take-home midterm due .....	March 23rd by 17:00
Data challenge begins .....	April 13th
Data challenge written due .....	May 4th by 17:00
Data challenge presentations ..	May 9th 8:00 – 10:00am

### Course Schedule:

Date(s)	Material	To Do
Jan. 17th & 19st	Introduction to comp. statistics & R	Install R and Rstudio and swirl 1–4 <sup>3</sup> .
Jan. 24th & 26th	Working with data	swirl 5 – 7 & 12. <b>HW1 due Thurs.</b> <sup>4</sup>
Jan. 31st & Feb. 2nd	Plotting and custom functions	swirl 8, 9, 15. <b>R quiz on Tuesday</b>
Feb. 7th & 9th	tapply, lapply, etc.	swirl 10 & 11
Feb. 14th & 16th	Simulations I	swirl 13. <b>HW2 due Thurs.</b>
Feb. 21st & 23rd	Simulations II	swirl Getting and Cleaning Data
Feb. 28th & Mar. 1st	Infectious diseases	Attend class
Mar. 7th & 9th	Infectious diseases cont.	<b>HW3 due Tues. &amp; midterm assigned</b>
Mar. 14th & 16th	Spring Break	—
Mar. 21st & 23rd	The Stock Market	<b>Take home midterm Thurs.</b>
Mar. 28th & 30th	The Stock Market cont.	Attend class
Apr. 4th & 6th	Stories	<b>HW4 due Mon. Apr. 10th at 8am</b>
Apr. 11th & 13th	SAS and data challenge	Attend class
Apr. 18th & 20th	SAS cont.	<b>HW5 due Thurs.</b>
Apr. 25th & 27th	SAS cont. & Work on data challenge	Attend class
May 2nd & 4th	<b>Data challenge final presentations</b>	<b>Data challenge report due Thurs.</b>
May 9th	<b>NO CLASS</b>	<b>NO CLASS</b>

<sup>3</sup>R: <https://cran.r-project.org/> RStudio: <https://www.rstudio.com/>

<sup>4</sup>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/csese/> for guidelines.

---