Syllabus: STAT 297

General Information

Location and Time

- **Location:** Huck Life Science Bldg 10
- **Time:** Tuesdays and Thursdays 12:05PM - 1:20PM

Course Websites

- **Course website:** http://stat297.smac-group.com
- **Online textbook:** http://r.smac-group.com
- **Discussion Forum:** http://piazza.com/psu/fall2017/stat297/home

Course Staff

Instructor

- **Name:** Stéphane Guerrier
- **Email:** stephane@psu.edu
- **Office:** Thomas Bldg 310
- **Office Hours:** Tuesdays 3PM to 5PM (in Thomas Bldg 310)

Teaching Assistant

- **Name:** Justin Lee
- **Email:** justinlee@psu.edu
- **Office Hours:** Tuesdays 3PM to 5PM (in Thomas Bldg 310)

Course Description

This course is intended to provide an introduction to statistical programming using the R language. It will also provide students with notions of data management, manipulation and analysis as well as of reproducible research, result-sharing and version control (using GitHub). At the end of the class, student should be able to construct the own R package, make it available on GitHub and document it using literate programming.

Tentative list of topics that will be discussed in this class are listed below:

- **Reproducible Research:** `knitr` and `rmarkdown`;
- **Version control:** `Github`;
- **Introduction to programming:** Data structures, logical operators, control structures and functions;
- **Visualizations:** Exploratory data analysis with Base R and `ggplot2`;
- **R packages:** Construction of R-packages using `devtools`, `roxygen2`, `pkgdown`;
- **Webscrapping:** Automatic extraction of data from websites using `rvest` and `quantmod`;
- **High performance computing:** C++ (using `Rcpp`) and parallel computing.

Laptops, Textbooks & Software

Laptops

Bring a laptop to class each day if you have one. I encourage students to work collaboratively, so I'd like to have at least one laptop for every 2 or 3 students. Please your laptop each week if you have one, but there's no need to buy a laptop if you don't already have one.

Textbooks

This class will based on the textbook:

- **Required:** An Introduction to Statistical Programming Methods with R

It is also is available here: http://r.smac-group.com. This document is **under development** and it is therefore preferable to always access the text online to be sure you are using the most up-to-date version. Due to its current development, you may encounter errors ranging from broken code to typos or poorly explained topics. If you do, please let us know! Simply add an issue to the GitHub repository used for this document (which can be accessed here https://github.com/SMAC-Group/ds/issues) and we will make the changes as soon as possible. In addition, if you know RMarkdown and are familiar with GitHub, make a pull request and fix an issue yourself, otherwise, if you're not familiar with these tools, they will be explained later on in the book itself.

The textbooks below are also recommended and are legally available online for free. The following texts will be heavily referenced:

- **Recommended:** Advanced R Programming Hadley Wickham
- **Recommended:** R Packages Hadley Wickham
- **Recommended:** An Introduction to R W. N. Venables, D. M. Smith, and the R Core Team

The following textbooks are *helpful*, but not necessary to succeed in the course:

- **Supplemental:** ggplot2: Elegant Graphics for Data Analysis [2nd Edition - GitHub Only] Hadley Wickham
- **Supplemental:** R for Data Science Garrett Grolemund and Hadley Wickham
- **Supplemental:** The R Inferno Patrick Burns

Software

The course will use and present the **R** statistical computing language as well as different parts of C++ through Rcpp. The integrated developer environment that we will use to explore *R* is RStudio IDE made by RStudio Inc.

Communication

Piazza

We will be using Piazza for class-related discussion and questions, to help you benefit from each other's questions and the collective knowledge of your classmates and professor. Questions can be posted to the entire class (for content-related questions). I encourage you to ask questions if you are struggling to understand a concept, and to answer your classmates' questions when you can.

Do Not use Piazza for issues related to your grade or other private matters (not even an instructor post); email those questions or comments to the instructor directly or discuss them in person.

Email

Most issues about classroom activities can be posted to Piazza, but you should use email (or a conversation in person) for all personal or private matters.

Grading

Learning outcomes will be assessed based on performance in each of the following categories accompanied by their impact on the overall grade:

Type	Percentage
Semester Project	55
Homework	20
Attendance	10
Participation	15
Course Total	*100*

Final letter grades will be determined as follows:

Letter grade	Percentage
A	93-100%
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
C+	77-79.99%
C	70-76.99%
D	60-69.99%
F	$< 60\%$

Semester Project

The final project provides an opportunity to combine content learned throughout the course for use in some realistic application. All projects will be conducted in groups. The details of the semester will be discussed in class and students will have the opportunity to choose between several project formats.

Homeworks

There will be **four** to **six** homework assignments this semester which will realized in groups. A tentative schedule will be communicated in the coming weeks.

Participation

Participation is graded based on Piazza participation and in-class participation. In order to earn full credit for the Piazza portion, each student should make 1 or more substantive posts per week (except during the first week) related to the content of the course; *at least one post each week should be a reply to another student's post.* Grading will utilize Piazza meta-data that can be accessed only by an instructor.

Attendance

Due to the structure of this class, attendance is very important. Moreover, students will be encouraged to work in teams on in-class activities each week, so others are counting on you to be in class and contribute. Students with University excused absences (e.g. athletics) should notify the instructor as soon as possible and provide a minimum of one week notice.

Each lecture will have its own magic word that you are responsible for entering into a form link that will be provided on piazza. In order to be eligible to receive attendance points, you must be *in class*, enter the magic word within 5 minutes after it is given, and stay for entire duration of class. Periodically, checks will be performed to verify the attendance. Students who have entered the magic word but are not present, will fail class and will have an academic integrity violation opened against them. Please note, every student is eligible to miss *three* class periods without needing to provide a reason.

The exact method that we will use to calculate attendance at the end of the semester is given by:

```
my_attendence = function(N, N_max){
    round( min( c( 1, N/(N_max - 3) ) )*100, 2)
}
```

where `N` denotes the number of times a student has attended class and `N_max` denotes the maximum number of classes where attendance was taken.

Late Work

Late work is generally accepted for two days after the original due date (& time) for 75% credit unless otherwise stated; assignments and projects submitted more than two days late (or not at all) are not accepted, and will be given a zero.

Policies & Resources

Counseling and Psychological Services (CAPS)

Many students at Penn State face personal challenges or have psychological needs that may interfere with interfere with their academic progress, social development, or emotional well being. The university offers a variety of confidential services to help you through difficult times, including individual and group counseling, crisis intervention, consultations, online chats, and mental health screenings. These services are provided by staff who welcome all students and embrace a philosophy respectful of clients' cultural and religious backgrounds, and sensitive to differences in race, ability, gender identity and sexual orientation.

Counseling and Psychological Services at University Park (CAPS):

- Phone: 814-863-0395
- Web: http://studentaffairs.psu.edu/counseling/

Penn State Crisis Line (24 hours/7 days/week): 877-229-6400

Crisis Text Line (24 hours/7 days/week): Text LIONS to 741741

ECoS Code of Mutual Respect

The Eberly College of Science Code of Mutual Respect and Cooperation embodies the values that we hope our faculty, staff, and students possess and will endorse to make the Eberly College of Science a place where every individual feels respected and valued, as well as challenged and rewarded.

Academic Integrity Statement

Academic dishonesty is not limited to simply cheating on an exam or assignment. The following is quoted directly from the "PSU Faculty Senate Policies for Students" regarding academic integrity and academic dishonesty: *Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this institution. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating of information or citations, facilitating acts of academic dishonesty by others, having unauthorized possession of examinations, submitting work of another person or work previously used without informing the instructor, or tampering with the academic work of other students.*

All University and Eberly College of Science policies regarding academic integrity/academic dishonesty apply to this course and the students enrolled in this course. Refer to the following URL for further details on the academic integrity policies of the Eberly College of Science: http://www.science.psu.edu/academic/Integrity/index.html. Each student in this course is expected to work entirely on her/his own while taking any exam, to complete assignments on her/his own effort without the assistance of others unless directed otherwise by the instructor, and to abide by University and Eberly College of Science policies about academic integrity and academic dishonesty. Academic dishonesty can result in assignment of "F" by the course instructors or "XF" by Judicial Affairs as the final grade for the student.

Disability Policy

Penn State welcomes students with disabilities into the University's educational programs. If you have a disability-related need for reasonable academic adjustments in this course, contact Student Disability Resources (SDR; formerly ODS) at 814-863-1807, 116 Boucke, http://equity.psu.edu/student-disability-resources. In order to receive consideration for course accommodations, you must contact ODS and provide documentation (see the guidelines at http://equity.psu.edu/student-disability-resources/guidelines).

Syllabus Changes

This syllabus is subject to change as circumstances warrant; all changes will be distributed in writing (usually electronically or on Piazza) or announced in class.