Statistics 585X - Data Technologies for Statistical Analysis

Course Description: Not all data lives in nice, clean spreadsheets, not all data fits in a computer's main memory. As statisticians we cannot always rely on other people and sciences to get the data into formats that we can deal with: we will discuss aspects of statistical computing as they are relevant for data analysis. Read and work with data in different formats: flat files, databases, web technologies. Elements of literate programming help us with making our workflow transparent and analyses reproducible. We will discuss communication of results in form of R packages and interactive web applications.

Meeting times:	T/R 3:40-5:00
Location:	Snedecor 3121
Instructor:	Dr Heike Hofmann
Contact:	`hofmann@iastate.edu`

Optional Textbooks:

Wickham, H. *Advanced R Programming*, `https://github.com/hadley/adv-r`,

Murrell, P. *Introduction to Data Technologies*, `http://www.stat.auckland.ac.nz/~paul/ItDT/` ,

Xie, Y. *Dynamic Documents with R and knitr*, `https://github.com/RevolutionAnalytics/RHadoop/wiki`,

and a selection of other reading as needed.

Course Objectives:

- Able to read and combine data from flat files, SQL database, binary netCDF, and making use of web technologies as data source.
- Clean the data, check the quality, impute missing values.
- Write efficient code, reproducible code so others are able to replicate the analysis.
- Develop software, individually and collaboratively, debug, profile and package code.
- Experiment with event driven programming to build an interactive graphic, and a GUI.
- Provide experience in pulling data together to solve a contemporary problem.

Tentative Outline:

Week 1	Collaborative working environment: team work, version control systems, RStudio, RMarkdown, git, github
Week 2	Reading (non-)standard file formats: csv, xls, fwf, binary vs ascii, html, netCDF, hdf5, json, wav, images, spatial
Week 3-4	Numeric and graphic summaries: (review) of tidyverse, split-apply-combine, ggplot2
Week 5	Data structures in R: date, times, space, factors, lists, S3 classes and tree structures; review of functional programming
Week 6	Rearranging data: key-value formats, normal forms, purrr, broom
Week 7	Elements of reproducible research: knitr, markdown, file naming conventions, data, code storage, html
Week 8	Web-scraping, working with text, awk & grep
Week 9	Databases, setup, access, working with massive data, SQL, dplyr &rsqlite
Week 10	Structure of code, profiling, testing, debugging, devtools
Week 11	Package system in R, documentation & R CMD check
Weeks 12-13	Event driven programming, GUIs, interactive graphics, shiny, ggvis & plotly
Week 14	Critical analysis of code, comparison of results and time from different sources
Week 15	Project presentations

Computing: Software primarily used will be R (`www.R-project.org`) with the RStudio IDE.

Course Grades: Lab work, reading assignments & blogging; final project & presentation.

Component	**Weight**
Reading assignments & Blog posts	15%
Lab work	40%
Final project (report & presentation)	45% (35% + 10%)

Academic Misconduct All acts of dishonesty in any work constitute academic misconduct. Online courses are not exception. The Student Disciplinary Regulations (`http://policy.iastate.edu/policy/SDR`) will be followed in the event of academic misconduct. Depending on the act, a student could receive an F grade on the test/assignment, F grade for the course, and could be suspended or expelled from the University. Academic misconduct includes all acts of dishonesty in any academically related matter and any knowing attempt to help another student commit an act of academic dishonesty. Academic dishonesty includes, but is not limited to each of the following acts when performed in any type of academic or academically related matter, exercise, or activity:

Plagiarism: Plagiarism is the act of representing directly or indirectly another persons work as your own. It can involve presenting someones speech, wholly or partially, as your; quoting without acknowledging the true source of the quoted material; copying and handing in another persons work with your name on it; and similar infractions. Even indirect quotations, paraphrasing, etc., can be considered plagiarism unless sources are properly cited. Plagiarism will not be tolerated, and students could receive an F grade on the test/assignment or an F grade for the course. The Iowa State University policy for academic misconduct can be found in the Student Disciplinary Regulations.

Obtaining Unauthorized Information: Information is obtained dishonesty, for example, by copying graded homework assignments from another student, by working with another student on a take-??home test or homework when not specifically permitted to do so by the instructor, or by looking at your notes or other written work during an examination when not specifically permitted to do so.

Tendering of Information: Students may not give or sell their work to another person who plans to submit it as his or her own work. This includes giving their work to another student to be copied, giving someone answers to exam question during the exam, taking an exam and discussing its contents with students who will be taking the same exam, or giving or selling a term paper to another student.

Misrepresentation: Students misrepresent their work by handing in the work of someone else. The following are examples: purchasing a paper from a term paper service; reproducing another persons paper (even with modifications) and submitting it as their own; having another student do their computer program or having someone else take their exam.

Bribery: Offering money or any item or service to a faculty member or any other person to gain academic advantage for yourself or another is dishonest.

Special Accommodations Please address any special needs or special accommodations with the instructor the first day of class or as soon as you become aware of your needs. Those seeking accommodations based on disabilities should obtain a Student Academic Accommodation Request (SAAR) from the Disability Resources Office.

Student Disability Resources
1076 Student Services Building
Ames, IA 50011–2222
Phone: 515 294-7220

Fax: 515 294-2397
TTY: 515 294-6635
E–mail: disabilityresources@iastate.edu
Website: http://www.dso.iastate.edu/dr/