

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 Dishonesty

The class will follow Iowa State University's policy on academic dishonesty. Anyone suspected of academic dishonesty will be reported to the Dean of Students Office. <http://www.dso.iastate.edu/ja/academic/misconduct.html>.

Disability Accommodation

Iowa State University complies with the Americans with Disabilities Act and Sect 504 of the Rehabilitation Act. If you have a disability and anticipate needing accommodations in this course, please contact (instructor name) to set up a meeting within the first two weeks of the semester or as soon as you become aware of your need. Before meeting with (instructor name), you will need to obtain a SAAR form with recommendations for accommodations from the Disability Resources Office, located in Room 1076 on the main floor of the Student Services Building. Their telephone number is 515-294-7220 or email disabilityresources@iastate.edu. Retroactive requests for accommodations will not be honored.

Dead Week

This class follows the Iowa State University Dead Week guidelines as outlined in <http://catalog.iastate.edu/academiclife/#deadweek>.

Harassment and Discrimination

Iowa State University strives to maintain our campus as a place of work and study for faculty, staff, and students that is free of all forms of prohibited discrimination and harassment based upon race, ethnicity, sex (including sexual assault), pregnancy, color, religion, national origin, physical or mental disability, age, marital status, sexual orientation, gender identity, genetic information, or status as a U.S. veteran. Any student who has concerns about such behavior should contact his/her instructor, Student Assistance at 515-294-1020 or email dsosas@iastate.edu, or the Office of Equal Opportunity and Compliance at 515-294-7612.

Religious Accommodation

If an academic or work requirement conflicts with your religious practices and/or observances, you may request reasonable accommodations. Your request must be in writing, and your instructor or supervisor will review the request. You or your instructor may also seek assistance from the Dean of Students Office or the Office of Equal Opportunity and Compliance.

Contact Information

If you are experiencing, or have experienced, a problem with any of the above issues, email academicissues@iastate.edu.