Welcome

Prof. Eric A. Suess

August 23, 2019
STAT 694 - Applied Research in Statistics and Biostatistics

Units: 1-4

A collaborative research experience on a research topic(s) designated by the instructor in Statistics and Data Science. Students conduct a literature search (search the web), develop a research proposal, complete a research project leading to a professional presentation, paper, or report.

Prerequisites: Admission to the Statistics graduate program.

Repeatability: May be repeated for credit for a maximum of 8 units.

Possible Instructional Methods: Entirely On-ground.

Grading: A-F or CR/NC (student choice).
What is this course about?

Many things!
Innovation
Motivation
Leadership and Collaboration
Data Product Development
Communication
During this class you will be asked to innovate.

From the BusinessDictionary innovation

The process of translating an idea or invention into a good or service that creates value or for which customers will pay. . . . In business, innovation often results when ideas are applied by the company in order to further satisfy the needs and expectations of the customers.
Innovation

- Develop a new idea to work with data!
Motivation

- Develop a plan to learn about the new idea and how to implement it!
Leadership and Collaboration

- “Take the lead!”
- Use communication tools
  - slack
  - github
  - RStudio Cloud, Rstudio Community
  - rstudio::conf 2020
  - LinkedIn
Data Product Development

- Use data science tools
  - R Notebook, Jupyter Notebook
- Make a data product!
  - R Project
  - R Shiny
  - R package
Communication

- Share the data product!
  - report, blog
I would like to you think about what you are interested in and what sort of project(s) you might work on in a job you hope to get some day.

Come up with an idea for a project that you could discuss in an interview process.

The project is a way to prepare for data assignments that might be given as part of an interview process.
Data Science Workflow

- Idea
- Access data, probably through an API
- Transform the data, hopefully using a database or using SQL or related dbplyr or dtplyr
- Load the data into R or Python
- Model the data using ML
- Build an App from the data using Shiny or create an R package
- “Sell the product to me, your Mentor, and others in the class.”
Lots of interesting apps

- CityMapper
- Excerise apps
- Fires
- Earthquakes
- Parking
Some ideas

- How are share bikes used?
- Personal data from an app
- Music lyrics
- Stock price visualizer
- Ebola data
More ideas

- Help Nathan Lau, Flowing Data update his book Visualize This, Notebooks, modern R API packages, dynamic graphics, Visualize This Tidyverse Edition
- Help Joe Blitzstein develop some R Notebooks for the examples in his book and/or get the R prob package working, introduce him to dbplyr
Thank you, former students

- David Plotz, Data Engineer, machineVantage
- Enrique Emanuel Rodriguez, Data Scientist, FlowWest, CDECRetrieve
- Daniel Park, Data Scientist, United Business Bank
- Tony Lai, Data Scientist, LinkedIn
- Erik Lee, Consumer Lending Risk Management - Credit Scoring and Analytics, Wells Fargo
- Navdeep Gill, Data Scientist, Data Engineer, h2O
- Gabriela de Queiroz, Data & AI Developer Advocate, IBM, R Ladies
- Joe Rickert, RStudio, BARUG
- Richard Hom, Director at BlackRock
- Kaniz Rachid Lubana Mamun
Mentors

- Hadley
- Jenny
- JJ Allaire
- Joe Blitzstein Stat_v1
- Robert Shumway and David Stoffer
- Bruce Trumbo
- Phil Spector

... and many many others in the field.