Last Updated April 16, 2020

Data Science for the Social World

Instructor 1
Name: Dr. Mike Findley
Email: mikefindley@utexas.edu
Office: BAT 3.108
Office Hours: Day time
Website: michael-findley.com

Instructor 2
Name: Mike Denly
Email: mdenly@utexas.edu
Office: Mezes 3.232
Office Hours: Day time
Website: mikedenly.com

Course Information
Abbreviation: GOV 355M
Unique Number: XXXX
Time: Day time
Room: Online
Website: canvas.utexas.edu

Lab Information
Innovations for Peace and Development
Website: www.ipdutexas.org
Lab Location: BEL (Stadium) 214
Lab Access Code: Ask the instructor

DRAFT

1. Course Description

This course provides students with a comprehensive introduction to data science for the political and economic world. The skills that students acquire in the course will help them prepare for jobs by teaching practical data skills coupled with strong social science reasoning. Organized around a set of substantive themes and practical tasks, each class topic is motivated by real-world problems and then backed with data science skills to solve those problems. Emphasis is placed on developing proficiency in cleaning, manipulating, wrangling, scraping, visualizing, and mapping data. Most work is conducted in the software programs R and Excel, and to a lesser extent through introductory exercises in other programs including Python. In the process, students learn about good principles of working with data, including through version control with Github. The class takes place through asynchronous instruction, online coding practice problems, exams, and online instructor consultations.

2. Course Requirements

2.1. Prerequisite Coursework

There are no prerequisites to enroll in this course. We will work at a speed so that everyone should be able to adequately learn the materials if they do not miss class or required assignments. However, students who have previous knowledge of basic statistics and/or computer programming may find the course easier.

2.2. Required Software and Resources

This course makes use of the following programs:

- `Excel`. Students who do not already have Excel on their computers may obtain the full Microsoft Office Suite, including Excel, for $19.99 at through the Campus Computer Store. Note that `Google Sheets` cannot be used as a substitute for `Excel`.
- `Google Sheets`. Students can use Google Sheets for free through their utexas email accounts.
- R. It is a free, open-source statistics and data science program. To install R, see here.
- R `Studio` is a companion program for R For instructions on how to freely download R `Studio`, consult here.
- `Git`. It is the program underlying most version control—i.e., a system of tracking file tracking that facilitates collaboration. For instructions on how to freely install Git, consult here.
- `GitHub`. It is an online platform that facilitates version control through `Git`. To use GitHub, students need to create a free online account at www.github.com.
- `Git Bash`. Only students who use Microsoft Windows for their operating system will need to download and install `Git Bash`. Mac users can directly use the Shell (command line) and do not need to install `Git Bash`.
- `Python`. Note that the version of `Python` is very important. [We will decide upon specifics later.]
- `Jupyter Notebook`. We will be using `Jupyter Notebook` to interface with `Python`. For more on how to install `Jupyter Notebook`, see here.

Prior knowledge of any of these software programs/platforms is not required. We will teach you the basics of all of these programs during the course. To obtain help with these programs and others, there are two resources that we will utilize:

- `Data Camp`. It is an online platform that provides hundreds of courses to learn new skills. The courses are interactive and fun. We will be using some of these courses from Data Camp as required homework.
- `Lynda`. You can also access free courses through UT-Austin's Lynda Portal.

2.3. Required Textbooks

Healy, Kieran. 2019. "Data Visualization: A Practical Introduction." Princeton: Princeton University Press. [Draft freely-available here]

Wickham, Hadley and Grolemund, Garrett. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastol, California: O'Reilly Media. [Freely-available here]

DRAFT

2.4. Grading Rubric

- Attendance: 5%
- Homework: 35%
- Exams: 20 % each (3 total)

2.5. Grading Scale

- 92.50-100 (A)
- 92.49-89.50 (A-)
- 86.50-89.49 (B+)
- 82.50-86.49 (B)
- 79.50-82.49 (B-)
- 76.50-79.49 (C+)
- 72.50-76.49 (C)
- 69.50-72.49 (C-)
- 66.50-69.49 (D+)
- 62.50-66.49 (D)
- 59.50-62.49 (D-)
- 59.49 or below (F)

2.6. Grade Rounding

The above grading scale already incorporates very generous grade rounding, not to mention the multitude of extra credit opportunities. Accordingly, there will be no additional rounding of grades under any circumstance.

2.7. Grade Posting on Canvas

We will post all grades to the class website, Canvas. We will also use the option where students may discern the average score of the class. This way, students will know where they stand by the end of semester.

2.8. Grade Appeals

If students would like to appeal your grade on any assignment, it is necessary make the request to both professors in writing, over email, within 5 days of receiving your grade. In the grade appeal, students must specify the reason(s) why we may have misgraded the assignment. Acceptable reasons include those pertaining to the concepts and material covered during the course. We will not consider requests for grade changes that are not germane to the course.

2.9. Late Work

Unless you receive prior approval from either professor, we will discount all late work as follows:

- 1-15 minutes: 0% (grace period for last-minute issues)
- 15 minutes-24 hours late: -10%
- 24-48 hours late: -25%

DRAFT

- more than 2 days late: -50%
- more than one week: no credit offered

2.10. Students Rights and Responsibilities

- You have a right to a learning environment that supports mental and physical wellness.
- You have a right to respect.
- You have a right to be assessed and graded fairly.
- You have a right to freedom of opinion and expression.
- You have a right to privacy and confidentiality.
- You have a right to meaningful and equal participation, to self-organize groups to improve your learning environment.
- You have a right to learn in an environment that is welcoming to all people. No student shall be isolated, excluded or diminished in any way.

With these rights come these responsibilities:

- You are responsible for taking care of yourself, managing your time, and communicating with the instructor if things start to feel out of control or overwhelming.
- You are responsible for acting in a way that is worthy of respect and always respectful of others.

2.11. Personal Pronoun and Name Preferences

Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student's legal name. We will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that we may make appropriate changes to my records.

2.12. Academic Integrity

Each student in the course is expected to abide by the University of Texas Honor Code: "As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity." Plagiarism is taken very seriously at UT. Therefore, if you use words or ideas that are not your own (or that you have used in previous class), you must cite your sources. Otherwise you will be guilty of plagiarism and subject to academic disciplinary action, including failure of the course. You are responsible for understanding UT's Academic Honesty and the University Honor Code, which can be found at the following web address: http://deanofstudents.utexas.edu/sjs/acint_student.php

DRAFT

2.13. Drop Policy

If you want to drop a class after the 12th class day, you'll need to execute a Q drop before the Q-drop deadline, which typically occurs near the middle of the semester. Under Texas law, you are only allowed six Q drops while you are in college at any public Texas institution. For more information, see: http://www.utexas.edu/ugs/csacc/academic/adddrop/qdrop

2.14. University Resources for Students

Your success in this class is important to me. We will all need accommodations because we all learn differently. If there are aspects of this course that prevent you from learning or exclude you, please let me know as soon as possible. Together we'll develop strategies to meet both your needs and the requirements of the course. There are also a range of resources on campus:

2.14.1. Services for Students with Disabilities

This class respects and welcomes students of all backgrounds, identities, and abilities. If there are circumstances that make our learning environment and activities difficult, if you have medical information that you need to share with me, or if you need specific arrangements in case the building needs to be evacuated, please let me know. I am committed to creating an effective learning environment for all students, but I can only do so if you discuss your needs with me as early as possible. I promise to maintain the confidentiality of these discussions. If appropriate, also contact Services for Students with Disabilities, 512-471-6259 (voice) or 1-866-329- 3986 (video phone). http://ddce.utexas.edu/disability/about/

2.14.2. Counseling and Mental Health Center

Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support: http://www.cmhc.utexas.edu/individualcounseling.html

2.14.3. The Sanger Learning Center

Did you know that more than one-third of UT undergraduate students use the Sanger Learning Center each year to improve their academic performance? All students are welcome to take advantage of Sanger Center's classes and workshops, private learning specialist appointments, peer academic coaching, and tutoring for more than 70 courses in 15 differ-

DRAFT

ent subject areas. For more information, please visit http://www.utexas.edu/ugs/slc or call 512-471-3614 (JES A332).

Undergraduate Writing Center: http://uwc.utexas.edu/
Libraries: http://www.lib.utexas.edu/
ITS: http://www.utexas.edu/its/
Student Emergency Services: http://deanofstudents.utexas.edu/emergency/

2.14.4. Important Safety Information

If you have concerns about the safety or behavior of fellow students, TAs or Professors, call BCAL (the Behavior Concerns Advice Line): 512-232-5050. Your call can be anonymous. If something doesn't feel right – it probably isn't. Trust your instincts and share your concerns.

The following recommendations regarding emergency evacuation from the Office of Campus Safety and Security (512-471-5767, http://www.utexas.edu/safety/):

- Occupants of buildings on The University of Texas at Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside.
- Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building.
- Students requiring assistance in evacuation shall inform their instructor in writing during the first week of class.
- In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, The University of Texas at Austin Police Department, or Fire Prevention Services office.
- Link to information regarding emergency evacuation routes and emergency procedures can be found at: www.utexas.edu/emergency

DRAFT

3. Class Schedule, Readings, and Homework

3.1. Part 1: Basics

Module 1: Intro to Data and Excel Basics

Class:

- Course overview

- Learning the basics with Excel
  - Saving and file types (e.g., `.xlsx` vs. `.csv`)
  - Inspecting and filtering data
  - Merging cells, wraping text, and freezing panes
  - Sorting data
  - Pivot tables
  - Missing data
  - Making graphs and troubleshooting

Required Reading:

- Healy, Kieran. 2019. "Data Visualization: A Practical Introduction." Princeton: Princeton University Press.
  - Read: Preface and Chapter 1.

Required Assignment:

- All chapters from Data Camp: Spreadsheet Basics (2 hours)
  - Please complete this assignment by XXX

Module 2: Intermediate Excel

Class:

- Preparing files for analysis
- Identifying and creating unique identifiers
- Relative and absolute cell referencing
- Basic formulas (IF, SUM, AVERAGE)
- VLOOKUP

Required Assignments:

- Chapter 1 from Data Camp: Pivot Tables (1 hour)
  - Please complete this assignment by XXX
- Chapter 2 from Data Camp: Data Analysis with Spreadsheets (1.5 hours)
  - Please complete this assignment by XXX

DRAFT

Module 3: Intro to R (Basics)

Class:

- Setting the working directory
- Installing packages and loading libraries
- Objects and vectors
- Loading existing data frames
- Creating new data frames manually
- Inspecting the data (`head`, `View`, `dim`, `summary`)
- Classes (mumeric, character/strings, factors)
- Generating new variables
- Dealing with missing values
- Getting help
- Descriptive statistics (mean, median, mode, quantiles)
- Tables with `stargazer`
- Cross tabulations (cross tabs)
- Correlations
- t-test
- Saving data

Required Reading and Watching:

- Watch this Getting Started with R and `R Studio` video
  - This video help you get R and R Studio set up on your computer
- Watch this Introduction to R video.
  - This video will also help you with the setup but goes a bit deeper as well.

Assignment:

- All chapters from Data Camp: Intro to R (4 hours)
  - Please complete this assignment by XXX

DRAFT

Module 4: Data Visualization with `ggplot2` in R

Class:

- An introduction to `ggplot2`
- Histograms, bar plots, scatter plots, line graphs, box plots,
- Displaying graphs (`facets` and `ggarrange`)

Required Reading:

- Healy, Kieran. 2019. "Data Visualization: A Practical Introduction." Princeton: Princeton University Press.
  - Read: Chapters 3-5.

Required Assignment:

- HW: Chapter 1 from Data Camp: Data Visualization with ggplot (Part 1) (1 hour)
  - Please submit your assignment by XXX

Module 5: Exam for Part 1 of Course

Required Assignment:

- Study for the Exam
  - Review practice questions

3.2. Part 2: Replicability, Programming Basics, and Data Wrangling

Module 6: R Markdown and Version Control (Git/Github)

Class:

- R Markdown
  - Setting up `.rmd` files
  - Inserting code chunks with different features
  - Create new sections and text with different features (see cheatsheet)
- Version control (Git/Github)

DRAFT

  - Using the command line/shell (Mac users) or GitBash (Windows users)
  - Creating repositories/projects in R
  - Committing, adding, and status checking
  - Linking R with GitHub

Required Reading and Installations:

- Wickham, Hadley and Grolemund, Garrett. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastol, California: O'Reilly Media.
  - Read: Chapter 27 (R Markdown)
- Bryan, Jenny. 2019. "Happy Git with R."
  - Read: Chs. 1, 4-9, 12-14 and make all required installations for your computer
- Wickham, Hadley. 2019. "R Packages: Git and Github."
  - Read: all instructions and make all required installations for your computer

Required Assignments:

- Chapter 3 (Introduction to R Markdown) and Chapter 4 (Customizing your R Markdown Report) from Data Camp: Communicating with Data in the Tidyverse (1.5-2 hours)
  - Please submit your assignment by XXx
- All chapters from Data Camp: Intro to Git for Data Science (3 hours)
  - Please submit your assignment by XXx

Additional Resources:

- Git and Github for Poets Tutorial Series
- R Markdown Cheat Sheets
  - Overall Cheat Sheet
  - Reference Guide

Module 7: Intermediate R (Programming Basics)

Class:

- Conditionals and control flow
- Loops
- Functions

DRAFT

- Utilities and regular expressions (e.g., `grep`, `gsub`)
- Dates and times
- Apply commands (e.g., `lapply`, `sapply`)

Required Reading:

- Wickham, Hadley and Grolemund, Garrett. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastol, California: O'Reilly Media.
  - Read: Chapters 16, 19-21

Required Assignment:

- All Chapters from Data Camp: Intermediate R (4 hours)
  - Please submit your assignment by XXX

Module 8: Data Cleaning with Cross-Sectional Data in R (Part 1)

Class:

- Subsetting (i.e. creating new data frames)
- Creating new variables and indexing
- Conditional statements (`ifelse`)
- Merging data
- Converting characters/string variables to numeric variables
- Removing accents
- Changing file encodings
- Working with factor variables
- Recoding data
- Filtering data
- Sorting data
- Taking logs
- Labeling variables

Required Reading:

- Wickham, Hadley and Grolemund, Garrett. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastol, California: O'Reilly Media.
  - Read: Chapter 12, 13

DRAFT

Required Assignment:

- Chapters 1 and 2 from Data Camp: Cleaning Data in R (2 hours)
  - Please submit your assignment by XXX

Module 9: Data Cleaning with Panel Data in R (Part 2)

Class:

- Reshaping
- Appending
- Finding and removing duplicates
- Collapsing/summarizing
- Piping
- Importing World Bank World Development Indicators data directly from R
- Creating lag and lead variables
- Balancing panel data

Required Reading:

- Wickham, Hadley and Grolemund, Garrett. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. Sebastol, California: O'Reilly Media.
  - Read: Chapter 18

Required Assignment:

- Chapters 3 and 4 from Data Camp: Cleaning Data in R (2 hours)
  - Please submit your assignment by XXX

Module 10: Exam for Part 2 of the Course

Required Assignment:

- Study for the Exam
  - Review Practice Exercises

DRAFT

3.3. Part 3: Mapping, Scraping, and `Python`

Module 11: Mapping (Part 1)

Class:

- Coordinate systems
- Map projections
- Vector data (points, lines, shapes)
- Raster data
- Shapefiles
- Mapping conventions
- Working with the `sf` package
- Using `rnaturalearth` and `rnaturalearthdata` to acquire static maps
- Plotting point data (latitudes/longitudes) with `sf`
- Plotting point data with `ggplot2`
- Interactive mapping with `mapview`
- Spatial joins with `sf`
- Spatial sums and means (within polygons)
- Obtaining polygon centroids

Required Reading:

- Pebesma, Edzer. 2020. 1. Simple Features for R
- Pebesma, Edzer. 2020. 5. Plotting Simple Features

Required Assignment:

- All chapters from Data Camp: Working with Geospatial Data in R (4 hours)
  - Please submit the assignment by XXX

Additional Resources

- Pebesma, Edzer. 2019. `sf` plot reference manual

Module 12: Mapping (Part 2)

Class:

DRAFT

- Using `tidycensus` to work with US Census data
- Making a `shiny` app

Required Reading:

- Healy, Kieran. 2019. "Data Visualization: A Practical Introduction." Princeton: Princeton University Press.
  - Read: Chapter 7
- Walker, Kyle. 2020. `tidycensus`
  - Read: All sections in Basic Usage of `tidycensus` and Spatial Data with `tidycensus`

Required Assignment:

- All chapters from Analyzing US Census Data in R (4 hours)
  - Please submit the assignment by XXX

Module 13: Web Scraping

- All chapters from Data Camp: Working with Web Data in R (4 hours)

Module 14: Intro to Python

- TBD

Module 15: Final Exam

Required Assignment:

- Study for the Exam
  - Review Practice Exercise

DRAFT