DS 3003: Communicating with Data
University of Virginia / School of Data Science
Spring 2022

Lecture:	MW 3:30 - 4:45 PM	Credits:	3 (Carnegie Definition)
Room:	Mechanical Engr Bldg 341	Instructional Mode:	face-to-face
Instructor:	Youmi Suk	Course URL:	collab.its.virginia.edu
Office:	Elson, Room 167		
e-mail:	eub6uw@virginia.edu		
Office Hours:	W 10:30 - 11:30 AM and by appointment, Zoom (https://virginia.zoom.us/j/3462312823)		

TA: Soumya Chappidi, sc4cb@virginia.edu
TA Office Hours: F 9:00 – 10:00 AM and by appointment, Zoom
        (https://virginia.zoom.us/j/98690598687?pwd=b08veE1TbHVkR2pVQTVtSmFqNmY3dz09)

Course Description: This course provides practical experiences about how data is commonly used in communication. The general objective is to make you familiar with how to effectively summarize and visualize data in order to share your story. All the examples and analyses will be done mainly in R, with a particular focus on graphical presentations with the R visualization tools like ggplot2, Plotly, and Shiny.

Learning Outcomes: By the end of the course, I hope that you will
- be excited about summarizing and visualizing data using R.
- be able to construct meaningful messages from data and effectively deliver them with visualization tools.
- create a portfolio showcasing your communication and visualization skills.
- have confidence in interacting and collaborating with colleagues.

Textbooks
There is no textbook for this course, as the world is our textbook; however, as such, students will both supply and be supplied with content over the course from sources like

- ☐ [Our World in Data] (https://ourworldindata.org),
- ☐ [The New York Times] (https://www.nytimes.com),
- ☐ [FiveThirtyEight.com] (https://fivethirtyeight.com),
- ☐ [Medium.com] (http://medium.com),
- ☐ [The Atlantic] (https://www.theatlantic.com),
- ☐ [TIME] (https://time.com)
- ☐ [National Geographic](https://www.nationalgeographic.com), etc.

Lecture notes will be available online on Collab.

Software
Throughout the course, we use the free software package R (http://www.r-project.org/) and an R editor, RStudio (https://www.rstudio.com/). R is a very popular and highly ubiquitous programming language designed for statistical data analysis and visualization. Instructions, resources, and materials to help students use R -- such as those available from

- ☐ [R for Data Science](https://r4ds.had.co.nz)
- ☐ [R Cookbook](http://www.cookbook-r.com)
- ☐ [An Introduction to R](https://cran.r-project.org/doc/manuals/R-intro.pdf)
- ☐ [ggplot2 book](https://ggplot2-book.org/index.html)
- ☐ [Plotly](https://plotly.com/r/)
- ☐ [Shiny](https://shiny.rstudio.com)

and *plain old [google](https://www.google.com/?client=safari) web searches* -- will be made available at appropriate times to support student efforts at each stage of the course. R and all of these additional resources are available for free.

Data

For several assignments and your final project, you will create your own story with data. These days it's actually fairly easy to get your hands on interesting data. Here is a list of websites where you can find interesting data.

- ☐ [kaggle](https://www.kaggle.com/)
- ☐ [AWS Open Data](https://registry.opendata.aws/)
- ☐ [data.world](https://data.world/search)
- ☐ [ICPSR](https://www.icpsr.umich.edu/web/pages/)
- ☐ [The Google Dataset Search](https://datasetsearch.research.google.com/)
- ☐ [The UCIML Repo](http://archive.ics.uci.edu/ml/)
- ☐ [The CMU data repository](http://lib.stat.cmu.edu/datasets/)
- ☐ [The datasets subreddit](https://www.reddit.com/r/datasets)
- ☐ [Tycho](https://www.tycho.pitt.edu)
- ☐ [Data Portals](http://dataportals.org/)

Class Management

Email / Communication
- Ask questions on Piazza.
- Email is the best way to get in touch with the instructor.
- Please be sure to include "DS 3003" in your email subject line when sending email to any of the course staff: professor and TA.

Course Schedule: Topics, Assignments (A) & Presentations (P)

Week 1 Jan 19	Intro to the Course, Syllabus; Intro to R A1 (due Jan 23) – Learning articles (2 pts) & Communications articles (2 pts) A2 (due Jan 26) – Getting started with R (4 pts)
Week 2 Jan 24 & 26	Intro to R (cont'd); Learning & Communication P1 (Jan 24) – Share learning articles (2) P2 (Jan 26) – Share communication articles (2) A3 (due Jan 30) – Figures/data about learning or communication articles (4)
Week 3 Jan 31 & Feb 2	Visualizations with Base R; ggplot2 A4 (due Feb 6) – Rmd file with plots (4) & Group review (1)
Week 4 Feb 7 & 9	ggplot2; Scatter Plots with Smoother Lines P3 (Feb 7) – Share plots (2) A5 (due Feb 13) – Group bar ggplot (4) & Group review (1) A6 (due Feb 15) – Midterm proposal (2)
Week 5 Feb 14 & 16	Non-Parametric & Parametric Smoother Lines; data format P4 (Feb 14) – Share plots (2) A7 (due Feb 20) – Group loess ggplot (4) & Group review (1)
Week 6 Feb 21 & 23	Midterms P5 (Feb 21 & 23) – Midterm presentations (16) A8 (due Feb 27) – Create your own plots using data from presentations (4*2)
Week 7 Feb 28 & Mar 2	Midterms P5 (Feb 28 & Mar 2) – Midterm presentations A9 (due Mar 6) – Create your own plots (4*2)
Week 8 Mar 7 & 9	Spring recess
Week 9 Mar 21 & 23	Plotly, Interactive Presentations & Plots with Statistical Uncertainty A10 (due Mar 20) – Group time series & confidence intervals (4) & Group review (1)
Week 10 Mar 21 & 23	Shiny P6 (Mar 21) – Share plots (2) A11 (due Mar 27) – Group assign. with Shiny (4) & Group review (1) A12 (due Mar 27) – Individual assign. with Shiny (2)
Week 11 Mar 28 & 30	Static Maps & Interactive Maps (+ Shiny) P7 (Mar 28) – Share plots (2) A13 (due Apr 3) – Group maps plots (4) & Group review (1) A14 (due Apr 3) – Individual assign. with Shiny (2) A15 (due Apr 5) – Final proposal (2)

Week 12 Apr 4 & 6	Heatmap & Animations (+ Shiny) **P**8 (Apr 4) – Share plots (2) **A**16 (due Apr 10) – Group animations (4) & Group review (1)
Week 13 Apr 11 & 13	Finals **P**9 (Apr 11 & 13) – Final presentations (20) with a link/file on reproducing your work (2) **A**17 (due Apr 17) – Create your own plots (4*2)
Week 14 Apr 18 & 20	Finals **P**9 (Apr 18 & Apr 20) – Final presentations **A**18 (due Apr 26) – Create your own plots (4*2)
Week 15 Apr 27	Finals **P**9 (Apr 27) – Final presentations **A**19 (due May 1) – Create your own plots (4) Note that you can update your final documents by Apr 29.

Weekly Assignments and Projects

Weekly Assignments: The weekly assignments consist of focused exercises related to each week's lectures (e.g. finding articles/data/figures, making and describing plots, or reproducing peers' work). The assignments are uploaded on Collab, and each due is specified on the course schedule. Many of the assignments are group assignments. For group assignments, you will work with your group members, but each group member must be able to submit group assignments using the tools they've installed on their own computer. Our TA will view your submissions and grade them based on rubrics. At the end of your assignments, you may also address issues you want to see discussed or further clarified in class.

Midterm Project: The midterm project is to find an article and related datasets(s) to support or argue against your article that you would like to raise awareness about in our class. You can present controversial or hot button topics, e.g., Covid and/or Governmental and Healthcare Industry Response Ability; the Climate and/or Economic Progress; Vaccines and/or Relative Risks of Global Health Issues; Healthcare and/or Disparity; Vegetarian Heath and/or Meat Production Efficiency; Fluoride in Drinking Water and/or Socio Economic Disparate Impact; Social Justice (e.g., Black Lives Matter) and/or Gender Bias (e.g., Feminism: the radical notion that women are people). Of course, you can present any topic that is of interest to you, e.g., is coffee bad for you?

   The *midterm proposal* should briefly describe your topic along with an article you've chosen and dataset by creating a post on the discussion board with your article title as the title and (a) a link to access your article, (b) a link to a dataset which you can argue is relevant, related to, and informative of the article you've chosen, and (c) a plot that you like to create.

   Your *midterm presentation* will be evaluated based on (i) how well it explains the article, (ii) how much data is related to the article, (iii) how visually attractive it is, and (iv) how engaging it is.

Final Project: For the final project, you should also find a topic that you like to discuss. You can use the same topic as in your midterm project, or you can choose a new topic that is of interest to you. Your final project needs to leverage data to strengthen and highlight the points you are trying to make and present to your audience. If your point is to show the limitations of the data you have with respect to the given topic you address, that could also work.

   Your final project consists of (1) a proposal, (2) a standalone document report, (3) a document/guide for reproducing your work, and (4) a presentation. The *final proposal* should also briefly describe your topic along with an article you've chosen and dataset by creating a post on the discussion board with your topic as the title and (a) relevant article(s), (b) a link to a dataset which you can argue is relevant, related to, and informative of the article you've chosen, and (c) a plot that you like to create.

   The *standalone document report* is like an article you might read from a data literate publication and the presentations you've seen (and will continue to see) over the course of the class. Your final report will be a written article (formatted as a knitted .Rmd/.html, or similar file) that includes compelling presentations of data that help the reader understand, be better informed, or perhaps even convinced and won over to the views presented in your article.

   Also, you will create a *document/guide that helps peers reproduce your final report.* You should make detailed annotations on your R codes so that peers can easily reproduce your work.

   Lastly, you will present your report document in class. Your *final presentation* will be evaluated based on (i) how well it explains the article, (ii) how much data is related to the article, (iii) how visually attractive it is, (iv) how engaging it is, and (v) how compelling it is.

Due Dates

Assignments/Projects are expected to be completed by 11:59pm, ET on due date. For the late assignment, it will be accepted by 11:59pm of the following day, but 15% will be deducted from the assignment score. Any issues addressed after a due date will receive no consideration with respect to the due date in question: *report issues immediately as they arise*. If you are not going to meet a given deadline, this must be addressed before the due date in question, and barring incredibly unique and fantastic circumstances will receive no consideration after the fact: *report issues immediately as they arise*.

Grading
Courses carrying a Data Science subject area use the following grading system: A, A-; B+, B-; C+, C-; D+, D, D-; F. The symbol W is used when a student officially drops a course before its completion or if the student withdraws from an academic program of the University.

Requirement	max. points
1. Assignments	87
2. Presentations (Presenters & Audiences)	14
3. Midterm & Finals	42

Thus, the maximum number of points you can achieve is 143 points (87 + 14 + 42):

A: 135 pts or more	A-: 134-130 pts	B+: 129-125 pts	B-: 124-120 pts
C+: 119-115 pts	C-: 114-110 pts	F: <109 pts	

Note that A+ will be determined by the class curve and overall performance in the course.

Honor Code
Work which should be pledged in the spirit of the Honor System at the University of Virginia will be explicitly noted. Generally speaking, group work will require no pledge (and indeed collaboration is intended and expected for these assignments); while later individual assignments require either of the following statements:

- "I pledge that I have neither given nor received help on this [examination/quiz/assignment]."
- "I have [received/given] help on this assignment [to/from] [student]; the nature of this support was [elaborate]."

For more information, visit www.virginia.edu/honor.

Accommodations
Special accommodations are available through the Student Disability Access Center. If you anticipate any issues related to the format, materials, or requirements of this course, please get more information on the process of applying for special accommodations, and apply for such services online, at the SDAC website at sdac.studenthealth.virginia.edu. Once you are approved for accommodations through SDAC, please send me (a) your accommodation letter, and (b) your specific suggestions as to how we can best accommodate your needs in our class. Generally speaking, all students, regardless of accommodation status, will be required to maintain course cadence throughout the semester.

Respectful Learning Environment
Courses in Data Science are venues for the free, open, and respectful exchange of ideas. Class members are expected to respect others' backgrounds and to contribute to a healthy learning environment in all course activities. Concerns in this regard should be brought to the attention of the course instructor.