INTL 4004: Data Analytics for International Policy

Spring 2024

Instructor:	LIAN, Jie/Jason	**Time:**	T, TR 9:35 am - 10:50 pm (CRN: 67222)
Email:	jasonlian@uga.edu	**Place:**	Baldwin Hall 0311.

Office Hours

- **Office Hours:** Candler Hall 217 – 12:30 pm to 1:30 pm on Tuesday.

- **Online Office Hours:** Zoom (meeting ID: 4827643495) – 10:30 am to 11: 30 am on Wednesday

- **Other Times:** By appointment

Course Description and Objectives

This course will provide students with the tools to find, critically examine, analyze, understand, and present politics-related data to a broad community. Although the focus is on data analytics, **no background in statistics or programming is expect or required**. The course will provide an overview of using R, the leading statistical programming language, to help work with multiple data types interesting to the policy community. At the end of the course, students will be able to claim a working knowledge of R, Qualtrics, quantitative data analysis, networked data analysis, survey methods, and machine-learning-based text mining techniques. If time permits, at the end of the semester, we will explore the applications of cutting-edge Artificial Intelligence (AI) techniques in the political study on new data formats like video and radio.

Course Textbook

There is no required textbook for the course. However, the following books are highly recommended. Students can find the resources online or through the UGA library.

- *R for Data Science* by Hadley Wickham and Garrett Grolemund. The full text is available online.

- *Introduction to Computing with Data* by Gaston Sanchez. The full text is available online.

- *An Introduction to Statistical Learning with Applications in R* by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. The full text is available online.

Course Format

This course is in-person. All materials, including the slides and the additional materials, are available online through UGA's eLearning Commons (eLC) portal.

The course consists of **five modules**. Each module consists of **two to four topics**. Each topic will include:

1. One to three required readings of helpful tutorial texts, current or canonical academic research on the topic or classic ;

2. A set of slides on the topic, which will be uploaded onto eLC by the end of each week;

3. A set of sample code on the topic, which will be uploaded onto eLC by the end of each week;

4. A set of additional materials/video links you could explore for further information on the topic.

I am here to help however I can. Please feel free to email at any time (jasonlian@uga.edu). I strive to answer all emails within 24 hours. I am happy to schedule a time to chat in person or over video.

Grading

- 40% - Four homework assignments, 10% each

- 10% - Four pop-up quizzes

- 40% - Term Project Assignment

- 10% - Attendance (5%) and participation (5%)

Homework

There will be four homework assignments throughout the semester. The first three assignments will be on the materials of Topic 2 - 4 in Module 1. For the last assignment, students can choose one out of three assignments on the materials of Module 2 and Module 4. Each homework will be graded on a 100 scale and takes 10% of the final grade.

Each homework will be released at ELC on the Friday of the corresponding week when the topic is discussed. Students will have two weeks to finish each homework.

Important Dates for the Homework Assignments

Homework 1: Exploratory Data Analysis (EDA)
Assigned on Jan. 18 . . . . by 11:59pm (EST)
Due on Feb. 01 . . . . by 11:59pm (EST)
Homework 2: Linear Regression Analysis
Assigned on Feb 01 . . . . by 11:59pm (EST)
Due on Feb. 15 . . . . by 11:59pm (EST)
Homework 3: Linear Regression Diagnostics
Assigned on Feb. 15 . . . . by 11:59pm (EST)
Due on Feb. 29 . . . . by 11:59pm (EST)
Homework 4 (Option 1): Networked Data
Assigned on Feb 29 . . . . by 11:59pm (EST)
Due on Mar. 12 . . . . by 11:59pm (EST)
Homework 4 (Option 2): Text as Data 1
Assigned on Mar. 28 . . . . by 11:59pm (EST)
Due on Apr. 09 . . . . by 11:59pm (EST)
Homework 4 (Option 3): Text as Data 2
Assigned on Apr. 09 . . . . by 11:59pm (EST)
Due on Apr. 18 . . . . by 11:59pm (EST)

Pop-up Quizzes

There will be four in-class pop-up quizzes throughout the semester. Each quiz is graded on a 10 scale and takes 2.5% of the final grade. The purpose of the pop-up quizzes is to **reward the students who attend the classes, and pay attention and no make-up quiz will be provided even with excused absence**. The time of the quizzes and all questions will be mentioned or hinted at during the lectures but not on the notes or slides. Students who pay close attention will get a good grades.

Pop-up Quiz Waivers

Throughout the semester, I will give out "1/2 Pop-up Quiz Waivers". Two "Pop-up Quiz Waivers" can help you either drop the lowest pop-up quiz score or skip one, whichever you want! To win a pop-up quiz waiver, there are several things you can do:

- Answer your classmates' questions in the R playground under the discussion board. The first non-replicated right answer will win a waiver!

- Contribute to our class discussion, especially helping your classmates when we run code during the class (which we will do a lot in the coming classes). Significant contributions recognized by Jason will win you a pop-up quiz waiver.

- Correct Jason when he makes a mistake! We are supposed to construct an interactive learning community from which you guys and myself can all learn!

Once you get two pop-up quiz waivers, shoot me an email, and you will be able to "cash it"!

Term Project Assignment

This semester, you'll be trained to become a qualified researcher who can analyze different data types, extract valuable information, and present your findings to audiences in the policy community and beyond. The term project assignment offers students invaluable hands-on experience by engaging them in real data analytic projects centered around politics-related topics. Students can **work alone or in teams up to 4 for the term projects**. (Team work is highly encouraged.) There are two options for the projects.

Research Replication

Students can choose one of the published articles from the top journals in political science and public administration related areas and replicate the main quantitative findings from the piece. The main requirements are:

- The article should be an empirical piece with replicable results and a published online dataset.

- The article choice should get the permission of the instructor.

- The instructor might ask students to manipulate the original dataset and present results outside the paper, like introducing extra control variables, running model diagnostics (we will learn this in Topic 3!), etc.

Self-purposed Project

Students are welcome to propose their own projects. It could be the student's degree project, an unfinished research project from another course, or an ongoing intern project. (*Caution: Data acquisition is an*

extremely laborious and time-consuming task. Make sure you already possess the necessary data or know you can immediately acquire them in digital form, you will not have time to conduct a project starting with data collection). The main requirements are:

- The project topic must be based on empirical interests and involve data analytics.

- The project topic should get the permission of the instructor.

- The instructor might ask students to present data analytic results outside the original interests of the project, like different types of exploratory data analysis (we will learn this in Topic 2!), running model diagnostics (we will learn this in Topic 3!), etc.

No matter the student's choice of the term project, the assignment will be thoughtfully divided into four manageable steps, each scheduled at different points throughout the semester. The four steps are: the question statement, exploratory data analysis, empirical analysis and interpretation, and a final poster presentation of the empirical results geared to a wide audience. Each step will have a draft stage and a final stage. The final product of each step will be graded. And I will offer feedback on your works at the draft stage. We will talk more about assignments details, expectations, and assessment in class. **The rubrics will be released on eLC later**. The tentative due dates for each assignment are listed below.

<div align="center">

Important Due Dates for the Term Project Assignment (Tentative)

</div>

The Question Statement:
Draft Stage (assigned on Jan. 9)............Jan. 25, by 11:59pm (EST)
Final Product (assigned on Jan. 25)........Feb. 08, by 11:59pm (EST)
Exploratory Data Analysis:
Draft Stage (assigned on Feb. 08) ....Feb. 22, by 11:59pm (EST)
Final Product (assigned on Feb. 22)........Mar. 14, by 11:59pm (EST)
Empirical Analysis and Interpretation:
Draft Stage (assigned on Mar. 14) ....Mar. 28, by 11:59pm (EST)
Final Product (assigned on Mar. 28).......Apr. 09, by 11:59pm (EST)
Poster Presentation:
Draft Stage (assigned on Apr. 9)...........Apr. 16, by 11:59pm (EST)
Final Product (assigned on Apr. 16) ....Apr. 25, by 11:59pm (EST)

Attendance

Consistently attending lectures is crucial for an effective and enriching learning experience. I will take attendance every class. There are **1 unexcused absence**. I do not need to know the reason for the absence. **Any unexcused absences after the freebie** will result in **a half point deduction from your final attendance grade for each absence**.

Excused absences are given when I receive one of the following: an official UGA excuse, a medical note, or a student-athlete responsibility. Documentation is required for all of these (For more information, please see the Class Attendance Policies.).

Participation

Participation, including in-class interaction, in-class activities, and term project assignments is worth 5% of your final grade. Keep in mind that a not-so-great attendance record can erase participation, because

it's hard to participate if you are not around. **Even perfect attendance without active participation does not guarantee a good participation score.** What is more, **disruptive behaviors will result in one point deduction**. The details can be found on the table below.

A note on participation: Your active participation is very important. Always feel free to state your opinions in a way that invites discussion. Active participation through **asking questions and seeking clarifications** is a vital and strongly encouraged aspect of our class discussions.

5 points	Student attends classes and regularly participates in class discussions. Student contributions critically engage with the material and with class peers.
4 points	Student attends classes and regularly participates in class discussions. Student contributions critically engage with the material.
3 points	Student attends classes and regularly participates in class discussions. Student contributions involve examples from personal life but they lack critical engagement with the material.
2 points	Student attends classes and occasionally participates in class discussions.
1 points	Student attends classes but does not participate in the discussions.
0 points	Student does not attend the class.

INTL 4004 R Play Yard

You can find special public space under "Discussion" section on eLC called "INTL 4004 R Play Yard." You can post your **R-coding or data analytic-related questions** in this place. I will **publicly provide the answers** in case other people meet the same challenges.
You are **highly encouraged to help each other out**. For the students providing the **first correct solutions** will get a "pop-up quiz waiver"!

- You must be the first to provide a correct solution to a posted question (before me and your classmates).

- If one posted question already has a correct answer, but your solution is essentially different (and better), it still counts.

- You cannot answer your own post.

- This process is cumulative. (4 answers could waive two quizzes!)

- Send me the screenshots of your answers to get the quiz waiver.

Posting and answering questions on the Play Yard **count as your class participation**!

Grading Scale

Your final grade will be calculated on the following scale:

- 94 to 100 – A   90 to 93 – A-

- 87 to 89 – B+   84 to 87 – B   80 to 83 – B-

- 77 to 79 – C+   74 to 77 – C   70 to 73 – C-

- 67 to 69 – D+   64 to 67 – D   60 to 63 – D-

- 59 and below – F

Trouble-shooting Sources

Mastering data analytics and coding demands dedicated self-learning. To support your journey, here are some valuable sources to aid your self-teaching and troubleshoot any challenges you may encounter in this course or in the future.

- **Stackoverflow**: One of the largest online platforms for developers seeking coding-related solutions and support, covering various programming languages, including R. Available at https://stackoverflow.com/.

- **R-bloggers**: Another important online platform dedicated to R programming with lots of helpful tutorials. Available at https://www.r-bloggers.com/.

- **CRAN (The Comprehensive R Archive Network)**: The website comprehensively archives all published R libraries along with their corresponding documents. I particularly find the vignettes (short coding samples) posted on some of the R library pages to be exceptionally helpful. Available at https://cran.r-project.org/.

- **ChatGPT**: Though it may be difficult for some to admit, ChatGPT actually serves as a competent coding coach, especially for handling simple tasks. An important point to bear in mind is that ChatGPT relies on generative language models (we'll introduce this briefly in Topic 9 if time permits!) and cannot guarantee the production of flawless results (till the time I construct the syllabus). Always remember to execute the code generated by ChatGPT to ensure its functionality. Available at https://chat.openai.com/.

- **Google**: Sometimes, the right answer is just one "enter" away! Simply googling your question can surprise you with invaluable help. The trick lies in **asking the right question**! Available at https://www.google.com/.

Useful Information and University Policies

Make-up Quiz

Legitimate excuses for absence from a quiz (e.g., religious holiday, medical emergency, or illness) must be presented to the instructor and accepted prior to the quiz when feasible. Grades are due immediately after the course is over; late or missing assignments at this time may cause difficulties in grade reporting. For more information about legitimate excuses, please refer to the UGA class attendance policy.

Term Project Assignment Submission

If you miss a due date for a term project assignment, I will allow all groups one "freebie" or late submission request: you or your group can have extra 24 hours for the assignment. Please simply email the instructor (jasonlian@uga.edu). Grades are due immediately after the course is over; late or missing assignments at this time may cause difficulties in grade reporting. Beyond the one free late submission request, please refer to the UGA class attendance policy.

Changes to the Syllabus Could Occur

The following is taken verbatim from the University Council: "The course syllabus is a general plan for the course; deviations announced to the class by the instructor may be necessary."

University Honor Code and Academic Honesty Policy

UGA Student Honor Code: "I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others." A Culture of Honesty, the University's policy and procedures for handling cases of suspected dishonesty, can be found the website of Office of Instruction.

We expect that the Student Honor Code will guide your efforts in this course. A lack of knowledge of the academic honesty policy does not explain a violation. Please direct any questions to the instructor.

Prohibition on Recording Lecture

The following is taken verbatim from LINK:

"In the absence of written authorization from the UGA Disability Resource Center, students may not make a visual or audio recording of any aspect of this course. Students who have a recording accommodation agree in writing that they:

- Will use the records only for personal academic use during the specific course.

- Understand that faculty members have copyright interest in their class lectures and that they agree not to infringe on this right in any way.

- Understand that the faculty member and students in the class have privacy rights and agree not to violate those rights by using recordings for any reason other than their own personal study.

- Will not release, digitally upload, broadcast, transcribe, or otherwise share all or any part of the recordings. They also agree that they not profit financially and will not allow others to benefit personally or financially from lecture recordings or other course materials.

- Will erase/delete all recordings at the end of the semester.

- Understand that violation of these terms my subject them to discipline under the Student Code of Conduct or subject them to liability under copyright laws."

Preferred Name and Pronouns

The following is taken verbatim from the Inclusive Statements for Syllabi:

"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. I am eager to address you

by your preferred name and/or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records."

FERPA Statement

The following is taken verbatim from the Syllabus Checklist:

"The Federal Family Educational Rights and Privacy Act (FERPA) grants students certain information privacy rights. See the registrar's explanation. FERPA allows disclosure of directory information (name, address, telephone, email, date of birth, place of birth, major, activities, degrees, awards, prior schools),unless requested in a written letter to the registrar."

Disability Services

The following is taken verbatim from the Disability Resource Center:

"If you plan to request accommodations for a disability, please register with the Disability Resource Center. They can be reached by visiting Clark Howell Hall, calling 706-542-8719 (voice) or 706-542-8778 (TTY), or by visiting the Disability Resource Center"

The following is taken verbatim from the document of Syllabus Creation:

"Students with disabilities who require reasonable accommodations in order to participate in course activities or meet course requirements should contact the instructor or designate during regular office hours or by appointment."

I want to help all students succeed in this course!

Mental Health and Wellness Resources

The following is taken verbatim from the University Council:

"If you or someone you know needs assistance, you are encouraged to contact Student Care and Outreach in the Division of Student Affairs at 706-542-7774 or visit the website of Student Care and Outreach. They will help you navigate any difficult circumstances you may be facing by connecting you with the appropriate resources or services.

UGA has several resources for a student seeking mental health services or crisis support. If you need help managing stress anxiety, relationships, etc., please visit BeWellUGA for a list of FREE workshops, classes, mentoring, and health coaching led by licensed clinicians and health educators in the University Health Center. Additional resources can be accessed through the UGA App."

Coronavirus Information for Students

I can offer accommodations for students with positive COVID-19 tests. Students who test positive SHOULD BE accommodated.

I'll accommodate students who test positive for COVID as I would have accommodated ill students prior to the pandemic. I will assist students who may need to isolate per CDC guidelines, students with Disability Resource Center (DRC)-recommended accommodations, or other circumstances that may require a student to occasionally miss a class. For all other students, classes must continue to be offered in an in-person format.

Please email me (jasonlian@uga.edu) when you need the accommodation.

Useful Campus Resources

There are many campus resources that can help you with your academic performance and assist you during the semester. I urge you to check out the Division of Academic Enhancement and the Writing Center

Some material in this course contains references to violence. I urge you to take your mental health seriously. There are campus resources to help.

Course Outline

Module 1: Tabular Data Analysis

Topic 1: Intro to Course & the Conception of Data
Week 1: Jan. 09

- Reading:

  - Syllabus
  - *Introduction to Computing with Data*, Chapter 6 Intro to R and RStudio

- Slides:

  - Intro to Course & the Conception of Data
  - R & RStudio Installation
  - Term Project Prep

- Sample Code

- Recommended Readings/Links:

  - Cukier, Kenneth, and Viktor Mayer-Schoenberger. 2013."The rise of big data: How it's changing the way we think about the world." *Foreign Affairs* 92:28.

Topic 2: Exploratory Data Analysis (EDA)
Week 1-2: Jan. 11 - Jan. 18

- Readings:

  - *R for Data Science*, Chapter 7 Exploratory Data Analysis
  - *Introduction to Computing with Data*, Chapter 9 Exploratory Data Analysis with **dplyr**
  - *Introduction to Computing with Data*, Chapter 10 Exploratory Data Analysis with **ggplot2**

- Slides:

  - R Basics
  - Exploratory Data Analysis & the Application in R
  - Data Wrestling in R
  - Data Visualization in R

- Sample Code

- Recommended Readings/Links:

  - *Data Visualization with R* by Rob Kabacoff. The full text is available online.
  - A helpful post from Paula: https://rforpoliticalscience.com/2022/02/06/exploratory-data-analysis-and-descriptive-statistics-for-political-science-research-in-r/.
  - Cheat sheets in R on eLC.

Topic 3: Crash Course in Multivariate Analyses
Week 3-4: Jan.23 - Feb. 01

- Readings:

- – Gallo, Amy. 2015. "A refresher on regression analysis." *Harvard Business Review*
  https://hbr.org/2015/11/a-refresher-on-regression-analysis.
- – *An Introduction to Statistical Learning with Applications in R*, Chapter 3.1 - 3.2 Simple Linear Regression & Multiple Linear Regression.
- – Regression Diagnostics with R, Available online.

- Slides:

  - – Linear Regression Basics & Practices
  - – Multivariate Regression Introductory & Practices
  - – Regression Diagnostics in R
  - – Regression Results Reporting & Visualization in R
  - – Intro to Model Selection & Other Advanced Topics
  - – *Supplement: Generalized Linear Regression and Diagnostics in R*

- Sample Code

- Recommended Readings/Links:

  - – *The Elements of Statistical Learning Data Mining, Inference, and Prediction* by Trevor Hastie, Robert Tibshirani, Jerome Friedman. Full text available on line. Chapter 3 Linear Methods for Regression. (Good illustration of linear model from data science/machine learning perspective.)
  - – *Applied Linear Statistical Models,* by Michael H Kutner, Christopher J. Nachtsheim, John Neter and William Li. (Canonical textbook on linear regression from statistical perspective.)
  - – Excellent video series on linear regression in R available on YouTube.
    Available at https://youtu.be/wsi0jg$_g$H28.

Topic 4: Wrestling with Big Database
Week 5: Feb. 06 - Feb. 08

- Readings:

  - – *Exploring, Visualizing, and Modeling Big Data with R* by Okan Bulut, Christopher Desjardins Chapter 2 Introduction.
    Available online.

- Slides:

  - – Intro of Big Data and Database Management
  - – Using SQL in R
  - – Using Google BigQuery in R

- Sample Code

- Recommended Readings/Links:

  - – A useful tutorial for using SQL in R https://www.datacamp.com/tutorial/tutorial-how-to-execute-sql-queries-in-r-and-python
  - – Another useful online tutorial: *How to use R with BigQuery.*
    Available at https://www.infoworld.com/article/3622926/how-to-use-r-with-bigquery.html

Module 2: Networked Data

Topic 5: Maps and Geographic Data
Week 6: Feb. 13 - Feb. 15

- Readings:

    - *GIS and Spatial Analysis with R* by Manny Gimond.
    - Cohen, Dara Kay, and Ragnhild Nordås. 2014. "Sexual violence in armed conict: Introducing the SVAC dataset, 1989-2009." *Journal of Peace Research* 51(3): 418-428.
    - Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research*, 50(4): 523-532.

- Slides:

    - Intro of Geographic Information Systems in R
    - Making Maps in R

- Sample Code

- Recommended Readings/Links:

    - *Intro to GIS and Spatial Analysis* by Manuel Gimond Appendix A Reading and writing spatial data in R, Appendix B Mapping data in R.
    - Sobolev, A., Chen, M.K., Joo, J. and Steinert-Threlkeld, Z.C., 2020. News and geolocated social media accurately measure protest size variation. *American Political Science Review*, 114(4), pp.1343-1351.

Topic 6: Network Analysis Data
Week 7: Feb. 20 - Feb. 22

- Readings:

    - *Network Analysis and Visualization with R and igraph* by Katherine Ognyanova. Available at https://kateto.net/netscix2016.html
    - Murdie, Amanda. 2014. "The Ties that Bind: A Network Analysis of Human Rights INGOs." *British Journal of Political Science*. 44(1): 1-27.
    - Asal, Victor, and R. Karl Rethemeyer. "Researching terrorist networks." Journal of Security Education 1.4 (2006): 65-74.

- Slides:

    - Network Science Overview
    - Network Visualization in R
    - Intro of Network Inference in R

- Sample Code

- Recommended Readings/Links:

    - *Statistical Analysis of Network Data with R* by Eric D. Kolaczyk , Gábor Csárdi Chapter 4 Descriptive Analysis of Network Graph Characteristics.

Module 3: Talk-with-people Data

Topic 7: Interviews
Week 8: Feb. 27 - Feb. 29

- Readings:

    - Aberbach, Joel D., and Bert A. Rockman. "Conducting and coding elite interviews." *PS: Political Politics & Politics* 35.4 (2002): 673-676.

    - Barney, Morgan et al. "From Age to Agency: Frame Adoption and Diffusion Concerning the International Human Rights Norm Against Child, Early, and Forced Marriage." *Human Rights Review* 23(4), 503-528.

- Slides: Interview Overview

- Recommended Readings/Links:

Topic 8: Surveys and Survey Experiments
Week 10: Mar. 12 - Mar. 14 (Spring Break – No Class in Mar. 05-07)

- Readings:

    - Druckman, James N., and Donald P. Green. "A New Era of Experimental Political Science." *Advances in Experimental Political Science* (2021). Chapter 1.

    - *Data Analysis in R*, by Steve Midway Chapter 7 Understanding ANOVA in R.

- Slides:

    - Overview of Surveys

    - Survey Experiment Data Analysis in R

- Sample Code

- Recommended Readings/Links:

    - *An useful online tutorial on experimental design in R* by kaelen medeiros.
      Available online.

Module 4: Text as Data

Topic 9: Intro to Text Mining
Week 11: Mar. 19 - Mar. 21

- Readings:

    - Grimmer, J. and Stewart, B.M., 2013. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), pp.267-297.

    - Gilardi, F., Shipan, C.R. and Wüest, B., 2021. Policy diffusion: The issue-definition stage. American Journal of Political Science, 65(1), pp.21-35.

- Slides:

    - Text as Data

    - Applying Topic Model Using R

    - Applying Sentiment Analysis Using R

- Sample Code

- Recommended Readings/Links:

  - *Text Mining with R: A Tidy Approach*, by Julia Silge, David Robinson.
    Available online.
  - Wilkerson, J. and Casas, A., 2017. Large-scale computerized text analysis in political science:
    Opportunities and challenges. *Annual Review of Political Science*, 20, pp.529-544.

Topic 10: Social Media Data (Short Text)
Week 12: Mar. 26 - Mar. 28

- Readings:

  - Larson, J.M., Nagler, J., Ronen, J. and Tucker, J.A., 2019. Social networks and protest partic-
    ipation: Evidence from 130 million Twitter users. *American Journal of Political Science*, 63(3),
    pp.690-705.
  - Zeitzo, Thomas. 2017. "How Social Media Is Changing Conflict." *Journal of Conflict Resolution*
    61(9): 1970-91.
  - Guess, Andrew et al., 2023, "Reshares on social media amplify political news but do not detectably
    affect beliefs or opinions." *Science.* Available online.

- Slides:

  - Visualizing Short Text Corpora in R.
  - Text Network Analysis in R.
  - Social Media Data Scrapping

- Sample Codes

- Recommended Readings/Links:

  - A good web page with text analysis tutorials from the University of Pennsylvania:
    https://guides.library.upenn.edu/penntdm/r.
  - A great Youtube video made by Dr. Chris Bail on text network analysis in R: https://youtu.be/-
    2dBI2ULcQw

Topic 11: Archived Data (Long Text)
Week 14: Apr. 09 - Apr. 11 (ISA Week – No Class in Apr. 02-04)

- Readings:

  - Park, B., Murdie, A. and Davis, D.R., 2019. The (co) evolution of human rights advocacy:
    Understanding human rights issue emergence over time. *Cooperation and Conflict*, 54(3), pp.313-
    334.
  - Park, B., Greene, K. and Colaresi, M., 2020. Human rights are (increasingly) plural: Learning the
    changing taxonomy of human rights from large-scale text reveals information effects. *American
    Political Science Review*, 114(3), pp.888-910.

- Slides:

  - Applying Machine Learning in Text Analysis for Political Research
  - A (Very) Brief Intro of Large Language Model

- Recommended Readings/Links:

  - Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J.W., Kreps, S. and McCain, M., 2019. Release strategies and the social impacts of language models. arXiv preprint arXiv:1908.09203.

Module 5: A Quick Peek on the Wider World (AI Applications in Political Study)

Topic 12: Image Data
Week 15: Apr. 16

- Readings:

  - Casas, A. and Williams, N.W., 2019. Images that matter: Online protests and the mobilizing role of pictures. *Political Research Quarterly*, 72(2), pp.360-375.
  - Joo, J. and Steinert-Threlkeld, Z.C., 2022. Image as data: Automated content analysis for visual presentations of political actors and events. *Computational Communication Research*, 4(1).

- Slides:

  - Utilizing Image Data in Political Analysis
  - A (Very) Short Crush Course on Deep-Learning-based AI Algorithms
  - Studying AI-based Computer Vision Technologies: A Self-Learning Guide

- Sample Code in Python

- Recommended Readings/Links:

  - Williams, N.W., Casas, A. and Wilkerson, J.D., 2020. *Images as data for social science research: An introduction to convolutional neural nets for image classification.* Cambridge University Press.
  - Loken, M., 2021. Using images as data in political violence research. *Journal of Human Rights*, 20(3), pp.373-379.
  - Dietrich, B.J., 2021. Using motion detection to measure social polarization in the US House of Representatives. *Political Analysis*, 29(2), pp.250-259.

Topic 13: Voice Data
Week 15: Apr. 18

- Required Readings:

  - Dietrich, B.J., Hayes, M. and O'brien, D.Z., 2019. Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review*, 113(4), pp.941-962.

- Slides: Voice Data and Political Research

- Recommended Readings/Links:

  - Knox, D. and Lucas, C., 2021. A dynamic model of speech for the social sciences. *American Political Science Review*, 115(2), pp.649-666.
  - Klofstad, C.A., 2016. Candidate voice pitch influences election outcomes. *Political Psychology*, 37(5), pp.725-738.

Term Project Presentations
Week 16: Apr. 23 - Apr. 25