Social Media, Data Science, and Political Science Research
Brain Korea 21 Program in Yonsei Political Science
Wednesday August 24th & Thursday August 25th, 9:00AM - 12:00PM KST

James Bisbee (jameshbisbee@gmail.com)
Office Hours: Friday August 26th 9:00AM-10:00AM KST or by appointment

Course Description: The study of political science has undergone a radical change over the past 20 years with the advent of online social media. These websites generate abundant data which technically sophisticated researchers can exploit to investigate political science questions that had previously been intractable.

This course is fundamentally about how to conduct this research, starting with the identifying and gathering raw data, followed by cleaning and preparing the processed data, then turning to tools and methods for analyzing the data, before concluding with an overview of the limitations and ethical considerations of using these data. The lecture will be hands-on experiences, combining teacher presentations with in-class work using the free software package `R`.

Expectations: Throughout, students are expected to (1) read assigned articles prior to lecture, (2) participate actively during lecture discussions, and (3) install the free software package `R` prior to the first lecture, as well as the free GUI `RStudio`. `R` can be downloaded for free from `https://cran.r-project.org/`. `RStudio` can be downloaded for free from `https://www.rstudio.com/products/rstudio/download/`.

In the course outline below, assigned readings and other homework must be completed prior to the lecture. All readings will be shared with the students via email prior to the lecture. The "example.csv" dataset will be shared with the students during Lecture #1. The problem set will be handed out to students at the end of Lecture #1.

Tentative Course Outline:

Lecture	Content

#1 (8/24)	• What is data science? How can data science empower us to use online social media for research? And what are some examples of social science questions that can be answered with these data science methods applied to these social media data sources?
	
	  – **Assigned Reading:**
	    ∗ "How Data Science is Used in Social Media" `https://www.datasciencegraduateprograms.com/social-media/`
	    ∗ "Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber?" (2015). Barbera, Jost, Nagler, Tucker, Bonneau.
	  – **Other Homework:**
	    ∗ Download and install both `R` and `RStudio`
	    ∗ Create a class folder labeled "2022_social_media" with the following subfolders:
	      · Code
	      · Data
	      · Figures
	      · Tables
	    ∗ Copy the provided "example.csv" file into the "Data" folder

#2 (8/25)	• Stepping through the pipeline using two applied examples. Students will learn how to gather, process, analyze, and describe data using `R`.
	
	  – **Assigned Reading:**
	    ∗ "Election Fraud, YouTube, and Public Perception of the Legitimacy of President Biden" (2022). Bisbee, Brown, Lai, Bonneau, Nagler, Tucker.
	    ∗ "Echo Chambers, Rabbit Holes, and Algorithmic Bias: How YouTube Recommends Content to Real Users" (2022). Brown, Bisbee, Lai, Bonneau, Nagler, Tucker.
	  – **Other Homework:**
	    ∗ Problem set (handed out in class)

Detailed Course Outline:

Section	Content
#1 (8/24: 9-10AM)	• What is data science? How can data science empower us to use online social media for research? – 45 minute lecture on the first reading – 15 minute discussion
#2 (8/24: 10-11AM)	• Twitter as political science data – 45 minute lecture on the second reading – 15 minute discussion
#3 (8/24: 11-12PM)	• Setting up `R`, downloading and opening data, introduction to basic functions. – 1 hour tutorial on `R`
#4 (8/25: 9-10AM)	• Review of problem set homework. Recap of data science workflow and pipeline. – 15 minute review of homework – 30 minute lecture on pipeline – 15 minute discussion
#5 (8/25: 10-11AM)	• First applied example of data science pipeline. – 45 minute lecture on the first reading – 15 minute discussion + applied examples using `R`
#6 (8/25: 11-12PM)	• Second applied example of data science pipeline. Discussion of data science ethics. – 15 minute lecture on the second reading – 15 minute discussion + applied examples using `R` – 30 minute lecture + discussion on ethics in data science