

INTA 708 - Data Science and Visualization for Policy Analysis

Wednesdays, 1330-1620 in Allen 1055 with [Dr. Jesse Sowell](#)

*No prerequisites, **designed for students with little to no coding experience***

All texts, readings, and software for this course are freely available online

Modern policy analysts are faced with developing a compelling visual narrative from a surfeit of data. This course will equip you with the data science skills necessary to make sense of our increasingly data rich policy environment. Data science and visualization provides the tools for cleaning, integrating, and transforming this data into a form that can be tractably evaluated and effectively visualized. This course focuses on data management and exploratory data analysis (EDA) tools for hypothesis generation through an applied introduction to mapping, cluster analysis and principal component analysis, social network analysis, and text mining. The policy research project is structured for you to choose *your own* data adventure by delving into a policy issue and data sets that further your academic and professional interests and development. Previous projects have focused on cybersecurity, international trade flows and economic development, treaty networks, terrorism, water policy, performance of non-profits, hospital resource allocation, public procurement, comparative agriculture policy, and China's foreign policy, to name a few.

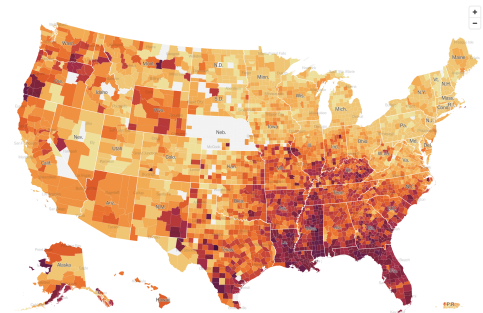
Starting with exploratory analyses driven by data visualization, you will develop and iteratively apply the data science skills and workflows necessary to identify and develop compelling visualizations of trends in your policy area, culminating in a final policy research project and analysis portfolio. At the end of this course, you will have established a strong foundation in data management and data science, *and* a portfolio of analyses and visualizations that previous students have used to provide compelling illustrations of their skills to potential employers.

Course Structure

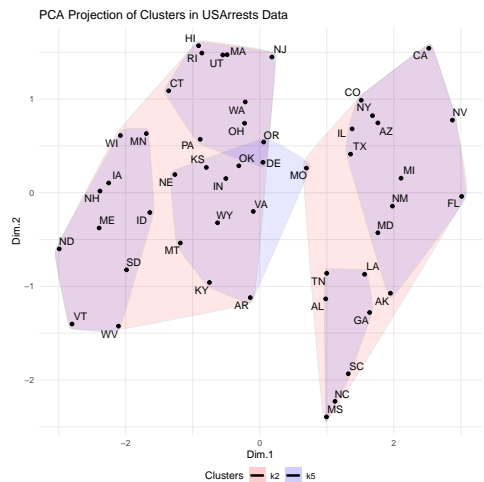
This project-based course (no exams!) develops concepts and skills through learning exercises, then applying these to your policy research project. Part I of the course introduces aesthetic strategies for developing compelling, meaningful visualizations, and how these concepts are used for hypothesis generation. The introduction also develops fundamental R, data management, data set integration (such as incorporating US Census and American Community Survey data), and data transformation skills. Throughout the course, skills are illustrated in class with small well-understood data sets, then applied to live data such as the NYTimes' COVID-19 data to create, critique, and improve on visualizations such as those in Panel 1a. Part II focuses on EDA and cluster analyses to help you understand the nuance of observed trends and outliers (such as in Panel 1b), which observations and variables contribute to different kinds of clusters, ultimately laying the foundation for effective “storytelling with data.” The remainder of the course integrates the fundamentals of common data science and visualization tools and methods: mapping, social network analysis, and text mining.

The best way to learn data science is through practice with pragmatic examples. Learning exercises give you the opportunity to work through variants of in-class examples individually *or in groups*. In your policy research project, you will apply these concepts and tools to a policy issue salient to your career path. Project milestones are structured to iteratively guide you through common data science workflows, helping you systematically work through the data cleaning, management, transformation, and visualization phases of your project. In addition to developing your data science skills, these milestones are structured to facilitate developing an analysis and visualization portfolio that effectively illustrates your analysis and understanding of the topic area you have selected. Previous students have used this course to work on topics that further their career development, topics and data for research projects they are already working on, and/or data from organizations and agencies supporting their studies. Recent graduates have indicated interviewers were quite impressed with their project portfolios.

If you have any questions about this course, please do not hesitate to contact Dr. Sowell at jsowell@tamu.edu. The current syllabus is available upon request.



(a) Example of a visualization from the media ([NYTimes](#)) we will initially analyze for aesthetics, then reproduce and improve upon in class.



(b) Visualization of a clustering and principal components analysis we work through in class.

Figure 1: Illustrations of analyses and visualizations used in class that you will learn how to design, adapt, interpret, and analyze.