The Data Science Landscape in Public Affairs

NASPAA 2022

Evolution of the Consortium

- **August 2019:** ASU/AYS host Data Science for Public Service Summit; half of top 30 NASPAA schools have courses or degrees
- **October 2019:** NASPAA conference has 15 sessions/meetings devoted to data science/civic technology
- **January 2020:** ASU/AYS share Coding in R course
- **The pandemic. . .**
- **October 2021:** ASU/AYS receive PIT-UN Challenge Grant for 2022
- **August 2022:** ASU/AYS host Data Science for Public Service Symposium; all of top 30 NASPAA schools have courses or degrees

Guiding Questions

- What is unique about data science in the public sector?
- What are working assumptions regarding the student body and career paths distinct to public sector?
- How should these needs inform curriculum for schools of public affairs?

A case for data science in public affairs. . .

Enhanced social impact through:

- Better targeting
- Actionable information
- Democratized knowledge creation
- Inclusive innovation

Compliments and Tensions

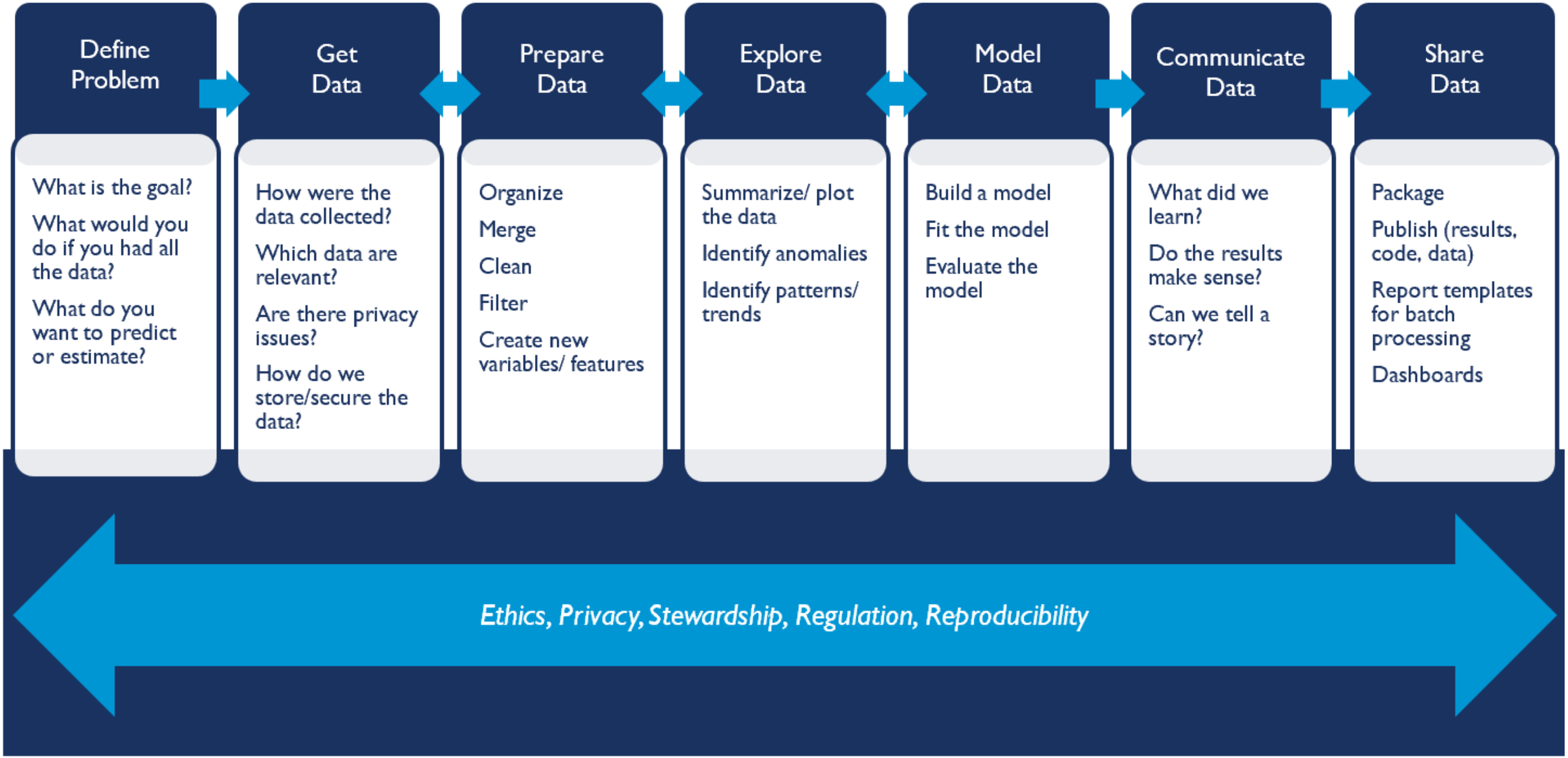
Data Science Themes

- Open Source
- Transparency
- Efficient / Profitable
- Big data
- Prediction / Automation (machine learning / AI)
- Model consumers
- Business analytics (people and orgs as units of analysis)

Public Affairs Themes

- Open Data (public goods)
- Privacy
- Accessible / Equitable
- Administrative / Survey data
- Monitoring and Evaluation
- Model citizens
- Community analytics (neighborhoods and populations as units of analysis)

DATA SCIENCE LIFE CYCLE



On-Ramp Courses

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
R Coding for Public Policy		●	●	●		●			NYU - Wagner	Install and set up R and RStudio; Understand data objects and how they relate to policy analysis; Read in, index and manipulate data objects; Find, install and use R packages; Calculate parameters such as mean, median, sum and standard deviation; Conduct T-test, ANOVA and Chi-Squared test, read the summary outputs and find the P-values; Conduct linear and logistic regressions and interpret the outputs; Plot simple graphics; Use basic SQL commands in R; Develop an effective R function for policy analysis
Python Coding for Public Policy			●	●		●	●		NYU - Wagner	Python fundamentals Common data types; Functions; How to read documentation' How to troubleshoot; Know how to use several Python packages for different kinds of data analysis, manipulation, and visualization
Introduction to Programming for Public Policy	●	●	●	●		●			Chicago - Harris	Introduction to the tools required to write and share code: text editors, the command line, the python shell, and version control (git): How to "think algorithmically," translating self-contained questions into python programs. Fundamentals of the language including types, control, functions, input/output, and scripts; debugging and (time-permitting) computability. Review of tools and recipes for retrieving, cleaning, visualizing, and analyzing ~~data~~.
Data Management & Visualization with R		●	●	●		●			Minnesota - Humphrey	Use RStudio to carry out R file and related database management; work with different types of databases and conduct basic data management; visualize data with different types of plots; carry out exploratory data analysis
Statistical Programming		●	●	●		●			American University	Programming and data analysis using the open-source statistical program R. Includes basic programming, basic data structures, data wrangling, data cleaning, data visualization, exploratory data analysis, data import and export, relational datasets, and data presentation. Emphasis is placed on the popular ~~tidyverse~~ suite of packages

CORE COURSES

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Introduction to Computational Thinking and Data Science		●		●		●		●	USC – Sol Price	Introduction to data analysis techniques and associated computing concepts for non-programmers. Topics include foundations for data analysis, visualization, parallel processing, metadata, provenance and data stewardship.
Large Scale Data Analysis I	●			●	●				NYU - Wagner	Identify and distinguish between large scale data analysis methods, focusing on three main problem paradigms (prediction, modeling, and detection). Translate policy questions into paradigms; Choose and apply the appropriate artificial intelligence and machine learning tools, with an emphasis on interpretable prediction (classification and regression) and data clustering; Interpret, evaluate, and apply the results for policy analysis and decision making.
Large Scale Data Analysis II	●			●	●				NYU - Wagner	Identify and distinguish between large scale data analysis methods, focusing on three main problem paradigms (prediction, modeling, and detection). Translate policy questions into paradigms. Choose and apply the appropriate artificial intelligence and machine learning tools, with an emphasis on Bayesian network modeling and on anomaly and pattern detection. Interpret, evaluate, and apply the results for policy analysis and decision making
Data, Evidence, Ethics, and Bias in an AI World		●			●			●	NYU - Wagner	This course will provide students with an overview of current approaches to examining potential biases and ethical challenges in that implementation, with a particular focus on data collection, model development and evidence building. It will focus in on particular use cases in social science, transportation, climate and industry.
Introduction to Database Design, Management, and Security		●	●					●	NYU - Wagner	Explain the value of databases, importance of database architecture, considerations for data integration, and provide an overview of database tools and trends. Learn the management policies, practices, and procedures required for maintaining information integrity, security, and privacy. Apply concepts and practices to database design and implementation

Extension Courses

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Foundations of Data Science III: Project Management		●	●			●	●		ASU - Watts	This course covers the main tools and practices of managing large or complex data projects, typically involving teams. You will learn about project management tools used by open-source software developers and useful tools for creating client reports.
Geographic Information Systems for Public Managers		●	●	●		●			UIC	Fundamental GIS tools and applications as well as the challenges in implementing and sustaining a GIS function in the public setting.
Social Network Analysis				●	●	●			UIC	Covers network data collection, research design, visualization, and inferential techniques for cross-sectional and longitudinal network data.
Geographic Data Programming		●	●	●		●			Georgetown - McCourt	Geographic Information Systems (GIS) are used as tools for describing, analyzing, managing, and presenting information about the relationships between geographical and spatial locations, sizes, and ~~shapes.~~ GIS data will be created through a variety of methods including those offered by global positioning system (GPS) technologies. This course will assume knowledge of R and Python.
GIS Applications to Planning and Policy Analysis		●	●	●		●			GSU - AYS	Students will have the skills to operate the basic functions of ArcMap software, integrate data from a variety of sources, conduct basic spatial analysis, and produce quality map products.

The case for a consortium

- Schools of public affairs are not competing with each other
- Many small programs = low capacity
- First movers/high-capacity programs are less accessible
- Data science requires an agile curriculum

Consortium Goals

Identify	Faculty to share courses
Build	Open educational resources in GitHub
Share	Mentor new faculty to teach courses Cross-registration