DS4PS

Data Analytics Course Summary

DS4PS

Preparing public service professionals to deploy data to protect and benefit society.

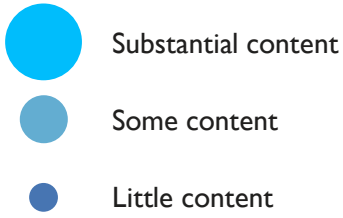
OVERVIEW

This summary reviews master's level courses offered by schools of public affairs that teach the knowledge and skills of data science. It maps curricula against the data science life cycle to identify areas of strength and where gaps remain.

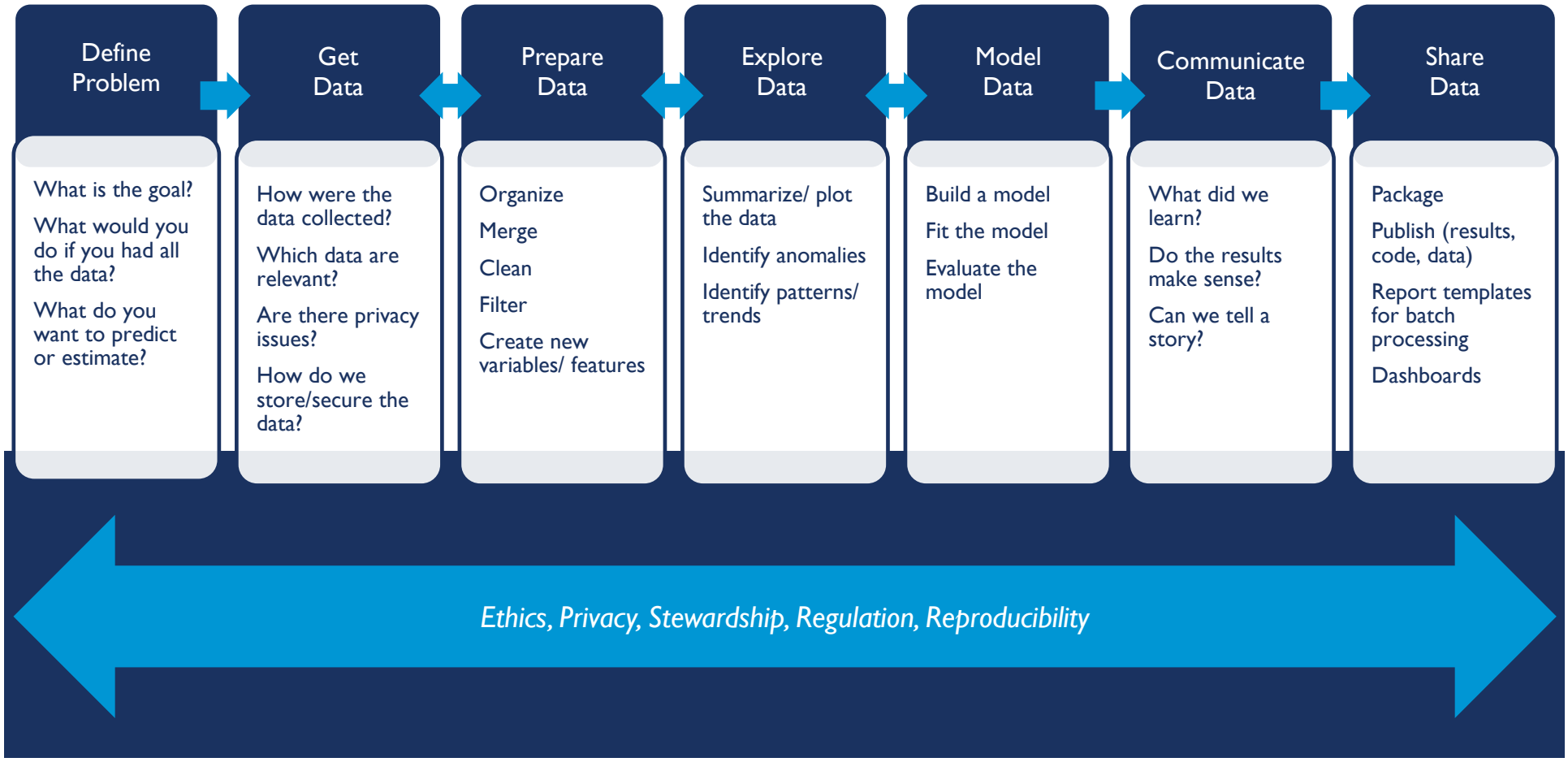
Note: The classification of course content is based on course descriptions from program websites, graduate catalogs, or course syllabi. It has not been reviewed by instructors who teach the courses to verify the classifications of content.

This summary **does not** include statistics/program evaluation courses required in most MPA/MPP programs. It also excludes courses that are offered by other disciplines that may be core to data science and are available as electives to students in public affairs programs.

KEY



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

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Foundations of Data Science I: Intro. to Data Programming in R		●	●	●		●	●		ASU - Watts GSU - AYS	Mastery of functions and arguments as the building blocks of R; Knowledge of variable types and data structures in R, including construction and manipulation; Use of logical statements to create and analyze groups within data; Ability to build custom visualizations through the base R graphics package; Creation of dynamic graphics and data dashboards using R shiny tools
Computer Programming & Information Systems		●	●	●					CMU - Heinz	Internet of Things Data Structures and Algorithms Data Warehousing Big Data and Large Scale Computing
Data Science I: Foundations		●	●	●			●		Georgetown - McCourt	Introduction to the fundamentals of programming in Python including data structures, data manipulation, and basic data management. Students work in Jupyter notebooks and use GitHub to submit coding assignments, developing literate programming and reproducible research.
Coding for Civic Data Applications		●	●	●		●			UIC	Advanced technical skills for coding, transforming, and displaying data within existing commercial platforms. Emphasis is placed on the use of and R and Python, SQL, data scraping, mining, etc.

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

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Machine Learning for Public Policy									UC - Harris	This course provides an introduction to machine learning and how it can be applied to public policy problems. It is intended for students who are interested in learning how to use modern, scalable, computational data analysis methods and tools for social impact and policy problems.
Data Visualization: Telling Stories with Numbers									Minnesota - Humphrey	The course is designed to be like a walk-through, starting from collecting and organizing data and ending with advanced graph designs.
Foundations of Data Science II: Data Wrangling									ASU - Watts	This course introduces how to use custom functions to make analysis more efficient, build simulations and animations, create R packages, learn text analysis functions in R, and build a website using the Jekyll framework in GitHub. Import data from multiple standard and nonstandard formats and data APIs. Apply string processing and regular expressions to clean data and analyze text. Generate new variables through transformations and group aggregation techniques. Create an R package; Build a website using the Jekyll framework in GitHub
Data Analytics Practicum									ASU - Watts	This course develops practical experience in building and analyzing custom research databases. You will receive a collection of raw data that needs to be linked together to conduct analysis, and guidance on how to implement a specific research question. You will merge the data, wrangle it into the proper format for analysis, apply the prescribed regression techniques, interpret results and create a report with key findings.
Data Mining and Machine Learning									CMU - Heinz	Large Scale Data Analysis for Public Policy Business Intelligence and Data Mining with SAS Unstructured Data Analytics for Policy Data Mining Techniques

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Data Visualization									GWU - T	How graphics can be used to obfuscate, illuminate, and compel. Focusing on the programming language R, includes working with large-scale data and distilling such data into pictures that communicate.
Using Big Data to Develop Public Policy									Columbia - SIPA	Data are a critical resource for understanding and solving public policy challenges. This course provides an applied understanding of data analytics tools and approaches to policy. This course is designed to bridge the gap between the statistical theory and real-world challenges of using data in public policy. The course leverages the DATA2GO.NYC data set. You will use the data set to conduct the in-depth analysis of an issue and ultimately develop a policy proposal or policy evaluation.
Intro to Infographics & Data Viz									Columbia - SIPA	This is a seven-week course that introduces students to design principles and techniques for effective data visualization. Visualizations graphically depict data to foster communication, improve comprehension and enhance decision-making. This course aims to help students: understand how visual representations can improve data comprehension, master techniques to facilitate the creation of visualizations as well as begin using widely available software and web-based, open-source frameworks.
Data Viz with R									GSU - AYS	This course shows how to use R, ggplot2 and the principles of graphic design to create beautiful and truthful visualizations of data.
Big Data for Public Good									ASU -Watts GSU - AYS	This course provides students with exposure to the opportunities and limitations of data science and its applications to public problems. It explores best practices for using existing data assets of public organizations and innovative approaches to application development while ensuring commitment to the public values of access, equity, and sustainability.

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Big Data and Public Policy									GA Tech - Ivan Allen	This course provides an introduction to "big data" for social science applications. Students learn to conduct experiments and to identify causal mechanisms in large-scale social and administrative data. Objectives: Identify compelling uses of big data to solve social and economic problems; Distinguish between research designs for prediction versus inference; Experiment with cloud technologies for storage, analysis and computation; Develop case studies related to big data and sustainable communities; Gain experience in presenting and defending research; Consider the protection of human subjects in commissioned studies.
Intro to Data Science									Georgetown - McCourt	This course teaches students how to synthesize disparate, possibly unstructured data to better understand and characterize the world around us and to draw meaningful inferences from data. Topics covered include fundamentals of programming, data wrangling and probing, data extraction (via web scraping and APIs), data visualization, and data ethics. The course will survey commonly used data science approaches, such as text analysis, machine learning, network, and geospatial analysis. Students will work on a policy-focused data science project that starts with question formulation and data collection and goes through all stages of the life cycle, culminating in data storytelling.
Data Visualization									Georgetown - McCourt	This course introduces students to the tools, methods, and skills necessary for making compelling quantitative graphics for the analysis and communication of public policy research. Students will be trained in programming and software applications useful for data visualization including R, Python, and Tableau. Includes: A foundation in the theoretical, practical, and aesthetic elements of data visualization. Skills in multiple software tools for making both static and dynamic visualizations. An understanding of different data visualization techniques, trends, and challenges they are likely to encounter.
Data Science II: Applied Statistical Learning									Georgetown - McCourt	This second course in the core data science sequence offers students an applied understanding of three key data science skills: data collection, data wrangling, and machine/statistical learning. Students will learn to gather raw data (using web scraping techniques and APIs); clean, structure, and manipulate data in a variety of formats; effectively explore and visualize data; and analyze datasets using a variety of machine learning models including regression, naive Bayes, K-nearest neighbors, decision trees and random forests, and support vector machines. Throughout the course, emphasis will be placed on effective visualization, model refinement and validation, and ethics. focused data science project.

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Data Science III: Advanced Modeling Techniques									Georgetown - McCourt	This final course in the core data science sequence focuses on unsupervised learning techniques, natural language processing, and network analysis. The course builds off of the modeling concepts covered in Data Science II by teaching students how to effectively explore, model, and predict with unstructured data, such as text or data streams like those encountered in social media. Students will engage with a number of policy-relevant data case studies throughout the course and will work on a policy-focused data science project in which to apply their statistical learning toolkit.
Massive Data Fundamentals									Georgetown - McCourt	In this course, students will learn the technology, business, science, and social implications of "big data" processing. In recent years there has been an explosion of tools, techniques, and technologies for working with massive data sets. Students will build real word systems, using stand-alone Hadoop/Spark environments running in VirtualBox on personal systems, and scalable clusters on Amazon Web Services. Topics: Big Data terminology, scaling from one computer to thousands, data storage and data privacy, Spark, data formats and data wrangling, text processing and web mining, streaming data, graph processing. Students will be provided Amazon Web accounts.
Data Driven Decision-Making									Syracuse Maxwell	Introduces students to using data for decision-making in the public sector. Covers possibilities and limitations of different data forms; fundamentals of data quality, measurement, and management; basics of algorithmic decision making; and introduction to Python.
Predictive Analytics									Syracuse Maxwell	This course introduces students to the field of predictive analytics with applications in the public policy field. Students will develop skills for training, testing, and validating predictive models for complex policy relationships.
Data Science Ethics and Information Security									UIC	Provides a review of the ethical considerations that arise from the use of data science and technology in the public sector, including location-based services and spatial intelligence.

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Data Management		●	●					●	UIC	An overview of the use and leveraging of data in order to allow government and nonprofit organizations to make better decisions and improve operations. Topics to be covered will include the concepts of data management and governance.
Data Analytics	●	●	●	●		●			UIC	An introduction to data analytics concepts, including the latest practices for gaining better value from data. An emphasis will be placed upon hands-on use and application of data analytics techniques.
Artificial Intelligence and Machine Learning in the Public Sector			●	●	●	●			UIC	Designed to provide students with the advanced technical skills for coding, transforming, and displaying data within existing commercial platforms. Emphasis is placed on the use of and R and Python, SQL, data scraping, mining, etc.
Database Management Systems		●	●					●	Indiana - O'Neill	Provides students an in-depth knowledge of database design and management in public organizations. The students create a conceptual, logical, and physical design of databases; build models of data required by users with modeling formalisms and computer-aided software engineering tools; and design queries using leading database software packages.

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.

Course	Define Problem	Gather Data	Prepare Data	Explore Data	Model Data	Communicate Data	Share Data	Ethics & Stewardship	Institution	Objectives/Description
Text as Data									Columbia - SIPA	This course is an introduction to the quantitative analysis of text as data a rapidly growing field within the social sciences. Throughout the course, students will learn different methods to acquire text, how to transform it to data, and how to analyze it to shed light on important research questions. Each week we will cover different methods, including dictionary construction and application, sentiment analysis, scaling and topic models, and machine learning classification of text. Lectures will be accompanied by hands-on exercises that will give students practical experience while working with real-world texts. By the end of the course, students will develop and write their own research projects using text as data