

Online Certificate Course

“DATA ANALYSIS AND VISUALISATION FOR SOCIAL SCIENCE AND POLICY RESEARCH”

This four week course develops foundational and applied skills in analysing, visualising, and communicating quantitative and textual data for social science and policy research, from formulating research questions to data management, descriptive and basic inferential analysis, visualisation, spatial communication, and NLP based text mining workflows. Using real Indian datasets and public data resources such as Census, PLFS, NFHS, NCRB, UDISE+, MoSPI and RBI, participants learn to clean and manage data, build descriptive measures, and communicate evidence through charts and maps, with a central emphasis on critical, responsible data literacy, examining what questions are asked, which measures are used, and how analytical and visual choices shape policy interpretation

 15 Sep – 15th Oct 2026 |  6:00 PM – 8:00 PM (IST)

INSTRUCTORS



Dr. Shamindra Nath Roy

ASSISTANT DIRECTOR & ASSISTANT
PROFESSOR, IDEAS, JGU



Dr. Pooja Sarin

ASSISTANT DIRECTOR & ASSISTANT
PROFESSOR, IDEAS, JGU

Register Now
Scan the QR Code

Total Sessions: 8 (Tuesday and Thursday), Approx. 16 Contact Hours.

No specific academic qualifications or prior disciplinary background are required

Participants who complete all evaluation components and maintain the required attendance will receive an official Certificate of Completion issued by the Office of Interdisciplinary Studies (IDEAS), O.P. Jindal Global University.



COURSE STRUCTURE

LEARNING OBJECTIVES:

- Translate social science and policy questions into measurable indicators and analytical tasks.
- Identify appropriate datasets, variables, units of analysis, and relevant public data sources.
- Read and critically interpret codebooks, metadata, sampling information, and data structures.
- Manage, document, clean, recode, reshape, and explore quantitative datasets.
- Generate and interpret descriptive statistics, frequency distributions, ratios, cross-tabulations and subgroup comparisons.
- Conduct and interpret introductory statistical comparisons, including t-tests, correlations and basic regression outputs.
- Distinguish statistical association from causal inference and statistical significance from substantive significance.
- Evaluate data quality, representativeness, missingness, measurement choices, and methodological limitations.
- Select appropriate charts and maps, recognise misleading visualisations, and communicate findings through clear evidence narratives.
- Understand introductory principles of spatial data visualisation and the implications of geographic aggregation and scale.
- Analyse qualitative and textual data using thematic and computational approaches.
- Apply selected text-mining and NLP techniques such as word frequencies, n-grams, TF-IDF, sentiment analysis and introductory topic exploration.
- Gain introductory exposure to NVivo and dictionary-based linguistic analysis using LIWC, subject to software availability.
- Use Generative AI to support data preparation, interpretation, visualisation and research workflows while critically verifying outputs.

SESSION-BY-SESSION SCHEDULE

WEEK	SESSION & THEME	CORE TOPICS	INDICATIVE TOOLS	LEAD
WEEK 1	SESSION 1 FROM RESEARCH QUESTIONS TO DATA AND EVIDENCE	RESEARCH AND POLICY QUESTIONS; VALUES AND OBJECTIVES; VARIABLES, INDICATORS AND UNITS OF ANALYSIS; DATA-ENABLED QUESTIONS; PUBLIC DATASETS; CODEBOOKS AND METADATA; SAMPLING AND REPRESENTATIVENESS; DATA SOURCES AND RESPONSIBLE CITATION.	CENSUS / PLFS / NFHS / PUBLIC REPOSITORIES	DR. SHAMINDRA NATH ROY
WEEK 1	SESSION 2 DATA MANAGEMENT, CLEANING AND EXPLORATORY ANALYSIS	DATA STRUCTURES; TIDY DATA PRINCIPLES; DATA DOCUMENTATION; MISSING VALUES; OUTLIERS; RECODING; DATA TYPES; RESHAPING; DESCRIPTIVE STATISTICS; PIVOT TABLES; CROSS-TABULATIONS; SUBGROUP COMPARISONS; REPRODUCIBLE DATA-MANAGEMENT PRACTICES.	EXCEL / GOOGLE SHEETS	DR. SHAMINDRA NATH ROY

WEEK 2	SESSION 3 DESCRIPTIVE AND BASIC INFERENTIAL STATISTICS	MEASURES OF CENTRAL TENDENCY AND DISPERSION; FREQUENCY DISTRIBUTIONS; RATIOS AND NORMALISATION; COMPARISON OF GROUPS; INTRODUCTORY T-TESTS; ASSOCIATION AND CORRELATION; INTERPRETATION OF BASIC REGRESSION OUTPUTS; STATISTICAL VERSUS SUBSTANTIVE SIGNIFICANCE; ASSOCIATION VERSUS CAUSATION.	EXCEL / GOOGLE SHEETS; DEMONSTRATION OUTPUTS	DR. POOJA SARIN
WEEK 2	SESSION 4 DATA VISUALISATION, SPATIAL VISUALISATION AND EVIDENCE COMMUNICATION	CHART SELECTION; LINE, BAR AND SCATTER PLOTS; MAPS AND SPATIAL VARIATION; GEOGRAPHIC UNITS AND AGGREGATION; VISUAL INTEGRITY; MISLEADING SCALES; COLOUR, SHAPE AND SIZE; ACCESSIBILITY; ANNOTATIONS; DATA STORYTELLING; COMMUNICATING EVIDENCE TO POLICY AUDIENCES.	DATAWRAPPER / FLOURISH / POWER BI	DR. SHAMINDRA NATH ROY
WEEK 3	SESSION 5 QUALITATIVE AND TEXTUAL DATA ANALYSIS	INTERVIEWS, OPEN-ENDED RESPONSES, REPORTS AND POLICY DOCUMENTS; CODING; CATEGORIES AND THEMES; WORD FREQUENCIES; QUALITATIVE INTERPRETATION; LINKING QUALITATIVE EVIDENCE TO RESEARCH QUESTIONS.	NVIVO / EQUIVALENT DEMONSTRATION	DR. POOJA SARIN
WEEK 3	SESSION 6 TEXT MINING AND NATURAL LANGUAGE PROCESSING	TEXT PREPROCESSING; TOKENISATION; N-GRAMS; TF-IDF; SENTIMENT ANALYSIS; INTRODUCTORY TOPIC EXPLORATION; INTERPRETATION AND LIMITATIONS OF COMPUTATIONAL TEXT ANALYSIS.	ORANGE DATA MINING / SELECTED NLP DEMONSTRATIONS	DR. POOJA SARIN
WEEK 4	SESSION 7 LINGUISTIC ANALYSIS AND AI-ASSISTED RESEARCH	DICTIONARY-BASED LINGUISTIC ANALYSIS USING LIWC; AI-ASSISTED SPREADSHEET FORMULAS, DATA CLEANING, CODING SUPPORT, INTERPRETATION AND VISUALISATION; PROMPT DESIGN FOR ANALYTICAL TASKS; OUTPUT VERIFICATION; BIAS, TRANSPARENCY AND RESPONSIBLE AI USE.	LIWC / GENERATIVE AI / SPREADSHEETS	DR. POOJA SARIN
WEEK 4	SESSION 8 FROM DATA TO EVIDENCE: INTEGRATED CAPSTONE WORKSHOP	INTEGRATED APPLICATION OF QUANTITATIVE OR TEXTUAL ANALYSIS; DATA VISUALISATION AND/OR MAPPING; EVIDENCE NARRATIVE; METHODOLOGICAL LIMITATIONS; RESPONSIBLE INTERPRETATION; PEER DISCUSSION AND INSTRUCTOR FEEDBACK.	PARTICIPANT-SELECTED TOOLS	DR. POOJA SARIN & DR. SHAMINDRA NATH ROY

TEACHING-LEARNING APPROACH

Each session will combine conceptual discussion, demonstration, applied examples and guided hands-on exercises. The course follows a ‘data, discussion and demonstration’ approach: participants will work with numbers and text as supporting evidence, while also examining the assumptions, choices and limitations behind the analysis.

- Short conceptual inputs linked to real social-science and policy questions.
- Hands-on work with public datasets and sample datasets.
- Demonstrations of good and misleading visualisations.
- Exercises on data cleaning, descriptive analysis, statistical interpretation and mapping.
- Applied textual analysis and AI-assisted workflows.
- Incremental development of a small capstone question or dataset from the middle of the course.
- Indicative Data Sources and Repositories
- Census of India and historical Census tables.
- Periodic Labour Force Survey (PLFS) and other MoSPI microdata resources.
- National Family Health Survey (NFHS).
- India Human Development Survey (IHDS).
- NCRB Crime in India reports and UDISE+ education data.
- RBI statistical resources and national accounts data.
- data.gov.in, NDAP, India Data Portal, Data for India and other public repositories.
- Electoral and geographic datasets where relevant for spatial or policy analysis.

INDICATIVE DATA SOURCES AND REPOSITORIES

- Census of India and historical Census tables.
- Periodic Labour Force Survey (PLFS) and other MoSPI microdata resources.
- National Family Health Survey (NFHS).
- India Human Development Survey (IHDS).
- NCRB Crime in India reports and UDISE+ education data.
- RBI statistical resources and national accounts data.
- data.gov.in, NDAP, India Data Portal, Data for India and other public repositories.
- Electoral and geographic datasets where relevant for spatial or policy analysis.

TENTATIVE READING AND RESOURCE LIST

TEACHING-LEARNING APPROACH

- Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.
- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as data: A new framework for machine learning and the social sciences. Princeton University Press.
- Healy, K. (2018). Data visualization: A practical introduction. Princeton University Press.
- Knafllic, C. N. (2015). Storytelling with data: A data visualization guide for business professionals. Wiley.
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Spiegelhalter, D. (2019). The art of statistics: How to learn from data. Basic Books.
- Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.

TOOLS AND PRACTICE RESOURCES

- Datawrapper Academy.
- Flourish training and data visualisation resources.
- Orange Data Mining documentation.
- Lumivero: Getting started with NVivo.
- LIWC documentation and psychometrics resources.

Note: NVivo and LIWC are proprietary tools. Hands-on use will depend on institutional or participant access; where full access is unavailable, selected workflows may be demonstrated using sample outputs or equivalent exercises.

WHO SHOULD ATTEND:

Postgraduate students, doctoral scholars, faculty members, early-career researchers, policy professionals, government and public-sector professionals, NGO and think-tank staff, journalists, development practitioners, management professionals and interdisciplinary researchers interested in applied data analysis, computational social science, data communication, spatial analysis and AI-assisted research.

ELIGIBILITY AND PREREQUISITES:

- Basic familiarity with spreadsheets and quantitative data.
- A basic understanding of research methods is desirable.
- Prior programming experience is not mandatory.
- Participants should have an interest in research, policy, social-sector data, or evidence-based decision-making.

DATES AND SCHEDULE:

Duration: 4 weeks, with two approximately two-hour sessions per week (8 sessions; approximately 16 contact hours). Each session will combine conceptual discussion, demonstrations, applied examples and guided hands-on exercises. Participants will begin identifying a capstone question or dataset by the middle of the course and develop it incrementally. The final session will be conducted as an interactive capstone workshop focused on analysis, interpretation, visual/spatial communication, methodological limitations and feedback.

INSTRUCTORS



Dr. Pooja Sarin

ASSISTANT DIRECTOR & ASSISTANT
PROFESSOR, IDEAS, JGU

Dr. Pooja Sarin is an Assistant Professor of Data Science and Assistant Director for Institutional Partnerships and Research Collaboration at the Office of Interdisciplinary Studies (IDEAS), O.P. Jindal Global University.

Ph.D. in Information Systems and Analytics from IIT Delhi. M.Tech. in Computer Engineering from Kurukshetra University, with thesis research undertaken at CSIR-CSIO, Chandigarh, and B.Tech. in Computer Science and Engineering.

Her interdisciplinary research focuses on computational social science, data analytics, digital platforms, responsible AI, AI governance, and digital wellbeing. Her methodological expertise spans mixed-methods research, quantitative analysis, machine learning, text mining, natural language processing, data visualisation, and computational research methods.



Dr. Shamindra Nath Roy

ASSISTANT DIRECTOR & ASSISTANT
PROFESSOR, IDEAS, JGU

Shamindra Nath Roy is an Assistant Professor in the IDEAS Program in JGU. He works at the interface of urbanization and public policy. He is a mixed-methods researcher research covers rural-urban transformations, urban local governance, internal migration in India, labor and employment issues, and socio-spatial inequalities in basic services in Delhi. His work is published in Economic and Political Weekly, Transactions of the Institute of British Geographers, and in a number of book chapters, policy briefs and opinion pieces.

Prior to JGU, Shamindra worked as a development professional in the World Resources Institute and Centre for Policy Research, New Delhi. He has also been a visiting researcher at CESSMA, Université de Paris, and has undertaken research consultancies with the Asian Development Bank, International Labour Organization, and the World Bank. He has also taught as an Adjunct faculty at Dr. B.R. Ambedkar University, Delhi.

Shamindra holds an M.Phil and Ph.D in Regional Development with a specialization in geography from Jawaharlal Nehru University, New Delhi.

I.D.E.A.S. *at a* GLANCE

IDEAS OFFICE OF INTERDISCIPLINARY STUDIES



100+

Ideas Advisory
Council



80+

Research
and Publications



31+

Research
Fellow



10+

Collaborative
Projects



10+

Core
Faculty



10+

Institutional
Partnership



6+

Initiatives



6+

Member
Associates



6+

Research
Constellations

ABOUT US

The Office of Interdisciplinary Studies (IDEAS) at O.P. Jindal Global University (JGU) serves as a university-wide platform for fostering interdisciplinary teaching, research, and community engagement across the social sciences. It brings together scholars, students, and practitioners to co-create knowledge and develop innovative responses to complex real-world challenges.

At its core, IDEAS operates as an intellectual incubator, supporting collaborative scholarship, experimental pedagogy, and research that transcends traditional disciplinary boundaries. Through its initiatives and field-based programmes, it promotes critical inquiry, applied learning, and engagement with pressing development concerns.

With a strong emphasis on interdisciplinary dialogue, global academic exchange, and community-linked research, IDEAS aims to cultivate a dynamic academic environment that connects theory with practice and advances socially relevant scholarship.