



UNIVERSITY OF
MARYLAND

Syllabus

Fundamentals of Survey and Data Science

3 credits/6 ECTS

Instructor: Dr. João Areal (jarealne@umd.edu)

Teaching Assistant: Paul Burlon (paul.burlon@campus.lmu.de)

Video lecture by

Prof. Florian Keusch, Prof. Frauke Kreuter,
Jennifer Sinibaldi PhD and Prof. Richard Valliant

September 2 - December 2, 2026

Short Course Description

The course introduces the student to a set of principles of survey and data science that are the basis of standard practices in these fields. The course exposes the student to key terminology and concepts of collecting and analyzing data from surveys and other data sources to gain insights and to test hypotheses about the nature of human and social behavior and interaction. It will also present a framework that will allow the student to evaluate the influence of different error sources on the quality of data.

Course Objectives

By the end of the course, students will...

- be able to apply the key terminology used by survey methodologists and data scientists.
- be able to assess the quality of data from different sources based on a data quality framework.
- be able to select an appropriate data source to answer different types of research questions.
- understand the influence of coverage, sampling, and nonresponse on data quality and know how to deal with deficiencies of the data.
- have a clear understanding of the steps involved in data preparation, data processing, data analysis, and data visualization.
- be able to comply with ethical standards in survey research and data science.

Prerequisites

Students are expected to be familiar with basic statistical concepts, such as mean, standard deviation, variance, and distributions (at the level of an undergraduate course) and have exposure to elements of social science perspectives on human behavior.

Class Structure and Course Concept

This is an online course using a flipped classroom design. It covers the same material and content as an on-site course but runs differently. In this course, you are responsible for watching video-recorded lectures and reading the required literature for each unit prior to participating in mandatory weekly one-hour online meetings where students have the chance to discuss the materials from a unit with the instructor.

Although this is an online course where students have more freedom in when they engage with the course materials, students are expected to spend the same amount of time overall on all activities in the course - including preparatory activities (readings, studying), in-class activities (watching prerecorded videos, attending the live online meetings), and follow-up activities (working on assignments and exams) - as in an onsite course. As a rule of thumb, you can expect to spend approximately 3h/week on in-class activities and 9h/week on out-of-class activities (preparing for class, readings, assignments, projects, studying for quizzes and exams). Therefore, the workload in all courses will be approximately 12h/week. Please note that the actual workload will depend on your personal knowledge.

Mandatory Weekly Online Meetings

Wednesdays, 11:00 AM US Eastern Time, starting September 2, 2026

Meetings will be held through Zoom. The link will be posted on the course website at umd.instructure.com.

In preparation for the weekly online meetings, students are expected to watch the lecture videos and read the assigned literature before the start of the meeting. In addition, students are encouraged to use the forum to share potential points for class discussion related to the week's topic—for example, an interesting observation about the videos or readings, a relevant real-world example, or a connection to a current event or news story.

The teaching assistant will also offer office hours during the hour before each weekly online meeting (Wednesdays, 10:00 AM - 11:00 AM US Eastern Time) through Zoom. The link will be posted in Canvas.

Students may use Canvas and other communication tools to connect with peers outside the scheduled weekly online meetings (e.g., for study groups).

Grading

Grading will be based on:

- Participation in discussion during the weekly online meetings and contributions to the weekly discussion forum (deadline is Tuesday, 11:00 AM US Eastern Time), demonstrating understanding of the required readings and video lectures, is worth 10% of the final grade.
- Ten online quizzes, worth 100 points each, reviewing specific aspects of the material covered in the pre-recorded lectures. The simple average of the points across all quizzes is worth 60% of the final grade.
- A final open-book online exam, worth 100 points, is worth 30% of the final grade.

Grading Scale:

A+	100 - 97
A	96 - 93
A-	92 - 90
B+	89-87
B	86-83
B-	82-80 Etc.

The final grade will be communicated under the assignment "Final Grade" in the Canvas course. Please note that the letter grade written in parentheses in Canvas is the correct final grade. The point grade displayed alongside the letter grade is irrelevant and can be ignored.

Dates of when assignments will be due are indicated in the syllabus. Quizzes and assignments will automatically close on the course website at the indicated time. Assignments can be completed at any time leading up to the deadline. Extensions will be granted sparingly and are at the instructor's discretion.

The final exam is open book and open notes, and students are on the honor system (see Academic Conduct below). **The exam will take place on Wednesday, December 2, 2026 at 11:00 AM EST/5:00 PM CET. Students have two days to complete the exam and must submit their exam before Friday, December 4, 2026 1:00 PM EST/7:00 PM CET.**

Technical Equipment Needs

The learning experience in this course will mainly rely on online interaction between the students and the instructor during the weekly online meetings. Students need a reliable Internet connection and a device capable of playing the lecture videos, accessing Canvas, and joining Zoom with working audio. A headset and web camera are recommended when they improve audio and video quality. Students who anticipate technical difficulties should contact the course team as early as possible.

Long Course Description

The fields of survey methodology and data science draw on theories and practices developed in several academic disciplines - mathematics, statistics, psychology, sociology, computer science, and economics. To become an accomplished professional in these fields requires a mastery of research literatures as well as experience designing, conducting, and analyzing surveys and data from other sources, such as administrative records, social media, or transactions.

This course introduces the student to a set of principles of survey design and data science that are the basis of standard practices in these fields. The course exposes the student to research literatures that use both observational and experimental methods to test key hypotheses about the nature of human behavior and other factors that affect the quality of data. It will also present important statistical concepts and techniques in sample design, execution, and estimation, as well as models of behavior describing errors in responding to survey questions. Thus, both social science and statistical concepts will be presented.

The course uses the concept of total error as a framework to discuss coverage properties of sampling frames and organic data, alternative sample designs and their impacts on standard errors of statistics, different modes of data collection and generation, the role of interviewers and respondents in surveys, impacts of nonresponse and missing data on statistics, measurement errors in data, data processing, and data/research ethics.

The course is intended as an introduction to the fields of survey methodology and data science, taught at a graduate level. Lectures and course readings assume that students understand basic statistical concepts (at the level of an undergraduate course) and have exposure to elements of social science perspectives on human behavior. For those lacking such a background, supplementary readings are recommended.

Readings

The primary textbook and additional freely available textbook are:

Groves, R.M., Fowler, F.J. Jr., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology*, 2nd Edition. New York: Wiley. (available at local bookstores and online retailers)

Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F., & Lane, J. (Eds.). (2021). *Big Data and Social Science: Data Science Methods and Tools for Research and Practice*, 2nd edition. Chapman and Hall/CRC. (available online at <https://textbook.coleridgeinitiative.org/>).

Required and Recommended Readings

Lists of required and recommended readings are provided below for each unit.

The weekly online quizzes assess material covered in the pre-recorded lectures. The final exam may also include material from the required readings. Recommended readings will not be assessed.

Academic Conduct

Clear definitions of the forms of academic misconduct, including cheating and plagiarism, as well as information about disciplinary sanctions for academic misconduct may be found in the [University of Maryland Code of Academic Integrity](#).

Knowledge of these rules is the responsibility of the student and ignorance of them does not excuse misconduct. The student is expected to be familiar with these guidelines before submitting any written work or taking any exams in this course. Lack of familiarity with these rules in no way constitutes an excuse for acts of misconduct. Charges of plagiarism and other forms of academic misconduct will be dealt with very seriously and may result in oral or written reprimands, a lower or failing grade on the assignment, a lower or failing grade for the course, suspension, and/or, in some cases, expulsion from the university.

Accommodations for Students with Disabilities

In order to receive services, students at the University of Maryland must contact the Accessibility & Disability Service (ADS) office and register for services. Please contact ADS to arrange an appointment. More information is available at <https://counseling.umd.edu/ads>.

Course Evaluation

In an effort to improve the learning experience for students in our online courses, students will be invited to participate in an online course evaluation at the end of the course. Participation is voluntary but highly appreciated.

Sessions

Please note that assignments and dates are subject to change. Information posted in Canvas supersedes the information noted here.

Unit 1: Introduction

Video lecture: available Wednesday, August 26, 2026

Online meeting: Wednesday, September 2, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 1: due Thursday, September 3, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Groves et al. (2009). Chapters 1.4 and 1.5

Leek, J.T. and Peng, R.D. (2015). What is the question? *Science*, 347, 1314-1315.

AAPOR (2026). Responsible AI Integration in Survey Research, executive summary, pp. 4-7. <https://aapor.org/wp-content/uploads/2026/05/Responsible-AI-Integration-In-Survey-Research.pdf>

Recommended Readings:

Peng & Matsui (2015). Chapters 1-3 <https://leanpub.com/artofdatascience>

Sturgis, P., & Luff, R. (2020). The demise of the survey? A research note on trends in the use of survey data in the social sciences, 1939 to 2015. *International Journal of Social Research Methodology*, 24(6), 691-696.

<https://doi.org/10.1080/13645579.2020.1844896>

Groves, R. (2011). Three eras of survey research. *Public Opinion Quarterly*, 75(5), 861-871. <https://doi.org/10.1093/poq/nfr057>

Unit 2: Data Quality

Video lecture: available Wednesday, September 2, 2026

Online meeting: Wednesday, September 9, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 2: due Thursday, September 10, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Groves et al. (2009). Chapter 2

Biemer, P. (2010). Total Survey Error. Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817-848.
Sen, I., Flöck, F., Weller, K., Weiß, B., & Wagner, C. (2021). A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. *Public Opinion Quarterly*, 85(S1), 399-422. <https://doi.org/10.1093/poq/nfab018>

Recommended Readings:

Lavrakas, P. (2013). Presidential Address: Applying a Total Error Perspective for Improving Research Quality in the Social, Behavioral, and Marketing Sciences. *Public Opinion Quarterly*, 77(3), 831-850.

Keusch, F., & Kreuter, F. (2021). Digital trace data. In *Handbook of Computational Social Science*, Volume 1, 100-118. <https://doi.org/10.4324/9781003024583-8>

Unit 3: Coverage

Video lecture: available Wednesday, September 9, 2026

Online meeting: Wednesday, September 16, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 3: due Thursday, September 17, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Groves et al. (2009). Chapter 3

Hargittai, E. (2015). Is bigger always better? Potential biases of Big Data derived from social network sites. *The Annals of the American Academy of Political and Social Science*, 659, 63-76.

Bosch, O. J., Sturgis, P., Kuha, J., & Revilla, M. (2024). Uncovering Digital Trace Data Biases: Tracking Undercoverage in Web Tracking Data. *Communication Methods and Measures*. <https://doi.org/10.1080/19312458.2024.2393165>

Recommended Readings:

Kreuter, F. and Peng, R.D. (2014). Extracting information from Big Data: Issues of measurement, inference and linkage. In *Privacy, Big Data, and the Public Good*, 257-275.

National Academies (2023). Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources, interactive overview. <https://nap.nationalacademies.org/resource/26688/interactive/>

Unit 4: Modes of Data Collection

Video lecture: available Wednesday, September 16, 2026

Online meeting: Wednesday, September 23, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 4: due Thursday, September 24, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Groves et al. (2009). Chapters 1.3 and 5

Wuttke, A. et al. (2025). AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers. Proceedings of LaTeCH-CLFL 2025, pp. 179-187 (Sections 1-6). <https://doi.org/10.18653/v1/2025.latechclfl-1.17>

Recommended Readings:

Couper, M.P., Antoun, C., & Mavletova, A. (2017). Mobile Web Surveys. In Total Survey Error in Practice, 133-154.

Blom, A.G. et al. (2016). A comparison of four probability-based online and mixed-mode panels in Europe. Social Science Computer Review, 34(1), 8-25. <https://doi.org/10.1177/0894439315574825>

de Leeuw, E.D. (2018). Mixed-Mode: Past, present, and future. Survey Research Methods, 12(2), 75-89.

<https://doi.org/10.18148/srm/2018.v12i2.7402>

Unit 5: Other Data Sources

Video lecture: available Wednesday, September 23, 2026

Online meeting: Wednesday, September 30, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 5: due Thursday, October 1, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Hölzl, J., Keusch, F., & Sajons, C. (2025). The (mis)use of Google Trends data in the social sciences - A systematic review, critique, and recommendations. Social Science Research, 126, 103099. <https://doi.org/10.1016/j.ssresearch.2024.103099>

Stier, S., Breuer, J., Siegers, P. & Thorson, K. (2020). Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field. Social Science Computer Review, 38(5), 503-516. <https://doi.org/10.1177/0894439319843669>

Ohme, J., Araujo, T., de Vreese, C.H., et al. (2024). Digital Trace Data Collection for Social Media Effects Research: APIs, Data Donation, and (Screen) Tracking. *Communication Methods and Measures*, 18(2), 124-141. <https://doi.org/10.1080/19312458.2023.2181319>

Recommended Readings:

boyd, d. & Crawford, K. (2012). Critical questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.

Unit 6: Sampling I

Video lecture: available Wednesday, September 30, 2026

Online meeting: Wednesday, October 7, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 6: due Thursday, October 8, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

Groves et al. (2009). Chapters 4.1-4.6

Unit 7: Sampling II

Video lecture: available Wednesday, October 7, 2026

Online meeting: Wednesday, October 14, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 7: due Thursday, October 15, 2026, 5:59 PM EDT/11:59 PM CEST

Required Readings:

AAPOR (2022). Data Quality Metrics for Online Samples: Considerations for Study Design and Analysis, pp. 2-13 and 47-52.

<https://aapor.org/wp-content/uploads/2023/02/Task-Force-Report-FINAL.pdf>

Rohr, B., Felderer, B., Silber, H., Daikeler, J., Roßmann, J., & Schröder, J. (2024). When are non-probability surveys fit for my purpose? GESIS Survey Guidelines. https://doi.org/10.15465/gesis-sg_en_050

Recommended Readings:

Cornesse, C. et al. (2020). A review of conceptual approaches and empirical evidence on probability and nonprobability sample survey research. *Journal of Survey Statistics and Methodology*, 8(1), 4-36. <https://doi.org/10.1093/jssam/smz041>

Unit 8: Questionnaires

Video lecture: available Wednesday, October 14, 2026

Online meeting: Wednesday, October 21, 2026, 11:00 AM EDT/5:00 PM CEST

Online quiz unit 8: due Thursday, October 22, 2026, 5:59 PM EDT/11:59 PM CEST

Daylight saving time ends in Europe on October 25, 2026, and clocks are turned back 1 hour.

Required Readings:

Groves et al. (2009). Chapters 8.1-8.8 and 9.1-9.7

AAPOR (2026). Responsible AI Integration in Survey Research, AI as a question writer, editor, and translator, pp. 25-27. <https://aapor.org/wp-content/uploads/2026/05/Responsible-AI-Integration-In-Survey-Research.pdf>

Recommended Readings:

West, B.T. & Blom, A. (2017). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 5(2), 175-211.

Olivos, F., & Liu, M. (2025). ChatGPTTest: Opportunities and Cautionary Tales of Utilizing AI for Questionnaire Pretesting. *Field Methods*, 37(4). <https://doi.org/10.1177/1525822X241280574>

Unit 9: Nonresponse

Video lecture: available Wednesday, October 21, 2026

Online meeting: Wednesday, October 28, 2026, 11:00 AM EDT/4:00 PM CET

Online quiz unit 9: due Thursday, October 29, 2026, 5:59 PM EDT/10:59 PM CET

Daylight saving time ends in the US on November 1, 2026, and clocks are turned back 1 hour.

Required Readings:

Groves et al. (2009). Chapters 6 and 10.6

Pew Research Center (2026). Q&A: Do AI and bogus respondents threaten polling's future? <https://www.pewresearch.org/short-reads/2026/05/12/qa-do-ai-and-bogus-respondents-threaten-pollings-future/>

Recommended Readings:

AAPOR (2023). *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 10th edition.

<https://aapor.org/wp-content/uploads/2024/03/Standards-Definitions-10th-edition.pdf>

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29(3), 329-353. <https://doi.org/10.2478/jos-2013-0026>

Unit 10: Data Processing

Video lecture: available Wednesday, October 28, 2026

Online meeting: Wednesday, November 4, 2026, 11:00 AM EST/5:00 PM CET

Online quiz unit 10: due Thursday, November 5, 2026, 5:59 PM EST/11:59 PM CET

Required Readings:

Groves et al. (2009). Chapters 10.1-10.5

Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F., & Lane, J. (Eds.). (2021). Big Data and Social Science, 2nd edition. Chapter 4: Databases. <https://textbook.coleridgeinitiative.org/chap-db.html>

AAPOR (2026). Responsible AI Integration in Survey Research, AI as transcriber, translator, cleaner, and labeler, pp. 17-22. <https://aapor.org/wp-content/uploads/2026/05/Responsible-AI-Integration-In-Survey-Research.pdf>

Unit 11: Analysis and Visualization

Video lecture: available Wednesday, November 4, 2026

Online meeting: Wednesday, November 11, 2026, 11:00 AM EST/5:00 PM CET

Required Readings:

Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F., & Lane, J. (Eds.). (2021). Big Data and Social Science, 2nd edition. Chapter 6: Information Visualization. <https://textbook.coleridgeinitiative.org/chap-viz.html>

Mellon, J. et al. (2024). Do AIs know what the most important issue is? Using language models to code open-text social survey responses at scale. Research & Politics, 11(1). <https://doi.org/10.1177/20531680241231468>

Recommended Readings:

Bisbee, J. et al. (2024). Synthetic Replacements for Human Survey Data? The Perils of Large Language Models. Political Analysis, 32(4), 401-416. <https://doi.org/10.1017/pan.2024.5>

Unit 12: Ethics

Video lecture: available Wednesday, November 11, 2026

Online meeting: Wednesday, November 18, 2026, 11:00 AM EST/5:00 PM CET

*** Thanksgiving Break, November 25 - November 29, 2026 ***

Required Readings:

Groves et al. (2009). Chapter 11

Barocas, S. & Nissenbaum, H. (2014). Big Data's end run around anonymity and consent. In Privacy, Big Data, and the Public Good: Frameworks for Engagement, 44-75.

AAPOR (2025). Ethical Issues in Linking Non-Survey Data to Survey Data: A Practical Guide. <https://aapor.org/wp-content/uploads/2025/07/25-Ethical-Issues-With-Data-Linkage-FINAL.pdf>

Recommended Readings:

AAPOR (2026). Responsible AI Integration in Survey Research, responsibility to human subjects, pp. 62-67.

<https://aapor.org/wp-content/uploads/2026/05/Responsible-AI-Integration-In-Survey-Research.pdf>

Final Exam

Wednesday, December 2, 2026, 11:00 AM EST/5:00 PM CET