



Stefan Müller, PhD

Associate Professor

School of Politics and International Relations

University College Dublin

Belfield, Dublin 4, Ireland

✉ stefan.mueller@ucd.ie

🌐 <https://stefanmueller.ie>

Level 4 Module; Autumn Trimester 2026

Applied Data Wrangling and Visualisation (POL42540)

Version: September 1, 2026

Latest version at: <https://stefanmueller.ie/teaching/2026-autumn-adwv.pdf>

Time: Thu, 09:00–10:50

Rooms: differ by week (see Course Structure)

Credits: 5.0

Format: Lecture and computer labs

Drop-In Sessions: Thu, 14:00 (Newman, F301)

Module Coordinator: Stefan Müller, PhD

stefan.mueller@ucd.ie | <https://stefanmueller.ie>

Office: Newman, G312

Office hours: Thu, 11:30–12:30 ([sign up here](#))

TA: Heba Elsahn | heba.elsahn@ucdconnect.ie

Course Content

Welcome to Applied Data Wrangling and Visualisation! The module offers an introduction to the essential techniques and tools required for effective data management and visualisation in R. This is a hands-on module: rather than working through toy examples, you will get your hands dirty with real-world datasets in every session, wrangling the same kind of messy, inconsistent data you will encounter outside the classroom. Students will also learn how to use AI tools as coding assistants, manage projects and handle data in various file formats, building a solid understanding of data cleaning, wrangling and merging.

The course covers the fundamentals of data visualisation, moving from principles to applied strategies for compelling data storytelling, including interactive visualisation techniques that let audiences explore data directly. It also covers data collection through web scraping and methods for merging data from multiple sources, so students can manage and analyse large datasets efficiently. Because every technique is practised on real-world data rather than simplified textbook examples, Applied Data Wrangling and Visualisation leaves you ready to hit the ground running on real-world projects, giving you the skills to use data visualisation and management tools effectively in a range of data-intensive roles.

Module Instructors

This module is coordinated and taught by Stefan Müller.

Heba Elsahn will serve as a Teaching Assistant for this module. Heba is the first point of contact for you. The communication in this module will take place on Slack, usually in channels that everyone can contribute to. More details on Slack are provided below.

Course Structure

Week 1: Introduction to R and Positron (10 September; Room L0.04-LEA)	11
Week 2: Central Terms and Concepts (17 September; Room L0.04-LEA)	11
Week 3: The Hidden Curriculum of Coding: Projects, Paths, and Folder Structure (24 September; Room L0.04-LEA)	11
Week 4: Summarising and Transforming Data (1 October; Room G-08.AG)	11
Week 5: Principles of Data Visualisation (8 October; Room G-08.AG)	12
Week 6: Explorative Data Visualisation (15 October; Room L2.17-LEA)	12
Week 7: Advanced, Intuitive, and Accessible Data Visualisation (22 October; Room G-08.AG)	12
Week 8: No Class – Reading Break (29 October)	13
Week 9: Maps and Interactive Data Visualisation (5 November; Room L0.04-LEA)	13
Week 10: Relational Databases (12 November; Room G-08.AG)	13
Week 11: Introduction to Web Scraping (19 November; Room L0.04-LEA)	13
Week 12: Advanced Web Scraping (26 November; Room G-08.AG)	13

Learning Outcomes

Upon successful completion of the course, students will be able to:

1. Manage and visualise data effectively in R, using Positron and GitHub, improving your ability to handle data in diverse file formats.
2. Create interactive data visualisations and merge data from multiple sources, allowing you to combine and analyse large datasets for a range of data-intensive roles.
3. Develop skills in using AI tools as coding assistants, improving your ability to manage projects efficiently and to make your work more replicable.
4. Build a solid understanding of data cleaning, wrangling and merging techniques, preparing you for the challenges of managing large and complex datasets.
5. Gain a solid foundation in the principles of data visualisation, and learn to apply these strategies to create compelling data stories that are both intuitive and accessible.

General Readings

The seminar does not build on a single textbook, but relies on papers and book chapters. All readings used in this module are freely available online or accessible through the [UCD Library](#).

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O'Reilly. URL: <https://r4ds.hadley.nz>.

- N. B. Weidmann (2023). *Data Management for Social Sciences: From Files to Databases*. Cambridge: Cambridge University Press. URL: <https://cambridge.org/9781108845670>.
- K. Healy (2019). *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. URL: <http://socviz.co>.
- R. Alexander (2023). *Telling Stories with Data: With Applications in R*. New York: CRC Press. URL: <https://tellingstorieswithdata.com>.
- C. O. Wilke (2019). *Fundamentals of Data Visualization: A Primer On Making Informative and Compelling Figures*. Sebastopol: O'Reilly. URL: <https://clauswilke.com/dataviz/>.
- C. Ismay and A. Y. Kim (2025). *Statistical Inference via Data Science: A Modern Dive into R and the tidyverse*. 2nd edition. Boca Raton: CRC Press. URL: <https://moderndive.com/v2/>.
- T. Monroe-White, D. Kozlowski, L. Domenech, C. Pradier, G. Rosati, and N. S. Shokida (2024). *Critical Computational Social Sciences*. URL: <https://critcss.github.io/CCSS/>.
- S. Dattani (2025). *Saloni's Guide to Data Visualization: Why Data Visualization Matters, and How to Make Charts More Effective, Clear, Transparent, and Sometimes, Beautiful*. URL: <https://www.scientificdiscovery.dev/p/salonis-guide-to-data-visualization>.
- D. Navarro (2026). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners*. URL: <https://learningstatisticswithr.com>.

Plagiarism

Although this should be obvious, plagiarism – copying someone else's text without acknowledgement or beyond 'fair use' quantities – is not allowed. Plagiarism is an issue we take very seriously here in UCD. Please familiarise yourself with the definition of plagiarism on UCD's website¹ and make sure not to engage in it.

Use of Generative Artificial Intelligence (AI) Tools

Artificial intelligence can speed up several tasks and is being used increasingly in universities and industry. Allowing AI use while raising expectations accordingly is, in my view, a good way to preserve the value of higher education in the face of AI.

You may use the AI tools listed above for the following aspects of your assignment:

- Conceptualisation: brainstorming topics, outlining, and planning study schedules
- Initial literature screening
- Code review, optimisation, and debugging
- Supporting your writing and editing (but no text generation)

You must always disclose your use of generative AI. To help you draft this declaration, please use the free, open-source AI Disclosure Generator at <https://disclose-ai.app>. You do not need to provide the exact prompts used for each task, but you must explain which tool and version you used, what you asked it to do, and how you validated the output.

For full transparency, every group member must complete their individual disclosure statement and add it at the end of your group assignment.

¹<https://libguides.ucd.ie/academicintegrity>.

Assignments submitted without a declaration from each member will not be accepted.

For more information on the AI tools supported by UCD, see:

- <https://www.ucd.ie/itservices/ourservices/communicationcollaboration/officeproductivity/aiservices/>
- <https://www.ucd.ie/teaching/resources/generativeai/>

Late Submission Policy: Additional Considerations for Assessment

From the 2026–27 academic year, UCD has introduced a new unified policy replacing the previous Extenuating Circumstances and Late Submission of Coursework policies. The policy offers three pathways for students facing difficulties that affect their assessments. Full details and up-to-date information are available at <https://www.ucd.ie/registry/projects/adcons/>.

1. **Self-Declare for Extension:** Students can request up to five working days' extension on eligible assessments (assignments, portfolios, reflective assignments, reports, and individual projects) via SISWeb before the submission deadline, without requiring documentation or academic approval. This may be extended by a further five days, subject to a one-grade penalty. Note: under this policy, self-declaration of extensions does **not** apply to the in-class MCQ assignments and does **not** apply to the group assignment.
2. **Certified Additional Considerations:** For longer-term or more serious circumstances (such as bereavement, health issues, or caregiving responsibilities), students submit supporting documentation to their Governing Board, which may approve a local solution, extension, IX grade, and/or WN grade. Information on these grades are available at: <https://www.ucd.ie/students/exams/gradingandremediation/understandinggrades/>.
3. **Retrospective Applications:** For cases where students were unable to apply on time due to exceptional circumstances, applications are reviewed case-by-case against clear criteria.

Release of Feedback and Grades

According to UCD's Academic Regulations, feedback on assessment must be provided no later than *20 working days* (roughly four weeks) after the deadline for submission of each piece of assessed work, except work submitted late or submitted as part of the final assessment component of the module. I will, of course, stick to these regulations.

Questions and Problems

In this module, we will discuss concepts, methods, and software you might not have heard of before. I am aware that parts of this module could be challenging, and I will assist you as best as I can.

We will use Slack in this module.² Please create a Slack account before the first seminar and join the Slack workspace. If you have a question that involve code or concepts, please share your question in the relevant channel.

If you struggle to solve problems, please follow the steps outlined below before contacting your peers or me. It is very likely that at least one other person faced the same problem before or received the same error message.

²I have had very positive experiences with Slack in my modules. Müller (2023) discusses both the advantages and shortcomings of Slack for teaching and learning.

1. You are welcome and encouraged to use AI tools as coding assistants.
2. Try to summarise the problem in your own words and then google this summary or use an AI tool, for example Microsoft Copilot, Claude Code, or ChatGPT. If the problem relates to R, add `rstats` to your search query. For example: `how to import csv file in rstats`. I am almost certain that you will find a solution to most of your questions.
3. If your R code returns an error, I would advise you to google the text of the error message. For example, you can google the error message `"Error: Can't subset columns that don't exist."`

→ If steps 1–3 still do not solve your problem or question, please ask your question in the Slack channel devoted to this module. Your peers and we will help you.

Contact and Office Hours

Heba Elsaahn, our Teaching Assistant for this module, is your first point of contact. If you have questions about software, code, or assignments, please use Slack first (see more details below) or check if one of your peers already asked this question. We monitor Slack regularly and will respond to your question if other module participants cannot help you.

In addition, Stefan Müller also offers office hours, either in person (Room G312, Newman Building) or online. Please sign up for a meeting at <https://calendly.com/mueller-ucd/office-hours>.

Proper communication practices are crucial for formal email exchanges. Please familiarise yourself with [this summary](#)³ on professional, efficient, and respectful communication.

Software

In this module, we will run R code in the code editor [Positron](#). We will also use [GitHub Copilot](#) (free to use for students).

Important: Make sure to install and set up R, Positron, and GitHub Copilot *before* our first lecture in Week 1. If you have any question related to (installing) the software, please ask this question in our Slack workspace.

Syllabus Modification Rights

I reserve the right to reasonably alter the elements of the syllabus at any time by adjusting the reading list to keep pace with the course schedule. Moreover, I may change the content of specific sessions, depending on the participants' prior knowledge and research interests. If I make adjustments, I will email all seminar participants and upload the revised syllabus to Brightspace.

Dignity and Respect

UCD is committed to the promotion of an environment for work and study which upholds the dignity and respect of all members of the UCD community and which supports your right to study and/or work in an environment which is free of any form of bullying, harassment or sexual misconduct (including sexual harassment and sexual violence).

³<https://www.beyondbera.org/blog/2022/05/03/email-etiquette-for-students/>.

There are a number of supports in place if you are experiencing bullying, harassment or sexual misconduct and you are strongly encouraged to come forward to seek confidential support and guidance on the range of informal options and formal options for resolving issues as appropriate. Reports of bullying, harassment or sexual misconduct can also be made anonymously through UCD's Report and Support tool.

UCD is actively promoting a culture where bullying, harassment and sexual misconduct is not tolerated, where everyone is respected and feels valued, included and that they belong in UCD.

You can find more details on UCD's Dignity and Respect Website at: <https://www.ucd.ie/equality/support/dignityrespect/>.

Expectations and Grading

The assignments consist of two Multiple Choice Questionnaires and a group project. Table 1 lists the deadlines and weight of each assignment for the final grade.

Assignments 1 and 2: Multiple Choice Questionnaires

Table 2 shows UCD's [Alternative Linear Conversion Grade Scale \(40% Pass\)](#), which is used for the two MCQ assignments. More details on the MCQ assignments will be provided in class.

Table 1: Overview of Assignments and Deadlines

Date	Assignment
Week 7 (22 October)	Multiple Choice Questionnaire #1: 25%
Week 11 (19 November)	Multiple Choice Questionnaire #2: 25%
4 December	Data Report: Descriptive analysis and visual representation of insights derived from a large dataset: 50%

Assignment 3: Data Report (Group Project)

In this final project, your group will craft a compelling data story based on a dataset that you find online (or optionally need to collect/scrape yourself). You will use descriptive analysis, visualisation techniques, and critical thinking to explore and explain trends, patterns, or insights from the data. The submission deadline for this assignment is **Friday, 4 December, 8pm**.

Groups will be allocated as follows: students registered for the Connected_Politics module will work in the same group as in Connected_Politics. The remaining students will be randomly assigned to a group. We try to ensure that all group members are registered for a degree in the same School. Further information will be provided in class.

Deliverable

A written report in the form of a blog post-style data story. The story should present a clear narrative that takes the reader through the data analysis process, supported by descriptive graphs and thoughtful commentary.

Below are websites and examples that can serve as an inspiration (the length and style of writing differ from the assignment brief, but all resources "tell stories with data").

Table 2: Alternative Linear Conversion Grade Scale for MCQ Assignments (with 15 questions)

Grades	Lower %	Upper %
A+	≥95	100
A	≥90	<95
A-	≥85	<90
B+	≥80	<85
B	≥75	<80
B-	≥70	<75
C+	≥65	<70
C	≥60	<65
C-	≥55	<60
D+	≥50	<55
D	≥45	<50
D-	≥40	<45
E+	≥35	<40
E	≥30	<35
E-	≥25	<30
F+	≥20	<25
F	≥15	<20
F-	≥10	<15
G+	≥5	<10
G	≥0.02	<5
G-	≥0.01	<0.02
NM	0	<0.01

ABS: No work was submitted
by the student or
the student was absent from assessment

- The [Data Stories](#) blog from the Swiss National Science Foundation, for example [this blog post on funding for early-career researchers](#).
- [Our World in Data](#), which publishes narrative articles combining data analysis with interactive visualisations on topics ranging from democracy and elections to inequality and public health. [This article on the rise of electoral democracy](#) is a strong example of the format.
- The [Pudding](#), which produces comprehensive data-driven visual essays on a wide range of social and cultural topics. Note that the interactive elements are what make the Pudding distinctive – you are not expected to produce anything similar – but the pieces are worth reading as examples of how to build a thorough, data-driven narrative.
- *The Economist* produces excellent data stories across politics, economics, and society. The website requires a subscription, but UCD students have access to *The Economist*'s full PDF archive through the [UCD Library](#).

We will also upload excellent examples from previous years to Brightspace to give you a better idea of the scope and structure.

Individual Accountability and Group Contribution Assessment

Group work is a core part of this module because it cultivates essential skills such as collaboration, problem-solving, and communication, which are vital for professional success. The policy on individual accountability within groups is designed to prevent issues related to uneven contributions in group assignments. Each group member is required to submit a document along with their group project that clearly specifies their individual contributions. This document must transparently detail each student's roles, responsibilities, and actual input. While the aim is to award the same grade to all members of the group, divergent contributions will result in varied grades to ensure fairness. This

policy not only promotes equitable grading but also deters disparities in the distribution of workload among team members. Students are encouraged to report any issues with group dynamics promptly to facilitate swift intervention and support from the course instructor.

Steps to Follow

1. Select a Dataset

- You may find a dataset online or collect your own data (this is optional). The dataset should *not* be one we have used in class. You can find links to large data repositories in the “Helpful Links” section on page 9 of this document.
- Ensure the dataset is complex enough to support meaningful analysis but manageable within the scope of the project.
- Depending on your research question, you might also collect and merge two or more datasets.
- Once you have identified a dataset, contact the TA Heba Elsahn [via Email](#) or Slack for quick approval before proceeding with the analysis. You must contact Heba three weeks before the submission deadline (i.e., no later than 13 November).

2. Define Your Research Question(s)

- Formulate 1–2 research questions that will guide your analysis. Your questions should help frame the narrative of your data story and be answerable using the data.
- Example(s): *“What are the key factors influencing student performance in higher education across different countries?” “What are the voting patterns in urban vs. rural areas over the last three U.S. presidential elections?” “How has government spending on social welfare changed in response to economic recessions?”*

3. Explore the Data

- Conduct an initial exploration of the dataset to understand its structure, key variables, and any missing data.
- Summarise your findings in a brief introductory section, giving readers context for what the dataset contains.

4. Analyse the Data

- Perform descriptive statistical analyses.
- Identify any notable relationships, patterns, or trends that help answer your research questions.

5. Create Visualisations

- Construct 3–5 clear and informative graphs that help illustrate your findings.
- Each graph should be accompanied by a brief description that explains what the graph shows and how it relates to your research question.
- Follow best practices for data visualisation (clear labels, appropriate scales, and use of colour) discussed in class. The readability and interpretability of the graphs will be a crucial component of this assignment.

6. Tell the Story

- Organise your findings into a coherent narrative. Your story should walk the reader through the data, describing your research question and how the analysis addresses it.
- Highlight interesting insights or unexpected findings. Where applicable, connect your findings to broader academic or policy debates.
- Keep your audience in mind – assume they have a basic understanding of the topic but not expert knowledge.

7. Conclusion

- Summarise your key takeaways. What did the data reveal? What questions remain unanswered?
- Discuss any limitations of your analysis and suggest potential avenues for future research.

8. Submission Requirements

- Word count: 1,000–1,300 words (excluding graphs and references).
- Include 3–5 descriptive graphs.
- Submit the assignment as a PDF document. Word documents will not be accepted. Make sure to save all graphs in a vector graphic format (ideally, PDF or EPS) and include these graphs in high-quality resolution.
- Submit the files containing the data and code to complete the project in a reproducible format. This means that you need to upload the raw dataset(s), and your R scripts to wrangle, (merge), and visualise the data. The code might be checked during the grading process if questions emerge.
- Include a reference section for any external data sources or literature cited.
- Make sure your team has completed the mandatory Declaration on the Use of Artificial Intelligence Tools (<https://disclose-ai.app>).

Grading Criteria

- **Data Exploration & Analysis:** Depth of analysis and insight into the dataset.
- **Visualisations:** Quality, clarity, and appropriateness of graphs.
- **Narrative & Storytelling:** Coherence and structure of the data story. Make sure to follow [this guide](#) on how to structure and write paragraphs.

Helpful Links

Here are some recommended repositories and datasets for your project. A good starting point is [PolData](#), which provides a comprehensive overview of political datasets across a wide range of topics and sub-fields – it is worth browsing before settling on a source.

- [Ireland's Open Data Portal](#): Irish public sector data in open, free, and reusable formats.
- [Harvard Dataverse](#): One of the world's largest open-access repositories, with data from various research disciplines including social sciences, law, and health.
- [Google Dataset Search](#): Google's search engine for publicly available datasets across a wide range of fields.
- [World Bank Open Data](#): Global development data on economics, demographics, and more.

- [UCI Machine Learning Repository](#): A large repository with datasets in health, economics, and more, often used for machine learning tasks.
- [United Nations Data](#): Global statistics covering population, education, employment, and development indicators.
- [European Union Open Data Portal](#): Public data on various sectors such as economics, transportation, and environment.
- [Pew Research Center](#): Survey data covering politics, social issues, media, and more.
- [Open Data UK](#): Datasets provided by UK government departments, covering various topics including health, crime, and transportation.
- [OECD Data](#): Data on global economic and social issues, including education and health.
- [WhoGov](#): A global dataset on cabinet members in 177 countries from 1966 to the present, including variables on gender, party affiliation, and ministerial experience.
- [Varieties of Democracy \(V-Dem\)](#): A comprehensive dataset measuring democracy across more than 400 indicators in all countries since 1900, covering electoral, liberal, participatory, deliberative, and egalitarian dimensions. Freely available and updated annually.
- [Manifesto Project \(MARPOR/CMP\)](#): Quantitative content analysis of party manifestos from over 1,000 parties in more than 50 countries since 1945, covering policy positions on issues such as welfare, European integration, and gender equality.
- [ParlGov](#): A database on parties, elections, and governments in EU member states and other established democracies, covering parliamentary and European elections as well as cabinet compositions.
- [Chapel Hill Expert Survey \(CHES\)](#): Expert estimates of party positioning on European integration, ideology, and policy issues for national parties across a growing number of European countries, available since 1999.
- [Quality of Government \(QoG\) Dataset](#): A compilation dataset from the University of Gothenburg drawing on more than 100 sources, covering governance, institutions, public administration, health, and social policy across countries and time.
- [GESIS – Leibniz Institute for the Social Sciences](#): One of Europe’s largest social science data archives, hosting over 7,000 studies including the Eurobarometer series, the European Values Study, and the International Social Survey Programme (ISSP), all with comprehensive documentation.
- [ICPSR – Inter-university Consortium for Political and Social Research](#): The world’s largest archive of social science data, covering politics, health, economics, education, and more, with thousands of well-documented studies available for download.
- [Comparative Study of Electoral Systems \(CSES\)](#): A collaborative programme of cross-national survey research covering voter behaviour, electoral institutions, and party competition across dozens of democracies.

Note: Whichever dataset you choose, it must be well-documented and accompanied by a codebook. Data uploaded informally by individual users – for example, on platforms such as Kaggle – often lacks the methodological transparency required for academic work and should not be used unless full documentation is available.

Week 1: Introduction to R and Positron (10 September; Room L0.04-LEA)

Mandatory Readings:

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O'Reilly. URL: <https://r4ds.hadley.nz>: chapters 2 and 27.
- C. O. Wilke (2019). *Fundamentals of Data Visualization: A Primer On Making Informative and Compelling Figures*. Sebastopol: O'Reilly. URL: <https://clauswilke.com/dataviz/>: chapter 2.

Important: Please follow the installation guide (see the **Installation Instructions** page under **My Learning**), available on Brightspace, in order to set up R, Positron, and GitHub Copilot. Make sure to install the software *before* our first lecture. If you have any question related to (installing) the software, please ask this question through our Slack workspace.

Week 2: Central Terms and Concepts (17 September; Room L0.04-LEA)

Mandatory Readings:

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O'Reilly. URL: <https://r4ds.hadley.nz>: chapter 3.
- C. Ismay and A. Y. Kim (2025). *Statistical Inference via Data Science: A Modern Dive into R and the tidyverse*. 2nd edition. Boca Raton: CRC Press. URL: <https://moderndive.com/v2/>: Appendix A.

Optional Readings:

- J. Bryan (2020). *Happy Git and GitHub for the useR*. URL: <https://happygitwithr.com> (skim if you are interested in version control).

Week 3: The Hidden Curriculum of Coding: Projects, Paths, and Folder Structure (24 September; Room L0.04-LEA)

Mandatory Readings:

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O'Reilly. URL: <https://r4ds.hadley.nz>: chapters 6, 20, and 28.
- Posit Team (2025). *Tutorial: Hello, Quarto*. 2025. URL: <https://quarto.org/docs/get-started/hello/positron.html>: select your preferred tool.

Week 4: Summarising and Transforming Data (1 October; Room G-08.AG)

Mandatory Readings:

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O'Reilly. URL: <https://r4ds.hadley.nz>: chapter 5.

Week 5: Principles of Data Visualisation (8 October; Room G-08.AG)

Mandatory Reading:

- C. O. Wilke (2019). *Fundamentals of Data Visualization: A Primer On Making Informative and Compelling Figures*. Sebastopol: O'Reilly. URL: <https://clauswilke.com/dataviz/>: chapters 17, 19, and 29.

Optional Reading:

- C. O. Wilke (2019). *Fundamentals of Data Visualization: A Primer On Making Informative and Compelling Figures*. Sebastopol: O'Reilly. URL: <https://clauswilke.com/dataviz/>: chapters 1, 4, 18, 26.

Week 6: Explorative Data Visualisation (15 October; Room L2.17-LEA)

Mandatory Readings:

- K. Healy (2019). *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. URL: <http://socviz.co>: chapters 1, 3, and 4.

Week 7: Advanced, Intuitive, and Accessible Data Visualisation (22 October; Room G-08.AG)

Mandatory Readings:

- K. Healy (2019). *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. URL: <http://socviz.co>: chapters 5 and 8.
- J. P. Kastellec (2025). *Practical Advice for Producing Better Graphs*. URL: https://jkastellec.scholar.princeton.edu/sites/g/files/toruqf3871/files/documents/kastellec_graphs_practical_tips_v2.pdf.

Optional Readings:

- S. Dattani (2025). *Saloni's Guide to Data Visualization: Why Data Visualization Matters, and How to Make Charts More Effective, Clear, Transparent, and Sometimes, Beautiful*. URL: <https://www.scientificdiscovery.dev/p/salonis-guide-to-data-visualization>.
- L. Theodosiou (2026). *ltc: Collection of Artistic and Nature-Inspired Color Palettes*. URL: <https://loukesio.github.io/ltc-color-palettes/>.
- J. L. Steenwyk and A. Rokas (2021). “ggpubfigs: Colorblind-Friendly Color Palettes and ggplot2 Graphic System Extensions for Publication-Quality Scientific Figures”. *Microbiology Resource Announcements* 10 (44): e00871–21. DOI: 10.1128/mra.00871-21.

Week 8: No Class – Reading Break (29 October)

Week 9: Maps and Interactive Data Visualisation (5 November; Room L0.04-LEA)

Mandatory Readings

- R. Alexander (2023). *Telling Stories with Data: With Applications in R*. New York: CRC Press. URL: <https://tellingstorieswithdata.com>: chapter 5.4.
- University of Gettysburg (2026). “Digital Humanities Toolkit: Flourish.Studio”. URL: <https://dh.sites.gettysburg.edu/toolkit/tools/flourish-studio/>.

Optional

- H. Wickham (2021). *Mastering Shiny: Build Interactive Apps, Reports & Dashboards Powered by R*. Sebastopol: O’Reilly. URL: <https://mastering-shiny.org>.
- G. Aden-Buie, C. Sievert, R. Iannone, J. Allaire, and B. Borges (2026). *flexdashboard: R Markdown Format for Flexible Dashboards*. R package version 0.6.3. DOI: 10.32614/CRAN.package.flexdashboard.

Week 10: Relational Databases (12 November; Room G-08.AG)

Mandatory Readings

- H. Wickham, M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O’Reilly. URL: <https://r4ds.hadley.nz>: chapters 19 and 21.

Optional

- N. B. Weidmann (2023). *Data Management for Social Sciences: From Files to Databases*. Cambridge: Cambridge University Press. URL: <https://cambridge.org/9781108845670>: chapters 8–9.
- H. Mühleisen and K. Müller (2025). *duckplyr: A ‘DuckDB’-Backed Version of ‘dplyr’*. R package version 1.1.1. URL: <https://CRAN.R-project.org/package=duckplyr>.

Week 11: Introduction to Web Scraping (19 November; Room L0.04-LEA)

Mandatory Readings:

- R. Alexander (2023). *Telling Stories with Data: With Applications in R*. New York: CRC Press. URL: <https://tellingstorieswithdata.com>: chapter 7.

Week 12: Advanced Web Scraping (26 November; Room G-08.AG)

Mandatory Reading:

- H. Wickham (2022). *rvest: Selector Gadget*. URL: <https://rvest.tidyverse.org/articles/selectorgadget.html>.
- M. A. Brown, A. Gruen, G. Maldoff, S. Messing, Z. Sanderson, and M. Zimmer (2025). “Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations”. *Big Data & Society* 12 (4): 1–17. DOI: 10.1177/20539517251381686.

References

- Aden-Buie, G., C. Sievert, R. Iannone, J. Allaire, and B. Borges (2026). *flexdashboard: R Markdown Format for Flexible Dashboards*. R package version 0.6.3. DOI: 10.32614/CRAN.package.flexdashboard.
- Alexander, R. (2023). *Telling Stories with Data: With Applications in R*. New York: CRC Press. URL: <https://tellingstorieswithdata.com>.
- Brown, M. A., A. Gruen, G. Maldoff, S. Messing, Z. Sanderson, and M. Zimmer (2025). “Web Scraping for Research: Legal, Ethical, Institutional, and Scientific Considerations”. *Big Data & Society* 12 (4): 1–17. DOI: 10.1177/20539517251381686.
- Bryan, J. (2020). *Happy Git and GitHub for the useR*. URL: <https://happygitwithr.com>.
- Dattani, S. (2025). *Saloni’s Guide to Data Visualization: Why Data Visualization Matters, and How to Make Charts More Effective, Clear, Transparent, and Sometimes, Beautiful*. URL: <https://www.scientificdiscovery.dev/p/salonis-guide-to-data-visualization>.
- Healy, K. (2019). *Data Visualization: A Practical Introduction*. Princeton: Princeton University Press. URL: <http://socviz.co>.
- Ismay, C. and A. Y. Kim (2025). *Statistical Inference via Data Science: A Modern Dive into R and the tidyverse*. 2nd edition. Boca Raton: CRC Press. URL: <https://moderndive.com/v2/>.
- Kastellec, J. P. (2025). *Practical Advice for Producing Better Graphs*. URL: https://jkastellec.scholar.princeton.edu/sites/g/files/toruqf3871/files/documents/kastellec_graphs_practical_tips_v2.pdf.
- Monroe-White, T., D. Kozlowski, L. Domenech, C. Pradier, G. Rosati, and N. S. Shokida (2024). *Critical Computational Social Sciences*. URL: <https://critcss.github.io/CCSS/>.
- Mühleisen, H. and K. Müller (2025). *duckplyr: A ‘DuckDB’-Backed Version of ‘dplyr’*. R package version 1.1.1. URL: <https://CRAN.R-project.org/package=duckplyr>.
- Müller, S. (2023). “How Slack Facilitates Communication and Collaboration in Seminars and Project-Based Courses”. *Journal of Educational Technology Systems* 51 (3): 303–316. DOI: 10.1177/00472395231151910.
- Navarro, D. (2026). *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners*. URL: <https://learningstatisticswithr.com>.
- Posit Team (2025). *Tutorial: Hello, Quarto*. 2025. URL: <https://quarto.org/docs/get-started/hello/positron.html>.
- Steenwyk, J. L. and A. Rokas (2021). “ggpubfigs: Colorblind-Friendly Color Palettes and ggplot2 Graphic System Extensions for Publication-Quality Scientific Figures”. *Microbiology Resource Announcements* 10 (44): e00871–21. DOI: 10.1128/mra.00871-21.
- Theodosiou, L. (2026). *ltc: Collection of Artistic and Nature-Inspired Color Palettes*. URL: <https://loukesio.github.io/ltc-color-palettes/>.
- University of Gettysburg (2026). “Digital Humanities Toolkit: Flourish.Studio”. URL: <https://dh.sites.gettysburg.edu/toolkit/tools/flourish-studio/>.
- Weidmann, N. B. (2023). *Data Management for Social Sciences: From Files to Databases*. Cambridge: Cambridge University Press. URL: <https://cambridge.org/9781108845670>.
- Wickham, H. (2021). *Mastering Shiny: Build Interactive Apps, Reports & Dashboards Powered by R*. Sebastopol: O’Reilly. URL: <https://mastering-shiny.org>.
- Wickham, H. (2022). *rvest: Selector Gadget*. URL: <https://rvest.tidyverse.org/articles/selectorgadget.html>.
- Wickham, H., M. Çetinkaya-Runde, and G. Grolemund (2023). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. 2nd edition. Sebastopol: O’Reilly. URL: <https://r4ds.hadley.nz>.
- Wilke, C. O. (2019). *Fundamentals of Data Visualization: A Primer On Making Informative and Compelling Figures*. Sebastopol: O’Reilly. URL: <https://clauswilke.com/dataviz/>.