SMI606 Introduction to Quantitative Research: Detailed Reading List

Dr. Calum Webb, Sheffield Methods Institute

2024-09-16

This reading list will help you to prepare for each session of the module. In addition to outlining the required, highly recommended, and recommended reading for each topic, there is also a more general breakdown of suggested resources, particularly those related to learning how to do quantitative social science using `R`.

Where possible, I have added stars to indicate how accessible I believe a given textbook is. One star (*) means the text should be *accessible for most people without any prior reading*; two stars (**) means that the text should be *mostly accessible, but may require some intensive study or prior reading*; and three stars (***) means that the text is *more complex, technical, or specialist than most, and might require working up to, cross-referencing terms with simpler texts, or going over multiple times.*

Highly recommended texts for this module

This module has one text that is highly recommended and is used for most of the preparatory reading. However, **for each week I have also supplied additional or alternative reading**, much of which is from textbooks that are available for free online or through the library. You should not feel constrained by these textbooks. You should also not feel like you need to complete all of the additional reading for each topic; my personal recommendation is to read one text for each week and then read further if you are still struggling with the topic or find it particularly interesting, or if the chosen reading did not work well for you.

You should aim to read one chapter of a textbook or relevant journal article for each week's topic.

The following textbook is strongly recommended as it will help you study the substantive content and provide you with additional practical activities in `R` to complement the skills we learn in class, and to ensure you are able to complete your assessments:

- **Mehmetoglu, Mehmet & Mittner, Matthias**. (2022). *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

This new textbook largely follows the `R` coding and analysis workflow that we will be following in class. It also covers most of the content from this module and some of the context for SMI601 Advanced Quantitative Methods, so may be a good investment. Copies are available to be taken out from the library and the book itself can be purchased from SAGE for £36.

You may also find the following classic text helpful for getting to grips with some of the underlying statistical theory and probability used in the module:

- **Rowntree, Derek**. (1981). *'Statistics without tears: An introduction for non-Mathematicians'*. Penguin.*

It is a very short book and does not complicate the subject matter too heavily by introducing other things like exercises in a programming language or piece of software at the same time. Available on bookshop.org

(£9.29) or older editions available for free on archive.org.

Week-by-week Preparatory Reading

This section outlines the required preparatory reading for each week and its purpose. You will also find additional reading that you might find interesting or useful.

Sometimes you will come across some topics or concepts out of order to how we cover them in the class. Don't worry about this too much! There are a great many pedagogical differences that mean that some people choose to introduce topics in different orders. If something you read about hasn't been explained in class, try not to panic, it will either be covered later or is not essential.

You should not feel restricted to only the texts in this list and I would encourage you to use your own initiative to find ways to work in `R` that work for you. There are large numbers of online tutorials, bootcamps, MOOCs, documentation, blogs, and other forms of content that can help you achieve the same goals in `R` using different tools. Don't be afraid to google a topic!

Week 1: What is quantitative social science?

The purpose of this week is to get you thinking about the nature of quantitative social science (what it posits about the way the social/human world works, its epistemological and ontological underpinning), what kinds of questions it might be useful for answering, and how it can be used. I would encourage you to regularly visit literature like this which will assist you in thinking critically about research methods and methodology.

Preparatory reading

- Powell, T. C. (2020). Can quantitative research solve social problems? Pragmatism and the ethics of social research. *Journal of Business Ethics*, 167(1), 41-48. https://doi.org/10.1007/s10551-019-04196-7 **
- Mehmetoglu, M. & Mittner, M. (2022). Chapter 1: Introduction to `R`. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*
- Mehmetoglu, M. & Mittner, M. (2022). Chapter 2: Importing and working with data in `R`. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*
- Mehmetoglu, M. & Mittner, M. (2022). Chapter 3: How does `R` work? *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative reading On the philosophy of science and quantitative social research:

- Walter, M. (2006). The nature of social science research. *Social research methods: An Australian perspective*, 1-28.*
- Bryman, A. (1984). The debate about quantitative and qualitative research: a question of method or epistemology?. *British Journal of Sociology*, 75-92. https://doi.org/10.2307/590553 *
- Boruch, R. F. (1984). Ideas about social research, evaluation, and statistics in medieval Arabic Literature: Ibn Khaldun and al-Biruni. *Evaluation Review*, 8(6), 823-842. https://doi.org/10.1177%2F0193841X8400800604
- Connell, C. M. (2016). *Introduction to quantitative methods.* Handbook of Methodological Approaches to Community-based Research: Qualitative, Quantitative, and Mixed Methods, 121-131. (Available through the library)
- Green, D. S., & Wortham, R. A. (2018). The sociological insight of WEB Du Bois. *Sociological Inquiry*, 88(1), 56-78. https://doi.org/10.1111/soin.12179

- Jenkins, R. (2018). Foundations of sociology: Towards a better understanding of the human world. Macmillan International Higher Education.
- Johnston, R., Harris, R., Jones, K., Manley, D., Wang, W. W., & Wolf, L. (2020). Quantitative methods II: How we moved on–Decades of change in philosophy, focus and methods. *Progress in Human Geography*, 44(5), 959-971. https://doi.org/10.1177/0309132519869451
- Mills, C. W. (1953). Two styles of research in current social studies. *Philosophy of Science*, 20(4), 266-275. https://doi.org/10.1086/287280
- Walter, M., & Suina, M. (2019). Indigenous data, indigenous methodologies and indigenous data sovereignty. *International Journal of Social Research Methodology*, 22(3), 233-243. https://doi.org/10.1080/13645579.2018.1531228
- Williams, R. W. (2006). The early social science of WEB Du Bois. *Du Bois Review: Social Science Research on Race*, 3(2), 365-394. https://doi.org/10.1017/S1742058X06060243

Additional reading for getting started with `R` (setting up and basic interaction with the console and scripts):

- Imai, K. (2017) Chapter 1.3: Introduction to `R`. *'Quantitative Social Science: An introduction'*. Princeton University Press.**

or

- Fogarty, B. (2019). Chapter 2: Introduction to `R` and RStudio. *Quantitative Social Science Data with `R`*. Sage.*

or

- Wickham, H. (2017) Chapter 1: Introduction. *R for Data Science*. O'Reilly (Free: https://r4ds.had.co.nz) *
- Wickham, H. (2017) Chapter 4: Workflow: basics. *R for Data Science*. O'Reilly (Free: https://r4ds.had.co.nz) *

or

- Estrellado, et al. (2020). Chapter 5: Getting Started with `R` and RStudio. *Data Science in Education Using R*. Routledge. (Free: https://datascienceineducation.com) *
- Estrellado, et al. (2020). Chapter 6: Foundational Skills. *Data Science in Education Using R*. Routledge. (Free: https://datascienceineducation.com) *

Week 2: Types of Quantification

Before we can use any kind of quantitative methods of analysis, we need to learn how we can actually quantify things. While this might sound extremely obvious, it is incredibly important to know the different types of quantification that exist as these dictate what kinds of data visualisation, summary statistics, and statistical tests and models we should use. They also allow us to check for patterns in our data that might violate some of the assumptions these tests, summary statistics, and models use.

By the end of this week, you should know the difference between continuous, ordinal, and categorical variables as well as appropriate summary statistics and visualisations for each and how to produce them using `R`.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 4: Data Management. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative reading Additional reading that applies these concepts in `R`:

- Fogarty, B. (2019). Chapter 5: Variables and manipulation. *Quantitative Social Science Data with R*. Sage.*
- Fogarty, B. (2019). Chapter 7: Univariate and descriptive statistics. *Quantitative Social Science Data with R*. Sage.*
- Fogarty, B. (2019). Chapter 8: Visualising data. *Quantitative Social Science Data with R*. Sage.*
- Navarro, D. (2020). Chapter 5: Descriptive Statistics. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **

Additional reading on statistics:

- Rowntree, D. (1981). Chapter 2: Describing our sample. *Statistics without Tears*. Penguin.*
- Rowntree, D. (1981). Chapter 3: Summarising our data. *Statistics without Tears*. Penguin.*
- Rowntree, D. (1981). Chapter 4: The shape of a distribution. *Statistics without Tears*. Penguin.*

Week 3: Relationships between variables

Now that we've learned how to quantify single variables and inspect their variation in `R` depending on their type, we can explore how to inspect relationships *between* two variables. By the end of this week, you should be able to demonstrate you know which kinds of visualisations and descriptive statistics to use to explore the relationships between different types of variable in your data.

This can include scatterplots and correlation statistics for two continuous/ordinal variables; contingency tables and heatmaps for two categorical variables; and mean-differences and boxplots/ridge plots for ordinal/categorical and ordinal/continuous variables.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 5: Data visualisation with `ggplot2`. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative Reading Additional examples in `R` for data visualisation in `ggplot`

- Fogarty, B. (2019). Chapter 8: Visualising data. *Quantitative Social Science Data with R*. Sage.*
- Fogarty, B. (2019). Chapter 10: Bivariate analysis. *Quantitative Social Science Data with R*. Sage.*
- Wickham, H. (2017) Chapter 2: Data visualisation. *R for Data Science*. O'Reilly (Free: https://r4ds.had.co.nz) *
- Navarro, D. (2020). Chapter 5: Descriptive Statistics. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **
- Navarro, D. (2020). Chapter 6: Drawing graphs. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **

More in depth treatment of the statistical theory behind some forms of comparison:

- Rowntree, D. (1981). Chapter 6: Comparing samples. *Statistics without Tears*. Penguin.*
- Rowntree, D. (1981). Chapter 8: Analysing relationships. *Statistics without Tears*. Penguin.*

Week 4: Inference

Often, an important goal for quantitative social scientists is being able to make generalisable claims about the patterns that exist in their data: that they apply to the entire population of interest and not just their specific sample. The most common way we achieve this is through inferential statistics and statistical hypothesis testing.

By the end of this week, you should be able to explain the logic behind statistical testing, how tests relate to specific hypotheses, and how to interpret a *p-value*. You should also have a working knowledge of what kinds

of samples and survey methodologies can lead to population inference. Between this week and the following week, you will have practiced running and interpreting the results of some bivariate statistical tests in `R`.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 6: Descriptive statistics. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative Reading Alternative treatment of statistical theory:

- Rowntree, D. (1981). Chapter 5: From sample to population. *Statistics without Tears*. Penguin.*
- Wheelan, C. (2012). Chapter 9: Inference. *Naked Statistics: Stripping the Dread from the Data*. W. W. Norton.*

Inferential statistics in `R`

- Fogarty, B. (2019). Chapter 6: Developing Hypotheses. *Quantitative Social Science Data with `R`*. Sage.*
- Fogarty, B. (2019). Chapter 9: Hypothesis Testing. *Quantitative Social Science Data with `R`*. Sage.*

or

- Imai, K. (2017) Chapter 7.1: Estimation. *'Quantitative Social Science: An introduction'*. Princeton University Press.**
- Imai, K. (2017) Chapter 7.2: Hypothesis Testing. *'Quantitative Social Science: An introduction'*. Princeton University Press.**

or

- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 4: Statistical inference. *Regression and Other Stories*. Cambridge University Press.**

or

- Navarro, D. (2020). Part IV: Statistical Theory. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **
- Navarro, D. (2020). Chapter 12: Categorical Data Analysis. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **
- Navarro, D. (2020). Chapter 13: Comparing Two Means. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **
- Navarro, D. (2020). Chapter 14: Comparing Several Means. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **

Week 5: Causality

We should now know when it is appropriate to generalise our findings to a wider population depending on how our data has been collected, but another major claim we might often want to make about quantitative research is whether we can argue that a relationship is *causal* or not.

As with inferential statistics, there are a number of conditions that need to be satisfied for different strengths of causal evidence. By the end of this week, you should be able to rate the degree to which causality can be inferred based on study design.

Preparatory Reading

- Goldthorpe, J. H. (2001). Causation, statistics, and sociology. *European Sociological Review*, 17(1), 1-20. https://doi.org/10.1093/esr/17.1.1 **

Additional/alternative Reading

- Imai, K. (2017) Chapter 2: Causality. *'Quantitative Social Science: An introduction'*. Princeton University Press.**
- Greenland, Pearl, & Robbins (1999). Causal Diagrams for Epidemiologic Research. *Epidemiology* https://www.jstor.org/stable/3702180 **
- Westreich & Greenland. (2013). The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. *American Journal of Epidemiology* https://doi.org/10.1093/aje/kws412 **
- Pearl, J. (2018). *The Book of Why: The new science of cause and effect*. London: Penguin Books. **
- Peters, Janzing, & Schölkopf (2017) *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press. ***
- McElreath, R. (2020) *Statistical Rethinking: A Bayesian course with examples in R and Stan*. Routledge. ***
- Cunningham, S. (2021). *Causal Inference: the Mixtape*. Yale University Press. ***

Week 6: Bivariate linear regression

Now that we have covered the bulk of the pre-requisite statistical theory and familiarity with working in `R` we can move onto the workhorse of contemporary quantitative social science: regression. Regression may seem daunting at first, but once you become familiar with its core concepts you will be able to easily run and interpret regression models and see how they share features with other types of statistical analysis.

By the end of this week you should be able to explain how linear regression works and interpret the output of a bivariate linear regression of a normally distributed dependent variable on a continuous independent variable. Further, you will be able to run a model like this in `R`.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 7: Simple (Bivariate) Regression. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative Reading

- Wheelan, C. (2012). Chapter 11: Regression analysis *Naked Statistics: Stripping the Dread from the Data*. W. W. Norton. (PDF available on Blackboard) *
- Wheelan, C. (2012). Chapter 12: Common regression mistakes *Naked Statistics: Stripping the Dread from the Data*. W. W. Norton. (PDF available on Blackboard) *

Week 7: Reading week

Week 7 is a reading week — my recommendation is for you to use this week to revisit the preparatory reading, or engage with the additional reading, of a topic from the previous weeks that you still feel somewhat challenged by. My suggestion below, for example, is to revisit some of the statistical theory literature on the topic of inference which many students tend to struggle with.

If you feel like you have a very good understanding of everything covered so far, I would recommend you either (a) engage with a practical learning resource to develop and reinforce your `R` skills, such as the #TidyTuesday project or (b) read ahead on the topic of regression.

Suggested reading:

- Rowntree, D. (1981). Chapter 6: Comparing samples. *Statistics without Tears*. Penguin.*
- Rowntree, D. (1981). Chapter 7: Further matters of significance. *Statistics without Tears*. Penguin.*
- Rowntree, D. (1981). Chapter 8: Analysing relationships. *Statistics without Tears*. Penguin.*

Week 8: Multiple linear regression

In week 8, you will take what you learned about linear regression and extend our regression models to include multiple predictors of an outcome. This will illustrate how powerful regression models can be for social science, especially where there may be confounding variables to control for *post hoc*.

By the end of this week, you will have be able to include multiple independent variables into regression models in `R`, including categorical variables, and interpret the output.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 8: Multiple Linear Regression. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*
- Mehmetoglu, M. & Mittner, M. (2022). Chapter 9: Dummy-variable Regression. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative Reading

- Fogarty, B. (2019). Chapter 11: Linear Regression & Model Building. *Quantitative Social Science Data with R*. Sage.*
- Fogarty, B. (2019). Chapter 12: OLS Assumptions & Diagnostic Testing. *Quantitative Social Science Data with R*. Sage.*
- Imai, K. (2017) Chapter 4.2: Linear regression *'Quantitative Social Science: An introduction'*. Princeton University Press.**
- Imai, K. (2017) Chapter 4.3.2: Regression with multiple predictors *'Quantitative Social Science: An introduction'*. Princeton University Press.**
- Navarro, D. (2020). Chapter 15: Linear regression. *Learning Statistics with R*. (Free online: https://learningstatisticswithr.com) **
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 6: Background on regression modeling. *Regression and Other Stories*. Cambridge University Press.**
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 7: Linear regression with a single predictor. *Regression and Other Stories*. Cambridge University Press.**
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 8: Fitting regression models. *Regression and Other Stories*. Cambridge University Press.**
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 10: Linear regression with multiple predictors. *Regression and Other Stories*. Cambridge University Press.**
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 11: Assumptions, diagnostics, and model evaluation. *Regression and Other Stories*. Cambridge University Press.**
- Dalpiaz, D. (2021). Chapter 7: Simple Linear Regression. *Applied Statistics with R* (Free online: http://daviddalpiaz.github.io/appliedstats/) ***
- Dalpiaz, D. (2021). Chapter 8: Inference for Simple Linear Regression. *Applied Statistics with R* (Free online: http://daviddalpiaz.github.io/appliedstats/) ***
- Dalpiaz, D. (2021). Chapter 9: Multiple Linear Regression. *Applied Statistics with R* (Free online: http://daviddalpiaz.github.io/appliedstats/) ***
- Dalpiaz, D. (2021). Chapter 11: Categorical Predictors and Interactions. *Applied Statistics with R* (Free online: http://daviddalpiaz.github.io/appliedstats/) ***

Week 9: Logistic regression

As we will have seen by week 9, multiple linear regression can be incredibly flexible to answer a number of research questions by including multiple predictors — but what if we need more flexibility around what kind of outcome variable we are interested in? In week 9 we will look at logistic regression, a type of Generalised Linear Model (GLMs) for predicting a binary categorical outcome.

By the end of this week you will be able to run and interpret the output of a logistic regression model in `R`.

Preparatory Reading

- Mehmetoglu, M. & Mittner, M. (2022). Chapter 11: Logistic Regression. *'Applied Statistics Using R: A guide for the social sciences'*. London: SAGE.*

Additional/alternative Reading

- Schumacker, R. (2021). Chapter 18: Logistic Regression. *Learning Statistics Using R*. Sage Research Methods. (Available online through the library) *
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 13: Logistic regression. *Regression and Other Stories*. Cambridge University Press.**
- Gelman, A., Hill, J. and Vehtari, A. (2020). Chapter 14: Working with logistic regression. *Regression and Other Stories*. Cambridge University Press.**
- Kaplan, D. Chapter 16: Models of Yes/No Variables. 'Statistical Modeling: A fresh approach' (Free online: https://dtkaplan.github.io/SM2-bookdown/models-of-yesno-variables.html) *
- Lottes, I. L., DeMaris, A., & Adler, M. A. (1996). Using and interpreting logistic regression: A guide for teachers and students. *Teaching Sociology*, 284-298. https://doi.org/10.2307/1318743 **
- Dalpiaz, D. (2021). Chapter 17: Logistic regression. *Applied Statistics with R* (Free online: http://daviddalpiaz.github.io/appliedstats/) ***

Week 10: Cluster Analysis

The second to last topic we will touch on is cluster analysis. There are many statistical and data learning methods that sit outside of the regression framework, especially those associated with machine learning, that are increasingly being used in the social sciences. Cluster analysis sits in one such area within a wider collection of methods under the umbrella of 'unsupervised machine learning'. It can be useful for identifying underlying 'groups' of observations based on their characteristics.

By the end of this week, you should be able to run two relatively simple algorithms for clustering observations by their features: k-means and agglomerative hierarchical clustering.

Note: Cluster Analysis should not be confused with Factor Analysis. Cluster analysis attempts to find clusters of ***observations*** *while factor analysis tries to reduce the number of dimensions in data by identifying a smaller number of latent factors associated with large groups of* ***variables****.*

Preparatory Reading

- UC Business Analytics `R` Programming Guide, 'k-Means Cluster Analysis' (https://uc-r.github.io/kmeans_clusterin) *
- UC Business Analytics `R` Programming Guide, 'Hierarchical Cluster Analysis' (https://uc-r.github.io/hc_clustering) *

Additional/alternative reading

- Imai, K. (2017) Chapter 3.7: Clustering. *'Quantitative Social Science: An introduction'*. Princeton University Press.**
- Waggoner, P. (2020). Chapter 4: k-means Clustering. *Unsupervised Machine Learning for Clustering in Political and Social Research*. Cambridge University Press. (Free online version here: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3693395) **
- Waggoner, P. (2020). Chapter 3: Agglomerative Hierarchical Clustering. *Unsupervised Machine Learning for Clustering in Political and Social Research*. Cambridge University Press. (Free online version here: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3693395) **
- Garson, G. D. (2021 *forthcoming*). Data Analytics for the Social Sciences: Applications in `R`. Routledge.
- James, G., Witten, D., Hastie, T. and Tibshirani, R. (2021). Chapter 12.4 Clustering Methods. *An Introduction to Statistical Learning with Applications in `R`. Second Edition*. Springer. (Available for free here: https://www.statlearning.com) **

Week 11: Spatial Analysis

When we do quantitative social research it's not uncommon for us to unthinkingly abstract our data from its spatial context, despite the impact that space and place has on lives. Spatial data can lead to critical new insights about segregation or integration, or can create powerful policy messages or strategies.

By the end of this final week, you will be able to plot spatial data in the form of choropleth maps and with data points to identify patterns using `R`. You will also learn a basic measure of spatial autocorrelation, Moran's *I*.

Preparatory reading

- Imai, K. (2017) Chapter 5.3 Spatial data. *'Quantitative Social Science: An introduction'*. Princeton University Press.**

Additional/alternative reading

- Pebesma, E., et al. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*. https://doi.org/10.32614/RJ-2018-009 **
- Pebesma, E., et al. (2021). `sf`: Simple Features for R. https://r-spatial.github.io/sf/index.html (See articles pages for walkthroughs on using `sf`)*

Additional Resources for Learning

One of the great things about learning `R` rather than a commercial statistical package is that it's free and has a large community of users who provide free resources for learners. The downside of this is that the amount of resources out there and the vast differences in how people code and teach can be quite overwhelming!

Below are a few resources that I think are helpful in addition to the books and articles used throughout the course. In particular, I would encourage you to use resources like the #TidyTuesday datasets and data from the UK Data Service to download and practice analysing data that you haven't seen before.

The easiest way to learn applied social science statistics is to embark on a project where you have to use them. I guarantee that if you were to spend a few hours a week applying the skills you've learned in class to a fresh dataset in an independent project, you will retain that knowledge much better (even if it is with a silly example of data!). What's more, because we are using `R` scripts, you will have a record of what you did to go back to when you need to do it again in future.

#TidyTuesday

TidyTuesday is a weekly recurring community data tidying, analysis, and visualisation project where `R` learners and users are encouraged to download, explore, visualise, and model open data and share their results (or seek help!) on social media (though you do not need to do this). There are now nearly four years worth of weekly-released datasets that can be revisited!

While this is not a social science-specific resource, many of the datasets submitted are related to social science. Some examples include:

- Bechdel Test data for movies on IMDB
- Wealth and income in the US over time
- The Big Mac Index
- African-American History
- Toronto Homeless Shelters

These data are often small in scope (number of variables) and consistent in file type, so can be easier to jump into working with than some of the much larger datasets like those from the UK Data Service.

UK Data Service

The UK Data Service is a national data service that provides access to social and economic data from censuses and surveys in the UK. It hosts many of the major surveys used for social science research in the country and internationally, including the Labour Force Survey, the UK Household Longitudinal Survey, the British Social Attitudes Survey, and the Crime Survey for England and Wales.

These are often quite large collections of data and you need to register with the UK Data Service to access and download them. This is an easy process and you can put your reason for downloading any data as learning purposes. You can also access Teaching Datasets, which can be helpful for learning as they tidy the data first (though in a quantitative research career you will have to learn how to do this yourself!)

Remember that you will have to make use of meta documents to know what each variable is. You might also need to use a package like `haven` to read in data that is only available in SPSS, Stata, or SAS files (I would recommend using Stata files if csv files are not available).

Some links to teaching datasets include:

- English Housing Survey, 2018-2019: Household Data Teaching Dataset
- Understanding Society: Ethnicity and Health Teaching Dataset
- Living Costs and Food Survey, 2013: Unrestricted Access Teaching Dataset
- National Survey of Sexual Attitudes and Lifestyles, 2010-2012: Open Access Teaching Dataset
- Crime Survey for England and Wales, 2017-2018: Teaching Dataset

Example data from textbooks

Two of the textbooks in this reading list also contain a wealth of small data examples that you can use alongside their content or independently.

You can find the data for chapters of Brian Fogarty's Quantitative Social Science Data with `R`: An introduction on the student companion website here: https://study.sagepub.com/fogarty

You can find the data for chapters of Kosuke Imai's Quantitative Social Science: An Introduction here: http://qss.princeton.press/student-resources-for-quantitative-social-science/

LEMMA

LEMMA is an online course developed by the University of Bristol. Despite its purpose being to train people to use Multilevel Modelling, it also has some excellent coverage and tutorials for introductory social statistics. It previously only had tutorials using MLWiN, but these have now grown to include tutorials using `R` and `Stata`.

To see an overview of the LEMMA course content, click here.

`R` for Data Science

`R` for Data Science has grown to be a go-to text for people learning `R`. Written by Hadley Wickham and Garrett Grolemund, it covers data manipulation and visualisation in detail. In addition, it also provides excellent practical advice on workflow and managing data science-related projects in `R`. Most importantly, this book follows Wickham and colleagues' 'tidy' approach to data science in `R`, which includes both the tidying of data into standard, rectangular formats, and the use of `tidyverse` tools, which emphasize the human readability of code.

The book is available for free online.

Supervised Machine Case Studies in `R`

Lastly, if you want some additional materials for a data scientist's approach to modelling in `R`, you might like Julia Silge's online course, Supervised Machine Learning Case Studies in `R`, and her associated screencasts. This course and screencasts provide an interesting view on statistical modelling and prediction from the perspective of a data scientist, and you will see some similarities and differences. In addition, these materials will also teach you how to use the packages in the `tidymodels` collection of packages.