Lancaster University

Quantitative Research Methods

January 16, 2023

Professor Patrick Rebuschat

p.rebuschat@lancaster.ac.uk

Our plan for today

Lancaster University

Introduction to this course

Quantitative research

- Two research strategies
- The PPDAC cycle
- The data science workflow

Introduction to R and RStudio

- First steps in R: Handout

Lancaster University

Introduction to this course

This course

Lancaster University

- A practical introduction to quantitative data analysis using R.

- R is a programming language and environment for statistical computing and graphics. The new standard.

This course

Lancaster University

We will cover two things:

1. How to design, analyze and interpret the results of statistical inquiry (up until and including multiple regression).
2. How to accomplish these tasks in R.

This course

Lancaster University

- Our Moodle site contains core information about the module (objectives, topics, suggested readings, assessment) as well as the materials we will use throughout the course.

This course

Lancaster University

- A **practical** introduction to quantitative data analysis using R.
- Each week, you need to work through a handout with practical tasks.
- **Please bring a laptop computer** to each class. You will need this to complete tasks during class.

Teaching strategy

Lancaster University

- We meet Mondays, 4pm to 6pm during Lent term.
- One exception: On Monday, February 13, there is no class. During this week, please complete the mid-term assignment on Moodle.
- Each session is roughly split between lecture material (explaining the main concepts of that week's work) and practical work (completing handouts using R).
- The balance between the theoretical and practical parts will not necessarily be 50:50 in every week.

Teaching strategy

Lancaster University

- You may also need to do some independent work with the R software, if you do not manage to finish the exercises during class time.
- For example, if you haven't completed the handout in class, you should complete this at home.
- This is important: We will assume that you know the material covered in the handout.

Our schedule

Lancaster University

Date	Topics
Jan 16	Introduction to quantitative research methods using R
Jan 23	Data management and data wrangling
Jan 30	Exploratory data analysis
Feb 6	Data visualization
Feb 13	No class, please complete the mid-term assignment (everybody)
Feb 20	Significance testing. Hypothesis tests for continuous variables: two groups.
Feb 27	Tests for discrete variables: Analysing contingency tables
Mar 6	Correlation and linear regression
Mar 13	Analysis of Variance (ANOVA) and tests for *N* groups
Mar 20	Multiple regression

Your responsibilities

Lancaster University

- To complete the assigned tasks during class and in between classes. Most tasks will be practical, requiring the use of R to complete statistical analyses.
- To engage in small group and whole class discussions during weekly sessions.
- To provide supportive, constructive peer feedback. This means, for example, helping each other while programming in R.

Learning R is like learning a new language.

Lancaster University

This means:

1. only you can do the learning, nobody can do this on your behalf, and
2. **the more you practice, the better you will become**.

So, please complete the tasks carefully, and write down any questions you have.

Autonomy and resilience

Lancaster University

- You will really on learn how to do quantitative research using R by doing it yourself **with your own data as part of your own project**.
- This course will prepare you for that, but it is impossible to prepare you for every single data challenge you will face.
- So, I am hoping that this course also further develops your ability to help yourself!

Assessment (for credit)

Lancaster University

Two assignments to assess your ability to analyze and interpret data set in R:

1. **Mid-term assignment** (1,500 words, 30% of total mark), due Friday, Feb 17, 2023, 12pm. Descriptive data analysis, numerical summaries, tables, graphics, etc
2. **Final assignment** (3,500 words, 70% of total mark), due Monday, April 24, 2023, 12pm. Confidence intervals and significance testing. Regression analysis and correlation.

Quantitative research

Quantitative research

Lancaster University

- A research strategy that focuses on quantification in the collection and analysis of data (Bryman, 2016).
- Quantitative research relies (depends) on computational tools and statistical methods.
- We collect and analyze large amounts of data to develop and test theories.
- Quantitative research tends to be the default across the sciences (cognitive, social, natural, clinical, etc.).
- Not better than qualitative research; complementary.

Two basic strategies in quantitative research

Lancaster University

1. Observational research
   - Simply observe and describe our phenomena of interest
   - Example: Systematic reviews, corpus analyses, correlational research, etc.
2. Experimental research
   - Systematically manipulate variables of interest and see what effect our manipulation has in the world
   - Example: Experimental studies in the lab or in the wild

Lancaster University

Observational research

Example: Systematic reviews

- A synthesis of the data on a given topic.
- Systematic reviews rely on quantitative methods to summarize, analyze and interpret data published on a given topic.
- Chalmers et al. (2021)
- Prisma Statement

Received: 1 June 2021 | Revised: 22 July 2021 | Accepted: 18 September 2021
DOI: 10.1111/lnc3.12440

REVIEW ARTICLE

WILEY

Sixty years of second language aptitude research: A systematic quantitative literature review

James Chalmers[1] | Susana A. Eisenchlas[2] | Andrew Munro[2] | Andrea C. Schalley[3]

[1]Library and Learning Services, Torrens University, Fortitude Valley, Queensland, Australia
[2]School of Humanities, Languages and Social Science, Griffith University, Nathan, Queensland, Australia
[3]Department of Language, Literature and Intercultural Studies, Karlstad University, Karlstad, Sweden

Correspondence
Andrea C. Schalley, Department of Language, Literature and Intercultural Studies, Karlstad University, 651 88 Karlstad, Sweden.
Email: andrea.schalley@kau.se

Abstract
Second language (L2) aptitude has been broadly defined as the rate and ease of initially acquiring a second language. Historically, L2 aptitude has been understood as a stable trait that predetermined L2 achievement, regardless of individual learners' efforts to acquire an L2. This traditional view of L2 aptitude as fixed and stable has led to it being a relatively neglected area of research within second language acquisition (SLA) studies. The little research that was in fact conducted was diagnostic in nature, and mostly used tests such as the Modern Language Aptitude Test (MLAT) to select potentially gifted L2 learners. Given that six decades have passed since the publication of the MLAT, now is a good time to revisit the literature and investigate whether L2 aptitude continues to be viewed as an individual difference of little interest to SLA research. While summative literature reviews of L2 aptitude research have been written, few systematic reviews exist. This article conducts a systematic quantitative literature review (SQLR) to provide a principled, comprehensive and reproducible synthesis of research into L2 aptitude published over the last 60 years (1959–2019). In this SQLR, close to one

Example: Corpus approach

Lancaster University

- A corpus is a (very) large collection of (machine-readable) texts.
- We use quantitative methods to analyze these texts.
- Lancaster's CASS
- Summer School
- LancsBox

Example: Corpus approach

Lancaster University

- If A is present, then B tends to be present, too, and vice versa. But we do not know if one causes the other.

Disadvantage

- Correlational research cannot tell us anything about causality! All we can say is that two things tend to co-occur.
- But: Sometimes, we cannot do experimental on certain topics, e.g. for ethical reasons, so correlational research has to suffice.

Slevc and Miyake (2006)

PSYCHOLOGICAL SCIENCE

Research Report

Individual Differences in Second-Language Proficiency

Does Musical Ability Matter?

L. Robert Slevc[1] and Akira Miyake[2]

[1]*University of California, San Diego, and* [2]*University of Colorado, Boulder*

ABSTRACT—*This study examined the relation between musical ability and second-language (L2) proficiency in adult learners. L2 ability was assessed in four domains: receptive phonology, productive phonology, syntax, and lexical knowledge. Also assessed were various other factors that might explain individual differences in L2 ability, including age of L2 immersion, patterns of language use and exposure, and phonological short-term memory. Hierarchical regression analyses were conducted to determine if musical ability explained any unique variance in each domain of L2 ability after controlling for other relevant factors. Musical ability predicted L2 phonological ability (both receptive and productive) even when controlling for other factors, but did not explain unique variance in L2 syntax or lexical knowledge. These results suggest that musical skills may facilitate the acquisition of L2 sound structure and add to a growing body of evidence linking language and music.*

People exhibit substantial individual differences in second-language (L2) proficiency. Learners' age of immersion is known to influence their ultimate level of L2 ability, but even when this factor is taken into account, striking individual differences still exist, especially among people who started acquiring an L2 after childhood. Although some adult L2 learners attain near-native proficiency, others speak with strong foreign accents and frequent grammatical errors long after their immersion in the L2. Why do some adult learners acquire an L2 more successfully than others? What characteristics differentiate good L2 learners from not-so-good ones?

One common answer to these questions (at least among laypersons) is that musical ability is an important determinant of such variation. According to this account, being skilled at music means having a "good ear" for analyzing and discriminating foreign speech sounds, so that musically talented individuals are better equipped than other people to pick up various aspects of an L2, especially pronunciations of L2 sounds.

There are good scientific reasons to expect a link between musical ability and L2 proficiency. First, like language, music is a human universal consisting of perceptually discrete elements organized into hierarchically structured sequences (Patel, 2003; Sloboda, 1985). Second, neuropsychological evidence indicates that some brain regions often assumed to be language-specific (e.g., the inferior frontal gyrus, including Broca's area) are also implicated in musical processing (Levitin & Menon, 2003; Maess, Koelsch, Gunter, & Friederici, 2001; Tillmann, Janata, & Bharucha, 2003), as are certain event-related potential signatures of language processing (Patel, Gibson, Ratner, Besson, & Holcomb, 1998). Third, musical ability can predict aspects of first-language (L1) verbal ability, such as reading ability in children (Anvari, Trainor, Woodside, & Levy, 2002; Atterbury, 1985).

Given such evidence for a music-language link, it is surprising that little evidence is available regarding the hypothesized relation between musical ability and L2 proficiency. Skehan (1989) provided a detailed monograph-length review of individual differences in L2 acquisition, but had nothing to say about the relevance (or irrelevance) of musical ability to L2 proficiency. Moreover, several studies failed to find a clear link between self-ratings of musical ability and L2 ability (Flege, Munro, & MacKay, 1995; Flege, Yeni-Komshian, & Liu, 1999; Tahta, Wood, & Loewenthal, 1981; Thompson, 1991). Although two recent studies reported a positive link between musical ability and aspects of L2 pronunciation ability (Nakata, 2002; Tanaka & Nakamura, 2004), they did not control for effects of other potentially correlated variables. Thus, it is unclear whether musical ability makes a unique contribution to explaining variance in L2 proficiency.

Address correspondence to L. Robert Slevc, UCSD Department of Psychology 0109, 9500 Gilman Dr., La Jolla, CA 92093-0109, e-mail: slevc@psy.ucsd.edu.

TABLE 1

Dependent Measures and Descriptive Statistics for the Tasks Used in This Study

Task	Dependent measure	Mean (*SD*)	Range
	Criterion (L2 proficiency) variables		
Receptive phonology[a]			
Word level	Number correct (of 26)	21.6 (2.8)	13–26
Sentence level	Number correct (of 26)	20.1 (3.2)	12–26
Passage level	Number correct (of 43)	32.1 (6.1)	16–43
Productive phonology[a]			
Word level	Raters' number correct (of 26)	22.1 (2.2)	17–26
Sentence level	Raters' number correct (of 26)	22.1 (2.2)	17–26
Passage level	Average ratings from 2 raters (1 = strong accent, 9 = none)	4.4 (1.9)	1.2–8.2
Syntax			
Grammaticality judgments	Number correct (of 72)	46.5 (6.4)	36–67
Lexical knowledge[a]			
Vocabulary	Number correct (of 25)	15.4 (4.4)	8–25
Listening comprehension	Number correct (of 30)	22.9 (5.9)	8–30
	Predictor variables		
Age-related variables			
Age of arrival	Age of arrival in the United States (in years)	25.0 (7.1)	11–47
Length of residence	Length of residence in the United States (in years)	4.4 (5.3)	0.5–25
Self-reported L2 use and exposure[a]			
When first arrived			
Use	Percentage English use (vs. Japanese)	64.2 (29.7)	5–100
Exposure	Percentage English exposure (vs. Japanese)	68.1 (27.3)	10–100
At time of testing			
Use	Percentage English use (vs. Japanese)	64.4 (22.6)	10–99
Exposure	Percentage English exposure (vs. Japanese)	68.6 (21.5)	20–100
Motivation to use L2[a]			
When first arrived	Average self-ratings from 4 questions (1 = not motivated, 5 = very motivated)	3.0 (1.1)	0.75–5
At time of testing	Average self-ratings from 3 questions (1 = not motivated, 5 = very motivated)	4.5 (0.6)	2.5–5
Nonverbal intelligence			
Cattell Culture Fair Test	Summed scores from the 4 subscales (of 50)	28.4 (4.7)	17–35
Phonological STM[a]			
Japanese digit span	Number of digits recalled (of 168)	113.7 (22.6)	71–166
Nonword repetition	Number of nonwords repeated accurately (of 40)	25.0 (4.7)	15–37
Musical ability[a]			
Chord analysis	Number correct (of 20)	11.1 (2.8)	5–19
Pitch change	Number correct (of 30)	20.2 (4.8)	13–30
Tonal memory	Number correct (of 30)	20.4 (3.8)	13–28
Tonal memory production	Number correct (of 75)	42.0 (20.3)	0–72

Note. L2 = second language; STM = short-term memory.

[a]For each of these measures, scores on almost all the component measures were highly correlated, and a *z*-score aggregate was calculated. The pattern of results did not change when all analyses were rerun after eliminating component measures that did not correlate significantly with the others.

TABLE 3

Summary of Hierarchical Regression Results for Four Different Domains of Second-Language (L2) Proficiency

Step and independent variable	R^2	ΔR^2	*df*	*F*	Final β
L2 receptive phonology					
Step 1: age of arrival	.06	.06	1, 48	2.92†	−.14
Step 2: length of residence	.23	.17	1, 47	10.25**	.23†
Step 3: language use and exposure	.23	.00	1, 46	0.09	.08
Step 4: phonological short-term memory	.30	.07	1, 45	4.90*	.23†
Step 5: musical ability	.42	.12	1, 44	8.82*	.37**
L2 productive phonology					
Step 1: age of arrival	.05	.05	1, 48	2.49	−.13
Step 2: length of residence	.29	.24	1, 47	15.59**	.37*
Step 3: language use and exposure	.29	.00	1, 46	0.46	.10
Step 4: phonological short-term memory	.30	.01	1, 45	0.31	.03
Step 5: musical ability	.38	.08	1, 44	5.53*	.30*
L2 syntax					
Step 1: age of arrival	.06	.06	1, 48	3.06†	−.18†
Step 2: length of residence	.32	.26	1, 47	18.04**	.31*
Step 3: language use and exposure	.36	.04	1, 46	3.16†	.28*
Step 4: phonological short-term memory	.51	.15	1, 45	13.90**	.39**
Step 5: musical ability	.53	.02	1, 44	1.34	.13
L2 lexical knowledge					
Step 1: age of arrival	.21	.21	1, 48	12.94**	−.42**
Step 2: length of residence	.39	.18	1, 47	13.28**	.27*
Step 3: language use and exposure	.48	.09	1, 46	8.30*	.35**
Step 4: phonological short-term memory	.52	.04	1, 45	3.92†	.21†
Step 5: musical ability	.52	.00	1, 44	0.14	.04

Note. "Final β" indicates the standardized beta weight for each variable when controlling for all other variables (i.e., standard beta weights in Step 5).

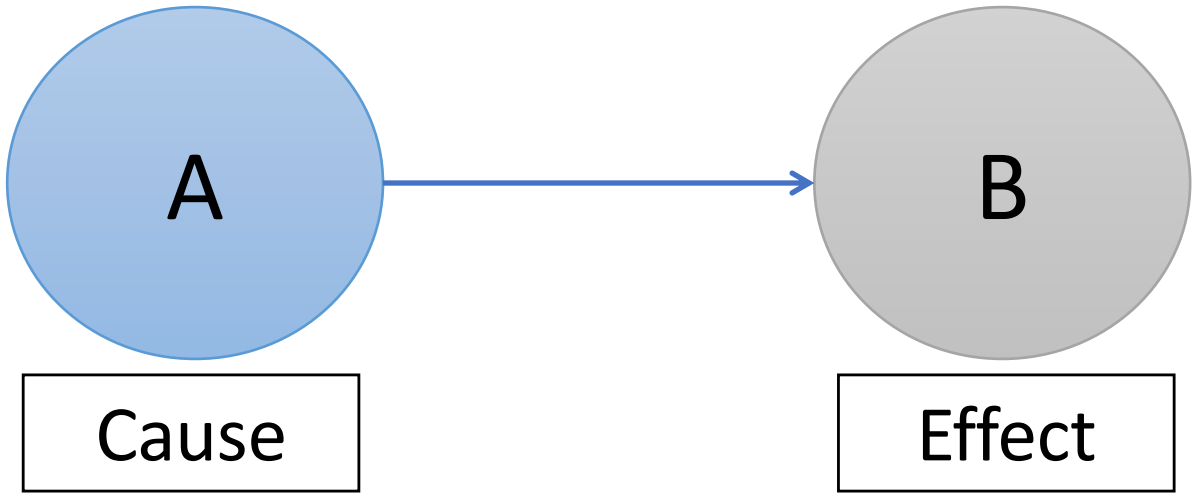
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .005$.

Experimental research

Example: Experimental research

Lancaster University

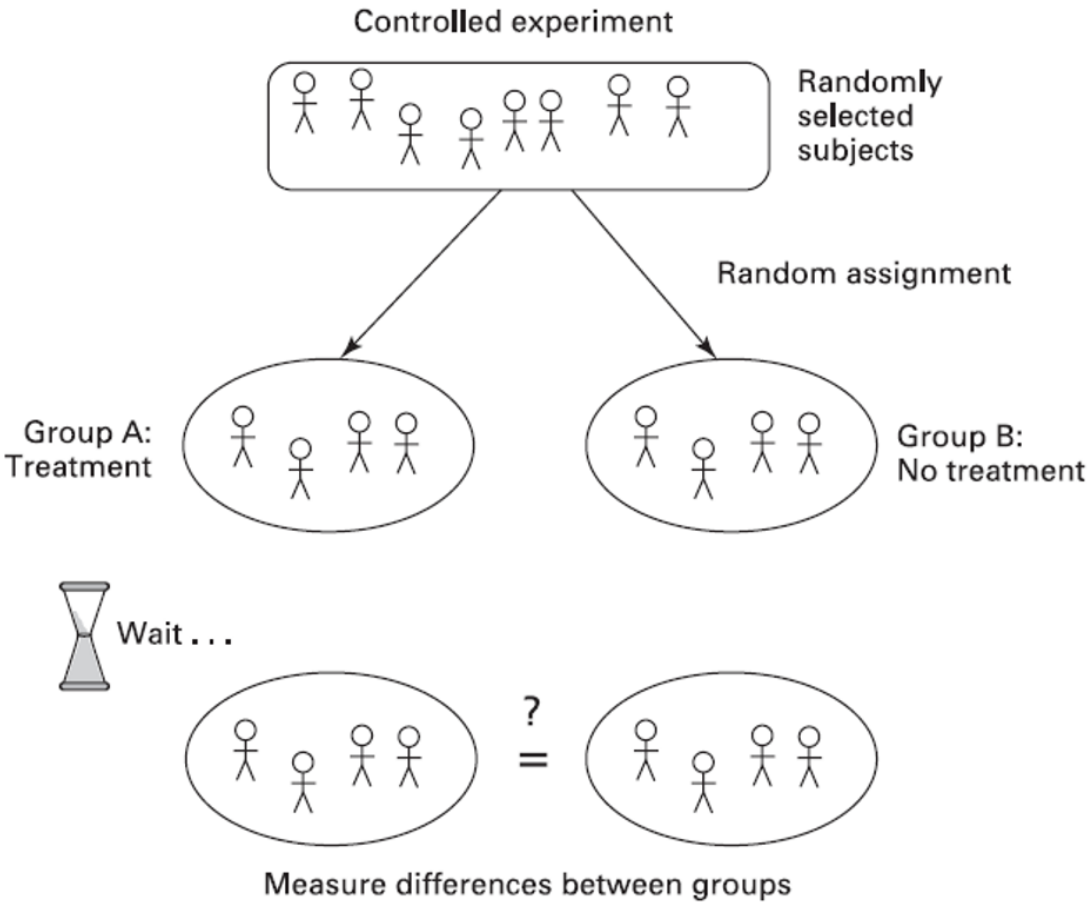
In experimental research, we try to discover causal relationships btw variables.



Example: Experimental research

Lancaster University

Reproduced from L



Lancaster University

Introduction to R and RStudio

What is R?

Lancaster University

- R is a programming language and environment doing statistics and data analysis.
- There are many programming languages that could do the trick (e.g., Python).
- But R is the particularly well-suited to perform the major components outlined in the data science workflow.

Advantages of R

Lancaster University

A very powerful tool for data analysis

- Range and depth of statistical analyses that can be done in R are immense.
- R has a standard set of “packages” that contains the entire repertoire of widely used statistical methods.
- R has an extensive graphics library for visualizing your data.
- R is extensible. You can add over 16,000 packages, which allow you to do more things and more easily.

Advantages of R

Lancaster University

Open source software

- R is free and open source software. Anyone can use it and develop it further.
- R, like many other open source software, has a large and self-sustaining community of users and developers.

Advantages of R

Lancaster University

Very widely used (the new standard)

- R is the standard computing platform in statistics. Almost every new or existing technique developed by statisticians is made available as a package in R.
- As a result, R has become the new standard in data science and in the social and behavioral sciences.

Why R and not SPSS?

Lancaster University

- Computing skills (e.g., reading and writing code) are an integral part of modem data analysis.
- Statistics programs with a graphical user interface (GUI) are sufficient for novices, but they are limited and inefficient for more advanced analyses in comparison to programming languages such as R.
- R and RStudio facilitate open and reproducible science (RMarkdown).
- Knowledge of R will improve your research and make you more employable.
- R is free and open source, and SPSS is not…

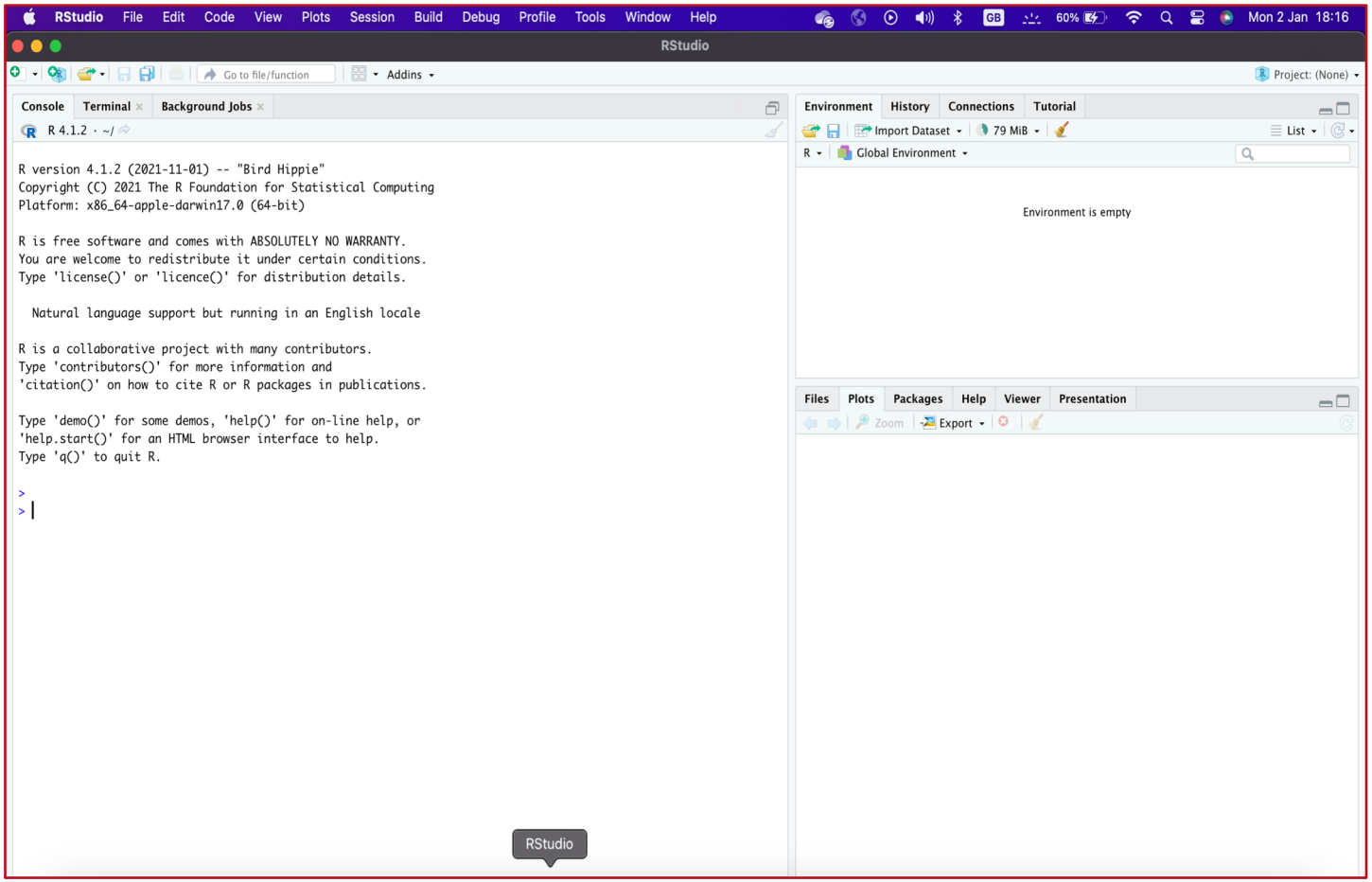
A guided tour of RStudio

Lancaster University

- Once you have installed R and RStudio, all of your computing will be done via RStudio.
- We don’t use R directly, it just runs in the background while we work with RStudio.

When you open RStudio for the first time, this is the typical layout.

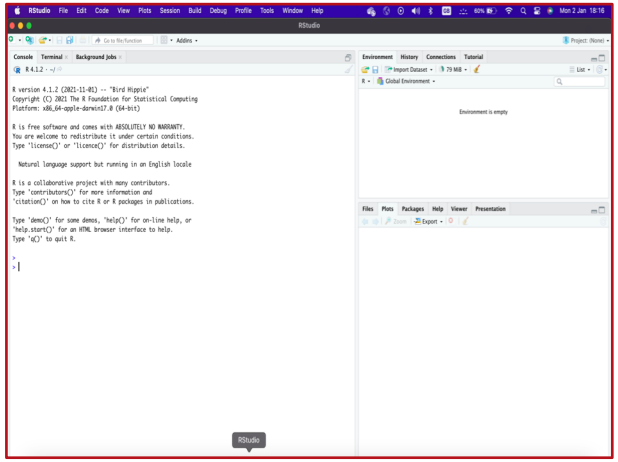
Lancaster University



A guided tour of RStudio

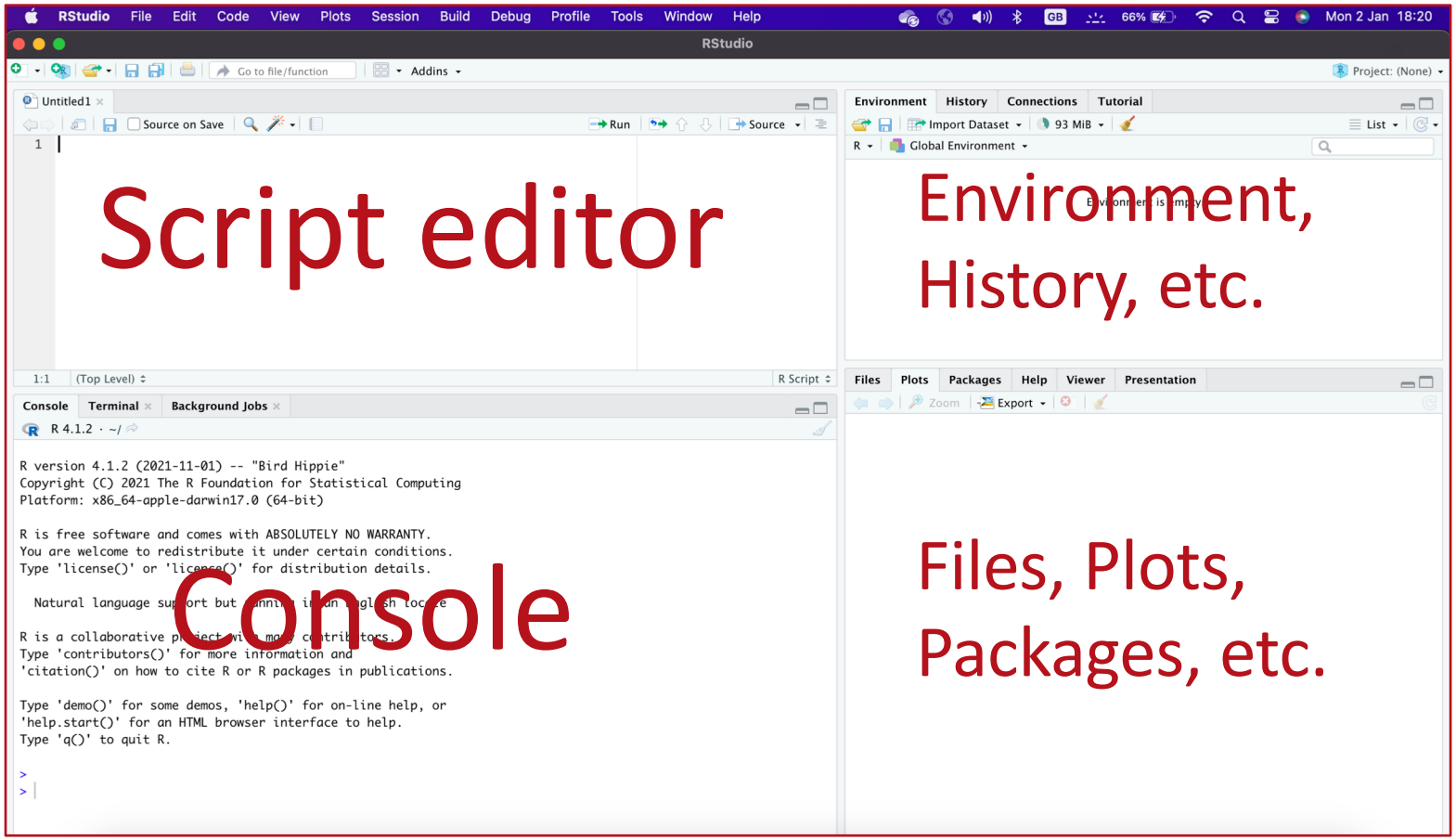
Lancaster University

- If RStudio looks like this, then please open a fourth important window, the script editor.

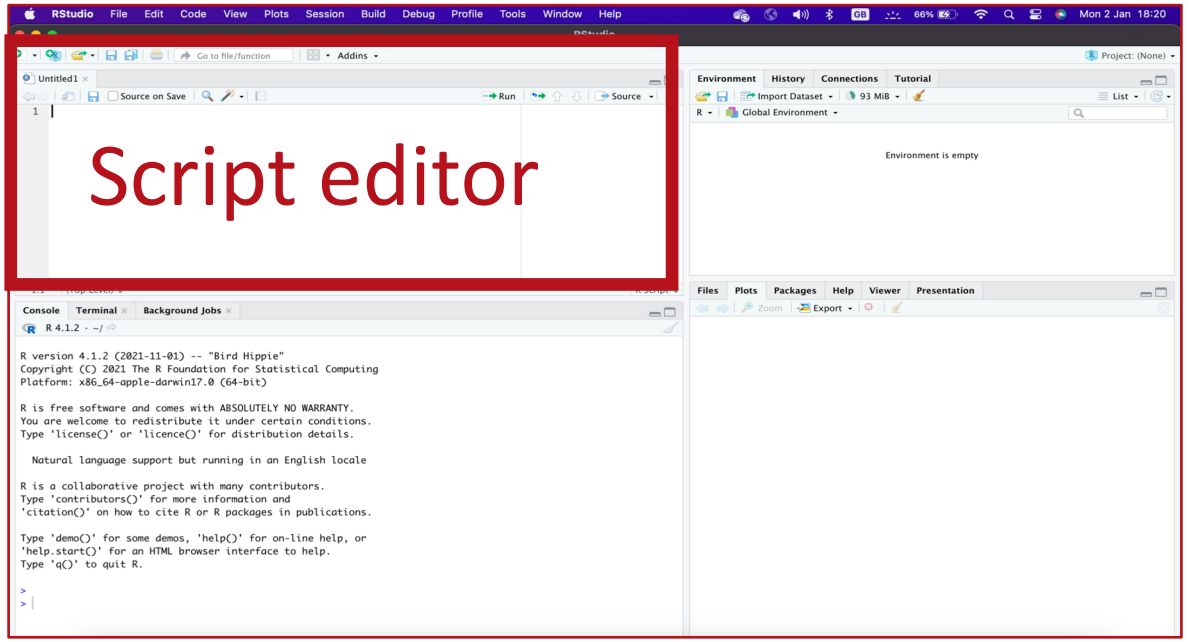


- To open the script editor, please go to File > New File > R script.
- Or use the key command Ctrl+Shift+N (Windows) or Cmd+Shift+N (Mac).
- (Get used to using shortcuts.)

Rstudio should now look like this



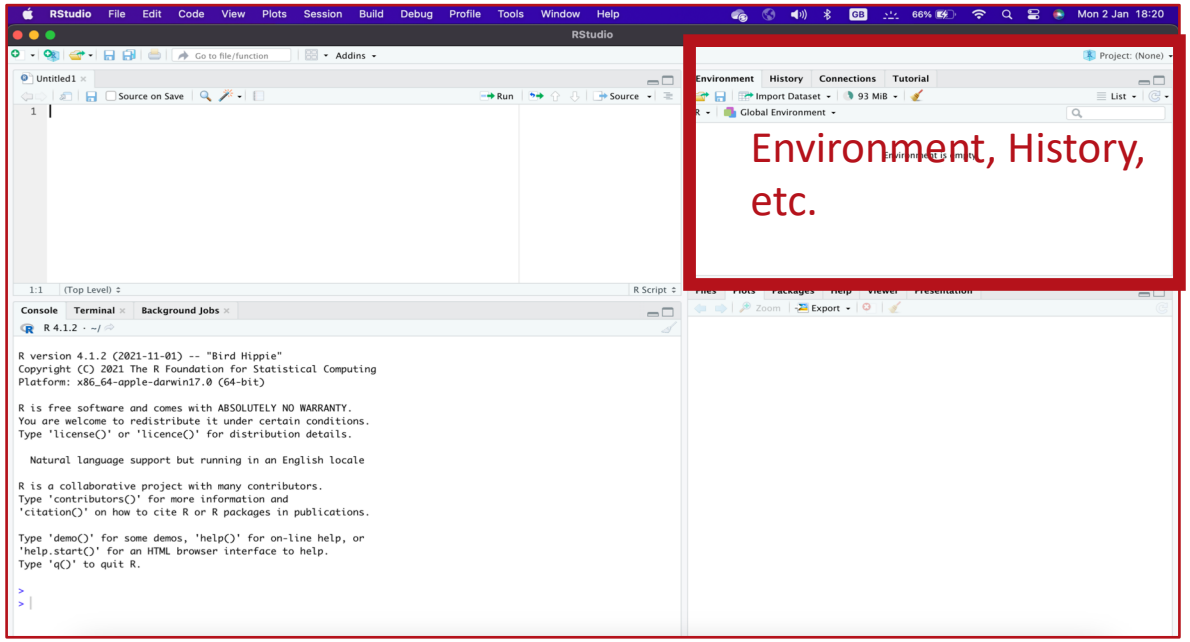
Script editor



This is where we write **scripts** of R commands.

- Scripts are a sequence of commands; these can be saved for later use and for sharing with other researchers.
- In the script editor, you can execute individual lines of R commands, specific sections of R commands, i.e. several lines, or the entire scripts (all the commands).

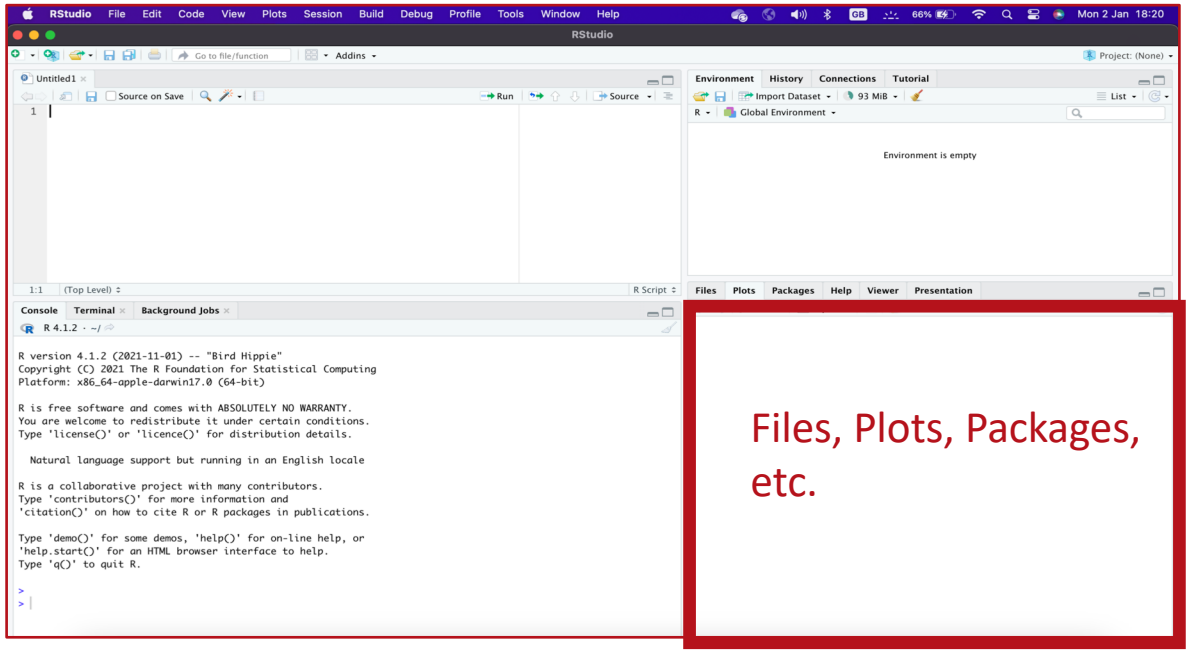
Environment, history, etc.



This has several tabs. The most important ones are:

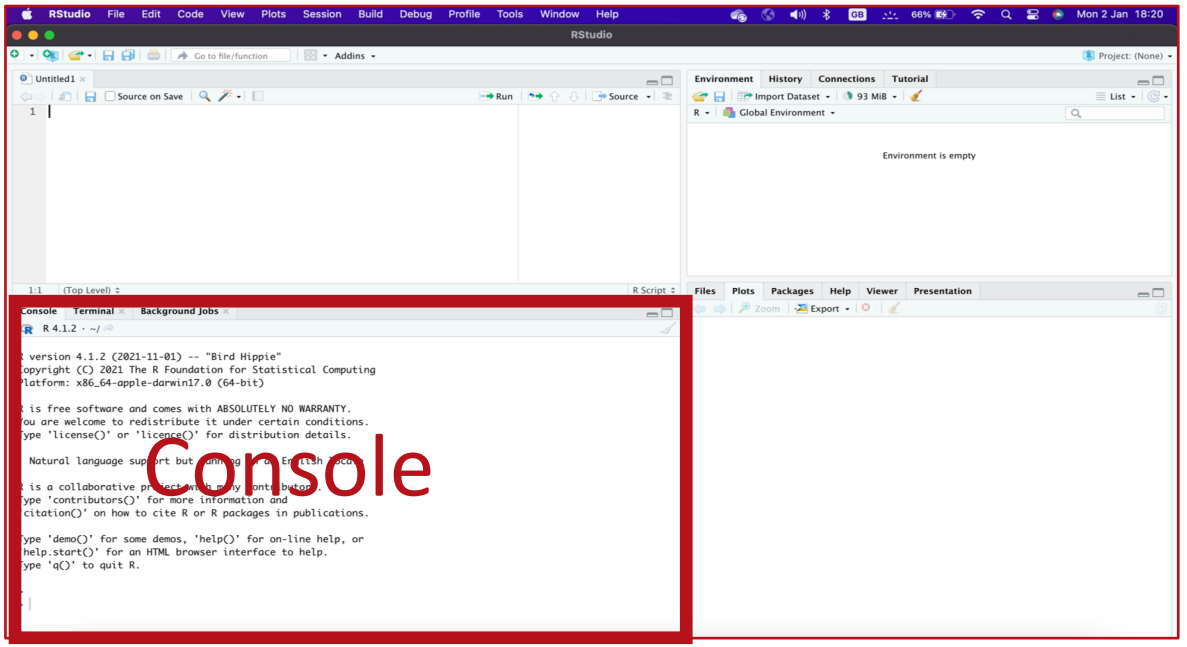
- Environment (lists variables and data structures from our current work session and allows us to easily import data files) and
- History (lists all the R commands we have typed in the session).

Files, Plots, Packages, etc.



- The Files tab is a file browser in which we can create, delete, view files and directories.
- The Plots tab is where our graphics will appear.
- The Packages tab shows us the packages we have installed in our session.
- The Help tab displays help pages for an R command or package.

Console



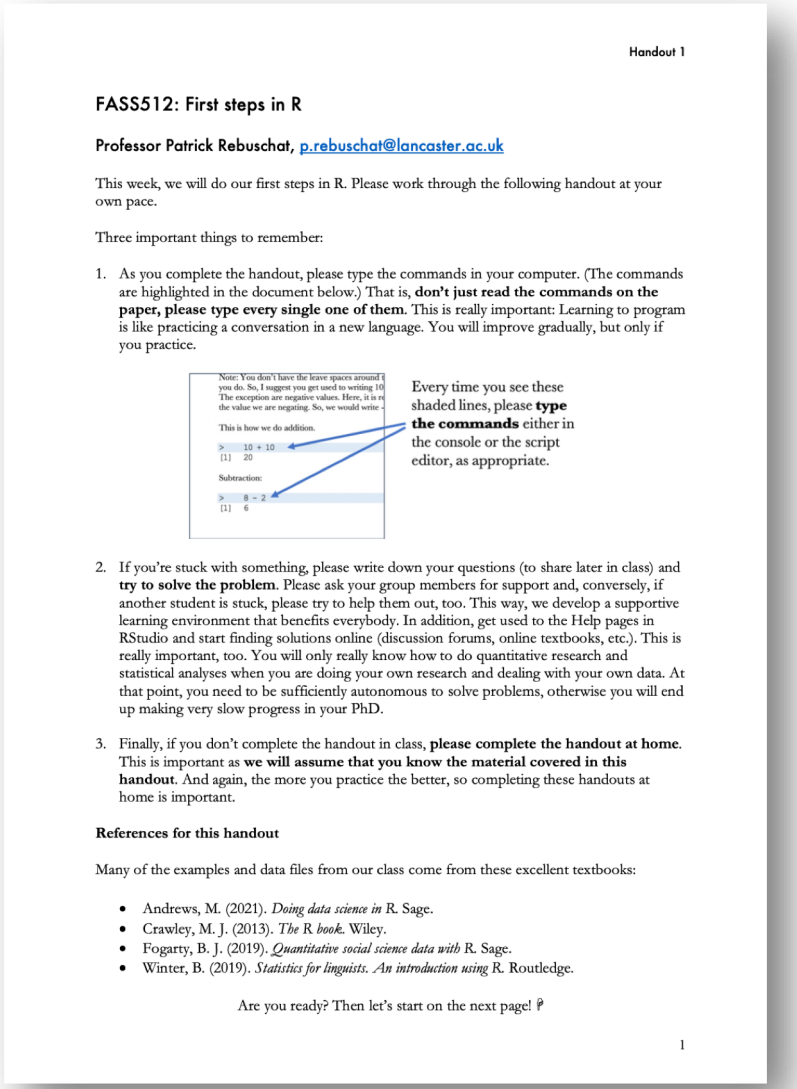
- This is the most important window.
- It also has multiple tabs, but we'll be mostly using the Console tab.
- This is where we can type our R commands next to the >, press Enter to execute them and see the output just below.
- Today, we will just use the Console, not the Script editor.

Lancaster University

First steps in R

Handout 1

Please download this handout from Moodle, then work through the document at your own pace.



Handout 1

FASS512: First steps in R

Professor Patrick Rebuschat, p.rebuschat@lancaster.ac.uk

This week, we will do our first steps in R. Please work through the following handout at your own pace.

Three important things to remember:

1. As you complete the handout, please type the commands in your computer. (The commands are highlighted in the document below.) That is, **don't just read the commands on the paper, please type every single one of them**. This is really important: Learning to program is like practicing a conversation in a new language. You will improve gradually, but only if you practice.

2. If you're stuck with something, please write down your questions (to share later in class) and **try to solve the problem**. Please ask your group members for support and, conversely, if another student is stuck, please try to help them out, too. This way, we develop a supportive learning environment that benefits everybody. In addition, get used to the Help pages in RStudio and start finding solutions online (discussion forums, online textbooks, etc.). This is really important, too. You will only really know how to do quantitative research and statistical analyses when you are doing your own research and dealing with your own data. At that point, you need to be sufficiently autonomous to solve problems, otherwise you will end up making very slow progress in your PhD.

3. Finally, if you don't complete the handout in class, **please complete the handout at home**. This is important as **we will assume that you know the material covered in this handout**. And again, the more you practice the better, so completing these handouts at home is important.

References for this handout

Many of the examples and data files from our class come from these excellent textbooks:

- Andrews, M. (2021). *Doing data science in* R. Sage.
- Crawley, M. J. (2013). *The R book*. Wiley.
- Fogarty, B. J. (2019). *Quantitative social science data with* R. Sage.
- Winter, B. (2019). *Statistics for linguists. An introduction using* R. Routledge.

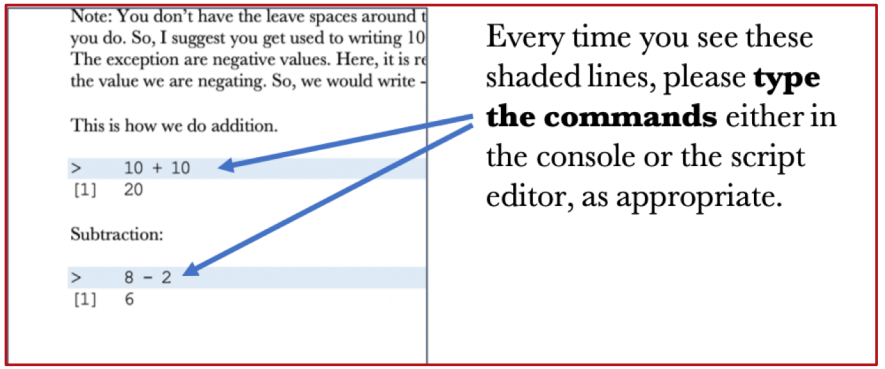
Are you ready? Then let's start on the next page!

1

Type the command lines

Lancaster University

- As you complete the handout, **please type every single one of them**.
- This is really important: Learning to program is like practicing a conversation in a new language. You will improve gradually, but only if you practice.



Try to solve the problem

Lancaster University

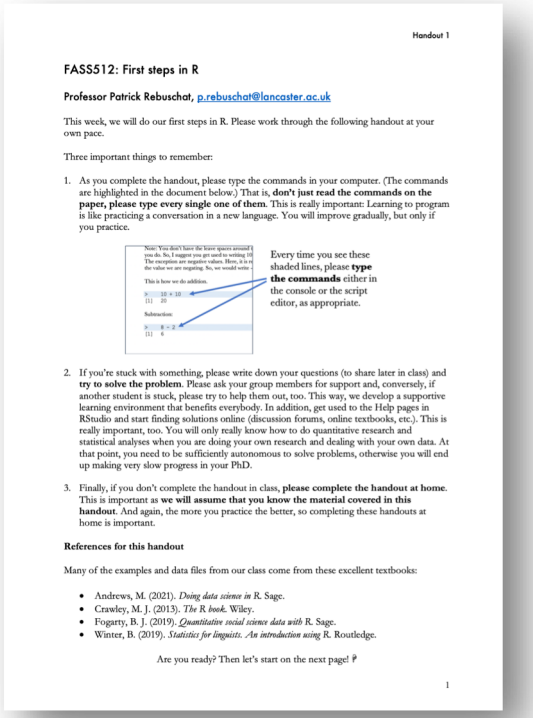
- If you're stuck with something, please write down your questions (to share later in class) and **try to solve the problem**.
- Please ask your group members for support and, conversely, if another student is stuck, please try to help them out, too.
- Get used to the Help pages in RStudio and start finding solutions online (discussion forums, online textbooks, etc.).

Complete the handout at home

- If you don’t complete the handout in class, **please complete the handout at home**.
- This is important as **we will assume that you know the material covered in this handout**.
- And again, the more you practice the better, so completing these handouts at home is important.

Steps in the handout

1. The R console
2. Using R as a calculator
3. Variables
4. Vectors
5. Data frames
6. Functions



Handout 1

FASS512: First steps in R

Professor Patrick Rebuschat, p.rebuschat@lancaster.ac.uk

This week, we will do our first steps in R. Please work through the following handout at your own pace.

Three important things to remember:

1. As you complete the handout, please type the commands in your computer. (The commands are highlighted in the document below.) That is, **don't just read the commands on the paper, please type every single one of them**. This is really important: Learning to program is like practicing a conversation in a new language. You will improve gradually, but only if you practice.

Every time you see these shaded lines, please **type the commands** either in the console or the script editor, as appropriate.

2. If you're stuck with something, please write down your questions (to share later in class) and **try to solve the problem**. Please ask your group members for support and, conversely, if another student is stuck, please try to help them out, too. This way, we develop a supportive learning environment that benefits everybody. In addition, get used to the Help pages in RStudio and start finding solutions online (discussion forums, online textbooks, etc.). This is really important, too. You will only really know how to do quantitative research and statistical analyses when you are doing your own research and dealing with your own data. At that point, you need to be sufficiently autonomous to solve problems, otherwise you will end up making very slow progress in your PhD.

3. Finally, if you don't complete the handout in class, **please complete the handout at home**. This is important as **we will assume that you know the material covered in this handout**. And again, the more you practice the better, so completing these handouts at home is important.

References for this handout

Many of the examples and data files from our class come from these excellent textbooks:

- Andrews, M. (2021). *Doing data science in* R. Sage.
- Crawley, M. J. (2013). *The R book*. Wiley.
- Fogarty, B. J. (2019). *Quantitative social science data with* R. Sage.
- Winter, B. (2019). *Statistics for linguists. An introduction using* R. Routledge.

Are you ready? Then let's start on the next page!

Lancaster University

Questions?

Lancaster University

Quantitative Research Methods

January 16, 2023

Professor Patrick Rebuschat

p.rebuschat@lancaster.ac.uk