Programming for Practical Data Analysis

Prof. Dr. Johannes Binswanger

Fall 2016

Lecture 1: Introduction


Organization of the course

LANGUAGE

- These slides are written in English, as is the online script and as will be the comments in the code that we are going to write.
- An important learning goal is that you will be able to help yourself in solving new programming and data analysis problems.
- For this you will have to google (a lot).
- And if you google in English, you get a lot more results.
- And so you need to know the terms in English…

STRUCTURE OF THE COURSE

- In total, we have 12 lecture units.
- During the first 9 units, we'll code and analyze together in class.
- At the end of October, you will choose the topics for your group projects (more on this on the next slide).
- During lectures 10 and 11, you will present the results of your group projects.
- During the last lecture you will write a written assignment.

GROUP ASSIGNMENTS

- For the group assignments, you'll work in groups of 3.
- You will analyze a data set that you can choose yourself (within limits).
- Every group needs to choose a different data set, one that you haven't already worked on in other courses.
- You'll get instructions on the type of analysis you should perform with your data.
- You'll submit a (short) written document.
- And you'll present the results in the form of an elevator pitch (about 5 minutes, depending on how many groups we will be).
- The presentation, together with the written document, counts for 50% of your grade for this course.

WRITTEN ASSIGNMENT

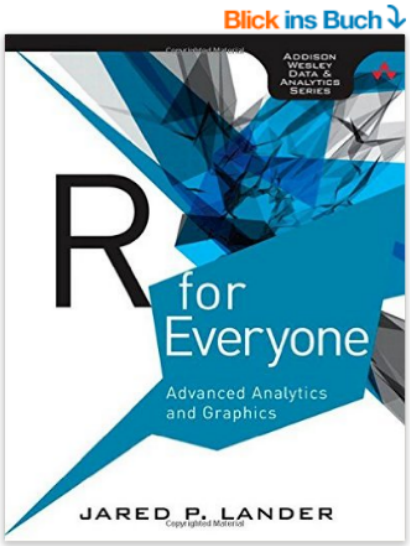
- You'll write a written assignment during the last lecture, on your own device.
  - You will write some short programs or pieces of code, copy the code and the results into a Word document, save the latter as pdf and upload it to StudyNet.
- You will have to solve some problems you haven't seen before.
- But you will have learned enough about coding and googling to be able to solve these problems.
- Technically, this is not an exam but just a home assignment that you happen to write during the last class!

HOME ASSIGNMENTS DURING THE COURSE

- Whenever suitable, there will be a home assignment that you need to solve for the following lecture.
- This will be available via the StudyNet «Test» function.

BOOK

Fremdsprachige Bücher › Computer & Internet › Programmieren



R for Everyone: Advanced Analytics and Graphics (Addison-Wesley Data and Analytics)

(Englisch) Taschenbuch – 19. Dezember 2013

von Jared P. Lander ▾ (Autor)

Geben Sie die erste Bewertung für diesen Artikel ab

▸ Alle Formate und Ausgaben anzeigen

Kindle Edition	**Taschenbuch**
EUR 23,59	**EUR 41,68**

Lesen Sie mit unserer **kostenfreien App**

Lieferung Dienstag, 20. Sept.: Bestellen Sie innerhalb 17 Stunden und 40 Minuten per **AmazonGlobal Express-Zustellung** an der Kasse. Siehe Details.

9 neu ab EUR 15,65 | 5 gebraucht ab EUR 32,90

Statistical Computation for Programmers, Scientists, Quants, Excel Users, and Other Professionals

- Purchasing a personal copy is strongly suggested!

FEEDBACK

- If you want to give anonymous feedback, you can do so using the “Forum” function on StudyNet.


Learning Goals

LEARNING GOALS

- Developing basic programming skills that are also useful for other programming languages.
- Being able to analyze data that are structured in “rows and columns”, i.e. data that can be arranged in spreadsheet format (all classical data sets come in this format).
- Automation of reporting of results.
- IT and data science are very dynamic fields. In 5 years, R may have many new functions or there may be other popular programming languages. An important goal is that you will be able to keep learning and teach new things to yourself.

LEARNING GOALS

- Finally, this course is not “just” about programming.
- Programming is never a goal in itself.
  - You program because you want to analyze some data.
  - You want to analyze some data because you want to answer an important question.
  - You may want to answer that question because you want to make an important decision (e.g. whether, as a government, to change the tax system; or, as a company, to invest in a certain market, etc.).
- --> It is always important to be aware of the context of your analysis.
- This is also an important learning goal for this course.


Why R?

WHY R?

- Here are a couple of links about the popularity of programming languages that are fun to visit:
  - http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages (first hit if you google “most popular programming languages 2016”!)
  - https://www.fastcompany.com/3030716/the-9-best-languages-for-crunching-data
  - http://www.kdnuggets.com/2015/05/r-vs-python-data-science.html
  - http://thenextweb.com/dd/2016/04/08/start-using-python-andor-r-data-science-one-best/#gref
- If you google “most important programming language for data science” you’ll get a lot more.

WHY R?

- After some googling (or talking to people in the field), you will easily notice that there are two languages that stand out in popularity for modern data science.
  - No, it's not Stata…
  - Nor Matlab…
  - Nor C, C++, Java (these are extremely popular in general, but a little less so for data science)
  - It's R and Python

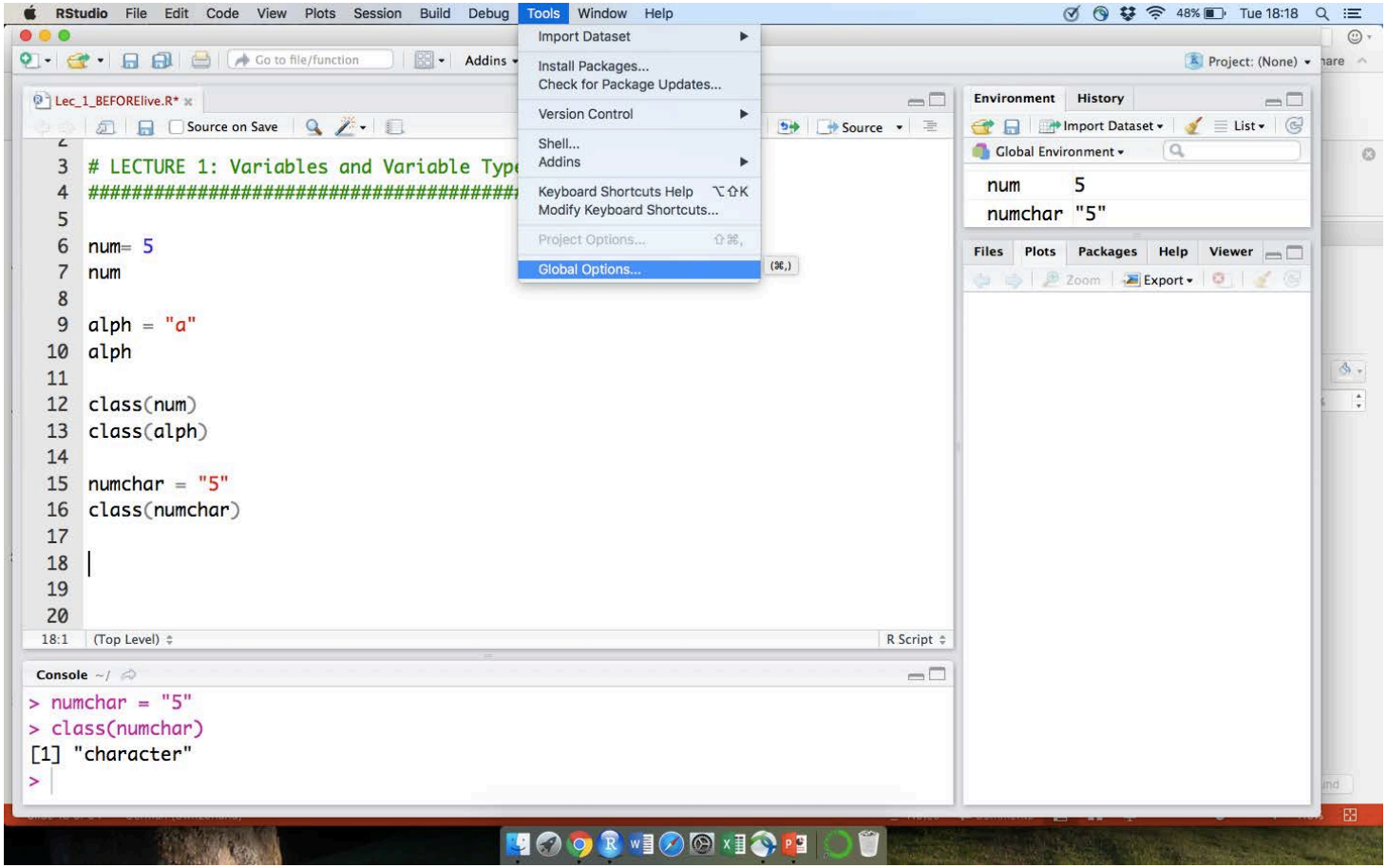
WHY R?

- There is no clear reason to either prefer R or Python.
- In fact, it's best to learn both.
- If we have time, we'll briefly look into Python. You will see, knowing R makes it a lot easier to learn and understand Python.
- Of course, the opposite is true as well.
- In the end, you just have to start with *something*.
- Python is also a general-purpose programming language, while R is more specialized to data analysis and simulations.
- Because of this, R may be slightly more natural as a starting point, at least for economists.
    - But it does not really matter, I could also just flip a coin; and my coin flip was showing the letter R…


Essentials about RStudio

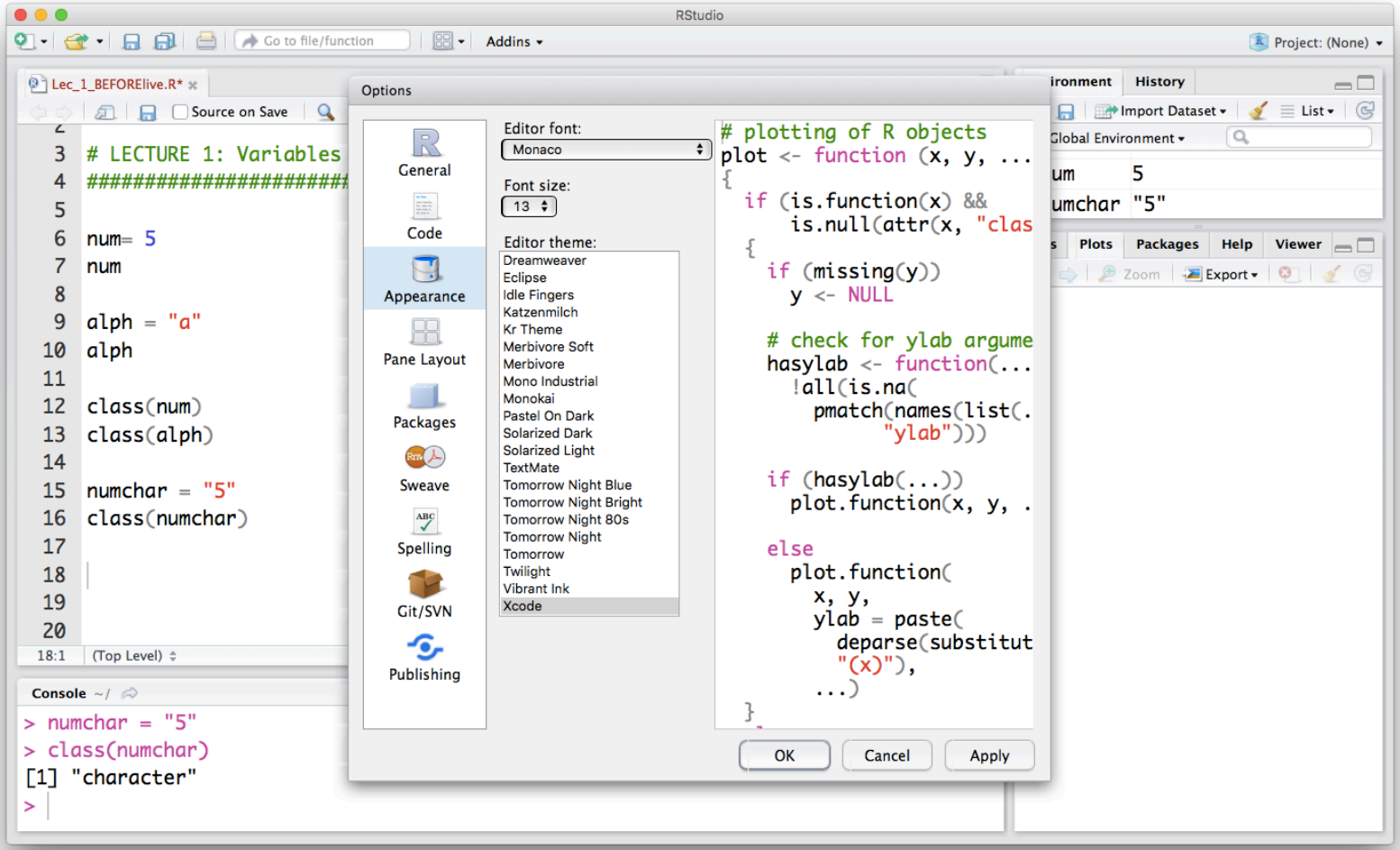
CUSTOMIZING RSTUDIO

- You can customize the looks of Rstudio via Tools --> Global Options.



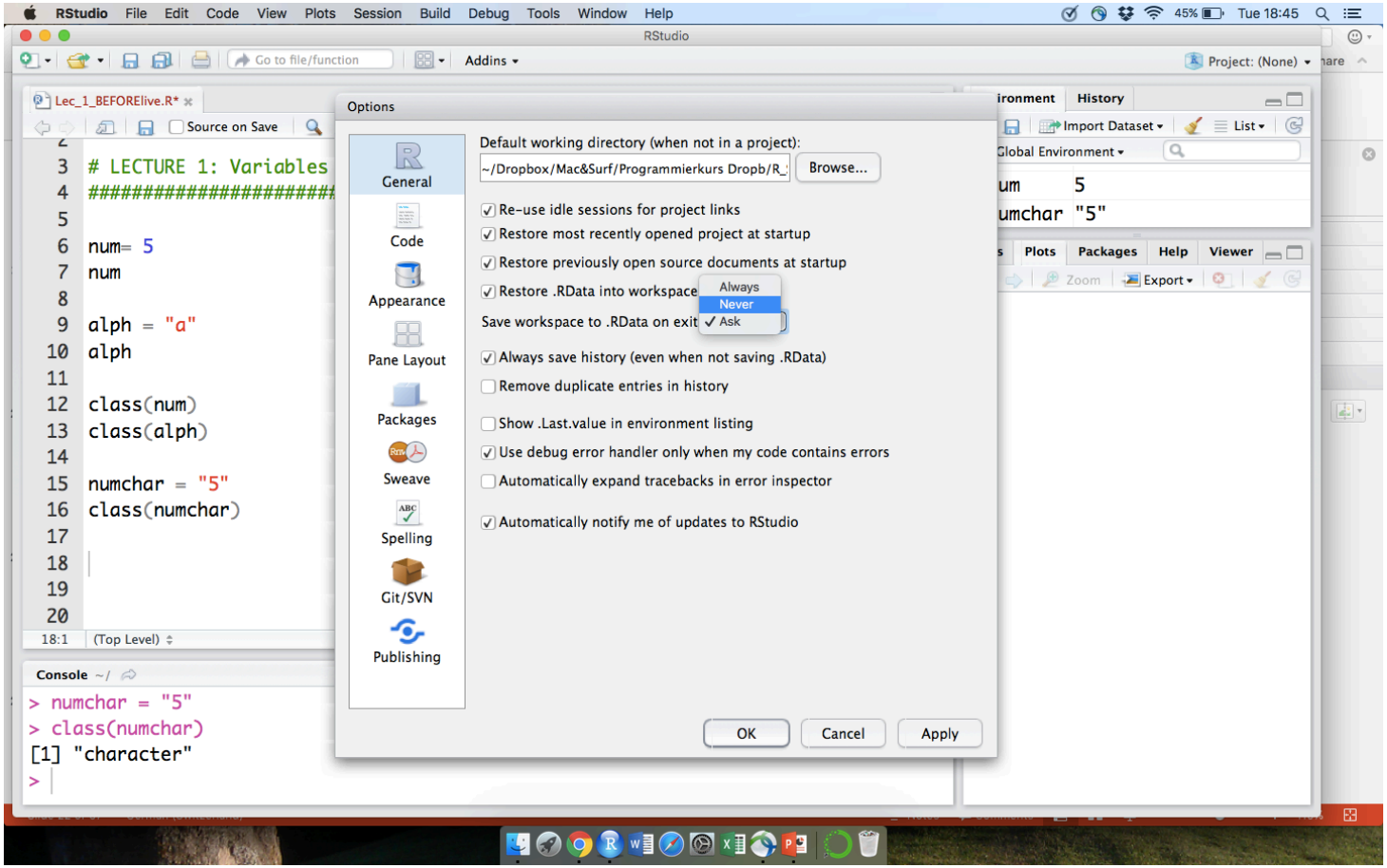
CUSTOMIZING RSTUDIO

- Under “Appearance” you can change the font size; and also choose a look you like for the editor.



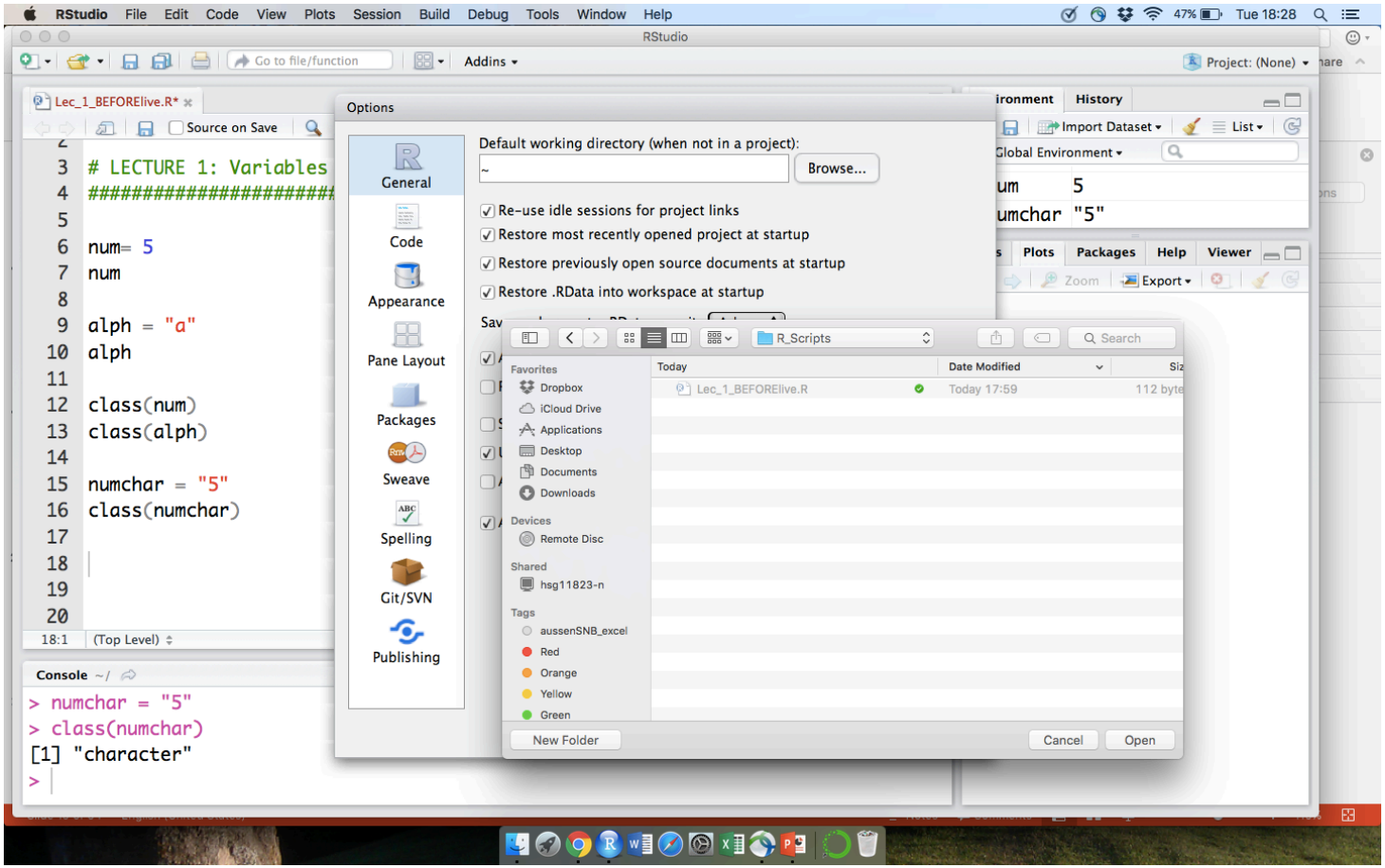
CUSTOMIZING RSTUDIO

- It is convenient to change some general settings:
  - "Save workspace to .Rdata on exit": choose NEVER
  - Uncheck "Always save history…"



CUSTOMIZING RSTUDIO

- It also makes sense to choose a default working directory

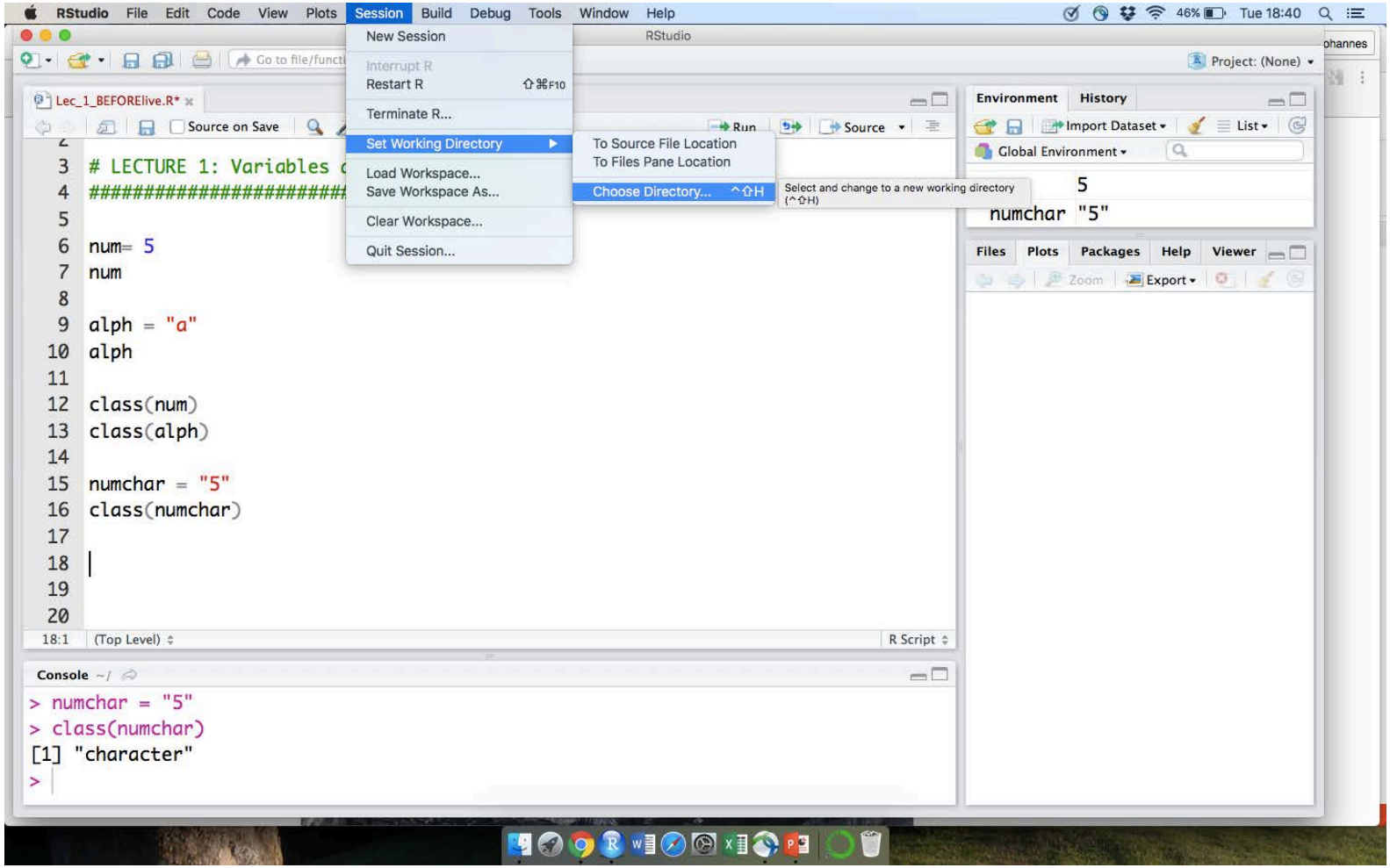


CUSTOMIZING RSTUDIO

- The default working directory is the directory in which your script files are going to be saved, it you do not specify otherwise…
- And were R looks for external data to load, if you do not specify a path.
  - We will look into this a little later in this course.

SETTING WORKING DIRECTORY

- Under “Session” you can set a customized working directory. We will use this quite frequently.




Practice Session: Writing the first lines of code…

WRITING THE FIRST LINES OF CODE…

- Today, we will start discovering the world of programming in quite a playful way.
- You will **not** be able to understand everything we do already today.
- The idea is that you get a first taste!
- Instructions and comments on the coding will be placed directly into the R scripts.
- I will post the scripts to the online script after the course.