R Resources

Alexander Gamerdinger

2022-02-04

Contents

Introduction **2**

Basic R exercises **2**
More comprehensive exercises 2

Workflow **3**

Data Manipulation **4**
data.table 4
dplyr 4

Data Visualization **5**
Network Visualization using ggraph and igraph 5

Introduction

Learning a new programming language, especially if it's your first, can be frustrating at times. This is completely normal.Even if the exercises shown in the links below are difficult to follow, please do not stop! Try to keep going, and remember, that wrestling with such initial frustrations can actually be helpful to learn the R programming language. It will also make it much easier for you to follow the R scripts in class.

However, please also remember, that this course is not *only* about coding, but also interpretation. So, try to keep your motivation up, and make sure to ask questions in the exercise classes in case you do not understand the course material.

Important note

This guide will be updated as we go along. New versions will have file names ending in higher version numbers, so the first one is **r-resources-v01.pdf**, the next one will be 'r-resources-v02.pdf' and so on. If you see a new version number than the one you have on your computer, this means that new guides have been added.
These will be useful as references, because you most likely can't remember all of this.

Basic R exercises

This is the very basic. Some of the common confusion beginners have can be clarified by reading this. Also: Instead of copy-pasting, type the commands yourself! Actually typing the commands will help you understand them better.

https://rstudio-education.github.io/hopr/basics.html

More comprehensive exercises

These three resources can be used as a supplement throughout the course. These three online tutorials can be recommended, but there is a wealth of resources out there to learn R, so if these are not up your alley then go on Google and pick your poison, as it were. If you find a course that you think is good, I'd appreciate it if you e-mail me and tell me: aga.ioa@cbs.dk.

- free and quite good interactive course:

https://www.udemy.com/course/r-basics

- all the links untill the "machine learning" chapter are relevant:

https://www.guru99.com/r-tutorial.html

- Probably the best of the three courses cited here, but it's only partly free.

https://www.datacamp.com/courses/free-introduction-to-r

Workflow

Figure 1 visualizes the workflow of the exercise classes.

In this course, we only work with cleaned data sets, so data cleaning is not necessary. When we begin to look at a data set, we have to manipulate the data. R has several packages that help us with data manipulation. The most popular ones are the `data.table` and `dplyr` package.

Once this step is done, we will use the `igraph` and `ggraph` package to visualize the content in networks. Lastly, we interpret the output using the theories of the course.

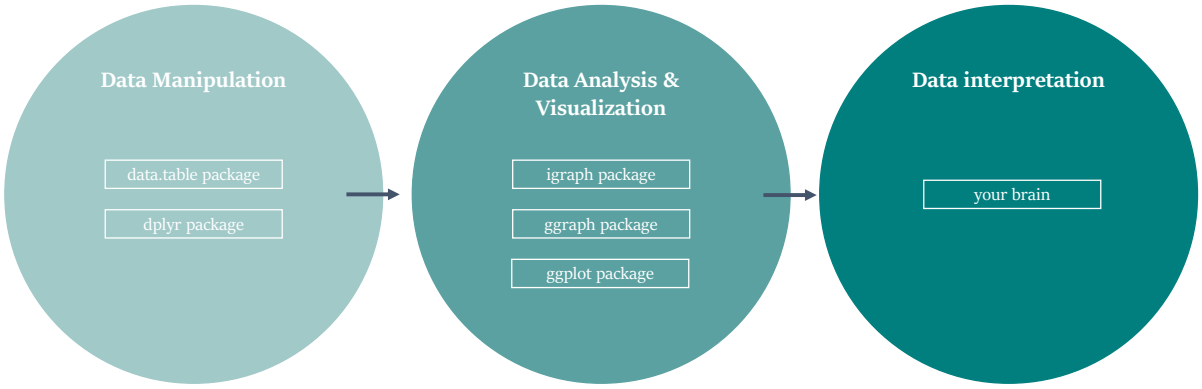

Figure 1: Workflow of this course

Data Manipulation

For each exercise class, you will have access to videos made by Emil Begtrup-Bright who taught the exercise classes last year. In his scripts, he uses the `data.table` package to manipulate data, partly, because it is the fastest and most efficient package for data manipulation in R.

Lately however, many R courses have taught students to manipulate data with the `dplyr` package, which is a part of the tidyverse. This package has a different syntax than `data.table` but might be easier to understand.

Important note

In this course, you are free to choose either `data.table` or `dplyr` for data manipulation. As the exercise class videos will be done in `data.table`, I will mainly stick to this one too. However, I will make sure to include the `dplyr` way in my slides too.

data.table

Data.table is the fastest, most efficient package / library for data analysis in R. These guides assume a certain level of comfort with using R already, but for the very interested, this is a good place to start.

All of these guides say the same things, but in different ways. I have tried to place them in an order I think is from 'easiest to understand' to 'hardest to understand' for a beginner.

- Good places to start:

https://rpubs.com/kykimeng/intro-to-data-table

https://hutsons-hacks.info/data-table-everything-you-need-to-know-to-get-you-started-in-r

https://cran.r-project.org/web/packages/data.table/vignettes/datatable-intro.html

- if you're used to the tidyverse and dplyr, this one is for you:

https://rstudio-pubs-static.s3.amazonaws.com/469157_754566dda9144a4dabec49512cff9cf1.html

- and two more:

https://www.datacamp.com/community/tutorials/data-table-r-tutorial

https://www.machinelearningplus.com/data-manipulation/datatable-in-r-complete-guide/

dplyr

Dplyr is a very popular package for data manipulation as it interacts well with other packages in the tidyverse. It makes use of pipe operators, which (once you have wrapped your head around it), make your code very intuitive.

Here are some guides:

https://dplyr.tidyverse.org/articles/dplyr.html

https://rpubs.com/hadley/dplyr-intro

Data Visualization

This course mostly uses network visualization packages, but sometimes also uses other data visualization techniques. Below, I present guides for the packages we are using.

Network Visualization using ggraph and igraph

The `ggraph` package is the most up-to-date way of visualizing networks. It uses the syntax of ggplot2, which is probably the best visualization library out there, for any programming language. Although the `igraph` package can also be used to visualize networks, it is less pretty.

We will mainly use the `igraph` package to analyse and manipulate graph objects.

ggraph

- very good and comprehensive guide, highly recommended:

http://mr.schochastics.net/netVizR.html

- Another good guide:

https://rpubs.com/neloe/ggraph_intro

- This guide uses the `tidygraph` package to manipulate graph-objects instead of `igraph`. So the manipulation of the graph-objects themselves will not look like the way we do it in `igraph`. But the explanation of graph-metrics, and the visualization with `ggraph` are really good:

https://evamaerey.github.io/ggplot2_grammar_guide/ggraph.html#1

- other useful guides:

https://cran.r-project.org/web/packages/ggraph/vignettes/

http://blog.schochastics.net/post/introducing-graphlayouts-with-got/

https://shirinsplayground.netlify.com/2018/03/got_network/

- these guides use `igraph`'s own plotting functions, but are quite good and thorough (and the one also cover how to make interactive network plots)

igraph

https://www.jessesadler.com/post/network-analysis-with-r/

https://kateto.net/netscix2016.html

Improvements

If you have any suggestions as to how to improve this guide, please reach out to: aga.ioa@cbs.dk. I would be more than happy to adapt the guide so it will be most helpful to you!