Course manual R programming 2020/2021

Misja Mikkers and Gertjan Verhoeven

2020-12-31

Contents

Welcome

This is the website for the course Programming in R at Tilburg University for the course year 2020/2021

Chapter 1

About this course

1.1 Aim of the course

After this course you will be able to import data, manipulate data and to visualize data. In addition you will be able to write (simple) functions to automate (boring) tasks. We will not dive into modeling.

Since visualization is the most interesting part, each lecture is working towards a plot. E.g. in the first lecture, we will start to import (prepared) data and work towards a nice looking graph.

We will teach you to work in Tidyverse. Tidyverse is a package (collection of functions) that is developed to manipulate and visualize data (and more!)

1.2 Set up of the course

This course is about programming. We believe that the only way to learn programming is by programming yourself. Therefore, we will introduce programming by online programming on the Datacamp website and work in class on notebooks.

Many concepts you will first see at Datacamp and then we apply them in class. Sometimes it will be the other way around: we used something in class and you learn more details about it at Datacamp. This is perfectly fine. However, it is important that you keep up-to-date with Datacamp otherwise you are going to get lost. Also programming is something that you need to practice. You can do the same the Datacamp two or three times. Also the notebooks that we do in class, you can play around with these. Plot different functions, solve equations for different parameter values etc. Just looking at the answers that we give you in class will not help you to learn programming.

Finally, we urge you to use google (or other search engines like DuckDuckGo) and stackoverflow with your assignments. Some students find this weird at the beginning: should we not teach you everything that you need to know? The answer is "no", for a number of reasons. First, even professional programmers use google and stackoverflow all the time. Second, python and R are open source and lots of people work with it. If you encounter a problem, chances are that someone else had the same problem and knows the solution to it. There is not need to "invent the wheel". Use the resources available to you. If you copy a lot of code, you should add a reference. Finally, because python and R are open source, they develop rapidly. The things that we teach you now, will be obsolete in a couple of years time. Hence, you need to be able to find your way around also in 10 years time. It will help you a lot to specifically google your error messages. To start practicing this, use google now.

1.3 Questions

here are no stupid questions, it's stupid not to ask questions. We encourage you to post your questions in the discussion section on Canvas.

Only when you need to include privately sensitive information ("my cat has passed away"), you can send an email. Always provide us with the following information: - say whether you are an ECO or EBE student - mention the group number of your tutorial and/or the name of your tutorial teacher - explain your question

1.4 Team

The R-part of the course Programming for Economists in 2020-2021 is taught by:

- Misja Mikkers
- Gertjan Verhoeven

1.5 Exam

This year the exam will be a multiple choice exam. You can use R-studio on your own computer and you are allowed to google.

Chapter 2

Schedule

2.1 Lecture 1: Visualisation

2.1.1 Week 8

February 22, 2021 - February 26, 2021

2.1.2 Preparation

Before class you should have

- installed the software
- cloned the repo XXX on your computer
- finished datacamp course "Introduction to the Tidyverse"

2.1.3 In class

In class we will go through the notebook of lecture 1

2.2 Lecture 2: Data wrangling 1

2.2.1 Week 9

March 1, 2021 - March 5, 2021

2.2.2   Preparation

Before class you should have

- finished datacamp course "Introduction to R"

2.2.3   In class

In class we will go through the notebook of lecture 2

2.3   Lecture 3: Data wrangling 2

2.3.1   Week 10

March 8, 2021 - March 12, 2021

2.3.2   Preparation

- downloaded some data
- made your own notebook with some data wrangling and a plot

The aim of this preparation is to practice with the skills required until now. This should be a good preparation for the exam with respect to the basis programming skills.

2.3.3   In class

In class we will have some time to answer your questions with respect to your prepared notebook.

Furthermore, we will go through the notebook of lecture 3.

2.4   Lecture 4: Functions and knitr

2.4.1   Week 11

March 15, 2021 - March 19 2021

2.4.2 Preparation

2.4.3 In class

2.5 Lecture 5: Generating data and some statistics

2.5.1 Week 12

March 22, 2021 - March 26, 2021

2.5.2 Preparation

2.5.3 In class

Chapter 3

Installing the software

Rstudio provides a good starting point for beginners to learn R and Rstudio.

In particular, the first chapter of moderndive and the first chapter of R for data science.

3.1   What do you need?

We assume you have a laptop (Windows or Mac). To be able to follow the course and use the software for other courses you need the following free software:

- R
- Rstudio
- TinyTex
- R packages

R is free software for importing data, manipulating data and statistical analysis. Once installed, you don't need to open the software.

We will use R in another program: Rstudio. Rstudio is a so called Integrated Development Environment (IDE), which allows you to write and run code.

Rstudio is able to transform your code to different nice outputs:

- notebook
- presentation
- article
- thesis
- much more (even this website and this course manual is build in Rstudio!)

Rstudio needs TinyTex to transform your code to pdf (which is our preferred format, however other formats are also possible).

If you need to do things more often, it is useful to write a function to do these things. Other R users also use functions and store them in packages. Some of these packages are published online and can be used by all users. We will depend in our course on some of these packages.

The first package we will install is the package TinyTex, which is basically a function to install the TinyTex software on your computer, taking into account your operating system.

We also show you how to install other packages.

In this chapter we will instruct how to install the software.

3.2   Installing R

In this paragraph we will take you through the steps to install R.

1. Go to this website
2. Choose download R for your operating system. You can choose from Windows, Mac OS X for apple computers and Linux. If you don't know which operating system you have and you don't have an apple computer, you may guess you have a Windows computer.
3. After downloading, open the downloaded file and follow the steps to install the program. Choose the default options.

3.3   R studio

You can download and install R studio from http://www.rstudio.com/ download. Choose the option: "Open Source License Free". The website will then recommend a version for your operating system.

You will see 3 "panes":

1. Console (left part of the screen)
2. Environment/history etc (top right of your screen)
3. Files/Plots/Packages (bottom right of your sreen)

$<>$

You can test the installation by typing

$1 + 1$

followed by an enter to test the installation.

3.4   Installation of packages

In the pane at the bottom right there is a tab called "Packages". After clicking this tab a new tab "Install" appears

<>

After clicking the "Install" button, you can type tinytex in the pop-up.

<>

Please install the package tidyverse.

Chapter 4

Practice exam