

Follow-up on R

Contents

About this course	1
1 Data sub-setting and logical conditions	1
1.1 Learning objectives	1
2 Tidyverse introduction	2
2.1 Seminar	2

About this course

Course Content

This course builds on introductory R and expands on essential concepts for R. We cover common problems in data manipulation to pre-process datasets and we will cover some visualization. 99.5% of a data scientist's job is data pre-processing and visualization. The rest is analysis. Therefore, brushing up on programming skills in R makes the experience of working with R a lot more enjoyable.

Course Objectives

Participants will learn data manipulation skills such as merging data, re-shaping it, aggregating it and more. Course participants should be a lot more comfortable in working with R upon completing this course.

Course Prerequisites

Participants are expected to have used R and RStudio before.

Agenda

1. Data sub-setting and logical conditions
2. Tidyverse introduction: Simple data manipulation, renaming, summarizing, aggregating
3. Merging data
4. Reshaping data
5. Regular expressions

Acknowledgements

The infrastructure for this website is in large parts adopted from work by UCL's Altaf Ali, Jack Blumenau, Lucas Leemann, Slava Jankin Mikhaylov, and Philipp Broniecki. The ESRC Business and Local Government Data Research Centre is funded by the Economic and Social Research Council (ESRC).

placeholder

placeholder

1 Data sub-setting and logical conditions

1.1 Learning objectives

ABC...

2 Tidyverse introduction

2.1 Seminar

2.1.1 Introduction to the tidyverse

The `tidyverse` package makes it easier to pre-process data. By pre-processing, we mean storing data in such a way that observations are in rows and variables are in columns. Data comes in different forms and is not always stored in this way. We will use data from the World Bank to illustrate this point.

Before getting started, we first need to install the `tidyverse` package like so: `install.packages("tidyverse")`. You only need to install once. However, doing this again is not a mistake. In fact, R, RStudio and R packages are regularly updated. It is good practice to update all of these on your computer as well. Just remember never to update before a deadline!

```
# clear workspace
rm(list = ls())

# load tidyverse package
library(tidyverse)
```

Let's check whether we have updates available by running `tidyverse_update()`

```
# check for available updates
tidyverse_update()
```

The following packages are out of date:

```
* httr      (1.4.0 -> 1.4.1)
* modelr    (0.1.4 -> 0.1.5)
* tidyr     (0.8.3 -> 1.0.0)
* xml2      (1.2.0 -> 1.2.2)
```

Start a clean R session then run:

```
install.packages(c("httr", "modelr", "tidyr", "xml2"))
```

R tells us that several packages are in indeed out of data (this may be different on your computer - it's possible that everything is up to date on your machine). Below, we update according to the console message.

```
install.packages(c("httr", "modelr", "tidyr", "xml2"))
tidyverse_update()
```

All tidyverse packages up-to-date

Now, everything is updated correctly.

Let's check the codebook of our data.

Variable Name	Description
IMMBRIT	Out of every 100 people in Britain, how many do you think are immigrants from Non-western countries?
over.estimate	1 if estimate is higher than 10.7%.
RSex	1 = male, 2 = female
RAge	Age of respondent
Househld	Number of people living in respondent's household
party_self	1 = Conservatives; 2 = Labour; 3 = SNP; 4 = Ukup; 5 = BNP; 6 = GP; 7 = party.other
paper	Do you normally read any daily morning newspaper 3+ times/week?
WWWhours	How many hours WWW per week?

Variable Name	Description
religious	Do you regard yourself as belonging to any particular religion?
employMonths	How many mnths w. present employer?
urban	Population density, 4 categories (highest density is 4, lowest is 1)
health.good	How is your health in general for someone of your age? (0: bad, 1: fair, 2: fairly good, 3: good)
HHInc	Income bands for household, high number = high HH income

The dataset is on your memory sticks and also available for download [here](#).

```
# load non-western foreigners data set
load("non_western_immigrants.RData")
```