

How To Tidy Up Your Dataset – Exercises



INTRODUCTION

In general data analysis includes four parts: Data collection, Data manipulation, Data visualization and Data Conclusion or Analysis. The `tidyr` package is one of the most useful packages for the second category of data manipulation as tidy data is the number one factor for a successful analysis.

Tidy data means that every column stands for a variable and every row represents for an observation.

The `tidyr` package offers useful functions, which we are going to see, that help us organize raw data.

Before proceeding, please follow our short [tutorial](#).

Look at the examples given and try to understand the logic behind them. Then try to solve the exercises below using R and without looking at the answers. Then check the [solutions](#) to check your answers.

Exercise 1

Gather “`day1points`” and “`day2points`” into a new column “`day`” and their values to a new column named “`points`”. **HINT:** Use `gather()`.

Exercise 2

Reverse the position of `day` and `points` to understand the

significance of their initial position.

Exercise 3

Reverse what you did in Exercise 1 by giving to the dataset its initial form. **HINT:** Use `spread()`.



Learn more about TidyR in the online course [R Data Pre-Processing & Data Management – Shape your Data!](#) This course has 5 lectures on TidyR, in addition to lectures on importing data, filtering and querying using dplyr, and SQL.

Exercise 4

Reverse the position of “day” and “points” in the answer of Exercise 3 to understand why the code is not working.

Exercise 5

Create two columns one for the “team” and the other for the “state” from the column “team”. Set the sep to 3. **HINT:** Use `separate()`.

Exercise 6

Change the sep argument from 3 to 2 and find the mistake.

Exercise 7

Unite the two columns you created in Exercise 6 to one as its initial form. **HINT:** Use `unite()`.

Exercise 8

Use the right commands to tidy up your dataset by creating 5 columns: “player”, “Team”, “State”, “day” and “points”.

Exercise 9

Plot your dataset by creating a scatterplot with day in x-axis and points in y-axis. **HINT:** Use `ggplot()`.

Exercise 10

Separate the plot of Exercise 9 according to "Team". **HINT:** Use `facet_wrap()`.