



Universidad
del País Vasco



Euskal Herriko
Unibertsitatea

ikerbasque
Basque Foundation for Science



Statistics
Korea



KOSTAT-UNFPA Summer Seminar on Population

Workshop 1. Demography in R

Day 2: The tidy data approach

Instructor: Tim Riffe

`tim.riffe@ehu.eus`

Assistants:

Jinyoen Jo: `jjyo43043@gmail.com`

Rustam Tursun-Zade: `rustam.tursunzade@gmail.com`

28 July 2022

Contents

1 Tidy data	1
1.1 Definition	1
1.2 Example	2

1 Tidy data

1.1 Definition

Tidy data follows a standard structure where each column is a variable, each row is an observation, and each cell is a value. Anything else is messy. It's literally that straightforward. A more complete definition can be found here: <https://cran.r-project.org/web/packages/tidyr/vignettes/tidy-data.html> Demographic data is often delivered in a tidy format. When it is not, then it can be reshaped into a tidy format.

Tidyverse packages work well together because they share a standard approach to formatting and working with datasets. Tidy datasets processed using tidyverse tools allow for fast and

understandable analyses that in many cases require no *programming*, whereas it often takes a certain amount of head-scratching (programming) to analyze not-tidy datasets.

Tidy datasets can also be visualized without further ado using a systematic *grammar* (Wilkinson 2012) implemented in the `ggplot2` package (Wickham (2016), this loads automatically with `tidyverse`). Today we will do just basic examples, but this will be made more explicit as the workshop progresses.

1.2 Example

The so-called `gapminder` dataset is an example of *tidy* data that allows to demonstrate some of the basic `tidyverse` concepts. Let's install this package and have a look. Remember to comment out the installation line of code using `#` after you install it once!

```
install.packages("gapminder")
```

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer.

Wilkinson, Leland. 2012. "The Grammar of Graphics." In *Handbook of Computational Statistics*, 375–414. Springer.