pandas (R)osetta: Tidyverse R

Intro to R and Tidyverse for Pandas users and vice versa.

Lisa Malins

Contents

This is a demonstration of basic data wrangling operations in Tidyverse R.

Sister notebooks demonstrate the exact same operations in base R and the Pandas library in Python, just like the Rosetta Stone.

I/O

Create dataframe from scratch

```
suppressPackageStartupMessages(library("dplyr"))
tbl <- tibble(
  letter = c("a", "b", "c", "d", "e"),
  number = c(1:5),
  fruit = c("apple", "banana", "coconut", "date", "elderberry"),
  vegetable = c("arugula", "beet", "carrot", "daikon", "eggplant"),
  name = c("Alice", "Bob", "Carol", "Dan", "Eve")
)
tbl
```

```
## # A tibble: 5 x 5
##   letter number fruit      vegetable name
##   <chr>   <int> <chr>      <chr>     <chr>
## 1 a           1 apple      arugula   Alice
## 2 b           2 banana     beet      Bob
## 3 c           3 coconut    carrot    Carol
## 4 d           4 date       daikon    Dan
## 5 e           5 elderberry eggplant  Eve
```

Write

```
library("readr")
write_csv(tbl, "data/tidy_letters.csv")
write_tsv(tbl, "data/tidy_letters.tsv")
```

Read

```
library("readr")
read_csv("data/tidy_letters.csv")
```

```
## Rows: 5 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (4): letter, fruit, vegetable, name
## dbl (1): number
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 5 x 5
##   letter number fruit      vegetable name
##   <chr>   <dbl> <chr>      <chr>     <chr>
## 1 a           1 apple      arugula   Alice
## 2 b           2 banana     beet      Bob
## 3 c           3 coconut    carrot    Carol
## 4 d           4 date       daikon    Dan
## 5 e           5 elderberry eggplant  Eve
```

```
read_tsv("data/tidy_letters.tsv")
```

```
## Rows: 5 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (4): letter, fruit, vegetable, name
## dbl (1): number
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 5 x 5
##   letter number fruit      vegetable name
##   <chr>   <dbl> <chr>      <chr>     <chr>
## 1 a           1 apple      arugula   Alice
## 2 b           2 banana     beet      Bob
## 3 c           3 coconut    carrot    Carol
## 4 d           4 date       daikon    Dan
## 5 e           5 elderberry eggplant  Eve
```