pandas (R)osetta: Pandas

Intro to Pandas for R users and vice versa.

Lisa Malins

Contents

I/O		1
	Create dataframe from scratch	1
	Write	2
	Read	2

This is a demonstration of basic data wrangling operations in the Pandas library for Python.

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

I/O

Create dataframe from scratch

```
import pandas as pd

df = pd.DataFrame(
    {"letter" : ["a", "b", "c", "d", "e"],
    "number" : range(1,6),
    "fruit" : ["apple", "banana", "coconut", "date", "elderberry"],
    "vegetable" : ["arugula", "beet", "carrot", "daikon", "eggplant"],
    "name" : ["Alice", "Bob", "Carol", "Dan", "Eve"]}
)
df
```

```
letter number
                          fruit vegetable
##
                                            name
## 0
                 1
                                  arugula Alice
         a
                          apple
                 2
## 1
         b
                         banana
                                     beet
                                             Bob
                 3
## 2
         С
                        coconut
                                   carrot Carol
## 3
          d
                           date
                                   daikon
                                             Dan
## 4
                 5 elderberry eggplant
                                             Eve
```

Write

```
df.to_csv("data/py_letters.csv", index=False)
df.to_csv("data/py_letters.tsv", sep="\t", index=False)
```

Read

```
pd.read_csv("data/py_letters.csv")
```

```
name
##
    letter number
                         fruit vegetable
## 0
         a
                1
                         apple
                                arugula Alice
## 1
                 2
                        banana
                                   beet
                                           Bob
         b
## 2
                 3
                                  carrot Carol
         С
                       coconut
## 3
         d
                 4
                          date
                                  daikon
                                           Dan
## 4
                 5 elderberry eggplant
                                           Eve
```

```
pd.read_table("data/py_letters.tsv")
```

name	vegetable	fruit	number	letter		##
Alice	arugula	apple	1	a	0	##
Bob	beet	banana	2	b	1	##
Carol	carrot	coconut	3	С	2	##
Dan	daikon	date	4	d	3	##
Eve	eggplant	elderberry	5	е	4	##