pandas (R)osetta: Pandas

Intro to Pandas for R users and vice versa.

Lisa Malins

Contents

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

```python
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      a       1       apple   arugula  Alice
## 1      b       2      banana      beet    Bob
## 2      c       3     coconut    carrot  Carol
## 3      d       4        date    daikon    Dan
## 4      e       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")
```

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

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

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