pandas_intro

December 16, 2023

1 Refresher of the Pandas Package

```
[ ]: import pandas as pd
```

1.1 Create a new DataFrame

1.1.1 Create DF from dictionary

```
[ ]: names = ['United States', 'Australia', 'Japan', 'India', 'Russia', 'Morocco',␣
      ↪'Egypt']
     dr =  [True, False, False, False, True, True, True]
     cpc = [809, 731, 588, 18, 200, 70, 45]
```

You build the dataframe from either a set of lists in a dictionary or import from external sources

```
[ ]: # Create dictionary
     my_dict = {'country':names, 'drives_right':dr, 'cars_per_cap':cpc}
```

```
[ ]: cars = pd.DataFrame(data=my_dict)
     print(cars)
```

```
         country  drives_right  cars_per_cap
0  United States          True           809
1      Australia         False           731
2          Japan         False           588
3          India         False            18
4         Russia          True           200
5        Morocco          True            70
6          Egypt          True            45
```

```
[ ]: row_labels = ['US', 'AUS', 'JPN', 'IN', 'RU', 'MOR', 'EG']
     cars.index = row_labels # Set row labels
     print(cars)
```

```
           country  drives_right  cars_per_cap
US   United States          True           809
AUS      Australia         False           731
JPN          Japan         False           588
IN           India         False            18
```

```
RU          Russia          True          200
MOR        Morocco          True           70
EG           Egypt          True           45
```

1.1.2 Create DF from csv file

Reading from an external source file

[ ]:
```
brics = pd.read_csv('brics.csv',index_col=0) # Set first column as row labels
print(brics)
```

```
         country    capital    area  population
BR        Brazil   Brasilia   8.516      200.40
RU        Russia     Moscow  17.100      143.50
IN         India  New Delhi   3.286     1252.00
CH         China    Beijing   9.597     1357.00
SA  South Africa   Pretoria   1.221       52.98
```

1.2 Selection of Elements from Pandas DataFrame

1.2.1 Using Square brackets

Selecting of columns

Selecting one column

[ ]:
```
country = brics['country']
print(country)
```

```
BR          Brazil
RU          Russia
IN           India
CH           China
SA    South Africa
Name: country, dtype: object
```

[ ]:
```
country = brics[['country']] # Double brackets to get a DataFrame
print(country)
```

```
         country
BR        Brazil
RU        Russia
IN         India
CH         China
SA  South Africa
```

Selecting multiple columns

[ ]:
```
brazil_russia = brics[['country','capital']] # Get multiple columns
print(brazil_russia)
```

```
          country    capital
BR         Brazil   Brasilia
RU         Russia     Moscow
IN          India  New Delhi
CH          China    Beijing
SA  South Africa   Pretoria
```

Selecting of rows

[ ]:
```
brics[1:2] # Get multiple columns by index
```

[ ]:
```
    country capital  area  population
RU  Russia  Moscow  17.1       143.5
```

1.2.2 Selecting by loc and iloc

Loc uses the labels to select

[ ]:
```
print(brics.loc[['RU','IN','CH']]) # Get multiple rows by label
```

```
   country    capital    area  population
RU  Russia     Moscow  17.100       143.5
IN   India  New Delhi   3.286      1252.0
CH   China    Beijing   9.597      1357.0
```

[ ]:
```
print(brics.loc[['RU','IN','CH'],['country','capital']]) # Get multiple rows␣
↪and columns by label
```

```
   country    capital
RU  Russia     Moscow
IN   India  New Delhi
CH   China    Beijing
```

iloc selects by the index

[ ]:
```
print(brics.iloc[[1,2,3]]) # Get multiple rows by index
```

```
   country    capital    area  population
RU  Russia     Moscow  17.100       143.5
IN   India  New Delhi   3.286      1252.0
CH   China    Beijing   9.597      1357.0
```

[ ]:
```
print(brics.iloc[[1,2,3],[1,2]]) # Get multiple rows and columns by index
```

```
      capital    area
RU     Moscow  17.100
IN  New Delhi   3.286
CH    Beijing   9.597
```