

# コンピュータリテラシ発展 ～Pythonを学ぶ～

## 第10回：表計算やデータ分析をやってみよう

( [shimizu@info.shonan-it.ac.jp](mailto:shimizu@info.shonan-it.ac.jp) )

# 今回の授業内容

## 今回の授業内容

- 
- 
- pandas
-

## 前回の課題解説

## 前回の課題解説

- 
- 

## 解答例

[https://colab.research.google.com/drive/1LhxzCzdAQu8Utigrpj2KZTrR1oMnll69?  
usp=sharing](https://colab.research.google.com/drive/1LhxzCzdAQu8Utigrpj2KZTrR1oMnll69?usp=sharing)

データ分析を始める前に

# ExcelとPythonの使い分け

## Excel

- 

## Python

- 

- JupyterLab (Google Colab)

- 

Python R

# データ処理の流れ

- 3ステップ
- Python ,

1. データ収集

2. 前処理

3. 集計・可視化



# データ分析の流れ

1.

○

○

Web

○

○

# データ分析の流れ

2.

- 
- 
- 
-

# データ分析の流れ

3.

- 
-

# pandasの使い方

# pandasの基本的な使い方

- pandas Python

- pandas

- 

- 

- 

- 

- 

<https://pandas.pydata.org/>

Qiita <https://qiita.com/tags/pandas>

# pandasの基本的な使い方

- pandas          Google Colab
- 
- pandas          pd

```
# pandasモジュールをインポートして「pd」と省略する  
import pandas as pd
```

# pandasの基本的な使い方

- Series
  - 
  -
- DataFrame
  - Excel
  -

( DataFrame

Series

)

# pandasの基本的な使い方

- DataFrame    Series

- Series

```
#一度インポートしたら毎回書かなくてもよい
```

```
import pandas as pd
```

```
# 名前のリストを元にpandasのSeriesオブジェクトを作成
```

```
# Seriesは一系列のデータを扱うためのデータ構造
```

```
name_series = pd.Series(['田中', '北野'])
```

```
# 作成したSeriesオブジェクトを表示
```

```
print(name_series)
```



# pandasの基本的な使い方

- DataFrame    Series

- DataFrame

- 名前    年齢

```
# 名前と年齢のデータを含むデータフレームを作成
# '名前' 列には事前に定義された Series オブジェクト 'name_series' を使用
# '年齢' 列には年齢のリスト [22, 28] を指定
people = pd.DataFrame({
    '名前': name_series,
    '年齢': [22, 28]
})

print(people)
```

# pandasの基本的な使い方

- DataFrame      Series
- DataFrame      Series

```
import pandas as pd
```

```
# 先に作成されたデータフレーム 'people' の '年齢' 列を表示  
print(people['年齢'])
```

# pandasの基本的な使い方

## インデックスラベル（行名）

- 

0

1

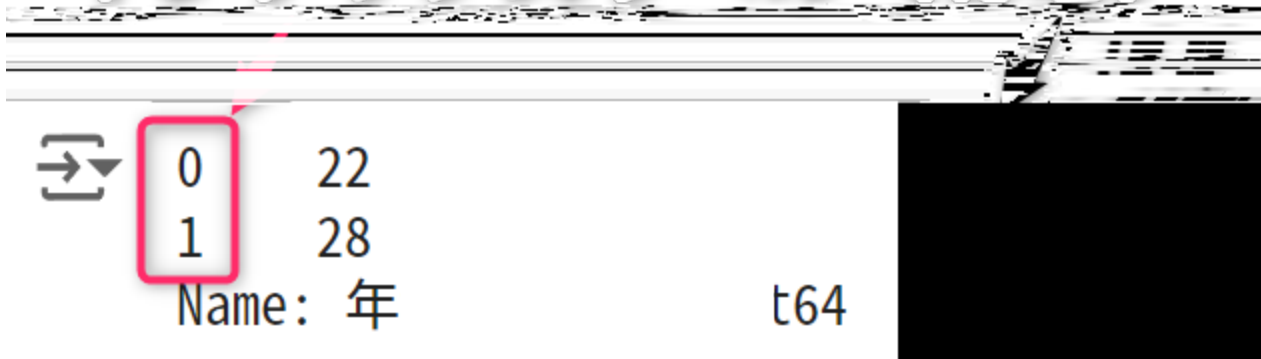
- 

インデックスラベル（行名）

- 

0

インデックスラベル（行名）



⇒	0	22	
	1	28	
	Name: 年	t64	

# pandasの基本的な使い方

- インデックスラベル（行名）

- Series

- DataFrame

- index

- 

- DataFrame

- 

```
# 名前のリストを元にpandasのSeriesオブジェクトを作成
# '田中' と '北野' という名前のリスト
# インデックスを [10, 20] に設定（デフォルトのインデックスを使わずに指定したインデックスを使用）
name_series = pd.Series(['田中', '北野'], index = [10, 20])

print(name_series)
```

# pandasの基本的な使い方

## 行インデックス（行番号）

- 行インデックス(行番号)
- 
- 0

# pandasの基本的な使い方

- `loc` `iloc`
  - `loc`属性
    - 
    -
  - `iloc`属性
    - 
    -

# pandasの基本的な使い方

- loc iloc

- 

DataFrame

```
# 年齢、身長、体重のデータを含むデータフレームを作成
# '年齢', '身長', '体重' という列を持つデータフレーム
# インデックスは ['田中', '北野', '高橋', '岡田', '長谷川'] に設定
people = pd.DataFrame({
    '年齢': [21, 34, 23, 44, 19],
    '身長': [180, 168, 174, 181, 169],
    '体重': [74, 61, 65, 82, 70]
},
    index=['田中', '北野', '高橋', '岡田', '長谷川'])

print(people)
```

# pandasの基本的な使い方

- loc    iloc

○

北野

loc

○

1

(iloc

○

```
# '北野'という名前のインデックスに対応する行をデータフレーム 'people' から取得し、表示  
print(people.loc['北野'])
```

```
# インデックス位置が1（2番目の行）に対応する行をデータフレーム 'people' から取得し、表示  
print(people.iloc[1])
```



# pandasの基本的な使い方

- `loc`   `iloc`
  - 
  -

```
# '高橋'という名前のインデックスに対応する行から、  
# 特定の列（'年齢'と'体重'）を取得し、表示  
print(people.loc['高橋', ['年齢', '体重']])
```

```
# インデックス位置が2（3番目の行）に対応する行から、  
# 特定の列（0番目と2番目の列：'年齢'と'体重'）を取得し、表示  
print(people.iloc[2, [0, 2]])
```

# pandasの基本的な使い方

- `loc`   `iloc`
  - `[start:stop]`
  - 
  -

```
# インデックスが '高橋' から '長谷川' までの行を  
# データフレーム 'people' から取得し、表示locプロパティを使用して、  
# インデックスラベルで範囲を指定  
print(people.loc['高橋':'長谷川'])
```

```
# インデックス位置が2から4（3番目から5番目の行）までの行を  
# データフレーム 'people' から取得し、表示ilocプロパティを使用して、  
# 整数位置で範囲を指定  
print(people.iloc[2:5])
```

# 課題

# 課題

- Moodle SCfCL-10th-prac.ipynb Colab
- File > Download > Download .ipynb .ipynb
- .ipynbファイル Prac09 Moodle
- 6月27日(木) 20時まで