



Jupyter Notebook で途中経過も共有する (IRSME17008)

平成 29 年 7 月 20 日 原田長州

データを利用してある物事を詳しく見ていく際に、試行錯誤しているうちに途中の情報を残せなくなってしまうことがある。また、結果にいたるまでの過程を共有したい場合には、一旦結果が判明した後、再現をするために画面キャプチャーをして保存し、パワーポイントなどのプレゼンテーションソフトに貼り付けなくてはならない。

「Jupyter Notebook¹」の仕組みを利用すれば、途中経過を残すこと、共有した相手が再現できる仕組みを入れること、グラフを出力することなどができる。データを処理する人とデータを共有する手間を省けるようになる。

■ Jupyter Notebook とは

Jupyter Notebook とは、オープンソースの Web アプリケーションである。プログラムのソースコードや数式、グラフなど含んだ文書を作成・共有をすることができる。大まかに言えば、Jupyter Notebook は以前「IPython²」と呼ばれていたものが複数のプログラミング言語に対応し、データサイエンス機能を強化したものだといえる。

Jupyter Notebook は修正 BSD ライセンスである。商用利用も可能で公開されているソースコードに改変を加えても、著作権など一定の表示をすれば改変後のソースコードを他人が入手可能な状態にすることは求められない。

■ 利用してみる

1. Anaconda³インストールする

Jupyter Notebook を利用するには、Python 環境が必要である。Continuum Analytics 社の Anaconda を利用して Python 環境を用意した。Anaconda は、ライブラリーとセットになった配布形態の一つである。別のやり方として Python を単体でインストールしておき、必要なものを追加していくこともできるが、一括で環境構築できるため Anaconda を選んだ。検証は、Windows7 の環境でおこなった。

Anaconda は、<https://www.continuum.io/downloads> からダウンロードしてインスト

¹ <http://jupyter.org/>

² <https://ipython.org/>

³ <https://www.continuum.io/>

平成 29 年 7 月 20 日

Jupyter Notebook で途中経過も共有する (IRSME170008)

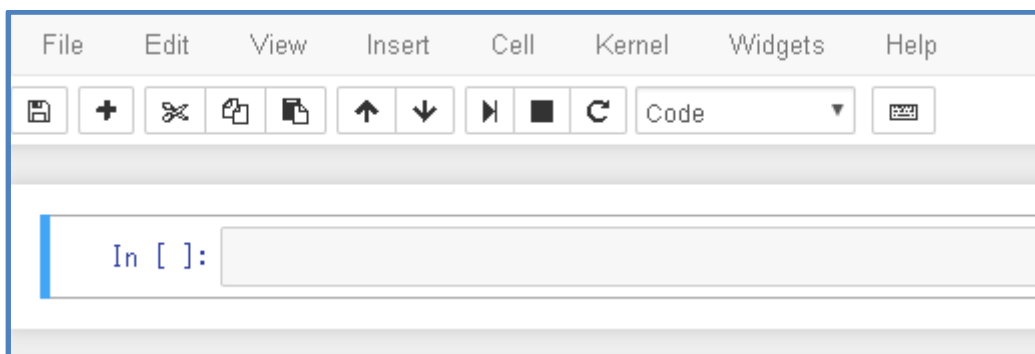
ールを行った。Python のバージョンは 3.6 か 2.7 を選択することができる。積極的に 2.7 を利用したいという理由がなければ 3.6 を選択すればよいだろう。

2. Jupyter Notebook を起動する

インストール後は、[スタート]-[Anaconda Prompt]を選択。プロンプト画面を表示させてから Jupyter Notebook と入力する。

ブラウザが立ち上がり、Jupyter Notebook が利用できるようになる。画面右上に「New」というボタンがありクリックすると Notebook や Text File/Folder などが選べる。Notebook の Python3 を選択する。

3. 実際に利用してみる



Python のコードを実行するには、In []:と表示されている横の「セル」にコードを入力して実行する。(セル内のコードを実行するにはメニューにある横向きの三角に棒が付いているボタンをクリックする。もしくは、Shift + Enter でも実行できる。)

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt

%matplotlib inline
```

上の 2 行は必要なライブラリーをインポートし短縮した名前で作るようにしている。4 行目は描画したグラフを別ファイルではなく、インラインで表示するように設定している。

```
In [2]: population = pd.read_excel('c:\\bin\\population.xlsx', sheetname='Sheet1')
```

平成 29 年 7 月 20 日

Jupyter Notebook で途中経過も共有する (IRSME170008)

エクセルのデータをインポートする。

```
In [3]: population
```

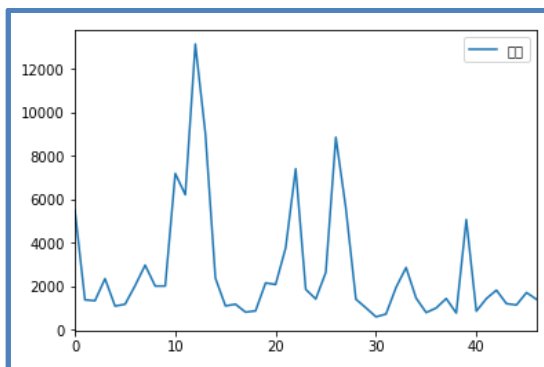
```
Out [3]:
```

	都道府県	人口
0	北海道	5506
1	青森	1373
2	岩手	1330
3	宮城	2348

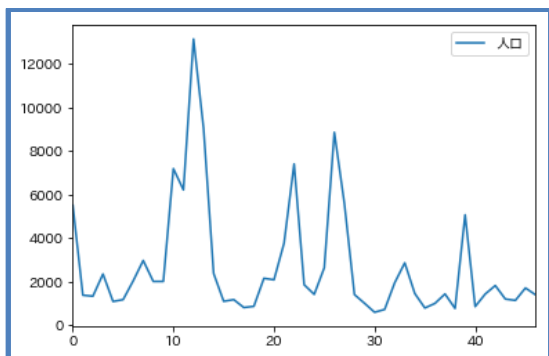
取り込んだ内容を表示する。都道府県名と人口のデータがある。

```
In [7]: population['人口'].plot()
```

人口の列を折れ線としてグラフを描画する。



グラフの表示ができた。(横軸は、都道府県の並び順に対応する番号。縦軸は人口を示している。)しかし、グラフの右上に表示される凡例部分の日本語が文字化けしているので、フォントを追加してからフォントの設定を変更した。



右上の凡例部分の文字が正しく表示された。

平成 29 年 7 月 20 日

Jupyter Notebook で途中経過も共有する (IRSME170008)

4. ヒストグラムを作る

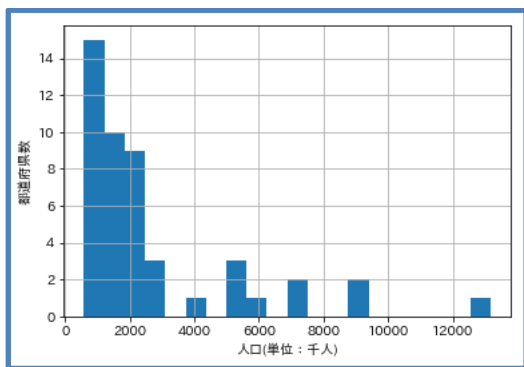
ヒストグラムを作成する。縦横のラベルを設定し、グリッド線を表示するよう設定した。

```
In [6]: plt.hist(population['人口'], bins=20)

plt.xlabel("人口(単位:千人)")
plt.ylabel("都道府県数")

plt.grid()

plt.plot()
```



結果が表示された。

■ まとめ

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline

In [2]: population = pd.read_excel('c:\bin\population.xlsx', sheetname='Sheet1')

In [7]: population['人口'].plot()
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0xa900f60>
```

```
In [5]: population['人口'].std()
Out[5]: 2682.1473373980016

In [6]: plt.hist(population['人口'], bins=20)
plt.xlabel("人口(単位:千人)")
plt.ylabel("都道府県数")
plt.grid()
plt.plot()
Out[6]: []
```

↑一連の入力と出力が連続して表示されている。

Jupyter Notebook は Web ブラウザ上で Python のコードを実行し結果を得ることができる。途中経過が自動保存され画面下に続けて表示される。(図 1 参照) 一つ前の入力に戻って、条件を変更することも可能だ。この機能を利用すれば、途中のデータから別の処理を追加するような枝分かれも可能なためデータを特定の範囲だけに限定してやり直すということもできる。

共有するには結果を ipynb 形式で保存しデータを送ることで Jupyter Notebook で開くことができる。表示するだけでよければ、HTML 形式で保存することもできる。(了)