

Visualisation

Ali Akbar Septiandri

Universitas Al-Azhar Indonesia

aliakbars@live.com

April 28, 2017

Overview

- 1 Pandas
- 2 Matplotlib
- 3 Seaborn
- 4 NetworkX

- ① VanderPlas, J. (2016). Python Data Science Handbook. O'Reilly Media Inc. (Chapter 3 & 4)
<https://github.com/jakevdp/PythonDataScienceHandbook>
- ② Dokumentasi pandas: <http://pandas.pydata.org/>
- ③ Dokumentasi matplotlib: <http://matplotlib.org/>
- ④ Dokumentasi Seaborn: <http://seaborn.pydata.org/>
- ⑤ Dokumentasi NetworkX: <https://networkx.github.io/>

Sebelum dimulai...

Google's Python Class
<https://developers.google.com/edu/python/>

Quick Quiz!

Quick Quiz

Diberikan sebuah *list of strings*, buatlah fungsi untuk:

- 1 Menghitung jumlah string yang panjangnya lebih dari 1 dan karakter pertama dan terakhirnya sama
- 2 Urutkan *list of strings* tersebut, tetapi dengan pengecualian bahwa setiap string yang diawali karakter 'x' harus ditaruh di awal *list*

Petunjuk: Buat dua *lists* terlebih dahulu, urutkan masing-masing *list*, lalu gabungkan kedua *lists* tersebut

e.g. ['mix', 'xyz', 'apple', 'xanadu', 'aardvark'] menghasilkan ['xanadu', 'xyz', 'aardvark', 'apple', 'mix']

Pandas

- Minggu lalu, kita sudah menggunakan NumPy untuk manipulasi matriks-vektor

- Minggu lalu, kita sudah menggunakan NumPy untuk manipulasi matriks-vektor
- Masalahnya, aksesnya (indeksnya) hanya bisa menggunakan angka

- Minggu lalu, kita sudah menggunakan NumPy untuk manipulasi matriks-vektor
- Masalahnya, aksesnya (indeksnya) hanya bisa menggunakan angka
- Padahal, representasi utama data adalah **tabel**!

- Pandas merupakan pustaka yang memudahkan pengolahan data
- NumPy array → Pandas Series
- NumPy matrix → Pandas DataFrame

Objek dalam Pandas

- ① Series dan DataFrame sebagai generalisasi NumPy array
- ② Series dan DataFrame sebagai spesialisasi dictionary
- ③ Index sebagai *immutable array* dan *ordered set*

loc, iloc, dan ix

- 1 loc: indeks eksplisit
- 2 iloc: indeks implisit
- 3 ix: *hybrid*

Operasi pada Data

- Bekerja seperti pada NumPy array
- Preservasi indeks
- Dapat menyebabkan NaN!

Operasi pada Nilai Null

Operasi yang dapat dilakukan pada nilai null (NaN) antara lain:

- 1 `isnull()`
- 2 `notnull()`
- 3 `dropna()`
- 4 `fillna()`

Beberapa metode untuk menggabungkan data:

- `concat()`
- `append()` [untuk objek]
- `merge()`
- `join()` [untuk objek]

Catatan: Lihat kembali materi aljabar relasional dari kuliah basis data!

Agregasi dan Pengelompokan

Untuk agregasi, seperti halnya pada SQL, Pandas mengenal `groupby()`. Lebih dari sekadar agregasi, Pandas juga mempunyai metode untuk:

- `filter()`
- `transform()`
- `apply()`

Catatan: Anda mungkin akan banyak memakai lambda function di sini!

Beberapa materi yang tidak dibahas di pertemuan ini:

- *Pivot tables*
- *Vectorized string operations*
- *Time series*

Harapannya, kalian tetap bisa mencobanya sendiri!

Matplotlib

Tentang Matplotlib

- Dibuat sebagai alternatif dari MATLAB, ada kemiripan sintaksis
- Salah satu pustaka standar untuk menggambar grafik
- Banyak pustaka lain yang membungkus matplotlib agar lebih mudah digunakan

Beberapa jenis grafik yang akan kita pelajari hari ini:

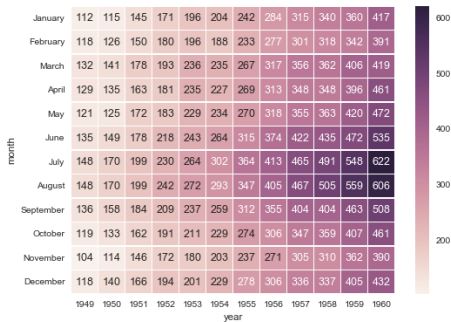
- 1 line plots
- 2 scatter plots
- 3 density & contour plots
- 4 histograms
- 5 3d plotting

Seaborn

*“Seaborn is a Python visualization library based on matplotlib.
It provides a high-level interface for drawing attractive statistical graphics.”*

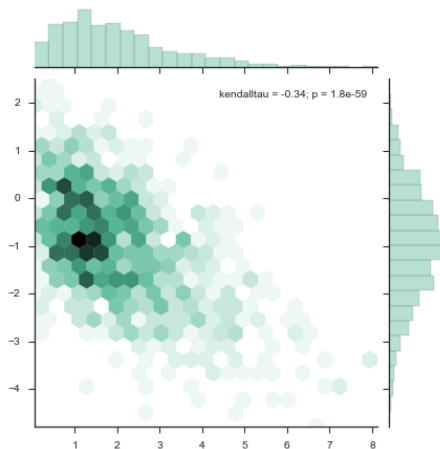
Beberapa plot menarik yang bisa dihasilkan...

Heatmap



Gambar: Annotated heatmaps [Waskom, 2015]

Joint Plot



Gambar: Hexbin plot with marginal distributions [Waskom, 2015]

NetworkX

Pembuatan dan manipulasi graf dan jejaring kompleks



Michael Waskom (2015)

Seaborn: statistical data visualization

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

Terima kasih