

Python for Data Science

Ch. 5 Visualisasi

Ahmad Rio Adriansyah

Visualisasi

- Untuk menyampaikan informasi, gambar mewakili ribuan kata
- Visualisasi digunakan untuk menyampaikan informasi yang didapatkan dari data dalam bentuk gambar
- Menampilkan pola yang sulit terdeteksi dengan tulisan saja
- Modul yang umum digunakan
 - Matplotlib
 - Seaborn
 - Ggplot
 - Plotly

Matplotlib

- Bisa digunakan untuk menggambar grafik-grafik sederhana
 - Diagram garis
 - Diagram batang
 - Histogram
 - Scatter plot
 - dll

```
▶ import matplotlib.pyplot as plt
```

Load Dataset

```
► import statsmodels.api as sm
mtcars = sm.datasets.get_rdataset('mtcars').data
mtcars.sample(5)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2

Plot

- Bandingkan jarak tempuh mobil per galon (mpg) dengan berat mobil (wt)

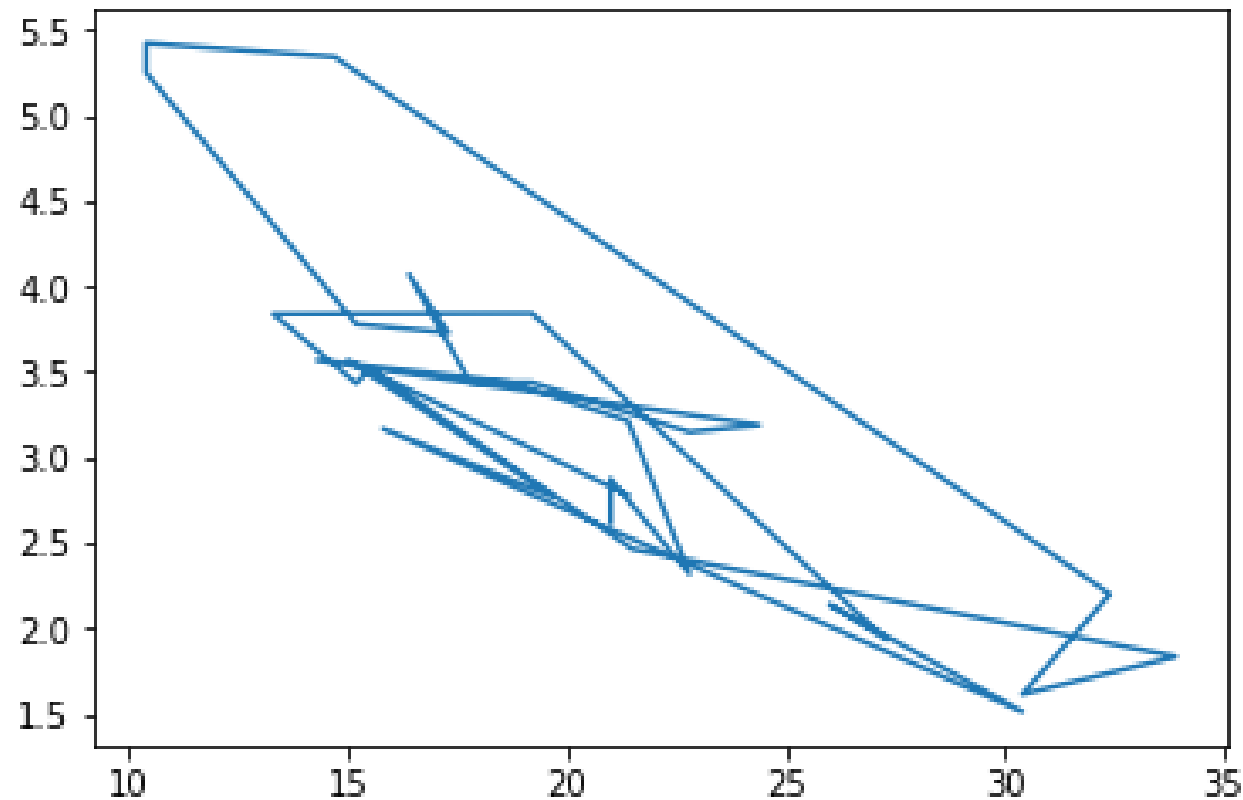
```
plt.plot(mtcars.mpg,mtcars.wt)  
plt.show()
```

- Note : jangan lupa plotnya dishow untuk menampilkan gambarnya

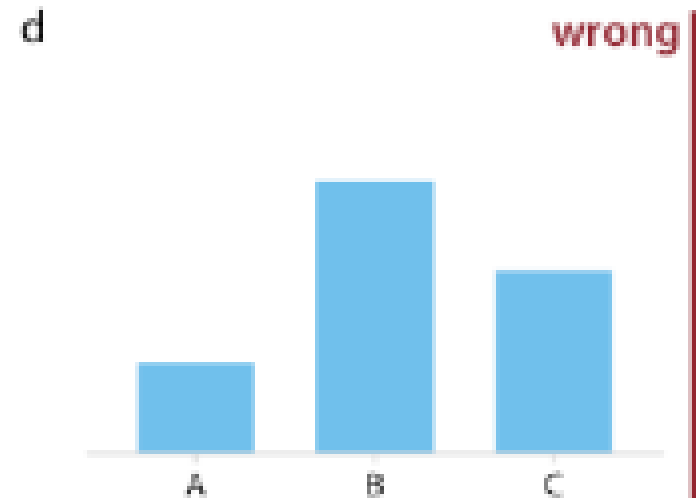
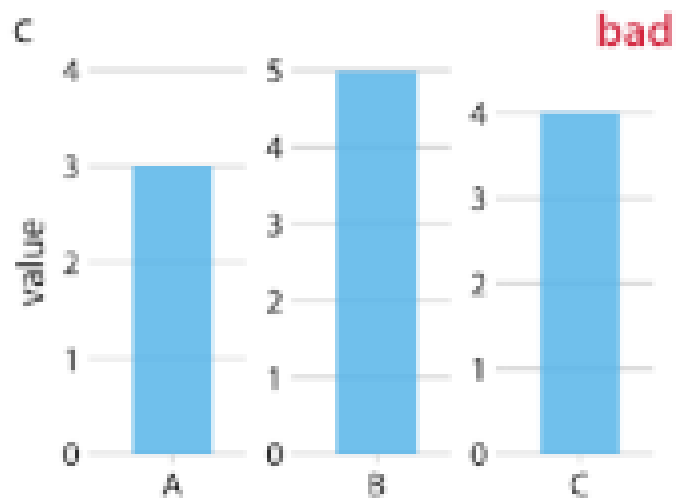
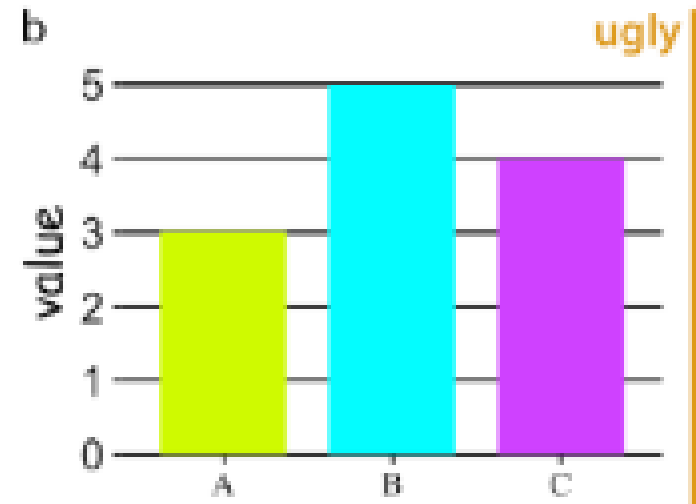
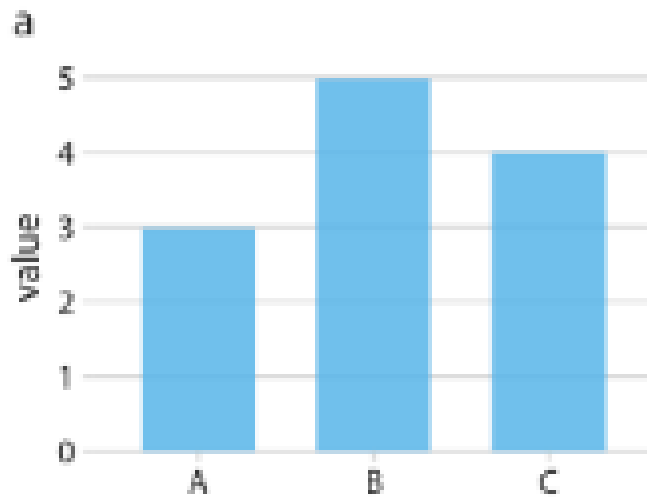
Hasil :

Makin Tidak Terbaca Informasinya

- Salah memilih jenis grafik/diagram



Bad Visualization

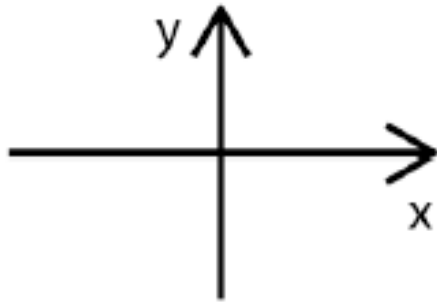


Bad Visualization

- Key principle :
 - Memahami konteks visualisasi
 - Memilih display yang sesuai
 - Menghilangkan *clutter*
 - Fokuskan perhatian ke tempat yang diinginkan
 - Berpikirlah seperti seorang desainer
 - Bercerita dengan data (tujuannya ke arah *insight*)

Estetika

position



shape



size



color



line width



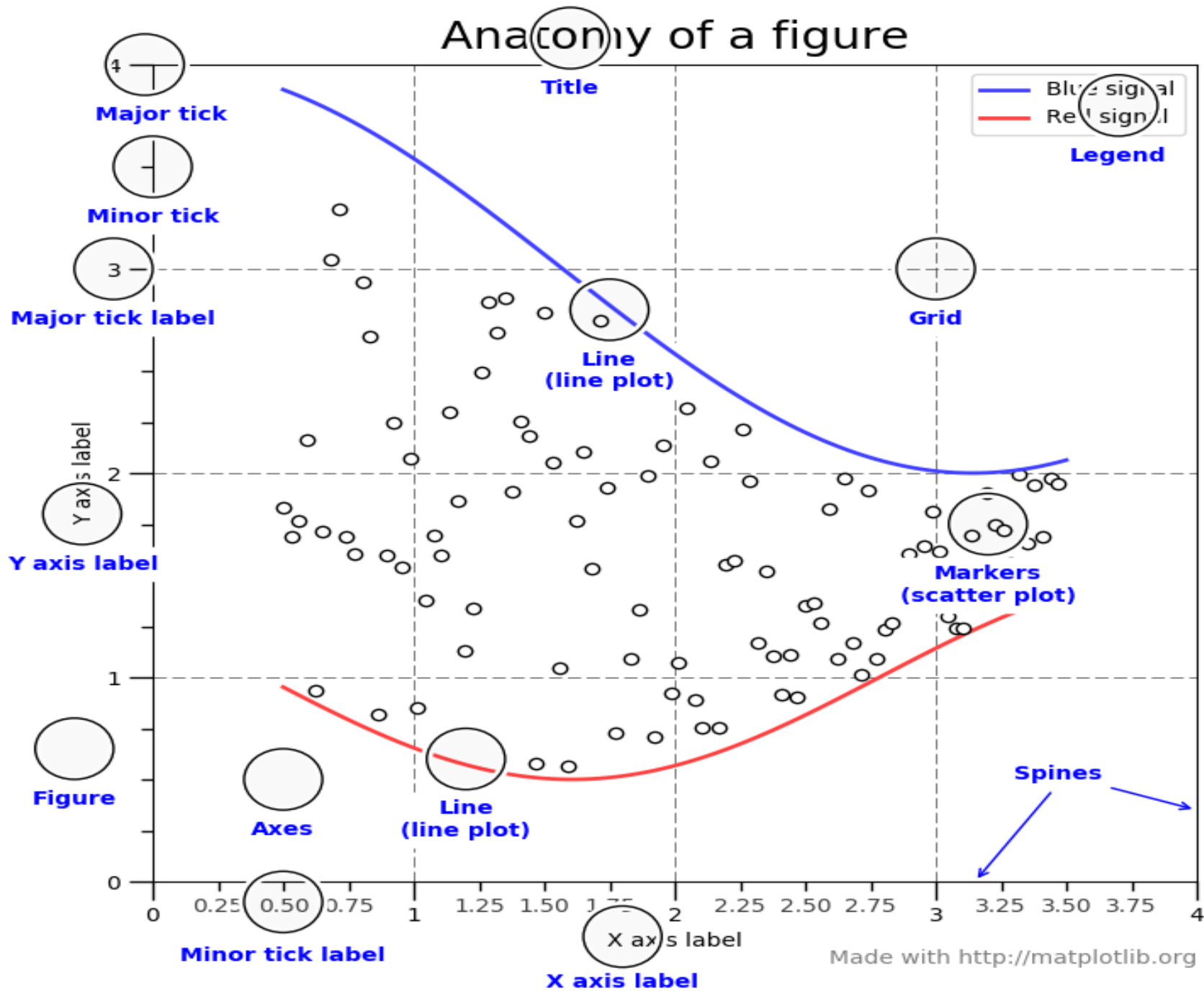
line type



Sumber : Slide Risman Adnan

Plot

- Defaultnya sebagai **diagram garis**
- Variabel pertama sebagai sumbu x, variabel kedua sebagai sumbu y
- Jika hanya diinput satu variabel, variabel x akan diambil dari index
- Bisa meletakkan beberapa plot dalam satu gambar



- <https://matplotlib.org/examples/showcase/anatomy.html>

Kostumisasi Plot

- Ada argumen ketiga yang dapat diberikan dalam bentuk string (optional) untuk mengubah warna dan tipe plotnya

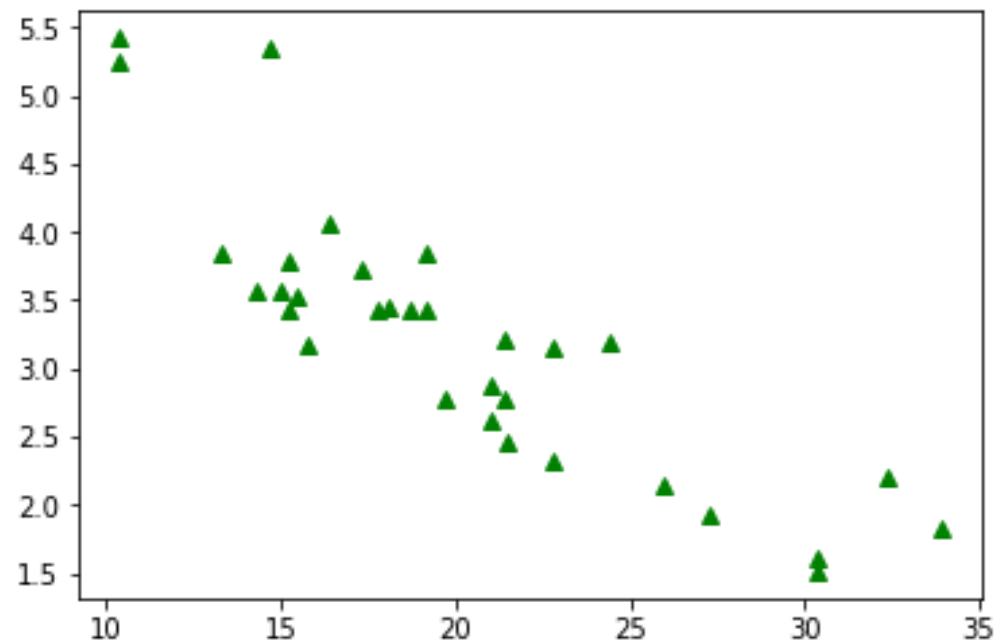
- Contoh :

```
plt.plot(mtcars.mpg,mtcars.wt,'g^')  
plt.show()
```

g^ untuk plot

berwarna hijau (g)

bentuk segitiga (^)



Kostumisasi Plot

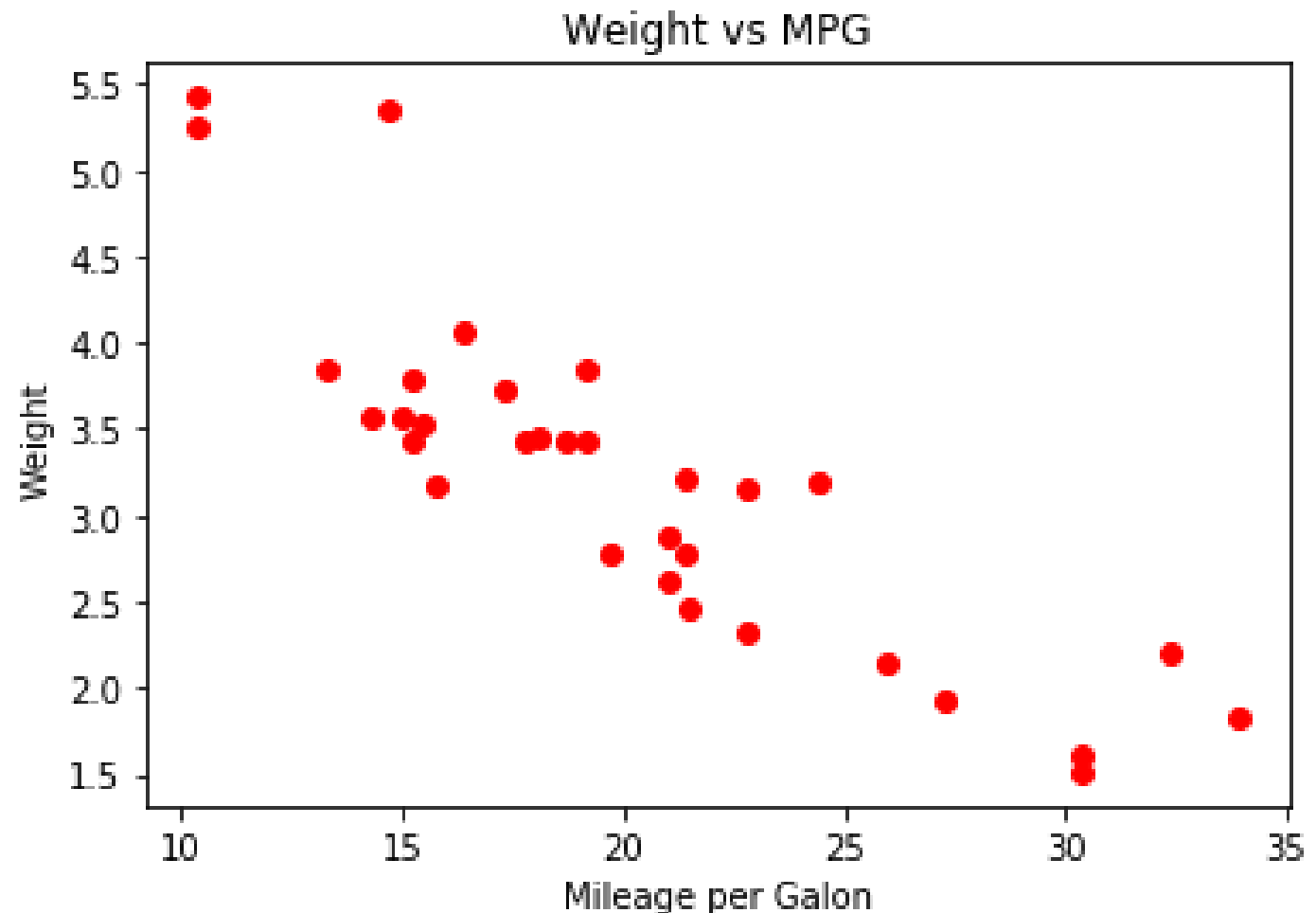
- Formatnya **[marker] [line] [color]**, tapi urutan yang lain juga bisa (parsing mungkin ambigu)

character	description	character	description	character	color
'.'	point marker	'-'	solid line style	'b'	blue
','	pixel marker	'--'	dashed line style	'g'	green
'o'	circle marker	'-.'	dash-dot line style	'r'	red
'v'	triangle_down marker	':'	dotted line style	'c'	cyan
'^'	triangle_up marker			'm'	magenta
'<'	triangle_left marker			'y'	yellow
'>'	triangle_right marker			'k'	black
'1'	tri_down marker			'w'	white
'2'	tri_up marker				

https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.plot.html

Judul dan Label

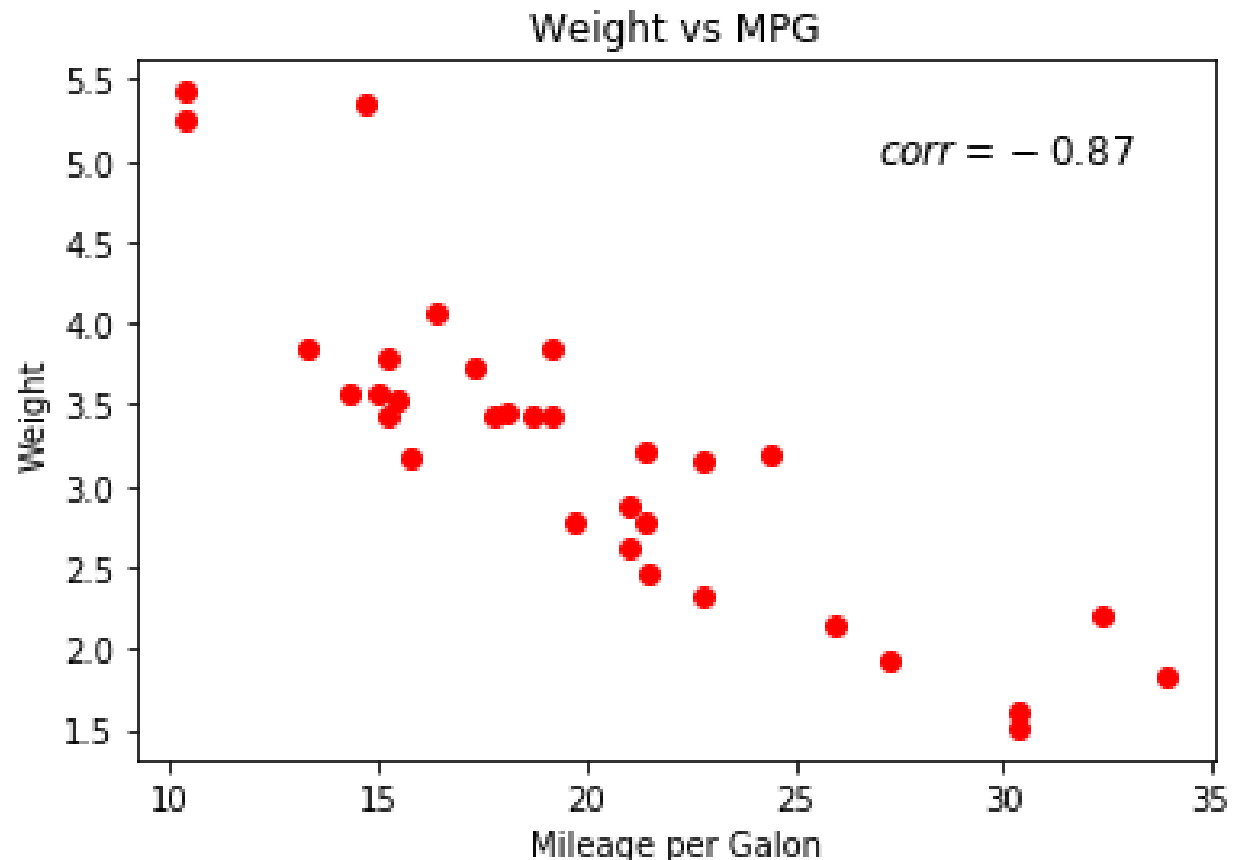
```
▶ plt.plot(mtcars.mpg,mtcars.wt,'or')  
plt.xlabel('Mileage per Galon')  
plt.ylabel('Weight')  
plt.title('Weight vs MPG')  
plt.show()
```



Menginput Teks

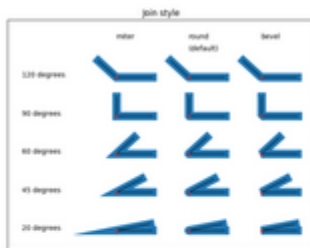
```
from scipy.stats import pearsonr
corr, _ = pearsonr(mtcars.mpg, mtcars.wt)

plt.plot(mtcars.mpg, mtcars.wt, 'or')
plt.xlabel('Mileage per Gallon')
plt.ylabel('Weight')
plt.title('Weight vs MPG')
plt.text(27, 5, r'$corr = {:.2f}$'.format(corr), fontsize=12)
plt.show()
```

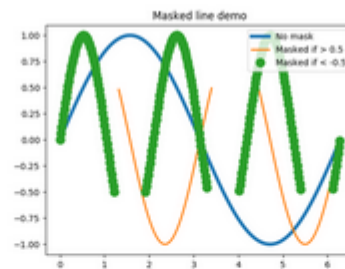


Varian Lain dan Demo

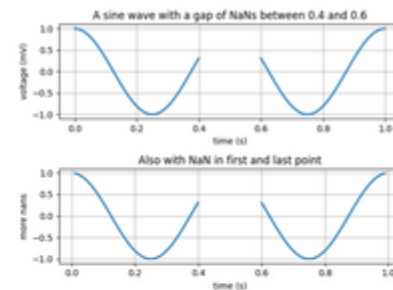
Dapat dilihat di <https://matplotlib.org/>



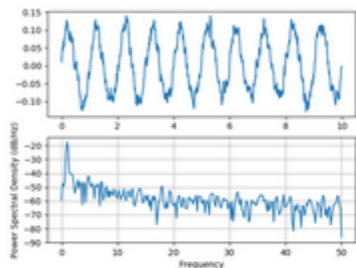
Join styles and cap styles



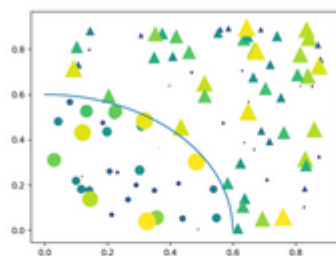
Masked Demo



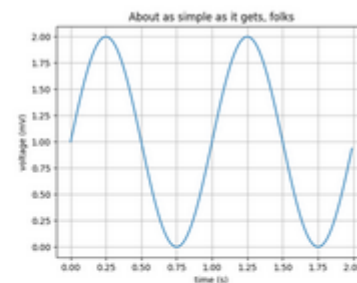
Nan Test



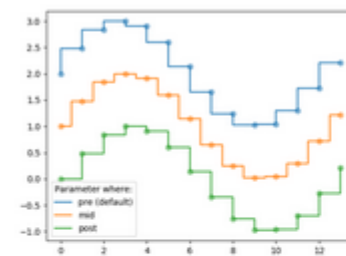
Psd Demo



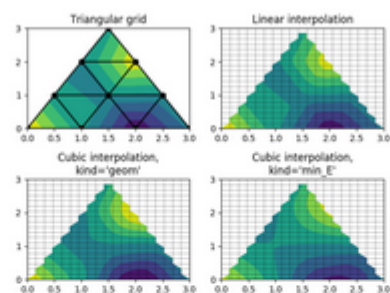
Scatter Masked



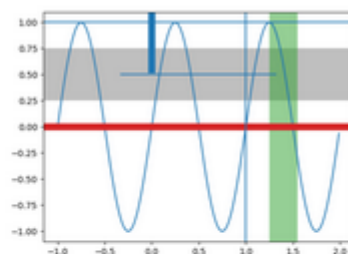
Simple Plot



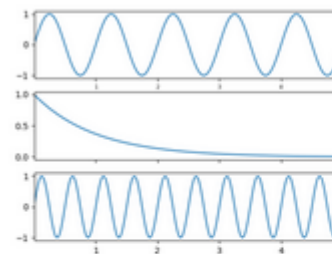
Step Demo



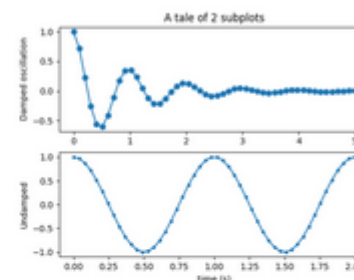
Triinterp Demo



axhspan Demo



Shared Axis Demo



Multiple subplots

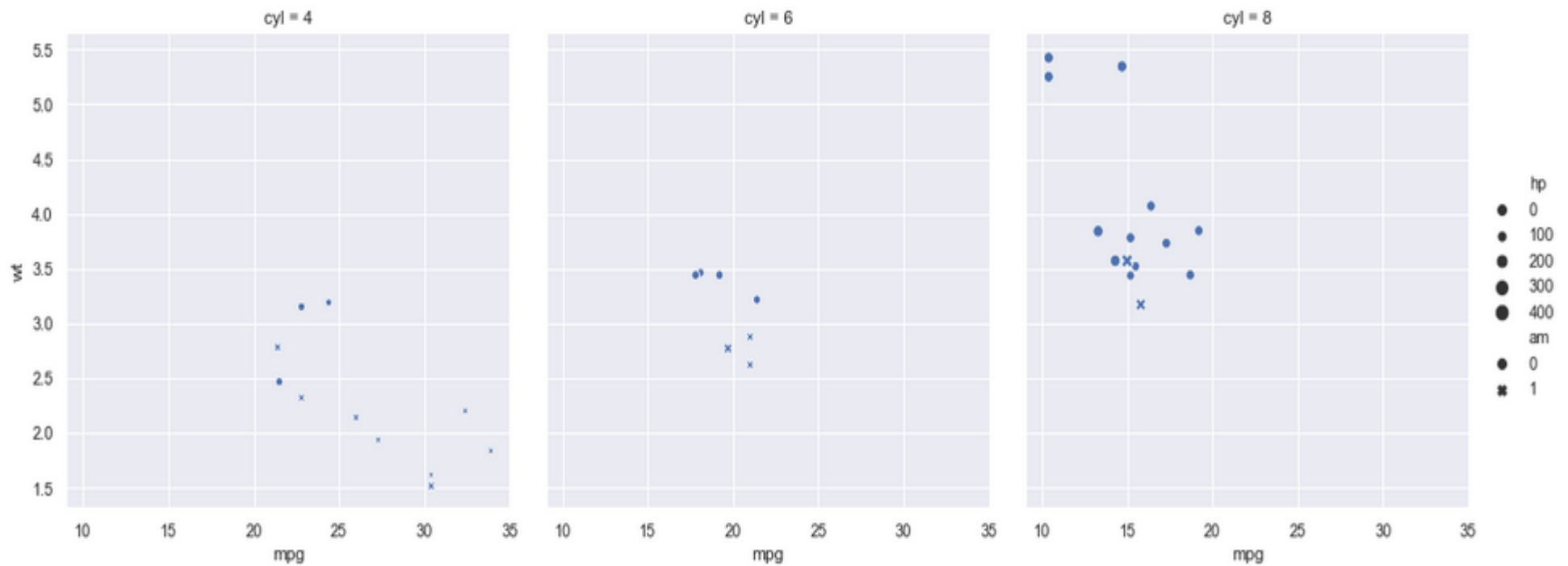
Seaborn

```
import seaborn as sns

sns.set()

sns.relplot(data=mtcars, x='mpg', y='wt', col='cyl', size='hp', style='am')
```

```
<seaborn.axisgrid.FacetGrid at 0xf792437ac8>
```

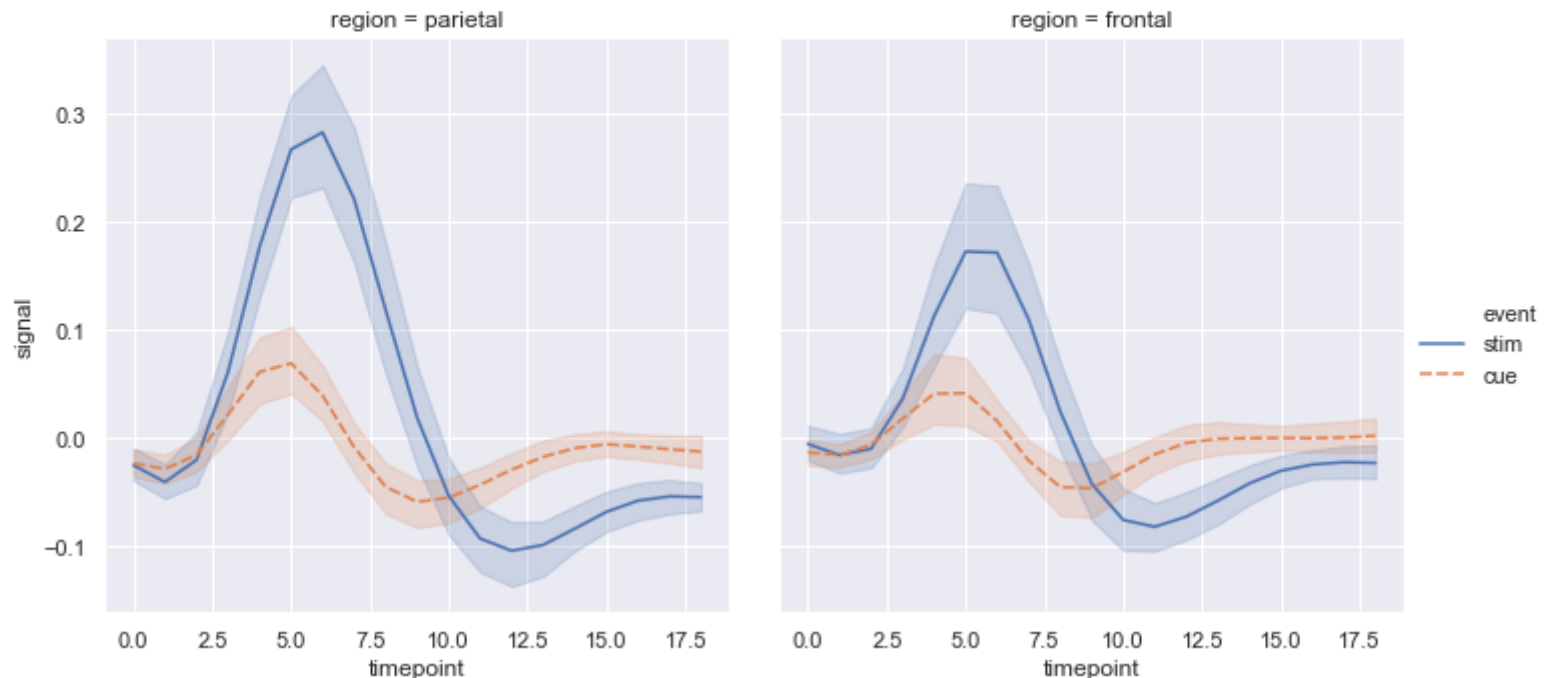


Estimasi Statistik dan Error Bars

- Seaborn dapat secara otomatis menampilkan rata-rata nilai suatu variabel sebagai fungsi dari variabel lainnya

```
fmri = sns.load_dataset('fmri')
sns.relplot(x='timepoint', y='signal', col='region',
            hue='event', style='event',
            kind='line', data=fmri)
```

```
<seaborn.axisgrid.FacetGrid at 0xf79230d080>
```



Boston House Price Dataset

- Sebagai contoh untuk memplot beberapa model berikutnya, kita gunakan boston house price dataset

```
▶ from sklearn.datasets import load_boston
boston = pd.DataFrame(load_boston().data, columns=load_boston().feature_names)
boston['MEDV'] = load_boston().target
boston.sample(5)
```

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
53	0.04981	21.0	5.64	0.0	0.439	5.998	21.4	6.8147	4.0	243.0	16.8	396.90	8.43	23.4
383	7.99248	0.0	18.10	0.0	0.700	5.520	100.0	1.5331	24.0	666.0	20.2	396.90	24.56	12.3
325	0.19186	0.0	7.38	0.0	0.493	6.431	14.7	5.4159	5.0	287.0	19.6	393.68	5.08	24.6
482	5.73116	0.0	18.10	0.0	0.532	7.061	77.0	3.4106	24.0	666.0	20.2	395.28	7.01	25.0
55	0.01311	90.0	1.22	0.0	0.403	7.249	21.9	8.6966	5.0	226.0	17.9	395.93	4.81	35.4

Description

- Bisa didapat dari fungsi `load_boston().DESCR`

```
Boston house prices dataset
```

```
-----
```

```
**Data Set Characteristics:**
```

```
:Number of Instances: 506
```

```
:Number of Attributes: 13 numeric/categorical predictive. Median Value (attribute 14) is usually the target.
```

```
:Attribute Information (in order):
```

- CRIM per capita crime rate by town
- ZN proportion of residential land zoned for lots over 25,000 sq.ft.
- INDUS proportion of non-retail business acres per town
- CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
- NOX nitric oxides concentration (parts per 10 million)
- RM average number of rooms per dwelling
- AGE proportion of owner-occupied units built prior to 1940
- DIS weighted distances to five Boston employment centres
- RAD index of accessibility to radial highways
- TAX full-value property-tax rate per \$10,000
- PTRATIO pupil-teacher ratio by town
- B $1000(B_k - 0.63)^2$ where B_k is the proportion of blacks by town
- LSTAT % lower status of the population
- MEDV Median value of owner-occupied homes in \$1000's

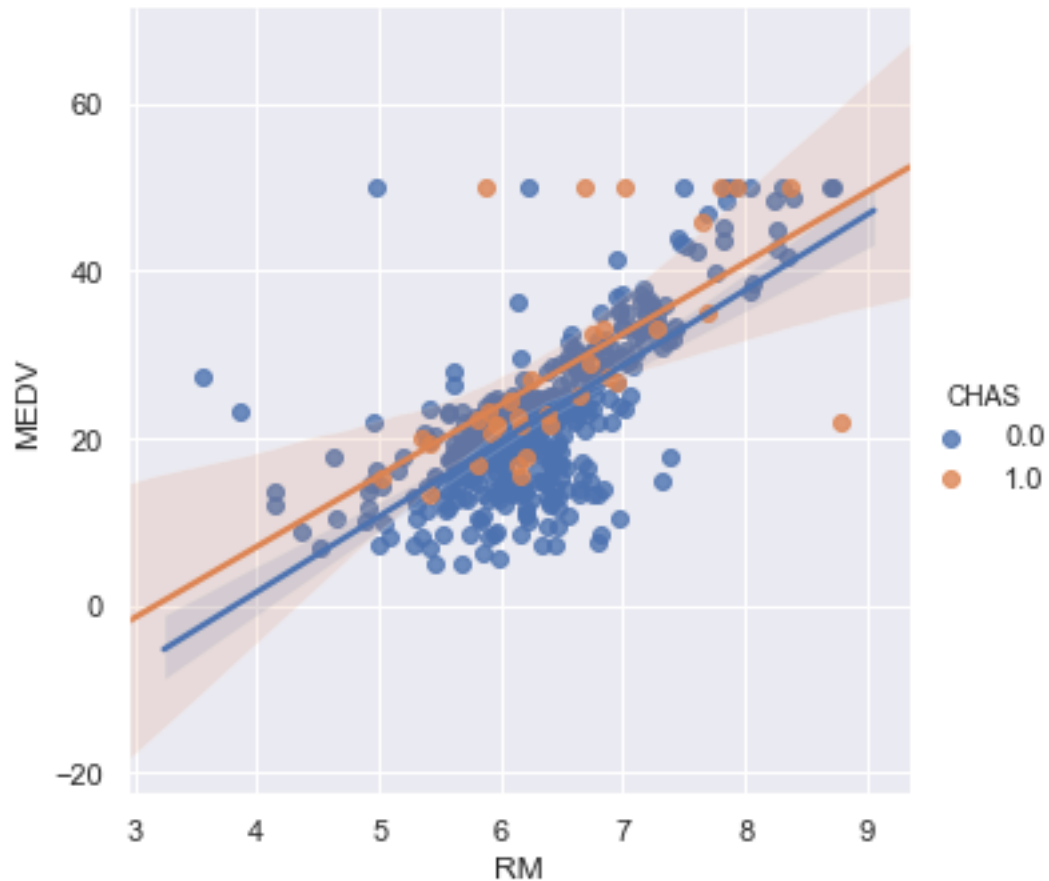
```
:Missing Attribute Values: None
```

```
:Creator: Harrison, D. and Rubinfeld, D.L.
```

LM Plot

```
sns.set()  
sns.lmplot(x='RM', y='MEDV', hue='CHAS', data=boston)
```

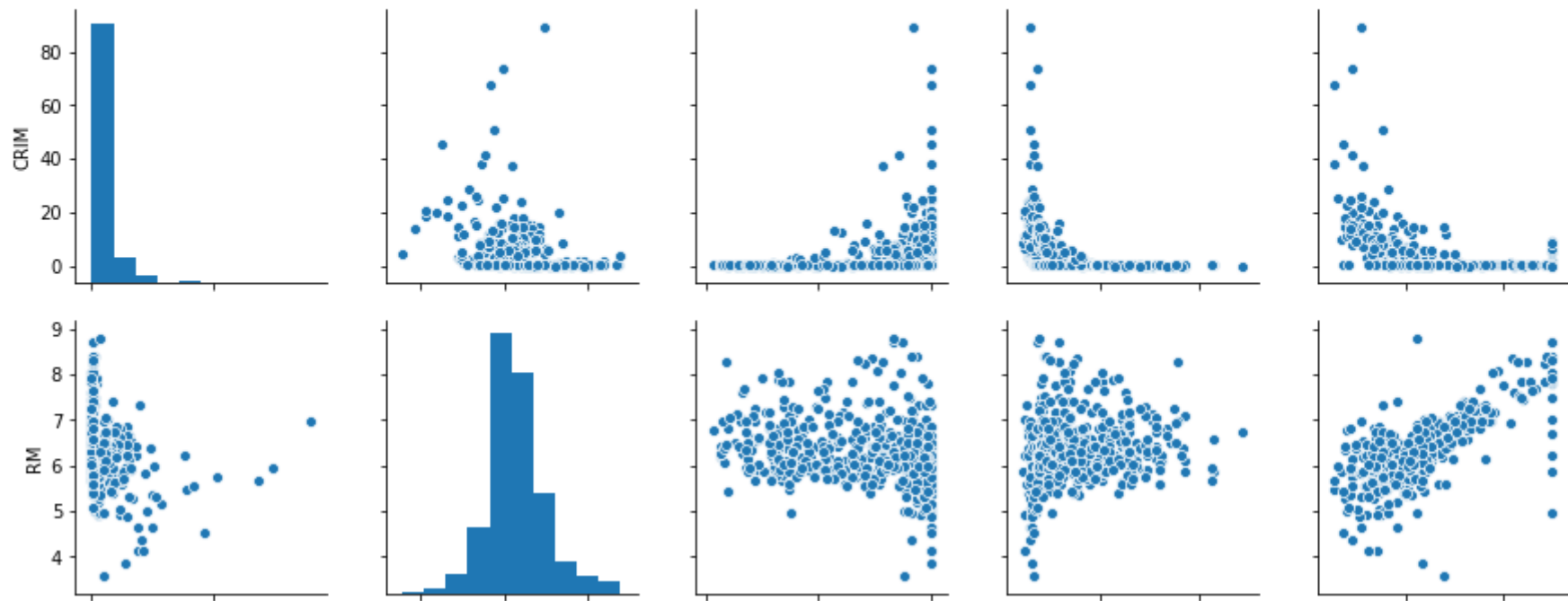
<seaborn.axisgrid.FacetGrid at 0xf79561bac8>



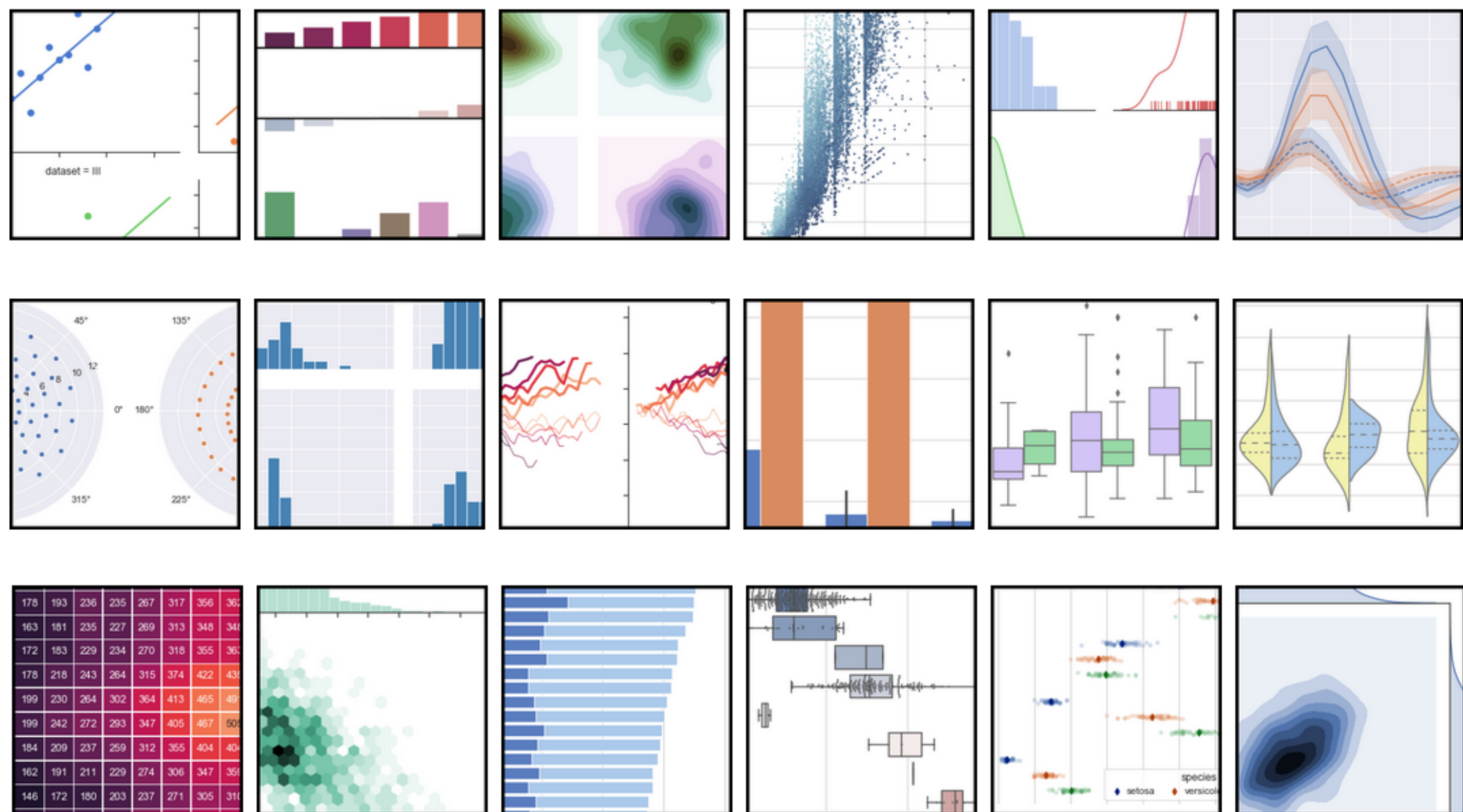
Pair Plot

```
sns.pairplot(data=boston[['CRIM', 'RM', 'AGE', 'DIS', 'MEDV']])
```

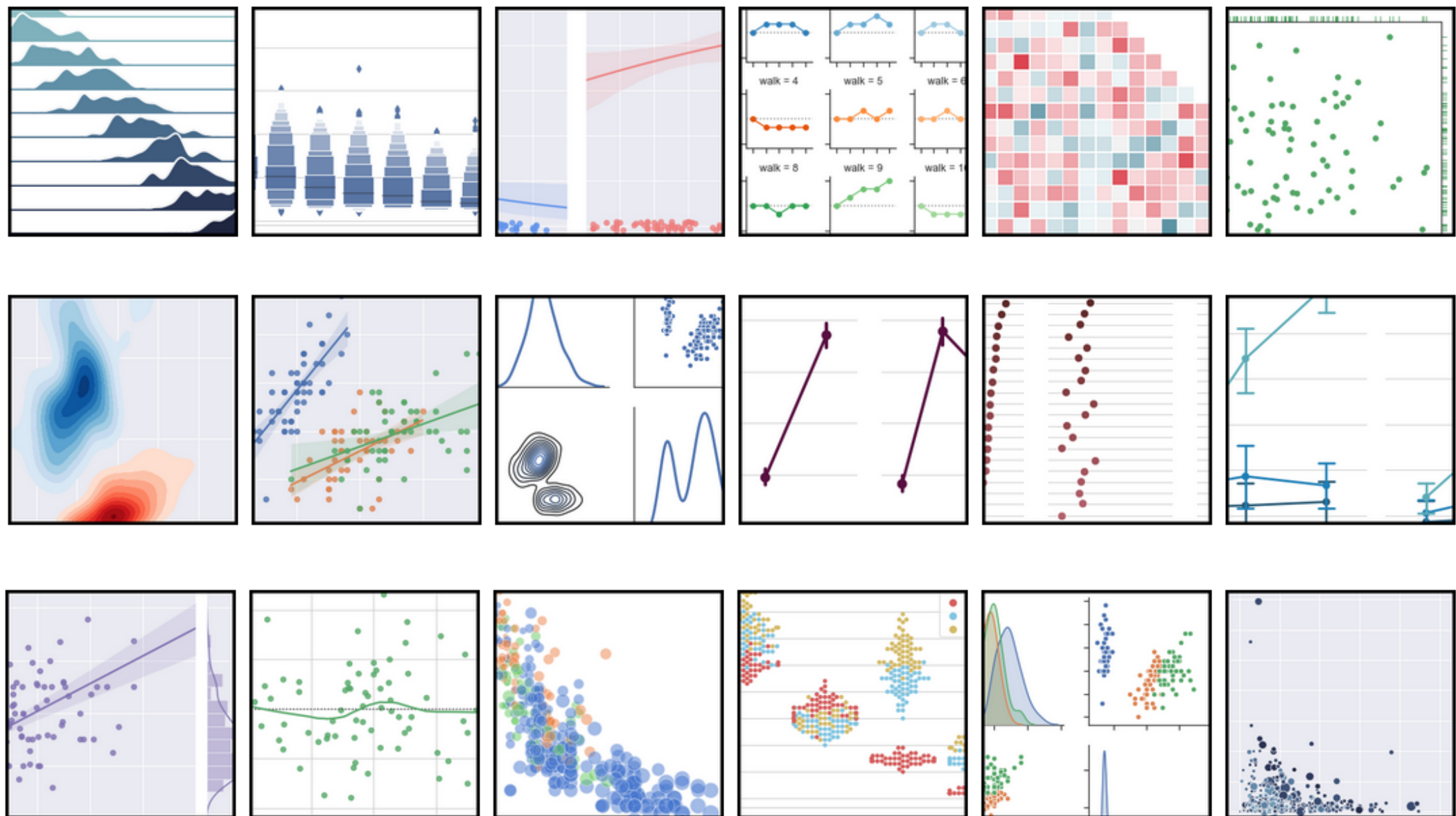
<seaborn.axisgrid.PairGrid at 0x726a218f98>



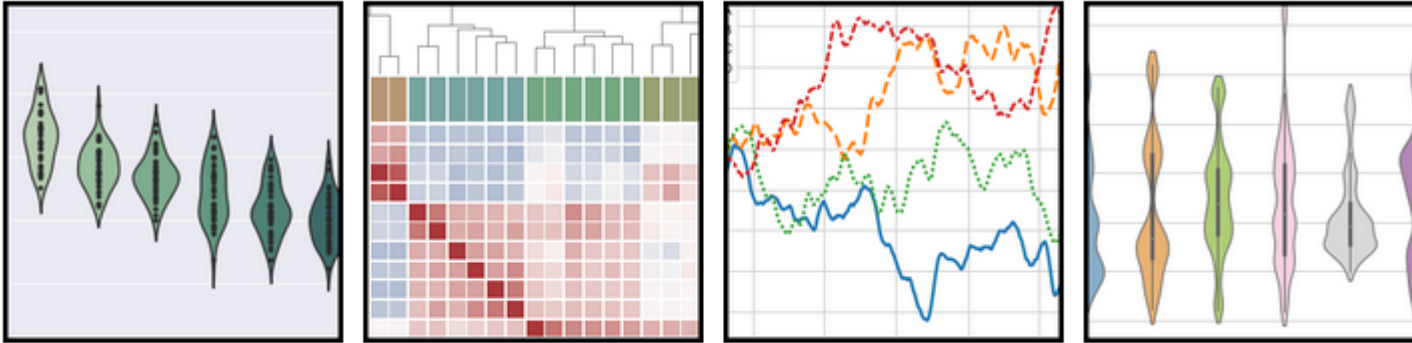
Galeri Contoh



Galeri Contoh



Galeri Contoh



<http://seaborn.pydata.org/examples/index.html>