

Python 분석 기초

모듈 - 3

강사: 장순용 박사

3. 파이썬 라이브러리 활용 분석:

3.1. Seaborn 통계 시각화.

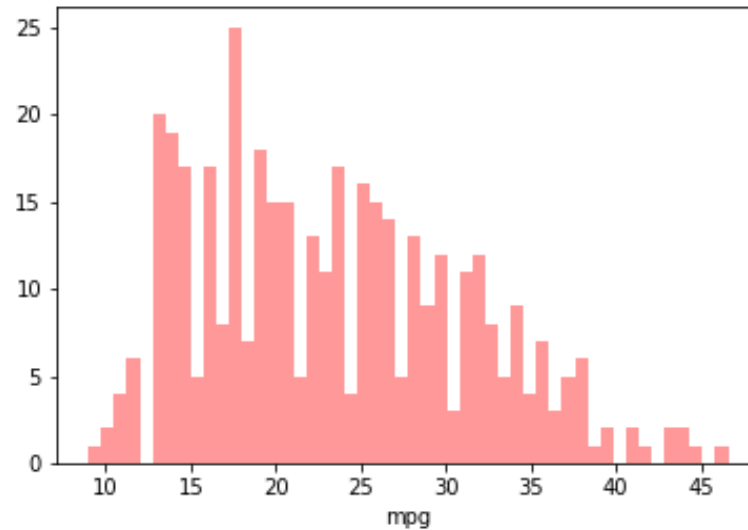
특장점:

- 내장 데이터를 제공한다: `load_dataset`. `read.csv`
- 기본 시각화 유형: `distplot`, `jointplot`, `kdeplot`, `rugplot`, `barplot`, `countplot`, 등.
- 다중 행렬 시각화 유형: `pairplot`, `PairGrid`, `FacetGrid`, 등.
- 회귀선 추가 기능: `lmplot`, `jointplot`, 등.
- 2D 특수 시각화: `heatmap`, `clustermap`, 등.
- 기본 시각화의 변형: `violinplot`, `swarmplot`, `stripplot`, 등.

Seaborn 패키지 : 히스토그램

히스토그램:

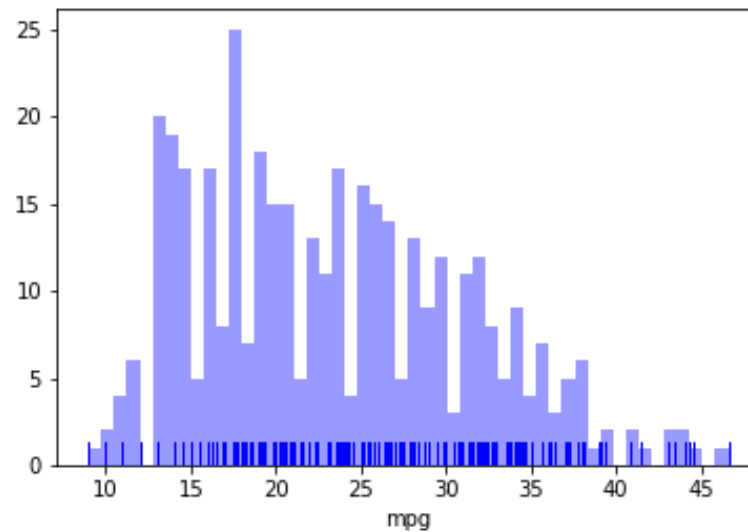
```
sns.distplot(dat.mpg, kde=False, rug=False, bins=50, color='red')  
plt.show()
```



Seaborn 패키지 : 히스토그램 + Rug

히스토그램 + Rug:

```
sns.distplot(dat.mpg, kde=False, rug=True, bins=50, color='blue')  
plt.show()
```



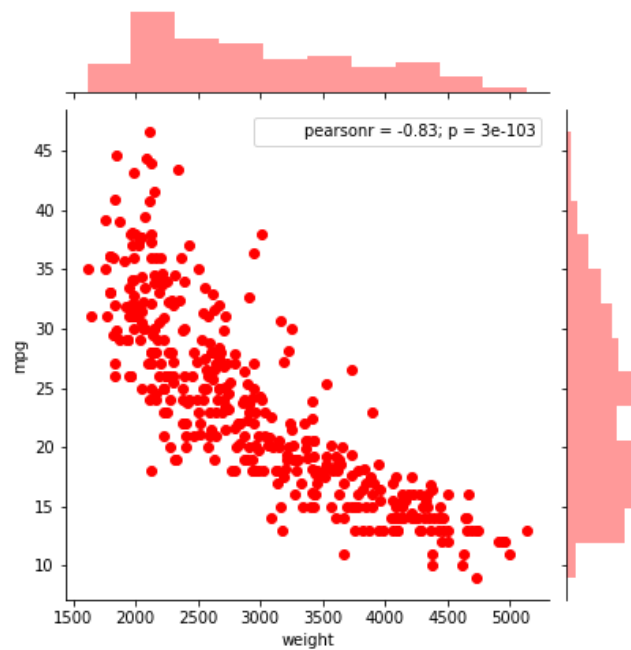
Seaborn 패키지 : 산점도

산점도:

2

가

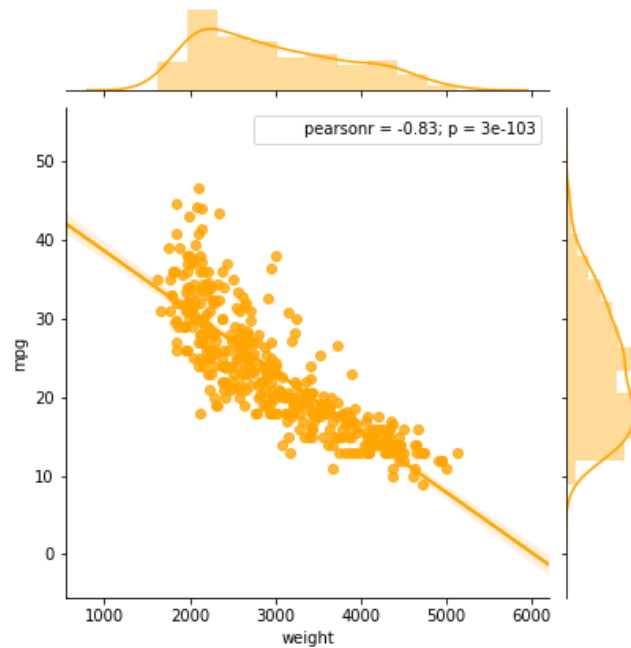
```
sns.jointplot(x='weight', y='mpg', data=dat, color='red', scatter kind='scatter' )  
plt.show()
```



Seaborn 패키지 : 산점도 + 회귀선

산점도 + 회귀선:

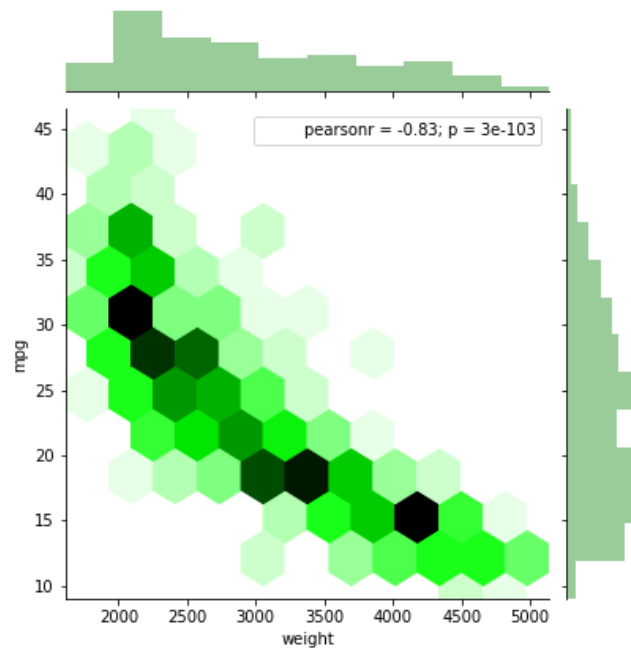
```
sns.jointplot(x='weight', y='mpg', data=dat, color='orange', kind='reg' )  
plt.show()
```



Seaborn 패키지 : Hex

Hex:

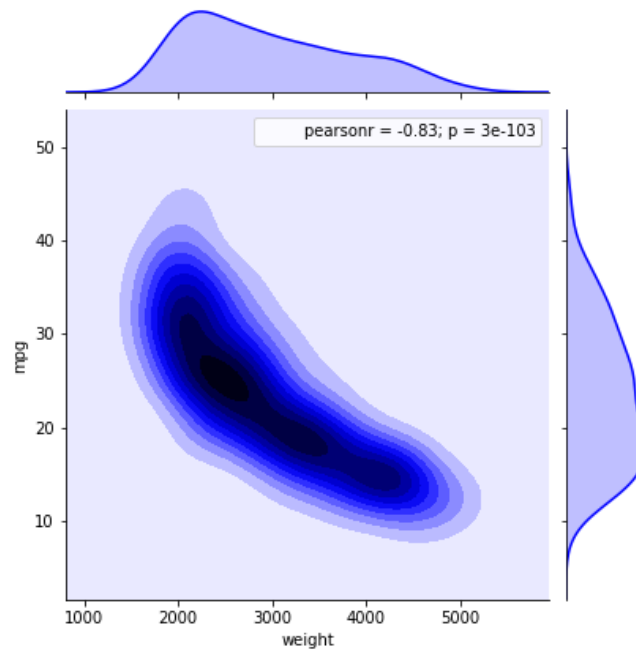
```
sns.jointplot(x='weight', y='mpg', data=dat, color='green', kind='hex' )  
plt.show()
```



Seaborn 패키지 : KDE

KDE:

```
sns.jointplot(x='weight', y='mpg', data=dat, color='blue', kind='kde' )  
plt.show()
```

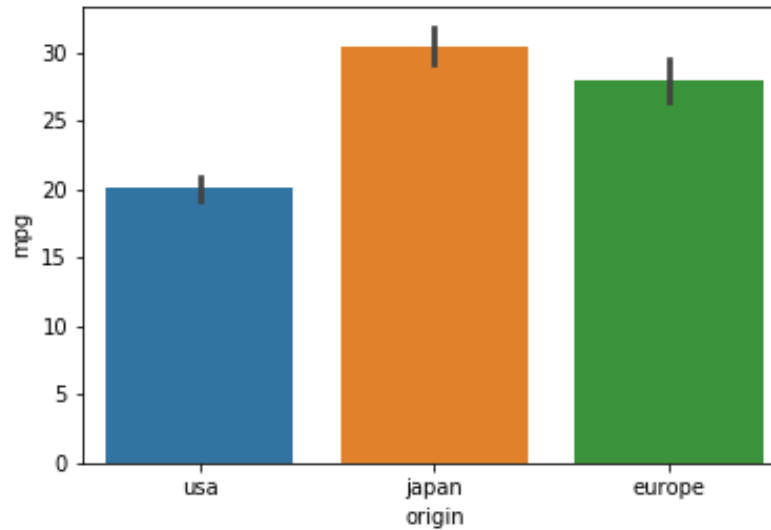


Seaborn 패키지 : 막대그림

막대그림:

- 1.
- 2.

```
sns.barplot(x='origin', y='mpg', data=dat)  
plt.show()
```



Seaborn 패키지 : 막대그림

막대그림:

가 2 가

?

또다른 변수를 인자 hue의 값으로 포함.

```
sns.countplot(x='origin', data=dat, hue='cylinders')
```

```
plt.show()
```

가

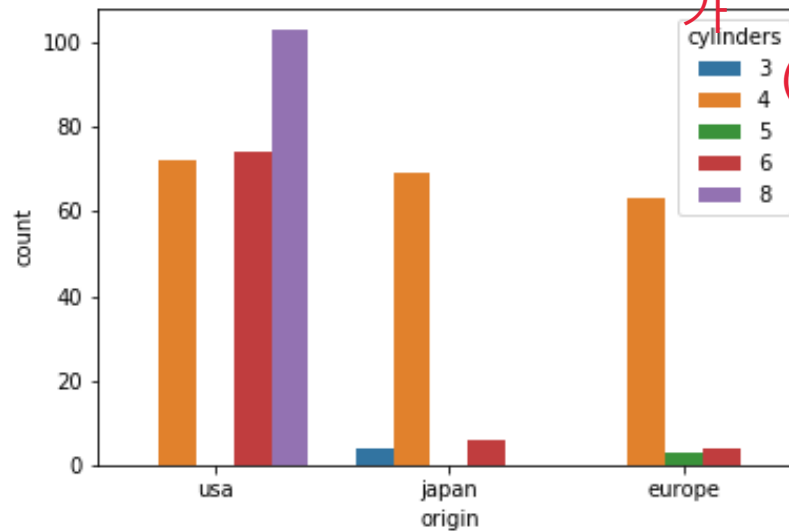
hue

2

a,b /

가

가

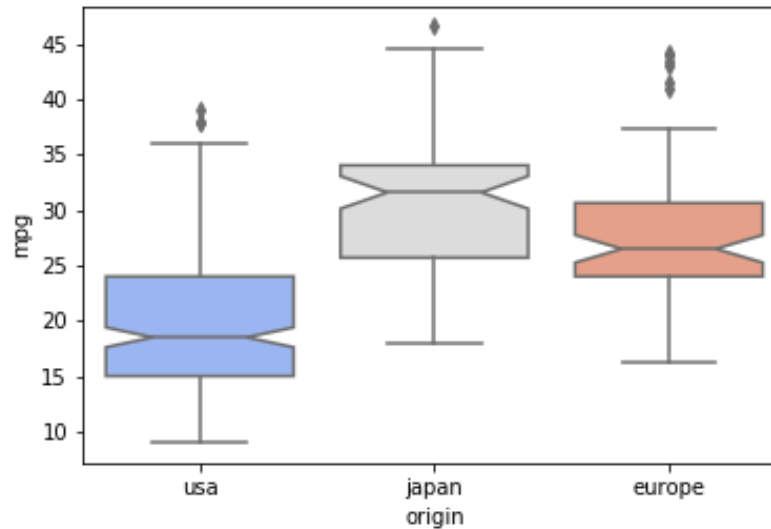


Seaborn 패키지 : 상자그림

상자그림:

다중 상자그림.

```
sns.boxplot(x='origin', y='mpg', data=dat, palette='coolwarm', notch=True)  
plt.show()
```

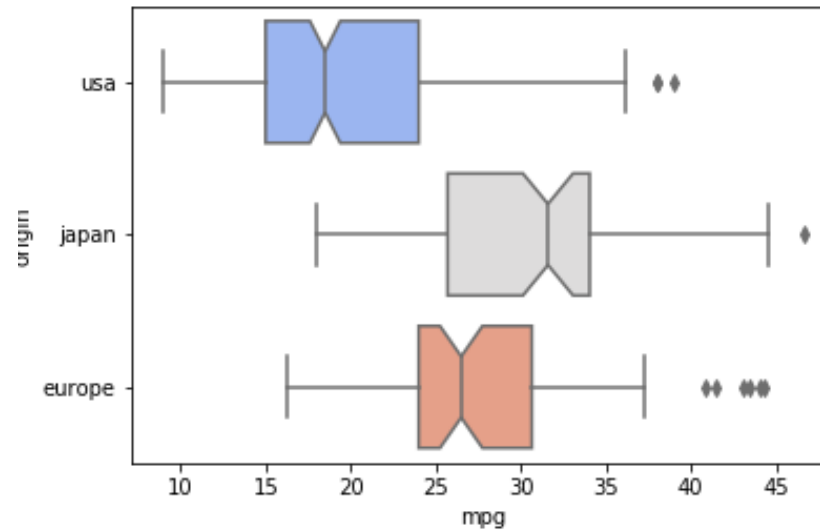


Seaborn 패키지 : 상자그림

상자그림:

다중 상자그림.

```
sns.boxplot(x='mpg', y='origin', data=dat, palette='coolwarm', notch=True)  
plt.show()
```

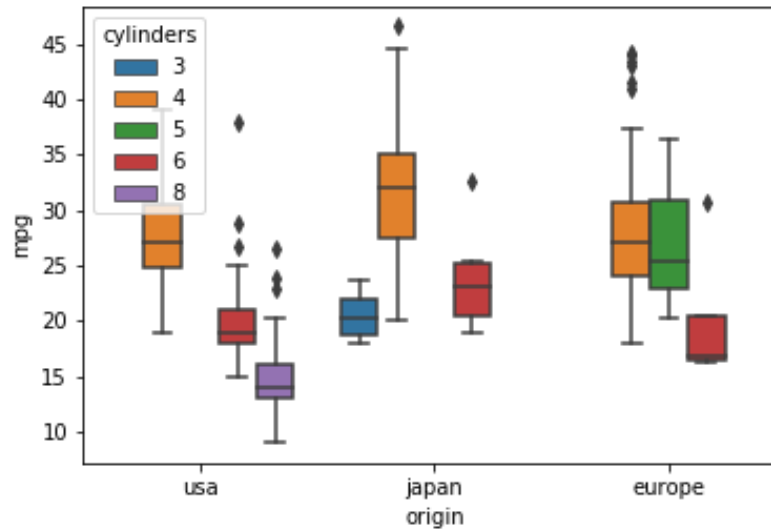


Seaborn 패키지 : 상자그림

상자그림:

인자 hue의 값으로 또 다른 변수 포함.

```
sns.boxplot(x='origin', y='mpg', data=dat, hue='cylinders' )  
plt.show()
```

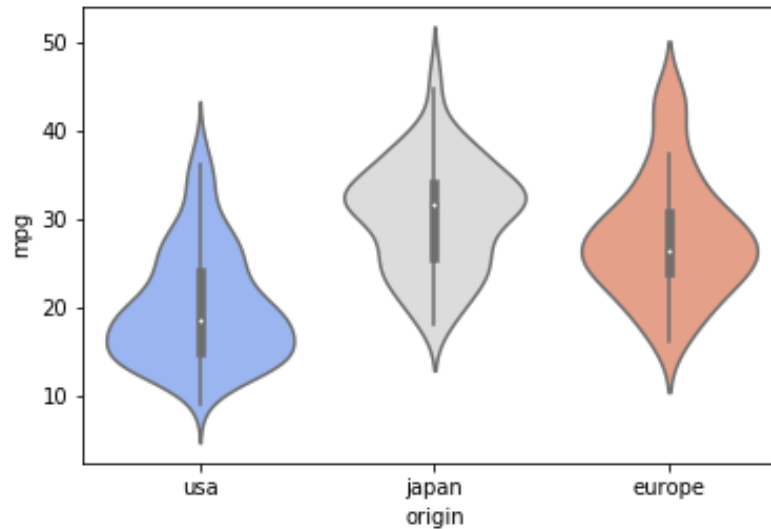


Seaborn 패키지 : 바이올린 플롯

바이올린 플롯:

boxplot과 인자가 같음.

```
sns.violinplot(x='origin', y='mpg', data=dat, palette='coolwarm' )  
plt.show()
```

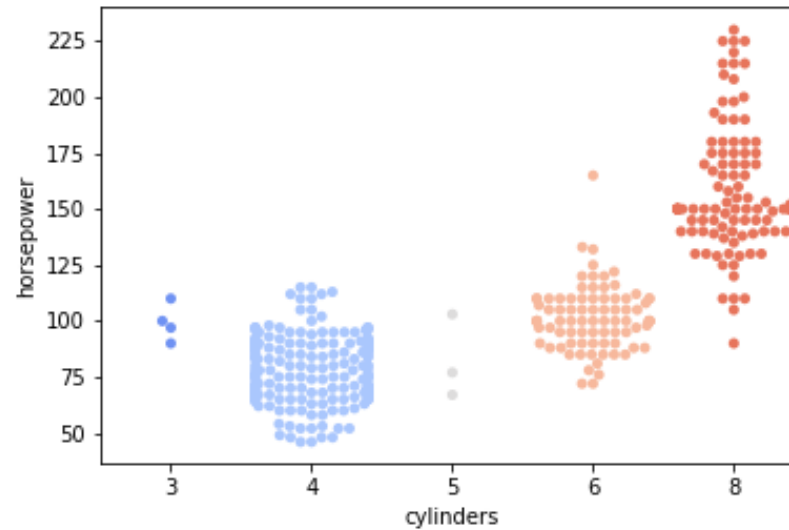


Seaborn 패키지 : Swarm plot

Swarm plot:

가

```
sns.swarmplot(x='cylinders', y='horsepower', data=dat, palette='coolwarm')  
plt.show()
```



실습 #0301

→ Seaborn 시각화 I. ←

→ 사용: **ex_0301.ipynb** ←

```
1 -  
df.value_counts()  
df.gender.value_counts()
```

```
2 -  
pd.crosstab(1, 2)  
df.crosstab(df.gender, df.bloodtype)
```

```
[
```

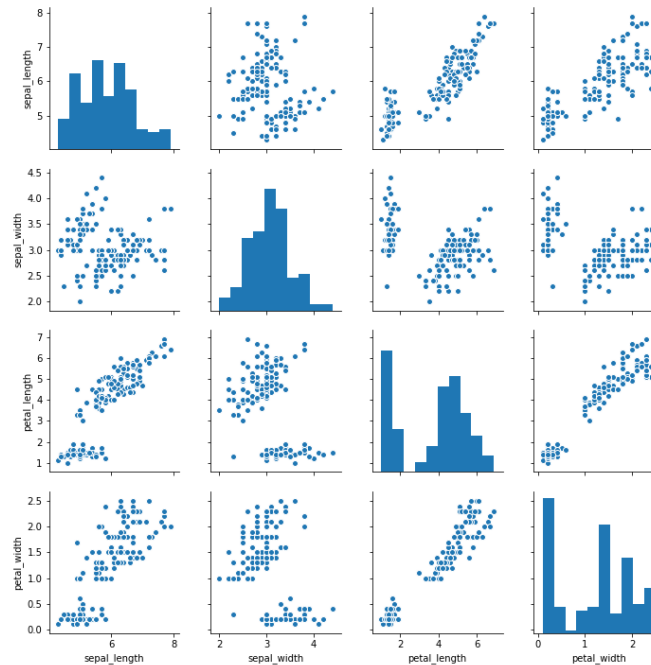
!!

Seaborn 패키지 : 산점도 행렬

산점도 행렬:

```
sns.pairplot(dat)
```

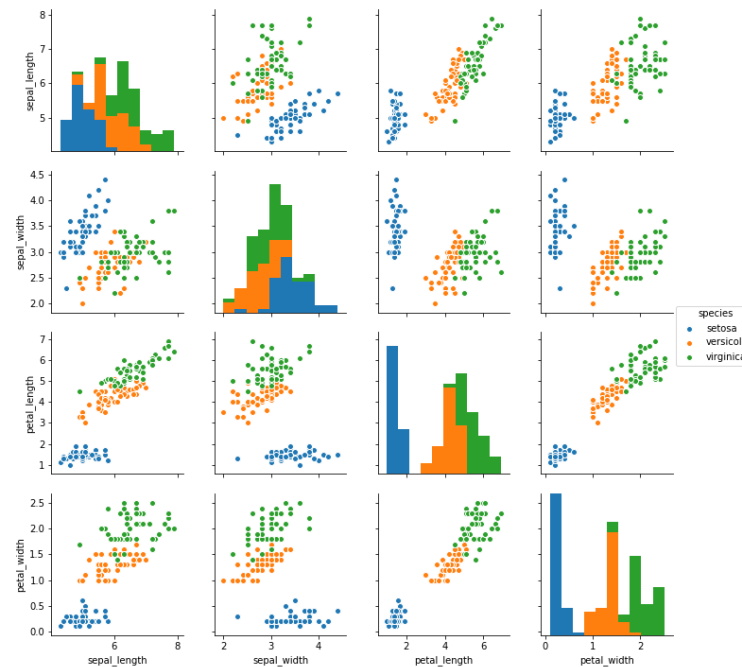
```
plt.show()
```



Seaborn 패키지 : 산점도 행렬

산점도 행렬:

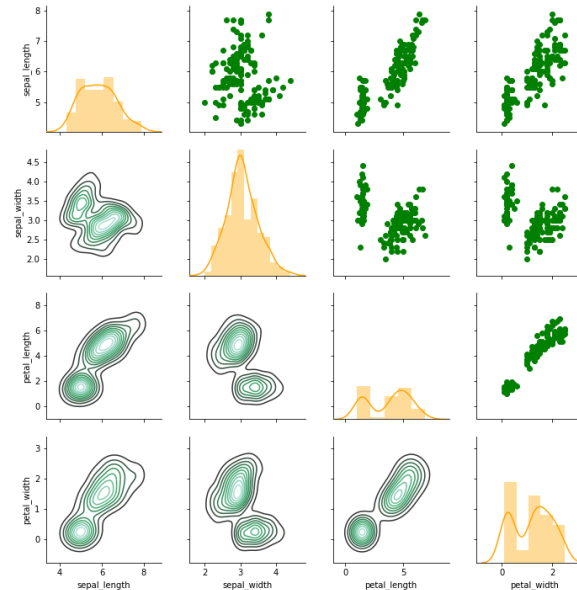
```
sns.pairplot(dat, hue='species')  
plt.show()
```



Seaborn 패키지 : 혼합 유형 시각화 행렬

혼합 유형 시각화 행렬:

```
g=sns.PairGrid(dat)
g.map_diag(sns.distplot, color='orange')      # 대각선 = 히스토그램.
g.map_upper(plt.scatter, color='green')       # 위 삼각 = 산점도.
g.map_lower(sns.kdeplot, color='blue')        # 아래 삼각 = KDE.
plt.show()
```



upper scatter

lower kdeplot

distplot

Seaborn 패키지 : 다중 시각화

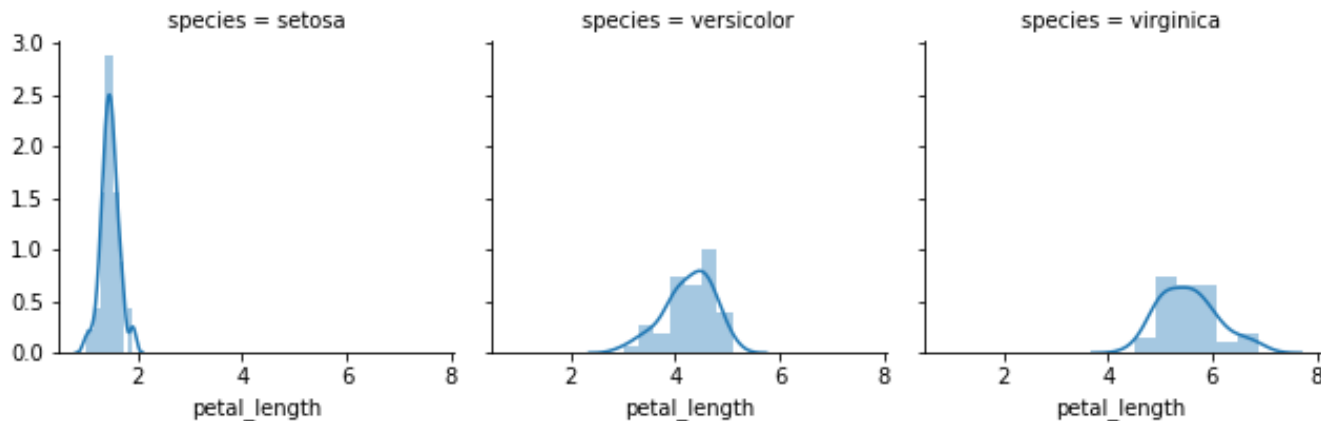
다중 시각화:

다중 히스토그램.

```
g=sns.FacetGrid(data=dat, col='species') species가
```

```
g.map(sns.distplot, 'petal_length') distplot
```

```
plt.show()
```



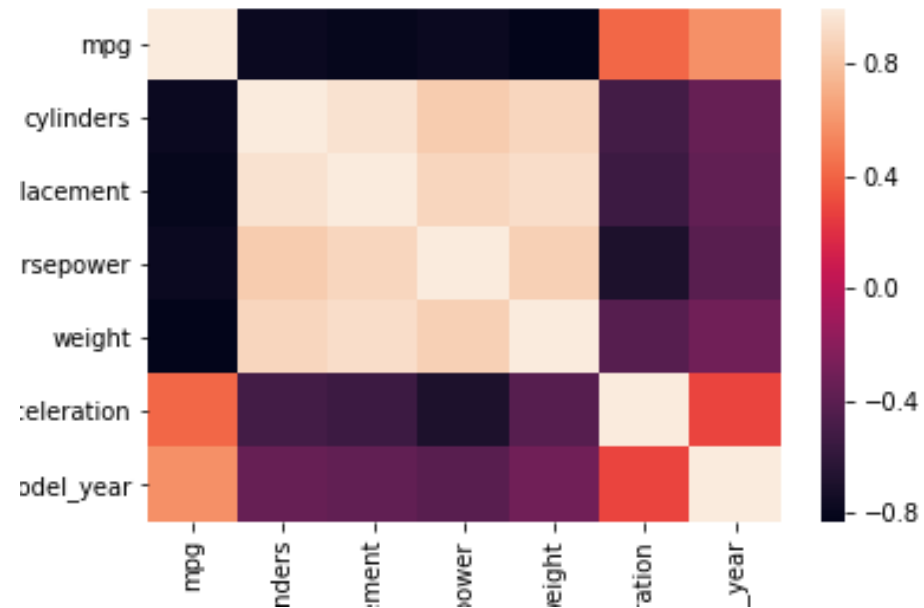
Seaborn 패키지 : Heatmap

Heatmap:

```
sns.heatmap(x)
plt.show()
```

x가

heat



실습 #0302

→ Seaborn 시각화 II. ←

→ 사용: **ex_0302.ipynb** ←

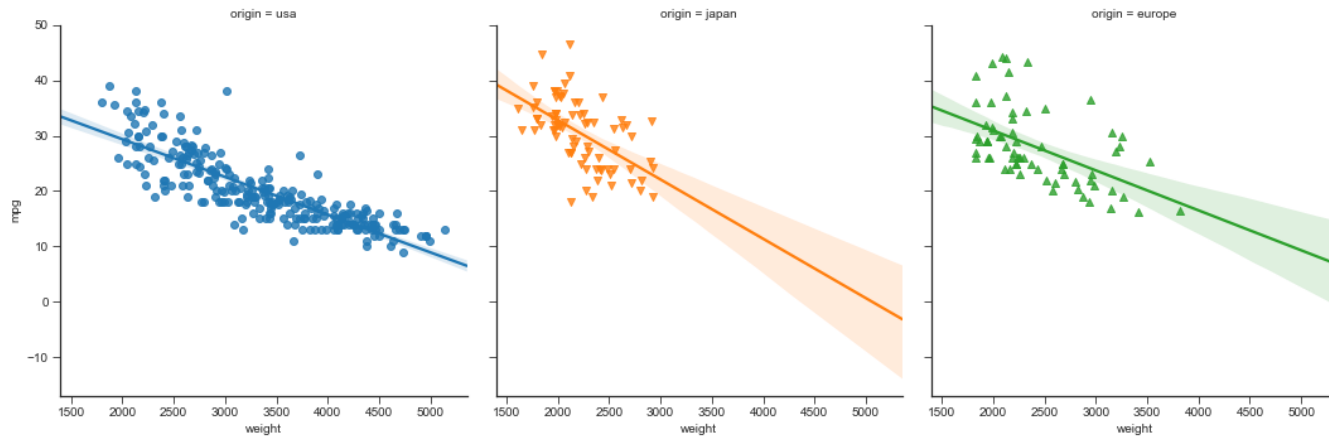
Seaborn 패키지 : 산점도 + 회귀선

산점도 + 회귀선:

다중 시각화.

```
sns.lmplot(data=dat, x='weight', y='mpg', col = 'origin', hue = 'origin')  
plt.show()
```

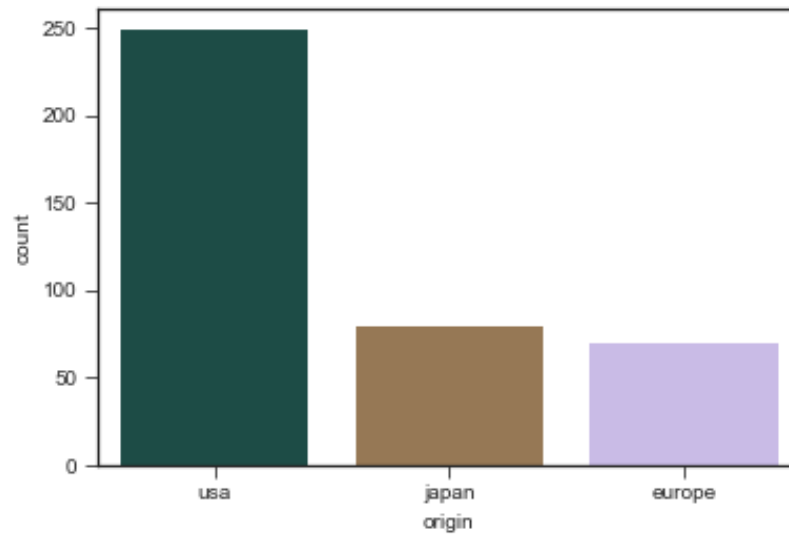
?



Seaborn 패키지 : 컬러 Palette

컬러 Palette:

```
sns.countplot(x='origin', data=dat, palette='cubehelix')  
# sns.countplot(x='origin', data=dat, palette='coolwarm')  
plt.show()
```



실습 #0303

→ Seaborn 시각화 III. ←

→ 사용: **ex_0303.ipynb** ←

문의:

sychang1@gmail.com