

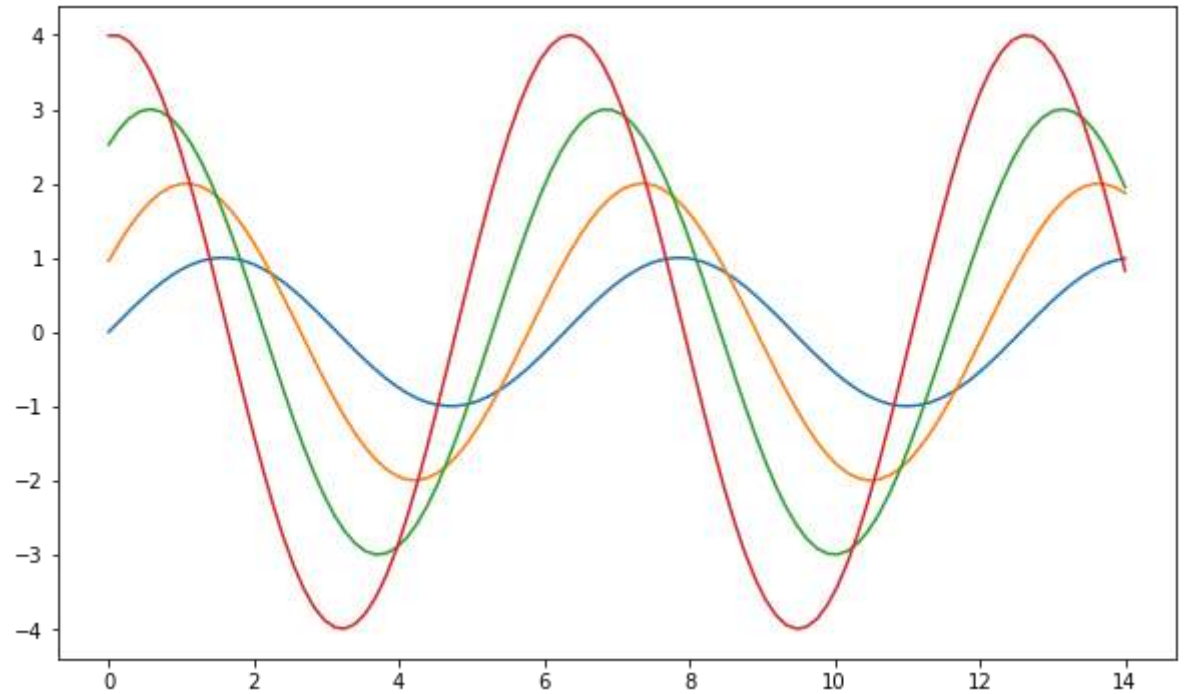
seaborn

- `import numpy as np`
- `import matplotlib.pyplot as plt`
- `%matplotlib inline`
- `import seaborn as sb`

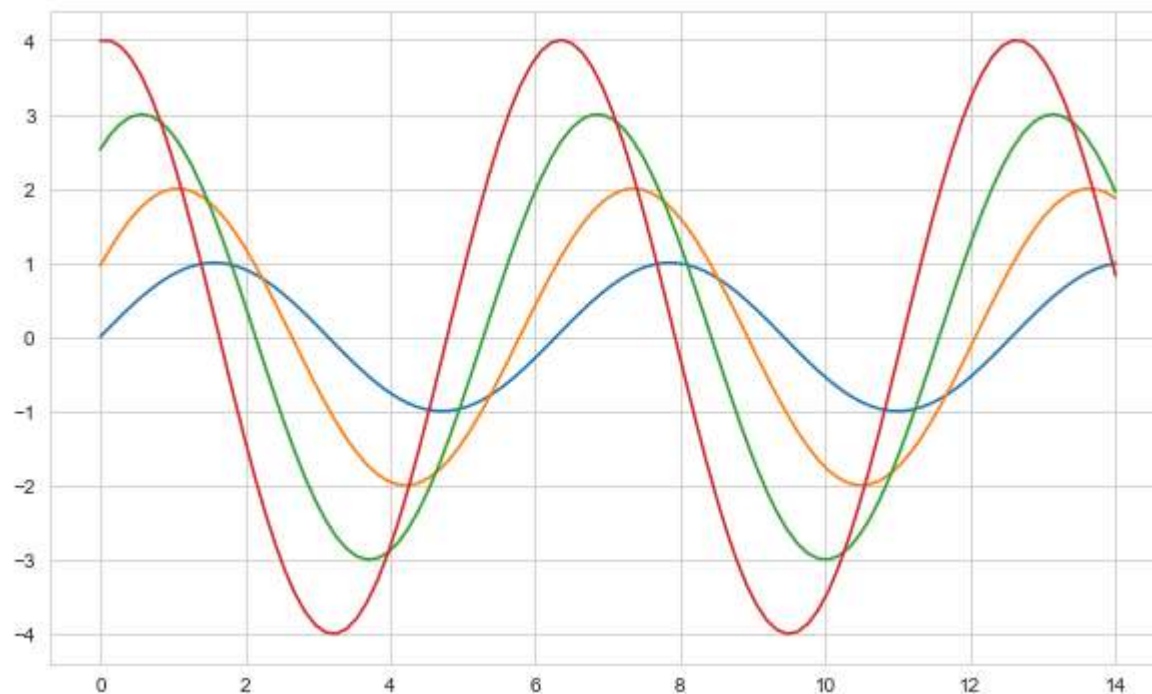
seaborn은 matplotlib보다 더욱 디자인적 요소를 가지고 있음

- `x = np.linspace(0,14,100)`
- `y1 = np.sin(x)`
- `y2 = 2*np.sin(x+0.5)`
- `y3 = 3*np.sin(x+1)`
- `y4 = 4*np.sin(x+1.5)`

- `plt.figure(figsize=(10,6))`
- `plt.plot(x,y1,x,y2,x,y3,x,y4)`



- `sb.set_style('whitegrid')`
*# seaborn의 whitegrid/darkgrid
스타일 적용*
- `plt.figure(figsize=(10,6))`
- `plt.plot(x,y1,x,y2,x,y3,x,y4)`
- `plt.show()`



*# tips: 요일별 점심 저녁 흡연여부와
식사금액과 팁을 정리한 데이터*

➤ `tips = sb.load_dataset('tips')`

➤ `tips.head()`

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

- `plt.figure(figsize=(10,6))`
- `sb.boxplot(x='day', y='total_bill', data=tips)`
- `plt.show()`

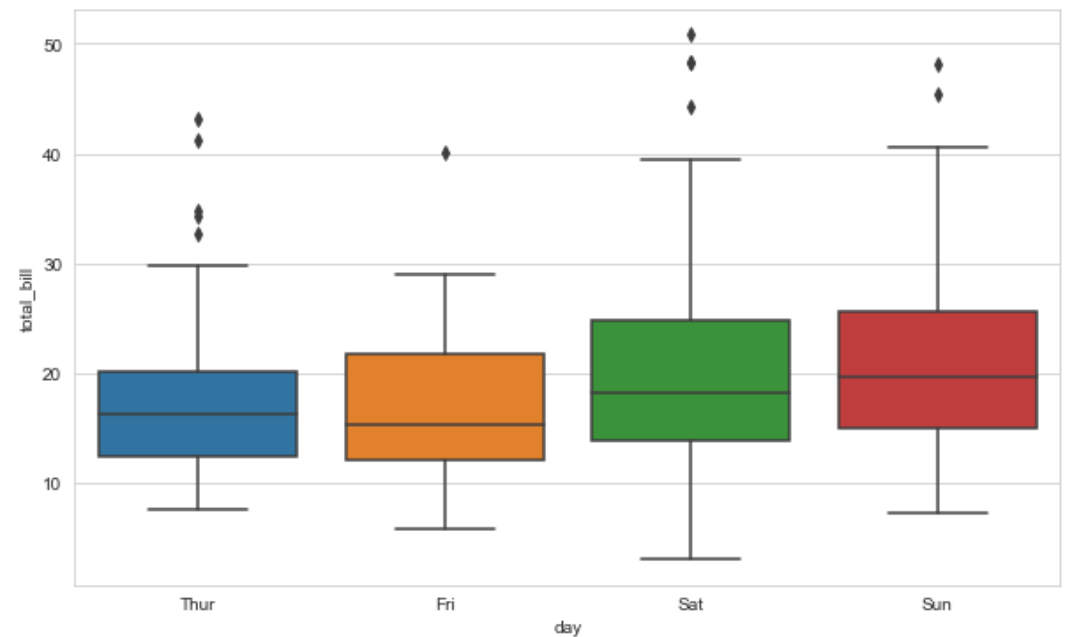
boxplot

선은 최대값, 최소값

점은 최대값, 최소값을 벗어난 이상값

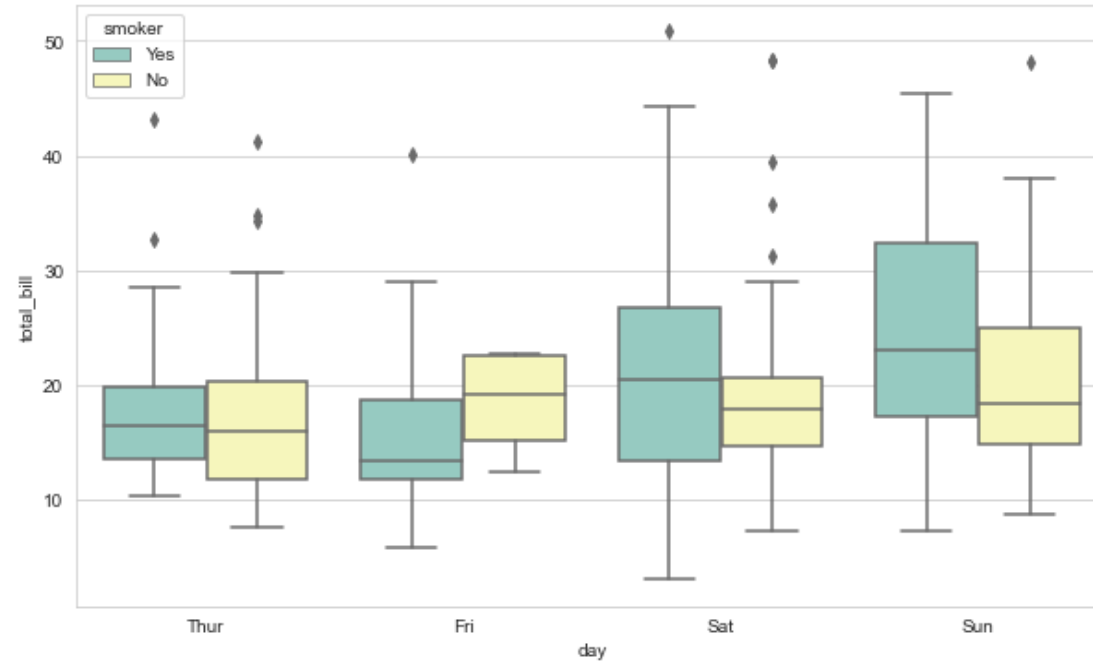
박스 영역은 25%~75%

박스 안의 선은 중앙값을 나타냄

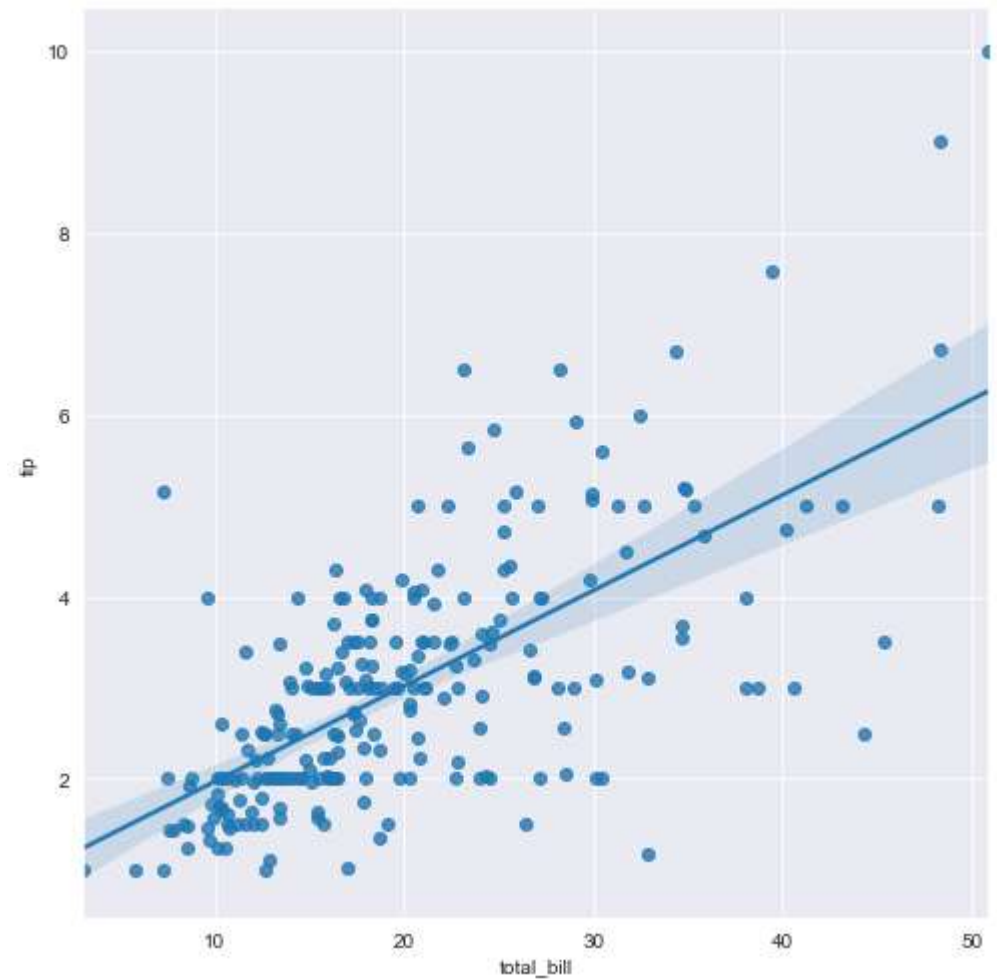


hue: 인수에 카테고리 변수 이름을 지정, 카테고리값에 따라 색상을 다르게 지정
palette: 준비된 디자인 중 하나를 선택함. Set1~3까지(smoker의 범례 색상 참조)

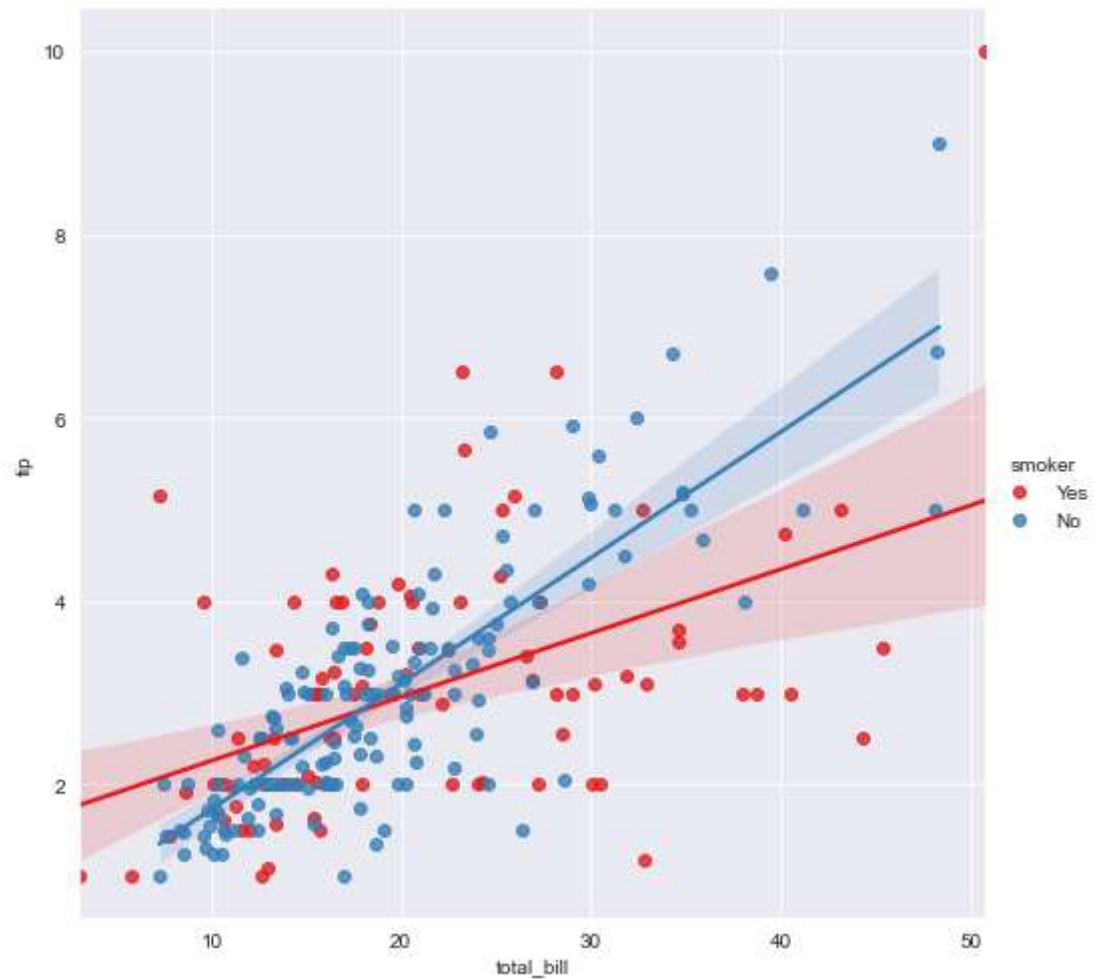
- plt.figure(figsize=(10,6))
- sb.boxplot(x='day', y='total_bill', data=tips, hue='smoker', palette='Set3')
- plt.show()



- `sb.set_style('darkgrid')`
- `sb.lmplot(x='total_bill', y='tip', data=tips, size=7)`
- # lmplot: 산포도와 추세선을 표시. 1가 아니라 L*
- `plt.show()`



- `sb.lmplot(x='total_bill', y='tip', hue='smoker', palette='Set1', data=tips, size=7)`
- `plt.show()`



연도별 월별 항공기 승객수 기록한 데이터

➤ flights = sb.load_dataset('flights')

➤ flights.head()

	year	month	passengers
0	1949	January	112
1	1949	February	118
2	1949	March	132
3	1949	April	129
4	1949	May	121

➤ flights =
flights.pivot('month','year','passengers')

pivot: 엑셀 피벗테이블과 유사(행, 열, 데이터 값)

➤ flights.head()

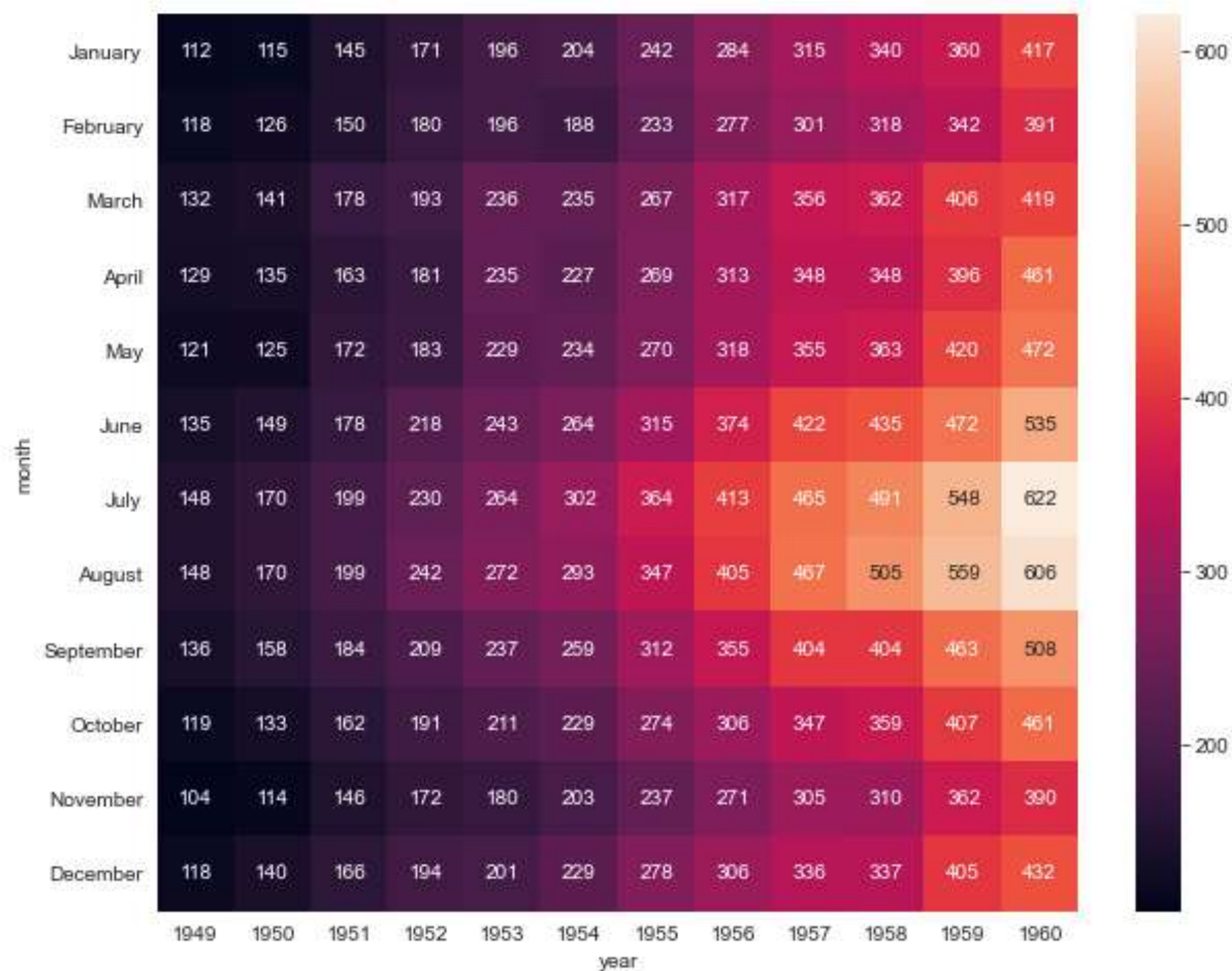
year	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
month												
January	112	115	145	171	196	204	242	284	315	340	360	417
February	118	126	150	180	196	188	233	277	301	318	342	391
March	132	141	178	193	236	235	267	317	356	362	406	419
April	129	135	163	181	235	227	269	313	348	348	396	461
May	121	125	172	183	229	234	270	318	355	363	420	472

- plt.figure(figsize=(10,8))
- sb.heatmap(flights, annot=True, fmt='d')

heatmap: 열분포 형태의 차트

annot=True: 셀에다 데이터 값을 씀

fmt='d': 정수형 선언, fmt='.2f' : 실수형 소수점 2자리까지

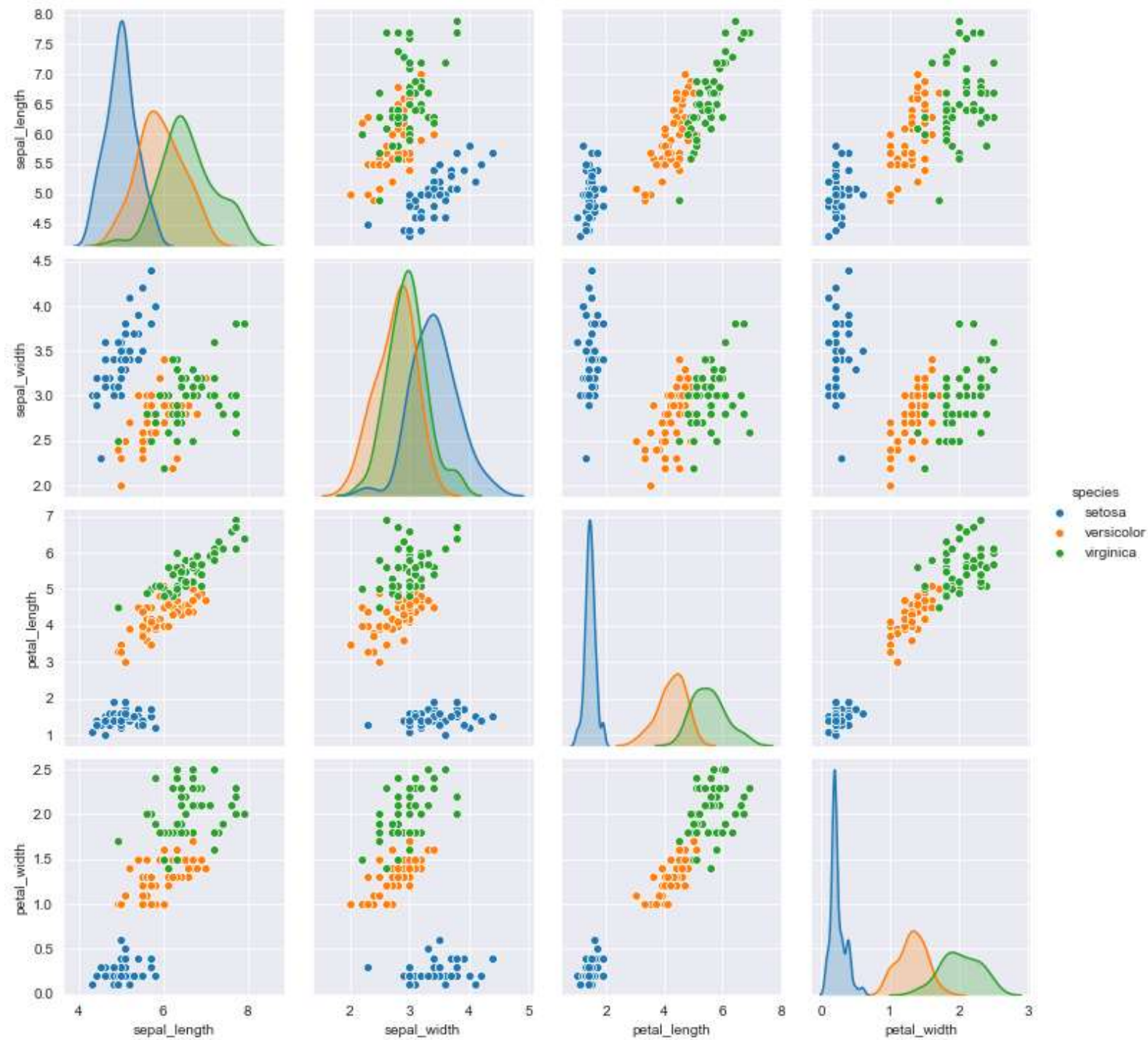


- `iris = sb.load_dataset('iris')`
- `iris.head()`

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

- `sb.pairplot(iris, hue='species')`
- `plt.show()`

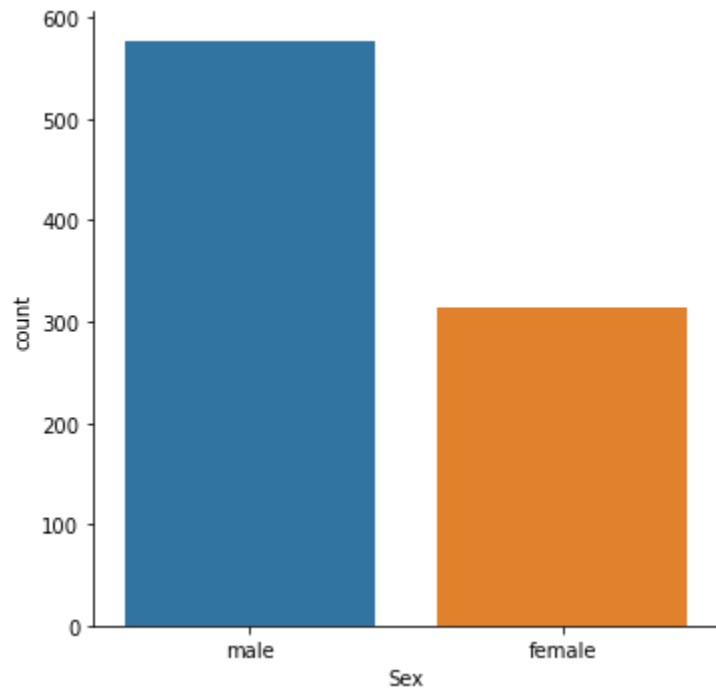
*# pairplot: column별 데이터에
대한 상관관계나 분류적 특성
표현*



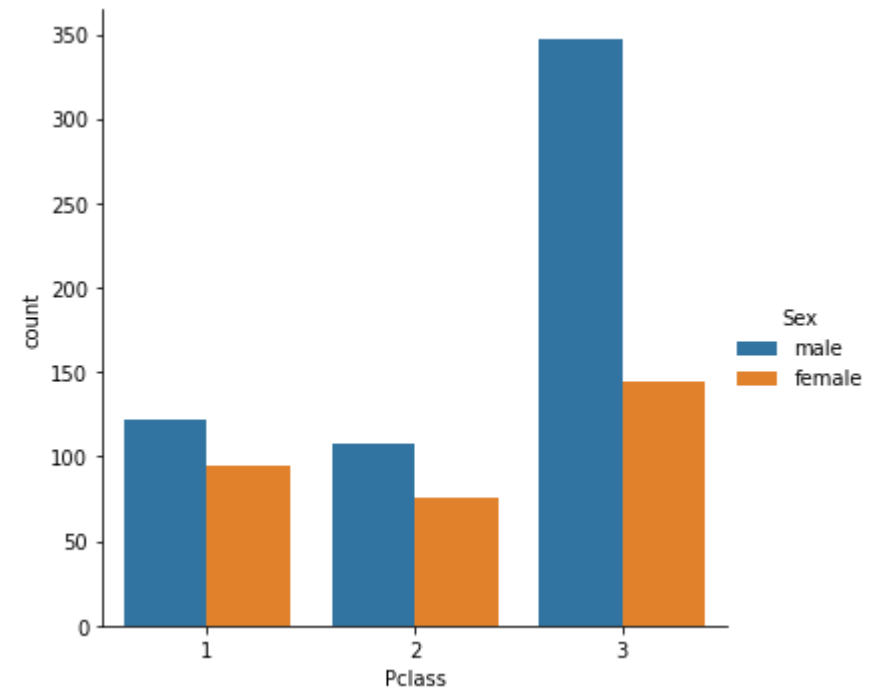
- `import seaborn as sb`
- `import numpy as np`
- `import pandas as pd`
- `import matplotlib.pyplot as plt`
- `%matplotlib inline`

- `titanic_df = pd.read_csv('train.csv')`
- `titanic_df.head()`

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S



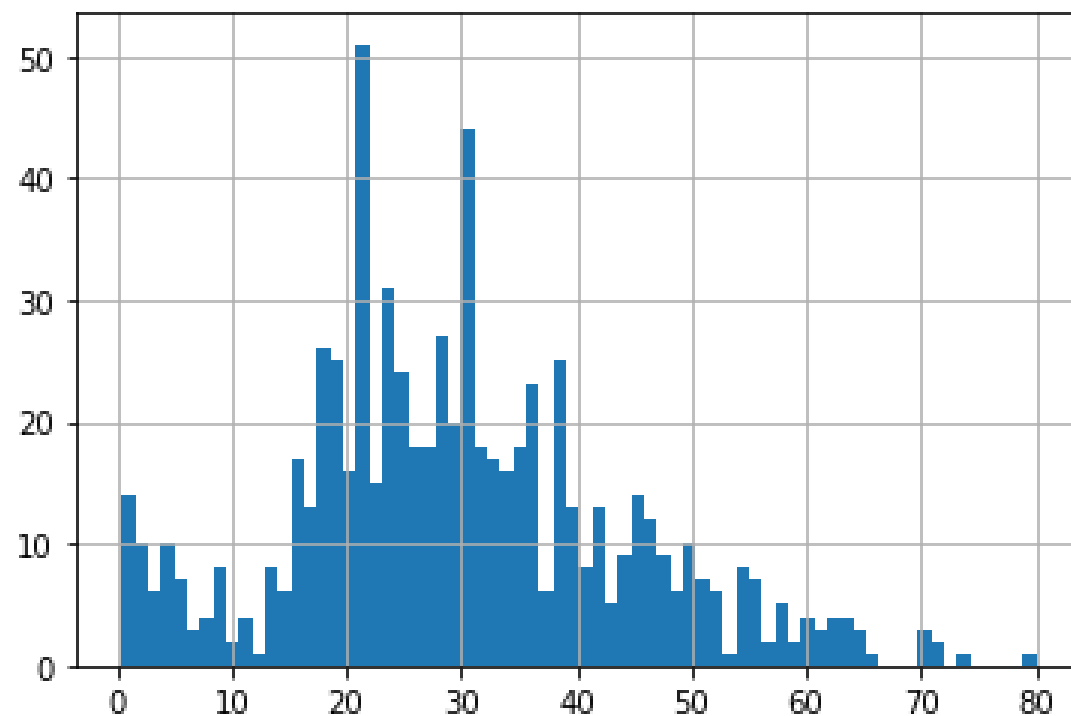
➤ `sb.factorplot('Sex', kind='count', data=titanic_df)`



➤ `sb.factorplot('Pclass', kind='count', hue='Sex', data=titanic_df)`

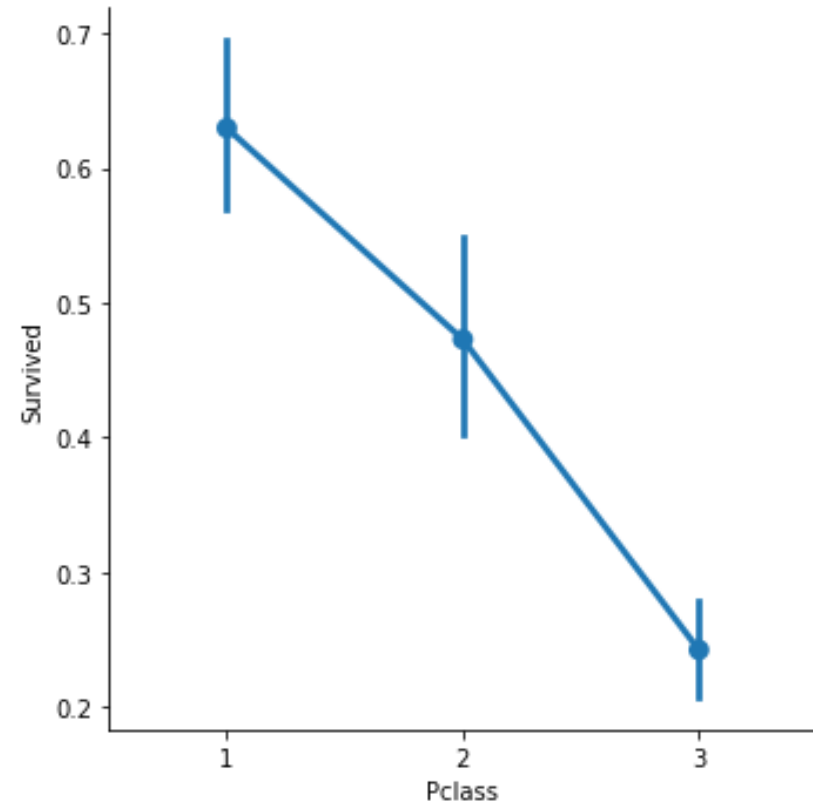
➤ `titanic_df['Age'].hist(bins=70)`

#hist(): 히스토그램 차트



```
0    549
1    342
Name: Survived, dtype: int64
```

➤ `titanic_df['Survived'].value_counts()`



➤ `sb.factorplot('Pclass', 'Survived', data=titanic_df, order=[1,2,3])`