Seaborn으로 시각화하기
– relplot, histplot, countplot

Soongsil University

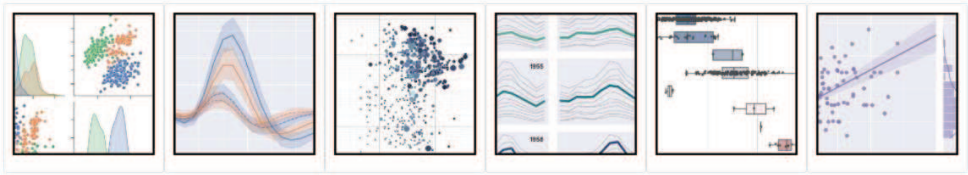
베어드교양대학
강의선
백마관 203호
02-828-7264
iami86@ssu.ac.kr

seaborn 라이브러리

https://seaborn.pydata.org/

- matplotlib 위에서 동작되는 시각화 라이브러리

```
import seaborn as sns
import matplotlib.pyplot as plt
```



Soongsil University

Seaborn에서 제공하는 titanic 데이타 읽어오기

```
1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4
5 titanic = sns.load_dataset('titanic')
6 titanic
```

그외에
Iris, tips, fmri, flights 등 제공

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
886	0	2	male	27.0	0	0	13.0000	S	Second	man	True	NaN	Southampton	no
887	1	1	female	19.0	0	0	30.0000	S	First	woman	False	B	Southampton	yes
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False	NaN	Southampton	no
889	1	1	male	26.0	0	0	30.0000	C	First	man	True	C	Cherbourg	yes
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True	NaN	Queenstown	no

891 rows × 15 columns

Soongsil University

titanic 데이타 정보 확인하기

```
1 titanic.info()
```

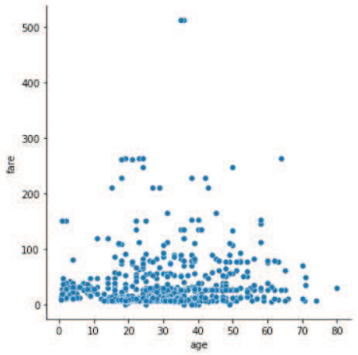
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 15 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   survived     891 non-null    int64
 1   pclass       891 non-null    int64
 2   sex          891 non-null    object
 3   age          714 non-null    float64
 4   sibsp        891 non-null    int64
 5   parch        891 non-null    int64
 6   fare         891 non-null    float64
 7   embarked     889 non-null    object
 8   class        891 non-null    category
 9   who          891 non-null    object
 10  adult_male   891 non-null    bool
 11  deck         203 non-null    category
 12  embark_town  889 non-null    object
 13  alive        891 non-null    object
 14  alone        891 non-null    bool
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.7+ KB
```


Soongsil University

relplot() 함수 : 기본형

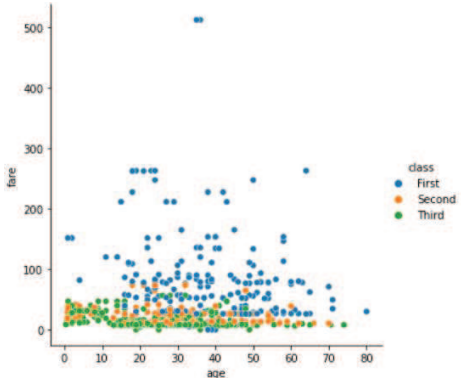
- 두 수치형 칼럼 간의 관계를 파악할 때 사용
- 기본적으로 스케터 차트 형태로 표현

```
sns.relplot(data=titanic, x='age', y='fare');
```



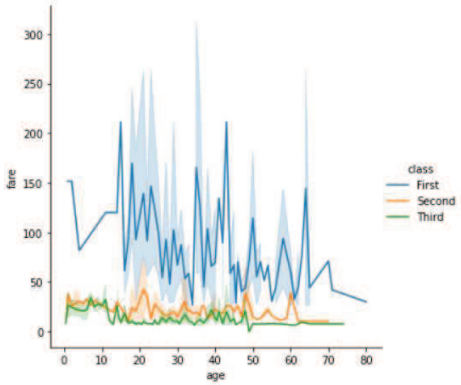
relplot() 함수 : hue 추가

```
sns.relplot(data=titanic, x='age', y='fare', hue='class');
```



relplot() 함수 : 라인 차트 변경

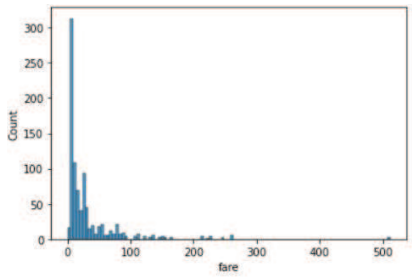
```
sns.relplot(data=titanic, x='age', y='fare', hue='class', kind='line');
```



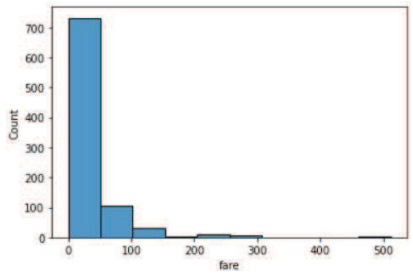
histplot() 함수

- 수치형 칼럼의 빈도수를 확인하고자 할 때 사용

```
sns.histplot(titanic.fare);
```



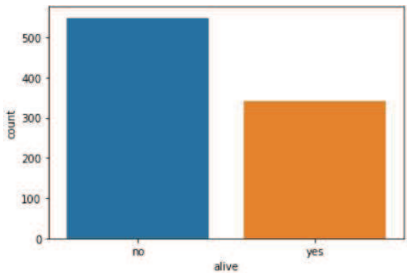

```
sns.histplot(titanic.fare, bins=10);
```



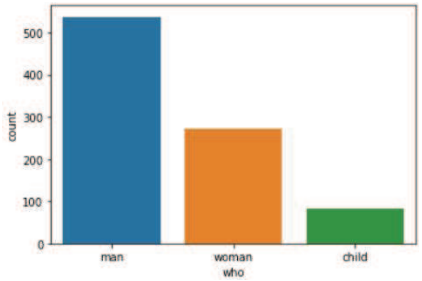
countplot () 함수

- 범주형 칼럼의 빈도수를 확인하고자 할 때 사용

```
sns.countplot(x = titanic.alive)
plt.show()
```




```
sns.countplot(x=titanic.who)
plt.show()
```



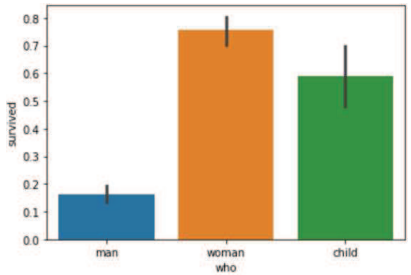
Soongsil University

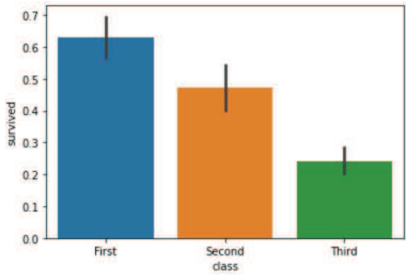
barplot() 함수

- 데이터의 분포를 백분율로 확인 가능

```
sns.barplot(data=titanic, x='who', y='survived');
```

```
sns.barplot(data=titanic, x='class', y='survived');
```





Soongsil University

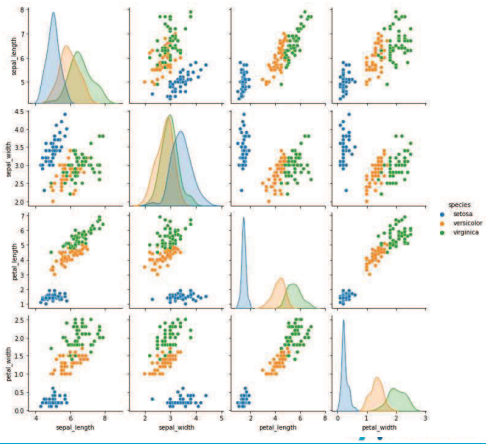
pairplot() 함수

- 모든 변수간의 관계를 일괄적으로 확인하고자 할 경우에 사용

```
iris = sns.load_dataset('iris')
sns.pairplot(iris, hue='species');
```

	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
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns



sity

수고하셨습니다.

Soongsil University

12