

## ▼ (4-1) Seaborn 소개

Seaborn : Matplotlib 기반 통계 시각화 라이브러리

- 통계 정보 : 구성, 분포, 관계 등
- Matplotlib 기반 -> Matplotlib으로 커스텀 가능
- 디자인 깔끔하고 문법이 쉬워 사용하기 쉬움
- 2021년 8월의 가장 최신 버전인 0.11로 실습
- 시각화의 목적과 방법에 따라 API를 제공
  - Categorical API
  - Distribution API
  - Relational API
  - Regression API
  - Matrix API
  - Multiples API : Facet과 관련
  - Theme API : Color과 관련

실습하면서 배운 것 --

```
pip install seaborn==0.11
```

```
Collecting seaborn==0.11
  Downloading seaborn-0.11.0-py3-none-any.whl (283 kB)
    |████████████████████████████████████████| 283 kB 9.1 MB/s
Requirement already satisfied: pandas>=0.23 in /usr/local/lib/python3.7/dist-packages (from s
Requirement already satisfied: matplotlib>=2.2 in /usr/local/lib/python3.7/dist-packages (frc
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.7/dist-packages (from se
Requirement already satisfied: scipy>=1.0 in /usr/local/lib/python3.7/dist-packages (from sea
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages (from r
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: pyparsing!=2.0.4,!2.1.2,!2.1.6,>=2.0.1 in /usr/local/lib/pyt
Requirement already satisfied: six in /usr/local/lib/python3.7/dist-packages (from cycler>=0.
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (from p
Installing collected packages: seaborn
  Attempting uninstall: seaborn
    Found existing installation: seaborn 0.11.1
    Uninstalling seaborn-0.11.1:
      Successfully uninstalled seaborn-0.11.1
  Successfully installed seaborn-0.11.0
```

```
import seaborn as sns
```

## ▼ (4-2) Seaborn 기초

## 통계 시각화 API

- Categorical API : 데이터의 기본 통계량
- Distribution API : 범주형/연속형 자료의 분포 시각화
- Relational API : 관계성 파악
- Regression API : 회귀 분석
- Matrix API : 히트맵

시각화를 위해 차원을 먼저 파악하자

## Categorical API

describe의 count : Null 값이 아닌 개수 반환

countplot

- hue\_order : 색 순서 지정

violin Plot : 분포를 곡선으로 표현, 연속적인 데이터 표현 때문에 데이터 손실과 오차가 존재한다.

boxenplot : 자주 사용하지 않음

swarmplot : 바이올린 그래프와 같이 사용하면 좋음

stripplot : 점의 밀도 확인 용이

## Distribution API

왜곡된 정보를 전달할 수 있어 multiple='fill' 사용 지양

histplot : 빈도 그래프

- 상관관계를 파악할 때 산점도보다 가독성이 높다.

kdeplot : histogram의 연속적 표현

rugplot : 데이터 갭 클 때, 한정된 공간 내에서 분포 표현할 때 사용

## Regression API

scatter Plot : 산점도

line Plot : 선 그래프

Regplot : 산점도에 추세선 추가한 그래프

- x\_estimator=np.mean : x에 따른 y 분포 확인 가능
- order=2 : 2차 곡선 사용 가능
- logx=True : 로그 곡선 사용 가능

## Matrix API

heatmap

- vmin=-1, vmax=1, center=0 : 도메인 지식을 활용해 범위 설정
- cmap='coolwarm' : diverse color map 이용
- annot=True, fmt='.2f' : 상관관계 수치 소수점 이하 2자리까지 표시, 정수형은 d 사용
- linewidth=0.1 : 라인 두께 추가해 가독성 증가
- square=True : 상관관계 네모를 정사각형으로 만들어줌
- mask=mask : 마스킹
- np.triu\_indices\_from : 상 삼각행렬 반환

## ✓ ▶ Seaborn 기초 실습 코드

[ ] ↳ 숨겨진 셀 122개

## ▼ (4-3) Seaborn 심화

시각화를 위해 차원을 먼저 파악하자

### jointplot

- kind의 값을 변경해 분포를 확인 가능

### pairplot

- corner=True : 데이터 많을 때 사용

### FacetGrid

- 정의 : 다중 패널을 사용하는 시각화
- 특징 : feature-feature와 feature's category-feature's category의 관계도 확인 가능
- 예 : catplot, displot, displot, lmpplot
- row/col로 다중 패널 사용, row\_order/col\_order로 순서지정 가능

## ✓ ▶ Seaborn 심화 실습 코드

▶ ↳ 숨겨진 셀 37개