



# Round 8

**PRESS  
START**



## 《 Round 8 》

- 과제 피드백
- 데이터 분석용 기초통계



New  
Assignment



## 《 Round 8 》

- 과제 피드백 《
- 데이터 분석용 기초통계



Let's  
Go



# Tips Dataset

```
tips = sns.load_dataset("tips")  
tips
```

	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
...	...	...	...	...	...	...	...
239	29.03	5.92	Male	No	Sat	Dinner	3
240	27.18	2.00	Female	Yes	Sat	Dinner	2
241	22.67	2.00	Male	Yes	Sat	Dinner	2
242	17.82	1.75	Male	No	Sat	Dinner	2
243	18.78	3.00	Female	No	Thur	Dinner	2

244 rows × 7 columns



```
# 1. 데이터정보와 요약통계 확인
# 1.1 데이터 형태 짚어보기 (Head, Tail)
tips = sns.load_dataset("tips")
tips
```

```
# 1.2 행과 열의 갯수
tips.shape
```

## Shape

## Head & Tail



# 1. Tips의 데이터 정보와 요약 통계 확인

- 필수적으로 확인해야 하는 내용: 타입, NaN, 요약통계 -

# 1.3 데이터 타입 및 Null

tips.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 244 entries, 0 to 243
Data columns (total 7 columns):
#   Column      Non-Null Count  Dtype
---  -
0   total_bill  244 non-null    float64
1   tip         244 non-null    float64
2   sex        244 non-null    category
3   smoker     244 non-null    category
4   day        244 non-null    category
5   time       244 non-null    category
6   size       244 non-null    int64
dtypes: category(4), float64(2), int64(1)
memory usage: 7.3 KB
```

Info

# 1.4 데이터 요약 통계

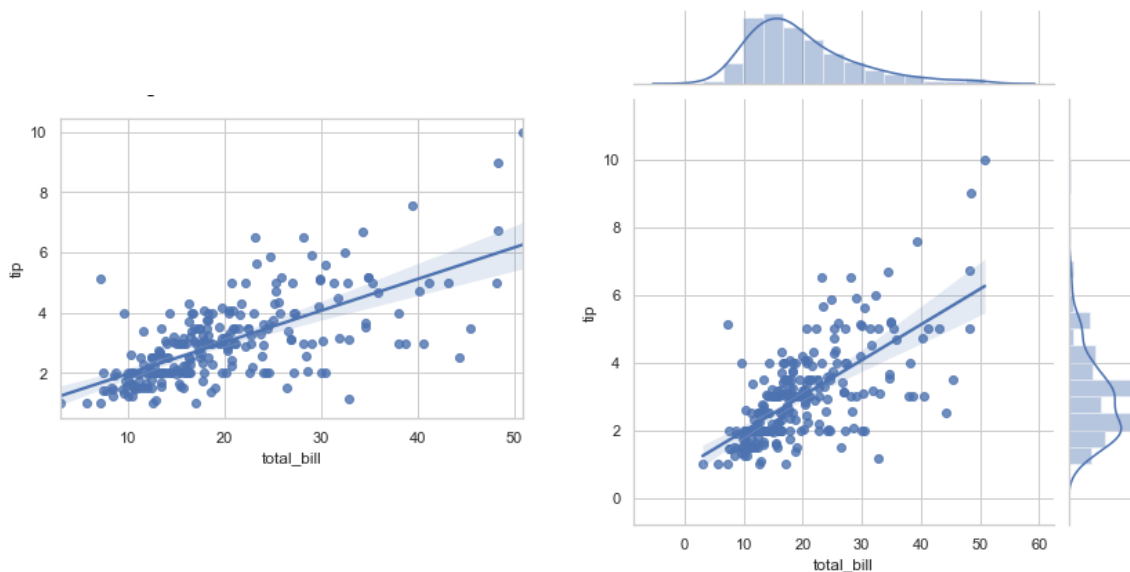
tips.describe(include='all')

	total_bill	tip	sex	smoker	day	time	size
count	244.000000	244.000000	244	244	244	244	244.000000
unique	NaN	NaN	2	2	4	2	NaN
top	NaN	NaN	Male	No	Sat	Dinner	NaN
freq	NaN	NaN	157	151	87	176	NaN
mean	19.785943	2.998279	NaN	NaN	NaN	NaN	2.569672
std	8.902412	1.383638	NaN	NaN	NaN	NaN	0.951100
min	3.070000	1.000000	NaN	NaN	NaN	NaN	1.000000
25%	13.347500	2.000000	NaN	NaN	NaN	NaN	2.000000
50%	17.795000	2.900000	NaN	NaN	NaN	NaN	2.000000
75%	24.127500	3.562500	NaN	NaN	NaN	NaN	3.000000
max	50.810000	10.000000	NaN	NaN	NaN	NaN	6.000000

Describe

## 2. 계산금액과 팁 사이의 상관관계?

```
# 2.
sns.regplot(x="total_bill", y="tip", data=tips)
sns.jointplot(x="total_bill", y="tip", kind='reg', data=tips)
```



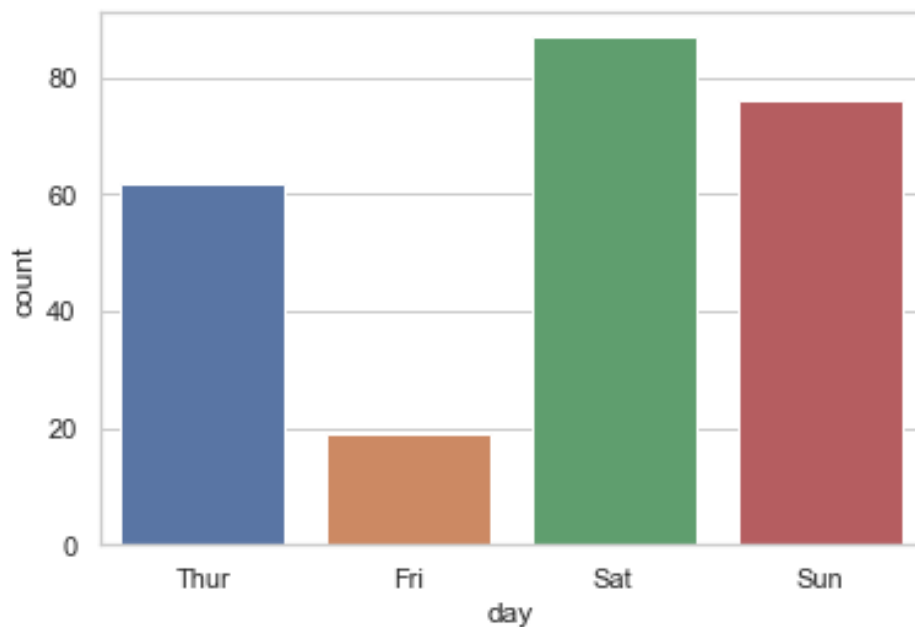
```
print(tips.corr())
```

	total_bill	tip	size
total_bill	1.000000	0.675734	0.598315
tip	<b>0.675734</b>	1.000000	0.489299
size	0.598315	0.489299	1.000000

### 3. 요일별 팁 발생횟수?

```
# 3. 요일별 팁 발생횟수  
sns.countplot(x='day', data=tips)
```

<AxesSubplot:xlabel='day', ylabel='count'>



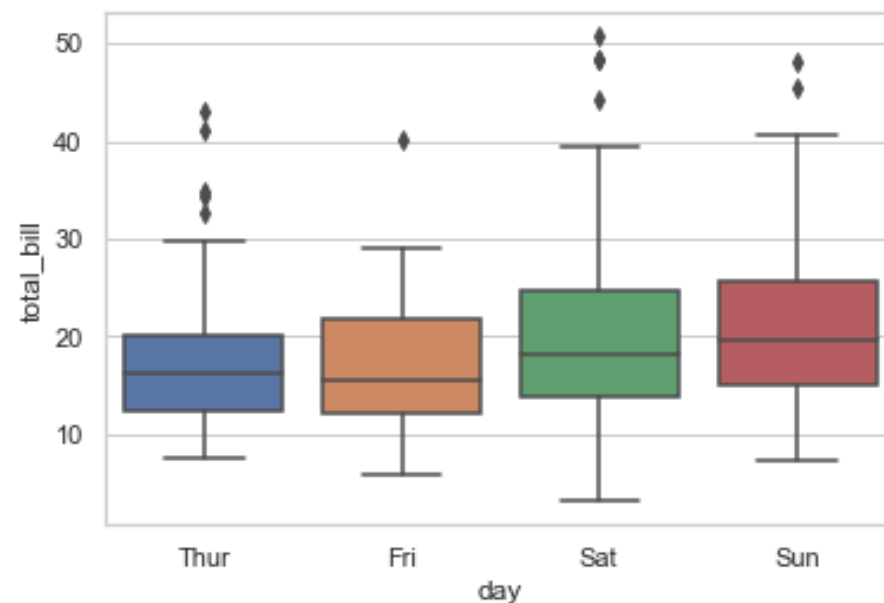


## 4. 요일별 계산금액의 통계적 분포?

# 4. 요일 별 계산금액의 통계적 분포

```
sns.boxplot(x="day", y="total_bill", data=tips)
```

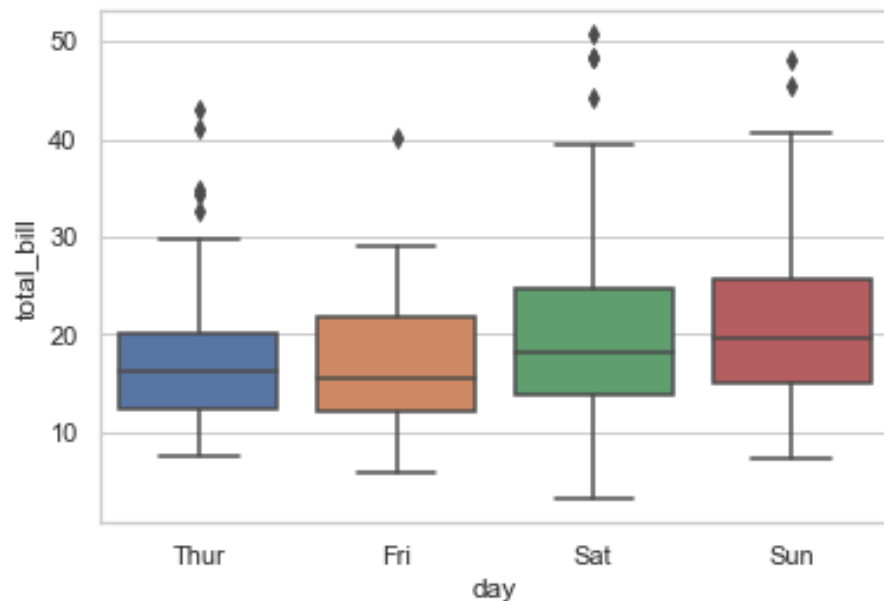
<AxesSubplot:xlabel='day', ylabel='total\_bill'>



## 4. 요일별 계산금액의 통계적 분포?

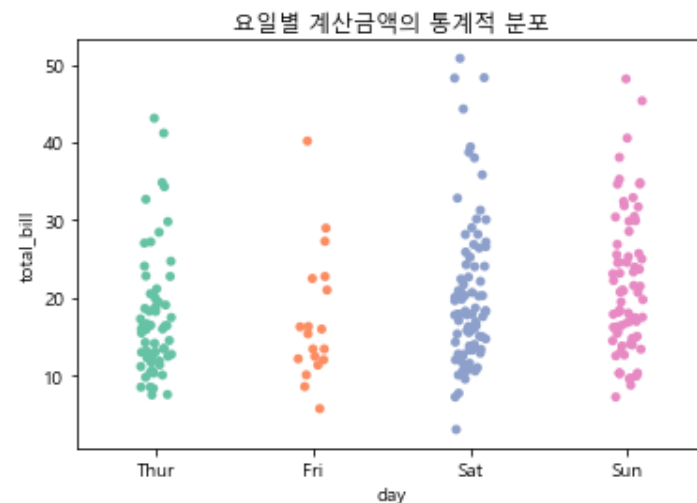
```
# 4. 요일 별 계산금액의 통계적 분포
sns.boxplot(x="day", y="total_bill", data=tips)
```

<AxesSubplot:xlabel='day', ylabel='total\_bill'>



VS

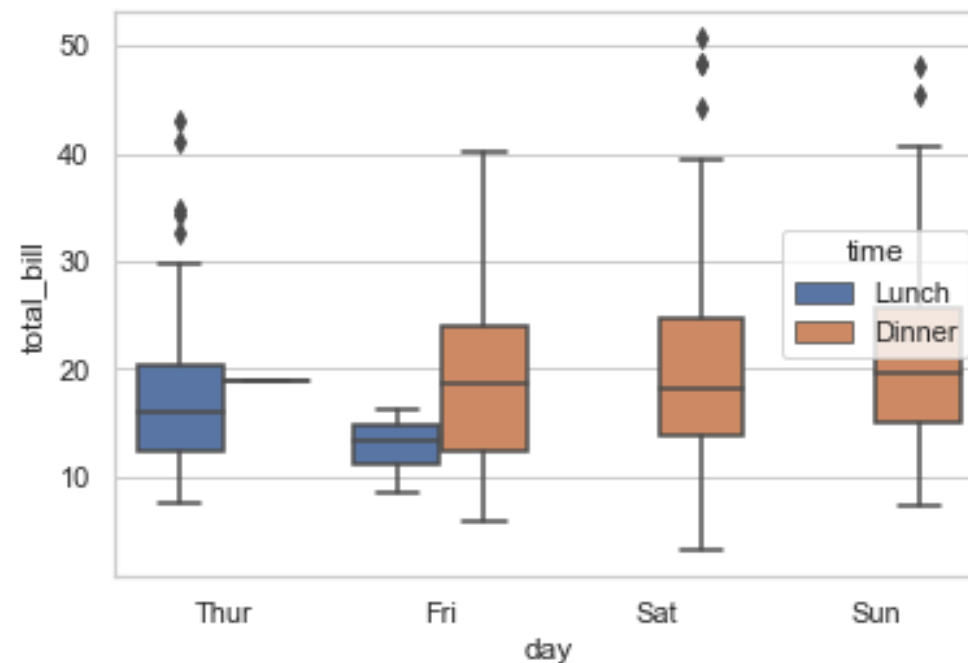
```
sns.stripplot(x='day', y='total_bill', data=tips_dataframe, palette='Set2')
plt.title("요일별 계산금액의 통계적 분포")
plt.show()
```



## 4+1. 요일+시간대 별 계산금액의 통계적 분포?

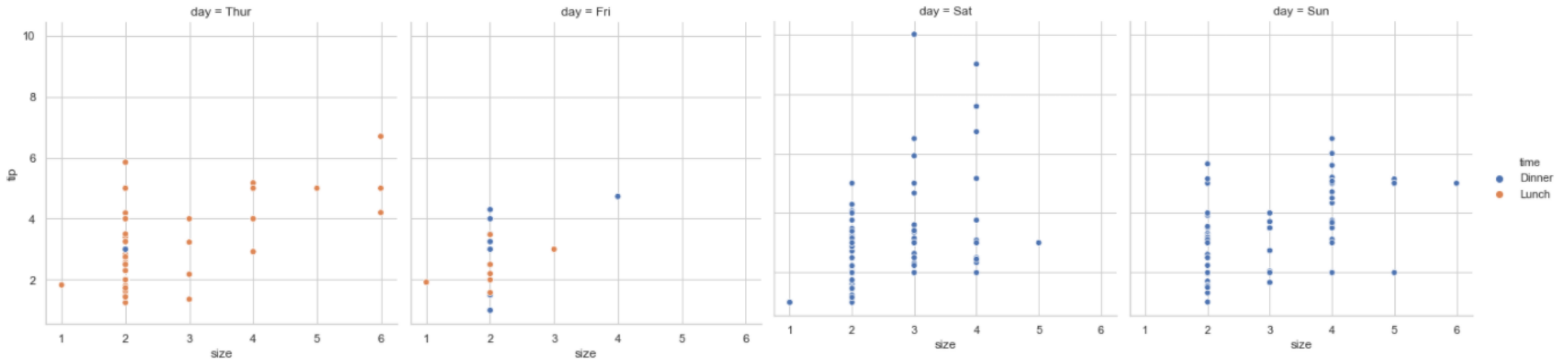
```
# 4+1. 요일+시간대별 계산금액의 통계적 분포
sns.boxplot(x="day", y="total_bill", hue="time", data=tips)
```

<AxesSubplot:xlabel='day', ylabel='total\_bill'>



## 5. 요일+시간대를 고려한 사이즈와 팁의 상관관계?

```
# 5. 요일과 시간대를 고려하여 테이블 사이즈가 팁 금액과 관계가 있는지?
sns.relplot(x="size", y="tip", hue="time", col="day", data=tips)
```



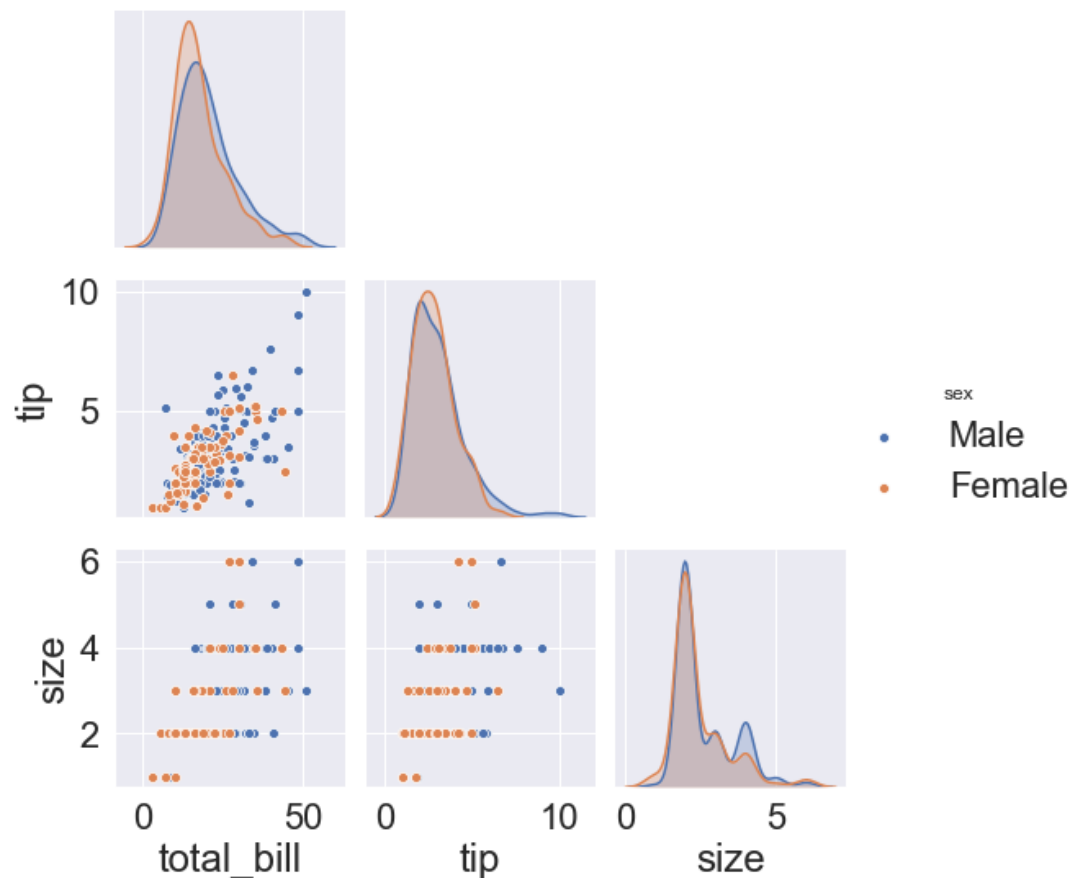


## 5. 계산금액, 팁, 사이즈 각각 상관관계를 가지는지 성별에 따라 구분하여 나타내라

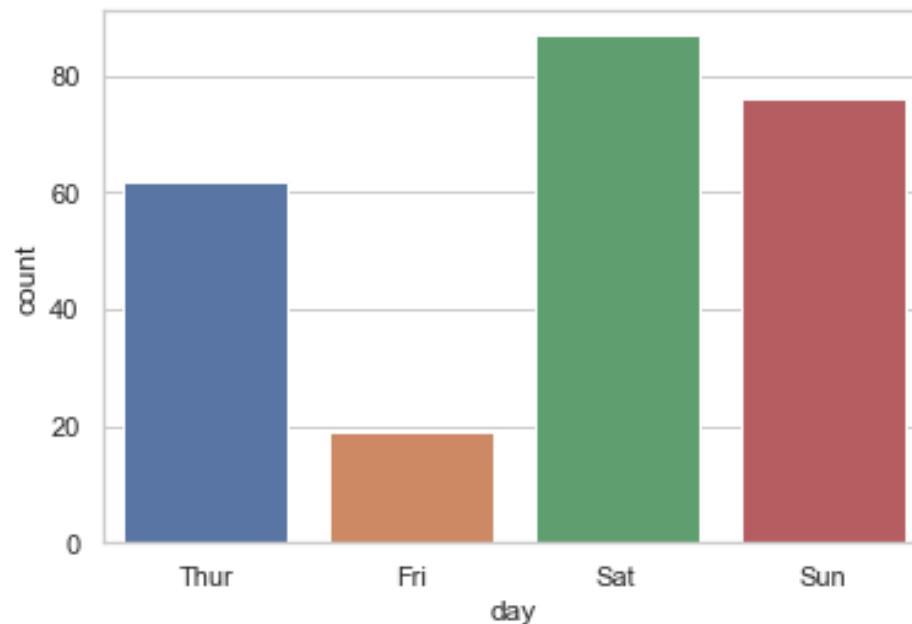
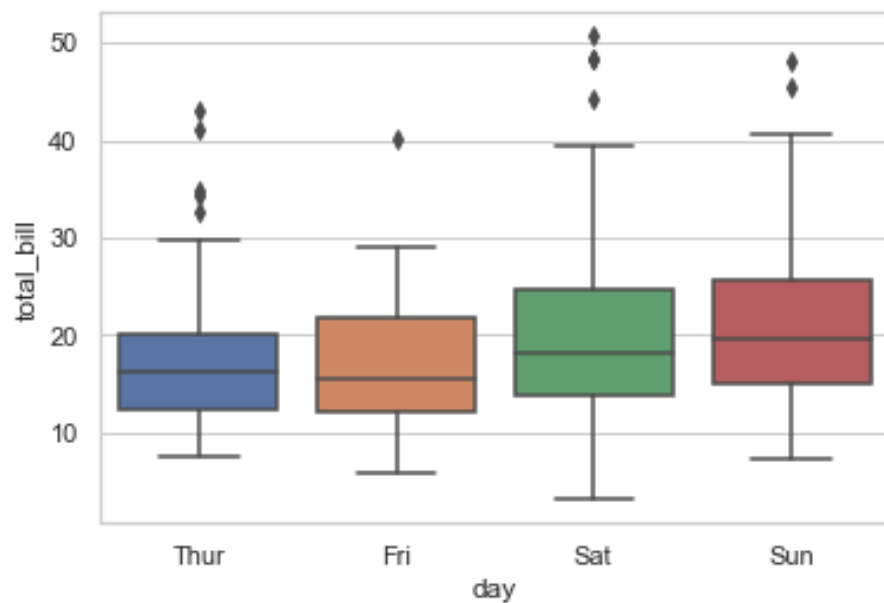
## 5. 계산금액, 팁, 사이즈 각각 상관관계를 가지는지 성별에 따라 구분하여 나타내라

```
[49]: tips_d = tips[["total_bill", "tip", "sex", "size"]]  
      sns.pairplot(data = tips_d, hue = "sex", corner = True)
```

[49]: <seaborn.axisgrid.PairGrid at 0x1d844d41048>



## 6. 추가 EDA 및 결론 도출

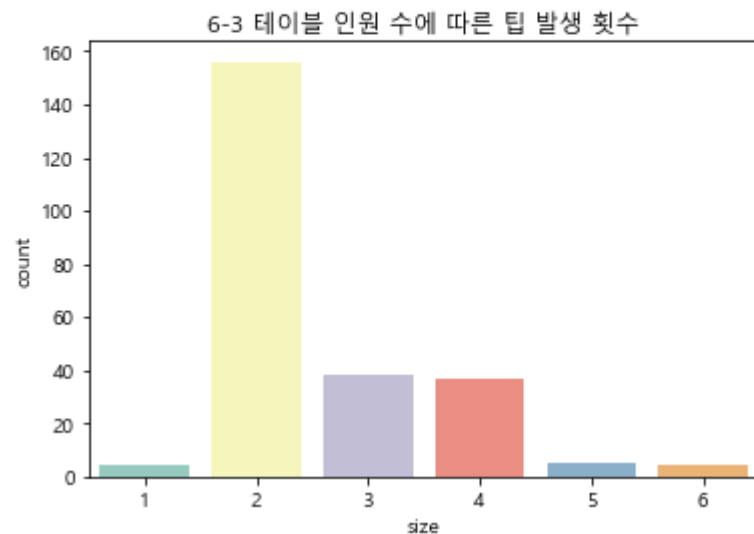


음식점의 매출액 혹은 팁 금액을 증대시키기 위한 방안

- 금요일에 매출이 적은 이유를 파악하고 이를 해결해야 한다. 같은 평일임에도 불구하고 목요일은 매출이 높고 금요일은 매출이 적다. 매주 금요일 한정 메뉴를 개발하거나 개발비용이 부담된다면 평일에 사람이 모일 수 있도록 조성해야 한다. 예를 들어 아울렛 내 매장인 경우 평일만 직원할인이 가능하고 주말은 직원할인이 불가능하다. 이렇게 평일 혜택을 두어 금요일도 매출이 오를 수 있도록 해야 한다.

## 6. 추가 EDA 및 결론 도출

```
sns.countplot(x='size', data=tips_dataframe, palette='Set3')
plt.title("6-3 테이블 인원 수에 따른 팁 발생 횟수")
plt.show()
```

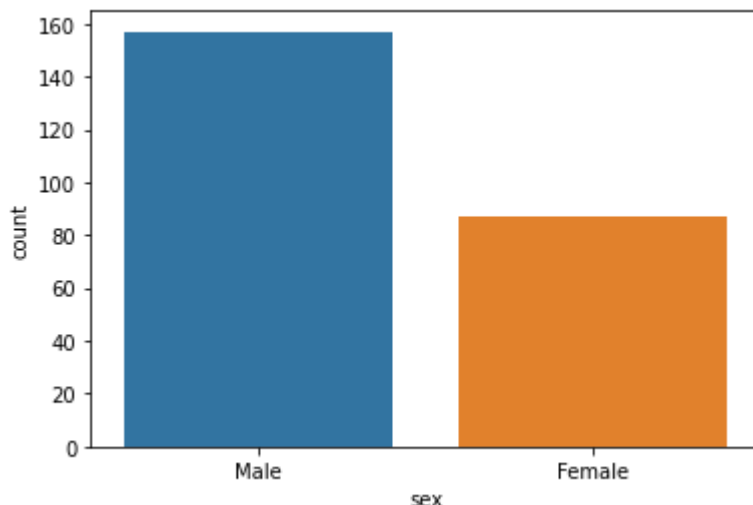


### 음식점의 매출액 혹은 팁 금액을 증대시키기 위한 방안

- 2인 팀이 주로 방문하는 식당인 것을 보아 2인석 테이블을 늘리면 좋을 것 같다. 웨이팅이 있을 경우 기다리지 않고 다른 식당으로 빠질 수도 있원 매출을 확인하여 다인석을 쪼개 2인석을 만드는 것이 효율적일 수도 있다. 1인 팀을 위한 1인석 구비는 조금 줄여도 되지만 1인 손님을 위한 혼밥 메뉴가 있으면 좋을 것 같다. 보통 2인 손님이 많다는 것은 가족단위로 오기 힘든 식당이라 판단되며, 그만큼 가격이 높을 것으로 예상된다. 여러 할인혜택을 준비하여 단체 손님이 부담 갖지 않고 방문할 수 있게 해야 한다. 2인 팀보다 3명 이상의 팀이 많은 팁과 계산 금액을 지불하기 때문에 2인 팀에 집중할 경우 오히려 매출이 저조할 수도 있다.



## 6. 생각해 볼 문제: 여성 고객의 유도가 필요?



음식점의 매출액 혹은 팁 금액을 증대시키기 위한 방안

### 1. 잠재고객층(여성) 유도

- 주로 남자에게서 tip발생 빈도가 높고, 남자고객이 많은 식당이다.  
이 식당에서 tip과 매출액을 증대시키려면 잠재고객인 여성고객층에 대한 공략이 필요해 보인다.  
이를 위하여 instagram을 통한 홍보나 감성적인 요소를 더한다.  
또한 menu에서 가격을 낮춘 small size와 아기자기한 menu를 추가하여 여성 고객층을 유도한다.

## 《 Round 8 》

- 과제 피드백
- 데이터 분석용 기초통계 《



Let's  
Go



# 변수의 종류?

## 범주형 데이터 vs 수치형 데이터

명목형 데이터 vs **순서형** 데이터

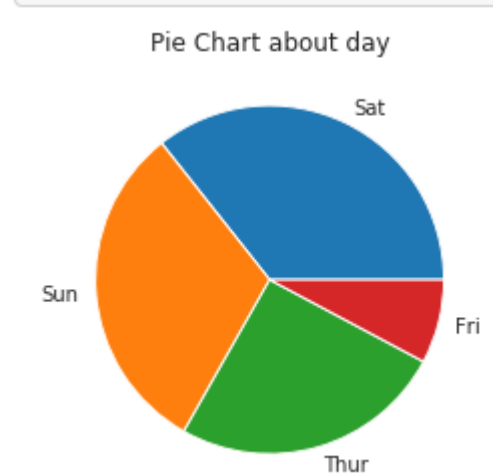
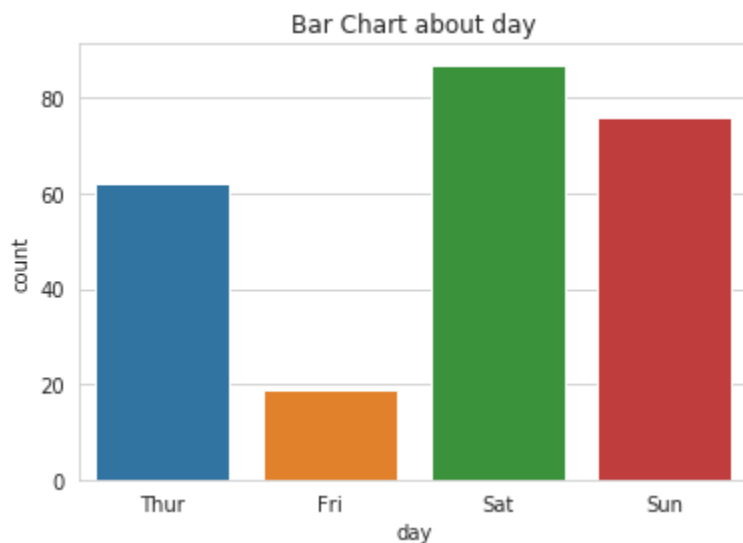
**이산형** 데이터 vs 연속형 데이터

학번	키	나이	남녀구분	신체검사만족도
1	169.1	25	남	상
2	184.2	22	남	중
3	172.4	26	여	중
4	159.0	24	여	하
5	178.9	21	남	상

# 변수의 종류에 따라 Plot을 그릴 때?

## 일변량 데이터

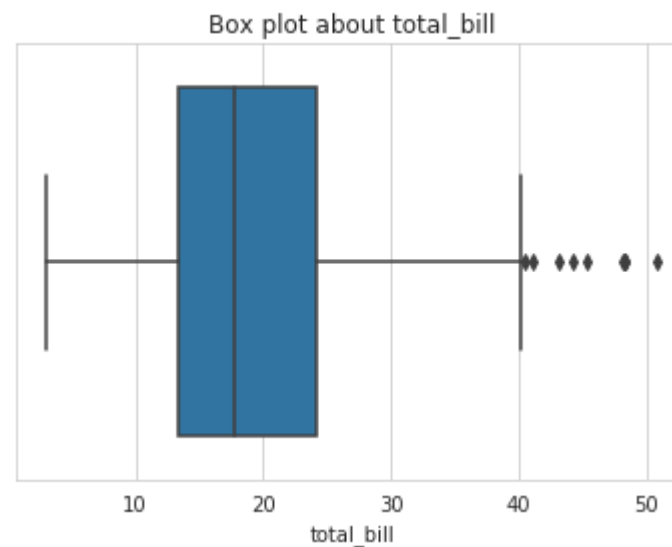
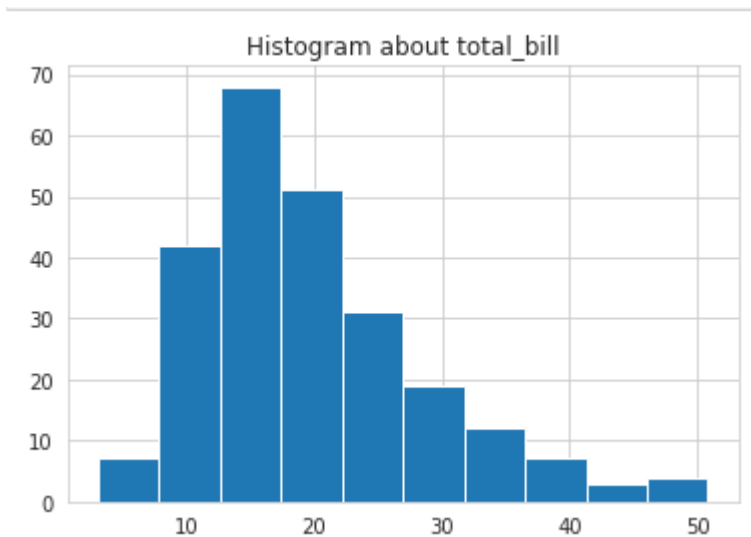
범주형 데이터(Tips-day)



# 변수의 종류에 따라 Plot을 그릴 때?

## 일변량 데이터

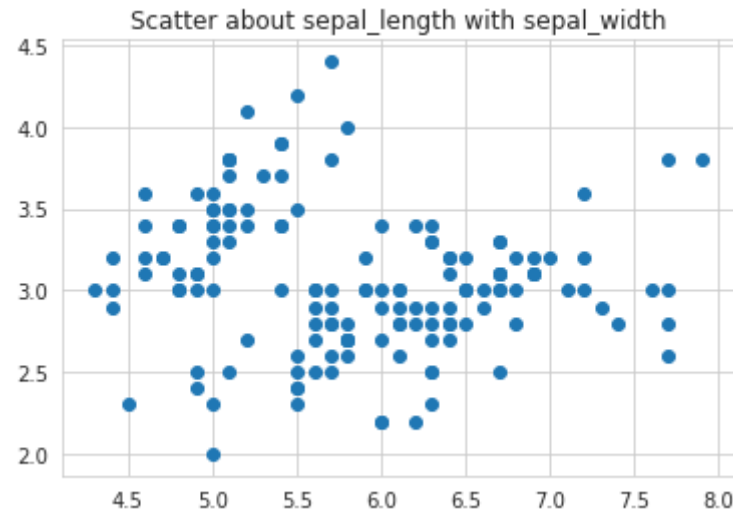
연속형 데이터(Tips-total\_bill)



# 변수의 종류에 따라 Plot을 그릴 때?

## 이변량 데이터

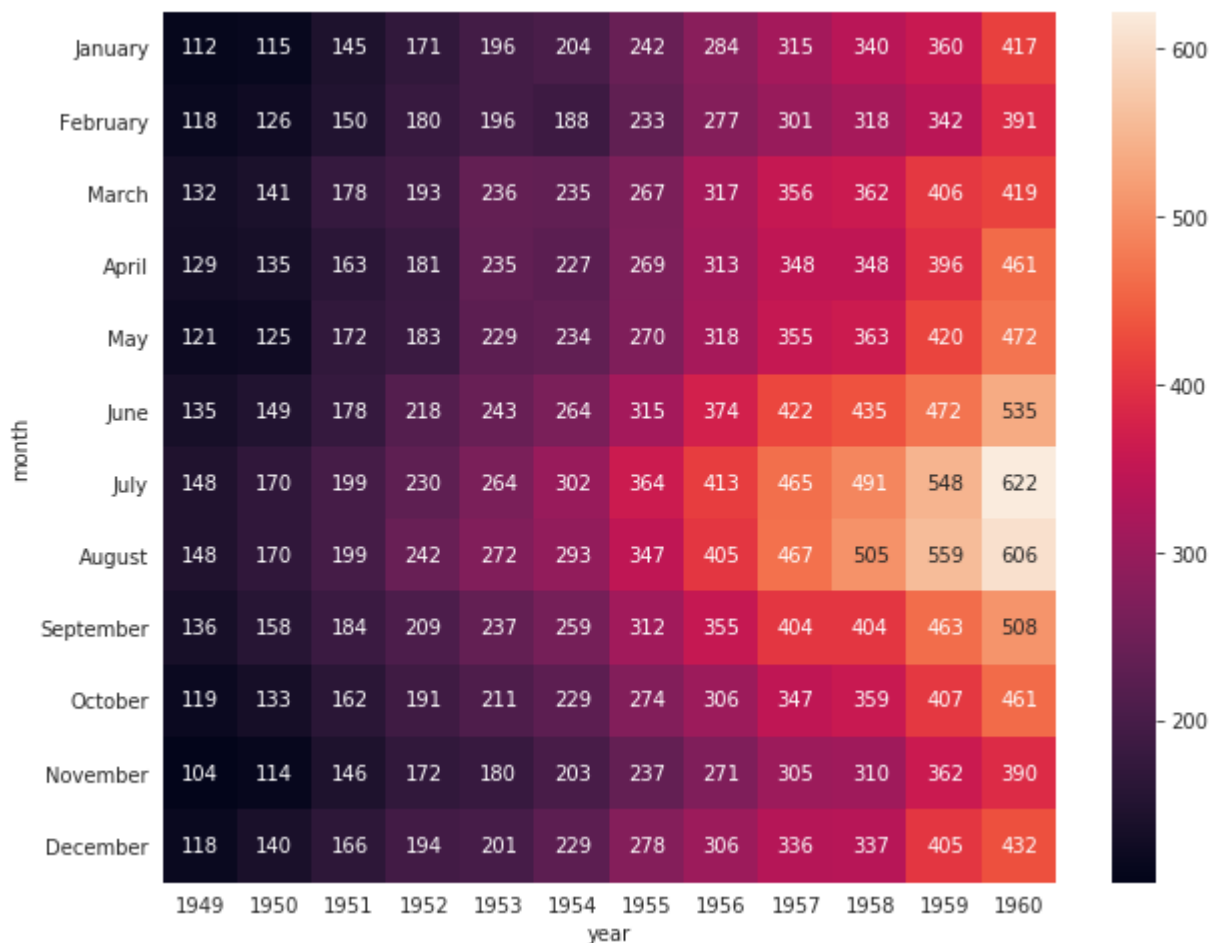
연속형 데이터(Iris-sepal\_length:sepal\_width)



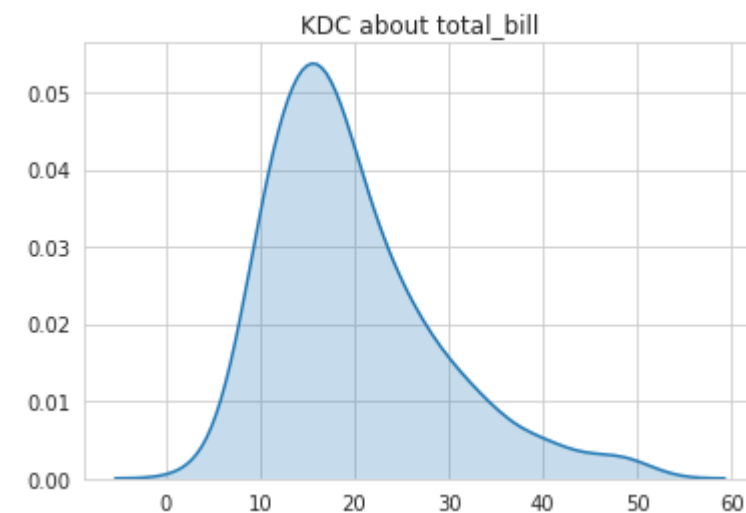
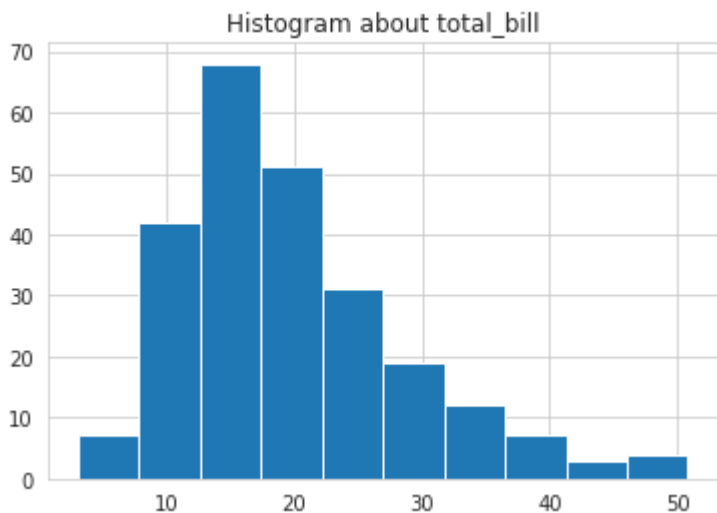
# 변수의 종류에 따라 Plot을 그릴 때?

## 이변량 데이터

범주형 데이터(Flights-year:month, passengers)



# 확률분포?





# 회귀 분석이란?

## 국어사전

---

**회귀** (回歸)

[회귀/훼귀] 

명사

한 바퀴 돌아 제자리로 돌아오거나 돌아감.

과거로의 회귀.

# NEXT STAGE

