In [52]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [53]:
```
data=pd.read_csv('tips2.csv')
data.head()
```

Out[53]:

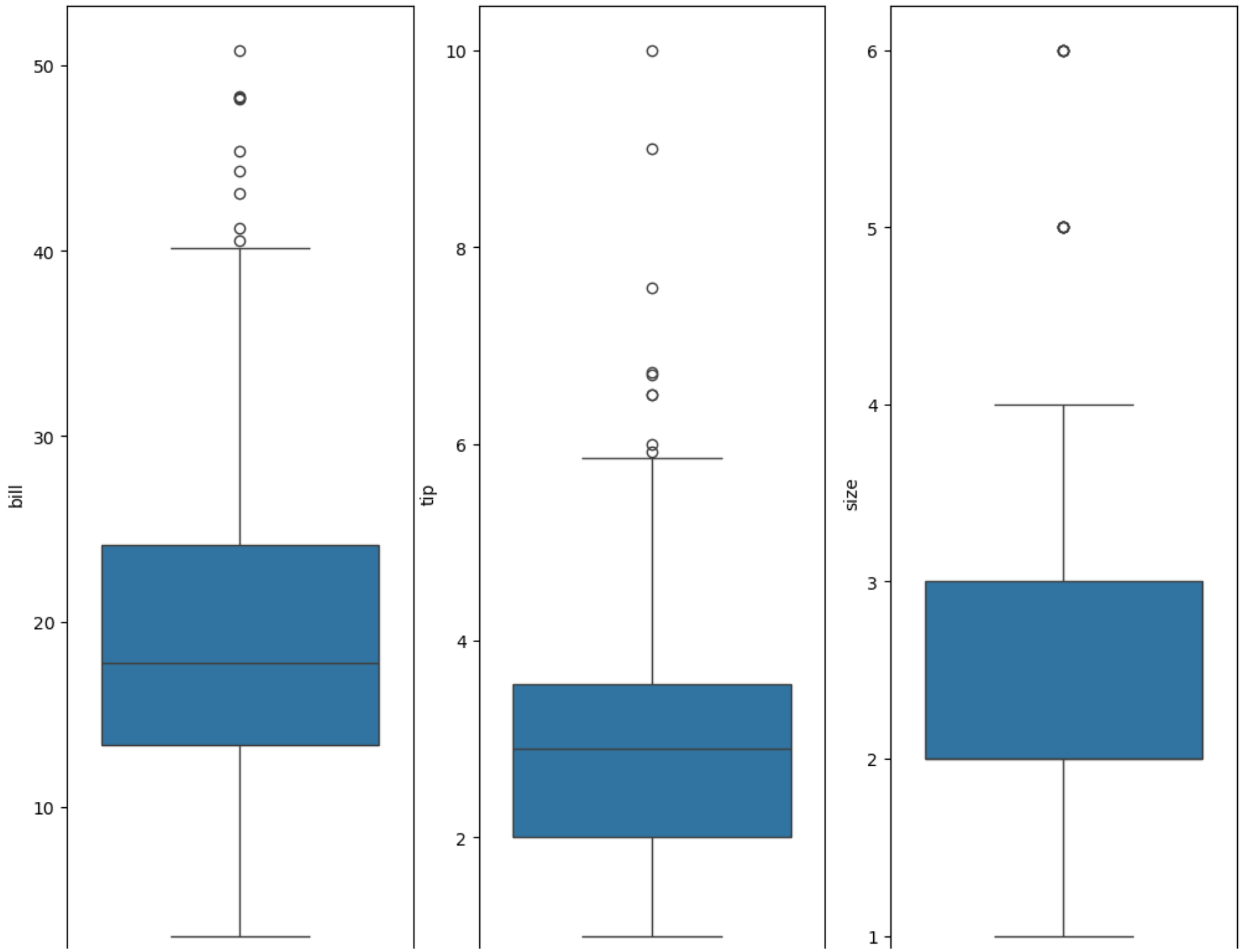
	bill	tip	gender	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

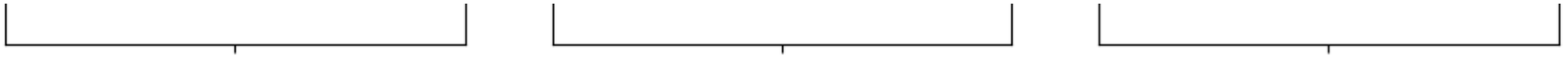
In [54]:
```
# Univariate analysis
data.describe(include="all")
```

Out[54]:

	bill	tip	gender	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

In [55]:
```
# Outlier analysis
fig,axes = plt.subplots(ncols=3, nrows=1, figsize=(10,8))
idx =0
axes=axes.flatten()
for k,v in data.items():
    if(v.dtype == "object"):
        continue
    sns.boxplot(data=data,y=k,ax=axes[idx])
    idx+=1
plt.tight_layout(pad=0.4, w_pad=0.4, h_pad=2.0)
```


bill
50
40
30
20
10
tip
10
8
6
4
2
size
6
5
4
3
2
1

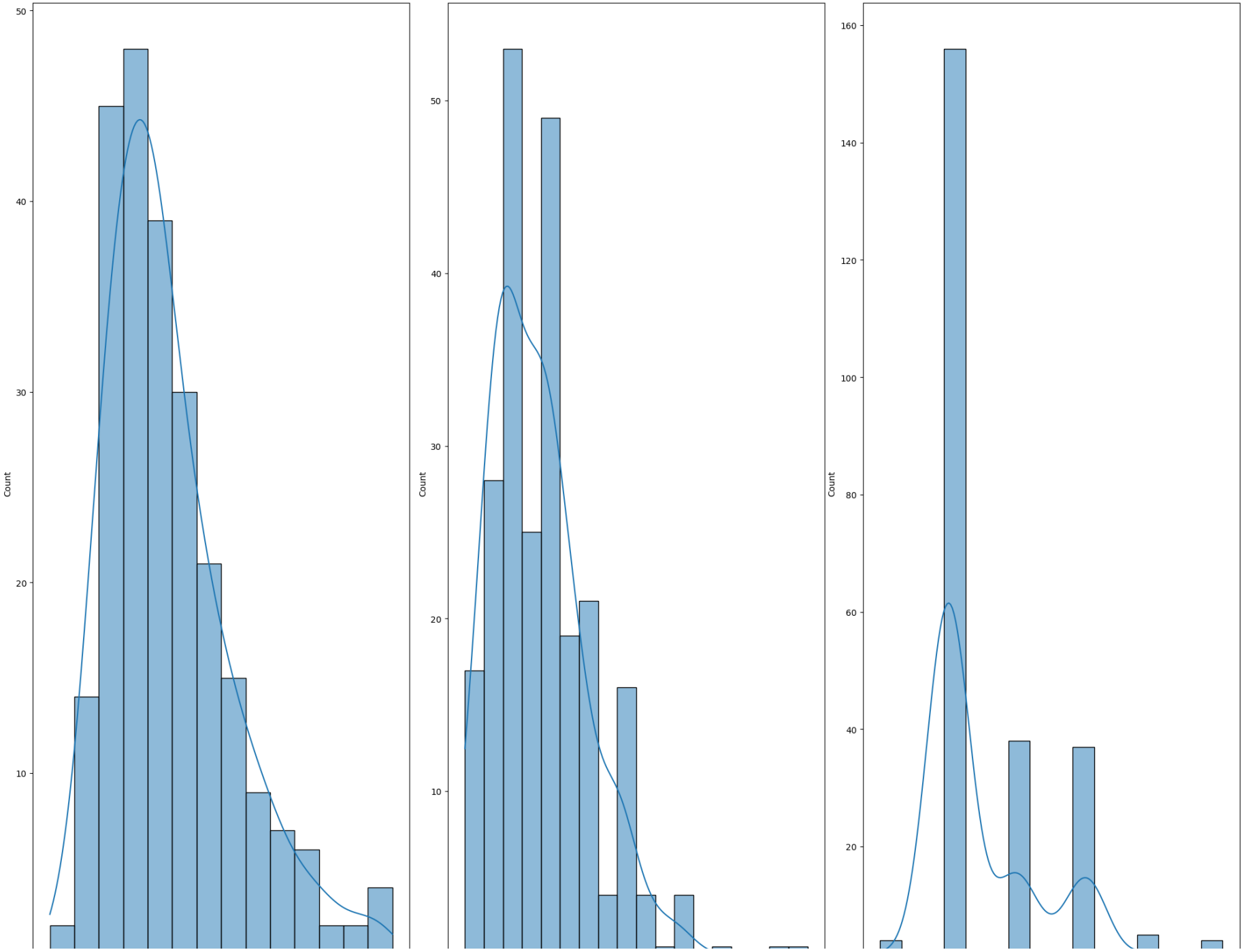


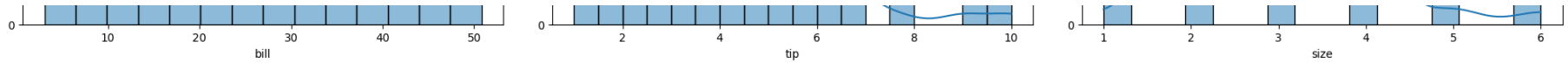
In [56]:
```
# skewness and kurtosis
fig,axes =plt.subplots(ncols=3, nrows=1, figsize=(20,16))
idx=0
axes=axes.flatten()
cols=["Column", "Skewness", "Kurtosis"]
values = []

for k, v in data.items():
    if(v.dtype =="object"):
        continue

    skew = v.skew()
    kurt = v.kurtosis()

    if int(skew)>=1:
        skew_val = "Right/Positively skewed"
    elif int(skew) <=-1:
        skew_val ="Left/Negatively skewed"
    else:
        skew_val="Normal"

    if round(kurt,0) >3:
        kurt_val = "Leptokurtic"
    elif round(kurt,0) ==3 :
        kurt_val ="mesokurtic"
    elif round(kurt,0) <3 and  round(kurt,0)!=0:
        kurt_val="platykurtic"
    else:
        kurt_val = "Normal"

    kurt_val += ":"+ str(kurt)
    skew_val += ":"+ str(skew)


    values.append([k,skew_val, kurt_val])
    sns.histplot(kde=True,data=v, ax=axes[idx])
    idx+=1
```

```
plt.tight_layout(pad=0.4, w_pad=0.4, h_pad=2.0)

df = pd.DataFrame(columns =cols, data=values)
df
```

Out[56]:

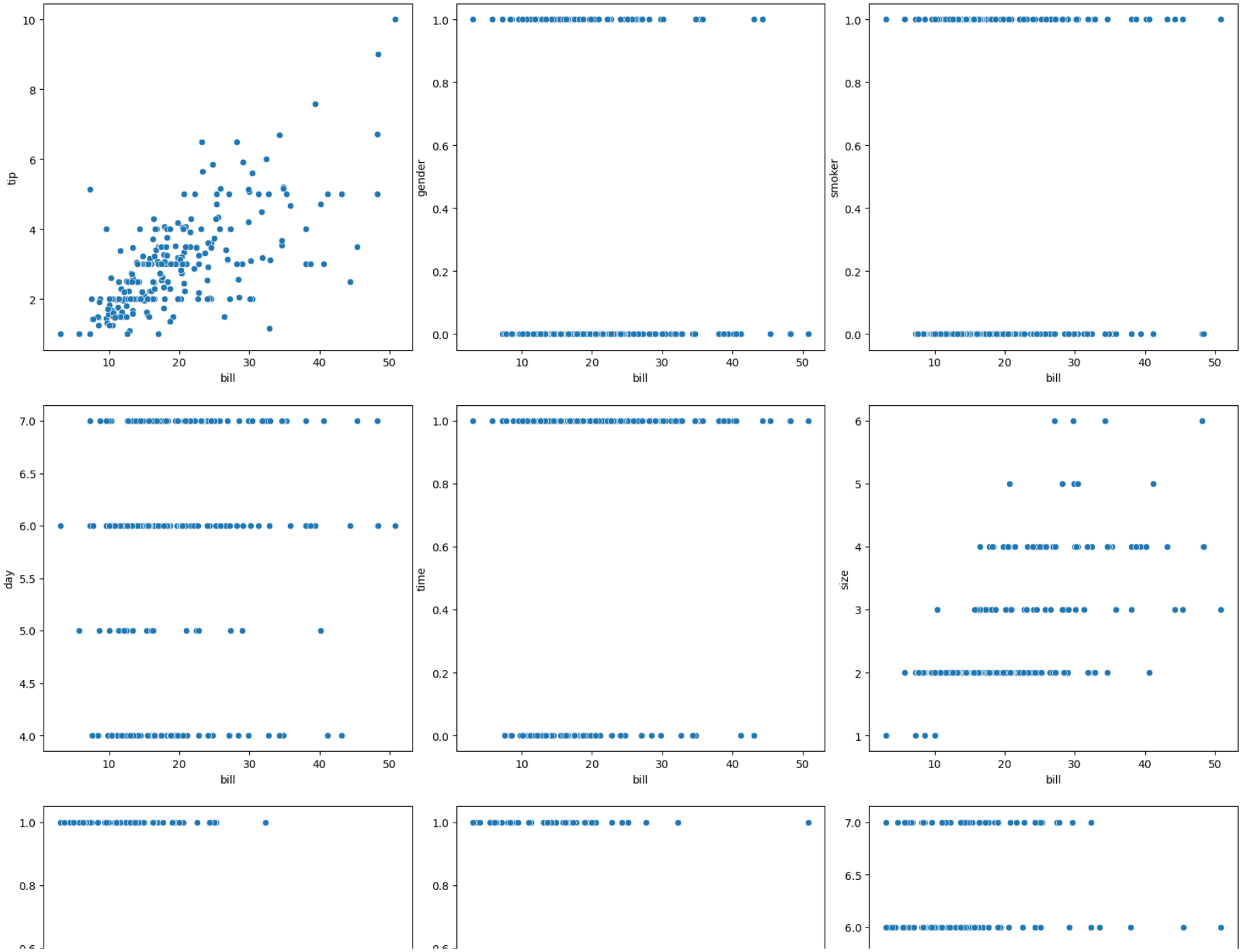
	Column	Skewness	Kurtosis
0	bill	Right/Positively skewed:1.1332130376158205	platykurtic:1.2184840156638854
1	tip	Right/Positively skewed:1.4654510370979401	Leptokurtic:3.648375873352852
2	size	Right/Positively skewed:1.4478815386834785	platykurtic:1.7317000657641097

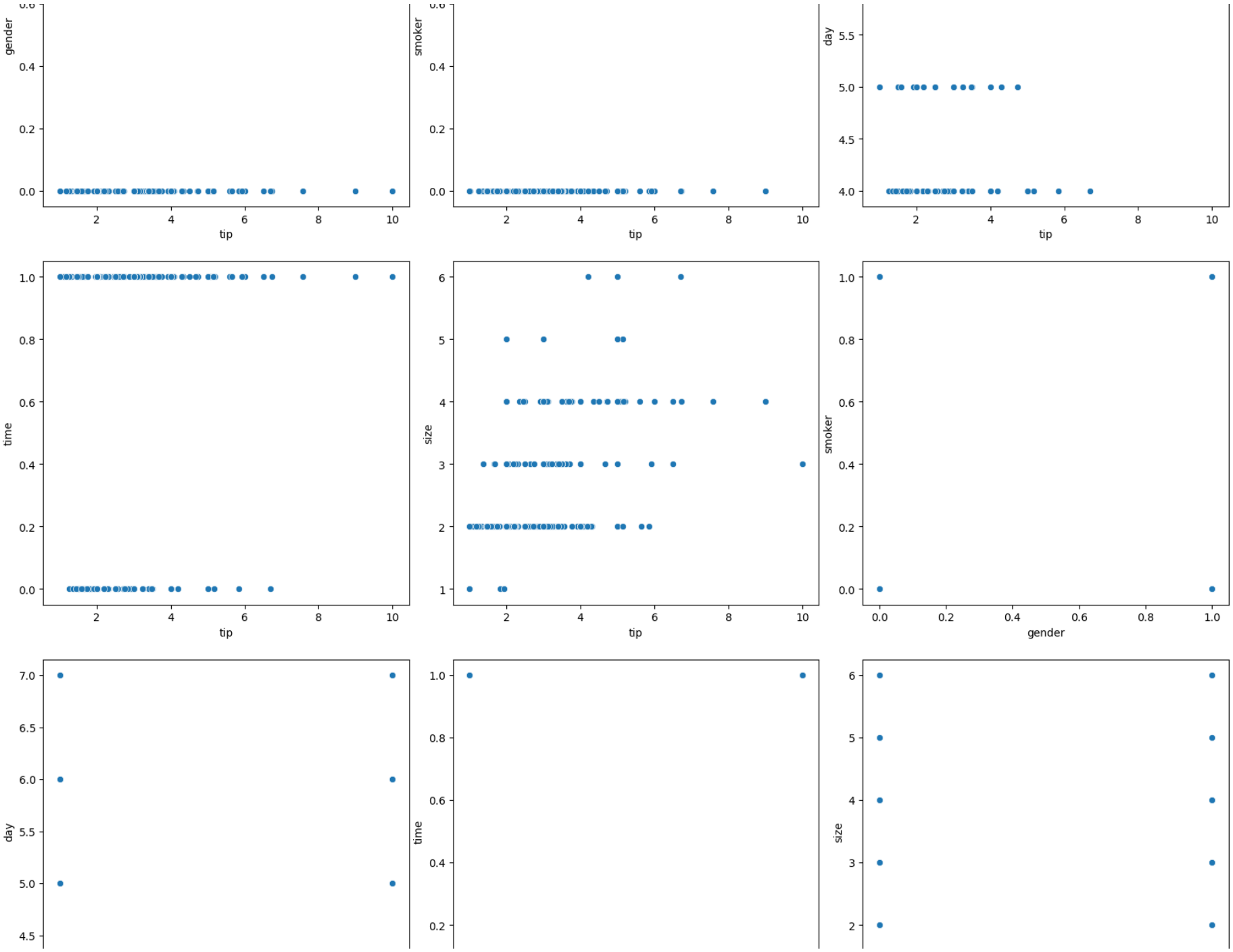

Count
Count
Count

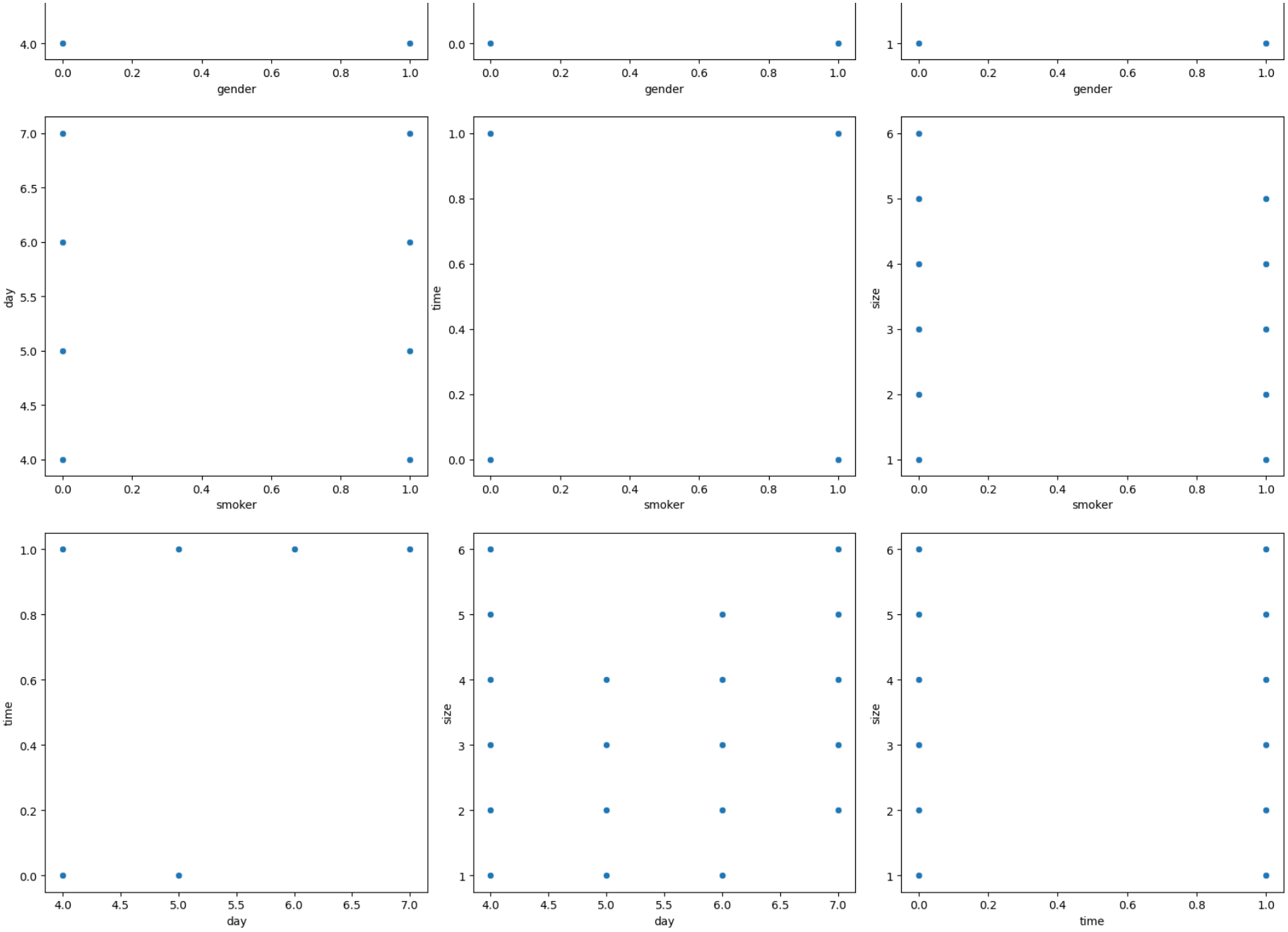


In [102…
```
# BiVariate  and Multivariate analysis
#convert categorical data to numerical representations
data.replace(to_replace=["Male","Female"],value=[0,1],inplace=True )
data=data.replace(to_replace=["No","Yes"],value=[0,1] )
data=data.replace({"day":{"Sun":7, "Sat":6,"Thur":4, "Fri":5}})
data=data.replace({"time":{"Lunch":0, "Dinner":1}})
```

In [103…
```
# scatter plots
fig,axes = plt.subplots(ncols=3, nrows=7, figsize=(16,36))
idx = 0
axes=axes.flatten()
done =[]
for k1, v1 in data.items():
    for k, v in data.items():
        if(k==k1):
            continue
        if ({k,k1} in done):
            continue

        sns.scatterplot(data=data, x=k1,y=k,ax=axes[idx])
        done.append({k,k1})
        idx+=1
plt.tight_layout(pad=0.4, w_pad=0.4, h_pad=2.0)
```


tip
bill
gender
bill
smoker
bill
day
bill
time
bill
size
bill


gender
tip
smoker
tip
day
tip
time
tip
size
tip
smoker
gender
day
time
size

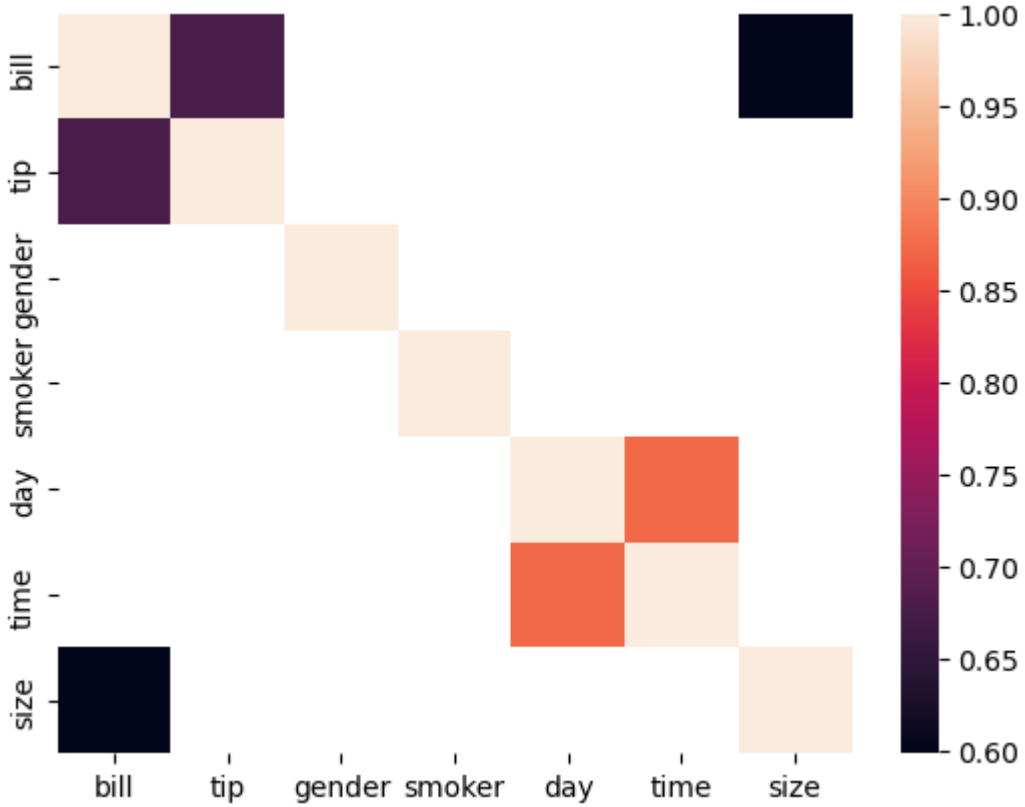


In [104…

```
# correlation matrix
```

```
#corr=data.select_dtypes(include="number").corr()
corr= corr[(corr>0.5)| (corr<-0.5)]
#print(corr)

sns.heatmap(corr)
```

Out[104… <Axes: >



In [ ]: