06_Correlation

April 2, 2022

1 Correlation

- What is Correlation
- Variable wihtin a dataset can be related for lots of reasons
- Types:
- Pearson
- Spearman's rho
- Kendall's tau

1.1 For Example

1. One variable could cause or depend on the value of another variable.
2. One variable could be lightly associated with another variable.
3. Two variables could depend on a third unknown variable.

Positive Correlation: both variables chnage in the same direction.

Nuetral Coorelation: No relationship in the change of the variables.

Negative Coorelation: variables change in opposite directions.

2 Coavriance

- Variables can be realted by a linear relationship. This is a relationship that is consistently additive across the two data samples.
- The realtionship can be summarized between two variables, called th covariance.
- The sign of the covariance can be interperated as whether the two variables change in the same dirction (positive) or change in different directions (negative)
- The magnitude of the is not easily interperated. A covariance value of zero indicates that both variables are completely independent.

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

#import dataset
kashti = sns.load_dataset('titanic')
phool = sns.load_dataset('iris')
```

[ ]:
```
kashti.head()
```

[ ]:
```
   survived  pclass     sex   age  sibsp  parch     fare embarked  class  \
0         0       3    male  22.0      1      0   7.2500        S  Third
1         1       1  female  38.0      1      0  71.2833        C  First
2         1       3  female  26.0      0      0   7.9250        S  Third
3         1       1  female  35.0      1      0  53.1000        S  First
4         0       3    male  35.0      0      0   8.0500        S  Third

     who  adult_male deck  embark_town alive  alone
0    man        True  NaN  Southampton    no  False
1  woman       False    C    Cherbourg   yes  False
2  woman       False  NaN  Southampton   yes   True
3  woman       False    C  Southampton   yes  False
4    man        True  NaN  Southampton    no   True
```

[ ]:
```
np.cov(kashti['age'], kashti['fare'])  #covarinace ko kisi or tariqey se dekh␣
↪skte hn likhen
```

[ ]:
```
array([[          nan,           nan],
       [          nan, 2469.43684574]])
```

[ ]:
```
# Python code to demonstarte the
# use of numpy

import numpy as np

x = [1.23, 2.12, 3.34, 4.5]
y = [2.56, 2.89, 3.76, 3.95]

# find out covariance with respect to columns
cov_mat = np.stack((x, y), axis = 0)
cov_mat
print(np.cov(cov_mat))
```

```
[[2.03629167 0.9313    ]
 [0.9313     0.4498    ]]
```

3 Coorelation instead of cov

[ ]:
```
kashti.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
```

[ ]:
```
kashti.corr() # 1 highly correlated -1 ngatively coorelated , 0 nuetral
```

[ ]:
```
             survived    pclass       age     sibsp     parch      fare  \
survived     1.000000 -0.338481 -0.077221 -0.035322  0.081629  0.257307
pclass      -0.338481  1.000000 -0.369226  0.083081  0.018443 -0.549500
age         -0.077221 -0.369226  1.000000 -0.308247 -0.189119  0.096067
sibsp       -0.035322  0.083081 -0.308247  1.000000  0.414838  0.159651
parch        0.081629  0.018443 -0.189119  0.414838  1.000000  0.216225
fare         0.257307 -0.549500  0.096067  0.159651  0.216225  1.000000
adult_male  -0.557080  0.094035  0.280328 -0.253586 -0.349943 -0.182024
alone       -0.203367  0.135207  0.198270 -0.584471 -0.583398 -0.271832

             adult_male     alone
survived      -0.557080 -0.203367
pclass         0.094035  0.135207
age            0.280328  0.198270
sibsp         -0.253586 -0.584471
parch         -0.349943 -0.583398
fare          -0.182024 -0.271832
adult_male     1.000000  0.404744
alone          0.404744  1.000000
```

[ ]:
```
corr = kashti.corr(method="pearson")  #for normal data
```

[ ]:
```
corr1 = kashti.corr(method="spearman") # for non-guassian distribution
```

[ ]:
```
corr
```

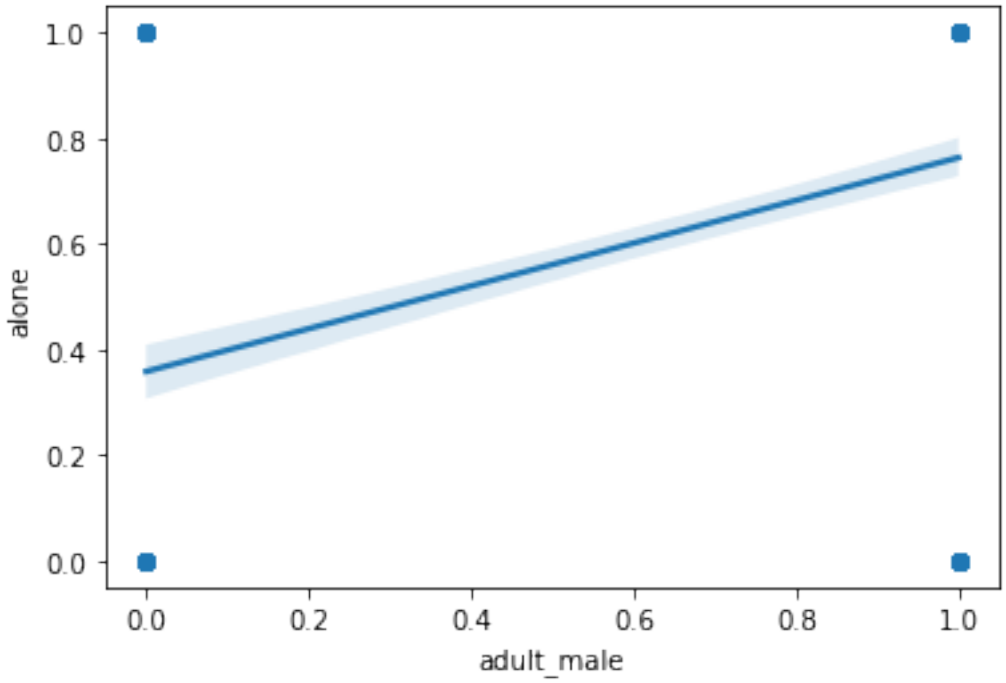
[ ]:
```
              survived    pclass       age     sibsp     parch      fare  \
survived      1.000000 -0.338481 -0.077221 -0.035322  0.081629  0.257307
pclass       -0.338481  1.000000 -0.369226  0.083081  0.018443 -0.549500
age          -0.077221 -0.369226  1.000000 -0.308247 -0.189119  0.096067
sibsp        -0.035322  0.083081 -0.308247  1.000000  0.414838  0.159651
parch         0.081629  0.018443 -0.189119  0.414838  1.000000  0.216225
fare          0.257307 -0.549500  0.096067  0.159651  0.216225  1.000000
adult_male   -0.557080  0.094035  0.280328 -0.253586 -0.349943 -0.182024
alone        -0.203367  0.135207  0.198270 -0.584471 -0.583398 -0.271832

             adult_male     alone
survived      -0.557080 -0.203367
pclass         0.094035  0.135207
age            0.280328  0.198270
sibsp         -0.253586 -0.584471
parch         -0.349943 -0.583398
fare          -0.182024 -0.271832
adult_male     1.000000  0.404744
alone          0.404744  1.000000
```

[ ]:
```
sns.regplot(kashti['adult_male'], kashti['alone'], data= kashti)
```

```
C:\Users\Faiza\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
  warnings.warn(
```

[ ]:
```
<AxesSubplot:xlabel='adult_male', ylabel='alone'>
```



[ ]:
```
phool.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

[ ]:
```
phool.corr()
```

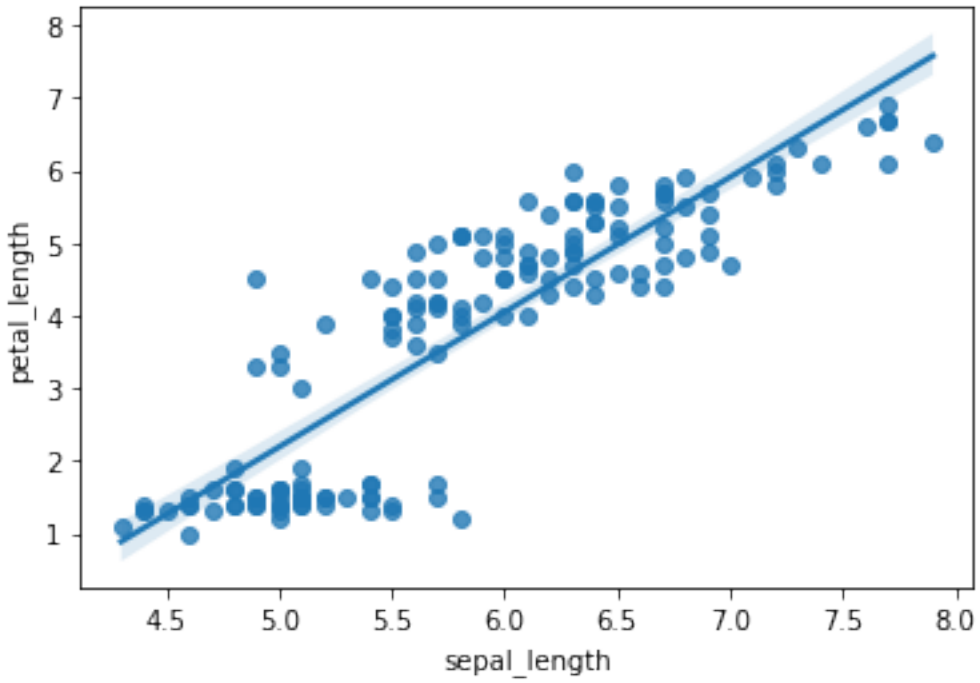
[ ]:

	sepal_length	sepal_width	petal_length	petal_width
sepal_length	1.000000	-0.117570	0.871754	0.817941
sepal_width	-0.117570	1.000000	-0.428440	-0.366126
petal_length	0.871754	-0.428440	1.000000	0.962865
petal_width	0.817941	-0.366126	0.962865	1.000000

[ ]:
```
sns.regplot(phool['sepal_length'], phool['petal_length'], data =phool)
```

```
C:\Users\Faiza\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
  warnings.warn(
```

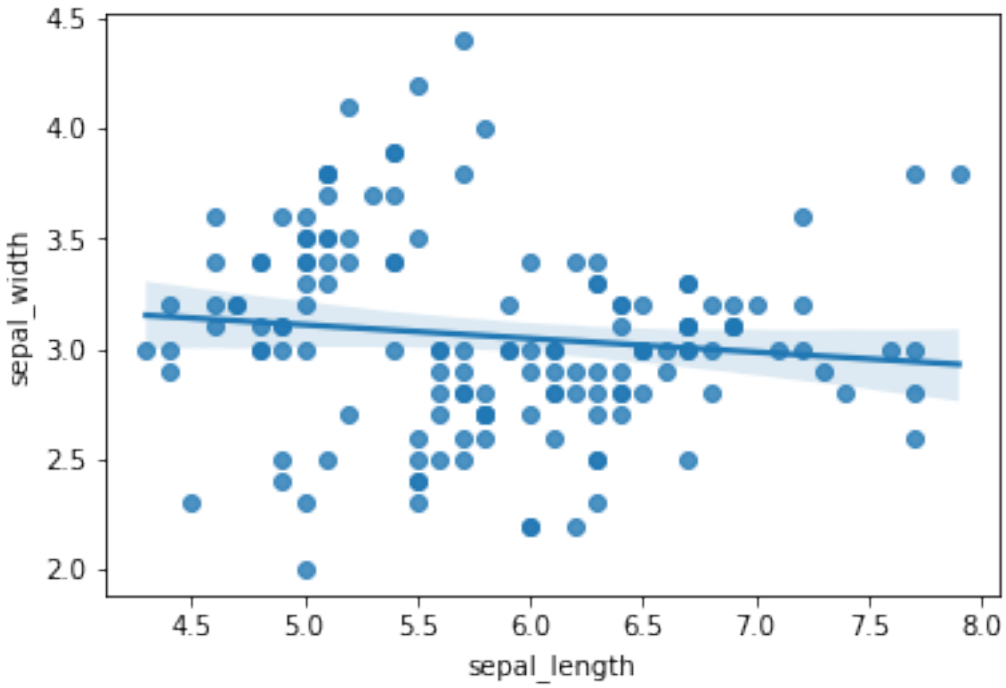
[ ]: <AxesSubplot:xlabel='sepal_length', ylabel='petal_length'>



[ ]:
```
sns.regplot(phool['sepal_length'], phool['sepal_width'], data =phool)
```

```
C:\Users\Faiza\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
  warnings.warn(
```

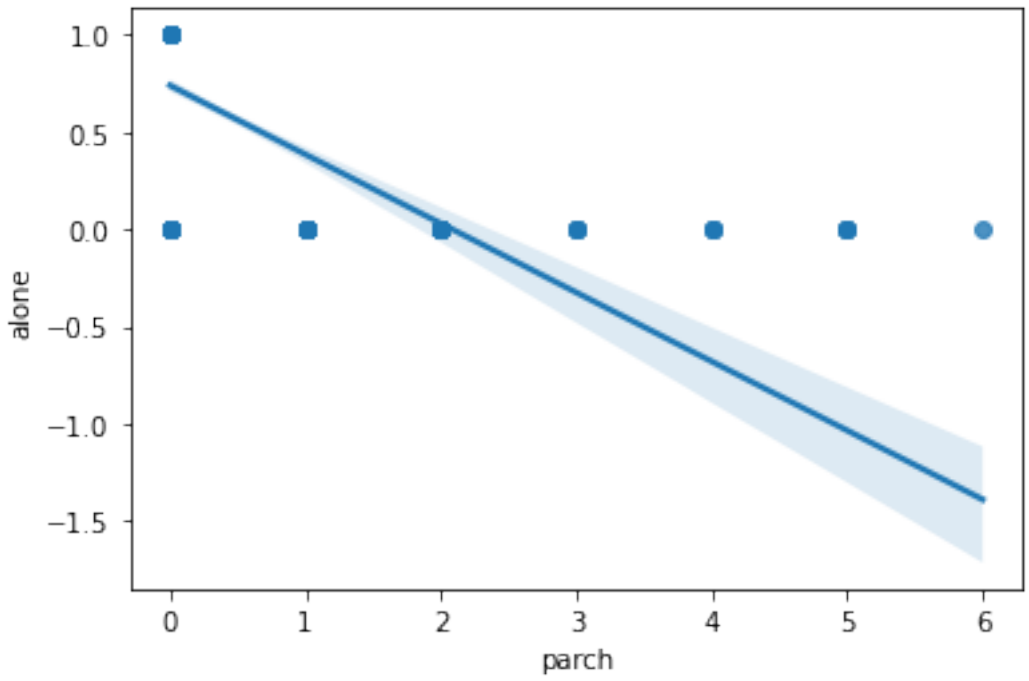
[ ]: <AxesSubplot:xlabel='sepal_length', ylabel='sepal_width'>



[ ]:
```
sns.regplot(kashti['parch'], kashti['alone'], data =kashti)
```

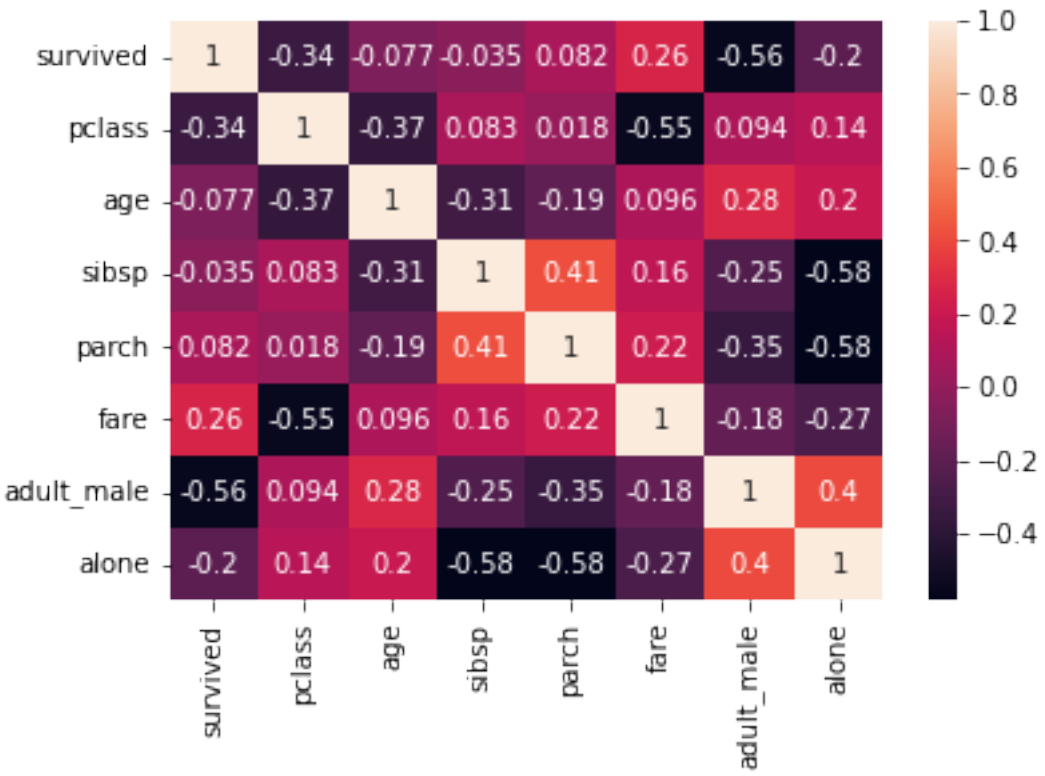
```
C:\Users\Faiza\AppData\Local\Programs\Python\Python310\lib\site-
packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variables
as keyword args: x, y. From version 0.12, the only valid positional argument
will be `data`, and passing other arguments without an explicit keyword will
result in an error or misinterpretation.
  warnings.warn(
```

[ ]:
```
<AxesSubplot:xlabel='parch', ylabel='alone'>
```



[ ]:
```
# apply corr function
corr = kashti.corr(method = 'pearson') # for normal data
# heat map
sns.heatmap(corr, annot =  True)
```

[ ]: <AxesSubplot:>



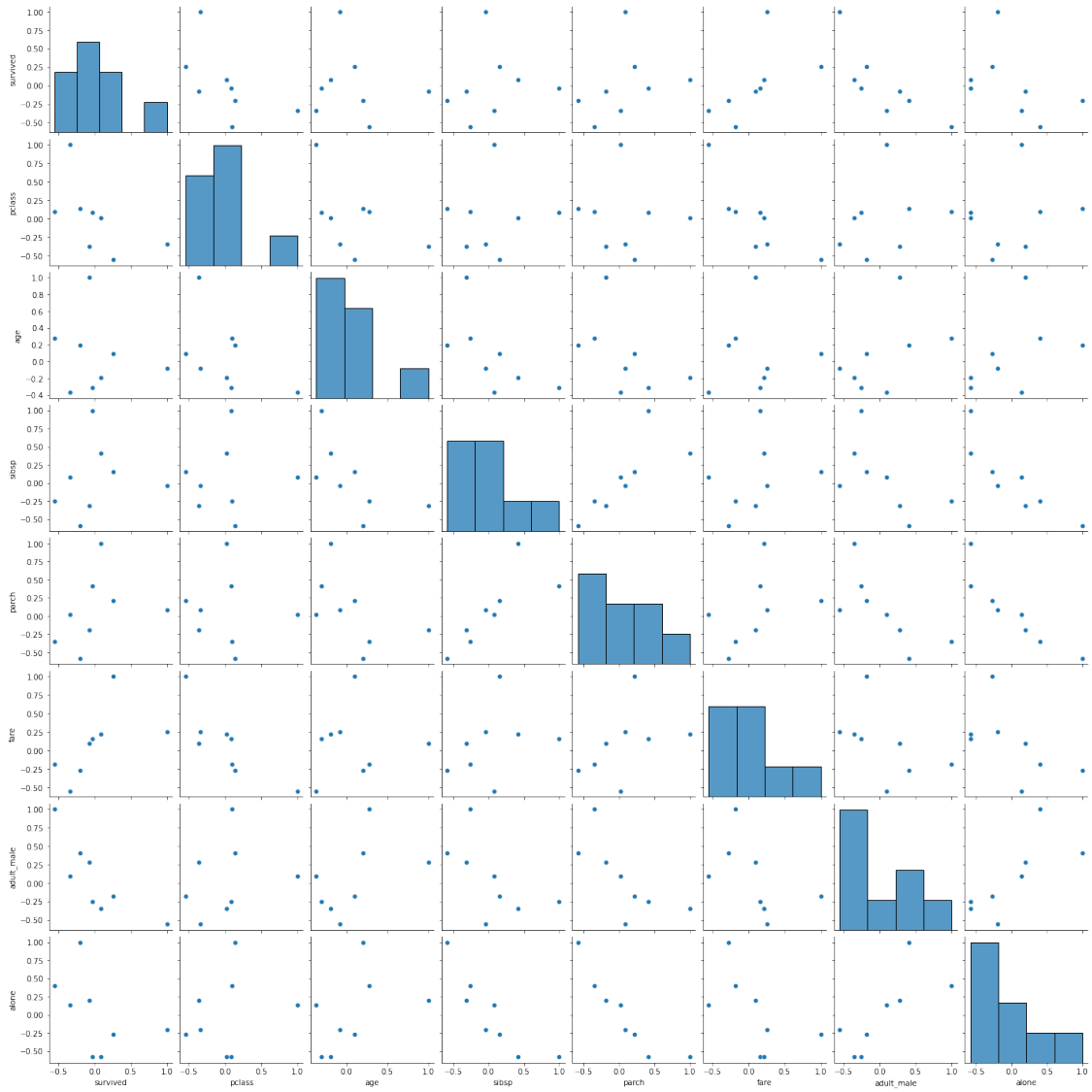
[ ]:
```
# jb ap ka 0.5,0.6 se 1 ki range  so higly positevely correlated  -0.5 se -0.6
↪do higly negativvely correlated
```

[ ]:
```
corr.style.background_gradient(cmap='coolwarm')
```

[ ]:
```
<pandas.io.formats.style.Styler at 0x1d6f7c02500>
```

[ ]:
```
sns.pairplot(corr)
```

[ ]:
```
<seaborn.axisgrid.PairGrid at 0x1d6f6af4ac0>
```



[ ]:
```
# we can change the points based on category
# import a new dataset
penguins = sns.load_dataset("penguins")
penguins.head()
```

[ ]:

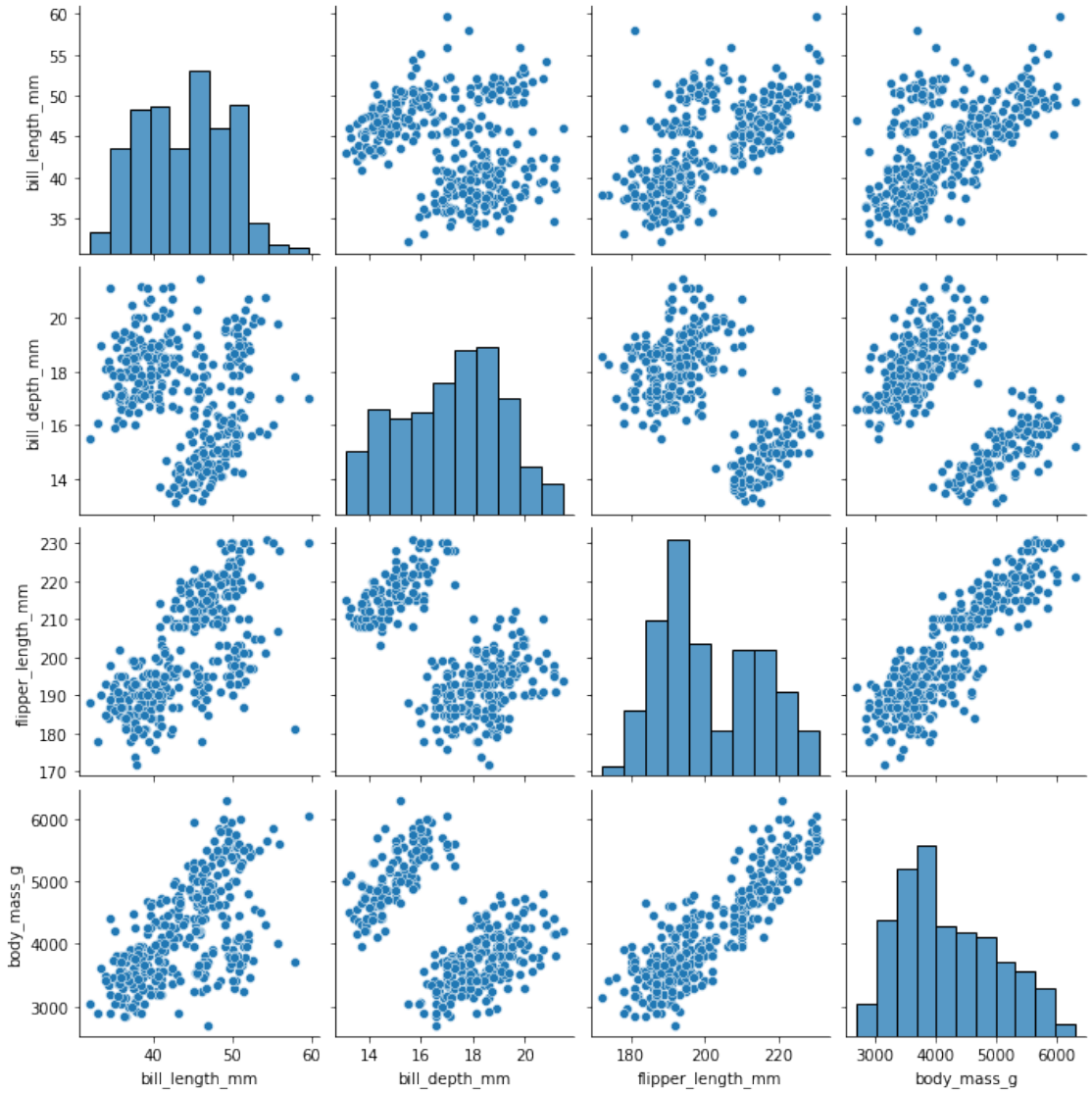
	species	island	bill_length_mm	bill_depth_mm	flipper_length_mm	\
0	Adelie	Torgersen	39.1	18.7	181.0	
1	Adelie	Torgersen	39.5	17.4	186.0	
2	Adelie	Torgersen	40.3	18.0	195.0	
3	Adelie	Torgersen	NaN	NaN	NaN	
4	Adelie	Torgersen	36.7	19.3	193.0	

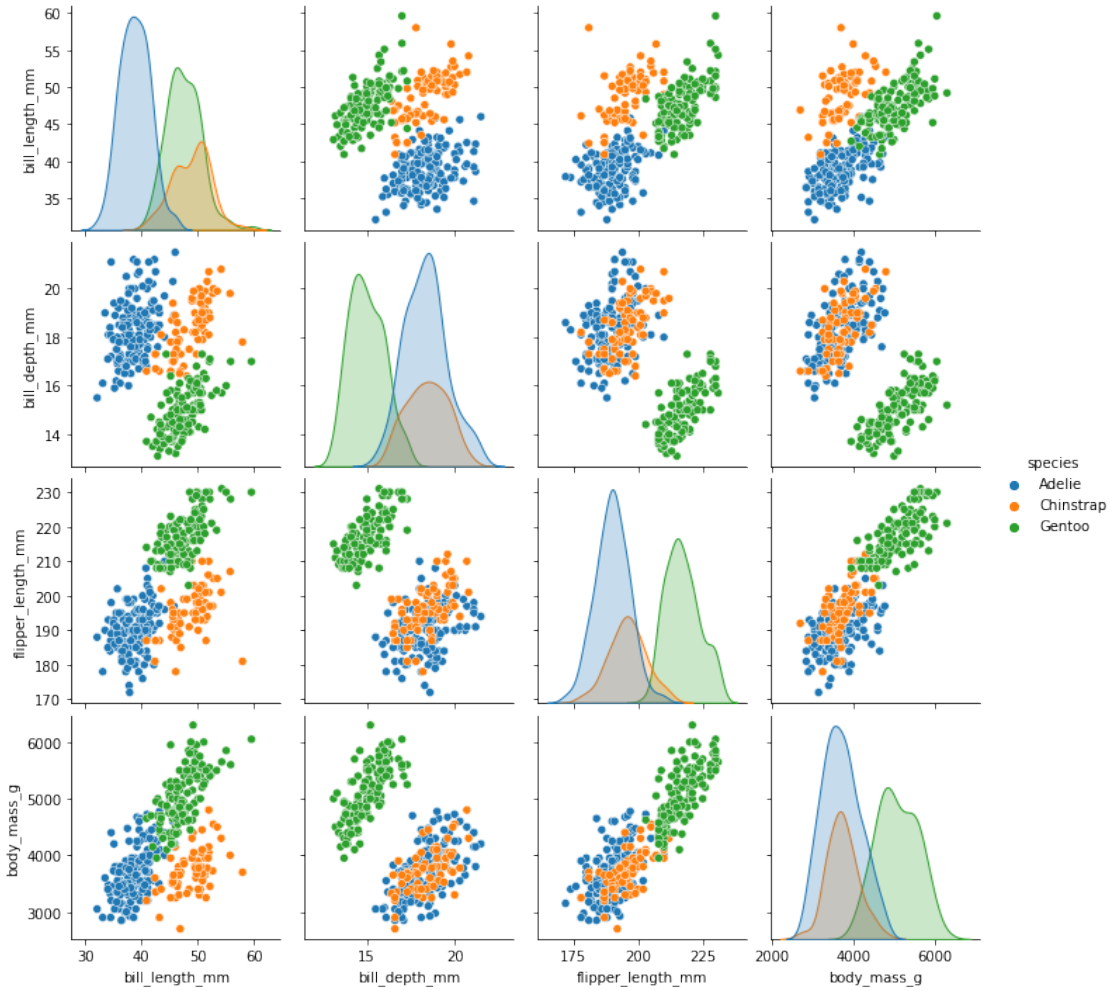
	body_mass_g	sex

```
0       3750.0    Male
1       3800.0  Female
2       3250.0  Female
3          NaN     NaN
4       3450.0  Female
```

[ ]:
```
sns.pairplot(penguins)
sns.pairplot(penguins, hue= 'species')
```

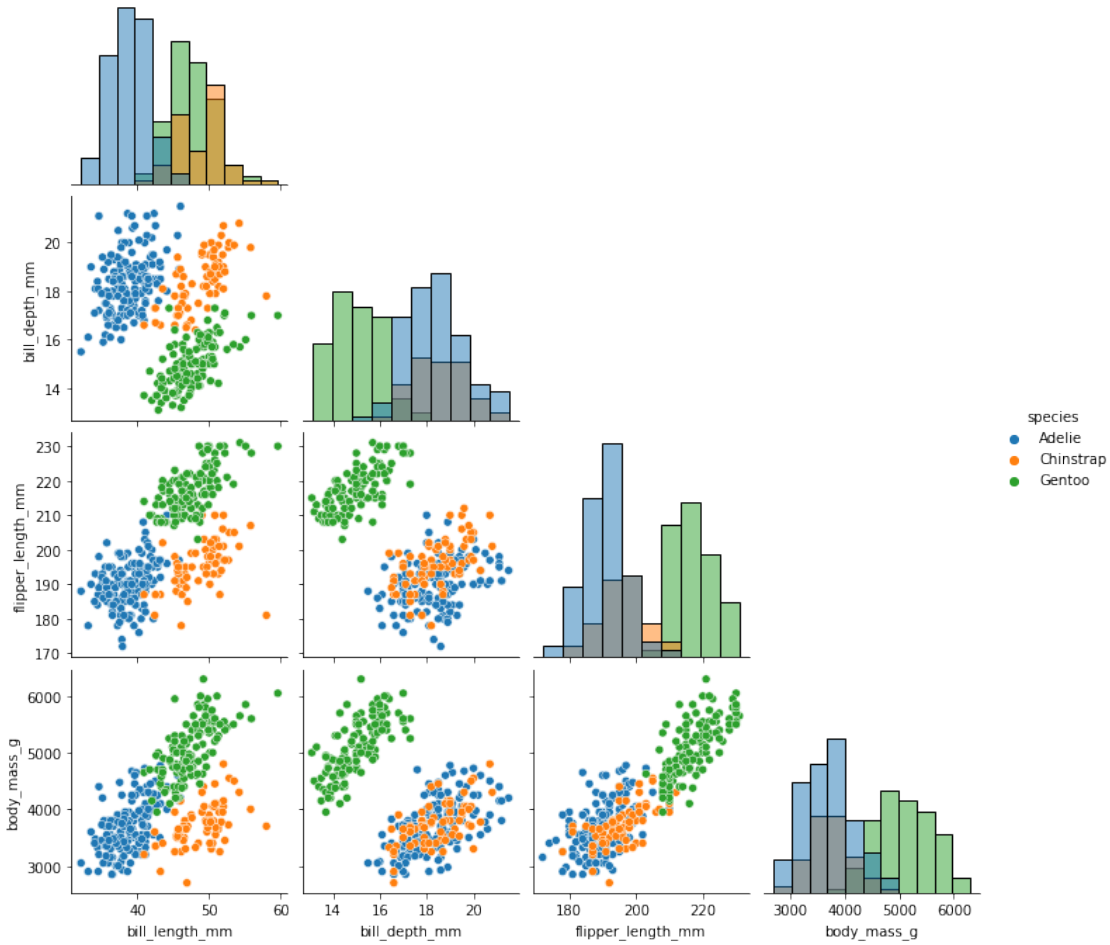
[ ]:
```
<seaborn.axisgrid.PairGrid at 0x1d6facda7a0>
```





[ ]:
```
#cwe can convert this into histograms
sns.pairplot(penguins, hue= "species", diag_kind="hist", corner= True)
```

[ ]: <seaborn.axisgrid.PairGrid at 0x1d6ffdff460>



[ ]:
```
phool.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

[ ]:
```
# clculate Pearson's correlation
from scipy.stats import pearsonr #spearman
corr, _ = pearsonr(phool['sepal_length'], phool['petal_width'])
print('Pearson correlation :%.3f' % corr)
```

```
Pearson correlation :0.818
```

[ ]:
```
#Assignemnt# 4 Qism k plot +ve corr, -ve corr, 0 corr, slightly +ve or slightly␣
↪-ve
```