Lecture 5 ¶

Using seaborn and Matplotlib library

Importing Library

In [1]:
```
import numpy as np
import pandas as pd


import matplotlib.pyplot as plt
#from pylab import rcParams
```

In [2]:
```
import seaborn as sns
%matplotlib inline
```

Import Dataset

In [3]:
```
cars = pd.read_csv('Downloads/mtcars.csv')
cars.columns
```

Out[3]:
```
Index(['name', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am',
       'gear', 'carb'],
      dtype='object')
```

In [4]:
```
cars.head()
```

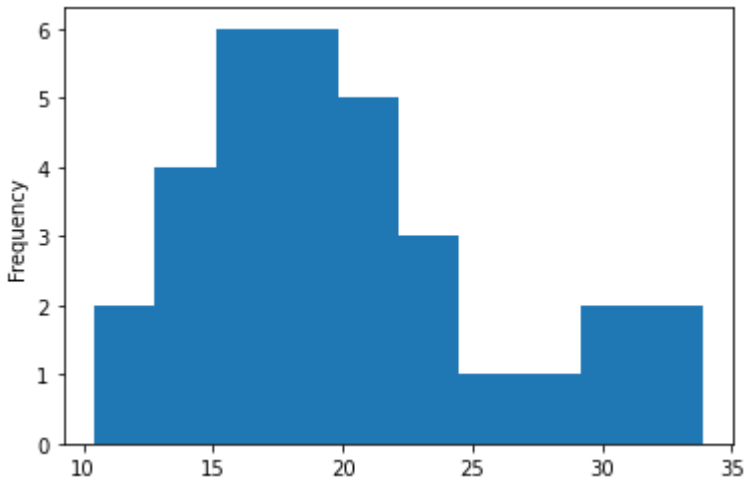
Out[4]:

	name	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
0	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
1	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
2	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
3	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
4	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2

In [6]:
```
mpg = cars['mpg']

mpg.plot(kind='hist')
```

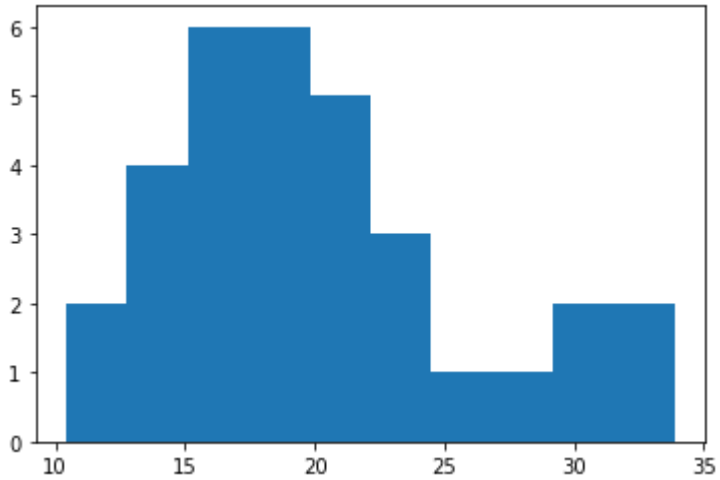
Out[6]: <AxesSubplot:ylabel='Frequency'>



Another way to create histogram in python is as follows

In [7]:
```
plt.hist(mpg)
plt.plot()
```

Out[7]: []



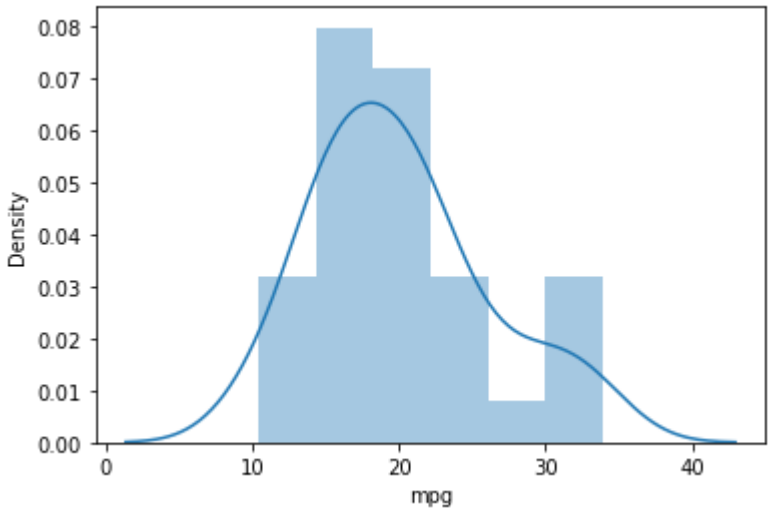
distplot

A Distplot or distribution plot, depicts the variation in the data distribution. Seaborn Distplot represents the overall distribution of continuous data variables. The Distplot depicts the data by a histogram and a line in combination to it.

In [8]:
```
1 sns.distplot(cars['mpg'])
2 # Safe to ignore warnings
```

```
C:\Users\prpou\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[8]: <AxesSubplot:xlabel='mpg', ylabel='Density'>



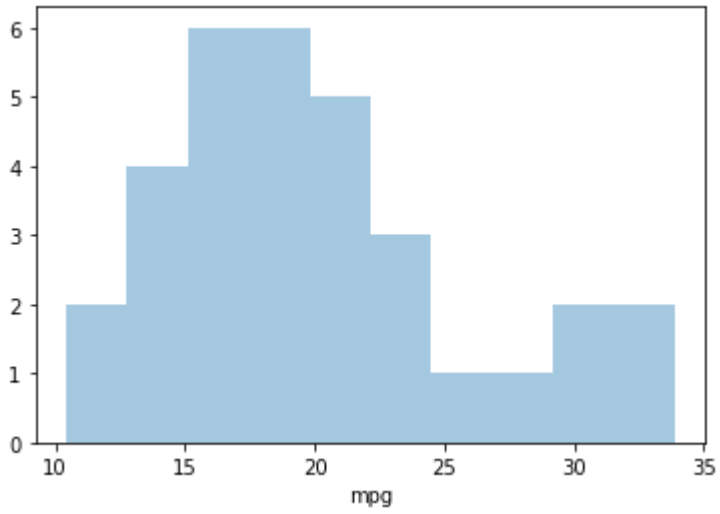
To remove the kde layer and just have the histogram use:

Bins

The bins parameter enables you to control the bins of the histogram (i.e., the number of bars). The most common way to do this is to set the number of bins by providing an integer as the argument to the parameter. For example, if you set bins = 30 , the function will create a histogram with 30 bars

In [11]:
```
1 sns.distplot(cars['mpg'],kde=False,bins=10)
```

Out[11]: <AxesSubplot:xlabel='mpg'>

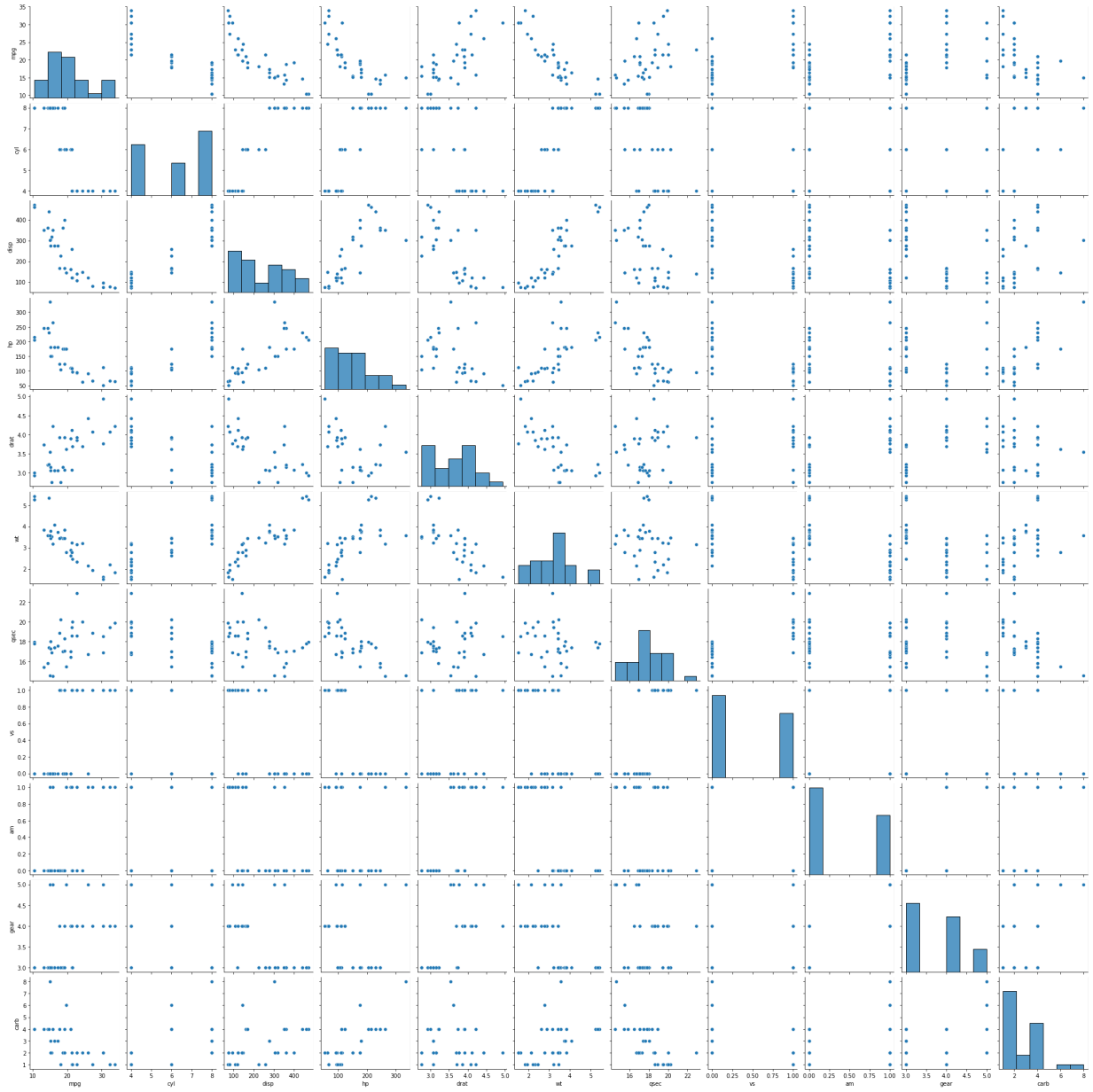


pairplot

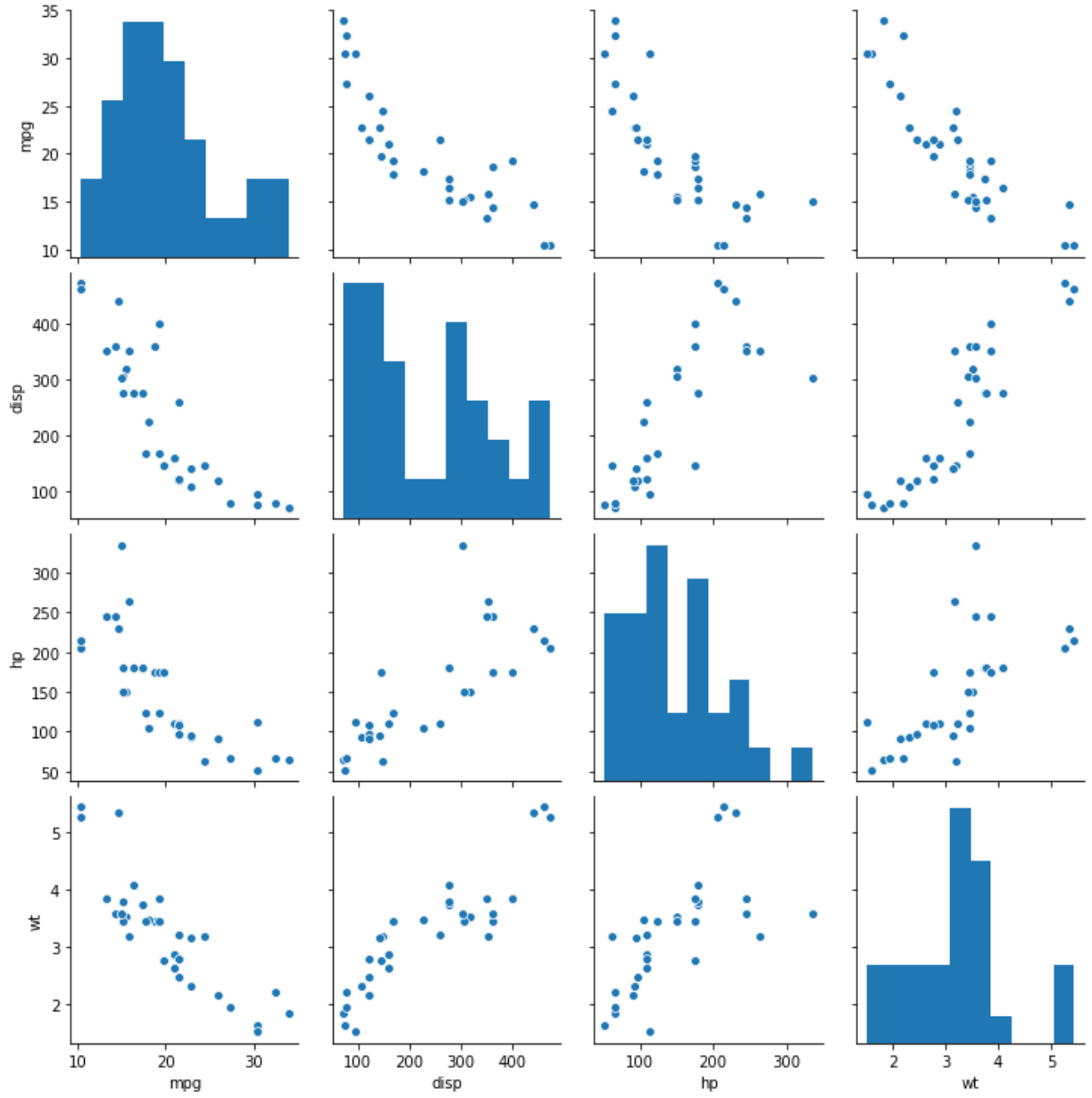
pairplot() : To plot multiple pairwise bivariate distributions in a dataset, you can use the pairplot() function. This shows the relationship for (n, 2) combination of variable in a DataFrame as a matrix of plots and the diagonal plots are the univariate plots. pairplot will plot pairwise relationships across an entire dataframe (for the numerical columns) and supports a color hue argument (for categorical columns).

In [12]:
```
1 sns.pairplot(cars)
```

Out[12]: <seaborn.axisgrid.PairGrid at 0x28384e57790>



In [37]:
```
cars_subset = cars[['mpg', 'disp', 'hp', 'wt']]
sb.pairplot(cars_subset)
plt.show()
```



Categorical Data Plots

In [17]:
```
import seaborn as sns
%matplotlib inline
```

In [18]:
```
df = pd.read_csv('Downloads/Titanic.csv')
```

In [19]:
```
1 df.head()
```

Out[19]:

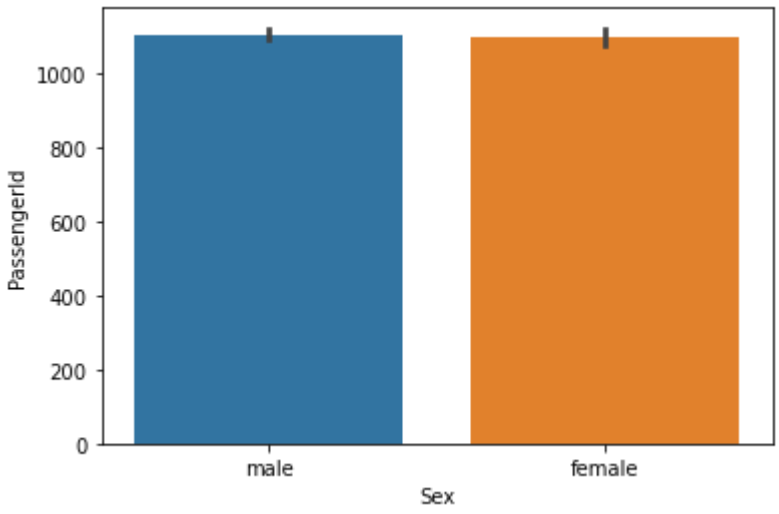
	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarke
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

barplot

These very similar plots allow you to get aggregate data off a categorical feature in your data.

In [20]:
```
1 sns.barplot(x='Sex',y='PassengerId',data=df)
```

Out[20]:
```
<AxesSubplot:xlabel='Sex', ylabel='PassengerId'>
```



boxplot

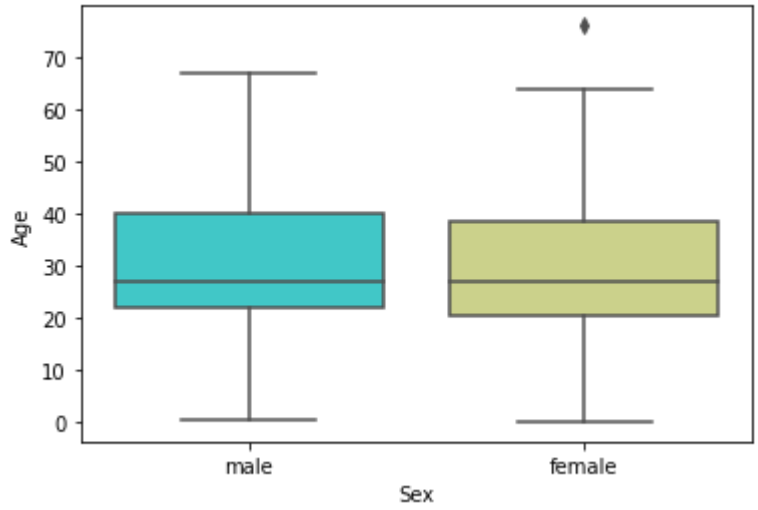
box plots allow us to choose a numerical variable, like age, and plot the distribution of age for each

category in a selected categorical variable.

A box plot (or box-and-whisker plot) shows the distribution of quantitative data in a way that facilitates comparisons between variables or across levels of a categorical variable.

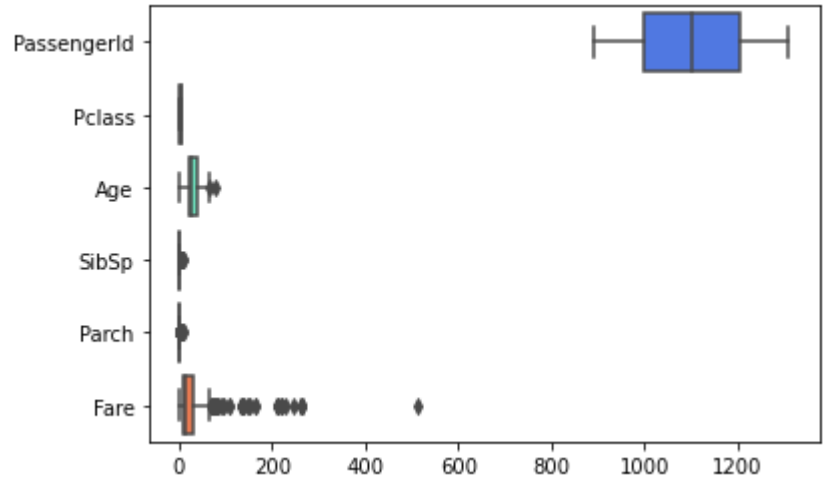
In [21]:
```
sns.boxplot(x="Sex", y="Age", data=df,palette='rainbow')
```

Out[21]: <AxesSubplot:xlabel='Sex', ylabel='Age'>



In [22]:
```
# Can do entire dataframe with orient='h'
sns.boxplot(data=df,palette='rainbow',orient='h')
```

Out[22]: <AxesSubplot:>



Grids

PairGrid allows us to draw a grid of subplots using the same plot type to visualize data.it uses different pair of variable for each subplot. It forms a matrix of sub-plots. It is also sometimes called as “scatterplot matrix”.

In [23]:
```
df = pd.read_csv('Downloads/Grade.csv')
df.head()
```

Out[23]:

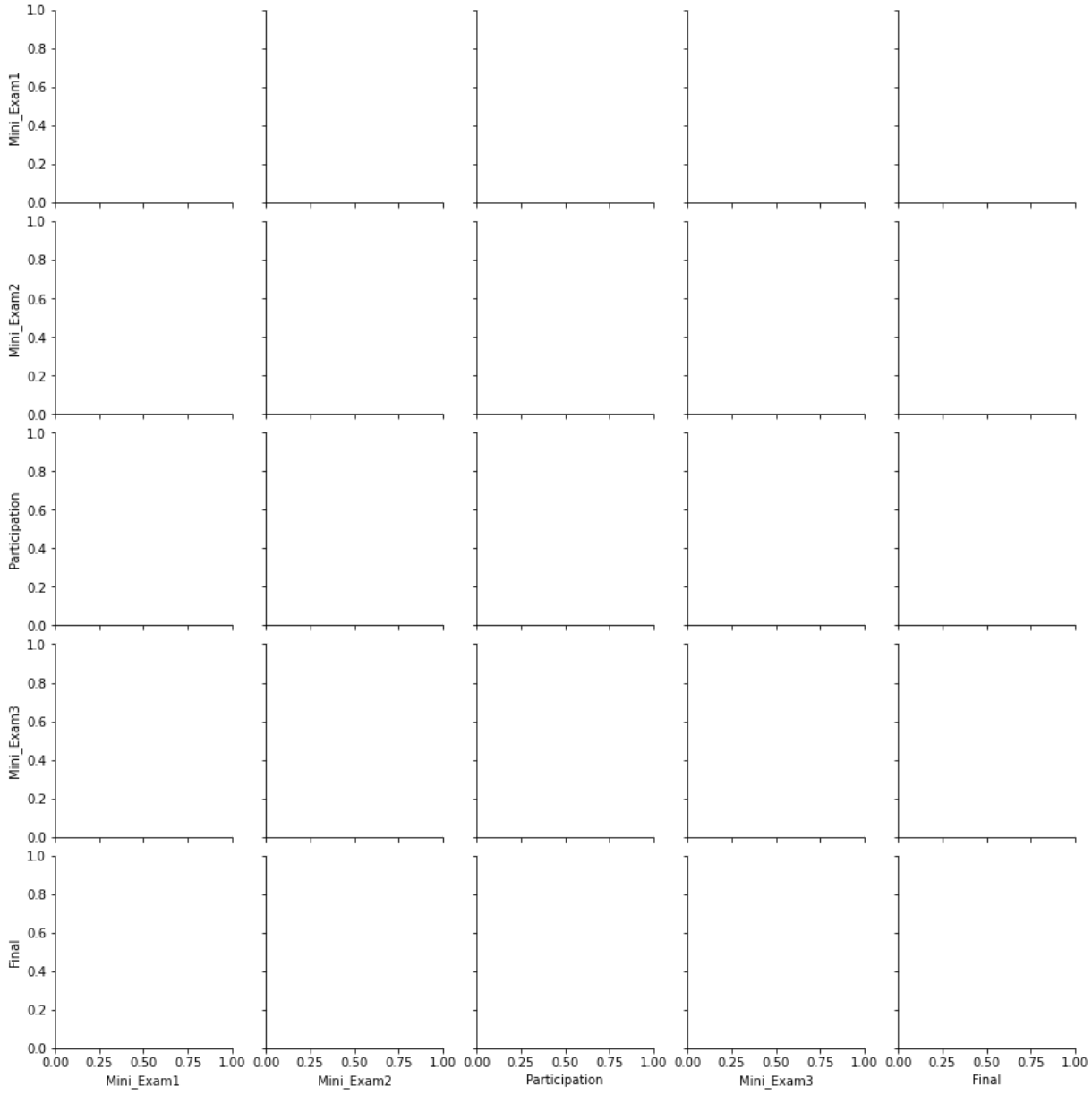
	Name	Mini_Exam1	Mini_Exam2	Participation	Mini_Exam3	Final	Grade
0	Jake	19.5	20.0	1	10.0	33.0	A
1	Joe	20.0	16.0	1	14.0	32.0	A
2	Susan	19.0	19.0	1	10.5	33.0	A-
3	Sol	22.0	13.0	1	13.0	34.0	A
4	Chris	19.0	17.0	1	12.5	33.5	A

PairGrid

Pairgrid is a subplot grid for plotting pairwise relationships in a dataset.PairGrid. Subplot grid for plotting pairwise relationships in a dataset. This object maps each variable in a dataset onto a column and row in a grid of multiple axes.

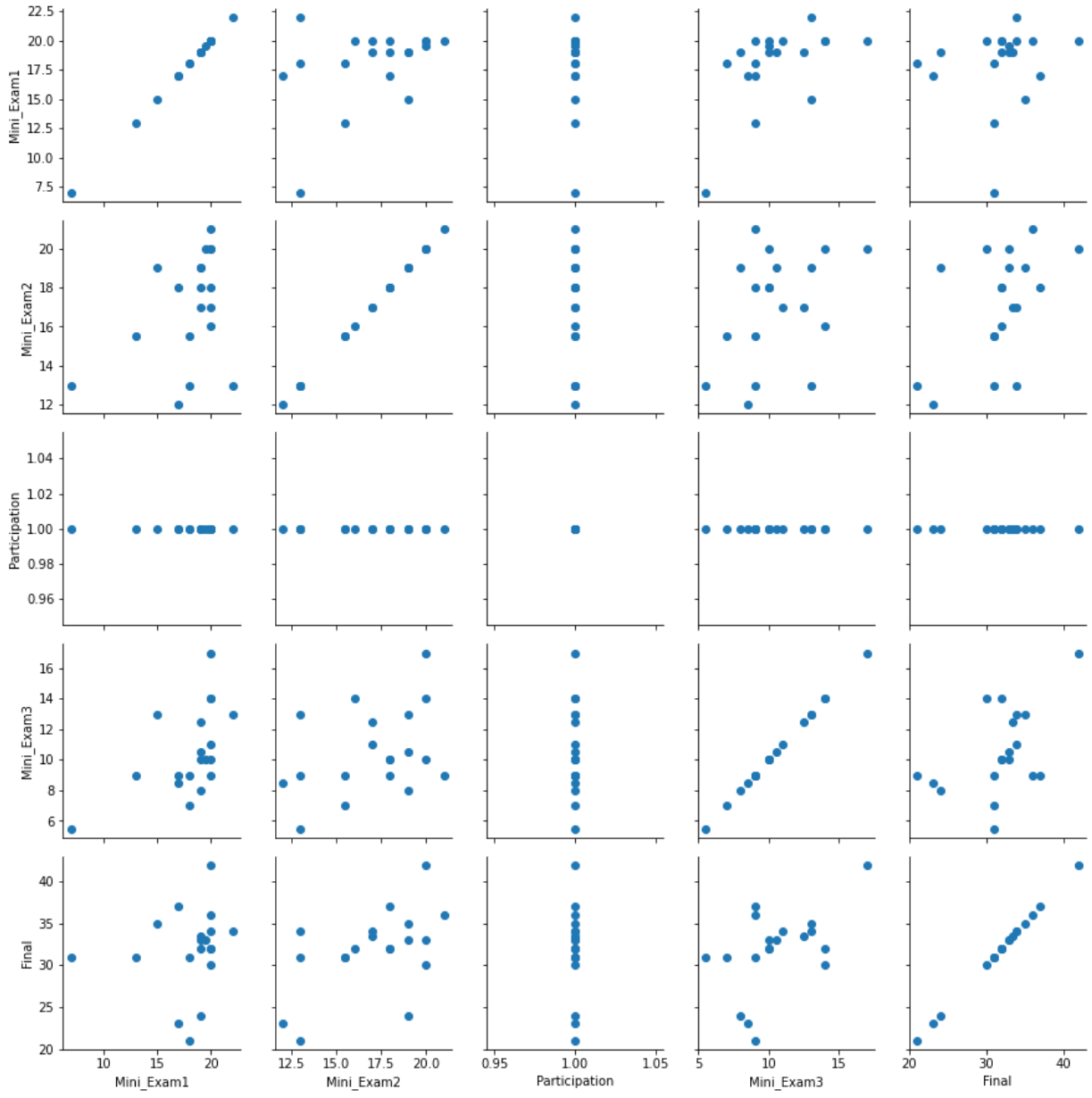
In [24]:
```
1 sns.PairGrid(df)
```

Out[24]: <seaborn.axisgrid.PairGrid at 0x28384e4abe0>



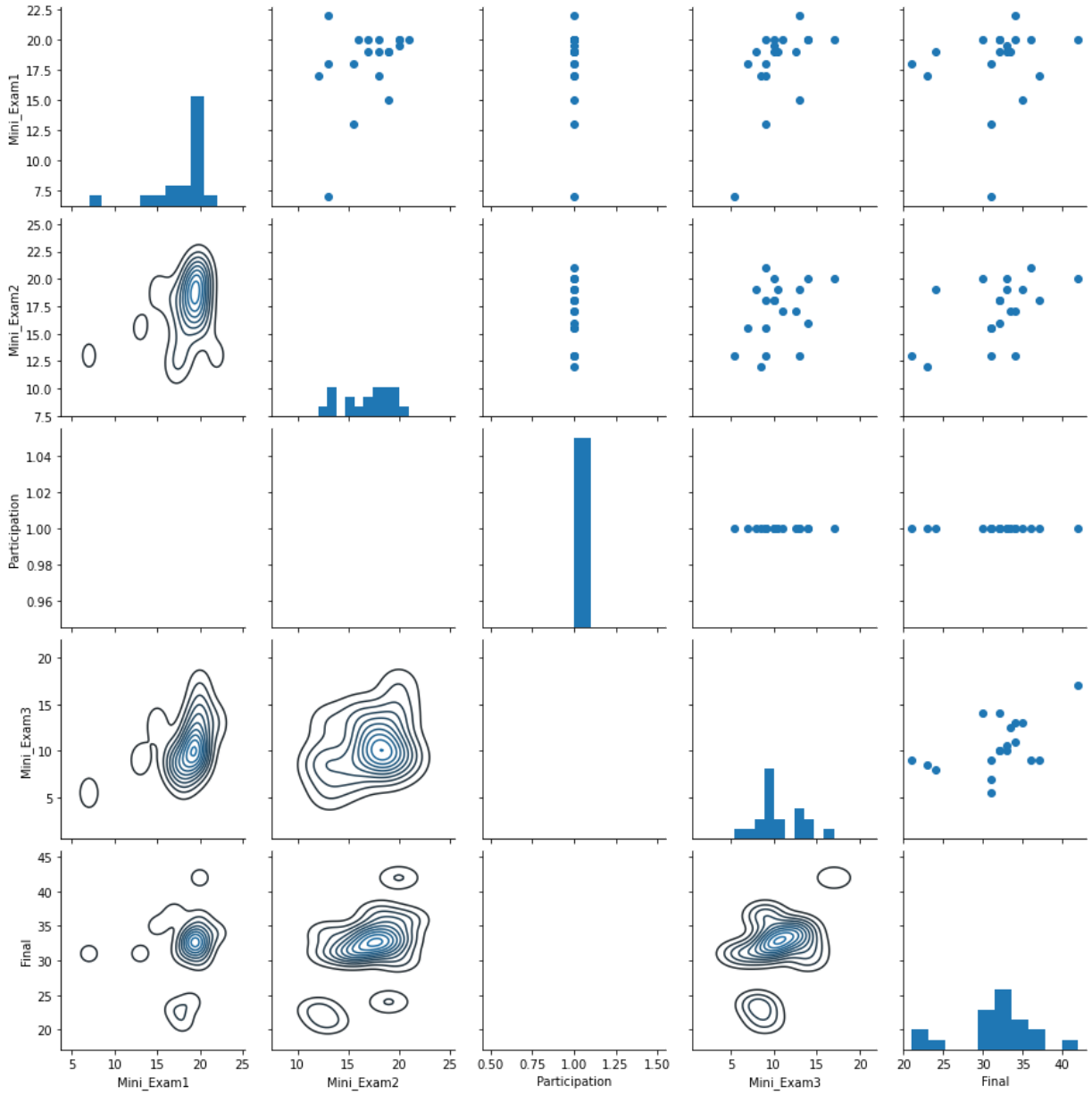
In [25]:
```
# map to the grid
g = sns.PairGrid(df)
g.map(plt.scatter)
```

Out[25]: <seaborn.axisgrid.PairGrid at 0x2838c0b07f0>



In [42]:
```
# Map to upper,lower, and diagonal
g = sns.PairGrid(df)
g.map_diag(plt.hist)
g.map_upper(plt.scatter)
g.map_lower(sns.kdeplot)
```

Out[42]: <seaborn.axisgrid.PairGrid at 0x1c35623b688>



In [ ]:
```
1
```