

Assignment 6

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
import os
import numpy as np
os.chdir("E:/GoogleDriveNew/PSU/DAAN862/Course contents/Lesson 4")
mtcars = pd.read_csv("mtcars.csv")
```

```
In [2]: mtcars.head()
```

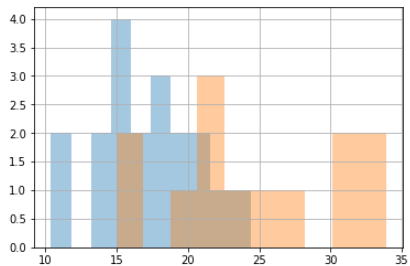
```
Out[2]:
```

	model	mpg	cyl	dis	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

1. Plot am-based histogram to compare mpg (20 points)

```
In [3]: plt.figure()
mtcars.groupby('am').mpg.hist(alpha = 0.4)
```

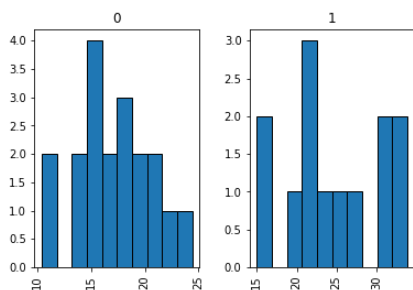
```
Out[3]: am
0      AxesSubplot(0.125,0.125;0.775x0.755)
1      AxesSubplot(0.125,0.125;0.775x0.755)
Name: mpg, dtype: object
```



```
In [4]: plt.figure()
mtcars.mpg.hist(by = mtcars.am, edgecolor = 'k')
```

```
Out[4]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x00002BA6D17C208>,
<matplotlib.axes._subplots.AxesSubplot object at 0x00002BA6D3C1860>],
dtype=object)
```

<Figure size 432x288 with 0 Axes>

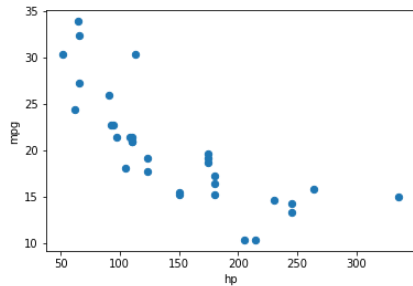


2. Use scatterplot to plot mpg VS. hp (20 points)

```
In [5]: plt.figure()
mtcars.plot.scatter(x = 'hp', y = 'mpg', s = 40)
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x2ba6d4688d0>
```

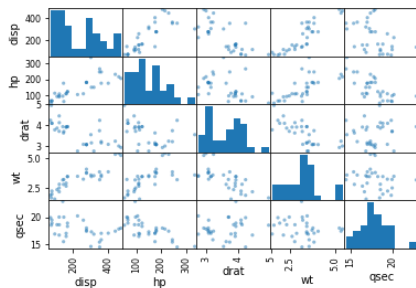
<Figure size 432x288 with 0 Axes>



3. Create a scatterplot matrix for a new data consisting of columns [disp, hp, drat, wt, qsec]. (20 points)

```
In [6]: new = mtcars[['disp', 'hp', 'drat', 'wt', 'qsec']]
pd.plotting.scatter_matrix(new, s = 40)
```

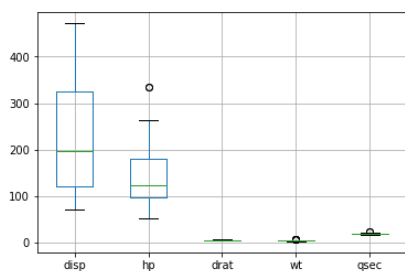
```
Out[6]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D4E6630>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D508F98>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D538668>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D55FC88>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D591358>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D591390>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D5EC088>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D613748>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D63DD08>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D66E4A8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D696B38>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D6C8208>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D6EF898>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D718F28>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D7485F8>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D772C88>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D7A3358>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D7CA9E8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D7FD088>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D826748>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D84EDD8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D87E4A8>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D8A6B38>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D8D9208>,
<matplotlib.axes._subplots.AxesSubplot object at 0x000002BA6D900898>]],
dtype=object)
```



4. Create barplots for a new data consisting of columns [disp, hp, drat, wt, qsec]. (20 points)

```
In [7]: new.plot.box()
new.boxplot()
```

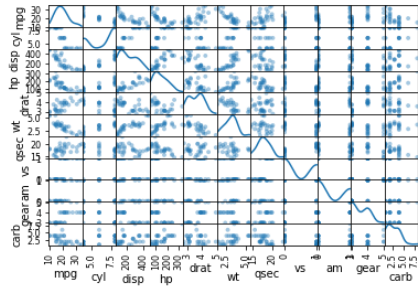
```
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x2ba6dabc908>
```



5 Determine which variable has most impact on mpg (20 points)

```
In [8]: plt.figure(figsize = (12, 12))
pd.plotting.scatter_matrix(mtcars, alpha = 0.4, s = 60, diagonal = 'kde');
```

<Figure size 864x864 with 0 Axes>



You can get general impression about the relationship of variables, but it doesn't provide quatitive analysis.

```
In [9]: cors = mtcars.corr().round(2)
cors.mpg.abs().sort_values(ascending = False)
```

```
Out[9]: mpg      1.00
wt         0.87
disp       0.85
cyl        0.85
hp         0.78
drat       0.68
vs         0.66
am         0.60
carb       0.55
gear       0.48
qsec       0.42
Name: mpg, dtype: float64
```

According to the correlation, wt is the most important variable to mpg.

```
In [10]: # Optional
fig, ax = plt.subplots(figsize = (8, 8))
cax = ax.matshow(cors)
plt.xticks(range(len(cors.columns)), cors.columns)
plt.yticks(range(len(cors.columns)), cors.columns)
fig.colorbar(cax)
```

Out[10]: <matplotlib.colorbar.Colorbar at 0x2ba7253a748>

