The Engineering World #DataScience 5 & 6

May 31, 2018

AKKAL BAHADUR BIST
DATA SCIENTIST AT
KATHMANDU INSTITUTE OF APPLIED SCIENCES (KIAS)
Center for Conservation Biology (CCB)

1 GROUPING AND AGGREGATE DATA

In [1]:
```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
```

1.0.1 Grouping data by column index

In [2]:
```
address = 'mtcars.csv'
```

In [3]:
```
cars = pd.read_csv(address)
```

In [4]:
```
cars.columns = ['car_name', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am'
```

In [5]:
```
cars.head()
```

Out[5]:

	car_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]:
```
cars_groups = cars.groupby(cars['cyl'])
```

In [7]:
```
cars_groups.mean()
```

Out[7]:
```
            mpg        disp          hp      drat        wt       qsec  \
cyl
4     26.663636  105.136364   82.636364  4.070909  2.285727  19.137273
6     19.742857  183.314286  122.285714  3.585714  3.117143  17.977143
8     15.100000  353.100000  209.214286  3.229286  3.999214  16.772143

            vs        am      gear      carb
cyl
4     0.909091  0.727273  4.090909  1.545455
6     0.571429  0.428571  3.857143  3.428571
8     0.000000  0.142857  3.285714  3.500000
```

2 LINE, BAR AND PIE PLOTS

In [8]:
```
import numpy as np
import pandas as pd
from pandas import Series, DataFrame
from numpy.random import randn
import matplotlib.pyplot as plt
from matplotlib import rcParams
import seaborn as sb
```

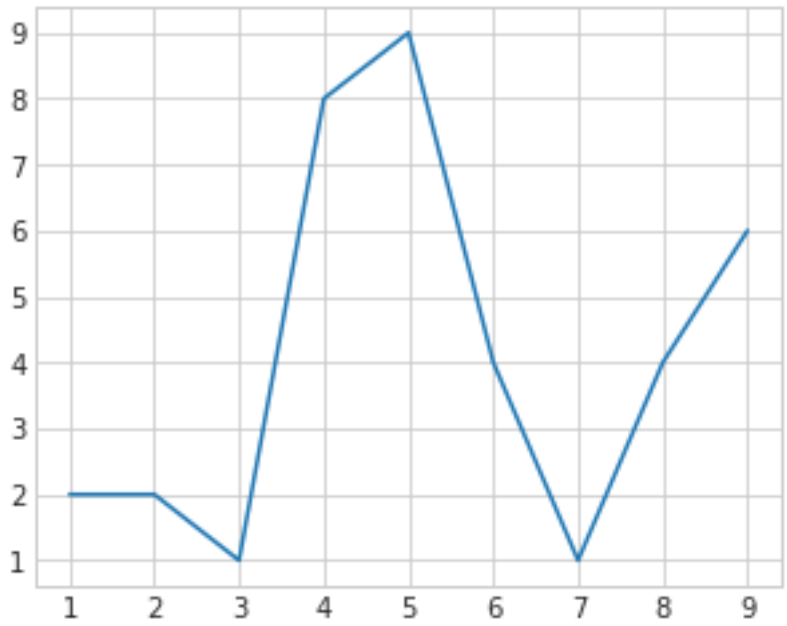
In [9]:
```
%matplotlib inline
rcParams['figure.figsize'] = 5, 4
sb.set_style('whitegrid')
```

2.0.1 Creating a line chart from a list object

2.0.2 Plotting line chart in matplotlib

In [10]:
```
x = range(1,10)
y = [2, 2, 1, 8, 9, 4, 1, 4, 6]
plt.plot(x,y)
```

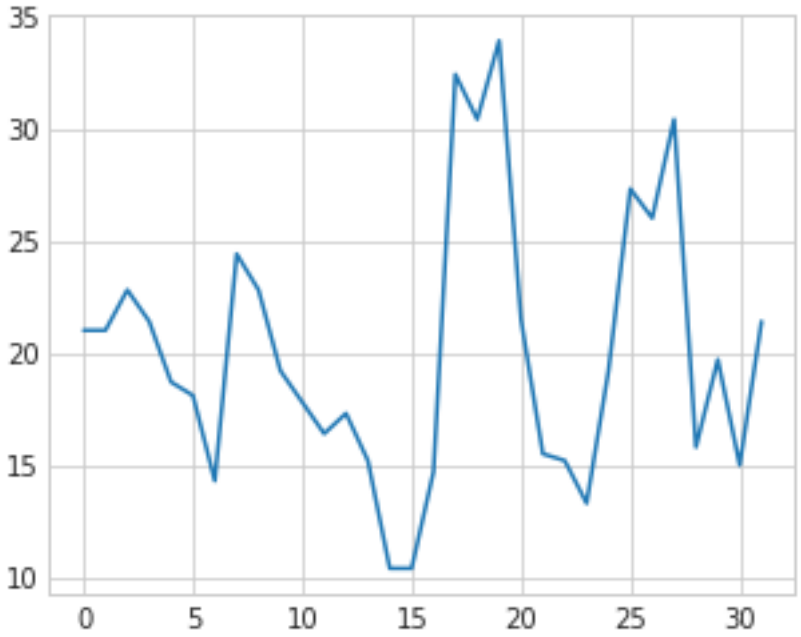
Out[10]:
```
[<matplotlib.lines.Line2D at 0x7fda1481a240>]
```



2.0.3 Plotting a line chart from a Pandas object

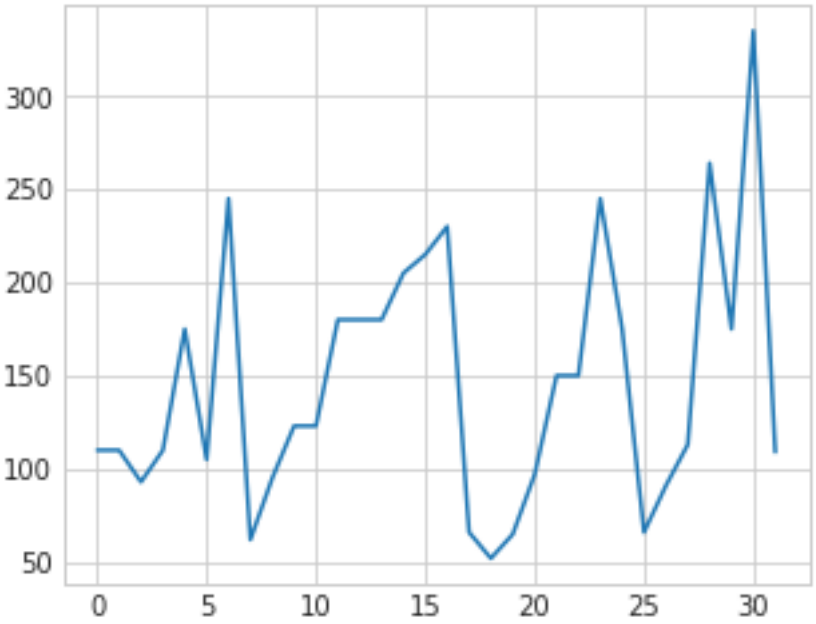
```
In [11]: address = 'mtcars.csv'
         cars = pd.read_csv(address)
         cars.columns = ['car_name', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am
         mpg = cars['mpg']
         mpg.plot()

Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda147c9710>
```



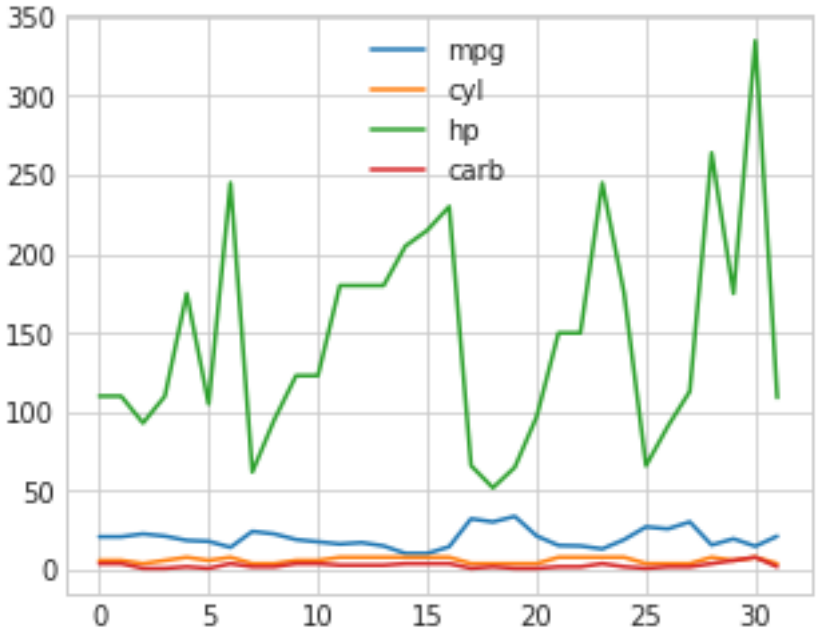
In [12]:
```
hp = cars['hp']
hp.plot()
```

Out[12]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda147b4160>



```
In [13]: df = cars[['mpg', 'cyl', 'hp', 'carb']]
         df.plot()
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda147eb898>
```

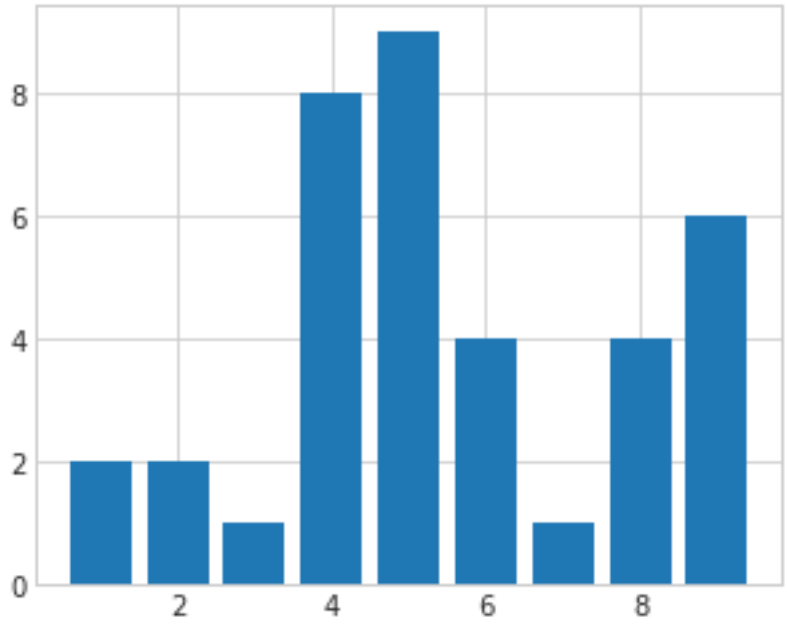


2.0.4 Creating Bar Chart

2.0.5 Creating a bar chart from a list

```
In [14]: plt.bar(x,y)
```

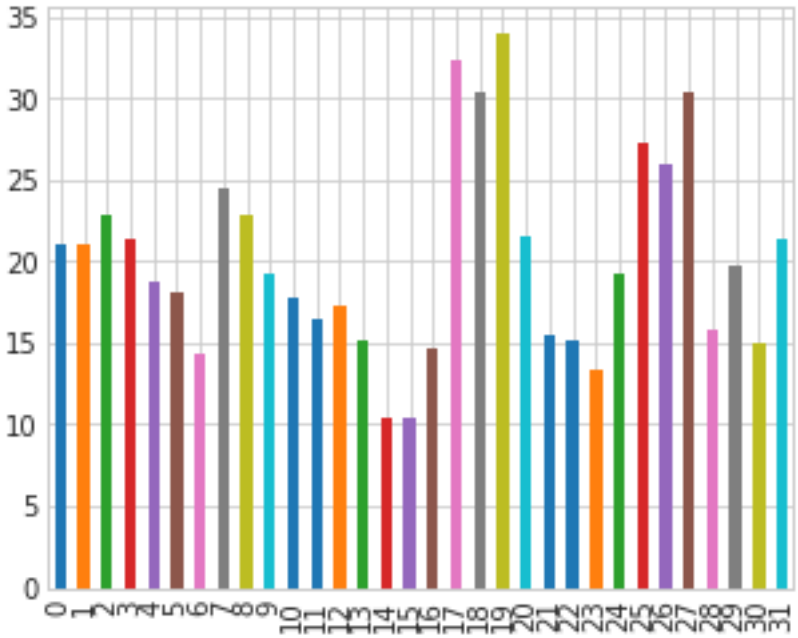
```
Out[14]: <Container object of 9 artists>
```



2.0.6 Creating bar chart from a pandas objects

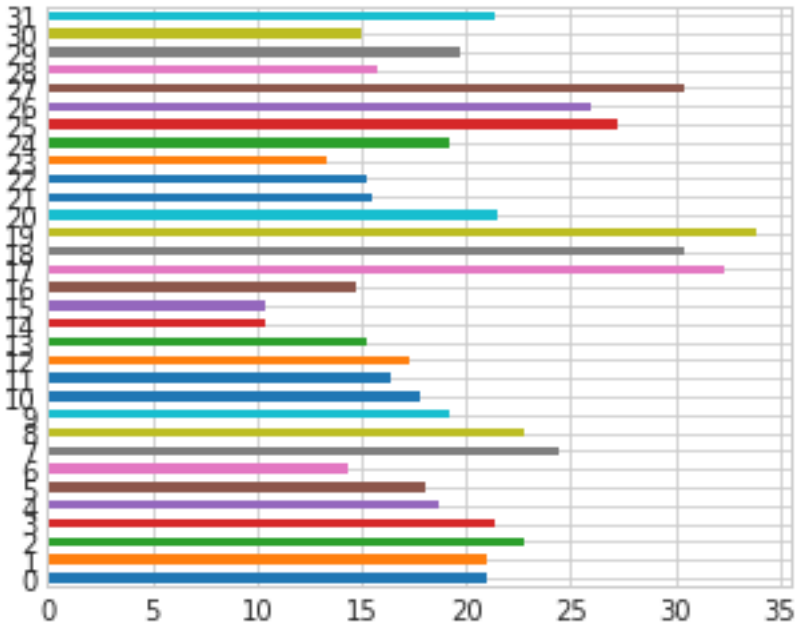
```
In [15]: mpg.plot(kind = 'bar')
```

```
Out[15]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda4c654940>
```



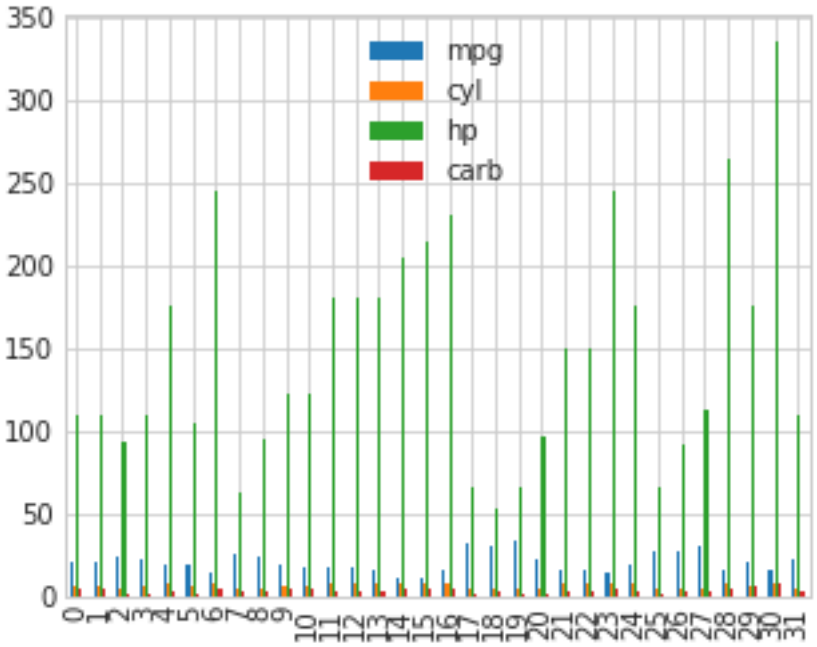
In [16]: mpg.plot(kind = 'barh')

Out[16]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda14530278>



In [17]: df.plot(kind = 'bar')

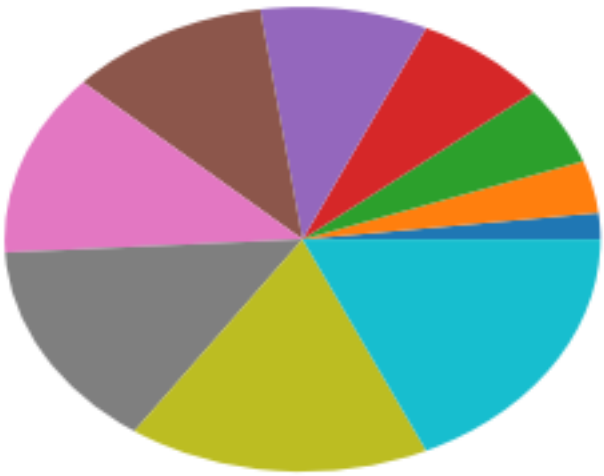
Out[17]: <matplotlib.axes._subplots.AxesSubplot at 0x7fda1441bcf8>



2.0.7 Creating a pie chart

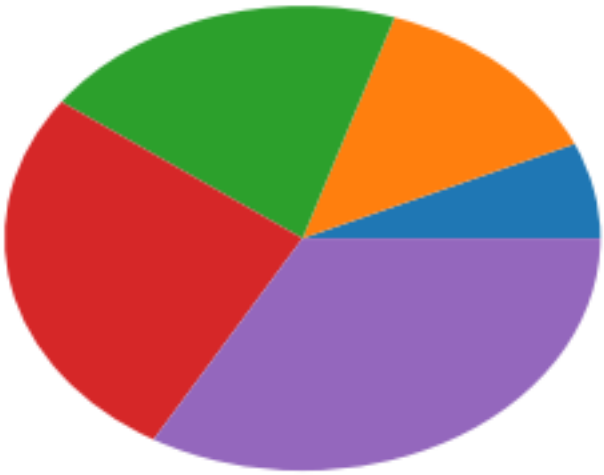
```
In [18]: x = [1,2,3,4,5,6,7,8,9,10]
         plt.pie(x)
```

```
Out[18]: ([<matplotlib.patches.Wedge at 0x7fda141deef0>,
           <matplotlib.patches.Wedge at 0x7fda141ef400>,
           <matplotlib.patches.Wedge at 0x7fda141ef940>,
           <matplotlib.patches.Wedge at 0x7fda141efe80>,
           <matplotlib.patches.Wedge at 0x7fda141f6400>,
           <matplotlib.patches.Wedge at 0x7fda141f6940>,
           <matplotlib.patches.Wedge at 0x7fda141f6e80>,
           <matplotlib.patches.Wedge at 0x7fda14180400>,
           <matplotlib.patches.Wedge at 0x7fda14180940>,
           <matplotlib.patches.Wedge at 0x7fda14180e80>],
          [Text(1.09821,0.0627977,''),
           Text(1.07141,0.249146,''),
           Text(0.957821,0.540906,''),
           Text(0.671713,0.871092,''),
           Text(0.156546,1.0888,''),
           Text(-0.513334,0.972876,''),
           Text(-1.03603,0.369654,''),
           Text(-0.957821,-0.540906,''),
           Text(-0.0941326,-1.09596,''),
           Text(0.925379,-0.594705,'')])
```



```
In [19]: y = [1,2,3,4,5]
         plt.pie(y)
```

```
Out[19]: ([<matplotlib.patches.Wedge at 0x7fda14144f28>,
           <matplotlib.patches.Wedge at 0x7fda1414d438>,
           <matplotlib.patches.Wedge at 0x7fda1414d978>,
           <matplotlib.patches.Wedge at 0x7fda1414deb8>,
           <matplotlib.patches.Wedge at 0x7fda14157438>],
          [Text(1.07596,0.228703,''),
           Text(0.736044,0.817459,''),
           Text(-0.339919,1.04616,''),
           Text(-1.07596,-0.228703,''),
           Text(0.55,-0.952628,'')])
```



2.0.8 Saving a Plot

```
In [20]: plt.savefig('pie_chart.png')
         plt.show()
```

```
<matplotlib.figure.Figure at 0x7fda4c6b4e48>
```

```
In [21]: %pwd
```

```
Out[21]: '/home/akkal/Documents/DS_PYTHON/TheEngineeringWorld'
```

df.plot(kind = 'bar') plt.savefig('bar_chart.png') plt.show()

```
In [22]: df.plot(kind = 'bar')
         plt.savefig('bar_chart.png')
         plt.show()
```