The Engineering World #DataScience 11

May 31, 2018

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

1 HISTOGRAMS, BOX PLOTS AND SCATTER PLOTS

1.0.1 Constructing Histogram, Box plots and Scatter plots

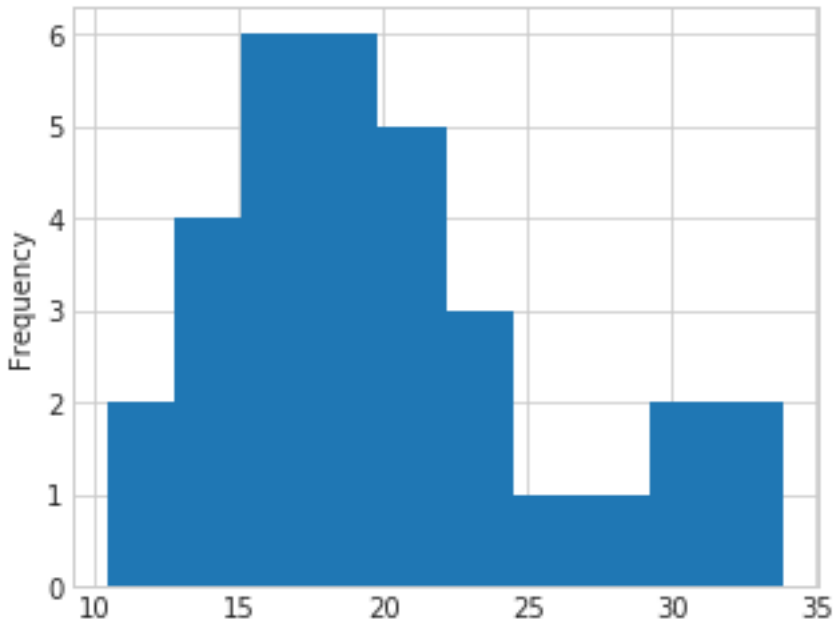
In [1]:
```
import numpy as np
from numpy.random import randn

import pandas as pd
from pandas import Series, DataFrame

import matplotlib.pyplot as plt
from matplotlib import rcParams

import seaborn as sb
```

In [2]:
```
%matplotlib inline
rcParams ['figure.figsize'] = 5,4
sb.set_style ('whitegrid')
```

1.0.2 Eyballing dataset distributions with histograms

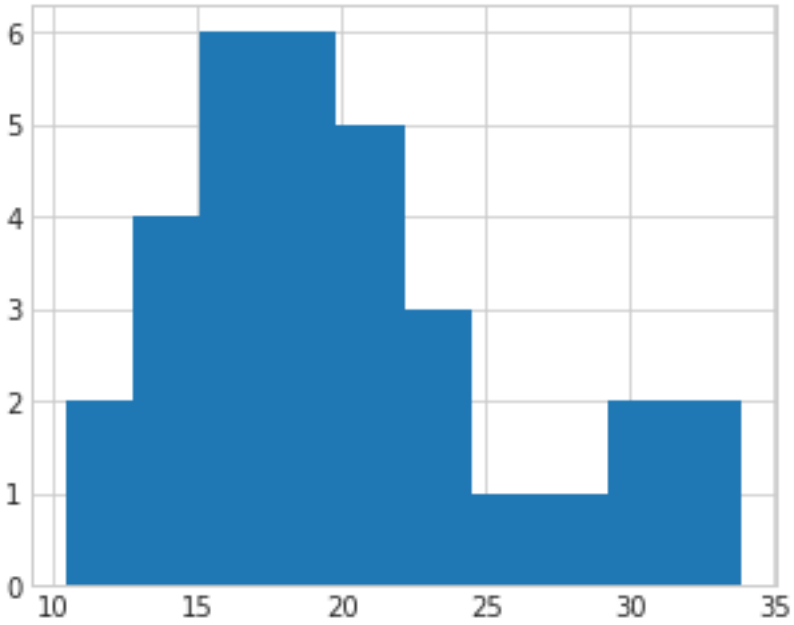
In [3]:
```
address = 'mtcars.csv'
cars = pd.read_csv(address)
cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'vs', 'am
cars.index = cars.car_names
mpg = cars['mpg']
mpg.plot(kind = 'hist')
```

Out[3]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fc6fae9bfd0>
```



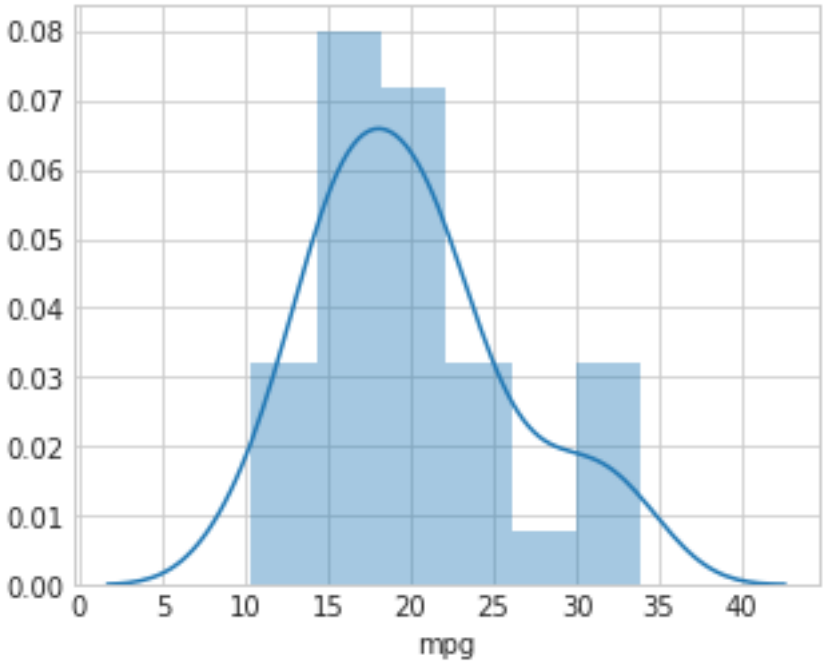
In [4]:
```
plt.hist(mpg)
plt.plot()
```

Out[4]: []



```
In [5]: sb.distplot(mpg) #shows the distribution line
```

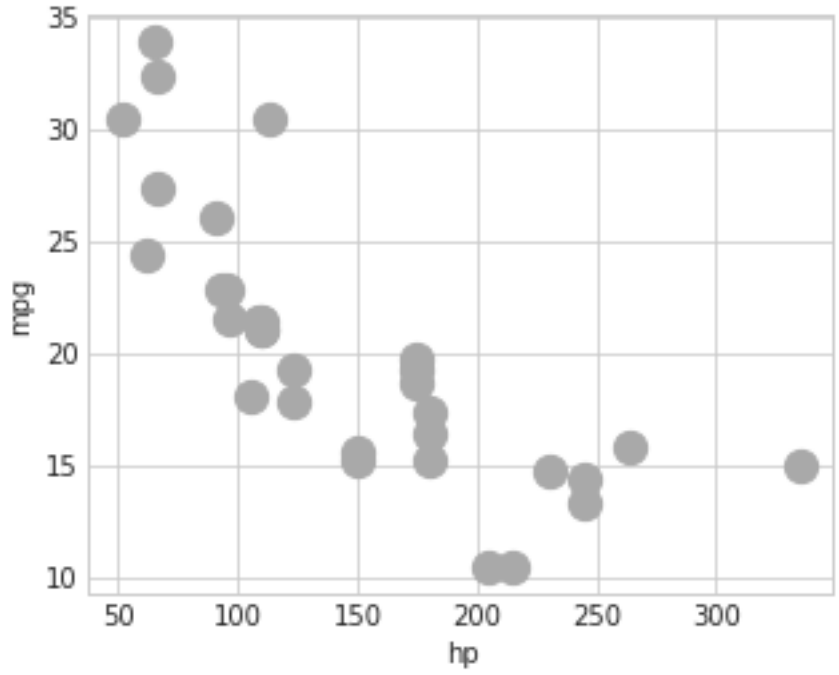
```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc6f0459470>
```



1.0.3 Seeing scatterplots in action

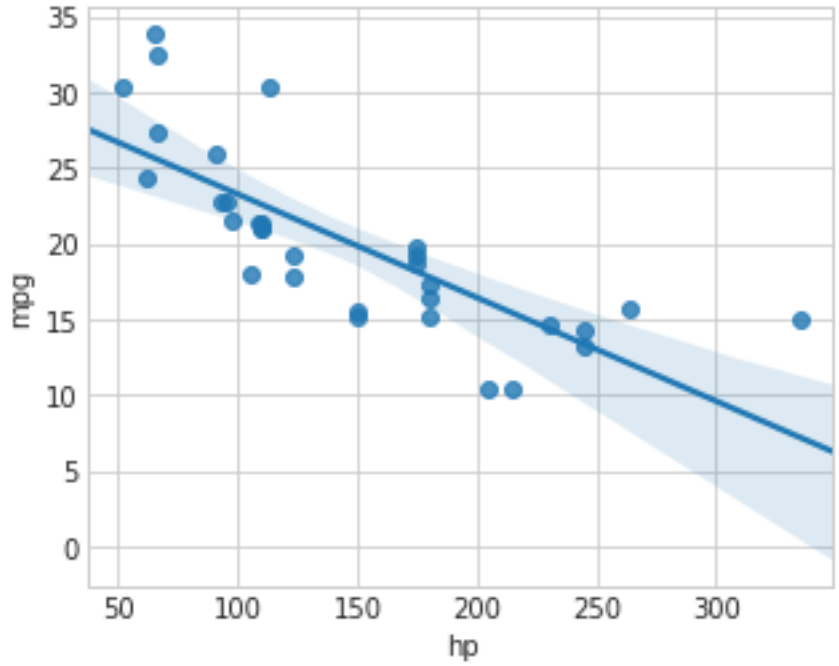
```
In [6]: cars.plot(kind='scatter', x = 'hp', y = 'mpg', c = ['darkgray'], s=150)
```

```
Out[6]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc6f03b5c50>
```



In [7]:
```
sb.regplot(x = 'hp', y = 'mpg', data = cars, scatter = True) #draw line in scatter point
```

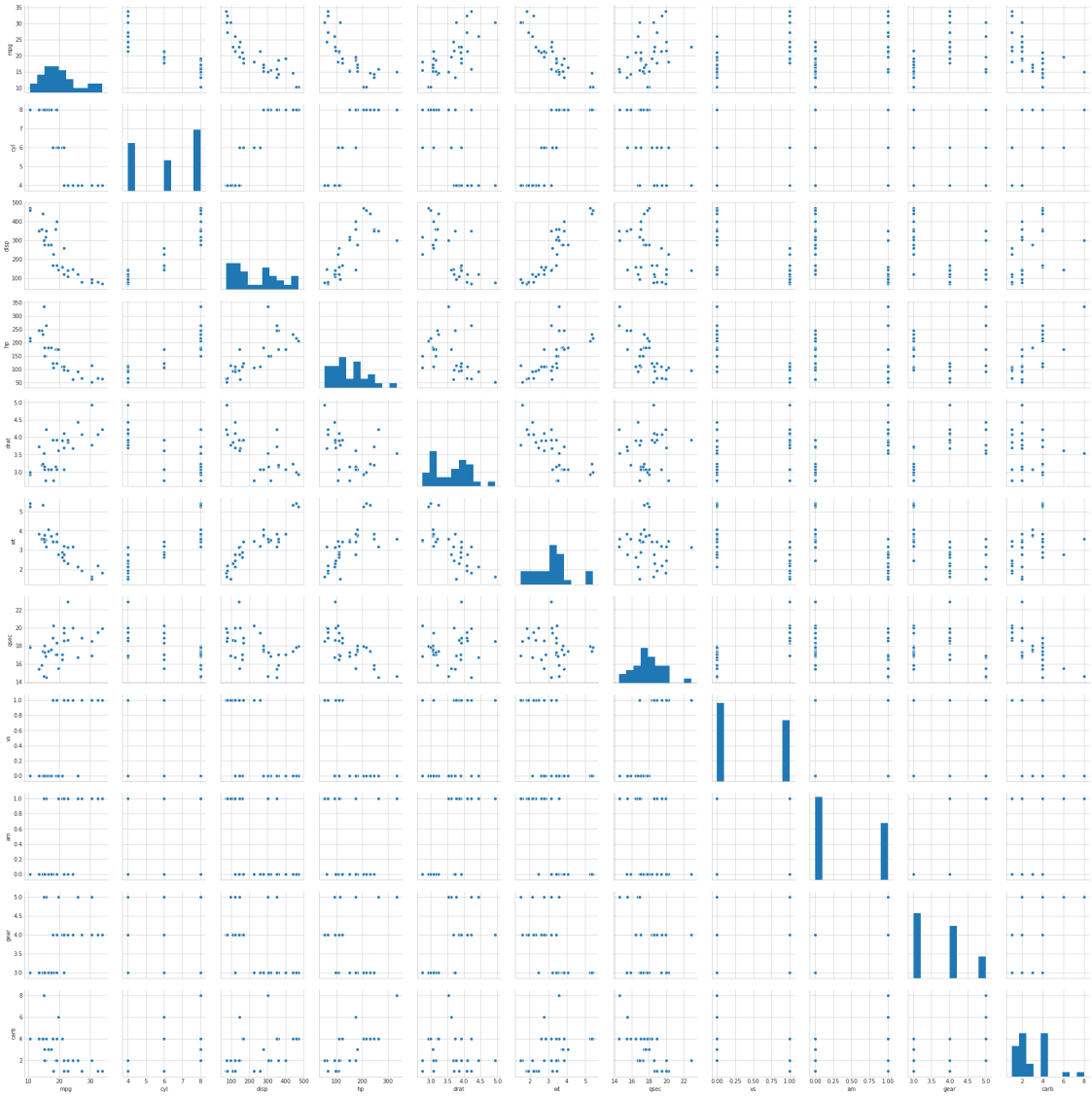
Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc6f0364cf8>



1.0.4 Generating a scatter plot matrix

In [8]:
```
sb.pairplot(cars)
```

Out[8]:
```
<seaborn.axisgrid.PairGrid at 0x7fc6faf3e550>
```



In [9]:
```
cars_df = pd.DataFrame((cars.ix[:,(1,3,4,6)].values), columns = ['mpg', 'disp', 'hp', 'w
cars_target = cars.ix[:,9].values
target_names = [0,1]

cars_df['group'] = pd.Series(cars_target, dtype = "category")
sb.pairplot(cars_df, hue = 'group', palette = 'hls')
```

```
/home/akkal/anaconda3/lib/python3.6/site-packages/ipykernel_launcher.py:1: DeprecationWarning:
.ix is deprecated. Please use
```

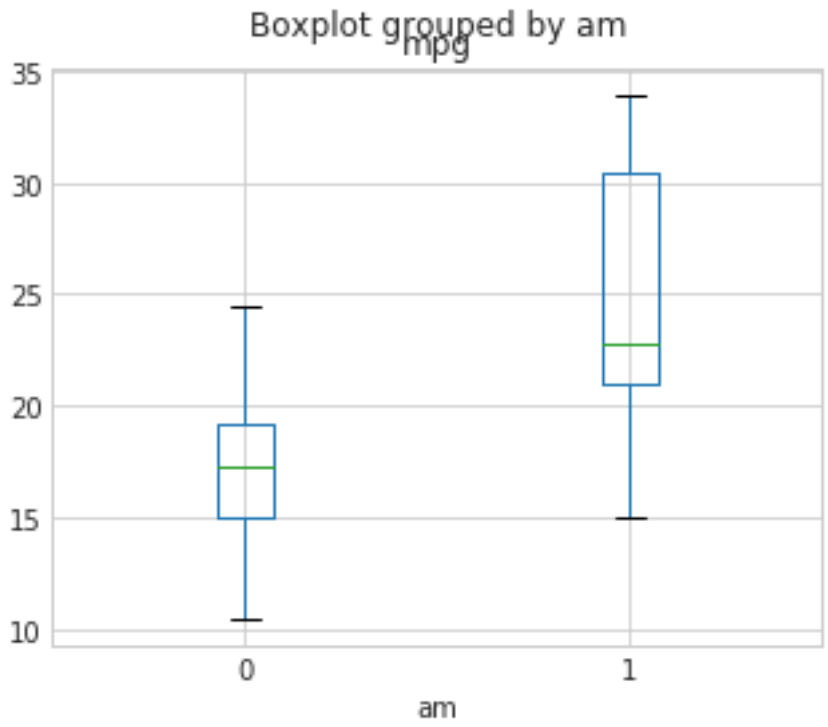
```
.loc for label based indexing or
.iloc for positional indexing

See the documentation here:
http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated
  """Entry point for launching an IPython kernel.
```

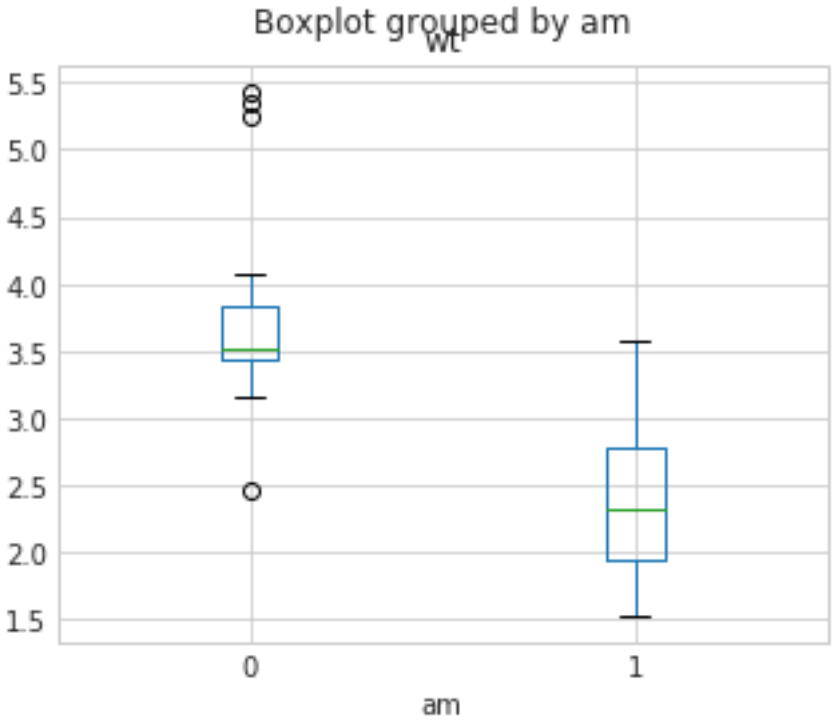
Out[9]: <seaborn.axisgrid.PairGrid at 0x7fc6ecd2fba8>

1.0.5 Building Box Plots

```
In [10]: cars.boxplot(column = 'mpg', by = 'am')
         cars.boxplot(column = 'wt', by = 'am')
```

Out[10]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc6e48ee668>


Boxplot grouped by am
mpg
35
30
25
20
15
10
0
1
am


Boxplot grouped by am
wt
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
0
1
am

```
In [11]: sb.boxplot(x = 'am', y = 'mpg', data = cars, palette = 'hls')
```

```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x7fc6e52a0978>
```