

In [4]:

```

import numpy as np
import pandas as pd
from pandas import Series, DataFrame
import scipy
from scipy.stats import spearmanr

from pylab import rcParams
import seaborn as sb
import matplotlib.pyplot as plt

import sklearn
from sklearn.preprocessing import scale
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn import preprocessing
%matplotlib inline
rcParams['figure.figsize'] = 5,4
sb.set_style('whitegrid')
address = '/home/sanyam/Desktop/dataset/mtcars.csv'
cars=pd.read_csv(address)
cars.columns = ['cars_names', 'mpg', 'cyl', 'disp', 'hp', 'qsec', 'drat', 'wt', 'vs', 'am', '
cars.head()

```

Out[4]:

	cars_names	mpg	cyl	disp	hp	qsec	drat	wt	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 [5]:

```
cars_data = cars.ix[:,(5,11)].values
cars_data_names = ['drat', 'carb']
```

```
y=cars.ix[:,9].values
sb.regplot(x='drat',y='carb', data = cars,scatter = True)
```

/usr/local/lib/python3.5/dist-packages/ipykernel_launcher.py:1: DeprecationWarning:

.ix is deprecated. Please use
.loc for label based indexing or
.iloc for positional indexing

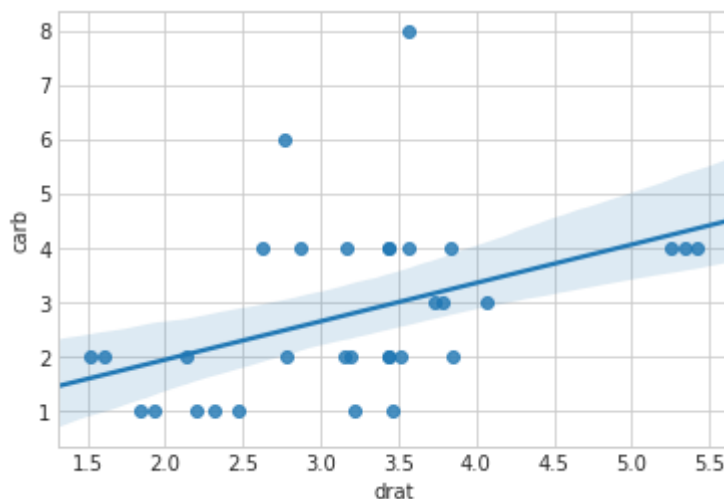
See the documentation here:

<http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated> (<http://pandas.pydata.org/pandas-docs/stable/indexing.html#ix-indexer-is-deprecated>)

"""Entry point for launching an IPython kernel.

Out[5]:

<matplotlib.axes._subplots.AxesSubplot at 0x7ff677a3d7f0>



In [6]:

```
drat = cars['drat']
carb = cars['carb']
spearmanr_coefficient, p_value=spearmanr(drat, carb)
print("Spearmanr Rank correlation coefficient",0.3 *(spearmanr_coefficient))
```

Spearmanr Rank correlation coefficient 0.149943614671

In [7]:

```
cars.isnull().sum()
```

Out[7]:

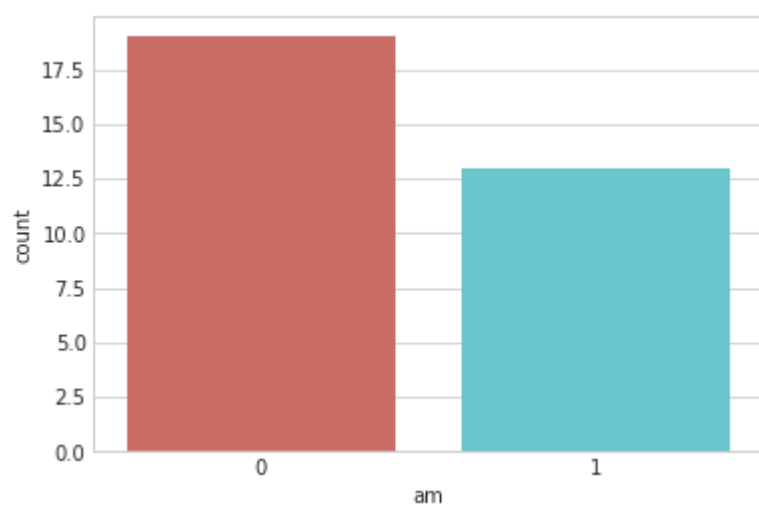
```
cars_names    0
mpg           0
cyl           0
disp          0
hp            0
qsec          0
drat          0
wt            0
vs            0
am            0
gear          0
carb         0
dtype: int64
```

In [11]:

```
sb.countplot(x='am', data=cars, palette = 'hls')
```

Out[11]:

<matplotlib.axes._subplots.AxesSubplot at 0x7ff66797e0b8>



In [12]:

cars.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 32 entries, 0 to 31
Data columns (total 12 columns):
cars_names      32 non-null object
mpg             32 non-null float64
cyl             32 non-null int64
disp           32 non-null float64
hp             32 non-null int64
qsec          32 non-null float64
drat          32 non-null float64
wt            32 non-null float64
vs            32 non-null int64
am            32 non-null int64
gear          32 non-null int64
carb         32 non-null int64
dtypes: float64(5), int64(6), object(1)
memory usage: 3.1+ KB
```

In [15]:

```
X= cars_data
LogReg=LogisticRegression()
LogReg.fit(X,y)
print(LogReg.score(X, y))
```

0.8125

In [16]:

```
y_pred = LogReg.predict(X)
from sklearn.metrics import classification_report
print(classification_report(y, y_pred))
```

	precision	recall	f1-score	support
0	0.76	1.00	0.86	19
1	1.00	0.54	0.70	13
avg / total	0.86	0.81	0.80	32

In []: