4.6.1 The Stock Market Data

May 18, 2018

0.1 The Stock Market Data - Logistic Regression

In [31]:
```
# conventional way to import pandas
import pandas as pd
# conventional way to import seaborn
import seaborn as sns
# conventional way to import numpy
import numpy as np

from sklearn import metrics
import matplotlib.pyplot as plt

data = pd.read_csv("https://raw.github.com/vincentarelbundock/Rdatasets/master/csv/ISL

data.head()
```

Out[31]:

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today	Direction
1	2001	0.381	-0.192	-2.624	-1.055	5.010	1.1913	0.959	Up
2	2001	0.959	0.381	-0.192	-2.624	-1.055	1.2965	1.032	Up
3	2001	1.032	0.959	0.381	-0.192	-2.624	1.4112	-0.623	Down
4	2001	-0.623	1.032	0.959	0.381	-0.192	1.2760	0.614	Up
5	2001	0.614	-0.623	1.032	0.959	0.381	1.2057	0.213	Up

In [32]:
```
data.describe()
```

Out[32]:

	Year	Lag1	Lag2	Lag3	Lag4
count	1250.000000	1250.000000	1250.000000	1250.000000	1250.000000
mean	2003.016000	0.003834	0.003919	0.001716	0.001636
std	1.409018	1.136299	1.136280	1.138703	1.138774
min	2001.000000	-4.922000	-4.922000	-4.922000	-4.922000
25%	2002.000000	-0.639500	-0.639500	-0.640000	-0.640000
50%	2003.000000	0.039000	0.039000	0.038500	0.038500
75%	2004.000000	0.596750	0.596750	0.596750	0.596750
max	2005.000000	5.733000	5.733000	5.733000	5.733000

	Lag5	Volume	Today
count	1250.00000	1250.000000	1250.000000
mean	0.00561	1.478305	0.003138

```
std       1.14755     0.360357     1.136334
min      -4.92200     0.356070    -4.922000
25%      -0.64000     1.257400    -0.639500
50%       0.03850     1.422950     0.038500
75%       0.59700     1.641675     0.596750
max       5.73300     3.152470     5.733000
```

As we can see the data is pretty uncorrlated exept Volume and year. As seen.

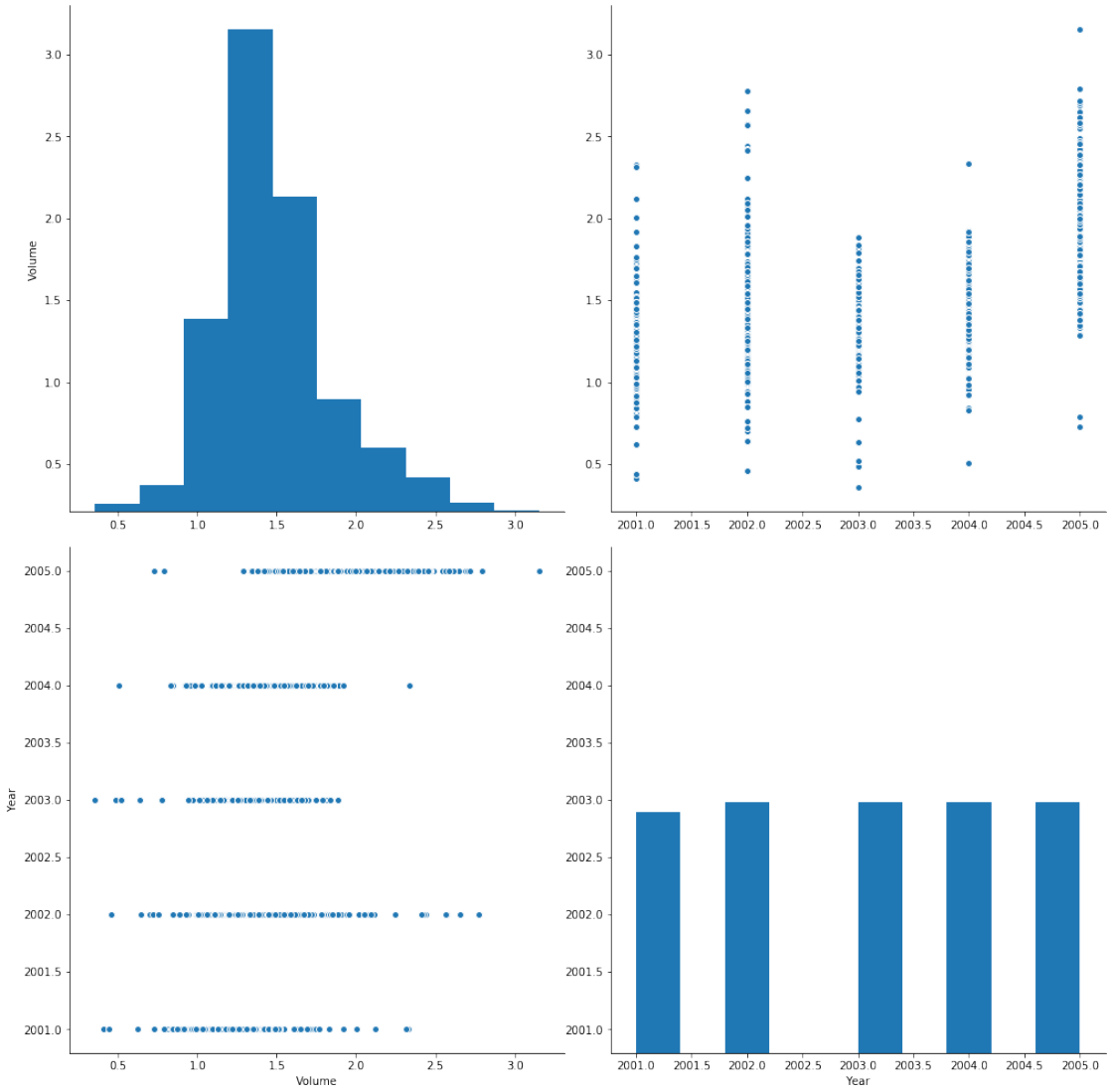
```
In [33]: data.corr()
```

```
Out[33]:              Year      Lag1      Lag2      Lag3      Lag4      Lag5    Volume  \
         Year    1.000000  0.029700  0.030596  0.033195  0.035689  0.029788  0.539006
         Lag1    0.029700  1.000000 -0.026294 -0.010803 -0.002986 -0.005675  0.040910
         Lag2    0.030596 -0.026294  1.000000 -0.025897 -0.010854 -0.003558 -0.043383
         Lag3    0.033195 -0.010803 -0.025897  1.000000 -0.024051 -0.018808 -0.041824
         Lag4    0.035689 -0.002986 -0.010854 -0.024051  1.000000 -0.027084 -0.048414
         Lag5    0.029788 -0.005675 -0.003558 -0.018808 -0.027084  1.000000 -0.022002
         Volume  0.539006  0.040910 -0.043383 -0.041824 -0.048414 -0.022002  1.000000
         Today   0.030095 -0.026155 -0.010250 -0.002448 -0.006900 -0.034860  0.014592

                    Today
         Year    0.030095
         Lag1   -0.026155
         Lag2   -0.010250
         Lag3   -0.002448
         Lag4   -0.006900
         Lag5   -0.034860
         Volume  0.014592
         Today   1.000000
```

Plotting Years onto Volume

```
In [42]: sns.pairplot(data[['Volume','Year']],size=7)
         plt.show()
```


Volume
Year
Volume
Year