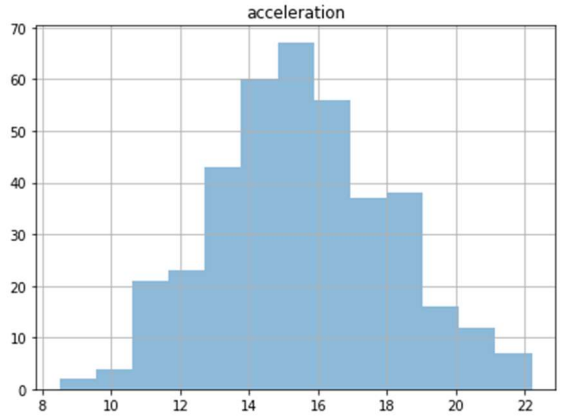
Numerical data

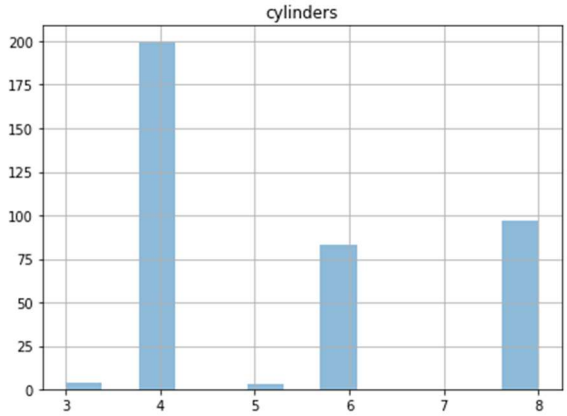
	mpg	cylinders	displacement	horsepower	weight	acceleration
count	379.000000	386.000000	386.000000	380.000000	386.000000	386.000000
mean	23.529551	5.432642	190.744819	103.157895	2957.603627	15.566321
std	7.569748	1.689347	100.392840	35.240692	832.882821	2.554315
min	9.000000	3.000000	68.000000	46.000000	1613.000000	8.500000
25%	17.600000	4.000000	105.000000	76.000000	2220.750000	14.000000
50%	23.000000	4.000000	146.000000	93.500000	2800.000000	15.500000
75%	29.000000	7.500000	260.000000	122.750000	3573.000000	17.075000
max	44.600000	8.000000	429.000000	210.000000	5140.000000	22.200000

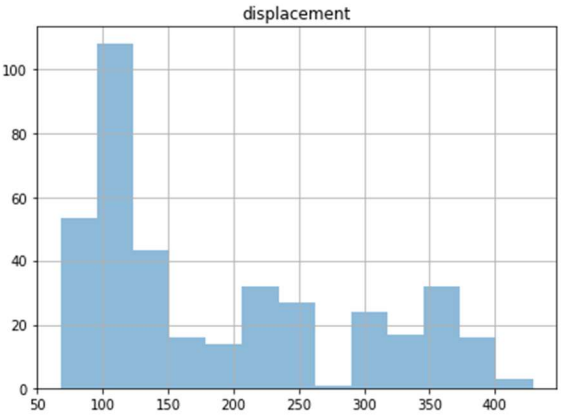
Categorical data

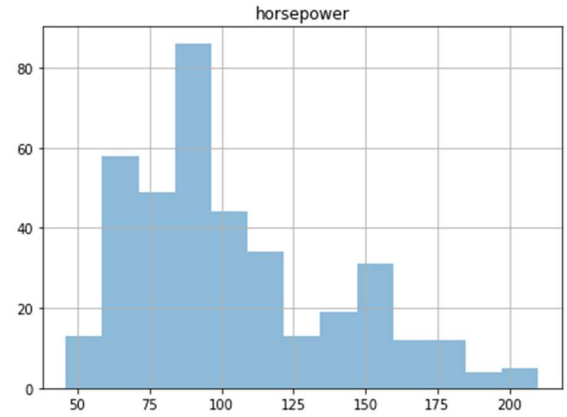
Origin

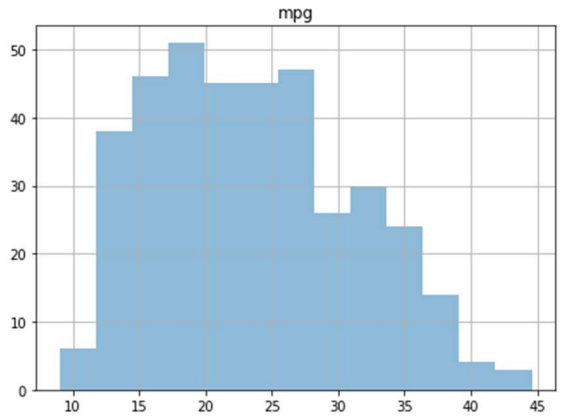
count 406.000000
mean 1.568966
std 0.797479
mode 1
mode freq 0.625616

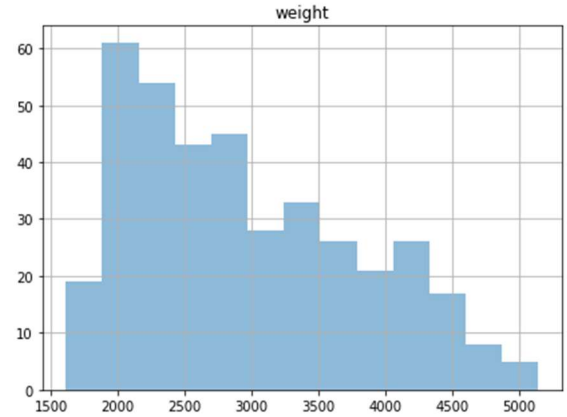

acceleration
70
60
50
40
30
20
10
0
8
10
12
14
16
18
20
22

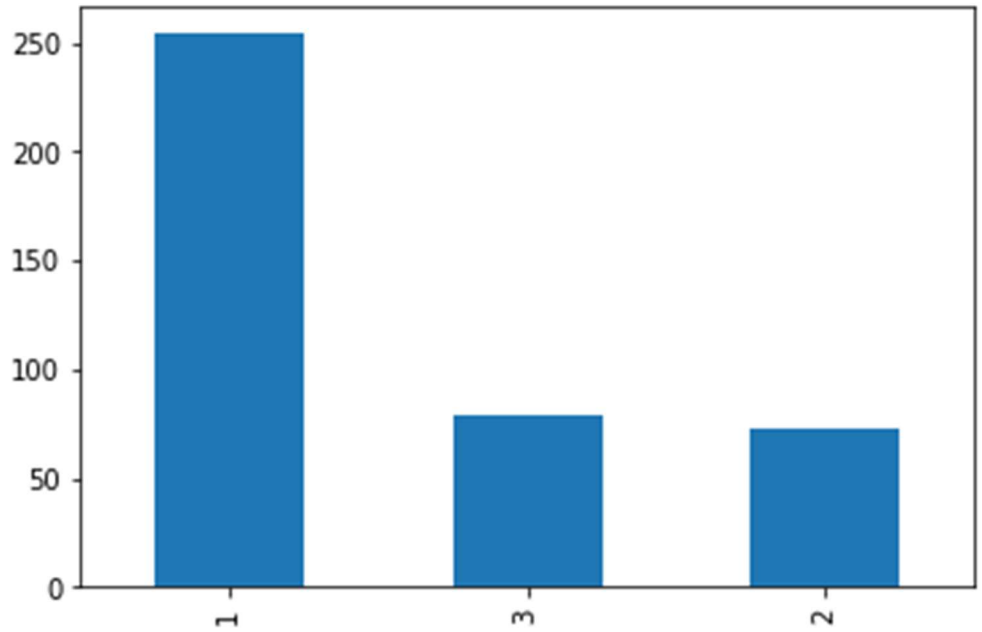

cylinders
200
175
150
125
100
75
50
25
0
3
4
5
6
7
8


displacement
100
80
60
40
20
0
50
100
150
200
250
300
350
400


horsepower
80
60
40
20
0
50
75
100
125
150
175
200


mpg
50
40
30
20
10
0
10
15
20
25
30
35
40
45


weight
60
50
40
30
20
10
0
1500
2000
2500
3000
3500
4000
4500
5000



Solutions to handling outliers and imputation

For outlier removal, I used the interquartile range and then removed all records which fall outside of the IQR * 1.5. For imputing the missing values, I used K-nearest-neighbour with k=3. This means that I used the other attributes of the record to predict the value of the missing value.