In [18]:

```
import pandas as pd
import matplotlib.pyplot as plt
```

In [4]:

```
df=pd.read_csv('Datasets/auto-mpg.csv')

```

In [11]:

```
df.dtypes
```

Out[11]:

```
mpg             float64
cylinders         int64
displacement    float64
horsepower        int64
weight            int64
acceleration    float64
model year        int64
origin            int64
car name         object
dtype: object
```

In [9]:

```
df['horsepower']=df["horsepower"].replace('?',100)
```

In [10]:

```
df['horsepower']=df['horsepower'].astype('int64')
```

In [12]:

```
df.nunique()
```

Out[12]:

```
mpg             129
cylinders         5
displacement     82
horsepower       93
weight          351
acceleration     95
model year       13
origin            3
car name        305
dtype: int64
```

In [17]:

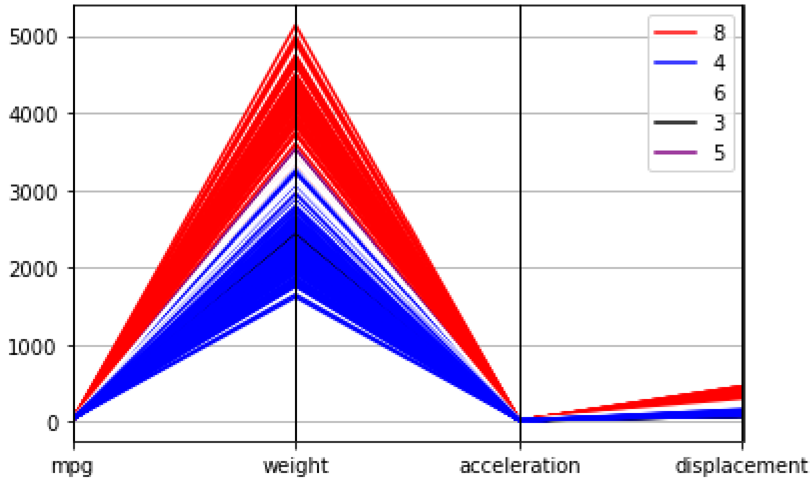
```
df.cylinders.unique()
```

Out[17]:

```
array([8, 4, 6, 3, 5], dtype=int64)
```

In [27]:

```
pd.plotting.parallel_coordinates(df,'cylinders',
                                 cols=['mpg','weight','acceleration','displacement'],
                                 color=['red','blue','white','black','purple'])
plt.show()
```



In [29]:

```
df=pd.read_csv('Datasets/chile.csv')
print(df)
```

```
     region  population sex   age education   income  statusquo vote
0         N      175000   M  65.0         P  35000.0    1.00820    Y
1         N      175000   M  29.0        PS   7500.0   -1.29617    N
2         N      175000   F  38.0         P  15000.0    1.23072    Y
3         N      175000   F  49.0         P  35000.0   -1.03163    N
4         N      175000   F  23.0         S  35000.0   -1.10496    N
...     ...         ...  ..   ...       ...      ...        ...  ...
2695      M       15000   M  42.0         P  15000.0   -1.26247    N
2696      M       15000   F  28.0         P  15000.0    1.32950    Y
2697      M       15000   F  44.0         P  75000.0    1.42045    Y
2698      M       15000   M  21.0         S  75000.0    0.18315  NaN
2699      M       15000   M  20.0        PS  35000.0    1.38179    Y

[2700 rows x 8 columns]
```

In [30]:

```
df.head()
```

Out[30]:

	region	population	sex	age	education	income	statusquo	vote
0	N	175000	M	65.0	P	35000.0	1.00820	Y
1	N	175000	M	29.0	PS	7500.0	-1.29617	N
2	N	175000	F	38.0	P	15000.0	1.23072	Y
3	N	175000	F	49.0	P	35000.0	-1.03163	N
4	N	175000	F	23.0	S	35000.0	-1.10496	N

In [32]:

```
df.shape
```

Out[32]:

```
(2700, 8)
```

In [35]:

```
df.duplicated().sum()
```

Out[35]:

```
9
```

```
In [37]: df=df.drop_duplicates()
```

```
In [38]: df.duplicated().sum()

Out[38]: 0
```

```
In [39]: df.shape

Out[39]: (2691, 8)
```

```
In [41]: df.isna().sum()

Out[41]: region          0
         population      0
         sex             0
         age             1
         education      11
         income         98
         statusquo      17
         vote          168
         dtype: int64
```

```
In [42]: df=df.dropna()
```

```
In [43]: df.isna().sum()

Out[43]: region        0
         population    0
         sex           0
         age           0
         education     0
         income        0
         statusquo     0
         vote          0
         dtype: int64
```

```
In [52]: pd.crosstab(df['sex'],df['education'],colnames=['ed'],rownames=['s'],margins=True,normalize=True)
```

```
Out[52]:
 ed   P         PS        S         All
  s
  F   0.227498  0.074732  0.210570  0.512799
  M   0.185797  0.097027  0.204377  0.487201
All   0.413295  0.171759  0.414946  1.000000
```

```
In [49]: pd.crosstab(df['sex'],df['education'])
```

```
Out[49]:
education  P    PS   S
      sex
        F  551  181  510
        M  450  235  495
```

```
In [55]: pd.crosstab(df['sex'],df['education'], values=df['income'],aggfunc={min,max})
```

```
Out[55]:
           max                           min
education  P         PS        S         P       PS      S
      sex
        F  125000.0  200000.0  200000.0  2500.0  2500.0  2500.0
        M  200000.0  200000.0  200000.0  2500.0  7500.0  2500.0
```

```
In [57]: pd.crosstab([df['sex'],df['education']],[df['region'],df['vote']])
```

```
Out[57]:
         region   C               M              N               S               SA
         vote     A   N   U   Y   A  N  U   Y    A   N   U   Y   A   N   U   Y   A   N   U   Y
sex  education
  F          P    5  22  38  53   0  3  10  15   6  13  14  45   8  33  43  81  12  40  62  48
            PS    1  15   3   8   0  1   1   2   1  10   2  12   3  17   7  13   8  40  16  21
             S   17  37  35  35   0  2   3   6   3  17  10  26  12  37  34  42  21  60  61  52
  M          P    6  48  34  32   1  5   3   6   2  15  11  27   5  35  36  69   4  47  31  33
            PS    2  33   3  11   0  0   0   2   5  19   4   8   5  32   3  15   5  50   7  31
             S   13  54  17  26   1  6   2   6  12  24   5  14   6  55  17  47  13  88  39  50
```

```
In [ ]:
```