```
Chapter 5 - Basic Math and Statistics

Segement 4 - Summarizing categorical data using pandas
```

In [1]:
```
import numpy as np
import pandas as pd
```

```
The basics
```

In [2]:
```
address = 'C:/Users/danal/Desktop/ExerciseFiles/Data/mtcars.csv'
cars = pd.read_csv(address)

cars.columns = ['car_names', 'mpg', 'cyl', 'disp', 'hp', 'drat', 'wt', 'qsec', 'v
cars.index = cars.car_names
cars.head(15)
```

Out[2]:

car_names	car_names	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4

In [4]:
```
carb = cars.carb
carb.value_counts()
```

Out[4]:
```
4    10
2    10
1     7
3     3
8     1
6     1
Name: carb, dtype: int64
```

In [5]:
```
cars_cat = cars[['cyl','vs','am','gear','carb']]
cars_cat.head()
```

Out[5]:

car_names	cyl	vs	am	gear	carb
Mazda RX4	6	0	1	4	4
Mazda RX4 Wag	6	0	1	4	4
Datsun 710	4	1	1	4	1
Hornet 4 Drive	6	1	0	3	1
Hornet Sportabout	8	0	0	3	2

In [6]:
```
gears_group = cars_cat.groupby('gear')
gears_group.describe()
```

Out[6]:

	cyl								vs		...	am		car
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	cou
gear														
3	15.0	7.466667	1.187234	4.0	8.0	8.0	8.0	8.0	15.0	0.200000	...	0.0	0.0	1
4	12.0	4.666667	0.984732	4.0	4.0	4.0	6.0	6.0	12.0	0.833333	...	1.0	1.0	1
5	5.0	6.000000	2.000000	4.0	4.0	6.0	8.0	8.0	5.0	0.200000	...	1.0	1.0	

3 rows × 32 columns

Transforming variables to categorical data type

In [7]:
```
cars['group'] = pd.Series(cars.gear, dtype="category")
cars['group'].dtypes
```

Out[7]:
```
CategoricalDtype(categories=[3, 4, 5], ordered=False)
```

In [9]:
```
cars['group'].value_counts()
```

Out[9]:
```
3    15
4    12
5     5
Name: group, dtype: int64
```

```
Describing categorical data with crosstabs
```

In [11]:
```
pd.crosstab(cars['am'], cars['gear'])
```

Out[11]:

gear	3	4	5
am			
0	15	4	0
1	0	8	5