Chapter 5 - Basic Math and Statistics

Segment 4 - Summarizing categorical data using pandas

```
import numpy as np
import pandas as pd
```

The basics

```
address = 'C:/Users/Lillian/Desktop/ExerciseFiles/Data/mtcars.csv'
cars = pd.read_csv(address)

cars.columns = ['car_names','mpg','cyl','disp','hp','drat','wt','qsec','vs','am','gear','carb']
cars.index = cars.car_names
cars.head(15)
```

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

```
carb = cars.carb
carb.value_counts()
```

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

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

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

```
gears_group = cars_cat.groupby('gear')
gears_group.describe()
```

	cyl								vs		...	am		carb					
	count	mean	std	min	25%	50%	75%	max	count	mean	...	75%	max	count	mean	std	min	25%	50%
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	15.0	2.666667	1.175139	1.0	2.0	3.0
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	12.0	2.333333	1.302678	1.0	1.0	2.0
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	5.0	4.400000	2.607681	2.0	2.0	4.0

Transforming variables to categorical data type

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

```
cars['group'].dtypes
```

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

```
cars['group'].value_counts()
```

```
3    15
4    12
5     5
Name: group, dtype: int64
```

Describing categorical data with crosstabs

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

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

Colab paid products - Cancel contracts here