```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import io

from google.colab import files


uploaded = files.upload()
```

```
<IPython.core.display.HTML object>

Saving abalone.csv to abalone.csv
```

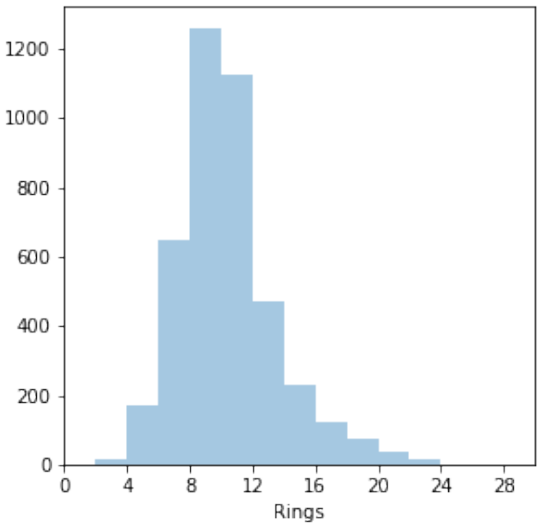
```
ds = pd.read_csv(io.BytesIO(uploaded['abalone.csv']))
ds.head()
```

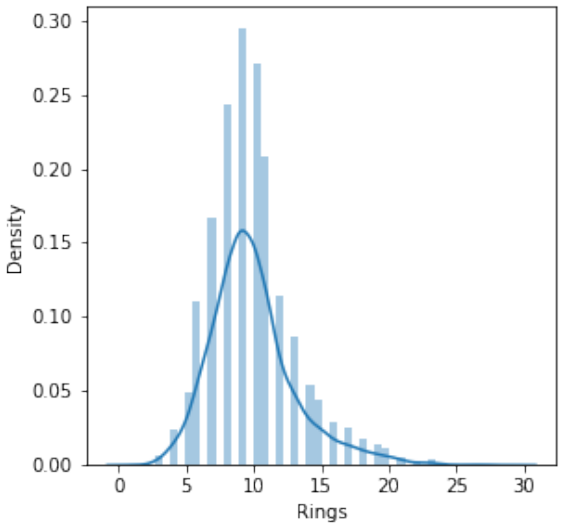
```
  Sex  Length  Diameter  Height  Whole weight  Shucked weight  Viscera
weight  \
0   M   0.455     0.365   0.095        0.5140          0.2245
0.1010
1   M   0.350     0.265   0.090        0.2255          0.0995
0.0485
2   F   0.530     0.420   0.135        0.6770          0.2565
0.1415
3   M   0.440     0.365   0.125        0.5160          0.2155
0.1140
4   I   0.330     0.255   0.080        0.2050          0.0895
0.0395

   Shell weight  Rings
0         0.150     15
1         0.070      7
2         0.210      9
3         0.155     10
4         0.055      7
```

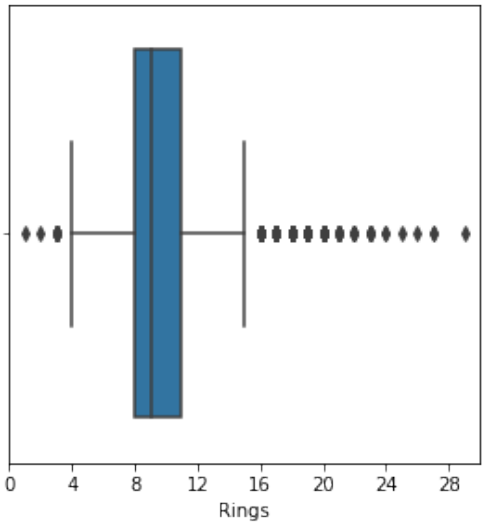
```
#Rings
rows = 2
cols = 2
i = 0

plt.figure(figsize=(cols * 5, rows * 5))

i += 1
plt.subplot(rows, cols, i)
plt.xticks(range(0, 31, 4))
plt.xlim(0, 30)
_ = sns.distplot(ds['Rings'], kde=False, bins=range(0, 31, 2))
```

```
i += 1
plt.subplot(rows, cols, i)
_ = sns.distplot(ds['Rings'])
i += 1
plt.subplot(rows, cols, i)
plt.xticks(range(0, 31, 4))
plt.xlim(0, 30)
_ = sns.boxplot(ds['Rings'])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```







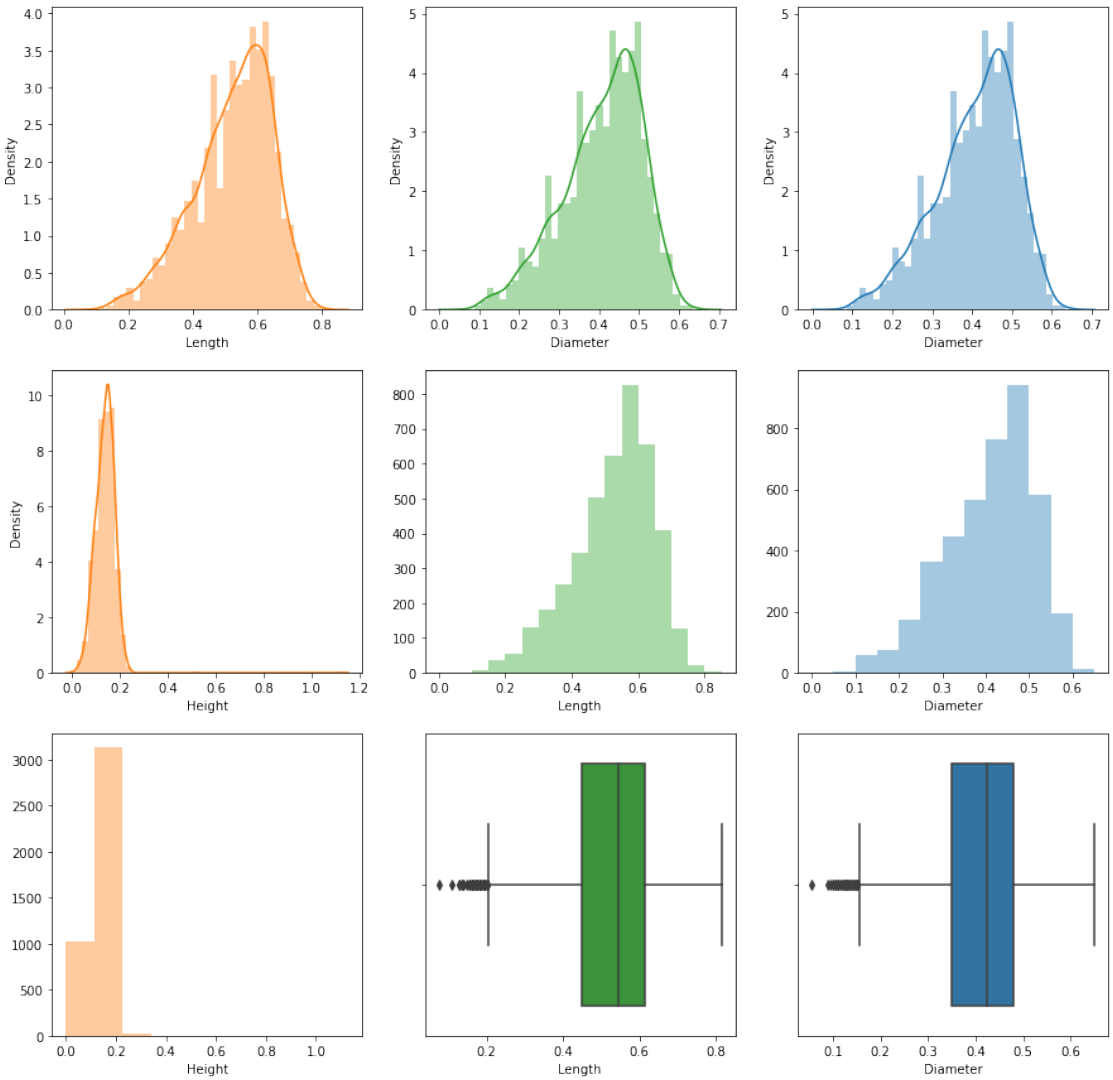

```
#Size
plt.figure(figsize=(15, 15))

colors = sns.color_palette()

lines = 3
rows = 3
i = 0

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Length'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Diameter'], color=colors[i % 3])
```

```
i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Diameter'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Height'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Length'], kde=False, bins=np.arange(0.0, 0.9,
0.05), color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Diameter'], kde=False, bins=np.arange(0.0, 0.7,
0.05), color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.distplot(ds['Height'], kde=False, bins=10, color=colors[i %
3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(ds['Length'], color=sns.color_palette()[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(ds['Diameter'], color=colors[i % 3])

i += 1
plt.subplot(lines, rows, i)
_ = sns.boxplot(ds['Height'], color=colors[i % 3])
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
```

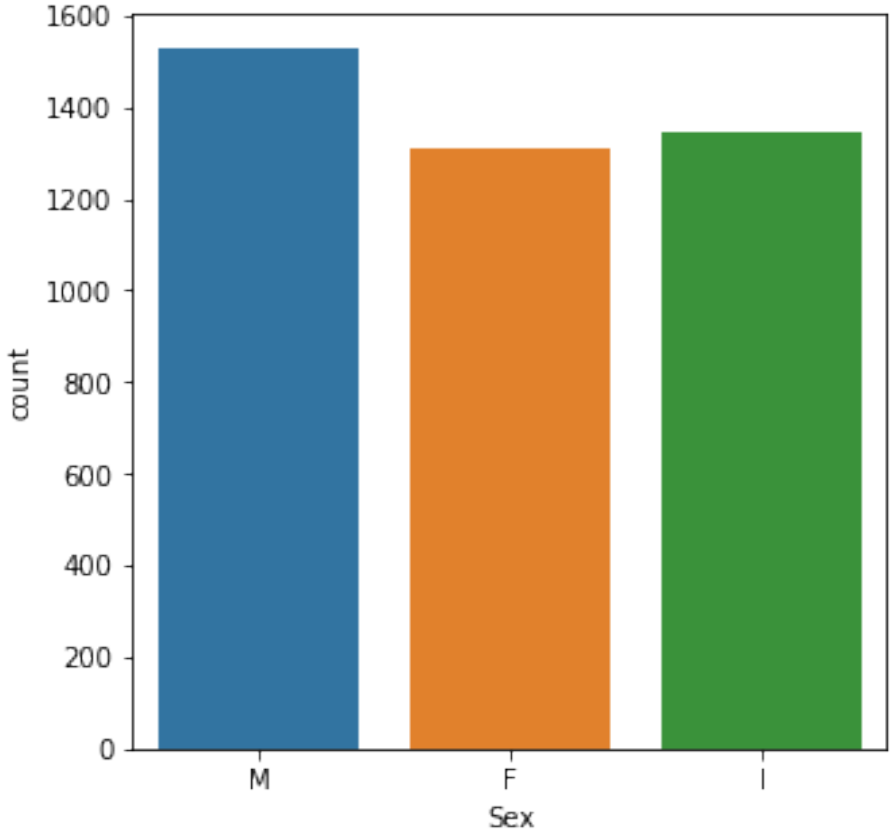
```
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning

---------------------------------------------------------------------
-----
ValueError                                Traceback (most recent call
last)
<ipython-input-19-69d3aec46ad8> in <module>
     45
     46 i += 1
---> 47 plt.subplot(lines, rows, i)
     48 _ = sns.boxplot(ds['Height'], color=colors[i % 3])

/usr/local/lib/python3.7/dist-packages/matplotlib/pyplot.py in
subplot(*args, **kwargs)
   1028
   1029     fig = gcf()
-> 1030     a = fig.add_subplot(*args, **kwargs)
   1031     bbox = a.bbox
   1032     byebye = []

/usr/local/lib/python3.7/dist-packages/matplotlib/figure.py in
add_subplot(self, *args, **kwargs)
   1417                     self._axstack.remove(ax)
   1418
```

```
-> 1419             a = subplot_class_factory(projection_class)(self,
*args, **kwargs)
   1420
   1421         return self._add_axes_internal(key, a)

/usr/local/lib/python3.7/dist-packages/matplotlib/axes/_subplots.py in
__init__(self, fig, *args, **kwargs)
     64                 if num < 1 or num > rows*cols:
     65                     raise ValueError(
---> 66                         f"num must be 1 <= num <= {rows*cols},
not {num}")
     67                 self._subplotspec = GridSpec(
     68                         rows, cols, figure=self.figure)
[int(num) - 1]

ValueError: num must be 1 <= num <= 9, not 10
```

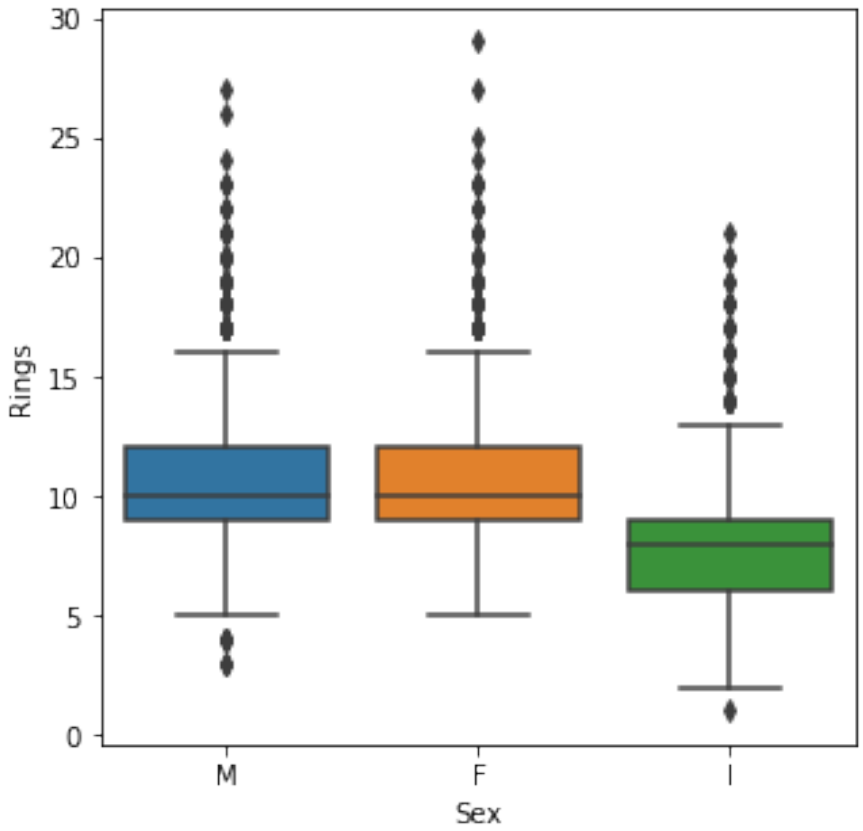


```
#Sex attribute
plt.figure(figsize=(5,5))
_ = sns.countplot(ds.Sex)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

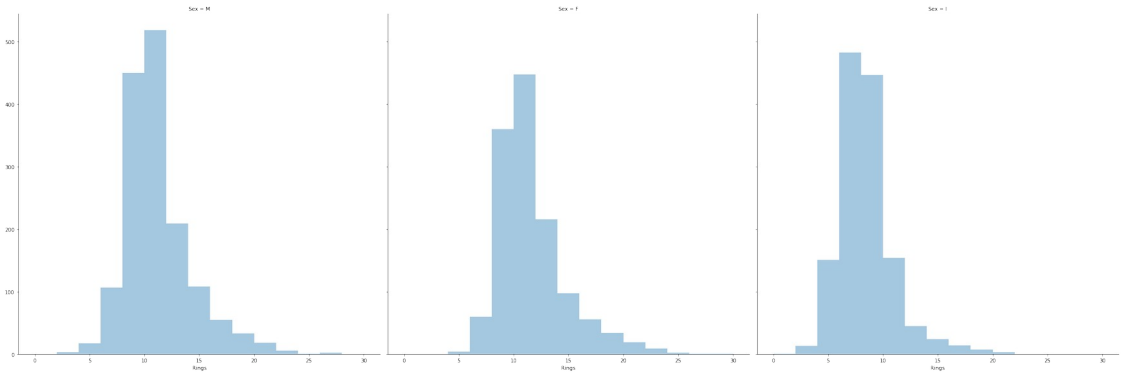



```
#sex influence number of rings
plt.figure(figsize=(5, 5))
_ = sns.boxplot(data=ds, x='Sex', y='Rings')
```

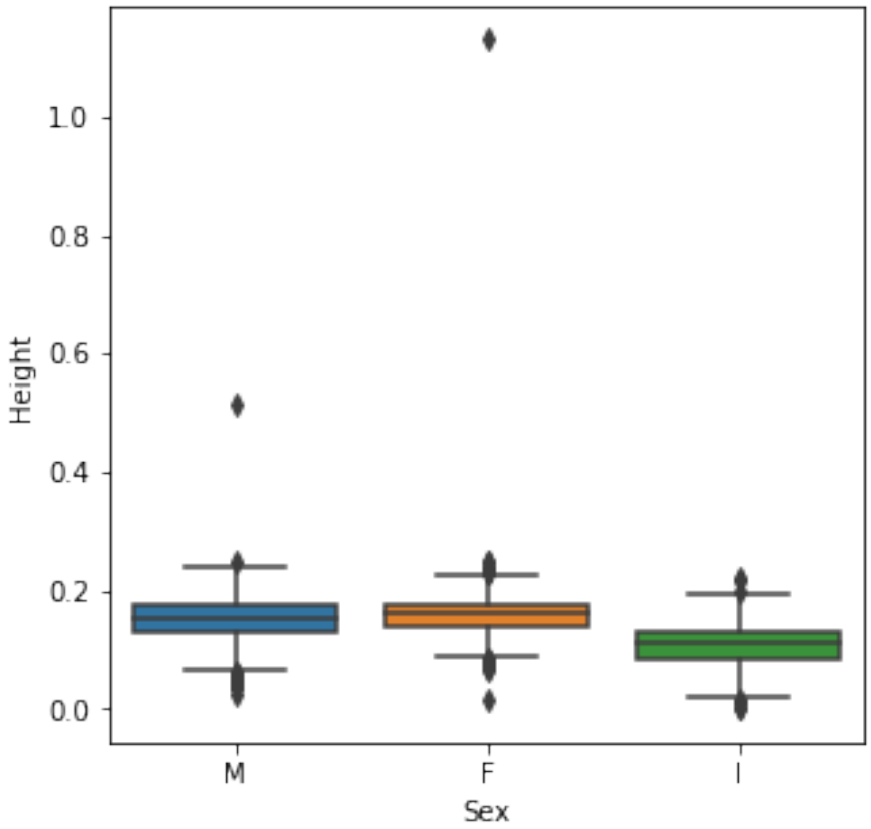



```
g = sns.FacetGrid(ds, col='Sex', margin_titles=True, size=10)
_ = g.map(sns.distplot, 'Rings', kde=False, bins=range(0, 31, 2))
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337:
UserWarning: The `size` parameter has been renamed to `height`; please
update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

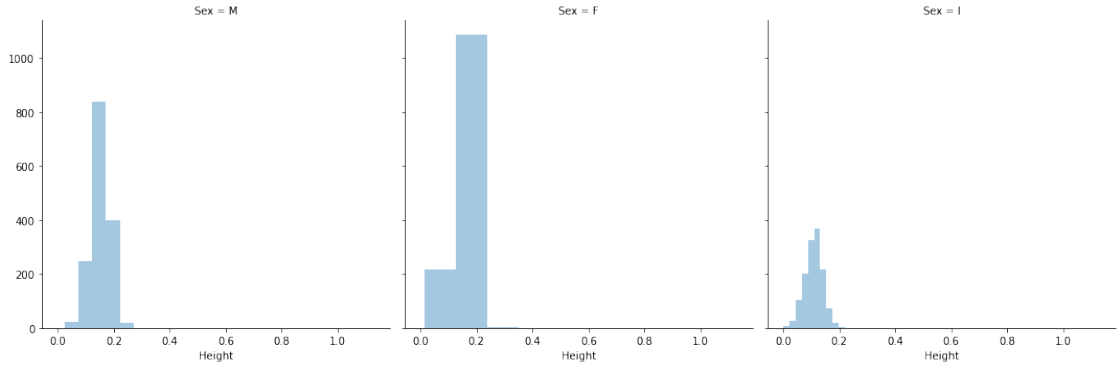


```
#sex influencing height
plt.figure(figsize=(5, 5))
_ = sns.boxplot(data=ds, x='Sex', y='Height')
```

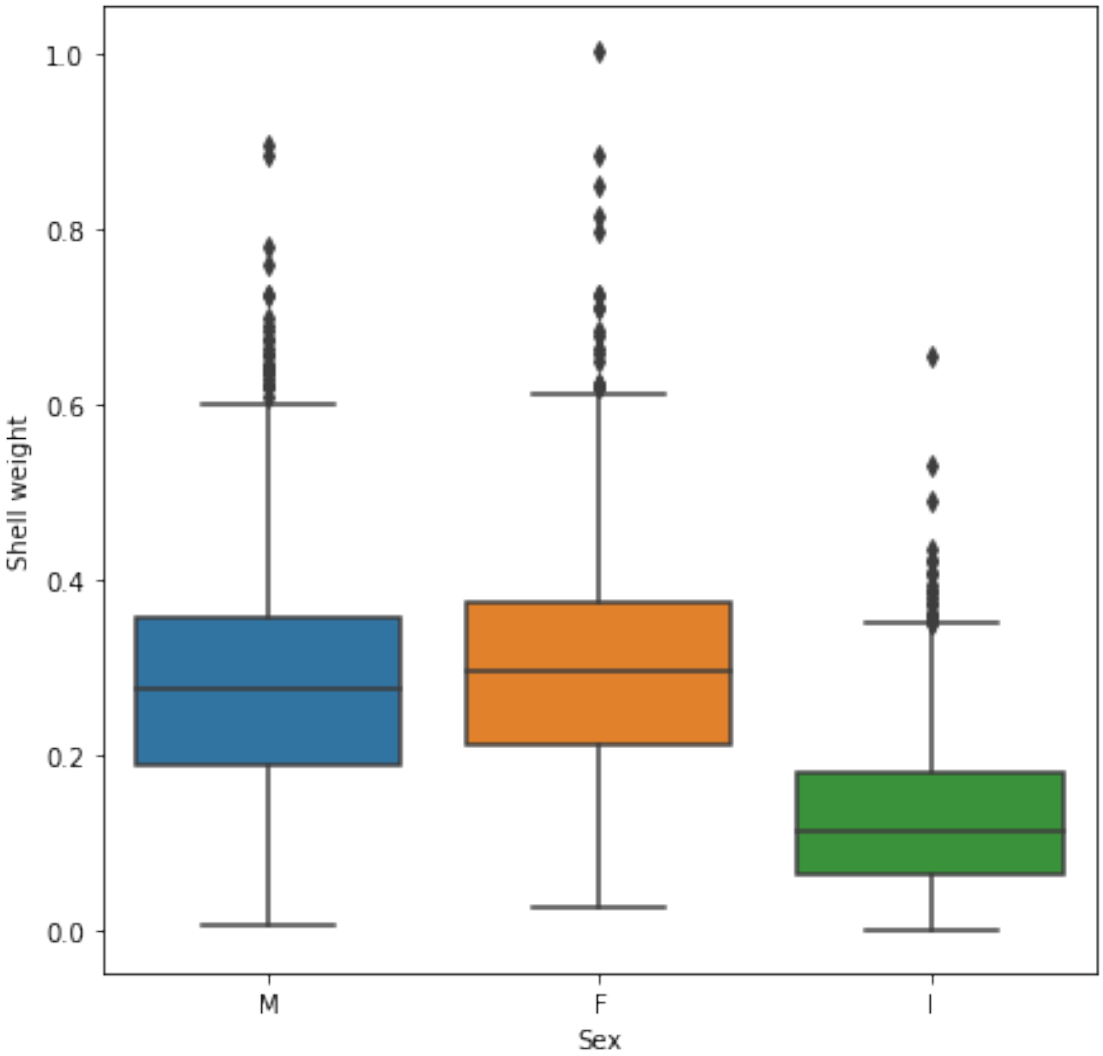



```
g = sns.FacetGrid(ds, col='Sex', margin_titles=True, size=5)
_ = g.map(sns.distplot, 'Height', kde=False, bins=10)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337:
UserWarning: The `size` parameter has been renamed to `height`; please
update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

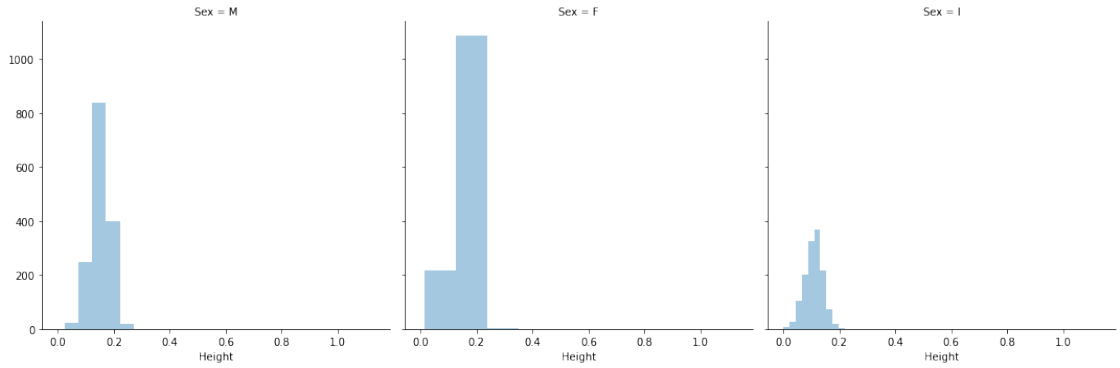



```
#sex influencing shell weight
plt.figure(figsize=(7, 7))
_ = sns.boxplot(data=ds, x='Sex', y='Shell weight')
```

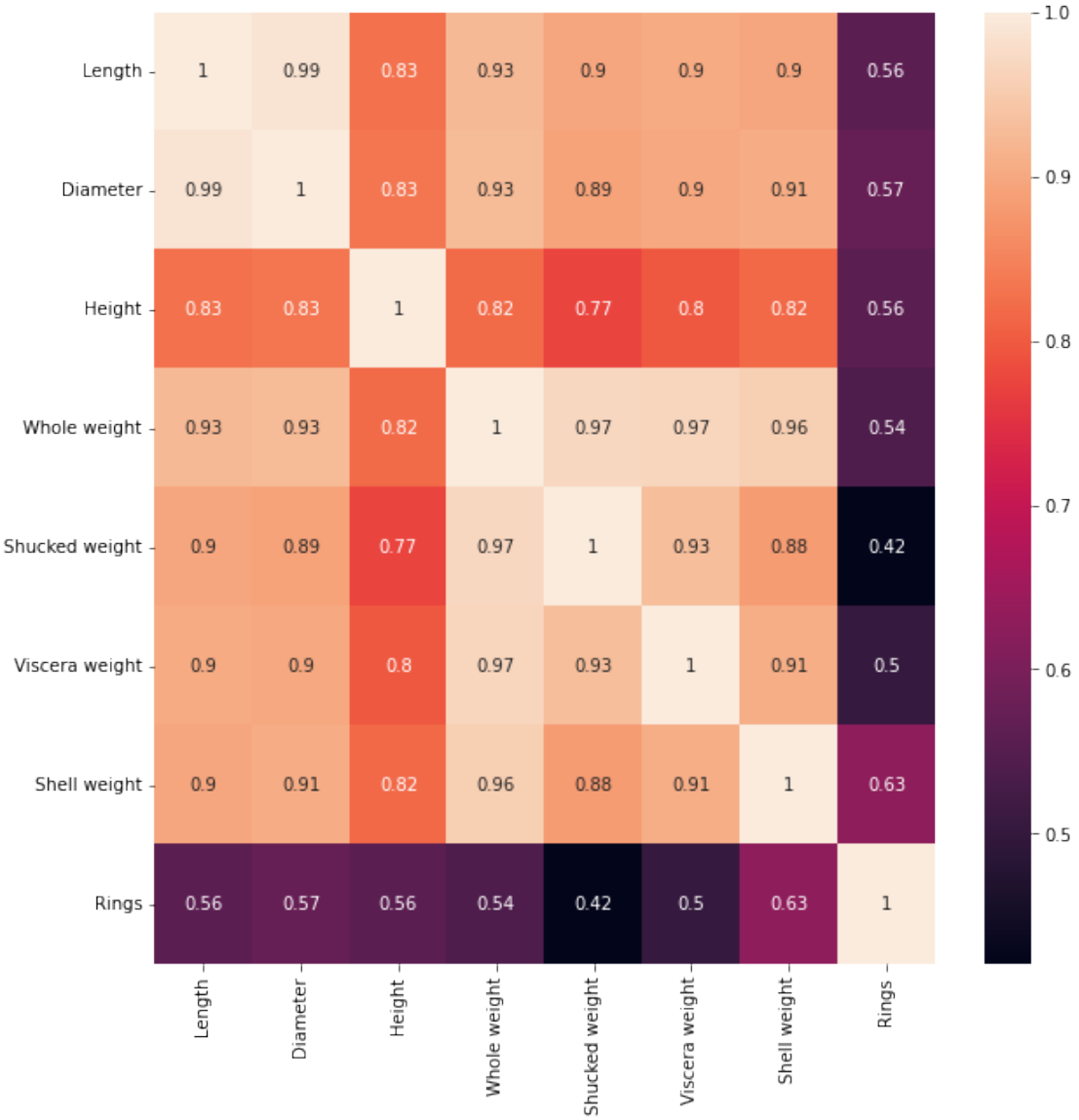



```
g = sns.FacetGrid(ds, col='Sex', margin_titles=True, size=5)
_ = g.map(sns.distplot, 'Height', kde=False, bins=10)
```

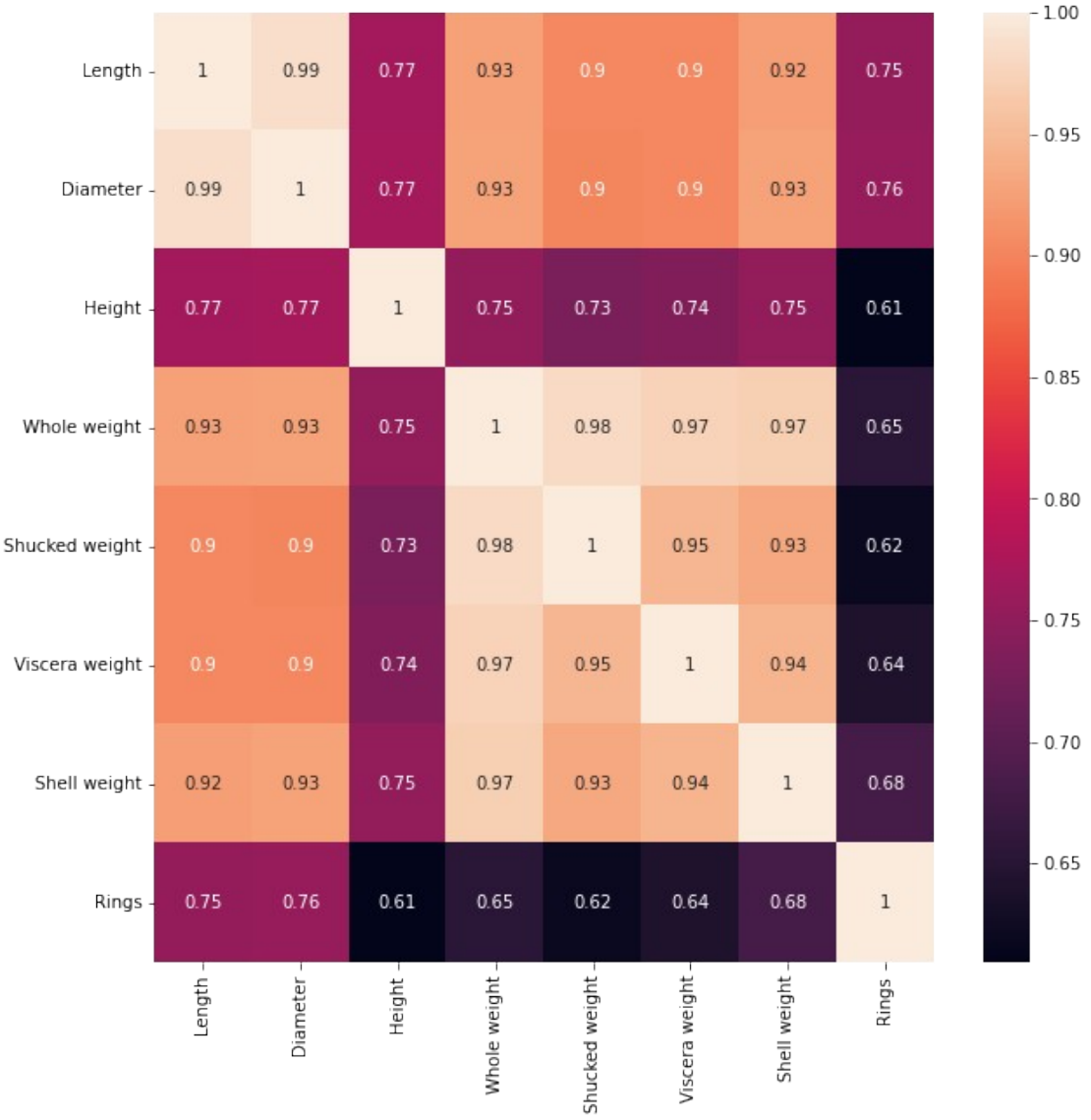
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:337:
UserWarning: The `size` parameter has been renamed to `height`; please
update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```




```
plt.figure(figsize=(10, 10))
corr = ds.corr()
_ = sns.heatmap(corr, annot=True)
```

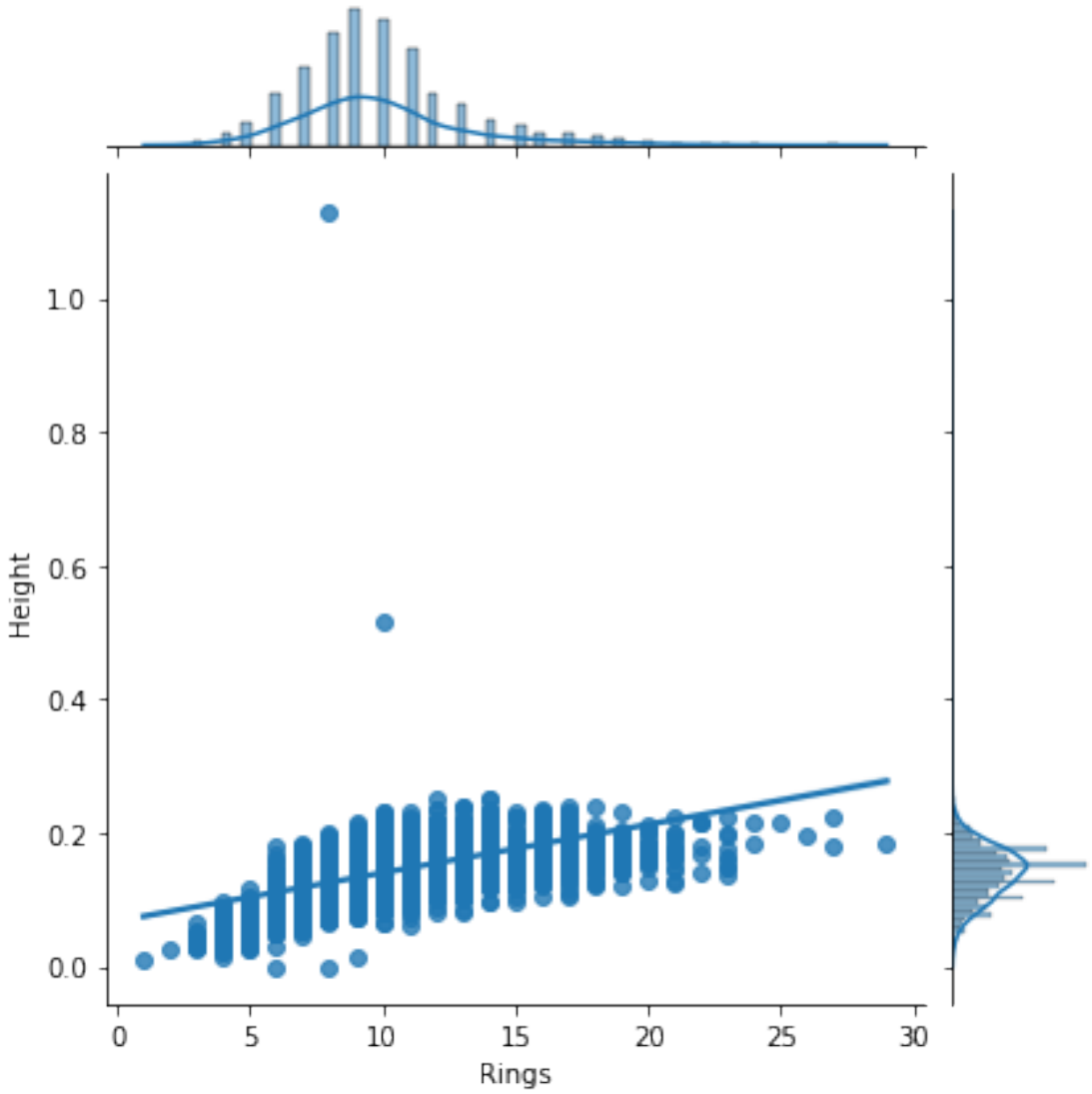


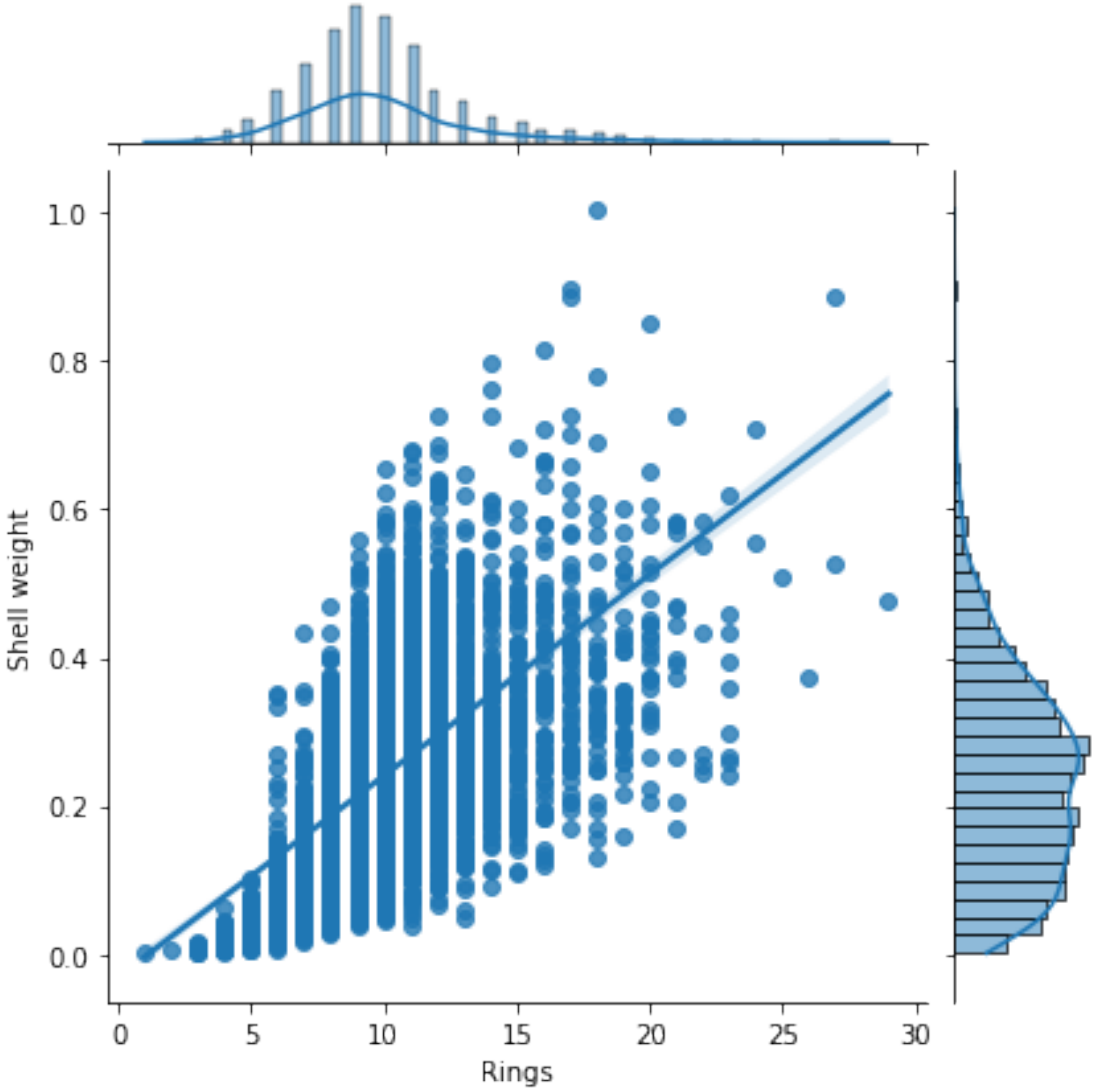

```
#variation of the correlation and the number of rings
i_abalone = ds[ds['Rings'] < 10]

plt.figure(figsize=(10, 10))
corr = i_abalone.corr()
_ = sns.heatmap(corr, annot=True)
```




```
#correlation of height and weight on rings
plt.figure(figsize=(20, 5))

_ = sns.jointplot(data=ds, x='Rings', y='Height', kind='reg')
_ = sns.jointplot(data=ds, x='Rings', y='Shell weight', kind='reg')
```

```
<Figure size 1440x360 with 0 Axes>
```


1.0
0.8
0.6
0.4
0.2
0.0
Height
0
5
10
15
20
25
30
Rings



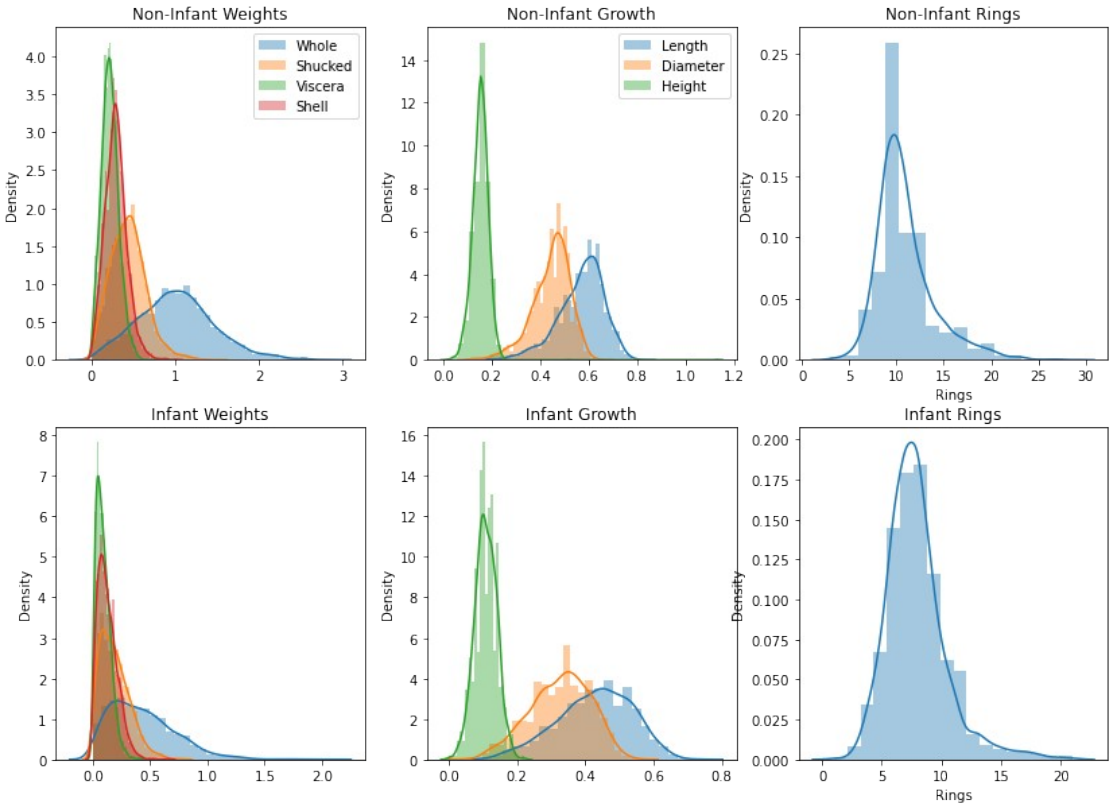

```
#Exploring the data
inf_df = ds[ds['Sex']=='I']
ni_df  = ds[ds['Sex']!='I']
fig = plt.figure(figsize=(14,10))
ax1 = fig.add_subplot(231)
ax2 = fig.add_subplot(232)
ax3 = fig.add_subplot(233)

ax4 = fig.add_subplot(234)
ax5 = fig.add_subplot(235)
ax6 = fig.add_subplot(236)

# Non infant weight info
sns.distplot(ni_df['Whole weight'],   label='Whole',   ax=ax1)
sns.distplot(ni_df['Shucked weight'], label='Shucked', ax=ax1)
sns.distplot(ni_df['Viscera weight'], label='Viscera', ax=ax1)
```

```
sns.distplot(ni_df['Shell weight'],   label='Shell',   ax=ax1)
ax1.legend()
ax1.set_title('Non-Infant Weights')


# Infant weight info
sns.distplot(inf_df['Whole weight'],   label='Whole',   ax=ax4)
sns.distplot(inf_df['Shucked weight'], label='Shucked', ax=ax4)
sns.distplot(inf_df['Viscera weight'], label='Viscera', ax=ax4)
sns.distplot(inf_df['Shell weight'],   label='Shell',   ax=ax4)
ax4.set_title('Infant Weights')


# Non-infant growth
sns.distplot(ni_df['Length'],   label='Length',   ax=ax2)
sns.distplot(ni_df['Diameter'], label='Diameter', ax=ax2)
sns.distplot(ni_df['Height'],   label='Height',   ax=ax2)
ax2.set_title('Non-Infant Growth')
ax2.legend()


# Infant growth
sns.distplot(inf_df['Length'],   label='Length',   ax=ax5)
sns.distplot(inf_df['Diameter'], label='Diameter', ax=ax5)
sns.distplot(inf_df['Height'],   label='Height',   ax=ax5)
ax5.set_title('Infant Growth')


# Plot system response
sns.distplot(ni_df['Rings'], bins=18, ax=ax3)
ax3.set_title('Non-Infant Rings')

sns.distplot(inf_df['Rings'], bins=18, ax=ax6)
ax6.set_title('Infant Rings')


ax1.set_xlabel('')
ax2.set_xlabel('')
ax3.set_xlabel('Rings')
ax4.set_xlabel('')
ax5.set_xlabel('')
ax6.set_xlabel('Rings')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
```

```
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619:
FutureWarning: `distplot` is a deprecated function and will be removed
in a future version. Please adapt your code to use either `displot` (a
figure-level function with similar flexibility) or `histplot` (an
axes-level function for histograms).
  warnings.warn(msg, FutureWarning)

Text(0.5, 0, 'Rings')
```



Perform descriptive statistics on the dataset

Whole weight has large variance compared to other weight factors, since it is integrating multiple weights.

Whole weight is therefore unlikely to correlate as strongly as other weight variables. Shell weight and viscera weight appear to be the two quantities that best reflect the size of the organism.

The length and diameter are both slightly negatively skewed.

A skewed distribution indicates a natural upper limit on the quantity, which would be expected for quantities related to size.

Infant and non-infant heights have very similar distributions Infant weights and lengths have smoother distributions with lower means and higher variances than the adult weight and length distributions.

Infants have measurements with smaller means and higher variance than adults. A model that makes predictions about the number of rings in infant shells will therefore have greater uncertainty and need to be more robust.

A good starting point is to build a model that does not include infants, reducing variability in the population.

```
ds.notnull()
```

```
       Sex  Length  Diameter  Height  Whole weight  Shucked weight  \
0     True    True      True    True          True            True
1     True    True      True    True          True            True
2     True    True      True    True          True            True
3     True    True      True    True          True            True
4     True    True      True    True          True            True
...    ...     ...       ...     ...           ...             ...
4172  True    True      True    True          True            True
4173  True    True      True    True          True            True
4174  True    True      True    True          True            True
4175  True    True      True    True          True            True
4176  True    True      True    True          True            True

      Viscera weight  Shell weight  Rings
0               True          True   True
1               True          True   True
2               True          True   True
3               True          True   True
4               True          True   True
...              ...           ...    ...
4172            True          True   True
4173            True          True   True
4174            True          True   True
4175            True          True   True
4176            True          True   True

[4177 rows x 9 columns]
```

```
ds.fillna(0)
```

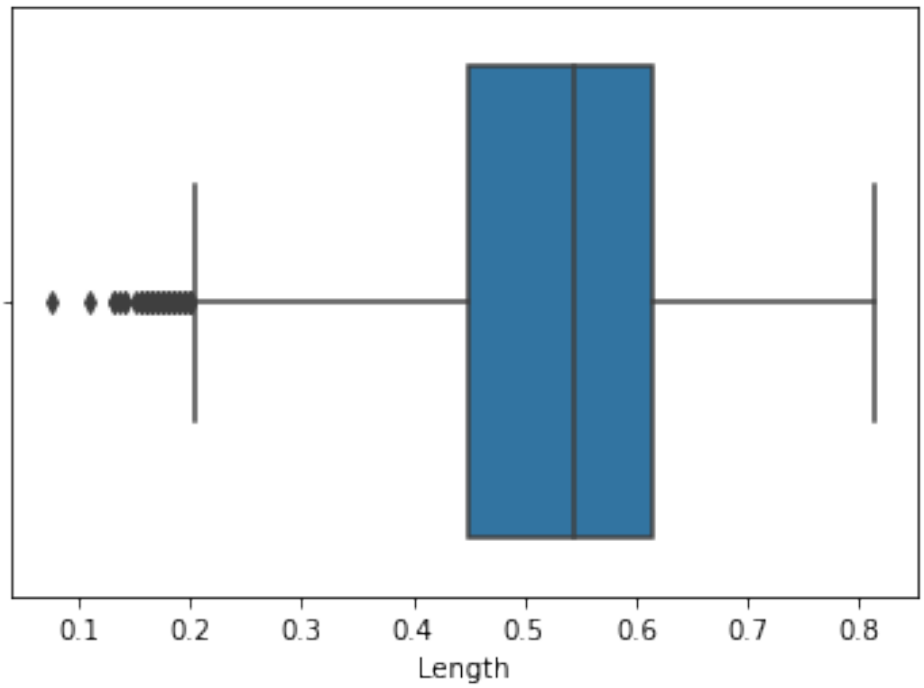
```
     Sex  Length  Diameter  Height  Whole weight  Shucked weight  \
0      M   0.455     0.365   0.095        0.5140          0.2245
1      M   0.350     0.265   0.090        0.2255          0.0995
2      F   0.530     0.420   0.135        0.6770          0.2565
3      M   0.440     0.365   0.125        0.5160          0.2155
4      I   0.330     0.255   0.080        0.2050          0.0895
...   ..     ...       ...     ...           ...             ...
4172   F   0.565     0.450   0.165        0.8870          0.3700
4173   M   0.590     0.440   0.135        0.9660          0.4390
4174   M   0.600     0.475   0.205        1.1760          0.5255
4175   F   0.625     0.485   0.150        1.0945          0.5310
4176   M   0.710     0.555   0.195        1.9485          0.9455

      Viscera weight  Shell weight  Rings
0             0.1010        0.1500     15
1             0.0485        0.0700      7
2             0.1415        0.2100      9
3             0.1140        0.1550     10
4             0.0395        0.0550      7
...              ...           ...    ...
```

```
4172          0.2390        0.2490     11
4173          0.2145        0.2605     10
4174          0.2875        0.3080      9
4175          0.2610        0.2960     10
4176          0.3765        0.4950     12

[4177 rows x 9 columns]
```

```
sns.boxplot(ds['Length'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

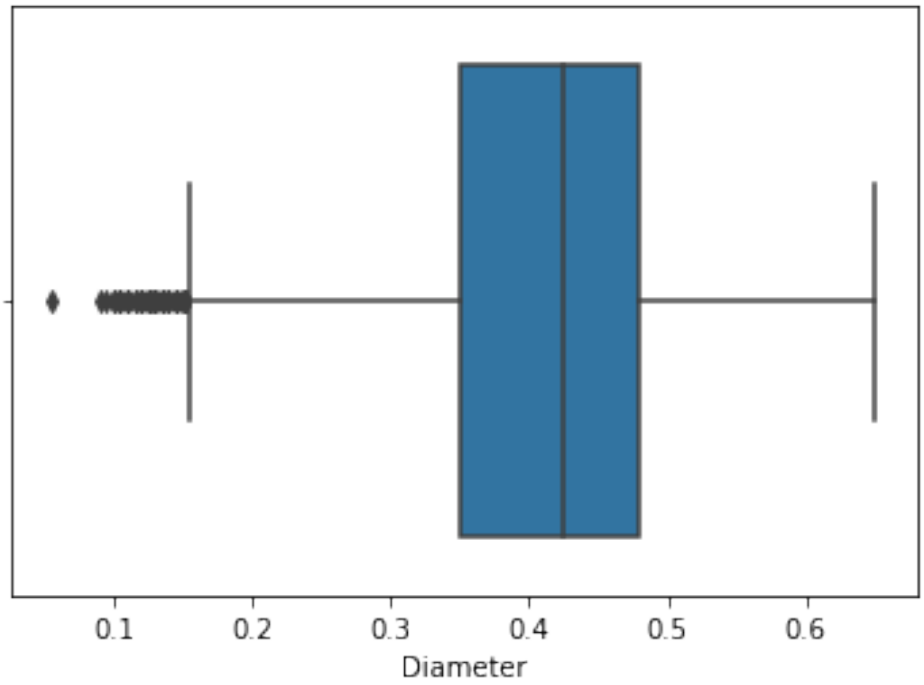
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f036272a710>
```




```
sns.boxplot(ds['Diameter'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

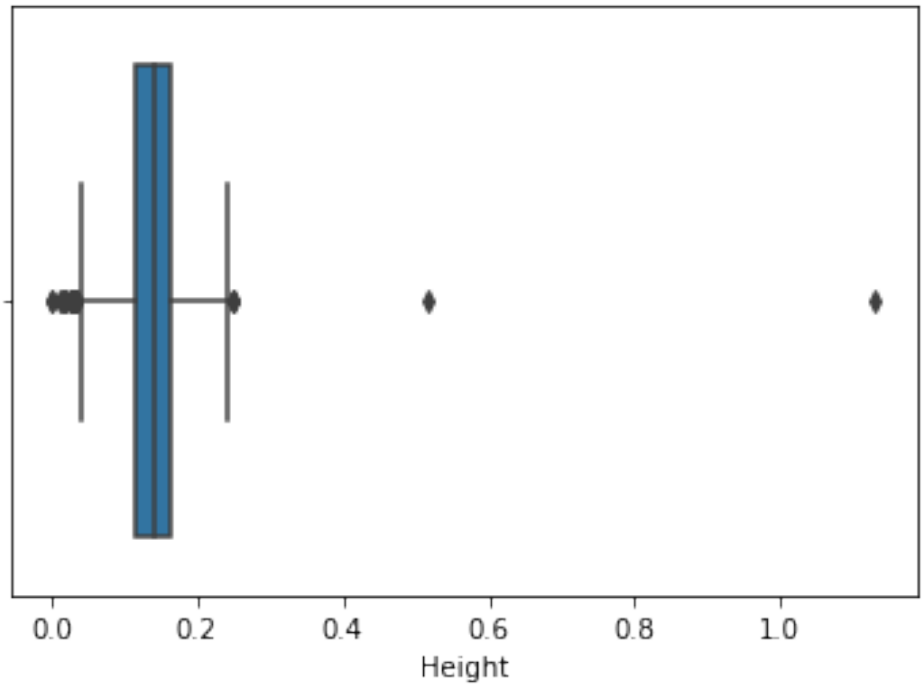
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360d32b50>
```




```
sns.boxplot(ds['Height'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

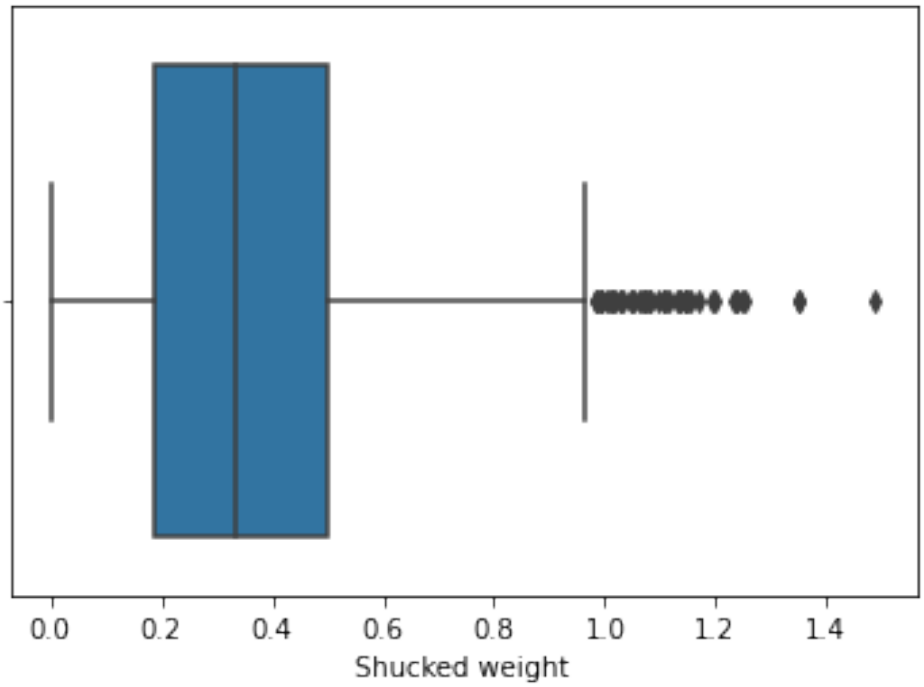
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360cef810>
```




```
sns.boxplot(ds['Shucked weight'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

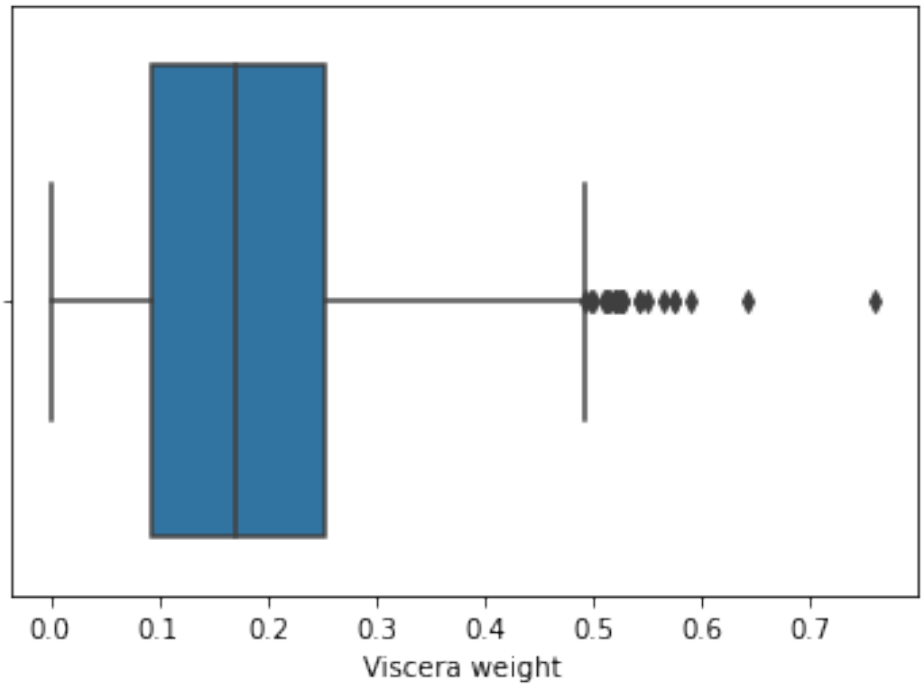
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360cf6790>
```




```
sns.boxplot(ds['Viscera weight'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

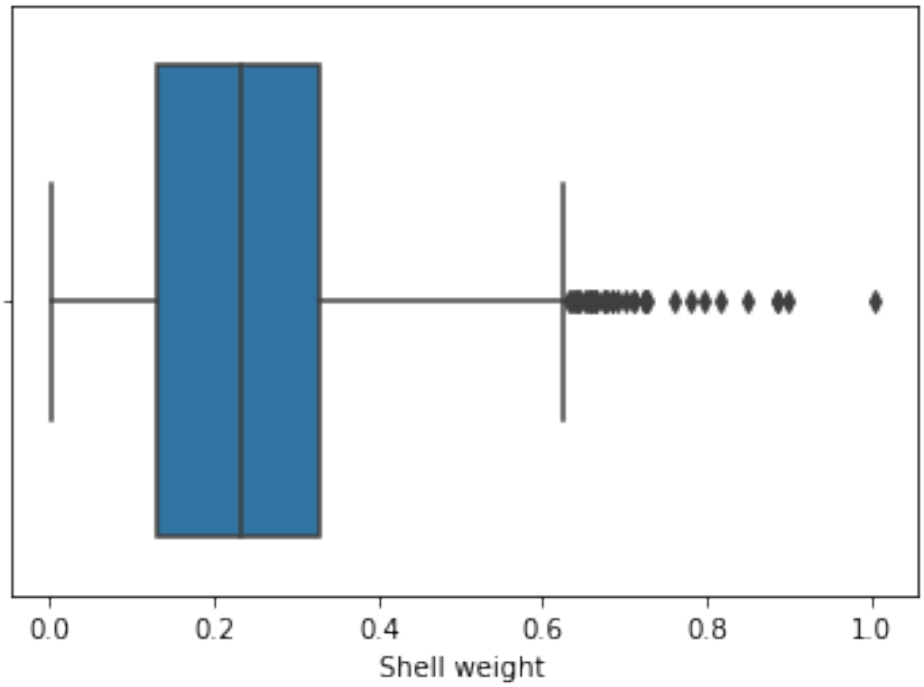
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360cd84d0>
```




```
sns.boxplot(ds['Shell weight'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

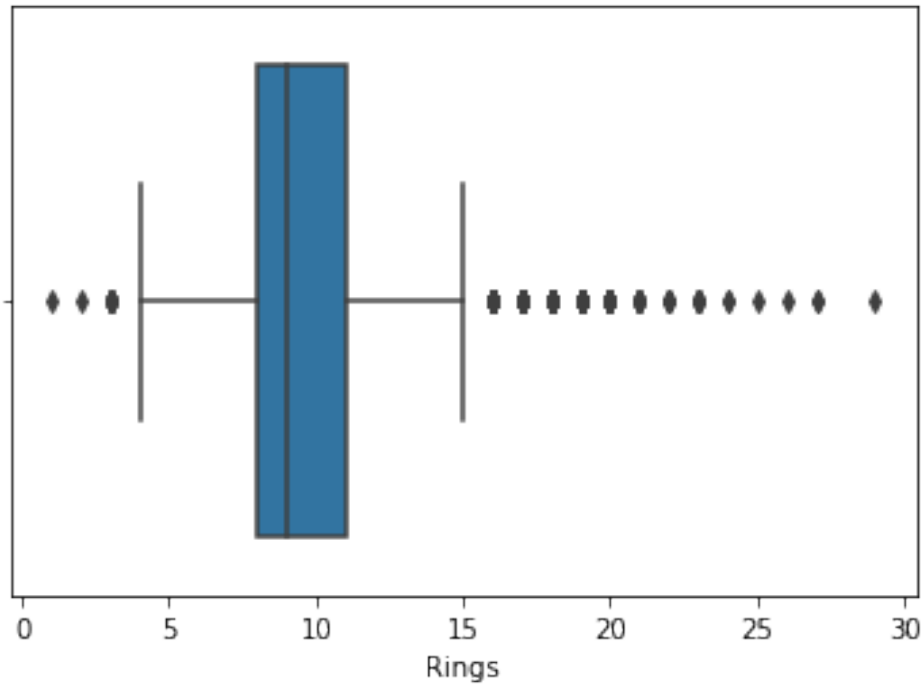
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0362ebc790>
```




```
sns.boxplot(ds['Rings'],data=ds)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0362c21c90>
```




```
#Inter Quartile Range
Q1 = ds.quantile(0.25)
Q3 = ds.quantile(0.75)
IQR = Q3-Q1
print(IQR)
```

```
Length            0.1650
Diameter          0.1300
Height            0.0500
Whole weight      0.7115
Shucked weight    0.3160
Viscera weight    0.1595
Shell weight      0.1990
Rings             3.0000
dtype: float64
```

```
#Remove outliers
abalone = ds[~((ds < (Q1 - 1.5 * IQR)) |(ds > (Q3 + 1.5 *
IQR))).any(axis=1)]
abalone.shape
```

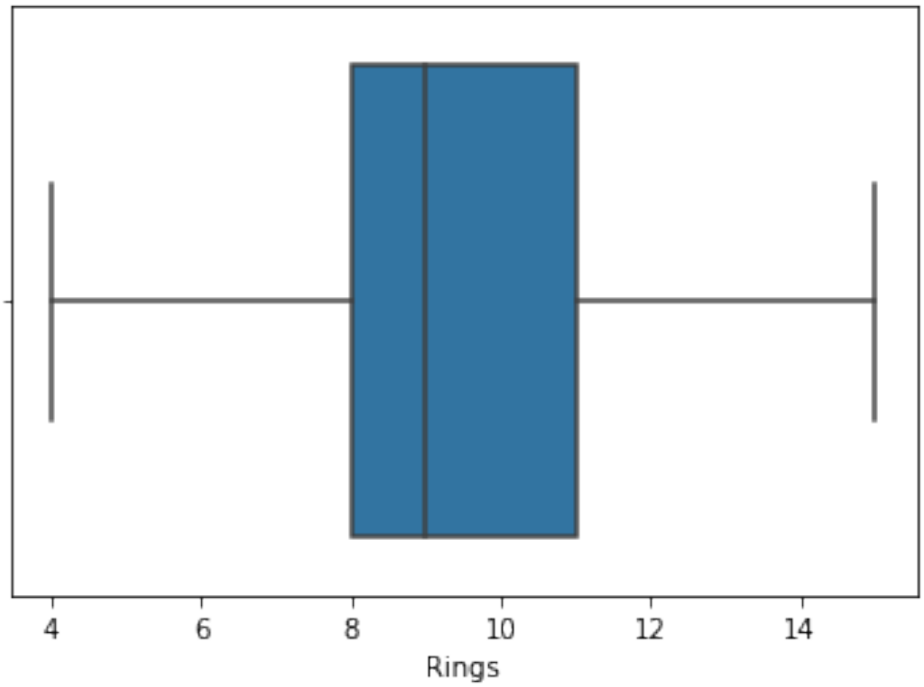
```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2:
FutureWarning: Automatic reindexing on DataFrame vs Series comparisons
is deprecated and will raise ValueError in a future version.  Do
`left, right = left.align(right, axis=1, copy=False)` before e.g.
`left == right`
```

```
(3781, 9)
```

```
sns.boxplot(abalone['Rings'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

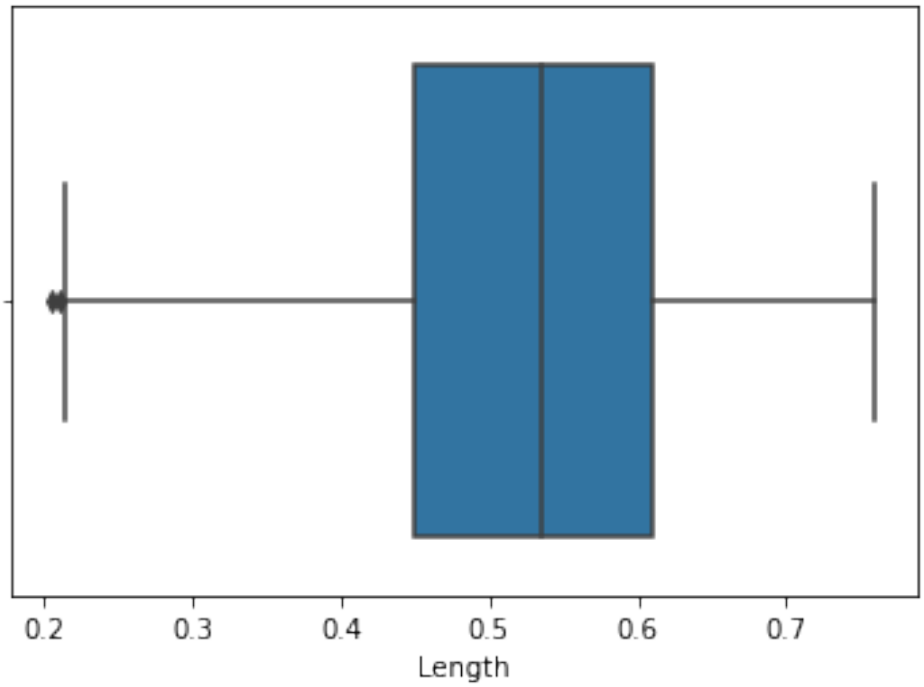
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360af2850>
```




```
sns.boxplot(abalone['Length'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

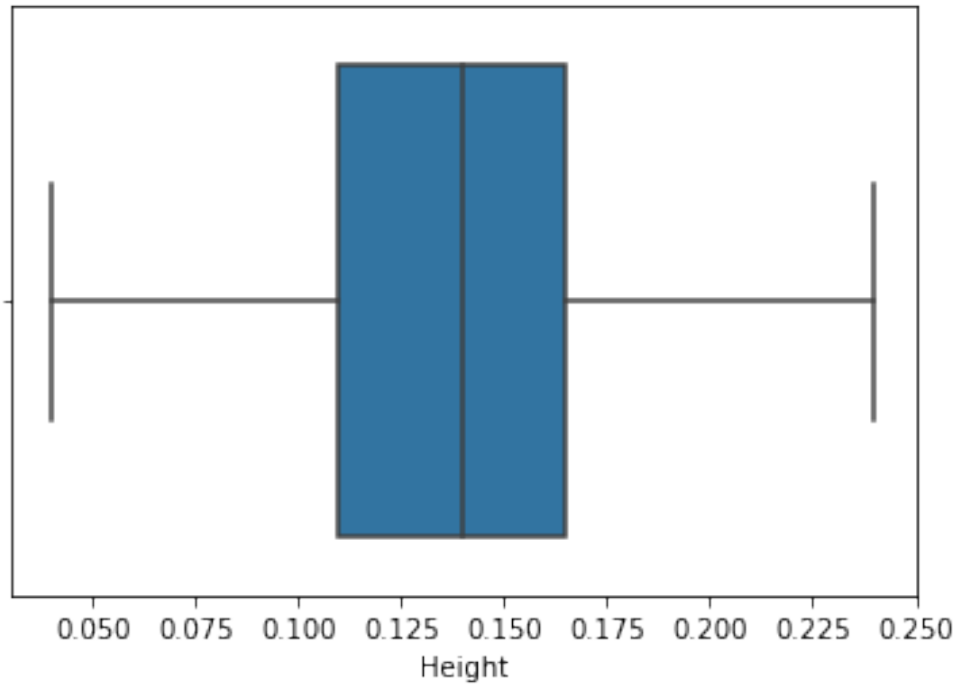
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360b63a50>
```




```
sns.boxplot(abalone['Height'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

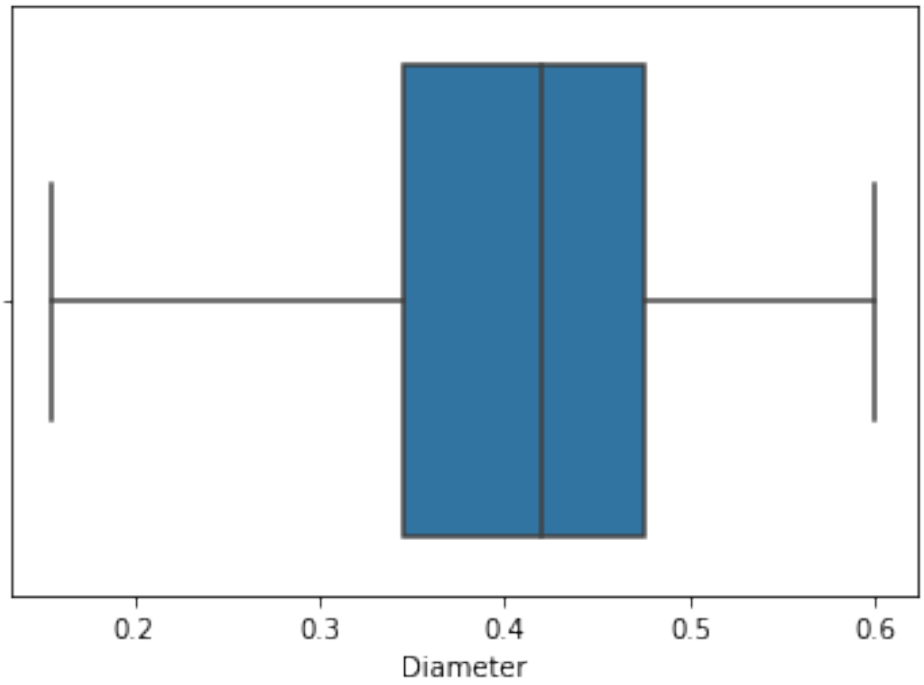
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f0360a492d0>
```




```
sns.boxplot(abalone['Diameter'],data=abalone)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/_decorators.py:43:
FutureWarning: Pass the following variable as a keyword arg: x. From
version 0.12, the only valid positional argument will be `data`, and
passing other arguments without an explicit keyword will result in an
error or misinterpretation.
  FutureWarning
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f03609b8d10>
```




```
ds["M"] = np.nan
ds["F"] = np.nan
ds["I"] = np.nan
columnName='Sex'
for i in range (len(ds[columnName])):
    if ds[columnName][i]=='M':
      ds['M'][i]=1
      ds['F'][i]=0
      ds['I'][i]=0
    elif ds[columnName][i]=='F':
      ds['M'][i]=0
      ds['F'][i]=1
      ds['I'][i]=0
    elif ds[columnName][i]=='I' :
      ds['M'][i]=0
      ds['F'][i]=0
      ds['I'][i]=1
df=ds.drop(['Sex'],axis=1)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:7:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  import sys
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:8:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:9:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  if __name__ == '__main__':
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:11:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  # This is added back by InteractiveShellApp.init_path()
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:12:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  if sys.path[0] == '':
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:13:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  del sys.path[0]
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:15:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  from ipykernel import kernelapp as app
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:16:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  app.launch_new_instance()
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:17:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
```

```
from sklearn.preprocessing import LabelEncoder
le=LabelEncoder()
ds['Sex']=le.fit_transform(ds['Sex'])
ds.head()
```

```
   Sex  Length  Diameter  Height  Whole weight  Shucked weight  \
0    2   0.455     0.365   0.095        0.5140          0.2245
1    2   0.350     0.265   0.090        0.2255          0.0995
2    0   0.530     0.420   0.135        0.6770          0.2565
3    2   0.440     0.365   0.125        0.5160          0.2155
4    1   0.330     0.255   0.080        0.2050          0.0895

   Viscera weight  Shell weight  Rings    M    F    I
0          0.1010         0.150     15  1.0  0.0  0.0
1          0.0485         0.070      7  1.0  0.0  0.0
2          0.1415         0.210      9  0.0  1.0  0.0
3          0.1140         0.155     10  1.0  0.0  0.0
4          0.0395         0.055      7  0.0  0.0  1.0
```

```
X = ds.iloc[:, :-1].values
Y = ds.iloc[:, -1].values
```

```
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
scaler.fit(ds)
```

```
StandardScaler()
```

```
from sklearn.model_selection import train_test_split
train_X,val_X,train_y,val_y = train_test_split(X, Y, test_size = 0.2,
random_state = 0)
```

```
print("Shape of Training X :",train_X.shape)
print("Shape of Validation X :",val_X.shape)
print("Shape of Training y :",train_y.shape)
print("Shape of Validation y :",val_y.shape)
```

```
Shape of Training X : (3341, 11)
Shape of Validation X : (836, 11)
```

```
Shape of Training y : (3341,)
Shape of Validation y : (836,)
```

```
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(train_X,train_y)
```

```
LinearRegression()
```

```
from sklearn import  metrics
```

```
%%time
y_pred_val_lr = lr.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y,
y_pred_val_lr))
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y,
y_pred_val_lr))
print("\n")
print('RMSE on Validation
set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_lr)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y,
y_pred_val_lr))
print("\n")
```

```
MAE on Validation set : 1.3373140978439386e-16


MSE on Validation set : 3.764133319058843e-32


RMSE on Validation set : 1.1564229753182607e-08


R2 Score on Validation set : 1.0


CPU times: user 11.1 ms, sys: 1.02 ms, total: 12.1 ms
Wall time: 14 ms
```

```
from sklearn.tree import DecisionTreeRegressor
dc = DecisionTreeRegressor(random_state = 0)
dc.fit(train_X,train_y)
```

```
DecisionTreeRegressor(random_state=0)
```

```
%%time
y_pred_val_dc = dc.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y,
y_pred_val_dc))
```

```
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y,
y_pred_val_dc))
print("\n")
print('RMSE on Validation
set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_dc)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y,
y_pred_val_dc))
print("\n")
```

```
MAE on Validation set : 0.0


MSE on Validation set : 0.0


RMSE on Validation set : 0.0


R2 Score on Validation set : 1.0


CPU times: user 11.3 ms, sys: 2.1 ms, total: 13.4 ms
Wall time: 15.3 ms
```

```
from sklearn.svm import SVR
svm = SVR()
svm.fit(train_X,train_y)
```

```
SVR()
```

```
%%time
y_pred_val_svm = svm.predict(val_X)
print('MAE on Validation set :',metrics.mean_absolute_error(val_y,
y_pred_val_svm))
print("\n")
print('MSE on Validation set :',metrics.mean_squared_error(val_y,
y_pred_val_svm))
print("\n")
print('RMSE on Validation
set :',np.sqrt(metrics.mean_absolute_error(val_y, y_pred_val_svm)))
print("\n")
print('R2 Score on Validation set :',metrics.r2_score(val_y,
y_pred_val_svm))
print("\n")
```

```
MAE on Validation set : 0.08812585132106524
```

```
MSE on Validation set : 0.008371038340690765


RMSE on Validation set : 0.29685998605582603


R2 Score on Validation set : 0.9631105317812326


CPU times: user 33.7 ms, sys: 760 µs, total: 34.4 ms
Wall time: 69.8 ms
```

```
print('Linear R2 Score on Validation set :',metrics.r2_score(val_y,
y_pred_val_lr))
print('SVR R2 Score on Validation set :',metrics.r2_score(val_y,
y_pred_val_svm))
print('Decision Tree Regressor R2 Score on Validation
set :',metrics.r2_score(val_y, y_pred_val_dc))
```

```
Linear R2 Score on Validation set : 1.0
SVR R2 Score on Validation set : 0.9631105317812326
Decision Tree Regressor R2 Score on Validation set : 1.0
```