

## 2. Exploratory Data Analysis

December 14, 2020

### 0.0.1 Importing libraries:

```
[1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib.inline
import numpy as np
```

```
C:\Users\Hari Prasad\anaconda3\lib\importlib\_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
    return f(*args, **kwargs)
C:\Users\Hari Prasad\anaconda3\lib\importlib\_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
    return f(*args, **kwargs)
C:\Users\Hari Prasad\anaconda3\lib\importlib\_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
    return f(*args, **kwargs)
C:\Users\Hari Prasad\anaconda3\lib\importlib\_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
    return f(*args, **kwargs)
C:\Users\Hari Prasad\anaconda3\lib\importlib\_bootstrap.py:219: RuntimeWarning:
numpy.ufunc size changed, may indicate binary incompatibility. Expected 192 from
C header, got 216 from PyObject
    return f(*args, **kwargs)
UsageError: Line magic function `%matplotlib.inline` not found.
```

### 0.0.2 Indian crops dataset:

```
[2]: df = pd.read_csv('Indian crops dataset.csv')
df.head(10)
```

```
[2]:
```

	Dist Code	Year	State Code	State Name	Dist Name	RICE AREA (1000 ha)	\
0	1	1966	14	Chhattisgarh	Durg	548.0	
1	1	1967	14	Chhattisgarh	Durg	547.0	

2	1	1968	14	Chhattisgarh	Durg	556.3
3	1	1969	14	Chhattisgarh	Durg	563.4
4	1	1970	14	Chhattisgarh	Durg	571.6
5	1	1971	14	Chhattisgarh	Durg	581.8
6	1	1972	14	Chhattisgarh	Durg	582.2
7	1	1973	14	Chhattisgarh	Durg	600.0
8	1	1974	14	Chhattisgarh	Durg	587.4
9	1	1975	14	Chhattisgarh	Durg	598.3

	RICE PRODUCTION (1000 tons)	RICE YIELD (Kg per ha) \
0	185.0	338
1	409.0	748
2	468.0	841
3	400.8	711
4	473.6	829
5	412.9	710
6	381.0	654
7	471.9	787
8	219.0	373
9	454.0	759

	RICE SHARE IN TOTAL (Percent)	WHEAT AREA (1000 ha) ... \
0	50.63	44.0 ...
1	42.16	50.0 ...
2	41.59	53.7 ...
3	42.04	49.4 ...
4	41.33	44.2 ...
5	41.56	44.4 ...
6	42.47	39.6 ...
7	42.43	37.3 ...
8	50.42	36.5 ...
9	42.89	49.2 ...

	COTTON YIELD (Kg per ha)	COTTON SHARE IN TOTAL (Percent) \
0	0	0.0
1	0	0.0
2	0	0.0
3	0	0.0
4	0	0.0
5	0	0.0
6	0	0.0
7	0	0.0
8	0	0.0
9	0	0.0

	FRUITS AREA (1000 ha)	FRUITS SHARE IN TOTAL (Percent) \
0	5.95	0.55

1	5.77	0.44
2	5.41	0.40
3	5.52	0.41
4	5.45	0.39
5	5.42	0.39
6	5.48	0.40
7	5.30	0.37
8	5.21	0.45
9	5.11	0.37

	VEGETABLES AREA (1000 ha)	VEGETABLES SHARE IN TOTAL (Percent)	\
0	6.64	0.61	
1	7.24	0.56	
2	7.40	0.55	
3	7.16	0.53	
4	7.19	0.52	
5	7.48	0.53	
6	7.53	0.55	
7	7.60	0.54	
8	7.44	0.64	
9	7.86	0.56	

	FRUITS AND VEGETABLES AREA (1000 ha)	POTATOES AREA (1000 ha)	\
0	12.59	0.01	
1	13.02	0.01	
2	12.81	0.10	
3	12.69	0.01	
4	12.64	0.02	
5	12.91	0.01	
6	13.01	0.01	
7	12.90	0.01	
8	12.64	0.03	
9	12.97	0.05	

	ONION AREA (1000 ha)	FODDER AREA (1000 ha)
0	0.60	0.47
1	0.56	1.23
2	0.58	1.02
3	0.56	0.84
4	0.52	0.42
5	0.54	0.38
6	0.55	0.26
7	0.53	0.14
8	0.45	0.06
9	0.52	0.08

[10 rows x 109 columns]

### 0.0.3 Dataset shortlisted:

```
[3]: df1 = pd.read_excel('Indian crops shortlisted.xlsx')
df1.head(10)
```

```
[3]:   Year  state_name dist_name  rice_area_1000ha  rice_production_1000tons \
0  1966  Chhattisgarh    Durg          548.0             185.0
1  1967  Chhattisgarh    Durg          547.0             409.0
2  1968  Chhattisgarh    Durg          556.3             468.0
3  1969  Chhattisgarh    Durg          563.4             400.8
4  1970  Chhattisgarh    Durg          571.6             473.6
5  1971  Chhattisgarh    Durg          581.8             412.9
6  1972  Chhattisgarh    Durg          582.2             381.0
7  1973  Chhattisgarh    Durg          600.0             471.9
8  1974  Chhattisgarh    Durg          587.4             219.0
9  1975  Chhattisgarh    Durg          598.3             454.0
```

```
   rice_yield_kgperha  rice_share_total_percent  wheat_area_1000ha \
0             338          50.63             44.0
1             748          42.16             50.0
2             841          41.59             53.7
3             711          42.04             49.4
4             829          41.33             44.2
5             710          41.56             44.4
6             654          42.47             39.6
7             787          42.43             37.3
8             373          50.42             36.5
9             759          42.89             49.2
```

```
   wheat_production_1000tons  wheat_yield_kgperha  ... \
0              20.0             455  ...
1              26.0             520  ...
2              30.0             559  ...
3              26.5             536  ...
4              29.0             656  ...
5              25.8             581  ...
6              20.6             520  ...
7              18.6             499  ...
8              22.4             614  ...
9              27.8             565  ...
```

```
   sugarcane_production_1000tons  sugarcane_yield_kgperha \
0              1.6             1778
1              1.2             1500
2              1.0             1000
3              1.9             1900
4              1.4             2000
```

5	1.0	2000
6	1.0	2000
7	1.2	6000
8	1.5	1875
9	1.4	1750

	sugarcane_share_total_percent	fruits_area_1000ha \
0	0.08	5.95
1	0.06	5.77
2	0.07	5.41
3	0.07	5.52
4	0.05	5.45
5	0.04	5.42
6	0.04	5.48
7	0.01	5.30
8	0.07	5.21
9	0.06	5.11

	fruits_share_total_percent	vegetables_area_1000ha \
0	0.55	6.64
1	0.44	7.24
2	0.40	7.40
3	0.41	7.16
4	0.39	7.19
5	0.39	7.48
6	0.40	7.53
7	0.37	7.60
8	0.45	7.44
9	0.37	7.86

	vegetables_share_total_percent	fruits_vegetables_area_1000ha \
0	0.61	12.59
1	0.56	13.02
2	0.55	12.81
3	0.53	12.69
4	0.52	12.64
5	0.53	12.91
6	0.55	13.01
7	0.54	12.90
8	0.64	12.64
9	0.56	12.97

	potatoes_area_1000ha	onion_area_1000ha
0	0.01	0.60
1	0.01	0.56
2	0.10	0.58
3	0.01	0.56

4	0.02	0.52
5	0.01	0.54
6	0.01	0.55
7	0.01	0.53
8	0.03	0.45
9	0.05	0.52

[10 rows x 22 columns]

#### 0.0.4 Correlation between Year & Rice area, Rice production, Rice yield, Rice share:

```
[4]: df1[['Year', 'rice_area_1000ha', 'rice_production_1000tons', 'rice_yield_kgperha', 'rice_share_tot
      ↪corr()
```

```
[4]:
```

	Year	rice_area_1000ha \
Year	1.000000	0.046335
rice_area_1000ha	0.046335	1.000000
rice_production_1000tons	0.225493	0.833241
rice_yield_kgperha	0.373542	0.213743
rice_share_total_percent	0.010535	0.798122

	rice_production_1000tons	rice_yield_kgperha \
Year	0.225493	0.373542
rice_area_1000ha	0.833241	0.213743
rice_production_1000tons	1.000000	0.524574
rice_yield_kgperha	0.524574	1.000000
rice_share_total_percent	0.637859	0.227043

	rice_share_total_percent
Year	0.010535
rice_area_1000ha	0.798122
rice_production_1000tons	0.637859
rice_yield_kgperha	0.227043
rice_share_total_percent	1.000000

#### 0.0.5 Correlation between Wheat & Wheat area, Wheat production, Wheat yield, Wheat share:

```
[5]: df1[['Year', 'wheat_area_1000ha', 'wheat_production_1000tons', 'wheat_yield_kgperha', 'wheat_share_tot
      ↪corr()
```

```
[5]:
```

	Year	wheat_area_1000ha \
Year	1.000000	0.140247
wheat_area_1000ha	0.140247	1.000000
wheat_production_1000tons	0.234321	0.909427
wheat_yield_kgperha	0.362038	0.610646

wheat_share_total_percent	0.108153	0.781800
---------------------------	----------	----------

	wheat_production_1000tons	wheat_yield_kgperha \
Year	0.234321	0.362038
wheat_area_1000ha	0.909427	0.610646
wheat_production_1000tons	1.000000	0.680656
wheat_yield_kgperha	0.680656	1.000000
wheat_share_total_percent	0.633834	0.574634

	wheat_share_total_percent
Year	0.108153
wheat_area_1000ha	0.781800
wheat_production_1000tons	0.633834
wheat_yield_kgperha	0.574634
wheat_share_total_percent	1.000000

### 0.0.6 Correlation between Year & Sugarcane area, production, yield, share:

```
[6]: df1[['Year', 'sugarcane_area_1000ha', 'sugarcane_production_1000tons', 'sugarcane_yield_kgperha',
        ↪corr()]
```

```
[6]:
```

	Year	sugarcane_area_1000ha \
Year	1.000000	0.094464
sugarcane_area_1000ha	0.094464	1.000000
sugarcane_production_1000tons	0.128651	0.961040
sugarcane_yield_kgperha	0.082547	0.227332
sugarcane_share_total_percent	0.073631	0.945497

	sugarcane_production_1000tons \
Year	0.128651
sugarcane_area_1000ha	0.961040
sugarcane_production_1000tons	1.000000
sugarcane_yield_kgperha	0.307622
sugarcane_share_total_percent	0.891960

	sugarcane_yield_kgperha \
Year	0.082547
sugarcane_area_1000ha	0.227332
sugarcane_production_1000tons	0.307622
sugarcane_yield_kgperha	1.000000
sugarcane_share_total_percent	0.207572

	sugarcane_share_total_percent
Year	0.073631
sugarcane_area_1000ha	0.945497
sugarcane_production_1000tons	0.891960
sugarcane_yield_kgperha	0.207572

sugarcane_share_total_percent	1.000000
-------------------------------	----------

### 0.0.7 Correlation between Year and Fruits area, share, Vegetables area, share, Potatoes area, Onion area:

```
[7]: df1[['Year', 'fruits_area_1000ha', 'fruits_share_total_percent', 'vegetables_area_1000ha', 'vegetables_share_total_percent', 'potatoes_area_1000ha', 'onion_area_1000ha']]
      ↪ corr()
```

```
[7]:
```

	Year	fruits_area_1000ha \
Year	1.000000	0.189445
fruits_area_1000ha	0.189445	1.000000
fruits_share_total_percent	0.153222	0.757910
vegetables_area_1000ha	0.185049	0.332415
vegetables_share_total_percent	0.118254	0.237564
fruits_vegetables_area_1000ha	0.217213	0.759885
potatoes_area_1000ha	0.143341	-0.020023
onion_area_1000ha	0.105068	0.081590

	fruits_share_total_percent \
Year	0.153222
fruits_area_1000ha	0.757910
fruits_share_total_percent	1.000000
vegetables_area_1000ha	0.156777
vegetables_share_total_percent	0.315590
fruits_vegetables_area_1000ha	0.510193
potatoes_area_1000ha	-0.056950
onion_area_1000ha	-0.024417

	vegetables_area_1000ha \
Year	0.185049
fruits_area_1000ha	0.332415
fruits_share_total_percent	0.156777
vegetables_area_1000ha	1.000000
vegetables_share_total_percent	0.663231
fruits_vegetables_area_1000ha	0.854747
potatoes_area_1000ha	0.461127
onion_area_1000ha	0.251201

	vegetables_share_total_percent \
Year	0.118254
fruits_area_1000ha	0.237564
fruits_share_total_percent	0.315590
vegetables_area_1000ha	0.663231
vegetables_share_total_percent	1.000000
fruits_vegetables_area_1000ha	0.576091
potatoes_area_1000ha	0.325209
onion_area_1000ha	0.052149



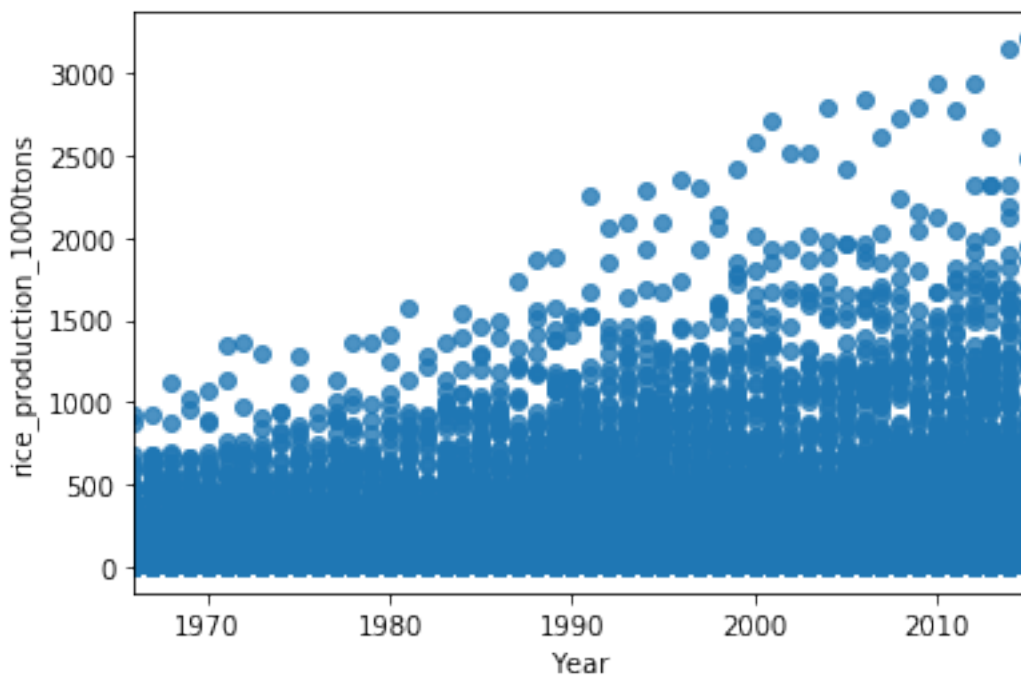
	fruits_vegetables_area_1000ha \
Year	0.217213
fruits_area_1000ha	0.759885
fruits_share_total_percent	0.510193
vegetables_area_1000ha	0.854747
vegetables_share_total_percent	0.576091
fruits_vegetables_area_1000ha	1.000000
potatoes_area_1000ha	0.308664
onion_area_1000ha	0.215406

	potatoes_area_1000ha	onion_area_1000ha
Year	0.143341	0.105068
fruits_area_1000ha	-0.020023	0.081590
fruits_share_total_percent	-0.056950	-0.024417
vegetables_area_1000ha	0.461127	0.251201
vegetables_share_total_percent	0.325209	0.052149
fruits_vegetables_area_1000ha	0.308664	0.215406
potatoes_area_1000ha	1.000000	-0.003818
onion_area_1000ha	-0.003818	1.000000

### 0.0.8 Year vs Rice production:

```
[16]: sns.regplot(x='Year',y='rice_production_1000tons',data=df1)
```

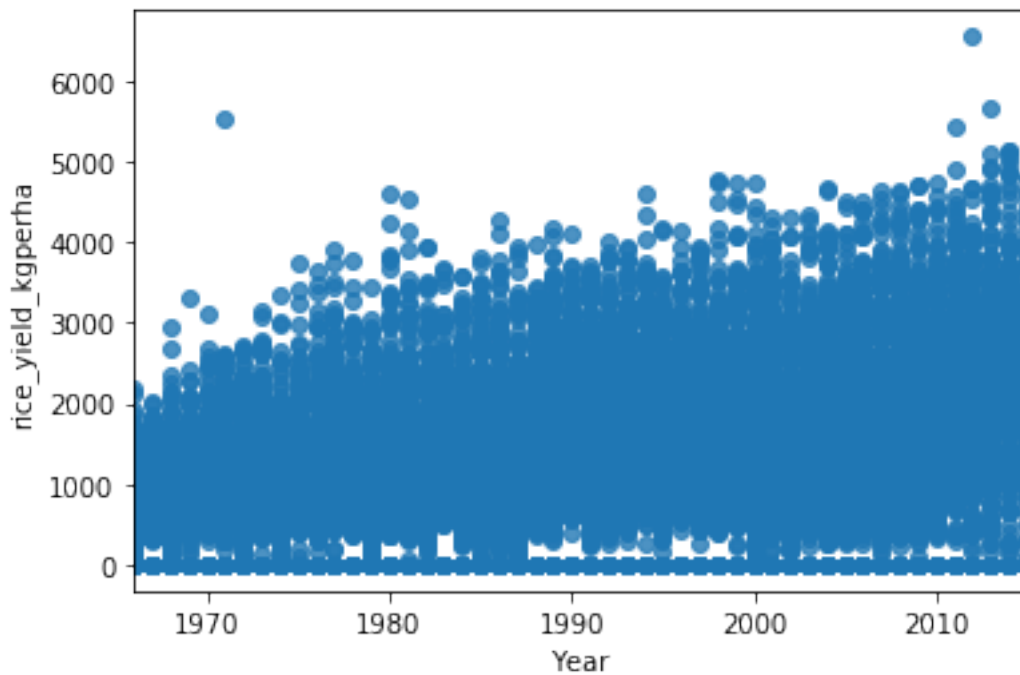
```
[16]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2acc8ec8>
```



### 0.0.9 Year vs Rice yield:

```
[9]: sns.regplot(x='Year',y='rice_yield_kgperha',data=df1)
```

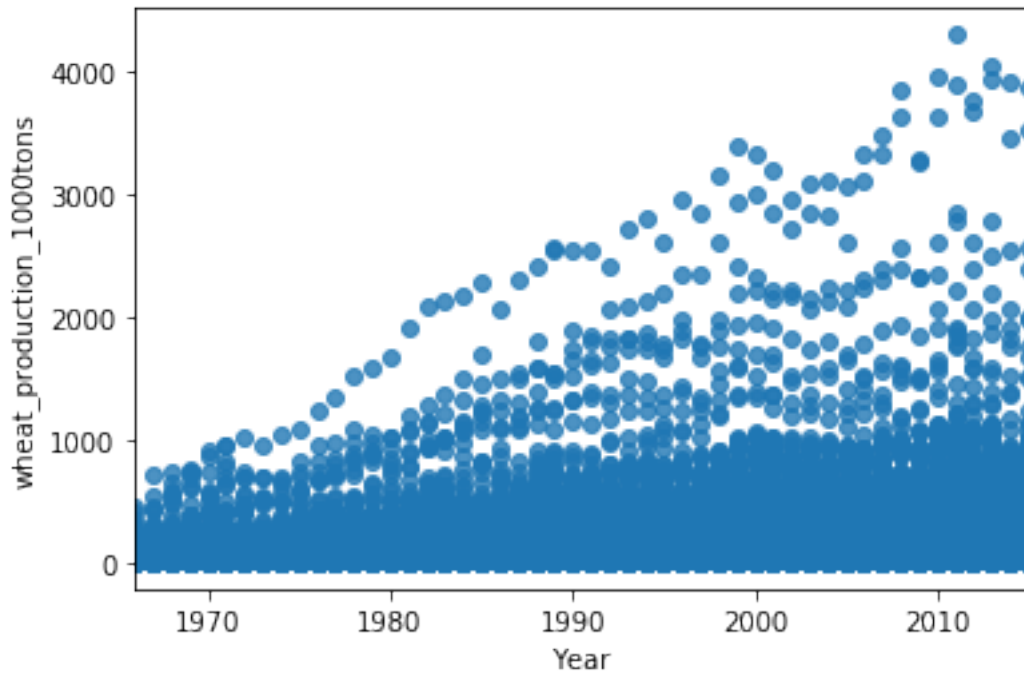
```
[9]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2bf6f8c8>
```



### 0.0.10 Year vs Wheat production:

```
[10]: sns.regplot(x='Year',y='wheat_production_1000tons',data=df1)
```

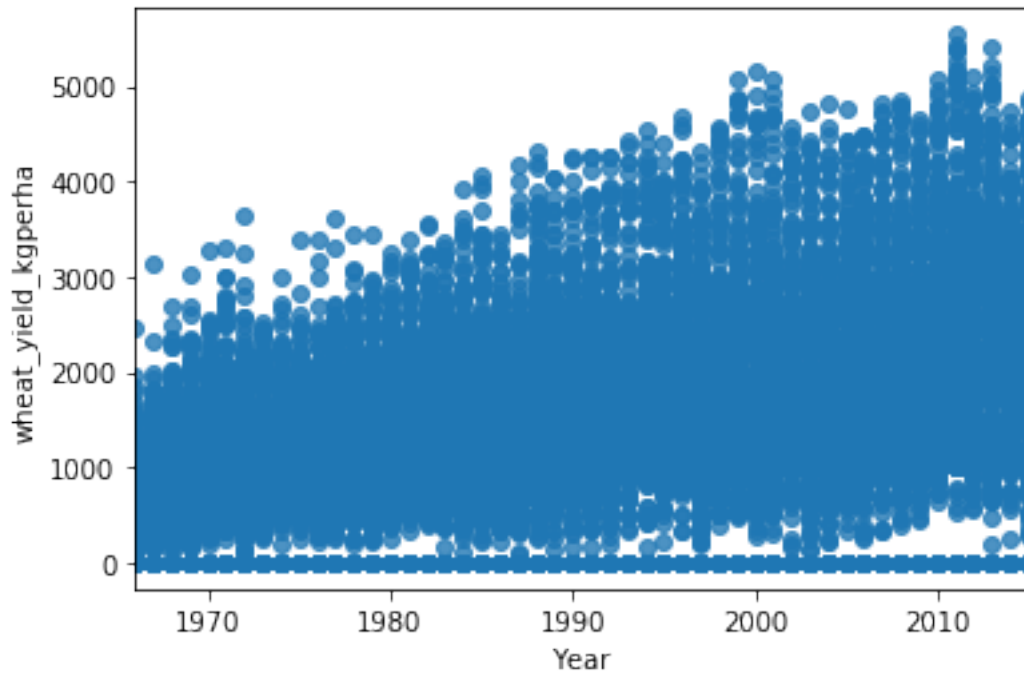
```
[10]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2d031cc8>
```



#### 0.0.11 Year vs Wheat yield:

```
[11]: sns.regplot(x='Year',y='wheat_yield_kgperha',data=df1)
```

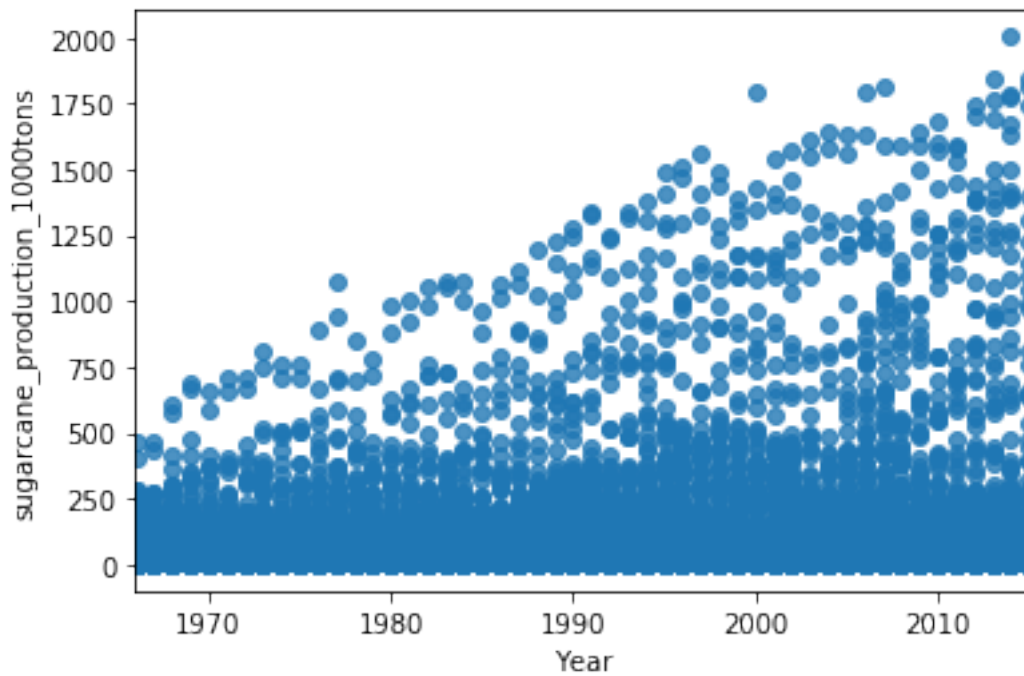
```
[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2adc6d88>
```



#### 0.0.12 Year vs Sugarcane production:

```
[12]: sns.regplot(x='Year',y='sugarcane_production_1000tons',data=df1)
```

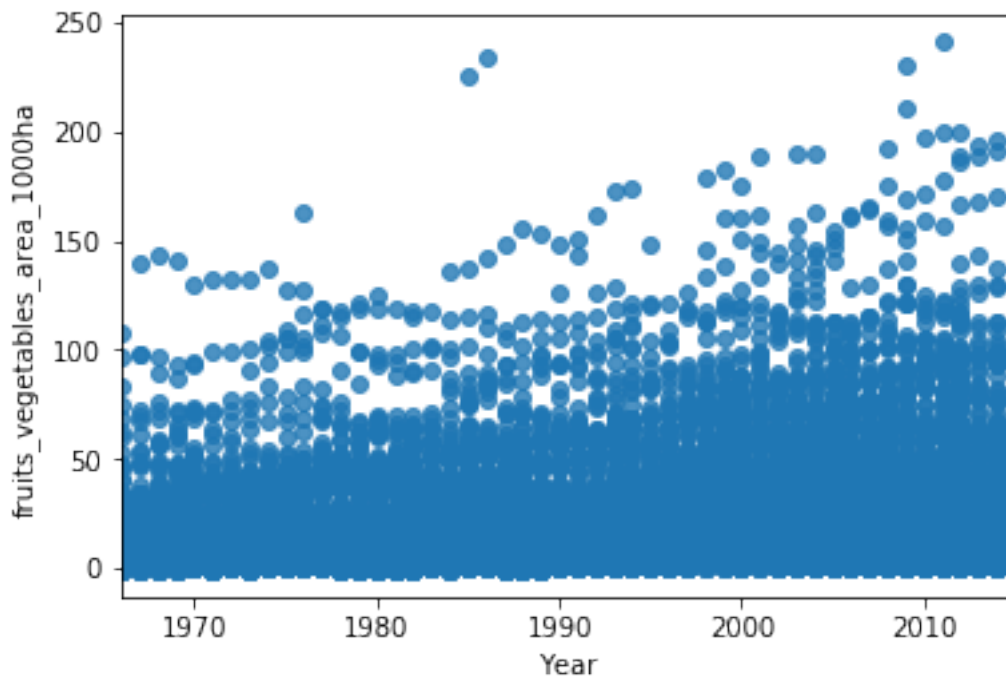
```
[12]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2abf3108>
```



#### 0.0.13 Year vs Fruits, Vegetables area:

```
[13]: sns.regplot(x='Year',y='fruits_vegetables_area_1000ha',data=df1)
```

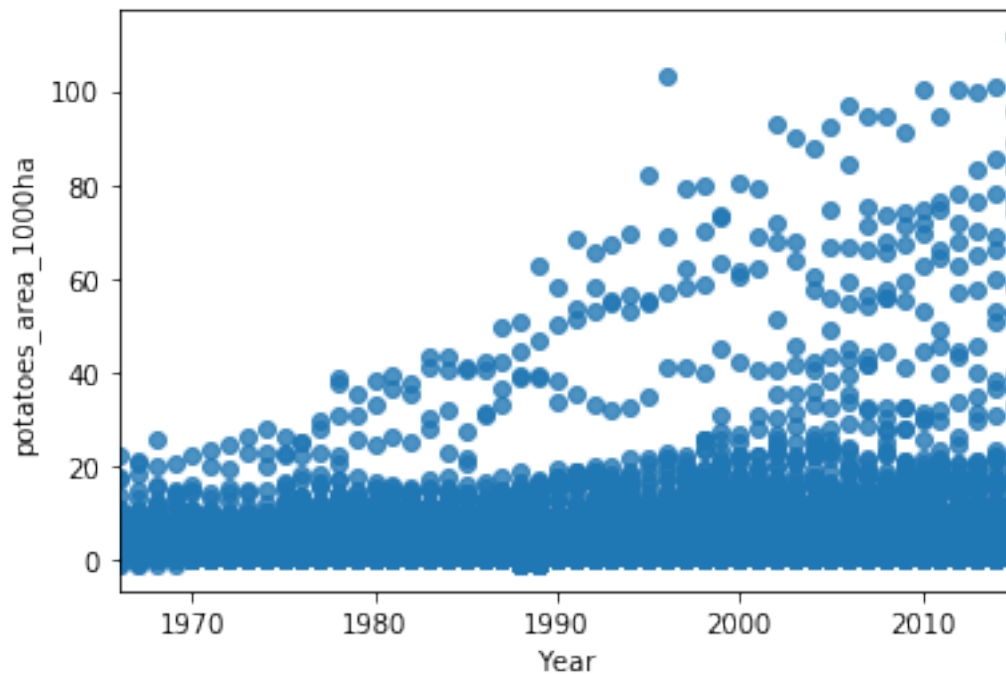
```
[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2afd6d88>
```



#### 0.0.14 Year vs Potatoes area:

```
[17]: sns.regplot(x='Year',y='potatoes_area_1000ha',data=df1)
```

```
[17]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2d72cb08>
```



#### 0.0.15 Year vs Onion area:

```
[15]: sns.regplot(x='Year',y='onion_area_1000ha',data=df1)
```

```
[15]: <matplotlib.axes._subplots.AxesSubplot at 0x1dc2ba89688>
```

