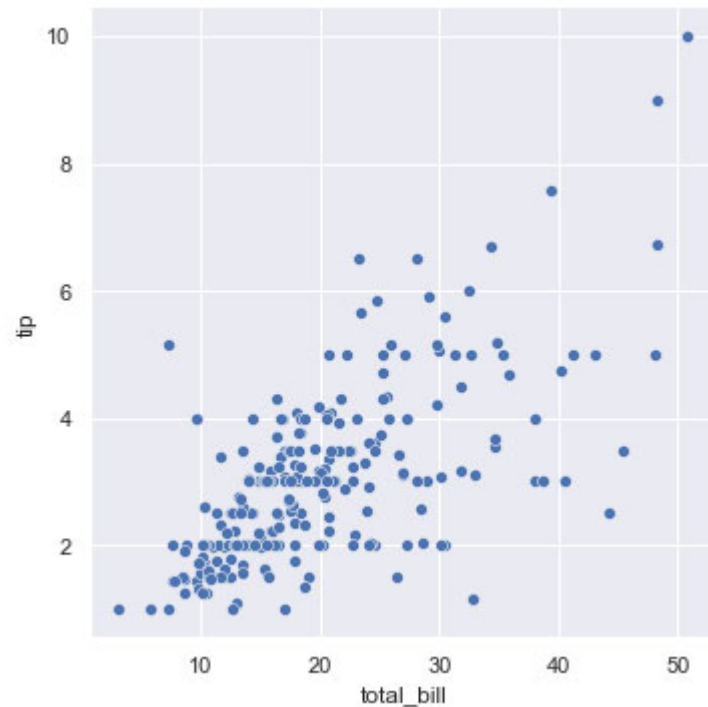


```
In [2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
sns.set_theme(style="darkgrid")
```

```
In [3]: tips = sns.load_dataset("tips")
sns.relplot(x="total_bill", y="tip", data=tips);
```



```
In [7]: import os
print(os.getcwd())
```

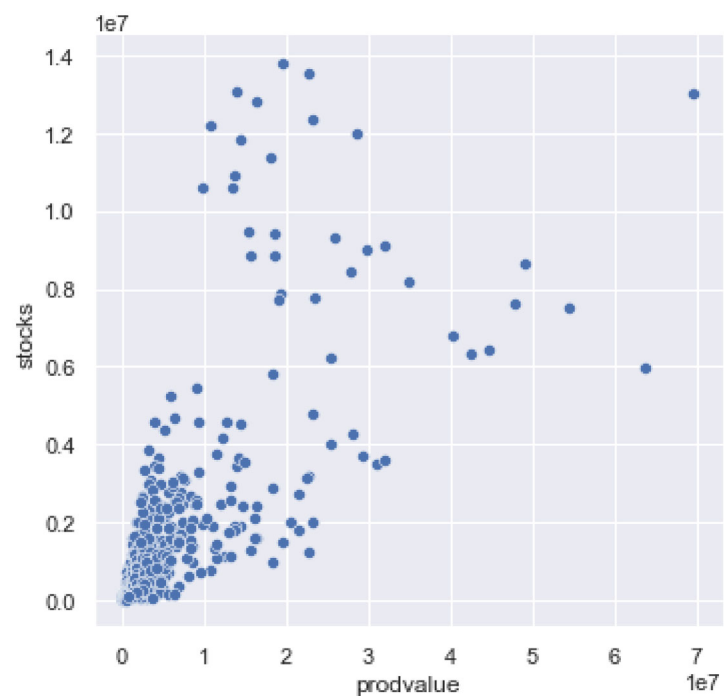
C:\Users\spark\Desktop\Data Science Projects\Honey Production in USA

```
In [8]: import matplotlib.pyplot as plt
import seaborn as sns #Uses matplotlib but works as an upgrade of matplotlib.
%matplotlib inline
os.chdir('C:\\Users\\spark\\Desktop\\Data Science Projects\\Honey Production in USA') #This is changing the working directory
Honeyproduction = pd.read_csv('honeyproduction.csv')
```

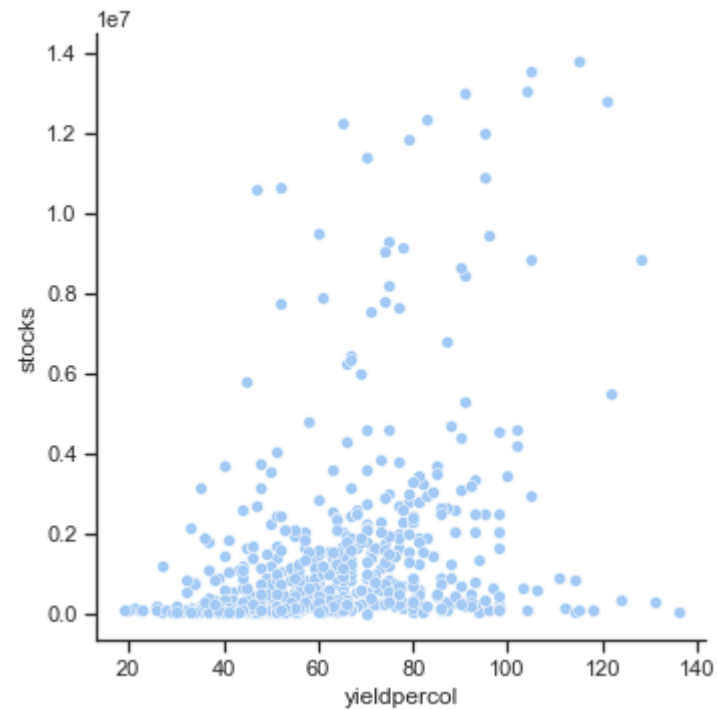
```
In [11]: a =Honeyproduction.stocks  
        b =Honeyproduction.prodvalue
```

```
Out[11]: 0      159000.0  
        1      1485000.0  
        2      1688000.0  
        3      12326000.0  
        4      1594000.0  
        ...  
        621      23000.0  
        622      1017000.0  
        623      95000.0  
        624      1863000.0  
        625      459000.0  
        Name: stocks, Length: 626, dtype: float64
```

```
In [10]: sns.relplot(x="prodvalue", y="stocks", data=Honeyproduction);
```

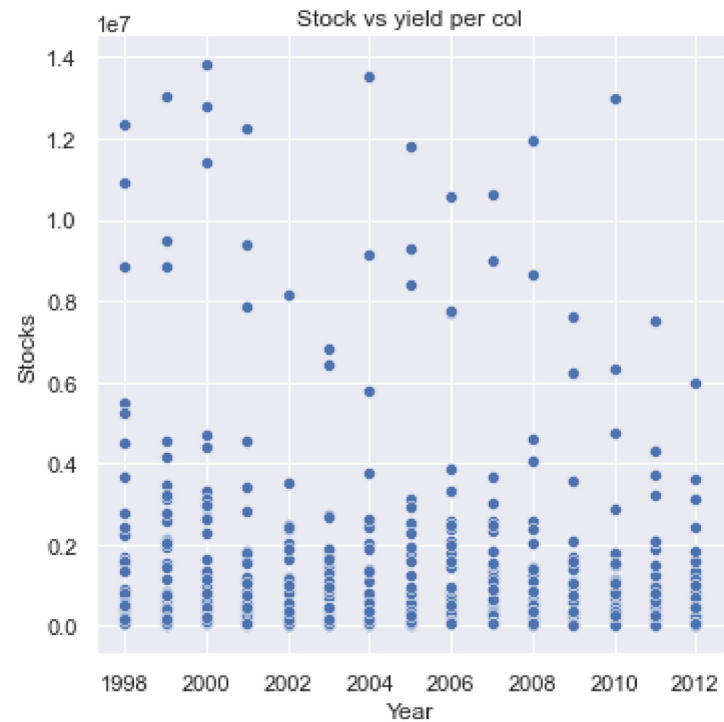


```
In [90]: sns.relplot(x="yieldpercol", y="stocks", data=Honeyproduction);
```



```
In [13]: sns.relplot(x="year", y="stocks", data=Honeyproduction)
plt.title("Stock vs yield per col")
plt.xlabel("Year")
plt.ylabel("Stocks")
#Would like to see this in different boxplot distribution.
```

```
Out[13]: Text(6.800000000000004, 0.5, 'Stocks')
```



```
In [14]: sns.set_theme(style="ticks", palette="pastel")
```

```
In [15]: Honeyproduction
```

```
Out[15]:
```

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
0	AL	16000.0	71	1136000.0	159000.0	0.72	818000.0	1998
1	AZ	55000.0	60	3300000.0	1485000.0	0.64	2112000.0	1998
2	AR	53000.0	65	3445000.0	1688000.0	0.59	2033000.0	1998
3	CA	450000.0	83	37350000.0	12326000.0	0.62	23157000.0	1998
4	CO	27000.0	72	1944000.0	1594000.0	0.70	1361000.0	1998
...
621	VA	4000.0	41	164000.0	23000.0	3.77	618000.0	2012
622	WA	62000.0	41	2542000.0	1017000.0	2.38	6050000.0	2012

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
623	WV	6000.0	48	288000.0	95000.0	2.91	838000.0	2012
624	WI	60000.0	69	4140000.0	1863000.0	2.05	8487000.0	2012
625	WY	50000.0	51	2550000.0	459000.0	1.87	4769000.0	2012

626 rows × 8 columns

```
In [18]: newhoneyproduction = pd.read_csv("honeyproduction.csv", index_col = "prodvalue" )
#This puts the prdo value as the first one. So we wenat to put the state as the first one.
newhoneyproduction = pd.read_csv("honeyproduction.csv", index_col = "year" )
newhoneyproduction = pd.read_csv("honeyproduction.csv")
```

```
In [22]: newhoneyproduction
```

```
Out[22]:
```

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
0	AL	16000.0	71	1136000.0	159000.0	0.72	818000.0	1998
1	AZ	55000.0	60	3300000.0	1485000.0	0.64	2112000.0	1998
2	AR	53000.0	65	3445000.0	1688000.0	0.59	2033000.0	1998
3	CA	450000.0	83	37350000.0	12326000.0	0.62	23157000.0	1998
4	CO	27000.0	72	1944000.0	1594000.0	0.70	1361000.0	1998
...
621	VA	4000.0	41	164000.0	23000.0	3.77	618000.0	2012
622	WA	62000.0	41	2542000.0	1017000.0	2.38	6050000.0	2012
623	WV	6000.0	48	288000.0	95000.0	2.91	838000.0	2012
624	WI	60000.0	69	4140000.0	1863000.0	2.05	8487000.0	2012
625	WY	50000.0	51	2550000.0	459000.0	1.87	4769000.0	2012

626 rows × 8 columns

```
In [23]: index_names = newhoneyproduction[newhoneyproduction['year'] == 1998 ].index
```

```

index_names1 = newhoneyproduction[newhoneyproduction['year'] == 1999 ].index
index_names2 = newhoneyproduction[newhoneyproduction['year'] == 2000 ].index
index_names3= newhoneyproduction[newhoneyproduction['year'] == 2001 ].index
index_names4 = newhoneyproduction[newhoneyproduction['year'] == 2002 ].index
index_names5 = newhoneyproduction[newhoneyproduction['year'] == 2003 ].index
index_names6 = newhoneyproduction[newhoneyproduction['year'] == 2004 ].index
index_names7 = newhoneyproduction[newhoneyproduction['year'] == 2005 ].index
index_names8 = newhoneyproduction[newhoneyproduction['year'] == 2006 ].index

```

[#https://www.geeksforgeeks.org/how-to-drop-rows-in-dataframe-by-conditions-on-column-values/](https://www.geeksforgeeks.org/how-to-drop-rows-in-dataframe-by-conditions-on-column-values/)

```

In [24]: # drop these row indexes
# from dataframe
newhoneyproduction.drop(index_names, inplace = True)
newhoneyproduction.drop(index_names1, inplace = True)
newhoneyproduction.drop(index_names2, inplace = True)
newhoneyproduction.drop(index_names3, inplace = True)
newhoneyproduction.drop(index_names4, inplace = True)
newhoneyproduction.drop(index_names5, inplace = True)
newhoneyproduction.drop(index_names6, inplace = True)
newhoneyproduction.drop(index_names7, inplace = True)
newhoneyproduction.drop(index_names8, inplace = True)

```

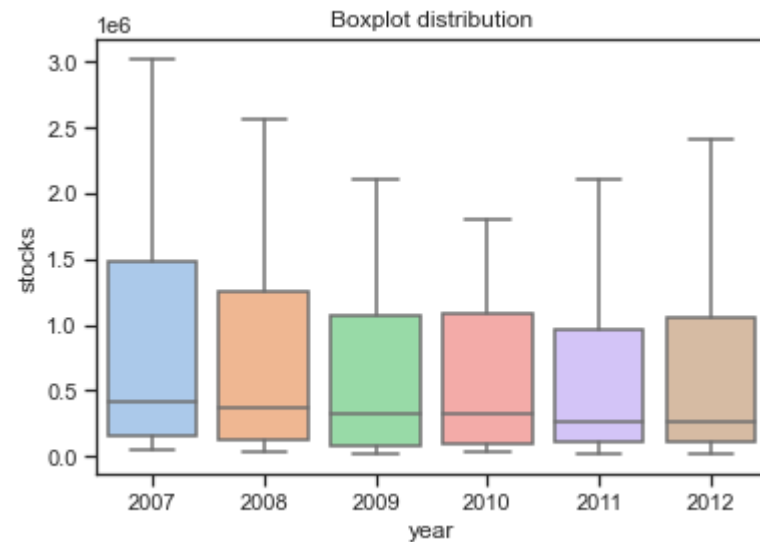
```

In [38]: ax = sns.boxplot(x="year", y="stocks", data=newhoneyproduction, showfliers=False).set(Title='Boxplot distribution')
#How to add a title for seaborn
a=plt.figure(figsize = (10,40))

```

<ipython-input-38-ad7da7350316>:1: MatplotlibDeprecationWarning: Case-insensitive properties were deprecated in 3.3 and support will be removed two minor releases later

```
ax = sns.boxplot(x="year", y="stocks", data=newhoneyproduction, showfliers=False).set(Title='Boxplot distribution')
```



<Figure size 720x2880 with 0 Axes>

```
In [25]: newhoneyproduction2 = pd.read_csv("honeyproduction.csv")
```

```
In [32]: newhoneyproduction2
```

```
Out[32]:
```

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
0	AL	16000.0	71	1136000.0	159000.0	0.72	818000.0	1998
1	AZ	55000.0	60	3300000.0	1485000.0	0.64	2112000.0	1998
2	AR	53000.0	65	3445000.0	1688000.0	0.59	2033000.0	1998
3	CA	450000.0	83	37350000.0	12326000.0	0.62	23157000.0	1998
4	CO	27000.0	72	1944000.0	1594000.0	0.70	1361000.0	1998
...
621	VA	4000.0	41	164000.0	23000.0	3.77	618000.0	2012
622	WA	62000.0	41	2542000.0	1017000.0	2.38	6050000.0	2012
623	WV	6000.0	48	288000.0	95000.0	2.91	838000.0	2012
624	WI	60000.0	69	4140000.0	1863000.0	2.05	8487000.0	2012

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
625	WY	50000.0	51	2550000.0	459000.0	1.87	4769000.0	2012

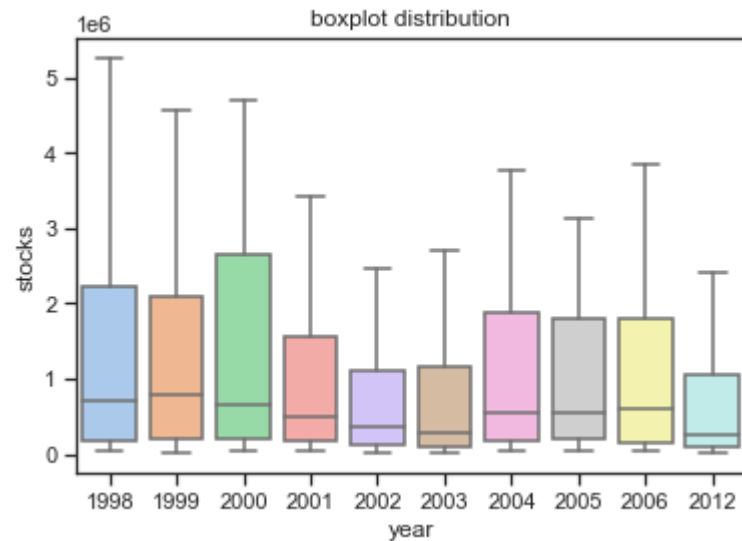
424 rows × 8 columns

```
In [33]: #Delete columns with years 2007:2012 and show a contrast of the graphs and indicate that there is a curve (up and down of the stock price)
#Also, data appears to get more stabilized as years went on.
```

```
In [34]: #Drop the years from 2007 to 2012
twoindex_names = newhoneyproduction2[newhoneyproduction2['year'] == 2007].index
twoindex_names1 = newhoneyproduction2[newhoneyproduction2['year'] == 2008].index
twoindex_names2 = newhoneyproduction2[newhoneyproduction2['year'] == 2009].index
twoindex_names3 = newhoneyproduction2[newhoneyproduction2['year'] == 2010].index
twoindex_names4 = newhoneyproduction2[newhoneyproduction2['year'] == 2011].index
twoindex_names5 = newhoneyproduction2[newhoneyproduction2['year'] == 2012].index
```

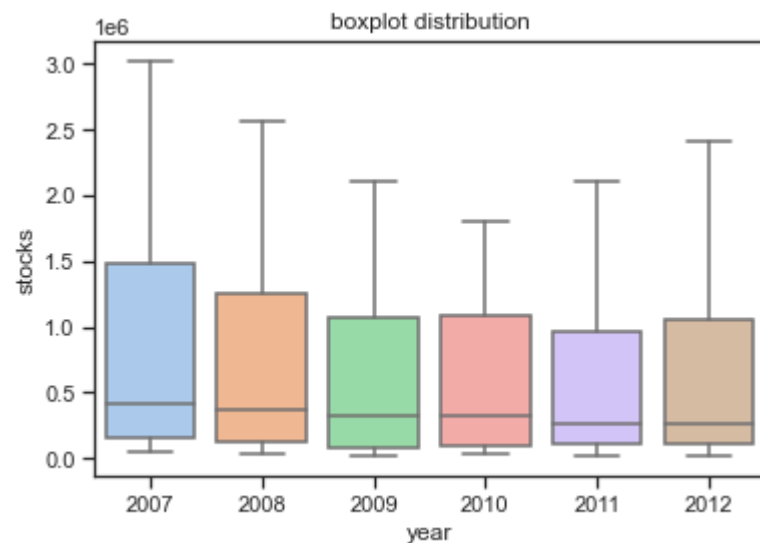
```
In [35]: newhoneyproduction2.drop(twoindex_names, inplace = True)
newhoneyproduction2.drop(twoindex_names1, inplace = True)
newhoneyproduction2.drop(twoindex_names2, inplace = True)
newhoneyproduction2.drop(twoindex_names3, inplace = True)
newhoneyproduction2.drop(twoindex_names4, inplace = True)
newhoneyproduction2.drop(twoindex_names5, inplace = True)
```

```
In [36]: showboxplotproduction2 = sns.boxplot(x="year", y="stocks", data=newhoneyproduction2, showfliers=False).set(title="boxplot distribution of stocks by year")
plt.figure(figsize = (10,40))
```

<Figure size 720x2880 with 0 Axes>

```
In [37]: showboxplotproduction1= sns.boxplot(x="year", y="stocks", data=newhoneyproduction, showfliers=False).set(title="boxplot distributio
a=plt.figure(figsize = (10,40))
```



<Figure size 720x2880 with 0 Axes>

```
In [38]: #Result analysis: The result indicates that the maximum value from 1998 to 2012 and 2007 to 2012 decreased. This indicates that t
#has been decreasing as time went by. So it is true with the previous assumption from the website Kaggle that as years went on,
```

```
#as years went on, the stock decreased and the possible reason is the decrease in the population of bees as mentioned by the Kaggl  
#The decrease in the stockprice definitely validates that.
```