In [2]:

```
import pandas as pd
import numpy as np

import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
```

In [4]:

```
data = pd.read_csv('honeyproduction.csv')
data.head(300)
```

Out[4]:

	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
...	...	...	...	...	...	...	...	...
295	UT	24000.0	70	1680000.0	554000.0	1.10	1848000.0	2004
296	VT	6000.0	68	408000.0	192000.0	1.51	616000.0	2004
297	VA	7000.0	38	266000.0	69000.0	2.10	559000.0	2004
298	WA	56000.0	63	3528000.0	1376000.0	0.98	3457000.0	2004
299	WV	9000.0	55	495000.0	183000.0	1.41	698000.0	2004

300 rows × 8 columns

In [6]:

```
data.tail()
```

Out[6]:

	state	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
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

In [11]:

```
data.describe()
```

Out[11]:

	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year

	numcol	yieldpercol	totalprod	stocks	priceperlb	prodvalue	year
count	626.000000	626.000000	6.260000e+02	6.260000e+02	626.000000	6.260000e+02	626.000000
mean	60284.345048	62.009585	4.169086e+06	1.318859e+06	1.409569	4.715741e+06	2004.864217
std	91077.087231	19.458754	6.883847e+06	2.272964e+06	0.638599	7.976110e+06	4.317306
min	2000.000000	19.000000	8.400000e+04	8.000000e+03	0.490000	1.620000e+05	1998.000000
25%	9000.000000	48.000000	4.750000e+05	1.430000e+05	0.932500	7.592500e+05	2001.000000
50%	26000.000000	60.000000	1.533000e+06	4.395000e+05	1.360000	1.841500e+06	2005.000000
75%	63750.000000	74.000000	4.175250e+06	1.489500e+06	1.680000	4.703250e+06	2009.000000
max	510000.000000	136.000000	4.641000e+07	1.380000e+07	4.150000	6.961500e+07	2012.000000

In [12]:
```
data.isnull().sum()
```

Out[12]:
```
state          0
numcol         0
yieldpercol    0
totalprod      0
stocks         0
priceperlb     0
prodvalue      0
year           0
dtype: int64
```

In [79]:
```
type(data)
```

Out[79]:
```
pandas.core.frame.DataFrame
```

In [80]:
```
data.shape
```

Out[80]:
```
(626, 8)
```

In [81]:
```
data.columns
```

Out[81]:
```
Index(['state', 'numcol', 'yieldpercol', 'totalprod', 'stocks', 'priceperlb',
       'prodvalue', 'year'],
      dtype='object')
```

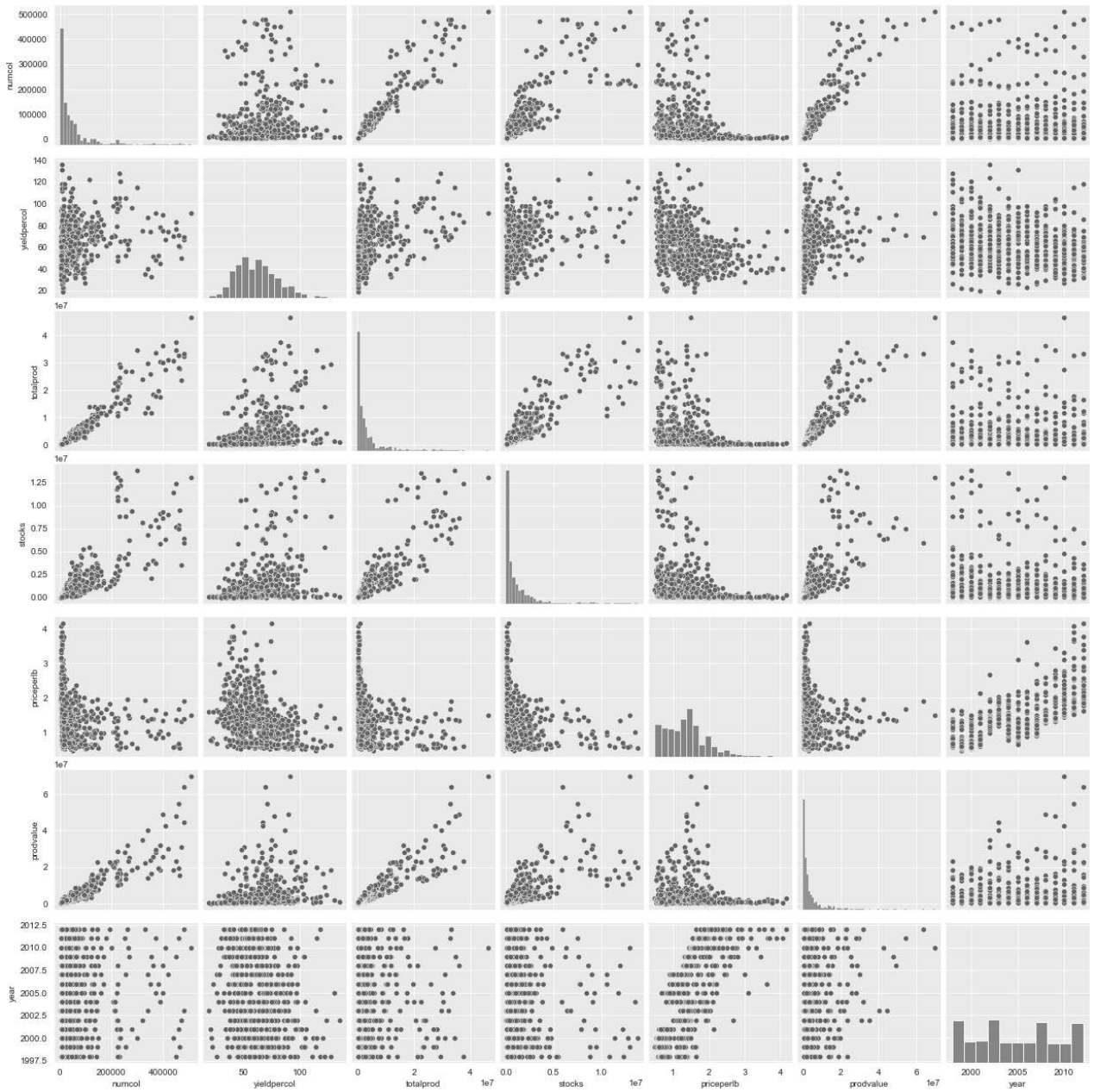
In [14]:
```
data.dtypes
```

Out[14]:
```
state           object
numcol         float64
yieldpercol      int64
totalprod      float64
stocks         float64
priceperlb     float64
prodvalue      float64
year             int64
dtype: object
```

In [15]:
```
data['numcol']
```

Out[15]:
```
0       16000.0
1       55000.0
2       53000.0
3      450000.0
4       27000.0
         ...
621      4000.0
622     62000.0
623      6000.0
624     60000.0
625     50000.0
Name: numcol, Length: 626, dtype: float64
```

In [23]:
```
# 2. Finding relationship between numerical variables using pair plot
sns.pairplot(data);
```

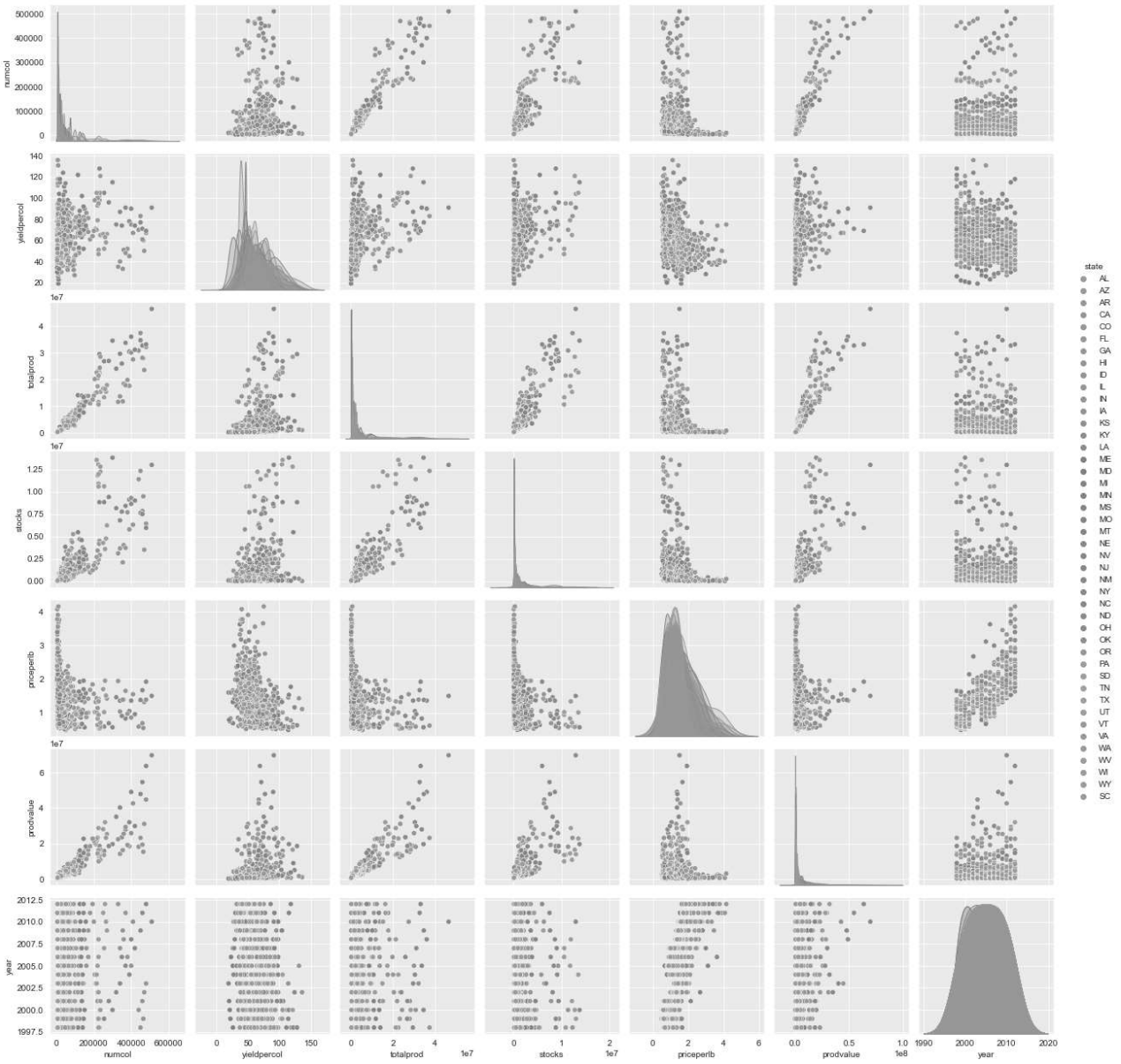


In [25]:

```
# 2. Finding relationship between numerical variables using pair plot
sns.pairplot(data, hue='state', palette="husl")
```

Out[25]:

```
<seaborn.axisgrid.PairGrid at 0x20455078d90>
```

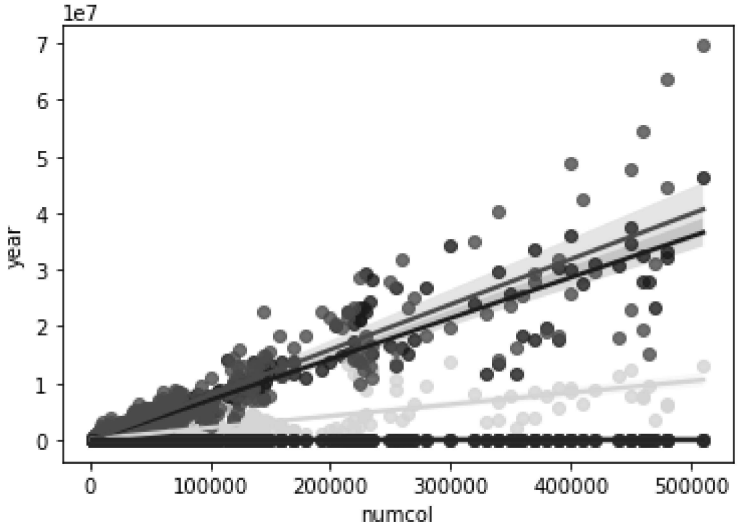


In [30]:

```
# 3. Correlation plots
sns.regplot(x=data["numcol"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["numcol"], y=data["totalprod"],color="blue")
sns.regplot(x=data["numcol"], y=data["stocks"],color="pink")
sns.regplot(x=data["numcol"], y=data["priceperlb"],color="purple")
sns.regplot(x=data["numcol"], y=data["prodvalue"],color="green")
sns.regplot(x=data["numcol"], y=data["year"],color="maroon")
```

Out[30]:

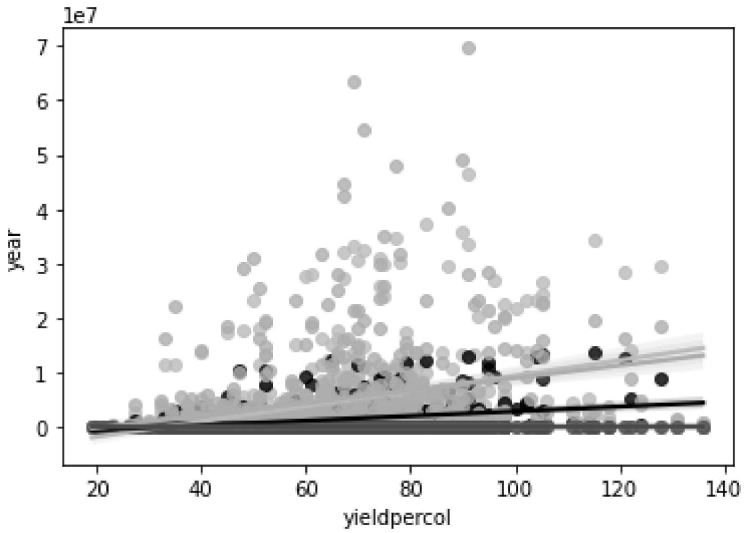
```
<AxesSubplot:xlabel='numcol', ylabel='year'>
```



In [29]:

```
sns.regplot(x=data["yieldpercol"], y=data["numcol"],color="red")
sns.regplot(x=data["yieldpercol"], y=data["totalprod"],color="skyblue")
sns.regplot(x=data["yieldpercol"], y=data["stocks"],color="black")
sns.regplot(x=data["yieldpercol"], y=data["priceperlb"],color="gold")
sns.regplot(x=data["yieldpercol"], y=data["prodvalue"],color="orange")
sns.regplot(x=data["yieldpercol"], y=data["year"],color="green")
```

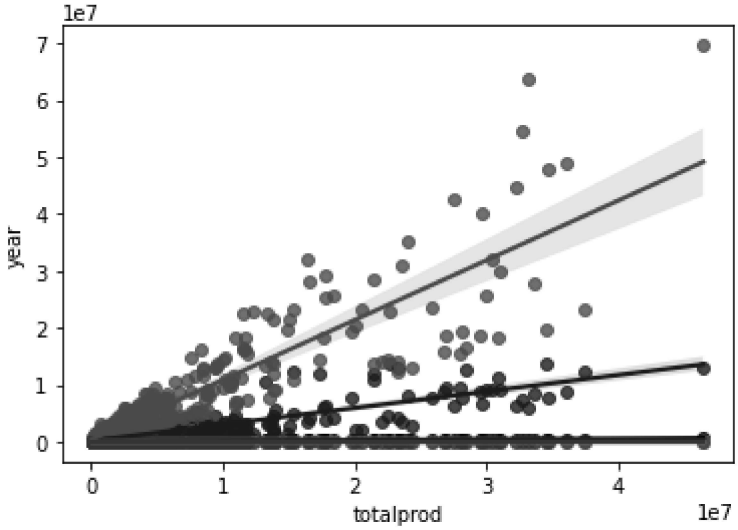
Out[29]: <AxesSubplot:xlabel='yieldpercol', ylabel='year'>



In [28]:

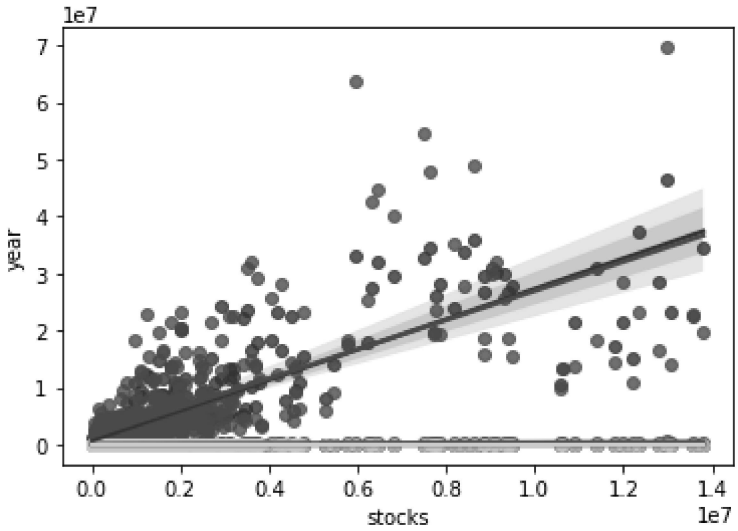
```
sns.regplot(x=data["totalprod"], y=data["numcol"],color="black")
sns.regplot(x=data["totalprod"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["totalprod"], y=data["stocks"],color="blue")
sns.regplot(x=data["totalprod"], y=data["priceperlb"],color="gold")
sns.regplot(x=data["totalprod"], y=data["prodvalue"],color="green")
sns.regplot(x=data["totalprod"], y=data["year"],color="purple")
```

Out[28]: <AxesSubplot:xlabel='totalprod', ylabel='year'>



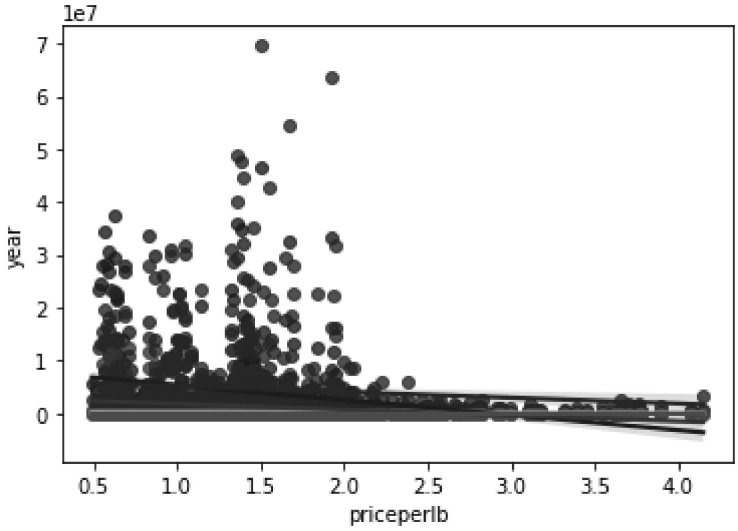
In [18]:
```
sns.regplot(x=data["stocks"], y=data["numcol"],color="red")
sns.regplot(x=data["stocks"], y=data["yieldpercol"],color="brown")
sns.regplot(x=data["stocks"], y=data["totalprod"],color="purple")
sns.regplot(x=data["stocks"], y=data["priceperlb"],color="black")
sns.regplot(x=data["stocks"], y=data["prodvalue"],color="green")
sns.regplot(x=data["stocks"], y=data["year"],color="gold")
```

Out[18]:
```
<AxesSubplot:xlabel='stocks', ylabel='year'>
```



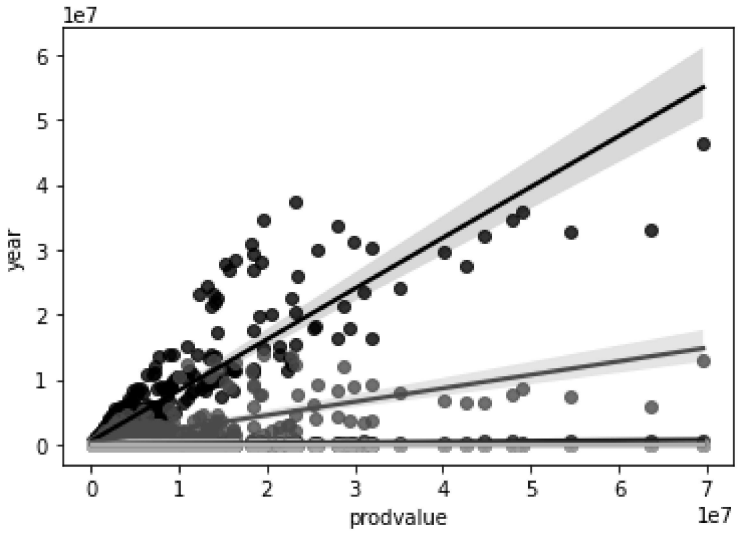
In [19]:
```
sns.regplot(x=data["priceperlb"], y=data["numcol"],color="gold")
sns.regplot(x=data["priceperlb"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["priceperlb"], y=data["totalprod"],color="blue")
sns.regplot(x=data["priceperlb"], y=data["stocks"],color="purple")
sns.regplot(x=data["priceperlb"], y=data["prodvalue"],color="maroon")
sns.regplot(x=data["priceperlb"], y=data["year"],color="green")
```

Out[19]:
```
<AxesSubplot:xlabel='priceperlb', ylabel='year'>
```



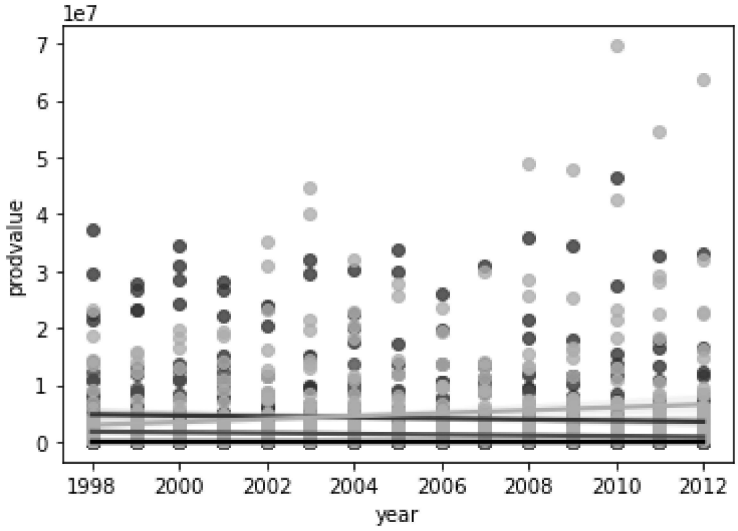
In [12]:
```
sns.regplot(x=data["prodvalue"], y=data["numcol"],color="blue")
sns.regplot(x=data["prodvalue"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["prodvalue"], y=data["totalprod"],color="black")
sns.regplot(x=data["prodvalue"], y=data["priceperlb"],color="purple")
sns.regplot(x=data["prodvalue"], y=data["stocks"],color="green")
sns.regplot(x=data["prodvalue"], y=data["year"],color="orange")
```

Out[12]: <AxesSubplot:xlabel='prodvalue', ylabel='year'>



In [10]:
```
sns.regplot(x=data["year"], y=data["numcol"],color="blue")
sns.regplot(x=data["year"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["year"], y=data["totalprod"],color="purple")
sns.regplot(x=data["year"], y=data["priceperlb"],color="black")
sns.regplot(x=data["year"], y=data["stocks"],color="green")
sns.regplot(x=data["year"], y=data["prodvalue"],color="orange")
```

Out[10]: <AxesSubplot:xlabel='year', ylabel='prodvalue'>



In [27]:

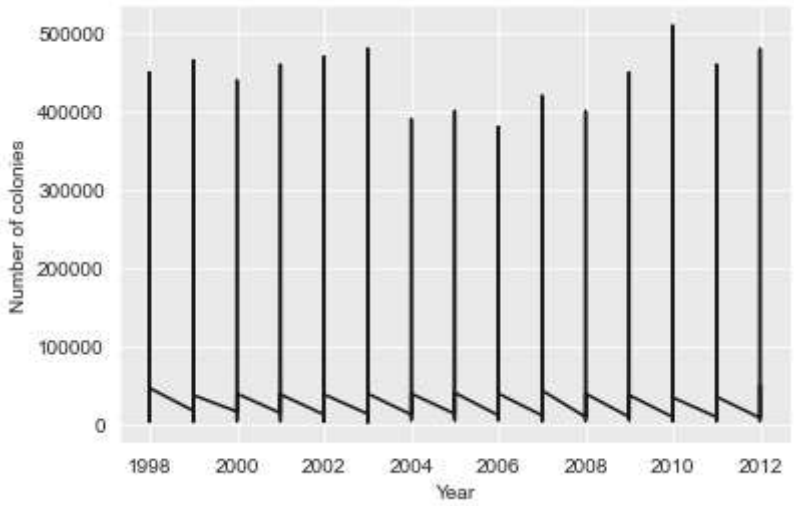
```
# 3.
a=np.array(data["state"])
print(a)
```

```
['AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY'
 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NV' 'NJ' 'NM' 'NY' 'NC'
 'ND' 'OH' 'OK' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI'
 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS'
 'KY' 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NV' 'NJ' 'NM' 'NY'
 'NC' 'ND' 'OH' 'OK' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV'
 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA'
 'KS' 'KY' 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NV' 'NJ' 'NM'
 'NY' 'NC' 'ND' 'OH' 'OK' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA'
 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN'
 'IA' 'KS' 'KY' 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NV' 'NJ'
 'NM' 'NY' 'NC' 'ND' 'OH' 'OK' 'OR' 'PA' 'SC' 'SD' 'TN' 'TX' 'UT' 'VT'
 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID'
 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE'
 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OK' 'OR' 'PA' 'SC' 'SD' 'TN' 'TX'
 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA'
 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MD' 'MI' 'MN' 'MS' 'MO'
 'MT' 'NE' 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OK' 'OR' 'PA' 'SC' 'SD'
 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO'
 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS'
 'MO' 'MT' 'NE' 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN'
 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL'
 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO'
 'MT' 'NE' 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX'
 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA'
 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT'
 'NE' 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT'
 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI'
 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE'
 'NV' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT'
 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID'
 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NV'
 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA'
 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL'
 'IN' 'IA' 'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NJ' 'NM'
 'NY' 'NC' 'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV'
 'WI' 'WY' 'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA'
```

```
'KS' 'KY' 'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NJ' 'NM' 'NY' 'NC'
'ND' 'OH' 'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY'
'AL' 'AZ' 'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY'
'LA' 'ME' 'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH'
'OR' 'PA' 'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY' 'AL' 'AZ'
'AR' 'CA' 'CO' 'FL' 'GA' 'HI' 'ID' 'IL' 'IN' 'IA' 'KS' 'KY' 'LA' 'ME'
'MI' 'MN' 'MS' 'MO' 'MT' 'NE' 'NJ' 'NM' 'NY' 'NC' 'ND' 'OH' 'OR' 'PA'
'SD' 'TN' 'TX' 'UT' 'VT' 'VA' 'WA' 'WV' 'WI' 'WY']
```

In [74]:

```
#5. Variation in number of colonies over the year
#This graph shows maximum no of colonies were in 2010 and the steep line between each y
#1999 to 2000 shows that they also decrease a little bit between every year.
plt.plot(data['year'],data['numcol'],color='blue')
plt.ylabel("Number of colonies")
plt.xlabel("Year")
plt.show()
```

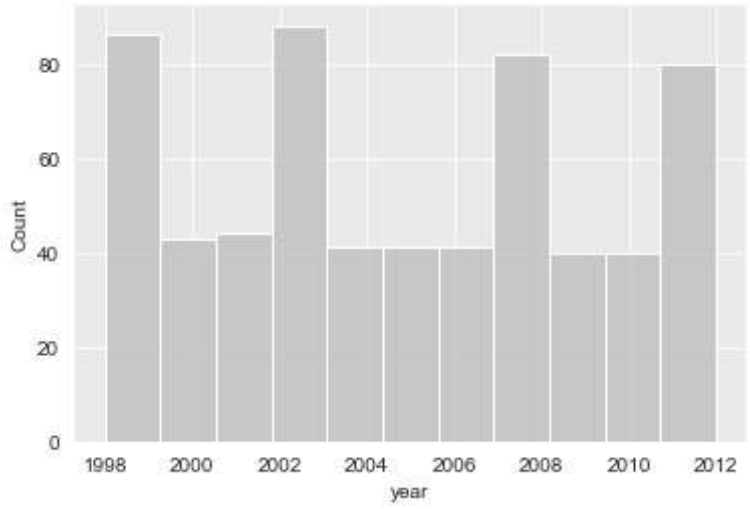


In [54]:

```
#This graph shows that overall trend of honey production was highest in 2002 followed b
sns.histplot(data, x="year", color="skyblue")
```

Out[54]:

```
<AxesSubplot:xlabel='year', ylabel='Count'>
```

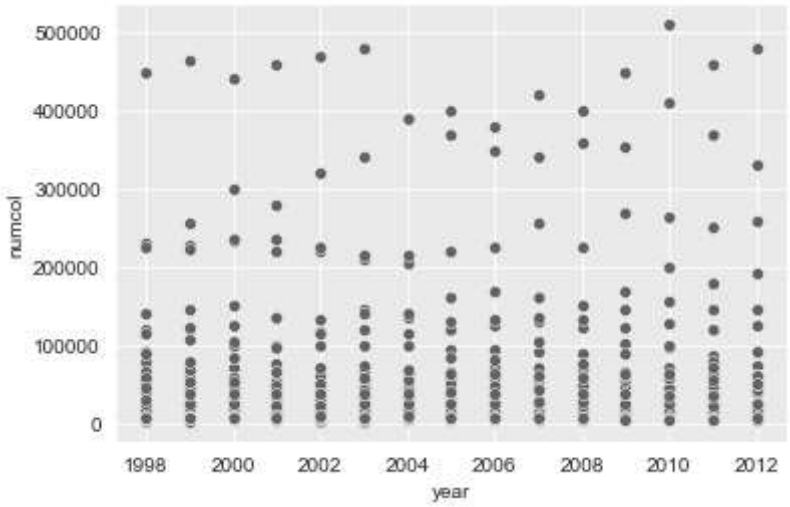


In [73]:

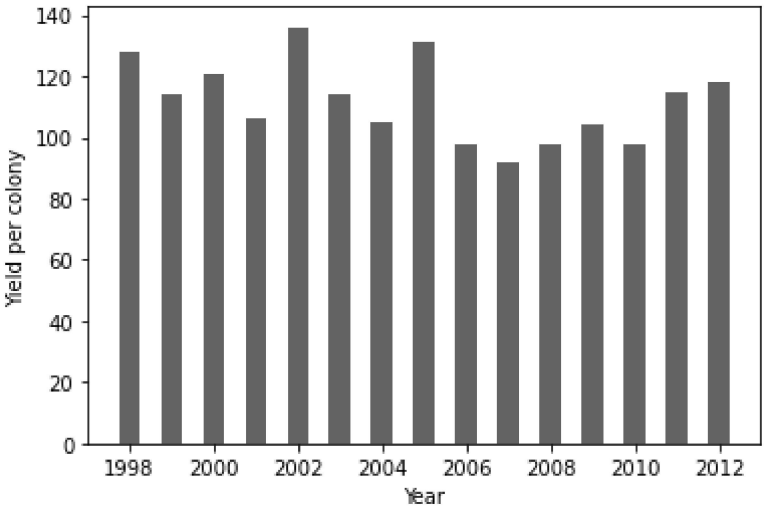
```
#This graph shows variation in number of colonies after every 2 years.And It shows that
#increase mostly from  a range of 0 to 100000. They also increase till 500000 but its v
```

```
sns.scatterplot(data['year'],data['numcol'])
```

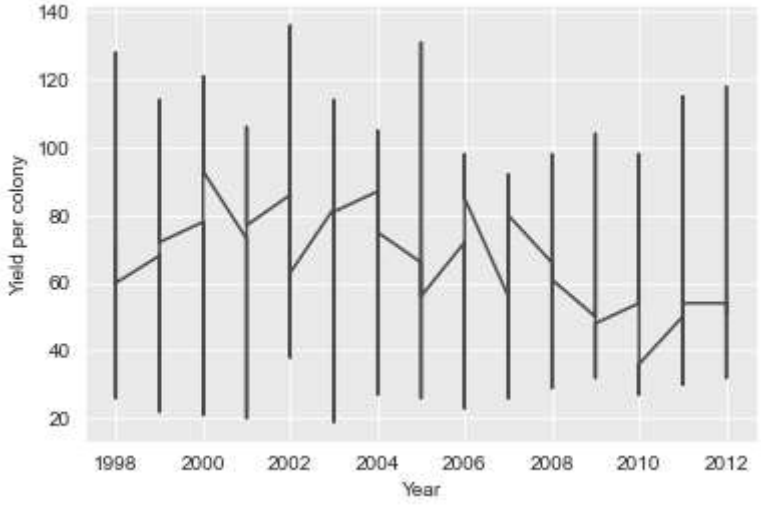
Out[73]: <AxesSubplot:xlabel='year', ylabel='numcol'>



In [6]:
```
#This graph shows that every year a minimum of 80 yield per colony is produced.
plt.bar(data['year'],data['yieldpercol'],width=0.5)
plt.xlabel("Year")
plt.ylabel("Yield per colony")
plt.show()
```

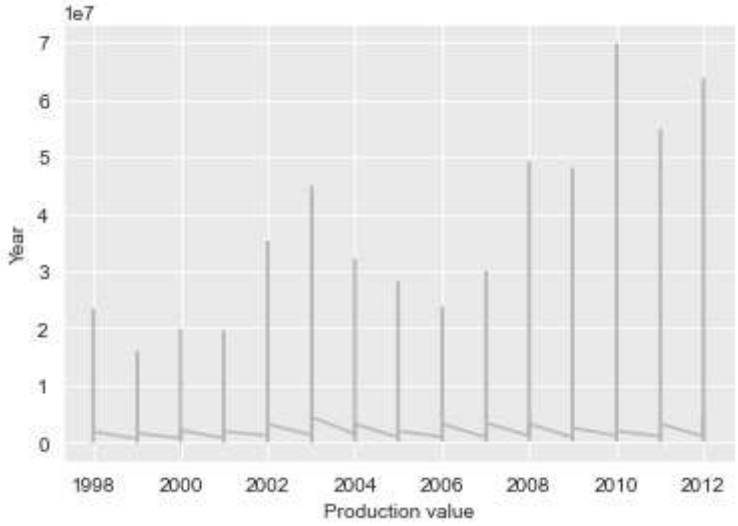


In [77]:
```
plt.plot(data['year'],data['yieldpercol'],color='red')
plt.ylabel("Yield per colony")
plt.xlabel("Year")
plt.show()
```



In [78]:

```
plt.plot(data['year'],data['prodvalue'],color='skyblue')
plt.xlabel("Production value")
plt.ylabel("Year")
plt.show()
```

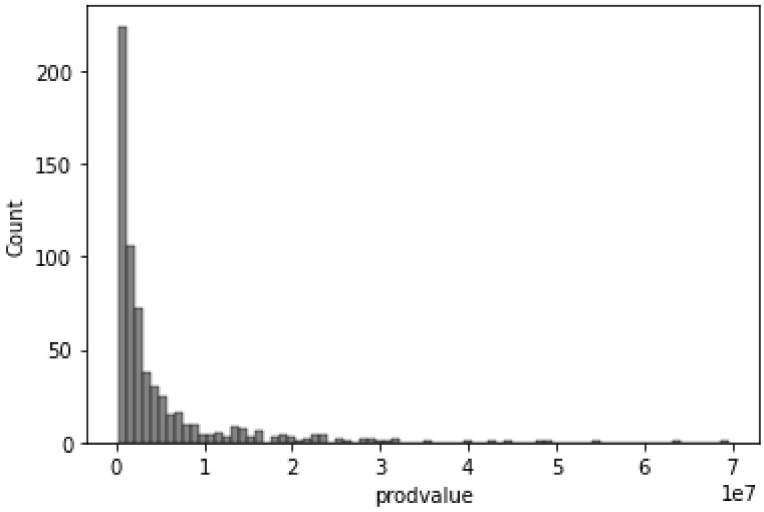


In [25]:

```
#7.
sns.histplot(data, x="prodvalue", color="teal")
```

Out[25]:

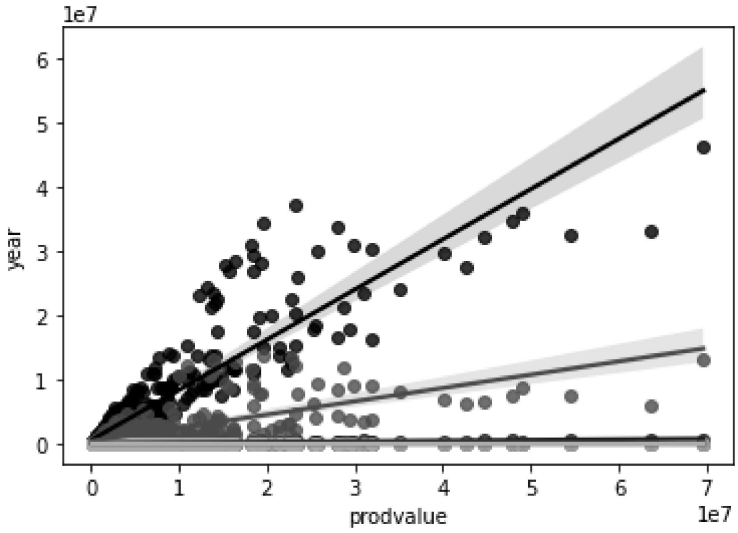
```
<AxesSubplot:xlabel='prodvalue', ylabel='Count'>
```



In [26]:

```
sns.regplot(x=data["prodvalue"], y=data["numcol"],color="blue")
sns.regplot(x=data["prodvalue"], y=data["yieldpercol"],color="red")
sns.regplot(x=data["prodvalue"], y=data["totalprod"],color="black")
sns.regplot(x=data["prodvalue"], y=data["priceperlb"],color="purple")
sns.regplot(x=data["prodvalue"], y=data["stocks"],color="green")
sns.regplot(x=data["prodvalue"], y=data["year"],color="orange")
```

Out[26]: <AxesSubplot:xlabel='prodvalue', ylabel='year'>



In [ ]: