To visualize how honey production is changed over the years (1998-2021) in the United States.

Import the libraries

In [1]:
```
import numpy as np  # matmatical computation
import pandas as pd  # data preprocessing
import matplotlib.pyplot as plt  # visualization
import seaborn as sns  # visualization
```

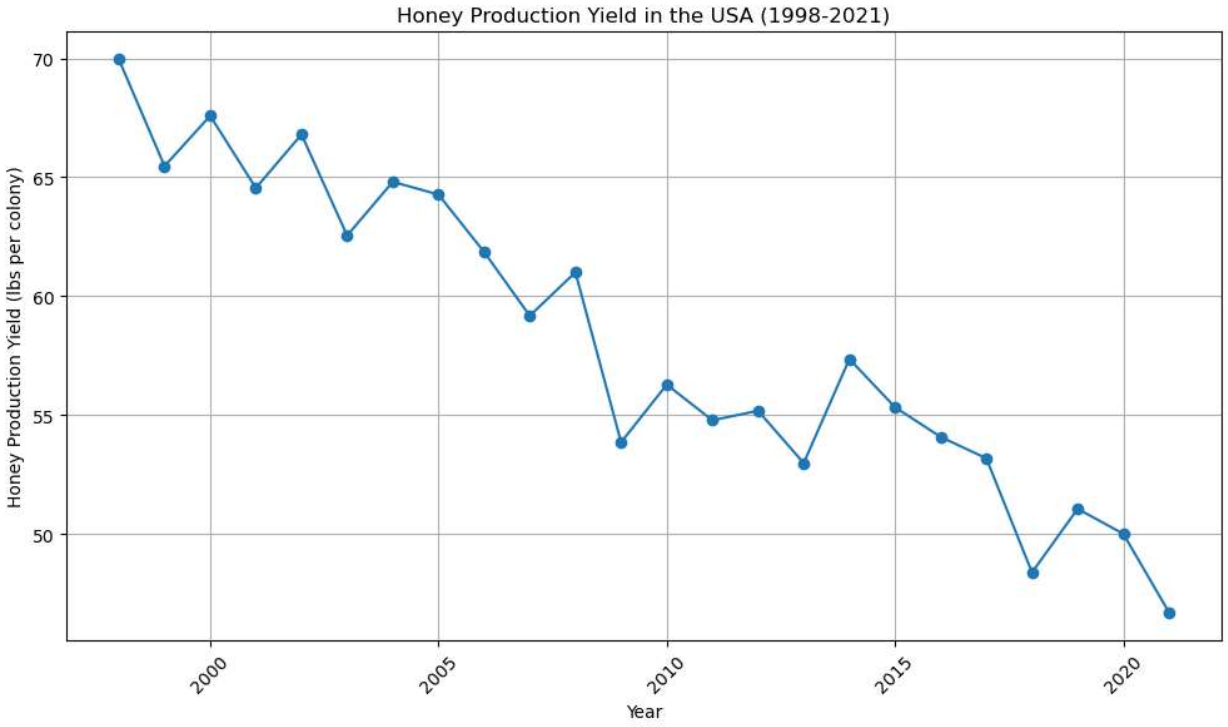
Feature Description

In [3]:
```
#Dataset:
#state: Various states of the U.S.
#numcol: Number of honey-producing colonies. Honey producing colonies are the maximum
#yieldpercol: Honey yield per colony. Unit is pounds
#totalprod: Total production (numcol x yieldpercol). Unit is pounds
#stocks: Refers to stocks held by producers. Unit is pounds
#priceperlb: Refers to average price per pound based on expanded sales.The unit is dol
#prodvalue: Value of production (totalprod x priceperlb). The unit is dollars. year: Y
```

In [24]:
```
# Load the dataset
df = pd.read_csv('honeyproduction 1998-2021.csv')
# Convert column names to lowercase
df.columns = df.columns.str.lower()
```

1. How has honey production yield changed from 1998 to 2021?

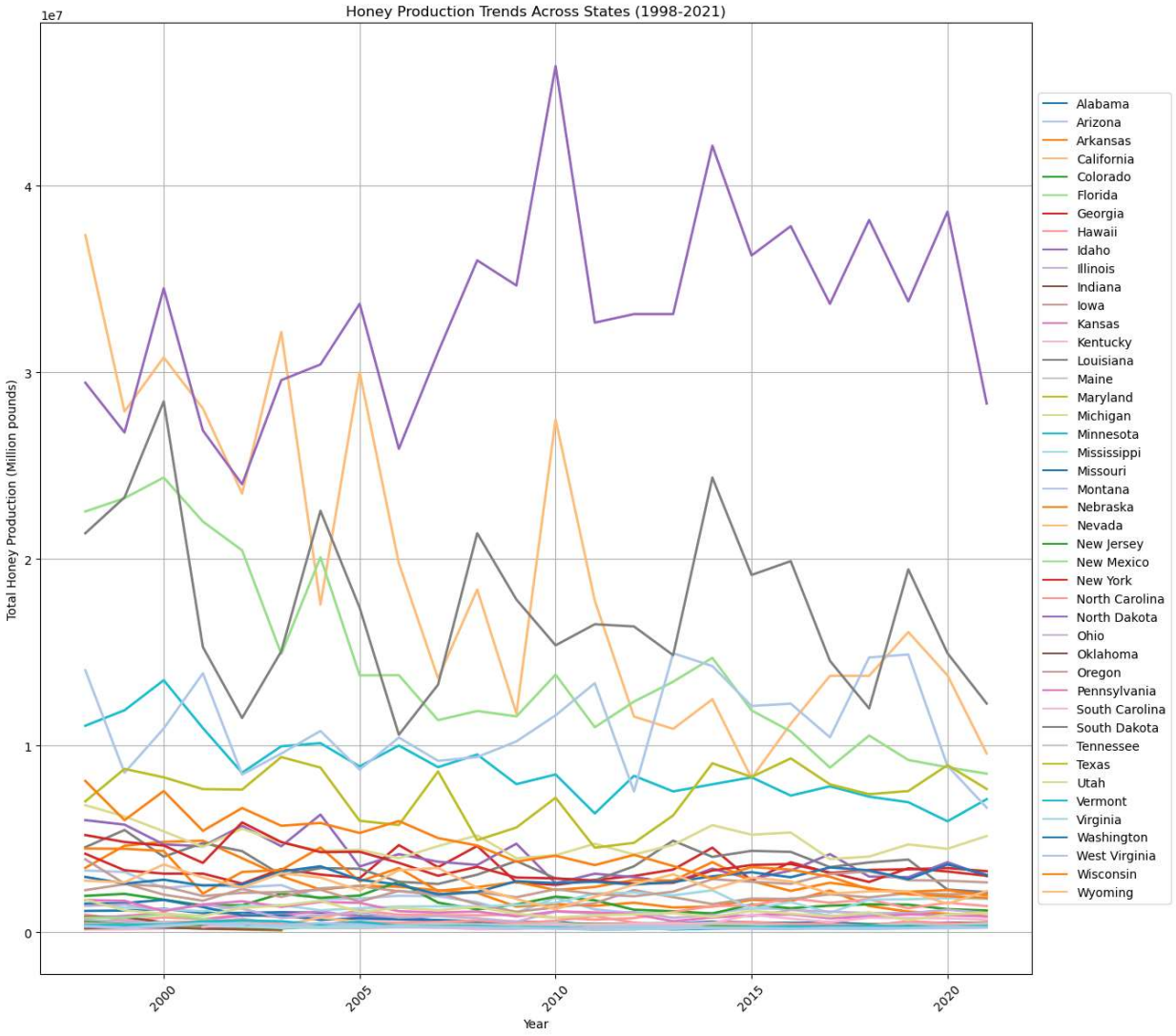
In [22]:
```
# Calculate the average honey production yield for each year
yearly_yield = df.groupby('year')['yieldpercol'].mean().reset_index()

# Plot the line chart
plt.figure(figsize=(10, 6))
plt.plot(yearly_yield['year'], yearly_yield['yieldpercol'], marker='o', linestyle='-')
plt.xlabel('Year')
plt.ylabel('Honey Production Yield (lbs per colony)')
plt.title('Honey Production Yield in the USA (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



2. Over time, what are the major production trends across the states?

In [25]:
```
# Check the column names to make sure the 'state' column is present and spelled correc
print(df.columns)

# Group the data by 'state' and 'year' and calculate the total honey production for ea
state_yearly_production = df.groupby(['state', 'year'])['totalprod'].sum().reset_index

# Plot the line chart using Seaborn to show the trends for each state
plt.figure(figsize=(12, 12))
sns.lineplot(data=state_yearly_production, x='year', y='totalprod', hue='state', palet
plt.xlabel('Year')
plt.ylabel('Total Honey Production (Million pounds)')
plt.title('Honey Production Trends Across States (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.legend(loc='center left', bbox_to_anchor=(1, 0.5))
plt.show()
```

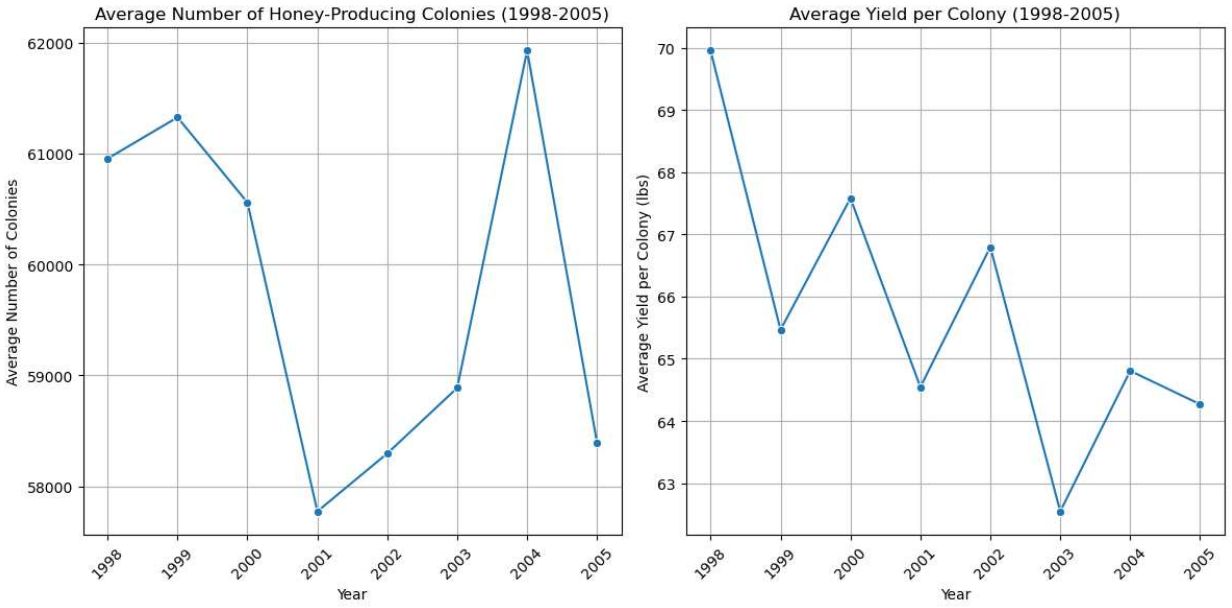
```
Index(['state', 'numcol', 'yieldpercol', 'totalprod', 'stocks', 'priceperlb',
       'prodvalue', 'year'],
      dtype='object')
```



3. Does the data show any trends in terms of the number of honeyproducing colonies and yield per colony before 2006, which was when concern over Colony Collapse Disorder spread nationwide?

In [26]:
```
# Filter the data for years before 2006
df_before_2006 = df[df['year'] < 2006]

# Group the data by year and calculate the average number of honey-producing colonies
yearly_colonies_avg = df_before_2006.groupby('year')['numcol'].mean().reset_index()
yearly_yield_avg = df_before_2006.groupby('year')['yieldpercol'].mean().reset_index()

# Plot the line charts
plt.figure(figsize=(12, 6))

# Plot the average number of honey-producing colonies
plt.subplot(1, 2, 1)
sns.lineplot(data=yearly_colonies_avg, x='year', y='numcol', marker='o', linestyle='-
plt.xlabel('Year')
plt.ylabel('Average Number of Colonies')
plt.title('Average Number of Honey-Producing Colonies (1998-2005)')
plt.grid(True)
plt.xticks(rotation=45)
```

```
# Plot the average yield per colony
plt.subplot(1, 2, 2)
sns.lineplot(data=yearly_yield_avg, x='year', y='yieldpercol', marker='o', linestyle=
plt.xlabel('Year')
plt.ylabel('Average Yield per Colony (lbs)')
plt.title('Average Yield per Colony (1998-2005)')
plt.grid(True)
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```

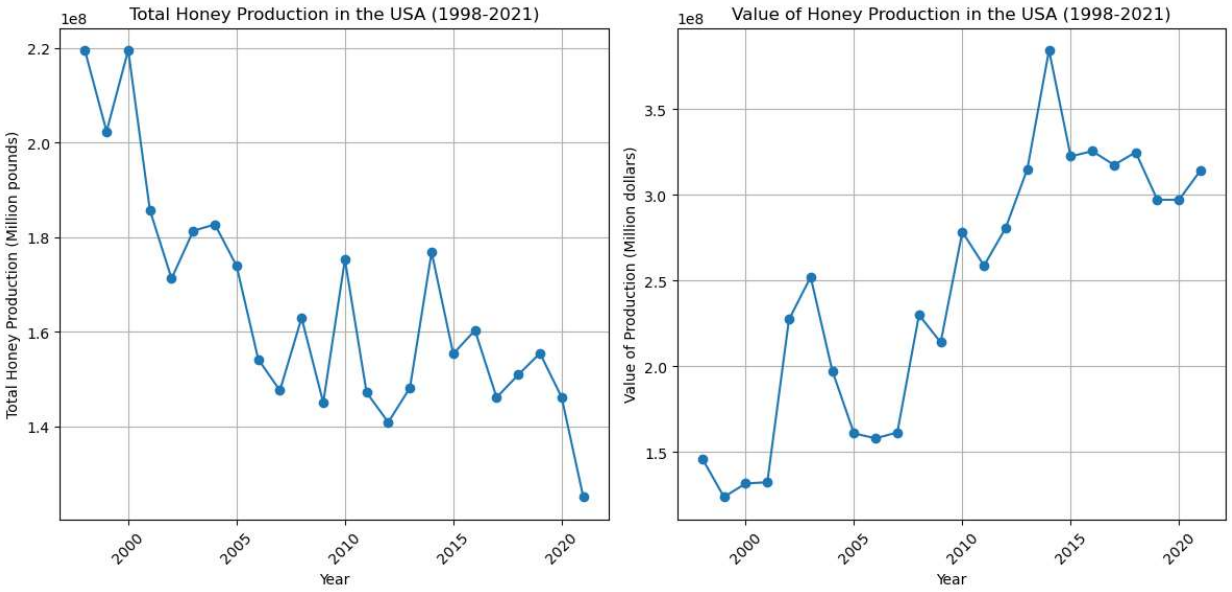


4. Are there any patterns that can be observed between total honey production and value of production every year?

In [27]:

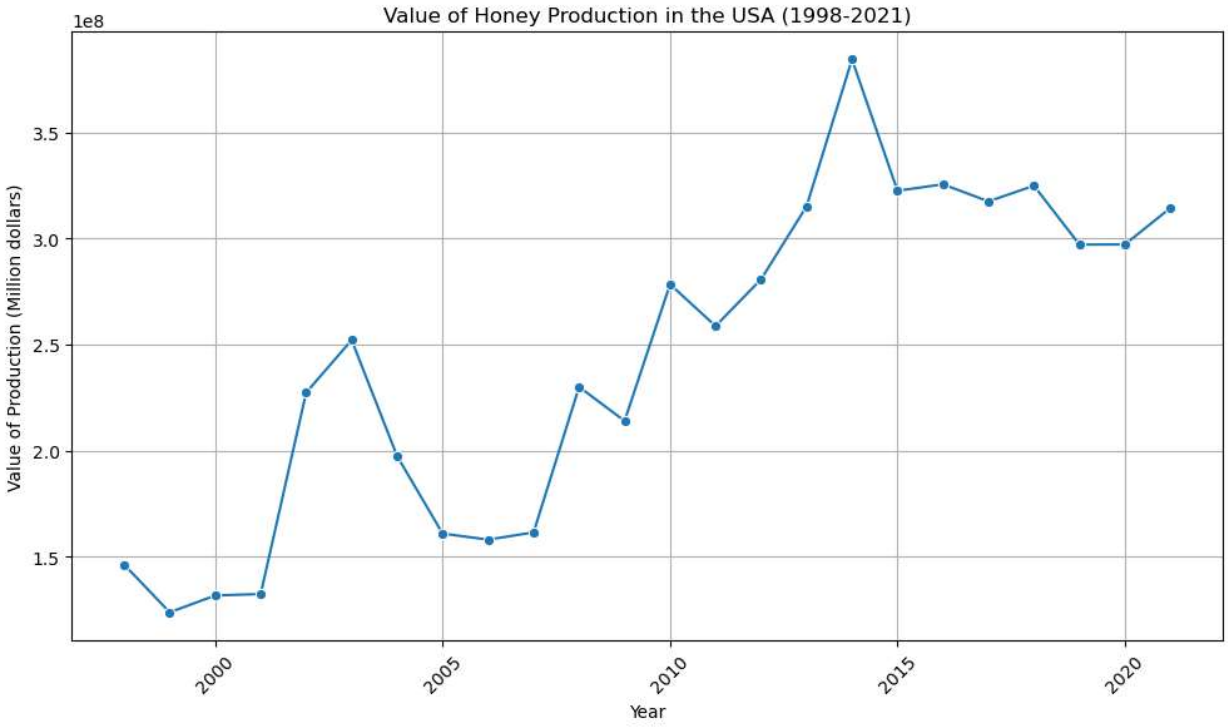
```
# Group the data by year and calculate the total honey production and value of product
yearly_production = df.groupby('year')['totalprod'].sum().reset_index()
yearly_value = df.groupby('year')['prodvalue'].sum().reset_index()

# Plot the line charts side by side for comparison
plt.figure(figsize=(12, 6))

# Plot the total honey production
plt.subplot(1, 2, 1)
plt.plot(yearly_production['year'], yearly_production['totalprod'], marker='o', linest
plt.xlabel('Year')
plt.ylabel('Total Honey Production (Million pounds)')
plt.title('Total Honey Production in the USA (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)

# Plot the value of production
plt.subplot(1, 2, 2)
plt.plot(yearly_value['year'], yearly_value['prodvalue'], marker='o', linestyle='-')
plt.xlabel('Year')
plt.ylabel('Value of Production (Million dollars)')
plt.title('Value of Honey Production in the USA (1998-2021)')
plt.grid(True)
```

```
plt.xticks(rotation=45)

plt.tight_layout()
plt.show()
```



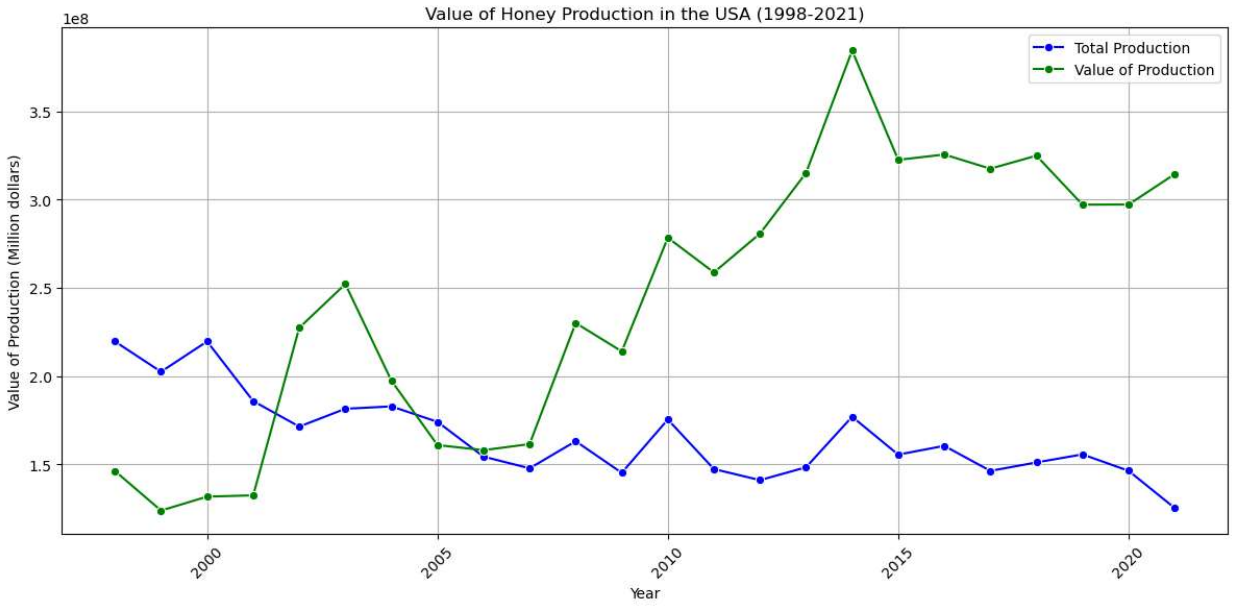
5. How has the value of production, which in some sense could be tied to demand, changed every year?

In [38]:
```
# Group the data by year and calculate the total value of production per year
yearly_value_production = df.groupby('year')['prodvalue'].sum().reset_index()

# Plot the line chart
plt.figure(figsize=(10, 6))
sns.lineplot(data=yearly_value_production, x='year', y='prodvalue', marker='o', linest
plt.xlabel('Year')
plt.ylabel('Value of Production (Million dollars)')
plt.title('Value of Honey Production in the USA (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



6. Constructs the related plots using Seaborn and Matplot apply customization and derive insights from the visualization.

In [39]:
```
# Group the data by year and calculate the total honey production and value of product
yearly_production = df.groupby('year')['totalprod'].sum().reset_index()
yearly_value_production = df.groupby('year')['prodvalue'].sum().reset_index()

# Plot the line charts side by side with customized styles
plt.figure(figsize=(12, 6))

# Plot the total honey production
sns.lineplot(data=yearly_production, x='year', y='totalprod', marker='o', linestyle='
plt.xlabel('Year')
plt.ylabel('Total Honey Production (Million pounds)')
plt.title('Total Honey Production in the USA (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)
plt.legend()

# Plot the value of production
sns.lineplot(data=yearly_value_production, x='year', y='prodvalue', marker='o', linest
plt.xlabel('Year')
plt.ylabel('Value of Production (Million dollars)')
plt.title('Value of Honey Production in the USA (1998-2021)')
plt.grid(True)
plt.xticks(rotation=45)
plt.legend()

plt.tight_layout()
plt.show()
```


Value of Honey Production in the USA (1998-2021)
1e8
Total Production
Value of Production
Value of Production (Million dollars)
3.5
3.0
2.5
2.0
1.5
2000
2005
2010
2015
2020
Year

In [ ]: