Iris Dataset

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

Inspection on dataset

iris_df = pd.read_csv('iris.csv') iris_df.head()

In [33]:
```
iris_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   sepal_length  150 non-null    float64
 1   sepal_width   150 non-null    float64
 2   petal_length  150 non-null    float64
 3   petal_width   150 non-null    float64
 4   species       150 non-null    object 
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

In [34]:
```
iris_df.describe()
```

Out[34]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.054000	3.758667	1.198667
std	0.828066	0.433594	1.764420	0.763161
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

In [35]:
```
iris_df.head()
```

Out[35]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

Data preprocessing

In [36]:
```
iris_df.columns
```

Out[36]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

In [37]:
```
iris_df.shape
```

Out[37]:
```
(150, 5)
```

In [38]:
```
iris_df.describe(include = 'object')
```

Out[38]:

	species
count	150
unique	3
top	setosa
freq	50

In [39]:
```
iris_df = iris_df.drop_duplicates()
iris_df.shape
```

Out[39]:
```
(147, 5)
```

In [40]:
```
iris_df.isnull().sum()
```

Out[40]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

In [41]:
```
iris_df.isnull().sum()
```

Out[41]:
```
sepal_length    0
sepal_width     0
petal_length    0
petal_width     0
species         0
dtype: int64
```

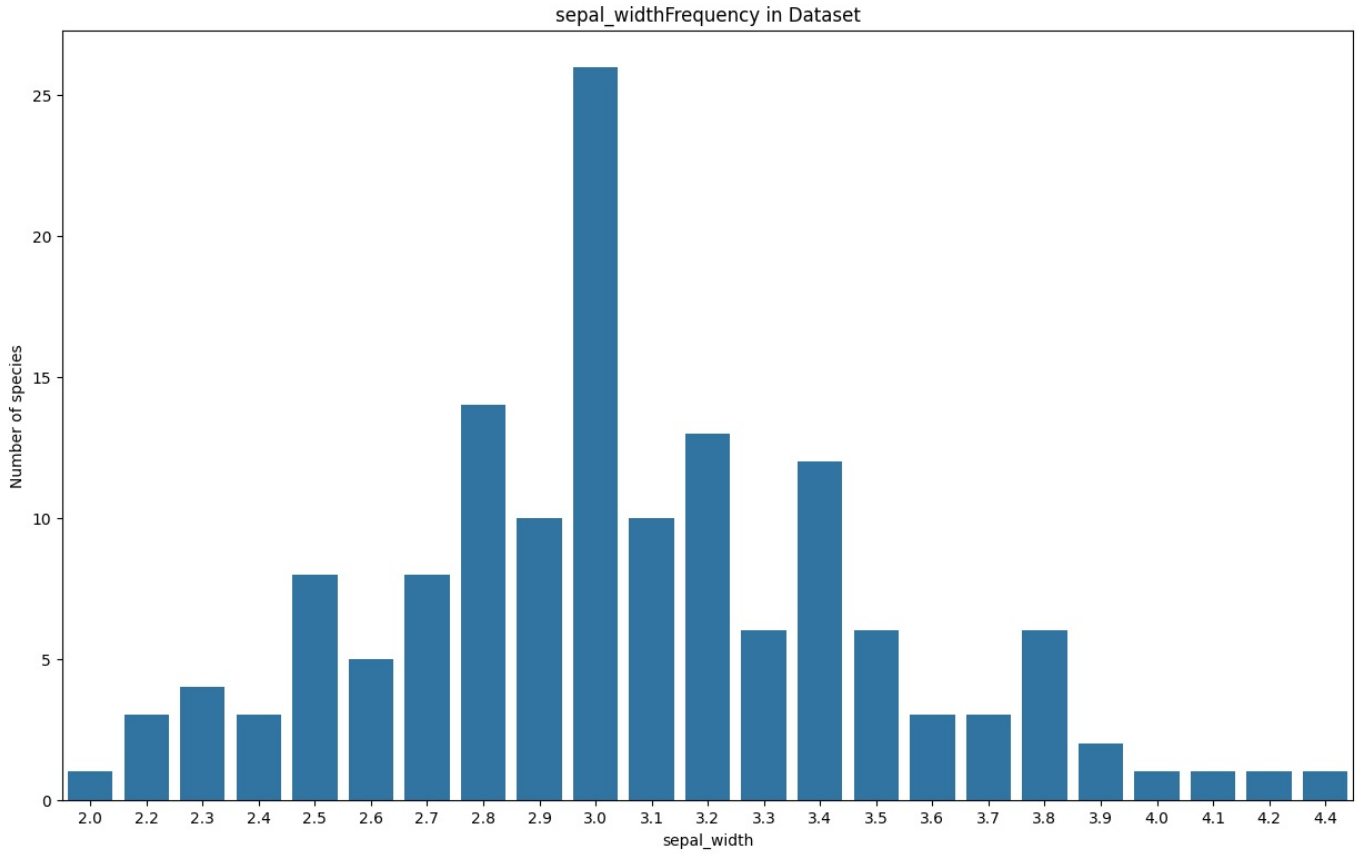
Data visualization

In [42]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [43]:
```
width_counts = iris_df['sepal_width'].value_counts()
```

Bar plot

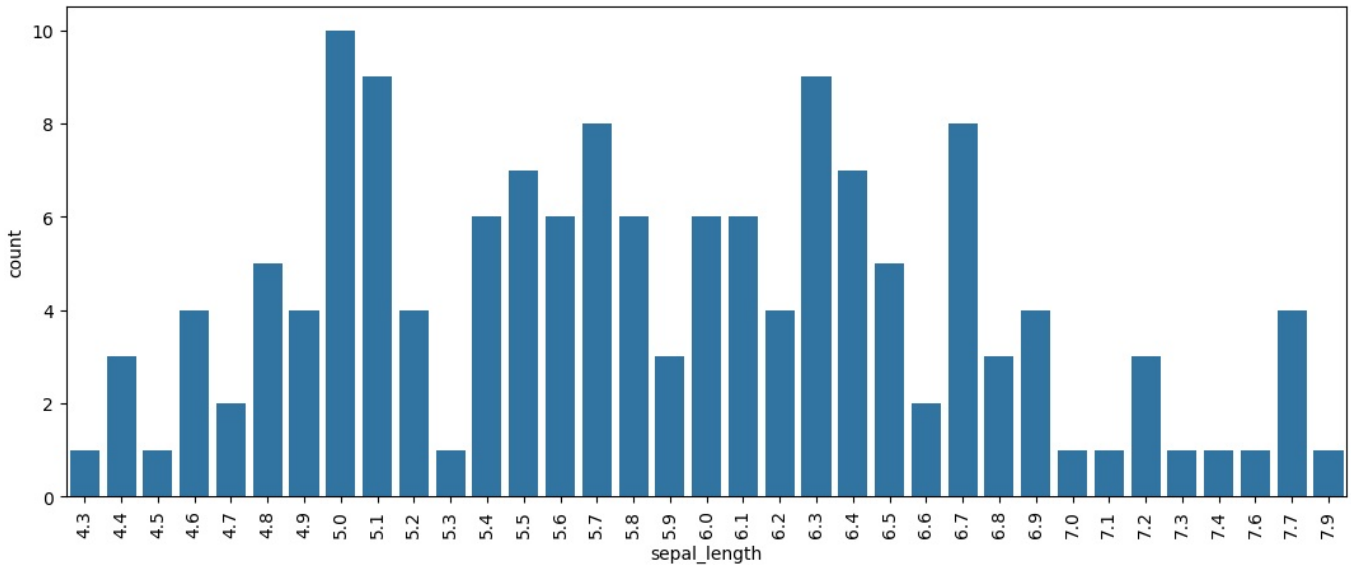
In [44]:
```
# Plotting Brand Frequency
plt.figure(figsize=(15,9))
sns.barplot(x=width_counts.index, y=width_counts.values)
plt.title('sepal_widthFrequency in Dataset')
plt.xlabel('sepal_width')
plt.ylabel('Number of species')
plt.xticks(rotation=360)
plt.show()
```



Count plot

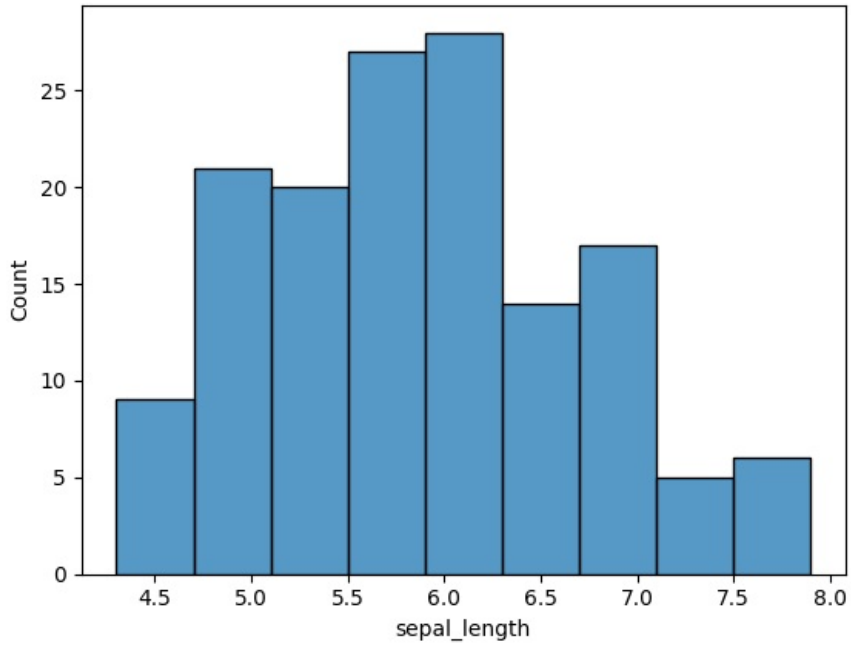
In [45]:
```
plt.figure(figsize=(13,5))

sns.countplot(x='sepal_length', data=iris_df)
plt.xticks(rotation=90)
```

```
Out[45]: ([0,
           1,
           2,
           3,
           4,
           5,
           6,
           7,
           8,
           9,
           10,
           11,
           12,
           13,
           14,
           15,
           16,
           17,
           18,
           19,
           20,
           21,
           22,
           23,
           24,
           25,
           26,
           27,
           28,
           29,
           30,
           31,
           32,
           33,
           34],
          [Text(0, 0, '4.3'),
           Text(1, 0, '4.4'),
           Text(2, 0, '4.5'),
           Text(3, 0, '4.6'),
           Text(4, 0, '4.7'),
           Text(5, 0, '4.8'),
           Text(6, 0, '4.9'),
           Text(7, 0, '5.0'),
           Text(8, 0, '5.1'),
           Text(9, 0, '5.2'),
           Text(10, 0, '5.3'),
           Text(11, 0, '5.4'),
           Text(12, 0, '5.5'),
           Text(13, 0, '5.6'),
           Text(14, 0, '5.7'),
           Text(15, 0, '5.8'),
           Text(16, 0, '5.9'),
           Text(17, 0, '6.0'),
           Text(18, 0, '6.1'),
           Text(19, 0, '6.2'),
           Text(20, 0, '6.3'),
           Text(21, 0, '6.4'),
           Text(22, 0, '6.5'),
           Text(23, 0, '6.6'),
           Text(24, 0, '6.7'),
           Text(25, 0, '6.8'),
           Text(26, 0, '6.9'),
           Text(27, 0, '7.0'),
           Text(28, 0, '7.1'),
           Text(29, 0, '7.2'),
           Text(30, 0, '7.3'),
           Text(31, 0, '7.4'),
           Text(32, 0, '7.6'),
           Text(33, 0, '7.7'),
           Text(34, 0, '7.9')])
```



In [53]:
```
sns.histplot(x='sepal_length', data = iris_df)
```

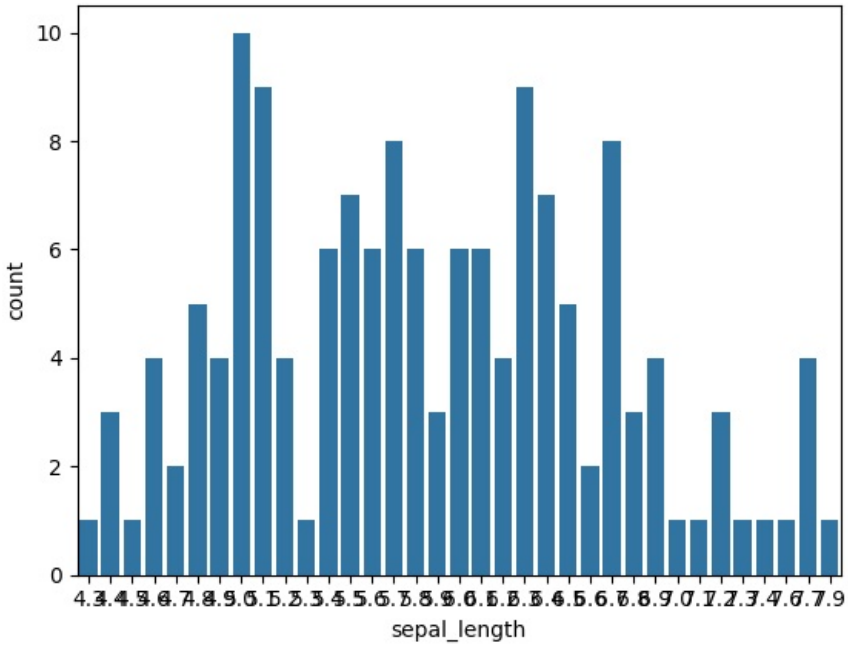
Out[53]:
```
<Axes: xlabel='sepal_length', ylabel='Count'>
```



Count plot

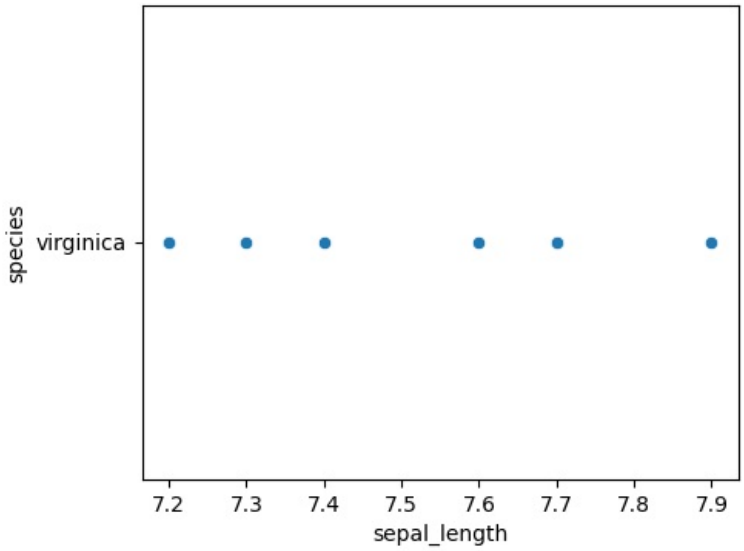
In [54]:
```
sns.countplot(x='sepal_length', data = iris_df)
```

Out[54]:
```
<Axes: xlabel='sepal_length', ylabel='count'>
```



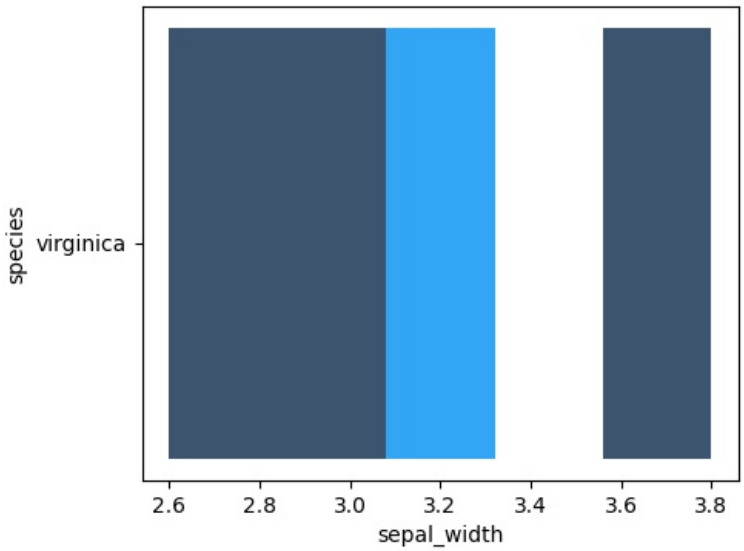
Scatter plot

In [67]:
```
plt.figure(figsize=(5,4))

sns.scatterplot(x='sepal_length',y='species',data = iris_df)
plt.show()
```



Histplot for sepal width & Species data

In [68]:
```
plt.figure(figsize=(5,4))

sns.histplot(x='sepal_width',y='species',data = iris_df)
plt.show()
```



In [69]:
```
iris_df = iris_df.nlargest(10, 'sepal_length')
iris_df
```

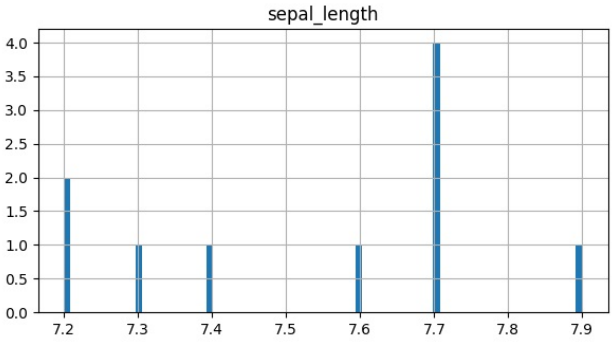
Out[69]:

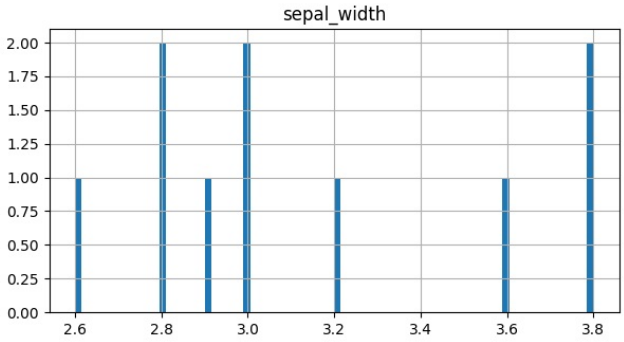
	sepal_length	sepal_width	petal_length	petal_width	species
131	7.9	3.8	6.4	2.0	virginica
117	7.7	3.8	6.7	2.2	virginica
118	7.7	2.6	6.9	2.3	virginica
122	7.7	2.8	6.7	2.0	virginica
135	7.7	3.0	6.1	2.3	virginica
105	7.6	3.0	6.6	2.1	virginica
130	7.4	2.8	6.1	1.9	virginica
107	7.3	2.9	6.3	1.8	virginica
109	7.2	3.6	6.1	2.5	virginica
125	7.2	3.2	6.0	1.8	virginica

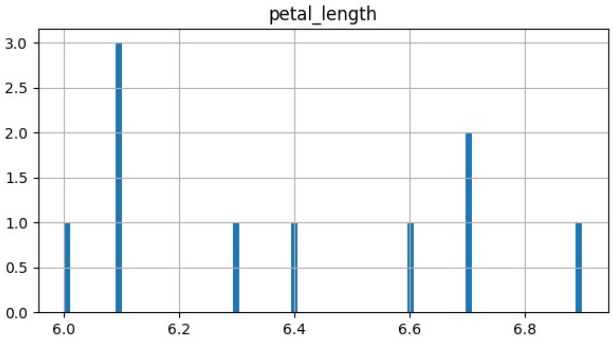
Histogram

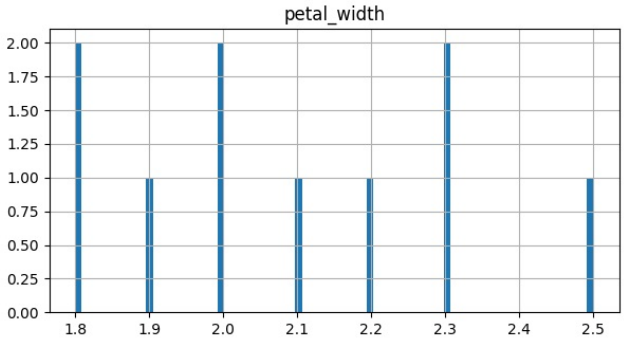
In [70]:
```
import matplotlib.pyplot as plt
import numpy as np
```

In [71]:
```
iris_df.hist(bins=80,figsize=(16,8))
plt.show()
```



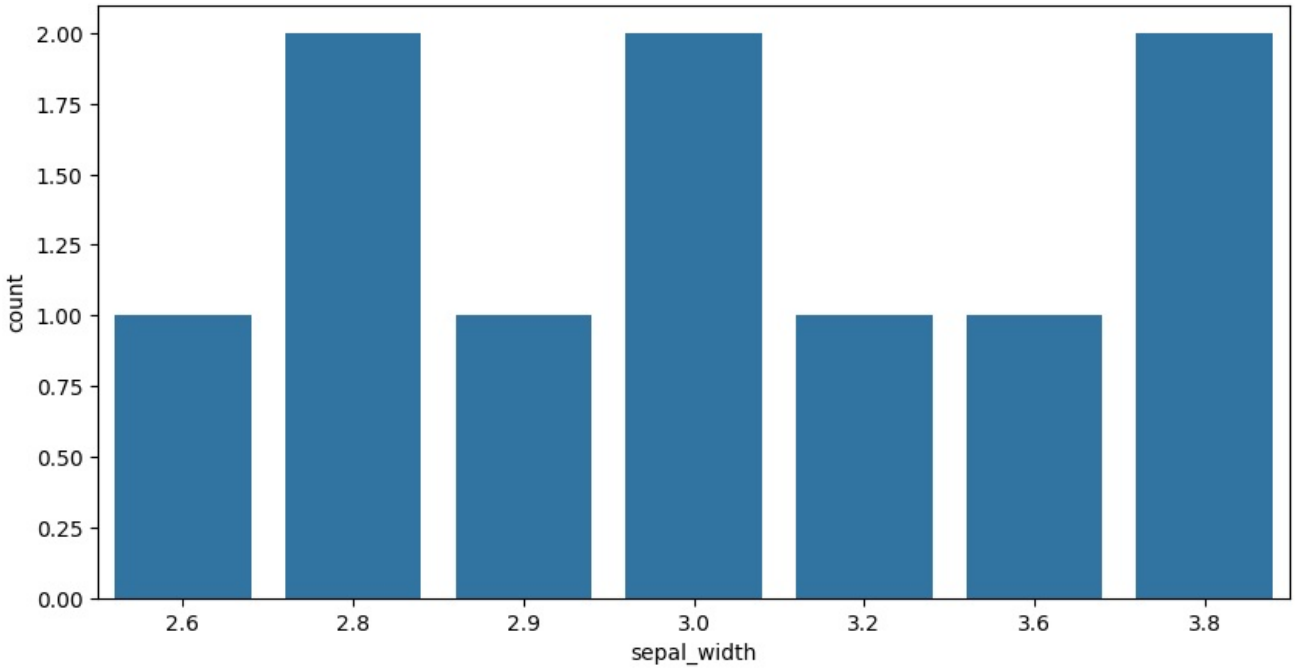






Data visualization

In [74]:
```
plt.figure(figsize=(10,5))
sns.countplot(x='sepal_width', data = iris_df)
plt.show()
```



In [75]:
```
iris_df.head()
```

Out[75]:

	sepal_length	sepal_width	petal_length	petal_width	species
131	7.9	3.8	6.4	2.0	virginica
117	7.7	3.8	6.7	2.2	virginica
118	7.7	2.6	6.9	2.3	virginica
122	7.7	2.8	6.7	2.0	virginica
135	7.7	3.0	6.1	2.3	virginica

In [76]:
```
iris_df.columns
```

Out[76]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

In [77]:
```
iris_df['sepal_length'].unique()
```

Out[77]:
```
array([7.9, 7.7, 7.6, 7.4, 7.3, 7.2])
```

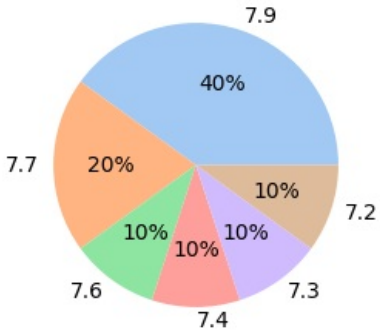
In [78]:
```
iris_df['sepal_width'].unique()
```

Out[78]:
```
array([3.8, 2.6, 2.8, 3. , 2.9, 3.6, 3.2])
```

In [79]:
```
iris_df['species'].unique()
```

Out[79]:
```
array(['virginica'], dtype=object)
```

In [80]:
```
plt.figure(figsize=(3,3))

# declaring data
data = iris_df["sepal_length"].value_counts()
keys = [7.9, 7.7, 7.6, 7.4, 7.3, 7.2]

# define Seaborn color palette to use
palette_color = sns.color_palette('pastel')

# plotting data on chart
plt.pie(data, labels=keys, colors=palette_color, autopct='%.0f%%')

# displaying chart
plt.show()
```



In [ ]:
```

```