In [4]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [6]:
```
train = pd.read_csv("train.csv")
```

In [7]:
```
train.head()
```

Out[7]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

In [9]:
```
train.isnull()
```

Out[9]:

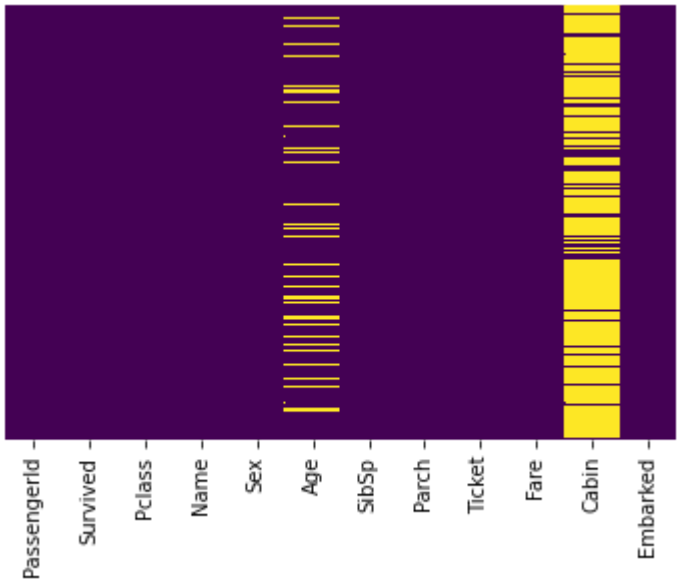
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	False	False	False	False	False	False	False	False	False	False	True	False
1	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	True	False
...	...	...	...	...	...	...	...	...	...	...	...	...

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	False	False	False	False	False	False	False	False	False	False	True	False
887	False	False	False	False	False	False	False	False	False	False	False	False
888	False	False	False	False	False	True	False	False	False	False	True	False
889	False	False	False	False	False	False	False	False	False	False	False	False
890	False	False	False	False	False	False	False	False	False	False	True	False

891 rows × 12 columns

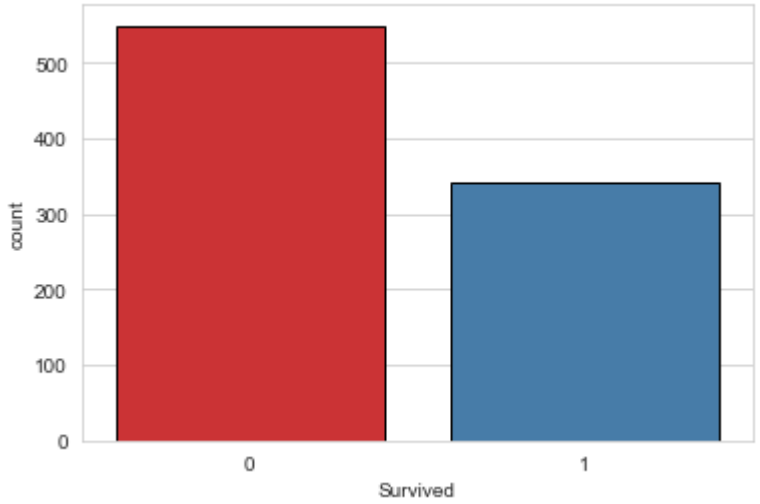
In [10]:
```
sns.heatmap(train.isnull(),yticklabels=False,cbar=False,cmap="viridis")
```

Out[10]: <AxesSubplot:>



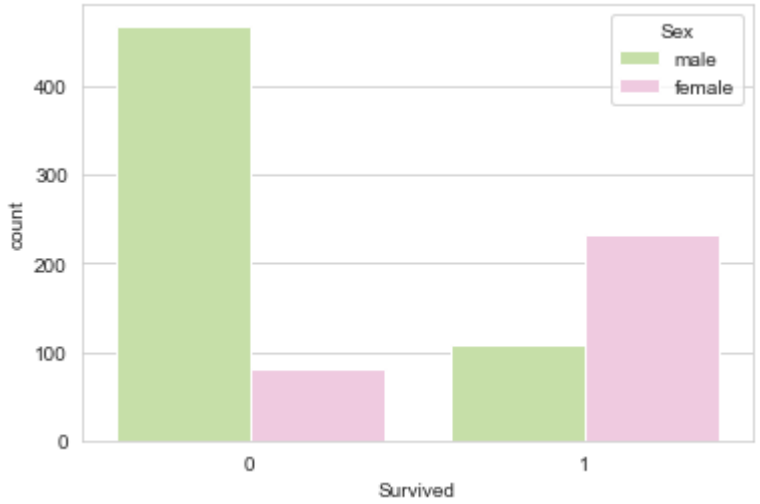
In [20]:
```
sns.set_style("whitegrid")
sns.countplot(x="Survived",data=train,palette = "Set1", edgecolor = "black")
```

Out[20]: <AxesSubplot:xlabel='Survived', ylabel='count'>



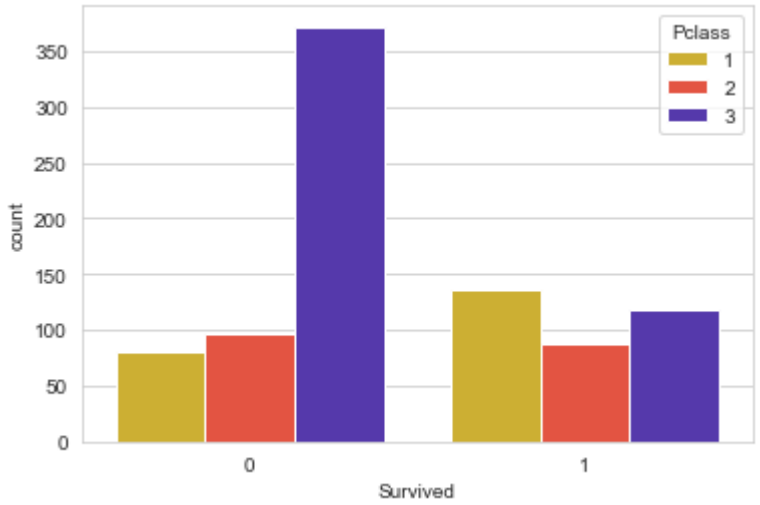
In [15]:
```
sns.set_style("whitegrid")
sns.countplot(x="Survived",hue="Sex",data=train,palette="PiYG_r")
```

Out[15]: <AxesSubplot:xlabel='Survived', ylabel='count'>



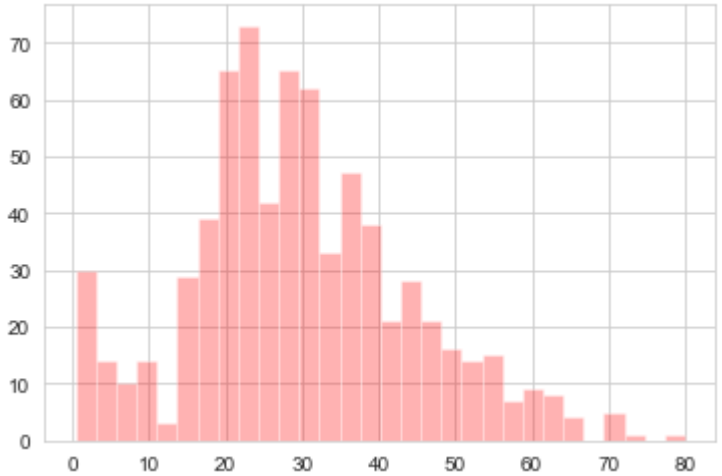
In [23]:
```
sns.set_style('whitegrid')
sns.countplot(x='Survived',hue='Pclass',data=train,palette='CMRmap_r')
```

Out[23]: <AxesSubplot:xlabel='Survived', ylabel='count'>



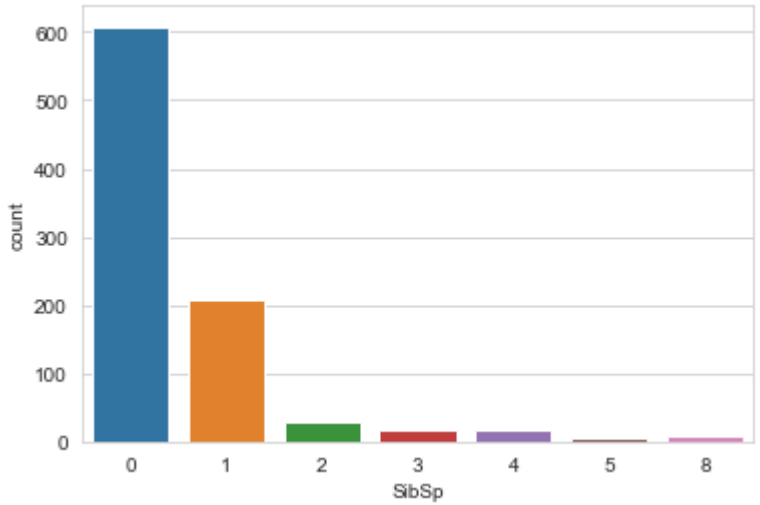
In [25]:
```
train['Age'].hist(bins=30,color="Red",alpha=0.3)
```

Out[25]: <AxesSubplot:>



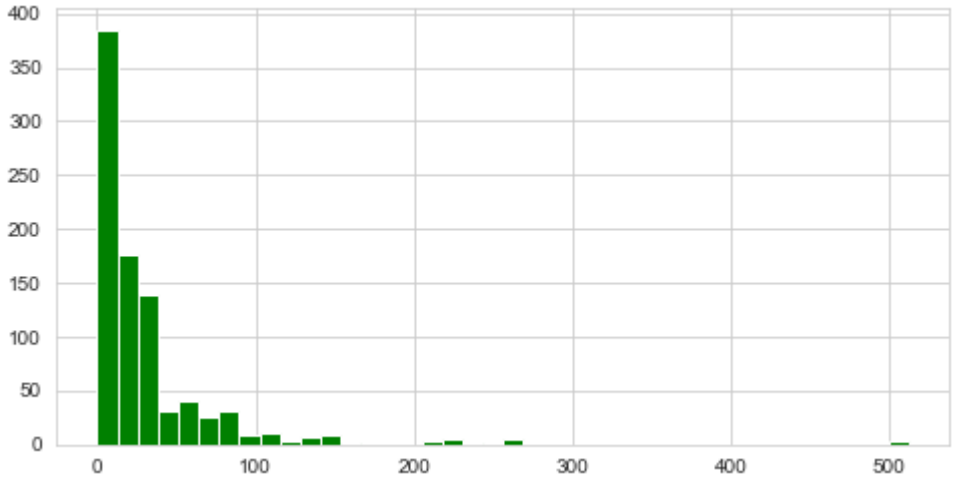
In [26]:
```
sns.countplot(x='SibSp',data=train)
```

Out[26]: <AxesSubplot:xlabel='SibSp', ylabel='count'>



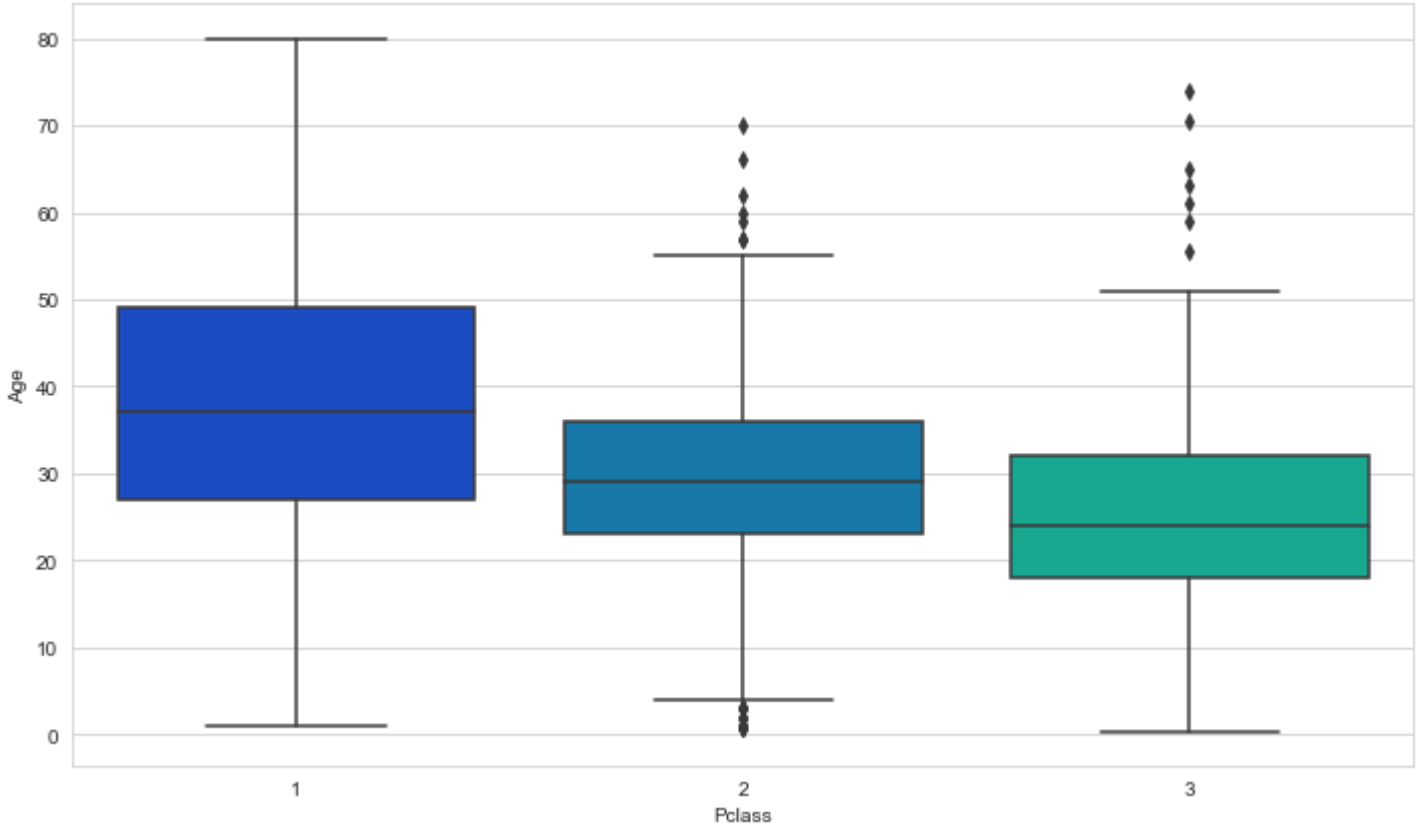
In [27]:
```
train['Fare'].hist(color="green",bins=40,figsize=(8,4))
```

Out[27]: <AxesSubplot:>



In [30]:
```
plt.figure(figsize=(12, 7))
sns.boxplot(x='Pclass',y="Age",data=train,palette='winter')
```

Out[30]: <AxesSubplot:xlabel='Pclass', ylabel='Age'>



In [31]:

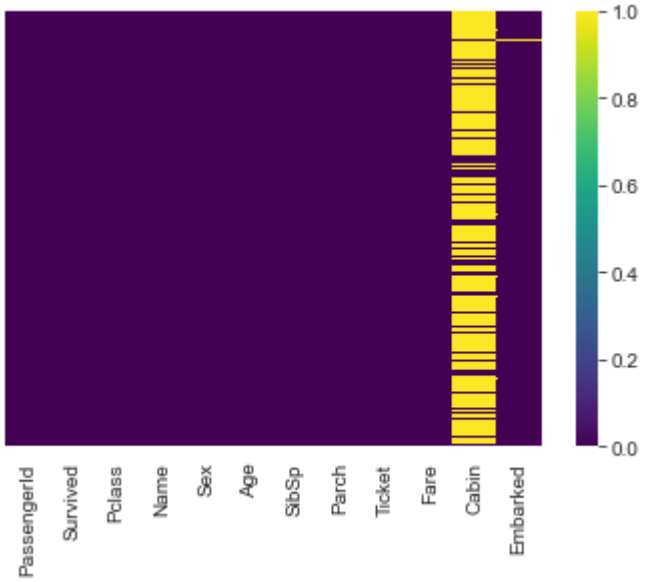
```
def impute_age(cols):
    Age = cols[0]
    Pclass = cols[1]

    if pd.isnull(Age):

        if Pclass == 1:
            return 37

        elif Pclass == 2:
            return 29

        else:
            return 24
```

```
    else:
        return Age
```

In [32]:
```
train['Age']= train[["Age","Pclass"]].apply(impute_age,axis=1)
```

In [33]:
```
sns.heatmap(train.isnull(),yticklabels=False,cmap='viridis')
```

Out[33]: <AxesSubplot:>



In [34]:
```
train.drop('Cabin',axis=1,inplace=True)
```

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

Out[35]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	S

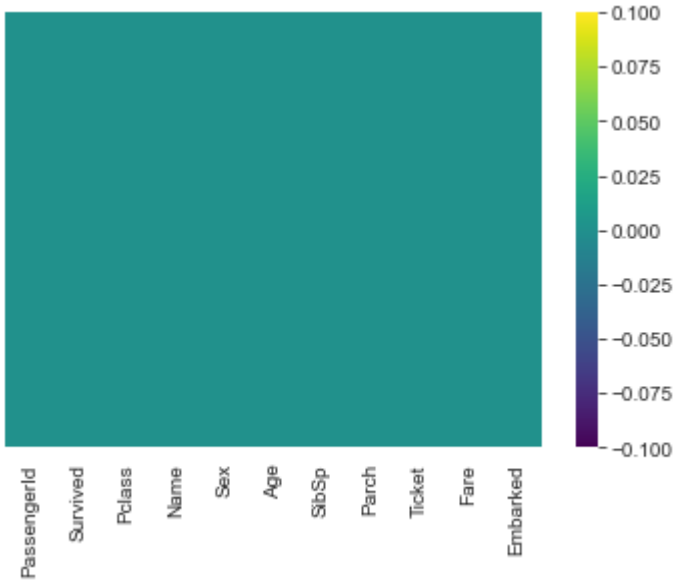
In [36]:
```
train.dropna(inplace=True)
```

In [37]:
```
train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 889 entries, 0 to 890
Data columns (total 11 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  889 non-null    int64  
 1   Survived     889 non-null    int64  
 2   Pclass       889 non-null    int64  
 3   Name         889 non-null    object 
 4   Sex          889 non-null    object 
 5   Age          889 non-null    float64
 6   SibSp        889 non-null    int64  
 7   Parch        889 non-null    int64  
 8   Ticket       889 non-null    object 
 9   Fare         889 non-null    float64
 10  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(4)
memory usage: 83.3+ KB
```

In [38]:
```
sns.heatmap(train.isnull(),yticklabels=False,cmap='viridis')
```

Out[38]: <AxesSubplot:>



In [39]:
```
sex = pd.get_dummies(train['Sex'],drop_first=True)
embark = pd.get_dummies(train['Embarked'],drop_first=True)
```

In [40]:
```
train.drop(['Sex',"Embarked",'Name','Ticket'],axis=1,inplace=True)
```

In [41]:
```
train.head()
```

Out[41]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
0	1	0	3	22.0	1	0	7.2500
1	2	1	1	38.0	1	0	71.2833
2	3	1	3	26.0	0	0	7.9250
3	4	1	1	35.0	1	0	53.1000
4	5	0	3	35.0	0	0	8.0500

In [42]:
```
train=pd.concat([train,sex,embark],axis=1)
```

In [43]:
```
train.head()
```

Out[43]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare	male	Q	S
0	1	0	3	22.0	1	0	7.2500	1	0	1
1	2	1	1	38.0	1	0	71.2833	0	0	0
2	3	1	3	26.0	0	0	7.9250	0	0	1
3	4	1	1	35.0	1	0	53.1000	0	0	1
4	5	0	3	35.0	0	0	8.0500	1	0	1

In [ ]:
```

```