In [1]:
```
import pandas as pd
from pandas import Series, DataFrame
```

In [2]:
```
titanic_df = pd.read_csv('train.csv')
```

In [3]:
```
titanic_df.head()
```

Out[3]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [4]:
```
titanic_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

First some basic questions:

1.) Who were the passengers on the Titanic? (Ages,Gender,Class,..etc)

2.) What deck were the passengers on and how does that relate to their class?

3.) Where did the passengers come from?

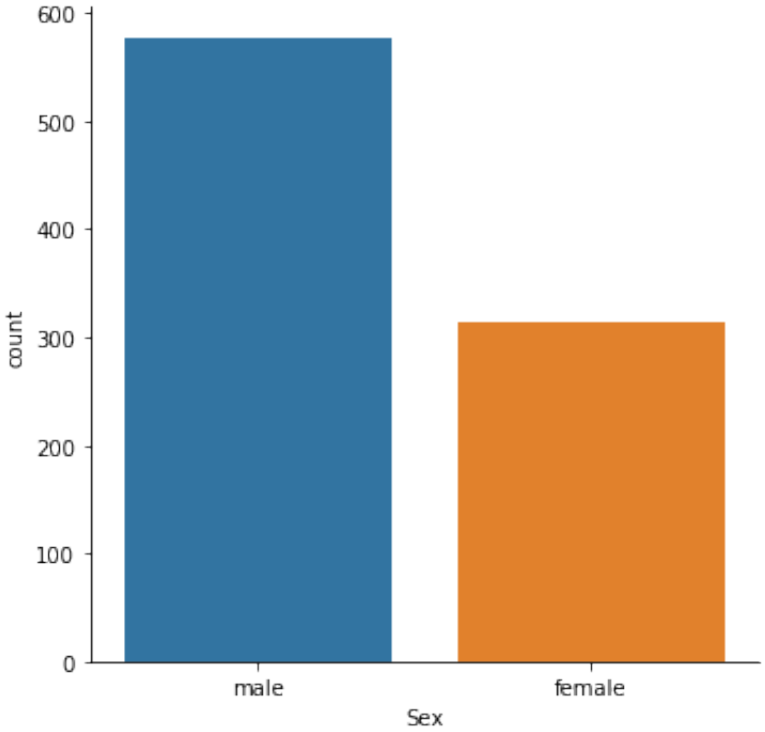
4.) Who was alone and who was with family?

5.) What factors helped someone survive the sinking?

In [5]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

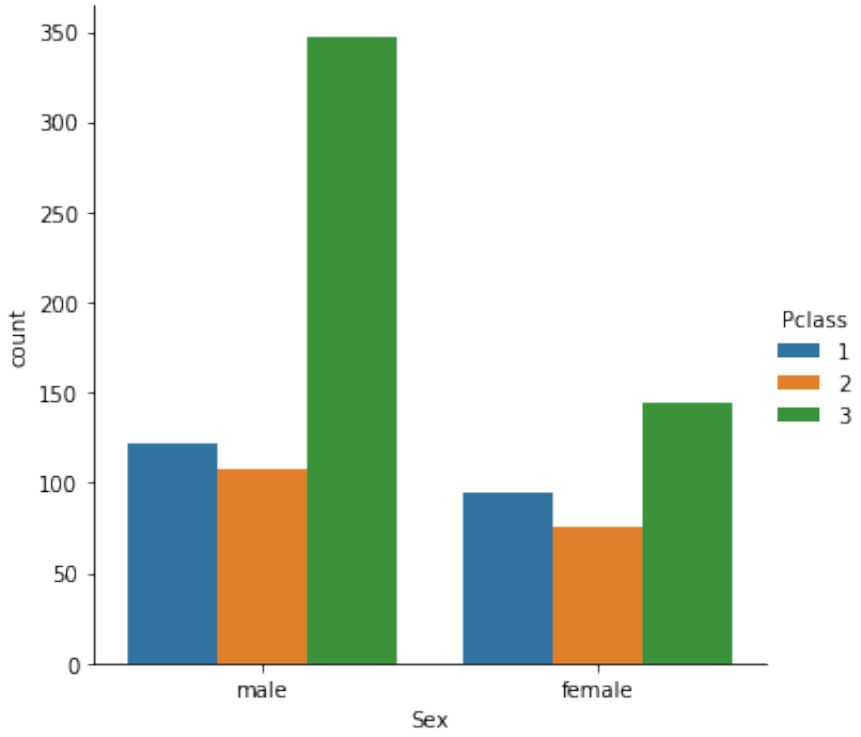
In [6]:
```
#1) who were the passengers on the titanic

#Spilt of male, female passengers on titanic across all classes

sns.catplot('Sex',data=titanic_df, kind='count')
```

Out[6]: <seaborn.axisgrid.FacetGrid at 0x1a20c71e90>



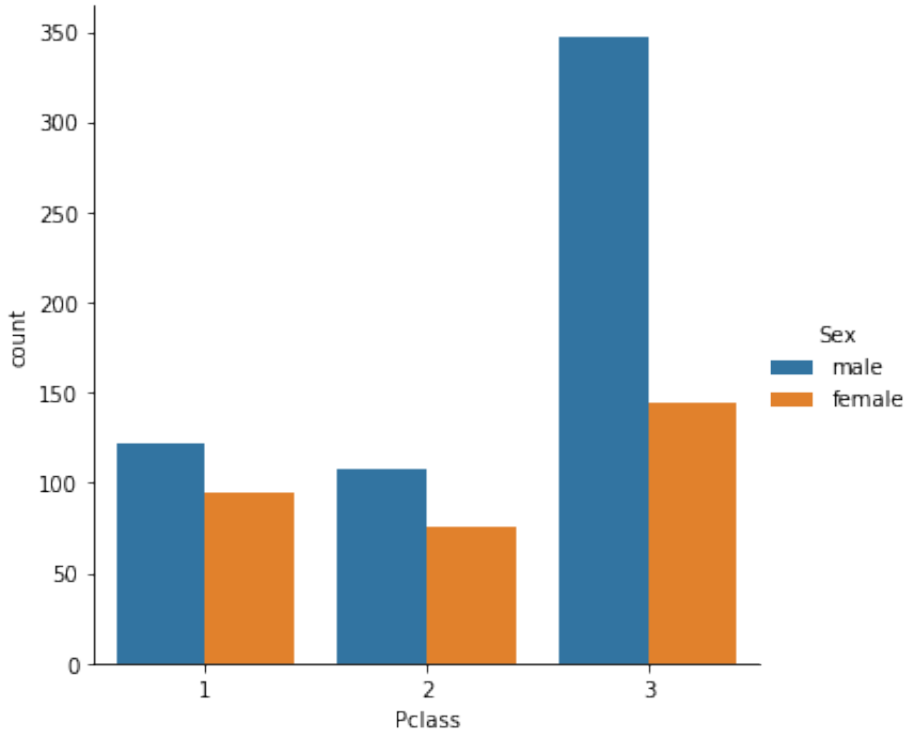
In [7]:
```
#Spilt of male, female (inc. children) in each class

sns.catplot('Sex', data= titanic_df, hue= 'Pclass', kind= 'count')
```

Out[7]: <seaborn.axisgrid.FacetGrid at 0x1a21555990>



In [8]:
```
#Spilt of male, female (inc. children) in each class

sns.catplot('Pclass', data= titanic_df, hue= 'Sex', kind= 'count')
```

Out[8]: <seaborn.axisgrid.FacetGrid at 0x1a20c7ff50>



In [9]:
```
#Function to determine if each passenger is a male, female of child

def male_female_child(passenger):
    age,sex = passenger

    if age < 16:
        return 'child'
    else:
        return sex
```

In [10]:
```
#adding a person column to data set using function male_female_chil

titanic_df['Person'] = titanic_df[['Age', 'Sex']].apply(male_female_
```

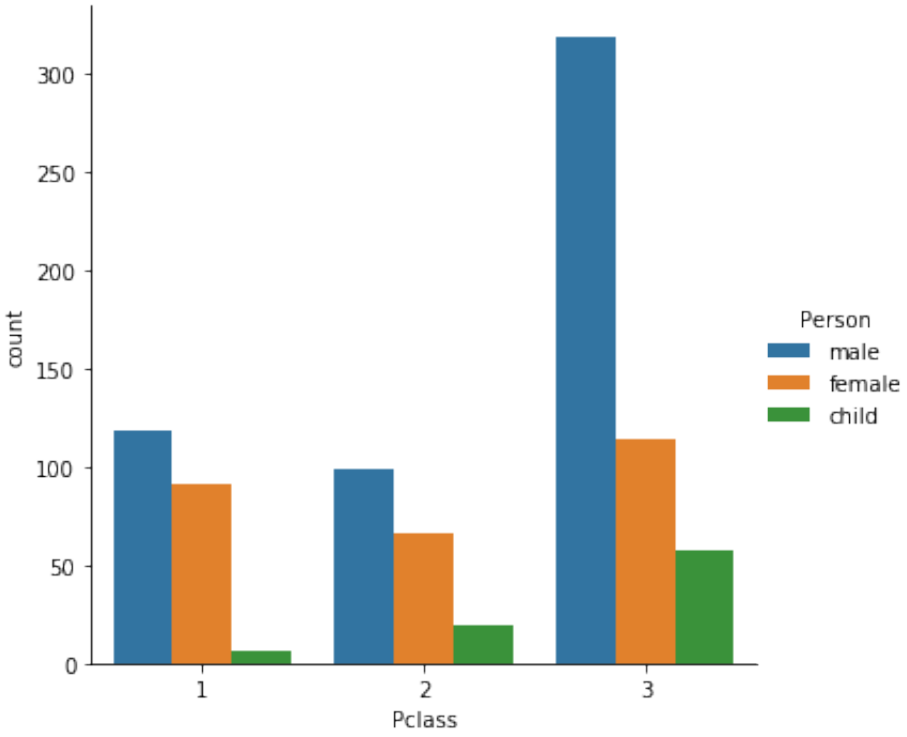
In [11]:
```
titanic_df.head(10)
```

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.458
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.862
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.075
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.133
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.070

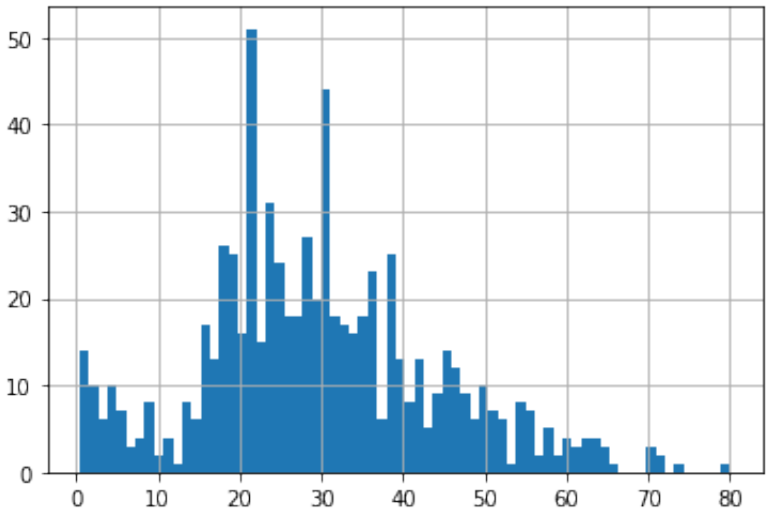
In [12]:
```
#Spilt of male, female and children in each class on the titanic

sns.catplot('Pclass', data= titanic_df, hue= 'Person', kind = 'coun
```

Out[12]: <seaborn.axisgrid.FacetGrid at 0x1a21555210>



In [13]:
```
#Histagram showing spread of ages on the titanic

titanic_df['Age'].hist(bins = 70)
```

Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x1a21ae0fd0>



In [15]:
```
#Average age of all passengers on the titanic

round(titanic_df['Age'].mean(),0)
```

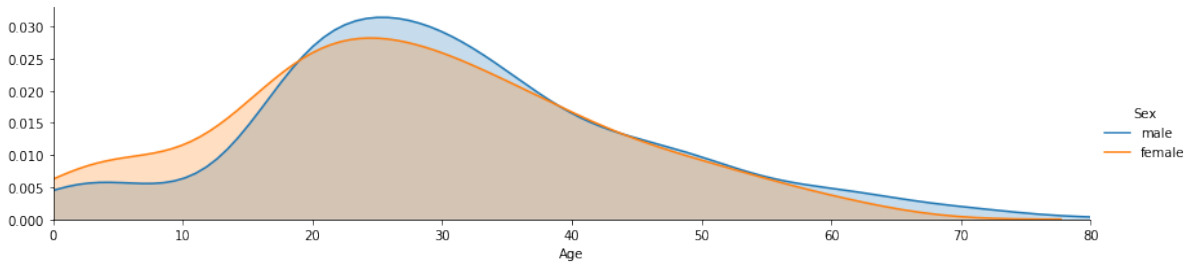
Out[15]:
```
30.0
```

In [16]:
```
#Passengers split between male, female and children

titanic_df['Person'].value_counts()
```

Out[16]:
```
male      537
female    271
child      83
Name: Person, dtype: int64
```

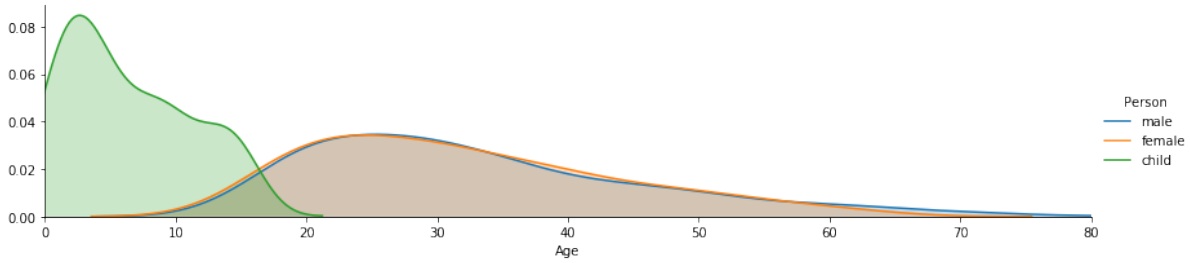
In [17]:
```
#kde plot showing the ages of male, female (inc. children) passenge

fig = sns.FacetGrid(titanic_df, hue = 'Sex', aspect = 4)
fig.map(sns.kdeplot, 'Age', shade = True)

oldest = titanic_df['Age'].max()

fig.set(xlim = (0, oldest))

fig.add_legend()
```

Out[17]:
```
<seaborn.axisgrid.FacetGrid at 0x1a21c1bfd0>
```



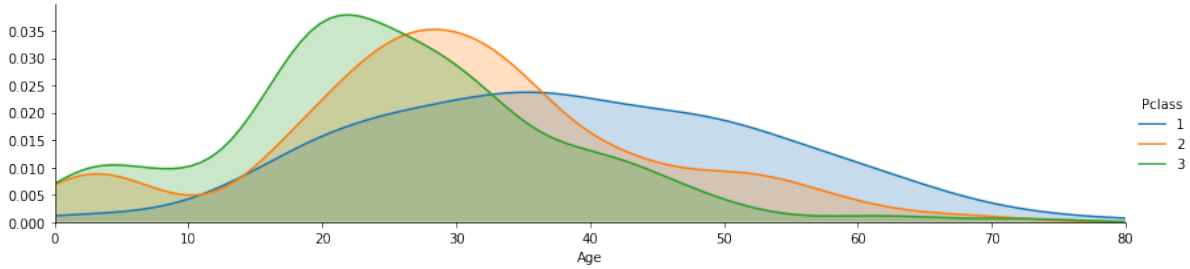
In [30]:
```
#kde plot showing the ages of male, female and children passengers (

fig = sns.FacetGrid(titanic_df, hue = 'Person', aspect = 4)
fig.map(sns.kdeplot, 'Age', shade = True)

oldest = titanic_df['Age'].max()

fig.set(xlim = (0, oldest))

fig.add_legend()
```

Out[30]: <seaborn.axisgrid.FacetGrid at 0x1a206ebd90>



In [31]:
```
#kde plot showing the ages of male, female by class on the titanic.

fig = sns.FacetGrid(titanic_df, hue = 'Pclass', aspect = 4)
fig.map(sns.kdeplot, 'Age', shade = True)

oldest = titanic_df['Age'].max()

fig.set(xlim = (0, oldest))

fig.add_legend()
```

Out[31]: <seaborn.axisgrid.FacetGrid at 0x1a2081ef10>



In [34]:
```
#2 What deck were the passengers on and how does that relate to the

titanic_df.head()
```

Out[34]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

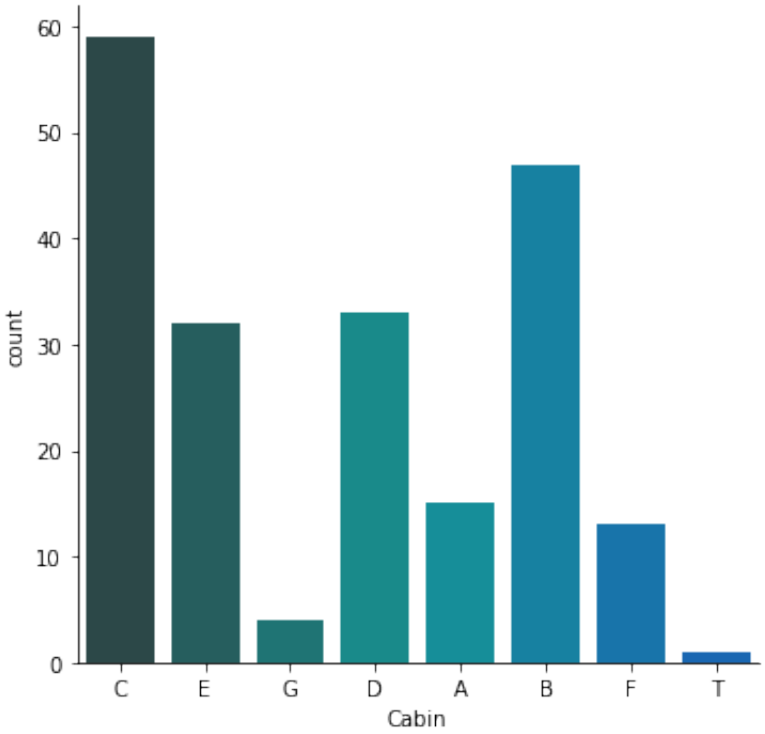
In [35]:
```
deck = titanic_df['Cabin'].dropna()
```

In [36]:
```
deck.head()
```

Out[36]:
```
1      C85
3     C123
6      E46
10      G6
11    C103
Name: Cabin, dtype: object
```

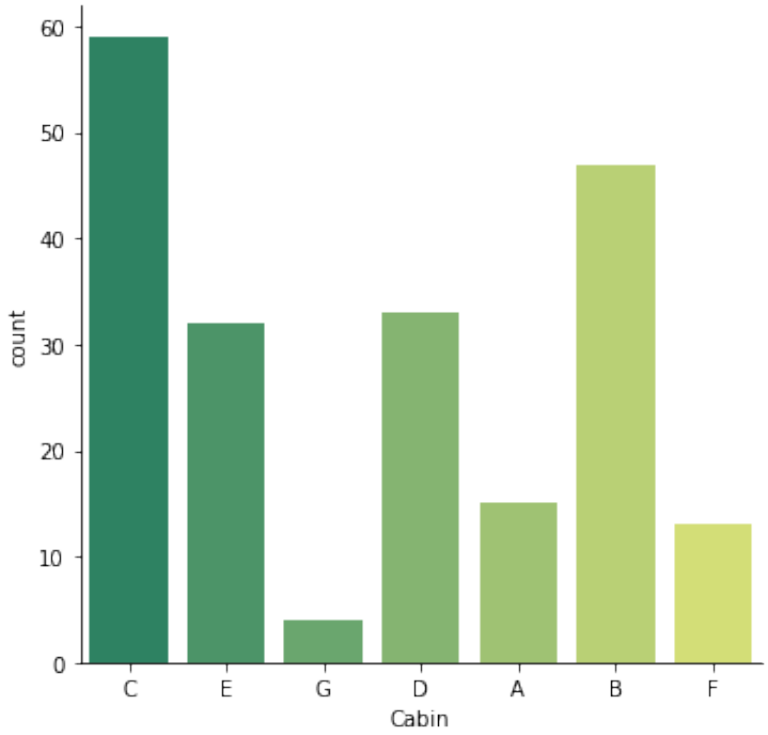
In [66]:
```
levels = []

for level in deck:
    levels.append(level[0])

cabin_df = DataFrame(levels)
cabin_df.columns = ['Cabin']
sns.catplot('Cabin', data = cabin_df, palette = 'winter_d', kind='c
```

Out[66]: <seaborn.axisgrid.FacetGrid at 0x1a21886a10>



In [65]:
```
cabin_df = cabin_df[cabin_df.Cabin != 'T']
sns.catplot('Cabin', data = cabin_df, palette = 'summer', kind= 'co
```

Out[65]: <seaborn.axisgrid.FacetGrid at 0x1a216d6250>



In [48]:
```
#3) Where did the passengers come from?
```

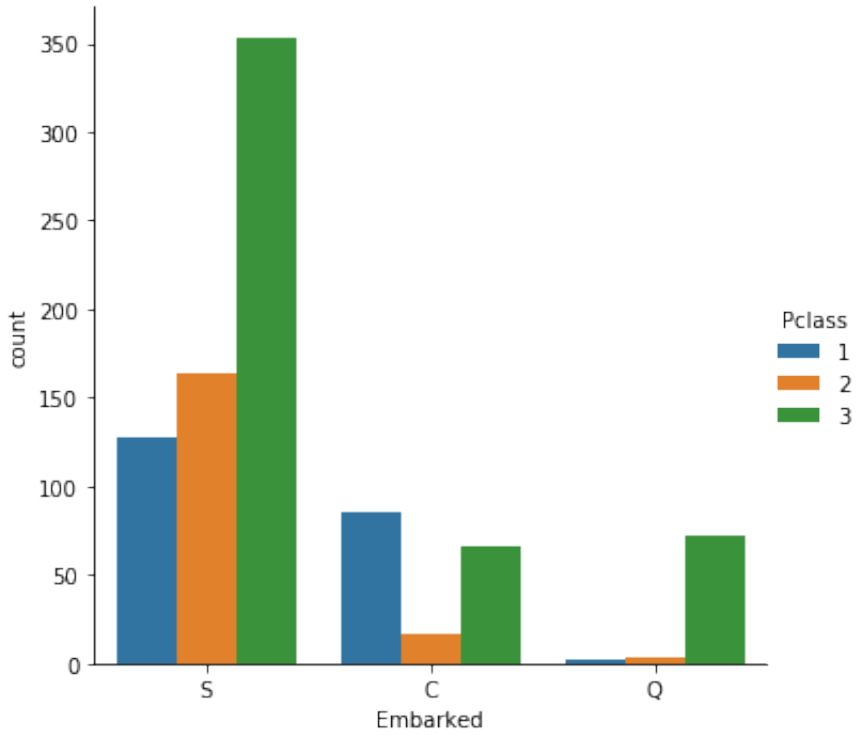
In [52]:
```
titanic_df.head()
```

Out[52]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [64]:
```
sns.catplot(x='Embarked',hue='Pclass', data = titanic_df, kind='cou
```

Out[64]:
```
<seaborn.axisgrid.FacetGrid at 0x1a216d62d0>
```



In [67]:
```
#4.) Who was alone and who was with family?
```

In [69]:
```
titanic_df.head()
```

Out[69]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [70]:
```
titanic_df['Alone'] = titanic_df.SibSp + titanic_df.Parch
```

In [71]:
```
titanic_df['Alone']
```

Out[71]:
```
0      1
1      1
2      0
3      1
4      0
      ..
886    0
887    0
888    3
889    0
890    0
Name: Alone, Length: 891, dtype: int64
```

In [72]:
```
titanic_df['Alone'].loc[titanic_df['Alone'] > 0] = 'With Family'
titanic_df['Alone'].loc[titanic_df['Alone'] == 0] = 'Alone'
```

```
/Users/Martin_Hopkins/opt/anaconda3/lib/python3.7/site-packages/pa
ndas/core/indexing.py:670: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pa
ndas-docs/stable/user_guide/indexing.html#returning-a-view-versus-
a-copy (https://pandas.pydata.org/pandas-docs/stable/user_guide/in
dexing.html#returning-a-view-versus-a-copy)
  self._setitem_with_indexer(indexer, value)
```

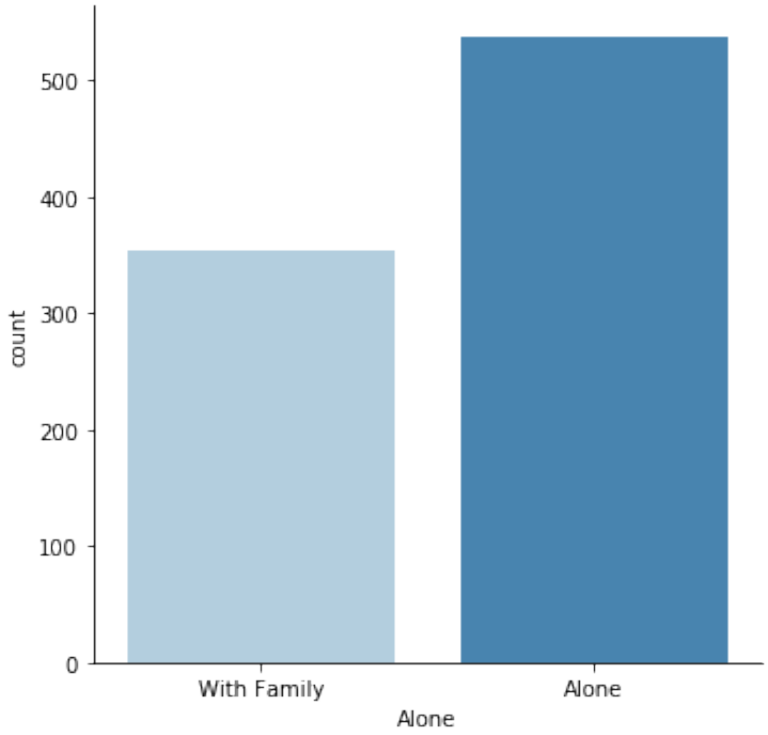
In [74]:
```
titanic_df.head()
```

Out[74]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [76]:
```
sns.catplot('Alone', data = titanic_df, palette = 'Blues', kind = '
```

Out[76]: <seaborn.axisgrid.FacetGrid at 0x1a21d08250>



In [77]:
```
#5) What factors helped someone survive the sinking?
```

In [78]:
```
titanic_df['Survivor'] = titanic_df.Survived.map({0: 'no', 1: 'yes'
```

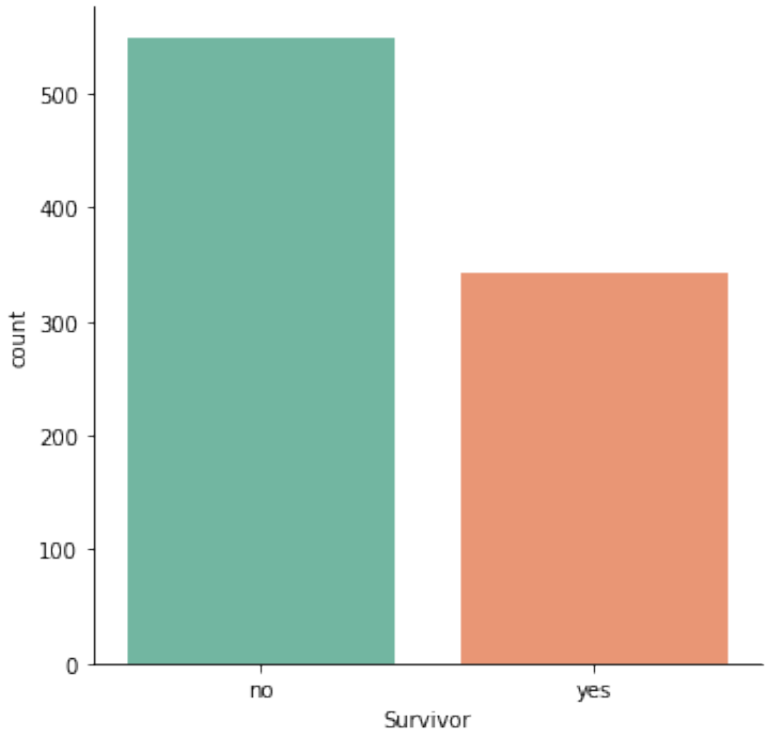
In [79]:
```
titanic_df.head()
```

Out[79]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

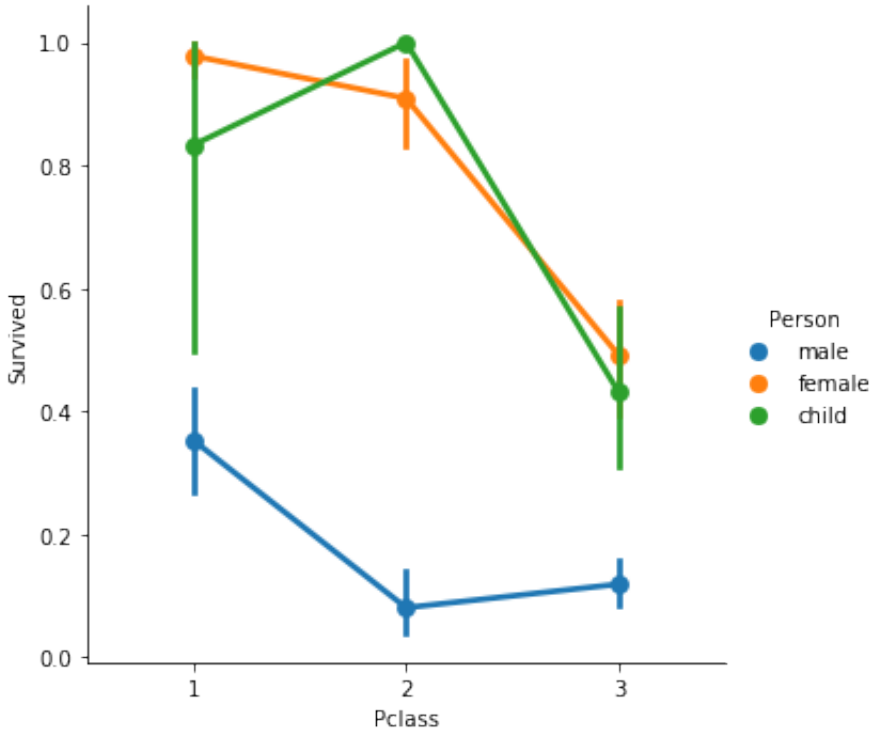
In [81]:
```
sns.catplot('Survivor', data = titanic_df, palette = 'Set2', kind =
```

Out[81]:
```
<seaborn.axisgrid.FacetGrid at 0x1a21f8bc10>
```



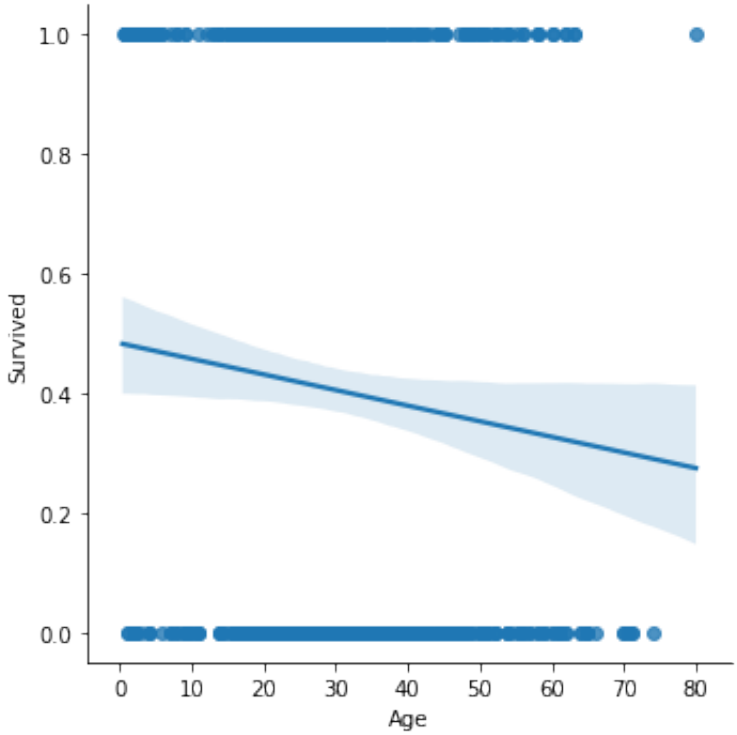
In [85]:
```
sns.catplot('Pclass', 'Survived', hue= 'Person', data= titanic_df,
```

Out[85]: <seaborn.axisgrid.FacetGrid at 0x1a224c27d0>



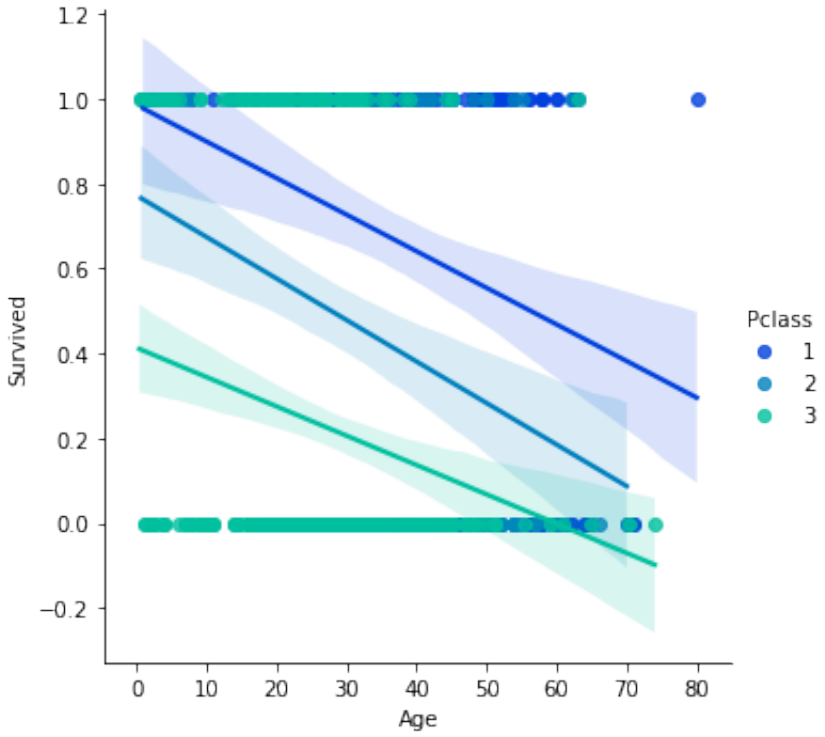
In [86]:
```
sns.lmplot('Age', 'Survived', data = titanic_df)
```

Out[86]: <seaborn.axisgrid.FacetGrid at 0x1a225d5450>



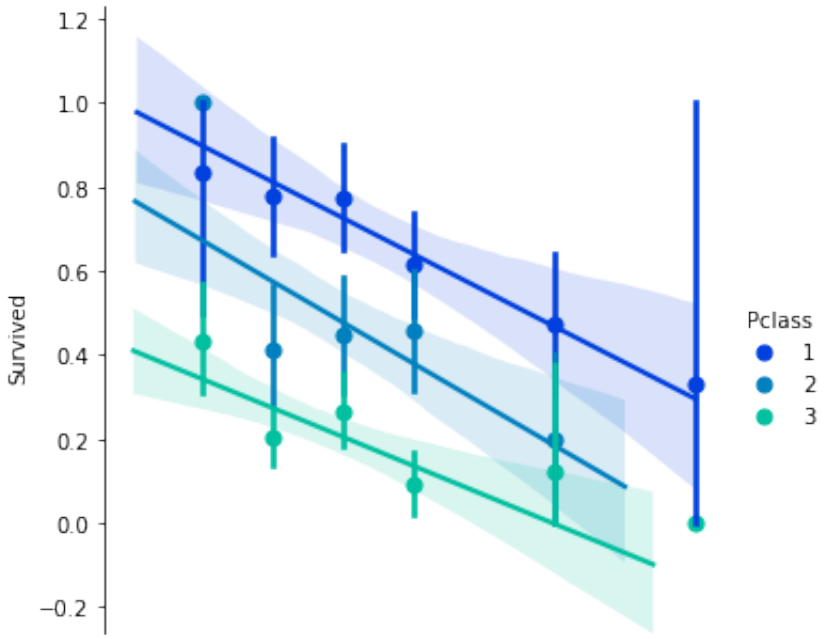
In [88]:
```
sns.lmplot('Age', 'Survived', hue= 'Pclass', data = titanic_df, pal
```

Out[88]: <seaborn.axisgrid.FacetGrid at 0x1a22438c90>



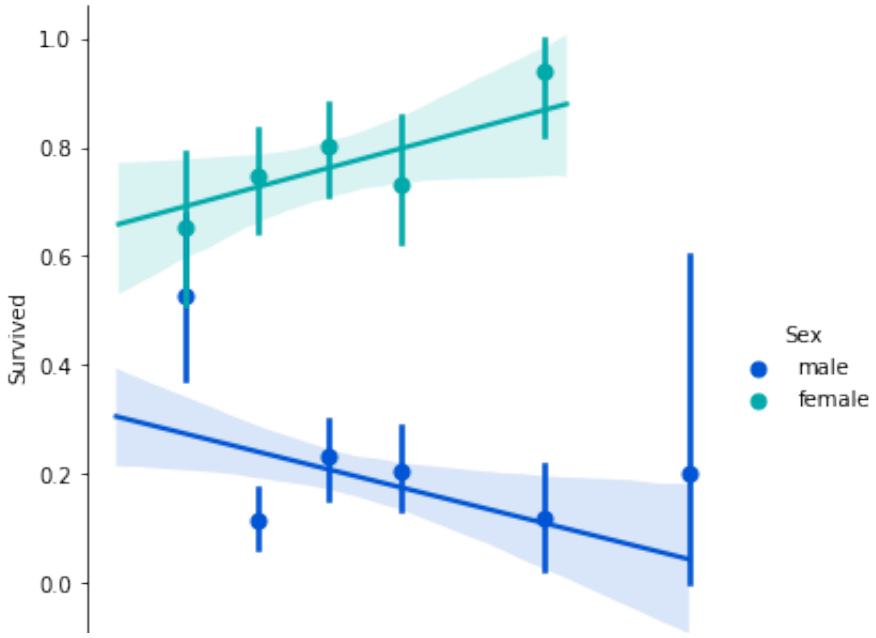
In [90]:
```
generations = [10, 20, 30, 40, 60, 80]

sns.lmplot("Age", 'Survived', hue='Pclass', data= titanic_df, palet
```

Out[90]: <seaborn.axisgrid.FacetGrid at 0x1a22a0b4d0>



In [91]:
```
sns.lmplot("Age", 'Survived', hue='Sex', data = titanic_df, palette

#6) Did the deck have an affect on passenger survival rate?
```

Out[91]: <seaborn.axisgrid.FacetGrid at 0x1a2291ac10>



In [93]:
```
titanic_df.head()
```

Out[93]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.250
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.283
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.100
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.050

In [ ]:
```
```

In [ ]: