dataset link= https://www.kaggle.com/cities/titanic123 (https://www.kaggle.com/cities/titanic123)

In [13]:

```
import import_ipynb
import pandas as pd
from pandas import Series, DataFrame
```

In [14]:

```
titanic_df=pd.read_csv('Titanic.csv')
```

In [15]:

```
titanic_df.head()
```

Out[15]:

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

In [16]:

```
titanic_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId    891 non-null int64
Survived       891 non-null int64
Pclass         891 non-null int64
Name           891 non-null object
Sex            891 non-null object
Age            714 non-null float64
SibSp          891 non-null int64
Parch          891 non-null int64
Ticket         891 non-null object
Fare           891 non-null float64
Cabin          204 non-null object
Embarked       889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```

Quetions to answer using data analysis skills:

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 classes? 3.Where did the passengers come from? 4.Who was alone and who was with family? 5.What factors helped someone survived sinking?

In [24]:

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [36]:

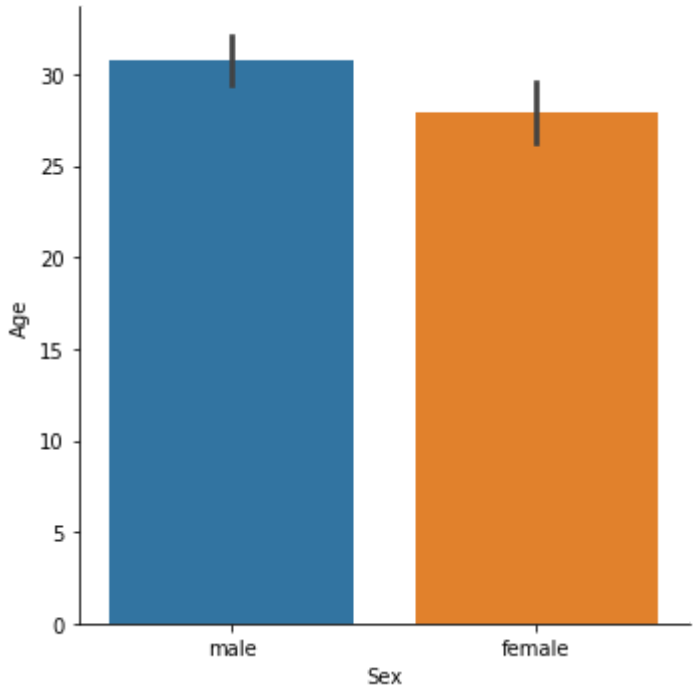
```
count=len(titanic_df.PassengerId)
```

In [64]:

```
sns.factorplot(x="Sex",y="Age",data=titanic_df,kind='bar',height=5)
```

Out[64]:

```
<seaborn.axisgrid.FacetGrid at 0x20194f37eb8>
```

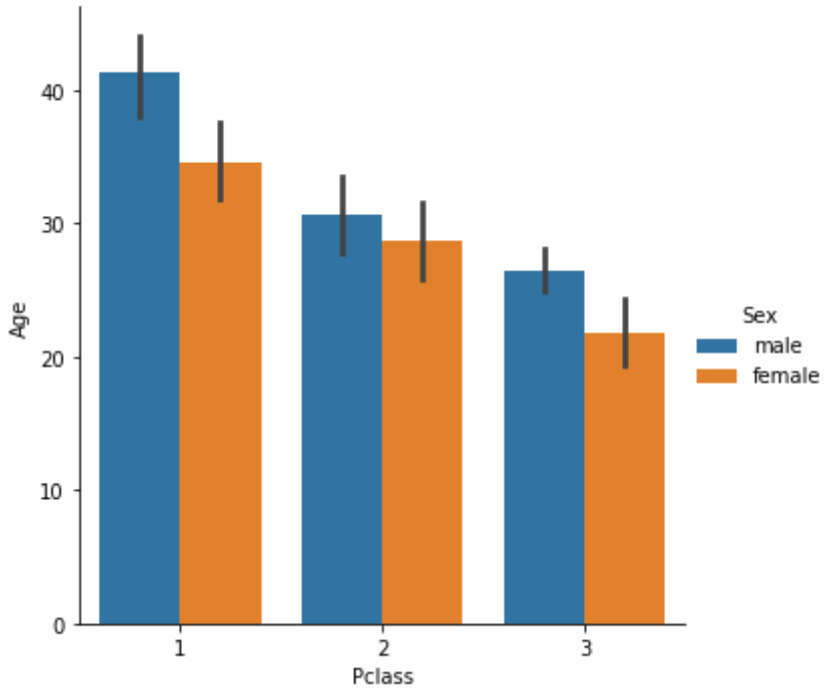


In [65]:

```
#1.Who were the passengers on the titanic?
sns.factorplot(x="Pclass",y="Age",data=titanic_df,hue='Sex',kind='bar')
```

Out[65]:

```
<seaborn.axisgrid.FacetGrid at 0x20194f37630>
```



In [53]:

```
#Function to mark passenger below age of 16 as child
def male_female_child(passenger):
    age,sex=passenger

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

In [55]:

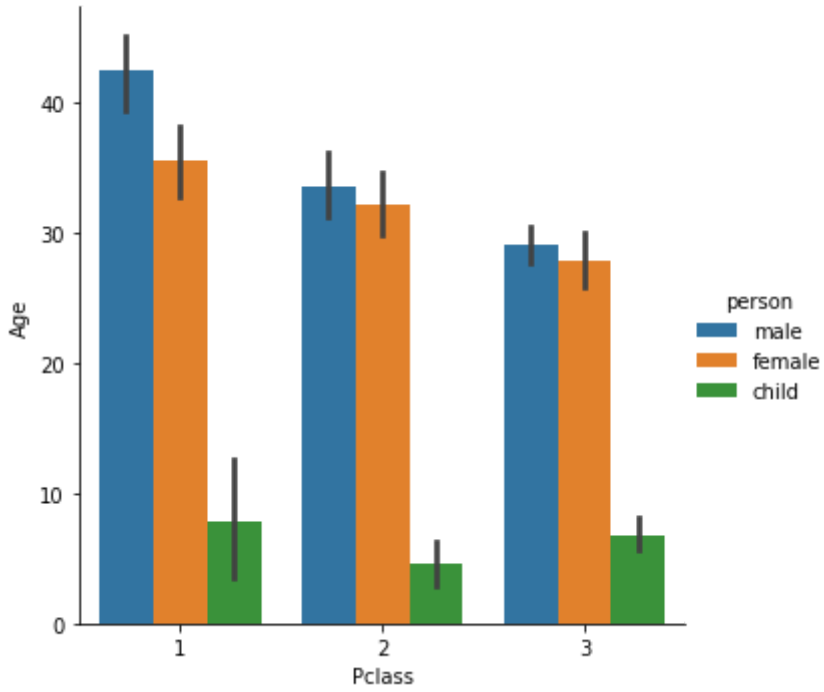
```
#Adding a column named person which categorizes passengers as male, female and child
titanic_df['person']=titanic_df[['Age','Sex']].apply(male_female_child,axis=1)
```

In [72]:

```
sns.factorplot('Pclass','Age',data=titanic_df,hue='person',kind='bar')
```

Out[72]:

```
<seaborn.axisgrid.FacetGrid at 0x2019611fdd8>
```

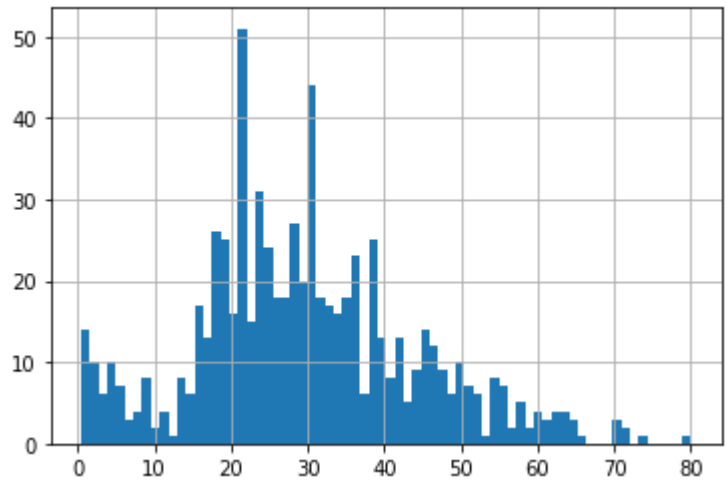


In [73]:

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

Out[73]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2019613f898>
```



In [74]:

```
#Calculating mean of the ages
titanic_df['Age'].mean()
```

Out[74]:

```
29.69911764705882
```

In [75]:

```
#Printing number of males,females and child passengers
titanic_df['person'].value_counts()
```

Out[75]:

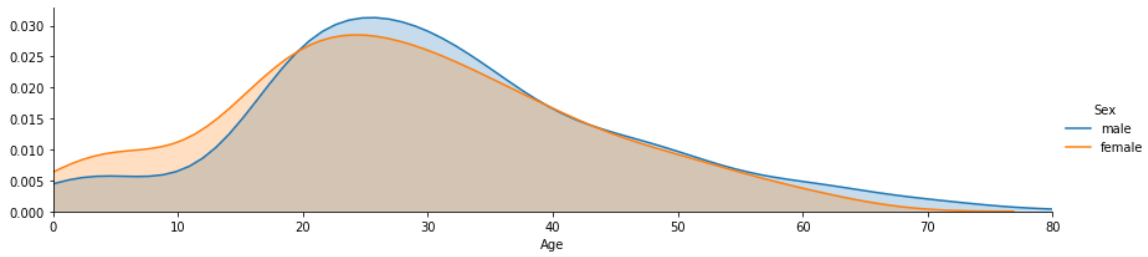
```
male      537
female    271
child      83
Name: person, dtype: int64
```

In [82]:

```
#Distribution of age by sex
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[82]:

```
<seaborn.axisgrid.FacetGrid at 0x201962f84e0>
```

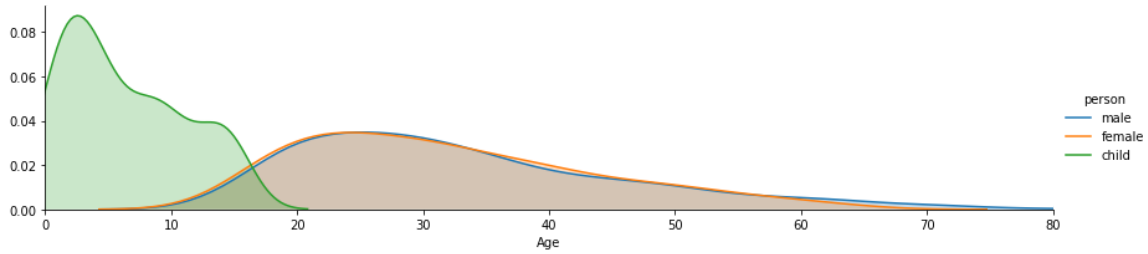


In [83]:

```
#Distribution of age by person
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[83]:

```
<seaborn.axisgrid.FacetGrid at 0x20196381908>
```

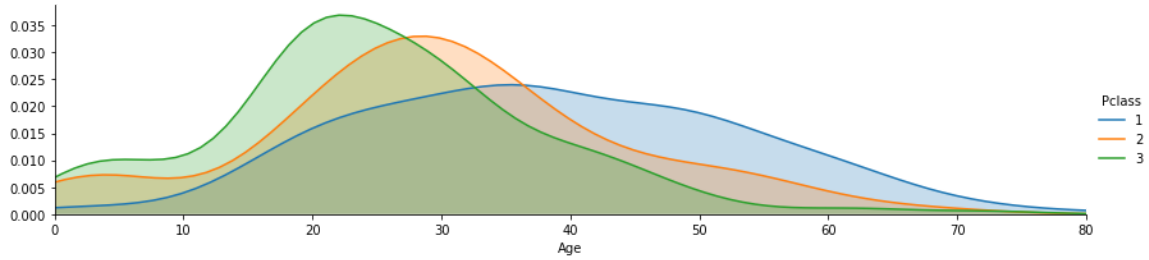


In [84]:

```
#Distribution of age by Pclass
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[84]:

```
<seaborn.axisgrid.FacetGrid at 0x20196434160>
```



In [85]:

```
#2.What deck were the passengers on?
#Dropping null values from the column Cabin
deck=titanic_df['Cabin'].dropna()
```

In [86]:

```
#Printing first values in the column Cabin
deck.head()
```

Out[86]:

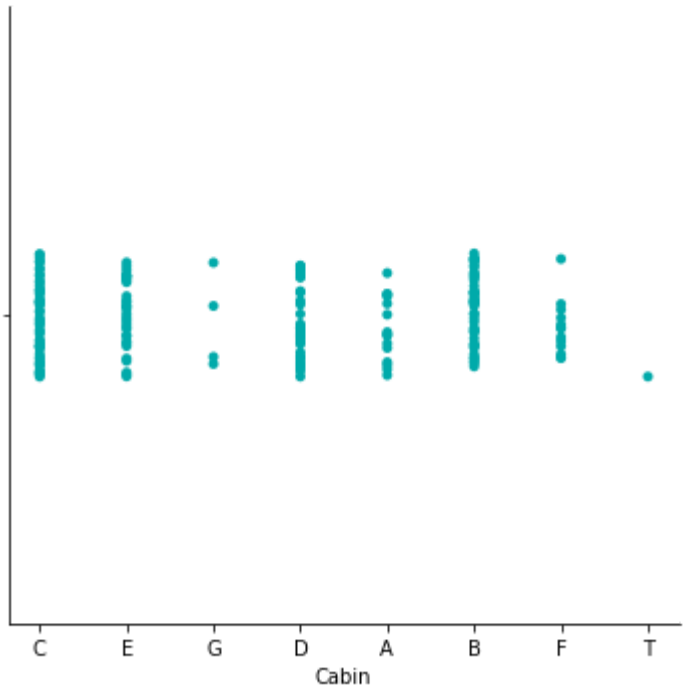
```
1      C85
3     C123
6      E46
10      G6
11    C103
Name: Cabin, dtype: object
```

In [118]:

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

cabin_df=DataFrame(levels)
cabin_df.columns=['Cabin']
#Plotting for the cabins occupied by paasengers
sns.catplot('Cabin', data=cabin_df,palette='winter_d')
```

Out[118]:

```
<seaborn.axisgrid.FacetGrid at 0x20198909b00>
```

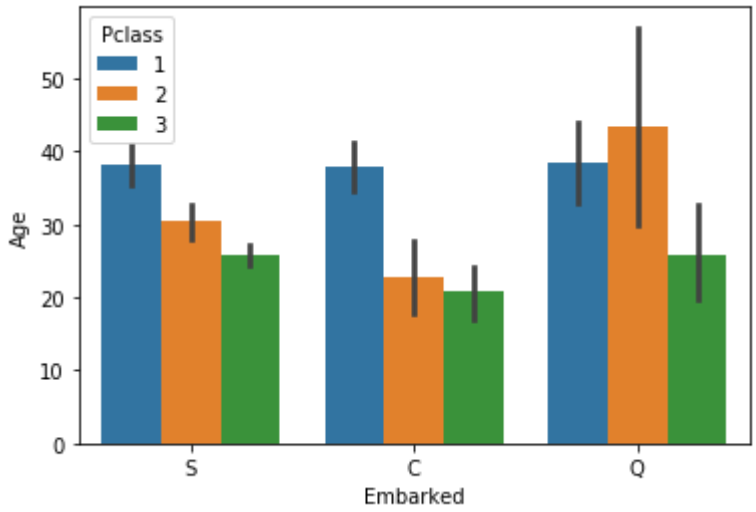


In [149]:

```
#3.Where did the passengers come from and in which class?
sns.barplot('Embarked','Age',data=titanic_df,hue='Pclass')
```

Out[149]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x2019a25b1d0>
```



In [18]:

```
#Adding a column to the tiatnic_df which contains information
#about passengers if they were travelling alone or not
titanic_df['Alone'] = titanic_df.SibSp + titanic_df.Parch
```

In [19]:

```
titanic_df['Alone']
```

Out[19]:

```
0       1
1       1
2       0
3       1
4       0
5       0
6       0
7       4
8       2
9       1
10      2
11      0
12      0
13      6
14      0
15      0
16      5
17      0
18      1
19      0
20      0
21      0
22      0
23      0
24      4
25      6
26      0
27      5
28      0
29      0
       ..
861     1
862     0
863    10
864     0
865     0
866     1
867     0
868     0
869     2
870     0
871     2
872     0
873     0
874     1
875     0
876     0
877     0
878     0
879     1
880     1
881     0
882     0
883     0
884     0
885     5
886     0
887     0
888     3
```

```
889     0
890     0
Name: Alone, Length: 891, dtype: int64
```

In [20]:

```
#4.Who was with family and without family?
#Changing value of the column Alone
titanic_df['Alone'].loc[titanic_df['Alone']>0] = 'With Family'
titanic_df['Alone'].loc[titanic_df['Alone']==0] = 'Alone'
```

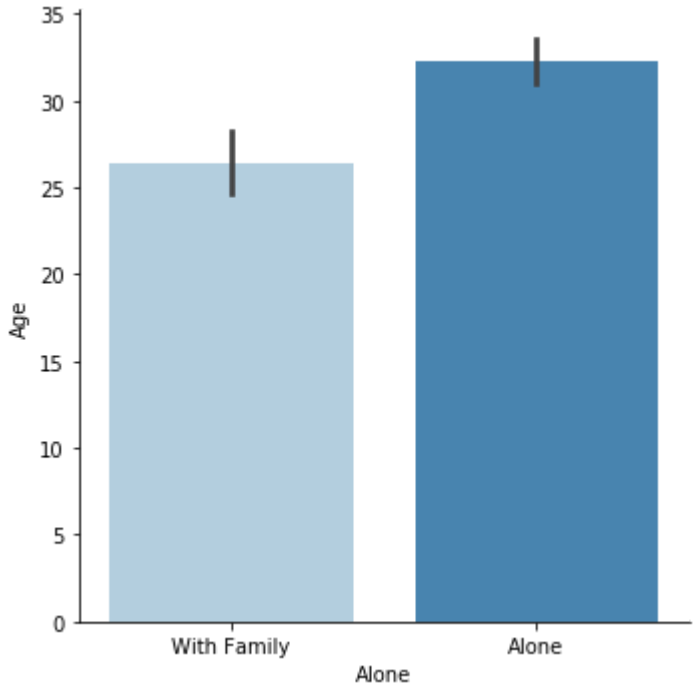
```
c:\users\ishita pamnani\appdata\local\programs\python\python37\lib\site-pa
ckages\pandas\core\indexing.py:190: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: http://pandas.pydata.org/pandas-doc
s/stable/indexing.html#indexing-view-versus-copy
  self._setitem_with_indexer(indexer, value)
```

In [27]:

```
sns.factorplot('Alone', 'Age',data=titanic_df,palette='Blues',kind='bar')
```

Out[27]:

```
<seaborn.axisgrid.FacetGrid at 0x14c14ff12b0>
```

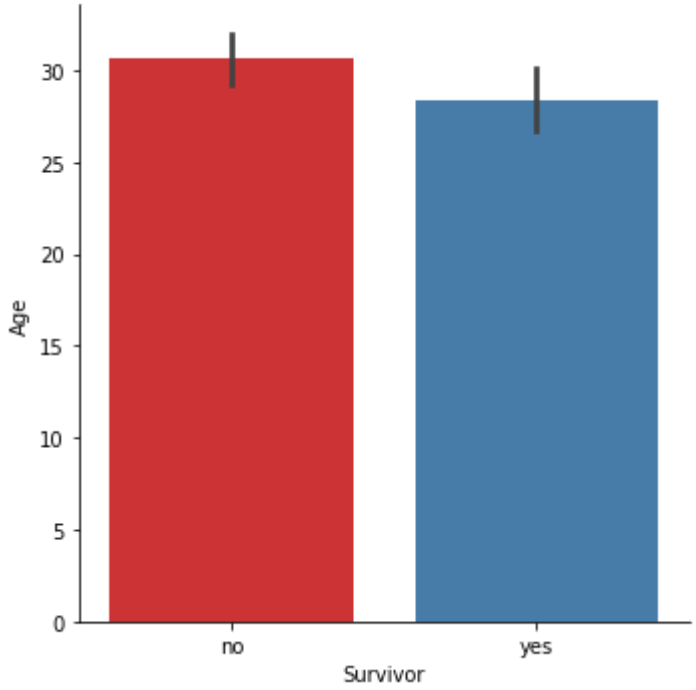


In [46]:

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

sns.factorplot('Survivor','Age',data=titanic_df,palette='Set1',kind='bar')
```

Out[46]:

```
<seaborn.axisgrid.FacetGrid at 0x14c15316550>
```

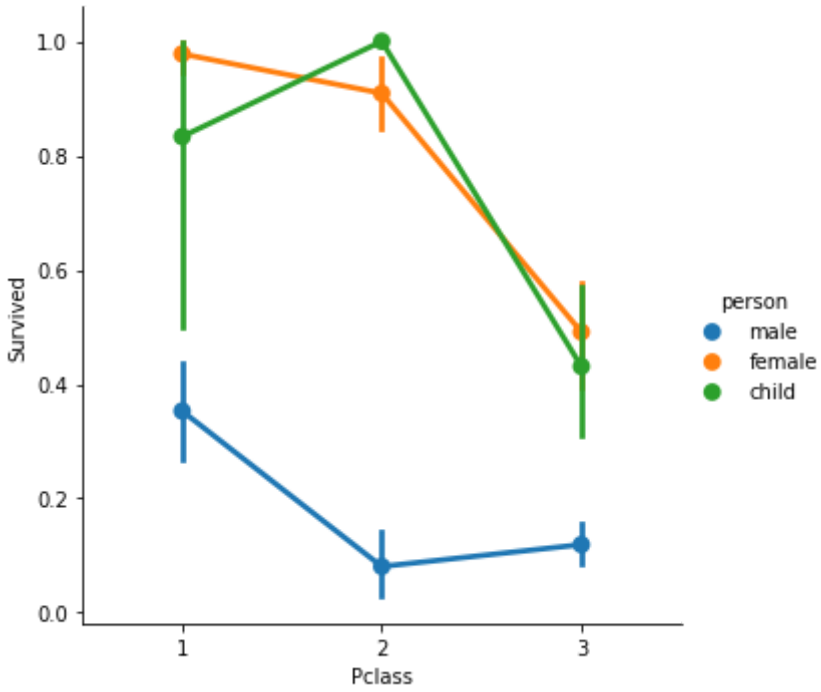


In [57]:

```
#5.What factors made passengers survived?
sns.factorplot('Pclass','Survived',hue='person',data=titanic_df)
```

Out[57]:

```
<seaborn.axisgrid.FacetGrid at 0x14c16447ac8>
```

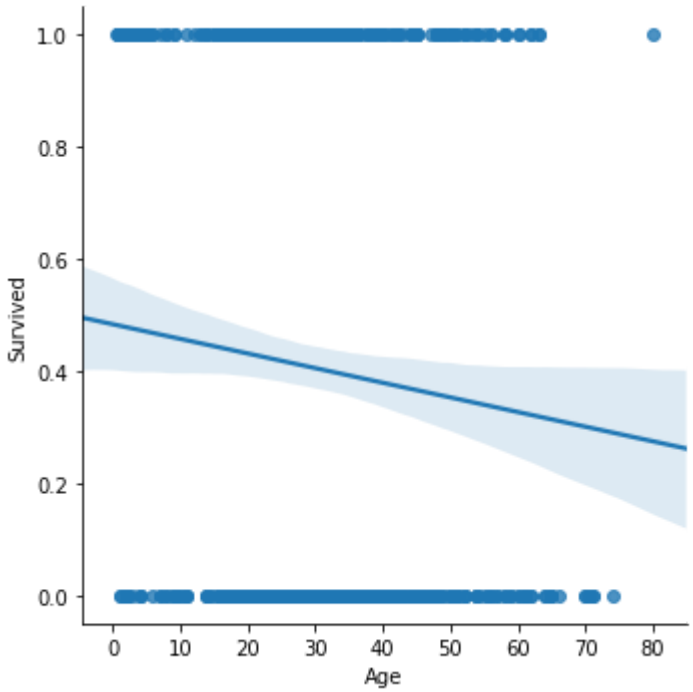


In [58]:

```
sns.lmplot('Age','Survived',data=titanic_df)
```

Out[58]:

```
<seaborn.axisgrid.FacetGrid at 0x14c16502cf8>
```

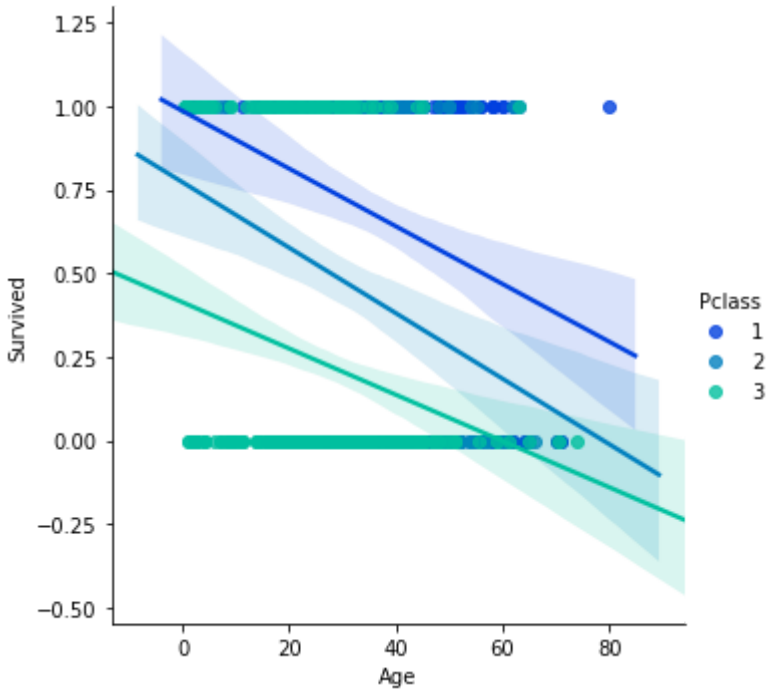


In [59]:

```
sns.lmplot('Age','Survived',hue='Pclass',data=titanic_df,palette='winter')
```

Out[59]:

```
<seaborn.axisgrid.FacetGrid at 0x14c165ad2e8>
```

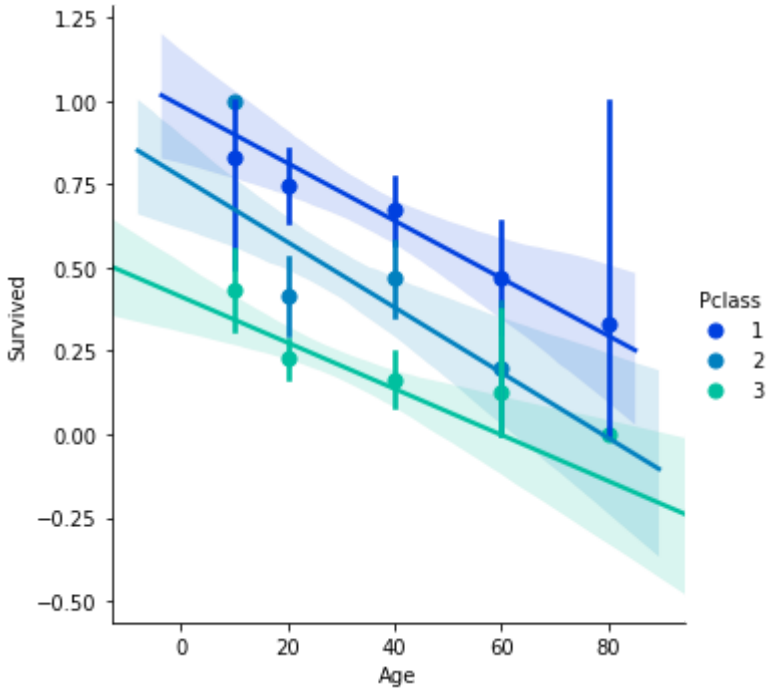


In [61]:

```
generations=[10,20,40,60,80]

sns.lmplot('Age','Survived',hue='Pclass',data=titanic_df,palette='winter',x_bins=genera
tions)
```

Out[61]:

```
<seaborn.axisgrid.FacetGrid at 0x14c16632748>
```



In [62]:

```
sns.lmplot('Age','Survived',hue='Sex',data=titanic_df,palette='winter',x_bins=generations)
```

Out[62]:

```
<seaborn.axisgrid.FacetGrid at 0x14c166ab7b8>
```