TitanicProjectpython

February 26, 2017

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

In [2]: titanic_df = pd.read_csv('https://kaggle2.blob.core.windows.net/competitior

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

Out[3]:

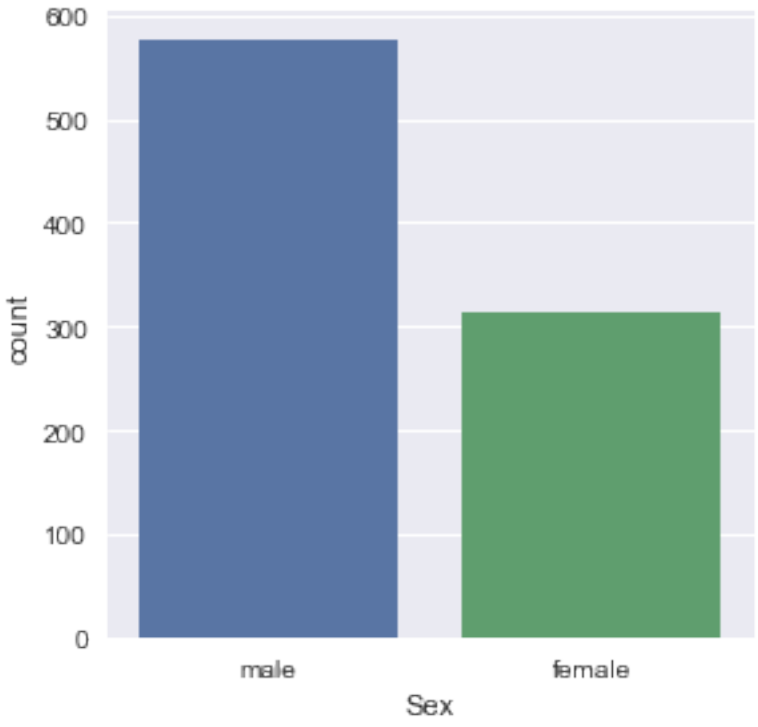
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 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 [4]: 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
```

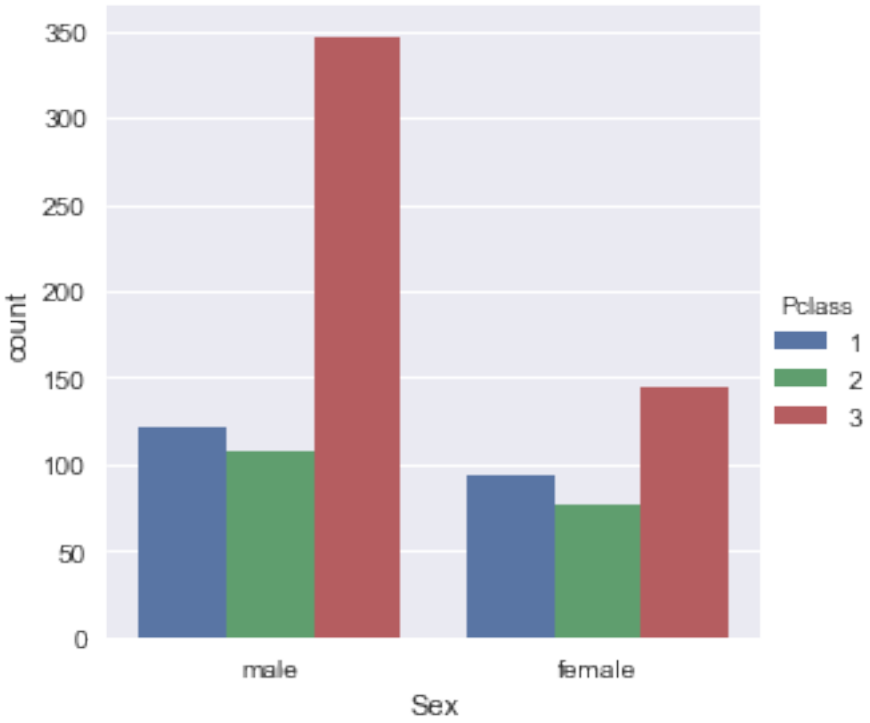
```
Age            714 non-null float64
SibSp          891 non-null int64
Parch          891 non-null int64
Ticket         891 non-null object
Fare           891 non-null float64
Cabin          204 non-null object
Embarked       889 non-null object
dtypes: float64(2), int64(5), object(5)
memory usage: 83.6+ KB
```

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?

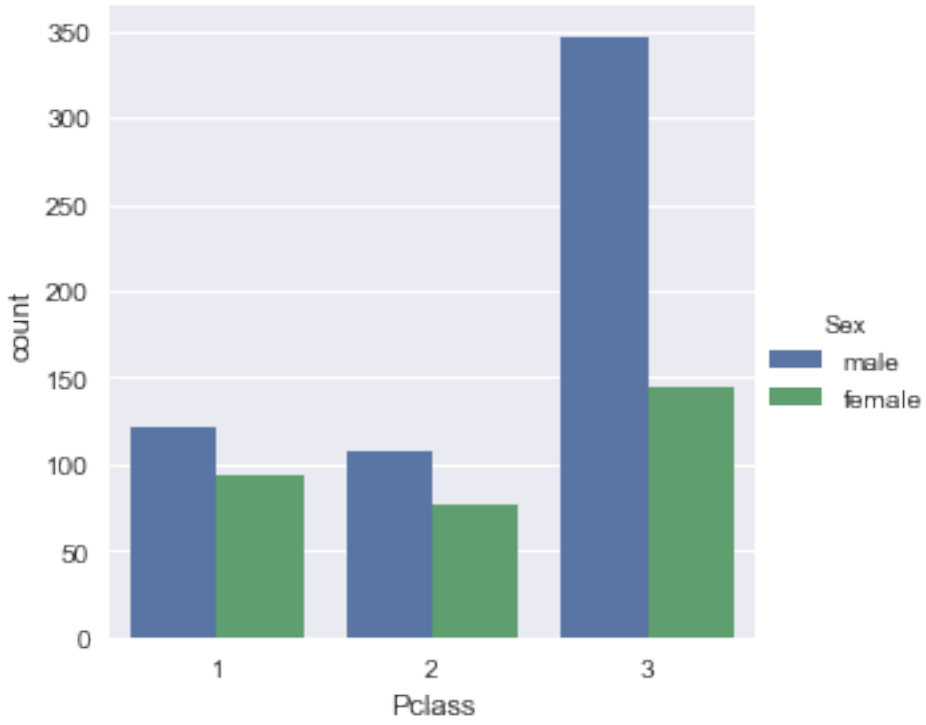
```
In [5]: #q1
        import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline

In [6]: sns.factorplot('Sex',data=titanic_df,kind='count')

Out[6]: <seaborn.axisgrid.FacetGrid at 0x1139ed2d0>
```



```
In [7]: sns.factorplot('Sex',data=titanic_df,hue='Pclass',kind='count')

Out[7]: <seaborn.axisgrid.FacetGrid at 0x117163710>
```




```
In [8]: sns.factorplot('Pclass',data=titanic_df,hue='Sex',kind='count')

Out[8]: <seaborn.axisgrid.FacetGrid at 0x11743ba50>
```




```
In [9]: # a column that has male, female and child based on the age and sex

        def male_female_child(passenger):
            age,sex=passenger

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

```
In [10]: titanic_df['person'] = titanic_df[['Age','Sex']].apply(male_female_child,a
```

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

Out[11]:

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	
5	6	0	3	
6	7	0	1	
7	8	0	3	
8	9	1	3	

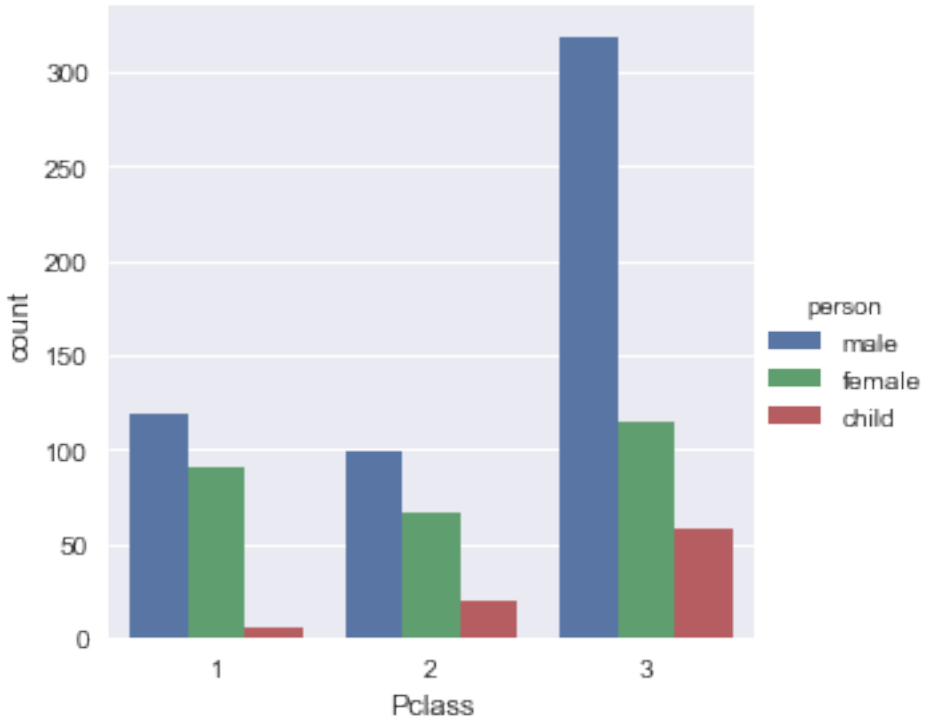
```
9           10         1       2
```

	Name	Sex	Age	SibSp
0	Braund, Mr. Owen Harris	male	22.0	1
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1
2	Heikkinen, Miss. Laina	female	26.0	0
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4	Allen, Mr. William Henry	male	35.0	0
5	Moran, Mr. James	male	NaN	0
6	McCarthy, Mr. Timothy J	male	54.0	0
7	Palsson, Master. Gosta Leonard	male	2.0	3
8	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0
9	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1

	Parch	Ticket	Fare	Cabin	Embarked	person
0	0	A/5 21171	7.2500	NaN	S	male
1	0	PC 17599	71.2833	C85	C	female
2	0	STON/O2. 3101282	7.9250	NaN	S	female
3	0	113803	53.1000	C123	S	female
4	0	373450	8.0500	NaN	S	male
5	0	330877	8.4583	NaN	Q	male
6	0	17463	51.8625	E46	S	male
7	1	349909	21.0750	NaN	S	child
8	2	347742	11.1333	NaN	S	female
9	0	237736	30.0708	NaN	C	child

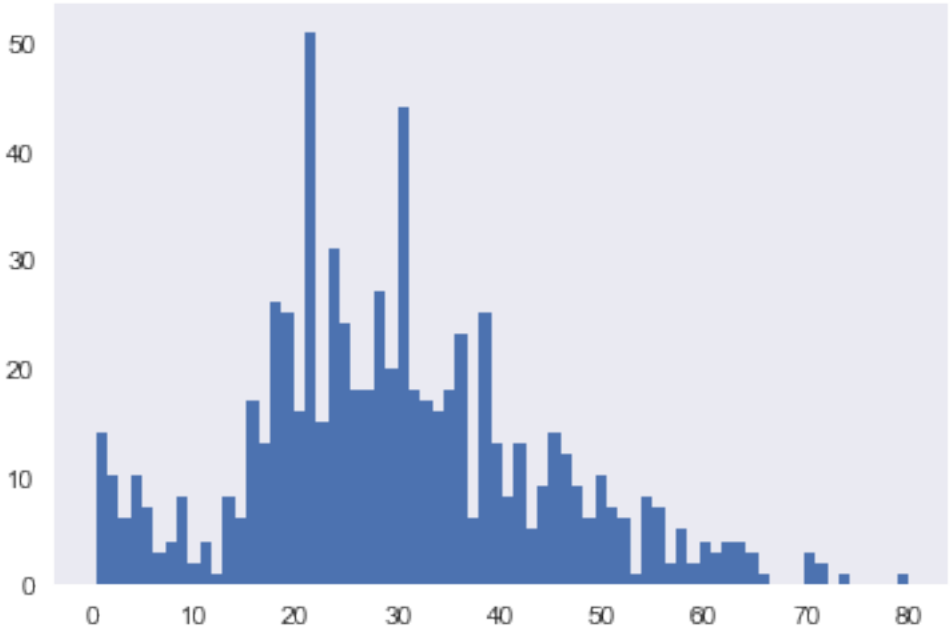
In [12]:
```
sns.factorplot('Pclass',data=titanic_df,kind='count',hue='person')
```

Out[12]:
```
<seaborn.axisgrid.FacetGrid at 0x11754be50>
```

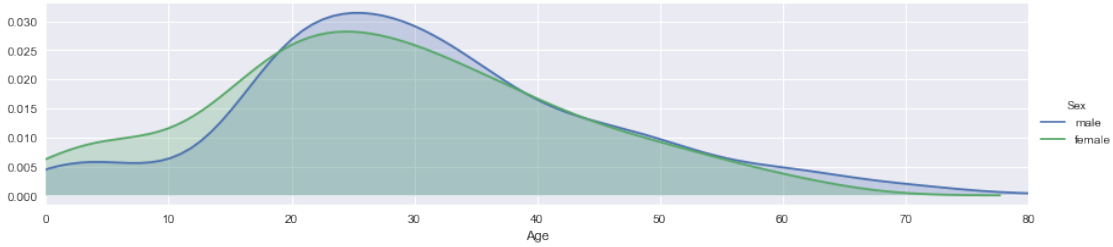



```
In [13]: titanic_df['Age'].hist(bins=70,grid=False)
```

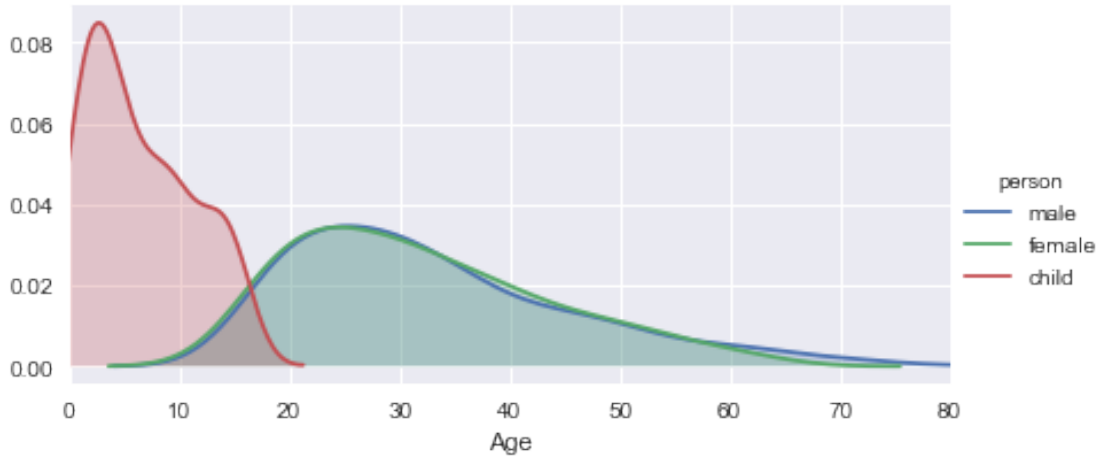
```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x117787090>
```



```
In [14]: titanic_df['Age'].mean()

Out[14]: 29.69911764705882

In [185]: titanic_df['person'].value_counts()

Out[185]: male      537
          female    271
          child      83
          Name: person, dtype: int64

In [18]: 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[18]: <seaborn.axisgrid.FacetGrid at 0x1133dc050>
```

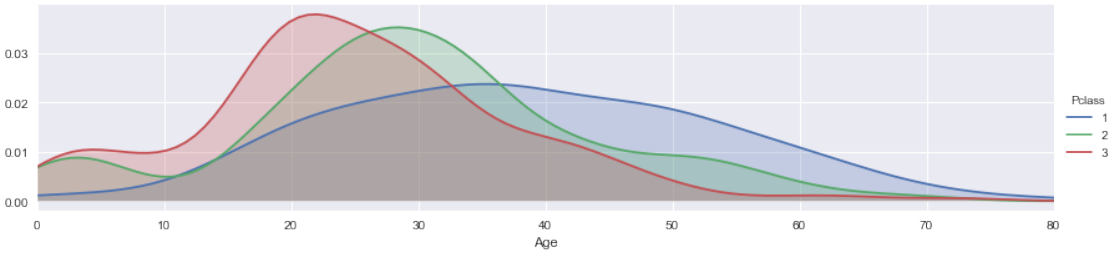



```
In [23]: fig = sns.FacetGrid(titanic_df,hue='person',aspect=2)

         fig.map(sns.kdeplot,'Age',shade=True)

         oldest = titanic_df['Age'].max()
         fig.set(xlim=(0,oldest))
         fig.add_legend()

Out[23]: <seaborn.axisgrid.FacetGrid at 0x1140d20d0>
```




```
In [217]: 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[217]: <seaborn.axisgrid.FacetGrid at 0x118450a90>
```




```
In [189]: #2.) What deck were the passengers on and how does that relate to their c

          #q2
```

```
In [190]: titanic_df.head()
```

```
Out[190]:    PassengerId  Survived  Pclass  \
          0            1         0       3
          1            2         1       1
          2            3         1       3
```

```
3            4         1       1
4            5         0       3

                                                 Name     Sex   Age  SibSp
0                             Braund, Mr. Owen Harris    male  22.0      1
1   Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                              Heikkinen, Miss. Laina  female  26.0      0
3        Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                            Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked  person
0      0         A/5 21171   7.2500   NaN        S    male
1      0          PC 17599  71.2833   C85        C  female
2      0  STON/O2. 3101282   7.9250   NaN        S  female
3      0            113803  53.1000  C123        S  female
4      0            373450   8.0500   NaN        S    male
```

In [191]:
```
deck=titanic_df['Cabin'].dropna()
```

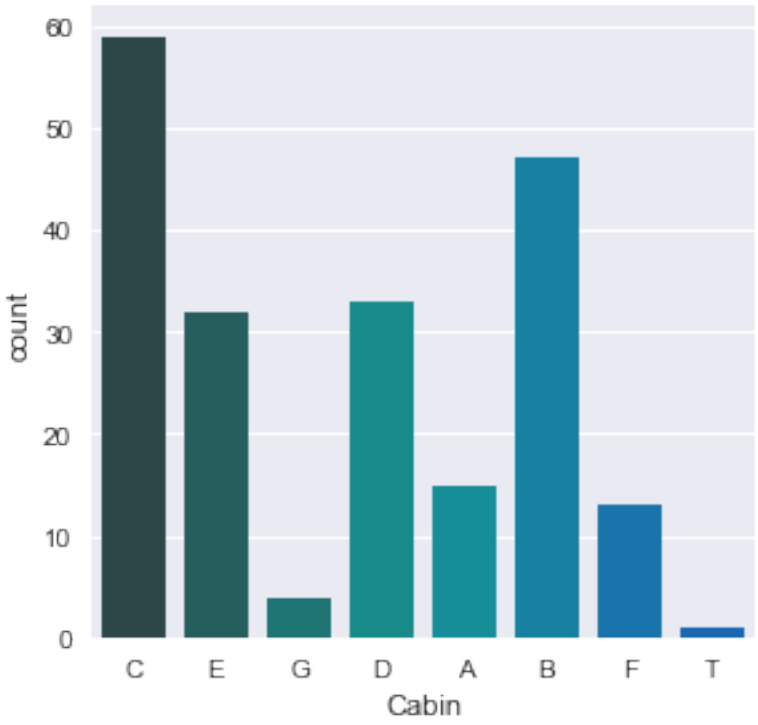
In [192]:
```
deck.head()
```

Out[192]:
```
1      C85
3     C123
6      E46
10      G6
11    C103
Name: Cabin, dtype: object
```

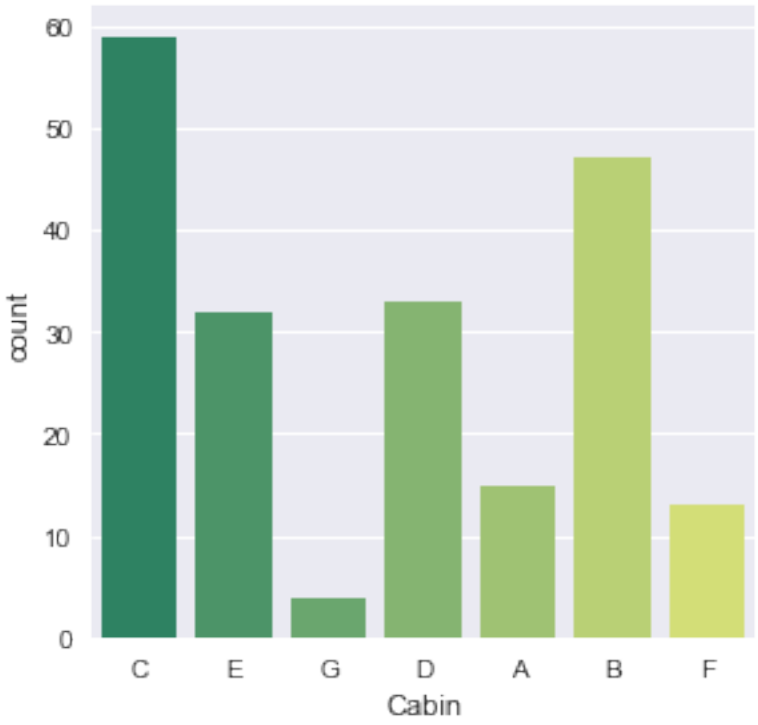
In [193]:
```
levels = []

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

cabin_df = DataFrame(levels)

cabin_df.columns=['Cabin']
cabin_df.sort()
sns.factorplot('Cabin', data = cabin_df, palette = 'winter_d',kind='count
```

```
/Users/arnavsomani/anaconda/lib/python2.7/site-packages/ipykernel/__main__.py:9: Fu
```

Out[193]:
```
<seaborn.axisgrid.FacetGrid at 0x116d26750>
```




```
In [194]: cabin_df = cabin_df[cabin_df.Cabin!="T"]
          sns.factorplot('Cabin', data = cabin_df, palette = 'summer',kind='count')

Out[194]: <seaborn.axisgrid.FacetGrid at 0x116543290>
```



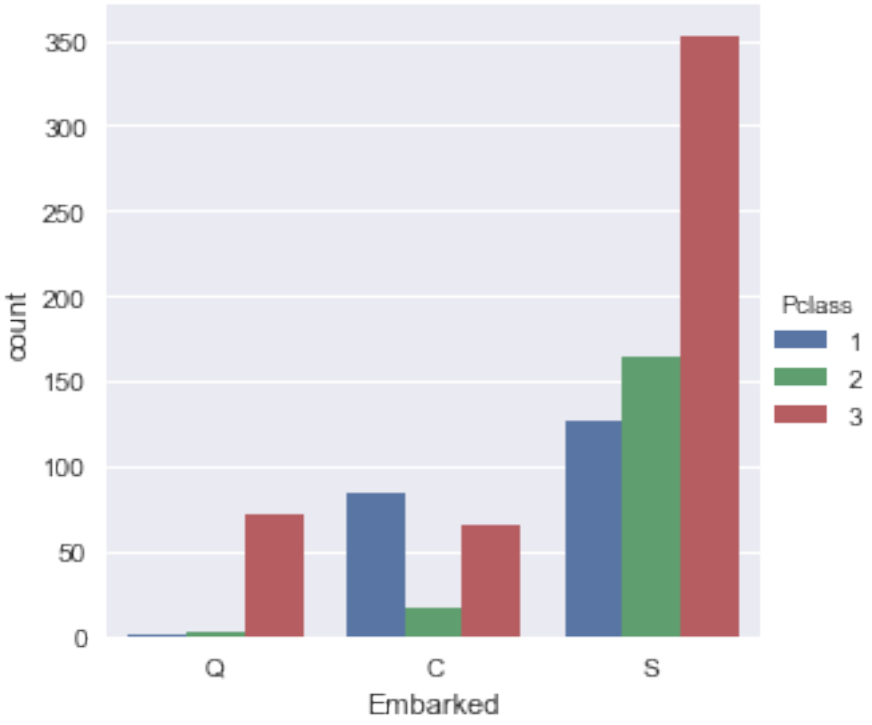

```
In [195]: titanic_df.head()
```

Out[195]:

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

```
In [196]: #q3
          sns.factorplot('Embarked',data=titanic_df,kind='count',hue='Pclass',x_orc
```

Out[196]: <seaborn.axisgrid.FacetGrid at 0x116fb6150>



In [197]: *#q4 who was alone and who was with family*

```
titanic_df.head()
```

Out[197]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1
4	5	0	3	Allen, Mr. William Henry	male	35.0	0

	Parch	Ticket	Fare	Cabin	Embarked	person
0	0	A/5 21171	7.2500	NaN	S	male
1	0	PC 17599	71.2833	C85	C	female
2	0	STON/O2. 3101282	7.9250	NaN	S	female

```
3      0            113803  53.1000  C123        S  female
4      0            373450   8.0500   NaN        S    male
```

```
In [198]: titanic_df['Alone'] = titanic_df.SibSp + titanic_df.Parch
```

```
In [199]: titanic_df['Alone'].head()
```

```
Out[199]: 0    1
          1    1
          2    0
          3    1
          4    0
          Name: Alone, dtype: int64
```

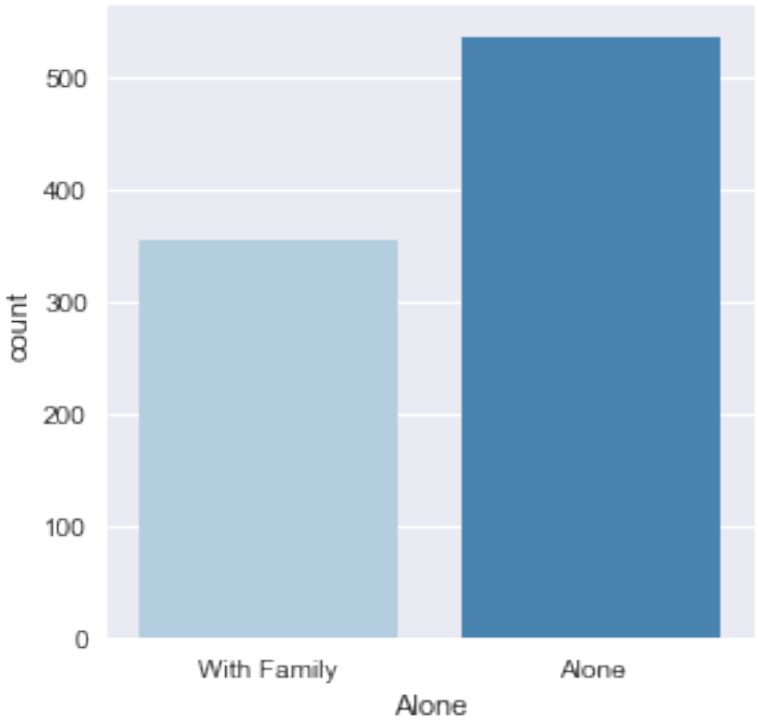
```
In [200]: titanic_df['Alone'].loc[titanic_df['Alone'] > 0] = 'With Family'

          titanic_df['Alone'].loc[titanic_df['Alone'] == 0] = 'Alone'
```

```
In [201]: titanic_df.head()
```

```
Out[201]:    PassengerId  Survived  Pclass  \
          0            1         0       3
          1            2         1       1
          2            3         1       3
          3            4         1       1
          4            5         0       3

                                                          Name     Sex   Age  SibSp
          0                            Braund, Mr. Owen Harris    male  22.0      1
          1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
          2                             Heikkinen, Miss. Laina  female  26.0      0
          3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
          4                           Allen, Mr. William Henry    male  35.0      0

             Parch            Ticket     Fare Cabin Embarked  person        Alone
          0      0         A/5 21171   7.2500   NaN        S    male  With Family
          1      0          PC 17599  71.2833   C85        C  female  With Family
          2      0  STON/O2. 3101282   7.9250   NaN        S  female        Alone
          3      0            113803  53.1000  C123        S  female  With Family
          4      0            373450   8.0500   NaN        S    male        Alone
```

```
In [202]: sns.factorplot('Alone',data=titanic_df,kind='count',palette='Blues')
```

```
Out[202]: <seaborn.axisgrid.FacetGrid at 0x116f70c90>
```




```
In [203]: # what factors helped someone survive the sining or crash of the Titanic

In [204]: titanic_df['Survivor'] = titanic_df.Survived.map({0:'no',1:'yes'})

In [205]: titanic_df.head()
```

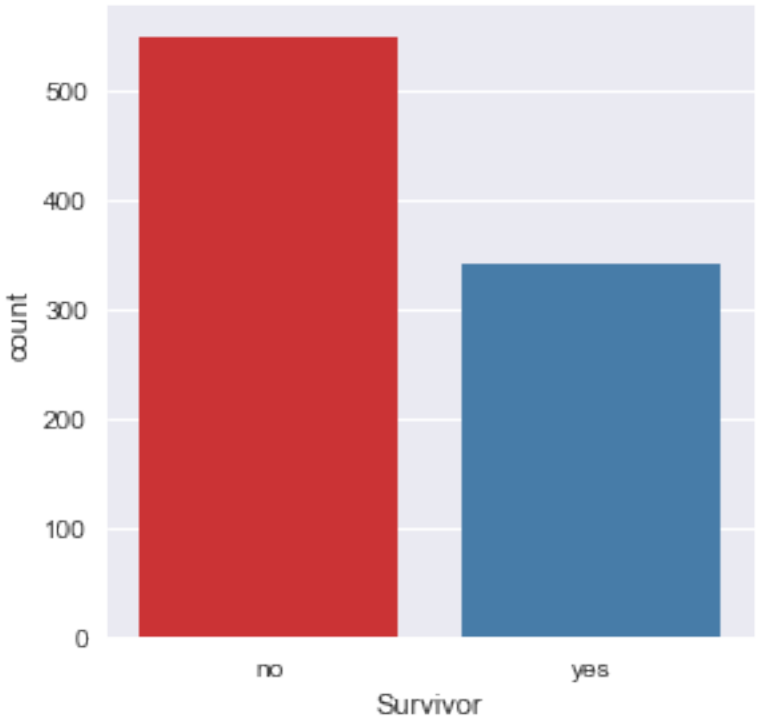
Out[205]:

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

```
3      0            113803  53.1000  C123        S  female  With Family
4      0            373450   8.0500   NaN        S    male        Alone

  Survivor
0       no
1      yes
2      yes
3      yes
4       no
```

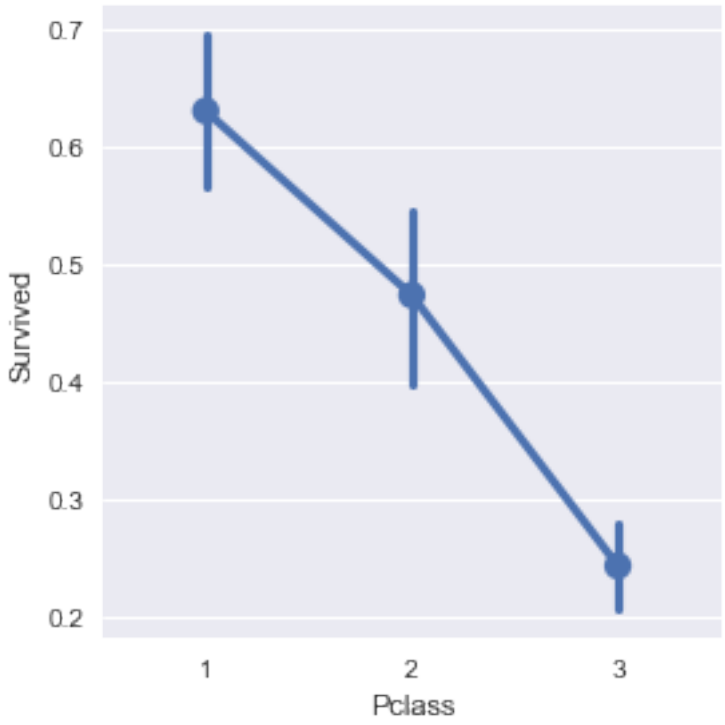
In [206]:
```
sns.factorplot('Survivor',data=titanic_df,kind='count',palette='Set1')
```

Out[206]: <seaborn.axisgrid.FacetGrid at 0x1175c0410>



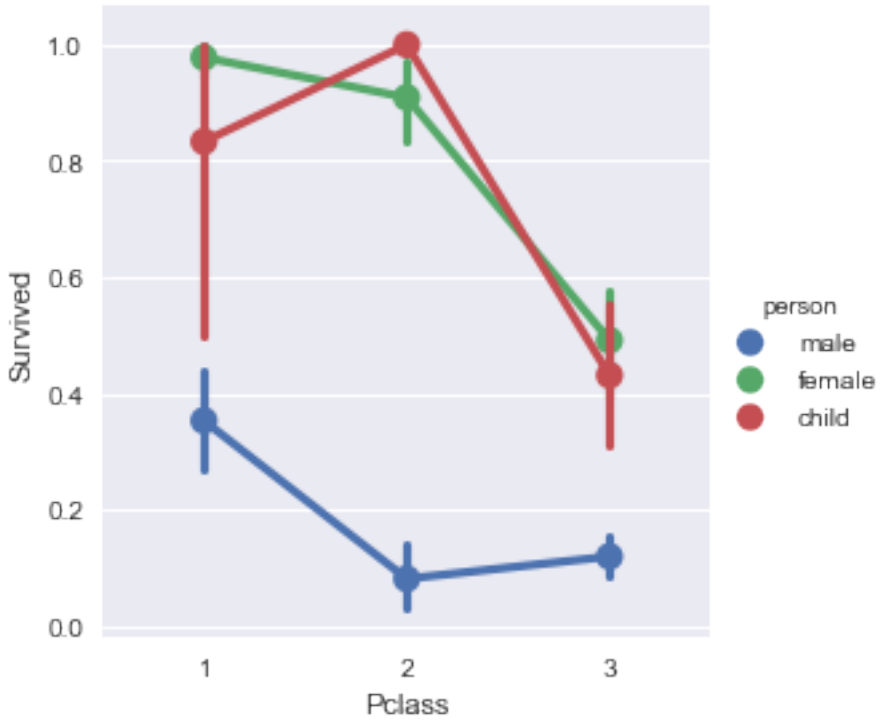
In [207]:
```
sns.factorplot('Pclass','Survived',data=titanic_df,)
```

Out[207]: <seaborn.axisgrid.FacetGrid at 0x116d04550>

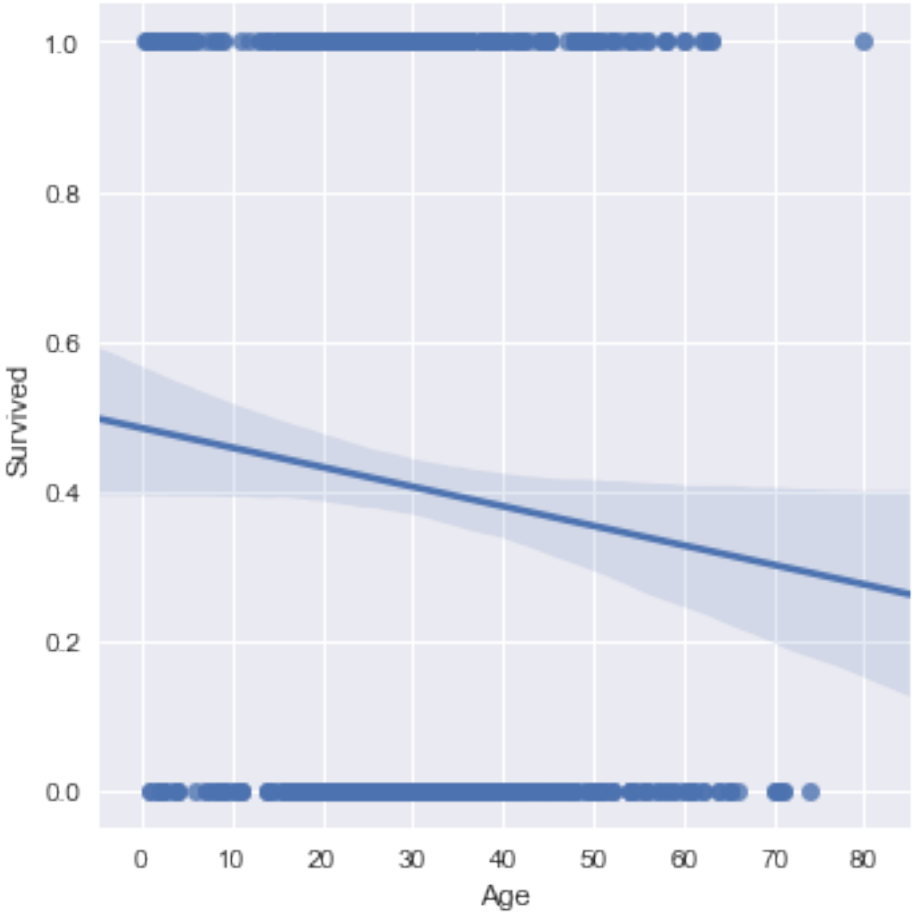



```
In [208]: sns.factorplot('Pclass','Survived',data=titanic_df,hue='person')
```

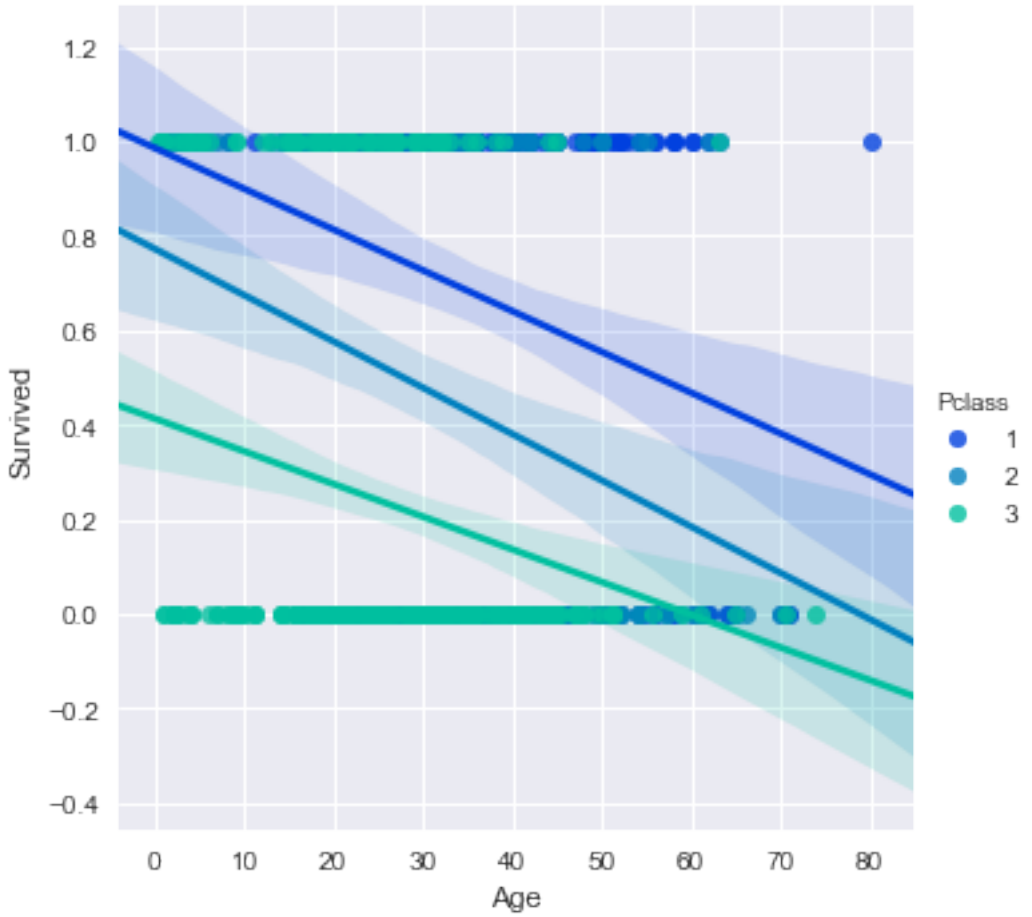
```
Out[208]: <seaborn.axisgrid.FacetGrid at 0x1176626d0>
```



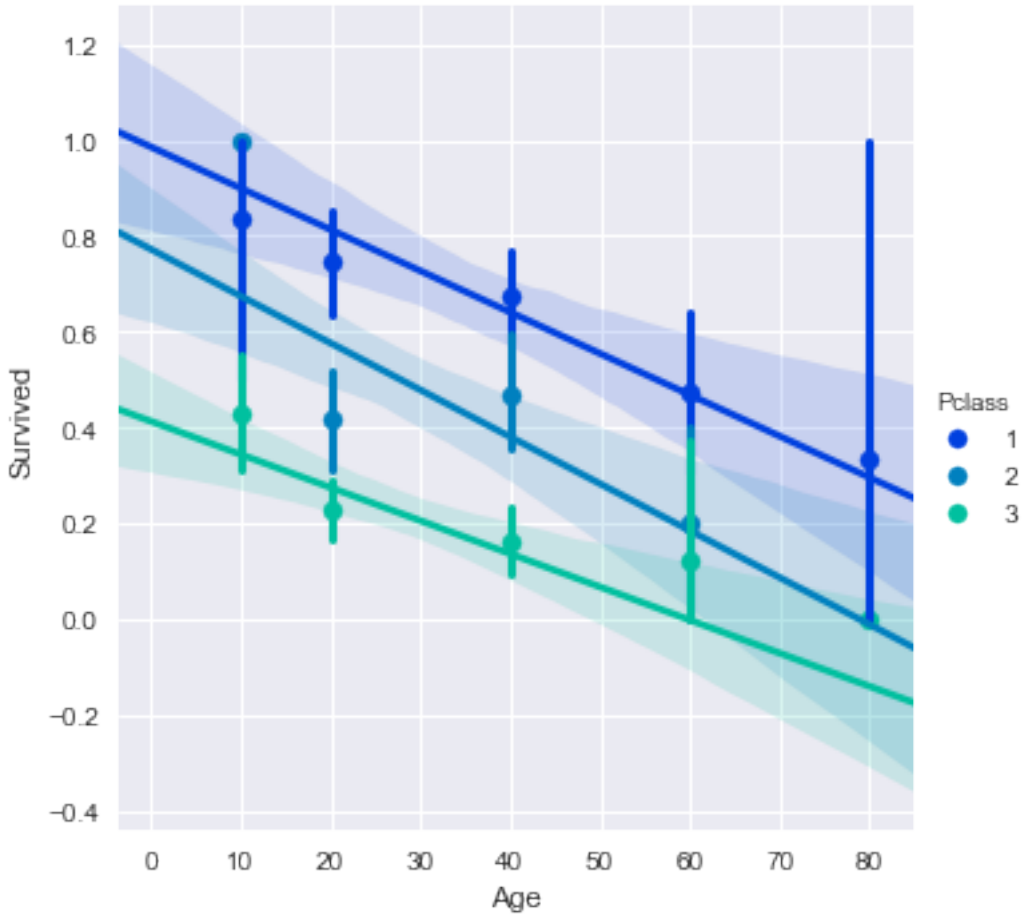
```
In [209]: sns.lmplot('Age','Survived',data=titanic_df)

Out[209]: <seaborn.axisgrid.FacetGrid at 0x11759fed0>
```




```
In [210]: sns.lmplot('Age','Survived',data=titanic_df,hue= 'Pclass',palette='winter

Out[210]: <seaborn.axisgrid.FacetGrid at 0x117b8a0d0>
```

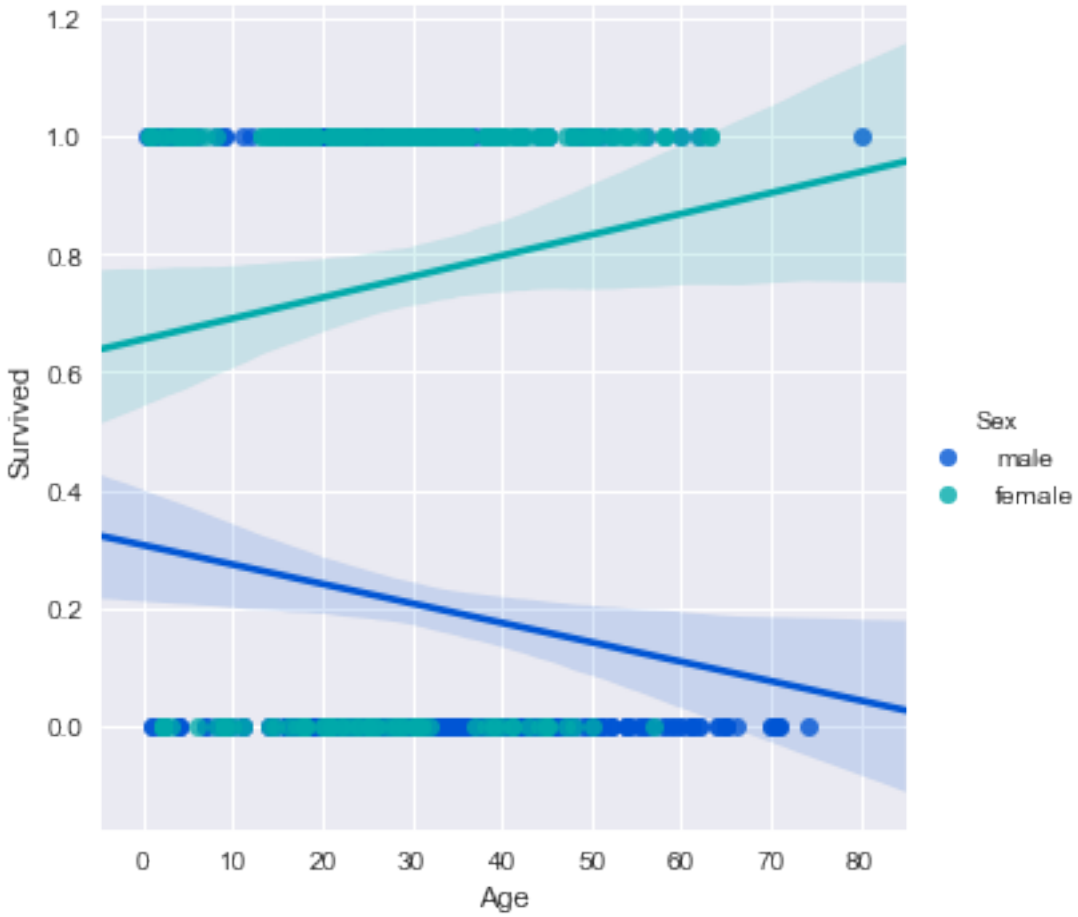



```
In [211]: generations = [10,20,40,60,80]

          sns.lmplot('Age','Survived',data=titanic_df,hue= 'Pclass',x_bins=generati

Out[211]: <seaborn.axisgrid.FacetGrid at 0x117ba1510>
```




```
In [212]: sns.lmplot('Age','Survived',data=titanic_df,hue= 'Sex',palette='winter')
```

```
Out[212]: <seaborn.axisgrid.FacetGrid at 0x117d2a590>
```


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Survived
0
10
20
30
40
50
60
70
80
Age
Sex
male
female