Titanic Project˜

November 15, 2016

In [2]:
```
import pandas as pd
from pandas import Series,DataFrame

titanic_df = pd.read_csv('train.csv')
titanic_df.head()
```

Out[2]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S

In [2]:
```
titanic_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 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: 90.5+ KB
```

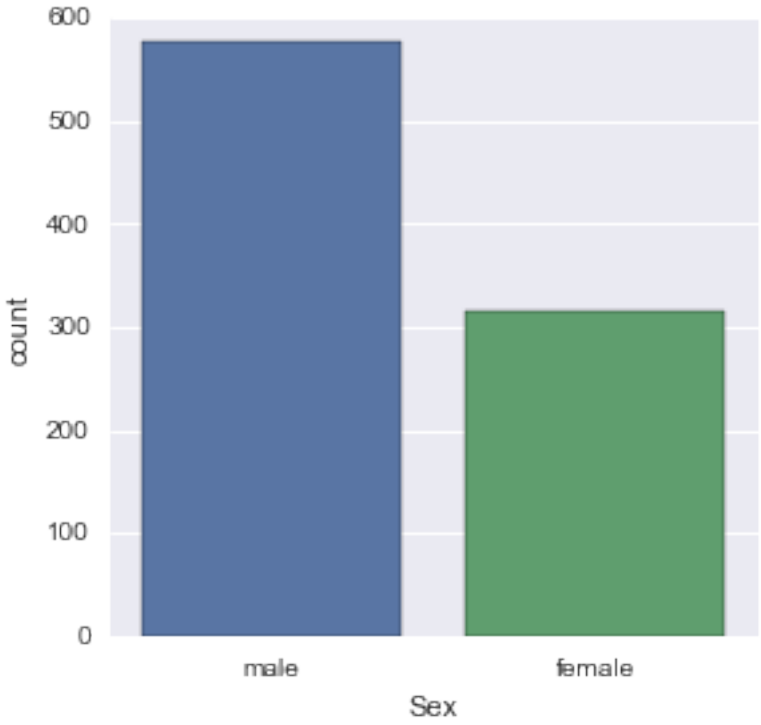
```
In [3]: import numpy as np
        import matplotlib.pyplot as plt
        import seaborn as sns
        %matplotlib inline
```

```
//anaconda/lib/python2.7/site-packages/matplotlib/__init__.py:872: UserWarning: axes.color_cycle is depre
  warnings.warn(self.msg_depr % (key, alt_key))
```

```
In [4]: sns.factorplot('Sex',data=titanic_df, kind='count')
```

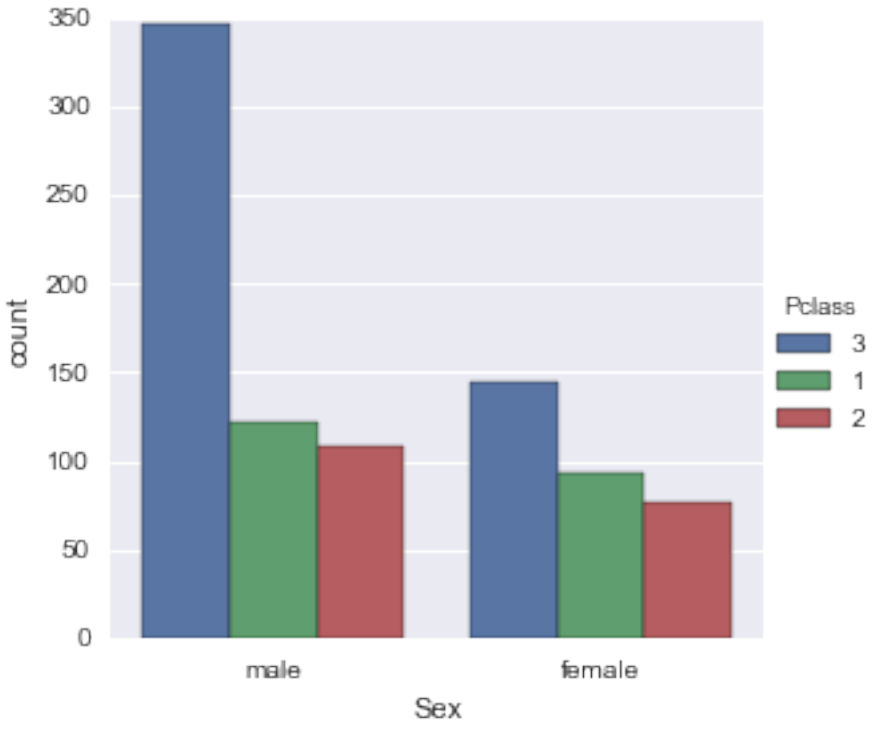
```
//anaconda/lib/python2.7/site-packages/matplotlib/__init__.py:892: UserWarning: axes.color_cycle is depre
  warnings.warn(self.msg_depr % (key, alt_key))
```

```
Out[4]: <seaborn.axisgrid.FacetGrid at 0x110181f10>
```



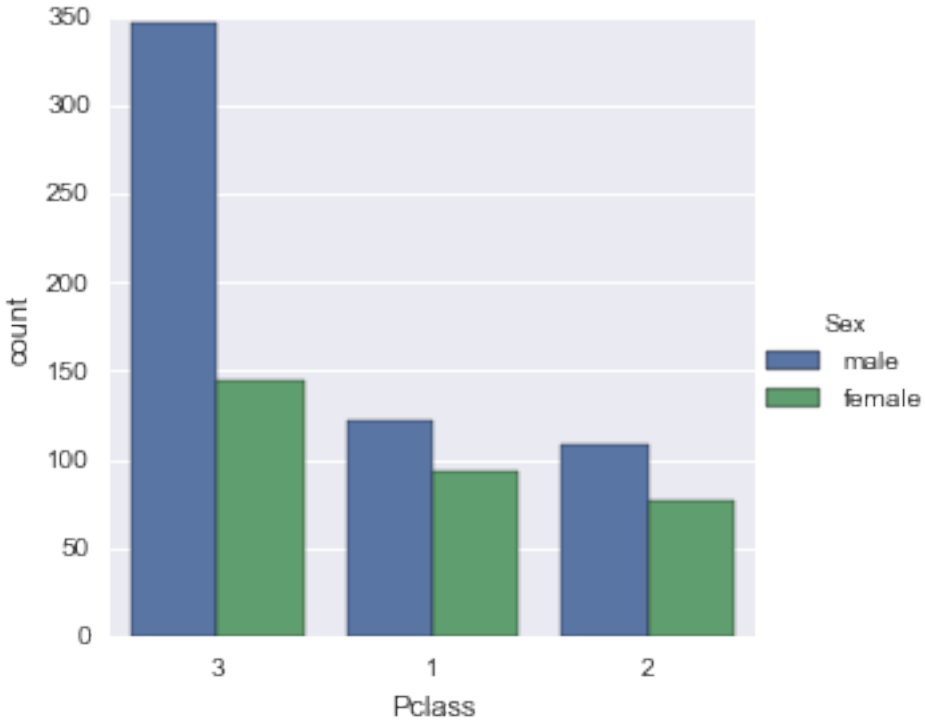

```
In [5]: sns.factorplot('Sex',data=titanic_df, hue='Pclass',kind='count')
```

```
Out[5]: <seaborn.axisgrid.FacetGrid at 0x11025f750>
```




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

```
Out[6]: <seaborn.axisgrid.FacetGrid at 0x11025f190>
```



```
In [7]: def male_female_child(passenger):
            age,sex = passenger

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

```
In [8]:  titanic_df['person'] = titanic_df[['Age','Sex']].apply(male_female_child,axis=1)
```

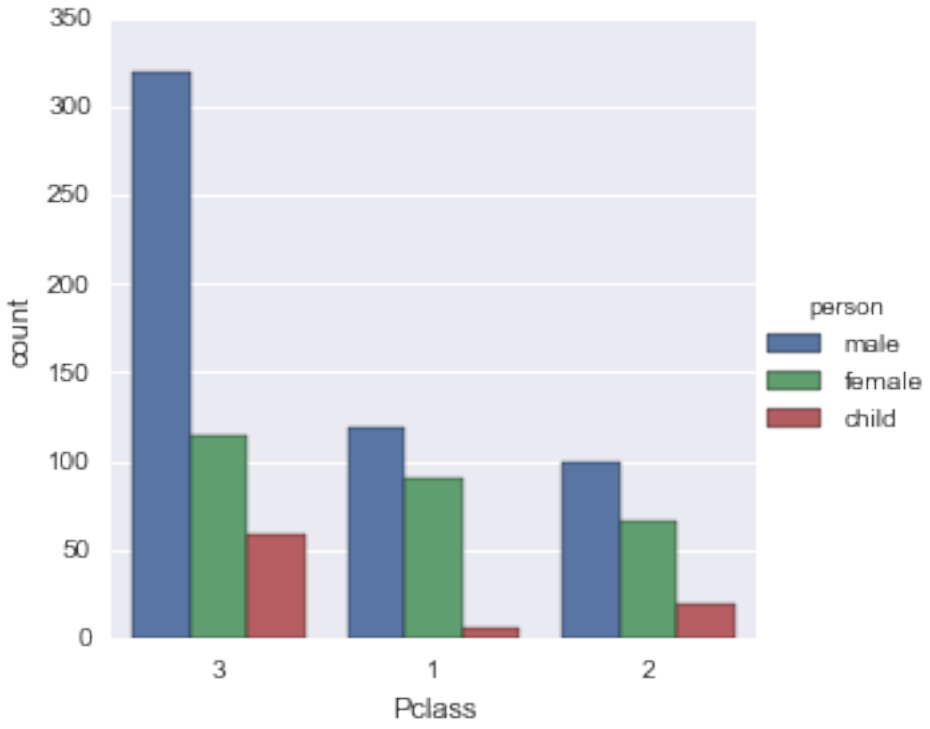
```
In [9]: titanic_df[0:10]
```

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NaN	S	male
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C	female
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S	female
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S	female
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S	male
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q	male
6	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S	male
7	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.0750	NaN	S	child
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333	NaN	S	female
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	30.0708	NaN	C	child

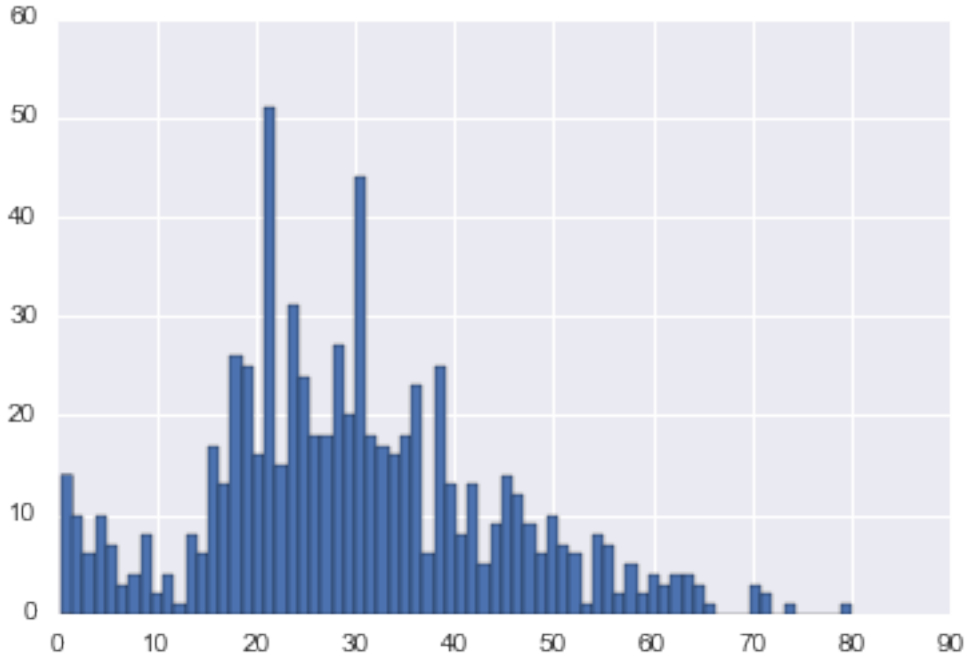
```
In [10]: sns.factorplot('Pclass',data=titanic_df, hue='person',kind='count')
```

```
Out[10]: <seaborn.axisgrid.FacetGrid at 0x10b141fd0>
```

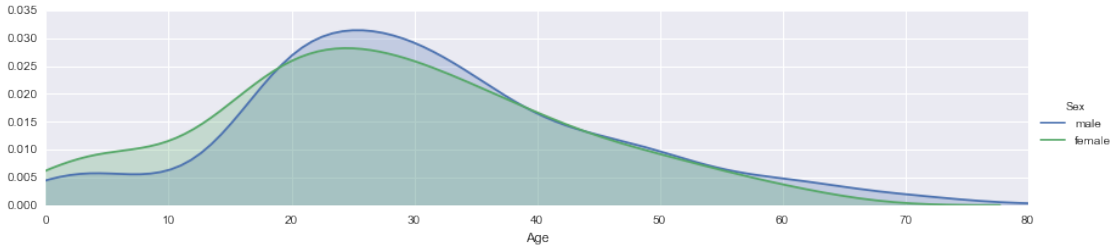



```
In [11]: titanic_df['Age'].hist(bins=70)
```

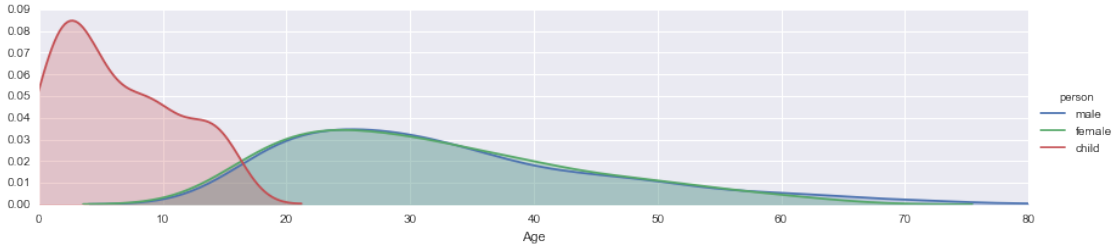
```
Out[11]: <matplotlib.axes._subplots.AxesSubplot at 0x1106aa650>
```



```
In [12]: titanic_df['Age'].mean()

Out[12]: 29.69911764705882

In [13]: titanic_df['person'].value_counts()

Out[13]: male      537
         female    271
         child      83
         Name: person, dtype: int64

In [14]: 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[14]: <seaborn.axisgrid.FacetGrid at 0x110495490>
```

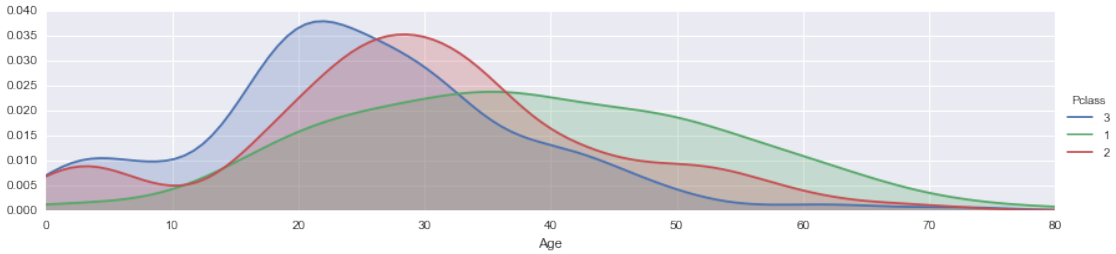



```
In [15]: fig = sns.FacetGrid(titanic_df, hue='person', aspect=4)

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

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

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

         fig.add_legend()

Out[15]: <seaborn.axisgrid.FacetGrid at 0x10b01c9d0>
```



In [16]:
```
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[16]: <seaborn.axisgrid.FacetGrid at 0x110c683d0>



In [17]:
```
titanic_df.head()
```

Out[17]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NaN	S	male
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C	female
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S	female
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S	female
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S	male

In [18]:
```
deck = titanic_df['Cabin'].dropna()
```

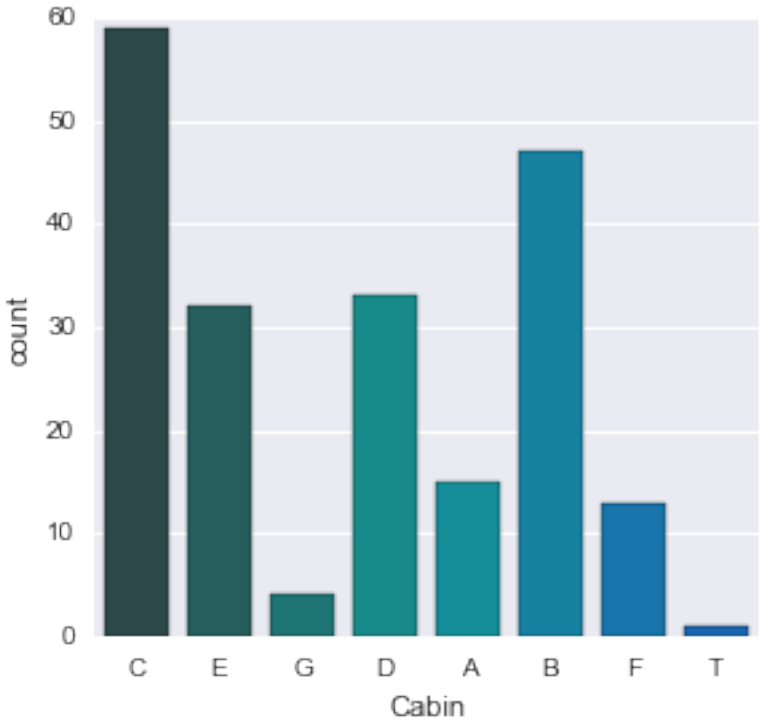
In [19]:
```
deck.head()
```

Out[19]:
```
1      C85
3     C123
```

```
6      E46
10      G6
11    C103
Name: Cabin, dtype: object
```

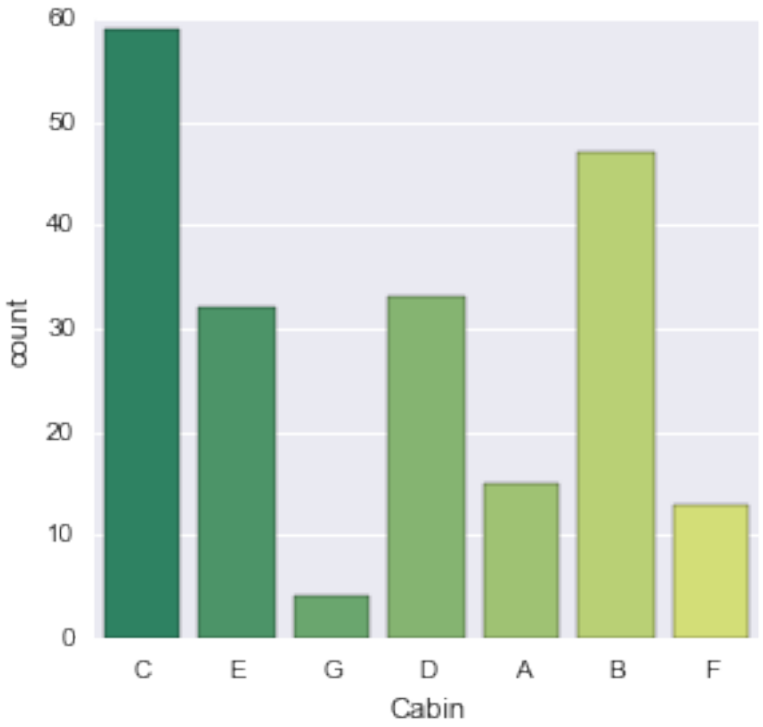
In [20]:
```
levels = []

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

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

Out[20]: <seaborn.axisgrid.FacetGrid at 0x1111828d0>



In [21]:
```
cabin_df = cabin_df[cabin_df.Cabin != 'T']
sns.factorplot('Cabin', data=cabin_df,palette='summer',kind="count")
```

Out[21]: <seaborn.axisgrid.FacetGrid at 0x11117a7d0>



In [22]: titanic_df.head()

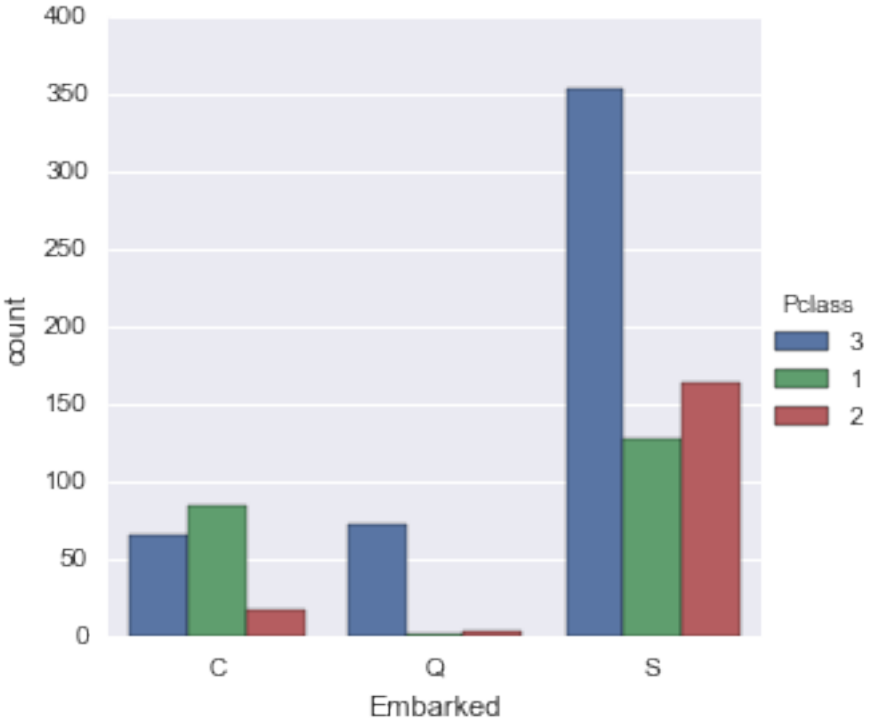
Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NaN	S	male
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C	female
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S	female
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S	female
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S	male

In [23]: sns.factorplot('Embarked',data=titanic_df,hue='Pclass',x_order=['C','Q','S'],kind='count')

//anaconda/lib/python2.7/site-packages/seaborn/categorical.py:2653: UserWarning: The ‘x_order‘ parameter
UserWarning)

Out[23]: <seaborn.axisgrid.FacetGrid at 0x1117f3710>



In [24]:
```
# who was alone and who was with family?
titanic_df.head()
```

Out[24]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
0	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.2500	NaN	S	male
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38	1	0	PC 17599	71.2833	C85	C	female
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S	female
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S	female
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S	male

In [25]:
```
titanic_df['Alone'] = titanic_df.SibSp + titanic_df.Parch
```

```
In [26]: titanic_df['Alone']
```

```
Out[26]: 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, dtype: int64
```

In [27]:
```
titanic_df['Alone'].loc[titanic_df['Alone'] >0] = 'With Family'

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

```
//anaconda/lib/python2.7/site-packages/pandas/core/indexing.py:117: 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-docs/stable/indexing.html#indexing
  self._setitem_with_indexer(indexer, value)
```

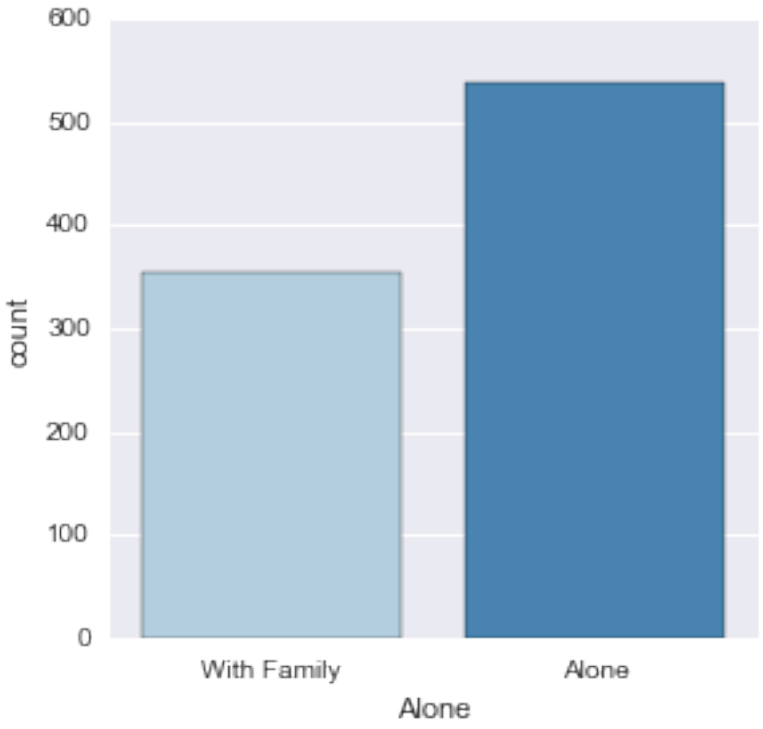
In [28]:
```
titanic_df.head()
```

Out[28]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person	Alone
0	1	0	3	Braund, Mr. Owen Harris	male	22	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	1	0	PC 17599	71.2833	C85	C	female	With Family
2	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.9250	NaN	S	female	Alone
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1000	C123	S	female	With Family
4	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.0500	NaN	S	male	Alone

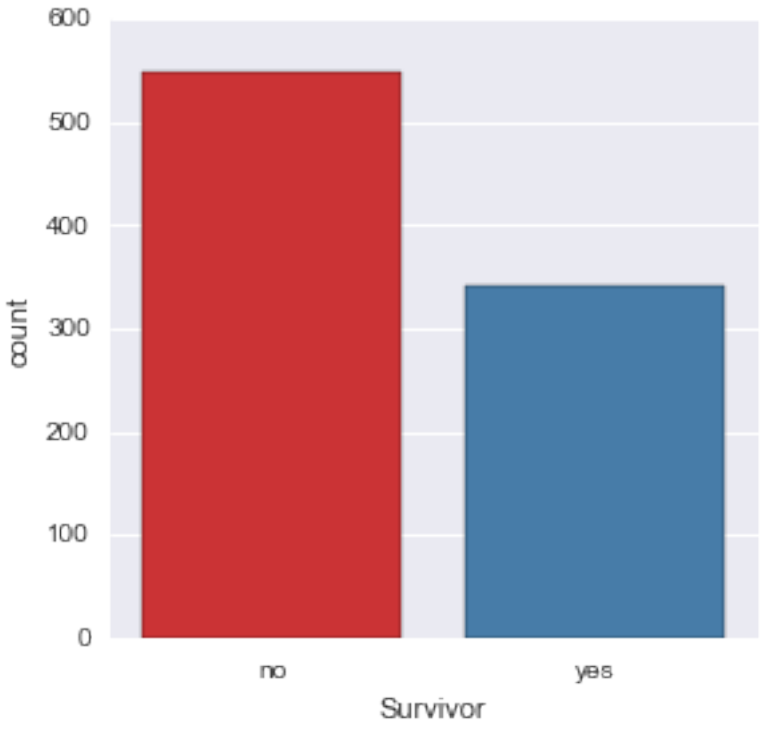
In [29]:
```
sns.factorplot('Alone',data=titanic_df,palette='Blues',kind='count')
```

Out[29]:
```
<seaborn.axisgrid.FacetGrid at 0x111430990>
```



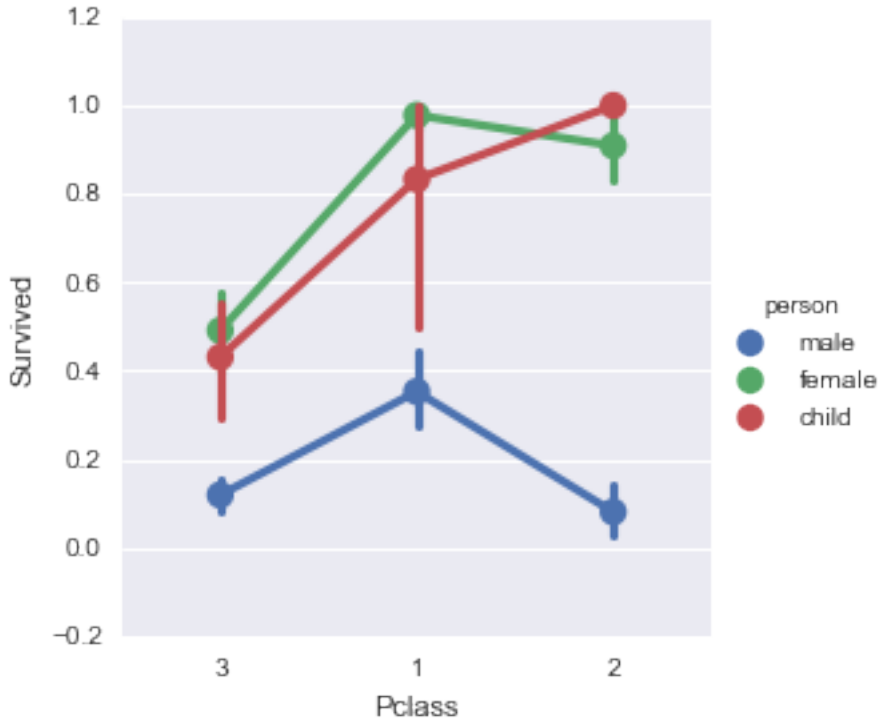

```
In [30]: titanic_df['Survivor'] = titanic_df.Survived.map({0:'no',1:'yes'})
         sns.factorplot('Survivor', data=titanic_df,palette='Set1',kind='count')
```

```
Out[30]: <seaborn.axisgrid.FacetGrid at 0x111b2b750>
```



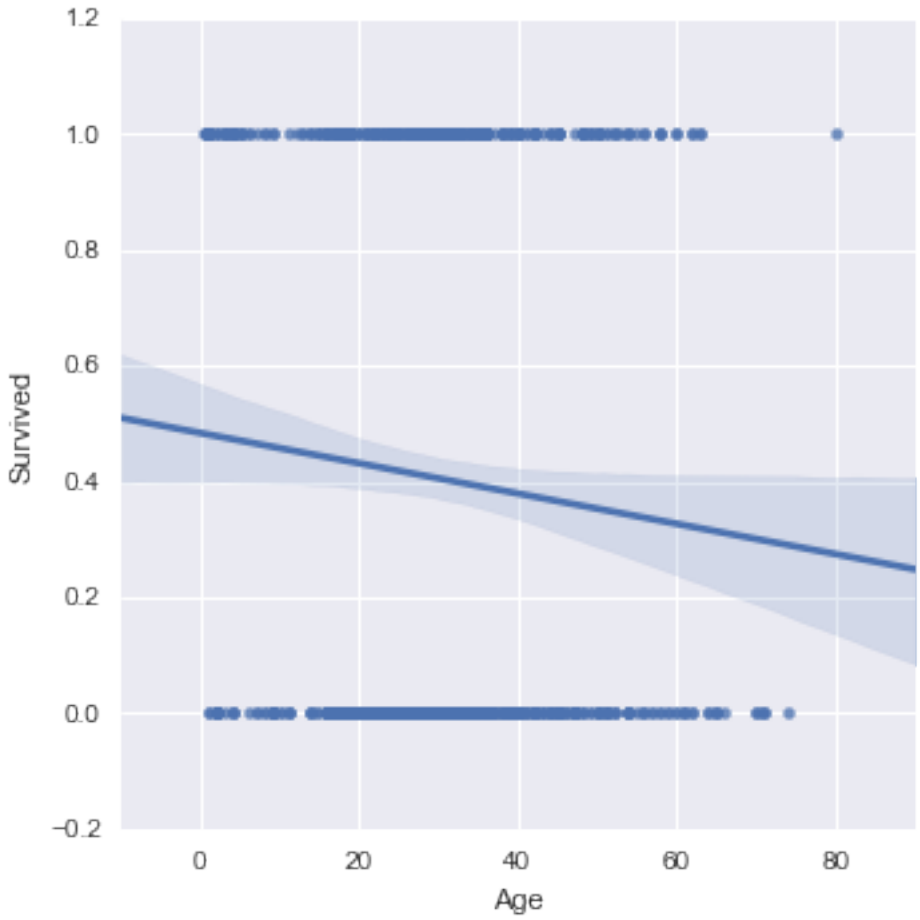

```
In [31]: sns.factorplot('Pclass','Survived',hue='person',data=titanic_df)
```

```
Out[31]: <seaborn.axisgrid.FacetGrid at 0x111b37210>
```



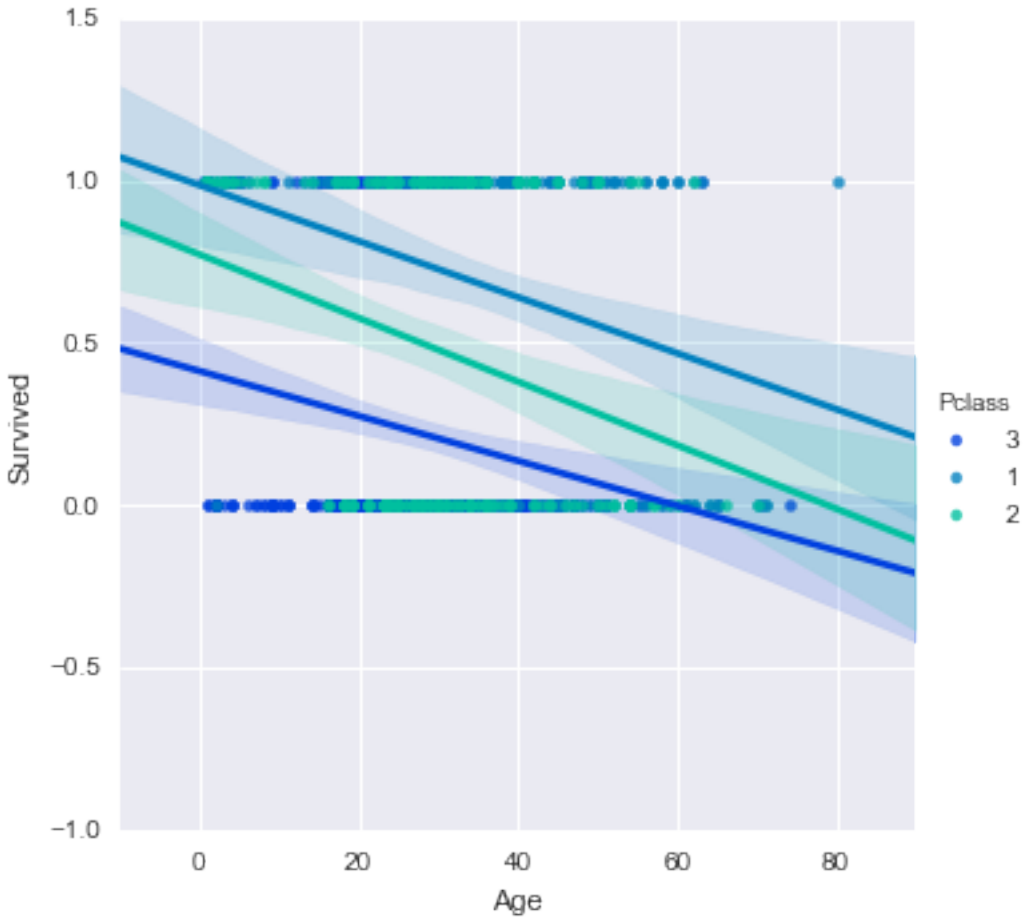
```
In [32]: sns.lmplot('Age','Survived', data=titanic_df)
```

```
Out[32]: <seaborn.axisgrid.FacetGrid at 0x111f50750>
```



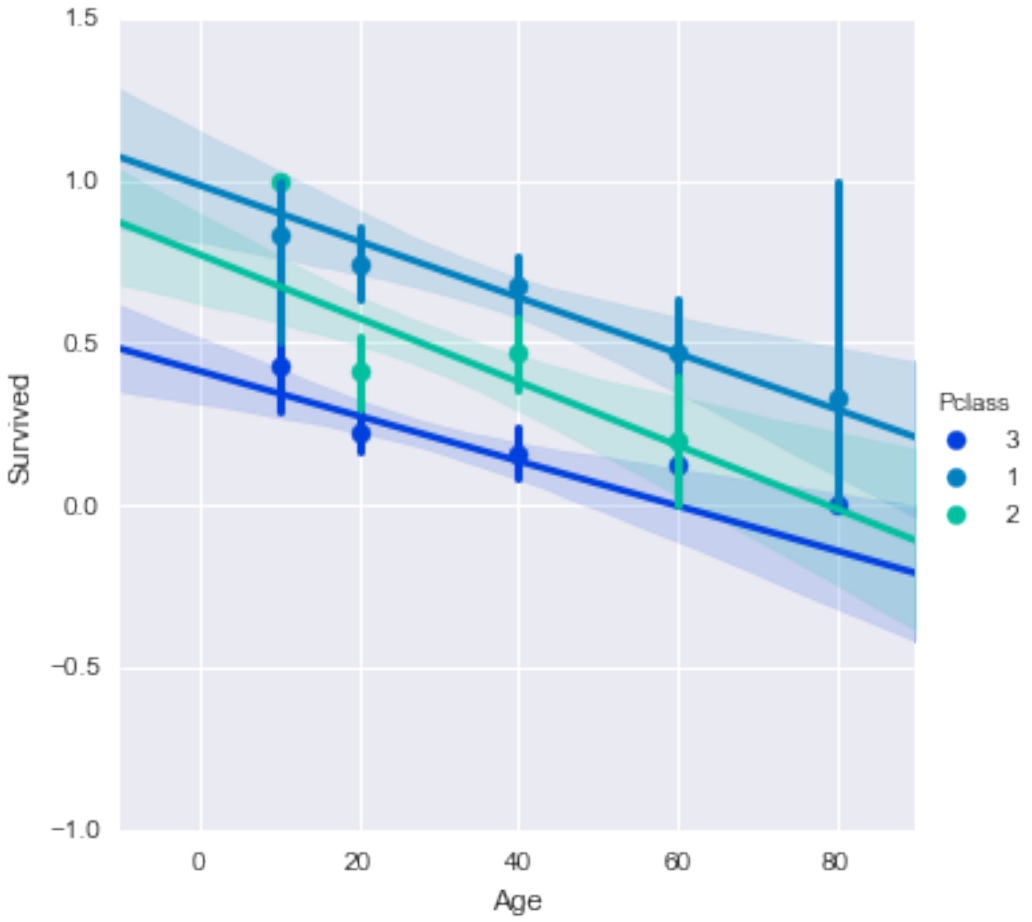

```
In [33]: sns.lmplot('Age','Survived',hue='Pclass', data=titanic_df,palette='winter')
```

```
Out[33]: <seaborn.axisgrid.FacetGrid at 0x111876290>
```



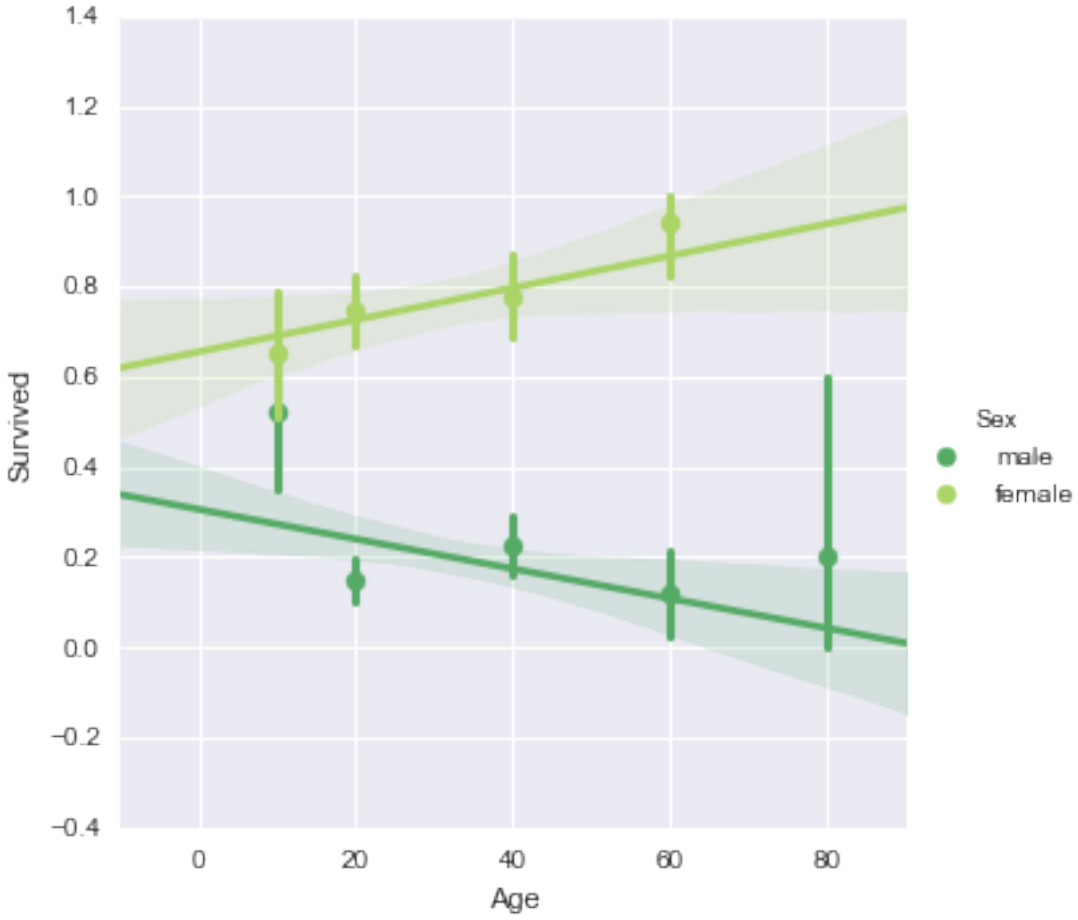

```
In [34]: generations= [10,20,40,60,80]

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

```
Out[34]: <seaborn.axisgrid.FacetGrid at 0x111e9d550>
```




```
In [35]: sns.lmplot('Age','Survived',hue='Sex',data=titanic_df,palette='summer',x_bins=generations)
```

```
Out[35]: <seaborn.axisgrid.FacetGrid at 0x1121e7650>
```



In [36]:
```
#1) Did the deck have an effect on passengers survivial rate? Did this answer match up with yo
#2) Did having a family member increase the odds of surviving the crash?
```

In [ ]:

In [ ]:

In [ ]:

In [ ]: