In [2]:
```
#General
import numpy as np
from numpy.random import randn
import pandas as pd
from pandas import Series,DataFrame

#Stats
from scipy import stats

#Plotting
import matplotlib as mpl
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

In [18]:
```
titanic=pd.read_csv('train.csv')#Loading the file in the form of dataframe
```

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

Out[19]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [20]:
```
titanic.info()#Summary of information on elements in a DF
```

```
<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
```

In [21]:
```
titanic['Name']
```

```
Out[21]: 0                                Braund, Mr. Owen Harris
         1      Cumings, Mrs. John Bradley (Florence Briggs Th...
         2                                Heikkinen, Miss. Laina
         3           Futrelle, Mrs. Jacques Heath (Lily May Peel)
         4                              Allen, Mr. William Henry
         5                                      Moran, Mr. James
         6                               McCarthy, Mr. Timothy J
         7                        Palsson, Master. Gosta Leonard
         8      Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)
         9                    Nasser, Mrs. Nicholas (Adele Achem)
         10                       Sandstrom, Miss. Marguerite Rut
         11                              Bonnell, Miss. Elizabeth
         12                        Saundercock, Mr. William Henry
         13                           Andersson, Mr. Anders Johan
         14                  Vestrom, Miss. Hulda Amanda Adolfina
         15                      Hewlett, Mrs. (Mary D Kingcome)
         16                                  Rice, Master. Eugene
         17                          Williams, Mr. Charles Eugene
         18      Vander Planke, Mrs. Julius (Emelia Maria Vande...
         19                               Masselmani, Mrs. Fatima
         20                                  Fynney, Mr. Joseph J
         21                                Beesley, Mr. Lawrence
         22                             McGowan, Miss. Anna "Annie"
         23                            Sloper, Mr. William Thompson
         24                           Palsson, Miss. Torborg Danira
         25      Asplund, Mrs. Carl Oscar (Selma Augusta Emilia...
         26                                Emir, Mr. Farred Chehab
         27                           Fortune, Mr. Charles Alexander
         28                            O'Dwyer, Miss. Ellen "Nellie"
         29                                  Todoroff, Mr. Lalio
                                             ...
         861                              Giles, Mr. Frederick Edward
         862     Swift, Mrs. Frederick Joel (Margaret Welles Ba...
         863                     Sage, Miss. Dorothy Edith "Dolly"
         864                               Gill, Mr. John William
         865                              Bystrom, Mrs. (Karolina)
         866                          Duran y More, Miss. Asuncion
         867                  Roebling, Mr. Washington Augustus II
         868                           van Melkebeke, Mr. Philemon
         869                       Johnson, Master. Harold Theodor
         870                                    Balkic, Mr. Cerin
         871      Beckwith, Mrs. Richard Leonard (Sallie Monypeny)
         872                               Carlsson, Mr. Frans Olof
         873                           Vander Cruyssen, Mr. Victor
         874                 Abelson, Mrs. Samuel (Hannah Wizosky)
         875                      Najib, Miss. Adele Kiamie "Jane"
         876                         Gustafsson, Mr. Alfred Ossian
         877                                  Petroff, Mr. Nedelio
         878                                    Laleff, Mr. Kristo
         879         Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)
         880          Shelley, Mrs. William (Imanita Parrish Hall)
         881                                    Markun, Mr. Johann
         882                           Dahlberg, Miss. Gerda Ulrika
         883                          Banfield, Mr. Frederick James
         884                                 Sutehall, Mr. Henry Jr
         885                  Rice, Mrs. William (Margaret Norton)
         886                                  Montvila, Rev. Juozas
         887                           Graham, Miss. Margaret Edith
         888              Johnston, Miss. Catherine Helen "Carrie"
         889                                  Behr, Mr. Karl Howell
         890                                    Dooley, Mr. Patrick
         Name: Name, Length: 891, dtype: object
```

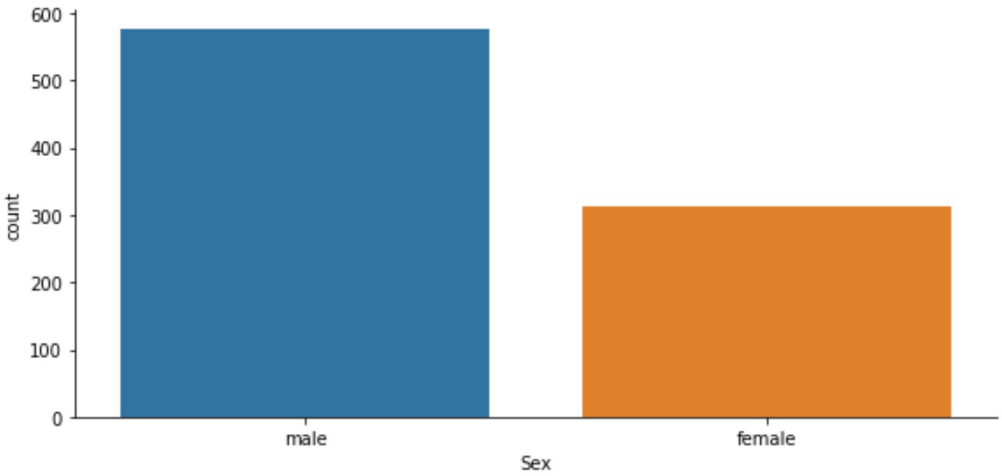
In [22]:
```
titanic.head()
```

Out[22]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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

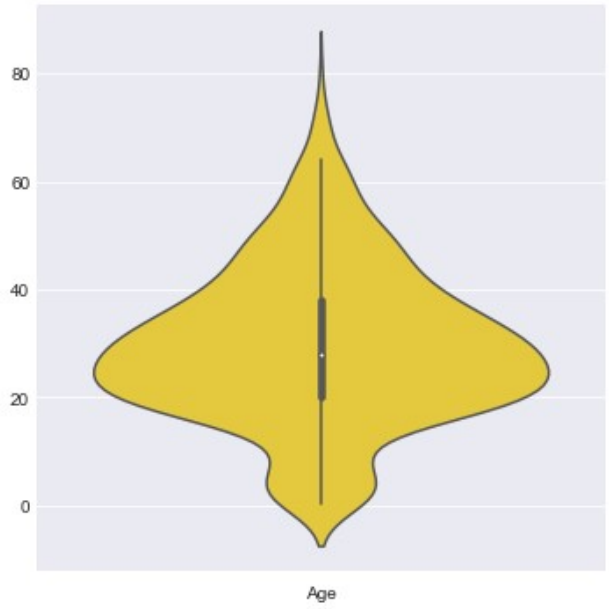
In [66]:
```
sns.factorplot('Sex',data=titanic,kind="count",margin_titles=True,aspect=2)
#kind : {``point``, ``bar``, ``count``, ``box``, ``violin``, ``strip``}
```

Out[66]: <seaborn.axisgrid.FacetGrid at 0x1f3779419b0>



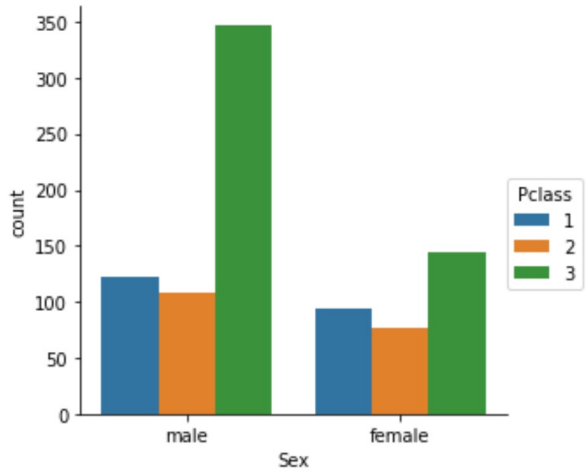
In [199]:
```
sns.set_style('darkgrid')
sns.set_palette('spring_r')
sns.factorplot(x='Age',data=titanic,kind="violin",orient='v',size=5)
#style must be one of white, dark, whitegrid, darkgrid, ticks
```

Out[199]: <seaborn.axisgrid.FacetGrid at 0x1f37bf32cf8>



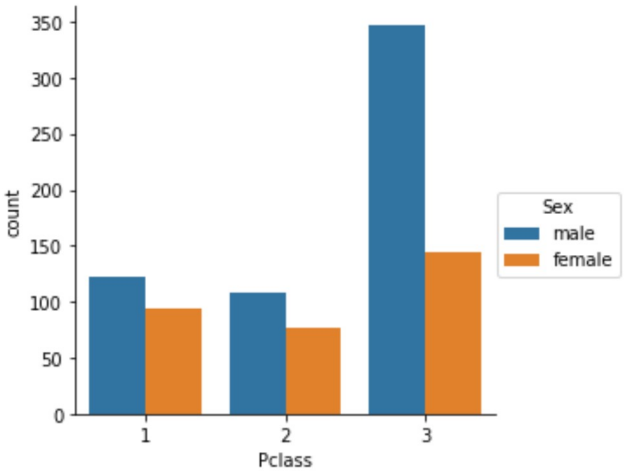
In [54]:
```
sns.factorplot('Sex',data=titanic,hue='Pclass',kind="count",)#hue is for splittign the data by
```

Out[54]: <seaborn.axisgrid.FacetGrid at 0x1f376505eb8>



In [69]:
```
sns.factorplot('Pclass',data=titanic,kind="count",hue='Sex')
```

Out[69]:
```
<seaborn.axisgrid.FacetGrid at 0x1f3765d5710>
```



In [76]:
```
def person(passng):
    Age,Sex=passng

    if Age < 16:
        return 'Child'
    else:
        return Sex
#Creating a function to calculate/ decide whether a person is a male/female/childre
n
```

In [77]:
```
titanic['Person']=titanic[['Age','Sex']].apply(person,axis=1)
#since the structure of entire DataFrame is Index:Column
#and we are making a change in the columns, axis=1 need to be specified.
```

In [78]:
```
titanic.head(10)
```

Out[78]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C

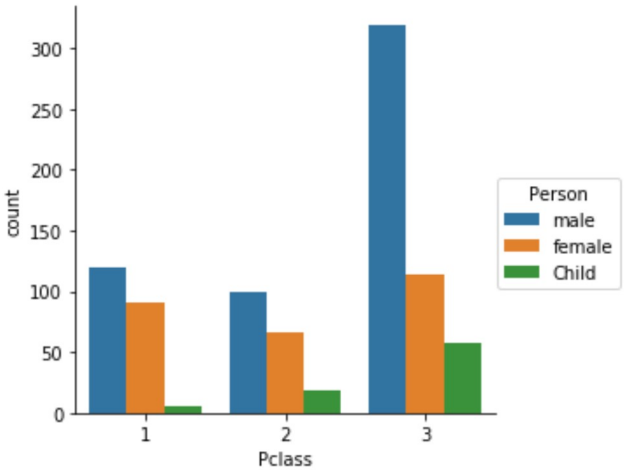
In [84]:
```
titanic[['Fare','Pclass']].tail(10)
```

Out[84]:

	Fare	Pclass
881	7.8958	3
882	10.5167	3
883	10.5000	2
884	7.0500	3
885	29.1250	3
886	13.0000	2
887	30.0000	1
888	23.4500	3
889	30.0000	1
890	7.7500	3

In [85]:
```
sns.factorplot('Pclass',kind="count",data=titanic,hue='Person')
```

Out[85]: `<seaborn.axisgrid.FacetGrid at 0x1f377b2be48>`



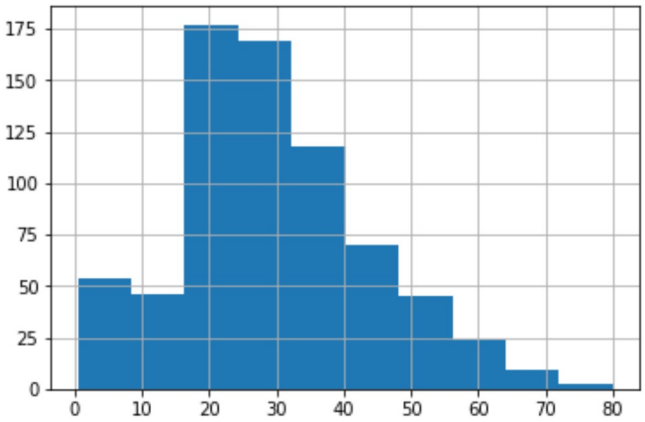
In [86]:
```
titanic.tail(10)
```

Out[86]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabi
881	882	0	3	Markun, Mr. Johann	male	33.0	0	0	349257	7.8958	NaN
882	883	0	3	Dahlberg, Miss. Gerda Ulrika	female	22.0	0	0	7552	10.5167	NaN
883	884	0	2	Banfield, Mr. Frederick James	male	28.0	0	0	C.A./SOTON 34068	10.5000	NaN
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN

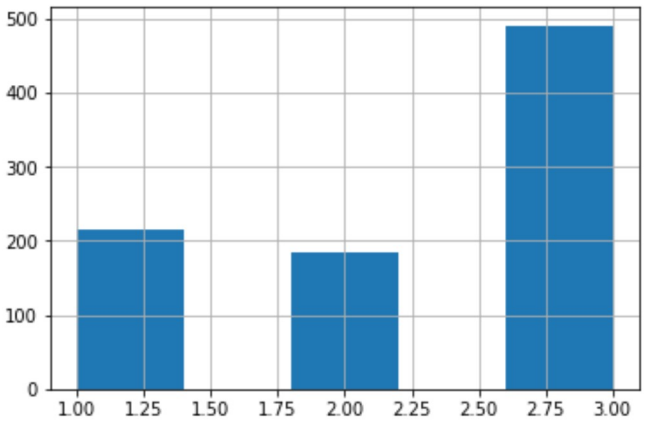
In [99]:
```
titanic['Age'].hist(bins=10)
```

Out[99]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1f379046828>
```



In [102]:
```
titanic['Pclass'].hist(bins=5)
```

Out[102]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1f3791bcda0>
```



In [103]:
```
titanic['Person'].value_counts()
```

Out[103]:
```
male      537
female    271
Child      83
Name: Person, dtype: int64
```

In [104]:
```
titanic['Sex'].value_counts()
```

Out[104]:
```
male      577
female    314
Name: Sex, dtype: int64
```

In [109]:
```
titanic['Pclass'].value_counts()
```

Out[109]:
```
3    491
1    216
2    184
Name: Pclass, dtype: int64
```

In [111]:
```
titanic.head()
```

Out[111]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [112]:
```
titanic['Deck']=titanic['Cabin'][0]
```

In [114]:
```
titanic.head()
```

Out[114]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [135]:
```
def deck(Cabin):
    Cabin=Cabin

    for letter in str(Cabin):
        return letter[0]
```

In [144]:
```
titanic_new=titanic[titanic.Deck!='n']#Eliminating NaN Records
titanic_new['Deck']=titanic_new['Cabin'].apply(deck)
```

```
C:\Users\acpimpar\Anaconda3\lib\site-packages\ipykernel_launcher.py:2: SettingWi
thCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stabl
e/indexing.html#indexing-view-versus-copy
```

In [145]:
```
titanic_new.head()
```

Out[145]:

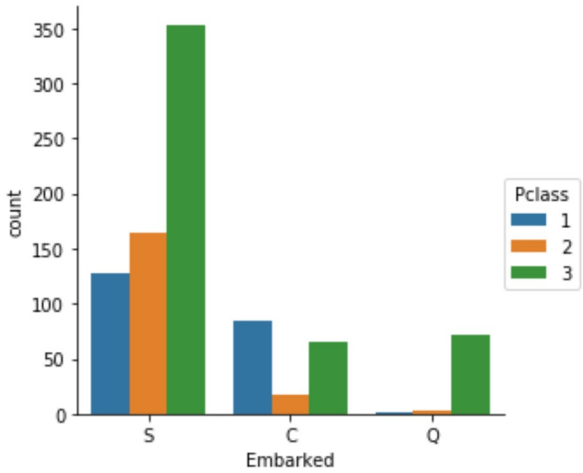
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	En
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S

In [146]:
```
titanic_new.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 204 entries, 1 to 889
Data columns (total 14 columns):
PassengerId    204 non-null int64
Survived       204 non-null int64
Pclass         204 non-null int64
Name           204 non-null object
Sex            204 non-null object
Age            185 non-null float64
SibSp          204 non-null int64
Parch          204 non-null int64
Ticket         204 non-null object
Fare           204 non-null float64
Cabin          204 non-null object
Embarked       202 non-null object
Person         204 non-null object
Deck           204 non-null object
dtypes: float64(2), int64(5), object(7)
memory usage: 23.9+ KB
```

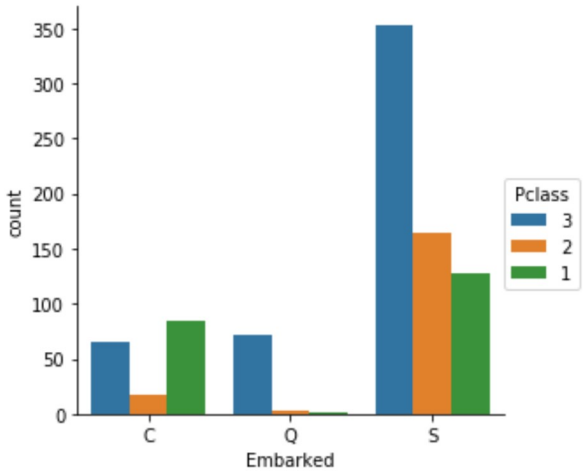
In [148]:
```
sns.factorplot('Embarked',data=titanic,hue='Pclass',kind="count")
```

Out[148]: <seaborn.axisgrid.FacetGrid at 0x1f379145198>



In [156]:
```
sns.factorplot('Embarked',data=titanic,hue='Pclass',kind="count",order=['C','Q','S
']
               ,hue_order=[3,2,1])#orderging the appearance and splitting by hue o
rder
```

Out[156]: <seaborn.axisgrid.FacetGrid at 0x1f377869dd8>



In [157]:
```
titanic.head()
```

Out[157]:

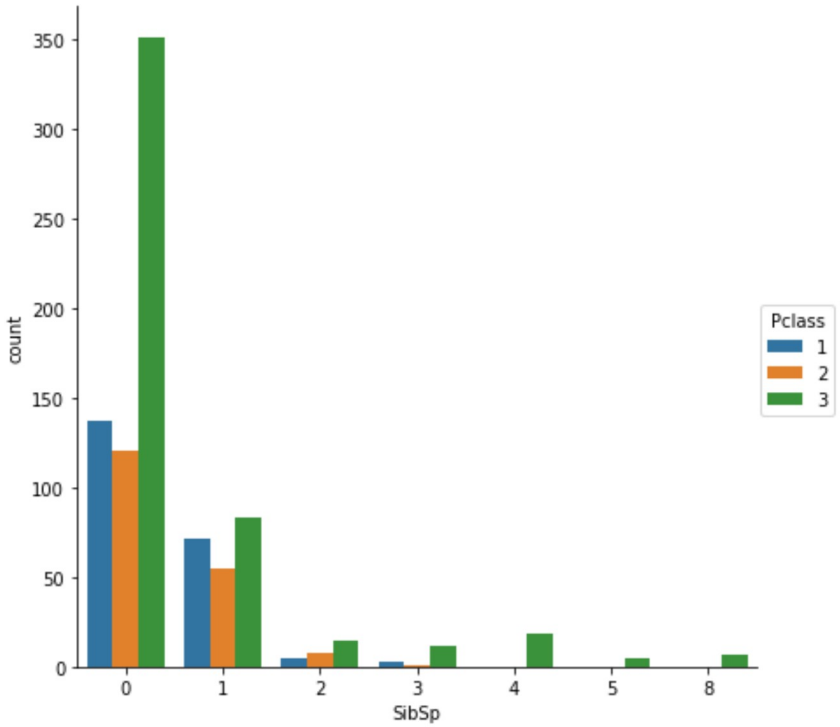
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	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 [159]:
```
titanic['SibSp'].value_counts()
```

Out[159]:
```
0    608
1    209
2     28
4     18
3     16
8      7
5      5
Name: SibSp, dtype: int64
```

In [164]:
```
sns.factorplot('SibSp',data=titanic,kind="count",hue='Pclass',size=6)
```

Out[164]: <seaborn.axisgrid.FacetGrid at 0x1f3793e01d0>



In [165]:
```
titanic.head()
```

Out[165]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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 [166]:
```
def alone(alone):
    Sibsp,Parch=alone
    if Sibsp==0 and Parch==0:
        return 'Y'
    else:
        return 'N'
```

In [167]:
```
titanic['Alone?']= titanic[['SibSp','Parch']].apply(alone,axis=1)
#Creating a New column by passing two columns as a input to a newly created functi
on above.
```

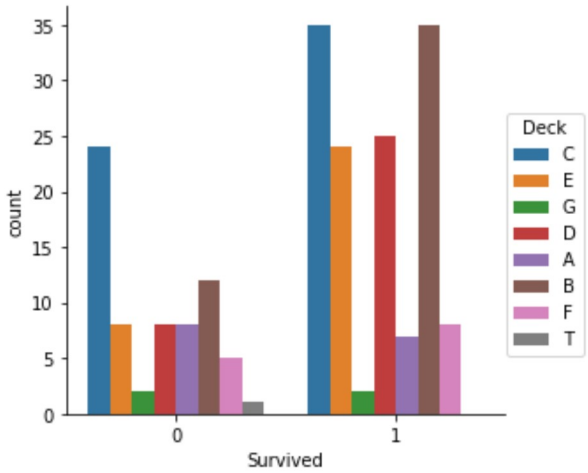
In [169]:
```
titanic.head()
```

Out[169]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	E
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

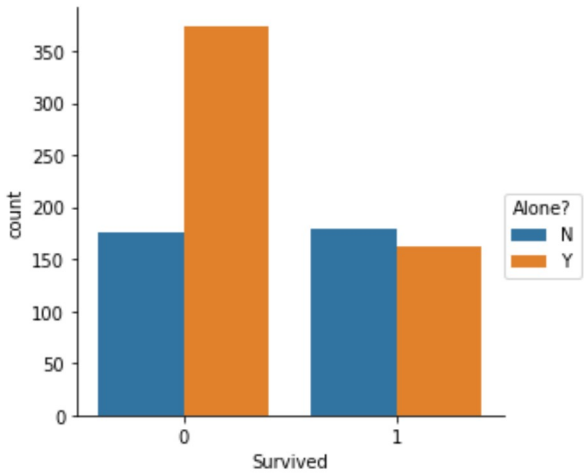
In [175]:
```
sns.factorplot('Survived',data=titanic_new,kind="count",hue='Deck')
```

Out[175]: <seaborn.axisgrid.FacetGrid at 0x1f379526b38>



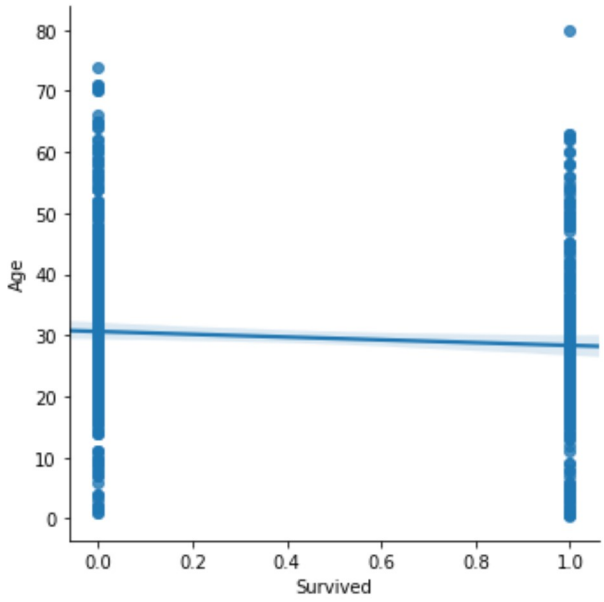
In [177]:
```
sns.factorplot('Survived',data=titanic,kind="count",hue='Alone?')
```

Out[177]: <seaborn.axisgrid.FacetGrid at 0x1f379749e10>



In [182]:
```
sns.lmplot('Survived','Age',titanic)
```

Out[182]:
```
<seaborn.axisgrid.FacetGrid at 0x1f37aac4fd0>
```



In [185]:
```
from IPython.display import Image
Image(url='http://i.imgur.com/DGNjT.gif')
```

Out[185]:

In [186]:
```
from IPython.display import Image
Image(url='http://i.imgur.com/EjgeEqn.jpg')
```

Out[186]: