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	22.0	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 [3]:
```
titanic_df.info()
```

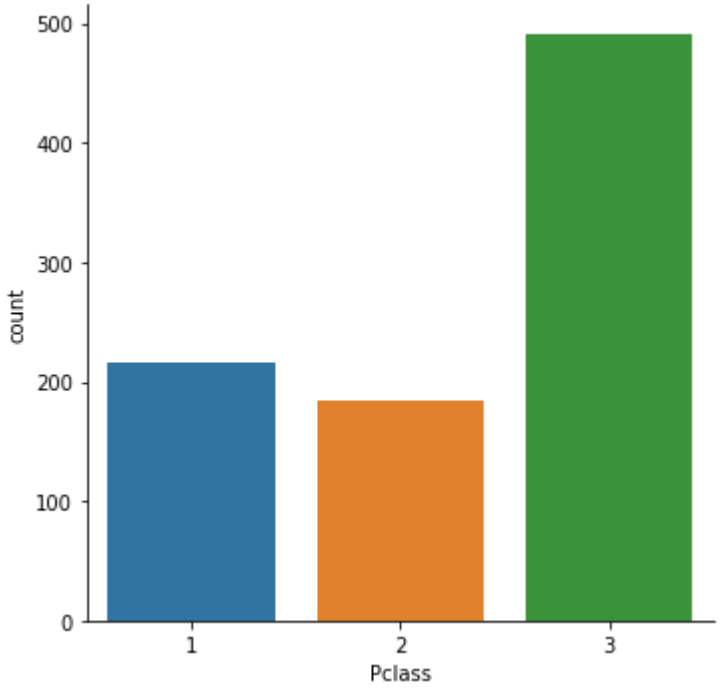
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
PassengerId    891 non-null int64
Survived       891 non-null int64
Pclass         891 non-null int64
Name           891 non-null object
Sex            891 non-null object
Age            714 non-null 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: 83.7+ KB
```

In [4]:
```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

In [5]:
```
# Find who were passengers like sex, class etc

# sns.factorplot('Sex', data=titanic_df, hue='Pclass') # giving type rror

sns.factorplot('Pclass', data=titanic_df, kind='count')
```

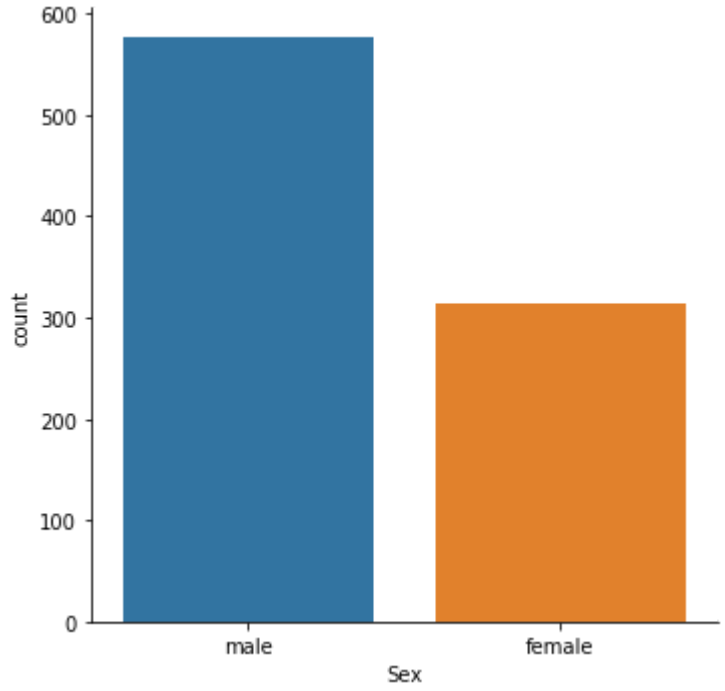
```
C:\Users\Lyci\Anaconda3\lib\site-packages\seaborn\categorical.py:3666: UserWarning: The `factorplot` function
has been renamed to `catplot`. The original name will be removed in a future release. Please update your code.
Note that the default `kind` in `factorplot` (`'point'`) has changed `'strip'` in `catplot`.
  warnings.warn(msg)
```

Out[5]: <seaborn.axisgrid.FacetGrid at 0x1291f5067b8>



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

Out[6]: <seaborn.axisgrid.FacetGrid at 0x1291d05f828>



In [7]:
```
# Find number of children
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.head() # see new column person is added to table
```

Out[9]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	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 [10]:
```
titanic_df[0:20] # can see top 20 rows with male, female and child entries in person
```

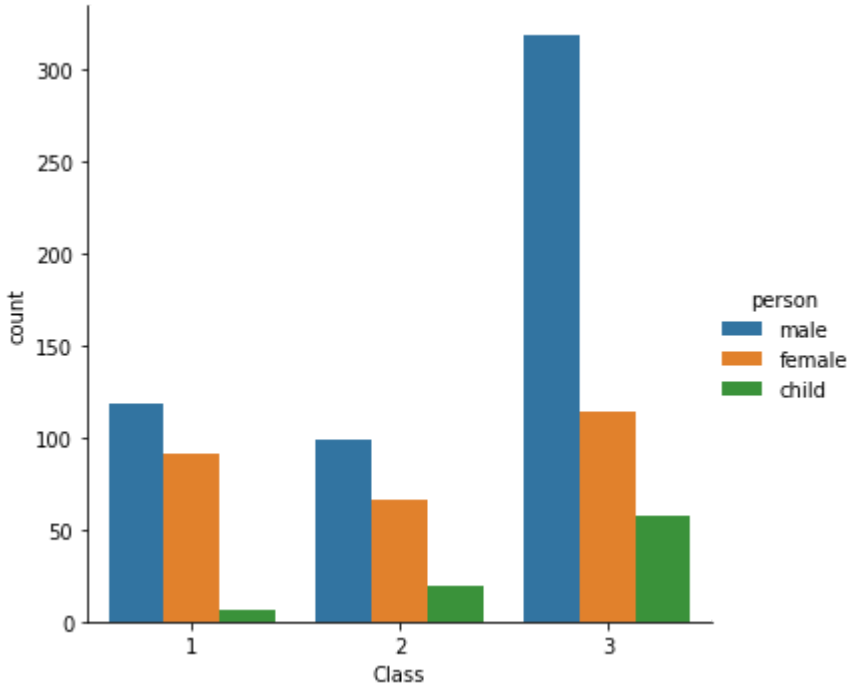
Out[10]:

	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
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	0	17463	51.8625	E46	S	male
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S	child
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S	female
9	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14.0	1	0	237736	30.0708	NaN	C	child
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S	child
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S	female
12	13	0	3	Saundercock, Mr. William Henry	male	20.0	0	0	A/5. 2151	8.0500	NaN	S	male
13	14	0	3	Andersson, Mr. Anders Johan	male	39.0	1	5	347082	31.2750	NaN	S	male
14	15	0	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14.0	0	0	350406	7.8542	NaN	S	child
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0000	NaN	S	female

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person
16	17	0	3	Rice, Master. Eugene	male	2.0	4	1	382652	29.1250	NaN	Q	child
17	18	1	2	Williams, Mr. Charles Eugene	male	NaN	0	0	244373	13.0000	NaN	S	male
18	19	0	3	Vander Planke, Mrs. Julius (Emelia Maria Vande...	female	31.0	1	0	345763	18.0000	NaN	S	female
19	20	1	3	Masselmani, Mrs. Fatima	female	NaN	0	0	2649	7.2250	NaN	C	female

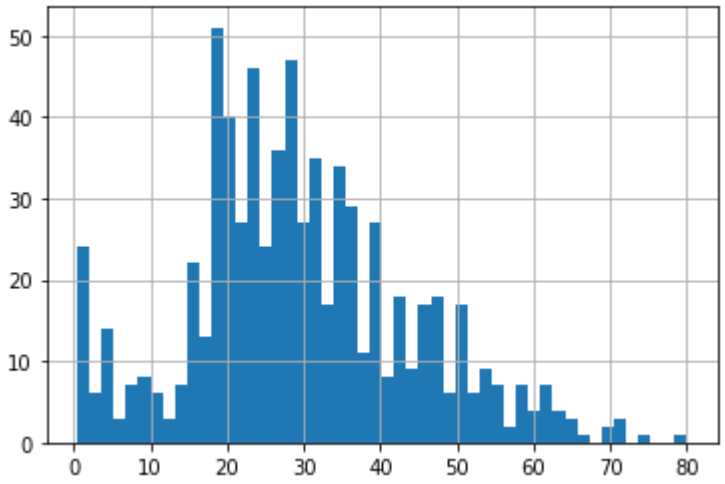
In [11]:
```
graph = sns.factorplot('Pclass', data=titanic_df, hue='person', kind='count')#, aspect=1.75)
graph.set_xlabels('Class')
```

Out[11]:
```
<seaborn.axisgrid.FacetGrid at 0x1291f4ee860>
```



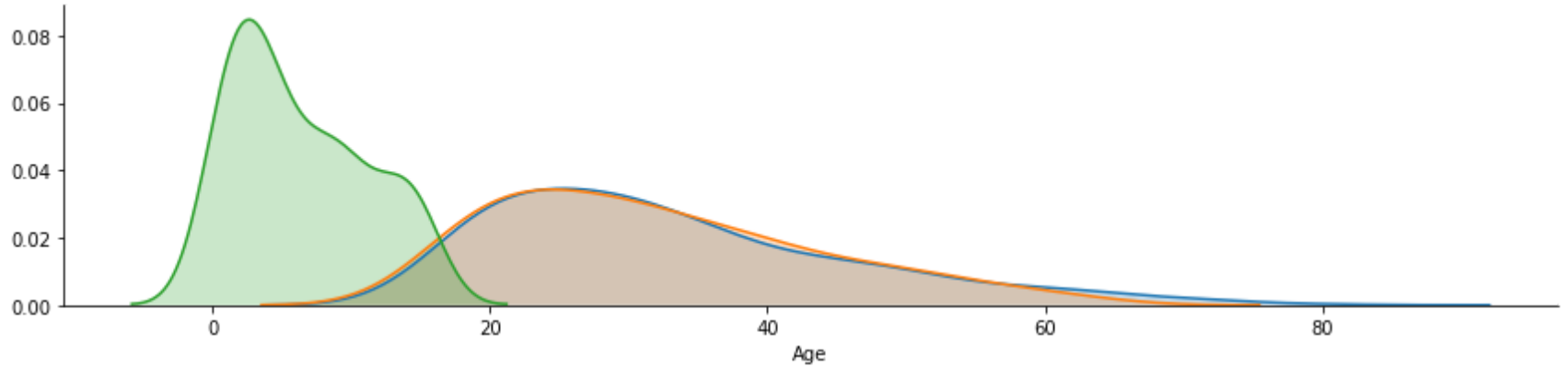
In [12]:
```
titanic_df['Age'].hist(bins=50)
titanic_df['person'].value_counts() # gives number of each category
```

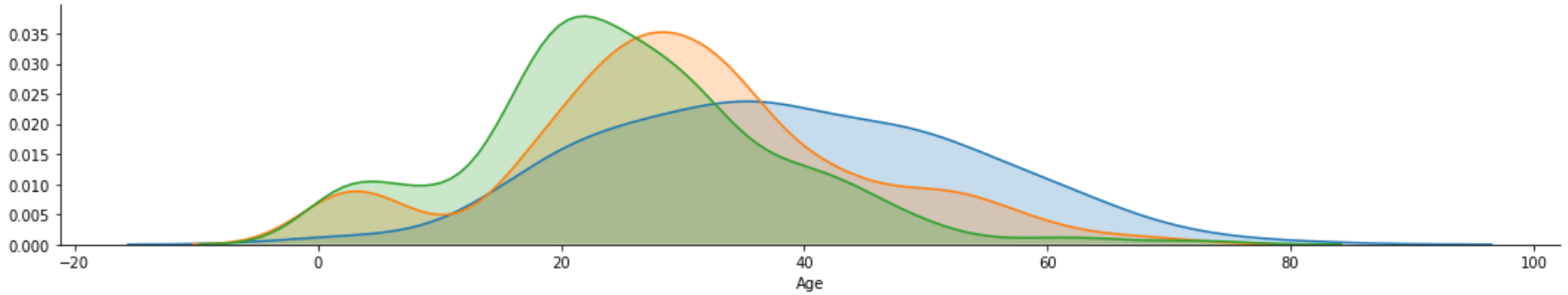
Out[12]:
```
male      537
female    271
child      83
Name: person, dtype: int64
```

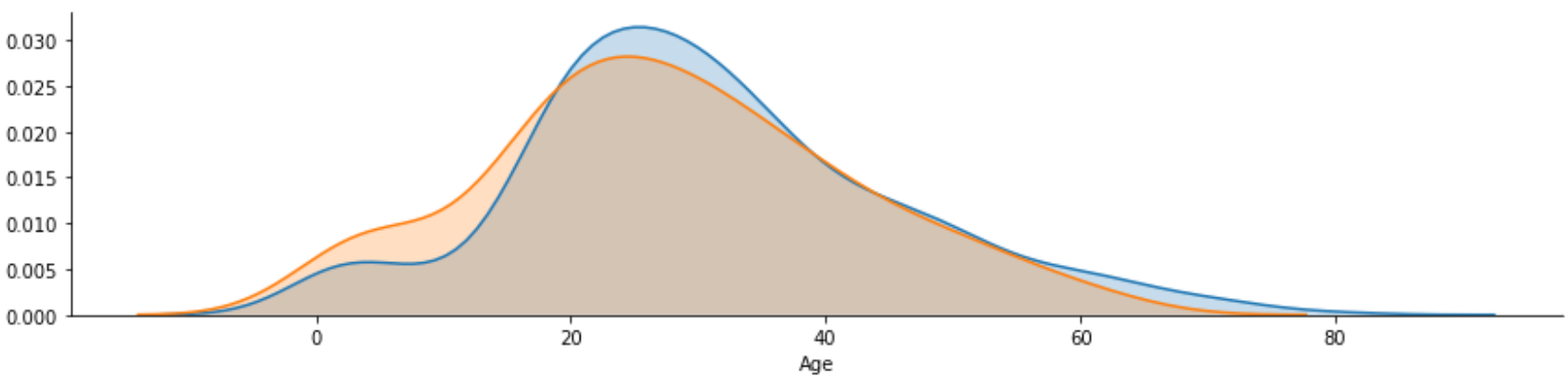


In [13]:
```
# using facetgrid for multiple plots
fig=sns.FacetGrid(titanic_df, hue='person',aspect=4)
fig.map(sns.kdeplot, 'Age', shade=True)
fig=sns.FacetGrid(titanic_df, hue='Pclass',aspect=5)
fig.map(sns.kdeplot, 'Age', shade=True)
fig=sns.FacetGrid(titanic_df, hue='Sex',aspect=4)
fig.map(sns.kdeplot, 'Age', shade=True)
```

Out[13]: <seaborn.axisgrid.FacetGrid at 0x1291f948a58>







In [14]:
```
oldest=titanic_df['Age'].max()
oldest
```

Out[14]:
```
80.0
```

In [15]:
```
fig.set(xlim=(0,oldest))
```

Out[15]:
```
<seaborn.axisgrid.FacetGrid at 0x1291f948a58>
```

In [16]:
```
fig.add_legend()
```

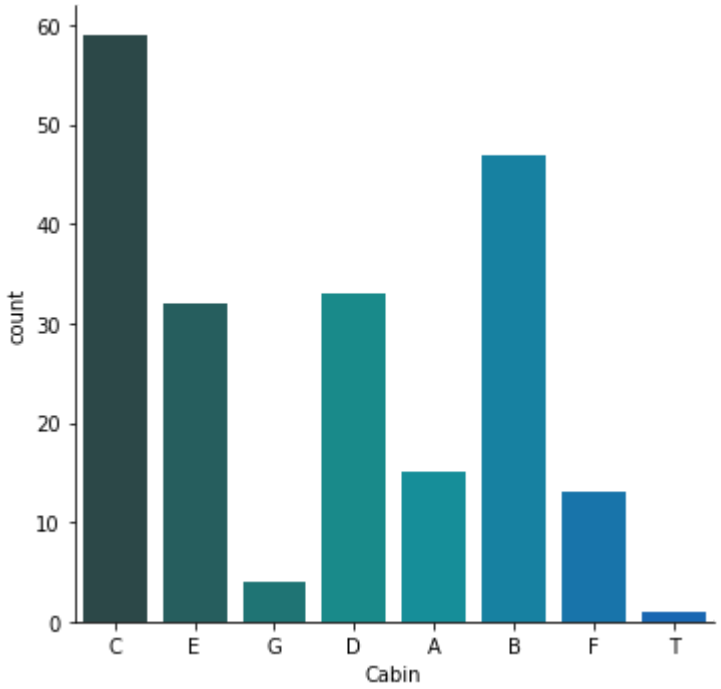
Out[16]:
```
<seaborn.axisgrid.FacetGrid at 0x1291f948a58>
```

In [17]:
```
#exclude cabin column due to many null values
deck=titanic_df['Cabin'].dropna()
deck.head() #cabin data not null
```

Out[17]:
```
1      C85
3     C123
6      E46
10      G6
11    C103
Name: Cabin, dtype: object
```

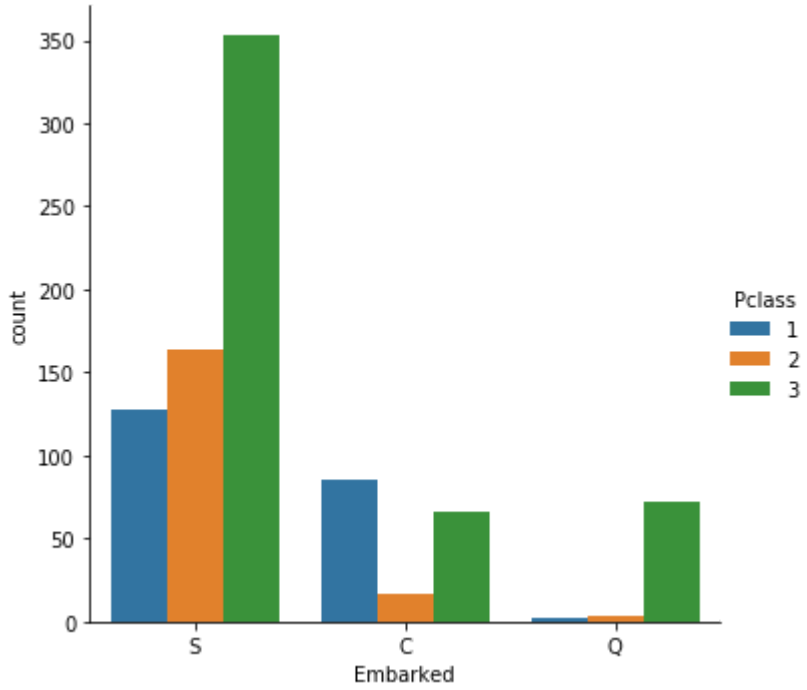
In [18]:
```
# find levels from cabin numbers by omitting last numbers and selecting only first alphabet
levels=[]
for level in deck:
    levels.append(level[0])
cabin_df=DataFrame(levels)
cabin_df.columns=['Cabin']
sns.factorplot('Cabin', data=cabin_df, kind='count', palette='winter_d') # palete color can be selected using ma
```

Out[18]: <seaborn.axisgrid.FacetGrid at 0x1291fc182e8>



In [19]:
```
# Find city of Origin/embarkation
sns.factorplot('Embarked', data=titanic_df, kind= 'count', hue='Pclass')#, x_order=['C','Q','S'])
```

Out[19]: <seaborn.axisgrid.FacetGrid at 0x1291fb3fd68>

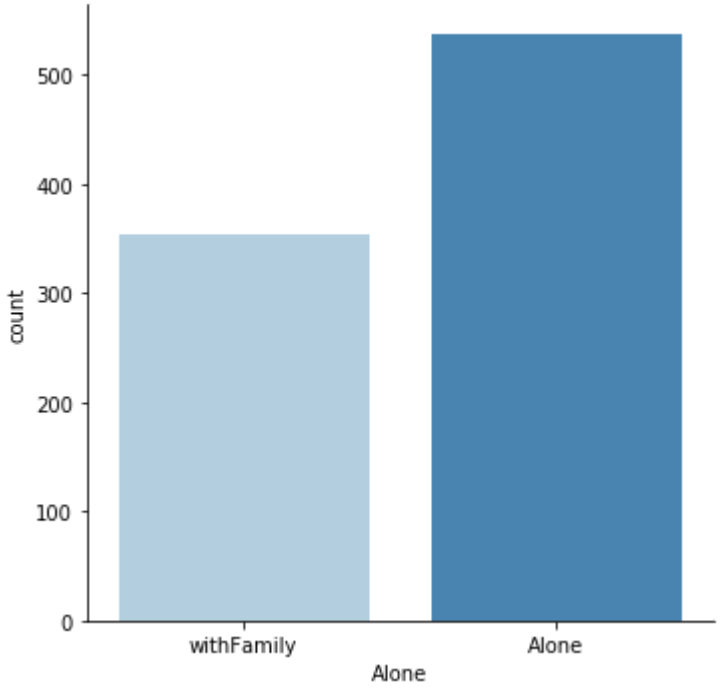


In [20]:
```
#find who was alone 0 values in Sibilings/spouse-SibSp, parch-parent/child
titanic_df['Alone']=titanic_df.SibSp+titanic_df.Parch
titanic_df['Alone'].loc[titanic_df['Alone']>0]='withFamily'
titanic_df['Alone'].loc[titanic_df['Alone']==0]='Alone'
```

```
C:\Users\Lyci\Anaconda3\lib\site-packages\pandas\core\indexing.py:205: 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/user_guide/indexing.html#returning-a-view-versus-a-copy (http://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)
  self._setitem_with_indexer(indexer, value)
```

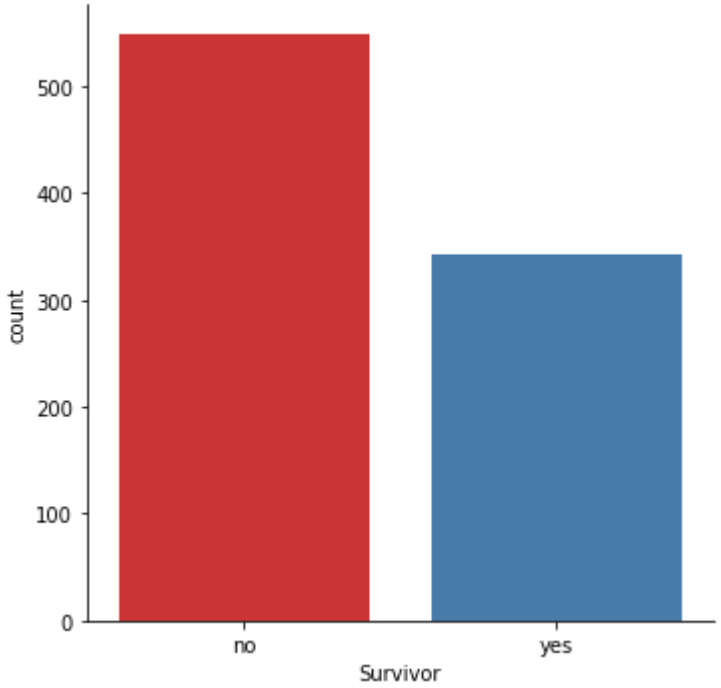
In [21]:
```
sns.factorplot('Alone', data=titanic_df, kind='count', palette='Blues')
```

Out[21]:
```
<seaborn.axisgrid.FacetGrid at 0x1291fad5d68>
```



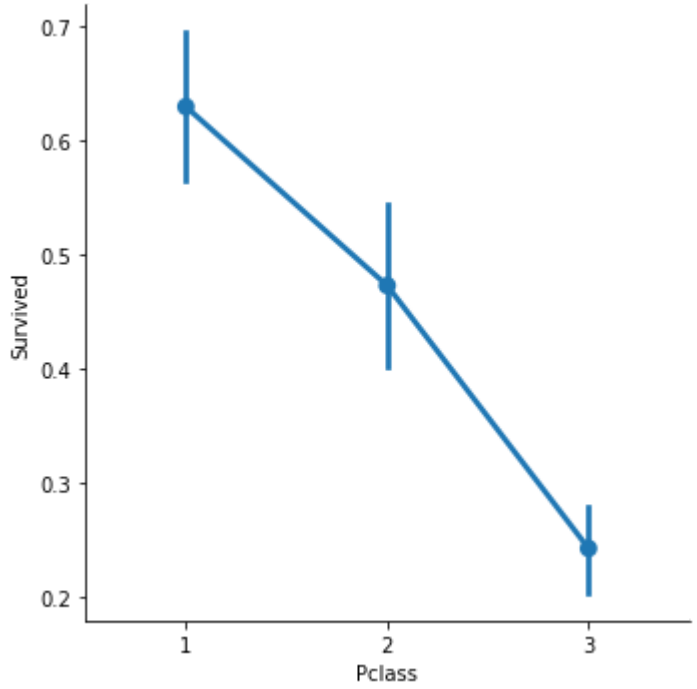
In [22]:
```
#finding survivors

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

Out[22]: <seaborn.axisgrid.FacetGrid at 0x1291fc8a240>



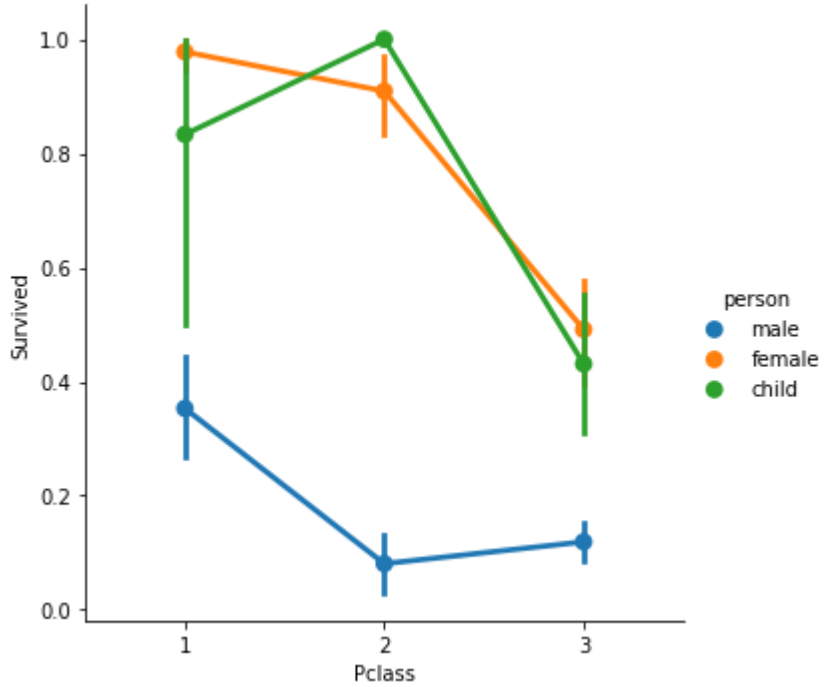
In [23]:
```
# we may predict men survival rat was low as they were rescued after children and females. Check this
sns.factorplot('Pclass', 'Survived', data=titanic_df)
```

Out[23]: <seaborn.axisgrid.FacetGrid at 0x1291fba6a90>



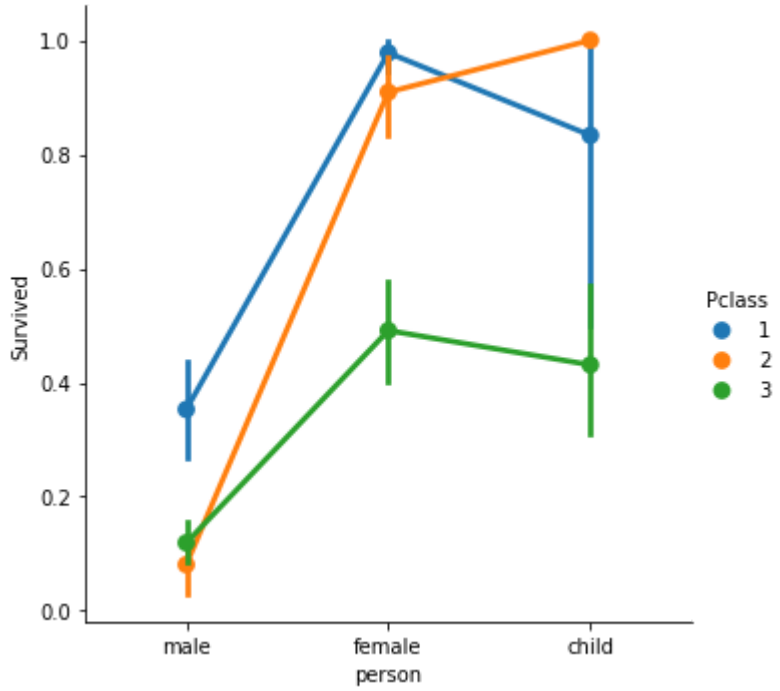
In [24]:
```
#3rd class had more men and less survival. Check for gender using hue

sns.factorplot('Pclass', 'Survived', hue='person', data=titanic_df)
```

Out[24]: <seaborn.axisgrid.FacetGrid at 0x1291fa16e80>



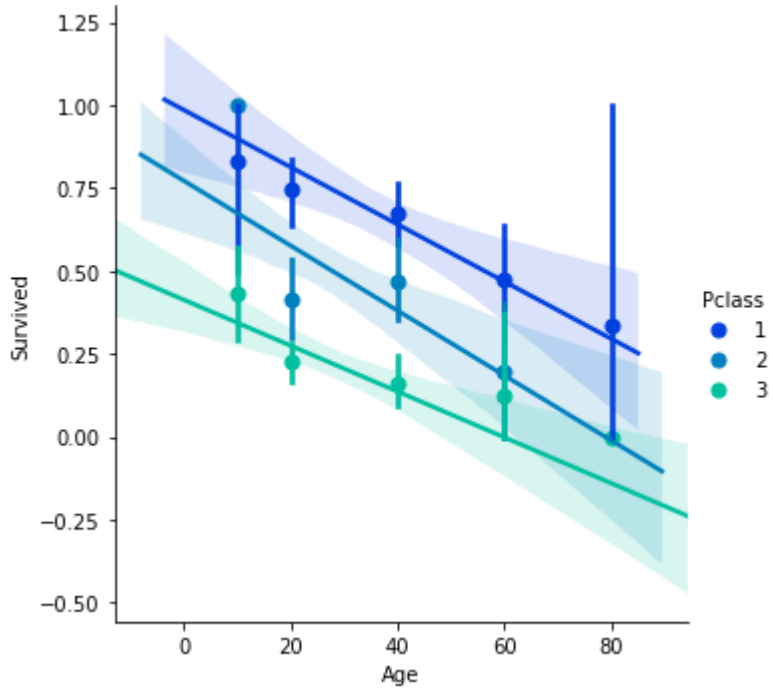
In [25]:
```
sns.factorplot('person', 'Survived', hue='Pclass', data=titanic_df)
```

Out[25]:
```
<seaborn.axisgrid.FacetGrid at 0x1291fa5be80>
```



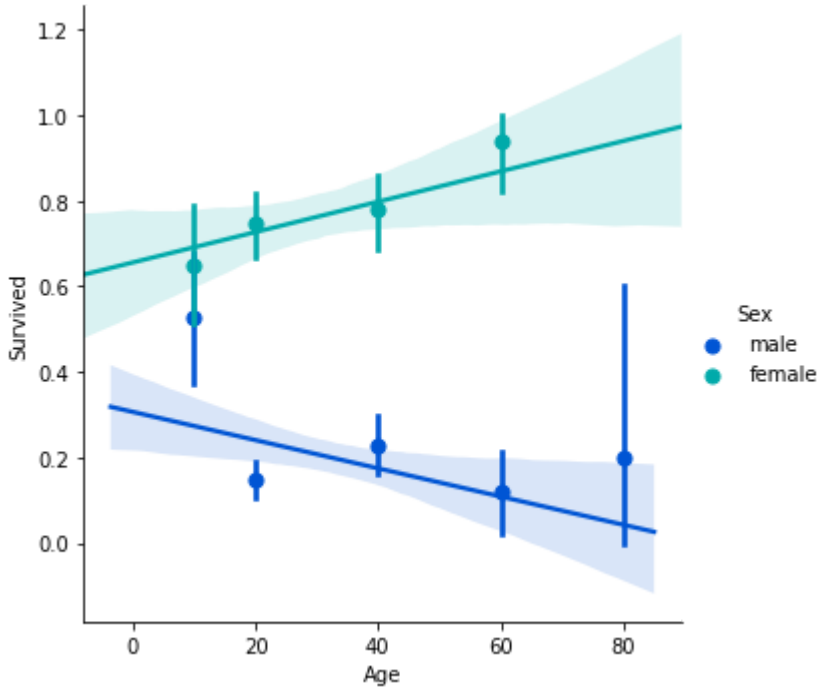
In [26]:
```
generations=[10,20,40,60,80]
sns.lmplot('Age', 'Survived', hue='Pclass', data=titanic_df, palette='winter', x_bins=generations)
```

Out[26]: <seaborn.axisgrid.FacetGrid at 0x12920e150b8>



In [27]:
```
sns.lmplot('Age', 'Survived', hue='Sex', data=titanic_df, palette='winter', x_bins=generations)
```

Out[27]:
```
<seaborn.axisgrid.FacetGrid at 0x12920ec30b8>
```



In [33]:
```
cabin_df
```

Out[33]:

	Cabin
0	C
1	C
2	E
3	G
4	C
...	...
199	D
200	B
201	C
202	B
203	C

204 rows × 1 columns

In [35]:
```
deck=titanic_df.dropna(subset=['Cabin'])
deck_df=DataFrame(deck) #create new dataframe
```

In [36]:
```
deck_df
```

Out[36]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person	Alone	Surv
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	female	withFamily	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	female	withFamily	
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S	male	Alone	
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7000	G6	S	child	withFamily	
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5500	C103	S	female	Alone	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.5542	D35	S	female	withFamily	
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.0000	B51 B53 B55	S	male	Alone	
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C	female	withFamily	

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person	Alone	Surv
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S	female	Alone	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C	male	Alone	

204 rows × 15 columns

In [45]:
```
deckLevel=titanic_df['Cabin'].dropna()

levels=[]
for level in deckLevel:
    levels.append(level[0])
Level_df=DataFrame(levels)
Level_df.columns=['deckLevel']
Level_df
```

Out[45]:

	deckLevel
0	C
1	C
2	E
3	G
4	C
...	...
199	D
200	B
201	C
202	B
203	C

204 rows × 1 columns

In [48]:
```
merged_df =deck_df.merge(Level_df, left_index=True, right_index=True) # mereges 2 data frames
merged_df # give error as some Cabin rowscontain multiple entries
```

166	167	1	1	Chibnall, Mrs. (Edith Martha Bowerman)	female	NaN	0	1	113505	55.0000	E33	S	female	withFamily
170	171	0	1	Van der hoef, Mr. Wyckoff	male	61.0	0	0	111240	33.5000	B19	S	male	Alone
174	175	0	1	Smith, Mr. James Clinch	male	56.0	0	0	17764	30.6958	A7	C	male	Alone
177	178	0	1	Isham, Miss. Ann Elizabeth	female	50.0	0	0	PC 17595	28.7125	C49	C	female	Alone
183	184	1	2	Becker, Master. Richard F	male	1.0	2	1	230136	39.0000	F4	S	child	withFamily

In [41]:
```
```

In [39]:
```
titanic_df.head()
```

Out[39]:

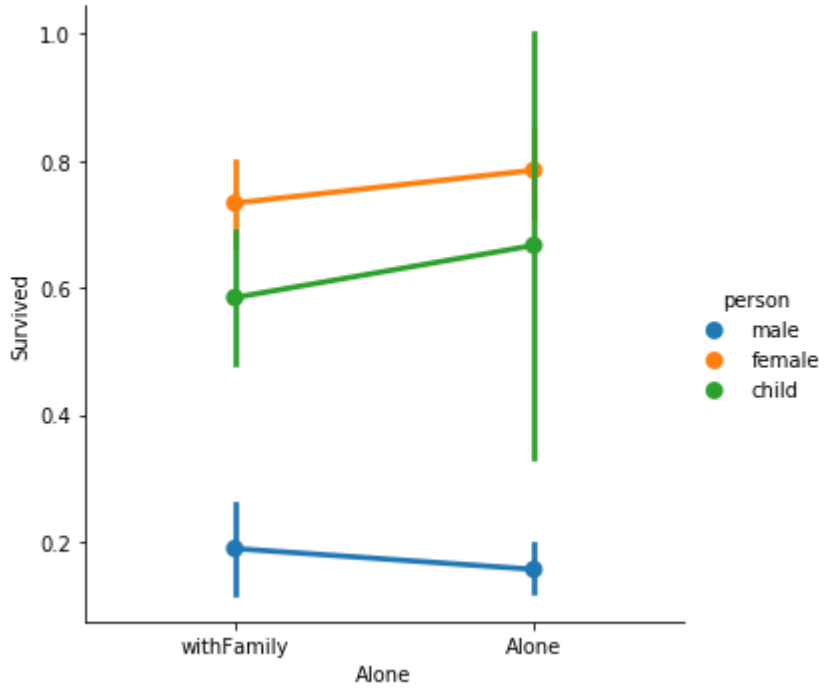
	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	person	Alone	Surv
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	male	withFamily	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	female	withFamily	
2	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	0	113803	53.1000	C123	S	female	withFamily	
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S	male	Alone	

In [131]:
```
#merged_df =titanic_df.merge(cabin_df, left_index=True, right_index=True, how='inner')
#merged_df.head()
```

In [ ]:
```
#sns.lmplot('DeckLevel', 'Survived', hue='Pclass', data=titanic_df, palette='winter', x_bins=generations)
```

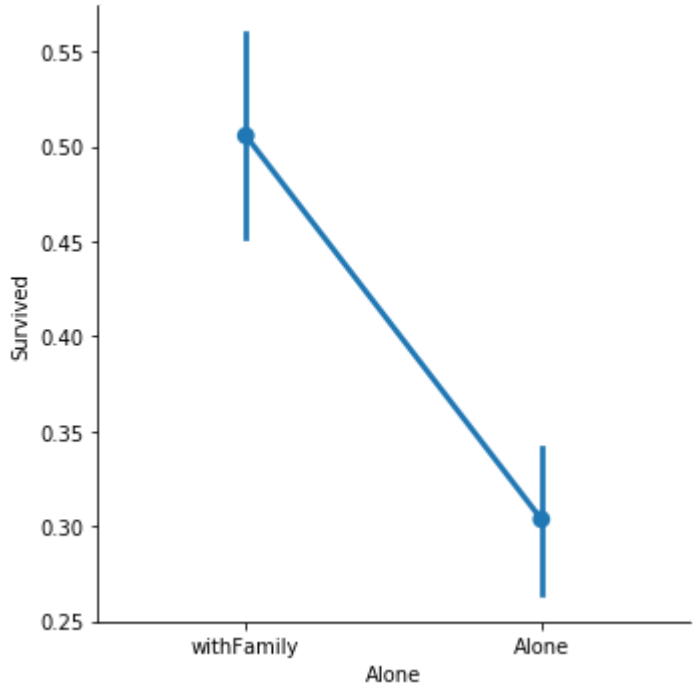
In [49]:
```
sns.factorplot('Alone', 'Survived', hue='person', data=titanic_df)
```

Out[49]: <seaborn.axisgrid.FacetGrid at 0x1291fd06978>



In [50]:
```
sns.factorplot('Alone', 'Survived', data=titanic_df)
```

Out[50]: <seaborn.axisgrid.FacetGrid at 0x12920ec3e10>



In [ ]:
```

```