

Titanic Intro Project

September 2, 2018

```
In [6]: import pandas as pd
        from pandas import Series, DataFrame
```

```
In [7]: titanic_df = pd.read_csv('train.csv')
```

```
In [8]: titanic_df.head()
```

```
Out[8]:
```

| | PassengerId | Survived | Pclass | \ | | Name | Sex | Age | SibSp | \ |
|---|-------------|----------|--------|---|--|------|-----|-----|-------|---|
| 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 |
|---|-------|------------------|---------|-------|----------|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S |
| 1 | 0 | PC 17599 | 71.2833 | C85 | C |
| 2 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S |
| 3 | 0 | 113803 | 53.1000 | C123 | S |
| 4 | 0 | 373450 | 8.0500 | NaN | S |

```
In [9]: 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

```

```

In [10]: import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

```

```

In [11]: def number(passenger):
passengerid=passenger
count=0
for passengernumber in range(passengerid):
    count=count+1
return count

```

```

In [12]: titanic_df['count']=titanic_df[['PassengerId']].apply(number,axis=1)

```

```

In [13]: titanic_df.head()

```

```

Out[13]:   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  count
0         0      A/5 21171    7.2500   NaN        S        1
1         0      PC 17599   71.2833   C85        C        2
2         0  STON/O2. 3101282    7.9250   NaN        S        3
3         0      113803   53.1000  C123        S        4
4         0      373450    8.0500   NaN        S        5

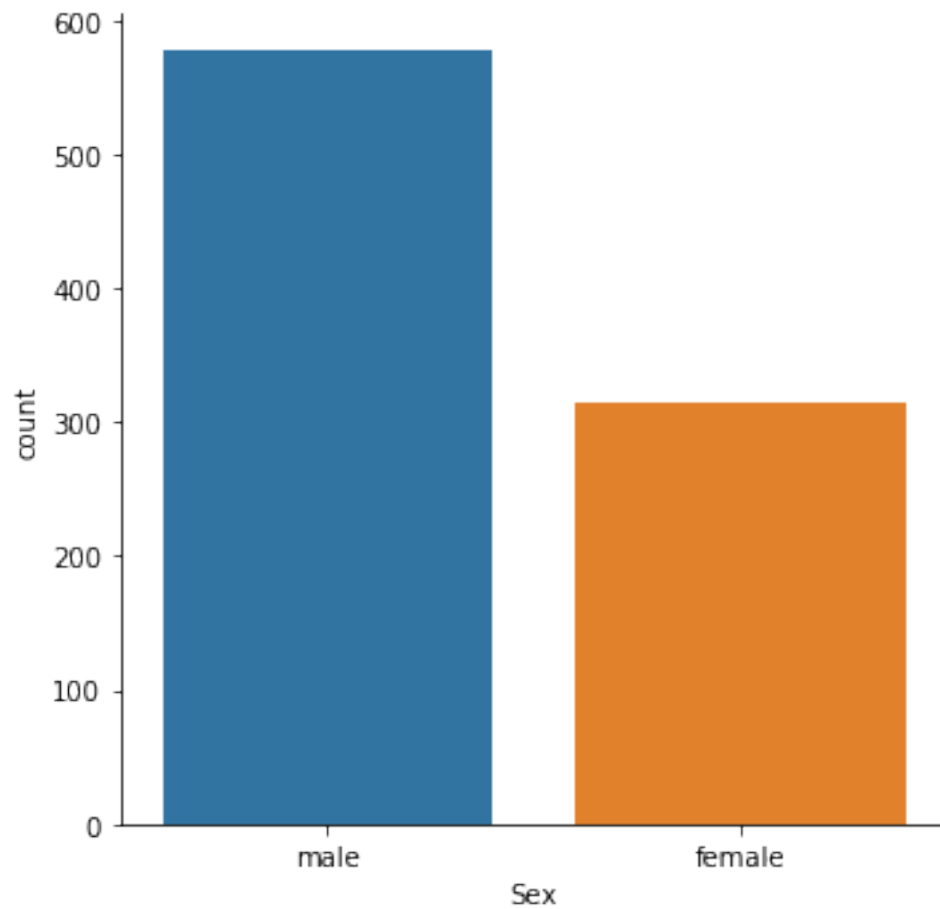
```

```

In [14]: sns.catplot('Sex',kind='count',data=titanic_df)

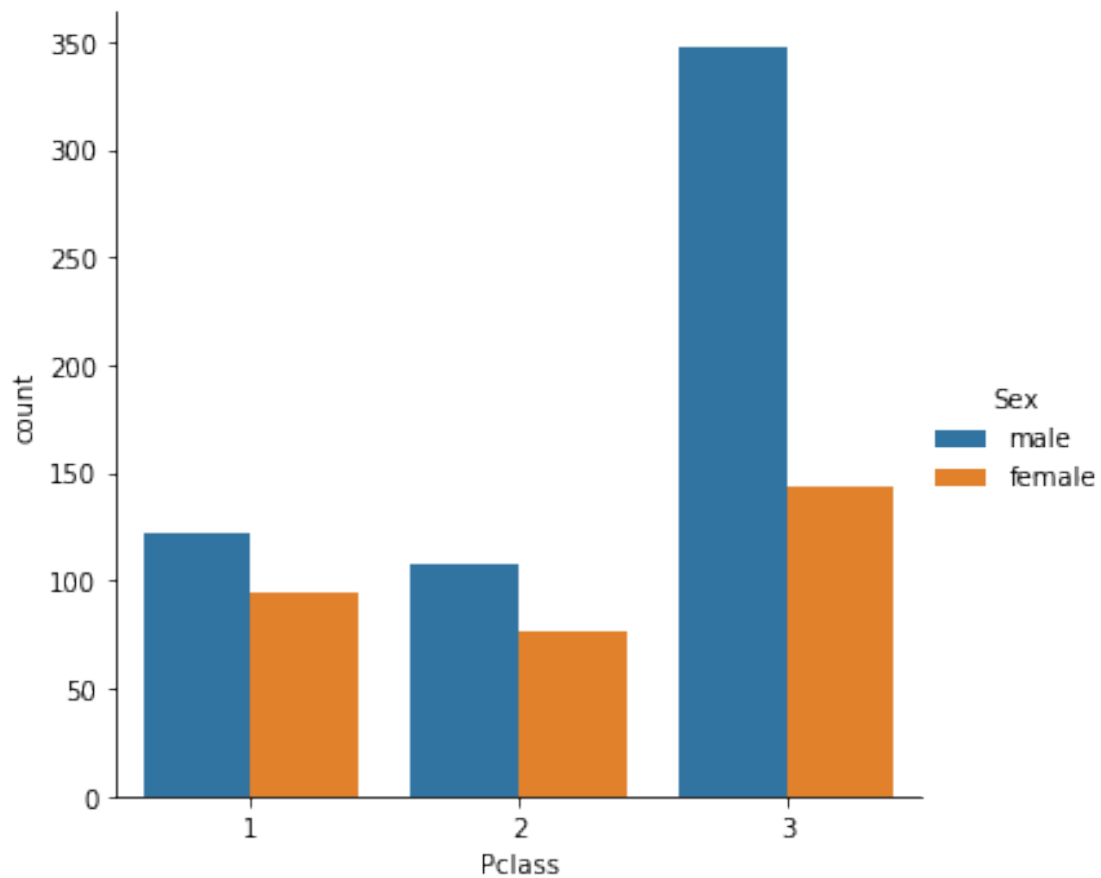
```

```
Out[14]: <seaborn.axisgrid.FacetGrid at 0x7f9d4eb29750>
```



```
In [15]: sns.catplot('Pclass',kind='count',hue='Sex',data=titanic_df)
```

```
Out[15]: <seaborn.axisgrid.FacetGrid at 0x7f9d905ed410>
```



```
In [16]: def male_female_child(passenger):
          age,sex=passenger
          if age<16:
              return 'child'
          else:
              return sex
```

```
In [17]: titanic_df['person']=titanic_df[['Age','Sex']].apply(male_female_child,axis=1)
```

```
In [18]: titanic_df[0:10]
```

```
Out[18]:
```

| | 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 | |

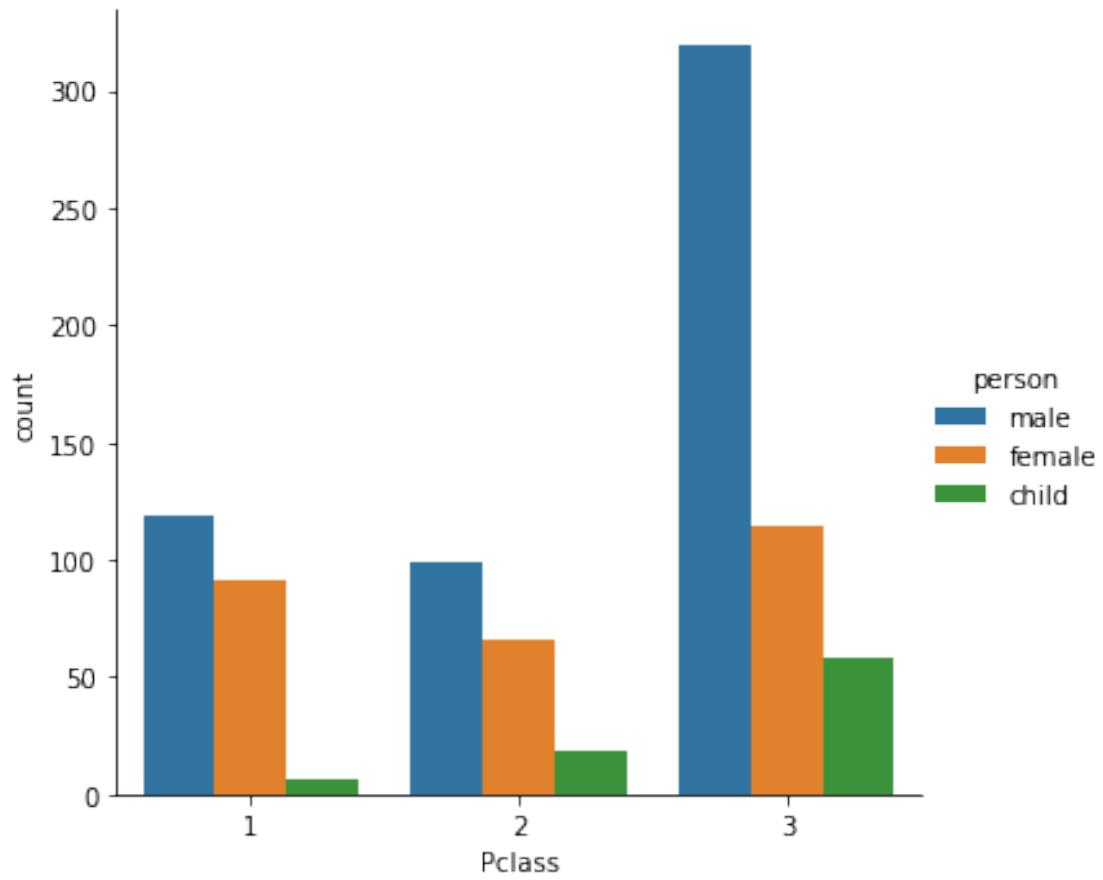
| | | | |
|---|----|---|---|
| 7 | 8 | 0 | 3 |
| 8 | 9 | 1 | 3 |
| 9 | 10 | 1 | 2 |

| | 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 | count | person |
|---|-------|------------------|---------|-------|----------|-------|--------|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | male |
| 1 | 0 | PC 17599 | 71.2833 | C85 | C | 2 | female |
| 2 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S | 3 | female |
| 3 | 0 | 113803 | 53.1000 | C123 | S | 4 | female |
| 4 | 0 | 373450 | 8.0500 | NaN | S | 5 | male |
| 5 | 0 | 330877 | 8.4583 | NaN | Q | 6 | male |
| 6 | 0 | 17463 | 51.8625 | E46 | S | 7 | male |
| 7 | 1 | 349909 | 21.0750 | NaN | S | 8 | child |
| 8 | 2 | 347742 | 11.1333 | NaN | S | 9 | female |
| 9 | 0 | 237736 | 30.0708 | NaN | C | 10 | child |

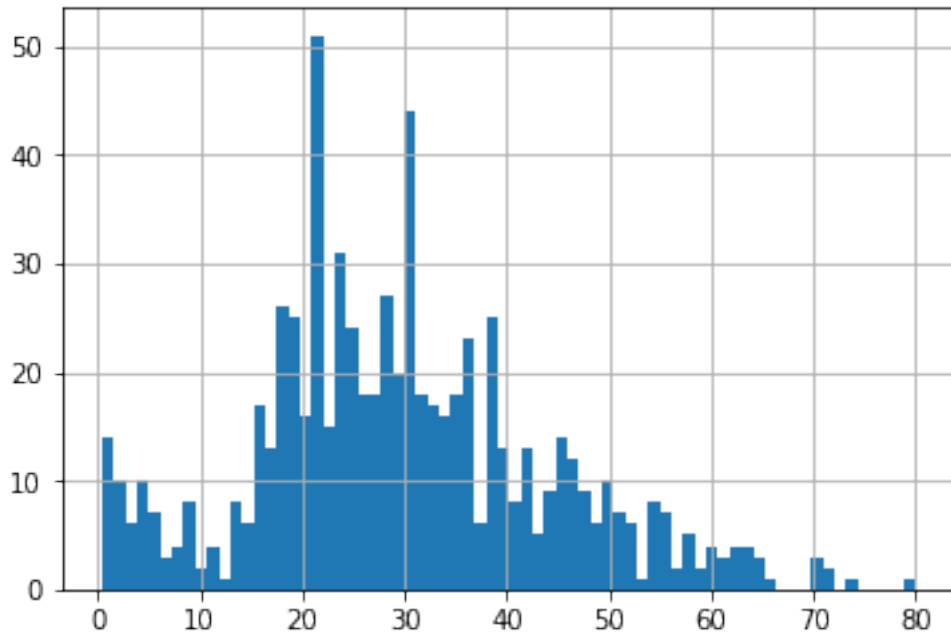
In [19]: sns.catplot('Pclass',kind='count',data=titanic_df,hue='person')

Out[19]: <seaborn.axisgrid.FacetGrid at 0x7f9d65edd790>



```
In [20]: titanic_df['Age'].hist(bins=70)
```

```
Out[20]: <matplotlib.axes._subplots.AxesSubplot at 0x7f9d90502950>
```



```
In [21]: titanic_df['Age'].mean()
```

```
Out[21]: 29.69911764705882
```

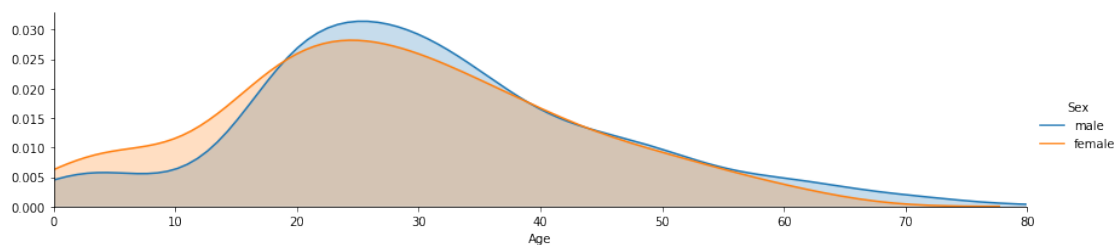
```
In [22]: titanic_df['person'].value_counts()
```

```
Out[22]: male      537
         female    271
         child     83
         Name: person, dtype: int64
```

```
In [23]: 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()
```

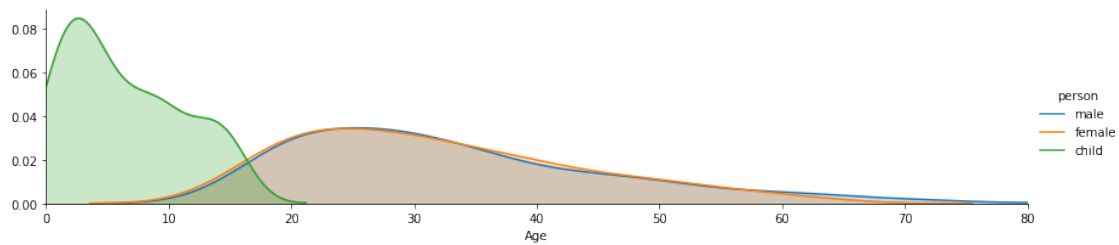
```
/home/naman/.local/lib/python2.7/site-packages/scipy/stats/stats.py:1713: FutureWarning: Using
return np.add.reduce(sorted[indexer] * weights, axis=axis) / sumval
```

```
Out[23]: <seaborn.axisgrid.FacetGrid at 0x7f9d65c45410>
```



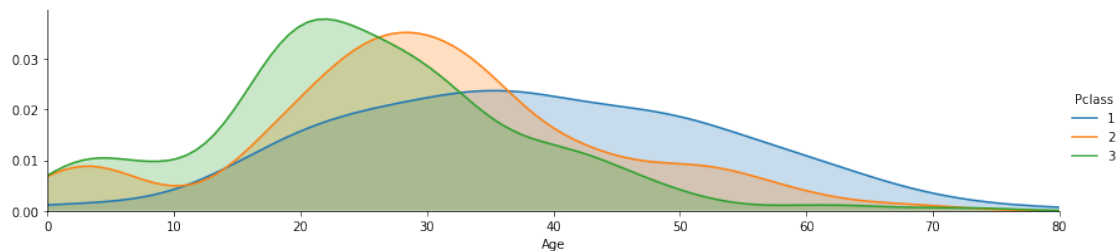
```
In [24]: fig=sns.FacetGrid(titanic_df,hue='person',aspect=4)
fig.map(sns.kdeplot,'Age',shade=True)
oldest=titanic_df['Age'].max()
fig.set(xlim=(0,oldest))
fig.add_legend()
```

Out [24]: <seaborn.axisgrid.FacetGrid at 0x7f9d65d9bb90>



```
In [25]: 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 [25]: <seaborn.axisgrid.FacetGrid at 0x7f9d4ec40fd0>



```
In [26]: titanic_df.head()
```

```
Out [26]:   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 | Cummings, 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 | count | person |
|---|-------|------------------|---------|-------|----------|-------|--------|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | male |
| 1 | 0 | PC 17599 | 71.2833 | C85 | C | 2 | female |
| 2 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S | 3 | female |
| 3 | 0 | 113803 | 53.1000 | C123 | S | 4 | female |
| 4 | 0 | 373450 | 8.0500 | NaN | S | 5 | male |

```
In [27]: deck=titanic_df['Cabin'].dropna()
```

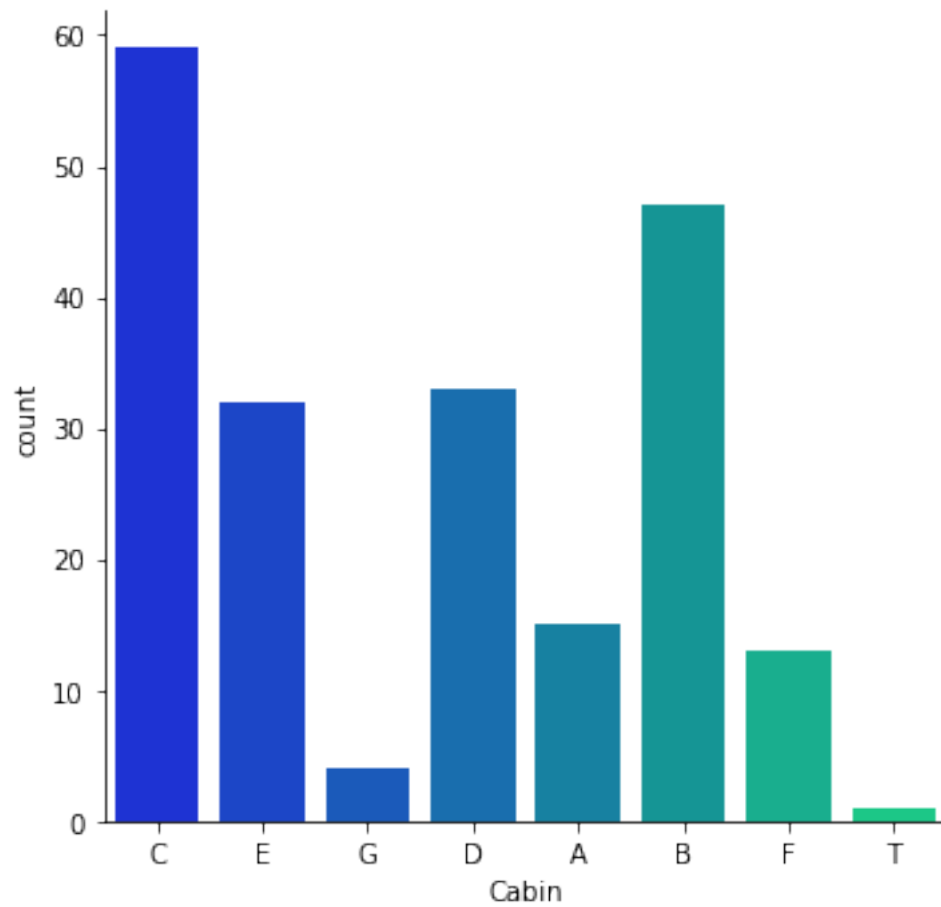
```
In [28]: deck.head()
```

```
Out[28]: 1      C85
3      C123
6      E46
10     G6
11     C103
Name: Cabin, dtype: object
```

```
In [29]: levels=[]
for level in deck:
    levels.append(level[0])

cabin_df=DataFrame(levels)
cabin_df.columns=['Cabin']
sns.catplot('Cabin',kind='count',data=cabin_df,palette='winter')
```

```
Out[29]: <seaborn.axisgrid.FacetGrid at 0x7f9d65cc8b10>
```

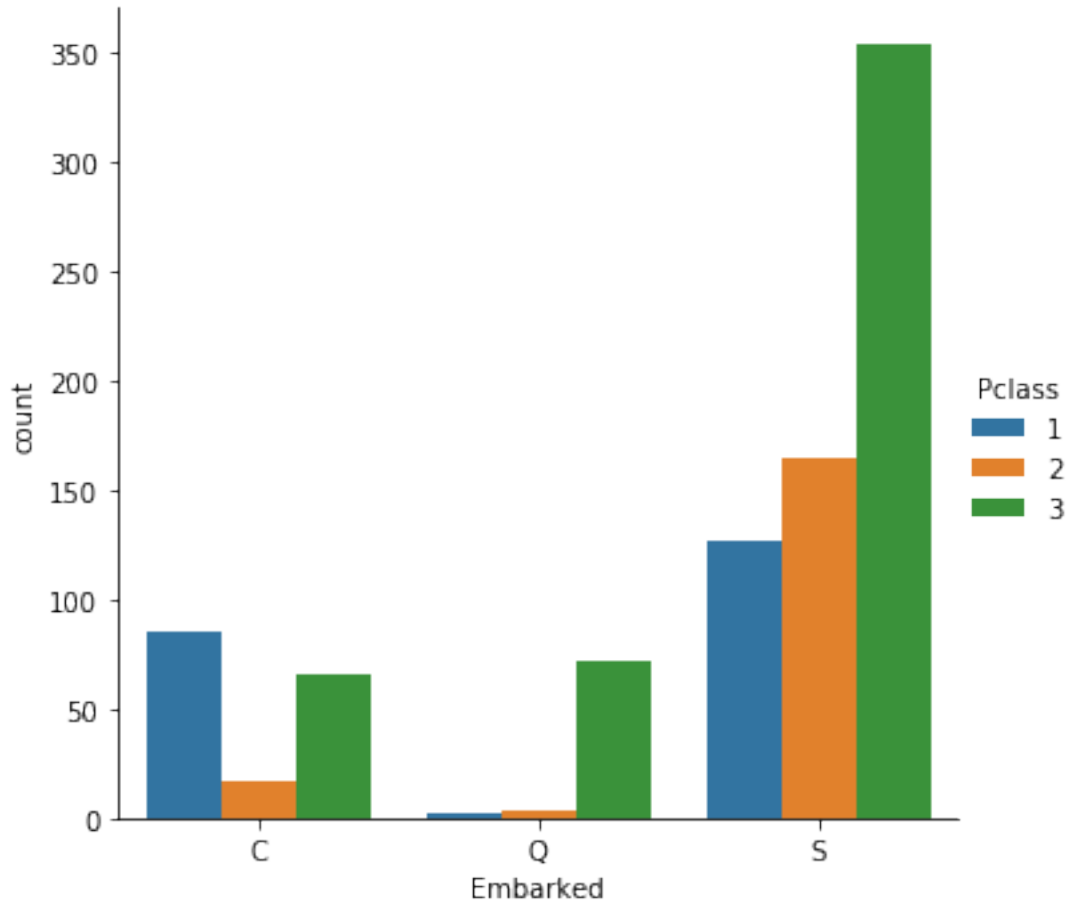


```
In [52]: cabin_df.head()
```

```
Out[52]:   Cabin
0      C
1      C
2      E
3      G
4      C
```

```
In [31]: sns.catplot('Embarked',data=titanic_df,kind='count',hue='Pclass',order=['C','Q','S'])
```

```
Out[31]: <seaborn.axisgrid.FacetGrid at 0x7f9d65c865d0>
```



```
In [32]: titanic_df['Alone']=titanic_df['SibSp']+titanic_df['Parch']
```

```
In [53]: titanic_df['Alone'].head()
```

```
Out[53]: 0    With Family
         1    With Family
         2         Alone
         3    With Family
         4         Alone
         Name: Alone, dtype: object
```

```
In [34]: titanic_df['Alone'].loc[titanic_df['Alone']>0] = 'With Family'
         titanic_df['Alone'].loc[titanic_df['Alone']==0] = 'Alone'
```

/home/naman/.local/lib/python2.7/site-packages/pandas/core/indexing.py:189: 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>
 self._setitem_with_indexer(indexer, value)

```
In [35]: titanic_df.head()
```

```
Out[35]:
```

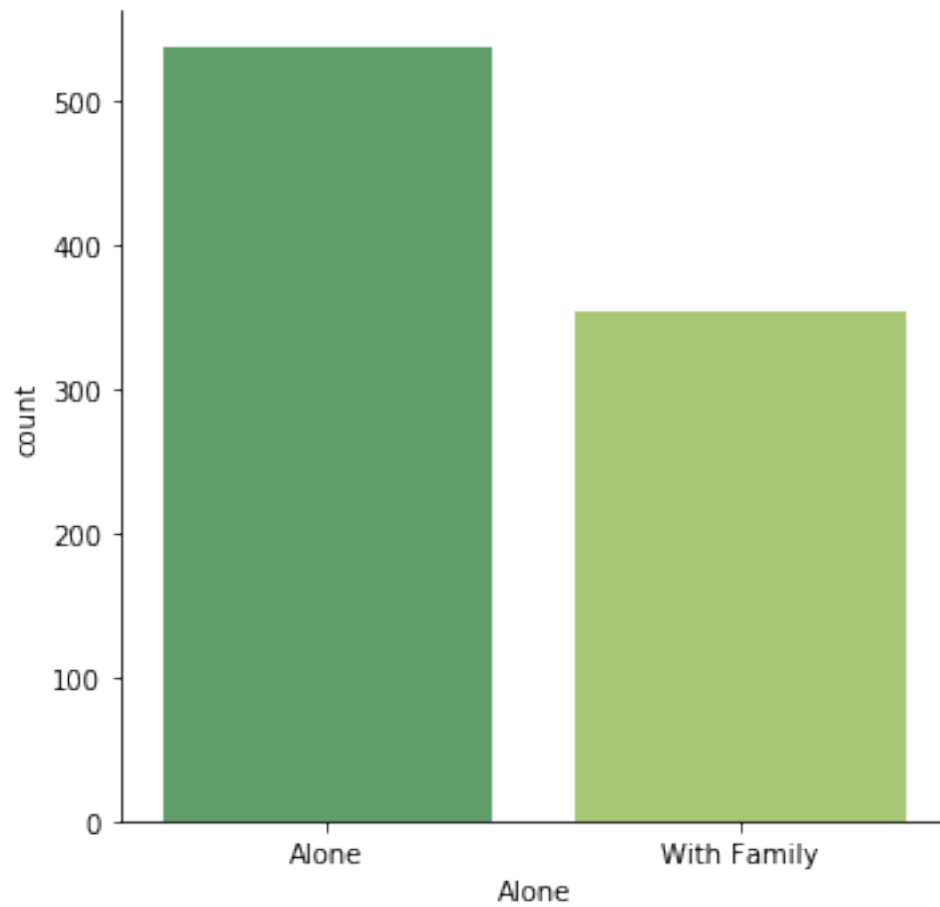
| | PassengerId | Survived | Pclass | \ |
|---|-------------|----------|--------|---|
| 0 | 1 | 0 | 3 | |
| 1 | 2 | 1 | 1 | |
| 2 | 3 | 1 | 3 | |
| 3 | 4 | 1 | 1 | |
| 4 | 5 | 0 | 3 | |

| | 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 | count | person | Alone |
|---|-------|------------------|---------|-------|----------|-------|--------|-------------|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | male | With Family |
| 1 | 0 | PC 17599 | 71.2833 | C85 | C | 2 | female | With Family |
| 2 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S | 3 | female | Alone |
| 3 | 0 | 113803 | 53.1000 | C123 | S | 4 | female | With Family |
| 4 | 0 | 373450 | 8.0500 | NaN | S | 5 | male | Alone |

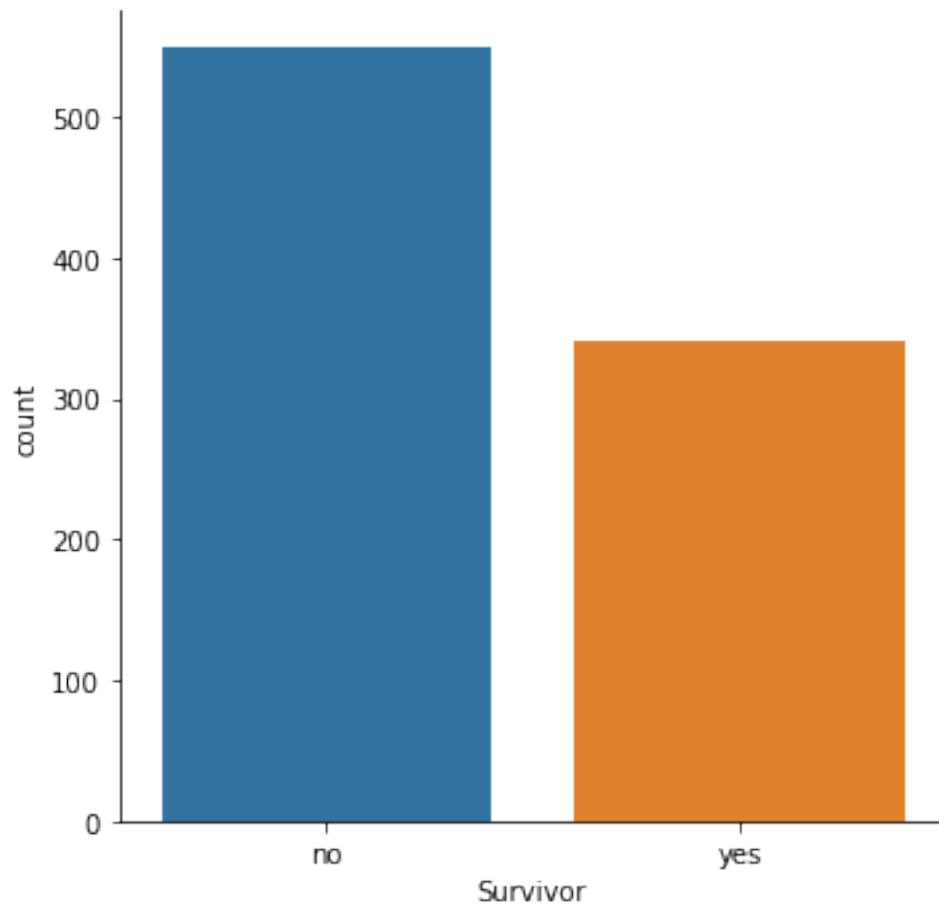
```
In [36]: sns.catplot('Alone',kind='count',data=titanic_df,order=['Alone','With Family'],palette='magma')
```

```
Out[36]: <seaborn.axisgrid.FacetGrid at 0x7f9d65c2c790>
```



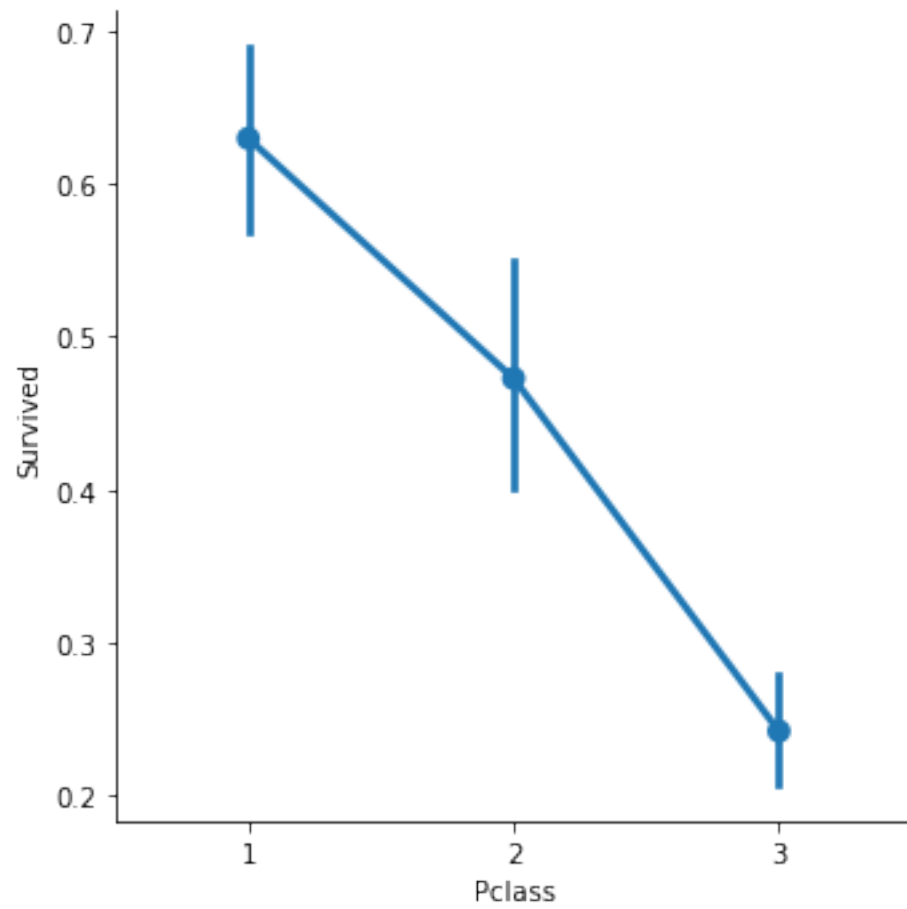
```
In [37]: titanic_df['Survivor']=titanic_df.Survived.map({0:'no',1:'yes'})
         sns.catplot('Survivor',kind='count',data=titanic_df)
```

```
Out[37]: <seaborn.axisgrid.FacetGrid at 0x7f9d65c45790>
```



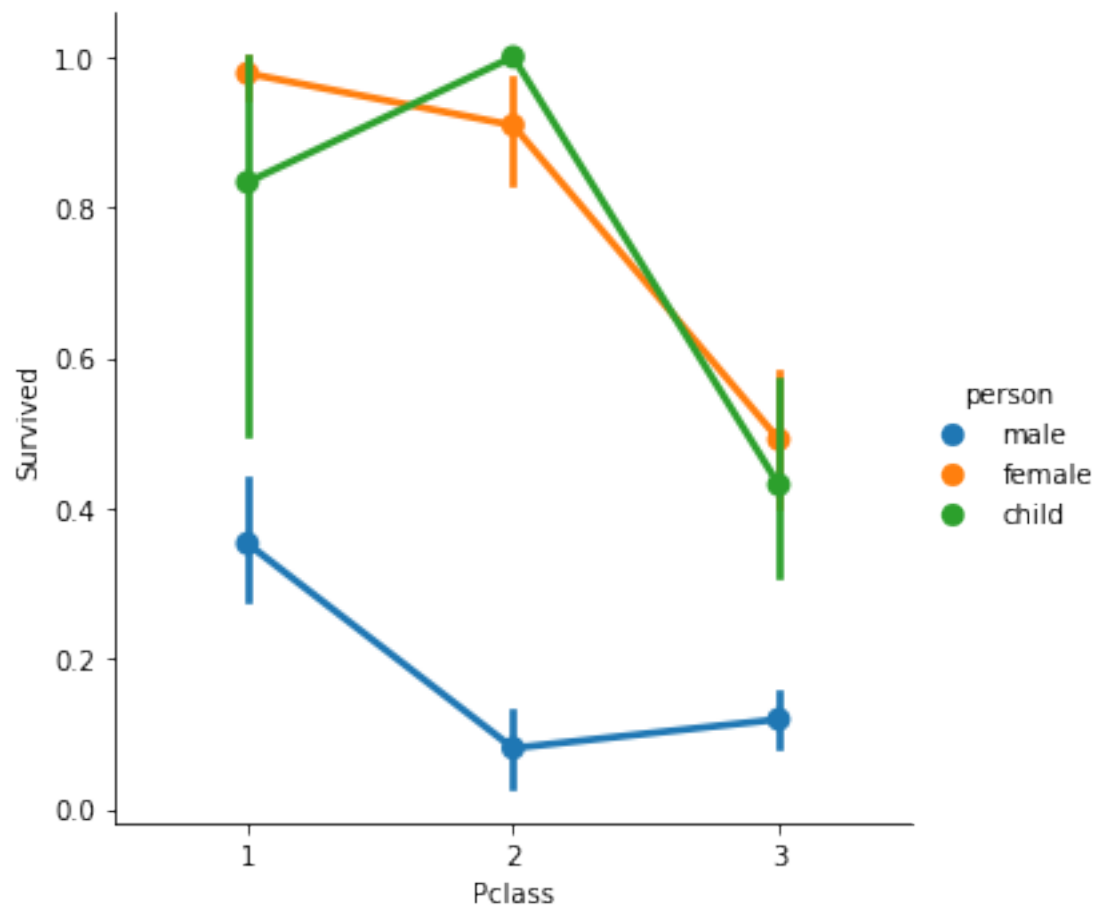
```
In [38]: sns.catplot(x='Pclass',y='Survived',kind='point',data=titanic_df,aspect=1)
```

```
Out[38]: <seaborn.axisgrid.FacetGrid at 0x7f9d6590ef10>
```



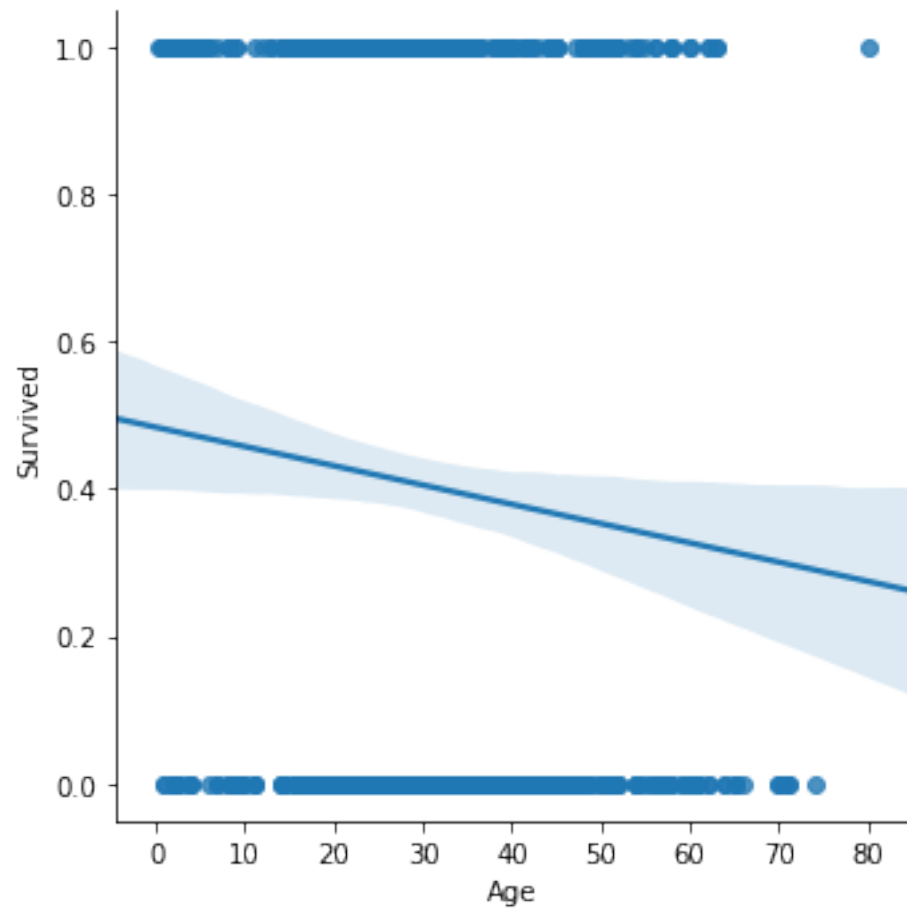
```
In [39]: sns.catplot(x='Pclass',y='Survived',hue='person',kind='point',data=titanic_df,aspect=
```

```
Out[39]: <seaborn.axisgrid.FacetGrid at 0x7f9d658dbf50>
```



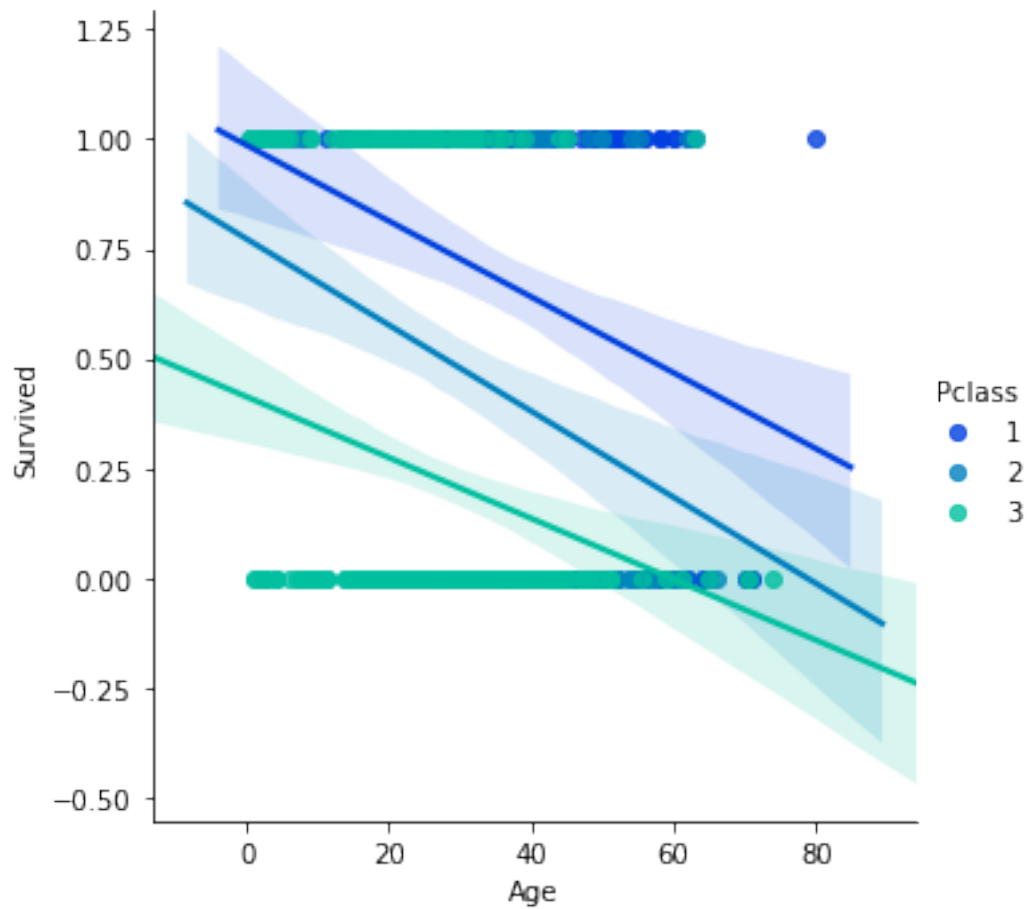
```
In [40]: sns.lmplot('Age', 'Survived', data=titanic_df)
```

```
Out[40]: <seaborn.axisgrid.FacetGrid at 0x7f9d43ff8350>
```

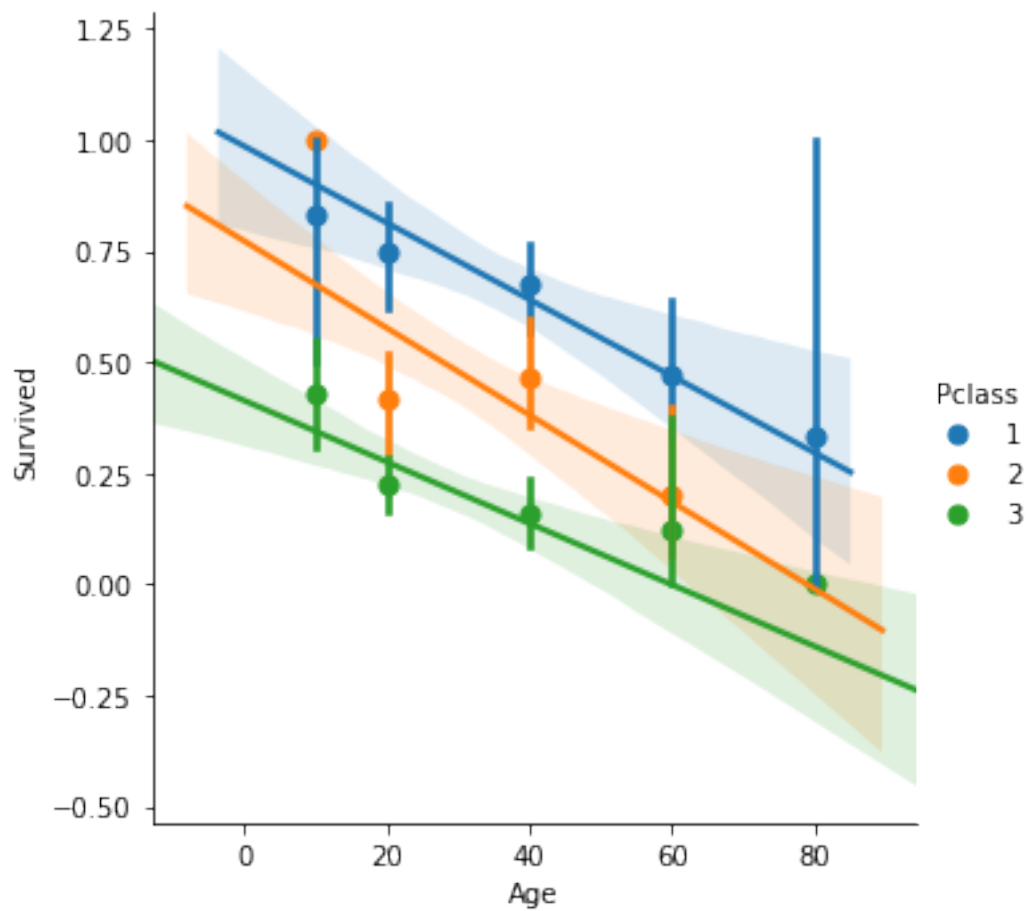
```
In [41]: sns.lmplot('Age', 'Survived', data=titanic_df, hue='Pclass', palette='winter')
```

```
Out[41]: <seaborn.axisgrid.FacetGrid at 0x7f9d4400de90>
```



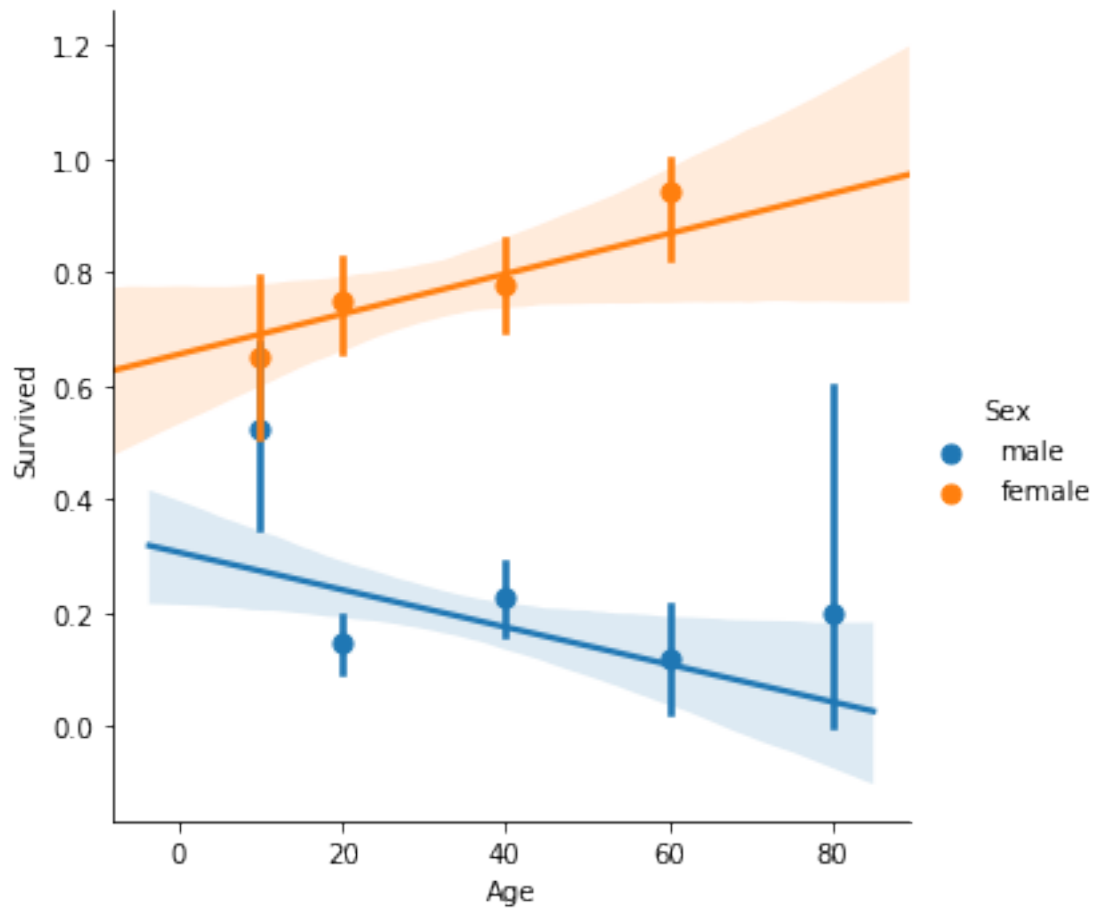
```
In [42]: generations = [10,20,40,60,80]
sns.lmplot('Age', 'Survived', data=titanic_df, hue='Pclass', x_bins=generations)
```

```
Out[42]: <seaborn.axisgrid.FacetGrid at 0x7f9d43e7c650>
```



```
In [43]: sns.lmplot('Age', 'Survived', data=titanic_df, hue='Sex', x_bins=generations)
```

```
Out[43]: <seaborn.axisgrid.FacetGrid at 0x7f9d43f537d0>
```



```
In [54]: titanic_df.head()
```

```
Out[54]:
```

| | 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 | |

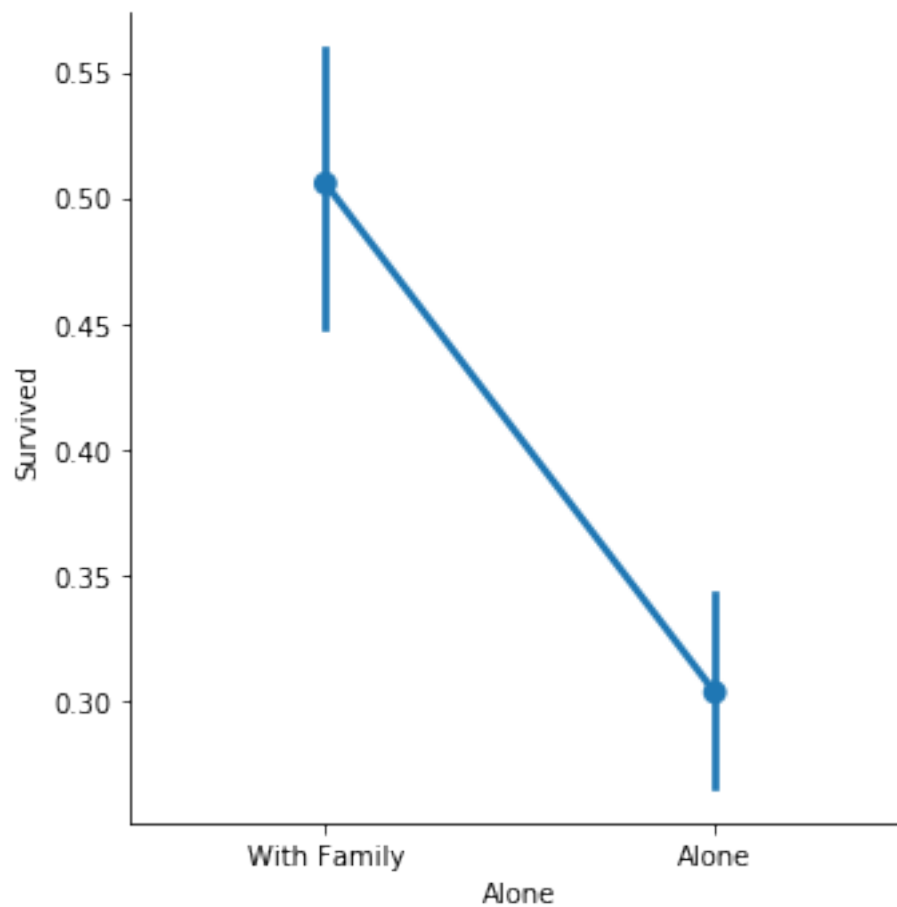
| | Parch | Ticket | Fare | Cabin | Embarked | count | person | \ |
|---|-------|-----------|--------|-------|----------|-------|--------|---|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | male | |

| | | | | | | | |
|---|---|----------|---------|---------|------|---|--------|
| 1 | 0 | PC 17599 | 71.2833 | C85 | C | 2 | female |
| 2 | 0 | STON/O2. | 3101282 | 7.9250 | NaN | 3 | female |
| 3 | 0 | | 113803 | 53.1000 | C123 | 4 | female |
| 4 | 0 | | 373450 | 8.0500 | NaN | 5 | male |

| Alone Survivor Deck | | | |
|---------------------|-------------|-----|---|
| 0 | With Family | no | n |
| 1 | With Family | yes | C |
| 2 | Alone | yes | n |
| 3 | With Family | yes | C |
| 4 | Alone | no | n |

```
In [51]: sns.catplot(x='Alone',y='Survived',kind='point',data=titanic_df)
```

```
Out[51]: <seaborn.axisgrid.FacetGrid at 0x7f9d43c00dd0>
```



```
In [46]: def dec(cabin):
cab=cabin
```

```

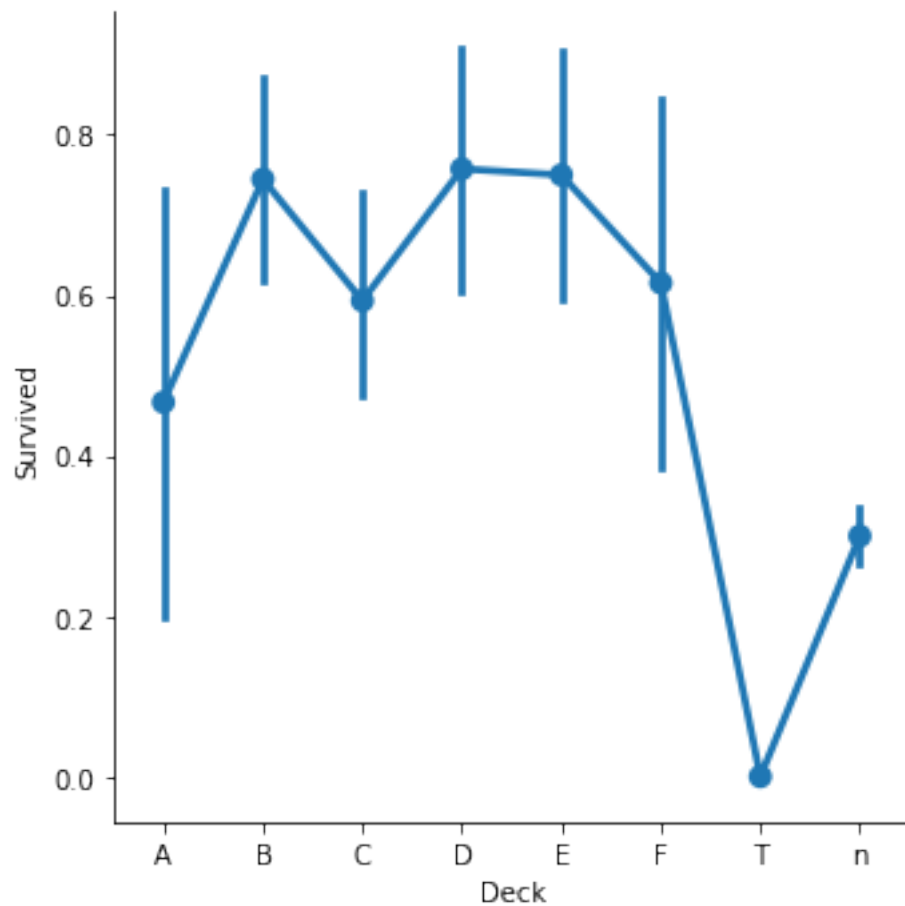
for c in cab:
    c=str(c)
    if c=='NaN':
        return 0
    else:
        return c[0]

```

```
In [47]: titanic_df['Deck']=titanic_df[['Cabin']].apply(dec,axis=1)
```

```
In [48]: sns.catplot('Deck', 'Survived', kind='point', order=['A', 'B', 'C', 'D', 'E', 'F', 'T', 'n'], data=titanic_df)
```

```
Out[48]: <seaborn.axisgrid.FacetGrid at 0x7f9d43d5d8d0>
```



```
In [55]: titanic_df.head()
```

```

Out[55]:
   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 | count | person | \ |
|---|-------|------------------|---------|-------|----------|-------|--------|---|
| 0 | 0 | A/5 21171 | 7.2500 | NaN | S | 1 | male | |
| 1 | 0 | PC 17599 | 71.2833 | C85 | C | 2 | female | |
| 2 | 0 | STON/O2. 3101282 | 7.9250 | NaN | S | 3 | female | |
| 3 | 0 | 113803 | 53.1000 | C123 | S | 4 | female | |
| 4 | 0 | 373450 | 8.0500 | NaN | S | 5 | male | |

| | Alone | Survivor | Deck |
|---|-------------|----------|------|
| 0 | With Family | no | n |
| 1 | With Family | yes | C |
| 2 | Alone | yes | n |
| 3 | With Family | yes | C |
| 4 | Alone | no | n |