Titanic Sample Data Investigation

In this investigation, I would like to know if the ticket class (hereafter referred to as 'Pclass') of the passengers influence the rate of survival? If so, was it worse for men or women?

In [23]:
```
import unicodecsv
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
%pylab inline
```

```
Populating the interactive namespace from numpy and matplotlib
```

In [24]:
```
## Converting data to a PANDAS dataframe
titanic_df = pd.read_csv('titanic_data.csv')
```

In [25]:
```
## Getting information about the dataframe
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
```

This data shows that there is data missing in the Age column (177 entries), Cabin column (604 entries), and Embarked column (2 entries). My investigation did not rely on this data, so I left it alone. However, it would be interesting to look at what age groups in each gender and class survived, and how that might be significant.

In [26]:
```
total_passengers = titanic_df.groupby(['Pclass', 'Sex']).count()
total_passengers
```

Out[26]:

		PassengerId	Survived	Name	Age	SibSp	Parch	Ticket	Fare	Cabin
Pclass	**Sex**									
1	**female**	94	94	94	85	94	94	94	94	81
	male	122	122	122	101	122	122	122	122	95
2	**female**	76	76	76	74	76	76	76	76	10
	male	108	108	108	99	108	108	108	108	6
3	**female**	144	144	144	102	144	144	144	144	6
	male	347	347	347	253	347	347	347	347	6

In [27]:
```
## Count of total males and females in each Pclass
total_surv= titanic_df.groupby(['Pclass','Sex'])[['Survived']].count()
total_surv
```

Out[27]:

		Survived
Pclass	**Sex**	
1	**female**	94
	male	122
2	**female**	76
	male	108
3	**female**	144
	male	347

In [28]:
```
## Count of both survivors (shown by '1') and nonsurvivors (shown by '0') grouped by Pclass and Sex
surv_nonsurv= titanic_df.groupby(['Pclass','Sex', 'Survived'])[['Survived']].count()
surv_nonsurv
```

Out[28]:

			Survived
Pclass	**Sex**	**Survived**	
1	**female**	**0**	3
		1	91
	male	**0**	77
		1	45
2	**female**	**0**	6
		1	70
	male	**0**	91
		1	17
3	**female**	**0**	72
		1	72
	male	**0**	300
		1	47

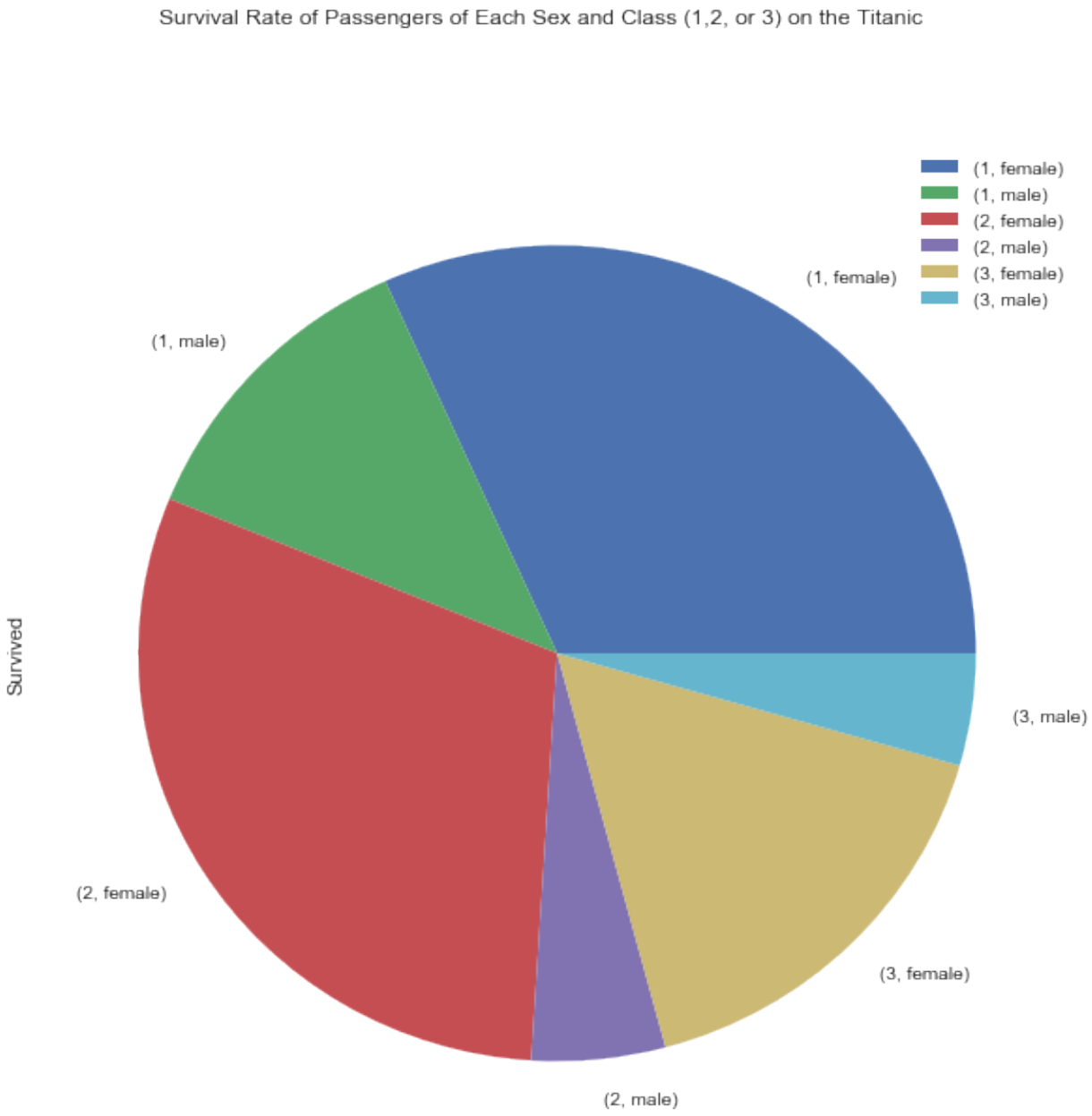
In [29]:
```
## Percent of survivors based on Pclass and Sex
percent_survived = titanic_df.groupby(['Pclass', 'Sex'])[['Survived']]
.mean()
percent_survived
```

Out[29]:

		Survived
Pclass	**Sex**	
1	**female**	0.968085
	male	0.368852
2	**female**	0.921053
	male	0.157407
3	**female**	0.500000
	male	0.135447

In [30]:
```
## Pie chart of total survivors based on Pclass and Sex
percent_survived.plot.pie(title='Survival Rate of Passengers of Each S
ex and Class (1,2, or 3) on the Titanic', subplots=True, figsize=(10,1
0))
```

Out[30]:
```
array([<matplotlib.axes._subplots.AxesSubplot object at 0x11de4e550>
], dtype=object)
```



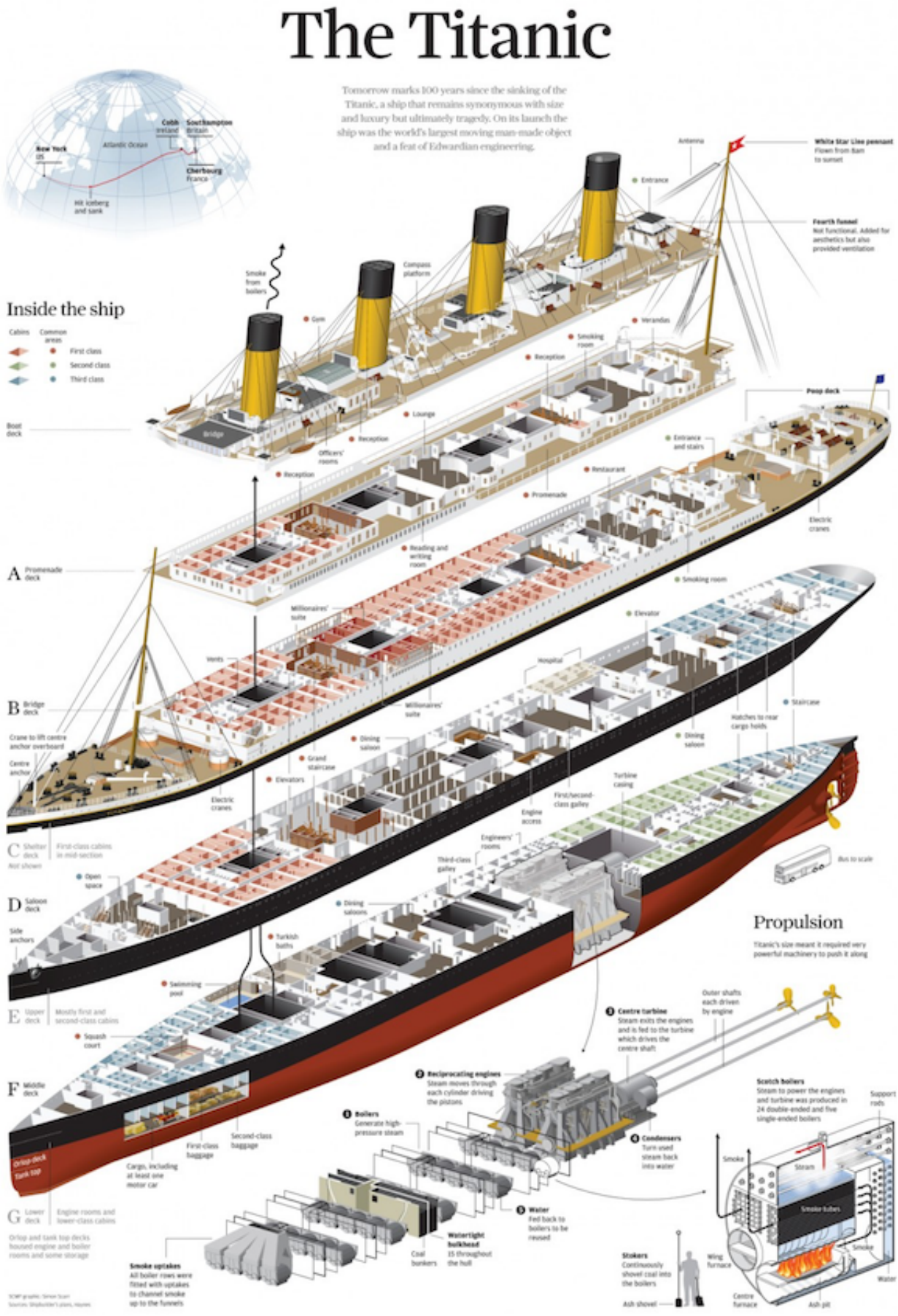
This pie chart illustrates the disproportional survivor rates amongst different ticket classes. The first class survivors account for almost half of the total survivors. If you consider both the first and second classes, they represent over 3/4 of the total survivors. Had class not been a factor, one would expect these figures to be closer to 1/3 and 2/3, respectively.

In [31]:
```
## Bar graph of total survivors based on Pclass and Sex
percent_survived.plot.bar(title='Percent of Survival of Males and Female Passengers in First, Second, and Third Class on the Titanic').set(xlabel='Ticket Class and Sex', ylabel='Percent of Survival')
```

Out[31]:
```
[<matplotlib.text.Text at 0x11d7a94a8>, <matplotlib.text.Text at 0x11e0c4be0>]
```

This bar graph emphasizes the difference in male and female survivors in each ticket class, where a much higher percentage of females survived in each class (particularly in first and second class), as well as overall. This is evidenced by the fact that 97% of first class female passengers survived, whereas only 50% of third class females survived. Of the third class males, only 47 out of 347 (14%) survived.

from IPython.display import Image


The Titanic
Tomorrow marks 100 years since the sinking of the Titanic, a ship that remains synonymous with size and luxury but ultimately tragedy. On its launch the ship was the world's largest moving man-made object and a feat of Edwardian engineering.
Cobh Ireland
Southampton Britain
New York US
Atlantic Ocean
Cherbourg France
Hit iceberg and sank
Inside the ship
Cabins
Common areas
First class
Second class
Third class
Boat deck
A Promenade deck
B Bridge deck
C Shelter deck
D Saloon deck
E Upper deck
F Middle deck
G Lower deck
Orlop and tank top decks housed engine and boiler rooms and some storage
Antenna
White Star Line pennant
Flown from 8am to sunset
Entrance
Fourth funnel
Not functional. Added for aesthetics but also provided ventilation
Smoke from boilers
Compass platform
Gym
Verandas
Smoking room
Reception
Bridge
Lounge
Officers' rooms
Poop deck
Entrance and stairs
Restaurant
Promenade
Electric cranes
Reading and writing room
Millionaires' suite
Vents
Elevator
Hospital
Staircase
Hatches to rear cargo holds
Crane to lift centre anchor overboard
Centre anchor
Grand staircase
Elevators
Dining saloon
Electric cranes
First/second-class galley
Turbine casing
Engine access
Engineers' rooms
Third-class galley
Bus to scale
Open space
Side anchors
Turkish baths
Swimming pool
Squash court
Mostly first and second-class cabins
First-class cabins in mid-section
Not shown
Engine rooms and lower-class cabins
Cargo, including at least one motor car
First-class baggage
Second-class baggage
Orlop deck
Tank top
Propulsion
Titanic's size meant it required very powerful machinery to push it along
1 Centre turbine
Steam exits the engines and is fed to the turbine which drives the centre shaft
Outer shafts each driven by engine
2 Reciprocating engines
Steam moves through each cylinder driving the pistons
3 Boilers
Generate high-pressure steam
4 Condensers
Turn used steam back into water
5 Water
Fed back to boilers to be reused
Scotch boilers
Steam to power the engines and turbine was produced in 24 double-ended and five single-ended boilers
Support rods
Smoke
Steam
Smoke tubes
Water
Wing furnace
Centre furnace
Ash pit
Ash shovel
Stokers
Continuously shovel coal into the boilers
Coal bunkers
Watertight bulkhead
15 throughout the hull
Smoke uptakes
All boiler rows were fitted with uptakes to channel smoke up to the funnels

This is an infographic of the RMS Titanic portraying the layout of the ship. It is interesting to note the proximity of each class's cabins to the deck of the ship with the lifeboats. It is also interesting to compare the number of cabins each Pclass had.

There is a correlation between the survival rate and Pclass of the passengers on the Titanic, especially amongst female passengers. While I cannot state for sure that there is causation here, the correlations are interesting. I would guess that women and children were likely to have been put on lifeboats first, which accounts for the gender bias. I would also suppose that, since the first class cabins were closer to the deck of the ship, they were likely to reach the lifeboats before any lower class persons. In addition, I would think that first class passengers were thought to be more important than lower class passengers, and therefore more readily allowed onto lifeboats.

However, since this is only a sample of the entire data, I cannot conclude anything for certain. There were a total of 2,224 passengers aboard the Titanic when it sailed, so I have less than half of the data. At the time the data was collected, perhaps the socioeconomic status of the passenger affected the "relevance" of the data. In which case, maybe gathering the data for lower class passengers did not seem as important? The Women Suffrage movement was still in progess, so maybe female passenger information was not as important, either.

It would be interesting, with this data, to see if the existence of siblings and spouses (SibSp) or parents and children (Parch) to the passengers affected their survival rate; if they had one or the other, were they more likely to survive? For example, if someone was standing with a child near a lifeboat - whether male or female, old or young adult - would they have been more likely to be let onto a lifeboat? Or, did a passenger's embarkment affect their chance of survival? For example, were there more Irish lower or upper class passengers, male or female, old or young, compared to British or French passengers? And, did this mean their chance of survival was higher or lower?

Resources: https://www.kaggle.com/c/titanic/data (https://www.kaggle.com/c/titanic/data), http://pandas.pydata.org/ (http://pandas.pydata.org/), http://stackoverflow.com (http://stackoverflow.com), http://www.scmp.com/infographics/article/1633333/infographic-titanic (http://www.scmp.com/infographics/article/1633333/infographic-titanic), https://en.wikipedia.org/wiki/RMS_Titanic (https://en.wikipedia.org/wiki/RMS_Titanic), https://en.wikipedia.org/wiki/Women's_suffrage (https://en.wikipedia.org/wiki/Women's_suffrage)