Surviving the Titanic

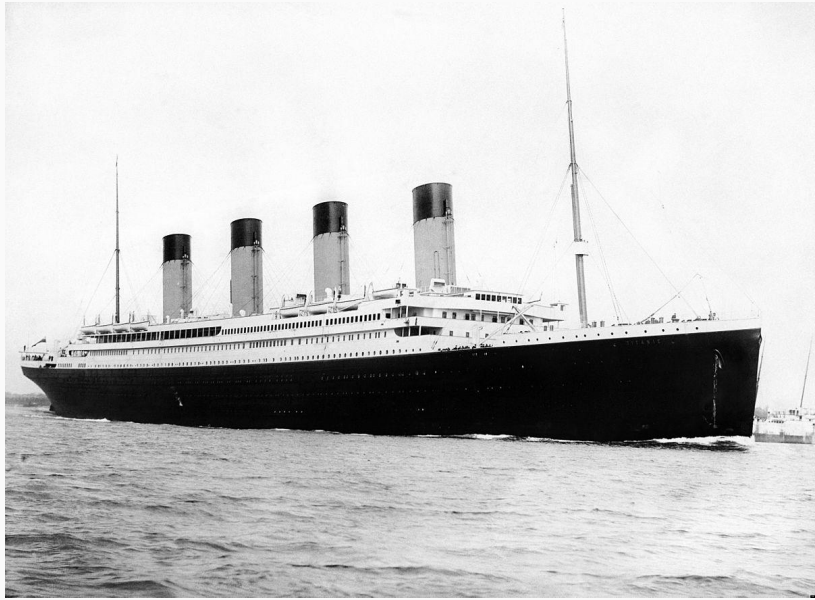
W200 Section 2
Project 2: Data Analysis

Authors:
Danish Iqbal, Matthew Holmes, Kalvin Kao

Background

The Titanic

- Built in 1912 in Belfast, Ireland
  - Largest vessel afloat!
- Transatlantic passenger and mail service
  - Southampton <--> New York
- Started maiden voyage on April 10, 1912
  - 2208 passengers and crew
- Collided with an iceberg four days later
  - 1496 people died
  - 712 survived



The Data

Features of Survival Status

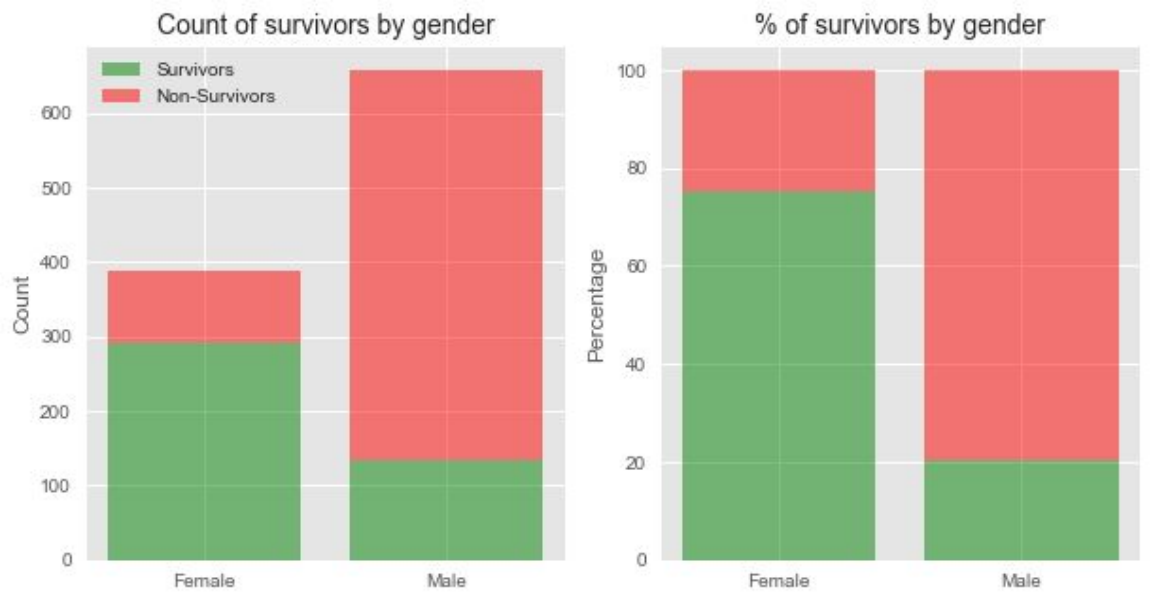
- 1309 passengers (out of 1317)
  - Crew excluded
- 14 variables
  - passenger class, gender, age, name, ticket ID, cabin, point of embarkment, lifeboat, body ID, home/destination, familial relations, fare, **survival**

- Vanderbilt University - Department of Biostatistics
- https://www.encyclopedia-titanica.org/

Questions

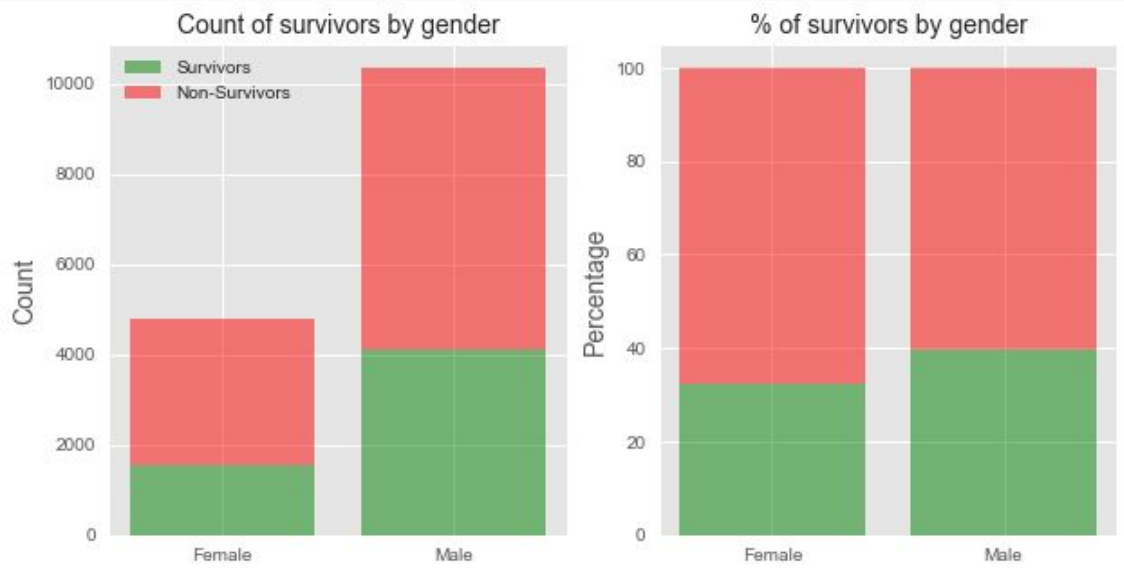
How did survival rates differ between men and women?

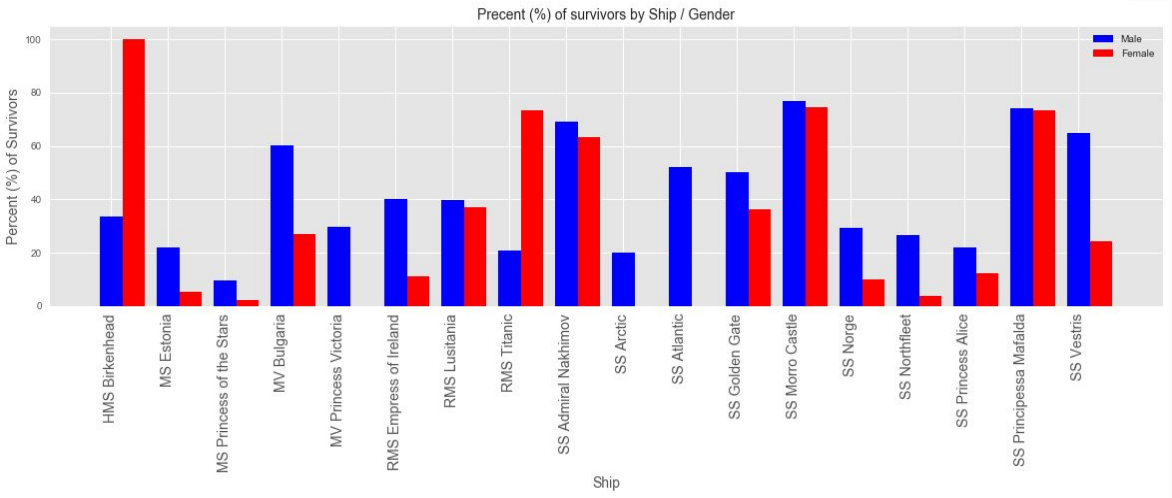
- No Missing Values.
- Womens survival rate is much higher than men's in terms of count and %.
- WCF was followed on the Titanic, sorry Leo fans!
- Was WCF standard practice???



How do gender survival rates on the Titanic compare with other shipwrecks?

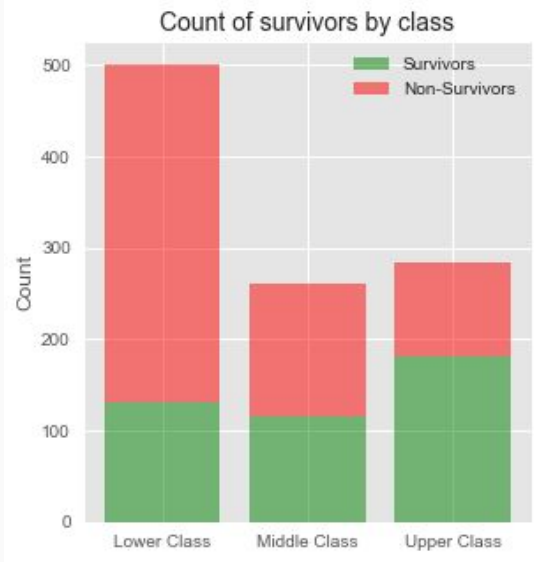
- 18 maritime disasters spanning 3 centuries, 15,000 passengers, and 30 nationalities.
- With the exception of the Birkenhead and Titanic, the men usually have a higher survival rate than women.

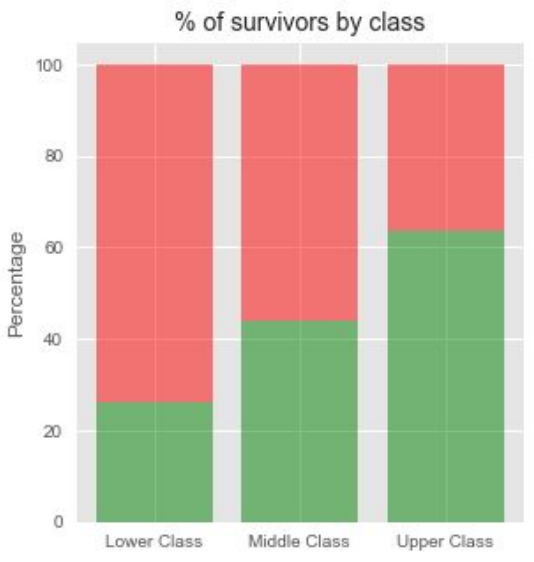


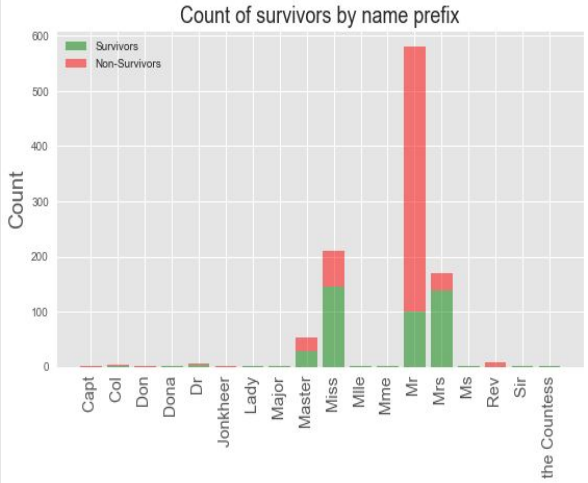


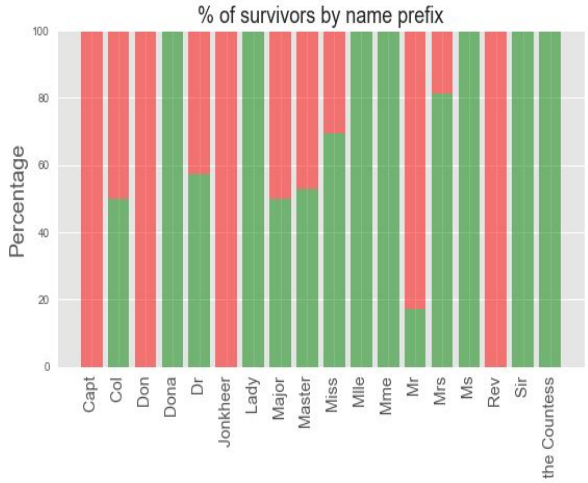
How does social status affect survival rate (pclass and name)?

- Lower class passengers had the worst survival rate
- Upper class passengers had the highest survival rate.
- The Mr.'s had the worst survival rates while the Miss's and Mrs.'s, Masters, and the Countess had high survival rates.





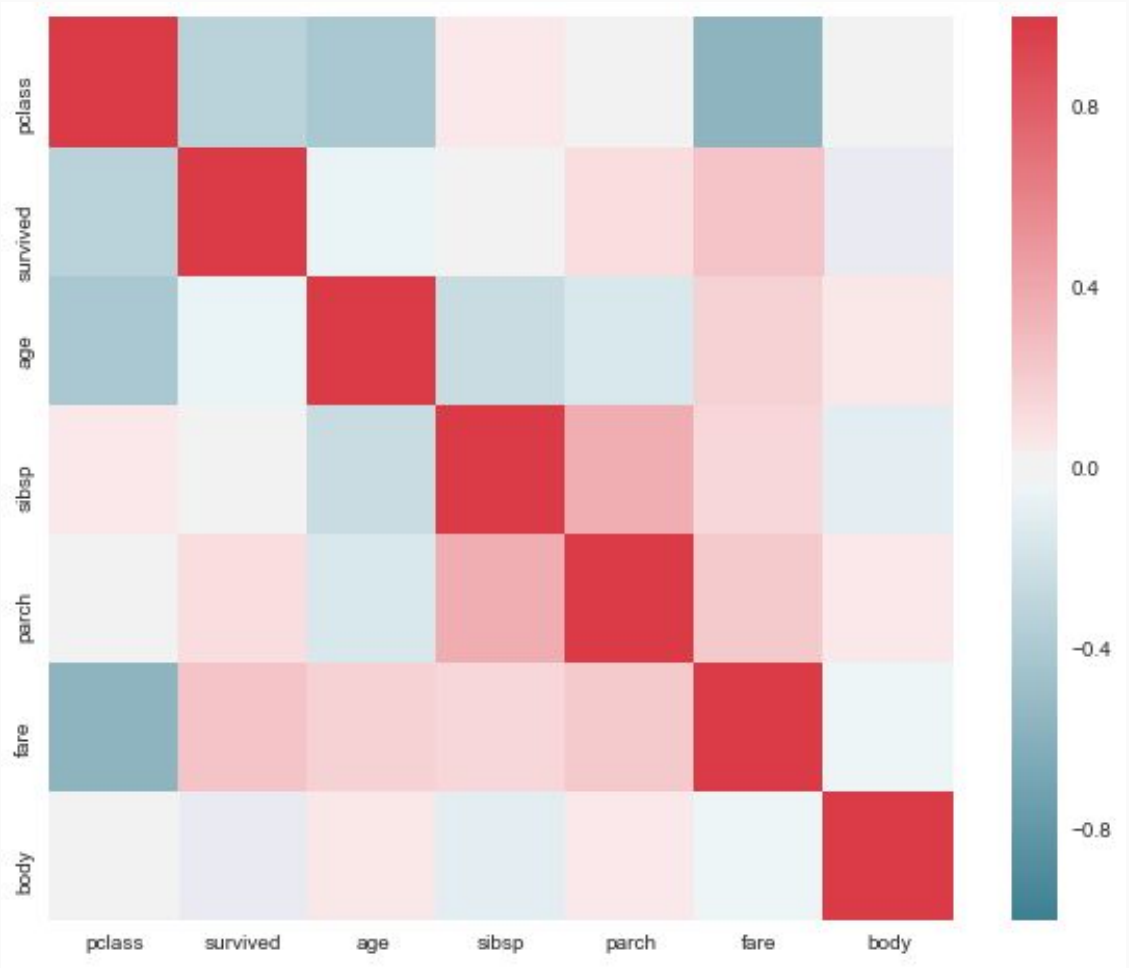




How did survival rates differ between children, adults and elderly?

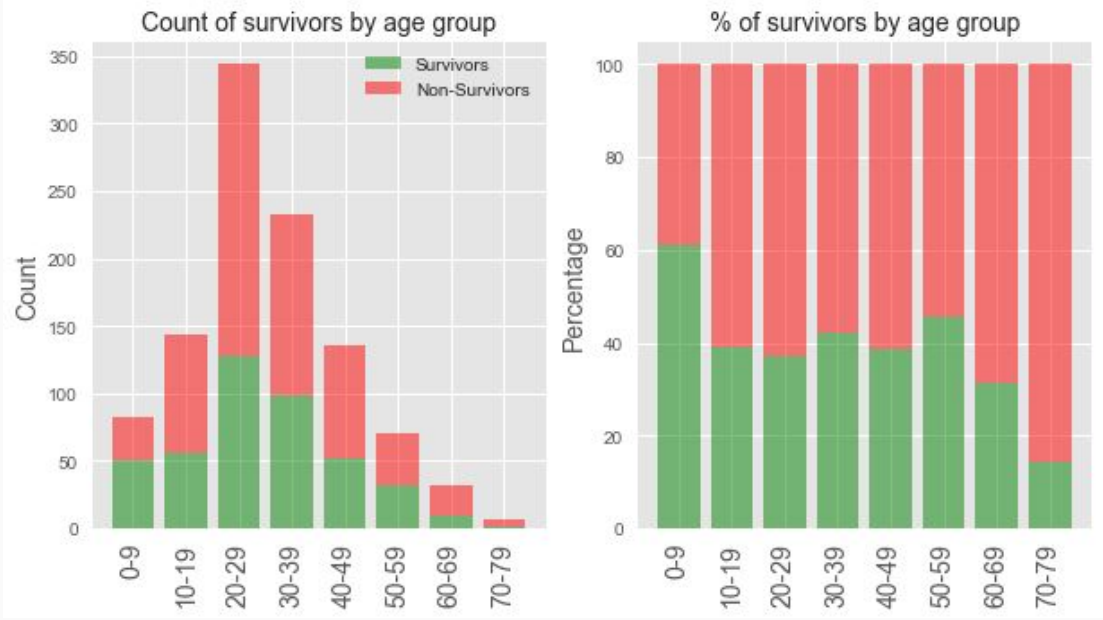
- Age has 263 missing values removed, no strong correlations.
- Tried mean and median, imputation filling in by class.
- Linear regression was also problematic creating negative ages.
- Project for future work . . .

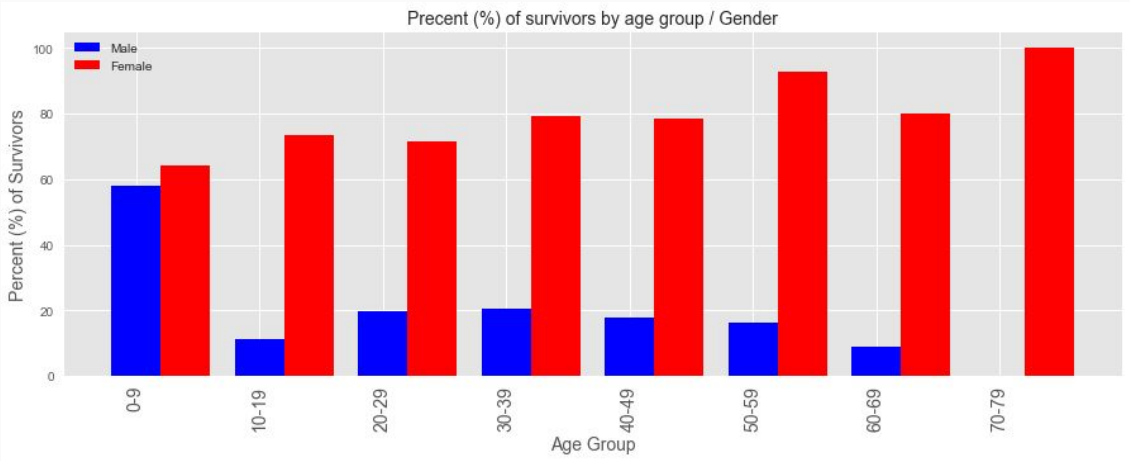
Correlation Heat Map



How did survival rates differ between elderly, adults and children?

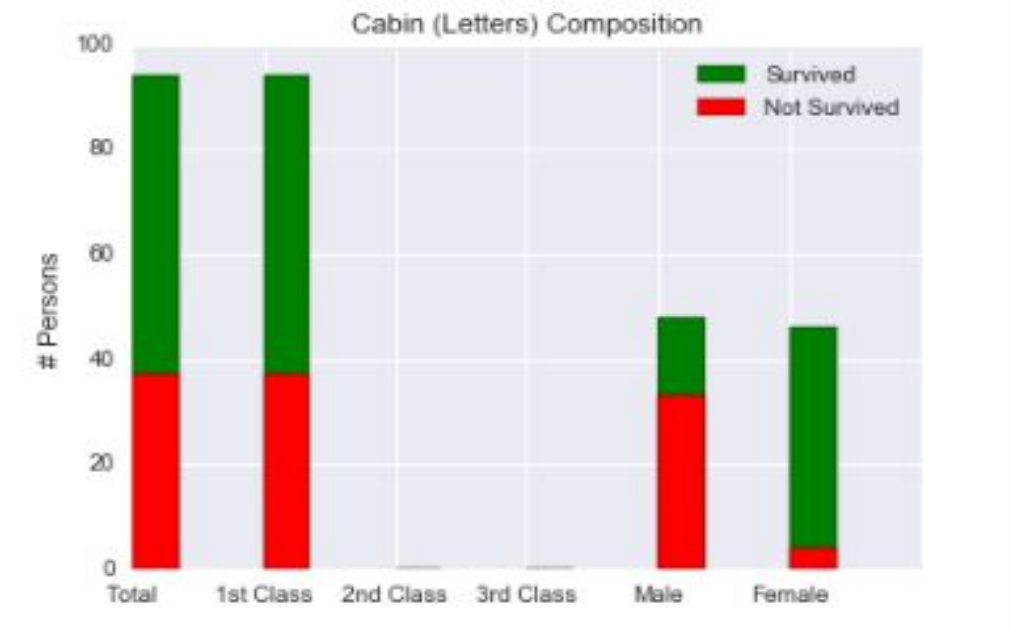
- Male Female children had the highest survival rates.
- Adults suffered the largest losses by count though their percentage was uniform.
- Elderly > 60yrs had a diminishing survival rate.





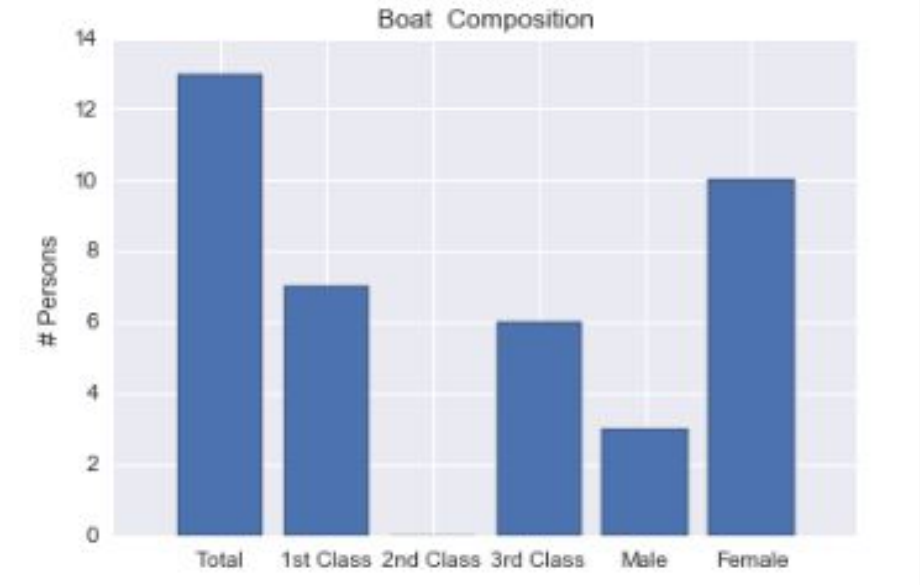
What were cabins like?

1. Creation of new variables, cabin letter, which groups cabins by their first letter (A, B, C, D, E, F, G, T). This roughly correlates to floor level
2. 1015 missg values, of which 693 come from 3rd class passengers
3. Interactive plot to show composition of cabin_letter by class and sex



Who was in the lifeboats?

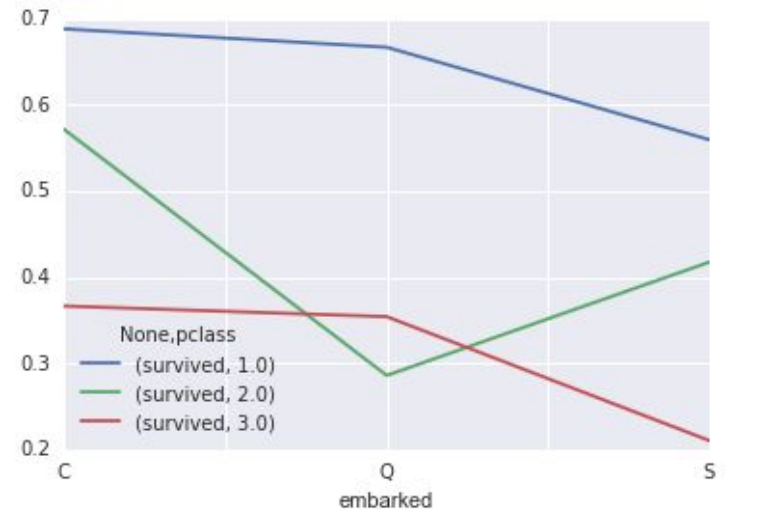
- 824 missing values
- Almost everyone who was in a boat survived.
- Only 9 people recorded in boats died.
- 23 people don't have a boat record but survived. Possible no-shows?
- Avg person per boat
-



Did points of embarkment affect survivorship?

Passengers embarking from Cherbourgh had the highest survival rates across all classes.

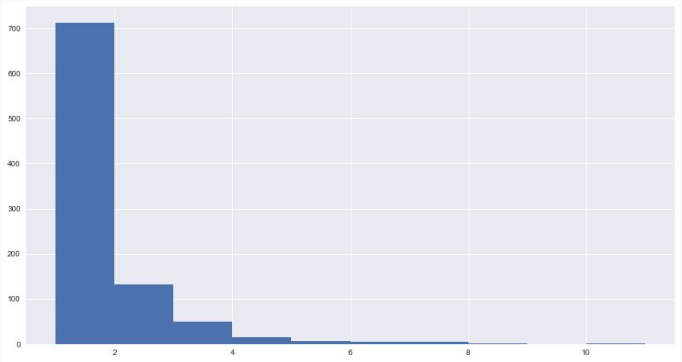
		survived		
		count	sum	mean
pclass	embarked			
1.00	C	141	97.00	0.69
	Q	3	2.00	0.67
	S	177	99.00	0.56
2.00	C	28	16.00	0.57
	Q	7	2.00	0.29
	S	242	101.00	0.42
3.00	C	101	37.00	0.37
	Q	113	40.00	0.35
	S	495	104.00	0.21



Did the group or family size affect survival rate?

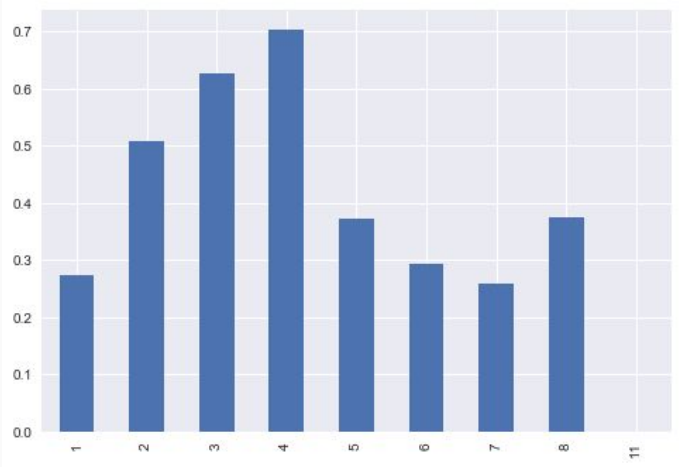
- Potential optimal group size (4)
  - Can assist one another
  - Can fit into a lifeboat

- Not likely that all members of a large group would survive
- But difference between small groups (2-4) and large groups (>4) is notable



Distribution of Group Sizes by Ticket ID Grouping

Survival Rates of Each Group Size

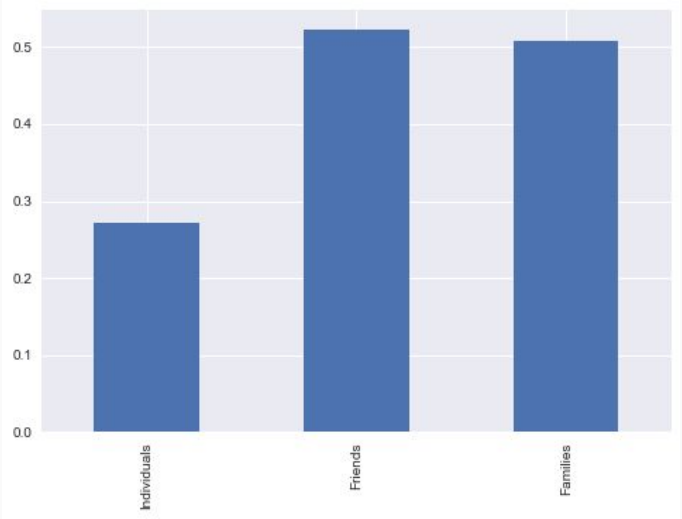


Did families have a higher survival rate than other group types?

- Difficult to determine which groups were families and which groups were not
- Being in a group does suggest a higher survival rate

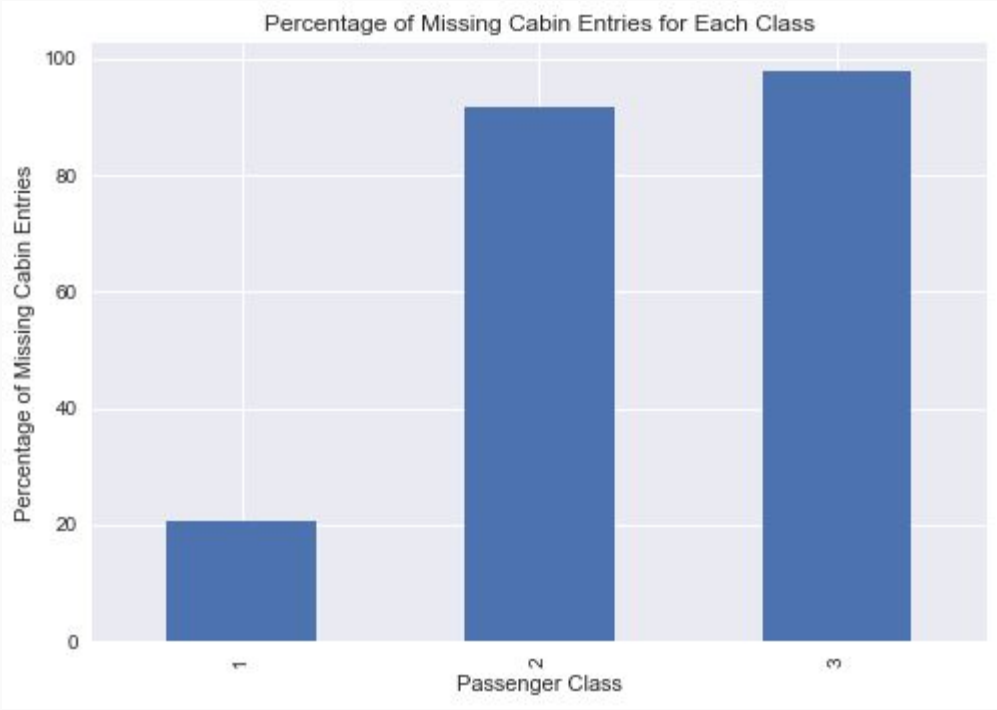
Data ID	Name	Ticket
0	Allen, Miss. Elisabeth Walton	24160
180	Kreuchen, Miss. Emilie	24160
193	Madill, Miss. Georgette Alexandra	24160
238	Robert, Mrs. Edward Scott (Elisabeth Walton Mc...	24160

Survival Rates of Different Group ‘Types’



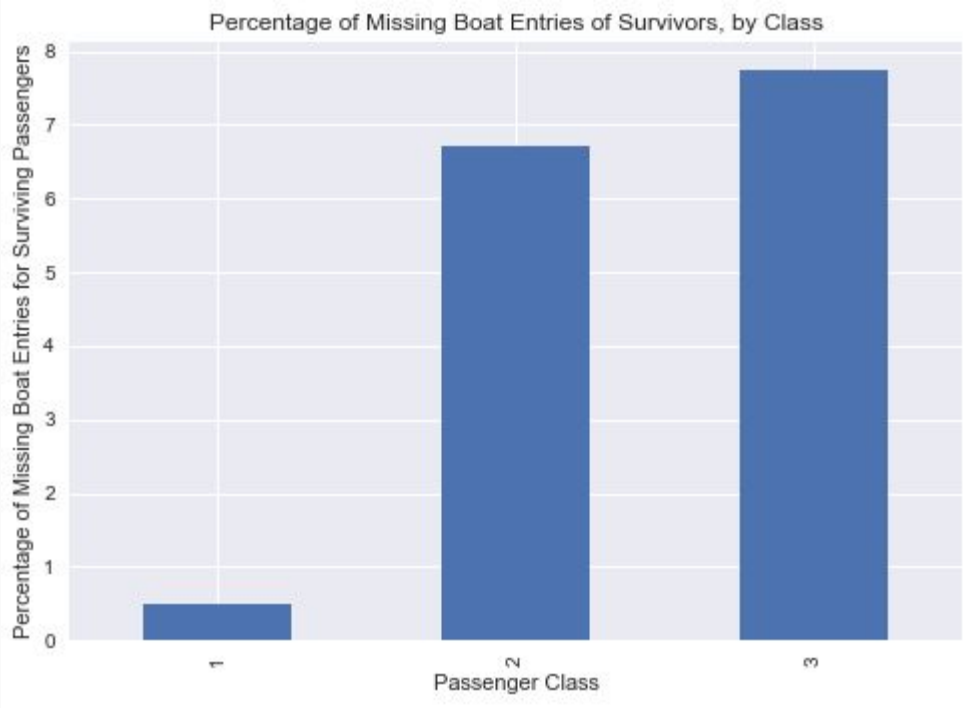
Was there less documentation for lower class passengers?

- Missing cabin numbers



Was there less documentation for lower class passengers?

- Missing boat numbers



Was there less documentation for lower class passengers?

- Missing body IDs

Was there less documentation for lower class passengers?

- Many data integrity issues
- Lower class passengers tended to have more missing information
- Was information about wealthy people considered more important?
- Is there a bias in data surrounding poorer people in the 1920s?

Percentage of missing data in each variable, by passenger class

class	age	fare	cabin	embarked	home.dest	boat (if survived)	body (if dead)
1	12.07%	0.00%	**20.74%**	0.62%	10.53%	**0.50%**	**71.54%**
2	5.78%	0.00%	**91.70%**	0.00%	5.78%	**6.72%**	**80.38%**
3	29.34%	0.14%	**97.74%**	0.00%	72.50%	**7.73%**	**89.58%**

Summary

- Don’t be a single, poor, adult, man from Southampton on the Titanic!!!

Also-- use tabs, not spaces.

Thanks for a great semester

Python is king of the world!