

# Exploring Titanic Passenger Data

**Demographics, Class, and Survival Analysis**

## Introduction

The dataset contains information about the passengers aboard the Titanic. This historical dataset provides details on various aspects such as gender, age, passenger class, and survival status. The primary goal is to gain insights into the demographics of the passengers, their distribution across different classes, and the survival rates within these categories.

**Problem or Questions:**

---

## Passenger Makeup:

- What is the overall distribution of passengers based on gender, age, and passenger class?
- How do these demographics contribute to the overall passenger makeup on the Titanic?

## Survival Analysis:

- What is the percentage of passengers who survived, and how does it compare to the overall percentage?
- How did survival rates vary across different genders, age groups, and passenger classes

# Visualizations

---

# Visualizations

## Overall Passenger Makeup:

- **Bar Charts:** Display the percentage distribution of passengers based on gender, age groups, and passenger class. This will provide a clear visual representation of the overall composition of the Titanic's passengers.
- **Scatter Plot:** Comparing passenger age to fare price. See if there is any correlation. Also try to spot the outliers for either category and there is the highest concentration.
- **Map:** Show from which countries the passengers were from.

## Survival Analysis:

- **Box plot:** Illustrate the mean survival age for each gender, and passenger class, highlighting the median age of survivors and non-survivors within each category.
- **Survival Rate Comparison Plot:** Create a bar plot to compare the survival rates of different demographics against the overall survival rate, allowing for a quick visual assessment of the impact of factors such as gender, age, and class on survival.

# Ethnic concern

# 1

Initially I had another titanic dataset with more columns but not all passengers. It only had 1300 passengers and I was concerned that the data was not random and unbiased. The surviving passengers probably have a greater chance of being in this dataset as more info exists of them. While the more famous people who died were also probably recorded. The ones without record I'm assuming are the poor people who died.

So in order to get unbiased results I used data with less columns but all passengers. The full list doesn't have info like cabin or lifeboat passengers were in their hometown and more. But this extra info won't necessarily make a difference to what I want to find out and it just simplifies stuff.



# Data Cleaning

---

Dataset has 11 columns and 2207 rows.

# Data Fields

**Name** - Passenger name, object passenger name stored as a string

**Gender** - Has two unique values either male or female for each person stored as a string

**Age** - how old a person was stored as an integer float64 as are some babies on board under 1.

**Class** - which class ticket the passenger had. There are 3 classes as well as 4 crew groups. String

**Embarked** -location where embarked on the ship. Stored as a string and have 4 unique values. S for Southampton, England, C for Cherbourg, France, Q for Queenstown, Ireland and B for crew members

**Country**- String, Passenger country of origin. Passengers were from 49 different countries.

**Ticketno** - ticket number stored as integer float64. I will drop this column. Not useful info and too many null values.

**Fare** - the price that was paid for the ticket in pounds. Is an integer.

**Sibsp** - integer float64 if had sibling on board 0 for none and how many siblings on board if had

**Parch** - integer float64 if parent or child is also on board.

**Survived** - String either yes or no if the passenger survived or not.

# Null values

sibsp 900

parch 900

ticketno 891

fare 916

Age 2

Country 81

- If they had a sibling or parent on board has a lot of null values and I won't be needing this info for the analysis I'm trying to do. So I will just drop these two columns.
- Ticket numbers will drop as well.
- For fare will check passenger class and where embarked and use that mean value.
- Age has 2 na's. Will drop na for age and
- Country has 81 null values. Will change null to unknown.

Most with na values for fare were crew members. Once those were filled with crew since they obviously didn't pay a fare there are only 17 nulls for fare.

Would also need to know which cabin stayed in for more accuracy but since don't have it will just go with mean for that class and embarked location.

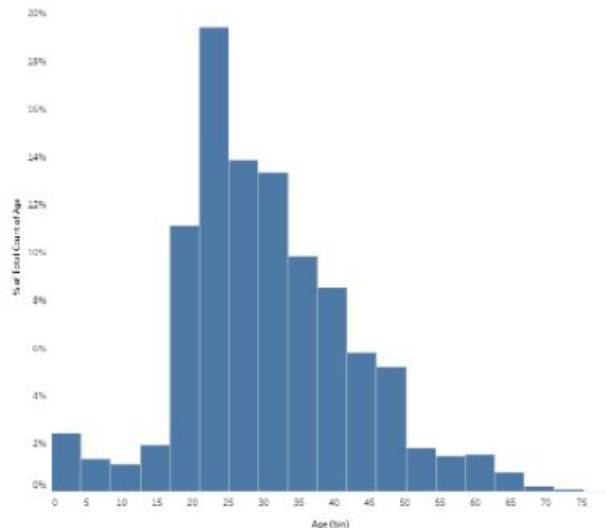
**Visuals**

---

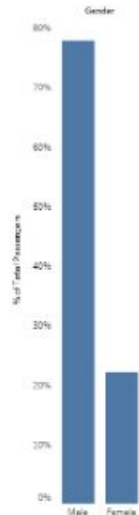
# Bar charts

- Set quick table calculations as percent of total passengers
- Set title of graphs
- Set y axis label
- Removed gridlines
- Set text color to black
- Sorted so larger column appears first
- For age distribution histogram did formula for number of bins and changed bin size to 4.18

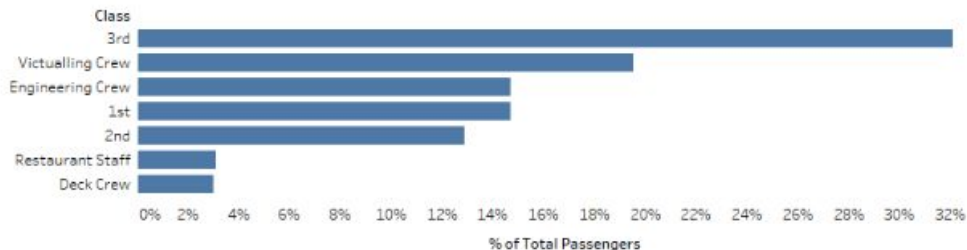
Distribution of Passengers by Age



Distribution of Passengers by Gender



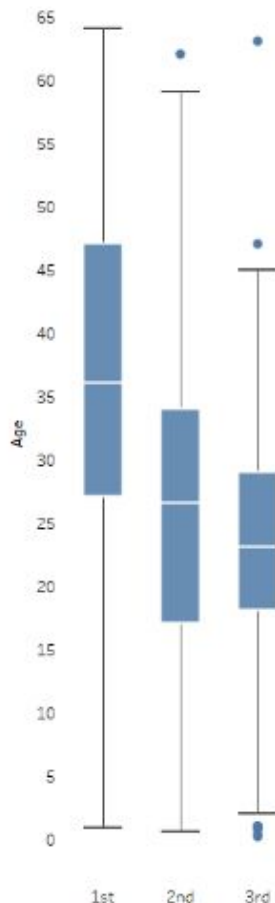
Distribution of Passengers by Class



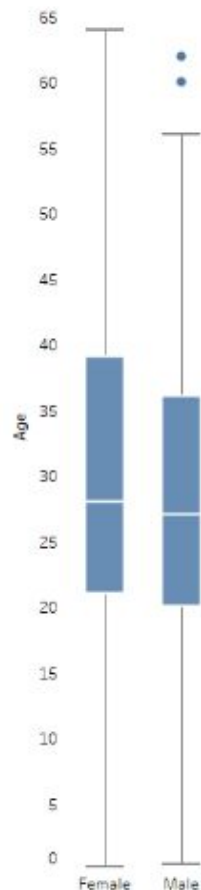
# Box plots

- Filtered by passengers survived
- Removed crew from list since the median age of crew was higher and mostly male to begin with
- Set to only show dots outside of whiskers
- Removed borders
- Removed gridlines
- Removed column headers since redundant

Mean Survival Age by Class

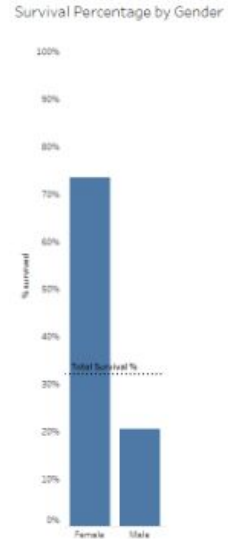
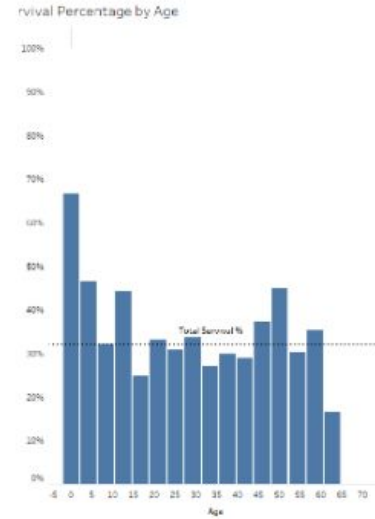
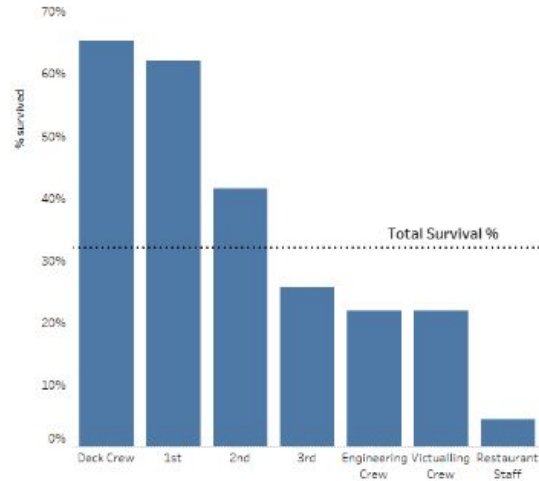


Mean Survival Age by Gender



## Survival Rate Comparison Plots:

- Removed grid lines
- Set to percent of total and compute using cell
- Added reference line for total percent survived
- Hid field label for column
- Filtered survived column by color and set not survived color to white





# Ethnic concern

# 2

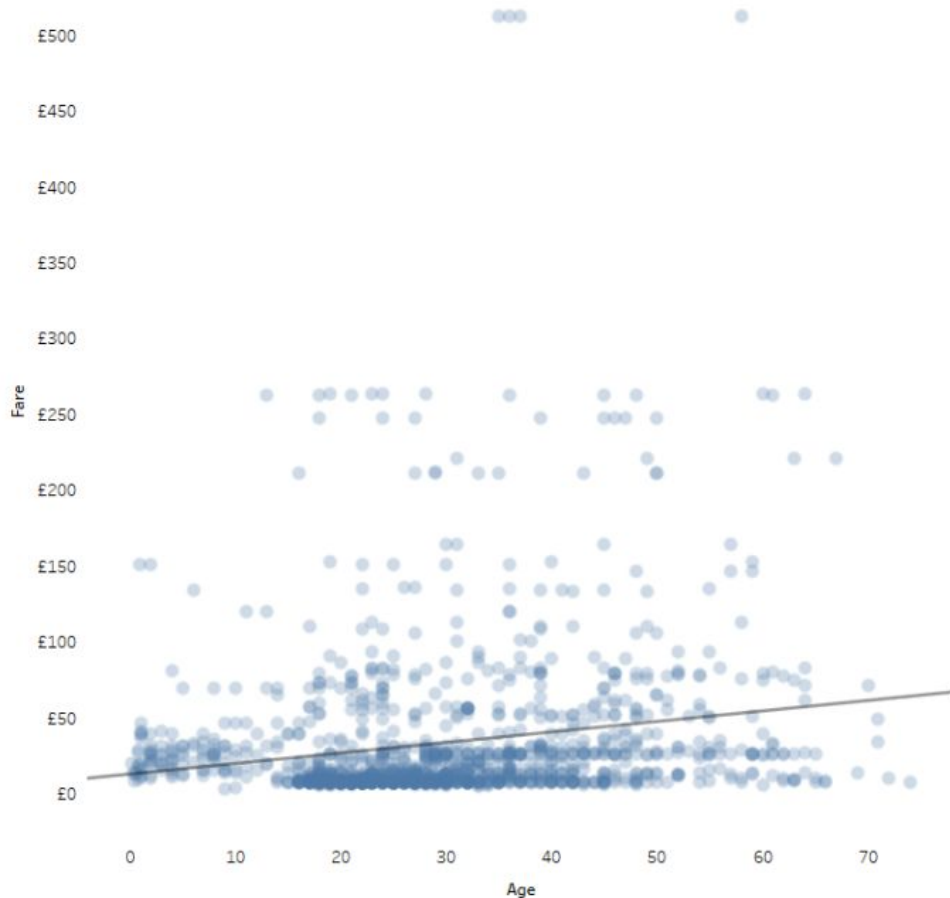
Initially I set survival for each category as the count or percentage of total survivors how many were from each class. But that wasn't an accurate representation because some classes had more people than others so of course more survivors.

I then switched it to the percentage of that class. From each class's total passengers what percentage survived. This is compared to the percentage of total passengers who survived which is around 32%

## Scatter plot

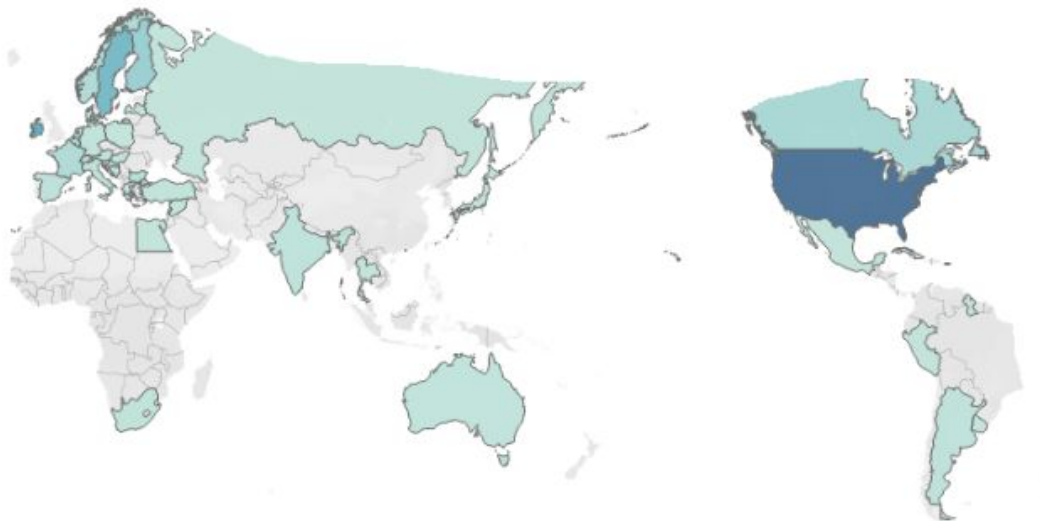
- Removed gridlines & chose black for text color
- Filled circle and set opacity lower to see concentrations
- Set the trendline
- Added currency sign to fare

Fare / Age Scatter Plot



## Geo-plot

- Changed color so that it ends at 265 the largest number for one country and not total of all so that different extreme values are more noticeable
- Chose a color palette that fits with the rest of the graphs
- Turned on tooltip for country and count so when hovering know which country and the number of passengers from that country



# Dashboard

---



Mr Francis Parkes, known as Frank. He was an apprentice plumber and worked by Harland & Wolff shipyard. He was 5ft 8½in tall, with brown hair and was missing the 1st finger of his right hand. Frank was aboard the *Olympic* for her maiden voyage arriving in New York on 21 June 1911

Frank repeated the voyage in April 1912 when he joined the nine-strong "[guarantee group](#)" of Harland and Wolff employees chosen to oversee the smooth running of the *Titanic's* maiden voyage.

### **What was Franks chance of survival when when the ship sank?**

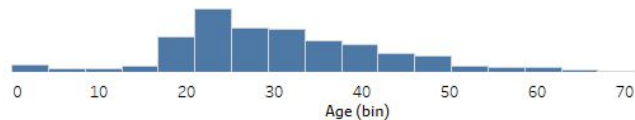
Frank, like the rest of his counterparts, died in the sinking. Though the nine-member guarantee group were given passenger accommodation, they were also regarded as members of the crew.

Most of the engineering crew remained below decks in the engine and boiler rooms: some fighting a losing battle to keep the ship afloat by operating the pumps and others keeping the generators running to maintain power up until two minutes before the ship sank. It is speculated that their actions delayed the sinking for over an hour and helped keep the ship afloat long enough for nearly all the lifeboats to be launched.

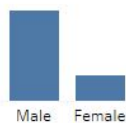
Total Onboard

2,205

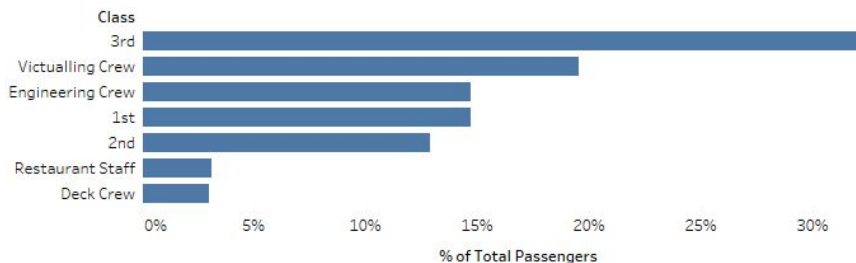
Distribution of Passengers by Age



by Gender



Distribution of Passengers by Class



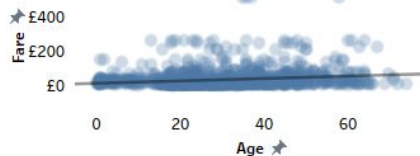
Passenger Country Map



© Mapbox © OSM

Count 1 265

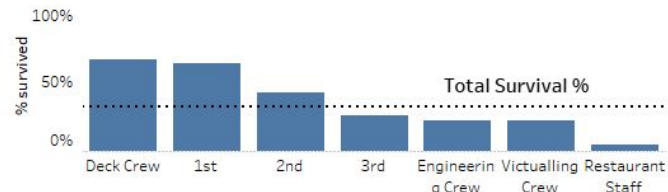
Fare / Age Scatter Plot



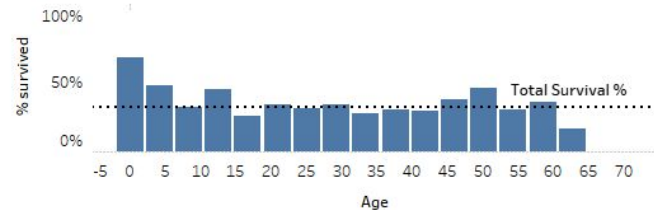
Drowned 1,494

Survived 711

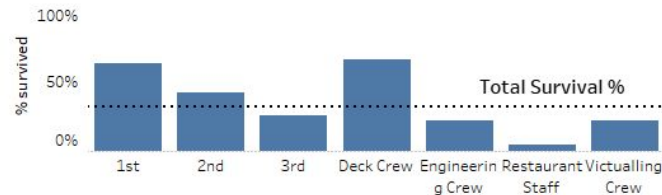
Survival Percentage by Class



Survival Percentage by Age



Survival Percentage by Gender



Mean Survival Age by Class



Mean Survival Age by Gender

