



# Analyzing the Titanic Dataset with a Business Analyst mindset

Using R(ggplot2)

Presented By: Aman Jain



# About Dataset

The survival status of the 1309 passengers who travelled on the RMS Titanic's inaugural voyage in 1912 is covered in this data set (the ship's crew are not included)



**Question 1: What  
was the survival  
rate?**



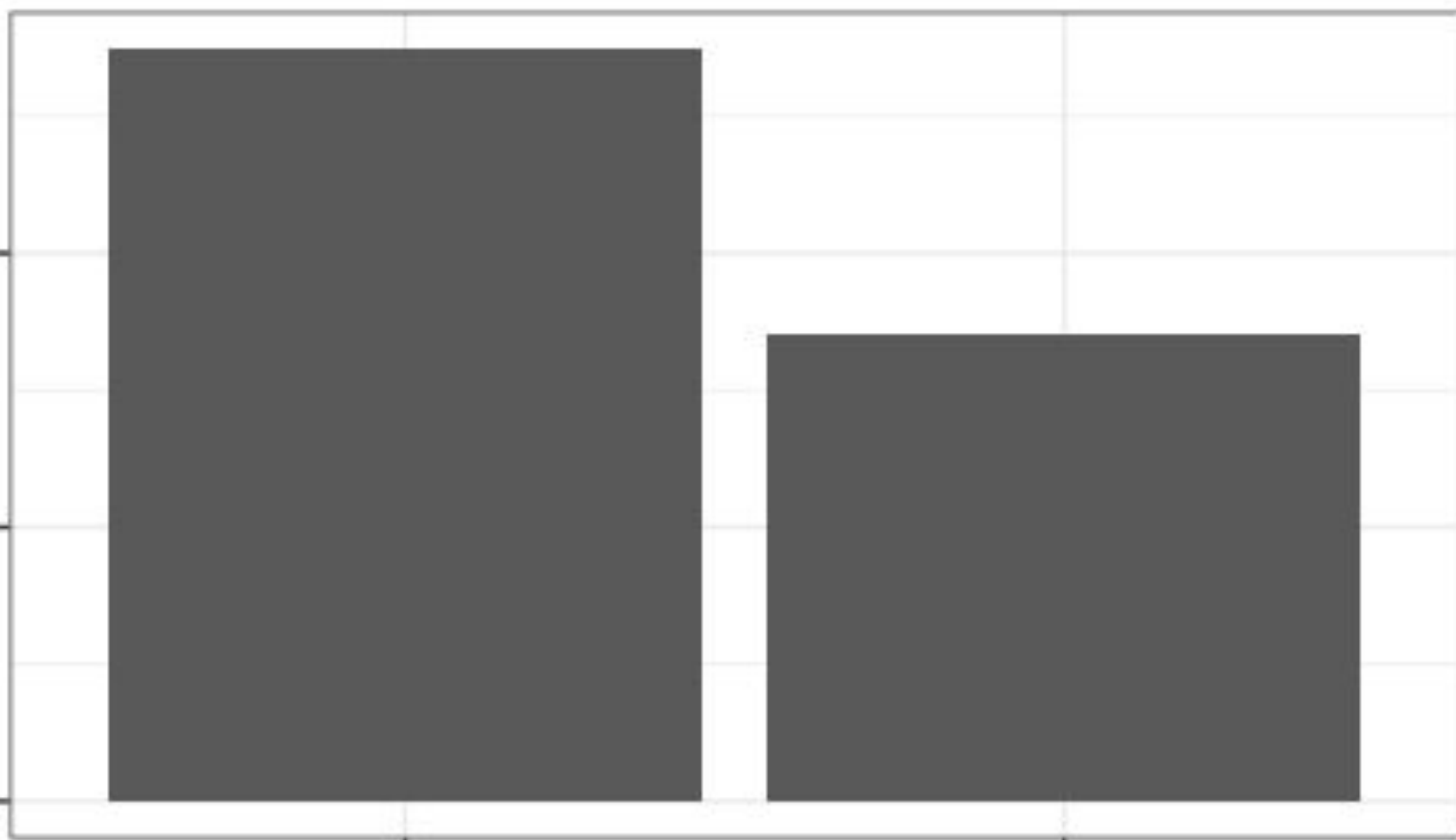
Passenger Count

400  
200  
0

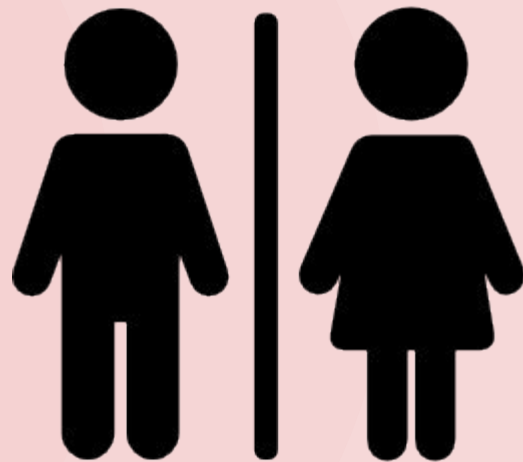
0

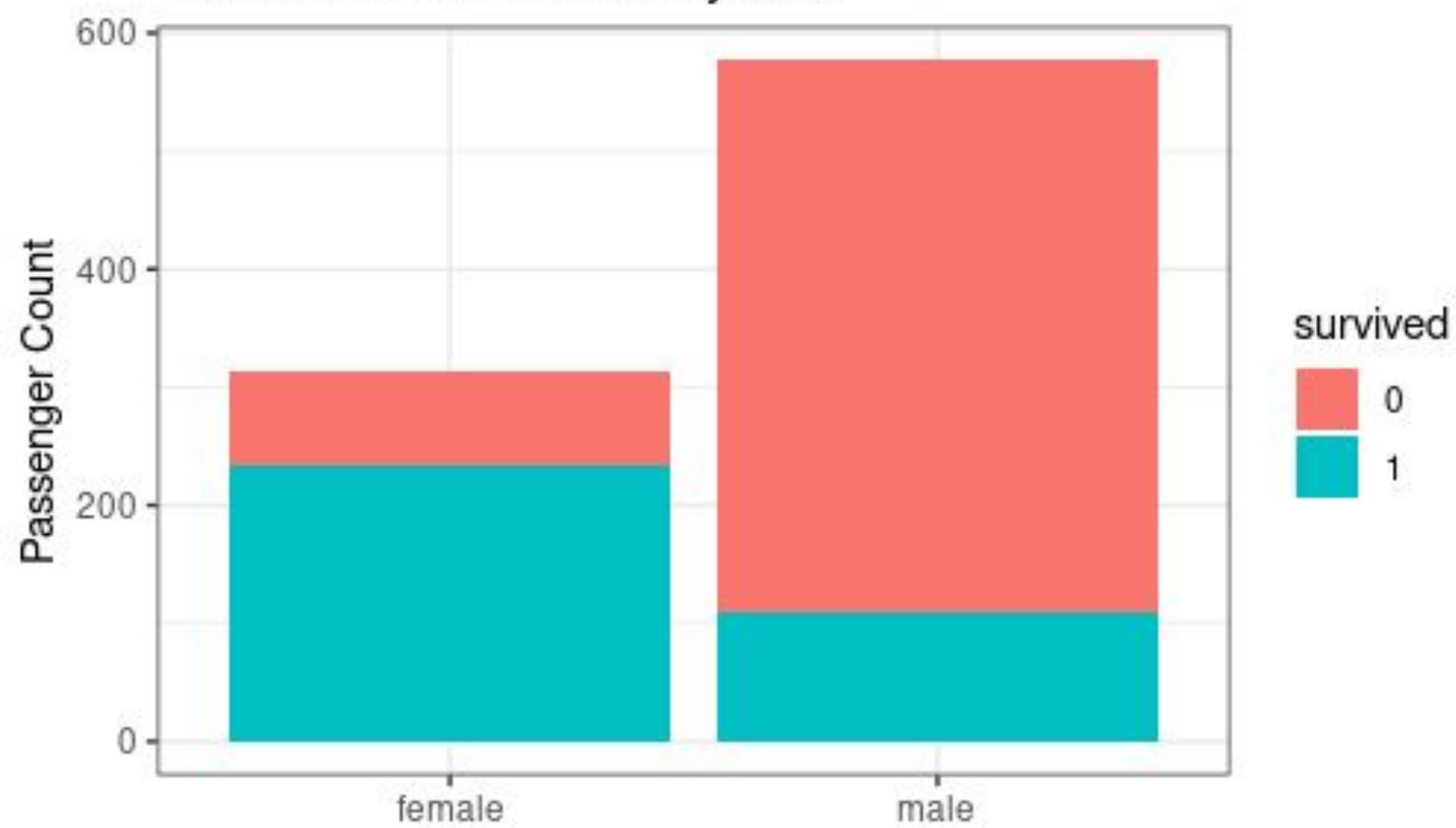
1

Age Group



**Question 2: What  
was the survival  
rate broken  
disaggregated for  
male and females?**

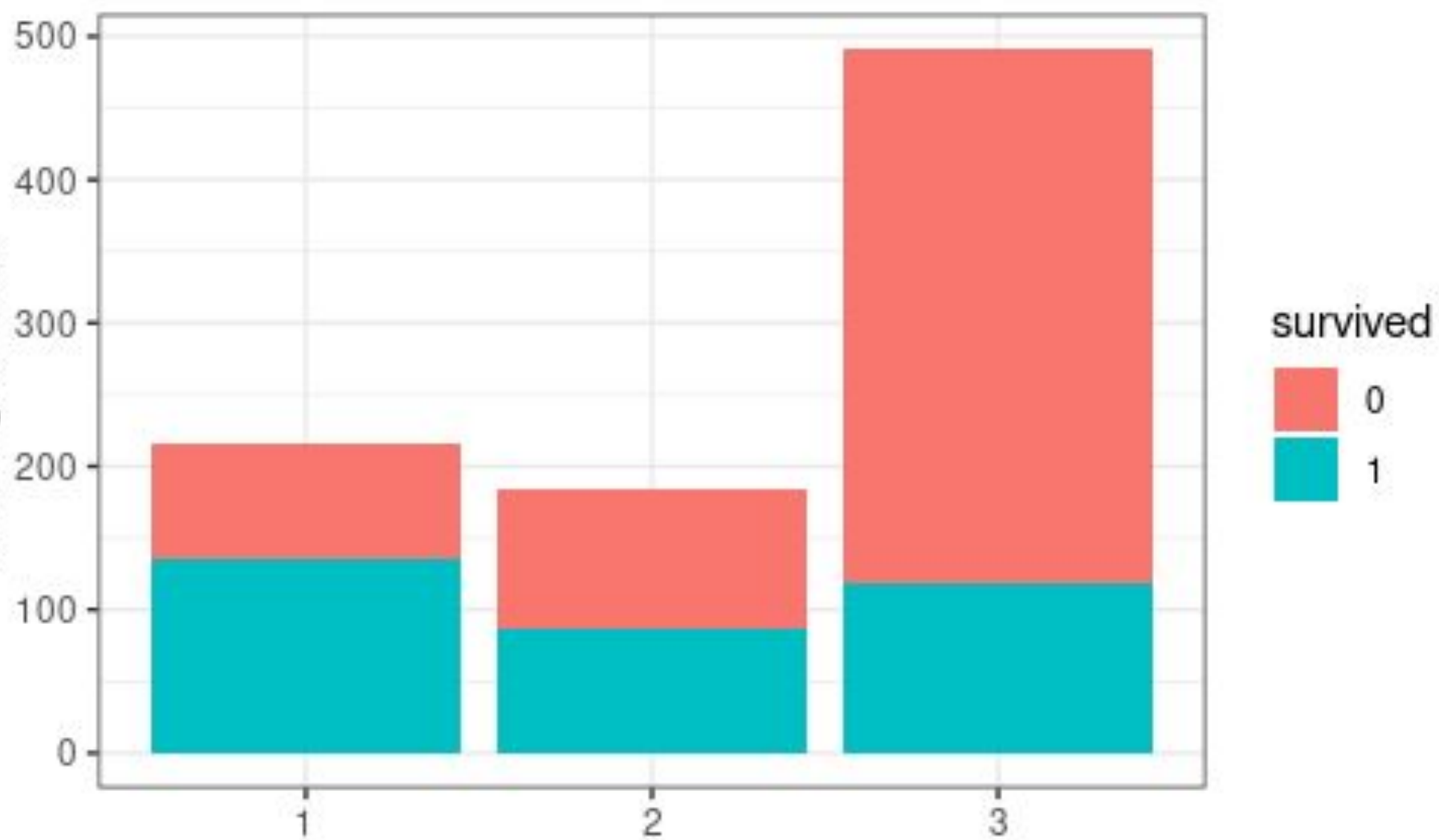




**Question 3: What was the survival rate by passenger class marked on ticket?**



Passenger Count

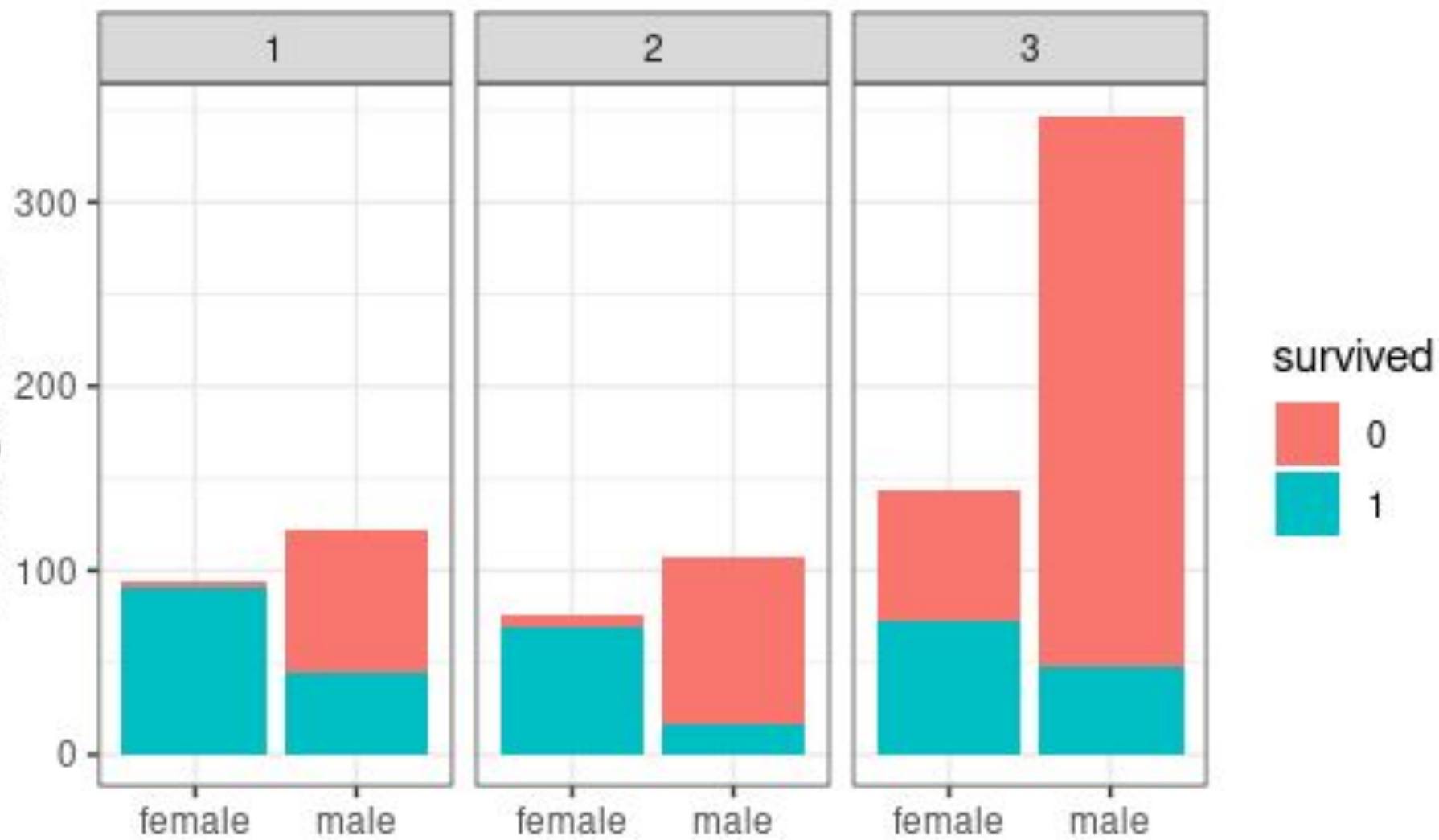




**Question 4: What  
was the survival  
rate by class of  
ticket and sex?**



Passenger Count

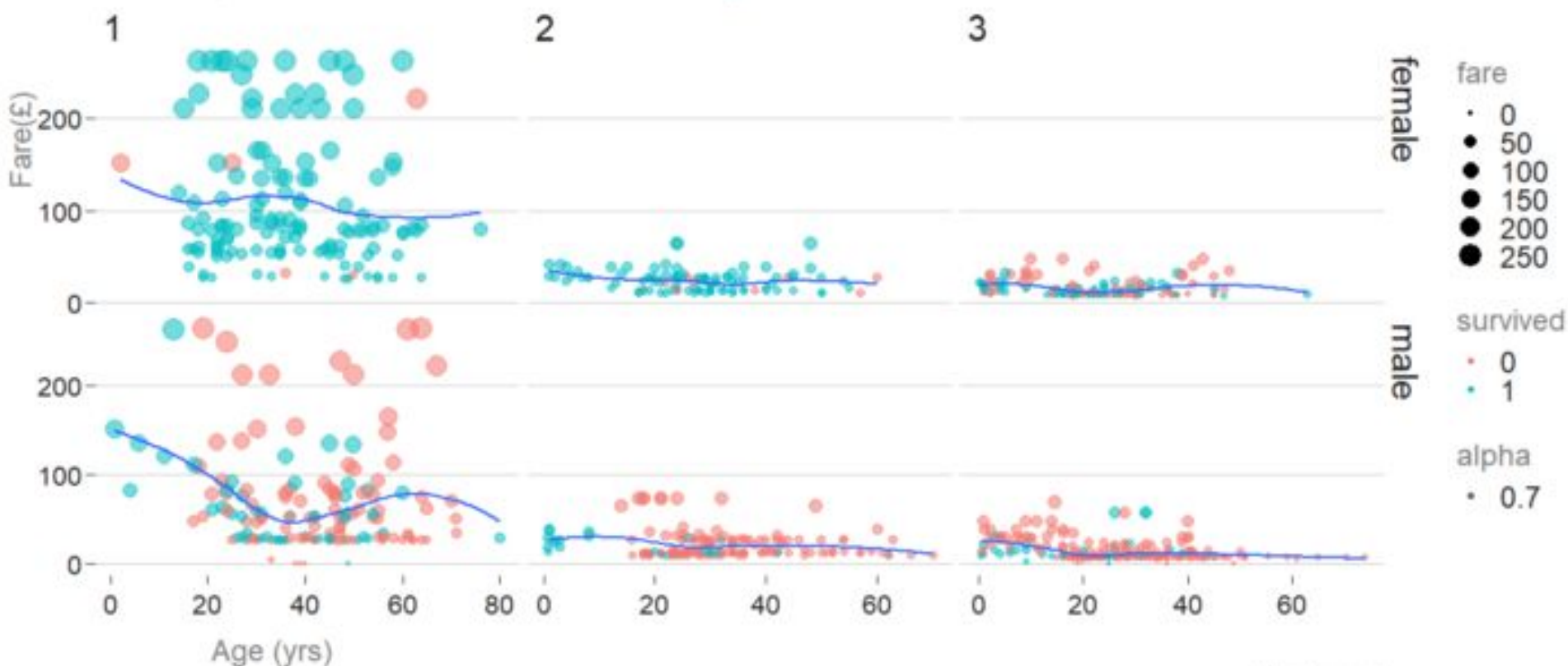


**Question 5: Is the initial strategy used by rescuers on the Titanic to focus on women and children?**



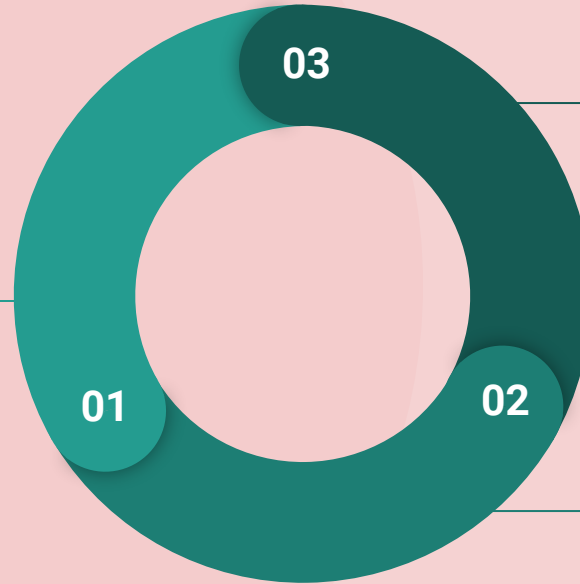
# Priority and pattern of rescue on the Titanic

Children and women in order of ticket class were considered first in the rescue plan with priority been women, children and older adults  $\geq 60$  yrs



## RESULTS ANALYSIS FOR QUESTION 5

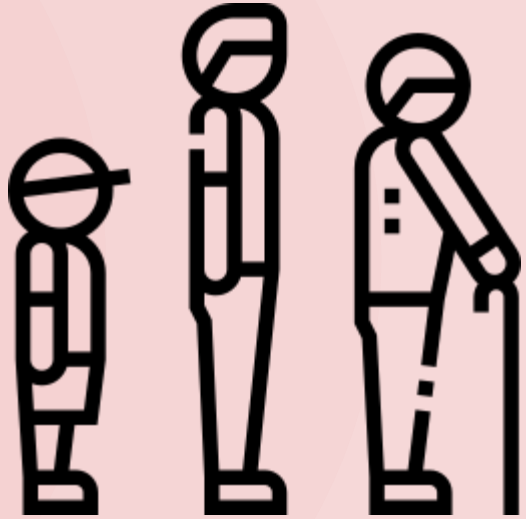
It seemed like the fares for children and teens seem unusually high when compared with average fares for non-children age groups.



Apparently little or no priority was given to male passengers by rescuers

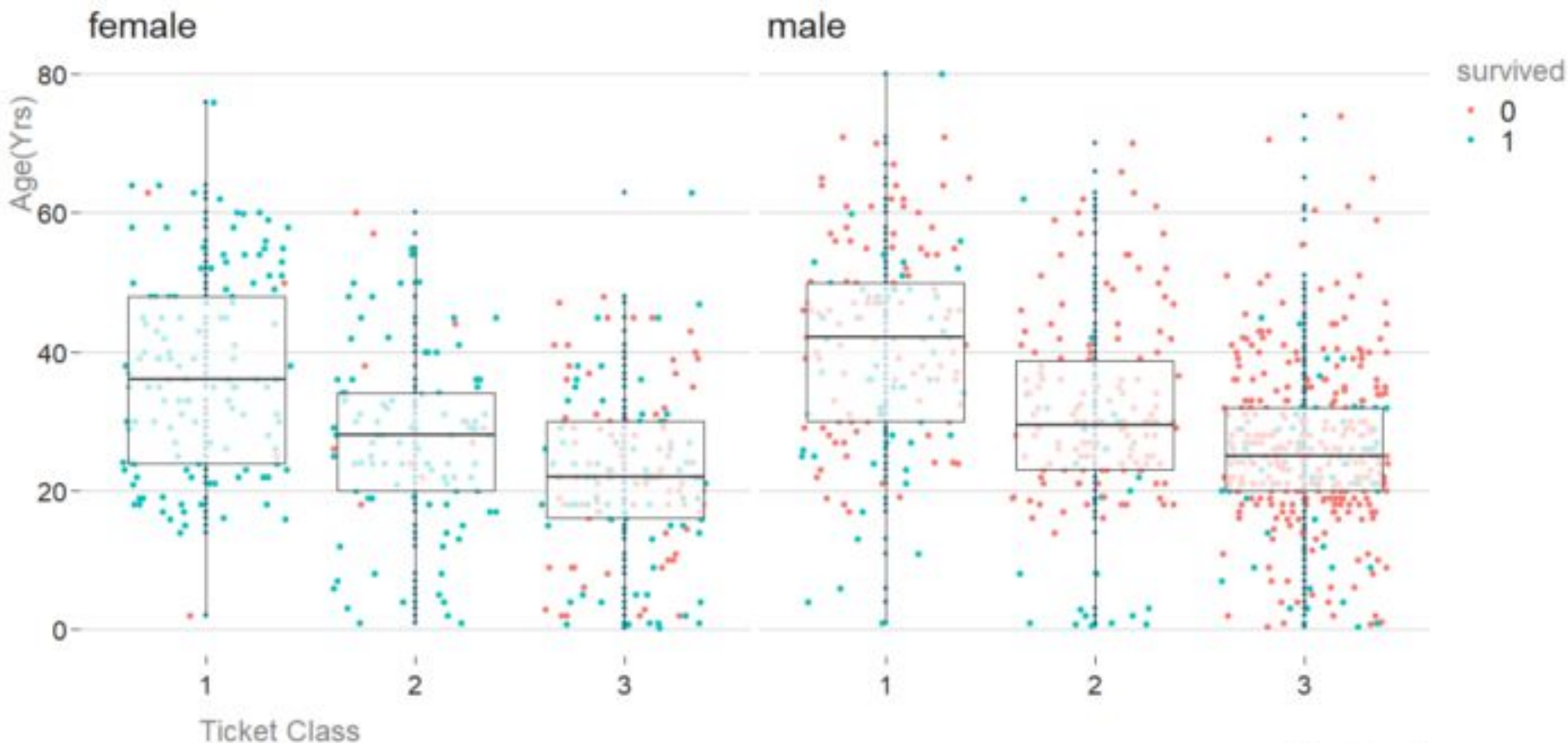
Evidently, Women and Children first approach in order of ticket class was adopted in the rescue plans by rescuers with priority been women and children and older adults at least 60yrs across all classes.

**Question 6: What  
is the distribution  
of passenger  
ages?**



# Age Distribution by Class on the Titanic

The males on the titanic were older than the females by an average of 3yrs across all ticket classes



## RESULTS ANALYSIS FOR QUESTION 6

### FEMALE

About 75% of the females in order of class (\*1st, 2nd, 3rd) were at least 22, 20 and 17 yrs old. The median age was 36 yrs (normally distributed), 28yrs (negatively skewed) and 22 yrs (positively skewed)

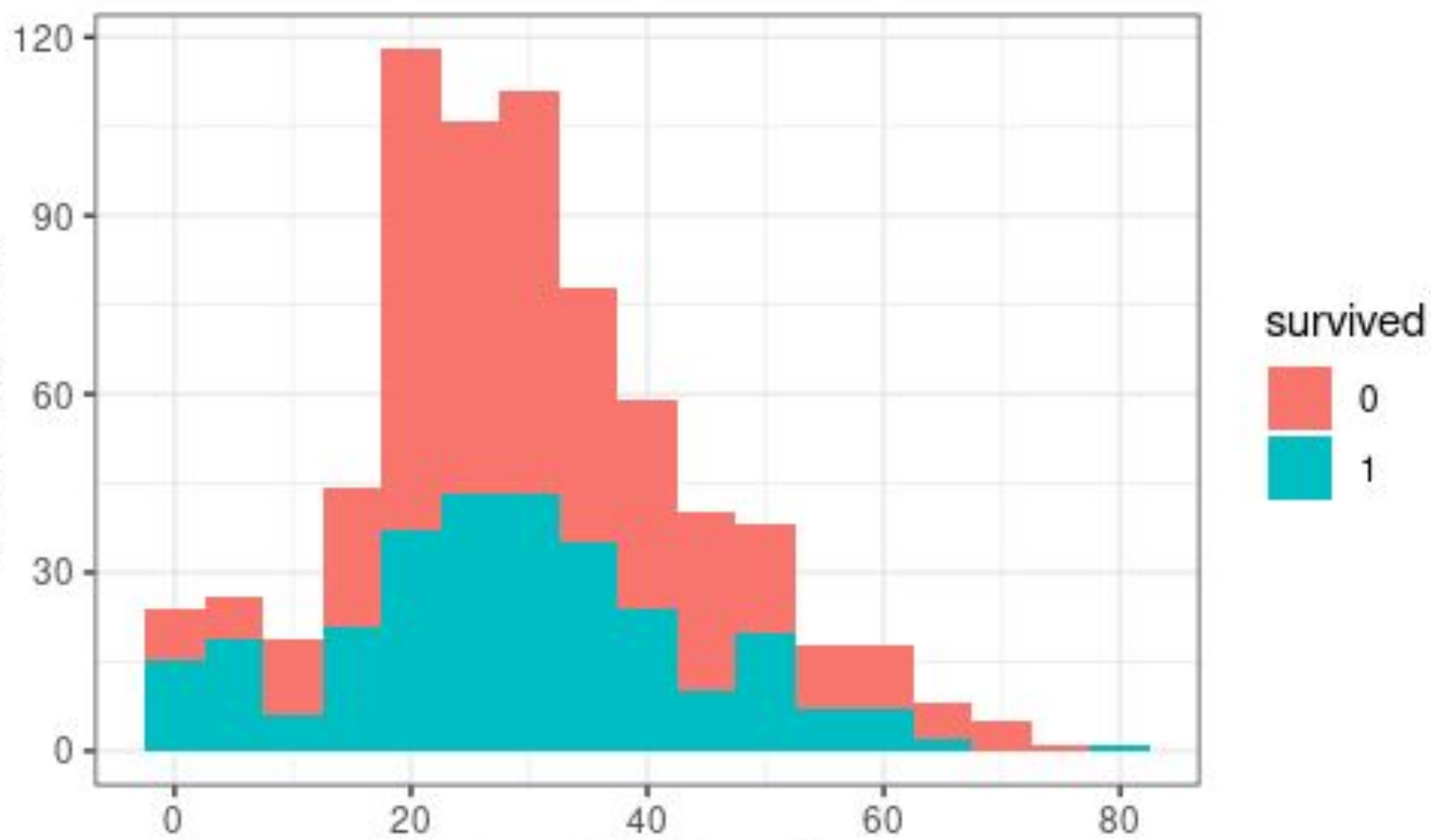
### MALE

About 75% of the males in order of class (\*1st, 2nd, 3rd) were at least 30, 24 and 20 yrs old. The median age was 42 yrs (negatively skewed), 30yrs (positively skewed) and 25 yrs (positively skewed)

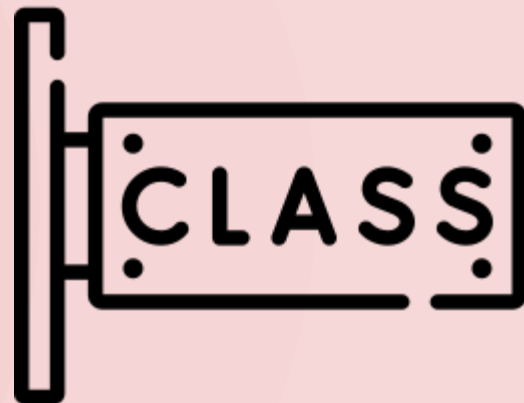
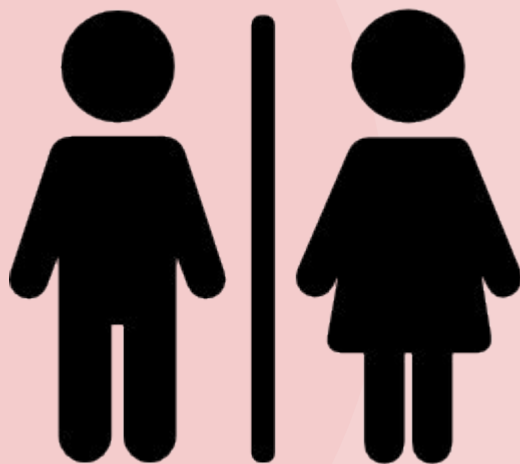


**Question 7: What  
are the survival  
rates  
disaggregated by  
age?**

Passenger Count

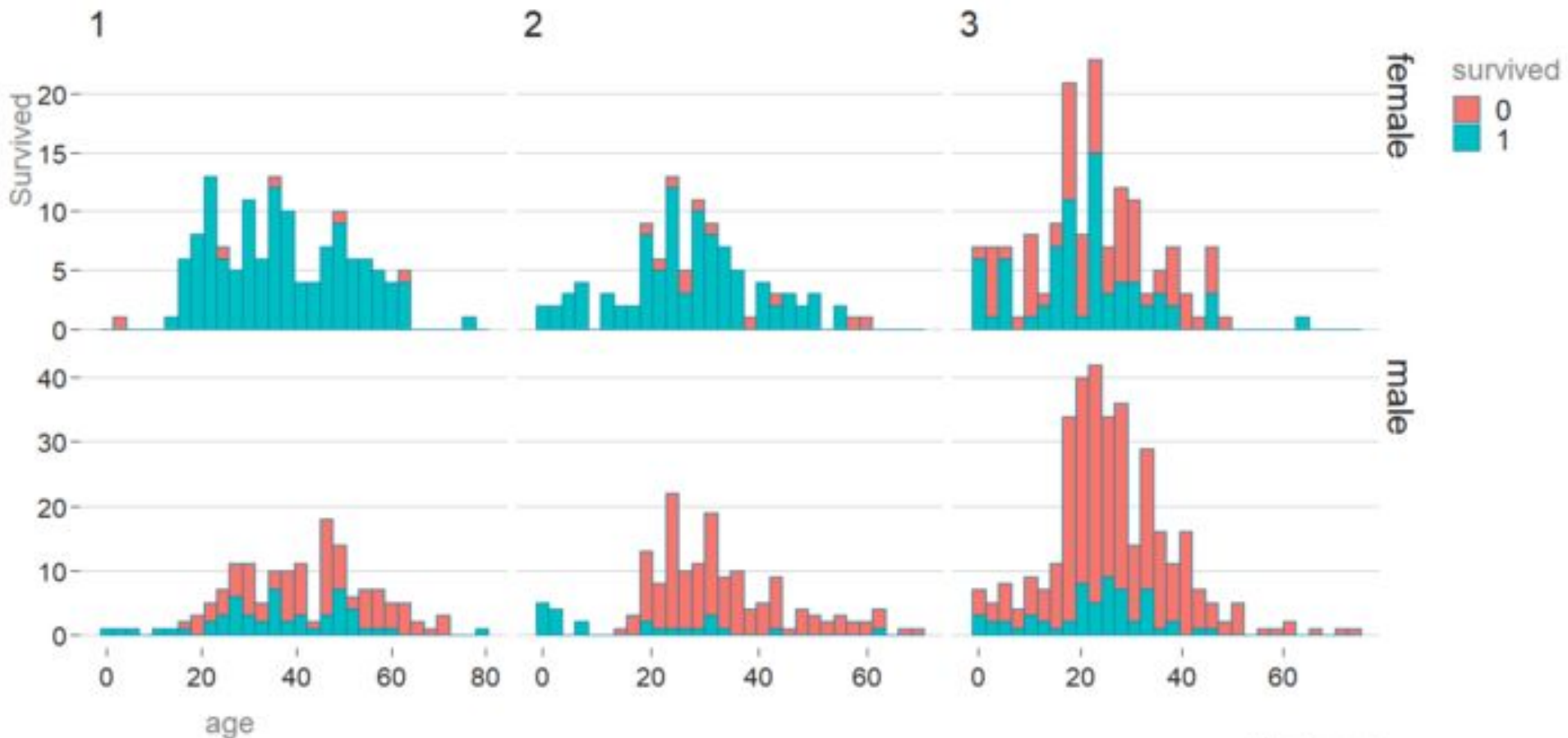


**Question 8: What is the survival rates by age when segmented by gender and class of ticket?**



# Survival rate on the Titanic

Distribution By Age, Gender and class of ticket



## RESULTS ANALYSIS FOR QUESTION 8

### 1ST Conclusion

In order of class (\*1st, 2nd and 3rd), the percentage of females that survived was 97%, 89% and 49%.

### 2ND Conclusion

In order of class (\*1st, 2nd and 3rd), the percentage of males that survived was 34%, 15(~14.6)% and 15(~15.2)%.

### 3RD Conclusion

Within 1st and 2nd class, all Children survived except one female child from 1st class. There were more children fatalities in 3rd class.

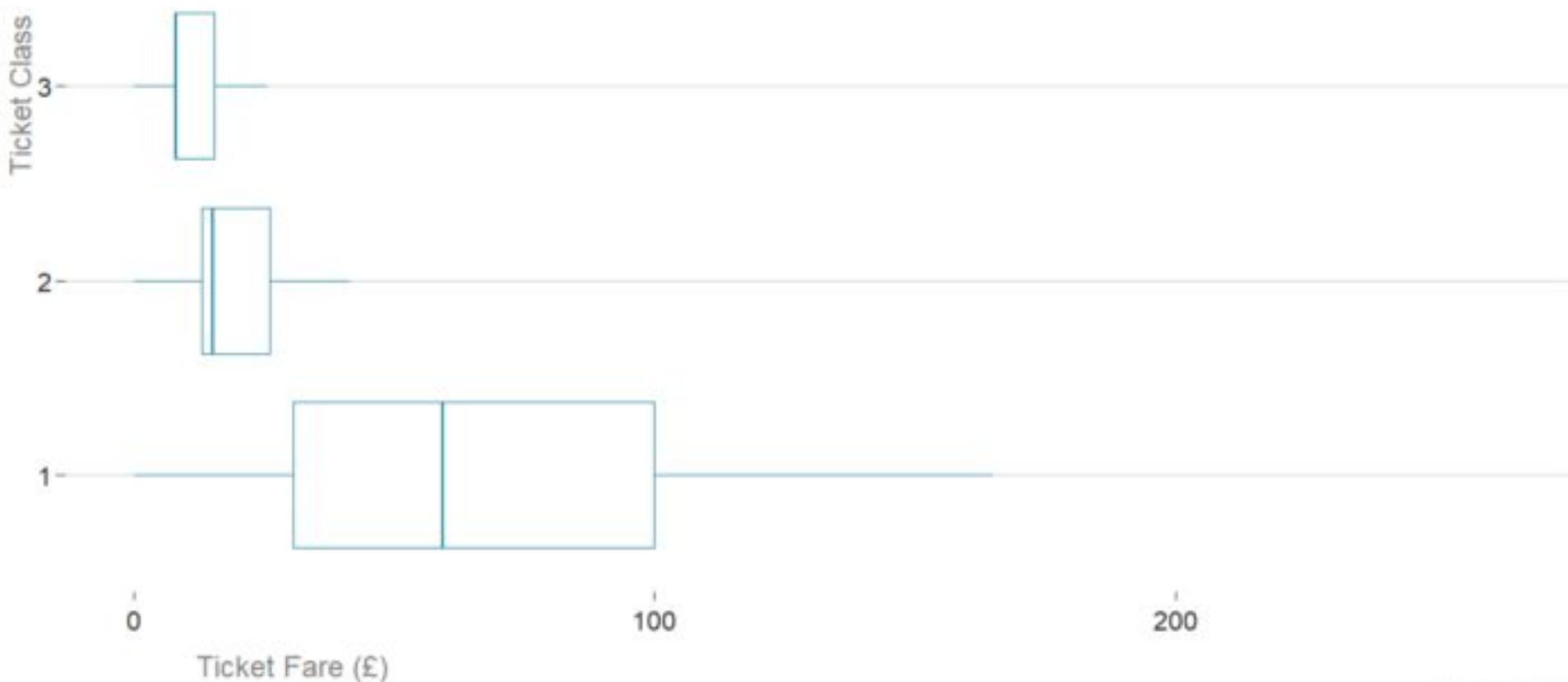


**Question 9: How  
does the fare value  
compare across all  
ticket classes?**



# Fare Value by Class

1st class ticket was worth 3 times a 2nd class ticket  
and 2nd class ticket was worth almost twice that of 3rd class





## RESULTS ANALYSIS FOR QUESTION 9

01

The distribution is skewed to the right. Mean fares are greater than median fares.

02

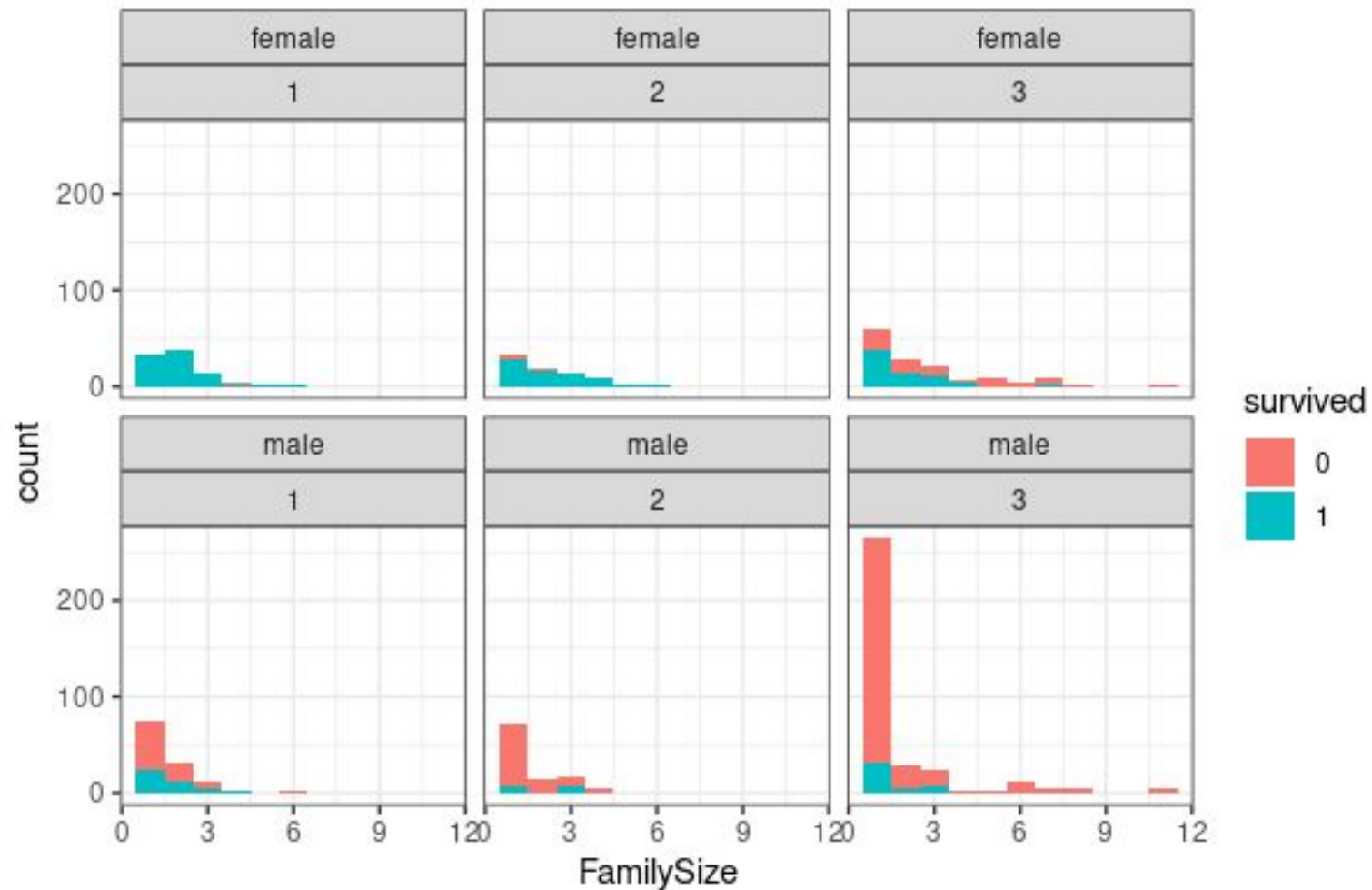
1st class ticket costs about 3 times a 2nd class ticket and 2nd class ticket was worth about twice that of 3rd class.

03

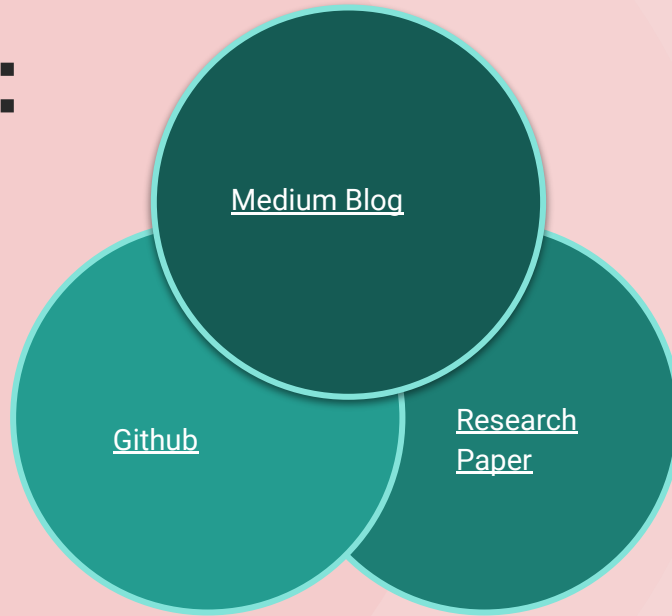
The average and median fare for children is higher when compared to that of adults in same class.

**Question 10: Is  
there a link  
between family  
size and survival**





# References:



**Thank you**

