INFO-633-002

INFORMATION VISUALIZATION

Assignment: Project A

Due Date: 01/31/2021

Team: Group-A3

Team Members

Anju Tyagi (at3358@drexel.edu)

Astha Jain (aj887@drexel.edu)

Madhu Bandru (mb4236@drexel.edu)

Vuthej Krishna (vv334@drexel.edu)

Description of the Dataset

The dataset that we have chosen for Project-A is 'Titanic's passengers' which is available in the sample datasets provided by the app.rawgraphs.io website. It revolves around the tragic journey of RMS Titanic which met with an accident on the early hours of April 15th, 1912 and later sank into the North Atlantic Ocean. This dataset talks about a portion of passengers that were aboard on RMS Titanic.

Current dataset consists of 1309 instances with around 15 different features. Amongst these 15 features, 8 are characterized as numerical and the remaining 7 as string data. Below mentioned is the list of features along with a short description about it.

1. **Class** – This feature talks about the class of the passenger's ticket.
2. **Survival** – It describes about the passenger's survival i.e. whether or not the passenger survived the tragedy.
3. **Name –** Passenger's name is conveyed under this feature.
4. **Gender –** It describes the sex of the Passenger i.e. Male or Female.
5. **Age –** Number of years since the birth of the passenger is conveyed here.
6. **Age Group –** Under this feature, age of the passenger is clubbed into different groups of size 10 starting from 0 until 99 i.e. 0-9, 10-19, 20-29 …. so on until 90-99.
7. **Siblings/Spouse Aboard –** This feature provides the count of siblings or spouse if any, that are travelling along with the passenger.
8. **Parents/Children Aboard –** Similar to the above feature, this too provides the count of parents/children if any, that are travelling with the passenger.
9. **Ticket Number –** This is the number allotted to the passenger upon purchasing a ticket.
10. **Fare –** This feature speaks about the price of the ticket that is being purchased by the passenger.
11. **Fare Group –** This field rounds the price of the ticket to the nearest ten.
12. **Cabin –** This is the room/cabin number that will be allotted to the passengers after purchasing a ticket.
13. **Port of Embarkation –** This is the source point where the passenger boards the RMS Titanic.
14. **Boat –** If the passenger survived the accident, this field conveys the information about the boat number by which the passenger was evacuated from the sinking RMS Titanic.
15. **Destination –** This is the end point of a passenger's journey and upon reaching, one should disembark.

Using this dataset, we would like to investigate the following points.

1. Is there a relation between the survival status of a passenger to the fare of their ticket?
2. Did the passengers age define their survival status?
3. Did the passengers gender play any role in securing one's life while evacuating?

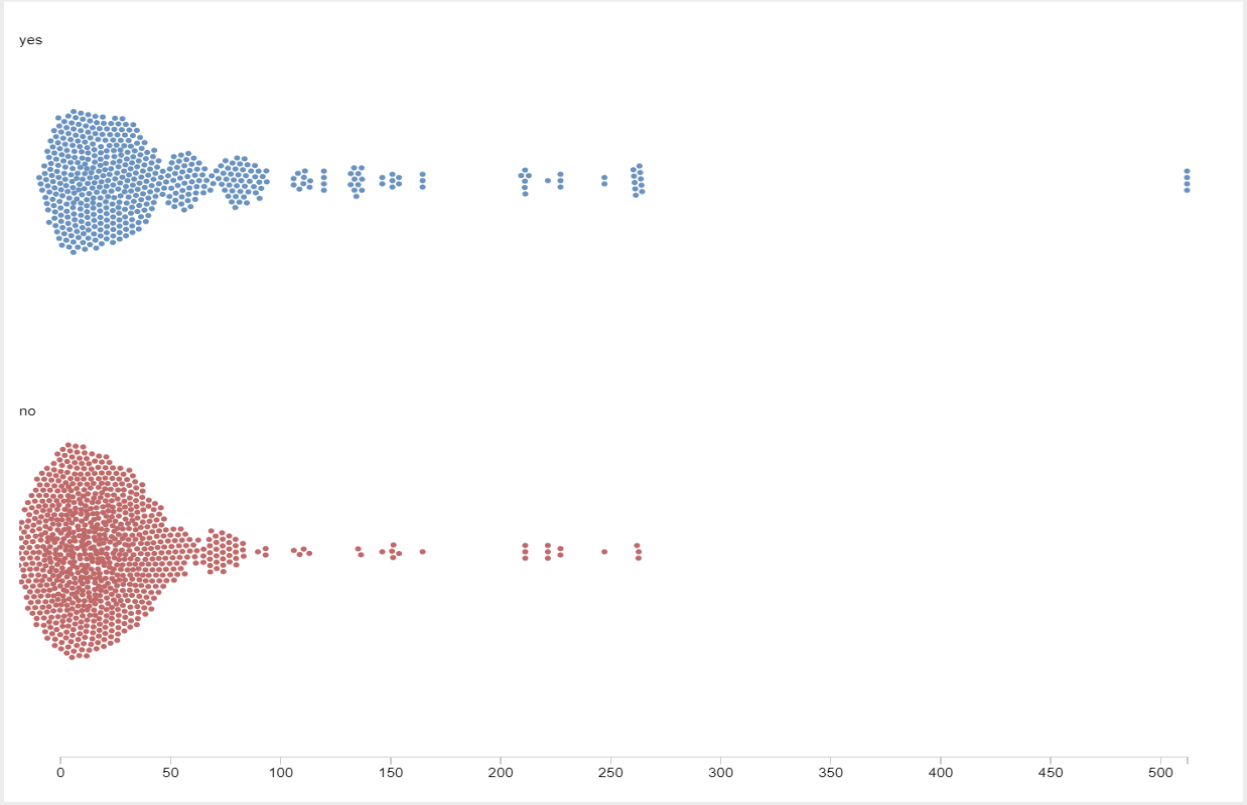

Figure 1 – Survival Status Vs Ticket Fare

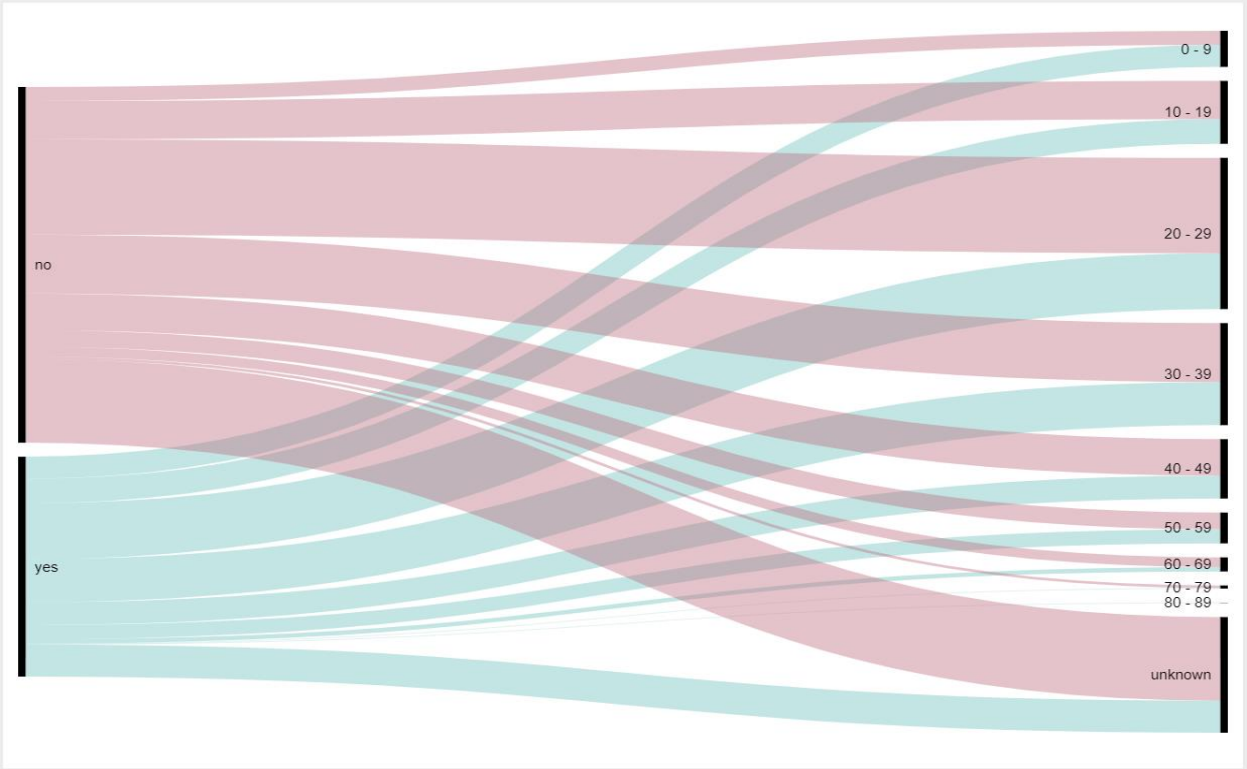

Figure 2 – Age Group Vs Survival Status

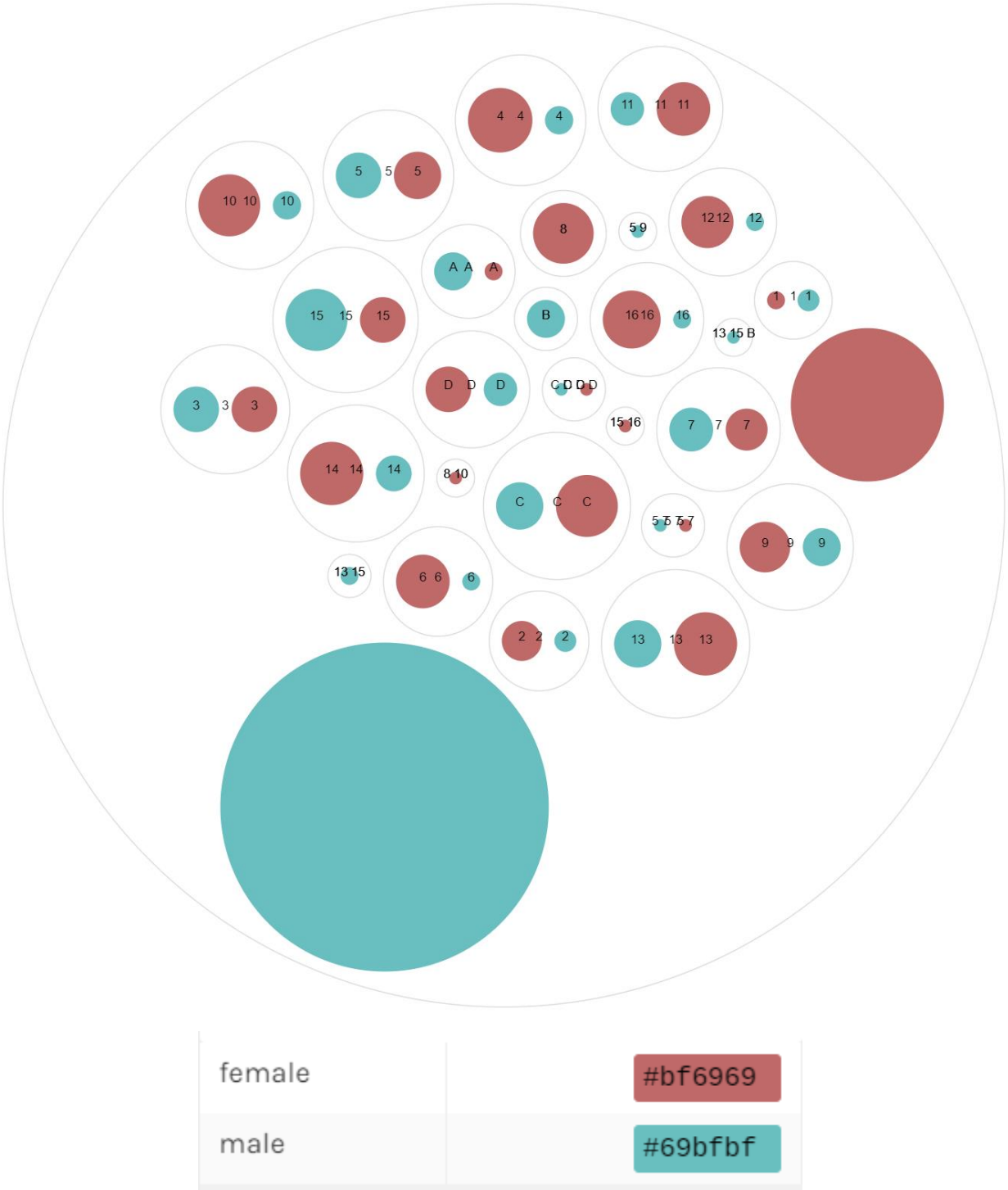

Figure 3 – Gender disparity amongst the survived passengers while evacuating in boats

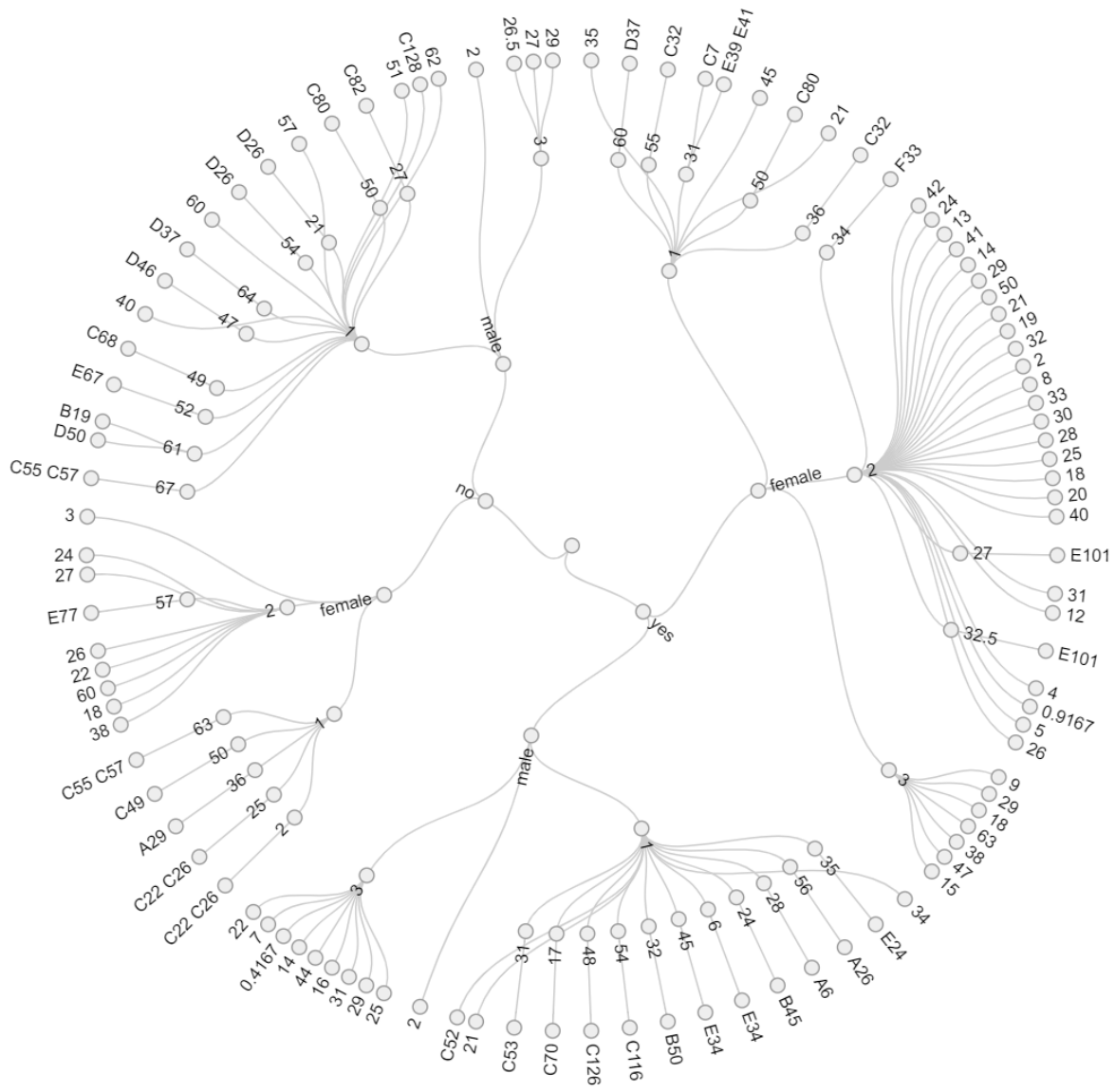

Figure 4 – Hierarchical representation of Survival Status, Gender, Class, Age and Cabin details

Is there a relation between the survival status of a passenger to the fare of their ticket?

The motivation behind this question is to understand how the ticket fare plays a role in the survival status of the passenger. After exploring different types of graphs to visualize this question, we have ended up selecting the 'Beeswarm Plot' to aptly depict the relation.

- Price for the tickets across the three given classes are ranging from $0 to $510.
- It can also be observed that most of the population is concentrated between the price range of $0 to $50.
- Second densely populated price range is from $51 to $100 followed by $101 to $300.

- Furthermore, there are only four passengers who have purchased the tickets at a higher price of $500 and all of them have survived the mishap.
- From the 'Figure 1' we can infer that there are less survivals and more deaths amongst the passengers.
- Off these deaths, 80% are observed in the price range of $0 to $50.
- Contrarily, a higher survival rate can be observed in the price range of $51 to $100 and $101 to $300.

We can observe that there are few instances in 'Figure 1' where the price of ticket is $0. It can be assumed that these people are the crew members and not really a passenger. It might also be a case where the tickets were offered freely owing to the fact that it is the maiden journey of RMS Titanic.

In conclusion, yes, fare of the ticket do have an impact on the passengers survival status as we can observe that survival to death ratio is lower in the price range of $0 to $50 while it is high in the price range of $51 to $500.

Did the passengers age define their survival status?

We have used 'Alluvial Plot' which we believe can trace all possible pathways that passengers take in a longitudinal dataset. This mapping is done between the survival status of the passenger with their corresponding age groups. Age from 0 to 100 years have been classified into 10 different groups with each group capturing 10 years. Considering the 'Figure 2', we can observe below points.

- Initial glance at the plot tells us about the higher deaths on comparison with survivals.
- Surprisingly, there are many unknowns for the value of passenger's age. At the same time, age group of 90 to 100 is missing from the data.
- Majority of the passengers fall under the age group of 20-29 and 30-39, whereas passengers aged 60 and beyond are very less in number.
- Infants and children below 9 might have been put on priority list as we can see a high survival rate in the age group 0 to 9.
- Teenagers (11 - 19), Youth (20 - 29) and Adults (30 - 49) have observed higher deaths than the rest.
- We can also observe that all the passengers in the age group 80 - 89 have survived.
- There were passengers with unknown age group and majority of them did not survive.

We might possibly infer that the Teens, Youth and Adults might have helped other infants, children and old passengers while evacuating RMS Titanic which resulted in their higher death rate.

All in all, people aged in the range of 0 - 9 and 80 - 89 have witnessed more survivals than deaths which conveys that these age groups have received some level of precedence than others. It is not incorrect to assume that the passengers surviving in the age group 20 - 39 is partially due to their accompaniment with their infants and children.

Did the passengers gender play any role in securing one's life while evacuating?

Evacuation boats played a crucial role in transporting the survivors aboard the sinking passenger liner to the nearest shores. To better understand the evacuation process during the mishap and to study the presence of gender disparity if any, we wanted to visualize the data using 'Circle Packing' plot as it best suited the agenda in question.

- As the legend suggests, blue depicts male population while light maroon corresponds to female. In addition to this, the boat number is mentioned in the smaller circles present within the larger circle.
- The two largest circles in 'Figure 3' for both male and female is corresponding to the passengers who never got a chance to board the evacuation boats. This list is populated with more males and lesser females.
- Almost in all the boats, we can observe that there are more females compared to males.
- Boat numbers '3', '5', '5 7', 'C D' seems to carry equal proportion of both the genders.
- It is very interesting to see some boats with all males and few with all females. These boats correspond to the number's 'B', '5 9', '13 15' and '13 15 B' for all males and '8', '8 10' and '15 16' for all females.

Capacity of the evacuation boats does not seem to be consistent as few carried many passengers (bigger circles within the smaller circles in 'Figure 3') while the rest remained a bit lower. This might also be due to overcrowding resulting due to anxiety.

To conclude, yes, we have observed higher precedence being given to the female passengers during the evacuation process since many of the boats are predominantly filled with female passengers.

Summary

Out of our interest, we wanted to try and depict the details about all the three investigating points in one plot and hence created 'Figure 4'. In here, we have started off with the survival status which is then divided into two sections based on the gender of the passengers. These values were further divided into three different classes based on the ticket fares. Furthermore, these classes have been divided into different age groups and finally into the cabins associated with the passengers of those age groups.

Many datasets and visualizations have been explored before finalizing on to the 'Titanic's Passengers' dataset. Rawgraph tool facilitated many graphs but few of them were restricted to certain points like having a mandatory feature with numbers or dates thus restricting us from applying it on our dataset. Various options available for customizing the plots proved to be helpful in defining the size, labels and colors for effective visualization.

References

Rawgraph - https://rawgraphs.io/