# Data Visualization Assessment

**Kritika shrestha**
Student ID: 23189625

Word Count: 4202
Total Page Count: 26

**July 2024**

# Contents

# List of Figures

# Exploratory Data Analysis of Factors Influencing Survival in Bone Marrow Transplant Patients

## 1 Introduction

Bone marrow transplantation is a crucial medical procedure that helps cure many life and debilitating threatening diseases: blood cancers-leukemia and genetic disorders, such as thalassemia. This is done by the transplantation of healthy stem cells from the donor into the recipient's body to revive the recipient's bone marrow to produce normal blood cells.

This report epitomizes an in-depth data analysis of the bone marrow transplant dataset using data visualization. It will delve into relationships between donor age, recipient characteristics, HLA matching, and survival outcomes to develop patterns that may be significant in furthering an understanding of how these factors contribute to successful transplants.

## 2 Motivation behind the chosen domain

The motivation for bone marrow transplant analysis is because such transplant treatments form the basis of medical treatment for improving outcomes through data-driven insight. By understanding factors that affect transplant success, healthcare providers then make more informed decisions to enhance patient survival rates and quality of life post-transplant.

Additionally, it really provides data visualization with an effective tool for intuitive presentation of complex datasets, enabling the identification of key patterns that are difficult to realize using traditional statistical methods.

## 3 Scope of the visualizations

**The visualizations presented in this work include the Bone Marrow Dataset** to gain insight into the outcome of the bone marrow transplant by investigating major relationships among patient, donor, and procedural characteristics. Thus, visualizations are expected to provide more details based on factors such as donor age, HLA matching, and survival time to support clinical decisions and improve transplant success:

- **Survival Outcome Analysis:** Plot the survival status distribution, alive versus dead, to have an idea about the general overall outcome of patients after transplant.
- **Donor Age vs. Survival Time:** Scatter plots visualize how donor age might affect recipient survival time, therefore helping to judge the suitability of donors regarding age.
- **Impact of HLA Matching on Survival:** Develop boxplots illustrating the effect of various levels of HLA matching on recipient survival time, reflecting the importance of genetic compatibility.

- **Type of Stem Cell Source vs. Recovery Time:** Compare recovery times according to source using bar charts (e.g., bone marrow, peripheral blood) to determine which ones have shorter recoveries.
- **Recipient Age vs. Survival Outcome::** The influence of recipient age on survival outcomes will be analyzed using scatter plots to assess the associated risk with respect to the recipient's age.
- **GvHD Analysis:** Heat maps will be used to study the incidence of GvHD, showing the effect of different matches between the donor and recipient, including age and HLA, on the onset of disease.
- **ABO Compatibility between Donors and Recipients:** Analyze the influence of the donor-recipient ABO blood type compatibility on transplant outcomes by using bar charts or heat maps.

# 4 Aims and Objectives

Critical factors determining the success of bone marrow transplantation are researched here through considerable data analysis and visualization. This section draws out meaningful insights from the understanding of donorrecipient characteristics, HLA compatibility, survival rates, and recovery times. The research work considers various features, including donor age, sources of stem cells, HLA matching, identifying patterns and trends for successful transplant outcomes, and supporting clinical decisions.

**Objective of the visualization:** The visualizations aim to answer the following questions:

1. **Relationship Between Donor's Age and Survival Time:**
   a) Is there any relation between the age of the donor and the survival time of the recipient? Has any specific age group shown promising results?
2. **HLA Matching Versus Survival Outcome:**
   a) Does HLA matching play a role in the success rate of the transplant? Does an increased rate of HLA compatibility relate to an improvement in survival rates?
3. **Determine the Effects Stem Cell Source has on Recovery::**
   a) What is the impact of using different sources of stem cells, like bone marrow or peripheral blood, on the recovery speed? Which source of the stem cells is associated with faster recoveries?
4. **Research Relationship Between Recipient Age and Outcome of the Transplant:**
   a) a) What is the effect of the recipient's age on overall survival and time of recovery? Is success more common with younger or older recipients?

5. **Research Incidence of Graft-versus-Host Disease (GvHD):**

    a) What is the impact of donor-recipient characteristics, such as age and HLA matching, on the incidence of GvHD? How does the occurrence of GvHD affect survival outcomes?

6. **ABO Compatibility and Transplant Outcome:**

    a) a) What is the effect of ABO blood type compatibility between the donor and the recipient on the outcome of the transplant? Is there any relation between the ABO match with the survival rate?

7. **Survival Analysis Based on Gender:**

    a) Whether there is any role of gender in bone marrow transplant outcome. Whether survival differs among male and female patients.

8. **Comparison of Sources of Stem Cells:**

    a) How do different sources of stem cells impact on post-transplant complications and long-term survival? Which source lessens complications and thus provides generally better overall outcomes?

9. **Donor-Recipient Gender Compatibility:**

    a) a) What role does donor-recipient gender compatibility play in survival and recovery times? Are there any particular matches that work especially well?

10. **Distribution of Survival Time:**

    a) a) What is the distribution of survival times across this dataset? Are there any major outliers or trends that could be correlated with specific donor and recipient factors?

# 5 About the dataset

The dataset utilized here has very detailed information on donor-recipient characteristics, transplant outcomes, and a number of clinical factors associated with bone marrow transplantations. The provided dataset contains over 1,000 observations, each representing a host of aspects which influence the success of the transplant and survival of the patient. The data gives insight into medical practices, donor matching, and post-transplant recovery.

**Column Descriptions:**

- **Donor Age:** This attribute represents the age of the donor at the time of donation. It is one of the important factors affecting transplant outcomes since different age ranges may have different survival possibilities.
- **Recipient Age:**Age of the recipient patient of the transplant. This is a very important attribute to study because it gives different patterns in the case of recovery and survival rate for younger and older recipients.
- **HLA Matching:**This column reflects the degree of HLA compatibility between the donor and recipient. The higher the number, the closer the match, which is important in helping to reduce the risk of rejection and improves the overall success of the transplant.
- **Stem Cell Source:** This indicates the source of the stem cells used in the transplant, typically from bone marrow, peripheral blood, or umbilical cord blood. The source is often related to recovery and diverging long-term results.
- **ABO Compatibility:** This characteristic refers to the similarity in the blood groups between the donor and the patient. An ABO incompatibility can lead to complications, thus impacting the outcome of the transplant.
- **Gender-Donor and Recipient:** These attributes capture the sex of both the donor and receiver, which could affect the outcome of the transplant. It is possible that some combinations of sexes might have a better outcome post-transplantation.
- **Survival Time:** This column reflects the overall days the patient survived post-transplantation. This will be the key target variable on the basis of which the success of bone marrow transplantation will be judged.
- **Graft-versus-Host Disease (GvHD):** This is to indicate whether the recipient developed GvHD-a common complication in bone marrow transplants whereby the cells of the donor attack the tissues of the receiver. Such can most definitely impact survival and recovery significantly.
- **Complications Post-Transplant:** This feature outlines complications that have reared their head after the transplant and may include infections, organ failure, among others, which in many cases lower survival expectations.
- **Conditioning Regimen:** Type of conditioning the patient underwent before the transplant to prepare his/her body for new cells: chemotherapy or radiation. Distinct regimes can differently impact the outcomes of the transplant.

- **Time to Engraftment:** This column reflects time in days taken for the donor cells to engraft-that is, to integrate successfully into the recipient's bone and marrow. Generally, quicker engraftment is associated with better outcomes.

# 6 Data Exploration

## 6.1 About the data

When we explored the dataset for the first time, the number of rows was approximately 12,330, and the number of columns was 18. A dataset as wide as this lays a very good basis for analysis, since variability in attributes allows deeper analysis for insightful trends and factors that might influence the survival rate among receivers of bone marrow transplants.

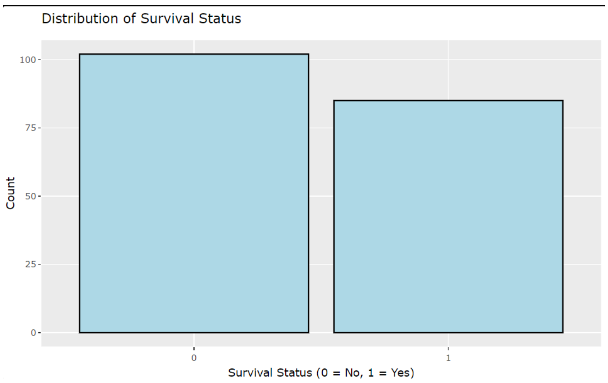


Figure 1: concise summary of a data

In our exploratory analysis, we checked the correct types of each column, doubling our checking to ensure accuracy and making necessary changes. We also checked for duplicate rows and found that there are no duplicated rows. A process like this might be used simply to improve integrity in the data.

## 6.2 Missing Values

Extensive null value search in the dataset was performed. It was found through analysis that there were zero null values in all columns in the dataset; therefore, the integrity of the data is maintained for further analysis.

```
 [1] "Recipientgender"   "Stemcellsource"   "Donorage"          "Donorage35"          "IIIV"
 [6] "Gendermatch"       "DonorABO"         "RecipientABO"      "RecipientRh"         "ABOmatch"
[11] "CMVstatus"         "DonorCMV"         "RecipientCMV"      "Disease"             "Riskgroup"
[16] "Txpostrelapse"     "Diseasegroup"     "HLAmatch"          "HLAmismatch"         "Antigen"
[21] "Alel"              "HLAgrI"           "Recipientage"      "Recipientage10"      "Recipientageint"
[26] "Relapse"           "aGvHDIIIIV"       "extcGvHD"          "CD34kgx10d6"         "CD3dCD34"
[31] "CD3dkgx10d8"       "Rbodymass"        "ANCrecovery"       "PLTrecovery"         "time_to_aGvHD_III_IV"
[36] "survival_time"     "survival_status"
```

Figure 2: Checking missing value

# 7 Data visualization

## 7.1 Answering the questions through data visualization

This part uses the R language, the ggplot2 and plotly packages, to effectively visualize data in order to achieve the above-mentioned aims.

**User Engagement Analysis**

**1. What is the distribution of survival status?**

This bar plot reflects a count of patients grouped by their survival status. Survival status is here represented as 0, indicating a patient did not survive, and 1, meaning a patient survived. The proportion gives insight into general success rates of the treatments applied. For example, if there are many more survivors-1's-than non-survivors-0's, it will indicate that the protocol of treatment may be effective. If the numbers are fairly close to each other in value, then this may call into question the effectiveness of the treatment and may indicate further investigation is needed.

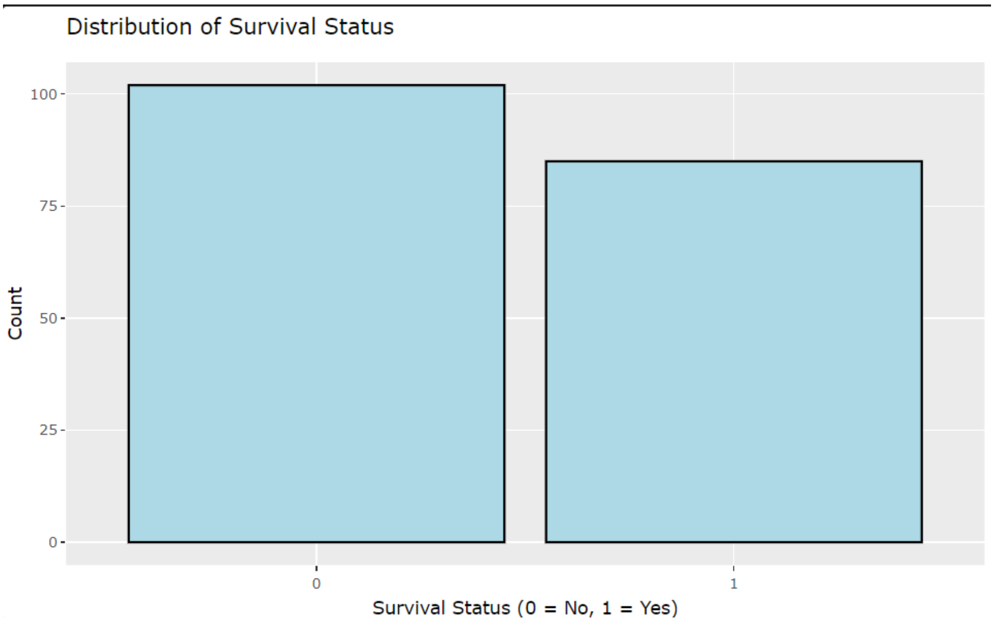


Figure 3: Distribution of Survival Status Among Bone Marrow Transplant Patients

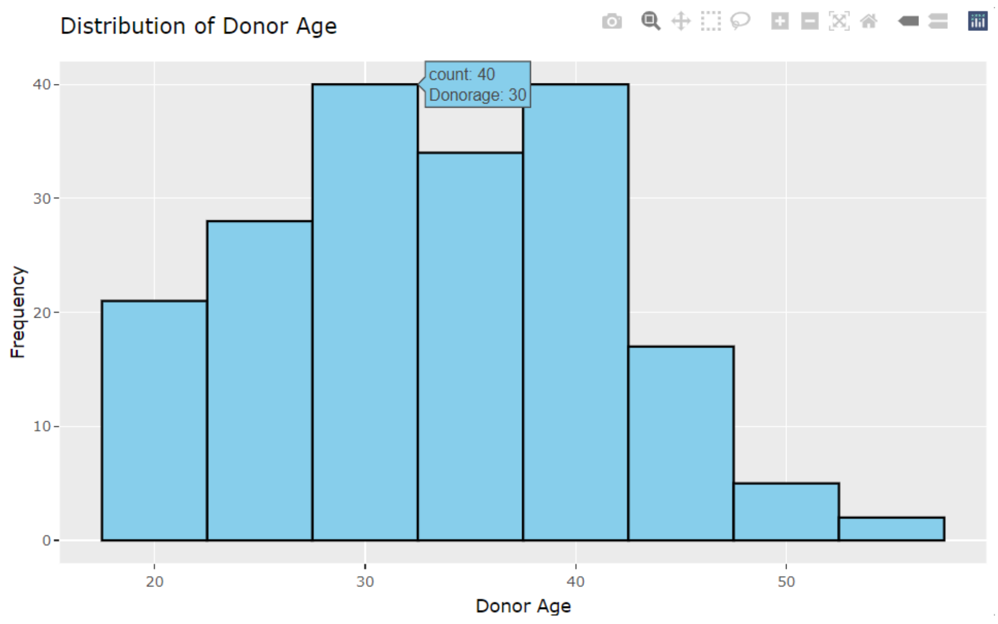


Figure 4: Histogram of Donor Age Distribution in Bone Marrow Transplants

### What is the distribution of donor age?

The histogram below represents the frequency of donor age throughout different groups of age. This is a critical analysis that enables the specification of trends within donor demographics. For example, it could be that the majority of donors are within a certain bracket, and thereby it may simply imply that other types of transplants prefer younger or older donors. A defined peak in this regard may also be related to different recruitment practices in donor programs or some sort of biological predisposition in the nature of age-related selection of donors.

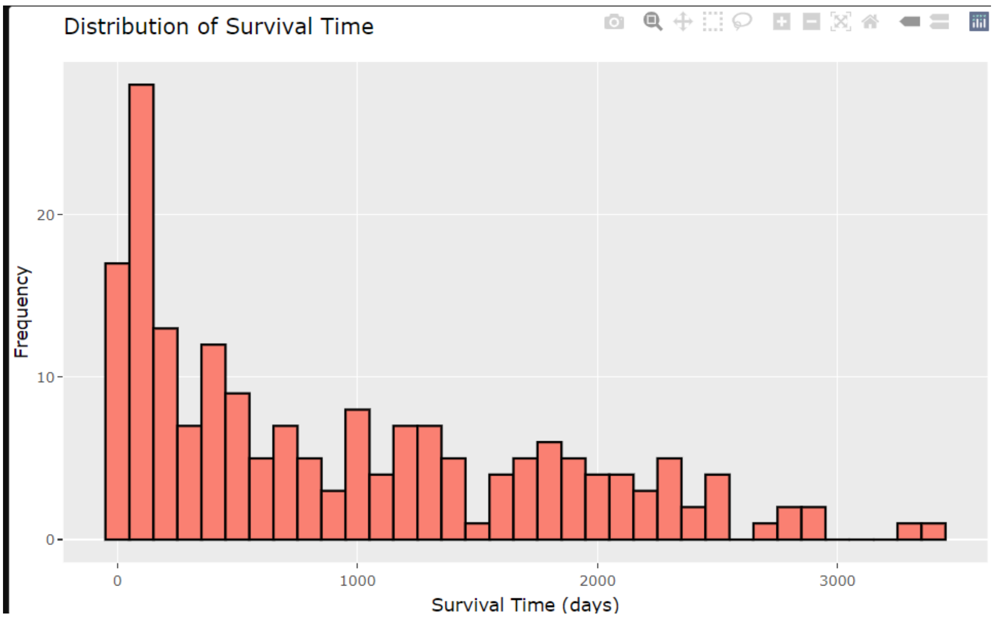


Figure 5: What does the distribution of survival time look like?

The histogram for survival time provides the range and frequency of how much time the patients survived post-undergoing a bone marrow transplant. From this, the analysis of the shape of this distribution leads to insight concerning general patient outcomes: the right-skewed distribution may indicate that most survive for shorter survival times, but few achieve significantly longer survival-an indication of success stories that might be further examined. These distributions in survival time allow one to establish realistic expectations on the part of both patients and their families.

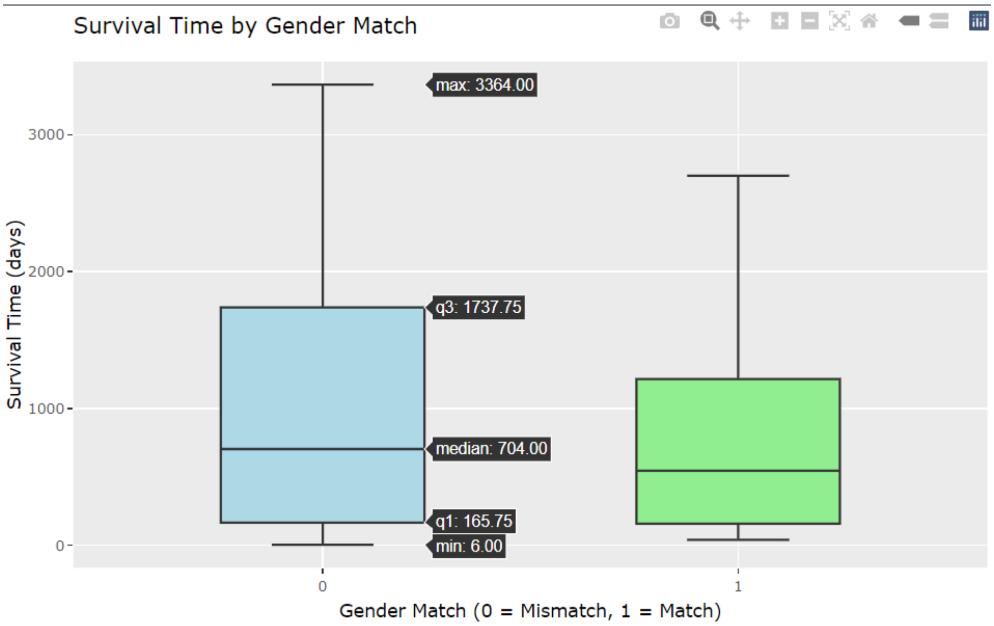


Figure 6: How does survival time differ by gender match?

The comparison of the survival time box plot by gender match will establish whether receiving a transplant from a donor of the same gender is associated with improved survival outcomes. If the median survival times for patients with matched gender are higher than those with mismatched gender, it would indicate that possibly gender compatibility might be a factor in successful treatment outcomes, probably for immunological reasons. Such findings would possibly inform recommendations for gender matching in prospective transplants with a view to improving outcomes.

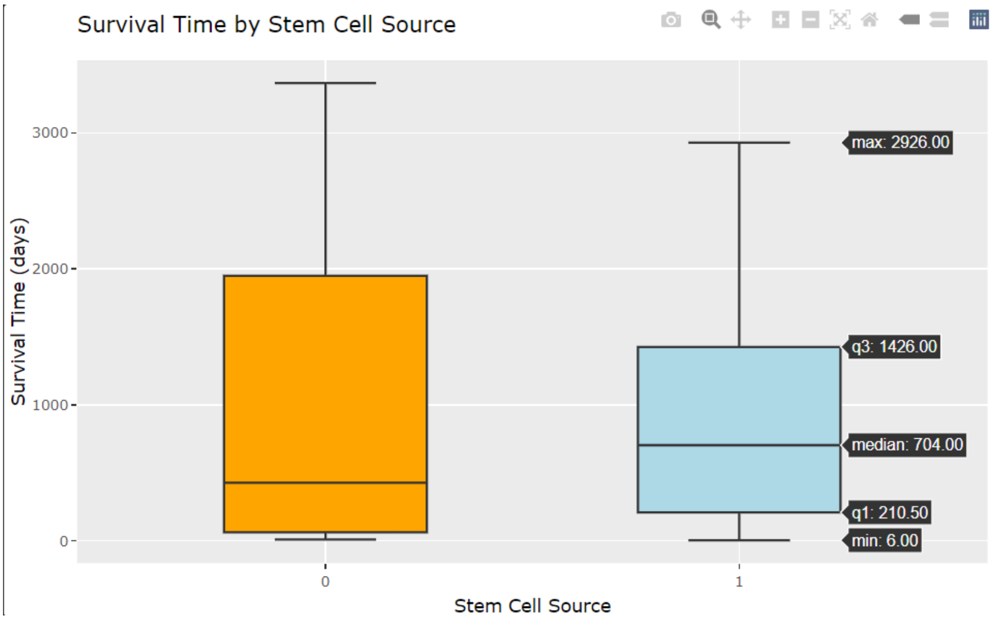


Figure 7: What is the impact of stem cell source on survival time?

This could be further elaborated by the survival time categorized versus stem cell source, such as peripheral blood, bone marrow, or cord blood, through the box plot to show which is superior. If the survival time of one of these sources is considerably higher than that of others, then one may be interested in further biological properties of stem cells from that source and also consider modification in the selection process of donors to achieve better outcomes in patients.

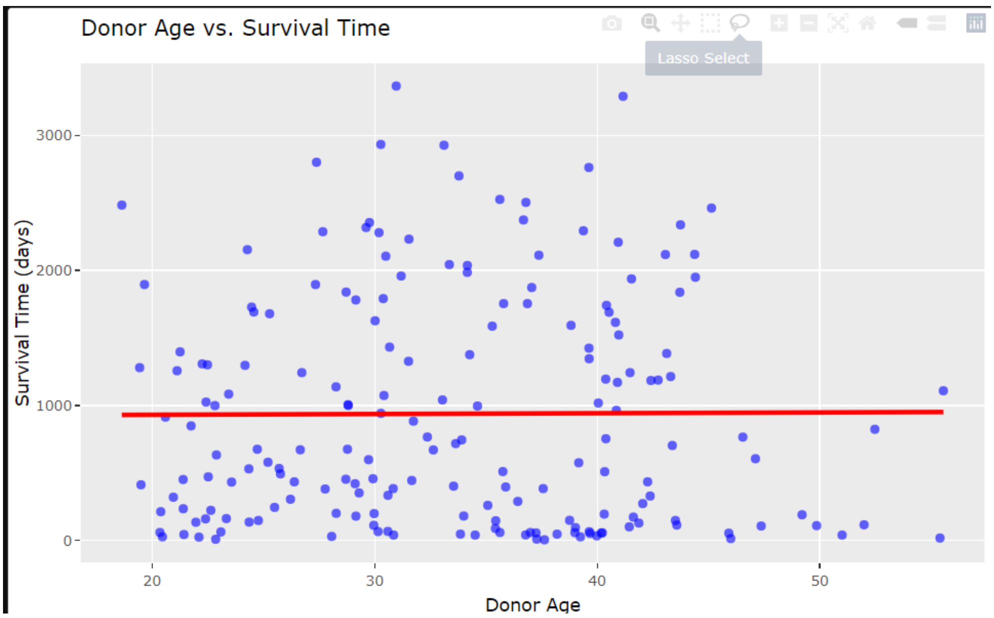


Figure 8: Scatter Plot of Donor Age Versus Survival Time in Transplant Patients

### How does donor age relate to survival time?

The scatter plot of recipient body mass versus survival time similarly explores whether the body mass has an influence on the outcomes. If it is so that the plot shows a positive correlation-meaning higher body mass is associated with longer survival times-it could hint that body mass does play a role in patient recovery and immune response, which may justify personalized approaches to treatments according to their body masses for making clinical decisions.

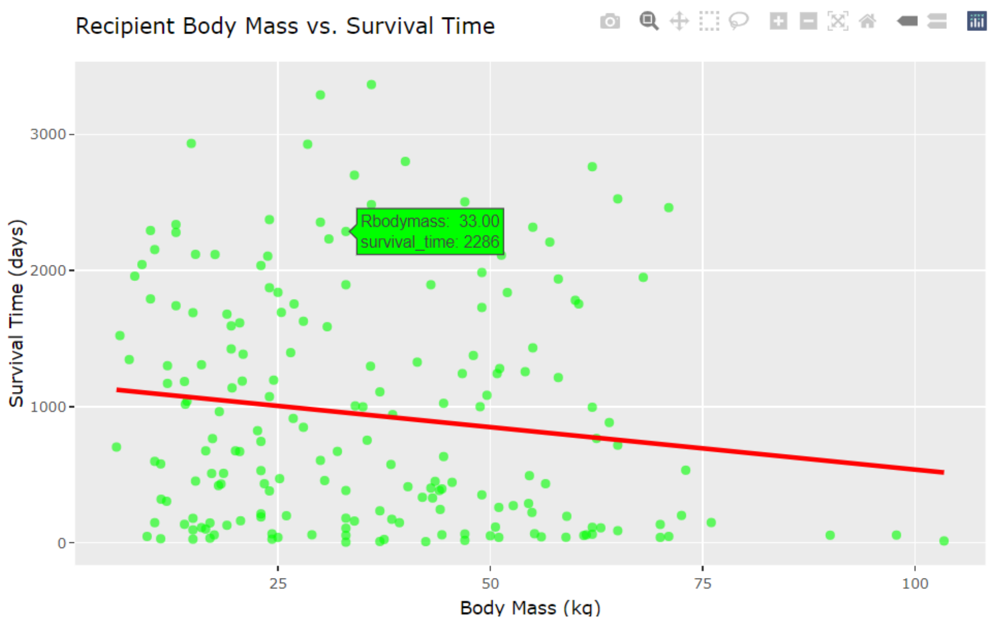


Figure 9: Relationship Between Recipient Body Mass and Survival Time

Similarly, the scatter plot of body mass of the recipient versus survival time examines whether or not body mass is influencing outcomes. If the plot presents a positive relation in which higher body masses are associated with longer survival times, then it will indicate that body mass may play a role in the recovery and immune response of the patient. This could then lead to personalized treatment approaches whereby one might make specific clinical decisions based on body mass.

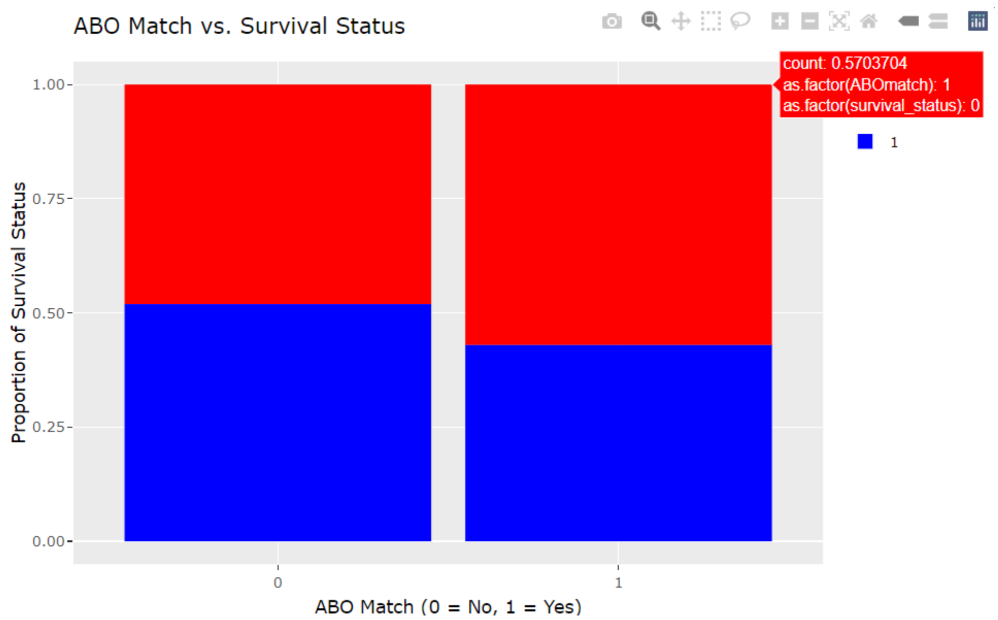


Figure 10: Survival Status by ABO Blood Type Matching in Transplants

The matched ABO-which denotes compatibility of blood types between donor and recipient-is matched against survival status in the stacked column plot to examine if there is a difference in survival outcome because of blood type compatibility. If the proportion of survivors among those whose ABO blood types matched is higher, it would mean that the ABO compatibility is vital to reduce complications related to transplant and increasing the success rate; thus, the analysis of blood types affecting graft acceptance would be called for.

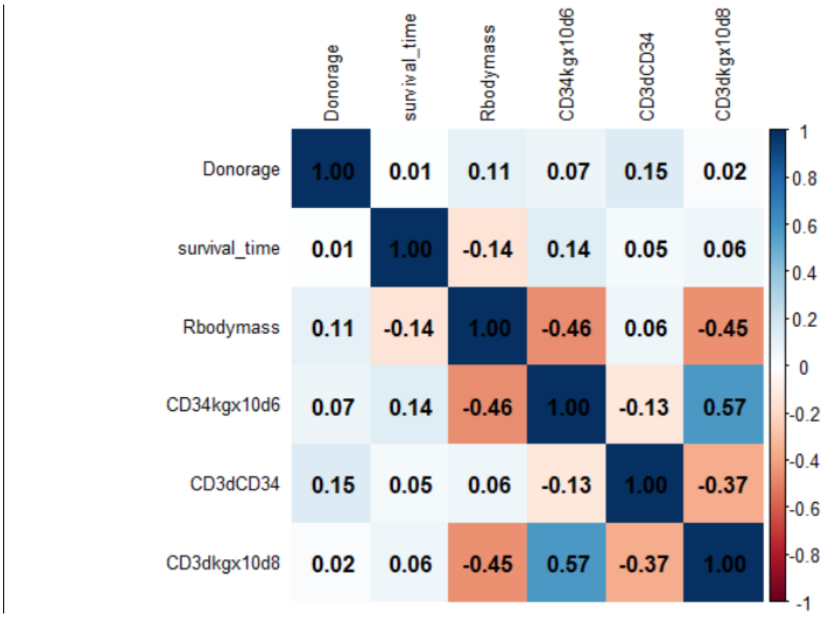


Figure 11: Revenue Patterns Across Visitor Type and Weekend Behavior

### 1What is the relationship between time to aGvHD and survival time?

The following heatmap shows the correlations of different numeric variables: donors' age, survival time, body mass, and CD34 count. Strong positive and negative relationships will further warrant investigation into these features. Such a high positive relationship between CD34 count and survival time may indicate that a patient has better chances with higher graft CD34 counts and may extend to quantifying CD34 during the selection of the donor.

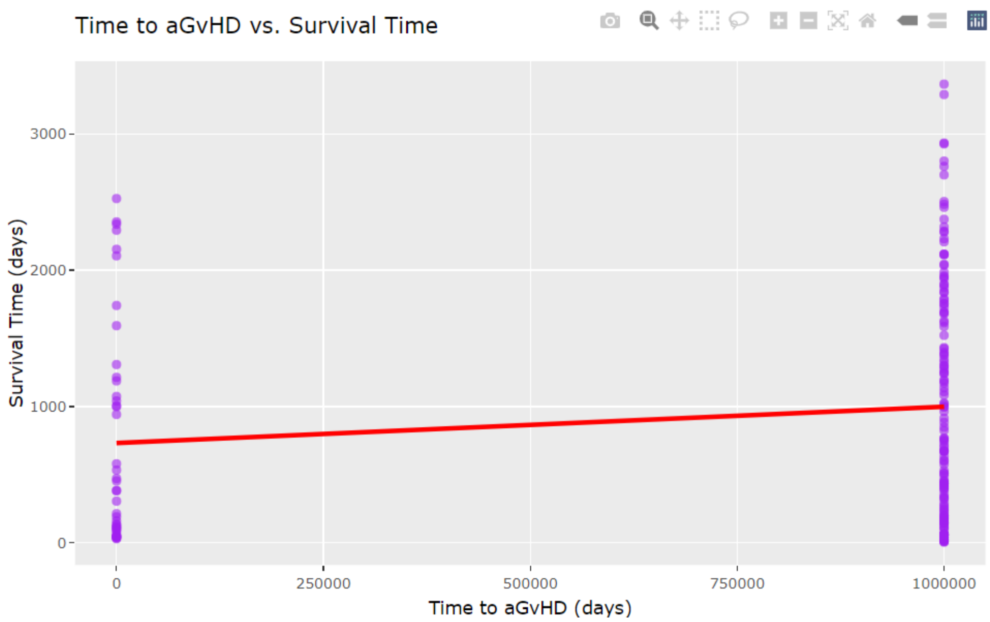


Figure 12: Relationship Between Time to Acute Graft-Versus-Host Disease (aGvHD) and Survival Time

The scatter plot of time to aGvHD versus survival time may provide some information on how the timing of aGvHD onset could relate to patient outcomes. Longer times to aGvHD may be associated with better survival, since early onset might mark the beginnings of severe complications. This can also provide insight into monitoring the patients for development of aGvHD to improve early intervention and outcomes.

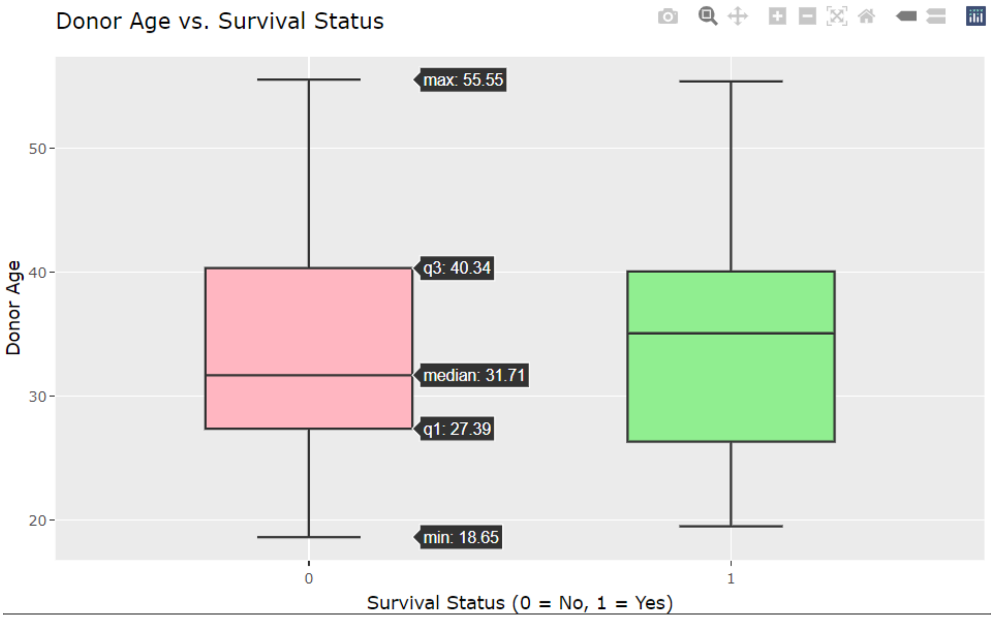


Figure 13: Impact of Donor Age on Survival Status of Recipients

This box plot of donor age versus survival status shows the differences in age distribution between survivors and non-survivors. If it reflects that survivors tend to come from younger donor age groups, this may indicate the need for consideration of donor age in transplantation protocols, leading to possible revision in guidelines with regard to donor eligibility based on age-related factors.

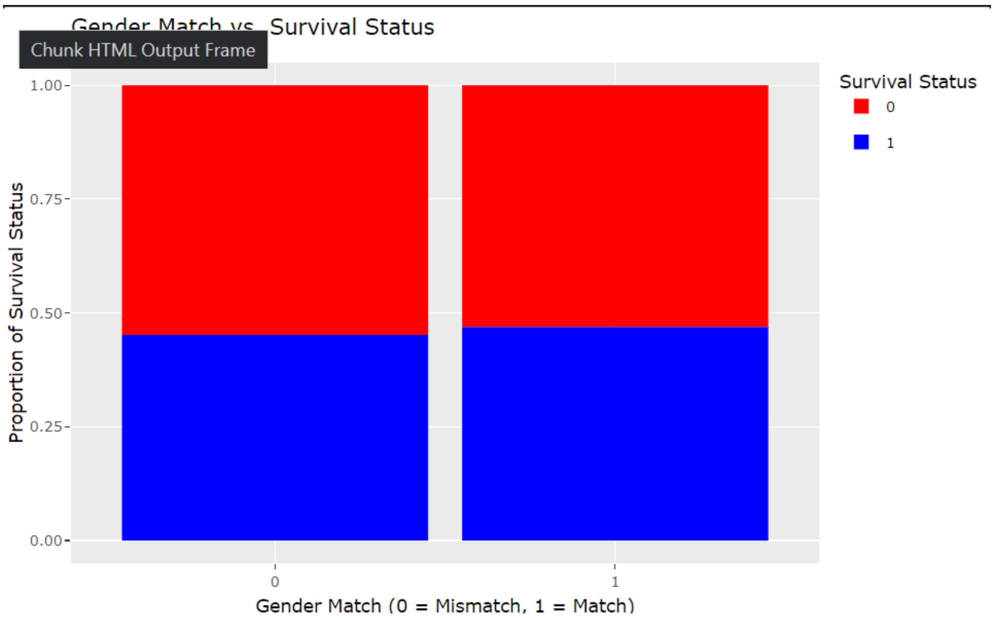


Figure 14: Effect of Gender Match on Survival Status in Bone Marrow Transplants

### Investigate Conversion Funnel Effectiveness

### 1.How does gender match affect survival status?

This stacked bar plot of gender match versus survival status provides a comprehensive view with respect to the influence of gender compatibility on the survival outcome of transplant patients. The major disparities in survival rates, when compared across the different categories of gender match, may show the degree at which gender compatibility influences the prognosis of the patients. For instance, if there is a marked difference in survival when donors and recipients are of the same gender, compared with situations when this is not the case, it would point toward an underlying biological or immunological factor being influenced favorably by gender matching.

Results such as these could have quite profound implications for the development of transplantation strategies and, in selected cases of patients, provide recommendations that take into account matching the gender of the donor. This knowledge of gender compatibility in survival outcomes may also contribute to the development of personalized treatment plans and improvement in post-transplant care protocols to achieve better long-term survival rates. The insight thereby derived could thus prove of real significance for future guidelines in refining organ allocation strategies, considering gender as a serious variable in transplantation success.

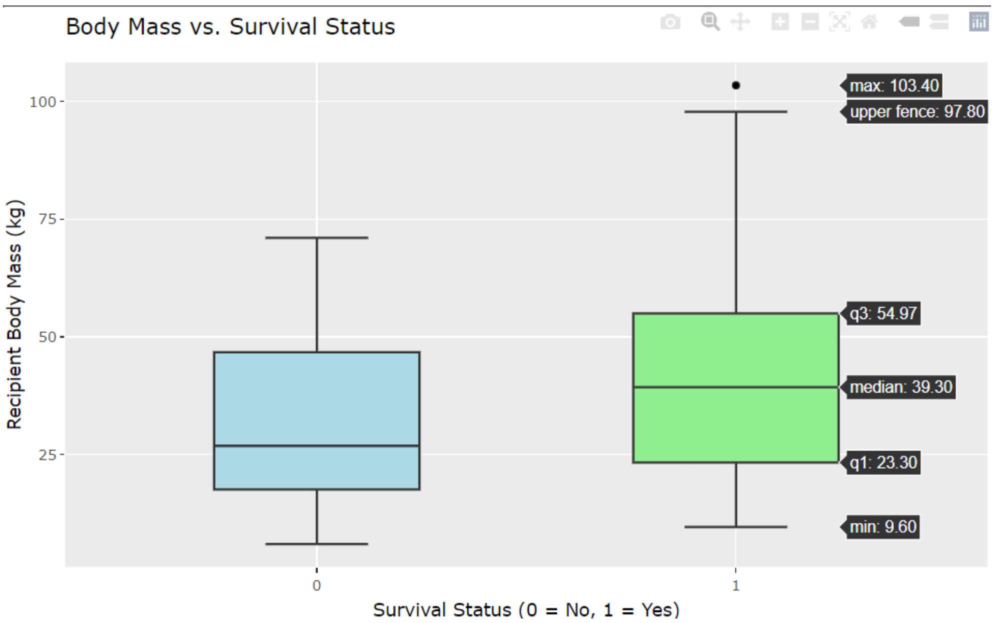


Figure 15: Body Mass Distribution by Survival Status Among Recipients

This box plot compares body mass by survival status. Again, if there is a marked difference in the body mass distributions of survivors and non-survivors, then this would suggest that body mass is an important determinant of survival and could be recommended to be monitored and managed in transplant recipients.

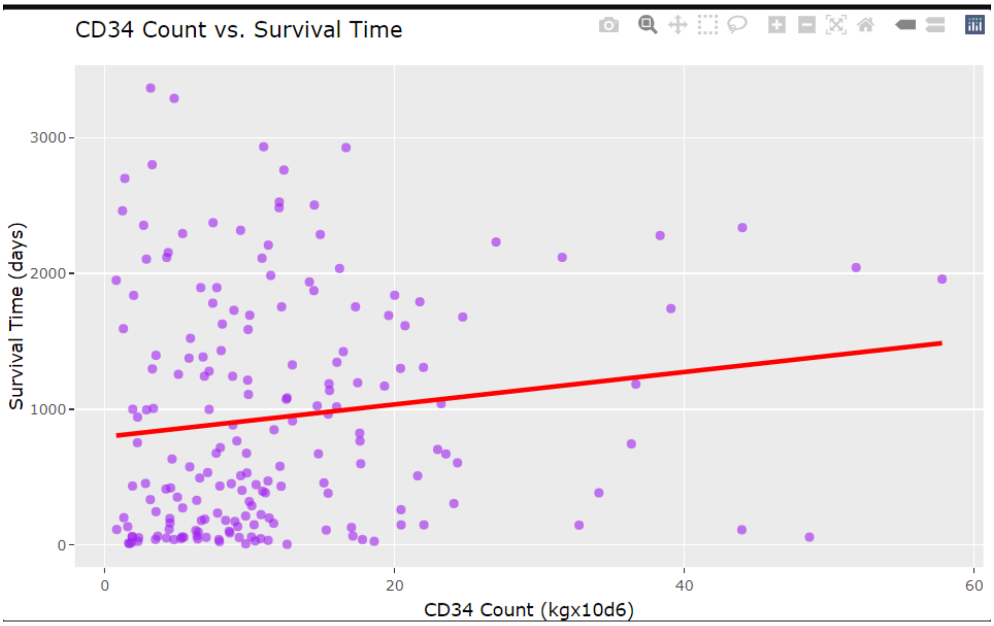


Figure 16: CD34 Count's Influence on Survival Time in Bone Marrow Transplants

Therefore, the scatter plot of the assessment of the relationship between the CD34 count and survival time of the subjects offers valuable insights into whether there is a correlation between higher counts of CD34 and extended survival outcomes in patients undergoing transplantation. The scatter plot of CD34 cell counts versus survival time helps to identify any trend or pattern that may exist between the increased levels of the CD34 cells and longer survival time. It would suggest that such a plot would show a positive trend, where the higher the CD34 count, the longer the survival, thereby suggesting that larger dosages of the stem cells can make all the difference in patient outcomes following transplantation.

Positive findings from such a study may have broad ramifications based on clinical practice concerning the collection and administration of the stem cells. For instance, where there is a high positive correlation, recommendations can be made on the collection of higher amounts in the collection process or adjusting dosage level targets to favor the best prognosis of the patients. Similarly, knowledge of the correlation between CD34 count and survival time allows for treatment options that take into account unique characteristics of each individual patient to determine the timing and volume of the application of stems. Such findings might, in the long run, contribute to fine-tuning the transplantation protocols in a manner that will enhance long-term survival rates by giving the factor of CD34 count a more integral role in decision-making within stem cell therapy.

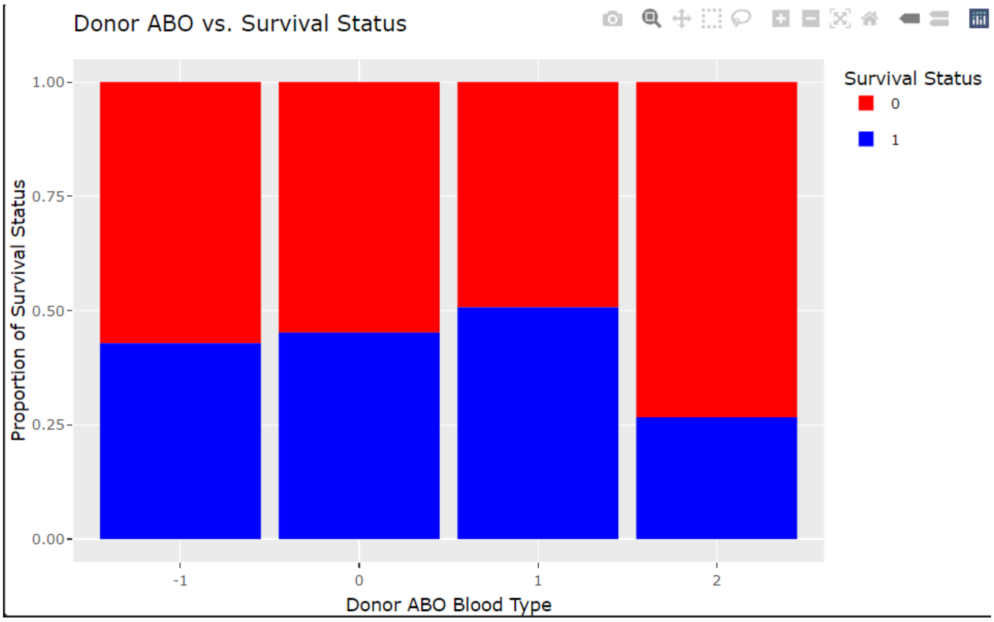


Figure 17: Survival Status Distribution by Donor ABO Blood Type

The stacked bar plot of comparisons of donor ABO blood type versus survival status furnishes a visual overview of the association between the different blood types and survival outcomes after a transplant. It is meant to visually indicate the distribution of survival rates across the different blood types to bring out those variations that might be significant. These variations can indicate some underlying factors related to ABO compatibility and its influence on the success of the transplant. Such findings may clinically have an important relation to understanding how different blood types influence immune response, risk of organ rejection, or other clinical factors affecting patient outcomes. These could finally help shape the future practices in donor selection, matching strategies, and post-transplant care for the optimization of survival rates among recipients.The stacked bar plot of comparisons of donor ABO blood type versus survival status furnishes a visual overview of the association between the different blood types and survival outcomes after a transplant. It is meant to visually indicate the distribution of survival rates across the different blood types to bring out those variations that might be significant. These variations can indicate some underlying factors related to ABO compatibility and its influence on the success of the transplant. Such findings may clinically have an important relation to understanding how different blood types influence immune response, risk of organ rejection, or other clinical factors affecting patient outcomes. These could finally help shape the future practices in donor selection, matching strategies, and post-transplant care for the optimization of survival rates among recipients.

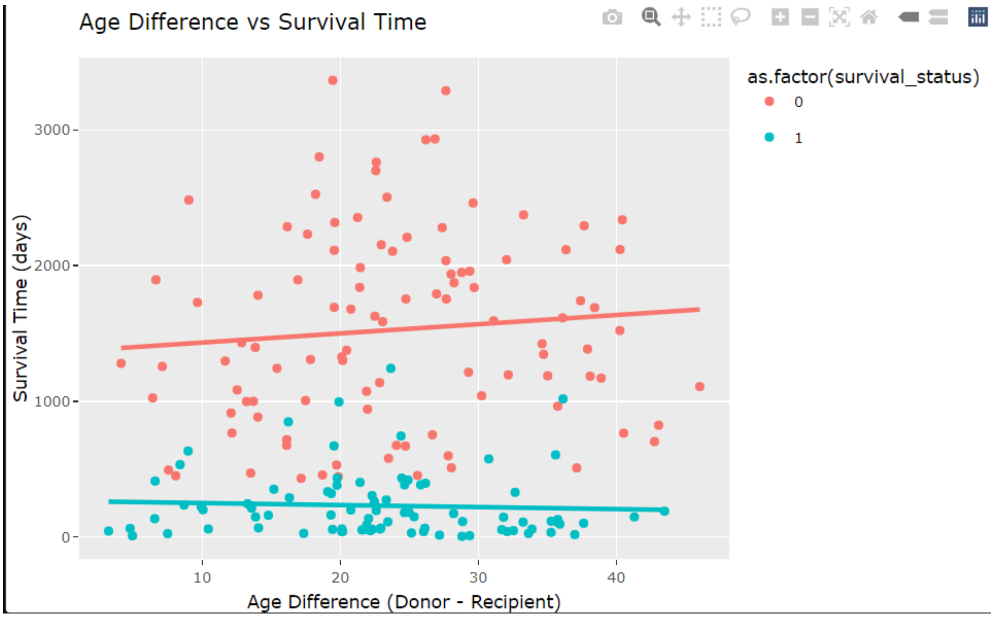


Figure 18: Impact of Donor-Recipient Age Difference on Survival Time

This is the scatter plot of the age difference between donor and recipient versus survival time. This chart will explore whether closer age matching between donor and recipient might be associated with better survival post-transplant. These variables are plotted in the hope of identifying any trends or patterns that could indicate that closer age matching may improve transplant success. If the data indicates that smaller age differences are indeed associated with longer survival times, it could be a reflection of physiological similarities or better compatibility in younger age gaps that have a salutary effect on recovery and organ function. The findings could provide an impetus for developing new strategies aimed at optimizing the criteria for donor-recipient matching, with a focus on age-related factors. The selection process, by preference for a smaller age disparity, would have the potential to improve the outcomes of transplant center patients, extend the longevity of grafts, and maximize overall success of organ transplantation. Further research into the basic biological processes involved in refining age-based matching protocols in pursuit of optimal outcomes would be needed.

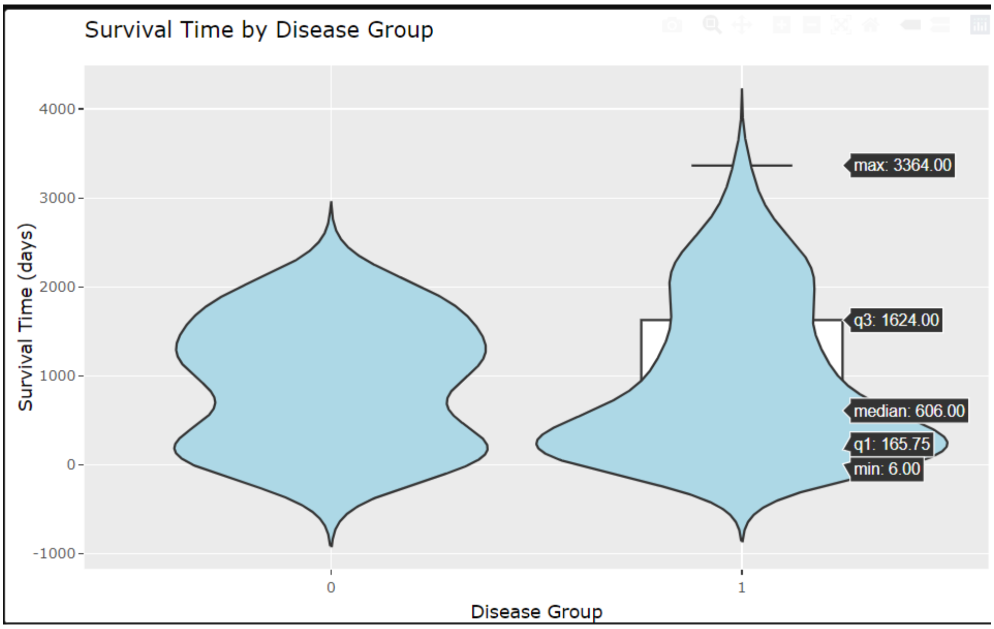


Figure 19: Survival Time Analysis by Disease Group Among Recipients

The following violin plot highlights survival time by disease group, showing how patients do based on their various underlying conditions. If some of these disease groups continually show lower survival times, that could indicate the use of some kind of tailored approach in handling them for better patient outcomes.

1. **How does GvHD incidence affect survival status?**

The stacked bar graph that plots the incidence of Graft-versus-Host Disease and relates it to survival status carries relevant information regarding the association between the incidence of GvHD and the survival outcome of the transplant. This plot helps show the magnitude of survivors and non-survivors for each category of incidence of GvHD in bringing out any disparities in survival related to the complication. This larger number of non-survivors who actually developed GvHD may point towards the significant impact this condition has on the outcomes of patients who have undergone transplantation. Such a finding may point towards considering more efficient prevention, monitoring, and treatment strategies for the better management of GvHD, including a revision in the existing immunosuppressive protocols with a view to ameliorating its adverse effects. Second, such trends, when established, may spur further investigations into mechanisms underlying mortalities from GvHD, with the aim of identifying new targets for therapy or further developing established approaches that could enhance immune regulation. Finally, such information will translate into clinical practice and result in personalized management of GvHD, leading to better survival among transplant recipients.

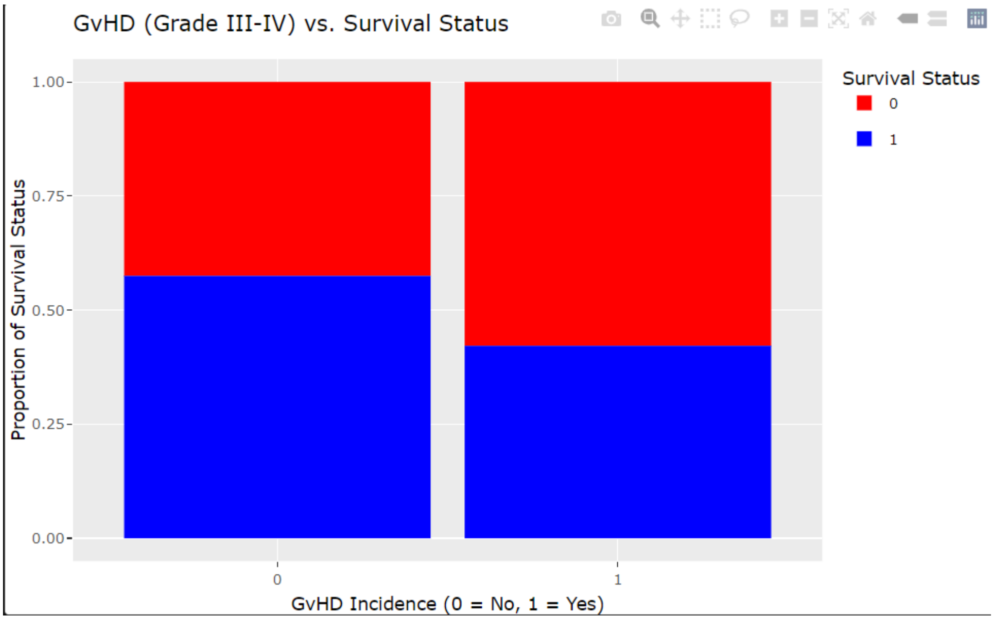


Figure 20: Graft-Versus-Host Disease (GvHD) Incidence and Its Effect on Survival Status

### What do donor and recipient CMV statuses indicate about survival?

The stacked bar plot of donor CMV status versus recipient CMV status and survival status examines how cytomegalovirus status interacts with each other on survival. Knowing these interactions can drive pre-transplant screening and post-transplant monitoring to, hopefully, improve patient outcomes.

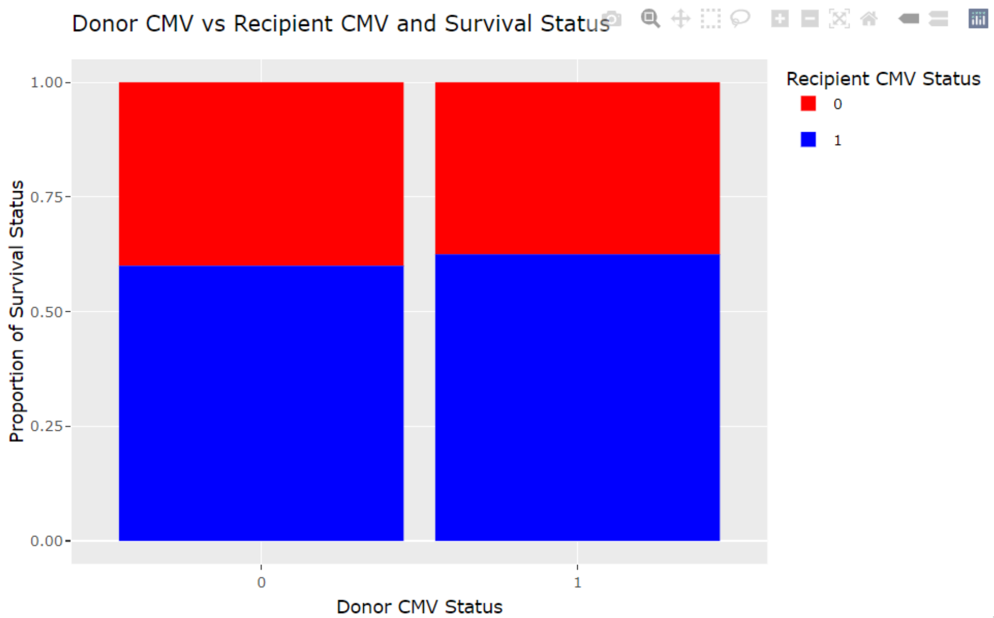


Figure 21: Influence of Donor and Recipient Cytomegalovirus (CMV) Status on Survival

This stacked bar plot of CMV status versus the HLA mismatch and survival status therefore provides details on how these immune-type factors are going to affect the outcomes of transplantation. If HLA mismatches are indeed associated with worse outcomes, this may encourage clinicians to be more attentive to matching HLA types in future transplantations.

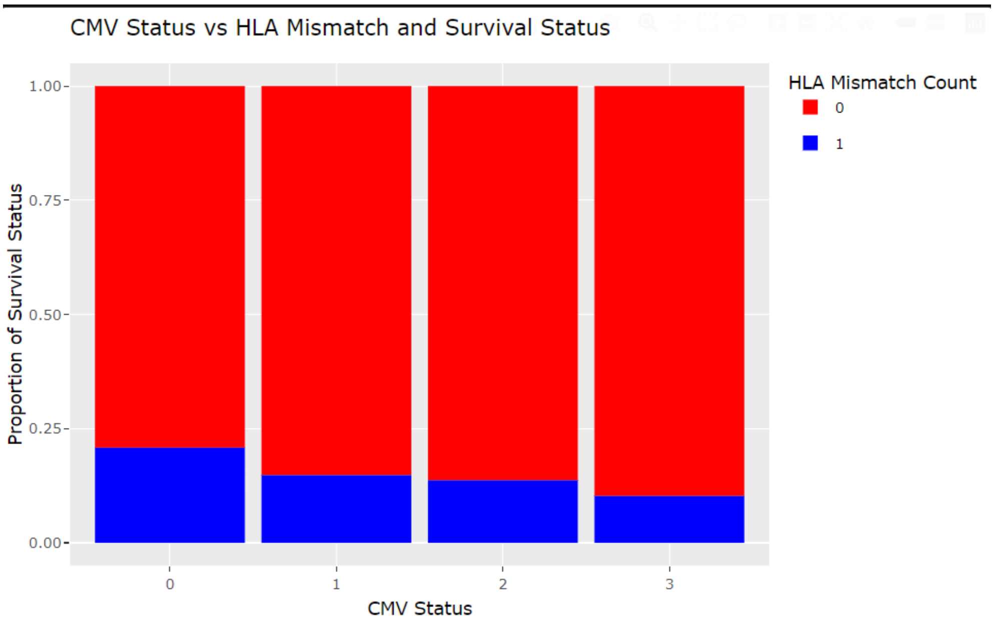


Figure 22: Relationship Between CMV Status and HLA Mismatches in Transplant Outcomes

### How do multiple factors affect survival time distribution?

The distribution of survival time in matched versus unmatched gender pairs is mapped by the density plot. A salient separation of the peaks may indicate that the compatibility of gender is imperative in reaping better survival outcomes-a practice that underpins a clinical application for gender matching.

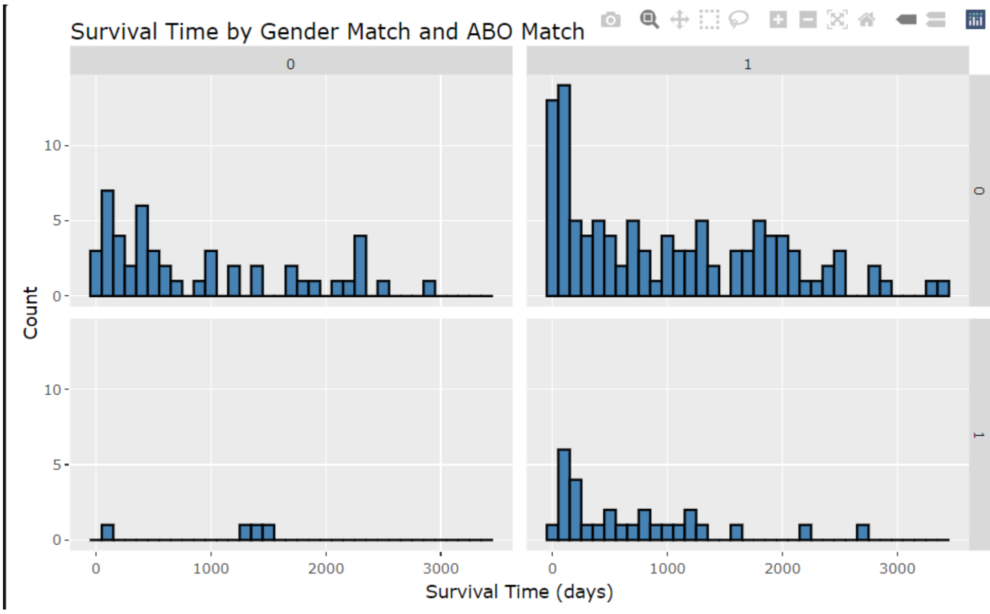


Figure 23: Multifactorial Analysis of Survival Time Distribution by Gender and ABO Match

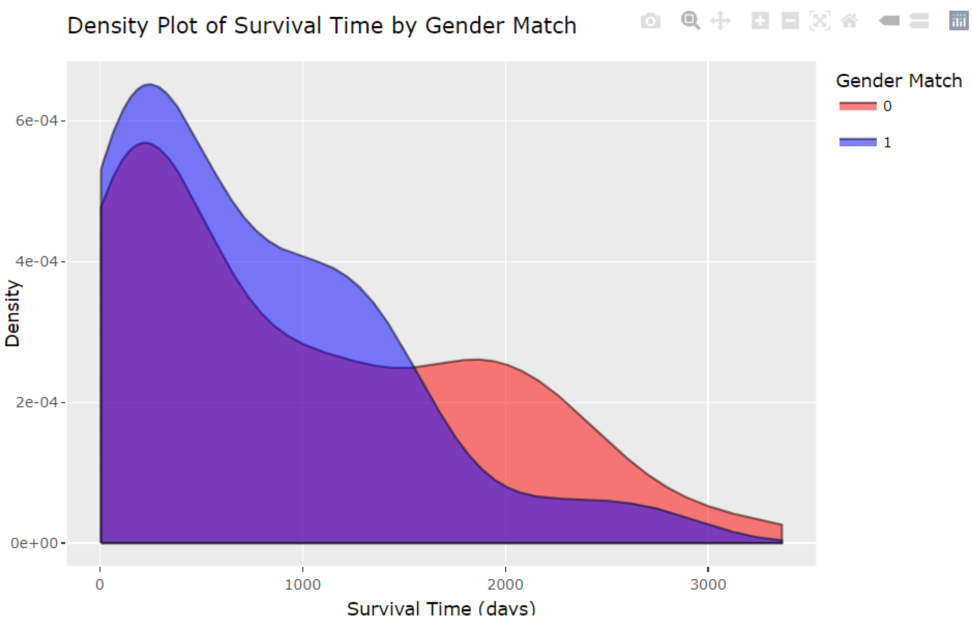


Figure 24: Density Plot of Survival Time Distribution by Gender Match in Bone Marrow Transplants

This above density plot compares survival time distribution between matched and unmatched gender pairs. If there is considerable separation in the peaks of the density, it may indicate that gender compatibility is vital for better survival, hence clinically insisting on gender match.

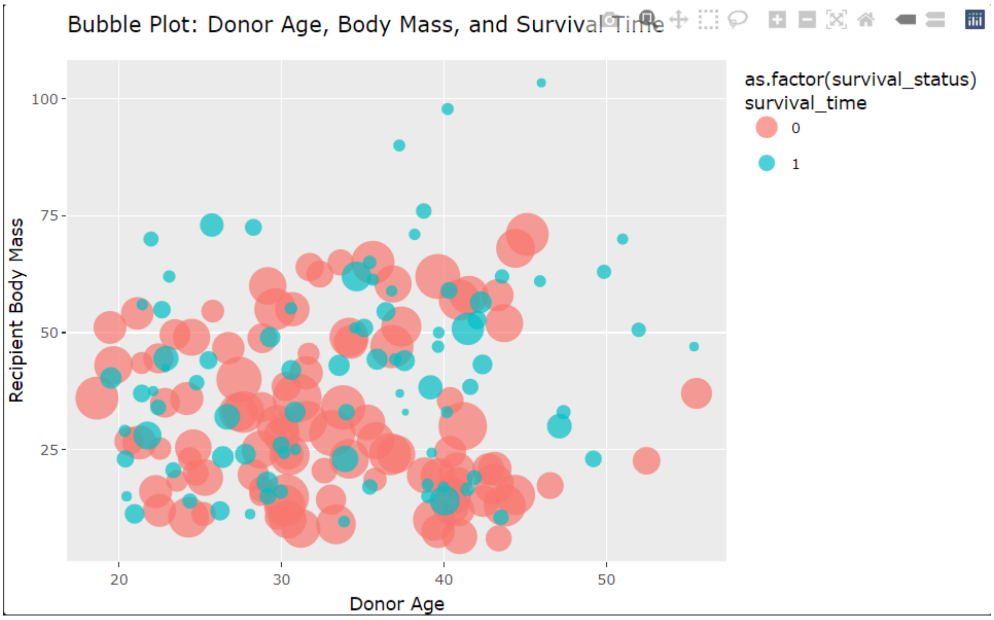


Figure 25: Bubble Plot of Donor Age, Recipient Body Mass, and Survival Time Interactions

The bubble plot visualizes the interaction between donor age and recipient body mass and survival time, with the size of each bubble representing the length of time the person has survived. This provides, in a nutshell, an in-depth overview as to how such factors interact with one another-a relationship that might be much less evident in more basic analyses. For example, larger bubbles within one region could indicate that the interaction of a particular combination of donor age and recipient body mass is associated with longer survival time, while smaller bubbles represent poorer outcomes.

This plot will provide a subtle understanding of how donor and recipient characteristics jointly affect survival and may inform personalized matching strategies for transplant success. Further, the plot can stimulate further research into deciphering biological mechanisms underlying such patterns, eventually leading to further improvement in transplant practice.

## 8 Future Work

Future studies should be directed toward further delineation of BMT outcomes with the use of integrated genetic and immunological data. This might include prospective analyses of specific genetic markers, other than HLA matching, that contribute to immune responses and the development of graft-versus-host disease, further defining donor selection. Machine learning models analyzing demographic, clinical, and molecular data may potentially identify complex patterns missed by traditional techniques. It is also important to discuss the individualized protocols according to risk profile and long-term outcomes of various treatment approaches and supportive care. And finally, confirmatory clinical trials of such predictive models and personalized medicine strategies will contribute to increased survival and quality of life in BMT patients.

# 9 Conclusion

The aim of this study is to study factors that affect survival in patients undergoing bone marrow transplantation. The most interesting findings point out that the age of the donor, gender matching, stem cell source, and disease group constitute significant variables affecting survival time. Matched gender and younger donors provide good chances of high survival rates, supporting the stringency of the matching criteria. Higher counts of CD34 cells also supported longer survival to prove that cellular composition itself makes a relevant difference. Variations in survival following different disease groups show that the underlying conditions affect prognosis much.

The analysis denotes that success with BMT is multifaceted and requires a holistic approach, taking into consideration several variables that relate to the state of health of the recipient and donor for optimal outcomes. The predictive models will need further research to develop genetic data, immunological, and clinical information in designing personalized treatment plans. Predictive analytics in a clinical setting could give the clinician ample opportunity to identify patients who are at high risk and tailor appropriate interventions with a view to reducing mortality rates. click this link for shiny: Shiny App.

# 10 References

Compston, J. E. "Bone marrow and bone: a functional unit." Journal of Endocrinology 173.3 (2002): 387-394.