# Analysis of Efficiency of Laser Treatment of Diabetic Retinopathy

Ananya Gangavarapu
Peter Tian
Data Science 1
1/20/2021

## Abstract

In this project, we attempt to analyze the efficiency of laser treatment in mitigating diabetic retinopathy. This analysis is being performed in order to gain more information about how laser treatments can influence survivability and prevent blindness, a key consequence for patients that have diabetic retinopathy. We use a diabetic retinopathy Kaggle dataset in order to conduct in-depth analysis into the effects of Diabetic Retinopathy. We utilize various methods in order to see how efficient the laser treatment is in mitigating the effect of diabetic retinopathy. Ultimately, our analysis of the dataset leads us to see that laser coagulation is effective in mitigating vision loss, one common complication of Diabetic Retinopathy. However, we cannot conclude that age affects the survival time rates of data subjects in the study and more effective sampling needs to take place.

## Introduction

Diabetic Retinopathy (DR) is a retinal disease that arises from various complications relating to diabetes (*2*). It involves the damaging of blood vessels in the retina for different individuals. Many of the common symptoms of Diabetic Retinopathy include spots, blurred vision, vision loss, and more. There are two stages to Diabetic Retinopathy: Early Diabetic Retinopathy and Advanced Diabetic Retinopathy. Early Diabetic Retinopathy is the most common type of Diabetic Retinopathy. In Early DR, the blood vessels in the retina begin to weaken, and blood vessels begin to get blocked. Diabetic Retinopathy can progress into Advanced Diabetic Retinopathy. In Advanced Diabetic Retinopathy, irregular blood vessels begin to form in the retina.  Complications arising from Diabetic Retinopathy include hemorrhages and vision loss.

Since the complications of Diabetic Retinopathy are so severe, many different approaches have been developed in order to mitigate the effects of DR. One common method of mitigating the effects of DR is utilizing laser coagulation. Laser coagulation is a treatment method that uses lasers to destroy various structures, such as irregular blood vessels in advanced DR, in the retina in order to prevent eye conditions (*3*).

In this project, we intend to take two different laser photocoagulation types: xenon laser and argon laser, to see their efficiency in preventing vision loss as a result of diabetic retinopathy. We also intend to see whether age of the data subjects within the sample affects the effectiveness of the laser treatment.

## Methodology

1. **Description**

In this project, we will utilize a Diabetic Retinopathy dataset provided in Kaggle, a popular forum where datasets and notebooks are shared (*1*). The dataset has 197 patients who had one of their eyes randomized for laser treatment while the other didn't. The dataset, in particular, is a 394 x 9 CSV file. There are 8 different columns within the dataset, which are labeled as id, laser, eye, age, trt, time, status, and risk. The data subjects in the dataset were a simple random sample of high risk patients. The data collection method was based on the Diabetic Retinopathy Study. Each variable is defined as follows (*4*)(*5*):

1. id - Specific data subject
2. laser - variable with 2 values: xenon and argon. These are the two types of laser coagulation that will be used in the project.
3. eye - variable with 2 values: left and right. This denotes which eye received the treatment.
4. age - age of the data subject when they were diagnosed for Diabetic Retinopathy
5. trt - variable that states whether the subject received the treatment in the eye or not
6. risk - Risk score for the respective eye that was treated
7. time - time taken for loss of vision (in months).
8. status - a variable that takes two values, 1 or 0.

I had utilized the Pearson correlation method in order to further analyze the association between the age and survival times of the sample. The Pearson correlation is computed by considering 2 linearly-related variables (*6*). If we have 2 vectors **x** and **y**, we can evaluate the Pearson correlation coefficient for a sample, where $x_i$ and $y_i$ are values inside the respective vectors,and $\overline{x}$ and $\overline{y}$ are the means of the values inside the vectors:

$$r_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})/\sqrt{}(\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2$$

**2. Preprocessing and Generation**

All of the graphs and values generated in this paper were generated by R programming language. I first imported the dataset into the environment. I then generated 9 different data frames based on certain factors I wanted to consider. Each data frame was defined in the following means:

DataFrame 1: age vs time
DataFrame 2: age vs time + status = 0
DataFrame 3: age vs time + status = 1
DataFrame 4: age vs time + xenon treatment
DataFrame 5: age vs time + argon treatment
DataFrame 6: age vs time + argon treatment in left eye
DataFrame 7: age vs time + xenon treatment in left eye
DataFrame 8: age vs time + argon treatment in right eye
DataFrame 9: age vs time + xenon treatment in right eye

For each data frame, a scatterplot of age vs time would be generated. For all data frames except the first, each data frame would filter specific values to use in the analysis. I then utilize R function ggplot in order to plot the scatter plot, and utilize R function cor in order to get the corresponding correlation coefficient (*7*).

## Results and Discussion

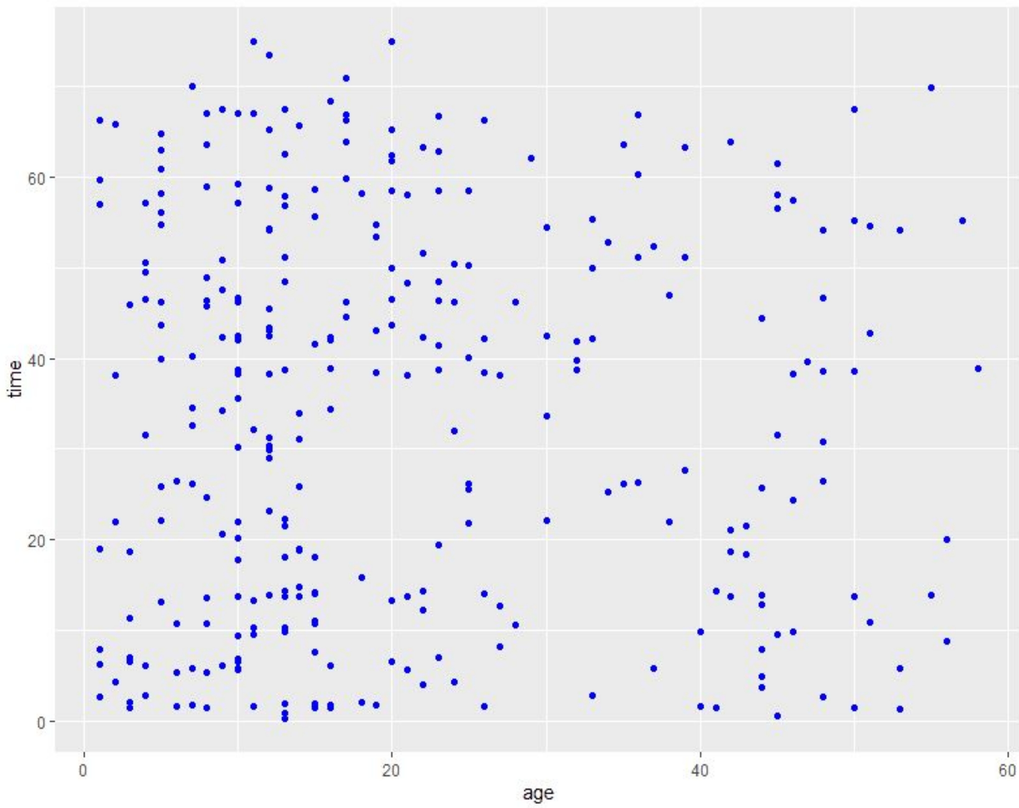


*Fig 1 - Scatterplot of age and time*

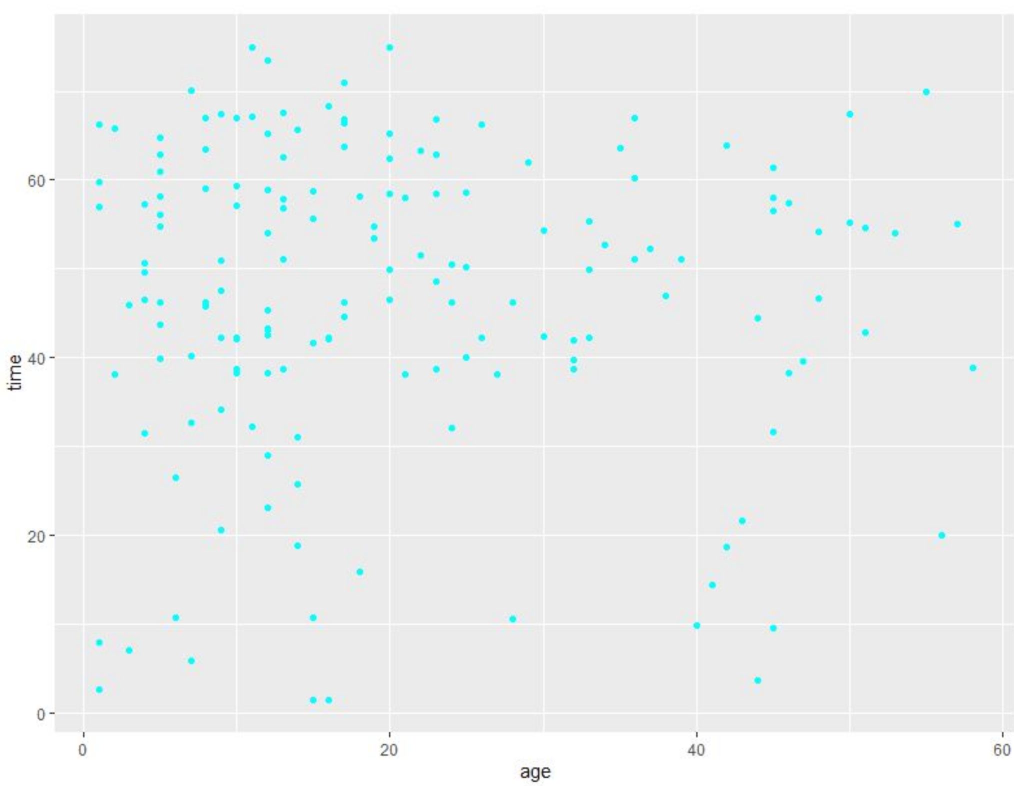

*Fig 2 - age vs time scatterplot with status = 0*

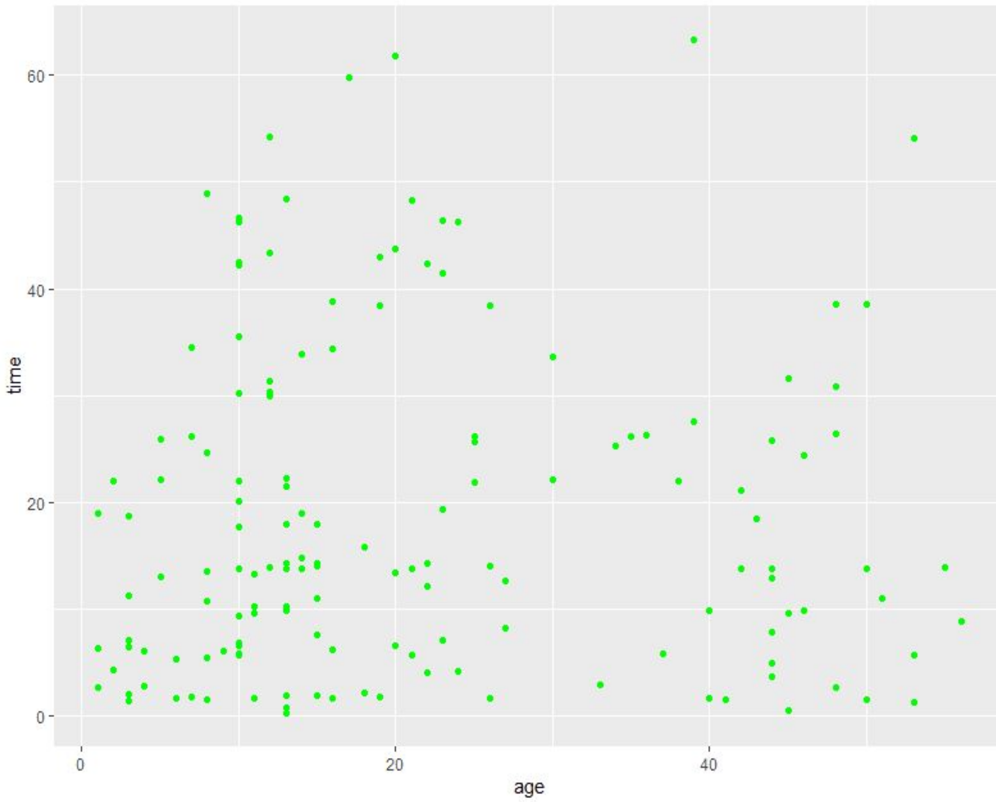


*Fig 3- scatterplot of age vs time with status = 1*

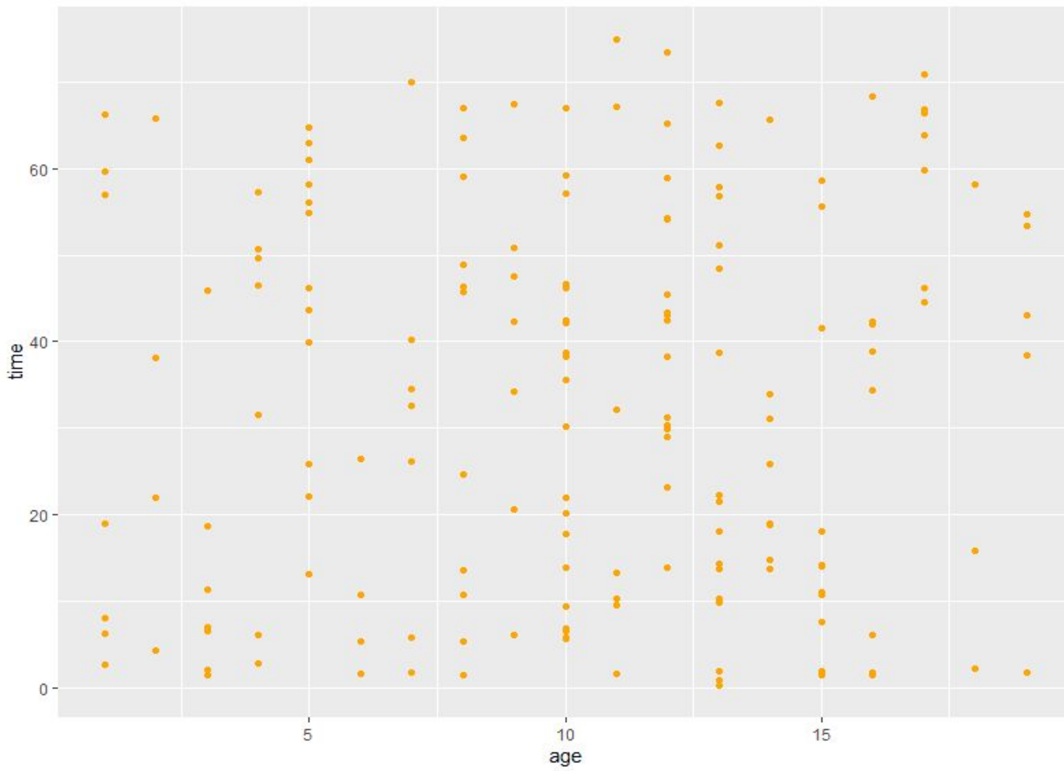


*Fig 4 - Xenon treatment*

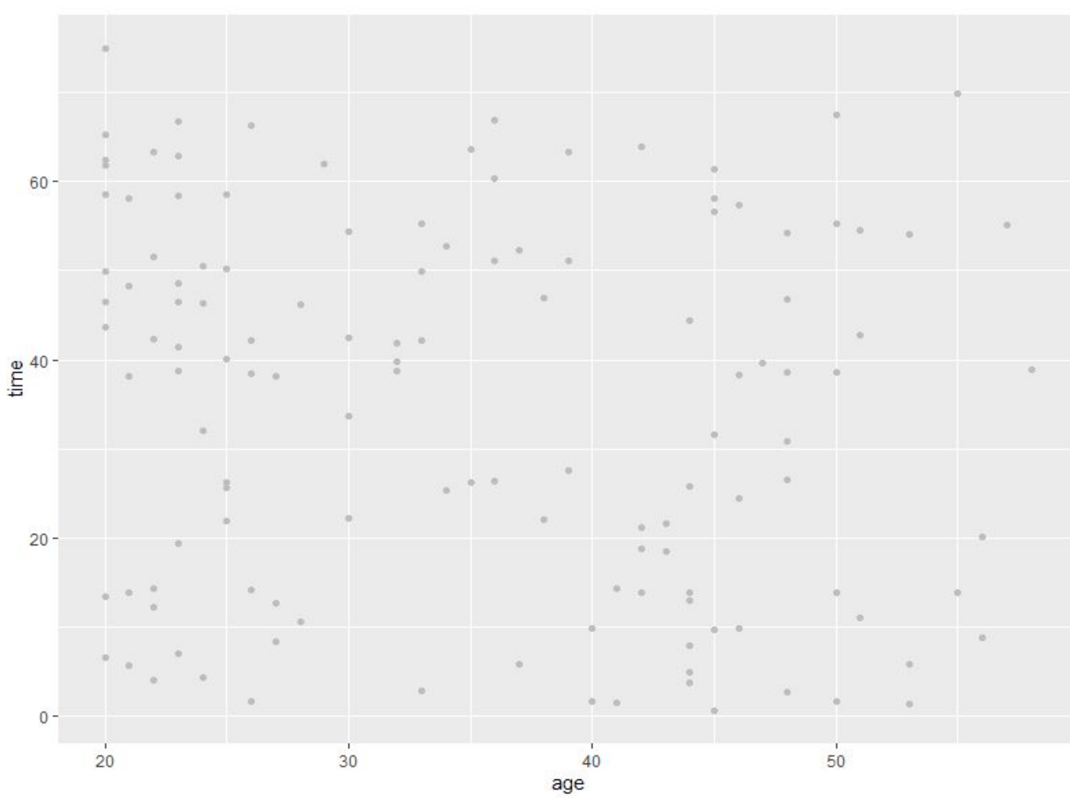

*Fig 5- Argon treatment*

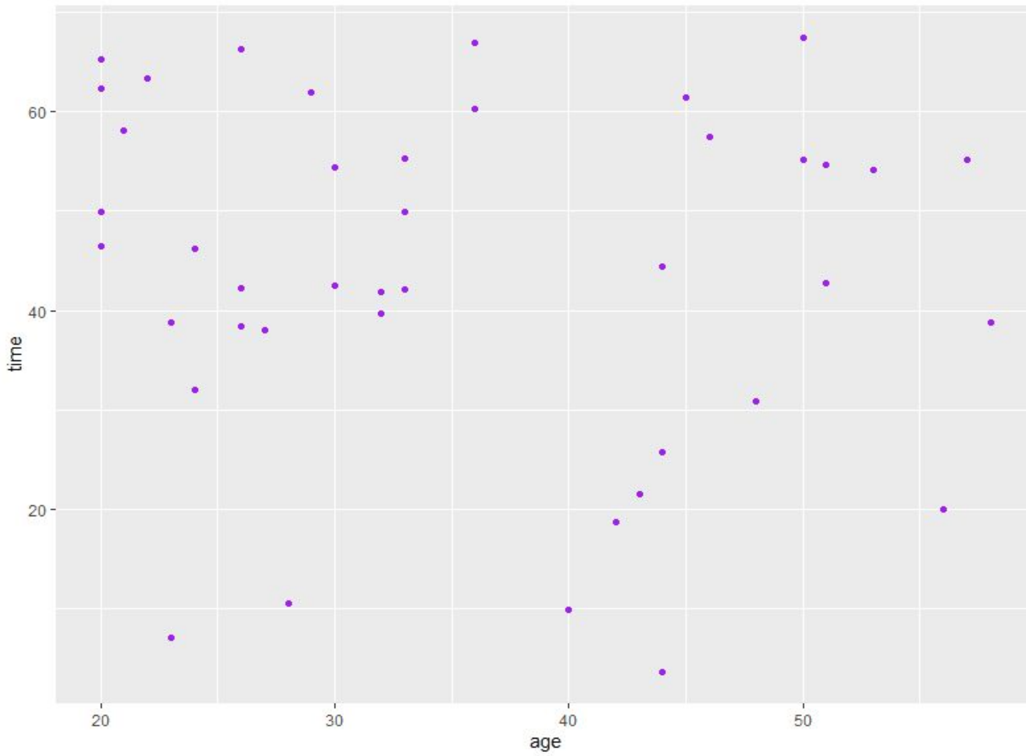


*Fig 4 - left eye treatment with argon*

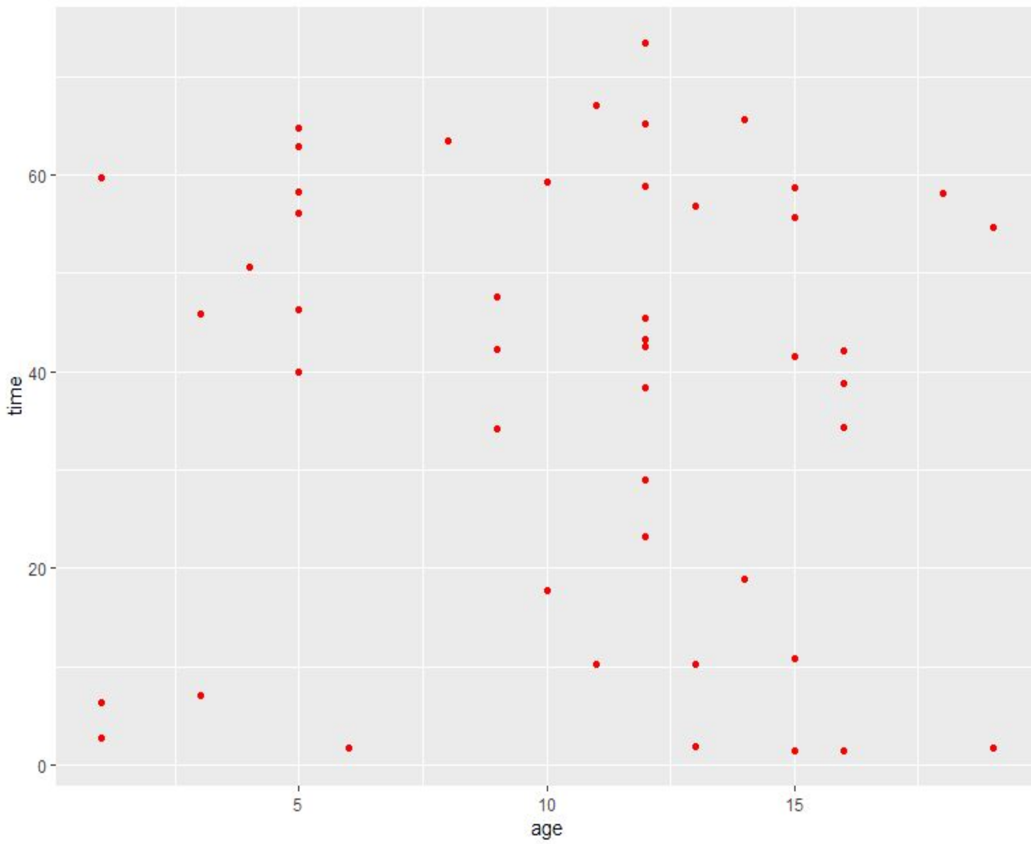


*Fig 5 - left eye treatment with xenon*

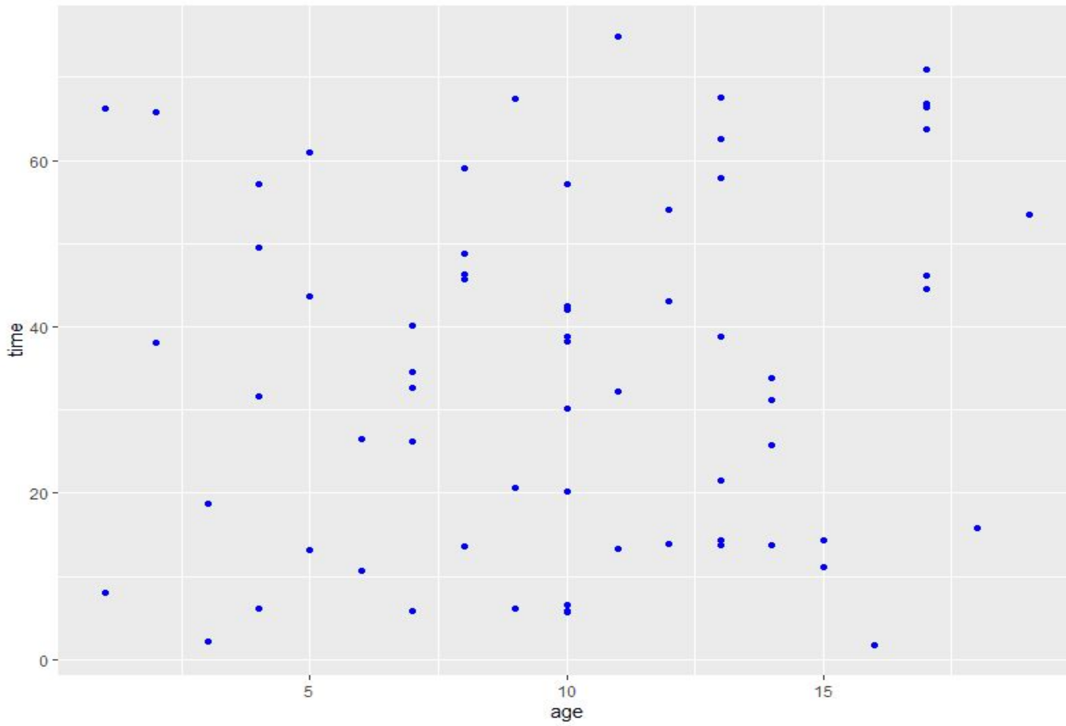


*Fig 6 - right eye treatment with xenon*

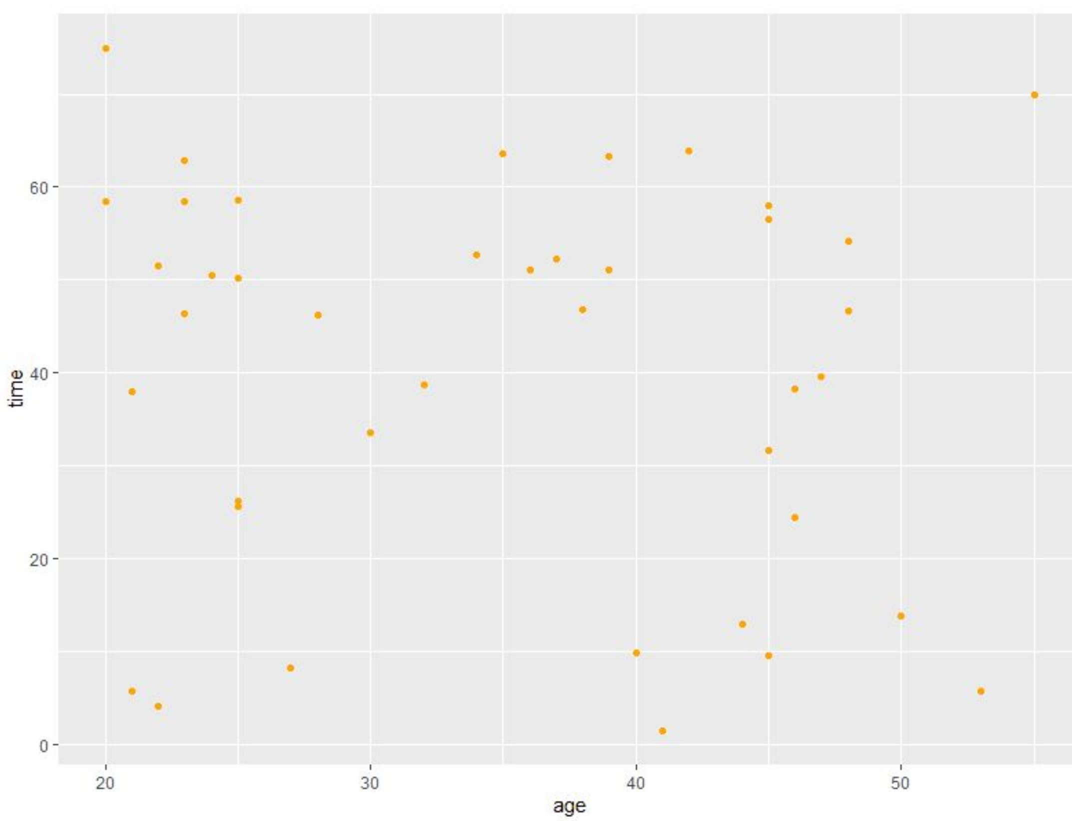

*Fig 7 - right eye treatment with argon*

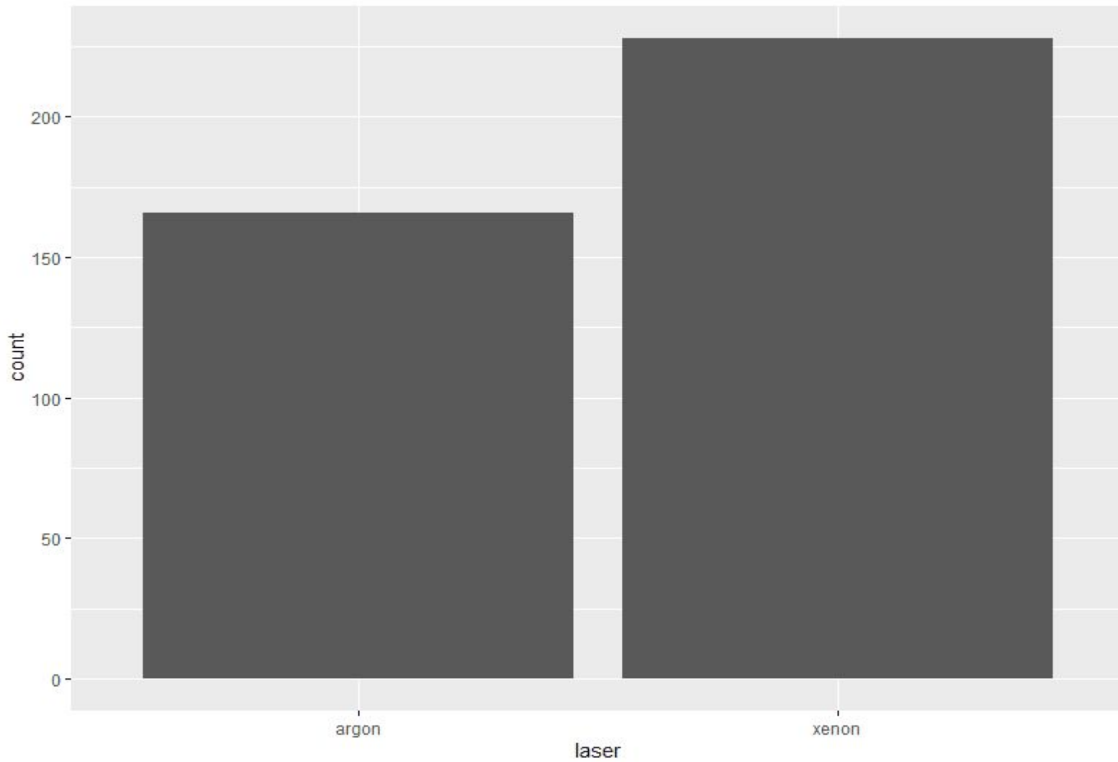


*Fig 8 - Distribution of laser treatments*

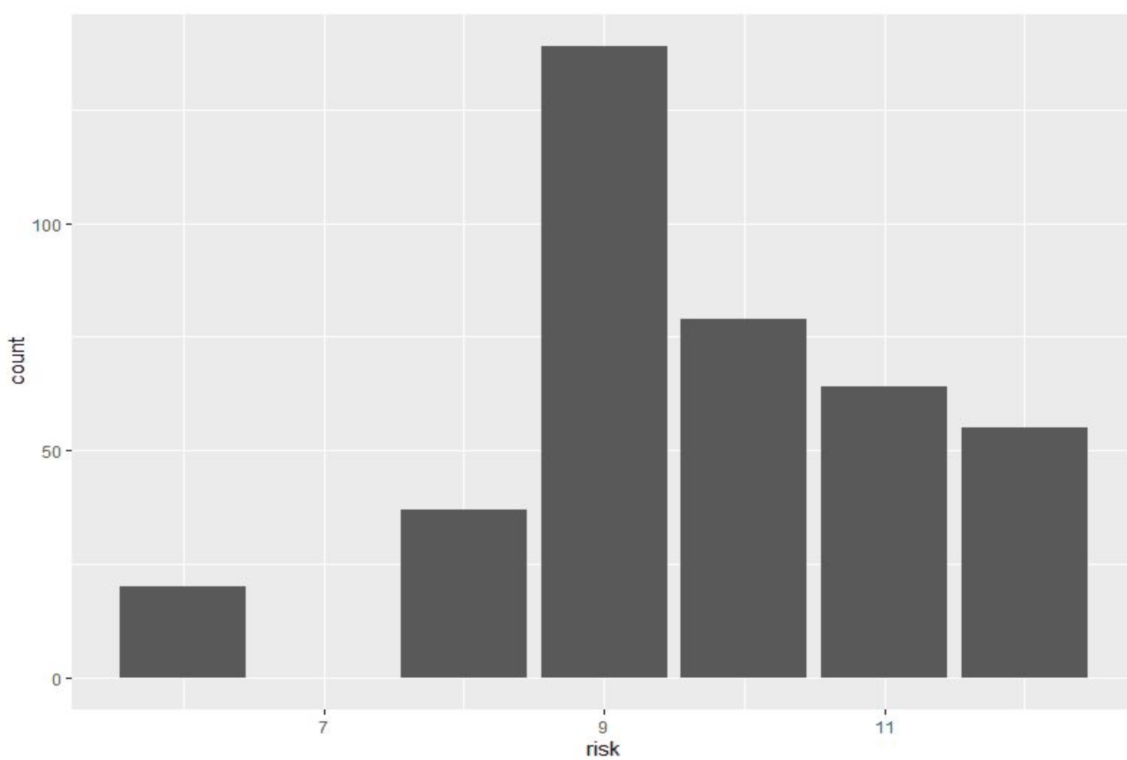

*Fig 9 - Distribution of risk factors*

| DataFrame (age vs time) | Correlation |
|---|---|
| 1 | -0.002297465 |
| 2 (Status = 0) | 0.02776837 |
| 3 (Status = 1) | 0.02565895 |
| 4 (Xenon) | 0.07083347 |
| 5 (Argon) | -0.1526307 |
| 6 (Left eye Argon treatment) | -0.1122114 |
| 7 (Left eye Xenon treatment) | -0.04304723 |
| 8 (Right eye Argon treatment) | 0.1143747 |
| 9 (Right eye Argon treatment) | -0.1209106 |

*Fig 10 - Table of correlation values*

In this project, we had attempted to see using a Diabetic Retinopathy dataset the effectiveness of laser treatment in preventing loss of vision, a common complication of Diabetic Retinopathy patients. We see that patients who didn't have vision loss had high vision rates for eyes on average as seen in *Fig 2,* where a massive cluster is present. In addition, both xenon and argon lasers were effective in ensuring high vision rates for both eyes, as xenon treatment had almost 68% of the sample having a vision time over 20 months while argon treatment had around 72% of the sample having vision time of over 20 months. However, data subjects received comparatively less argon treatments in comparison to xenon treatments, as seen in *Fig 8*. This sample was effective in having patients with high risk, as the risk factor distribution shown in *Fig 9* is unimodal with peak 9 and skewed to the left. We considered the age as another result in order to ensure that it wasn't a confounding variable. However, the association between age and time wasn't very high, as we can see from several of the correlation values in *Fig 10*. More efficient sampling is necessary to see how the age affects survival rates, since the number of data subjects over 50 years of age is only 11(5% of the sample).

Several sources of error presented themselves in the project. One was the existence of other retinal diseases that could cause the deterioration of vision. Other retinal conditions such as macular degeneration and others would have an effect on the vision and hence the survival times, which isn't accounted for in the sampling process of this dataset. Another source of error would be data subjects leaving the study, and hence having status value of 0. This ultimately would have an effect on the data since survival times would be augmented due to various data subjects being more involved within the analysis.

## Conclusion

In this project, we sought to analyze the efficiency of laser coagulation treatment on patients with Diabetic Retinopathy. We analyzed different types of laser coagulation such as xenon and argon and we analyzed the efficiency of these laser treatments using a Kaggle dataset. In the project, we had utilized various tools such as the R programming language to conduct my analysis. I was able to find that laser coagulation was extremely effective in preventing vision loss, one of the most common complications associated with Diabetic Retinopathy. In the future, I would like to see the effectiveness of laser treatment and other such treatments on other retinal conditions such as Macular Degeneration. I believe that this research is pressing as Diabetic Retinopathy has been a leading problem that is affecting many people across the world and analyzing how effective current treatments are is crucial in ensuring that Diabetic Retinopathy and other such conditions are mitigated as well as allow for new, novel treatments to arise.

## References


1. Ashfaque, J. M. (2019). Diabetic retinopathy. Retrieved from https://www.kaggle.com/ukveteran/diabetic-retinopathy
2. Diabetic Retinopathy. (n.d.). Retrieved from https://www.mayoclinic.org/diseases-conditions/diabetic-retinopathy/symptoms-causes/syc-20371611
3. Laser photocoagulation. (n.d.). Retrieved from https://medlineplus.gov/ency/article/007664.htm#:~:text=Laser photocoagulation is eye surgery,or to intentionally cause scarring.
4. Diabetic Retinopathy. (n.d.). Retrieved from https://stat.ethz.ch/R-manual/R-devel/library/survival/html/retinopathy.html

5. Diabetic Retinopathy. (n.d.). Retrieved from https://rdrr.io/cran/survival/man/diabetic.html
6. Pearson Correlation. (n.d.). Retrieved from https://en.wikipedia.org/wiki/Pearson_correlation_coefficient
7. Correlation Test Between Two Variables in R. (n.d.). Retrieved from http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r