Final Report

AUTHOR
Maria Oljaca

Introduction

Background:

[Note: “Protein X” and “Protein Y” are used in place of actual protein names for data security purposes.]

This dataset originates from an experiment focused on stem cell therapy for bone regeneration, using a rat animal model. Prior research has established the effectiveness of stem cells transduced with Protein X in treating critical-sized bone defects. In rat studies, it has been established that five million cells is the standard dose needed to heal critical-sized bone defects.

This study serves a dual purpose: it seeks to ascertain whether Protein Y, an upstream protein of Protein X, can enhance the bone-healing capabilities of Protein X. Furthermore, it aims to determine whether a reduced dosage of three million cells can produce comparable bone-healing outcomes.

The study includes four distinct experimental groups: X+Y5M, X+Y3M, X5M, and X3M. Critical-sized femoral defects, created surgically in rats, were treated with either three or five million stem cells transduced with a virus encoding Protein X, or a combination of both Protein X and Y. Subsequently, the quality and quantity of bone formation in rat femurs were assessed through various methods, including biomechanical strength testing, the results of which are contained in this dataset. The key variables measured with biomechanical testing are max torque (N), torsional stiffness (Nm/deg), total energy to failure (Nm*deg), and rotation at failure or max displacement (deg). Torsional stiffness is considered to be the most important variable in measuring the integrity of regenerated bone.

Question:

Which experimental group showed the greatest levels of bone regeneration, as measured by torsional stiffness, max torque, total energy to failure, and max displacement?

Methods

The dataset used for this project was obtained from a colleague who recently conducted the above experiment. I received the results of the biomechanical testing in an Excel file via email and read them in using the readxl R package.

For data exploration and wrangling, I mainly used base R and the dplyr package within tidyverse. Kable, ggplot2, and cowplot were used to display summary statistics and visualize the data.

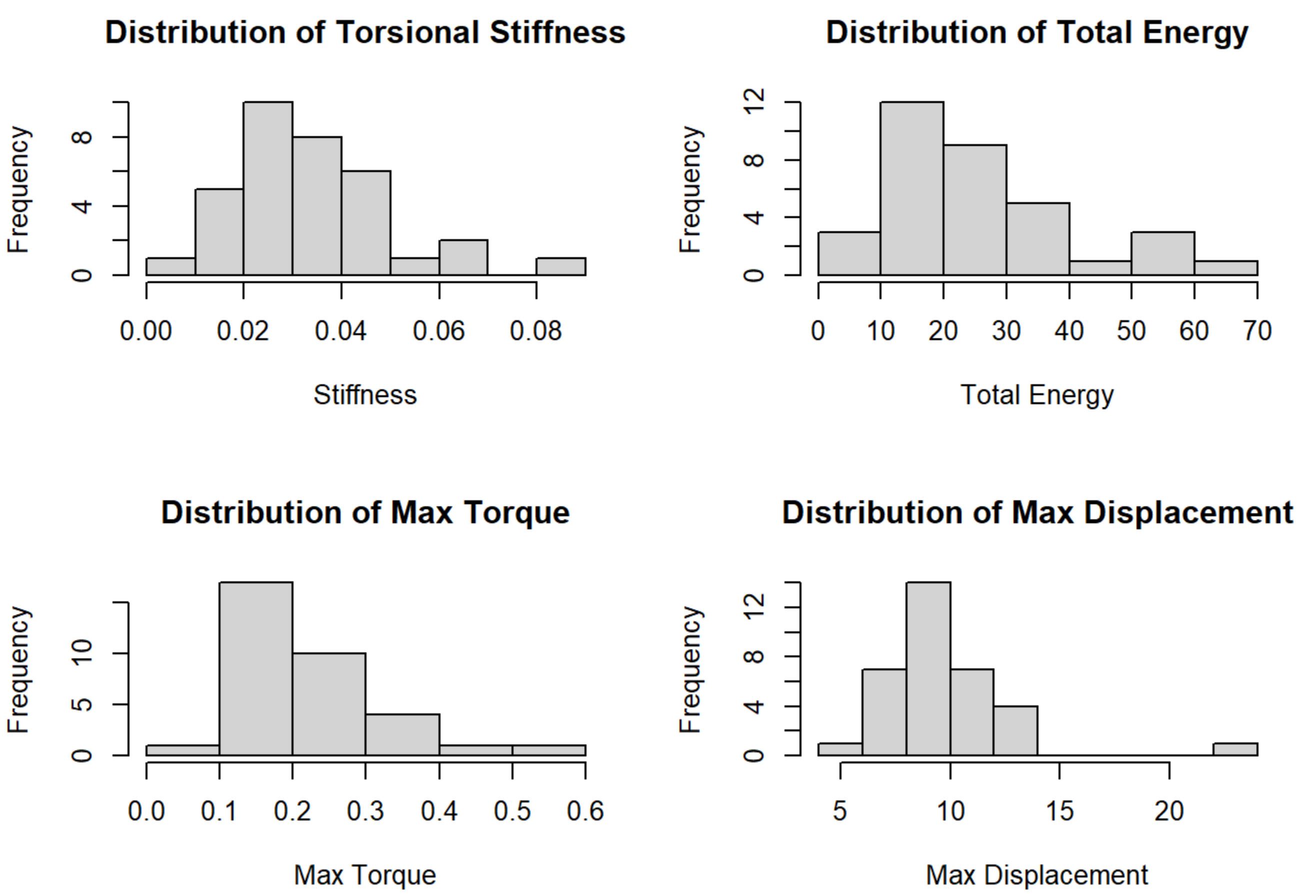
First, I evaluated the data's integrity using the "dim," "head," and "tail" functions. These actions showed that there were no apparent import issues and that the data set contains 34 rows and 5 columns: SID, Stiffness, Total Energy, Max Torque, and Max Displacement. Next, I checked the variables using the "str" function, which revealed no major issues and variable types as expected.

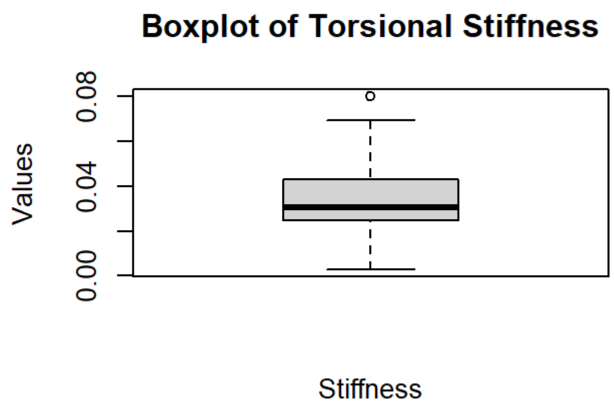
The imported dataset did not include a column for experimental groups. However, that information was recorded in another document provided to me by my colleague. I created a new data frame mapping each observation to its correct experimental group, and then used the "left_join" function to join this data frame to my dataset (xy_bmech) by "SID."

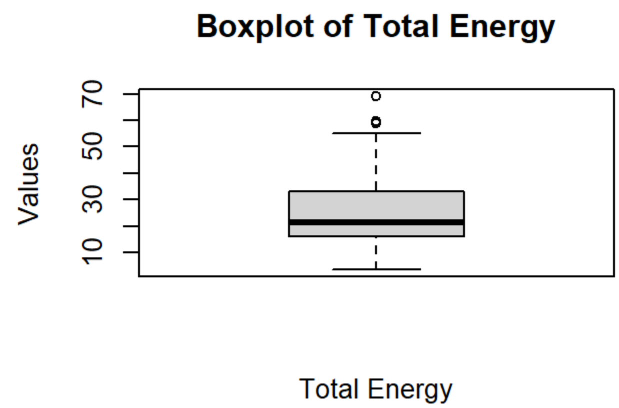
Next, I renamed some variables for ease of coding using the "rename" function.

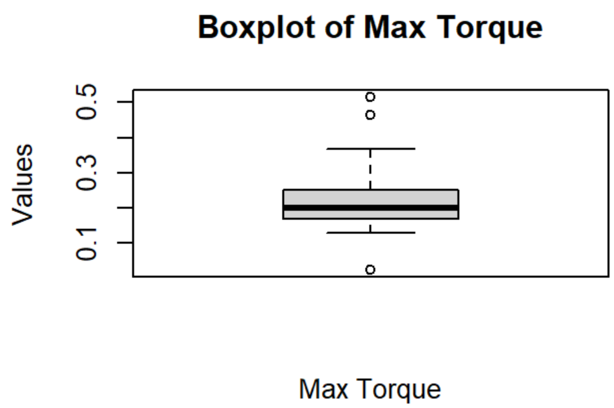
Returning to the EDA checklist, I used the "summary" and "table" functions to take a closer look at the key variables, which showed 9 observations per experimental group. There also seemed to be a wide range of values for each variable (although plausible), and "mean(is.na())" showed there to be 3.7% missing data.

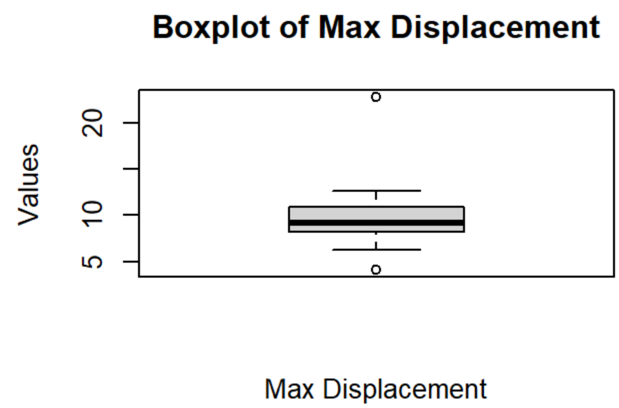
Histograms and boxplots revealed that most of the key biomechanical variables had a right-skewed distribution in addition to two or more outliers, which indicated that I should use median and IQR to summarize my data rather than the mean and standard deviation.











Per my colleague, a priori power analysis for this experiment had determined that 9 animals per group were needed, for a total of 36. However, there were 34 in this experiment because one rat was mislabeled and the other was not strong enough to undergo biomechanical testing.

Although 9 animals per group are needed for adequate statistical power, I chose to remove the missing values to avoid creating false positives, and given that less than 5% of the data is missing.

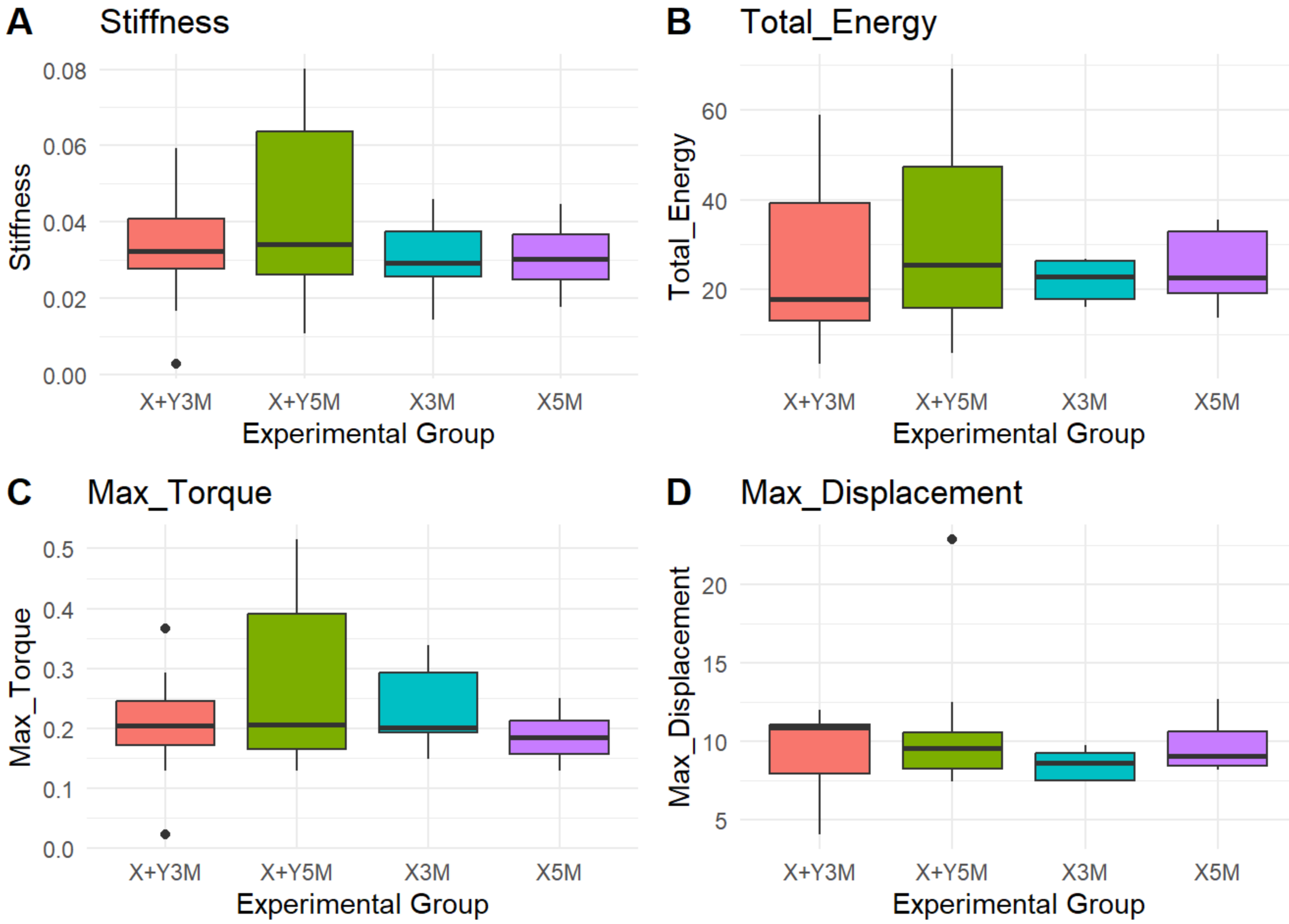
Preliminary Results

Summary Statistics

Experimental Group	Stiffness (Nm/deg)	Total Energy (Nm*deg)	Max Torque (N)	Max Displacement (deg)
X+Y3M	0.0324 (0.0129)	17.78 (26.22)	0.2041 (0.0742)	10.89 (3.17)
X+Y5M	0.0341 (0.0375)	25.46 (31.38)	0.2064 (0.2261)	9.53 (2.32)
X3M	0.0292 (0.012)	22.95 (8.54)	0.2018 (0.1009)	8.6 (1.76)
X5M	0.0301 (0.0118)	22.56 (13.84)	0.1855 (0.0557)	9.09 (2.22)

[a] Values represent median (IQR)

Data Visualization



Conclusion

Observing the summary table and boxplots, it appears that bone in the X+Y5M group performed the best on most measures of biomechanical strength, including torsional stiffness, total energy to failure, and max torque.

For the most important variable in measuring bone integrity, torsional stiffness, X+Y5M minimally outperformed X+Y3M (median stiffnesses of 0.0341 Nm/deg vs 0.0324 Nm/deg, respectively). However, X+Y5M had much greater varability in torsional stiffness.

X+Y3M outperformed X+Y5M on rotation at failure, or max displacement. The X+Y5M group also generally had the greatest variability in markers of biomechanical strength other than max displacement.

In general, it also appears that both of the X+Y experimental groups outperformed the X only groups on measures of biomechanical strength other than total energy to failure, in which the X+Y3M group performed the worst.

On visualization, many of these differences appear minuscule. It is not possible to form a full conclusion on which experimental treatment was most successful at bone healing until analytic statistics are applied.