

Math 244: Final Report

AUTHOR

Mia Zottoli, Vishnu Lakshman, Julia Vitale

Introduction and Data

The data set includes statistics from the 2024 Division 1 and 3 Men's and Women's Ultimate Frisbee Championships. The statistics were found on USA Ultimate, the non-profit organization serving as the governing body for ultimate in the United States, and were taken from a data visualization titled "USA Ultimate 2024 Nationals Stats Dashboard", which was created by Ben Ayres. The data set includes 1665 rows which each correspond to an individual player, and it includes 15 variables which categorize the players by Division, Gender, and Team, and provide game statistics for each player.

Ultimate Frisbee is a sport growing in popularity at the collegiate level and within the Vassar student body as well. However, not much data analysis is available for Ultimate compared to other popular sports. We want to fit a model that would help players analyze their game performance. Therefore, via linear, LASSO, and ridge regression we will fit a prediction model that we train under supervised conditions to be able to predict a player's plus/minus score (AKA individual impact, calculated as the difference between the points scored and the points allowed while the player is on the court) based on the variables `turns_per_game`, `ds_per_game` and `ast_per_game`. These variables all relate to a player's effectiveness on the field, which is why we will use them to predict `pls_mns_per_game`. We will also stratify by division and gender, so that we ensure that we have an accurate model for each group. We will have four models in the final product, DI Men, DI Women, DIII Men, and DIII Women.

Since the data was already pretty clean, we did not have to do much data tidying. We just did some string manipulation to extract the school name from the team name, creating a new column called 'school', and had to convert some character variables to factors.

Methodology

Summary Statistics

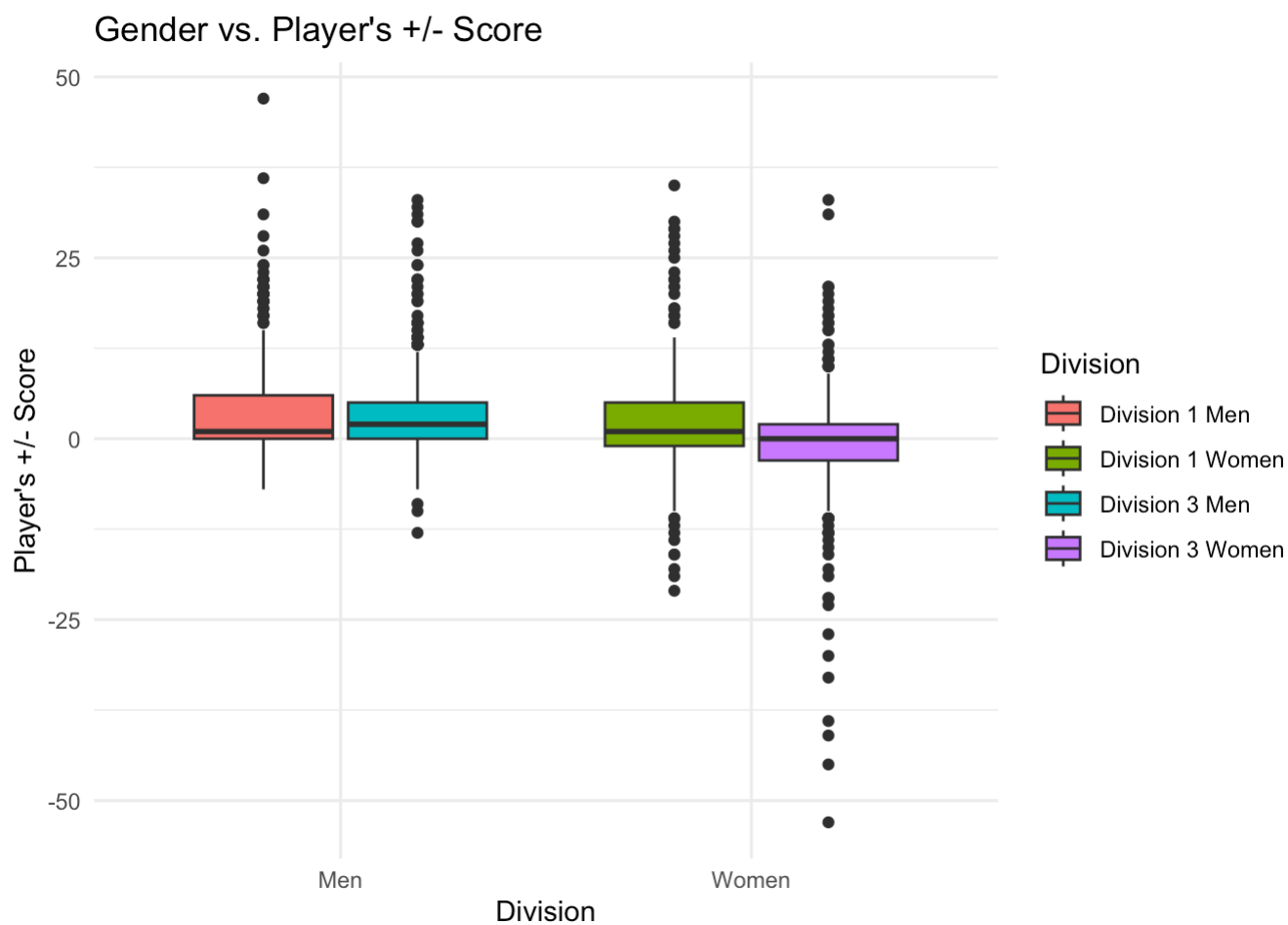
<code>pls_mns_per_game</code>	<code>turns_per_game</code>	<code>ds_per_game</code>	<code>ast_per_game</code>
Min. : -8.8333	Min. : 0.0000	Min. : 0.0000	Min. : 0.0000
1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000	1st Qu.: 0.0000
Median : 0.2000	Median : 0.4000	Median : 0.2000	Median : 0.1667
Mean : 0.4430	Mean : 0.9229	Mean : 0.3543	Mean : 0.5050
3rd Qu.: 0.8333	3rd Qu.: 1.1667	3rd Qu.: 0.5000	3rd Qu.: 0.6000
Max. : 5.8750	Max. : 14.8333	Max. : 7.1667	Max. : 7.6000

<code>pts_per_game</code>
Min. : 0.0000
1st Qu.: 0.0000
Median : 0.2000
Mean : 0.5066
3rd Qu.: 0.7143
Max. : 4.4000

	pls_mns_per_game	turns_per_game	ds_per_game	ast_per_game
pls_mns_per_game	1.0000000	-0.2626174	0.1892149	0.3830593
turns_per_game	-0.2626174	1.0000000	0.6852003	0.6550684
ds_per_game	0.1892149	0.6852003	1.0000000	0.5028222
ast_per_game	0.3830593	0.6550684	0.5028222	1.0000000
pts_per_game	0.6805235	0.2856562	0.4187957	0.4498837

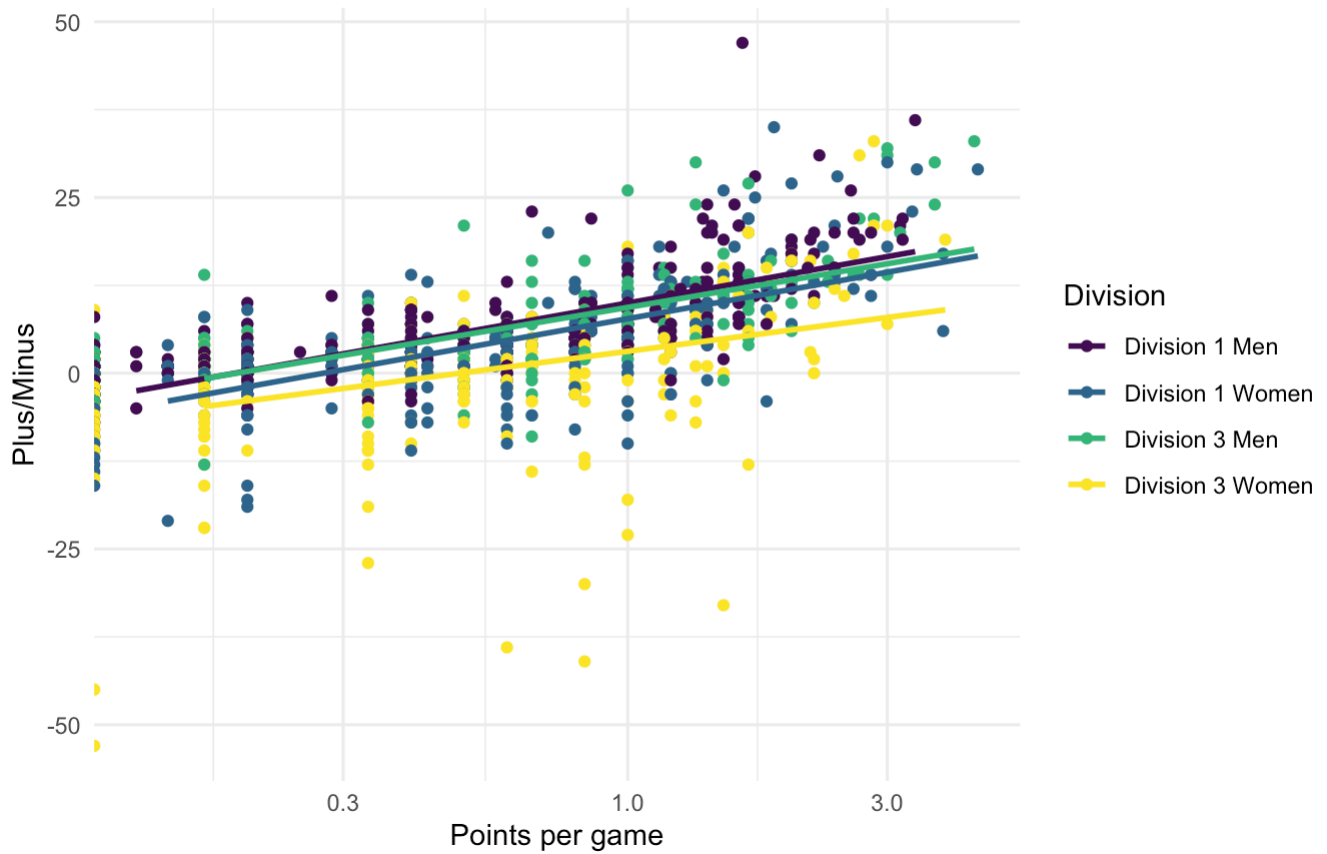
	pts_per_game
pls_mns_per_game	0.6805235
turns_per_game	0.2856562
ds_per_game	0.4187957
ast_per_game	0.4498837
pts_per_game	1.0000000

Visualizations/EDA



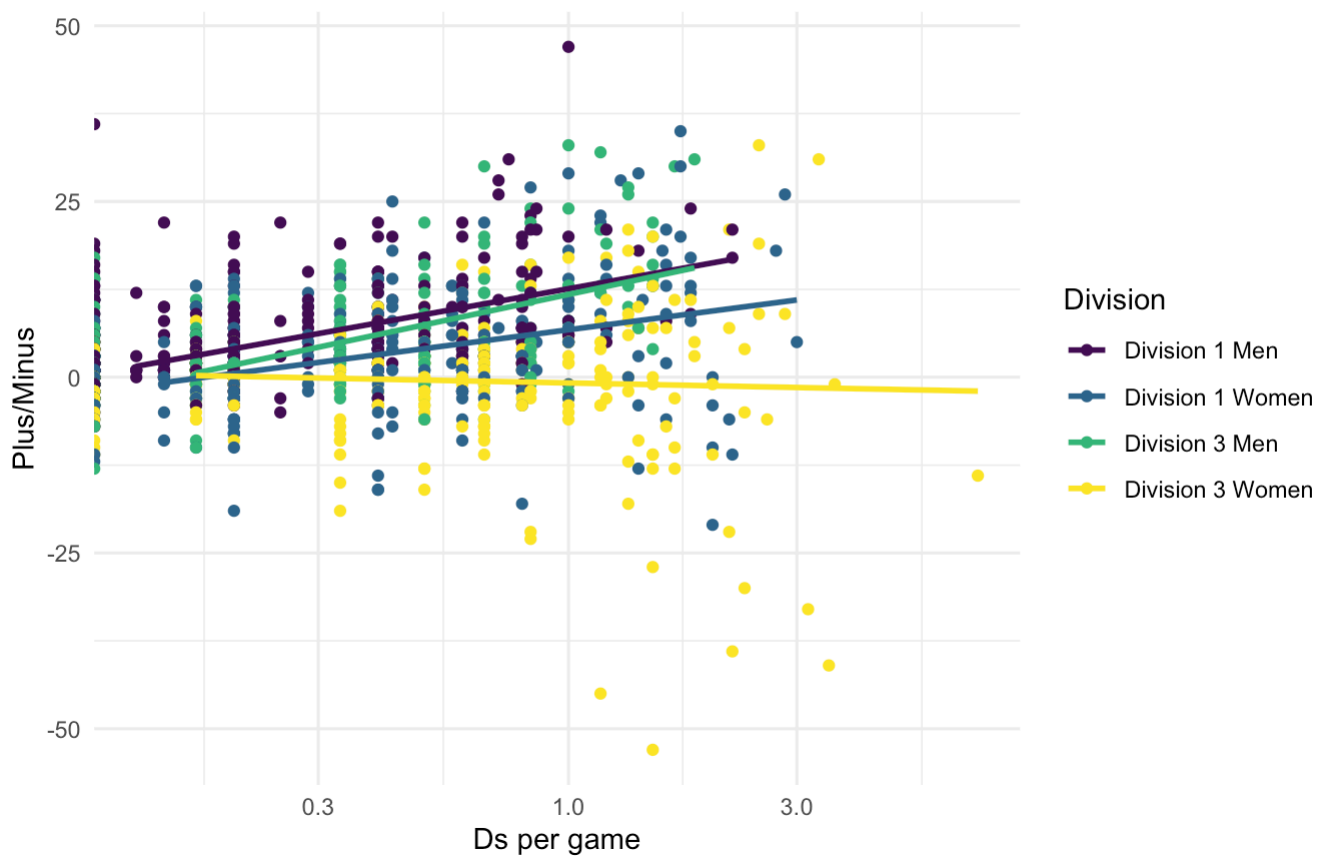
Points per game vs. player impact

By division and gender



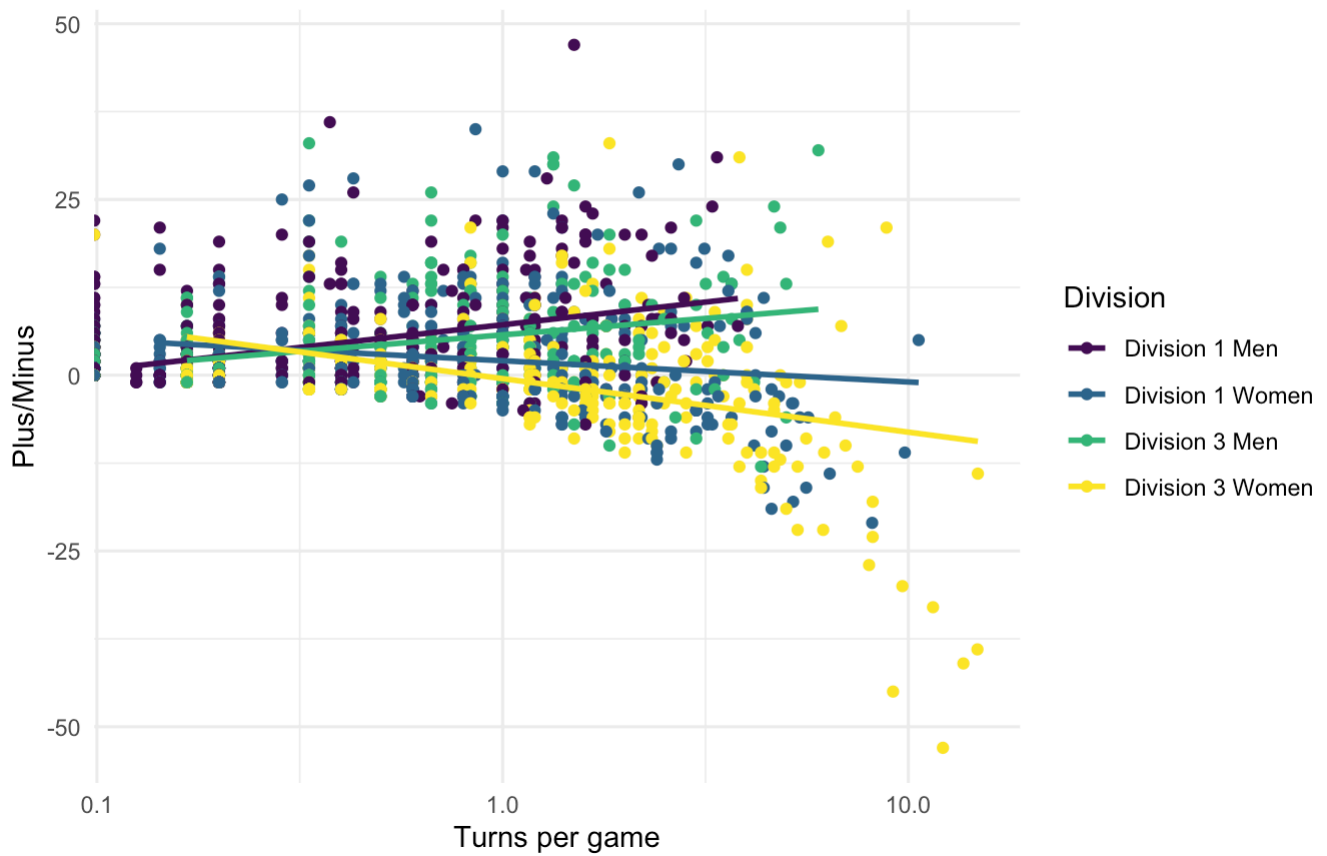
Ds per game vs. player impact

By division and gender



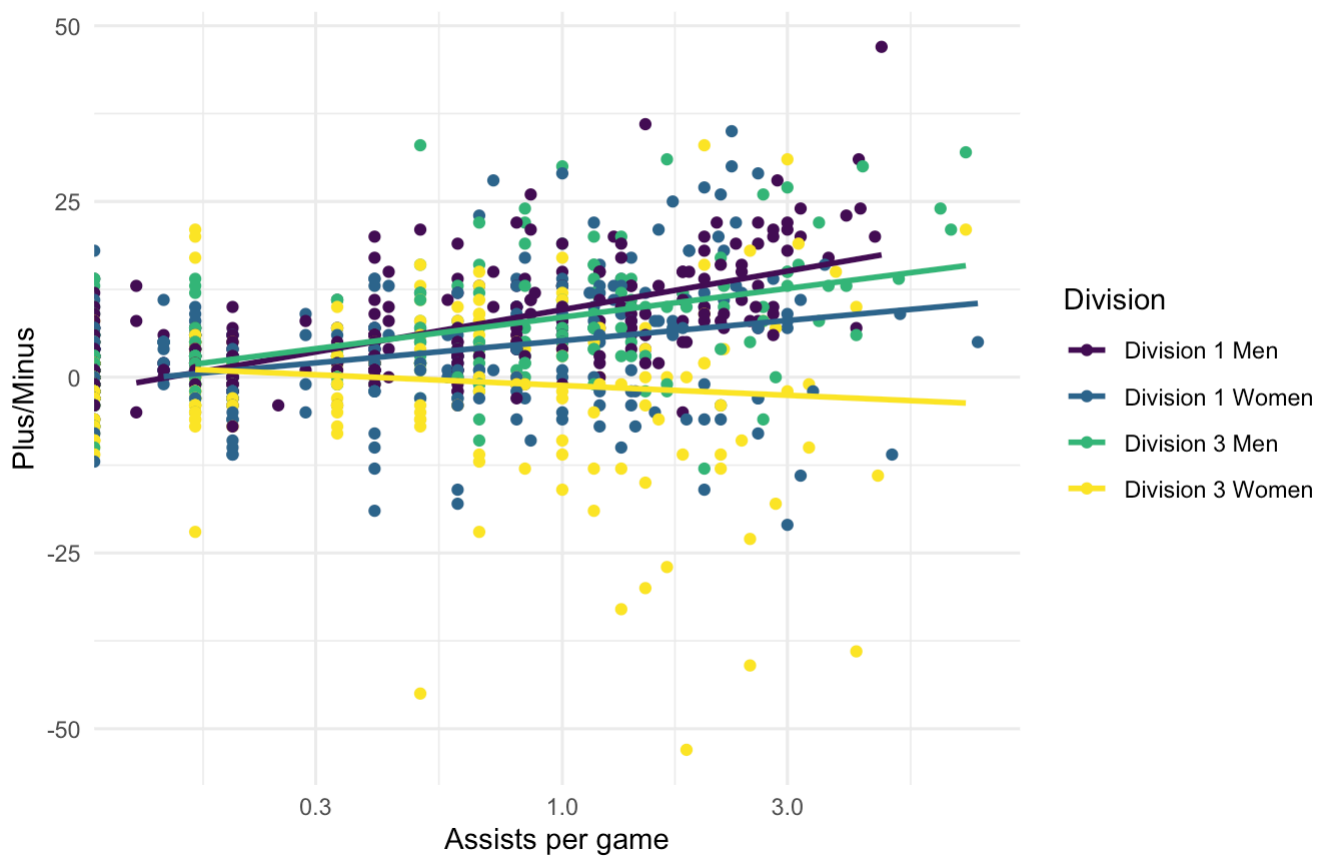
Turns per game vs. player impact

By division and gender



Assists per game vs. player impact

By division and gender



Modeling methodology

We are fitting all three models that predict continuous variables: linear, LASSO, and ridge regression. Based on the results, we will select the model that is most accurate. From the scatter plots for our four predictor variables against the outcome variable of +/-, we see a fairly linear relationship between the variables. For pts_per_game, all of the relationships are positively linear. For ds_per_game and ast_per_game, all of the relationships are positive except for Division 3 Women. Additionally, for turns_per_game, the regression lines for both Division 1 and 3 have negative slopes. These differences, along with the slight variations in the box plots for player's +/- score by Division and Gender, suggest that these predictors act differently by Division and Gender. Therefore, we will fit the model for each combination of division and gender, rather than utilizing a blanket model for all types of players.

Results

Linear Regression

Division 1 Men – Mean R^2 (CV): 0.799 (n=521)

Coefficients:

turns_per_game: -0.634

ds_per_game: 0.450

ast_per_game: 1.106

Intercept: 0.708

Division 1 Women – Mean R^2 (CV): 0.753 (n=452)

Coefficients:

turns_per_game: -1.620

ds_per_game: 0.804

ast_per_game: 1.074

Intercept: 0.321

Division 3 Men – Mean R^2 (CV): 0.707 (n=372)

Coefficients:

turns_per_game: -0.935

ds_per_game: 0.663

ast_per_game: 1.063

Intercept: 0.604

Division 3 Women – Mean R^2 (CV): 0.788 (n=320)

Coefficients:

turns_per_game: -2.516

ds_per_game: 1.034

ast_per_game: 1.113

Intercept: -0.105

Ridge Regression

Division 1 Men – Mean R^2 (CV): 0.799 (n=521)

Coefficients:

turns_per_game: -0.633

ds_per_game: 0.450

ast_per_game: 1.105
Intercept: 0.708

Division 1 Women – Mean R^2 (CV): 0.754 (n=452)

Coefficients:
turns_per_game: -1.618
ds_per_game: 0.803
ast_per_game: 1.072
Intercept: 0.321

Division 3 Men – Mean R^2 (CV): 0.707 (n=372)

Coefficients:
turns_per_game: -0.932
ds_per_game: 0.663
ast_per_game: 1.061
Intercept: 0.604

Division 3 Women – Mean R^2 (CV): 0.788 (n=320)

Coefficients:
turns_per_game: -2.512
ds_per_game: 1.032
ast_per_game: 1.111
Intercept: -0.106

Lasso Regression

Function grouped by division and gender

Division 1 Men – Mean R^2 (CV): 0.799 (n=521)

Coefficients:
turns_per_game: -0.629
ds_per_game: 0.448
ast_per_game: 1.101
Intercept: 0.708

Division 1 Women – Mean R^2 (CV): 0.754 (n=452)

Coefficients:
turns_per_game: -1.615
ds_per_game: 0.802
ast_per_game: 1.070
Intercept: 0.321

Division 3 Men – Mean R^2 (CV): 0.707 (n=372)

Coefficients:
turns_per_game: -0.928
ds_per_game: 0.661
ast_per_game: 1.057
Intercept: 0.604

Division 3 Women – Mean R^2 (CV): 0.782 (n=320)

Coefficients:
turns_per_game: -2.461

```
ds_per_game: 1.006
ast_per_game: 1.075
Intercept: -0.107
```

Cross-validated R-squared values				
level	gender	Linear Regression	Lasso	Ridge
Division 1	Men	0.7987	0.7988	0.7988
Division 1	Women	0.7535	0.7539	0.7537
Division 3	Men	0.7069	0.7067	0.7069
Division 3	Women	0.7884	0.7816	0.7881

Prediction Model

We can use our model to predict player impact based on historical performance statistics. For example, we can predict the plus/minus score of a Division 1 Male athlete with an average of 2.5 turns per game, 1.0 Ds per game, and 4.0 assists per game.

ridge_prediction	linear_prediction	lasso_prediction
4.626153	4.627888	4.620286

```
[1] "Predictions for Division 1 Male Player:"
```

```
[1] "Ridge Prediction: 4.626"
```

```
[1] "Linear Prediction: 4.628"
```

```
[1] "Lasso Prediction: 4.620"
```

All three models predict that this player has a plus/minus of approximately 4.6 (pretty good!). The models behave similarly, but have slight variations in prediction.

Discussion and Conclusion

In this project, we explored the feasibility of building a predictive model to estimate a player's plus-minus score in college Ultimate Frisbee based on individual game statistics: turnovers per game, defensive blocks per game, and assists per game. We implemented three distinct regression models—Linear Regression, Ridge Regression, and Lasso Regression—to evaluate how well these features could predict the target outcome.

All three models (**Linear, Lasso and Ridge**) produced very similar predictions across multiple gender-division combinations, suggesting that the underlying relationships in the data are fairly linear and do not require strong regularization to prevent overfitting.

It should be mentioned that during the time of our preliminary analysis and presentation, our R^2 value was 1 in most cases, and close to 1 (e.g., 0.997) in the remaining cases, across all three models and categories. We later were able to identify that this was due to the inclusion of the explanatory variable `pts_per_game`, which we learnt was used directly in calculating the outcome variable. To address this, we excluded `pts_per_game` from the model and reran the analysis. After its removal, all subgroups across the models produced more realistic R^2 values, ranging from 0.7 to 0.8.

Additionally, in the preliminary analysis the **Lasso Regression** model was under performing. It was predicting plus-minus scores of 0.68 when the other two models, Linear and Ridge, were predicting 8.28 and 8.30 respectively. However, this was fixed through using cross-validation to find the best alpha (it was too high).

In summary, our results suggest that individual game statistics like turnovers, blocks, and assists can reasonably predict a player's plus-minus score using standard regression techniques. After refining the feature set and tuning model parameters through cross-validation, we achieved consistent and realistic performance across all models.

Methodological Evaluation

To ensure consistent model evaluation, we split the dataset by both **division (D-I vs. D-III)** and **gender (men's vs. women's)**, and trained separate models for each subgroup. This decision was grounded in the understanding that the dynamics of ultimate frisbee can vary substantially across these categories, and that a model trained on pooled data might obscure those differences.

Limitations

1. **Small feature set:** Only three predictor variables were used. While these are intuitively related to performance, plus-minus is inherently a team-dependent metric influenced by many contextual factors (e.g., opponent strength, line composition, playing time) that are not captured here.
2. **Potential data imbalance:** The performance of the models across different (division, gender) groups could vary depending on how many observations we had per group. If some subgroups had fewer examples, model stability would be affected.
3. **No non-linear modeling:** While most of the relationships we observe appear somewhat linear in exploratory analysis, fitting a nonlinear model, such as a random forest, may be a better analysis route.

Data Reliability and Validity

The dataset was sourced from a publicly available website and published within the past year. While not a first-party dataset, its reliability is bolstered by several key factors:

1. It reflects official statistics from the most recent USA Ultimate College National Championships, meaning it's drawn from the highest college-level competition.
2. As active Frisbee players ourselves, we were able to recognize nearly all the teams listed and even some individual players, affirming the data's authenticity and face validity.
3. The dataset has had multiple downloads from the site, indicating community engagement and a degree of scrutiny.

Appropriateness of the Analysis

Our analytical choices—using regularized regression methods and controlling for gender and division—are appropriate given our objective and dataset size. We standardized our features using `StandardScaler()` and implemented a consistent 80/20 train-test split across all models (as shown in the `lasso_reg` function), which helped avoid data leakage and ensured fair comparison between models.

Using multiple models allowed us to assess the sensitivity of plus-minus predictions to different regularization schemes. Ridge, in particular, proved effective in slightly dampening noisy relationships while preserving feature contributions.

Ethical Considerations

Benefits:

- **Enhanced Performance Insights:** Our model can provide players and coaches with data-driven insights, potentially informing training and strategy.
- **Objective Evaluation:** By quantifying performance metrics, we aim to reduce subjective biases in player assessments.

Risks:

- **Privacy Concerns:** Even though our data is publicly available, using it for predictive modeling raises questions about consent and the extent to which athletes are aware of or agree to such analyses.
- **Potential for Misuse:** There's a risk that our model could be used to make decisions about player selection or playing time without considering the broader context of an athlete's performance or potential.
- **Bias and Fairness:** If our model inadvertently reflects existing biases in the data, it could perpetuate inequalities, especially if used in decision-making processes.

Overall, this project demonstrates the potential of data-driven approaches to enhance our understanding of individual performance in Ultimate Frisbee. While our models show promise within the scope of elite college-level play, future work must address broader data inclusion to effective application.

Given a second chance what we would have done differently

If we were to restart the project, we would have done a far more a rigorous analysis of the explanatory variables used. While all the variables currently chosen did show evidence affecting the target variable, we choose them based on our advanced knowledge and experience playing the sport and not through empirical means. It would strengthen the accuracy and reliability of our models results if could empirically confirm that these four are the most influential variables out of all the possible options in the dataset. To empirically test would include calculations such as making a correlation heatmap of all the explanatory variables against the outcome variable. Also, there could be a lot to gain if we modified our outcome variable so that it could be more informative. The plus-minus statistic is taken from basketball and while useful in Frisbee as well, it has its limitations. One of them is that in most Frisbee games there is an Offensive-line and Defensive-line with totally

different players on each line. And the team on offensive has a far greater probability of scoring. Therefore, there could be some really good players on the Defensive-line who keep getting scored on and this statistic would unfairly represent them as worse than Offensive-line players even if that may not necessarily be the case.