EDLD 654 Final Project: Fiction Transportation Modeling Proof of Concept

Eliott Doyle[1]

[1] University of Oregon

EDLD 654 Final Project: Fiction Transportation Modeling Proof of Concept

Research Problem

Last year I collected data for my master's project, in which participants in three experimental conditions wrote short stories, letters, or journal entries and responded to some surveys and demographic questions. Subsequently, several research assistants coded these writing samples for *transportation*: the extent to which the reader (i.e., the coder) feels that they were immersed or swept up in the text.

Transportation is one of the main operationalizations of audience involvement when reading a text; the degree to which someone experiences transportation while reading generally predicts how much they are emotionally affected by the text, how much they identify with the characters, and how much they are persuaded by the text's message (Green, Brock, & Kaufman, 2004).

I am interested in how author features contribute to their ability to write a text that is transporting to readers. My research more broadly focuses on empathic accuracy, or, how well people are able to understand what another person is feeling during interactions. The manner in which the target (or, person being perceived, who the perceiver is attempting to empathize with) expresses themselves affects how well the perceived is able to pick up on their thoughts and feelings, and features like how invested the target is in what they are communicating, experience communicating with others, and emotional expressivity can influence how easy their communication is to parse for a perceiver. While transportation does not directly tie to empathic accuracy, it is related, in that it is a measure of how compelled the perceiver is by the target's words.

Most past research on transportation has focused on the features of the perceiver/audience rather than of the author. I am interested in investigating author features. Identifying characteristics of authors that make them better at writing texts that produce

high transportation in others will help illuminate what makes communication more effective. Being able to model the transportation scores of texts based on the text and characteristics of the author will be informative in my understanding of what aspects of authors' experience while writing leads to more effective writing that has an impact on readers/perceivers.

Description of the Data

264 University of Oregon undergraduates participated in this study and produced data that was usable for the current analyses. The sample size was adequate for the ANOVA and linear regression analyses I conducted for hypothesis testing, but because it is relatively small for machine learning purposes, I will be considering this a proof of concept for future applications similar to this, rather than a definitive model for predicting transportation.

Predicted variable: Transportation. Seven research assistants coded the writing samples for transportation on a Likert scale ranging from 1 (no transportation) to 5 (very high transportation). The mean scores from all seven research assistants make up the transportation variable.

Distribution of Observations by Transportation Score
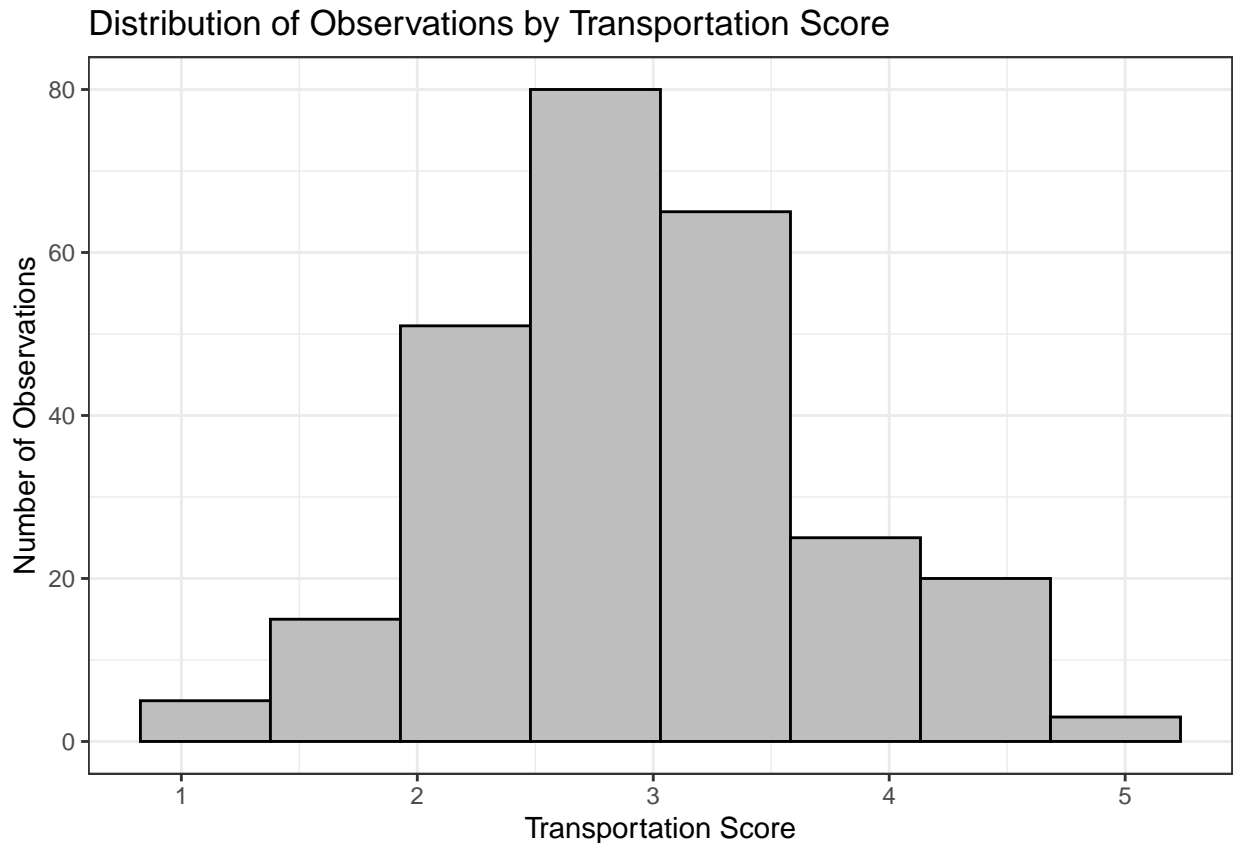
Writing samples. The writing samples were personal in nature and so are confidential. I have gotten the embeddings from the samples without the samples themselves, and saved those for the purposes of this project.

Writing experience. Authors reported how much experience they have with writing fiction and journaling, within the past year and prior to the past year when they participated in the study. These were four separate questions that here will be treated as one measure with mean scores for the four questions representing writing experience. 1 indicates no writing experience; 3 indicates frequent and long-time writing experience in different forms. Mean writing experience was 1.610 with standard deviation 0.434, indicating that authors generally had some writing experience but not a lot, with a lot of variability.

Distribution of Observations by Writing Experience
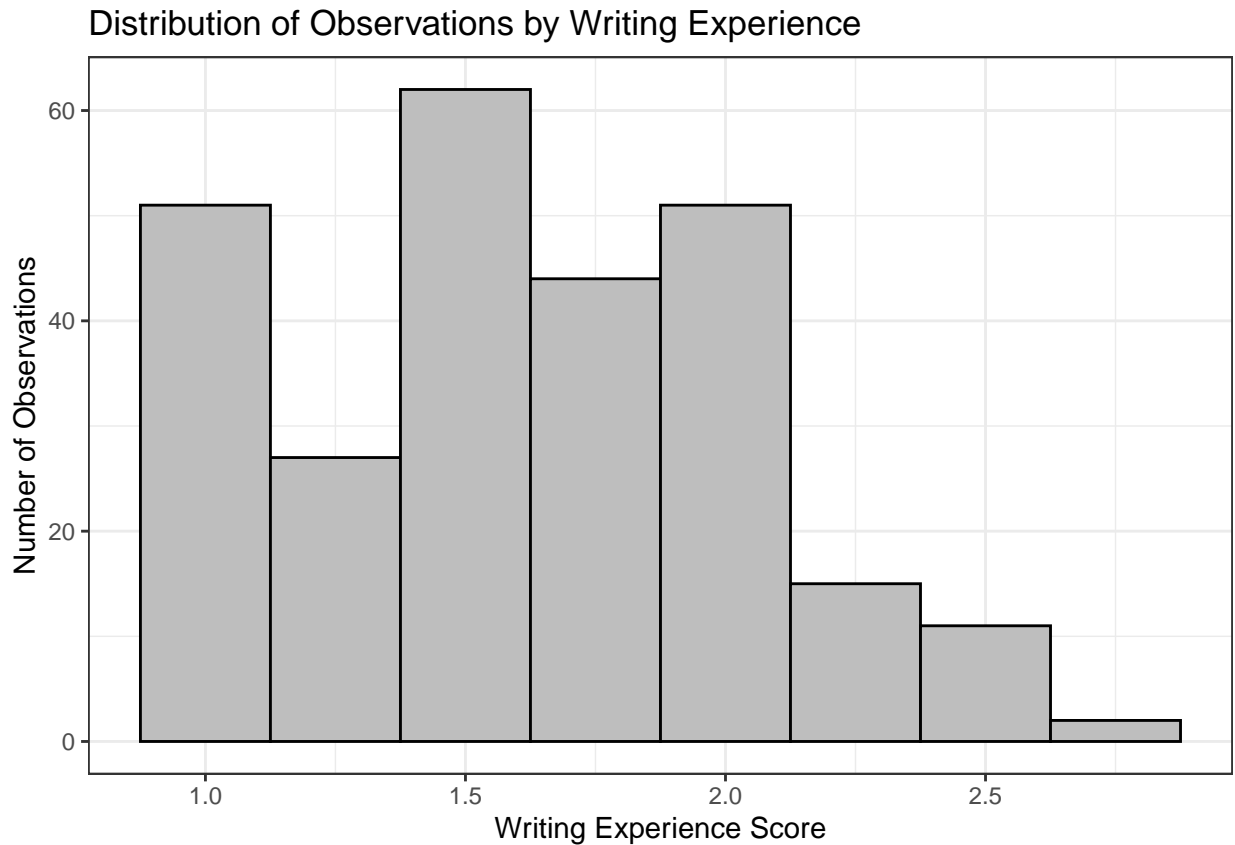
Interpersonal Reactivity Index (IRI). The IRI (Davis, 1983) is a popular measure of trait empathy. The authors of the writing samples completed this survey, so this is a measure of author trait empathy, not perceiver empathy. The IRI comprises 28 questions; the mean score for each other is reported with 1 being very low empathy and 5 being very high empathy. Mean IRI score was 3.468, with SD = 0.444. The distribution of responses was normal.

Distribution of Observations per Interpersonal Reactivity Index (IRI) Score
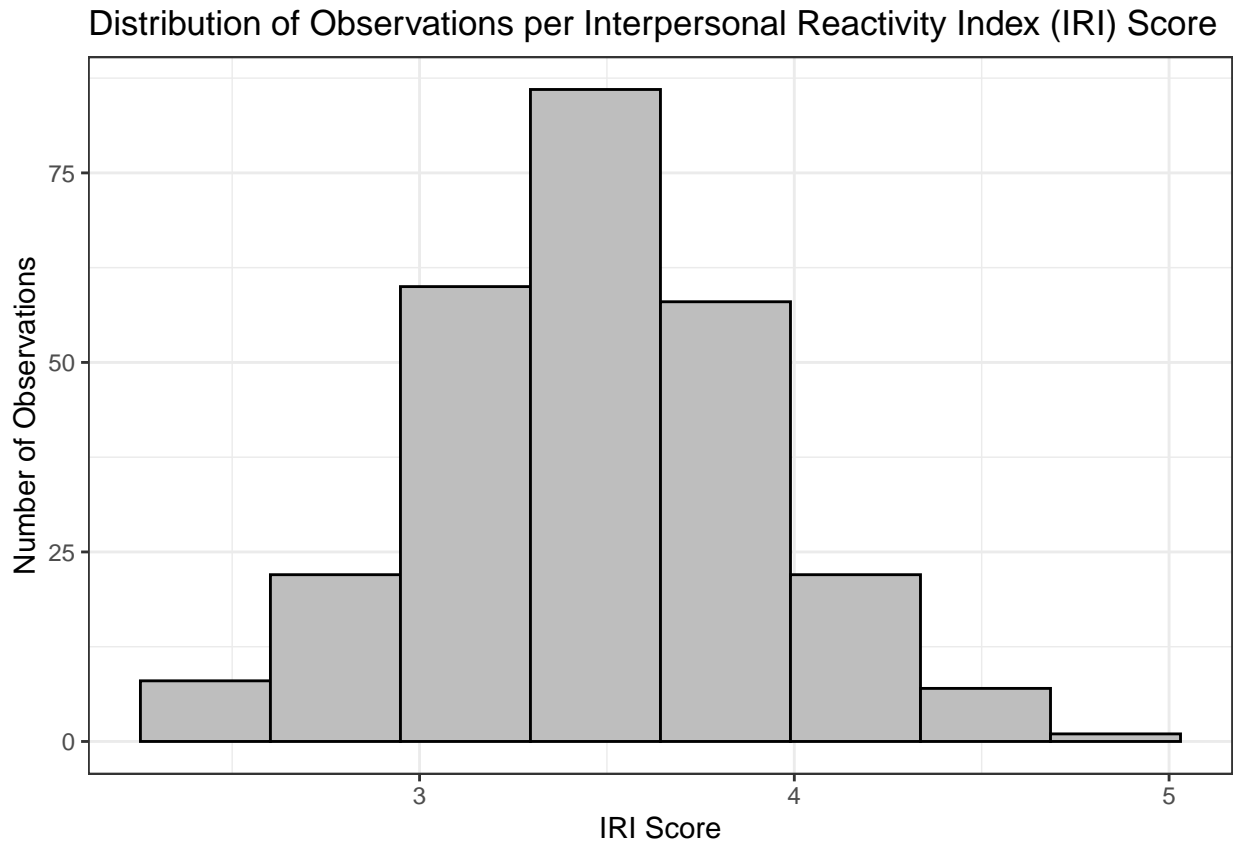
Character liking. Authors rated how much they liked the characters they were writing about on a measure of liking comprising 9 questions. The mean score for each author is reported with 1 being strong dislike for the character and 5 being strong liking for the character. Mean character liking was 3.566 with SD = 0.771. Liking was fairly high presumably because authors were instructed to choose their favorite fictional character.

Distribution of Observations per Score on Character Liking
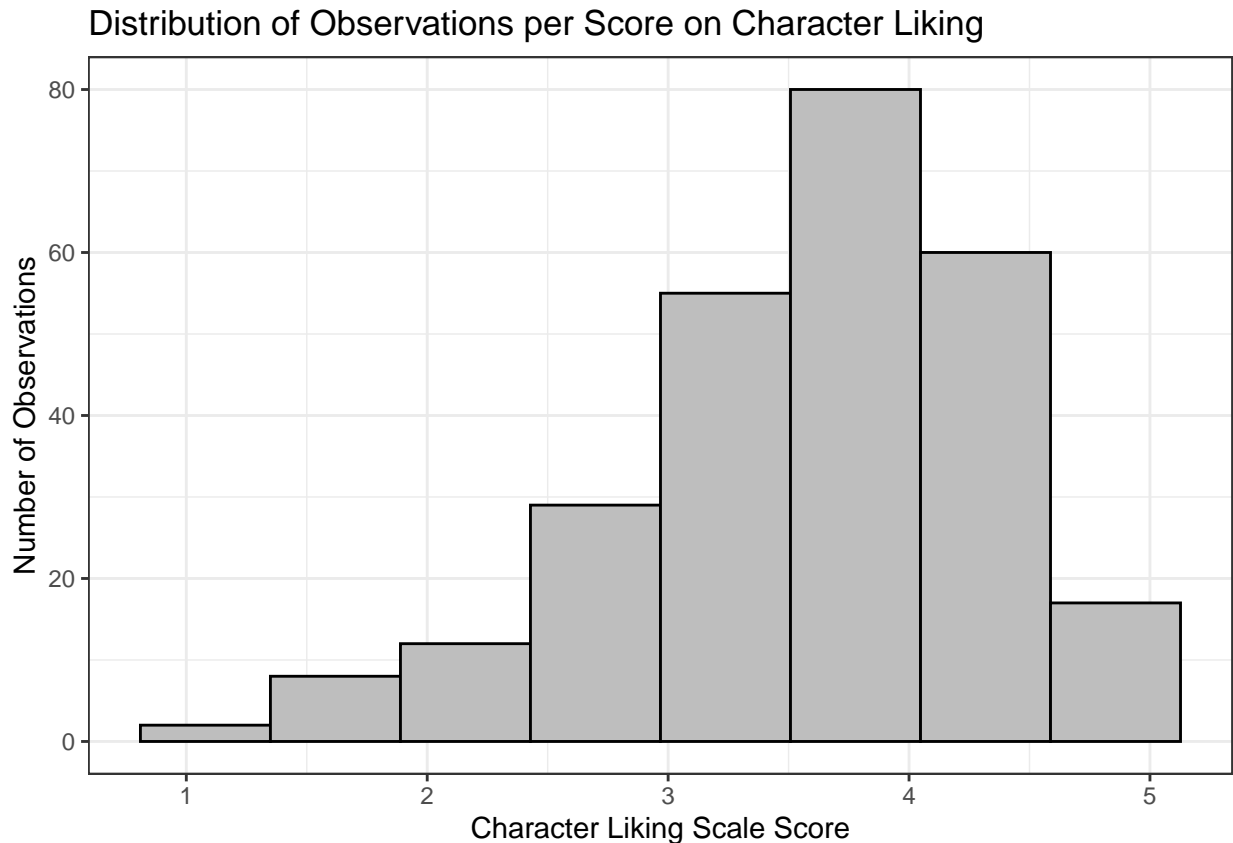
Missingness.

These data have already been cleaned, so there are not many missing values. No variable is missing more than a few data points, so I have not dropped any of my variables from the analyses.

Description of the Models

Describe each model fit, why the given model was selected, which hyperparameters to be optimized and how, assumptions of the model, and a high-level (think broad audience) description of what the model is doing and why it is appropriate (even as an initial starting point). Also, discuss how you plan to evaluate model performance.

Model 1: Linear Regression. I chose to start with a linear regression model without penalty because I am trying to predict a continuous variable using predictor variables, and I want to know which of those predictor variables are the most helpful in

predicting transportation scores. This model looks at all the predictor variables I have, and tells us how much of the variance in transportation scores is predicted by the linear relationship between the predictor variables and transportation. I want to start with this model because it is relatively straightforward, but it might have issues with modeling patterns that aren't there because I have more predictor variables than I have samples. There are no hyperparameters to tune for this model.

Model 2: Ridge Regression.   My next model is regression with ridge penalty. I will need to tune the lambda value for this model. It might be a good idea to implement a penalty because I have a lot of predictor variables, some of which I expect to be correlated with each other, and because I do not have as many samples as I have predictor variables. Using ridge regression will reduce sampling variation to help avoid modeling the noise of the data, and hopefully only capture signal. In other words, I have a lot of predictors and not a lot of samples, so there will be a tendency for a linear regression without penalty to make predictions of patterns that aren't there. Doing a ridge regression will put some constraints on the model to cut out meaningless patterns that might show up in the plain linear regression model.

Model 3: Bagged Trees.   Finally, I will model the data using bagged trees. This will provide a non-linear model of the data, in case a linear approach is not the best way to look at the relationships between variables in my data. This model is based on decision trees: in a decision tree analysis, individual continuous variables are split at some optimal point, below which the analysis is conducted differently from the observations above the cutoff. This can be done for any number of variables, any number of times. It is difficult to predict which variables and which cut points will be best (hyperparameters that would need to be tuned, in a decision tree analysis), so it is best to do a lot of decision trees with cuts at different places to determine which is best. I opted to use bagged trees instead of random forests because random forests randomizes rows as well as columns and I have a very limited number of rows. For this analysis, the hyperparameter that will need to be tuned is the

number of trees.

Evaluation of Model Performance. I will compare the root mean squared error
(RMSE), mean absolute error (MAE), and R-square of the test data. I will be looking for
low values for the error terms (RMSE and MAE) and high values of R-square to determine
the best model for my data.

Model Fit

Linear Regression Model.

```
## glmnet
##
## 198 samples
## 772 predictors
##
## Recipe steps: normalize, impute_mean
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 178, 178, 178, 179, 178, 178, ...
## Resampling results:
##
##   RMSE      Rsquared   MAE
##   0.480031  0.6164308  0.3721775
##
## Tuning parameter 'alpha' was held constant at a value of 0
## Tuning
##  parameter 'lambda' was held constant at a value of 0

##   alpha lambda    RMSE    Rsquared       MAE    RMSESD RsquaredSD     MAESD
## 1     0      0 0.480031 0.6164308 0.3721775 0.07266484  0.1184717 0.06785649
```
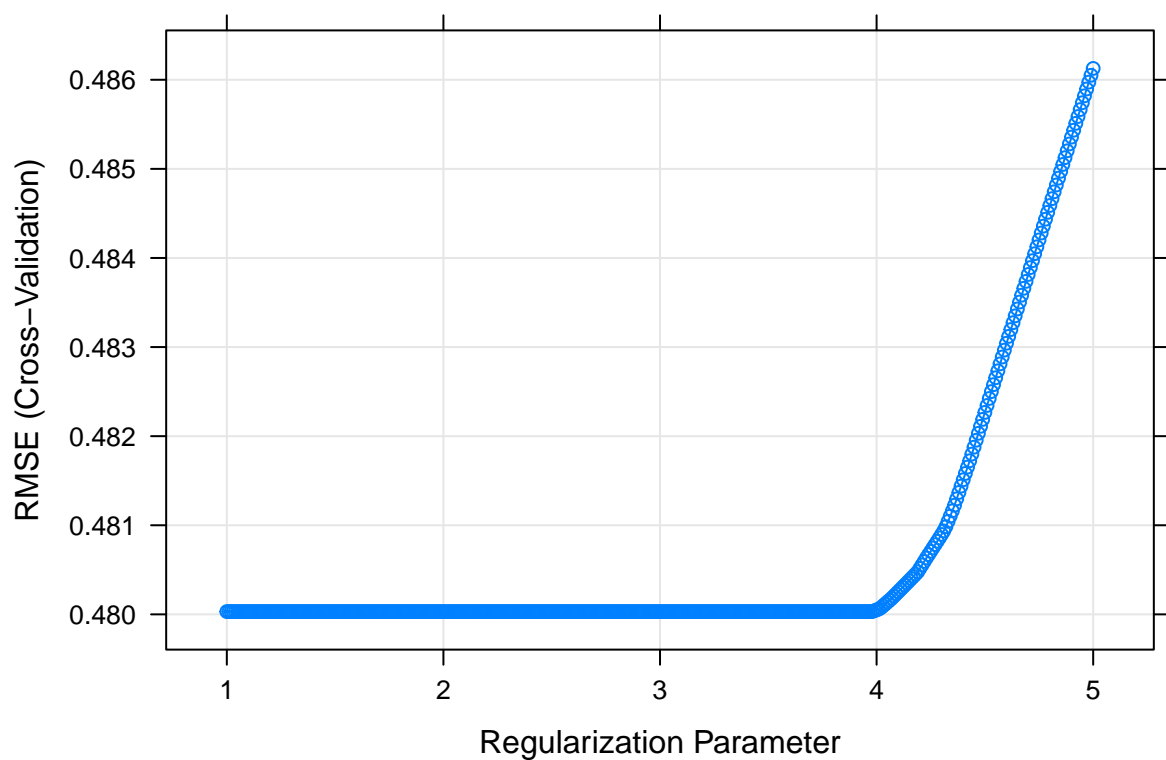
```
## [1] 0.4526108
```

```
## [1] 0.4005285
```

```
## [1] 0.5219195
```

Ridge Regression Model.   Below is a plot of RMSE for different values of lambda. RMSE is flat until lambda is greater than 4, at which point RMSE goes up rapidly.

The best lambda value for this model is 3.97, so I will be proceeding with that. The results for the model using that lambda value are:

```
##     alpha lambda    RMSE  Rsquared       MAE    RMSESD RsquaredSD      MAESD
## 298     0   3.97 0.480031 0.6164308 0.3721775 0.07266484  0.1184717 0.06785649
```

Below are the importance evaluations of each of the predictors in this model:
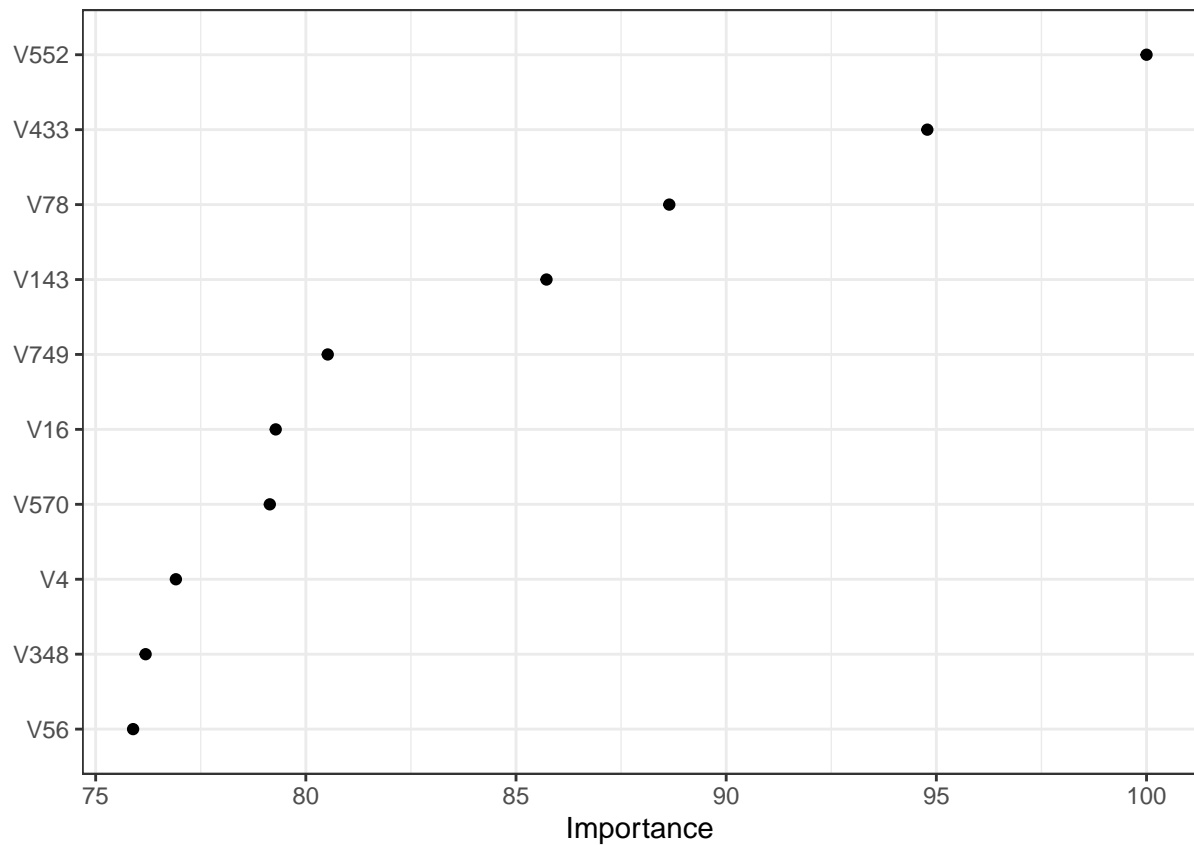
The top ten most important predictors in this model are all encodings of the text data

Bagged Trees.

```
##   mtry splitrule min.node.size      RMSE  Rsquared       MAE    RMSESD
## 1  771  variance             2 0.5815234 0.4051429 0.4583493 0.1088807
##   RsquaredSD      MAESD
## 1  0.1418285 0.08519205
```

Below is a plot of RMSEs for the model using different numbers of trees. The minimum RMSE comes from a model using 160 trees, so I will proceed with that.

Root Mean Square Error (RMSE) by Number of Trees in Bagged Trees Mo
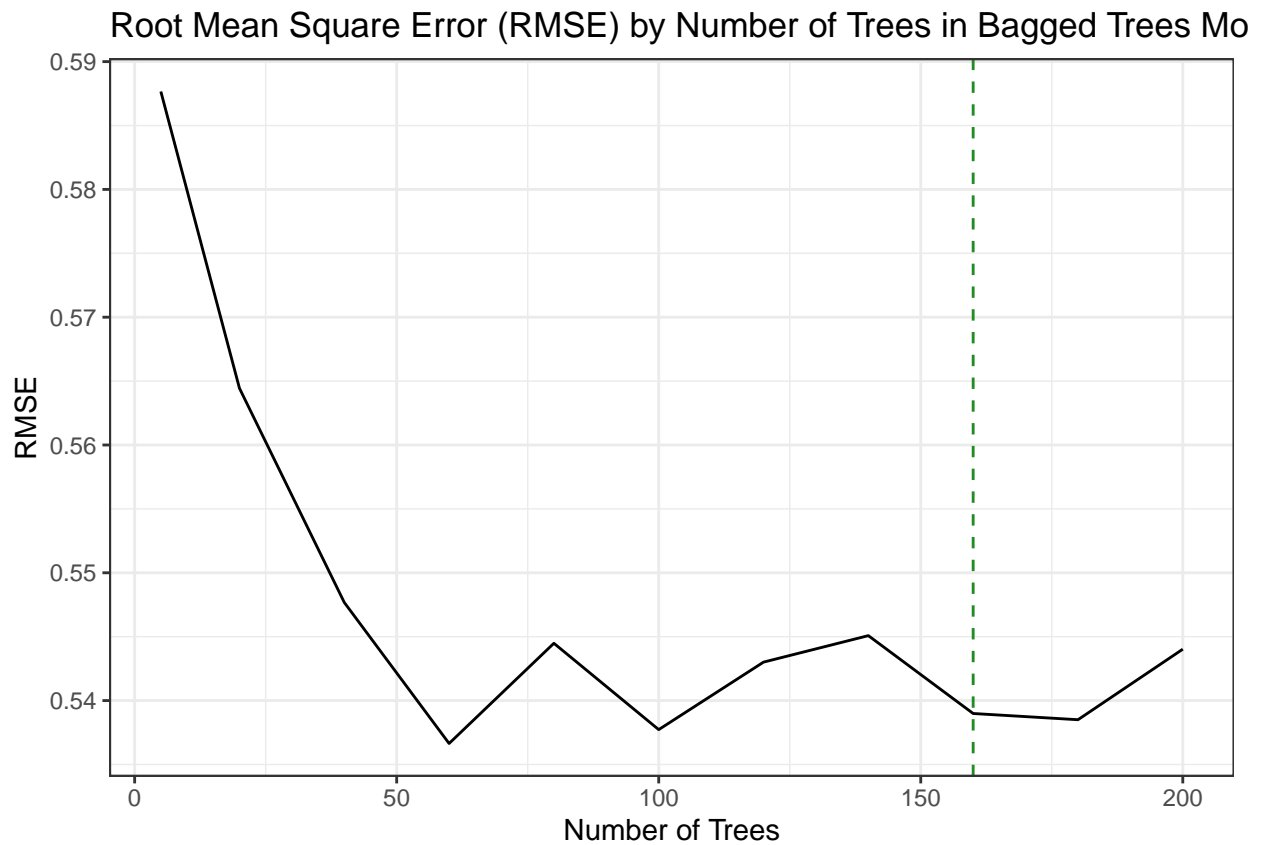

```
nbags[which.min(rmses)]
```

```
## [1] 60
```

Model Comparison

```
##                    R-sq        MAE        RMSE
## Linear        0.4526108 0.4005285 0.5219195
## Ridge         0.4526108 0.4005285 0.5219195
## Bagged Trees  0.1406842 0.5370130 0.6835069
```

The bagged trees model yields a lower R-square and higher MAE and higher RMSE than either regression model. I would therefore choose not to use bagged trees for this analysis.

All terms are identical for linear regression without penalty and for linear regression

with ridge penalty. The RMSE was the same in the ridge regression model for all lambda values up to the optimal cutoff point, and non-penalized linear regression uses a lambda value of 0, which is why these models aren't different. This tells us that overfitting is not an issue for linear regression in these data. I would favor linear regression over ridge regression because it's simpler to run, but functionally it makes no real difference.

Discussion

The regression models were the best for predicting the transportation score of the writing samples provided by the participants in my study. Linear regression without penalty and with ridge penalty were equally predictive; neither was excellent, but both were better at predicting transportation than the bagged trees model.

Some of the text encodings were the most important in predicting transportation score: the text samples themselves seem to do a better job of predicting how transporting readers will find the writing than features of the writer. This is not necessarily unexpected, as transportation scores came from research assistants reading the writing samples (with no knowledge of the participants' backgrounds or personalities). This is interesting because it suggests that people can write stories that are transporting to readers more or less regardless of their personal qualities — at least, in the case of these very short stories. This also indicates that I should look into other features of the writing itself that can be operationalized that underlies transportation scores — in other words, what do the encodings mean, and could we code for/name them so as to develop a better model with or without these encodings? It is somewhat inconvenient not to know what exactly the encodings mean, since it is hard to come to any real conclusions about what features of the text lead to transportation that are psychologically meaningful. It is helpful to have a model that can predict transportation based on text, but it does not mean very much, and might not be very satisfying if I were to write it up in a report for a psychological journal. However, the small sample size makes it so that using this model would not be a good idea.

Overall, the conclusion I get from this is that the participant features included in this model (empathy scale scores, writing experience, and liking of the character they were writing about) probably are not the most important features to look at when determining how transporting their writing will be. In the future, it might be useful to build a model like this that uses text encodings that could predict how transporting text samples are so that human coders would not have to code thousands upon thousands of new text samples, and then to use the predictions from the model instead of transportation scores from human coders as the transportation variable in larger-scale studies — if we were not reliant on human coders to determine transportation because we had a model that could assign transportation scores reasonably well, then we could use the model's predictions as a variable, and then use participant self-report to determine the relationship between empathy score or writing experience or other variables. So, going forward, if I planned to do large-scale studies trying to explain what features of a writer lead them to write transporting stories, it would be helpful to collect a lot of text samples and have human coders rate them on transportation in order to build a model in a pilot study, and then use the model's scores instead of human coders' transportation score in the actual study.

REFERENCES

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a

    multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113.

Green, M. C., Brock, T. C., & Kaufman, G. F. (2004). Understanding media enjoyment:

    The role of transportation into narrative worlds. *Communication Theory, 14*, 311–327.