# Enjoyability, Trends, and Real World Relevance of Detective Fiction

Aleksey Panas, Rohit Shetty, Melinda Chen

December 3, 2021

# Introduction

To explore the fascinating rise of detective stories, we will be working with a dataset of over three hundred short detective stories from the early 1800s to the early 1900s. We will focus on three research questions which might capture insights on these ideas:

- How relevant is detective fiction to the real world crime scenarios?

- What makes an enjoyable detective fiction story?

- Are there any common trends among detective fiction stories?

# Is a more sophisticated story a more enjoyable/satisfying story?

We would like to analyze whether there is a correlation between detective narrative complexity and enjoyability. To do this, we numerically define sophistication and enjoyability. Then we would try to see if a linear relationship can be observed between the two, using which we could make predictions about enjoyability, using sophistication, for detective stories in the 1800s/1900s not in the data.

**Motivation:**

A popular genre in Russian cinema today is criminal detective shows. The most enjoyable aspect of these shows is the constantly evolving narrative, which is usually full of nuances and unexpected plot twists while also containing a significant amount of usable clues allowing the viewer to work towards solving the crime. This motivated the analysis of detective fiction in terms of these factors to see if during the 1800s and 1900s, a similar pattern could be found where enjoyment is positively influenced by sophistication.

# Data Summary

In this analysis. . .

- We define enjoyment score as a numerical value which takes the 0-5 value from the fiction satisfaction variable and adds 2.5 if the annotator would recommend to a friend, and another 2.5 if the annotator was satisfied with the narrative.

- The sophistication score combines 7 variables, such as the presence of usable clues, presence of fabricated evidence, crime types, detective's choice not to alert police, and more, and forms a numerical score ranging roughly from 3 to 14.

The resulting enjoyability and sophistication scores are then used to analyze a correlation between them.

**Figure 1** This table shows the minimum, maximum, mean, and median enjoyment score values for sophistication scores below 5, above 11, and between 5 and 11.

```
## # A tibble: 3 x 5
##   sophistication_score_categories   min   max median  mean
##   <chr>                           <dbl> <dbl>  <dbl> <dbl>
## 1 Above 11                            2    10    5.5  6.42
## 2 Between 5 and 11                    1    10    5.5  5.8
## 3 Less than 5                         1     8    4.5  4.93
```
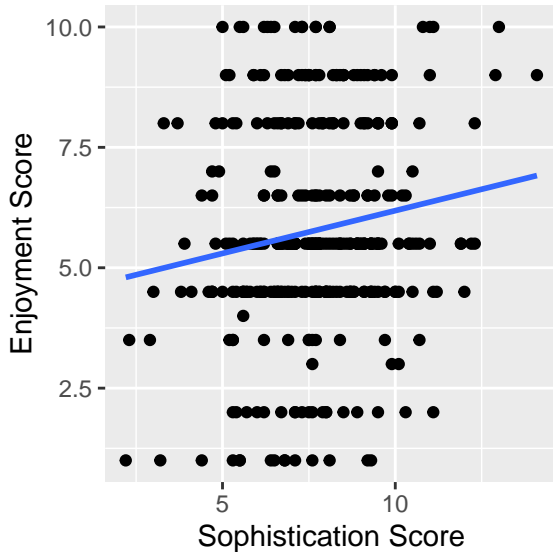
**Figure 2** This scatterplot shows relation points between sophistication and enjoyment for each piece of detective fiction in the data. A fitted linear regression line is overlayed

# Method Results

- The linear regression model, as seen by the line in Figure 2, has an R value of **0.147** which means that the correlation is positive but pretty weak.

- Furthermore, the RMSE has a value of **2.238**, which is the amount by which the actual enjoyment scores vary on average from the enjoyment scores predicted by the linear model.

- An $R^2$ value of **0.0216** tells us that little to none of this variability is explained by the model.

- Figure 1 however, does show that mean enjoyment score is about 1.5 higher for greater sophistication scores than for lower ones.

All of this means that although a small positive relationship can be observed between sophistication and enjoyability, the model is not very viable for prediction on new data.

# Knowing the motive of a crime can we predict the type of crime?

We would like to analyze whether there is a way to predict the type of crime given the motive. To do this, we have produced classification trees corresponding to each crime with the motives as the predictor variables. These tree models will allow us to see if there are any interesting relationships between crimes and motives for detective stories in the 1800s/1900s.

## Motivation:

There are many categories of motives for each of our corresponding types of crimes. We like putting ourselves into the detective roles when reading the story. When the story begins with the investigation, can we also predict the crimes using the motives. This also aids the study of criminology. If police officers can correctly correlate different crimes to their motives. It could help police officers to find the bodies of the dead and return them to their family, or even stop an attempted crime to prevent more injuries and dead.

# Data Summary

In this analysis, we took the *motives* and
*types_of_crimes_or_quasi_crimes_present_in_story* variables and split them into individual variables. The new variables simply produce a binary response indicating if the crime/motive was present for that story or not. The idea behind creating such binary categorical variables was to provide correct data to our classification tree method.

**Figure 3** This is one of the trees obtained it shows the motives behind one of our crimes which is fraud.
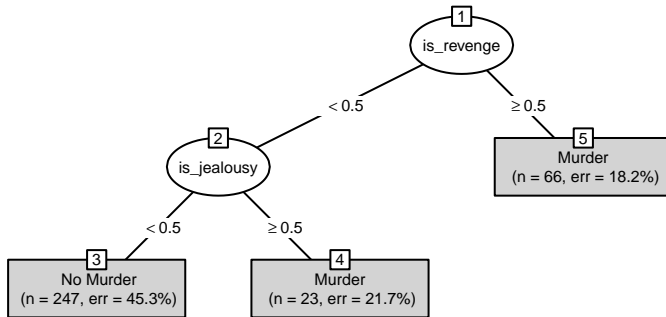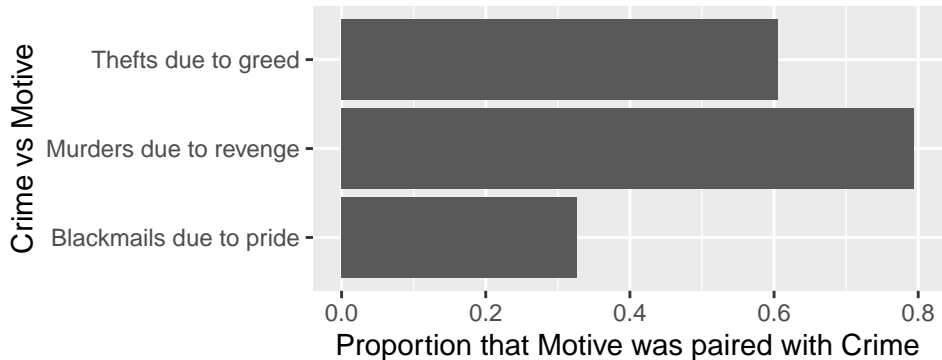
**Figure 4** This plot shows the 3 strongest crime/motive relationships. The bar shows the proportion of how many stories which had a particular motive also had the given crime.

# Method Results

There were many interesting observations we noticed after producing our classification trees for the 5 crimes, here are some:

- We noticed that of all the cases where the motive was revenge, 80% of the crimes committed were murder. Suggesting a strong positive correlation between the two, with a very low rate of error.

- We also notice a weak but interesting relationship between the motive of greed and the crime of theft. With over 60% of crimes motivated by greed being thefts. The error rate for the predictability in our tree for this category is however quite high(47%).

# Is the proportion of female victims in detective stories 80%?

According to police-reported data from the Canadian government, approximately 8 in 10 victims who encountered gender-based violence are female. We will make the hypothesis that in the detective fiction stories in our data, the proportion of female victims is 0.8, reflecting the police-reported data. Then we will run a test using the detective stories data to see if this hypothesis is true or not.

**Motivation**

This question reflects the levels of vulnerability among two gender groups, which is an important aspect of sociology and criminology. The stories reflect the authors' perspective. Thus, it shows us some insights into the gender-related problems between the 1800s and 1900s. By revealing the exact figure of the proportion, we can find the similarities of different culprits' mindsets. This research is also meaningful to current-day society. It alerts the female group and fosters their awareness of self-protection. Sometimes it aids officers to prevent crimes caused by copycats who are intoxicated by the murderers from the detective stories.

# Data Summary

For this question, we will be using the victim variables from the dataset,

- We will use *number_of_victims_of_gender_male*

- We will use *number_of_victims_of_gender_female*

- We will calculate the proportion of female victims to total number of victims for each detective fiction story and use that as a new variable
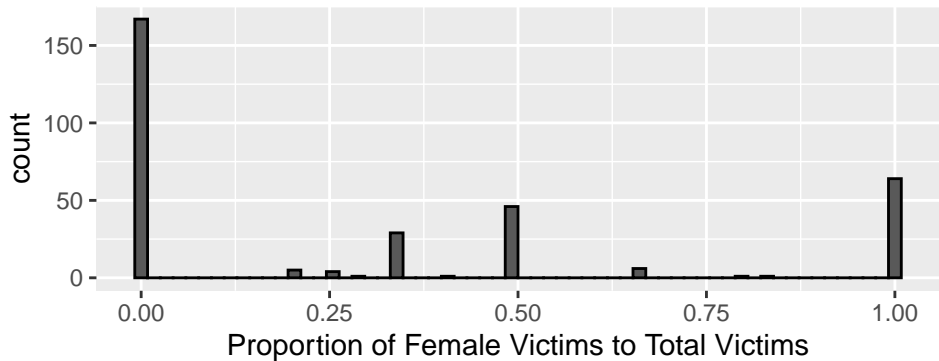
**Figure 5** This histogram shows the proportion of female victims to total victims for each story in the data.

# Method Results

- From the detective stories data, the average proportion of female victims is found to be **0.323033**

- This value is significantly different from our hypothesized proportion of **0.8**.

- From the hypothesis test that was run, a p-value of **0** was calculated. This value of 0 indicates that there is no evidence whatsoever supporting that across all detective stories in the 1800/1900s, a proportion of 0.8 for female victims is observed.

Together, this means that there is no association between the number of female victims in detective stories written in 1800-1900 and the real world in 2021.

# Limitations

- Having more stories in the data might have yielded a stronger correlation between sophistication and enjoyability, and a lower error rate in prediction of crime from motive.

- Each story in the data was rated on enjoyability by a single reader. This creates more potential for outliers and bias. If the satisfaction variables for each story in the data were instead an average rating from many readers, this could have yielded more accurate results.

- Lastly, extrapolation is a present issue within this analysis, that is, this prediction model in only valid for detective fiction in the 1800s and 1900s. The data would have to include stories from later times for our models to make accurate predictions for the present time period.

# Conclusion

In conclusion, our three research questions have provided us with some insights:

- Our research shows that there might not be a relationship between today's real world crime statistics and detective stories, as shown by the difference in female victim proportions between the two. It makes sense that fiction wouldn't follow patterns of real world current-day society, especially writing from a much earlier time period.

- When searching for what makes an enjoyable story, we found a weak but positive correlation between enjoyment and sophistication. Although a more sophisticated story is more likely to be enjoyed, this relationship is not strong enough to provide accurate predictions.

- There are trends between crimes and motives within detective stories. In particular, it seems that murders tend to be motivated by revenge most often, and thefts are commonly motivated by greed. This reflects human nature where greed tends to provoke theft while serious desire for revenge provokes ideas of murder.

# References and Acknowledgements

According to police-reported data from the Canadian government, approximately 8 in 10 victims who encountered gender-based violence are female. Therefore, the proportion number is set as 0.8 based on a reliable source.

Canada, W. and G. E. (2021, May 10). Government of Canada - About gender-based violence. Women and Gender Equality Canada. Retrieved October 20, 2021, from https://women-gender-equality.canada.ca/en/gender-based-violence-knowledge-centre/about-gender-based-violence.html