

# Magic Vs Evolution\*

Or more specifically language in Harry Potter vs Darwin's Natural Selection

Shreya Sakura Noskor

April 2, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Source . . . . .	2
2.2	Variables . . . . .	3
2.2.1	Distribution of each word with each text . . . . .	3
<b>3</b>	<b>Model</b>	<b>3</b>
3.1	Natural and the 2 texts . . . . .	3
3.1.1	Model justification . . . . .	9
3.2	“Magic” and the 2 texts . . . . .	9
3.2.1	Model justification . . . . .	16
<b>4</b>	<b>Results</b>	<b>16</b>
<b>5</b>	<b>Discussion</b>	<b>18</b>
5.1	Why is the result the way it is . . . . .	18
5.2	Importance . . . . .	18
5.3	Third discussion point . . . . .	18
5.4	Weaknesses and next steps . . . . .	18
	<b>Appendix</b>	<b>19</b>
	<b>References</b>	<b>20</b>

---

\*Code and data are available at: [HarryDarwin](#)

```
# A tibble: 21,566 x 2
  gutenber_id text
    <int> <chr>
1      2009 "1228 1859, First Edition"
2      2009 "22764 1860, Second Edition"
3      2009 "2009 1872, Sixth Edition, considered the definitive edition~
4      2009 ""
5      2009 ""
6      2009 ""
7      2009 ""
8      2009 "On the Origin of Species"
9      2009 ""
10     2009 "BY MEANS OF NATURAL SELECTION,"
# i 21,556 more rows

[1] "character"
```

## 1 Introduction

## 2 Data

### 2.1 Source

The data utilized was from Project Gutenberg and PDF Drives and with the help of R (R Core Team 2023) we were able to create this paper. Also code for making the models were made referencing Telling Stories by Rohan Alexander(Wickham et al. 2019a). Other R packages were used to clean, process and model the data such as, Wickham et al. (2019b), Johnston and Robinson (2023), Goodrich et al. (2022), Ooms (2023), Wickham et al. (2023), Richardson et al. (2024), Wickham (2023), Arel-Bundock (2024).

## 2.2 Variables

### 2.2.1 Distribution of each word with each text

## 3 Model

### 3.1 Natural and the 2 texts

Define  $y_i$  is the number of times “natural” appeared in each text and the explanatory variable is the number of words in the line. We predict to see that there is a positive correlation in Darwin’s text but not in Harry Potter due to the difference in topics.

$$y_i | \lambda_i \sim \text{Poisson}(\lambda_i) \quad (1)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \times \text{Number of Words}_i \quad (2)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (3)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (4)$$

$$(5)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

```
SAMPLING FOR MODEL 'count' NOW (CHAIN 1).
```

```
Chain 1:
```

```
Chain 1: Gradient evaluation took 0.004216 seconds
```

```
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 42.16 seconds.
```

```
Chain 1: Adjust your expectations accordingly!
```

```
Chain 1:
```

```
Chain 1:
```

```
Chain 1: Iteration:    1 / 2000 [  0%] (Warmup)
```

```
Chain 1: Iteration:  200 / 2000 [ 10%] (Warmup)
```

```
Chain 1: Iteration:  400 / 2000 [ 20%] (Warmup)
```

```
Chain 1: Iteration:  600 / 2000 [ 30%] (Warmup)
```

```
Chain 1: Iteration:  800 / 2000 [ 40%] (Warmup)
```

```
Chain 1: Iteration: 1000 / 2000 [ 50%] (Warmup)
```

```
Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)
```

```
Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)
```

```
Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)
```

```
Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)
```

```
Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)
```

Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 1:  
Chain 1: Elapsed Time: 23.97 seconds (Warm-up)  
Chain 1: 25.809 seconds (Sampling)  
Chain 1: 49.779 seconds (Total)  
Chain 1:

SAMPLING FOR MODEL 'count' NOW (CHAIN 2).

Chain 2:  
Chain 2: Gradient evaluation took 0.005256 seconds  
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 52.56 seconds.  
Chain 2: Adjust your expectations accordingly!  
Chain 2:  
Chain 2:  
Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 2: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 2: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 2: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 2: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 2:  
Chain 2: Elapsed Time: 25.551 seconds (Warm-up)  
Chain 2: 27.559 seconds (Sampling)  
Chain 2: 53.11 seconds (Total)  
Chain 2:

SAMPLING FOR MODEL 'count' NOW (CHAIN 3).

Chain 3:  
Chain 3: Gradient evaluation took 0.004519 seconds  
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 45.19 seconds.  
Chain 3: Adjust your expectations accordingly!  
Chain 3:  
Chain 3:  
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 3: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 3: Iteration: 600 / 2000 [ 30%] (Warmup)

```

Chain 3: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 3: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 3:
Chain 3: Elapsed Time: 20.49 seconds (Warm-up)
Chain 3:                19.635 seconds (Sampling)
Chain 3:                40.125 seconds (Total)
Chain 3:

```

SAMPLING FOR MODEL 'count' NOW (CHAIN 4).

```

Chain 4:
Chain 4: Gradient evaluation took 0.002219 seconds
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 22.19 seconds.
Chain 4: Adjust your expectations accordingly!
Chain 4:
Chain 4:
Chain 4: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 4: Iteration: 400 / 2000 [ 20%] (Warmup)
Chain 4: Iteration: 600 / 2000 [ 30%] (Warmup)
Chain 4: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 4:
Chain 4: Elapsed Time: 17.211 seconds (Warm-up)
Chain 4:                13.55 seconds (Sampling)
Chain 4:                30.761 seconds (Total)
Chain 4:

```

SAMPLING FOR MODEL 'count' NOW (CHAIN 1).

```

Chain 1:
Chain 1: Gradient evaluation took 0.001825 seconds

```

Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 18.25 seconds.  
Chain 1: Adjust your expectations accordingly!  
Chain 1:  
Chain 1:  
Chain 1: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 1: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 1: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 1: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 1: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 1: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 1:  
Chain 1: Elapsed Time: 11.344 seconds (Warm-up)  
Chain 1: 10.096 seconds (Sampling)  
Chain 1: 21.44 seconds (Total)  
Chain 1:

SAMPLING FOR MODEL 'count' NOW (CHAIN 2).

Chain 2:  
Chain 2: Gradient evaluation took 0.001159 seconds  
Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 11.59 seconds.  
Chain 2: Adjust your expectations accordingly!  
Chain 2:  
Chain 2:  
Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 2: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 2: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 2: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 2: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 2:  
Chain 2: Elapsed Time: 18.018 seconds (Warm-up)

Chain 2: 14.392 seconds (Sampling)  
Chain 2: 32.41 seconds (Total)  
Chain 2:

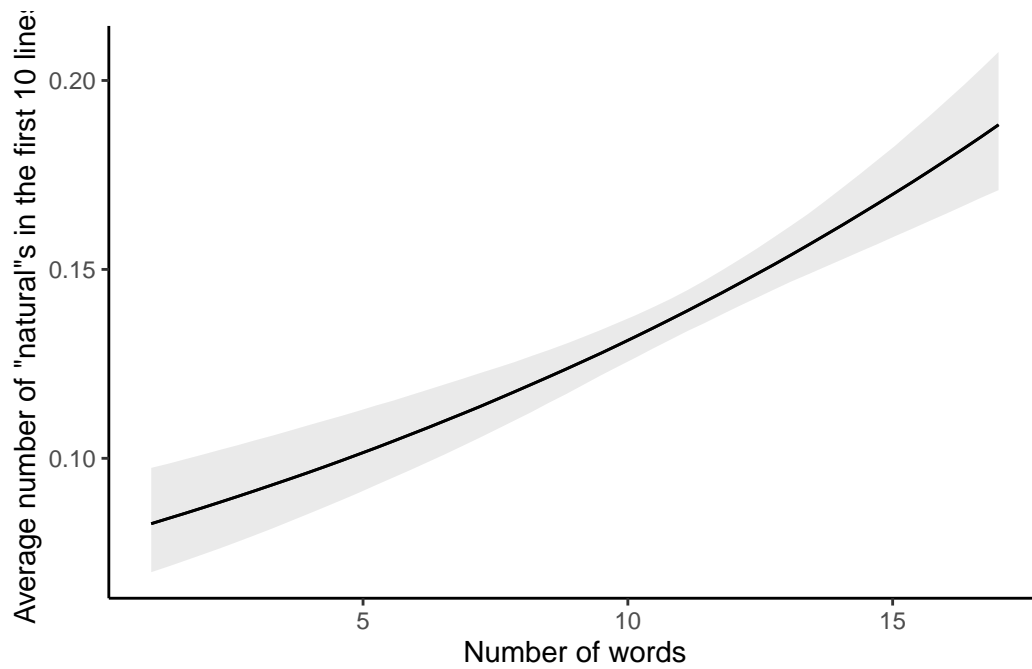
SAMPLING FOR MODEL 'count' NOW (CHAIN 3).

Chain 3:  
Chain 3: Gradient evaluation took 0.001802 seconds  
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 18.02 seconds.  
Chain 3: Adjust your expectations accordingly!  
Chain 3:  
Chain 3:  
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 3: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 3: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 3: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 3: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 3:  
Chain 3: Elapsed Time: 15.491 seconds (Warm-up)  
Chain 3: 9.905 seconds (Sampling)  
Chain 3: 25.396 seconds (Total)  
Chain 3:

SAMPLING FOR MODEL 'count' NOW (CHAIN 4).

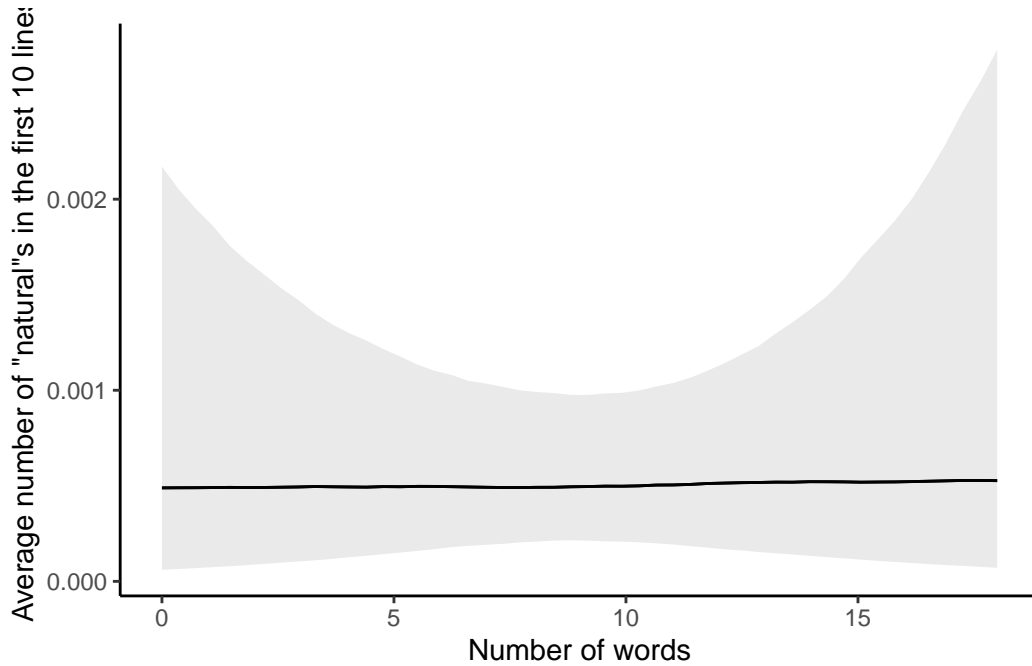
Chain 4:  
Chain 4: Gradient evaluation took 0.001235 seconds  
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 12.35 seconds.  
Chain 4: Adjust your expectations accordingly!  
Chain 4:  
Chain 4:  
Chain 4: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 4: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 4: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 4: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 4: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)

Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 4:  
Chain 4: Elapsed Time: 12.137 seconds (Warm-up)  
Chain 4: 8.989 seconds (Sampling)  
Chain 4: 21.126 seconds (Total)  
Chain 4:



positive correlation it seems





No correlation at all

### 3.1.1 Model justification

We predict to see that there is a positive correlation in Darwin's text but not in Harry Potter due to the difference in topics.

## 3.2 “Magic” and the 2 texts

Define  $y_i$  is the number of times “magic” or “miracle” appeared in the text and the explanatory variable is the number of words in the line.

$$y_i | \lambda_i \sim \text{Poisson}(\lambda_i) \quad (6)$$

$$\log(\lambda_i) = \beta_0 + \beta_1 \times \text{Number of Words}_i \quad (7)$$

$$\beta_0 \sim \text{Normal}(0, 2.5) \quad (8)$$

$$\beta_1 \sim \text{Normal}(0, 2.5) \quad (9)$$

$$(10)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

SAMPLING FOR MODEL 'count' NOW (CHAIN 1).

Chain 1:

Chain 1: Gradient evaluation took 0.002579 seconds

Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 25.79 seconds.

Chain 1: Adjust your expectations accordingly!

Chain 1:

Chain 1:

Chain 1: Iteration: 1 / 2000 [ 0%] (Warmup)

Chain 1: Iteration: 200 / 2000 [ 10%] (Warmup)

Chain 1: Iteration: 400 / 2000 [ 20%] (Warmup)

Chain 1: Iteration: 600 / 2000 [ 30%] (Warmup)

Chain 1: Iteration: 800 / 2000 [ 40%] (Warmup)

Chain 1: Iteration: 1000 / 2000 [ 50%] (Warmup)

Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)

Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)

Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)

Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)

Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)

Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)

Chain 1:

Chain 1: Elapsed Time: 22.826 seconds (Warm-up)

Chain 1: 15.473 seconds (Sampling)

Chain 1: 38.299 seconds (Total)

Chain 1:

SAMPLING FOR MODEL 'count' NOW (CHAIN 2).

Chain 2:

Chain 2: Gradient evaluation took 0.002317 seconds

Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 23.17 seconds.

Chain 2: Adjust your expectations accordingly!

Chain 2:

Chain 2:

Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)

Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)

Chain 2: Iteration: 400 / 2000 [ 20%] (Warmup)

Chain 2: Iteration: 600 / 2000 [ 30%] (Warmup)

Chain 2: Iteration: 800 / 2000 [ 40%] (Warmup)

Chain 2: Iteration: 1000 / 2000 [ 50%] (Warmup)

Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)

Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)

Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)

Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)

Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)

Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 2:  
Chain 2: Elapsed Time: 16.989 seconds (Warm-up)  
Chain 2: 21.725 seconds (Sampling)  
Chain 2: 38.714 seconds (Total)  
Chain 2:

SAMPLING FOR MODEL 'count' NOW (CHAIN 3).

Chain 3:  
Chain 3: Gradient evaluation took 0.002862 seconds  
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 28.62 seconds.  
Chain 3: Adjust your expectations accordingly!  
Chain 3:  
Chain 3:  
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 3: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 3: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 3: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 3: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 3:  
Chain 3: Elapsed Time: 15.949 seconds (Warm-up)  
Chain 3: 16.314 seconds (Sampling)  
Chain 3: 32.263 seconds (Total)  
Chain 3:

SAMPLING FOR MODEL 'count' NOW (CHAIN 4).

Chain 4:  
Chain 4: Gradient evaluation took 0.002022 seconds  
Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 20.22 seconds.  
Chain 4: Adjust your expectations accordingly!  
Chain 4:  
Chain 4:  
Chain 4: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 4: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 4: Iteration: 600 / 2000 [ 30%] (Warmup)

```

Chain 4: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 4: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 4:
Chain 4: Elapsed Time: 15.158 seconds (Warm-up)
Chain 4:                22.223 seconds (Sampling)
Chain 4:                37.381 seconds (Total)
Chain 4:

```

SAMPLING FOR MODEL 'count' NOW (CHAIN 1).

```

Chain 1:
Chain 1: Gradient evaluation took 0.001012 seconds
Chain 1: 1000 transitions using 10 leapfrog steps per transition would take 10.12 seconds.
Chain 1: Adjust your expectations accordingly!
Chain 1:
Chain 1:
Chain 1: Iteration: 1 / 2000 [ 0%] (Warmup)
Chain 1: Iteration: 200 / 2000 [ 10%] (Warmup)
Chain 1: Iteration: 400 / 2000 [ 20%] (Warmup)
Chain 1: Iteration: 600 / 2000 [ 30%] (Warmup)
Chain 1: Iteration: 800 / 2000 [ 40%] (Warmup)
Chain 1: Iteration: 1000 / 2000 [ 50%] (Warmup)
Chain 1: Iteration: 1001 / 2000 [ 50%] (Sampling)
Chain 1: Iteration: 1200 / 2000 [ 60%] (Sampling)
Chain 1: Iteration: 1400 / 2000 [ 70%] (Sampling)
Chain 1: Iteration: 1600 / 2000 [ 80%] (Sampling)
Chain 1: Iteration: 1800 / 2000 [ 90%] (Sampling)
Chain 1: Iteration: 2000 / 2000 [100%] (Sampling)
Chain 1:
Chain 1: Elapsed Time: 9.197 seconds (Warm-up)
Chain 1:                9.112 seconds (Sampling)
Chain 1:                18.309 seconds (Total)
Chain 1:

```

SAMPLING FOR MODEL 'count' NOW (CHAIN 2).

```

Chain 2:
Chain 2: Gradient evaluation took 0.002846 seconds

```

Chain 2: 1000 transitions using 10 leapfrog steps per transition would take 28.46 seconds.  
Chain 2: Adjust your expectations accordingly!  
Chain 2:  
Chain 2:  
Chain 2: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 2: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 2: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 2: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 2: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 2: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 2: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 2: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 2: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 2: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 2: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 2: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 2:  
Chain 2: Elapsed Time: 9.124 seconds (Warm-up)  
Chain 2: 10.786 seconds (Sampling)  
Chain 2: 19.91 seconds (Total)  
Chain 2:

SAMPLING FOR MODEL 'count' NOW (CHAIN 3).

Chain 3:  
Chain 3: Gradient evaluation took 0.001248 seconds  
Chain 3: 1000 transitions using 10 leapfrog steps per transition would take 12.48 seconds.  
Chain 3: Adjust your expectations accordingly!  
Chain 3:  
Chain 3:  
Chain 3: Iteration: 1 / 2000 [ 0%] (Warmup)  
Chain 3: Iteration: 200 / 2000 [ 10%] (Warmup)  
Chain 3: Iteration: 400 / 2000 [ 20%] (Warmup)  
Chain 3: Iteration: 600 / 2000 [ 30%] (Warmup)  
Chain 3: Iteration: 800 / 2000 [ 40%] (Warmup)  
Chain 3: Iteration: 1000 / 2000 [ 50%] (Warmup)  
Chain 3: Iteration: 1001 / 2000 [ 50%] (Sampling)  
Chain 3: Iteration: 1200 / 2000 [ 60%] (Sampling)  
Chain 3: Iteration: 1400 / 2000 [ 70%] (Sampling)  
Chain 3: Iteration: 1600 / 2000 [ 80%] (Sampling)  
Chain 3: Iteration: 1800 / 2000 [ 90%] (Sampling)  
Chain 3: Iteration: 2000 / 2000 [100%] (Sampling)  
Chain 3:  
Chain 3: Elapsed Time: 8.134 seconds (Warm-up)

Chain 3: 9.502 seconds (Sampling)  
Chain 3: 17.636 seconds (Total)  
Chain 3:

SAMPLING FOR MODEL 'count' NOW (CHAIN 4).

Chain 4:

Chain 4: Gradient evaluation took 0.001369 seconds

Chain 4: 1000 transitions using 10 leapfrog steps per transition would take 13.69 seconds.

Chain 4: Adjust your expectations accordingly!

Chain 4:

Chain 4:

Chain 4: Iteration: 1 / 2000 [ 0%] (Warmup)

Chain 4: Iteration: 200 / 2000 [ 10%] (Warmup)

Chain 4: Iteration: 400 / 2000 [ 20%] (Warmup)

Chain 4: Iteration: 600 / 2000 [ 30%] (Warmup)

Chain 4: Iteration: 800 / 2000 [ 40%] (Warmup)

Chain 4: Iteration: 1000 / 2000 [ 50%] (Warmup)

Chain 4: Iteration: 1001 / 2000 [ 50%] (Sampling)

Chain 4: Iteration: 1200 / 2000 [ 60%] (Sampling)

Chain 4: Iteration: 1400 / 2000 [ 70%] (Sampling)

Chain 4: Iteration: 1600 / 2000 [ 80%] (Sampling)

Chain 4: Iteration: 1800 / 2000 [ 90%] (Sampling)

Chain 4: Iteration: 2000 / 2000 [100%] (Sampling)

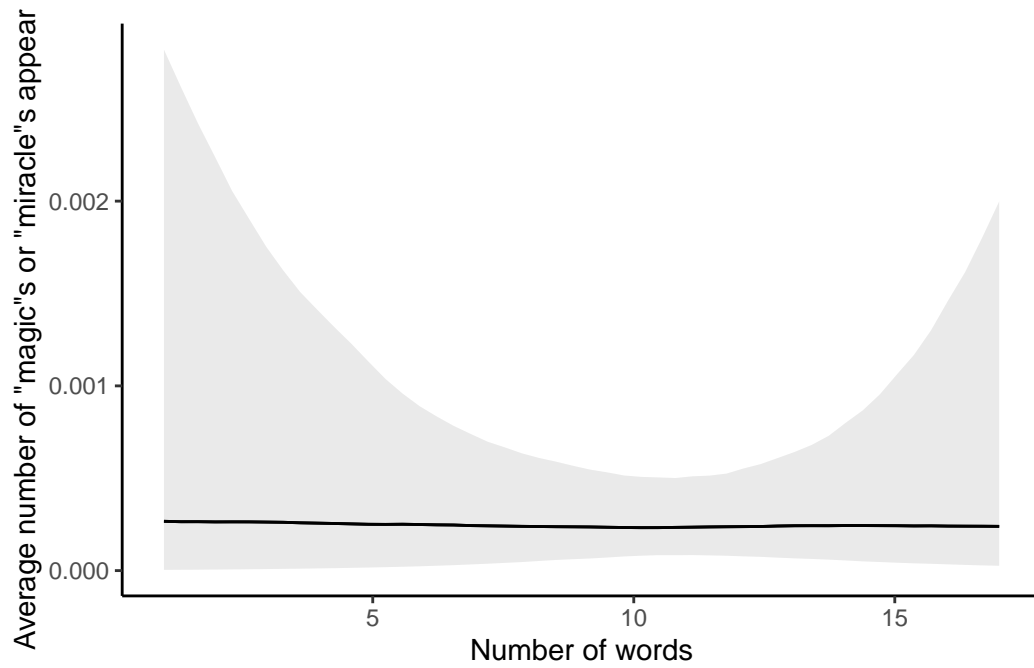
Chain 4:

Chain 4: Elapsed Time: 7.494 seconds (Warm-up)

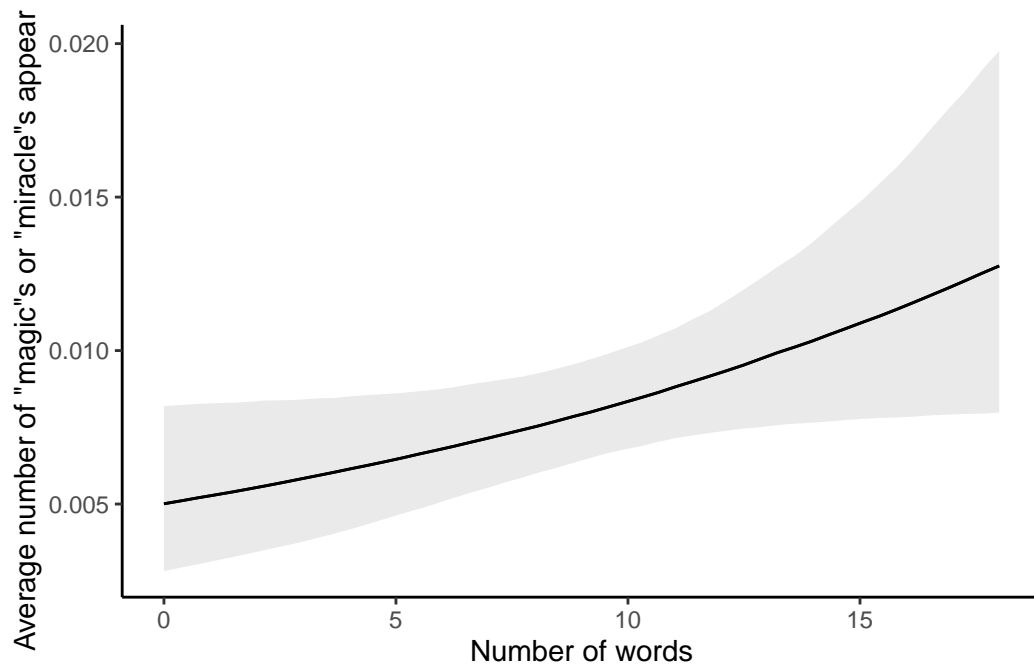
Chain 4: 9.655 seconds (Sampling)

Chain 4: 17.149 seconds (Total)

Chain 4:



As we can see here there is almost no correlation at all.



### 3.2.1 Model justification

We predict to see that there is a positive correlation in Harry Potter but not in Darwin’s text due to the difference in topics.

## 4 Results

Our results are summarized in `?@tbl-modelresults`.

Table 1: Model Summary showcasing the correlation coefficient for Darwin’s evolution book and the word nature

darwin and nature	
(Intercept)	−2.54 (0.09)
word_count	0.05 (0.01)
Num.Obs.	19 413
Log.Lik.	−8095.481
ELPD	−8097.2
ELPD s.e.	106.3
LOOIC	16 194.4
LOOIC s.e.	212.7
WAIC	16 194.4
RMSE	0.37



Table 2: Model Summary showcasing the correlation coefficient for Darwin’s evolution book and the word magic

darwin and magic	
(Intercept)	−8.22 (1.74)
word_count	−0.01 (0.16)
Num.Obs.	19 413
Log.Lik.	−37.980
ELPD	−39.8
ELPD s.e.	17.3
LOOIC	79.6
LOOIC s.e.	34.5
WAIC	79.5
RMSE	0.01

Table 3: Model Summary showcasing the correlation coefficient for Harry Potter and The Prisoner of Azkaban and the word nature

harry and nature	
(Intercept)	−7.62 (0.93)
word_count	0.00 (0.10)
Num.Obs.	12 767
Log.Lik.	−52.022
ELPD	−53.6
ELPD s.e.	19.0
LOOIC	107.2
LOOIC s.e.	38.0
WAIC	107.2
RMSE	0.02

Table 4: Model Summary showcasing the correlation coefficient for Harry Potter and The Prisoner of Azkaban and the word magic

	harry and magic
(Intercept)	−5.30 (0.28)
word_count	0.05 (0.03)
Num.Obs.	12 767
Log.Lik.	−597.437
ELPD	−599.1
ELPD s.e.	48.6
LOOIC	1198.3
LOOIC s.e.	97.2
WAIC	1198.3
RMSE	0.09

## 5 Discussion

### 5.1 Why is the result the way it is

### 5.2 Importance

### 5.3 Third discussion point

### 5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

## Appendix

## References

- Arel-Bundock, Vincent. 2024. *MarginalEffects: Predictions, Comparisons, Slopes, Marginal Means, and Hypothesis Tests*. <https://CRAN.R-project.org/package=marginalEffects>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- Johnston, Myfanwy, and David Robinson. 2023. *Gutenbergr: Download and Process Public Domain Works from Project Gutenberg*. <https://CRAN.R-project.org/package=gutenbergr>.
- Ooms, Jeroen. 2023. *Pdftools: Text Extraction, Rendering and Converting of PDF Documents*. <https://CRAN.R-project.org/package=pdfutils>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoş Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. <https://CRAN.R-project.org/package=arrow>.
- Wickham, Hadley. 2023. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019b. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- , et al. 2019a. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.