Reading Speed Simulator Analysis

Jonny Giordano

January 24, 2019

```
load("reading_speed_environment.RData")
```

Reading Speed Simulator

A Reading Speed Simulator was built in R Version 3.2.3. The simulation was to model the effects of eye dominancy and native language, respective to the text, on reading speed. The simulation was based on a partially completed reading speed experiment. The intention of this analysis is to show that the parameter settings in the simulation can be retrieved, and to create a linear mixed model on the data.

Effects on reading speeds were modeled with normal distributions in all cases except one, the effect caused by the prosthetic eye replacement. However, this case is not covered in this analysis.

Native speakers were modeled as having an average reading speed of 250 words per minute (WPM), with a standard deviation (STD) of 12. Foreign speakers were modeled as having an average reading speed of 230 WPM, with a STD of 16.

Using only the dominant eye had an effect of reducing reading speed by an average of 3 WPM, with a STD of 0.75. Using only the weak eye reduced the reading speed by an average of 5 WPM, with a STD of 1. Using both eyes had no effect on reading speed. The prosthetic "enhancement" was modeled with an even distribution having effects on WPM from -7 to +5.

Based on the experiment leading to the simulation, texts were modeled on an even distribution containing 45 to 90 words. A text difficulty effect was added, being a normal distribution having no average effect on texts, but having a STD of 5 WPM.

The simulation was carried out with 30 native and non-native subjects (60 total), and 15 texts.

The output for consideration in the experiment is the time in minutes used to read each text.

Here is presentation of a portion of the final output of the simulation. The only data present are those that would be available to a researcher. The exception is that each subject was able to read each text in all conditions. In the real world having a subject reread text would most likely affect reading speed.

```
head(data_figs)
```

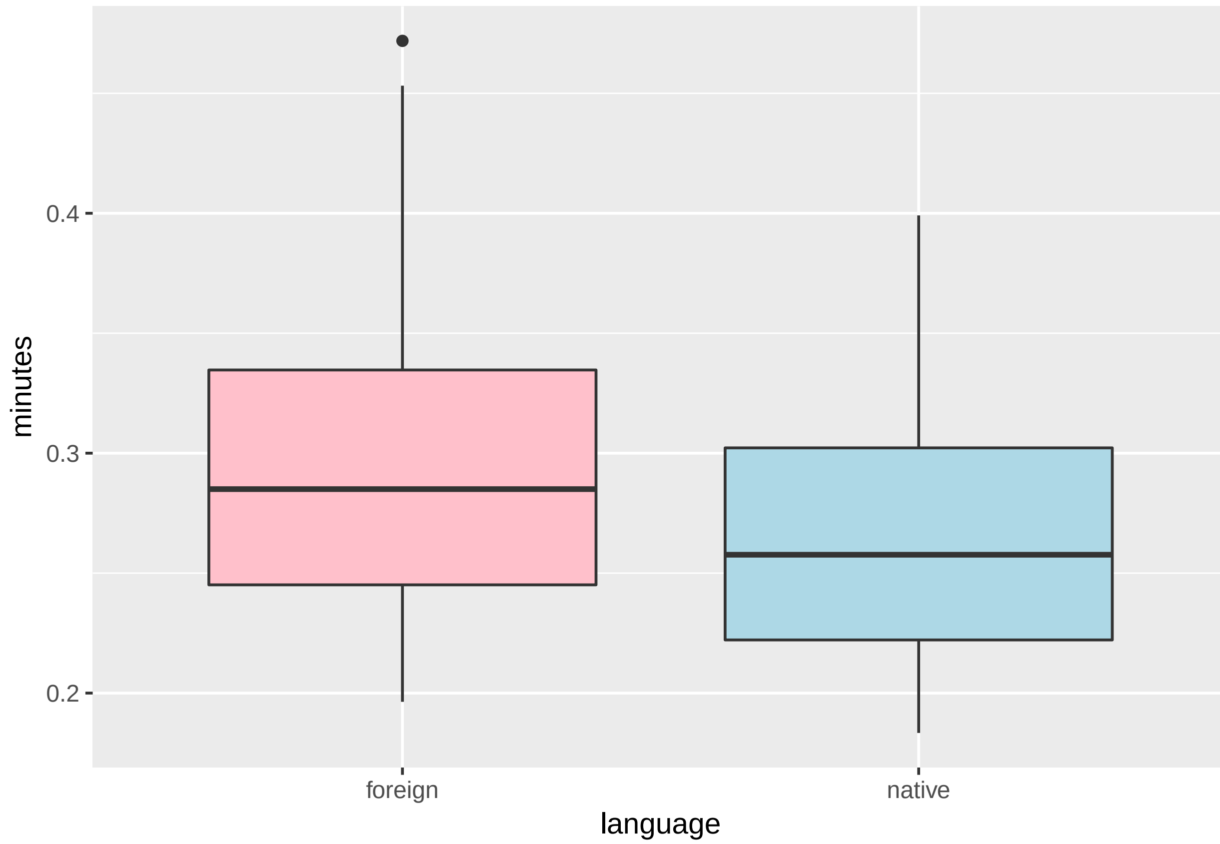
```
##   text_id subject_id text_length language       eye   minutes
## 1       1          1          79   native both_eyes 0.3032771
## 2       2          1          51   native both_eyes 0.1986656
## 3       3          1          76   native both_eyes 0.3015110
## 4       4          1          66   native both_eyes 0.2606506
## 5       5          1          58   native both_eyes 0.2199015
## 6       6          1          54   native both_eyes 0.2121878
```

Figures

Now the simulated data will be presented.

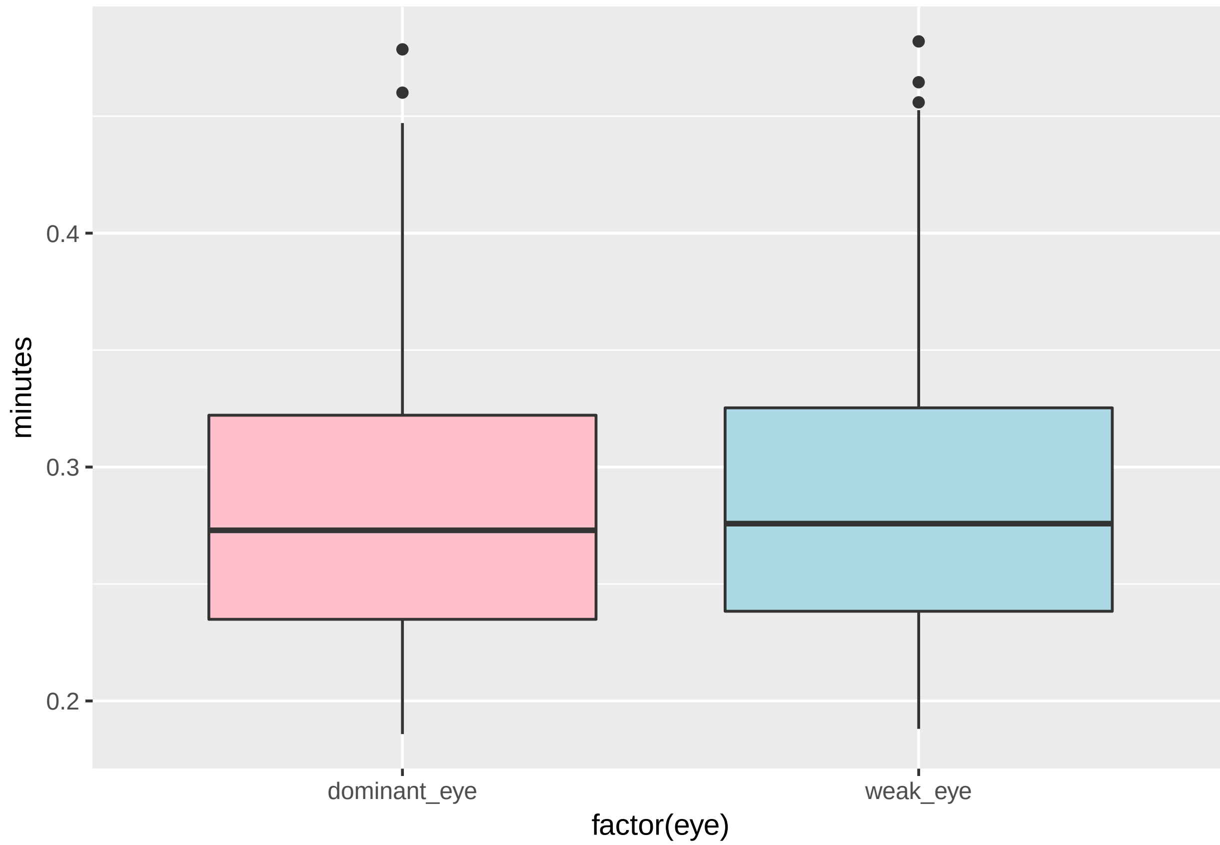
Here is a boxplot demonstrating the average reading time between native and non-native readers while reading with both eyes. The reading condition for both eyes had no effect on the subjects reading speed. We observe that natives have a lower reading time, meaning they tend to read faster.

```
language_box.plot
```



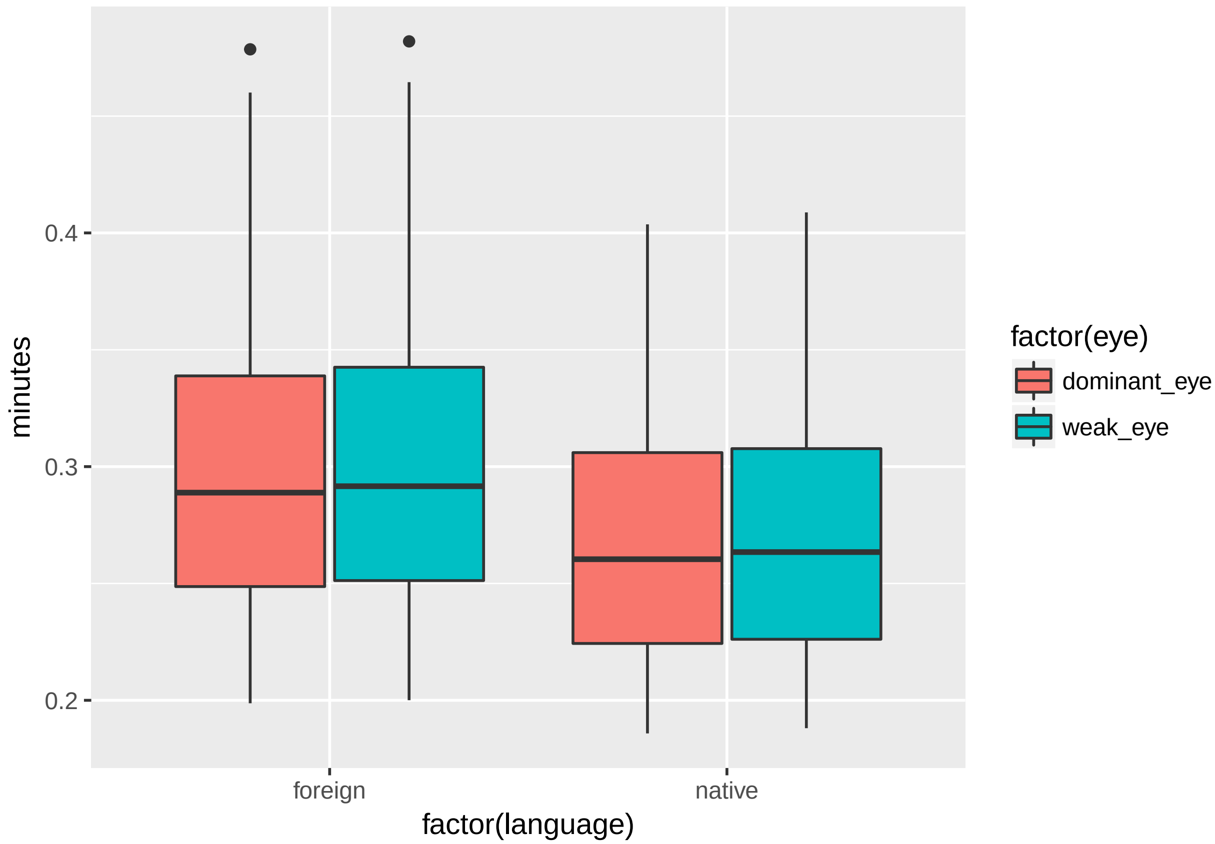
He we show the reading times based on the conditions of reading with the dominant or weak eye.

```
eyes_box.plot
```



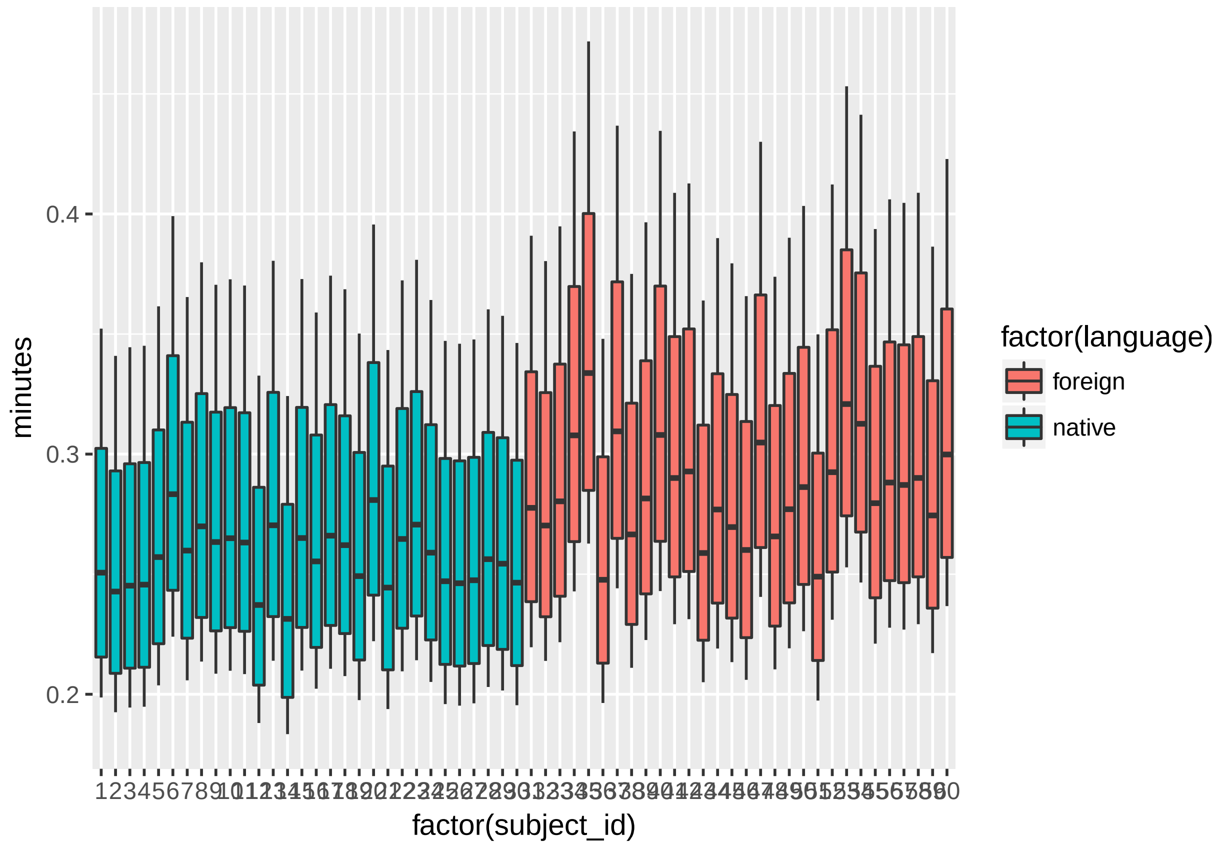
Next we show the reading times of foreign and native speakers, considering the factor of dominant and weak eyes. We see across the foreign vs native case that non-natives speakers have longer reading times. The dominant eye has a slightly lower reading time for both types of subjects.

```
eye_native_box.plot
```



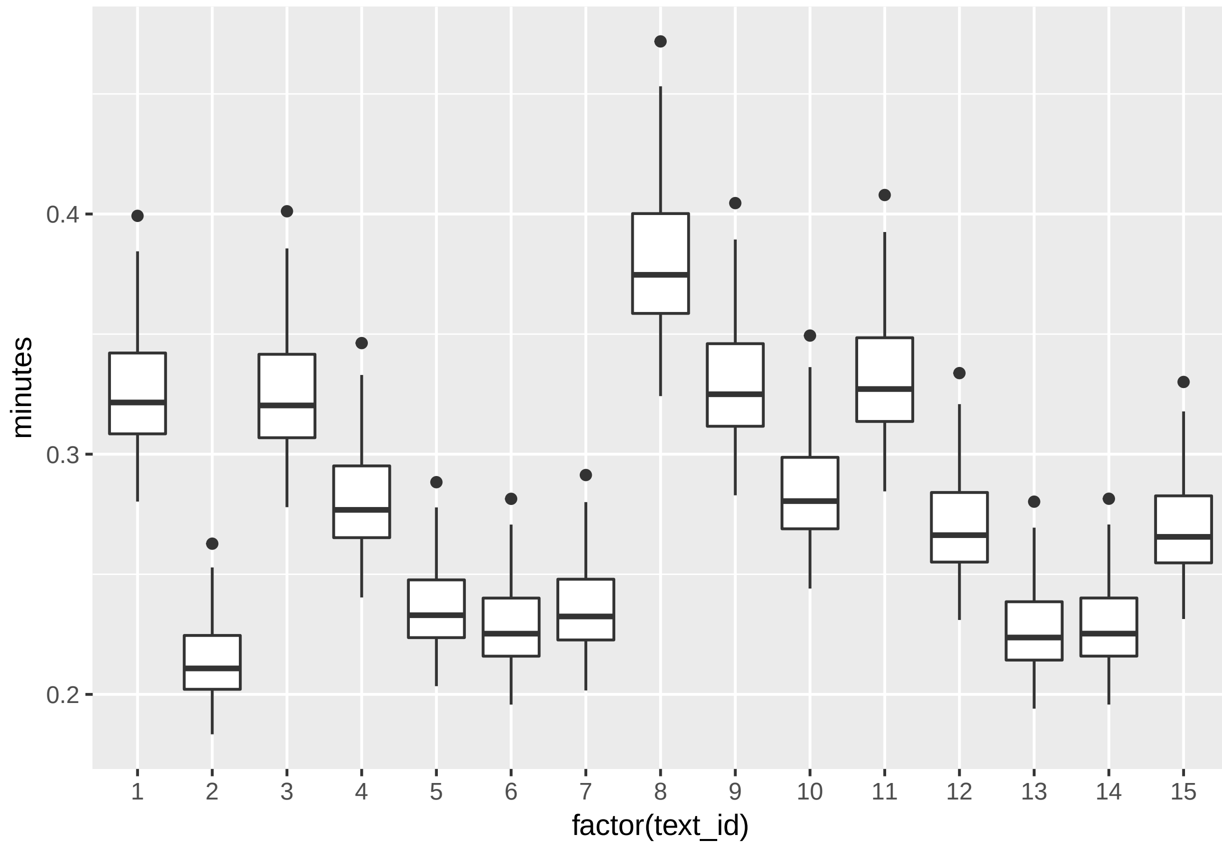
The diagram below shows the reading times by subject.

```
subjects_box.plot
```



The diagram below shows the reading times per text. These measurements are from all subject, with both eyes.

```
texts_box.plot
```



Statistics

There are two goals of this analysis. First, identify from the data some of the parameters used in the model. Second, create a linear mixed model with variables of native language (native vs non-native) and eye dominancy (dominant vs weak).

The first step in the analysis was to filter the data set to include only dominant eye and weak eye conditions from the four possible eye conditions. Next we could begin with some summaries of the data. In the first chart below, we found the summarized data for native speakers. In the second chart, the summarized data for non-native speakers.

```
summary_native.stats
```

```
##     text_id     subject_id    text_length      language
##  Min.   : 1   Min.   : 1.0   Min.   :51.0   foreign:  0
##  1st Qu.: 4   1st Qu.: 8.0   1st Qu.:54.0   native :900
##  Median : 8   Median :15.5   Median :65.0
##  Mean   : 8   Mean   :15.5   Mean   :65.8
##  3rd Qu.:12   3rd Qu.:23.0   3rd Qu.:79.0
##  Max.   :15   Max.   :30.0   Max.   :87.0
##             eye         minutes
##  both_eyes     :  0   Min.   :0.1858
##  dominant_eye  :450   1st Qu.:0.2249
##  prosthetic_eye:  0   Median :0.2620
##  weak_eye      :450   Mean   :0.2683
##                       3rd Qu.:0.3075
##                       Max.   :0.4088
```

```
summary_foreign.stats
```

```
##     text_id     subject_id    text_length      language  
##  Min.   : 1   Min.   :31.0   Min.   :51.0   foreign:900  
##  1st Qu.: 4   1st Qu.:38.0   1st Qu.:54.0   native :  0  
##  Median : 8   Median :45.5   Median :65.0                
##  Mean   : 8   Mean   :45.5   Mean   :65.8                
##  3rd Qu.:12   3rd Qu.:53.0   3rd Qu.:79.0                
##  Max.   :15   Max.   :60.0   Max.   :87.0                
##              eye         minutes      
##  both_eyes     :  0   Min.   :0.1987  
##  dominant_eye  :450   1st Qu.:0.2496  
##  prosthetic_eye:  0   Median :0.2898  
##  weak_eye      :450   Mean   :0.2986  
##                       3rd Qu.:0.3416  
##                       Max.   :0.4820
```

Some math will help us to confirm that the two populations follow the average WPM set in the model.

The Native readers had a mean reading time of 0.2683 minutes. Non-native readers took on average 0.2986. In both cases the average text came out to 65 words.

Words Per Minute = Words / Minutes

This gives Natives readers an average WPM = 242 compared with the models 250, and Non-natives a WPM = 217 compared with the models 230. What can account for these differences? The models set parameters of 250 and 230 account for no effects, where as these numbers account for cases of single eye use, both which carried WPM "penalties" in the simulation. So, though it may seem low, it would only be surprising if we had the models expected 250 WPM or higher.

The next step is to create our linear mixed model.

The formula as seen below: minutes ~ eye + language + (1 | subject_id) + (1 | text_id)

Here we are predicting the reading time, denoted "minutes", by the fixed effects of "eye", which eye was used, and "language", if a subject is native or not. The random effects then are the subject shown by "subject_id" and the text, shown by "text_id".

From the "Fixed Effects:" the model tells us two important things. First, going from the dominant eye to the week eye adds ~0.0028 minutes to the reading time. Second, being a native has an effect of decreasing the reading time by ~0.0302 minutes.

```
summary(minutes.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: minutes ~ eye + language + (1 | subject_id) + (1 | text_id)
##    Data: data_stats
## 
##      AIC      BIC   logLik deviance df.resid 
## -13884.8 -13851.8   6948.4 -13896.8     1794 
## 
## Scaled residuals: 
##     Min      1Q  Median      3Q     Max 
## -3.9586 -0.4807 -0.0119  0.4735  6.3196 
## 
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  subject_id (Intercept) 3.071e-04 0.017525
```

```
##  text_id    (Intercept) 2.497e-03 0.049966
##  Residual               1.959e-05 0.004426
## Number of obs: 1800, groups:  subject_id, 60; text_id, 15
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)     0.2971607  0.0132933  22.354
## eyeweak_eye     0.0028376  0.0002087  13.599
## languagenative -0.0302822  0.0045297  -6.685
##
## Correlation of Fixed Effects:
##             (Intr) eywk_y
## eyeweak_eye -0.008
## languagentv -0.170  0.000
```

Now we want to know if these two fixed effects of eye dominancy and language has any significant effect on the model. We can test this by creating null models that do not account for these variables and "assume" they do not make a difference. We can compare our complete model against these two null models using ANOVA.

We create our first null model, excluding eye dominancy as a fixed effect, then use an ANOVA to test for signifiance between our first model.

What we find is that eye dominancy had an effect on the reading time with a chi square = 175.68, or a p = 2.2e-16.

```r
summary(minutes_null_eye.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: minutes ~ language + (1 | subject_id) + (1 | text_id)
##    Data: data_stats
##
##      AIC      BIC   logLik deviance df.resid
## -13711.1 -13683.6   6860.6 -13721.1     1795
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0679 -0.5217  0.0192  0.5281  6.3132
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  subject_id (Intercept) 3.070e-04 0.017523
##  text_id    (Intercept) 2.497e-03 0.049966
##  Residual               2.169e-05 0.004657
## Number of obs: 1800, groups:  subject_id, 60; text_id, 15
##
## Fixed effects:
##                Estimate Std. Error t value
## (Intercept)     0.29858    0.01329  22.462
## languagenative -0.03028    0.00453  -6.685
##
## Correlation of Fixed Effects:
##             (Intr)
## languagentv -0.170
```

```r
anova(minutes.model, minutes_null_eye.model)
```

```
## Data: data_stats
## Models:
## minutes_null_eye.model: minutes ~ language + (1 | subject_id) + (1 | text_id)
## minutes.model: minutes ~ eye + language + (1 | subject_id) + (1 | text_id)
##                         Df    AIC    BIC logLik deviance  Chisq Chi Df
## minutes_null_eye.model  5 -13711 -13684 6860.6   -13721
## minutes.model           6 -13885 -13852 6948.4   -13897 175.68      1
##                         Pr(>Chisq)
## minutes_null_eye.model
## minutes.model           < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Next we create our null model, excluding language as a fixed effect and again test with an ANOVA.

What we find is that language affected reading time with a chi square of 33.28 and a $p = 7.981e\text{-}09$. This is surprising because the eye dominancy, though having a smaller effect on the data, test with a higher signifiance (ie lower p value). The reason for this might be because the effects between dominant and weak eye share less of a cross over due to small standard deviations, thereby acting as a weak, but seperative factor.

```r
summary(minutes_null_language.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: minutes ~ eye + (1 | subject_id) + (1 | text_id)
##    Data: data_stats
##
##      AIC      BIC   logLik deviance df.resid
## -13853.5 -13826.0   6931.8 -13863.5     1795
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9574 -0.4874 -0.0119  0.4687  6.3135
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.
##  subject_id (Intercept) 5.399e-04 0.023237
##  text_id    (Intercept) 2.500e-03 0.050002
##  Residual               1.959e-05 0.004426
## Number of obs: 1800, groups:  subject_id, 60; text_id, 15
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept) 0.2820196  0.0132552   21.28
## eyeweak_eye 0.0028376  0.0002087   13.60
##
## Correlation of Fixed Effects:
##             (Intr)
## eyeweak_eye -0.008
```

```r
anova(minutes.model, minutes_null_language.model)
```

```
## Data: data_stats
## Models:
## minutes_null_language.model: minutes ~ eye + (1 | subject_id) + (1 | text_id)
```

```
## minutes.model: minutes ~ eye + language + (1 | subject_id) + (1 | text_id)
##                            Df    AIC    BIC logLik deviance Chisq Chi Df
## minutes_null_language.model  5 -13854 -13826 6931.8   -13864
## minutes.model                6 -13885 -13852 6948.4   -13897 33.28      1
##                            Pr(>Chisq)
## minutes_null_language.model
## minutes.model                7.981e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To end the analysis, we will build a model with random slopes for eye dominancy. Below is the summary, followed by the list of coefficients. We see that there is consistency in the slopes, and that in each case the weak eye adds to the reading time.

```r
summary(minutes_random.model)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: minutes ~ eye + language + (1 + eye | subject_id) + (1 + eye |
##     text_id)
##    Data: data_stats
##
##      AIC      BIC   logLik deviance df.resid
## -13890.4 -13835.4   6955.2 -13910.4     1790
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.9079 -0.4777  0.0092  0.4492  6.4258
##
## Random effects:
##  Groups     Name        Variance  Std.Dev.  Corr
##  subject_id (Intercept) 3.029e-04 0.0174050
##             eyeweak_eye 1.405e-06 0.0011853 0.20
##  text_id    (Intercept) 2.471e-03 0.0497085
##             eyeweak_eye 2.652e-07 0.0005149 1.00
##  Residual               1.916e-05 0.0043769
## Number of obs: 1800, groups:  subject_id, 60; text_id, 15
##
## Fixed effects:
##                  Estimate Std. Error t value
## (Intercept)     0.2966226  0.0132225  22.433
## eyeweak_eye     0.0028376  0.0002892   9.810
## languagenative -0.0292060  0.0044882  -6.507
##
## Correlation of Fixed Effects:
##             (Intr) eywk_y
## eyeweak_eye  0.459
## languagentv -0.170  0.000
```

```r
coef(minutes_random.model)
```

```
## $subject_id
##    (Intercept)  eyeweak_eye languagenative
## 1    0.2903251 0.0019540921    -0.02920599
## 2    0.2825733 0.0026927988    -0.02920599
## 3    0.2849670 0.0025955925    -0.02920599
## 4    0.2862203 0.0023111582    -0.02920599
```

```
## 5    0.2958167 0.0029605828    -0.02920599
## 6    0.3235853 0.0033775402    -0.02920599
## 7    0.3006805 0.0023327125    -0.02920599
## 8    0.3096600 0.0020719496    -0.02920599
## 9    0.3029134 0.0020863973    -0.02920599
## 10   0.3048274 0.0032721958    -0.02920599
## 11   0.3016753 0.0023823436    -0.02920599
## 12   0.2764138 0.0028395357    -0.02920599
## 13   0.3098929 0.0028377407    -0.02920599
## 14   0.2706630 0.0025945841    -0.02920599
## 15   0.3045193 0.0037400987    -0.02920599
## 16   0.2939759 0.0028067265    -0.02920599
## 17   0.3063395 0.0024481261    -0.02920599
## 18   0.3013967 0.0028026685    -0.02920599
## 19   0.2883639 0.0023728858    -0.02920599
## 20   0.3206078 0.0040250257    -0.02920599
## 21   0.2823954 0.0031284674    -0.02920599
## 22   0.3039002 0.0017519900    -0.02920599
## 23   0.3106668 0.0027533349    -0.02920599
## 24   0.2984916 0.0033264870    -0.02920599
## 25   0.2864072 0.0023264406    -0.02920599
## 26   0.2840250 0.0026723605    -0.02920599
## 27   0.2873534 0.0028764867    -0.02920599
## 28   0.2955048 0.0032046411    -0.02920599
## 29   0.2953189 0.0019469749    -0.02920599
## 30   0.2854019 0.0019394359    -0.02920599
## 31   0.2899552 0.0036608392    -0.02920599
## 32   0.2802894 0.0023847765    -0.02920599
## 33   0.2920134 0.0017381277    -0.02920599
## 34   0.3204319 0.0016914037    -0.02920599
## 35   0.3461274 0.0031877030    -0.02920599
## 36   0.2586631 0.0017147518    -0.02920599
## 37   0.3236360 0.0040955319    -0.02920599
## 38   0.2766370 0.0030015575    -0.02920599
## 39   0.2929264 0.0020234578    -0.02920599
## 40   0.3207158 0.0038185380    -0.02920599
## 41   0.3022704 0.0039975545    -0.02920599
## 42   0.3037709 0.0022537099    -0.02920599
## 43   0.2703453 0.0032439818    -0.02920599
## 44   0.2878966 0.0032824725    -0.02920599
## 45   0.2800120 0.0029832889    -0.02920599
## 46   0.2712898 0.0022130242    -0.02920599
## 47   0.3162931 0.0039922034    -0.02920599
## 48   0.2769211 0.0036750879    -0.02920599
## 49   0.2887254 0.0028514841    -0.02920599
## 50   0.2988301 0.0031312092    -0.02920599
## 51   0.2586299 0.0027100297    -0.02920599
## 52   0.3032016 0.0032710268    -0.02920599
## 53   0.3335993 0.0033119085    -0.02920599
## 54   0.3256092 0.0042896537    -0.02920599
## 55   0.2925322 0.0009742572    -0.02920599
## 56   0.3006208 0.0026425588    -0.02920599
## 57   0.2985056 0.0023534907    -0.02920599
## 58   0.3028965 0.0039233888    -0.02920599
```

```
## 59   0.2859418 0.0030731705    -0.02920599
## 60   0.3131852 0.0043315780    -0.02920599
##
## $text_id
##    (Intercept) eyeweak_eye languagenative
## 1    0.3453633 0.003342468    -0.02920599
## 2    0.2316219 0.002164195    -0.02920599
## 3    0.3445774 0.003334327    -0.02920599
## 4    0.2997289 0.002869731    -0.02920599
## 5    0.2541430 0.002397497    -0.02920599
## 6    0.2466020 0.002319378    -0.02920599
## 7    0.2541144 0.002397200    -0.02920599
## 8    0.4009469 0.003918272    -0.02920599
## 9    0.3490392 0.003380548    -0.02920599
## 10   0.3032932 0.002906655    -0.02920599
## 11   0.3513652 0.003404643    -0.02920599
## 12   0.2889854 0.002758437    -0.02920599
## 13   0.2450669 0.002303475    -0.02920599
## 14   0.2466297 0.002319665    -0.02920599
## 15   0.2878614 0.002746793    -0.02920599
##
## attr(,"class")
## [1] "coef.mer"
```

Conclusion

A Reading Speed Simulator was created in R with fixed parameters for eye dominancy and nativeness. Then a visual analysis was performed of the data, a simple statistics analysis, and a more advanced mixed model analysis. Last, we created a model with random slopes for eye dominancy. In conclusion, we found that the effects of both eye dominancy and nativeness are significant factors.