My Notebook

true

```
source("My_Script.R")
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.4.4     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Loading required package: tinylabels
```

```
# Call the hello_world function
result1 <- hello_world("Dr.Dowling")
result2 <- hello_world("Voldemort")
result3 <- hello_world("Wubalubadubdub")

# Print the results
print(result1)
```

```
## [1] "Good morning,Dr.Dowling!"
```

```
print(result2)
```

```
## [1] "Voldemort, Expelliarmus!"
```

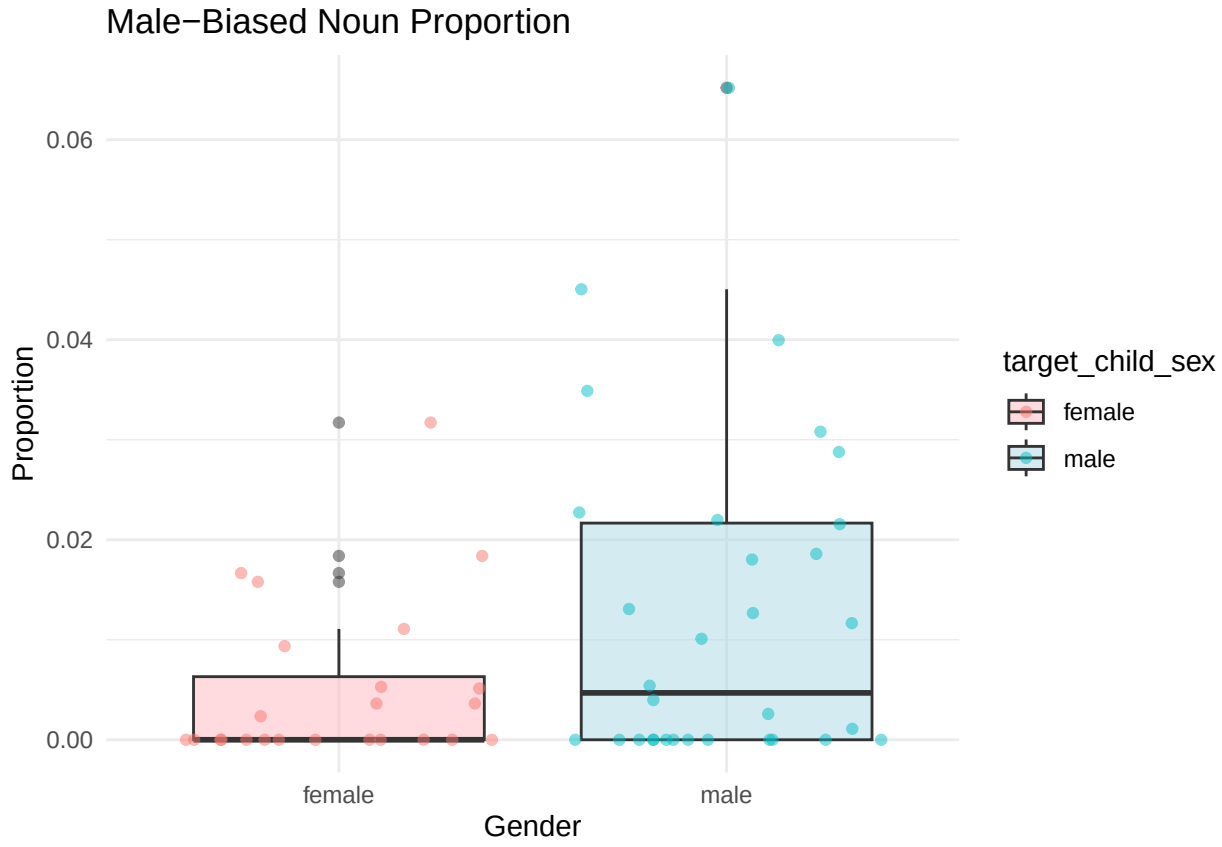
```
print(result3)
```

```
## [1] "I don't know you:("
```

dataset

Assignment 10 & 12

```r
# Data analysis

m.data <- read.csv("//Users/katia/Desktop/D2M/CHILDES_utterance/Childes dataset/Proportion data/noun_pr

ggplot(m.data, aes(x = target_child_sex, y = mnoun_proportion, fill = target_child_sex)) +
  geom_boxplot(alpha = 0.5) +
  geom_jitter(alpha = 0.5, aes(colour = target_child_sex)) +
  scale_fill_manual(values = c("female" = "lightpink", "male" = "lightblue")) +
  labs(x = "Gender", y = "Proportion", title = "Male-Biased Noun Proportion") +
  theme_minimal()
```



X-axis: The gloss variable, representing different words (like “dress”, “doll”, “necklace”, etc.). Y-axis: The count or frequency of each word, indicating how often each word appears in the dataset. Type of Plot: Bar Chart. Comparison Across Groups?: Yes Anticipated Findings: mothers speak more female-biased nouns to female infants than to male infants

As seen in Figure @ref(fig:my-figure), There’s gender bias in the nouns of mother’s infant-directed speech.

Assignment 13. Prettified Plot

```r
combined.data <- read.csv("//Users/katia/Desktop/D2M/CHILDES_utterance/Childes dataset/Proportion data/

library(tidyr)
```

```r
long_data <- combined.data %>%
  pivot_longer(cols = c(mnoun_proportion, fnoun_proportion),
               names_to = "type",
               values_to = "proportion")
labels_gender = c('fnoun_proportion' = "Female-Biased", 'mnoun_proportion' = "Male-Biased")

ggplot(long_data, aes(x = target_child_sex, y = proportion, fill = target_child_sex)) +
  geom_boxplot(alpha = 0.5) +
  geom_jitter(alpha = 0.5, aes(colour = target_child_sex)) +
  scale_fill_manual(values = c("female" = "lightpink", "male" = "lightblue")) +
  labs(x = "Gender", y = "Proportion") +
  facet_wrap(~type, scales = "fixed", labeller = as_labeller(labels_gender)) +
  labs(title = "Proportion of Gender-Biased Nouns")+
  theme(
  strip.background = element_rect(fill = "gray60"),
  strip.text = element_text(colour = "white", family = "Courier", hjust = 0),
  legend.position = "bottom"
)
```

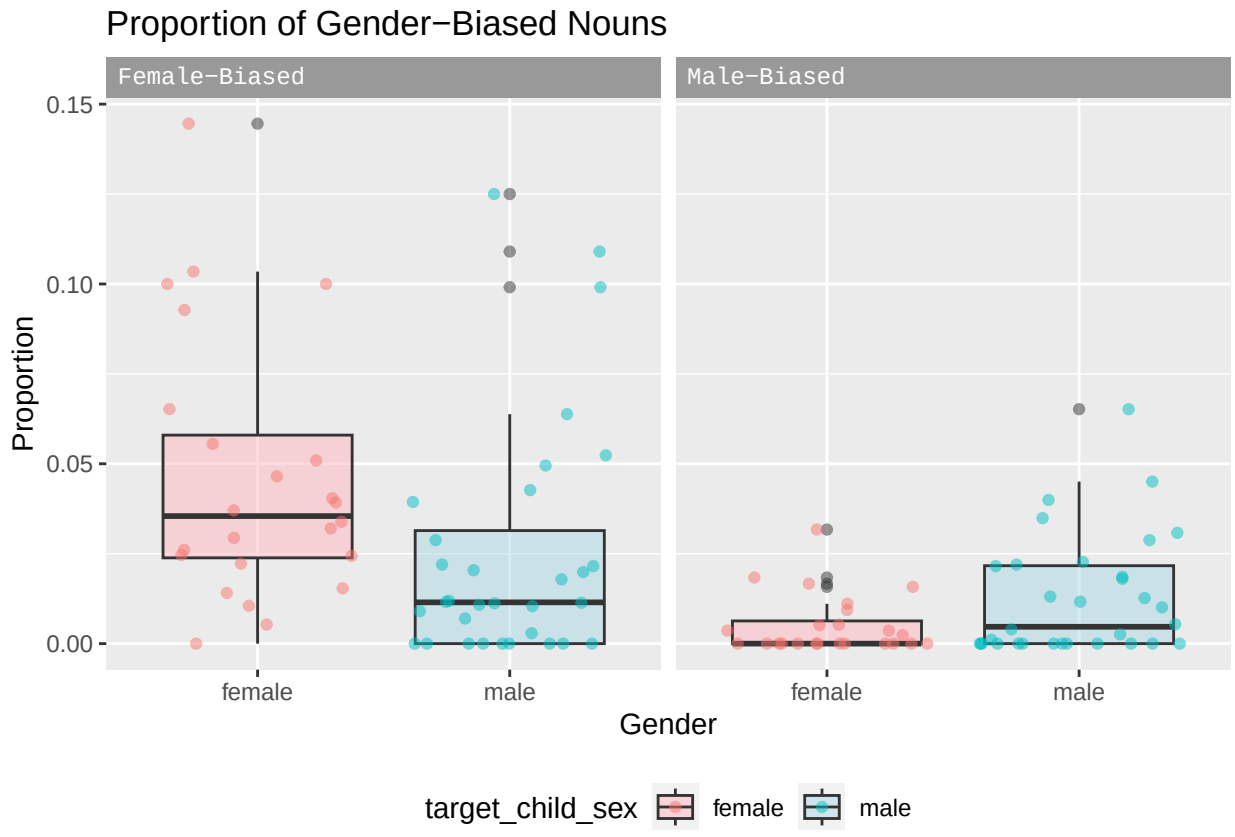


Figure 1: Proportion of Gender-Biased Nouns by Gender

X-axis: female/male infants
Y-axis: counts of frequency of female-biased nouns Type of Plot: box plot
Comparison Across Groups?: Yes
Anticipated Findings: mothers speak more female-biased nouns in CDS to female infants than to male infants

As shown in Table @ref(tab:my-table), the gender distribution in the dataset varies, with a higher percentage of Female compared to Male.

Assignment 14

```
library(kableExtra)
my_table <- kable(m.data,
                  format = "latex",
                  booktabs = TRUE,
                  caption = "Male-Infant Data",
                  label = "my-table") %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = FALSE) %>%
  column_spec(1, bold = TRUE, color = "blue") %>%
  row_spec(0, bold = TRUE, background = "#D3D3D3")

print(my_table)
```

```
## \begin{table}
## \centering
## \caption{\label{tab:my-table}Male-Infant Data}
## \centering
## \begin{tabular}[t]{>{}llrrrrr}
## \toprule
## \cellcolor[HTML]{D3D3D3}{\textbf{target\_child\_name}} & \cellcolor[HTML]{D3D3D3}{\textbf{target\_ch
## \midrule
## \textcolor{blue}{\textbf{Aaron}} & male & 57 & 0 & 0 & 0.0000000 & 0.0000000\\
## \textcolor{blue}{\textbf{Adam}} & male & 201 & 4 & 0 & 0.0000000 & 0.0199005\\
## \textcolor{blue}{\textbf{Alex}} & male & 19296 & 135 & 225 & 0.0116604 & 0.0069963\\
## \textcolor{blue}{\textbf{Alfred}} & male & 91 & 2 & 2 & 0.0219780 & 0.0219780\\
## \textcolor{blue}{\textbf{Alice}} & female & 2081 & 106 & 66 & 0.0317155 & 0.0509370\\
## \addlinespace
## \textcolor{blue}{\textbf{Allen}} & male & 108 & 0 & 0 & 0.0000000 & 0.0000000\\
## \textcolor{blue}{\textbf{Anthony}} & male & 39 & 0 & 0 & 0.0000000 & 0.0000000\\
## \textcolor{blue}{\textbf{Benjamin}} & male & 914 & 36 & 1 & 0.0010941 & 0.0393873\\
## \textcolor{blue}{\textbf{Brian}} & male & 89 & 1 & 0 & 0.0000000 & 0.0112360\\
## \textcolor{blue}{\textbf{Brooke}} & female & 97 & 9 & 0 & 0.0000000 & 0.0927835\\
## \addlinespace
## \textcolor{blue}{\textbf{Carol}} & female & 83 & 12 & 0 & 0.0000000 & 0.1445783\\
## \textcolor{blue}{\textbf{Danielle}} & female & 90 & 5 & 0 & 0.0000000 & 0.0555556\\
## \textcolor{blue}{\textbf{David}} & male & 688 & 75 & 24 & 0.0348837 & 0.1090116\\
## \textcolor{blue}{\textbf{Doug}} & male & 88 & 1 & 2 & 0.0227273 & 0.0113636\\
## \textcolor{blue}{\textbf{Emily}} & female & 142 & 2 & 0 & 0.0000000 & 0.0140845\\
## \addlinespace
## \textcolor{blue}{\textbf{Emma}} & female & 180 & 4 & 3 & 0.0166667 & 0.0222222\\
## \textcolor{blue}{\textbf{Erica}} & female & 123 & 3 & 0 & 0.0000000 & 0.0243902\\
## \textcolor{blue}{\textbf{Ethan}} & male & 23210 & 269 & 715 & 0.0308057 & 0.0115898\\
## \textcolor{blue}{\textbf{Jarret}} & male & 99 & 0 & 1 & 0.0101010 & 0.0000000\\
## \textcolor{blue}{\textbf{Jas}} & male & 185 & 2 & 1 & 0.0054054 & 0.0108108\\
## \addlinespace
## \textcolor{blue}{\textbf{Jase}} & male & 1777 & 88 & 71 & 0.0399550 & 0.0495217\\
## \textcolor{blue}{\textbf{Jeff}} & male & 88 & 0 & 0 & 0.0000000 & 0.0000000\\
```

```
## \textcolor{blue}{\textbf{Jessica}} & female & 120 & 12 & 0 & 0.0000000 & 0.1000000\\
## \textcolor{blue}{\textbf{Jillian}} & female & 2550 & 75 & 6 & 0.0023529 & 0.0294118\\
## \textcolor{blue}{\textbf{Johnnie}} & male & 222 & 22 & 10 & 0.0450450 & 0.0990991\\
## \addlinespace
## \textcolor{blue}{\textbf{Katie}} & female & 81 & 3 & 0 & 0.0000000 & 0.0370370\\
## \textcolor{blue}{\textbf{Kevin}} & male & 139 & 4 & 4 & 0.0287770 & 0.0287770\\
## \textcolor{blue}{\textbf{Kimberly}} & female & 86 & 4 & 0 & 0.0000000 & 0.0465116\\
## \textcolor{blue}{\textbf{Laura}} & female & 1360 & 55 & 25 & 0.0183824 & 0.0404412\\
## \textcolor{blue}{\textbf{Laura\_Aurie}} & female & 4 & 0 & 0 & 0.0000000 & 0.0000000\\
## \addlinespace
## \textcolor{blue}{\textbf{Laurel}} & female & 276 & 18 & 1 & 0.0036232 & 0.0652174\\
## \textcolor{blue}{\textbf{Lily}} & female & 38035 & 1289 & 356 & 0.0093598 & 0.0338898\\
## \textcolor{blue}{\textbf{Martin}} & male & 773 & 33 & 2 & 0.0025873 & 0.0426908\\
## \textcolor{blue}{\textbf{Matt}} & male & 3947 & 85 & 50 & 0.0126678 & 0.0215353\\
## \textcolor{blue}{\textbf{Matthew}} & male & 49 & 0 & 0 & 0.0000000 & 0.0000000\\
## \addlinespace
## \textcolor{blue}{\textbf{Megan}} & female & 190 & 2 & 3 & 0.0157895 & 0.0105263\\
## \textcolor{blue}{\textbf{Naima}} & female & 36986 & 964 & 410 & 0.0110853 & 0.0260639\\
## \textcolor{blue}{\textbf{Nanette}} & female & 609 & 15 & 0 & 0.0000000 & 0.0246305\\
## \textcolor{blue}{\textbf{Nicole}} & female & 100 & 10 & 0 & 0.0000000 & 0.1000000\\
## \textcolor{blue}{\textbf{Patricia}} & female & 567 & 3 & 3 & 0.0052910 & 0.0052910\\
## \addlinespace
## \textcolor{blue}{\textbf{Patrick}} & male & 8 & 1 & 0 & 0.0000000 & 0.1250000\\
## \textcolor{blue}{\textbf{Peter}} & male & 49 & 1 & 0 & 0.0000000 & 0.0204082\\
## \textcolor{blue}{\textbf{Richard}} & male & 345 & 1 & 0 & 0.0000000 & 0.0028986\\
## \textcolor{blue}{\textbf{Rick}} & male & 974 & 51 & 21 & 0.0215606 & 0.0523614\\
## \textcolor{blue}{\textbf{Robert}} & male & 56 & 1 & 0 & 0.0000000 & 0.0178571\\
## \addlinespace
## \textcolor{blue}{\textbf{Roman}} & male & 111 & 1 & 2 & 0.0180180 & 0.0090090\\
## \textcolor{blue}{\textbf{Ronny}} & male & 721 & 46 & 47 & 0.0651872 & 0.0638003\\
## \textcolor{blue}{\textbf{Sarah}} & female & 51 & 2 & 0 & 0.0000000 & 0.0392157\\
## \textcolor{blue}{\textbf{Scott}} & male & 91 & 0 & 0 & 0.0000000 & 0.0000000\\
## \textcolor{blue}{\textbf{Shawna}} & female & 29 & 3 & 0 & 0.0000000 & 0.1034483\\
## \addlinespace
## \textcolor{blue}{\textbf{Stephen}} & male & 153 & 0 & 2 & 0.0130719 & 0.0000000\\
## \textcolor{blue}{\textbf{Tommy}} & male & 83 & 0 & 0 & 0.0000000 & 0.0000000\\
## \textcolor{blue}{\textbf{Victor}} & male & 504 & 6 & 2 & 0.0039683 & 0.0119048\\
## \textcolor{blue}{\textbf{Violet}} & female & 14361 & 460 & 52 & 0.0036209 & 0.0320312\\
## \textcolor{blue}{\textbf{Wendy}} & female & 195 & 3 & 1 & 0.0051282 & 0.0153846\\
## \addlinespace
## \textcolor{blue}{\textbf{William}} & male & 15388 & 160 & 286 & 0.0185859 & 0.0103977\\
## \bottomrule
## \end{tabular}
## \end{table}
```

Assignment 15 & 16

Descriptive Analysis: Examining Maternal Gender-Biased Language Imput Patterns

```r
library(psych)
```

```
##
## Attaching package: 'psych'

## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```r
library(lme4)
```

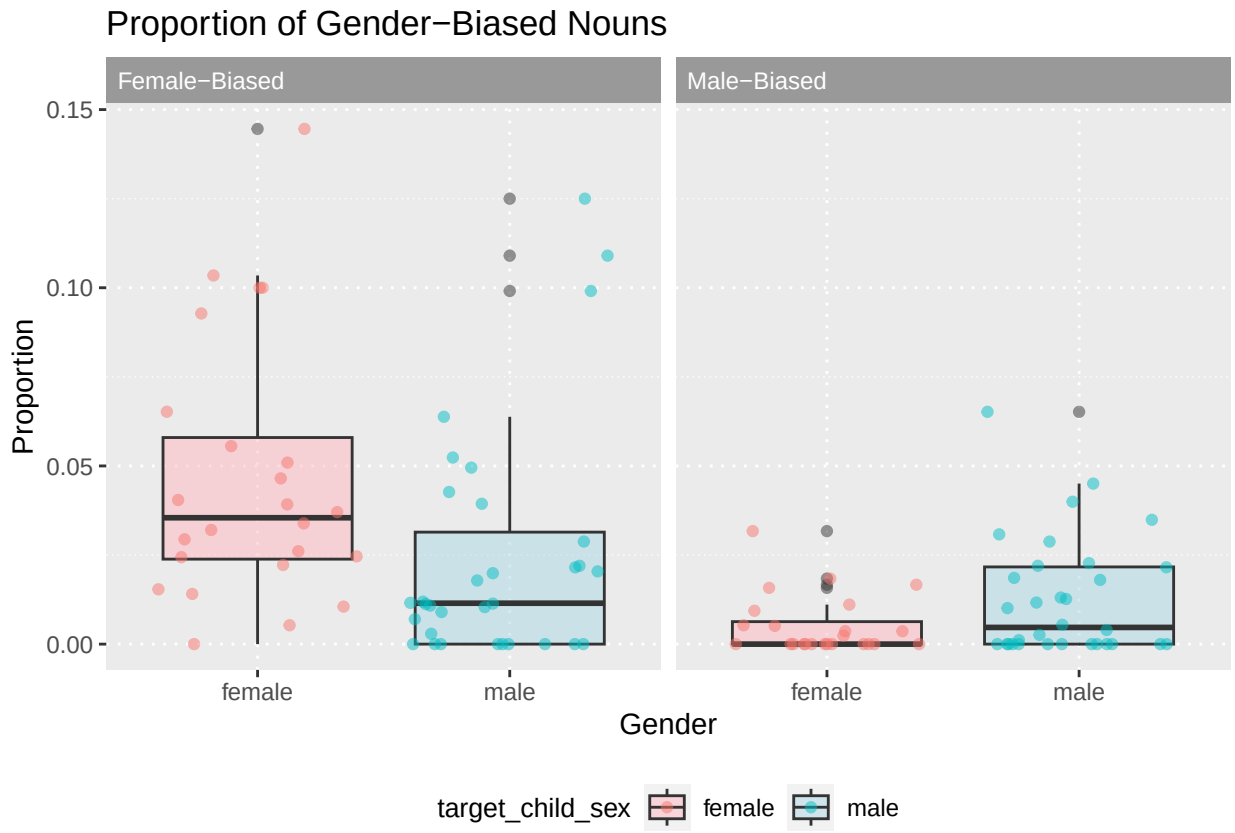
```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```r
   ## In this analysis, I aim to describe the patterns of gender-biased language usage by mothers in int
   ### - The frequency of appearance-related adjectives (e.g., "pretty", "beautiful") and competence-re
   ### - The frequency of gender-biased nouns (e.g., "truck" for boys, "doll" for girls) used by mother

  ## Independent Variables: Child's gender (binary: male or female)

## Assignment 16.1 *Descriptive Statistics* for male-biased nouns
# Calculate descriptive statistics for male-biased nouns
noun_tidied_count <- read.csv("/Users/katia/Desktop/D2M/CHILDES_utterance/Childes dataset/Proportion dat

summary_mnoun <- noun_tidied_count %>%
  group_by(target_child_sex) %>%
  summarise(mean_mnoun = mean(mnoun_proportion),
            median_mnoun = median(mnoun_proportion),
            sd_mnoun = sd(mnoun_proportion),
            .groups = 'drop')

# Print the summary statistics
print(summary_mnoun)
```

```
## # A tibble: 2 x 4
##   target_child_sex mean_mnoun median_mnoun sd_mnoun
##   <chr>                <dbl>        <dbl>    <dbl>
## 1 female             0.00513      0        0.00817
## 2 male               0.0128       0.00469  0.0165
```

```r
#Visualization of all stats
library(tidyr)
long_data <- noun_tidied_count %>%
  pivot_longer(cols = c(mnoun_proportion, fnoun_proportion),
```

```
               names_to = "type",
               values_to = "proportion")
labels_gender = c('fnoun_proportion' = "Female-Biased", 'mnoun_proportion' = "Male-Biased")

ggplot(long_data, aes(x = target_child_sex, y = proportion, fill = target_child_sex)) +
  geom_boxplot(alpha = 0.5) +
  geom_jitter(alpha = 0.5, aes(colour = target_child_sex)) +
  scale_fill_manual(values = c("female" = "lightpink", "male" = "lightblue")) +
  labs(x = "Gender", y = "Proportion") +
  facet_wrap(~type, scales = "fixed", labeller = as_labeller(labels_gender)) +
  labs(title = "Proportion of Gender-Biased Nouns")+
  theme(
  panel.grid = element_line(colour = "white", linetype = 3),
  legend.title = element_text(size = 10),
  strip.background = element_rect(fill = "gray60"),
  strip.text = element_text(colour = "white", hjust = 0),
  legend.position = "bottom"
)
```




```
  ## Dependent Variables:
  ### - Percentage of gender-biased adjectives (appearance-related and competence-related) in maternal
  ### - Percentage of gender-biased nouns in maternal language input.

  ## This descriptive analysis will involve calculating means and standard deviations for the usage of
```

Hypothesis Testing Analysis: Testing the Influence of Child's Gender on Maternal Language Input

```
library(stats)
noun_tidied_count <- read.csv("/Users/katia/Desktop/D2M/CHILDES_utterance/Childes dataset/Proportion da
# ANOVA for male-biased nouns
anova_mnoun <- aov(mnoun_proportion ~ target_child_sex, data = noun_tidied_count)
summary(anova_mnoun)
```

```
##                  Df   Sum Sq   Mean Sq F value Pr(>F)
## target_child_sex  1 0.000798 0.0007977   4.326 0.0423 *
## Residuals        54 0.009958 0.0001844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# ANOVA for female-biased nouns
anova_fnoun <- aov(fnoun_proportion ~ target_child_sex, data = noun_tidied_count)
summary(anova_fnoun)
```

```
##                  Df  Sum Sq  Mean Sq F value Pr(>F)
## target_child_sex  1 0.00633 0.006327   5.252 0.0259 *
## Residuals        54 0.06506 0.001205
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
  ## Hypothesized Results: It is hypothesized that there will be a significant difference in the percen

  ## Appropriate Tests: For comparing multiple variables between two groups (boys vs. girls), *ANOVA* c

  ### This analysis will help determine if there is a statistically significant difference in maternal
```

The ANOVA tests conducted on the proportions of gender-biased nouns used by mothers reveal the following: The F-statistic is $F() = .$

P-value: 0.0423 The results suggest that there is a statistically significant difference in the usage of male-biased nouns by mothers when speaking to male children compared to female children, at the conventional 0.05 significance level.

For female-biased nouns:

F-statistic: 5.25 P-value: 0.0259 Similarly, there is a statistically significant difference in the usage of female-biased nouns by mothers when speaking to male children compared to female children, also at the 0.05 significance level.