Count Data Linguistic Analysis

Dagmar Heintze

2024-09-10

```
#Set a Cran monitor
options(repos = c(CRAN = "https://cran.r-project.org"))
setwd("/Users/Dagmar Heintze/Desktop/R-Directory")
install.packages("ggplot2")
```

```
##
## The downloaded binary packages are in
##  /var/folders/gf/_b3p3ndn77s6fct3p08sxzq80000gp/T//RtmptbRiTi/downloaded_packages
```

```
library(ggplot2)
```

#Ensure long format # Load the tidyr package for converting to long format #load relevant library

```
library(tidyr)
```

#########Count Analyses for Arabic########## #Import the data

```
ar_count_data <- read.csv("/Users/Dagmar Heintze/Downloads/ar_counts.csv")
View(ar_count_data)
```

#Transform to long table and select variables # Lemma counts

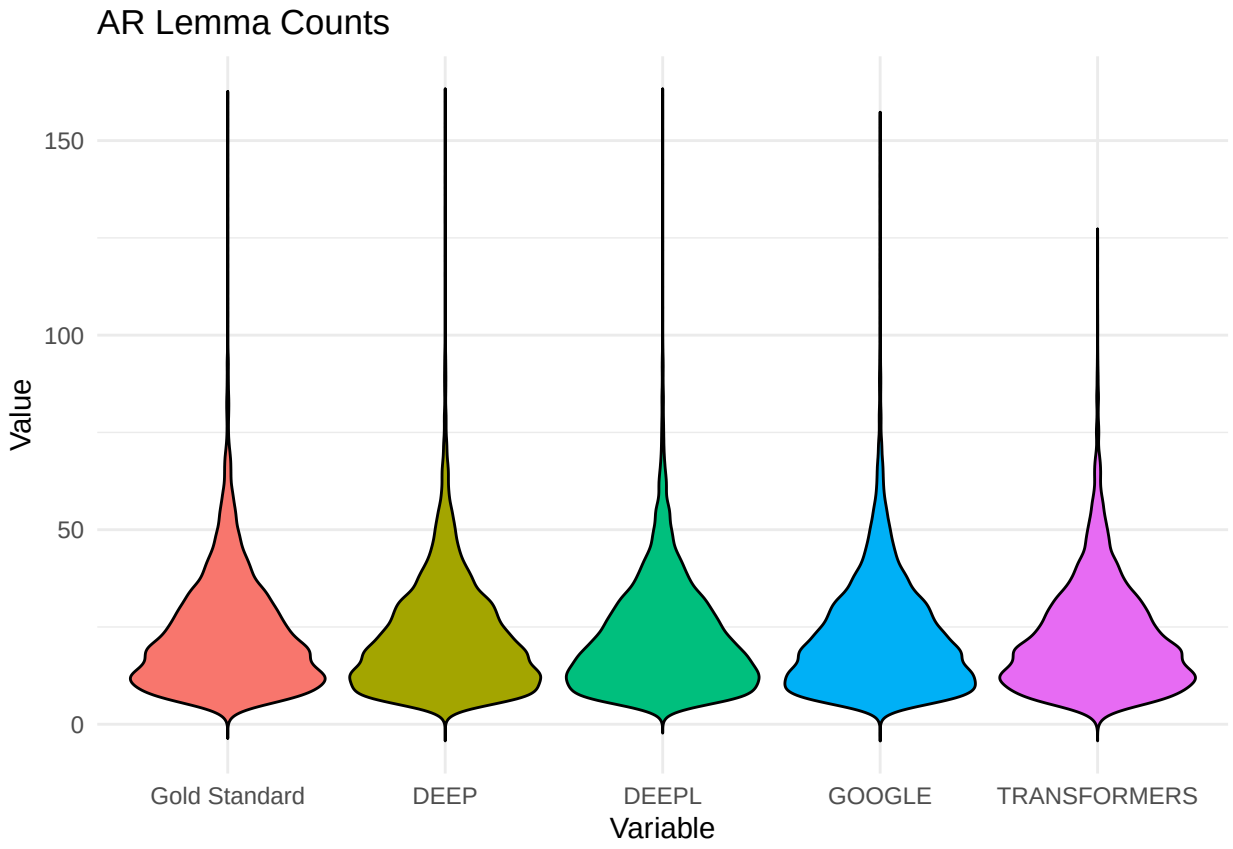
```
df_lemma_long <- pivot_longer(ar_count_data,
                              cols = c(ar_lemma_counts,
                                       en_ar_DEEP_lemma_counts,
                                       en_ar_DEEPL_lemma_counts,
                                       en_ar_GOOGLE_lemma_counts,
                                       en_ar_TRANSFORMERS_lemma_counts),
                              names_to = "Variable",
                              values_to = "Value")
```

Create the violin plot

AR Lemma counts

```
ggplot(df_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR Lemma Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_ydensity()‘).
```



#Transform to long table and select variables # Verb counts

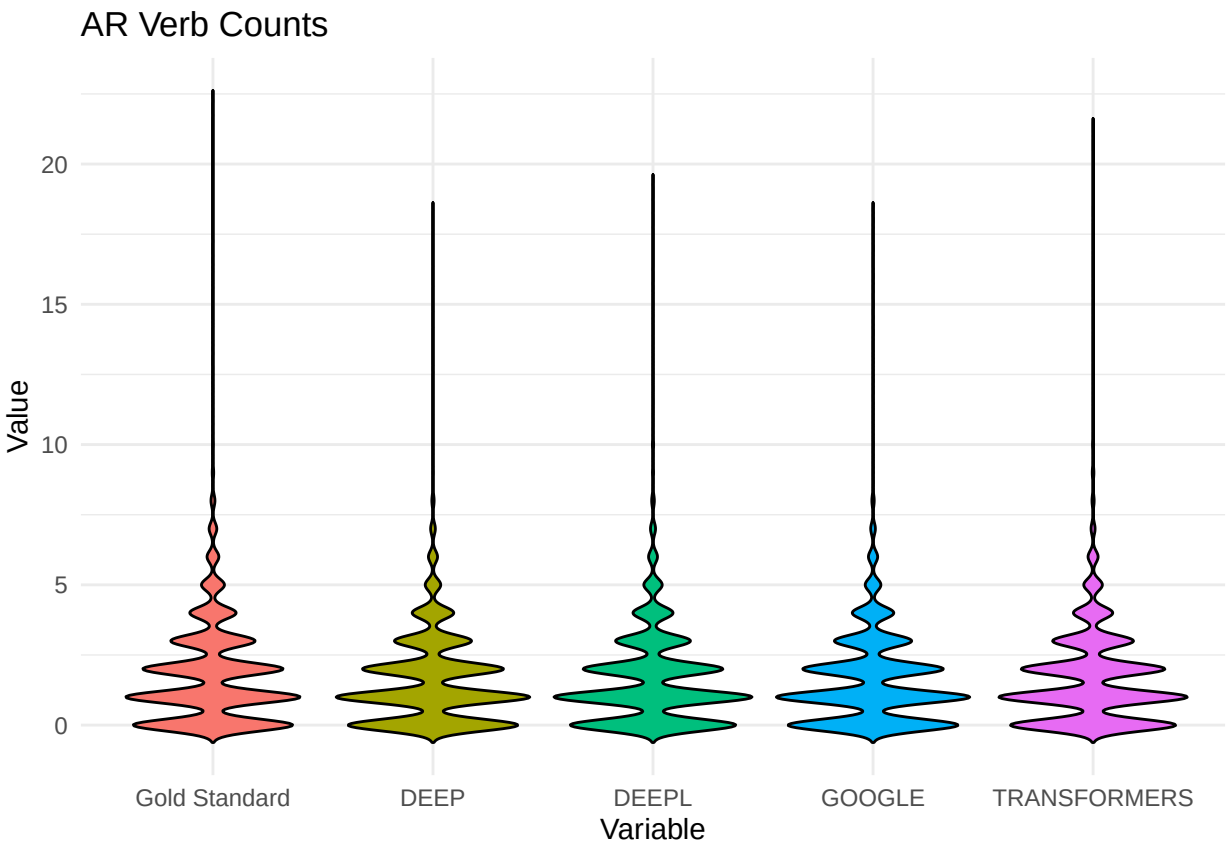
```
df_verb_long <- pivot_longer(ar_count_data,
                             cols = c(ar_verb_counts,
                                      en_ar_DEEP_verb_counts,
                                      en_ar_DEEPL_verb_counts,
                                      en_ar_GOOGLE_verb_counts,
                                      en_ar_TRANSFORMERS_verb_counts),
                             names_to = "Variable",
                             values_to = "Value")
```

Create the violin plot

AR Verb counts

```
ggplot(df_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR Verb Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_ydensity()‘).
```



#Transform to long table and select variables # Noun counts

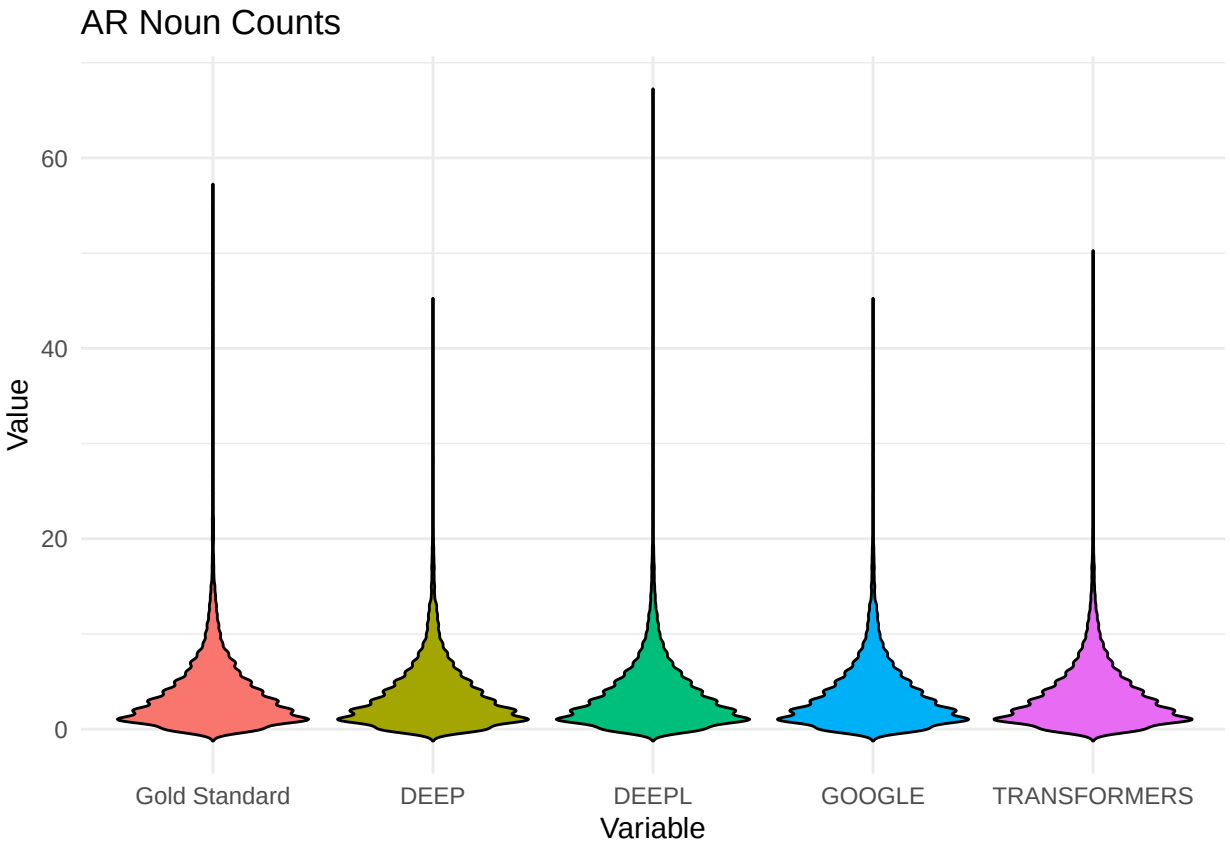
```
df_noun_long <- pivot_longer(ar_count_data,
                             cols = c(ar_noun_counts,
                                      en_ar_DEEP_noun_counts,
                                      en_ar_DEEPL_noun_counts,
                                      en_ar_GOOGLE_noun_counts,
                                      en_ar_TRANSFORMERS_noun_counts),
                             names_to = "Variable",
                             values_to = "Value")
```

Create the violin plot

AR Noun counts

```
ggplot(df_noun_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR Noun Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## (‘stat_ydensity()‘).
```



End Count Analysis Arabic

Count Analyses for English

#Import the data

```
en_count_data <- read.csv("/Users/Dagmar Heintze/Downloads/en_counts.csv")
View(en_count_data)
```

Analysis for ES-EN

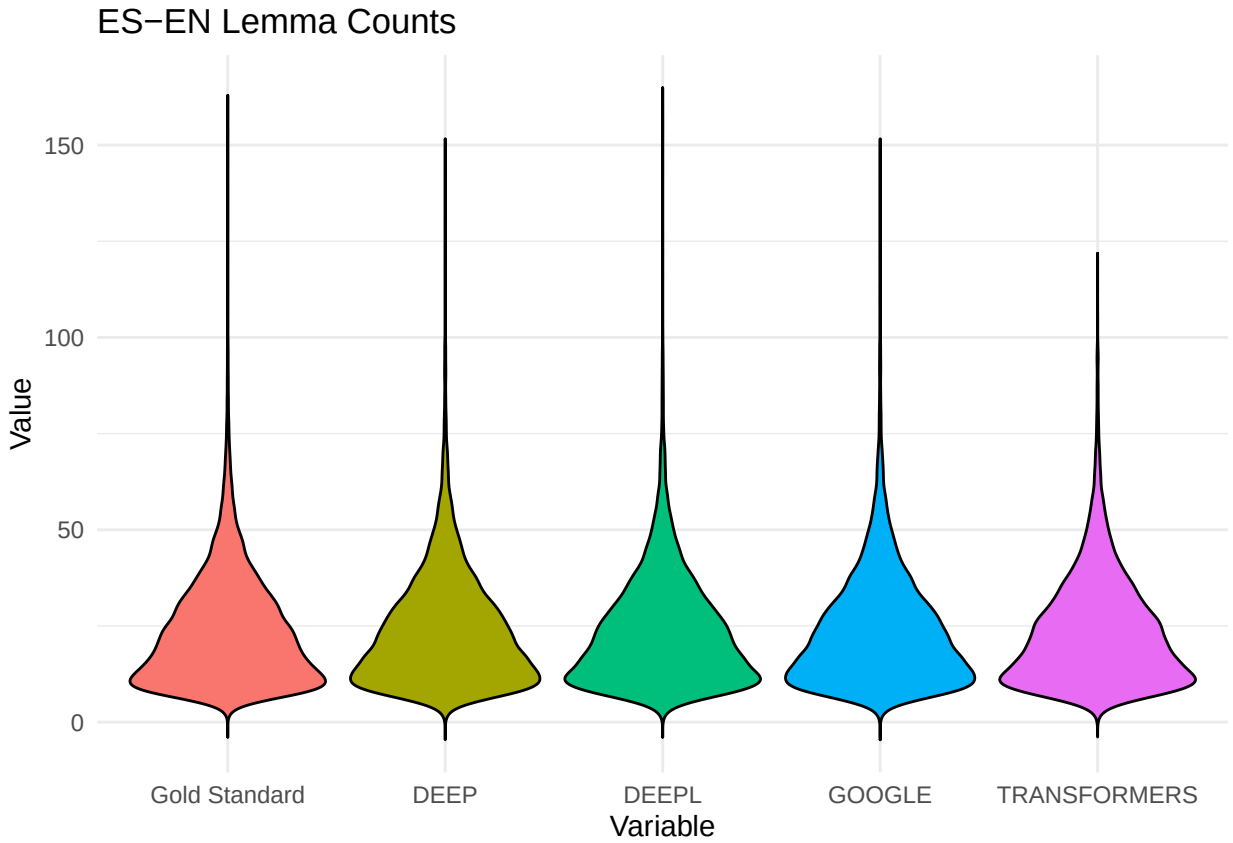
#Transform to long table and select variables # Lemma counts ES-EN

```r
df_es_en_lemma_long <- pivot_longer(en_count_data,
                                    cols = c(en_lemma_counts,
                                             es_en_DEEP_lemma_counts,
                                             es_en_DEEPL_lemma_counts,
                                             es_en_GOOGLE_lemma_counts,
                                             es_en_TRANSFORMERS_lemma_counts),
                                    names_to = "Variable",
                                    values_to = "Value")
```

Create the violin plot

ES-EN Lemma counts

```r
ggplot(df_es_en_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Lemma Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



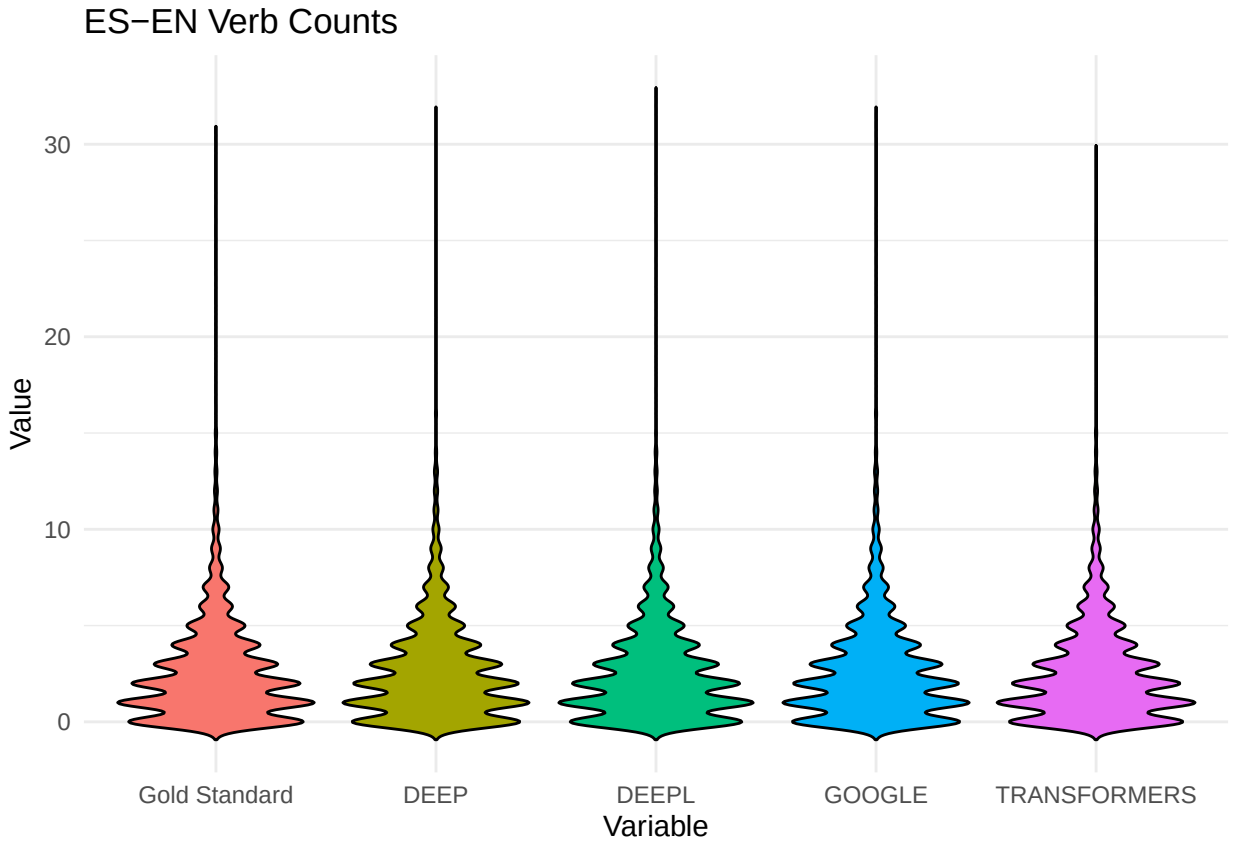
#Transform to long table and select variables # ES-EN Verb counts

```
df_es_en_verb_long <- pivot_longer(en_count_data,
                                   cols = c(en_verb_counts,
                                            es_en_DEEP_verb_counts,
                                            es_en_DEEPL_verb_counts,
                                            es_en_GOOGLE_verb_counts,
                                            es_en_TRANSFORMERS_verb_counts),
                                   names_to = "Variable",
                                   values_to = "Value")
```

Create the violin plot

ES-EN Verb counts

```
ggplot(df_es_en_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Verb Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



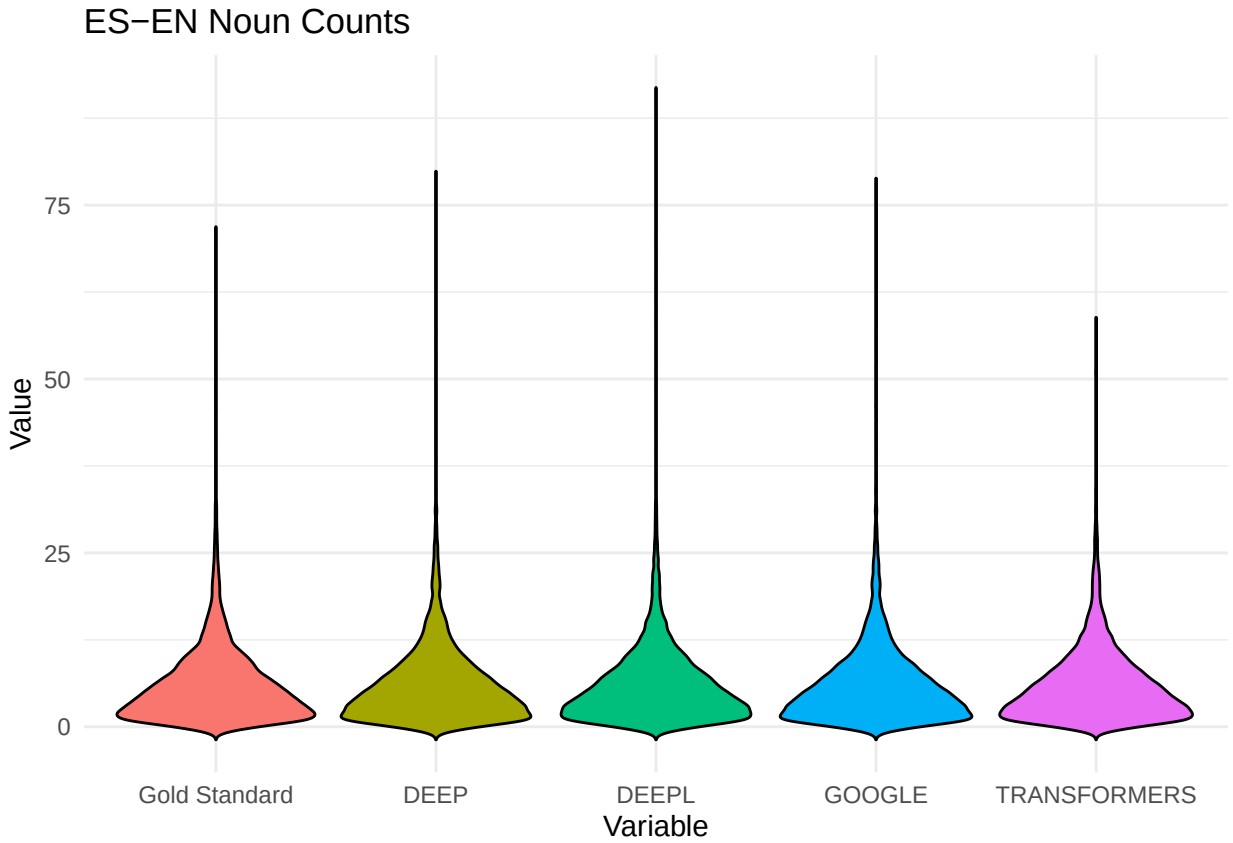
#Transform to long table and select variables # Noun counts ES-EN

```r
df_es_en_verb_long <- pivot_longer(en_count_data,
                                   cols = c(en_noun_counts,
                                            es_en_DEEP_noun_counts,
                                            es_en_DEEPL_noun_counts,
                                            es_en_GOOGLE_noun_counts,
                                            es_en_TRANSFORMERS_noun_counts),
                                   names_to = "Variable",
                                   values_to = "Value")
```

Create the violin plot

ES-EN Noun counts

```r
ggplot(df_es_en_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Noun Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



Analysis for AR-EN

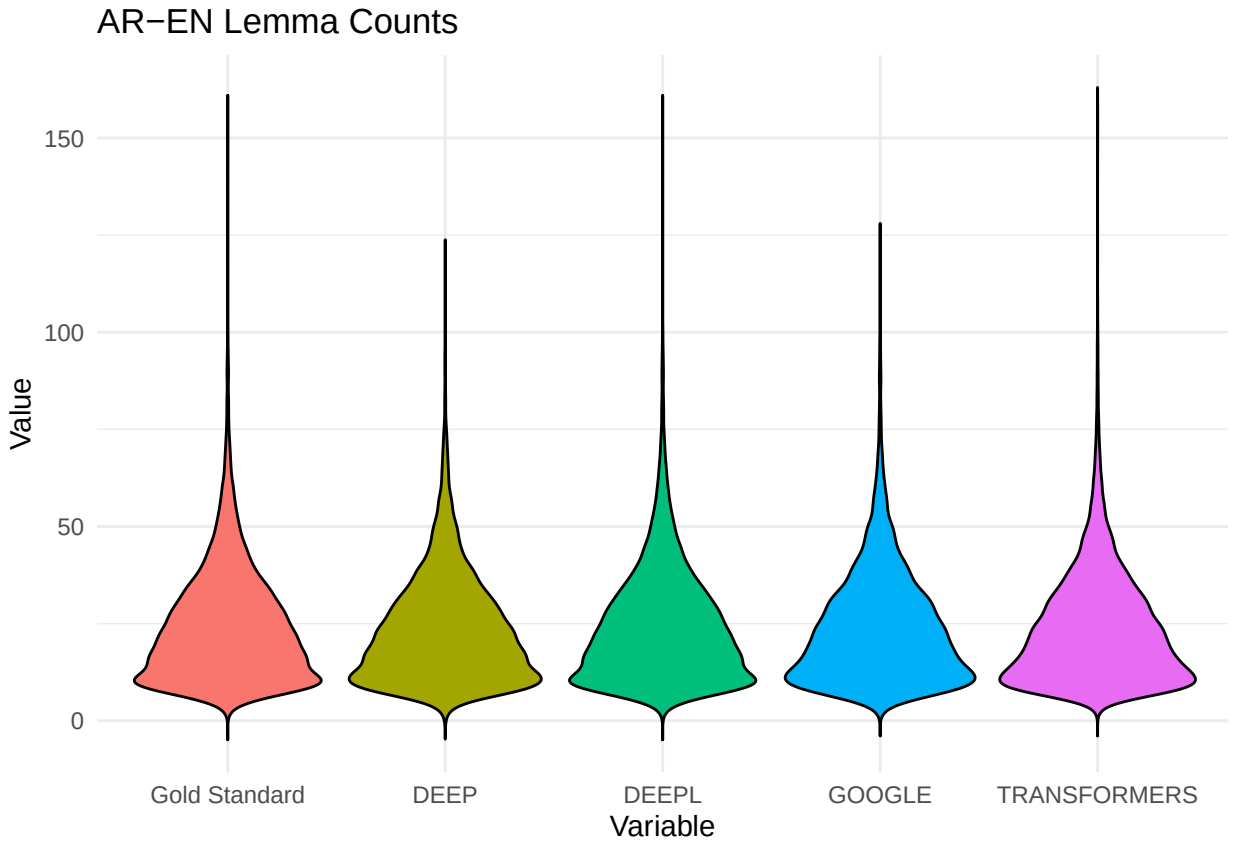
#Transform to long table and select variables # AR-EN Lemma counts

```r
df_ar_en_lemma_long <- pivot_longer(en_count_data,
                                    cols = c(en_lemma_counts,
                                             ar_en_DEEP_lemma_counts,
                                             ar_en_DEEPL_lemma_counts,
                                             ar_en_GOOGLE_lemma_counts,
                                             ar_en_TRANSFORMERS_lemma_counts),
                                    names_to = "Variable",
                                    values_to = "Value")
```

Create the violin plot

AR-EN Lemma counts

```r
ggplot(df_ar_en_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR-EN Lemma Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



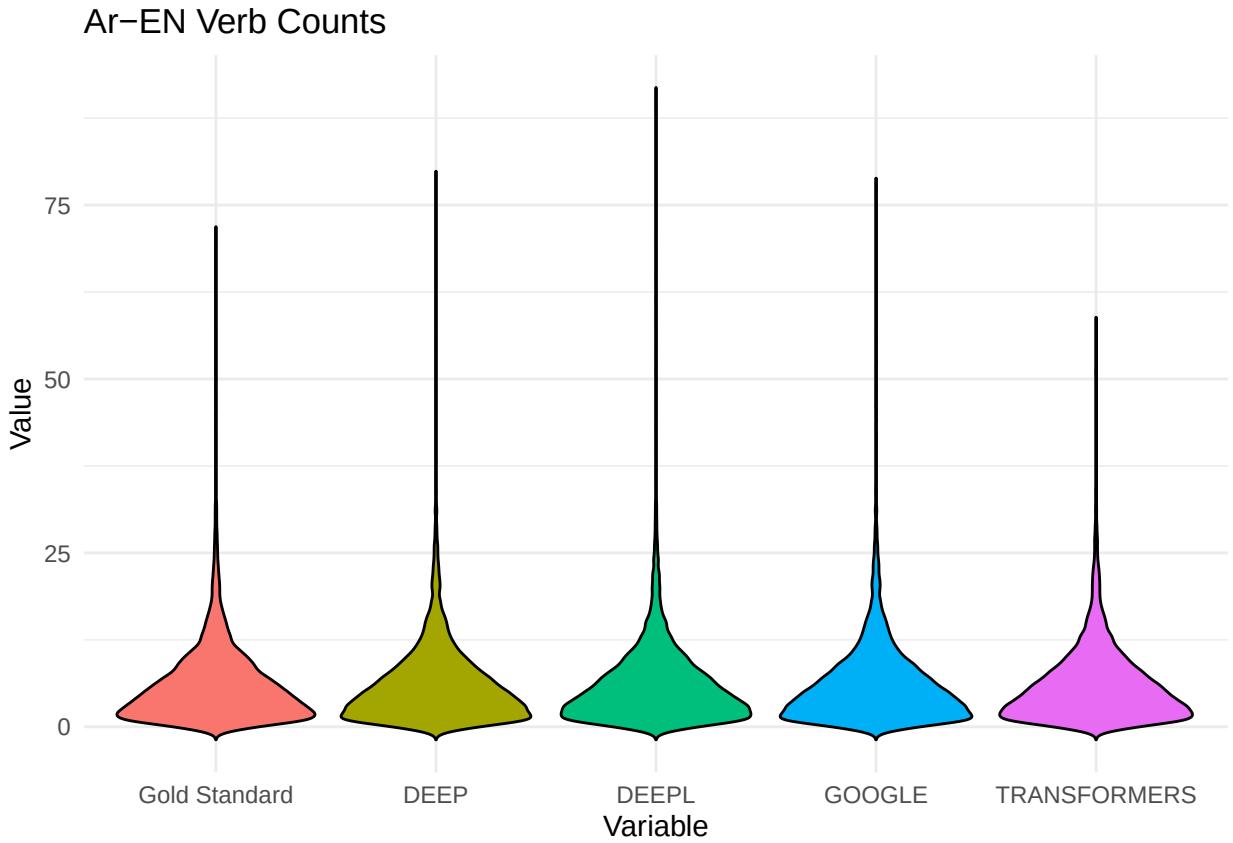
#Transform to long table and select variables # AR-EN Verb counts

```r
df_ar_en_verb_long <- pivot_longer(en_count_data,
                                   cols = c(en_verb_counts,
                                            ar_en_DEEP_verb_counts,
                                            ar_en_DEEPL_verb_counts,
                                            ar_en_GOOGLE_verb_counts,
                                            ar_en_TRANSFORMERS_verb_counts),
                                   names_to = "Variable",
                                   values_to = "Value")
```

Create the violin plot

AR-EN Verb counts

```r
ggplot(df_es_en_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "Ar-EN Verb Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



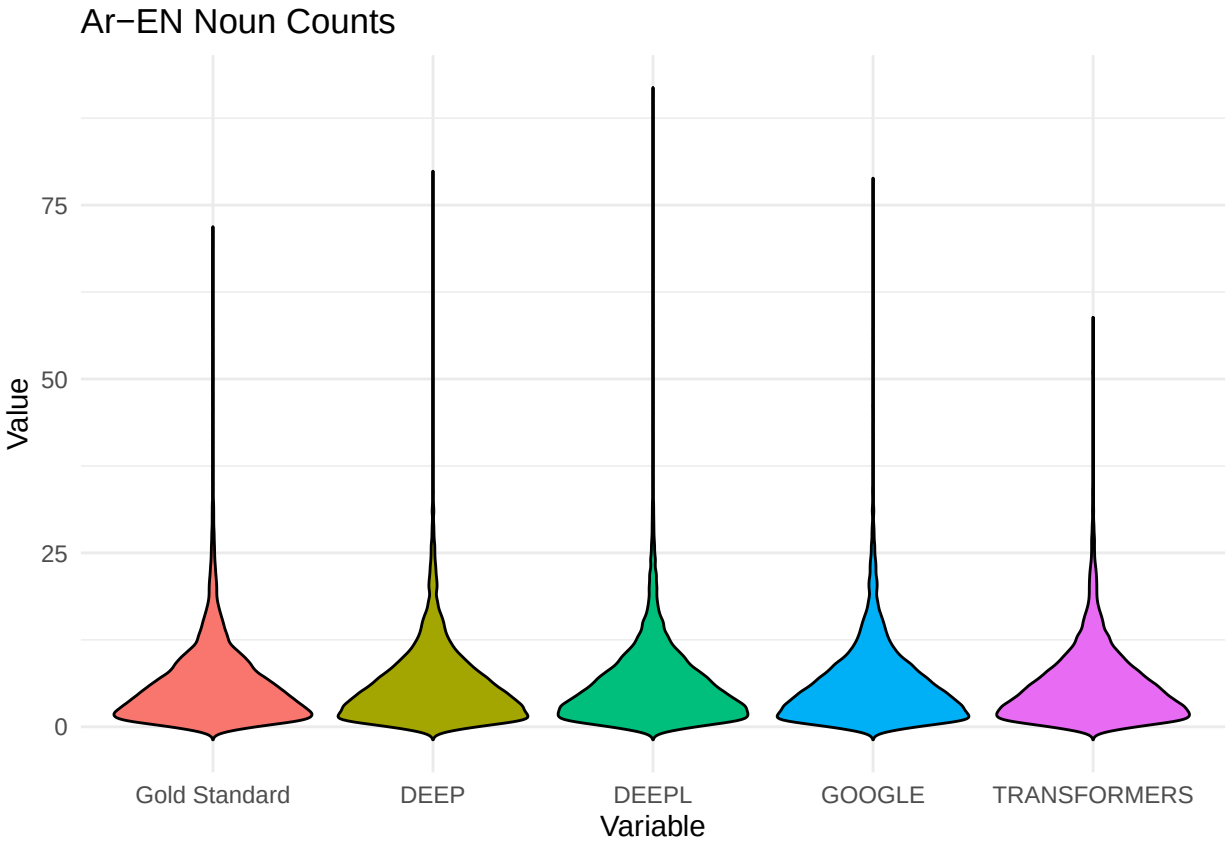
#Transform to long table and select variables # AR-EN Noun counts

```r
df_ar_en_verb_long <- pivot_longer(en_count_data,
                                   cols = c(en_noun_counts,
                                            ar_en_DEEP_noun_counts,
                                            ar_en_DEEPL_noun_counts,
                                            ar_en_GOOGLE_noun_counts,
                                            ar_en_TRANSFORMERS_noun_counts),
                                   names_to = "Variable",
                                   values_to = "Value")
```

Create the violin plot

AR-EN Noun counts

```r
ggplot(df_es_en_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "Ar-EN Noun Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



End Count Analyses for English

Count Analyses for Spanish

#Import the data

```
es_count_data <- read.csv("/Users/Dagmar Heintze/Downloads/es_counts.csv")
View(es_count_data)
```

#Transform to long table and select variables

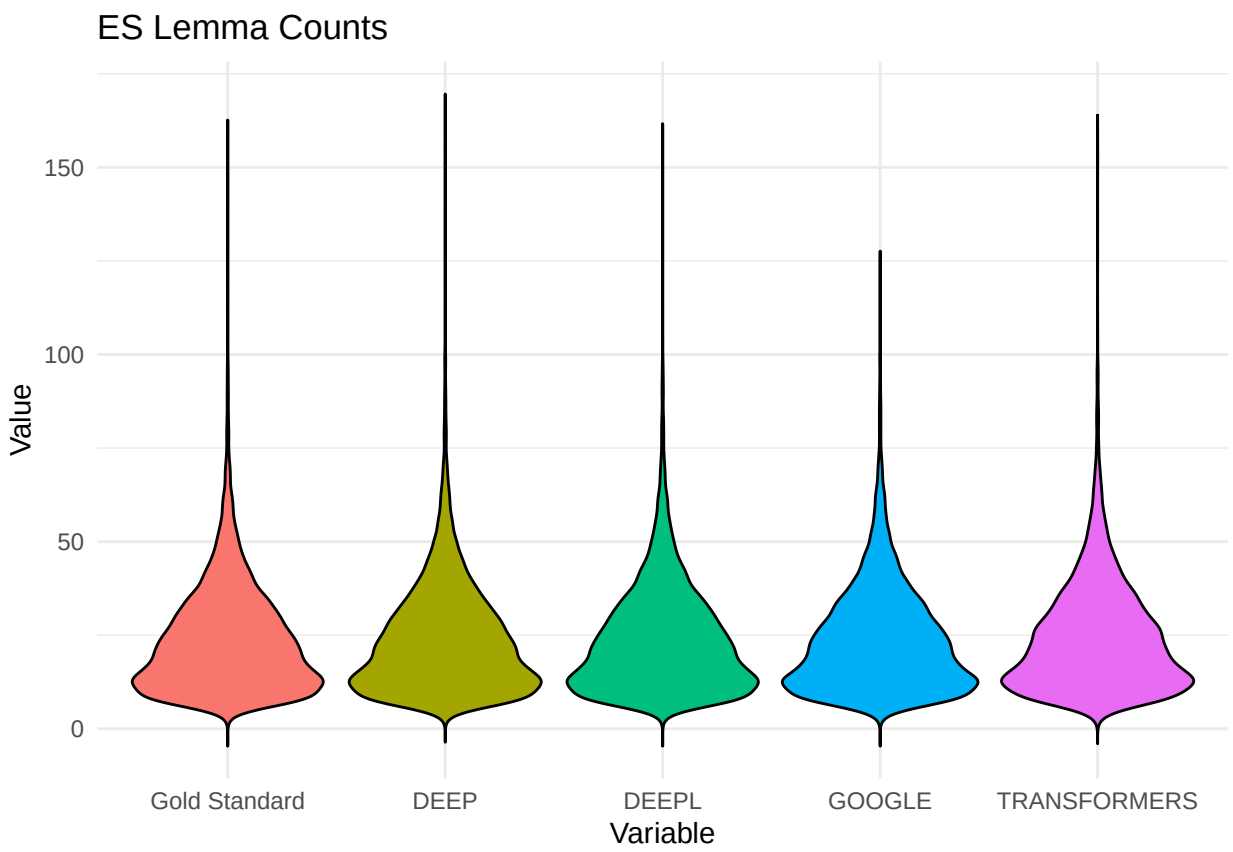
ES Lemma counts

```
df_es_lemma_long <- pivot_longer(es_count_data,
                                 cols = c(es_lemma_counts,
                                          en_es_DEEP_lemma_counts,
                                          en_es_DEEPL_lemma_counts,
                                          en_es_GOOGLE_lemma_counts,
                                          en_es_TRANSFORMERS_lemma_counts),
                                 names_to = "Variable",
                                 values_to = "Value")
```

Create the violin plot

ES Lemma counts

```
ggplot(df_es_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES Lemma Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



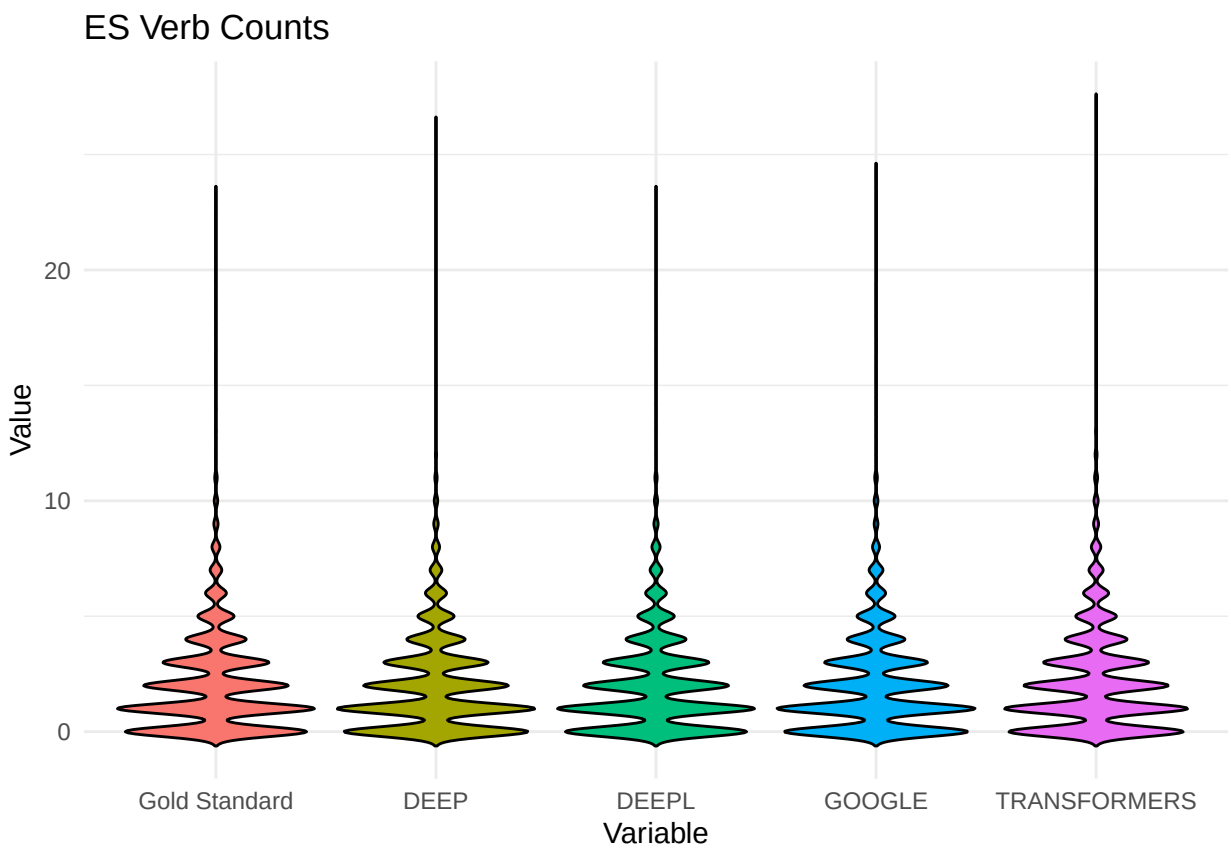
#Transform to long table and select variables # ES Verb counts

```
df_es_verb_long <- pivot_longer(es_count_data,
                                cols = c(es_verb_counts,
                                         en_es_DEEP_verb_counts,
                                         en_es_DEEPL_verb_counts,
                                         en_es_GOOGLE_verb_counts,
                                         en_es_TRANSFORMERS_verb_counts),
                                names_to = "Variable",
                                values_to = "Value")
```

Create the violin plot

ES Verb counts

```
ggplot(df_es_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES Verb Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



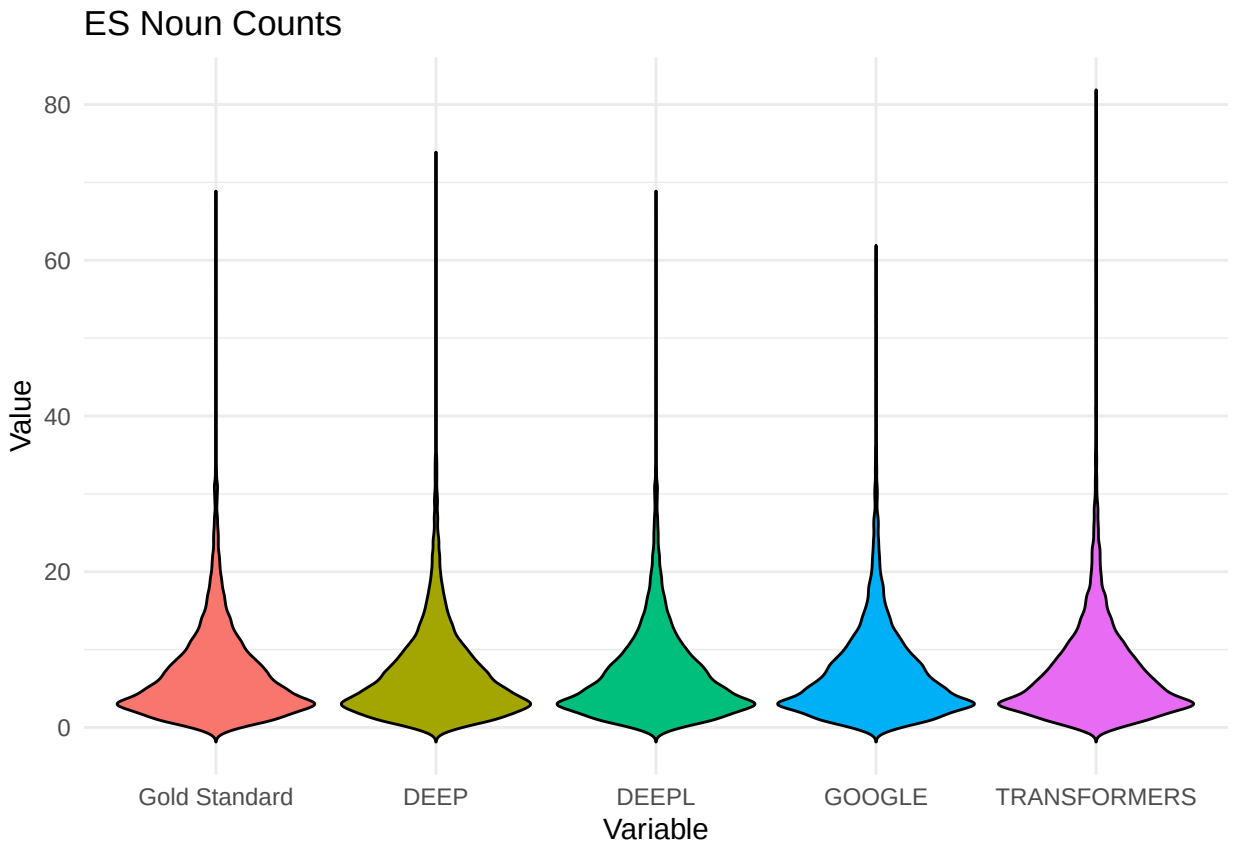
#Transform to long table and select variables # ES Noun counts

```
df_es_noun_long <- pivot_longer(es_count_data,
                                cols = c(es_noun_counts,
                                         en_es_DEEP_noun_counts,
                                         en_es_DEEPL_noun_counts,
                                         en_es_GOOGLE_noun_counts,
                                         en_es_TRANSFORMERS_noun_counts),
                                names_to = "Variable",
                                values_to = "Value")
```

Create the violin plot

ES Noun counts

```
ggplot(df_es_noun_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES Noun Counts", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



End Count Analysis Spanish

Begin Count Difference Analysis from Arabic

#Import the data

```
ar_diff_data <- read.csv("/Users/Dagmar Heintze/Downloads/from_ar_counts_difference.csv")
View(ar_diff_data)
```

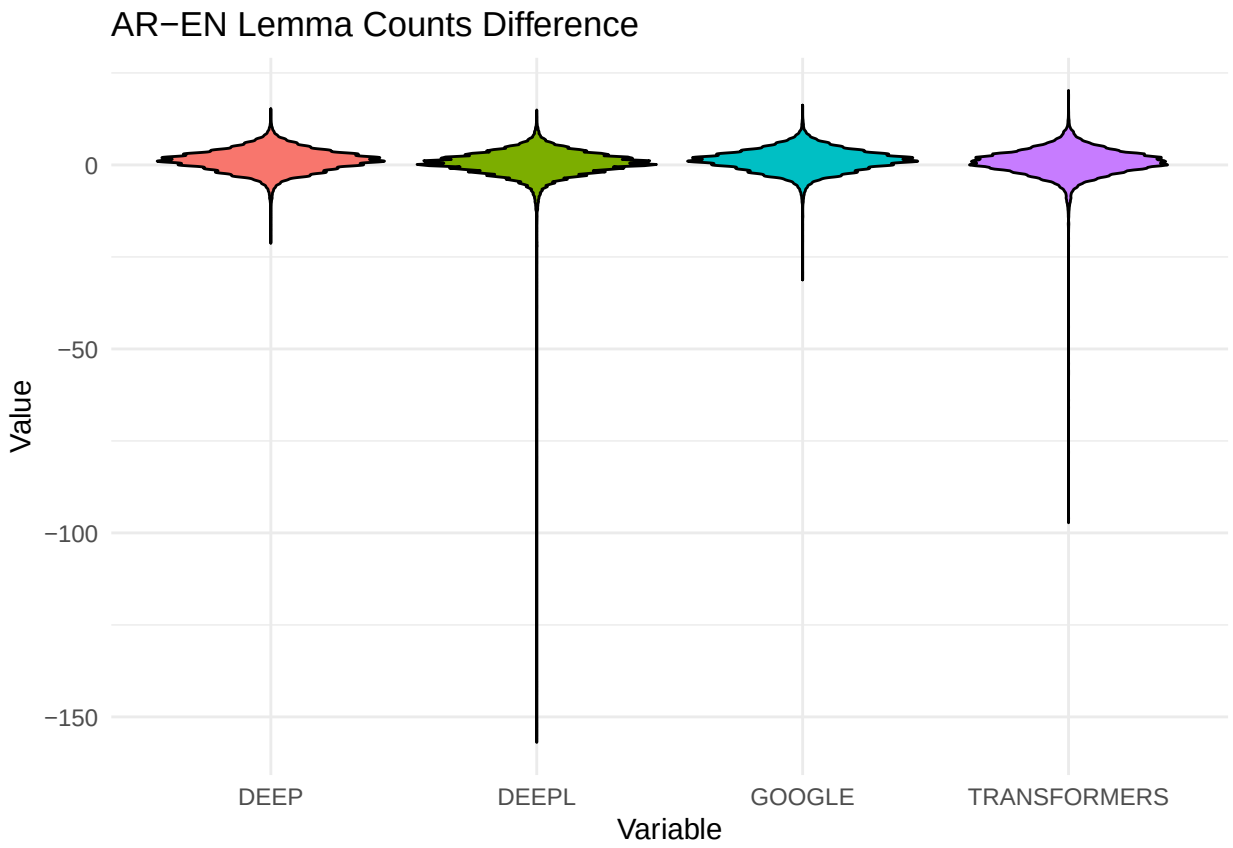
#Transform to long table and select variables # AR-EN Lemma difference

```
df_ar_diff_lemma_long <- pivot_longer(ar_diff_data,
                                      cols = c(ar_en_DEEP_difference_lemma_counts,
                                               ar_en_DEEPL_difference_lemma_counts,
                                               ar_en_GOOGLE_difference_lemma_counts,
                                             ar_en_TRANSFORMERS_difference_lemma_counts),
                                      names_to = "Variable",
                                      values_to = "Value")
```

Create the violin plot

AR-EN Lemma counts differnce

```r
ggplot(df_ar_diff_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR-EN Lemma Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



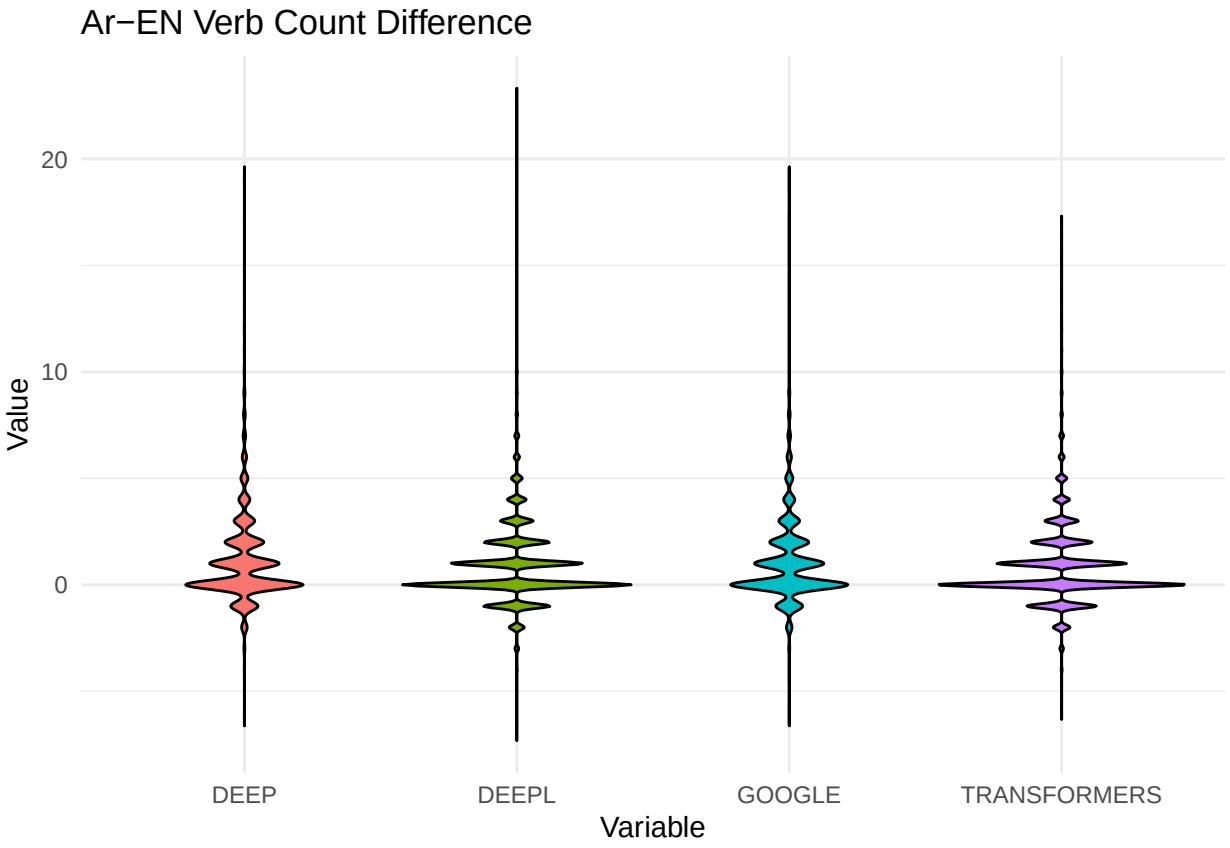
#Transform to long table and select variables # AR-EN Verb counts Difference

```r
df_ar_diff_verb_long <- pivot_longer(ar_diff_data,
                                     cols = c(
                                              ar_en_DEEP_difference_verb_counts,
                                              ar_en_DEEPL_difference_verb_counts,
                                              ar_en_GOOGLE_difference_verb_counts,
                                              ar_en_TRANSFORMERS_difference_verb_counts),
                                     names_to = "Variable",
                                     values_to = "Value")
```

Create the violin plot

AR-EN Verb counts Difference

```
ggplot(df_ar_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "Ar-EN Verb Count Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



#Transform to long table and select variables

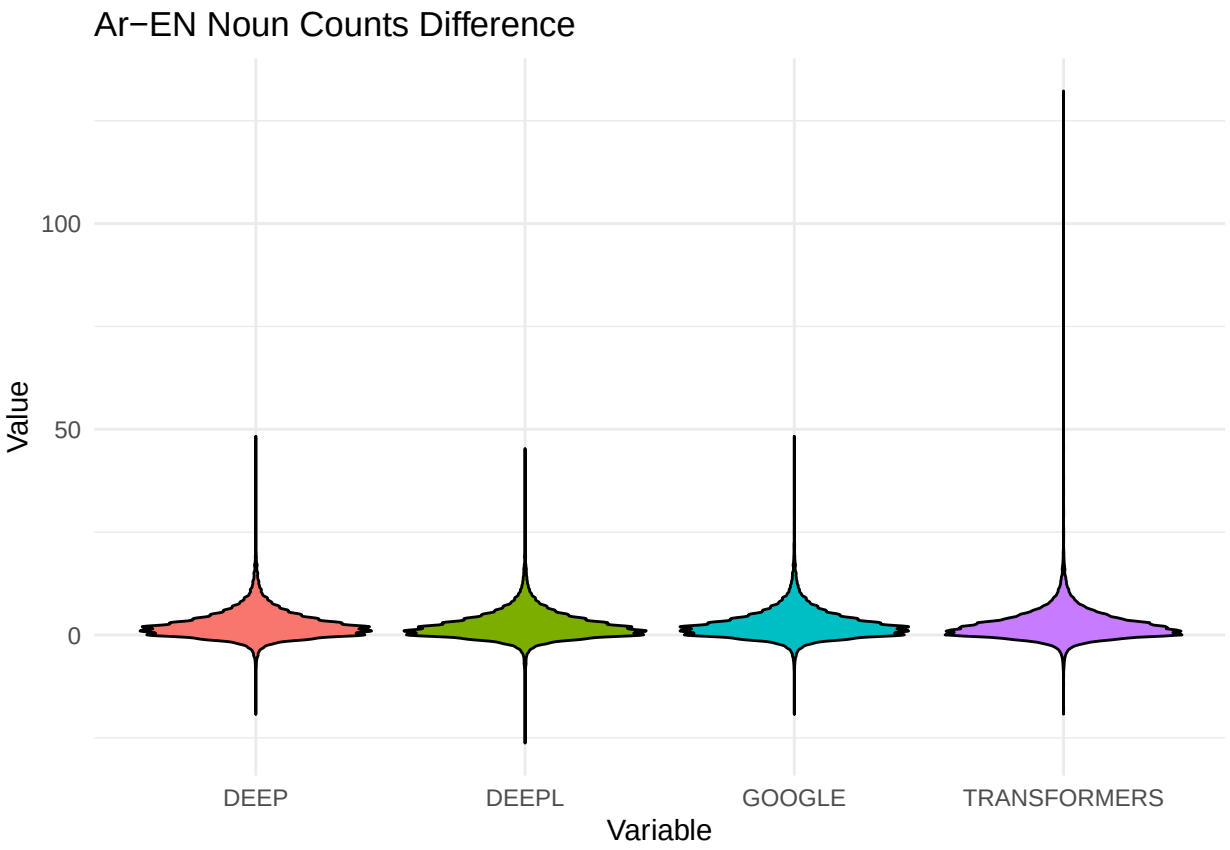
AR-EN Noun counts Difference

```
df_ar_diff_verb_long <- pivot_longer(ar_diff_data,
                                     cols = c(
                                              ar_en_DEEP_difference_noun_counts,
                                              ar_en_DEEPL_difference_noun_counts,
                                              ar_en_GOOGLE_difference_noun_counts,
                                              ar_en_TRANSFORMERS_difference_noun_counts),
                                     names_to = "Variable",
                                     values_to = "Value")
```

Create the violin plot

AR-EN Noun counts Difference

```
ggplot(df_ar_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "Ar-EN Noun Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c( "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



End Count Difference Analysis Arabic

Begin Count Difference Analysis English

#Import the data

```
en_diff_data <- read.csv("/Users/Dagmar Heintze/Downloads/en_counts_difference.csv")
View(en_diff_data)
```

#Transform to long table and select variables # ES-EN Lemma counts Difference

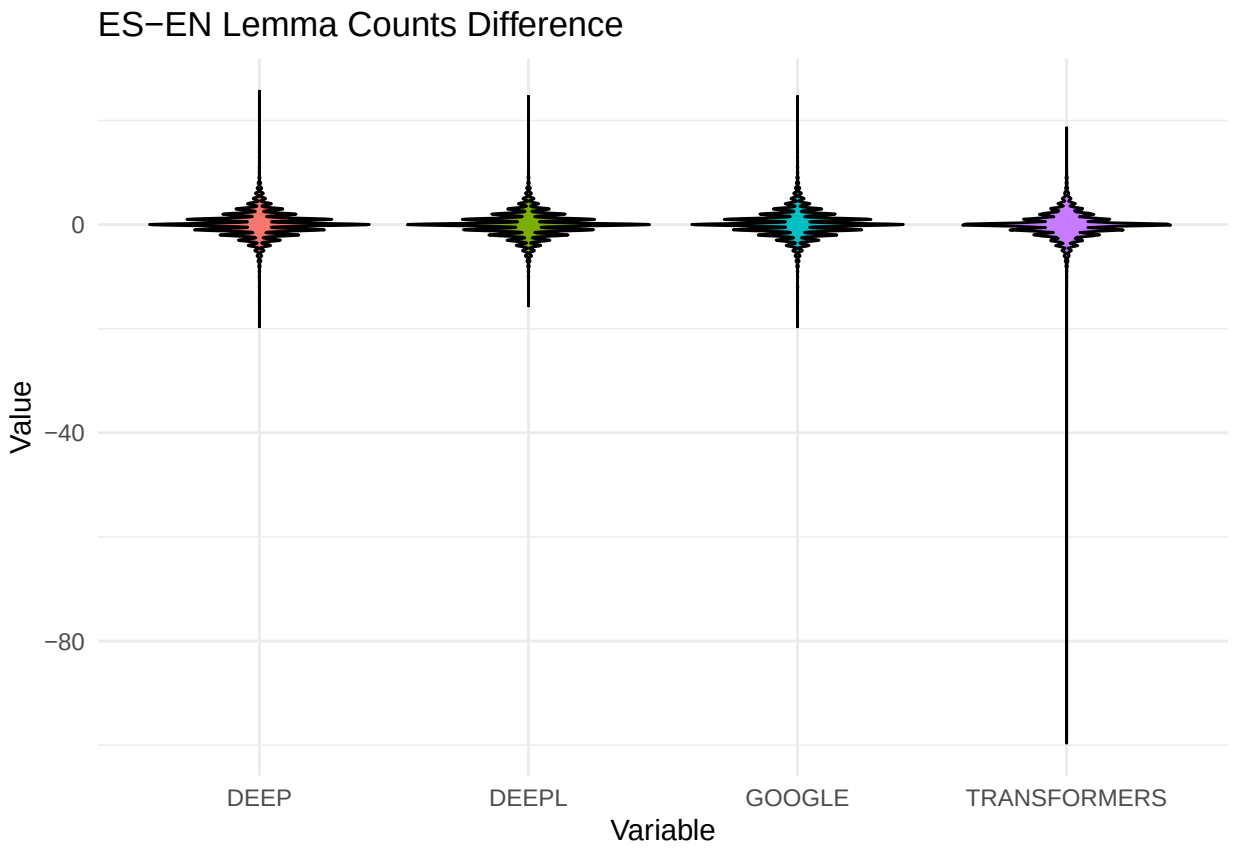
```
df_es_en_diff_lemma_long <- pivot_longer(en_diff_data,
                                cols = c(
                                        es_en_DEEP_difference_lemma_counts,
```

```
                                    es_en_DEEPL_difference_lemma_counts,
                                    es_en_GOOGLE_difference_lemma_counts,
                                    es_en_TRANSFORMERS_difference_lemma_counts),
                           names_to = "Variable",
                           values_to = "Value")
```

Create the violin plot

ES-EN Lemma counts Difference

```
ggplot(df_es_en_diff_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Lemma Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



#Transform to long table and select variables # ES-EN Verb counts Difference

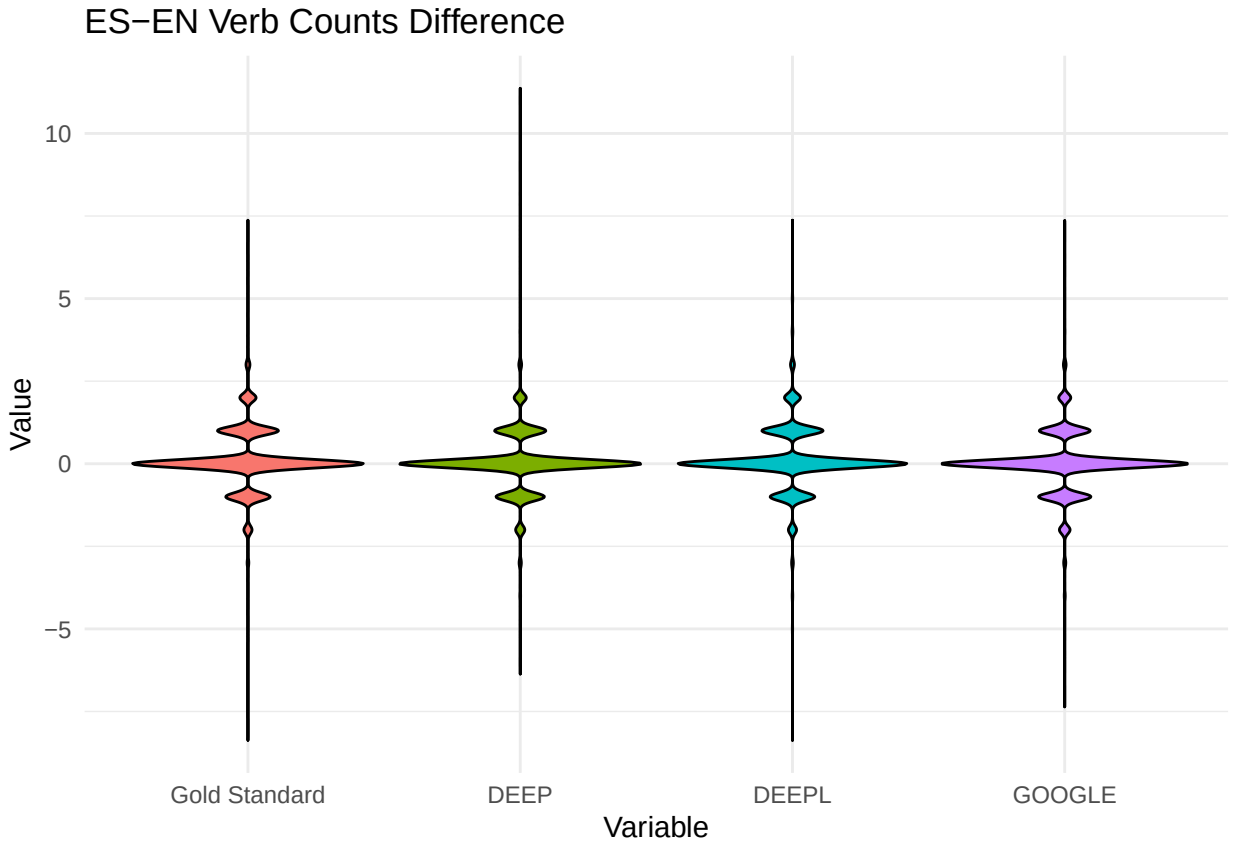
```
df_es_en_diff_verb_long <- pivot_longer(en_diff_data,
                                cols = c(
                                        es_en_DEEP_difference_verb_counts,
                                        es_en_DEEPL_difference_verb_counts,
```

```
                                    es_en_GOOGLE_difference_verb_counts,
                                    es_en_TRANSFORMERS_difference_verb_counts),
                           names_to = "Variable",
                           values_to = "Value")
```

Create the violin plot

ES-EN Verb counts difference

```
ggplot(df_es_en_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Verb Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



#Transform to long table and select variables # ES-EN Noun counts difference

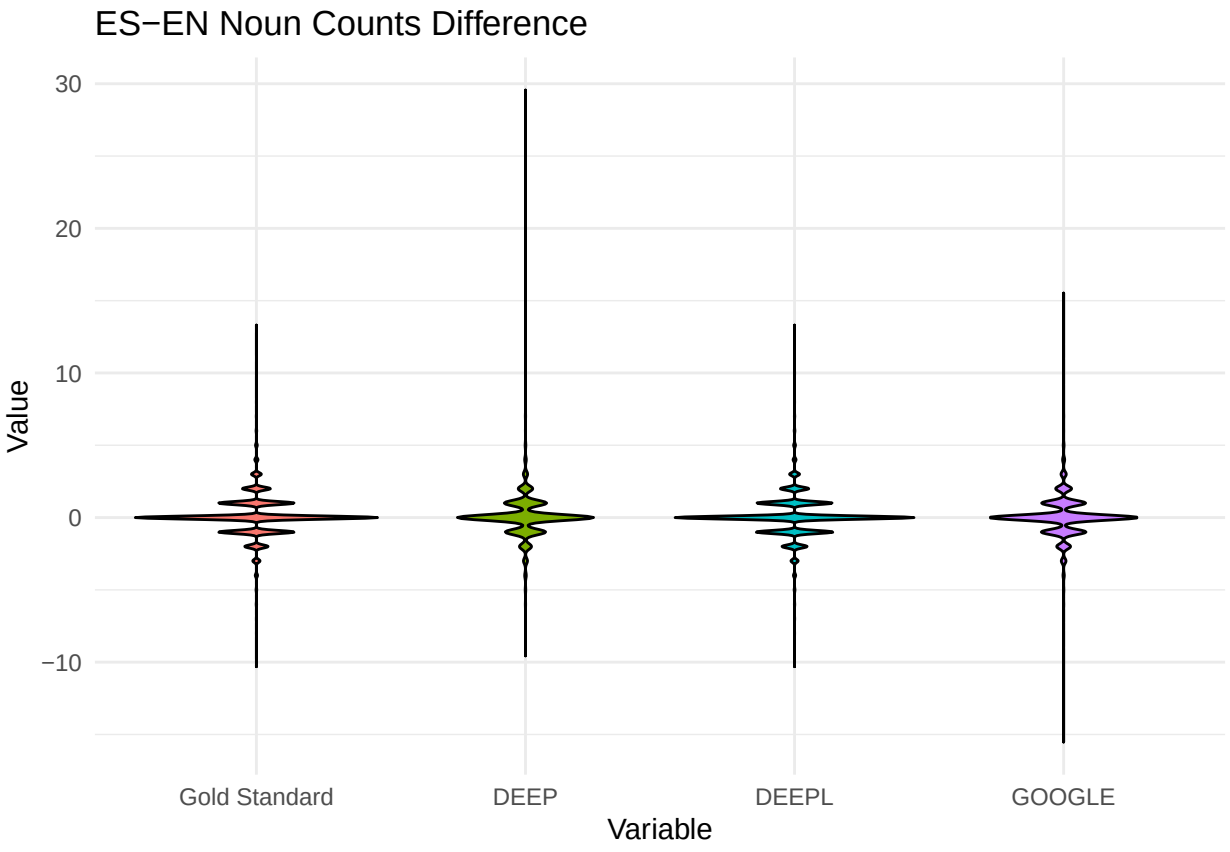
```
df_es_en_diff_noun_long <- pivot_longer(en_diff_data,
                           cols = c(
                                    es_en_DEEP_difference_noun_counts,
                                    es_en_DEEPL_difference_noun_counts,
                                    es_en_GOOGLE_difference_noun_counts,
```

```
                                        es_en_TRANSFORMERS_difference_noun_counts),
                                names_to = "Variable",
                                values_to = "Value")
```

Create the violin plot

ES-EN Noun counts difference

```
ggplot(df_es_en_diff_noun_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "ES-EN Noun Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



End ES-EN Count Difference Analysis

Beginning AR-EN Count Difference Analysis

#Transform to long table and select variables # AR-EN Lemma counts Difference

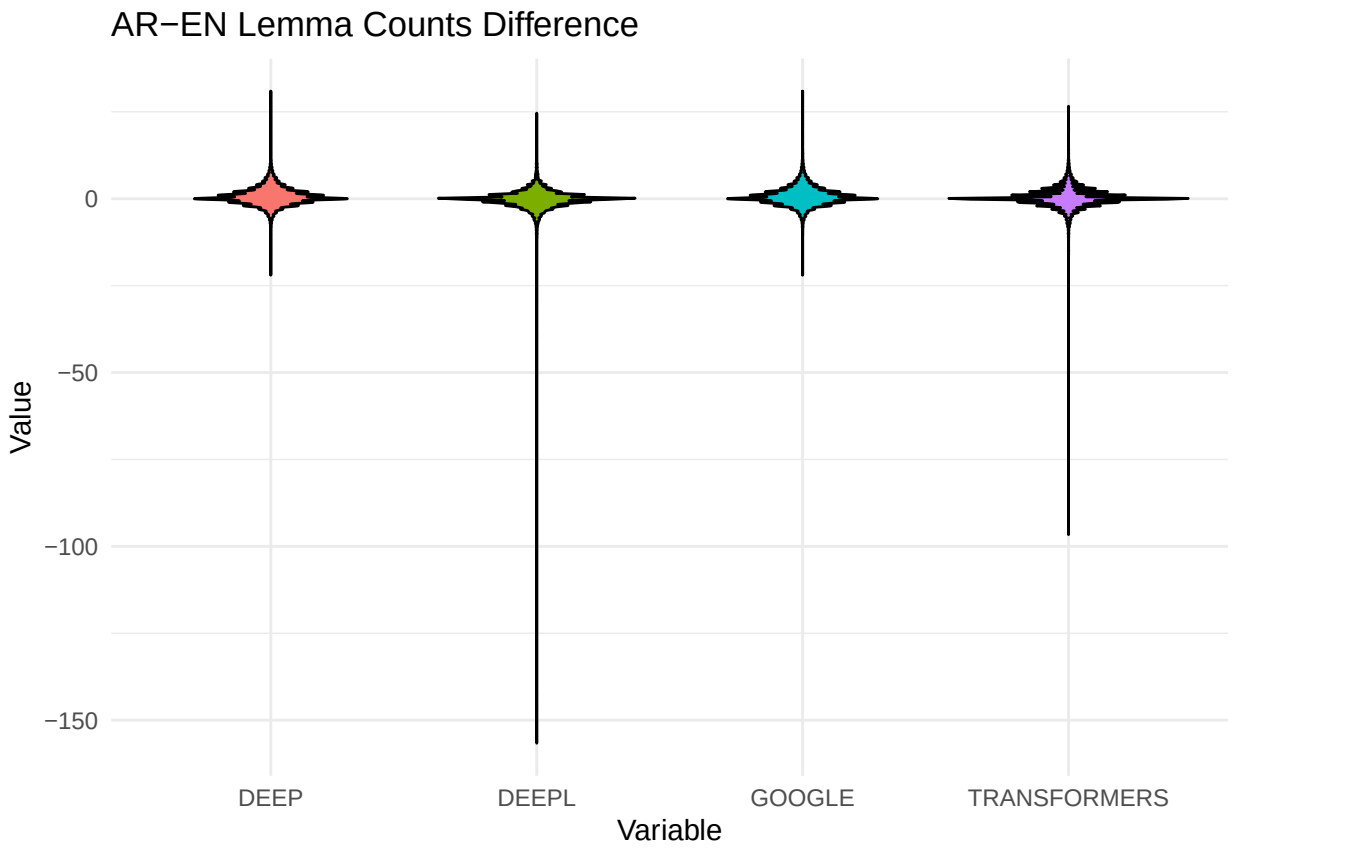
```
df_ar_en_diff_lemma_long <- pivot_longer(en_diff_data,
                                cols = c(
                                        ar_en_DEEP_difference_lemma_counts,
```

```
                                    ar_en_DEEPL_difference_lemma_counts,
                                    ar_en_GOOGLE_difference_lemma_counts,
                                    ar_en_TRANSFORMERS_difference_lemma_counts),
                            names_to = "Variable",
                            values_to = "Value")
```

Create the violin plot

AR-EN Lemma counts Difference

```
ggplot(df_ar_en_diff_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR-EN Lemma Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



#Transform to long table and select variables # AR-EN Verb counts difference

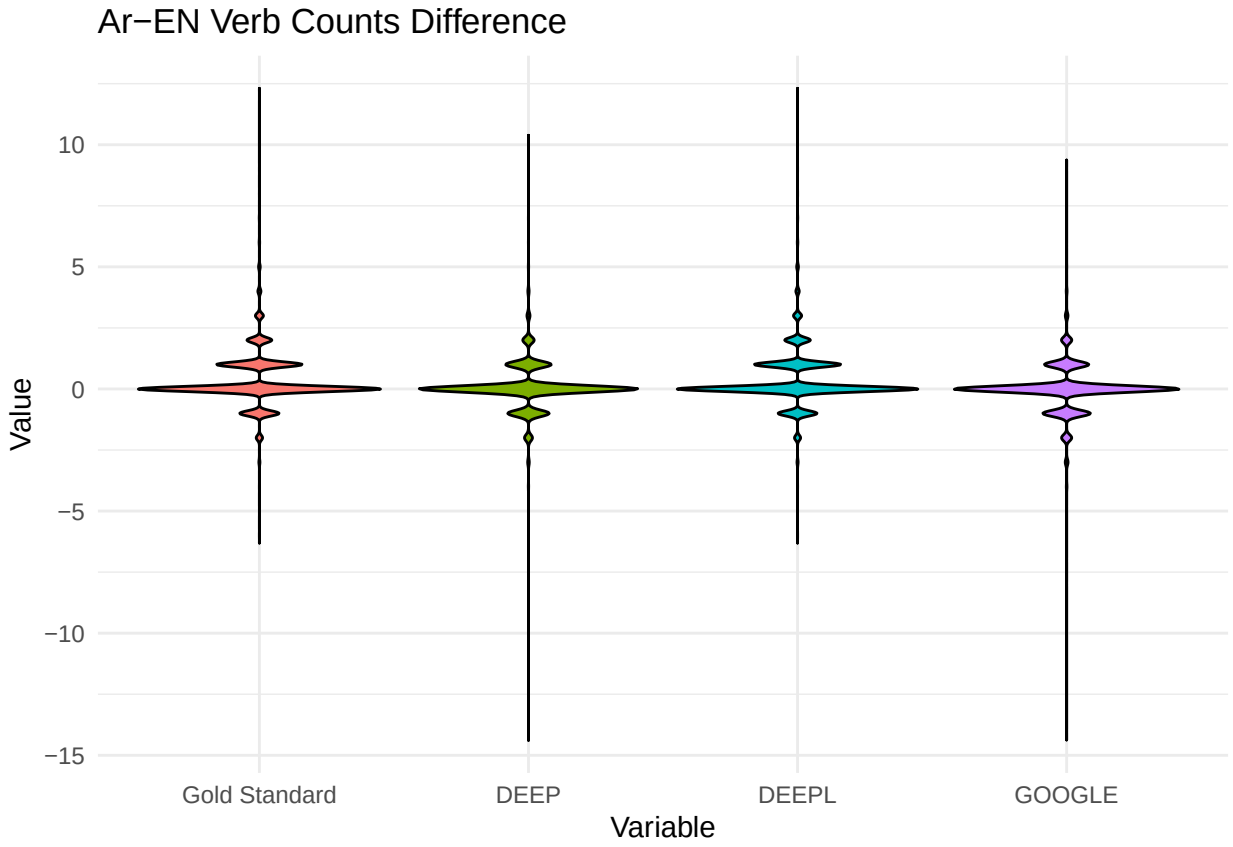
```
df_ar_en_diff_verb_long <- pivot_longer(en_diff_data,
                                cols = c(
                                        ar_en_DEEP_difference_verb_counts,
                                        ar_en_DEEPL_difference_verb_counts,
```

```
                                    ar_en_GOOGLE_difference_verb_counts,
                                    ar_en_TRANSFORMERS_difference_verb_counts),
                           names_to = "Variable",
                           values_to = "Value")
```

Create the violin plot

AR-EN Verb counts difference

```
ggplot(df_ar_en_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "Ar-EN Verb Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



#Transform to long table and select variables # AR-EN Noun counts Difference

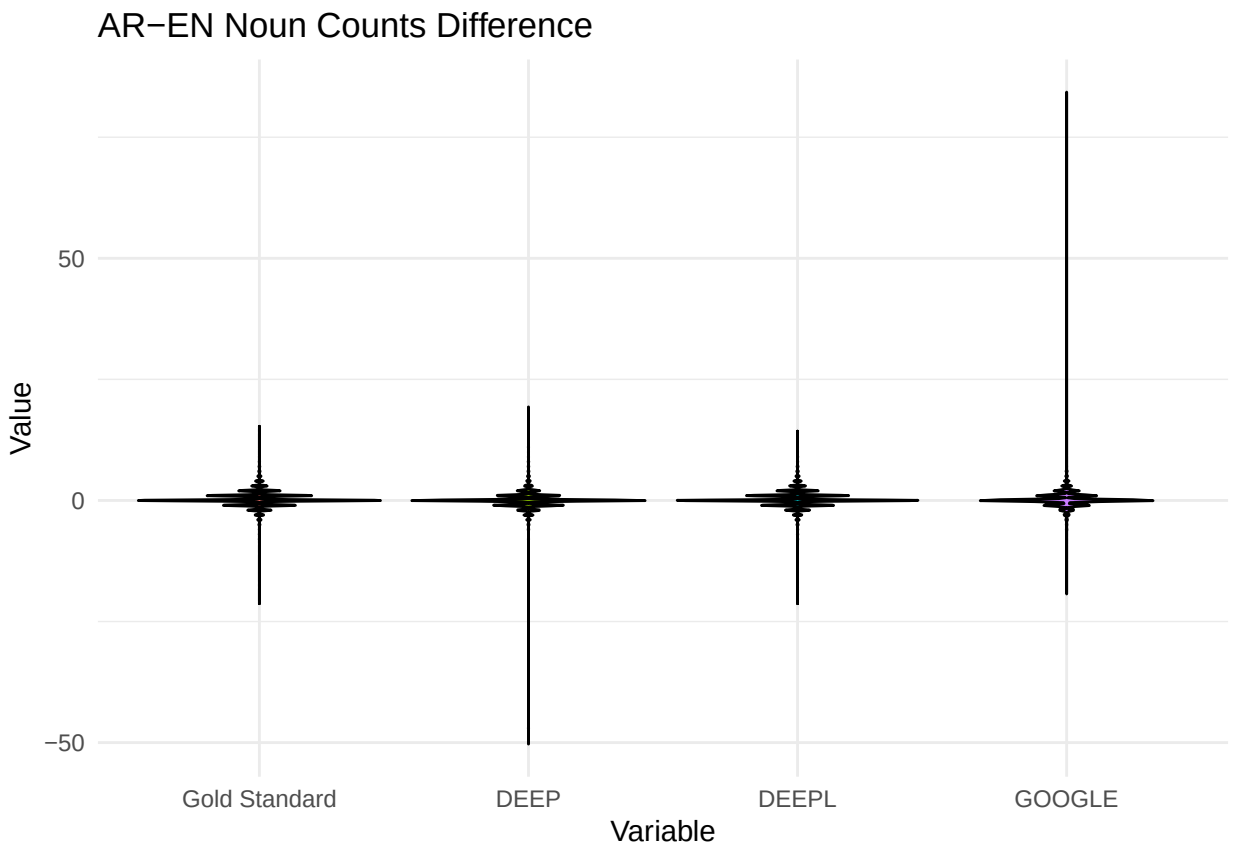
```
df_ar_en_diff_noun_long <- pivot_longer(en_diff_data,
                           cols = c(
                                    ar_en_DEEP_difference_noun_counts,
                                    ar_en_DEEPL_difference_noun_counts,
                                    ar_en_GOOGLE_difference_noun_counts,
```

```
                                    ar_en_TRANSFORMERS_difference_noun_counts),
                        names_to = "Variable",
                        values_to = "Value")
```

Create the violin plot

AR-EN Noun counts Difference

```
ggplot(df_ar_en_diff_noun_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "AR-EN Noun Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("Gold Standard", "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-
```



End AR-EN Count Difference Analysis

Begin Count Difference Analysis from Spanish

#Import the data

```
es_diff_data <- read.csv("/Users/Dagmar Heintze/Downloads/es_counts_difference.csv")
View(es_diff_data)
```

#Transform to long table and select variables

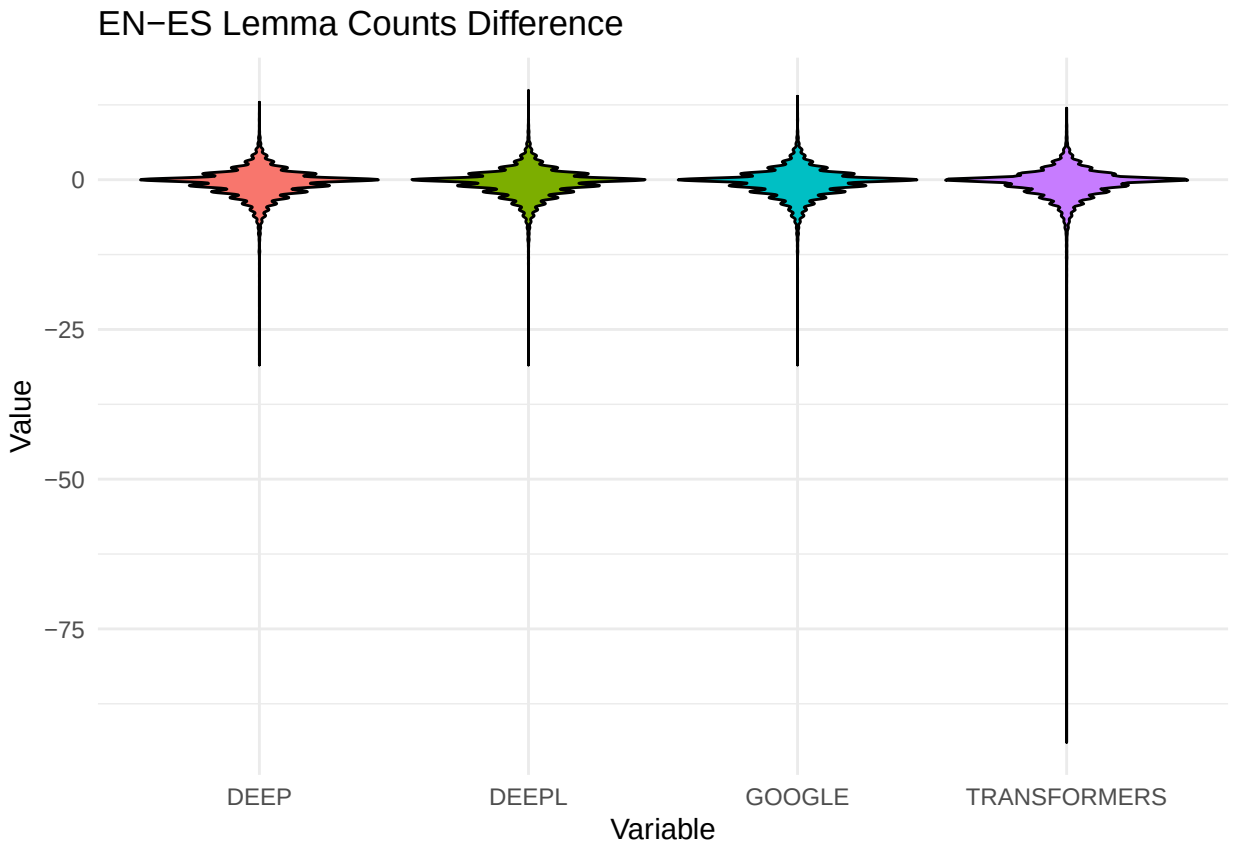
EN-ES Lemma count Difference

```
df_es_diff_lemma_long <- pivot_longer(es_diff_data,
                                      cols = c(en_es_DEEP_difference_lemma_counts,
                                               en_es_DEEPL_difference_lemma_counts,
                                               en_es_GOOGLE_difference_lemma_counts,
                                             en_es_TRANSFORMERS_difference_lemma_counts),
                                      names_to = "Variable",
                                      values_to = "Value")
```

Create the violin plot

EN-ES Lemma counts differnce

```
ggplot(df_es_diff_lemma_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "EN-ES Lemma Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



#Transform to long table and select variables

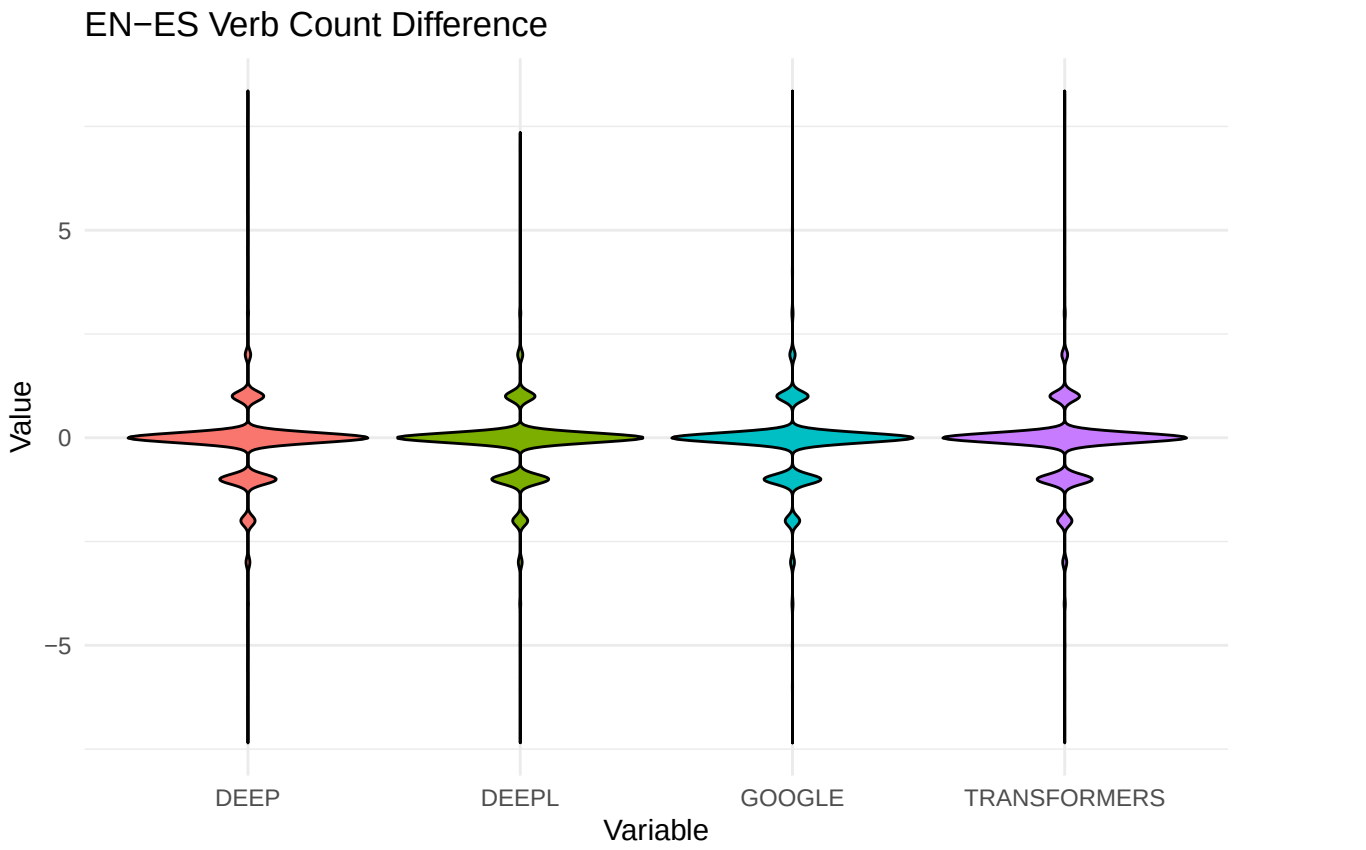
EN-ES Verb counts Difference

```
df_es_diff_verb_long <- pivot_longer(es_diff_data,
                                     cols = c(
                                              en_es_DEEP_difference_verb_counts,
                                              en_es_DEEPL_difference_verb_counts,
                                              en_es_GOOGLE_difference_verb_counts,
                                              en_es_TRANSFORMERS_difference_verb_counts),
                                     names_to = "Variable",
                                     values_to = "Value")
```

Create the violin plot

EN-ES Verb counts Difference

```
ggplot(df_es_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "EN-ES Verb Count Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c("DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



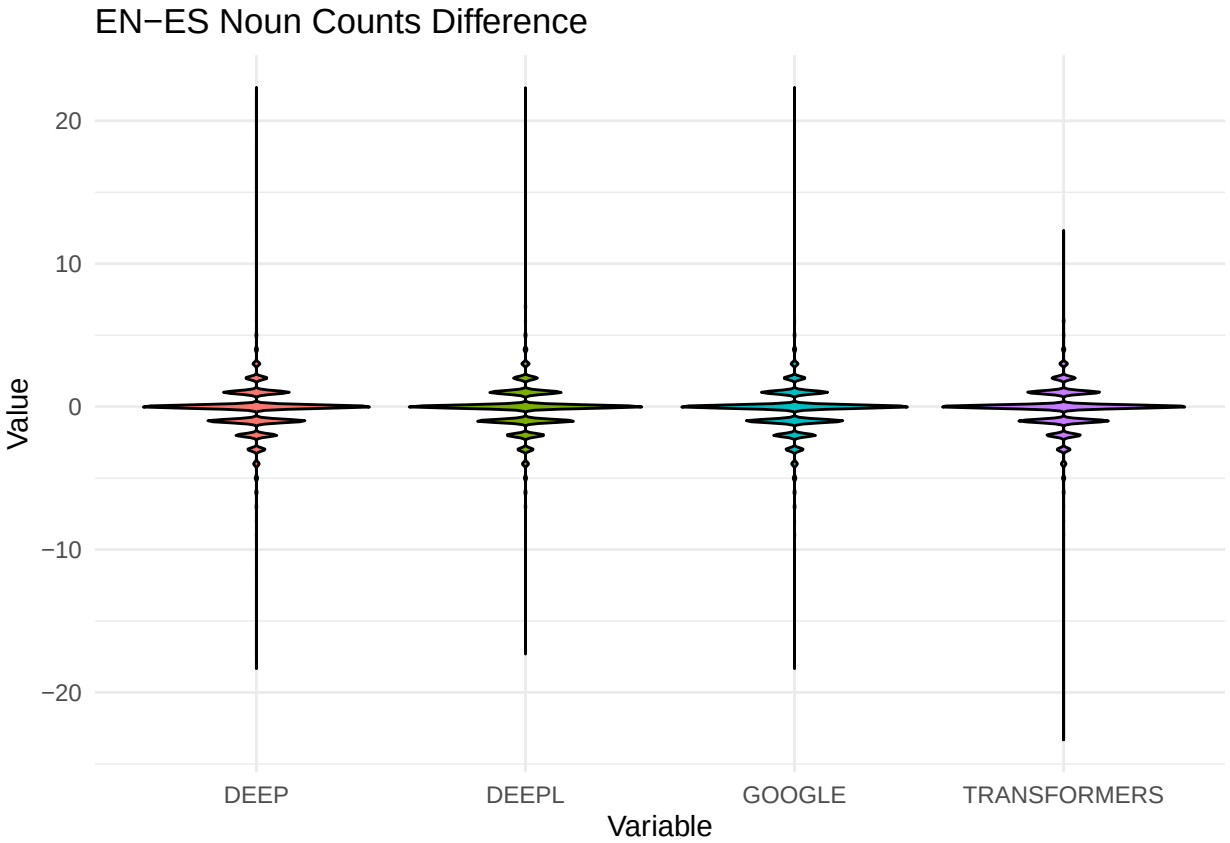
#Transform to long table and select variables # EN-ES Noun counts Difference

```r
df_es_diff_verb_long <- pivot_longer(es_diff_data,
                                     cols = c(
                                              en_es_DEEP_difference_noun_counts,
                                              en_es_DEEPL_difference_noun_counts,
                                              en_es_GOOGLE_difference_noun_counts,
                                              en_es_TRANSFORMERS_difference_noun_counts),
                                     names_to = "Variable",
                                     values_to = "Value")
```

Create the violin plot

EN-ES Noun counts Difference

```r
ggplot(df_es_diff_verb_long, aes(x = Variable, y = Value, fill = Variable)) +
  geom_violin(trim = FALSE, color = "black") +
  theme_minimal() +
  labs(title = "EN-ES Noun Counts Difference", y = "Value") +
  theme(legend.position = "none") +
  scale_x_discrete(labels = c( "DEEP", "DEEPL", "GOOGLE", "TRANSFORMERS"))  # Custom x-axis labels
```



End Count Difference Analysis Spanish