5291 Final Project: Country-Level Influence on Engagement Rate and Likes using Hierarchical Linear Models

Francis Zhang

November 20, 2024

Motivation

From the EDA results (Section 10), we observe that the United States has the highest number of influencers compared to other countries, as shown in the chart. This raises an important question: does the dominance in the number of influencers translate into higher average Engagement_Rate and Likes for influencers in the United States compared to other countries? Additionally, it is worth exploring whether countries with fewer influencers demonstrate unique patterns in these metrics. By analyzing Engagement_Rate and Likes across countries, we aim to uncover potential country-level effects and evaluate whether geographic factors significantly influence these key performance metrics. This study can provide actionable insights into the regional dynamics of influencer performance.

Steps for Analysis

1.Data Preparation

1.Ensure the data has the necessary variables:

- Engagement_Rate: Outcome variable 1.
- Likes: Outcome variable 2.
- Country: Grouping variable for hierarchical modeling.
- Followers and Influence_Score: Predictors.

2.Convert the Country variable to a factor for hierarchical modeling.

```
# Load the data (replace 'file_path' with the actual file path)
data <- read.csv("/Users/franciszhang/Downloads/top_insta_influencers_data.csv")

# Load necessary libraries
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(lmerTest)  # For p-values in mixed models
```

```
##
## Attaching package: 'lmerTest'
```

```
## The following object is masked from 'package:lme4':
##
##     lmer
```

```
## The following object is masked from 'package:stats':
##
```

```
##     step
```

```
# Function to convert 'k' and 'm' abbreviations to numeric
convert_to_numeric <- function(x) {
  as.numeric(ifelse(
    grepl("k", x), as.numeric(gsub("k", "", x)) * 1e3,
    ifelse(grepl("m", x), as.numeric(gsub("m", "", x)) * 1e6, x)
  ))
}

# Apply to relevant columns
data$followers <- convert_to_numeric(data$followers)
data$avg_likes <- convert_to_numeric(data$avg_likes)
```

```
## Warning in ifelse(grepl("k", x), as.numeric(gsub("k", "", x)) * 1000,
## ifelse(grepl("m", : NAs introduced by coercion
```

```
## Warning in ifelse(grepl("m", x), as.numeric(gsub("m", "", x)) * 1e+06, x): NAs
## introduced by coercion
```

```
data$new_post_avg_like <- convert_to_numeric(data$new_post_avg_like)
```

```
## Warning in ifelse(grepl("k", x), as.numeric(gsub("k", "", x)) * 1000,
## ifelse(grepl("m", : NAs introduced by coercion
## Warning in ifelse(grepl("k", x), as.numeric(gsub("k", "", x)) * 1000,
## ifelse(grepl("m", : NAs introduced by coercion
```

```
data$total_likes <- convert_to_numeric(data$total_likes)
```

```
## Warning in ifelse(grepl("m", x), as.numeric(gsub("m", "", x)) * 1e+06, x): NAs
## introduced by coercion
```

```
## Warning in convert_to_numeric(data$total_likes): NAs introduced by coercion
```

```
data$posts <- convert_to_numeric(data$posts)

# Remove the percentage symbol and convert to numeric
data$X60_day_eng_rate <- as.numeric(gsub("%", "", data$X60_day_eng_rate)) / 100

# Now, Engagement_Rate is a numeric value (e.g., 1.39 instead of "1.39%")

# Convert country to a factor if not already
data$country <- as.factor(data$country)

# Scale predictors for better model convergence
data$Followers_Scaled <- scale(data$followers)
data$Influence_Score_Scaled <- scale(data$influence_score)

head(data)
```

```
##   rank  channel_info influence_score posts followers avg_likes X60_day_eng_rate
## 1    1     cristiano              92  3300 475800000   8700000           0.0139
## 2    2   kyliejenner              91  6900 366200000   8300000           0.0162
## 3    3      leomessi              90   890 357300000   6800000           0.0124
## 4    4   selenagomez              93  1800 342700000   6200000           0.0097
## 5    5       therock              91  6800 334100000   1900000           0.0020
## 6    6 kimkardashian              91  5600 329200000   3500000           0.0088
##   new_post_avg_like total_likes       country Followers_Scaled
```

```
## 1           6500000          NA         Spain         5.406504
## 2           5900000          NA United States         3.919137
## 3           4400000          NA                       3.798356
## 4           3300000          NA United States         3.600222
## 5            665300          NA United States         3.483512
## 6           2900000          NA United States         3.417015
##   Influence_Score_Scaled
## 1              1.1466341
## 2              1.0339981
## 3              0.9213622
## 4              1.2592701
## 5              1.0339981
## 6              1.0339981
```

2.Model Construction

2.1.Null Model: • Test whether there is significant variation in Engagement_Rate across countries.

```
# Null model for Engagement Rate
null_model_eng <- lmer(X60_day_eng_rate ~ 1 + (1 | country), data = data)
summary(null_model_eng)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: X60_day_eng_rate ~ 1 + (1 | country)
##    Data: data
##
## REML criterion at convergence: -785.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.8593 -0.4015 -0.2419 -0.0122  7.2847
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  country  (Intercept) 5.16e-05 0.007183
##  Residual             1.05e-03 0.032404
## Number of obs: 199, groups:  country, 26
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  0.016410   0.003483 18.266338   4.712 0.000168 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Null model for Average Likes
null_model_avglikes <- lmer(avg_likes ~ 1 + (1 | country), data = data)
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 1.7e-10
```

```
summary(null_model_avglikes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: avg_likes ~ 1 + (1 | country)
##    Data: data
```

```
##
## REML criterion at convergence: 6377.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -0.9891 -0.5453 -0.3045  0.1309  6.1428
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  country  (Intercept) 2.265e+11  475962
##  Residual             4.609e+12 2146855
## Number of obs: 200, groups:  country, 26
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1.511e+06  2.307e+05 1.536e+01    6.55 8.18e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Null model for Total Likes
null_model_totallikes <- lmer(total_likes ~ 1 + (1 | country), data = data)
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 2.4e-15
```

```
summary(null_model_totallikes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: total_likes ~ 1 + (1 | country)
##    Data: data
##
## REML criterion at convergence: 2048.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.01573 -0.60375  0.05277  0.76843  1.54399
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  country  (Intercept) 2.671e+15  51679452
##  Residual             7.619e+16 276021320
## Number of obs: 50, groups:  country, 11
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 5.451e+08  4.643e+07 3.610e+00   11.74 0.000526 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The country-level variance in Engagement_Rate is negligible (5.16e-05), indicating that geographic factors have minimal influence. Most variability is at the individual level (Residual Variance = 1.05e-03), suggesting engagement rates are relatively consistent across countries. The average engagement rate across countries is small but statistically significant (0.0164, $p < 0.001$).

For Average_Likes, significant country-level variance (2.265e+11) highlights that geographic factors play a

meaningful role. However, individual-level variability is dominant (Residual Variance = 4.609e+12). The average likes across countries is substantial (1.511e+06, $p < 0.001$), indicating meaningful differences in influencer performance by geography.

Total_Likes shows considerable country-level variance (2.671e+15), with even larger residual variability (7.619e+16). This suggests that while country-specific factors strongly influence total likes, individual characteristics remain the primary drivers. The average total likes is very high (5.451e+08, $p < 0.001$), demonstrating the scale of influence at a global level.

2.2.Random Intercepts Model • Add Followers and Influence_Score as fixed effects to explain Engagement_Rate and Likes.

• Allow the intercepts to vary by country.

```
# Random intercepts model
model_eng <- lmer(X60_day_eng_rate ~ Followers_Scaled + Influence_Score_Scaled
                  + (1 | country), data = data)
summary(model_eng)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: X60_day_eng_rate ~ Followers_Scaled + Influence_Score_Scaled +
##     (1 | country)
##    Data: data
##
## REML criterion at convergence: -767.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.0391 -0.4097 -0.2428  0.0269  7.2796
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  country  (Intercept) 5.524e-05 0.007432
##  Residual             1.048e-03 0.032366
## Number of obs: 199, groups:  country, 26
##
## Fixed effects:
##                           Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)               0.015820   0.003545  17.248741   4.463 0.000331 ***
## Followers_Scaled         -0.001842   0.002543 195.963653  -0.724 0.469796
## Influence_Score_Scaled   -0.002405   0.002504 195.908031  -0.960 0.338047
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld  0.040
## Inflnc_Sc_S  0.069 -0.349
```

```
model_avglikes <- lmer(avg_likes ~ Followers_Scaled + Influence_Score_Scaled
                       + (1 | country), data = data)
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 1.9e-10
```

```r
summary(model_avglikes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: avg_likes ~ Followers_Scaled + Influence_Score_Scaled + (1 |
##     country)
##    Data: data
##
## REML criterion at convergence: 6295.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4765 -0.4845 -0.2532  0.1377  6.7088
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  country  (Intercept) 2.247e+11  474054
##  Residual             3.982e+12 1995499
## Number of obs: 200, groups:  country, 26
##
## Fixed effects:
##                            Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)             1555128.86  221591.33      19.58   7.018 9.35e-07 ***
## Followers_Scaled         894361.10  156833.28    6914.31   5.703 1.23e-08 ***
## Influence_Score_Scaled -242250.69  154441.17   10191.18  -1.569    0.117
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld  0.041
## Inflnc_Sc_S  0.070 -0.348
```

```r
model_totallikes <- lmer(total_likes ~ Followers_Scaled + Influence_Score_Scaled
                         + (1 | country), data = data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 2.5e-15
```

```r
summary(model_totallikes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: total_likes ~ Followers_Scaled + Influence_Score_Scaled + (1 |
##     country)
##    Data: data
##
## REML criterion at convergence: 1969.8
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1257 -0.5136  0.0413  0.6893  1.7336
##
## Random effects:
```

```
##  Groups   Name        Variance  Std.Dev.
##  country  (Intercept) 0.000e+00         0
##  Residual             7.479e+16 273475327
## Number of obs: 50, groups:  country, 11
##
## Fixed effects:
##                          Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)             4.988e+08  5.241e+07  2.543e+47   9.516   <2e-16 ***
## Followers_Scaled       -1.655e+08  9.477e+07  2.543e+47  -1.746   0.0808 .
## Influence_Score_Scaled  5.054e+07  3.397e+07  2.543e+47   1.488   0.1368
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld  0.668
## Inflnc_Sc_S -0.068 -0.240
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

The random intercept variance for Engagement_Rate is negligible (5.524e-05), suggesting minimal differences across countries. Neither Followers ($p = 0.470$) nor Influence_Score ($p = 0.338$) significantly predicts Engagement_Rate, indicating that engagement levels are relatively consistent globally and not strongly influenced by these factors.

Significant country-level variance (2.247e+11) indicates geographic differences in Average Likes. Followers is a strong predictor ($p < 0.001$), with higher follower counts leading to significantly increased average likes. However, Influence_Score is not significant ($p = 0.117$), suggesting limited impact on likes compared to follower count.

The model for Total Likes highlights a high residual variance (7.479e+16) with no detectable country-level random effects (Variance = 0). Neither Followers ($p = 0.0808$) nor Influence_Score ($p = 0.1368$) reaches significance, indicating that total likes are likely driven by other factors beyond those included in the model.

2.3.Random Slopes Model: • Allow the effects of Followers and Influence_Score to vary across countries.

```r
# Random slopes model
model_eng_slopes <- lmer(X60_day_eng_rate ~ Followers_Scaled
                         + Influence_Score_Scaled
                         + (Followers_Scaled
                            + Influence_Score_Scaled | country), data = data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```r
summary(model_eng_slopes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: X60_day_eng_rate ~ Followers_Scaled + Influence_Score_Scaled +
##     (Followers_Scaled + Influence_Score_Scaled | country)
##    Data: data
##
## REML criterion at convergence: -768.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.1508 -0.4018 -0.2277  0.0183  7.2710
```

```
##
## Random effects:
##  Groups   Name                   Variance  Std.Dev.  Corr
##  country  (Intercept)            5.043e-05 0.0071015
##           Followers_Scaled       3.744e-06 0.0019349 -1.00
##           Influence_Score_Scaled 3.916e-07 0.0006258 -1.00  1.00
##  Residual                        1.042e-03 0.0322794
## Number of obs: 199, groups:  country, 26
##
## Fixed effects:
##                          Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)              0.015618   0.003374  15.342731   4.628  0.00031 ***
## Followers_Scaled        -0.001500   0.002557  34.900974  -0.587  0.56127
## Influence_Score_Scaled  -0.002241   0.002505 171.605116  -0.895  0.37225
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld -0.194
## Inflnc_Sc_S  0.008 -0.355
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
model_avglikes_slopes <- lmer(avg_likes ~ Followers_Scaled
                              + Influence_Score_Scaled
                              + (Followers_Scaled
                                 + Influence_Score_Scaled | country),
                              data = data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 1.9e-10
```

```
summary(model_avglikes_slopes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## avg_likes ~ Followers_Scaled + Influence_Score_Scaled + (Followers_Scaled +
##     Influence_Score_Scaled | country)
##    Data: data
##
## REML criterion at convergence: 6294.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5065 -0.4902 -0.2286  0.1295  6.7500
##
## Random effects:
##  Groups   Name                   Variance  Std.Dev. Corr
##  country  (Intercept)            2.060e+11  453874
##           Followers_Scaled       2.504e+10  158248  -0.68
##           Influence_Score_Scaled 1.626e+10  127508  -0.85  0.97
##  Residual                        3.929e+12 1982118
```

```
## Number of obs: 200, groups:  country, 26
##
## Fixed effects:
##                           Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)             1.546e+06  2.100e+05  1.682e+01   7.361 1.19e-06 ***
## Followers_Scaled        9.572e+05  1.736e+05  3.672e+00   5.514  0.00675 **
## Influence_Score_Scaled -2.005e+05  1.599e+05  3.294e+01  -1.254  0.21871
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld -0.103
## Inflnc_Sc_S -0.090 -0.248
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
model_totallikes_slopes <- lmer(total_likes ~ Followers_Scaled
                                + Influence_Score_Scaled
                                + (Followers_Scaled
                                   + Influence_Score_Scaled | country),
                                data = data)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
## Warning: Model failed to converge with 2 negative eigenvalues: -1.4e-02
## -2.5e-02
```

```
## Warning in as_lmerModLT(model, devfun): Model may not have converged with 1
## eigenvalue close to zero: 3.5e-15
```

```
summary(model_totallikes_slopes)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## total_likes ~ Followers_Scaled + Influence_Score_Scaled + (Followers_Scaled +
##     Influence_Score_Scaled | country)
##    Data: data
##
## REML criterion at convergence: 1969.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2227 -0.5529  0.1701  0.6706  1.5562
##
## Random effects:
##  Groups   Name                    Variance  Std.Dev.  Corr
##  country  (Intercept)             0.000e+00         0
##           Followers_Scaled        4.246e+15  65165144   NaN
##           Influence_Score_Scaled  3.681e+15  60675191   NaN -1.00
##  Residual                         7.071e+16 265911943
## Number of obs: 50, groups:  country, 11
##
## Fixed effects:
##                           Estimate Std. Error         df t value Pr(>|t|)
```

```
## (Intercept)             4.922e+08  5.218e+07  1.764e+02   9.433   <2e-16 ***
## Followers_Scaled       -1.728e+08  9.739e+07  1.342e+02  -1.774   0.0783 .
## Influence_Score_Scaled  5.589e+07  4.309e+07  4.303e+00   1.297   0.2598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) Fllw_S
## Fllwrs_Scld  0.619
## Inflnc_Sc_S -0.078 -0.310
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

The random slopes model for Engagement Rate shows negligible variance in slopes for Followers (3.744e-06) and Influence_Score (3.916e-07) across countries, indicating that these factors have a consistent (non-significant) effect globally. Neither Followers ($p = 0.561$) nor Influence_Score ($p = 0.372$) significantly predicts Engagement Rate, further confirming the consistent and minimal impact of these predictors across countries.

For Average Likes, the random slopes for Followers (2.504e+10) and Influence_Score (1.626e+10) indicate some variability in their effects across countries. Followers remains a significant predictor ($p = 0.007$), while Influence_Score is not ($p = 0.219$), suggesting that follower counts consistently influence average likes globally, with moderate variability by country.

The random slopes model for Total Likes exhibits substantial variance for Followers (4.246e+15) and Influence_Score (3.681e+15), indicating notable geographic differences in how these factors impact total likes. However, neither Followers ($p = 0.078$) nor Influence_Score ($p = 0.260$) is statistically significant in this model, suggesting that other unaccounted factors may drive total likes.

3.Model Comparison

Compare models to evaluate the added benefit of including random slopes over random intercepts.

```
anova(null_model_eng, model_eng, model_eng_slopes)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
## Models:
## null_model_eng: X60_day_eng_rate ~ 1 + (1 | country)
## model_eng: X60_day_eng_rate ~ Followers_Scaled + Influence_Score_Scaled + (1 | country)
## model_eng_slopes: X60_day_eng_rate ~ Followers_Scaled + Influence_Score_Scaled + (Followers_Scaled +
##                   npar     AIC     BIC logLik deviance  Chisq Df Pr(>Chisq)
## null_model_eng       3 -789.30 -779.42 397.65  -795.30
## model_eng            5 -787.48 -771.02 398.74  -797.48 2.1877  2     0.3349
## model_eng_slopes    10 -778.24 -745.31 399.12  -798.24 0.7569  5     0.9797
```

```
anova(null_model_avglikes, model_avglikes, model_avglikes_slopes)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
## Models:
## null_model_avglikes: avg_likes ~ 1 + (1 | country)
## model_avglikes: avg_likes ~ Followers_Scaled + Influence_Score_Scaled + (1 | country)
## model_avglikes_slopes: avg_likes ~ Followers_Scaled + Influence_Score_Scaled + (Followers_Scaled + I
##                        npar    AIC    BIC  logLik deviance   Chisq Df Pr(>Chisq)
## null_model_avglikes       3 6410.3 6420.2 -3202.1   6404.3
```

```
## model_avglikes           5 6383.6 6400.1 -3186.8   6373.6 30.6432  2  2.218e-07
## model_avglikes_slopes   10 6392.4 6425.4 -3186.2   6372.4  1.2073  5     0.9442
##
## null_model_avglikes
## model_avglikes        ***
## model_avglikes_slopes
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(null_model_totallikes, model_totallikes, model_totallikes_slopes)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: data
## Models:
## null_model_totallikes: total_likes ~ 1 + (1 | country)
## model_totallikes: total_likes ~ Followers_Scaled + Influence_Score_Scaled + (1 | country)
## model_totallikes_slopes: total_likes ~ Followers_Scaled + Influence_Score_Scaled + (Followers_Scaled
##                         npar    AIC    BIC  logLik deviance  Chisq Df
## null_model_totallikes      3 2091.8 2097.5 -1042.9   2085.8
## model_totallikes           5 2091.5 2101.0 -1040.7   2081.5 4.3344  2
## model_totallikes_slopes   10 2101.2 2120.4 -1040.6   2081.2 0.2317  5
##                         Pr(>Chisq)
## null_model_totallikes
## model_totallikes            0.1145
## model_totallikes_slopes     0.9987
```

The comparison of models for Engagement Rate indicates no significant improvement when adding fixed effects (Followers and Influence_Score) or allowing their slopes to vary across countries. The random slopes model (AIC = -778.24, p = 0.9797) does not significantly outperform the simpler random intercept model or the null model. This suggests that Engagement Rate is not meaningfully influenced by Followers, Influence_Score, or their variation across countries.

For Average Likes, the model with fixed effects (AIC = 6383.6, $p < 0.001$) significantly improves upon the null model, indicating that Followers and Influence_Score play a role in explaining variability. However, adding random slopes (AIC = 6392.4, p = 0.944) does not improve the model further, suggesting that the effects of Followers and Influence_Score are consistent across countries.

The models for Total Likes show no significant improvement when adding fixed effects (AIC = 2091.5, p = 0.1145) or random slopes (AIC = 2101.2, p = 0.9987). This indicates that neither Followers nor Influence_Score sufficiently explains variability in total likes, and their effects do not vary meaningfully across countries.

4.Extract Variance Components

Evaluate the proportion of variance in the outcome explained at the country level.

```
summary(model_eng)$varcor
```

```
##  Groups   Name        Std.Dev.
##  country  (Intercept) 0.0074321
##  Residual             0.0323658
```

```
summary(model_avglikes)$varcor
```

```
##  Groups   Name        Std.Dev.
##  country  (Intercept)  474054
##  Residual             1995499
```

```
summary(model_totallikes)$varcor
```

```
##  Groups   Name        Std.Dev.
##  country  (Intercept)         0
##  Residual             273475327
```

The variance decomposition for Engagement Rate shows a small country-level standard deviation (0.0074) compared to the residual variability (0.0324). This indicates that most of the variability in Engagement Rate is at the individual level, with minimal influence from country-level differences.

For Average Likes, the country-level standard deviation is substantial (474,054), but the residual standard deviation is much larger (1,995,499). This suggests that while country-level factors contribute meaningfully to the variability in average likes, individual-level factors still dominate.

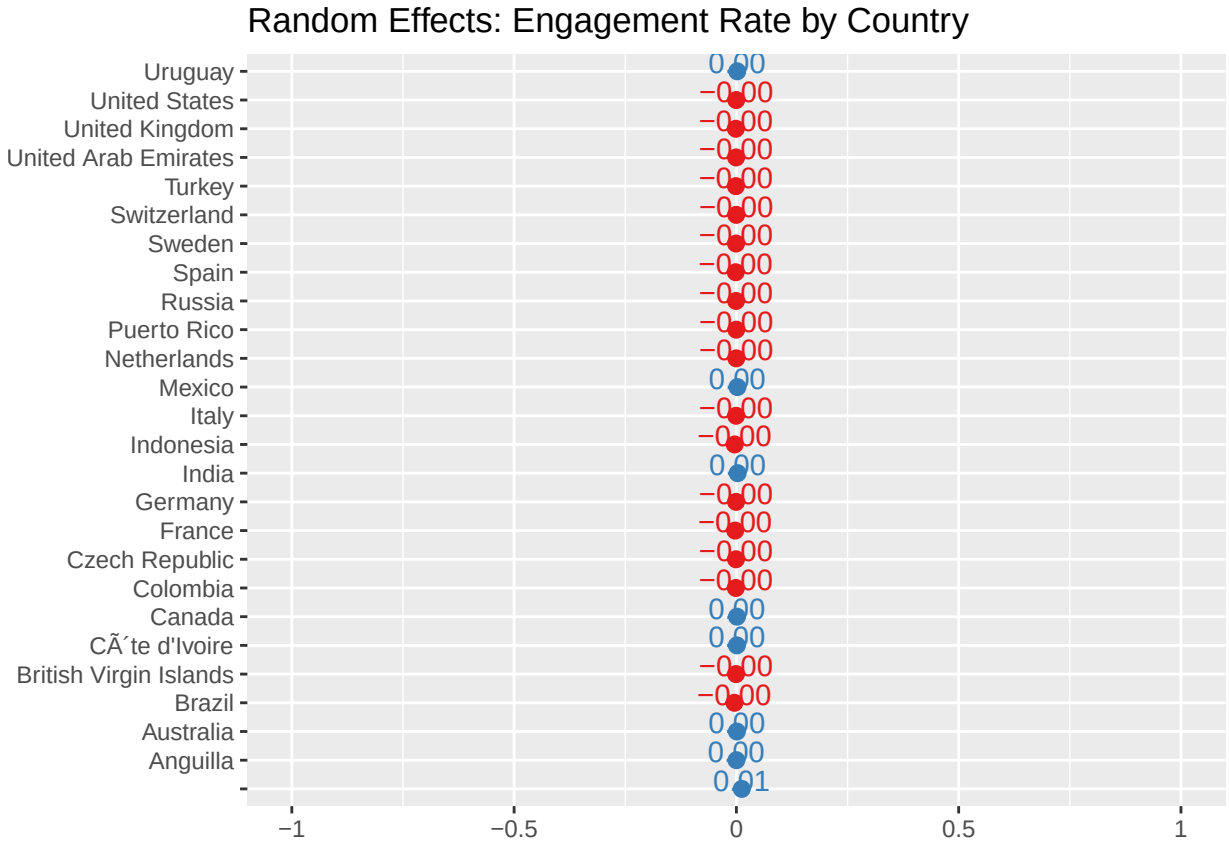
The variance decomposition for Total Likes reveals no detectable country-level variance (0), with all variability attributed to the residual standard deviation (273,475,327). This suggests that geographic factors have no significant impact on total likes, and the differences are entirely driven by individual-level factors.

5.Visualize Results

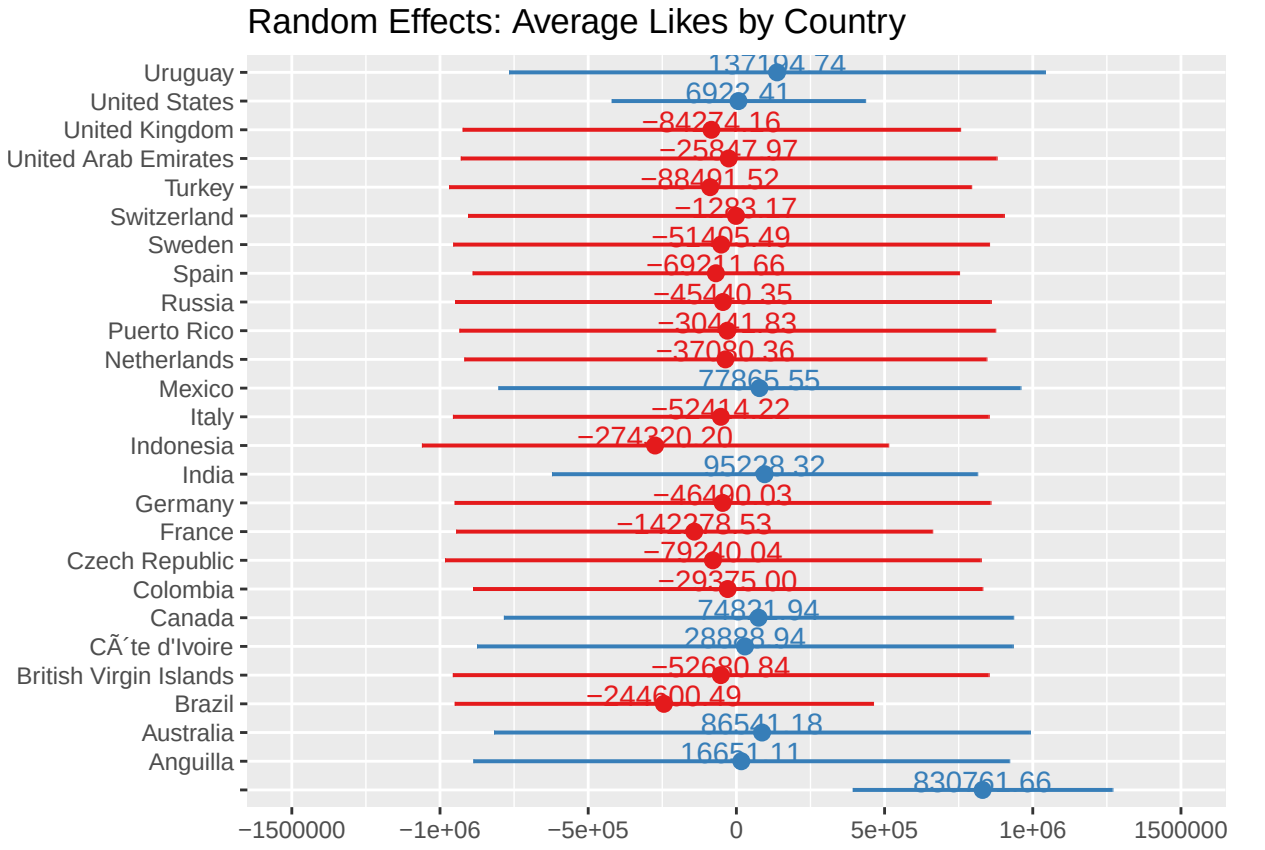
1.Plot the random effects to see country-level variability.

2.Plot predicted engagement rates and likes by country.

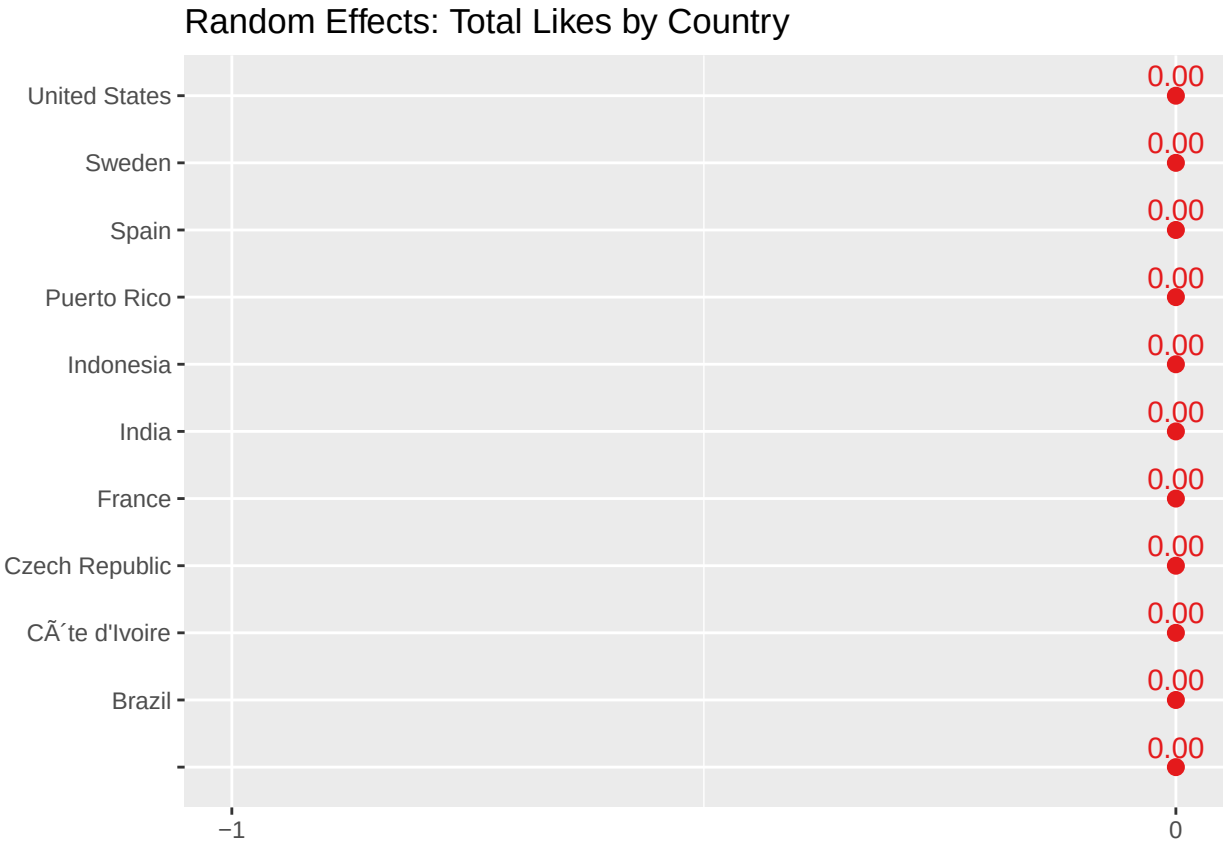
```
# Plot random effects
library(sjPlot)

# Random effects for Engagement Rate
plot_model(model_eng, type = "re", show.values = TRUE, value.offset = 0.3,
           title = "Random Effects: Engagement Rate by Country")
```



```r
# Random effects for Likes
plot_model(model_avglikes, type = "re", show.values = TRUE, value.offset = 0.3,
           title = "Random Effects: Average Likes by Country")
```




```r
plot_model(model_totallikes, type = "re", show.values = TRUE,
           value.offset = 0.3, title = "Random Effects: Total Likes by Country")
```



The random effects for Engagement Rate reveal minimal variability at the country level, as most random effects are close to zero. This aligns with the earlier findings that Engagement Rate is predominantly influenced by individual-level factors, with little to no geographic differentiation.

For Average Likes, there is notable country-level variability. Countries like Anguilla and Uruguay show positive random effects, indicating higher average likes than the global average. In contrast, countries like Indonesia, Brazil, and France show negative random effects, suggesting lower-than-average performance. This highlights significant geographic differences in the distribution of average likes.

The random effects for Total Likes are uniformly zero across countries, confirming the absence of country-level variance. This indicates that Total Likes are entirely determined by individual-level factors, and geographic characteristics play no role in influencing this outcome.

Conclusion

Overall, the findings suggest that geographic factors have minimal impact on Engagement Rate, as country-level variance is negligible, and individual-level factors predominantly drive this metric. For Average Likes, country-level differences are significant, with some countries consistently outperforming others, highlighting notable regional disparities. In contrast, Total Likes is entirely determined by individual-level characteristics, with no observable country-level effects, indicating that global audience behaviors, rather than geographic differences, shape this metric. These results emphasize the varying roles of geographic factors across different engagement metrics.