Analysis of the Original CLT/FLE Study

1. Exploration of the Original Dataset

This first section aims to explore the key variables upon which we can base our replication study of Díaz-Lago and Matute (2019) . Their open dataset includes data from two experiments that examine the relationship between the Causality Bias and the Foreign Language Effect (FLE). In this section, we will also investigate critical differences between the two experiments in order to determine which one more closely aligns with the study we intend to develop.

1.1 Importing Data

As the first step, we will import the dataset and convert the variables into appropriate formats for analysis.

```
# Importing the dataset
datacomplete <- read.csv2("datasetFLE.csv")

# Converting to factor
datacomplete$experiment <- as.factor(datacomplete$experiment)
datacomplete$gender <- as.factor(datacomplete$gender)
datacomplete$nativeLanguage <- as.factor(datacomplete$nativeLanguage)
datacomplete$experimentLanguage <- as.factor(datacomplete$experimentLanguage)
datacomplete$contingency <- as.factor(datacomplete$contingency)
```

The variables are as follows:

- `Experiment`: nominal variable with two levels (first experiment or second experiment);
- `Age`: numerical variable (age expressed in years);
- `Gender`: nominal variable with two levels (M or F);
- `Native Language` (NL): nominal variable with two levels (English or Spanish);

- `Language` used in the experiment: nominal variable with two levels (NL or FL);
- `Contingency`: nominal variable with two levels (causal illusion or true causality);
- `Self-assessed fluency` in the `native` language (NL; scale from 1 to 40), interval numerical variable;
- `Self-assessed fluency` in the `foreign` language (FL; scale from 1 to 40), interval numerical variable;
- Age of acquisition of FL (`AoA`): numerical variable (expressed in years);
- `Comprehension Test`: 5 true/false questions based on a text to read; it has been treated as a numerical variable (numbers of corrected responses), but it could be evaluated as a dichotomous variable (True or False);
- `CRT` Test (level of System 1 or 2 usage): count of correct answers, from 0 to 3, numerical variable.

```
# Column names of the dataframe
names(datacomplete)
```

```
 [1] "experiment"          "age"                 "gender"
 [4] "nativeLanguage"      "experimentLanguage" "contingency"
 [7] "SelfEvalNative"      "AoAForeign"          "SelfEvalForeign"
[10] "ComprehensionTest"   "CRT"                 "CausalJudgment"
```

```
# Structure of the dataframe
str(datacomplete)
```

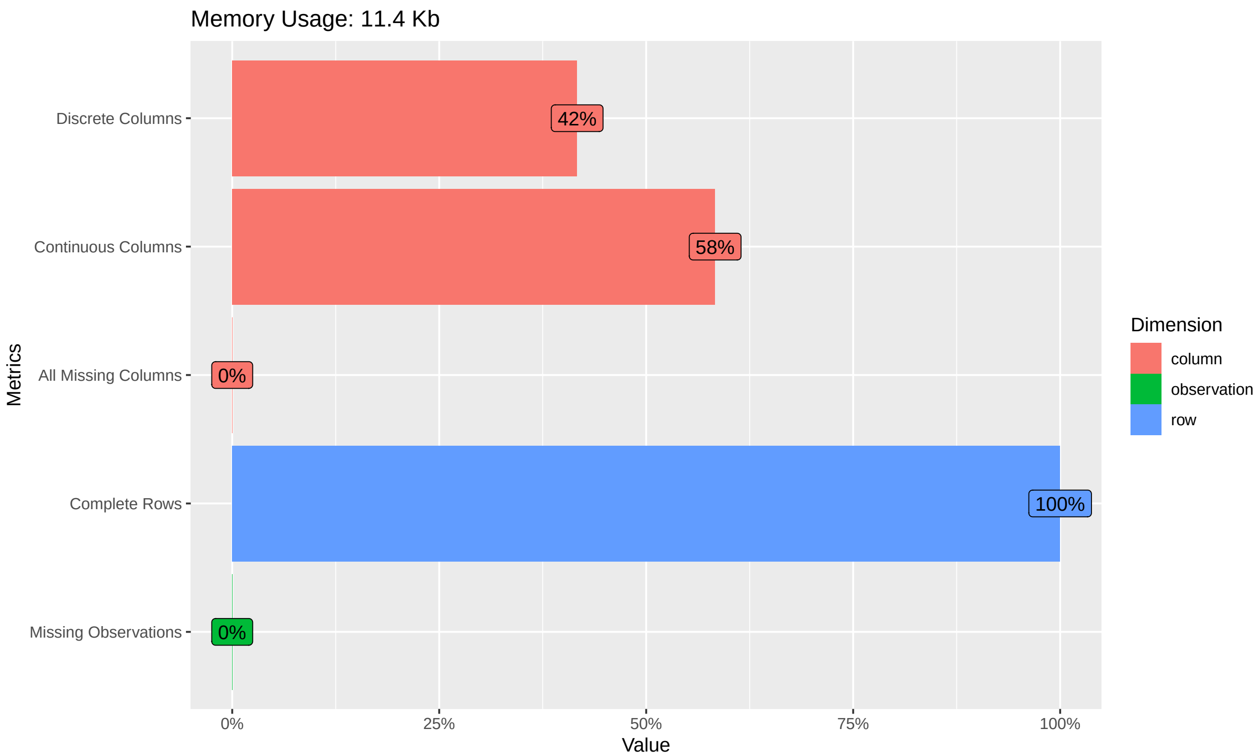
```
'data.frame':   116 obs. of  12 variables:
 $ experiment        : Factor w/ 2 levels "Experiment1",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
 $ age                : int  20 21 21 20 21 21 20 21 21 20 ...
 $ gender             : Factor w/ 2 levels "man","woman": 2 2 2 2 2 2 2 2 2 2 ...
 $ nativeLanguage     : Factor w/ 2 levels "English","Spanish": 1 1 1 1 1 1 1 1 1 1 ...
 $ experimentLanguage: Factor w/ 2 levels "Foreign","Native": 1 1 1 2 1 1 2 2 2 1 ...
 $ contingency        : Factor w/ 2 levels "contingent","non-continent": 2 2 2 2 2 2 2 2
 $ SelfEvalNative     : int  40 40 40 38 40 40 40 40 40 40 ...
 $ AoAForeign         : int  14 5 12 4 6 6 14 19 7 10 ...
 $ SelfEvalForeign    : int  22 33 25 23 28 29 19 7 27 20 ...
 $ ComprehensionTest : int  3 4 5 2 3 4 3 4 5 3 ...
 $ CRT                : int  0 0 2 0 1 0 1 0 1 0 ...
 $ CausalJudgment     : int  50 44 32 57 60 75 65 73 70 44 ...
```

1.2 Some preliminary observations

First, we check the dataset for any missing or incomplete data. Based on the exploration, we confirm that there are no missing values in the dataset.

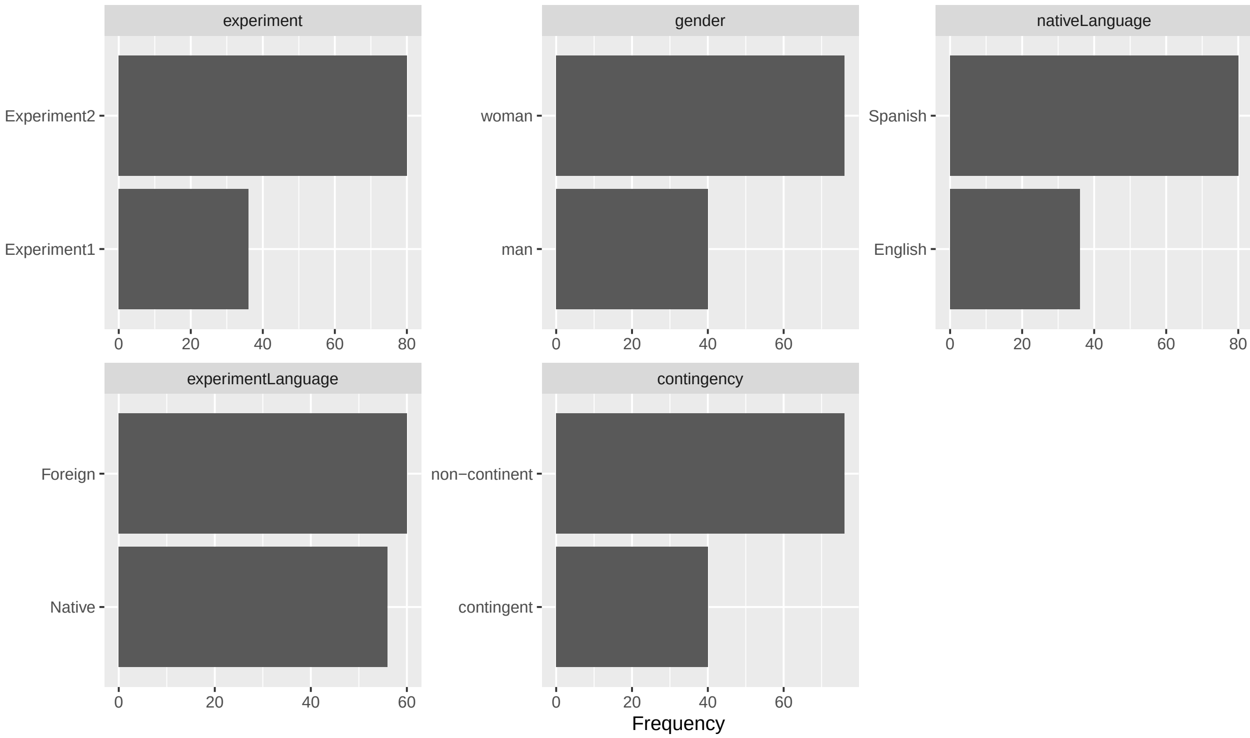
```
# Check for missing values in the dataset
DataExplorer::plot_intro(datacomplete)
```



1.3 Qualitative variables

In this section, we examine the qualitative (categorical) variables in the dataset.

```
# General exploration of qualitative variables
DataExplorer::plot_bar(datacomplete)
```



From the initial exploration, we observe that the first experiment has fewer participants compared to the second (36 vs. 80). The first experiment follows a factorial design with two groups based on experiment language (FL vs. NL). The second experiment includes four groups, based on experiment language and contingency (null vs. true). Additionally, more females (76) participated compared to males (40), with experiment 2 showing a more balanced sex distribution (3:1 ratio vs. ~1.5:1 ratio).

```
# Frequencies tables of Experiment and Gender
table(datacomplete$experiment); table(datacomplete$gender)
```

```
Experiment1 Experiment2
         36          80

  man woman
   40    76
```

```r
table(datacomplete$experiment,datacomplete$gender)
```

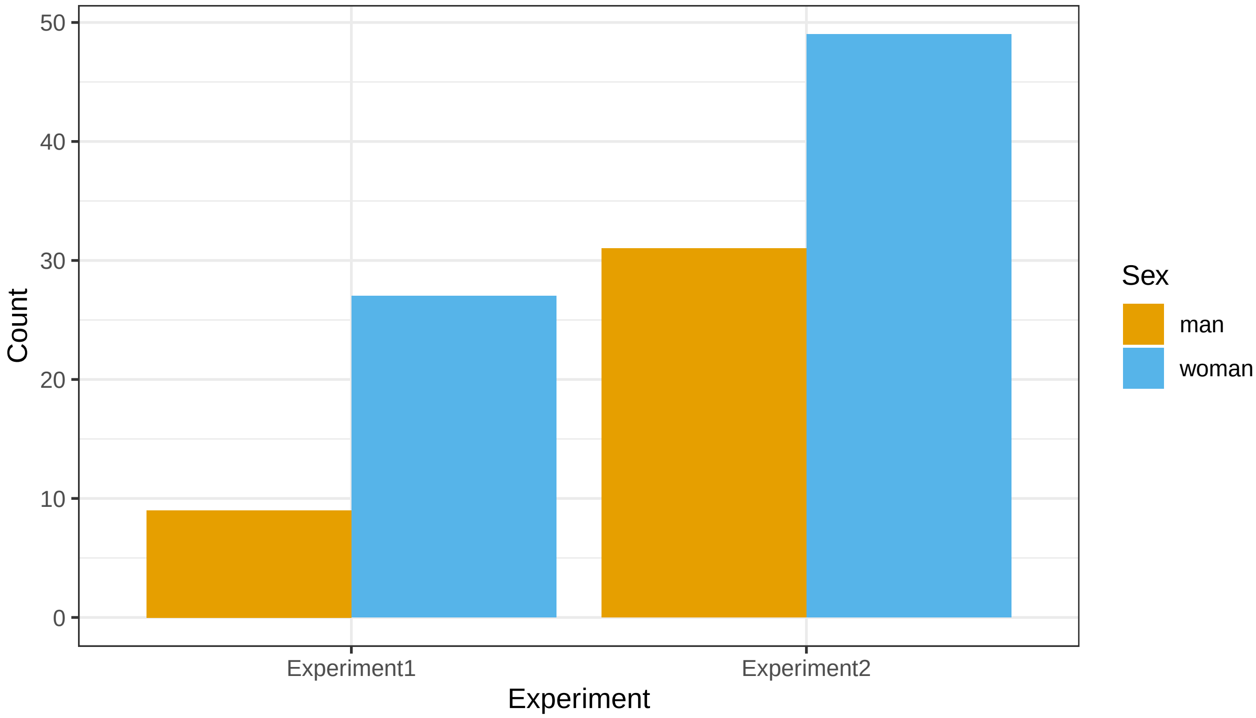
```
              man woman
  Experiment1   9    27
  Experiment2  31    49
```

```r
# Ratio
table(datacomplete$experiment,datacomplete$gender)[,2]/
  table(datacomplete$experiment,datacomplete$gender)[,1]
```

```
Experiment1 Experiment2
   3.000000    1.580645
```

```r
# Bar plot
library(ggplot2); library(ggokabeito)

ggplot(datacomplete, aes(x = experiment, fill = gender)) +
  geom_bar(position = position_dodge(preserve = "single")) +
  scale_fill_okabe_ito() +  # Color-blind friendly palette
  labs(x = "Experiment", y = "Count") +
  guides(fill = guide_legend(title = "Sex")) +
  theme_bw()
```



In the first experiment, participants were English students with Spanish as a FL, whereas in the second experiment, participants were Spanish students with English as a FL. This explains the difference in the first explorative bar plot concerning the differences in terms of NL (36 vs 80). In the first experiment, we have only the null contingency condition (N=36), whereas in the second experiment, we also have the true contingency condition (N=40 vs N=40), explaining the differences in the imbalance between the bars of the contingency variable.

```
# Frequencies tables of NL and Contingency x Experiment
table(datacomplete$nativeLanguage)
```

```
English Spanish
     36      80
```

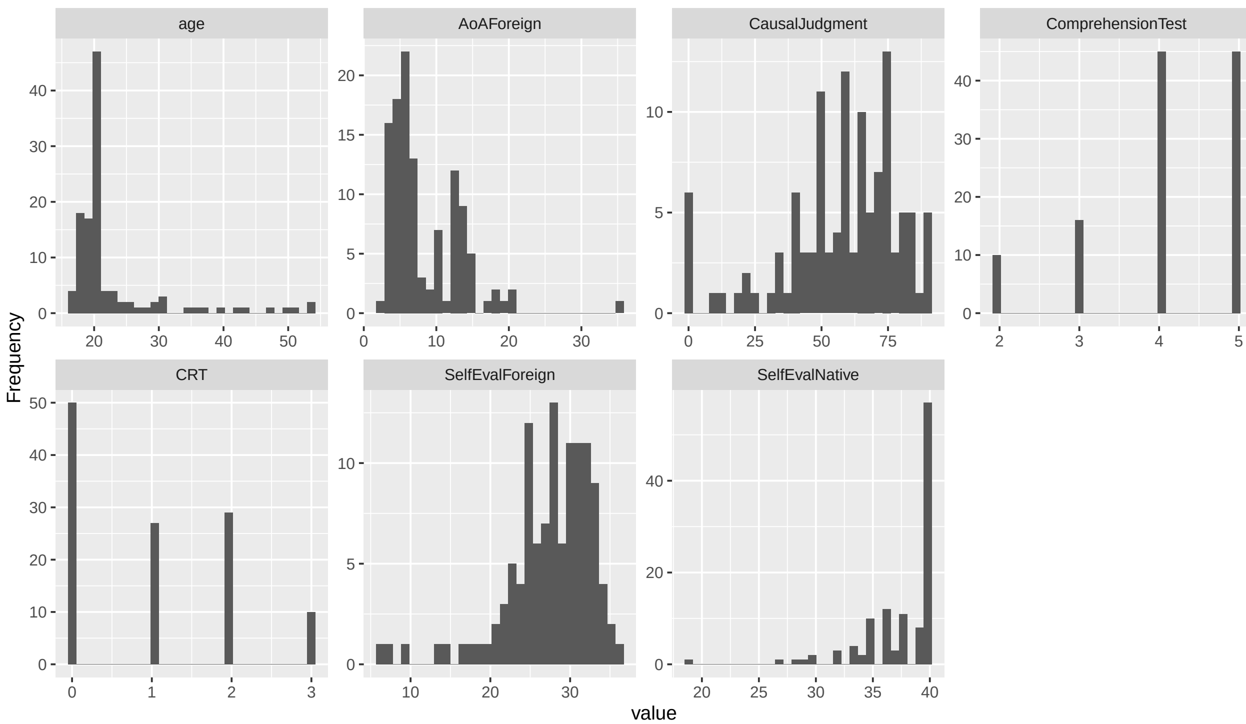
```
table(datacomplete$contingency, datacomplete$experiment)
```

```
                Experiment1 Experiment2
  contingent              0          40
  non-continent          36          40
```

1.4 Quantitative variables

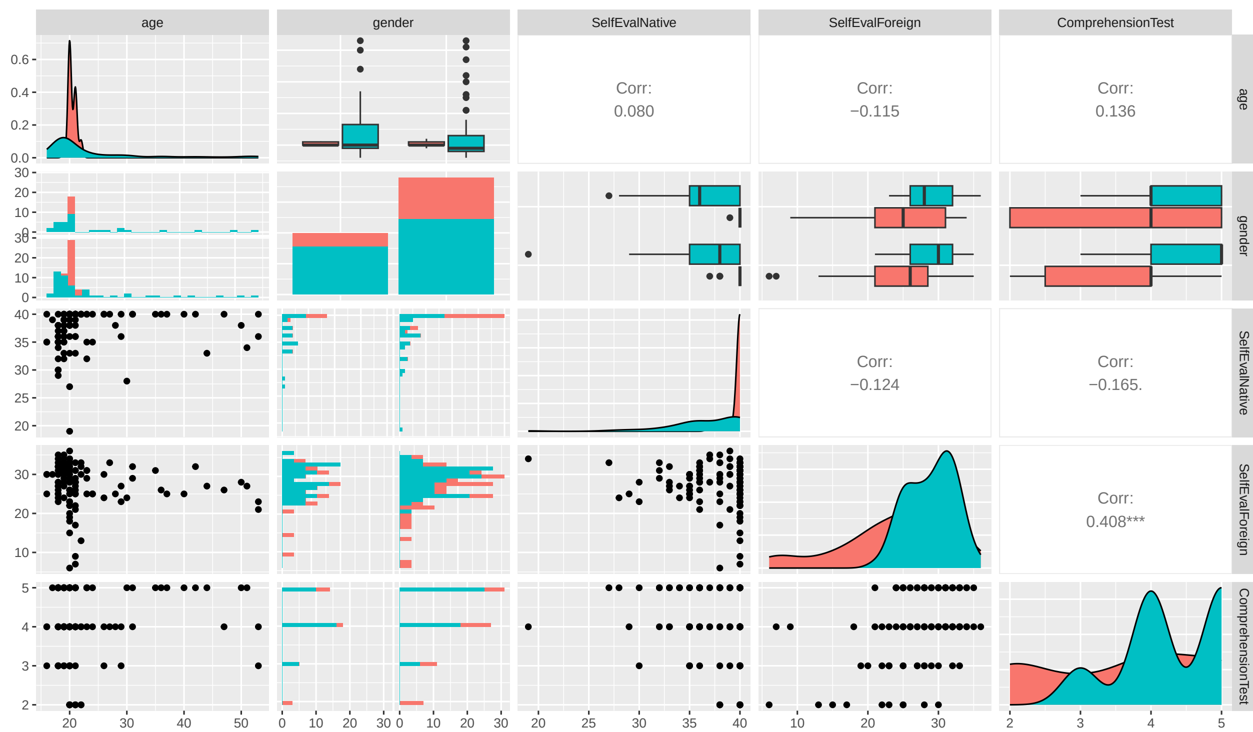
In this section, we explore the distributions of the continuous variables across the two experiments. We visualize the aggregated histograms to observe the general distribution of these variables.

```
# Plotting histograms for all continuous variables
DataExplorer::plot_histogram(datacomplete)
```



We also want to explore in greater detail the differences between the two experiments, as in the first experiment we have English students and in the second experiment, Spanish students.

```
# Plotting bivariate graphs for the continuous variables
library(GGally)
ggpairs(datacomplete[, c("age", "gender", "SelfEvalNative",
                          "SelfEvalForeign",
                     "ComprehensionTest", "experiment")], columns= 1:5,
        mapping = aes(fill=experiment))
```



1.4.1 Age

In this section, we focus on the Age variable. The average age of participants is 22.78 years (M = 22.78, Mdn = 20, SD = 7.82, MAD = 1), which is typical of university students, though there are participants outside the typical range, with ages spanning from 16 to 53. Notably, 80% of the participants fall within the age range of 16 to 23.

```
library(pastecs); library(ggdist); library(ggthemes)


# Descriptive statistics for Age
summary(datacomplete$age)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  16.00   19.00   20.00   22.78   22.00   53.00
```

```
round(stat.desc(datacomplete$age, norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
      116.00         0.00         0.00        16.00        53.00        37.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
     2642.00        20.00        22.78         0.73         1.44        61.22 
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE 
        7.82         0.34         2.49         5.55         5.64         6.33 
  normtest.W   normtest.p 
        0.62         0.00 
```

```
# MAD and range
median(abs(datacomplete$age - median(datacomplete$age))) # MAD
```

```
[1] 1
```

```
mad(datacomplete$age)  # MAD function
```

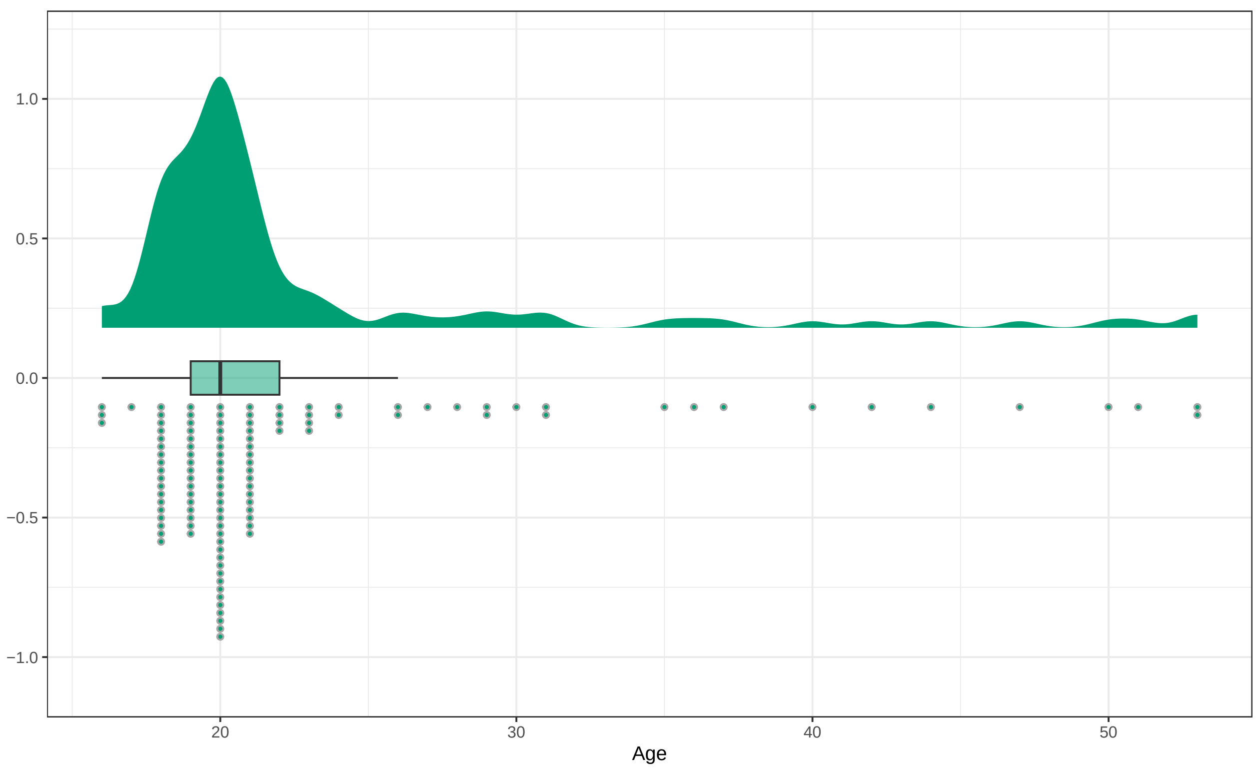
```
[1] 1.4826
```

```
quantile(datacomplete$age, probs = c(0, 0.80)) # Quantile for 80% range
```

```
 0% 80% 
 16  23 
```

```
# Distribution of Age
ggplot(datacomplete, aes(y = age, fill = factor(1))) +
  scale_fill_okabe_ito(order=3) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
```

```
               .width = 0, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(y = "Age", x = "") +
  coord_flip() +
  guides(fill = guide_legend(title = "")) +
  theme_bw() +
  theme(legend.position = "none")
```



We observe that the age distributions differ between the two experiments. English students (M = 20.47, Mdn = 20, 1st Quartile = 20, 3rd Quartile = 21, range = 3, SD = 0.7) on Erasmus tend to be more homogeneous in age compared to the Spanish students (M = 23.81, Mdn = 20, 1st Quartile = 18, 3rd Quartile = 24.5, range = 37, SD = 9.24). The Spanish students show more variability, with a wider age range and more outliers.

```r
# Age for each experiment
aggregate(datacomplete$age, list(datacomplete$experiment), summary)
```

```
       Group.1    x.Min.  x.1st Qu.  x.Median    x.Mean  x.3rd Qu.    x.Max.
1  Experiment1  19.00000   20.00000  20.00000  20.47222   21.00000  22.00000
2  Experiment2  16.00000   18.00000  20.00000  23.81250   24.50000  53.00000
```

```r
round(stat.desc(datacomplete$age[datacomplete$experiment == "Experiment1"],
                norm = TRUE), 2)
```

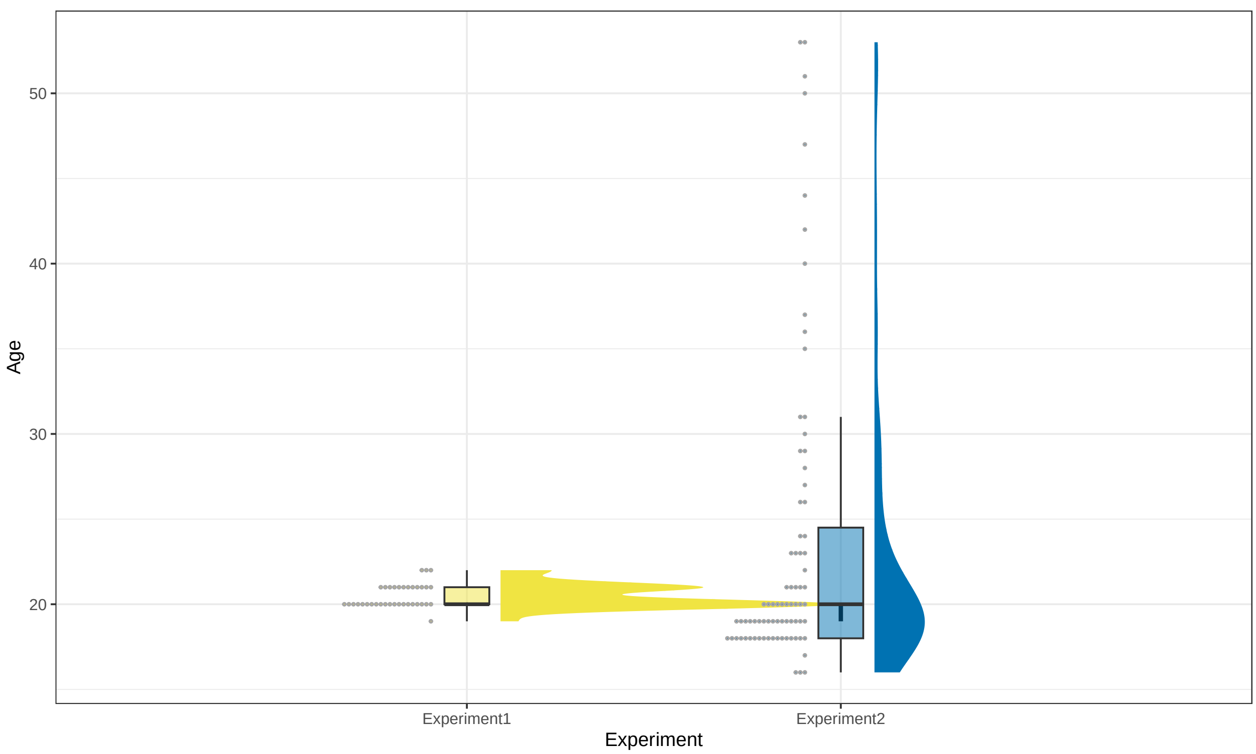
```
     nbr.val     nbr.null       nbr.na          min          max        range
       36.00         0.00         0.00        19.00        22.00         3.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      737.00        20.00        20.47         0.12         0.24         0.48
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        0.70         0.03         0.59         0.75        -0.29        -0.19
  normtest.W   normtest.p
        0.79         0.00
```

```r
round(stat.desc(datacomplete$age[datacomplete$experiment == "Experiment2"],
                norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       80.00         0.00         0.00        16.00        53.00        37.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     1905.00        20.00        23.81         1.03         2.06        85.39
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        9.24         0.39         1.87         3.47         2.52         2.37
  normtest.W   normtest.p
```

```
      0.70          0.00
```

```
# Age distributions
ggplot(datacomplete, aes(y = age, x = experiment, fill = experiment)) +
  stat_halfeye(adjust = 2, justification = -0.1,
               .width = 0.1, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x = "Experiment", y = "Age") +
  theme_bw() + theme(legend.position = "none")  +
  scale_fill_okabe_ito(order=c(4,5))
```



1.4.2 Age of FL acquisition

The Age of FL Acquisition (AoA) has an average of 8.45 years (M = 8.45, Mdn = 7, SD = 5.07, MAD = 3). The majority of participants started learning a foreign language during primary or secondary school, with the first quartile at age 5 and the third quartile at age 12. However, there are a few outliers (min = 2, max = 35). Over 99% of participants began studying a foreign language before turning 20.

```r
# AoA descriptive statistics
summary(datacomplete$AoAForeign)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.000   5.000   7.000   8.448  12.000  35.000 
```

```r
round(stat.desc(datacomplete$AoAForeign, norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
      116.00         0.00         0.00         2.00        35.00        33.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
      980.00         7.00         8.45         0.47         0.93        25.75 
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE 
        5.07         0.60         1.65         3.68         4.91         5.51 
  normtest.W   normtest.p 
        0.85         0.00 
```

```r
# MAD and proportion of participants with AoA > 20
median(abs(datacomplete$AoAForeign - median(datacomplete$AoAForeign)))
```

```
[1] 3
```

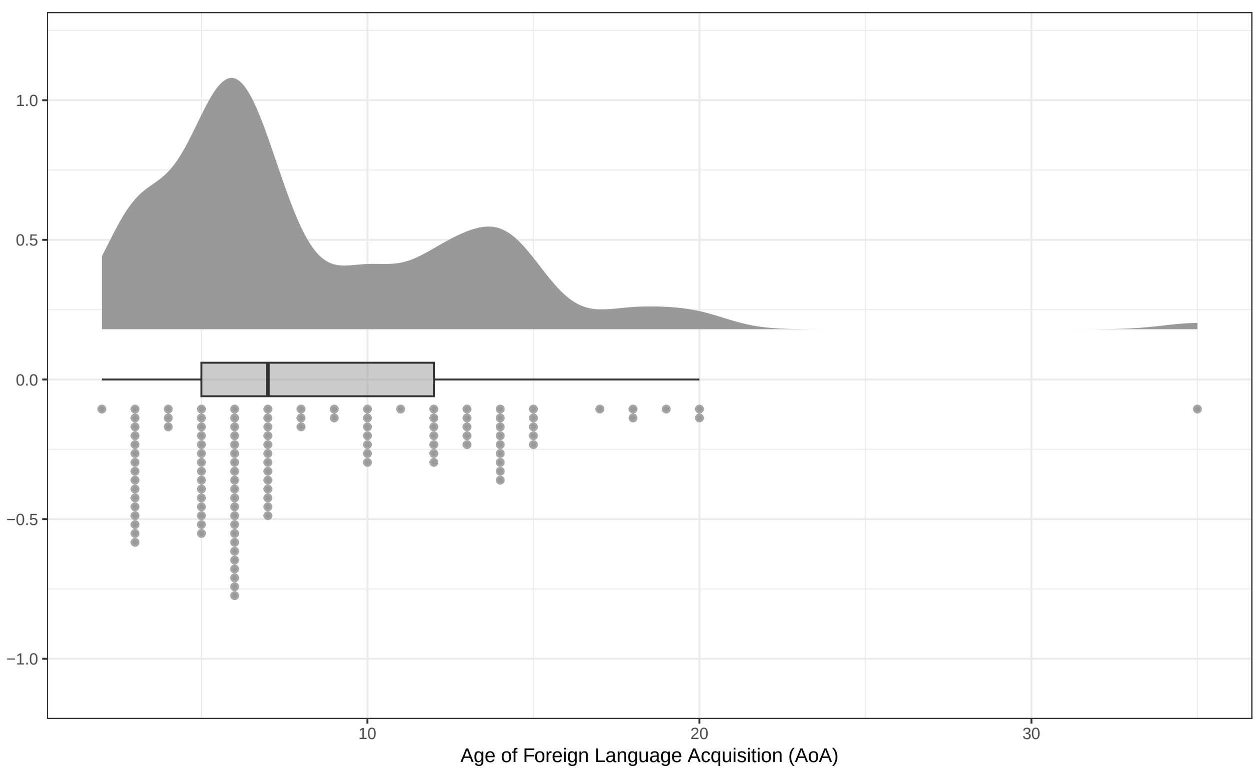
```r
mad(datacomplete$AoAForeign)
```

```
[1] 4.4478
```

```
sum(datacomplete$AoAForeign > 20) / length(datacomplete$AoAForeign)
```

```
[1] 0.00862069
```

```
# AoA distribution
ggplot(datacomplete, aes(y = AoAForeign, fill = factor(1))) +
  scale_fill_okabe_ito(order=8) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
               .width = 0, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x = "", y = "Age of Foreign Language Acquisition (AoA)") +
  coord_flip() +
  guides(fill = guide_legend(title = "")) +
  theme_bw() +
  theme(legend.position = "none")
```



An important consideration is whether there are differences in the AoA between the two groups of students. English-speaking students (M = 12.61, Mdn = 13.5, SD = 4.08, min = 4, max = 20) began learning Spanish at a later age compared to Spanish-speaking students (M = 6.57, Mdn = 6, SD = 4.32, min = 2, max = 35).

The Cohen's d, as standardized measure of effect size, is 1.43, suggesting a large effect. Additionally, the Cliff's delta, a non-parametric measure of effect size that is more robust to small sample sizes, non-normality, and heteroscedasticity, is 0.74, confirming the large effect.

```
# AoA descriptive statistics
aggregate(datacomplete$AoAForeign, list(datacomplete$experiment), summary)
```

```
      Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
1 Experiment1  4.00000  10.75000 13.50000 12.61111  15.00000 20.00000
2 Experiment2  2.00000   5.00000  6.00000  6.57500   7.00000 35.00000
```

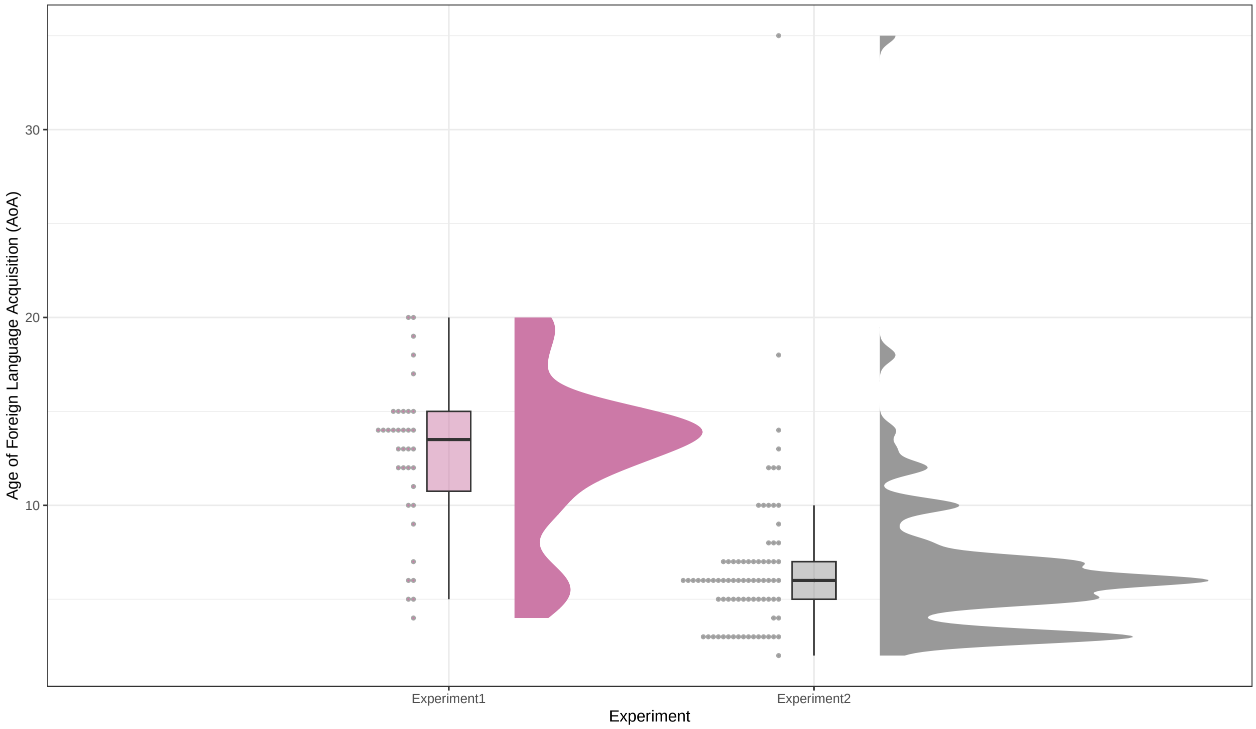
```
round(stat.desc(
  datacomplete$AoAForeign[datacomplete$experiment == "Experiment1"],
              norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       36.00         0.00         0.00         4.00        20.00        16.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      454.00        13.50        12.61         0.68         1.38        16.64
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        4.08         0.32        -0.39        -0.50        -0.36        -0.23
  normtest.W   normtest.p
        0.94         0.04
```

```
round(stat.desc(
  datacomplete$AoAForeign[datacomplete$experiment == "Experiment2"],
              norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       80.00         0.00         0.00         2.00        35.00        33.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      526.00         6.00         6.58         0.48         0.96        18.65
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        4.32         0.66         3.87         7.19        21.47        20.19
  normtest.W   normtest.p
        0.65         0.00
```

```
# AoA distributions
ggplot(datacomplete, aes(y = AoAForeign, x = experiment,
                         fill = experiment)) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
              .width = 0, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  scale_fill_okabe_ito(order = c(7, 8)) +
  labs(x = "Experiment",
       y = "Age of Foreign Language Acquisition (AoA)") +
  theme_bw() +
  theme(legend.position = "none")
```




```
# Overlap
library(overlapping)

overlap_stats <- list(
  x1 = datacomplete$AoAForeign[datacomplete$experiment == "Experiment1"],
  x2 = datacomplete$AoAForeign[datacomplete$experiment == "Experiment2"]
)

overlap_result <- overlapping::overlap(overlap_stats, type = "2")
overlap_result
```

```
$OV
[1] 0.2003548
```

```r
# Cohen's d
library(effectsize)


cohens_d_result <- cohens_d(
  x = datacomplete$AoAForeign[datacomplete$experiment == "Experiment1"],
  y = datacomplete$AoAForeign[datacomplete$experiment == "Experiment2"]
)
cohens_d_result
```

```
Cohen's d |       95% CI
------------------------
1.42      | [0.98, 1.85]


- Estimated using pooled SD.
```

```r
# Cliff's delta
cliffs_delta_result <- cliffs_delta(
  x = datacomplete$AoAForeign[datacomplete$experiment == "Experiment1"],
  y = datacomplete$AoAForeign[datacomplete$experiment == "Experiment2"]
)
cliffs_delta_result
```

```
r (rank biserial) |       95% CI
--------------------------------
0.74              | [0.61, 0.82]
```

1.4.3 Age - Age of foreign language acquisition

The difference between age and AoA provides insight into how many years participants have been practicing a foreign language. On average, participants have been speaking a foreign language for over 10 years, although there is considerable variability in the data (M = 14.33, Mdn = 14, SD = 7.8, 1st quartile = 10, 3rd quartile = 16).

```
# Calculating the difference
datacomplete$y <- datacomplete$age - datacomplete$AoAForeign


# Summary statistics of the difference
summary(datacomplete$y)
```

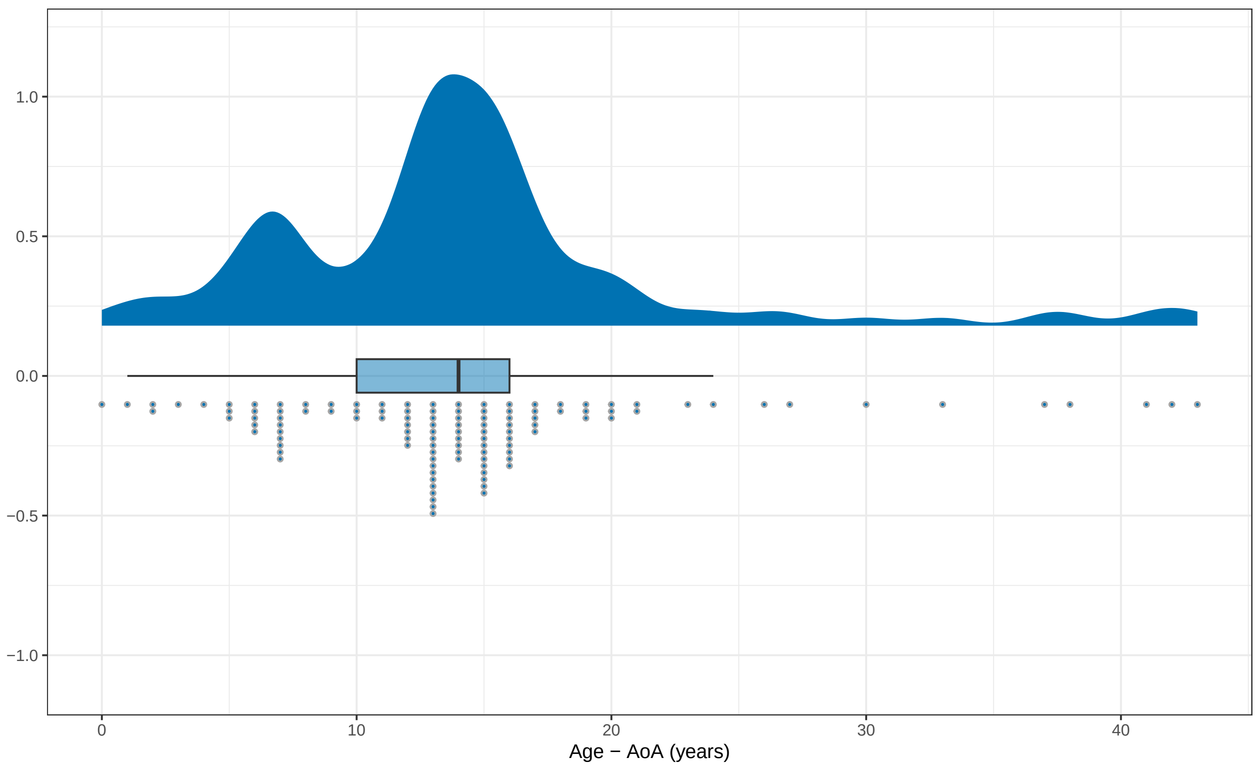
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   10.00   14.00   14.33   16.00   43.00 
```

```
round(stat.desc(datacomplete$y, norm=T),2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
      116.00         1.00         0.00         0.00        43.00        43.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
     1662.00        14.00        14.33         0.73         1.44        61.65 
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE 
        7.85         0.55         1.50         3.34         3.40         3.82 
  normtest.W   normtest.p 
        0.86         0.00 
```

```
# Distribution of the difference
ggplot(datacomplete, aes(y = y, fill=factor(1))) +
  scale_fill_okabe_ito(order=5) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
```

```
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x = "", y = "Age - AoA (years)") +
  coord_flip() +
  guides(fill = guide_legend(title = "")) +
  theme_bw() +
  theme(legend.position = "none")
```




```
# Summary statistics of the difference by group
aggregate(datacomplete$y, list(datacomplete$experiment), summary)
```

```
      Group.1    x.Min. x.1st Qu.  x.Median    x.Mean x.3rd Qu.    x.Max.
1 Experiment1  0.000000  6.000000  7.000000  7.861111 10.000000 16.000000
2 Experiment2  4.000000 13.000000 15.000000 17.237500 18.000000 43.000000
```

```
round(stat.desc(
  datacomplete$y[datacomplete$experiment == "Experiment1"], norm = TRUE), 2)
```

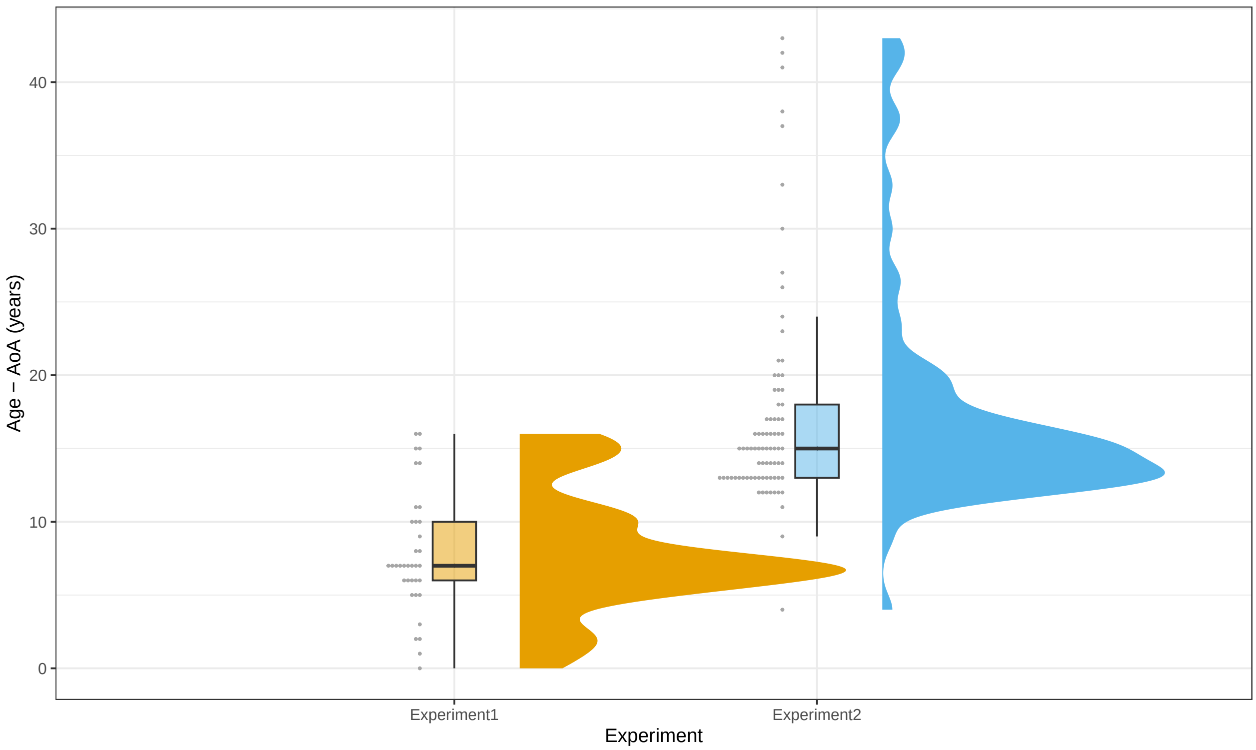
```
     nbr.val     nbr.null       nbr.na          min          max        range
       36.00         1.00         0.00         0.00        16.00        16.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      283.00         7.00         7.86         0.69         1.39        16.98
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        4.12         0.52         0.37         0.47        -0.47        -0.30
  normtest.W   normtest.p
        0.94         0.05
```

```
round(stat.desc(
  datacomplete$y[datacomplete$experiment == "Experiment2"], norm = TRUE), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       80.00         0.00         0.00         4.00        43.00        39.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     1379.00        15.00        17.24         0.83         1.64        54.59
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        7.39         0.43         2.01         3.74         3.83         3.61
  normtest.W   normtest.p
        0.73         0.00
```

```
# Distribution of the difference by group
ggplot(datacomplete, aes(y = y, x = experiment, fill = experiment)) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
               .width = 0, point_colour = NA) +
```

```
geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
scale_fill_okabe_ito(order = c(1, 2)) +
labs(x = "Experiment", y = "Age - AoA (years)") +
theme_bw() +
theme(legend.position = "none")
```



1.4.4 Comprehension test

The results of the comprehension test indicate that most participants were able to complete the test successfully. The median score was 4 (1st Quartile = 4, 3rd Quartile = 5). A few participants obtained slightly lower scores, with the minimum score being 2, but no participant scored 1 or 0. The comprehension test was designed to assess whether participants could understand information presented in a FL, providing a measure of their FL comprehension abilities.

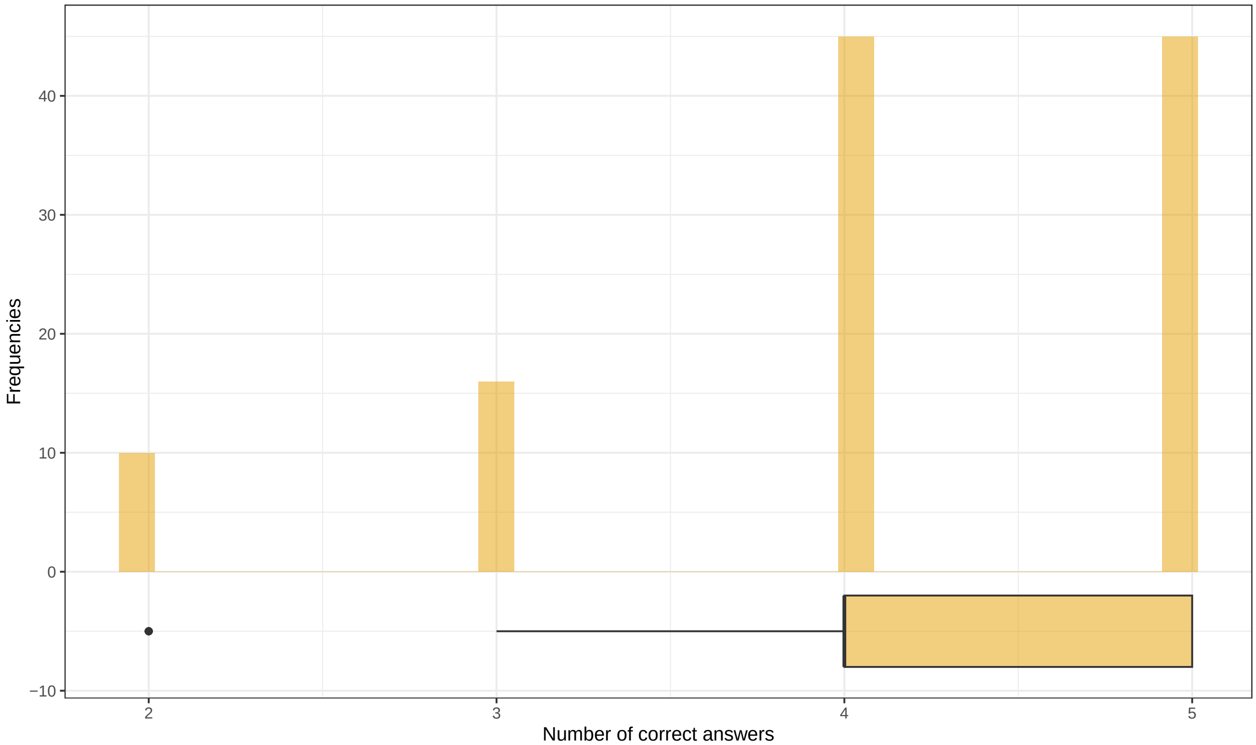
```
# Comprehension test description
summary(datacomplete$ComprehensionTest)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  2.000   4.000   4.000   4.078   5.000   5.000 
```

```
round(stat.desc(datacomplete$ComprehensionTest, norm=T),2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
      116.00         0.00         0.00         2.00         5.00         3.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
      473.00         4.00         4.08         0.09         0.17         0.87 
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE 
        0.93         0.23        -0.79        -1.75        -0.27        -0.31 
  normtest.W   normtest.p 
        0.82         0.00 
```

```
# Distribution of comprehension test scores
ggplot(datacomplete, aes(x = ComprehensionTest, fill=factor(1))) +
  geom_histogram(alpha = 0.5)+
  geom_boxplot(width=6, alpha=.5, position = position_nudge(y=-5))+
  scale_fill_okabe_ito(order=1)+
  labs(x="Number of correct answers", y="Frequencies")+
  theme_bw()+theme(legend.position="none")
```



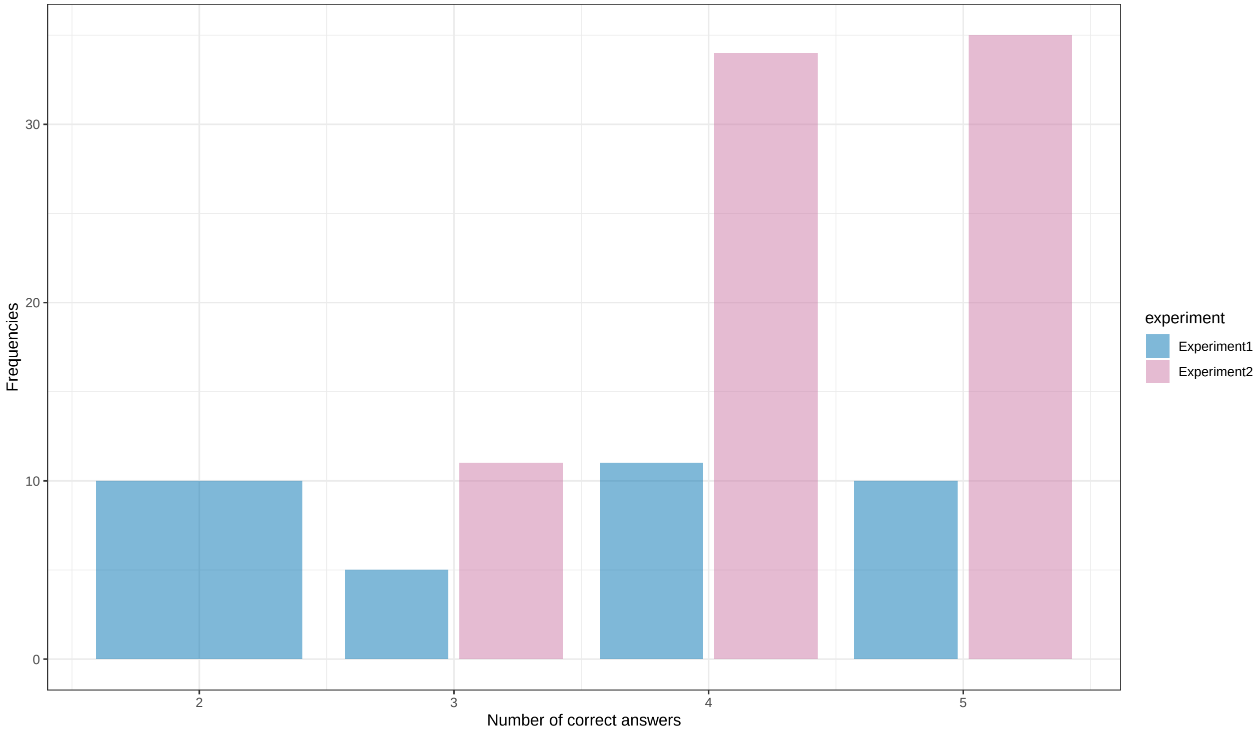
We now examine whether there are any differences between the two experimental groups in their comprehension test scores. English-speaking students performed slightly worse (M = 3.58, Mdn = 4) compared to Spanish-speaking students, who had a higher mean score (M = 4.30, Mdn = 4).

```
# Comparison of comprehension test scores between the two experiments
aggregate(datacomplete$ComprehensionTest, list(datacomplete$experiment),
          summary)
```

```
      Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
1 Experiment1 2.000000  2.000000 4.000000 3.583333  5.000000 5.000000
2 Experiment2 3.000000  4.000000 4.000000 4.300000  5.000000 5.000000
```

```
# Distributions of comprehension test scores by experiment group
ggplot(datacomplete, aes(x = ComprehensionTest, fill=experiment)) +
  geom_bar(alpha = 0.5, position=position_dodge2())+
  labs(x="Number of correct answers", y="Frequencies")+
```

```
  scale_fill_okabe_ito(order=c(5,7))+
  theme_bw()
```



1.4.4 Self-assessment of language fluency

The self-assessment scores of FL fluency (M = 27.48, Mdn = 28, SD = 5.44) are generally lower than those for NL fluency (M = 37.6, Mdn = 39, SD = 3.48). The Cohen's d value of 2.21 suggests a very large effect, indicating a substantial difference between the self-assessments of native and foreign language fluency. Moreover, the Cliff's Delta value of -0.92 suggests minimal overlap between the two distributions, further reinforcing the strong distinction between the self-reported fluency in the two languages.

```
# Self-assessment of NL fluency
summary(datacomplete$SelfEvalNative)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   19.0    36.0    39.0    37.6    40.0    40.0
```

```r
round(stat.desc(datacomplete$SelfEvalNative, norm=T),2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
      116.00         0.00         0.00        19.00        40.00        21.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     4362.00        39.00        37.60         0.32         0.64        12.08
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        3.48         0.09        -2.15        -4.78         6.48         7.27
  normtest.W   normtest.p
        0.72         0.00
```

```r
# Self-assessment of FL fluency
summary(datacomplete$SelfEvalForeign)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   6.00   25.00   28.00   27.48   31.00   36.00
```

```r
round(stat.desc(datacomplete$SelfEvalForeign, norm=T),2)
```

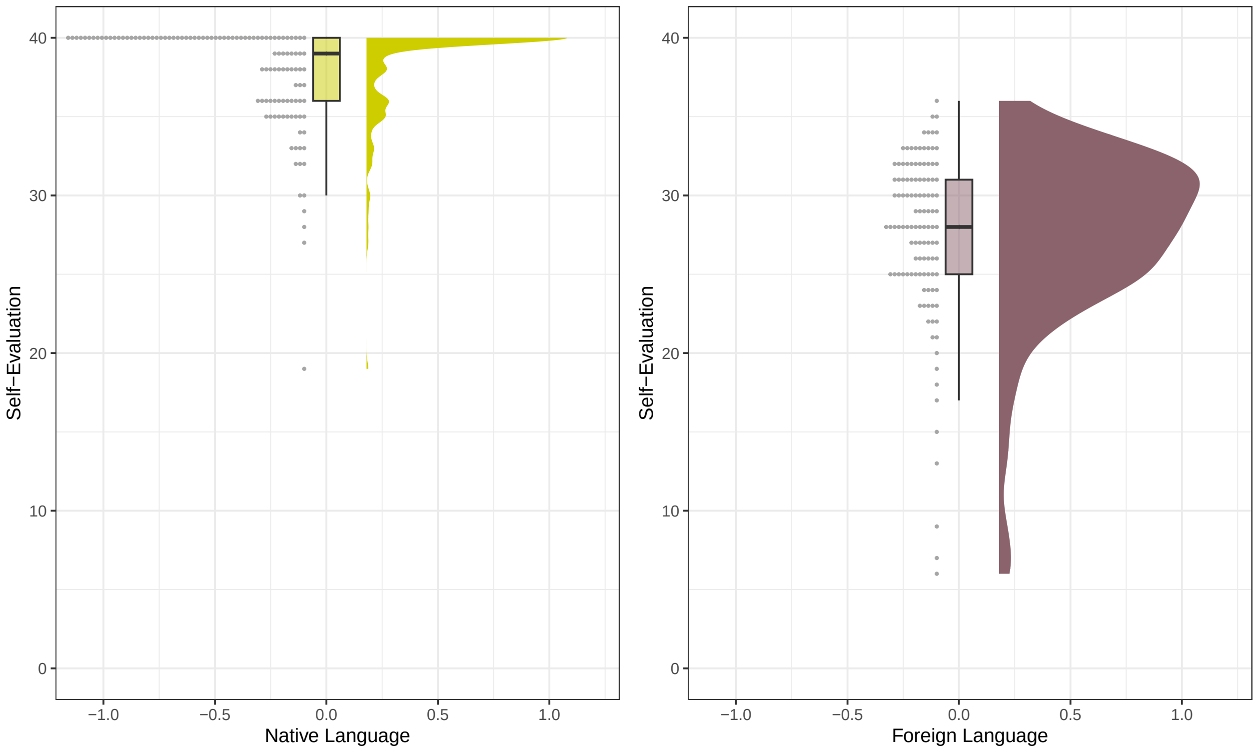
```
     nbr.val     nbr.null       nbr.na          min          max        range
      116.00         0.00         0.00         6.00        36.00        30.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     3188.00        28.00        27.48         0.51         1.00        29.61
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        5.44         0.20        -1.55        -3.45         3.35         3.76
  normtest.W   normtest.p
        0.88         0.00
```

```
# Visualizing the self-assessment of NL fluency
a <- ggplot(datacomplete, aes(y = SelfEvalNative, fill=factor(1))) +
  scale_fill_manual(values=c("yellow3")) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x="Native Language", y="Self-Evaluation") +
  guides(fill=guide_legend(title="")) +
  ylim(0,40) +
  theme_bw() +
  theme(legend.position="none")

# Visualizing the self-assessment of FL fluency
b <- ggplot(datacomplete, aes(y = SelfEvalForeign, fill=factor(1))) +
  scale_fill_manual(values=c("pink4")) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x="Foreign Language", y="Self-Evaluation") +
  guides(fill=guide_legend(title="")) +
  ylim(0,40) +
  theme_bw() +
  theme(legend.position="none")


# Displaying both plots together
```

```r
library(gridExtra)
grid.arrange(a, b, nrow = 1)
```




```r
# Cohen's d and Cliff's delta
cohens_d(datacomplete$SelfEvalNative, datacomplete$SelfEvalForeign)
```

```
Cohen's d |       95% CI
------------------------
2.22      | [1.89, 2.54]

- Estimated using pooled SD.
```

```r
cliffs_delta(datacomplete$SelfEvalNative, datacomplete$SelfEvalForeign)
```

```
r (rank biserial) |       95% CI
--------------------------------
0.92              | [0.90, 0.94]
```

When examining the self-assessment of NL fluency, English-speaking students rated their fluency higher (M = 39.72, Mdn = 40, SD = 0.74, min = 37, max = 40) compared to their Spanish-speaking counterparts (M = 36.65, Mdn = 38, SD = 3.79, min = 19, max = 40). It is notable that the Spanish group exhibited a greater variability in their self-assessments.
The standardized effect size is 0.97, which indicates a large effect, suggesting a meaningful difference in the self-assessed NL fluency between the two groups.

```
# Self-assessment of NL fluency in the two experiments
aggregate(datacomplete$SelfEvalNative, list(datacomplete$experiment), summary)
```

```
      Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
1 Experiment1 37.00000  40.00000 40.00000 39.72222  40.00000 40.00000
2 Experiment2 19.00000  35.00000 38.00000 36.65000  40.00000 40.00000
```

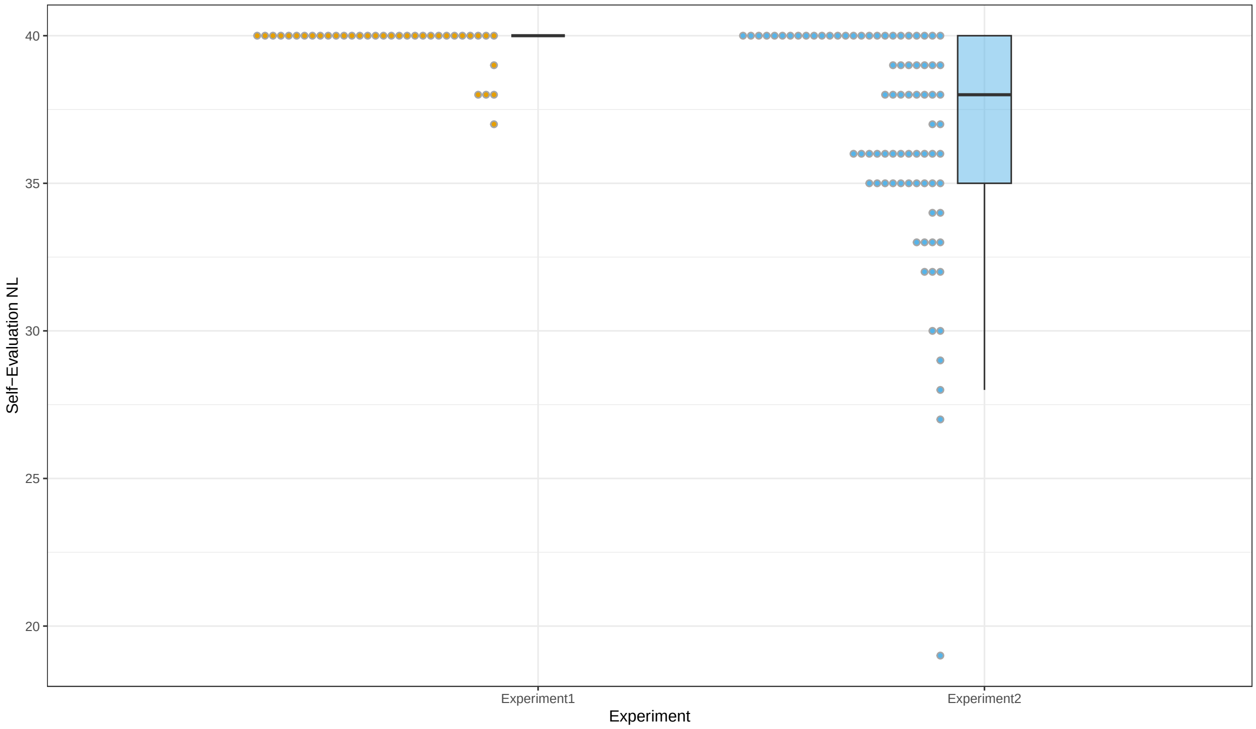
```
round(stat.desc(
  datacomplete$SelfEvalNative[datacomplete$experiment=="Experiment1"], norm=T),2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       36.00         0.00         0.00        37.00        40.00         3.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     1430.00        40.00        39.72         0.12         0.25         0.55
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
        0.74         0.02        -2.40        -3.06         4.53         2.95
  normtest.W   normtest.p
        0.43         0.00
```

```
round(stat.desc(
  datacomplete$SelfEvalNative[datacomplete$experiment=="Experiment2"], norm=T),2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
       80.00         0.00         0.00        19.00        40.00        21.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
     2932.00        38.00        36.65         0.42         0.84        14.38 
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE 
        3.79         0.10        -1.74        -3.23         4.52         4.25 
  normtest.W   normtest.p 
        0.81         0.00 
```

```r
# Visualizing the self-assessment of NL fluency
ggplot(datacomplete, aes(y = SelfEvalNative, x= experiment,
                         fill = experiment)) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  scale_fill_okabe_ito() +
  labs(x = "Experiment", y = "Self-Evaluation NL") +
  theme_bw() +
  theme(legend.position = "none")
```




```
# Cohen's d
cohens_d(x = datacomplete$SelfEvalNative[
  datacomplete$experiment == "Experiment1"],
         y = datacomplete$SelfEvalNative[
           datacomplete$experiment == "Experiment2"])
```

```
Cohen's d |       95% CI
------------------------
0.97      | [0.55, 1.38]

- Estimated using pooled SD.
```

```
# Cliff's Delta
cliffs_delta(datacomplete$SelfEvalNative[
  datacomplete$experiment == "Experiment1"],
```

```
                 datacomplete$SelfEvalNative[
                   datacomplete$experiment == "Experiment2"])
```

```
r (rank biserial) |       95% CI
--------------------------------
0.59              | [0.43, 0.72]
```

Regarding the self-assessment of FL fluency, English-speaking students rated their proficiency lower (M = 24.19, Mdn = 25.5, SD = 7.45, min = 6, max = 35) compared to their Spanish-speaking counterparts (M = 28.96, Mdn = 29.5, SD = 3.37, min = 21, max = 36), although the English group displayed greater variability in their ratings.
The standardized effect size is 0.96, which indicates a large effect, suggesting that the difference in self-assessed FL fluency between the two groups is meaningful.

```
# Self-assessment of FL fluency in the two experiments
aggregate(datacomplete$SelfEvalForeign, list(datacomplete$experiment), summary)
```

```
      Group.1   x.Min. x.1st Qu. x.Median   x.Mean x.3rd Qu.   x.Max.
1 Experiment1  6.00000  20.75000 25.50000 24.19444  29.25000 35.00000
2 Experiment2 21.00000  26.00000 29.50000 28.96250  32.00000 36.00000
```

```
round(stat.desc(
  datacomplete$SelfEvalForeign[datacomplete$experiment=="Experiment1"], norm=T),
  2)
```

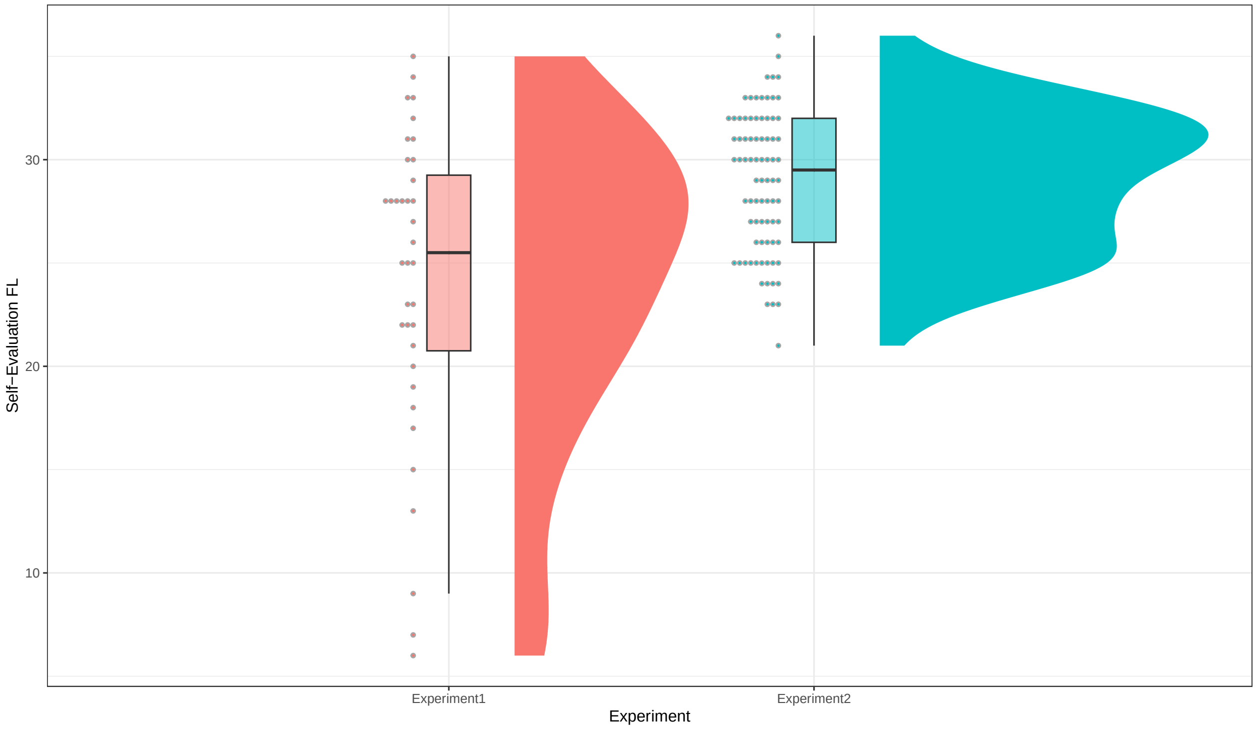
```
     nbr.val     nbr.null       nbr.na          min          max        range
       36.00         0.00         0.00         6.00        35.00        29.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      871.00        25.50        24.19         1.24         2.52        55.48
```

```
     std.dev      coef.var      skewness      skew.2SE      kurtosis      kurt.2SE
        7.45          0.31         -0.82         -1.04         -0.06         -0.04
  normtest.W    normtest.p
        0.93          0.02
```

```
round(stat.desc(
  datacomplete$SelfEvalForeign[datacomplete$experiment=="Experiment2"], norm=T),
  2)
```

```
     nbr.val      nbr.null        nbr.na           min           max         range
       80.00          0.00          0.00         21.00         36.00         15.00
         sum        median          mean       SE.mean  CI.mean.0.95           var
     2317.00         29.50         28.96          0.38          0.75         11.38
     std.dev      coef.var      skewness      skew.2SE      kurtosis      kurt.2SE
        3.37          0.12         -0.19         -0.36         -0.92         -0.87
  normtest.W    normtest.p
        0.97          0.04
```

```
# Visualizing the self-assessment of FL fluency
ggplot(datacomplete, aes(y = SelfEvalForeign, x= experiment,
                         fill = experiment)) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
               .width = 0, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 0.25) +
  labs(x = "Experiment", y = "Self-Evaluation FL") +
  theme_bw() +
  theme(legend.position = "none")
```




```
# Cohen's d
cohens_d(x = datacomplete$SelfEvalForeign[
  datacomplete$experiment == "Experiment1"],
         y = datacomplete$SelfEvalForeign[
           datacomplete$experiment == "Experiment2"])
```

```
Cohen's d |         95% CI
--------------------------
-0.96     | [-1.37, -0.54]

- Estimated using pooled SD.
```

```
# Cliff's Delta
cliffs_delta(datacomplete$SelfEvalForeign[
  datacomplete$experiment == "Experiment1"],
```

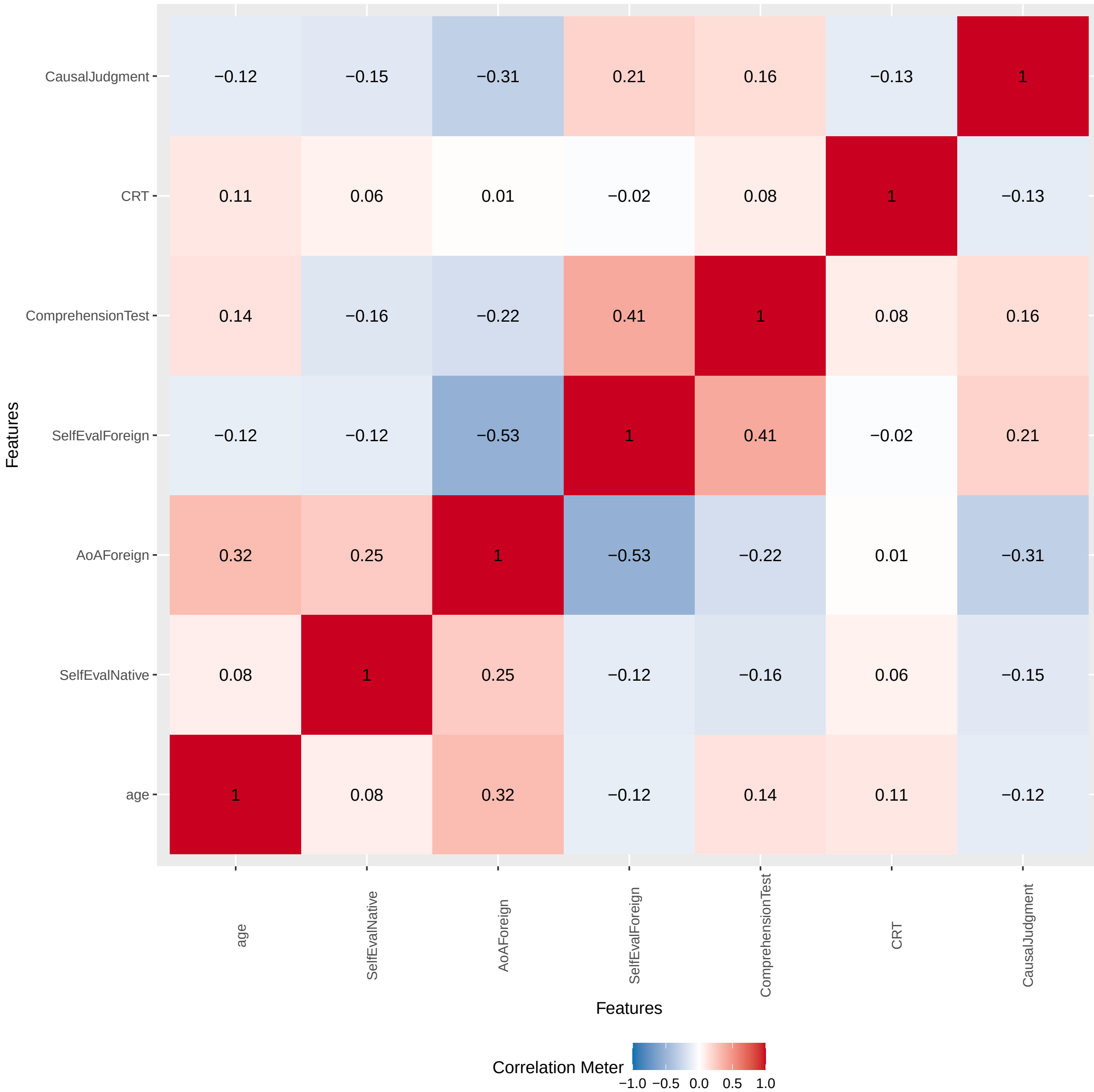
```
                datacomplete$SelfEvalForeign[
                  datacomplete$experiment == "Experiment2"])
```

```
r (rank biserial) |         95% CI
----------------------------------
-0.39             | [-0.57, -0.19]
```

1.4.5 Other bivariate relationships

We briefly observe the correlation matrix between continuous variables. Among other correlations worth noting, there is a moderate positive correlation (r = 0.41) between self-assessed foreign language fluency and the comprehension test. Consistent with expectations, those who acquired their foreign language at a younger age also rated their proficiency higher in the foreign language (r = -0.53).

```
DataExplorer::plot_correlation(datacomplete[,c(2,7,8,9,10,11,12)])
```



2. First experiment main results

2.1 Descriptive statistics

The first experiment included 36 participants, and the only condition tested was the null contingency condition.

```
# Extract data for the first experiment
data1 <- datacomplete[datacomplete$experiment == "Experiment1", ]
```

```r
nrow(data1)  # Confirming the number of participants
```

```
[1] 36
```

We now describe the subjective ratings of causality provided by participants, which were measured on a 101-point Likert scale. The scores ranged from 1 to 83, with the median (Mdn = 53.5) being fairly close to the mean (M = 51.31) and indicating a somewhat symmetrical distribution. The standard deviation (SD = 20.63) reflects moderate variability in participants' responses. The first and third quartiles (Q1 = 39.25, Q3 = 65.5) are relatively symmetric, suggesting a distribution that does not have extreme skewness.

The results of the Shapiro-Wilk normality test (W = 0.96, p = 0.25) indicate that the distribution does not significantly deviate from normality, meaning we do not reject the null hypothesis of normality.

```r
# Summary statistics
data1$CausalJudgment
```

```
 [1] 50 44 32 57 60 75 65 73 70 44 57 41 25 40 35 12 35 50 53 10  1 60 54 37 65
[26] 70 65 80 51 83 75 55 82 52 22 67
```

```r
summary(data1$CausalJudgment)
```

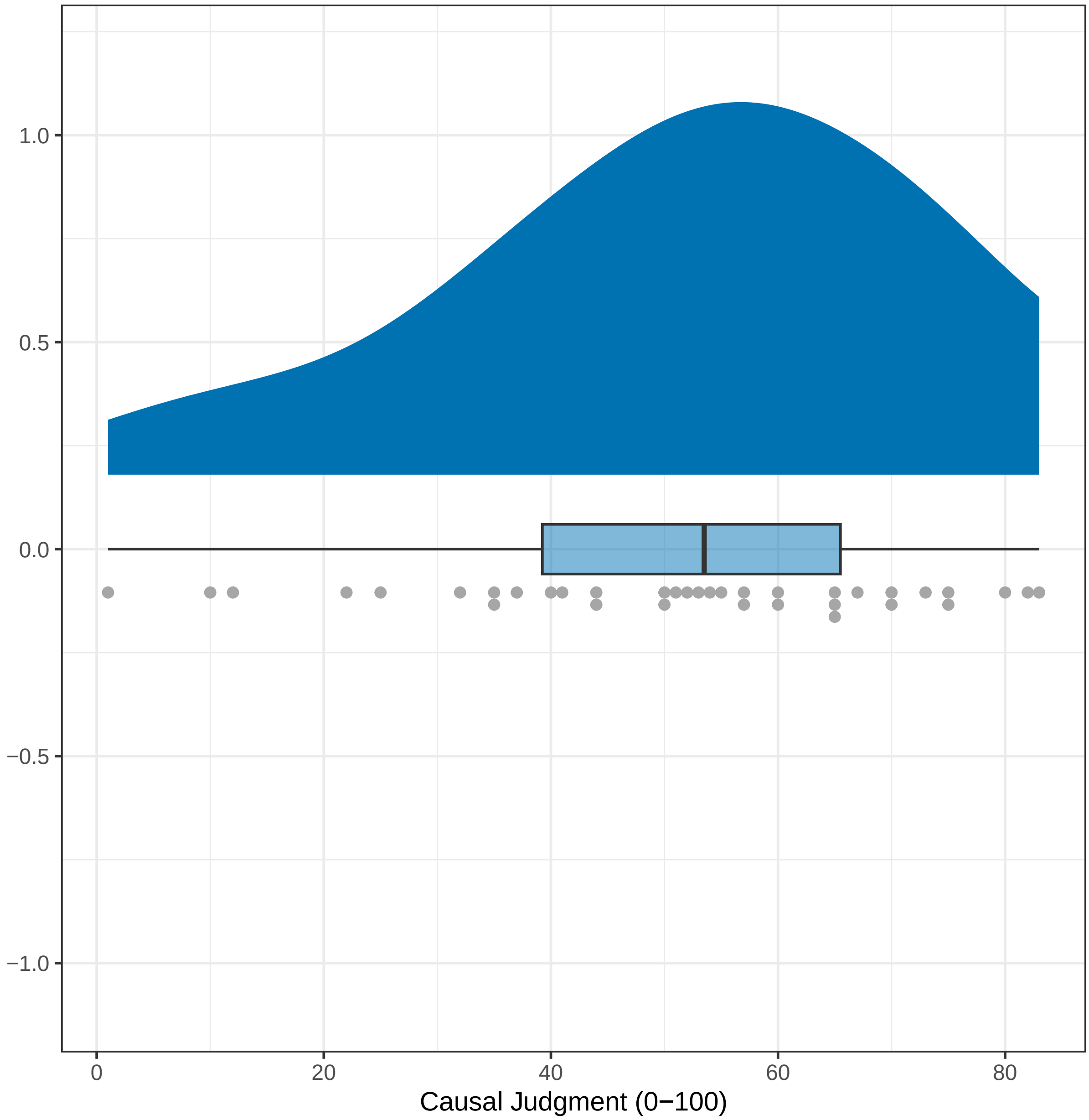
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   1.00   39.25   53.50   51.31   65.50   83.00 
```

```r
round(stat.desc(data1$CausalJudgment, norm = T), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range 
       36.00         0.00         0.00         1.00        83.00        82.00 
         sum       median         mean      SE.mean CI.mean.0.95          var 
```

```
    1847.00        53.50        51.31         3.44         6.98       425.65
    std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
      20.63         0.40        -0.56        -0.71        -0.37        -0.24
 normtest.W   normtest.p
       0.96         0.25
```

```
# Distribution of Causal Judgment
ggplot(data1, aes(y = CausalJudgment, fill = factor(1))) +
  scale_fill_okabe_ito(order = 5) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 1, size=2) +
  labs(x = "", y = "Causal Judgment (0-100)") +
  coord_flip() +
  guides(fill = guide_legend(title = "")) +
  theme_bw() +
  theme(legend.position = "none")
```



The goal now is to compare the causality scores between the two groups that completed the task in either their native or foreign language. There are 20 participants in the foreign language group and 16 participants in the native language group. NL is English, and FL is Spanish.

```
# Number of participants in each group
length(data1$CausalJudgment[data1$experimentLanguage == "Foreign"])
```

```
[1] 20
```

```r
length(data1$CausalJudgment[data1$experimentLanguage == "Native"])
```

```
[1] 16
```

The range of causality scores is wider in the FL condition (from 1 to 75) compared to the NL condition (from 22 to 83). This suggests that participants in the FL condition have a broader spread of judgments.

In the NL condition, the distribution of causality scores has a mean of 64.5 with a standard deviation of 15.12. The median score of 66 is very close to the mean, indicating a symmetrical distribution.

In contrast, in the FL condition, the mean causality score is 42.5 with a standard deviation of 18.43, indicating more variability in this group. The median score of 42 is very close to the mean, suggesting a fairly symmetric distribution. The first and third quartiles are 34 and 53.

```r
# Descriptive statistics
aggregate(x=data1$CausalJudgment, by=list(data1$experimentLanguage),
          FUN=summary)
```

```
  Group.1 x.Min. x.1st Qu. x.Median x.Mean x.3rd Qu. x.Max.
1 Foreign   1.00     34.25    42.50  40.75     53.25  75.00
2  Native  22.00     56.50    66.00  64.50     73.50  83.00
```

```r
round(stat.desc(data1$CausalJudgment[data1$experimentLanguage=="Foreign"], norm=T),2)
```

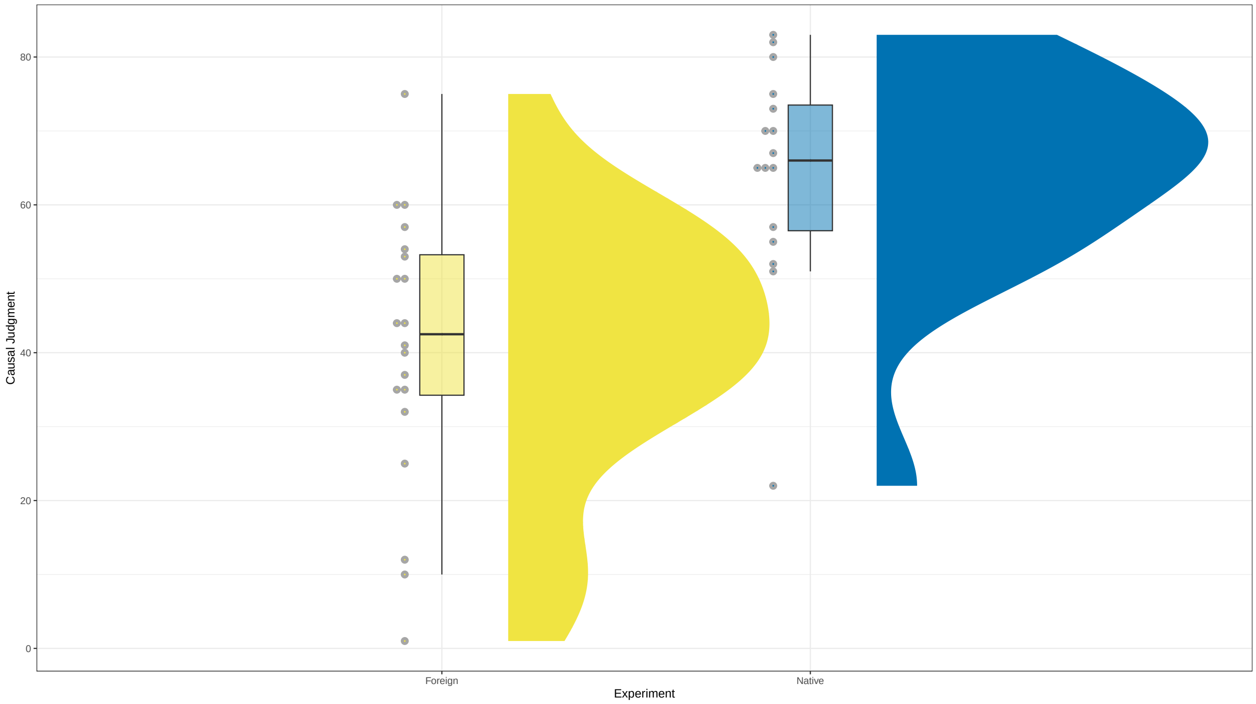
```
     nbr.val     nbr.null       nbr.na          min          max        range
       20.00         0.00         0.00         1.00        75.00        74.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      815.00        42.50        40.75         4.12         8.63       339.67
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
```

```
      18.43        0.45       -0.45       -0.44       -0.43       -0.22
 normtest.W   normtest.p
       0.96        0.60
```

```r
round(stat.desc(data1$CausalJudgment[data1$experimentLanguage=="Native"], norm=T),2)
```

```
    nbr.val     nbr.null       nbr.na          min          max        range
      16.00         0.00         0.00        22.00        83.00        61.00
        sum       median         mean      SE.mean CI.mean.0.95          var
    1032.00        66.00        64.50         3.78         8.06       228.67
    std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
      15.12         0.23        -1.17        -1.04         1.33         0.61
 normtest.W   normtest.p
       0.89         0.05
```

```r
# Visualization
ggplot(data1, aes(y = CausalJudgment, x = experimentLanguage,
                  fill = experimentLanguage)) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 1, size=2) +
  scale_fill_okabe_ito(order = c(4, 5)) +
  labs(x = "Experiment", y = "Causal Judgment") +
  theme_bw() +
  theme(legend.position = "none")
```



2.2 Effect size

The Cohen's d effect size is 1.39, which is very large.

```
# Cohen's d
cohens_d(x = data1$CausalJudgment[
  data1$experimentLanguage == "Foreign"],
         y = data1$CausalJudgment[
           data1$experimentLanguage == "Native"])
```

```
Cohen's d |         95% CI
--------------------------
-1.39     | [-2.12, -0.65]

- Estimated using pooled SD.
```

```
# Cliff's Delta
cliffs_delta(x = data1$CausalJudgment[
  data1$experimentLanguage == "Foreign"],
             y = data1$CausalJudgment[
               data1$experimentLanguage == "Native"])
```

```
r (rank biserial) |         95% CI
----------------------------------
-0.73             | [-0.86, -0.49]
```

We examine the BF. In this case, the data are 114 times more likely under the alternative hypothesis (H1) than under the null hypothesis (H0), suggesting strong evidence in favor of the effect.

```
# BF
library(BayesFactor)
ttestBF(x = data1$CausalJudgment[
  data1$experimentLanguage == "Foreign"],
        y = data1$CausalJudgment[
          data1$experimentLanguage == "Native"])
```

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 114.429 ±0%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

3. Second experiment main results

3.1 Descriptive statistics

The second experiment involved 80 participants. To analyze the data, we focus on the null contingency condition, which includes 40 data points.

```
# Extract data for the second experiment
data2 <- datacomplete[datacomplete$experiment == "Experiment2", ]
nrow(data2)
```

```
[1] 80
```

```
data2 <- data2[data2$contingency == "non-continent", ]
```

The subjective ratings of causality in the null contingency condition show a wide range of scores from 0 to 90. The mean is 53.65, with a median of 60, and the standard deviation is 25.06. The first and third quartiles are somewhat asymmetric (1st Quartile = 46.75, 3rd Quartile = 69.50). The distribution has a negative skew of -0.95 and a kurtosis value of 0.09. The Shapiro-Wilk test for normality ($W = 0.88$, $p < 0.05$) indicates that the distribution deviates from normality.

```
# Summary statistics
data2$CausalJudgment
```

```
 [1] 78 60 85 60 65 40 80 67 90  0 60 47 50 79 60 58 50 80 86 65 49 62  0 50 65
[26] 41 46 51  0 65 61 48 71  0  0 40 21 72 69 75
```

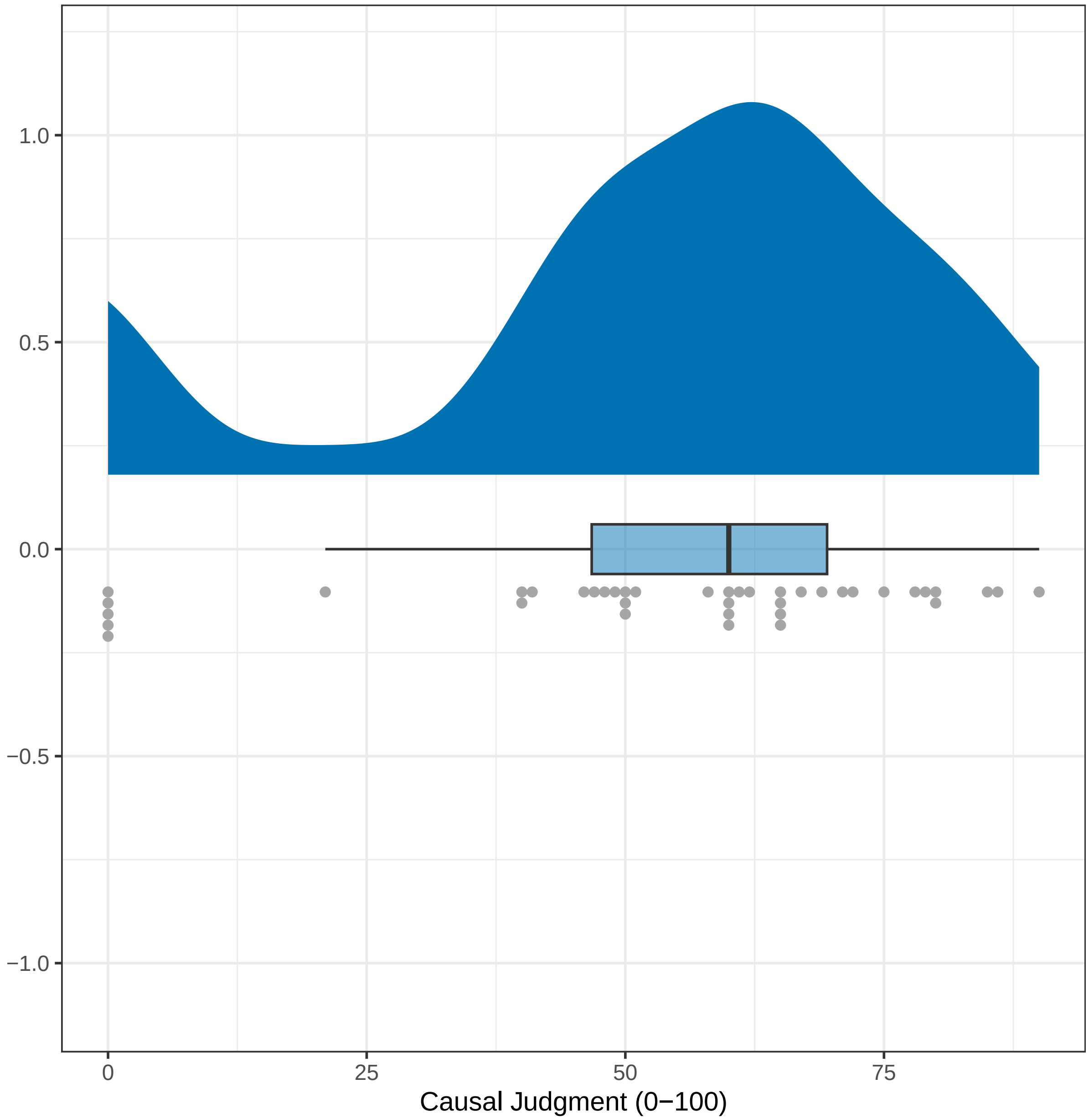
```
summary(data2$CausalJudgment)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   0.00   46.75   60.00   53.65   69.50   90.00 
```

```r
round(stat.desc(data2$CausalJudgment, norm = T), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       40.00         5.00         0.00         0.00        90.00        90.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     2146.00        60.00        53.65         3.96         8.02       628.08
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
       25.06         0.47        -0.95        -1.28         0.09         0.06
  normtest.W   normtest.p
        0.88         0.00
```

```r
# Distribution of Causal Judgment
ggplot(data2, aes(y = CausalJudgment, fill = factor(1))) +
  scale_fill_okabe_ito(order = 5) +
  stat_halfeye(adjust = 0.9, justification = -0.2,
               .width = 0, point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 1, size=2) +
  labs(x = "", y = "Causal Judgment (0-100)") +
  coord_flip() +
  guides(fill = guide_legend(title = "")) +
  theme_bw() +
  theme(legend.position = "none")
```



Now, let's focus on comparing Causal judgment scores between the two groups that completed the task in a null contingency condition. There are 20 participants in each group (NL = Spanish, FL = English).

```
# Number of participants in each  group
length(data2$CausalJudgment[data2$experimentLanguage == "Foreign"])
```

```
[1] 20
```

```r
length(data2$CausalJudgment[data2$experimentLanguage == "Native"])
```

```
[1] 20
```

In the NL condition, the distribution has a mean of 63 (M = 63, SD = 20.49), and the median of 62.5 is close to the mean. The first and third quartiles are somewhat asymmetric (56 and 63). The distribution has a negative skew of -1.23 and is leptokurtic (1.96).
In the FL condition, the distribution has a mean of 44 (M = 44.30, SD = 26.18), with a median of 49.5, which is not close to the mean. The first and third quartiles are symmetric around the median (35 and 65), indicating a fairly symmetric distribution. The distribution shows a slight negative skew of -0.69 and is platykurtic (kurtosis=-1.02).

```r
# Descriptive statistics
aggregate(x = data2$CausalJudgment, by = list(data2$experimentLanguage), FUN = summary)
```

```
  Group.1 x.Min. x.1st Qu. x.Median x.Mean x.3rd Qu. x.Max.
1 Foreign   0.00     35.25    49.50  44.30     65.00  75.00
2  Native   0.00     56.00    62.50  63.00     79.25  90.00
```

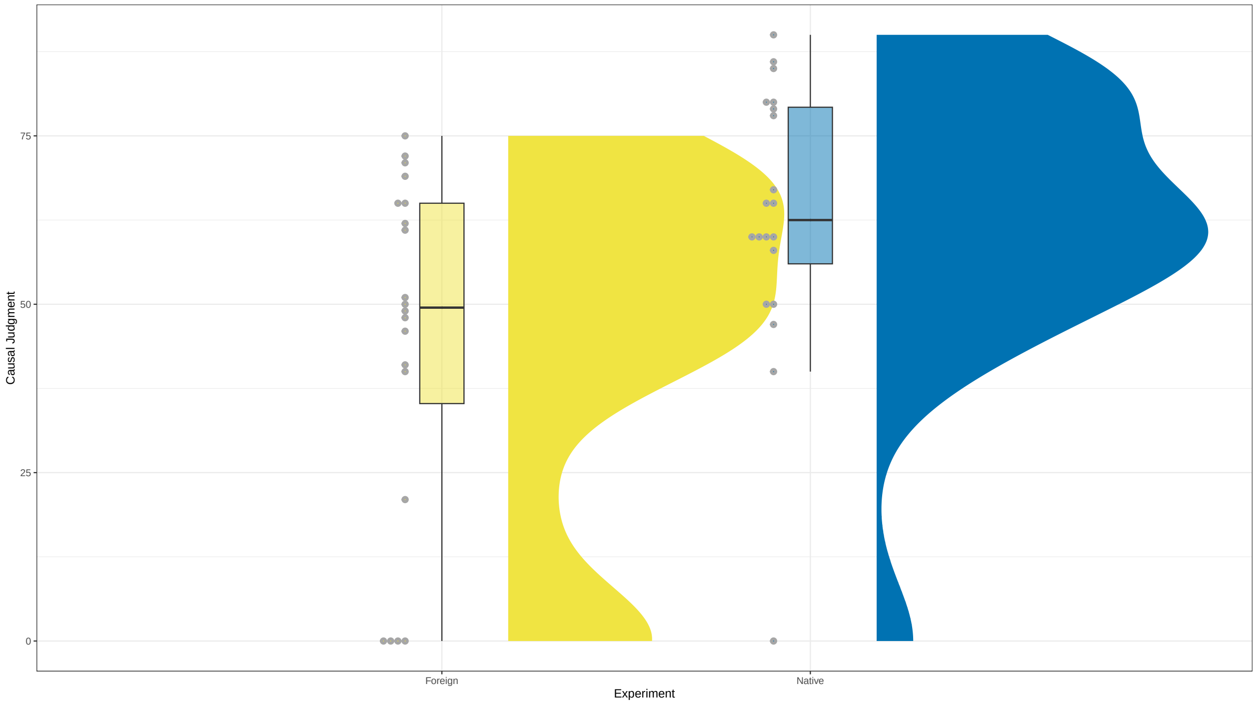
```r
round(stat.desc(data2$CausalJudgment[data2$experimentLanguage == "Foreign"],
                norm = T), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       20.00         4.00         0.00         0.00        75.00        75.00
         sum       median         mean      SE.mean CI.mean.0.95          var
      886.00        49.50        44.30         5.85        12.25       685.27
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
       26.18         0.59        -0.69        -0.67        -1.02        -0.51
  normtest.W   normtest.p
        0.85         0.01
```

```r
round(stat.desc(data2$CausalJudgment[data2$experimentLanguage == "Native"],
                norm = T), 2)
```

```
     nbr.val     nbr.null       nbr.na          min          max        range
       20.00         1.00         0.00         0.00        90.00        90.00
         sum       median         mean      SE.mean CI.mean.0.95          var
     1260.00        62.50        63.00         4.58         9.59       419.89
     std.dev     coef.var     skewness     skew.2SE     kurtosis     kurt.2SE
       20.49         0.33        -1.23        -1.20         1.96         0.99
  normtest.W   normtest.p
        0.88         0.02
```

```r
# Visualization
ggplot(data2, aes(y = CausalJudgment, x = experimentLanguage,
                  fill = experimentLanguage)) +
  stat_halfeye(adjust = 0.9, justification = -0.2, .width = 0,
               point_colour = NA) +
  geom_boxplot(width = 0.12, outlier.color = NA, alpha = 0.5) +
  stat_dots(side = "left", justification = 1.1, binwidth = 1, size=2) +
  scale_fill_okabe_ito(order = c(4, 5)) +
  labs(x = "Experiment", y = "Causal Judgment") +
  theme_bw() +
  theme(legend.position = "none")
```



3.2 Effect size

The standardized effect size is 0.8, which is considered a large effect size.

```
library(effectsize)
cohens_d(x = data2$CausalJudgment[data2$experimentLanguage == "Foreign"],
         y = data2$CausalJudgment[data2$experimentLanguage == "Native"])
```

```
Cohen's d |         95% CI
--------------------------
-0.80     | [-1.44, -0.15]

- Estimated using pooled SD.
```

```
# Cliff's Delta
cliffs_delta(x = data2$CausalJudgment[data2$experimentLanguage == "Foreign"],
             y = data2$CausalJudgment[data2$experimentLanguage == "Native"])
```

```
r (rank biserial) |         95% CI
----------------------------------
-0.41             | [-0.66, -0.08]
```

Next, we examine the Bayes Factor. The data are 3 times more likely under the alternative hypothesis (H1) than under the null hypothesis (H0), providing moderate evidence for the alternative hypothesis.

```
# BF
ttestBF(x = data2$CausalJudgment[data2$experimentLanguage == "Foreign"],
        y = data2$CausalJudgment[data2$experimentLanguage == "Native"])
```

```
Bayes factor analysis
--------------
[1] Alt., r=0.707 : 3.433997 ±0.01%

Against denominator:
  Null, mu1-mu2 = 0
---
Bayes factor type: BFindepSample, JZS
```

Arnold, J. (2024). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. https://CRAN.R-project.org/package=ggthemes

Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics*. https://CRAN.R-project.org/package=gridExtra

Barrett, M. (2024). *Ggokabeito: 'Okabe-ito' scales for 'ggplot2' and 'ggraph'*. https://github.com/malcolmbarrett/ggokabeito

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). Effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software*, *5*(56), 2815. https://doi.org/10.21105/joss.02815

Cui, B. (2024). *DataExplorer: Automate data exploration and treatment*. https://CRAN.R-project.org/package=DataExplorer

Díaz-Lago, M., & Matute, H. (2019). Thinking in a foreign language reduces the causality bias. *Quarterly Journal of Experimental Psychology*, *72*(1), 41–51. https://doi.org/10.1177/1747021818755326

Fox, J., & Weisberg, S. (2019). *An r companion to applied regression, third edition*. Sage. https://www.john-fox.ca/Companion/

Grosjean, P., & Ibanez, F. (2024). *Pastecs: Package for analysis of space-time ecological series*. https://CRAN.R-project.org/package=pastecs

Kay, M. (2024). Ggdist: Visualizations of distributions and uncertainty in the grammar of graphics. *IEEE Transactions on Visualization and Computer Graphics*, *30*(1), 414–424. https://doi.org/10.1109/TVCG.2023.3327195

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174.

Masked Citation. (n.d.). *Masked Title*.

Morey, R., & Rouder, J. (2024). *BayesFactor: Computation of bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Pastore, M., Di Loro, P. A., Mingione, M., & Calcagni', A. (2022). *Overlapping: Estimation of overlapping in empirical distributions*. https://CRAN.R-project.org/package=overlapping

Schloerke, B., Cook, D., Larmarange, J., Briatte, F., Marbach, M., Thoen, E., Elberg, A., & Crowley, J. (2024). *GGally: Extension to 'ggplot2'*. https://CRAN.R-project.org/package=GGally

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.