# Analysis Script Pilot Indirect Source v1

Hening Wang

2024-11-18

Supress messages in this document so that the output is cleaner

Import packages.

```
rm(list = ls())

options(warn = -1)
library(tidyverse)
library(lme4)
library(lmerTest)
library(ggplot2)
library(aida)
```

Set up theme for ggplot.

```
theme_set(theme_aida())

##################################################
## CSP-colors
##################################################
CSP_colors = c(
  "#7581B3", "#99C2C2", "#C65353", "#E2BA78", "#5C7457", "#575463",
  "#B0B7D4", "#66A3A3", "#DB9494", "#D49735", "#9BB096", "#D4D3D9",
  "#414C76", "#993333"
  )
# setting theme colors globally
scale_colour_discrete <- function(...) {
  scale_colour_manual(..., values = CSP_colors)
}
scale_fill_discrete <- function(...) {
  scale_fill_manual(..., values = CSP_colors)
}
```

## Data Import & Preprocessing

Import data

```
data <- read.csv("../data/pilot-v1/results.csv")
```

Factorise variables: id, item, informationSource, listenerRole

```
data$id <- as.factor(1:nrow(data))
data$item <- as.factor(data$condition)
data$informationSource <- as.factor(data$informationSource)
data$listenerRole <- as.factor(data$listenerRole)
```

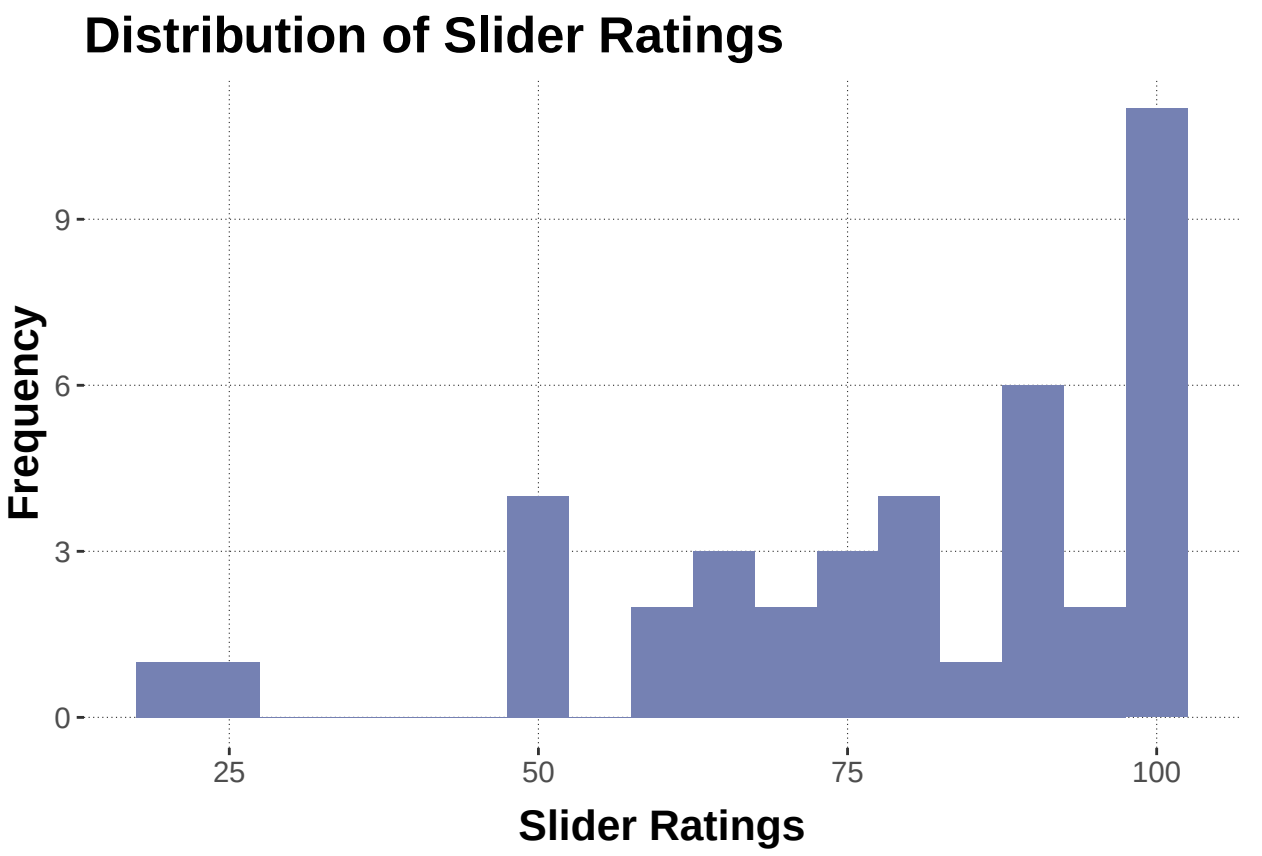
Recruitment criteria: 1. English native speakers 2. Approval rate > 90% 3. Completion > 5 times 4. Not included in the prior studies of CommuniCause

# Plots

### A histogram showing the distribution of probs.

**Interpretation: Data is left-skewed.**

```
ggplot(data, aes(x = probs)) +
  geom_histogram(binwidth = 5, fill = CSP_colors[1]) +
  labs(title = "Distribution of Slider Ratings",
       x = "Slider Ratings",
       y = "Frequency")
```




### Plot the dependent variable "probs" against conditions.

**Interpretation:**

1. No significant difference between conditions.
2. Perhaps due to:

- not enough sample size.
- the empirical distribution is severe skewed. This is problem of slider rating in general.
- manipulation too subtle.

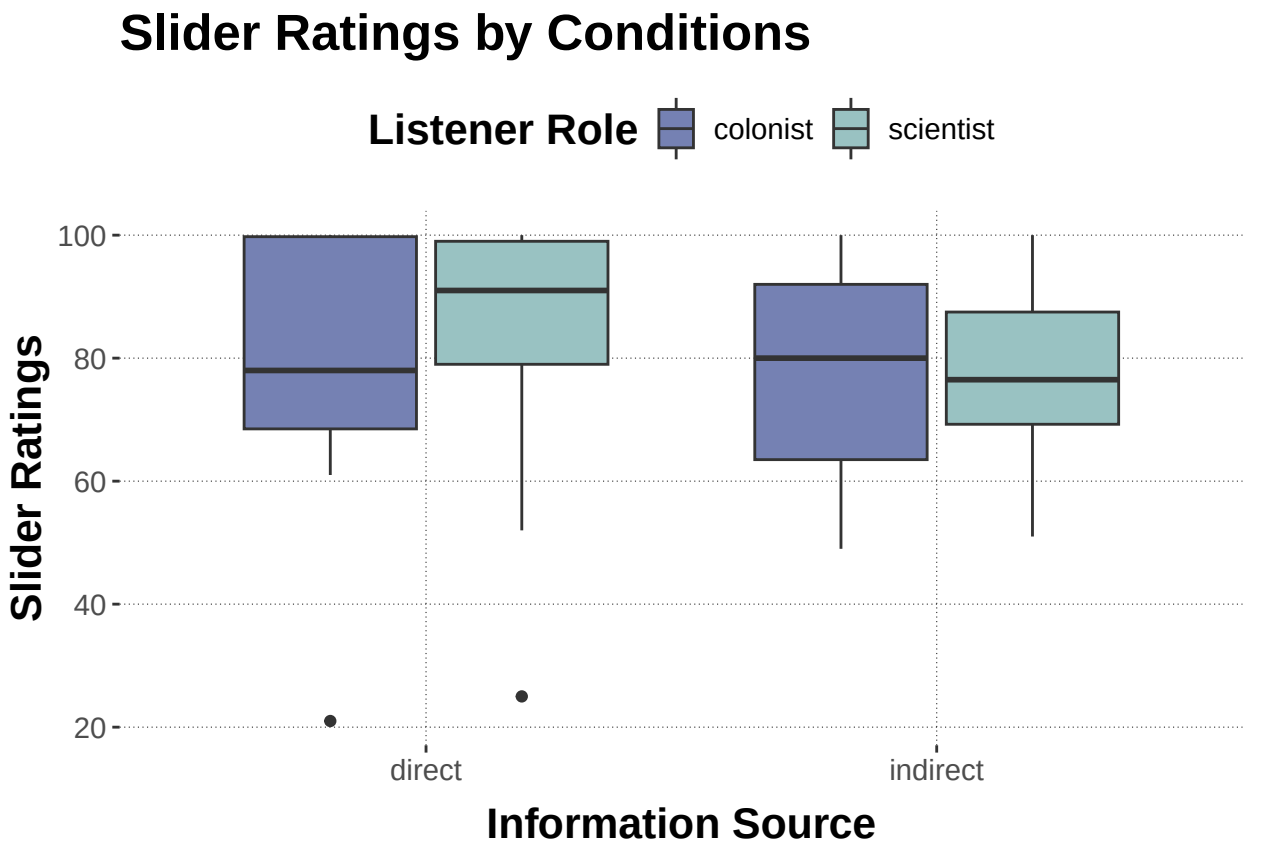
3. We can try to observe some trends though:

- "Direct" information seems to have a higher mean rating than "indirect", as expected.

- The role "scientist" seems to be more sensitive to indirectness of information. This could lead to a potential interaction as expected.

Next step: It seems that manipulation did have some effects given 3.2, but it was too subtle to be statistically significant. We should probably rework the design of the study by addressing the three points in 2.

```
ggplot(data, aes(x = informationSource, y = probs, fill = listenerRole)) +
  geom_boxplot() +
  labs(title = "Slider Ratings by Conditions",
       x = "Information Source",
       y = "Slider Ratings",
       fill = "Listener Role")
```

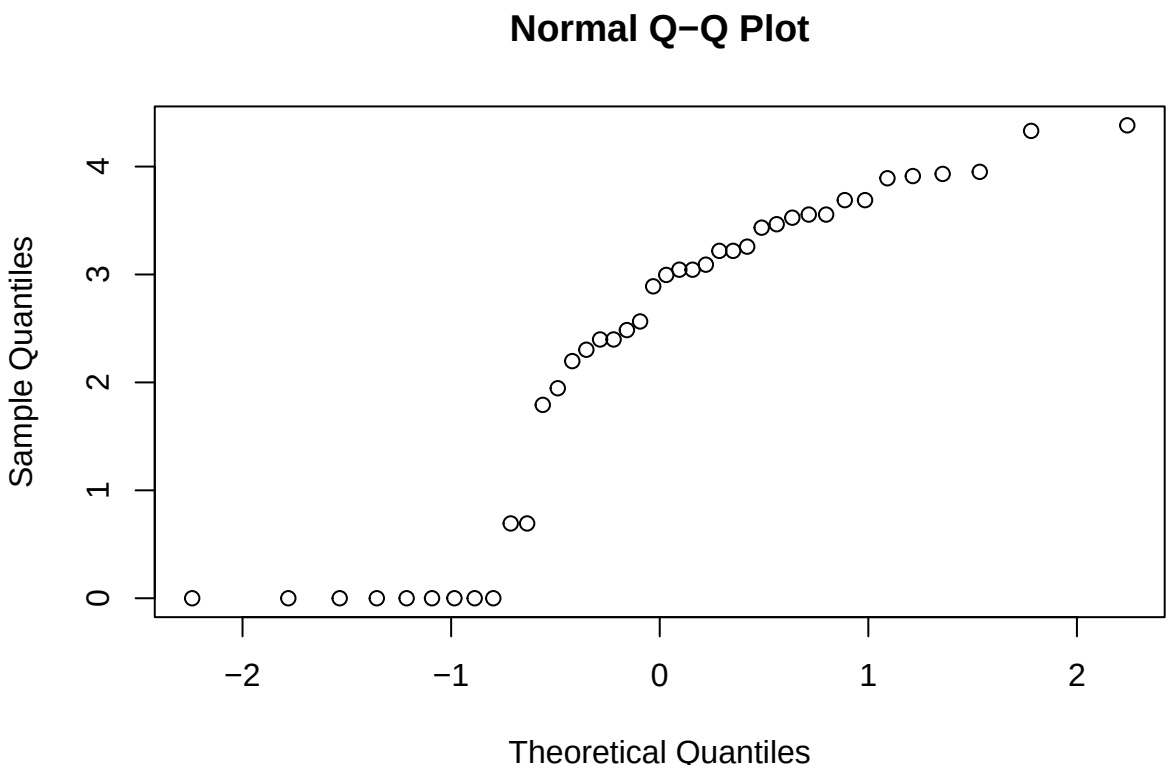



### Plot transformed data

Given that data is left-skewed. We perform a log transformation, and show a QQ plot after data transformation.

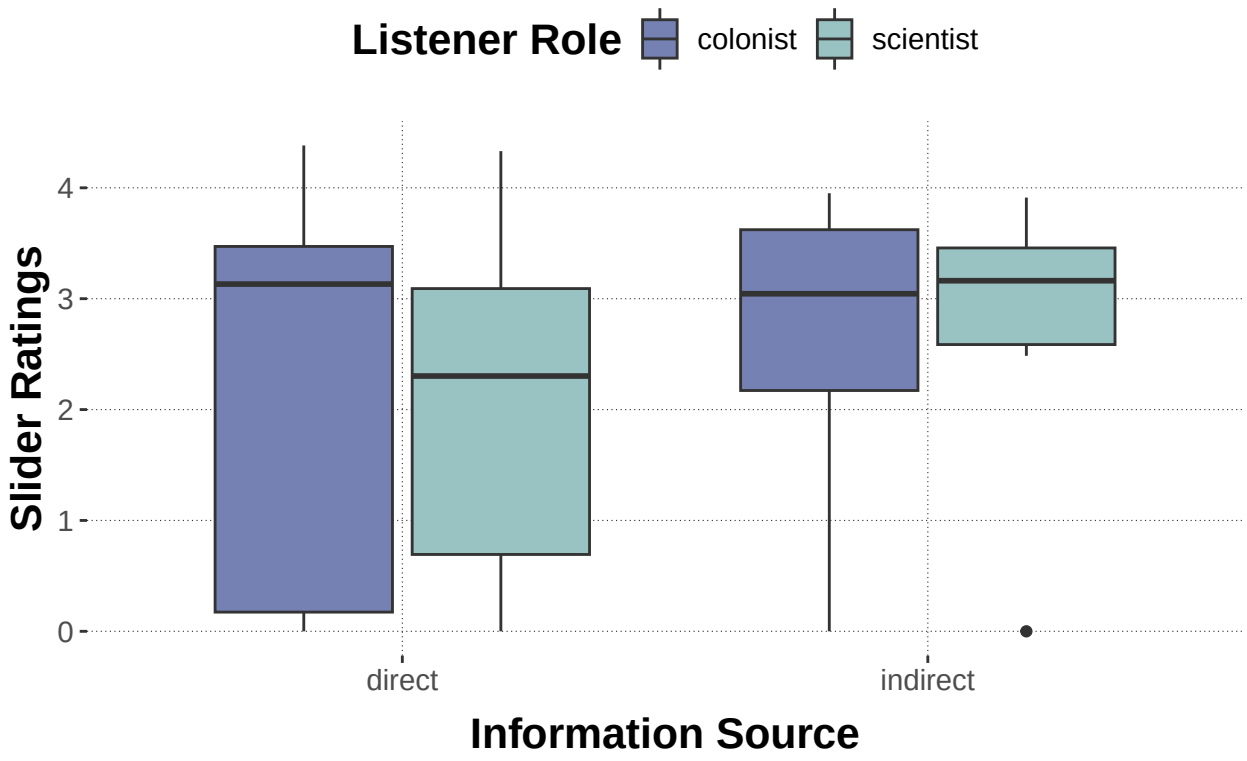
It does not really help. Now the data is right-skewed.

```
data$transformed_probs <- log(max(data$probs+1) - data$probs)
qqnorm(data$transformed_probs)
```




Another empirical plot with transformed data. Nothing really interesting here.

```
ggplot(data, aes(x = informationSource, y = transformed_probs, fill = listenerRole)) +
  geom_boxplot() +
  labs(title = "log transformed Slider Ratings by Conditions",
       x = "Information Source",
       y = "Slider Ratings",
       fill = "Listener Role")
```




## Statistical Analysis

**Note: Nothing is significant here.**

Fit a linear model with full interaction.

```
model_full <- lm(probs ~ informationSource * listenerRole, data = data)
summary(model_full)
```

```
##
## Call:
## lm(formula = probs ~ informationSource * listenerRole, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.923 -11.157   4.584  17.077  23.091
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                     77.9000     6.7911  11.471
## informationSourceindirect                       -0.9909     9.3832  -0.106
## listenerRolescientist                            5.0231     9.0330   0.556
## informationSourceindirect:listenerRolescientist -4.9322    14.1557  -0.348
##                                                Pr(>|t|)
## (Intercept)                                     1.4e-13 ***
## informationSourceindirect                         0.916
## listenerRolescientist                             0.582
```

```
## informationSourceindirect:listenerRolescientist    0.730
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.48 on 36 degrees of freedom
## Multiple R-squared:  0.0168, Adjusted R-squared:  -0.06514
## F-statistic: 0.205 on 3 and 36 DF,  p-value: 0.8923
```

Fit another lm with transformed data.

```
model_full_transformed <- lm(transformed_probs ~ informationSource * listenerRole, data = data)
summary(model_full_transformed)
```

```
##
## Call:
## lm(formula = transformed_probs ~ informationSource * listenerRole,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.6978 -1.3973  0.4047  1.0474  2.2702
##
## Coefficients:
##                                                Estimate Std. Error t value
## (Intercept)                                      2.1802     0.4878   4.469
## informationSourceindirect                        0.4352     0.6740   0.646
## listenerRolescientist                           -0.1196     0.6489  -0.184
## informationSourceindirect:listenerRolescientist  0.2020     1.0169   0.199
##                                                Pr(>|t|)
## (Intercept)                                    7.51e-05 ***
## informationSourceindirect                         0.523
## listenerRolescientist                             0.855
## informationSourceindirect:listenerRolescientist   0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.543 on 36 degrees of freedom
## Multiple R-squared:  0.03245,    Adjusted R-squared:  -0.04818
## F-statistic: 0.4024 on 3 and 36 DF,  p-value: 0.7521
```