Fisher's MD

```
data <- readr::read_csv("Final_Main_Coder.csv")
```

```
## Rows: 394 Columns: 3-- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): Video ID, Type of Video, Risk of Bullshit
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
contingency_table <- table(data$`Type of Video`, data$`Risk of Bullshit`)
fisher_result <- fisher.test(contingency_table, simulate.p.value = TRUE)
p_value <- fisher_result$p.value
cat("p-value:", p_value, "\n")
```

```
## p-value: 0.0004997501
```

```
# Extract observed values
observed_values <- contingency_table
print(observed_values)
```

```
##
##                      Bullshit High Low Reducing Risk
##   Advertisement             0    7  31             0
##   Cautionary                1    0  11             0
##   Education                 6   67 117            16
##   Entertainment             7    1  24             0
##   Events                    0    2  36             4
##   Experiment                3    4  38             4
##   Experiment Illegal        0    0   4             0
##   Hype                      0    1  10             0
```

```
# Perform chi-square test to calculate expected values
chi_result <- chisq.test(contingency_table)
```

```
## Warning in chisq.test(contingency_table): Chi-squared approximation may be
## incorrect
```

```
expected_values <- chi_result$expected

print(expected_values)
```

```
##
##                       Bullshit       High        Low Reducing Risk
##   Advertisement      1.6395939  7.9086294  26.137056     2.3147208
##   Cautionary         0.5177665  2.4974619   8.253807     0.7309645
##   Education          8.8883249 42.8730964 141.690355    12.5482234
##   Entertainment      1.3807107  6.6598985  22.010152     1.9492386
##   Events             1.8121827  8.7411168  28.888325     2.5583756
##   Experiment         2.1142132 10.1979695  33.703046     2.9847716
##   Experiment Illegal 0.1725888  0.8324873   2.751269     0.2436548
##   Hype               0.4746193  2.2893401   7.565990     0.6700508
```

```r
# Convert observed_values to a data frame
observed_data <- as.data.frame.table(observed_values)

# Perform chi-square test to calculate expected values
chi_result <- chisq.test(observed_values)
```

```
## Warning in chisq.test(observed_values): Chi-squared approximation may be
## incorrect
```

```r
expected_values <- chi_result$expected

# Convert expected_values to a data frame
expected_data <- as.data.frame.table(expected_values)

# Rename the columns
names(observed_data) <- c("Type_of_Video", "Risk_of_Bullshit", "Observed")
names(expected_data) <- c("Type_of_Video", "Risk_of_Bullshit", "Expected")

# Reorder the factor levels
observed_data$Type_of_Video <- factor(observed_data$Type_of_Video, levels = c("Advertisement", "Educati
observed_data$Risk_of_Bullshit <- factor(observed_data$Risk_of_Bullshit, levels = c("Reducing Risk", "L
expected_data$Type_of_Video <- factor(expected_data$Type_of_Video, levels = c("Advertisement", "Educati
expected_data$Risk_of_Bullshit <- factor(expected_data$Risk_of_Bullshit, levels = c("Reducing Risk", "L
```

```r
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

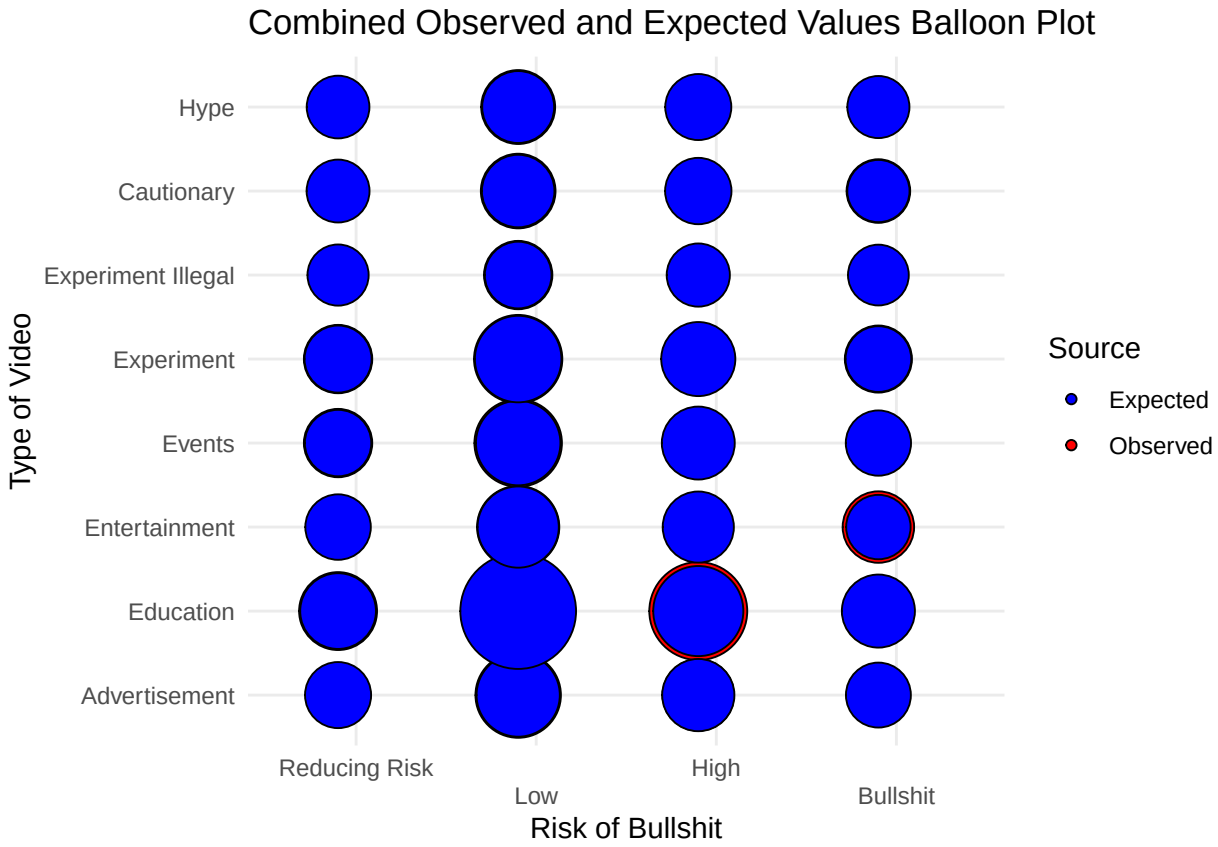
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# Combine observed and expected data
combined_data <- bind_rows(
  observed_data %>% mutate(Source = "Observed"),
  expected_data %>% mutate(Source = "Expected")
)

# Reorder levels for better visualization
combined_data$Type_of_Video <- factor(
  combined_data$Type_of_Video,
  levels = c(
    "Advertisement", "Education", "Entertainment", "Events",
    "Experiment", "Experiment Illegal", "Cautionary", "Hype"
  )
)
combined_data$Risk_of_Bullshit <- factor(
  combined_data$Risk_of_Bullshit,
  levels = c("Reducing Risk", "Low", "High", "Bullshit")
)
```

```r
# Balloon plot for combined observed and expected values
ggplot(combined_data, aes(x = Risk_of_Bullshit, y = Type_of_Video)) +
  geom_point(aes(size = ifelse(Source == "Observed", Observed, Expected),
                 fill = Source), shape = 21, position = position_nudge(x = -0.1)) +
  labs(title = "Combined Observed and Expected Values Balloon Plot",
       x = "Risk of Bullshit", y = "Type of Video") +
  theme_minimal() +
  scale_size(range = c(10, 20)) +  # Increase bubble sizes
  scale_x_discrete(guide = guide_axis(n.dodge = 2)) +
  guides(size = FALSE, fill = guide_legend(title = "Source")) +
  scale_fill_manual(values = c("blue", "red"),  # Set colors for observed and expected
                    guide = guide_legend(override.aes = list(size = 4, alpha = 0.5)))  # Decrease darkn
```

```
## Warning: The `<scale>` argument of `guides()` cannot be `FALSE`. Use "none"
## instead as of ggplot2 3.3.4.
```




```r
# Balloon plot for combined observed and expected values
ggplot(combined_data, aes(x = factor(Risk_of_Bullshit), y = Type_of_Video)) +
  geom_point(aes(size = ifelse(Source == "Observed", Observed, Expected),
                 fill = Source), shape = 21, position = position_dodge(width = 0.75)) +
  labs(title = "Combined Observed and Expected Values Balloon Plot",
       x = "Risk of Bullshit", y = "Type of Video") +
  theme_minimal() +
  scale_size(range = c(5, 15)) +
  scale_x_discrete(
    breaks = levels(combined_data$Risk_of_Bullshit),
    labels = levels(combined_data$Risk_of_Bullshit),
```

```
  expand = expansion(mult = c(0.05, 0.05))
) +
guides(size = FALSE, fill = guide_legend(title = "Source")) +
scale_fill_manual(values = c("blue", "red")) +
coord_cartesian(clip = "off")
```