DATA-2010_Project

Jillian Manguba, Thien Pham, Kelly Villamayor, Abhinav Jain

2024-02-12

DATA 2010 Project

```
library(ggplot2)
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 4.3.2
```

```
nv_tourist <- read.csv("non-verbal tourist data.csv")
```

Note: Original dataset switched TAudio1 and Proxemics new_nv_tourist rectified this error

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.1     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::combine() masks gridExtra::combine()
## x dplyr::filter()  masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
new_nv_tourist <- nv_tourist |>
  rename(
    handshake_indiff = GImg1,
    hug_indiff = GImg2,
    kiss_indiff = GImg3,
    consent_post = PImg1,
    interest_post = PImg2,
    neutral_post = PImg3,
    reflexive_post = PImg4,
    negative_post = PImg5,
    relax_score = Tense...relaxed,
    anarchic_score = Authoritative..anarchic,
    friendly_score = Hostile...friendly,
    authoritative = Proxemics,
```

```
    sarcastic = TAudio2,
    friendly = TAudio3,
    spitting = QAudio1,
    hum = QAudio2,
    sigh = QAudio3,
    proxemic = TAudio1,
    client_type = Type.of.Client
  )
```

```
# checking for missing values...

# the dataset uses "?" to indicate a missing value
contains_missing <- function(row) {
  any(row == "?") || any(is.na(row))
}

rows_w_missing_val <- apply(new_nv_tourist, 1, contains_missing)
print(new_nv_tourist[rows_w_missing_val,])
```

```
##    sex age country returning handshake_indiff hug_indiff kiss_indiff
## 2    M  60  brasil        no            likes indiferent  indiferent
## 39   F  49  canada       yes            likes      likes  indiferent
## 44   F  40 england       yes            likes      likes       likes
## 72   M  39 germany        no         dislikes   dislikes           ?
##    consent_post interest_post neutral_post reflexive_post negative_post
## 2         likes         likes        likes          likes             ?
## 39        likes         likes        likes          likes             ?
## 44        likes         likes        likes          likes      dislikes
## 72        likes         likes     dislikes       dislikes      dislikes
##    relax_score anarchic_score friendly_score proxemic sarcastic friendly
## 2            2              9              1        B  dislikes dislikes
## 39           2              2              2        B  dislikes dislikes
## 44           1              ?              1        A  dislikes dislikes
## 72           5              4              4        C  dislikes dislikes
##      spitting        hum     sigh authoritative client_type
## 2       likes      likes dislikes      dislikes           0
## 39      likes   dislikes dislikes      dislikes           2
## 44      likes indiferent dislikes      dislikes           2
## 72 indiferent   dislikes dislikes      dislikes           5
```

Based off the results, it seems the missing values fall under MAR (missing at random). We decided to simply delete the rows that contained any missing values, since it is just 4 rows out of the total 73. Thus, it meets the 5-10% range for deletion in MAR.

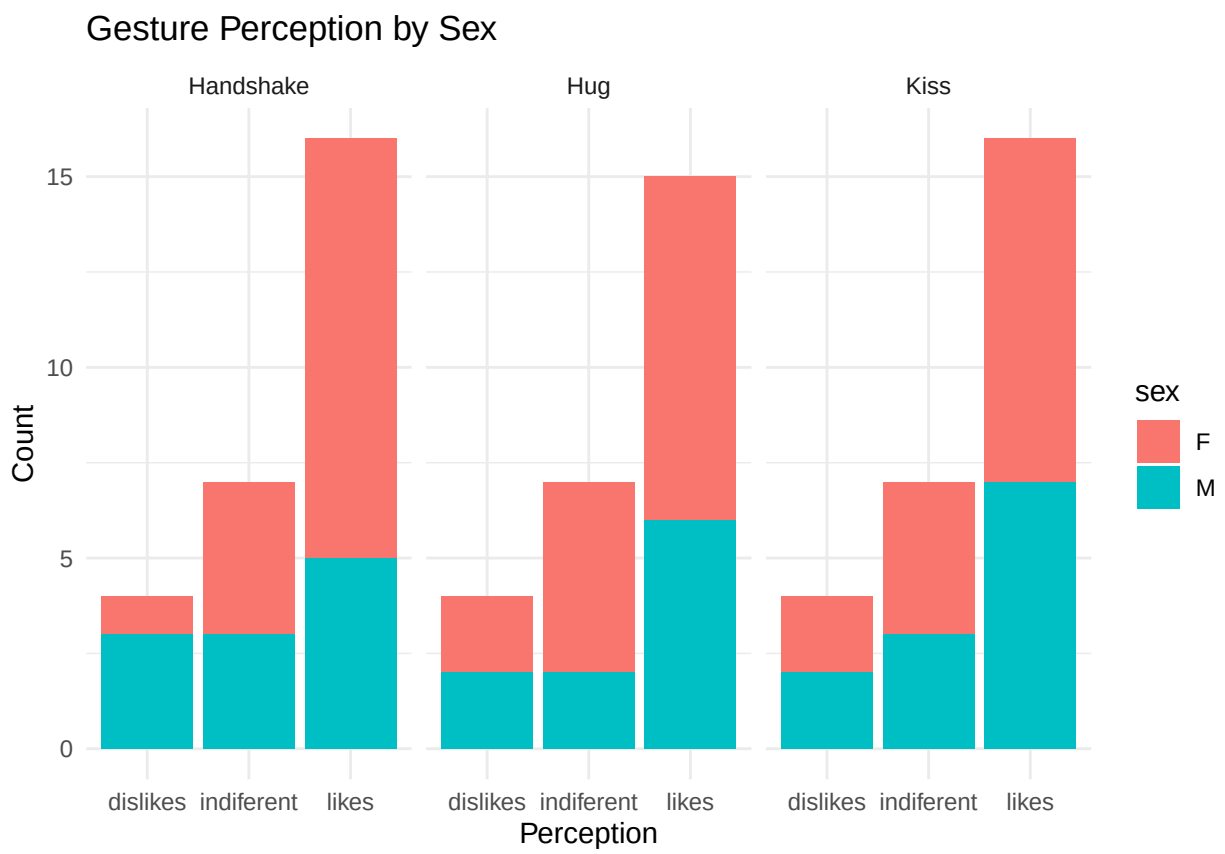
```
new_nv_tourist <- new_nv_tourist[!rows_w_missing_val,]
```

```
new_nv_tourist$anarchic_score <- as.numeric(new_nv_tourist$anarchic_score)
```

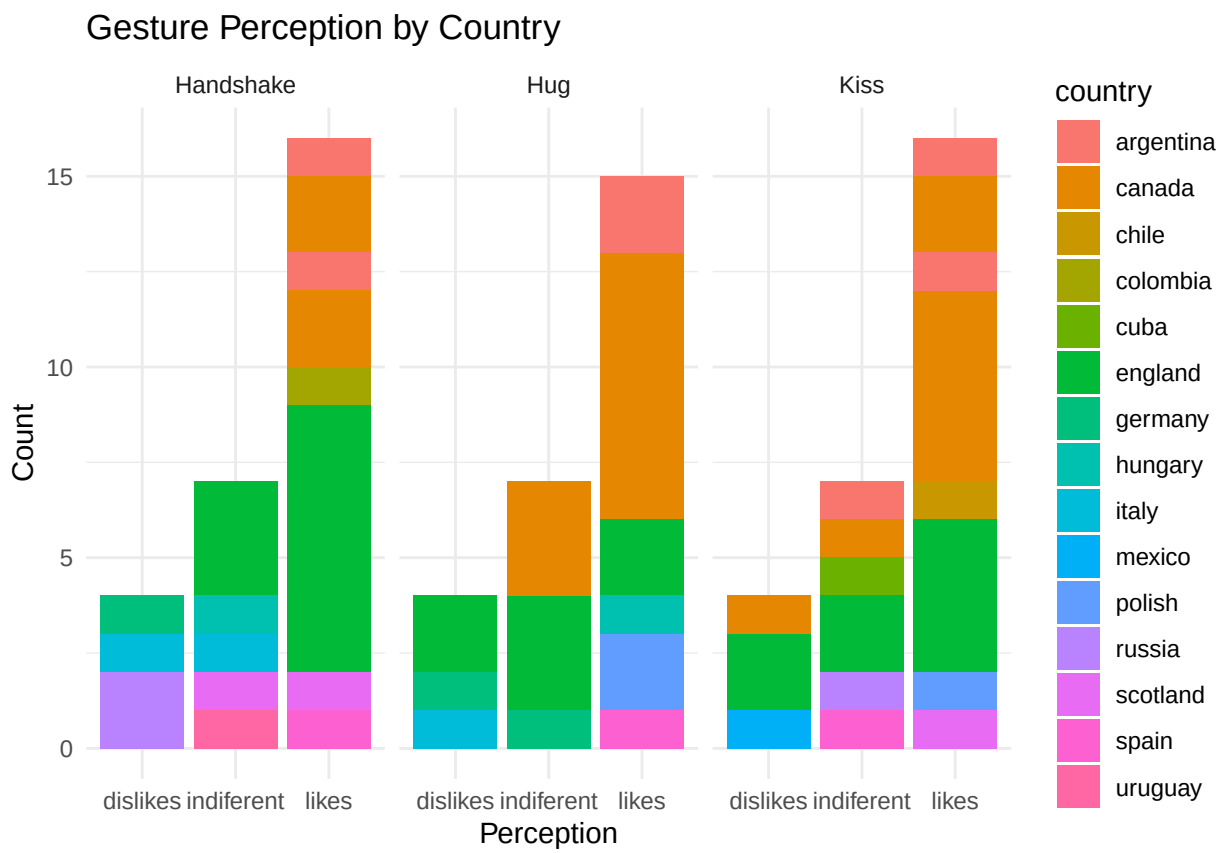
Data Exploration

Gesture

```
# this barplot is acting weird with the stacks...

# GRAPH 1
new_nv_tourist |>
  ggplot() +
  geom_bar(aes(x = handshake_indiff, fill = sex), position = "stack") +
  geom_bar(aes(x = hug_indiff, fill = sex), position = "stack") +
  geom_bar(aes(x = kiss_indiff, fill = sex), position = "stack") +
  labs(x = "Perception", y = "Count", title = "Gesture Perception by Sex") +
  facet_wrap(~ c("Handshake", "Hug", "Kiss"), nrow = 1) +
  theme_minimal()
```

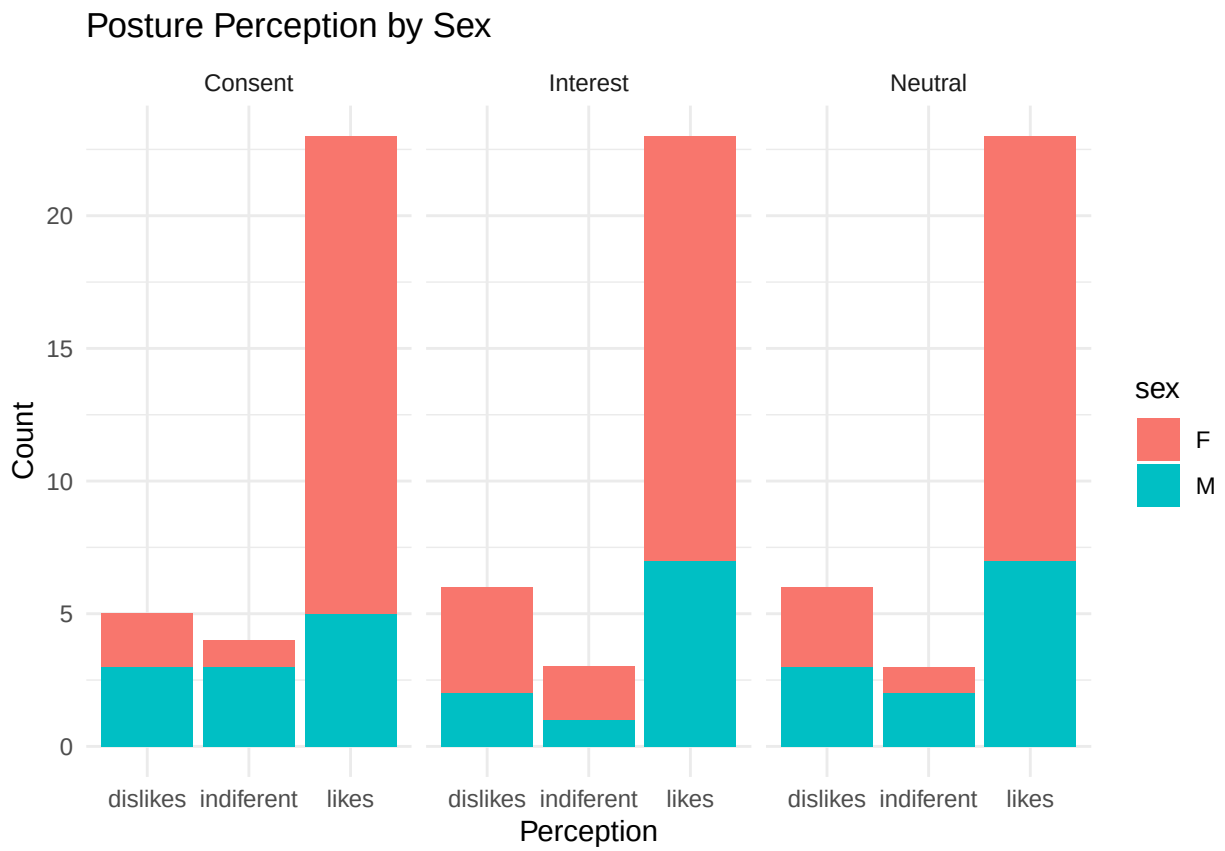



```
# GRAPH 2
new_nv_tourist |>
  ggplot() +
  geom_bar(aes(x = handshake_indiff, fill = country), position = "stack") +
  geom_bar(aes(x = hug_indiff, fill = country), position = "stack") +
  geom_bar(aes(x = kiss_indiff, fill = country), position = "stack") +
  labs(x = "Perception", y = "Count", title = "Gesture Perception by Country") +
  facet_wrap(~ c("Handshake", "Hug", "Kiss"), nrow = 1) +
  theme_minimal()
```

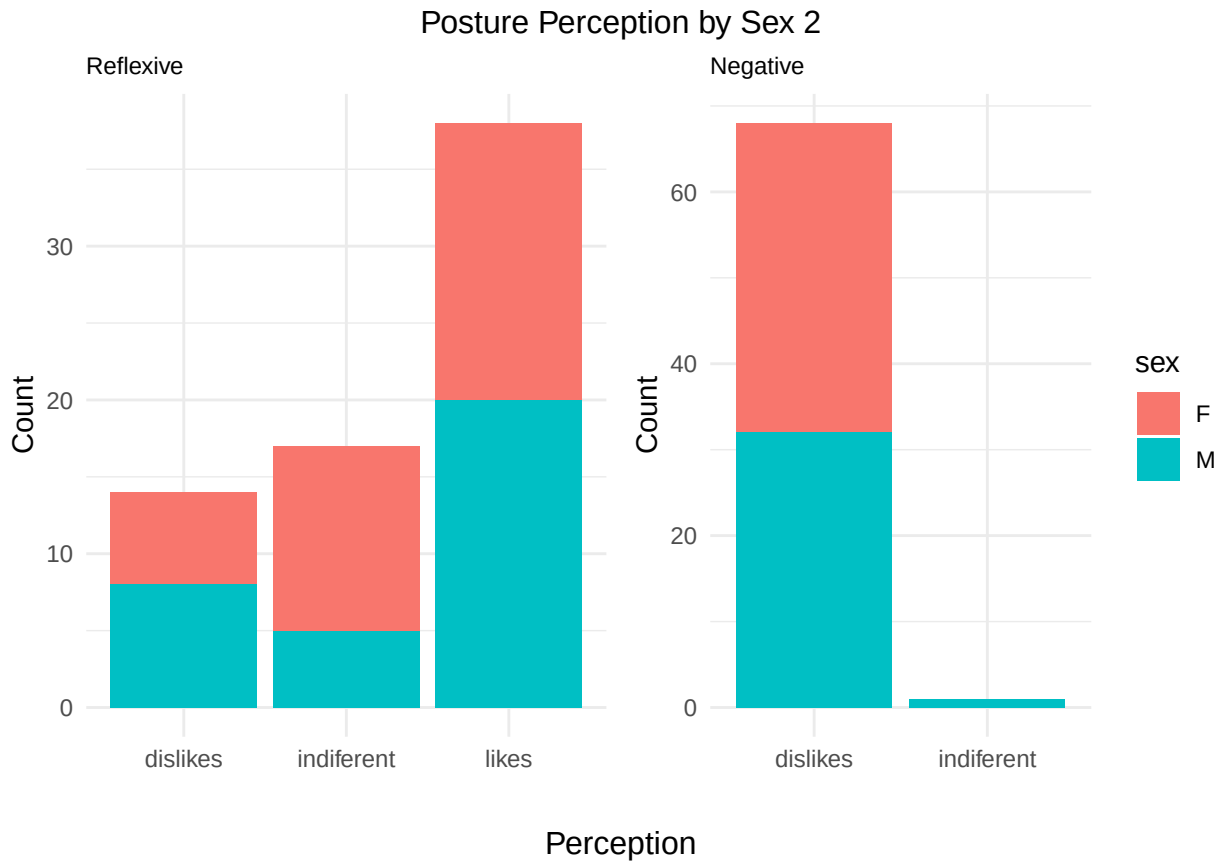


Posture

```
new_nv_tourist |>
  ggplot() +
  geom_bar(aes(x = consent_post, fill = sex), position = "stack") +
  geom_bar(aes(x = interest_post, fill = sex), position = "stack") +
  geom_bar(aes(x = neutral_post, fill = sex), position = "stack") +
  labs(x = "Perception", y = "Count", title = "Posture Perception by Sex") +
  facet_wrap(~c("Consent", "Interest", "Neutral"), nrow = 1) +
  theme_minimal()
```




```
plot1 <- ggplot(new_nv_tourist, aes(x = reflexive_post, fill = sex)) +
  geom_bar(position = "stack") +
  labs(x = "", y = "Count", title = "Reflexive") +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(size = 9))

plot2 <- ggplot(new_nv_tourist, aes(x = negative_post, fill = sex)) +
  geom_bar(position = "stack") +
  labs(x = "", y = "Count", title = "Negative") +
  theme_minimal() +
  theme(plot.title = element_text(size = 9))

grid.arrange(plot1, plot2, nrow = 1,
             top = "Posture Perception by Sex 2",
             bottom = "Perception")
```



Emotional Atmosphere

```
find_mode <- function(vector) {
  names(sort(table(as.character(vector)), decreasing =TRUE)[1])
}

new_nv_tourist |>
  group_by(sex) |>
  summarise(avg_relax = mean(relax_score),
            avg_anarchic = mean(anarchic_score),
            avg_friendly = mean(friendly_score),
            mode_relax = find_mode(relax_score),
            mode_anarchic = find_mode(anarchic_score),
            mode_friendly = find_mode(friendly_score))
```

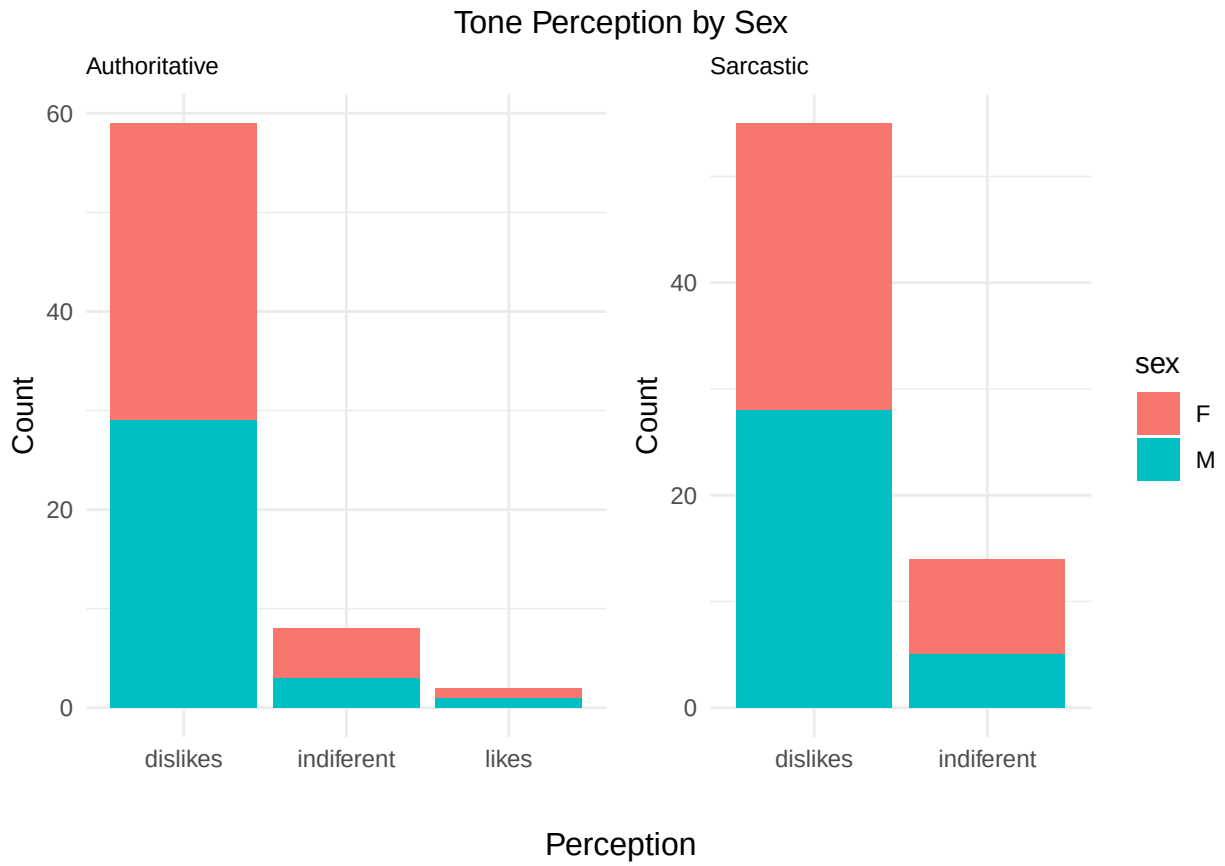
```
## # A tibble: 2 x 7
##   sex   avg_relax avg_anarchic avg_friendly mode_relax mode_anarchic
##   <chr>     <dbl>        <dbl>        <dbl> <chr>      <chr>
## 1 F          2.33         4.94         2.36 1          1
## 2 M          2.09         5.21         2.12 1          2
## # i 1 more variable: mode_friendly <chr>
```

Tone

```
plot1 <- ggplot(new_nv_tourist, aes(x = authoritative, fill = sex)) +
  geom_bar(position = "stack", legend = FALSE) +
  labs(x = "", y = "Count", title = "Authoritative") +
  theme_minimal() +
  theme(legend.position = "none",
        plot.title = element_text(size = 9))
```

```
## Warning in geom_bar(position = "stack", legend = FALSE): Ignoring unknown
## parameters: `legend`
```

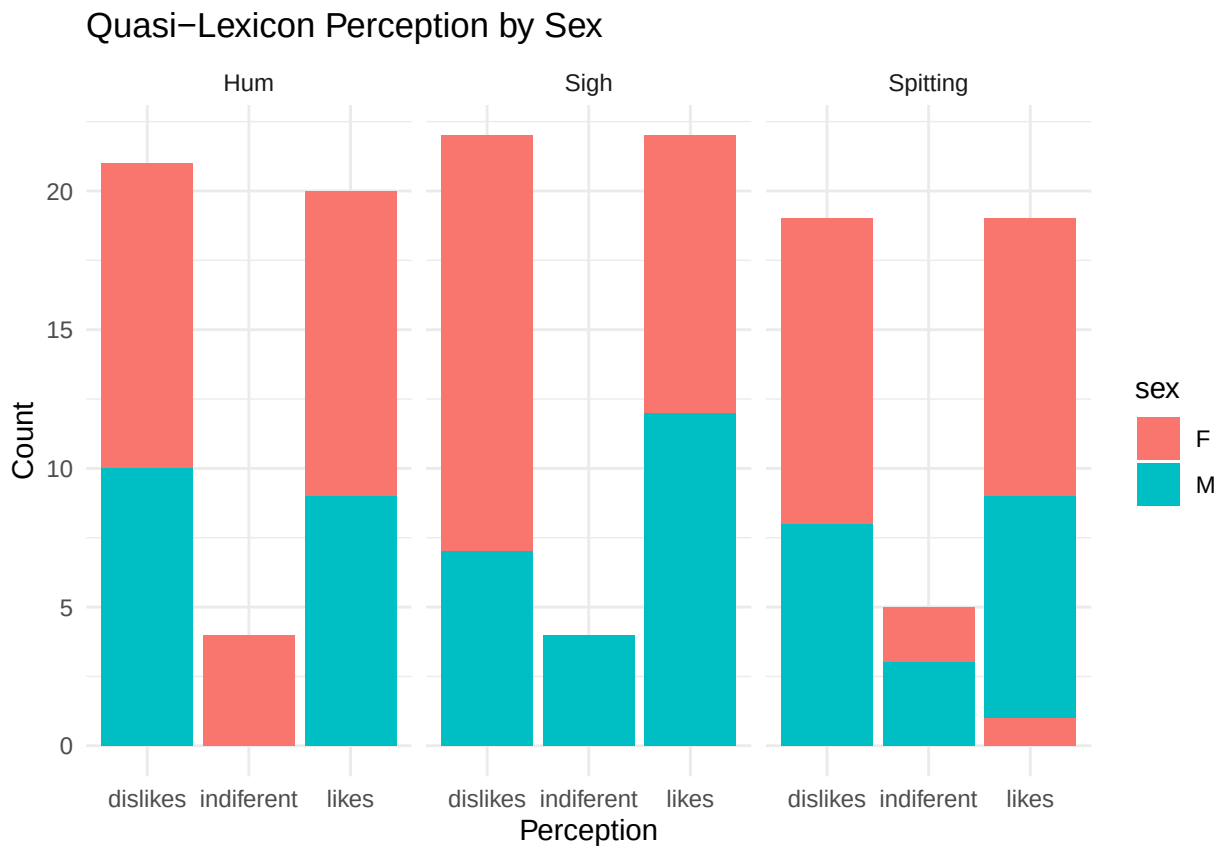
```
plot2 <- ggplot(new_nv_tourist, aes(x = sarcastic, fill = sex)) +
  geom_bar(position = "stack") +
  labs(x = "", y = "Count", title = "Sarcastic") +
  theme_minimal() +
  theme(plot.title = element_text(size = 9))

grid.arrange(plot1, plot2, nrow = 1,
             top = "Tone Perception by Sex",
             bottom = "Perception")
```



Quasi Lexicon

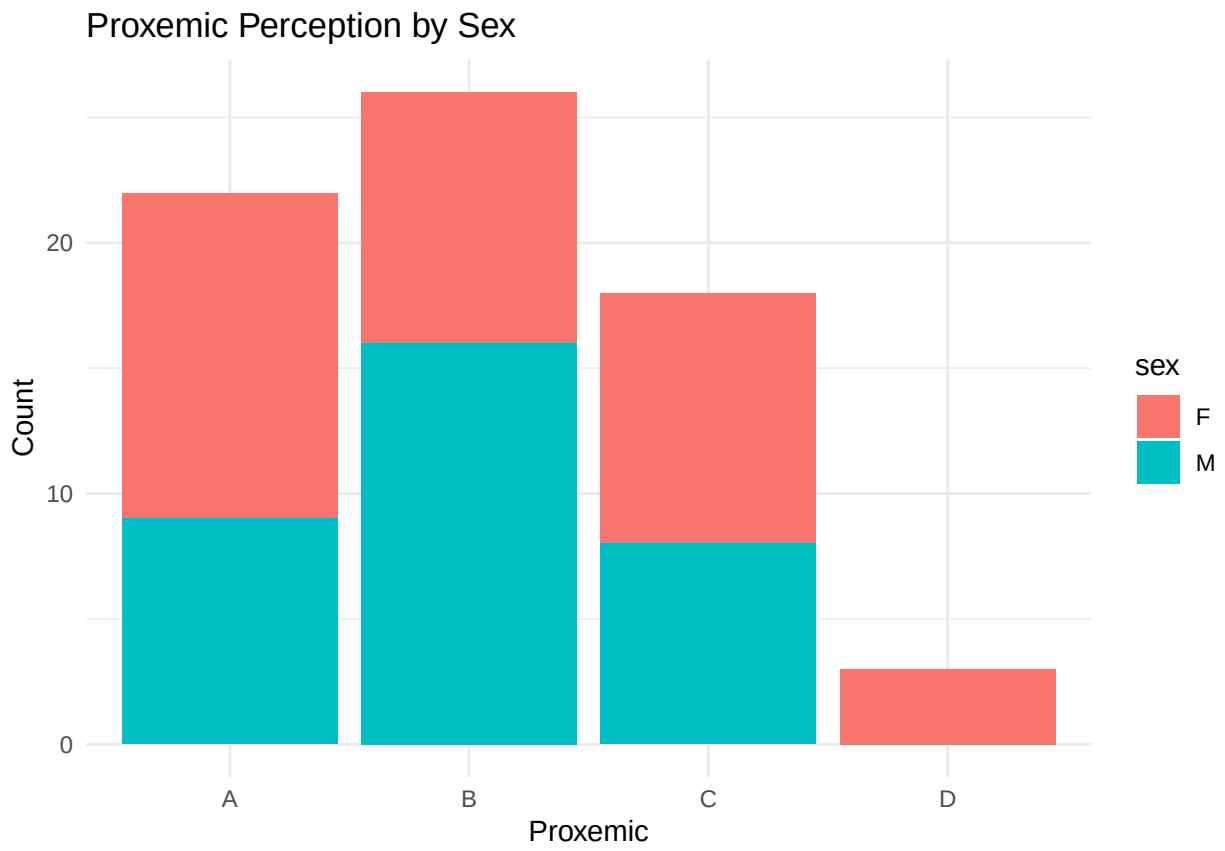
```
# there's something weird happening in this bar plot

new_nv_tourist |>
```

```
  ggplot() +
  geom_bar(aes(x = spitting, fill = sex), position = "stack") +
  geom_bar(aes(x = hum, fill = sex), position = "stack") +
  geom_bar(aes(x = sigh, fill = sex), position = "stack") +
  labs(x = "Perception", y = "Count", title = "Quasi-Lexicon Perception by Sex") +
  facet_wrap(~ c("Spitting", "Hum", "Sigh"), nrow = 1) +
  theme_minimal()
```



Proxemic

```
new_nv_tourist |>
  ggplot() +
  geom_bar(aes(x = proxemic, fill = sex), position = "stack") +
  labs(x = "Proxemic", y = "Count", title = "Proxemic Perception by Sex") +
  theme_minimal()
```


Proxemic Perception by Sex
Count
20
10
0
A
B
C
D
Proxemic
sex
F
M