Lab 6.A

Contingency Table Analyses

AUTHOR
KOMAL BHOSLE

RStudio Link

https://posit.cloud/spaces/603138/content/10210537

Library Calls

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ───────────────────── tidyverse 2.0.0 ──
✓ dplyr     1.1.4     ✓ readr     2.1.5
✓ forcats   1.0.0     ✓ stringr   1.5.1
✓ ggplot2   3.5.1     ✓ tibble    3.2.1
✓ lubridate 1.9.3     ✓ tidyr     1.3.1
✓ purrr     1.0.2
── Conflicts ─────────────────────────────────── tidyverse_conflicts() ──
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()    masks stats::lag()
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
errors
```

```
library(gt)

library(gtsummary)

library(tidymodels)
```

```
── Attaching packages ───────────────────────────────── tidymodels 1.2.0 ──
✓ broom        1.0.6     ✓ rsample      1.2.1
✓ dials        1.2.1     ✓ tune         1.2.1
✓ infer        1.0.7     ✓ workflows    1.1.4
✓ modeldata    1.3.0     ✓ workflowsets 1.1.0
✓ parsnip      1.2.1     ✓ yardstick    1.3.1
✓ recipes      1.0.10
── Conflicts ─────────────────────────────────── tidymodels_conflicts() ──
✗ recipes::all_double()  masks gtsummary::all_double()
✗ recipes::all_factor()  masks gtsummary::all_factor()
✗ recipes::all_integer() masks gtsummary::all_integer()
✗ recipes::all_logical() masks gtsummary::all_logical()
✗ recipes::all_numeric() masks gtsummary::all_numeric()
✗ scales::discard()      masks purrr::discard()
```

✗ dplyr::filter()   masks stats::filter()
✗ recipes::fixed()   masks stringr::fixed()
✗ dplyr::lag()   masks stats::lag()
✗ yardstick::spec()   masks readr::spec()
✗ recipes::step()   masks stats::step()
• Use tidymodels_prefer() to resolve common conflicts.

```
library(openintro)
```

Loading required package: airports
Loading required package: cherryblossom
Loading required package: usdata

Attaching package: 'openintro'

The following object is masked from 'package:modeldata':

    ames

The following object is masked from 'package:gt':

    sp500

```
library(easystats)
```

Attaching packages: easystats 0.7.2 (red = needs update)
✗ bayestestR 0.13.2   ✗ correlation 0.8.5
✗ datawizard 0.11.0   ✗ effectsize 0.8.8
✗ insight   0.20.1   ✗ modelbased 0.8.8
✗ performance 0.12.0   ✗ parameters 0.22.0
✗ report   0.5.8   ✗ see   0.8.4

Restart the R-Session and update packages with `easystats::easystats_update()`.

```
library(DescTools)
```

Attaching package: 'DescTools'

The following object is masked from 'package:openintro':

    cards

```
library(gtExtras)

library(webshot2)
```

Dichotomous-Dichotomous Association

The data set selected is Malaria Vaccine trial, which has two binary variables, one is the Explanatory Variable and the other is Response Variable. EV is the Treatment Variable which has a Vaccine factor and a Placebo which not clearly dichotomous. The RV is the Outcome which is the presence or absence of disease malaria.

Contingency Table

```r
malaria_clean <-
  malaria |>
  rename(
    vaccinated = treatment,
    infected   = outcome
  ) |>
  mutate(
    vaccinated = case_match(
      .x         = vaccinated,
      "vaccine" ~ "Yes",
      "placebo" ~ "No"
    ),
    vaccinated = factor(
      x        = vaccinated,
      levels   = c("Yes", "No")
    ),
    infected = case_match(
      .x      = infected,
      "infection" ~ "Yes",
      "no infection" ~ "No"
    ),
    infected = factor(
      x       = infected,
      levels = c("Yes", "No")
    )
  ) |>
  select(
    vaccinated,
    infected
  )

malaria_clean_table <-
  malaria_clean |>
  tbl_cross(
    row      = vaccinated,
    col      = infected,
    percent = "row",
    label    = list(
      vaccinated = "Vaccination Received",
      infected   = "malaria infection"
    )
```

```
  ) |>
  bold_labels() |>
  as_gt() |>
  tab_source_note(
    source_note = md(
      "Data from the `malaria` dataset(in **openintro** package)."
    )
  ) |>
  tab_header(
    title = md(
      "**2x2 Contingency Table of Malaria Infection by Vaccination Status**"
    ),
    subtitle = md(
      "Clinical Trial with 20 Patients"
    )
  )

gtsave(
  data     = malaria_clean_table,
  filename = "malaria_clean_tbl_cross.png"
)
```

```
file:////tmp/RtmpH1OCIl/file69f6dd28d24.html screenshot completed
```

```
malaria_clean_table
```

2x2 Contingency Table of Malaria Infection by Vaccination Status

Clinical Trial with 20 Patients

	malaria infection		
	Yes	No	Total
Vaccination Received			
Yes	5 (36%)	9 (64%)	14 (100%)
No	6 (100%)	0 (0%)	6 (100%)
Total	11 (55%)	9 (45%)	20 (100%)

Data from the `malaria` dataset(in **openintro** package).

Measure(s) of Association

```
oddsratio(
  x = table(malaria_clean) + 0.5
```

```
  )
```
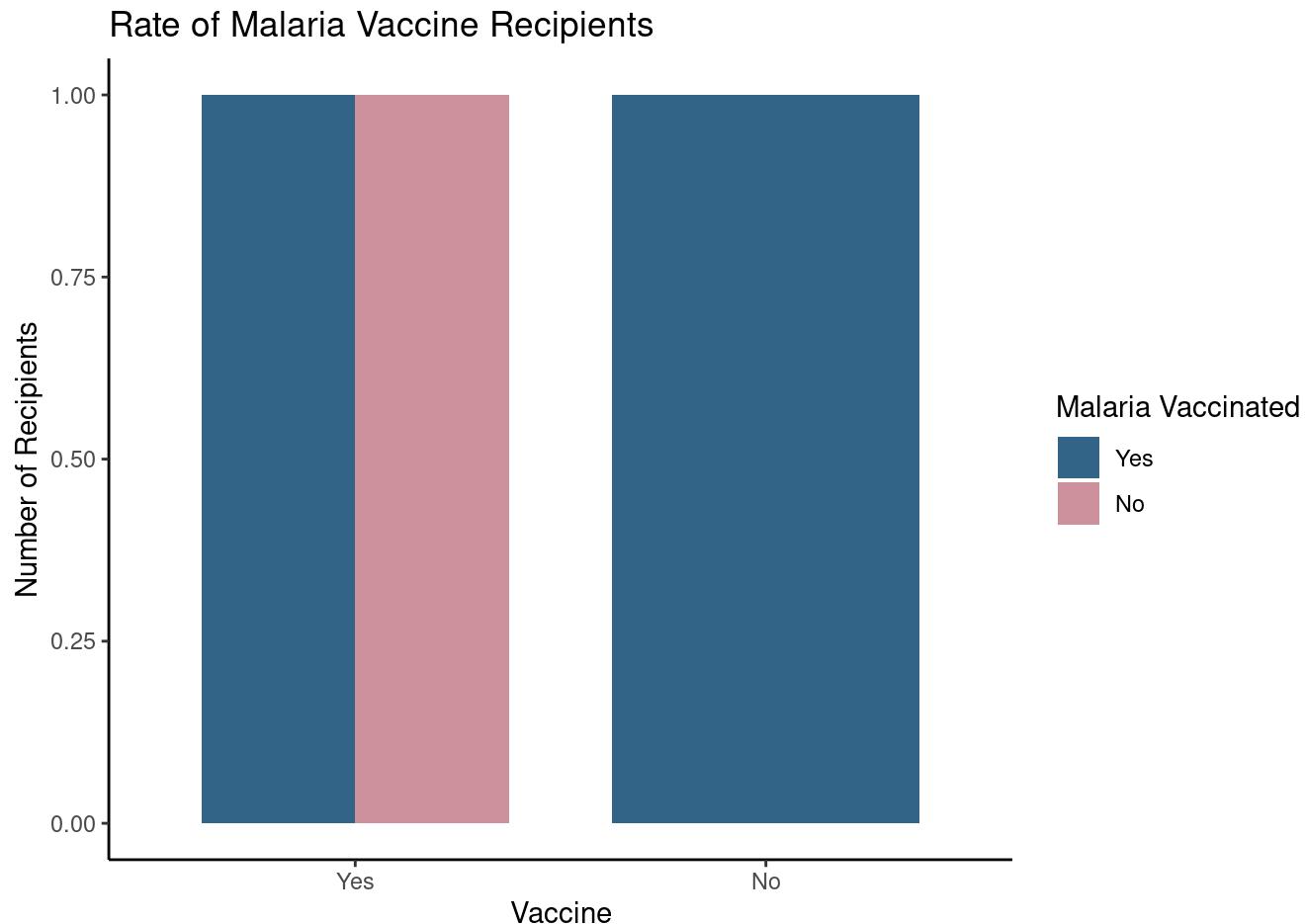
```
Odds ratio |         95% CI
--------------------------
0.04       | [0.00, 0.95]
```

Plot

```r
malaria_clean |>
  group_by(
    vaccinated,
    infected
  ) |>
  summarise(
    n       = n(),
    .groups = "drop"
  ) |>
  group_by(vaccinated) |>
  ggplot(
    mapping = aes(
      x    = vaccinated,
      fill = infected
    )
  )+
  geom_bar(
    position = position_dodge(),
    width    = 0.75
  ) +
  scale_fill_manual(
    values = c("steelblue4", "pink3")
  ) +
  labs(
    x     = "Vaccine",
    y     = "Number of Recipients",
    fill  = "Malaria Vaccinated",
    title = "Rate of Malaria Vaccine Recipients"
  ) +
  theme_classic()
```

Rate of Malaria Vaccine Recipients

Insights

The Contingency table provides a clear numeric details on the proportion of variables. The plot provides visual presentation of the outcome of the data set. On observing both these, it shows that people who were vaccinated were less infected by malaria, compared to individuals who did not take the vaccine. It is also clear that there were 100 percent of malaria infections observed among people who did not get vaccinated.

Ordinal-Ordinal Association

Ordinal Variables with Same Dimension Lengths

The data set selected for Ordinal-Ordinal Association is the Dream Act data. It has the Explanatory Variable Ideology and the Response Variable as Stance. The Ideology variable is categorized as Liberal, Conservative and Moderate. Where as Stance has variables No, Not Sure, and Yes.

Contingency Table

```
dream_clean <-
  dream |>
  as_tibble() |>
  select(
```

```r
    ideology,
    stance
  ) |>
  rename(
    Ideology = ideology,
    Stance   = stance
  ) |>
  drop_na() |>
  mutate(
    Ideology = case_when(
    Ideology == "Conservative" ~ "Conservative",
    Ideology == "Liberal" ~ "Liberal",
    Ideology == "Moderate" ~ "Moderate"
),
Ideology = factor(
  x = Ideology,
  levels = c(
    "Conservative",
    "Liberal",
    "Moderate"
  )
),
Stance = case_when(
Stance == "No" ~ "Low Stance",
Stance == "Not Sure" ~ "Moderate Stance",
Stance == "Yes" ~ "High Stance"
),
Stance = factor(
  x = Stance,
  levels = c(
    "Low Stance",
    "Moderate Stance",
    "High Stance"
  ),
 ),
 )

  dream_clean_table <-
    dream_clean |>
    tbl_cross(
      row   = Ideology,
      col   = Stance,
      label = list(
        Ideology ~ "Idology Levels",
        Stance ~ "Stance Level"
    ),
      percent = "row"
      ) |>
    bold_labels()
```

```
dream_clean_table
```

	Stance Level				
	Low Stance	Moderate Stance	High Stance	Unknown	**Total**
Idology Levels					
Conservative	151 (41%)	0 (0%)	186 (50%)	35 (9.4%)	372 (100%)
Liberal	52 (30%)	0 (0%)	114 (65%)	9 (5.1%)	175 (100%)
Moderate	161 (44%)	0 (0%)	174 (48%)	28 (7.7%)	363 (100%)
Total	364 (40%)	0 (0%)	474 (52%)	72 (7.9%)	910 (100%)

Measure(s) of Association

```
dream_clean_table_gamma <-
  GoodmanKruskalGamma(
    x          = table(dream_clean),
    conf.level = 0.95
  ) |>
  round(
    digits = 2
  )
```
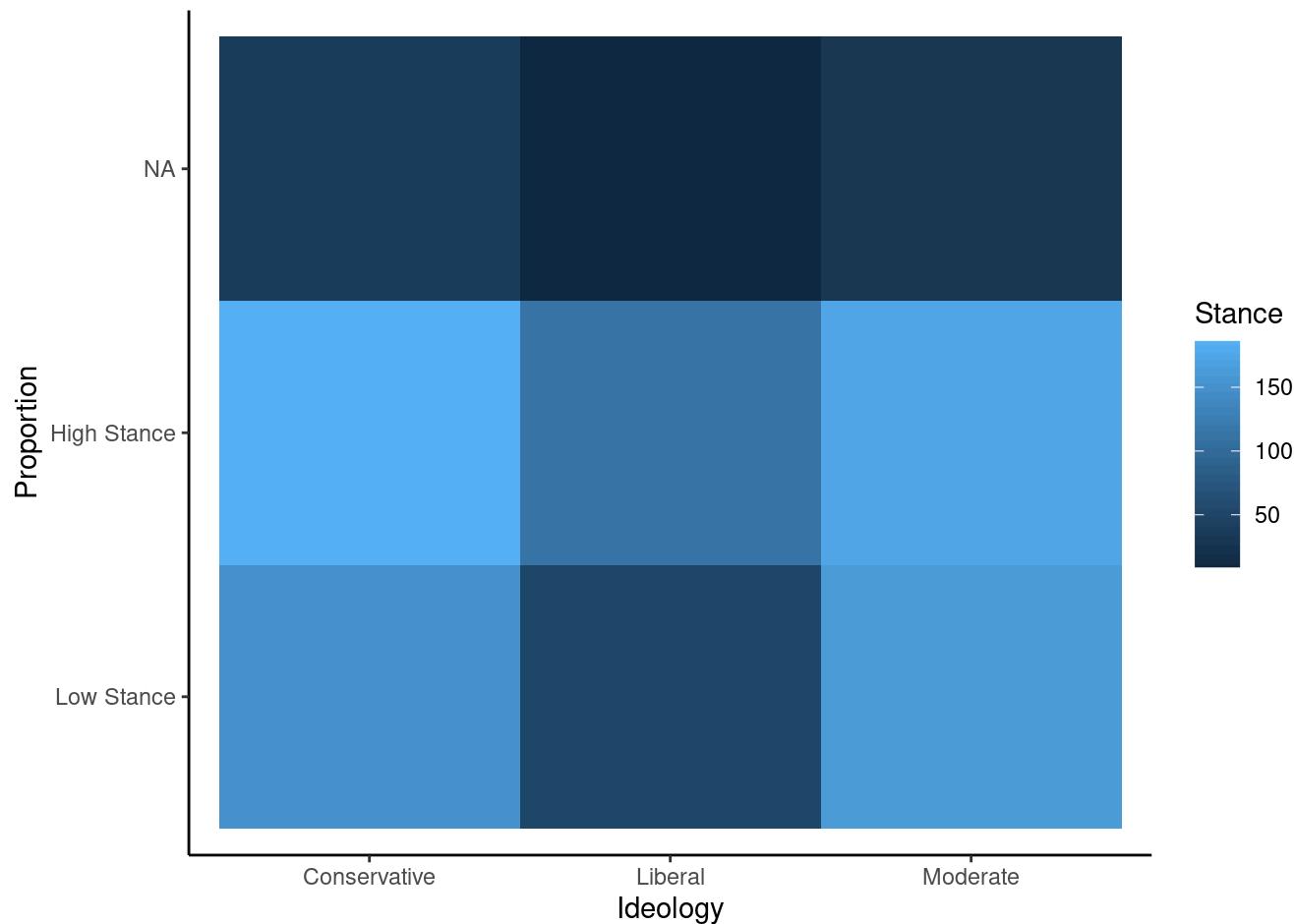
Plot

```
dream_clean |>
  group_by(Ideology, Stance) |>
  summarize(
    n = n(),
    .groups = "drop"
  ) |>
  ggplot(
    mapping = aes(
      x    = Ideology,
      y    = Stance,
      fill = n,
      group = interaction(Ideology, Stance)
    )
  ) +
  geom_tile() +
  labs(
```

```
    x = "Ideology",
    y = "Proportion",
    fill = "Stance"
) +
theme_classic()
```


Insights

Analysis of the contingency table and the corresponding plot reveals a clear relationship between ideology and support for the dream. Participants classified as liberal were significantly more likely to support the Act, while conservatives showed greater opposition.

Ordinal Variables with Different Dimension Lengths

The sleep deprivation data set used here is on hours of sleep and if they are related to different types of professions.The Explanatory Variable is Hours of sleep with ordered variables Less, Moderate and More sleep and Response Variable is Profession which has data about hours of sleep of professions such as Control, Public Transportation, Pilot, Truck drivers and Train Operators.

Contingency Table

```r
sleep_deprivation_clean <-
  sleep_deprivation |>
  as_tibble() |>
  select(
    sleep,
    profession
  ) |>
  rename(
    Sleep      = sleep,
    Profession = profession
  ) |>
  drop_na() |>
  mutate(
    Sleep = case_when(
      Sleep == "<6" ~ "Less Sleep",
      Sleep == "6-8" ~ "Moderate Sleep",
      Sleep == ">8" ~ "More Sleep",
      TRUE ~ NA_character_
    ),
    Sleep= factor(
      x  = Sleep,
      levels = c(
        "Less Sleep",
        "Moderate Sleep",
        "More Sleep"
      ),
    ),
    Profession = case_when(
      Profession == "control" ~ "Control Groups",
      Profession == "bus / taxi / limo drivers" ~ "Public Transport Drivers",
      Profession == "truck drivers" ~ "Truck Drivers",
      Profession == "pilots" ~ "Pilots",
      Profession == "train operators" ~ "Train Operators"
    ),
    Profession = factor(
      x = Profession,
      levels = c(
        "Control Groups",
        "Public Transport Drivers",
        "Truck Drivers",
        "Pilots",
        "Train Operators"
      ),
    ),
  )

sleep_deprivation_clean_table <-
  sleep_deprivation_clean |>
  tbl_cross(
    row   = Sleep,
```

```
    col   = Profession,
    label = list(
      Sleep ~ "Hours of Sleep",
      Profession ~ "Profession of participants"
    ),
    percent = "row"
    ) |>
  bold_labels()

sleep_deprivation_clean_table
```

	Profession of participants					
	Control Groups	Public Transport Drivers	Truck Drivers	Pilots	Train Operators	**Total**
Hours of Sleep						
Less Sleep	35 (25%)	21 (15%)	35 (25%)	19 (14%)	29 (21%)	139 (100%)
Moderate Sleep	193 (28%)	131 (19%)	117 (17%)	132 (19%)	119 (17%)	692 (100%)
More Sleep	64 (25%)	58 (23%)	51 (20%)	51 (20%)	32 (13%)	256 (100%)
Total	292 (27%)	210 (19%)	203 (19%)	202 (19%)	180 (17%)	1,087 (100%)

Measure(s) of Association
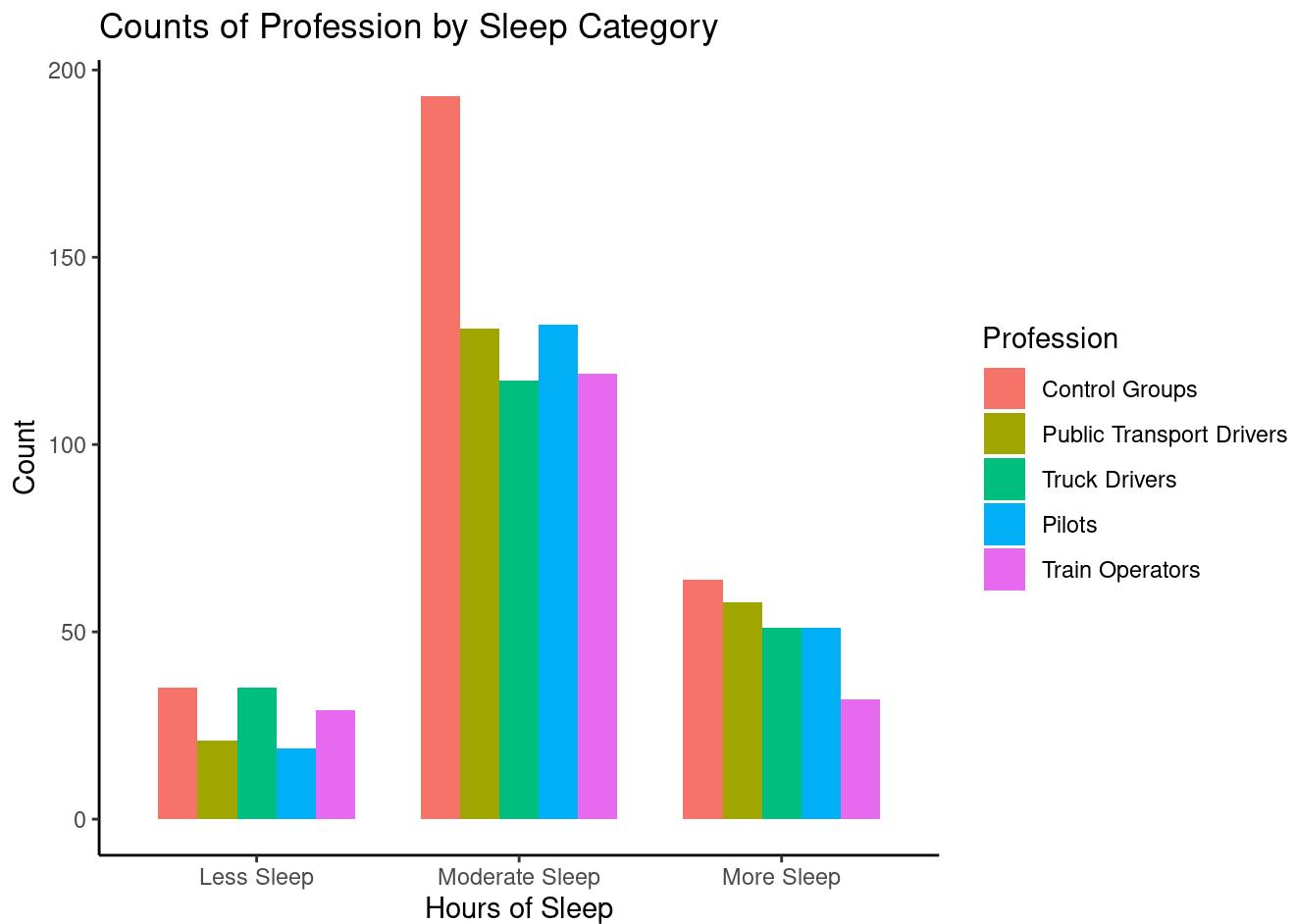
```
sleep_deprivation_clean_table_gamma <-
  GoodmanKruskalGamma(
    x         = table(sleep_deprivation_clean),
    conf.level = 0.95
  ) |>
  round(
    digits = 2
  )
```

Plot

```r
sleep_deprivation_clean |>
  ggplot(
    mapping = aes(
      x = Sleep,
      fill = Profession
    )
  ) +
  geom_bar(
    position = position_dodge(),
    width    = 0.75) +
  labs(
    x     = "Hours of Sleep",
    y     = "Count",
    fill  = "Profession",
    title = "Counts of Profession by Sleep Category"
  ) +
  theme_classic()
```


Counts of Profession by Sleep Category

Insights

The contingency table and visual plot indicate a meaningful relationship between profession type and sleep duration. Workers in high-risk transportation roles are significantly more likely to experience less sleep

compared to control or less demanding roles. As job-related stress or irregular hours increase, the probability of reduced sleep also rises.

Nominal-Nominal Association

The gender_discrimination data set is related to promotion decisions based on gender. For this analysis, the explanatory variable is Gender (Male, Female), and the response variable is Promotion Decision (Promoted, Not Promoted). The data set helps in analyzing whether gender influences promotion outcomes.

Contingency Table

```r
gender_descrimination_clean <-
 gender_discrimination |>
  select(
    gender,
    decision
  ) |>
  drop_na() |>
  mutate(
    gender = case_when(
    gender == "male" ~ "Male",
    gender == "female" ~ "Female"
  ),
  gender = as.factor(gender),
  decision = case_when(
    decision == "promoted" ~ "Promoted",
    decision == "not promoted" ~ "Not Promoted"
  ),
  decision = as.factor(decision)
  )

gender_descrimination_clean |>
  tbl_cross(
    row = gender,
    col = decision,
    label = list(
      gender ~ "Gender",
      decision ~ "Decision of Promotion"
    ),
    percent = "row"
  ) |>
  bold_labels()
```

	Decision of Promotion		Total
	Not Promoted	Promoted	
Gender			
Female	10 (42%)	14 (58%)	24 (100%)
Male	3 (13%)	21 (88%)	24 (100%)
Total	13 (27%)	35 (73%)	48 (100%)

## Measure(s) of Association
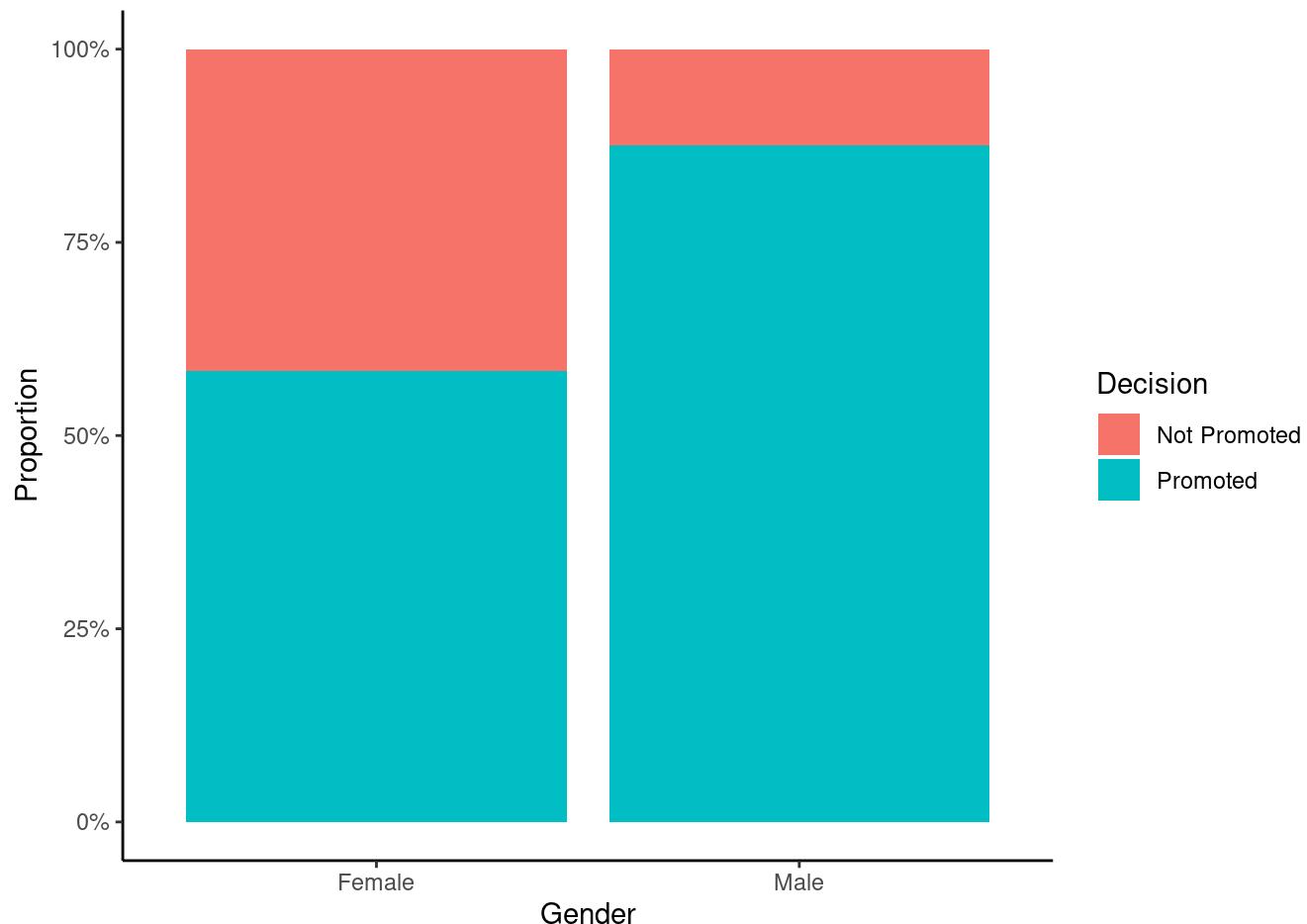
```
CramerV(
  x = gender_descrimination_clean$gender,
  y = gender_descrimination_clean$decision,
  method = "ncchisq",
  conf.level = 0.95
) |>
  round(
    digits = 2
  )
```

```
Cramer V   lwr.ci   upr.ci
   0.33     0.02     0.61
```

Plot

```
gender_descrimination_clean |>
  ggplot(
    mapping = aes(
      x    = gender,
      fill = decision
    )
  )+
  geom_bar(
    position = "fill"
  ) +
  labs(
    x = "Gender",
    y = "Proportion",
    fill = "Decision"
  ) +
  scale_y_continuous(
    labels = scales::percent
```

```
) +
theme_classic()
```


Insights

The contingency table and associated plot show the relationship between gender and promotion. Females are less likely to be promoted compared to males, showing a negative association between gender and promotion decision. The strength of this association suggests bias in promotion, where gender may influence the promotion decision, with females experiencing fewer promotions.