Lab 5.A

Summary Tables with gt

AUTHOR
KOMAL BHOSLE

RStudio Link

https://posit.cloud/spaces/603138/content/10139048

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(easystats)
```

```
# Attaching packages: easystats 0.7.2 (red = needs update)
✗ bayestestR  0.13.2   ✗ correlation 0.8.4
✗ datawizard  0.11.0   ✗ effectsize  0.8.8
✗ insight     0.20.1   ✗ modelbased  0.8.8
✗ performance 0.12.0   ✗ parameters  0.21.7
✗ report      0.5.8    ✗ see         0.8.4

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

```
library(ggpubr)
```

```
Attaching package: 'ggpubr'
```

The following objects are masked from 'package:datawizard':

    mean_sd, median_mad

Data Import

```
melanoma <-
  read_csv(
    file = "melanoma.csv"
  )
```

Rows: 205 Columns: 7
── Column specification ──
Delimiter: ","
dbl (7): time, status, sex, age, year, thickness, ulcer

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.

Frequency Distribution Table

```
melanoma_diagnosis_sex <-
  melanoma |>
  select(
    sex,
    ulcer
  ) |>
  mutate(
    sex = recode(
      sex,
      `1` = "Male",
      `0` = "Female"
    ),
    ulcer = recode(
      ulcer,
      `1` = "Ulcer Present",
      `0` = "Ulcer Absent"
    )
  )
```

```
melanoma_diagnosis_sex |>
  tbl_summary(
    label = list(
      sex = "Sex",
      ulcer = "Ulcer Diagnosis"
```

```
    )
  ) |>
  bold_labels() |>
  italicize_levels() |>
  as_gt() |>
  tab_source_note(
    source_note = md(
      "Data from 'Melanoma' dataset (in **MASS** package)."
    )
  ) |>
  tab_header(
    title = md(
      "**Frequency of Malignant Melanoma Patients by Diagnosis and Sex"
    ),
    subtitle = "Based of 205 patients in Denamrk"
  )
```

**Frequency of Malignant Melanoma Patients by Diagnosis and Sex

Based of 205 patients in Denamrk

Characteristic	N = 205[1]
Sex	
Female	126 (61%)
Male	79 (39%)
Ulcer Diagnosis	
Ulcer Absent	115 (56%)
Ulcer Present	90 (44%)

[1] n (%)

Data from 'Melanoma' dataset (in **MASS** package).
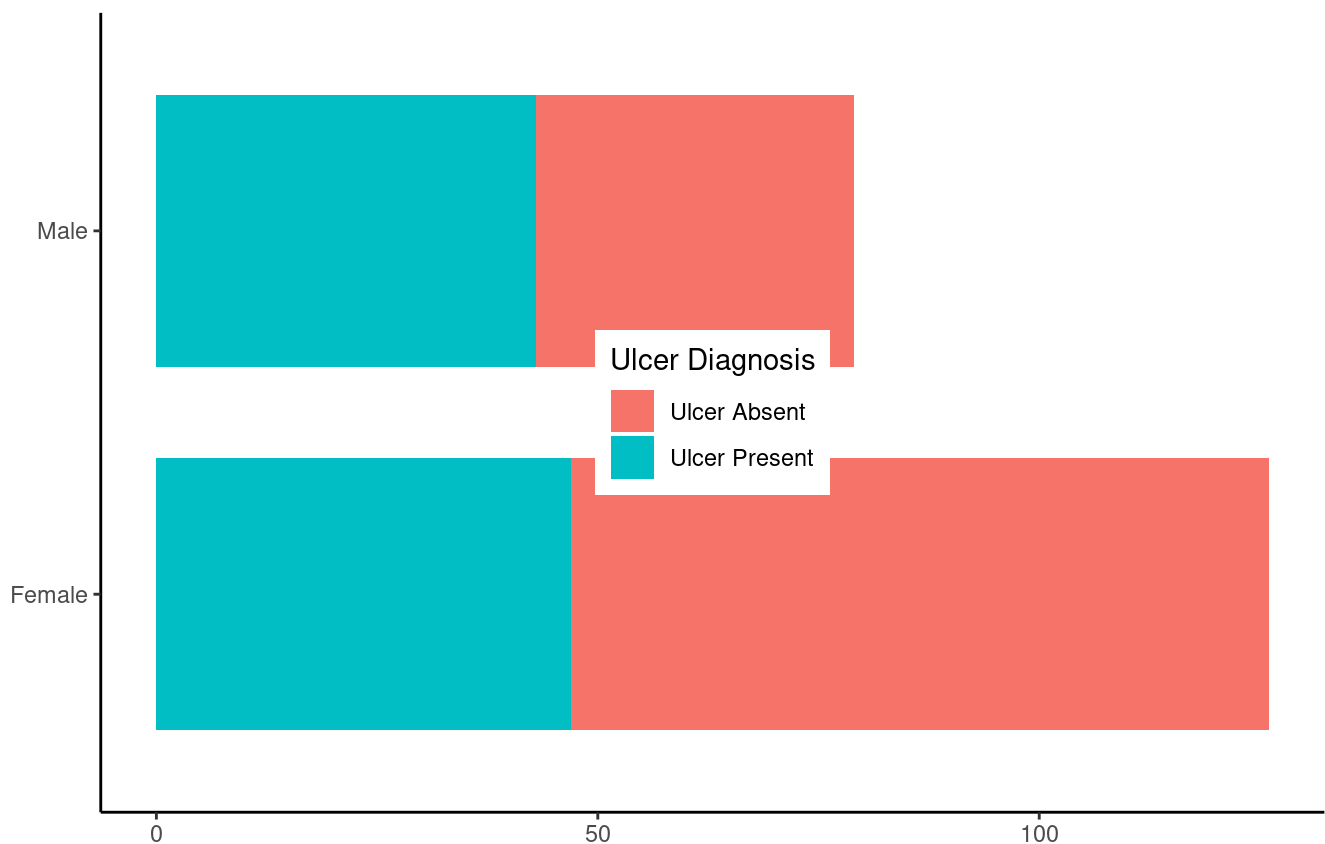
Grouped Bar Plot

```
melanoma_diagnosis_sex |>
  ggplot(
    mapping = aes(
      x = sex,
      fill = ulcer
    )
  ) +
```

```
geom_bar(
    width = 0.75
) +
scale_color_manual(
    "steelblue4",
    "hotpink3"
) +
coord_flip() +
labs(
    x = NULL,
    y = NULL,
    fill = "Ulcer Diagnosis",
    title = "Grouped Bar Plot of Melanoma Ulcer Diagnosis by Sex",
    subtitle = "(Based on 205 patients in Denmark)"
) +
theme_classic() +
theme (
    legend.position = c("top", "left"),
    legend.title = NULL,
    plot.title = element_text(
        face = "bold",
        size = 15
    )
)
```

Grouped Bar Plot of Melanoma Ulcer Diagnosis by Sex
(Based on 205 patients in Denmark)

Descriptive Statistics Table

```
melanoma_clean <-
  melanoma |>
  mutate(
    status = recode(
      status,
      `1` = "Melanoma Death",
      `2` = "Not Dead",
      `3` = "Non-Melanoma Death"
    ),
    sex = recode(
      sex,
      `1` = "Male",
      `0` = "Female"
    ),
    ulcer = recode(
      ulcer,
      `1` = "Ulcer Present",
      `0` = "Ulcer Absent"
    )
  )

library(e1071)
```

```
Attaching package: 'e1071'

The following objects are masked from 'package:parameters':

    kurtosis, skewness

The following objects are masked from 'package:datawizard':

    kurtosis, skewness
```

```
melanoma_clean |>
  select(
    sex,
    ulcer,
    thickness
  ) |>
  group_by(
    sex,
    ulcer
  ) |>
  summarise(
    mean = mean(thickness),
    sd = sd(thickness),
```

```r
    skewness = skewness(thickness),
    kurtosis = kurtosis(thickness),
    .groups = "drop"
  ) |>
  gt(
    rowname_col = "ulcer",
    groupname_col = "sex"
  ) |>
  fmt_number(
    columns = c(mean, sd, skewness, kurtosis),
    decimals = 2
  ) |>
  tab_style(
    style = list(
      cell_text(weight = "bold"),
      cell_fill(color = "#F2F2F2")
    ),
    locations = cells_row_groups()
  ) |>
  cols_label(
    mean = md("**Mean**"),
    sd = md("**SD**"),
    skewness = md("**γ1**"),
    kurtosis = md("**γ2**")
  ) |>
  tab_stub_indent(
    rows = everything(),
    indent = 2
  ) |>
  tab_footnote(
    footnote = "arithmetic mean",
    locations = cells_column_labels(
      columns = "mean"
    )
  ) |>
  tab_footnote(
    footnote = "standard deviation",
    locations = cells_column_labels(
      columns = "sd"
    )
  ) |>
  tab_footnote(
    footnote = "skewness",
    locations = cells_column_labels(
      columns = "skewness"
    )
  ) |>
  tab_footnote(
    footnote = "kurtosis",
    locations = cells_column_labels(
      columns = "kurtosis"
```

```
  )
 ) |>
 opt_footnote_marks(
   marks = "letters"
 ) |>
 tab_source_note(
   source_note = md(
     "Data from 'Melanoma' dataset (in **MASS** package)."
   )
 ) |>
 tab_header(
   title = md(
     "**Descriptive Statistics for Malignant Melanoma Thickness by Patient Sex and Ulcer Diagnos:
   ) ,
   subtitle = "Based of 205 patients in Denmark"
 )
```
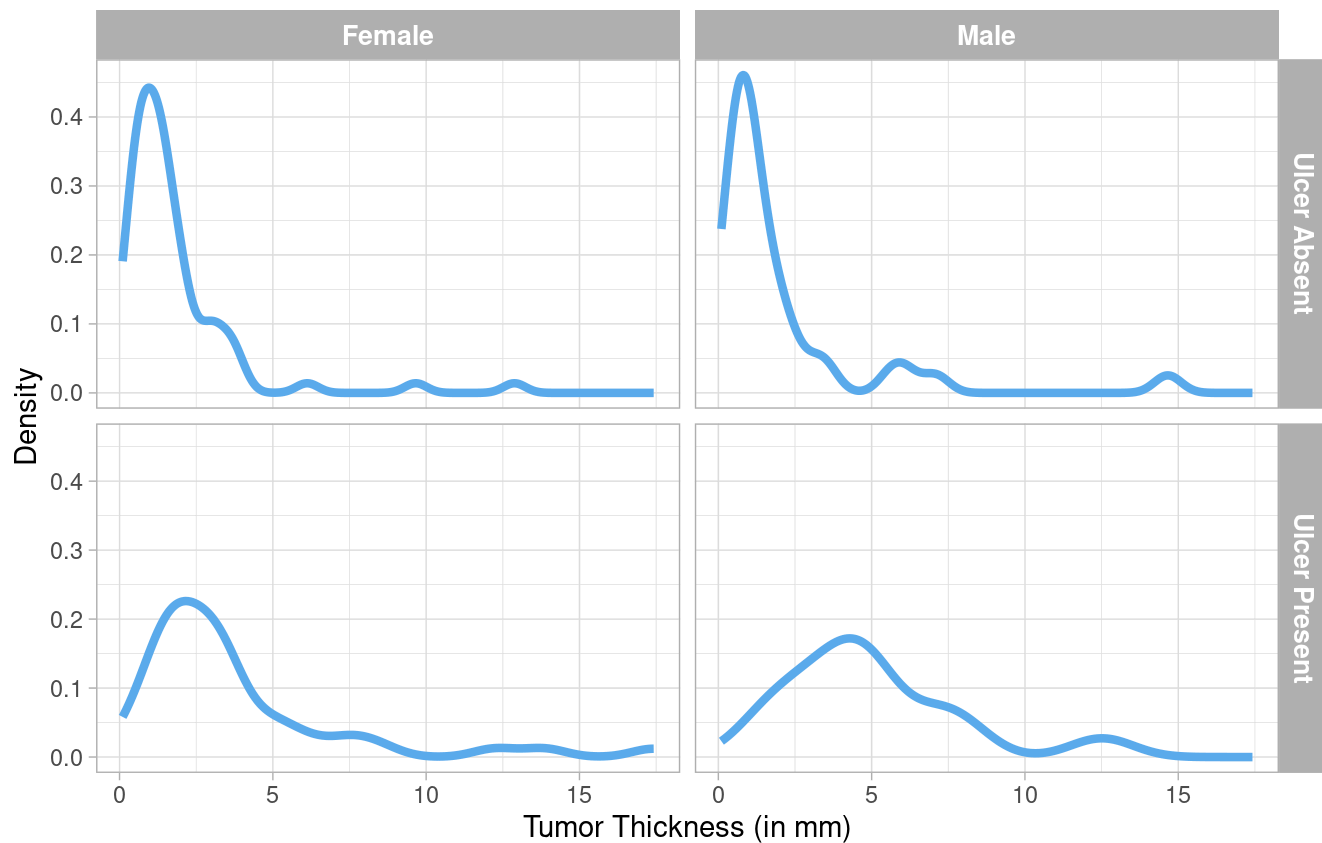
Descriptive Statistics for Malignant Melanoma Thickness by Patient Sex and Ulcer Diagnosis

Based of 205 patients in Denmark

	Mean[a]	SD[b]	γ1[c]	γ2[d]
Female				
Ulcer Absent	1.74	1.90	3.59	16.14
Ulcer Present	3.75	3.45	2.17	4.89
Male				
Ulcer Absent	1.97	2.73	3.08	10.63
Ulcer Present	4.98	2.84	1.11	1.02

[a] arithmetic mean

[b] standard deviation

[c] skewness

[d] kurtosis

Data from 'Melanoma' dataset (in **MASS** package).

Density Plot via Facet Grid

```
melanoma_clean |>
  ggplot(
    mapping = aes(
      x = thickness
    )
  ) +
  geom_density(
    linewidth = 1.5,
    color = "steelblue2"
  ) +
  facet_grid(
    rows = vars(ulcer),
    cols = vars(sex)
  ) +
  labs(
    x = "Tumor Thickness (in mm)",
    y = "Density",
    title = "Density Plots of Melanoma Tumor Thickness (in mm)",
    subtitle = "Facet Grid by Ulcer Diagnosis and Sex"
  ) +
  theme_light() +
  theme(
    strip.text = element_text(
      face = "bold",
      size = 10
    ),
    plot.title = element_text(
      face = "bold",
      size = 15
    )
  )
```

Density Plots of Melanoma Tumor Thickness (in mm)
Facet Grid by Ulcer Diagnosis and Sex

Contingency Table

```
melanoma_transform <-
  melanoma_clean |>
  mutate(
    sex = recode(
      sex,
      "Male" = "Yes",
      "Female" = "No"
    ),
    sex = factor(
      sex,
      levels = c("Yes", "No")
    ),
    ulcer = recode(
      ulcer,
      "Ulcer Present" = "Yes",
      "Ulcer Absent" = "No"
    ),
    ulcer = factor(
      ulcer,
      levels = c("Yes", "No")
```

```
    )
  )
```

```
melanoma_transform |>
  tbl_cross(
    row = sex,
    col = ulcer,
    percent = "cell",
    label = list(
      sex = "Male sex",
      ulcer = "Ulcer Presence"
    )
  ) |>
  bold_labels() |>
  italicize_levels() |>
  as_gt() |>
  tab_source_note(
    source_note = md(
      "Data from 'Melanoma' dataset (in **MASS** package)."
    )
  ) |>
  tab_header(
    title = md(
      "**2 x 2 Contingency Table of Melanoma Ulcer Present in Males**"
    ) ,
    subtitle = "**A cross-tabulation based upon 205 patients in Denmark**"
  )
```
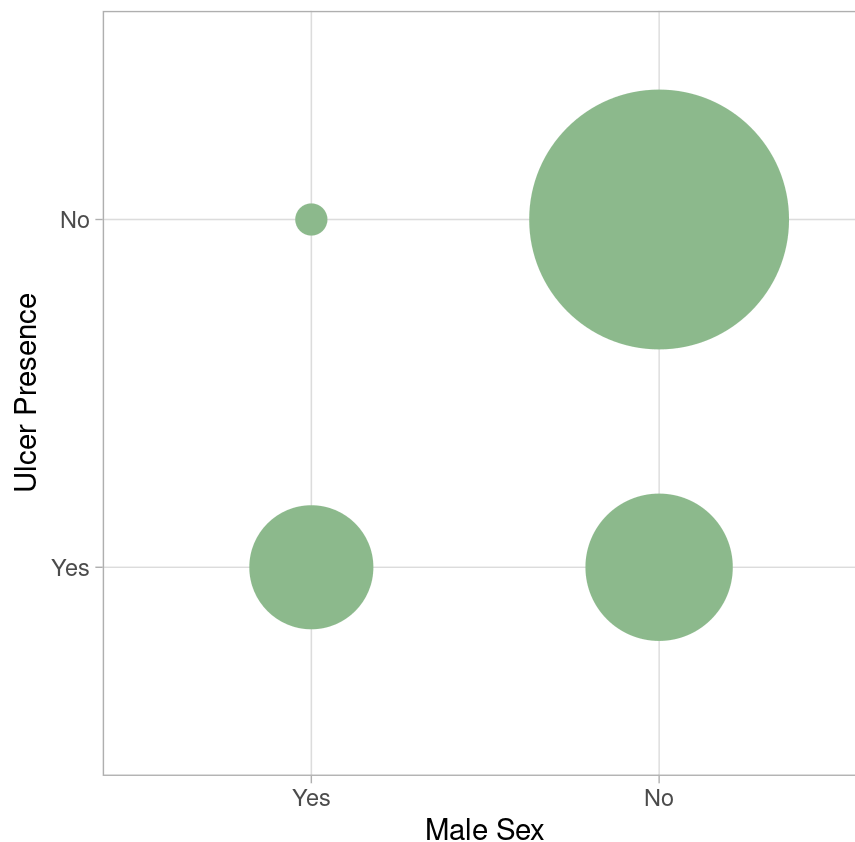
2 x 2 Contingency Table of Melanoma Ulcer Present in Males

A cross-tabulation based upon 205 patients in Denmark

| | Ulcer Presence | | |
	Yes	*No*	**Total**
Male sex			
Yes	43 (21%)	36 (18%)	79 (39%)
No	47 (23%)	79 (39%)	126 (61%)
Total	90 (44%)	115 (56%)	205 (100%)

Data from 'Melanoma' dataset (in **MASS** package).

Balloon Plot

```r
melanoma_transform |>
  select(
    sex,
    ulcer
  ) |>
  group_by(
    sex, ulcer
  ) |>
  summarise(
    n = n(),
    .groups = "drop"
  ) |>
  ggplot(
    mapping = aes(
      x = sex,
      y = ulcer,
      size = n
    )
  ) +
  geom_point(
    color = "darkseagreen"
  ) +
  scale_size_continuous(
    range = c(5, 45)
  ) +
  coord_fixed() +
  labs(
    x = "Male Sex",
    y = "Ulcer Presence",
    title = "Ulcer Incidence in Male Melanoma Patients",
    subtitle = "Balloon Plot of 205 Danish Patients"
  ) +
  theme_light() +
  theme(
    legend.position = "none",
    plot.title = element_text(
      face = "bold",
      size = 14
    )
  )
```

Ulcer Incidence in Male Melanoma Patients
Balloon Plot of 205 Danish Patients

Questions on Data Insights

Frequency Distribution Table

1. Why is the frequency differential for sex concerning?

The frequency differential is concerning as the table shows 61% of ulcer presence diagnosed in females. Where as percentage of male diagnosed are 39%.

2. How would over-representation of one sex affect the interpretation of the ulcer diagnosis frequency differential?

Over-representation of one sex can be misleading in interpreting the ulcer diagnosis. One sex can divert the percentage making it appear as one gender more prone to ulcer than the other, where as the result is due to misproportion of one sex.

Grouped Bar Plot

3. With the knowledge of males being under-represented in the sample, how would you interpret the group bar plot?

If males are under-represented, interpreting the bar plot needs to be done cautiously. It should be noted that if there is representation of males lower than females, it could be due to lesser percentage of male representation. To avoid this, focus should be on equal proportions.

Descriptive Statistics Table

4. What trend can be logically deduced from kurtosis, as calculated in both males and females?

Ulcers can be associated with tumor thickness in males where as females might indicate fewer patients with ulcer with possibility of presence of tumor.

5. What other trends can you determine from the descriptive statistics table?

The tumor thickness is increased when there is ulcer present. Males have more thick tumors than females.

Density Plot Facet Grid

6. What insights does the density plot facet grid offer?

The density plot facet grid provides insights on comparison of different variables and combinations increasing insights. The thickness of the tumor increases with presence of ulcer. In males the tumor thickness is more compared to females.

2 x 2 Contingency Table

7. If males were not under-represented in the sample, how would the table change?

If males were not under-represented, there would be different frequencies and percentages. Increase in percentage of male would slightly affet the female percentage to balance out the total.

Balloon Plot

8. Why is the comparison of ulcer presence for males insightful, but the comparison across males and females when ulcers are present not insightful?

It is insightful because it provides isolated result on presence of ulcers where as it gets un-insightful as it gets uncleaer on what is actually making an impact.