Vis

Edward Bosko

2025-12-19

Load libraries

```
library(tidyverse)
library(readxl)
library(ggplot2)
```

Read in data

```
#directory of file
wd <- "C:/Users/Eddie/OneDrive/Desktop/BIOSTAT/Thesis/Journal/DataProcessed/ModelingData.xlsx"

#load in individual sheets
tib <- read_excel(wd, sheet = "tib.long")
ln <- read_excel(wd, sheet = "ln.long")
ggo <- read_excel(wd, sheet = "ggo.long")
cons <- read_excel(wd, sheet = "cons.long")
bronch <- read_excel(wd, sheet = "bronch.long")
atel <- read_excel(wd, sheet = "atel.long")
thin_bin <- read_excel(wd, sheet = "thin.long")
thick_bin <- read_excel(wd, sheet = "thick.long")
thin <- read_excel(wd, sheet = "thin.long")
thick <- read_excel(wd, sheet = "thick.long")
```

Visualize the Distribution of Scores

Tree-in-bud

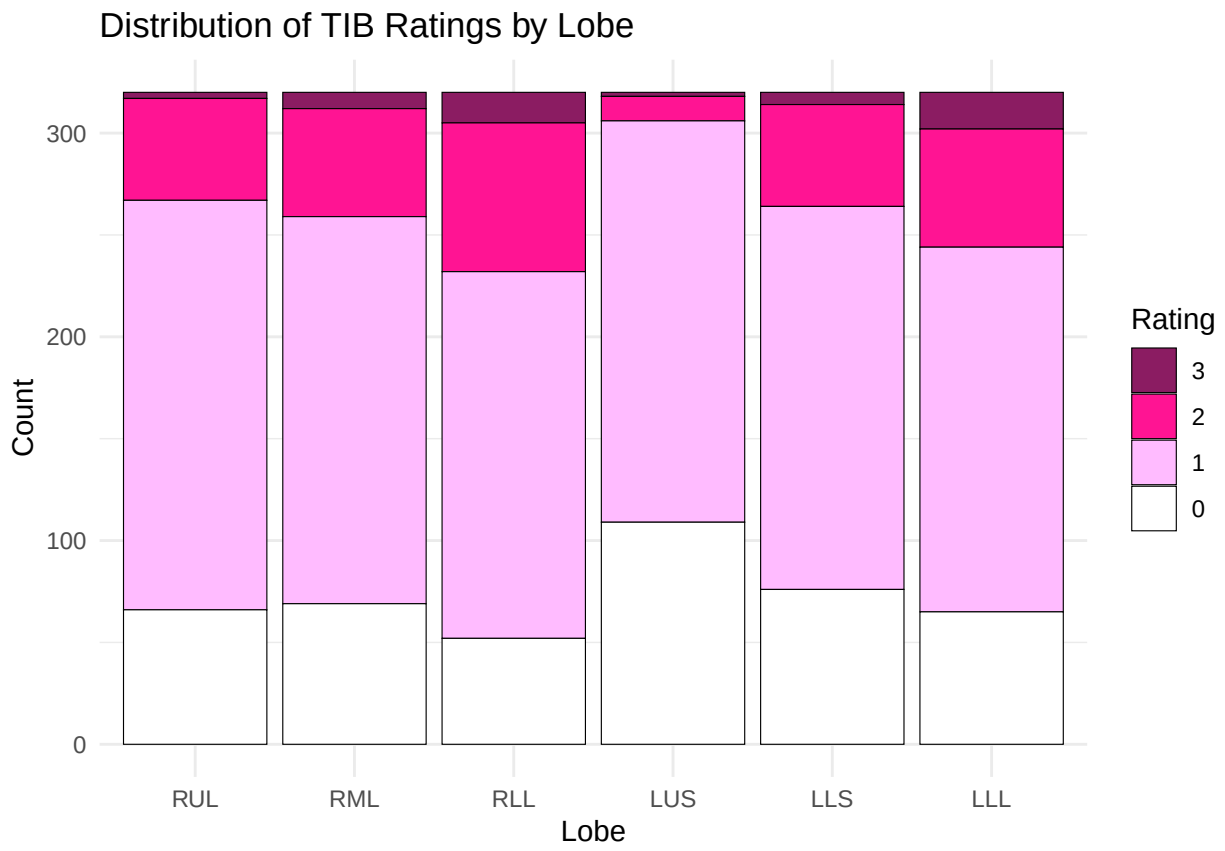
```
score_colors_new <- c(
  "3" = "maroon4",  # maroon for Score 3
  "2" = "deeppink",  # "strawberry for Score 2
  "1" = "plum1",  # Light pink for Score 1
  "0" = "white"   # white for Score 0
)

#by lobe
tib %>%
```

```
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 24 x 4
## # Groups:   lobe [6]
##    lobe  score     n   prop
##    <chr> <dbl> <int>  <dbl>
##  1 LLL       0    65 0.203
##  2 LLL       1   179 0.559
##  3 LLL       2    58 0.181
##  4 LLL       3    18 0.0562
##  5 LLS       0    76 0.238
##  6 LLS       1   188 0.588
##  7 LLS       2    50 0.156
##  8 LLS       3     6 0.0188
##  9 LUS       0   109 0.341
## 10 LUS       1   197 0.616
## # i 14 more rows
```

```
tib %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of TIB Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Large Nodules

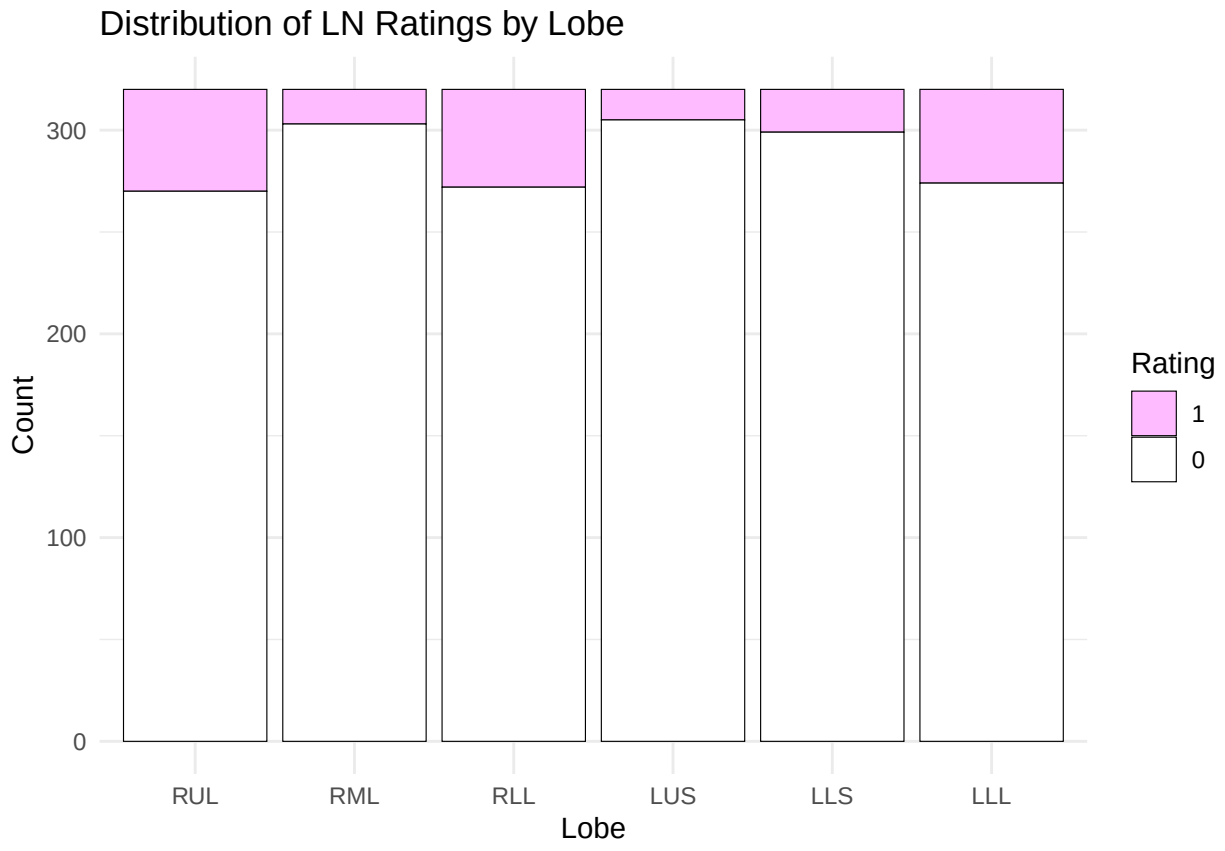
```
#by lobe
ln %>%
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 12 x 4
## # Groups:   lobe [6]
##    lobe  score     n   prop
##    <chr> <dbl> <int>  <dbl>
##  1 LLL       0   274 0.856
##  2 LLL       1    46 0.144
##  3 LLS       0   299 0.934
##  4 LLS       1    21 0.0656
##  5 LUS       0   305 0.953
##  6 LUS       1    15 0.0469
##  7 RLL       0   272 0.85
##  8 RLL       1    48 0.15
```

```
##  9 RML       0   303 0.947
## 10 RML       1    17 0.0531
## 11 RUL       0   270 0.844
## 12 RUL       1    50 0.156
```

```
ln %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of LN Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Ground-glass Opacities

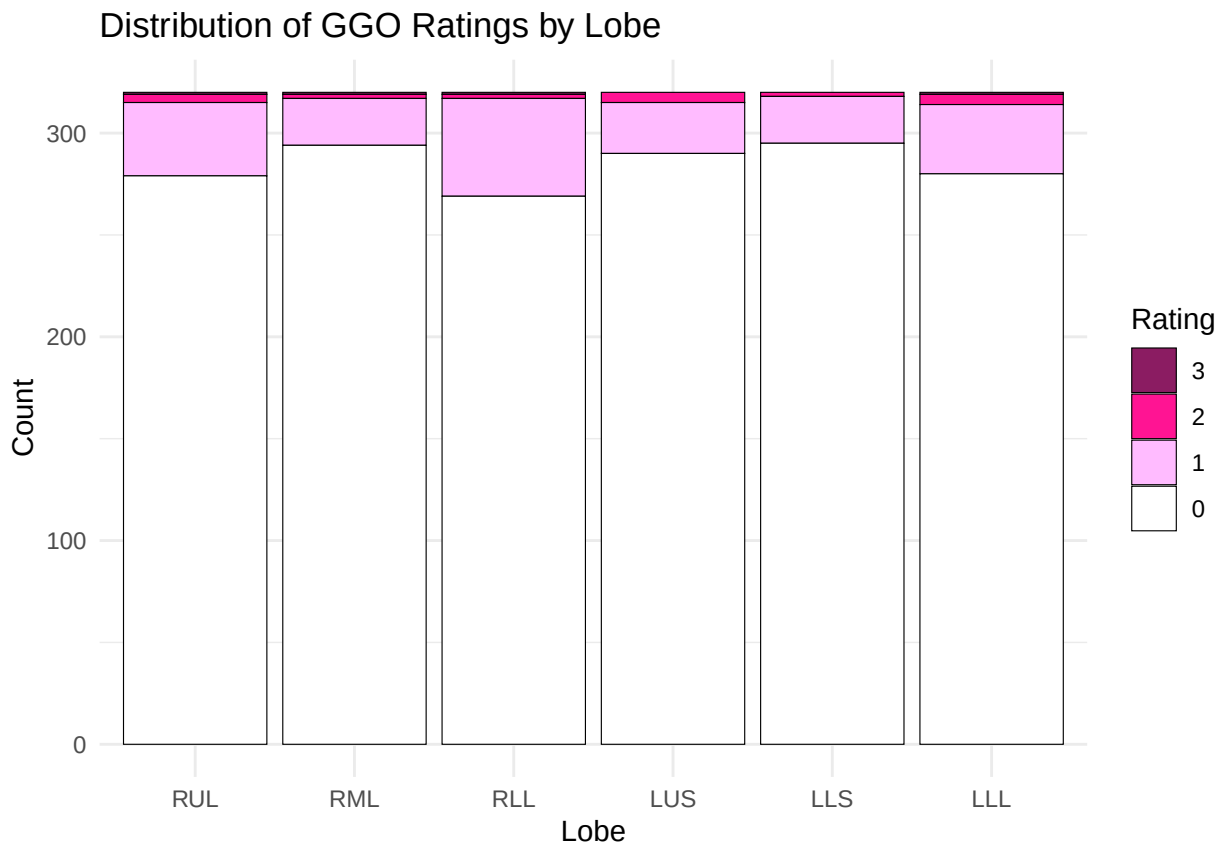
```
#by lobe
ggo %>%
  group_by(lobe, score) %>%
```

```
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 22 x 4
## # Groups:   lobe [6]
##    lobe  score     n    prop
##    <chr> <dbl> <int>   <dbl>
##  1 LLL       0   280 0.875
##  2 LLL       1    34 0.106
##  3 LLL       2     5 0.0156
##  4 LLL       3     1 0.00312
##  5 LLS       0   295 0.922
##  6 LLS       1    23 0.0719
##  7 LLS       2     2 0.00625
##  8 LUS       0   290 0.906
##  9 LUS       1    25 0.0781
## 10 LUS       2     5 0.0156
## # i 12 more rows
```

```
ggo %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of GGO Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Consolidations

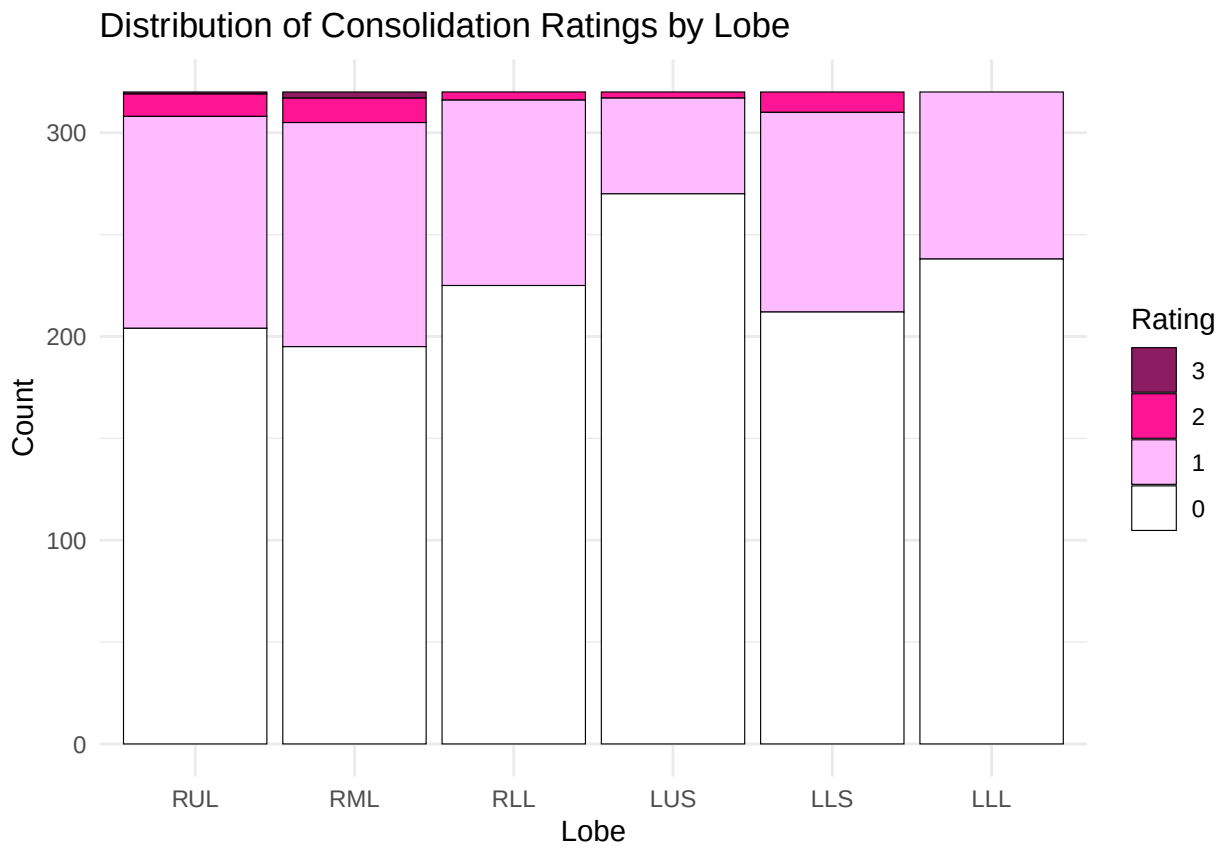
```
#by lobe
cons %>%
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 19 x 4
## # Groups:   lobe [6]
##    lobe  score     n    prop
##    <chr> <dbl> <int>   <dbl>
##  1 LLL       0   238 0.744
##  2 LLL       1    82 0.256
##  3 LLS       0   212 0.662
##  4 LLS       1    98 0.306
##  5 LLS       2    10 0.0312
##  6 LUS       0   270 0.844
##  7 LUS       1    47 0.147
##  8 LUS       2     3 0.00938
```

```
##  9 RLL       0   225 0.703
## 10 RLL       1    91 0.284
## 11 RLL       2     4 0.0125
## 12 RML       0   195 0.609
## 13 RML       1   110 0.344
## 14 RML       2    12 0.0375
## 15 RML       3     3 0.00938
## 16 RUL       0   204 0.638
## 17 RUL       1   104 0.325
## 18 RUL       2    11 0.0344
## 19 RUL       3     1 0.00312
```

```
cons %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of Consolidation Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



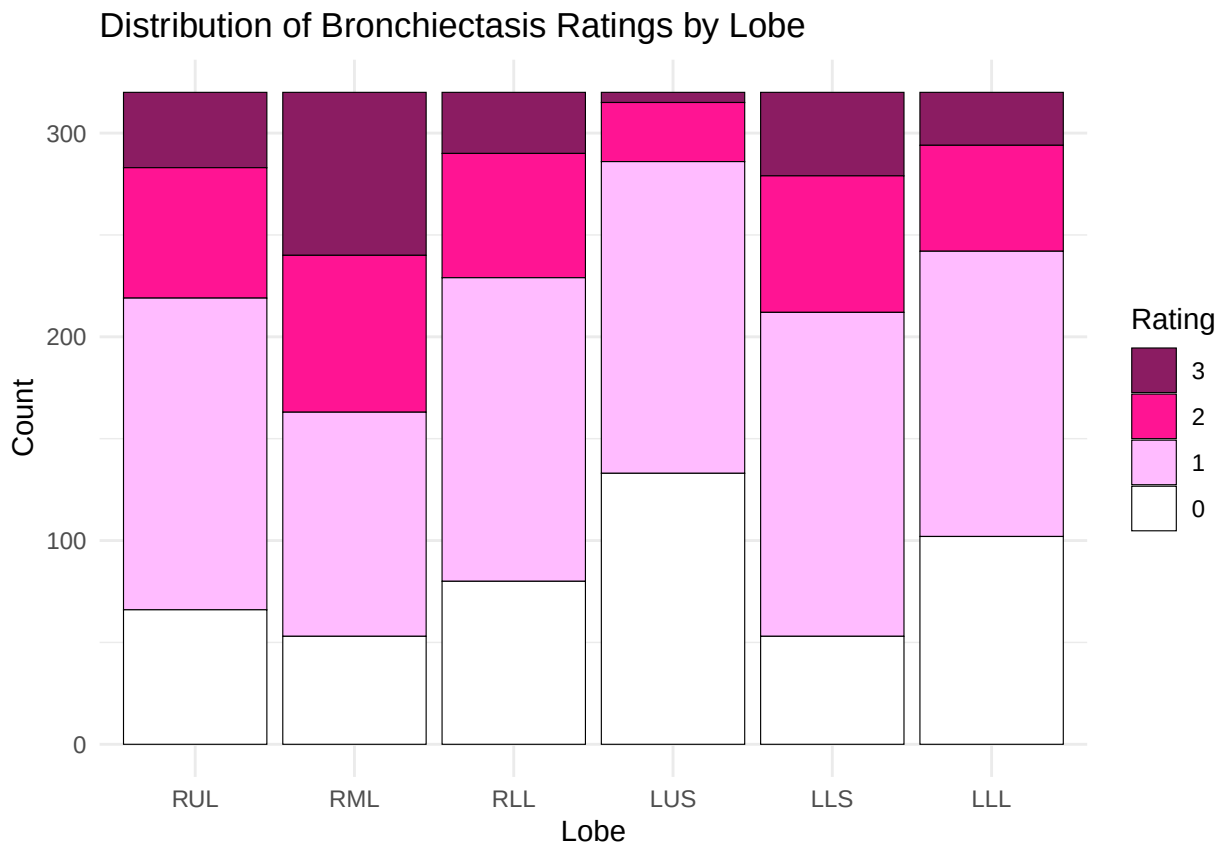
Bronchiectasis

```
#by lobe
bronch %>%
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 24 x 4
## # Groups:   lobe [6]
##    lobe  score     n   prop
##    <chr> <dbl> <int>  <dbl>
##  1 LLL       0   102 0.319
##  2 LLL       1   140 0.438
##  3 LLL       2    52 0.162
##  4 LLL       3    26 0.0812
##  5 LLS       0    53 0.166
##  6 LLS       1   159 0.497
##  7 LLS       2    67 0.209
##  8 LLS       3    41 0.128
##  9 LUS       0   133 0.416
## 10 LUS       1   153 0.478
## # i 14 more rows
```

```
bronch %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of Bronchiectasis Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Atelectasis

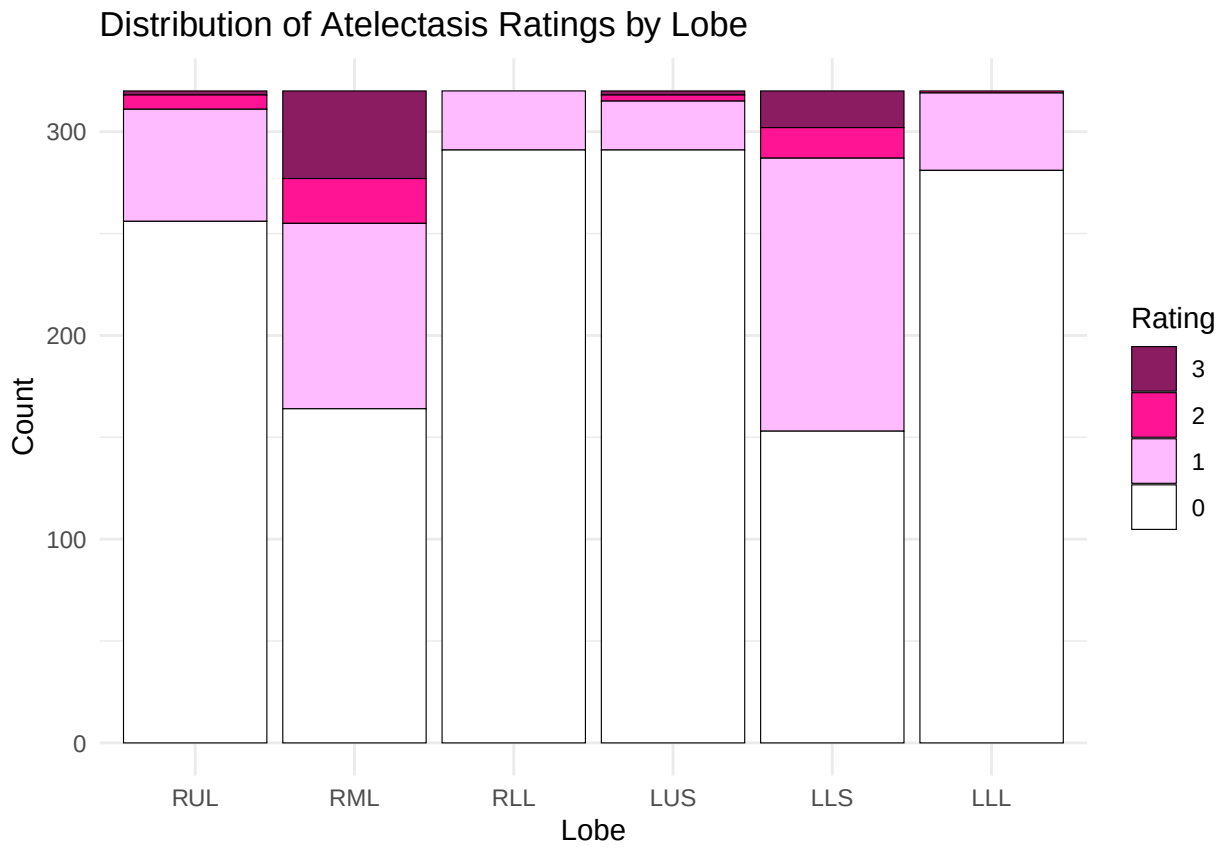
```
#by lobe
atel %>%
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 21 x 4
## # Groups:   lobe [6]
##    lobe  score     n    prop
##    <chr> <dbl> <int>   <dbl>
##  1 LLL       0   281 0.878
##  2 LLL       1    38 0.119
##  3 LLL       2     1 0.00312
##  4 LLS       0   153 0.478
##  5 LLS       1   134 0.419
##  6 LLS       2    15 0.0469
##  7 LLS       3    18 0.0562
##  8 LUS       0   291 0.909
```

```
##  9 LUS       1    24 0.075
## 10 LUS       2     3 0.00938
## # i 11 more rows
```

```
atel %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of Atelectasis Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Thin Wall Cavity

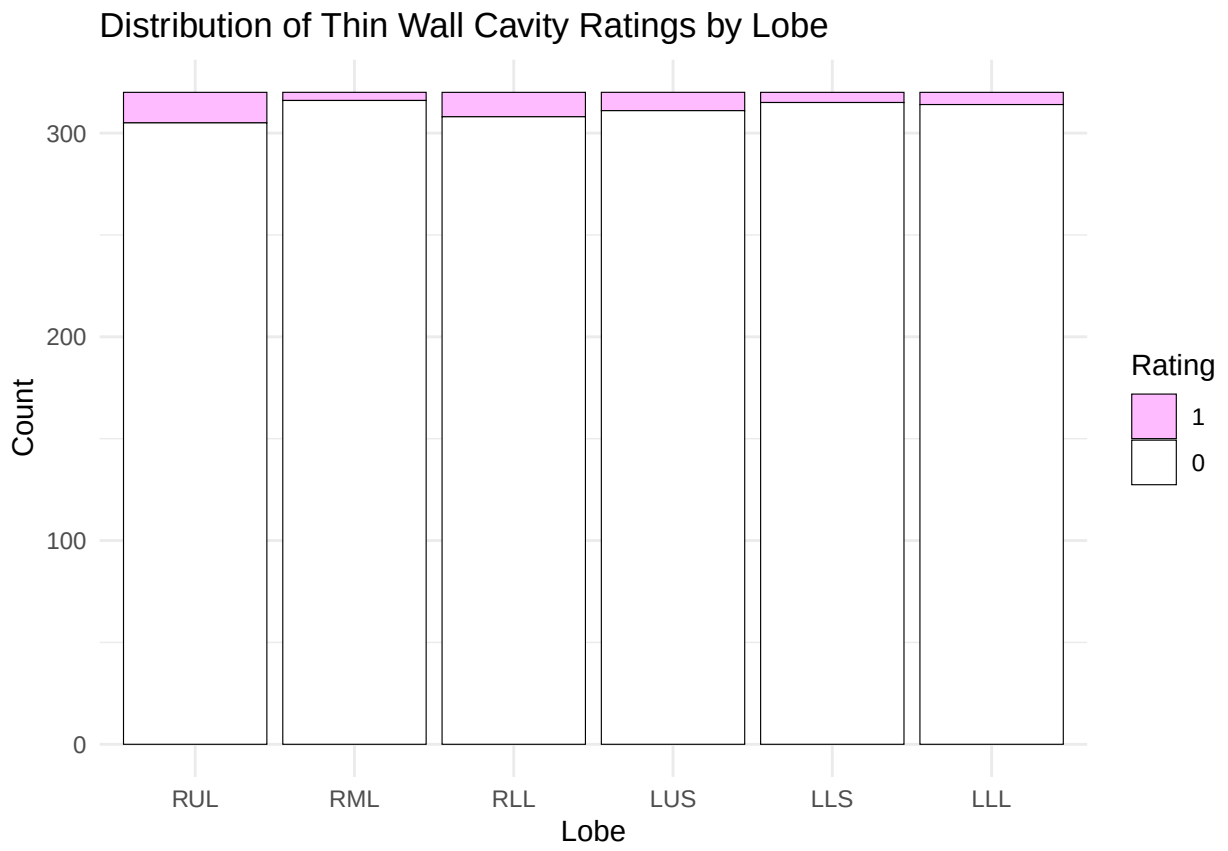
```
#by lobe
thin %>%
  group_by(lobe, score) %>%
```

```
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 12 x 4
## # Groups:   lobe [6]
##    lobe  score     n   prop
##    <chr> <dbl> <int>  <dbl>
##  1 LLL       0   314 0.981
##  2 LLL       1     6 0.0188
##  3 LLS       0   315 0.984
##  4 LLS       1     5 0.0156
##  5 LUS       0   311 0.972
##  6 LUS       1     9 0.0281
##  7 RLL       0   308 0.962
##  8 RLL       1    12 0.0375
##  9 RML       0   316 0.988
## 10 RML       1     4 0.0125
## 11 RUL       0   305 0.953
## 12 RUL       1    15 0.0469
```

```
thin %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of Thin Wall Cavity Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```



Thick Wall Cavity

```
#by lobe
thick %>%
  group_by(lobe, score) %>%
  summarize(n=n()) %>%
  mutate(prop = n / sum(n))
```

```
## `summarise()` has grouped output by 'lobe'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 12 x 4
## # Groups:   lobe [6]
##    lobe  score     n   prop
##    <chr> <dbl> <int>  <dbl>
##  1 LLL       0   306 0.956
##  2 LLL       1    14 0.0438
##  3 LLS       0   312 0.975
##  4 LLS       1     8 0.025
##  5 LUS       0   302 0.944
##  6 LUS       1    18 0.0562
##  7 RLL       0   290 0.906
##  8 RLL       1    30 0.0938
```

```
##  9 RML       0   316 0.988
## 10 RML       1     4 0.0125
## 11 RUL       0   265 0.828
## 12 RUL       1    55 0.172
```

```
thick %>%
  mutate(score = factor(score, levels = c(3,2,1,0)),
         lobe = factor(lobe, levels = c('RUL','RML','RLL','LUS','LLS','LLL'))) %>%
  ggplot(aes (x=lobe, fill = factor(score))) +
  geom_bar(position = 'stack', color='black', linewidth=0.2) +
  scale_fill_manual(values = score_colors_new) +
  labs(
    title = 'Distribution of Thick Wall Cavity Ratings by Lobe',
    x='Lobe',
    y='Count',
    fill='Rating'
  ) +
  theme_minimal()
```