FYS

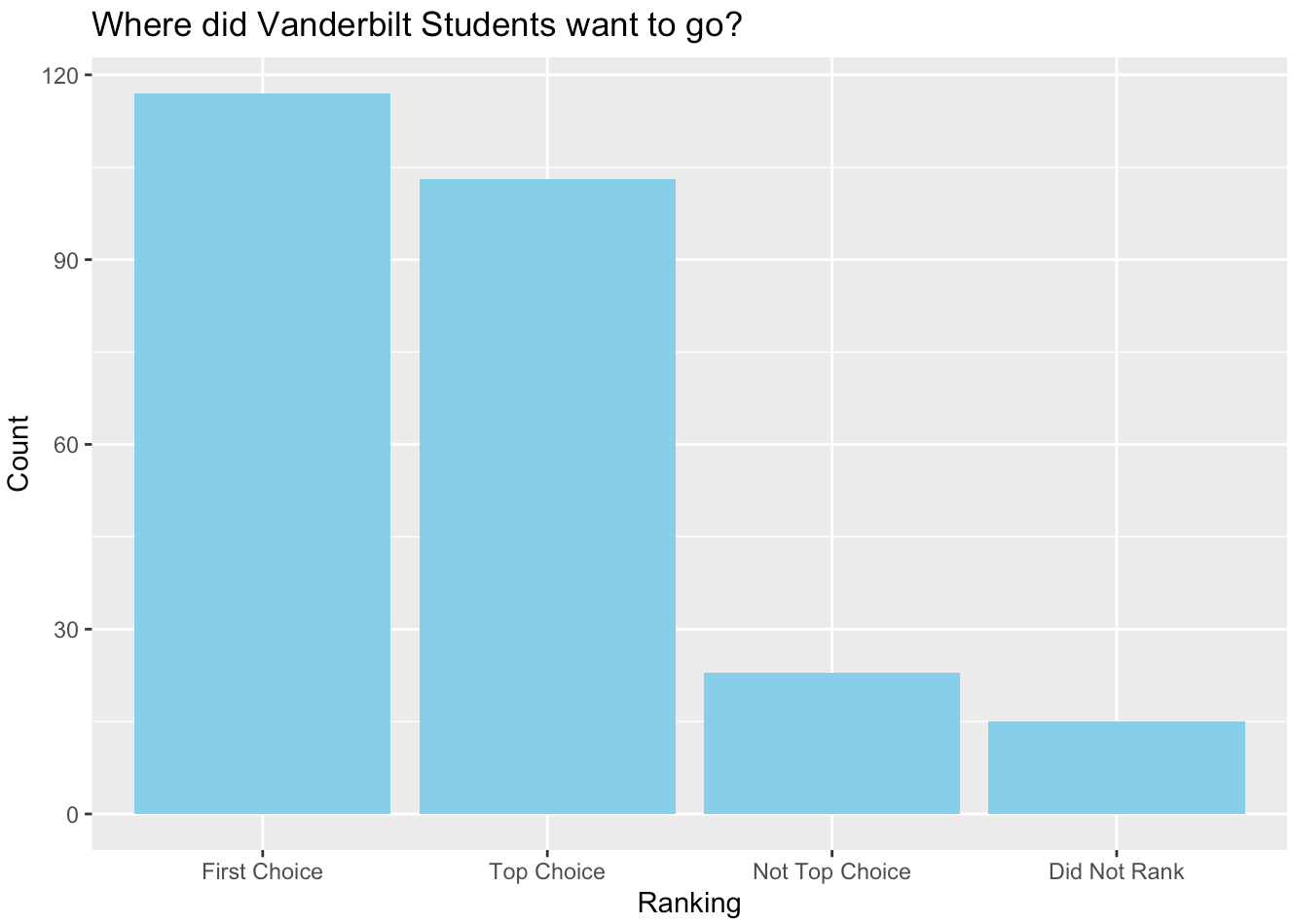
```
df <- read_csv('/Users/ethanxu/Downloads/responses.csv')
```

```
## Rows: 260 Columns: 69
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (55): Timestamp, I confirm that I am a member of Vanderbilt's Class of 2...
## dbl (14): How many colleges and universities did you apply to (including Van...
##
## 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.
```

```
df_cleaned <- df %>%
  filter(`I confirm that I am a member of Vanderbilt's Class of 2028.` == "Yes, I a
m") %>%
  select(
    `Where did Vanderbilt fall in your ranking of all of the schools you applied to?
`,
    `If you took the SAT, what was your score?`,
    `If you took the ACT, what was your score?`,
    `Did you pay for admissions assistance in your college application process?`,
    `Select the top three reasons why you chose to attend Vanderbilt.`,
    `Which of these activities did you attend before coming to Vanderbilt?`
  )

df_cleaned
```

```
## # A tibble: 258 × 6
##    Where did Vanderbilt fall in …¹ If you took the SAT,…² If you took the ACT,…³
##    <chr>                                            <dbl>                   <dbl>
##  1 Vanderbilt was my first choice                    1580                      NA
##  2 Vanderbilt was my first choice                      NA                      36
##  3 Vanderbilt was not my first ch…                   1540                      36
##  4 Vanderbilt was not my first ch…                   1560                      35
##  5 Vanderbilt was my first choice                      NA                      35
##  6 Vanderbilt was not my first ch…                     NA                      NA
##  7 Vanderbilt was my first choice                      NA                      35
##  8 Vanderbilt was my first choice                      NA                      24
##  9 Vanderbilt was my first choice                      NA                      35
## 10 Vanderbilt was my first choice                      NA                      35
## # ℹ 248 more rows
## # ℹ abbreviated names:
## #   ¹`Where did Vanderbilt fall in your ranking of all of the schools you applied
to?`,
## #   ²`If you took the SAT, what was your score?`,
## #   ³`If you took the ACT, what was your score?`
## # ℹ 3 more variables:
## #   `Did you pay for admissions assistance in your college application process?` <
chr>, …
```

```
df_ranking <- df_cleaned %>%
  count(`Where did Vanderbilt fall in your ranking of all of the schools you applied
to?`)

custom_labels <- c("First Choice", "Top Choice", "Not Top Choice", "Did Not Rank")
ggplot(df_ranking, aes(x = reorder(`Where did Vanderbilt fall in your ranking of all
of the schools you applied to?`, -n), y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(title = "Where did Vanderbilt Students want to go?", x = "Ranking", y = "Coun
t") +
  scale_x_discrete(labels = custom_labels)
```



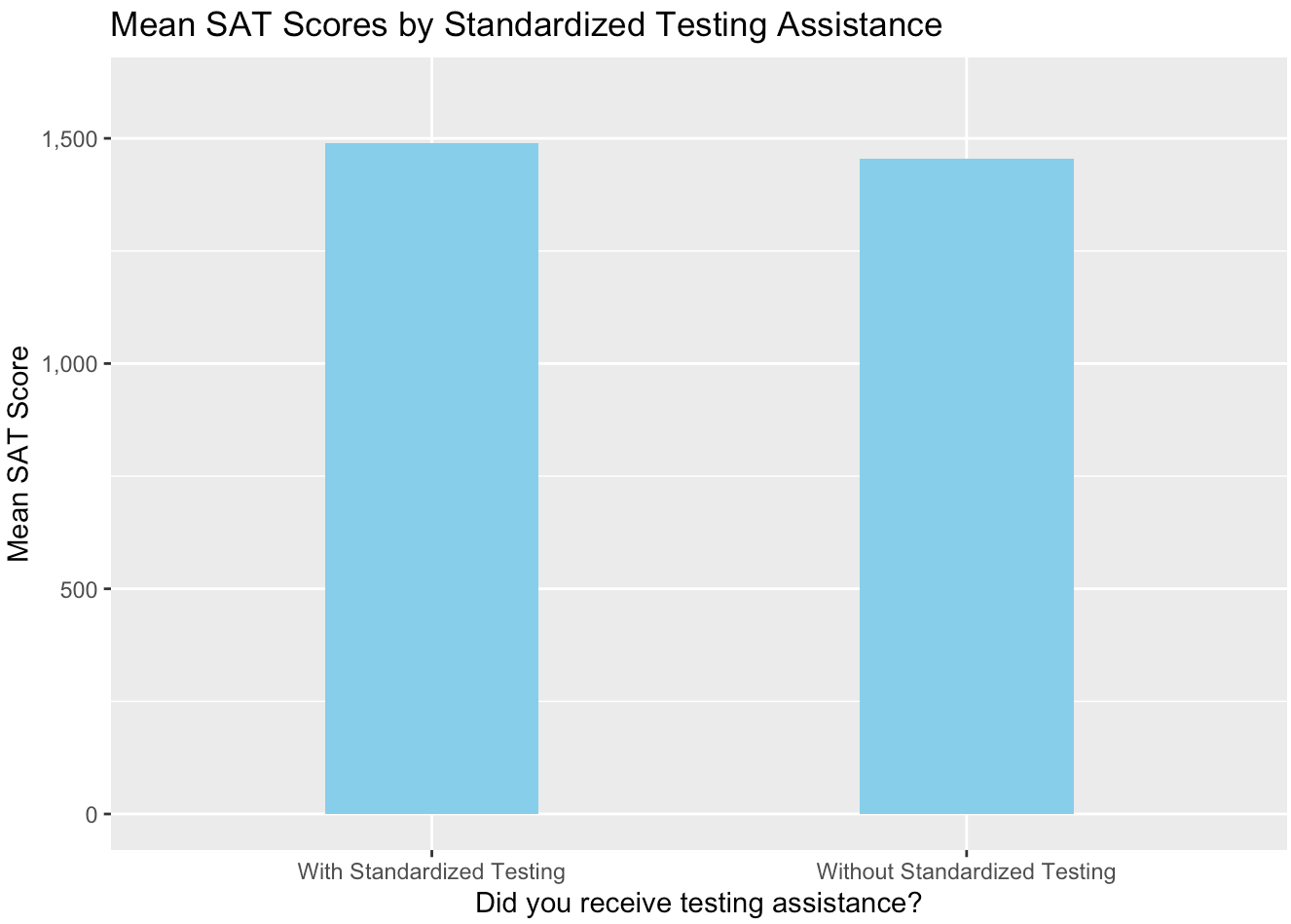

```
adm_counts <- df_cleaned %>%
  count(`Did you pay for admissions assistance in your college application process?`)

adm_counts
```

```
## # A tibble: 7 × 2
##   Did you pay for admissions assistance in your college application proc…¹     n
##   <chr>                                                                   <int>
## 1 No                                                                        147
## 2 No,                                                                         1
## 3 Yes, college counseling                                                    23
## 4 Yes, standardized testing (ACT/SAT) tutoring                               47
## 5 Yes, standardized testing (ACT/SAT) tutoring, Yes, college counseling      38
## 6 Yes, standardized testing (ACT/SAT) tutoring, Yes, college counseling, …    1
## 7 idk                                                                         1
## # ℹ abbreviated name:
## #   ¹`Did you pay for admissions assistance in your college application process?`
```

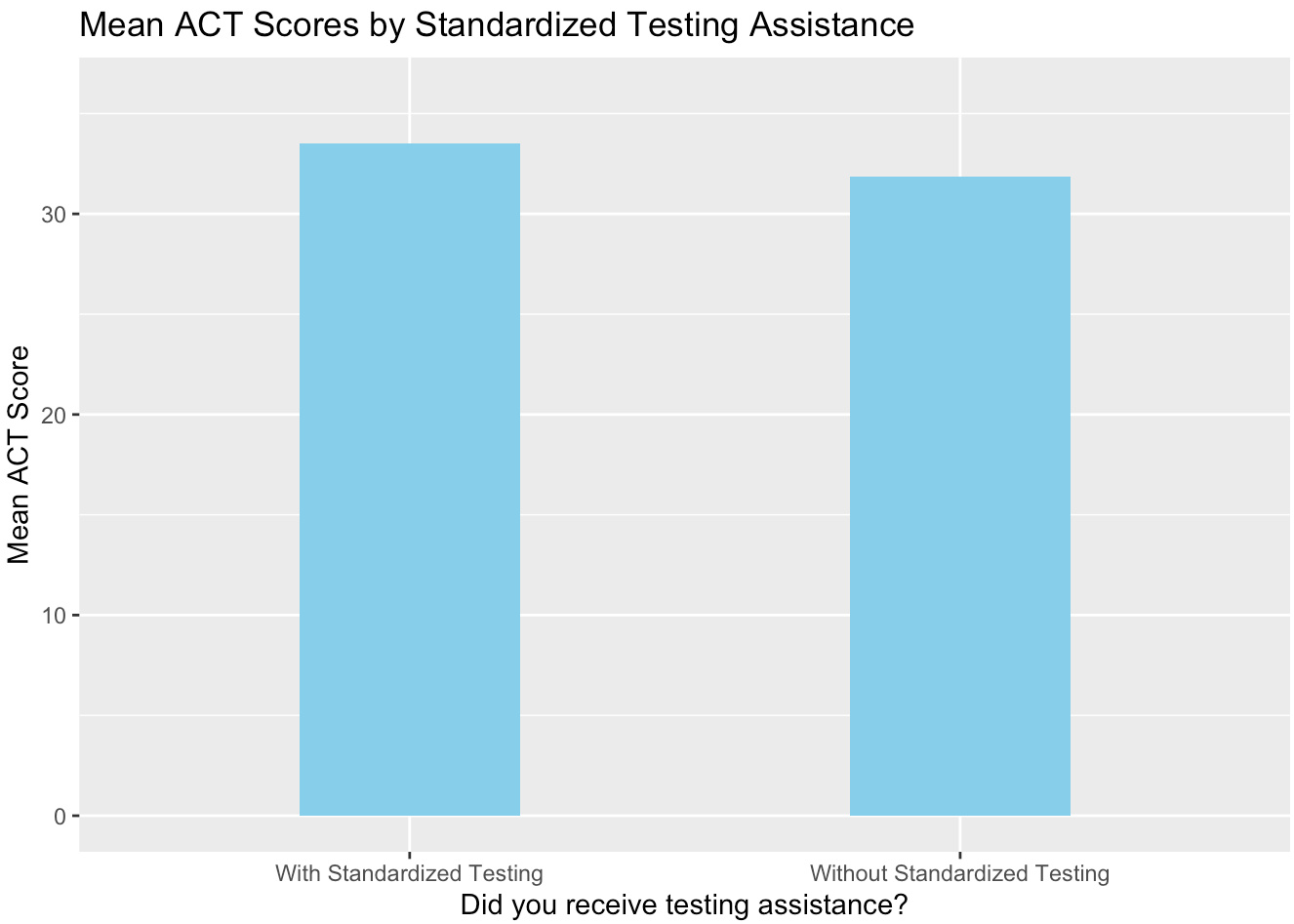
```
df_cleaned <- df_cleaned %>%
  mutate(standardized_testing = if_else(str_detect(`Did you pay for admissions assist
ance in your college application process?`, "standardized testing"), "With Standardiz
ed Testing", "Without Standardized Testing"))

grouped_testing_scores <- df_cleaned %>%
  group_by(standardized_testing) %>%
  summarise(
    mean_sat = mean(`If you took the SAT, what was your score?`, na.rm = TRUE),
    mean_act = mean(`If you took the ACT, what was your score?`, na.rm = TRUE)
  )

grouped_testing_scores
```

```
## # A tibble: 2 × 3
##   standardized_testing         mean_sat mean_act
##   <chr>                           <dbl>    <dbl>
## 1 With Standardized Testing       1490.     33.5
## 2 Without Standardized Testing    1455.     31.9
```

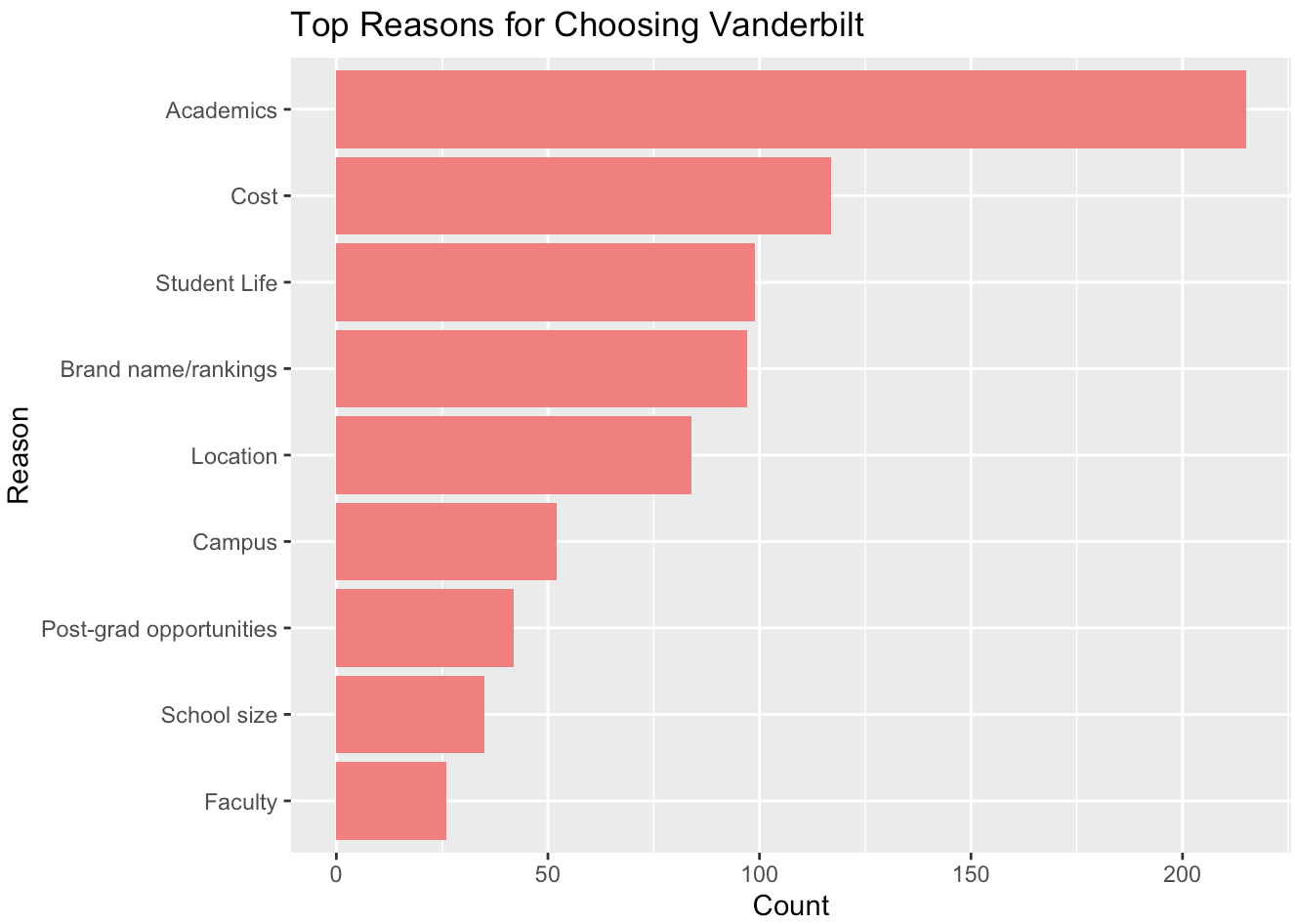
```
ggplot(grouped_testing_scores, aes(x = standardized_testing)) +
  geom_bar(aes(y = mean_sat), stat = "identity", position = position_dodge(), fill =
"skyblue", width = 0.4) +
  labs(title = "Mean SAT Scores by Standardized Testing Assistance",
       x = "Did you receive testing assistance?",
       y = "Mean SAT Score") +
  coord_cartesian(ylim = c(0, 1600)) +
  scale_y_continuous(labels = scales::comma)
```




```
ggplot(grouped_testing_scores, aes(x = standardized_testing)) +
  geom_bar(aes(y = mean_act), stat = "identity", position = position_dodge(), fill =
"skyblue", width = 0.4) +
  labs(title = "Mean ACT Scores by Standardized Testing Assistance",
       x = "Did you receive testing assistance?",
       y = "Mean ACT Score") +
  coord_cartesian(ylim = c(0, 36)) +
  scale_y_continuous(labels = scales::comma)
```




```
reasons <- df_cleaned %>%
  separate_rows(`Select the top three reasons why you chose to attend Vanderbilt.`, s
ep = ", ") %>%
  count(`Select the top three reasons why you chose to attend Vanderbilt.`) %>%
  arrange(desc(n))

ggplot(reasons, aes(x = reorder(`Select the top three reasons why you chose to attend
Vanderbilt.`, n), y = n)) +
  geom_bar(stat = "identity", fill = "lightcoral") +
  labs(title = "Top Reasons for Choosing Vanderbilt", x = "Reason", y = "Count") +
  coord_flip()
```




```
activities <- df_cleaned %>%
  separate_rows(`Which of these activities did you attend before coming to Vanderbil
t?`, sep = ", ") %>%
  count(`Which of these activities did you attend before coming to Vanderbilt?`) %>%
  arrange(desc(n))

activities
```

```
## # A tibble: 6 × 2
##   `Which of these activities did you attend before coming to Vanderbilt?`     n
##   <chr>                                                                   <int>
## 1 Campus tour                                                               129
## 2 Anchor Day (admitted students day)                                         83
## 3 None                                                                       76
## 4 Commodore Launch                                                           50
## 5 'Dore For A Day                                                            28
## 6 Dore For A Day                                                              8
```

```
activities <- activities %>%
  mutate(`Which of these activities did you attend before coming to Vanderbilt?` =
           if_else(`Which of these activities did you attend before coming to Vanderb
ilt?` %in% c("Dore For A Day", "'Dore For A Day"), "'Dore For A Day", `Which of these
activities did you attend before coming to Vanderbilt?`))
activities
```

```
## # A tibble: 6 × 2
##   `Which of these activities did you attend before coming to Vanderbilt?`     n
##   <chr>                                                                   <int>
## 1 Campus tour                                                               129
## 2 Anchor Day (admitted students day)                                         83
## 3 None                                                                       76
## 4 Commodore Launch                                                           50
## 5 'Dore For A Day                                                            28
## 6 'Dore For A Day                                                             8
```

```
ggplot(activities, aes(x = reorder(`Which of these activities did you attend before c
oming to Vanderbilt?`, n), y = n)) +
  geom_bar(stat = "identity", fill = "dodgerblue") +
  labs(title = "Events Attended Before Vanderbilt", x = "Activity", y = "Count") +
  coord_flip()
```