FINAL_PROJECT

Nanda Vihari,Kesav,Bringesh

RESEARCH QUESTION:

Student's perceptions of the effectiveness of college provided resources(in-person classes, library books, and office hours with professors) versus non-college resources (YouTube, ChatGPT, and online articles) on students' academic performance and engagement.

```
#Importing dataset
library(readxl)
res<-read_excel("stats_methods_response.xlsx")
```

DATA EXPLORATION

```
#Shape of the dataset
dim(res)
```

```
## [1] 103   7
```

```
#column names
names(res)
```

```
## [1] "Timestamp"
## [2] "Year of Study"
## [3] "College Department"
## [4] "Frequency of Use of College Resources( classes, library, office hours)"
## [5] "Frequency of Use of Non-College Resources(YouTube, ChatGPT, online articles)\n"
## [6] "Perceived Effectiveness of College Resources in Supporting Academic Success\n"
## [7] "Perceived Effectiveness of Non-College Resources in Supporting Academic Success"
```

```
# Null Values
sum(is.na(res))
```

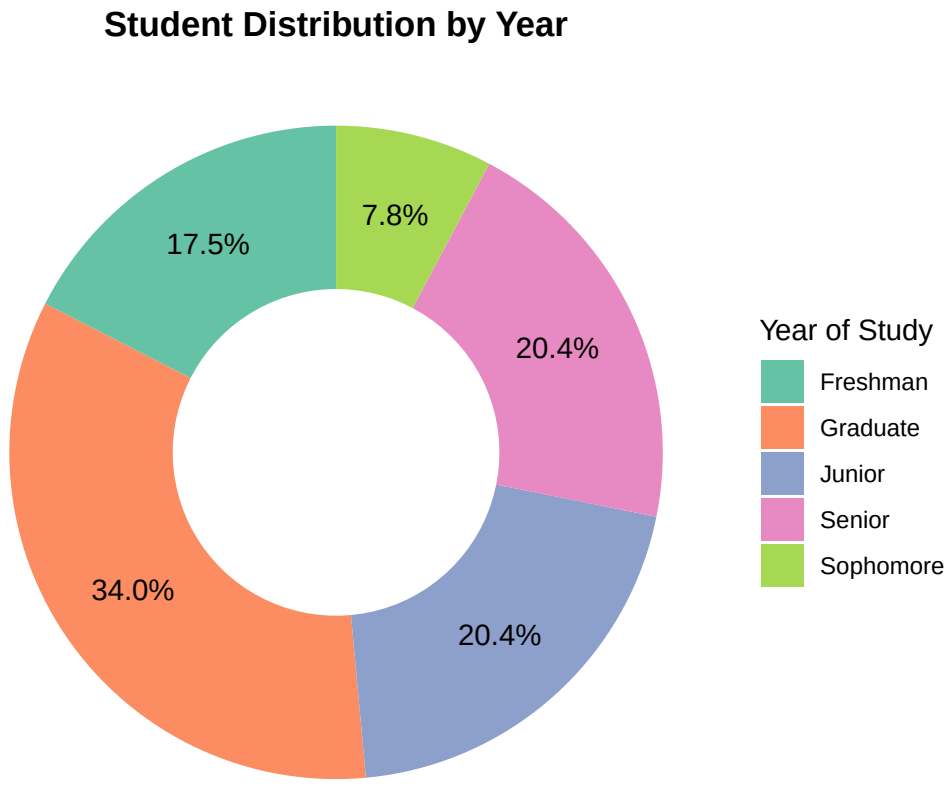
```
## [1] 0
```

```
#Responses grouped by Year of study
responses_by_year <- table(res$`Year of Study`)
responses_by_year
```

```
##
##  Freshman  Graduate    Junior    Senior Sophomore
##        18        35        21        21         8
```

```r
library(dplyr)
library(ggplot2)
year_percentages <- res %>%
  count(`Year of Study`) %>%
  mutate(percentage = n/sum(n) * 100) %>%
  arrange(desc(percentage))

# donut chart
ggplot(year_percentages, aes(x = 2, y = percentage, fill = `Year of Study`)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  geom_text(aes(label = sprintf("%.1f%%", percentage)),
            position = position_stack(vjust = 0.5)) +
  scale_fill_brewer(palette = "Set2") +
  labs(
    title = "Student Distribution by Year",
    fill = "Year of Study"
  ) +
  theme_void() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.position = "right"
  ) +
  xlim(0.5, 2.5)
```



```r
#Responses grouped by College Department
responses_by_dept <- table(res$`College Department`)
responses_by_dept
```

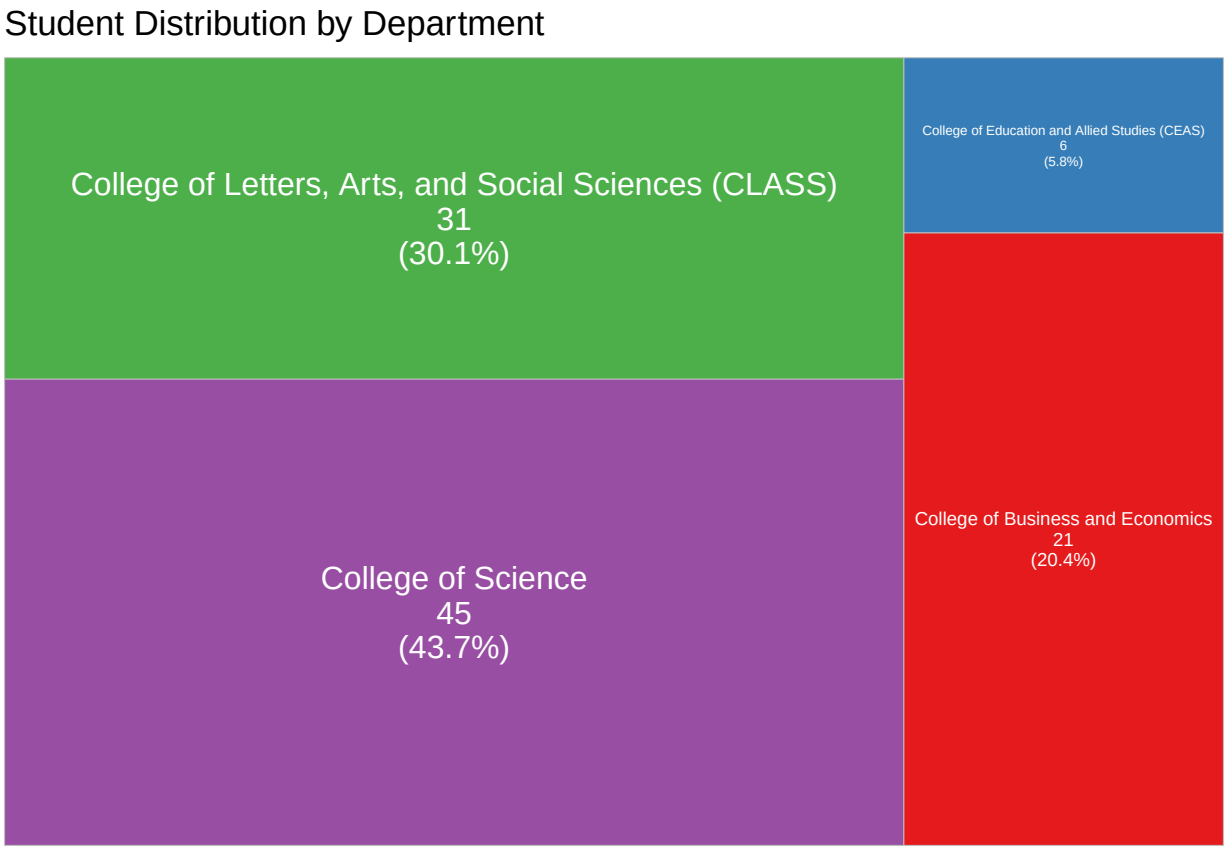
```
##
##                     College of Business and Economics
##                                                    21
##        College of Education and Allied Studies (CEAS)
##                                                     6
## College of Letters, Arts, and Social Sciences (CLASS)
##                                                    31
##                                     College of Science
##                                                    45
```

```r
library(treemapify)

# counts and percentages
dept_summary <- res %>%
    count(`College Department`, name = "count") %>%
    mutate(percentage = round(count/sum(count) * 100, 1)) %>%
    arrange(desc(count))



ggplot(dept_summary,
       aes(area = count,
           fill = `College Department`,
           label = paste0(`College Department`, "\n", count, "\n(", percentage, "%)"))) +
    geom_treemap() +
    geom_treemap_text(colour = "white",
                      place = "centre",
                      size = 12) +
    scale_fill_brewer(palette = "Set1") +
    theme_minimal() +
    labs(title = "Student Distribution by Department") +
    theme(legend.position = "none")
```



DATA CLEANING

```
#removing time stamp column
resp <- res[, -which(names(res) == "Timestamp")]
```

```
#changing categorical variables to factors
resp <- resp %>%
  mutate(across(everything(), factor))
```

GROUPING RESPONSES

Grouping Responses into One variable for better interpretation and easiness.

```
gr <- resp %>%
  mutate(
    student_response = case_when(
      # Both resources are perceived as Effective/Very Effective
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` %in%
        c("Effective", "Very Effective") &
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` %in%
        c("Effective", "Very Effective") ~ "Both_Effective",
```

```
      # Only College resources are Effective/Very Effective, Non-College is not
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` %in%
        c("Effective", "Very Effective") &
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` %in%
        c("Neutral", "Ineffective", "Very Ineffective") ~ "College_Effective",

      # Only Non-College resources are Effective/Very Effective, College is not
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` %in%
        c("Effective", "Very Effective") &
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` %in%
        c("Neutral", "Ineffective", "Very Ineffective") ~ "NonCollege_Effective",

      # Both resources are Ineffective/Very Ineffective
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` %in%
        c("Ineffective", "Very Ineffective") &
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` %in%
        c("Ineffective", "Very Ineffective") ~ "Both_Ineffective",

      # Both resources are Neutral
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` ==
        "Neutral" &
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` ==
        "Neutral" ~ "Both_Neutral",

      # College resources are Neutral, Non-College is Ineffective/Very Ineffective
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` ==
        "Neutral" &
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` %in%
        c("Ineffective", "Very Ineffective") ~ "Neutral_College_NonCollege_Ineffective",

      # Non-College resources are Neutral, College is Ineffective/Very Ineffective
      `Perceived Effectiveness of Non-College Resources in Supporting Academic Success` ==
        "Neutral" &
      `Perceived Effectiveness of College Resources in Supporting Academic Success\n` %in%
        c("Ineffective", "Very Ineffective") ~ "Neutral_NonCollege_College_Ineffective",

      # Any other unspecified cases
      TRUE ~ "Uncategorized"
    )
  )
```

Summary table before bootstrapping

```
library(gt)
library(gtsummary)
gr %>%
  tbl_summary(
    statistic = list(
      all_categorical() ~ "{n} ({p}%)"
    )
  ) %>%
```

```
bold_labels() %>%
italicize_levels() %>%
modify_header(label = "**Variable**") %>%
as_gt() %>%
tab_options(
  table.width = px(600),
  table.font.size = px(12)
)
```

Variable	N = 103[1]
Year of Study	
Freshman	18 (17%)
Graduate	35 (34%)
Junior	21 (20%)
Senior	21 (20%)
Sophomore	8 (7.8%)
College Department	
College of Business and Economics	21 (20%)
College of Education and Allied Studies (CEAS)	6 (5.8%)
College of Letters, Arts, and Social Sciences (CLASS)	31 (30%)
College of Science	45 (44%)
Frequency of Use of College Resources(classes, library, office hours)	
Daily	36 (35%)
Monthly	4 (3.9%)
Never	3 (2.9%)
Rarely	11 (11%)
Weekly	49 (48%)
Frequency of Use of Non-College Resources(YouTube, ChatGPT, online articles)	
Daily	56 (54%)
Monthly	11 (11%)
Never	2 (1.9%)
Rarely	11 (11%)
Weekly	23 (22%)
Perceived Effectiveness of College Resources in Supporting Academic Success	
Effective	55 (53%)
Ineffective	6 (5.8%)
Neutral	29 (28%)
Very Effective	9 (8.7%)
Very Ineffective	4 (3.9%)
Perceived Effectiveness of Non-College Resources in Supporting Academic Success	
Effective	31 (30%)
Ineffective	4 (3.9%)
Neutral	36 (35%)
Very Effective	25 (24%)
Very Ineffective	7 (6.8%)
student_response	
Both_Effective	41 (40%)
Both_Ineffective	5 (4.9%)
Both_Neutral	17 (17%)
College_Effective	23 (22%)
Neutral_College_NonCollege_Ineffective	1 (1.0%)
Neutral_NonCollege_College_Ineffective	1 (1.0%)
NonCollege_Effective	15 (15%)

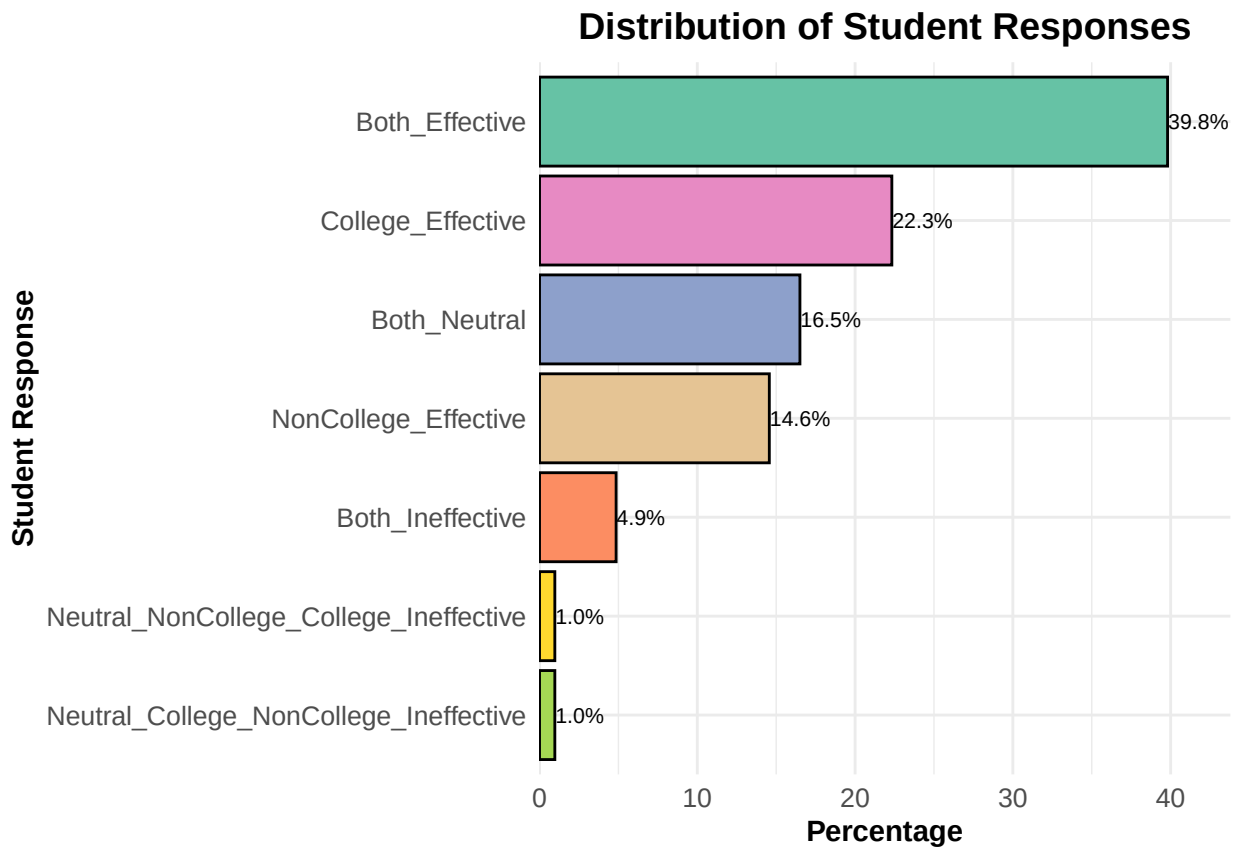
[1] n (%)

```r
library(RColorBrewer)
response_summary <- gr %>%
  count(student_response) %>%
  mutate(percentage = (n / sum(n)) * 100)


colors <- brewer.pal(n = length(unique(response_summary$student_response)), name = "Set2")



ggplot(response_summary, aes(x = reorder(student_response, percentage), y = percentage, fill = student_
  geom_bar(stat = "identity", color = "black") +
  geom_text(aes(label = sprintf("%.1f%%", percentage)), hjust = -0.01, size = 2.7) +
  coord_flip() +
  labs(
    title = "Distribution of Student Responses",
    x = "Student Response",
    y = "Percentage"
  ) +
  theme_minimal() +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.title = element_text(face = "bold"),
    axis.text = element_text(size = 10),
    legend.position = "none"
  ) +
  scale_fill_manual(values = colors)
```



BOOTSTRAP SAMPLING

Generating 1,500 bootstrap samples each of size 103 for introducing more uncertainty and also to get more accurate estimates of the statistic.

```
library(dplyr)

#Bootstraping with double stratification
set.seed(12112001)
bootstrapped_data <- do.call(rbind, lapply(1:1500, function(i) {
  gr %>%
    group_by(`Year of Study`, `College Department`) %>%
    slice_sample(prop = 1, replace = TRUE) %>%
    ungroup()
}))


nrow(bootstrapped_data)
```

```
## [1] 154500
```

Summary table after bootstrap sampling

```
library(gt)
library(gtsummary)
bootstrapped_data %>%
  tbl_summary(
    statistic = list(
      all_categorical() ~ "{n} ({p}%)"
    )
  ) %>%
  bold_labels() %>%
  italicize_levels() %>%
  modify_header(label = "**Variable**") %>%
  as_gt() %>%
  tab_options(
    table.width = px(600),
    table.font.size = px(12)
  )
```

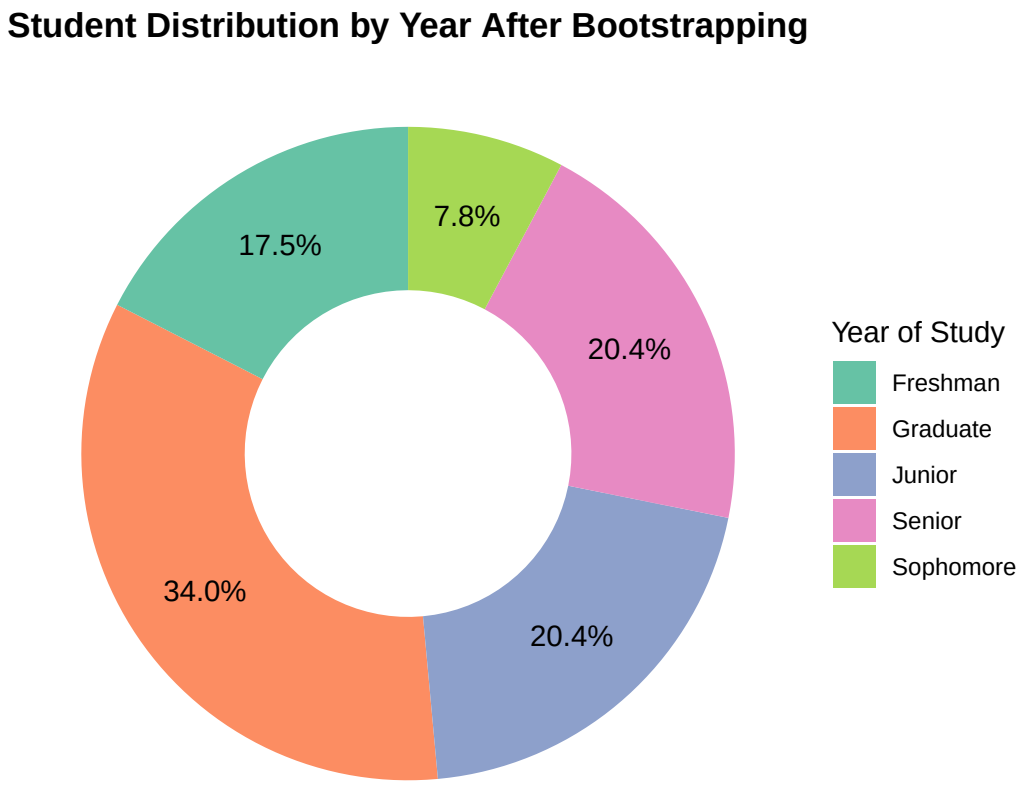
Variable	**N = 154,500**[1]
Year of Study	
Freshman	27,000 (17%)
Graduate	52,500 (34%)
Junior	31,500 (20%)
Senior	31,500 (20%)
Sophomore	12,000 (7.8%)
College Department	
College of Business and Economics	31,500 (20%)
College of Education and Allied Studies (CEAS)	9,000 (5.8%)
College of Letters, Arts, and Social Sciences (CLASS)	46,500 (30%)
College of Science	67,500 (44%)
Frequency of Use of College Resources(classes, library, office hours)	
Daily	54,285 (35%)
Monthly	5,950 (3.9%)
Never	4,380 (2.8%)
Rarely	16,499 (11%)
Weekly	73,386 (47%)
Frequency of Use of Non-College Resources(YouTube, ChatGPT, online articles)	
Daily	83,927 (54%)
Monthly	16,387 (11%)
Never	2,978 (1.9%)
Rarely	16,592 (11%)
Weekly	34,616 (22%)
Perceived Effectiveness of College Resources in Supporting Academic Success	
Effective	82,653 (53%)
Ineffective	9,105 (5.9%)
Neutral	43,438 (28%)
Very Effective	13,394 (8.7%)
Very Ineffective	5,910 (3.8%)
Perceived Effectiveness of Non-College Resources in Supporting Academic Success	
Effective	46,402 (30%)
Ineffective	5,973 (3.9%)
Neutral	54,109 (35%)
Very Effective	37,668 (24%)
Very Ineffective	10,348 (6.7%)

student__response	
Both_Effective	61,435 (40%)
Both_Ineffective	7,443 (4.8%)
Both_Neutral	25,412 (16%)
College_Effective	34,612 (22%)
Neutral_College_NonCollege_Ineffective	1,444 (0.9%)
Neutral_NonCollege_College_Ineffective	1,519 (1.0%)
NonCollege_Effective	22,635 (15%)

[1] n (%)

```
#Dist of year of study after bootstrapping
year_percentage <- bootstrapped_data %>%
  count(`Year of Study`) %>%
  mutate(percentage = n/sum(n) * 100) %>%
  arrange(desc(percentage))

# donut chart
ggplot(year_percentage, aes(x = 2, y = percentage, fill = `Year of Study`)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y", start = 0) +
  geom_text(aes(label = sprintf("%.1f%%", percentage)),
            position = position_stack(vjust = 0.5)) +
  scale_fill_brewer(palette = "Set2") +
  labs(
    title = "Student Distribution by Year After Bootstrapping",
    fill = "Year of Study"
  ) +
  theme_void() +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold"),
    legend.position = "right"
  ) +
  xlim(0.5, 2.5)
```

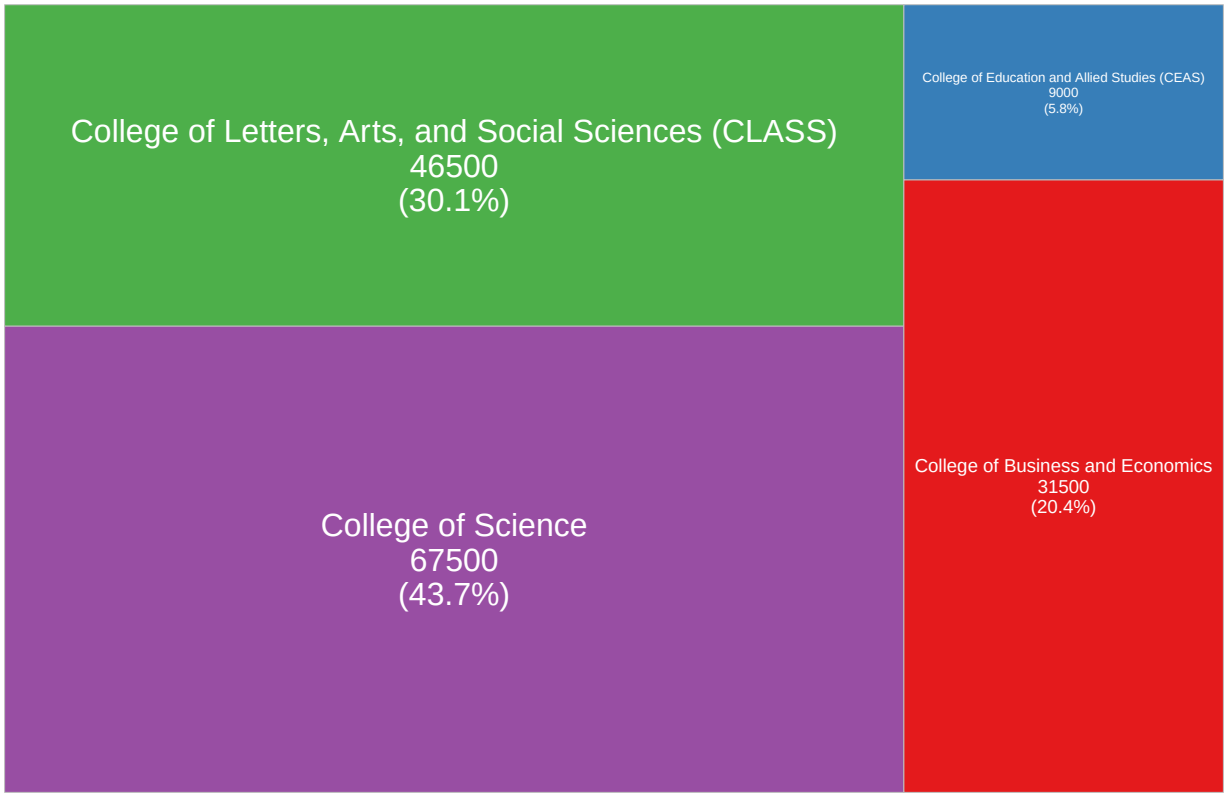



```
dep_summary <- bootstrapped_data %>%
    count(`College Department`, name = "count") %>%
    mutate(percentage = round(count/sum(count) * 100, 1)) %>%
    arrange(desc(count))



ggplot(dep_summary,
       aes(area = count,
           fill = `College Department`,
           label = paste0(`College Department`, "\n", count, "\n(", percentage, "%)"))) +
    geom_treemap() +
    geom_treemap_text(colour = "white",
                      place = "centre",
                      size = 12) +
    scale_fill_brewer(palette = "Set1") +
    theme_minimal() +
    labs(title = "Student Distribution by Department") +
    theme(legend.position = "none")
```

Student Distribution by Department



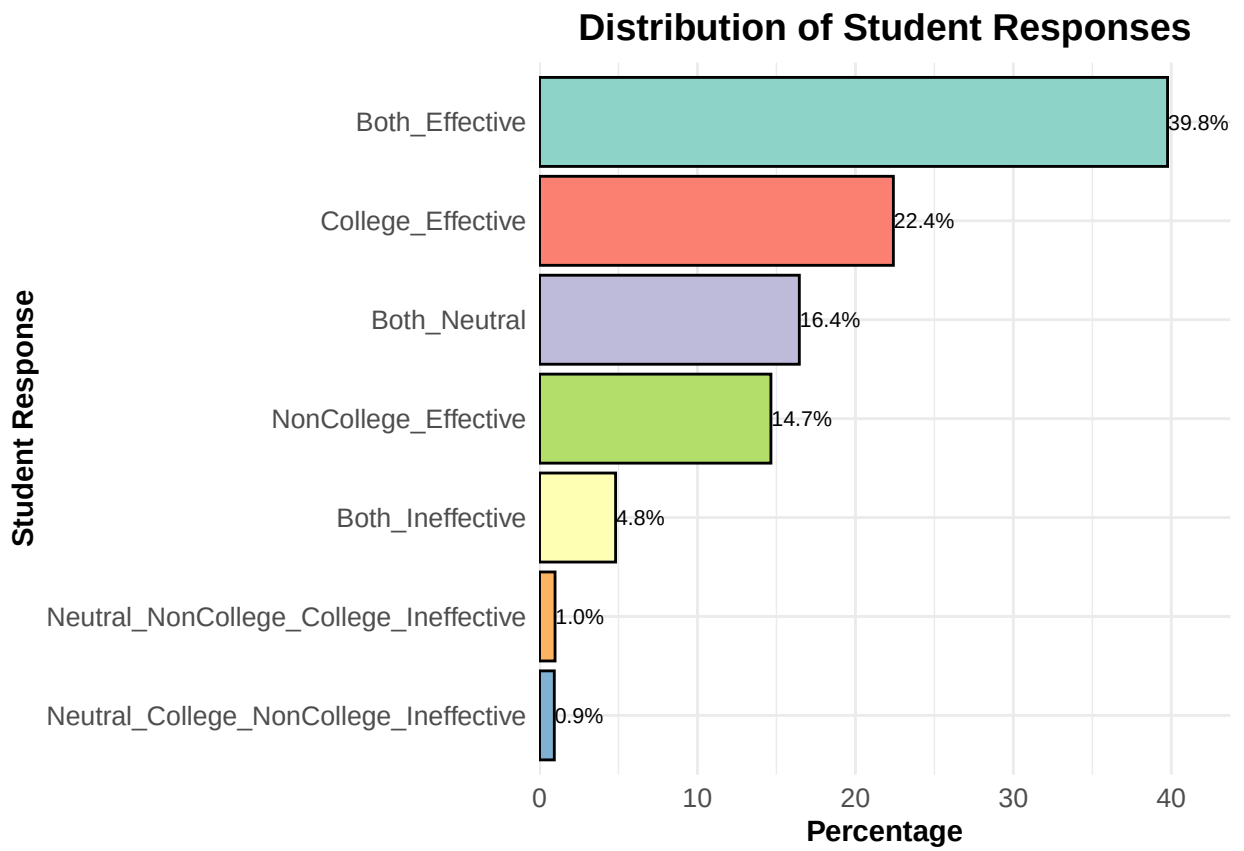

```r
# Distribution of student responses after bootstrap sampling
library(RColorBrewer)
response_summary <- bootstrapped_data %>%
  count(student_response) %>%
  mutate(percentage = (n / sum(n)) * 100)


colors <- brewer.pal(n = length(unique(response_summary$student_response)), name = "Set3")



ggplot(response_summary, aes(x = reorder(student_response, percentage), y = percentage, fill = student_
  geom_bar(stat = "identity", color = "black") +
  geom_text(aes(label = sprintf("%.1f%%", percentage)), hjust = -0.01, size = 2.7) +
  coord_flip() +
  labs(
    title = "Distribution of Student Responses",
    x = "Student Response",
    y = "Percentage"
  ) +
  theme_minimal() +
  scale_y_continuous(expand = expansion(mult = c(0, 0.1))) +
  theme(
    plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
    axis.title = element_text(face = "bold"),
    axis.text = element_text(size = 10),
```

```
    legend.position = "none"
  ) +
  scale_fill_manual(values = colors)
```



Hypothesis Testing

Which do students consider more effective: college or non-college resources?

H_0: The proportion of students who responded college effective is the same as those who considered non-college effective.
H_A: There is difference in proportion of students who responded college effective is the same as those who considered non-college effective.

Conditions:

Independence: Independence is mostly satisfied,Because we took responses from random people at random places around the campus.

Large sample size:

```
table(bootstrapped_data$student_response)
```

```
##
##                          Both_Effective                          Both_Ineffective
##                                   61435                                      7443
##                            Both_Neutral                       College_Effective
##                                   25412                                     34612
## Neutral_College_NonCollege_Ineffective Neutral_NonCollege_College_Ineffective
##                                    1444                                     1519
##                     NonCollege_Effective
##                                   22635
```

```r
# Counts of responses in each category
both_effective <- 61435
college_effective <- 34612
noncollege_effective <- 22635
both_neutral <- 25412

# Spliting Both_Effective evenly between college and non-college
college_success <- college_effective + (both_effective / 2)
noncollege_success <- noncollege_effective + (both_effective / 2)

# Totals for each group include successes + Both_Neutral
college_total <- college_success + both_neutral
noncollege_total <- noncollege_success + both_neutral

# Combining successes and totals into vectors for prop.test
successes <- c(college_success, noncollege_success)
totals <- c(college_total, noncollege_total)

# Checking large sample size conditions
p_hat <- successes / totals
n_p <- totals * p_hat
n_1_minus_p <- totals * (1 - p_hat)


if (all(n_p >= 5) & all(n_1_minus_p >= 5)) {
  print("Large sample size condition met for both groups")
} else {
  print("Large sample size condition NOT met")
}
```

```
## [1] "Large sample size condition met for both groups"
```

Test and Results

```r
#Proportion test
prop.test(successes, totals, alternative = "two.sided", correct = FALSE)
```

```
##
##  2-sample test for equality of proportions without continuity correction
##
## data:  successes out of totals
```

```
## X-squared = 364.22, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.03820329 0.04696549
## sample estimates:
##    prop 1    prop 2
## 0.7199517 0.6773673
```

Since p-value is less than our significance level,we can conclude that there is strong evidence that the proportions of college_effective and Non-college effective responses are different.
.

We are 95% confident that difference between the proportions of the two groups lies b/w 3.82% and 4.70% i.e college_effective will have 3.82%-4.70% higher responses than Non-college_effective responses

Do students really Prefer college resources and online tools are equally helpful, or do they have a clear favorite?

H_0:The proportion of students finding both resources effective equals the proportion finding only one type effective.
H_A:There is a significant difference in these proportions.

Conditions

Independence: Independence is satisfied above.

Large sample size:

```
#total
total_responses <- both_effective + college_effective + noncollege_effective

# proportions
p_both <- both_effective / total_responses
p_college <- college_effective / total_responses
p_noncollege <- noncollege_effective / total_responses

n_pr <- c(
  both = total_responses * p_both,
  college = total_responses * p_college,
  noncollege = total_responses * p_noncollege
)

n_1_minus_pr <- c(
  both = total_responses * (1 - p_both),
  college = total_responses * (1 - p_college),
  noncollege = total_responses * (1 - p_noncollege)
)


# Checking large sample size condition
if (all(n_pr >= 5) & all(n_1_minus_pr >= 5)) {
```

```
  cat("Large sample size condition met for all groups.\n")
} else {
  cat("Large sample size condition NOT met.\n")
}
```

```
## Large sample size condition met for all groups.
```

Tests and Results

```
#Proportion test b/w both_effective and college_effective
prop.test(
  x = c(both_effective, college_effective),
  n = c(total_responses,total_responses)
)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(both_effective, college_effective) out of c(total_responses, total_responses)
## X-squared = 12581, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.2221559 0.2298587
## sample estimates:
##    prop 1    prop 2
## 0.5176438 0.2916365
```

Since p-value is less than our significance level,we can conclude that there is strong evidence that the proportions of both_effective responses and college_effective are different.

.

We are 95% confident that difference between the proportions of the two groups lies b/w 22.2% and 22.9% i.e both_effective will have 22.2%-22.9% higher responses than college_effective responses.

```
#Proportion test b/w both_effective and non-college_effective
prop.test(
  x = c(both_effective, noncollege_effective),
  n = c(total_responses,total_responses)
)
```

```
##
##  2-sample test for equality of proportions with continuity correction
##
## data:  c(both_effective, noncollege_effective) out of c(total_responses, total_responses)
## X-squared = 27726, df = 1, p-value < 2.2e-16
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.3232993 0.3305488
## sample estimates:
##    prop 1    prop 2
## 0.5176438 0.1907197
```

Since p-value is less than our significance level,we can conclude that there is strong evidence that the proportions of both_effective responses and Non-college_effective are different.
.

We are 95% confident that difference between the proportions of the two groups lies b/w 32.32% and 33.05% i.e both_effective will have 32.32%-33.05% higher responses than college_effective responses.

By both hypothesis we can say that we have evidence that there is significant difference in proportion of students finding both resources effective and the proportion finding only one type effective and also we can say that proportion of students finding both resources effective is greater than only one type effective.

Does the effectiveness of college and non-college resources vary across different years of study?

H_0:There is no association b/w the Year of study and Effectiveness. H_A:There is association b/w the Year of study and Effectiveness.

Conditions

Independence : Independence is satisfied **Expected frequencies** :

```
# contingency table
mytable <- table(bootstrapped_data$`Year of Study`, bootstrapped_data$student_response)

row_totals <- rowSums(mytable)
col_totals <- colSums(mytable)
n <- sum(mytable)

# expected frequencies
expected <- outer(row_totals, col_totals) / n

print(expected)
```

```
##           Both_Effective Both_Ineffective Both_Neutral College_Effective
## Freshman        10736.21        1300.7184     4440.932          6048.699
## Graduate        20875.97        2529.1748     8635.146         11761.359
## Junior          12525.58        1517.5049     5181.087          7056.816
## Senior          12525.58        1517.5049     5181.087          7056.816
## Sophomore        4771.65         578.0971     1973.748          2688.311
##           Neutral_College_NonCollege_Ineffective
## Freshman                                252.3495
## Graduate                                490.6796
## Junior                                  294.4078
## Senior                                  294.4078
## Sophomore                               112.1553
##           Neutral_NonCollege_College_Ineffective NonCollege_Effective
## Freshman                                265.4563             3955.631
## Graduate                                516.1650             7691.505
## Junior                                  309.6990             4614.903
## Senior                                  309.6990             4614.903
## Sophomore                               117.9806             1758.058
```

All are greater than 5,This condition is satisfied.

Test and Results

```
#chi-sq test
chisq.test(mytable)
```

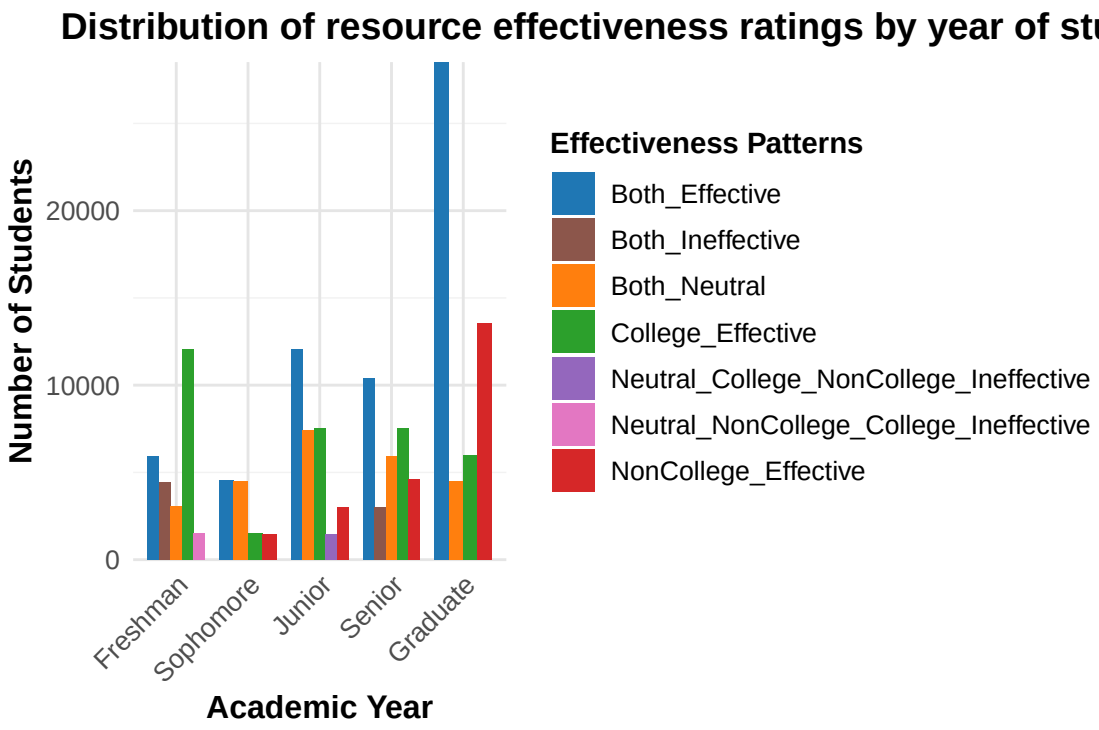
```
##
##  Pearson's Chi-squared test
##
## data:  mytable
## X-squared = 56941, df = 24, p-value < 2.2e-16
```

Since p-value is less than our significance level,we can conclude that there is association b/w the variables Year of study and student response.

```
bootstrapped_data$`Year of Study` <- factor(
  bootstrapped_data$`Year of Study`,
  levels = c("Freshman", "Sophomore", "Junior", "Senior", "Graduate"),
  ordered = TRUE
)

#plot
ggplot(bootstrapped_data, aes(x = `Year of Study`, fill = student_response)) +
  geom_bar(position = "dodge", width = 0.8) +
  scale_fill_manual(values = c(
    "Both_Effective" = "#1f77b4",
    "Both_Neutral" = "#ff7f0e",
    "College_Effective" = "#2ca02c",
    "NonCollege_Effective" = "#d62728",
    "Both_Ineffective" = "#8c564b",
    "Neutral_College_NonCollege_Ineffective" = "#9467bd",
    "Neutral_NonCollege_College_Ineffective" = "#e377c2"
  )) +
  theme_minimal() +
  labs(
    title = "Distribution of resource effectiveness ratings by year of study",
    x = "Academic Year",
    y = "Number of Students",
    fill = "Effectiveness Patterns"
  ) +
  theme(
    plot.title = element_text(size = 14, face = "bold", hjust = 0.1),
    axis.title = element_text(size = 12, face = "bold"),
    axis.text = element_text(size = 10),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.title = element_text(size = 11, face = "bold"),
    legend.text = element_text(size = 10),
    legend.position = "right",
    panel.grid.major = element_line(color = "grey90"),
    panel.grid.minor = element_line(color = "grey95"),
```

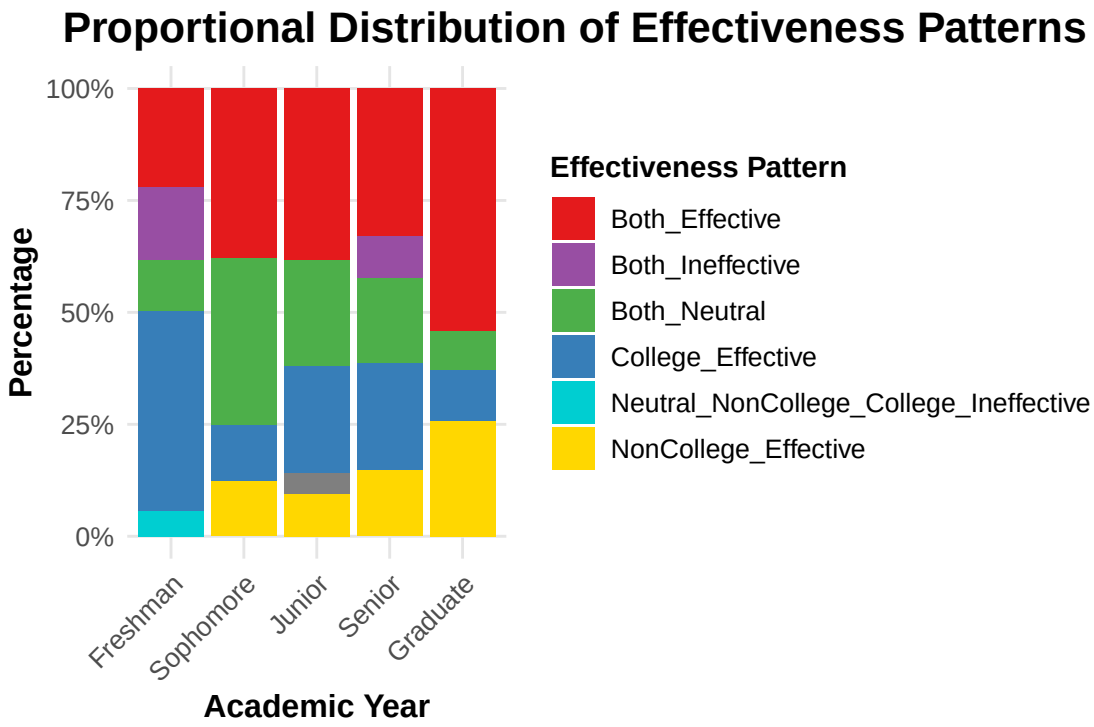
```
    plot.margin = margin(1, 1, 1, 1, "cm")
  ) +
  scale_y_continuous(expand = c(0, 0)) +
  coord_cartesian(clip = "off")
```



From the plot, we can see as the students are advance in Year of study , there is upward trend in both effective counts.

```
# stacked percentage plot
ggplot(bootstrapped_data, aes(x = `Year of Study`, fill = student_response)) +
  geom_bar(position = "fill") +
  scale_y_continuous(labels = scales::percent) +
  scale_fill_manual(values = c(
    "Both_Effective" = "#E41A1C",
    "Both_Neutral" = "#4DAF4A",
    "College_Effective" = "#377EB8",
    "NonCollege_Effective" = "#FFD700",
    "Both_Ineffective" = "#984EA3",
    "Other" = "#FF7F00",
    "Neutral_NonCollege_College_Ineffective" = "#00CED1"
  )) +
  theme_minimal() +
  labs(
    title = "Proportional Distribution of Effectiveness Patterns",
    x = "Academic Year",
    y = "Percentage",
```

```
    fill = "Effectiveness Pattern"
  ) +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 0.1),
    plot.subtitle = element_text(size = 12, hjust = 0.5),
    axis.title = element_text(size = 12, face = "bold"),
    axis.text = element_text(size = 10),
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.title = element_text(size = 11, face = "bold"),
    legend.text = element_text(size = 10),
    legend.position = "right",
    panel.grid.major = element_line(color = "grey90"),
    panel.grid.minor = element_blank(),
    plot.margin = margin(1, 1, 1, 1, "cm")
  )
```



Does the effectiveness of college and non-college resources vary across different Departments?

H_0:There is no association b/w college department and student response. H_A:There is association b/w college department and response.

Conditions

Independence : Independence is satisfied **Expected frequencies** :

```
#contingency table
contingency_tab <- table(bootstrapped_data$`College Department`, bootstrapped_data$student_response)

row_total <- rowSums(contingency_tab)
col_total <- colSums(contingency_tab)
n_t <- sum(mytable)

# expected frequencies
expected_freq <- outer(row_total, col_total) / n_t

print(expected_freq)
```

```
##                                                    Both_Effective
## College of Business and Economics                       12525.583
## College of Education and Allied Studies (CEAS)           3578.738
## College of Letters, Arts, and Social Sciences (CLASS)   18490.146
## College of Science                                      26840.534
##                                                    Both_Ineffective
## College of Business and Economics                        1517.5049
## College of Education and Allied Studies (CEAS)            433.5728
## College of Letters, Arts, and Social Sciences (CLASS)    2240.1262
## College of Science                                       3251.7961
##                                                    Both_Neutral
## College of Business and Economics                      5181.087
## College of Education and Allied Studies (CEAS)         1480.311
## College of Letters, Arts, and Social Sciences (CLASS)  7648.272
## College of Science                                    11102.330
##                                                    College_Effective
## College of Business and Economics                          7056.816
## College of Education and Allied Studies (CEAS)             2016.233
## College of Letters, Arts, and Social Sciences (CLASS)     10417.204
## College of Science                                        15121.748
##                                                    Neutral_College_NonCollege_Ineffective
## College of Business and Economics                                                294.4078
## College of Education and Allied Studies (CEAS)                                    84.1165
## College of Letters, Arts, and Social Sciences (CLASS)                            434.6019
## College of Science                                                               630.8738
##                                                    Neutral_NonCollege_College_Ineffective
## College of Business and Economics                                               309.69903
## College of Education and Allied Studies (CEAS)                                   88.48544
## College of Letters, Arts, and Social Sciences (CLASS)                           457.17476
## College of Science                                                              663.64078
##                                                    NonCollege_Effective
## College of Business and Economics                              4614.903
## College of Education and Allied Studies (CEAS)                 1318.544
## College of Letters, Arts, and Social Sciences (CLASS)          6812.476
## College of Science                                             9889.078
```

All are greater than 5,This condition is satisfied.

Test and Results

```r
#chi-sq test
chisq.test(contingency_tab)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_tab
## X-squared = 24437, df = 18, p-value < 2.2e-16
```

Since p-value is less than our significance level,we can conclude that there is association b/w the variables college department and student response.

```r
library(forcats)

# Combining less significant categories into "Other"
gr <- gr %>%
  mutate(
    student_response = fct_lump(student_response, n = 6) # Group smaller categories into "Other"
  )

# Reordering the departments by total student count
gr <- gr %>%
  group_by(`College Department`) %>%
  mutate(
    total_count = n()
  ) %>%
  ungroup() %>%
  mutate(
    `College Department` = fct_reorder(`College Department`, total_count) # Reorder by total count
  )

#plot
ggplot(gr, aes(x = `College Department`, fill = student_response)) +
  geom_bar(position = "dodge", width = 0.8) +
  scale_fill_manual(
    name = "Effectiveness Pattern",
    values = c(
      "Both_Effective" = "#FF5733",
      "Both_Neutral" = "#33FF57",
      "Both_Ineffective" = "#5733FF",
      "College_Effective" = "#FFC300",
      "NonCollege_Effective" = "#FF33A8",
      "Other" = "#33D4FF"
    )
  ) +
  theme_minimal() +
  labs(
    title = "Distribution of Resource Effectiveness Across Departments",
    x = "College Department",
    y = "Number of Students"
```

```
  ) +
  theme(
    plot.title = element_text(size = 16, face = "bold", hjust = 1),
    axis.text.x = element_text(size = 10, angle = 45, hjust = 1),
    axis.text.y = element_text(size = 10),
    legend.position = "bottom",
    legend.title = element_text(size = 10),
    legend.text = element_text(size = 8),
    legend.key.width = unit(0.5, "cm"),
    legend.margin = margin(t = 5, r = 150, b = 5, l = 2)
  ) +
  coord_flip()
```