ACLU Questions

Shelby Chapa

2023-07-18

1. Please provide a table of the top 3 banned books. The table should have a column for the title of the book, its author(s), the number of occurrences of that book, and the percentage of occurrences of each book in the dataset. Please arrange the table in ascending order. Do you notice any commonalities among the books in this list?

In my research on the top 3 banned books, I discovered a commonality among all of them: they explore LGBTQIA+ themes or content. I also observed a shared characteristic among the authors, as each author belongs to a marginalized community or possesses an intersectional identity.

```r
# Read the dataset
bannedbook<-read.csv("C:/Users/shelb/OneDrive/Desktop/aclu_analytics_skills_test/banned_books.csv")
# Load libraries
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.3
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 4.2.3
```

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.2.3
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```r
# Create a dataframe with title and authors
books <- data.frame(
  Title = bannedbook$title,
```

```r
  Authors = bannedbook$authors
)

# Calculate occurrences of each book
book_count <- count(books, Title, Authors, sort = TRUE)

# Total occurrence calculation (sum of all book titles)
total_occurrence <- sum(book_count$n)

# Calculate percentage of total occurrence and add it to the dataframe
book_count <- mutate(book_count, Percentage = n / total_occurrence * 100)

# Top 3 banned books
top_books <- head(book_count, 3)

# Arrange in ascending order by title
top_books <- top_books[order(top_books$Percentage), ]

# Format the percentage column to percentage values to two decimal places
top_books$Percentage <- sprintf("%.2f", top_books$Percentage)

# Remove row names
rownames(top_books) <- NULL

# Print the table for "Title", "Authors", "n"(occurrences),and "Percentage_%"
print(top_books, row.names = FALSE)

## Title                    Authors              n      Percentage_%
## Out of Darkness          Pérez, Ashley Hope   23        0.91
## All Boys Aren't Blue     Johnson, George M.   29        1.15
## Gender Queer: A Memoir   Kobabe, Maia         41        1.62
```

2. Graph the number of bans over time (from July 1 , 2021 to June 30, 2022) by the type of ban (type_of_ban). Briefly describe the trends that you see and explain why you chose the particular graph type (line graph, bar graph, pie chart, etc.) used in your answer. The graph does not need to be pretty, as long as the trends are clear to see!

The most prominent trend I observed is on 05/31/2022, this date exhibited the highest number of bans in the classroom and library, as well as the greatest number of pending banned books. To effectively convey the changes over time, I opted for a line graph. This graph type allows for the identification of trends, facilitates visual comparison of various types of bans across time, and aids in detecting any outliers that may be present.

```r
library(ggplot2)
# Convert the publishing_date column to Date type using as.Date format
bannedbook$publishing_date <- as.Date(bannedbook$publishing_date)
# Filter the data for the specified time period (July 1, 2021, to June 30, 20
22)
start_date <- as.Date("2021-07-01")
end_date <- as.Date("2022-06-30")
#Filter the dates of the banned books and publishing dates to the start and e
```

```
nd dates
filtered_data<-bannedbook[bannedbook$publishing_date>=start_date &bannedbook$
publishing_date<=end_date,]
# Group the filtered data by the type of ban, publishing date, and count the
number of bans for each type
ban_counts <- with(filtered_data, table(publishing_date, type_of_ban))
# Convert the table to a data frame
ban_counts_df<- as.data.frame.table(ban_counts)
names(ban_counts_df) <- c("Date", "Type_of_Ban", "Number_of_Bans")
```


Number of Bans Over Time (July 2021 - June 2022)
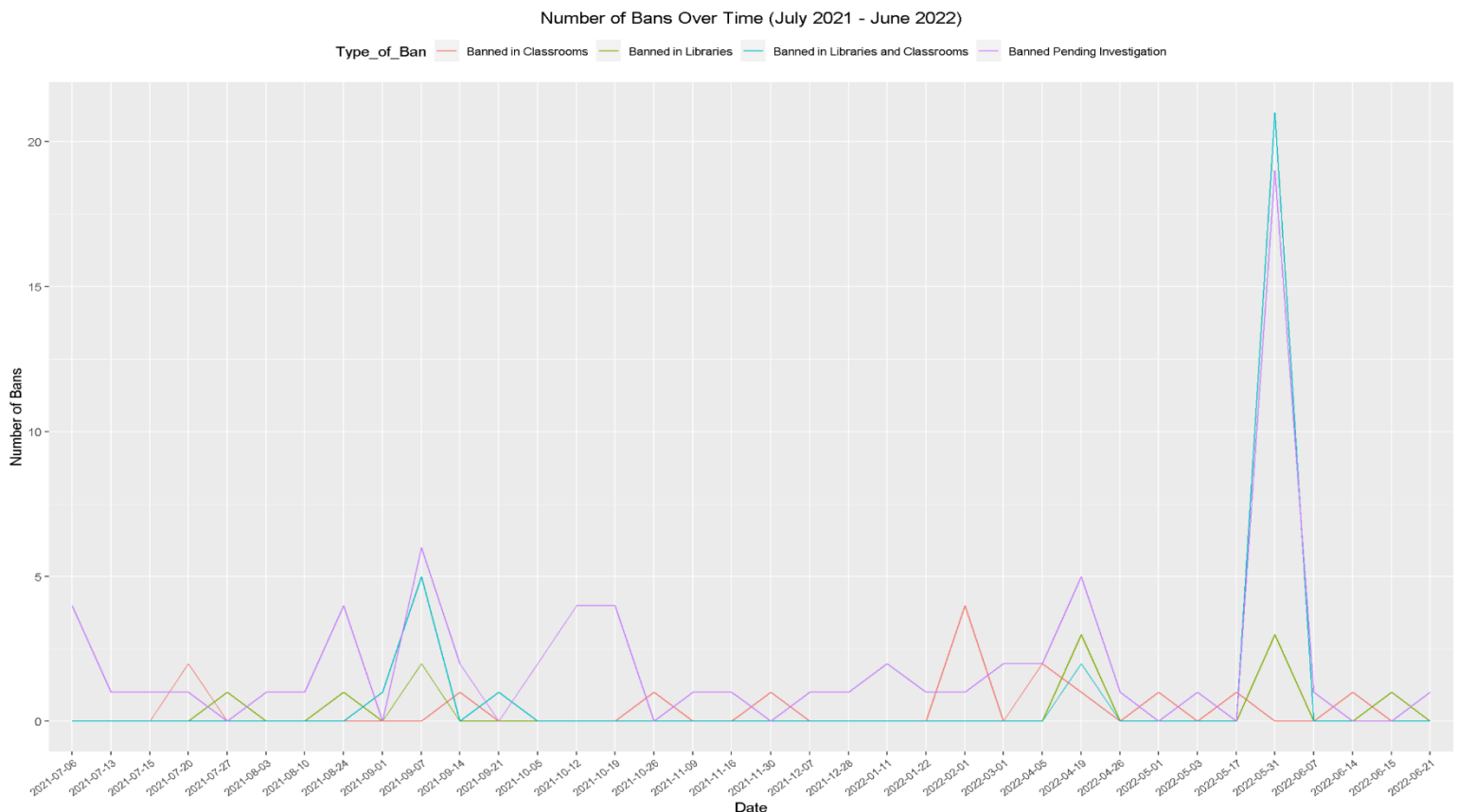
```
# Plot the line graph - Best to identify changes over time. Use angle and hig
ht adjustments for legibility, change font to prevent date crowding
ggplot(ban_counts_df, aes(Date, Number_of_Bans, color = Type_of_Ban, group=Ty
pe_of_Ban))+
  geom_line()+labs(x = "Date", y = "Number of Bans", title = "Number of Bans
Over Time (July 2021 - June 2022)") + theme(legend.position = "top", axis.tex
t.x = element_text(angle = 40, hjust = 1, size = 8), plot.title = element_tex
t(hjust = 0.5))
```

3. Find the author (or authors) who have written or contributed to the greatest number of unique titles that have been banned. Remember that some books have multiple authors. How many books by that author (or authors) have been banned? In addition to providing the

name of the author (or authors), please make sure to provide the code that you wrote to get the correct answer.

The Authors with the greatest number of unique banned titles are as follows: Anh Do, Saadia Faruqi, and Jacqueline Jules. According to the author_count table all three authors have 17 unique banned book titles. The code is as follows:

```
# Seperate authors by semicolons from the rows for each author
author_titles<-bannedbook %>%
  separate_rows(authors, sep = ";") %>%
  select(Author = authors, Title = title)
# Count the number of unique banned (distinct) titles for each author

# Used distinct to count unique combinations of variables.
author_counts<-author_titles %>%
  distinct(Author, Title) %>%
  group_by(Author) %>%
  summarise(Count = n()) %>%
  arrange(desc(Count))
# Find the authors with the maximum count
max_count <- max(author_counts$Count)
top_authors <- author_counts$Author[author_counts$Count == max_count]
# Display the authors with the maximum count and unique titles
print(paste("Authors with the greatest number of unique banned titles:", top_
authors))

## [1] "Authors with the greatest number of unique banned titles: Do, Anh"
## [2] "Authors with the greatest number of unique banned titles: Faruqi, Saa
dia"
## [3] "Authors with the greatest number of unique banned titles: Jules, Jacq
ueline"

print(paste("Number of unique banned titles:", max_count))

## [1] "Number of unique banned titles: 17"
```

#4. 4.1 Imagine you have been asked to advise a lawyer or advocate trying to build a strategy to protect free speech and freedom of expression rights in K-12 schools in a particular state mentioned in this dataset. Based on this dataset, What state would you choose to analyze and why?

Upon examining the dataset, I have identified that Texas is the state with the highest number of banned books, with a total of 801 books banned. Considering this significant quantity, it would be advisable to conduct an analysis of the state. This analysis would support a lawyer or advocate in formulating a strategy to protect free speech and freedom of expression rights in K-12 schools. Moreover, as a Texas resident who has firsthand experience in the state's public school system, I am particularly interested in undertaking this comprehensive analysis.

4.2 What research question would you explore?

Which district is most affected by the book ban? This research question is crucial as it provides insights into the consequences of restricted free speech and expression at a local level. Identifying the district with the greatest impact allows for targeted legal strategies

and advocacy to address specific district and state challenges. Understanding the impact across districts contributes to comprehending the implications of book bans on students' access to diverse ideas and perspectives in education. I was able to asess that North East Independent School District had the highest number of bans at 435 of the 801 total Texas bans, which is roughly 54% of Texas bans.

4.3 What additional data outside of this dataset would you seek to add to this dataset to build a stronger picture?

Alongside the dataset on banned books, it would be valuable to include additional datasets on standardized testing and literacy scores assessing the impact of literacy on Texas K-12 statewide and within North East I.S.D. This data would provide a more robust foundation for understanding potential correlations between book bans, literacy outcomes, and the infringement of free speech rights. The lawyer or advocate can craft a stronger strategy tailored to address the unique challenges faced by K-12 schools in Texas. Promoting literacy education helps foster an informed population capable of engaging in robust free speech.

```
# Calculate the total number of banned books for each state Q 4.1
state_counts <- bannedbook %>%
  group_by(state) %>%
  summarise(total_banned_books = n()) %>%
  arrange(desc(total_banned_books))

# Get the state with the highest number of banned books
max_count_state <- max(state_counts$total_banned_books)
top_state <- state_counts$state[state_counts$total_banned_books == max_count_
state]

# Display the state with the highest number of banned books
print(paste("State with the highest number of banned books:", top_state))

## [1] "State with the highest number of banned books: Texas"

print(paste("Total banned books in the state:", max_count_state))

## [1] "Total banned books in the state: 801"

# Filter the dataset for only the state with the highest number of banned boo
ks
state_data <- filter(bannedbook, state %in% top_state)

# Calculate the total number of banned books for each district
district_counts <- state_data %>%
  group_by(district) %>%
  summarise(total_banned_books = n()) %>%
  arrange(desc(total_banned_books))

# Get the district with the highest number of banned books in Texas- Q 4.2
max_count_district <- max(district_counts$total_banned_books)
top_district <- district_counts$district[district_counts$total_banned_books =
= max_count_district]
```

```
# State and District with the highest number of banned books in Texas, Print
print(paste("District in", top_state, "with the highest number of banned book
s:", top_district))
```

[1] "District in Texas with the highest number of banned books: North East
Independent School District"

```
print(paste("Total banned books in the district:", max_count_district))
```

[1] "Total banned books in the district: 435"

#4.4 (Extra Visualizations for Texas and district banned books)

```
# Bar graph for Texas banned books
state_bar <- ggplot(state_counts, aes(x = state, y = total_banned_books)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  xlab("State") +
  ylab("Total Banned Books") +
  ggtitle("Total Banned Books per State") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, size=8))
print(state_bar)
```
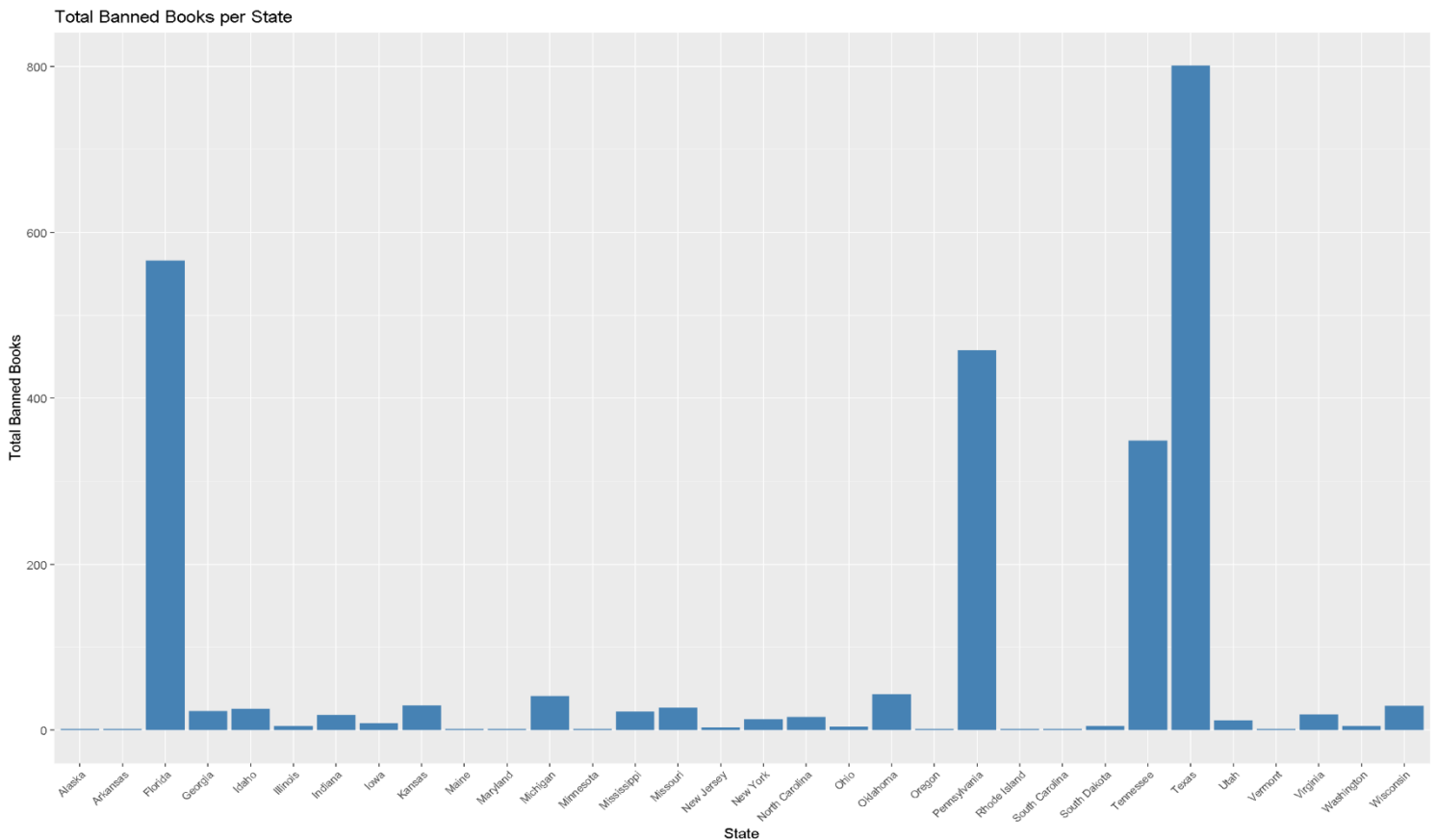
```
# Bar graph for district-level banned books in Texas
district_bar <- ggplot(district_counts, aes(x = district, y = total_banned_bo
oks)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  xlab("District") +
  ylab("Total Banned Books") +
  ggtitle(paste("Total Banned Books per District in", top_state)) +
  theme(axis.text.x = element_text(angle = 65, hjust = 1, size=8))
print(district_bar)
```


Total Banned Books per District in Texas