

tidyverse: using ggplot2 functions to visualize data

Julia Ferris

2023-11-06

Introduction

In this demonstration, we will be using the packages tidyverse and ggplot2. ggplot2 is one of the packages that automatically loads with the tidyverse. This package is used for data visualization. We will focus on visualizing data with two variables. Some options include `geom_violin()`, `geom_density_2d()`, and `geom_rug()`. These are less frequently used, so this tutorial will explain how to use them with the chosen dataset.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.2      v readr      2.1.4
## v forcats    1.0.0      v stringr   1.5.0
## v ggplot2    3.4.3      v tibble    3.2.1
## v lubridate  1.9.2      v tidyr     1.3.0
## v purrr      1.0.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Load the Data

The dataset I used was called guns-polls. This data contained information about poll results for Democrats and Republicans about guns. When I loaded the data from the csv file, I removed the support column to focus on the differences between parties.

This is where you can find the data: <https://github.com/fivethirtyeight/data/tree/master/poll-quiz-guns>

```
## Rows: 57 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (6): Question, Start, End, Pollster, Population, URL
## dbl (3): Support, Republican Support, Democratic Support
##
## 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.
```

geom_violin() Demonstration

How did Democrats and Republicans vary in their answers to the questions?

In this section, `geom_violin()` is demonstrated. To use this for my data, I wanted to see how Democrats answered the questions compared to how Republicans answered the questions. The format of `geom_violin()`, which is a violin plot, is to create the `ggplot` and then add `geom_violin()`.

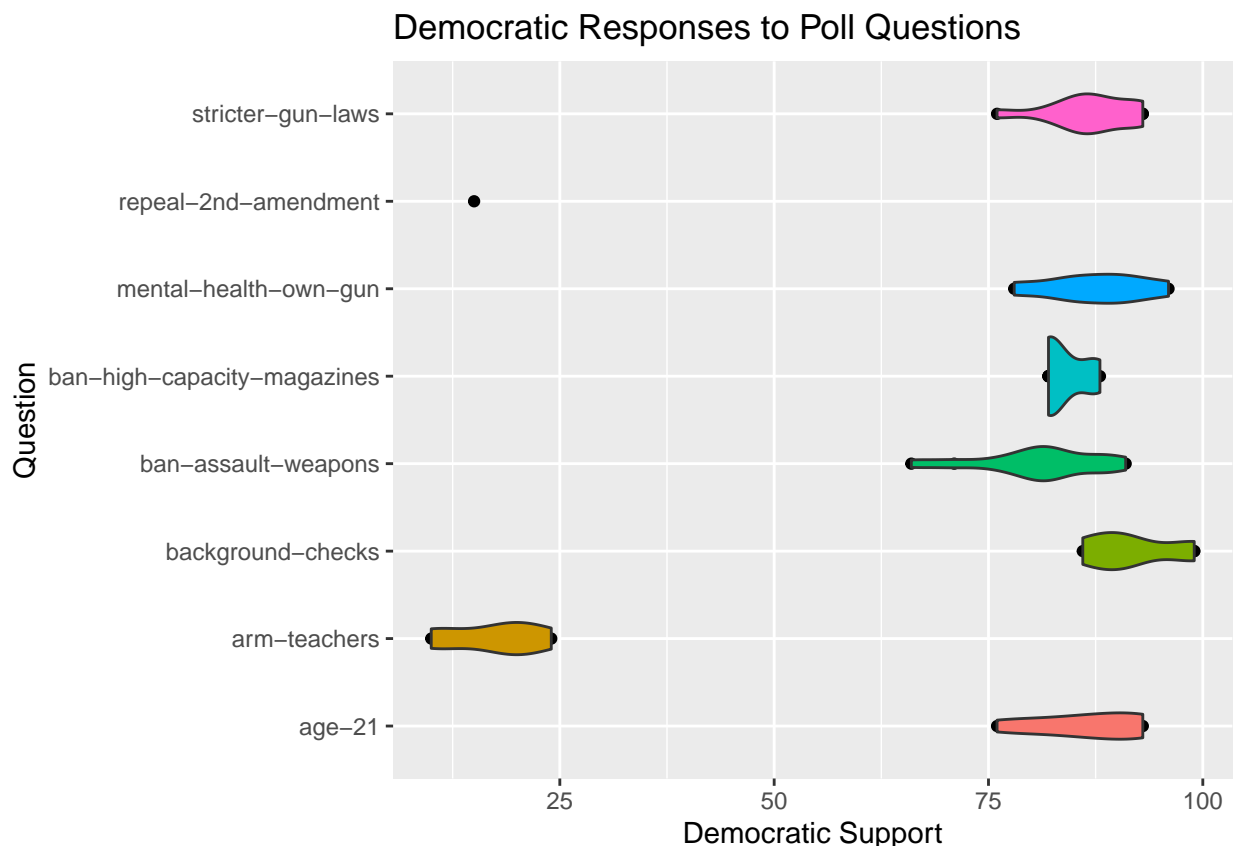
```
ggplot(data, aes(x, y)) + geom_violin()
```

I added `geom_point()` because I find it easier to read when it gives the points as a dot, and I added `coord_flip()` because that makes the poll questions easier to read. Also, adding color with `fill` makes the plot more aesthetically pleasing.

Based on these graphs, Democrats preferred stricter gun laws, Democrats had a stronger support for banning high-capacity magazines, and Republicans were much more in support of teachers being armed. Those were the top three differences I noticed from the results.

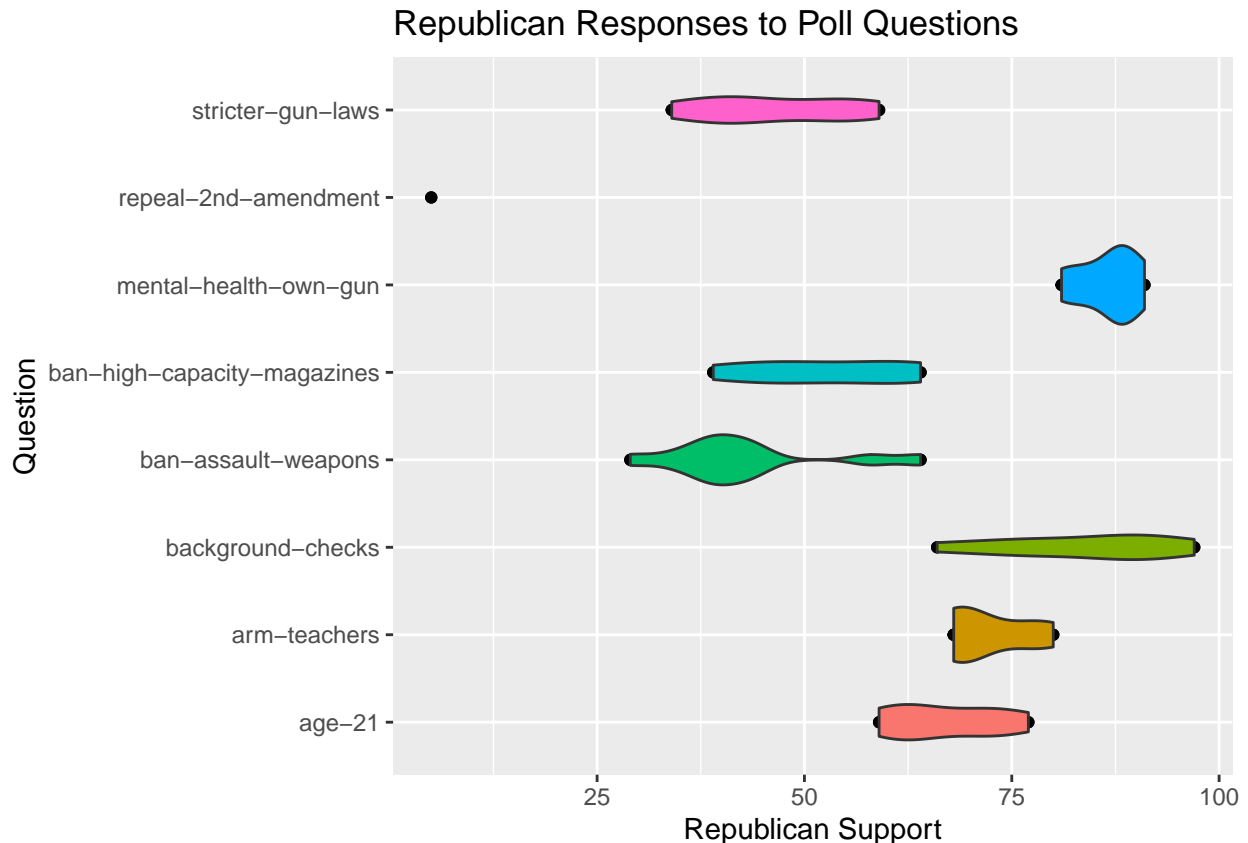
```
ggplot(guns_polls, aes(`Question`, `Democratic Support`, fill = Question)) +  
  geom_point() +  
  geom_violin() +  
  coord_flip() +  
  theme(legend.position = "none") +  
  labs(title = "Democratic Responses to Poll Questions")
```

Warning: Groups with fewer than two data points have been dropped.



```
ggplot(guns_polls, aes(`Question`, `Republican Support`, fill = Question)) + geom_point() +
  geom_point() +
  geom_violin() +
  coord_flip() +
  theme(legend.position = "none") +
  labs(title = "Republican Responses to Poll Questions")
```

Warning: Groups with fewer than two data points have been dropped.



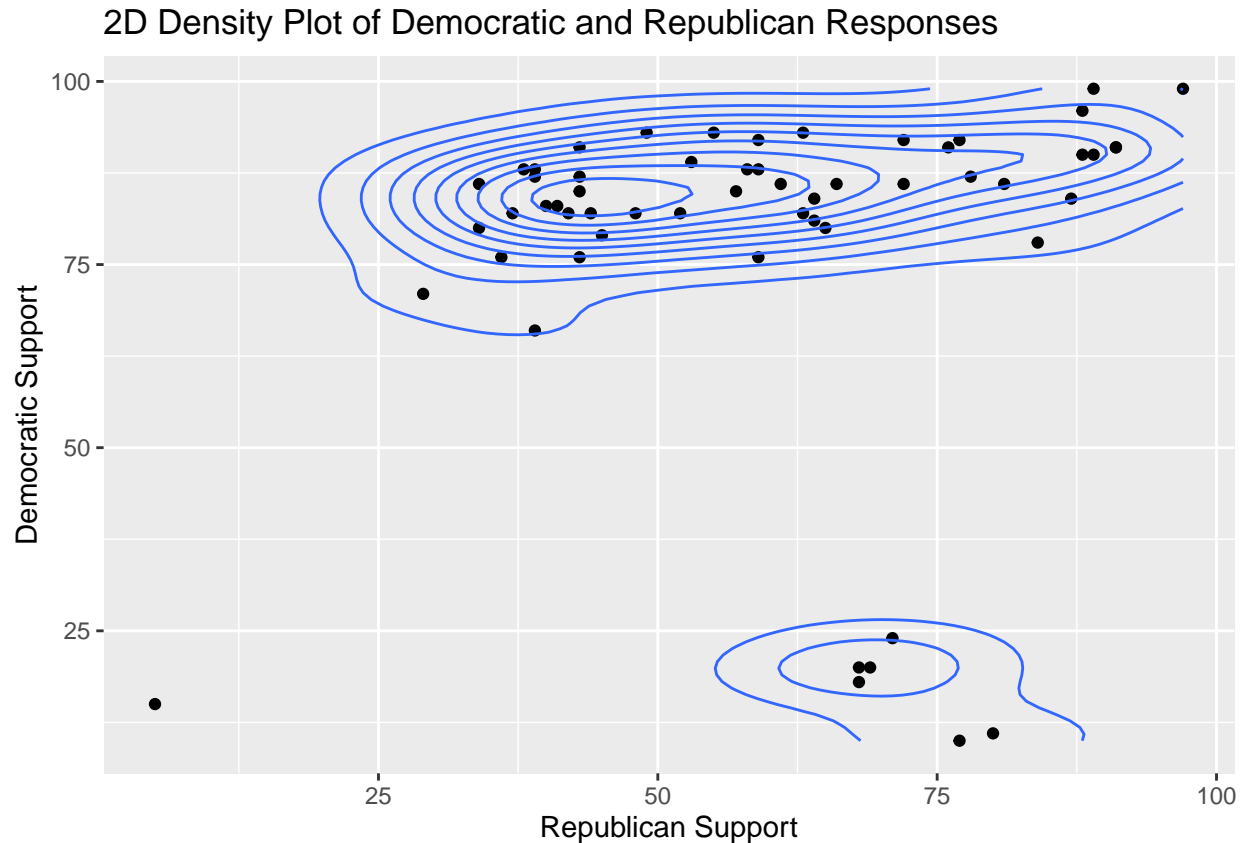
geom_density_2d() Demonstration

Did Republicans and Democrats answer the same overall, regardless of question?

To answer this question, I compared how the percentages of people answered from each group. If both answered with similar percentages, it would appear to be positive and linear. This is showing the function `geom_density_2d()` not because it is the best way to show this information, but because it is another option in the package `ggplot2`. Typically, a contour map is good for showing geographical data with different variables like temperature or altitude.

As we can see from the graph below, the responses were different. When Republicans responded with low to medium support, Democrats responded with high support. Also, when Republicans responded with about 75% support, Democrats responded less than 25% some of the time. Based on previous graphs, this difference is probably showing data about the question that asked about arming teachers. If the two groups responded similarly, both groups would answer low at the same time and high at the same time.

```
ggplot(guns_polls, aes(`Republican Support`, `Democratic Support`)) +
  geom_point() +
  geom_density_2d() +
  labs(title = "2D Density Plot of Democratic and Republican Responses")
```



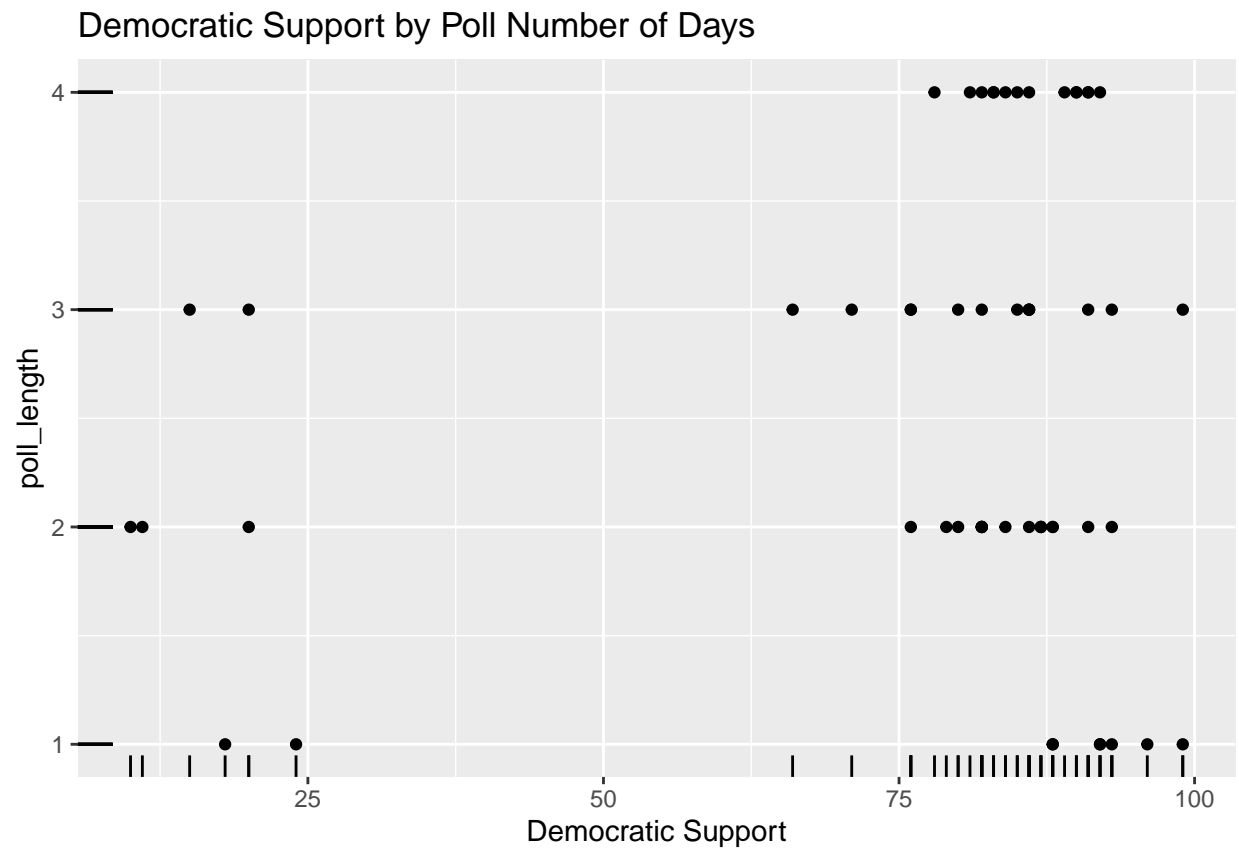
geom_rug() Demonstration

Did a longer amount of time for a question cause a higher percentage of support regardless of the question for Democrats and Republicans?

This question is important because if more people responded over more days, it could affect the results. I thought this question would help determine if the results were valid. If the responses did not change when the number of days of the poll for each question changed, then it would be valid. If the responses changed when the number of days of the poll for each question changed, then it would not be valid.

Based on the rug plots, the number of days for each poll did not affect the responses of the people in either party. Therefore, the results seem valid. I verified with correlations, and they confirmed the results from the plots.

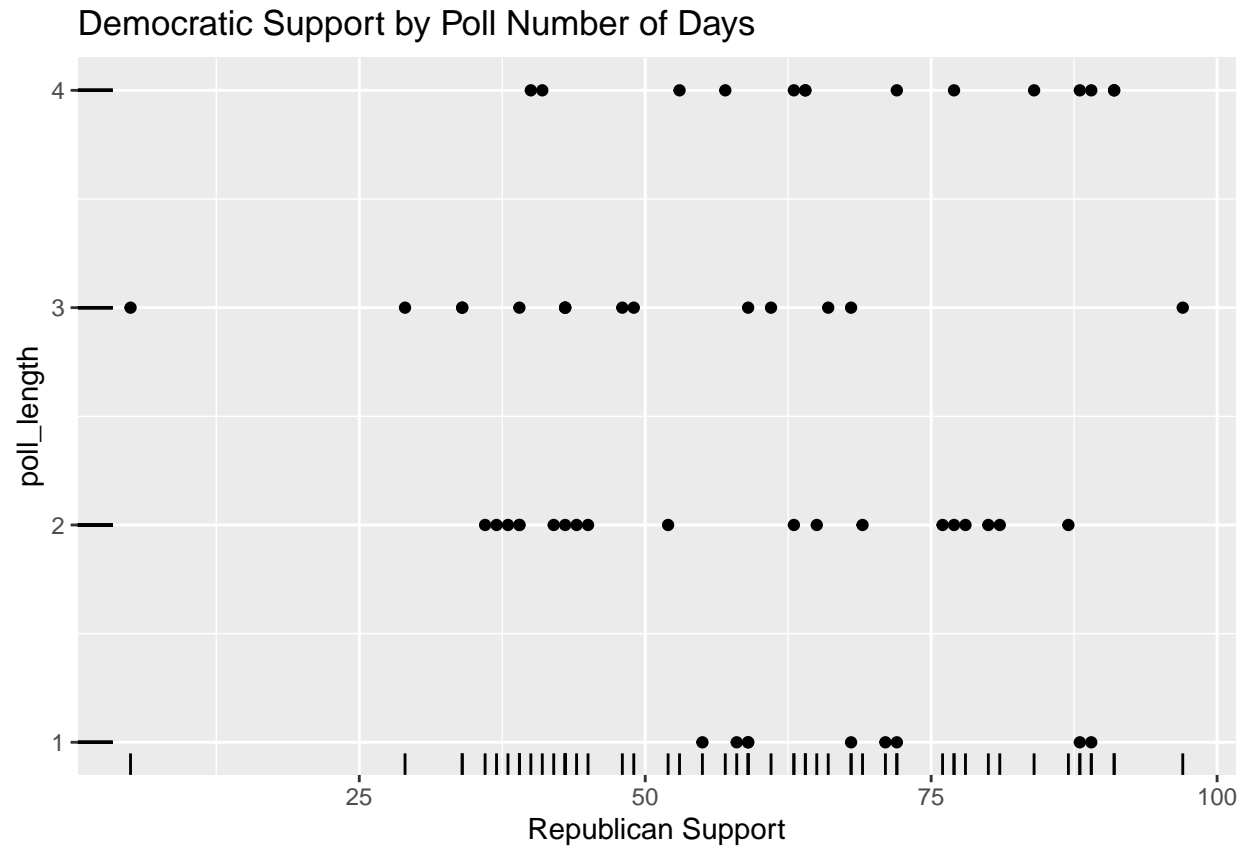
```
ggplot(guns_polls, aes(`Democratic Support`, poll_length)) +
  geom_point() +
  geom_rug() +
  labs(title = "Democratic Support by Poll Number of Days")
```



```
cor(guns_polls$`Democratic Support`, guns_polls$poll_length)
```

```
## [1] 0.1507378
```

```
ggplot(guns_polls, aes(`Republican Support`, poll_length)) +  
  geom_point() +  
  geom_rug() +  
  labs(title = "Democratic Support by Poll Number of Days")
```



```
cor(guns_polls$`Republican Support`, guns_polls$poll_length)
```

```
## [1] 0.01548815
```

Sources:

<https://github.com/fivethirtyeight/data/tree/master/poll-quiz-guns>

<https://github.com/rstudio/cheatsheets/blob/main/data-visualization.pdf>