New Voter Registration

Eric Carrieri Taylor

2024-06-12

Section 1: Description of the Data

The dataset under analysis is focused on the registration of new voters across various jurisdictions in the United States. It measures the number of new registered voters each month from different jurisdictions, providing insights into voter registration trends over time. This data was collected to answer research questions related to electoral engagement, the impact of policy changes on voter registration, and demographic shifts in voter registration patterns.

Section 2: Reading the Data into R

```
# Load necessary packages
library(readr)

# Read the data into R
data <- read_csv("/Users/erictaylor/Downloads/New Voter Registration.csv")
```

```
## Rows: 106 Columns: 4
## — Column specification ————————————————————————————————
## Delimiter: ","
## chr (2): Jurisdiction, Month
## dbl (2): Year, New registered voters
##
## 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.
```

Section 3: Cleaning the Data

```
# Load necessary packages
library(dplyr)
```

```
##
## 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
```

```
# Clean the data
# Rename columns for consistency and convert date columns to proper Date format
data <- data %>%
  rename(Jurisdiction = Jurisdiction,
         Year = Year,
         Month = Month,
         New_Registered_Voters = `New registered voters`)

# Convert Month to numeric, handling non-numeric values
data$Month <- as.numeric(as.character(data$Month))
```

```
## Warning: NAs introduced by coercion
```

Section 4: Characteristics of the Data

This dataframe has 106 rows and 4 columns. The names of the columns and a brief description of each are in the table below:

```
# Load necessary package for displaying table
library(knitr)

# Column names and descriptions
column_descriptions <- data.frame(
  Column_Name = c("Jurisdiction", "Year", "Month", "New_Registered_Voters"),
  Description = c("The region where the data was collected",
                  "The year of data collection",
                  "The month of data collection",
                  "The number of new registered voters")
)

# Display the table
kable(column_descriptions, col.names = c("Column Name", "Description"))
```

Column Name	Description
Jurisdiction	The region where the data was collected
Year	The year of data collection
Month	The month of data collection
New_Registered_Voters	The number of new registered voters

Section 5: Summary Statistics

```r
# Summary statistics for selected columns
# Using summary functions to get min, max, mean, and number of missing values

# For column 'Year'
year_min <- min(data$Year, na.rm = TRUE)
year_max <- max(data$Year, na.rm = TRUE)
year_mean <- mean(data$Year, na.rm = TRUE)
year_missing <- sum(is.na(data$Year))

# For column 'Month'
month_min <- min(data$Month, na.rm = TRUE)
```

```
## Warning in min(data$Month, na.rm = TRUE): no non-missing arguments to min;
## returning Inf
```

```r
month_max <- max(data$Month, na.rm = TRUE)
```

```
## Warning in max(data$Month, na.rm = TRUE): no non-missing arguments to max;
## returning -Inf
```

```r
month_mean <- mean(data$Month, na.rm = TRUE)
month_missing <- sum(is.na(data$Month))

# For column 'New_Registered_Voters'
new_reg_voters_min <- min(data$New_Registered_Voters, na.rm = TRUE)
new_reg_voters_max <- max(data$New_Registered_Voters, na.rm = TRUE)
new_reg_voters_mean <- mean(data$New_Registered_Voters, na.rm = TRUE)
new_reg_voters_missing <- sum(is.na(data$New_Registered_Voters))

# Output the results
list(
  Year = list(min = year_min, max = year_max, mean = year_mean, missing = year_missing),
  Month = list(min = month_min, max = month_max, mean = month_mean, missing = month_miss
ing),
  New_Registered_Voters = list(min = new_reg_voters_min, max = new_reg_voters_max, mean
= new_reg_voters_mean, missing = new_reg_voters_missing)
)
```

```
## $Year
## $Year$min
## [1] 2016
##
## $Year$max
## [1] 2020
##
## $Year$mean
## [1] 2018
##
## $Year$missing
## [1] 0
##
##
## $Month
## $Month$min
## [1] Inf
##
## $Month$max
## [1] -Inf
##
## $Month$mean
## [1] NaN
##
## $Month$missing
## [1] 106
##
##
## $New_Registered_Voters
## $New_Registered_Voters$min
## [1] 589
##
## $New_Registered_Voters$max
## [1] 238281
##
## $New_Registered_Voters$mean
## [1] 48223.46
##
## $New_Registered_Voters$missing
## [1] 0
```