Voting Patterns in Massachusetts

2023-06-27

General Voter Turnout

Import Data

```
library(readxl)

# Set the file path of the Excel file
file_path <- "data/Voter-Turnout-Statistics.xlsx"
file_path1 <- "data/voter2.xlsx"
file_path2 <- "data/voter3.xlsx"

# Read the Excel file
data <- read_excel(file_path)
data1 <- read_excel(file_path1)
data2 <- read_excel(file_path2) ## CHECK WHICH IS WHICH BEFORE ANALYSIS

# Display the data
print(data)
```

```
## # A tibble: 38 x 4
##    `State Election` `Registered Voters` `Total Votes Cast` `Turnout Percentage`
##               <dbl>               <dbl>              <dbl>                <dbl>
##  1             1948             2484938            2155347                0.867
##  2             1950             2475396            1947071                0.787
##  3             1952             2555025            2424548                0.949
##  4             1954             2523414            1942071                0.770
##  5             1956             2671369            2388129                0.894
##  6             1958             2556300            1952588                0.764
##  7             1960             2720359            2495504                0.917
##  8             1962             2635086            2144051                0.814
##  9             1964             2723598            2388230                0.877
## 10             1966             2641538            2076826                0.786
## # i 28 more rows
```

```
print(data1)
```

```
## # A tibble: 38 x 4
##    `State Primary` `Registered Voters` `Total Votes Cast` `Turnout Percentage`
##              <dbl>               <dbl>              <dbl>                <dbl>
##  1            1948             2484938             591248                0.238
##  2            1950             2475396             827158                0.334
##  3            1952             2555025             960580                0.376
##  4            1954             2523414             604804                0.240
##  5            1956             2671369             848880                0.318
##  6            1958             2556300             768456                0.301
```

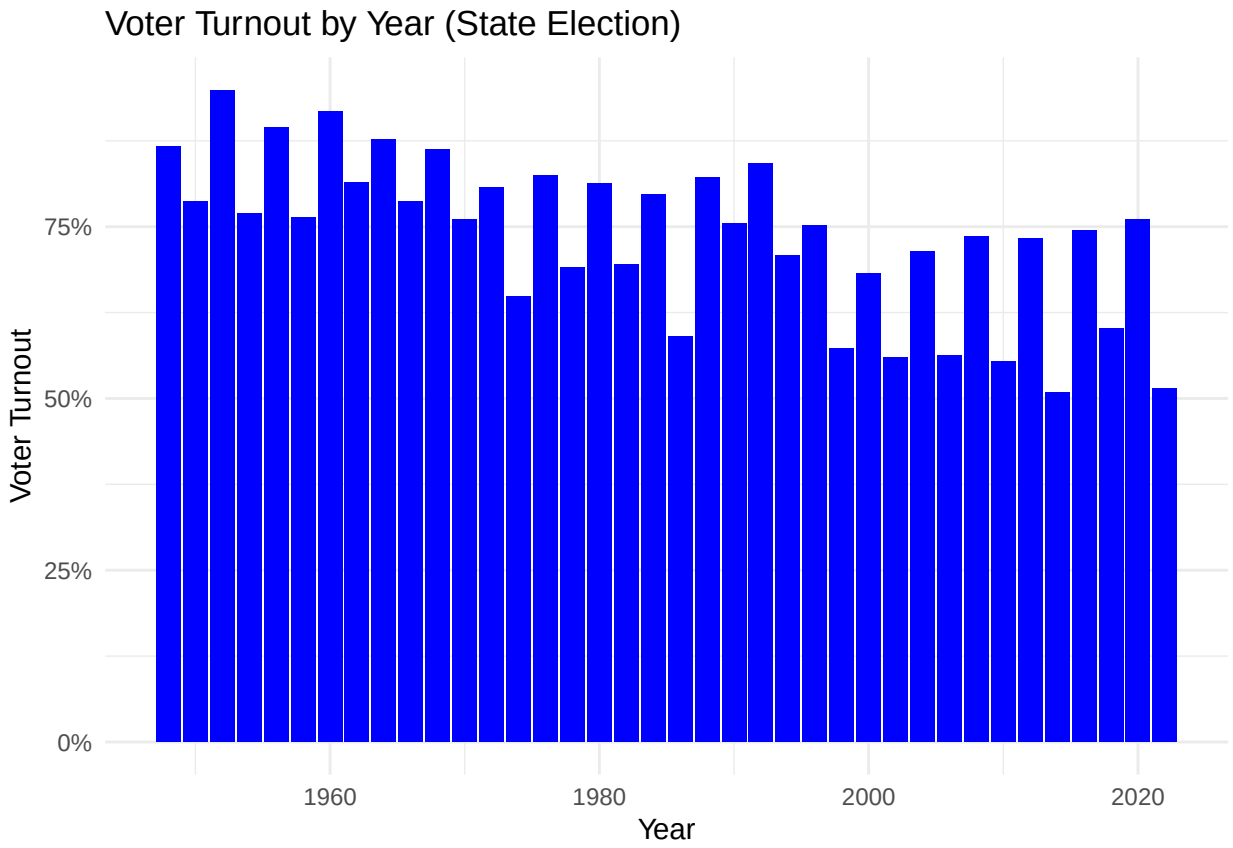
```
##  7            1960             2720359             860474                0.316
##  8            1962             2635086            1293764                0.491
##  9            1964             2723598             946864                0.348
## 10            1966             2641538             846094                0.320
## # i 28 more rows
```

```
print(data2)
```

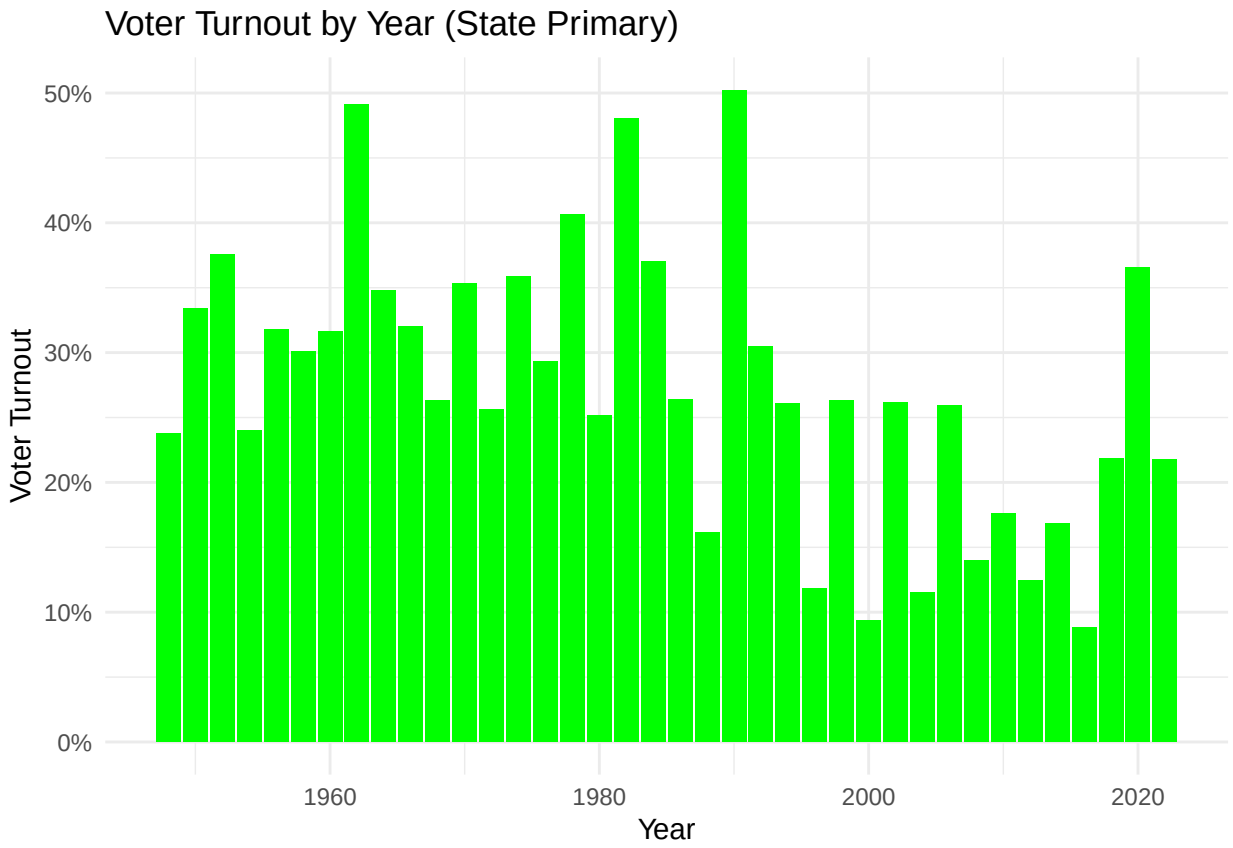
```
## # A tibble: 18 x 4
##    `State Primary` `Registered Voters` `Total Votes Cast` `Turnout Percentage`
##              <dbl>               <dbl>              <dbl>                 <dbl>
##  1            1952             2666025             573973                0.215
##  2            1956             2671936             183660                0.0687
##  3            1960             2720359             252244                0.0927
##  4            1964             2723598             345598                0.127
##  5            1968             2725058             471397                0.173
##  6            1972             2775538             768981                0.277
##  7            1976             2872483             941943                0.328
##  8            1980             3026097            1330727                0.440
##  9            1984             3054129             711171                0.233
## 10            1988             2965272             975106                0.329
## 11            1992             3130272            1086359                0.347
## 12            1996             3166047             455362                0.144
## 13            2000             3794046             360064                0.0949
## 14            2004             3903810             696636                0.178
## 15            2008             4308228            1883846                0.437
## 16            2012             4111128             529542                0.129
## 17            2016             4271835            1863339                0.436
## 18            2020             4581319            1700087                0.371
```

Turnout Percentage Per Each Voting Year

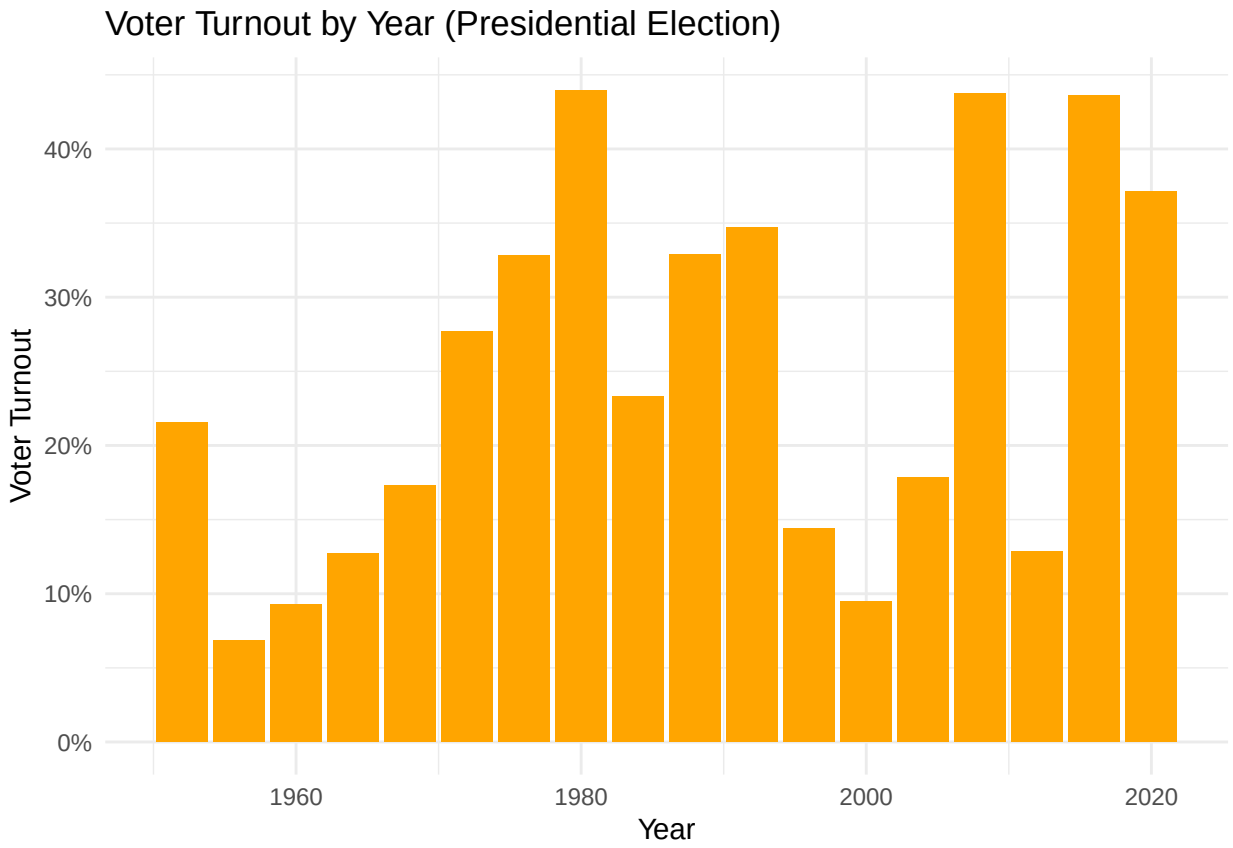
```
library(ggplot2)

# Create a bar plot for data
ggplot(data, aes(x = `State Election`, y = `Turnout Percentage`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Election)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::percent)
```




```
# Create a bar plot for data1
ggplot(data1, aes(x = `State Primary`, y = `Turnout Percentage`)) +
  geom_bar(stat = "identity", fill = "green") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Primary)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::percent)
```

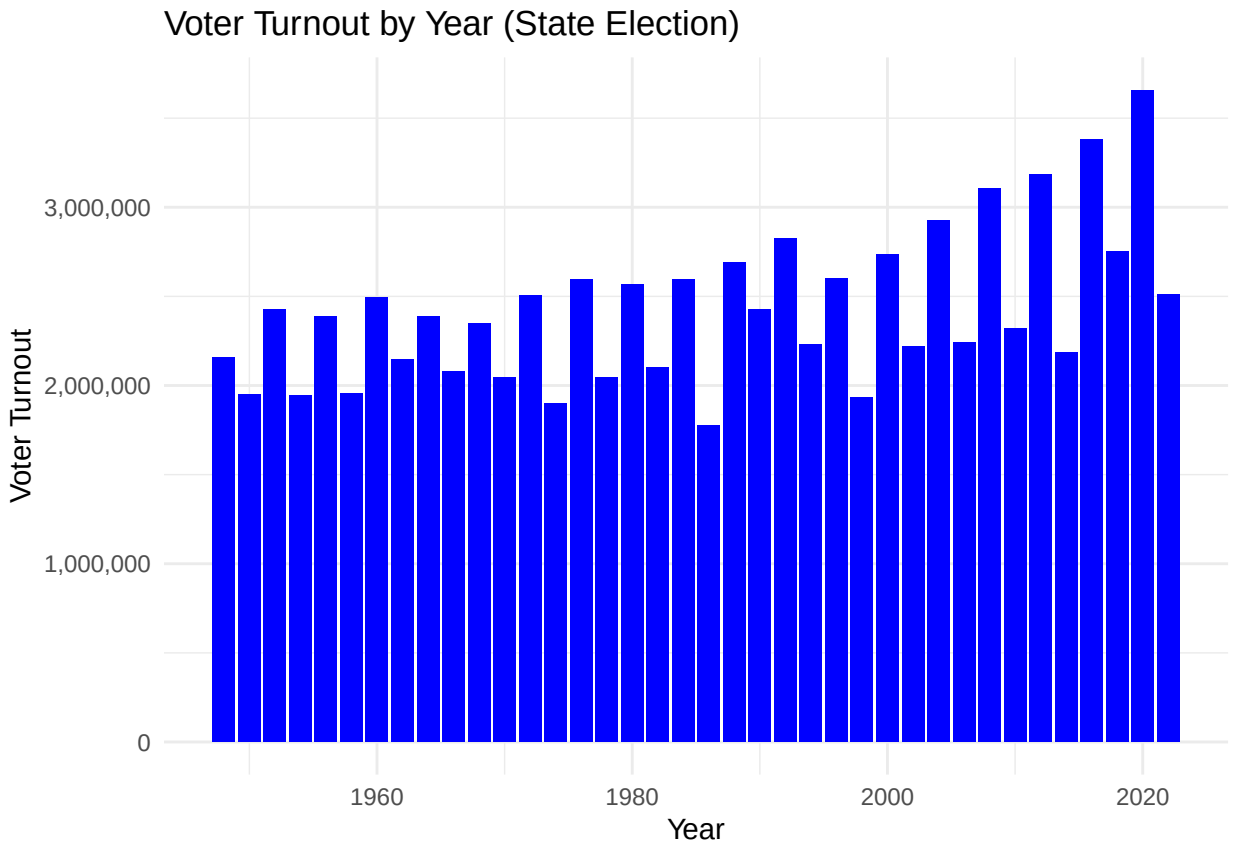



```
# Create a bar plot for data2
ggplot(data2, aes(x = `State Primary`, y = `Turnout Percentage`)) +
  geom_bar(stat = "identity", fill = "orange") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (Presidential Election)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::percent)
```

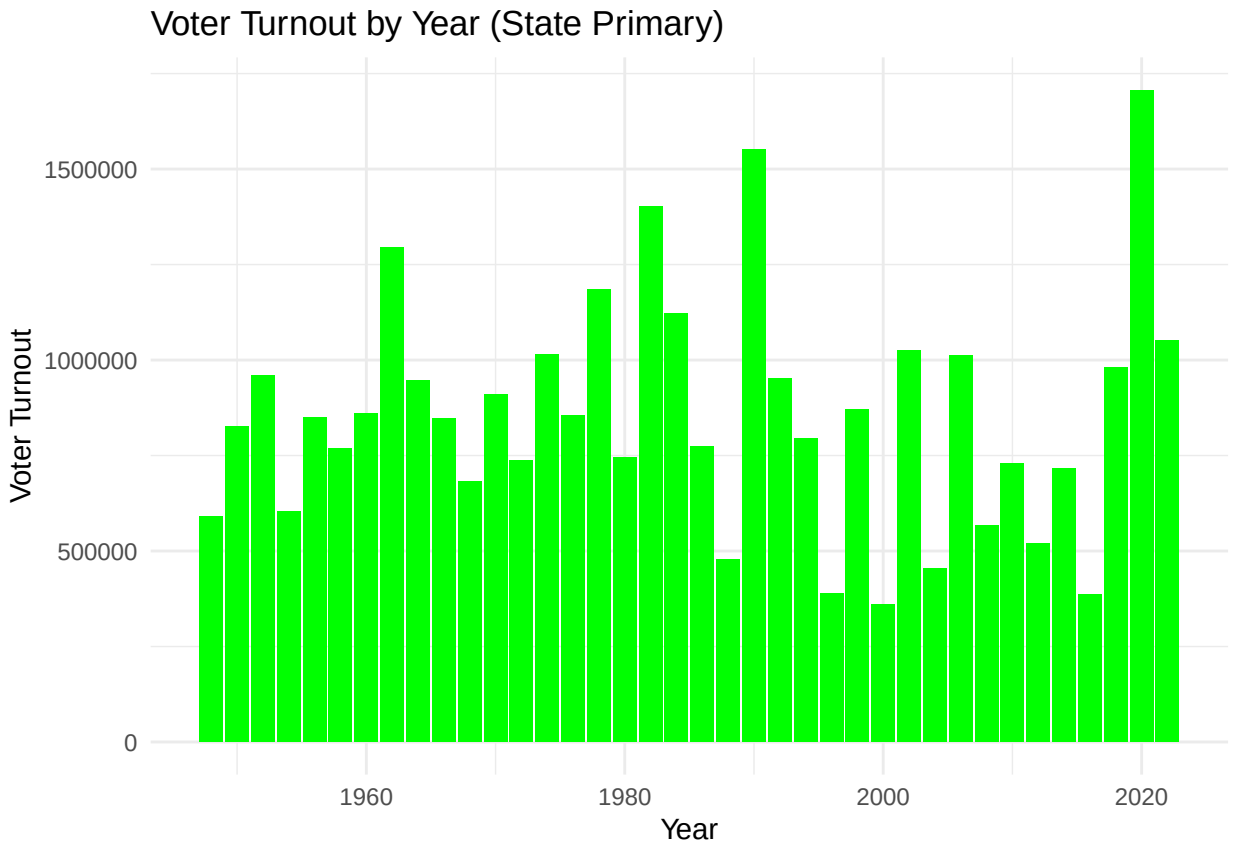


Total Votes Cast Per Each Voting Year

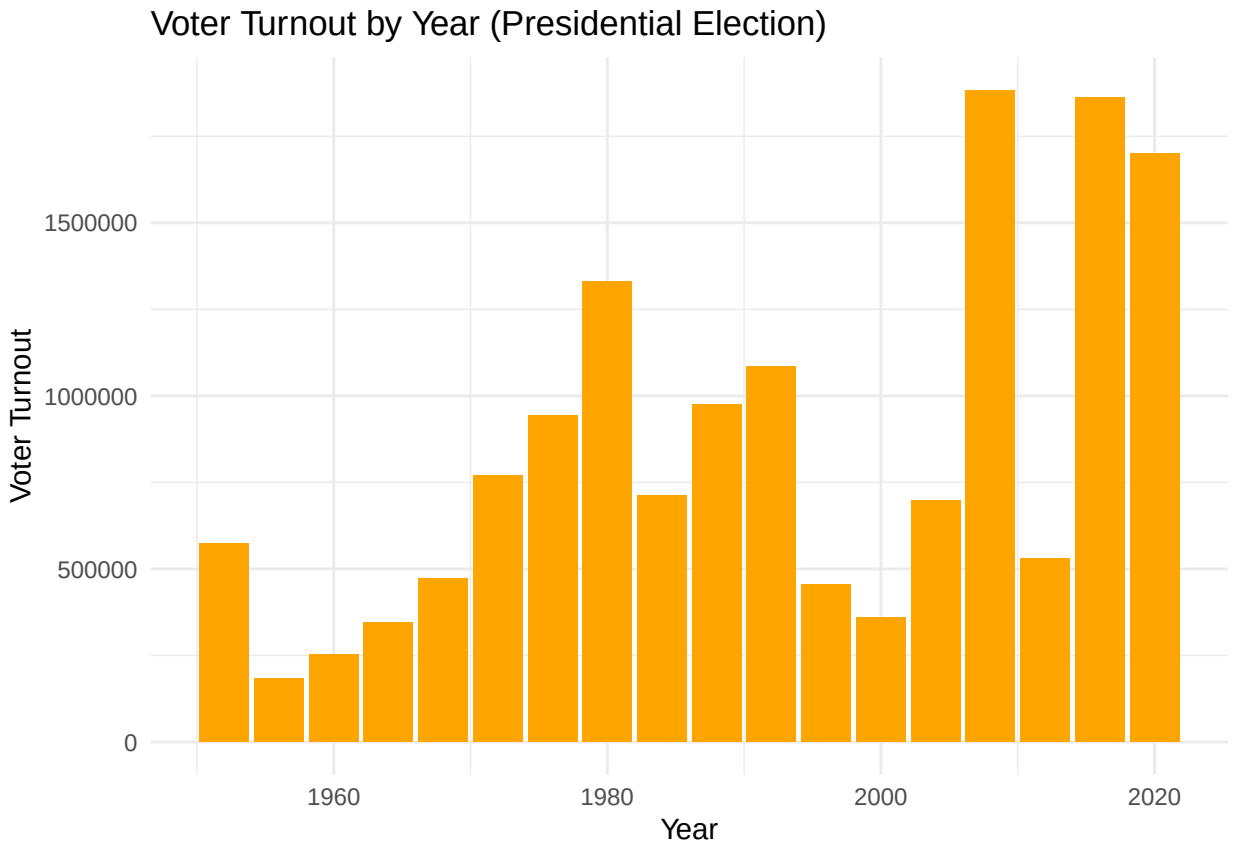
```
# Create a bar plot for data
ggplot(data, aes(x = `State Election`, y = `Total Votes Cast`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Election)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma)
```




```
# Create a bar plot for data1
ggplot(data1, aes(x = `State Primary`, y = `Total Votes Cast`)) +
  geom_bar(stat = "identity", fill = "green") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Primary)") +
  theme_minimal()
```

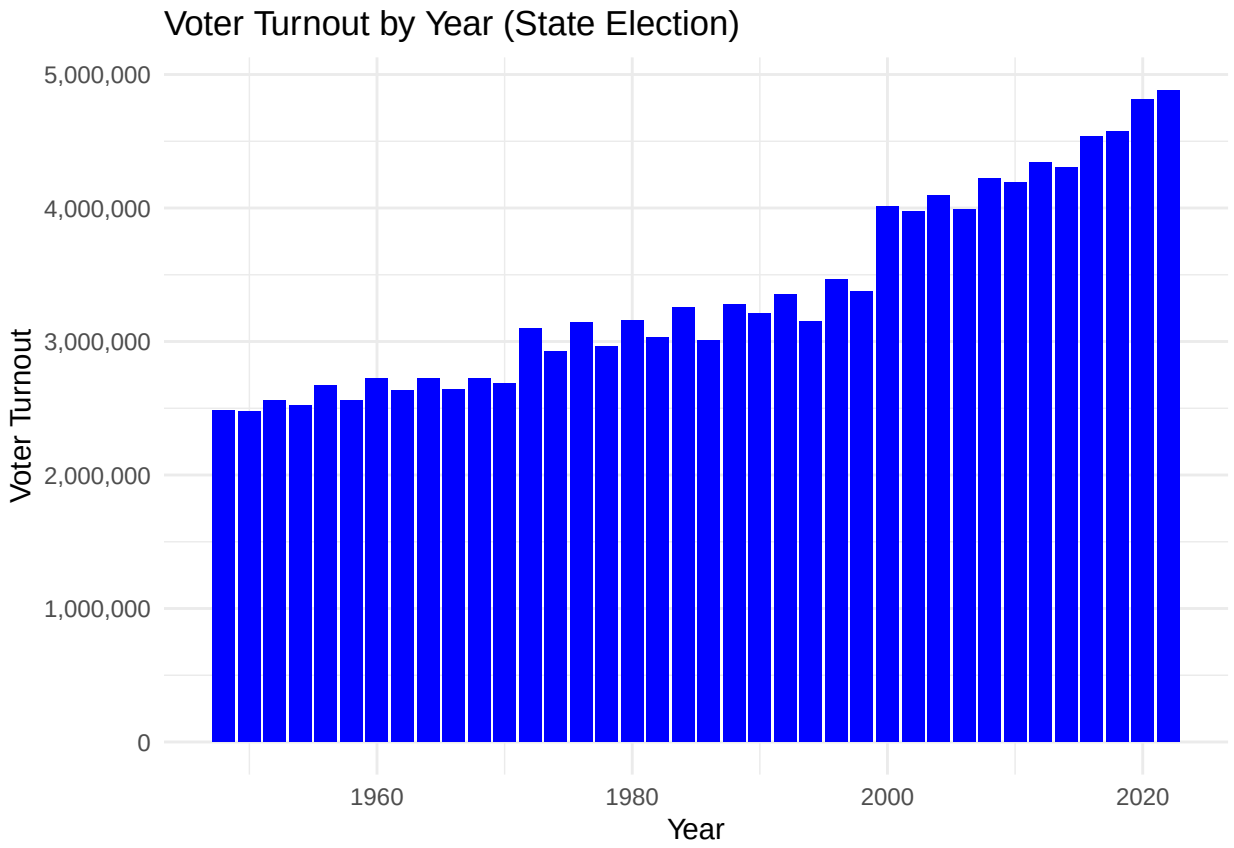



```
# Create a bar plot for data2
ggplot(data2, aes(x = `State Primary`, y = `Total Votes Cast`)) +
  geom_bar(stat = "identity", fill = "orange") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (Presidential Election)") +
  theme_minimal()
```

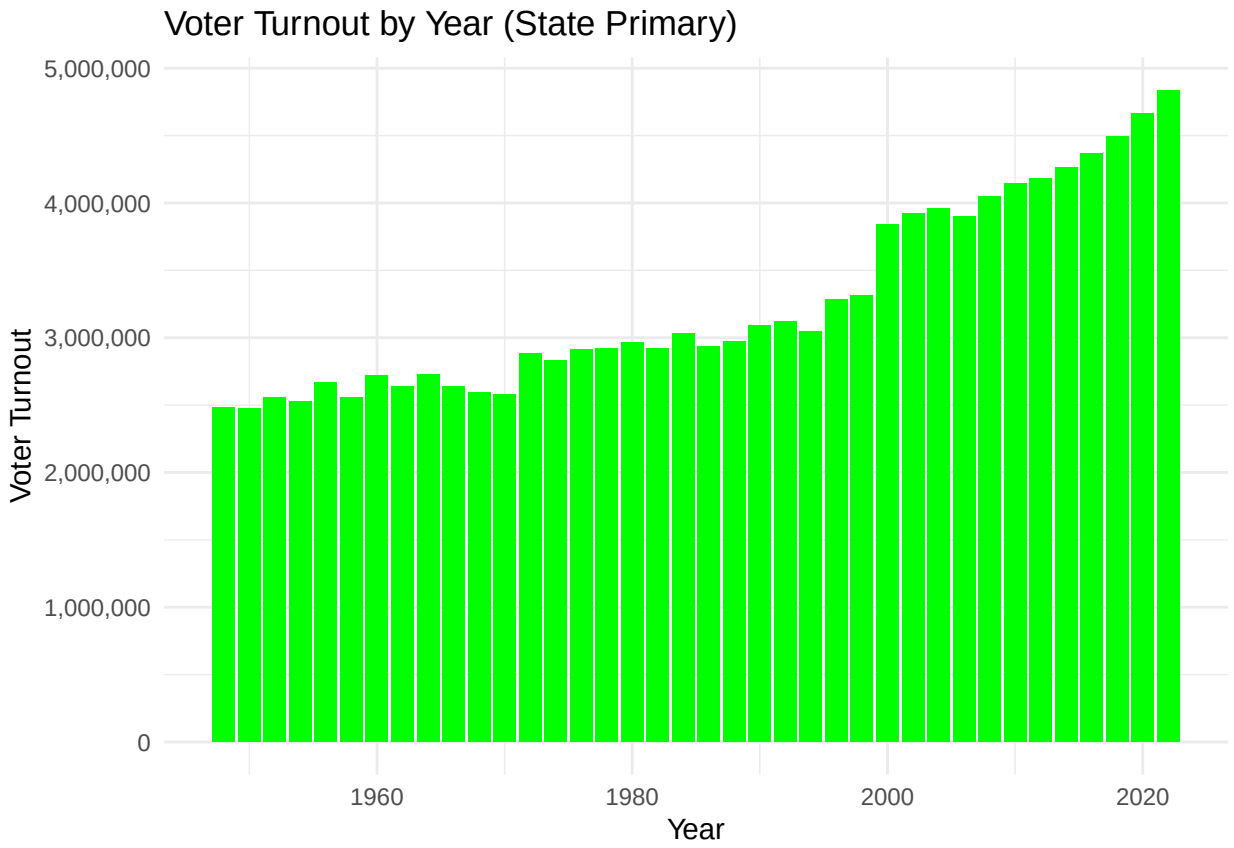


Registered Voters Per Each Voting Year

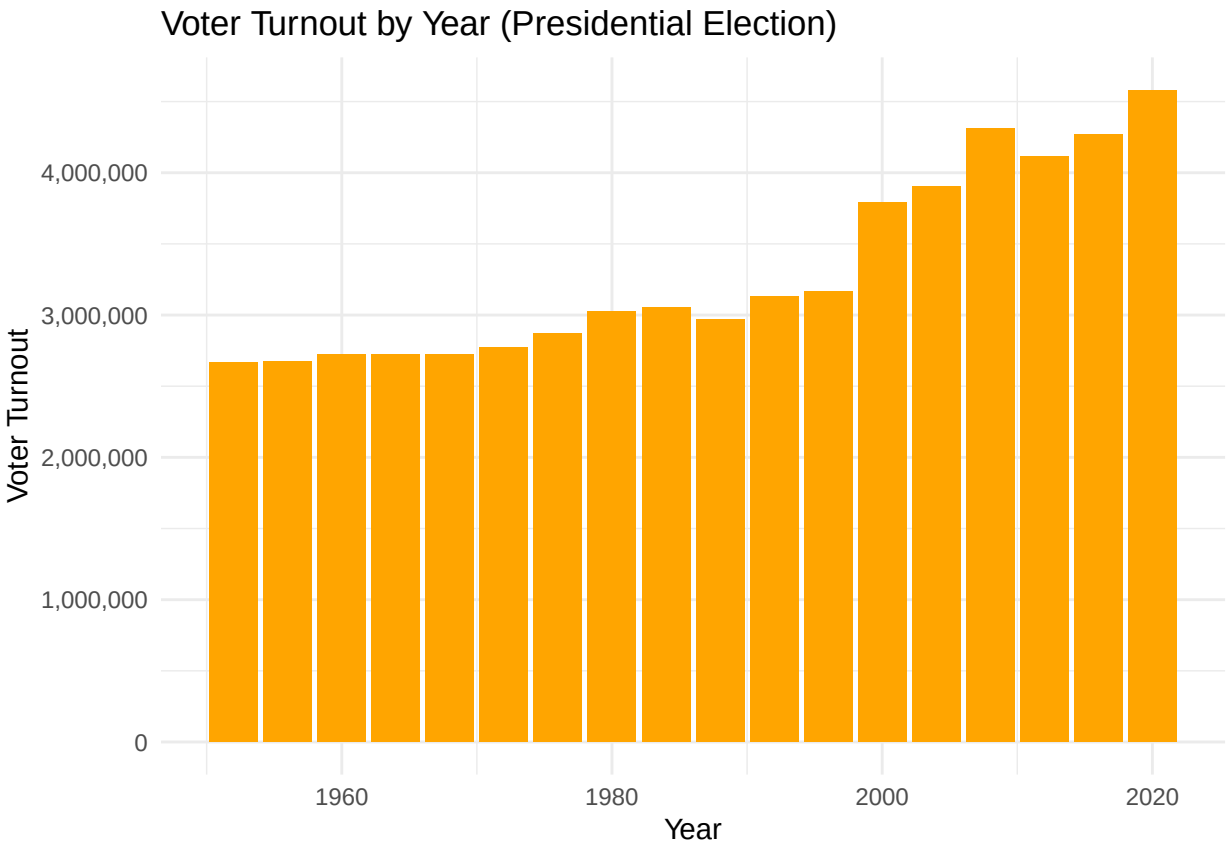
```
library(ggplot2)

# Create a bar plot for data
ggplot(data, aes(x = `State Election`, y = `Registered Voters`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Election)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma)
```




```
# Create a bar plot for data1
ggplot(data1, aes(x = `State Primary`, y = `Registered Voters`)) +
  geom_bar(stat = "identity", fill = "green") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (State Primary)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma)
```




```
# Create a bar plot for data2
ggplot(data2, aes(x = `State Primary`, y = `Registered Voters`)) +
  geom_bar(stat = "identity", fill = "orange") +
  labs(x = "Year", y = "Voter Turnout") +
  ggtitle("Voter Turnout by Year (Presidential Election)") +
  theme_minimal() +
  scale_y_continuous(labels = scales::comma)
```



Census

Read data

```
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
```

```
# Remove the first row from the dataset
file_path2022 <- "data/census2022.xlsx"

data2022 <- read_excel(file_path2022,skip = 1)

data2022 <- data2022 %>%
  mutate(precinct = as.integer(substr(`Ward and Precinct (Updated 2022)`, 1, 2)))
```

```
## Warning: There was 1 warning in `mutate()`.
## i In argument: `precinct = as.integer(substr(`Ward and Precinct (Updated
##   2022)`, 1, 2))`.
```

```
## Caused by warning:
## ! NAs introduced by coercion
```

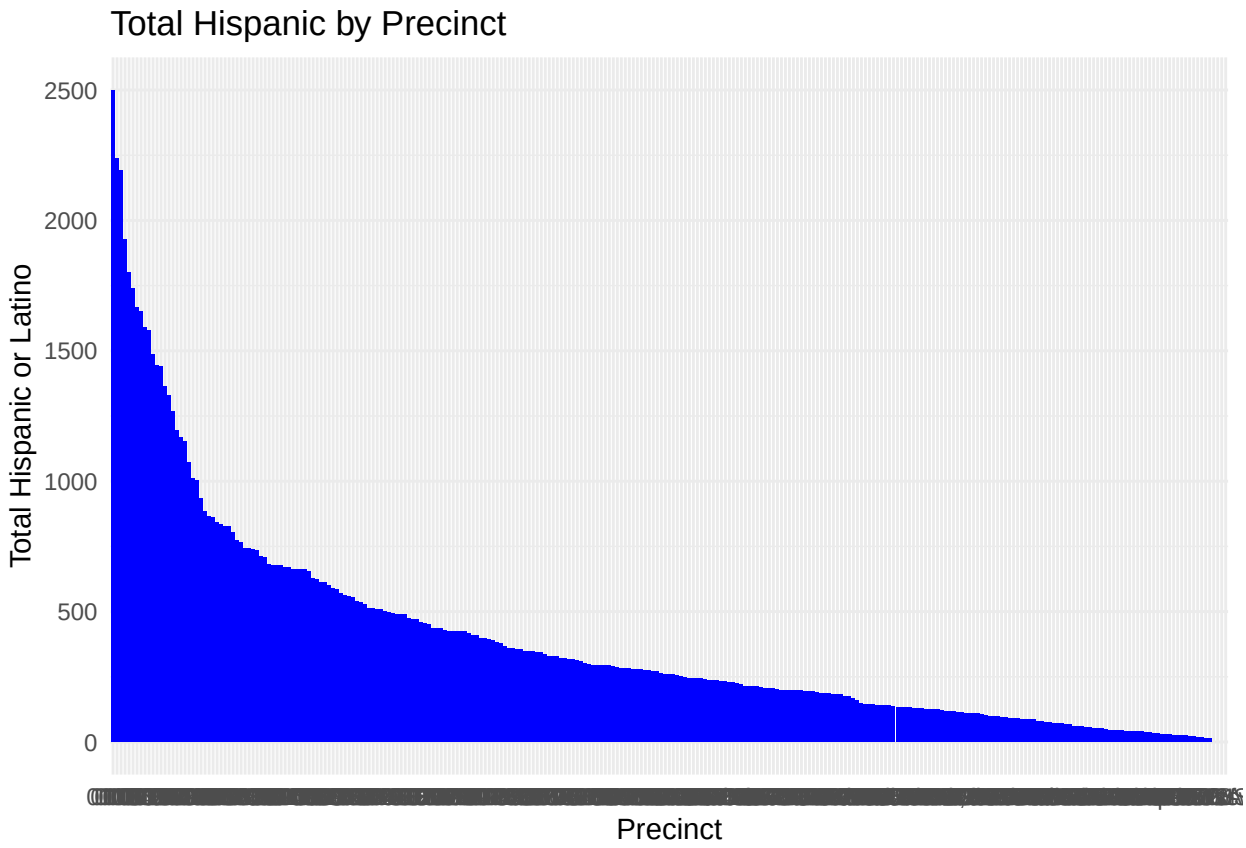
```
data2022 <- data2022 %>%
  mutate(realPrecinct = `Ward and Precinct (Updated 2022)`)

print(data2022)
```

```
## # A tibble: 279 x 32
##    `Ward and Precinct (Updated 2022)` `Total Population` `White alone`
##    <chr>                                           <dbl>         <dbl>
##  1 0101                                             2047          1200
##  2 0102                                             2608           968
##  3 0103                                             3965          1417
##  4 0104                                             2397           606
##  5 0105                                             3649          1105
##  6 0106                                             3060           883
##  7 0107                                             3438           910
##  8 0108                                             3984          1051
##  9 0109                                             2764           855
## 10 0110                                             2414           651
## # i 269 more rows
## # i 29 more variables: `Black or African American` <dbl>,
## #   `Hispanic or Latino (of any Race)` <dbl>, Asian <dbl>,
## #   `American Indian and Alaska Native` <dbl>,
## #   `Native Hawaiian and Other Pacific Islander` <dbl>,
## #   `Some Other Race` <dbl>, `Two or More Non-White Races` <dbl>,
## #   `Total Population,\r\nAge 18 and over` <dbl>, ...
```

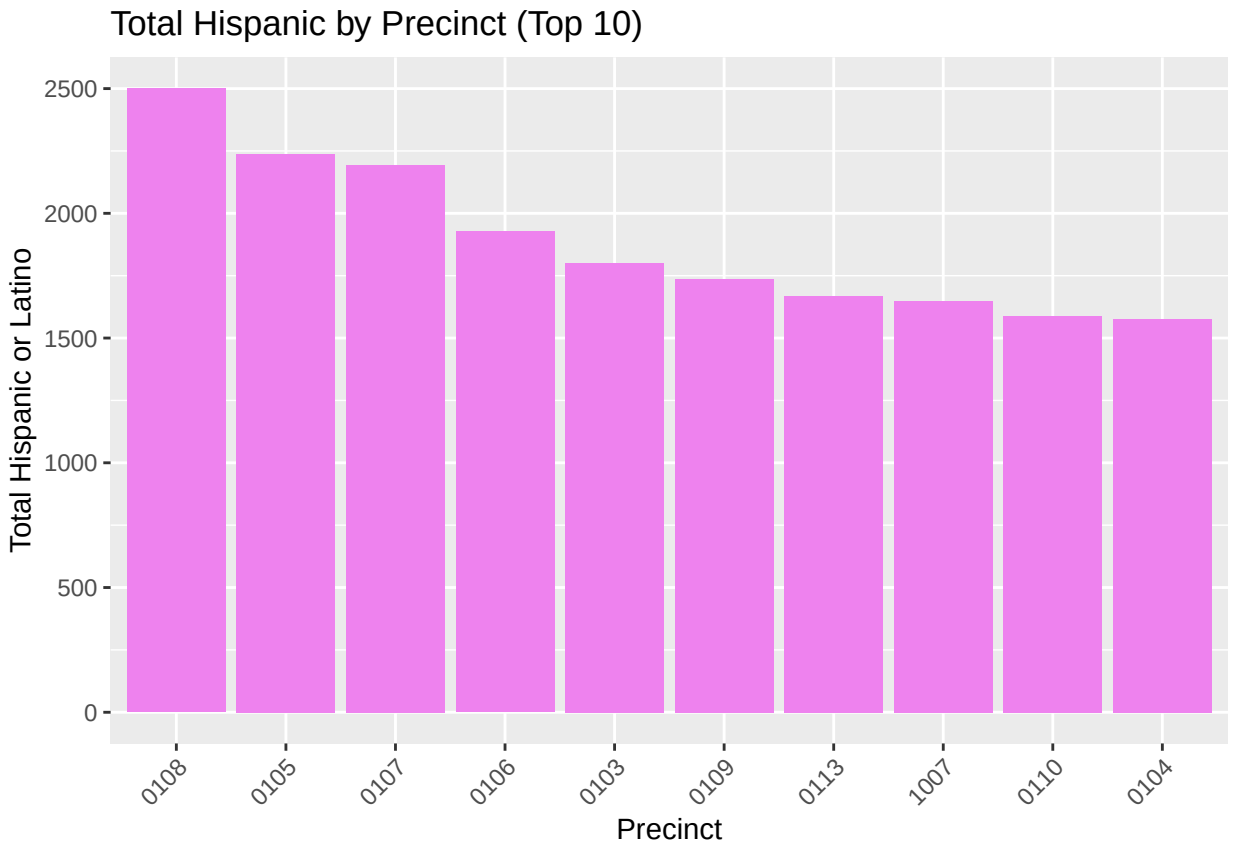
Hispanic by precinct

```
library(ggplot2)
library(dplyr)

# Group the data by Precinct and summarize the total number of "Hispanic or Latino (of any Race)"
#summarized_data <- data2022 %>%
 # group_by(precinct) %>%
  #summarize(Total_Hispanic_Latino = sum(`Hispanic or Latino (of any Race)`))

# Reordering the levels of the variable based on the values
data2022$`Ward and Precinct (Updated 2022)` <- factor(data2022$`Ward and Precinct (Updated 2022)`,
          levels = data2022$`Ward and Precinct (Updated 2022)`[order(-data2022$`Hispanic or Latino (of a

# Plotting with descending bars
ggplot(data2022, aes(x = `Ward and Precinct (Updated 2022)`, y = `Hispanic or Latino (of any Race)`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Precinct", y = "Total Hispanic or Latino") +
  ggtitle("Total Hispanic by Precinct") +
  theme_minimal()
```

```
## Warning: Removed 4 rows containing missing values (`position_stack()`).
```

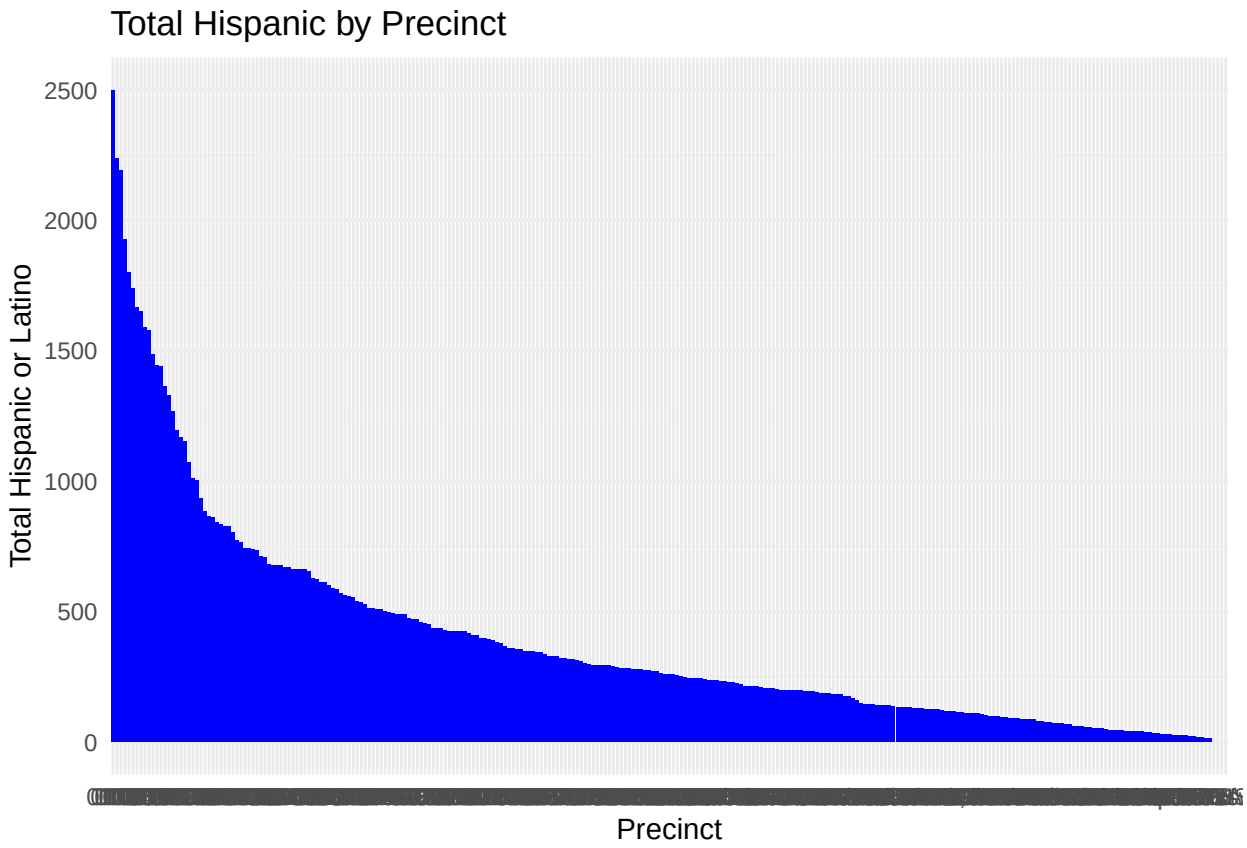



```
library(dplyr)

# Reordering the levels of the variable based on the values
data2022$`Ward and Precinct (Updated 2022)` <- factor(data2022$`Ward and Precinct (Updated 2022)`,
                                                      levels = data2022$`Ward and Precinct (Updated 2022`

# Selecting the top 10 ward/precincts
top_10_precincts <- data2022 %>%
  top_n(10, `Hispanic or Latino (of any Race)`)

# Plotting with descending bars for the top 10 ward/precincts
ggplot(top_10_precincts, aes(x = `Ward and Precinct (Updated 2022)`, y = `Hispanic or Latino (of any Ra
  geom_bar(stat = "identity", fill = "violet") +
  labs(x = "Precinct", y = "Total Hispanic or Latino") +
  ggtitle("Total Hispanic by Precinct (Top 10)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
# Reordering the levels of the variable based on the values
data2022$`Ward and Precinct (Updated 2022)` <- factor(data2022$`Ward and Precinct (Updated 2022)`,
                                                       levels = data2022$`Ward and Precinct (Updated 2022)

# Plotting with descending bars
ggplot(data2022, aes(x = `Ward and Precinct (Updated 2022)`, y = `Hispanic or Latino (of any Race)`)) +
  geom_bar(stat = "identity", fill = "blue") +
  labs(x = "Precinct", y = "Total Hispanic or Latino") +
  ggtitle("Total Hispanic by Precinct") +
  theme_minimal()
```

```
## Warning: Removed 4 rows containing missing values (`position_stack()`).
```


Total Hispanic by Precinct
Total Hispanic or Latino
2500
2000
1500
1000
500
0
Precinct