voting-rights-v2

margaret

3/30/2022

```r
library(tidyverse)
library(janitor)
library(scales)
```

data from: https://dl.ncsbe.gov/index.html?prefix=data/Snapshots/

```r
durham21 <- read_csv("data/duke_2021.csv")
```

```
## Warning: One or more parsing issues, see `problems()` for details
```

```
## Rows: 11801 Columns: 92
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (52): county_desc, voter_reg_num, ncid, status_cd, voter_status_desc, r...
## dbl  (12): county_id, house_num, zip_code, mail_zipcode, area_cd, phone_num,...
## lgl  (24): absent_ind, name_prefx_cd, half_code, street_dir, street_sufx_cd,...
## date  (4): snapshot_dt, registr_dt, cancellation_dt, load_dt
##
## 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.
```

```r
durhamhist <- read_tsv("data/ncvhis32.txt")
```

```
## Rows: 1269492 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: "\t"
## chr (13): county_desc, voter_reg_num, election_lbl, election_desc, voting_me...
## dbl  (2): county_id, voted_county_id
##
## 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.
```

```r
duke_students <- durham21 %>%
  filter(
    street_name %in% c("DUKE UNIVERSITY WEST CAMPUS",
                       "DUKE UNIVERSITY EAST CAMPUS", "TOWERVIEW", "ALEXANDER",
                       "YEARBY") |
      (street_name %in% c("BASSETT", "EPWORTH", "FEW QUAD", "GILBERT ADDOMS",
                          "GILES", "JARVIS", "MAXWELL", "MIRECOURT", "PEGRAM",
```

```
                             "RANDOLPH", "RANDOLPH HALL", "ROUND TABLE",
                             "SOUTHGATE", "WAYNE MANOR", "WILSON", "WANNAMAKER") &&
        street_type_cd == "DORM") |
      (street_name == "SWIFT" & house_num == "300") |
      (street_name == "CAMPUS"  & str_detect(house_num, "13\\d\\d"))
  )

full_data <- inner_join(
  duke_students,
  durhamhist,
  by = c("voter_reg_num" = "voter_reg_num")
  ) %>%
  mutate(
    registr_yr = lubridate::year(registr_dt),
    datev = as.Date(election_lbl, format = "%m/%d/%Y"),
    voted_year = lubridate::year(datev)
    )
election_year <- full_data %>%
  distinct(voter_reg_num, .keep_all = T) %>%
  count(voted_year) %>%
  filter(voted_year %in% c(2004, 2008, 2012, 2016, 2020))
election_yr_reg <- duke_students %>%
  distinct(voter_reg_num, .keep_all = T) %>%
  count(registr_yr) %>%
  filter(registr_yr %in% c(2004, 2008, 2012, 2016, 2020))
```

```
duke_students %>%
  mutate(registr_yr = lubridate::year(registr_dt)) %>%
  count(registr_yr) %>%
  filter(registr_yr >= 2000) %>%
  ggplot(aes(x = registr_yr, y = n)) +
  geom_line() +
  geom_line(data = election_yr_reg, color = "red") +
  geom_point(data = election_yr_reg, color = "red", size = 3) +
  scale_y_continuous(labels = scales::label_comma()) +
  scale_x_continuous(breaks = seq(2000, 2020, by = 4)) +
  labs(
    x = "Year",
    y = "Number of Duke students registered"
  ) +
  theme_minimal()
```
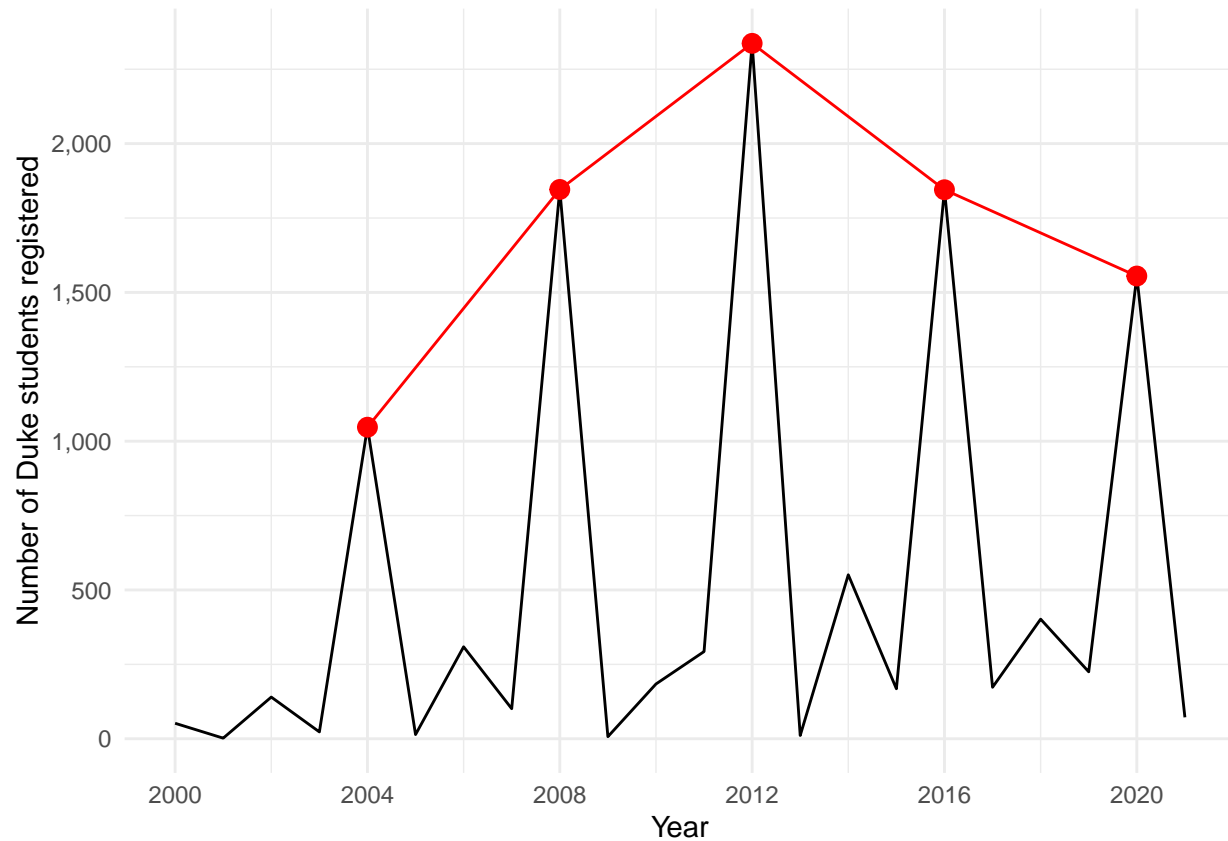
```
durham21 %>%
  count(street_name, street_type_cd) %>%
  arrange(desc(n))
```

```
## # A tibble: 17 x 3
##    street_name               street_type_cd     n
##    <chr>                     <chr>          <int>
##  1 DUKE UNIVERSITY WEST CAMPUS <NA>           5551
##  2 DUKE UNIVERSITY EAST CAMPUS <NA>           3151
##  3 DUKE UNIVERSITY WEST CAMPUS DORM            929
##  4 DUKE UNIVERSITY EAST CAMPUS DORM            726
##  5 ALEXANDER                 AVE              499
##  6 YEARBY                    AVE              399
##  7 SWIFT                     AVE              383
##  8 YEARBY                    ST                58
##  9 CAMPUS                    DR                39
## 10 TOWERVIEW                 DR                35
## 11 ALEXANDER                 ST                13
## 12 ALEXANDER                 DR                10
## 13 ALEXANDER                 <NA>               2
## 14 TOWERVIEW                 RD                 2
## 15 YEARBY                    <NA>               2
## 16 ALEXANDER                 DORM               1
## 17 YEARBY                    RD                 1
```

```
durham21 %>%
  count(ncid) %>%
  arrange(desc(n))
```

```
## # A tibble: 11,513 x 2
##    ncid          n
##    <chr>     <int>
##  1 BL349867      5
##  2 BL471211      5
##  3 BL366593      3
##  4 BL373586      3
##  5 BL379711      3
##  6 BL397153      3
##  7 BL398874      3
##  8 BL398938      3
##  9 BL400822      3
## 10 BL401104      3
## # ... with 11,503 more rows
```

```
duke_students %>%
  count(registr_yr, party_desc) %>%
  filter(registr_yr >= 2000,
         registr_yr %% 4 == 0) %>%
  ggplot(aes(x = registr_yr, y = n, color = party_desc)) +
  geom_line() +
  geom_point() +
  scale_y_continuous(labels = scales::label_comma()) +
  scale_x_continuous(breaks = seq(2000, 2020, by = 4)) +
  scale_color_manual(values = c("blue4", "green3", "red3", "grey")) +
  labs(
    x = "Year",
    y = "Number of Duke students registered"
  ) +
  theme_minimal()
```
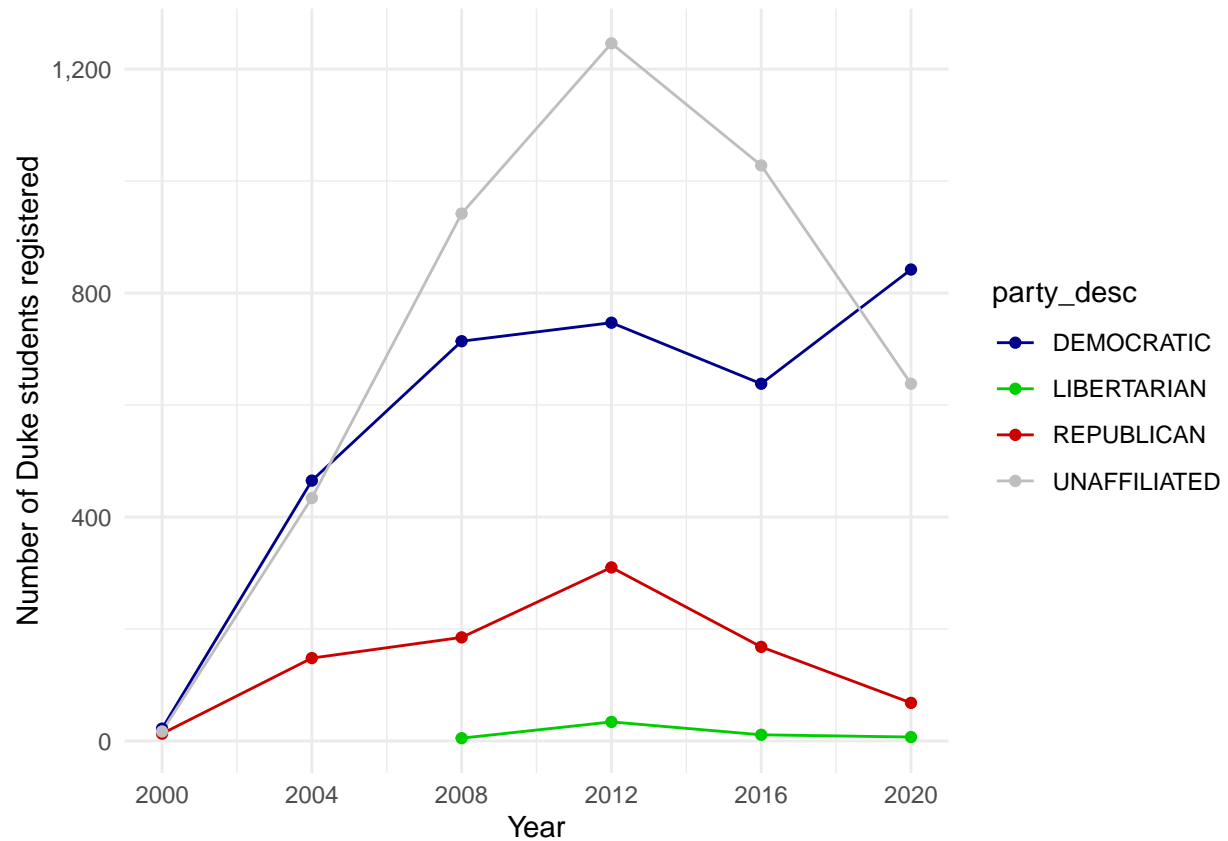
```
duke_students %>%
  filter(registr_yr >= 2000, registr_yr %% 4 == 0) %>%
  ggplot(aes(x = registr_yr, fill = fct_rev(fct_infreq(party_desc)))) +
  geom_bar(position = "fill") +
  scale_x_continuous(breaks = seq(2000, 2020, by = 4)) +
  scale_fill_manual(values = c("green3", "red3", "blue4", "grey")) +
  labs(
    x = "Year",
    y = NULL,
    fill = "Party",
    title = "Duke students' registered party over time"
  ) +
  theme_minimal()
```

Duke students' registered party over time
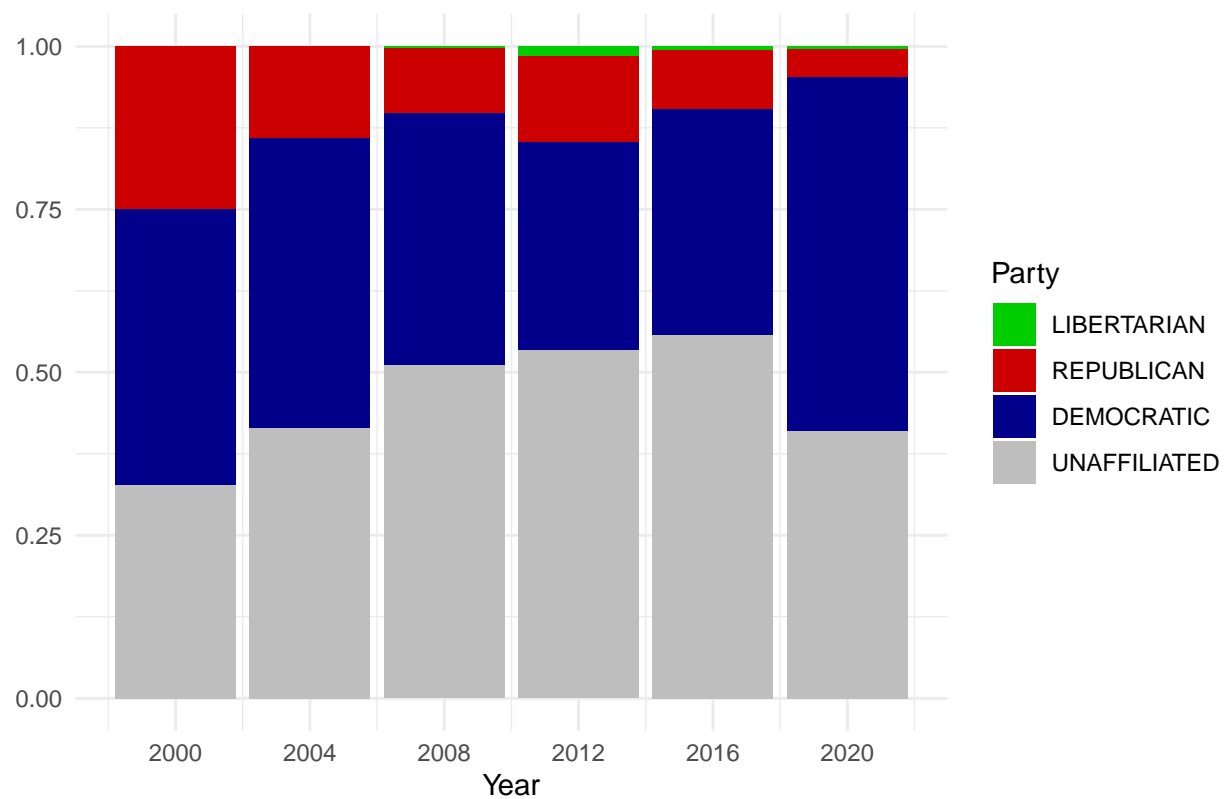

```
full_data %>%
  distinct(voter_reg_num, .keep_all = T) %>%
  count(voted_year) %>%
  ggplot(aes(x = voted_year, y = n)) +
  geom_line() +
  geom_line(data = election_year, color = "red") +
  geom_point(data = election_year, color = "red", size = 3) +
  scale_y_continuous(labels = scales::label_comma()) +
  scale_x_continuous(breaks = seq(2000, 2020, by = 4)) +
  labs(
    x = "Year",
    y = "Number of Duke students who voted"
  ) +
  theme_minimal()
```
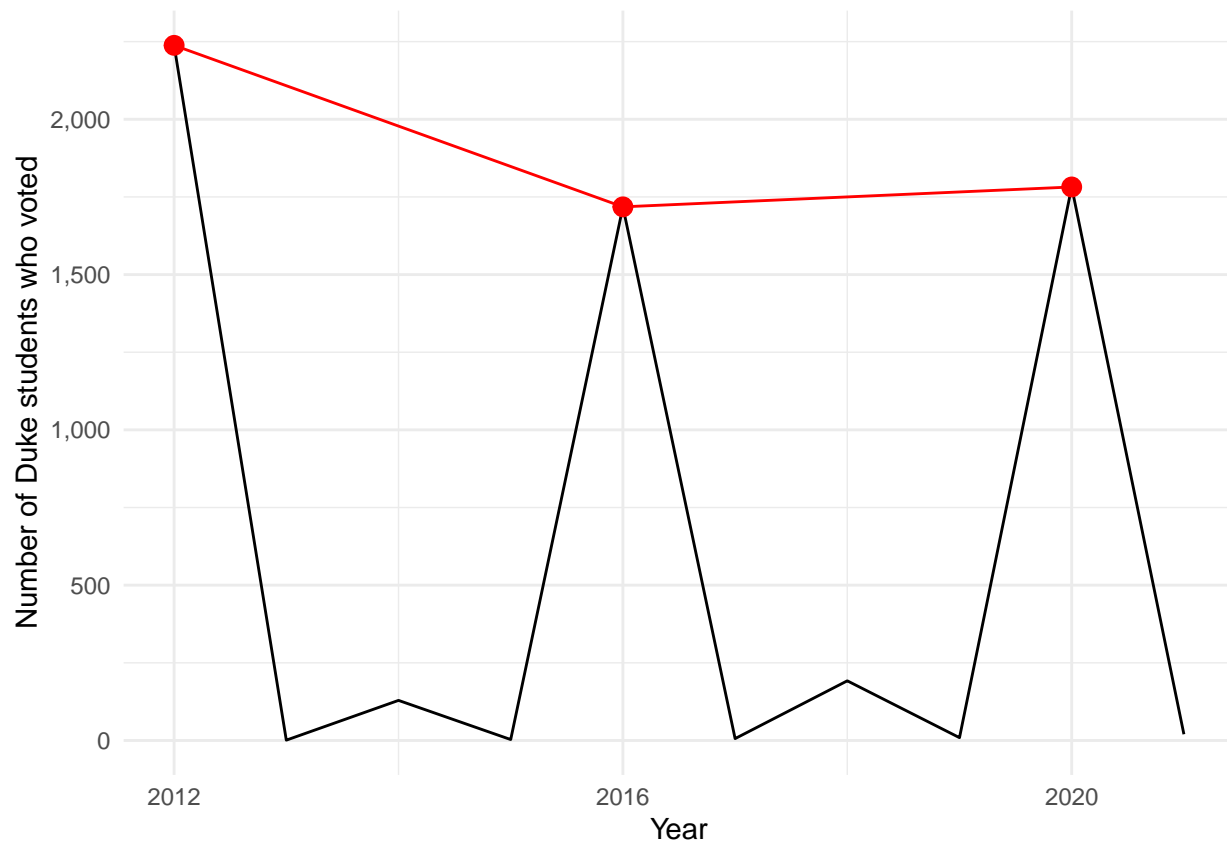
```
full_data %>%
  mutate(
    datev = as.Date(election_lbl, format = "%m/%d/%Y"),
    voted_year = lubridate::year(datev)
  ) %>%
  select(datev, voted_year, everything())
```

```
## # A tibble: 8,955 x 108
##     datev      voted_year snapshot_dt county_id.x county_desc.x voter_reg_num
##     <date>          <dbl> <date>            <dbl> <chr>         <chr>
##  1 2012-11-06       2012 2021-11-02           32 DURHAM        000030187158
##  2 2012-11-06       2012 2021-11-02           32 DURHAM        000030187136
##  3 2012-05-08       2012 2021-11-02           32 DURHAM        000030187136
##  4 2012-11-06       2012 2021-11-02           32 DURHAM        000030187135
##  5 2012-05-08       2012 2021-11-02           32 DURHAM        000030187135
##  6 2014-11-04       2014 2021-11-02           32 DURHAM        000030187135
##  7 2012-11-06       2012 2021-11-02           32 DURHAM        000030217456
##  8 2016-11-08       2016 2021-11-02           32 DURHAM        000030246068
##  9 2016-11-08       2016 2021-11-02           32 DURHAM        000030246078
## 10 2016-11-08       2016 2021-11-02           32 DURHAM        000030300681
## # ... with 8,945 more rows, and 102 more variables: ncid.x <chr>,
## #   status_cd <chr>, voter_status_desc <chr>, reason_cd <chr>,
## #   voter_status_reason_desc <chr>, absent_ind <lgl>, name_prefx_cd <lgl>,
## #   last_name <chr>, first_name <chr>, midl_name <chr>, name_sufx_cd <chr>,
## #   house_num <dbl>, half_code <lgl>, street_dir <lgl>, street_name <chr>,
```

```
## #    street_type_cd <chr>, street_sufx_cd <lgl>, unit_designator <lgl>,
## #    unit_num <chr>, res_city_desc <chr>, state_cd <chr>, zip_code <dbl>, ...
```

```
ggplot() +
  geom_line(data = election_yr_reg, aes(x = registr_yr, y = n, color = "Registered")) +
  geom_line(data = election_year, aes(x = voted_year, y = n, color = "Voted")) +
  geom_point(data = election_yr_reg, aes(x = registr_yr, y = n, color = "Registered")) +
  geom_point(data = election_year, aes(x = voted_year, y = n, color = "Voted")) +
  scale_y_continuous(labels = scales::label_comma()) +
  scale_x_continuous(breaks = seq(2012, 2020, by = 4), limits = c(2012, 2020)) +
  scale_color_manual(
    name = "Duke Students",
    breaks = c("Voted", "Registered"),
    values = c("Voted" = "blue", "Registered" = "red")
  ) +
  labs(
    x = "Year",
    y = "Number of Duke students",
    title = "Number of Duke students who voted vs who registered to vote"
  ) +
  theme_minimal()
```

```
## Warning: Removed 2 row(s) containing missing values (geom_path).
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```
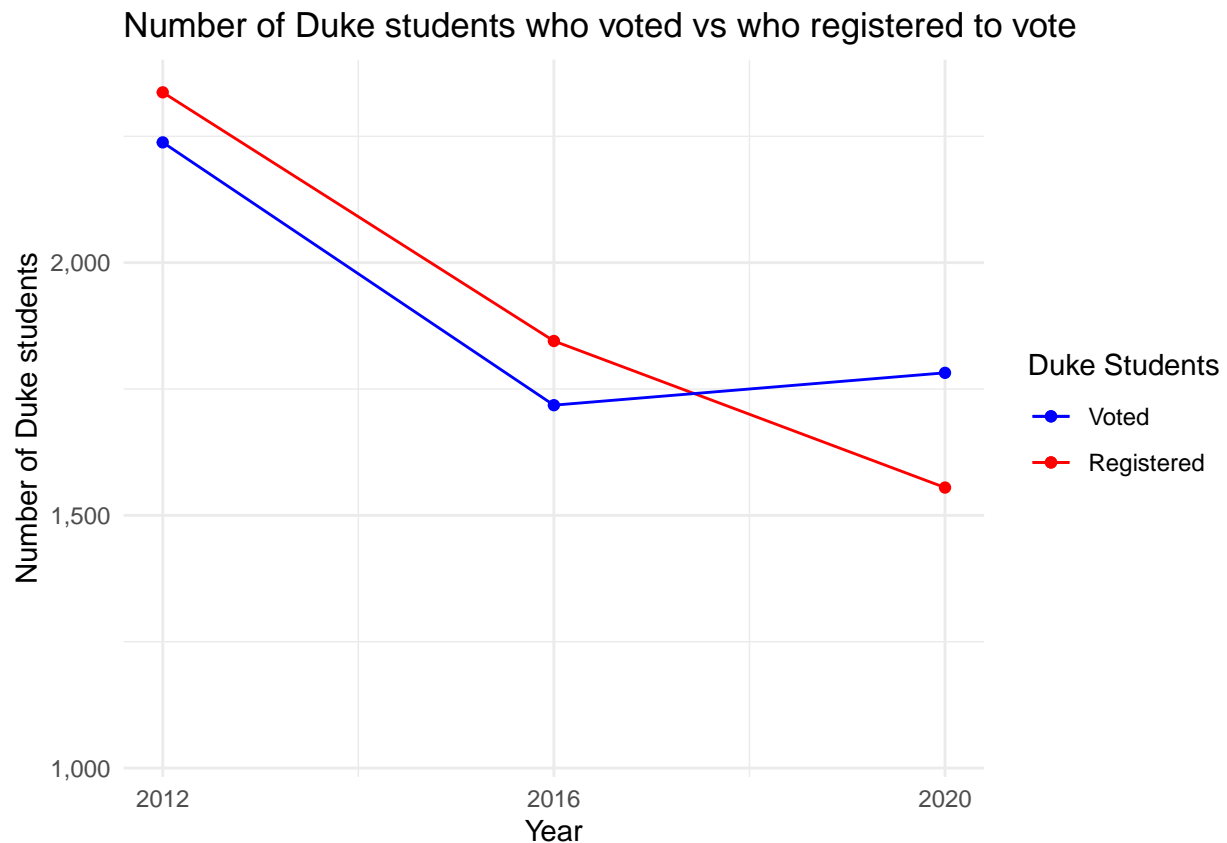

Number of Duke students who voted vs who registered to vote

```r
full_data %>%
  filter(voted_year %in% seq(2012, 2020, 4)) %>%
  ggplot(
    aes(x = voted_year, fill = fct_rev(fct_infreq(voting_method)))
  ) +
  geom_bar(position = "fill") +
  colorblindr::scale_fill_OkabeIto() +
  labs(
    x = "Year",
    y = NULL,
    fill = "Voting method",
    title = "Duke students' voting method over time"
  ) +
  theme_minimal()
```
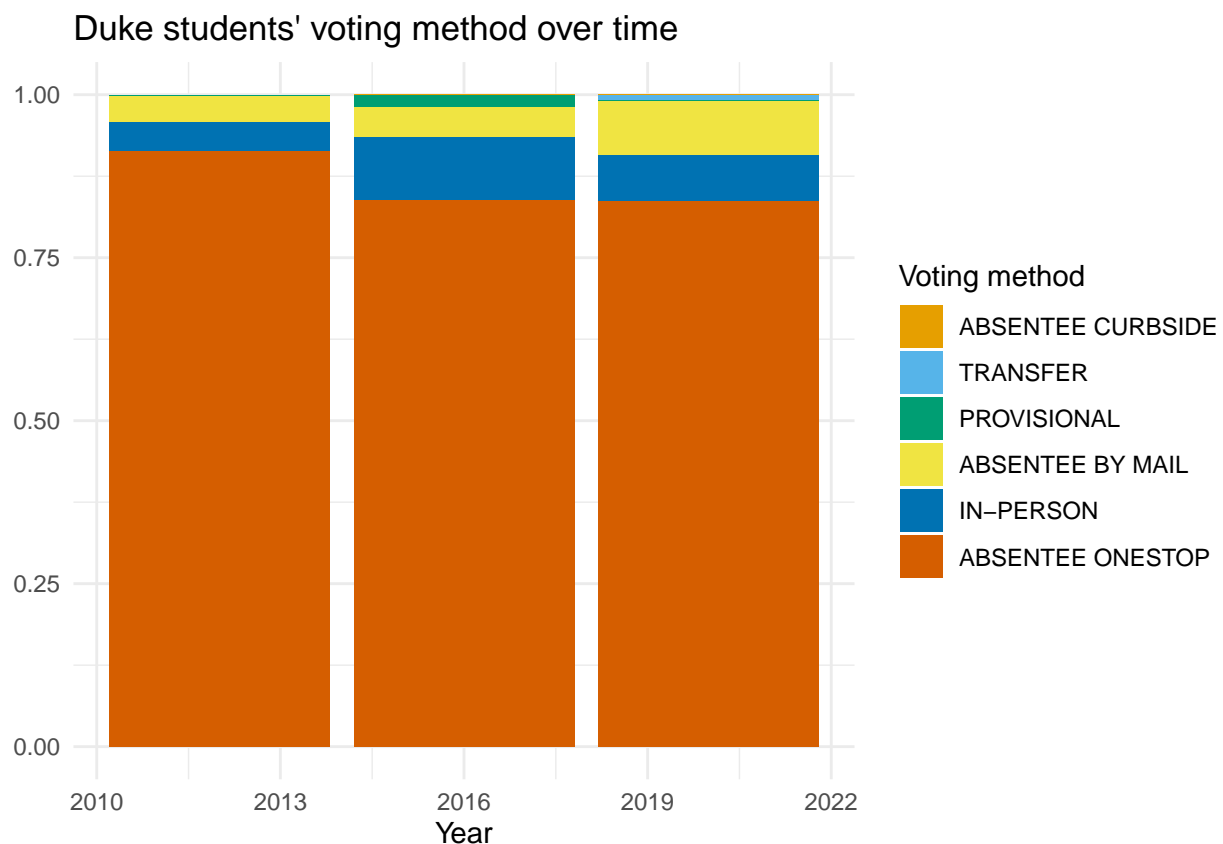


```r
full_data %>%
  filter(voted_year %in% seq(2012, 2020, 4)) %>%
  ggplot(
    aes(x = voted_year, fill = fct_rev(fct_infreq(voted_party_desc)))
  ) +
  geom_bar(position = "fill") +
  scale_fill_manual(values = c("green3", "red3", "grey", "blue4")) +
  labs(
    x = "Year",
    y = NULL,
```

```
    fill = "Party",
    title = "Duke students' voted party over time"
) +
theme_minimal()
```

Duke students' voted party over time