US Polling Places 2012-2020

Libraries/Packages

EDA

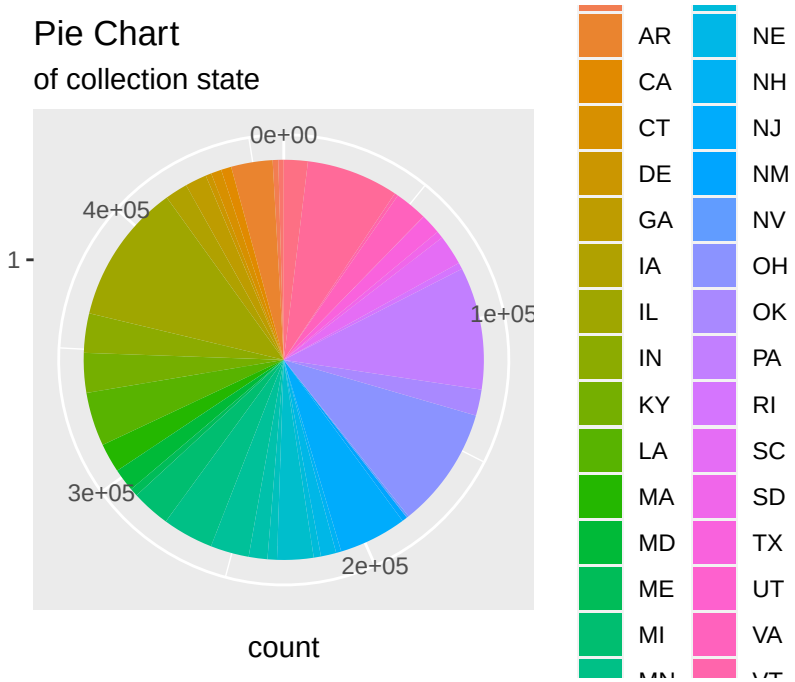
```
glimpse(polling_places)
```

```
Rows: 461,445
Columns: 15
$ election_date     <date> 2020-11-03, 2020-11-03, 2020-11-03, 2020-11-03, 202~
$ state             <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL"~
$ county_name       <chr> "AUTAUGA", "AUTAUGA", "AUTAUGA", "AUTAUGA", "AUTAUGA~
$ jurisdiction      <chr> "AUTAUGA", "AUTAUGA", "AUTAUGA", "AUTAUGA", "AUTAUGA~
$ jurisdiction_type <chr> "county", "county", "county", "county", "county", "c~
$ precinct_id       <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ precinct_name     <chr> "AUTAUGAVILLE VOL FIRE DEPT", "BILLINGSLEY COMMUNITY~
$ polling_place_id  <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ location_type     <chr> "election_day", "election_day", "election_day", "ele~
$ name              <chr> "AUTAUGAVILLE VOL FIRE DEPT", "BILLINGSLEY COMMUNITY~
$ address           <chr> "2610 HIGHWAY 14 W, AUTAUGAVILLE, AL 36003", "2159 C~
$ notes             <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ source            <chr> "ORR", "ORR", "ORR", "ORR", "ORR", "ORR", "ORR", "OR~
$ source_date       <date> 2020-10-21, 2020-10-21, 2020-10-21, 2020-10-21, 202~
$ source_notes      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
```

```
polling_places <- polling_places %>%
  mutate(year = substring(election_date, 0, 4))
```

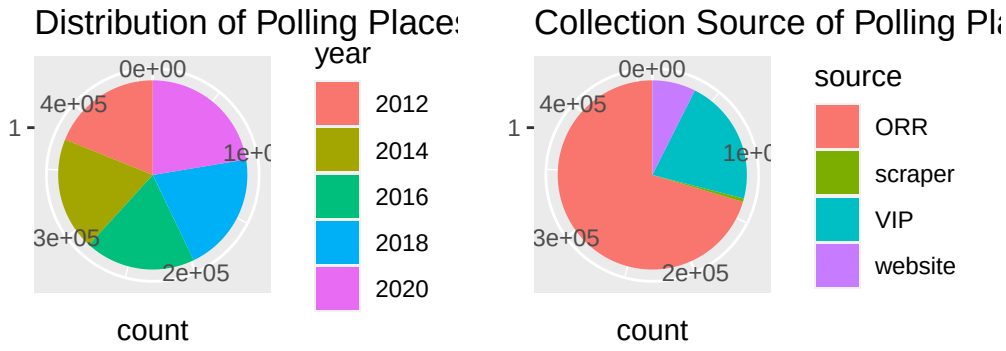
```
ggplot(data = polling_places, aes(x = factor(1), fill = state)) +
  geom_bar() +
```

```
  coord_polar("y") +
  labs(x = "",
       title = "Pie Chart",
       subtitle = "of collection state")
```



While colorful, there's not much to be gleamed from this chart.

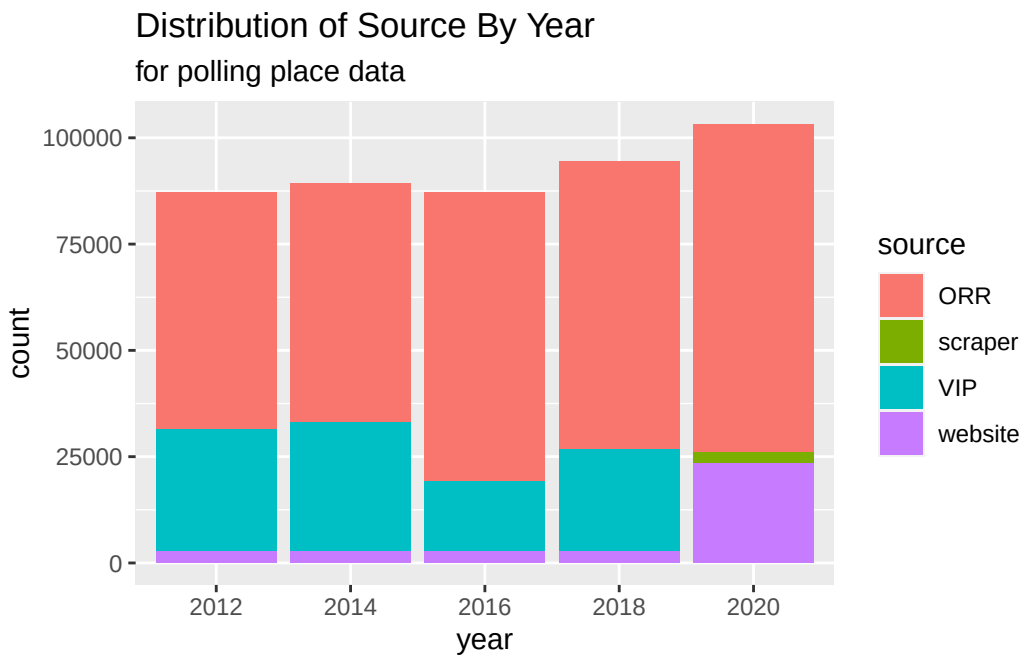
```
p1 <- ggplot(data = polling_places, aes(x = factor(1), fill = year)) +
  geom_bar() +
  coord_polar("y") +
  labs(x = "",
       title = "Distribution of Polling Places By Year")

p2 <- ggplot(data = polling_places, aes(x = factor(1), fill = source)) +
  geom_bar() +
  coord_polar("y") +
  labs(x = "",
       title = "Collection Source of Polling Places")

p1+p2
```



Looks like the distribution of year is very even, about the same number of votes collected from each year.

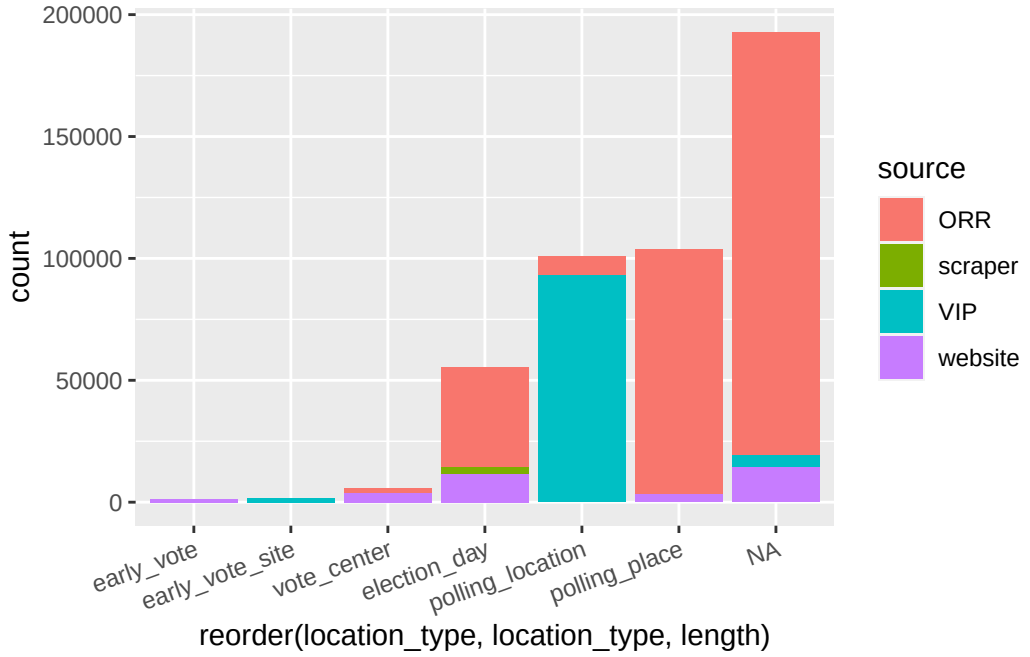
The distribution of source on the other hand, is very dominantly ORR

```
ggplot(data = polling_places, aes(x = year, fill = source)) +
  geom_bar() +
  labs(title = "Distribution of Source By Year",
       subtitle = "for polling place data"
  )
```



It looks like the distribution of source by year is mostly the same until 2020 where the distribution of website increased and scraper was introduced. ORR remains the dominant source, but the distribution of the others changed dramatically.

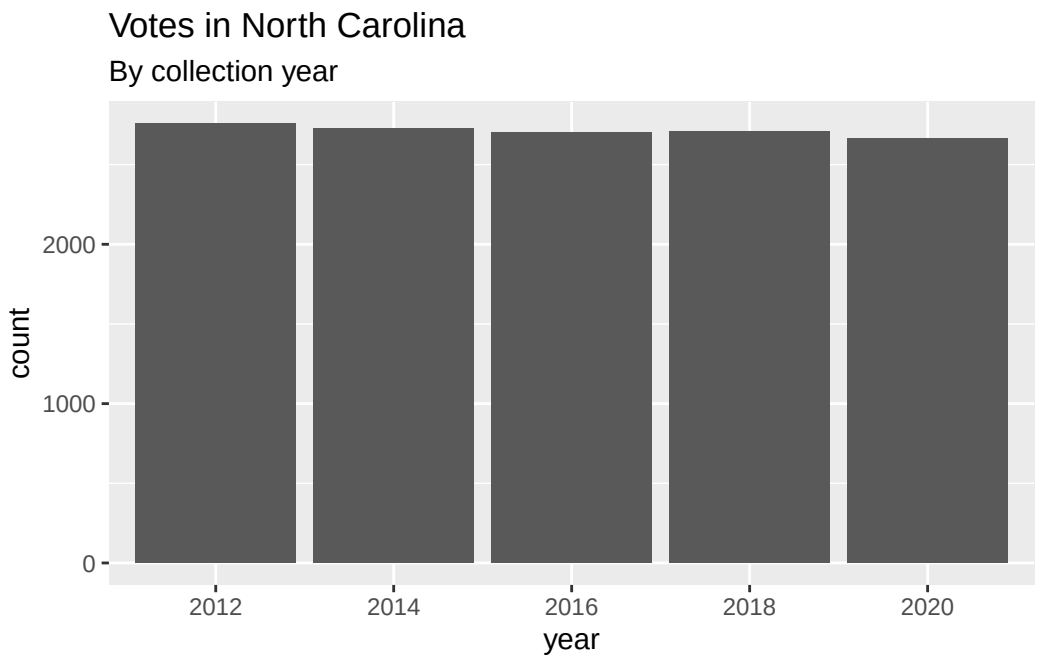
```
ggplot(data = polling_places, aes(x = reorder(location_type, location_type, length), fill
  geom_bar() +
  theme(axis.text.x = element_text(angle = 20, hjust = 1, vjust = 1))
```



It looks like almost half of the location types aren't classified so this doesn't seem like a variable of great interest

Deeper Analysis

```
b1 <- ggplot(data = filter(polling_places, state == "NC"), aes(x = year)) +
  geom_bar() +
  labs(title = "Votes in North Carolina",
       subtitle = "By collection year")
b1
```




```
b2 <- layer_data(b1, 1)
b2
```

```
   y count prop x flipped_aes PANEL group ymin ymax xmin xmax colour   fill
1 2756  2756    1 1       FALSE     1     1    0 2756 0.55 1.45     NA grey35
2 2726  2726    1 2       FALSE     1     2    0 2726 1.55 2.45     NA grey35
3 2704  2704    1 3       FALSE     1     3    0 2704 2.55 3.45     NA grey35
4 2706  2706    1 4       FALSE     1     4    0 2706 3.55 4.45     NA grey35
5 2662  2662    1 5       FALSE     1     5    0 2662 4.55 5.45     NA grey35
  linewidth linetype alpha
1       0.5        1    NA
2       0.5        1    NA
3       0.5        1    NA
4       0.5        1    NA
5       0.5        1    NA
```

```
uniquelist <- unique(polling_places$state) |> tibble()

names(uniquelist)[1] <- "States"

mutate(uniquelist, total2020 = as.numeric(count(filter(polling_places, year == "2020"))))
```

```
# A tibble: 39 x 2
   States total2020
   <chr>      <dbl>
 1 AL        103290
 2 AK        103290
 3 AR        103290
 4 CA        103290
 5 CT        103290
 6 DE        103290
 7 GA        103290
 8 IL        103290
 9 IN        103290
10 IA        103290
# i 29 more rows
```

```
count(filter(polling_places, state == "NC", year == "2020")) - count(filter(polling_places
```

```
   n
1 -94
```

```
ggplot(data = filter(polling_places, year == "2020"), aes(x = state)) +
  geom_bar()
```