Yale Voting Process Survey Analysis

Darwin Do

10/30/2020

```
##
## 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
```

```
##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:tidyr':
##
##     extract
```

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
d <- read_csv("voting_survey.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
names(d)
```

```
##  [1] "StartDate"             "EndDate"               "Status"
##  [4] "IPAddress"             "Progress"              "Duration (in seconds)"
##  [7] "Finished"              "RecordedDate"          "ResponseId"
## [10] "RecipientLastName"     "RecipientFirstName"    "RecipientEmail"
## [13] "ExternalReference"     "LocationLatitude"      "LocationLongitude"
## [16] "DistributionChannel"   "UserLanguage"          "undergrad"
## [19] "registered"            "party"                 "party_3_TEXT"
## [22] "voted"                 "plan_to_vote"          "vote_difficulty"
## [25] "vote_diff_expect"      "vote_satisfied"        "vote_resources"
## [28] "vote_resources_4_TEXT" "vote_method"           "home"
## [31] "enrollment"            "new_haven"             "current_residence"
## [34] "current_registration"  "email"
```

```
# only take relevant columns
d %<>% select(grep("^[a-z]", names(d)), EndDate)

# remove the first two rows
d %<>% slice_tail(n=dim(d)[1] - 2)

# get rid of test responses
d %<>% mutate(EndDate = ymd_hms(EndDate)) %>%
  filter(EndDate > ymd(20201028))

# there are 2 NAs for undergrads, not sure how but I'm tossing them out
d %<>% filter(!is.na(undergrad))
```

Survey Summaries:

```
# total number of valid participants:
d %>% group_by(undergrad) %>% summarise(count = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   undergrad count
##   <chr>     <int>
## 1 No            2
## 2 Yes        1081
```

```
undergrads_only <- d %>% filter(undergrad == "Yes")
num_undergrad <- d %>% group_by(undergrad) %>% summarise(count = n()) %>% filter(undergr
ad == "Yes") %>% pull()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# registration
d %>% group_by(registered) %>% summarise(count = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
##   registered count
##   <chr>      <int>
## 1 No            75
## 2 Yes         1006
## 3 <NA>           2
```

```
registered_only <- d %>% filter(registered == "Yes")
registered_voters <- d %>% group_by(registered) %>%summarise(count = n()) %>% filter(registered == "Yes") %>% pull()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

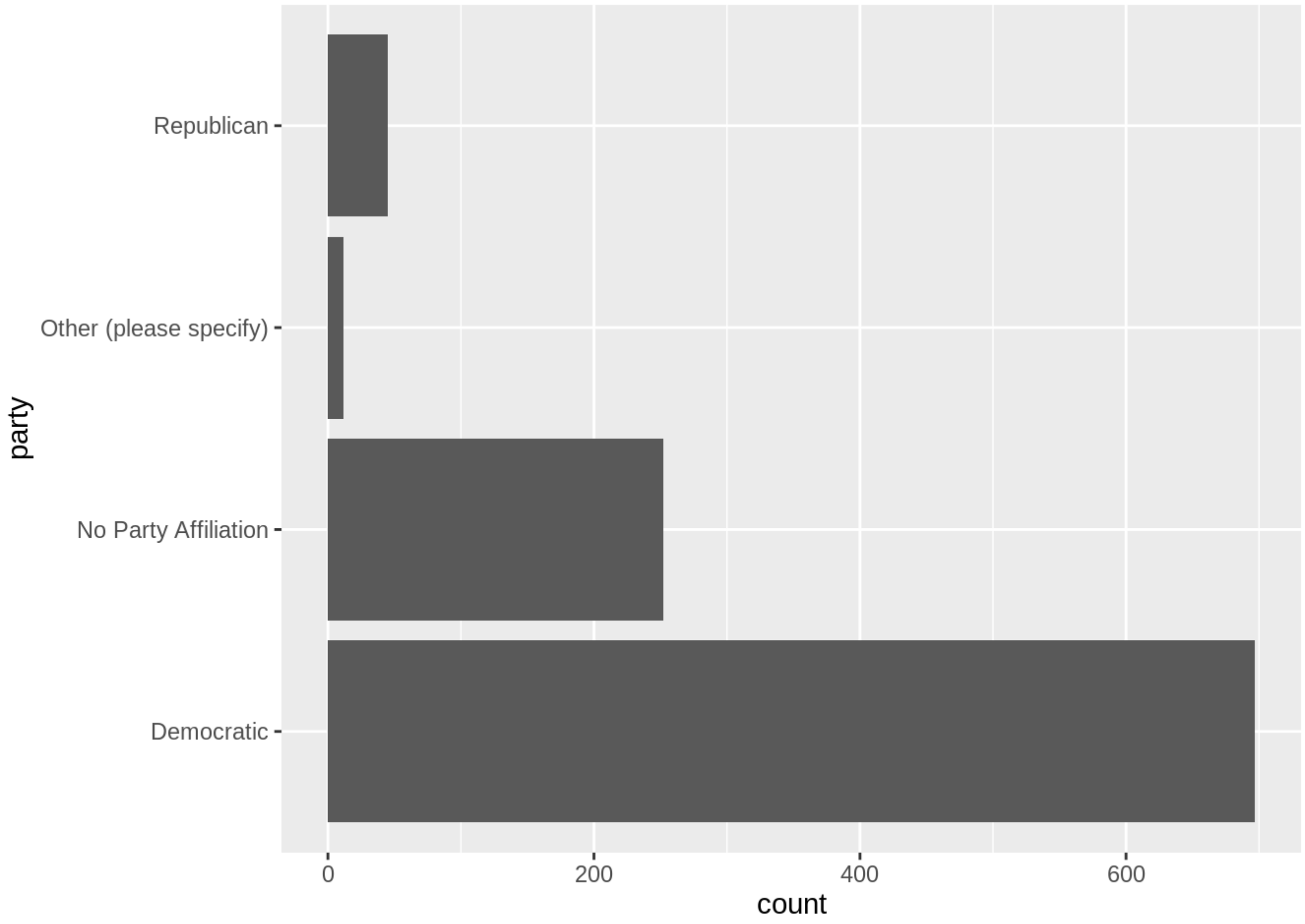
```
# plan to vote?
registered_only %>%
  group_by(voted, plan_to_vote) %>%
  summarise(count = n())
```

```
## `summarise()` regrouping output by 'voted' (override with `.groups` argument)
```

```
## # A tibble: 3 x 3
## # Groups:   voted [2]
##   voted plan_to_vote count
##   <chr> <chr>        <int>
## 1 No    No              10
## 2 No    Yes            125
## 3 Yes   <NA>           871
```

Breakdown by Party

```
# breakdown by party
d %>% filter(!is.na(party)) %>%
  ggplot() +
  geom_bar(aes(y = party))
```



```
registered_only %>% group_by(party) %>% summarise(n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 4 x 2
##   party                  `n()`
##   <chr>                  <int>
## 1 Democratic               697
## 2 No Party Affiliation     252
## 3 Other (please specify)    12
## 4 Republican                45
```

Voting Turnout

```
registered_tbl <- d %>% group_by(registered) %>% summarise(count=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# num of registered voters
(registered_voters <- registered_tbl %>% filter(registered == "Yes") %>% pull())
```

```
## [1] 1006
```

```
d %>% filter(registered == "Yes") %>%
  group_by(voted, plan_to_vote) %>%
  summarise(count = n())
```

```
## `summarise()` regrouping output by 'voted' (override with `.groups` argument)
```

```
## # A tibble: 3 x 3
## # Groups:   voted [2]
##   voted plan_to_vote count
##   <chr> <chr>        <int>
## 1 No    No              10
## 2 No    Yes            125
## 3 Yes   <NA>           871
```

```
# out of the 1007 registered voters, 872 have already voted, 125 plan to vote, and 10 do
not plan to vote

voted_only <- registered_only %>% filter(voted == "Yes")
num_voted <- dim(voted_only)[1]
voted_only %>% group_by(vote_difficulty) %>% summarise(count = n(), pct = count / num_vo
ted * 100)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 3
##   vote_difficulty            count    pct
##   <chr>                      <int>  <dbl>
## 1 Extremely difficult           16  1.84
## 2 Extremely easy               246 28.2
## 3 Neither easy nor difficult    92 10.6
## 4 Somewhat difficult           133 15.3
## 5 Somewhat easy                383 44.0
## 6 <NA>                           1  0.115
```

```
plan_to_vote_only <- registered_only %>% filter(plan_to_vote == "Yes")
num_plan_to_vote <- dim(plan_to_vote_only)[1]
plan_to_vote_only %>% group_by(vote_diff_expect) %>% summarise(count = n(), pct = count
 / num_plan_to_vote * 100)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 5 x 3
##   vote_diff_expect           count   pct
##   <chr>                      <int> <dbl>
## 1 Extremely difficult            8   6.4
## 2 Extremely easy                18  14.4
## 3 Neither easy nor difficult    27  21.6
## 4 Somewhat difficult            28  22.4
## 5 Somewhat easy                 44  35.2
```

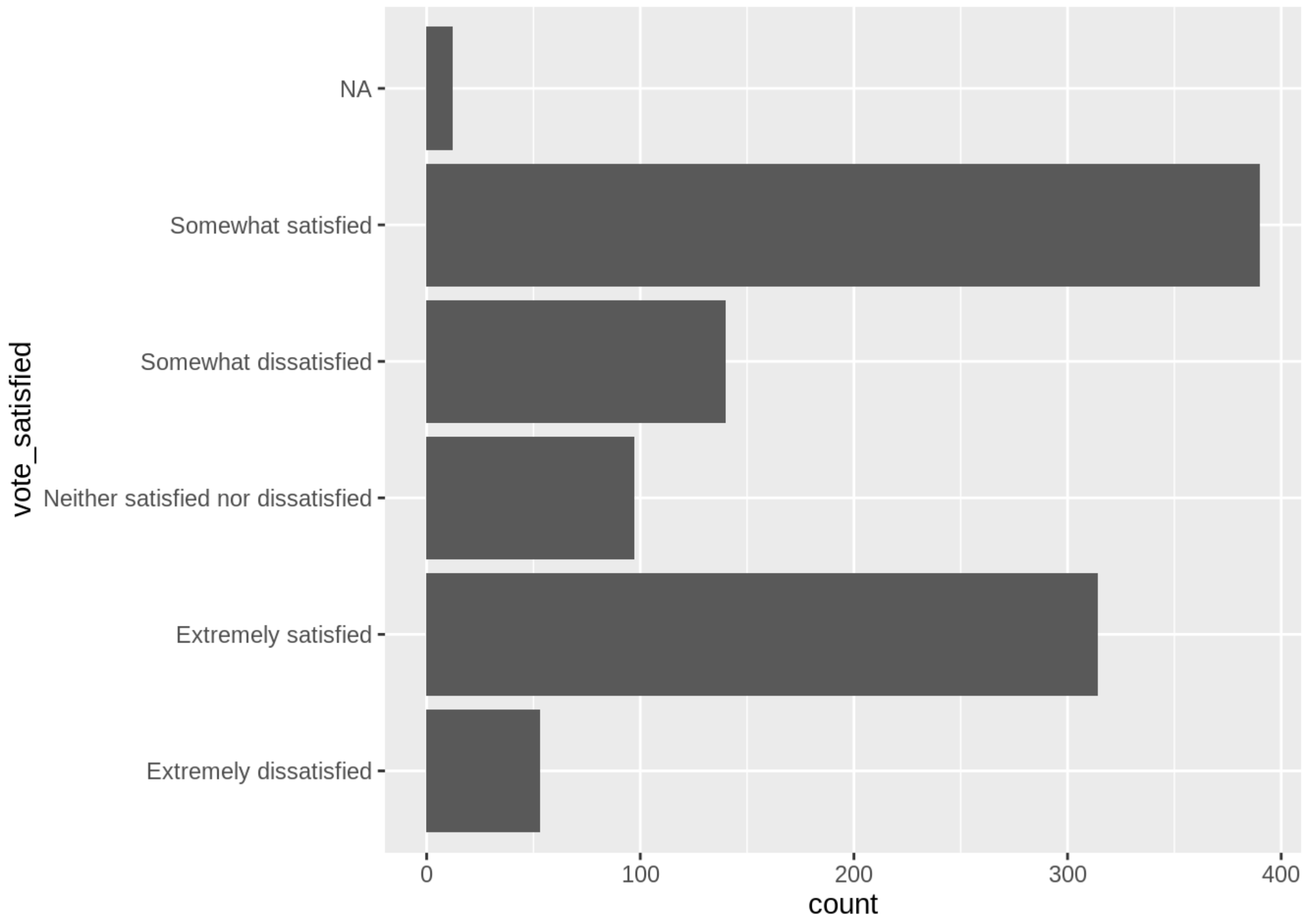
Satisfaction

```
registered_only %>%
  group_by(vote_satisfied) %>%
  summarise(count = n(), pct = count / registered_voters * 100)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 3
##   vote_satisfied                     count   pct
##   <chr>                              <int> <dbl>
## 1 Extremely dissatisfied                53  5.27
## 2 Extremely satisfied                  314 31.2 
## 3 Neither satisfied nor dissatisfied    97  9.64
## 4 Somewhat dissatisfied                140 13.9 
## 5 Somewhat satisfied                   390 38.8 
## 6 <NA>                                  12  1.19
```

```
registered_only %>%
  ggplot() +
  geom_bar(aes(y = vote_satisfied))
```



Resources

```
registered_only %>%
  group_by(vote_resources) %>%
  summarise(count = n(), pct = count / registered_voters * 100)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 13 x 3
##    vote_resources                                                  count     pct
##    <chr>                                                           <int>   <dbl>
##  1 Every Vote Counts                                                  17  1.69  
##  2 Every Vote Counts,Yale Administration                               1  0.0994
##  3 Other (please specify)                                             18  1.79  
##  4 Yale Administration                                                11  1.09  
##  5 Yale Votes                                                        166 16.5   
##  6 Yale Votes,Every Vote Counts                                       33  3.28  
##  7 Yale Votes,Every Vote Counts,Other (please specify)                 1  0.0994
##  8 Yale Votes,Every Vote Counts,Yale Administration                   19  1.89  
##  9 Yale Votes,Every Vote Counts,Yale Administration,Other (please…     1  0.0994
## 10 Yale Votes,Other (please specify)                                   3  0.298 
## 11 Yale Votes,Yale Administration                                     14  1.39  
## 12 Yale Votes,Yale Administration,Other (please specify)               1  0.0994
## 13 <NA>                                                              721 71.7   
```

```
registered_only %>% filter(!is.na(vote_resources))
```

```
## # A tibble: 285 x 19
##    undergrad registered party party_3_TEXT voted plan_to_vote vote_difficulty
##    <chr>     <chr>      <chr> <chr>        <chr> <chr>        <chr>          
##  1 Yes       Yes        Demo… <NA>         Yes   <NA>         Extremely easy 
##  2 Yes       Yes        Demo… <NA>         No    Yes          <NA>           
##  3 Yes       Yes        Demo… <NA>         Yes   <NA>         Somewhat easy  
##  4 Yes       Yes        No P… <NA>         No    Yes          <NA>           
##  5 Yes       Yes        Demo… <NA>         Yes   <NA>         Extremely easy 
##  6 Yes       Yes        Demo… <NA>         Yes   <NA>         Somewhat diffi…
##  7 Yes       Yes        Demo… <NA>         Yes   <NA>         Extremely easy 
##  8 Yes       Yes        Demo… <NA>         No    Yes          <NA>           
##  9 Yes       Yes        Demo… <NA>         Yes   <NA>         Somewhat diffi…
## 10 Yes       Yes        Demo… <NA>         Yes   <NA>         Somewhat diffi…
## # … with 275 more rows, and 12 more variables: vote_diff_expect <chr>,
## #   vote_satisfied <chr>, vote_resources <chr>, vote_resources_4_TEXT <chr>,
## #   vote_method <chr>, home <chr>, enrollment <chr>, new_haven <chr>,
## #   current_residence <chr>, current_registration <chr>, email <chr>,
## #   EndDate <dttm>
```

Vote Method

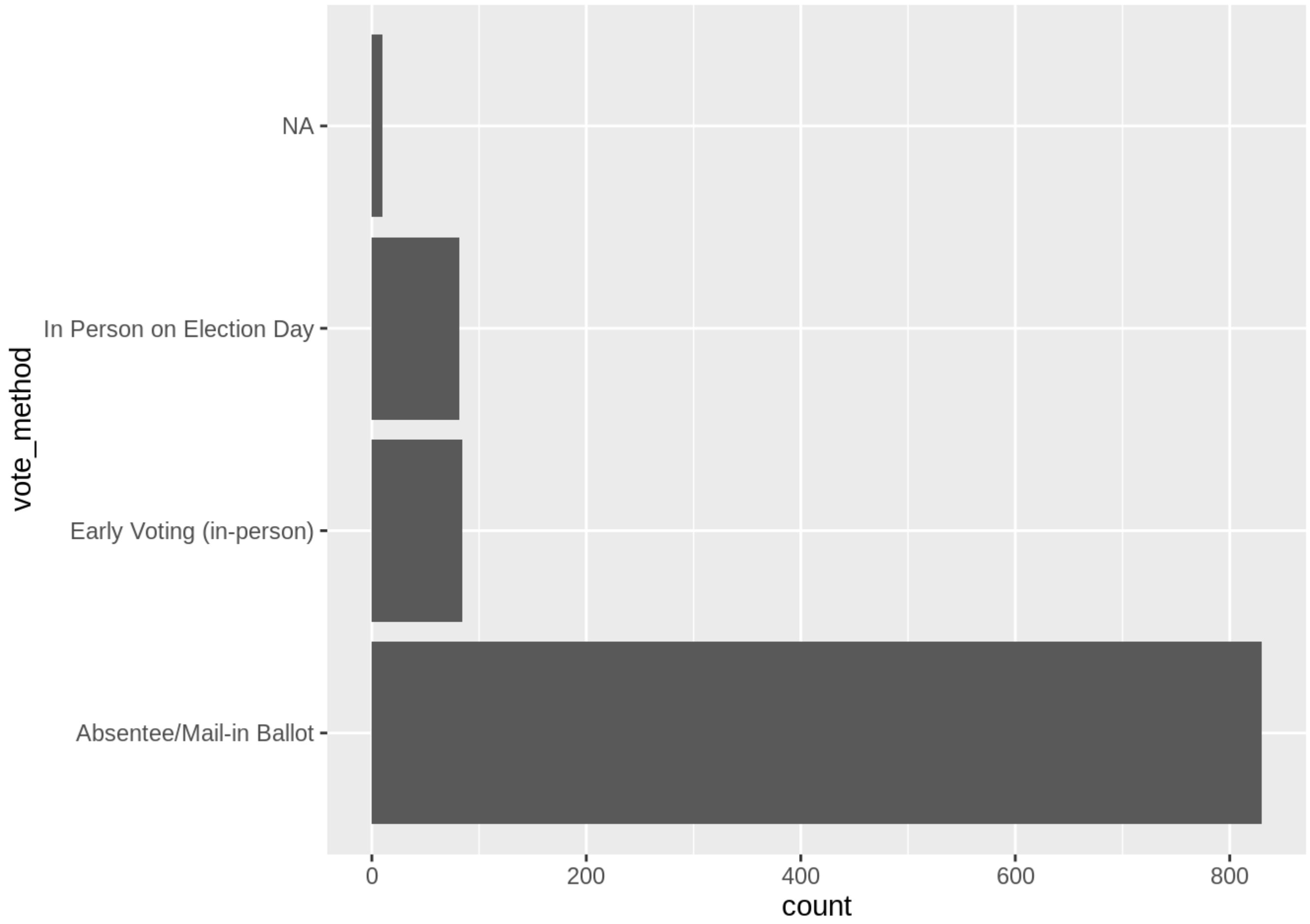
```
registered_only %>%
  group_by(vote_method) %>%
  summarise(count = n(), pct = count / registered_voters * 100)
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
## # A tibble: 4 x 3
##   vote_method               count    pct
##   <chr>                     <int>  <dbl>
## 1 Absentee/Mail-in Ballot     830 82.5
## 2 Early Voting (in-person)     84  8.35
## 3 In Person on Election Day    82  8.15
## 4 <NA>                         10  0.994
```

```
registered_only %>% filter(vote_method == "Absentee/Mail-in Ballot" & voted == "No")
```

```
## # A tibble: 40 x 19
##    undergrad registered party party_3_TEXT voted plan_to_vote vote_difficulty
##    <chr>     <chr>      <chr> <chr>        <chr> <chr>        <chr>
##  1 Yes       Yes        No P… <NA>         No    Yes          <NA>
##  2 Yes       Yes        Demo… <NA>         No    Yes          <NA>
##  3 Yes       Yes        Demo… <NA>         No    Yes          <NA>
##  4 Yes       Yes        Demo… <NA>         No    Yes          <NA>
##  5 Yes       Yes        No P… <NA>         No    Yes          <NA>
##  6 Yes       Yes        Demo… <NA>         No    Yes          <NA>
##  7 Yes       Yes        Demo… <NA>         No    Yes          <NA>
##  8 Yes       Yes        No P… <NA>         No    Yes          <NA>
##  9 Yes       Yes        Demo… <NA>         No    Yes          <NA>
## 10 Yes       Yes        No P… <NA>         No    Yes          <NA>
## # … with 30 more rows, and 12 more variables: vote_diff_expect <chr>,
## #   vote_satisfied <chr>, vote_resources <chr>, vote_resources_4_TEXT <chr>,
## #   vote_method <chr>, home <chr>, enrollment <chr>, new_haven <chr>,
## #   current_residence <chr>, current_registration <chr>, email <chr>,
## #   EndDate <dttm>
```

```
registered_only %>%
  ggplot() +
  geom_bar(aes(y = vote_method))
```


NA
In Person on Election Day
Early Voting (in-person)
Absentee/Mail-in Ballot
vote_method
0
200
400
600
800
count