

```

library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

poll <- read.csv("~/12Nov) 2024 exit poll data - MASTER.csv") %>% mutate(Z_REPRESENT = recode(Z_REPRESENT,
  `1` = "BIDEN",
  `2` = "HARRIS",
  `3` = "SANDERS",
  `4` = "PHILLIPS",
  `5` = "OTHER",
  `6` = "TRUMP",
  `7` = "HALEY",
  `8` = "DESANTIS",
  `9` = "RFK_JR"),
  B_PRES = recode(B_PRES,
    `1` = "TRUMP",
    `2` = "HARRIS",
    `3` = "OTHER"),
  C_HOUSE = recode(C_HOUSE,
    `1` = "TEIRAB",
    `2` = "CRAIG",
    `3` = "OTHER"),
  D_SENATE = recode(D_SENATE,
    `1` = "KLOBUCHAR",
    `2` = "WHITE",
    `3` = "OTHER"),
  F_PID = recode(F_PID,
    `1` = "STRONG DEM",
    `2` = "LEAN DEM",
    `3` = "INDEPENDENT",
    `3.5` = "LEAN REP",
    `4` = "LEAN REP",
    `5` = "STRONG REP"))

```

Hypothesis A: A voter who prefers Biden, Trump, or Harris is less likely to split-ticket vote than a voter who prefers one of the other candidates.

Hypothesis B: Voters from more conservative precincts will have a higher ratio of Biden support to Harris support, relative to more suburban or urban districts. RESULT: INSUFFICIENT DATA. Only five people preferred Biden.

Hypothesis C: Support for Biden/Harris and Trump in this survey will exceed the amount of support they got relative to their primary opponents (Dean Phillips / RFK / Other (Dem vote); Haley / DeSantis / Other (R vote)) in the 2024 primaries.

#EDA

```
#poll %>% select(Number, B_PRES, C_HOUSE, D_SENATE, Z_REPRESENT)
```

```
# 5 Americans preferred one of the nominees but voted against him / her:
```

```
poll %>% filter(Z_REPRESENT == "HARRIS" | Z_REPRESENT == "BIDEN" | Z_REPRESENT == "TRUMP") %>% count(Z_REPRESENT)
```

```
##      Z_REPRESENT B_PRES      n
## 1      BIDEN HARRIS      4
## 2      BIDEN  OTHER      1
## 3      HARRIS HARRIS    214
## 4      HARRIS  OTHER      1
## 5      HARRIS  TRUMP       3
## 6      HARRIS   <NA>       1
## 7      TRUMP HARRIS       2
## 8      TRUMP  OTHER       1
## 9      TRUMP  TRUMP    144
## 10     TRUMP   <NA>       2
```

```
#Exploring this:
```

```
#Wanted Trump, voted Harris
```

```
poll %>% select(Number, B_PRES, C_HOUSE, D_SENATE, Z_REPRESENT, F_PID) %>% filter(Z_REPRESENT == "TRUMP")
```

```
##      Number B_PRES C_HOUSE D_SENATE Z_REPRESENT      F_PID
## 1      38 HARRIS  TEIRAB KLOBUCHAR      TRUMP LEAN REP
## 2      72 HARRIS  TEIRAB KLOBUCHAR      TRUMP LEAN REP
```

```
#Wanted Harris, voted Trump
```

```
poll %>% select(Number, B_PRES, C_HOUSE, D_SENATE, Z_REPRESENT, F_PID) %>% filter(Z_REPRESENT == "HARRIS")
```

```
##      Number B_PRES C_HOUSE D_SENATE Z_REPRESENT      F_PID
## 1      41  TRUMP  TEIRAB      WHITE      HARRIS  LEAN DEM
## 2      42  TRUMP   CRAIG KLOBUCHAR      HARRIS  LEAN DEM
## 3       8  TRUMP  TEIRAB KLOBUCHAR      HARRIS INDEPENDENT
```

```
#PID vs Represent
```

```
poll %>% count(F_PID, Z_REPRESENT) %>% pivot_wider(values_from = n, names_from = F_PID) %>% mutate(across(everything(), ~ as.numeric(.)))
```

```
## # A tibble: 10 x 7
```

```
##      Z_REPRESENT `STRONG DEM` `LEAN DEM` INDEPENDENT `LEAN REP` `STRONG REP` `NA`
##      <chr>          <int>      <int>      <int>      <int>      <int> <int>
## 1 BIDEN              1          3          1          0          0      0
## 2 DESANTIS            0          0          2         13          4      0
## 3 HALEY              0          1         16         15          5      0
## 4 HARRIS            92         83         35          6          3      0
## 5 OTHER              4          6         13         15          7      2
## 6 PHILLIPS           1          2          5          1          0      0
## 7 RFK_JR             0          5         17         11          3      0
## 8 SANDERS            19         14          6          3          0      0
## 9 TRUMP              0          5         20         66         57      1
## 10 <NA>             20         15         16         26         18      7
```

```
Vote vs. Preference
```

```
poll %>% count(Z_REPRESENT, B_PRES) %>% pivot_wider(names_from = B_PRES, values_from = n)
```

```
## # A tibble: 10 x 5
```

```
##      Z_REPRESENT HARRIS OTHER TRUMP `NA`
##      <chr>          <int> <int> <int> <int>
```

```
## 1 BIDEN          4      1    NA    NA
## 2 DESANTIS       3      1    15    NA
## 3 HALEY          14      5    18    NA
## 4 HARRIS        214     1     3     1
## 5 OTHER          14      3    29     1
## 6 PHILLIPS        7      1     1    NA
## 7 RFK_JR          8      4    23     1
## 8 SANDERS        38      2     2    NA
## 9 TRUMP           2      1   144     2
## 10 <NA>          45      3    42    12
```

#Hypothesis A: Split-Ticket Voting

#View of split vote rate by candidate

```
poll %>% select(Number, B_PRES, C_HOUSE, D_SENATE, Z_REPRESENT, F_PID) %>% mutate(across(2:6, ~replace_na(
  mutate(DemVote = ifelse(B_PRES == "HARRIS" | C_HOUSE == "CRAIG" | D_SENATE == "KLOBUCHAR", 1, 0),
    RepVote = ifelse(B_PRES == "TRUMP" | C_HOUSE == "TEIRAB" | D_SENATE == "WHITE", 1, 0),
    SplitVote = DemVote * RepVote) %>% group_by(Z_REPRESENT) %>% summarise(SplitVote = mean(SplitVote)
```

A tibble: 10 x 3

```
##   Z_REPRESENT SplitVote      n
##   <chr>          <dbl> <int>
## 1 BIDEN          0         5
## 2 DESANTIS       0.368     19
## 3 HALEY          0.243     37
## 4 HARRIS         0.0365    219
## 5 NONE          0.0980    102
## 6 OTHER          0.170     47
## 7 PHILLIPS       0.111      9
## 8 RFK_JR         0.194     36
## 9 SANDERS        0.0714     42
## 10 TRUMP         0.154    149
```

#View by Nominee vs Not Nominee

```
SplTicket_poll <-poll %>% select(Number, B_PRES, C_HOUSE, D_SENATE, Z_REPRESENT, F_PID) %>% mutate(across(
  mutate(PartyNominee = ifelse(Z_REPRESENT == "BIDEN" | Z_REPRESENT == "HARRIS" | Z_REPRESENT == "TRUMP",
    DemVote = ifelse(B_PRES == "HARRIS" | C_HOUSE == "CRAIG" | D_SENATE == "KLOBUCHAR", 1, 0),
    RepVote = ifelse(B_PRES == "TRUMP" | C_HOUSE == "TEIRAB" | D_SENATE == "WHITE", 1, 0),
    SplitVote = DemVote * RepVote)
```

```
SplTicket_poll %>% group_by(PartyNominee) %>% summarise(SplitVoteTotal = sum(SplitVote), SplitVoteRate =
```

A tibble: 2 x 4

```
##   PartyNominee SplitVoteTotal SplitVoteRate      n
##   <chr>          <dbl>          <dbl> <int>
## 1 Nominee          31          0.0831    373
## 2 Not_Nominee      45          0.154     292
```

```
Nom_Tick <-SplTicket_poll %>% filter(PartyNominee == "Nominee") %>% select(SplitVote)
```

```
Not_Nom_Tick <-SplTicket_poll %>% filter(PartyNominee == "Not_Nominee") %>% select(SplitVote)
```

```
t.test(Nom_Tick, Not_Nom_Tick, var.equal = TRUE)
```

##

Two Sample t-test

##

data: Nom_Tick and Not_Nom_Tick

```
## t = -2.8693, df = 663, p-value = 0.004245
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.11958707 -0.02241227
## sample estimates:
## mean of x mean of y
## 0.08310992 0.15410959
```

With a p-value of 0.0042, we reject the null hypothesis that the split ticket voting rate is identical for voters who preferred the party nominee, and voters who preferred a different candidate. Therefore, we have statistically significant evidence that there is a higher rate of split ticket voting among voters whose first choice for president was not the eventual nominee.

#Hypothesis C: Biden/Harris and Trump have *more* support among their voters than in primary.

##Harris / Biden

Primary results: Biden 171278/244281 (70.1%); Trump 232848/337014 (68.94%)

I realised that my question was not worded super well, so I will include those who voted for the presidential nominee, as well as leaners / strong partisans.

```
poll %>% filter(B_PRES == "HARRIS" | F_PID == "STRONG DEM" | F_PID == "LEAN DEM") %>%
  mutate(across(c(4:6 | 38), ~replace_na(.x, "NONE"))) %>%
  mutate(Support_Nominee = ifelse(Z_REPRESENT == "BIDEN" | Z_REPRESENT == "HARRIS", 1, 0)) %>%
  summarise(n=n(), Nom_Sup = sum(Support_Nominee), Nom_Sup_Rate = mean(Support_Nominee), Vote_Prop = me
```

```
##      n Nom_Sup Nom_Sup_Rate Vote_Prop Vote_Rate
## 1 368      223    0.6059783 0.9483696      349
```

```
poll %>% filter(B_PRES == "TRUMP" | F_PID == "STRONG REP" | F_PID == "LEAN REP") %>%
  mutate(across(c(4:6 | 38), ~replace_na(.x, "NONE"))) %>%
  mutate(Support_Nominee = ifelse(Z_REPRESENT == "TRUMP", 1, 0)) %>%
  summarise(n=n(), Nom_Sup = sum(Support_Nominee), Nom_Sup_Rate = mean(Support_Nominee), Vote_Prop = me
```

```
##      n Nom_Sup Nom_Sup_Rate Vote_Prop Vote_Total
## 1 321      148    0.4610592 0.8629283      277
```

TESTS: Harris-Biden:

```
t.test(x = c(rep(1, 223), rep(0, 349-223)),
       y = c(rep(1, 171278), rep(0, 244281 - 171218)),
       var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: c(rep(1, 223), rep(0, 349 - 223)) and c(rep(1, 171278), rep(0, 244281 - 171218))
## t = -2.4069, df = 348.9, p-value = 0.01661
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.11268203 -0.01133974
## sample estimates:
## mean of x mean of y
## 0.6389685 0.7009794
```

Poll: 223/349 = 63.9% Primary: 171278/244281 = 70.1% p-value: 0.0166

Trump:

```
t.test(x = c(rep(1, 148), rep(0, 277-148)),
       y = c(rep(1, 232848), rep(0, 337014-232848)),
       var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: c(rep(1, 148), rep(0, 277 - 148)) and c(rep(1, 232848), rep(0, 337014 - 232848))
## t = -5.2143, df = 276.39, p-value = 3.619e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.21574754 -0.09749024
## sample estimates:
## mean of x mean of y
## 0.5342960 0.6909149
```

Poll: $148/277 = 53.4\%$ Primary: $232848/337014 = 69.09\%$ p-value: $3.62e-07$

#Hypothesis C recalculation

We are now going to re-calculate the statistics, regarding just the responses for Preference, and assuming party based on this alone. We are assuming a supporter of Sanders, Harris, Biden, or Phillips is a Democrat, and one of RFK, Trump, Haley, or DeSantis is a Republican. We are not considering anyone who answered “Other”, or failed to answer the Representation question.

Biden/Harris:

```
poll %>% filter(Z_REPRESENT == "HARRIS" | Z_REPRESENT == "BIDEN" | Z_REPRESENT == "SANDERS" | Z_REPRESENT == "PHILLIPS")
```

```
## HarrisTotal HarrisProp n
## 1 224 0.8145455 275
```

```
t.test(x = c(rep(1, 224), rep(0, 275-223)),
       y = c(rep(1, 171278), rep(0, 244281 - 171218)),
       var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: c(rep(1, 224), rep(0, 275 - 223)) and c(rep(1, 171278), rep(0, 244281 - 171218))
## t = 4.6874, df = 275.85, p-value = 4.354e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.06415874 0.15707093
## sample estimates:
## mean of x mean of y
## 0.8115942 0.7009794
```

Poll: $224/275 = 81.5\%$ Primary: $171278/244281 = 70.1\%$ p-value: $4.354e-06$

Trump:

```
poll %>% filter(Z_REPRESENT == "TRUMP" | Z_REPRESENT == "DESANTIS" | Z_REPRESENT == "HALEY" | Z_REPRESENT == "RFK")
```

```
## TrumpTotal TrumpProp n
## 1 149 0.6182573 241
```

```
t.test(x = c(rep(1, 149), rep(0, 241-148)),
       y = c(rep(1, 232848), rep(0, 337014-232848)),
       var.equal = FALSE)
```

```
var.equal = FALSE)
```

```
##  
## Welch Two Sample t-test  
##  
## data: c(rep(1, 149), rep(0, 241 - 148)) and c(rep(1, 232848), rep(0, 337014 - 232848))  
## t = -2.3996, df = 241.31, p-value = 0.01717  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.13695470 -0.01347018  
## sample estimates:  
## mean of x mean of y  
## 0.6157025 0.6909149
```

Poll: $149/241 = 61.8\%$ Primary: $232848/337014 = 69.09\%$ p-value: 0.0172

First calculation of Hypothesis C showed that both candidates had significantly lower rates of support on Election Day than in the primary; the second calculation method showed marked improvement for both candidates; Harris performs significantly better on Election Day than Biden on Super Tuesday, while Trump remains significantly worse [this is to be expected, as Trump had a much more competitive primary than Biden did]. Therefore, our hypothesis is rejected.