Final Projects; Winter Institute in Data Science

Adam Weber, Ali Amini, John Burzawa

2023-01-11

Load packages and import two data set

```
library(devtools)
```

```
## Warning: package 'devtools' was built under R version 4.2.2
```

```
## Loading required package: usethis
```

```
## Warning: package 'usethis' was built under R version 4.2.2
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.2.2
```

```
##
## 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
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.2.2
```

```
## here() starts at C:/Users/aa6718a/Desktop/git/American-Public-Opinion-Project-based-on-ANES-
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.2.2
```

```
## -- Attaching packages ------------------------------------ tidyverse 1.3.2 --
```

```
## v ggplot2 3.4.0      v purrr   1.0.0
## v tibble  3.1.8      v stringr 1.5.0
## v tidyr   1.2.1      v forcats 0.5.2
## v readr   2.1.3
```

```
## Warning: package 'ggplot2' was built under R version 4.2.2
```

```
## Warning: package 'tidyr' was built under R version 4.2.2
```

```
## Warning: package 'readr' was built under R version 4.2.2
```

```
## Warning: package 'purrr' was built under R version 4.2.2
```

```
## Warning: package 'stringr' was built under R version 4.2.2
```

```
## Warning: package 'forcats' was built under R version 4.2.2
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(haven)
library(ggplot2)
```

```
anes_2020 <- read.csv(here("anes_2020.csv"))
anes_2016 <- read_dta("~/anes_timeseries_2016.dta")
```

cleaning data

In the following steps we will clean data and create approprite varibles to show diffrences in mean and standard deviation(sd) for each, we use sd as apporximate for polrization; we do sampe porcess for free trade, health care, persident job, and wall at Mexico, we will include all codes for the free trdae and we will do the same process for others and only the output and t test will be shown.

#FAVOR OR OPPOSE FREE TRADE AGREEMENTS WITH OTHER COUNTRIES

```
anes_2020 <- anes_2020 %>%
  mutate (opinion_free_trade = case_when( free_trade == 3 ~ 0,
                           free_trade == 1 & free_trade_intensity == 1 ~ 3,
                          free_trade == 1 &  free_trade_intensity == 2 ~ 2,
                           free_trade == 1 &  free_trade_intensity == 3 ~ 1,
                            free_trade == 2 &  free_trade_intensity == 1 ~ -3,
                           free_trade == 2 &  free_trade_intensity == 2 ~ -2,
                          free_trade == 2 &  free_trade_intensity == 3 ~ -1))

anes_2016 <- anes_2016 %>%
  mutate(opinion_free_trade = case_when( free_trade == 3 ~ 0,
                          free_trade == 1 & free_trade_intensity == 1 ~ 3,
```

```
                        free_trade == 1 &  free_trade_intensity == 2 ~ 2,
                        free_trade == 1 &  free_trade_intensity == 3 ~ 1,
                        free_trade == 2 &  free_trade_intensity == 1 ~ -3,
                        free_trade == 2 &  free_trade_intensity == 2 ~ -2,
                        free_trade == 2 &  free_trade_intensity == 3 ~ -1))


#calculation of Standard Deviation and mean for each party in free  trade----
anes_2020 %>% group_by(pid_chr) %>%
  summarise(sd_free_trade = sd(opinion_free_trade, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   pid_chr sd_free_trade
##   <chr>           <dbl>
## 1 dem              1.42
## 2 ind              1.43
## 3 rep              1.57
## 4 <NA>             1.45
```

```
anes_2020 %>% group_by(pid_chr) %>%
  summarise(mean_free_trade = mean(opinion_free_trade, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   pid_chr mean_free_trade
##   <chr>             <dbl>
## 1 dem               1.00
## 2 ind               0.837
## 3 rep               0.713
## 4 <NA>              0.767
```

```
anes_2016 %>% group_by(pid_chr) %>%
  summarise(sd_free_trade = sd(opinion_free_trade, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   pid_chr sd_free_trade
##   <chr>           <dbl>
## 1 dem              1.66
## 2 ind              1.57
## 3 rep              1.79
## 4 <NA>             1.62
```

```
anes_2016 %>% group_by(pid_chr) %>%
  summarise(mean_free_trade = mean(opinion_free_trade, na.rm = TRUE))
```

```
## # A tibble: 4 x 2
##   pid_chr mean_free_trade
##   <chr>             <dbl>
## 1 dem               0.631
## 2 ind               0.570
## 3 rep               0.213
## 4 <NA>              0.296
```

```r
# t-test----
t.test (anes_2016$opinion_free_trade, anes_2020$opinion_free_trade)
```

```
##
##  Welch Two Sample t-test
##
## data:  anes_2016$opinion_free_trade and anes_2020$opinion_free_trade
## t = -13.367, df = 6393.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4988572 -0.3712543
## sample estimates:
## mean of x mean of y
## 0.3859210 0.8209767
```
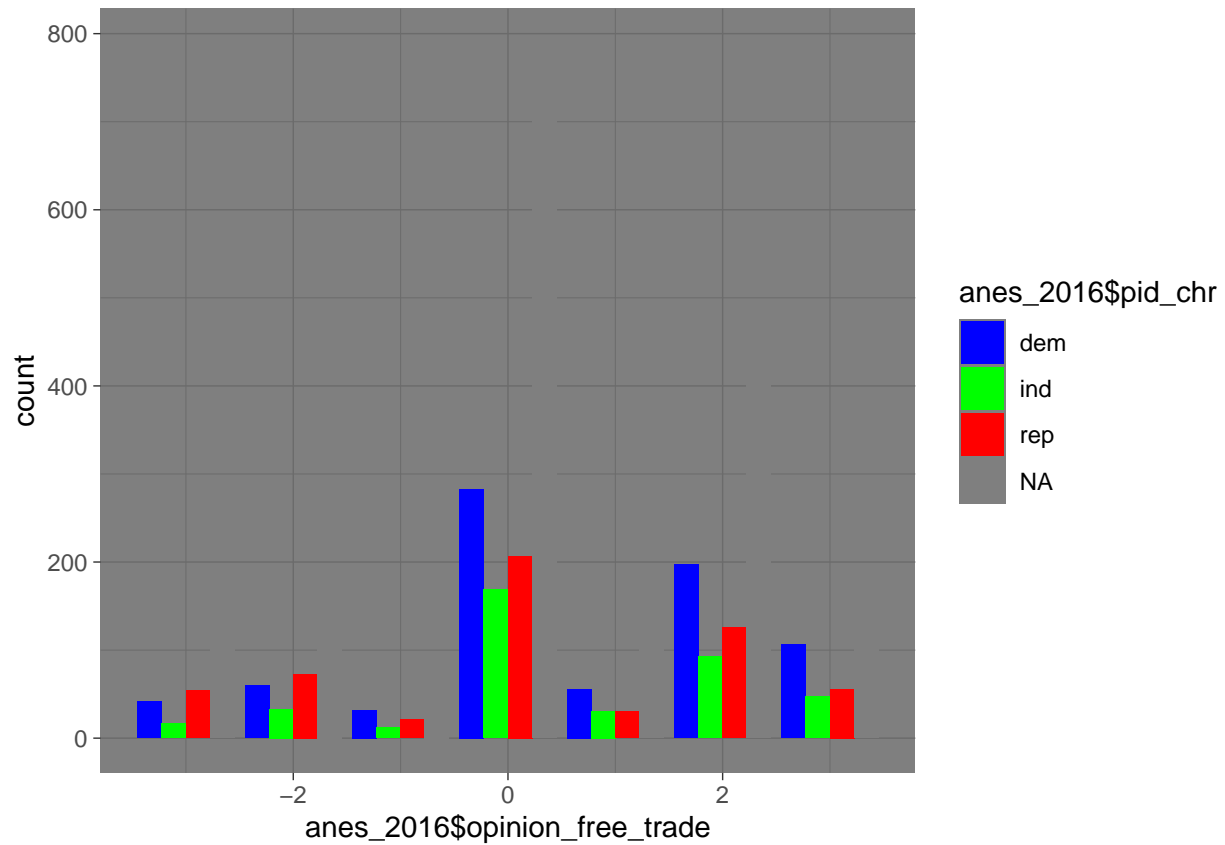
```r
#visualization----
ggplot(data = anes_2016,
       mapping = aes(x=anes_2016$opinion_free_trade, fill = anes_2016$pid_chr)) +
  geom_bar(position = "dodge") + scale_fill_manual(values=c('blue', 'green', 'red', 'white')) +theme_da
```

```
## Warning: Use of `anes_2016$opinion_free_trade` is discouraged.
## i Use `opinion_free_trade` instead.
```
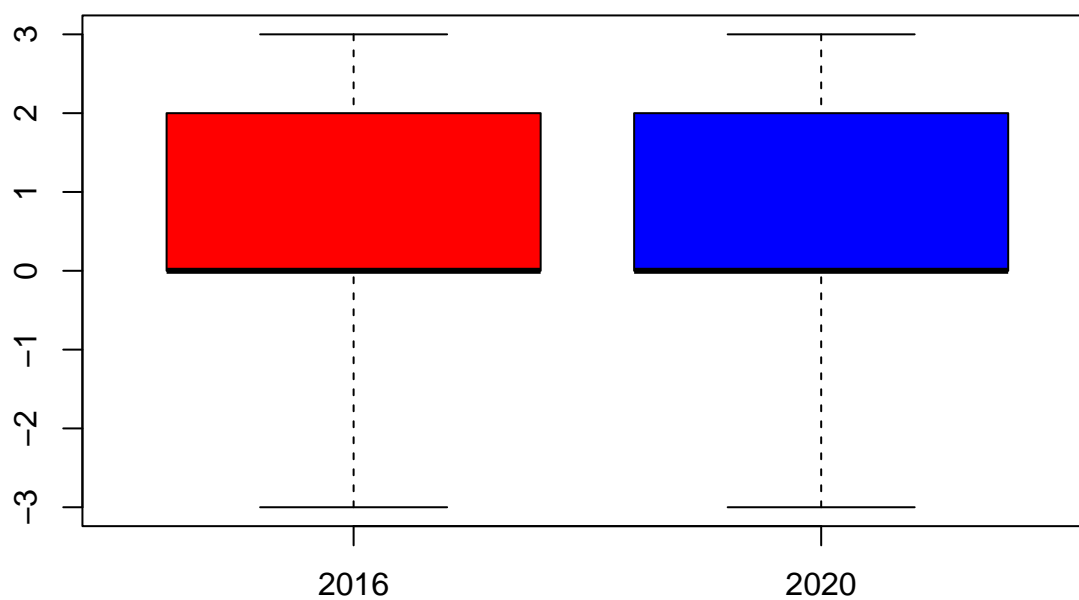
```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```

```
## Warning: Removed 676 rows containing non-finite values (`stat_count()`).
```

```
ggplot(data = anes_2020,
       mapping = aes(x=anes_2020$opinion_free_trade, fill = anes_2020$pid_chr)) +
  geom_bar(position = "dodge") + scale_fill_manual(values=c('blue', 'green', 'red', 'white')) +theme_da
```

```
## Warning: Use of `anes_2020$opinion_free_trade` is discouraged.
## i Use `opinion_free_trade` instead.
```

```
## Warning: Use of `anes_2020$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```

```
## Warning: Removed 929 rows containing non-finite values (`stat_count()`).
```
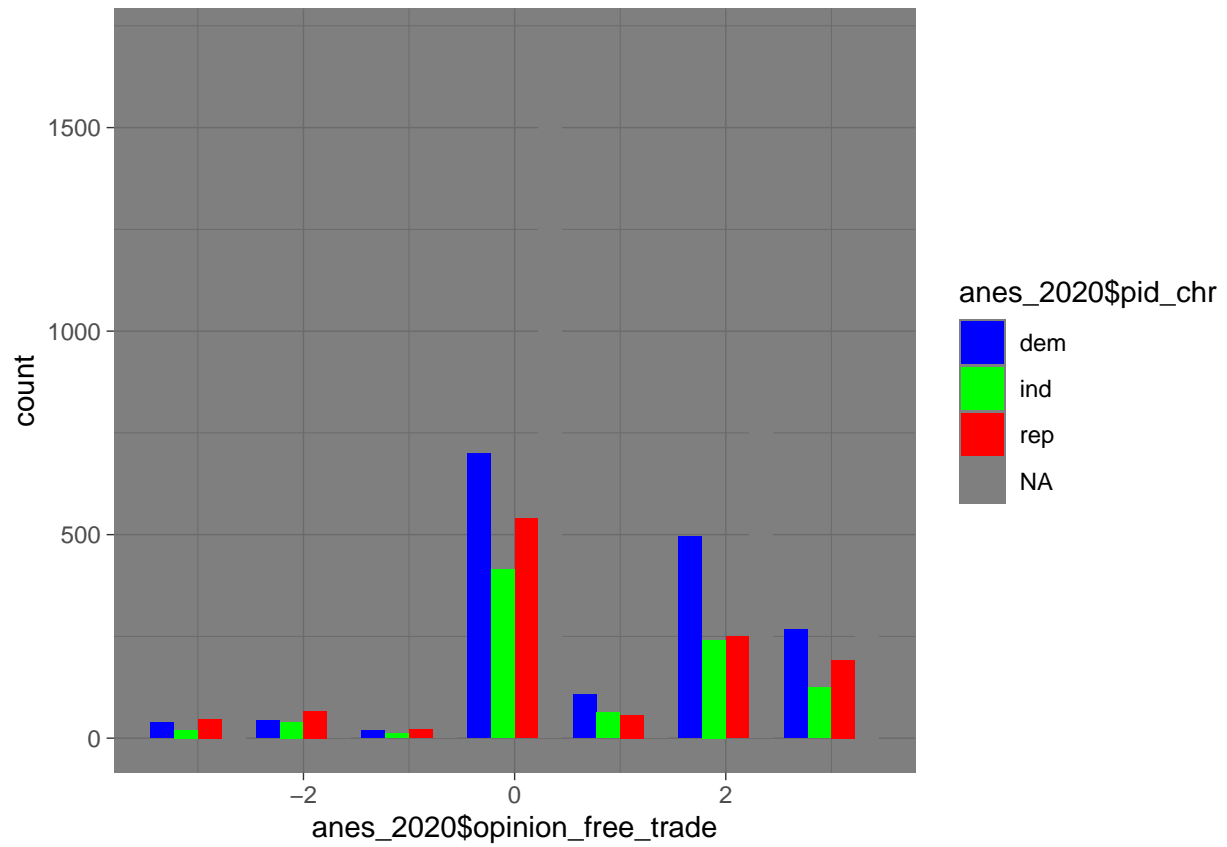
```
boxplot(anes_2016$opinion_free_trade, anes_2020$opinion_free_trade,
        names = c('2016', '2020'), col=c('red', 'blue'), main='Public oppinion in free trade')
```

Public oppinion in free trade


```
ggplot(anes_2016, aes(anes_2016$pid_chr,anes_2016$opinion_free_trade)) +
  geom_boxplot(col=c('blue','green', 'red', 'white'))
```

```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```

```
## Warning: Use of `anes_2016$opinion_free_trade` is discouraged.
## i Use `opinion_free_trade` instead.
```

```
## Warning: Removed 676 rows containing non-finite values (`stat_boxplot()`).
```
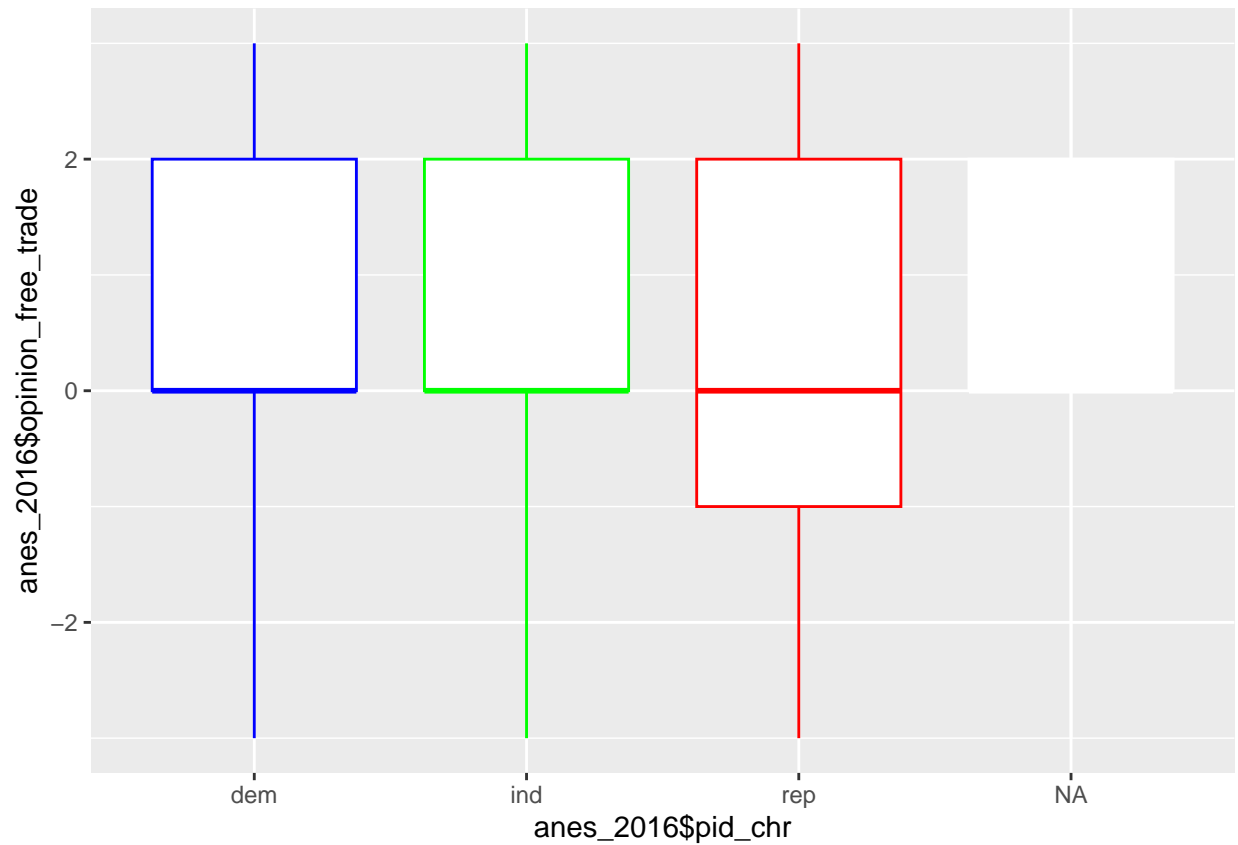
```
ggplot(anes_2020, aes(anes_2020$pid_chr,anes_2020$opinion_free_trade)) +
  geom_boxplot(col=c('blue','green', 'red', 'white'))
```

```
## Warning: Use of `anes_2020$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```

```
## Warning: Use of `anes_2020$opinion_free_trade` is discouraged.
## i Use `opinion_free_trade` instead.
```

```
## Warning: Removed 929 rows containing non-finite values (`stat_boxplot()`).
```
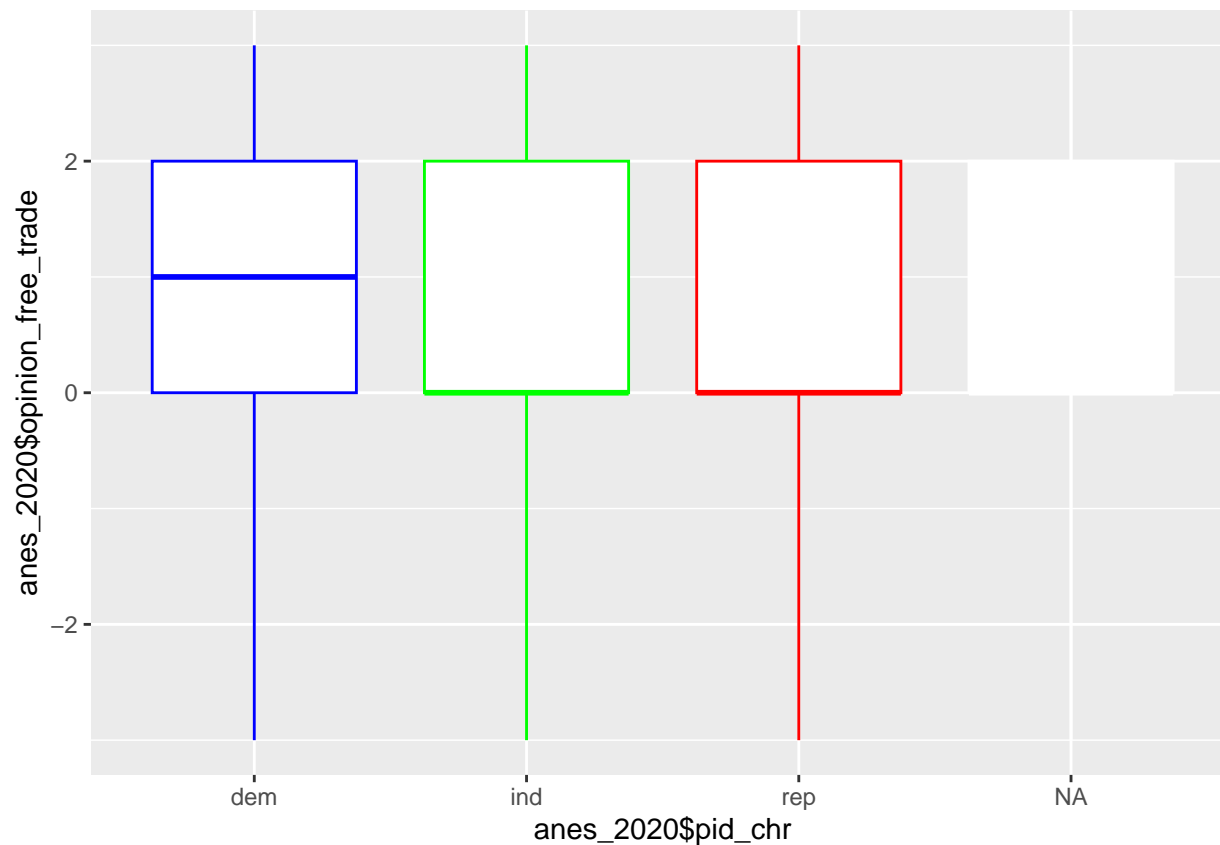
#INCREASE/DECREASE GOVERNMENT SPENDING TO HELP PAY FOR HEALTH CARE

```
## # A tibble: 4 x 2
##   pid_chr sd_health_care
##   <chr>            <dbl>
## 1 dem               1.40
## 2 ind               1.58
## 3 rep               1.52
## 4 <NA>              1.68
```

```
## # A tibble: 4 x 2
##   pid_chr mean_health_care
##   <chr>              <dbl>
## 1 dem                1.68
## 2 ind                0.986
## 3 rep                0.112
## 4 <NA>               0.865
```
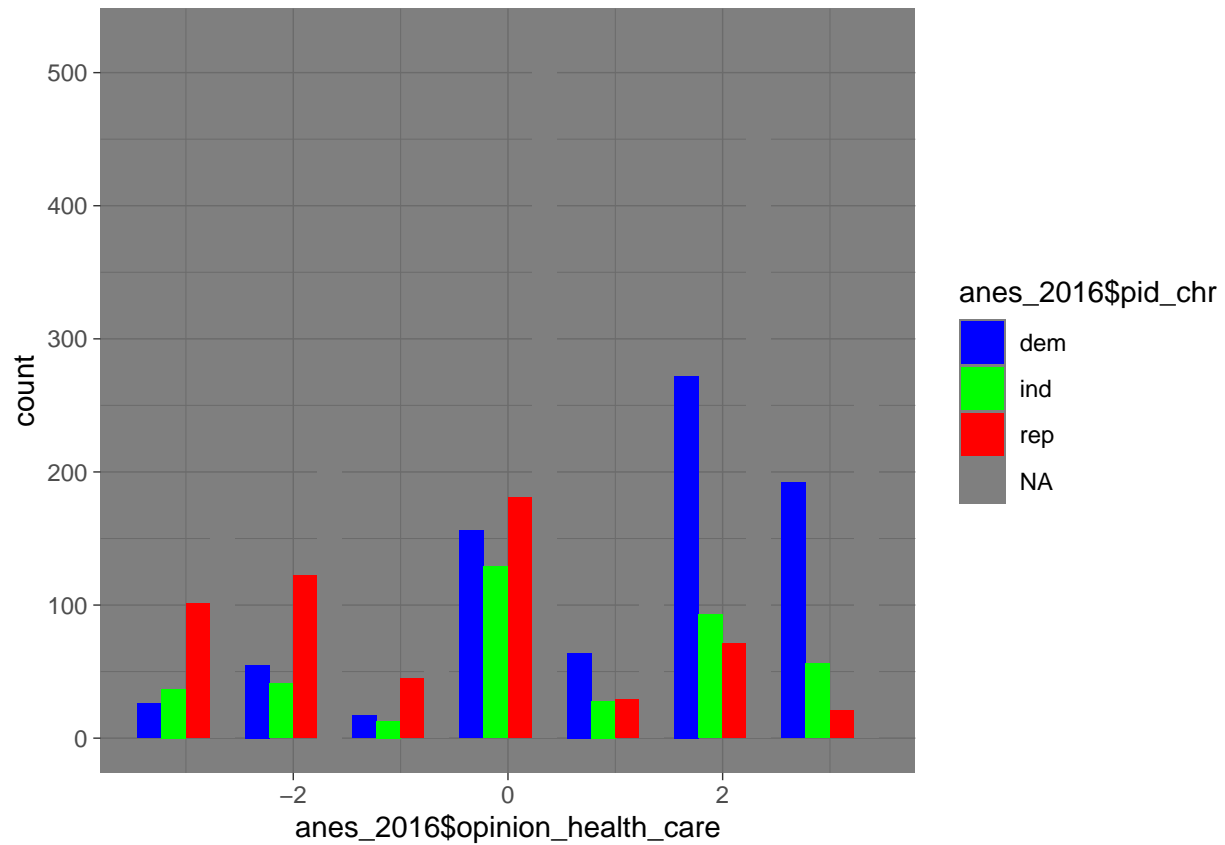
```
## # A tibble: 4 x 2
##   pid_chr sd_health_care
##   <chr>            <dbl>
## 1 dem               1.65
## 2 ind               1.84
## 3 rep               1.74
## 4 <NA>              1.90
```

```
## # A tibble: 4 x 2
##   pid_chr mean_health_care
##   <chr>              <dbl>
## 1 dem                 1.25
## 2 ind                 0.443
## 3 rep                -0.628
## 4 <NA>                0.460


##
##  Welch Two Sample t-test
##
## data:  anes_2016$opinion_health_care and anes_2020$opinion_health_care
## t = -13.085, df = 6340.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5589705 -0.4133119
## sample estimates:
## mean of x mean of y
## 0.4581363 0.9442775


## Warning: Use of 'anes_2016$opinion_health_care' is discouraged.
## i Use 'opinion_health_care' instead.


## Warning: Use of 'anes_2016$pid_chr' is discouraged.
## i Use 'pid_chr' instead.


## Warning: Removed 675 rows containing non-finite values ('stat_count()').
```
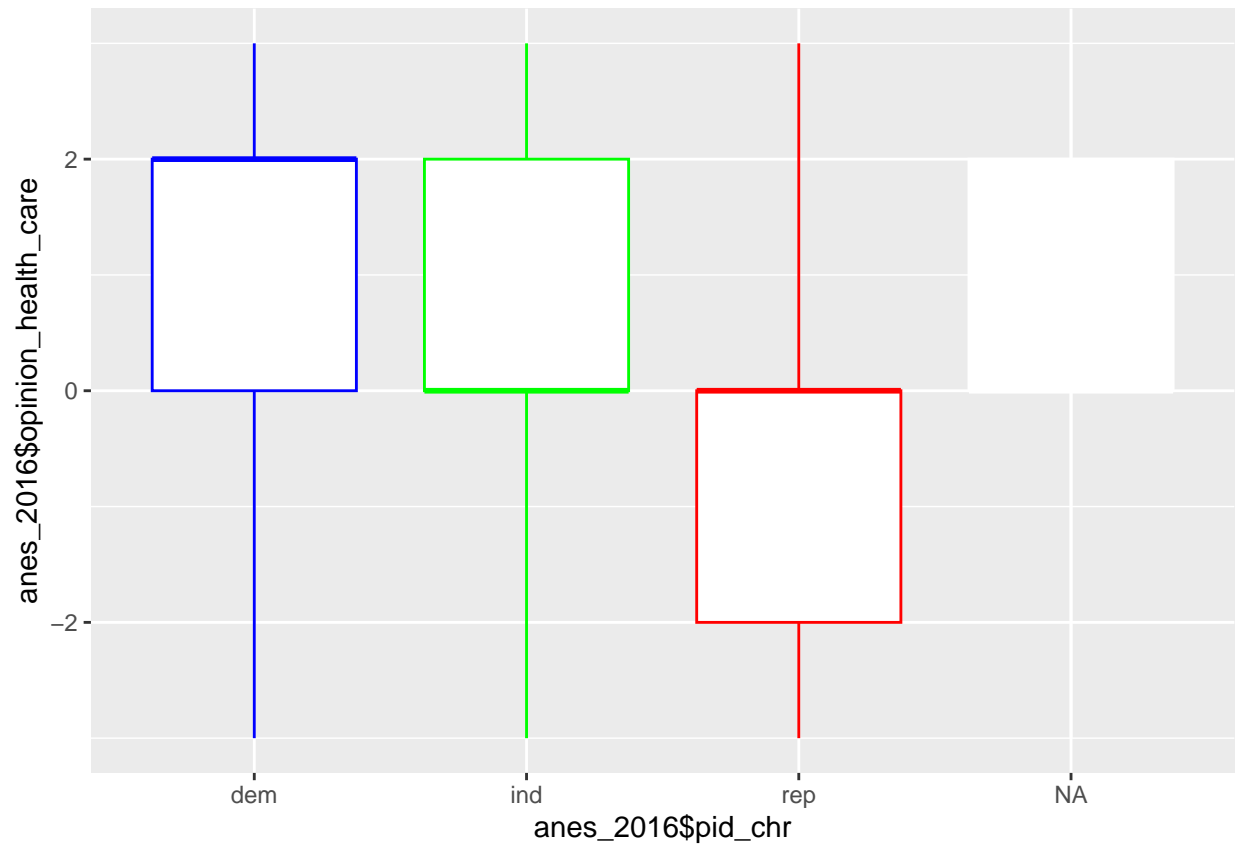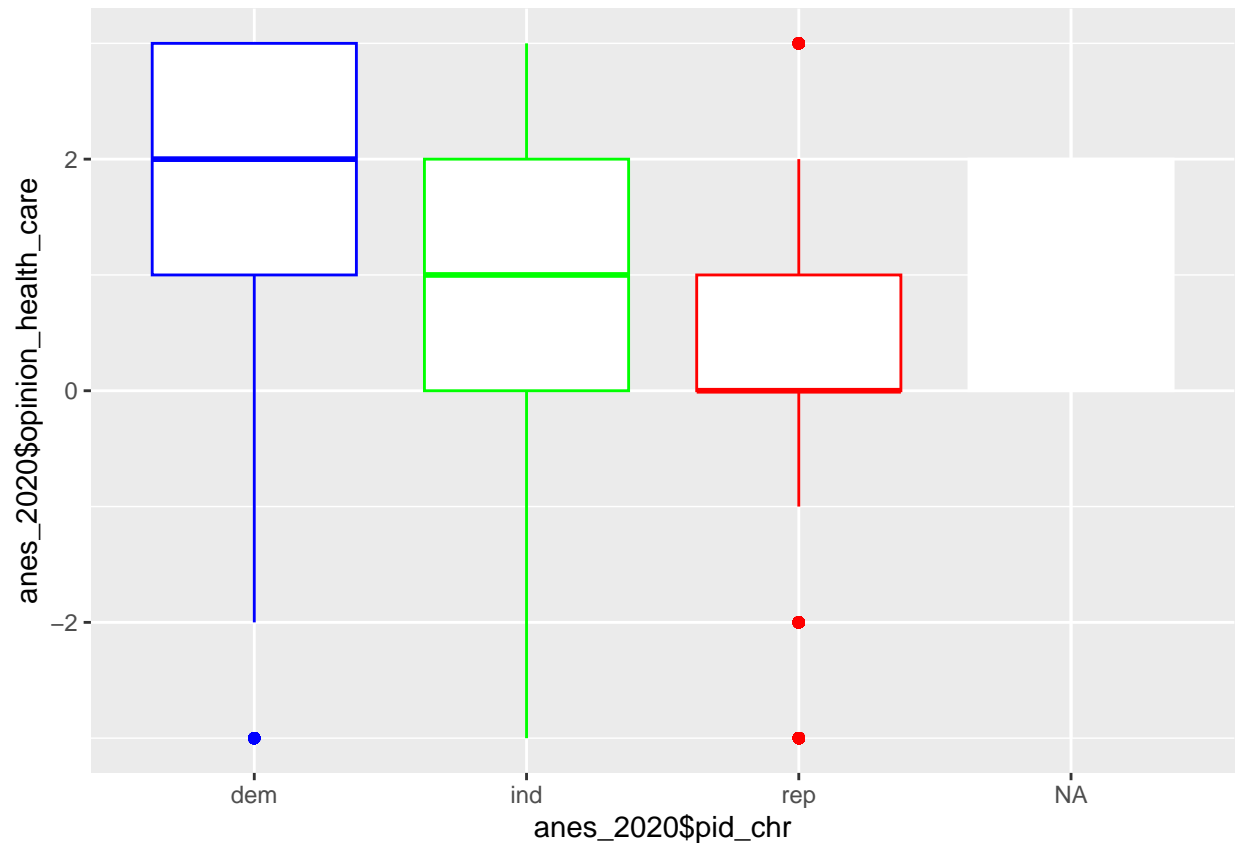
```
## Warning: Use of 'anes_2020$opinion_health_care' is discouraged.
## i Use 'opinion_health_care' instead.

## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.

## Warning: Removed 958 rows containing non-finite values ('stat_count()').
```

Public oppinion in health care


```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.

## Warning: Use of `anes_2016$opinion_health_care` is discouraged.
## i Use `opinion_health_care` instead.

## Warning: Removed 675 rows containing non-finite values (`stat_boxplot()`).
```

```
## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.
```

```
## Warning: Use of 'anes_2020$opinion_health_care' is discouraged.
## i Use 'opinion_health_care' instead.
```

```
## Warning: Removed 958 rows containing non-finite values ('stat_boxplot()').
```

#FAVOR OR OPPOSE BUILDING A WALL ON BORDER WITH MEXICO

```
## # A tibble: 4 x 2
##   pid_chr sd_wall_Mexico
##   <chr>            <dbl>
## 1 dem               1.85
## 2 ind               2.30
## 3 rep               1.90
## 4 <NA>              2.44


## # A tibble: 4 x 2
##   pid_chr mean_wall_Mexico
##   <chr>              <dbl>
## 1 dem                -1.83
## 2 ind                -0.518
## 3 rep                 1.73
## 4 <NA>               -0.200


## # A tibble: 4 x 2
##   pid_chr sd_wall_Mexico
##   <chr>            <dbl>
## 1 dem               2.08
## 2 ind               2.35
## 3 rep               2.14
## 4 <NA>              2.32
```

```
## # A tibble: 4 x 2
##   pid_chr mean_wall_Mexico
##   <chr>              <dbl>
## 1 dem               -1.53
## 2 ind               -0.645
## 3 rep                1.02
## 4 <NA>              -0.350


##
##   Welch Two Sample t-test
##
## data:  anes_2016$opinion_wall_Mexico and anes_2020$opinion_wall_Mexico
## t = -2.704, df = 8890.4, p-value = 0.006863
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.21223528 -0.03384511
## sample estimates:
##  mean of x  mean of y
## -0.4185499 -0.2955097


## Warning: Use of 'anes_2016$opinion_wall_Mexico' is discouraged.
## i Use 'opinion_wall_Mexico' instead.


## Warning: Use of 'anes_2016$pid_chr' is discouraged.
## i Use 'pid_chr' instead.


## Warning: Removed 22 rows containing non-finite values ('stat_count()').
```
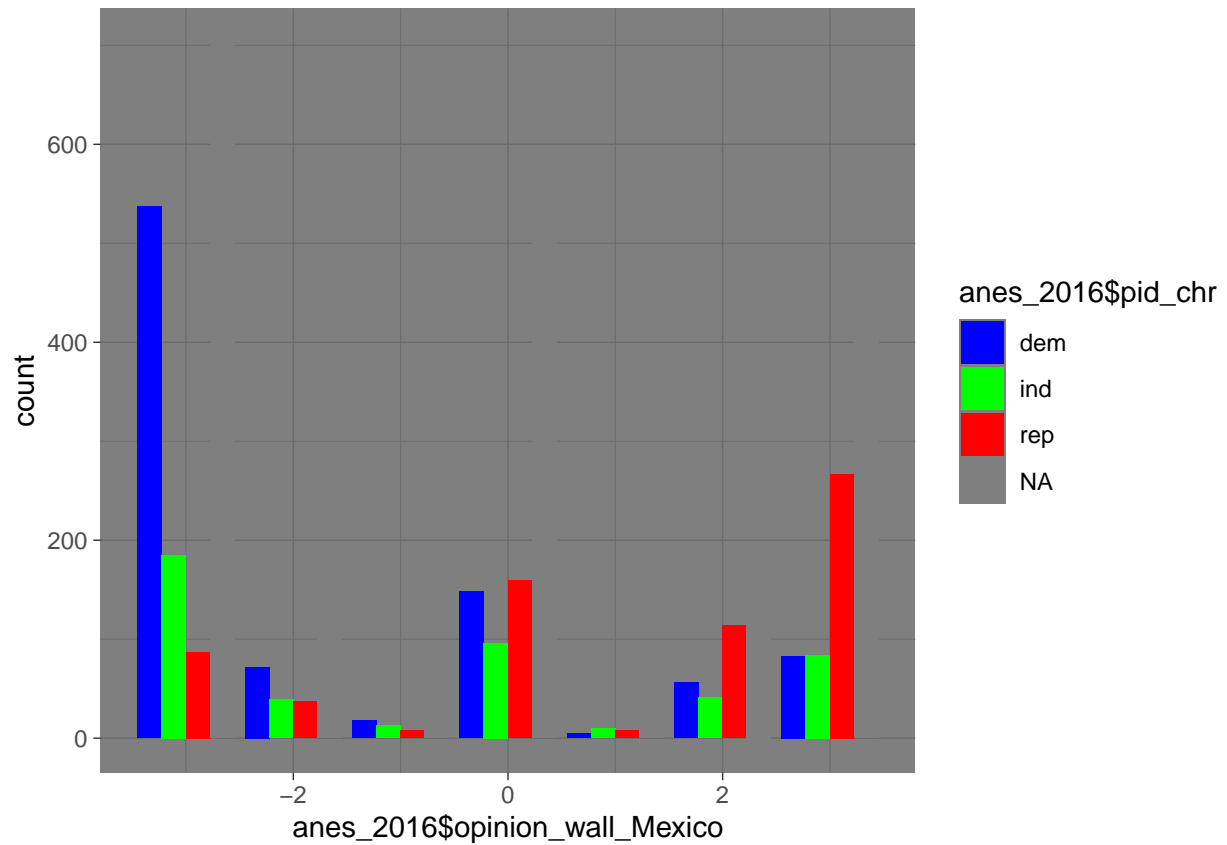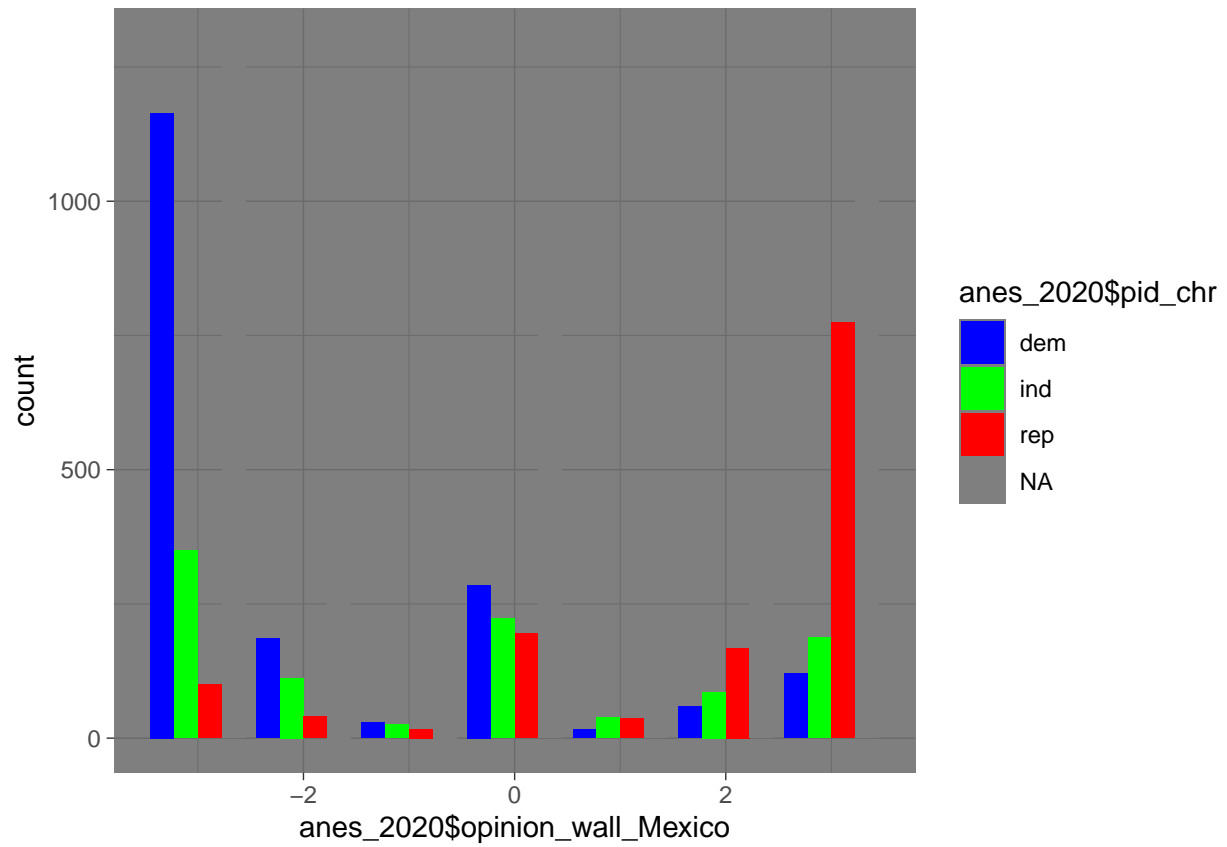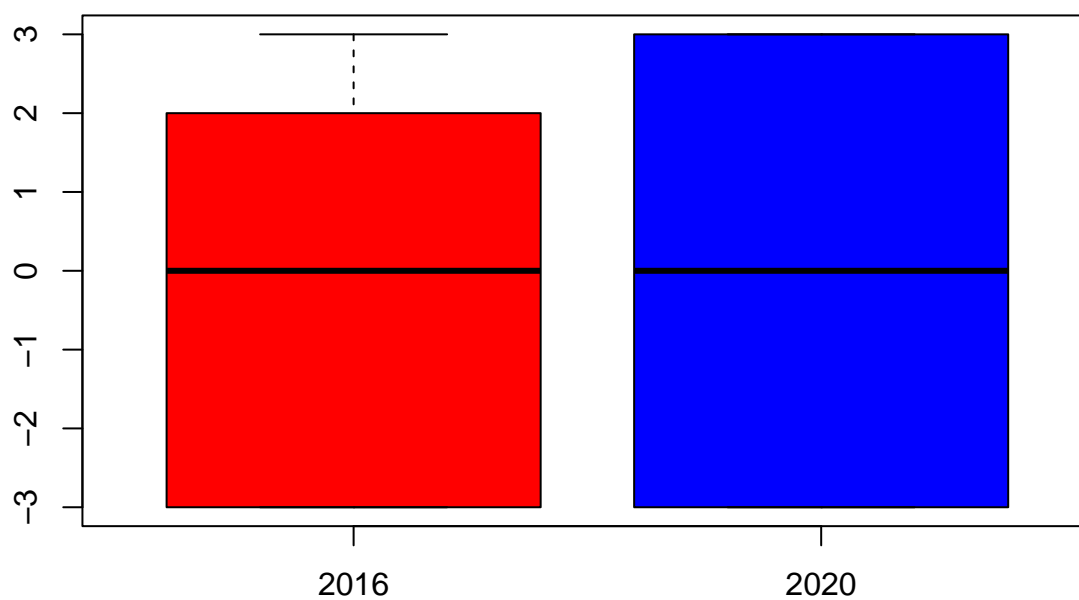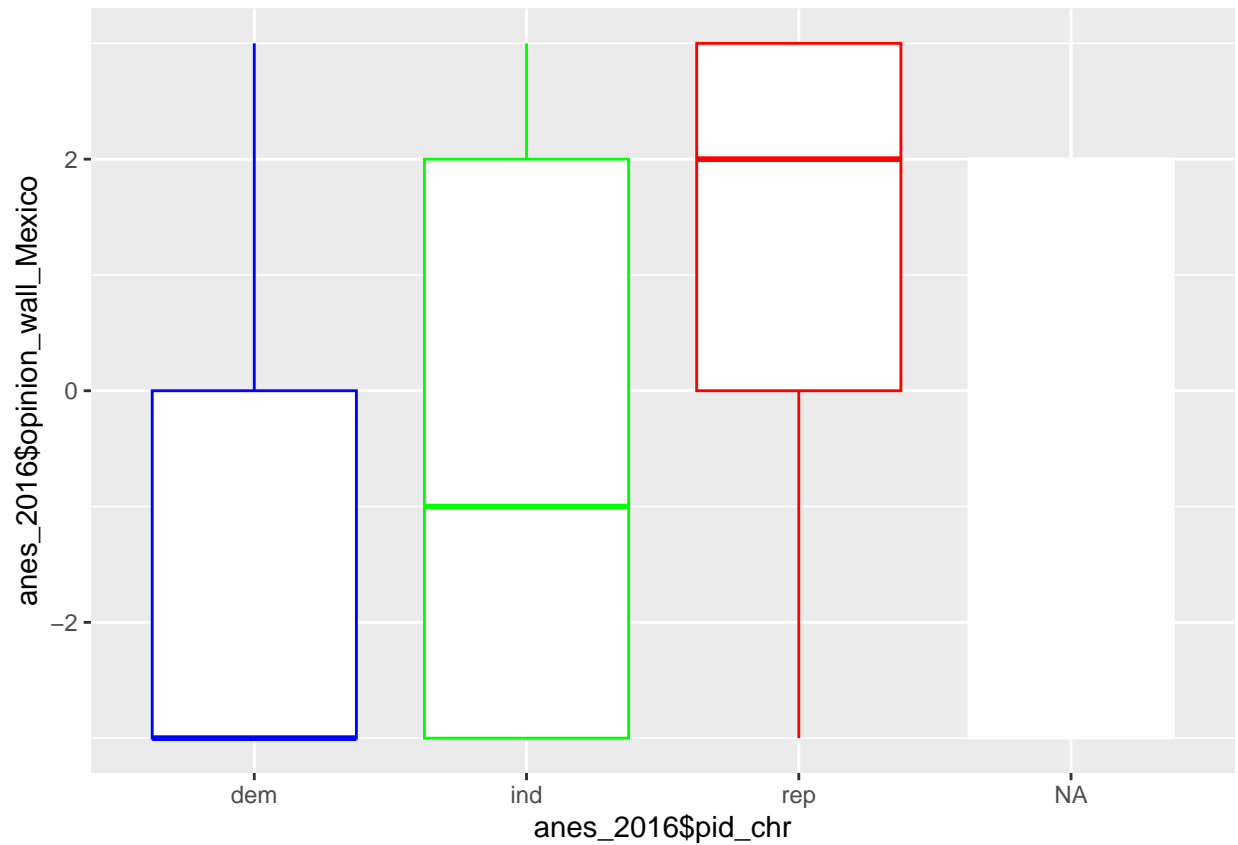
```
## Warning: Use of `anes_2020$opinion_wall_Mexico` is discouraged.
## i Use `opinion_wall_Mexico` instead.

## Warning: Use of `anes_2020$pid_chr` is discouraged.
## i Use `pid_chr` instead.

## Warning: Removed 40 rows containing non-finite values (`stat_count()`).
```

Public oppinion in wall on Mexico


```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.

## Warning: Use of `anes_2016$opinion_wall_Mexico` is discouraged.
## i Use `opinion_wall_Mexico` instead.

## Warning: Removed 22 rows containing non-finite values (`stat_boxplot()`).
```
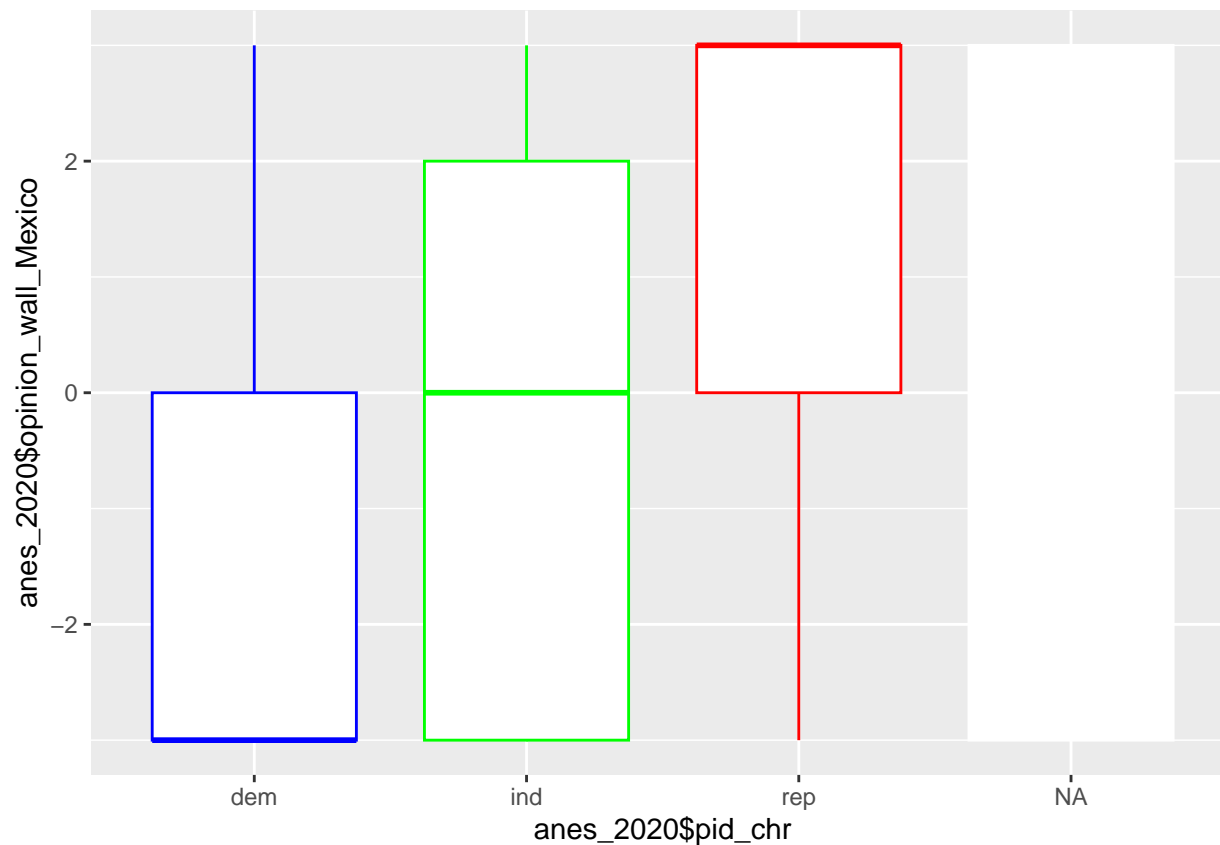
```
## Warning: Use of `anes_2020$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```

```
## Warning: Use of `anes_2020$opinion_wall_Mexico` is discouraged.
## i Use `opinion_wall_Mexico` instead.
```

```
## Warning: Removed 40 rows containing non-finite values (`stat_boxplot()`).
```

#APPROVE OR DISAPPROVE PRESIDENT HANDLING JOB

```
## # A tibble: 4 x 2
##   pid_chr sd_opinion_persident_job
##   <chr>                     <dbl>
## 1 dem                        1.07
## 2 ind                        1.73
## 3 rep                        1.45
## 4 <NA>                       1.84
```

```
## # A tibble: 4 x 2
##   pid_chr mean_opinion_persident_job
##   <chr>                        <dbl>
## 1 dem                         -1.63
## 2 ind                         -0.618
## 3 rep                          1.18
## 4 <NA>                        -0.290
```

```
## # A tibble: 4 x 2
##   pid_chr sd_opinion_persident_job
##   <chr>                     <dbl>
## 1 dem                        1.42
## 2 ind                        1.74
## 3 rep                        1.26
## 4 <NA>                       1.76
```

```
## # A tibble: 4 x 2
##   pid_chr mean_opinion_persident_job
##   <chr>                        <dbl>
## 1 dem                         1.14
## 2 ind                         0.0409
## 3 rep                        -1.35
## 4 <NA>                       -0.0718


##
##  Welch Two Sample t-test
##
## data:  anes_2016$opinion_persident_job and anes_2020$opinion_persident_job
## t = 11.482, df = 8712.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.3256703 0.4597587
## sample estimates:
##    mean of x    mean of y
## -0.003321471 -0.396035992


## Warning: Use of 'anes_2016$opinion_persident_job' is discouraged.
## i Use 'opinion_persident_job' instead.


## Warning: Use of 'anes_2016$pid_chr' is discouraged.
## i Use 'pid_chr' instead.


## Warning: Removed 55 rows containing non-finite values ('stat_count()').
```
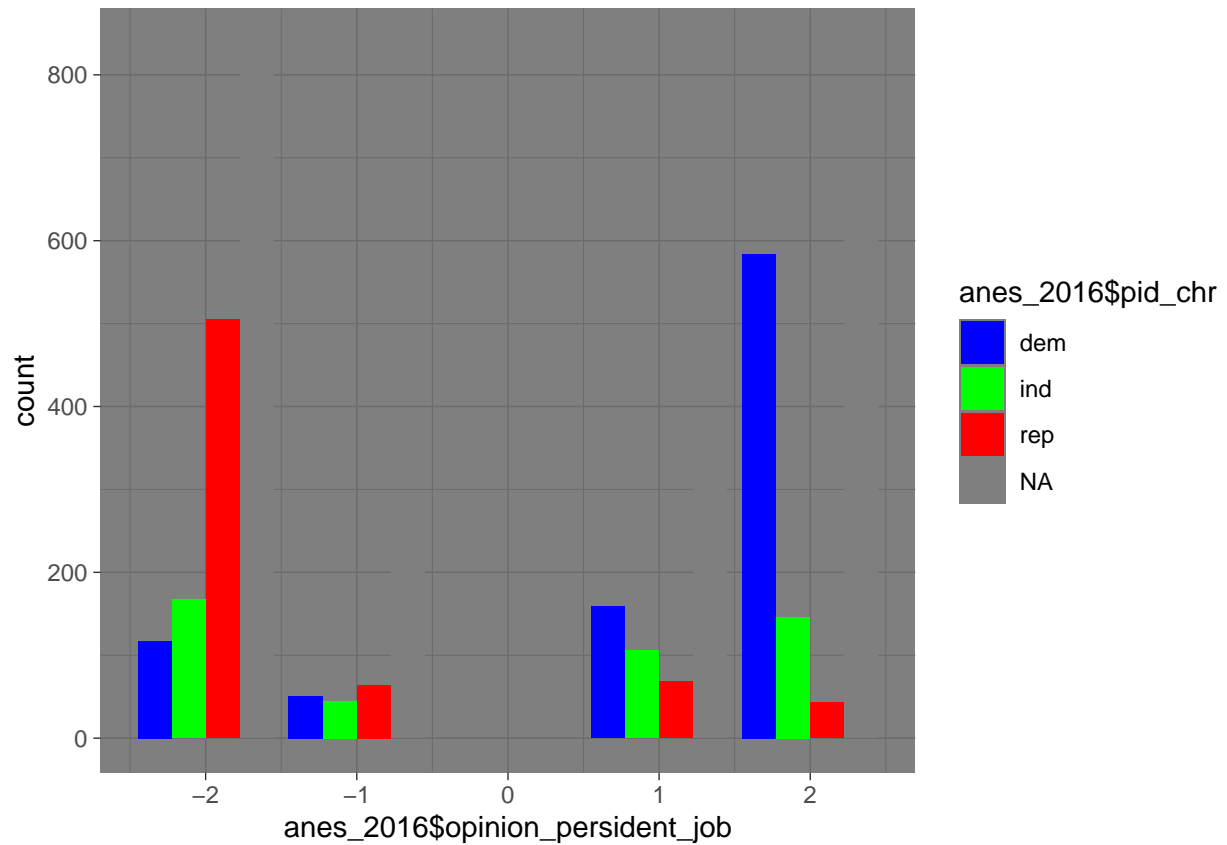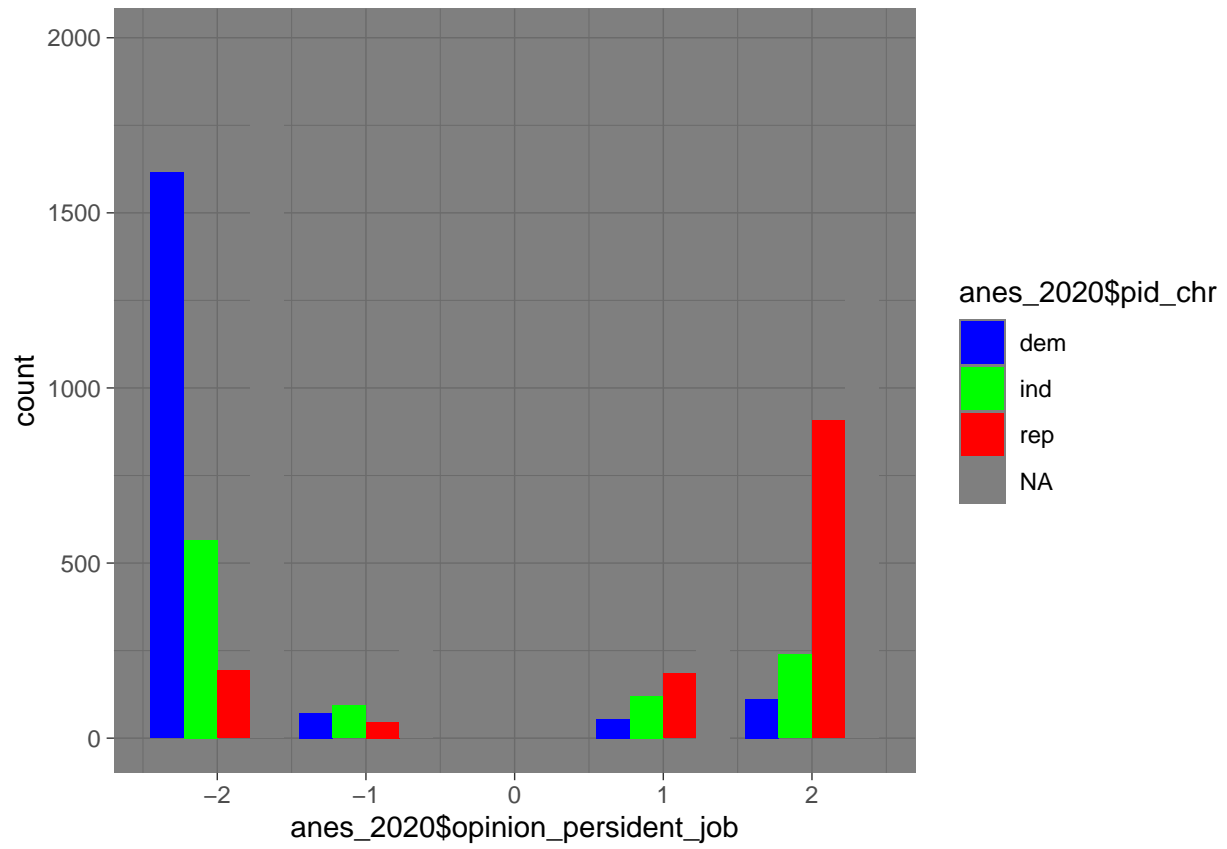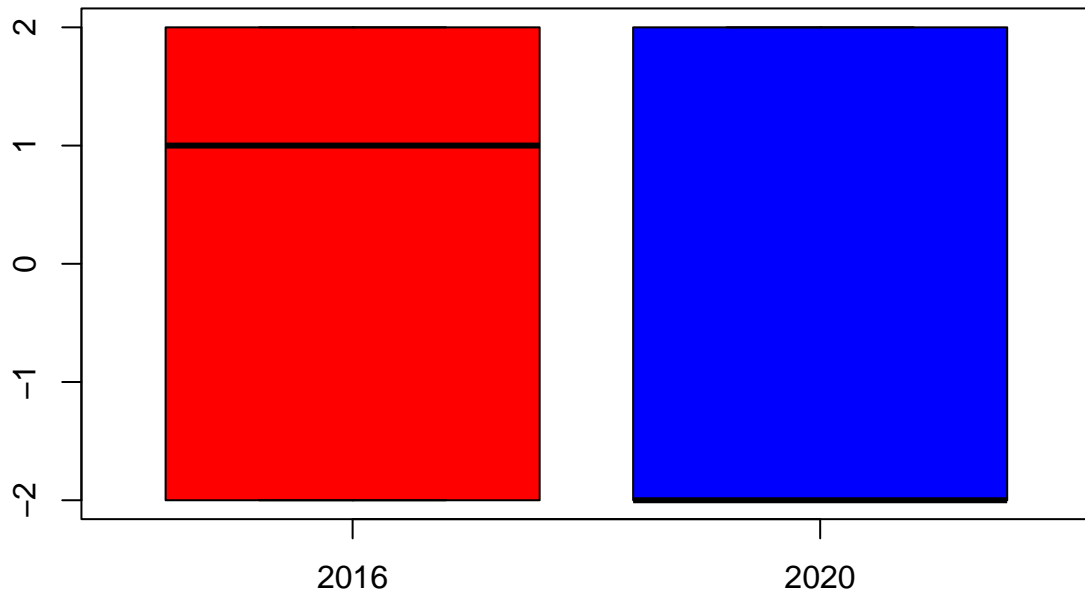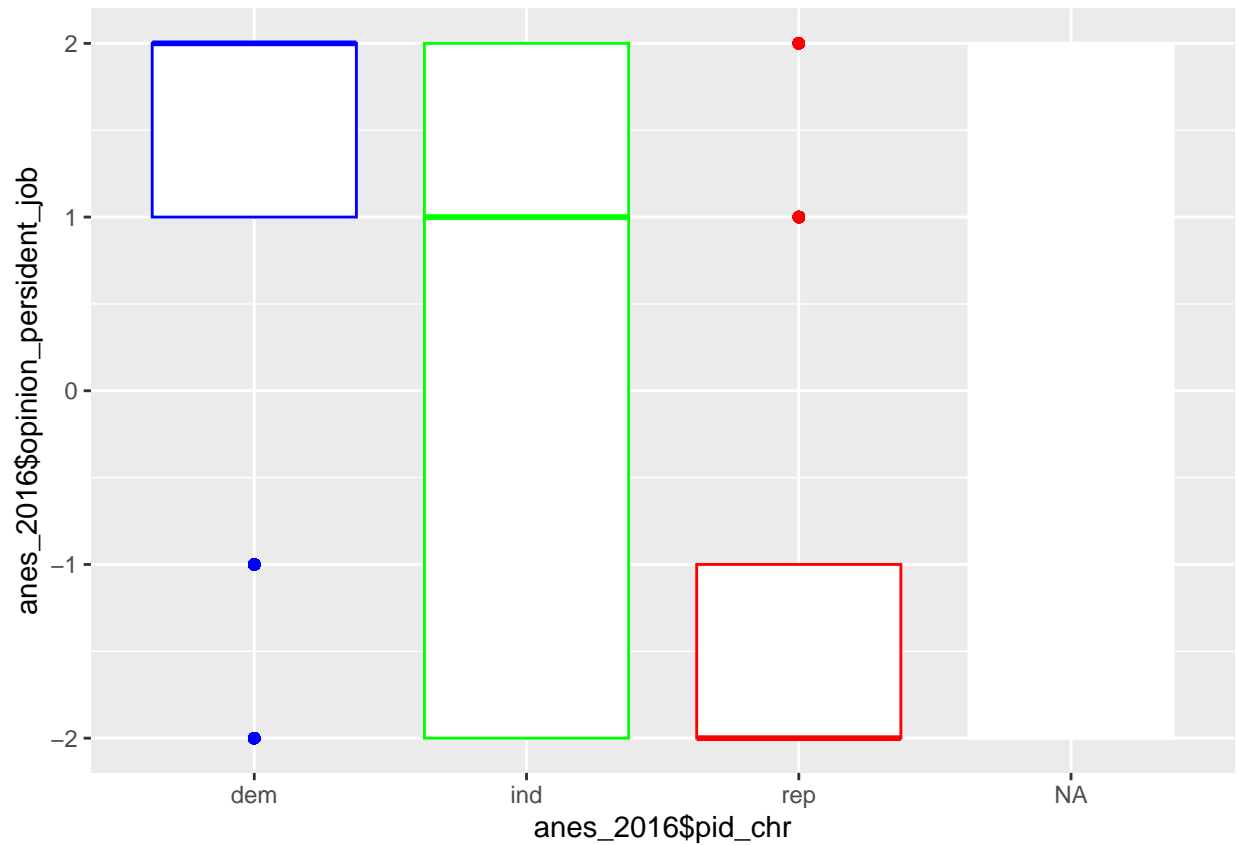
```
## Warning: Use of 'anes_2020$opinion_persident_job' is discouraged.
## i Use 'opinion_persident_job' instead.

## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.

## Warning: Removed 56 rows containing non-finite values ('stat_count()').
```

Public oppinion in persident job


```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.

## Warning: Use of `anes_2016$opinion_persident_job` is discouraged.
## i Use `opinion_persident_job` instead.

## Warning: Removed 55 rows containing non-finite values (`stat_boxplot()`).
```
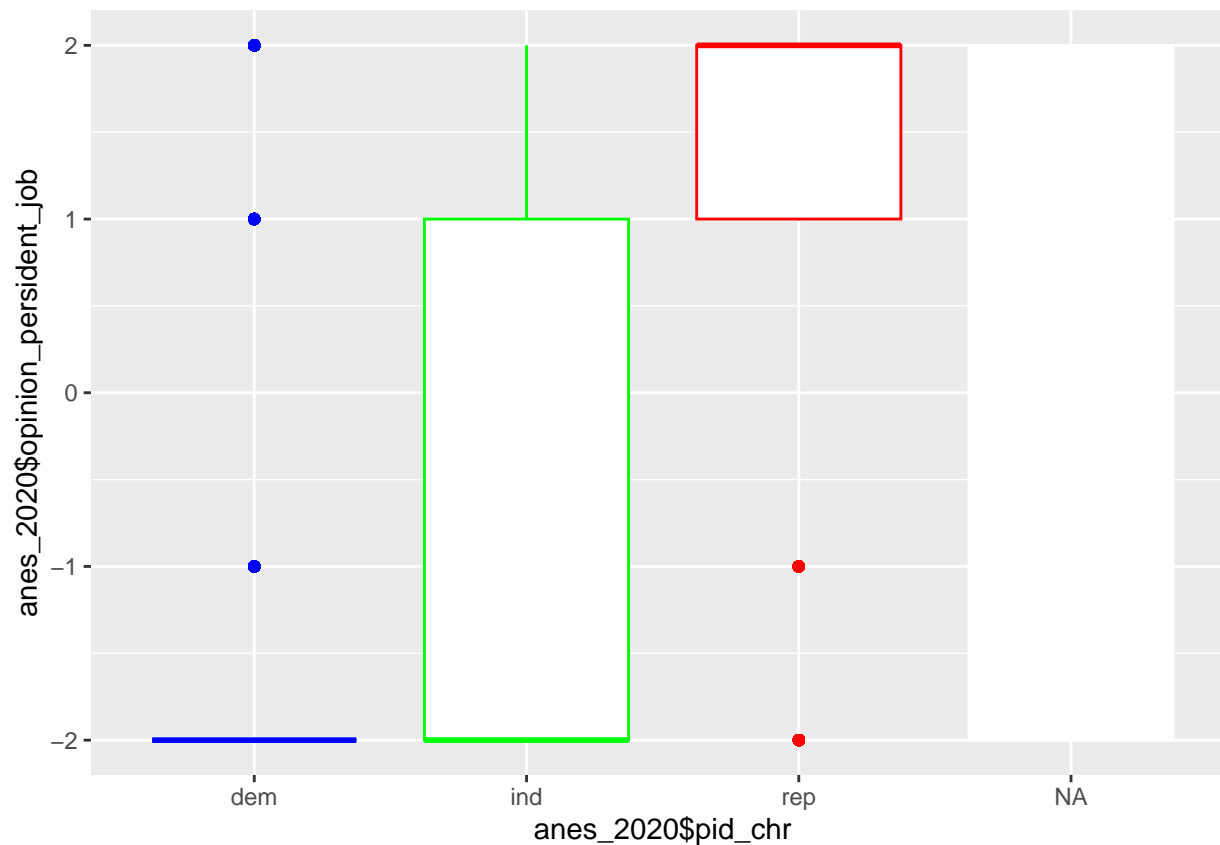
```
## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.

## Warning: Use of 'anes_2020$opinion_persident_job' is discouraged.
## i Use 'opinion_persident_job' instead.

## Warning: Removed 56 rows containing non-finite values ('stat_boxplot()').
```

#GOVERNMENT ACTION ABOUT RISING TEMPERATURES

```
## # A tibble: 4 x 2
##   pid_chr sd_rising_temperatures
##   <chr>                    <dbl>
## 1 dem                       1.28
## 2 ind                       1.64
## 3 rep                       1.68
## 4 <NA>                      1.70


## # A tibble: 4 x 2
##   pid_chr mean_rising_temperatures
##   <chr>                      <dbl>
## 1 dem                        2.26
## 2 ind                        1.47
## 3 rep                       0.182
## 4 <NA>                       1.25


## # A tibble: 4 x 2
##   pid_chr sd_rising_temperatures
##   <chr>                    <dbl>
## 1 dem                       1.54
## 2 ind                       1.84
## 3 rep                       1.86
## 4 <NA>                      1.85
```

```
## # A tibble: 4 x 2
##   pid_chr mean_rising_temperatures
##   <chr>                     <dbl>
## 1 dem                        1.72
## 2 ind                        1.13
## 3 rep                      -0.270
## 4 <NA>                       0.763


##
##   Welch Two Sample t-test
##
## data:  anes_2016$opinion_rising_temperatures and anes_2020$opinion_rising_temperatures
## t = -13.918, df = 7792.5, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.5542027 -0.4173636
## sample estimates:
## mean of x mean of y
##  0.844939  1.330722


## Warning: Use of `anes_2016$opinion_rising_temperatures` is discouraged.
## i Use `opinion_rising_temperatures` instead.


## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.


## Warning: Removed 91 rows containing non-finite values (`stat_count()`).
```
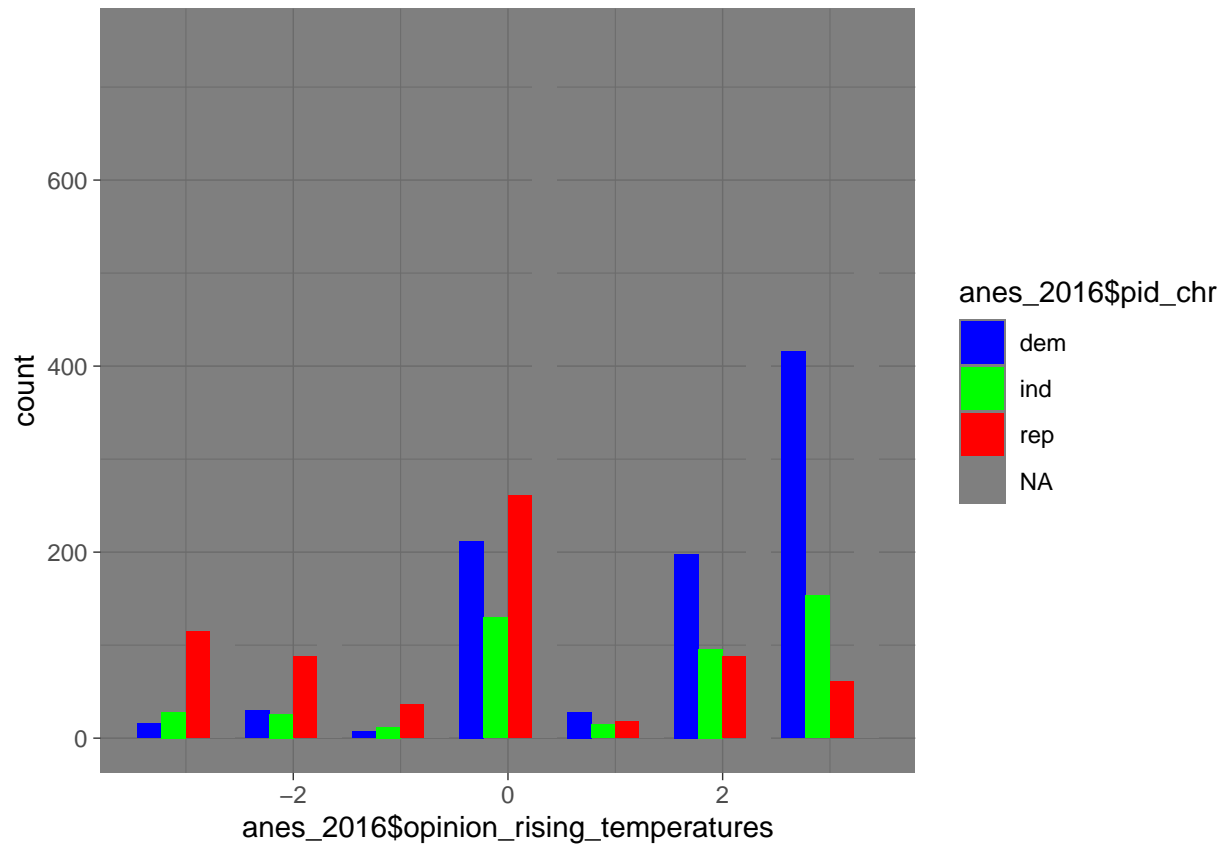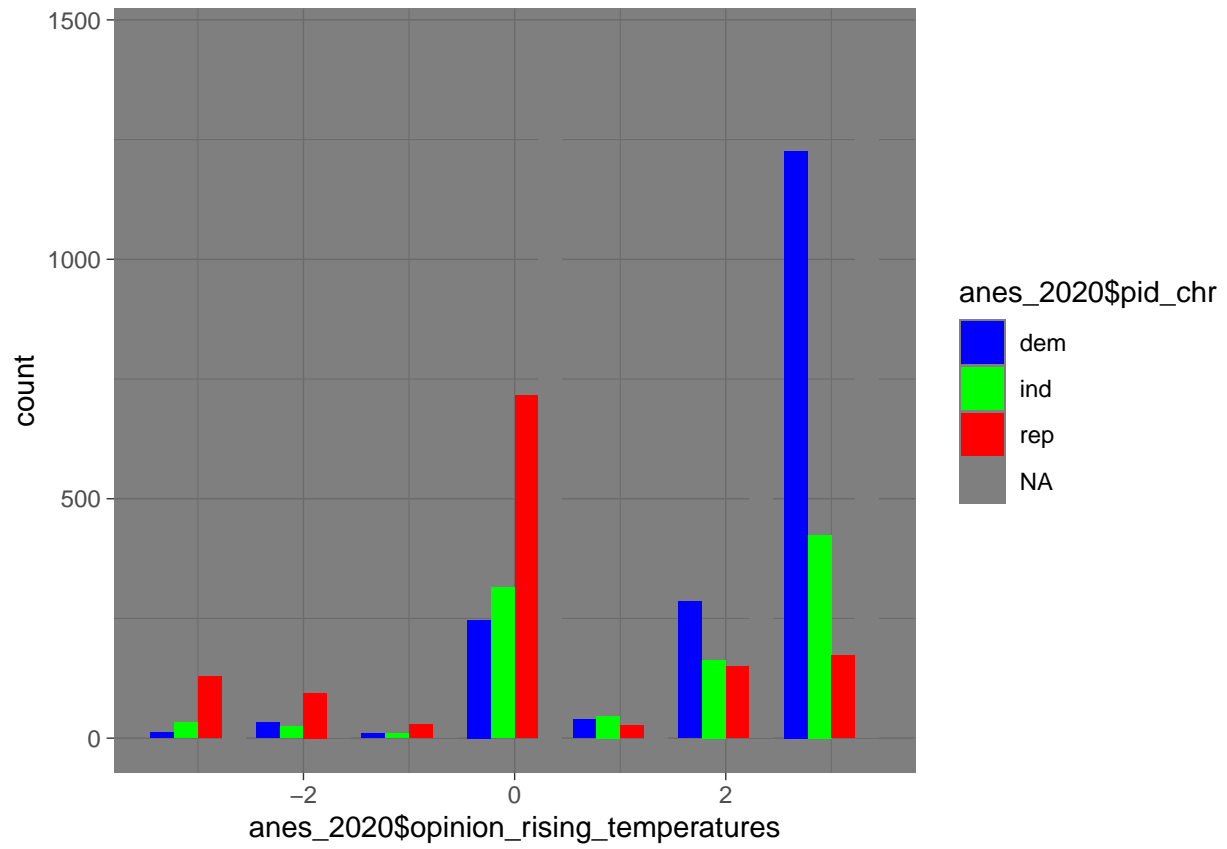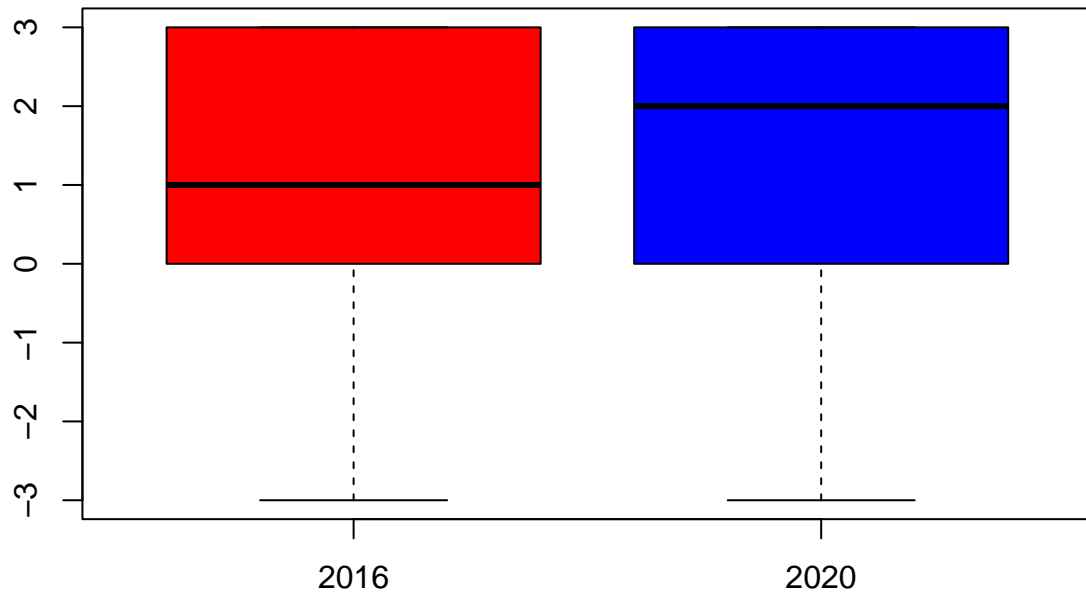
```
## Warning: Use of 'anes_2020$opinion_rising_temperatures' is discouraged.
## i Use 'opinion_rising_temperatures' instead.

## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.

## Warning: Removed 110 rows containing non-finite values ('stat_count()').
```
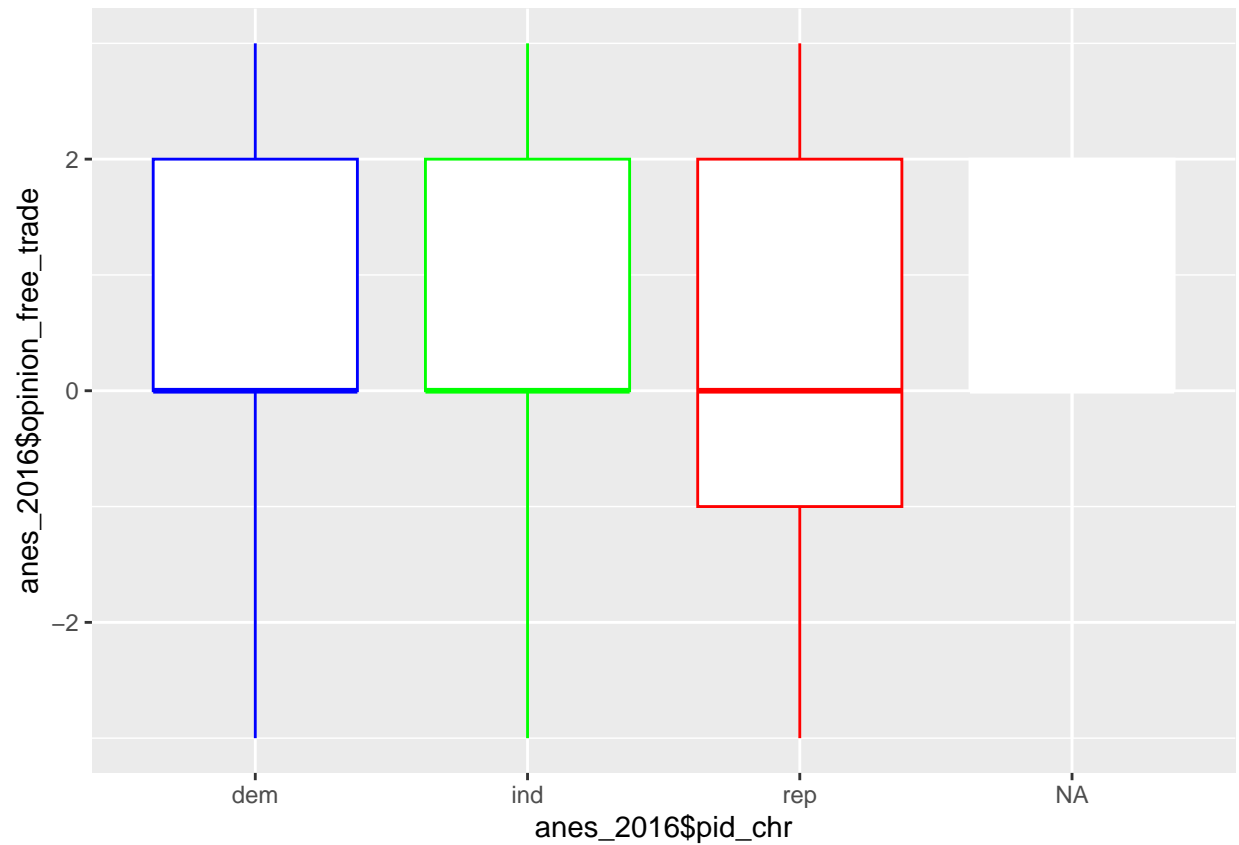
Public oppinion in health care


```
## Warning: Use of `anes_2016$pid_chr` is discouraged.
## i Use `pid_chr` instead.

## Warning: Use of `anes_2016$opinion_free_trade` is discouraged.
## i Use `opinion_free_trade` instead.

## Warning: Removed 676 rows containing non-finite values (`stat_boxplot()`).
```
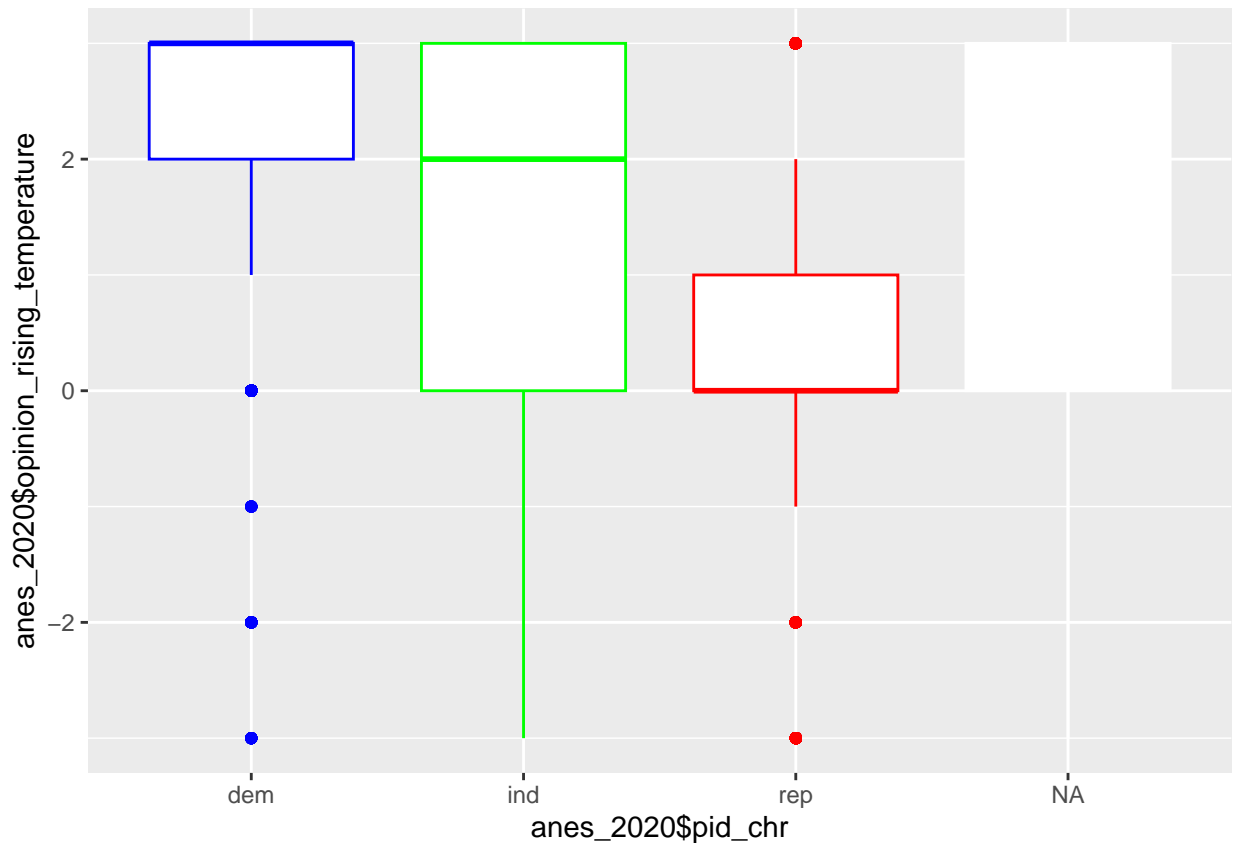
```
## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.

## Warning: Use of 'anes_2020$opinion_rising_temperature' is discouraged.
## i Use 'opinion_rising_temperature' instead.

## Warning: Removed 110 rows containing non-finite values ('stat_boxplot()').
```

FAVOR OR OPPOSE VOTE BY MAIL

```
## # A tibble: 4 x 2
##   pid_chr sd_vote_mail
##   <chr>          <dbl>
## 1 dem             1.89
## 2 ind             2.01
## 3 rep             1.84
## 4 <NA>            2.06
```

```
## # A tibble: 4 x 2
##   pid_chr mean_vote_mail
##   <chr>            <dbl>
## 1 dem              0.716
## 2 ind             -0.169
## 3 rep             -1.57
## 4 <NA>            -0.465
```

```
## Warning: Use of 'anes_2020$opinion_vote_mail' is discouraged.
## i Use 'opinion_vote_mail' instead.
```

```
## Warning: Use of 'anes_2020$pid_chr' is discouraged.
## i Use 'pid_chr' instead.
```
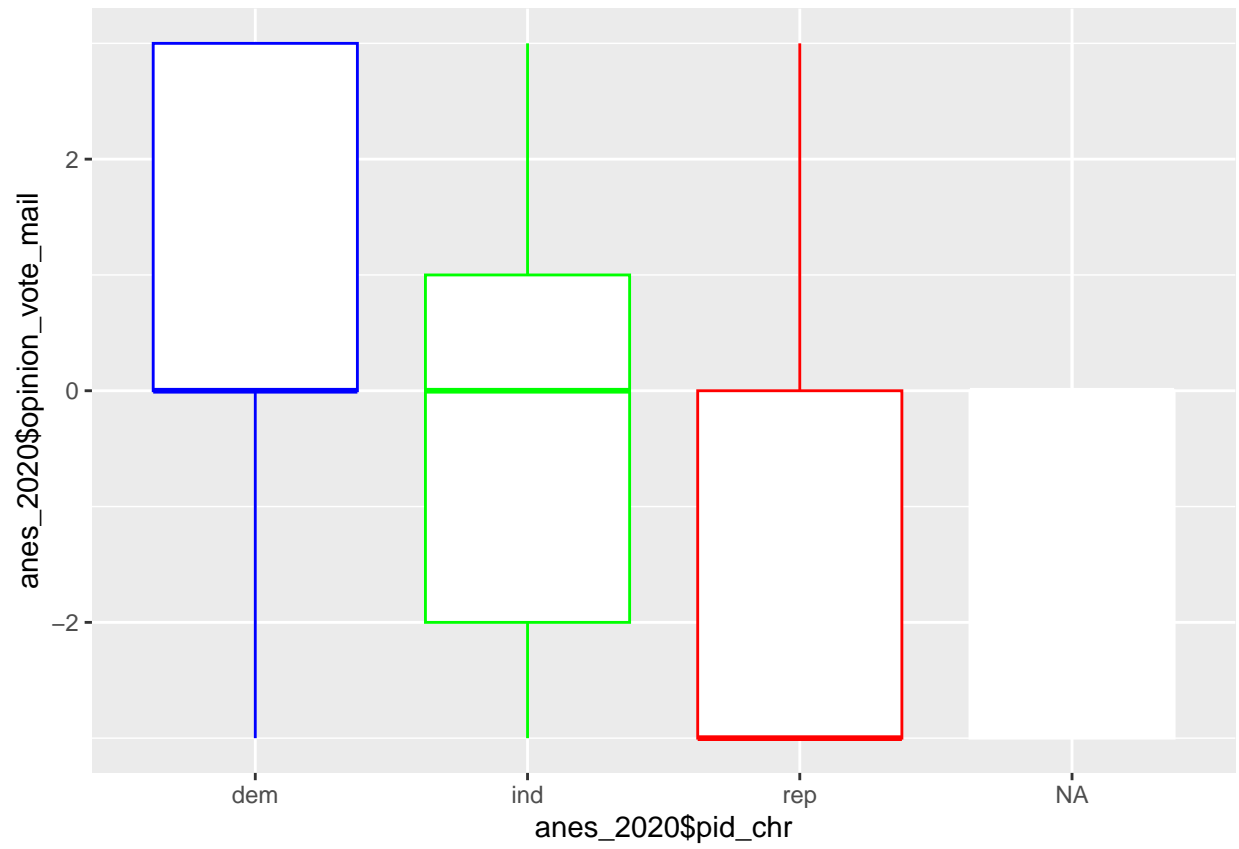
```
## Warning: Removed 23 rows containing non-finite values (`stat_count()`).
```
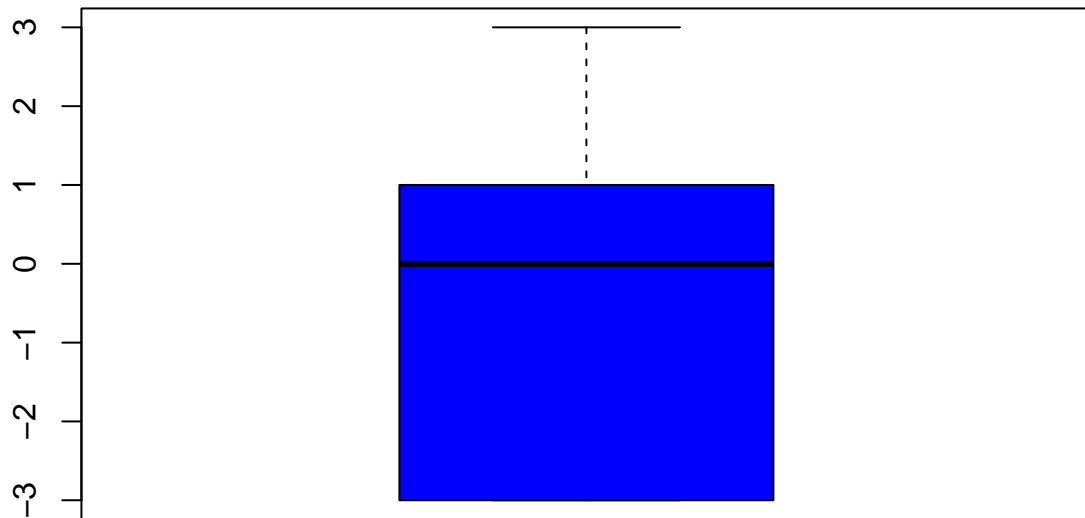


```
## Warning: Use of `anes_2020$pid_chr` is discouraged.
## i Use `pid_chr` instead.
```
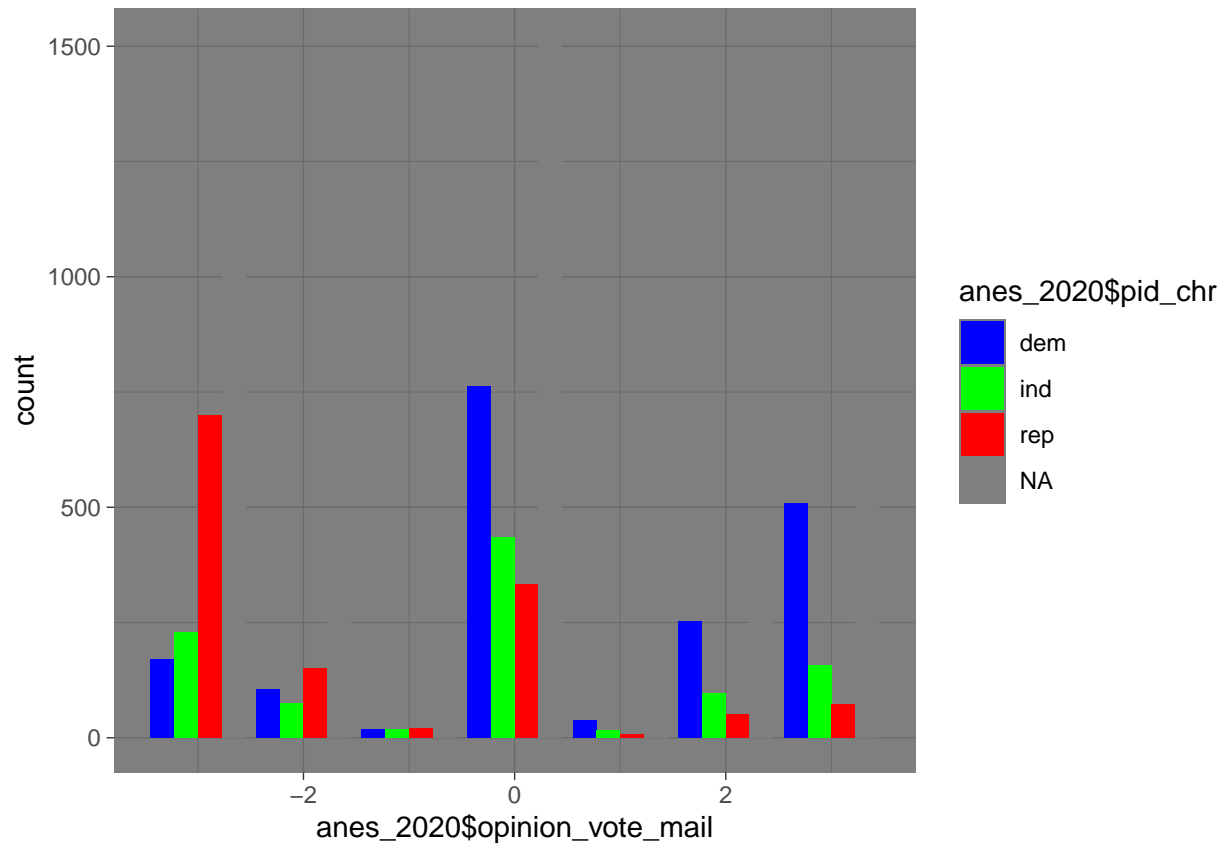
```
## Warning: Use of `anes_2020$opinion_vote_mail` is discouraged.
## i Use `opinion_vote_mail` instead.
```

```
## Warning: Removed 23 rows containing non-finite values (`stat_boxplot()`).
```

Public oppinion vote by mail

Thank you!