Data 607 - Connin HW1

Overview

The ‘Atlas Of Redistricting’ is a project published online by Nate Silver’s “FiveThirtyEight” website early 2018. The project describes and maps various congressional redistricting scenarios in the U.S. in order to explore how changes in district boundaries impact the racial and partisan makeup of congress.

A description and data used in the project can be found at the following websites.

Project description:

https://fivethirtyeight.com/features/we-drew-2568-congressional-districts-by-hand-heres-how/ (https://fivethirtyeight.com/features/we-drew-2568-congressional-districts-by-hand-heres-how/)

Redistricting atlas:

https://projects.fivethirtyeight.com/redistricting-maps/ (https://projects.fivethirtyeight.com/redistricting-maps/)

Redistricting atlas data:

https://github.com/fivethirtyeight/redistricting-atlas-data (https://github.com/fivethirtyeight/redistricting-atlas-data)

The code below supports an initial review of the data sets (‘districts’, ‘county_assignments’, ‘states’).

```
# Load R packages

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

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

```
library(magrittr)
```

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

```
## The following object is masked from 'package:purrr':
##
##     set_names
```

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

```
library(readr)
library(ggplot2)
library(cowplot)

# Read in 358.com project data

districts <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/redistricting-atlas-dat
a/master/districts.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   statefp = col_character(),
##   state = col_character(),
##   maptype = col_character(),
##   district = col_character(),
##   population = col_double(),
##   population_18_over = col_double(),
##   PVI = col_double(),
##   dem_chance = col_double(),
##   `Non-Hispanic White` = col_double(),
##   `African-American` = col_double(),
##   `Hispanic/Latino` = col_double(),
##   Asian = col_double(),
##   `Native American` = col_double(),
##   `Pacific Islander` = col_double(),
##   Other = col_character(),
##   race_category = col_character(),
##   minority_chance = col_double(),
##   current_map = col_logical(),
##   impossible = col_logical()
## )
```

```
county_files <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/redistricting-atlas-
data/master/county_assignments.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   statefp = col_character(),
##   state = col_character(),
##   maptype = col_character(),
##   countyfp = col_character(),
##   county = col_character(),
##   cd = col_character()
## )
```

```
states <- read_csv("https://raw.githubusercontent.com/fivethirtyeight/redistricting-atlas-data/master/states.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   statefp = col_character(),
##   state = col_character(),
##   maptype = col_character(),
##   districts = col_double(),
##   county_splits = col_double(),
##   efficiency_gap = col_double(),
##   efficiency_gap_extra_seats = col_character(),
##   district_perimeters = col_double(),
##   state_perimeter = col_double(),
##   interior_perimeter_measure = col_double(),
##   compactness_rank = col_double()
## )
```

The Datasets

```
# review dataframe dimensions and components

districts%>%glimpse()
```

```
## Rows: 3,480
## Columns: 19
## $ statefp              <chr> "02", "02", "02", "02", "02", "02", "02", "02"...
## $ state                <chr> "AK", "AK", "AK", "AK", "AK", "AK", "AK", "AK"...
## $ maptype              <chr> "Compact", "Competitive", "Dem", "GOP", "MajMi...
## $ district             <chr> "00", "00", "00", "00", "00", "00", "00", "00"...
## $ population           <dbl> 710231, 710231, 710231, 710231, 710231, 710231...
## $ population_18_over   <dbl> 522853, 522853, 522853, 522853, 522853, 522853...
## $ PVI                  <dbl> -9.39, -9.39, -9.39, -9.39, -9.39, -9.39, -9.3...
## $ dem_chance           <dbl> 5.40942673, 5.40942673, 5.40942673, 5.40942673...
## $ `Non-Hispanic White` <dbl> 68.27674, 68.27674, 68.27674, 68.27674, 68.276...
## $ `African-American`   <dbl> 3.084806, 3.084806, 3.084806, 3.084806, 3.0848...
## $ `Hispanic/Latino`    <dbl> 4.673780, 4.673780, 4.673780, 4.673780, 4.6737...
## $ Asian                <dbl> 5.3328565, 5.3328565, 5.3328565, 5.3328565, 5....
## $ `Native American`    <dbl> 13.2700778, 13.2700778, 13.2700778, 13.2700778...
## $ `Pacific Islander`   <dbl> 0.85989752, 0.85989752, 0.85989752, 0.85989752...
## $ Other                <chr> "4.501838948997137%", "4.501838948997137%", "4...
## $ race_category        <chr> "Non-Hispanic White Majority", "Non-Hispanic W...
## $ minority_chance      <dbl> 9.8894051, 9.8894051, 9.8894051, 9.8894051, 9....
## $ current_map          <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
## $ impossible           <lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA...
```

```
county_files%>%glimpse()
```

```
## Rows: 27,959
## Columns: 6
## $ statefp  <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01", "01"...
## $ state    <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL"...
## $ maptype  <chr> "Compact", "Compact", "Compact", "Compact", "Compact", "Co...
## $ countyfp <chr> "01001", "01003", "01005", "01007", "01009", "01011", "010...
## $ county   <chr> "Autauga County", "Baldwin County", "Barbour County", "Bib...
## $ cd       <chr> "07", "01", "02", "07", "04", "02", "07", "03", "02", "03"...
```

```
states%>%glimpse()
```

```
## Rows: 400
## Columns: 11
## $ statefp                   <chr> "02", "02", "02", "02", "02", "02", "02"...
## $ state                     <chr> "AK", "AK", "AK", "AK", "AK", "AK", "AK"...
## $ maptype                   <chr> "Compact", "Competitive", "Dem", "GOP", ...
## $ districts                 <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 7, 7, 7, 7, 7, 7...
## $ county_splits             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 5, 12, 8, 8, 8, ...
## $ efficiency_gap            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, -0.05672...
## $ efficiency_gap_extra_seats <chr> NA, NA, NA, NA, NA, NA, NA, NA, "D+0", "...
## $ district_perimeters       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 55.39691...
## $ state_perimeter           <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 18.98561...
## $ interior_perimeter_measure <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 18.20565...
## $ compactness_rank          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, 1, 6, 3,...
```

Guiding Question

Are political outcomes (republican vs democratic congressional seats) in the redistricting scenarios related to changing percentages of minority voters at the state level?

I have organized/cleaned the ‘districts’ dataset in order to assess the question above.

```
# create a view of the districts dataframe

districts%>%view()

# return total number of missing values

sprintf("The total number of NA and NAN is %d", sum(is.na(districts)))
```

```
## [1] "The total number of NA and NAN is 6714"
```

```
# identify/count missing values in district by column

map(districts, ~sum(is.na(.))) #-- > using purrr, note (.) refers to cols
```

```
## $statefp
## [1] 0
##
## $state
## [1] 0
##
## $maptype
## [1] 0
##
## $district
## [1] 0
##
## $population
## [1] 0
##
## $population_18_over
## [1] 0
##
## $PVI
## [1] 0
##
## $dem_chance
## [1] 0
##
## $`Non-Hispanic White`
## [1] 0
##
## $`African-American`
## [1] 0
##
## $`Hispanic/Latino`
## [1] 0
##
## $Asian
## [1] 0
##
## $`Native American`
## [1] 0
##
## $`Pacific Islander`
## [1] 0
##
## $Other
## [1] 0
##
## $race_category
## [1] 0
##
## $minority_chance
## [1] 0
##
## $current_map
## [1] 3321
```

```
##
## $impossible
## [1] 3393
```

```
# count the number of duplicate rows

sprintf("The number of duplicate rows is %d", sum(duplicated(districts)))
```

```
## [1] "The number of duplicate rows is 0"
```

```
# select subset of columns for new dataframe

d <- districts%>%select(-c(current_map, impossible))

#set column names to lower case

names(d)%<>%tolower

#update column names for districts dataframe

d%<>%dplyr::rename(state_fips_code=statefp, district_number=district,cook_partisan_index=pvi, no
n_hispanic_white=`non-hispanic white`, african_american =`african-american`, hispanic_latino =`h
ispanic/latino`, native_american =`native american`, pacific_islander =`pacific islander`)

# remove trailing '%' from values in Other column

d<-separate(data = d, col = other, into = c("other"), sep = "%")
```

```
## Warning: Expected 1 pieces. Additional pieces discarded in 3480 rows [1, 2, 3,
## 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, ...].
```

```
# change 'other' col to dbl format

d%<>%mutate(other = as.numeric(other))

# pivot select cols to long form

d<-pivot_longer(d, cols=9:15, names_to = 'ethnicity', values_to ='percent_of_voters')

#reduce number of categorical variables in race_category column

d <- mutate(d, race_category = ifelse(race_category == "Non-Hispanic White Majority", "white_maj
ority", "non_white_majority"))

# subset dataframe on maptypes, rows: current, democrat, republican, competitive ("|" --> 'or')

d <- filter(d, maptype == "current" | maptype =="Dem" | maptype =="GOP" | maptype == "Competitiv
e")

# rename category values in maptype column

d<-d%>%mutate(maptype=recode(maptype, 'Competitive'='competitive', 'Dem'='democrat', 'GOP'='repu
blican'))

# review updates to dataframe

head(d, 5)
```

```
## # A tibble: 5 x 12
##   state_fips_code state maptype district_number population population_18_o~
##   <chr>           <chr> <chr>   <chr>                <dbl>            <dbl>
## 1 02              AK    compet~ 00                  710231           522853
## 2 02              AK    compet~ 00                  710231           522853
## 3 02              AK    compet~ 00                  710231           522853
## 4 02              AK    compet~ 00                  710231           522853
## 5 02              AK    compet~ 00                  710231           522853
## # ... with 6 more variables: cook_partisan_index <dbl>, dem_chance <dbl>,
## #   race_category <chr>, minority_chance <dbl>, ethnicity <chr>,
## #   percent_of_voters <dbl>
```

Exploratory Data Analyis

Here we compute basic statistical measures for all numerical variables in the dataset.

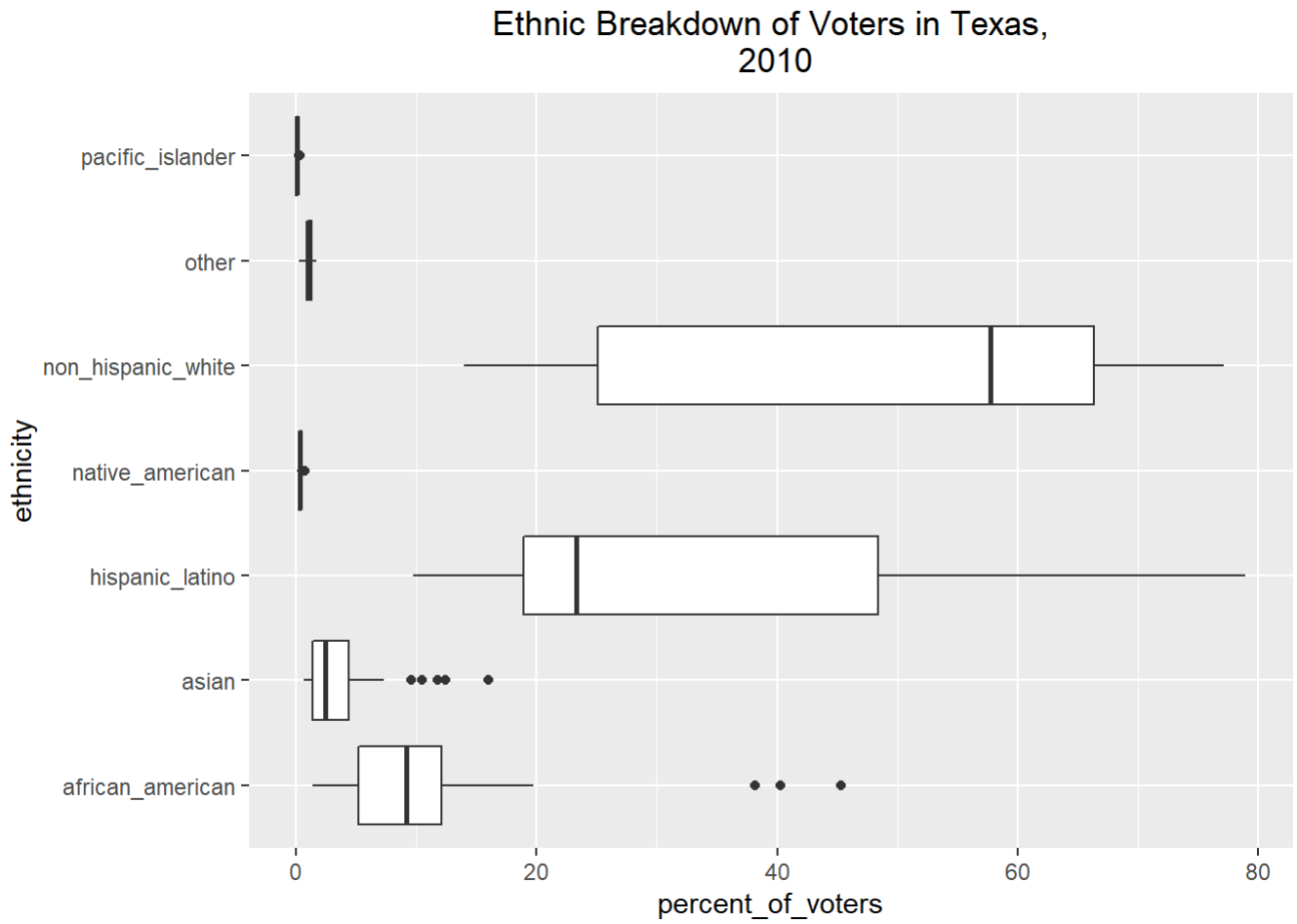
In addition, we also compare changes in the percent of minorities voting in Texas (a potential swing state) under two scenarios: ‘current’ vs. ‘competitive’.

The latter data are drawn from 2010 census results while the former reflects estimates based on redistricting to enhance two-party competitiveness at the district level.

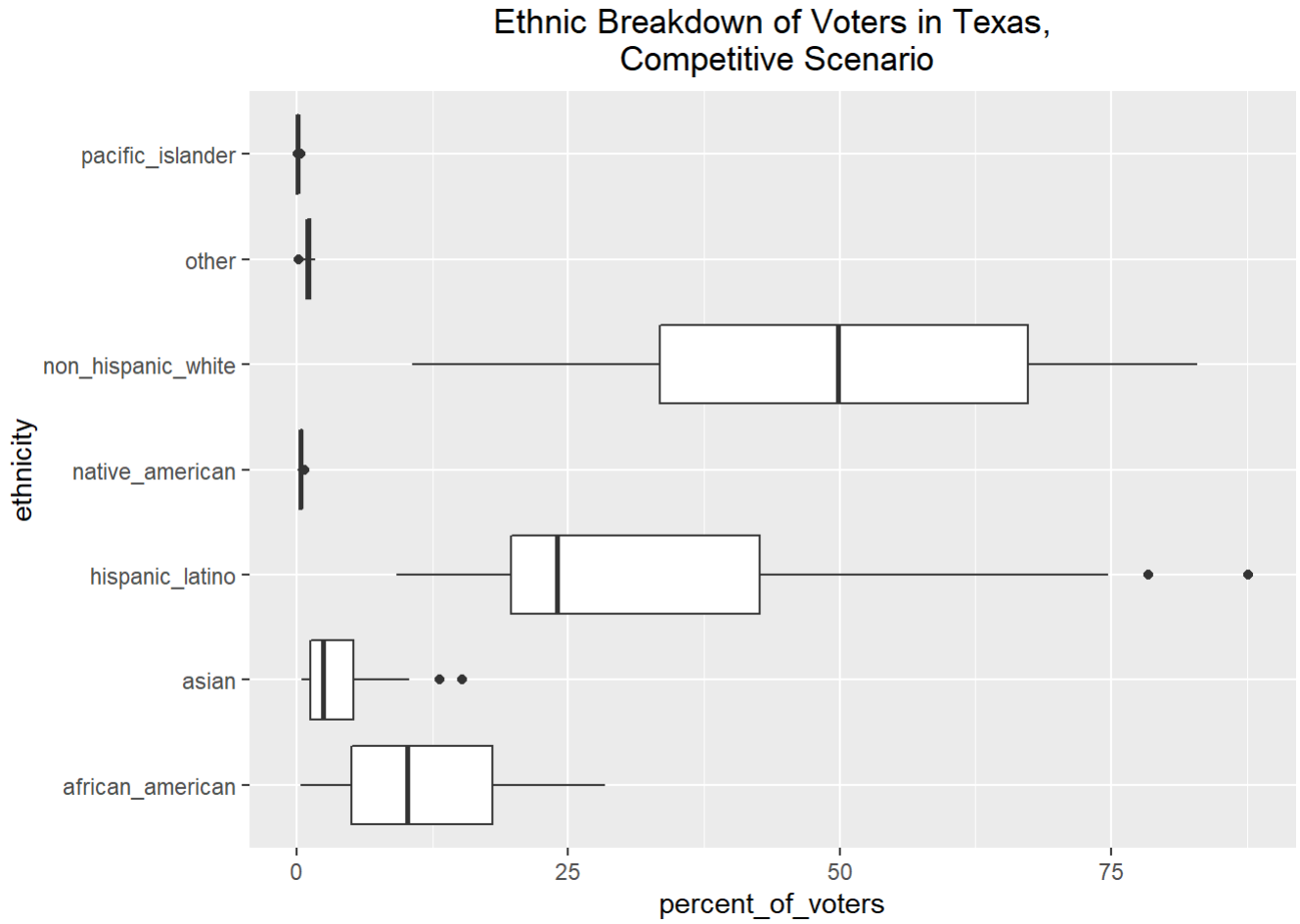
```
# return statistical measures for numerical variables

summary(d)
```

```
##  state_fips_code       state              maptype          district_number
##  Length:12180       Length:12180       Length:12180       Length:12180
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##    population     population_18_over cook_partisan_index   dem_chance
##  Min.   :525777   Min.   :410765     Min.   :-34.1800    Min.   :  0.003
##  1st Qu.:698180   1st Qu.:519620     1st Qu.:-11.1025    1st Qu.:  3.280
##  Median :705974   Median :542062     Median : -1.4500    Median : 37.832
##  Mean   :708375   Mean   :538075     Mean   :  0.4453    Mean   : 47.488
##  3rd Qu.:720932   3rd Qu.:557644     3rd Qu.:  9.7950    3rd Qu.: 94.897
##  Max.   :989415   Max.   :765852     Max.   : 44.4800    Max.   :100.000
##  race_category      minority_chance    ethnicity         percent_of_voters
##  Length:12180       Min.   : 0.2844   Length:12180       Min.   : 0.00674
##  Class :character   1st Qu.: 1.5536   Class :character   1st Qu.: 0.50337
##  Mode  :character   Median : 4.5669   Mode  :character   Median : 2.12516
##                     Mean   :19.3643                      Mean   :14.28587
##                     3rd Qu.:19.5905                      3rd Qu.:12.47718
##                     Max.   :99.3074                      Max.   :96.97341
```

```
# a boxplot graph of percent of votes by ethnicity in 2010

(ethnic1 <- d%>%group_by(state)%>%filter(state =='TX')%>%filter(maptype == 'current')%>%group_by
(ethnicity) %>%ggplot(aes(x=ethnicity,y=percent_of_voters))+geom_boxplot()+coord_flip()+ggtitle(
'Ethnic Breakdown of Voters in Texas,\n 2010')+theme(plot.title = element_text(hjust=0.5)))
```




```
# a boxplot graph of percent of votes by ethnicity in the competitive scenario

(ethnic2 <- d%>%group_by(state)%>%filter(state =='TX')%>%filter(maptype == 'competitive')%>%grou
p_by(ethnicity)%>%ggplot(aes(x=ethnicity,y=percent_of_voters))+geom_boxplot()+coord_flip()+ggtit
le('Ethnic Breakdown of Voters in Texas,\n Competitive Scenario')+theme(plot.title = element_tex
t(hjust=0.5)))
```

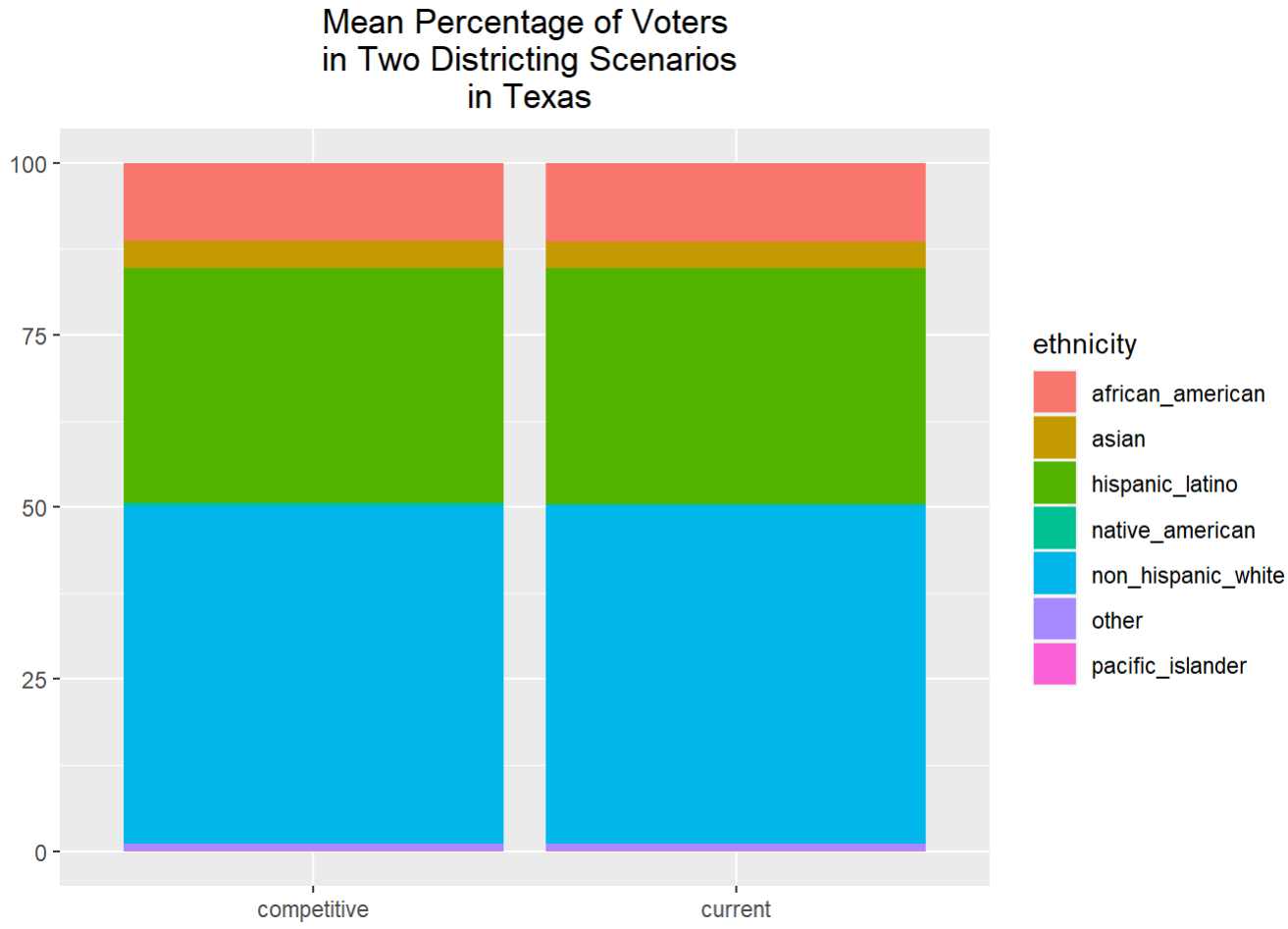



```
# calculate the mean/median percent of voting by ethnic groups compared between current and comp
etitive scenarios

(scenario <- d%>%group_by(maptype, ethnicity)%>%filter(state =='TX')%>%filter(maptype == 'curren
t'| maptype=='competitive')%>%summarize(mean_pct = mean(percent_of_voters), median_pct=median(pe
rcent_of_voters)))
```

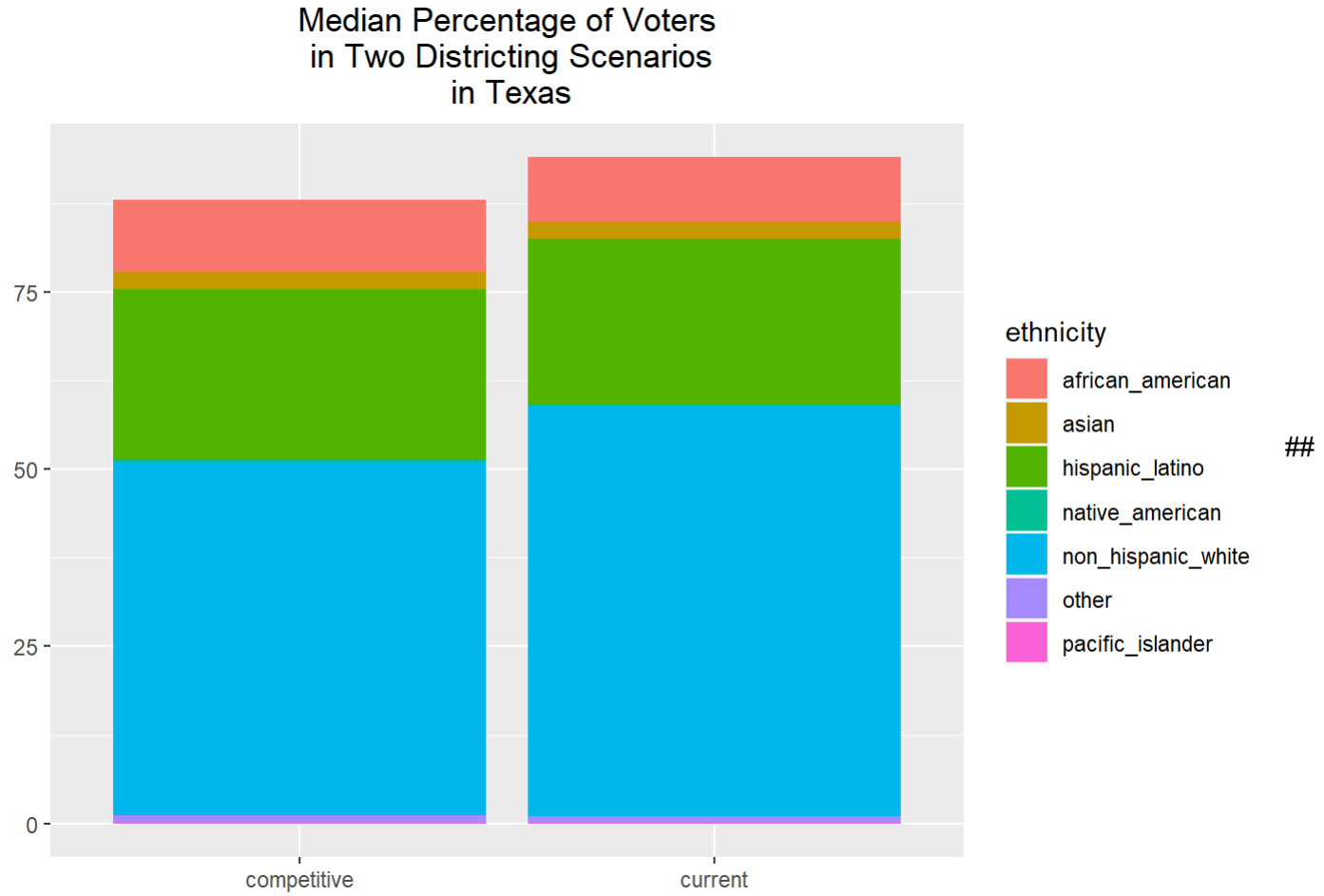
```
## `summarise()` has grouped output by 'maptype'. You can override using the `.groups` argument.
```

```
## # A tibble: 14 x 4
## # Groups:   maptype [2]
##    maptype     ethnicity          mean_pct median_pct
##    <chr>       <chr>                 <dbl>      <dbl>
##  1 competitive african_american    11.4       10.2
##  2 competitive asian                3.91       2.44
##  3 competitive hispanic_latino     34.2       24.0
##  4 competitive native_american      0.336      0.334
##  5 competitive non_hispanic_white  49.1       49.9
##  6 competitive other                1.05       1.08
##  7 competitive pacific_islander     0.0702     0.0469
##  8 current     african_american    11.4        9.15
##  9 current     asian                3.91       2.44
## 10 current     hispanic_latino     34.2       23.3
## 11 current     native_american      0.336      0.332
## 12 current     non_hispanic_white  49.1       57.7
## 13 current     other                1.05       1.05
## 14 current     pacific_islander     0.0703     0.0486
```

```
# compare scenarios using stacked barplot

ggplot(scenario, aes(fill=ethnicity, y=mean_pct, x=maptype))+geom_bar(position='stack', stat='id
entity')+ggtitle('Mean Percentage of Voters\n in Two Districting Scenarios\n in Texas')+theme(pl
ot.title = element_text(hjust=0.5))+theme(axis.title.x = element_blank())+theme(axis.title.y = e
lement_blank())
```



```
ggplot(scenario, aes(fill=ethnicity, y=median_pct, x=maptype))+geom_bar(position='stack', stat='identity')+ggtitle('Median Percentage of Voters\n in Two Districting Scenarios\n in Texas')+theme(plot.title = element_text(hjust=0.5))+theme(axis.title.x = element_blank())+theme(axis.title.y = element_blank())
```



##

Findings and Recommendations

Initial review of the data indicates the following:

1. Non-hispanic whites comprise the largest voting block in Texas followed by Latinos and African Americans. Variance in voting between districts by ethnic groups, correlates with the percent of votes attributable to each group.
2. There are no state-level differences in the mean percent of votes by ethnic group compared between the ‘current’ vs. ‘competitive’ scenarios.
3. The median value for percent of votes by non_hispanic whites decreased from 57.7% to 49.8% (state level) compared between the ‘current’ vs. ‘competitive’ scenarios - indicating a decrease in the relative percentage of white voters in select districts as a result of redistricting. In contrast, the median values (state level) for other ethnic groups remained relatively unchanged.
4. Additional analyses should focus at the district level in order to explicate factors that shape election outcomes at the state level. These factors may include voter population densities, district geometries, etc.
5. The three sets of data provided by the authors lack a common variable to enable joins between these sets. Additional information should be acquired to link these sets for more extensive analyses.