Homework 4 Data Preparation

Ashwin Malshe

```
ele_2020 <- read_csv("https://raw.githubusercontent.com/tonmcg/US_County_Level_Election_Results_08-20/
                     master/2020_US_County_Level_Presidential_Results.csv")
ele_2016 <- read_csv("https://raw.githubusercontent.com/tonmcg/US_County_Level_Election_Results_08-20/
                     master/2016_US_County_Level_Presidential_Results.csv")
```

```
## Rows: 3152 Columns: 10
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): state_name, county_fips, county_name
## dbl (7): votes_gop, votes_dem, total_votes, diff, per_gop, per_dem, per_poin...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
## New names:
## Rows: 3141 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): per_point_diff, state_abbr, county_name
## dbl (7): ...1, votes_dem, votes_gop, total_votes, per_dem, per_gop, combined...
## num (1): diff
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
names(ele_2016)
```

```
##  [1] "...1"           "votes_dem"      "votes_gop"      "total_votes"
##  [5] "per_dem"        "per_gop"        "diff"           "per_point_diff"
##  [9] "state_abbr"     "county_name"    "combined_fips"
```

```
names(ele_2020)
```

```
##  [1] "state_name"     "county_fips"    "county_name"    "votes_gop"
##  [5] "votes_dem"      "total_votes"    "diff"           "per_gop"
##  [9] "per_dem"        "per_point_diff"
```

```
ele_2016 <- ele_2016 %>%
  select(-1, -per_point_diff, -county_name, -diff) %>%
  mutate(combined_fips = sprintf("%05d", combined_fips)) %>%
  rename(votes_dem_16 = votes_dem,
         votes_gop_16 = votes_gop,
         total_votes_16 = total_votes,
```

```
         per_dem_16 = per_dem,
         per_gop_16 = per_gop)
```

```
ele_2020 <- ele_2020 %>%
  select(-diff) %>%
    rename(votes_dem_20 = votes_dem,
         votes_gop_20 = votes_gop,
         total_votes_20 = total_votes,
         per_dem_20 = per_dem,
         per_gop_20 = per_gop,
         per_point_diff_20 = per_point_diff)
```

```
total_election <- ele_2020 %>% left_join(ele_2016, by = c("county_fips" = "combined_fips"))
```

Variable description

```
knitr::kable(var_desc)
```

state_name	Name.of.the.state
county_fips	County FIPS code. Not relevant for you.
county_name	Name of the county
votes_gop_20	GOP votes in 2020 (Trump) in the county
votes_dem_20	Dem votes in 2020 (Biden) in the county
total_votes_20	Total votes cast in 2020 in the county
per_gop_20	Percentage GOP votes in 2020 in the county
per_dem_20	Percentage Dem votes in 2020 in the county (I recommend recreating this variable in Tableau)
per_point_diff_20	Percentage vote difference in 2020 (Don't use this variable)
votes_dem_16	GOP votes in 2016 (Trump) in the county
votes_gop_16	Dem votes in 2016 (Clinton) in the county
total_votes_16	Total votes cast in 2016 in the county
per_dem_16	Percentage GOP votes in 2016 in the county
per_gop_16	Percentage Dem votes in 2016 in the county
state_abbr	Two letter abbreviation of state name