County Level Data

2025-11-20

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.6
## v forcats   1.0.1     v stringr   1.6.0
## v ggplot2   4.0.1     v tibble    3.3.0
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.2.0
## -- 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
```

```
library(dplyr)
library(ggplot2)
library(stringr)
library(readr)
library(tigris)
```

```
## To enable caching of data, set `options(tigris_use_cache = TRUE)`
## in your R script or .Rprofile.
```

```
library(sf)
```

```
## Linking to GEOS 3.13.0, GDAL 3.8.5, PROJ 9.5.1; sf_use_s2() is TRUE
```

Load in election data and clean

```
props <- read_csv("Ohio Ballot Issues 2023 copy.csv")
```

```
## New names:
## Rows: 8880 Columns: 12
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (12): November 7, 2023 General Election Official Canvass, ...2, ...3, .....
## 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.
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
```

```
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...10`
## * `` -> `...12`
```

```
props <- props[-c(1:3), ]

colnames(props) <- c(
"County Name", "Precinct Name", "Precinct Code", "Region Name", "Media Market", "Registered Voters", "Ba
)

props_county <- props |>
  mutate(across(
    c("Registered Voters", "Ballots Counted",
      "abortion_yes", "abortion_no", "weed_yes", "weed_no"),
    ~ readr::parse_number(as.character(.x))
  )) |>
  group_by(`County Name`) |>
  summarise(
    `Registered Voters` = sum(`Registered Voters`, na.rm = TRUE),
    `Ballots Counted`   = sum(`Ballots Counted`, na.rm = TRUE),
    abortion_yes        = sum(abortion_yes, na.rm = TRUE),
    abortion_no         = sum(abortion_no, na.rm = TRUE),
    weed_yes            = sum(weed_yes, na.rm = TRUE),
    weed_no             = sum(weed_no, na.rm = TRUE)
  )

props_county <- props_county |>
  mutate(voter_turnout = (`Ballots Counted`/`Registered Voters`) * 100,
         abortion_diff = abortion_yes - abortion_no,
         abortion_turnout = abortion_yes + abortion_no,
         weed_diff = weed_yes - weed_no,
         weed_turnout = weed_yes + weed_no,
         abortion_weed_diff = abortion_diff - weed_diff,
         prop_turnout_diff = abortion_turnout - weed_turnout,
         abortion_margin = (abortion_diff / abortion_turnout) * 100,
         weed_margin = (weed_diff / weed_turnout) * 100,
         diff_margin = (abortion_weed_diff / `Ballots Counted`) * 100,
         across(c(abortion_margin, weed_margin, diff_margin), ~ round(.x, 2)))
```

Load in shapefile data

```
ohio_sf <- st_read("ohio_county_level.shp")
```

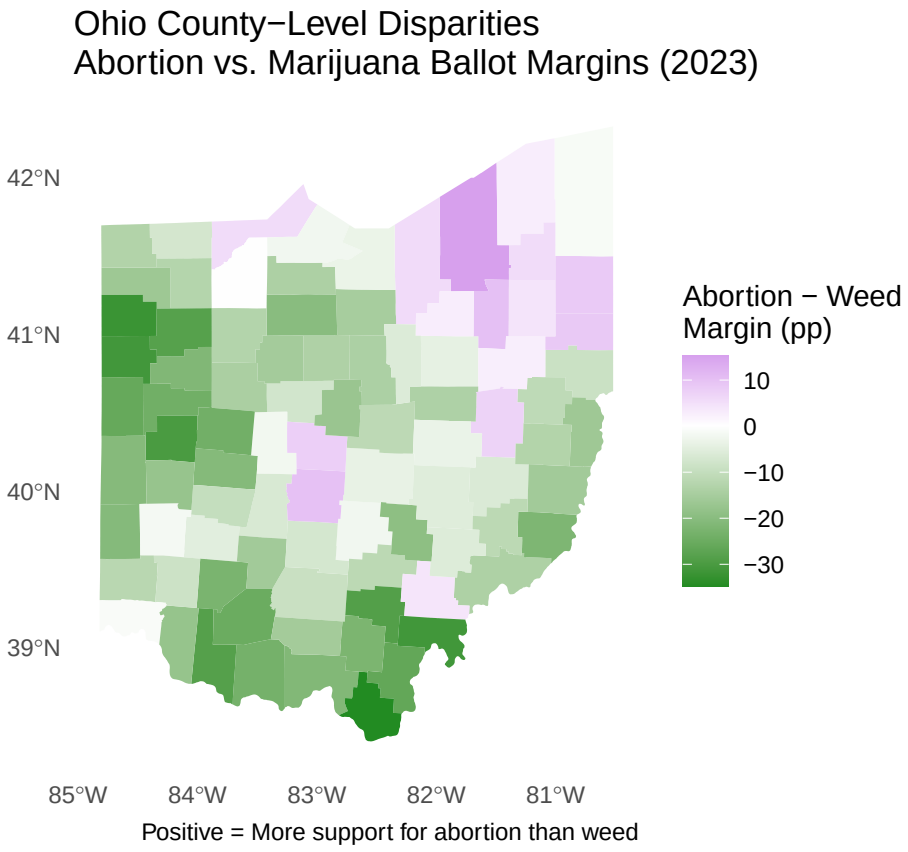
```
## Reading layer `ohio_county_level' from data source
##   `/Users/jackholland/Downloads/School/GOV 1372/Final Project/Gov Final Repository/ohio_county_level
##   using driver `ESRI Shapefile'
## Simple feature collection with 88 features and 4 fields
```

```
## Geometry type: POLYGON
## Dimension:     XY
## Bounding box:  xmin: -84.8203 ymin: 38.40342 xmax: -80.5187 ymax: 42.32713
## Geodetic CRS:  NAD83
```

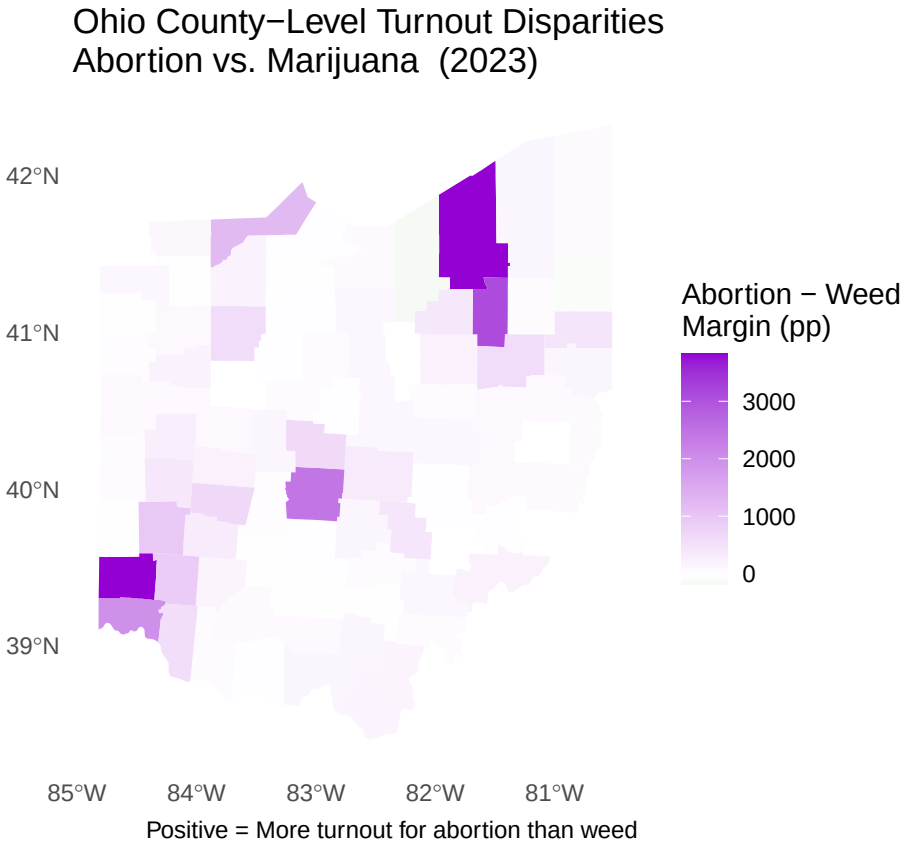
```
ohio_prop_map_df <- ohio_sf |>
  left_join(props_county, by = c("NAME" = "County Name"))
```

First Map!

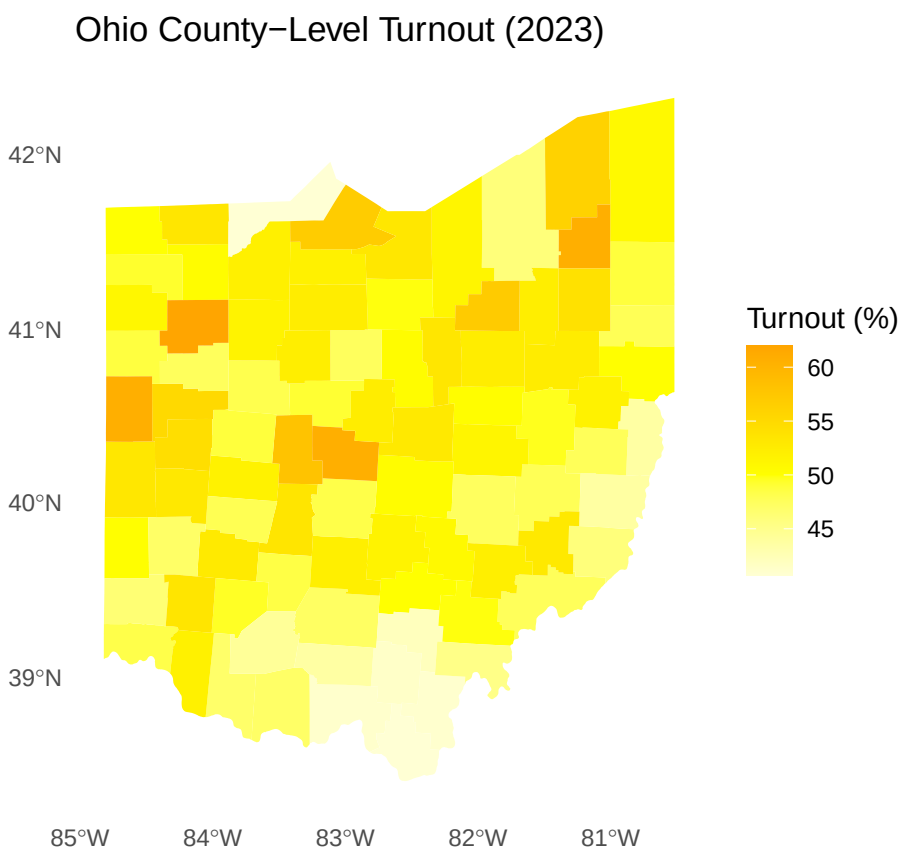
```
ggplot(ohio_prop_map_df) +
geom_sf(aes(fill = diff_margin), color = NA) + scale_fill_gradient2(
    low = "forestgreen",
    mid = "white",
    high = "darkviolet",
    midpoint = 0,
name = "Abortion - Weed\nMargin (pp)" )+
labs(
title = "Ohio County-Level Disparities\nAbortion vs. Marijuana Ballot Margins (2023)",
caption = "Positive = More support for abortion than weed") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
    legend.position = "right"
  )
```



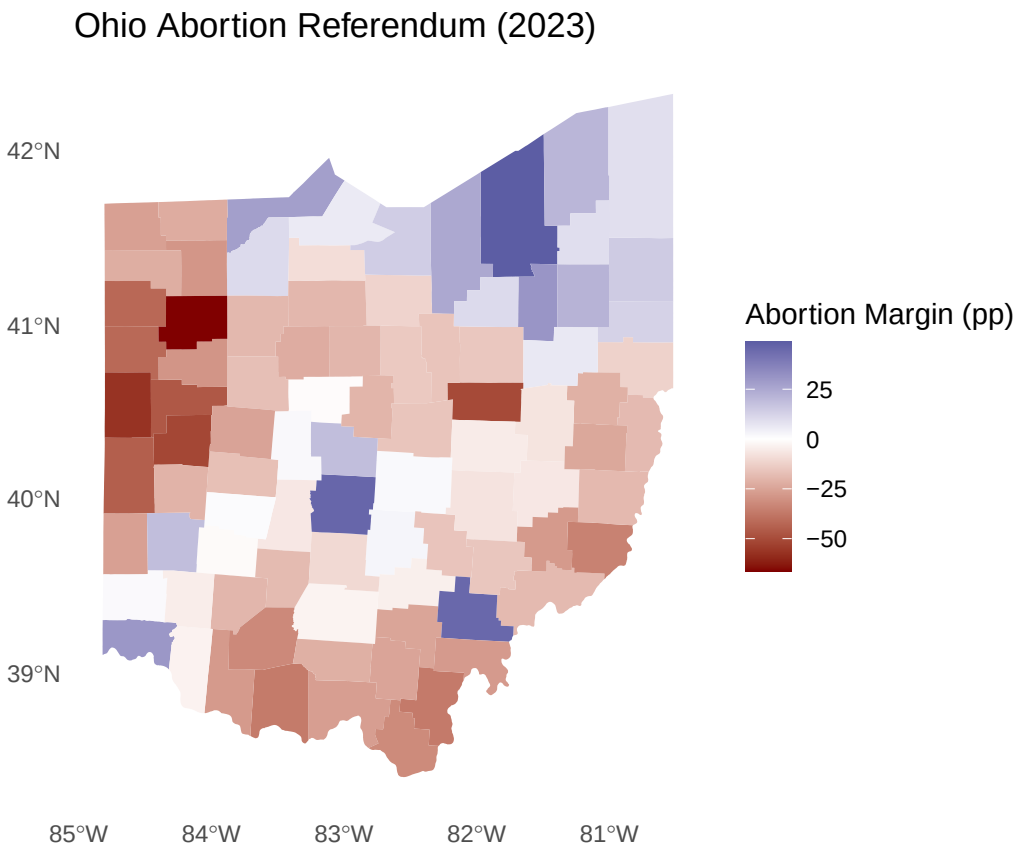

```
ggplot(ohio_prop_map_df) +
geom_sf(aes(fill = prop_turnout_diff), color = NA) + scale_fill_gradient2(
    low = "forestgreen",
    mid = "white",
    high = "darkviolet",
    midpoint = 0,
name = "Abortion - Weed\nMargin (pp)" )+
labs(
title = "Ohio County-Level Turnout Disparities\nAbortion vs. Marijuana  (2023)",
caption = "Positive = More turnout for abortion than weed") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
    legend.position = "right"
  )
```




```
ggplot(ohio_prop_map_df) +
geom_sf(aes(fill = voter_turnout), color = NA) + scale_fill_gradient2(
    low = "white",
    mid = "yellow1",
    high = "orange",
    midpoint = 50,
name = "Turnout (%)" )+
labs(
title = "Ohio County-Level Turnout (2023)") +
  theme_minimal() +
  theme(panel.grid = element_blank(),
    legend.position = "right"
  )
```



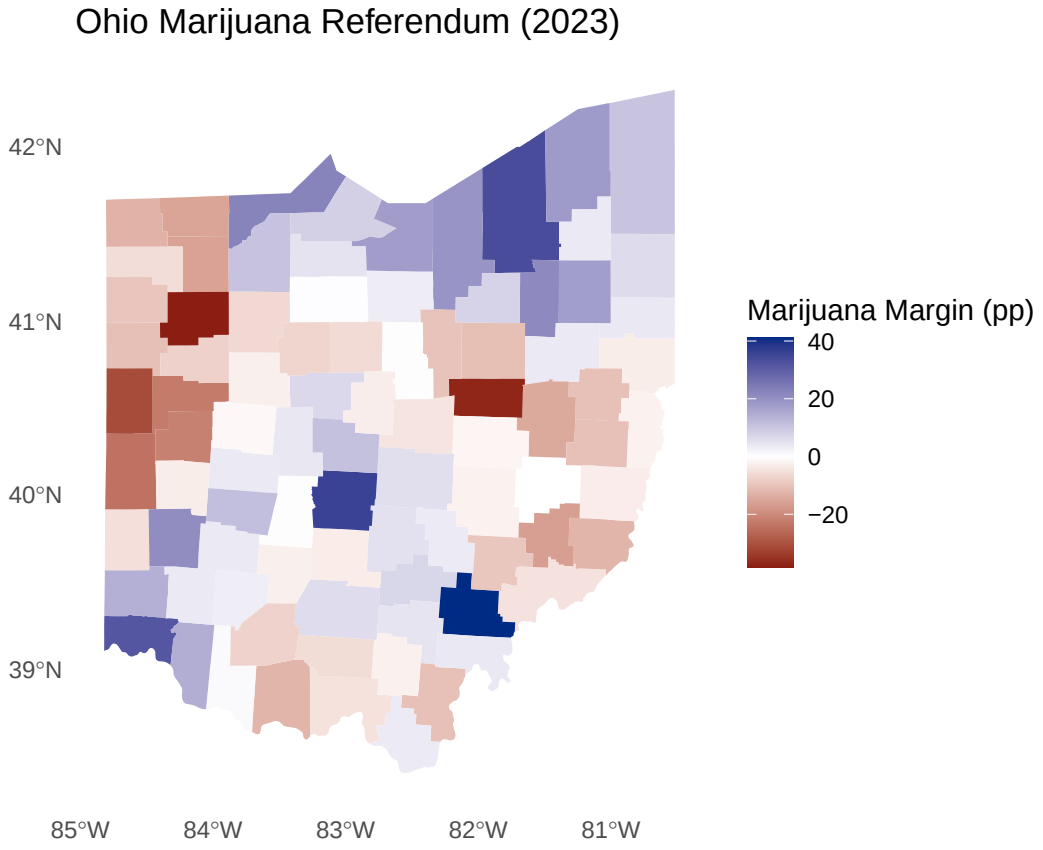

```
ggplot(ohio_prop_map_df) +
geom_sf(aes(fill = abortion_margin), color = NA) + scale_fill_gradient2(
    low = "#800000",
    high = "#002B84",
    midpoint = 0,
    name = "Abortion Margin (pp)"
)+ labs(
title = "Ohio Abortion Referendum (2023)" )+
theme_minimal() + theme(
panel.grid = element_blank(),
    legend.position = "right"
  )
```




```
min(ohio_prop_map_df$weed_margin)
```

```
## [1] -38.4
```

```
ggplot(ohio_prop_map_df) +
geom_sf(aes(fill = weed_margin), color = NA) + scale_fill_gradient2(
    low = "#800000",
    high = "#002B84",
    midpoint = 0,
    name = "Marijuana Margin (pp)"
)+ labs(
title = "Ohio Marijuana Referendum (2023)" )+
theme_minimal() + theme(
panel.grid = element_blank(),
    legend.position = "right"
  )
```



Looking at census data

```
library(tidyverse)
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
library(stringr)

age_raw <- read_csv("AgeRace.csv", show_col_types = FALSE)
```

```
## New names:
## * `` -> `...379`
```

```
meta_raw <- read_csv("ACSDP5Y2023.DP05-Column-Metadata.csv", show_col_types = FALSE)

age_raw <- age_raw[-1, ]
```

```r
estimate_cols <- names(age_raw)[str_detect(names(age_raw), "DP\\d{2}_.+E$")]
id_cols <- intersect(c("GEO_ID", "NAME"), names(age_raw))

age_est <- age_raw |>
  select(all_of(id_cols), all_of(estimate_cols))

rename_map <- meta_raw |>
  clean_names() |>
  transmute(
    var = column_name,
    label = label
  ) |>
  filter(str_detect(var, "DP\\d{2}_.+E$")) |>
  mutate(
    label = label |>
      str_replace_all("(?i)^Estimate!!", "") |>
      str_replace_all("!!", " / ") |>
      str_replace_all("[^A-Za-z0-9/ &-]+", " ") |>
      str_squish(),
    new_name = label |>
      str_to_lower() |>
      str_replace_all("/", " ") |>
      str_replace_all("&", "and") |>
      str_replace_all("-", " ") |>
      str_replace_all("\\s+", "_")
  ) |>
  select(var, new_name) |>
  distinct()

rename_map <- rename_map |>
  filter(str_detect(new_name, "percent"))

rename_map_keep <- rename_map |>
  filter(var %in% names(age_est)) |>
  mutate(new_name = make.unique(new_name, sep = "_"))

keep_vars <- c(id_cols, rename_map_keep$var)
valid_map <- setNames(rename_map_keep$var, rename_map_keep$new_name)

age_clean <- age_est |>
  select(all_of(keep_vars)) |>               # only keep id + percent vars
  rename(!!!valid_map) |>
  mutate(
    NAME = str_squish(NAME),
    county = NAME |>
      str_replace(",\\s*Ohio$", "") |>
      str_replace("\\s*County$", "") |>
      str_to_title()
  ) |>
  relocate(county, .after = NAME)

age_clean <- age_clean |>
  mutate(across(all_of(names(valid_map)), ~ as.numeric(.x)))
```

```
## Warning: There were 5 warnings in `mutate()`.
## The first warning was:
## i In argument: `across(all_of(names(valid_map)), ~as.numeric(.x))`.
## Caused by warning:
## ! NAs introduced by coercion
## i Run `dplyr::last_dplyr_warnings()` to see the 4 remaining warnings.
```

```
age_clean_selective <- age_clean |>
  select(county, percent_sex_and_age_total_population, percent_sex_and_age_total_population_male, percer
  rename(
    pct_total_pop = percent_sex_and_age_total_population,
    pct_male = percent_sex_and_age_total_population_male,
    pct_female = percent_sex_and_age_total_population_female,
    pct_under5 = percent_sex_and_age_total_population_under_5_years,
    pct_5_9 = percent_sex_and_age_total_population_5_to_9_years,
    pct_10_14 = percent_sex_and_age_total_population_10_to_14_years,
    pct_15_19 = percent_sex_and_age_total_population_15_to_19_years,
    pct_20_24 = percent_sex_and_age_total_population_20_to_24_years,
    pct_25_34 = percent_sex_and_age_total_population_25_to_34_years,
    pct_35_44 = percent_sex_and_age_total_population_35_to_44_years,
    pct_45_54 = percent_sex_and_age_total_population_45_to_54_years,
    pct_55_59 = percent_sex_and_age_total_population_55_to_59_years,
    pct_60_64 = percent_sex_and_age_total_population_60_to_64_years,
    pct_65_74 = percent_sex_and_age_total_population_65_to_74_years,
    pct_75_84 = percent_sex_and_age_total_population_75_to_84_years,
    pct_85plus = percent_sex_and_age_total_population_85_years_and_over
  )

race_clean <- age_clean |>
  select(county, percent_race_total_population_one_race, percent_race_total_population_one_race_white, p
  rename(
    pct_one_race = percent_race_total_population_one_race,
    pct_white = percent_race_total_population_one_race_white,
    pct_black = percent_race_total_population_one_race_black_or_african_american,
    pct_native_american = percent_race_total_population_one_race_american_indian_and_alaska_native,
    pct_asian = percent_race_total_population_one_race_asian,
    pct_pacific_islander = percent_race_total_population_one_race_native_hawaiian_and_other_pacific_isla
    pct_hispanic = percent_hispanic_or_latino_and_race_total_population_hispanic_or_latino_of_any_race
  )
```

##Merge age and race data with existing data

```
ohio_joined_age <- ohio_prop_map_df |>
  left_join(
    age_clean_selective,
    by = c("NAME" = "county")
  )

ohio_joined_race <- ohio_joined_age |>
  left_join(
    race_clean,
    by = c("NAME" = "county")
  )
```

```r
library(tidyverse)
library(stringr)

edu_raw <- read_csv("EduAttain.csv", col_names = FALSE, show_col_types = FALSE)

new_names <- edu_raw %>%
  slice(2) %>%
  unlist() %>%
  as.character()

edu <- edu_raw %>%
  slice(-c(1, 2))     # remove first two rows
colnames(edu) <- new_names

nm <- names(edu)
nm <- replace_na(nm, "")

estimate_cols2 <- nm[str_detect(nm, regex("Estimate!!Total!!", ignore_case = TRUE))]

id_candidates <- c("Geographic Area Name")
id_cols2 <- intersect(id_candidates, nm)

edu_estimate <- edu %>%
  select(any_of(id_cols2), all_of(estimate_cols2))

nm2 <- names(edu_estimate)
nm2 <- replace_na(nm2, "")

estimate_cols3 <- nm2[str_detect(nm2, regex("AGE BY EDUCATIONAL ATTAINMENT", ignore_case = TRUE))]

edu_estimate2 <- edu_estimate %>%
  select(any_of(id_cols2), all_of(estimate_cols3))

# 4) Clean county name; optionally keep only county rows
edu_estimate2 <- edu_estimate2 %>%
  mutate(
    `Geographic Area Name` = str_squish(`Geographic Area Name`),
    county = `Geographic Area Name` %>%
      str_replace(",\\s*Ohio$", "") %>%
      str_replace("\\s*County$", "") %>%
      str_squish() %>%
      str_to_title()
  ) %>%
  relocate(county, .after = `Geographic Area Name`)

edu_totals <- edu_estimate2 %>%
  mutate(across(
    contains("Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT"),
    ~ readr::parse_number(as.character(.x))
  )) %>%
  mutate(
    pop_18_24      = `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 18 to 24 years`,
    pop_25plus     = `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over`,
```

```
    pop_18plus     = pop_18_24 + pop_25plus,

    # Less than high school (18-24 "less than HS") + (25+ "less than 9th" + "9-12 no diploma")
    lt_hs_18plus   = `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 18 to 24 years!!Less th
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Less
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!9th

    # High school graduate (18-24 HS grad) + (25+ HS grad)
    hs_grad_18plus = `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 18 to 24 years!!High sc
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!High

    # Some college or associate's (18-24 some college/AA) + (25+ some college no degree + AA)
    some_coll_18plus =
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 18 to 24 years!!Some co
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Some
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Asso

    # Bachelor's or higher (18-24 BA+) + (25+ BA + Grad/Prof)
    ba_plus_18plus = `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 18 to 24 years!!Bachelo
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Bach
                     `Estimate!!Total!!AGE BY EDUCATIONAL ATTAINMENT!!Population 25 years and over!!Grad


  # 4) keep a clean output set with better labels
  select(
    `Geographic Area Name`, county,
    pop_18_24, pop_25plus, pop_18plus,
    lt_hs_18plus, hs_grad_18plus, some_coll_18plus, ba_plus_18plus
  )
```

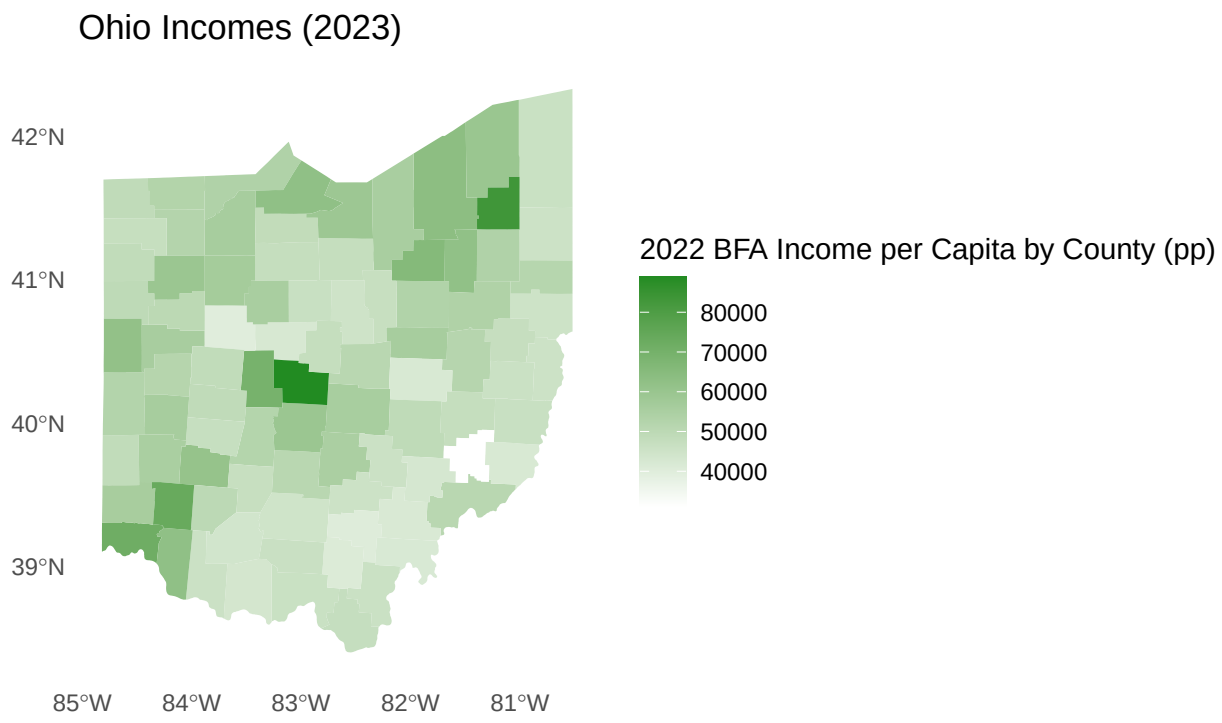
```
library(sf)
library(tidyverse)

inc_raw <- read_csv("Incomes_2.csv", show_col_types = FALSE) %>%
  as_tibble()

ohio_joined_race <- ohio_joined_race %>%
  left_join(inc_raw, by = "NAME")

class(ohio_joined_race)
```

```
## [1] "sf"         "data.frame"
```

```
ohio_joined_race <- ohio_joined_race %>%
  mutate(`2022_BFA_Percap_Income` = as.numeric(`2022_BFA_Percap_Income`))
```

```
ggplot(ohio_joined_race) +
geom_sf(aes(fill = `2022_BFA_Percap_Income`), color = NA) + scale_fill_gradient2(
    low = "white",
    high = "forestgreen",
    midpoint = min(ohio_joined_race$`2022_BFA_Percap_Income`, na.rm = TRUE),
    name = "2022 BFA Income per Capita by County (pp)"
```

```
)+ labs(
title = "Ohio Incomes (2023)" )+
theme_minimal() + theme(
panel.grid = element_blank(),
    legend.position = "right"
  )
```



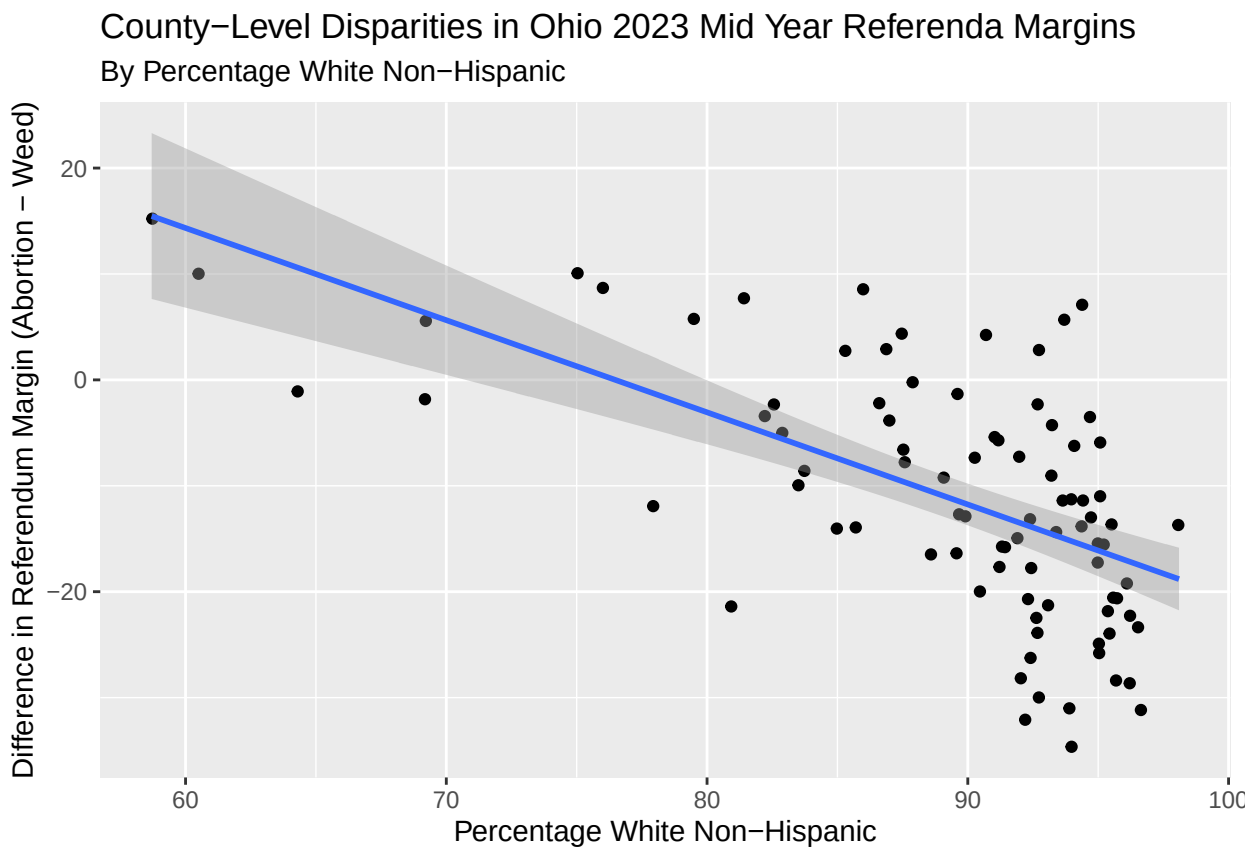

```
#Our first mode: Regressing the margin difference between abortion and weed on some of our covariates:

diff_model <- lm(diff_margin ~ voter_turnout + pct_female + pct_white + `2022_BFA_Percap_Income`, data =

summary(diff_model)
```

```
##
## Call:
## lm(formula = diff_margin ~ voter_turnout + pct_female + pct_white +
##     `2022_BFA_Percap_Income`, data = ohio_joined_race)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -23.7003  -5.2420   0.6845   5.5716  23.3134
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)               2.463e+01  4.041e+01   0.610   0.5438
```

```
## voter_turnout             7.891e-01  3.105e-01   2.541   0.0129 *
## pct_female                1.202e-01  7.070e-01   0.170   0.8654
## pct_white                -9.025e-01  1.429e-01  -6.317 1.26e-08 ***
## `2022_BFA_Percap_Income` -1.624e-05  1.701e-04  -0.096   0.9241
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.764 on 83 degrees of freedom
## Multiple R-squared:  0.4468, Adjusted R-squared:  0.4202
## F-statistic: 16.76 on 4 and 83 DF,  p-value: 4.163e-10
```

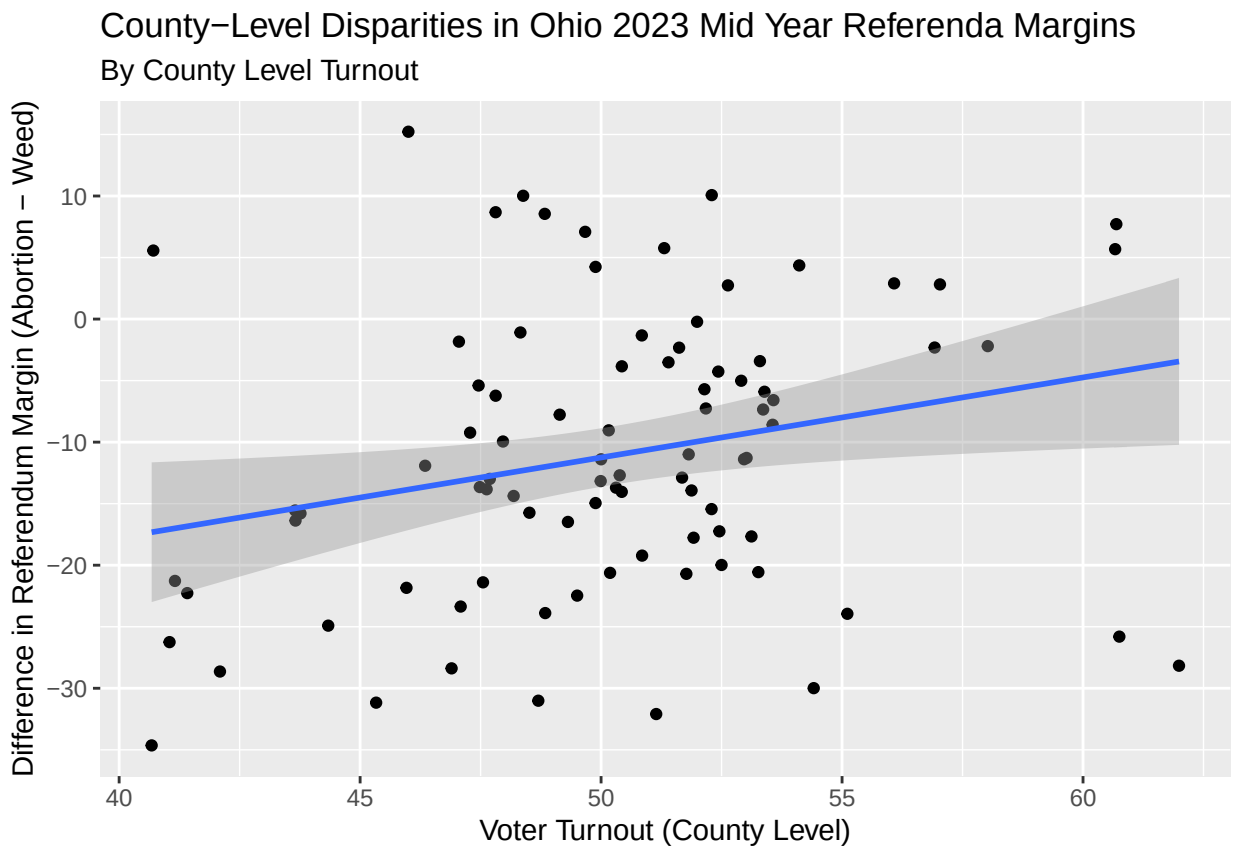
```
disparity_plot <- ggplot(data = ohio_joined_race, aes(x = pct_white, y = diff_margin)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    x = "Percentage White Non-Hispanic",
    y = "Difference in Referendum Margin (Abortion - Weed)",
    title = "County-Level Disparities in Ohio 2023 Mid Year Referenda Margins",
    subtitle = "By Percentage White Non-Hispanic"
  )

disparity_plot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



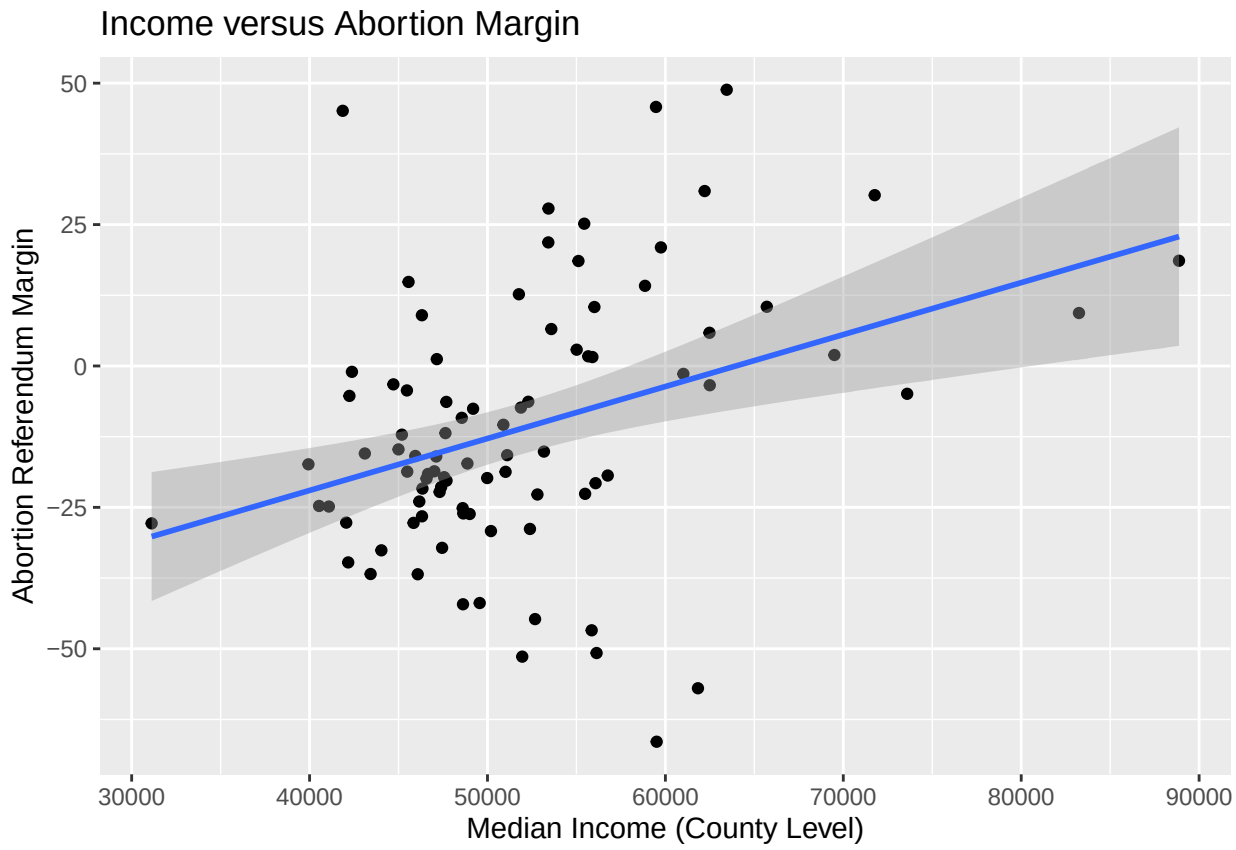
```
turnout_disparity_plot <- ggplot(data = ohio_joined_race, aes(x = voter_turnout, y = diff_margin)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    x = "Voter Turnout (County Level)",
    y = "Difference in Referendum Margin (Abortion - Weed)",
    title = "County-Level Disparities in Ohio 2023 Mid Year Referenda Margins",
    subtitle = "By County Level Turnout"
  )

turnout_disparity_plot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
abortion_income_plot <- ggplot(data = ohio_joined_race, aes(x = `2022_BFA_Percap_Income`, y = abortion_m
  geom_jitter() +
  geom_smooth(method = "lm", se = TRUE) +
  labs(
    x = "Median Income (County Level)",
    y = "Abortion Referendum Margin",
    title = "Income versus Abortion Margin"
  )

abortion_income_plot
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
#population-weighted regression -- trying to fix some of our ecological inference issues

model_weighted <- lm(diff_margin ~ voter_turnout + pct_female + pct_white + `2022_BFA_Percap_Income`,
                     data = ohio_joined_race,
                     weights = `Registered Voters`)

summary(model_weighted)
```

```
##
## Call:
## lm(formula = diff_margin ~ voter_turnout + pct_female + pct_white +
##     `2022_BFA_Percap_Income`, data = ohio_joined_race, weights = `Registered Voters`)
##
## Weighted Residuals:
##     Min      1Q  Median      3Q     Max
## -5677.5 -1628.4  -309.8  1274.5  5249.4
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)              -95.927771  57.145614  -1.679   0.0970 .
## voter_turnout              1.371576   0.305838   4.485 2.32e-05 ***
## pct_female                 2.112326   1.025269   2.060   0.0425 *
## pct_white                 -0.825584   0.103463  -7.980 7.15e-12 ***
```

```
## `2022_BFA_Percap_Income`  -0.000273   0.000134  -2.037   0.0449 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2220 on 83 degrees of freedom
## Multiple R-squared:  0.6217, Adjusted R-squared:  0.6035
## F-statistic: 34.11 on 4 and 83 DF,  p-value: < 2.2e-16
```