Geospatial Mapping

Libraries/Packages

How many counties have multiple school districts?

```
county_district_count <- school_districts |>
  group_by(`State FIPS`,`County FIPS`) |>
  count() |>
  group_by(n) |>
  count()
```

```
Storing counts in `nn`, as `n` already present in input
i Use `name = "new_name"` to pick a new name.
```

```
district_count_plot <- ggplot() +
  geom_point(county_district_count, mapping = aes(x = n, y = nn))
```

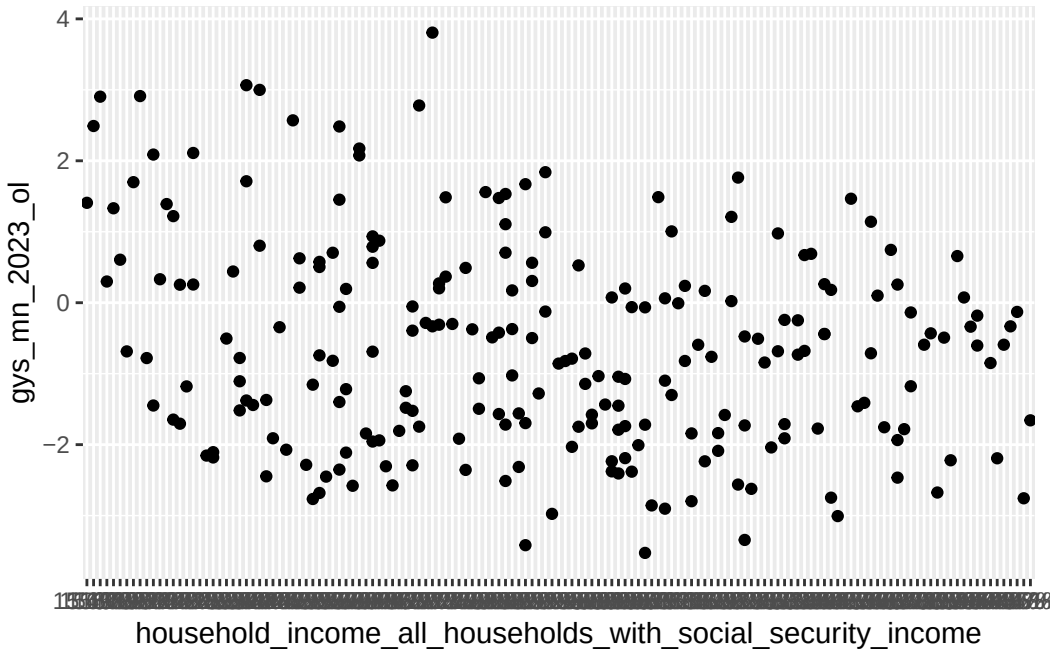
```
#Data Cleaning to Fix Column Names and Allignment
#Only 737 school districts in this dataset, the education data has 7390 school districts

income_district_cleaned <- income_district |>
  clean_names()

income_district_estimates <- income_district_cleaned |>
  rename(
    school_district = label_grouping
  ) |>
  mutate(
    SchoolDistrict = lag(school_district, n = 2, default = NULL)
  ) |>
  filter(
```

```
    grepl("District", school_district)
  )

income_district_percentages <- income_district_cleaned |>
  rename(
    school_district = label_grouping
  ) |>
  mutate(
    school_district = lag(school_district, n = 5, default = NULL)
  ) |>
  filter(
    grepl("District", school_district)
  )
```

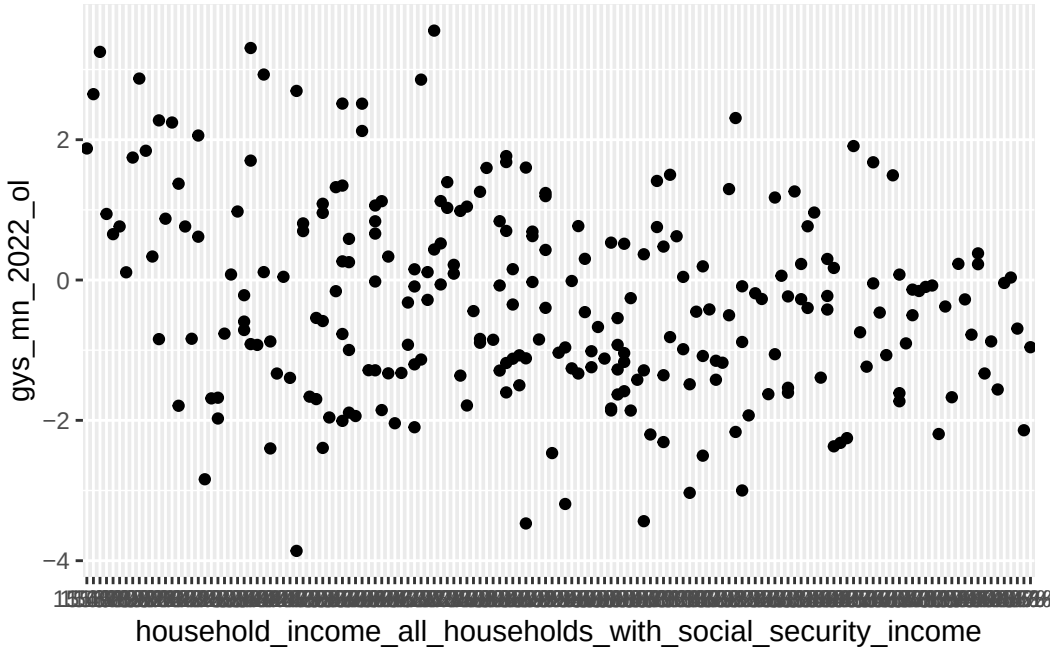
```
#Percentage of people that receive social security income


ssi_percent_district <- income_district_percentages |>
  mutate(
    state = word(school_district, -1),
    school_district = word(school_district, start = 1, end = -4),
    state = state.abb[match(state,state.name)]
  )

write.csv(ssi_percent_district, "data/acs_income_2021.csv")


ssi_percent_joined <- ssi_percent_district |>
  left_join(
    test_admindist_gys,
    by = join_by(school_district == sedaadminname, state == stateabb)
  )


# After filtering, only 248 school districts
ssi_math <- ssi_percent_joined |>
  filter(
    subject == "mth",
    subgroup == "all"
  ) |>
  mutate(
```

```
    ol_diff_2022_2023 = gys_mn_2023_ol - gys_mn_2022_ol,
    ol_diff_2019_2022 = gys_mn_2022_ol - gys_mn_2019_ol,
  )

ssi_rla <- ssi_percent_joined |>
  filter(
    subject == "rla",
    subgroup == "all"
  ) |>
  mutate(
    ol_diff_2022_2023 = gys_mn_2023_ol - gys_mn_2022_ol,
    ol_diff_2019_2022 = gys_mn_2022_ol - gys_mn_2019_ol,
  )

ssi_math |>
  ggplot() +
  geom_point(
    aes(
      x = household_income_all_households_with_social_security_income,
      y = gys_mn_2023_ol)
    )
```

```
Warning: Removed 17 rows containing missing values (`geom_point()`).
```



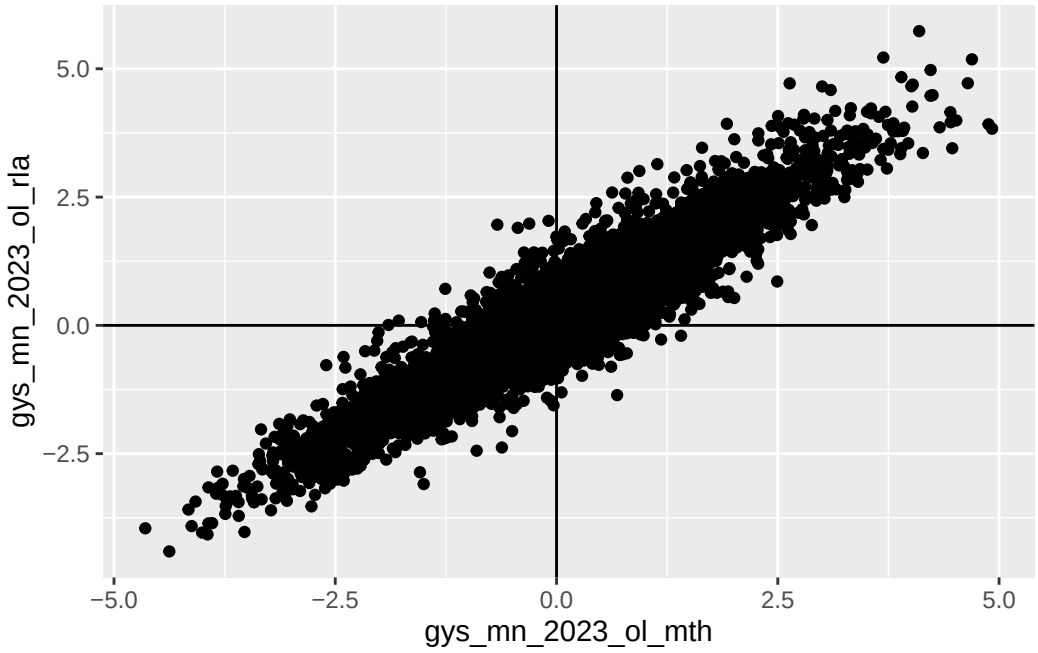

```
ssi_rla |>
  ggplot() +
  geom_point(
    aes(
      x = household_income_all_households_with_social_security_income,
      y = gys_mn_2022_ol)
    )
```




```
test_selected_years <- test_admindist_gys |>
  select(
    sedaadmin, sedaadminname, stateabb, subject, subgroup,
    gys_mn_2019_ol, gys_mn_2022_ol, gys_mn_2023_ol
  )

test_wider <- test_selected_years |>
  pivot_wider(
    names_from = subject,
    values_from = c(gys_mn_2019_ol, gys_mn_2022_ol, gys_mn_2023_ol)
  )


ggplot(test_wider |>  filter(subgroup == "all")) +
  geom_point(
    aes(
      x = gys_mn_2023_ol_mth,
      y = gys_mn_2023_ol_rla,
    )
  ) +
  geom_hline(yintercept = 0) +
```

```
  geom_vline(xintercept = 0)
```

```
Warning: Removed 2498 rows containing missing values (`geom_point()`).
```




```
write.csv(test_wider, "data/test_admindist_wider.csv")
```

Todo (4/19/2024)

- Potentially pandemic affected survey coverage
  - 2019 may have most coverage
  - 2022 would be post pandemic
- ACS data from one year is usually generalizable to other years
- Shiny App Development
  - Figure out what other variables to filter/customize by
  - Drop-down menu variables
    - Make all the labels human readable (not default)

        * Make choices also affect axis + plot titles.
        * Keep the configuration menu on the side and affect all pages
        * Different ACS variables on the x axis
        * Math or Reading Scores
        * Others?
        * Keep subgroup (demphasize)
    - Considerations about plots
        * Add in variable correlation
        * Fix the y-axis scale
        * Make it easier to distinguish above and below zero (improvement/non-improvement)
        * For dynamic axis/plot titles
            · Try to automate label creation (replace underscores with , capitalize every word)
        * Subtitle: For _____ students in _____ districts
        * Side: Consider making x-axis static within a variable
        * Side: Interactivity with points showing popups
    - About Page
        * Explain the project, where the data comes from, its coverage, etc.
        * Concise version of the report
        * Coverage Map
            · State map colored by count of school district
        * Explain limitations and issues with datasets and plots

Todo (4/26/24)

App

- Look into Expansion with other variables or datasets
- Make all labels human readable
    - Make them more intuitive

- Interactivity?
- About page
- Use the fixed color scale for all plots going forward
- Plots
  - Make the points bigger
  - Point outlining
  - Trend Line
    - Option to toggle it
    - Change trend line color?
  - Potentially plot year to year (2022 to 2023) to see more variation in Math vs Reading
- Sidebar Menu Expansion Options
  - Possibility to add a new drop down or exxpand the variable menu
  - Make menus hide if they don't apply to the current tab
- Download button on the data page

Writeup:

- Project Report
  - Summarizing reading and research
  - Lit review/background
  - Dataset descriptions
    - Academic performance data
      - Explain their grade calculation
    - ACS Data
  - Methodology
    - Data prep + Data cleaning
    - Consideration for district converage
  - Results
    - Description of the app

        * Takeaways + Conclusions
    - Discussion/Conclusion
        * Overall takeways answering the research question
            · The state of education post-pandemic
        * LImitations of my analysis/research
        * Future research steps
- Reflection on the project and experience
- Citing code/libraries
    - Only big ones like tidyverse and shiny

Future?

- Competitions, contests, etc.
- Ask Mine