California School District Gender Achievement Gaps (SEDA 2018)

A reproducible exploratory analysis of reading and math gaps and their socioeconomic correlates

Aryan Bhojani

```
library(tidyverse)
library(ggplot2)
library(colorspace)
library(scales)
```

Project overview

This project uses the Stanford Education Data Archive (SEDA) to study gender achievement gaps across California school districts in 2018. I focus on two subjects:

- Reading (RLA)
- Math (MTH)

For each district and grade, the dataset provides standardized mean test score estimates. I compute a district-level gender gap as:

Gender gap = male mean - female mean

So: - Positive gap means boys score higher on average. - Negative gap means girls score higher on average.

I then examine how these gaps relate to district socioeconomic measures (income, poverty, unemployment, SNAP, and a composite SES index), and how patterns vary by grade and locale.

Data source

SEDA is a public database that standardizes test score estimates across districts, grades, and years. Scores are reported in standard deviation units relative to national norms, enabling consistent comparisons.

Import data

```
main_path <- file.path(params$data_dir, params$main_file)
cov_path  <- file.path(params$data_dir, params$cov_file)
check_path <- file.path(params$data_dir, params$check_file)

ca_main <- read_csv(main_path, show_col_types = FALSE)
ca_cov  <- read_csv(cov_path,  show_col_types = FALSE)

ca_main |> head(5)
```

```
# A tibble: 5 x 59
  sedalea grade stateabb sedaleaname    subject cs_mn_all cs_mnse_all totgyb_all
    <dbl> <dbl> <chr>    <chr>          <chr>       <dbl>       <dbl>      <dbl>
1  600001     4 CA       ACTON-AGUA DU~ mth      -3.67e-1      0.109         86
2  600001     4 CA       ACTON-AGUA DU~ rla       5.69e-3      0.117         85
3  600001     6 CA       ACTON-AGUA DU~ rla      -4.02e-5      0.0922       114
4  600001     8 CA       ACTON-AGUA DU~ mth      -9.77e-2      0.103         98
5  600001     8 CA       ACTON-AGUA DU~ rla      -4.86e-1      0.111         99
# i 51 more variables: cs_mn_asn <dbl>, cs_mnse_asn <dbl>, totgyb_asn <dbl>,
#   cs_mn_blk <dbl>, cs_mnse_blk <dbl>, totgyb_blk <dbl>, cs_mn_ecd <dbl>,
#   cs_mnse_ecd <dbl>, totgyb_ecd <dbl>, cs_mn_fem <dbl>, cs_mnse_fem <dbl>,
#   totgyb_fem <dbl>, cs_mn_hsp <dbl>, cs_mnse_hsp <dbl>, totgyb_hsp <dbl>,
#   cs_mn_mal <dbl>, cs_mnse_mal <dbl>, totgyb_mal <dbl>, cs_mn_mfg <dbl>,
#   cs_mnse_mfg <dbl>, totgyb_mfg <dbl>, cs_mn_mtr <dbl>, cs_mnse_mtr <dbl>,
#   totgyb_mtr <dbl>, cs_mn_nam <lgl>, cs_mnse_nam <lgl>, totgyb_nam <dbl>, ...
```

```
ca_cov |> head(5)
```

```
# A tibble: 5 x 60
  sedalea grade sedaleanm  urban suburb  town rural locale  perind perasn perhsp
    <dbl> <dbl> <chr>      <dbl>  <dbl> <dbl> <dbl> <chr>    <dbl>  <dbl>  <dbl>
1  600001     4 ACTON-AGU~     0      0     0     1 Rural~ 0.00389 0.0459  0.308
```

```
2  600001     5 ACTON-AGU~     0      0     0     1 Rural~ 0.00379 0.0467  0.304
3  600001     6 ACTON-AGU~     0      0     0     1 Rural~ 0.00322 0.0437  0.309
4  600001     7 ACTON-AGU~     0      0     0     1 Rural~ 0.00341 0.0441  0.291
5  600001     8 ACTON-AGU~     0      0     0     1 Rural~ 0.00379 0.0453  0.309
# i 49 more variables: perblk <dbl>, perwht <dbl>, perfl <dbl>, perrl <dbl>,
#   perecd <dbl>, perell <dbl>, perspeced <dbl>, totenrl <dbl>, gslo <chr>,
#   gshi <dbl>, hswhtblk <dbl>, hswhthsp <dbl>, hsflnfl <dbl>, hsecdnec <dbl>,
#   rswhtblk <dbl>, rswhthsp <dbl>, rsflnfl <dbl>, rsecdnec <dbl>,
#   perfrl <dbl>, sesall <dbl>, sesblk <dbl>, seshsp <dbl>, seswht <dbl>,
#   lninc50all <dbl>, lninc50blk <dbl>, lninc50hsp <dbl>, lninc50wht <dbl>,
#   baplusall <dbl>, baplusblk <dbl>, baplushsp <dbl>, bapluswht <dbl>, ...
```

Unit of analysis and dataset scope

Each row in the test score table corresponds to one district-grade-subject combination. The covariate table corresponds to one district-grade combination (covariates do not vary by subject in these files).

```r
ca_main_units <- ca_main |>
  distinct(sedalea, grade, subject) |>
  nrow()

ca_cov_units <- ca_cov |>
  distinct(sedalea, grade) |>
  nrow()

tibble(
  dataset = c("ca_main (test scores)", "ca_cov (covariates)"),
  unique_units = c(ca_main_units, ca_cov_units)
)
```

```
# A tibble: 2 x 2
  dataset               unique_units
  <chr>                        <int>
1 ca_main (test scores)         7513
2 ca_cov (covariates)           4340
```

Build the analysis dataset

Select variables

```
main_vars <- c(
  "sedalea",
  "sedaleaname",
  "grade",
  "subject",
  "cs_mn_mal",
  "cs_mn_fem"
)

cov_vars <- c(
  "sedalea",
  "grade",
  "locale",
  "sesall",
  "lninc50all",
  "povertyall",
  "unempall",
  "snapall"
)

main_sub <- ca_main |>
  select(all_of(main_vars))

cov_sub <- ca_cov |>
  select(all_of(cov_vars))
```

Merge covariates onto test scores

```
rawdata <- main_sub |>
  left_join(cov_sub, by = c("sedalea", "grade"))

head(rawdata, 4)
```

```
# A tibble: 4 x 12
  sedalea sedaleaname grade subject cs_mn_mal cs_mn_fem locale sesall lninc50all
```

```
    <dbl> <chr>       <dbl> <chr>       <dbl>     <dbl> <chr>   <dbl>      <dbl>
1  600001 ACTON-AGUA~     4 mth        NA        -0.251 Rural~   1.24       11.4
2  600001 ACTON-AGUA~     4 rla        NA         0.226 Rural~   1.24       11.4
3  600001 ACTON-AGUA~     6 rla        -0.211    NA     Rural~   1.24       11.4
4  600001 ACTON-AGUA~     8 mth        -0.363     0.170 Rural~   1.24       11.4
# i 3 more variables: povertyall <dbl>, unempall <dbl>, snapall <dbl>
```

Compute gap, rename columns, reorder

```
name_dict <- c(
  "District ID"           = "sedalea",
  "District"              = "sedaleaname",
  "Locale"                = "locale",
  "Log(Median income)"    = "lninc50all",
  "Poverty rate"          = "povertyall",
  "Unemployment rate"     = "unempall",
  "SNAP rate"             = "snapall",
  "Socioeconomic index"   = "sesall",
  "Grade"                 = "grade",
  "Subject"               = "subject"
)

col_order <- c(
  "District ID", "District", "Locale",
  "Log(Median income)", "Poverty rate", "Unemployment rate", "SNAP rate",
  "Socioeconomic index", "Grade", "Subject", "Gender gap"
)

rawdata_mod1 <- rawdata |>
  mutate(`Gender gap` = cs_mn_mal - cs_mn_fem) |>
  rename(!!!name_dict) |>
  select(all_of(col_order))

head(rawdata_mod1, 4)
```

```
# A tibble: 4 x 11
  `District ID` District              Locale `Log(Median income)` `Poverty rate`
          <dbl> <chr>                 <chr>                 <dbl>          <dbl>
1        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
2        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
```

```
3        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
4        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
# i 6 more variables: `Unemployment rate` <dbl>, `SNAP rate` <dbl>,
#   `Socioeconomic index` <dbl>, Grade <dbl>, Subject <chr>, `Gender gap` <dbl>
```

Pivot to a tidy, wide dataset (Math and Reading in separate columns)

```
seda_data <- rawdata_mod1 |>
  pivot_wider(
    names_from  = Subject,
    values_from = `Gender gap`
  ) |>
  rename(
    Math    = mth,
    Reading = rla
  ) |>
  select(
    `District ID`, District, Locale,
    `Log(Median income)`, `Poverty rate`, `Unemployment rate`, `SNAP rate`,
    `Socioeconomic index`, Grade, Math, Reading
  )

head(seda_data, 4)
```

```
# A tibble: 4 x 11
  `District ID` District             Locale `Log(Median income)` `Poverty rate`
          <dbl> <chr>                <chr>                 <dbl>          <dbl>
1        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
2        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
3        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
4        600006 ROSS VALLEY ELEMENTA~ Subur~                 11.6         0.0414
# i 6 more variables: `Unemployment rate` <dbl>, `SNAP rate` <dbl>,
#   `Socioeconomic index` <dbl>, Grade <dbl>, Math <dbl>, Reading <dbl>
```

Optional structural check against a reference file

```
if (file.exists(check_path)) {
  data_reference <- read_csv(check_path, show_col_types = FALSE)
```

```
  data_reference
}
```

```
# A tibble: 10 x 11
   `District ID` District            Locale `log(Median income)` `Poverty rate`
           <dbl> <chr>               <chr>                 <dbl>          <dbl>
 1        600001 ACTON-AGUA DULCE UN~ Rural~                 11.4         0.0919
 2        600001 ACTON-AGUA DULCE UN~ Rural~                 11.4         0.0919
 3        600001 ACTON-AGUA DULCE UN~ Rural~                 11.4         0.0919
 4        600006 ROSS VALLEY ELEMENT~ Subur~                 11.6         0.0414
 5        600006 ROSS VALLEY ELEMENT~ Subur~                 11.6         0.0414
 6        600006 ROSS VALLEY ELEMENT~ Subur~                 11.6         0.0414
 7        600006 ROSS VALLEY ELEMENT~ Subur~                 11.6         0.0414
 8        600011 FORT SAGE UNIFIED    Rural~                 10.7         0.160
 9        600011 FORT SAGE UNIFIED    Rural~                 10.7         0.160
10        600011 FORT SAGE UNIFIED    Rural~                 10.7         0.160
# i 6 more variables: `Unemployment rate` <dbl>, `SNAP rate` <dbl>,
#   `Socioeconomic index` <dbl>, Grade <dbl>, Math <dbl>, Reading <dbl>
```

Missingness assessment

SEDA does not compute gap estimates for some district-grade-subject combinations due to small sample sizes. I quantify missingness at both the row level and district level.

```
row_missing_props <- seda_data |>
  summarise(
    prop_missing_math    = mean(is.na(Math)),
    prop_missing_reading = mean(is.na(Reading))
  )

district_missing <- seda_data |>
  group_by(`District ID`) |>
  summarise(any_missing = any(is.na(Math) | is.na(Reading)), .groups = "drop")

prop_districts_with_missing <- mean(district_missing$any_missing)

list(
  row_missing_props = row_missing_props,
  prop_districts_with_any_missing = prop_districts_with_missing
)
```

```
$row_missing_props
# A tibble: 1 x 2
  prop_missing_math prop_missing_reading
              <dbl>                <dbl>
1             0.303                0.307

$prop_districts_with_any_missing
[1] 0.516055
```

Exploratory visualization setup

To visualize relationships between gaps and socioeconomic measures, I reshape to a long format with:

- Subject: Math or Reading
- Gap: gender gap value
- Socioeconomic type: which measure (income, poverty, etc.)
- Socioeconomic measure: the numeric value

```
plot_df <- seda_data |>
  pivot_longer(
    cols = c(Math, Reading),
    names_to = "Subject",
    values_to = "Gap"
  ) |>
  pivot_longer(
    cols = c(`Log(Median income)`, `Poverty rate`, `Unemployment rate`, `SNAP rate`, `Socioe
    names_to = "Socioeconomic type",
    values_to = "Socioeconomic measure"
  )

head(plot_df)
```

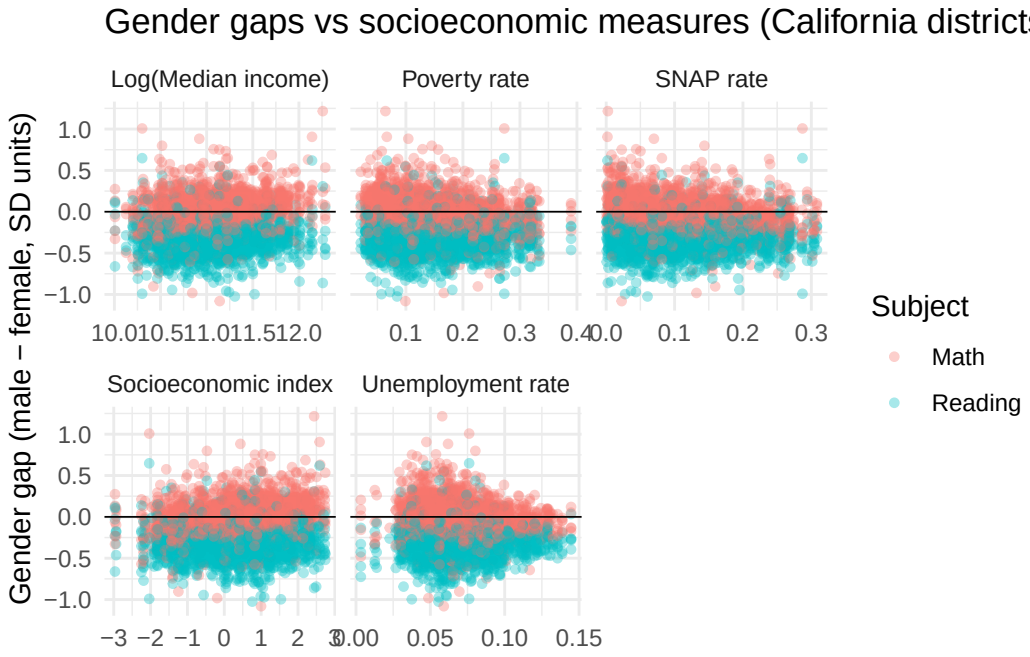
```
# A tibble: 6 x 8
  `District ID` District         Locale Grade Subject   Gap `Socioeconomic type`
          <dbl> <chr>            <chr>  <dbl> <chr>   <dbl> <chr>
1        600001 ACTON-AGUA DULC~ Rural~     4 Math       NA Log(Median income)
2        600001 ACTON-AGUA DULC~ Rural~     4 Math       NA Poverty rate
3        600001 ACTON-AGUA DULC~ Rural~     4 Math       NA Unemployment rate
4        600001 ACTON-AGUA DULC~ Rural~     4 Math       NA SNAP rate
5        600001 ACTON-AGUA DULC~ Rural~     4 Math       NA Socioeconomic index
```

```
6        600001 ACTON-AGUA DULC~ Rural~     4 Reading    NA Log(Median income)
# i 1 more variable: `Socioeconomic measure` <dbl>
```

Results: gaps vs socioeconomic measures

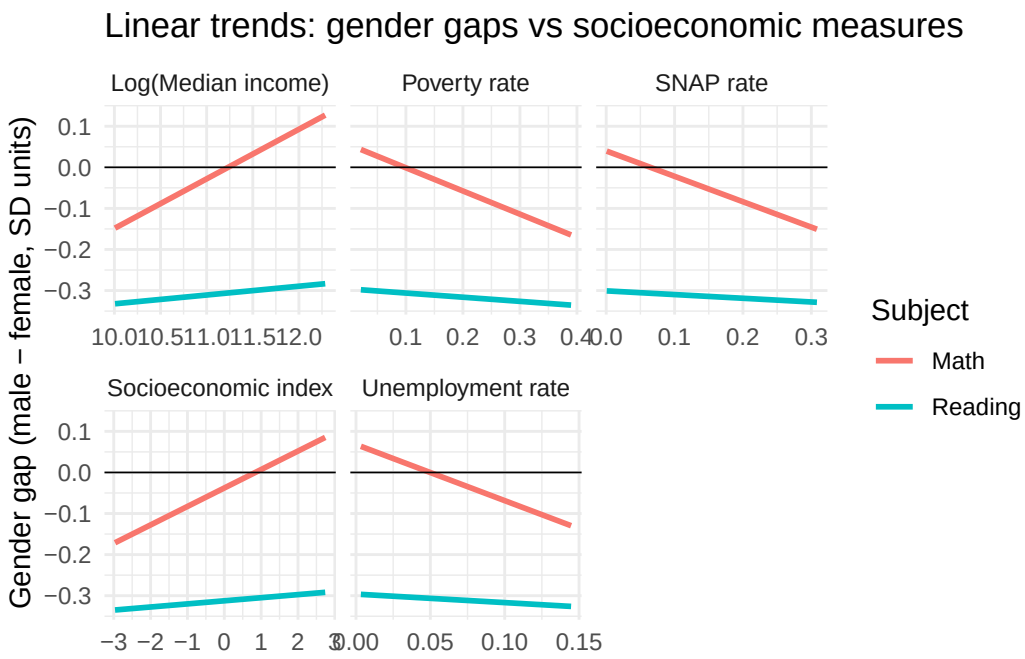
Scatter panels

```
fig_scatter <- ggplot(
  plot_df,
  aes(x = `Socioeconomic measure`, y = Gap, col = Subject)
) +
  geom_point(alpha = 0.35, size = 1.2) +
  facet_wrap(~ `Socioeconomic type`, scales = "free_x") +
  geom_hline(yintercept = 0, linewidth = 0.3) +
  labs(
    x = NULL,
    y = "Gender gap (male - female, SD units)",
    col = "Subject",
    title = "Gender gaps vs socioeconomic measures (California districts, 2018)"
  ) +
  theme_minimal()

fig_scatter
```



Linear trend panels

```
fig_lm <- ggplot(
  plot_df,
  aes(x = `Socioeconomic measure`, y = Gap, col = Subject)
) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 1) +
  facet_wrap(~ `Socioeconomic type`, scales = "free_x") +
  geom_hline(yintercept = 0, linewidth = 0.3) +
  labs(
    x = NULL,
    y = "Gender gap (male - female, SD units)",
    col = "Subject",
    title = "Linear trends: gender gaps vs socioeconomic measures"
  ) +
  theme_minimal()

fig_lm
```

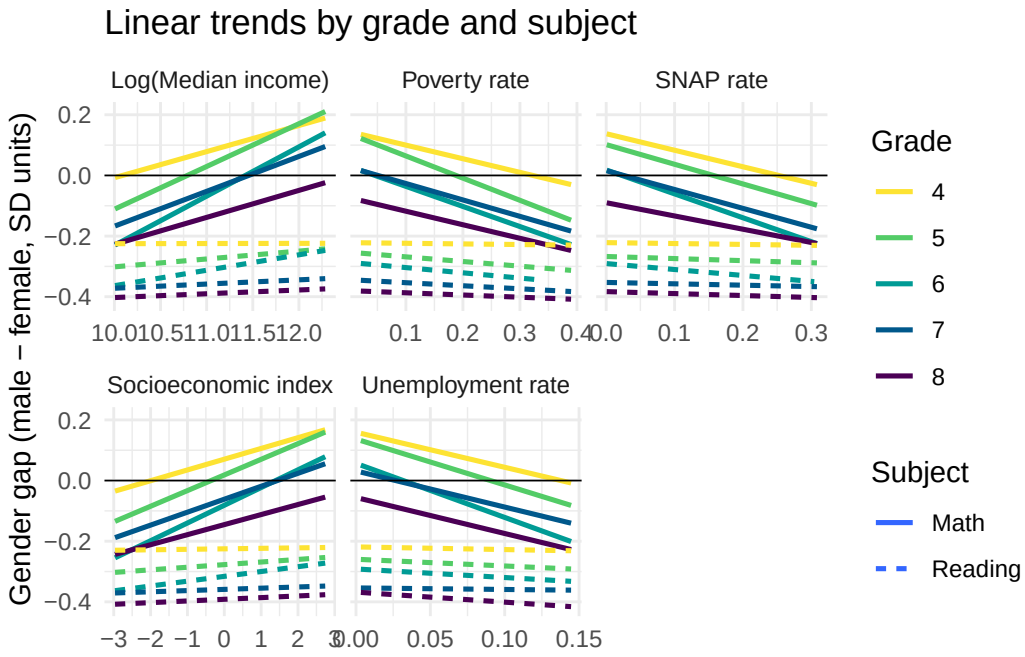


Results: relationships by grade

Here I examine whether the socioeconomic patterns look similar within each grade, and whether gaps vary in magnitude by grade.

```
fig_grade <- ggplot(
  plot_df,
  aes(
    x = `Socioeconomic measure`,
    y = Gap,
    linetype = Subject,
    col = as.factor(Grade)
  )
) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 0.9) +
  facet_wrap(~ `Socioeconomic type`, scales = "free_x") +
  geom_hline(yintercept = 0, linewidth = 0.3) +
  colorspace::scale_color_discrete_sequential(palette = "Viridis") +
  labs(
    x = NULL,
    y = "Gender gap (male - female, SD units)",
```

```
    linetype = "Subject",
    col = "Grade",
    title = "Linear trends by grade and subject"
  ) +
  theme_minimal()

fig_grade
```



Results: locale and math gaps

I create a coarser locale label and then compare math gaps across locales.

```
locale_levels <- sort(unique(plot_df$Locale))
locale_levels
```

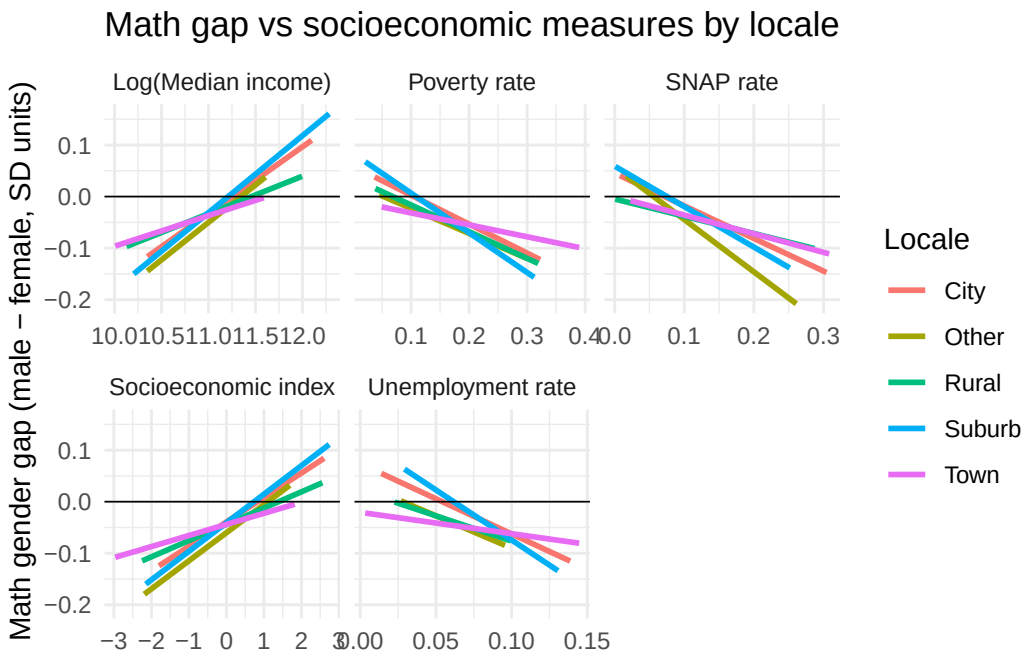
```
 [1] "City, Large"     "City, Midsize"   "City, Small"     "Rural, Distant"
 [5] "Rural, Fringe"   "Rural, Remote"   "Suburb, Large"   "Suburb, Midsize"
 [9] "Sururb, Small"   "Town, Distant"   "Town, Fringe"    "Town, Remote"
```

```r
plot_df <- plot_df |>
  mutate(
    locale_coarse = case_when(
      str_detect(Locale, regex("^Urban|^City", ignore_case = TRUE)) ~ "City",
      str_detect(Locale, regex("^Subur", ignore_case = TRUE))       ~ "Suburb",
      str_detect(Locale, regex("^Town",  ignore_case = TRUE))       ~ "Town",
      str_detect(Locale, regex("^Rural", ignore_case = TRUE))       ~ "Rural",
      TRUE                                                         ~ "Other"
    )
  )

plot_df |>
  count(locale_coarse, Locale) |>
  arrange(locale_coarse, desc(n))
```

```
# A tibble: 13 x 3
   locale_coarse Locale              n
   <chr>         <chr>           <int>
 1 City          City, Large      3040
 2 City          City, Midsize    2620
 3 City          City, Small      1980
 4 Other         Sururb, Small    1200
 5 Other         <NA>              100
 6 Rural         Rural, Fringe    5130
 7 Rural         Rural, Distant   4780
 8 Rural         Rural, Remote    1990
 9 Suburb        Suburb, Large    9610
10 Suburb        Suburb, Midsize  1570
11 Town          Town, Distant    2680
12 Town          Town, Fringe     2470
13 Town          Town, Remote     1230
```

```r
math_locale <- plot_df |>
  filter(Subject == "Math")

fig_locale <- ggplot(
  math_locale,
  aes(x = `Socioeconomic measure`, y = Gap, col = locale_coarse)
) +
  geom_smooth(method = "lm", se = FALSE, linewidth = 1) +
  facet_wrap(~ `Socioeconomic type`, scales = "free_x") +
```

```
  geom_hline(yintercept = 0, linewidth = 0.3) +
  labs(
    x = NULL,
    y = "Math gender gap (male - female, SD units)",
    col = "Locale",
    title = "Math gap vs socioeconomic measures by locale"
  ) +
  theme_minimal()

fig_locale
```



Aggregation: district-level averages across grades

I average outcomes and socioeconomic variables across grades within a district to create a district-level summary dataset.

```
seda_data_agg <- seda_data |>
  group_by(`District ID`, District, Locale) |>
  summarise(
    across(
      c(`Log(Median income)`, `Poverty rate`, `Unemployment rate`,
```

```
        `SNAP rate`, `Socioeconomic index`, Math, Reading),
      ~ mean(.x, na.rm = TRUE)
    ),
    .groups = "drop"
  )

seda_data_agg |> head(5)
```

```
# A tibble: 5 x 10
  `District ID` District             Locale `Log(Median income)` `Poverty rate`
          <dbl> <chr>                <chr>                 <dbl>          <dbl>
1        600001 ACTON-AGUA DULCE UNI~ Rural~                 11.4         0.0919
2        600006 ROSS VALLEY ELEMENTA~ Subur~                 11.6         0.0414
3        600011 FORT SAGE UNIFIED     Rural~                 10.7         0.160
4        600012 TWIN RIDGES ELEMENTA~ Rural~                 10.6         0.179
5        600013 ROCKLIN UNIFIED       Subur~                 11.4         0.0603
# i 5 more variables: `Unemployment rate` <dbl>, `SNAP rate` <dbl>,
#   `Socioeconomic index` <dbl>, Math <dbl>, Reading <dbl>
```

Income brackets

I convert log median income back to dollars (exp) and create 8 contiguous income brackets.

```
seda_data_agg <- seda_data_agg |>
  drop_na(Math, Reading, `Log(Median income)`) |>
  mutate(
    income_dollars = exp(`Log(Median income)`),
    `Income bracket` = cut(income_dollars, breaks = 8)
  )

seda_data_agg |>
  count(`Income bracket`) |>
  arrange(`Income bracket`)
```

```
# A tibble: 8 x 2
  `Income bracket`        n
  <fct>               <int>
1 (2.2e+04,4.65e+04]    149
2 (4.65e+04,7.07e+04]   244
3 (7.07e+04,9.5e+04]    128
```

```
4 (9.5e+04,1.19e+05]     52
5 (1.19e+05,1.44e+05]    16
6 (1.44e+05,1.68e+05]     9
7 (1.68e+05,1.92e+05]     4
8 (1.92e+05,2.17e+05]     2
```

Summary table by income bracket

```
income_bracket_summary <- seda_data_agg |>
  group_by(`Income bracket`) |>
  summarise(
    across(
      c(Math, Reading, `Poverty rate`, `Unemployment rate`, `SNAP rate`, `Socioeconomic index
      ~ mean(.x, na.rm = TRUE)
    ),
    n_districts = n(),
    .groups = "drop"
  ) |>
  arrange(`Income bracket`)

income_bracket_summary
```

```
# A tibble: 8 x 8
  `Income bracket`        Math Reading `Poverty rate` `Unemployment rate`
  <fct>                  <dbl>   <dbl>          <dbl>               <dbl>
1 (2.2e+04,4.65e+04]  -0.0711   -0.319         0.223               0.0833
2 (4.65e+04,7.07e+04] -0.0408   -0.312         0.139               0.0674
3 (7.07e+04,9.5e+04]   0.00116  -0.306         0.0868              0.0536
4 (9.5e+04,1.19e+05]   0.0514   -0.291         0.0634              0.0468
5 (1.19e+05,1.44e+05]  0.0885   -0.285         0.0475              0.0424
6 (1.44e+05,1.68e+05]  0.0998   -0.336         0.0439              0.0424
7 (1.68e+05,1.92e+05]  0.203    -0.214         0.0406              0.0401
8 (1.92e+05,2.17e+05]  0.333    -0.307         0.0471              0.0541
# i 3 more variables: `SNAP rate` <dbl>, `Socioeconomic index` <dbl>,
#   n_districts <int>
```

Proportion of districts with notable math gaps

Here I define a "notable math gap favoring boys" as Math > 0.1 SD (male - female) after averaging across grades.

```r
income_bracket_boys_favored <- seda_data_agg |>
  mutate(math_gap_favoring_boys = Math > 0.1) |>
  group_by(`Income bracket`) |>
  summarise(
    prop_boys_favored = mean(math_gap_favoring_boys, na.rm = TRUE),
    n_districts = n(),
    .groups = "drop"
  ) |>
  arrange(`Income bracket`)

income_bracket_boys_favored
```

```
# A tibble: 8 x 3
  `Income bracket`    prop_boys_favored n_districts
  <fct>                           <dbl>       <int>
1 (2.2e+04,4.65e+04]             0.0604         149
2 (4.65e+04,7.07e+04]            0.0697         244
3 (7.07e+04,9.5e+04]             0.109          128
4 (9.5e+04,1.19e+05]             0.25            52
5 (1.19e+05,1.44e+05]            0.5             16
6 (1.44e+05,1.68e+05]            0.556            9
7 (1.68e+05,1.92e+05]            0.5              4
8 (1.92e+05,2.17e+05]            1                2
```

Discussion and interpretation (ASCII-friendly)

This section prints a concise narrative discussion using computed summaries, without inline code.

Key takeaways

- Across districts (averaged across grades), the average Reading gap is -0.31 SD (male - female). Negative values indicate a girl advantage.
- Across districts, the average Math gap is -0.023 SD (male - female). This is typically much closer to zero.

A useful way to compare magnitude is the average absolute gap. The mean absolute Reading gap is 0.31 SD versus 0.087 SD for Math. This supports the pattern that Reading gaps are larger and more consistent, while Math gaps are smaller and often near zero.

Socioeconomic patterns

In the socioeconomic panels, Reading tends to show a clearer and more systematic relationship with district socioeconomic conditions than Math. Reading gaps more often shift away from zero across the socioeconomic range, while Math gaps cluster tightly around zero with only modest variation.

Grade patterns

When separating trends by grade, the overall subject contrast persists within grades: Reading gaps favor girls at most grades, while Math remains closer to zero. In this dataset, the grade with the largest average Reading gap magnitude is grade 8 (mean absolute Reading gap about 0.396 SD).

Locale patterns (Math focus)

When focusing on Math and comparing fitted lines by locale, average differences across locales are small. Any locale differences should be interpreted cautiously because the effect sizes are generally close to zero relative to the Reading patterns.

Missing data note

Gap estimates are missing for some district-grade-subject combinations, typically due to small sample sizes. Dropping every district with any missing value can bias the analysis toward larger districts and change the socioeconomic mix of the sample. In this project, I keep available observations and report missingness explicitly.

Summary

Overall, the descriptive patterns align with prior SEDA-based findings: a robust Reading gap favoring girls and no strong statewide Math gap on average, alongside local variation that correlates with socioeconomic context. These results are exploratory and describe associations, not causal effects.