Data behind criticisms against Chicago's School Quality Rating Policy (SQRP)

Aulia Larasati

6/1/2021

```
# SQRP ratings in 2018-2020
sqrp_rating <- read_excel("sqrp_rating_2018-2020.xlsx",
    sheet = "long_data")  %>% rename(school_year = schoo_year)
```

```
# racial composition in CPS 2019-2020
race <- read_excel("race_ethnic_2019-2020.xlsx")
```

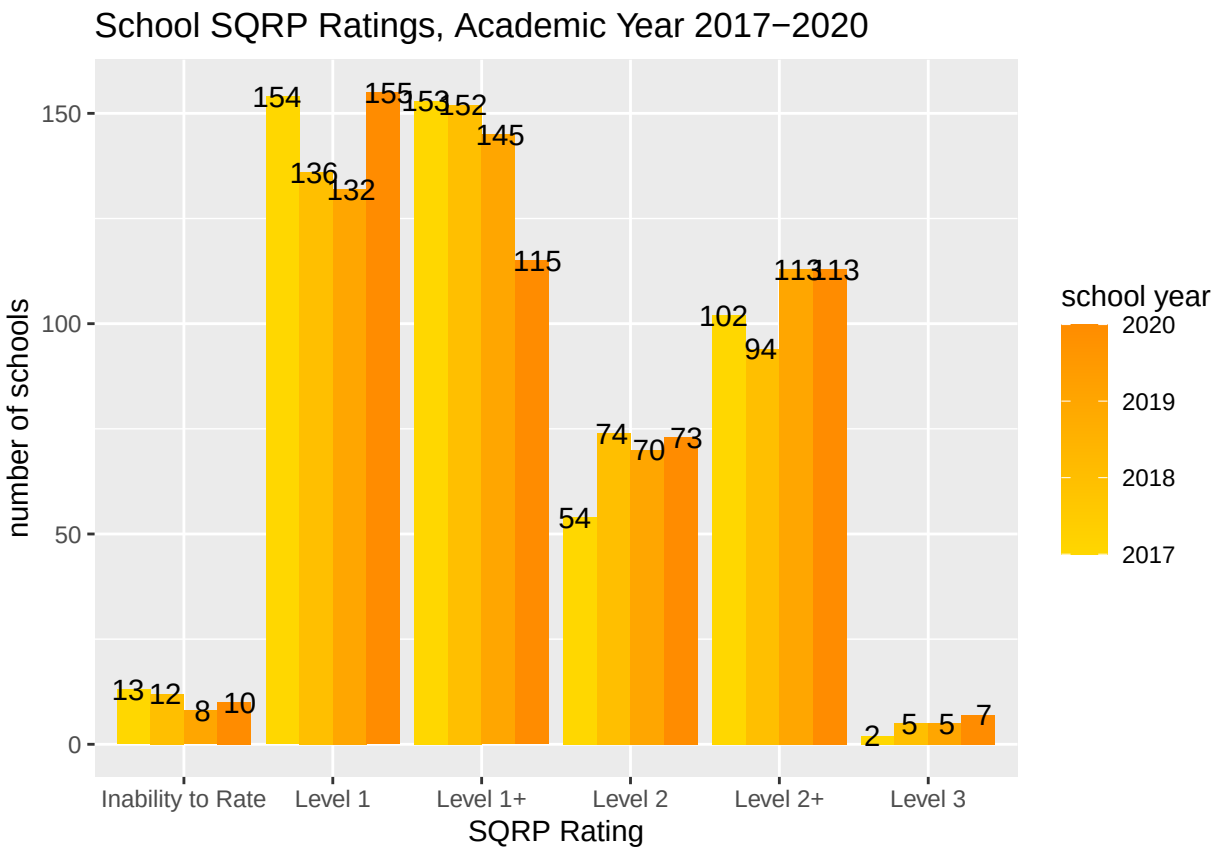
School Accountability Status and Rating 2017-2020

Measuring and reporting a school's quality to parents provides the ability for parents to choose the best school to meet their child's needs and advocate for improvements. The coexistence between school rating and choice provides the incentives for schools to improve. Based on 2017-2020 data, most elementary schools under Chicago Public Schools are in good standing. The number of schools that need intensive support has also decreased.

```
rating_graph <- sqrp_rating %>%
  group_by(school_year, rating) %>%
  summarize(n_school = n()) %>%
  ggplot(aes(x = rating, y = n_school, group = school_year)) +
  geom_col(aes(fill= school_year), position = "dodge") +
  geom_text(aes(x = rating, label = n_school),
            position = position_dodge(width = 1), size = 4) +
  scale_fill_continuous(low = "gold", high = "darkorange") +
  labs(title = "School SQRP Ratings, Academic Year 2017-2020",
       x = "SQRP Rating", fill = "school year", y = "number of schools")
```

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

```
rating_graph
```

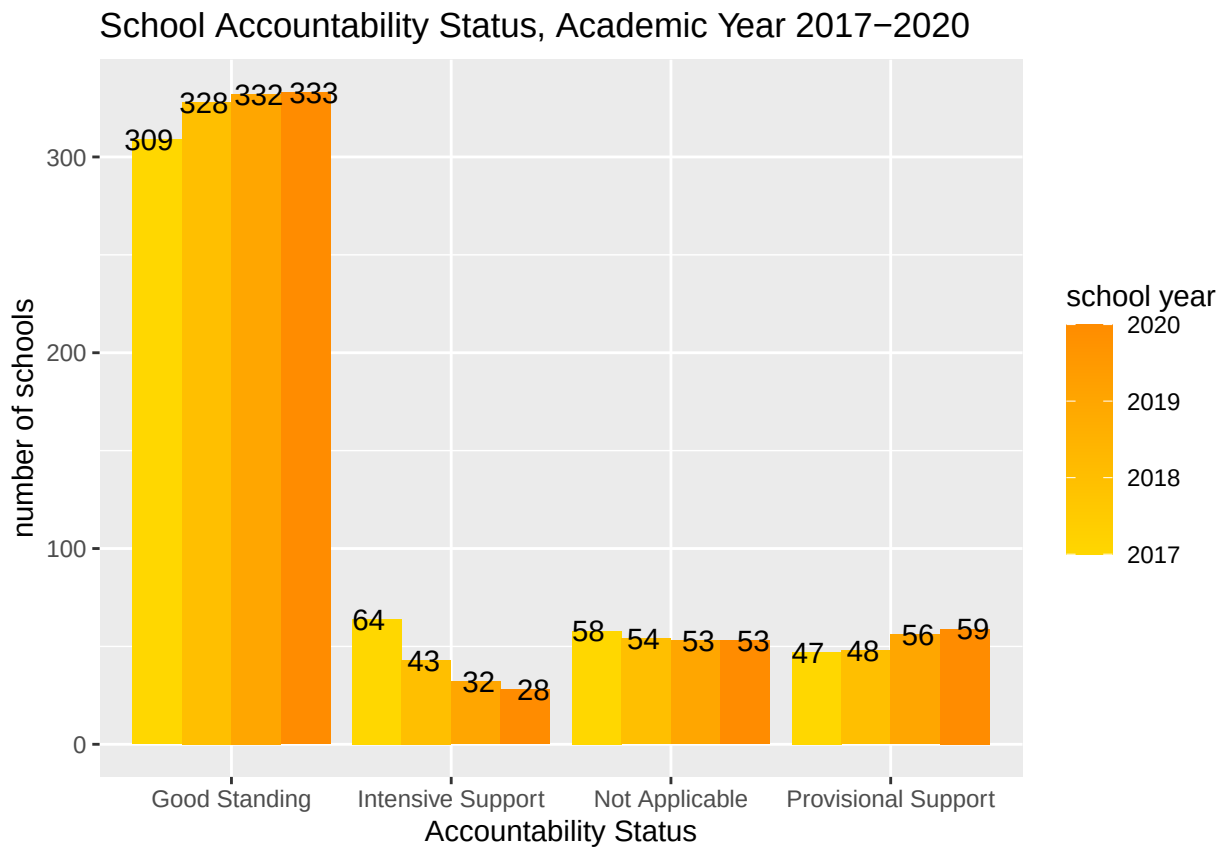


When we look more closely at school's rating annually in the past four years, the number of schools that obtain level 3 ratings has only increased by a handful. However, level 2 schools have increased from 2017 to 2018, and the number has persisted in the past two years.

```
status_graph <- sqrp_rating %>%
  group_by(school_year, status) %>%
  summarize(n_school = n()) %>%
  ggplot(aes(x = status, y = n_school, group = school_year)) +
  geom_col(aes(fill= school_year), position = "dodge") +
  geom_text(aes(x = status, label = n_school),
            position = position_dodge(width = 1), size = 4) +
  scale_fill_continuous(low = "gold", high = "darkorange") +
  labs(title = "School Accountability Status, Academic Year 2017-2020",
       x = "Accountability Status", fill = "school year", y = "number of schools")
```

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

```
status_graph
```




```
tabel1_ranking <- sqrp_rating %>%
  group_by(school_year, status) %>%
  summarize(n_school = n()) %>%
  pivot_wider(names_from = school_year, values_from = n_school)
```

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

```
tabel1_ranking
```

```
## # A tibble: 4 x 5
##   status              `2017` `2018` `2019` `2020`
##   <chr>                <int>  <int>  <int>  <int>
## 1 Good Standing          309    328    332    333
## 2 Intensive Support       64     43     32     28
## 3 Not Applicable          58     54     53     53
## 4 Provisional Support     47     48     56     59
```

Relationship between attendance and school rating

```
y2020 <- sqrp_rating %>%
  mutate(attandance_rate = as.numeric(attandance_rate),
         pct_read = as.numeric(pct_read),
```

```
        pct_math = as.numeric(pct_math),
        sqrp_total = as.numeric(sqrp_total)) %>%
  filter(school_year == 2020)
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion

## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

When we look at elementary school attendance on the percentile of NWEA growth in reading and math, there seems to be a strong positive correlation between the two variables. An increase in attendance rate by 1 percent among schools increases the school's percentile by 5 points in reading and math. Despite the strong correlation, attendance can only explain a small variation of the school's NWEA growth. Specifically, attendance can only explain the variation in NWEA growth percentile in math and reading by eight and nine percent, respectively. Such evidence suggests there are other factors that influence students' test scores.

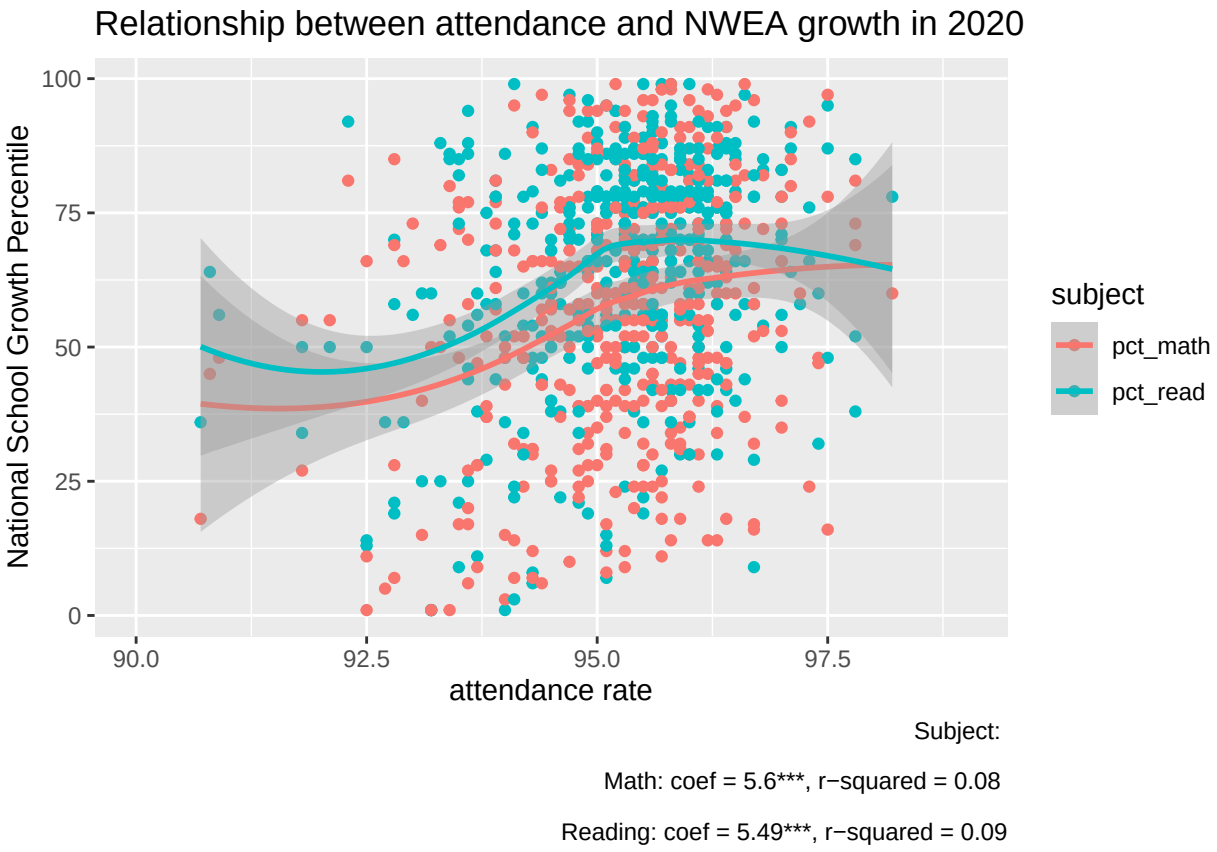
```
attendance_math_read <- y2020 %>%
  pivot_longer(pct_read:pct_math, names_to = "subject", values_to = "percentile") %>%
  group_by(subject) %>%
  ggplot(aes(x = attandance_rate, y = percentile)) +
  geom_point(aes(color = subject)) +
  geom_smooth(aes(color = subject)) + xlim(90,99) +
  labs(title = "Relationship between attendance and NWEA growth in 2020",
       x = "attendance rate",
       y = "National School Growth Percentile",
       label = "test subjects",
       caption = "Subject: \n
       Math: coef = 5.6***, r-squared = 0.08 \n
       Reading: coef = 5.49***, r-squared = 0.09")

attendance_math_read
```

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

## Warning: Removed 24 rows containing non-finite values (stat_smooth).

## Warning: Removed 24 rows containing missing values (geom_point).
```




```
reg_attend_read <- lm(pct_read ~ attandance_rate, data = y2020)
summary(reg_attend_read)
```

```
##
## Call:
## lm(formula = pct_read ~ attandance_rate, data = y2020)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.131 -11.164   3.111  13.759  43.044
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -458.1738    74.8485  -6.121 1.99e-09 ***
## attandance_rate    5.4944     0.7857   6.993 9.52e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.51 on 461 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.0959, Adjusted R-squared:  0.09394
## F-statistic:  48.9 on 1 and 461 DF,  p-value: 9.52e-12
```

```
reg_attend_math <- lm(pct_math ~ attandance_rate, data = y2020)
summary(reg_attend_math)
```

```
##
## Call:
## lm(formula = pct_math ~ attandance_rate, data = y2020)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.909 -16.557   1.714  17.321  44.688
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -483.541     86.874  -5.566 4.43e-08 ***
## attandance_rate    5.676      0.912   6.224 1.09e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22.64 on 461 degrees of freedom
##   (10 observations deleted due to missingness)
## Multiple R-squared:  0.07753,    Adjusted R-squared:  0.07553
## F-statistic: 38.74 on 1 and 461 DF,  p-value: 1.088e-09
```

Relationship between FiveEssential Survey to School Ranking

To measure school's climate, a survey titled Five Essential are conducted to teachers, students, and parents grade 4 to 12. In the past four years, most schools are categorized as "well organized" or "organized." Based on the measurement released by SQRP annually, schools have little information and incentives for school equity policies. To look at the school environment in detail, one could search for schools individually through the Illinois report card website, download an application or go to the five-essential website.

```
five_ess <- read_excel("Five Essential Survey_SQRP.xlsx")

five_ess_table <- five_ess %>%
  mutate(school_year = as.numeric(school_year)) %>%
  rename(five_ess = five_ess_2020) %>%
  filter(!is.na(five_ess)) %>%
  group_by(five_ess, school_year) %>%
  summarize(n_school = n()) %>%
  pivot_wider(names_from = school_year, values_from = n_school)
```

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

```
five_ess_table
```

```
## # A tibble: 6 x 5
## # Groups:   five_ess [6]
##   five_ess             `2017` `2018` `2019` `2020`
##   <chr>                 <int>  <int>  <int>  <int>
## 1 Moderately Organized     41     16     44     56
## 2 Not Enough Data           3      3      1      3
## 3 Not Yet Organized        30      8     28     21
## 4 Organized               114     72    120    107
## 5 Partially Organized      39     17     43     49
## 6 Well Organized          251    172    236    237
```

Relationship between Race and SQRP Ratings

One key argument against the current SQRP highlights the racial achievement gap and further reinforces segregation. Based on the SQRP 2020 result, when combined with the school's demographic data, there is a strong negative correlation between the share of African American students with SQRP total score. The share of African American students explains a 13 percent variation in the total SQRP score, which could be considered large compared to other racial/ethnic backgrounds. A larger share of white students strongly correlates with a higher SQRP score, further highlighting the racial gap in the SQRP scoring system. In other words, schools with low SQRP scores are likely schools dominated by African Americans.

```
race_2019 <- race %>%
  select(-school_name) %>%
  mutate(school_id = as.numeric(school_id))

race_sqrp <- sqrp_rating %>%
  filter(school_year == 2019) %>%
  left_join(race_2019, by = "school_id")
```

```
race_sqrp_2019 <- race_sqrp %>%
  mutate(pct_african_american = round(african_american/Total, 2),
         pct_white = round(white/Total,2),
         pct_hispanic = round(hispanic/Total,2),
         pct_asian = round(asian/Total,2),
         other = multi_racial+hawaii_pacific_islander+native_american_alaskan,
         pct_other = round(other/Total,2),
         sqrp_total = as.numeric(sqrp_total)
         )
```

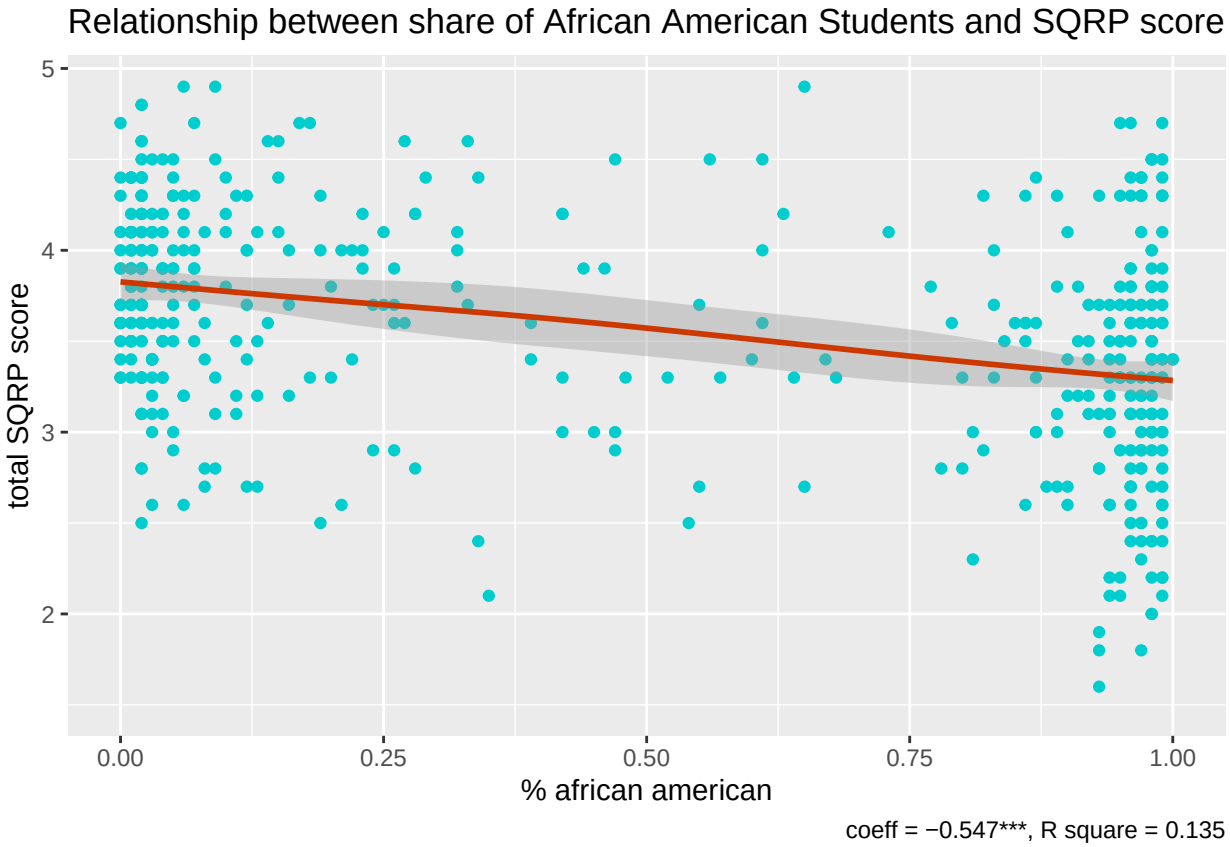
```
#relationship between share of african american and sqrp score
race_plot <- race_sqrp_2019 %>%
 ggplot(aes(x = pct_african_american, y = sqrp_total)) + geom_point(color = "cyan3") +
  geom_smooth(aes(x = pct_african_american, y = sqrp_total), color = "orangered3") +
  labs(x = "% african american", y = "total SQRP score",
       caption = "coeff = -0.547***, R square = 0.135",
       title = "Relationship between share of African American Students and SQRP score")

race_plot
```

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

## Warning: Removed 12 rows containing non-finite values (stat_smooth).

## Warning: Removed 12 rows containing missing values (geom_point).
```




```
reg_race_sqrp <- race_sqrp_2019 %>%
  select(sqrp_total, pct_african_american:pct_other, -other) %>%
  cor(use = "pairwise.complete.obs")

#african american
reg_aa <- lm(sqrp_total ~ pct_african_american, data = race_sqrp_2019)
summary(reg_aa)
```

```
##
## Call:
## lm(formula = sqrp_total ~ pct_african_american, data = race_sqrp_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.72124 -0.38688  0.01158  0.38643  1.42559
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.82998    0.04125  92.841  < 2e-16 ***
## pct_african_american -0.54703    0.06464  -8.463 3.53e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5908 on 459 degrees of freedom
##   (12 observations deleted due to missingness)
```

```
## Multiple R-squared:  0.135,  Adjusted R-squared:  0.1331
## F-statistic: 71.62 on 1 and 459 DF,  p-value: 3.526e-16
```

```
#white
reg_white <- lm(sqrp_total ~ pct_white, data = race_sqrp_2019)
summary(reg_white)
```

```
##
## Call:
## lm(formula = sqrp_total ~ pct_white, data = race_sqrp_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.86741 -0.38047  0.03259  0.41953  1.44565
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.45435    0.03133 110.264  < 2e-16 ***
## pct_white    1.30635    0.16461   7.936 1.62e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5956 on 459 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.1207, Adjusted R-squared:  0.1187
## F-statistic: 62.98 on 1 and 459 DF,  p-value: 1.615e-14
```

```
#hispanic
reg_his <- lm(sqrp_total ~ pct_hispanic, data = race_sqrp_2019)
summary(reg_his)
```

```
##
## Call:
## lm(formula = sqrp_total ~ pct_hispanic, data = race_sqrp_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.87620 -0.38818  0.03541  0.44957  1.34539
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.45878    0.04150  83.343  < 2e-16 ***
## pct_hispanic  0.29037    0.07726   3.759 0.000193 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6256 on 459 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.02986,    Adjusted R-squared:  0.02774
## F-statistic: 14.13 on 1 and 459 DF,  p-value: 0.0001929
```

```r
#asian
reg_asian <- lm(sqrp_total ~ pct_asian, data = race_sqrp_2019)
summary(reg_asian)
```

```
##
## Call:
## lm(formula = sqrp_total ~ pct_asian, data = race_sqrp_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.90140 -0.40140  0.00153  0.39860  1.39860
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.50140    0.02991  117.08  < 2e-16 ***
## pct_asian    2.18966    0.32014    6.84 2.55e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6051 on 459 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.09249,    Adjusted R-squared:  0.09051
## F-statistic: 46.78 on 1 and 459 DF,  p-value: 2.545e-11
```

```r
#other
reg_other <- lm(sqrp_total ~ pct_other, data = race_sqrp_2019)
summary(reg_other)
```

```
##
## Call:
## lm(formula = sqrp_total ~ pct_other, data = race_sqrp_2019)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.95218 -0.38729  0.03904  0.44782  1.34782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.55218    0.03281 108.255   <2e-16 ***
## pct_other    0.87784    0.71023   1.236    0.217
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6341 on 459 degrees of freedom
##   (12 observations deleted due to missingness)
## Multiple R-squared:  0.003317,   Adjusted R-squared:  0.001146
## F-statistic: 1.528 on 1 and 459 DF,  p-value: 0.2171
```