

STAR EXPLORATION

Importing Data and Recoding Categorical Variables into Numeric

Recoding kinder into new column

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr  0.3.4
## v tibble  3.1.8      v dplyr  1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

star <- read_csv("STAR.csv")

## Rows: 6325 Columns: 6
## -- Column specification -----
## Delimiter: ","
## dbl (6): race, classtype, yearssmall, hsgrad, g4math, g4reading
##
## 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.

star <- star %>% mutate(classtype=recode(classtype,
'1'="small",
'2'="regular",
'3'="regular with aid"))

print(star)

## # A tibble: 6,325 x 6
##   race classtype yearssmall hsgrad g4math g4reading
##   <dbl> <chr>      <dbl>   <dbl>   <dbl>   <dbl>
## 1 1 regular with aid 0 NA NA NA
## 2 2 regular with aid 0 NA 706 661
## 3 1 regular with aid 0 1 711 750
## 4 2 small 4 NA 672 659
## 5 1 regular 0 NA NA NA
## 6 1 regular with aid 0 NA NA NA
## 7 1 small 4 NA 668 657
## 8 1 regular with aid 0 NA NA NA
## 9 1 small 4 1 709 725
## 10 1 regular 0 1 698 692
## # ... with 6,315 more rows
```

Recoding race in same column

```
star <- read_csv("STAR.csv")

## Rows: 6325 Columns: 6
## -- Column specification -----
## Delimiter: ","
## dbl (6): race, classtype, yearssmall, hsgrad, g4math, g4reading
##
## 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.
```

```
star <- star %>% mutate(race=recode(race,
  '1'="White",
  '2'="Black",
  '3'="Hispanic",
  '4'="Other",
  '5'="Other",
  '6'="Other"))

print(star, n = 20)
```

```
## # A tibble: 6,325 x 6
##   race classtype yearssmall hsgrad g4math g4reading
##   <chr>      <dbl>      <dbl>  <dbl>  <dbl>    <dbl>
## 1 White         3         0    NA     NA     NA
## 2 Black         3         0    NA    706    661
## 3 White         3         0     1    711    750
## 4 Black         1         4    NA    672    659
## 5 White         2         0    NA     NA     NA
## 6 White         3         0    NA     NA     NA
## 7 White         1         4    NA    668    657
## 8 White         3         0    NA     NA     NA
## 9 White         1         4     1    709    725
## 10 White        2         0     1    698    692
## 11 White         3         0    NA     NA     NA
## 12 White         3         0    NA     NA     NA
## 13 Black         2         0     1    NA     NA
## 14 White         3         0    NA     NA     NA
## 15 White         3         0     1    NA     NA
## 16 White         3         0    NA     NA     NA
## 17 White         1         1    NA     NA     NA
## 18 White         2         2    NA     NA     NA
## 19 White         3         0    NA    733    672
## 20 White         3         0    NA     NA     NA
## # ... with 6,305 more rows
```

Difference in Means between Small and Regular Size Classes

Filter all students in small classes

```
small <- star[star$classtype == "1",]  
print(small, n = 20)
```

```
## # A tibble: 1,900 x 6  
##   race   classtype yearssmall hsgrad g4math g4reading  
##   <chr>      <dbl>      <dbl>  <dbl>  <dbl>    <dbl>  
## 1 Black         1         4    NA    672     659  
## 2 White         1         4    NA    668     657  
## 3 White         1         4     1    709     725  
## 4 White         1         1    NA    NA      NA  
## 5 Black         1         4    NA    NA      NA  
## 6 White         1         1    NA    NA      NA  
## 7 White         1         2    NA    NA      NA  
## 8 White         1         4     1   740     836  
## 9 White         1         4    NA   685     785  
## 10 White        1         4     1   771     725  
## 11 Black        1         1     1    NA      NA  
## 12 Black        1         1    NA    NA      NA  
## 13 Black        1         2     0    NA      NA  
## 14 Black        1         3     1    NA      NA  
## 15 White        1         4     1   709     761  
## 16 Black        1         3    NA    NA      NA  
## 17 Black        1         2     0    NA      NA  
## 18 Black        1         1     0    NA      NA  
## 19 White        1         3    NA    NA      NA  
## 20 White        1         1     1    NA      NA  
## # ... with 1,880 more rows
```

Find the means for math and reading test scores of small classes

```
small_math_mean <- mean(small$g4math, na.rm = TRUE)  
print(paste("small_math_mean:", small_math_mean))
```

```
## [1] "small_math_mean: 709.185135135135"
```

```
small_reading_mean <- mean(small$g4reading, na.rm = TRUE)  
print(paste("small_reading_mean:", small_reading_mean))
```

```
## [1] "small_reading_mean: 723.391184573003"
```

Filter all students in regular sized classes

```
regular <- star[star$classtype == "2",]  
print(regular, n = 20)
```

```
## # A tibble: 2,194 x 6  
##   race   classtype yearssmall hsgrad g4math g4reading  
##   <chr>      <dbl>      <dbl>  <dbl>  <dbl>    <dbl>  
## 1 White         2          0    NA     NA      NA  
## 2 White         2          0     1    698    692  
## 3 Black         2          0     1    NA     NA  
## 4 White         2          2    NA     NA     NA  
## 5 White         2          1    NA     NA     NA  
## 6 Black         2          0     1    NA     NA  
## 7 White         2          0     1    NA     NA  
## 8 White         2          0    NA     NA     NA  
## 9 White         2          0    NA    758    836  
## 10 Black        2          0     1    NA     NA  
## 11 Black        2          0    NA     NA     NA  
## 12 White        2          0    NA     NA     NA  
## 13 White        2          0    NA     NA     NA  
## 14 White        2          0    NA     NA     NA  
## 15 White        2          1     1    NA     NA  
## 16 White        2          0     1    698    778  
## 17 White        2          0    NA     NA     NA  
## 18 White        2          0    NA     NA     NA  
## 19 Black        2          0     0    NA     NA  
## 20 Black        2          0    NA     NA     NA  
## # ... with 2,174 more rows
```

Find the means for math and reading test scores of regular sized classes

```
regular_math_mean <- mean(regular$g4math, na.rm = TRUE)  
print(paste("regular_math_mean:", regular_math_mean))
```

```
## [1] "regular_math_mean: 709.521377672209"
```

```
regular_reading_mean <- mean(regular$g4reading, na.rm = TRUE)  
print(paste("regular_reading_mean:", regular_reading_mean))
```

```
## [1] "regular_reading_mean: 719.88995215311"
```

Find the difference in means for math and reading tests scores between small and regular sized classes

```
dif_math <- small_math_mean - regular_math_mean
print(paste("dif_math:", dif_math))
```

```
## [1] "dif_math: -0.336242537073872"
```

```
dif_reading <- small_reading_mean - regular_reading_mean
print(paste("dif_reading:", dif_reading))
```

```
## [1] "dif_reading: 3.50123241989274"
```

Find standard deviation of math and reading scores

```
sd_math <- sd(star$g4math, na.rm = TRUE)
print(paste("sd_math:", sd_math))
```

```
## [1] "sd_math: 43.0921739963875"
```

```
sd_reading <- sd(star$g4reading, na.rm = TRUE)
print(paste("sd_reading:", sd_reading))
```

```
## [1] "sd_reading: 52.425921784878"
```

Considering the standard deviations of test scores for both math and reading are above 40, the difference in the mean test scores between small and regular classes for both math and reading doesn't seem significant enough to attribute any real difference between the two class sizes.

Quantile Treatment Effect Analysis

Find high and low score for small classes

```
high_low_small_reading <- quantile(small$g4reading, probs=c(0.33,0.66), na.rm = TRUE)
print(high_low_small_reading)
```

```
## 33% 66%
## 705 741
```

```
high_low_small_math <- quantile(small$g4math, probs=c(0.33,0.66), na.rm = TRUE)
print(high_low_small_math)
```

```
## 33% 66%
## 694 726
```

Find high and low score for regular classes

```
high_low_regular_reading <- quantile(small$g4reading, probs=c(0.33,0.66), na.rm = TRUE)
print(high_low_regular_reading)
```

```
## 33% 66%
## 705 741
```

```
high_low_regular_math <- quantile(small$g4math, probs=c(0.33,0.66), na.rm = TRUE)
print(high_low_regular_math)
```

```
## 33% 66%
## 694 726
```

Difference in Reading & Math Scores between Minority and White Stuedents in Different Class Sizes

Find average reading and math test scores for white students in small classes

```
white_small <- small[small$race == "White",]
white_small_reading_mean <- mean(white_small$g4reading, na.rm = TRUE)
white_small_math_mean <- mean(white_small$g4math, na.rm = TRUE)
print(paste("white_small_reading_mean:", white_small_reading_mean))
```

```
## [1] "white_small_reading_mean: 727.838815789474"
```

```
print(paste("white_small_math_mean:", white_small_math_mean))
```

```
## [1] "white_small_math_mean: 711.19001610306"
```

Find average reading and math test scores for minority students in small classes

```
minority_small <- small[small$race %in% c("Black","Hispanic"),]
minority_small_reading_mean <- mean(minority_small$g4reading, na.rm = TRUE)
minority_small_math_mean <- mean(minority_small$g4math, na.rm = TRUE)
print(paste("minority_small_reading_mean:", minority_small_reading_mean))
```

```
## [1] "minority_small_reading_mean: 698.652173913043"
```

```
print(paste("minority_small_math_mean:", minority_small_math_mean))
```

```
## [1] "minority_small_math_mean: 697.508620689655"
```

Find average reading and math test scores for white students in regular classes

```
white_regular <- regular[regular$race == "White",]  
white_regular_reading_mean <- mean(white_regular$g4reading, na.rm = TRUE)  
white_regular_math_mean <- mean(white_regular$g4math, na.rm = TRUE)  
print(paste("white_regular_reading_mean:", white_regular_reading_mean))
```

```
## [1] "white_regular_reading_mean: 725.11581920904"
```

```
print(paste("white_regular_math_mean:", white_regular_math_mean))
```

```
## [1] "white_regular_math_mean: 711.410364145658"
```

Find average reading and math test scores for minority students in regular classes

```
minority_regular <- regular[regular$race %in% c("Black", "Hispanic"),]  
minority_regular_reading_mean <- mean(minority_regular$g4reading, na.rm = TRUE)  
minority_regular_math_mean <- mean(minority_regular$g4math, na.rm = TRUE)  
print(paste("minority_regular_reading_mean:", minority_regular_reading_mean))
```

```
## [1] "minority_regular_reading_mean: 690.984375"
```

```
print(paste("minority_regular_math_mean:", minority_regular_math_mean))
```

```
## [1] "minority_regular_math_mean: 698.984375"
```

Find difference between white and minority students in small and regular classes

```
dif_means_reading_race_small <- white_small_reading_mean - minority_small_reading_mean  
dif_means_math_race_small <- white_small_math_mean - minority_small_math_mean  
dif_means_reading_race_regular <- white_regular_reading_mean - minority_regular_reading_mean  
dif_means_math_race_regular <- white_regular_math_mean - minority_regular_math_mean
```

```

variables <- c("dif_means_reading_race_small", "dif_means_math_race_small", "dif_means_reading_race_reg", "dif_means_math_race_reg")
values <- c(dif_means_reading_race_small, dif_means_math_race_small, dif_means_reading_race_regular, dif_means_math_race_regular)

matrix_data <- matrix(c(variables, values), ncol = 4, byrow = TRUE)
print(matrix_data)

```

```

##      [,1]                [,2]
## [1,] "dif_means_reading_race_small" "dif_means_math_race_small"
## [2,] "29.1866418764301"            "13.6813954134045"
##      [,3]                [,4]
## [1,] "dif_means_reading_race_regular" "dif_means_math_race_regular"
## [2,] "34.1314442090395"            "12.4259891456583"

```

For small classes, there seems to be a noticeable difference between white and minority students in all class sizes and subjects (with white students having higher test scores). Taking into account the standard deviation of the test scores, these differences are still a portion of the standard deviations. However, the differences are much more than the differences between class sizes. In regards to race, the class sizes don't significantly change the difference between test scores between white and minority students.

Graduation Rates Across Different Class Sizes

Comparing high school graduation rates across students in different class types

```

grad_rate_small <- mean(small$hsgrad, na.rm = TRUE)

grad_rate_regular <- mean(regular$hsgrad, na.rm = TRUE)

regular_w_aid <- star[star$classtype == '3',]

grad_rate_regular_w_aid <- mean(regular_w_aid$hsgrad, na.rm = TRUE)

print(paste("grad_rate_regular_w_aid:", grad_rate_regular_w_aid))

## [1] "grad_rate_regular_w_aid: 0.839285714285714"

```

The graduation rates between the different classtypes are all relatively equal, suggesting there is no significant difference in graduation rate due to classtype.

Graduation Rates based on number of years spent in small classes

```

zero_year <- star[star$yearssmall == '0',]

one_year <- star[star$yearssmall == '1',]

two_year <- star[star$yearssmall == '2',]

```



```

three_year <- star[star$yearssmall == '3',]
four_year <- star[star$yearssmall == '4',]

grad_rate_zero_year <- mean(zero_year$hsgrad, na.rm = TRUE)
grad_rate_one_year <- mean(one_year$hsgrad, na.rm = TRUE)
grad_rate_two_year <- mean(two_year$hsgrad, na.rm = TRUE)
grad_rate_three_year <- mean(three_year$hsgrad, na.rm = TRUE)
grad_rate_four_year <- mean(four_year$hsgrad, na.rm = TRUE)

variables <- c("grad_rate_zero_year", "grad_rate_one_year", "grad_rate_two_year", "grad_rate_three_year")
values <- c(grad_rate_zero_year, grad_rate_one_year, grad_rate_two_year, grad_rate_three_year, grad_rate_four_year)

matrix_data <- matrix(c(variables, values), ncol = 5, byrow = TRUE)
print(matrix_data)

```

```

##      [,1]      [,2]      [,3]
## [1,] "grad_rate_zero_year" "grad_rate_one_year" "grad_rate_two_year"
## [2,] "0.82860203535083"    "0.791044776119403"    "0.813186813186813"
##      [,4]      [,5]
## [1,] "grad_rate_three_year" "grad_rate_four_year"
## [2,] "0.832460732984293"    "0.877551020408163"

```

While the graduation rates for spending zero to three years in small classes are all relatively similar (with one year in small classes having the lowest graduation rate), spending four years in small classes seems to have a significant increase in graduation rates, suggesting small classes may be better if students spend the maximum amount (4 years) in them.

Graduation Rates among white and minority students in small classes

```

grad_rate_white_small <- mean(white_small$hsgrad, na.rm = TRUE)
grad_rate_minority_small <- mean(minority_small$hsgrad, na.rm = TRUE)

dif_grad_rate_small <- grad_rate_white_small - grad_rate_minority_small

print(paste("dif_grad_rate_small:", dif_grad_rate_small))

## [1] "dif_grad_rate_small: 0.120634436480098"

```

Graduation Rates among white and minority students in regular classes

```
grad_rate_white_regular <- mean(white_regular$hsgrad, na.rm = TRUE)
grad_rate_minority_regular <- mean(minority_regular$hsgrad, na.rm = TRUE)
dif_grad_rate_regular <- grad_rate_white_regular - grad_rate_minority_regular
print(paste("dif_grad_rate_regular:", dif_grad_rate_regular))
```

```
## [1] "dif_grad_rate_regular: 0.119030990833697"
```

Graduation Rates among white and minority students in regular classes with aids

```
white_regular_w_aid <- regular_w_aid[regular_w_aid$race == "White",]
minority_regular_w_aid <- regular_w_aid[regular_w_aid$race %in% c("Black", "Hispanic"),]
grad_rate_white_regular_w_aid <- mean(white_regular_w_aid$hsgrad, na.rm = TRUE)
grad_rate_minority_regular_w_aid <- mean(minority_regular_w_aid$hsgrad, na.rm = TRUE)
dif_grad_rate_regular_w_aid <- grad_rate_white_regular_w_aid - grad_rate_minority_regular_w_aid
print(paste("dif_grad_rate_regular_w_aid:", dif_grad_rate_regular_w_aid))
```

```
## [1] "dif_grad_rate_regular_w_aid: 0.142261883787308"
```

While the regular classtype had the lowest difference in graduation rates between white and minority students, all the difference in graduation rates for the different classtypes were fairly close in value, suggesting the STAR program has not significantly reduced the racial gap between white and minority student's graduation rates. Small classtypes have a higher difference in mean graduation rates than regular sized classes, suggesting a lower student to teacher ratio doesn't mean a higher graduation rate. Moreover, a regular sized class with an aid, which should have the lowest student to teach ratio, has the highest difference in mean graduation rates, further suggesting the aforementioned.