601Final

Joey Wolpert

8/27/2021

Homework 2

Let's read in the data and see what it looks like:

«««< HEAD

```
data <- read.csv("C:/Users/wolpe/DACSS601August2021/_data/Public_School_Characteristics_2017-18.csv")
```

What's in the data?

This dataset looks at the characteristics of various public schools across the United States. Among the variables in the dataset are identifying characteristics such as the name of the school, its school district, and its location; there are also several quantitative variables such as the number of students in each grade, as well as the overall number of students broken down in categories such as race and gender.

The numbers of the data

The data has 79 total columns and just over 100,000 rows. It is unlikely that all of this info will be useful, so in the next section we can see if the data can be cleaned and subset to be more useful to the project.

Homework 3

Cleaning Data

The data is definitely unclean, let's filter out some stuff so that we have more complete data.

```
data_clean <- data %>%
  filter(!is.na(TOTAL) & TOTAL > 0 & !is.na(FTE) & FTE > 0 & STUTERATIO < 500) %>%
  filter(SCHOOL_TYPE_TEXT == "Regular school" & VIRTUAL == "Not a virtual school") %>%
  filter(SCHOOL_LEVEL == "High" | SCHOOL_LEVEL == "Middle" | SCHOOL_LEVEL == "Elementary") %>%
  filter((is.na(G13) | G13 == 0) & (is.na(PK) | PK == 0)) %>%
  filter(!is.na(STABR) & !is.na(SCH_NAME))
head(data_clean)
```

```
##        ï..X        Y OBJECTID     NCESSCH                     NMCNTY  SURVYEAR
## 1 -151.0701 60.49144        3 20039000448    Kenai Peninsula Borough 2017-2018
## 2 -151.2791 60.56828        4 20039000463    Kenai Peninsula Borough 2017-2018
```

```
## 3 -166.5224 53.86895       15 20072000340 Aleutians West Census Area 2017-2018
## 4 -166.5296 53.87267       16 20072000661 Aleutians West Census Area 2017-2018
## 5 -161.7707 60.80436       18 20000100207         Bethel Census Area 2017-2018
## 6 -161.7704 60.80258       19 20000100208         Bethel Census Area 2017-2018
##   STABR  LEAID ST_LEAID                                LEA_NAME
## 1    AK 200390    AK-24 Kenai Peninsula Borough School District
## 2    AK 200390    AK-24 Kenai Peninsula Borough School District
## 3    AK 200720    AK-47           Unalaska City School District
## 4    AK 200720    AK-47           Unalaska City School District
## 5    AK 200001    AK-31         Lower Kuskokwim School District
## 6    AK 200001    AK-31         Lower Kuskokwim School District
##                                SCH_NAME                    LSTREET1 LSTREET2
## 1    Soldotna Montessori Charter School               158 E Park Ave
## 2 Kaleidoscope School of Arts & Science              549 N Forest Dr
## 3        Eagle's View Elementary School            503 East Broadway
## 4            Unalaska Jr/Sr High School               55 E Broadway
## 5                Gladys Jung Elementary 1007 Ron Edwards Memorial Dr
## 6           Bethel Regional High School 1006 Ron Edwards Memorial Dr
##   LSTREET3    LCITY LSTATE  LZIP LZIP4         PHONE GSLO GSHI
## 1          Soldotna     AK 99669    NA (907)260-9221   KG   06
## 2             Kenai     AK 99611    NA (907)283-0804   KG   05
## 3          Unalaska     AK 99685    NA (907)581-3979   PK   06
## 4          Unalaska     AK 99685    NA (907)581-1222   07   12
## 5            Bethel     AK 99559    NA (907)543-4440   03   06
## 6            Bethel     AK 99559    NA (907)543-3957   07   12
##                 VIRTUAL TOTFRL FRELCH REDLCH PK KG G01 G02 G03 G04 G05 G06 G07
## 1 Not a virtual school     43     23     20 NA 23  23  27  22  25  28  19  NA
## 2 Not a virtual school     69     50     19 NA 40  43  42  46  46  43  NA  NA
## 3 Not a virtual school     53     35     18  0 32  30  36  33  31  26  29  NA
## 4 Not a virtual school     38     27     11 NA NA  NA  NA  NA  NA  NA  NA  30
## 5 Not a virtual school    294    294      0 NA NA  NA  NA  97  75  79  90  NA
## 6 Not a virtual school    373    373      0 NA NA  NA  NA  NA  NA  NA  NA  94
##   G08 G09 G10 G11 G12 G13 TOTAL MEMBER  AM HI BL  WH HP TR   FTE   LATCOD
## 1  NA  NA  NA  NA  NA  NA   167    167   8  5  0 136  0 15 10.35 60.49144
## 2  NA  NA  NA  NA  NA  NA   260    260  16 14  3 168  0 56 16.75 60.56828
## 3  NA  NA  NA  NA  NA  NA   217    217  23 30  2  56 13  3 13.50 53.86895
## 4  25  26  38  36  29  NA   184    184  24 21  0  38  8  0 14.50 53.87267
## 5  NA  NA  NA  NA  NA  NA   341    341 284  6  1  44  0  1 22.13 60.80436
## 6  90 106  52  63  70  NA   475    475 418  7  1  38  0  2 33.05 60.80258
##      LONCOD          ULOCALE STUTERATIO        STITLEI AMALM AMALF ASALM ASALF
## 1 -151.0702  33-Town: Remote      16.14 Not Applicable     4     4     0     3
## 2 -151.2791  33-Town: Remote      15.52 Not Applicable    10     6     1     2
## 3 -166.5225 43-Rural: Remote      16.07 Not Applicable    11    12    52    38
## 4 -166.5296 43-Rural: Remote      12.69 Not Applicable    12    12    52    41
## 5 -161.7707 41-Rural: Fringe      15.41            Yes   141   143     2     3
## 6 -161.7704 41-Rural: Fringe      14.37            Yes   221   197     5     4
##   HIALM HIALF BLALM BLALF WHALM WHALF HPALM HPALF TRALM TRALF TOTMENROL
## 1     2     3     0     0    58    78     0     0     7     8        71
## 2     6     8     3     0    82    86     0     0    26    30       128
## 3    14    16     0     2    26    30     7     6     1     2       111
## 4    12     9     0     0    23    15     4     4     0     0       103
## 5     4     2     0     1    21    23     0     0     0     1       168
## 6     4     3     0     1    20    18     0     0     1     1       251
##   TOTFENROL STATUS UG AE SCHOOL_TYPE_TEXT         SY_STATUS_TEXT SCHOOL_LEVEL
```

```
## 1        96      1 NA NA   Regular school Currently operational    Elementary
## 2       132      1 NA NA   Regular school Currently operational    Elementary
## 3       106      1 NA NA   Regular school Currently operational    Elementary
## 4        81      1 NA NA   Regular school Currently operational          High
## 5       173      1 NA NA   Regular school Currently operational    Elementary
## 6       224      1 NA NA   Regular school Currently operational          High
##   AS CHARTER_TEXT MAGNET_TEXT
## 1  3          Yes          No
## 2  3          Yes          No
## 3 90           No          No
## 4 93           No          No
## 5  5           No          No
## 6  9           No          No
```

With filtering, we now have a subset of the original data that will be much more useful for analysis. The schools were filtered to include only regular, non-virtual schools at the elementary, middle, and high school levels. It also removed any schools that had students younger than kindergarten (PK) or those past their senior year of high school (G13).

There is still an issue however with student to teacher ratios, so let's look at a distribution of that:

```
data_clean %>%
  summarise(quantile = c(0, 0.005, 0.025, 0.5, 0.975, 0.995, 1), quant.val = quantile(STUTERATIO, c(0, 0
```

```
##   quantile quant.val
## 1    0.000    0.1900
## 2    0.005    4.8200
## 3    0.025    8.1800
## 4    0.500   15.7700
## 5    0.975   27.0000
## 6    0.995   34.4274
## 7    1.000  485.0000
```

Looking at the distribution, we see that 99% of the dataset falls between a student to teacher ratio of 4.82 and 34.4274, which we can use as a boundary to remove outliers of this variable from the dataset.

```
data_clean <- data_clean %>%
  filter(STUTERATIO >= 4.82 & STUTERATIO <= 34.4274)
```

Subsetting Columns

Now let's subset the columns in the dataset to only include those that interest this project.

```
data_sub <- data_clean %>%
  select(SCH_NAME, STABR, GSLO, GSHI, G01, G02, G03, G04, G05, G06, G07, G08, G09, G10, G11, G12, TOTAL
head(data_sub)
```

```
##                              SCH_NAME STABR GSLO GSHI G01 G02 G03 G04 G05 G06
## 1    Soldotna Montessori Charter School    AK   KG   06  23  27  22  25  28  19
## 2 Kaleidoscope School of Arts & Science    AK   KG   05  43  42  46  46  43  NA
## 3        Eagle's View Elementary School    AK   PK   06  30  36  33  31  26  29
## 4            Unalaska Jr/Sr High School    AK   07   12  NA  NA  NA  NA  NA  NA
```

```
## 5                Gladys Jung Elementary    AK   03   06  NA  NA  97  75  79  90
## 6           Bethel Regional High School    AK   07   12  NA  NA  NA  NA  NA  NA
##   G07 G08 G09 G10 G11 G12 TOTAL  AM HI BL  WH HP TR   FTE STUTERATIO AMALM
## 1  NA  NA  NA  NA  NA  NA   167   8  5  0 136  0 15 10.35      16.14     4
## 2  NA  NA  NA  NA  NA  NA   260  16 14  3 168  0 56 16.75      15.52    10
## 3  NA  NA  NA  NA  NA  NA   217  23 30  2  56 13  3 13.50      16.07    11
## 4  30  25  26  38  36  29   184  24 21  0  38  8  0 14.50      12.69    12
## 5  NA  NA  NA  NA  NA  NA   341 284  6  1  44  0  1 22.13      15.41   141
## 6  94  90 106  52  63  70   475 418  7  1  38  0  2 33.05      14.37   221
##   AMALF ASALM ASALF HIALM HIALF BLALM BLALF WHALM WHALF HPALM HPALF TRALM TRALF
## 1     4     0     3     2     3     0     0    58    78     0     0     7     8
## 2     6     1     2     6     8     3     0    82    86     0     0    26    30
## 3    12    52    38    14    16     0     2    26    30     7     6     1     2
## 4    12    52    41    12     9     0     0    23    15     4     4     0     0
## 5   143     2     3     4     2     0     1    21    23     0     0     0     1
## 6   197     5     4     4     3     0     1    20    18     0     0     1     1
##   TOTMENROL TOTFENROL SCHOOL_LEVEL
## 1        71        96   Elementary
## 2       128       132   Elementary
## 3       111       106   Elementary
## 4       103        81         High
## 5       168       173   Elementary
## 6       251       224         High
```

Arranging Some Data

It may be useful to be the top several rows for certain columns, in this script, we look at the top 6 schools ordered based on their student to teacher ratio.

```
data_stuteratio <- data_sub %>%
  arrange(STUTERATIO, by_group = TRUE)
head(data_stuteratio)
```

```
##                         SCH_NAME STABR GSLO GSHI G01 G02 G03 G04 G05 G06 G07
## 1           George Jr Republic HS    PA   09   12  NA  NA  NA  NA  NA  NA  NA
## 2         Medicine Bow Elementary    WY   KG   06   1   4   0   3   2   1  NA
## 3 LONE STAR UNDIVIDED HIGH SCHOOL    CO   09   12  NA  NA  NA  NA  NA  NA  NA
## 4          VERDIGRE MIDDLE SCHOOL    NE   07   08  NA  NA  NA  NA  NA  NA   4
## 5          Judith Gap High School    MT   09   12  NA  NA  NA  NA  NA  NA  NA
## 6        Mobile Elementary School    AZ   KG   08   5   0   5   1   3   1   0
##   G08 G09 G10 G11 G12 TOTAL AM HI  BL WH HP TR   FTE STUTERATIO AMALM AMALF
## 1  NA  54  54  88  83   279  1 36 135 80  0 26 57.94       4.82     1     0
## 2  NA  NA  NA  NA  NA    11  0  0   0 11  0  0  2.28       4.82     0     0
## 3  NA  16   5   4   4    29 NA  3  NA 26 NA NA  6.01       4.83    NA    NA
## 4  10  NA  NA  NA  NA    14  2  0   0 12  0  0  2.90       4.83     1     1
## 5  NA   4   3   3   3    13  0  1   0 12  0  0  2.68       4.85     0     0
## 6   2  NA  NA  NA  NA    19  0  4   5  9  0  1  3.90       4.87     0     0
##   ASALM ASALF HIALM HIALF BLALM BLALF WHALM WHALF HPALM HPALF TRALM TRALF
## 1     1     0    36     0   135     0    80     0     0     0    26     0
## 2     0     0     0     0     0     0     4     7     0     0     0     0
## 3    NA    NA     2     1    NA    NA    13    13    NA    NA    NA    NA
## 4     0     0     0     0     0     0     5     7     0     0     0     0
## 5     0     0     1     0     0     0     5     7     0     0     0     0
```

```
## 6     0     0     3     1     2     3     5     4     0     0     0     1
##   TOTMENROL TOTFENROL SCHOOL_LEVEL
## 1       279         0         High
## 2         4         7   Elementary
## 3        15        14         High
## 4         6         8       Middle
## 5         6         7         High
## 6        10         9   Elementary
```

And now we look at the bottom 6 schools with the highest student-teacher ratio.

```
data_stuteratio <- data_sub %>%
  arrange(STUTERATIO)
tail(data_stuteratio)
```

```
##                                                SCH_NAME STABR GSLO GSHI G01 G02
## 58631                        College of So. NV HS East    NV   11   12  NA  NA
## 58632                             Everest High School    OH   09   12  NA  NA
## 58633                           Birch Grove Intermediate    CA   03   06  NA  NA
## 58634                          John C. Fremont Elementary    CA   KG   08  85  76
## 58635 Aspire Benjamin Holt College Preparatory Academy    CA   09   12  NA  NA
## 58636                               Hamburg Middle School    IA   06   08  NA  NA
##       G03 G04 G05 G06 G07 G08 G09 G10 G11 G12 TOTAL AM  HI BL WH HP TR   FTE
## 58631  NA  NA  NA  NA  NA  NA  NA  NA 102  68   170 NA  85 23 42  4  5  5.00
## 58632  NA  NA  NA  NA  NA  NA  20  24  17   7    68  0   4 37 24  0  3  2.00
## 58633 109 118 119 112  NA  NA  NA  NA  NA  NA   458  1 180 28 52  6 29 13.40
## 58634  77  95 115 113  91 102  NA  NA  NA  NA   867 13 759 26 33  4  6 25.24
## 58635  NA  NA  NA  NA  NA  NA 122 130  94  72   418  5 184 23 83  3 16 12.17
## 58636  NA  NA  NA  14  18  13  NA  NA  NA  NA    45 NA   7 NA 37 NA  1  1.31
##       STUTERATIO AMALM AMALF ASALM ASALF HIALM HIALF BLALM BLALF WHALM WHALF
## 58631      34.00    NA    NA     3     8    42    43     5    18    17    25
## 58632      34.00     0     0     0     0     1     3    27    10    15     9
## 58633      34.18     1    NA    93    69    95    85    12    16    30    22
## 58634      34.35     8     5    14    12   400   359    16    10    19    14
## 58635      34.35     1     4    55    49    87    97    10    13    35    48
## 58636      34.35    NA    NA    NA    NA     3     4    NA    NA    17    20
##       HPALM HPALF TRALM TRALF TOTMENROL TOTFENROL SCHOOL_LEVEL
## 58631     1     3     2     3        70       100         High
## 58632     0     0     1     2        44        24         High
## 58633     5     1    18    11       254       204   Elementary
## 58634     1     3     4     2       462       405   Elementary
## 58635     3    NA     4    12       195       223         High
## 58636    NA    NA    NA     1        20        25       Middle
```

Summary Data

Now let's take a look at some summary of the dataset in terms of its student-teacher ratio in each state.

```
data_stats <- data_sub %>%
  group_by(STABR) %>%
  summarise(STABR = STABR, AvgRatio = mean(STUTERATIO), SDRatio = sd(STUTERATIO))
```

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

```
data_summary <- distinct(data_stats)
data_summary
```

```
## # A tibble: 53 x 3
## # Groups:   STABR [53]
##    STABR AvgRatio SDRatio
##    <chr>    <dbl>   <dbl>
##  1 AK        16.0    3.02
##  2 AL        18.1    2.64
##  3 AR        13.2    2.81
##  4 AZ        18.9    3.77
##  5 CA        23.4    3.47
##  6 CO        16.4    3.71
##  7 CT        12.0    1.92
##  8 DC        13.1    4.02
##  9 DE        14.9    2.63
## 10 FL        17.2    3.66
## # ... with 43 more rows
```

Homework 4

For this section we need to look deeper into two variables from the dataset, for this I want to look at school type and student to teacher ratio.

How This Data Was Collected

Taken from the NCES website: "The National Center for Education Statistics' (NCES) Common Core of Data (CCD) program is an annual collection of basic administrative characteristics for all public schools, school districts, and state education agencies in the United States. These characteristics are reported by state education officials and include directory information, number of students, number of teachers, grade span, and other conditions. The NCES Education Demographic and Geographic Estimate (EDGE) program develops annually updated point locations (latitude and longitude) for public elementary and secondary schools included in the CCD. The NCES EDGE program collaborates with the U.S. Census Bureau's Education Demographic, Geographic, and Economic Statistics (EDGE) Branch to develop point locations for schools and school district administrative offices based on reported physical addresses" (NCES).

School Type

Let's take a look at all the different levels of schools from the original, uncleaned dataset.

```
data %>%
  group_by(SCHOOL_LEVEL) %>%
  summarise(SCHOOL_LEVEL = SCHOOL_LEVEL, count.type = n()) %>%
  distinct()
```

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

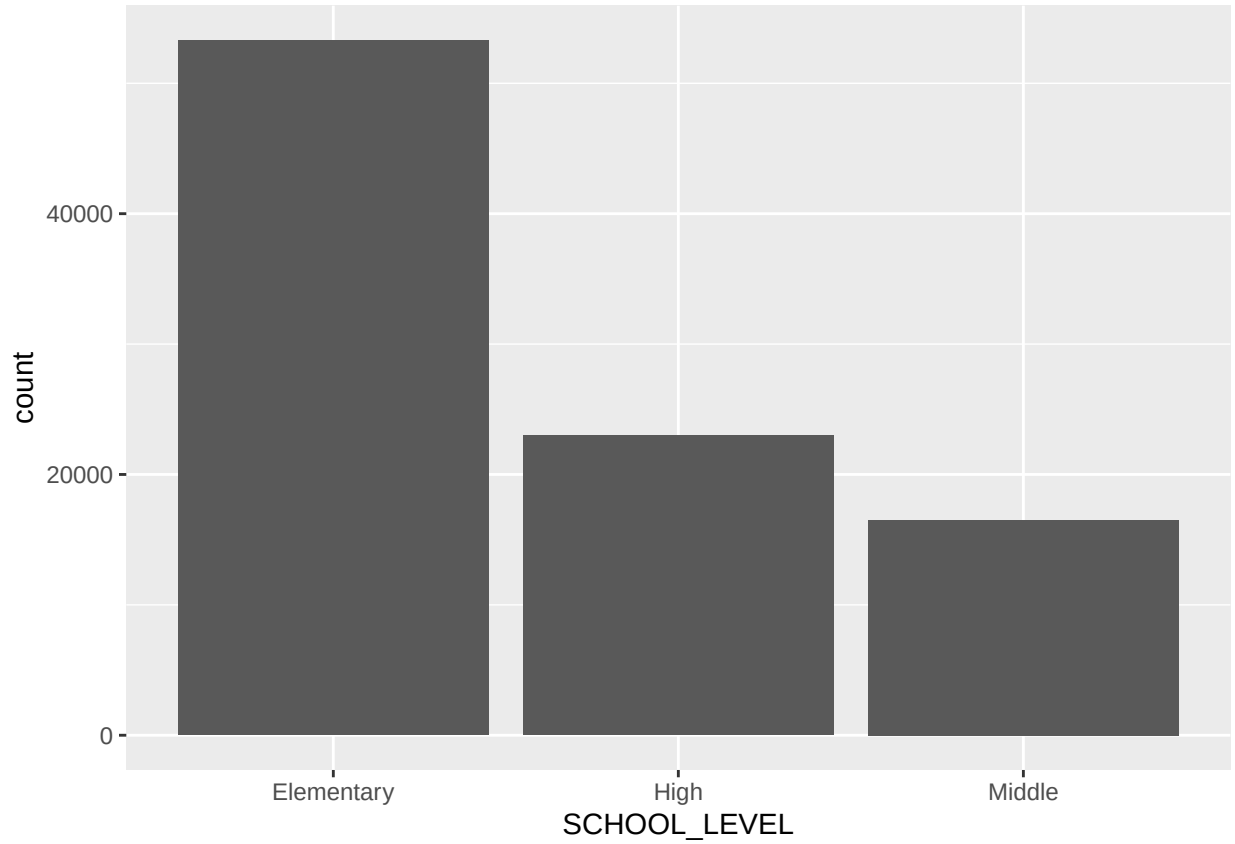
```
## # A tibble: 10 x 2
## # Groups:   SCHOOL_LEVEL [10]
```

```
##    SCHOOL_LEVEL    count.type
##    <chr>                <int>
##  1 Adult Education         28
##  2 Elementary           53287
##  3 High                 22977
##  4 Middle               16506
##  5 Not Applicable         796
##  6 Not Reported          1113
##  7 Other                 3824
##  8 Prekindergarten       1430
##  9 Secondary              602
## 10 Ungraded               166
```

For the dataset I wanted to focus on the most common types of school, which immediately meant filtering out adult education, secondary, and ungraded schools. This also left not applicable, not reported, and other in the dataset, which also should be removed. The last four remaining categories were elementary, high, middle, and prekindergarten. Prekindergarten schools were then removed to focus on the main three types of public schools as well as because of the number of fewer schools of this type.

Now let's filter out these unwanted categories and look at the breakdown of schook type.

```
data %>%
  filter(SCHOOL_LEVEL == "High" | SCHOOL_LEVEL == "Middle" | SCHOOL_LEVEL == "Elementary") %>%
  ggplot(aes(x = SCHOOL_LEVEL)) + geom_bar()
```



This looks significantly cleaner and more organized than with all the other unused variables.

Student to Teacher Ratio

Let's start by looking at some summary statistics and a barplot of student to teacher ratio from the unclean dataset:

```
data %>%
  summarise(count = n(), mean.val = mean(STUTERATIO), sd.val = sd(STUTERATIO), median.val = median(STUT
```

```
##    count mean.val sd.val median.val
## 1 100729       NA     NA         NA
```

We see that there are probably a large amount of NA values, so we first need to clean that.

```
data_stuteratio <- data %>%
  filter(!is.na(STUTERATIO))
data_stuteratio %>%
  summarise(count = n(), mean.val = mean(STUTERATIO), sd.val = sd(STUTERATIO), median.val = median(STUT
```

```
##   count mean.val   sd.val median.val
## 1 93894 16.94477 85.73974      15.33
```

Now we are getting valid statistics, but the standard deviation appears to be very large, over 5 times the mean and median! Let's take a look at some quantiles to see what the breakdown is.

```
quantiles <- c(0, 0.005, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95, 0.975, 0.995, 1)
data_stuteratio %>%
  summarise(quantiles = quantiles, value = quantile(STUTERATIO, quantiles))
```

```
##    quantiles       value
## 1      0.000     0.00000
## 2      0.005     0.00000
## 3      0.025     5.10325
## 4      0.050     8.02000
## 5      0.100    10.33000
## 6      0.250    12.85000
## 7      0.500    15.33000
## 8      0.750    18.18000
## 9      0.900    22.50000
## 10     0.950    25.11000
## 11     0.975    27.50000
## 12     0.995    50.30070
## 13     1.000 22350.00000
```

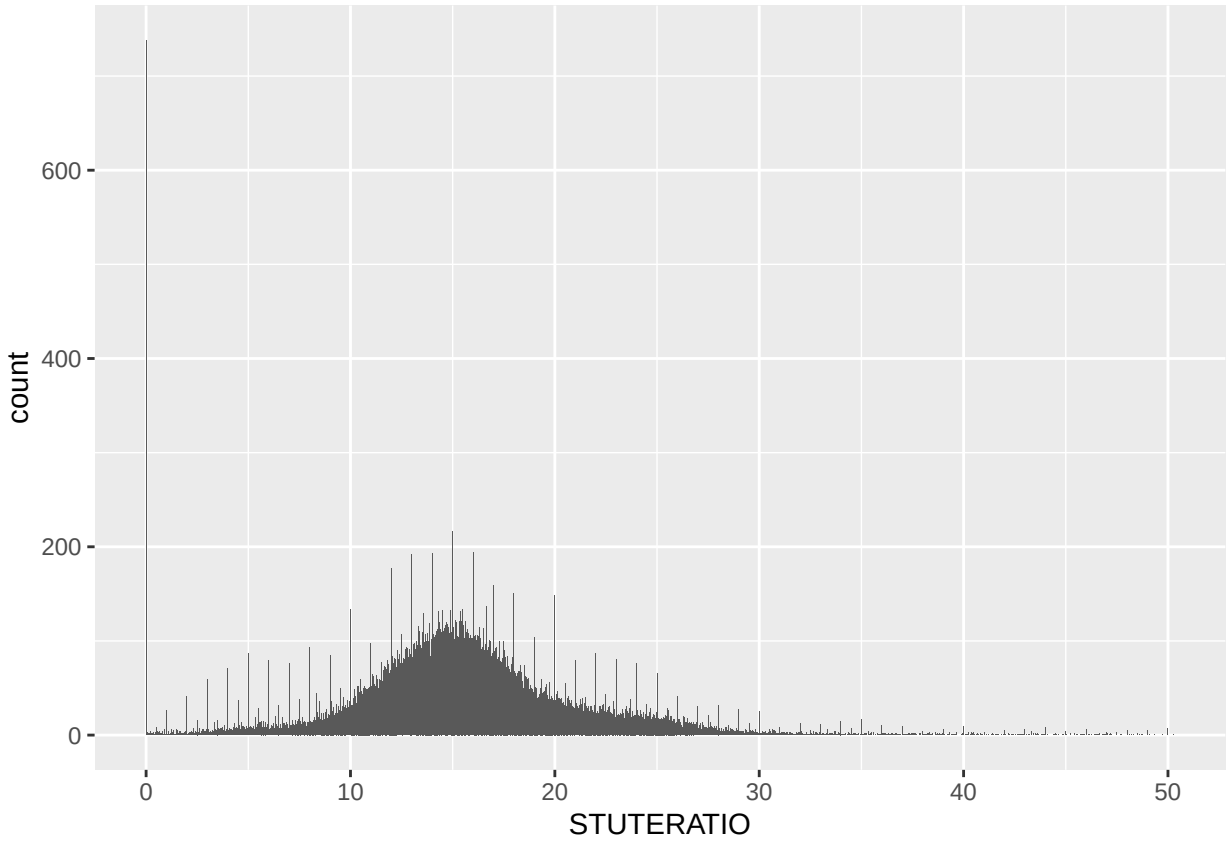
We see that our wide range of values that cause a large standard deviation is probably due to the top 0.5% of values for student to teacher ration, as the bottom 99.5% are all equal to or less than 50.3, with the top 0.5% ranging from that to over 20,000! To look at more meaningful data, we can select the middle 99% of the data to use.

```
data_stuteratio <- data_stuteratio %>%
  filter(STUTERATIO <= 50.3007)
data_stuteratio %>%
  summarise(count = n(), mean.val = mean(STUTERATIO), sd.val = sd(STUTERATIO), median.val = median(STUT
```

```
##   count mean.val sd.val median.val
## 1 93424 15.71938 5.2405      15.31
```

The standard deviation for this value now makes a lot more sense, and the median has moved closer to the mean. Let's take a look at a barplot to see the distribution.

```
data_stuteratio %>%
  ggplot(aes(x = STUTERATIO)) + geom_bar()
```



The data still appears somewhat skewed, but not in a way that will negatively affect our analysis.

NOTE: In the actual filtering of data the cutoffs used for STUTERATIO are different because of other filtering, in actuality the cutoffs of 4.82 and 34.4274 will be used.

Homework 5

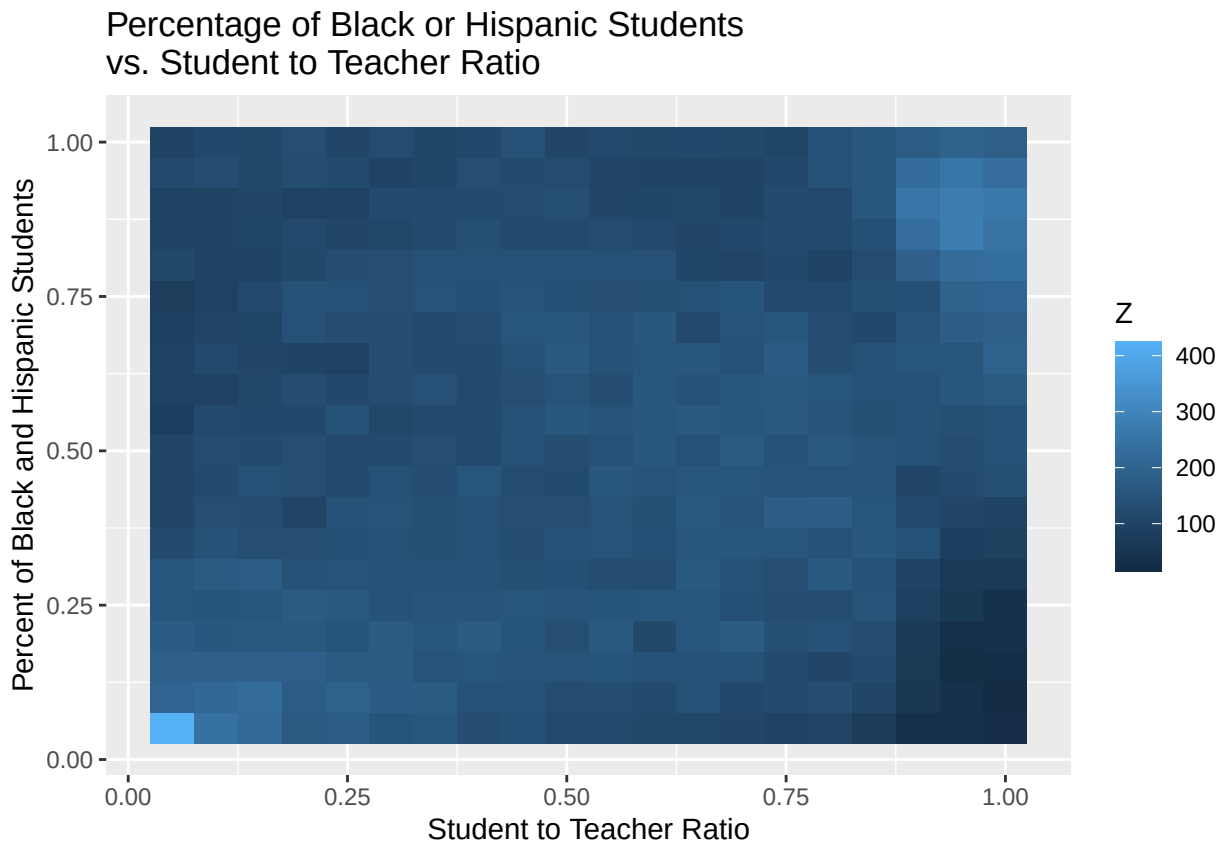
Let's take a look at student to teacher ratio in relation to the percentage of black and hispanic students in the school.

```
quant <- c()
axis_val <- c()
count <- 1
heat.data <- data_sub %>%
  summarise(SCH_NAME, STUTERATIO, BLHI_PCT = (BL + HI)/TOTAL) %>%
```

```
  filter(!is.na(BLHI_PCT))
while (count <= 20) {
  count2 <- 1
  while (count2 <= 20) {
    cur.data <- heat.data %>%
      filter(STUTERATIO >= quantile(STUTERATIO, prob = c((count2-1)*0.05)) & STUTERATIO < quantile(STUT
    quant <- append(quant, nrow(cur.data))
    count2 <- count2 + 1
  }
  axis_val <- append(axis_val, 0.05*count)
  count <- count + 1
}
data <- expand.grid(X=axis_val, Y=axis_val)
data$Z <- quant
print(axis_val)
```

```
##  [1] 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75
## [16] 0.80 0.85 0.90 0.95 1.00
```

```
ggplot(data, aes(X, Y, fill= Z)) +
  geom_tile() + labs(title = "Percentage of Black or Hispanic Students \nvs. Student to Teacher Ratio",
```




```
quant
```

```
##   [1] 425 241 214 172 180 150 158 129 136 118 122 113 110 108  98 102  73  33
##  [19]  31  20 201 209 222 180 192 176 175 143 144 127 127 124 141 115 122 128
##  [37] 102  57  35  15 188 189 185 187 169 172 147 155 151 147 150 144 140 145
##  [55] 124 106 122  62  24  26 176 162 165 166 150 171 156 169 150 131 165 113
##  [73] 155 171 134 143 126  68  28  36 154 150 153 173 166 146 151 152 153 148
##  [91] 150 156 154 137 130 129 147  88  59  35 153 170 179 145 149 144 145 146
## [109] 134 137 127 125 166 142 133 167 143 100  67  65 122 146 133 133 136 142
## [127] 135 141 128 144 148 136 153 159 155 146 161 142  85  89 108 131 128 102
## [145] 145 148 135 140 131 132 149 134 166 148 181 183 153 118 106  96 103 121
## [163] 143 131 123 142 130 155 127 122 161 149 158 153 151 147 151 107 124 135
## [181] 106 125 123 131 121 122 132 121 141 128 138 157 142 171 146 164 151 142
## [199] 129 145  85 118 116 114 145 110 120 123 140 161 147 158 163 159 164 152
## [217] 134 145 134 146  93  94 112 130 115 128 139 119 131 149 130 157 146 160
## [235] 164 155 141 140 159 172  93 118 108 101  93 129 118 122 138 164 146 153
## [253] 159 143 174 130 142 153 155 195  88 102 105 139 129 128 123 129 153 157
## [271] 140 160 120 147 154 126 115 148 179 190  78  94 119 144 138 133 148 136
## [289] 149 136 132 134 139 150 120 118 134 137 195 201 112  99  98 113 129 131
## [307] 138 142 142 141 138 139 109 104 117  97 126 190 225 235  99 100 105 120
## [325] 107 110 122 135 123 121 125 119 108 117 123 118 137 233 277 244  98  99
## [343] 107  93  93 124 124 123 125 135 106 109 110  99 123 119 158 252 275 261
## [361] 121 129 111 125 122  98 109 133 122 125 106  99  96 100 114 145 160 230
## [379] 256 233  99 115 113 131 102 127 109 114 138 107 121 111 114 112 105 144
## [397] 160 180 193 186
```

Looking at the generated heatmap, we can see that the bottom right, where schools have the highest student to teacher ratio and lowest percentage of black and hispanic are the least common. However, it appears that the most common combnation is a high student to teacher ratio and a high percentage of black and hispanic students. The heat map was used to convey the number of schools whose student to teacher ratio and black and hispanic percentage matched the x and y axes for those values. It would have been nice if there was more robust racial statistics available for the schools as more specific heatmaps could be created.

Works Cited

*NCES.PublicSchoolcharacteristics.*2017−18.*CommonCoreDataSet.https* : *//data−nces.opendata.arcgis.com/datasets/n*
public − *school* − *characteristics* − 2017 − 18/*about*