

# Comparing F&P Proficiency Levels of Bushwick and Crown Heights

Ben Foglia

6/23/2020

## Data Cleaning

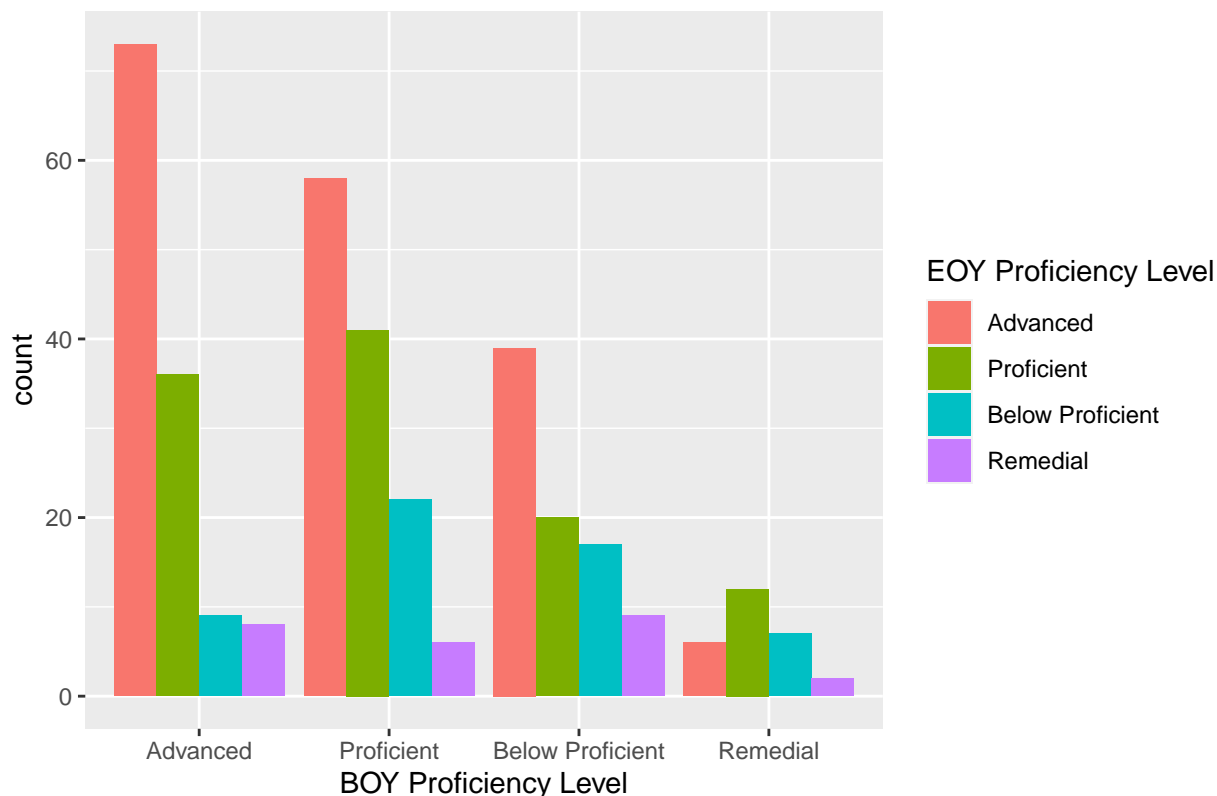
For the data cleaning process, I began by checking for duplicate rows. I found one in row 270 and removed it. I then went through each column to find irregularities in the character data, which I found in the grade level and school name columns, so I changed them to match up. I also coded in Bushwick Middle School as 1 and Crown Heights Middle School as 2 to make them easier to work with. The next step was dealing with missing values. There were 98 BOY (beginning of year) scores missing and 10 EOY (end of year) scores. Since the rate at which the EOY scores were missing was so low, I decided to remove them and only fill in values for the BOY scores using the mice package in R. After this was done, my final step was adding columns for the beginning and end of year proficiency levels (ex. if a 5th grader has a beginning score of 8, this column would have “Remedial” in their row) based on the chart below.

	Proficiency Level			
	Remedial	Below Proficient	Proficient	Advanced
4th Grade EOY / 5th Grade BOY	1-9*	10-11	12-13	14+
5th Grade EOY / 6th Grade BOY	1-11	12-13	14-15	16+
6th Grade EOY / 7th Grade BOY	1-13	14-15	16-17	18+

## Visualizations

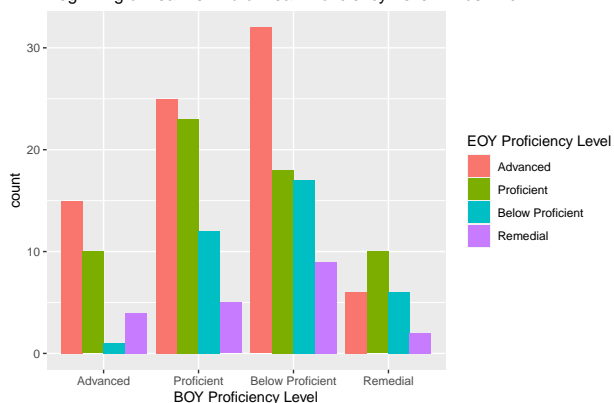
The first visualization is a bar chart counting the students in each proficiency level they were in at the end of the year based on where they were at the beginning of it (ex. the green bar furthest to the left above the “Advanced” label represents students who started the year at an advanced level and ended the year at a proficient level).

Beginning of Year vs End of Year Proficiency Level

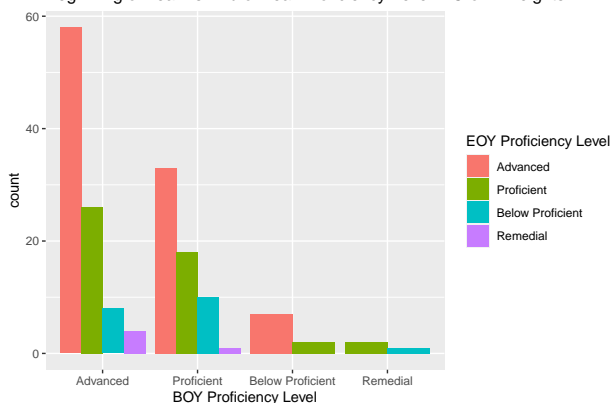


The chart shows us that a large number of students who began in the advance column stayed there. Even better, the majority of students in the proficient and below proficient groups managed to work their way up to advanced as well. Below is the same chart for each of the two schools.

Beginning of Year vs End of Year Proficiency Level – Bushwick

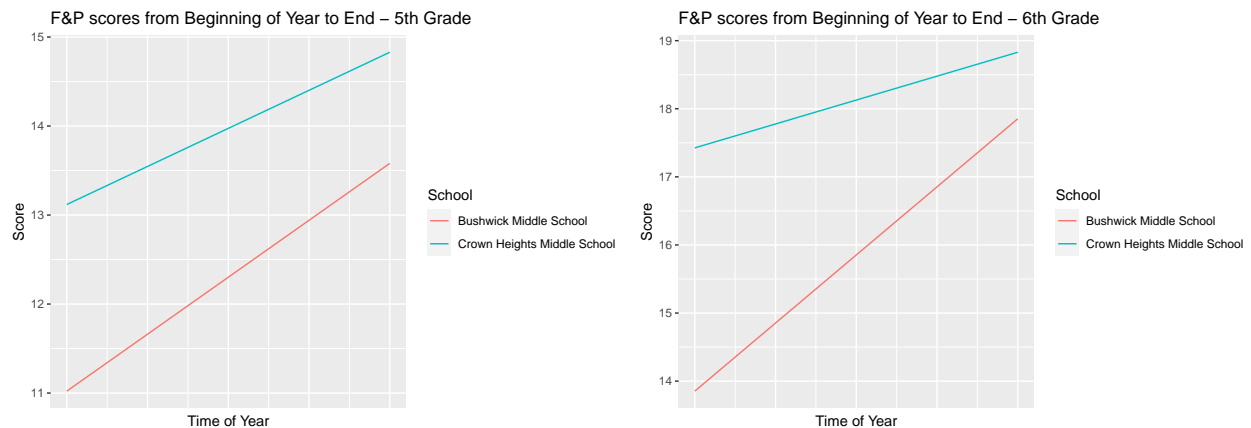


Beginning of Year vs End of Year Proficiency Level – Crown Heights



The results remain consistent for each school. Furthermore, we can see that Crown Heights had a very low number of remedial cases by the end of the year.

The final graph below compares the two school's beginning and end of year average scores for both 5th and 6th grade students in a line graph.



We can see that both grades for both schools saw an overall improvement from the beginning of the year to the end. Just as important, we can see that Bushwick has done a good job of narrowing the score gap between the two schools, with both their 5th and 6th grade students increasing their scores by at least 2.5 points on average.

## Conclusion

The visualizations above and the chart below show that both Bushwick and Crown Heights have been very successful in improving students reading level. Across the two schools, 78% of students were reading at a proficient or advanced level. If the current trends continue, these numbers will surely go up even more.

```
## # A tibble: 4 x 2
##   eoy_prof_level count
##   <fct>          <int>
## 1 Advanced         176
## 2 Proficient       109
## 3 Below Proficient  55
## 4 Remedial         25
```

## Code

```
library(readr)
library(dplyr)
library(ggplot2)
library(mice)
library(stringr)
library(knitr)

sample_data <- read_csv("F:/F&P Sample Data Set.xlsx - Sample F&P Data (1).csv")

sample_data <- sample_data %>% rename(
  student_id = 'Student ID',
  school_name = 'School Name',
  grade_level = 'Grade Level',
  boy_score = 'BOY F&P Score',
```

```

    eoy_score = 'EOY F&P Score'
  )

  duplicated(sample_data)

  sample_data[270,]
  sample_data[269,]
  sample_data <- sample_data[-270,]

  multi.unique <- function(a,b,c,d) {
    print(unique(a))
    print(unique(b))
    print(unique(c))
    print(unique(d))
  }

  multi.unique(sample_data$school_name,
               sample_data$grade_level,
               sample_data$boy_score,
               sample_data$eoy_score)

  sample_data$grade_level <- str_remove_all(sample_data$grade_level, "th")
  sample_data$school_name <- gsub("Crown Hghts Middle School", "Crown Heights Middle School",
                                sample_data$school_name)
  sample_data$school_name <- gsub("Bushwick MS", "Bushwick Middle School", sample_data$school_name)

  sample_data$school_name <- gsub("Bushwick Middle School", 1, sample_data$school_name)
  sample_data$school_name <- gsub("Crown Heights Middle School", 2, sample_data$school_name)
  sample_data$school_name <- as.numeric(sample_data$school_name)
  sample_data$student_id <- as.character(sample_data$student_id)

  sample_data[sample_data == 0] <- NA

  multi.unique(sample_data$school_name,
               sample_data$grade_level,
               sample_data$boy_score,
               sample_data$eoy_score)

  sample_data$grade_level <- as.numeric(sample_data$grade_level)

  colSums(is.na(sample_data))

  filled_values <- sample_data %>% na.omit()

  sum(sample_data$school_name == 1)
  sum(sample_data$school_name == 2)

  sum(filled_values$school_name == 1)
  sum(filled_values$school_name == 2)

  na_boy_values <- sample_data[!is.na(sample_data$eoy_score),]
  na_boy_values <- subset(na_boy_values, select = -student_id)

```

```

temp_values <- mice(na_boy_values, m = 5, maxit = 50, meth = 'pmm', seed = 500)
temp_values$imp$boy_score

filled_boy <- complete(temp_values,1)

boy_prof_levels <- function(grade_level, score){
  if((grade_level == 5 && score < 10) || (grade_level == 6 && score < 12))
    print("Remedial")
  else if((grade_level == 5 && score < 12) || (grade_level == 6 && score < 14))
    print("Below Proficient")
  else if((grade_level == 5 && score < 14) || (grade_level == 6 && score < 16))
    print("Proficient")
  else
    print("Advanced")
}

boy_prof_levels(6,19)

var1 <- filled_boy$grade_level
var2 <- filled_boy$boy_score

filled_boy$boy_prof_level <- mapply(boy_prof_levels, var1, var2)

eoy_prof_levels <- function(grade_level, score){
  if((grade_level == 5 && score < 12) || (grade_level == 6 && score < 14))
    print("Remedial")
  else if((grade_level == 5 && score < 14) || (grade_level == 6 && score < 16))
    print("Below Proficient")
  else if((grade_level == 5 && score < 16) || (grade_level == 6 && score < 18))
    print("Proficient")
  else
    print("Advanced")
}

eoy_prof_levels(5,13)

newvar2 <- filled_boy$eoy_score

filled_boy$eoy_prof_level <- mapply(eoy_prof_levels, var1,newvar2)

levels <- c("Advanced", "Proficient", "Below Proficient", "Remedial")
filled_boy$boy_prof_level <- factor(filled_boy$boy_prof_level, levels = levels)
filled_boy$eoy_prof_level <- factor(filled_boy$eoy_prof_level, levels = levels)

ggplot(filled_boy, aes(x=boy_prof_level, fill = factor(eoy_prof_level))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete(name = "EOY Proficiency Level") +
  xlab("BOY Proficiency Level") +
  ggtitle("Beginning of Year vs End of Year Proficiency Level")

bushwick <- filled_boy %>% filter(school_name == 1)
crown_heights <- filled_boy %>% filter(school_name == 2)

```

```

ggplot(bushwick, aes(x=boy_prof_level, fill = factor(eoy_prof_level))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete(name = "EOY Proficiency Level") +
  xlab("BOY Proficiency Level") +
  ggtitle("Beginning of Year vs End of Year Proficiency Level - Bushwick")

ggplot(crown_heights, aes(x=boy_prof_level, fill = factor(eoy_prof_level))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete(name = "EOY Proficiency Level") +
  xlab("BOY Proficiency Level") +
  ggtitle("Beginning of Year vs End of Year Proficiency Level - Crown Heights")

bushwick5 <- bushwick %>% filter(grade_level == 5)
bushwick6 <- bushwick %>% filter(grade_level == 6)
crown_heights5 <- crown_heights %>% filter(grade_level == 5)
crown_heights6 <- crown_heights %>% filter(grade_level == 6)

School <- c("Bushwick Middle School", "Bushwick Middle School", "Crown Heights Middle School", "Crown Heights Middle School")
TimeOfYear <- c(0, 1, 0, 1)
ScoreFifth <- c(mean(bushwick5$boy_score), mean(bushwick5$eoy_score), mean(crown_heights5$boy_score), mean(crown_heights5$eoy_score))
ScoreSixth <- c(mean(bushwick6$boy_score), mean(bushwick6$eoy_score), mean(crown_heights6$boy_score), mean(crown_heights6$eoy_score))

time_series_fifth <- data.frame(School = School, TimeOfYear = TimeOfYear, Score = ScoreFifth)
time_series_sixth <- data.frame(School = School, TimeOfYear = TimeOfYear, Score = ScoreSixth)

ggplot(time_series_fifth, aes(x = TimeOfYear, y = Score, color = factor(School))) +
  geom_line() +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  xlab("Time of Year") +
  scale_color_discrete(name = "School") +
  ggtitle("F&P scores from Beginning of Year to End - 5th Grade")
ggplot(time_series_sixth, aes(x = TimeOfYear, y = Score, color = factor(School))) +
  geom_line() +
  theme(axis.text.x = element_blank(), axis.ticks.x = element_blank()) +
  xlab("Time of Year") +
  scale_color_discrete(name = "School") +
  ggtitle("F&P scores from Beginning of Year to End - 6th Grade")

count_levels <- filled_boy %>% group_by(eoy_prof_level) %>% summarize(count = n())
count_levels

```