1   EDLD 651 Final Project

2   Anwesha Guha[1], Heidi Iwashita[1], Christopher Loan[1], Adam Nielsen[1], & Aaron Rothbart[1]

3   [1] University of Oregon

4   Author Note

₇   Abstract

₈   FILL IN ABSTRACT IF WANTED

₉   *Keywords:* keywords

₁₀   Word count: X

₁₁   EDLD 651 Final Project

Introduction

₁₃   We explore proportion of graduation (outcome), across several categorical variables.

₁₄ In particular, we plan to focus on comparisons of two groups who have historically had

₁₅ unequal access to resources: English language learners (ELL) vs. English proficient (EP)

₁₆ students & Special Education (SPED) status vs. non-SPED status.

₁₇   Not only will we report these outcomes across different groups, we will also explore

₁₈ these across boroughs, too, to see if these groups are succeeding equally across

₁₉ boroughs—-as measured by graduation outcomes—-compared to the English proficient

₂₀ students in their boroughs.

Methods

₂₂   We retrieved data collected by the Department of Education from *NYC OpenData*.

₂₃ The data contains four-year graduation outcomes for the cohorts of 2001 through 2006

₂₄ (classes of 2005 - 2010). According to the website, graduates are defined as those students

₂₅ earning either a Local or Regents diploma and exclude those earning either a special

₂₆ education (IEP) diploma or GED. Additionally, students who were in a school for less than

₂₇ five months are not included in the school cohort data. The data was last updated in April

₂₈ 2019.

Participants

₃₀   The original dataset of high school students contained 22 variables and 385 rows.

₃₁   First, we import and clean our data to begin our analyses using the `import()` and

₃₂ `here()` functions from `rio` and `here` packages.

Pivots

₃₄   The data we are starting with are already tidy, but for the purposes of demonstrating

₃₅ our rather acute proficiency in our *ability* to tidy data, in this segment will make the data

₃₆ untidy and then tidy it once more.

```r
messy_grad <- grad %>%

  pivot_wider(names_from = borough,

              values_from = total_cohort)


clean_grad <- messy_grad %>%

  pivot_longer(cols = c("Bronx":"Staten Island"),

               names_to = "borough",

               values_to = "total_cohort",

               values_drop_na = TRUE)



clean_grad <- clean_grad[, c(1,21,2,22,3:20)]
```

37   After tidying the entire dataset, we can focus on our variables of interest: enrollment

38 and graduation outcomes for specific boroughs, cohort years and student classifications.

39   Through this re-coding, and the variables recorded for each, we find that each

40 graduation outcome category (number graduated, dropped out, still enrolled) do not sum

41 to the total number of students in the cohort. This made clear the data was not as tidy as

42 it initially seemed.

43   As shown above, we mutated the data to create a new column named

44 unclassified_n, which holds the number of students in each level that are unaccounted

45 for.

46   We can see those results visually here, first by student classification:

student_characteristic	mean_grad_pct	mean_dropout_pct	mean_enrolled_pct	mean_unc
ELL	34.23714	23.10286	36.58571	
EP	61.58000	11.98000	23.31714	
SPED	24.74857	23.24000	37.77714	
Non-SPED	64.27143	11.53714	22.56571	

48

⁴⁹   The table shows that, of the student classifications, the SPED classification contains

⁵⁰ the largest percentage of unclassified students at roughly 14%. Notably, 6% of students

⁵¹ with ELL classification also are also unclassified, compared to 3% of English proficient
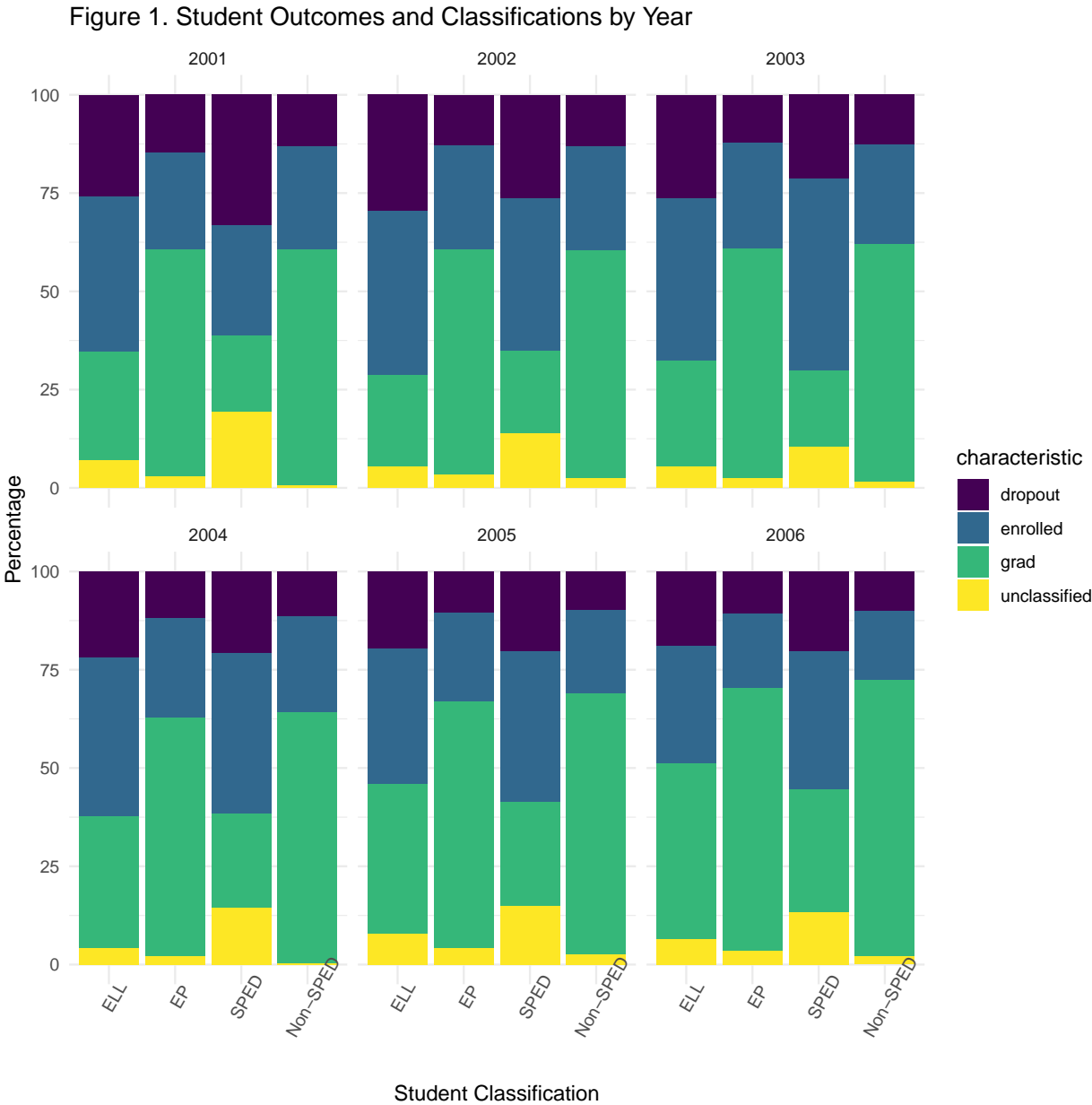
⁵² students.

⁵³   We also examine outcomes by borough here:

borough	mean_local	mean_grad_pct	mean_dropout_pct	mean_enrolled_pct	mean
Bronx	17.38571	40.28571	20.38214	31.38214	
Brooklyn	14.71429	42.75714	18.38929	33.10000	
Manhattan	15.38929	47.38214	16.79643	30.05357	
Queens	14.70714	46.96429	17.94286	29.92500	
Staten Island	16.89643	53.65714	13.81429	25.84643	

⁵⁶   These unclassified students were not concentrated in any one year, shown in the table

⁵⁷ here:

cohort	mean_local	mean_grad_pct	mean_dropout_pct	mean_enrolled_pct	mean_uncla
2001	17.805	41.360	21.8100	29.4950	
2002	16.015	40.005	20.4750	33.2950	
2003	15.130	41.555	18.1950	35.3800	
2004	16.080	45.485	16.5700	32.6100	
2005	15.770	48.520	15.0100	29.0750	
2006	14.965	53.270	15.0975	25.2875	

⁶⁰   Additional use of the `pivot_longer()` function allowed us to further tidy the data to

⁶¹ create a **graduation outcomes** variable, so we could see the average percentage of

⁶² students for each graduation outcome, by student classification (see Figure 1).

Figure 1. Student Outcomes and Classifications by Year

Data analysis

All analysis were conducted in R, with heavy reliance upon the `{tidyverse}`
packages to manipulate and visualize the data.

Results

Visual inspection of Figure 2 demonstrates heterogeneity in the gaps between ELL
and EP students. Certain Boroughs (e.g., Staten Island) have much larger gaps between
proficiency than others (e.g., Brooklyn). This is most easily seen from difference in the

71 dashed lines which represent the means of each group within a given Borough.

72   From an equity framework, it is concerning that this gap is not equal across

73 boroughs. Ideally, any difference in graduation based on EL status between students would

74 be constant despite location. Without follow-ups to gather more qualitative (or highly

75 dimensional quantitative) data, it is difficult to explain why Boroughs with higher EP

76 graduation rates do not have correspondingly higher ELL graduation rates.

Figure 2. Graduation Rates in NYC by EL Status

77

78   Visual inspection of Figure 3 suggests less variability in the difference between

79 average graduation rates of SPED vs. non-SPED students across Boroughs. Unlike ELL

80 vs. EP students, SPED students appear to succeed at rates relative to their Borough. This

81 may suggest support for SPED students increases proportionately to non-SPED students,

82 which is essential from an equity framework.

83   This trend is evident when we compare the difference between these groups across

84 districts (e.g., Staten Island & Manhattan) which perform well. Visually this looks like a

85 shift in both mean lines, rather than in increase only for non-SPED students.

Figure 3. Graduation Rates in NYC by SPED Status
Dashed lines represent Borough average for each group

86
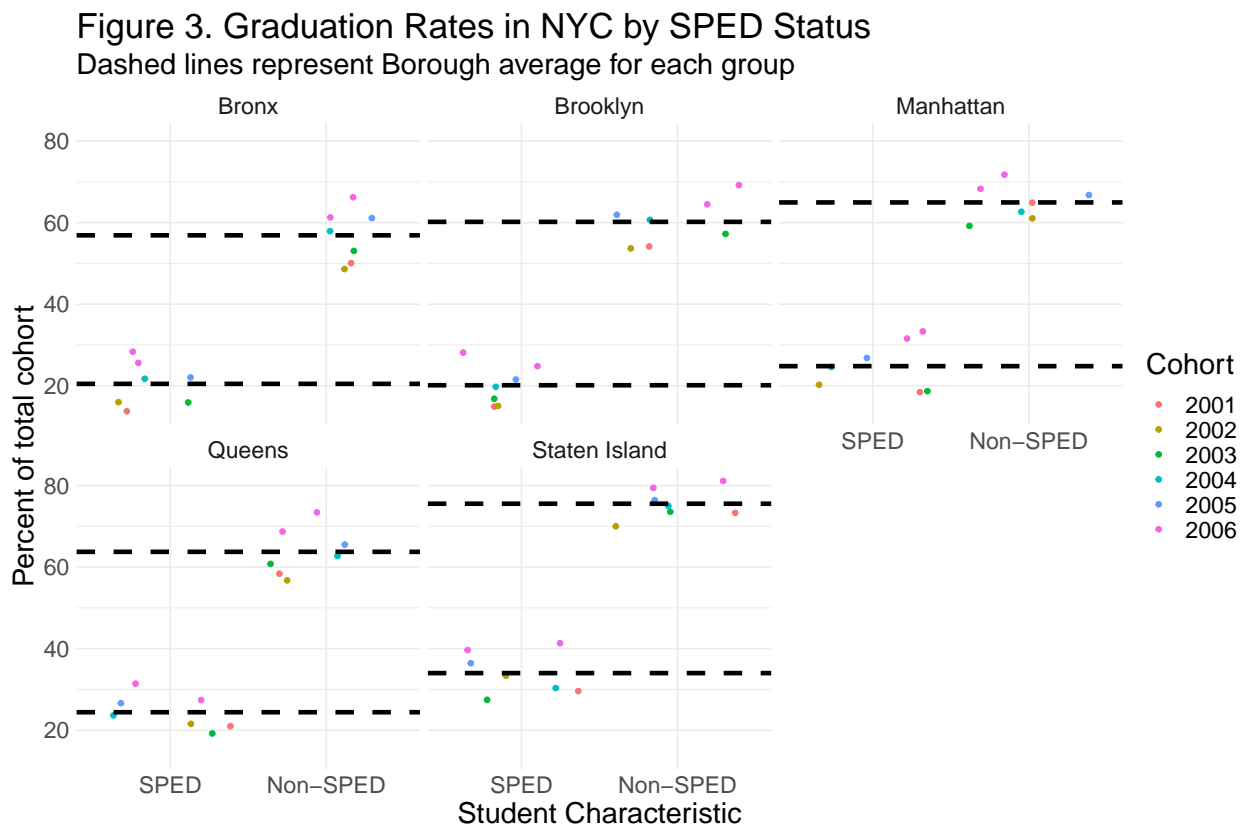
For in both visual analyses——though some variability over years is
evident——annual differences are much smaller than the differences between groups (and
in some cases differences between Boroughs). Overall, though, graduation rates seem to
not be too different over time for either group comparison.

Discussion

This project leveraged public data to determine differential graduation outcomes
across student classification status. Specifically, we compared differences between (i) ELL
and EP as well as (ii) SPED and non-SPED across boroughs. We incorporated several
years of data and visualized these as a jittered scatter plot with mean lines at each
Borough and student-classification Cluster to clearly identify differences between groups.

As this was a purely descriptive analysis, we recommend inferential statistics to
explore the significance of these group differences. Furthermore, we suggest future research
incorporate experts and educational theorists with the purpose of explaining these

differences. Through greater explanation and exploration, we hope differences in support
for ELL vs. EP students across borough can be minimized in a way that all students are
supported adequately.

We pose a few considerations in an exploration of potentially confounding variables
which explain the variation in these averages (particularly for ELL/EP). These include but
are not limited to: unequal access to resources by all portions of the school, percent of
teachers that speak languages other than English, predominant non-English language(s)
is/are spoken in each borough, Average qualification of SPED/ELL teachers, and parental
access to (and leverage of) extra-curricular educational resources.

109

References