EDLD 651 Final Project

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

[1] University of Oregon

Author Note

All work done herein represents contributions from all authors equally. Author order is alphabetical.

Abstract

FILL IN ABSTRACT IF WANTED

Keywords: keywords

Word count: X

EDLD 651 Final Project

Introduction

Now more than ever, it is crucial to address systematic disparities in education. Though there has been modest improvement in recent years (Disability Scoop, 2019), high school graduation rates still are significantly lower in specific groups who have historically received unequal access to resources, in particular students with disabilities and students with limited English proficiency (Burke, 2015). This is especially troubling considering the lower rates of employment among adults who do not complete high school (Schifter, 2016). Approximately 67% of students receiving special education services in the U.S. obtain their high school diploma, and only 66% of English learners graduate on time from secondary education (Johnson, 2019; National Center for Education Statistics, n.d.). These percentages fall short of the approximately 85% of all high school students who graduate (National Center for Education Statistics, n.d.).

In light of these concerns, we decided to explore high school graduation rates for students with disabilities and with limited English proficiency as compared to general education populations. This data can be assessed to identify systematic breakdowns and disparities impacting marginalized student populations. In particular, we wanted to see if the national data was represented regionally by looking at data from New York City schools. Furthermore, to examine the impact of geographical disparities in access to educational resources, this paper will also examine graduation outcomes across New York City boroughs.

Methods

We retrieved data collected by the Department of Education from *NYC OpenData*(NYC Open Data, 2019). The data contains four-year graduation outcomes for the cohorts of 2001 through 2006 (classes of 2005 - 2010). According to the website, graduates are defined as those students earning either a Local or Regents diploma and exclude those earning either a special education (IEP) diploma or GED. Additionally, students who were in a school for less than five months are not included in the school cohort data. The data was last updated in April 2019.

Participants

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

First, we import and clean our data to begin our analyses using the `import()` and `here()` functions from `rio` and `here` packages.

Pivots

The data we are starting with are already tidy, but for the purposes of demonstrating our rather acute proficiency in our *ability* to tidy data, in this segment will make the data untidy and then tidy it once more.

```
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)]
```

After tidying the entire dataset, we can focus on our variables of interest: enrollment and graduation outcomes for specific boroughs, cohort years and student classifications.

Through this re-coding, and the variables recorded for each, we find that each graduation outcome category (number graduated, dropped out, still enrolled) do not sum to the total number of students in the cohort. This made clear the data was not as tidy as it initially seemed.

As shown above, we mutated the data to create a new column named `unclassified_n`, which holds the number of students in each level that are unaccounted for.

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	

.

The table shows that, of the student classifications, the SPED classification contains the largest percentage of unclassified students at roughly 14%. Notably, 6% of students with ELL classification also are also unclassified, compared to 3% of English proficient students.

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	

.

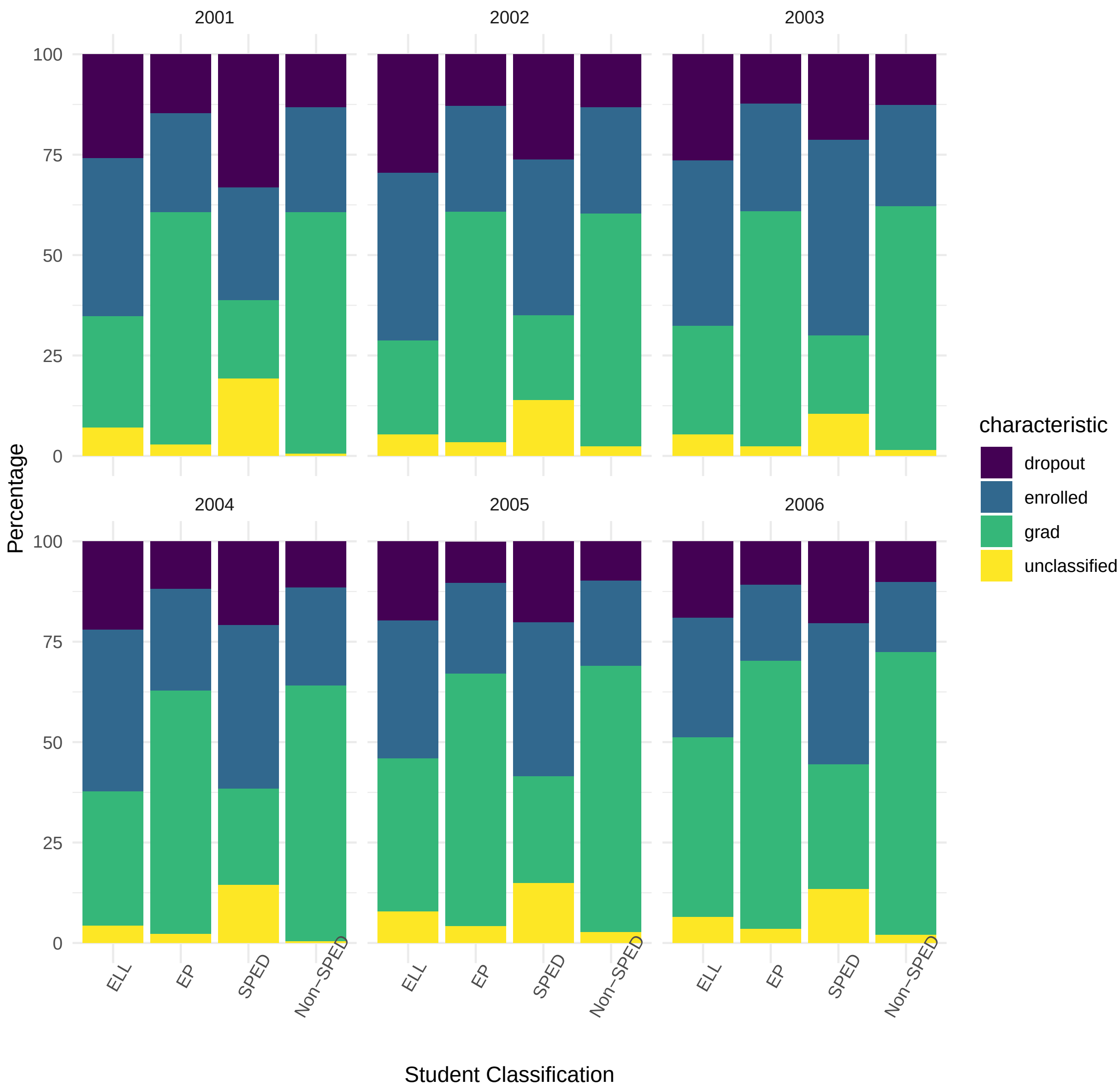
These unclassified students were not concentrated in any one year, shown in the table here:

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	

.

Additional use of the `pivot_longer()` function allowed us to further tidy the data to create a **graduation outcomes** variable, so we could see the average percentage of 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. We used the following R versions and packages for this project: R (Version 4.0.2; R Core Team, 2020) and the R-packages *dplyr* (Version 1.0.2; Wickham et al., 2020), *forcats* (Version 0.5.0; Wickham, 2020a), *ggplot2* (Version 3.3.2; Wickham, 2016), *here* (Version 1.0.0; Müller, 2017), *janitor* (Version 2.0.1; Firke, 2020), *kableExtra* (Version 1.3.1; Zhu, 2020), *knitr* (Version 1.30; Xie, 2015), *papaja*

(Version 0.1.0.9997; Aust & Barth, 2020), *purrr* (Version 0.3.4; Henry & Wickham, 2020), *readr* (Version 1.4.0; Wickham, Hester, & Francois, 2018), *rio* (Version 0.5.16; Chan, Chan, Leeper, & Becker, 2018), *stringr* (Version 1.4.0; Wickham, 2019), *tibble* (Version 3.0.4; Müller & Wickham, 2020), *tidyr* (Version 1.1.2; Wickham, 2020b), and *tidyverse* (Version 1.3.0; Wickham, Averick, et al., 2019).

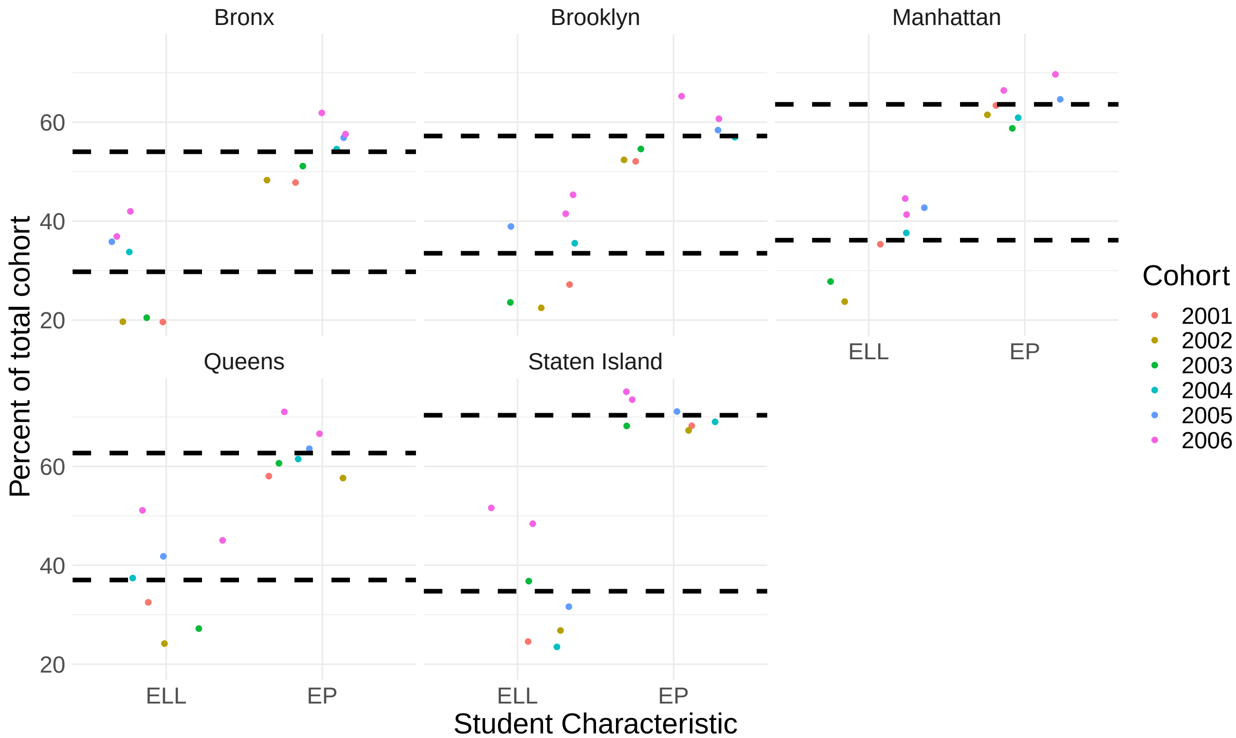
Results

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

From an equity framework, it is concerning that this gap is not equal across boroughs. Ideally, any difference in graduation based on EL status between students would be constant despite location. Without follow-ups to gather more qualitative (or highly dimensional quantitative) data, it is difficult to explain why Boroughs with higher EP graduation rates do not have correspondingly higher ELL graduation rates.

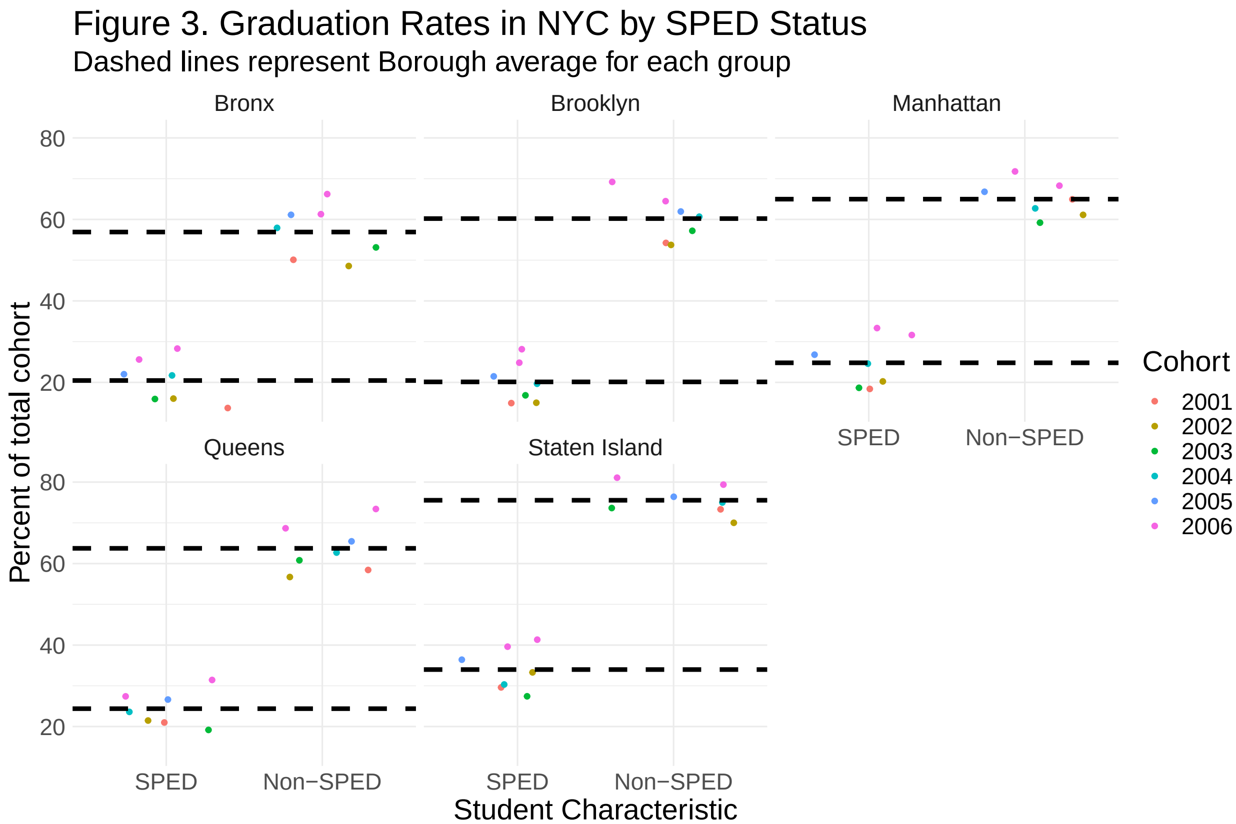
Figure 2. Graduation Rates in NYC by EL Status

Dashed lines represent Borough average for each group



Visual inspection of Figure 3 suggests less variability in the difference between average graduation rates of SPED vs. non-SPED students across Boroughs. Unlike ELL vs. EP students, SPED students appear to succeed at rates relative to their Borough. This may suggest support for SPED students increases proportionately to non-SPED students, which is essential from an equity framework.

This trend is evident when we compare the difference between these groups across districts (e.g., Staten Island & Manhattan) which perform well. Visually this looks like a shift in both mean lines, rather than in increase only for non-SPED students.



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.

References

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Burke, A. (2015). Early identification of high school graduation outcomes in oregon leadership network schools. *Regional Educational Laboratory Northwest.*

Chan, C.-h., Chan, G. C., Leeper, T. J., & Becker, J. (2018). *Rio: A swiss-army knife for data file i/o.*

Disability Scoop. (2019). *Graduation rate for students with disabilities shows improvement.* https://www.disabilityscoop.com/2019/01/30/graduation-rate-shows-improvement/25962/#:~:text=For%20the%202016%2D2017%20school,that%20the%20rate%20has%20increased.

Firke, S. (2020). *Janitor: Simple tools for examining and cleaning dirty data.* Retrieved from https://CRAN.R-project.org/package=janitor

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools.* Retrieved from https://CRAN.R-project.org/package=purrr

Johnson, A. (2019). The effects of english learner classification on high school graduation and college attendance. *AERA Open*, *5*(2), 2332858419850801.

Müller, K. (2017). *Here: A simpler way to find your files.* Retrieved from https://CRAN.R-project.org/package=here

Müller, K., & Wickham, H. (2020). *Tibble: Simple data frames.* Retrieved from https://CRAN.R-project.org/package=tibble

National Center for Education Statistics. (n.d.). *Table 1. Public high school 4-year adjusted cohort graduation rate (acgr), by race/ethnicity and selected demographic characteristics for the united states, the 50 states, and the district of columbia: School year 2016–17.* https://nces.ed.gov/ccd/tables/ACGR_RE_and_characteristics_2016-17.asp.

NYC Open Data. (2019). *2005-2010 graduation outcomes - by borough.*

https://data.cityofnewyork.us/Education/2005-2010-Graduation-Outcomes-By-Borough/avir-tzek.

R Core Team. (2020). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Schifter, L. A. (2016). Using survival analysis to understand graduation of students with disabilities. *Exceptional Children*, *82*(4), 479–496.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H. (2020a). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2020b). *Tidyr: Tidy messy data.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Francois, R. (2018). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://yihui.org/knitr/

Zhu, H. (2020). *KableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra