

EDLD 651 Final Project Draft

Anwesha Guha<sup>1</sup>, Heidi Iwashita<sup>1</sup>, Christopher Loan<sup>1</sup>, Adam Nielsen<sup>1</sup>, & Aaron Rothbart<sup>1</sup>

<sup>1</sup> 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 Draft

**Introduction**

We explore proportion of graduation (outcome), across several categorical variables. In particular, we plan to focus on comparisons of two groups who have historically had unequal access to resources: English language learners (ELL) vs. English proficient (EP) students & Special Education (SPED) status vs. non-SPED status.

Not only will we report these outcomes across different groups, we will also explore these across boroughs, too, to see if these groups are succeeding equally across boroughs—as measured by graduation outcomes—compared to the English proficient students in their boroughs.

**Methods**

We retrieved the data collected by the Department of Education from  
Information about variables, how they were measured here  
Information about regents examinations here

**Participants**

Explain participants' from what we have in data.

First, we import and clean our data:

```
##      Length      Class      Mode
##      385 character character
## # A tibble: 6 x 22
##   demographic borough cohort total_cohort total_grads_n total_grads_per~
##   <chr>          <chr>    <dbl>         <int>         <int>         <dbl>
## 1 Borough To~ Bronx      2001         11453          4913          42.9
## 2 Borough To~ Bronx      2002         12032          5328          44.3
## 3 Borough To~ Bronx      2003         13632          6389          46.9
## 4 Borough To~ Bronx      2004         14364          7448          51.9
```

```

37 ## 5 Borough To~ Bronx      2005      15175      8229      54.2
38 ## 6 Borough To~ Bronx      2006      15579      8524      54.7
39 ## # ... with 16 more variables: total_regents_n <int>,
40 ## #   total_regents_percent_of_cohort <dbl>,
41 ## #   total_regents_percent_of_grads <dbl>, advanced_regents_n <int>,
42 ## #   advanced_regents_percent_of_cohort <dbl>,
43 ## #   advanced_regents_percent_of_grads <dbl>, regents_w_o_advanced_n <int>,
44 ## #   regents_w_o_advanced_percent_of_cohort <dbl>,
45 ## #   regents_w_o_advanced_percent_of_grads <dbl>, local_n <int>,
46 ## #   local_percent_of_cohort <dbl>, local_percent_of_grads <dbl>,
47 ## #   still_enrolled_n <int>, still_enrolled_percent_of_cohort <dbl>,
48 ## #   dropped_out_n <int>, dropped_out_percent_of_cohort <dbl>

```

## 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)]

```

Now that we have tidied the entire dataset, we can focus on our variables of interest:  
enrollment and graduation for specific boroughs, cohorts and demographics.

```
## # A tibble: 6 x 16
```

```
##   demographic borough cohort total_cohort total_grads_n total_grads_per~ local_n
```

```
##   <chr>          <chr>    <dbl>          <int>          <int>          <dbl>    <int>
```

```
## 1 English La~ Bronx    2001          1984          388          19.6    311
```

```
## 2 English La~ Bronx    2002          1693          333          19.7    257
```

```
## 3 English La~ Bronx    2003          1905          391          20.5    296
```

```
## 4 English La~ Bronx    2004          1894          640          33.8    426
```

```
## 5 English La~ Bronx    2005          1940          694          35.8    377
```

```
## 6 English La~ Bronx    2006          2143          791          36.9    395
```

```
## # ... with 9 more variables: local_percent_of_cohort <dbl>,
```

```
## #   local_percent_of_grads <dbl>, still_enrolled_n <int>,
```

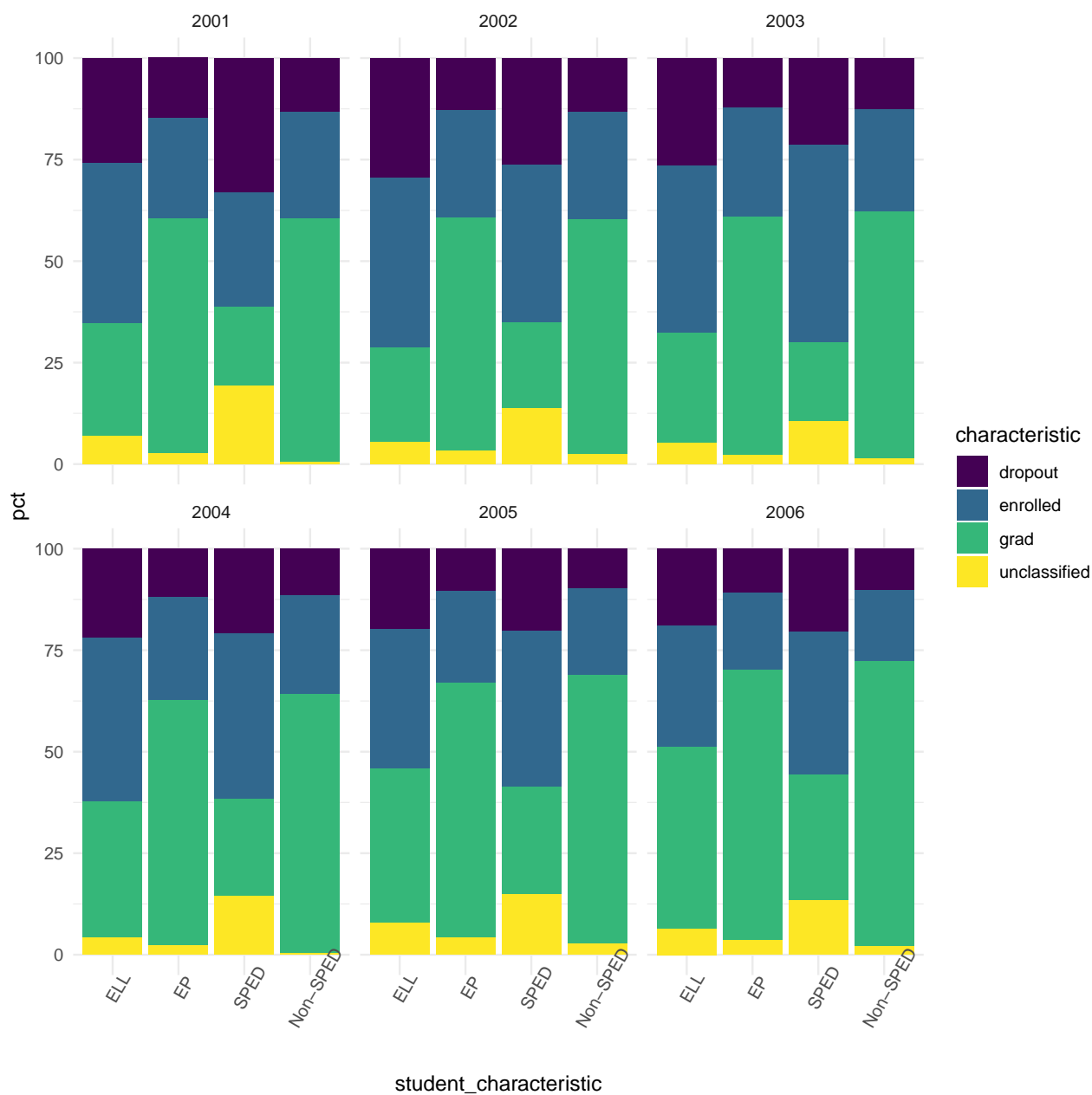
```
## #   still_enrolled_percent_of_cohort <dbl>, dropped_out_n <int>,
```

```
## #   dropped_out_percent_of_cohort <dbl>, student_characteristic <fct>,
```

```
## #   unclassified_n <int>, unclassified_percent_of_cohort <dbl>
```

# NEW YORK CITY GRADUATION OUTCOME BY BOROUGH AND STUDENT CLASSIFICATION

6



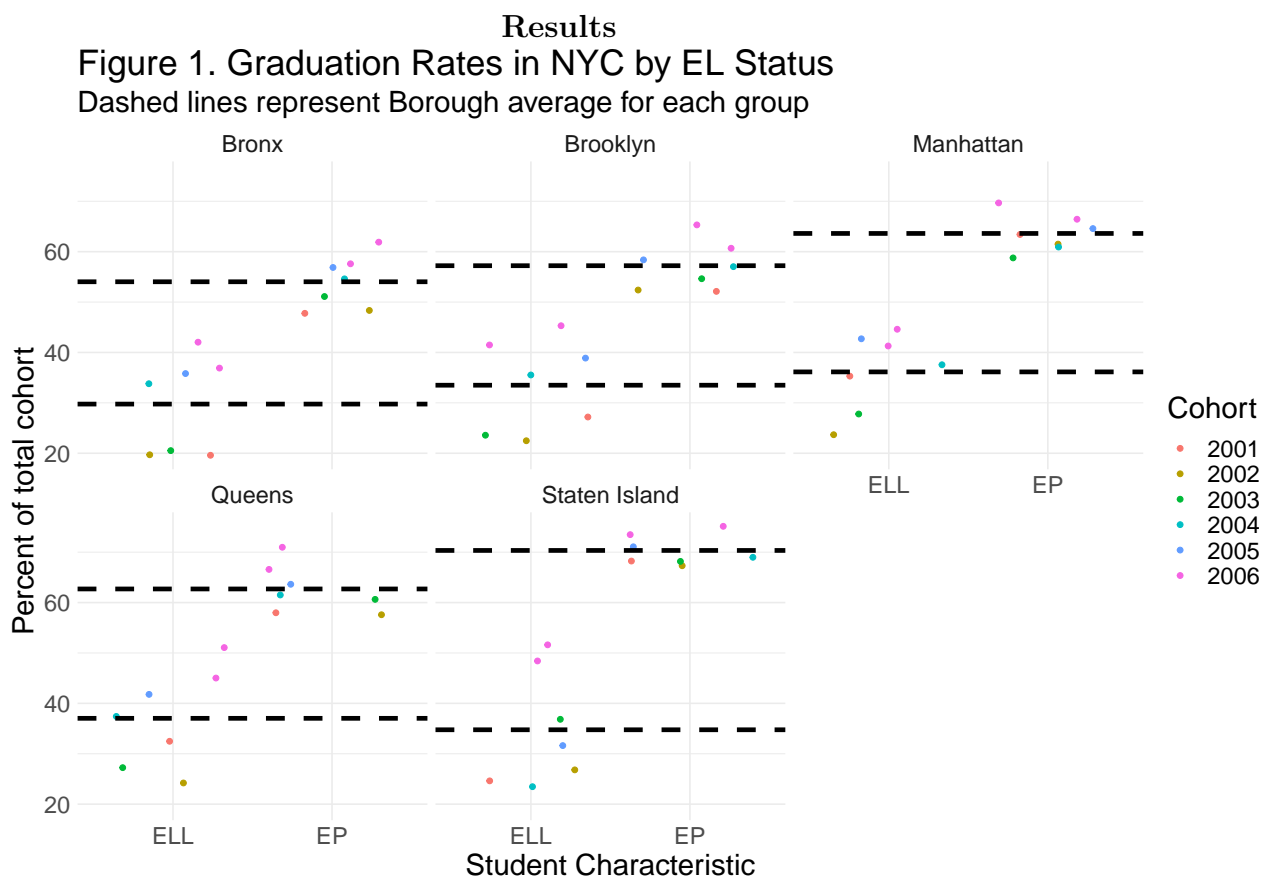
69

## Data analysis

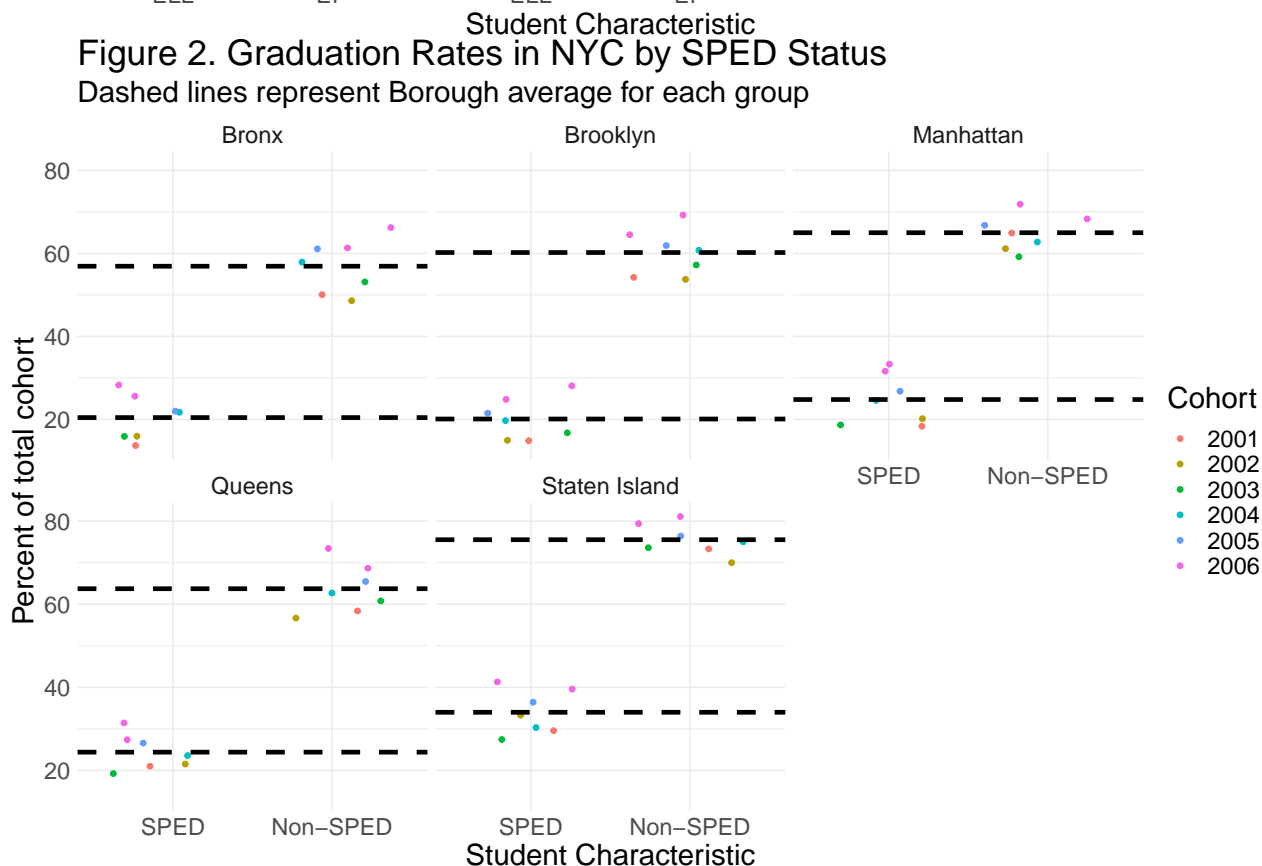
All analysis were conducted in R, with heavy reliance upon the `{tidyverse}` packages to manipulate and visualize the data.

72

73



74



75

## Discussion

76

77 Differences appear to be blah by blah for blah. XYZ boroughs should consider blah

78 blah blah, based on the results. Inferential tests are recommended for next directions.



## References