

EDLD 651 Final Project Draft

Anwesha Guha¹, Heidi Iwashita¹, Christopher Loan¹, Adam Nielsen¹, & Aaron Rothbart¹

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 Draft

Introduction

We explore proportion of graduation (outcome), across several categorical variables. In particular, we plan to focus on English learners vs. English proficient students.

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

Methods

Data was taken from INSERT LINK

Participants

Explain participants' from what we have in data.

```
#clean names done here
grad <- import(here("data", "2005-2010__Graduation_Outcomes_-_By_Borough.csv"))
grad <- grad %>%
  clean_names() %>%
  as_tibble()

summary(grad$cohort) # needs to be cleaned in new df, change Aug 2006 to 2006
```

```
##      Length      Class      Mode
##      385 character character
```

```
clean_grad <- grad
clean_grad$cohort <- as.numeric(sub("Aug 2006", "2006", grad$cohort))

clean_grad
```

```
## # A tibble: 385 x 22
```

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

```

26 ##      <chr>      <chr>      <dbl>      <int>      <int>      <dbl>
27 ##  1 Borough To~ Bronx      2001      11453      4913      42.9
28 ##  2 Borough To~ Bronx      2002      12032      5328      44.3
29 ##  3 Borough To~ Bronx      2003      13632      6389      46.9
30 ##  4 Borough To~ Bronx      2004      14364      7448      51.9
31 ##  5 Borough To~ Bronx      2005      15175      8229      54.2
32 ##  6 Borough To~ Bronx      2006      15579      8524      54.7
33 ##  7 Borough To~ Bronx      2006      15579      9215      59.2
34 ##  8 Borough To~ Brookl~    2001      19961      9758      48.9
35 ##  9 Borough To~ Brookl~    2002      20808     10337      49.7
36 ## 10 Borough To~ Brookl~    2003      21334     11064      51.9
37 ## # ... with 375 more rows, and 16 more variables: total_regents_n <int>,
38 ## #   total_regents_percent_of_cohort <dbl>,
39 ## #   total_regents_percent_of_grads <dbl>, advanced_regents_n <int>,
40 ## #   advanced_regents_percent_of_cohort <dbl>,
41 ## #   advanced_regents_percent_of_grads <dbl>, regents_w_o_advanced_n <int>,
42 ## #   regents_w_o_advanced_percent_of_cohort <dbl>,
43 ## #   regents_w_o_advanced_percent_of_grads <dbl>, local_n <int>,
44 ## #   local_percent_of_cohort <dbl>, local_percent_of_grads <dbl>,
45 ## #   still_enrolled_n <int>, still_enrolled_percent_of_cohort <dbl>,
46 ## #   dropped_out_n <int>, dropped_out_percent_of_cohort <dbl>

```

47 PIVOTS

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

```

messy_grad <- clean_grad %>%
  pivot_wider(names_from = borough,
              values_from = total_cohort)
messy_grad

```

```

51 ## # A tibble: 385 x 25
52 ##   demographic cohort total_grads_n total_grads_per~ total_regents_n
53 ##   <chr>          <dbl>          <int>          <dbl>          <int>
54 ## 1 Borough To~    2001            4913            42.9            2644
55 ## 2 Borough To~    2002            5328            44.3            3118
56 ## 3 Borough To~    2003            6389            46.9            3861
57 ## 4 Borough To~    2004            7448            51.9            4625
58 ## 5 Borough To~    2005            8229            54.2            5618
59 ## 6 Borough To~    2006            8524            54.7            6312
60 ## 7 Borough To~    2006            9215            59.2            6605
61 ## 8 Borough To~    2001            9758            48.9            6177
62 ## 9 Borough To~    2002           10337            49.7            7050
63 ## 10 Borough To~   2003           11064            51.9            7711
64 ## # ... with 375 more rows, and 20 more variables:
65 ## #   total_regents_percent_of_cohort <dbl>,
66 ## #   total_regents_percent_of_grads <dbl>, advanced_regents_n <int>,
67 ## #   advanced_regents_percent_of_cohort <dbl>,
68 ## #   advanced_regents_percent_of_grads <dbl>, regents_w_o_advanced_n <int>,
69 ## #   regents_w_o_advanced_percent_of_cohort <dbl>,
70 ## #   regents_w_o_advanced_percent_of_grads <dbl>, local_n <int>,
71 ## #   local_percent_of_cohort <dbl>, local_percent_of_grads <dbl>,
72 ## #   still_enrolled_n <int>, still_enrolled_percent_of_cohort <dbl>,
73 ## #   dropped_out_n <int>, dropped_out_percent_of_cohort <dbl>, Bronx <int>,

```

74 **## #** Brooklyn <int>, Manhattan <int>, Queens <int>, `Staten Island` <int>

```
clean_grad_2 <- messy_grad %>%
  pivot_longer(cols = c("Bronx":"Staten Island"),
               names_to = "borough",
               values_to = "total_cohort",
               values_drop_na = TRUE)

clean_grad_2 <- clean_grad_2[, c(1,21,2,22,3:20)]
apa_table(clean_grad_2)
```

#select() relevant variables to make subsetting dataset

#filter() out cases that are of interest

#descriptive stats (counts of demographics reported by borough in a table() call?) we

75 Data analysis

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

78 Results

```
#group_by()
#summarize()
#report graduation by borough
#report graduation by english language status
#report graduation by borough & english learner status
```

#Chris Loan would like to do this part: graphing.

my code works assuming we use the "clean_grad" dataset in the `final_project.Rmd` fi

#graph outcomes by English language status

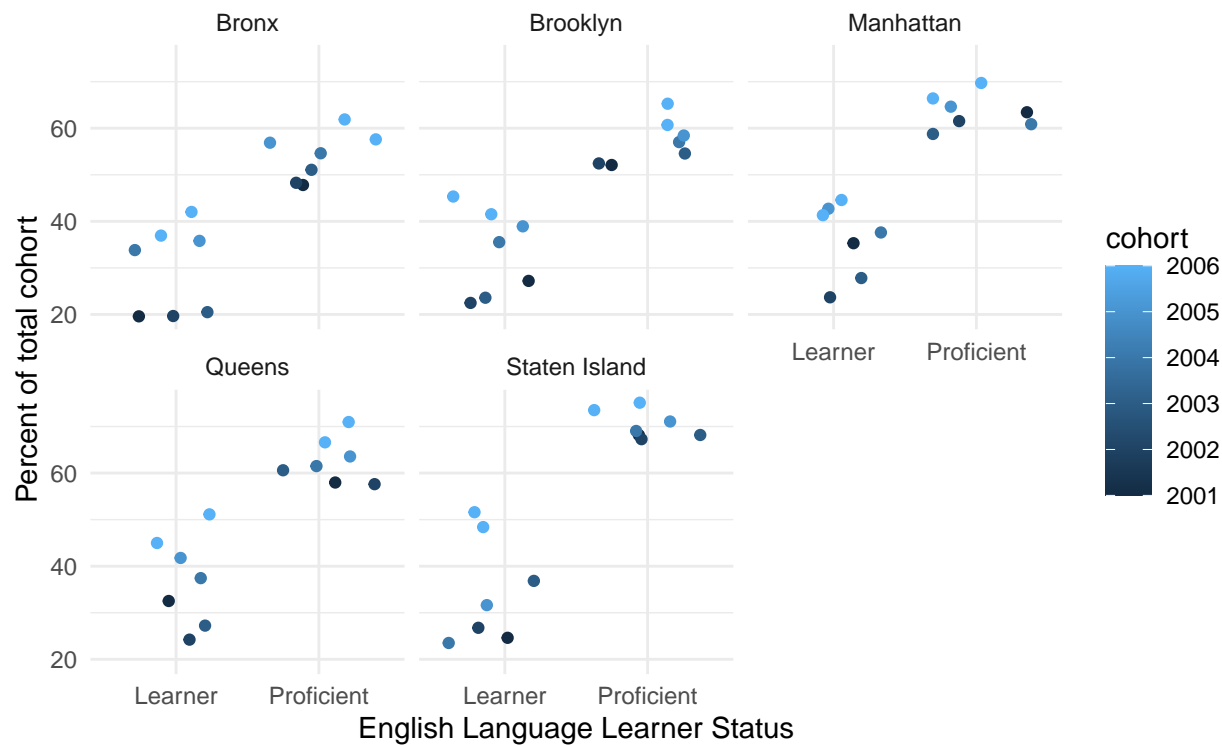
```
#facet wrap by borough

#jitter the points so we can see all the years

#give color to all the years so we can differentiate them

clean_grad %>%
  filter(demographic == "English Language Learners" |
         demographic == "English Proficient Students") %>%
  mutate(`English Language Learner Status` =
         factor(demographic,
               levels = c("English Language Learners",
                          "English Proficient Students"),
               labels = c('Learner', 'Proficient')
         )
  ) %>% group_by(`English Language Learner Status`, borough) %>%
  ggplot(aes(x = `English Language Learner Status`,
             y = total_grads_percent_of_cohort)) +
  geom_jitter(aes(color = cohort)) + facet_wrap(~borough) +
  labs(title = 'Figure 1. Graduation Rates in NYC by English Learner Status',
       subtitle = 'Boroughs are reported separately with lighter dots indicating more re
       y = 'Percent of total cohort')
```

Figure 1. Graduation Rates in NYC by English Learner Status
Boroughs are reported separately with lighter dots indicating more recent years



Discussion

Differences appear to be blah by blah for blah. XYZ boroughs should consider blah
blah blah, based on the results. Inferential tests are recommended for next directions.

References

Table 1

demographic	borough	cohort	total_cohort	total_grads_n	total_grads_per
Borough Total	Bronx	2,001.00	11453	4913	42.90
Borough Total	Bronx	2,002.00	12032	5328	44.30
Borough Total	Bronx	2,003.00	13632	6389	46.90
Borough Total	Bronx	2,004.00	14364	7448	51.90
Borough Total	Bronx	2,005.00	15175	8229	54.20
Borough Total	Bronx	2,006.00	15579	8524	54.70
Borough Total	Bronx	2,006.00	15579	9215	59.20
Borough Total	Brooklyn	2,001.00	19961	9758	48.90
Borough Total	Brooklyn	2,002.00	20808	10337	49.70
Borough Total	Brooklyn	2,003.00	21334	11064	51.90
Borough Total	Brooklyn	2,004.00	22353	12303	55.00
Borough Total	Brooklyn	2,005.00	22331	12603	56.40
Borough Total	Brooklyn	2,006.00	22177	13040	58.80
Borough Total	Brooklyn	2,006.00	22177	14043	63.30
Borough Total	Manhattan	2,001.00	12670	7480	59.00
Borough Total	Manhattan	2,002.00	13463	7746	57.50
Borough Total	Manhattan	2,003.00	13879	7613	54.90
Borough Total	Manhattan	2,004.00	15127	8780	58.00
Borough Total	Manhattan	2,005.00	15843	9816	62.00
Borough Total	Manhattan	2,006.00	16416	10411	63.40
Borough Total	Manhattan	2,006.00	16416	10947	66.70
Borough Total	Queens	2,001.00	17011	9180	54.00
Borough Total	Queens	2,002.00	18262	9869	54.00
Borough Total	Queens	2,003.00	18415	10455	56.80
Borough Total	Queens	2,004.00	18725	10922	58.30
Borough Total	Queens	2,005.00	19511	11863	60.80