Final Project

Anwesha Guha, Heidi Iwashita, Chris Loan, Adam Nielsen & Aaron Rothbart

15 November, 2020

```
grad <- import(here("data", "2005-2010__Graduation_Outcomes_-__By_Borough.csv"))
grad <- grad %>%
  clean_names() %>%
  as_tibble()
grad
```

```
## # A tibble: 385 x 22
##    demographic borough cohort total_cohort total_grads_n total_grads_per~
##    <chr>       <chr>   <chr>         <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
##  5 Borough To~ Bronx   2005          15175          8229             54.2
##  6 Borough To~ Bronx   2006          15579          8524             54.7
##  7 Borough To~ Bronx   Aug 2~        15579          9215             59.2
##  8 Borough To~ Brookl~ 2001          19961          9758             48.9
##  9 Borough To~ Brookl~ 2002          20808         10337             49.7
## 10 Borough To~ Brookl~ 2003          21334         11064             51.9
## # ... with 375 more rows, and 16 more variables: total_regents_n <int>,
## #   total_regents_percent_of_cohort <dbl>,
## #   total_regents_percent_of_grads <dbl>, advanced_regents_n <int>,
## #   advanced_regents_percent_of_cohort <dbl>,
## #   advanced_regents_percent_of_grads <dbl>, regents_w_o_advanced_n <int>,
## #   regents_w_o_advanced_percent_of_cohort <dbl>,
## #   regents_w_o_advanced_percent_of_grads <dbl>, local_n <int>,
## #   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>
```

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.

```
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~
##    <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
##  5 Borough To~ Bronx     2005        15175          8229             54.2
##  6 Borough To~ Bronx     2006        15579          8524             54.7
##  7 Borough To~ Bronx     2006        15579          9215             59.2
##  8 Borough To~ Brookl~   2001        19961          9758             48.9
##  9 Borough To~ Brookl~   2002        20808         10337             49.7
## 10 Borough To~ Brookl~   2003        21334         11064             51.9
## # ... with 375 more rows, and 16 more variables: total_regents_n <int>,
## #   total_regents_percent_of_cohort <dbl>,
## #   total_regents_percent_of_grads <dbl>, advanced_regents_n <int>,
## #   advanced_regents_percent_of_cohort <dbl>,
## #   advanced_regents_percent_of_grads <dbl>, regents_w_o_advanced_n <int>,
## #   regents_w_o_advanced_percent_of_cohort <dbl>,
## #   regents_w_o_advanced_percent_of_grads <dbl>, local_n <int>,
## #   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>
```

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

```
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 separetely with lighter dots indicating more recent years',
     y = 'Percent of Total Cohort')
```