

# Final

Molly Bratton and Audrey Moyer

## Reading in datasets

```
library(ggplot2)
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## vforcats   1.0.0     v stringr   1.5.1
## v lubridate 1.9.3     v tibble    3.2.1
## v purrr    1.0.2     v tidyverse  1.3.1
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
#reading in the datasets
public_schools <- read_csv("~/Desktop/RStudio/STAT_220/quizzes and portfolio projects/final-beast-mode/public-schools.csv")

## Rows: 101390 Columns: 77
## -- Column specification -----
## Delimiter: ","
## chr (23): NCESSCH, SURVYEAR, STABR, LEAID, ST_LEAID, LEA_NAME, SCH_NAME, LST...
## dbl (54): X, Y, OBJECTID, STATUS, TOTFRL, FRELCH, REDLCH, DIRECTCERT, PK, KG...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.

neighborhood_poverty <- read_csv("~/Desktop/RStudio/STAT_220/quizzes and portfolio projects/final-beast-mode/neighborhood-poverty.csv")

## Rows: 100682 Columns: 10
## -- Column specification -----
## Delimiter: ","
## chr (3): NCESSCH, NAME, SCHOOLYEAR
## dbl (7): X, Y, OBJECTID, IPR_EST, IPR_SE, LAT, LON
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

## Research Questions

1. Does a higher school neighborhood poverty estimate correlate with a higher or lower free lunch program at schools?
2. Are more densely populated areas more likely to have a higher or lower teacher to student ratio?

3. Does the school neighborhood poverty estimate correlate to the percentage/ratio of black students of a school (or non-white students)?

4. Does a higher student to teacher ratio correlate with higher or lower free lunch programs at schools?

5. Does the male to female ratio of students correlate with a higher or lower school neighborhood poverty estimate?

*#need to join the datasets - they are lined up in the same order (checked by school name and NCESSCH)*

```
schools_poverty <- public_schools %>%
  left_join(neighborhood_poverty, by = "NCESSCH")
```

Does a higher school neighborhood poverty estimate correlate with a higher or lower free lunch program at schools?

```
ggplot(schools_poverty, aes(x = IPR_EST, y = TOTFRL)) +
  geom_point() +
  geom_smooth(method = lm, se = FALSE) +
  labs(y = "Total number of students eligible
    for free and reduced lunch", x = "Income to poverty ratio, where lower IPR means higher poverty
    reduced/free lunch at schools in the US")

## `geom_smooth()` using formula = 'y ~ x'
## Warning: Removed 3394 rows containing non-finite outside the scale range
## (`stat_smooth()`).
## Warning: Removed 3394 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

Income to poverty ratio (IPR) and total students eligible for reduced/free lunch at schools in the US

