

# NYC Schools Perceptions

Eda AYDIN

30 04 2021

## Contents

New York City Schools Survey Data	1
Simplifying the Dataframes	3
Creating a Single Dataframe for Analysis	3
Look for Interesting Correlations and Examine Relationships Using Scatter Plots	4
Differences in Student, Parent, and Teacher Perceptions: Reshape the Data	8

## New York City Schools Survey Data

```
# Load the packages
```

```
library(readr)  
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(stringr)  
library(purrr)  
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.0.4
```

```
library(ggplot2)
```

```
# Import the data
```

```
combined <- read_csv("combined.csv")
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   DBN = col_character(),
##   school_name = col_character(),
##   boro = col_character()
## )
## i Use 'spec()' for the full column specifications.
```

```
survey <- read_tsv("masterfile11_gened_final.txt")
```

```
##
## -- Column specification -----
## cols(
##   .default = col_double(),
##   dbn = col_character(),
##   bn = col_character(),
##   schoolname = col_character(),
##   studentssurveyed = col_character(),
##   schooltype = col_character(),
##   p_q1 = col_logical(),
##   p_q3d = col_logical(),
##   p_q9 = col_logical(),
##   p_q10 = col_logical(),
##   p_q12aa = col_logical(),
##   p_q12ab = col_logical(),
##   p_q12ac = col_logical(),
##   p_q12ad = col_logical(),
##   p_q12ba = col_logical(),
##   p_q12bb = col_logical(),
##   p_q12bc = col_logical(),
##   p_q12bd = col_logical(),
##   t_q6m = col_logical(),
##   t_q9 = col_logical(),
##   t_q10a = col_logical()
##   # ... with 18 more columns
## )
## i Use 'spec()' for the full column specifications.
```

```
survey_d75 <- read_tsv("masterfile11_d75_final.txt")
```

```
##
```

```
## -- Column specification -----
## cols(
##   .default = col_double(),
##   dbn = col_character(),
##   bn = col_character(),
##   schoolname = col_character(),
##   studentssurveyed = col_character(),
##   schooltype = col_character(),
##   p_q5 = col_logical(),
##   p_q9 = col_logical(),
##   p_q13a = col_logical(),
##   p_q13b = col_logical(),
##   p_q13c = col_logical(),
##   p_q13d = col_logical(),
##   p_q14a = col_logical(),
##   p_q14b = col_logical(),
##   p_q14c = col_logical(),
##   p_q14d = col_logical(),
##   t_q11a = col_logical(),
##   t_q11b = col_logical(),
##   t_q14 = col_logical(),
##   t_q15a = col_logical(),
##   t_q15b = col_logical()
##   # ... with 14 more columns
## )
## i Use 'spec()' for the full column specifications.
```

## Simplifying the Dataframes

```
# Select columns needed for analysis
survey_select <- survey %>%
  filter(schooltype == "High School") %>%
  select(dbn:saf_p_11)

survey_d75_select <- survey_d75 %>%
  select(dbn:com_s_11)
```

## Creating a Single Dataframe for Analysis

```
# Combine 'survey' and 'survey_d75' data frames
survey_total <- survey_select %>%
  bind_rows(survey_d75_select)

# Rename 'survey_total' variable 'dbn' to 'DBN' so can use as key
# to join with 'combined' data frame

survey_total <- survey_total %>%
  rename (DBN = dbn)
```

```
# Join the 'combined' and 'survey_total' data frames.
combined_survey <- combined %>%
  left_join(survey_total, by="DBN")
```

## Look for Interesting Correlations and Examine Relationships Using Scatter Plots

```
cor_mat <- combined_survey %>%
  select(avg_sat_score, saf_p_11:com_s_11) %>%
  cor(use = "pairwise.complete.obs")

cor_tib <- cor_mat %>%
  as_tibble(rownames = "variable")
```

```
strong_cors <- cor_tib %>%
  select(variable, avg_sat_score) %>%
  filter(avg_sat_score > 0.25 | avg_sat_score < -0.25)
```

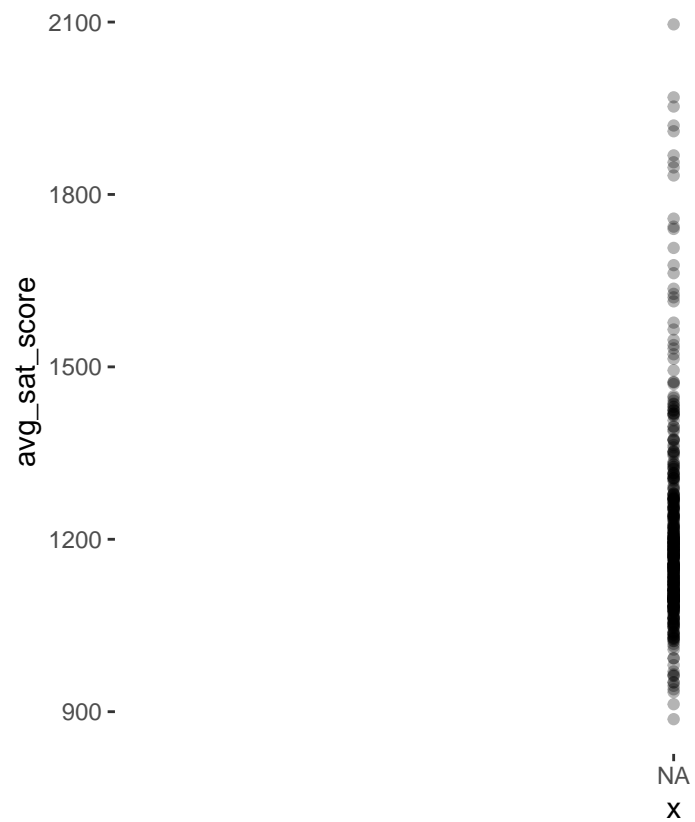
```
create_scatter <- function(x, y) {
  ggplot(data = combined_survey) +
    aes_string(x = x, y = y) +
    geom_point(alpha = 0.3) +
    theme(panel.background = element_rect(fill = "white"))
}
```

```
x_var <- strong_cors$variable[2:5]
y_var <- "avg_sat_score"

map2(x_var, y_var, create_scatter)
```

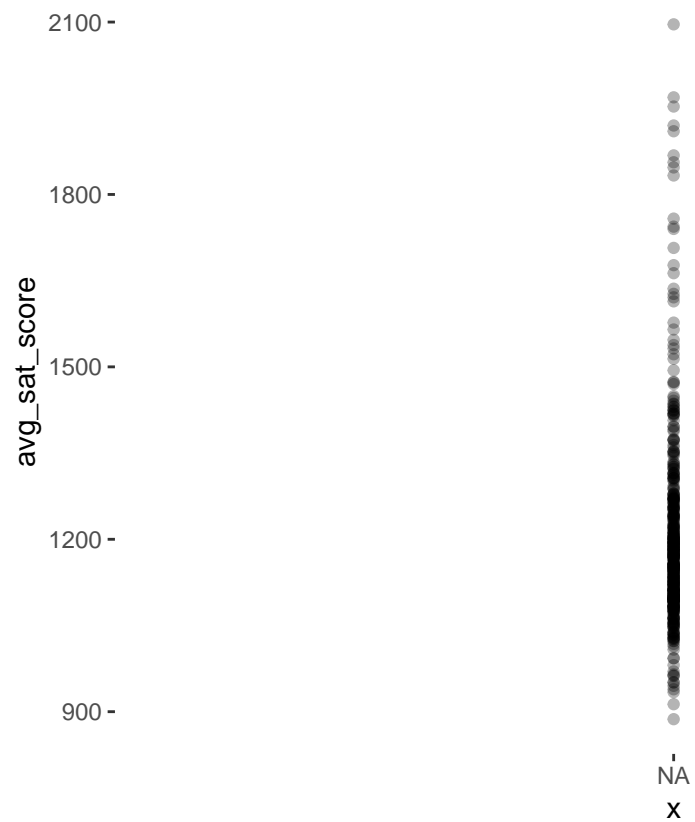
```
## [[1]]
```

```
## Warning: Removed 57 rows containing missing values (geom_point).
```



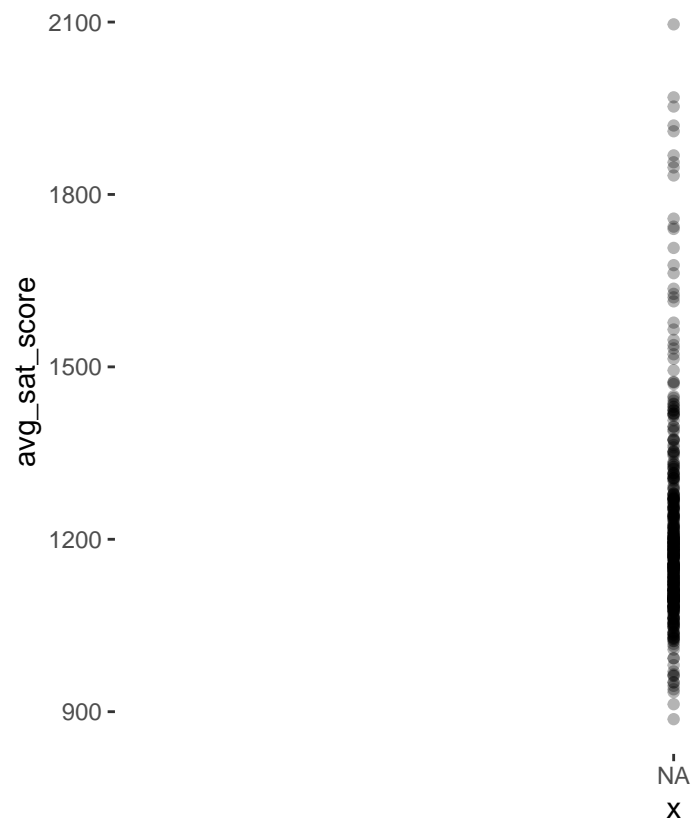
```
##
## [[2]]
```

```
## Warning: Removed 57 rows containing missing values (geom_point).
```



```
##
## [[3]]
```

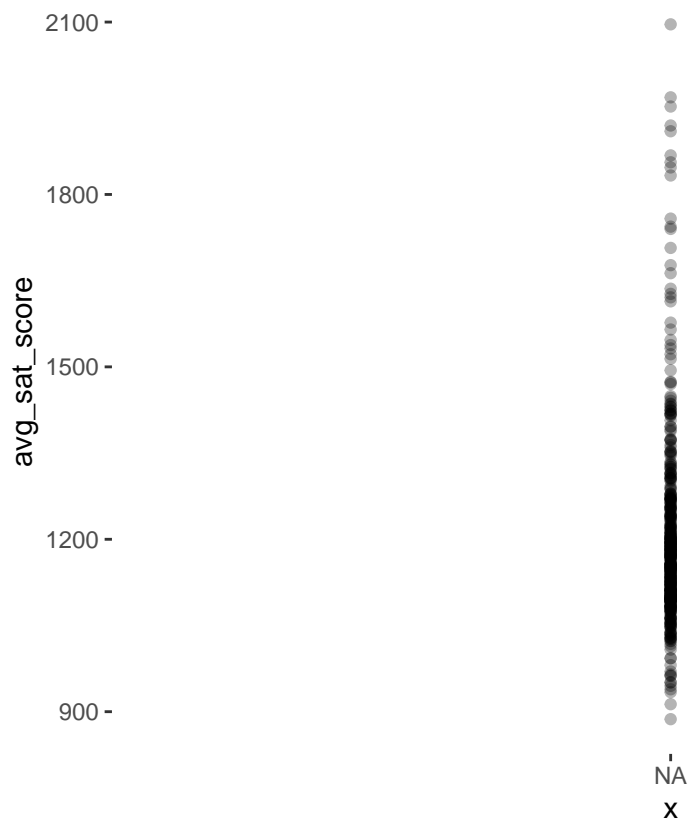
```
## Warning: Removed 57 rows containing missing values (geom_point).
```



```
##
```

```
## [[4]]
```

```
## Warning: Removed 57 rows containing missing values (geom_point).
```



```
combined_survey_gather <- combined_survey %>%
  pivot_longer(cols = saf_p_11:com_s_11,
               names_to = "survey_question",
               values_to = "score")
```

```
combined_survey_gather <- combined_survey_gather %>%
  mutate(response_type = str_sub(survey_question, 4, 6)) %>%
  mutate(question = str_sub(survey_question, 1, 3))
```

## Differences in Student, Parent, and Teacher Perceptions: Reshape the Data

```
combined_survey_gather <- combined_survey_gather %>%
  mutate(response_type = ifelse(response_type == "_p_", "parent",
                                ifelse(response_type == "_t_", "teacher",
                                          ifelse(response_type == "_s_", "student",
                                                  ifelse(response_type == "_to", "total", "NA")))))
```

```
combined_survey_gather %>%
  filter(response_type != "total") %>%
  ggplot(aes(x = question, y = score, fill = response_type)) +
  geom_boxplot()
```



## Warning: Removed 4353 rows containing non-finite values (stat\_boxplot).

