Lab 7 Activity

We will continue using the same dataset from Lab 7. Run the following code to load the data into your `R` environment.

```
dat <- rio::import("https://fabio-setti.netlify.app/data/NYC_SAT_data.csv")
```

1. We think that the race distribution across different borough may be different. In particular, we want to test whether the percentage of black students in each school is different across boroughs. The variable indicating the percentage of black students in each school is `Percent_Black`. Attempt to run a one-way ANOVA testing whether there are significant differences in mean `Percent_Black` across `Borough`.

- Do you encounter any issues? what does the `Percent_Black` look like? Why do you think the analysis does not run?

2. Whoever made this dataset, entered the variables in a very suboptimal way. Appropriately turn the `Percent_Black` variable into a `numeric` variable such that we can run the one-way ANOVA.

- Transform the `Percent_Black` variable into a `numeric` variable. There are many ways of doing it, but try looking into the `sub()` function. here is an example of how it works:

```
sub(pattern = "hi",
    x = c("hi there", "you can change the pattern to hi"),
    replacement = "anything you want")
```

```
[1] "anything you want there"
[2] "you can change the pattern to anything you want"
```

- Run a one-way ANOVA testing whether there are significant differences in mean `Percent_Black` across `Borough`. Remember to first turn `Borough` into a `factor` variable. Interpret the ANOVA result.

3. Calculate the effect size for the effect of `Borough` on `Percent_Black`. How would you interpret it?

4. Evaluate the ANOVA assumptions by creating violin plots of `Percent_Black` by `Borough`. What do you think about the equal variances and the normality assumption?

5. conduct some contrasts. Test all the following contrasts at once:

- Test the mean difference in `Percent_Black` between the Bronx and Manhattan.
- Test the mean difference in `Percent_Black` between the Bronx and Manhattan plus Brooklyn.
- Test the mean difference in `Percent_Black` between Queens plus Staten Island and Manhattan plus the Bronx.

Run all contrasts together twice; the first time apply a Bonferroni corerction to the p-values. The second time apply Sidak adjustment to the p-values. Would you conclusions change depending on the adjustment?