**AN ANALYSIS REPORT ON SCHOOL SURVEYS IN KENYA**

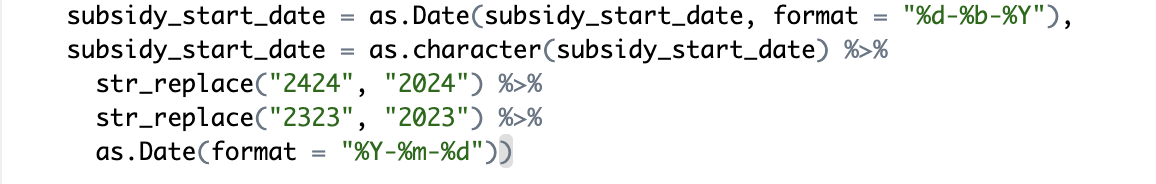
**TITUS OLANG’**

Data Quality Issues:

1. **Data format**

By observing the data at a glance, I could tell the ‘subsidy\_start\_date’ had erroneous dates.

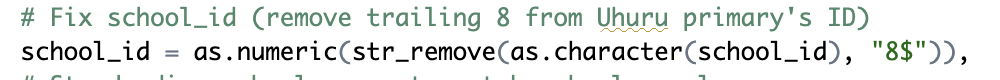
I formatted the ‘subsidy\_start\_date’ to a more likely date.



1. **Extra characters**

The ‘school\_id’ in one of the student scores tables had extra characters.

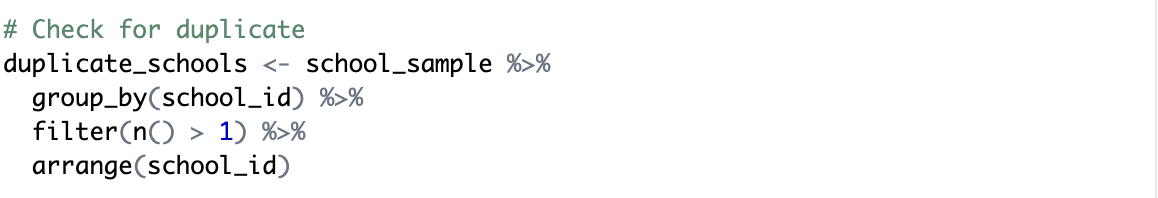
I trimmed the extra to match the school sample id.



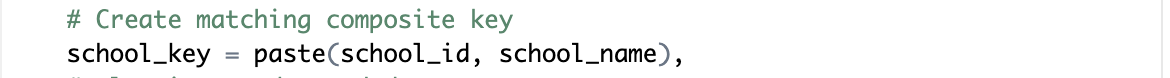
* This ensured that the id’s matched in both tables.

1. **ID Duplicates**

There were duplicates in ‘school\_ids’, checked for duplicates by running:

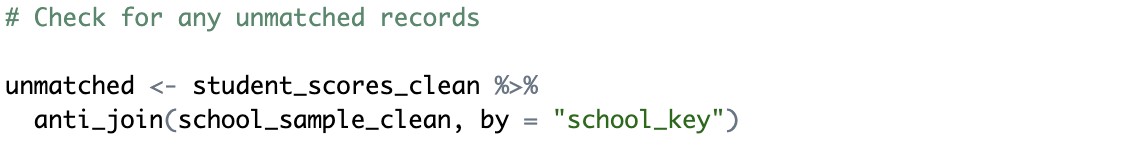


I combined it with the ‘school name’ to make a composite key for the join.



This is to ensure accurate mapping and that no data is lost as a result of duplication.

1. **After joining, I checked if there were any unmatched records.**

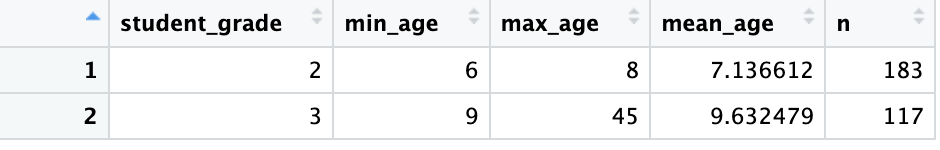
****

* There are other data validations, I just pinpointed some major ones.

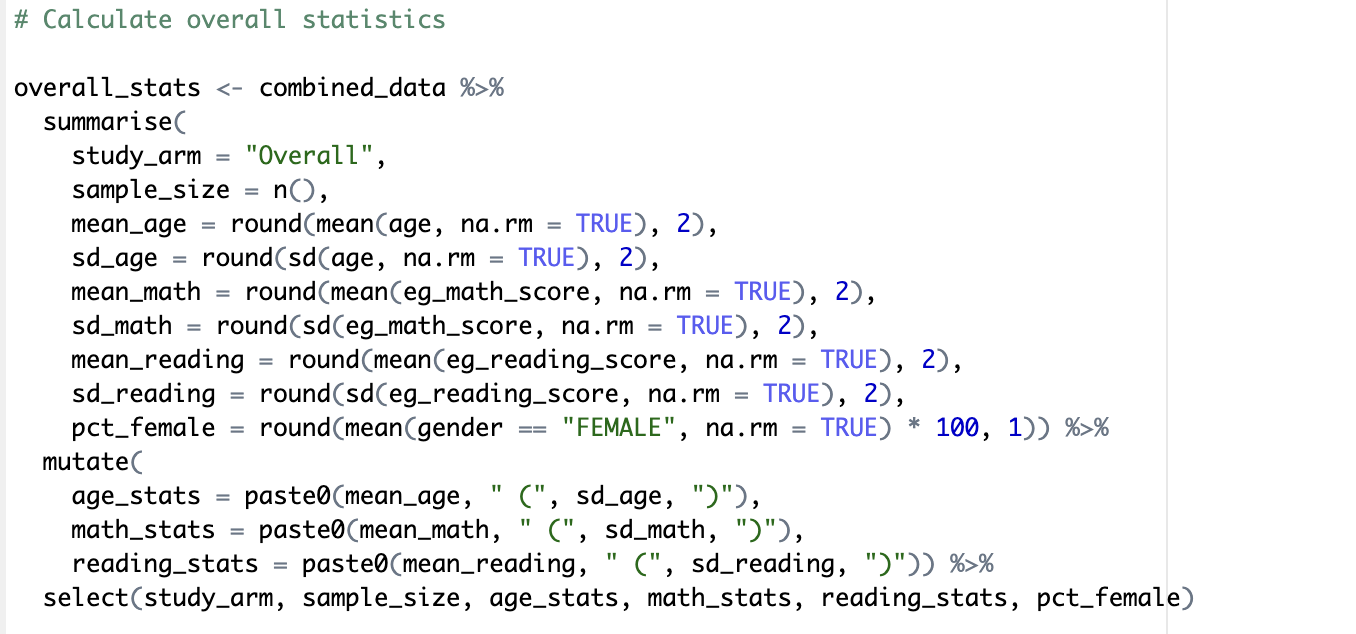
Attraction Analysis and Sample Balance

1. **Age distribution**

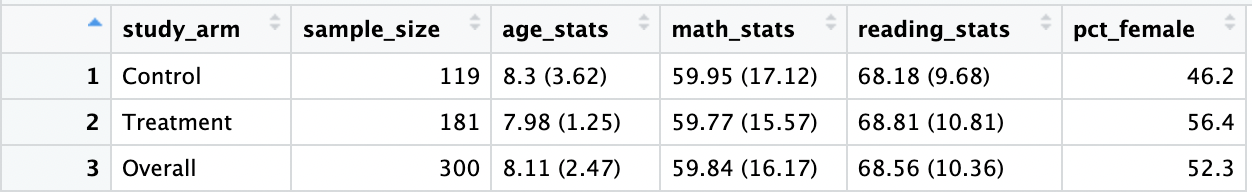




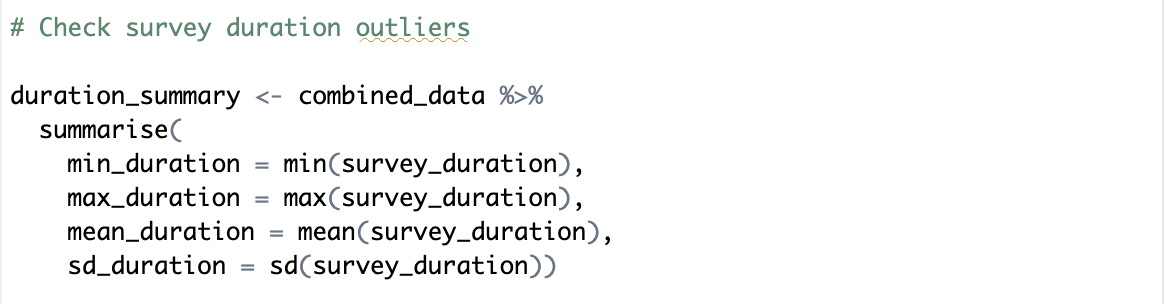
1. **Summary statistics for sample and control groups**



Results:



1. **Survey Duration and Outliers**

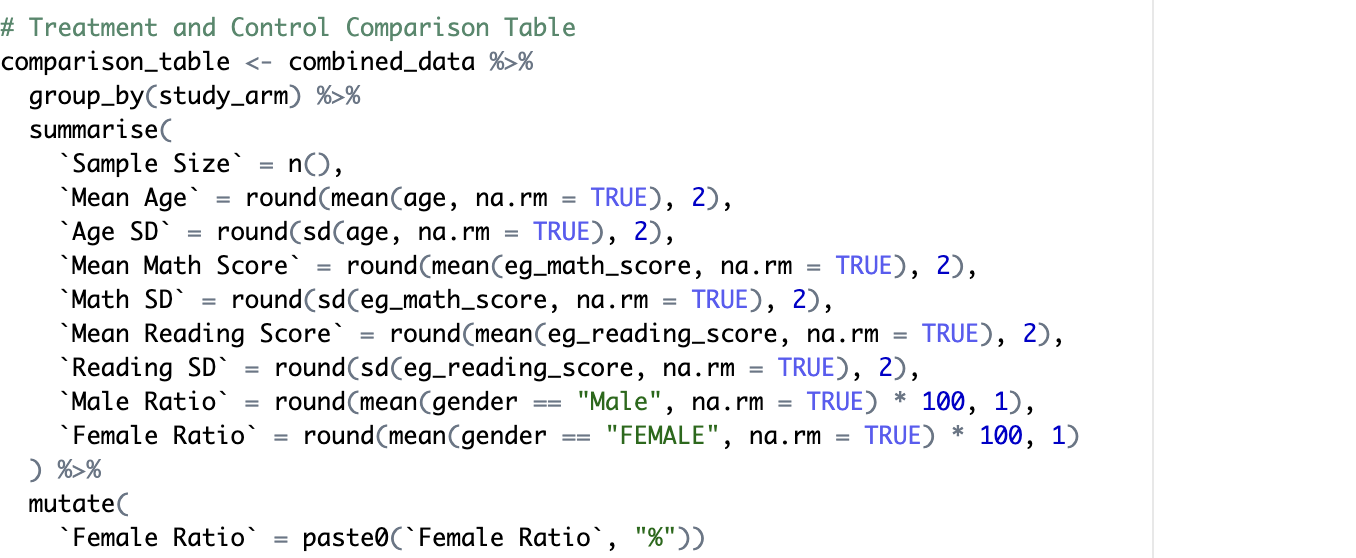


Results



1. **Comparison Table**

Includes means of age, sample size, test scores, and gender ratios



Results

