Max Pursche
CS124 Project 5

Data

The dataset used was a rated list of different quick-cook ramen brands from around the world. It was updated last as of 2017, it has 2,580 different types of ramen. I removed a field that was unnecessary which identified different top ten year placements however the top tens weren't complete sets of year and date.

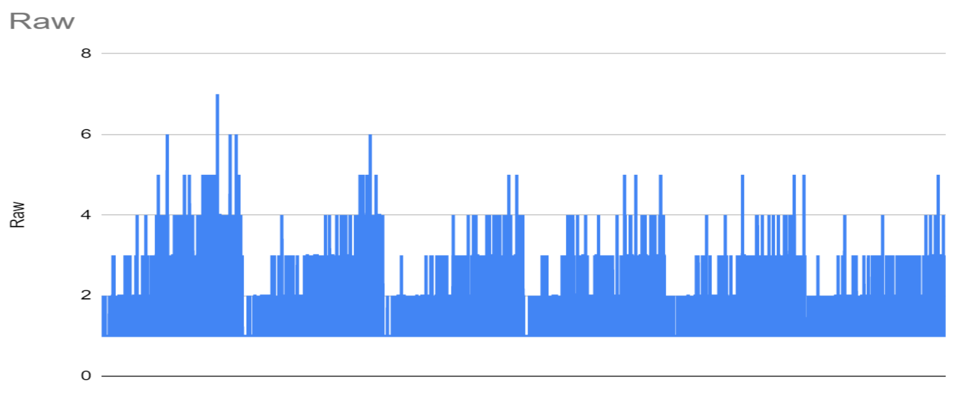
The following 6 fields are in the order of the CSV:

- Number - Integer, tracks the row number of a ramen
- Brand - String, The brand of a ramen
- Variety - String, The flavor of a ramen
- Style - String, The type of instant ramen that it is; a cup, a pack, a tray, or a bowl.
- Country - String, holds the country that a ramen is from
- Rating - Float, holds the rating of a ramen

Entries are ordered from 2580 to 1.

Analysis

The reads are much larger the smaller the dataset is, they are fairly similar between separate chaining and Linear Probing. The separate chaining graph is below, the reads became more efficient with more space available for the hash table. The best hash table for my set was definitely the 5000+ size. The horner hash with a country and number header concatenated together was the most efficient key.



.

Questions

The Insert and find operations are easily going to be the most used, each piece of data would only need to be added to the hashtable. Remove would only need to be used if a review was updated for an existing ramen.

Separate chaining is easiest because the dataset isnt likely to need to chain very many nodes together.

Other

I couldnt get the reads for Linear probing, the rehash function was causing a sigsev when reinserting into the table so it wouldnt record the data.