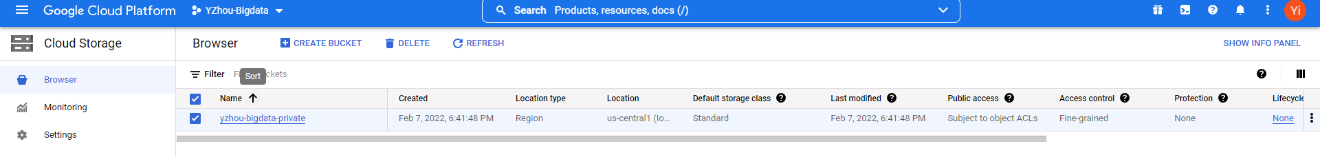
HW2

Yi Zhou

Feb. 9th. 2022

   1.



2.

First I randomly took 50000 line of data from the original file so that it at least contains 5000 lines of data from 2006:

$zcat arcos\_all\_washpost.tsv.gz | shuf -n50000 > hw2.txt

And I find all data from 2006

$ awk -F '\t' '$31 ~ /2006$/ {print $0}' hw2.txt > temp.txt

Finally I random took 5000 line of data:

$ shuf -n5000 temp.txt>q2.txt

3. 

4.

A screenshot of a computer

Description automatically generated with medium confidence

5.

yzhou20@cloudshell:~ (nimble-factor-340723)$ gcloud config set project nimble-factor-340723

Updated property [core/project].

yzhou20@cloudshell:~ (nimble-factor-340723)$ bq load --field\_delimiter=tab --source\_format=CSV nimble-factor-340723:alice\_2022.arcos\_2006 gs://yzhou-bigdata-private/arcos\_2006.tsv

Waiting on bqjob\_r56cfa81afae2fb9e\_0000017ee55bc43e\_1 ... (1s) Current status: DONE

6.

[

{

"DRUG\_NAME": "HYDROCODONE"

},

{

"DRUG\_NAME": "OXYCODONE"

}

]