

Assignment 1 – twitter

CS 751 Introduction to Digital Libraries

Rashmin Patel

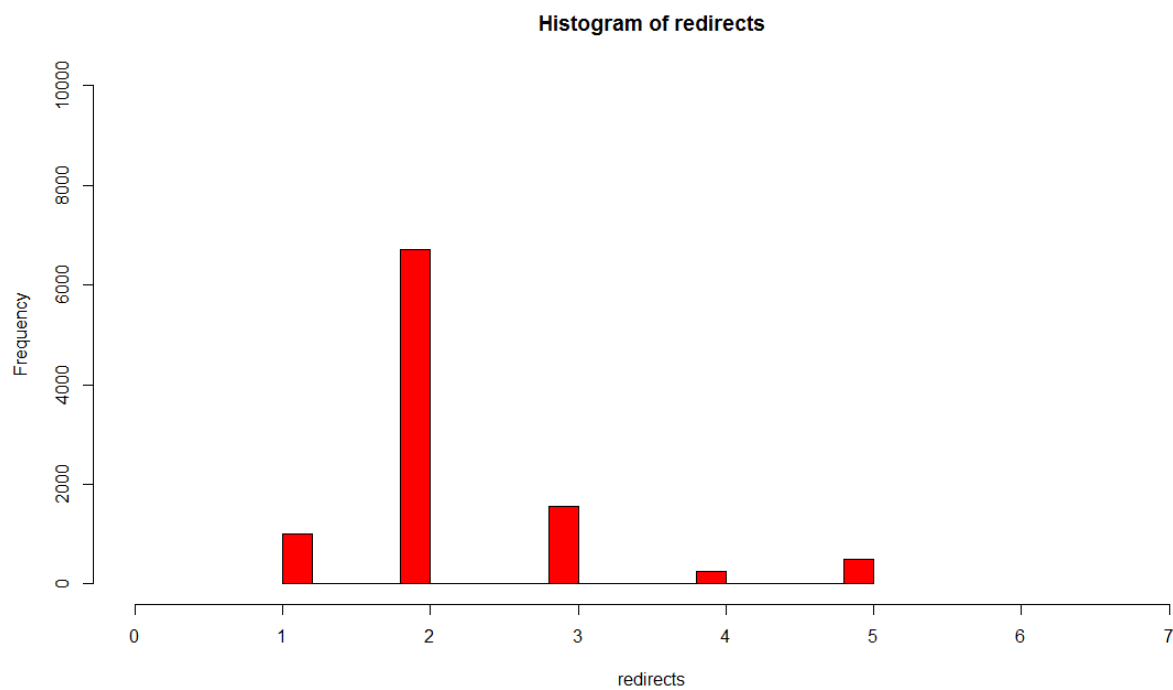
Question 1: Build a histogram of how many redirects and HTTP response codes?

Answer:

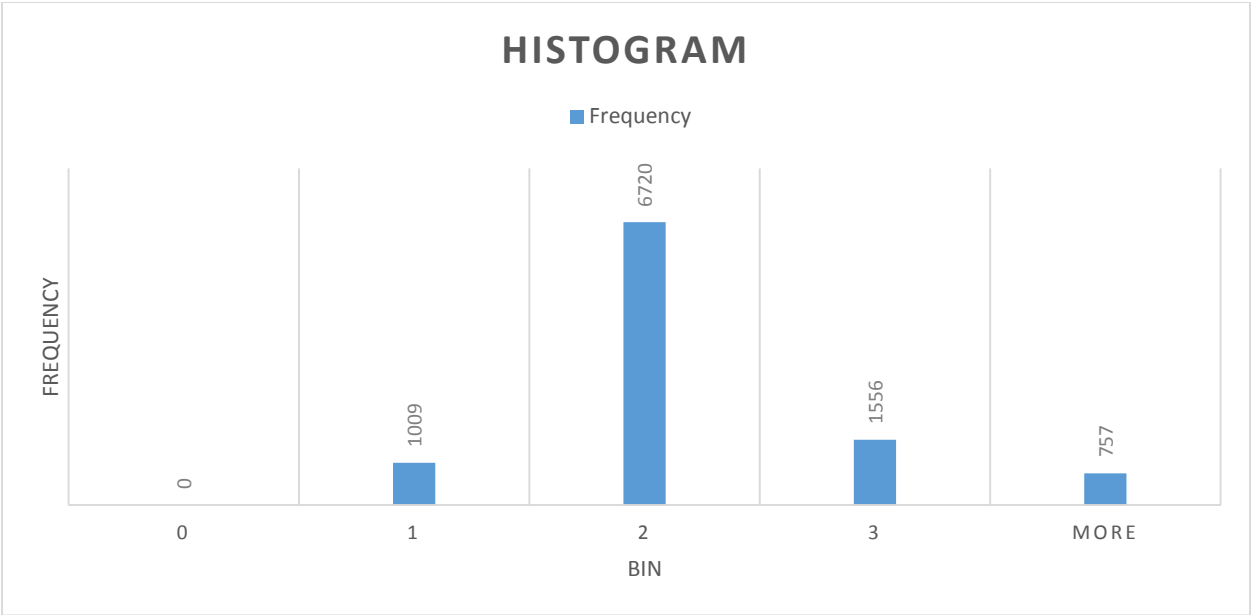
For each twitter link, curl operation has been performed and analyzed the number of redirects performed for each link operation.

Below histogram represents the average number of redirects.

<i>No. Of Redirects</i>	<i>Frequency</i>
0	0
1	1009
2	6720
3	1556
More	757



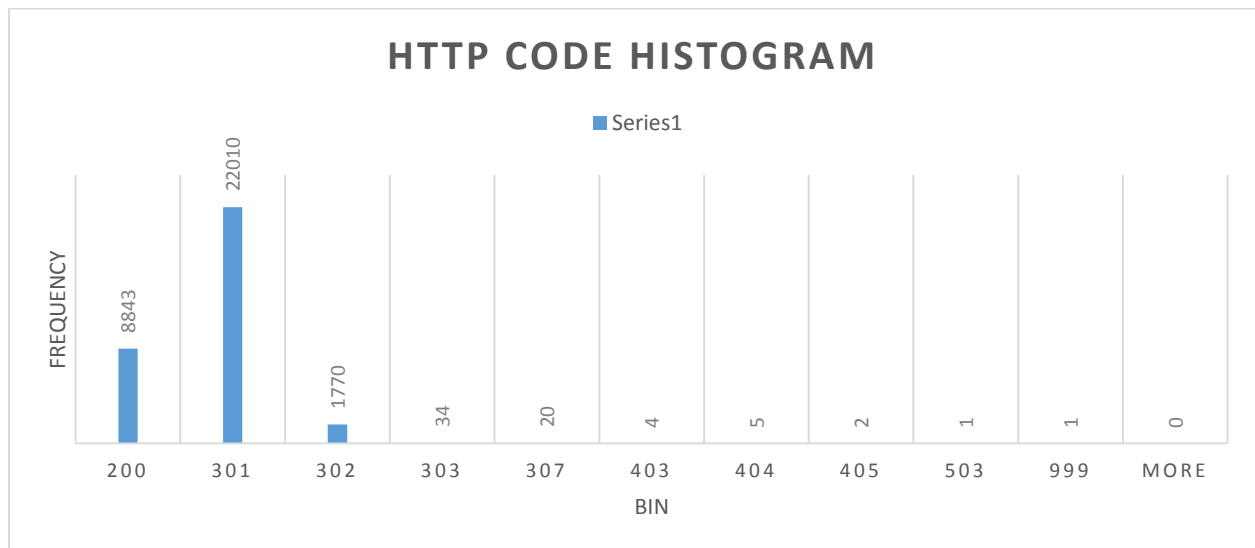
(Histogram output from R)



(Histogram output from Excel)

HTTP Response Code:

<i>HTTP Code</i>	<i>Frequency</i>
200	8843
301	22010
302	1770
303	34
307	20
403	4
404	5
405	2
503	1
999	1
More	0

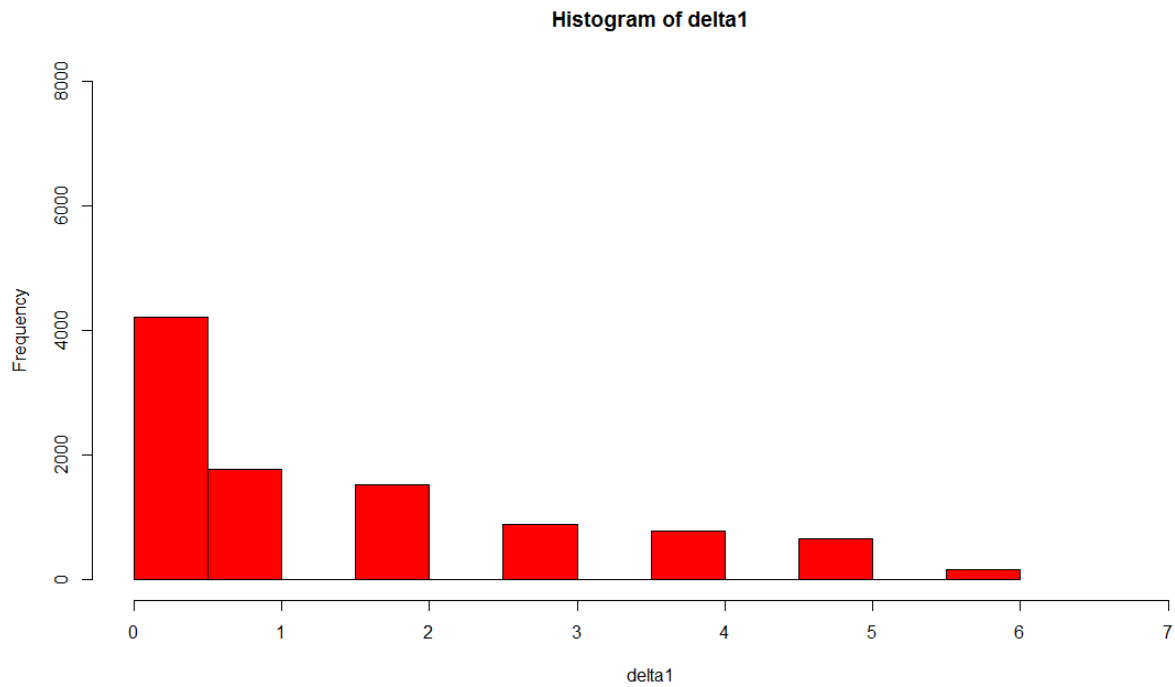


How many duplicate URIs found?

Among all 10043 URIs, there are total 1464 duplicate URIs.

Question 2: Build a histogram of delta between tweet age and link age?

<i>Bin</i>	<i>Frequency</i>
0	4217
1	1785
2	1530
3	897
4	784
5	658
6	167
7	0
More	0



(Histogram Output from R)

Mean	Median	Standard Deviation	Standard Error
1.490735	1	1.690294	0.016871