Anaylsis of Shanghai Data from NCBI SARS-CoV-2 SRA Bioproject ID 627662

Mikaela Kuan

10 May, 2021

Background and Overview

This is a report on SARS-CoV-2, including some variant analysis (Koyama *et al.*, 2020).

Methods

See the set of tutorials on the vcfR package website.

You may also want to use any of a range of different COVID data packages and data sources:

- https://kjhealy.github.io/covdata/
- https://github.com/como-ph/oxcovid19
- https://ropensci.org/blog/2020/10/20/searching-medrxivr-and-biorxiv-preprint-data/
- https://covidtracking.com/data/api

    - `readr::read_csv("https://api.covidtracking.com/v1/states/daily.csv")`

Subsections are ok too

Results and Discussion

Figures

Frequency of the Quality Scores in the SNPs from Shanghai
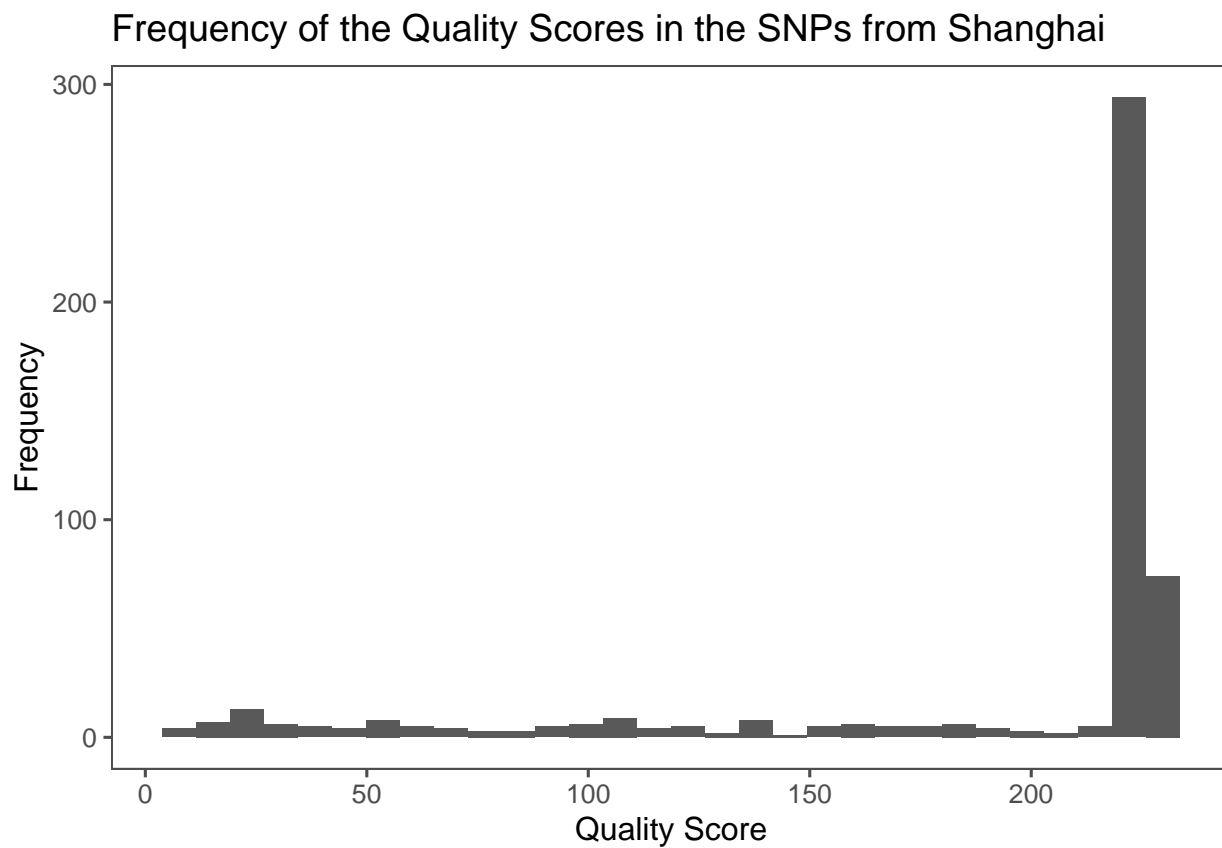
Figure 1. The frequencies of the quality scores from Shanghai data. Quality scores less than 200 are not of high quality.

COVID−19 Samples Collected in Shanghai from January to February
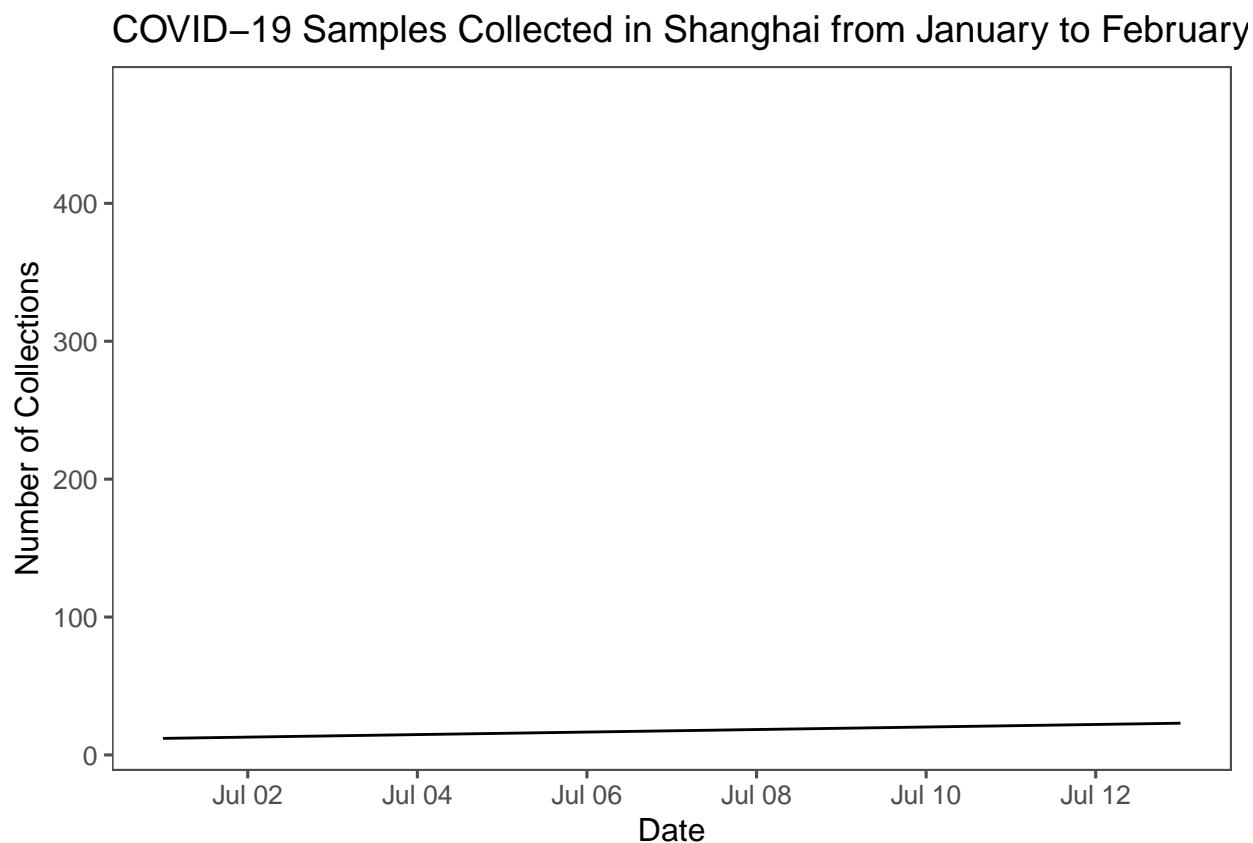
Figure 2. The frequency of samples collected in Shanghai from late January to early February.

```
## Error in (function (classes, fdef, mtable) : unable to find an inherited method for function 'dbList'
```

Figure 3. Hospitalization records in China from late January to early February.

Tables

SNP Position	Frequency
241	6
1397	7
2293	8
3037	6
8782	34
11083	11
23403	6
24034	6
26144	9
26729	6
28077	6
28144	34
28688	7
29095	6
29449	6
29742	11

Table 1. Frequency of each SNP position. The highest frequency was the SNPs in position 28144 and 8782. SNPs with frequencies less than five were removed from the table.

SNP Position	Reference Nucleotide	Altered Nucleotide	Frequency
28144	T	C	34
8782	C	T	34
29742	G	A	4
29742	G	T	7
11083	G	T	11

Table 2. Frequency of the SNPs and its different changes of the top four frequent SNPs.

SNP Position	Quality Score	Frequency
28144	225	31
28144	228	3
8782	225	32
8782	228	1
8782	65	1
29742	225	7
29742	228	4
11083	105	1
11083	107	1
11083	117	1
11083	139	1
11083	150	1
11083	169	1
11083	194	1
11083	210	1
11083	225	2
11083	65	1

Table 3. The quality score of the top 4 most frequent SNPs in the Shanghai dataset. SNPs in position 28144, 8782, and 29742 had much better quality scores than SNPs in position 11083.

Sources Cited

Koyama,T. *et al.* (2020) Variant analysis of sars-cov-2 genomes. *Bulletin of the World Health Organization*, **98**, 495.