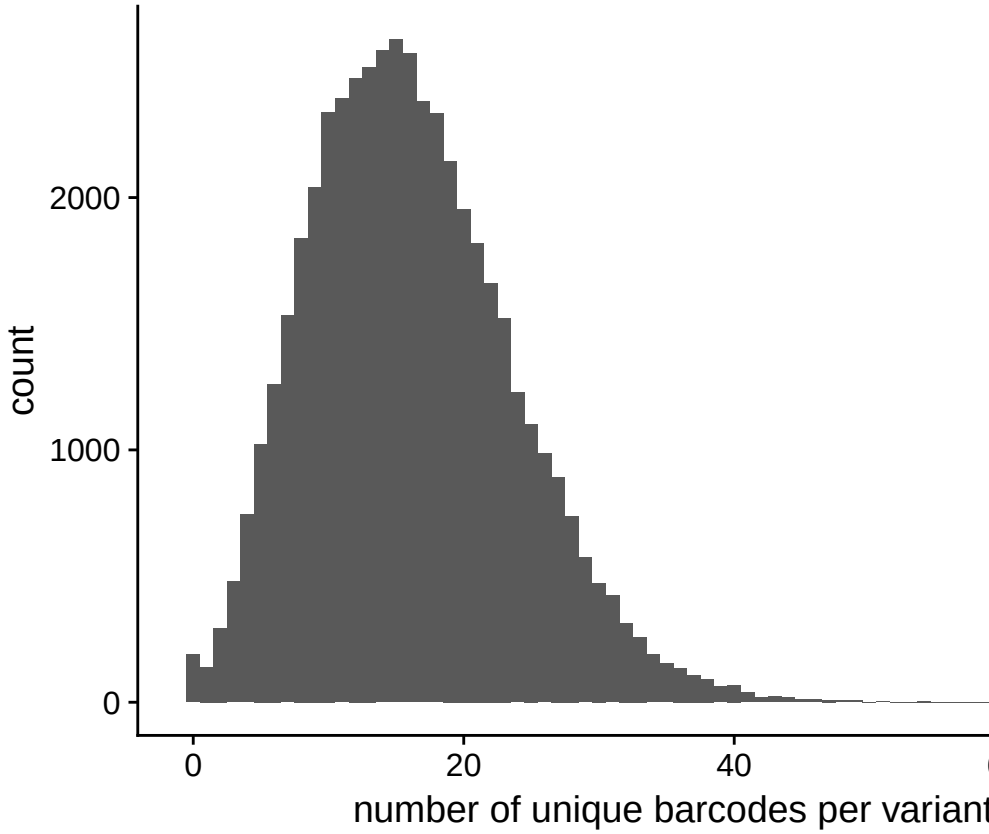
Quality Control for Peak Tiling Barcode Mapping

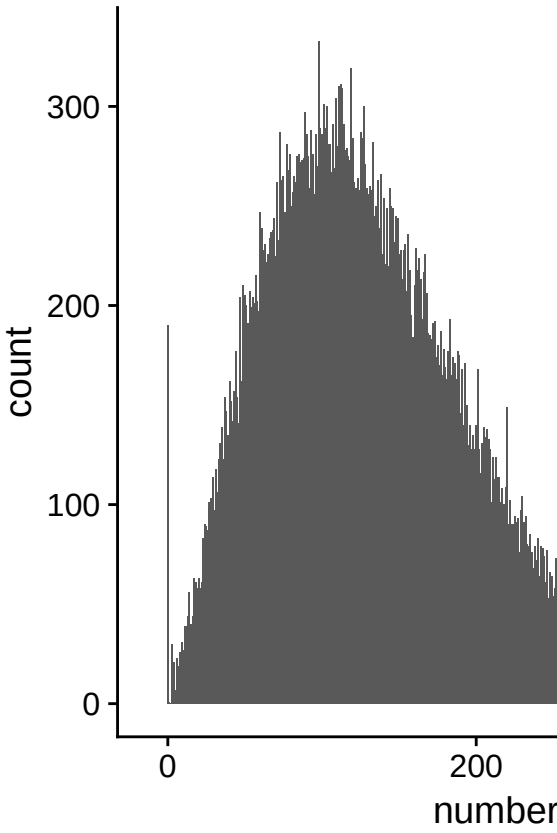
Mapping results:

```
Number of reads: 42985175
Reading in library reference...
Extracting perfect reads...
Percent perfect: 0.173709028752
Number of unique barcodes for perfect reads:  1235317
Filter by barcode frequency...
Number of barcodes > 2: 807898
Mapping...
Bootstrapping reference sequences to obtain cutoff... cutoff is Levenshtein distance  74.0
Filtering and writing results...
Percent of library represented by final barcodes: 0.98002821443
Number of final barcodes:  788822
```



How many unique barcodes per variant?

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   11.00   15.00   16.16   21.00   71.00
```

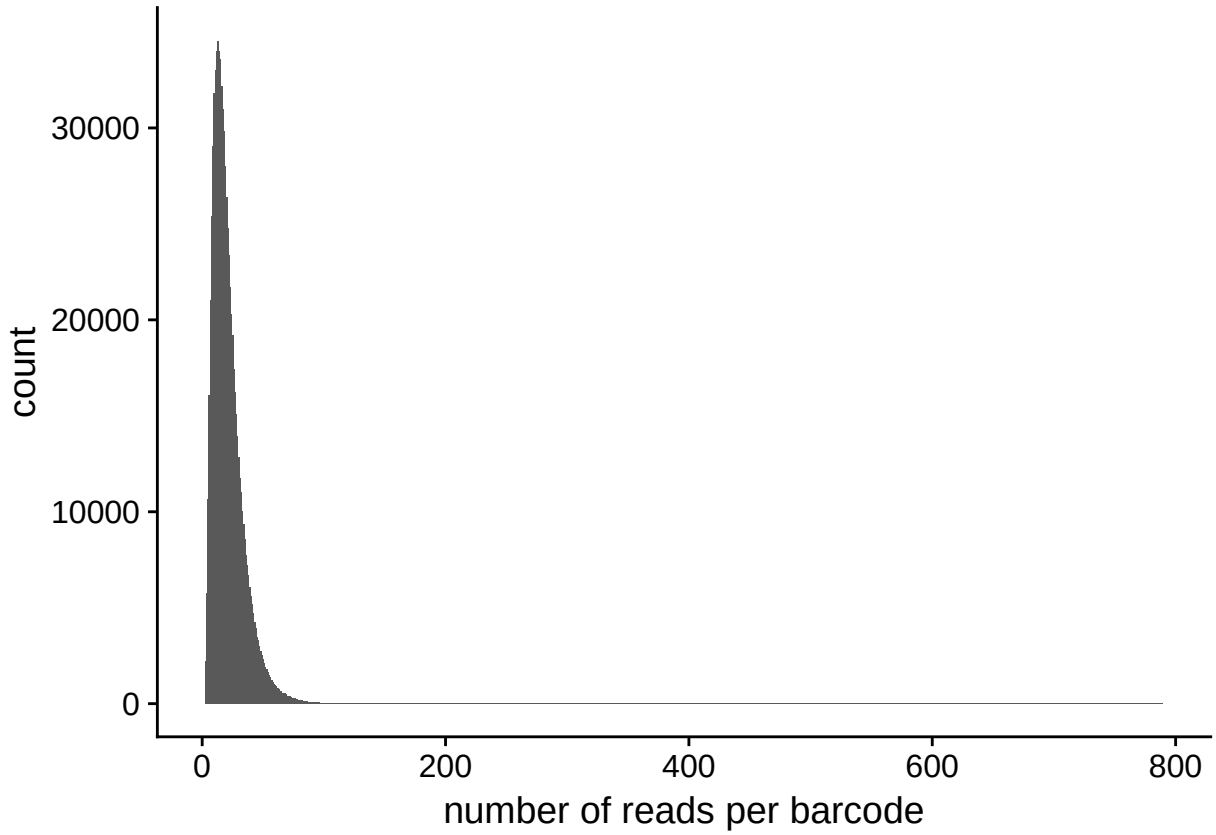


What is the number of reads per variant (aka number of non-unique barcodes)?

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0      82     125     137     179     641
```

How many reads per barcode?

```
ggplot(bc_stats, aes(num_reads)) + geom_histogram(binwidth = 1) +
    labs(x = 'number of reads per barcode')
```




```
summary(bc_stats$num_reads)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00   12.00   18.00   21.17   27.00  789.00
```