

QTLseqrRiceColdTolerance

Michael Hall

2022-06-08

```
# install devtools and all other dependent packages first
#utils::install.packages("devtools")
#utils::install.packages("tidyr")
#utils::install.packages("ggplot2")
#utils::install.packages("dplyr")
#utils::install.packages("data.table")
#utils::install.packages("vcfR")

# use devtools to install QTLseqr from my "Forked" Repository at PBGL
devtools::install_github("PBGLMichaelHall/QTLseqr")
```

Skipping install of 'QTLseqr' from a github remote, the SHA1 (9647160e) has not changed since last install.
Use `force = TRUE` to force installation

```
# Load them into your session so they are available for use
base::library("QTLseqr")
base::library("data.table")
base::library("dplyr")
```

Attaching package: 'dplyr'

The following objects are masked from 'package:data.table':

between, first, last

The following objects are masked from 'package:stats':

filter, lag

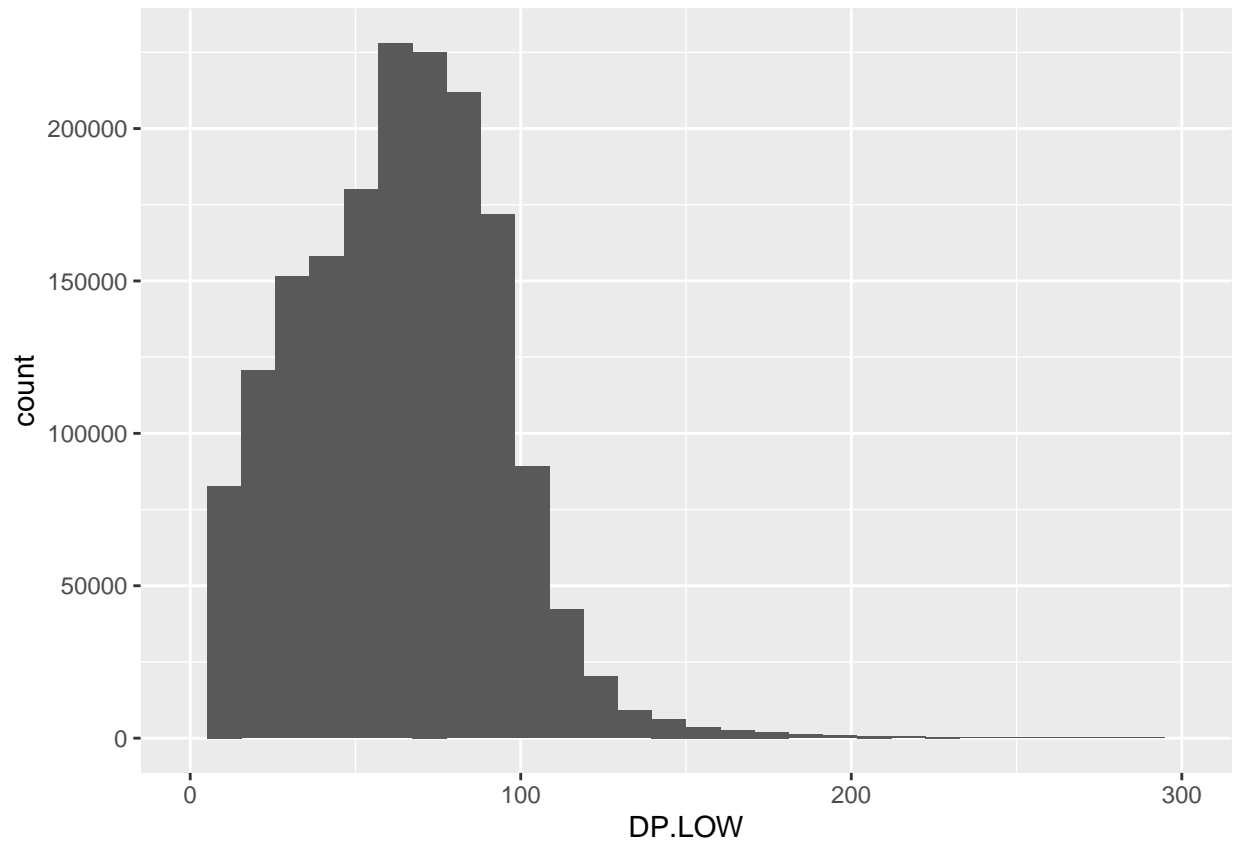
The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
base::library("tidyr")
base::library("vcfR")
```

```
*****      ***   vcfR   ***      *****
This is vcfR 1.12.0
  browseVignettes('vcfR') # Documentation
  citation('vcfR') # Citation
*****      *****      *****      *****
```

```
base::library("ggplot2")
```

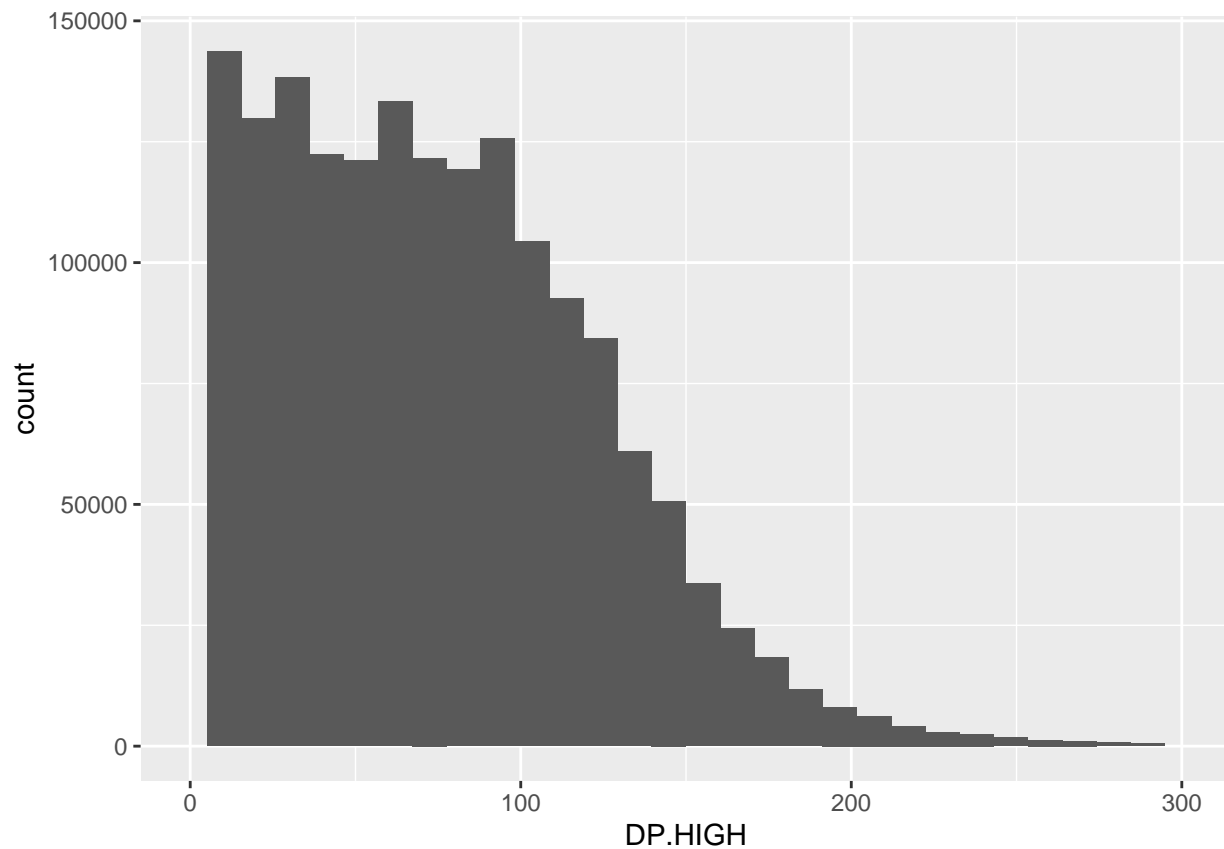



```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = DP.HIGH)) + ggplot2::xlim(0,300)
```

`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

Warning: Removed 3790 rows containing non-finite values (stat_bin).

Removed 2 rows containing missing values (geom_bar).

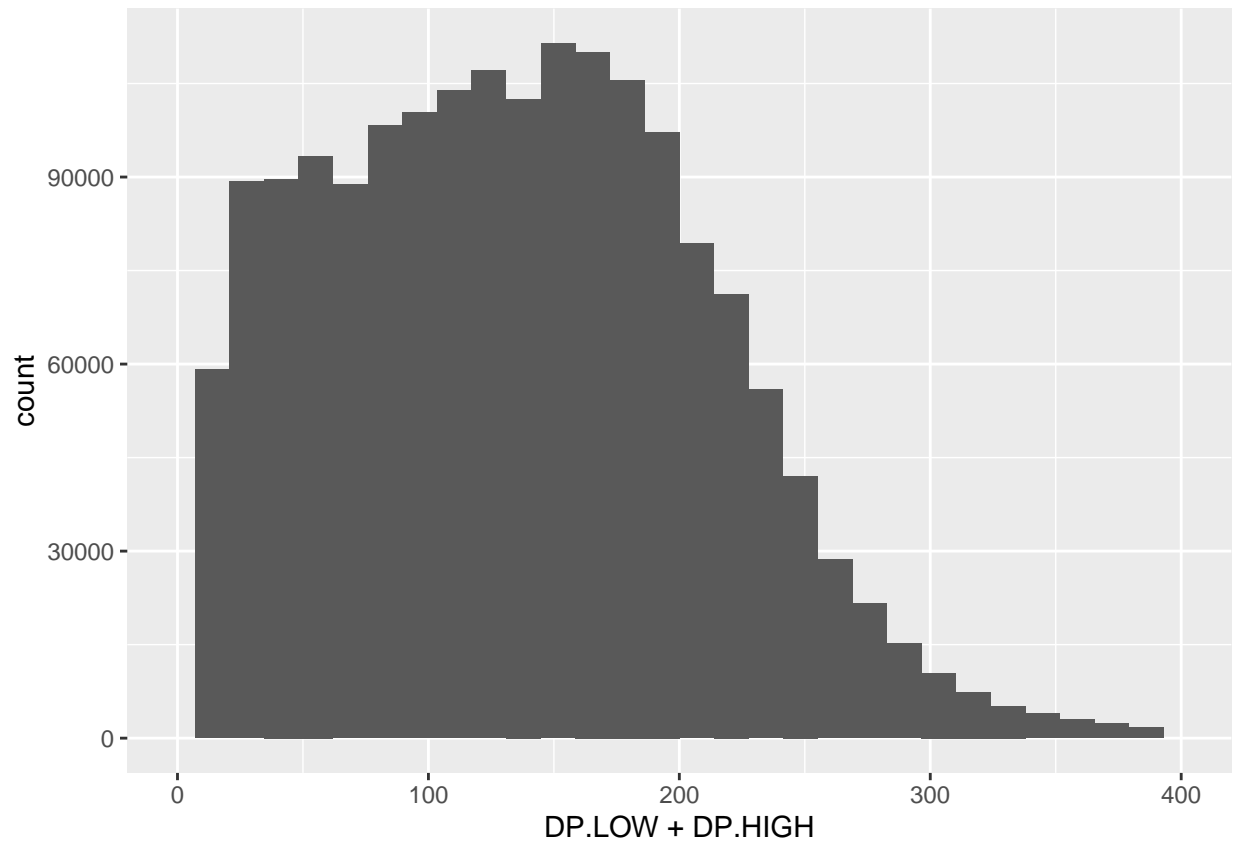


```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = DP.LOW + DP.HIGH)) + ggplot2::xlim(
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
Warning: Removed 9027 rows containing non-finite values (stat_bin).
```

```
Removed 2 rows containing missing values (geom_bar).
```

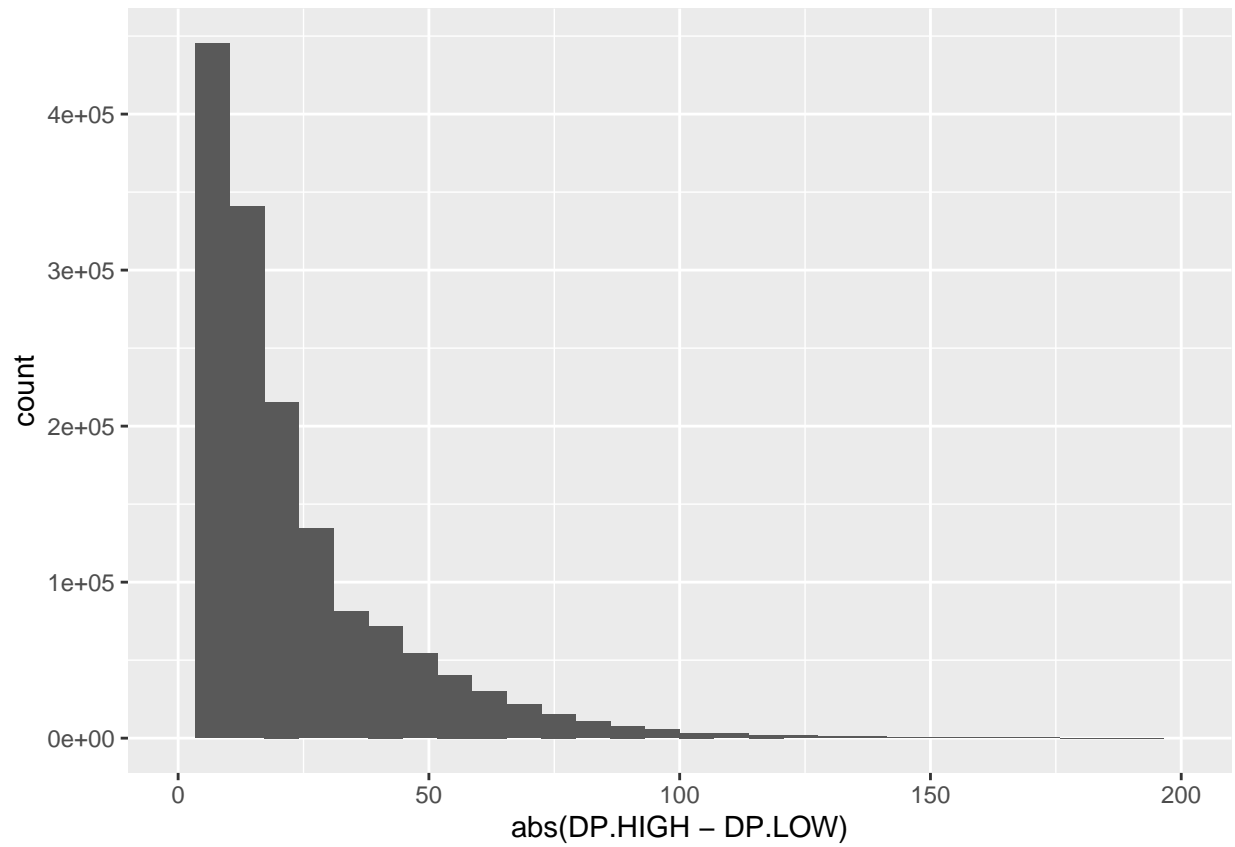


```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = abs(DP.HIGH - DP.LOW))) + ggplot2::
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

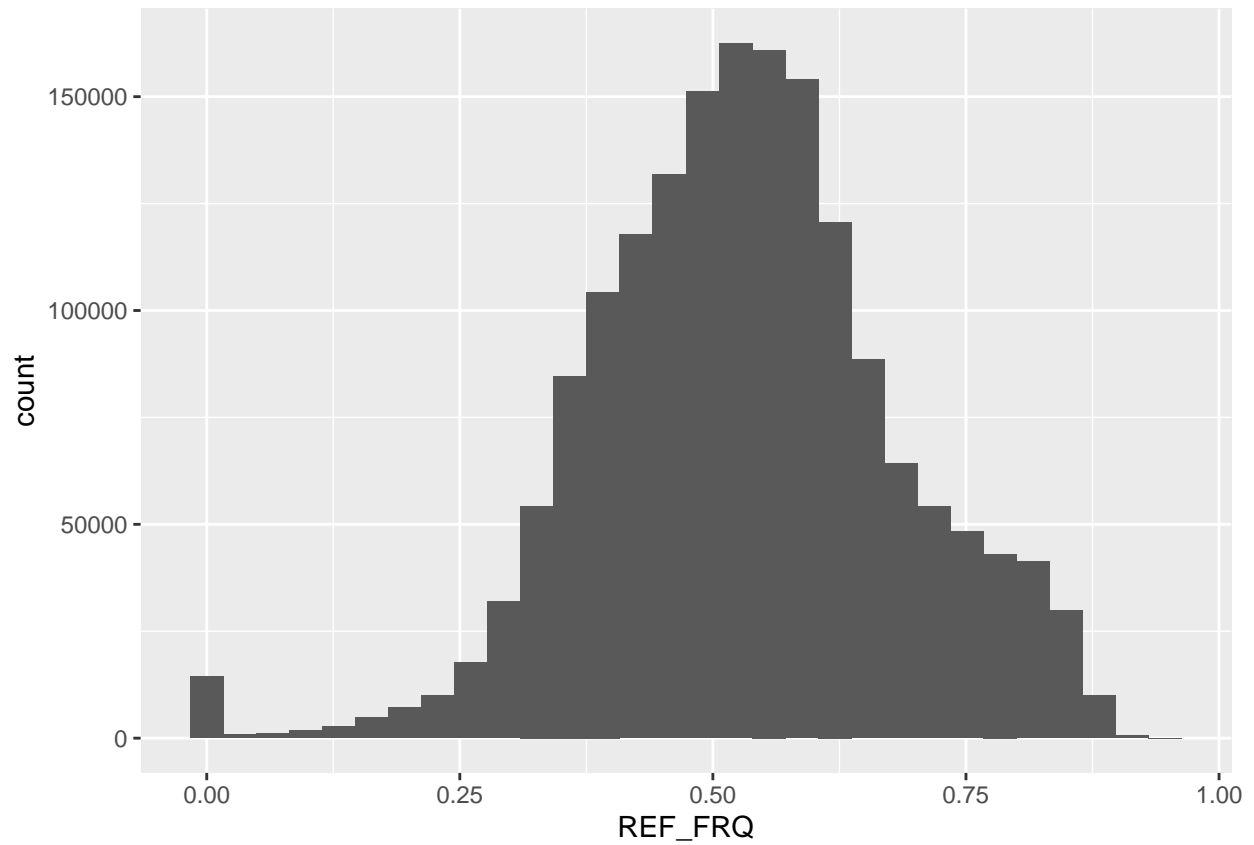
```
Warning: Removed 755 rows containing non-finite values (stat_bin).
```

```
Removed 2 rows containing missing values (geom_bar).
```



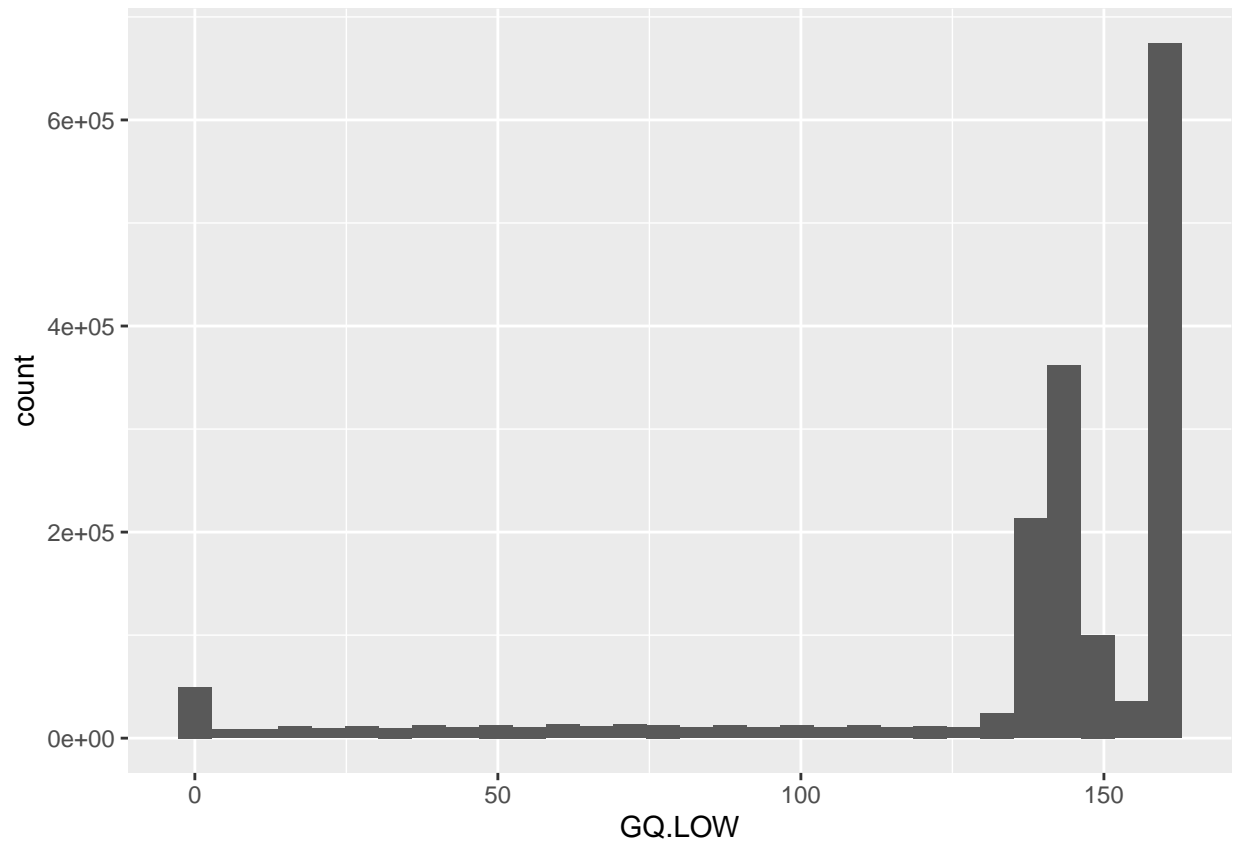
```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = REF_FRQ))
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



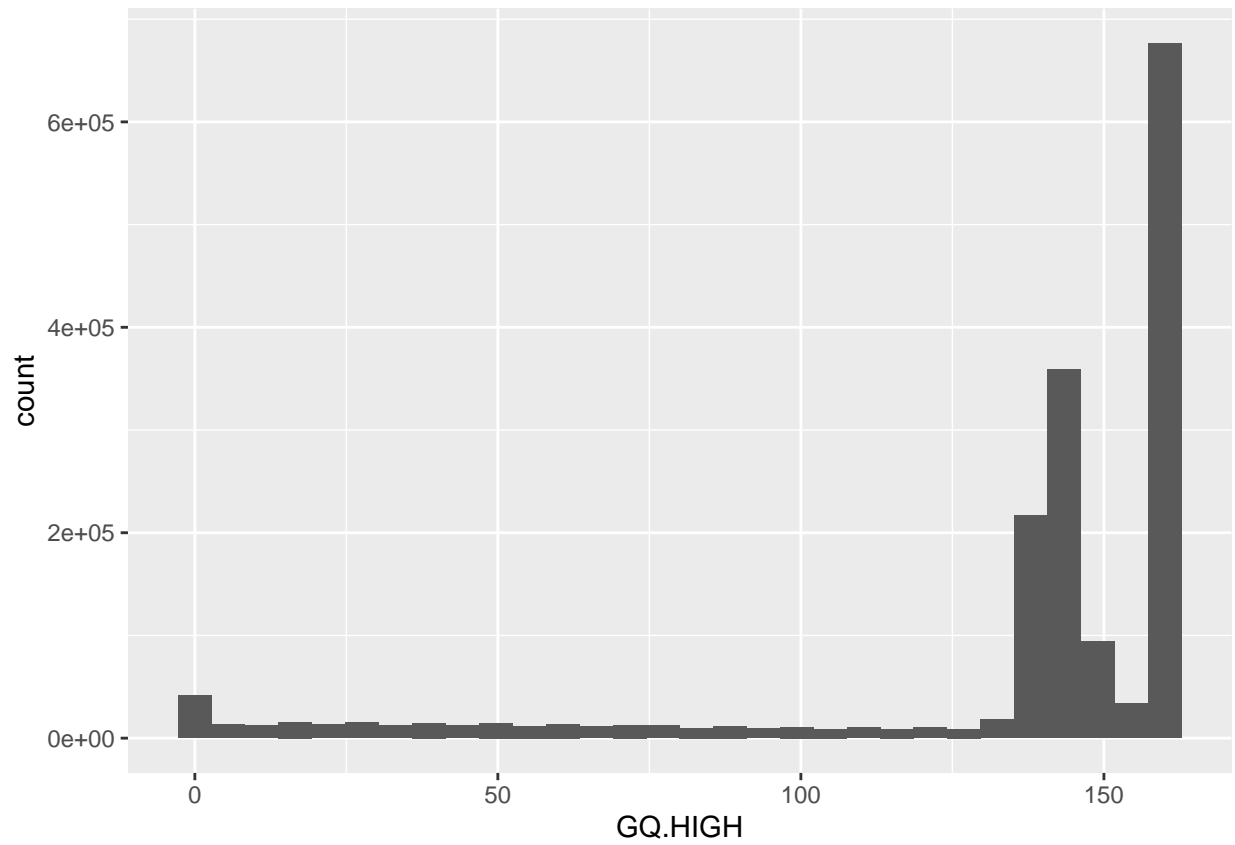
```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = GQ.LOW))
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.

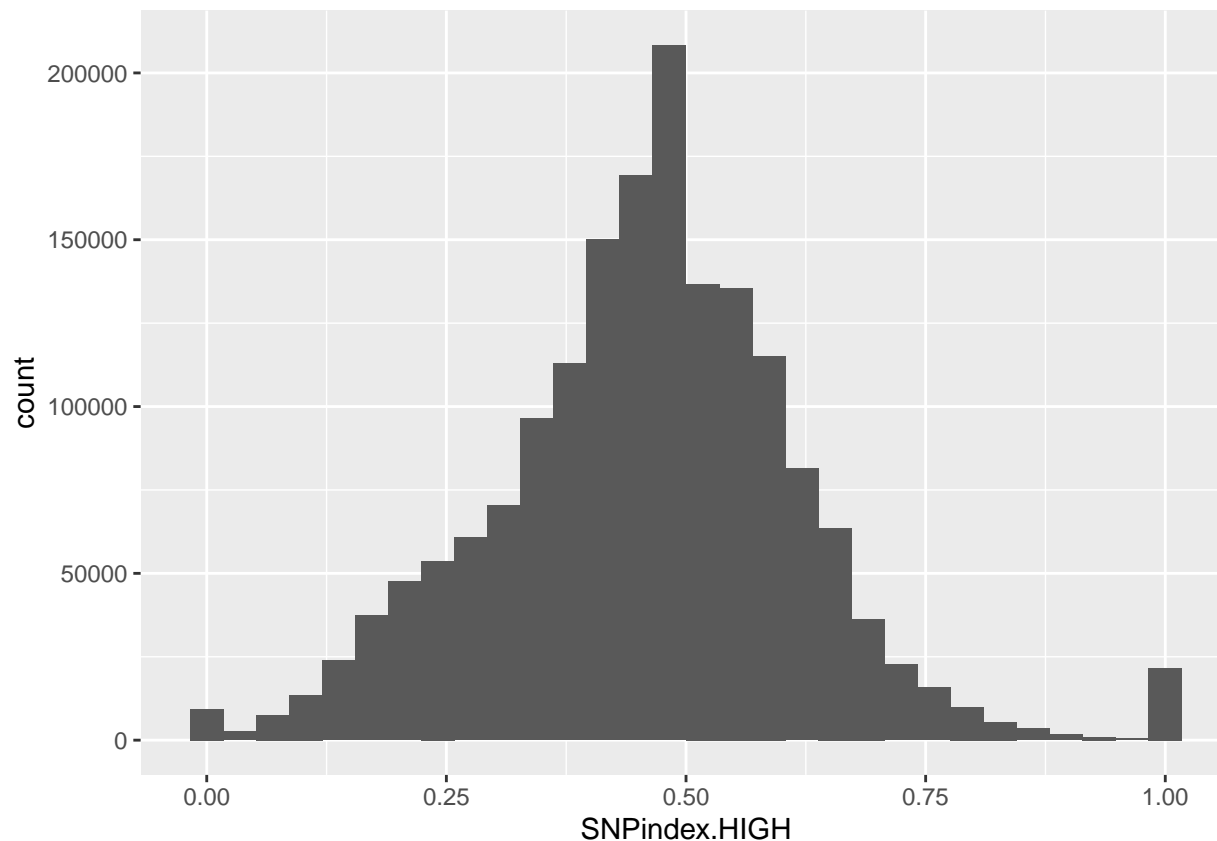


```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = GQ.HIGH))
```

``stat_bin()`` using ``bins = 30``. Pick better value with ``binwidth``.



```
ggplot2::ggplot(data = df) + ggplot2::geom_histogram(ggplot2::aes(x = SNPindex.HIGH))  
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



#Filter SNPs based on some criteria

```
df_filt <-
  QTLseqr::filterSNPs(
    SNPset = df,
    refAlleleFreq = 0.20,
    minTotalDepth = 100,
    maxTotalDepth = 400,
    depthDifference = 100,
    minSampleDepth = 40,
    minGQ = 99,
    verbose = TRUE
  )
```

Filtering by reference allele frequency: $0.2 \leq \text{REF_FRQ} \leq 0.8$

...Filtered 112443 SNPs

Filtering by total sample read depth: Total DP ≥ 100

...Filtered 537579 SNPs

Filtering by total sample read depth: Total DP ≤ 400

...Filtered 6975 SNPs

Filtering by per sample read depth: DP ≥ 40

...Filtered 8015 SNPs

Filtering by Genotype Quality: GQ ≥ 99

...Filtered 37905 SNPs

Filtering by difference between bulks <= 100

...Filtered 8746 SNPs

Original SNP number: 1714745, Filtered: 711663, Remaining: 1003082

#Run G' analysis

```
df_filt<-QTLseqr::runGprimeAnalysis(  
  SNPset = df_filt,  
  windowSize = 1e6,  
  outlierFilter = "deltaSNP",  
  filterThreshold = 0.1)
```

Counting SNPs in each window...

Calculating tricube smoothed delta SNP index...

Calculating G and G' statistics...

Using deltaSNP-index to filter outlier regions with a threshold of 0.1

Estimating the mode of a trimmed G prime set using the 'modeest' package...

Calculating p-values...

#Run QTLseq analysis

```
df_filt <- QTLseqr::runQTLseqAnalysis(  
  SNPset = df_filt,  
  windowSize = 1e6,  
  popStruc = "F2",  
  bulkSize = c(385, 430),  
  replications = 10000,  
  intervals = c(95, 99)  
)
```

Counting SNPs in each window...

Calculating tricube smoothed delta SNP index...

Returning the following two sided confidence intervals: 95, 99

Variable 'depth' not defined, using min and max depth from data: 40-199

Assuming bulks selected from F2 population, with 385 and 430 individuals per bulk.

Simulating 10000 SNPs with reads at each depth: 40-199

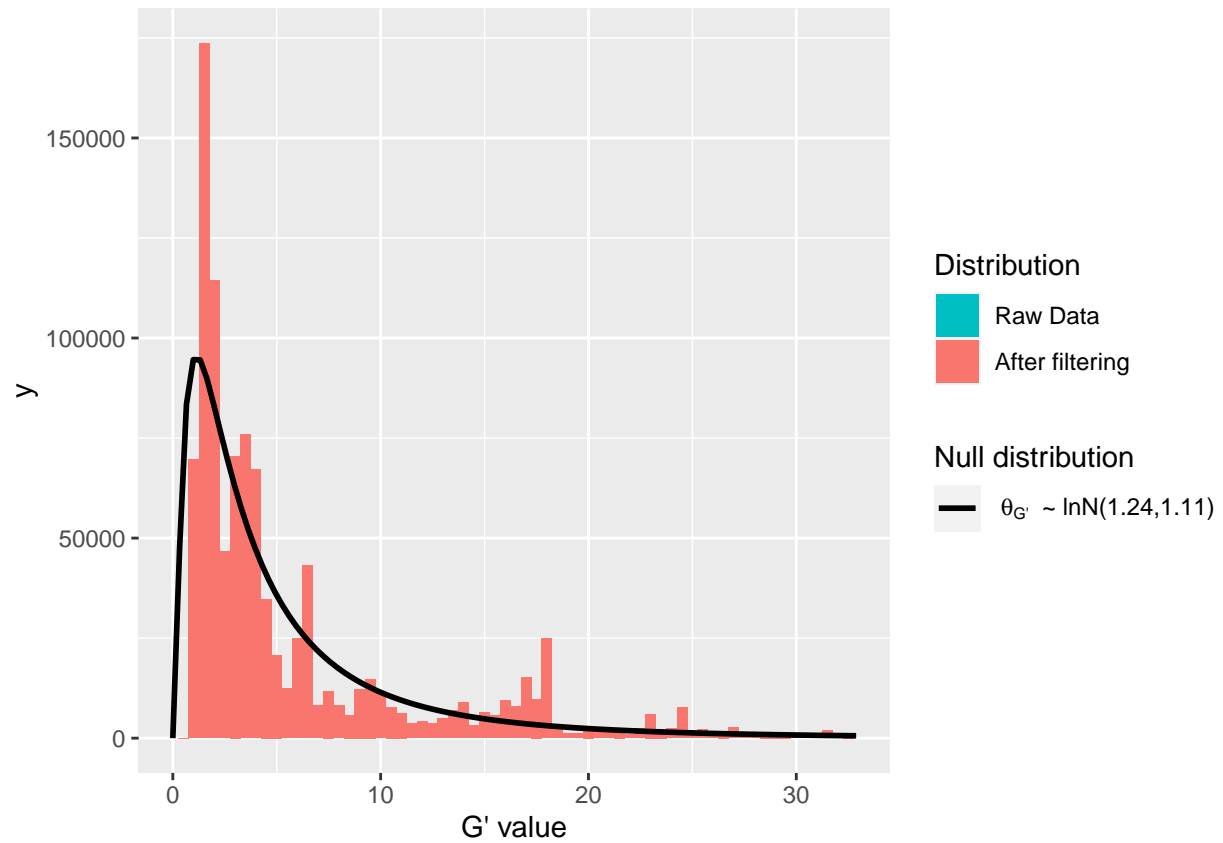
Keeping SNPs with >= 0.3 SNP-index in both simulated bulks

Joining, by = "tricubeDP"

```
QTLseqr::plotGprimeDist(SNPset = df_filt, outlierFilter = "Hampel")
```

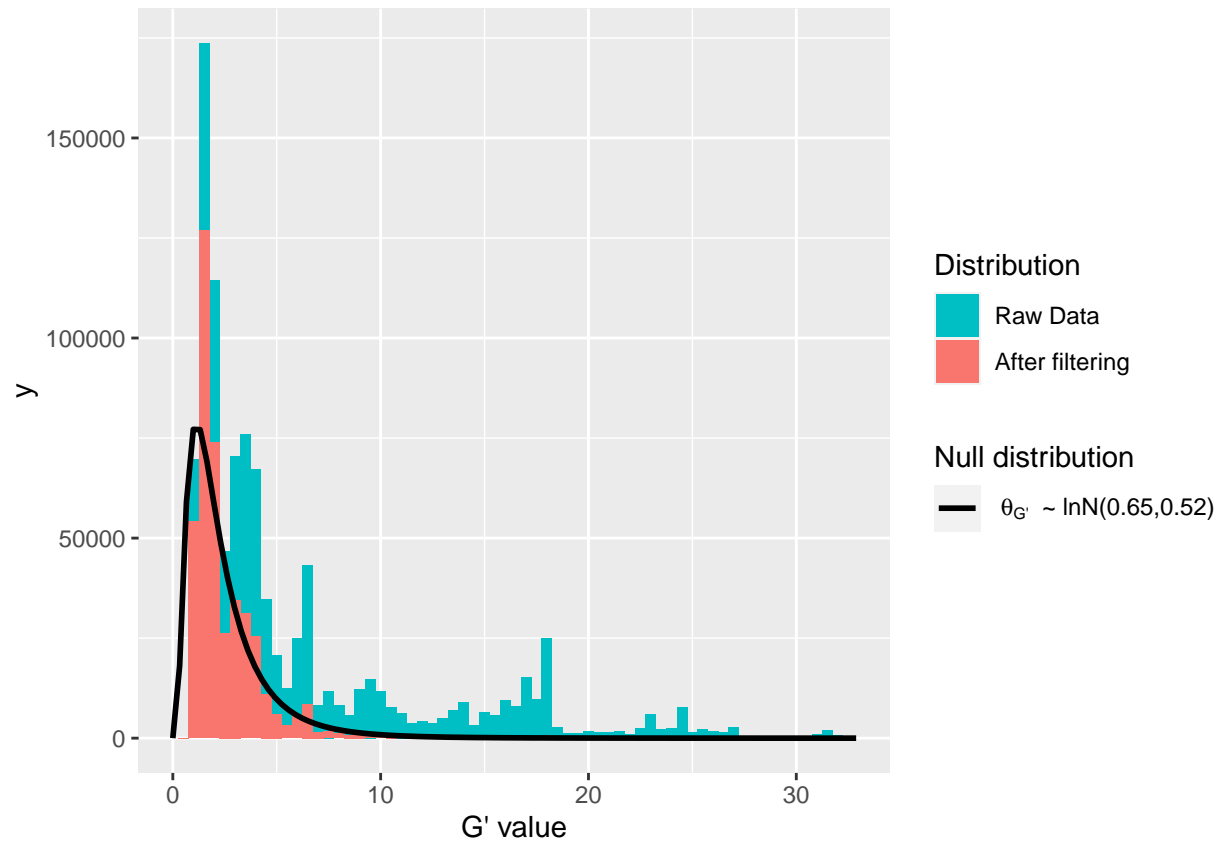
Warning: Removed 2 rows containing missing values (geom_bar).

Removed 2 rows containing missing values (geom_bar).

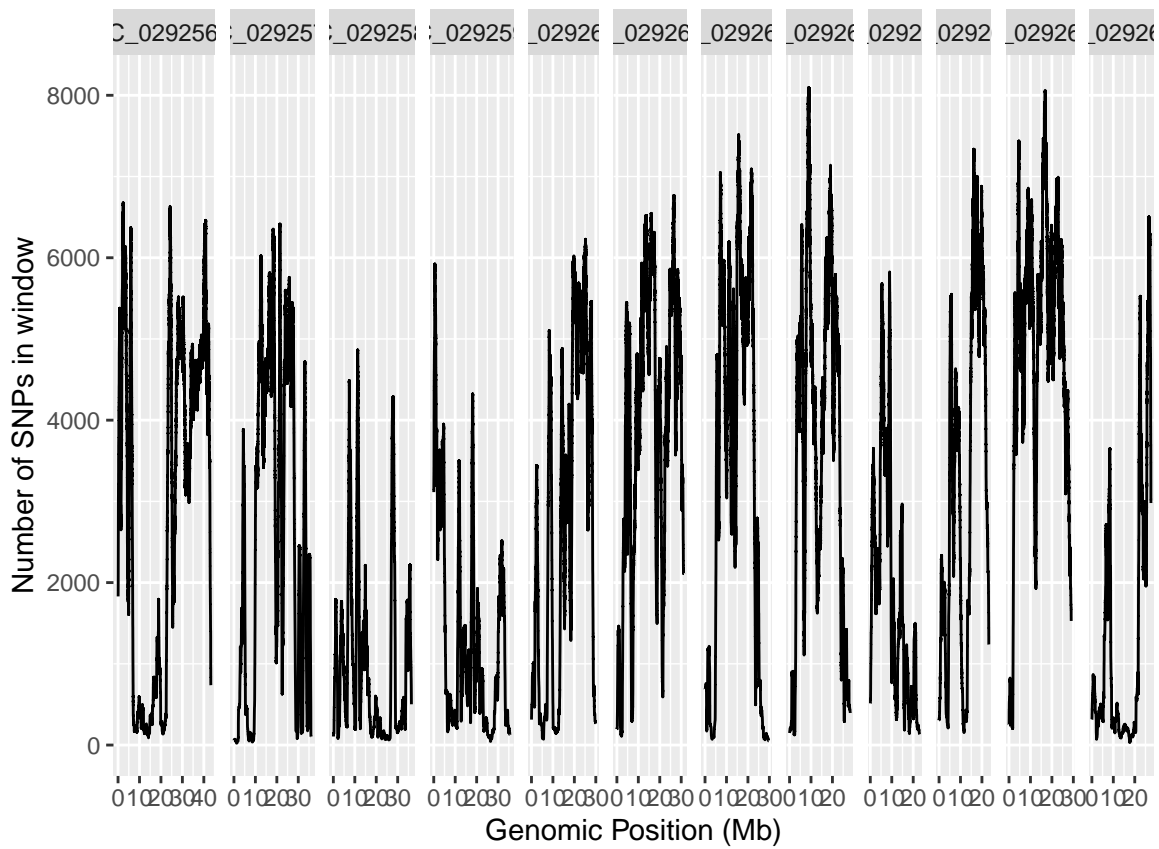


```
QTLseqr::plotGprimeDist(SNPset = df_filt, outlierFilter = "deltaSNP", filterThreshold = 0.1)
```

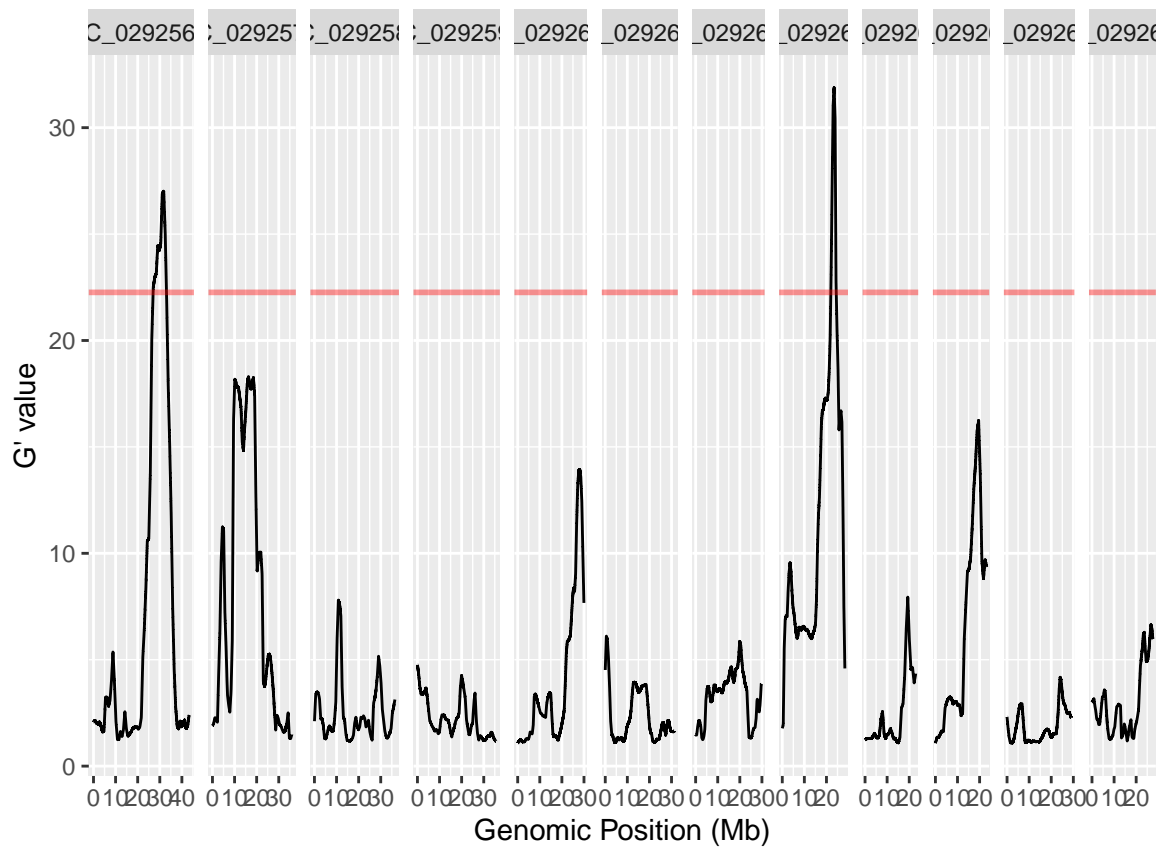
Warning: Removed 2 rows containing missing values (geom_bar).
 Removed 2 rows containing missing values (geom_bar).



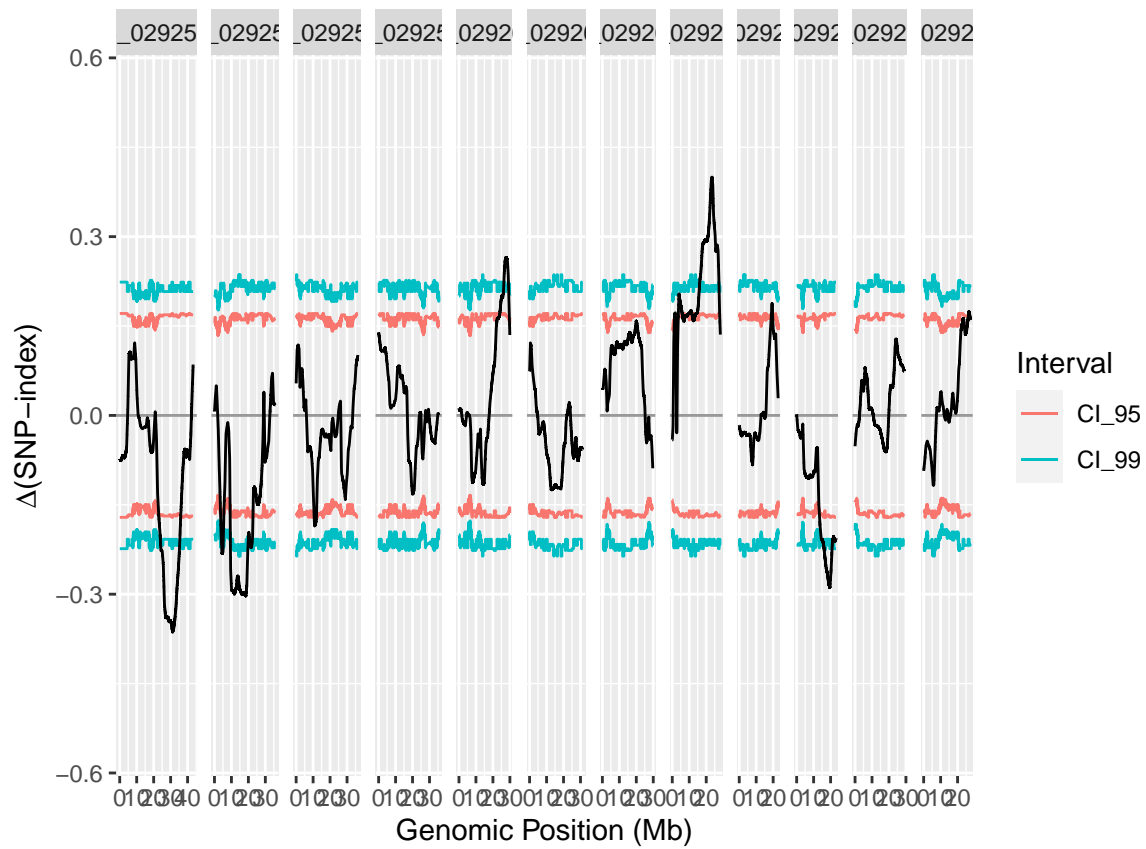
```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "nSNPs")
```



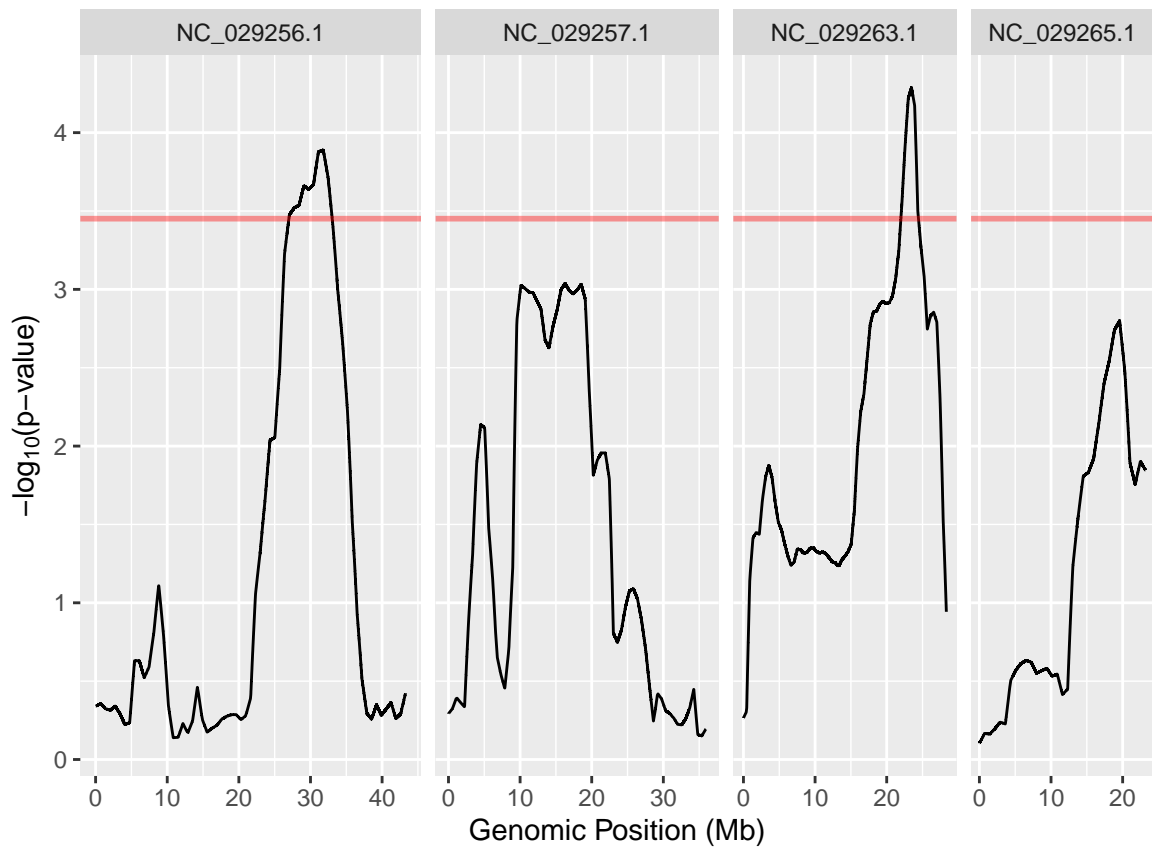
```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "Gprime", plotThreshold = TRUE, q = 0.01)
```



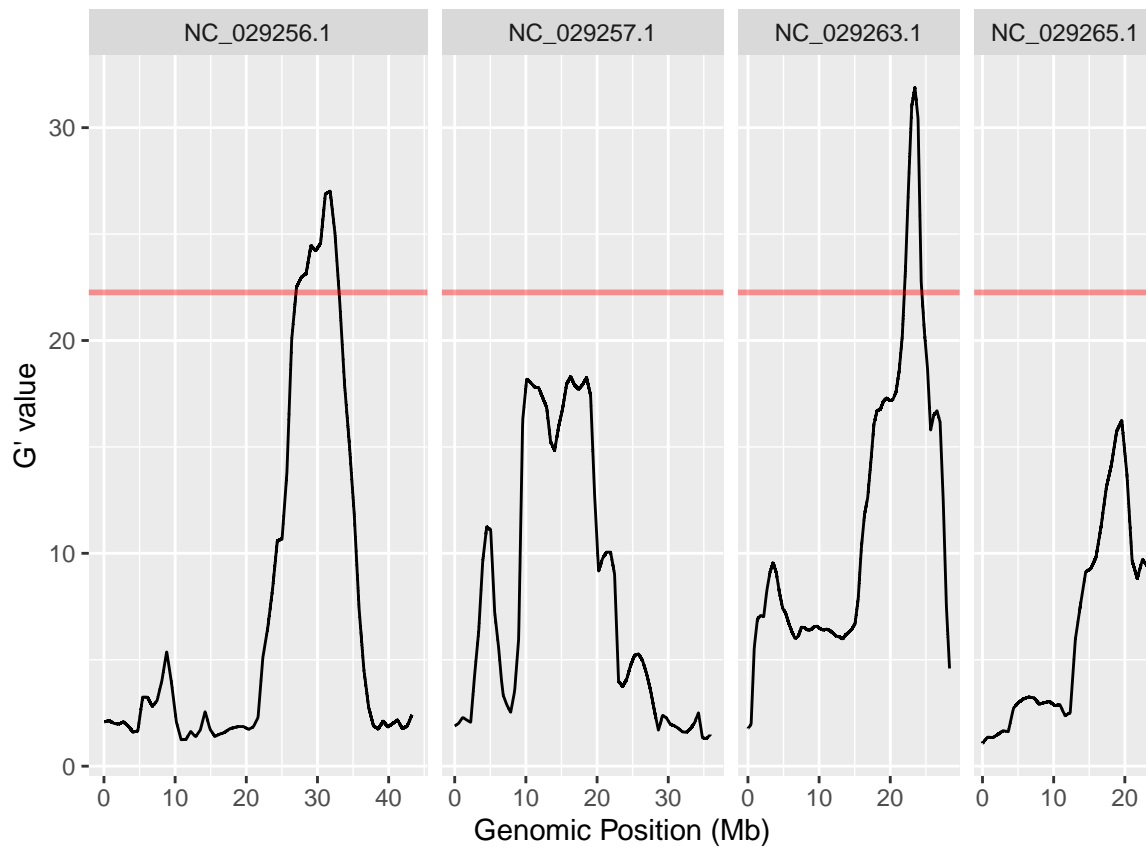
```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "deltaSNP", plotIntervals = TRUE)
```



```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "negLog10Pval", plotThreshold = TRUE, q=0.01, subset = c("NC
```

```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "Gprime", plotThreshold = TRUE, q=0.01, subset = c("NC_029256.1", "NC_029257.1", "NC_029263.1", "NC_029265.1"))
```



```
QTLseqr::getQTLTable(SNPset = df_filt, alpha = 0.01, export = TRUE, fileName = "my_BSA_QTL.csv")
```

```
##          CHROM qtl      start      end  length nSNPs avgSNPs_Mb peakDeltaSNP
## 1 NC_029256.1   1 26986793 33033424 6046631 26816      4435  -0.3636383
## 2 NC_029263.1   2 21980124 24413447 2433323  8704      3577   0.3999064
## posPeakDeltaSNP avgDeltaSNP maxGprime posMaxGprime meanGprime sdGprime
## 1          31110319 -0.3456919  27.01544    31780847   24.38459  1.321401
## 2          23458666  0.3756486  31.87973    23458666   28.26922  3.219559
##          AUCaT      meanPval      meanQval
## 1 13606901 0.0002315438 0.009554921
## 2 14451588 0.0001277682 0.009555889
```