# 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")

# Load them into your session so they are available for use
base::library("QTLseqr")
base::library("data.table")
base::library("dplyr")
base::library("tidyr")
base::library("vcfR")
base::library("ggplot2")
```

```
QTLseqr::importFromVCF(file = "wGQ-Filt-freebayes~bwa~IRGSP-1.0~both-segregant_bulks~filtered-default.vc
```

```
Scanning file to determine attributes.
File attributes:
  meta lines: 126
  header_line: 127
  variant count: 1714745
  column count: 11
Meta line 126 read in.
All meta lines processed.
gt matrix initialized.
Character matrix gt created.
  Character matrix gt rows: 1714745
  Character matrix gt cols: 11
  skip: 0
  nrows: 1714745
  row_num: 0
Processed variant 1000Processed variant 2000Processed variant 3000Processed variant 4000Processed varian
All variants processed
```

```
Keeping SNPs that pass all filters Either PASS or No Filter
```

```
Extracting gt element AD
Extracting gt element DP
```

```
Extracting gt element GQ
```

```
Removing the following chromosomes:
```

```r
HighBulk <- "ET-pool-385"
LowBulk <- "ES-pool-430"
file <- "Hall.csv"

#Choose which chromosomes/contigs will be included in the analysis,

chromList <- c("NC_029256.1","NC_029257.1","NC_029258.1","NC_029259.1","NC_029260.1","NC_029261.1","NC_

df <-
  QTLseqr::importFromTable(
    file = file,
    highBulk = HighBulk,
    lowBulk = LowBulk,
    chromList = chromList,
    sep = ","
  )
```

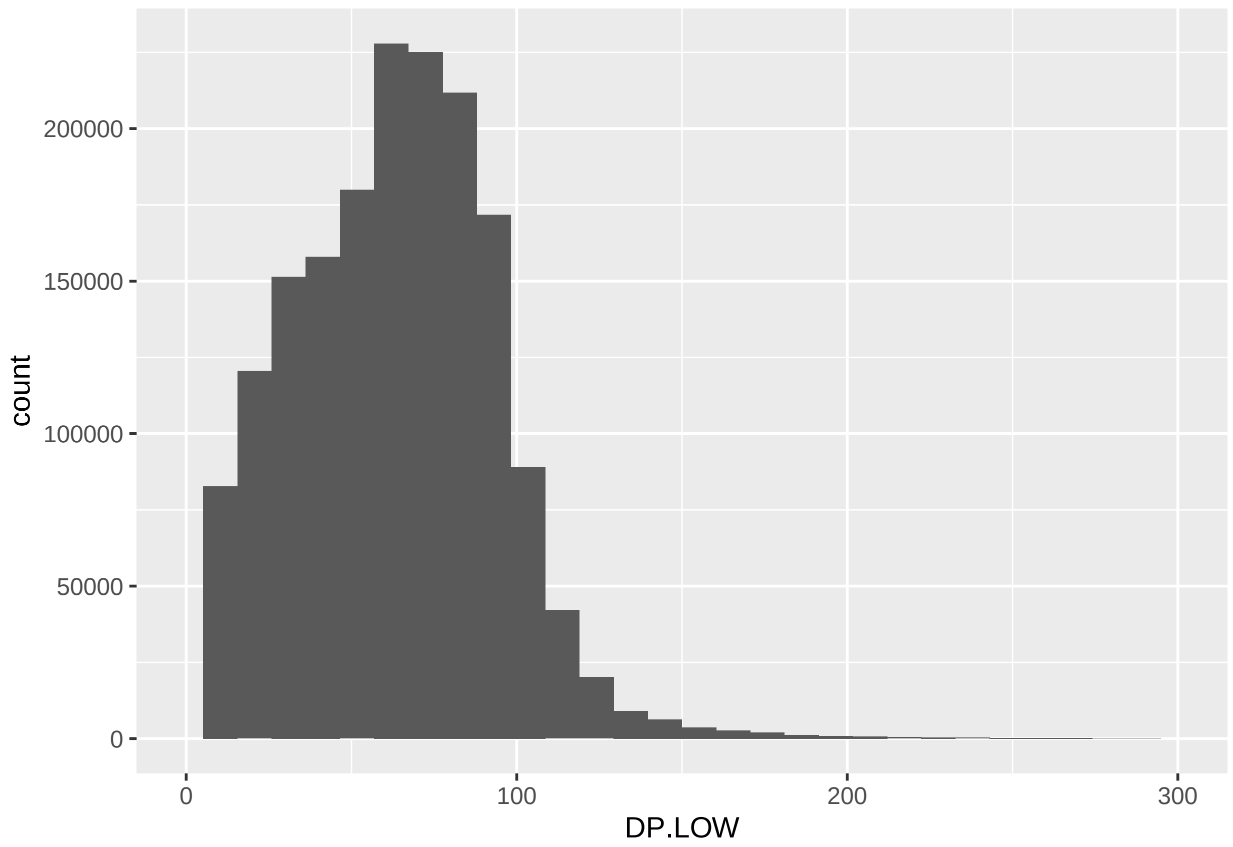
```
Removing the following chromosomes:
```

```
Renaming the following columns: AD_REF.ET-pool-385, AD_ALT.ET-pool-385
```

```
Renaming the following columns: AD_REF.ES-pool-430, AD_ALT.ES-pool-430
```

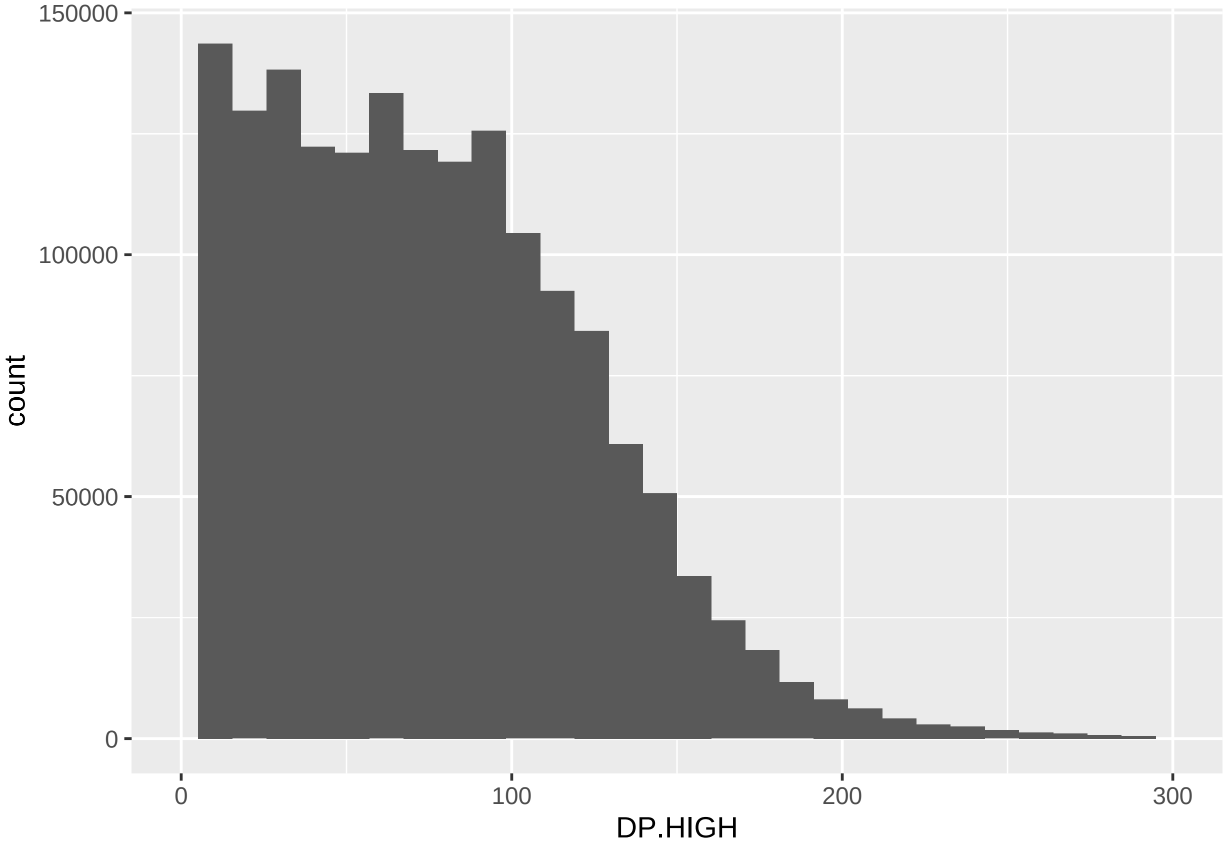
```r
ggplot(data = df) + geom_histogram(aes(x = DP.LOW)) + xlim(0,300)
```

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



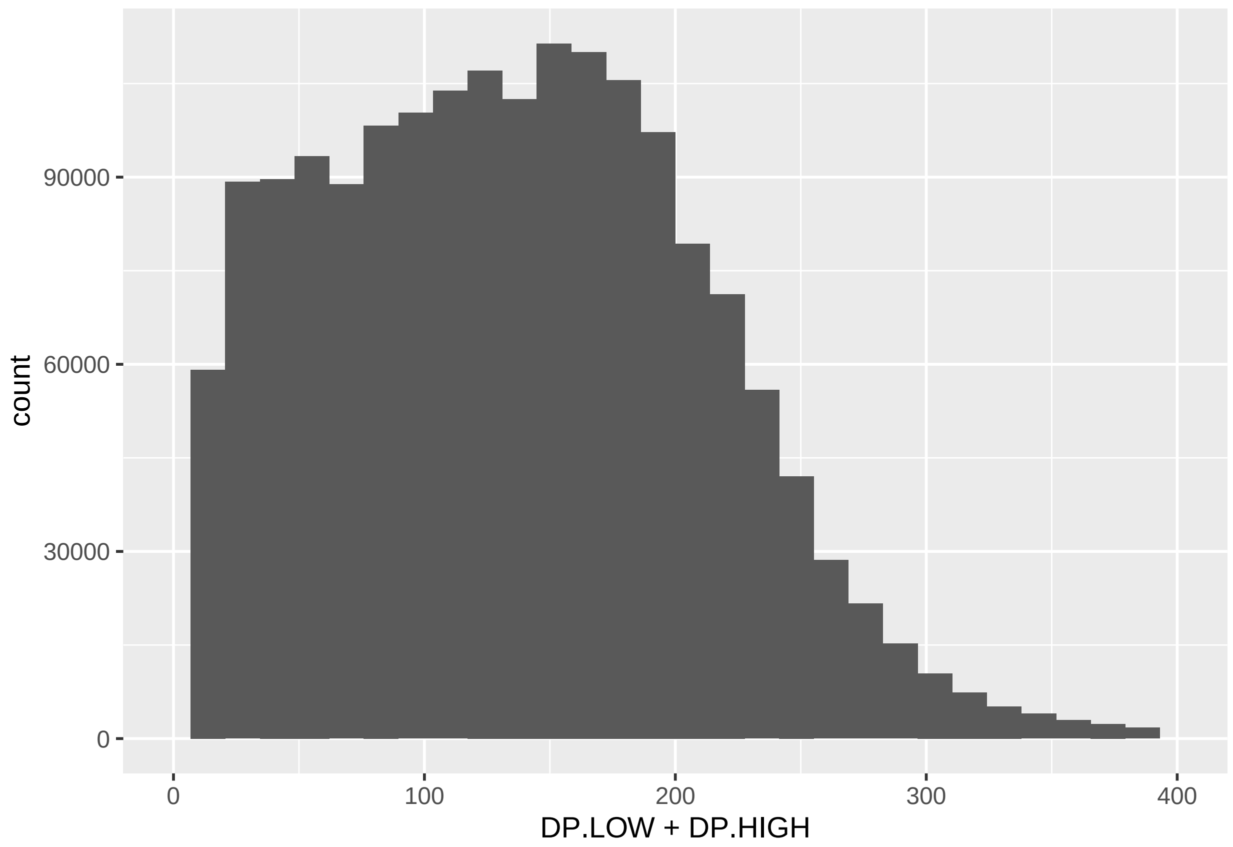
```
ggplot(data = df) + geom_histogram(aes(x = DP.HIGH)) + xlim(0,300)
```

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



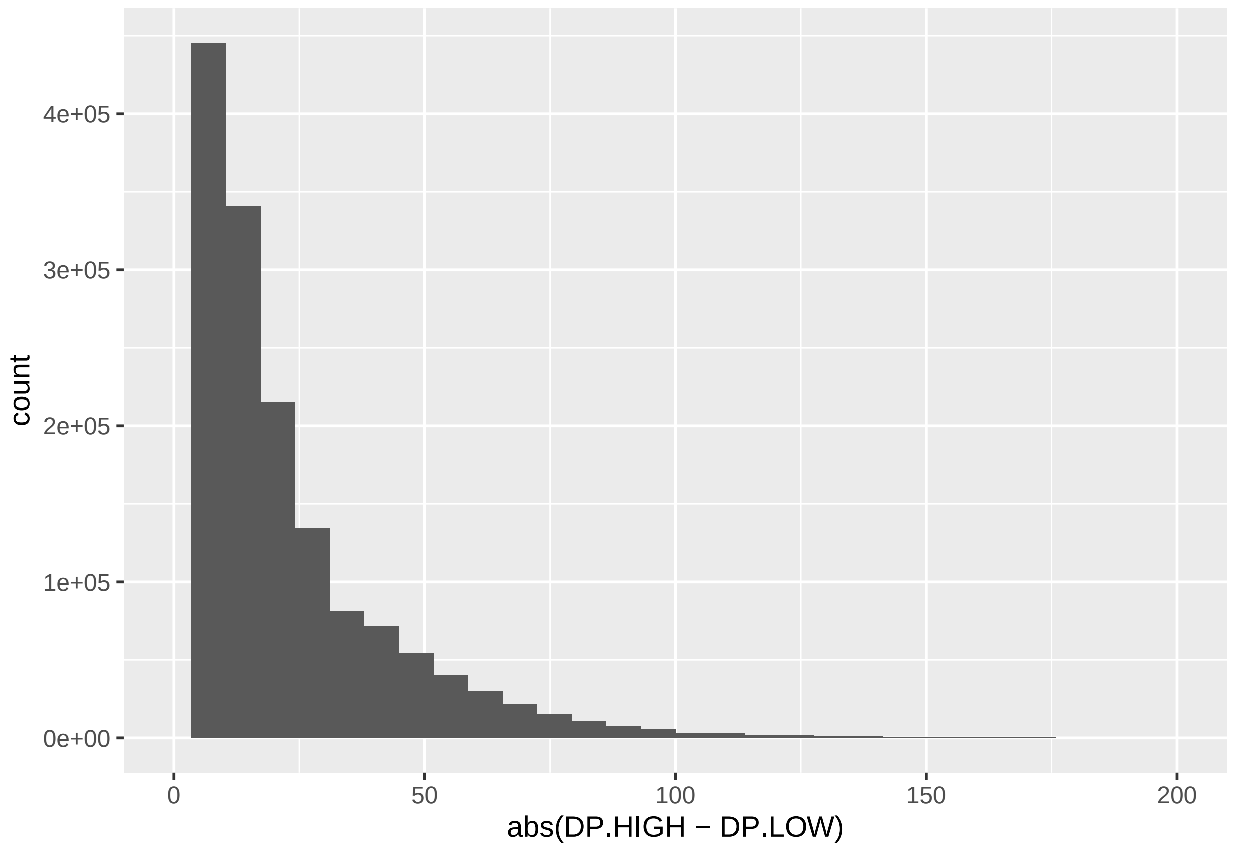
```
ggplot(data = df) + geom_histogram(aes(x = DP.LOW + DP.HIGH)) + xlim(0,400)
```

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



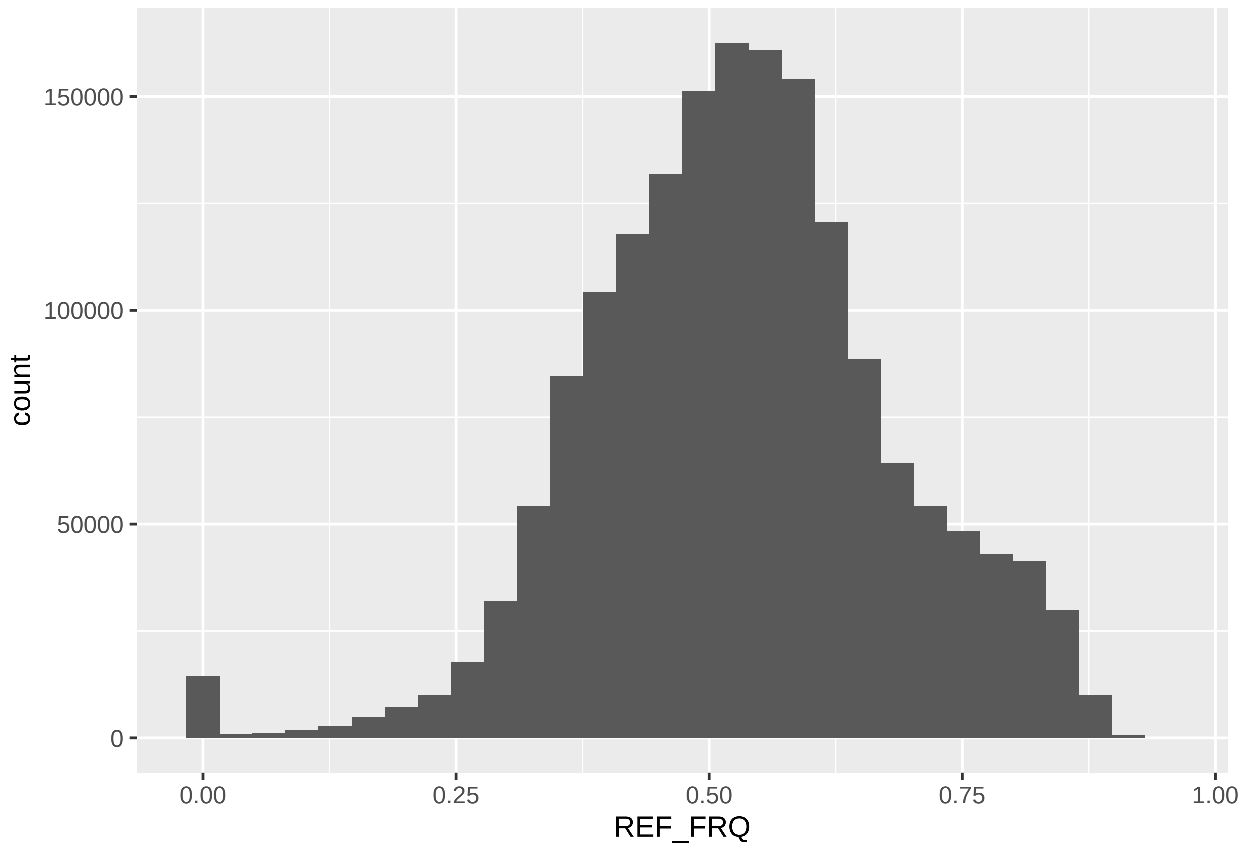
```
ggplot(data = df) + geom_histogram(aes(x =  abs(DP.HIGH - DP.LOW))) + xlim(0, 200)
```

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



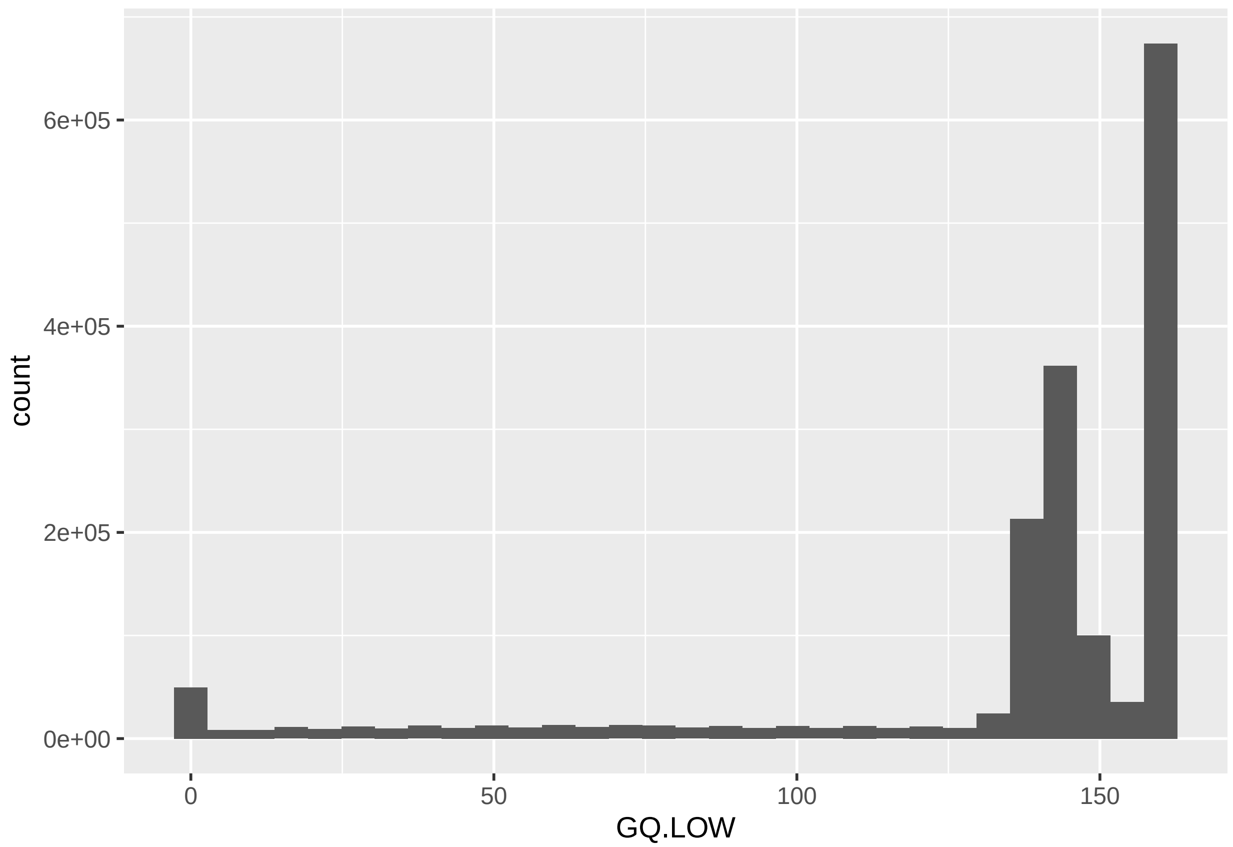
```
ggplot(data = df) + geom_histogram(aes(x = REF_FRQ))
```

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



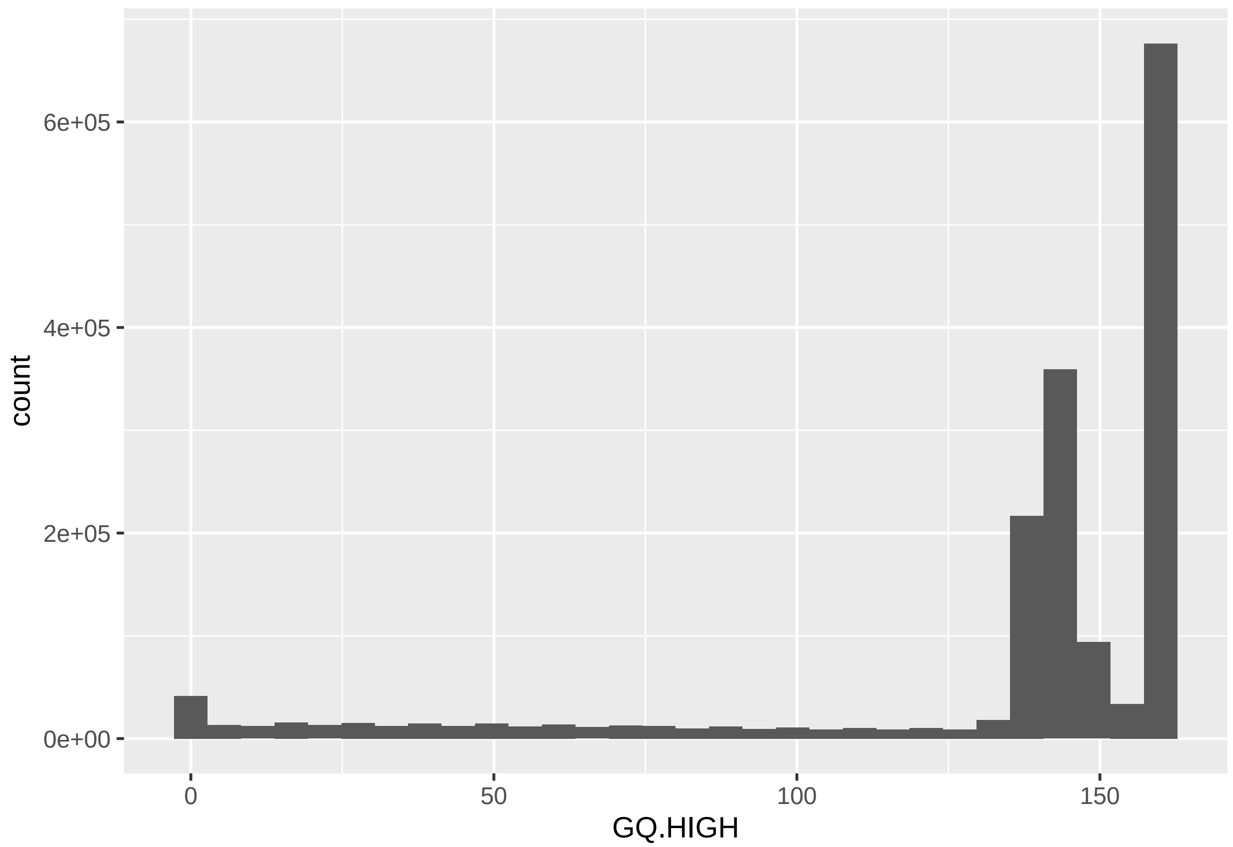
```
ggplot(data = df) + geom_histogram(aes(x = GQ.LOW))
```

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



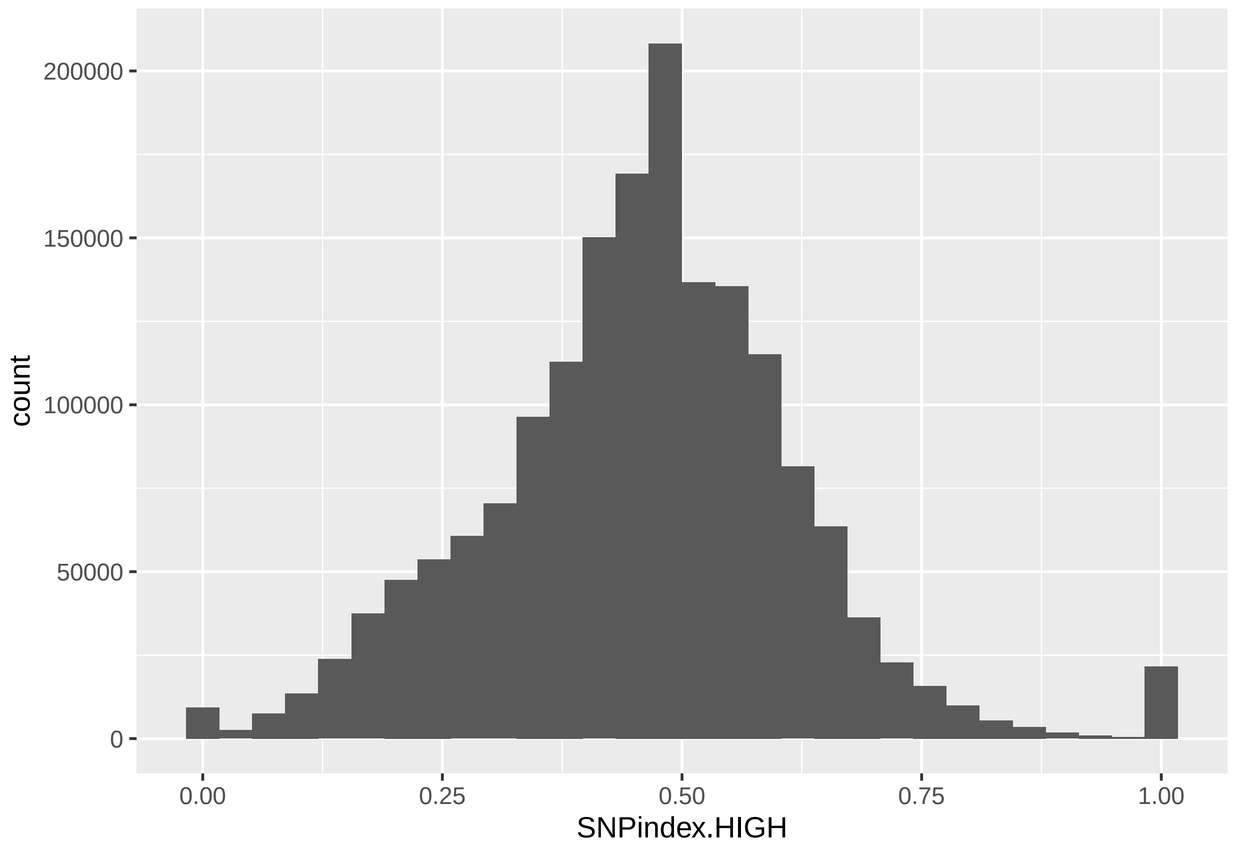
```
ggplot(data = df) + geom_histogram(aes(x = GQ.HIGH))
```

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



```
ggplot(data = df) + geom_histogram(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 <= REF_FRQ <= 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
```

```r
#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...
```

```r
#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"
```

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

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

```r
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).
```

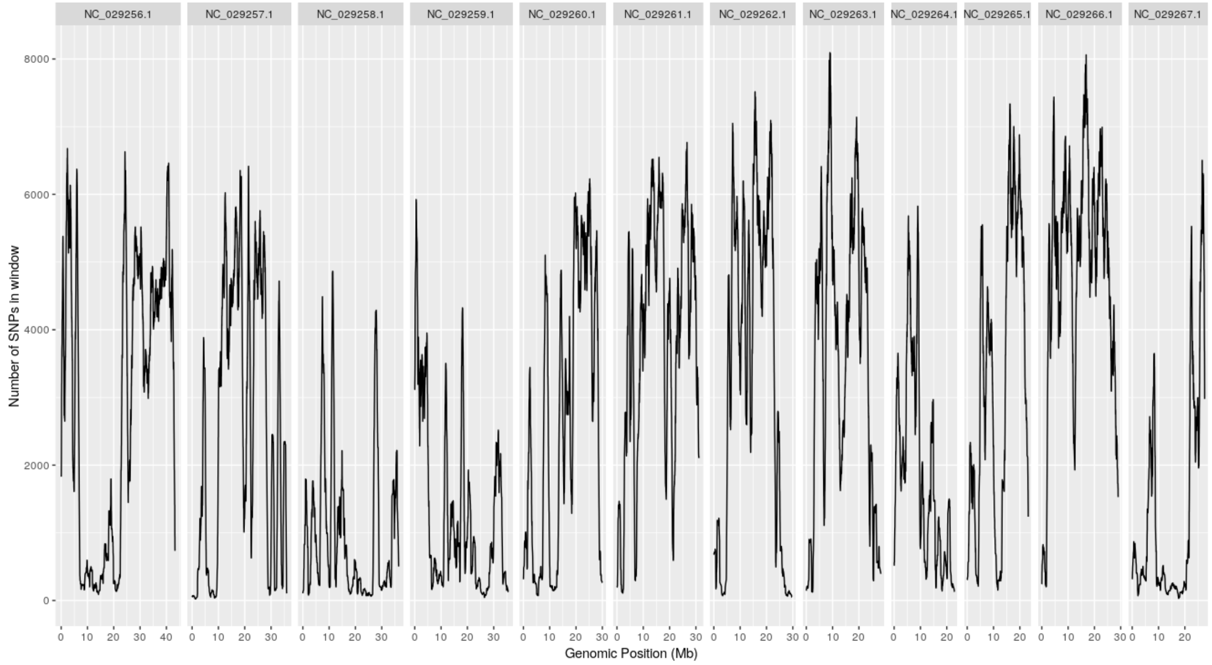
```r
QTLseqr::plotQTLStats(SNPset = df_filt, var = "nSNPs")
```

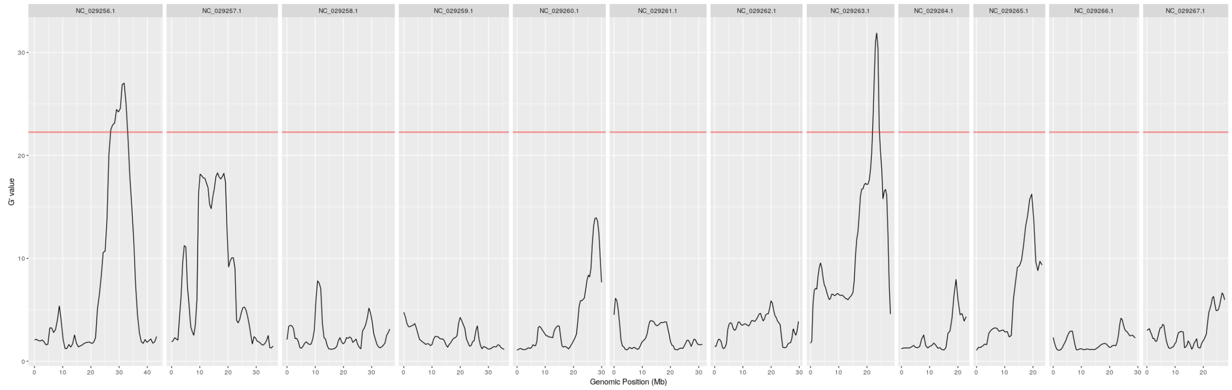
```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "Gprime", plotThreshold = TRUE, q = 0.01)
```

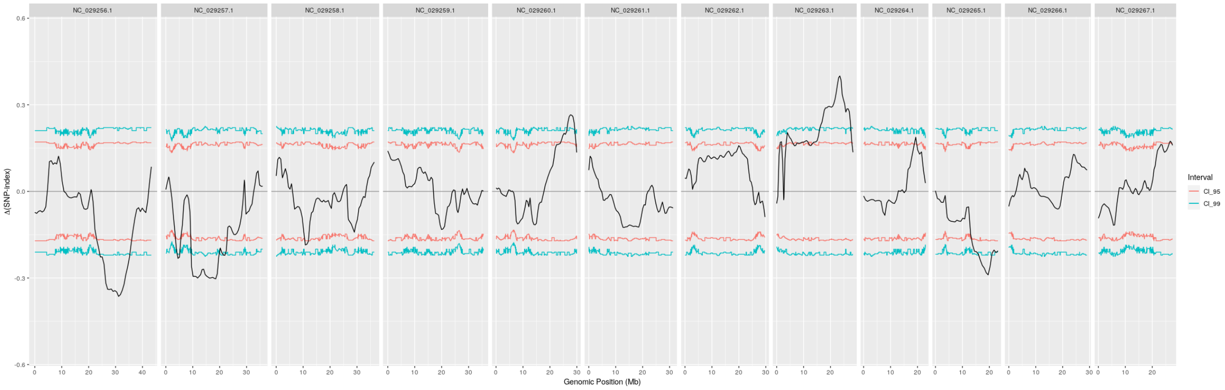
```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "deltaSNP", plotIntervals  = TRUE)
```

```
QTLseqr::plotQTLStats(SNPset = df_filt, var = "negLog10Pval",plotThreshold = TRUE,q=0.01,
                      subset = c("NC_029256.1","NC_029257.1","NC_029263.1","NC_029265.1"))
```

```
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"))
```







```r
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
```