

class13

```
library(readr)
col <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "sstart", "send", "eval", "bitscore")
blastresults <- read_tsv("mm-second.x.zebrafish.tsv", colnames <- col)

## Rows: 67387 Columns: 12

## -- Column specification -----
## Delimiter: "\t"
## chr (2): qseqid, sseqid
## dbl (10): pident, length, mismatch, gapopen, qstart, qend, sstart, send, eval, bitscore

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

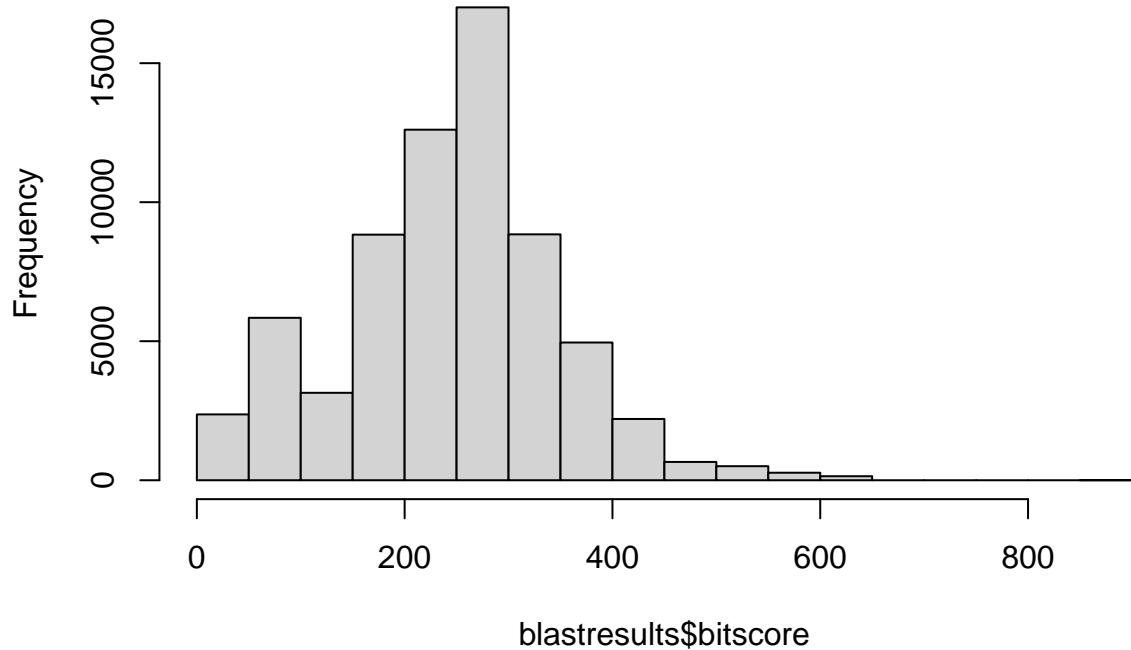
#Check if the colnames are correct
colnames(blastresults)

## [1] "qseqid"    "sseqid"    "pident"    "length"    "mismatch"   "gapopen"
## [7] "qstart"    "qend"      "sstart"    "send"      "eval"       "bitscore"
```

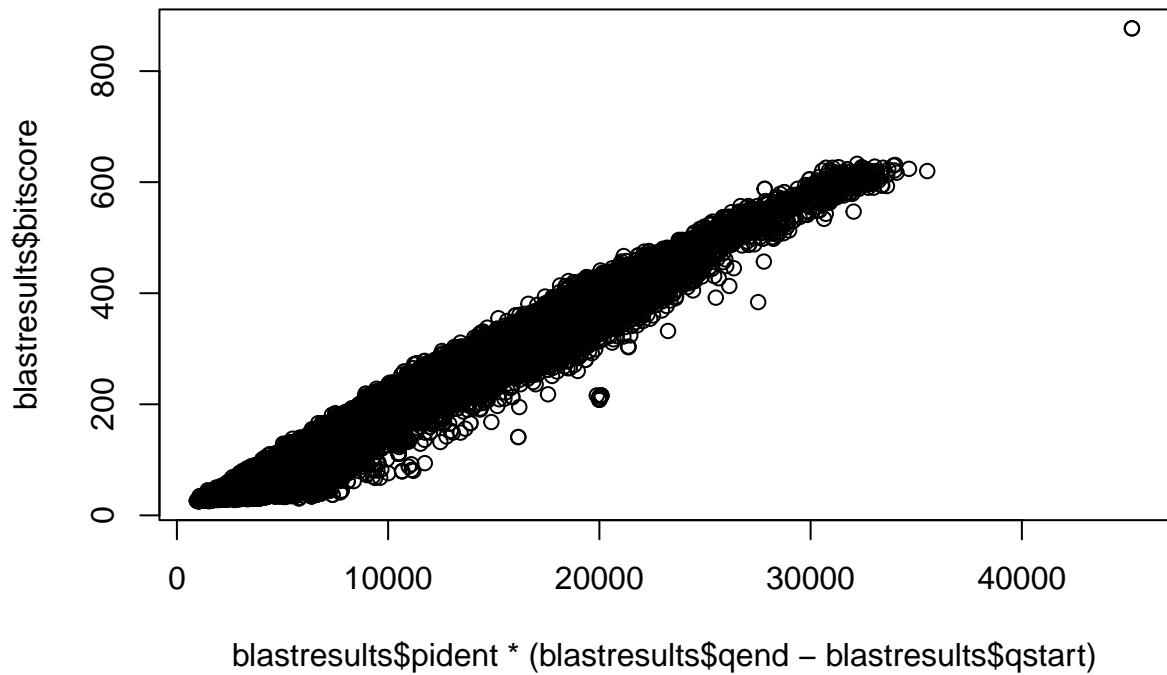
Now let's plot this.

```
#Using histogram
hist(blastresults$bitscore, breaks = 30)
```

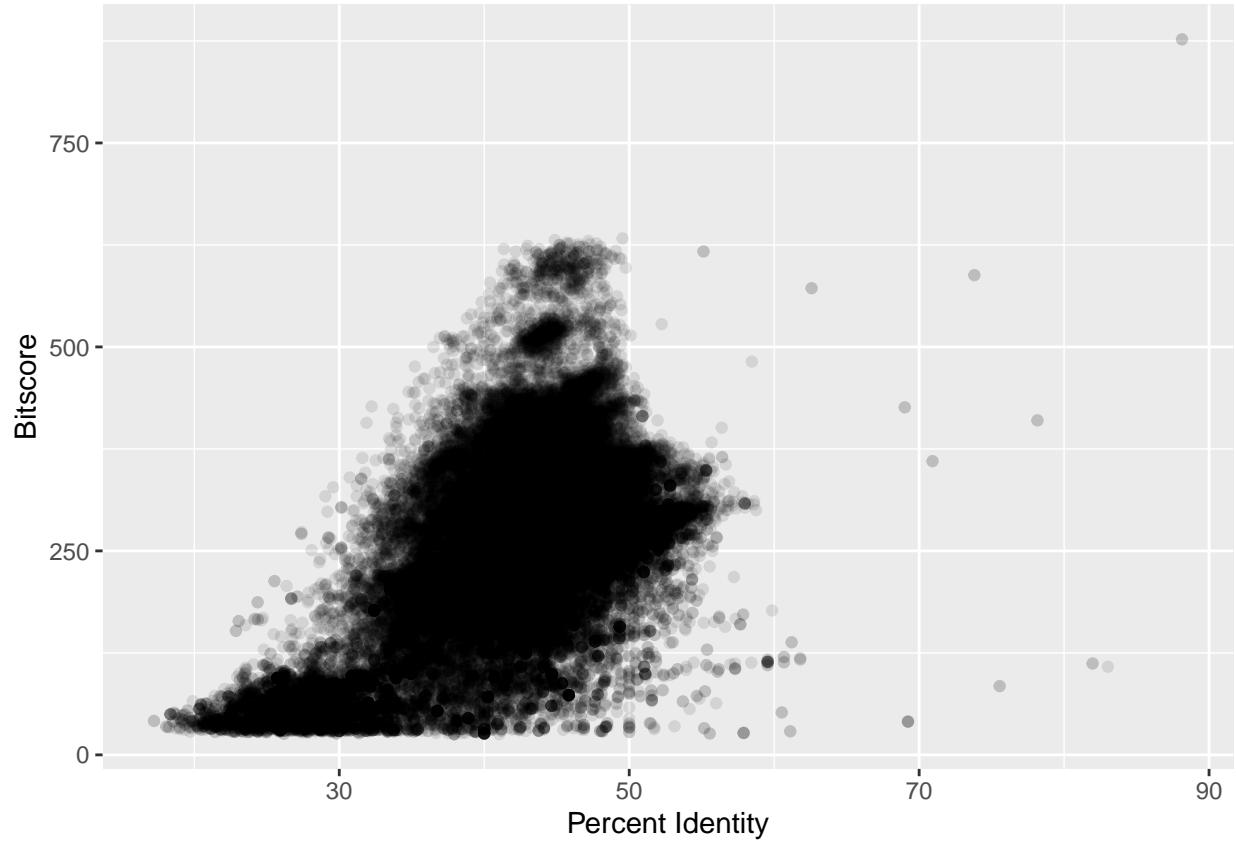
Histogram of blastresults\$bitscore



```
plot(blastresults$pident * (blastresults$qend - blastresults$qstart), blastresults$bitscore)
```



```
#Using ggplot
library(ggplot2)
ggplot(blastresults, aes(pident, bitscore)) + geom_point(alpha=0.1) + labs(x = "Percent Identity", y = "Bitscore")
```



```
ggplot(blastresults, aes((blastresults$pident * (blastresults$qend - blastresults$qstart)), bitscore)) +  
  ## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

