

BIMM 143 Class 13

Anika Bhattacharjya (A15459876)

Read Zebrafish TSV File

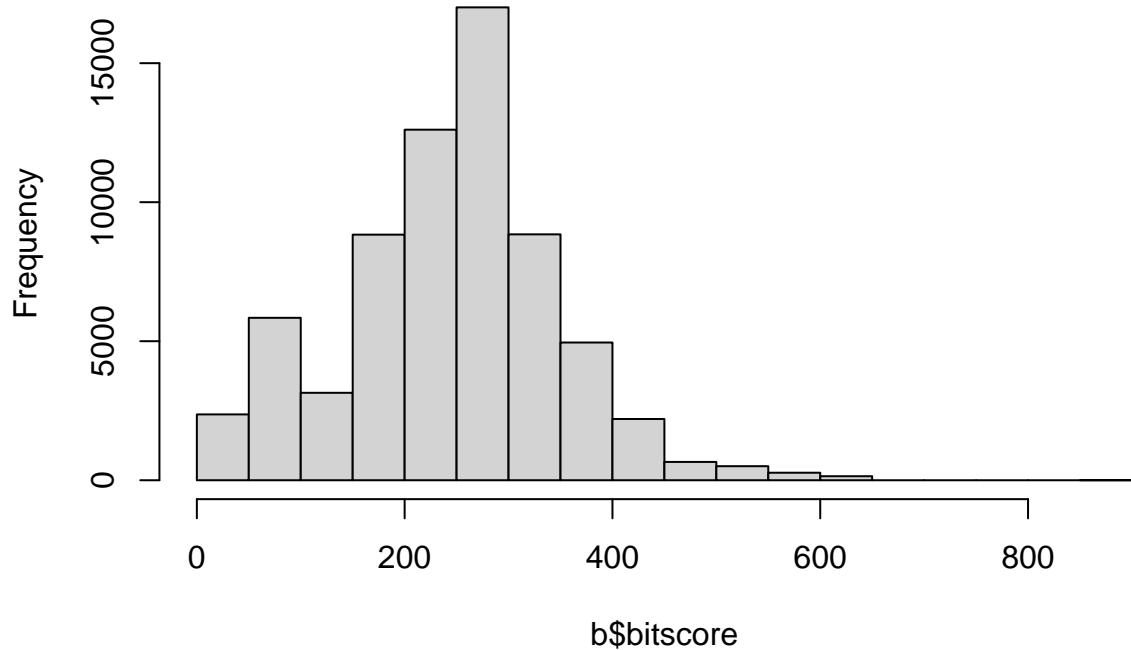
```
b <- read.delim("mm-second.x.zebrafish.tsv")
head(b)

##   YP_220550.1      NP_059331.1 X69.010 X313 X97 X0  X4  X316 X10 X322 X1.24e.150
## 1 YP_220551.1      NP_059332.1  44.509  346 188  3   1  344   1  344   8.62e-92
## 2 YP_220551.1      NP_059341.1  24.540  163 112  3  112  263 231  393   5.14e-06
## 3 YP_220551.1      NP_059340.1  26.804   97  65  2   98  188 200  296   1.00e-01
## 4 YP_220552.1      NP_059333.1  88.132  514  61  0   1  514   1  514   0.00e+00
## 5 YP_220552.1     XP_021326074.1 31.818   66  32  2  427  482  16   78   6.70e+00
## 6 YP_220552.1     NP_001373511.1 31.818   66  32  2  427  482   48  110  7.50e+00
##          X426
## 1  279.0
## 2  49.7
## 3  35.8
## 4 877.0
## 5  29.3
## 6  29.6

Assign column names
colnames(b) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "ssta

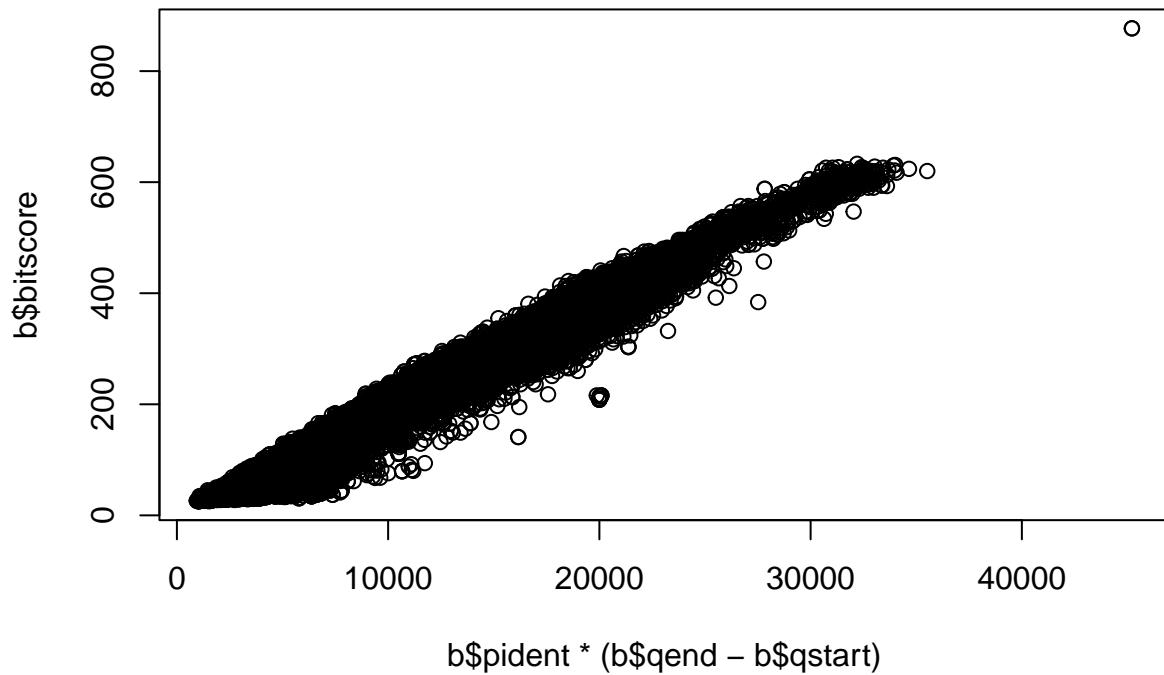
Make a Histogram
hist(b$bitscore)
```

Histogram of b\$bitscore



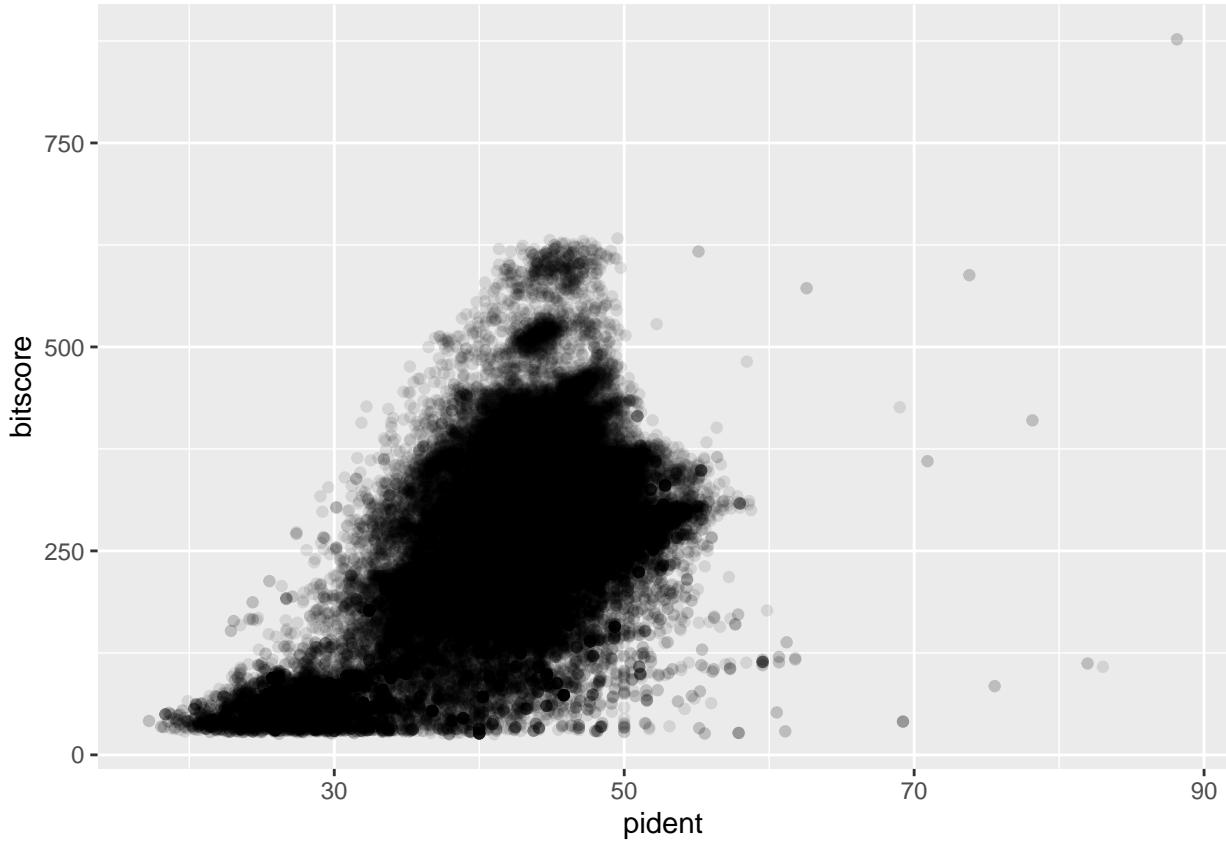
Making plots of the data

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



Use ggplot for better plot

```
library(ggplot2)
ggplot(b, aes(pident, bitscore)) + geom_point(alpha=0.1)
```



```
ggplot(b, aes((b$pident * (b$qend - b$qstart)), bitscore)) + geom_point(alpha=0.1) + geom_smooth()  
## Warning: Use of `b$pident` is discouraged. Use `pident` instead.  
## Warning: Use of `b$qend` is discouraged. Use `qend` instead.  
## Warning: Use of `b$qstart` is discouraged. Use `qstart` instead.  
## Warning: Use of `b$pident` is discouraged. Use `pident` instead.  
## Warning: Use of `b$qend` is discouraged. Use `qend` instead.  
## Warning: Use of `b$qstart` is discouraged. Use `qstart` instead.  
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

