

class 13

Ramola Baviskar (PID A12228297)

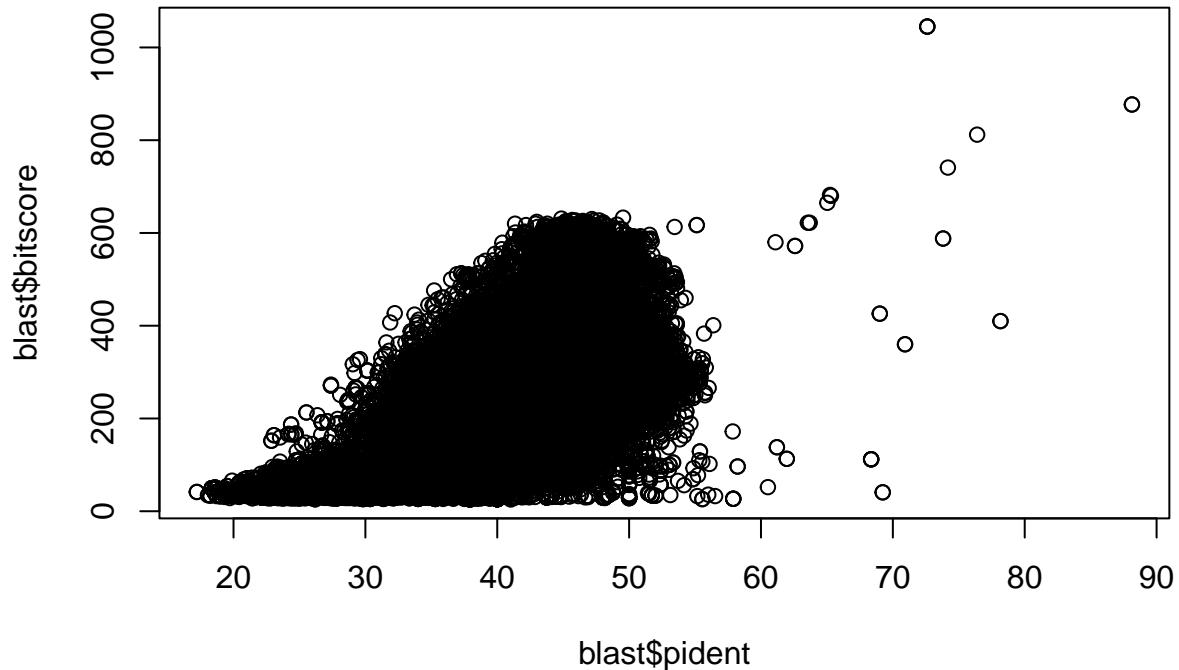
3/1/2022

```
blast <- read.table("mm-second.x.zebrafish.tsv")
```

Set the column names

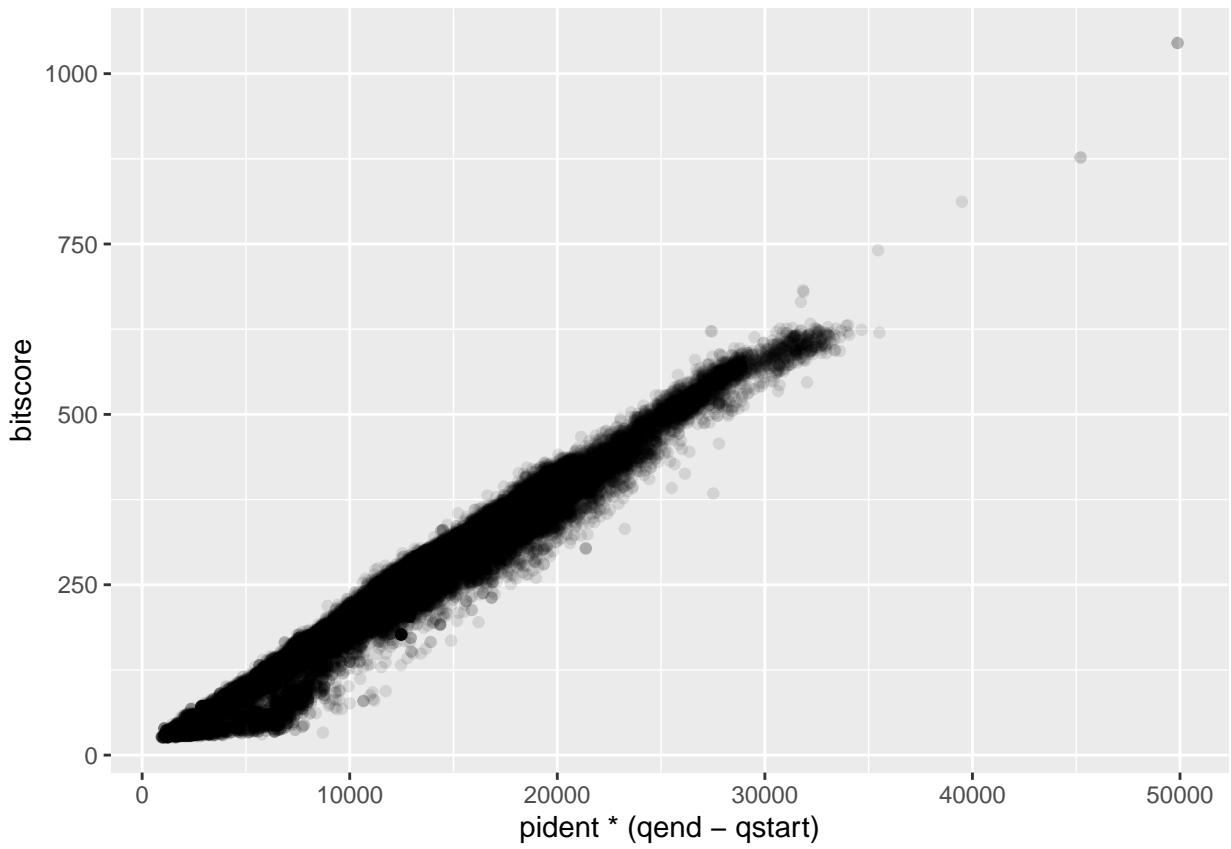
```
colnames(blast) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "qseq", "sseq", "bitscore")
```

```
plot(blast$pident, blast$bitscore)
```



We can have very high percent identity even for short alignments where the bitscore (alignment score) won't be that high. Let's try factoring that into our analysis.

```
library(ggplot2)
ggplot(blast) + aes(pident * (qend - qstart), bitscore) + geom_point(alpha=0.1)
```



```
ggplot(blast) + aes(pident * (qend - qstart), bitscore) + geom_point(alpha=0.1) + geom_smooth()

## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

