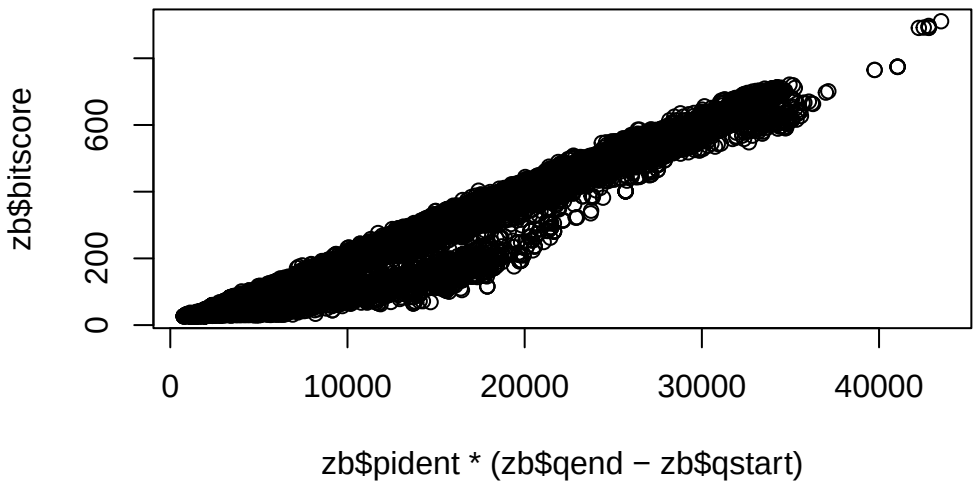
Class16

James Garza

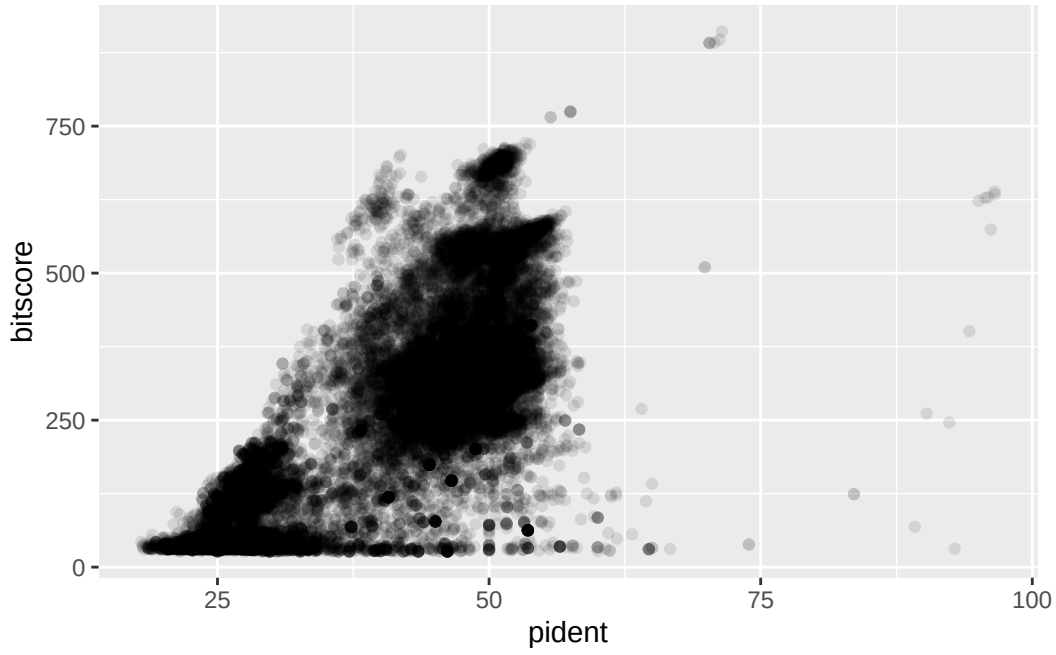
```
zb <- read.table(file = 'mm-second.x.zebrafish.tsv', sep='\t')
colnames(zb) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart",
head(zb)
```

```
      qseqid         sseqid pident length mismatch gapopen qstart qend sstart
1 NP_598866.1 XP_009294521.1 46.154    273      130       6      4  267    420
2 NP_598866.1 NP_001313634.1 46.154    273      130       6      4  267    476
3 NP_598866.1 XP_009294513.1 46.154    273      130       6      4  267    475
4 NP_598866.1 NP_001186666.1 33.071    127       76       5      4  126    338
5 NP_598866.1 NP_001003517.1 30.400    125       82       4      4  126    344
6 NP_598866.1 NP_001003517.1 30.645     62       41       2     53  113     43
  send   evalue bitscore
1  684 1.70e-63    214.0
2  740 4.51e-63    214.0
3  739 4.69e-63    214.0
4  459 5.19e-12     67.8
5  465 2.67e-11     65.5
6  103 4.40e-01     33.9
```

```
## Asuming your blast results are stored in an object called 'b'
plot(zb$pident  * (zb$qend - zb$qstart), zb$bitscore)
```




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



There appears to be a positive relationship between pident and bitscore as the data is positively correlated.

```
ggplot(zb, aes((zb$pident * (zb$qend - zb$qstart)), bitscore)) + geom_point(alpha=0.1) + g
```

```
Warning: Use of `zb$pident` is discouraged.
i Use `pident` instead.
```

```
Warning: Use of `zb$qend` is discouraged.
i Use `qend` instead.
```

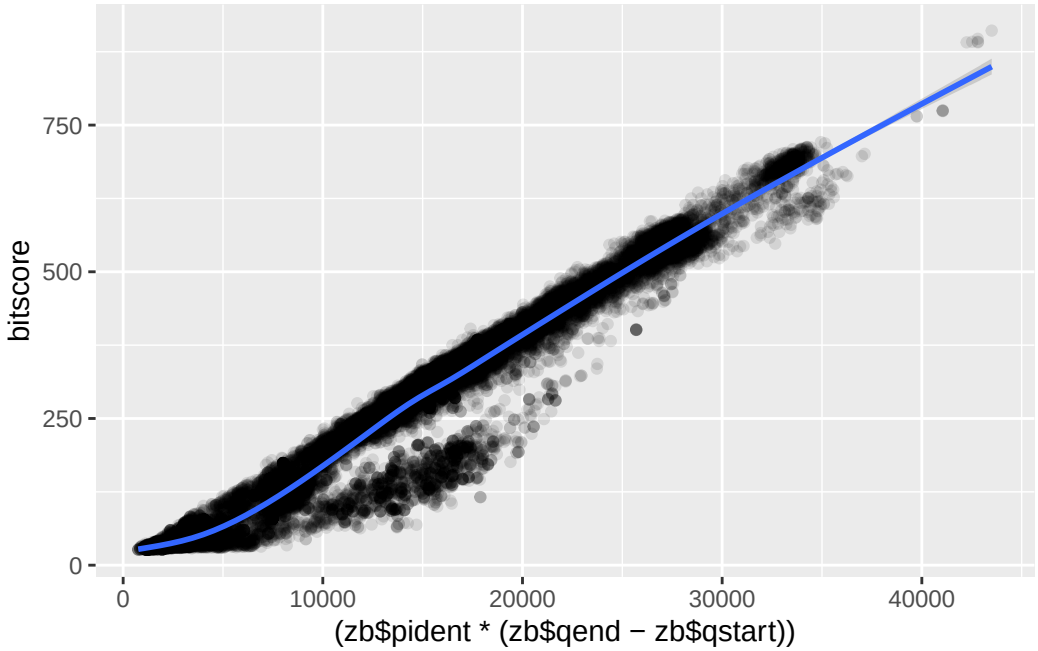
```
Warning: Use of `zb$qstart` is discouraged.
i Use `qstart` instead.
```

```
Warning: Use of `zb$pident` is discouraged.
i Use `pident` instead.
```

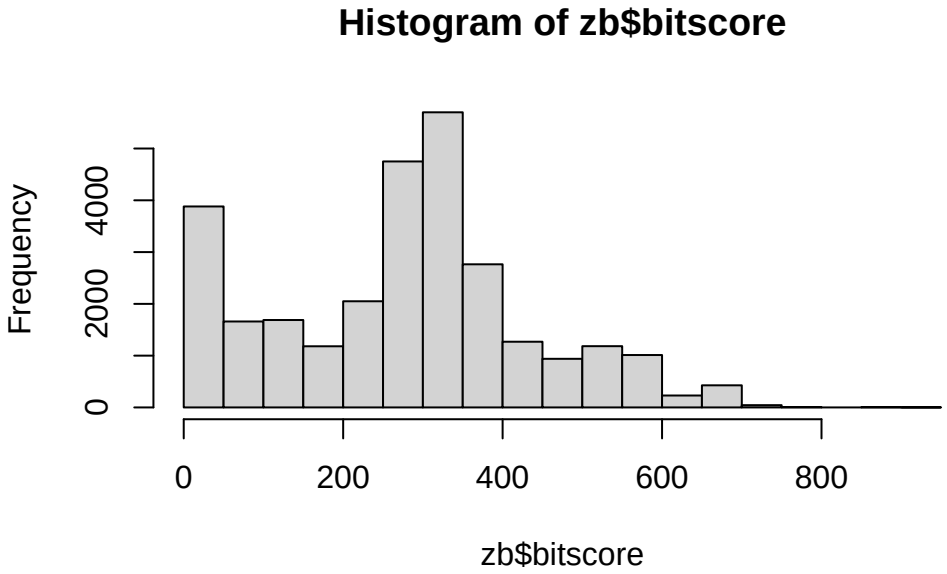
```
Warning: Use of `zb$qend` is discouraged.
i Use `qend` instead.
```

```
Warning: Use of `zb$qstart` is discouraged.
i Use `qstart` instead.
```

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




```
hist(zb$bitscore, breaks = 30)
```



As the bitscores increase, the frequency of high bitscores tends to decrease.