Class 16 Lab

Using RStudio to read data output

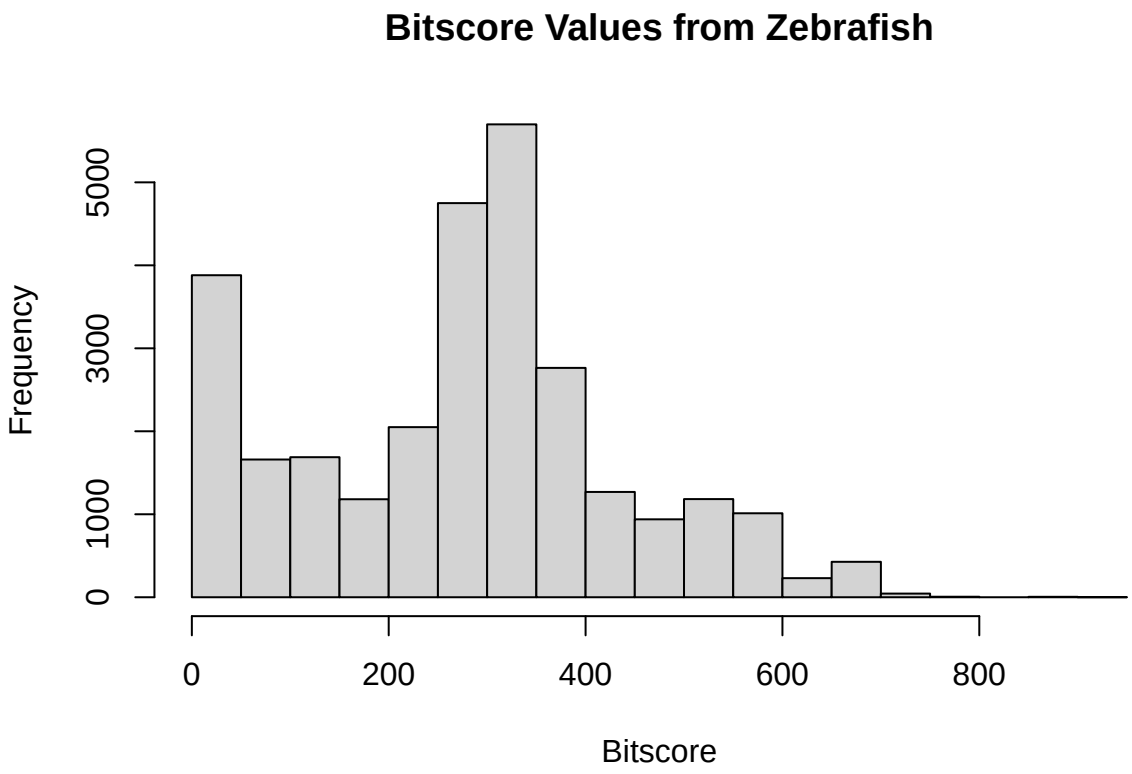
First, I need to read in my tsv file that I obtained earlier in the class:

```
data <- read.table("mm-second.x.zebrafish.tsv", col.names = c("qseqid", "sseqid", "pident", "length", "m

head(data)
```

```
##        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
```

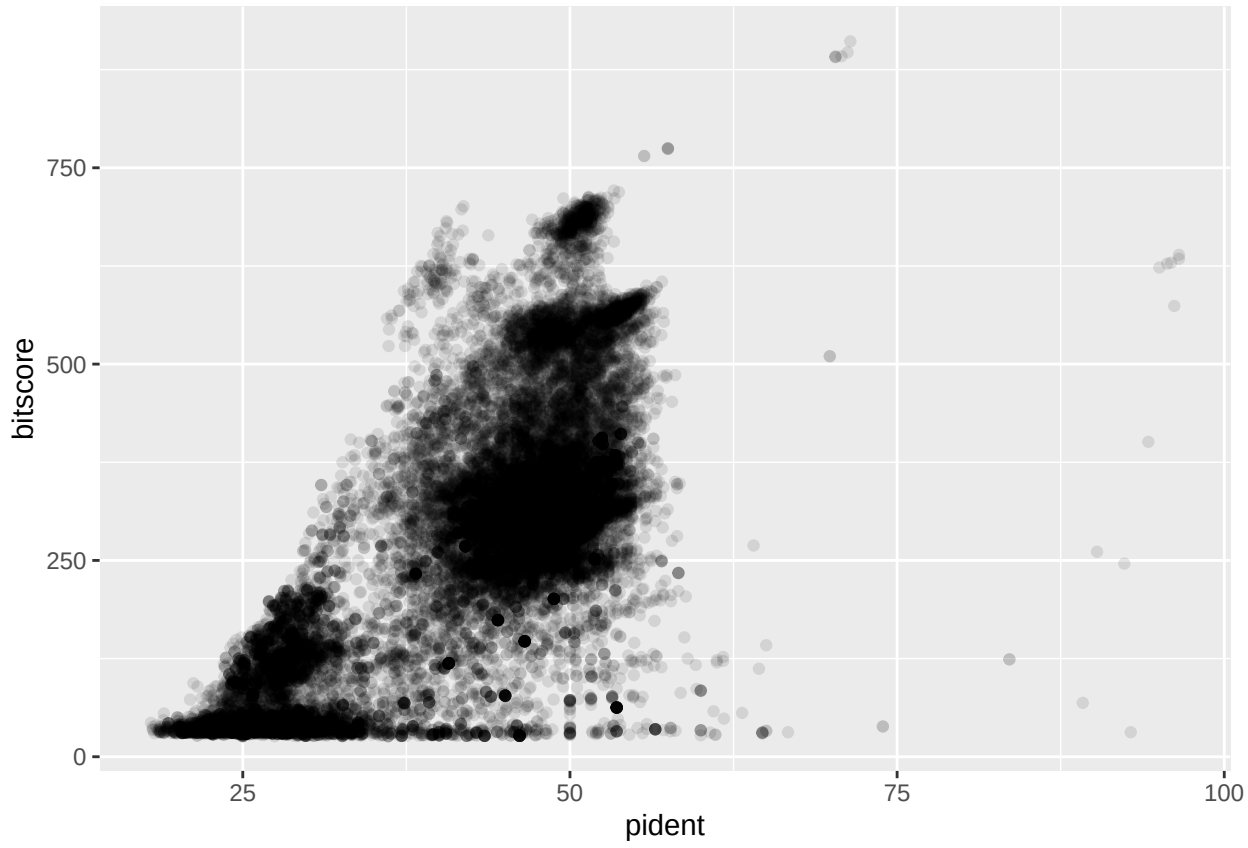
Make a histogram of the bitscore values:

```
hist(data$bitscore, breaks = 30, main = "Bitscore Values from Zebrafish",xlab = "Bitscore")
```



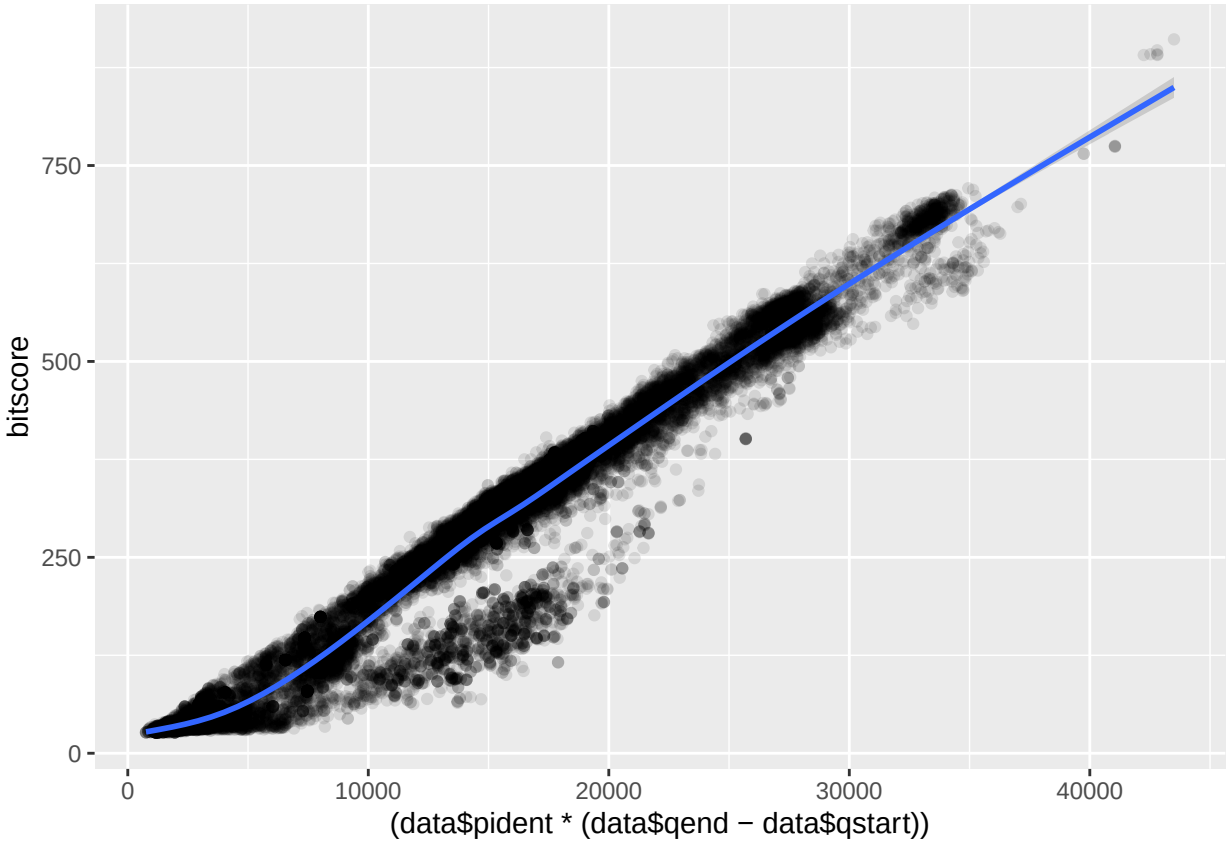
Now we will plot percent identity against bitscore to see the relationship between them using ggplot:

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




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

```
## Warning: Use of ‘data$pident‘ is discouraged.
## i Use ‘pident‘ instead.

## Warning: Use of ‘data$qend‘ is discouraged.
## i Use ‘qend‘ instead.

## Warning: Use of ‘data$qstart‘ is discouraged.
## i Use ‘qstart‘ instead.

## Warning: Use of ‘data$pident‘ is discouraged.
## i Use ‘pident‘ instead.

## Warning: Use of ‘data$qend‘ is discouraged.
## i Use ‘qend‘ instead.

## Warning: Use of ‘data$qstart‘ is discouraged.
## i Use ‘qstart‘ instead.

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


bitscore
750
500
250
0
0
10000
20000
30000
40000
(data$pident * (data$qend − data$qstart))