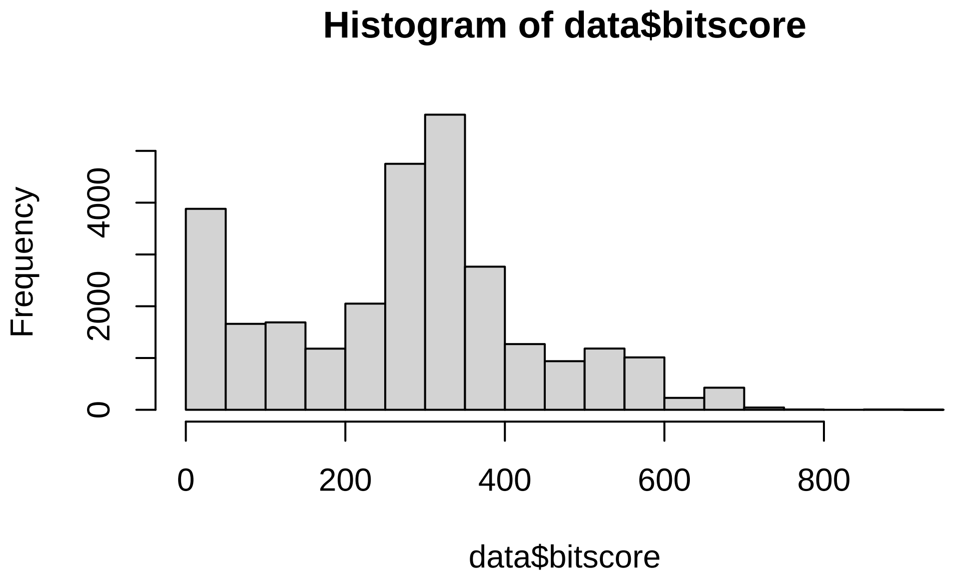
Class16 - R Graph

Richard Gao (PID: A16490010)

```
data <- read.table("~/Desktop/class16/mm-second.x.zebrafish.tsv")
colnames(data) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart
```

Make a histogram of the $bitscore values. You may want to set the optional breaks to be a larger number (e.g. breaks=30).

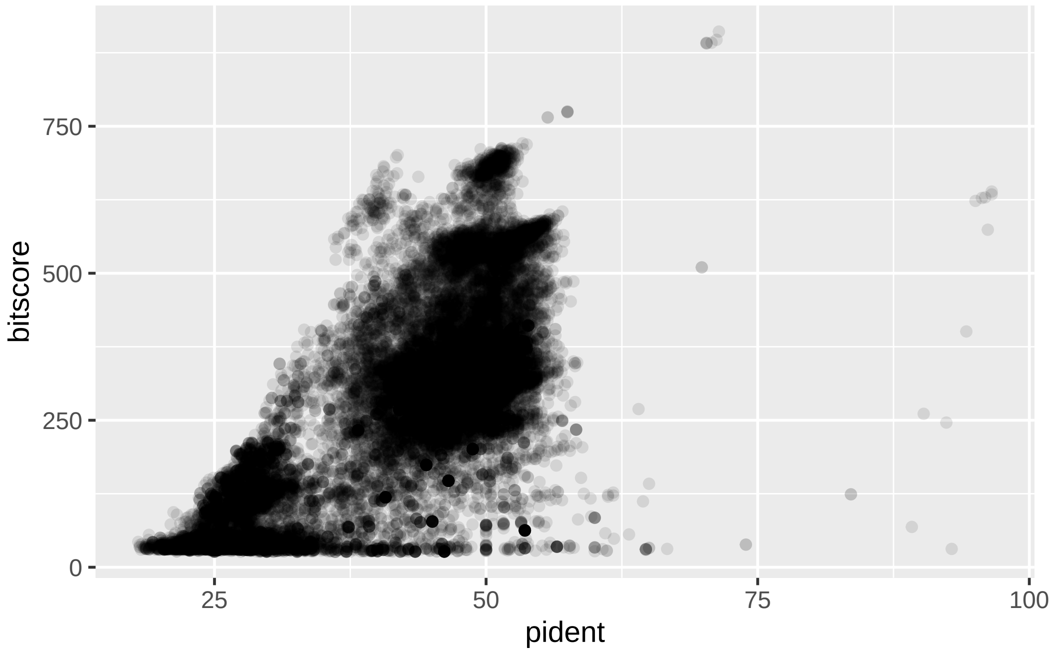
```
hist(data$bitscore, breaks=30)
```



> What do you notice here? Note that larger bitscores are better.

I notice that the histogram is bimodal and right-skewed, there are not many larger bitscores which makes sense since they are “better” so will probably be rarer.

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



> Is there a straightforward relationship between percent identity (*pident*)*andbitscore*(bitscore) for the alignments we generated?

No there seems to be a slight positive correlation but there must be another factor involved (length of alignment).

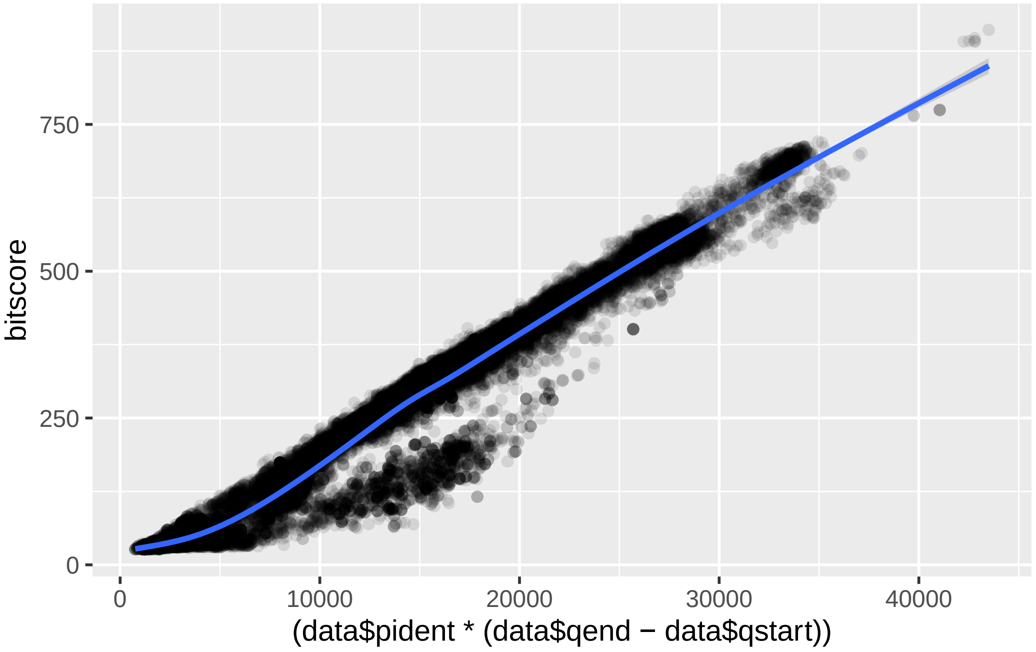
```
ggplot(data, aes((data$pident * (data$qend - data$qstart)), bitscore)) + geom_point(alpha=
```

```
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
(data$pident * (data$qend − data$qstart))