Class 17: BLAST on AWS

Nicolò (PID: 18109144)

10. Using RStudio online (or locally) to read your output

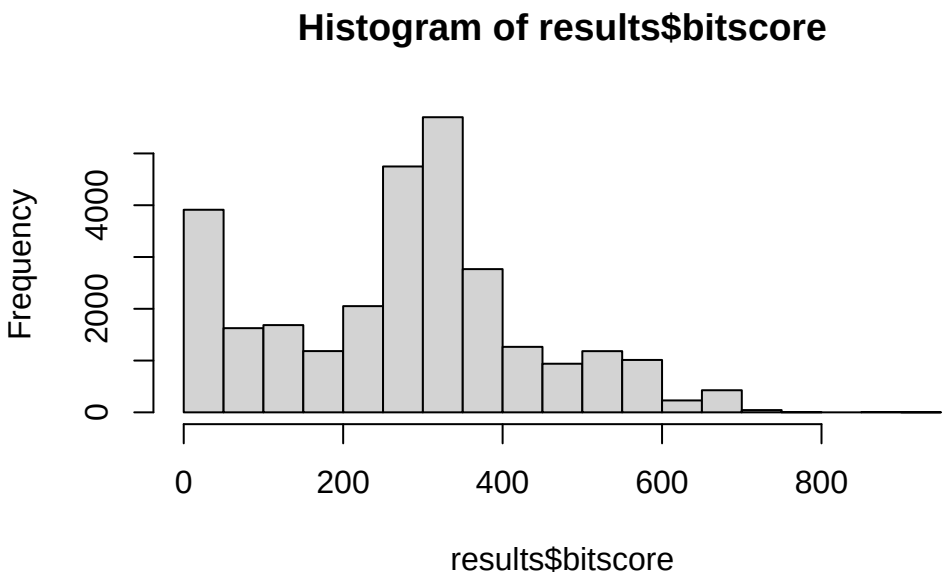
Read your mm-second.x.zebrafish.tsv. Set the colnames to be:

```
col_names <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "q
```

```
results <- read.delim("results.tsv", col.names = col_names)
```

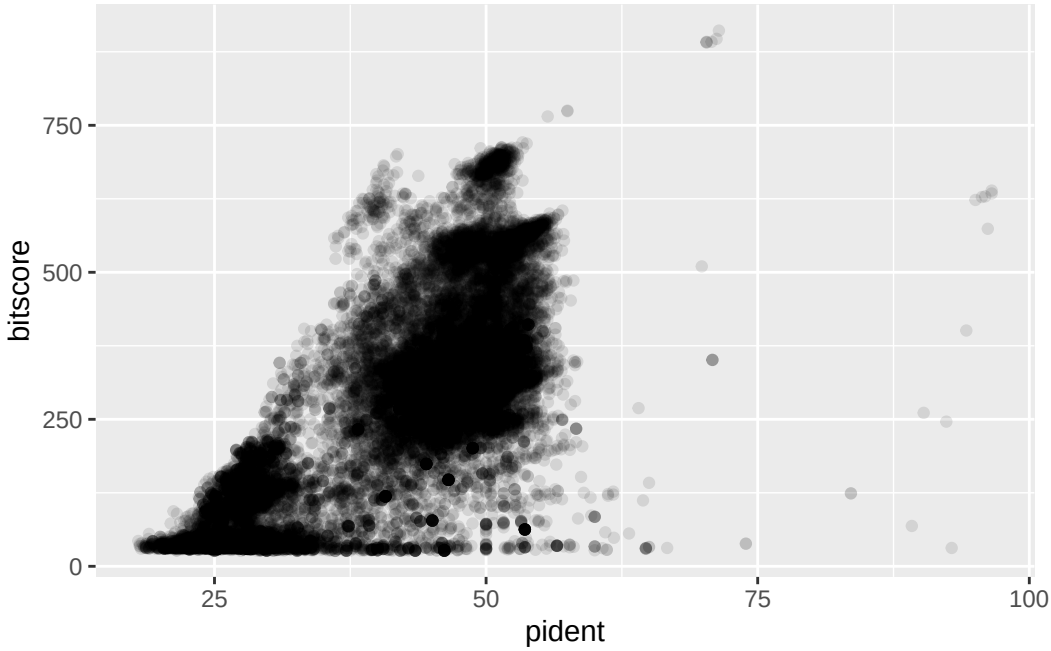
Make a histogram of the `$bitscore` values. You may want to set the optional breaks to be a larger number (e.g. `breaks=30`).

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



Is there a straightforward relationship between percent identity (`$pident`) and bitscore (`$bitscore`) for the alignments we generated?

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




```
ggplot(results, aes((results$pident * (results$qend - results$qstart)), bitscore)) + geom_
```

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

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

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

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

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

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

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