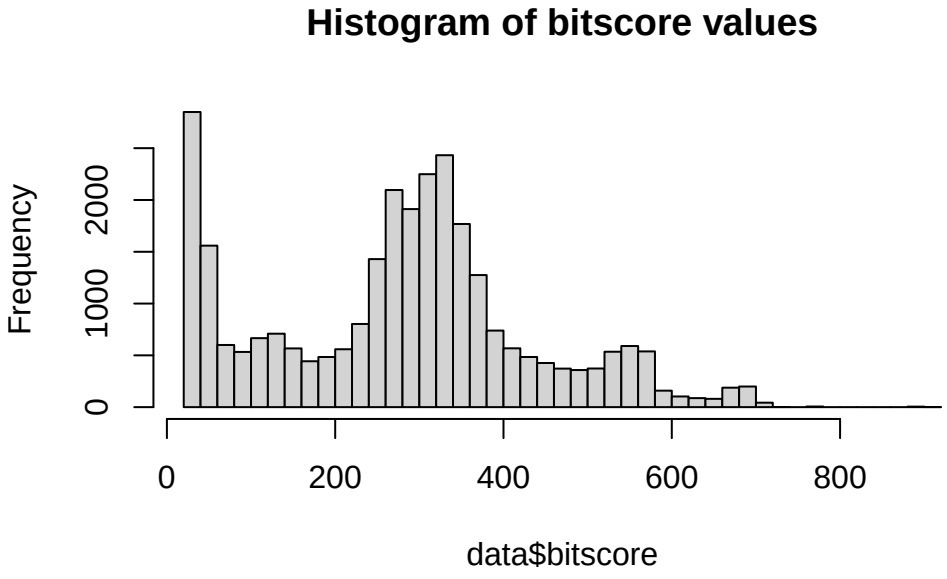
Class 16

Irene Hsieh

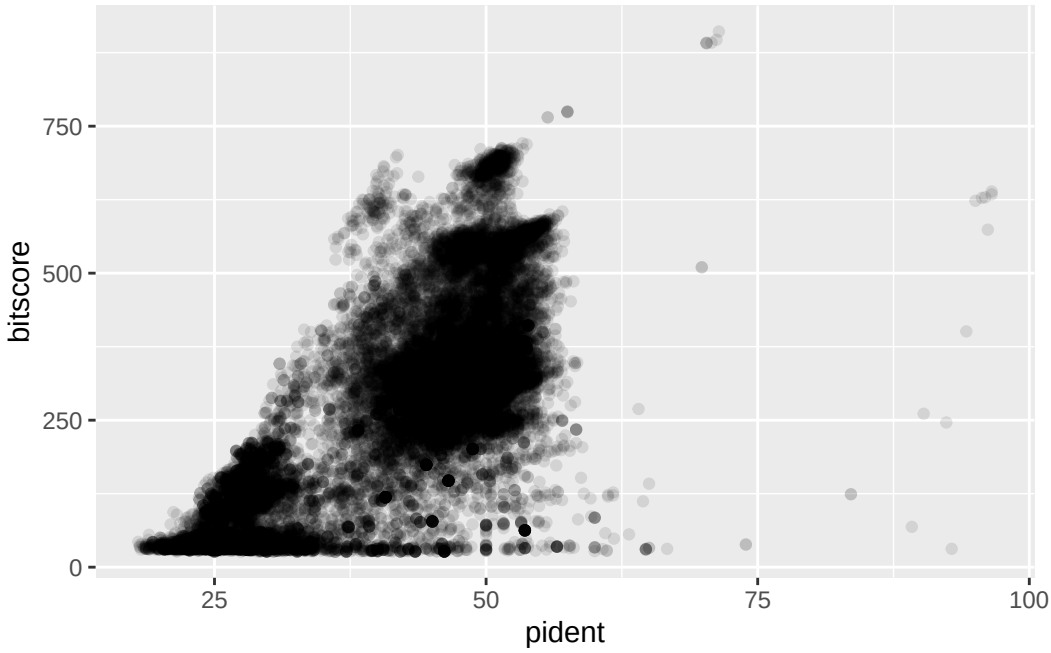
```
# Read the TSV file
data <- read.delim("mm-second.x.zebrafish.tsv", header = TRUE)

# Set column names
colnames(data) <- c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen",
"qstart", "qend", "sstart",
"send", "evalue", "bitscore")
```

```
# Make histogram of bitscore values
hist(data$bitscore, breaks = 50, main = "Histogram of bitscore values")
```



```
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_smooth()
```

```
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")'
```