Class 13

Tasnia Sharia (PID: A15931128)

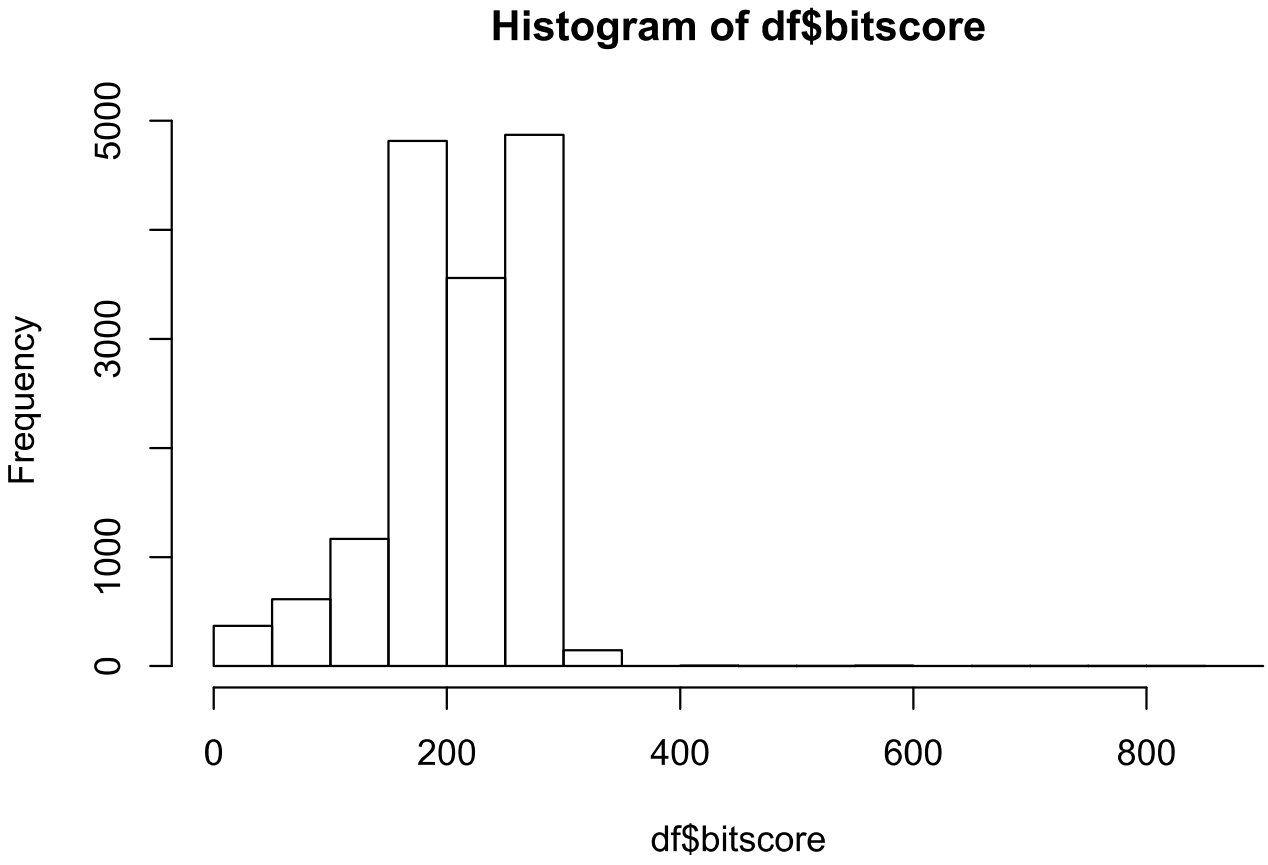
11/9/2021

```
df <- read.table("mm-second.x.zebrafish.tsv")
df
```

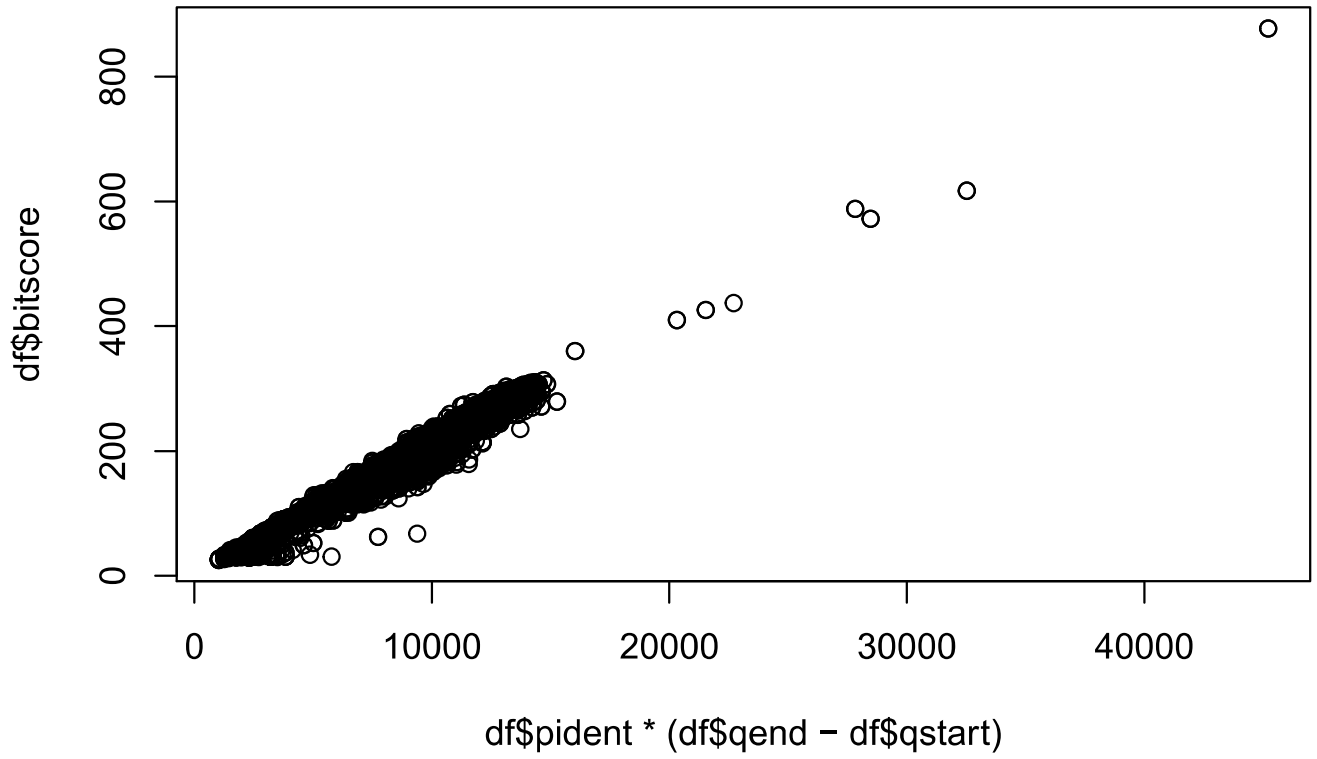
```
##                  V1             V2     V3  V4  V5 V6  V7  V8   V9  V10
## 1       YP_220550.1    NP_059331.1 69.010 313  97  0   4 316   10  322
## 2       YP_220551.1    NP_059332.1 44.509 346 188  3   1 344    1  344
## 3       YP_220551.1    NP_059341.1 24.540 163 112  3 112 263  231  393
## 4       YP_220551.1    NP_059340.1 26.804  97  65  2  98 188  200  296
## 5       YP_220552.1    NP_059333.1 88.132 514  61  0   1 514    1  514
## 6       YP_220552.1 XP_021326074.1 31.818  66  32  2 427 482   16   78
## 7       YP_220552.1 NP_001373511.1 31.818  66  32  2 427 482   48  110
## 8       YP_220553.1    NP_059334.1 70.925 227  66  0   1 227    1  227
## 9       YP_220553.1 NP_001038725.1 28.571  70  46  2  15  84  261  326
## 10      YP_220553.1 XP_005161478.1 28.571  70  46  2  15  84  261  326
## 11      YP_220553.1 XP_021324580.1 28.571  70  46  2  15  84  163  228
## 12      YP_220554.1 XP_002665692.3 25.862  58  42  1   5  61 2588 2645
## 13      YP_220554.1 XP_021324264.1 25.862  58  42  1   5  61 2278 2335
## 14      YP_220554.1 XP_009290033.1 50.000  30  13  1  26  55   42   69
## 15      YP_220554.1 XP_005156656.1 37.931  29  17  1  34  61  167  195
## 16      YP_220554.1 NP_001007391.1 37.931  29  17  1  34  61  167  195
## 17      YP_220555.1    NP_059336.1 52.212 226 107  1   1 225    1  226
## 18      YP_220556.1    NP_059337.1 78.161 261  57  0   1 261    1  261
## 19      YP_220556.1 XP_021330598.1 37.838  37  23  0   3  39   37   73
## 20      YP_220556.1 NP_001139036.1 37.838  37  23  0   3  39   37   73
## 21      YP_220556.1 NP_001070643.1 36.111  36  19  1 173 208  675  706
## 22      YP_220557.1    NP_059338.1 61.207 116  44  1   1 115    1  116
## 23      YP_220558.1    NP_059339.1 53.061  98  46  0   1  98    1   98
## 24      YP_220558.1 NP_001002376.1 57.895  19   8  0  20  38  136  154
## 25      YP_220558.1 XP_005158565.1 57.895  19   8  0  20  38  149  167
## 26      YP_220558.1    XP_706537.4 31.373  51  25  2  40  80   65  115
## 27      YP_220559.1    NP_059340.1 62.582 457 170  1   1 456    1  457
## 28      YP_220559.1    NP_059341.1 23.039 204 146  6 106 306  128  323
## 29      YP_220559.1    NP_059332.1 32.432 111  59  6 198 304   97  195
## 30      YP_220559.1 XP_009301289.1 24.074 108  65  4 169 276  148  238
## 31      YP_220560.1    NP_059341.1 55.126 595 257  8   7 597    9  597
## 32      YP_220560.1    NP_059340.1 23.077 351 246 11 107 442   84  425
## 33      YP_220560.1    NP_059332.1 24.171 211 145  6 226 433  105  303
## 34      YP_220560.1 XP_017210528.1 26.087  69  44  1 365 433  129  190
## 35      YP_220560.1 NP_001159387.1 26.087  69  44  1 365 433  142  203
## 36      YP_220560.1 XP_021330470.1 30.645  62  38  2 164 223  379  437
```

```
## 8293    105  368  4.41e-79    248.0
## 8294     37  340  6.89e-77    243.0
## 8295    124  398  1.78e-71    229.0
## 8296    223  399  9.04e-48    167.0
## 8297     57  121  1.00e-03     41.6
## 8298     74  362  4.47e-79    248.0
## 8299     70  326  4.67e-79    246.0
## 8300     21  267  3.15e-67    216.0
## 8301     21  246  5.42e-61    200.0
## 8302     20  216  4.92e-51    174.0
## 8303     82  330  4.76e-79    246.0
## 8304     82  304  1.73e-64    209.0
## 8305    131  330  2.77e-55    185.0
## 8306     81  211  1.42e-26    108.0
## 8307     80  185  8.54e-20     90.1
## 8308    235  330  6.42e-18     84.7
## 8309    289  330  4.40e-01     33.9
## 8310     82  330  4.76e-79    246.0
## 8311     82  304  1.73e-64    209.0
## 8312    131  330  2.77e-55    185.0
## 8313     81  211  1.42e-26    108.0
## 8314     80  185  8.54e-20     90.1
## 8315    235  330  6.42e-18     84.7
## 8316    289  330  4.40e-01     33.9
## 8317     71  324  5.31e-79    246.0
## 8318     94  349  8.44e-69    220.0
## 8319     70  237  8.06e-50    171.0
## 8320     82  332  6.24e-79    246.0
## 8321     82  307  1.24e-68    219.0
## 8322    123  332  6.59e-56    186.0
## 8323     81  271  6.95e-50    171.0
## 8324     79  220  1.78e-39    144.0
## 8325    245  333  1.10e-13     72.4
## 8326     75  374  8.12e-79    247.0
## 8327     37  311  8.43e-79    245.0
## 8328     22  274  2.75e-72    228.0
## 8329     78  305  3.46e-53    179.0
## 8330     26   79  2.40e-11     65.5
## 8331     37  304  8.79e-79    246.0
## 8332     22  276  1.95e-74    234.0
## 8333    149  347  7.80e-46    160.0
##  [ reached 'max' / getOption("max.print") -- omitted 7221 rows ]
```

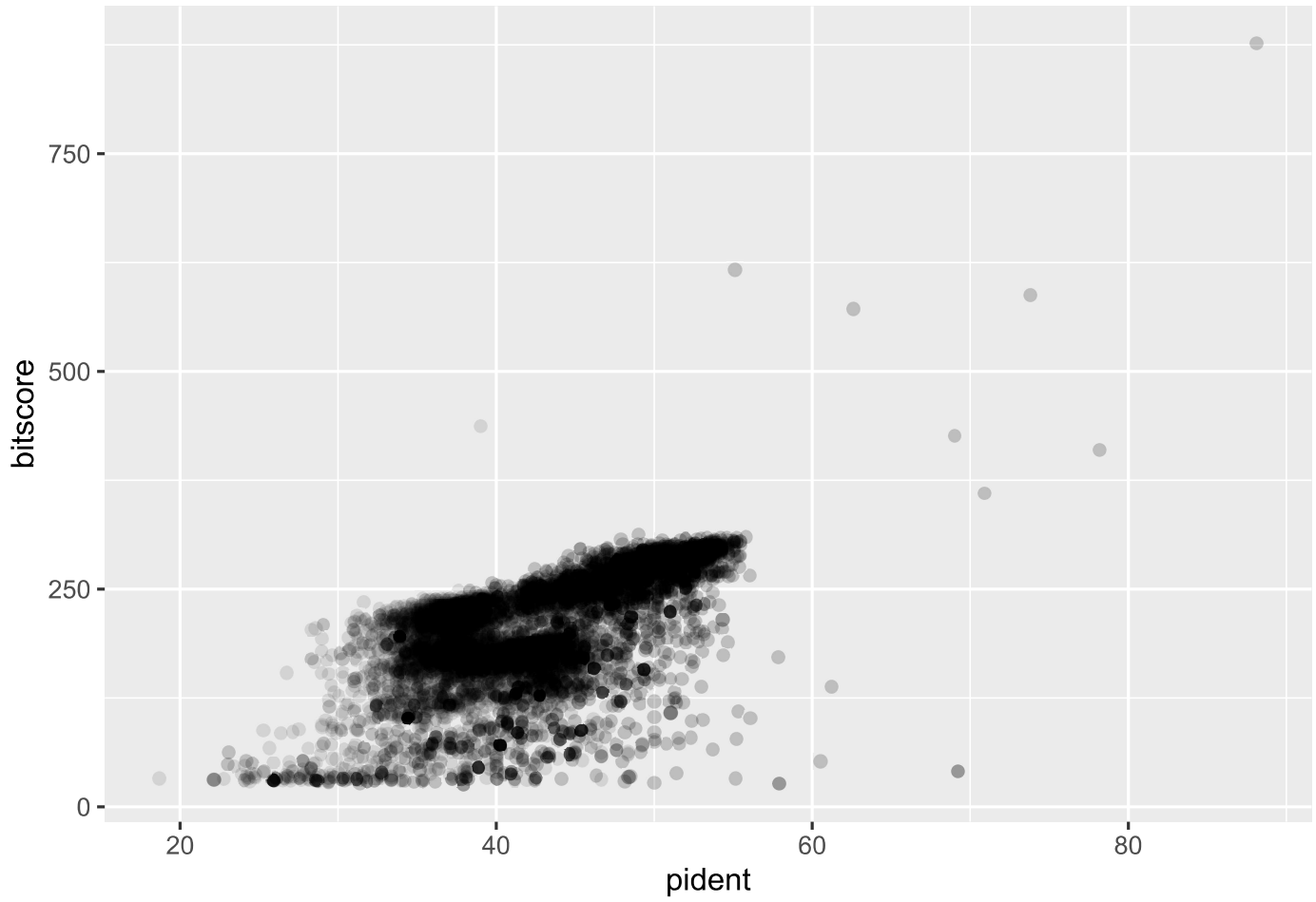
```
#producing a histrogram
hist(df$bitscore, breaks=30)
```




```
#seeing if there is a relationship between percent identity and bitscore
plot(df$pident  * (df$qend - df$qstart), df$bitscore)
```

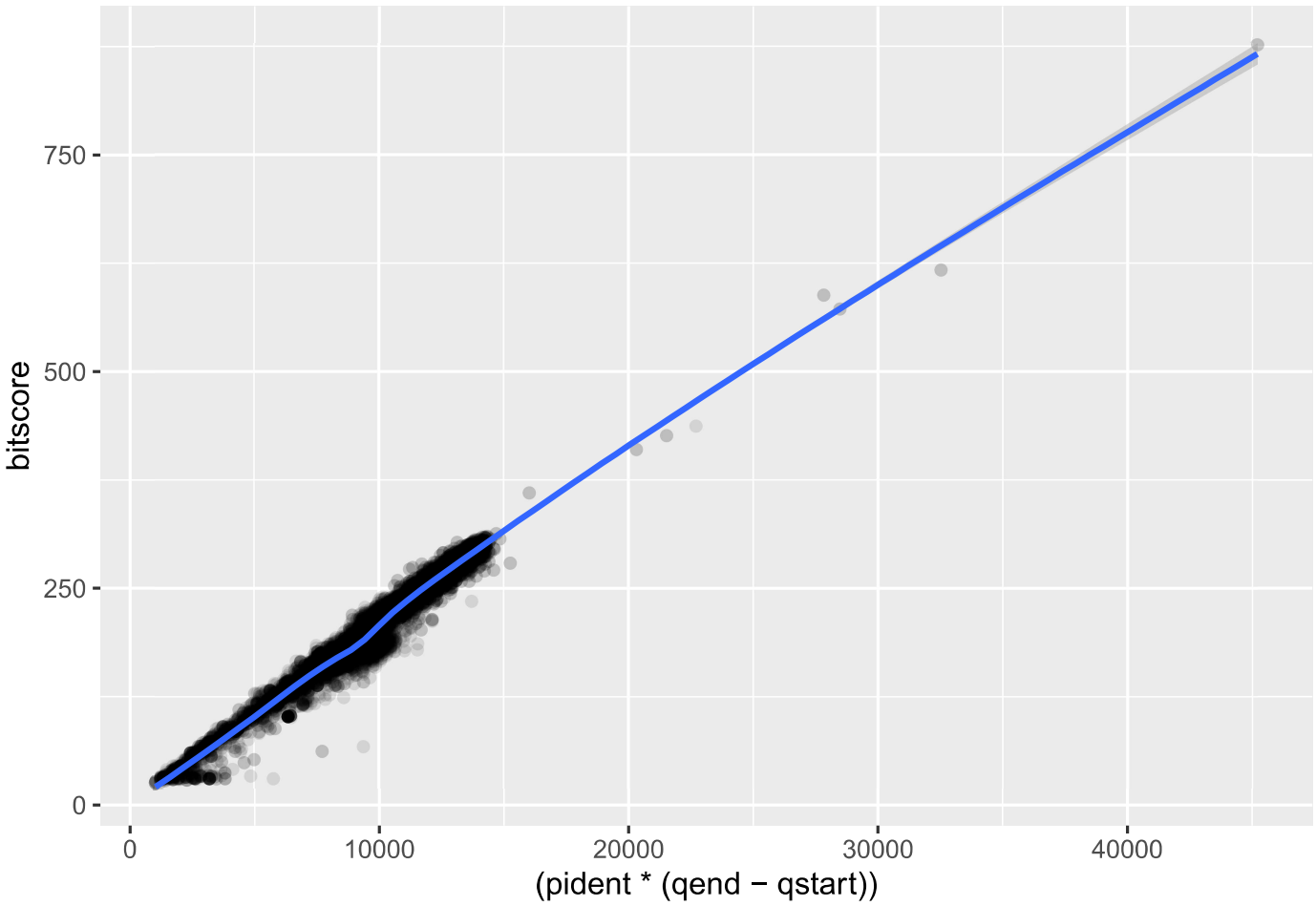



```
#had to install ggplot in AWS
library(ggplot2)
# producing a nicer plot using ggplot
ggplot(df, aes(pident, bitscore)) + geom_point(alpha=0.1)
```




```
ggplot(df, aes((pident * (qend - qstart)), bitscore)) + geom_point(alpha=0.1) + geom_smooth
```

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




```
#had to install tinytex package for AWS
library(tinytex)
```