

class13

Kai Movellan

11/9/2021

#do in terminal

```
#file.txt my_text_file.txt
#zico:class13>rm file.txt
#zico:class14> ls
#my_text_file.txt
#zico:class13>rm file.txt
#zico:class14> ls
#my_text_file.txt
#zico:class13> less my_text_file.txt
#zico:class13> cp my_text)file.txt another_file.txt
#zico:class13> ls
#my_text_file another_file
#zico:class13> mv sanother_file.txt tmp
#zico:class13> ls

#this is just a test editor and not a very fancy one.
#(END)
```

#Read TSV blast result file

```
blast<-read.delim("mm-second.x.zebrafish.tsv")
head(blast)

##    YP_220550.1      NP_059331.1  X69.010  X313  X97  X0   X4  X316  X10  X322  X1.24e.150
## 1 YP_220551.1      NP_059332.1  44.509   346  188   3   1  344   1  344  8.62e-92
## 2 YP_220551.1      NP_059341.1  24.540   163  112   3  112  263  231  393  5.14e-06
## 3 YP_220551.1      NP_059340.1  26.804   97   65   2   98  188  200  296  1.00e-01
## 4 YP_220552.1      NP_059333.1  88.132   514   61   0   1  514   1  514  0.00e+00
## 5 YP_220552.1  XP_021326074.1  31.818   66   32   2  427  482   16   78  6.70e+00
## 6 YP_220552.1  NP_001373511.1  31.818   66   32   2  427  482   48  110  7.50e+00
##          X426
## 1  279.0
## 2  49.7
## 3  35.8
## 4 877.0
## 5  29.3
## 6  29.6
```

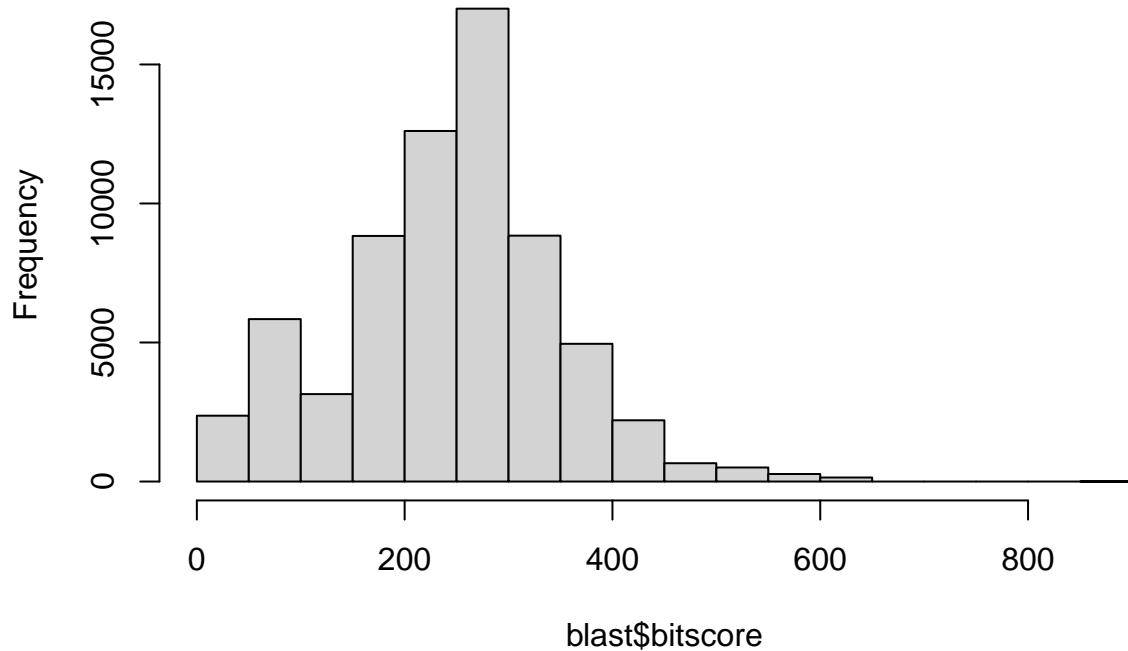
Set the colnames to be:

```
colnames(blast)<-c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "ss  
head(blast)

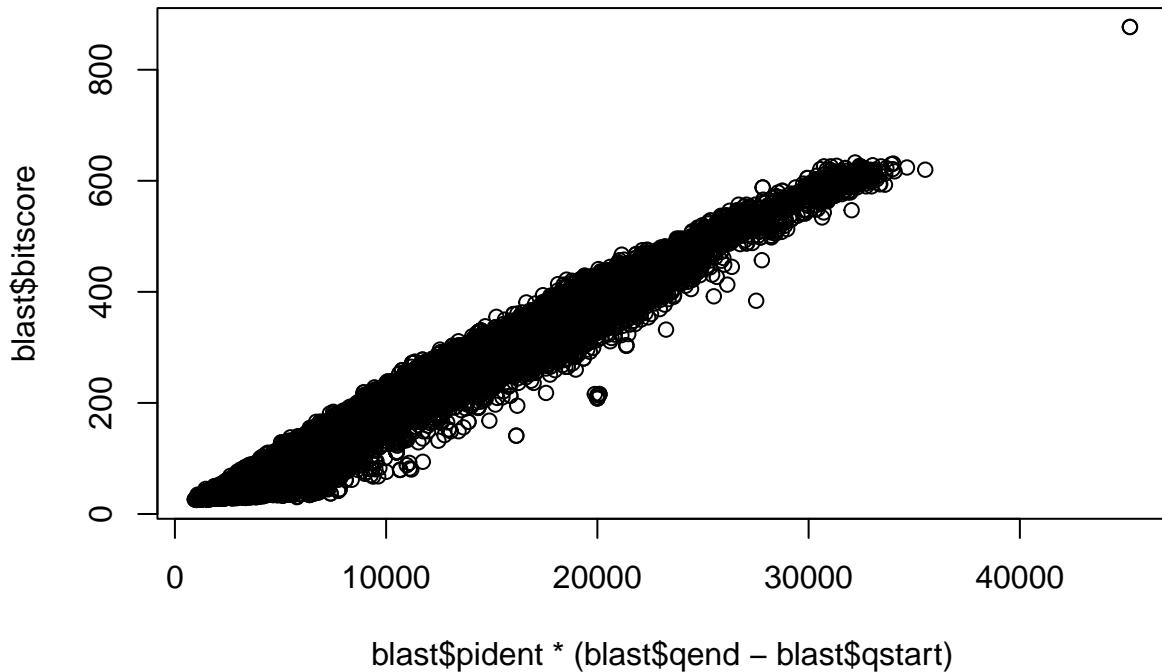
##      qseqid      sseqid pident length mismatch gapopen qstart qend sstart
## 1 YP_220551.1 NP_059332.1 44.509    346     188       3     1 344     1
## 2 YP_220551.1 NP_059341.1 24.540    163     112       3   112 263   231
## 3 YP_220551.1 NP_059340.1 26.804     97      65       2    98 188   200
## 4 YP_220552.1 NP_059333.1 88.132    514      61       0     1 514     1
## 5 YP_220552.1 XP_021326074.1 31.818     66      32       2   427 482    16
## 6 YP_220552.1 NP_001373511.1 31.818     66      32       2   427 482    48
##      send      evalue bitscore
## 1 344 8.62e-92    279.0
## 2 393 5.14e-06    49.7
## 3 296 1.00e-01    35.8
## 4 514 0.00e+00   877.0
## 5  78 6.70e+00    29.3
## 6 110 7.50e+00    29.6

hist(blast$bitscore)
```

Histogram of blast\$bitscore

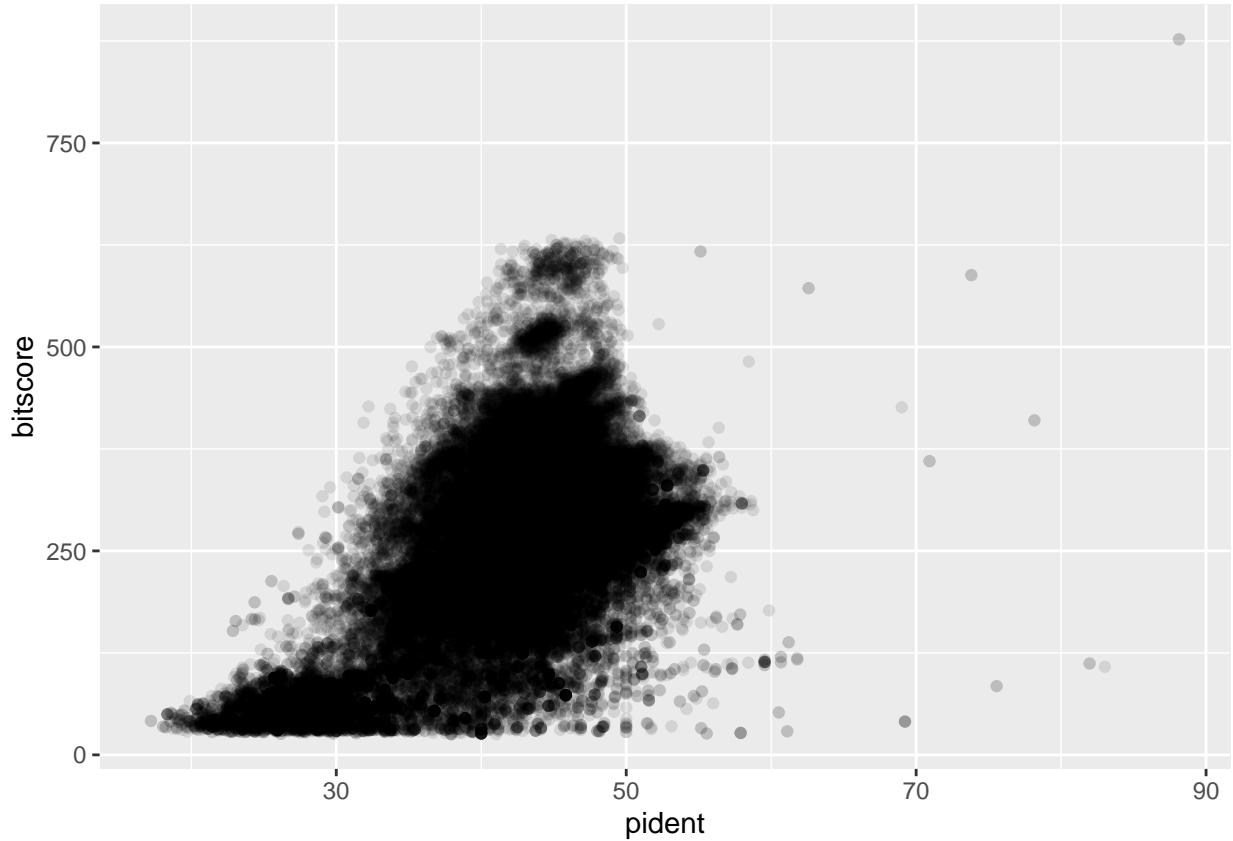


```
## Assuming your blast results are stored in an object called 'blast'
plot(blast$pident * (blast$qend - blast$qstart), blast$bitscore)
```



```
#Use ggplot for a nicer figure
```

```
#install.packages("ggplot2")
library(ggplot2)
ggplot(blast, aes(pident, bitscore)) + geom_point(alpha=0.1)
```



```
ggplot(blast, aes((pident * (qend - qstart)), bitscore)) + geom_point(alpha=0.1) + geom_smooth()  
## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

