

Unix command review homework questions (Week 8)

Instructions: Type your answers to the following **3 questions** in the boxed blue form space provided. Save your answers and upload to our GradeScope site. There are 10 total points on offer. You can use any resources but please no communication (electronic or otherwise) with your fellow students regarding these questions. Thank you!

Q1. [6pt] List the UNIX bash shell commands to:

- open a **secure shell** on a remote machine:

- make a **new folder** in your home area called "test":

- **download** this file "<https://files.rcsb.org/download/5P21.pdb.gz>":

- **unzip/decompress** it the file:

- **print to screen** the first 6 lines:

- print to lines beginning with ATOM to a **new file** called "coords.pdb":

Q2. [3pt] List the UNIX commands to **copy securely** the file "myaln.fa" in your current working directory to your home area on the remote machine "**biglabcluster.ucsd.edu**":

	<u>File</u>
Key	User@SupercomputerHost:RemoteDirectory

Q3. [1pt] The alignment file “***myaln.fa***” is not in your current working directory but it is in your “**Downloads**” directory. Write the R code to import this alignment to the named object “***aln***” using a function from the bio3d package.

Q30. [5pt] The

class16_Homework_R_ggplot

Gabriella Tanoto (A18024184)

Table of contents

Loading Packages & Transfer File to R	1
Making Histogram of BitScore	1
Answering Questions	6

Loading Packages & Transfer File to R

Load up ggplot2 for graphing.

```
library(ggplot2)
```

Read my .tsv file, and change its column names.

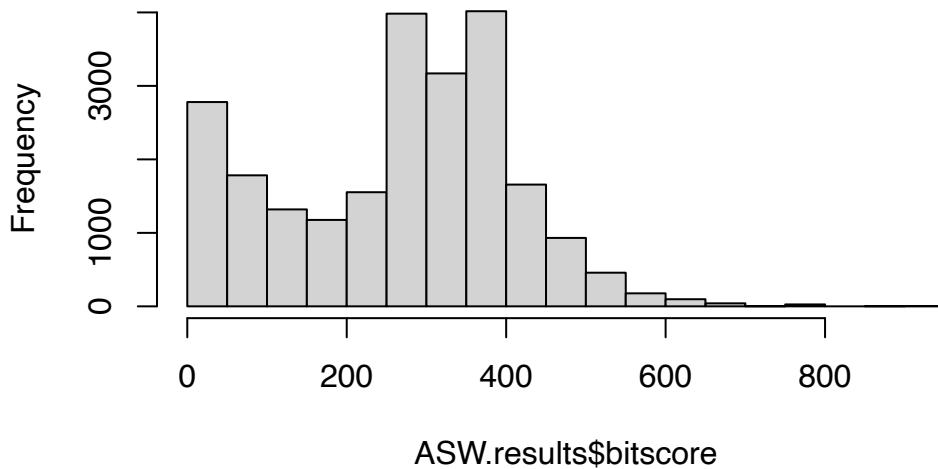
```
ASW.results <- read.csv("results_AWSclass16.tsv", sep = "\t")
colnames(ASW.results) <-c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qs
```

Making Histogram of BitScore

You may want to set the optional breaks to be a larger number (e.g. breaks=30).

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

Histogram of ASW.results\$bitscore

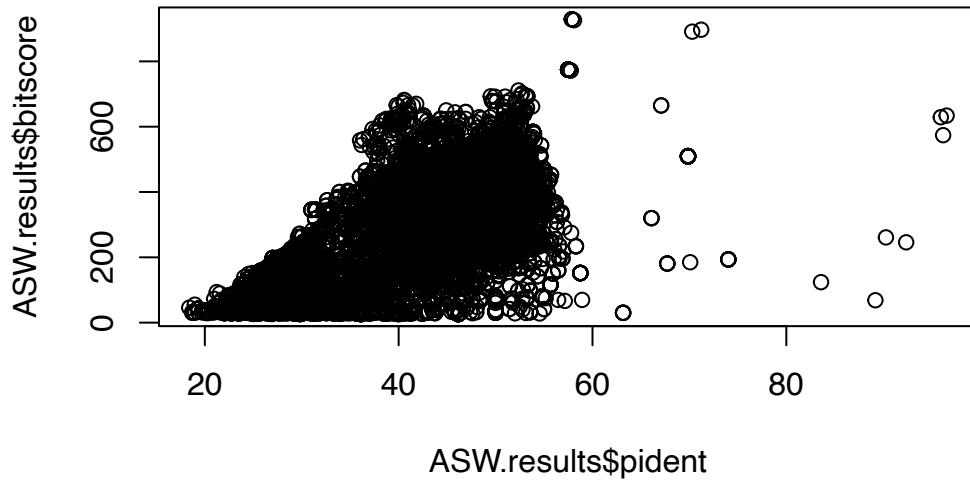


What do you notice here? Note that larger bitscores are better.

From the histogram, we can see that the bitscore is sort of normally distributed, but there are also many that is on the lower side. Only a few alignment results have higher bitscore than 500.

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

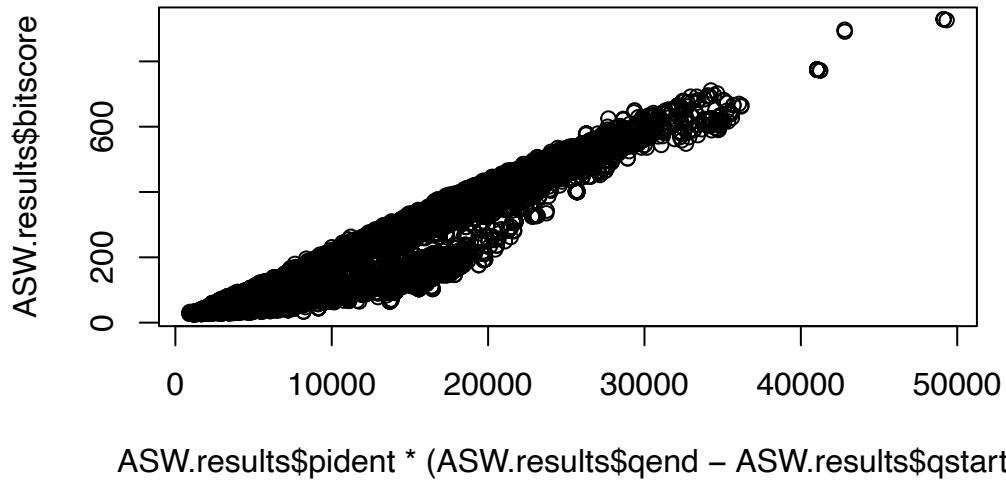
```
plot(ASW.results$pident, ASW.results$bitscore)
```



From the scatterplot above, we can see that it's somehow related, but not much.

However, if we take into consideration the length of the sequence ($\$qend - \$qstart$), we get a much nicer relationship:

```
plot(ASW.results$pident * (ASW.results$qend - ASW.results$qstart), ASW.results$bitscore)
```

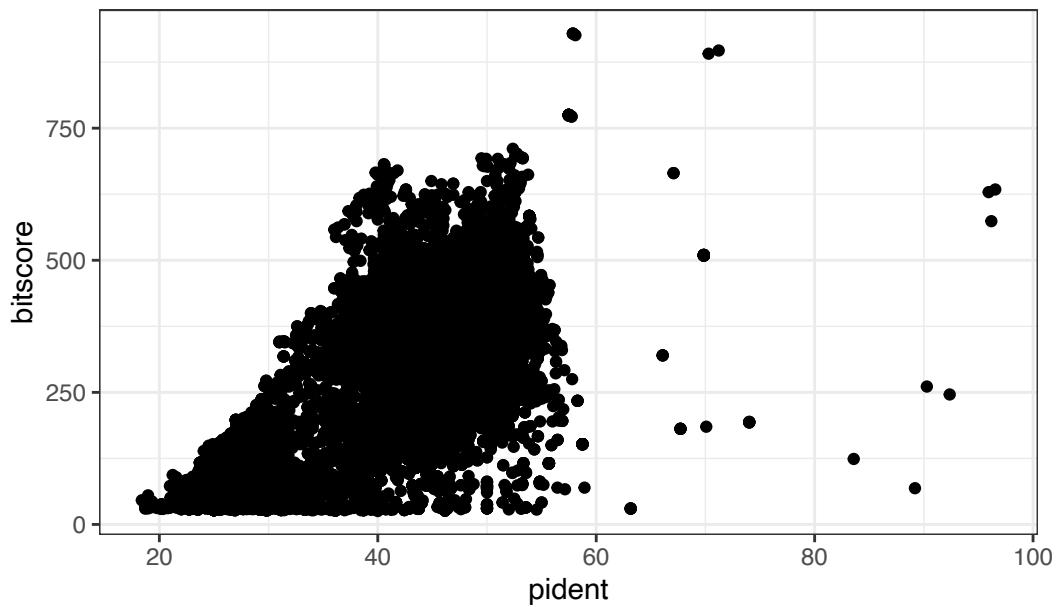


This graph suggests that we will get a higher bitscore as we have a longer match and a better percent identity.

These graphs can also be made in ggplot as follows:

```
ggplot(ASW.results) +
  aes (x = pident, y = bitscore) +
  geom_point() +
  ggtitle("Scatterplot of Percent Identity v. BitScore") +
  theme_bw()
```

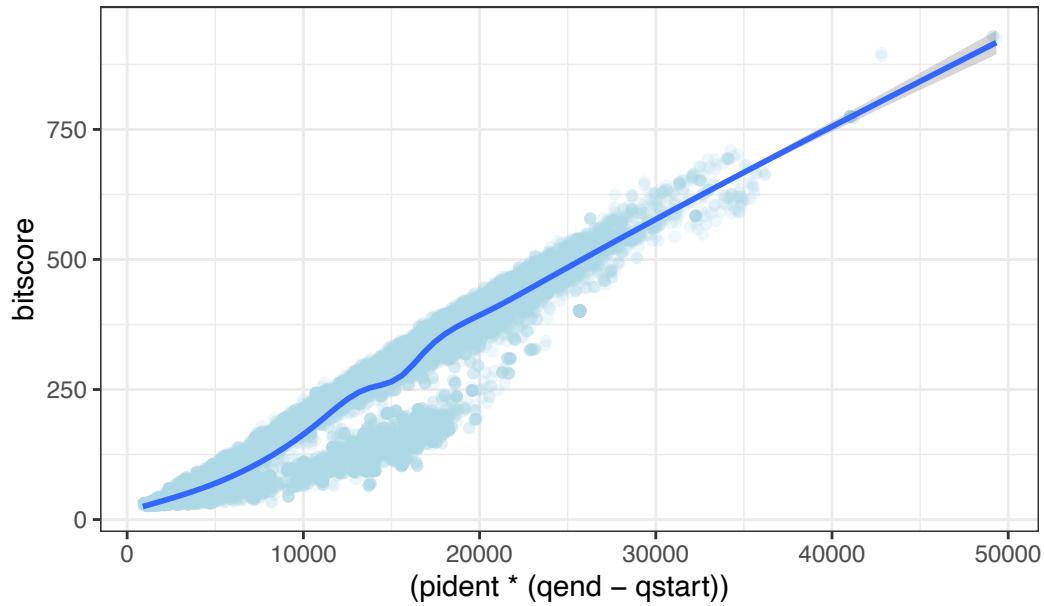
Scatterplot of Percent Identity v. BitScore



Then, when we include also the length of the sequence that matches:

```
ggplot(ASW.results) +  
  aes(x = (pident * (qend - qstart)),  
      y = bitscore) +  
  geom_point(alpha = 0.2, color = "lightblue") +  
  geom_smooth() +  
  ggtitle("Scatterplot of Percent Identity*Length v. BitScore") +  
  theme_bw()  
  
`geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```

Scatterplot of Percent Identity*Length v. BitScore



Answering Questions

If we were doing this graphing in the supercomputer AWS server, we would have to transfer it to our local laptop using the codes below (into our laptop's terminal):

```
scp -i ~/Downloads/barry_bioinf.pem -r ubuntu@YOUR_IP_ADDRESS:~/work/* .
```

Q. Note the addition of the -r option here: What is its purpose? Also what about the *, what is its purpose here?

The -r addition stands for recursively copying the directory (/work/), so we can make sure everything within that directory is copied into our personal computers. The * will copy ALL of the things we have in that directory, both the subdirectories and files. It basically copies everything that has the indicated absolute path (i.e., ~/work/).