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:

```
ssh -i "key.pem" username@address
```

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

```
mkdir ~/test
```

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

```
curl -O https://files.rcsb.org/download/5P21.pdb.gz
```

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

```
gzip -d 5P21.pdb.gz
```

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

```
head -6 5P21.pdb
```

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

```
grep '^ATOM' 5P21.pdb > 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**":

```
scp -i "key.pem" ~/myaln.fa username@biglabcluster.ucsd.edu:~/
```

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.

```
library(bio3d)
aln <- read.fasta("~/Downloads/myaln.fa")
```

Q30. [5pt] The

Class 17: BLAST on AWS

Nicolò (PID: 18109144)

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

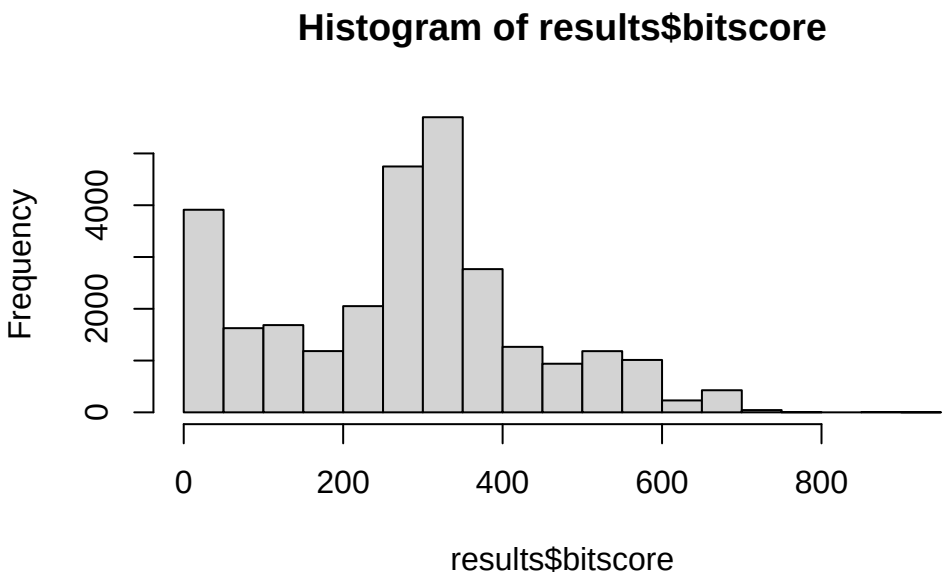
Read your mm-second.x.zebrafish.tsv. Set the colnames to be:

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

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

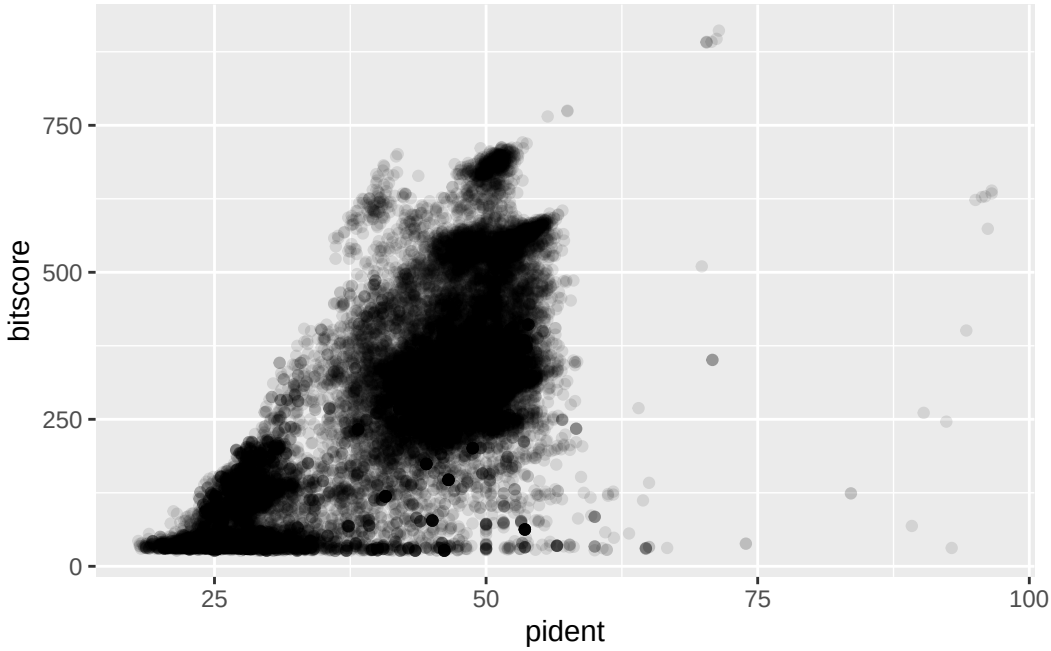
Make a histogram of the `$bitscore` values. You may want to set the optional breaks to be a larger number (e.g. `breaks=30`).

```
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")'
```