Odds and Ends

GZIP

Compression of files with gzip

```
$ gzip file.fa
# will produce
file.fa.gz
```

To uncompress

```
$ gunzip file.fa.gz
# will produce
file.fa
```

Searching for text with grep

Powerful pattern seaching with `grep`

Simple search for a text string:

```
$ grep Chr11 /bigdata/gen220/shared/data-examples/examples/random_exons.csv
Chr11,14656670,14656778
Chr11,3528895,3530426
Chr11,16238576,16239304
```

To get the count of number of lines that match a pattern use the `-c` option.

```
$ grep -c Chr11 /bigdata/gen220/shared/data-examples/examples/random_exons.csv
3
```

What if we wanted to count the number of times Chr1 showed up?

```
$ grep Chr1 /bigdata/gen220/shared/data-examples/examples/random_exons.csv
Chr11,14656670,14656778
Chr1,1147485,1147562
Chr12,22130532,22130707
Chr10,19029658,19029760
Chr11,3528895,3530426
Chr12,23125462,23125634
Chr1,4249358,4249468
Chr11,16238576,16239304
Chr12,9264478,9264617
Chr1,18658403,18658693
Chr12,9488597,9489239
```

```
Chr1,12152,12435
Chr1,43214981,43215253
```

How can we make this a more specific query? Well we know the ',' comes after
so we can include that in the search.

```
$ grep Chr1, /bigdata/gen220/shared/data-examples/examples/random_exons.csv
Chr1,1147485,1147562
Chr1,4249358,4249468
Chr1,18658403,18658693
Chr1,12152,12435
Chr1,43214981,43215253
```

If you want to invert the search and find lines that DO NOT match the pattern
use the -v option.

```
$ grep -c Chr1, /bigdata/gen220/shared/data-examples/examples/random_exons.csv
5
$ grep -v -c Chr1, /bigdata/gen220/shared/data-examples/examples/random_exons.csv
25
```