Exercise 2 - Pipes, filters, and finding things with The Unix Shell

Work with others to answer the questions below and each of you return your answers at the beginning of lecture on Monday.

1. What is the difference between `>` and `>>` in Unix? What is the difference between `*` and `?` in Unix?

2. Assuming we start with the file `data-shell/data/planets.txt`, what will be the content of the file *fiveLittlePlanets.txt* after running the code below?

```
cat planets.txt | tr -d '"' | sed 's/ /_/g' | sed 's/,/ /g' | sort -k 2 -n | head -n 6 >
fiveLittlePlanets.txt
```

Why is each filter included in this line of code?

3. Provide the Unix code required to return the first five lines containing the word 'beetle' from a file *insectSightings.txt*. Give us two versions of code. One using multiple Unix commands and no flags when using `grep`, and a second version that only uses `grep`.

4. Your advisor has given you a text file (*dataV5_final.txt*) containing data from a large experiment run by a former graduate student in the lab. Unfortunately, it has a number of typos, mixture of capitalizations used for site and treatment names, and other errors. All your advisor can remember is that there were 6 sites (column 1), 4 treatments (column 2), and ten replicates of each site-treatment combination (column 3). Use your new found Unix skills to cleanup this digital mess! On the back of this page describe your strategy for finding errors and write the code necessary to produce a clean text file below.