**Bootcamp notes**

**Day 1**

* \*.bed files
  + One line per gene
* \*.vcf
  + One line per SNP/variant
* With awk
  + NR == FNR checks whether reading first file
    - NR = number of input records
    - FNR = number of input records in current file
  + NF is number of fields! Useful for iterating through every field (column)
  + awk '{for (i=1; i <=NF; i++) {if($i=="gene\_type") print $(i+1)}}'



**Awk**

* $0 = entire row
* /^#/{next} 🡪 ignore header
* {} 🡪 encases commands
  + Put if in brackets too
* ^ 🡪 line startswith

**Day 2**

**Run sublime from terminal:**

$ /Applications/Sublime\ Text.app/Contents/SharedSupport/bin/subl

Add to bash\_profile: export PATH=/bin:/sbin:/usr/bin:/usr/local/sbin:/usr/local/bin:$PATH

$ source ~/.bash\_profile

$ ln -s /Applications/Sublime\ Text.app/Contents/SharedSupport/bin/subl /usr/local/bin/subl

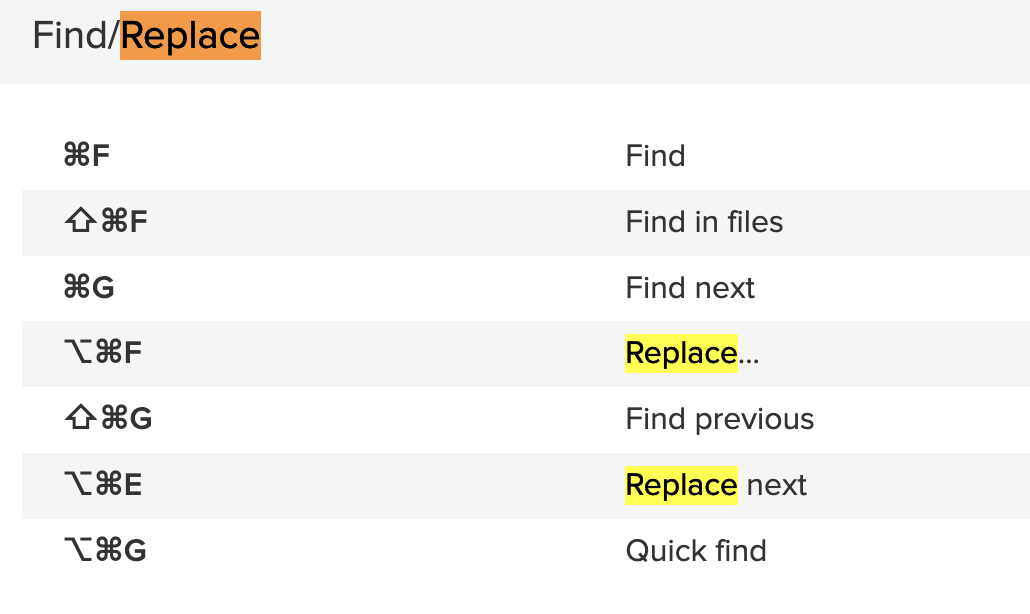
Now use subl to open

**Running scripts as programs**

* chmod a+x script.py
* Also add #!/usr/bin/env python3 to beginning of script
* Now can run as standalone script

**Truth statements in python**

* You can actually do 1 < 2 < 3 now! :O

****

**Day 3**

* Find max of one column and report the value of that max as well as the corresponding value from the same row of another column
  + awk -v max=-100 '{if($3>max){max=$3;largest=$1}} END {print largest,max}' plink.eigenvec
* conda environment stuff  
  $ conda create -n bcftools -y python=3 bcftools  
  $ conda activate bcftools  
  $ conda deactivate
* bcftools powerful tool for giant vcf files
* use gzcat bash command to take a look at a binary file without unzipping it

**Day 4**

* genfromtxt will automatically give column names, f0, f1, f2 …
* if you load into data, you can get this list by doing data.dtype.names
* conda info --env
* conda install something==1.8 # this specifies version
* viridis seaborn color palette has been checked for like colorblindness.

Poisson distribution

* for mean > 5, well approximated by normal distribution