

Unix Commands

<u>Unix Commands</u>	
wc	Count the number of words or lines
grep	Find the lines that contain the keyword
head, tail	Print the first few and last few lines
cat	Print the entire content of the file to the screen
awk	Process text files one line at a time Can scale up to massive datasets
program1 &	Run programs in parallel
<u>Flags and arguments - modifies the behavior of the files</u>	
sort	Sort the content alphabetically
sort -n	Sort the content numerically
sort -k 1,2,3	Sort the content by column 1, 2 then 3
<u>Wildcards</u>	
Cat	Taking multiple files at one time cat test1.txt test2.txt

Pipes

- Pipe is buffered (memory required to process the data is limited)
- Output one program to be used as the input of the second program (second command)
- it allows us to scale up large files
- each program in the sub-pipe only generates data for the command in the next stage of the pipeline

Redirects

- saves the results of the filtered command by using the symbol > with a filename

R commands

<u>Summary Statistics</u>	
min(data\$column) mean(data\$column)	Finding the minimum and average values
sd(data\$column)	Finding the standard deviation
summary(data\$column)	Finding the medians and quartiles (1 st quartile, 3 rd quartile, median, mean, max)
<u>Visualizing data</u>	
hist(x, data)	Plotting histogram
boxplot(x, data)	Plotting boxplot
plot(x, data)	Plotting scatter plots