STATS 769
Prospering in Linux

Paul Murrell

The University of Auckland

July 20, 2021

Overview

- In the last lecture we learned how to **survive** in the linux shell (how to do things we already do in other operating system environments)
- In this lecture we will learn how to **prosper** in the shell (how to work better/smarter than we do in other operating system environments)

- NOTE that we are working on the filesystem rather than in RAM, so the consequences are more permanent (compared to running R code).

Globs and command substitution

- Filenames can contain "wildcards" that specify patterns.
- `*` matches any number of characters.
- `?` matches any one character.

- Anything within `$()` is executed first and the result is used

Escape sequences

- Space is important in shell commands
  (it is used between arguments)
- Use backslash (\) to escape a space
  (or a wildcard).
- Use double quotes (") to escape spaces and wildcards
  (but keep command substitution)
- Use single quotes (') to escape everything.

Standard Input and Standard Output

- If a program produces output, we typically see that output on the screen.
- If a program produces output, what actually happens is that it sends the output to "standard output" and that is **by default** connected to the screen.

- If a program requires input, we typically type that input via the keyboard.
- If a program requires input, what acually happens is that is reads input from "standard input" and that is **by default** connected to the keyboard.

Redirection

- > redirects output from a program to a file
  (for programs that normally print output to the screen)
- >> redirects output from a program **and appends it** to a file.
- < redirects input to a program from a file
  (for programs that normally accept input from the keyboard)
- | "pipes" output from one program to another program.

Loops

```
for variable in list
do
    cmd $variable
done
```

- `list` is a space-separated list of one or more values (typically file names).
- `variable` is a shell variable that gets each value in `list`, one at a time.
- `cmd` is called for each value in `list`.

GNU tools

Every Linux distribution comes with a basic set of useful programs, including ...

`wc`	Count lines, words, and characters in text files.
`grep`	(Regular expression) text search (across multiple files).
`awk`	Text processor.

AWK

- awk implicitly loops over each line in a file and breaks each line into fields
- predefined variables for each field in the line
  - `$0` is the whole line, `$1` is field 1, `$2` field 2, ...
  - `NR` is current row, `NF` is num fields
- an awk program is a series of ...

```
condition { action }
```

- conditions are `A == B`, or `/regular expression/`, or `BEGIN`, or `END`
- most common action is `print()`

Resources

- The GNU Awk User's Guide
  `https://www.gnu.org/software/gawk/manual/gawk.html`
- The Unix Shell (Sections 4, 5, 6 and 7)
  `http://swcarpentry.github.io/shell-novice/`