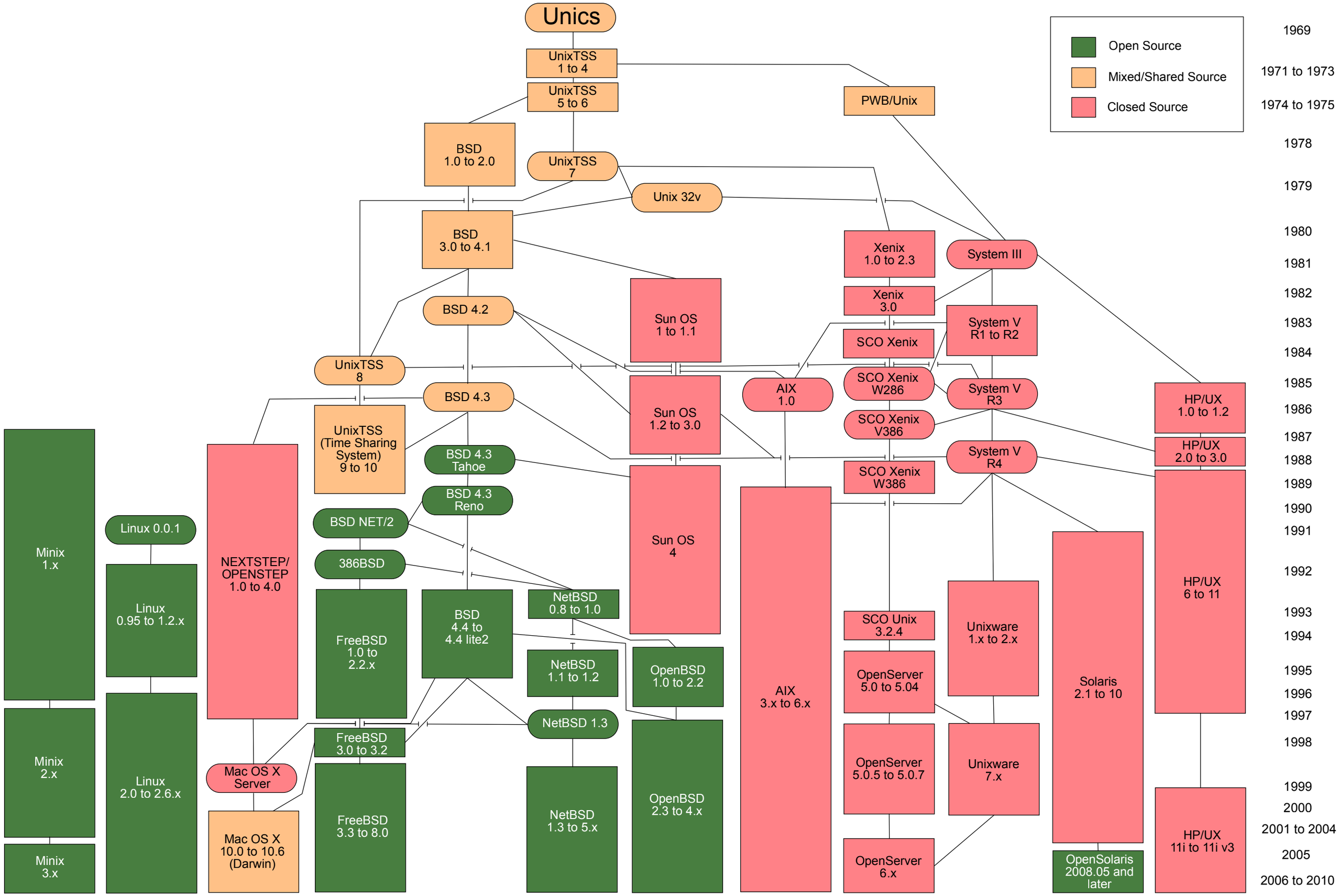
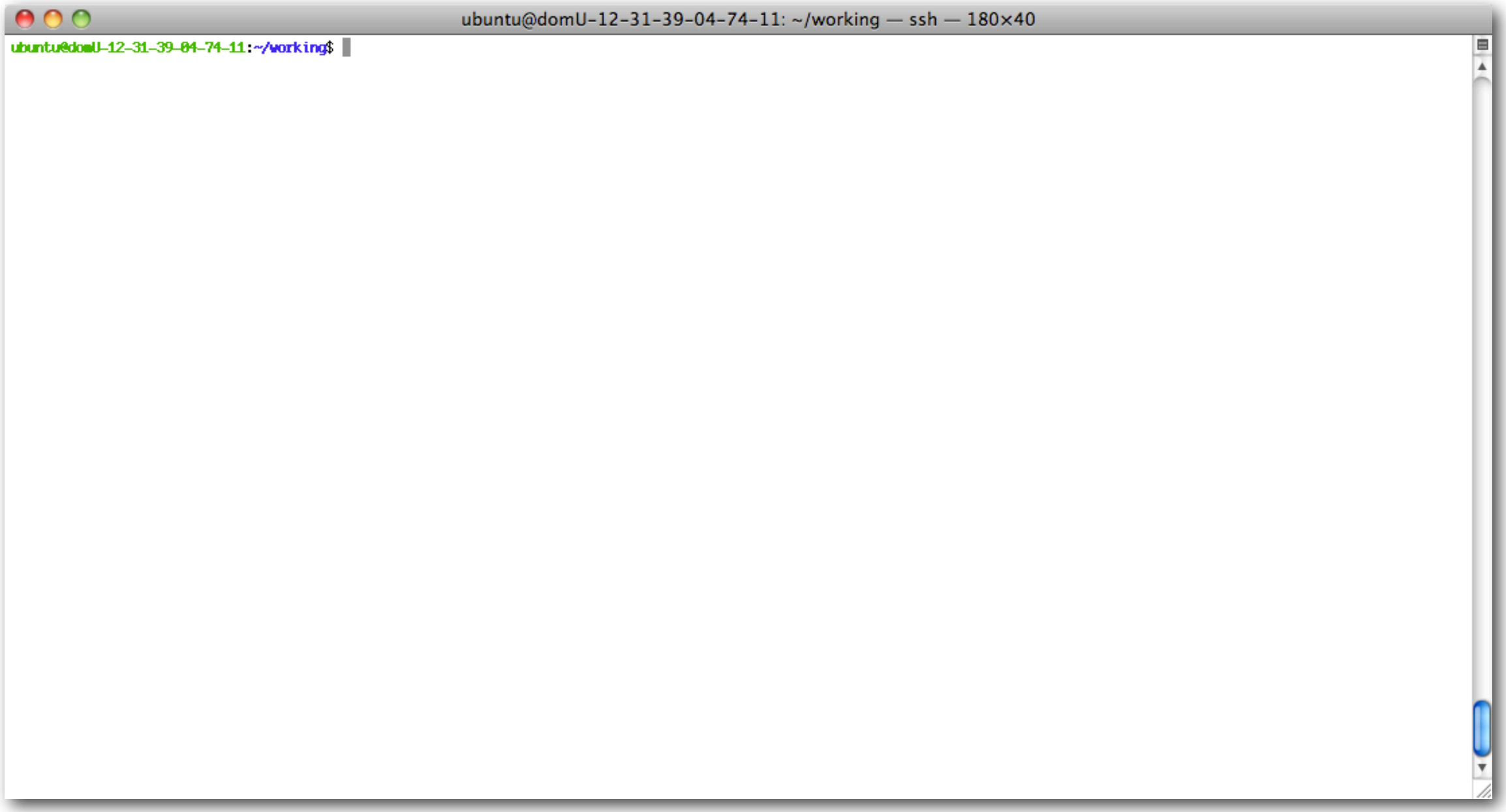
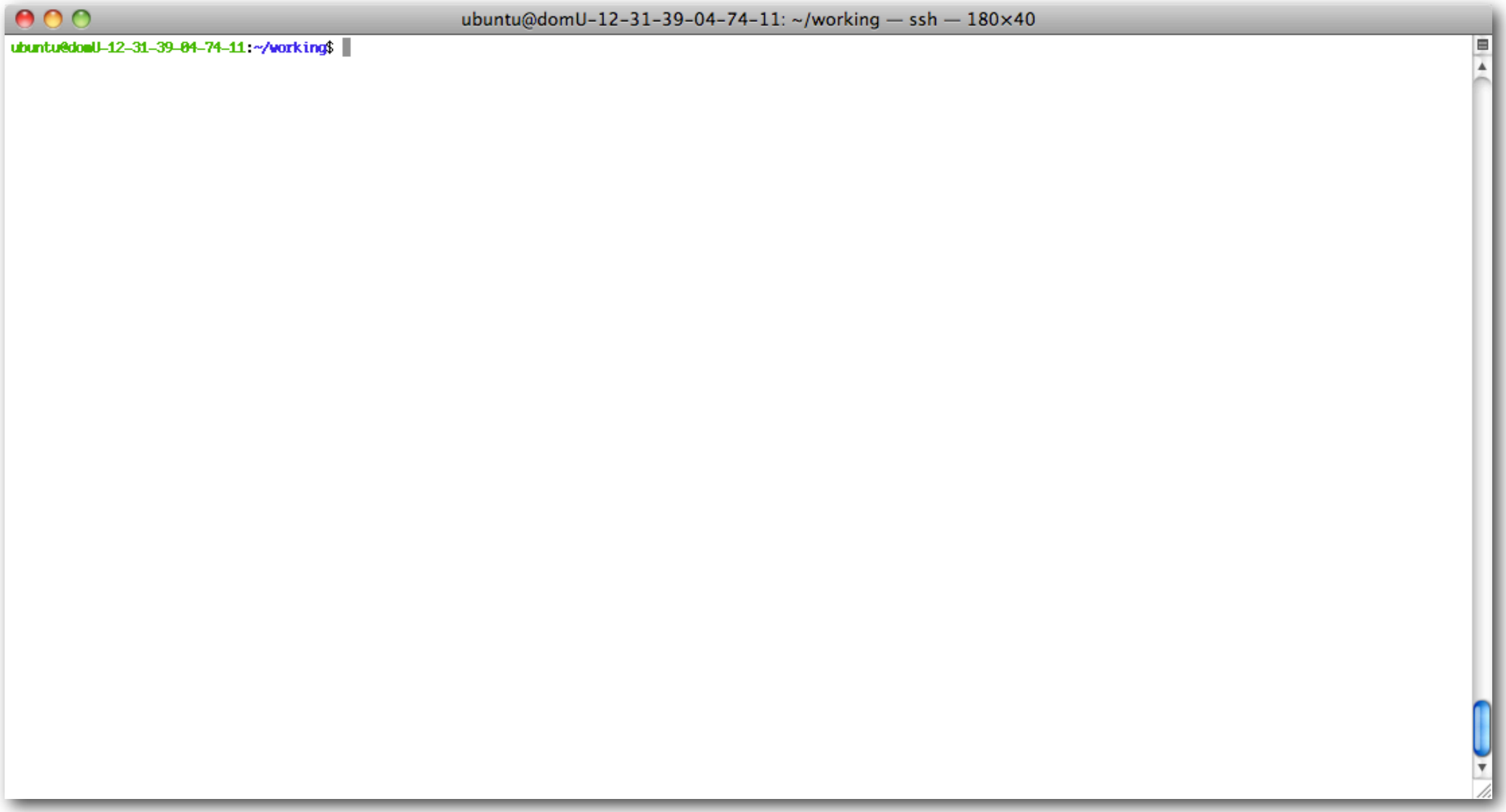
# Unix History



# What computers can run Unix?

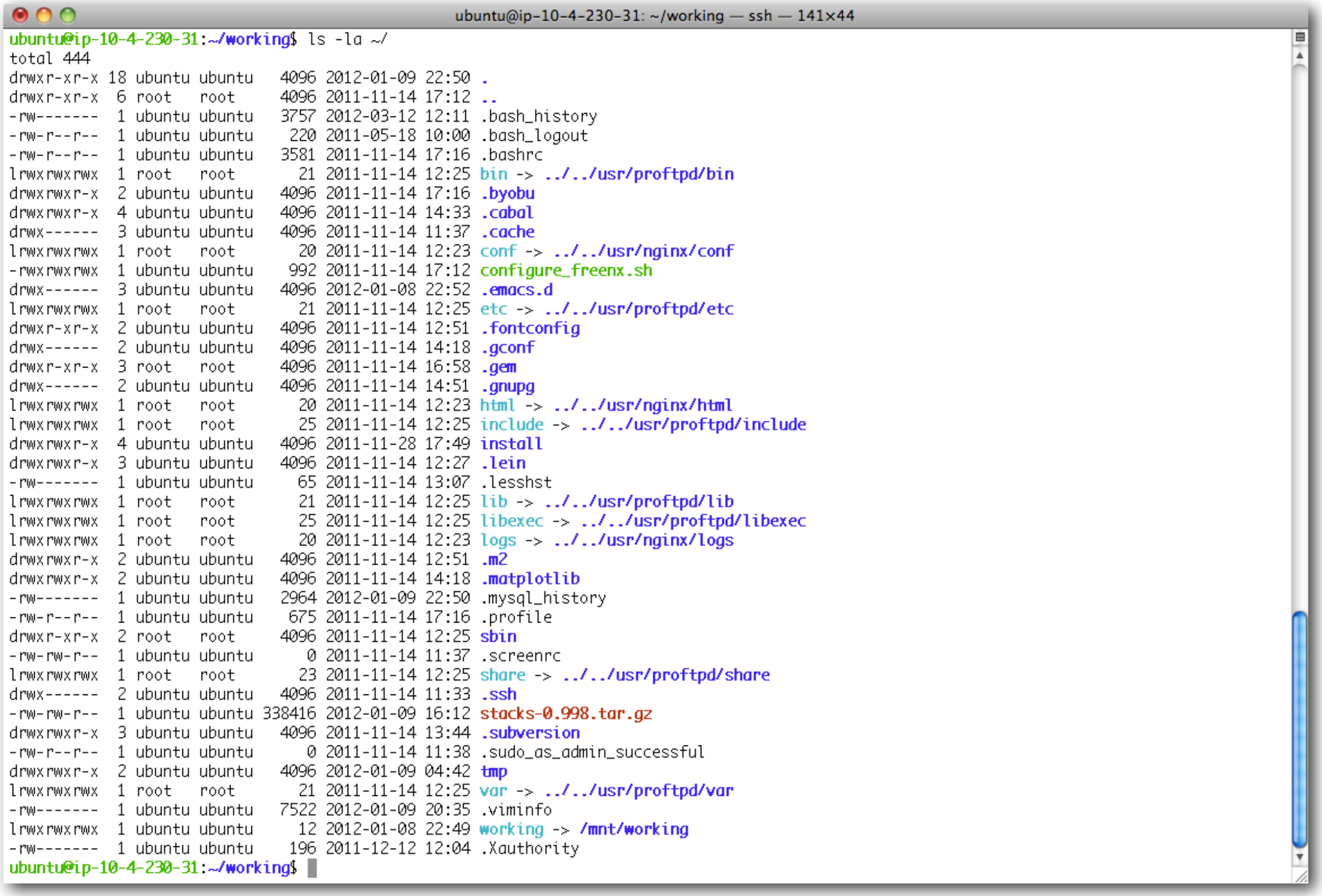


# What computers can run Unix?

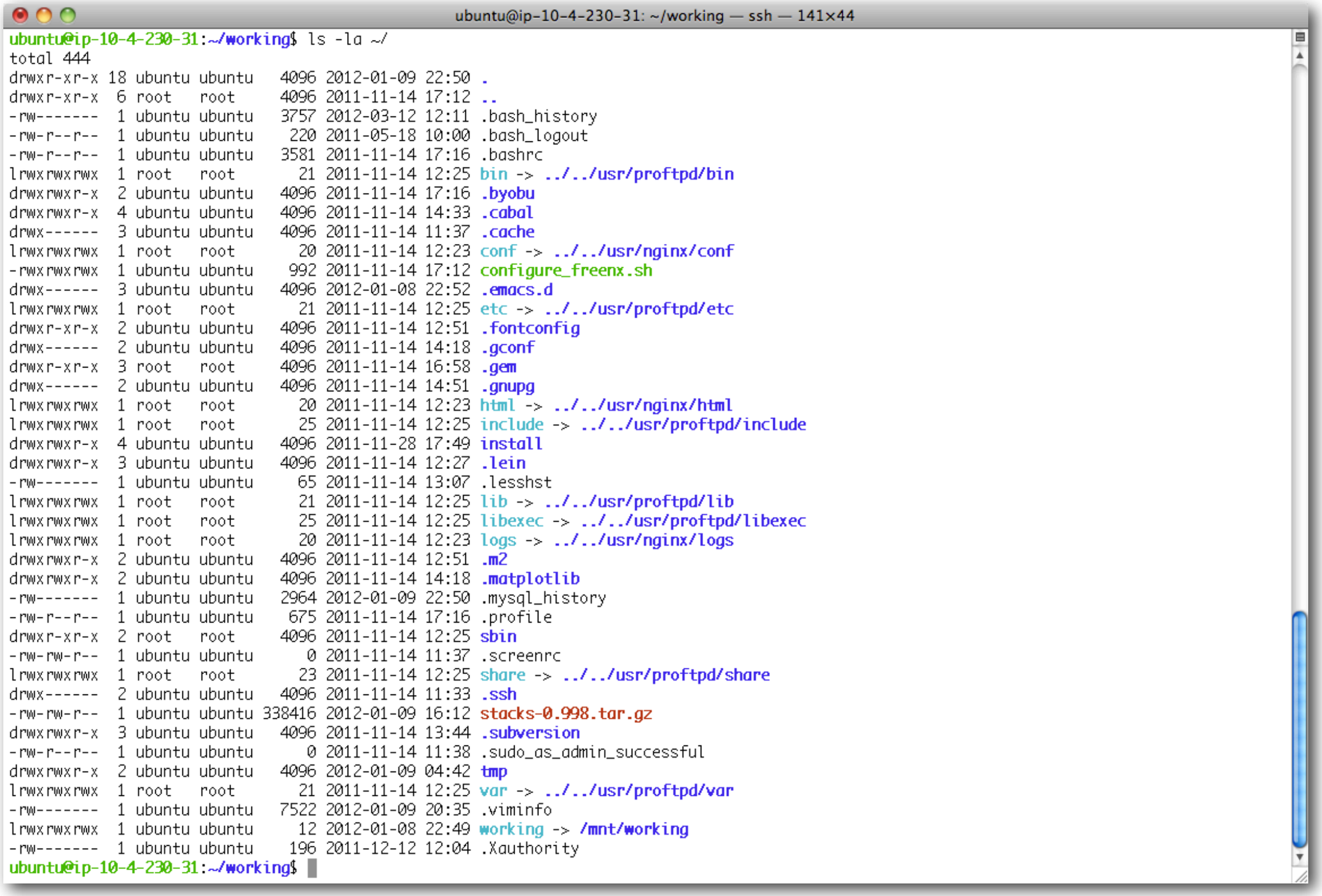



- Apple OS X Macs
- Google's Android phones
- Most airplane entertainment systems
- Wireless internet routers

# The Terminal Window



# The Terminal Window

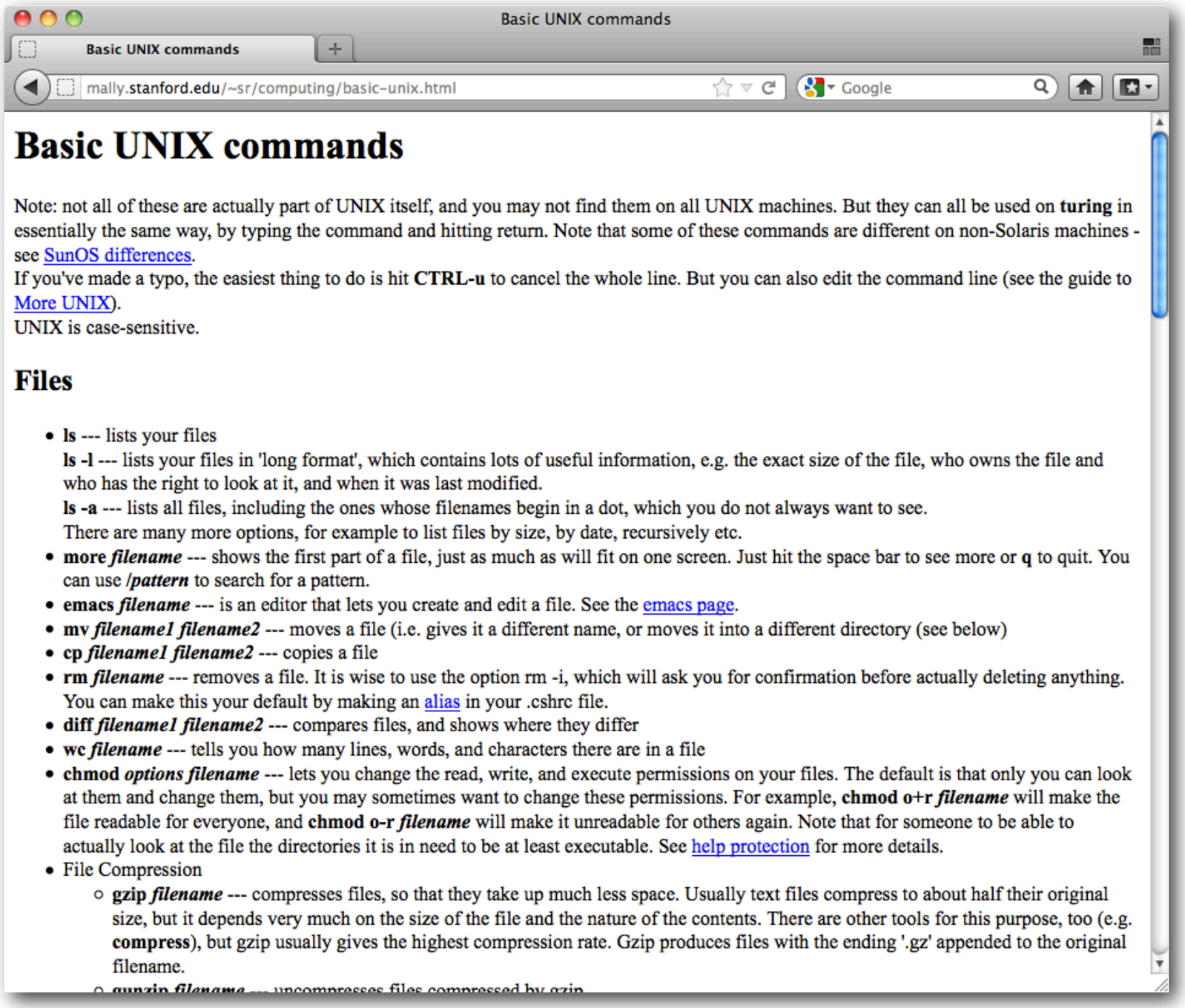



Make it comfortable to work in:

- Resize the window
- Change your font size
- Open multiple terminal windows

# Obtain a cheat sheet

google “unix commands”



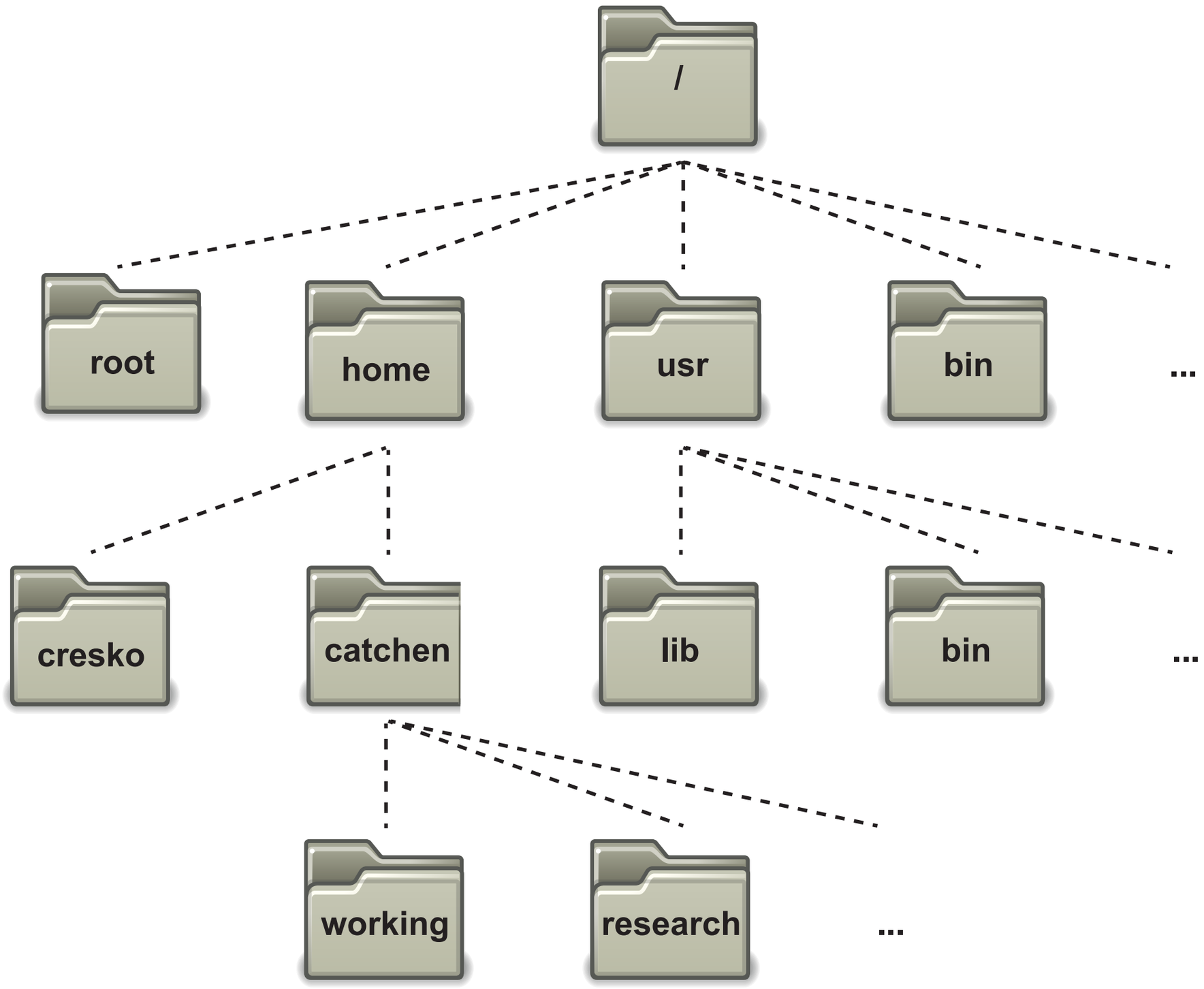

## Basic UNIX commands

Note: not all of these are actually part of UNIX itself, and you may not find them on all UNIX machines. But they can all be used on **turing** in essentially the same way, by typing the command and hitting return. Note that some of these commands are different on non-Solaris machines - see SunOS differences.
If you've made a typo, the easiest thing to do is hit **CTRL-u** to cancel the whole line. But you can also edit the command line (see the guide to More UNIX).
UNIX is case-sensitive.

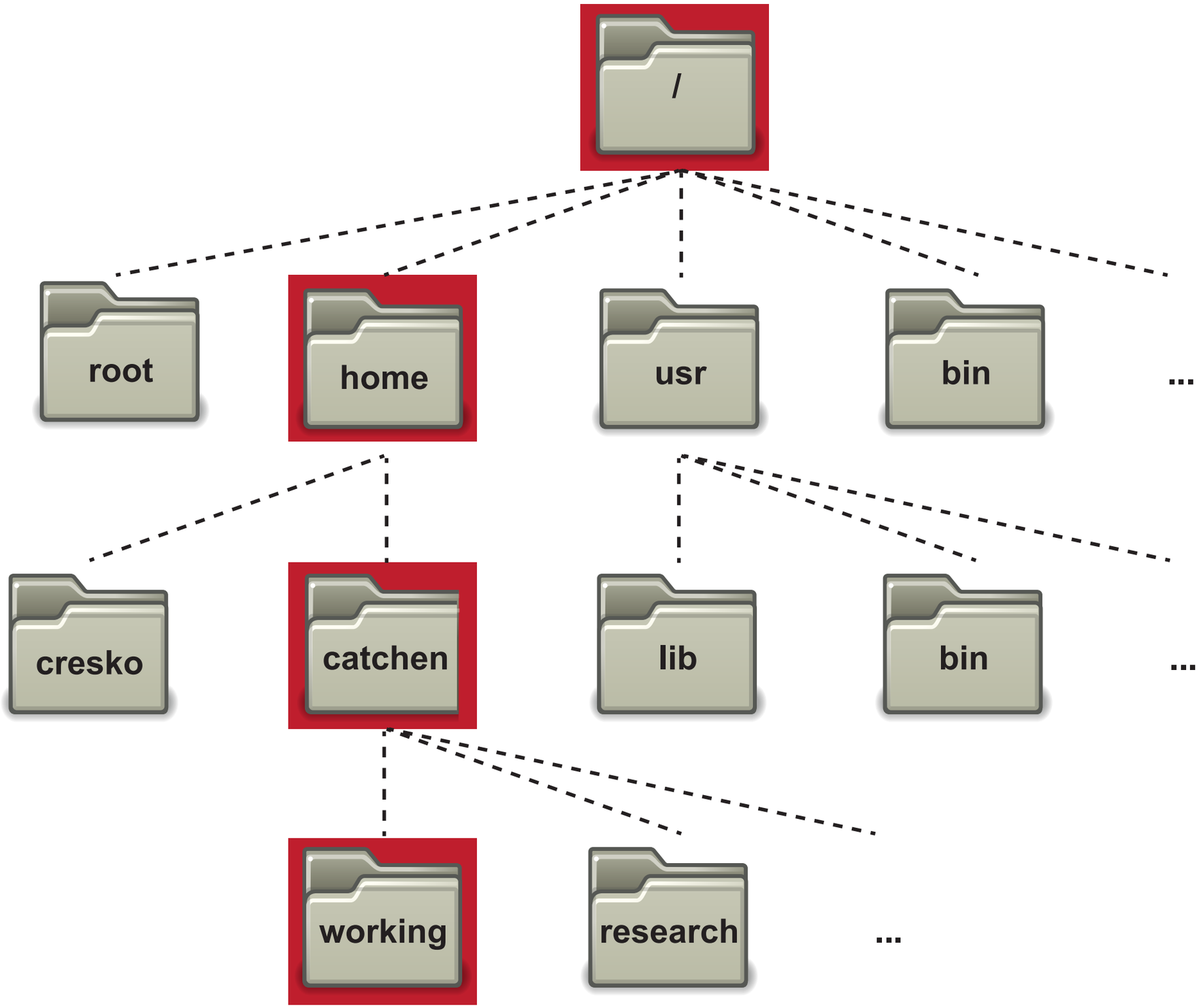
## Files

- **ls** --- lists your files
  **ls -l** --- lists your files in 'long format', which contains lots of useful information, e.g. the exact size of the file, who owns the file and who has the right to look at it, and when it was last modified.
  **ls -a** --- lists all files, including the ones whose filenames begin in a dot, which you do not always want to see.
  There are many more options, for example to list files by size, by date, recursively etc.
- **more** ***filename*** --- shows the first part of a file, just as much as will fit on one screen. Just hit the space bar to see more or **q** to quit. You can use ***/pattern*** to search for a pattern.
- **emacs** ***filename*** --- is an editor that lets you create and edit a file. See the emacs page.
- **mv** ***filename1 filename2*** --- moves a file (i.e. gives it a different name, or moves it into a different directory (see below)
- **cp** ***filename1 filename2*** --- copies a file
- **rm** ***filename*** --- removes a file. It is wise to use the option rm -i, which will ask you for confirmation before actually deleting anything. You can make this your default by making an alias in your .cshrc file.
- **diff** ***filename1 filename2*** --- compares files, and shows where they differ
- **wc** ***filename*** --- tells you how many lines, words, and characters there are in a file
- **chmod** ***options filename*** --- lets you change the read, write, and execute permissions on your files. The default is that only you can look at them and change them, but you may sometimes want to change these permissions. For example, **chmod o+r** ***filename*** will make the file readable for everyone, and **chmod o-r** ***filename*** will make it unreadable for others again. Note that for someone to be able to actually look at the file the directories it is in need to be at least executable. See help protection for more details.
- File Compression
  - **gzip** ***filename*** --- compresses files, so that they take up much less space. Usually text files compress to about half their original size, but it depends very much on the size of the file and the nature of the contents. There are other tools for this purpose, too (e.g. **compress**), but gzip usually gives the highest compression rate. Gzip produces files with the ending '.gz' appended to the original filename.
  - **gunzip** ***filename*** --- uncompresses files compressed by gzip

# In UNIX everything is a file organized in a hierarchy



# Paths



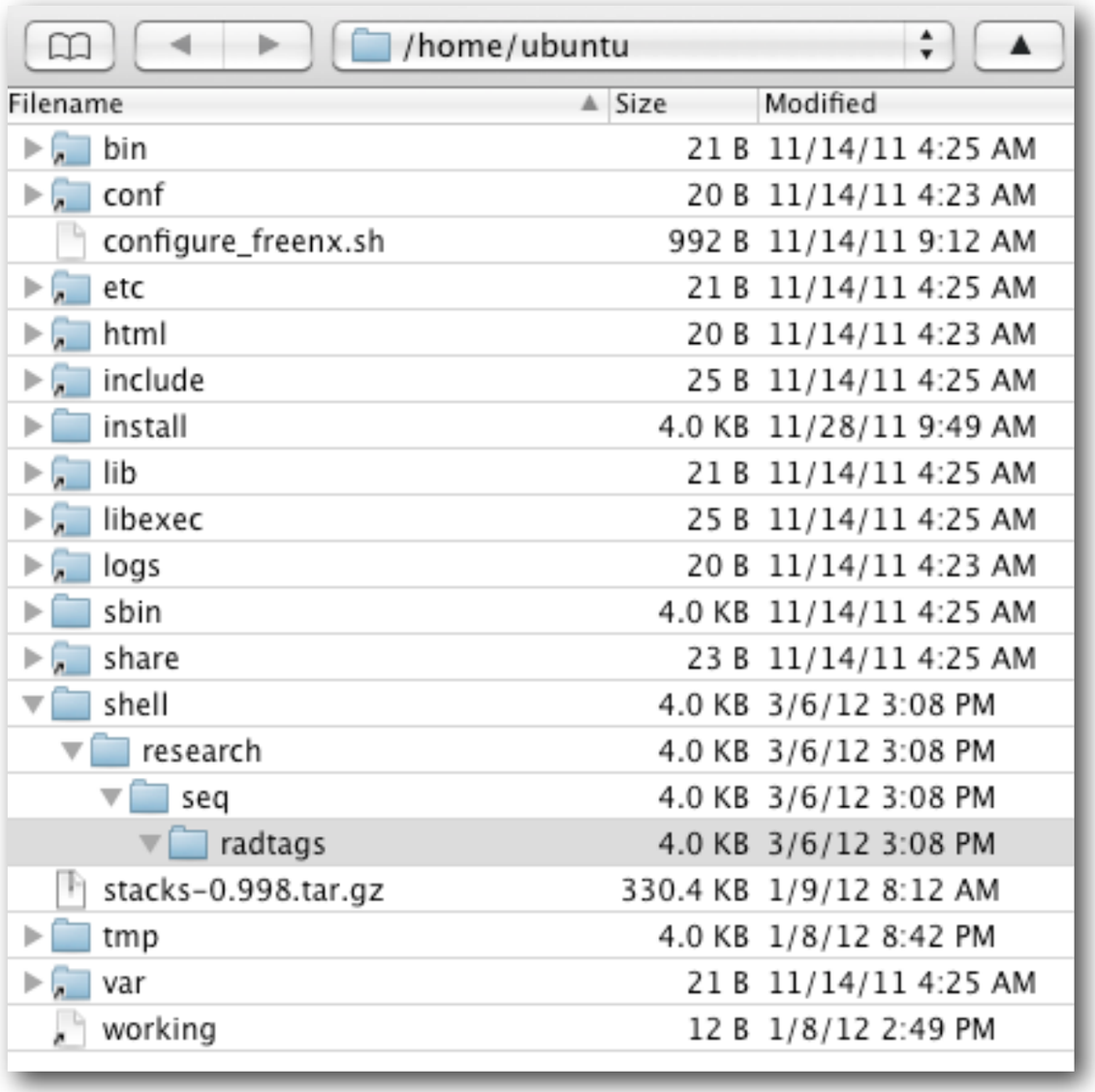

**`/home/catchen/working`**

# Paths, cont

This shell view of the nested directories shell, research, seq, and radtags…..

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

…. is equivalent to this GUI view of the same directories




And the `radtags` directory is uniquely identified by its path:

`/home/ubuntu/shell/research/seq/radtags`

# Create a series of directories

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

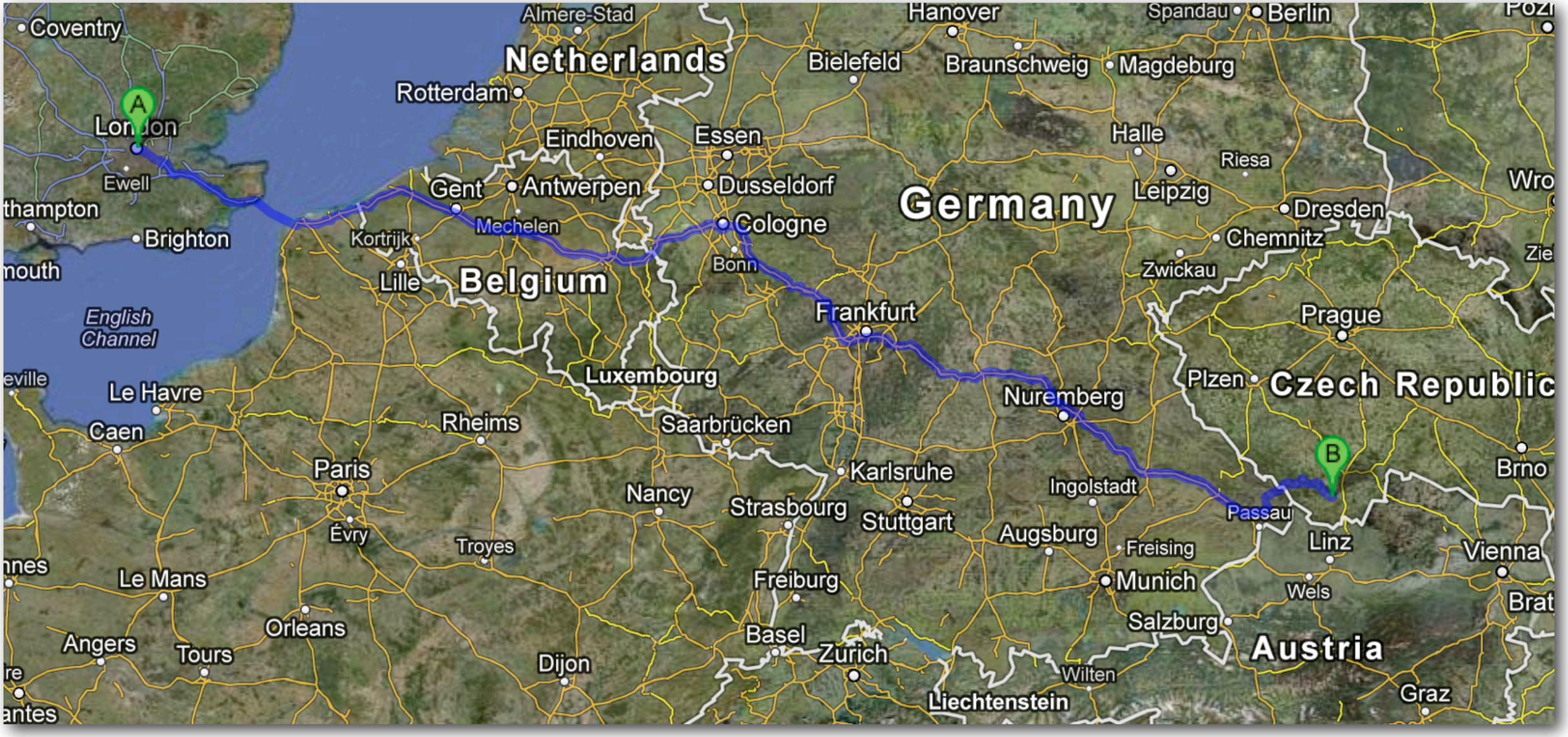
# Create a series of directories

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

```
% mkdir shell        % cd shell        % ls
```

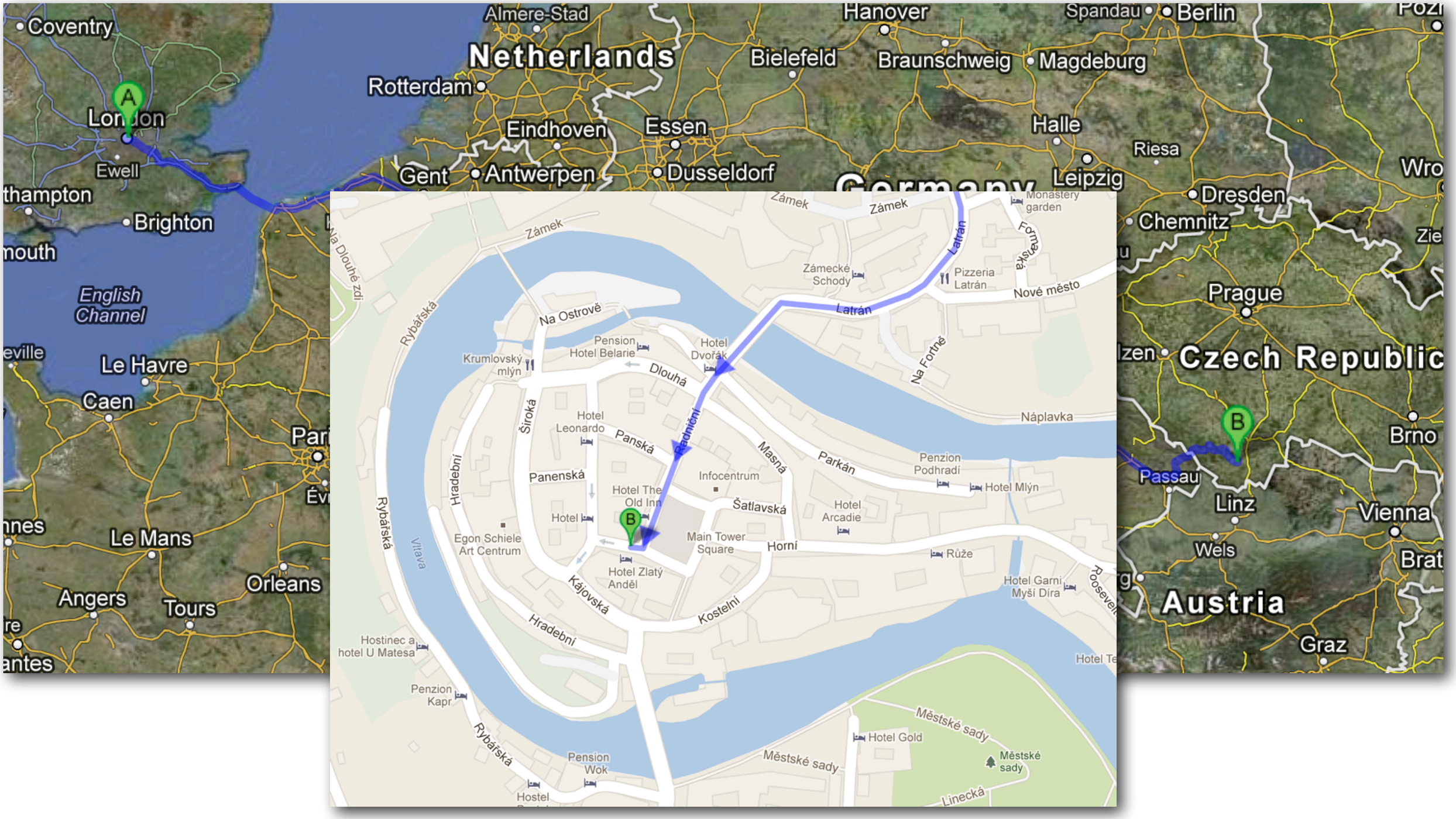
# Absolute and relative paths

## *How do I get to the Hotel Zlaty Andel?*



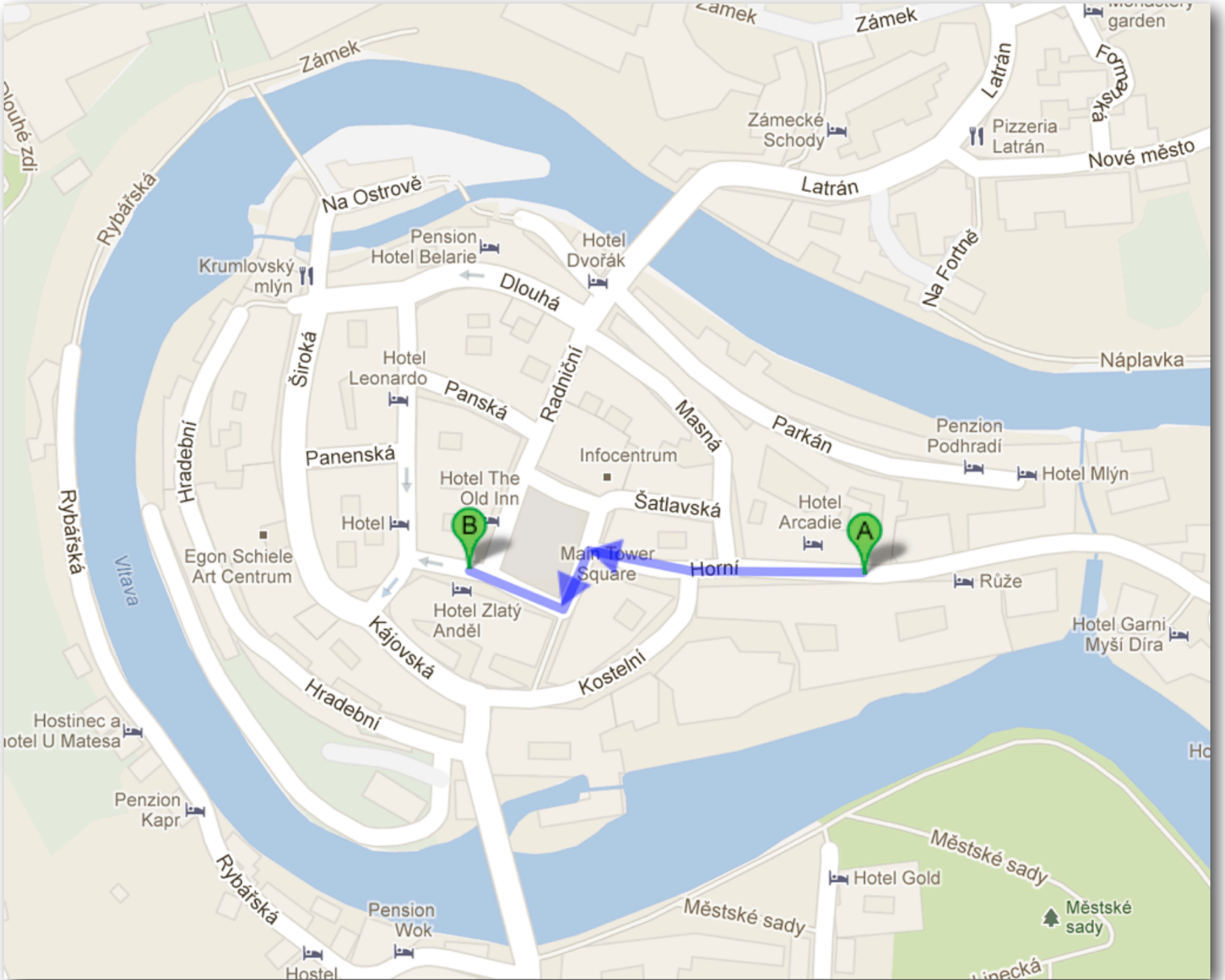
# Absolute and relative paths

## *How do I get to the Hotel Zlaty Andel?*

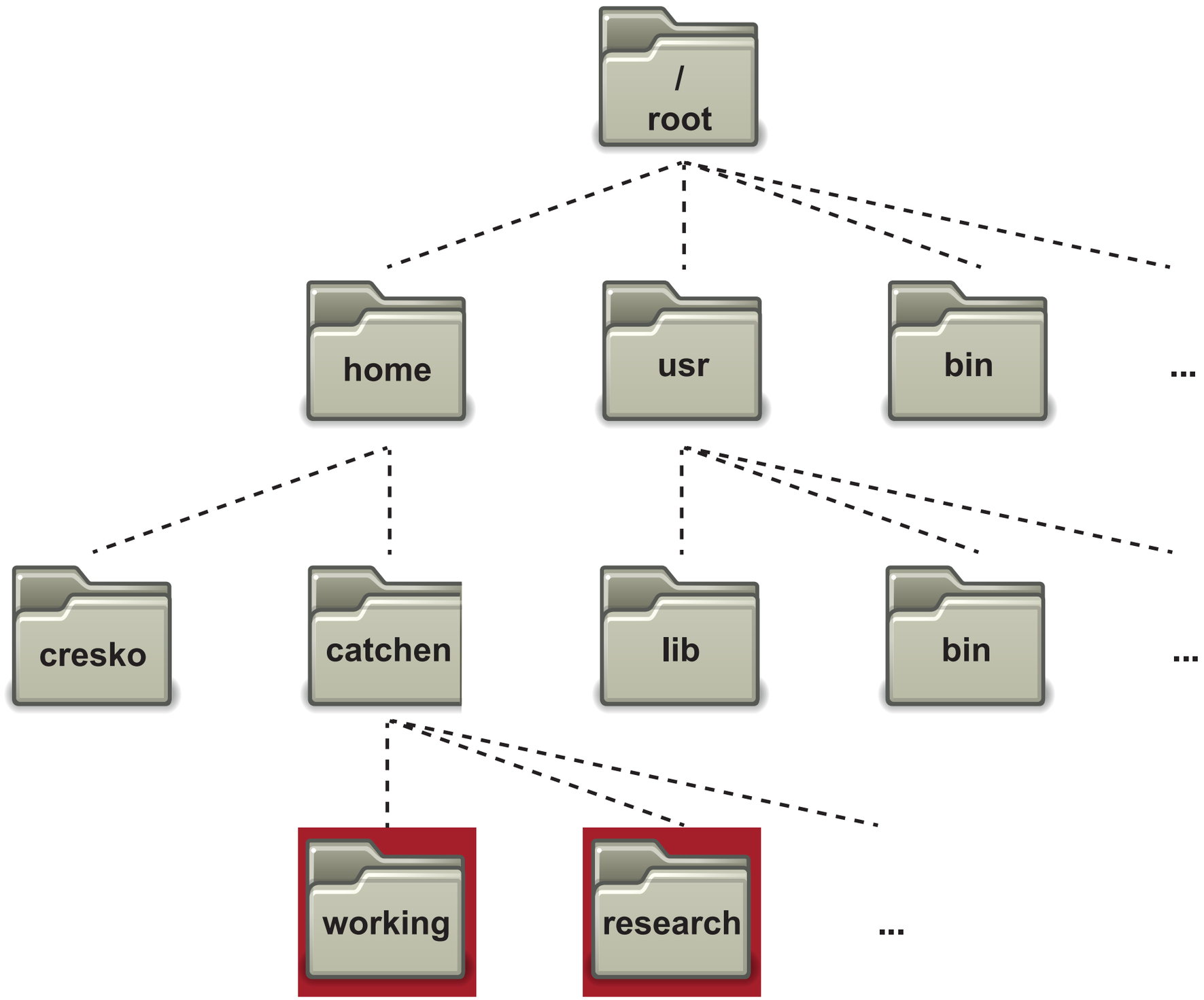


# Absolute and relative paths

## *How do I get to the Hotel Zlaty Andel?*

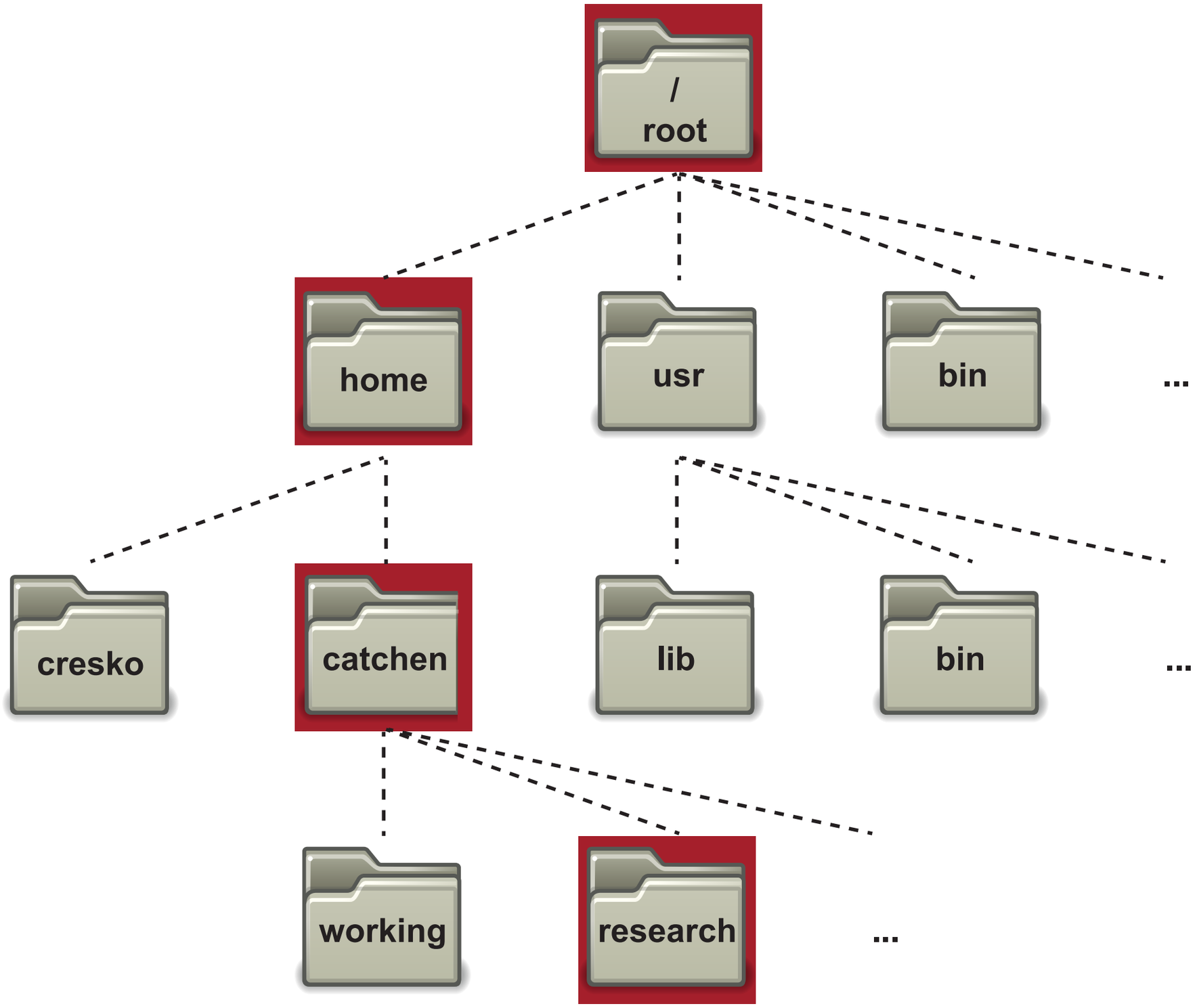


# Paths



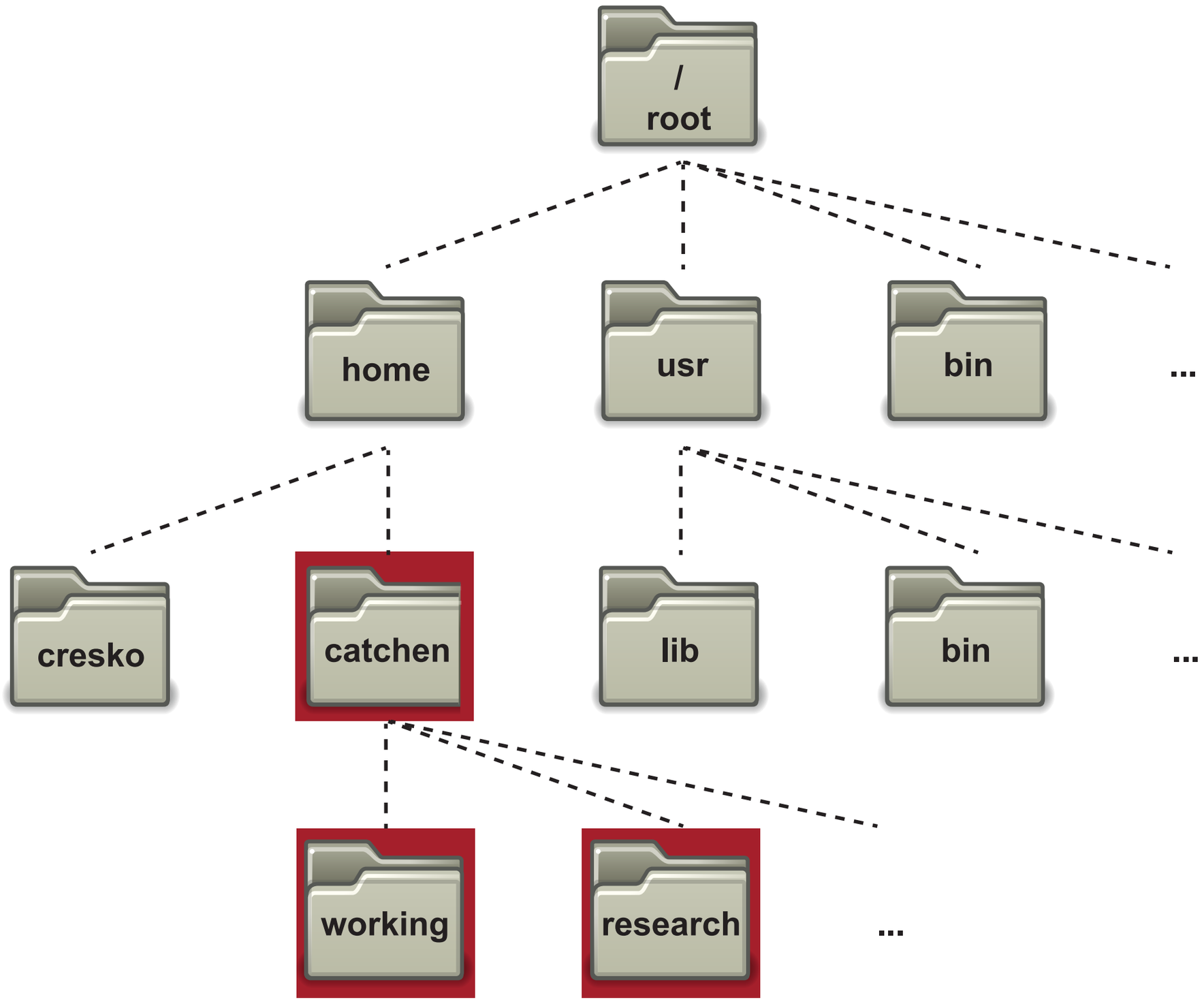

**`/home/catchen/working`**

# Relative Path



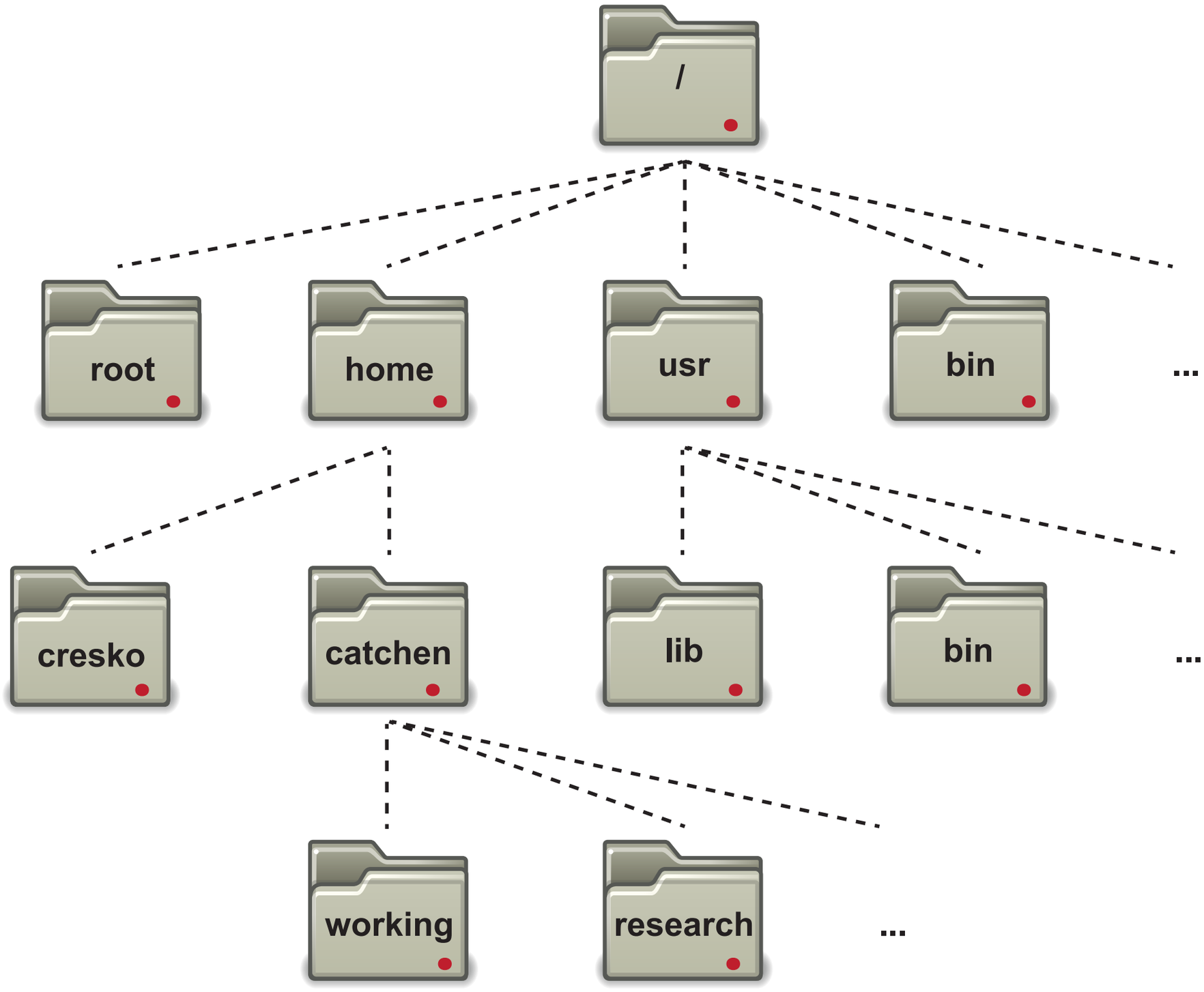

`/home/catchen/research`

# Relative Path

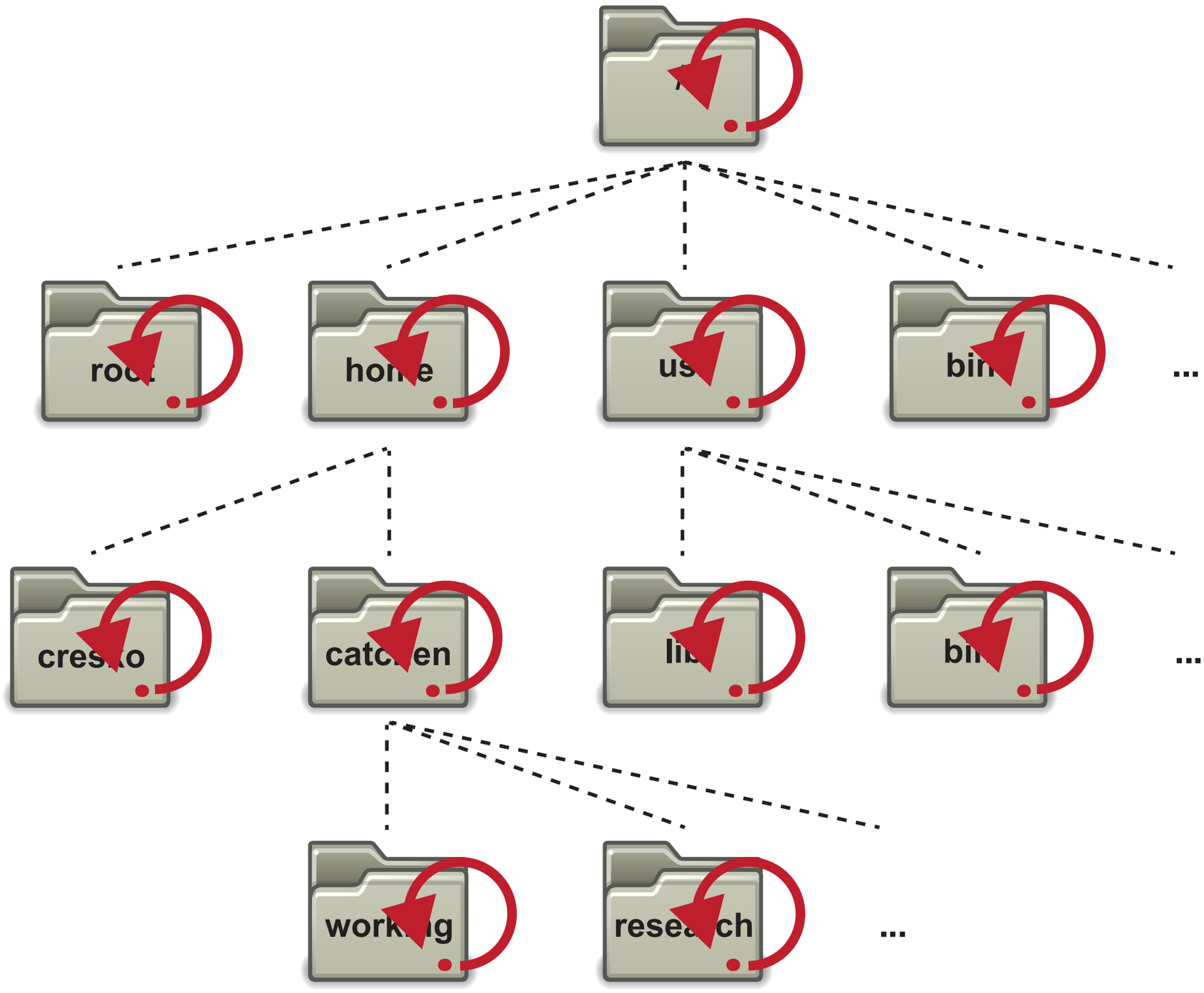



```
../working
```

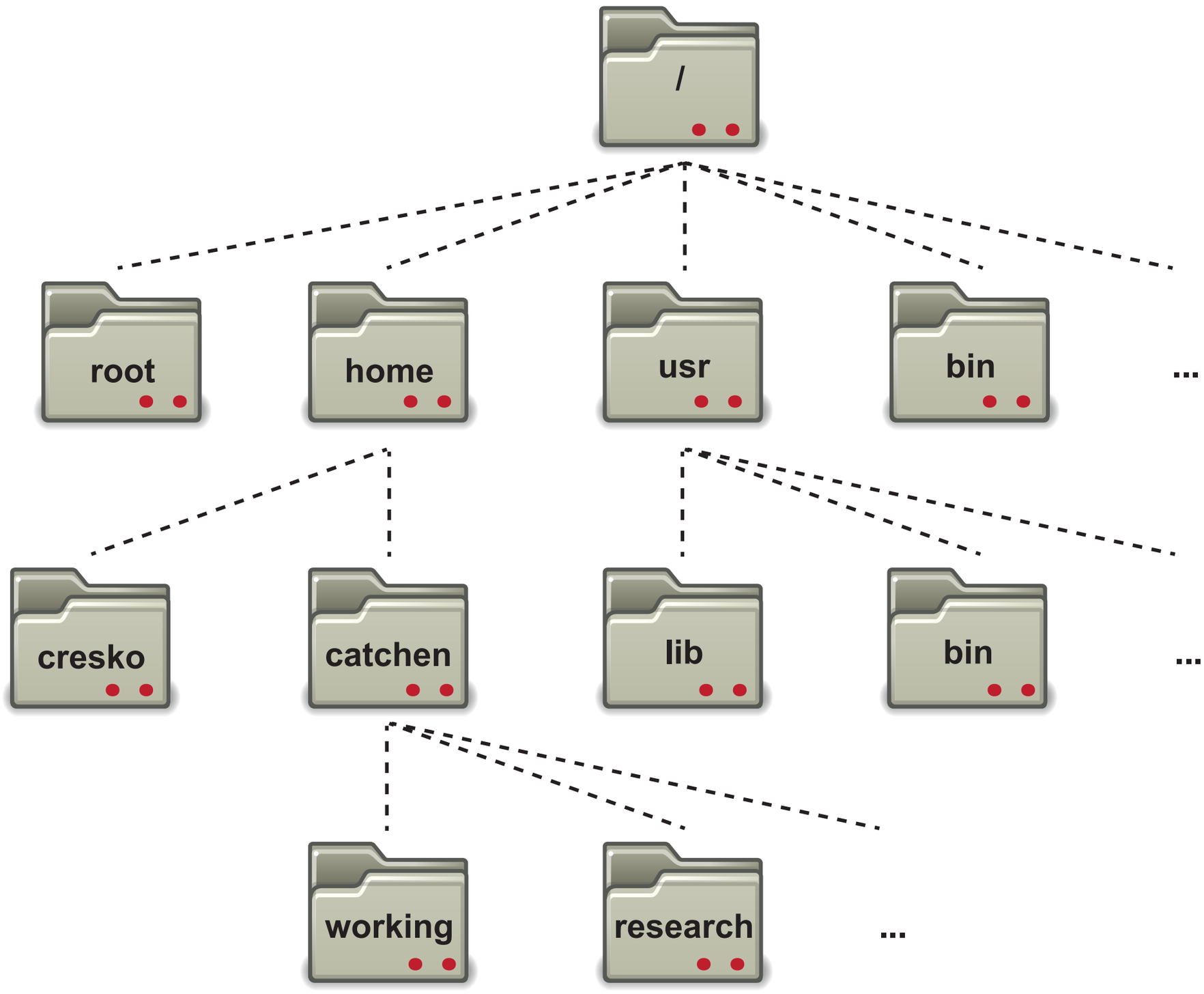
# Special files -- '*dot*'



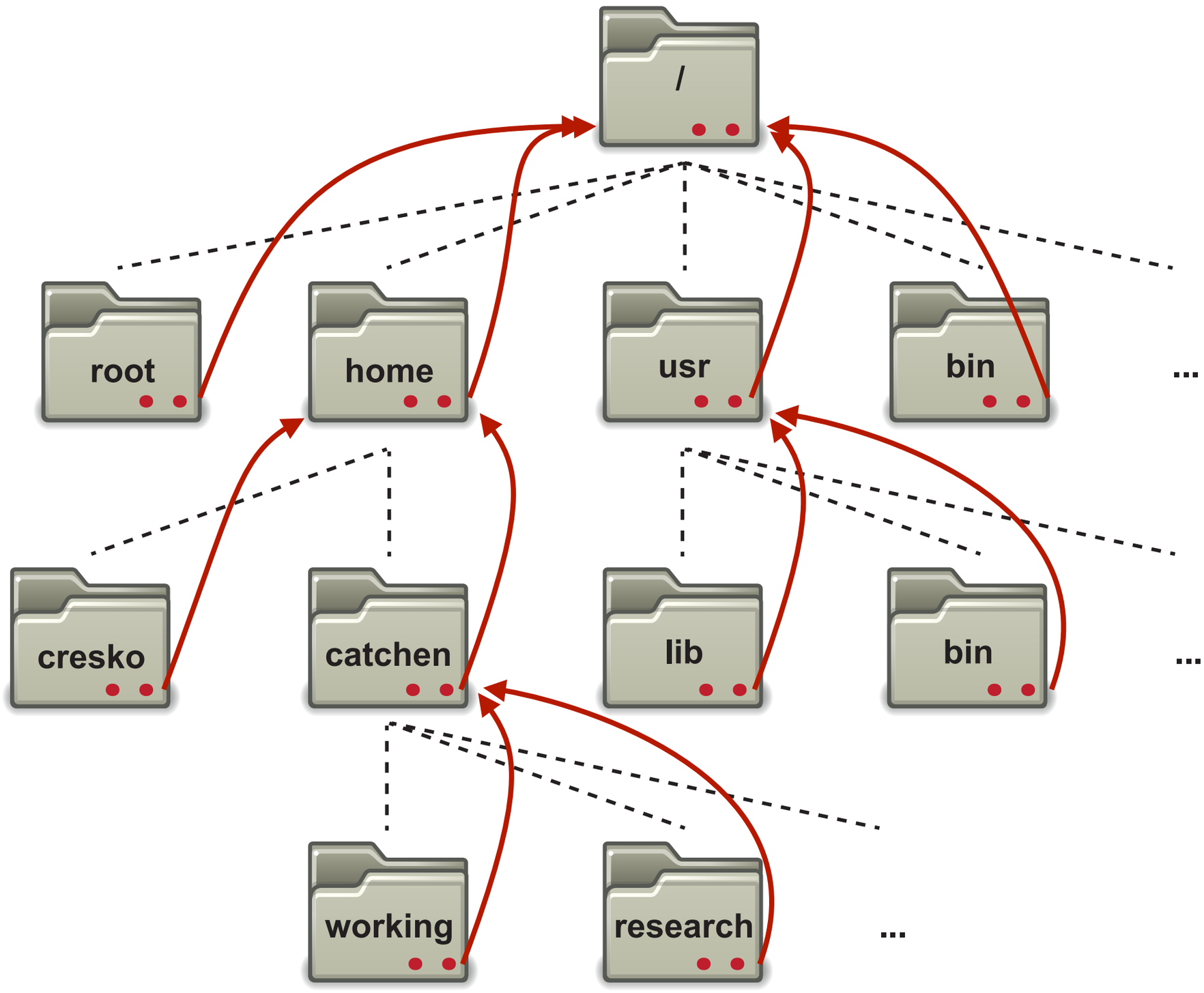
# Special files -- '*dot*'



# Special files -- '*dot dot*'



# Special files -- '*dot dot*'



# Absolute and relative paths

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

## Special Files

*dot*

*dot dot*

# Absolute and relative paths

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

## Special Files

*dot*

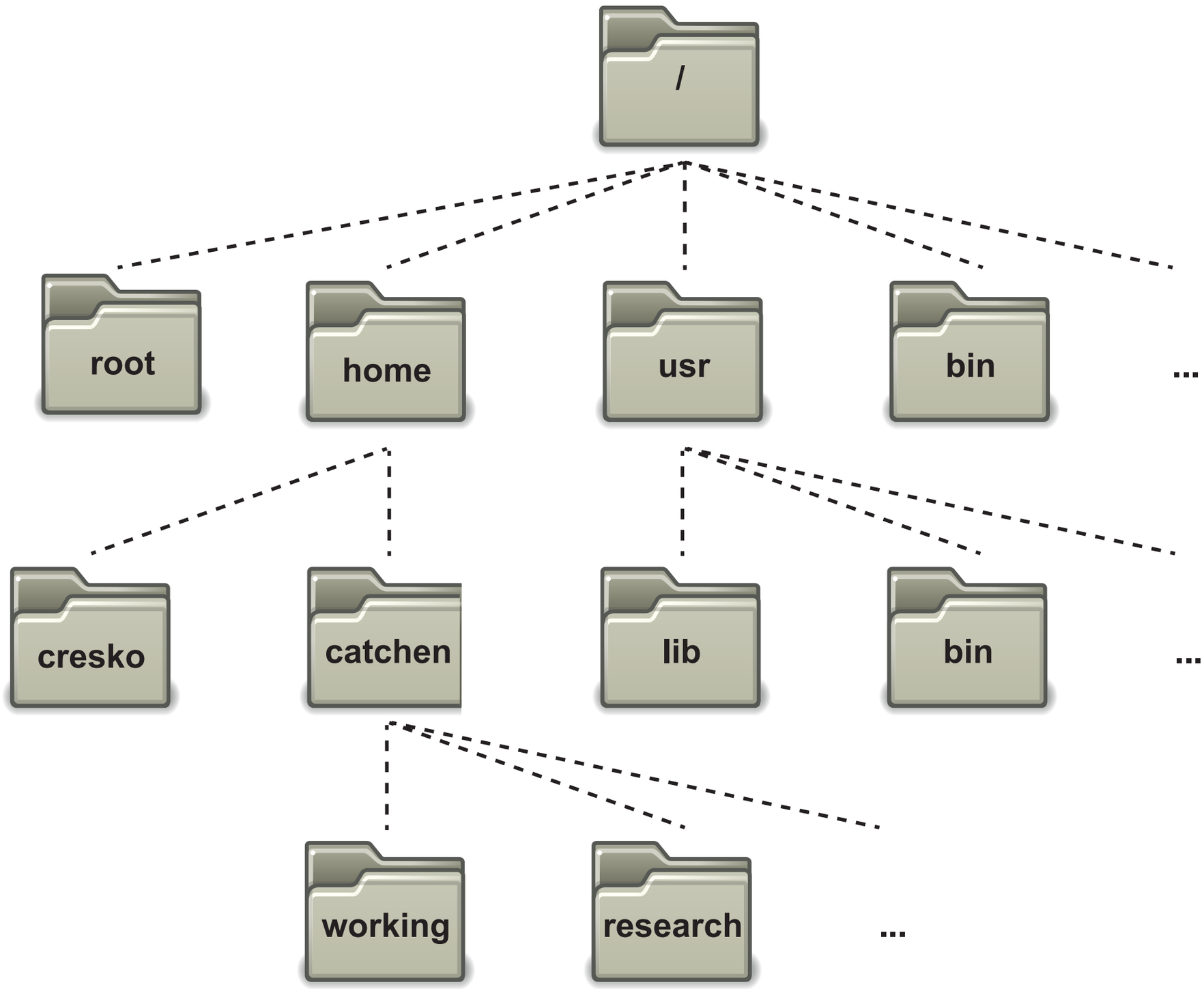
*dot dot*

```
% ls .
```

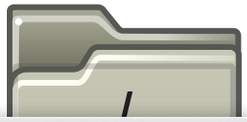
```
% ls ..
```

```
% ls ../../
```

# Binary programs - ls, cp, mkdir, etc.



# Binary programs - ls, cp, mkdir, etc.



ubuntu@ip-10-140-6-74: ~ — ssh — 133×24

```
ubuntu@ip-10-140-6-74:~$ ls /bin
bash                csh                 getfacl             lsblk               ntfsck                  readlink            true
bunzip2             dash                grep                lsmod               ntfscluster             rm                  ulockmgr_server
busybox             date                gunzip              mkdir               ntfscmp                 rmdir               umount
bzcat               dbus-cleanup-sockets gzexe              mknod               ntfsdecrypt             rnano               uname
bzcmp               dbus-daemon         gzip                mktemp              ntfsdump_logfile        run-parts           uncompress
bzdiff              dbus-uuidgen        hostname            more                ntfsfix                 rzsh                unicode_start
bzegrep             dd                  init-checkconf      mount               ntfsinfo                sed                 vdir
bzexe               df                  initctl2dot         mountpoint          ntfsls                  setfacl             which
bzfgrep             dir                 ip                  mt                  ntfsmftalloc            setfont             ypdomainname
bzgrep              dmesg               kbd_mode            mt-gnu              ntfsmove                setupcon            zcat
bzip2               dnsdomainname       kill                mv                  ntfstruncate            sh                  zcmp
bzip2recover        domainname          ksh                 nano                ntfswipe                sh.distrib          zdiff
bzless              dumpkeys            less                nc                  open                    sleep               zegrep
bzmore              echo                lessecho            nc.openbsd          openvt                  static-sh           zfgrep
cat                 ed                  lessfile            netcat              pidof                   stty                zforce
chacl               egrep               lesskey             netstat             ping                    su                  zgrep
chgrp               false               lesspipe            nisdomainname       ping6                   sync                zless
chmod               fgconsole           ln                  ntfs-3g             plymouth                tailf               zmore
chown               fgrep               loadkeys            ntfs-3g.probe       plymouth-upstart-bridge tar                 znew
chvt                findmnt             login               ntfs-3g.secaudit    ps                      tcsh                zsh
cp                  fuser               lowntfs-3g          ntfs-3g.usermap     pwd                     tempfile            zsh4
cpio                fusermount          ls                  ntfscat             rbash                   touch
ubuntu@ip-10-140-6-74:~$
```



...

**% ls /bin**

# Relative and absolute paths

A shortcut to your ‘home’, tilde:

~

Moving through the filesystem:

**cd**

Knowing where you are:

**pwd**

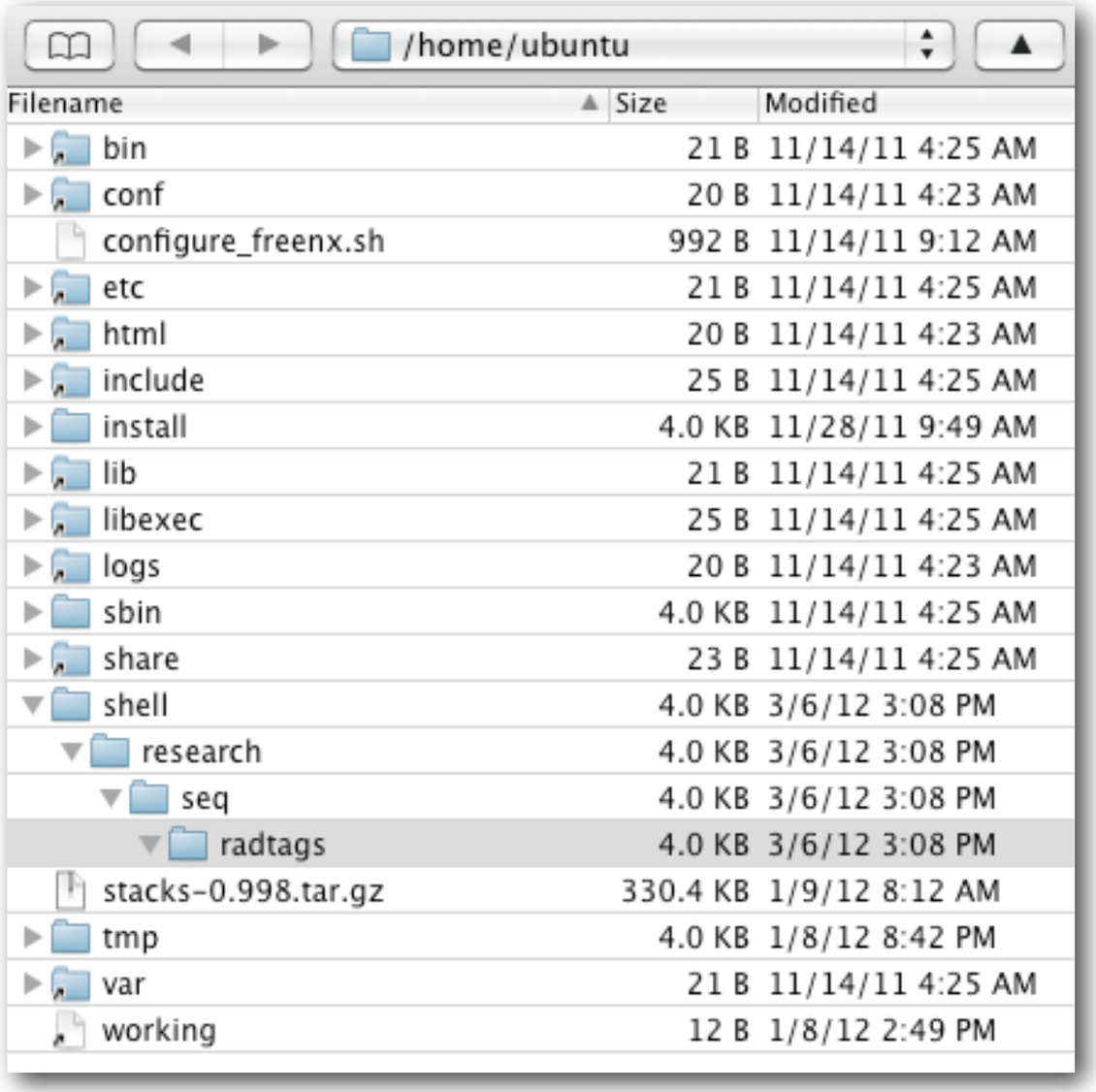
```
% ls ~/          % cd ~/          % cd          % pwd
```

# Relative and absolute paths

`/home/ubuntu/shell/research/seq/radtags`

Create a series of directories under shell:

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

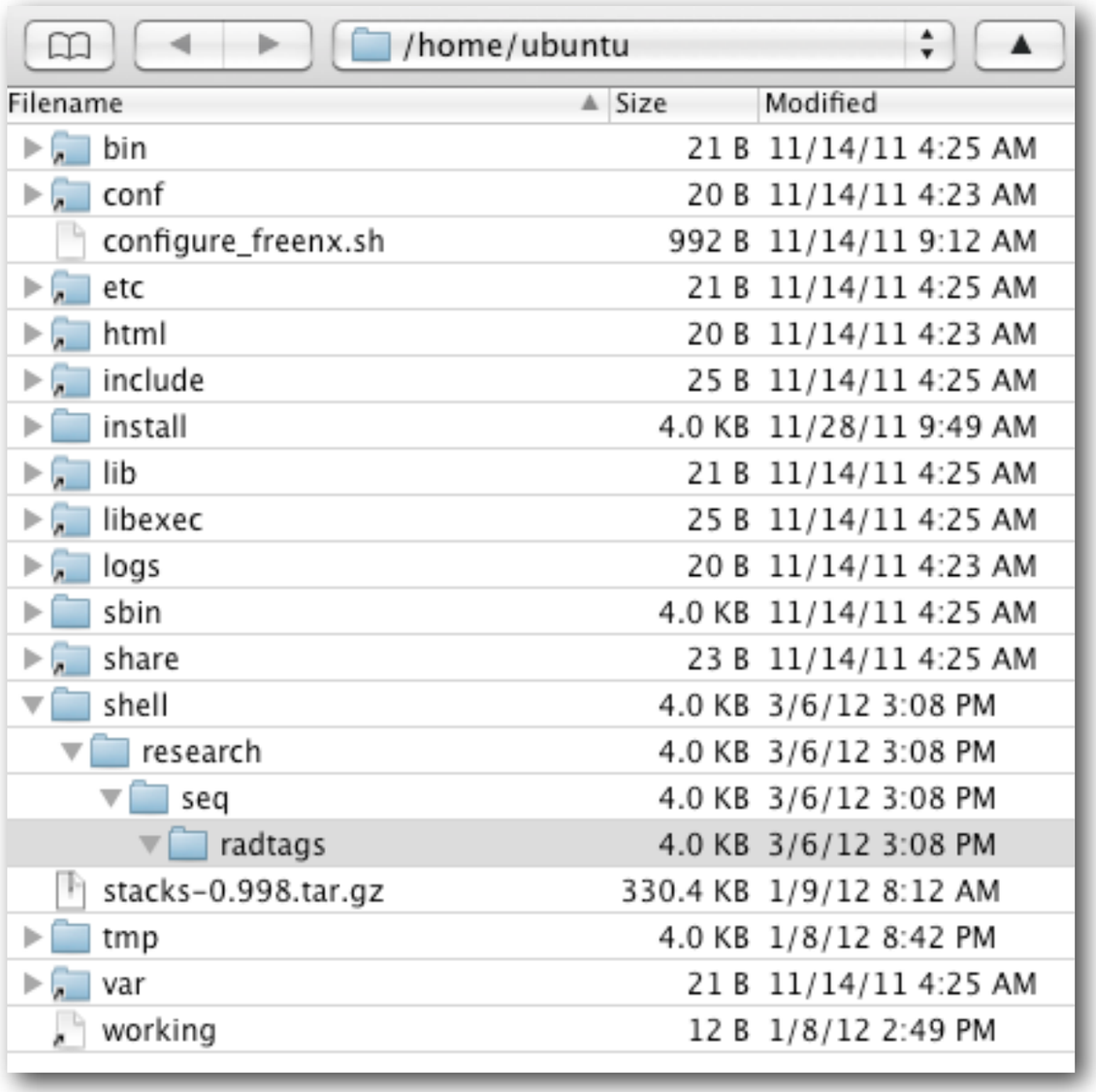


# Relative and absolute paths

`/home/ubuntu/shell/research/seq/radtags`

Create a series of directories under shell:

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```




`% ls .`

`% ls ..`

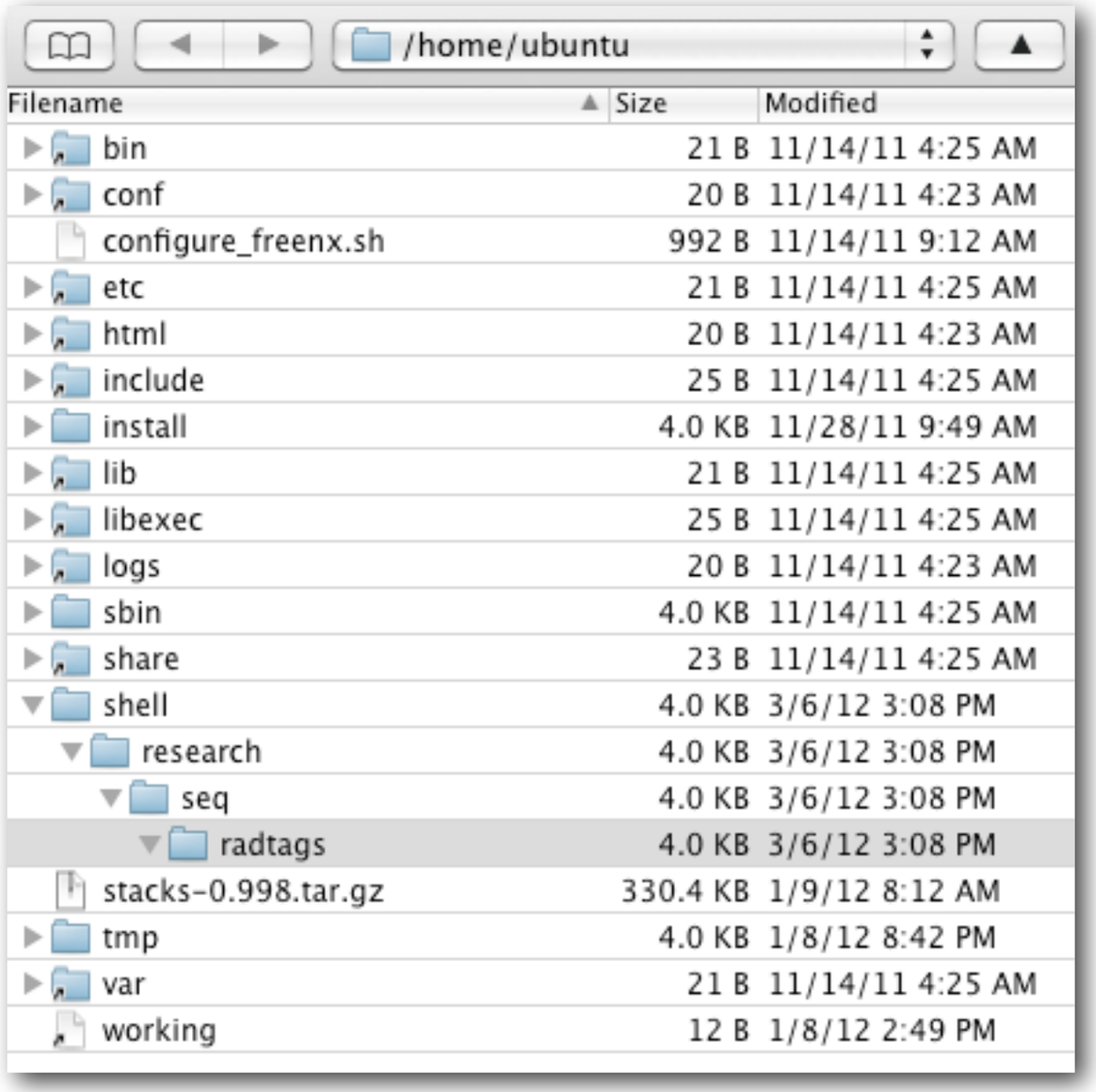
`% ls ../../`

# Relative and absolute paths

`/home/ubuntu/shell/research/seq/radtags`

Create a series of directories under shell:

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```




```
% ls .          % ls ..          % ls ../../

% cd ~/         % cd shell/research      % pwd
```

# Are you typing? You're doing it wrong.

## Tab-completion:

- Tab once to complete uniquely
- Tab twice to see all possible completions

## Up-arrow:

- Previous commands can be found by pressing "up-arrow"

## 'history'

```
root@ubuntu:~# cd /etc/
root@ubuntu:/etc# ls -la
total 788
drwxr-xr-x 82 root root    4096 2012-06-25 01:56 .
drwxr-xr-x 23 root root    4096 2012-04-18 09:17 ..
-rw-r--r--  1 root root    2981 2012-04-16 10:09 adduser.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 alternatives
drwxr-xr-x  3 root root    4096 2012-04-16 10:19 apm
drwxr-xr-x  3 root root    4096 2012-04-16 10:21 apparmor
drwxr-xr-x  8 root root    4096 2012-04-16 12:09 apparmor.d
drwxr-xr-x  3 root root    4096 2012-04-16 12:09 apport
drwxr-xr-x  6 root root    4096 2012-04-16 12:08 apt
-rw-r-----  1 root daemon   144 2011-05-16 03:32 at.deny
-rw-r--r--  1 root root    1975 2011-05-18 03:00 bash.bashrc
-rw-r--r--  1 root root   58753 2011-10-04 13:53 bash_completion
drwxr-xr-x  3 root root   12288 2012-04-16 12:09 bash_completion.d
-rw-r--r--  1 root root     344 2011-10-04 12:56 bindresvport.blacklist
-rw-r--r--  1 root root     321 2011-08-09 09:16 blkid.conf
lrwxrwxrwx  1 root root      15 2011-08-09 09:16 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 byobu
drwxr-xr-x  3 root root    4096 2012-04-16 10:18 ca-certificates
-rw-r--r--  1 root root    7014 2012-04-16 10:21 ca-certificates.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 calendar
drwxr-s---  2 root dip     4096 2012-04-16 10:22 chatscripts
drwxr-xr-x  4 root root    4096 2012-04-16 12:10 cloud
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 console-setup
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.d
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 cron.daily
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.hourly
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.monthly
-rw-r--r--  1 root root     724 2011-09-19 17:04 crontab
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 cron.weekly
drwxr-xr-x  4 root root    4096 2012-04-16 10:21 dbus-1
-rw-r--r--  1 root root    2969 2011-06-23 00:01 debconf.conf
-rw-r--r--  1 root root      11 2011-07-08 10:13 debian_version
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 default
```

# Are you typing? You're doing it wrong.

## Tab-completion:

- Tab once to complete uniquely
- Tab twice to see all possible completions

## Up-arrow:

- Previous commands can be found by pressing "up-arrow"

## 'history'

```
root@ubuntu:~# cd /etc/
root@ubuntu:/etc# ls -la
total 788
drwxr-xr-x 82 root root    4096 2012-06-25 01:56 .
drwxr-xr-x 23 root root    4096 2012-04-18 09:17 ..
-rw-r--r--  1 root root    2981 2012-04-16 10:09 adduser.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 alternatives
drwxr-xr-x  3 root root    4096 2012-04-16 10:19 apm
drwxr-xr-x  3 root root    4096 2012-04-16 10:21 apparmor
drwxr-xr-x  8 root root    4096 2012-04-16 12:09 apparmor.d
drwxr-xr-x  3 root root    4096 2012-04-16 12:09 apport
drwxr-xr-x  6 root root    4096 2012-04-16 12:08 apt
-rw-r-----  1 root daemon   144 2011-05-16 03:32 at.deny
-rw-r--r--  1 root root    1975 2011-05-18 03:00 bash.bashrc
-rw-r--r--  1 root root   58753 2011-10-04 13:53 bash_completion
drwxr-xr-x  3 root root   12288 2012-04-16 12:09 bash_completion.d
-rw-r--r--  1 root root     344 2011-10-04 12:56 bindresvport.blacklist
-rw-r--r--  1 root root     321 2011-08-09 09:16 blkid.conf
lrwxrwxrwx  1 root root      15 2011-08-09 09:16 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 byobu
drwxr-xr-x  3 root root    4096 2012-04-16 10:18 ca-certificates
-rw-r--r--  1 root root    7014 2012-04-16 10:21 ca-certificates.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 calendar
drwxr-s---  2 root dip     4096 2012-04-16 10:22 chatscripts
drwxr-xr-x  4 root root    4096 2012-04-16 12:10 cloud
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 console-setup
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.d
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 cron.daily
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.hourly
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.monthly
-rw-r--r--  1 root root     724 2011-09-19 17:04 crontab
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 cron.weekly
drwxr-xr-x  4 root root    4096 2012-04-16 10:21 dbus-1
-rw-r--r--  1 root root    2969 2011-06-23 00:01 debconf.conf
-rw-r--r--  1 root root      11 2011-07-08 10:13 debian_version
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 default
```

```
% ls c <tab>
```

```
% ls c <tab><tab>
```

# Are you typing? You're doing it wrong.

## Tab-completion:

- Tab once to complete uniquely
- Tab twice to see all possible completions

## Up-arrow:

- Previous commands can be found by pressing "up-arrow"

## 'history'

```
root@ubuntu:~# cd /etc/
root@ubuntu:/etc# ls -la
total 788
drwxr-xr-x 82 root root    4096 2012-06-25 01:56 .
drwxr-xr-x 23 root root    4096 2012-04-18 09:17 ..
-rw-r--r--  1 root root    2981 2012-04-16 10:09 adduser.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 alternatives
drwxr-xr-x  3 root root    4096 2012-04-16 10:19 apm
drwxr-xr-x  3 root root    4096 2012-04-16 10:21 apparmor
drwxr-xr-x  8 root root    4096 2012-04-16 12:09 apparmor.d
drwxr-xr-x  3 root root    4096 2012-04-16 12:09 apport
drwxr-xr-x  6 root root    4096 2012-04-16 12:08 apt
-rw-r-----  1 root daemon   144 2011-05-16 03:32 at.deny
-rw-r--r--  1 root root    1975 2011-05-18 03:00 bash.bashrc
-rw-r--r--  1 root root   58753 2011-10-04 13:53 bash_completion
drwxr-xr-x  3 root root   12288 2012-04-16 12:09 bash_completion.d
-rw-r--r--  1 root root     344 2011-10-04 12:56 bindresvport.blacklist
-rw-r--r--  1 root root     321 2011-08-09 09:16 blkid.conf
lrwxrwxrwx  1 root root      15 2011-08-09 09:16 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 byobu
drwxr-xr-x  3 root root    4096 2012-04-16 10:18 ca-certificates
-rw-r--r--  1 root root    7014 2012-04-16 10:21 ca-certificates.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 calendar
drwxr-s---  2 root dip     4096 2012-04-16 10:22 chatscripts
drwxr-xr-x  4 root root    4096 2012-04-16 12:10 cloud
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 console-setup
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.d
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 cron.daily
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.hourly
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.monthly
-rw-r--r--  1 root root     724 2011-09-19 17:04 crontab
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 cron.weekly
drwxr-xr-x  4 root root    4096 2012-04-16 10:21 dbus-1
-rw-r--r--  1 root root    2969 2011-06-23 00:01 debconf.conf
-rw-r--r--  1 root root      11 2011-07-08 10:13 debian_version
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 default
```

```
% cd /etc
% ls c <tab>
```

```
% pwd
% ls c <tab><tab>
```

# Three variants to `ls`

| `ls -l` | `ls -la` | `ls -lh` |
| --- | --- | --- |
| provides a *long* listing | includes *all* files, even hidden files | displays file sizes in *human* readable numbers |

# Four ways to view a text file

| `more` | `head` | `tail` | `cat` |
|---|---|---|---|
| view a text file one screen full at a time | view the top 15 lines of a file | view the last 15 lines of a file | spit the whole file at once |
| space-bar: scroll<br>q: quit | -n num controls the number of lines | -n num controls the number of lines | |

# Explore the file hierarchy

1. Move to the directory /etc
    - What is the first line of the files '`hosts`' in the directory `/etc`?
    - What is the relative file path to get to `/var/log` from here?
    - What is the absolute path?
2. Move to the directory `/var/log/`
    - What is the contents on line 73 of the dmesg file?
    - Without changing directories, what is the second line of the `cpuinfo` file in the `proc` directory?
        - What is the command to read this file with a relative path?
        - An absolute path?
3. Move back to the root, what directories do you see?
4. Move back home, what are three ways to move home from the root?

# Download example files using `wget`

Return to the directory in your home called ‘shell’.

TSV file:
`http://creskolab.uoregon.edu/cesky/batch_1.genotypes_1.loc.gz`

FASTQ file:
`http://creskolab.uoregon.edu/cesky/s_1_sequence.txt.gz`

Tar Archive:
`http://creskolab.uoregon.edu/cesky/samples.tar.gz`

# What is a tar archive?


tar = tape archive

## Compress / Decompress

`gzip` / `gunzip`

batch_1.genotypes_1.loc.gz

s_1_sequence.txt.gz

## Gzipped Tar archive

`tar xvfz`

samples.tar.gz

## Tar archive

`tar xvf`

samples.tar

# The FASTQ File Format

## FASTA

```
>HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
ACATACTGCAGGGTCAGGAGGATTATCTCCTCTGCAAGGTAACGCCTGCTGTAACCGTTGTTCTTCATCCTTTT
CCTAACTGCAGGGCTGTCTTGTCAGGTCTGACAAGACATATGCAGGGCTCAATTTGAGATAATTGCTCAATATA
```

## FASTQ

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# Quality Scores

```
  SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....................................
  ........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....................
  ............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....................
  ..............................JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ.....................
 LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....................................
  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
  |                          |    |        |                                 |                       |
 33                         59   64       73                               104                     126

 S - Sanger        Phred+33,  raw reads typically (0, 40)
 X - Solexa        Solexa+64, raw reads typically (-5, 40)
 I - Illumina 1.3+ Phred+64,  raw reads typically (0, 40)
 J - Illumina 1.5+ Phred+64,  raw reads typically (3, 40)
    with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)
    (Note: See discussion above).
 L - Illumina 1.8+ Phred+33,  raw reads typically (0, 41)
```

# ASCII values 64 - 104 = 0 - 40

http://en.wikipedia.org/wiki/FASTQ_format

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# Quality Scores

```
  SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....................................
  ........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....................
  ............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....................
  ..............................JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ.....................
 LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....................................
  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
  |                         |    |        |                                     |                     |
 33                        59   64       73                                   104                   126

 S - Sanger        Phred+33,  raw reads typically (0, 40)
 X - Solexa        Solexa+64, raw reads typically (-5, 40)
 I - Illumina 1.3+ Phred+64,  raw reads typically (0, 40)
 J - Illumina 1.5+ Phred+64,  raw reads typically (3, 40)
    with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)
    (Note: See discussion above).
 L - Illumina 1.8+ Phred+33,  raw reads typically (0, 41)
```

ASCII values 64 - 104 = 0 - 40

‘F’ = 70

http://en.wikipedia.org/wiki/FASTQ_format

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# Quality Scores

```
  SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....................................
  ........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....................
  ............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....................
  ..............................JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ.....................
 LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....................................
  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
  |                          |    |        |                                 |                     |
 33                         59   64       73                               104                   126

 S - Sanger        Phred+33,  raw reads typically (0, 40)
 X - Solexa        Solexa+64, raw reads typically (-5, 40)
 I - Illumina 1.3+ Phred+64,  raw reads typically (0, 40)
 J - Illumina 1.5+ Phred+64,  raw reads typically (3, 40)
    with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)
    (Note: See discussion above).
 L - Illumina 1.8+ Phred+33,  raw reads typically (0, 41)
```

## ASCII values 64 - 104 = 0 - 40

## ‘F’ = 70 70 - 64 = **14**

http://en.wikipedia.org/wiki/FASTQ_format

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

| Quality Score | Probability of incorrect base call | Base call accuracy |
|---|---|---|
| 10 | 1 in 10 | 90% |
| 20 | 1 in 100 | 99% |
| 30 | 1 in 1000 | 99.9% |
| 40 | 1 in 10000 | 99.99% |

http://en.wikipedia.org/wiki/Phred_quality_score

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

**70 - 64 = 14**

| Quality Score | Probability of incorrect base call | Base call accuracy |
|---|---|---|
| 10 | 1 in 10 | 90% |
| 20 | 1 in 100 | 99% |
| 30 | 1 in 1000 | 99.9% |
| 40 | 1 in 10000 | 99.99% |

http://en.wikipedia.org/wiki/Phred_quality_score

## Count raw reads:

```
wc -l s_1_sequence.txt

grep "@" s_1_sequence.txt
grep -c "@" s_1_sequence.txt

grep -v "@" s_1_sequence.txt
grep -v -c "@" s_1_sequence.txt
```

## Count reads with barcode:

```
grep -c "^CGATA" s_1_sequence.txt
```

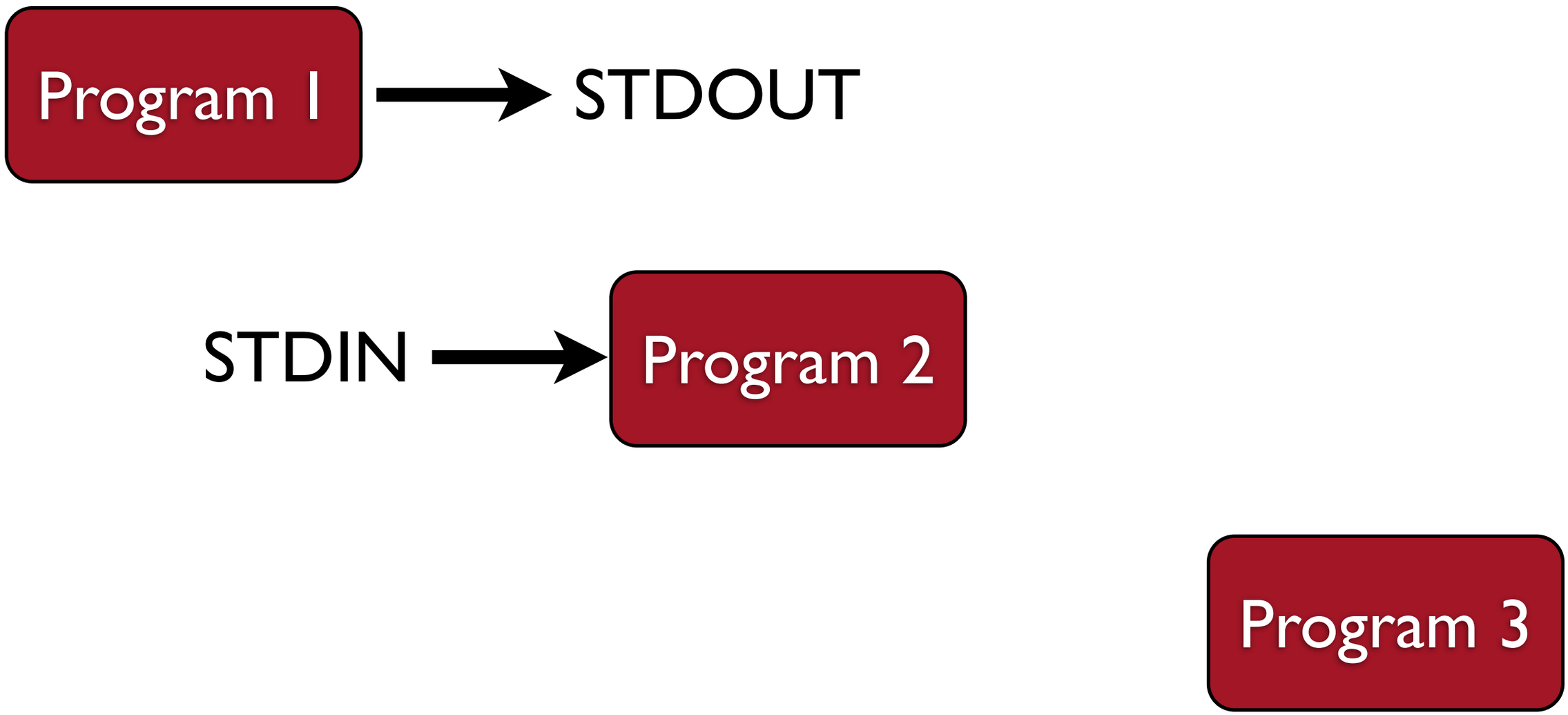
# Special Files

STDIN, STDOUT, STDERR

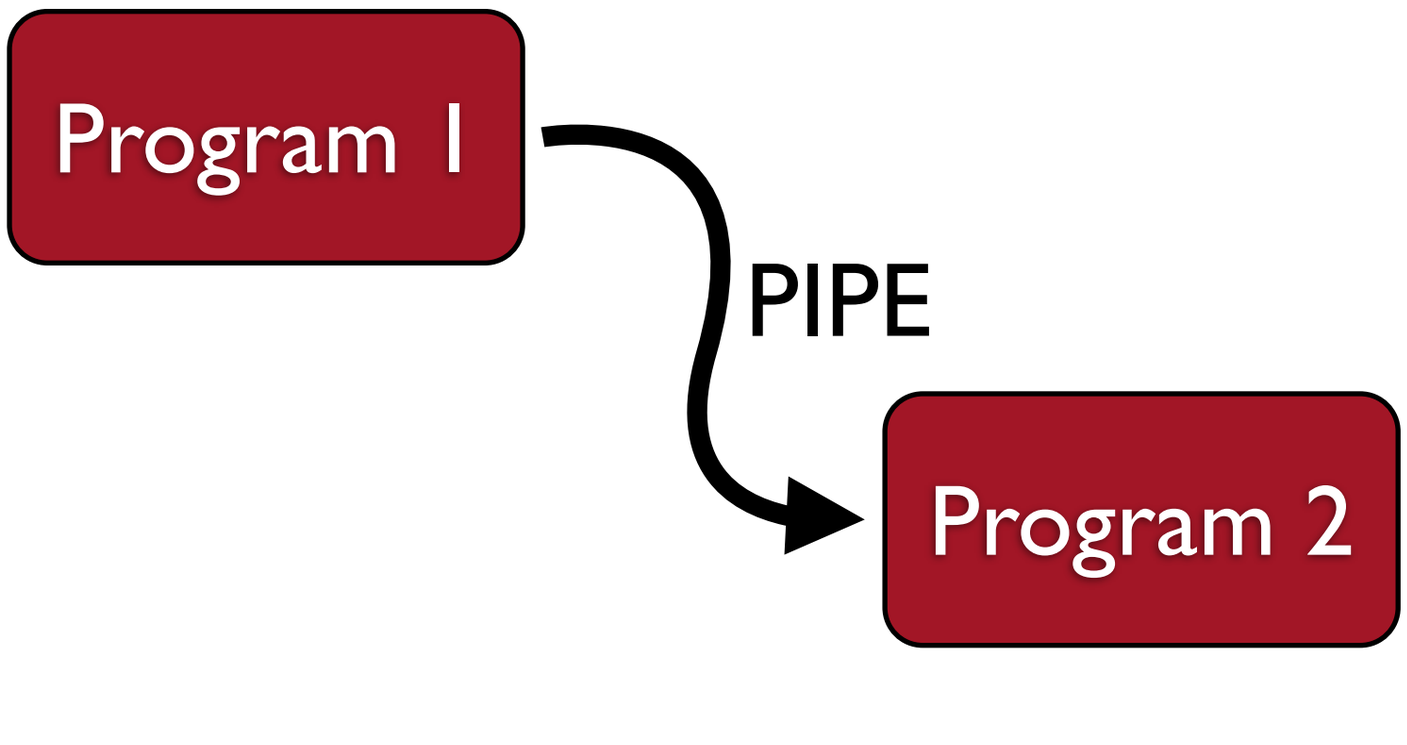
The Shell's Killer App: **Pipes**

# The Shell's Killer App: **Pipes**, ctd.

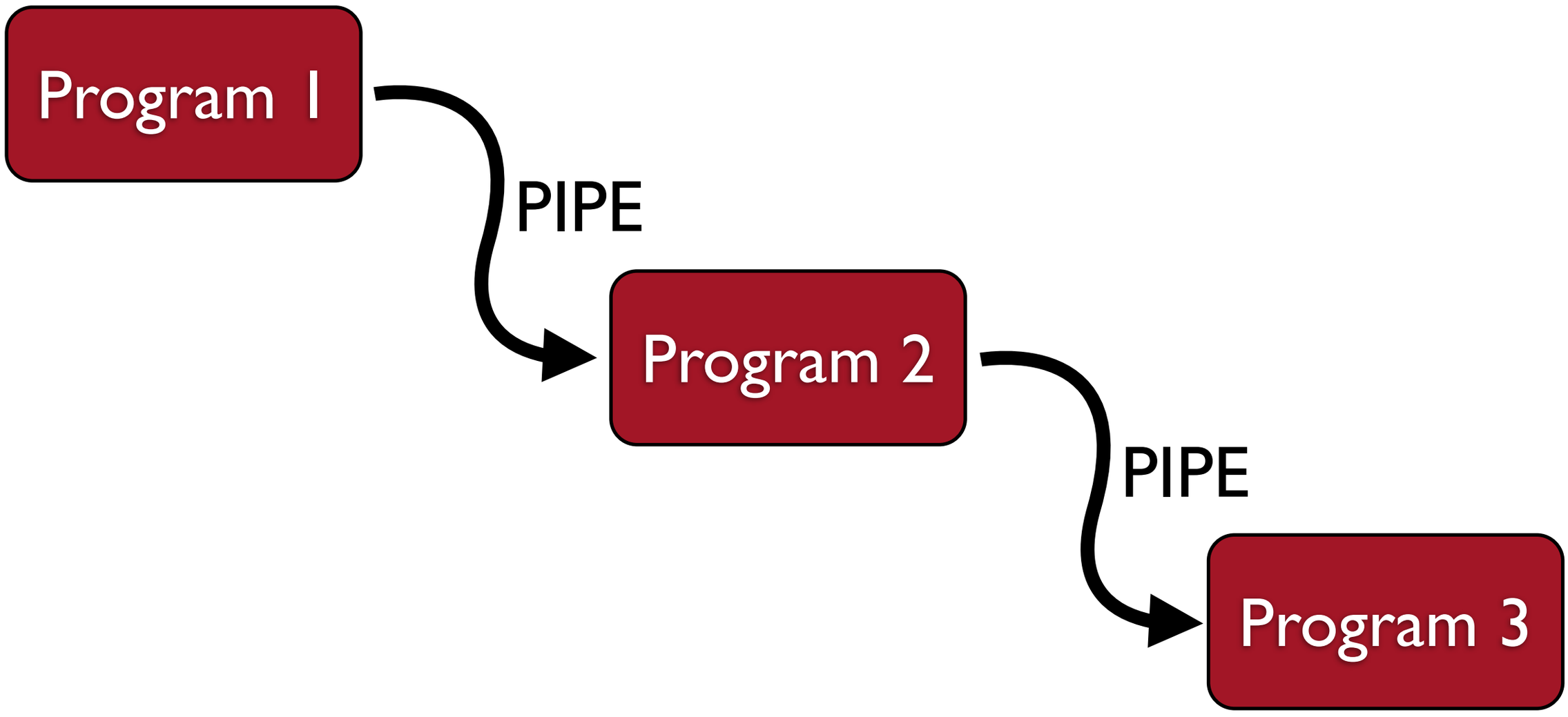
# The Shell's Killer App: **Pipes**, ctd.



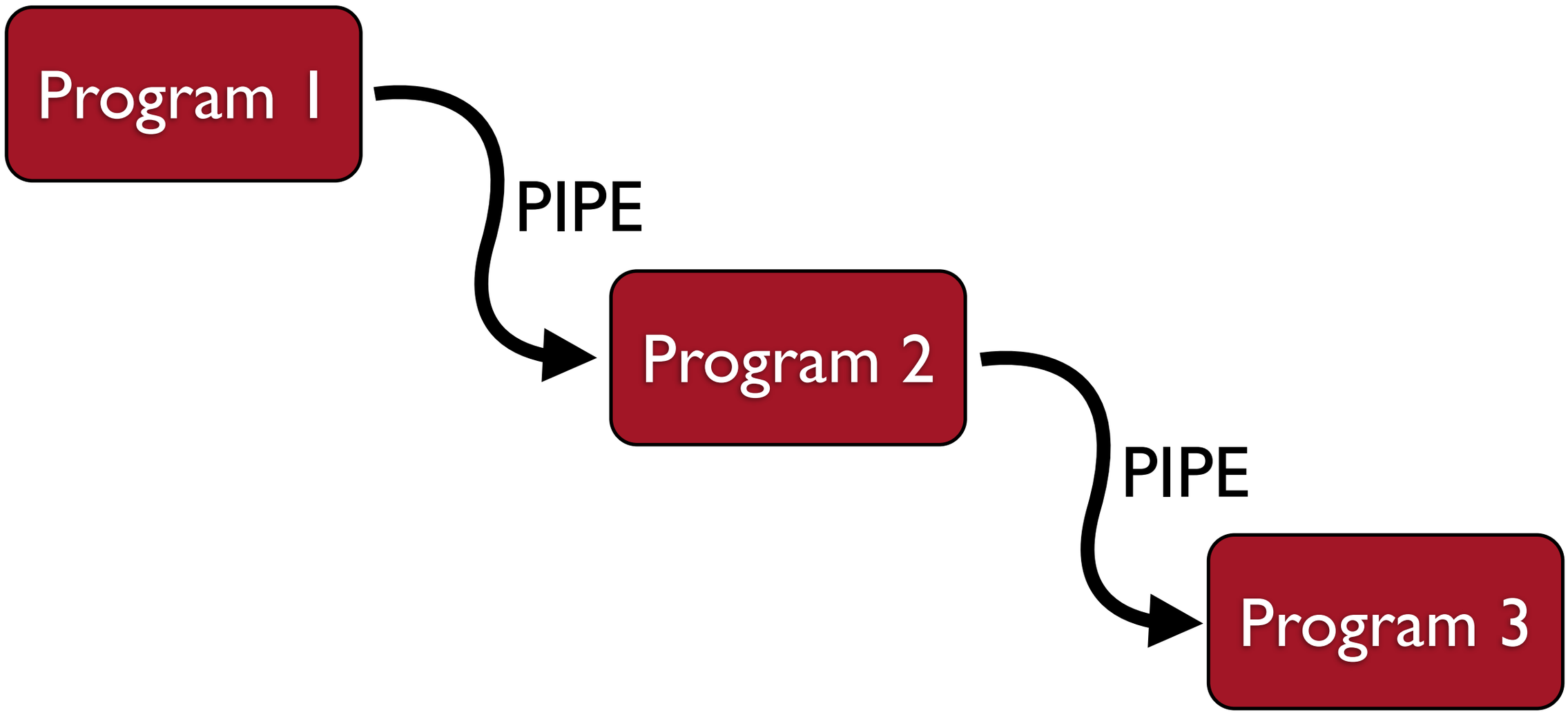
# The Shell's Killer App: **Pipes**, ctd.




# The Shell's Killer App: **Pipes**, ctd.



# The Shell's Killer App: **Pipes**, ctd.




So what is the purpose of the program **`cat`**?

## cut

```
cut -f 10 batch_1.genotypes_1.loc
```

## cut, capture the output

```
cut -f 1-10 batch_1.genotypes_1.loc > genos
```

## cut, pipe the output to grep

```
cut -f 2 batch_1.genotypes_1.loc | grep -c "nnxnp"
```

```
cut -f 1-10,15,17 batch_1.genotypes_1.loc | grep "nnxnp" > genos2
```

## Examine a marker, translating the output

```
cat batch_1.genotypes_1.loc | tr "     " "," | grep "^96053"
```

# s_1_sequence.txt.gz

1. Decompress the file
2. Count the number of raw reads (250,000)
3. Count the number of reads with barcode `CGATA` (19,501)
4. Capture all FASTQ records for `ACCAT` into a file called sample_01.fq (you should get 18352 records, 73408 lines)
5. Determine the count of all barcodes in the file

```
286 CTAGT
7900 TCAGA
10659 ACTGC
10931 TGACC
11536 GAGAT
11871 CTGAA
14409 CGGCG
14508 TGGTT
18226 GAAGC
18352 ACCAT
18375 TCGAG
19501 CGATA
23012 AATTT
26336 GCATT
31136 CTAGG
```

`ls`
`gunzip`
`man`
`more`
`cat`
`wc`
`head`
`cut`
`grep`
`sort`
`uniq`
`>`
`|`

1. Use **`head`** when building a command, **`cat`** once the command is working
2. Look at the **-n** option for the **head** command, the **-l** option for **wc**
3. The "^" character means "must occur at beginning of line" in a grep search
4. Look at the **grep** options: **-c**, **-v, -A**, **-B**
5. Read the man pages for **sort** and **uniq** to learn how to combine them

## cut

```
cut -f 10 batch_1.genotypes_1.loc
```

## cut, capture the output

```
cut -f 1-10 batch_1.genotypes_1.loc > genos
```

## cut, pipe the output to grep

```
cut -f 2 batch_1.genotypes_1.loc | grep -c "nnxnp"
```

```
cut -f 1-10,15,17 batch_1.genotypes_1.loc | grep "nnxnp" > genos2
```

## Examine a marker, translating the output

```
cat batch_1.genotypes_1.loc | tr "    " "," | grep "^96053"
```

Count raw reads:

```
wc -l s_1_sequence.txt

grep "@" s_1_sequence.txt
grep -c "@" s_1_sequence.txt

grep -v "@" s_1_sequence.txt
grep -v -c "@" s_1_sequence.txt
```

Count reads with barcode:

```
grep -c "^CGATA" s_1_sequence.txt
```