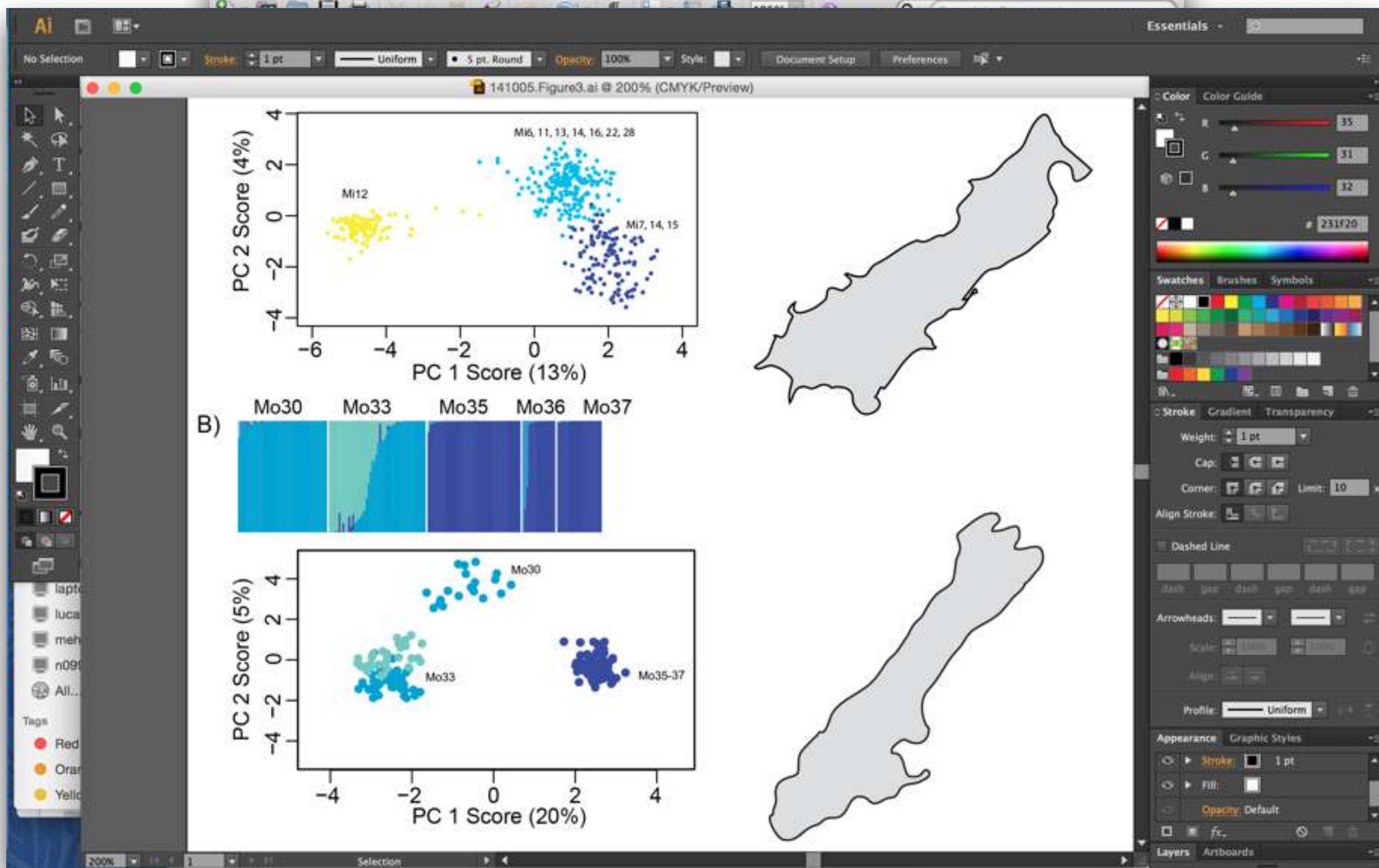


digital™

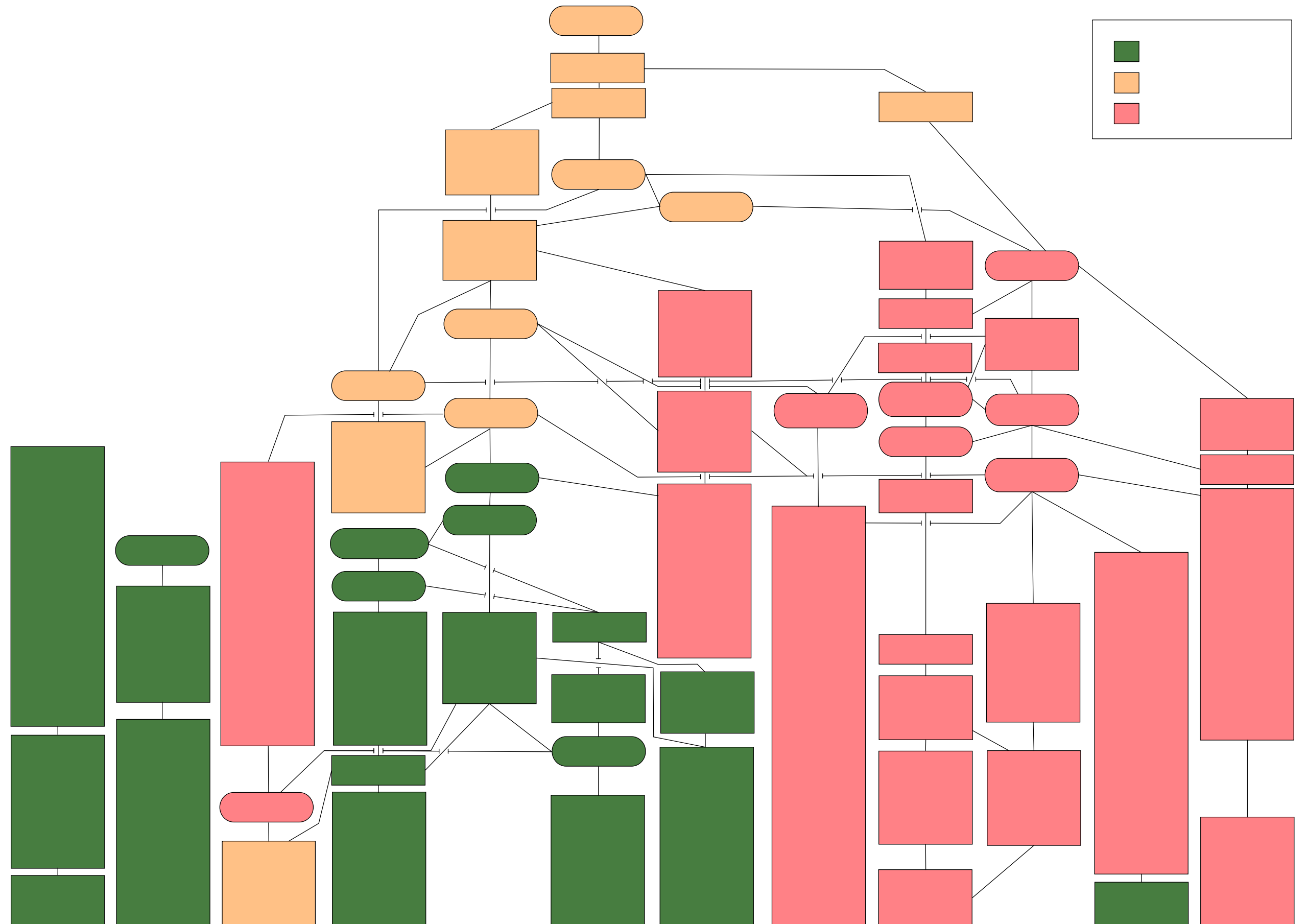
LIVE FREE OR DIE

UNIX

Registered TM of UNIX System Laboratories,
A wholly owned subsidiary of Novell, Inc.



Unix History



What computers can run Unix?



Apple OS X Macs



Wireless internet routers

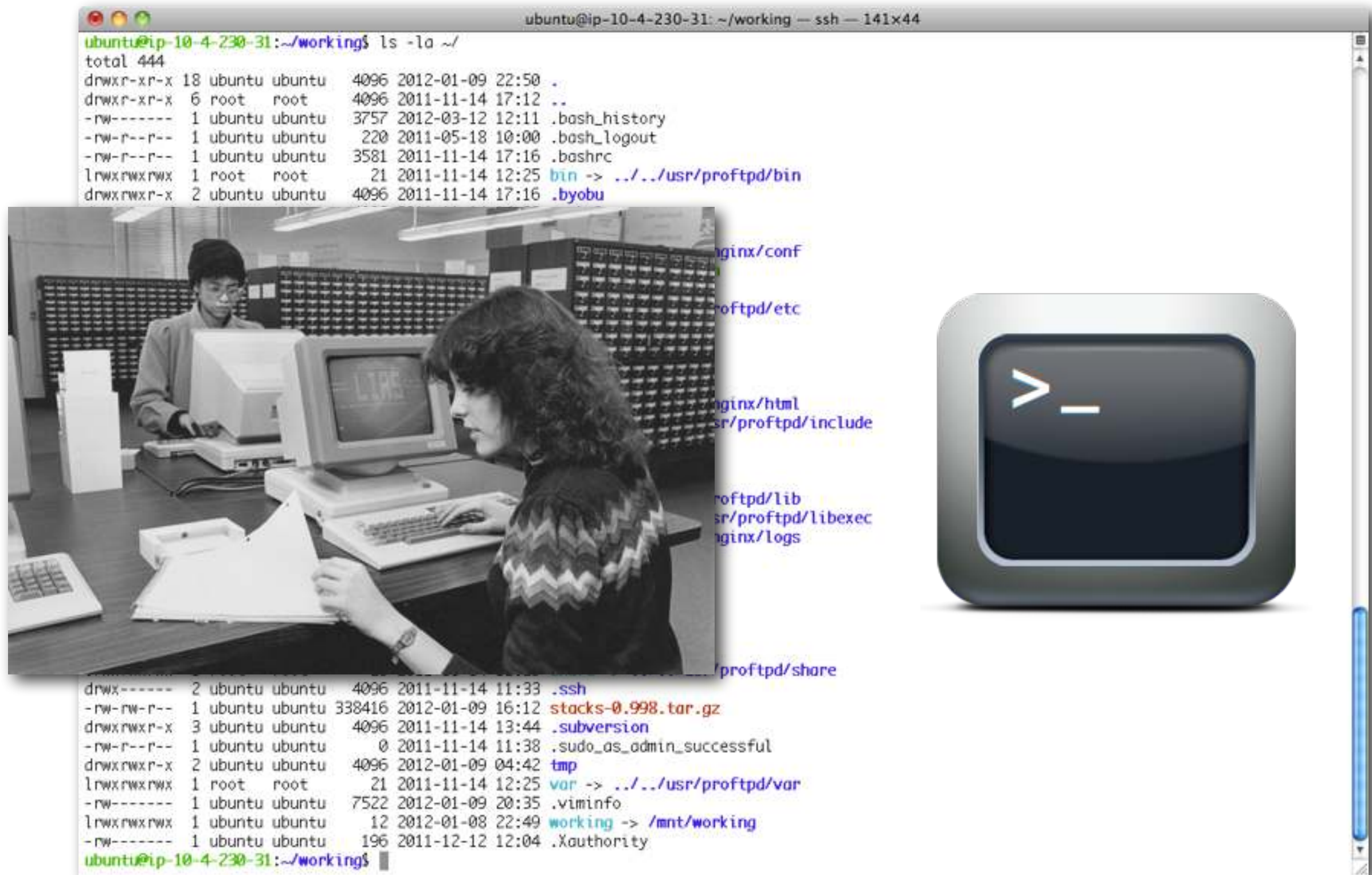


Google's Android phones



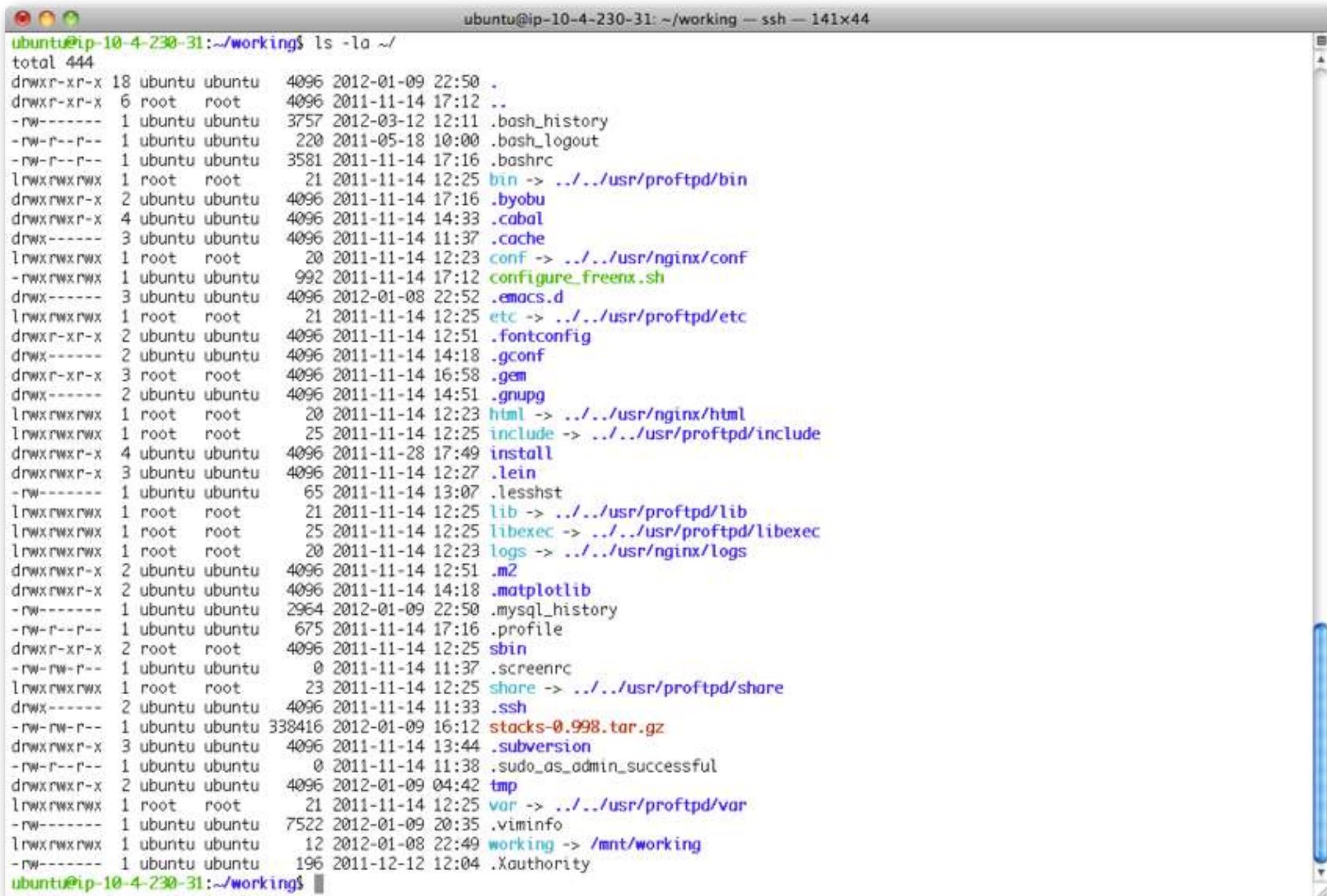
Airplane
entertainment
systems

The Terminal Window



the shell, the prompt, the command line

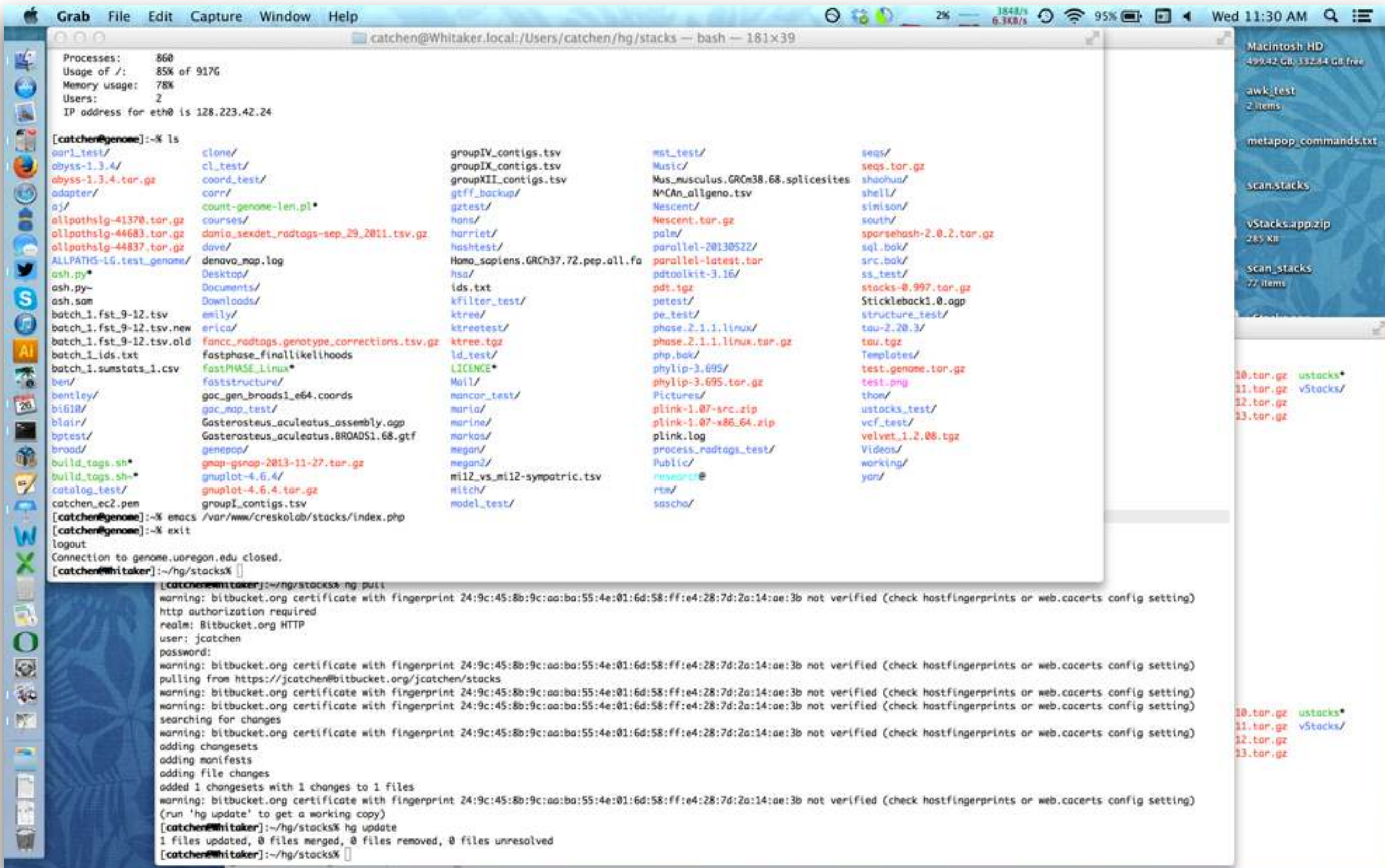
The Terminal Window

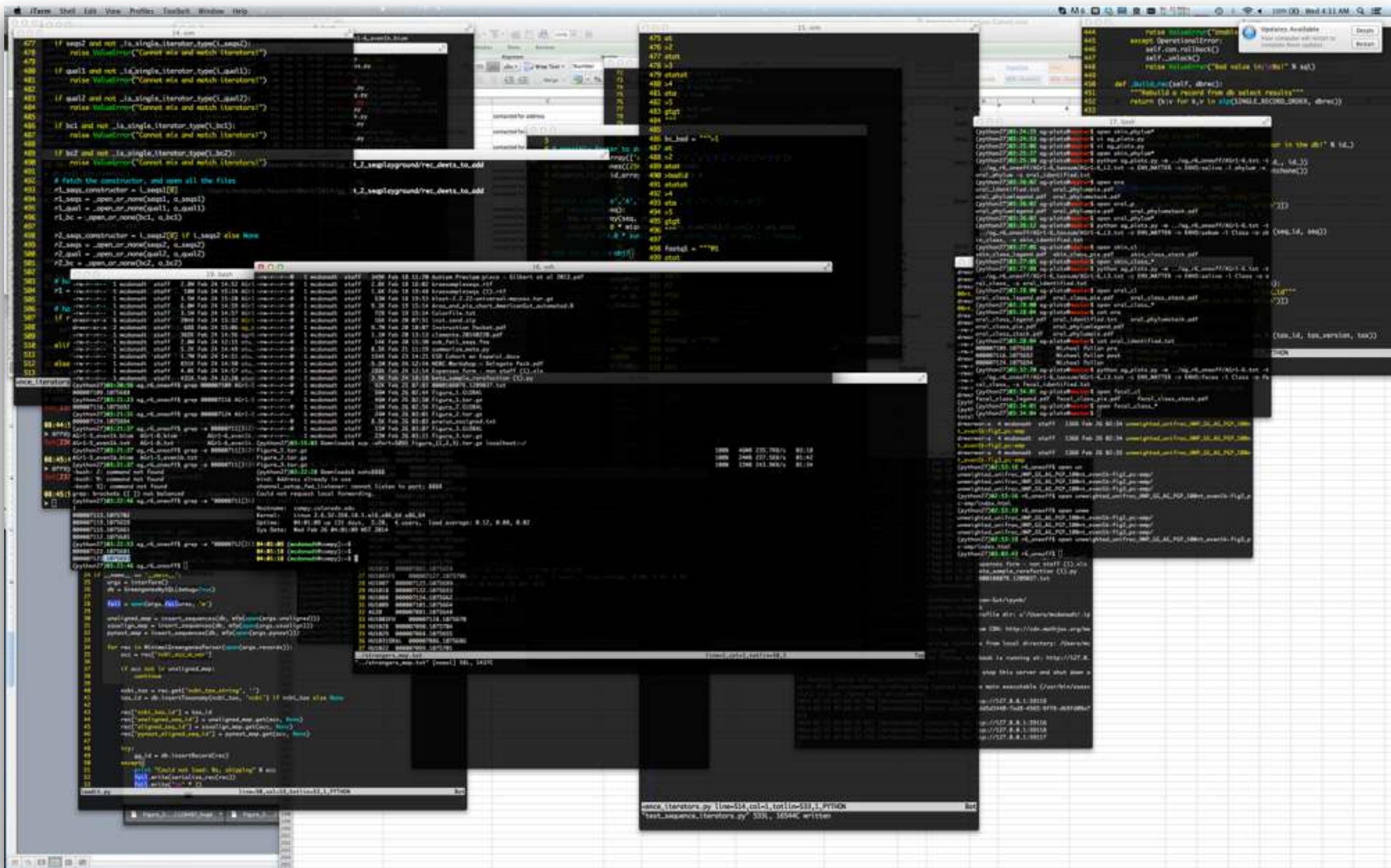


```
ubuntu@ip-10-4-230-31: ~/working — ssh — 141x44
ubuntu@ip-10-4-230-31:~/working$ ls -la ~/
total 444
drwxr-xr-x 18 ubuntu ubuntu 4096 2012-01-09 22:50 .
drwxr-xr-x  6 root    root   4096 2011-11-14 17:12 ..
-rw-r----- 1 ubuntu ubuntu 3757 2012-03-12 12:11 .bash_history
-rw-r--r--  1 ubuntu ubuntu  220 2011-05-18 10:00 .bash_logout
-rw-r--r--  1 ubuntu ubuntu 3581 2011-11-14 17:16 .bashrc
lrwxrwxrwx  1 root    root    21 2011-11-14 12:25 bin -> ../../usr/proftpd/bin
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 17:16 .byobu
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-14 14:33 .cabal
drwx----- 3 ubuntu ubuntu 4096 2011-11-14 11:37 .cache
lrwxrwxrwx  1 root    root    20 2011-11-14 12:23 conf -> ../../usr/nginx/conf
-rwxrwxrwx  1 ubuntu ubuntu  992 2011-11-14 17:12 configure_freenx.sh
drwx----- 3 ubuntu ubuntu 4096 2012-01-08 22:52 .emacs.d
lrwxrwxrwx  1 root    root    21 2011-11-14 12:25 etc -> ../../usr/proftpd/etc
drwxr-xr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .fontconfig
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 14:18 .gconf
drwxr-xr-x  3 root    root   4096 2011-11-14 16:58 .gem
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 14:51 .gnupg
lrwxrwxrwx  1 root    root    20 2011-11-14 12:23 html -> ../../usr/nginx/html
lrwxrwxrwx  1 root    root    25 2011-11-14 12:25 include -> ../../usr/proftpd/include
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-28 17:49 install
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 12:27 .lein
-rw-r----- 1 ubuntu ubuntu  65 2011-11-14 13:07 .lessht
lrwxrwxrwx  1 root    root    21 2011-11-14 12:25 lib -> ../../usr/proftpd/lib
lrwxrwxrwx  1 root    root    25 2011-11-14 12:25 libexec -> ../../usr/proftpd/libexec
lrwxrwxrwx  1 root    root    20 2011-11-14 12:23 logs -> ../../usr/nginx/logs
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .m2
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 14:18 .matplotlib
-rw-r----- 1 ubuntu ubuntu 2964 2012-01-09 22:50 .mysql_history
-rw-r--r--  1 ubuntu ubuntu  675 2011-11-14 17:16 .profile
drwxr-xr-x  2 root    root   4096 2011-11-14 12:25 sbin
-rw-rw-r--  1 ubuntu ubuntu  0 2011-11-14 11:37 .screenrc
lrwxrwxrwx  1 root    root    23 2011-11-14 12:25 share -> ../../usr/proftpd/share
drwx----- 2 ubuntu ubuntu 4096 2011-11-14 11:33 .ssh
-rw-rw-r--  1 ubuntu ubuntu 338416 2012-01-09 16:12 stacks-0.998.tar.gz
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 13:44 .subversion
-rw-r--r--  1 ubuntu ubuntu  0 2011-11-14 11:38 .sudo_as_admin_successful
drwxrwxr-x  2 ubuntu ubuntu 4096 2012-01-09 04:42 tmp
lrwxrwxrwx  1 root    root    21 2011-11-14 12:25 var -> ../../usr/proftpd/var
-rw-r----- 1 ubuntu ubuntu 7522 2012-01-09 20:35 .viminfo
lrwxrwxrwx  1 ubuntu ubuntu  12 2012-01-08 22:49 working -> /mnt/working
-rw-r----- 1 ubuntu ubuntu  196 2011-12-12 12:04 .Xauthority
ubuntu@ip-10-4-230-31:~/working$
```

Make it comfortable to work in:

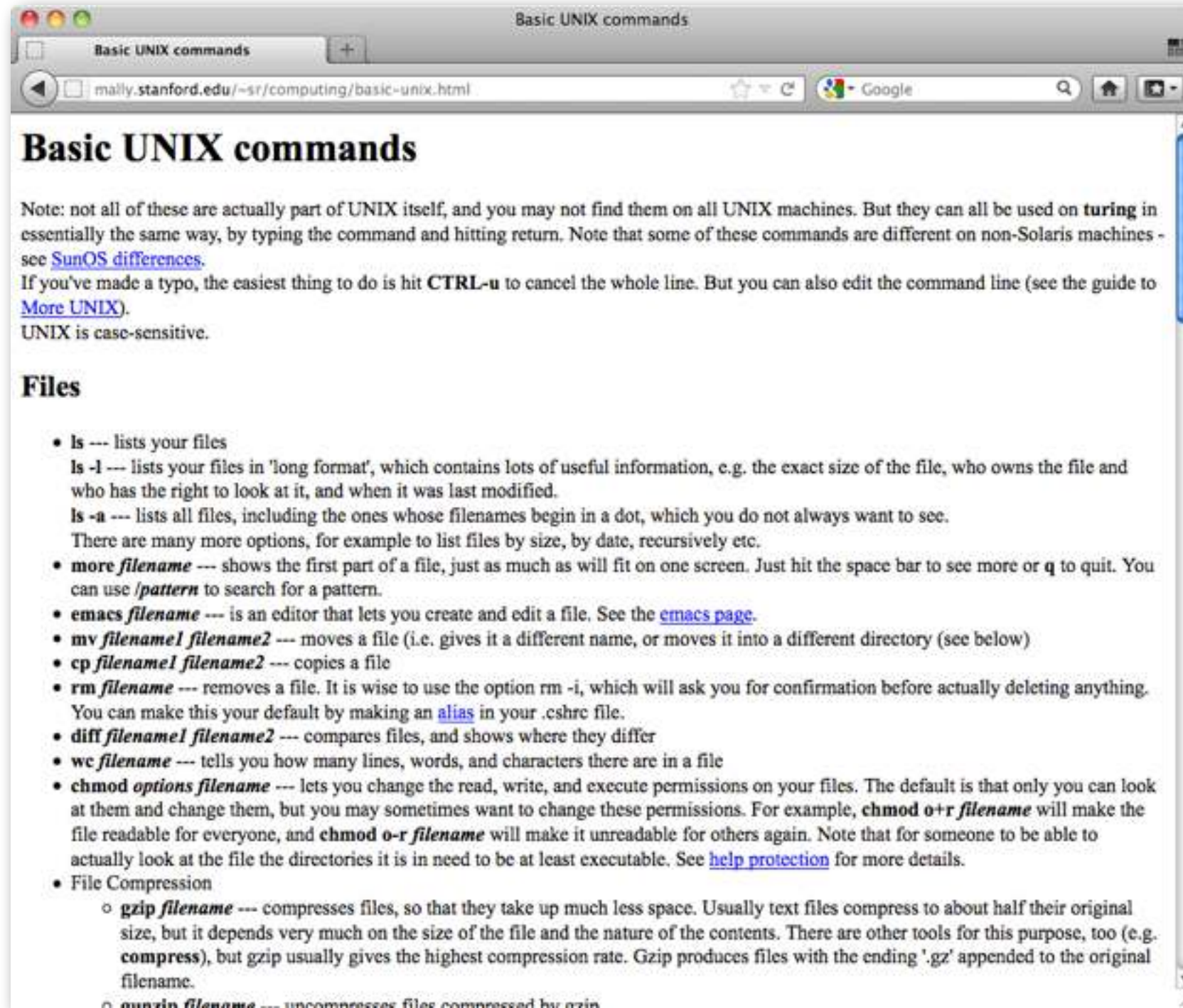
- Resize the window
- Change your font size
- Open multiple terminal windows





Obtain a cheat sheet

google “unix commands”



In UNIX everything is a file organized in a hierarchy

Paths

`/home/catchen/working`

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

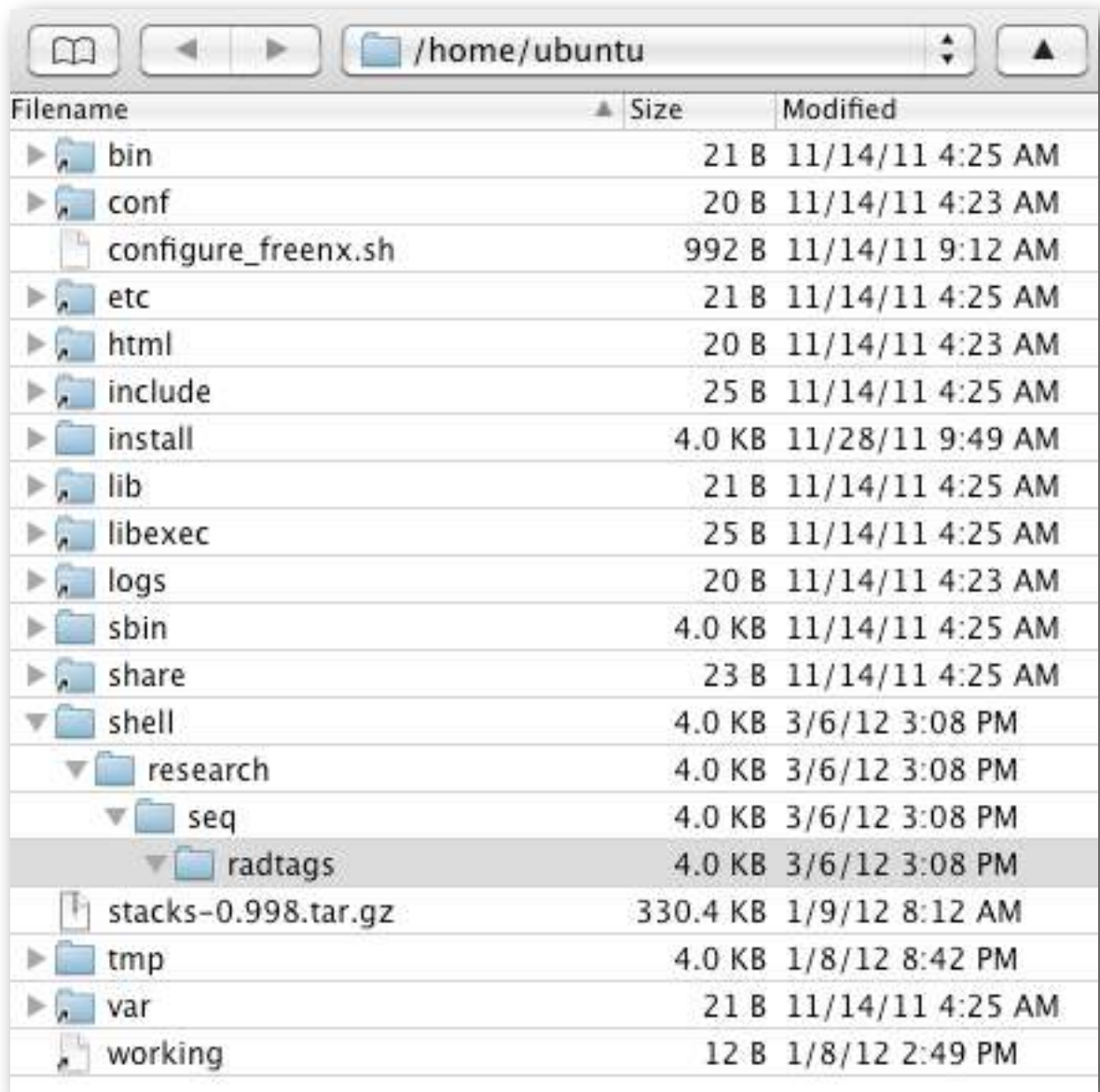
% ls

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
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

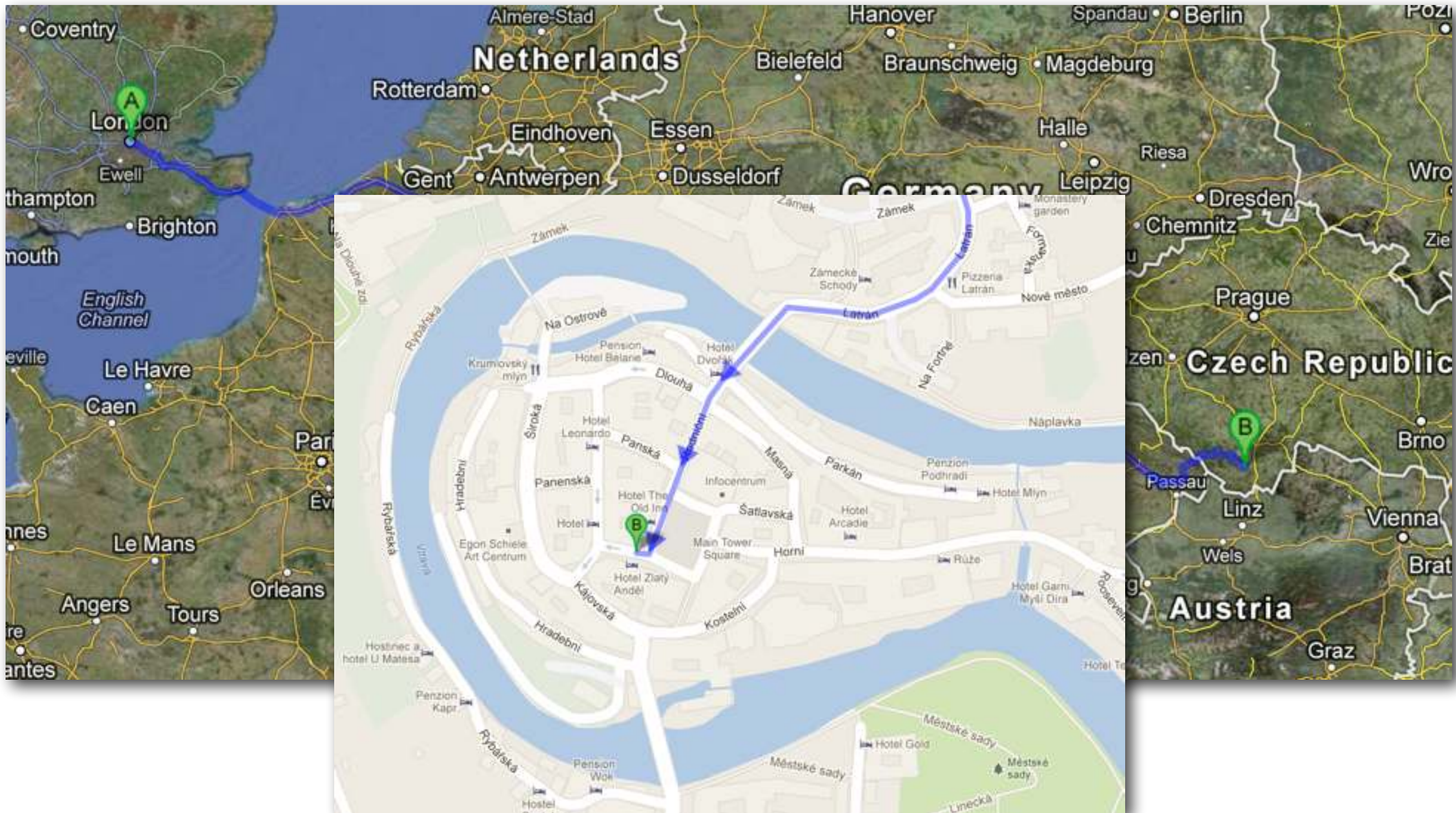


Filename	Size	Modified
bin	21 B	11/14/11 4:25 AM
conf	20 B	11/14/11 4:23 AM
configure_freenx.sh	992 B	11/14/11 9:12 AM
etc	21 B	11/14/11 4:25 AM
html	20 B	11/14/11 4:23 AM
include	25 B	11/14/11 4:25 AM
install	4.0 KB	11/28/11 9:49 AM
lib	21 B	11/14/11 4:25 AM
libexec	25 B	11/14/11 4:25 AM
logs	20 B	11/14/11 4:23 AM
sbin	4.0 KB	11/14/11 4:25 AM
share	23 B	11/14/11 4:25 AM
shell	4.0 KB	3/6/12 3:08 PM
research	4.0 KB	3/6/12 3:08 PM
seq	4.0 KB	3/6/12 3:08 PM
radtags	4.0 KB	3/6/12 3:08 PM
stacks-0.998.tar.gz	330.4 KB	1/9/12 8:12 AM
tmp	4.0 KB	1/8/12 8:42 PM
var	21 B	11/14/11 4:25 AM
working	12 B	1/8/12 2:49 PM

And the radtags directory is uniquely identified by its path:
/home/ubuntu/shell/research/seq/radtags

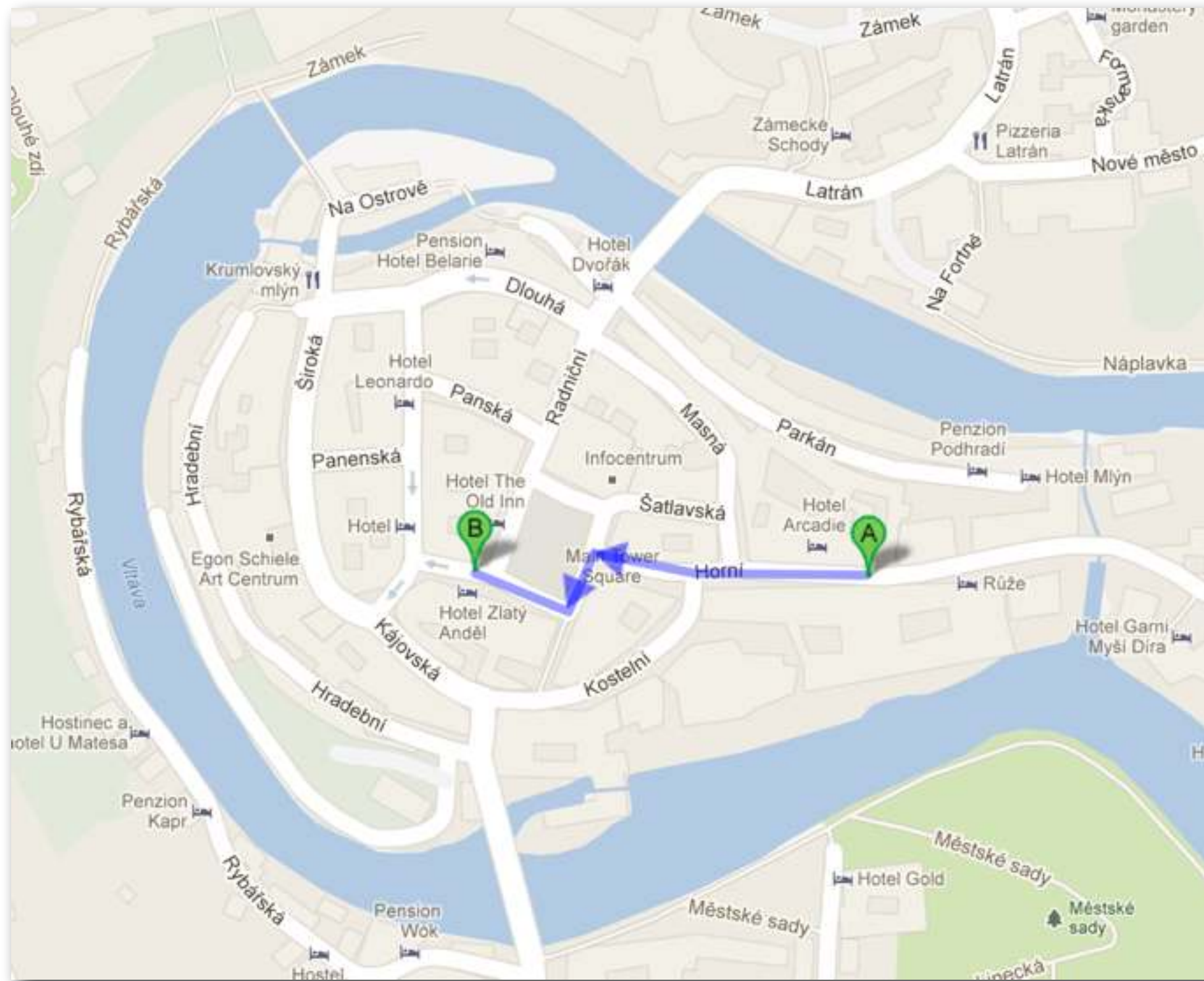
Absolute and relative paths

How do I get to the Hotel Zlatý Anděl?

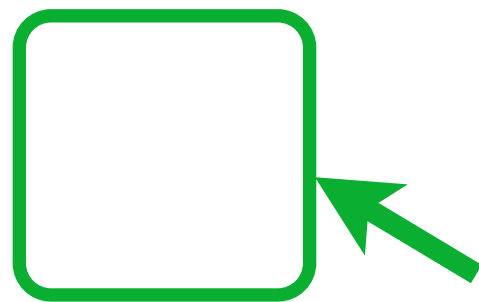


Absolute and relative paths

How do I get to the Hotel Zlatý Anděl?



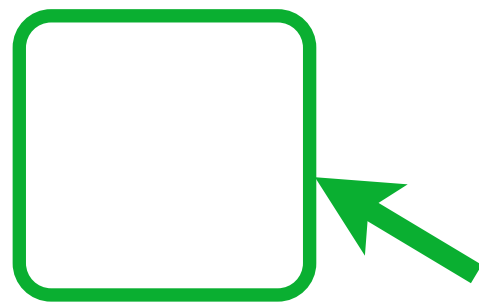
Paths



Absolute Path

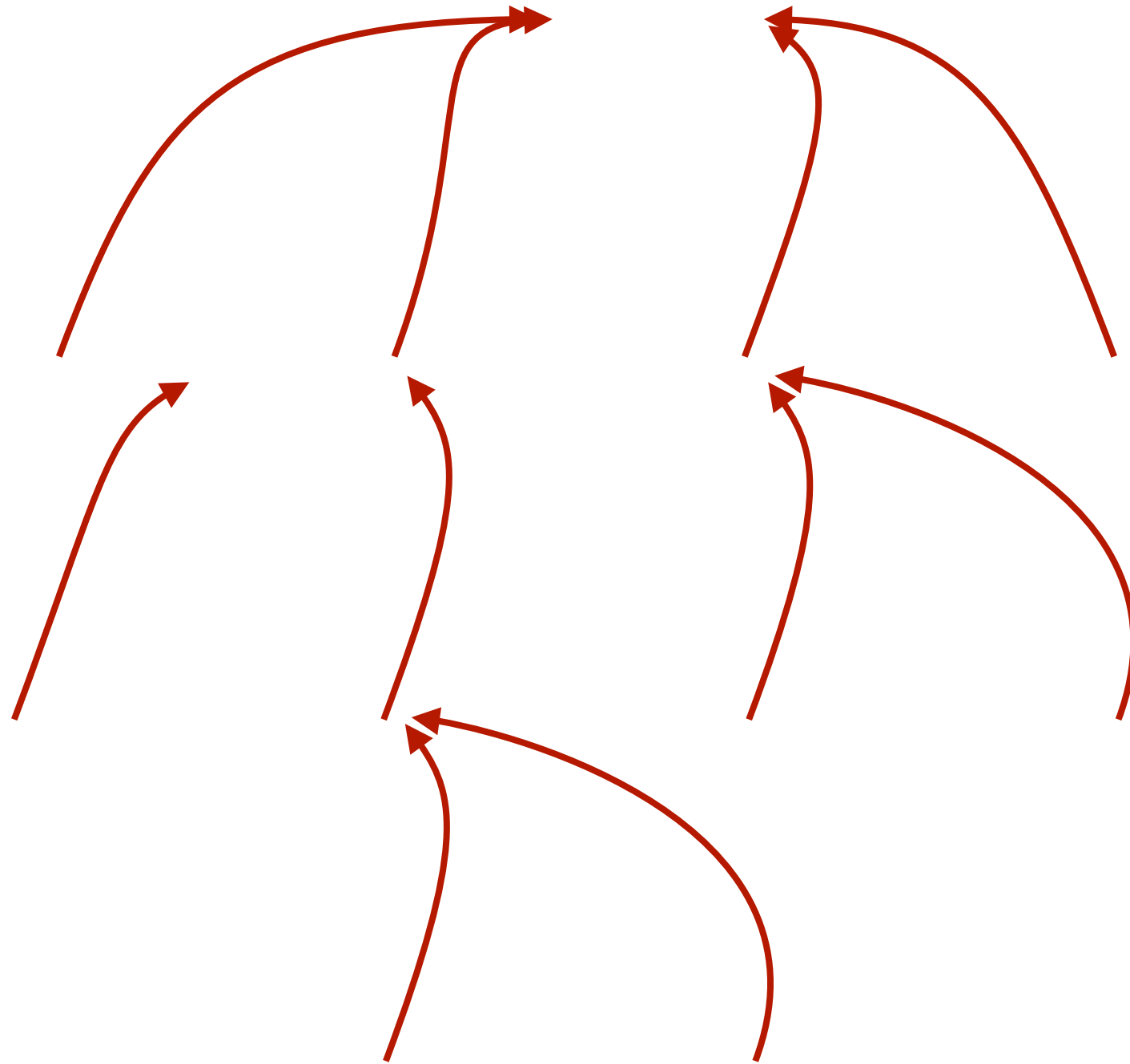
`/home/catchen/working`

Relative Path?

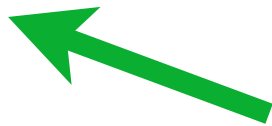


Special files -- '*dot*'

Special files -- '*dot dot*'



Relative Path



`../working`

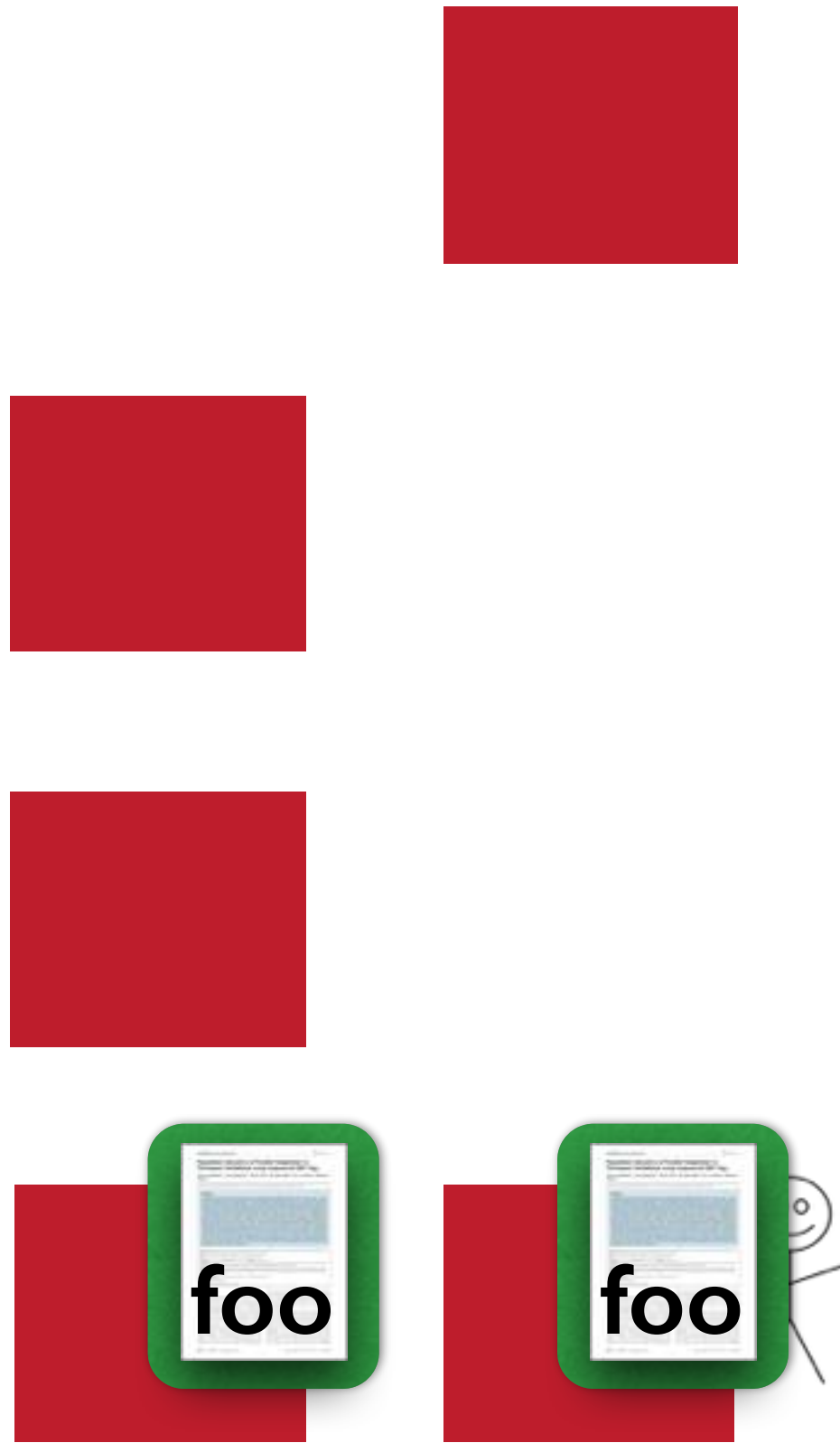


Absolute Path:

`/home/catchen/working/foo`

Relative Path:

`../working/foo`



Absolute Path: `/home/catchen/working/foo`

Absolute Path: `/home/catchen/research/foo`



Relative Path: `../working/foo`

Relative Path: `./foo`

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.

```
ubuntu@ip-10-140-6-74: ~ — ssh — 133x24
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    ntfsnnp  rmdir     umount
bzipcat   dbus-cleanup-sockets gzexe    mknod    ntfsdecrypt  nano      uname
bzip      dbus-daemon  gzip     mktemp   ntfsdump_logfile  run-parts  uncompress
bzdiff    dbus-uuidgen  hostname  more     ntfsfix    rzsh      unicode_start
bzegrep   dd        init-checkconf  mountpoint  ntfsinfo  sed        vdir
bzexe     df         initctl2dot    mount      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
bzip      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          ntfsckat   rbash      touch
```

% ls /bin

Relative and absolute paths

A shortcut to your 'home', tilde:

~

Moving through the filesystem:

cd

Knowing where you are:

pwd

% ls ~/

% cd ~/

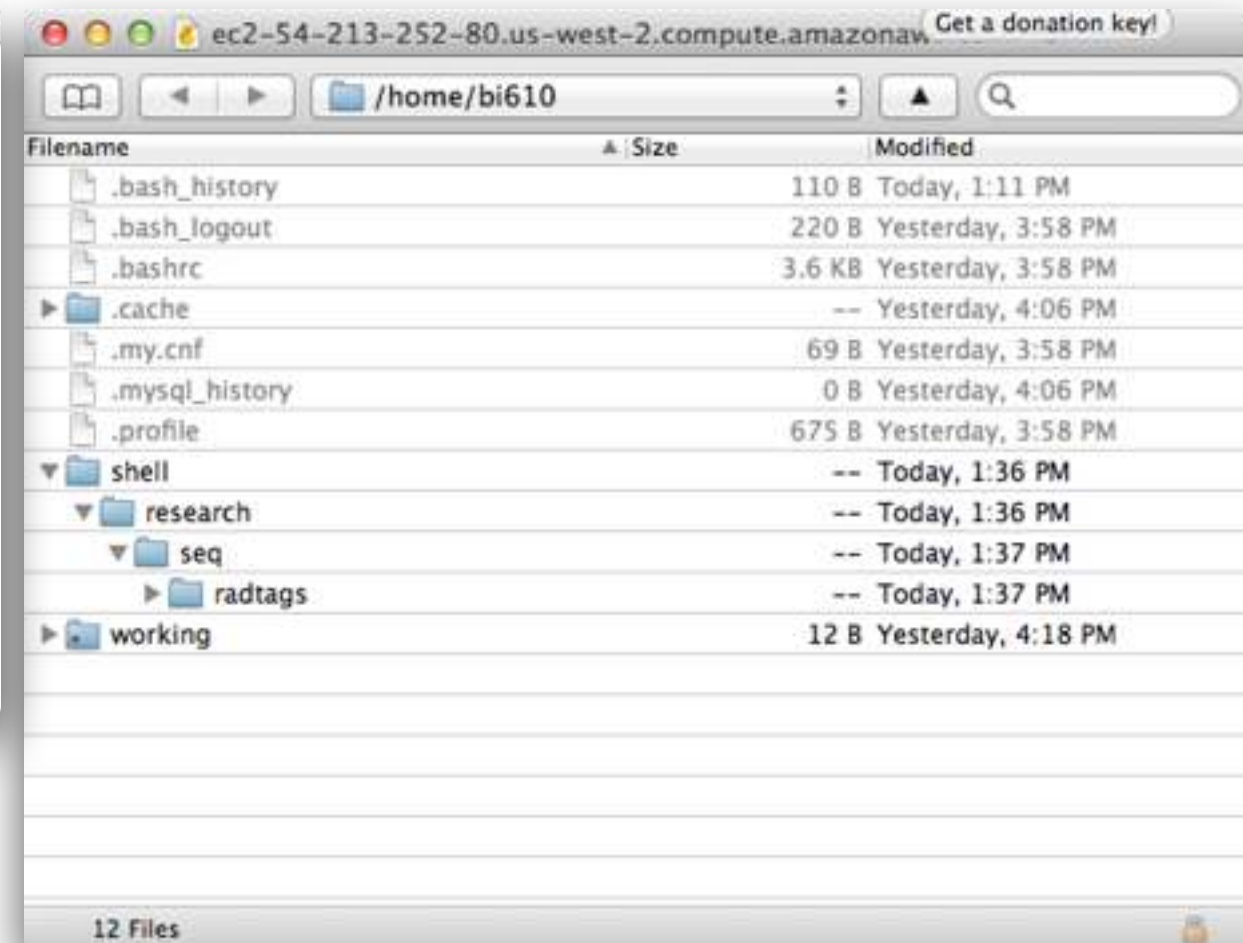
% cd

% pwd

Relative and absolute paths

/home/tgac/shell/research/seq/radtags

```
tgac@Genotyping2: ~/shell/research/seq/radtags — ssh — 105x23
tgac@Genotyping2:~$ mkdir shell
tgac@Genotyping2:~$ cd shell
tgac@Genotyping2:~/shell$ mkdir research
tgac@Genotyping2:~/shell$ ls
research
tgac@Genotyping2:~/shell$ cd research/
tgac@Genotyping2:~/shell/research$ mkdir seq
tgac@Genotyping2:~/shell/research$ ls
seq
tgac@Genotyping2:~/shell/research$ cd seq/
tgac@Genotyping2:~/shell/research/seq$ ls
tgac@Genotyping2:~/shell/research/seq$ mkdir radtags
tgac@Genotyping2:~/shell/research/seq$ ls
radtags
tgac@Genotyping2:~/shell/research/seq$ ls -la
total 12
drwxrwxr-x 3 tgac tgac 4096 Nov  4 09:34 .
drwxrwxr-x 3 tgac tgac 4096 Nov  4 09:34 ..
drwxrwxr-x 2 tgac tgac 4096 Nov  4 09:34 radtags
tgac@Genotyping2:~/shell/research/seq$ cd radtags/
tgac@Genotyping2:~/shell/research/seq/radtags$ pwd
/home/tgac/shell/research/seq/radtags
tgac@Genotyping2:~/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’

```
% cd /etc
```

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

```
% pwd
```

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

```
bi610@ip-172-31-17-121: /etc — ssh — 92x40
bi610@ip-172-31-17-121:~/shell/research/seq/radtags$ cd /etc/
bi610@ip-172-31-17-121:/etc$ ls -l
total 868
-rw-r--r-- 1 root root 15070 Apr  4 04:30 a2ps.cfg
-rw-r--r-- 1 root root 2563 Apr  4 04:30 a2ps-site.cfg
drwxr-xr-x 3 root root 4096 Apr 16 18:26 acpi
-rw-r--r-- 1 root root 2981 Apr 16 18:25 adduser.conf
-rw-r--r-- 1 root root 10 Apr 16 18:27 adjtime
drwxr-xr-x 2 root root 4096 Jun 20 22:23 alternatives
drwxr-xr-x 8 root root 4096 Jun 20 21:30 apache2
drwxr-xr-x 3 root root 4096 Apr 16 18:25 apm
drwxr-xr-x 3 root root 4096 Apr 16 18:26 apparmor
drwxr-xr-x 8 root root 4096 Jun 19 22:10 apparmor.d
drwxr-xr-x 3 root root 4096 Jun 19 22:00 appport
drwxr-xr-x 6 root root 4096 Apr 16 18:27 apt
-rw-r----- 1 root daemon 144 Oct 21 2013 at.deny
-rw-r--r-- 1 root root 2177 Apr  9 01:03 bash.bashrc
-rw-r--r-- 1 root root 45 Mar 22 20:57 bash_completion
drwxr-xr-x 2 root root 4096 Jun 20 21:44 bash_completion.d
-rw-r--r-- 1 root root 356 Jan  1 2012 bindresvport.blacklist
-rw-r--r-- 1 root root 321 Apr 16 16:07 blkid.conf
lrwxrwxrwx 1 root root 15 Jun  3 20:54 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x 2 root root 4096 Apr 16 18:26 byobu
drwxr-xr-x 3 root root 4096 Apr 16 18:25 ca-certificates
-rw-r--r-- 1 root root 7464 Apr 16 18:26 ca-certificates.conf
drwxr-xr-x 2 root root 4096 Apr 16 18:26 calendar
drwxr-s--- 2 root dip 4096 Apr 16 18:26 chatscripts
drwxr-xr-x 4 root root 4096 Apr 16 18:27 cloud
drwxr-xr-x 2 root root 4096 Apr 16 18:25 console-setup
drwxr-xr-x 2 root root 4096 Jun 19 22:10 cron.d
drwxr-xr-x 2 root root 4096 Jun 19 22:10 cron.daily
drwxr-xr-x 2 root root 4096 Apr 16 18:25 cron.hourly
drwxr-xr-x 2 root root 4096 Apr 16 18:25 cron.monthly
-rw-r--r-- 1 root root 722 Feb  9 2013 crontab
drwxr-xr-x 2 root root 4096 Jun 20 22:23 cron.weekly
-rw-r--r-- 1 root root 54 Apr 16 18:26 crypttab
drwxr-xr-x 4 root root 4096 Apr 16 18:26 dbus-1
-rw-r--r-- 1 root root 2969 Feb 23 14:30 debconf.conf
-rw-r--r-- 1 root root 11 Feb 20 02:43 debian_version
drwxr-xr-x 3 root root 4096 Jun 20 22:23 default
```

Three variants to ls

```
ls -l
```

provides a *long*
listing

```
ls -la
```

includes *all* files,
even hidden files

```
ls -lh
```

displays file sizes
in *human* readable
numbers

Four ways to view a text file

<code>more</code>	<code>head</code>	<code>tail</code>	<code>cat</code>
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 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 file `'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?

Copy example files

Return to the directory in your home called 'shell'.

TSV file:

`~/workshop_data/unix/batch_1.genotypes_1.loc.gz`

FASTQ file:

`~/workshop_data/unix/s_1_sequence.txt.gz`

Tar Archive:

`~/workshop_data/unix/samples.tar.gz`

What is a tar archive?



tar = tape archive

Compress / Decompress

`gzip / gunzip`

`batch_l.genotypes_l.loc.gz`

`s_l_sequence.txt.gz`

Gzipped Tar archive

`tar xvfz`

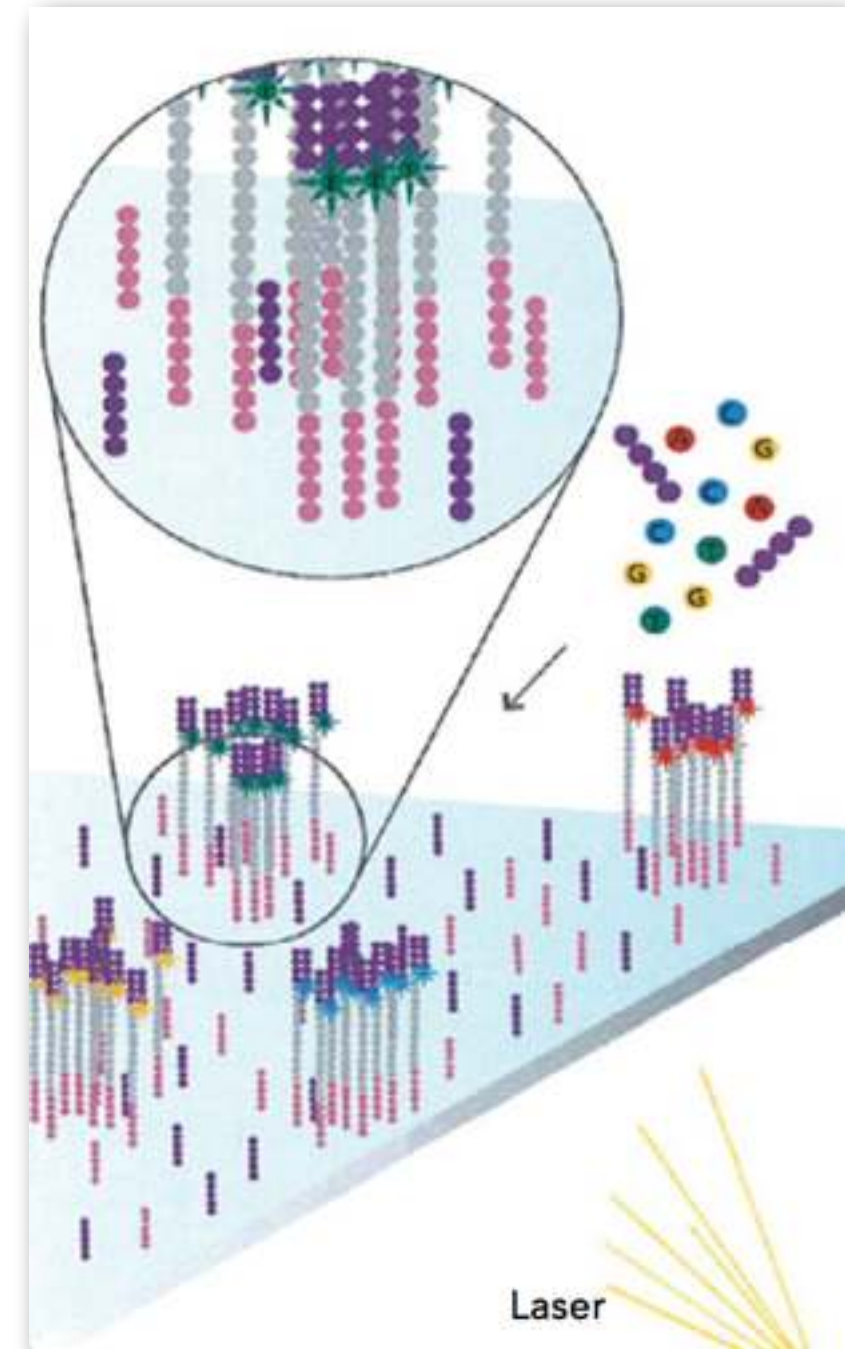
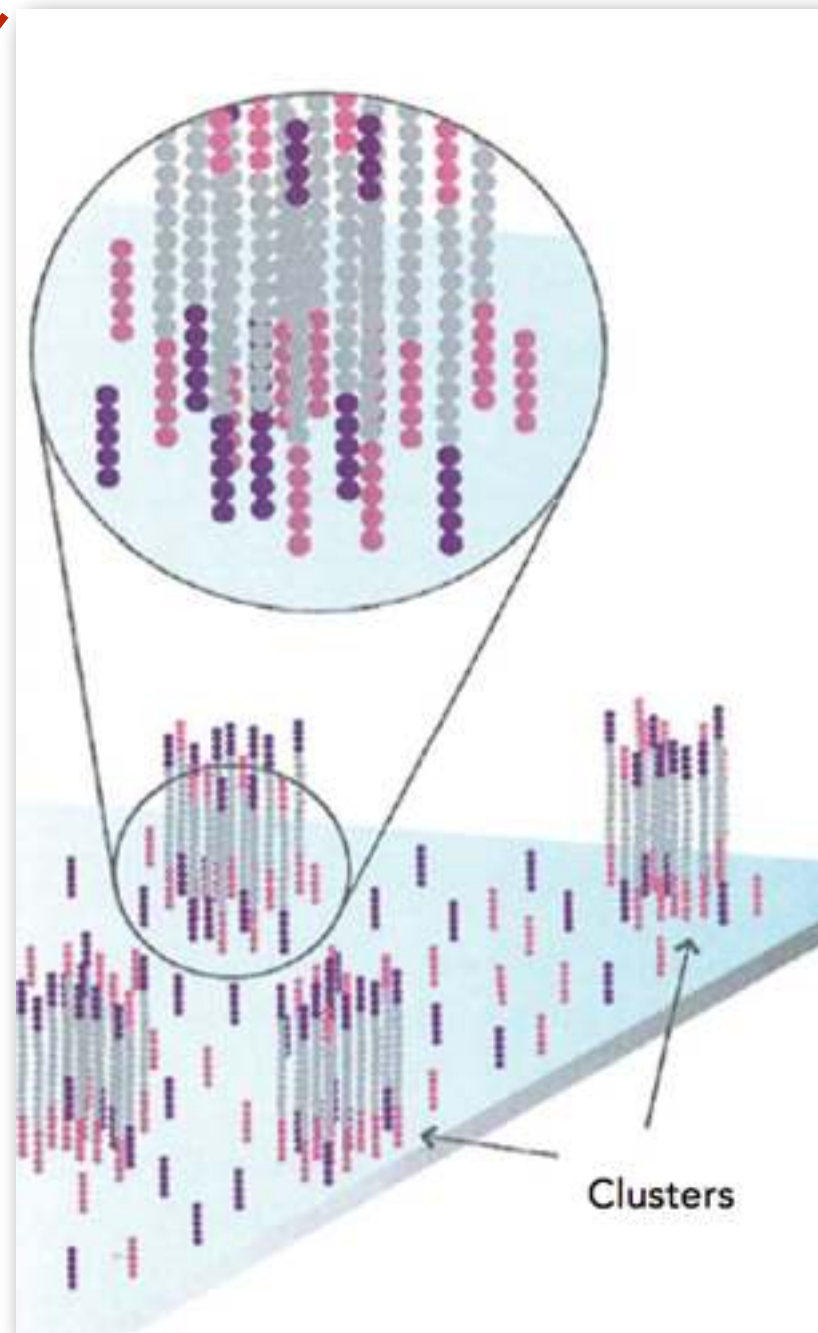
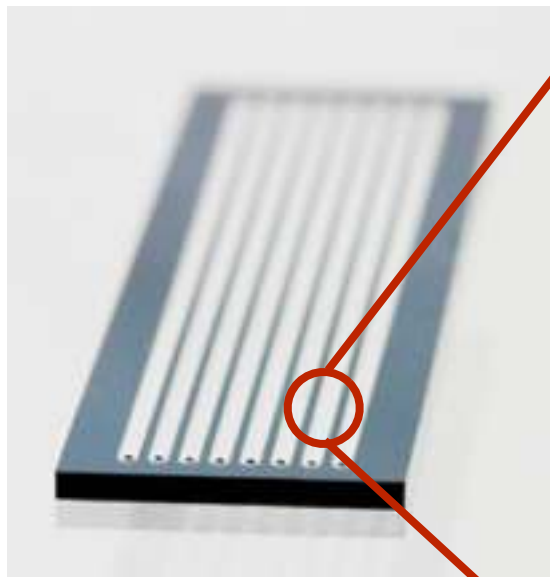
`samples.tar.gz`

Tar archive

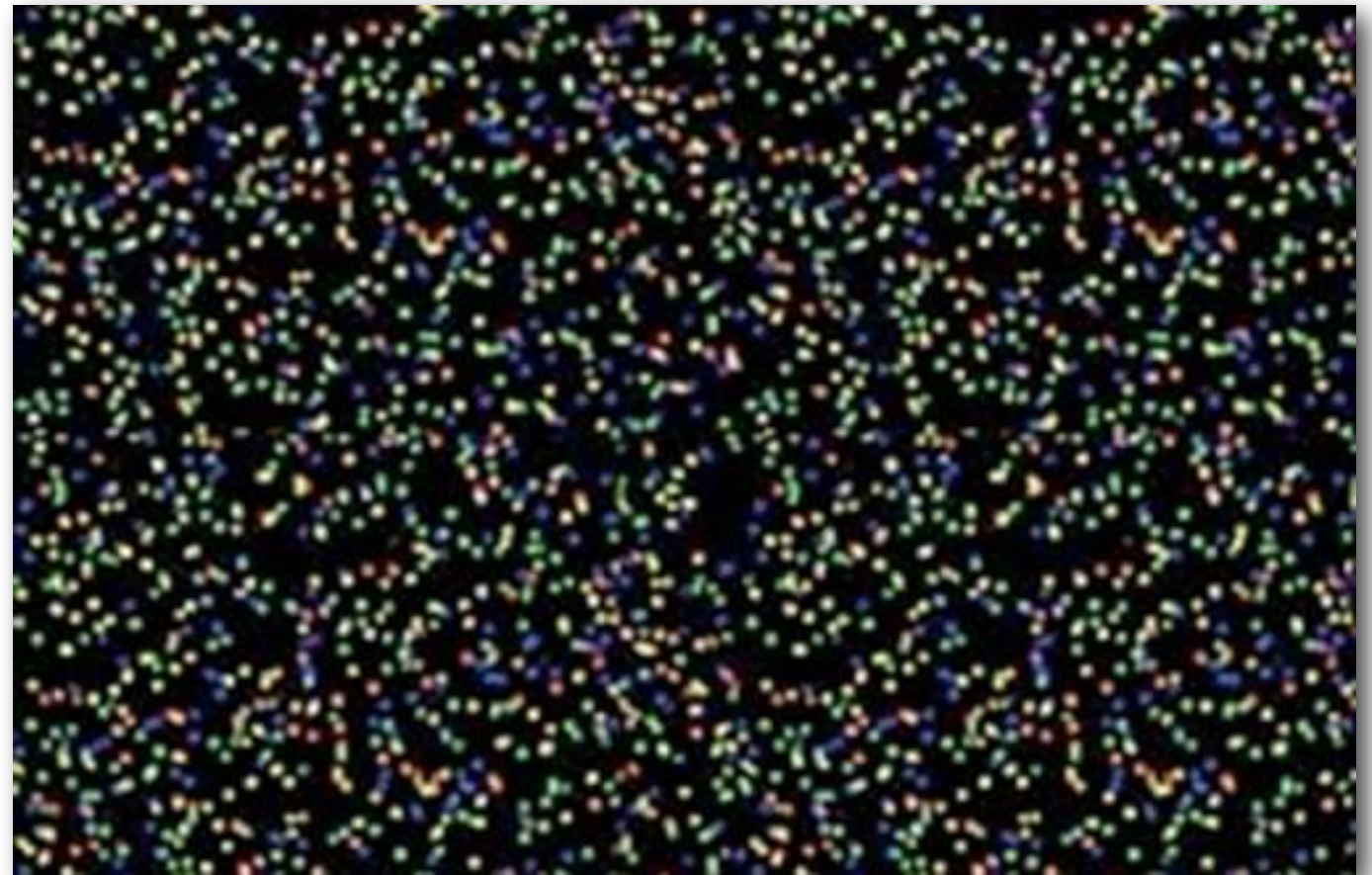
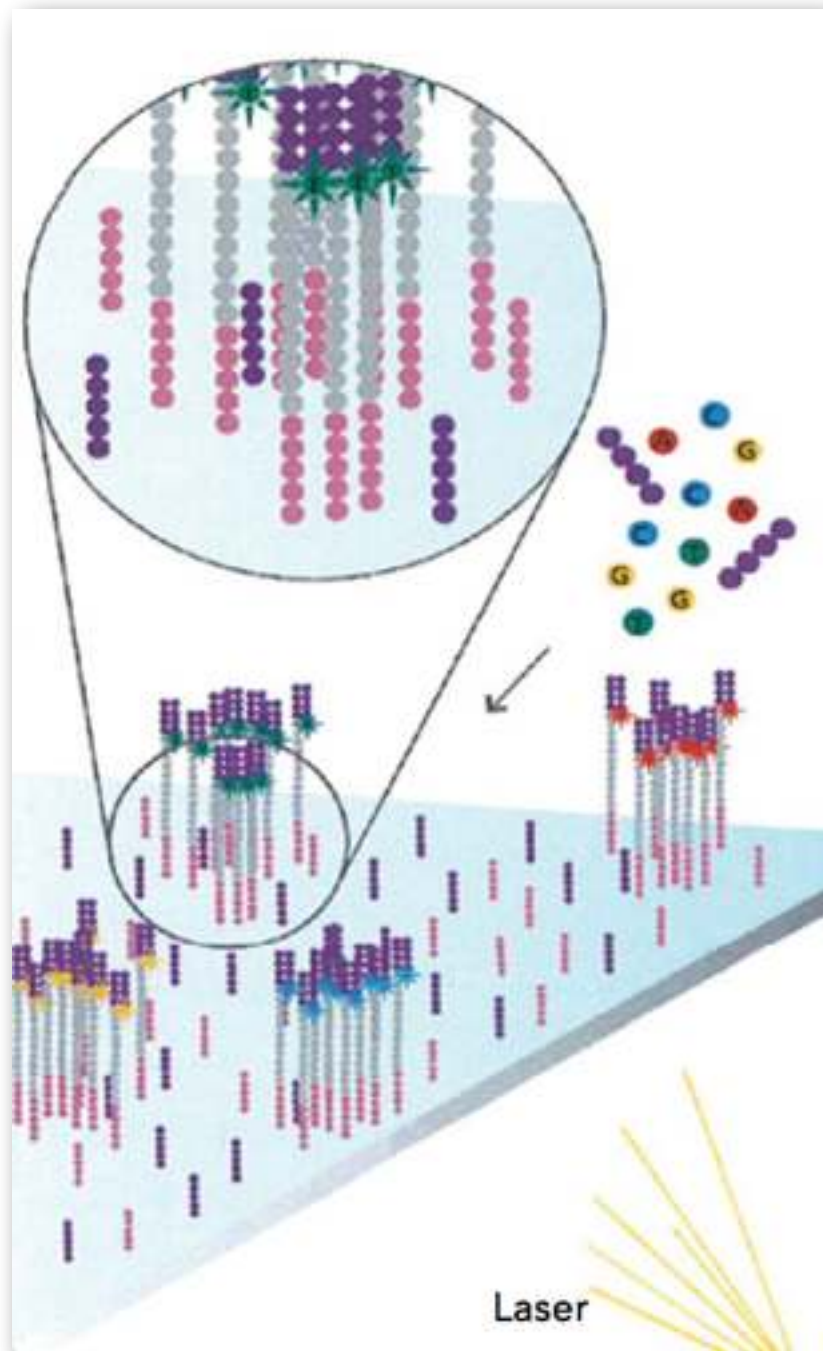
`tar xvf`

`samples.tar`

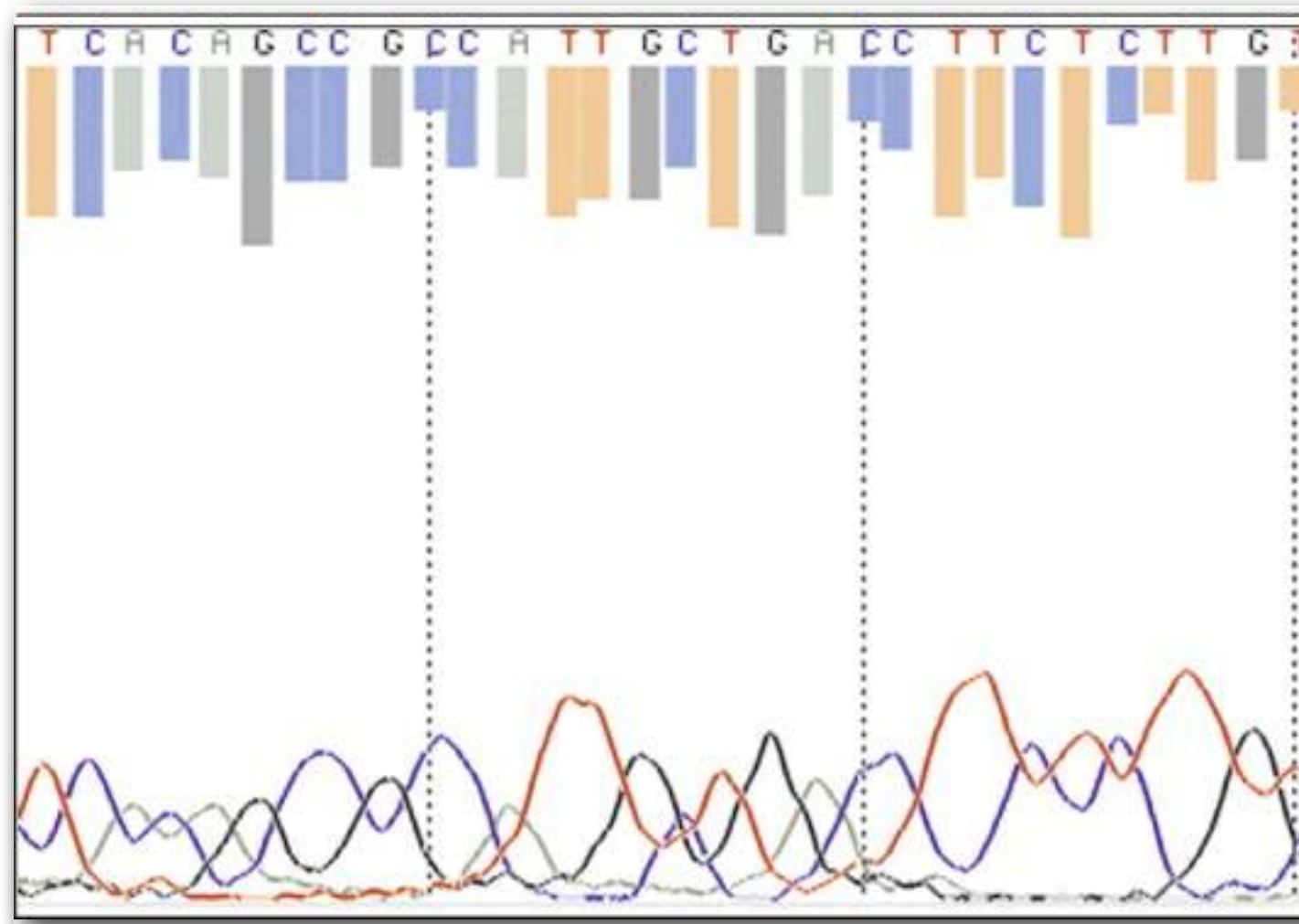
Sequencing on Illumina's Flow cell



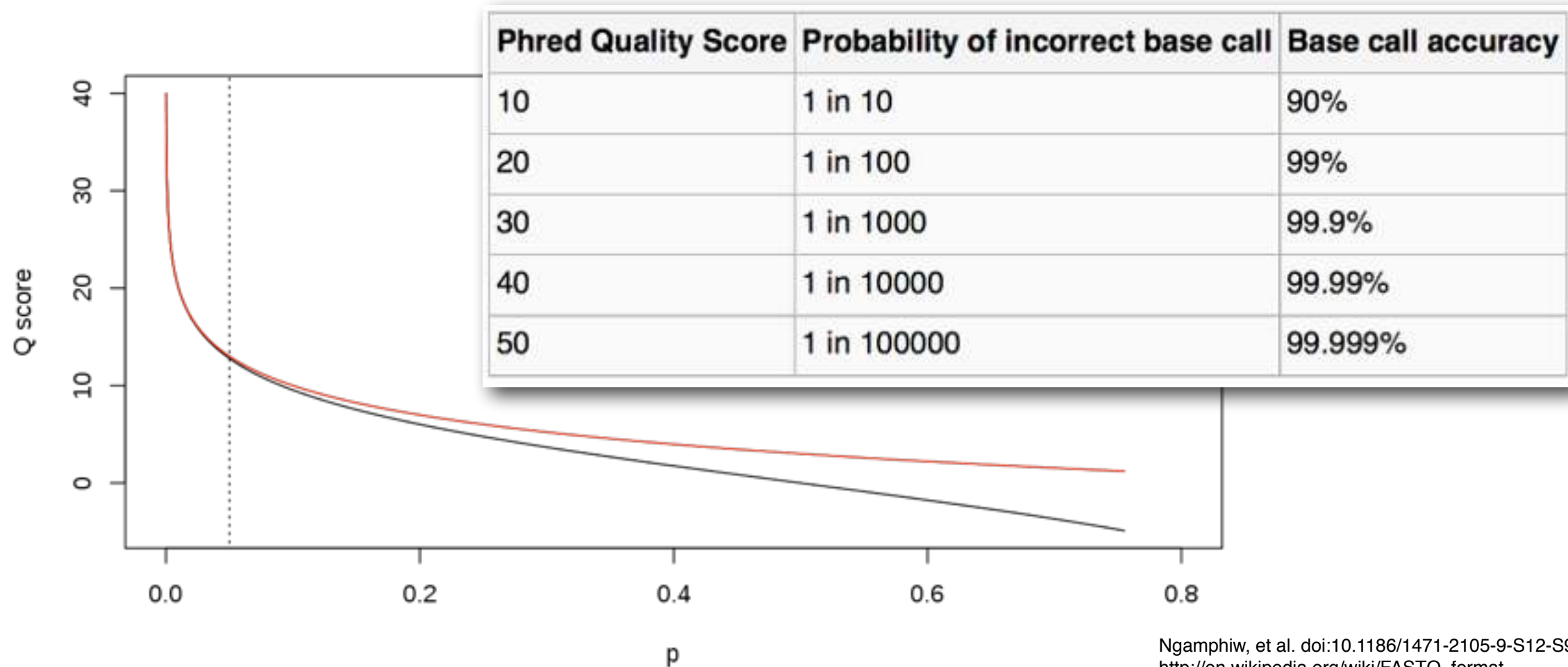
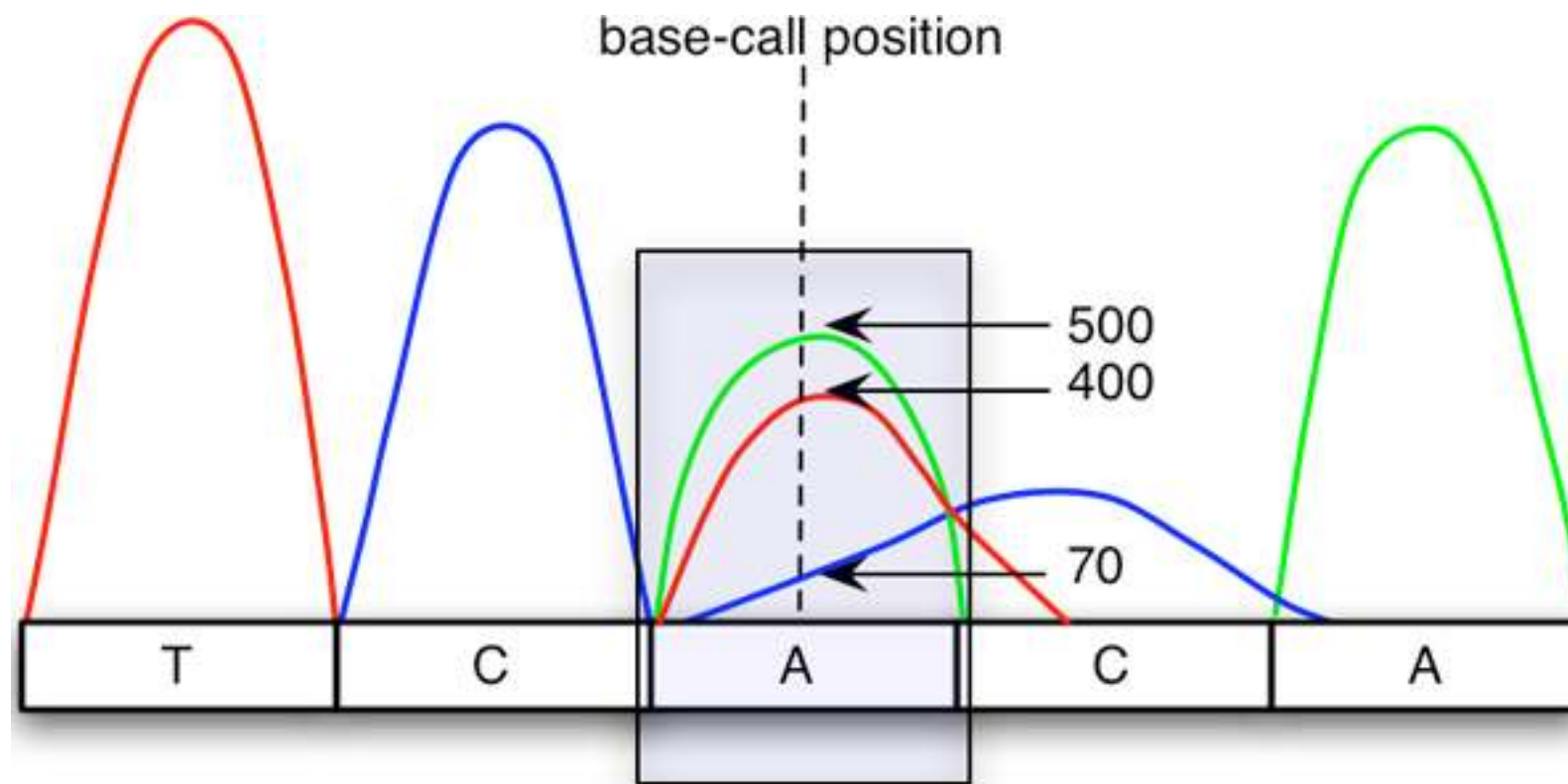
Sequencing on Illumina's Flow cell, ctd.



Sequencing on Illumina's Flow cell, ctd.



Phred Quality Score



The FASTQ File Format

FASTA

```
>chromosome7
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
ACATACTGCAGGGGTCAGGAGGATTATCTCCTCTGCAAGGTAACGCCTGCTGTAACCGTTGTTCTTCATCCTTTT
CCTAACTGCAGGGGCTGTCTTGTCAGGTCTGACAAGACATATGCAGGGGCTCAATTTGAGATAATTGCTCAATATA
```

FASTQ

```
@Sequence_137
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+Sequence_137
<?@DDDDDDHFHHFBB@GGIACFHGGHGBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

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

ASCII Code

0	<NUL>	32	<SPC>	64	@	96	`	128	Ä	160	†	192	¿	224	‡
1	<SOH>	33	!	65	A	97	a	129	Å	161	°	193	ì	225	·
2	<STX>	34	"	66	B	98	b	130	Ç	162	¢	194	í	226	,
3	<ETX>	35	#	67	C	99	c	131	É	163	£	195	√	227	„
4	<EOT>	36	\$	68	D	100	d	132	Ñ	164	§	196	ƒ	228	‰
5	<ENQ>	37	%	69	E	101	e	133	Ö	165	•	197	≈	229	Â
6	<ACK>	38	&	70	F	102	f	134	Ü	166	¶	198	Δ	230	Ê
7	<BEL>	39	'	71	G	103	g	135	á	167	ß	199	«	231	Á
8	<BS>	40	(	72	H	104	h	136	à	168	®	200	»	232	Ë
9	<TAB>	41	)	73	I	105	i	137	â	169	©	201	...	233	È
10	<LF>	42	*	74	J	106	j	138	ä	170	™	202		234	Í
11	<VT>	43	+	75	K	107	k	139	å	171	'	203	À	235	Î
12	<FF>	44	,	76	L	108	l	140	ä	172	“	204	Ã	236	Ï
13	<CR>	45	-	77	M	109	m	141	ç	173	≠	205	Ö	237	Ì
14	<SO>	46	.	78	N	110	n	142	é	174	Æ	206	Œ	238	Ó
15	<SI>	47	/	79	O	111	o	143	è	175	Ø	207	œ	239	Ô
16	<DLE>	48	0	80	P	112	p	144	ê	176	∞	208	-	240	Ⓜ
17	<DC1>	49	1	81	Q	113	q	145	ë	177	±	209	—	241	Ò
18	<DC2>	50	2	82	R	114	r	146	í	178	≤	210	“	242	Ú
19	<DC3>	51	3	83	S	115	s	147	ì	179	≥	211	”	243	Û
20	<DC4>	52	4	84	T	116	t	148	î	180	¥	212	`	244	Ü
21	<NAK>	53	5	85	U	117	u	149	ï	181	μ	213	'	245	ı
22	<SYN>	54	6	86	V	118	v	150	ñ	182	ð	214	÷	246	ˆ
23	<ETB>	55	7	87	W	119	w	151	ó	183	Σ	215	◊	247	˜
24	<CAN>	56	8	88	X	120	x	152	ò	184	Π	216	ÿ	248	—
25		57	9	89	Y	121	y	153	ô	185	π	217	ÿ	249	˘
26	<SUB>	58	:	90	Z	122	z	154	ö	186	ƒ	218	/	250	·
27	<ESC>	59	;	91	[	123	{	155	õ	187	ª	219	€	251	°
28	<FS>	60	<	92	\	124		156	ú	188	º	220	<	252	¸
29	<GS>	61	=	93	]	125	}	157	ù	189	Ω	221	>	253	”
30	<RS>	62	>	94	^	126	~	158	û	190	æ	222	fi	254	˚
31	<US>	63	?	95	_	127		159	ü	191	ø	223	fi	255	˛

8 bits = 2^8 combinations = 256

0	1	1	1	1	0	0	1
---	---	---	---	---	---	---	---

$$0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 1 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 121 = y$$

$$1 \times 10^2 + 2 \times 10^1 + 1 \times 10^0 = 121 = y$$

The FASTQ File Format, ctd

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

Quality Scores

```
SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....  
.....XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....  
.....IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....  
.....JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ.....  
LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....  
!"#$%&'()*+,-./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 33 - 73 = 0 - 40

‘F’ = 70

$$70 - 33 = 37$$

The FASTQ File Format, ctd

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

$$70 - 33 = 37$$

Phred 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%
50	1 in 100000	99.999%

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
```

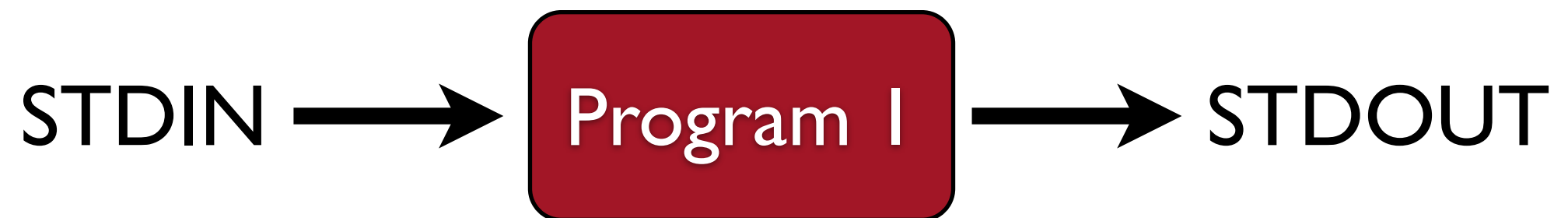
Special Files

STDIN, STDOUT, STDERR

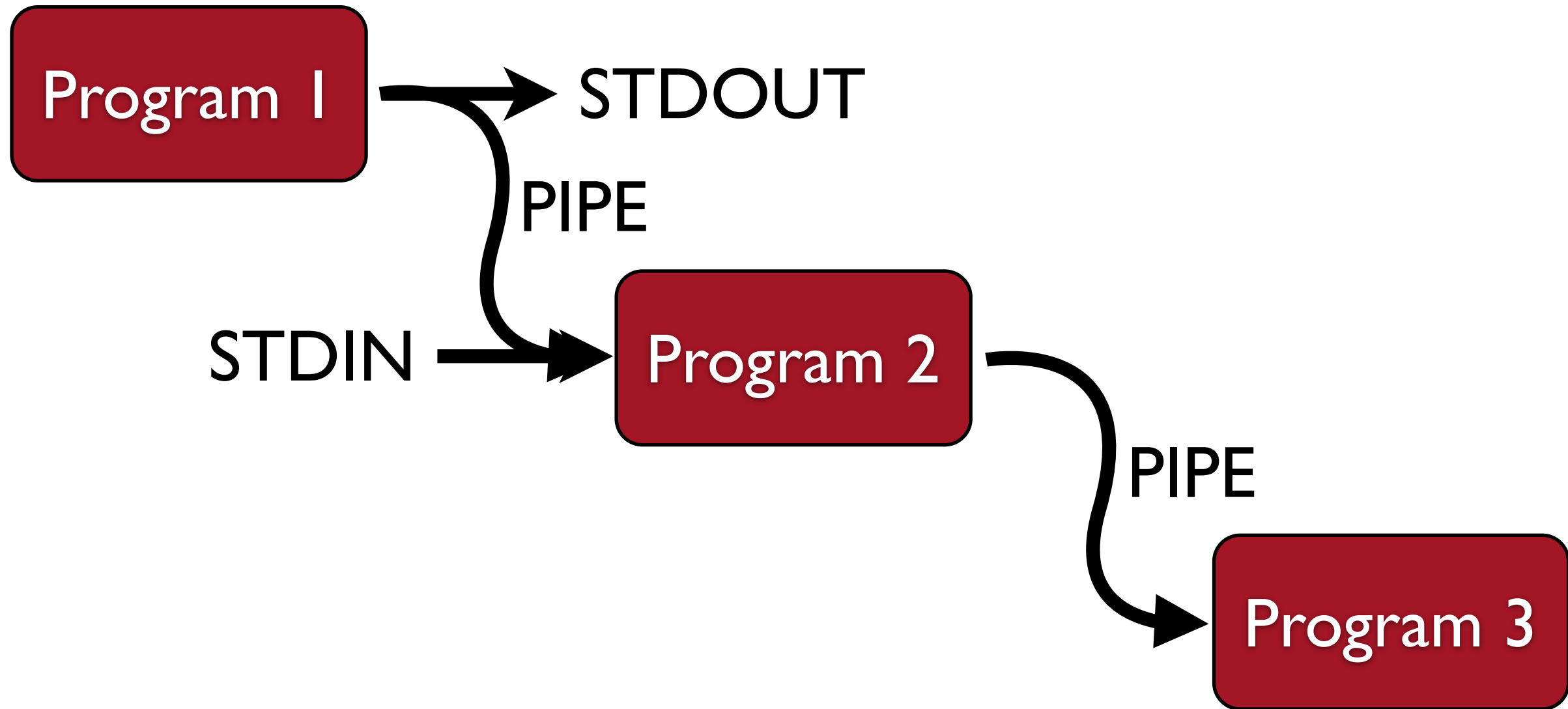
The Shell's Killer App: Pipes



The Shell's Killer App: Pipes, ctd.



The Shell's Killer App: Pipes, ctd.



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_l_sequence.txt.gz

Decompress the file

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

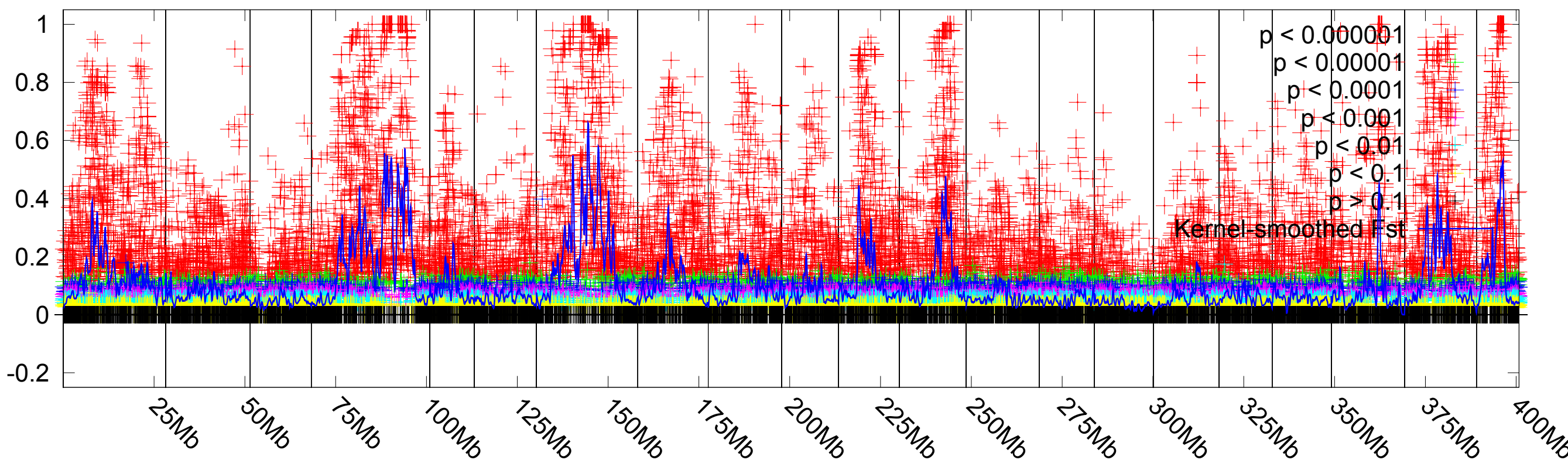
```
286 CTAGT
7900 TCAGA
10659 ACTGC
10931 TGACC
11536 GAGAT
11871 CTGAA
14409 CGGCG
14508 TGGTT
18226 GAAGC
18352 ACCAT
18375 TCGAG
19501 CGATA
23012 AATTT
26336 GCATT
31136 CTAGG
```

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

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

Problem Set #1

Danger Is. #04 vs Middleton Is. #16 Fst



#	Batch	ID	Locus	ID	Pop 1	ID	Pop 2	ID	Chr	BP	Column	Overall	Pi	Fst	Fisher's	P	Odds	Ratio
2	7894	2	3	groupI	11832	19	0.428182	-0.0076252913						0.191294			0.687192	
2	7896	2	3	groupI	11900	83	0.328622	0.1775694587						4.35747e-08			5.44667	
2	9611	2	3	groupI	49756	48	0.090426	-0.1127451906						0.00109115			0.0898072	
2	9611	2	3	groupI	49765	57	0.0132887	-0.1522407447						0.518395			0.493113	
2	9611	2	3	groupI	49766	58	0.0133776	-0.1563725438						0.520033			0.501401	
...																		
2	7083	2	3	groupIX	20172984	57	0.0480227	-0.1249752728						0.0230205			0.152263	
2	7083	2	3	groupIX	20173004	77	0.0181808	-0.1405892833						0.27288			0.355556	
2	7083	2	3	groupIX	20173016	89	0.0424072	-0.1243625038						0.0442229			0.17037	
2	7083	2	3	groupIX	20173018	91	0.148072	0.2305903127						2.24832e-07			13.2936	
2	7084	2	3	groupIX	20172960	29	0.388239	-0.0102170169						0.307185			0.758001	
2	7084	2	3	groupIX	20172995	64	0.368259	-0.0130602425						0.233889			0.70084	
2	7084	2	3	groupIX	20173013	82	0.0121578	-0.1359057575						0.51309			0.46472	