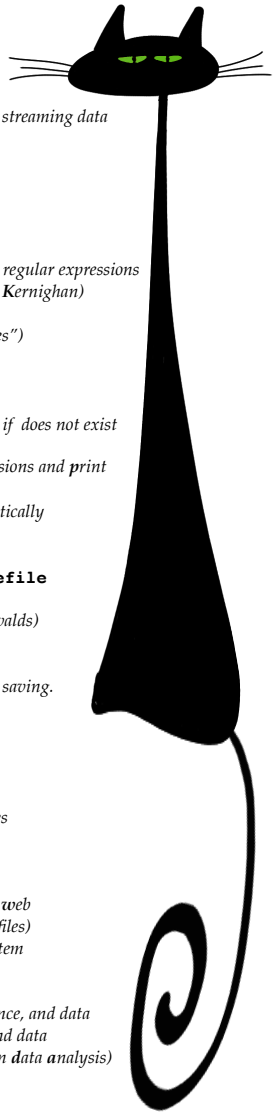
etc. *nix Survival Guide for Data Science

head	*show the first lines in file ("**head** of file")*
cat	*show ("con**cat**enate") files*
more	*show file, paging with space bar to get **more***
less	*more-or-**less** the same as **more**, but remembers streaming data*
echo	*print message*
rm	*r**m**ove file*
cp	*c**p**y file*
mv	***m**ove or rename file*
ln	*create **l**in**k** to file. (**ln -s** for **s**ymbolic link)*
sed	***s**tream **ed**itor: replace matching patterns using regular expressions*
awk	*a kind-of souped-up **sed** (**A**ho, **W**einberger & **K**ernighan)*
sort	***sort** lines in file*
uniq	*remove duplicate adjacent lines in file ("**uniq**ues")*
man	***man**splain a unix command*
cut	*extract (**cut**) fields from flat file*
wc	*count characters, words, lines ("**w**ord **c**ount")*
touch	*update timestamp on file by **touch**ing it; create if does not exist*
grep	*find lines in files (**g**lobally) with **r**egular **e**xpressions and **p**rint*
find	***find** files with specified properties*
xargs	*construct **arg**ument**s** for command programmatically*
.	*execute script in current shell*
sh	*execute script in new **sh**ell*
make	*build ("**make**") something from recipe in **Makefile***
git	*version control (named after creator, Linus Torvalds)*
tar	*create/extract archives (**t**ape **ar**chive)*
vi/vim	*edit file with **V**isual **I**nterface to editor*
	*If alarmed, hit escape then **:q!** to exit without saving.*
pwd	***p**rint **w**orking **d**irecory (where am I)*
ls -lrt	*list files ("**l**ist **s**orted")*
cd	*change **d**irectory*
mkdir	***m**ake a **dir**ectory*
chmod	***ch**ange permissions (**mod**e) of files & directories*
chown	***ch**ange **own**ership of files & directories*
su -	*set **u**ser to a (normally different) user*
curl/wget	*download (**get**) files from internet **URL** on the **w**eb*
sftp	***s**ecure **f**ile **t**ransfer **p**rotocol (download/upload files)*
rsync	*(remote) **syn**chronize files from/to another system*
ssh	***s**ecurely log into another server's **sh**ell*
python	*the best programming language for maths, science, and data*
R	*the best programming language for statistics and data*
tdda	*generate constraints and check data (**t**est-**d**riven **d**ata **a**nalysis)*
tail	*show the last lines in file (**tail** of file)*



INPUT AND OUTPUT STREAMS

c > a	*write output (stdout) from command **c** to file **a***
c 2>&1 > a	*write all output from command **c** to file **a** (stdout + stderr)*
c >> a	*append output from **c** to file **a***
c < a	*take input from file **a** (replacing stdin) for command **c***
c1 | c2	*pipe output from command **c1** to command **c2***
c &	*run command **c** in background*
^D	*end of file/end of input/log out of shell (control+D key)*
^Z	*stop job (suspend job; can be reactivated — control+Z key)*
^C	*interrupt program (control+C key)*
jobs	*show **jobs** (inc stopped jobs)*
fg %1	*restart stopped job 1 in **f**oreground*
bg %1	*restart stopped job 1 in **b**ack**g**round*
kill -9 %1	***kill** job 1 with extreme prejudice (**-9** is extreme prejudice)*
ps egu	*show running **p**rocesses' **s**tatus*

DIRECTORIES

.	*current directory **$CWD / $(cwd)***
..	*parent of current directory*
/	*root of the file system*
~	*current user's (**whoami**'s) home directory*
~user	***user**'s home directory*

SHELL VARIABLES

$?	*exit status of last command (**0** is normal/success; **1–127** are error codes)*
$PATH	*colon-separated list of directories to search for commands*
$TMPDIR	*override for temporary directory **/tmp.***

QUOTING & ESCAPING

'star* *filename'	*single quotes protect arguments strongly*
"spacey filename"	*double quotes protect less (shell variable expansion continues)*
spacey\ filename	*backslash escapes special characters like whitespace, backslash*
A=`ls -l`	*assign output of **ls -l** to variable **A***

GLOBBING

*****	*matches any file/portion of filename e.g. **ls *.txt** to match **.txt** files*
?	*matches any character in filename e.g. **ls f?.txt** for **f1.txt**, **f2.txt** etc.*
[0-9]	*matches characters **0** to **9**, e.g. **cat [0-9].txt > a***

RECIPES

tar cvzf foo.tgz foo	*create compressed archive of **foo** in **foo.tgz***
tar xvd foo.tar	*extract contents of **foo.tar***
sort a | uniq | wc -l	*count the number of different lines in file **a***
find . -name '*.tgz'	*find **.tgz** files in **. and** subdirectories*
find . -name '*.py' | grep '40.'	*search for **40.** in **.py** files in directory tree*
ls -a	***l**ist **a**ll files, including hidden files (which start with **.**)*
rm -rf dir	*r**m** ove **dir** and all its directories **r**ecursively, **f**orcefully*

