Awk :

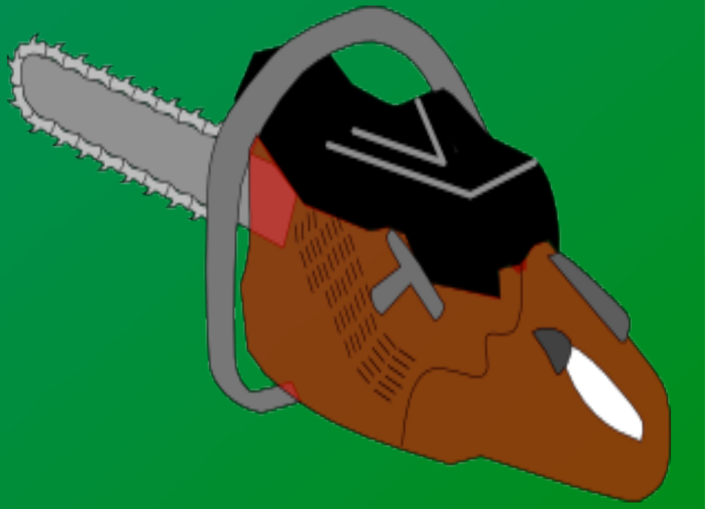
This will change your life

By : Brittany Kamai
03 June 2010

Solving a problem requires choosing the correct tool

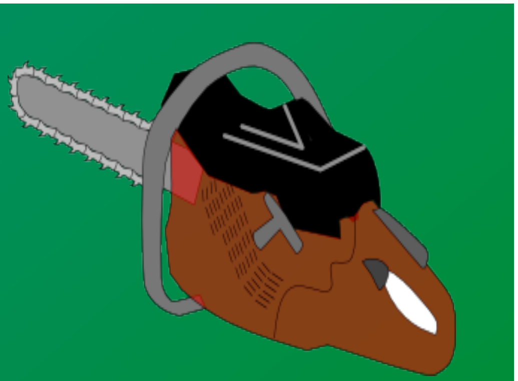
GREP

SED



AWK

What is AWK?



A simple,elegant pattern scanning and processing language
A scripting language that can handle text processing tasks

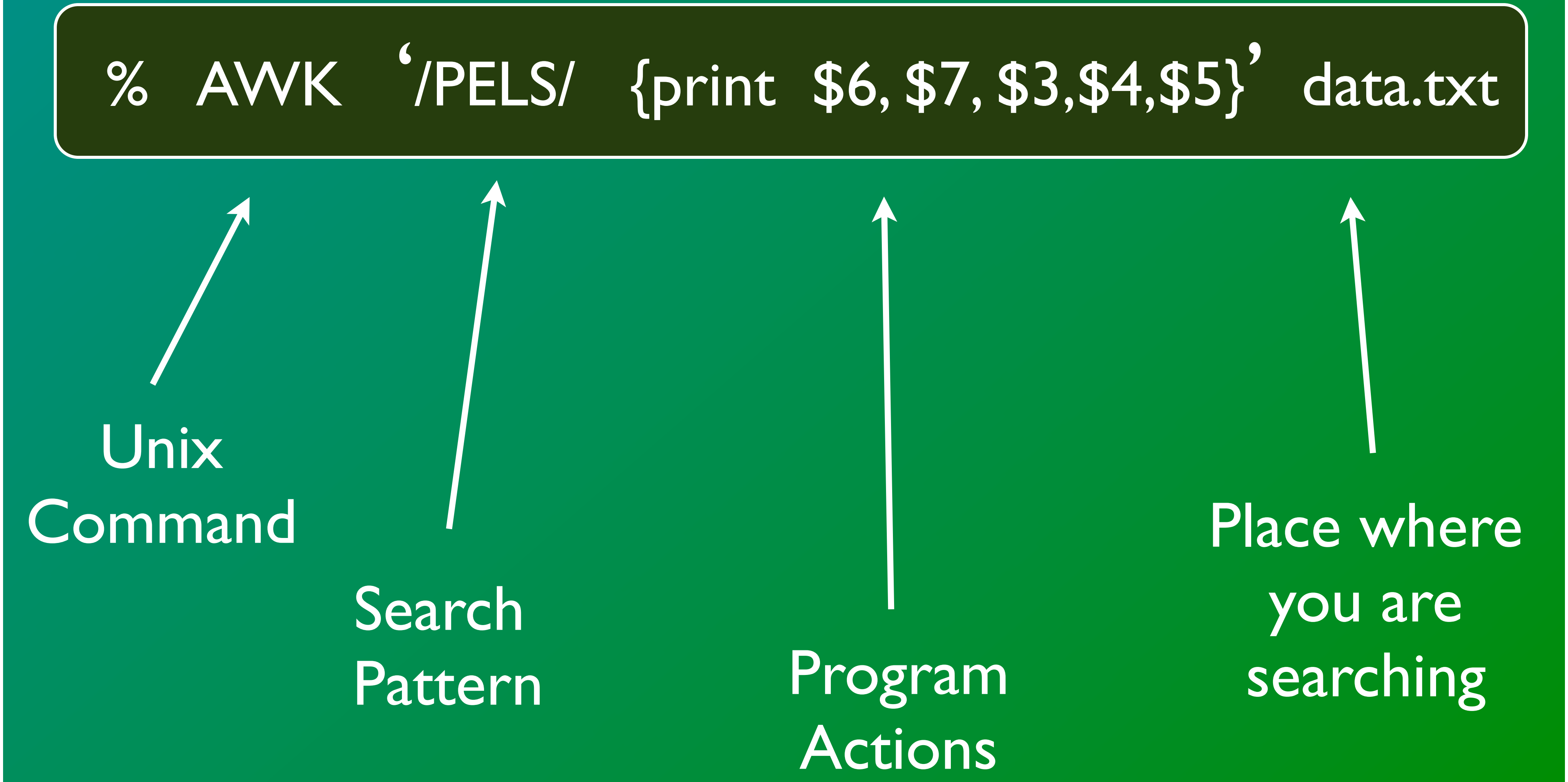
Where will I use AWK?

Great for quick and dirty calculations
Finding out information in a large data file
Creating a small database

AWK pep talk

Anyone who can write a tiny basic program can write a command using awk
Users of C will find the notation of this utility easy

Awk Syntax



```
%  AWK  ‘/PELS/  {print  $6, $7, $3,$4,$5}’  data.txt
```

Unix
Command

Search
Pattern

Program
Actions

Place where
you are
searching

Awk Syntax : Searching

1. Search can be Regular Expressions

```
% AWK   /[a-z][A-Z]*/  data.txt
```

2. Search *DOESN'T* have to be

- **~** : Searches for that string
- !~ : Searches **not** for that string

```
% AWK  $1  ~  /star/
% AWK  $1  !~ /^$asteroid/
```

Data.txt

```
RA         DEC        J      H      K      B     V     ST  m     g     name

56.256958  19.559372  8.404  8.225  8.152   9.92  9.42  O   18.1  3201  PELS135
60.934029  22.944309  8.519  8.262  8.183  10.25  9.67  B    8.6  8901  PELS174
59.507069  20.676823  8.436  8.248  8.197  10.13  9.63  K   0.72  4262  PELS086
55.961143  25.268934  8.511  8.284  8.214  10.15  9.63  M   0.12  4423  AKIB146   HD23170
57.470463  25.647491  8.489  8.332  8.254  10.09  9.57  M   0.33  4158  PELS150   HD23935  TrS144y
52.355843  25.652304  8.459  8.314  8.270   9.90  9.43  F   1.04  6821  AKII179
53.530495  24.344501  8.553  8.312  8.274  10.08  9.57  G   0.81  9932  PELS006   TrS4
56.245522  22.032455  8.520  8.349  8.278   9.97  9.48  A   1.50  4612  PELS140   HD23312  TrS61
55.012829  27.740541  8.638  8.479  8.373  10.11  9.62  A   2.06  5123  PELS025
57.573582  25.379555  8.645  8.469  8.382  10.14  9.64  G   0.95  4187  TrS151x   HD23975
58.589973  24.075645  8.687  8.480  8.424  10.29  9.75  K   0.45  1102  AKIA317   HD24463
56.837021  25.525862  8.720  8.463  8.435  10.41  9.85  M   0.30  3124  PELS060   HD23598
60.221325  23.193913  8.676  8.558  8.488  10.06  9.61  F   1.08  8019  PELS173
55.599987  21.473492  8.797  8.604  8.526  10.40  9.85  O   17.8  7162  PELS035   TrS42
53.882030  22.823627  8.831  8.599  8.541  10.40  9.86  G   1.02  4331  PELS124
56.291157  21.704592  8.902  8.694  8.586  10.52  9.95  M   0.24  3131  PELS040
55.400673  25.619331  8.959  8.657  8.593  10.69 10.10  M   0.18  9881  PELS023   DH181
```

Fields

$1 $2 $3 $4 $5 $6 $7 $8 $9 $10 $11 $12 $13

Numbers : Recognizes integer & floating point numbers

Strings : Expressed in double quotes

How many PELS stars do I have?

```
RA         DEC         J      H      K      B      V     ST  m    g     name
$1         $2          $3     $4     $5     $6     $7    $8  $9   $10   $11       $12      $13
56.256958  19.559372  8.404  8.225  8.152   9.92  9.42   O  18.1  3201  PELS135
60.934029  22.944309  8.519  8.262  8.183  10.25  9.67   B   8.6  8901  PELS174
59.507069  20.676823  8.436  8.248  8.197  10.13  9.63   K  0.72  4262  PELS086
55.961143  25.268934  8.511  8.284  8.214  10.15  9.63   M  0.12  4423  AK1B146   HD23170
57.470463  25.647491  8.489  8.332  8.254  10.09  9.57   M  0.33  4158  PELS150   HD23935  TrS144y
52.355843  25.652304  8.459  8.314  8.270   9.90  9.43   F  1.04  6821  AKIII79
53.530495  24.344501  8.553  8.312  8.274  10.08  9.57   G  0.81  9932  PELS006   TrS4
56.245522  22.032455  8.520  8.349  8.278   9.97  9.48   A  1.50  4612  PELS140   HD23312  TrS61
55.012829  27.740541  8.638  8.479  8.373  10.11  9.62   A  2.06  5123  PELS025
57.573582  25.379555  8.645  8.469  8.382  10.14  9.64   G  0.95  4187  TrS151x   HD23975
58.589973  24.075645  8.687  8.480  8.424  10.29  9.75   K  0.45  1102  AK1A317   HD24463
```

```
%  AWK  '/PELS/  {print  $11,$6, $7, $3,$4,$5}'  data.txt
```

```
Names      B      V     J      H      K
$11        $6     $7    $3     $4     $5

PELS135    9.92  9.42  8.404  8.225  8.152
PELS174   10.25  9.67  8.519  8.262  8.183
PELS086   10.13  9.63  8.436  8.248  8.197
PELS150   10.09  9.57  8.489  8.332  8.254
PELS006   10.08  9.57  8.553  8.312  8.274
PELS140    9.97  9.48  8.520  8.349  8.278
PELS025   10.11  9.62  8.638  8.479  8.373
```

How many PELS stars do I have?

```
%   AWK   ‘/PELS/  END {print  NR, ” PELS stars”}’  data.txt
```

END : Read through entire file and do the following command
(Similarly there is BEGIN which initiates variables,etc.)

NR : Number of Records (Lines)
(NF : Number of Fields)

Built-in Variables

, : Important to ensure a space is printed

```
700 PELS stars
```

What AWK returns

Total Stars in File

```
%   AWK  ‘END {print  NR, ” stars”}’  data.txt
```

Which stars have a g value greater than 6000?

RA	DEC	J	H	K	B	V	ST	m	g	name		
$1	$2	$3	$4	$5	$6	$7	$8	$9	$10	$11	$12	$13
56.256958	19.559372	8.404	8.225	8.152	9.92	9.42	O	18.1	3201	PELS135		
60.934029	22.944309	8.519	8.262	8.183	10.25	9.67	B	8.6	8901	PELS174		
59.507069	20.676823	8.436	8.248	8.197	10.13	9.63	K	0.72	4262	PELS086		
55.961143	25.268934	8.511	8.284	8.214	10.15	9.63	M	0.12	4423	AK1B146	HD23170	
57.470463	25.647491	8.489	8.332	8.254	10.09	9.57	M	0.33	4158	PELS150	HD23935	TrS144y
52.355843	25.652304	8.459	8.314	8.270	9.90	9.43	F	1.04	6821	AKIII79		
53.530495	24.344501	8.553	8.312	8.274	10.08	9.57	G	0.81	9932	PELS006	TrS4	
56.245522	22.032455	8.520	8.349	8.278	9.97	9.48	A	1.50	4612	PELS140	HD23312	TrS61
55.012829	27.740541	8.638	8.479	8.373	10.11	9.62	A	2.06	5123	PELS025		
57.573582	25.379555	8.645	8.469	8.382	10.14	9.64	G	0.95	4187	TrS151x	HD23975	
58.589973	24.075645	8.687	8.480	8.424	10.29	9.75	K	0.45	1102	AK1A317	HD24463	

```
% AWK '{if ($10 >=6000) print $11, $8, $9,$10}' data.txt
```

name	ST	m	g
PELS174	B	8.6	8901
AKIII79	F	1.04	6821
PELS006	G	0.81	9932
PELS025	A	2.06	5123

Comparison Statements

== : equal to
!= : not equal to
> : greater than
>= : greater than or equal
< : less than
<=: less than or equal

Conditional Statements: If, Else, *For*, While

Unconditional Statements : *Break*, *Continue*, Next, Exit

Which stars have a g value greater than 6000 ?

* A word of caution about comparison statements *

```
%   AWK   {if ($10 >= 6000)  print  $12, $8, $9,$10}  data.txt
```

$10

```
3201
8901
4262
4423
4158
word
9932
4612
5123
4187
1102
```

Will have a BIG problem here

What do you do about it?

```
((( $10 + 0 ) == $10 ) && ( if $10 >= 6000 print $12,$8,$9,$10 ))
```

Combine search statements
&& : AND
|| : OR

If $10 is a **number** than this will hold since it will interpret it as a **1**
If $10 is a **string** than it will interpret it as a 0 so this condition will = 0

```
(( $10 + 0 ) != $10 )
```

Another way to catch strings

Average Mass of M stars

```
awk '/M/ {++ma  avg+=$9/ma} ; END{print "Avg M Mass : ",avg}' data.txt > avgM.dat
```

1. Search for M Stars

```
AK1B146 M 0.12
PELS150  M 0.33
PELS060  M 0.30
PELS040  M 0.24
PELS023  M 0.18
```

2. Increment every match and calculate an average

```
AK1B146 M 0.12 1 0.12
PELS150  M 0.33 2 0.285
PELS060  M 0.30 3 0.385
PELS040  M 0.24 4 0.445
PELS023  M 0.18 5 0.481
```

3. At the END print the average value for M stars

```
Avg M Mass : 0.481
```

4. Output to file called avgM.dat

> output to file

>> append to the file

This is a sub-database that is printed just for an illustration of what AWK is doing for each step

EQUATIONS

RA	DEC	Z
56.256958	19.559372	2.31
60.934029	22.944309	0.20
59.507069	20.676823	0.88
55.961143	25.268934	1.24
57.470463	25.647491	2.88
52.355843	25.652304	10.6
53.530495	24.344501	1.77
56.245522	22.032455	0.98
55.012829	27.740541	0.24
57.573582	25.379555	0.85
58.589973	24.075645	0.77
56.837021	25.525862	1.23

$$x = z \; x \; \cos(ra) \; x \; \cos(dec)$$

$$y = z \; x \; \sin(ra) \; x \; \cos(dec)$$

$$z = z \; x \; \sin(dec)$$

X	Y	Z
1.67807	-0.503885	1.50541
0.03720	0.109683	-0.16305
0.21949	-0.040665	0.851216
1.02251	-0.680972	0.168357
1.51509	1.997146	1.41787
-4.5684	7.986964	5.26291
-1.2385	-0.153639	-1.25513
-0.9345	0.292304	-0.0404688
-0.0072	0.206486	0.122113
0.42792	0.704454	0.207668
-0.1715	0.337265	-0.670618
1.08926	0.323098	0.47118

```
awk '{x = $3*cos($1)*cos($2);  y = $3*sin($1)*cos($2);  z = $3*sin($2) } {print x,y,z}'  radecz.dat > xyz.dat
```

OPERATIONS

+	Addition.
-	Subtraction.
*	Multiplication.
/	Division.
%	Mod.
++	Increment.
--	Decrement.

SHORTHAND

x **+=** 2 ---> x = x + 2

x **-=** 2 ---> x = x - 2

x ***=** 2 ---> x = x * 2

x **/=** 2 ---> x = x / 2

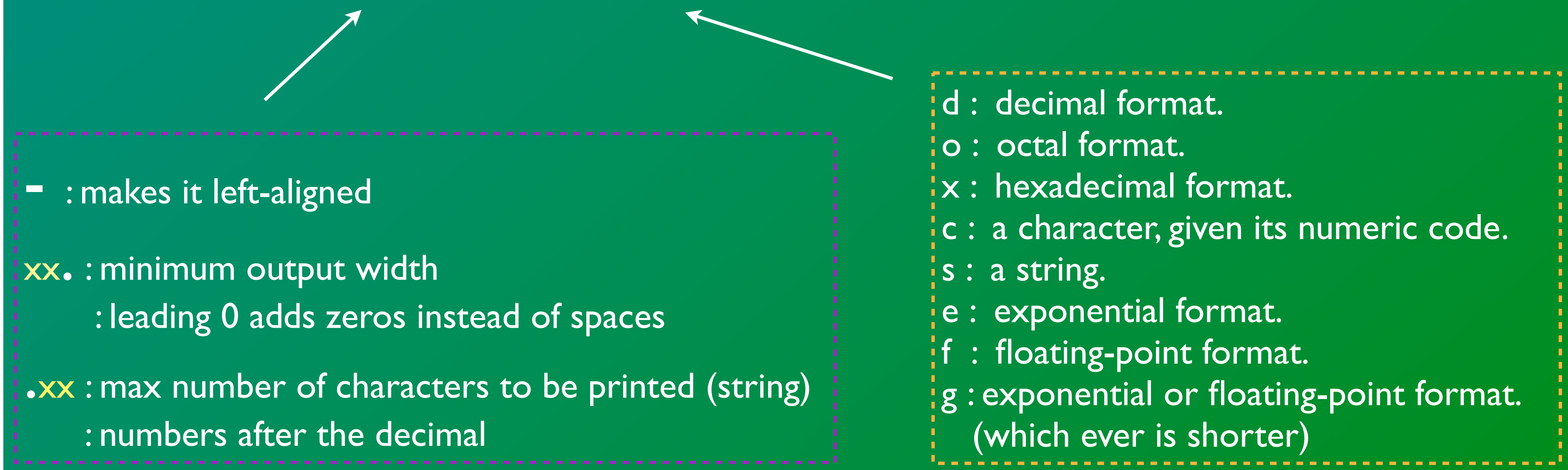
x **%=** 2 ---> x = x % 2

Print Statements

PRINT $5, $6, $8 : prints fields 5, 6 and 8 separated by **spaces**

PRINT $5 $6 $8 : prints fields 5, 6 and 8 concatenated (no spaces)

PRINTF (“[number format code]” , parameters) : General Syntax



- : makes it left-aligned

xx. : minimum output width

: leading 0 adds zeros instead of spaces

.xx : max number of characters to be printed (string)

: numbers after the decimal

d : decimal format.
o : octal format.
x : hexadecimal format.
c : a character, given its numeric code.
s : a string.
e : exponential format.
f : floating-point format.
g : exponential or floating-point format.
(which ever is shorter)

```
RA = 56.256958 ; printf ”(" RA :  %-8.3f\n",RA)”
```

↓

```
RA :  56.256
```

1. How many PELS stars are there?
2. What is the total number of stars that I have?
3. What is average mass of M stars?
4. What is average mass of G stars?

RA	DEC	J	H	K	B	V	ST	m	g	name		
56.256958	19.559372	8.404	8.225	8.152	9.92	9.42	O	18.1	3201	PELS135		
60.934029	22.944309	8.519	8.262	8.183	10.25	9.67	B	8.6	8901	PELS174		
59.507069	20.676823	8.436	8.248	8.197	10.13	9.63	K	0.72	4262	PELS086		
55.961143	25.268934	8.511	8.284	8.214	10.15	9.63	M	0.12	4423	AK1B146	HD23170	
57.470463	25.647491	8.489	8.332	8.254	10.09	9.57	M	0.33	4158	PELS150	HD23935	TrS144y
52.355843	25.652304	8.459	8.314	8.270	9.90	9.43	F	1.04	6821	AKII179		
53.530495	24.344501	8.553	8.312	8.274	10.08	9.57	G	0.81	9932	PELS006	TrS4	
56.245522	22.032455	8.520	8.349	8.278	9.97	9.48	A	1.50	4612	PELS140	HD23312	TrS61
55.012829	27.740541	8.638	8.479	8.373	10.11	9.62	A	2.06	5123	PELS025		
57.573582	25.379555	8.645	8.469	8.382	10.14	9.64	G	0.95	4187	TrS151x	HD23975	
58.589973	24.075645	8.687	8.480	8.424	10.29	9.75	K	0.45	1102	AK1A317	HD24463	
56.837021	25.525862	8.720	8.463	8.435	10.41	9.85	M	0.30	3124	PELS060	HD23598	
60.221325	23.193913	8.676	8.558	8.488	10.06	9.61	F	1.08	8019	PELS173		
55.599987	21.473492	8.797	8.604	8.526	10.40	9.85	O	17.8	7162	PELS035	TrS42	
53.882030	22.823627	8.831	8.599	8.541	10.40	9.86	G	1.02	4331	PELS124		
56.291157	21.704592	8.902	8.694	8.586	10.52	9.95	M	0.24	3131	PELS040		
55.400673	25.619331	8.959	8.657	8.593	10.69	10.10	M	0.18	9881	PELS023	DH181	

An Awk Program

info.awk

```
/PELS/ {++num_PELS}                          #  Number of PELS stars
/M/ {++num_M;  mass_M+=$9}                   #  Number of M stars and masses
/G/ {++num_G  mass_G=$9}                     # Number of G stars and total mass
END { avg_M = mass_M / num_M;                # Compute Average M Mass
      avg_G = mass_G / num_G;                # Compute Average G Mass
      total = NR ;                           #  Total Number of Stars
      print "Data Summary:";                 # Print results.
      printf ("\n");
      printf ("   Total Number of Stars:              %2d\n", total);
      printf ("   Number of PELS Stars:               %5.2f\n", num_PELS);
      printf ("\n");
      printf ("   Avg. Mass of M Stars:               %2d\n", avg_M);
      printf ("   Avg. Mass of G Stars:               %5.2f\n", avg_G);
      printf ("\n");    }
```

```
AWK    -f info.awk     data.txt
```

An Awk Program

1. How many PELS stars are there?
2. What is the total number of stars that I have?
3. What is average mass of M stars?
4. What is average mass of G stars?

Questions you asked

```
Data Summary:

Total Number of Stars:              18
Number of PELS Stars:               13

Avg. Mass of M Stars:           0.2340
Avg. Mass of G Stars:           0.9267
```

What Awk Returns

Websites

http://www.vectorsite.net/tsawk_1.html ***awesome***

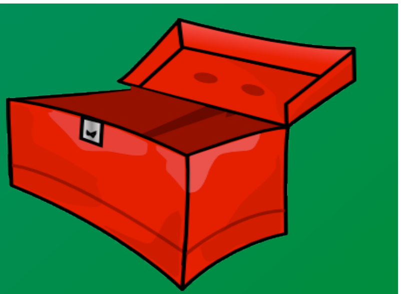
http://oreilly.com/catalog/unixnut3/chapter/ch11.html

http://www.grymoire.com/Unix/Awk.html

http://www.cs.sjsu.edu/web_mater/cs46b/cs46blab/awk.html

http://www.gnu.org/manual/gawk/html_node/Very-Simple.html

Recap of your UNIX tool box

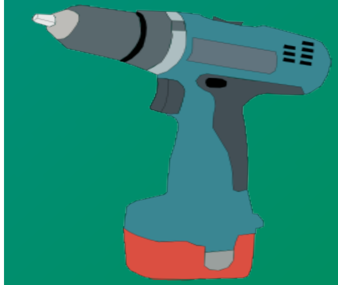


GREP

```
GREP  -n  -B3  -A2  ‘oplot’ code.txt
```

Uses

- Finding pieces of code
- Finding files has specific uses
- Looking through data files



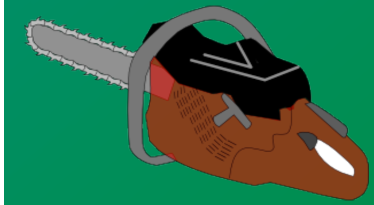
SED

```
SED 's_http://www.foo.com/_http://bar.net/_'  file.txt
```

Uses

- Replace a common mistake throughout your files
- Inserting multiple characters through each line

(remember your favorite sed command)



AWK

```
AWK -f info.awk g=9.81  data.txt
```

Uses

- Find out a bunch of information in my files
- Using it to do equations
  - Converting RA and DEC
  - Converting between filter sets
- Looking through data files

A Few Tips

- If you are writing on the command line you need to use single quotes instead of double

- CTRL-D : escapes you out if AWK is hung up