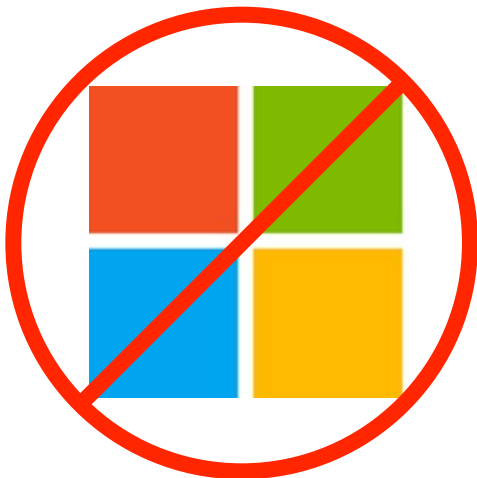# AWK

Why use awk?

- Many datasets come in text files arranged into rows (records) and columns (fields)
- Tasks for this dataset:
  - organize
  - extract (some or all)
  - do math on it
- Awk does all of this and (much) more
- Plus, it's fast and portable

Why not Excel?

- It's made by Microsoft

- Requires payment (awk is FREE!)

- Have to open separate program, import data, and export after analysis

- Much slower, especially for big files

Before you start...

- There are a couple datasets used in this tutorial. You can download them at:

  https://geodyn.psu.edu/tutorials/awk_datasets.zip

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Who was the opponent?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Who was the opponent?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
```

```
awk '{print $3}' 2014_orioles.txt
```

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Who was the opponent?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
```

```
awk '{print $3}' 2014_orioles.txt
```

"Print the third field (column) in each record (line) to the terminal"

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What about the date in addition to the opponent?


```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1  Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2  Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3  Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4  Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5  Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6  Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7  Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8  Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9  Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What about the date in addition to the opponent?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

awk '{print $2,$3}' 2014_orioles.txt

"Print the second and third fields separated by a space to the terminal"

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What about the date in addition to the opponent?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

```
awk '{print $2,$3}' 2014_orioles.txt
```

Try putting a space instead of a comma (I make this mistake a lot).

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Rearrange the dataset to W-L record then date

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Rearrange the dataset to W-L record then date

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

awk '{print $7,$2}' 2014_orioles.txt

The fields don't need to be in order, and they can be repeated

Print Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What if we wanted everything?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

Print Statement

- Baltimore Orioles 2014 season results

    (baseball-reference.com)
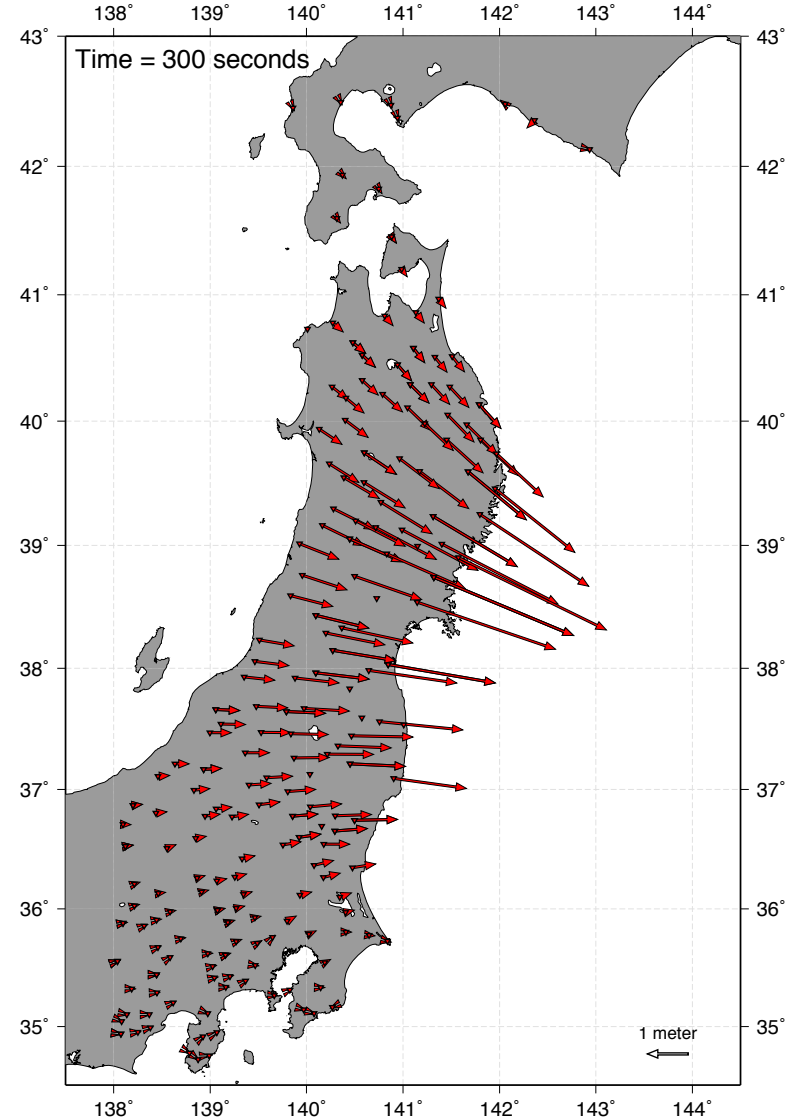
- What if we wanted everything?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

awk '{print $0}' 2014_orioles.txt

"$0" stands for the entire record

Math in Awk

- Displacements from 2011 M 9.0 Tohoku earthquake

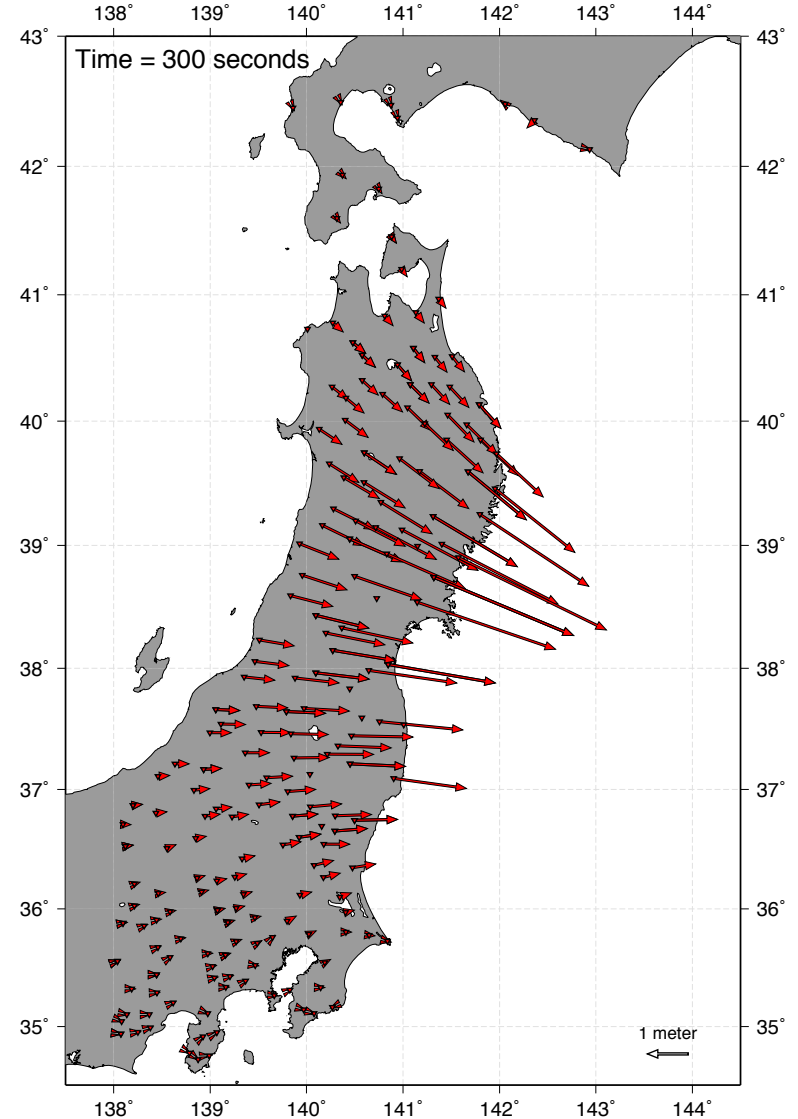- What do we need to make this figure?

Math in Awk

- Location, direction, and magnitude

```
0016 142.365 42.3569 0.0492 0.0125 -0.0808
0017 139.858 42.4507 -0.0781 0.0162 -0.1077
0018 140.942 42.3742 -0.0684 0.0156 -0.0664
0020 139.52 42.1535 -0.0452 0.0478 -0.1154
0021 140.107 42.0081 -0.0850 0.0285 -0.0719
0022 140.748 41.8261 -0.1011 0.0600 -0.0485
0023 140.041 41.4664 -0.1293 0.0876 -0.0408
0024 141.213 41.3008 -0.1236 0.0754 -0.1034
0026 140.273 40.779 -0.2116 0.2328 -0.0321
0027 141.789 40.1334 -0.5972 0.5436 -0.1059
0028 141.94 39.5724 -1.3633 1.5499 -0.3729
0029 141.204 39.1106 -1.2539 2.2863 -0.2054
0030 139.776 39.9679 -0.2606 0.4403 -0.0303
0031 140.048 39.3987 -0.3954 0.7776 -0.0068
0032 139.809 38.8946 -0.2658 0.8980 -0.0413
```

Station ID   Coordinates   N E Z
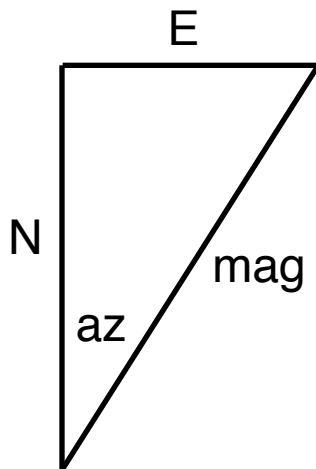   Displacement

Math in Awk

- Addition   x+y
- Subtraction   x−y
- Multiplication   x*y
- Division   x/y
- Modulo   x%y
- Exponentiation   x^y
- Square root   sqrt(x)
- Trig functions   cos(x), sin(x), atan2(y,x)
- Exponential fns   exp(x), log(x)

Arguments are in radians, not degrees. No tangent function!

Natural logarithm

Math in Awk

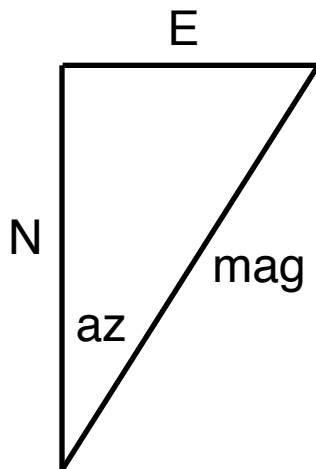- Location, direction, and magnitude


```
0016 142.365 42.3569 0.0492 0.0125 -0.0808
0017 139.858 42.4507 -0.0781 0.0162 -0.1077
0018 140.942 42.3742 -0.0684 0.0156 -0.0664
0020 139.52 42.1535 -0.0452 0.0478 -0.1154
0021 140.107 42.0081 -0.0850 0.0285 -0.0719
0022 140.748 41.8261 -0.1011 0.0600 -0.0485
0023 140.041 41.4664 -0.1293 0.0876 -0.0408
0024 141.213 41.3008 -0.1236 0.0754 -0.1034
0026 140.273 40.779 -0.2116 0.2328 -0.0321
0027 141.789 40.1334 -0.5972 0.5436 -0.1059
0028 141.94 39.5724 -1.3633 1.5499 -0.3729
0029 141.204 39.1106 -1.2539 2.2863 -0.2054
0030 139.776 39.9679 -0.2606 0.4403 -0.0303
0031 140.048 39.3987 -0.3954 0.7776 -0.0068
0032 139.809 38.8946 -0.2658 0.8980 -0.0413
```

Station ID Coordinates   N E Z Displacement

Math in Awk

- Location, direction, and magnitude

```
0016 142.365 42.3569 0.0492 0.0125 -0.0808
0017 139.858 42.4507 -0.0781 0.0162 -0.1077
0018 140.942 42.3742 -0.0684 0.0156 -0.0664
0020 139.52 42.1535 -0.0452 0.0478 -0.1154
0021 140.107 42.0081 -0.0850 0.0285 -0.0719
0022 140.748 41.8261 -0.1011 0.0600 -0.0485
0023 140.041 41.4664 -0.1293 0.0876 -0.0408
0024 141.213 41.3008 -0.1236 0.0754 -0.1034
0026 140.273 40.779 -0.2116 0.2328 -0.0321
0027 141.789 40.1334 -0.5972 0.5436 -0.1059
0028 141.94 39.5724 -1.3633 1.5499 -0.3729
0029 141.204 39.1106 -1.2539 2.2863 -0.2054
0030 139.776 39.9679 -0.2606 0.4403 -0.0303
0031 140.048 39.3987 -0.3954 0.7776 -0.0068
0032 139.809 38.8946 -0.2658 0.8980 -0.0413
```
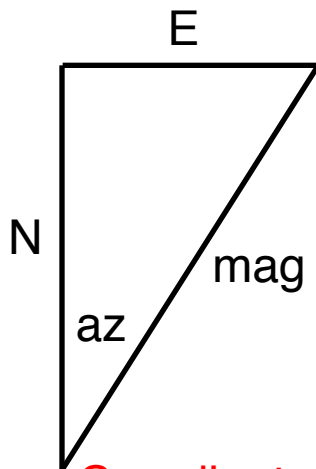
Station ID   Coordinates   N E Z Displacement

E

N

mag

az

```
awk '{
    print $2,$3,atan2($5,$4),sqrt($4*$4+$5*$5)
}' TIME_300.dat
```

Math in Awk

- Location, direction, and magnitude

```
0016 142.365 42.3569 0.0492 0.0125 -0.0808
0017 139.858 42.4507 -0.0781 0.0162 -0.1077
0018 140.942 42.3742 -0.0684 0.0156 -0.0664
0020 139.52 42.1535 -0.0452 0.0478 -0.1154
0021 140.107 42.0081 -0.0850 0.0285 -0.0719
0022 140.748 41.8261 -0.1011 0.0600 -0.0485
0023 140.041 41.4664 -0.1293 0.0876 -0.0408
0024 141.213 41.3008 -0.1236 0.0754 -0.1034
0026 140.273 40.779 -0.2116 0.2328 -0.0321
0027 141.789 40.1334 -0.5972 0.5436 -0.1059
0028 141.94 39.5724 -1.3633 1.5499 -0.3729
0029 141.204 39.1106 -1.2539 2.2863 -0.2054
0030 139.776 39.9679 -0.2606 0.4403 -0.0303
0031 140.048 39.3987 -0.3954 0.7776 -0.0068
0032 139.809 38.8946 -0.2658 0.8980 -0.0413
```

Station ID   Coordinates   N E Z Displacement

E

N

az   mag

Coordinates are easy

Returns angle in radians!   Pythagoras strikes again

```
awk '{
  print $2,$3,atan2($5,$4),sqrt($4*$4+$5*$5)
}' TIME_300.dat
```

Note the different formatting; awk is somewhat flexible in this regard.

Conditional Statement

- Grade assignments

  >= 100: A+

  90-100: A

  80-90: B

  70-80: C

  60-70: D

  < 60: F

```
Last_name Percentage
Ammon 102
Anandakrishnan 97
Beethoven 69
Columbus 71
Einstein 76
Gaga 64
Gandhi 85
Herman 63
Leeman 87
Mandela 82
Picasso 94
Presley 24
```

Conditional Statement

- Grade assignments

  >= 100: A+

  90-100: A

  80-90: B

  70-80: C

  60-70: D

  < 60: F

```
Last_name Percentage
Ammon 102
Anandakrishnan 97
Beethoven 69
Columbus 71
Einstein 76
Gaga 64
Gandhi 85
Herman 63
Leeman 87
Mandela 82
Picasso 94
Presley 24
```

```
awk '{if ($2>=100) print $1,"A+"}' grades.txt
```

Conditional Statement

- Grade assignments

  >= 100: A+

  90-100: A

  80-90: B

  70-80: C

  60-70: D

  < 60: F

Only print to terminal if the second field is greater than or equal to 100

```
awk '{if ($2>=100) print $1,"A+"}' grades.txt
```

To print text, enclose it in double quotes

Last_name	Percentage
Ammon	102
Anandakrishnan	97
Beethoven	69
Columbus	71
Einstein	76
Gaga	64
Gandhi	85
Herman	63
Leeman	87
Mandela	82
Picasso	94
Presley	24

Conditional Statement

• Grade assignments

```
awk '{
  if ($2>=100) print $1,"A+"
  else if ($2>=90) print $1,"A"
  else if ($2>=80) print $1,"B"
  else if ($2>=70) print $1,"C"
  else if ($2>=60) print $1,"D"
  else print $1,"F"
}' grades.txt
```

```
Last_name Percentage
Ammon 102
Anandakrishnan 97
Beethoven 69
Columbus 71
Einstein 76
Gaga 64
Gandhi 85
Herman 63
Leeman 87
Mandela 82
Picasso 94
Presley 24
```

Conditional Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What were the attendance values?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
```

Conditional Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What were the attendance values?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
```

```
awk '{
    if (substr($3,1,1) != "@")
        print $11
}' 2014_orioles.txt
```

Conditional Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What were the attendance values?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

`Gm# Date Opp W/L R RA W-L Win Loss Time Attendance`

```
awk '{
   if (substr($3,1,1) != "@")
      print $11
}' 2014_orioles.txt
```

"substr" takes three arguments:
(1) the field, (2) the position to
start, (3) number of characters

Conditional Statement

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What were the attendance values?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
```

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
```

```
awk '{
    if (substr($3,1,1) != "@") {
        {print $11}
    }
}' 2014_orioles.txt
```

Home Games

BEGIN and END

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- What was the average home attendance?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

```
awk '
BEGIN{tot=0;count=0}
{if (substr($3,1,1) != "@") {tot=tot+$11;count=count+1}}
END{print tot/count}' 2014_orioles.txt
```

BEGIN and END

- Baltimore Orioles 2014 season results

  (baseball-reference.com)

- Why do we need BEGIN and END here?

```
Gm# Date Opp W/L R RA W-L Win Loss Time Attendance
1 Monday_Mar_31 BOS W 2 1 1-0 Britton Lester 2:53 46685
2 Wednesday_Apr_2 BOS L 2 6 1-1 Lackey Jimenez 2:43 25708
3 Thursday_Apr_3 BOS L 3 4 1-2 Doubront Chen 2:55 20880
4 Friday_Apr_4 @DET L 4 10 1-3 Smyly Gonzalez 3:07 23625
5 Saturday_Apr_5 @DET L 6 7 1-4 Porcello Norris 3:20 32041
6 Sunday_Apr_6 @DET W 3 1 2-4 Tillman Verlander 2:34 34261
7 Monday_Apr_7 @NYY L 2 4 2-5 Kuroda Jimenez 3:05 48142
8 Tuesday_Apr_8 @NYY W 14 5 3-5 Chen Nova 3:28 35864
9 Wednesday_Apr_9 @NYY W 5 4 4-5 Matusz Kelley 3:06 39412
10 Friday_Apr_11 TOR L 0 2 4-6 McGowan Tillman 2:20 22327
11 Saturday_Apr_12 TOR W-wo 2 1 5-6 Britton Redmond 3:39 30446
12 Sunday_Apr_13 TOR L 3 11 5-7 Buehrle Jimenez 2:45 39281
13 Monday_Apr_14 TBR W 7 1 6-7 Chen Archer 2:46 15799
14 Wednesday_Apr_16 TBR W 3 0 7-7 Gonzalez Odorizzi 2:53 22611
15 Friday_Apr_18 @BOS W 8 4 8-7 Tillman Lackey 3:26 36408
16 Saturday_Apr_19 @BOS L 2 4 8-8 Tazawa Norris 2:47 37689
17 Sunday_Apr_20 @BOS L-wo 5 6 8-9 Mujica Matusz 3:32 33947
18 Monday_Apr_21 @BOS W 7 6 9-9 Chen Buchholz 3:18 37513
19 Tuesday_Apr_22 @TOR L 3 9 9-10 Delabar Meek 2:54 14866
20 Wednesday_Apr_23 @TOR W 10 8 10-10 Tillman Redmond 3:11 15202
21 Thursday_Apr_24 @TOR W 11 4 11-10 Norris Cecil 2:54 16283
22 Friday_Apr_25 KCR L 0 5 11-11 Ventura Jimenez 2:40 22478
23 Saturday_Apr_26 KCR W-wo 3 2 12-11 Britton Duffy 3:09 34941
24 Sunday_Apr_27 KCR L 3 9 12-12 Shields Gonzalez 3:08 38368
25 Thursday_May_1_(1) PIT W 5 1 13-12 Norris Morton 3:08 0
```

```
awk '                          Without BEGIN, variables tot and count are uninitialized
BEGIN{tot=0;count=0}
{if (substr($3,1,1) != "@") {tot=tot+$11;count=count+1}}
END{print tot/count}' 2014_orioles.txt
```

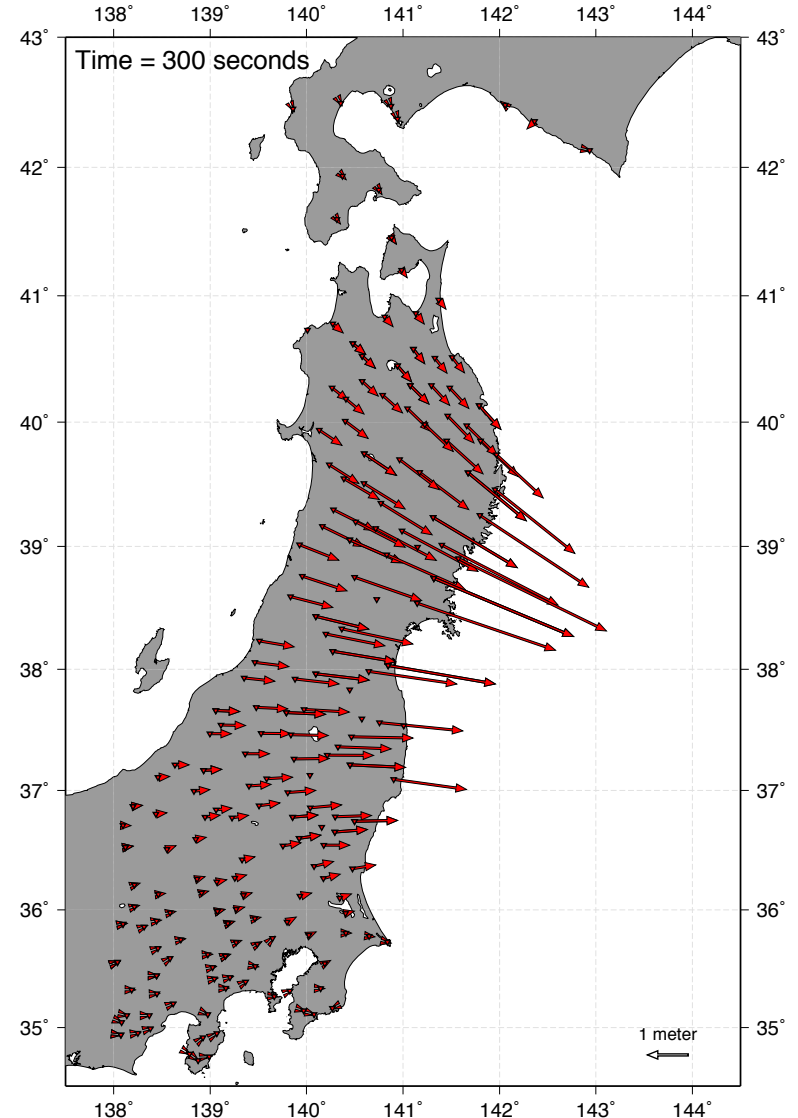END allows you to do operations after finishing file

Maximum Displacement

- Largest horizontal displacement

```
0016 142.365 42.3569  0.0492 0.0125 -0.0808
0017 139.858 42.4507 -0.0781 0.0162 -0.1077
0018 140.942 42.3742 -0.0684 0.0156 -0.0664
0020 139.52  42.1535 -0.0452 0.0478 -0.1154
0021 140.107 42.0081 -0.0850 0.0285 -0.0719
0022 140.748 41.8261 -0.1011 0.0600 -0.0485
0023 140.041 41.4664 -0.1293 0.0876 -0.0408
0024 141.213 41.3008 -0.1236 0.0754 -0.1034
0026 140.273 40.779  -0.2116 0.2328 -0.0321
0027 141.789 40.1334 -0.5972 0.5436 -0.1059
0028 141.94  39.5724 -1.3633 1.5499 -0.3729
0029 141.204 39.1106 -1.2539 2.2863 -0.2054
0030 139.776 39.9679 -0.2606 0.4403 -0.0303
0031 140.048 39.3987 -0.3954 0.7776 -0.0068
0032 139.809 38.8946 -0.2658 0.8980 -0.0413
```
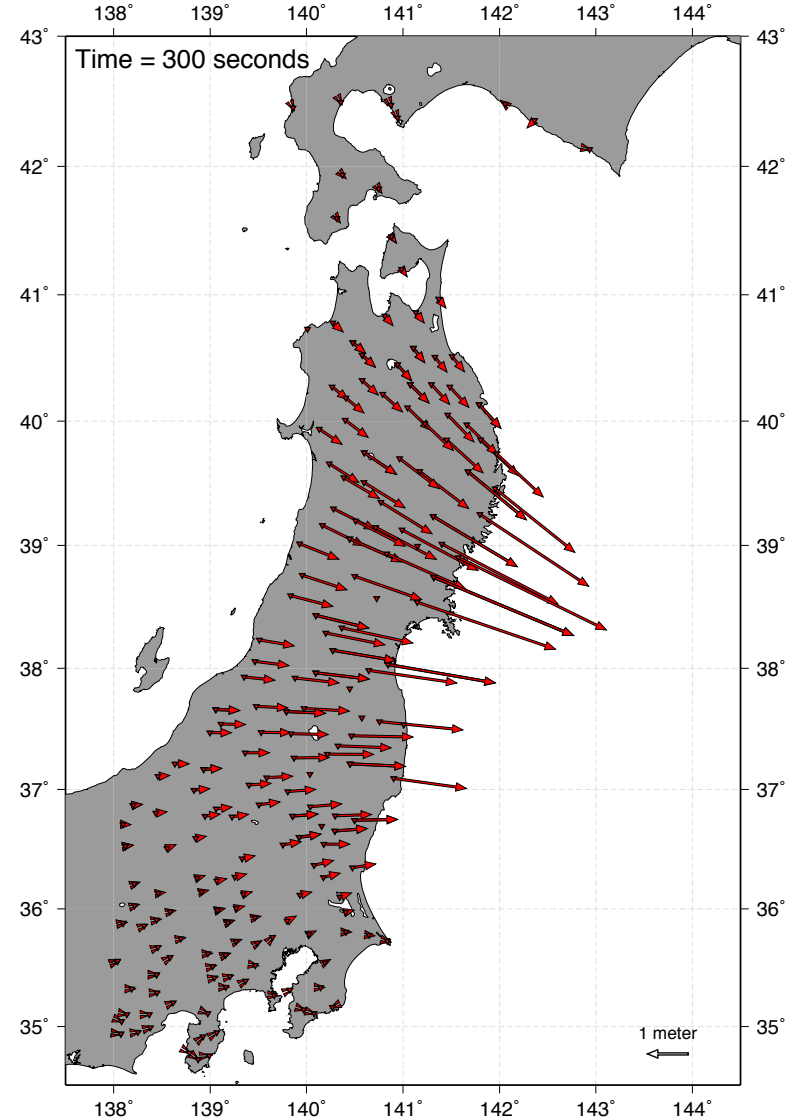
Station ID   Coordinates   N E Z
                                    Displacement

Time = 300 seconds

1 meter

Maximum Displacement

```
awk '
  BEGIN{max=0}
  {
  mag = sqrt($4*$4+$5*$5)
  if (mag > max) {
    stid = $1
    stlo = $2
    stla = $3
    max = mag
  }
  }
  END{
    print "STID",stid
    print "STLO",stlo
    print "STLA",stla
    print "MAX",max
  }' TIME_300.dat
```


Time = 300 seconds

1 meter

Last Thing: Delimiters

- What if fields are separated by something besides whitespace?
  - Contents of *file.csv*: 1,2,3,4,5,6,7,8,9

- Use -F*fs* (where *fs* is the file separator)

- For a comma-separated variable (.csv) file:
```
awk –F, '{print $3}' file.csv
```

What's Next?

- Piping into awk (or from awk)

```
echo 3 4| awk '{print sqrt($1*$1+$2*$2)}'
```

Echo prints text to the terminal.
The vertical line "pipes" the output of echo into awk.

- Scripting
- Other powerful Unix commands (sed, grep)
- Regular expressions (steep learning curve, incredible utility)
- Plotting data (GMT, Python, gnuplot, etc.)