Now that we know how to select specific columns, we can select a column and pipe it to the **csvstat**tool to calculate summary statistics for that column:



csvcut -c 4 Combined\_hud.csv | csvstat

This calculates a full suite of summary statistics, including:

* max,
* min,
* sum,
* mean,
* median,
* standard deviation.

Depending on the size of the data, the full summary statistics for a column can take a long time and you often just want a specific summary statistic. You can use -- flags to choose specific summary statistics, which will greatly improve the speed:



# Just the max value.

csvcut -c 2 Combined\_hud.csv | csvstat --max

# Just the mean value.

csvcut -c 2 Combined\_hud.csv | csvstat --mean

# Just the number of null values.

csvcut -c 2 Combined\_hud.csv | csvstat --nulls

You can see a full list of flags in the [documentation](http://csvkit.readthedocs.io/en/0.9.1/scripts/csvstat.html#description). If you want to calculate summary statistics over all the columns in a CSV file, you can pass the file to csvstat directly:



csvstat Combined\_hud.csv

Instructions

* Use csvstat to calculate just the mean for each column in Combined\_hud.csv.

/home/dq$ csvstat Combined\_hud.csv --mean

1. year: 2008.9044232628457

2. AGE1: 46.511215505103266

3. BURDEN: 5.303764743668771

4. FMR: 1037.1186695822005

5. FMTBEDRMS: None

6. FMTBUILT: None

7. TOTSAL: 44041.841931779105