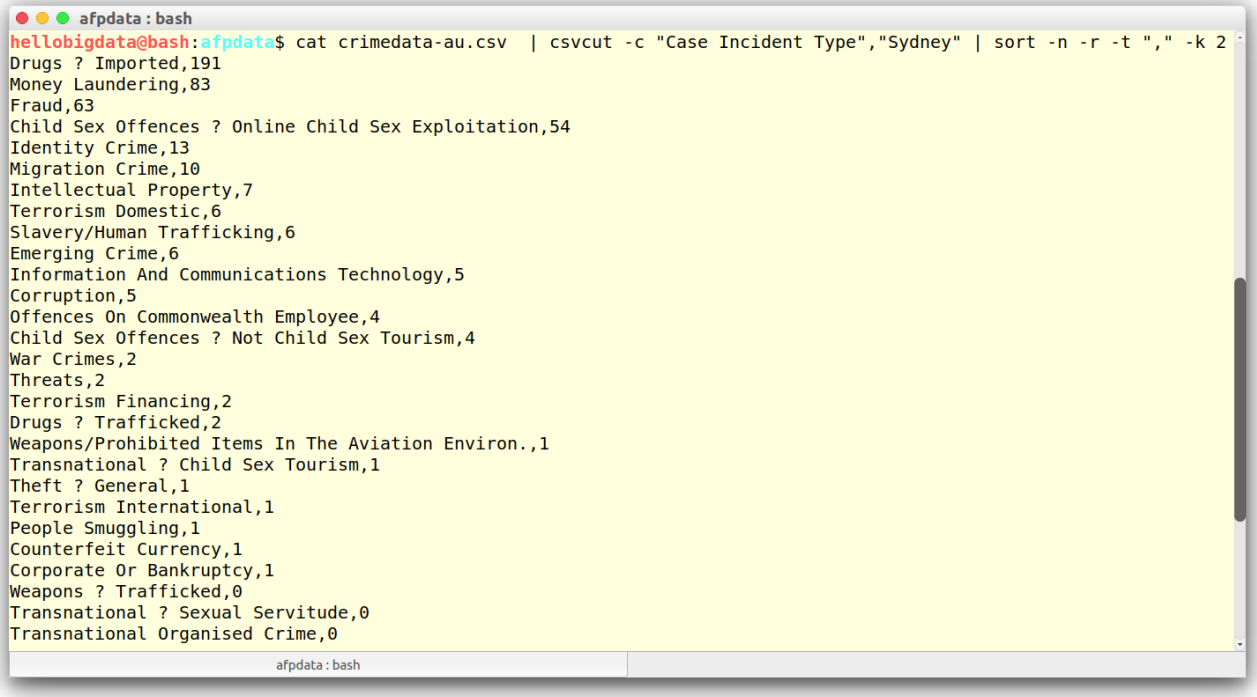
Finding the top most crime per city (csvcut, sort, tail, awk)

Our next compute quest is to find the total number of crimes, given a city name. Let's watch the following video lecture first!

Video lecture: Finding the top most crime per city

This essentially will require us to compute the summation of a column which contains the city names. Ideally, we would like to have a function like `top_crime ()` that would take an argument (e.g., Sydney) and tell us the top crime for that city. Let's start with a basic version first:

```
cat crimedata-au.csv  | \
csvcut -c "Case Incident Type","Sydney" | \
sort -n -t "," -k 2
```



Finding the total number of crimes, for Sydney

Note that the `csvcut` function can take column titles with the `-c` option. Now, if we just want the top crime we add `tail -n 1` at the end, as follows:

```
cat crimedata-au.csv  | \
csvcut -c "Case Incident Type","Sydney" |  \
sort -n -t "," -k 2 | \
tail -n 1
```

Now, let's build our desired function. For knowing more about bash function, please refer to the tutorials section.

```
function topcrime() { cat $1 | \
  csvcut -c "Case Incident Type",$2 | \
  sort -n -t "," -k 2 |  tail -n 1| \
  awk -F',' '{print $1 "= " $2}' ; }

topcrime crimedata-au.csv Sydney
```

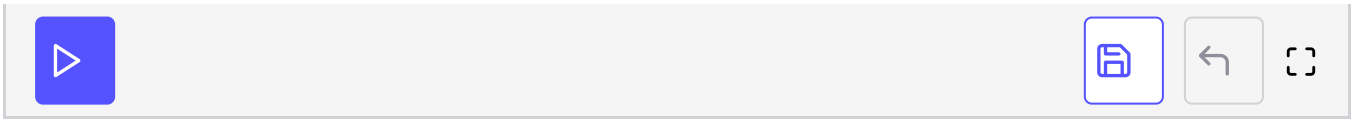
Note that the first (`crimedata-au.csv`) and second (city name: `Sydney`)

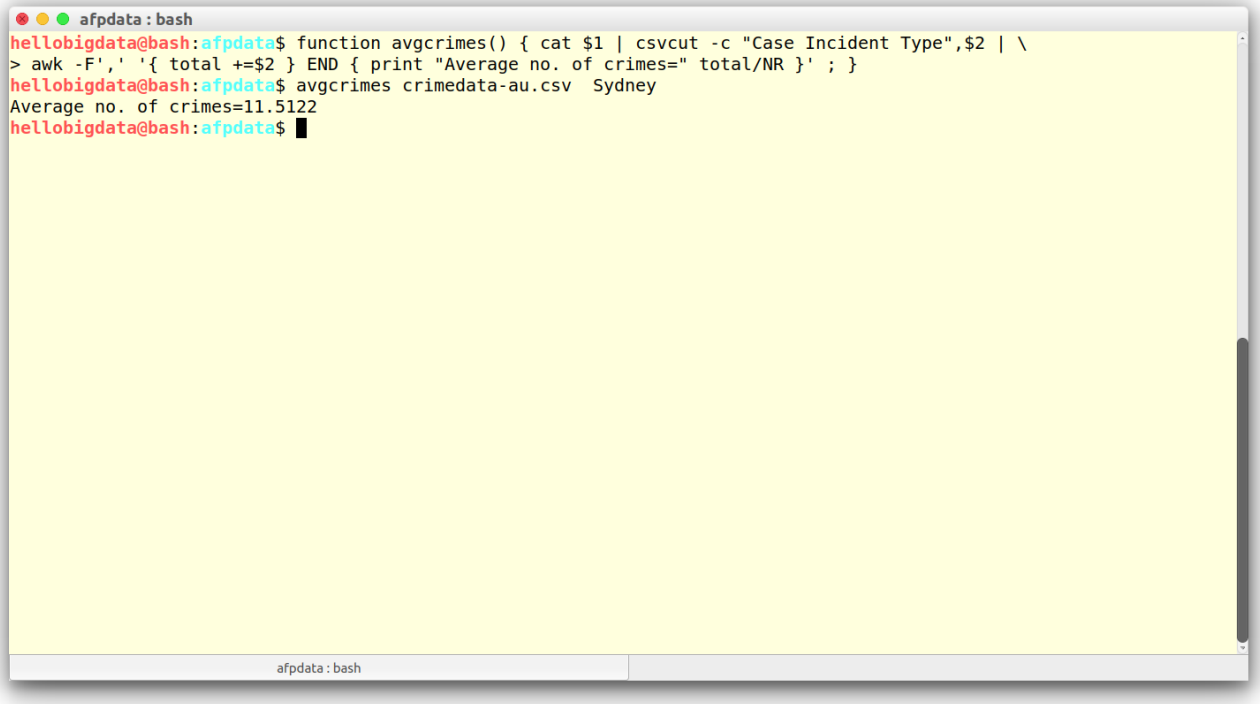
arguements were automatically placed into the `$1` and `$2` variables inside

the function, however indside `awk`'s call `$1` and `$2` represented the first (crime type) and second (number of crimes) columns of the piped output.

Now, we can slightly modify the above function so that it can also compute, given any city name, the average number of crimes.

```
function avgcrimes() { cat $1 | \
csvcut -c "Case Incident Type",$2 | \
awk -F',' '{ total +=$2 } END { print "Average no. of crimes=" total/NR }' ; }
avgcrimes crimedata-au.csv Sydney
```





Finding the average number of crimes, for Sydney

Here, everything remains as above, except there's no sort function and the `awk` function computes the sum of crimes (col 2) and then divide the sum by the total number of rows (`NR`).