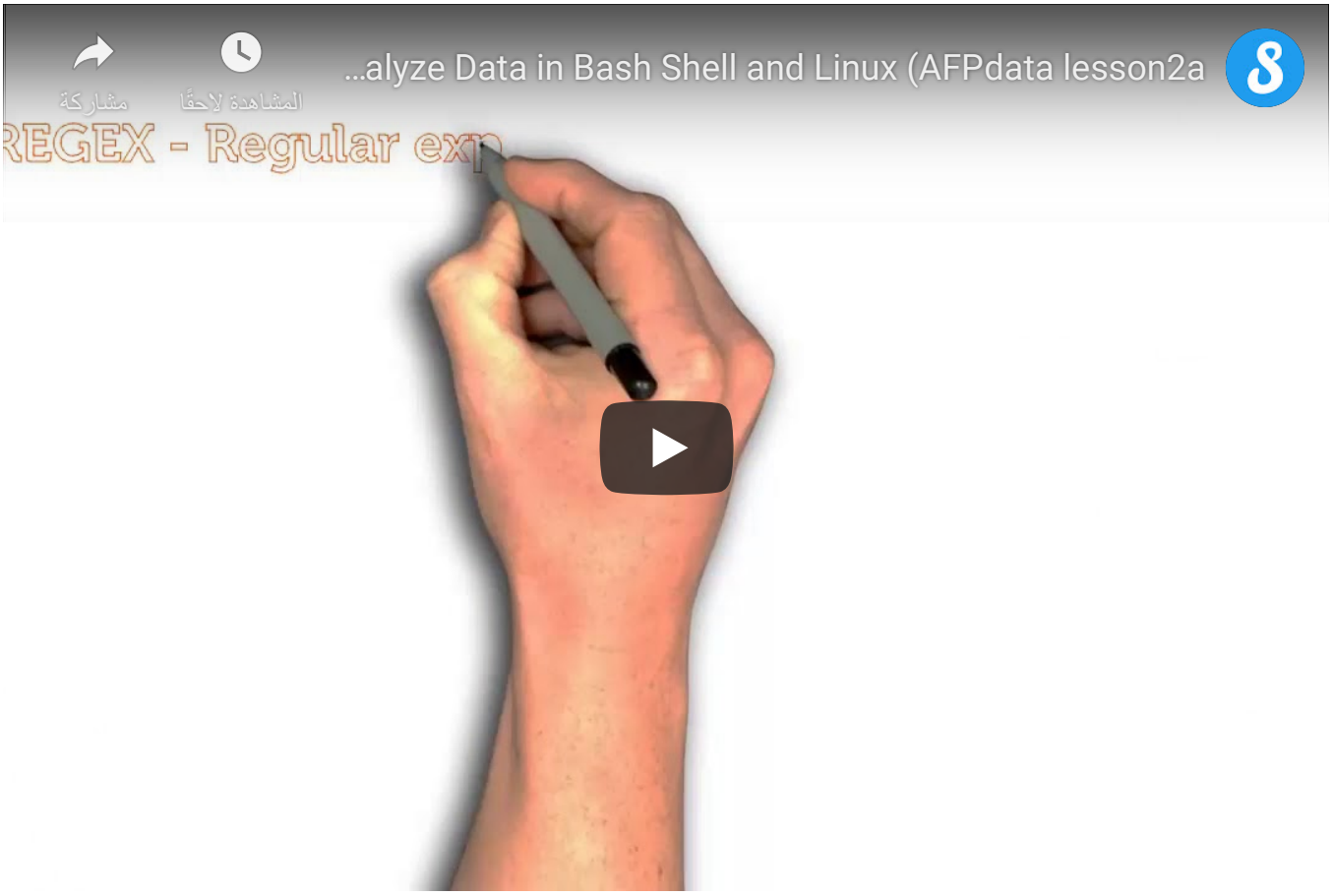
Row and Columns Statistics (sed, wc, head, cat)

WE'LL COVER THE FOLLOWING

Now we want to get some some statistics. Let's first find the total number of columns in the file. There are two ways to do this, let's first check the hard way:



Video lecture: Finding some data stats

The hard way

Count columns:

Step 1. First get only the first row using `head` command:

```
head -1 crimedata-au.csv
```

Step 2. Next use sed to remove everything except commas. To do this, we use the regex (regular expression) pattern `[^,]`, See the appendix for more details on the regex.

```
head -1 crimedata-au.csv | sed 's/[^,]//g'
```

Step 3. All what has left is to simply use `wc` command to count number of characters (commas).

```
echo "Cols stat"
head -1 crimedata-au.csv | sed 's/[^,]//g' | wc -c
```

Count rows:

Now, let's find the total number of rows in the file well:

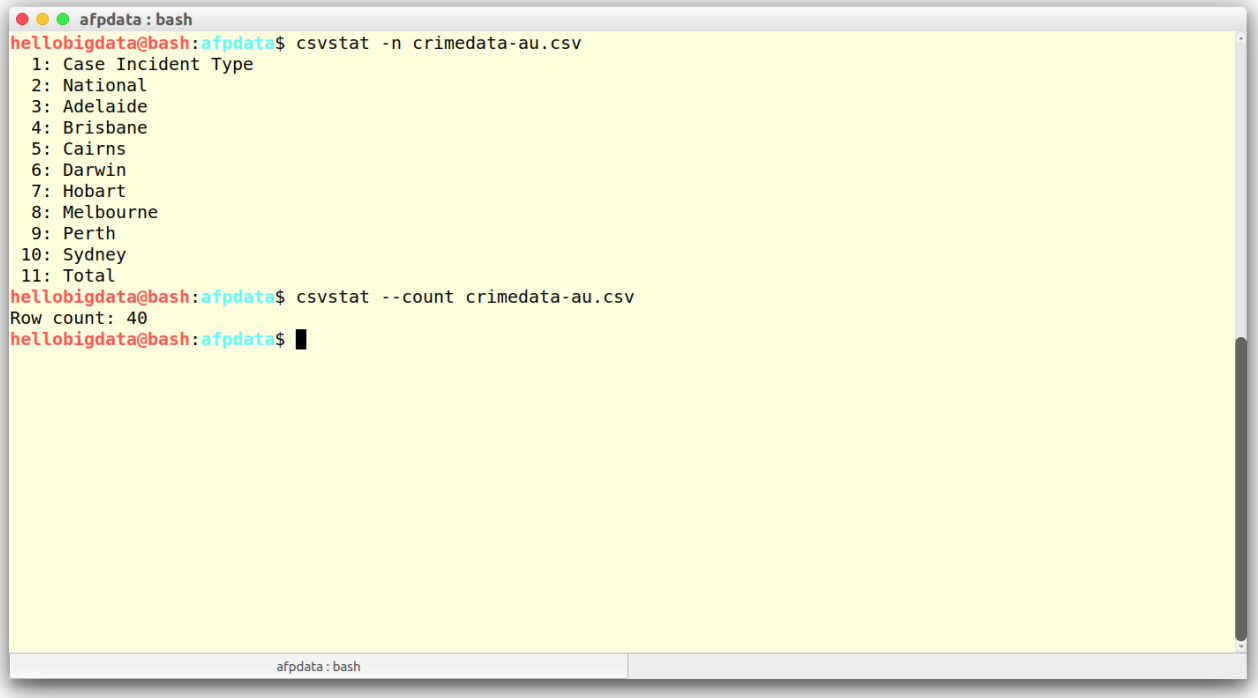
```
echo "Rows stat:"
cat crimedata-au.csv | wc -l
```

The easy way

As the previous chapter, we could also use the `csvkit`'s command `csvstat` to get the same stats, but with much simpler way:

```
echo "Cols stat:"
csvstat -n crimedata-au.csv
```

```
echo "Rows stat:"
csvstat --count crimedata-au.csv
```



Use `csvstat` to find #rows and #cols.

Note that the `csvstat` is smarter that `wc -l` as it omitted the titles line and counted only the data rows!

Do you want to know more?

'sed' man page