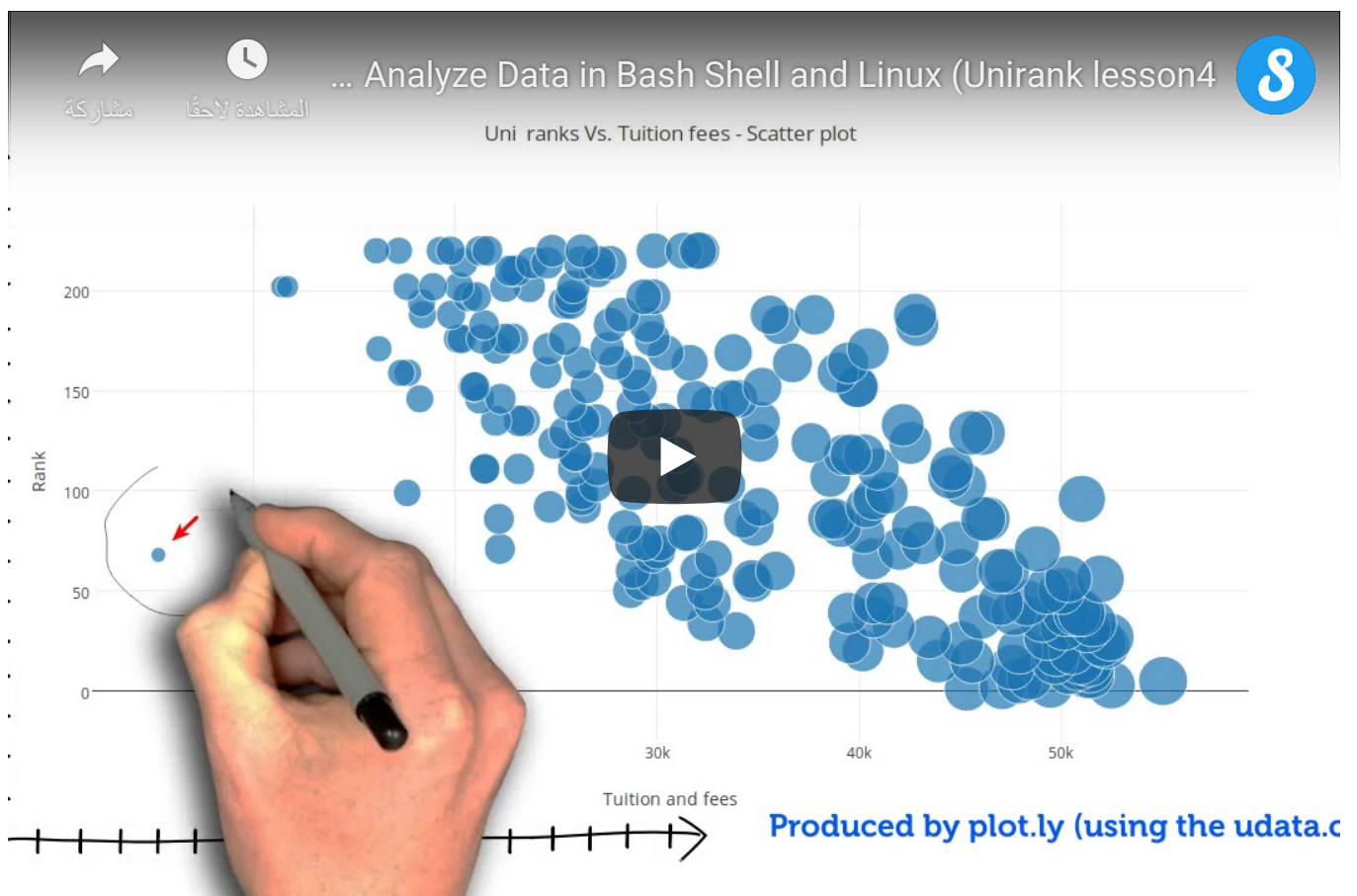


Finding the correlation between university tuition and ranks

WE'LL COVER THE FOLLOWING ^

- Do you want to know more?

We already know the ranks and the tuition fees per university (given). An interesting question to investigate would be to find what's the correlation of uni ranks with tuition and fees?



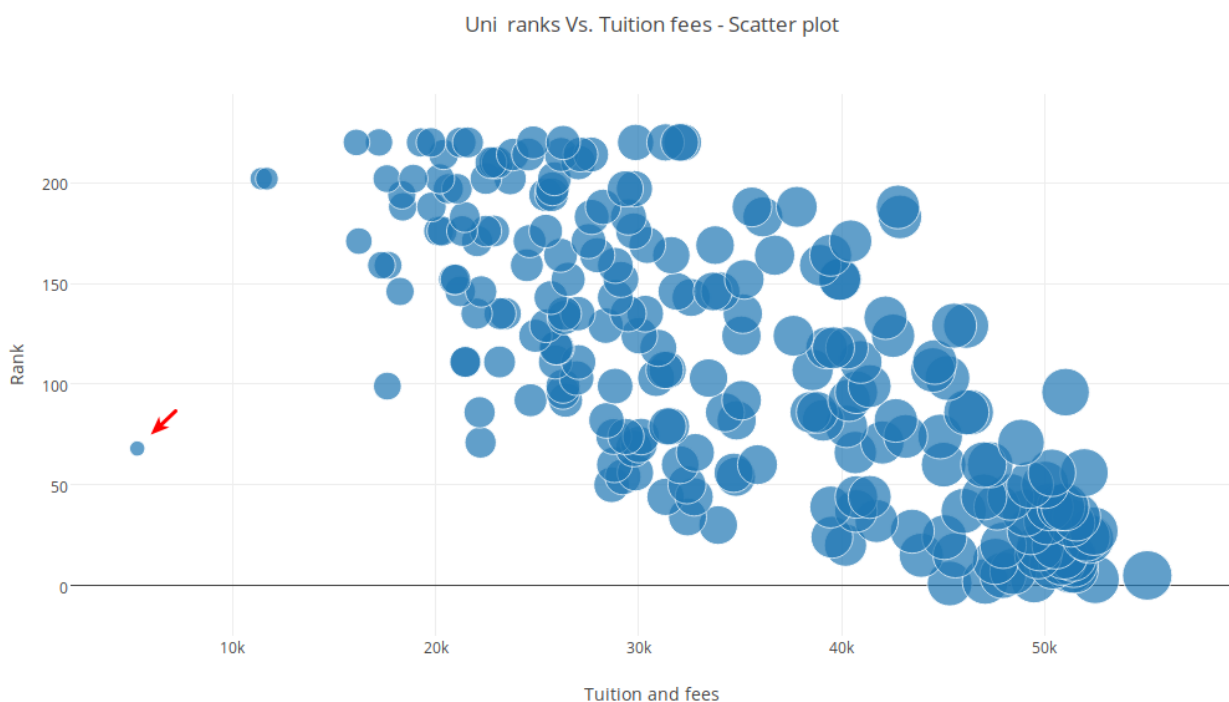
To carry out this, we first `cat` and `cut` out the `Tuition and fees` (col 4) and `Ranks` (col 6) from the data into the new dataset called `udata.csv`:

```
cat unirank.csv | cut -f4,6 -d, > udata.csv
```

Note that the redirection symbol (`>`), helped us to save the output. Now this data can simply be plotted using a scatterplot tool called `scatter` (install `sudo pip install bashplotlib`).

```
cat udata.csv | tail -n +2 | scatter
```


Note that `tail -n +2` excludes the first row i.e., column titles prior to passing the output all the way to end to `scatter` . However, this tool's output doesn't make much sense, as it doesn't show any x,y- axes legends. Therefore, we uploaded the data (`udata.csv`) to an online tool called plot.ly, which produced the following beautiful scatter plot:




Ranks vs. tuition: a scatterplot via plot.ly

It's a no brainer to understand from the plot above that highly ranked universities have higher tuition fees! However, the scatterplot also depicts one university (*Brigham Young University-Provo*) that had a higher rank (`rank=68`) with an extremely low tuition fees (`$5300 USD p/a`). Is this an anomaly (**outlier**) in the dataset? We leave the question for you to further investigate!

Do you want to know more?

 'tail' man page  

 bash redirections (GNU manual) 