Data Science toolbox (CSVKit) exercises

All the exercises are done using the file optd_aircraft.csv as input

Also, it might be useful to follow this tutorial:
https://csvkit.readthedocs.org/en/0.9.1/tutorial.html

1. Use csvlook to evaluate what columns and what kind of data is included in the CSV file
2. Use csvstat to find out how many different manufactures are in the file. What are the top 5 manufacturers?
3. Extract the column manufacturer and using pipes, use sort and uniq –c to find the top 10 manufacturers
4. Extract the column manufacturer and using pipes, use sort, uniq and wc –l to find out how many manufacturers are in the file. Why does this number differ to the number reported in csvstat?
5. Using csvgrep, get only the records with manufacturer equal to *Airbus*
6. What is the aircraft with the maximum number of engines? And if we focus only on *Airbus* aircrafts, what are the aircrafts with the maximum number of engines?
7. Create a script that accepts a CSV filename as input ($1 inside your script) and returns the model of the aircraft with the highest number of engines.
8. Repeat script 7, but add a second argument to accept the manufacturer. The script will output the model with the highest number of engines, but limited to that manufacturer. If several planes have the highest number of engines, then the script will only show one of them.
9. Create a script that accepts as input arguments the name of the CSV file, and a number (number of engines) and returns the number of aircrafts that have that number of engines.
10. Using the file optd_por_public.csv, compress it with bzip2 and extracts all the lines starting with MAD (with the compressed file, using bzcat and grep)