📖 **project_01.md**

Project 01 Solution

Created: 08/19/2019

Question 3: Investigating some Data on Scholar

> NOTE: these problems can be solved by navigating the file browser or using commands. In some questions they yield differing results.

Question 3a

Some airline data is located at: `/class/datamine/data/flights` *Roughly how much data in stored in this directory? We do not need an exact count (yet). Just give a sense: several kilobytes? megabytes? gigabytes? petabytes?*

Use the `du` command with the `-b` option for approximate disk usage and `-h` option for human-readable output:

```
du -bh /class/datamine/data/flights
```

- **ANSWER (with file browser): 83.1G**
- **ANSWER (with commands): 78G**

Question 3b

Find the map files for the taxi data (these are jpg files). How many files are there, in that maps directory?

Use `ls` to list the files in `.../taxi/maps` and count the number of lines in the output:

```
ls /class/datamine/data/taxi/maps | wc -l
```

- **ANSWER: 5 files**

Question 3c

*Roughly how much data is stored (altogether) in the directory about **yellow** taxi cabs?*

Use the `du` command with the `-b` option for approximate disk usage and `-h` option for human-readable output:

```
du -bh /class/datamine/data/taxi/yellow
```

- **ANSWER (with file browser): 241.9G**
- **ANSWER (with commands): 226G**

Question 3d

For which years do we have election data?

Use `ls` to list the files in `.../data/election` :

```
ls /class/datamine/data/election
```

- **ANSWER: 1980 through 2020**

Question 3e

For which cities in California do we have AirBnB data?

Use `ls` to list the files in `.../data/airbnb/united-states/ca`:

```
ls /class/datamine/data/airbnb/united-states/ca
```

- **ANSWER: Los Angeles, Oakland, Pacific Grove, San Diego, San Francisco**