

Data Science Techniques (MAT 339) Homework 2 - - 10 Points - - SOLUTIONS

Submit a **hard copy** of your work at the beginning of class on Wednesday, January 28th. There is no **electronic submission** for this homework.

1. Download the free book *The Linux Command Line* from <https://linuxcommand.org/tlcl.php>. Although we are not using Linux, much of the content of this book applies to us in Git Bash.
 - (a) Read pages xvi to xviii of the **Introduction**.
 - (b) Spend at least another 10-15 minute skimming the book and jot down a few pieces of information you learned as your answer to this question.
2. On your computer find (or create) a folder with numerous files in it (and possibly folders). Use an `ls` command with some of its options (use `-l` and others) to list the files. Then use the approach shown in class to redirect the listing to a file (name it `file-listing.txt` or similar). For your answer to this question include:

- (a) your line of code with the redirect

SOLUTION: `ls -l > filelisting.txt`

- (b) a screen shot of the resulting text file open in Notepad or similar

SOLUTION: a snippet of the output file `file-listing.txt`:

```
-rw-r--r-- 1 IFrommer 1049089 23721 Feb 12 2024 2024 ORDA
catalog updates.docx
-rw-r--r-- 1 IFrommer 1049089 16902 Apr 9 2024 202425 HPE
Curriculum Update Advisor Guidance.docx
drwxr-xr-x 1 IFrommer 1049089 0 Dec 30 2024 Admin/
drwxr-xr-x 1 IFrommer 1049089 0 Jul 14 2023 Apps/
drwxr-xr-x 1 IFrommer 1049089 0 Jan 4 2023 Attachments/
-rw-r--r-- 1 IFrommer 1049089 14239 May 20 2025 Awards luncheon.
docx
-rw-r--r-- 1 IFrommer 1049089 10007 Mar 7 2024 Book1.xlsx
-rw-r--r-- 1 IFrommer 1049089 10281 May 13 2025 Book2.xlsx
-rw-r--r-- 1 IFrommer 1049089 92966 Feb 7 2025 CAP question
writing.docx
-rw-r--r-- 1 IFrommer 1049089 11633 Jan 15 2024 Calc 3111-101
Section.xlsx
```

3. Jake VanderPlas' *Python Data Science Handbook* is available from a GitHub repo at <https://github.com/jakevdp/PythonDataScienceHandbook>. Start by cloning this repo to your machine. (Note it is around 80 Mb, so if you are low on space, find a different smaller repo online to use.) Using shell commands covered in class (including `ls` with options), find all Jupyter files in this repo and list their names with details in decreasing order of file size.

SOLUTION:

From the root directory of the repo:

```
find . -name *ipynb -print0 | xargs -0 ls -lhS | less
```

-S to sort by decreasing size, -H for human-readable sizes. Here are the beginnings of the results:

```

-rw-r--r-- 1 IFrommer 1049089 2.7M Feb 5 12:52 ./notebooks_v1/04.13-
    Geographic-Data-With-Basemap.ipynb
-rw-r--r-- 1 IFrommer 1049089 2.1M Feb 5 12:52 ./notebooks_v1/05.11-
    K-Means.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.8M Feb 5 12:52 ./notebooks_v1/06.00-
    Figure-Code.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.7M Feb 5 12:52 ./notebooks/05.11-K-
    Means.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.7M Feb 5 12:52 ./notebooks/06.00-
    Figure-Code.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.5M Feb 5 12:52 ./notebooks_v1/05.10-
    Manifold-Learning.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.1M Feb 5 12:52 ./notebooks_v1/05.12-
    Gaussian-Mixtures.ipynb
-rw-r--r-- 1 IFrommer 1049089 1.1M Feb 5 12:52 ./notebooks_v1/04.14-
    Visualization-With-Seaborn.ipynb
-rw-r--r-- 1 IFrommer 1049089 1010K Feb 5 12:52 ./notebooks_v1/05.07-
    Support-Vector-Machines.ipynb
-rw-r--r-- 1 IFrommer 1049089 953K Feb 5 12:52 ./notebooks/05.10-
    Manifold-Learning.ipynb

```

4. Go a folder on your machine with regular activity of files being saved there, such as your **Downloads** or **Documents** folder.

- (a) Look up the `-mtime` option of the `find` command in the *The Linux Command Line* book you downloaded at the beginning of this homework and read about it. What page was it on?

SOLUTION: p237 of the book (p261 of the pdf)

- (b) Use the `-mtime` option to list files in the folder modified within some recent amount of time, enough to return some but not too many results. Do not submit this result.

SOLUTION: `find . -type f -mtime -1` (using `-1` to capture files modified within the past 1 day)

- (c) Now repeat the previous step but pipe it to the `tr` command changing the letters (and possibly numbers as well) to something that would cloak them, e.g. change all letters to B's. There are examples of the usage of `tr` in the slides from Day 2 and in the book. Include your command for this part and a screen shot (or pasted text) of the result in your homework submission.

SOLUTION: `find . -type f -mtime -1 | tr [a-z] 7`

```

./..777/777777
./..777/FETCH_HEAD
./..777/7777/7777/7777777/777777/HEAD
./..777/7777/7777777/777777/HEAD
./_7777777777/HW/7702-77777777/77-77777777-777777-77777.777
./_7777777777/HW/7702-77777777/77-77777777-777777-77777.777
./_7777777777/HW/7702-77777777/77-77777777-777777-77777.777
./_7777777777/HW/7702-77777777/77-77777777-777777-77777.777
./_7777777777/HW
    /7702-77777777/77-77777777-777777-77777.777777.77
./_7777777777/HW/7702-77777777/77-77777777-777777-77777.777
./_7777777777/HW/7702-77777777/777/1.777
./_7777777777/HW/7702-77777777/777/2.777

```