



Script Walkthrough





What each script does

DIFF (Two files)

- Can compare two Excel files
- Highlights differences: Green = new, Red = changed, Grey = removed

Format Check (One file)

- Setup: (only needs to be run once) Uses a sample file as a reference for formatting - Creates json file
- Running: Analyzes given file with json as reference for formatting

Number Check (One file)

- Setup: Uses 3x Standard Deviation from numeric column as default values (Covers ~99.7%)
- Running: Compares values in numeric column with range created by setup



**Be sure to have
Python or Anaconda**



Links to Python Distributions (Pick one)

Anaconda:

<https://www.anaconda.com/distribution/>

Miniconda:

<https://docs.conda.io/en/latest/miniconda.html>

Vanilla Python:

<https://www.python.org/downloads/release/python-374/>



Command Line stuff - Changing Directory (Folder)

Type `cd` followed by a path (Case insensitive) to change directories

Note: If the path contains white space, put it in quotes

E.g. `cd`

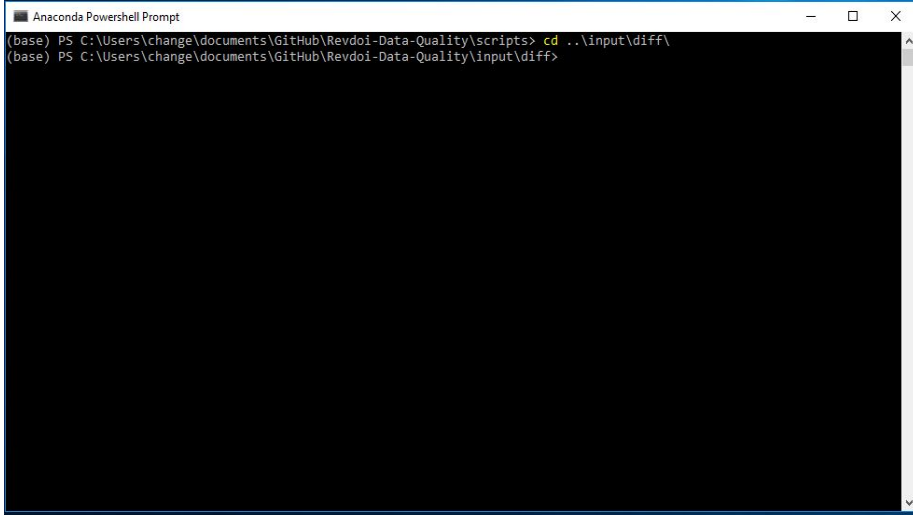
```
documents/github/revdoi-data-quality/scripts
```



Command Line stuff - Changing Directory (Folder)

To navigate backwards, type `cd ..`

- Can be followed by any path within that directory

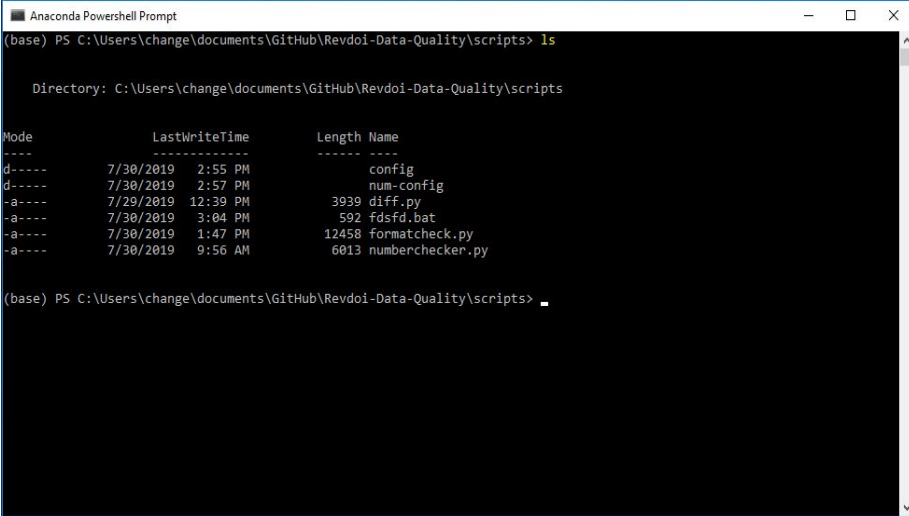


```
Anaconda PowerShell Prompt
(base) PS C:\Users\change\documents\GitHub\Revdoi-Data-Quality\scripts> cd ..\input\diff\
(base) PS C:\Users\change\documents\GitHub\Revdoi-Data-Quality\input\diff>
```



Command Line stuff - dir / ls

`dir` or `ls` will print out anything listed under the current directory



```

Anaconda PowerShell Prompt
(base) PS C:\Users\change\documents\GitHub\Revdoi-Data-Quality\scripts> ls

Directory: C:\Users\change\documents\GitHub\Revdoi-Data-Quality\scripts

Mode                LastWriteTime         Length Name
----                -
d-----          7/30/2019   2:55 PM              config
d-----          7/30/2019   2:57 PM             num-config
-a-----          7/29/2019  12:39 PM           3939 diff.py
-a-----          7/30/2019   3:04 PM            592 fdsfd.bat
-a-----          7/30/2019   1:47 PM          12458 formatcheck.py
-a-----          7/30/2019   9:56 AM            6013 numberchecker.py

(base) PS C:\Users\change\documents\GitHub\Revdoi-Data-Quality\scripts>

```



Command Line stuff - Running Python scripts

`python name.py`

- Might require additional input after the script name
 - Diff needs two excel files
 - Formatcheck and numcheck need one

Sample script available



Running the Scripts - diff.py

```
python diff.py [old_file.xlsx] [new_file.xlsx]
```

```
python diff.py ../input/diff/monthly_production_06-2019.xlsx  
../input/diff/monthly_production_07-2019.xlsx
```

Should export new Excel, "[old] vs [new].xlsx" in output



Running the Scripts - formatcheck.py

Run setup first (Will create json for monthly_production)

Python formatcheck.py setup [sampleFile.xlsx]

```
python formatcheck.py setup ../input/setup/monthly_production_05-2019.xlsx
```

python formatcheck.py [fileToCheck.xlsx]

```
python formatcheck.py ../input/formatTest/monthly_production_test.xlsx
```

Should Export new Excel, “[new] filename.xlsx” in output



Running the Scripts - numberchecker.py

Can work without calling setup

Python numberchecker.py setup [sampleFile.xlsx]

```
python numberchecker.py setup ../input/numTest/monthly_revenue_05-2019.xlsx
```

python numberchecker.py [fileToCheck.xlsx]

```
python numberchecker.py ../input/numTest/monthly_revenue_05-2019.xlsx
```

Should Export new Excel, “NumChecked-fileName.xlsx” in output



Sample Locations

Diff samples

`../input/diff`

Format Sample

`../input/setup`

`../input/formatTest`

Number Sample

`../input/numTest`



In case I forgot something

Link to Data Script Wiki:

<https://github.com/ONRR/Revdoi-Data-Quality/wiki>

Glossary:

<https://github.com/ONRR/Revdoi-Data-Quality/wiki/Glossary>