[Update: SQL of ~1GB converted to csv with: (2 websites – file too large; pip install sqlite3 failed; solution: 1) Manually download the sqlite3 **tools** package <https://stackoverflow.com/questions/43050795/no-sqlite3-exe-in-sqlite3-download-folder-for-windows-64-bit> ; copy sql file to same folder; then code directly in sqlite3 window, from <https://stackoverflow.com/questions/33467031/export-very-large-sql-file-into-csv-with-python-or-r> . In sqlite3 shell: ‘.open “openparliament2public.sql”’ [whatever filename]; ‘.output canadianparl.csv’; ‘.mode csv’; ‘select \* from openparliament2public’; ‘.quit’. Solution edited for typos and filename – CM Aug. 19, 2022. Update – open file without quot marks around filename; view headers with ‘.show’ (in Command shell but while in sqlite mode there); table content spears to be ‘stdout’; trying command shell one-liner (<https://stackoverflow.com/questions/5776660/export-from-sqlite-to-csv-using-shell-script/21741408#21741408>) w that name:   
[In Command Shell]  
>cd C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts

>sqlite3  
>.open openparliament2public.sql #file

>.show #show headers, table name? (“output”?)

> TRY: sqlite3 -header -csv “openparliament2public.sql” “select \* from stdout” > “out.csv”

[or without quots]. [ W quots, did not work ] [no luck w command shell].

Or > TRY: sqlite3 -header -csv openparliament2public.sql “select \* from stdout;” > cdnout.csv

\*\* Update from sqlite3 .help: try . instead of – in commands above. !! (.exit, .help, .import FILE TABLE, .open FILE, .mode MODE ?TABLE?, .show, )

-NVM – Code outdated – trying sqlite3 in Python…

-Or – w <https://www.philiplb.de/sqldumpsplitter3/> lol ;(

-Or: Downloading the mysqldump\_to\_csv.py script (save in Notepad as .py) from <https://github.com/yashsmehta/mysqldump-to-csv> ; open Cmd shell > cd set > python mysqldump\_to\_csv.py;

-Update – Modify the mysqldump\_to\_csv.py script to has out/remove the SIGPIPE lines (forming “mysqldump\_to\_csv\_mod.py”; need to set decimal/precision level lower (problem on Windows only) cf here (<https://stackoverflow.com/questions/38314118/overflowerror-python-int-too-large-to-convert-to-c-long-on-windows-but-not-ma>). [In Cmd shell:]

>cd C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts

[[ DON’T then load python as its own line ]]. – Hashed out the sys.maxsize line to test.

>python mysqldump\_to\_csv\_mod.py openparliament2public.sql

-Update: Run either in cmd shell, cmd shell with Postgresl dl’d/loaded/called, or some Postgresl shell:

<https://www.postgresql.org/docs/current/app-pgdump.html>

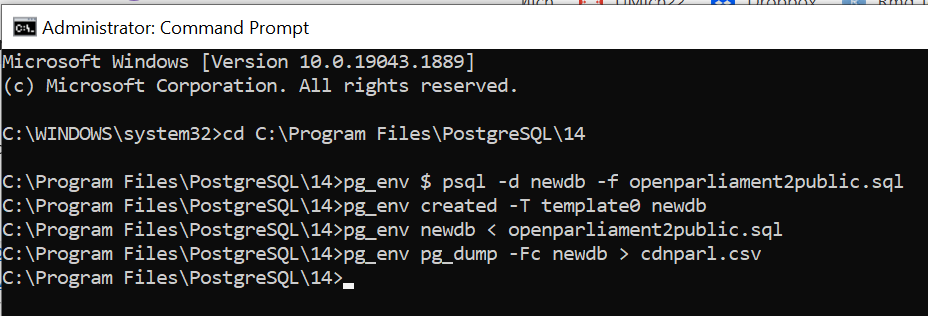
$ psql -d newdb -f db.sql

1. Dl from <https://www.postgresql.org/download/>
2. Run the downloaded .exe. (Couldn’t save to Python or Scripts folder bc non-empty; password for super-user set to pxpx…3; server it’s using is Port 5432). Default locale.
3. Site says only “command-line” re the functions shown – there also doesn’t seem to be an app/shell in the c:/ program files / PostgreSQL folder. Trying in regular shell after setting cd to C:\Program Files\PostgreSQL\14 . \*\*Note: Running ‘pg\_env’ after this will get the batch file.
4. ‘pg\_env –help’ also works or at least doesn’t return error.
5. Line as given is ‘$ psql -d newdb -f db.sql’ (w sql file in same folder and real file name. sql)
6. Trying ‘pg\_env $ psql -d newdb -f openparliament2public.sql’
7. No errors…
8. Return to orig python file to get the ‘newdb’ database? No errors but need table name
9. Table names <https://www.tutorialspoint.com/how-to-show-all-the-tables-present-in-the-database-and-server-in-mysql-using-python> : Download MySQL, etc. (checked – costs money and/or requires user account to dl free trial).
10. (Alt) Get ‘psql’ which has a terminal/front end program. <https://www.postgresql.org/docs/current/app-psql.html> . <https://www.timescale.com/blog/how-to-install-psql-on-mac-ubuntu-debian-windows/> . <https://www.postgresql.org/download/windows/> (looks like regular dl though). Looks to be already at: C:\ProgramData\Microsoft\Windows\Start Menu\Programs\PostgreSQL 14 . Copied to c / program files / postgresql / 14 folder to match… Opened here and typed in psql shell:
11. >’-d newdb’ [database file already created w postgresql]
12. >’\dt’ to ideally get tables list…

Restarting in command shell: <https://stackoverflow.com/questions/6842393/import-sql-dump-into-postgresql-database>

1. setting cd to C:\Program Files\PostgreSQL\14
2. run: pg\_env $ psql -d newdb -f openparliament2public.sql
3. (This time, copied an empty ‘newdb’ file to same folder first.)
4. Run: psql --username=postgres newdb < openparliament2public.sql
5. Or: pg\_env --username=postgres newdb < openparliament2public.sql ?

Again <https://www.postgresql.org/docs/14/backup-dump.html>

1. Right-click open Command Shell to Run as Administrator. (for filewriting access).
2. In Command Shell. Run: cd C:\Program Files\PostgreSQL\14
3. Run: pg\_env $ psql -d newdb -f openparliament2public.sql
4. Create new data base with (cm edits) (<https://www.postgresql.org/docs/14/backup-dump.html>).   
   Run: pg\_env created -T template0 newdb
5. (Same link code, edited re psql vs pg-env by cm)   
   Run: pg\_env newdb < openparliament2public.sql
6. (try to save directly from shell)   
   Run: pg\_env pg\_dump -Fc newdb > cdnparl.csv
7. Results: Runs, creates file, but file is empty.   
   (Image below of steps) 
8. Or pg\_env pg\_dump -Fc newdb > C:\Program Files\PostgreSQL\14\cdnparl2.csv (No file created)

Or pg\_env pg\_dump -Fc --user=root --host=localhost --password=”pxpx3023” newdb’.’ > cdnparl3.csv (File created but empty)  
Or pg\_env $command = 'C:\Program Files\PostgreSQL\14' --opt --user=root --host=localhost --password="pxpx3023" newdb'.' > cdnparl4.csv; exec($command); (File created but empty)

1. Not working – online help is abt a specific mysql command. – tried pg\_env newdb, (next line) pg\_env \d or pg\_env \dt but no table names printed.

MySQL for tablenames… advice seemed more up to date, but paid/user account subscription req…

<https://edelivery.oracle.com/osdc/faces/Home.jspx>

New script //

<https://github.com/darrentu/convert-db-to-csv>

1. Command shell set cd back to Scripts folder,
2. **create db with sqlite3 mydata.db < openparliament2public.sql**\*\*Note – some errors even in this; unrecognized chars/tokens… but did seem to fully run (errors returned up to line 3,806,976). (~2h?).
3. run the script… and/or try to get table names; save to csv now that mydata.db exists? For the Darren Tu script: Run (assume in Cmd shell):   
   $ git clone <https://github.com/darrentu/convert-db-to-csv>  
   Update: Apparently in Bash Shell.
4. Any attempts to save ‘mydata’ or ‘mydata.db’ to DB file in command shell/sqlite3: (<https://stackoverflow.com/questions/28024810/how-we-can-save-database-created-in-sqlite3>) – some empty files print with .save, .backup, ‘.save mydata C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts\mydata.db’   
   = ‘Error: unknown database mydata’ or syntax error. – Long line quoted above most successful… (.save mydata C:\...).
5. .save mydata.db C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts\mydata.db  
   = ‘Error: unknown database mydata.db’.
6. ‘.show’ (in sqlite) says last filename is ‘.backup’…
7. Tables from saved .db: looks promising <https://www.sqlitetutorial.net/sqlite-show-tables/>
8. Export file to csv from database (w/ its table names): <https://www.sqlitetutorial.net/sqlite-export-csv/> and other similar sources…
9. Or.. (sqlite) -csv C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts\mydata.db < openparliament2public.sql > cdnparl111.csv (no … Full pathway filenames above also worse).
10. Trying SQL Lite Studio?? (<https://sqlitestudio.pl/>); (<https://www.sqlitetutorial.net/sqlite-export-csv/>). Can’t add database as it is not in format: .db, .sdb, .sqlite, .sqlite3, etc. (no .sql and would not open either .sql or .bzz).
11. Wait.. <https://database.guide/export-entire-sqlite-database-to-sql-file/> good explanation of making the db.. Looks good but tried all steps in cmd shell sqlite mode AND in sqlite app and they did not run. (“syntax error”). (sqlite3 Store2.db, next line: .read openparliament2public.sql).

Darren Tu script --

1. DL the Bash Shell (<https://www.gnu.org/software/bash/>). Downloaded most recent ‘HTTPS’ version; copied to Scripts folder. (Already had cd set to this in command shell).
2. Ran ‘bash’ in Windows command shell and it seems like it started.
3. Run: $ git clone <https://github.com/darrentu/convert-db-to-csv>  
   Ran both this without the $ (as already visible in pre-loaded line info) and with it; both returned error: “bash: git:: command not found”.
4. Solution pages: must download GIT (<https://superuser.com/questions/568964/git-command-not-found-windows>) and add to path (<https://stackoverflow.com/questions/66129948/windows-10-git-bash-bash-git-command-not-found>).   
   Git dl (<https://git-scm.com/download/win>) (.exe file, selected all defaults to install)
5. Assume I have to edit the ‘environment variables’ on my computer and add ‘C:\Program Files\Git’ to PATH. Tried in ‘git bash’ shell:   
   $ export PATH=$PATH:/c/Program\ Files/Git/bin:/c/Program\ Files/Git/cmd
6. Then re-open git-bash shell and try ‘git’, ‘ls’, ‘env’ (w/o quots) (seemed to work) (prev 2 steps from <https://stackoverflow.com/questions/66129948/windows-10-git-bash-bash-git-command-not-found>.
7. Back to main Windows Command Shell: Run [$ already showing]:  
   git clone <https://github.com/darrentu/convert-db-to-csv>
8. Re-opened Windows Command Shell and re-set dir (cd) to C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts
9. Re-type ‘bash’ in main cmd shell
10. Try git line again (git clone <https://github.com/darrentu/convert-db-to-csv>)
11. Still ‘git command not found’, ‘$ command not found’ if including.
12. RUN IN ‘GIT BASH’ APP: git clone <https://github.com/darrentu/convert-db-to-csv> (Worked)
13. Run: cd convert-db-to-csv
14. Run: ‘./convert-db-to-csv.sh example.db’ [if db file in same folder] or ‘./convert-db-to-csv.sh ~/Library/Safari/History.db’ [if file in different folder]. [cm: typing ‘cd’ alone reveals the folder is: ‘katie@DESKTOP-SN009NK MINGW64 ~’ [??]. – Also need a .db file, which I don’t have. NVM.

(Aug 19, 12:30pm – 9:47 pm minus maybe 30min lunch (3:15-3:45), 1h walk). ☹

-Remaining issues: -Just have a .sql file and can’t save it as .db

-Due to not knowing the ‘table names’ – official SQLite documentation says .tables ; some special schema\_table will exist – but calls will not work to get these. (if loading in the open parl .sql file to mydata.db, for example). – Could try again w/o changing to .backup when trying to save as I think this over-wrote the current database being mydata. – Mydata.db loading returned many errors but was most successful at getting to 3.8m lines +. – Also not sure if errors break it.

-IE need the .sql file to be a .db file, then either SQL studio program downloaded or the Darren Tu script are supposed to be able to navigate, find tables, save tables to csv.

Or via R:

<https://stackoverflow.com/questions/52138048/how-to-load-large-datasets-to-r-from-bigquery?noredirect=1&lq=1>

-Monday: Try (1) Via R link above; (2) the semi-successful mydata.db creation in command line – try not to overwrite, try saving w admin permissions, etc. ; (3) mysql (proprietary software but possibly free trial). If not – (4) email site owner (5) check w cc.

-Or

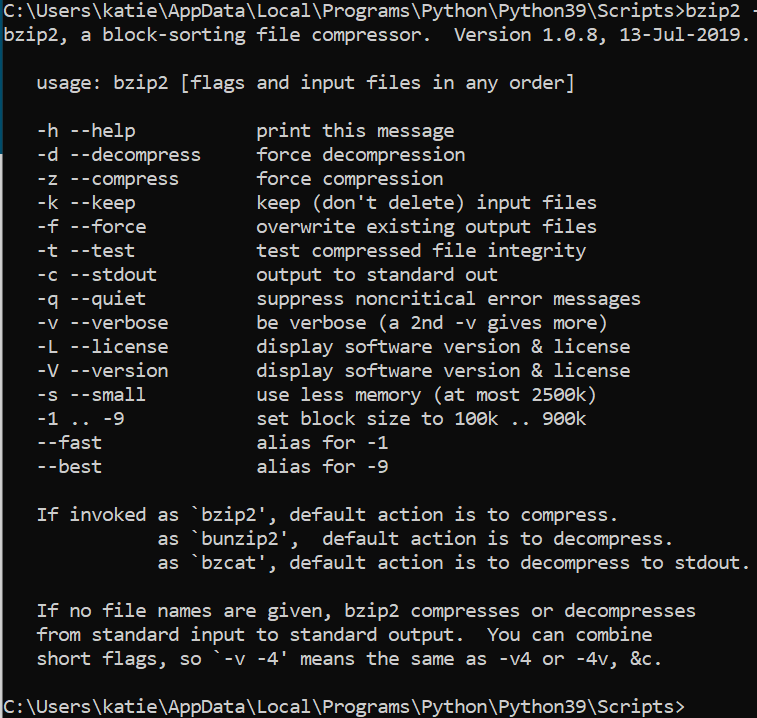
1. redownload .bz2 file. (remove .sql ending to name)

2. open 7zip (found online: WinZip) and “open” .bz2 file from within the 7-zip program

3. ‘extract’ in 7zip which creates a new 5gb file (no file type/extension) in the folder. [[This was because I changed the original name from filename.sql.bz2 to just filename.bz2]].

4. search ‘split large bz2 file’. Via command line w bzcat download and ‘split’ function: <https://madflex.de/splitting-a-big-file-with-split/> [Apple/Unix only?]

5. get bzip2 for windows version, <https://superuser.com/questions/864673/uncompress-bz2-file-from-command-line-on-windows> . (DL Bzip2 manually, copy app to regular cd folder). Setting cd then typing ‘bzip2’ gets a response, but it says “I won’t write compressed data to a terminal.” -checking help w ‘bzip2 –help’.  
Help message:



6. trying in cmd shell… (copy paste .bz2 1GB file to same folder 1st). ‘bunzip2 openparliamentpublic2022b.bz2 -c --stdout -9 bzoutput’ ? [bunzip2 not recognized]

**bzip2 --decompress openparliamentpublic2022b.bz2 -c --stdout -9 bzoutput**  
[ran and printed all contents to screen …] [ some saved via copying ].

Note: It printed “Name: hansards\_statement\_mentioned\_politicians\_id\_seq; Type: SEQUENCE SET; Schema: public; Owner: op

SELECT pg\_catalog.setval('public.hansards\_statement\_mentioned\_politicians\_id\_seq', 297474, true);

-- PostgreSQL database dump complete” at end.  
  
  
7. try (<https://www.rootusers.com/10-simple-bzip2-examples/>)   
  
bzip2 -d -c openparliamentpublic2022b.bz2 > file.txt

[-d is same as --decompress; -c means keep original file]

[Did not work – made empty file.txt.

6. [^ if this doesn’t work try re-downloading as filename.sql.bz2 then unzipping, etc. ]  
  
bzip2 -d -c openparliamentpublic2022.sql   
[THIS PRINTS TO SCREEN AGAIN]

bzip2 -d -c openparliamentpublic2022.sql > cdnfile.sql

[creates empty file :/]

[wait it was running]

[to decompress to multi files try : ]

bzip2 -d -c openparliamentpublic2022.sql > cdnfiles.bz2

[creates actually empty file]

[it did load to full size after closing cmd shell, 1.5 gb…]

[to get non sql output… ]

**bzip2 -d -c openparliamentpublic2022.sql > cdnfiles.txt**

[tbd try]

[works on closing, ~10 min = 4.5 gb of txt ]

- 6. Get vcruntime.dll to enable the program (video on how to get: <https://www.youtube.com/watch?v=6PTf1ywq6g0> ).

6. Winzip apparently splits large bz2 into smaller files. Paid software though.

7. Trying splitting file with 7zip. (<https://www.webhostinghub.com/help/learn/website/managing-files/split-file>).

(could also try to re-dl as filename.sql.bz2 and then unzip w 7zip as this may produce a usable small file/several??)

-trying aug 22 / 7:45pm … open/extract 1st, then right click the one extracted file (in a folder) > split

-! Splitting the EXTRACTED SQL file (openparl folder), right click, 7Zip – Add to Archive

-brings up ‘Split to volumes, bytes’ option. !! (Choosing save to .zip, 100M bytes size).

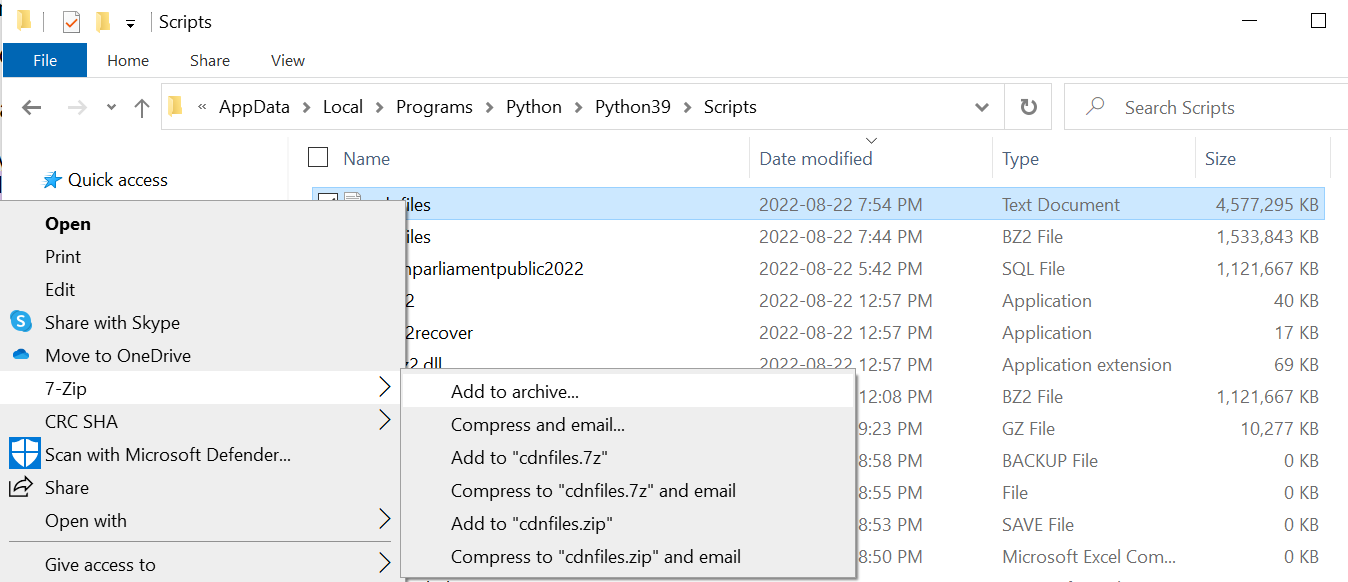
-Then just had to delete “.001” off end of file name, which changes it to .zip

-but then just has the 7kb poorly encoded one line…

-Or: **Splitting the large .txt file directly** (**right click.. 7 zip.. 100m file size ) – did split but saved as .025 filetype, etc. (Split2 folder), and 100k kb each… [UPDATE THIS WORKED- SET TO 650 MB AND MANUALLY EDIT FILE NAMES WHEN DONE].**

-re-checking with a 1000M size, then manually renaming… (2nd time just showed right-click – 7Zip – Add to Archive)… Selected to create .zip, 1000M file size. X Cancelled and changed the setting to 100M on re-run. Saved to ‘split txt’ folder. (see eg also here <https://tuf.hds.com/big_file_instructions.php#:~:text=Once%20you%20have%20installed%207zip,%2D%3E%20Add%20To%20Archive%22.&text=This%20will%20compress%20and%20create,data%20in%20the%20large%20file>. )

-i.e. (First download the filename.sql.bz2 file. Use command shell and program (download) bzip2 to convert one file (load as filename.sql) to output filename.txt format. Then use 7zip on the large .txt file (right-click it in folder, 7zip -> Add to Archive -> save as .zip format and split into 100MB chunks. Simply delete the ‘.001’ off .zip output filename to get the actual .zip folder. Can then open, etc. <3



-update – 13 zipped folders after, but can’t open/extract/open with 7zip after as ‘not an archive’

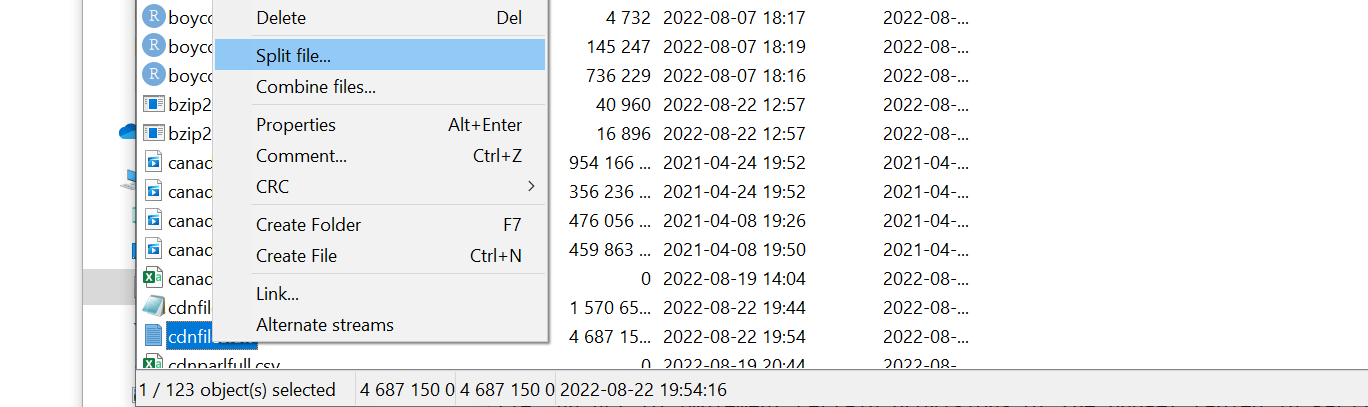
-right click file – 7zip ‘Extract files’ worked better in making a new unzipped folder. Actually fine 😊

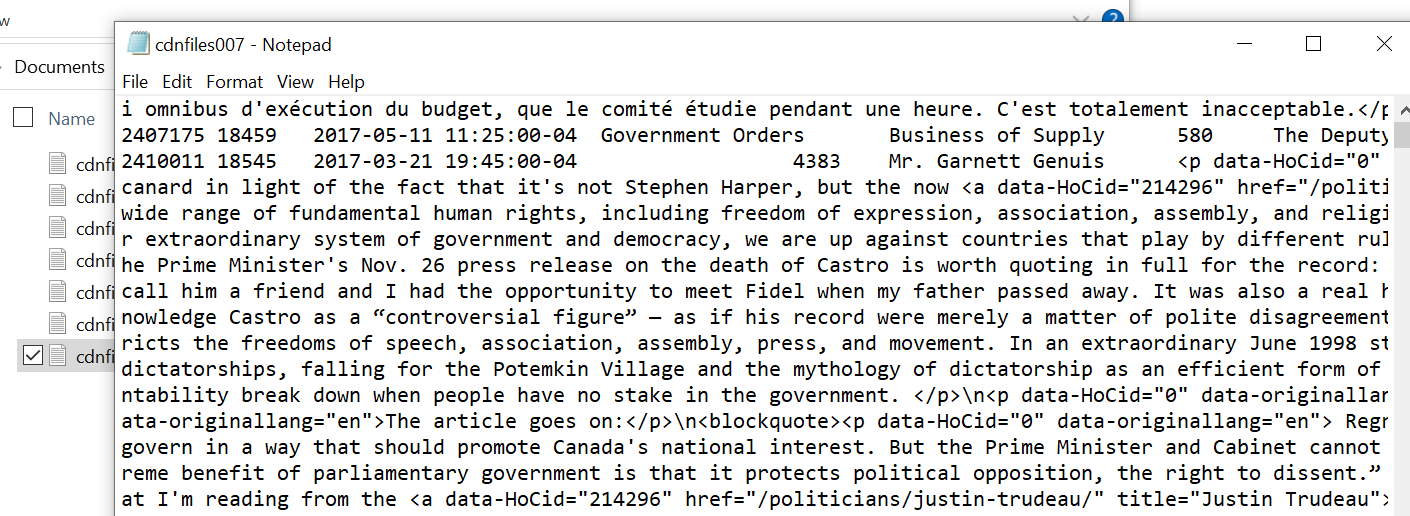
-manually rename all txt files. Abt 400K each.

-Extract files not working past 1st folder… or Open archive (error) then Extract files… (maybe try diff file type; or could be problem w headers <https://7ziphelp.com/7zip-cannot-open-file-as-archive> ).

-retrying from large .txt file to 7zip archive in diff format.

-Actually just **right-clicking large .txt file in 7zip app, Split**, selected 650 MB file size. Saved to Split3 folder. Manually take off invalid endings to the names. WORKED !!





-Update: Need to rerun though as it only went to 2019. Give cmd shell 20min+, probably like 50min.

FINAL INSTRUCTIONS:

1. Manual DL of data (openparliamentpublic2022.sql.bz2) from OpenParliament.ca/datadownload. Save to C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts (usual cmd shell cd). Have 7zip and ‘bzip2’ programs manually downloaded (for Windows).
2. Open command shell, run cd C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts
3. Run in cmd shell: bzip2 -d -c openparliamentpublic2022.sql > cdnfiles2.txt  
   (Can’t see progress but leave for 25+min, can see upon closing command shell.)  
   (Started 8:52pm, August 22, 2022).
4. In 7zip program, navigate to folder, right click the large .txt file (cdnfiles2.txt), ‘Split’. Set to 650MB and choose end folder. (Now in txtsplit2).
5. Rename individual output files (take number off end). 😊

ALT APPROACH

-After reading files to openable size .txt, header info on the first doc shows (probably) main table in the SQL is: “public.hansards\_statement”.

-Re-extracting one large .SQL file from the sql.bz2 file w 7zip. (File looks like .sql in folder). “Extract” from within the 7zip program. (Then could split to multi small .sqls as one option … ) (or one large .sql to dataframe in python then split).

-Going to new 5.7 GB .sql in 7zip and right-click “Split”.. ? Trying 650MB size per file.

-Remember to manually fix file name endings.

-(1 option) try auto .sql to .csv websites… (<https://tableconvert.com/sql-to-csv> many pos reviews); -- Failed on 1st one and another = ‘Out of memory’ even after reloading site.

-1.5 Try manually renaming theextracted 5.7gb”file” (no extension) to .sql? Re-split and try? –Also remember to rename the split files. Re-upload on website (Error: ‘something went wrong while parsing’ on first 2).

-(2nd option) Try to convert one large unzipped .sql (5.7gb) into one csv in command shell, then split.:

-Cmd shell: set cd to C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts

-Cmd shell: sqlite3 -csv openparliamentpublic2022unzip.sql “SELECT \* FROM public.hansards\_statement;” > cdnout.csv

-error: ‘file is encrypted or is not a database’. – trying running shell as admin…

sqlite3 -csv openparliamentpublic2022unzip.sql “SELECT \* FROM public.hansards\_statement;” > cdnout.csv (same error)

sqlite3 -csv openparliamentpublic2022unzip.sql “SELECT \* FROM public.hansards\_statement” > cdnout.csv

sqlite3 -csv openparliamentpublic2022.sql “SELECT \* FROM public.hansards\_statement” > cdnout.csv

(Same error on all – try .open on file first… )

sqlite3 .open openparliamentpublic2022unzip.sql

(‘syntax error’)

sqlite3 mydata.db > .open openparliamentpublic2022unzip.sql

(also tried w/o .sql ending, and w filename in quots (w sql ending) -- no luck).

[[ nvm, use text? Or r, python… ]] [[ ^ Retrying w/ table name “hansards\_statement”? ]]

<https://stackoverflow.com/questions/5776660/export-from-sqlite-to-csv-using-shell-script/21741408#21741408>) w that name:   
[In Command Shell]  
>cd C:\Users\katie\AppData\Local\Programs\Python\Python39\Scripts

>sqlite3  
>.open openparliament2public.sql #file

>.show #show headers, table name? (“output”?)

> TRY: sqlite3 -header -csv “openparliament2public.sql” “select \* from stdout” > “out.csv”

[or without quots]. [ W quots, did not work ] [no luck w command shell].

Or > TRY: sqlite3 -header -csv openparliament2public.sql “select \* from stdout;” > cdnout.csv

\*\* Update from sqlite3 .help: try . instead of – in commands above. !! (.exit, .help, .import FILE TABLE, .open FILE, .mode MODE ?TABLE?, .show, )