Lab Assignment 7: Database Queries

DS 6001: Practice and Application of Data Science

H. Diana McSpadden (hdm5s)

Instructions

Please answer the following questions as completely as possible using text, code, and the results of code as needed. Format your answers in a Jupyter notebook. To receive full credit, make sure you address every part of the problem, and make sure your document is formatted in a clean and professional way.

Problem 0

Import the following libraries, load the `.env` file where you store your passwords (see the notebook for module 4 for details), and turn off the error tracebacks to make errors easier to read:

In [ ]:
```
import numpy as np
import pandas as pd
import sys
import os
import requests
import psycopg2
import pymongo
import json
from bson.json_util import dumps, loads
from sqlalchemy import create_engine
import dotenv

# change to the directory where your .env file is
#os.chdir("/Users/jk8sd/Box Sync/Practice and Applications 1 online/Module 7 - Data

dotenv.load_dotenv() # register the .env file where passwords are stored
sys.tracebacklimit = 0 # turn off the error tracebacks
```

Problem 1

For this problem, we will be building a PostgreSQL database that contains the collected works of Shakespeare.

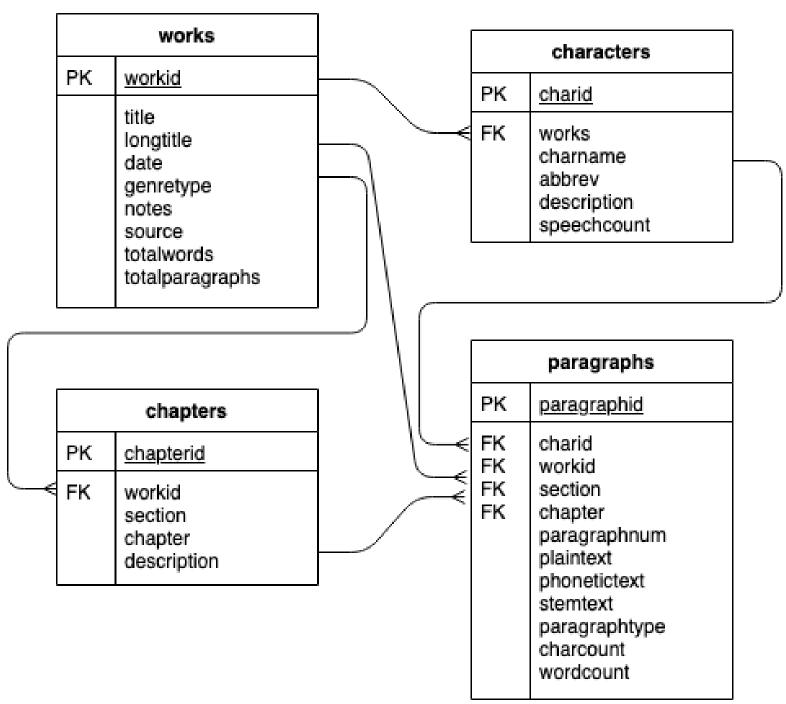
The data were collected by Catherine Devlin from the repository at https://opensourceshakespeare.org/. The database will have four tables, one representing works by Shakespeare, one for characters that appear in Shakespeare's plays, one for chapters (this is, scenes within acts), and one for paragraphs (that is, lines of dialogue). The data to populate these four tables are here:

In [ ]:
```
works = pd.read_csv("https://github.com/jkropko/DS-6001/raw/master/localdata/Works.
characters = pd.read_csv("https://github.com/jkropko/DS-6001/raw/master/localdata/C
chapters = pd.read_csv("https://github.com/jkropko/DS-6001/raw/master/localdata/Cha
paragraphs = pd.read_csv("https://github.com/jkropko/DS-6001/raw/master/localdata/P
```

In PostgreSQL, it is best practice to convert all column names to lower-case, as case sensitive column names will require extraneous double-quotes in any query. We first convert the column names in all four dataframe to lowercase:

In [ ]:
```
works.columns = works.columns.str.lower()
characters.columns = characters.columns.str.lower()
chapters.columns = chapters.columns.str.lower()
paragraphs.columns = paragraphs.columns.str.lower()
```

You will build a database and populate it with these data. The ER diagram for the database is:



There's no codebook, unfortunately, but the values in the columns are mostly self-explanatory:

Confirming the column names matching the IE diagram

In [ ]:
```
works.columns
```

Out[ ]:
```
Index(['workid', 'title', 'longtitle', 'date', 'genretype', 'notes', 'source',
       'totalwords', 'totalparagraphs'],
      dtype='object')
```

In [ ]:
```
characters.columns
```

Out[ ]:
```
Index(['charid', 'charname', 'abbrev', 'works', 'description', 'speechcount'], dty
pe='object')
```

In [ ]:
```
chapters.columns
```

Out[ ]:
```
Index(['workid', 'chapterid', 'section', 'chapter', 'description'], dtype='objec
t')
```

In [ ]:
```
paragraphs.columns
```

Out[ ]:
```
Index(['workid', 'paragraphid', 'paragraphnum', 'charid', 'plaintext',
       'phonetictext', 'stemtext', 'paragraphtype', 'section', 'chapter',
       'charcount', 'wordcount'],
      dtype='object')
```

In [ ]:
```
works.head()
```

Out[ ]:

	workid	title	longtitle	date	genretype	notes	source	totalwords	totalparagra
0	12night	Twelfth Night	Twelfth Night, Or What You Will	1599	c	NaN	Moby	19837	1
1	allswell	All's Well That Ends Well	All's Well That Ends Well	1602	c	NaN	Moby	22997	1
2	antonycleo	Antony and Cleopatra	Antony and Cleopatra	1606	t	NaN	Moby	24905	1
3	asyoulikeit	As You Like It	As You Like It	1599	c	NaN	Gutenberg	21690	
4	comedyerrors	Comedy of Errors	The Comedy of Errors	1589	c	NaN	Moby	14692	

In [ ]:
```
characters.head()
```

Out[ ]:

	charid	charname	abbrev	works	description	speechcount
0	1apparition-mac	First Apparition	First Apparition	macbeth	NaN	1.0
1	1citizen	First Citizen	First Citizen	romeojuliet	NaN	3.0
2	1conspirator	First Conspirator	First Conspirator	coriolanus	NaN	3.0
3	1gentleman-oth	First Gentleman	First Gentleman	othello	NaN	1.0
4	1goth	First Goth	First Goth	titus	NaN	4.0

In [ ]:
```
chapters.head()
```

Out[ ]:

	workid	chapterid	section	chapter	description
0	12night	18704.0	1.0	1.0	DUKE ORSINO's palace.
1	12night	18705.0	1.0	2.0	The sea-coast.
2	12night	18706.0	1.0	3.0	OLIVIA'S house.
3	12night	18707.0	1.0	4.0	DUKE ORSINO's palace.
4	12night	18708.0	1.0	5.0	OLIVIA'S house.

In [ ]:
```
paragraphs.head()
```

Out[ ]:

	workid	paragraphid	paragraphnum	charid	plaintext	phonetictext	stemtext	paragrapht
0	12night	630863	3	xxx	[Enter DUKE ORSINO, CURIO, and other Lords; Mu...	ENTR TK ORSN KR ANT O0R LRTS MSXNS ATNTNK	enter duke orsino curio and other lord musicia...	
1	12night	630864	4	ORSINO	If music be the food of love, play on;\n[p]Giv...	IF MSK B 0 FT OF LF PL ON JF M EKSSS OF IT 0T ...	if music be the food of love plai on give me e...	
2	12night	630865	19	CURIO	Will you go hunt, my lord?\n	WL Y K HNT M LRT	will you go hunt my lord	
3	12night	630866	20	ORSINO	What, Curio?\n	HT KR	what curio	
4	12night	630867	21	CURIO	The hart.\n	0 HRT	the hart	

Part a

Connect to your local PostgreSQL server (take steps to hide your password!), create a new database for the Shakespeare data, use `create_engine()` from `sqlalchemy` to connect to the database, and create the works, characters, chapters, and paragraphs tables populated with the data from the four dataframes shown above. [2 points]

In [ ]:
```
dotenv.load_dotenv("mod.env")
dbpassword = os.getenv("dbpassword")
```

In [ ]:
```
dbserver = psycopg2.connect(
    user='postgres',
    password=dbpassword,
    host="localhost"
)
# autocommit == True so we can create databases
dbserver.autocommit = True
cursor = dbserver.cursor()
# create the Shakespeare database
try:
    cursor.execute("CREATE DATABASE shakes;")
except:
    cursor.execute("DROP DATABASE shakes;")
    cursor.execute("CREATE DATABASE shakes;")

cursor = dbserver.cursor()
```

In [ ]:
```
# connect to the new database (same connection variable)
dbserver = psycopg2.connect(
        user='postgres',
        password=dbpassword,
        host="localhost",
        database="shakes"
)

engine = create_engine("postgresql+psycopg2://{user}:{pw}@localhost/{db}"
        .format(user="postgres", pw=dbpassword, db="shakes"))
```

In [ ]:
```
works.to_sql('works', con = engine, index=False, chunksize=1000, if_exists = 'repla
characters.to_sql('characters', con = engine, index=False, chunksize=1000, if_exist
chapters.to_sql('chapters', con = engine, index=False, chunksize=1000, if_exists =
paragraphs.to_sql('paragraphs', con = engine, index=False, chunksize=1000, if_exist
```

Out[ ]: 35475

Part b

Write a query to display `title`, `date`, and `totalwords` from the `works` table. Rename `date` to `year`, and sort the output by `totalwords` in descending order. Also create a new column called `era` which is equal to "early" for works created before 1600, "middle" for works created between 1600 and 1607, and "late" for works created after 1607. Finally, display only the 7th through 11th rows of the output data. [1 point]

First, do the query of title, date and totalword from the works table:

In [ ]:
```
cursor = dbserver.cursor()
t_sql = """
SELECT
    title,
    date AS year,
    totalwords,
    CASE WHEN date < 1600 THEN 'early'
        WHEN date > 1607 THEN 'late'
        ELSE 'middle' END AS era
FROM works
ORDER BY totalwords DESC
LIMIT 5 OFFSET 6;
"""


cursor.execute(t_sql)
rs_works = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_works = pd.DataFrame(rs_works, columns=colnames)
df_works
```

Out[ ]:

	title	year	totalwords	era
0	King Lear	1605	26119	middle
1	Troilus and Cressida	1601	26089	middle
2	Henry IV, Part II	1597	25692	early
3	Henry VI, Part II	1590	25411	early
4	The Winter's Tale	1610	24914	late

Part c

The `genretype` column in the "works" table designates five types of Shakespearean work:

- `t` is a tragedy, such as *Romeo and Juliet* and *Hamlet*
- `c` is a comedy, such as *A Midsummer Night's Dream* and *As You Like It*
- `h` is a history, such as *Henry V* and *Richard III*
- `s` refers to Shakespeare's sonnets
- `p` is a narrative (non-sonnet) poem, such as *Venus and Adonis* and *Passionate Pilgrim*

Write a query that generates a table that reports the average number of words in Shakepeare's works by genre type. Display the genre type and the average wordcount within genre, use appropriate aliases, and sort by the average in descending order. [1 point]

Answer:

Works table has the **totalwords** column, which I can use to get the **average wordcount** by genretype.

```
In [ ]: genre_map = {'t':'tradegy',
                     'c':'comedy',
                     'h':'history',
                     's':'sonnet',
                     'p': 'narrative poem'}
```

```
In [ ]: t_sql = """
SELECT
    CASE WHEN genretype = 't' THEN 'tragedy'
        WHEN genretype = 'c' THEN 'comedy'
        WHEN genretype = 'h' THEN 'history'
        WHEN genretype = 's' THEN 'sonnet'
        WHEN genretype = 'p' THEN 'narrative poem' END AS genretype,
    AVG(totalwords)::NUMERIC(10,0) AS average_words
FROM works
GROUP BY genretype
ORDER BY average_words DESC;
"""

cursor.execute(t_sql)
rs_works = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_genre_works = pd.DataFrame(rs_works, columns=colnames)
df_genre_works
```

Out[ ]:

	genretype	average_words
0	history	24236
1	tragedy	23817
2	comedy	20212
3	sonnet	17515
4	narrative poem	6182

Part d

Use a query to generate a table that contains the text of Hamlet's (the character, not just the play) longest speech, and use the `print()` function to display this text. [1 point]

Answer:

```
In [ ]: t_sql = """
SELECT
    sub.charname,
    sub.plaintext,
    sub.wordcount
FROM (
    SELECT
```

```
        p.plaintext,
        p.wordcount,
        c.charname
    FROM paragraphs p
    JOIN characters c ON c.charid = p.charid
    WHERE
        c.charname LIKE 'Hamlet'
        AND p.wordcount > 1.
    ORDER BY p.wordcount DESC
    ) AS sub
WHERE
    sub.wordcount = (SELECT MAX(g.wordcount) FROM paragraphs g JOIN characters d ON
"""
cursor.execute(t_sql)
rs_para = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_para = pd.DataFrame(rs_para, columns=colnames)

print(df_para.iloc[0]['plaintext'])
```

```
Ay, so, God b' wi' ye!                        [Exeunt Rosencrantz and Guildenstern
[p]Now I am alone.
[p]O what a rogue and peasant slave am I!
[p]Is it not monstrous that this player here,
[p]But in a fiction, in a dream of passion,
[p]Could force his soul so to his own conceit
[p]That, from her working, all his visage wann'd,
[p]Tears in his eyes, distraction in's aspect,
[p]A broken voice, and his whole function suiting
[p]With forms to his conceit? And all for nothing!
[p]For Hecuba!
[p]What's Hecuba to him, or he to Hecuba,
[p]That he should weep for her? What would he do,
[p]Had he the motive and the cue for passion
[p]That I have? He would drown the stage with tears
[p]And cleave the general ear with horrid speech;
[p]Make mad the guilty and appal the free,
[p]Confound the ignorant, and amaze indeed
[p]The very faculties of eyes and ears.
[p]Yet I,
[p]A dull and muddy-mettled rascal, peak
[p]Like John-a-dreams, unpregnant of my cause,
[p]And can say nothing! No, not for a king,
[p]Upon whose property and most dear life
[p]A damn'd defeat was made. Am I a coward?
[p]Who calls me villain? breaks my pate across?
[p]Plucks off my beard and blows it in my face?
[p]Tweaks me by th' nose? gives me the lie i' th' throat
[p]As deep as to the lungs? Who does me this, ha?
[p]'Swounds, I should take it! for it cannot be
[p]But I am pigeon-liver'd and lack gall
[p]To make oppression bitter, or ere this
[p]I should have fatted all the region kites
[p]With this slave's offal. Bloody bawdy villain!
[p]Remorseless, treacherous, lecherous, kindless villain!
[p]O, vengeance!
[p]Why, what an ass am I! This is most brave,
[p]That I, the son of a dear father murther'd,
[p]Prompted to my revenge by heaven and hell,
[p]Must (like a whore) unpack my heart with words
[p]And fall a-cursing like a very drab,
[p]A scullion!
[p]Fie upon't! foh! About, my brain! Hum, I have heard
[p]That guilty creatures, sitting at a play,
[p]Have by the very cunning of the scene
[p]Been struck so to the soul that presently
[p]They have proclaim'd their malefactions;
[p]For murther, though it have no tongue, will speak
[p]With most miraculous organ, I'll have these Players
[p]Play something like the murther of my father
[p]Before mine uncle. I'll observe his looks;
[p]I'll tent him to the quick. If he but blench,
[p]I know my course. The spirit that I have seen
[p]May be a devil; and the devil hath power
[p]T' assume a pleasing shape; yea, and perhaps
[p]Out of my weakness and my melancholy,
```

```
[p]As he is very potent with such spirits,
[p]Abuses me to damn me. I'll have grounds
[p]More relative than this. The play's the thing
[p]Wherein I'll catch the conscience of the King. Exit.
```

Part e

Many scenes in Shakespeare's works take place in palaces or castles. Use a query to create a table that lists all of the chapters that take place in a palace. Include the work's title, the section (renamed to "act"), the chapter (renamed to "scene"), and the description of these chapters. The setting of each scene is listed in the `description` column of the "chapters" table. [Hint: be sure to account for case sensitivity] [2 points]

In [ ]:
```
#LOWER(chapters.description) LIKE '%palace%'
# Index(['workid', 'chapterid', 'section', 'chapter', 'description'], dtype='object

t_sql = """
SELECT
    w.title,
    c.section AS act,
    c.chapter AS scene,
    c.description
FROM chapters c
JOIN works w ON w.workid = c.workid
WHERE LOWER(c.description) LIKE '%palace%'
ORDER BY w.title ASC;
"""
cursor.execute(t_sql)
rs_palace = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_palace = pd.DataFrame(rs_palace, columns=colnames)
```

In [ ]:
```
pd.set_option('display.max_row',None)
```

In [ ]:
```
print(df_palace.shape)
df_palace
```

```
(125, 4)
```

Out[ ]:

	title	act	scene	description
0	All's Well That Ends Well	1.0	3.0	Rousillon. The COUNT's palace.
1	All's Well That Ends Well	1.0	2.0	Paris. The KING's palace.
2	All's Well That Ends Well	1.0	1.0	Rousillon. The COUNT's palace.
3	All's Well That Ends Well	5.0	3.0	Rousillon. The COUNT's palace.
4	All's Well That Ends Well	5.0	2.0	Rousillon. Before the COUNT's palace.
5	All's Well That Ends Well	4.0	5.0	Rousillon. The COUNT's palace.
6	All's Well That Ends Well	3.0	4.0	Rousillon. The COUNT's palace.
7	All's Well That Ends Well	3.0	3.0	Florence. Before the DUKE's palace.
8	All's Well That Ends Well	3.0	2.0	Rousillon. The COUNT's palace.
9	All's Well That Ends Well	3.0	1.0	Florence. The DUKE's palace.
10	All's Well That Ends Well	2.0	5.0	Paris. The KING's palace.
11	All's Well That Ends Well	2.0	4.0	Paris. The KING's palace.
12	All's Well That Ends Well	2.0	3.0	Paris. The KING's palace.
13	All's Well That Ends Well	2.0	2.0	Rousillon. The COUNT's palace.
14	All's Well That Ends Well	2.0	1.0	Paris. The KING's palace.
15	Antony and Cleopatra	3.0	3.0	Alexandria. CLEOPATRA's palace.
16	Antony and Cleopatra	1.0	1.0	Alexandria. A room in CLEOPATRA's palace.
17	Antony and Cleopatra	1.0	5.0	Alexandria. CLEOPATRA's palace.
18	Antony and Cleopatra	4.0	13.0	Alexandria. Cleopatra's palace.
19	Antony and Cleopatra	4.0	4.0	The same. A room in the palace.
20	Antony and Cleopatra	4.0	3.0	The same. Before the palace.
21	Antony and Cleopatra	4.0	2.0	Alexandria. CLEOPATRA's palace.
22	Antony and Cleopatra	3.0	13.0	Alexandria. CLEOPATRA's palace.
23	Antony and Cleopatra	3.0	11.0	Alexandria. CLEOPATRA's palace.
24	Antony and Cleopatra	2.0	5.0	Alexandria. CLEOPATRA's palace.
25	As You Like It	3.0	1.0	The palace
26	As You Like It	2.0	2.0	The DUKE'S palace
27	As You Like It	1.0	3.0	The DUKE's palace
28	As You Like It	1.0	2.0	A lawn before the DUKE'S palace
29	Comedy of Errors	1.0	1.0	A hall in DUKE SOLINUS'S palace.
30	Cymbeline	1.0	1.0	Britain. The garden of Cymbeline's palace.
31	Cymbeline	1.0	3.0	A room in Cymbeline's palace.
32	Cymbeline	4.0	3.0	A room in Cymbeline's palace.

	title	act	scene	description
33	Cymbeline	3.0	5.0	A room in Cymbeline's palace.
34	Cymbeline	3.0	2.0	Another room in the palace.
35	Cymbeline	3.0	1.0	Britain. A hall in Cymbeline's palace.
36	Cymbeline	2.0	2.0	Imogen's bedchamber in Cymbeline's palace:
37	Cymbeline	2.0	1.0	Britain. Before Cymbeline's palace.
38	Cymbeline	1.0	5.0	Britain. A room in Cymbeline's palace.
39	Cymbeline	1.0	6.0	The same. Another room in the palace.
40	Henry IV, Part I	1.0	1.0	London. The palace.
41	Henry IV, Part I	4.0	4.0	York. The ARCHBISHOP'S palace.
42	Henry IV, Part I	3.0	2.0	London. The palace.
43	Henry IV, Part I	1.0	3.0	London. The palace.
44	Henry IV, Part II	1.0	3.0	York. The ARCHBISHOP'S palace
45	Henry IV, Part II	3.0	1.0	Westminster. The palace
46	Henry IV, Part II	5.0	2.0	Westminster. The palace
47	Henry V	5.0	2.0	France. A royal palace.
48	Henry V	2.0	4.0	France. The KING'S palace.
49	Henry V	1.0	1.0	London. An ante-chamber in the KING'S palace.
50	Henry V	3.0	4.0	The FRENCH KING's palace.
51	Henry VI, Part I	5.0	5.0	London. The palace.
52	Henry VI, Part I	5.0	1.0	London. The palace.
53	Henry VI, Part I	3.0	4.0	Paris. The palace.
54	Henry VI, Part II	1.0	3.0	The palace.
55	Henry VI, Part II	1.0	1.0	London. The palace.
56	Henry VI, Part II	4.0	4.0	London. The palace.
57	Henry VI, Part III	4.0	4.0	London. The palace.
58	Henry VI, Part III	5.0	7.0	London. The palace.
59	Henry VI, Part III	4.0	8.0	London. The palace.
60	Henry VI, Part III	4.0	1.0	London. The palace.
61	Henry VI, Part III	3.0	3.0	France. KING LEWIS XI's palace.
62	Henry VI, Part III	3.0	2.0	London. The palace.
63	Henry VIII	1.0	1.0	London. An ante-chamber in the palace.
64	Henry VIII	5.0	1.0	London. A gallery in the palace.
65	Henry VIII	2.0	2.0	An ante-chamber in the palace.

	title	act	scene	description
66	Henry VIII	1.0	3.0	An ante-chamber in the palace.
67	Henry VIII	5.0	5.0	The palace.
68	Henry VIII	5.0	4.0	The palace yard.
69	King John	4.0	2.0	KING JOHN'S palace.
70	King John	1.0	1.0	KING JOHN'S palace.
71	King John	5.0	1.0	KING JOHN'S palace.
72	King Lear	1.0	5.0	Court before the Duke of Albany's Palace. Ente...
73	King Lear	1.0	1.0	King Lear's Palace.
74	King Lear	1.0	3.0	The Duke of Albany's Palace.
75	King Lear	4.0	2.0	Before the Duke of Albany's Palace.
76	King Lear	1.0	4.0	The Duke of Albany's Palace. Enter Kent, [disg...
77	Macbeth	1.0	4.0	Forres. The palace.
78	Macbeth	3.0	3.0	A park near the palace.
79	Macbeth	3.0	2.0	The palace.
80	Macbeth	3.0	1.0	Forres. The palace.
81	Macbeth	4.0	3.0	England. Before the King's palace.
82	Macbeth	3.0	6.0	Forres. The palace.
83	Macbeth	3.0	4.0	The same. Hall in the palace.
84	Measure for Measure	1.0	1.0	An apartment in the DUKE'S palace.
85	Midsummer Night's Dream	1.0	1.0	Athens. The palace of THESEUS.
86	Midsummer Night's Dream	5.0	1.0	Athens. The palace of THESEUS.
87	Pericles	2.0	5.0	Pentapolis. A room in the palace.
88	Pericles	1.0	1.0	Antioch. A room in the palace.
89	Pericles	1.0	2.0	Tyre. A room in the palace.
90	Pericles	1.0	3.0	Tyre. An ante-chamber in the palace.
91	Richard II	5.0	3.0	A royal palace.
92	Richard II	5.0	2.0	The DUKE OF YORK's palace.
93	Richard II	2.0	2.0	The palace.
94	Richard II	1.0	2.0	The DUKE OF LANCASTER'S palace.
95	Richard II	1.0	1.0	London. KING RICHARD II's palace.
96	Richard III	4.0	4.0	Before the palace.
97	Richard III	4.0	2.0	London. The palace.
98	Richard III	2.0	4.0	London. The palace.

	title	act	scene	description
99	Richard III	2.0	2.0	The palace.
100	Richard III	2.0	1.0	London. The palace.
101	Richard III	1.0	3.0	The palace.
102	The Winter's Tale	4.0	2.0	Bohemia. The palace of POLIXENES.
103	The Winter's Tale	2.0	3.0	A room in LEONTES' palace.
104	The Winter's Tale	5.0	1.0	A room in LEONTES' palace.
105	The Winter's Tale	2.0	1.0	A room in LEONTES' palace.
106	The Winter's Tale	1.0	1.0	Antechamber in LEONTES' palace.
107	The Winter's Tale	5.0	2.0	Before LEONTES' palace.
108	Titus Andronicus	4.0	4.0	The same. Before the palace.
109	Titus Andronicus	2.0	1.0	Rome. Before the Palace.
110	Titus Andronicus	4.0	2.0	The same. A room in the palace.
111	Troilus and Cressida	2.0	2.0	Troy. A room in Priam's palace.
112	Troilus and Cressida	5.0	3.0	Troy. Before Priam's palace.
113	Troilus and Cressida	3.0	1.0	Troy. Priam's palace.
114	Troilus and Cressida	1.0	1.0	Troy. Before Priam's palace.
115	Twelfth Night	1.0	4.0	DUKE ORSINO's palace.
116	Twelfth Night	1.0	1.0	DUKE ORSINO's palace.
117	Twelfth Night	2.0	4.0	DUKE ORSINO's palace.
118	Two Gentlemen of Verona	2.0	1.0	Milan. The DUKE's palace.
119	Two Gentlemen of Verona	3.0	2.0	The same. The DUKE's palace.
120	Two Gentlemen of Verona	3.0	1.0	Milan. The DUKE's palace.
121	Two Gentlemen of Verona	4.0	2.0	Milan. Outside the DUKE's palace, under SILVIA...
122	Two Gentlemen of Verona	5.0	2.0	The same. The DUKE's palace.
123	Two Gentlemen of Verona	2.0	6.0	The same. The DUKE'S palace.
124	Two Gentlemen of Verona	2.0	4.0	Milan. The DUKE's palace.

Part f

Create a table that lists characters, the plays that the characters appear in, the number of speeches the character gives, and the average length of the speeches that the character gives. Display the character description and the work title, not the ID values. Sort the table by average speech length, and restrict the table to only those characters that give at least 20 speeches. [Hint: you will need to use a subquery.] [2 points]

Answer Number One:

Below is a first answer to this question. I also provide a second answer.

In [ ]:
```
t_sql = """
SELECT
    sub.charname,
    sub.title as plays,
    c.speechcount AS total_speeches,
    AVG(wordcount) AS avg_wordcount
FROM (
    SELECT DISTINCT
        c.charid,
        w.title AS title,
        w.workid AS workid,
        c.charname AS charname,
        c.speechcount AS speechcount,
        p.wordcount AS wordcount
    FROM works w
    JOIN characters c ON c.works LIKE w.workid
    JOIN paragraphs p ON c.charid = p.charid AND c.works LIKE p.workid
    WHERE c.speechcount > 19) AS sub
JOIN characters c ON c.charid = sub.charid
GROUP BY sub.charname, c.speechcount, sub.title
ORDER BY sub.charname;
"""

#(SELECT c.charid from characters c WHERE speechcount > 19)
cursor.execute(t_sql)
rs_chars = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_chars = pd.DataFrame(rs_chars, columns=colnames)
df_chars
```

Out[ ]:

	charname	plays	total_speeches	avg_wordcount
0	(stage directions)	As You Like It	126.0	6.454545
1	Aaron	Titus Andronicus	57.0	58.146341
2	Achilles	Troilus and Cressida	74.0	29.272727
3	Adriana	Comedy of Errors	79.0	45.142857
4	Aeneas	Troilus and Cressida	44.0	32.807692
5	Agamemnon	Troilus and Cressida	52.0	36.151515
6	Agrippa	Antony and Cleopatra	28.0	19.000000
7	Ajax	Troilus and Cressida	55.0	15.000000
8	Alcibiades	Timon of Athens	39.0	45.608696
9	Alice	Henry V	22.0	10.000000
10	Alonso	Tempest	40.0	26.560000
11	Angelo	Comedy of Errors	31.0	26.000000
12	Angelo	Measure for Measure	83.0	43.888889
13	Antipholus of Ephesus	Comedy of Errors	76.0	41.875000
14	Antipholus of Syracuse	Comedy of Errors	103.0	37.500000
15	Antonio	Much Ado about Nothing	23.0	24.285714
16	Antonio	Twelfth Night	26.0	35.947368
17	Antonio	Merchant of Venice	47.0	39.111111
18	Antonio	Tempest	57.0	28.730769
19	Apemantus	Timon of Athens	100.0	37.187500
20	Archbishop Cranmer	Henry VIII	21.0	66.642857
21	Ariel	Tempest	45.0	43.695652
22	Arthur	King John	23.0	44.272727
23	Arviragus	Cymbeline	46.0	28.703704
24	Autolycus	The Winter's Tale	67.0	50.951220
25	Banquo	Macbeth	33.0	29.954545
26	Baptista Minola	Taming of the Shrew	68.0	30.187500
27	Bassanio	Merchant of Venice	73.0	48.372093
28	Bawd	Pericles	43.0	22.766667
29	Beatrice	Much Ado about Nothing	106.0	32.318182
30	Belarius	Cymbeline	58.0	55.487805
31	Benedick	Much Ado about Nothing	134.0	47.611111
32	Benvolio	Romeo and Juliet	64.0	30.107143

	charname	plays	total_speeches	avg_wordcount
33	Bertram	All's Well That Ends Well	102.0	33.500000
34	Bianca	Taming of the Shrew	29.0	22.812500
35	Biondello	Taming of the Shrew	39.0	27.523810
36	Biron	Love's Labour's Lost	159.0	70.980769
37	Borachio	Much Ado about Nothing	38.0	31.689655
38	Bottom	Midsummer Night's Dream	59.0	42.743590
39	Boult	Pericles	38.0	18.769231
40	Boyet	Love's Labour's Lost	80.0	40.171429
41	Brabantio	Othello	30.0	42.652174
42	Brutus	Julius Caesar	194.0	56.492537
43	Caesar	Julius Caesar	42.0	31.000000
44	Caius Lucius	Cymbeline	25.0	33.944444
45	Caliban	Tempest	50.0	35.033333
46	Camillo	The Winter's Tale	72.0	40.463415
47	Capulet	Romeo and Juliet	51.0	50.027027
48	Cardinal Pandulph	King John	23.0	61.950000
49	Cardinal Wolsey	Henry VIII	79.0	60.739130
50	Casca	Julius Caesar	39.0	37.370370
51	Cassio	Othello	110.0	29.000000
52	Cassius	Julius Caesar	140.0	52.592593
53	Celia	As You Like It	108.0	31.075000
54	Cerimon	Pericles	23.0	34.550000
55	Charles, King of France	Henry VI, Part I	41.0	31.629630
56	Charmian	Antony and Cleopatra	63.0	18.454545
57	Chiron	Titus Andronicus	30.0	16.647059
58	Christopher Sly	Taming of the Shrew	24.0	23.266667
59	Claudio	Measure for Measure	35.0	33.571429
60	Claudio	Much Ado about Nothing	125.0	28.880952
61	Claudius	Hamlet	106.0	66.018182
62	Cleopatra	Antony and Cleopatra	204.0	45.507937
63	Cloten	Cymbeline	77.0	43.486486
64	Clown	All's Well That Ends Well	58.0	35.058824
65	Clown	The Winter's Tale	64.0	32.470588

	charname	plays	total_speeches	avg_wordcount
66	Cominius	Coriolanus	67.0	41.050000
67	Conrade	Much Ado about Nothing	23.0	14.071429
68	Constable of France	Henry V	40.0	28.814815
69	Constance	King John	36.0	66.800000
70	Cordelia	King Lear	31.0	32.304348
71	Corin	As You Like It	24.0	30.687500
72	Coriolanus	Coriolanus	189.0	64.415584
73	Costard	Love's Labour's Lost	83.0	28.611111
74	Countess	All's Well That Ends Well	87.0	41.452381
75	Cressida	Troilus and Cressida	152.0	33.325000
76	Cromwell	Henry VIII	21.0	17.600000
77	Curtis	Taming of the Shrew	20.0	10.083333
78	Cymbeline	Cymbeline	81.0	37.487179
79	Demetrius	Titus Andronicus	39.0	23.695652
80	Demetrius	Midsummer Night's Dream	48.0	26.533333
81	Desdemona	Othello	165.0	37.413043
82	Diana	All's Well That Ends Well	44.0	29.038462
83	Dick the Butcher	Henry VI, Part II	24.0	10.733333
84	Diomedes	Troilus and Cressida	54.0	23.500000
85	Doctor	Macbeth	20.0	18.375000
86	Doctor Caius	Merry Wives of Windsor	49.0	19.750000
87	Dogberry	Much Ado about Nothing	52.0	35.057143
88	Dolabella	Antony and Cleopatra	23.0	18.857143
89	Doll Tearsheet	Henry IV, Part II	31.0	24.181818
90	Domitius Enobarus	Antony and Cleopatra	113.0	35.177778
91	Don Adriano de Armado	Love's Labour's Lost	102.0	38.472222
92	Don John	Much Ado about Nothing	40.0	27.148148
93	Don Pedro	Much Ado about Nothing	135.0	33.372093
94	Dromio of Ephesus	Comedy of Errors	63.0	28.705882
95	Dromio of Syracuse	Comedy of Errors	99.0	29.761905
96	Duchess of York	Richard II	28.0	29.695652
97	Duchess of York	Richard III	43.0	34.178571
98	Duke	As You Like It	32.0	35.400000

	charname	plays	total_speeches	avg_wordcount
99	Duke of Albany	King Lear	58.0	27.354839
100	Duke of Aumerle	Richard II	38.0	23.150000
101	Duke of Buckingham	Henry VIII	26.0	57.307692
102	Duke of Cornwall	King Lear	53.0	20.500000
103	Duke of Milan	Two Gentlemen of Verona	48.0	43.033333
104	Duke of Norfolk	Henry VIII	48.0	45.464286
105	Duke of Orleans	Henry V	29.0	11.625000
106	Duke of Suffolk	Henry VIII	30.0	22.913043
107	Duke of Venice	Othello	25.0	26.000000
108	Dumain	Love's Labour's Lost	54.0	20.733333
109	Earl of Glouchester	King Lear	118.0	38.936170
110	Earl of Kent	King Lear	127.0	38.367347
111	Earl of Surrey	Henry VIII	24.0	31.400000
112	Earl of Worcester	Henry IV, Part I	35.0	53.791667
113	Edgar	King Lear	98.0	49.695652
114	Edmund	King Lear	79.0	46.772727
115	Edmund of Langley	Richard II	54.0	52.513514
116	Elbow	Measure for Measure	28.0	22.047619
117	Eleanor	Henry VI, Part II	21.0	47.400000
118	Emilia	Othello	103.0	30.769231
119	Eros	Antony and Cleopatra	27.0	14.400000
120	Escalus	Measure for Measure	78.0	28.388889
121	Fabian	Twelfth Night	51.0	25.227273
122	Fenton	Merry Wives of Windsor	20.0	47.071429
123	Ferdinand	Tempest	31.0	38.038462
124	Ferdinand	Love's Labour's Lost	117.0	46.885714
125	Feste	Twelfth Night	104.0	33.062500
126	First Citizen	Coriolanus	33.0	25.380952
127	First Clown	Hamlet	34.0	25.160000
128	First Gentleman	Henry VIII	34.0	26.956522
129	First Lord	Timon of Athens	28.0	14.187500
130	First Lord	All's Well That Ends Well	48.0	21.633333
131	First Murderer	Macbeth	21.0	11.727273

	charname	plays	total_speeches	avg_wordcount
132	First Murderer	Richard III	33.0	15.777778
133	First Senator	Timon of Athens	27.0	28.222222
134	First Senator	Coriolanus	33.0	23.952381
135	First Soldier	All's Well That Ends Well	37.0	21.454545
136	First Witch	Macbeth	23.0	20.066667
137	Flavius	Timon of Athens	41.0	45.500000
138	Florizel	The Winter's Tale	45.0	39.064516
139	Fluellen	Henry V	68.0	43.460000
140	Fool	King Lear	58.0	39.617647
141	Ford	Merry Wives of Windsor	99.0	44.869565
142	Frederick	As You Like It	20.0	28.777778
143	Friar Laurence	Romeo and Juliet	55.0	67.416667
144	Gardiner	Henry VIII	22.0	34.833333
145	Gentleman	King Lear	41.0	22.652174
146	Gertrude	Hamlet	72.0	30.400000
147	Glendower	Henry IV, Part I	23.0	30.222222
148	Goneril	King Lear	53.0	35.138889
149	Gonzalo	Tempest	52.0	28.194444
150	Gratiano	Othello	20.0	13.416667
151	Gratiano	Merchant of Venice	48.0	36.343750
152	Gremio	Taming of the Shrew	58.0	35.678571
153	Grumio	Taming of the Shrew	63.0	33.781250
154	Guiderius	Cymbeline	62.0	23.882353
155	Guildenstern	Hamlet	29.0	15.058824
156	Hamlet	Hamlet	358.0	78.980000
157	Hector	Troilus and Cressida	57.0	40.058824
158	Helena	Midsummer Night's Dream	36.0	62.400000
159	Helena	All's Well That Ends Well	109.0	52.377358
160	Helicanus	Pericles	37.0	29.615385
161	Henry VIII	Henry VIII	81.0	57.900000
162	Hermia	Midsummer Night's Dream	48.0	38.482759
163	Hermione	The Winter's Tale	35.0	51.633333
164	Hero	Much Ado about Nothing	44.0	30.541667

	charname	plays	total_speeches	avg_wordcount
165	Holofernes	Love's Labour's Lost	54.0	35.068966
166	Horatio	Hamlet	109.0	45.916667
167	Hortensio	Taming of the Shrew	70.0	30.702703
168	Host	Merry Wives of Windsor	46.0	23.423077
169	Hubert de Burgh	King John	52.0	26.900000
170	Iachimo	Cymbeline	77.0	52.300000
171	Iago	Othello	272.0	67.638554
172	Imogen	Cymbeline	118.0	60.859649
173	Isabella	Measure for Measure	129.0	35.803571
174	Jack Cade	Henry VI, Part II	61.0	43.500000
175	Jaques (lord)	As You Like It	57.0	47.468750
176	Jessica	Merchant of Venice	26.0	27.478261
177	Joan la Pucelle	Henry VI, Part I	46.0	53.606061
178	John of Gaunt	Richard II	28.0	64.086957
179	Julia	Two Gentlemen of Verona	107.0	44.452381
180	Juliet	Romeo and Juliet	118.0	55.727273
181	Junius Brutus	Coriolanus	91.0	30.305556
182	Katharine	Love's Labour's Lost	25.0	18.142857
183	Katharine	Henry V	33.0	17.187500
184	Katherina	Taming of the Shrew	82.0	41.833333
185	King John	King John	95.0	47.872727
186	King Lewis XI	Henry VI, Part III	21.0	30.533333
187	King of France	All's Well That Ends Well	87.0	47.960784
188	King Phillip	King John	43.0	51.653846
189	King Richard II	Richard II	98.0	77.704225
190	Lady Anne	Richard III	51.0	38.185185
191	Lady Capulet	Romeo and Juliet	45.0	25.407407
192	Lady Macbeth	Macbeth	59.0	45.000000
193	Laertes	Hamlet	62.0	36.500000
194	Lafeu	All's Well That Ends Well	97.0	33.390244
195	Launce	Two Gentlemen of Verona	68.0	43.941176
196	Launcelot Gobbo	Merchant of Venice	44.0	38.750000
197	Lear	King Lear	188.0	55.671053

	charname	plays	total_speeches	avg_wordcount
198	Lennox	Macbeth	21.0	31.666667
199	Leonato	Much Ado about Nothing	120.0	38.155556
200	Leontes	The Winter's Tale	125.0	59.208955
201	Lewis	King John	29.0	52.750000
202	Lewis the Dauphin	Henry V	31.0	33.291667
203	Lodovico	Othello	33.0	21.181818
204	Longaville	Love's Labour's Lost	40.0	18.687500
205	Lord Chamberlain	Henry VIII	38.0	34.518519
206	Lord Chief Justice	Henry IV, Part II	56.0	29.607143
207	Lord Talbot/Earl of Shrewsbury	Henry VI, Part I	70.0	65.127660
208	Lorenzo	Merchant of Venice	47.0	33.575758
209	Lucentio	Taming of the Shrew	61.0	35.193548
210	Lucetta	Two Gentlemen of Verona	48.0	13.950000
211	Luciana	Comedy of Errors	43.0	29.555556
212	Lucio	Measure for Measure	111.0	32.822222
213	Lucius	Julius Caesar	24.0	12.272727
214	Lucius	Titus Andronicus	51.0	41.266667
215	Lysander	Midsummer Night's Dream	50.0	33.636364
216	Lysimachus	Pericles	40.0	24.291667
217	Macbeth	Macbeth	146.0	55.057971
218	Macduff	Macbeth	59.0	30.000000
219	Malcolm	Macbeth	40.0	43.866667
220	Malvolio	Twelfth Night	87.0	43.897436
221	Marcellus	Hamlet	37.0	22.000000
222	Marcus Andronicus	Titus Andronicus	63.0	59.945946
223	Margaret	Much Ado about Nothing	26.0	27.125000
224	Maria	Love's Labour's Lost	22.0	16.642857
225	Maria	Twelfth Night	59.0	28.137931
226	Mariana	Measure for Measure	24.0	28.375000
227	Marina	Pericles	63.0	29.939394
228	Menas	Antony and Cleopatra	35.0	13.550000
229	Menenius Agrippa	Coriolanus	162.0	46.050000
230	Mercutio	Romeo and Juliet	62.0	46.243243

	charname	plays	total_speeches	avg_wordcount
231	Messala	Julius Caesar	20.0	15.166667
232	Messenger	Antony and Cleopatra	42.0	18.875000
233	Miranda	Tempest	49.0	26.040000
234	Mistress Ford	Merry Wives of Windsor	85.0	23.000000
235	Mistress Page	Merry Wives of Windsor	101.0	38.288889
236	Montano	Othello	24.0	22.058824
237	Moth	Love's Labour's Lost	78.0	24.142857
238	Nerissa	Merchant of Venice	36.0	24.952381
239	Nestor	Troilus and Cressida	38.0	47.090909
240	Nurse	Romeo and Juliet	90.0	40.930233
241	Oberon	Midsummer Night's Dream	29.0	58.629630
242	Old Shepherd	The Winter's Tale	42.0	31.933333
243	Oliver	As You Like It	37.0	42.708333
244	Olivia	Twelfth Night	118.0	36.636364
245	Ophelia	Hamlet	58.0	32.000000
246	Orlando	As You Like It	120.0	41.157895
247	Orsino	Twelfth Night	59.0	38.702703
248	Osric	Hamlet	25.0	19.125000
249	Oswald	King Lear	38.0	21.631579
250	Othello	Othello	274.0	58.549296
251	Page	Merry Wives of Windsor	75.0	20.566667
252	Painter	Timon of Athens	30.0	20.473684
253	Pandarus	Troilus and Cressida	153.0	35.686275
254	Paris	Romeo and Juliet	23.0	26.800000
255	Paris	Troilus and Cressida	27.0	32.095238
256	Parolles	All's Well That Ends Well	141.0	40.739130
257	Patroclus	Troilus and Cressida	37.0	18.250000
258	Paulina	The Winter's Tale	59.0	46.600000
259	Pedant	Taming of the Shrew	20.0	20.937500
260	Pembroke	King John	20.0	30.600000
261	Perdita	The Winter's Tale	25.0	38.391304
262	Pericles	Pericles	121.0	57.328125
263	Petruchio	Taming of the Shrew	158.0	53.687500

	charname	plays	total_speeches	avg_wordcount
264	Phebe	As You Like It	23.0	38.470588
265	Philip the Bastard	King John	89.0	68.061224
266	Pisanio	Cymbeline	58.0	39.312500
267	Poet	Timon of Athens	30.0	28.608696
268	Polixenes	The Winter's Tale	57.0	48.361111
269	Polonius	Hamlet	86.0	54.536585
270	Pompey	Antony and Cleopatra	41.0	27.555556
271	Pompey	Measure for Measure	60.0	29.636364
272	Portia	Merchant of Venice	117.0	59.864407
273	Posthumus Leonatus	Cymbeline	77.0	63.304348
274	Princess of France	Love's Labour's Lost	102.0	38.333333
275	Prospero	Tempest	115.0	56.138462
276	Proteus	Two Gentlemen of Verona	147.0	50.913043
277	Provost	Measure for Measure	65.0	27.090909
278	Puck	Midsummer Night's Dream	33.0	48.777778
279	Queen	Richard II	25.0	41.090909
280	Queen	Cymbeline	27.0	48.840000
281	Queen Elinor	King John	22.0	23.000000
282	Queen Katharine	Henry VIII	50.0	63.975000
283	Quince	Midsummer Night's Dream	40.0	32.160000
284	Regan	King Lear	73.0	28.714286
285	Reignier	Henry VI, Part I	24.0	20.947368
286	Roderigo	Othello	59.0	25.480000
287	Romeo	Romeo and Juliet	163.0	52.359375
288	Rosalind	As You Like It	201.0	53.794118
289	Rosaline	Love's Labour's Lost	75.0	31.000000
290	Rosencrantz	Hamlet	48.0	21.240000
291	Ross	Macbeth	39.0	31.600000
292	Salarino	Merchant of Venice	27.0	33.571429
293	Salisbury	King John	36.0	41.111111
294	Sampson	Romeo and Juliet	20.0	13.812500
295	Sebastian	Twelfth Night	31.0	36.153846
296	Sebastian	Tempest	67.0	17.083333

	charname	plays	total_speeches	avg_wordcount
297	Second Gentleman	Henry VIII	37.0	27.043478
298	Second Lord	Cymbeline	20.0	18.357143
299	Second Lord	Timon of Athens	29.0	13.642857
300	Second Lord	All's Well That Ends Well	57.0	28.129032
301	Second Murderer	Richard III	30.0	20.941176
302	Shylock	Merchant of Venice	79.0	55.772727
303	Sicinius Velutus	Coriolanus	117.0	27.952381
304	Silence	Henry IV, Part II	22.0	14.000000
305	Silvia	Two Gentlemen of Verona	58.0	29.866667
306	Silvius	As You Like It	24.0	26.800000
307	Simonides	Pericles	42.0	32.413793
308	Simple	Merry Wives of Windsor	25.0	15.444444
309	Sir Andrew Aguecheek	Twelfth Night	88.0	19.964286
310	Sir Hugh Evans	Merry Wives of Windsor	87.0	28.047619
311	Sir Thomas Lovell	Henry VIII	21.0	26.333333
312	Sir Toby Belch	Twelfth Night	152.0	32.914894
313	Sir William Catesby	Richard III	31.0	19.631579
314	Slender	Merry Wives of Windsor	56.0	27.166667
315	Solinus	Comedy of Errors	22.0	40.187500
316	Speed	Two Gentlemen of Verona	117.0	25.343750
317	Stephano	Tempest	60.0	27.270270
318	Tamora	Titus Andronicus	49.0	55.757576
319	Thaisa	Pericles	32.0	19.434783
320	Thersites	Troilus and Cressida	90.0	42.136364
321	Theseus	Midsummer Night's Dream	48.0	48.468750
322	Third Servingman	Coriolanus	20.0	23.923077
323	Thurio	Two Gentlemen of Verona	36.0	15.235294
324	Timon	Timon of Athens	210.0	67.402985
325	Titania	Midsummer Night's Dream	23.0	56.700000
326	Titus Andronicus	Titus Andronicus	117.0	68.910448
327	Titus Lartius	Coriolanus	23.0	17.736842
328	Touchstone	As You Like It	74.0	44.590909
329	Tranio	Taming of the Shrew	90.0	34.902439

	charname	plays	total_speeches	avg_wordcount
330	Trinculo	Tempest	39.0	26.230769
331	Troilus	Troilus and Cressida	131.0	54.448276
332	Tullus Aufidius	Coriolanus	45.0	56.222222
333	Ulysses	Troilus and Cressida	80.0	70.600000
334	Valentine	Two Gentlemen of Verona	149.0	38.020833
335	Vincentio	Taming of the Shrew	23.0	16.882353
336	Vincentio	Measure for Measure	194.0	57.184211
337	Viola	Twelfth Night	121.0	32.319149
338	Virgilia	Coriolanus	26.0	9.785714
339	Volumnia	Coriolanus	57.0	52.560976
340	Widow	All's Well That Ends Well	21.0	25.214286
341	Williams	Henry V	28.0	30.538462

Answer Number Two:

The use of "Plays" and "Average" made me wonder if the request was for a comma separated list of plays and the total average word count. This is my second answer for the question.

In [ ]:
```
t_sql = """
SELECT
    sub2.charname,
    STRING_AGG(sub2.plays, '; ') AS plays,
    SUM(sub2.total_speeches) AS total_speeches,
    AVG(sub2.avg_wordcount) AS avg_wordcount
FROM (
    SELECT
        sub.charname,
        sub.title as plays,
        c.speechcount AS total_speeches,
        AVG(wordcount) AS avg_wordcount
    FROM (
        SELECT DISTINCT
            c.charid,
            w.title AS title,
            w.workid AS workid,
            c.charname AS charname,
            c.speechcount AS speechcount,
            p.wordcount AS wordcount
        FROM works w
        JOIN characters c ON c.works LIKE w.workid
        JOIN paragraphs p ON c.charid = p.charid AND c.works LIKE p.workid
        WHERE c.speechcount > 19) AS sub
    JOIN characters c ON c.charid = sub.charid
    GROUP BY sub.charname, c.speechcount, sub.title
```

```
    ORDER BY sub.charname
) AS sub2
GROUP BY sub2.charname;
"""

#(SELECT c.charid from characters c WHERE speechcount > 19)
cursor.execute(t_sql)
rs_chars = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_chars = pd.DataFrame(rs_chars, columns=colnames)
df_chars
```

Out[ ]:

	charname	plays	total_speeches	avg_wordcount
0	(stage directions)	As You Like It	126.0	6.454545
1	Aaron	Titus Andronicus	57.0	58.146341
2	Achilles	Troilus and Cressida	74.0	29.272727
3	Adriana	Comedy of Errors	79.0	45.142857
4	Aeneas	Troilus and Cressida	44.0	32.807692
5	Agamemnon	Troilus and Cressida	52.0	36.151515
6	Agrippa	Antony and Cleopatra	28.0	19.000000
7	Ajax	Troilus and Cressida	55.0	15.000000
8	Alcibiades	Timon of Athens	39.0	45.608696
9	Alice	Henry V	22.0	10.000000
10	Alonso	Tempest	40.0	26.560000
11	Angelo	Comedy of Errors; Measure for Measure	114.0	34.944444
12	Antipholus of Ephesus	Comedy of Errors	76.0	41.875000
13	Antipholus of Syracuse	Comedy of Errors	103.0	37.500000
14	Antonio	Much Ado about Nothing; Twelfth Night; Merchan...	153.0	32.018741
15	Apemantus	Timon of Athens	100.0	37.187500
16	Archbishop Cranmer	Henry VIII	21.0	66.642857
17	Ariel	Tempest	45.0	43.695652
18	Arthur	King John	23.0	44.272727
19	Arviragus	Cymbeline	46.0	28.703704
20	Autolycus	The Winter's Tale	67.0	50.951220
21	Banquo	Macbeth	33.0	29.954545
22	Baptista Minola	Taming of the Shrew	68.0	30.187500
23	Bassanio	Merchant of Venice	73.0	48.372093
24	Bawd	Pericles	43.0	22.766667
25	Beatrice	Much Ado about Nothing	106.0	32.318182
26	Belarius	Cymbeline	58.0	55.487805
27	Benedick	Much Ado about Nothing	134.0	47.611111
28	Benvolio	Romeo and Juliet	64.0	30.107143
29	Bertram	All's Well That Ends Well	102.0	33.500000
30	Bianca	Taming of the Shrew	29.0	22.812500
31	Biondello	Taming of the Shrew	39.0	27.523810

	charname	plays	total_speeches	avg_wordcount
32	Biron	Love's Labour's Lost	159.0	70.980769
33	Borachio	Much Ado about Nothing	38.0	31.689655
34	Bottom	Midsummer Night's Dream	59.0	42.743590
35	Boult	Pericles	38.0	18.769231
36	Boyet	Love's Labour's Lost	80.0	40.171429
37	Brabantio	Othello	30.0	42.652174
38	Brutus	Julius Caesar	194.0	56.492537
39	Caesar	Julius Caesar	42.0	31.000000
40	Caius Lucius	Cymbeline	25.0	33.944444
41	Caliban	Tempest	50.0	35.033333
42	Camillo	The Winter's Tale	72.0	40.463415
43	Capulet	Romeo and Juliet	51.0	50.027027
44	Cardinal Pandulph	King John	23.0	61.950000
45	Cardinal Wolsey	Henry VIII	79.0	60.739130
46	Casca	Julius Caesar	39.0	37.370370
47	Cassio	Othello	110.0	29.000000
48	Cassius	Julius Caesar	140.0	52.592593
49	Celia	As You Like It	108.0	31.075000
50	Cerimon	Pericles	23.0	34.550000
51	Charles, King of France	Henry VI, Part I	41.0	31.629630
52	Charmian	Antony and Cleopatra	63.0	18.454545
53	Chiron	Titus Andronicus	30.0	16.647059
54	Christopher Sly	Taming of the Shrew	24.0	23.266667
55	Claudio	Measure for Measure; Much Ado about Nothing	160.0	31.226190
56	Claudius	Hamlet	106.0	66.018182
57	Cleopatra	Antony and Cleopatra	204.0	45.507937
58	Cloten	Cymbeline	77.0	43.486486
59	Clown	All's Well That Ends Well; The Winter's Tale	122.0	33.764706
60	Cominius	Coriolanus	67.0	41.050000
61	Conrade	Much Ado about Nothing	23.0	14.071429
62	Constable of France	Henry V	40.0	28.814815
63	Constance	King John	36.0	66.800000

	charname	plays	total_speeches	avg_wordcount
64	Cordelia	King Lear	31.0	32.304348
65	Corin	As You Like It	24.0	30.687500
66	Coriolanus	Coriolanus	189.0	64.415584
67	Costard	Love's Labour's Lost	83.0	28.611111
68	Countess	All's Well That Ends Well	87.0	41.452381
69	Cressida	Troilus and Cressida	152.0	33.325000
70	Cromwell	Henry VIII	21.0	17.600000
71	Curtis	Taming of the Shrew	20.0	10.083333
72	Cymbeline	Cymbeline	81.0	37.487179
73	Demetrius	Titus Andronicus; Midsummer Night's Dream	87.0	25.114493
74	Desdemona	Othello	165.0	37.413043
75	Diana	All's Well That Ends Well	44.0	29.038462
76	Dick the Butcher	Henry VI, Part II	24.0	10.733333
77	Diomedes	Troilus and Cressida	54.0	23.500000
78	Doctor	Macbeth	20.0	18.375000
79	Doctor Caius	Merry Wives of Windsor	49.0	19.750000
80	Dogberry	Much Ado about Nothing	52.0	35.057143
81	Dolabella	Antony and Cleopatra	23.0	18.857143
82	Doll Tearsheet	Henry IV, Part II	31.0	24.181818
83	Domitius Enobarus	Antony and Cleopatra	113.0	35.177778
84	Don Adriano de Armado	Love's Labour's Lost	102.0	38.472222
85	Don John	Much Ado about Nothing	40.0	27.148148
86	Don Pedro	Much Ado about Nothing	135.0	33.372093
87	Dromio of Ephesus	Comedy of Errors	63.0	28.705882
88	Dromio of Syracuse	Comedy of Errors	99.0	29.761905
89	Duchess of York	Richard II; Richard III	71.0	31.937112
90	Duke	As You Like It	32.0	35.400000
91	Duke of Albany	King Lear	58.0	27.354839
92	Duke of Aumerle	Richard II	38.0	23.150000
93	Duke of Buckingham	Henry VIII	26.0	57.307692
94	Duke of Cornwall	King Lear	53.0	20.500000
95	Duke of Milan	Two Gentlemen of Verona	48.0	43.033333

	charname	plays	total_speeches	avg_wordcount
96	Duke of Norfolk	Henry VIII	48.0	45.464286
97	Duke of Orleans	Henry V	29.0	11.625000
98	Duke of Suffolk	Henry VIII	30.0	22.913043
99	Duke of Venice	Othello	25.0	26.000000
100	Dumain	Love's Labour's Lost	54.0	20.733333
101	Earl of Glouchester	King Lear	118.0	38.936170
102	Earl of Kent	King Lear	127.0	38.367347
103	Earl of Surrey	Henry VIII	24.0	31.400000
104	Earl of Worcester	Henry IV, Part I	35.0	53.791667
105	Edgar	King Lear	98.0	49.695652
106	Edmund	King Lear	79.0	46.772727
107	Edmund of Langley	Richard II	54.0	52.513514
108	Elbow	Measure for Measure	28.0	22.047619
109	Eleanor	Henry VI, Part II	21.0	47.400000
110	Emilia	Othello	103.0	30.769231
111	Eros	Antony and Cleopatra	27.0	14.400000
112	Escalus	Measure for Measure	78.0	28.388889
113	Fabian	Twelfth Night	51.0	25.227273
114	Fenton	Merry Wives of Windsor	20.0	47.071429
115	Ferdinand	Tempest; Love's Labour's Lost	148.0	42.462088
116	Feste	Twelfth Night	104.0	33.062500
117	First Citizen	Coriolanus	33.0	25.380952
118	First Clown	Hamlet	34.0	25.160000
119	First Gentleman	Henry VIII	34.0	26.956522
120	First Lord	Timon of Athens; All's Well That Ends Well	76.0	17.910417
121	First Murderer	Macbeth; Richard III	54.0	13.752525
122	First Senator	Timon of Athens; Coriolanus	60.0	26.087302
123	First Soldier	All's Well That Ends Well	37.0	21.454545
124	First Witch	Macbeth	23.0	20.066667
125	Flavius	Timon of Athens	41.0	45.500000
126	Florizel	The Winter's Tale	45.0	39.064516
127	Fluellen	Henry V	68.0	43.460000

	charname	plays	total_speeches	avg_wordcount
128	Fool	King Lear	58.0	39.617647
129	Ford	Merry Wives of Windsor	99.0	44.869565
130	Frederick	As You Like It	20.0	28.777778
131	Friar Laurence	Romeo and Juliet	55.0	67.416667
132	Gardiner	Henry VIII	22.0	34.833333
133	Gentleman	King Lear	41.0	22.652174
134	Gertrude	Hamlet	72.0	30.400000
135	Glendower	Henry IV, Part I	23.0	30.222222
136	Goneril	King Lear	53.0	35.138889
137	Gonzalo	Tempest	52.0	28.194444
138	Gratiano	Othello; Merchant of Venice	68.0	24.880208
139	Gremio	Taming of the Shrew	58.0	35.678571
140	Grumio	Taming of the Shrew	63.0	33.781250
141	Guiderius	Cymbeline	62.0	23.882353
142	Guildenstern	Hamlet	29.0	15.058824
143	Hamlet	Hamlet	358.0	78.980000
144	Hector	Troilus and Cressida	57.0	40.058824
145	Helena	Midsummer Night's Dream; All's Well That Ends ...	145.0	57.388679
146	Helicanus	Pericles	37.0	29.615385
147	Henry VIII	Henry VIII	81.0	57.900000
148	Hermia	Midsummer Night's Dream	48.0	38.482759
149	Hermione	The Winter's Tale	35.0	51.633333
150	Hero	Much Ado about Nothing	44.0	30.541667
151	Holofernes	Love's Labour's Lost	54.0	35.068966
152	Horatio	Hamlet	109.0	45.916667
153	Hortensio	Taming of the Shrew	70.0	30.702703
154	Host	Merry Wives of Windsor	46.0	23.423077
155	Hubert de Burgh	King John	52.0	26.900000
156	Iachimo	Cymbeline	77.0	52.300000
157	Iago	Othello	272.0	67.638554
158	Imogen	Cymbeline	118.0	60.859649
159	Isabella	Measure for Measure	129.0	35.803571

	charname	plays	total_speeches	avg_wordcount
160	Jack Cade	Henry VI, Part II	61.0	43.500000
161	Jaques (lord)	As You Like It	57.0	47.468750
162	Jessica	Merchant of Venice	26.0	27.478261
163	Joan la Pucelle	Henry VI, Part I	46.0	53.606061
164	John of Gaunt	Richard II	28.0	64.086957
165	Julia	Two Gentlemen of Verona	107.0	44.452381
166	Juliet	Romeo and Juliet	118.0	55.727273
167	Junius Brutus	Coriolanus	91.0	30.305556
168	Katharine	Love's Labour's Lost; Henry V	58.0	17.665179
169	Katherina	Taming of the Shrew	82.0	41.833333
170	King John	King John	95.0	47.872727
171	King Lewis XI	Henry VI, Part III	21.0	30.533333
172	King of France	All's Well That Ends Well	87.0	47.960784
173	King Phillip	King John	43.0	51.653846
174	King Richard II	Richard II	98.0	77.704225
175	Lady Anne	Richard III	51.0	38.185185
176	Lady Capulet	Romeo and Juliet	45.0	25.407407
177	Lady Macbeth	Macbeth	59.0	45.000000
178	Laertes	Hamlet	62.0	36.500000
179	Lafeu	All's Well That Ends Well	97.0	33.390244
180	Launce	Two Gentlemen of Verona	68.0	43.941176
181	Launcelot Gobbo	Merchant of Venice	44.0	38.750000
182	Lear	King Lear	188.0	55.671053
183	Lennox	Macbeth	21.0	31.666667
184	Leonato	Much Ado about Nothing	120.0	38.155556
185	Leontes	The Winter's Tale	125.0	59.208955
186	Lewis	King John	29.0	52.750000
187	Lewis the Dauphin	Henry V	31.0	33.291667
188	Lodovico	Othello	33.0	21.181818
189	Longaville	Love's Labour's Lost	40.0	18.687500
190	Lord Chamberlain	Henry VIII	38.0	34.518519
191	Lord Chief Justice	Henry IV, Part II	56.0	29.607143

	charname	plays	total_speeches	avg_wordcount
192	Lord Talbot/Earl of Shrewsbury	Henry VI, Part I	70.0	65.127660
193	Lorenzo	Merchant of Venice	47.0	33.575758
194	Lucentio	Taming of the Shrew	61.0	35.193548
195	Lucetta	Two Gentlemen of Verona	48.0	13.950000
196	Luciana	Comedy of Errors	43.0	29.555556
197	Lucio	Measure for Measure	111.0	32.822222
198	Lucius	Julius Caesar; Titus Andronicus	75.0	26.769697
199	Lysander	Midsummer Night's Dream	50.0	33.636364
200	Lysimachus	Pericles	40.0	24.291667
201	Macbeth	Macbeth	146.0	55.057971
202	Macduff	Macbeth	59.0	30.000000
203	Malcolm	Macbeth	40.0	43.866667
204	Malvolio	Twelfth Night	87.0	43.897436
205	Marcellus	Hamlet	37.0	22.000000
206	Marcus Andronicus	Titus Andronicus	63.0	59.945946
207	Margaret	Much Ado about Nothing	26.0	27.125000
208	Maria	Love's Labour's Lost; Twelfth Night	81.0	22.390394
209	Mariana	Measure for Measure	24.0	28.375000
210	Marina	Pericles	63.0	29.939394
211	Menas	Antony and Cleopatra	35.0	13.550000
212	Menenius Agrippa	Coriolanus	162.0	46.050000
213	Mercutio	Romeo and Juliet	62.0	46.243243
214	Messala	Julius Caesar	20.0	15.166667
215	Messenger	Antony and Cleopatra	42.0	18.875000
216	Miranda	Tempest	49.0	26.040000
217	Mistress Ford	Merry Wives of Windsor	85.0	23.000000
218	Mistress Page	Merry Wives of Windsor	101.0	38.288889
219	Montano	Othello	24.0	22.058824
220	Moth	Love's Labour's Lost	78.0	24.142857
221	Nerissa	Merchant of Venice	36.0	24.952381
222	Nestor	Troilus and Cressida	38.0	47.090909
223	Nurse	Romeo and Juliet	90.0	40.930233

	charname	plays	total_speeches	avg_wordcount
224	Oberon	Midsummer Night's Dream	29.0	58.629630
225	Old Shepherd	The Winter's Tale	42.0	31.933333
226	Oliver	As You Like It	37.0	42.708333
227	Olivia	Twelfth Night	118.0	36.636364
228	Ophelia	Hamlet	58.0	32.000000
229	Orlando	As You Like It	120.0	41.157895
230	Orsino	Twelfth Night	59.0	38.702703
231	Osric	Hamlet	25.0	19.125000
232	Oswald	King Lear	38.0	21.631579
233	Othello	Othello	274.0	58.549296
234	Page	Merry Wives of Windsor	75.0	20.566667
235	Painter	Timon of Athens	30.0	20.473684
236	Pandarus	Troilus and Cressida	153.0	35.686275
237	Paris	Romeo and Juliet; Troilus and Cressida	50.0	29.447619
238	Parolles	All's Well That Ends Well	141.0	40.739130
239	Patroclus	Troilus and Cressida	37.0	18.250000
240	Paulina	The Winter's Tale	59.0	46.600000
241	Pedant	Taming of the Shrew	20.0	20.937500
242	Pembroke	King John	20.0	30.600000
243	Perdita	The Winter's Tale	25.0	38.391304
244	Pericles	Pericles	121.0	57.328125
245	Petruchio	Taming of the Shrew	158.0	53.687500
246	Phebe	As You Like It	23.0	38.470588
247	Philip the Bastard	King John	89.0	68.061224
248	Pisanio	Cymbeline	58.0	39.312500
249	Poet	Timon of Athens	30.0	28.608696
250	Polixenes	The Winter's Tale	57.0	48.361111
251	Polonius	Hamlet	86.0	54.536585
252	Pompey	Antony and Cleopatra; Measure for Measure	101.0	28.595960
253	Portia	Merchant of Venice	117.0	59.864407
254	Posthumus Leonatus	Cymbeline	77.0	63.304348
255	Princess of France	Love's Labour's Lost	102.0	38.333333

	charname	plays	total_speeches	avg_wordcount
256	Prospero	Tempest	115.0	56.138462
257	Proteus	Two Gentlemen of Verona	147.0	50.913043
258	Provost	Measure for Measure	65.0	27.090909
259	Puck	Midsummer Night's Dream	33.0	48.777778
260	Queen	Richard II; Cymbeline	52.0	44.965455
261	Queen Elinor	King John	22.0	23.000000
262	Queen Katharine	Henry VIII	50.0	63.975000
263	Quince	Midsummer Night's Dream	40.0	32.160000
264	Regan	King Lear	73.0	28.714286
265	Reignier	Henry VI, Part I	24.0	20.947368
266	Roderigo	Othello	59.0	25.480000
267	Romeo	Romeo and Juliet	163.0	52.359375
268	Rosalind	As You Like It	201.0	53.794118
269	Rosaline	Love's Labour's Lost	75.0	31.000000
270	Rosencrantz	Hamlet	48.0	21.240000
271	Ross	Macbeth	39.0	31.600000
272	Salarino	Merchant of Venice	27.0	33.571429
273	Salisbury	King John	36.0	41.111111
274	Sampson	Romeo and Juliet	20.0	13.812500
275	Sebastian	Twelfth Night; Tempest	98.0	26.618590
276	Second Gentleman	Henry VIII	37.0	27.043478
277	Second Lord	Cymbeline; Timon of Athens; All's Well That En...	106.0	20.043011
278	Second Murderer	Richard III	30.0	20.941176
279	Shylock	Merchant of Venice	79.0	55.772727
280	Sicinius Velutus	Coriolanus	117.0	27.952381
281	Silence	Henry IV, Part II	22.0	14.000000
282	Silvia	Two Gentlemen of Verona	58.0	29.866667
283	Silvius	As You Like It	24.0	26.800000
284	Simonides	Pericles	42.0	32.413793
285	Simple	Merry Wives of Windsor	25.0	15.444444
286	Sir Andrew Aguecheek	Twelfth Night	88.0	19.964286
287	Sir Hugh Evans	Merry Wives of Windsor	87.0	28.047619

	charname	plays	total_speeches	avg_wordcount
288	Sir Thomas Lovell	Henry VIII	21.0	26.333333
289	Sir Toby Belch	Twelfth Night	152.0	32.914894
290	Sir William Catesby	Richard III	31.0	19.631579
291	Slender	Merry Wives of Windsor	56.0	27.166667
292	Solinus	Comedy of Errors	22.0	40.187500
293	Speed	Two Gentlemen of Verona	117.0	25.343750
294	Stephano	Tempest	60.0	27.270270
295	Tamora	Titus Andronicus	49.0	55.757576
296	Thaisa	Pericles	32.0	19.434783
297	Thersites	Troilus and Cressida	90.0	42.136364
298	Theseus	Midsummer Night's Dream	48.0	48.468750
299	Third Servingman	Coriolanus	20.0	23.923077
300	Thurio	Two Gentlemen of Verona	36.0	15.235294
301	Timon	Timon of Athens	210.0	67.402985
302	Titania	Midsummer Night's Dream	23.0	56.700000
303	Titus Andronicus	Titus Andronicus	117.0	68.910448
304	Titus Lartius	Coriolanus	23.0	17.736842
305	Touchstone	As You Like It	74.0	44.590909
306	Tranio	Taming of the Shrew	90.0	34.902439
307	Trinculo	Tempest	39.0	26.230769
308	Troilus	Troilus and Cressida	131.0	54.448276
309	Tullus Aufidius	Coriolanus	45.0	56.222222
310	Ulysses	Troilus and Cressida	80.0	70.600000
311	Valentine	Two Gentlemen of Verona	149.0	38.020833
312	Vincentio	Taming of the Shrew; Measure for Measure	217.0	37.033282
313	Viola	Twelfth Night	121.0	32.319149
314	Virgilia	Coriolanus	26.0	9.785714
315	Volumnia	Coriolanus	57.0	52.560976
316	Widow	All's Well That Ends Well	21.0	25.214286
317	Williams	Henry V	28.0	30.538462

Part g

Which Shakepearean works do not contain any scenes in a palace or a castle? Use a query that displays the title, genre type, and publication date of works that do not contain any scenes that take place in a palace or castle. [Hint: use your work in part e as a starting point. You will need a subquery, and you will need to think carefully about the type of join that you need to perform.][2 points]

In [ ]:
```
t_sql = """
SELECT
    w.title,
    w.genretype AS genre,
    w.date AS publication_date
FROM works w
WHERE w.workid NOT IN
(SELECT
    DISTINCT
    c.workid
FROM chapters c
WHERE
    LOWER(c.description) LIKE '%palace%'
    OR LOWER(c.description) LIKE '%castle%'
)
ORDER BY w.title;
"""
cursor.execute(t_sql)
rs_no_palace_castle = cursor.fetchall()
colnames = [x[0] for x in cursor.description]
df_no = pd.DataFrame(rs_no_palace_castle, columns=colnames)
```

In [ ]:
```
df_no.genre = df_no.genre.map(genre_map)
print(df_no.shape)
df_no
```

```
(16, 3)
```

Out[ ]:

	title	genre	publication_date
0	Coriolanus	tradegy	1607
1	Julius Caesar	tradegy	1599
2	Love's Labour's Lost	comedy	1594
3	Lover's Complaint	narrative poem	1609
4	Merchant of Venice	comedy	1596
5	Merry Wives of Windsor	comedy	1600
6	Much Ado about Nothing	comedy	1598
7	Passionate Pilgrim	narrative poem	1598
8	Phoenix and the Turtle	narrative poem	1601
9	Rape of Lucrece	narrative poem	1594
10	Romeo and Juliet	tradegy	1594
11	Sonnets	sonnet	1609
12	Taming of the Shrew	comedy	1593
13	Tempest	comedy	1611
14	Timon of Athens	tradegy	1607
15	Venus and Adonis	narrative poem	1593

Problem 2

The following file contains JSON formatted data of the official English-language translations of every constitution currently in effect in the world:

In [ ]:

```
url = 'http://httpbin.org/user-agent'
r = requests.get(url)
user_agent = json.loads(r.text)['user-agent']
user_agent
```

Out[ ]: 'python-requests/2.27.1'

In [ ]:

```
url = 'http://httpbin.org/user-agent'
r = requests.get(url)
user_agent = json.loads(r.text)['user-agent']
#user_agent

# getting the big block of html text from the webpage
url = 'https://github.com/jkropko/DS-6001/raw/master/localdata/const.json'

headers = {'User-Agent': user_agent}
print(headers)
r = requests.get(url, headers=headers, verify=False) # added verify = False to avoi
const_json = json.loads(r.text)
```

```
#const_json
#consts = const_json['data']
```

```
{'User-Agent': 'python-requests/2.27.1'}
```

```
c:\Users\dianam\Anaconda3\envs\ds6001\lib\site-packages\urllib3\connectionpool.py:
1043: InsecureRequestWarning: Unverified HTTPS request is being made to host 'gith
ub.com'. Adding certificate verification is strongly advised. See: https://urllib
3.readthedocs.io/en/1.26.x/advanced-usage.html#ssl-warnings
  warnings.warn(
c:\Users\dianam\Anaconda3\envs\ds6001\lib\site-packages\urllib3\connectionpool.py:
1043: InsecureRequestWarning: Unverified HTTPS request is being made to host 'raw.
githubusercontent.com'. Adding certificate verification is strongly advised. See:
https://urllib3.readthedocs.io/en/1.26.x/advanced-usage.html#ssl-warnings
  warnings.warn(
```

In [ ]:

```
#const = requests.get("https://github.com/jkropko/DS-6001/raw/master/localdata/cons
#const_json = json.loads(const.text)
pd.DataFrame.from_records(const_json)
```

Out[ ]:

	text	country	adopted	revised	reinstated	democracy
0	'Afghanistan 2004 Preamble \nIn the na...	Afghanistan	2004	NaN	NaN	0.372201
1	'Albania 1998 (rev. 2012) Preamble \nWe...	Albania	1998	2012.0	NaN	0.535111
2	'Andorra 1993 Preamble \nThe Andorran P...	Andorra	1993	NaN	NaN	NaN
3	'Angola 2010 Preamble \nWe, the people ...	Angola	2010	NaN	NaN	0.315043
4	'Antigua and Barbuda 1981 Preamble \nWH...	Antigua and Barbuda	1981	NaN	NaN	NaN
5	'Armenia 1995 (rev. 2005) Preamble \nTh...	Armenia	1995	2005.0	NaN	0.393278
6	'Australia 1901 (rev. 1985) Chapter I. The...	Australia	1901	1985.0	NaN	0.879540
7	'Austria 1920 (reinst. 1945, rev. 2013) Ch...	Austria	1920	2013.0	1945.0	0.820705
8	'Bahamas 1973 (rev. 2002) Preamble \nWh...	Bahamas	1973	2002.0	NaN	NaN
9	'Bahrain 2002 (rev. 2012) Preamble \nIn...	Bahrain	2002	2012.0	NaN	NaN
10	'Bangladesh 1972 (reinst. 1986, rev. 2014) ...	Bangladesh	1972	2014.0	1986.0	0.369978
11	'Barbados 1966 (rev. 2007) Preamble \nW...	Barbados	1966	2007.0	NaN	0.706611
12	'Belarus 1994 (rev. 2004) Preamble \nWe...	Belarus	1994	2004.0	NaN	0.289968
13	'Belgium 1831 (rev. 2014) TITLE I. On Fede...	Belgium	1831	2014.0	NaN	0.858361
14	'Belize 1981 (rev. 2011) Preamble \nWHE...	Belize	1981	2011.0	NaN	NaN
15	'Bhutan 2008 Preamble \nWE, the people ...	Bhutan	2008	NaN	NaN	0.537041
16	'Bosnia and Herzegovina 1995 (rev. 2009) P...	Bosnia and Herzegovina	1995	2009.0	NaN	0.338267
17	'Botswana 1966 (rev. 2005) CHAPTER I. The ...	Botswana	1966	2005.0	NaN	0.682246
18	'Brazil 1988 (rev. 2015) Preamble \nWe ...	Brazil	1988	2015.0	NaN	0.731678
19	'Brunei Darussalam 1959 (rev. 2006) Preamb...	Brunei Darussalam	1959	2006.0	NaN	NaN
20	'Bulgaria 1991 (rev. 2007) Preamble \nW...	Bulgaria	1991	2007.0	NaN	0.767290

	text	country	adopted	revised	reinstated	democracy
21	'Cambodia 1993 (rev. 2008) Preamble \nW...	Cambodia	1993	2008.0	NaN	0.313738
22	'Cameroon 1972 (rev. 2008) Preamble \nW...	Cameroon	1972	2008.0	NaN	0.363031
23	'Canada 1867 (rev. 2011) CONSTITUTION ACT ...	Canada	1867	2011.0	NaN	0.867248
24	'Central African Republic 2013 Preamble ...	Central African Republic	2013	NaN	NaN	0.504033
25	'Chile 1980 (rev. 2015) Chapter I. Bases o...	Chile	1980	2015.0	NaN	0.836397
26	'China 1982 (rev. 2004) Preamble \nChin...	China	1982	2004.0	NaN	0.096066
27	'Croatia 1991 (rev. 2010) I. Historical Fo...	Croatia	1991	2010.0	NaN	0.710922
28	'Cyprus 1960 (rev. 2013) Part I. GENERAL P...	Cyprus	1960	2013.0	NaN	0.810509
29	'Czech Republic 1993 (rev. 2013) Preamble ...	Czech Republic	1993	2013.0	NaN	0.859101
30	'Denmark 1953 Part I \n1. This Consti...	Denmark	1953	NaN	NaN	0.883552
31	'Dominica 1978 (rev. 1984) Preamble \nW...	Dominica	1978	1984.0	NaN	NaN
32	'Dominican Republic 2015 Preamble \nWe,...	Dominican Republic	2015	NaN	NaN	0.583654
33	'Ecuador 2008 (rev. 2015) Preamble \nWe...	Ecuador	2008	2015.0	NaN	0.631449
34	'Egypt 2014 Preamble \nIn the Name of G...	Egypt	2014	NaN	NaN	0.218600
35	'El Salvador 1983 (rev. 2014) TITLE I ...	El Salvador	1983	2014.0	NaN	0.661989
36	'Equatorial Guinea 1991 (rev. 2012) Preamb...	Equatorial Guinea	1991	2012.0	NaN	0.217861
37	'Eritrea 1997 Preamble \nWe, the people...	Eritrea	1997	NaN	NaN	0.075621
38	'Estonia 1992 (rev. 2011) Preamble \nWi...	Estonia	1992	2011.0	NaN	0.909233
39	'Ethiopia 1994 Preamble \nWe, the Natio...	Ethiopia	1994	NaN	NaN	0.254865
40	'Fiji 2013 Preamble \nWE, THE PEOPLE OF...	Fiji	2013	NaN	NaN	0.473559
41	'Finland 1999 (rev. 2011) Chapter 1. Funda...	Finland	1999	2011.0	NaN	0.856265

	text	country	adopted	revised	reinstated	democracy
42	'France 1958 (rev. 2008) Preamble \nThe...	France	1958	2008.0	NaN	0.918962
43	'Gambia 1996 (rev. 2004) Preamble \nIn ...	Gambia	1996	2004.0	NaN	0.348132
44	'Georgia 1995 (rev. 2013) Preamble \nWe...	Georgia	1995	2013.0	NaN	0.757486
45	'Germany 1949 (rev. 2012) Preamble \nCo...	Germany	1949	2012.0	NaN	0.856599
46	'Ghana 1992 (rev. 1996) Preamble \nIN T...	Ghana	1992	1996.0	NaN	0.670849
47	'Greece 1975 (rev. 2008) Preamble \nIn ...	Greece	1975	2008.0	NaN	0.829714
48	'Grenada 1973 (reinst. 1991, rev. 1992) Pr...	Grenada	1973	1992.0	1991.0	NaN
49	'Guyana 1980 (rev. 2009) Preamble \nWE,...	Guyana	1980	2009.0	NaN	0.652792
50	'Hungary 2011 (rev. 2013) Preamble \nGo...	Hungary	2011	2013.0	NaN	0.697058
51	'Iceland 1944 (rev. 2013) I. Article 1...	Iceland	1944	2013.0	NaN	0.836434
52	'India 1949 (rev. 2015) Preamble \nWE, ...	India	1949	2015.0	NaN	0.632527
53	'Indonesia 1945 (reinst. 1959, rev. 2002) ...	Indonesia	1945	2002.0	1959.0	0.667082
54	'Iran (Islamic Republic of) 1979 (rev. 1989) ...	Iran (Islamic Republic of)	1979	1989.0	NaN	0.228920
55	'Iraq 2005 Preamble \nIn the name of Go...	Iraq	2005	NaN	NaN	0.455402
56	'Ireland 1937 (rev. 2015) Preamble \nIn...	Ireland	1937	2015.0	NaN	0.840020
57	'Israel 1958 (rev. 2013) Basic Law. The Kn...	Israel	1958	2013.0	NaN	0.729912
58	'Italy 1947 (rev. 2012) FUNDAMENTAL PRINCI...	Italy	1947	2012.0	NaN	0.826405
59	'Côte d'Ivoire 2000 (rev. 2004) Preamble ...	Côte d'Ivoire	2000	2004.0	NaN	0.578146
60	'Jamaica 1962 (rev. 2011) CHAPTER I. PRELI...	Jamaica	1962	2011.0	NaN	0.818984
61	'Japan 1946 Preamble \nWe, the Japanese...	Japan	1946	NaN	NaN	0.777828
62	'Jordan 1952 (rev. 2014) Preamble \nWe,...	Jordan	1952	2014.0	NaN	0.270614

	text	country	adopted	revised	reinstated	democracy
63	'Kazakhstan 1995 (rev. 2011) Preamble \...	Kazakhstan	1995	2011.0	NaN	0.262596
64	'Kenya 2010 Preamble \nWe, the people o...	Kenya	2010	NaN	NaN	0.531911
65	'Kiribati 1979 (rev. 1995) Preamble \nW...	Kiribati	1979	1995.0	NaN	NaN
66	'Kosovo 2008 Preamble \nWe, the people ...	Kosovo	2008	NaN	NaN	0.510447
67	'Kuwait 1962 (reinst. 1992) Preamble \n...	Kuwait	1962	NaN	1992.0	0.311841
68	'Lao People's Democratic Republic 1991 (rev. 2...	Lao People's Democratic Republic	1991	2003.0	NaN	0.094434
69	'Latvia 1922 (reinst. 1991, rev. 2014) Pre...	Latvia	1922	2014.0	1991.0	0.837859
70	'Lesotho 1993 (rev. 1998) CHAPTER I. THE K...	Lesotho	1993	1998.0	NaN	0.672989
71	'Liberia 1986 Preamble \nWe the People ...	Liberia	1986	NaN	NaN	0.629972
72	'Libya 2011 (rev. 2012) Preamble \nIn T...	Libya	2011	2012.0	NaN	0.294716
73	'Liechtenstein 1921 (rev. 2003) Preamble ...	Liechtenstein	1921	2003.0	NaN	NaN
74	'Lithuania 1992 (rev. 2006) Preamble \n...	Lithuania	1992	2006.0	NaN	0.830487
75	'Macedonia (The former Yugoslav Republic of) 1...	Macedonia (The former Yugoslav Republic of)	1991	2011.0	NaN	0.510983
76	'Malawi 1994 (rev. 1999) Preamble \nTHE...	Malawi	1994	1999.0	NaN	0.510355
77	'Malaysia 1957 (rev. 2007) PART I. THE STA...	Malaysia	1957	2007.0	NaN	0.345091
78	'Maldives 2008 CHAPTER I. STATE, SOVEREIGN...	Maldives	2008	NaN	NaN	0.386754
79	'Malta 1964 (rev. 2014) CHAPTER I. The Rep...	Malta	1964	2014.0	NaN	NaN
80	'Marshall Islands 1979 (rev. 1995) Preambl...	Marshall Islands	1979	1995.0	NaN	NaN
81	'Mauritius 1968 (rev. 2011) CHAPTER I. THE...	Mauritius	1968	2011.0	NaN	0.790856
82	'Mexico 1917 (rev. 2015) TITLE ONE CHA...	Mexico	1917	2015.0	NaN	0.672567
83	'Micronesia (Federated States of) 1978 (rev. 1...	Micronesia (Federated States of)	1978	1990.0	NaN	NaN

	text	country	adopted	revised	reinstated	democracy
84	'Moldova (Republic of) 1994 (rev. 2006) Pr...	Moldova (Republic of)	1994	2006.0	NaN	0.571357
85	'Monaco 1962 (rev. 2002) Chapter I. The Pr...	Monaco	1962	2002.0	NaN	NaN
86	'Mongolia 1992 (rev. 2001) Preamble \nW...	Mongolia	1992	2001.0	NaN	0.675772
87	'Montenegro 2007 Preamble \nStemming fr...	Montenegro	2007	NaN	NaN	0.455338
88	'Myanmar 2008 Preamble \nMyanmar is a N...	Myanmar	2008	NaN	NaN	0.405772
89	'Namibia 1990 (rev. 2010) Preamble \nWh...	Namibia	1990	2010.0	NaN	0.745421
90	'Nauru 1968 Preamble \nWHEREAS we the p...	Nauru	1968	NaN	NaN	NaN
91	'Netherlands 1815 (rev. 2008) CHAPTER 1. F...	Netherlands	1815	2008.0	NaN	0.859255
92	'New Zealand 1852 (rev. 2014) Legislature ...	New Zealand	1852	2014.0	NaN	0.863429
93	'Nigeria 1999 Preamble \nWe the people ...	Nigeria	1999	NaN	NaN	0.628442
94	'Korea (Democratic People's Republic of) 1972 ...	Korea (Democratic People's Republic of)	1972	1998.0	NaN	0.090438
95	'Norway 1814 (rev. 2015) A. Form of govern...	Norway	1814	2015.0	NaN	0.901217
96	'Oman 1996 (rev. 2011) CHAPTER ONE. The St...	Oman	1996	2011.0	NaN	0.191211
97	'Pakistan 1973 (reinst. 2002, rev. 2015) P...	Pakistan	1973	2015.0	2002.0	0.430273
98	'Palau 1981 (rev. 1992) Preamble \nIn e...	Palau	1981	1992.0	NaN	NaN
99	'Panama 1972 (rev. 2004) Preamble \nWit...	Panama	1972	2004.0	NaN	0.739508
100	'Papua New Guinea 1975 (rev. 2014) Preambl...	Papua New Guinea	1975	2014.0	NaN	0.488884
101	'Peru 1993 (rev. 2009) Preamble \nThe D...	Peru	1993	2009.0	NaN	0.730816
102	'Philippines 1987 Preamble \nWe, the so...	Philippines	1987	NaN	NaN	0.567393
103	'Poland 1997 (rev. 2009) Preamble \nHav...	Poland	1997	2009.0	NaN	0.682208

	text	country	adopted	revised	reinstated	democracy
104	'Portugal 1976 (rev. 2005) Preamble \nO...	Portugal	1976	2005.0	NaN	0.873393
105	'Romania 1991 (rev. 2003) TITLE I. GENERAL...	Romania	1991	2003.0	NaN	0.767288
106	'Russian Federation 1993 (rev. 2014) Pream...	Russian Federation	1993	2014.0	NaN	0.275516
107	'Rwanda 2003 (rev. 2010) Preamble \nWe,...	Rwanda	2003	2010.0	NaN	0.274476
108	'Saint Kitts and Nevis 1983 Preamble \n...	Saint Kitts and Nevis	1983	NaN	NaN	NaN
109	'Saint Lucia 1978 Preamble \nWHEREAS th...	Saint Lucia	1978	NaN	NaN	NaN
110	'Saint Vincent and the Grenadines 1979 Pre...	Saint Vincent and the Grenadines	1979	NaN	NaN	NaN
111	'Samoa 1962 (rev. 2013) Preamble \nIN T...	Samoa	1962	2013.0	NaN	NaN
112	'Saudi Arabia 1992 (rev. 2013) Basic Law ...	Saudi Arabia	1992	2013.0	NaN	0.024049
113	'Serbia 2006 Preamble \nConsidering the...	Serbia	2006	NaN	NaN	0.474443
114	'Seychelles 1993 (rev. 2011) Preamble \...	Seychelles	1993	2011.0	NaN	0.589139
115	'Sierra Leone 1991 (reinst. 1996, rev. 2008) ...	Sierra Leone	1991	2008.0	1996.0	0.564657
116	'Singapore 1963 (rev. 2010) PART I. PRELIM...	Singapore	1963	2010.0	NaN	0.446464
117	'Slovakia 1992 (rev. 2014) Preamble \nW...	Slovakia	1992	2014.0	NaN	0.797398
118	'Slovenia 1991 (rev. 2013) Preamble \nP...	Slovenia	1991	2013.0	NaN	0.861380
119	'Somalia 2012 CHAPTER 1. DECLARATION OF TH...	Somalia	2012	NaN	NaN	0.177772
120	'South Africa 1996 (rev. 2012) Preamble ...	South Africa	1996	2012.0	NaN	0.727070
121	'Korea (Republic of) 1948 (rev. 1987) Prea...	Korea (Republic of)	1948	1987.0	NaN	0.757692
122	'South Sudan 2011 (rev. 2013) Preamble ...	South Sudan	2011	2013.0	NaN	0.183267
123	'Spain 1978 (rev. 2011) Preamble \nThe ...	Spain	1978	2011.0	NaN	0.834466
124	'Sri Lanka 1978 (rev. 2015) Preamble \n...	Sri Lanka	1978	2015.0	NaN	0.647035

	text	country	adopted	revised	reinstated	democracy
125	'Sudan 2005 Preamble \nWe the people of...	Sudan	2005	NaN	NaN	0.311799
126	'Suriname 1987 (rev. 1992) Preamble \nW...	Suriname	1987	1992.0	NaN	0.799657
127	'Swaziland 2005 Preamble \nWhereas We t...	Swaziland	2005	NaN	NaN	0.136008
128	'Sweden 1974 (rev. 2012) The Instrument of...	Sweden	1974	2012.0	NaN	0.902575
129	'Syrian Arab Republic 2012 Preamble \nA...	Syrian Arab Republic	2012	NaN	NaN	0.148212
130	'Tonga 1875 (rev. 1988) Preamble \nGran...	Tonga	1875	1988.0	NaN	NaN
131	'Trinidad and Tobago 1976 (rev. 2007) Prea...	Trinidad and Tobago	1976	2007.0	NaN	0.730927
132	'Tunisia 2014 Preamble \nIn the Name of...	Tunisia	2014	NaN	NaN	0.748064
133	'Turkey 1982 (rev. 2011) Preamble \nAff...	Turkey	1982	2011.0	NaN	0.341745
134	'Turkmenistan 2008 Preamble \nWe, the p...	Turkmenistan	2008	NaN	NaN	0.154887
135	'Tuvalu 1986 Preamble \nWHEREAS in adop...	Tuvalu	1986	NaN	NaN	NaN
136	'Uganda 1995 (rev. 2005) Preamble \nWE ...	Uganda	1995	2005.0	NaN	0.338308
137	'Ukraine 1996 (rev. 2014) Preamble \nTh...	Ukraine	1996	2014.0	NaN	0.361911
138	'United Arab Emirates 1971 (rev. 2009) Pre...	United Arab Emirates	1971	2009.0	NaN	NaN
139	'United States of America 1789 (rev. 1992) ...	United States of America	1789	1992.0	NaN	0.849155
140	'Uzbekistan 1992 (rev. 2011) Preamble \...	Uzbekistan	1992	2011.0	NaN	0.195932
141	'Viet Nam 1992 (rev. 2013) Preamble \nI...	Viet Nam	1992	2013.0	NaN	0.251461
142	'Yemen 1991 (rev. 2001) PART ONE. THE FOUN...	Yemen	1991	2001.0	NaN	0.125708
143	'Zambia 1991 (rev. 2009) Preamble \nWE,...	Zambia	1991	2009.0	NaN	0.405497
144	'Zimbabwe 2013 Preamble \nWe the people...	Zimbabwe	2013	NaN	NaN	0.315359

The text of the constitutions are available from the Wolfram Data Repository. I also included scores that represent the level of democractic quality in each country as of 2016. These scores are compiled by the Varieties of Democracy (V-Dem) project. Higher scores indicate greater levels of democratic openness and competition.

Part a

Connect to your local MongoDB server and create a new collection for the constitution data. Use `.delete_many({})` to remove any existing data from this collection, and insert the data in `const_json` into this collection. [2 points]

In [ ]:
```
import pymongo
# connect to my local MongoDB and create a new database called "history" My localho
mongodb_client = pymongo.MongoClient("mongodb://localhost:27017/")
```

In [ ]:
```
# connect to my local MongoDB and create a new database called "politics"
politicsdb = mongodb_client["globalpolitics"]
constitutionscollection = politicsdb["constitutions"]
```

In [ ]:
```
constitutionscollection.delete_many({}) # remove any existing data in the collectio
```

Out[ ]:
```
<pymongo.results.DeleteResult at 0x1580bf85e40>
```

In [ ]:
```
allConsts = constitutionscollection.insert_many(const_json)
```

Part b

Use MongoDB queries and the `dumps()` and `loads()` functions from the `bson` package to produce dataframes with the following restrictions:

- The country, adoption year, and democracy features (and not `_id`, text, revised, or reinstated) for countries with constitutions that were written after 1990
- The country, adoption year, and democracy features (and not `_id`, text, revised, or reinstated) for countries with constitutions that were written after 1990 AND have a democracy score of less than 0.5
- The country, adoption year, and democracy features (and not `_id`, text, revised, or reinstated) for countries with constitutions that were written after 1990 OR have a democracy score of less than 0.5

[1 point]

In [ ]:
```
# as recommended by the Professor, I am creating the mongo_read_query function
"""
This function takes a collection and a query with features and returns a dataframe
It returns a dataframe of the results.
"""
def mongo_read_query(col, q, f):
    qtext = dumps(col.find(q,f))
```

```
    qrec = loads(qtext)
    qdf = pd.DataFrame.from_records(qrec)
    return qdf
```

In [ ]:
```
df_post1990 = mongo_read_query(constitutionscollection, {'adopted' : {'$gt':1990}},
print(df_post1990.shape)
df_post1990
```

```
(71, 3)
```

Out[ ]:

	country	adopted	democracy
0	Afghanistan	2004	0.372201
1	Albania	1998	0.535111
2	Andorra	1993	NaN
3	Angola	2010	0.315043
4	Armenia	1995	0.393278
5	Bahrain	2002	NaN
6	Belarus	1994	0.289968
7	Bhutan	2008	0.537041
8	Bosnia and Herzegovina	1995	0.338267
9	Bulgaria	1991	0.767290
10	Cambodia	1993	0.313738
11	Central African Republic	2013	0.504033
12	Croatia	1991	0.710922
13	Czech Republic	1993	0.859101
14	Dominican Republic	2015	0.583654
15	Ecuador	2008	0.631449
16	Egypt	2014	0.218600
17	Equatorial Guinea	1991	0.217861
18	Eritrea	1997	0.075621
19	Estonia	1992	0.909233
20	Ethiopia	1994	0.254865
21	Fiji	2013	0.473559
22	Finland	1999	0.856265
23	Gambia	1996	0.348132
24	Georgia	1995	0.757486
25	Ghana	1992	0.670849
26	Hungary	2011	0.697058
27	Iraq	2005	0.455402
28	Côte d'Ivoire	2000	0.578146
29	Kazakhstan	1995	0.262596
30	Kenya	2010	0.531911
31	Kosovo	2008	0.510447
32	Lao People's Democratic Republic	1991	0.094434

	country	adopted	democracy
33	Lesotho	1993	0.672989
34	Libya	2011	0.294716
35	Lithuania	1992	0.830487
36	Macedonia (The former Yugoslav Republic of)	1991	0.510983
37	Malawi	1994	0.510355
38	Maldives	2008	0.386754
39	Moldova (Republic of)	1994	0.571357
40	Mongolia	1992	0.675772
41	Montenegro	2007	0.455338
42	Myanmar	2008	0.405772
43	Nigeria	1999	0.628442
44	Oman	1996	0.191211
45	Peru	1993	0.730816
46	Poland	1997	0.682208
47	Romania	1991	0.767288
48	Russian Federation	1993	0.275516
49	Rwanda	2003	0.274476
50	Saudi Arabia	1992	0.024049
51	Serbia	2006	0.474443
52	Seychelles	1993	0.589139
53	Sierra Leone	1991	0.564657
54	Slovakia	1992	0.797398
55	Slovenia	1991	0.861380
56	Somalia	2012	0.177772
57	South Africa	1996	0.727070
58	South Sudan	2011	0.183267
59	Sudan	2005	0.311799
60	Swaziland	2005	0.136008
61	Syrian Arab Republic	2012	0.148212
62	Tunisia	2014	0.748064
63	Turkmenistan	2008	0.154887
64	Uganda	1995	0.338308
65	Ukraine	1996	0.361911

	country	adopted	democracy
66	Uzbekistan	1992	0.195932
67	Viet Nam	1992	0.251461
68	Yemen	1991	0.125708
69	Zambia	1991	0.405497
70	Zimbabwe	2013	0.315359

In [ ]:
```
#The country, adoption year, and democracy features (and not `_id`, text, revised,
df_post1990_dlt05 = mongo_read_query(constitutionscollection, {'adopted':{'$gt':199
print(df_post1990_dlt05.shape)
df_post1990_dlt05
```

```
(37, 3)
```

Out[ ]:

	country	adopted	democracy
0	Afghanistan	2004	0.372201
1	Angola	2010	0.315043
2	Armenia	1995	0.393278
3	Belarus	1994	0.289968
4	Bosnia and Herzegovina	1995	0.338267
5	Cambodia	1993	0.313738
6	Egypt	2014	0.218600
7	Equatorial Guinea	1991	0.217861
8	Eritrea	1997	0.075621
9	Ethiopia	1994	0.254865
10	Fiji	2013	0.473559
11	Gambia	1996	0.348132
12	Iraq	2005	0.455402
13	Kazakhstan	1995	0.262596
14	Lao People's Democratic Republic	1991	0.094434
15	Libya	2011	0.294716
16	Maldives	2008	0.386754
17	Montenegro	2007	0.455338
18	Myanmar	2008	0.405772
19	Oman	1996	0.191211
20	Russian Federation	1993	0.275516
21	Rwanda	2003	0.274476
22	Saudi Arabia	1992	0.024049
23	Serbia	2006	0.474443
24	Somalia	2012	0.177772
25	South Sudan	2011	0.183267
26	Sudan	2005	0.311799
27	Swaziland	2005	0.136008
28	Syrian Arab Republic	2012	0.148212
29	Turkmenistan	2008	0.154887
30	Uganda	1995	0.338308
31	Ukraine	1996	0.361911
32	Uzbekistan	1992	0.195932

	country	adopted	democracy
33	Viet Nam	1992	0.251461
34	Yemen	1991	0.125708
35	Zambia	1991	0.405497
36	Zimbabwe	2013	0.315359

In [ ]:

```
#* The country, adoption year, and democracy features (and not `_id`, text, revised
df_post1990_dlt05or = mongo_read_query(constitutionscollection, {'$or': [{'adopted'
#df_post1990_dlt05or = mongo_read_query(constitutionscollection, {'$or': [{'adopted
print(df_post1990_dlt05or.shape)
df_post1990_dlt05or
```

```
(83, 3)
```

Out[ ]:

	country	adopted	democracy
0	Afghanistan	2004	0.372201
1	Albania	1998	0.535111
2	Andorra	1993	NaN
3	Angola	2010	0.315043
4	Armenia	1995	0.393278
5	Bahrain	2002	NaN
6	Bangladesh	1972	0.369978
7	Belarus	1994	0.289968
8	Bhutan	2008	0.537041
9	Bosnia and Herzegovina	1995	0.338267
10	Bulgaria	1991	0.767290
11	Cambodia	1993	0.313738
12	Cameroon	1972	0.363031
13	Central African Republic	2013	0.504033
14	China	1982	0.096066
15	Croatia	1991	0.710922
16	Czech Republic	1993	0.859101
17	Dominican Republic	2015	0.583654
18	Ecuador	2008	0.631449
19	Egypt	2014	0.218600
20	Equatorial Guinea	1991	0.217861
21	Eritrea	1997	0.075621
22	Estonia	1992	0.909233
23	Ethiopia	1994	0.254865
24	Fiji	2013	0.473559
25	Finland	1999	0.856265
26	Gambia	1996	0.348132
27	Georgia	1995	0.757486
28	Ghana	1992	0.670849
29	Hungary	2011	0.697058
30	Iran (Islamic Republic of)	1979	0.228920
31	Iraq	2005	0.455402
32	Côte d'Ivoire	2000	0.578146

	country	adopted	democracy
33	Jordan	1952	0.270614
34	Kazakhstan	1995	0.262596
35	Kenya	2010	0.531911
36	Kosovo	2008	0.510447
37	Kuwait	1962	0.311841
38	Lao People's Democratic Republic	1991	0.094434
39	Lesotho	1993	0.672989
40	Libya	2011	0.294716
41	Lithuania	1992	0.830487
42	Macedonia (The former Yugoslav Republic of)	1991	0.510983
43	Malawi	1994	0.510355
44	Malaysia	1957	0.345091
45	Maldives	2008	0.386754
46	Moldova (Republic of)	1994	0.571357
47	Mongolia	1992	0.675772
48	Montenegro	2007	0.455338
49	Myanmar	2008	0.405772
50	Nigeria	1999	0.628442
51	Korea (Democratic People's Republic of)	1972	0.090438
52	Oman	1996	0.191211
53	Pakistan	1973	0.430273
54	Papua New Guinea	1975	0.488884
55	Peru	1993	0.730816
56	Poland	1997	0.682208
57	Romania	1991	0.767288
58	Russian Federation	1993	0.275516
59	Rwanda	2003	0.274476
60	Saudi Arabia	1992	0.024049
61	Serbia	2006	0.474443
62	Seychelles	1993	0.589139
63	Sierra Leone	1991	0.564657
64	Singapore	1963	0.446464
65	Slovakia	1992	0.797398

	country	adopted	democracy
66	Slovenia	1991	0.861380
67	Somalia	2012	0.177772
68	South Africa	1996	0.727070
69	South Sudan	2011	0.183267
70	Sudan	2005	0.311799
71	Swaziland	2005	0.136008
72	Syrian Arab Republic	2012	0.148212
73	Tunisia	2014	0.748064
74	Turkey	1982	0.341745
75	Turkmenistan	2008	0.154887
76	Uganda	1995	0.338308
77	Ukraine	1996	0.361911
78	Uzbekistan	1992	0.195932
79	Viet Nam	1992	0.251461
80	Yemen	1991	0.125708
81	Zambia	1991	0.405497
82	Zimbabwe	2013	0.315359

Part c

According to the Varieties of Democracy project, Hungary has become less democratic over the last few years, and can no longer be considered a democracy. Update the record for Hungary to set the democracy score at 0.4. Then query the database to extract the record for Hungary and display the data in a dataframe. [1 point]

In [ ]:

```
# here is the MongoDB update command
constitutionscollection.update_one({'country': 'Hungary'},{'$set' : {'democracy': 0
# here I do the MongoDB select and show the dataframe with Hungary
mongo_read_query(constitutionscollection, {'country': 'Hungary'}, {'country':1, 'ad
```

Out[ ]:

	country	adopted	democracy
0	Hungary	2011	0.4

Part d

Set the `text` field in the database as a text index. Then query the database to find all constitutions that contain the exact phrase "freedom of speech". Display the country name, adoption year, and democracy scores in a dataframe for the constitutions that match this query. [2 points]

In [ ]:
```
# first, set the text field to an index - I think we are setting the 'keytosearch'
constitutionscollection.create_index([('text', 'text')])
```

Out[ ]: 'text_text'

In [ ]:
```
# now we can do the search using the $text operator
#{'$text': {'$search': 'searchterms', '$caseSensitive': False}}
search_text = '\"freedom of speech\"' # added the double-quote with an \ escape cha

df = mongo_read_query(constitutionscollection,
    {'$text': {'$search':search_text, '$caseSensitive': True}},
    {'country':1, 'adopted':1, 'democracy': 1, 'text': 1, '_id':0})
print(df.shape)
df
```

(46, 4)

Out[ ]:

	text	country	adopted	democracy
0	'Slovenia 1991 (rev. 2013) Preamble \nP...	Slovenia	1991	0.861380
1	'Poland 1997 (rev. 2009) Preamble \nHav...	Poland	1997	0.682208
2	'Eritrea 1997 Preamble \nWe, the people...	Eritrea	1997	0.075621
3	'Croatia 1991 (rev. 2010) I. Historical Fo...	Croatia	1991	0.710922
4	'Macedonia (The former Yugoslav Republic of) 1...	Macedonia (The former Yugoslav Republic of)	1991	0.510983
5	'Kazakhstan 1995 (rev. 2011) Preamble \...	Kazakhstan	1995	0.262596
6	'Zimbabwe 2013 Preamble \nWe the people...	Zimbabwe	2013	0.315359
7	'Kenya 2010 Preamble \nWe, the people o...	Kenya	2010	0.531911
8	'Fiji 2013 Preamble \nWE, THE PEOPLE OF...	Fiji	2013	0.473559
9	'Georgia 1995 (rev. 2013) Preamble \nWe...	Georgia	1995	0.757486
10	'Namibia 1990 (rev. 2010) Preamble \nWh...	Namibia	1990	0.745421
11	'Spain 1978 (rev. 2011) Preamble \nThe ...	Spain	1978	0.834466
12	'Korea (Republic of) 1948 (rev. 1987) Prea...	Korea (Republic of)	1948	0.757692
13	'Antigua and Barbuda 1981 Preamble \nWH...	Antigua and Barbuda	1981	NaN
14	'Swaziland 2005 Preamble \nWhereas We t...	Swaziland	2005	0.136008
15	'Ghana 1992 (rev. 1996) Preamble \nIN T...	Ghana	1992	0.670849
16	'Gambia 1996 (rev. 2004) Preamble \nIn ...	Gambia	1996	0.348132
17	'Peru 1993 (rev. 2009) Preamble \nThe D...	Peru	1993	0.730816
18	'Papua New Guinea 1975 (rev. 2014) Preambl...	Papua New Guinea	1975	0.488884
19	'Sierra Leone 1991 (reinst. 1996, rev. 2008) ...	Sierra Leone	1991	0.564657
20	'Hungary 2011 (rev. 2013) Preamble \nGo...	Hungary	2011	0.400000

	text	country	adopted	democracy
21	'Seychelles 1993 (rev. 2011) Preamble \...	Seychelles	1993	0.589139
22	'China 1982 (rev. 2004) Preamble \nChin...	China	1982	0.096066
23	'Uganda 1995 (rev. 2005) Preamble \nWE ...	Uganda	1995	0.338308
24	'Bangladesh 1972 (reinst. 1986, rev. 2014) ...	Bangladesh	1972	0.369978
25	'Somalia 2012 CHAPTER 1. DECLARATION OF TH...	Somalia	2012	0.177772
26	'South Africa 1996 (rev. 2012) Preamble ...	South Africa	1996	0.727070
27	'Trinidad and Tobago 1976 (rev. 2007) Prea...	Trinidad and Tobago	1976	0.730927
28	'Jordan 1952 (rev. 2014) Preamble \nWe,...	Jordan	1952	0.270614
29	'Samoa 1962 (rev. 2013) Preamble \nIN T...	Samoa	1962	NaN
30	'Sri Lanka 1978 (rev. 2015) Preamble \n...	Sri Lanka	1978	0.647035
31	'Liberia 1986 Preamble \nWe the People ...	Liberia	1986	0.629972
32	'Mexico 1917 (rev. 2015) TITLE ONE CHA...	Mexico	1917	0.672567
33	'Lao People's Democratic Republic 1991 (rev. 2...	Lao People's Democratic Republic	1991	0.094434
34	'Pakistan 1973 (reinst. 2002, rev. 2015) P...	Pakistan	1973	0.430273
35	'India 1949 (rev. 2015) Preamble \nWE, ...	India	1949	0.632527
36	'Myanmar 2008 Preamble \nMyanmar is a N...	Myanmar	2008	0.405772
37	'Bhutan 2008 Preamble \nWE, the people ...	Bhutan	2008	0.537041
38	'Korea (Democratic People's Republic of) 1972 ...	Korea (Democratic People's Republic of)	1972	0.090438
39	'Philippines 1987 Preamble \nWe, the so...	Philippines	1987	0.567393
40	'Tonga 1875 (rev. 1988) Preamble \nGran...	Tonga	1875	NaN
41	'Marshall Islands 1979 (rev. 1995) Preambl...	Marshall Islands	1979	NaN

	text	country	adopted	democracy
42	'Cyprus 1960 (rev. 2013) Part I. GENERAL P...	Cyprus	1960	0.810509
43	'Singapore 1963 (rev. 2010) PART I. PRELIM...	Singapore	1963	0.446464
44	'Malaysia 1957 (rev. 2007) PART I. THE STA...	Malaysia	1957	0.345091
45	'United States of America 1789 (rev. 1992) ...	United States of America	1789	0.849155

Part e

Use a query to search for the terms "freedom", "liberty", "legal", "justice", and "rights". Generate a text score for all of the countries, and display the data for the countries with the top 10 relevancy scores in a dataframe. [2 points]

In [ ]:
```
# now we can do the search using the $text operator
#{'$text': {'$search': 'searchterms', '$caseSensitive': False}}
search_text = 'freedom liberty legal justice'

#df = mongo_read_query(constitutionscollection,
#    {'$text': {'$search':search_text, '$caseSensitive': True}},
#    {'country':1, 'adopted':1, 'democracy': 1, 'text': 1, '_id':0})
#print(df.shape)
#df

cursor = constitutionscollection.find(
            {'$text': {'$search': search_text}},
            {'score': {'$meta': 'textScore'}})
# need to sort the text scores
cursor.sort([('score', {'$meta': 'textScore'})])

# load the results into a dataframe
qtext = dumps(cursor)
qrec = loads(qtext)
df = pd.DataFrame.from_records(qrec)
```

In [ ]:
```
df.head(10)
```

Out[ ]:

	_id	text	country	adopted	revised	reinstated	democracy
0	63f56a3f6cc8dd00d6f87e94	'Finland 1999 (rev. 2011) Chapter 1. Funda...	Finland	1999	2011.0	NaN	0.856265
1	63f56a3f6cc8dd00d6f87edc	'Serbia 2006 Preamble \nConsidering the...	Serbia	2006	NaN	NaN	0.474443
2	63f56a3f6cc8dd00d6f87eeb	'Sweden 1974 (rev. 2012) The Instrument of...	Sweden	1974	2012.0	NaN	0.902575
3	63f56a3f6cc8dd00d6f87e70	'Armenia 1995 (rev. 2005) Preamble \nTh...	Armenia	1995	2005.0	NaN	0.393278
4	63f56a3f6cc8dd00d6f87e6c	'Albania 1998 (rev. 2012) Preamble \nWe...	Albania	1998	2012.0	NaN	0.535111
5	63f56a3f6cc8dd00d6f87e8b	'Dominican Republic 2015 Preamble \nWe,...	Dominican Republic	2015	NaN	NaN	0.583654
6	63f56a3f6cc8dd00d6f87e8e	'El Salvador 1983 (rev. 2014) TITLE I ...	El Salvador	1983	2014.0	NaN	0.661989
7	63f56a3f6cc8dd00d6f87ef0	'Turkey 1982 (rev. 2011) Preamble \nAff...	Turkey	1982	2011.0	NaN	0.341745
8	63f56a3f6cc8dd00d6f87e91	'Estonia 1992 (rev. 2011) Preamble \nWi...	Estonia	1992	2011.0	NaN	0.909233
9	63f56a3f6cc8dd00d6f87e97	'Georgia 1995 (rev. 2013) Preamble \nWe...	Georgia	1995	2013.0	NaN	0.757486

Question 3

Close the connections to the PostgreSQL and MongoDB databases. [1 point]

MAKE SURE TO UNCOMMENT THESE FOR FINAL SUBMIT

In [ ]:
```
# postgres connection
dbserver.close()

# close the mondgodb client
mongodb_client.close()
```

In [ ]: