BST 263 HOMEWORK 5

Jenny Wang (71401898)

April 14, 2020

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        from sklearn.tree import DecisionTreeRegressor, plot_tree
        from sklearn.ensemble import RandomForestRegressor
        from sklearn.feature_extraction.text import TfidfTransformer
        from sklearn.cluster import KMeans
        from sklearn.preprocessing import normalize
        from hw5_helper import *
        import warnings
        warnings.filterwarnings("ignore")
```

```
Bad key "text.kerning_factor" on line 4 in
/opt/anaconda3/lib/python3.7/site-packages/matplotlib/mpl-data/styleli
b/_classic_test_patch.mplstyle.
You probably need to get an updated matplotlibrc file from
https://github.com/matplotlib/matplotlib/blob/v3.1.3/matplotlibrc.templ
ate
or from the matplotlib source distribution
```

Question 1a)

```
In [2]: # Load data
        data = pd.read_csv("baseball.csv")
```
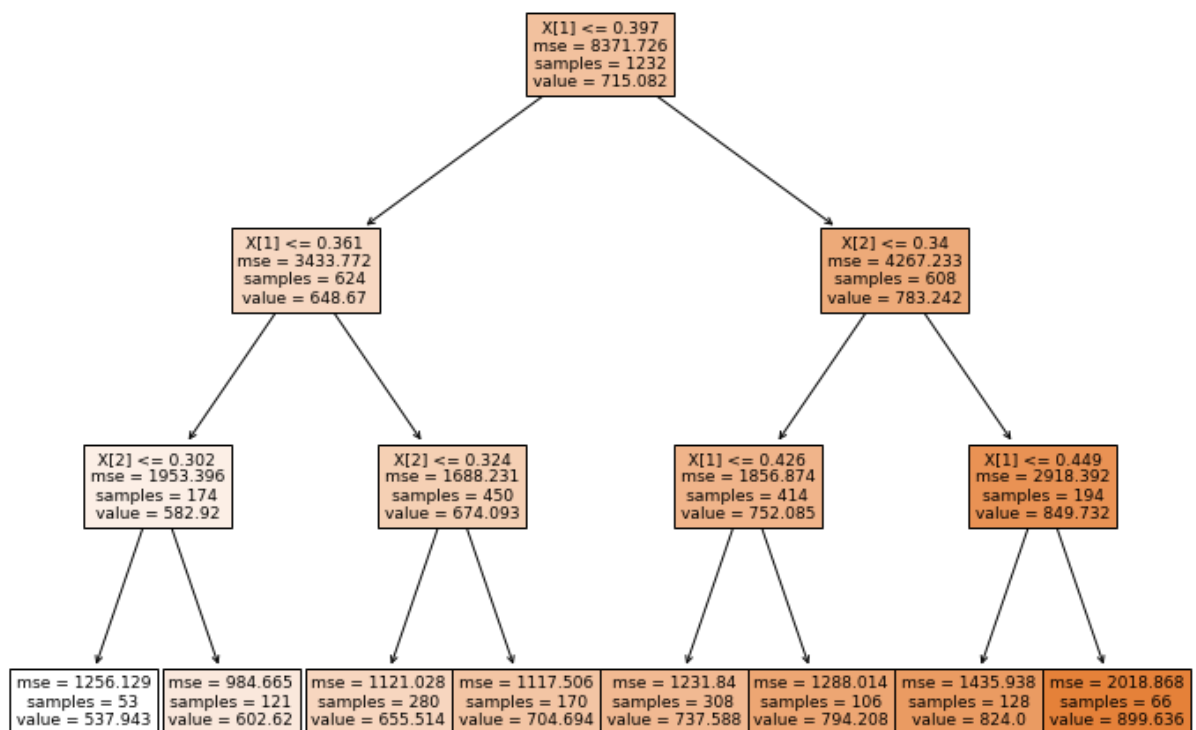
```
In [3]:  # RS ~ BA + SLG + OBP
         ## Predictors and outcome
         X1 = np.array(data[["BA", "SLG", "OBP"]])
         y = np.array(data["RS"])

         ## Build and fit the decision tree
         np.random.seed(263)
         regressor1 = DecisionTreeRegressor(max_depth=3)
         regressor1.fit(X1, y)

         ## Plot the tree
         plt.figure(figsize=(12, 9))
         plot_tree(regressor1, filled=True)
         plt.show()
```
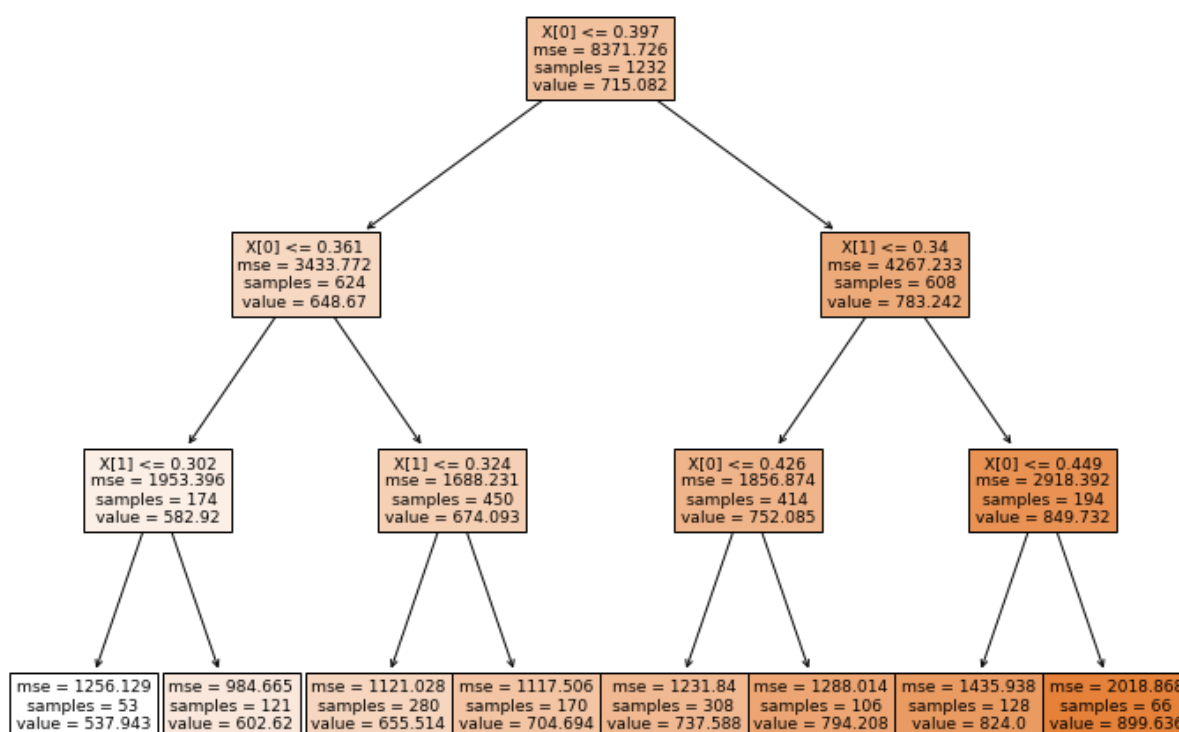
```
In [4]:   # RS ~ SLG + OBP
          ## Predictors and outcome
          X2 = np.array(data[["SLG", "OBP"]])
          y = np.array(data["RS"])

          ## Build and fit the decision tree
          np.random.seed(263)
          regressor2 = DecisionTreeRegressor(max_depth=3)
          regressor2.fit(X2, y)

          ## Plot the tree
          plt.figure(figsize=(12, 9))
          plot_tree(regressor2, filled=True)
          plt.show()
```



```
In [5]:   # Compare the two models
          print("The R\u00b2 of model 1 is " + str(regressor1.score(X1, y)) + ".")
          print("The R\u00b2 of model 2 is " + str(regressor2.score(X2, y)) + ".")
```

```
The R² of model 1 is 0.852378356387466.
The R² of model 2 is 0.852378356387466.
```

There is no difference in the R^2 value of the models. Therefore, both models explain the same amount of variability in the data regardless of whether we include batting average or not. I prefer the second model (without BA) as it is more parsimonious yet gives the same performance as the first one.
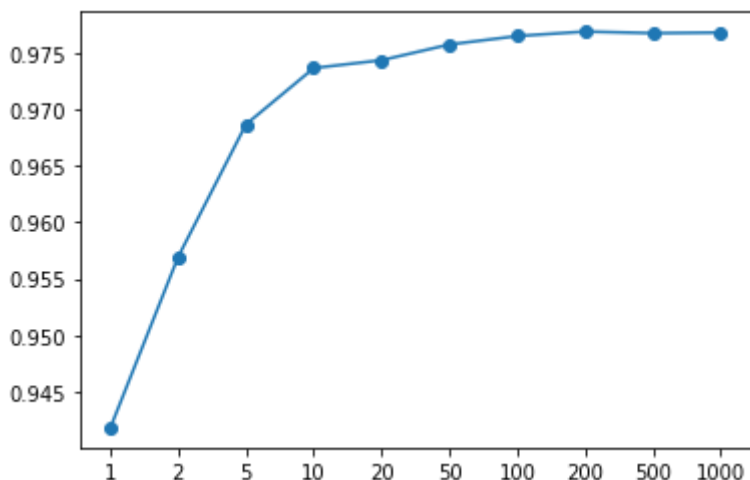
Question 1.2)

```
In [6]:  # Include only data from before 2002
         train = data[data.Year < 2002]
         X_train = np.array(train[["SLG", "OBP"]])
         y_train = np.array(train["RS"])

         # Create RD
         train["RD"] = train["RS"] - train["RA"]
         train = train.dropna()
```

```
In [7]:  # Random forest
         np.random.seed(263)
         Bs = [1, 2, 5, 10, 20, 50, 100, 200, 500, 1000]
         Rsqs = []
         for B in Bs:
             regressor1 = RandomForestRegressor(n_estimators=B, max_depth=10)
             regressor1.fit(X_train, y_train)
             Rsqs.append(regressor1.score(X_train, y_train))

         # Make the plot
         plt.plot(list(map(str, Bs)), Rsqs, "-o")
         plt.show()
```



```
In [8]:  # Number of baggings that maximized R-squared
         max_Rsq = max(Rsqs)
         max_index = Rsqs.index(max_Rsq)
         max_index
         B = Bs[max_index]
```

```
In [9]:  # Use these statistics to predict RS, RA, and W
         SLG = 0.430
         OBP = 0.349
         OOBP = 0.307
         OSLG = 0.373

         # RS ~ SLG + OBP
         regressor1 = RandomForestRegressor(n_estimators=B, max_depth=10)
         regressor1.fit(train[["SLG", "OBP"]], train.RS)
         RS_pred = regressor1.predict(np.array([[SLG, OBP]]))

         # RA ~ OOBP + OSLG
         regressor2 = RandomForestRegressor(n_estimators=B, max_depth=10)
         regressor2.fit(train[["OSLG", "OOBP"]], train.RA)
         RA_pred = regressor2.predict(np.array([[OSLG, OOBP]]))

         # W ~ RD
         RD_pred = RS_pred - RA_pred
         regressor3 = RandomForestRegressor(n_estimators=B, max_depth=10)
         regressor3.fit(train[["RD"]], train.W)
         W_pred = regressor3.predict(np.array([RD_pred]))

         print("The random forest predicts", int(W_pred), "wins in 2002.")
```

```
The random forest predicts 98 wins in 2002.
```

```
In [10]:  # True results
          data[(data["Year"]==2002) & (data["Team"]=="OAK")]
```

Out[10]:

	Team	League	Year	RS	RA	W	OBP	SLG	BA	Playoffs	RankSeason	RankPlayo
320	OAK	AL	2002	800	654	103	0.339	0.432	0.261	1	1.0	

In 2002, Oakland won 103 games. The random forest predicted 98 wins, which is fairly close.

Question 2.1)

```
In [11]:  # Load data
          wikipedia = pd.read_csv("Wikipedia.csv")
          wikipedia.head()
```

Out[11]:

	name	text
0	Michel Che	professor michel che born in lyon france compl...
1	Hossein Modarressi	hossein modarressi is a leading muslim jurist ...
2	Xiao-Gang Wen	xiaogang wen simplified chinese traditional ch...
3	Nicholas Doumanis	nicholas doumanis is an historian of europe an...
4	Don Cahoon	don toot cahoon is a retired american ice hock...

```
In [12]:   # Unique word count
           X, vocab = vectorize(wikipedia["text"])
           print(str(np.shape(X)[0]) + ' individuals; ' + str(len(vocab)) + ' word
           s')
           print(vocab)
```

812 individuals; 8034 words
['aaa', 'aaron', 'ab', 'aba', 'ababa', 'abandoned', 'abandoning', 'abandonment', 'abba', 'abbott', 'abc', 'abelian', 'aberdeen', 'abigail', 'ability', 'able', 'abraham', 'abrams', 'abroad', 'absence', 'absolute', 'abstract', 'abu', 'abundance', 'abuse', 'ac', 'academia', 'academic', 'academically', 'academicborn', 'academician', 'academy', 'acadmie', 'accelerating', 'accelerator', 'accept', 'accepted', 'accepting', 'access', 'accessible', 'accident', 'acclaim', 'acclaimed', 'accolade', 'accompany', 'accompanying', 'accomplish', 'accomplished', 'accomplishment', 'according', 'account', 'accountability', 'accountant', 'accounting', 'accredited', 'accretion', 'accumulated', 'accuracy', 'accurate', 'accurately', 'accusation', 'accused', 'ace', 'achieve', 'achieved', 'achievement', 'achiever', 'achieving', 'acid', 'acknowledged', 'acknowledges', 'acknowledging', 'acm', 'acquired', 'acquiring', 'acquisition', 'act', 'acted', 'acting', 'action', 'activated', 'active', 'actively', 'activism', 'activist', 'activity', 'acton', 'actor', 'actress', 'actual', 'actually', 'ad', 'adam', 'adaptation', 'adapted', 'adapting', 'adaptive', 'adapts', 'added', 'adding', 'addis', 'addition', 'additional', 'additionally', 'address', 'addressed', 'addressing', 'adelaide', 'adept', 'adequate', 'adjunct', 'adjustment', 'adler', 'adm', 'administered', 'administration', 'administrative', 'administrator', 'admired', 'admission', 'admitted', 'admonished', 'adobe', 'adopt', 'adopted', 'adopting', 'adscft', 'adult', 'advance', 'advanced', 'advancement', 'advancing', 'adventure', 'advertising', 'advice', 'advised', 'adviser', 'advises', 'advising', 'advisor', 'advisory', 'advocacy', 'advocate', 'advocating', 'aea', 'aei', 'aeronautics', 'aerospace', 'aesthetic', 'affair', 'affairshe', 'affect', 'affected', 'affecting', 'affiliate', 'affiliated', 'affiliation', 'afghan', 'afghanistan', 'aflcio', 'afraid', 'africa', 'africahe', 'african', 'africanamerican', 'afroamerican', 'aftermath', 'afternoon', 'age', 'aged', 'agency', 'agenda', 'agent', 'agile', 'aging', 'agnes', 'agon', 'agreed', 'agreement', 'agricultural', 'agriculture', 'ahead', 'ai', 'aia', 'aid', 'aide', 'aided', 'aiga', 'aim', 'aimed', 'air', 'airborne', 'aircraft', 'aired', 'airline', 'airport', 'aixmarseille', 'aka', 'akademie', 'akamai', 'akin', 'akron', 'al', 'alabama', 'alain', 'alamo', 'alan', 'alaska', 'albanese', 'albany', 'albert', 'alberta', 'alberto', 'albion', 'albritton', 'album', 'alcott', 'alden', 'aldo', 'aldrich', 'alex', 'alexander', 'alexandra', 'alexandre', 'alexandria', 'alfred', 'algebra', 'algebraic', 'algeria', 'algorithm', 'ali', 'alice', 'alicia', 'alien', 'alike', 'alison', 'alive', 'allamerican', 'allan', 'allegation', 'alleged', 'allen', 'alliance', 'allivy', 'allow', 'allowed', 'allowing', 'allows', 'allrookie', 'allstar', 'allstate', 'alltime', 'allusion', 'alma', 'alongside', 'alonzo', 'alsace', 'altered', 'alternate', 'alternative', 'alto', 'alum', 'alumna', 'alumnus', 'alvin', 'amateur', 'amazing', 'amazon', 'ambassador', 'ambassadorship', 'ambiguous', 'amendment', 'america', 'american', 'americanborn', 'ames', 'amherst', 'amnesty', 'amsterdam', 'amy', 'analog', 'analysis', 'analysisin', 'analyst', 'analytic', 'analytical', 'analyzed', 'analyzes', 'analyzing', 'anatomy', 'ancestor', 'ancestry', 'anchor', 'anchorage', 'ancient', 'anderson', 'andor', 'andover', 'andr', 'andre', 'andreas', 'andrei', 'andrew', 'andriessen', 'andromeda', 'andrs', 'andy', 'angel', 'angeles', 'angelesin', 'anglia', 'anglican', 'anglophone', 'angry', 'animal', 'anisotropy', 'anita', 'ankara', 'ann', 'anna', 'annals', 'annapolis', 'anne', 'annenberg', 'annie', 'anniversary', 'annotated', 'annotation', 'announced', 'anns', 'annual', 'annually', 'anomaly', 'anonymous', 'anschutz', 'answer', 'antarctic', 'antarctica', 'anthologized', 'anthology', 'anthony', 'anthropological', 'anthropologist', 'anthropology', 'antiapartheid', 'anticipated', 'antique', 'a

ntiquity', 'antisemitic', 'antitrust', 'antonio', 'antony', 'ao', 'ao
l', 'aosdna', 'apart', 'apartment', 'apocalypse', 'apollo', 'apostle',
'appalachian', 'apparatus', 'apparently', 'appeal', 'appear', 'appearan
ce', 'appeared', 'appearing', 'appears', 'appel', 'appellate', 'apple',
'applicant', 'application', 'applied', 'apply', 'applying', 'appointe
d', 'appointment', 'appreciation', 'apprentice', 'apprenticed', 'approa
ch', 'appropriation', 'approval', 'approved', 'approximately', 'april',
'aquinas', 'ar', 'arab', 'arabia', 'arabian', 'arabic', 'arabisraeli',
'arbitrary', 'arbitration', 'arbor', 'arc', 'arch', 'archaeological',
'archaeology', 'archeology', 'archer', 'archibald', 'archie', 'architec
t', 'architectural', 'architecture', 'archival', 'archive', 'arctic',
'arden', 'area', 'arendt', 'argentina', 'arguably', 'argued', 'argues',
'arguing', 'argument', 'ariel', 'arise', 'arising', 'aristotelian', 'ar
istotle', 'arithmetic', 'arizona', 'arkansas', 'arlington', 'arm', 'arm
ed', 'armenian', 'armor', 'armstrong', 'army', 'arnold', 'arora', 'aros
e', 'arranged', 'arranger', 'array', 'arrest', 'arrival', 'arrived', 'a
rriving', 'arrow', 'arsenal', 'art', 'artforum', 'arthur', 'article',
'articulated', 'articulation', 'artificial', 'artisan', 'artist', 'arti
stic', 'artstor', 'artwork', 'arvind', 'ascap', 'ash', 'asher', 'ashlan
d', 'asia', 'asian', 'asiapacific', 'asiatic', 'asimov', 'asked', 'aski
ng', 'asme', 'aspect', 'aspen', 'aspiring', 'assassination', 'assault',
'assembly', 'asserted', 'asserts', 'assessed', 'assessment', 'asset',
'assigned', 'assignment', 'assimilation', 'assist', 'assistance', 'assi
stant', 'assisted', 'assisting', 'associate', 'associated', 'associatio
n', 'associative', 'assume', 'assumed', 'assuming', 'assumption', 'aste
roid', 'astronaut', 'astronomer', 'astronomical', 'astronomy', 'astroph
ysical', 'astrophysicist', 'astrophysics', 'asu', 'asylum', 'asymptoti
c', 'athens', 'atherton', 'athlete', 'athletic', 'athletics', 'atlant
a', 'atlantic', 'atlantis', 'atmosphere', 'atmospheric', 'atomic', 'atr
ocity', 'att', 'attached', 'attack', 'attacked', 'attain', 'attained',
'attaining', 'attempt', 'attempted', 'attend', 'attendance', 'attende
d', 'attending', 'attention', 'attitude', 'attorney', 'attracted', 'att
ribute', 'attributed', 'attribution', 'au', 'auction', 'audacity', 'aud
ience', 'audio', 'audiovisual', 'audition', 'aug', 'augsburg', 'augus
t', 'augusto', 'aurora', 'auschwitz', 'auspex', 'austin', 'australia',
'australian', 'austria', 'austrian', 'authentic', 'author', 'authored',
'authorhe', 'authoritarian', 'authoritative', 'authority', 'authorizati
on', 'authorized', 'authorship', 'autism', 'auto', 'autobiography', 'au
tomatic', 'automatically', 'automation', 'automobile', 'autonoma', 'aut
onomous', 'autonomy', 'autopsy', 'autumn', 'available', 'avalanche', 'a
vantgarde', 'avenue', 'average', 'averaged', 'averaging', 'aviation',
'avid', 'aviv', 'avoid', 'avoiding', 'award', 'awarded', 'awardhe', 'aw
ardin', 'awardshe', 'awardwinning', 'awareness', 'away', 'awk', 'axia
l', 'axiom', 'axiomatic', 'axon', 'b', 'ba', 'babbitt', 'baccalaureat
e', 'bach', 'bachelor', 'backed', 'background', 'backing', 'backtobac
k', 'bacon', 'bacterial', 'bad', 'badge', 'bag', 'bailout', 'bainton',
'baker', 'balance', 'balanced', 'balanchine', 'balboa', 'balkan', 'bal
l', 'ballerina', 'ballet', 'balliol', 'ballistic', 'ballot', 'baltimor
e', 'balzan', 'ban', 'band', 'bang', 'bangalore', 'bangkok', 'banglades
h', 'bangor', 'bank', 'banker', 'banking', 'bankruptcy', 'baptist', 'ba
r', 'barack', 'barbara', 'barber', 'barcelona', 'bard', 'barilan', 'bar
ksdale', 'barlow', 'barn', 'barnard', 'barnes', 'baroque', 'barrett',
'barrier', 'barry', 'barth', 'bartlett', 'barton', 'barwise', 'baryoni
c', 'baryshnikov', 'base', 'baseball', 'based', 'basel', 'basic', 'basi
n', 'basis', 'basketball', 'bass', 'bassist', 'bat', 'bates', 'bath',
'baton', 'battalion', 'batted', 'batter', 'battersea', 'batting', 'batt
le', 'bavaria', 'bavarian', 'baxter', 'bay', 'bayard', 'baylor', 'bbc',

'bd', 'beach', 'beak', 'beam', 'bearing', 'beast', 'beat', 'beating', 'beatrice', 'beautiful', 'beauty', 'bebe', 'beck', 'becker', 'beckett', 'bed', 'bedford', 'beethoven', 'began', 'begin', 'beginner', 'beginning', 'behalf', 'behavior', 'behavioral', 'behaviour', 'behrman', 'beijing', 'beirut', 'belfast', 'belfer', 'belgian', 'belgium', 'belgrade', 'belief', 'believe', 'believed', 'believer', 'bell', 'bellagio', 'belly', 'beloved', 'ben', 'bench', 'bend', 'beneath', 'benedict', 'benefit', 'bengurion', 'benjamin', 'bennett', 'bennington', 'benson', 'bentley', 'berg', 'berger', 'bergman', 'berkeley', 'berkeleyhe', 'berklee', 'berkshire', 'berlin', 'berman', 'bern', 'bernanke', 'bernard', 'bernstein', 'berra', 'berry', 'berth', 'besley', 'best', 'bestknown', 'bestowed', 'bestseller', 'bestselling', 'bet', 'beta', 'beth', 'bethesda', 'betrayal', 'better', 'beverly', 'bfa', 'bhargava', 'bhatnagar', 'bianchi', 'bias', 'bible', 'biblical', 'bibliography', 'bicentennial', 'bid', 'bidding', 'biennial', 'big', 'bigger', 'biggest', 'bilateral', 'billing', 'billion', 'billy', 'biloxi', 'bin', 'binding', 'bingham', 'biochemical', 'biochemistry', 'biodiversity', 'bioethics', 'biographer', 'biographical', 'biography', 'bioinformatics', 'biological', 'biologist', 'biology', 'biomedical', 'biomolecular', 'biophysical', 'biophysics', 'biosynthetic', 'biotechnology', 'bipartisan', 'bird', 'birkbeck', 'birmingham', 'birth', 'birthday', 'biscuit', 'bishop', 'bit', 'black', 'blackburn', 'blackwell', 'blade', 'blake', 'blend', 'blent', 'blind', 'bloc', 'block', 'blocked', 'blog', 'blogger', 'blood', 'bloom', 'bloomberg', 'bloomington', 'blowing', 'blue', 'bluegrass', 'blunt', 'board', 'boardin', 'boarding', 'boat', 'boazii', 'bob', 'bobby', 'boca', 'bochum', 'body', 'boeing', 'boggs', 'bohr', 'bok', 'bold', 'bologna', 'bolshoi', 'bolster', 'bomb', 'bond', 'bonn', 'bonnie', 'book', 'boom', 'booth', 'bordeaux', 'bordentown', 'border', 'borisov', 'born', 'borough', 'bosnia', 'bosonic', 'boston', 'botanical', 'botany', 'bott', 'bottle', 'bought', 'boulder', 'bound', 'boundary', 'boutique', 'bow', 'bowdoin', 'bowen', 'bowl', 'bowling', 'box', 'boy', 'boycott', 'boyle', 'brad', 'bradley', 'bradshaw', 'brain', 'branch', 'brand', 'brandeis', 'branding', 'brane', 'brasil', 'brass', 'brave', 'bravo', 'bray', 'brazil', 'brazilhe', 'brazilian', 'bread', 'break', 'breaking', 'breakout', 'breakthrough', 'breath', 'breeder', 'brendan', 'brennan', 'brent', 'brett', 'brian', 'bribery', 'brick', 'bride', 'bridge', 'brief', 'briefly', 'brigham', 'bright', 'brill', 'brilliant', 'bring', 'bringing', 'brings', 'brisbane', 'bristol', 'britain', 'britannica', 'british', 'britishamerican', 'broad', 'broadcast', 'broadcaster', 'broadcasting', 'broader', 'broadly', 'broadway', 'broke', 'broken', 'brokerage', 'bronx', 'bronze', 'brook', 'brooke', 'brookings', 'brooklyn', 'brother', 'brought', 'brouwer', 'browder', 'brown', 'bruce', 'brun', 'bruno', 'brunswick', 'brussels', 'brutality', 'bryan', 'bryn', 'brzezinski', 'bsc', 'bse', 'bubble', 'bucharest', 'buck', 'buckley', 'bucknell', 'budapest', 'buddha', 'buddhism', 'buddhist', 'buddy', 'budget', 'buffalo', 'build', 'builder', 'building', 'built', 'bulk', 'bull', 'bulletin', 'bundle', 'bureau', 'buressuryvette', 'burke', 'burlington', 'burn', 'burning', 'burton', 'burundi', 'bus', 'bush', 'business', 'businessman', 'butch', 'butler', 'buying', 'bypass', 'byron', 'c', 'ca', 'caa', 'cabinet', 'cable', 'cadet', 'cage', 'cain', 'cairn', 'cairo', 'calabiyau', 'calculation', 'calculus', 'calcutta', 'calendar', 'california', 'called', 'calling', 'caltech', 'calvert', 'calvin', 'cambodia', 'cambridge', 'came', 'camera', 'camp', 'campaign', 'campbell', 'campus', 'canada', 'canadian', 'canadianborn', 'canberra', 'cancer', 'candid', 'candidate', 'candor', 'candy', 'cannon', 'canon', 'canonical', 'canopy', 'canton', 'cantor', 'canyon', 'capacity', 'cape', 'capital', 'capitalism', 'capitalist', 'capitol', 'capote', 'captain', 'captained', 'caption', 'capture', 'car', 'c

arbon', 'card', 'cardinal', 'cardiovascular', 'cardozo', 'care', 'cared', 'career', 'caribbean', 'carl', 'carleton', 'carlos', 'carlson', 'carmen', 'carnegie', 'carol', 'carolina', 'caroline', 'carolyn', 'carpenter', 'carr', 'carrie', 'carried', 'carrier', 'carril', 'carroll', 'carry', 'carrying', 'carter', 'cartography', 'cartoon', 'cartoonist', 'cary', 'cascade', 'case', 'cash', 'casino', 'cassette', 'cast', 'castle', 'cat', 'catalog', 'catalogue', 'catalyst', 'category', 'cathedral', 'cather', 'catherine', 'catholic', 'cathryn', 'caucus', 'caught', 'causa', 'causal', 'causally', 'cause', 'caused', 'causing', 'caustic', 'cb', 'cbc', 'cbe', 'cbs', 'cc', 'cd', 'cdrom', 'cecil', 'cedar', 'celebrated', 'celebrating', 'celebration', 'celebrity', 'cell', 'cellist', 'cello', 'cellular', 'centennial', 'center', 'centered', 'central', 'centre', 'centred', 'centro', 'centrum', 'century', 'ceo', 'ceremony', 'cergeei', 'cern', 'certain', 'certificate', 'certified', 'cf', 'chad', 'chain', 'chair', 'chaired', 'chairing', 'chairman', 'chairperson', 'challenge', 'challenged', 'challenging', 'chamber', 'champion', 'championing', 'championship', 'chan', 'chance', 'chancellor', 'change', 'changed', 'changing', 'channel', 'chaos', 'chapel', 'chaplain', 'chapter', 'character', 'characteristic', 'characterized', 'characterizing', 'charg', 'charge', 'charged', 'charismatic', 'charitable', 'charity', 'charles', 'charleston', 'charlesworth', 'charlie', 'charlotte', 'charter', 'charting', 'chase', 'chatham', 'chautauqua', 'check', 'cheerful', 'chef', 'chemical', 'chemist', 'chemistry', 'chemistryin', 'cheney', 'chengdu', 'chennai', 'chern', 'chernsimons', 'cherry', 'ches', 'chester', 'chi', 'chicago', 'chief', 'chiefly', 'child', 'childhood', 'childrenin', 'childrens', 'chile', 'chilean', 'china', 'chinese', 'chineseamerican', 'chip', 'chiral', 'chittenden', 'choate', 'choice', 'choir', 'choose', 'chooses', 'choosing', 'chopin', 'choral', 'choreographed', 'choreographer', 'choreography', 'chorus', 'chose', 'chosen', 'chris', 'christ', 'christensen', 'christi', 'christian', 'christiane', 'christianity', 'christie', 'christine', 'christmas', 'christology', 'christopher', 'chromosome', 'chronic', 'chronicle', 'chteau', 'chuck', 'church', 'churchill', 'cia', 'cincinnati', 'cinema', 'circle', 'circuit', 'circulated', 'circulation', 'circumstance', 'circumvent', 'citation', 'cite', 'cited', 'citing', 'citizen', 'citizenship', 'city', 'citybased', 'cityin', 'civic', 'civil', 'civilian', 'civilization', 'claim', 'claimed', 'claiming', 'claire', 'clara', 'claremont', 'clarence', 'clarinet', 'clarinetist', 'clark', 'clarke', 'class', 'classic', 'classical', 'classicism', 'classicist', 'classification', 'classified', 'classmate', 'classroom', 'claude', 'claudio', 'clause', 'clay', 'clean', 'cleaning', 'cleanup', 'clear', 'cleared', 'clearly', 'clerk', 'clerked', 'cleveland', 'client', 'clifford', 'climate', 'clinic', 'clinical', 'clinician', 'clinton', 'clipper', 'clive', 'cloning', 'close', 'closed', 'closely', 'clothing', 'cloud', 'club', 'cluster', 'clustering', 'cmo', 'cnn', 'cnrs', 'coach', 'coached', 'coaching', 'coal', 'coalition', 'coanchor', 'coast', 'coastal', 'coauthor', 'coauthored', 'coauthoring', 'coauthorship', 'cochair', 'cocreator', 'cod', 'codd', 'code', 'codeveloped', 'coding', 'codirector', 'codirects', 'coedited', 'coeditor', 'coedits', 'coffee', 'cofounded', 'cofounder', 'cognition', 'cognitive', 'cohen', 'cohomology', 'cohosted', 'coif', 'coincided', 'coined', 'coining', 'coinvented', 'coinventor', 'colbert', 'cold', 'cole', 'coleman', 'colgate', 'colin', 'collaborated', 'collaborates', 'collaborating', 'collaboration', 'collaborative', 'collaborator', 'collapse', 'collar', 'colleague', 'collect', 'collected', 'collecting', 'collection', 'collective', 'collector', 'college', 'collegiate', 'collegium', 'collge', 'collins', 'cologne', 'colonel', 'colonial', 'colonialism', 'colonization', 'colony', 'color', 'colorado', 'columbia', 'columbian', 'columbiaprinceton', 'columbus', 'colum

n', 'columnist', 'combat', 'combating', 'combination', 'combinatorial',
'combinatorics', 'combine', 'combined', 'combining', 'comcast', 'come',
'comedy', 'comfortable', 'comic', 'coming', 'command', 'commander', 'co
mmanding', 'commemorating', 'commemoration', 'commencement', 'commendat
ion', 'comment', 'commentary', 'commentator', 'commented', 'commentin
g', 'commerce', 'commercial', 'commercialization', 'commercially', 'com
mission', 'commissioned', 'commissioner', 'commitment', 'committed', 'c
ommittee', 'common', 'commonly', 'commonweal', 'commonwealth', 'communa
l', 'communicate', 'communication', 'communicator', 'communism', 'commu
nist', 'community', 'compact', 'companion', 'company', 'comparative',
'comparing', 'comparison', 'compass', 'compensation', 'compete', 'compe
ted', 'competing', 'competition', 'competitive', 'competitor', 'compila
tion', 'compiled', 'compiler', 'compiling', 'complained', 'complaint',
'complement', 'complete', 'completed', 'completely', 'completing', 'com
pletion', 'complex', 'complexity', 'compliance', 'component', 'compose
d', 'composer', 'composerinresidence', 'composing', 'composition', 'com
positional', 'compound', 'comprehensive', 'comprises', 'comprising', 'c
omptroller', 'compuserve', 'computability', 'computation', 'computation
al', 'computer', 'computing', 'conceded', 'concentrate', 'concentrate
d', 'concentration', 'concept', 'conception', 'conceptual', 'concern',
'concerned', 'concerning', 'concert', 'concerto', 'concluded', 'conclus
ion', 'concordia', 'concrete', 'concurrently', 'condensed', 'conditio
n', 'conduct', 'conducted', 'conducting', 'conductor', 'conduit', 'con
e', 'conference', 'confessed', 'confirm', 'confirmation', 'confirmed',
'conflict', 'conformal', 'confronting', 'confusion', 'congo', 'congrega
tion', 'congregational', 'congress', 'congressional', 'congressman', 'c
onjecture', 'conjunction', 'connect', 'connected', 'connecticut', 'conn
ecting', 'connection', 'conner', 'connie', 'conquest', 'conrad', 'consc
ience', 'consciousness', 'consecrated', 'consecutive', 'consensus', 'co
nsent', 'consequence', 'conservation', 'conservatism', 'conservative',
'conservatory', 'conserved', 'consider', 'considerable', 'considerabl
y', 'considered', 'considering', 'considers', 'consistency', 'consisten
t', 'consisting', 'consists', 'consortium', 'conspiracy', 'constance',
'constant', 'constantine', 'constitution', 'constitutional', 'constrain
t', 'construct', 'constructed', 'construction', 'constructive', 'consu
l', 'consult', 'consultancy', 'consultant', 'consultative', 'consulte
d', 'consulting', 'consumer', 'contact', 'contain', 'contained', 'conta
ining', 'contains', 'contemporary', 'content', 'contest', 'context', 'c
ontinent', 'continental', 'contingent', 'continue', 'continued', 'conti
nues', 'continuing', 'continuity', 'continuous', 'continuum', 'contra',
'contract', 'contracting', 'contraction', 'contrary', 'contrast', 'cont
ribute', 'contributed', 'contributes', 'contributing', 'contribution',
'contributor', 'control', 'controlled', 'controlling', 'controversial',
'controversy', 'convened', 'convention', 'conventional', 'conversatio
n', 'conversion', 'convert', 'converted', 'converting', 'convicted', 'c
onviction', 'convinced', 'coo', 'cook', 'cookbook', 'cooking', 'coole
y', 'cooper', 'cooperate', 'cooperation', 'cooperhewitt', 'coordinate',
'coordinated', 'coordinating', 'coordinator', 'coowned', 'copenhagen',
'copper', 'coproduced', 'copy', 'copyright', 'core', 'coriolis', 'corne
l', 'cornell', 'corner', 'cornish', 'corp', 'corporate', 'corporation',
'corporationhe', 'corpus', 'correct', 'correctly', 'correlated', 'corre
lation', 'correspondence', 'correspondent', 'corresponding', 'corruptio
n', 'corzine', 'cosmic', 'cosmological', 'cosmology', 'cosmopolitan',
'cosponsored', 'cost', 'costa', 'cottage', 'cough', 'council', 'counse
l', 'counselor', 'counterexample', 'counterpart', 'counterpoint', 'coun
terterrorism', 'counting', 'country', 'county', 'couple', 'coupled', 'c
ourage', 'courageous', 'courant', 'course', 'court', 'courtappointed',

'cousin', 'covenant', 'cover', 'coverage', 'covered', 'covering', 'cowb
oy', 'cowles', 'cowrite', 'cowritten', 'cowrote', 'cox', 'crack', 'crac
ked', 'cracow', 'craddock', 'craft', 'craig', 'cram', 'crash', 'crawfor
d', 'cream', 'create', 'created', 'creating', 'creation', 'creative',
'creativity', 'creator', 'credential', 'credit', 'credited', 'creek',
'crest', 'crew', 'cri', 'crime', 'criminal', 'criminology', 'crisis',
'crispin', 'criterion', 'critic', 'critical', 'critically', 'criticis
m', 'criticized', 'criticizes', 'critique', 'croatia', 'cronkite', 'cro
oked', 'crosby', 'cross', 'crosscultural', 'crossing', 'crossroad', 'cr
ossword', 'crown', 'crucial', 'cruel', 'crust', 'cruz', 'cryptographi
c', 'cryptography', 'crystal', 'cs', 'cspan', 'ct', 'cuba', 'cullom',
'culminated', 'culminating', 'cult', 'cultural', 'culture', 'cum', 'cum
mings', 'cumulative', 'cuny', 'cup', 'curated', 'curator', 'curie', 'cu
riosity', 'currency', 'current', 'currently', 'curriculum', 'curse', 'c
urtis', 'curvature', 'curve', 'custodian', 'custom', 'customer', 'custo
mized', 'cut', 'cutler', 'cutting', 'cv', 'cyberspace', 'cycle', 'cynth
ia', 'cyprus', 'czech', 'd', 'da', 'daad', 'dad', 'daffaires', 'daily',
'dakota', 'dal', 'dalai', 'dale', 'dallas', 'damage', 'dame', 'damon',
'dan', 'dana', 'dance', 'danced', 'dancer', 'dancing', 'danforth', 'dan
gerous', 'daniel', 'danish', 'dannie', 'dans', 'danse', 'darby', 'darc
y', 'darius', 'dark', 'darkness', 'darmstadt', 'dart', 'dartmouth', 'da
rwin', 'data', 'database', 'date', 'dateline', 'daughter', 'dave', 'dav
id', 'davidovsky', 'davidson', 'davidsons', 'davie', 'davis', 'day', 'd
ayan', 'dayin', 'daytoday', 'dayton', 'dc', 'dci', 'deacon', 'dead', 'd
eadlock', 'deadly', 'deal', 'dealing', 'dealt', 'dean', 'dear', 'deat
h', 'debate', 'deborah', 'debt', 'debussy', 'debut', 'debuted', 'dec',
'decade', 'december', 'decentralized', 'deception', 'decided', 'decide
s', 'decision', 'decisionmaking', 'deck', 'decline', 'declined', 'decor
ated', 'decorative', 'dedicated', 'dedication', 'deemed', 'deep', 'deep
er', 'deeply', 'deepwater', 'deerfield', 'defamation', 'defeat', 'defea
ted', 'defeating', 'defence', 'defend', 'defendant', 'defended', 'defen
der', 'defending', 'defends', 'defense', 'defensive', 'deferred', 'defi
cit', 'define', 'defined', 'defines', 'definition', 'definitive', 'defl
ationary', 'degree', 'del', 'delaware', 'delay', 'delegate', 'delegatio
n', 'delhi', 'deliberate', 'deliver', 'delivered', 'delta', 'demand',
'demanded', 'demetrios', 'demise', 'democracy', 'democrat', 'democrati
c', 'demographer', 'demographic', 'demography', 'demonstrably', 'demons
trate', 'demonstrated', 'demonstrating', 'demonstration', 'den', 'denia
l', 'denied', 'denmark', 'dennis', 'denomination', 'dent', 'dentistry',
'denunciation', 'denver', 'department', 'departure', 'deploy', 'deploye
d', 'deployment', 'deported', 'depot', 'depression', 'dept', 'depth',
'deputy', 'der', 'derek', 'descended', 'descent', 'described', 'describ
es', 'describing', 'description', 'desegregation', 'desert', 'design',
'designated', 'designation', 'designed', 'designer', 'designing', 'desi
re', 'desk', 'despair', 'despite', 'destroyed', 'destruction', 'detaile
d', 'detainee', 'detat', 'detection', 'detector', 'determination', 'det
ermine', 'determined', 'deterministic', 'detroit', 'detudes', 'deutsch
e', 'develop', 'developed', 'developer', 'developing', 'development',
'developmental', 'develops', 'device', 'devil', 'devon', 'devoted', 'de
wey', 'dhabi', 'di', 'dia', 'diagnosis', 'diagram', 'dialogue', 'diamon
d', 'diane', 'diaspora', 'dick', 'dickinson', 'dickson', 'dictionary',
'did', 'didnt', 'die', 'died', 'diego', 'dielectric', 'difference', 'di
fferent', 'differential', 'difficult', 'difficulty', 'digest', 'digita
l', 'dijkstra', 'dilemma', 'dillon', 'dimension', 'dimensional', 'dimin
ished', 'dinkins', 'diocese', 'dioxide', 'diploma', 'diplomacy', 'diplo
mat', 'diplomatic', 'diplomatinresidence', 'diptych', 'dirac', 'direc
t', 'directed', 'directeur', 'directing', 'direction', 'directional',

'directly', 'director', 'directorsin', 'directs', 'dirty', 'disabilit
y', 'disabled', 'disappearance', 'disarmament', 'disaster', 'disc', 'di
scharge', 'discharged', 'disciple', 'disciplinary', 'discipline', 'disc
losed', 'disclosing', 'discography', 'discourse', 'discovered', 'discov
ering', 'discovery', 'discrete', 'discrimination', 'discus', 'discusse
d', 'discussing', 'discussion', 'disease', 'disenchanted', 'disillusion
ed', 'disk', 'disney', 'disorder', 'displaced', 'display', 'dispute',
'disputed', 'disruption', 'dissent', 'dissertation', 'dissident', 'diss
olved', 'distant', 'distinct', 'distinction', 'distinctive', 'distingui
shed', 'distinguishes', 'distortion', 'distract', 'distributed', 'distr
ibution', 'district', 'districtin', 'divergence', 'diverse', 'diversit
y', 'divide', 'divided', 'divine', 'divinity', 'division', 'divisiona
l', 'divorce', 'divorced', 'dixon', 'dj', 'djerassi', 'dna', 'dobkin',
'dobson', 'doc', 'doctor', 'doctoral', 'doctorate', 'doctrine', 'docume
nt', 'documenta', 'documentary', 'documenting', 'dodge', 'doe', 'dog',
'doherty', 'doing', 'doll', 'dollar', 'domain', 'domestic', 'dominant',
'dominated', 'dominican', 'dominion', 'don', 'donald', 'donated', 'dona
tion', 'donna', 'donor', 'donovan', 'dont', 'door', 'dora', 'dorfman',
'doris', 'dorm', 'dormitory', 'dorothy', 'dorsey', 'double', 'doublea',
'doubled', 'doubt', 'doug', 'douglas', 'downing', 'download', 'downtow
n', 'doyle', 'dozen', 'dphil', 'dr', 'draft', 'drafted', 'drafting', 'd
ragon', 'drake', 'drama', 'dramatic', 'dramatically', 'draper', 'draw',
'drawing', 'dream', 'dreamworks', 'dressing', 'drew', 'drexel', 'driv
e', 'driven', 'driver', 'dropped', 'drosophila', 'drug', 'drum', 'drumm
er', 'dtudes', 'du', 'dual', 'duality', 'dubai', 'dubbed', 'dublin', 'd
uet', 'duffy', 'duke', 'dummy', 'dun', 'duncan', 'dune', 'dunn', 'duo',
'dupont', 'durham', 'dust', 'dutch', 'duty', 'dvd', 'dwarf', 'dying',
'dylan', 'dynamic', 'dynamical', 'dynasty', 'dyson', 'e', 'eagle', 'ea
r', 'earl', 'earlier', 'earliest', 'early', 'earn', 'earned', 'earnin
g', 'earth', 'earthquake', 'easily', 'east', 'eastcentral', 'easter',
'eastern', 'eastman', 'eastwest', 'eberhard', 'ebony', 'ecac', 'eccentr
icity', 'eckert', 'ecm', 'ecole', 'ecological', 'ecologist', 'ecology',
'econometric', 'econometrics', 'economic', 'economics', 'economicshe',
'economicsin', 'economist', 'economy', 'ecuador', 'ecumenism', 'ed', 'e
ddie', 'eddington', 'eden', 'edgar', 'edge', 'edinburgh', 'edison', 'ed
ited', 'edith', 'edition', 'editor', 'editorial', 'editorin', 'editorin
chief', 'edmonton', 'edmund', 'edn', 'edna', 'edsger', 'educated', 'edu
cation', 'educational', 'educator', 'edward', 'edwin', 'effect', 'effec
tive', 'effectively', 'effectiveness', 'efficiency', 'effort', 'egg',
'egypt', 'egyptian', 'ei', 'eighteen', 'eighteenth', 'eighteenthcentur
y', 'eighth', 'eine', 'einstein', 'eisenhower', 'el', 'elder', 'eleano
r', 'elected', 'election', 'electric', 'electrical', 'electro', 'electr
on', 'electronic', 'electronica', 'electronics', 'element', 'elementa
l', 'elementary', 'elena', 'elevated', 'eleventh', 'elgin', 'eli', 'eli
a', 'elijah', 'eliminated', 'eliminating', 'elimination', 'elinor', 'el
iot', 'elite', 'eliyahu', 'elizabeth', 'ellen', 'elli', 'elliot', 'elli
ott', 'elliptic', 'ellison', 'elucidated', 'elucidation', 'ely', 'emai
l', 'emanate', 'emanuel', 'embarking', 'embassy', 'embodied', 'embrac
e', 'embryo', 'embryonic', 'emerged', 'emergence', 'emergency', 'emerge
nt', 'emerging', 'emerita', 'emeritus', 'emerson', 'emi', 'emigrated',
'emily', 'eminent', 'emission', 'emma', 'emmy', 'emory', 'emotion', 'em
otional', 'emperor', 'emphasis', 'emphasized', 'empire', 'empirical',
'employ', 'employed', 'employee', 'employer', 'employment', 'empowermen
t', 'en', 'enable', 'enabled', 'encoded', 'encompassed', 'encompasses',
'encounter', 'encountering', 'encourage', 'encouraged', 'encourages',
'encryption', 'encyclopaedia', 'encyclopdia', 'encyclopedia', 'end', 'e
ndeavor', 'ended', 'ending', 'endo', 'endorsed', 'endorsement', 'endowe

d', 'endowment', 'energy', 'enforcement', 'enforcing', 'engage', 'engag
ed', 'engagement', 'engaging', 'engine', 'engineer', 'engineering', 'en
gineeringhe', 'england', 'english', 'englishspeaking', 'enhanced', 'enh
ancement', 'enhancing', 'enjoy', 'enjoyed', 'enlarged', 'enlargement',
'enlightenment', 'enlisted', 'enlisting', 'enrico', 'enroll', 'enrolle
d', 'enrolling', 'enschede', 'ensemble', 'ensure', 'ensuring', 'enter',
'entered', 'entering', 'enterprise', 'entertainment', 'enthusiast', 'en
tire', 'entirely', 'entitled', 'entity', 'entrepreneur', 'entry', 'envi
ronment', 'environmental', 'envoy', 'enzyme', 'ep', 'epa', 'epic', 'epi
scopal', 'episode', 'epistemic', 'epistemological', 'epistemology', 'ep
onymous', 'epstein', 'equal', 'equality', 'equally', 'equation', 'eques
trian', 'equilibrium', 'equipment', 'equity', 'equivalence', 'equivalen
t', 'er', 'era', 'erds', 'eric', 'erich', 'erik', 'erlangen', 'ernest',
'ernie', 'ernst', 'error', 'erskine', 'escalation', 'escape', 'esoteri
c', 'espaola', 'especially', 'esquire', 'essay', 'essayist', 'essence',
'essential', 'essentially', 'essex', 'establish', 'established', 'estab
lishing', 'establishment', 'estate', 'esther', 'estimate', 'estimated',
'et', 'eth', 'ethan', 'ethic', 'ethical', 'ethiopia', 'ethiopian', 'eth
nic', 'ethnicity', 'ethnomusicology', 'eton', 'ets', 'etude', 'eu', 'eu
clidean', 'eugene', 'euro', 'europaea', 'europe', 'european', 'eva', 'e
valuate', 'evaluation', 'evan', 'evangelical', 'evangelist', 'evans',
'evanston', 'evansville', 'eve', 'evening', 'eveninglength', 'event',
'eventsin', 'eventual', 'eventually', 'eventuated', 'everett', 'everyda
y', 'evidence', 'evidenced', 'evident', 'evil', 'evolution', 'evolution
ary', 'evolve', 'evolved', 'exact', 'exam', 'examination', 'examine',
'examined', 'examiner', 'examines', 'examining', 'example', 'excelled',
'excellence', 'excellent', 'exceptional', 'exceptionalism', 'exchange',
'excitation', 'excited', 'exciting', 'exclude', 'exclusion', 'exclusiv
e', 'exclusively', 'execution', 'executive', 'exegetical', 'exemplifie
d', 'exercise', 'exerted', 'exeter', 'exhibit', 'exhibited', 'exhibitio
n', 'exile', 'existence', 'existing', 'exists', 'exotic', 'expanded',
'expanding', 'expansion', 'expectation', 'expected', 'expecting', 'expe
dition', 'experience', 'experienced', 'experiment', 'experimental', 'ex
pert', 'expertise', 'explain', 'explained', 'explaining', 'explains',
'explanation', 'explicit', 'exploitation', 'exploration', 'explore', 'e
xplored', 'explorer', 'explores', 'exploring', 'expose', 'exposed', 'ex
position', 'express', 'expressed', 'expression', 'expressiveness', 'exp
ulsion', 'extend', 'extended', 'extending', 'extension', 'extensive',
'extensively', 'external', 'externalism', 'extra', 'extracurricular',
'extragalactic', 'extraordinarily', 'extraordinary', 'extreme', 'extrem
ely', 'extremism', 'eye', 'eyewitness', 'f', 'fable', 'fabra', 'face',
'facebook', 'faced', 'facilitate', 'facilitating', 'facility', 'facin
g', 'fact', 'factor', 'faculty', 'fail', 'failed', 'failing', 'failur
e', 'fair', 'fairfield', 'fairleigh', 'fairness', 'faith', 'faithful',
'falcon', 'fall', 'false', 'fame', 'famed', 'familiar', 'family', 'famo
us', 'famously', 'fan', 'fanciful', 'fanin', 'fannie', 'fantasy', 'fa
r', 'farabi', 'fare', 'fareed', 'farewell', 'farm', 'farrar', 'fascinat
ed', 'fascinating', 'fascination', 'fashion', 'faso', 'fast', 'faster',
'fatal', 'fate', 'father', 'fatherhood', 'fault', 'favor', 'favorite',
'favour', 'favourable', 'fba', 'fbi', 'fe', 'fear', 'feast', 'feature',
'featured', 'featuring', 'feb', 'february', 'federal', 'federalist', 'f
ederation', 'feedback', 'feel', 'feeling', 'felix', 'fell', 'fellow',
'fellowhe', 'fellowship', 'felony', 'felt', 'female', 'feminine', 'femi
nism', 'feminist', 'fenton', 'ferguson', 'fermi', 'fermion', 'fernand
o', 'ferrari', 'ferris', 'festival', 'festschrift', 'fiber', 'fiction',
'fictional', 'fictitious', 'field', 'fieldwork', 'fifth', 'fight', 'fig
hting', 'figure', 'file', 'filed', 'filled', 'film', 'filmmaker', 'fina

l', 'finale', 'finalist', 'finally', 'finance', 'financed', 'financehe', 'financial', 'financing', 'finch', 'finding', 'fine', 'finest', 'finish', 'finished', 'finishing', 'finite', 'finland', 'finn', 'finnish', 'fired', 'firm', 'firmly', 'firstclass', 'firstever', 'firstteam', 'fiscal', 'fischer', 'fish', 'fitzgerald', 'fitzpatrick', 'fiveyear', 'fixed', 'flag', 'flagship', 'flame', 'flash', 'flaw', 'flea', 'flee', 'fleet', 'fleming', 'flemming', 'fletcher', 'flexible', 'flight', 'flip', 'floor', 'florence', 'florida', 'flourish', 'flow', 'floyd', 'fluent', 'fluid', 'flute', 'flux', 'fly', 'flying', 'fo', 'focus', 'focused', 'focusing', 'fogg', 'folk', 'follow', 'followed', 'follower', 'following', 'follows', 'folly', 'fong', 'food', 'foot', 'football', 'foote', 'forbes', 'forbescom', 'force', 'forced', 'ford', 'fordham', 'forecasting', 'forefront', 'foreign', 'foreigner', 'foremost', 'forest', 'foreword', 'forgotten', 'form', 'formal', 'formalism', 'formally', 'format', 'formation', 'formative', 'formed', 'forming', 'formula', 'formulate', 'formulation', 'fort', 'forthcoming', 'fortress', 'fortune', 'forum', 'forward', 'foster', 'fostering', 'foucault', 'foundation', 'foundational', 'foundationin', 'founded', 'founder', 'founding', 'fountain', 'fourier', 'fourteen', 'fourteenth', 'fourth', 'fourtime', 'fouryear', 'fox', 'fr', 'fractal', 'frame', 'framework', 'franaise', 'france', 'francehe', 'franchise', 'francis', 'francisco', 'franciscobased', 'franco', 'frank', 'frankenstein', 'frankfurt', 'franklin', 'franois', 'franz', 'fraser', 'fraud', 'fred', 'freddie', 'frederic', 'frederick', 'fredric', 'free', 'freedman', 'freedom', 'freelance', 'freely', 'freeman', 'freespeech', 'frege', 'freiburg', 'french', 'frequent', 'frequently', 'fresh', 'freshman', 'fresno', 'freud', 'frey', 'friday', 'friedrich', 'friend', 'friendship', 'fringe', 'fromm', 'frontier', 'frsc', 'frse', 'fruit', 'fruitful', 'fsg', 'fu', 'fuel', 'fugue', 'fulbright', 'fulbrighthays', 'fulfilling', 'fulfillment', 'fulkerson', 'fullback', 'fuller', 'fulllength', 'fulltime', 'fully', 'fulton', 'function', 'functional', 'fund', 'fundamental', 'fundamentally', 'fundao', 'funded', 'funding', 'fundraising', 'fungi', 'funk', 'furnace', 'furniss', 'fuse', 'fusion', 'future', 'g', 'gabriel', 'gabriele', 'gained', 'gainesville', 'gala', 'galaxy', 'gale', 'gallery', 'gallup', 'galpagos', 'gamble', 'gambling', 'game', 'gamewinning', 'gandhi', 'gang', 'gap', 'garage', 'garden', 'garnering', 'garrison', 'garth', 'gary', 'gas', 'gate', 'gateway', 'gather', 'gathering', 'gauge', 'gauss', 'gave', 'gay', 'gdel', 'geist', 'gender', 'gene', 'genealogy', 'general', 'generalization', 'generalize', 'generally', 'generated', 'generating', 'generation', 'generator', 'generically', 'genesis', 'genetic', 'geneticist', 'genetics', 'geneva', 'genital', 'genius', 'genocide', 'genome', 'genomic', 'genomics', 'genre', 'gentnerkastler', 'genuine', 'genus', 'geo', 'geoff', 'geoffrey', 'geographic', 'geographical', 'geography', 'geologic', 'geological', 'geologist', 'geology', 'geometric', 'geometry', 'geophysical', 'geophysicist', 'geophysics', 'george', 'georgetown', 'georgia', 'geosciences', 'gerald', 'gerard', 'gerhard', 'german', 'germanic', 'germany', 'gertrud', 'gesellschaft', 'gestapo', 'gesture', 'getting', 'getty', 'ghana', 'ghent', 'ghost', 'ghraib', 'giant', 'gibbon', 'gibson', 'gifford', 'gift', 'gifted', 'gig', 'gilbert', 'gill', 'gillian', 'gilman', 'gilmore', 'gina', 'girl', 'giroux', 'giuseppe', 'given', 'giving', 'glaeser', 'glasgow', 'glass', 'glen', 'glenn', 'global', 'globalization', 'globally', 'globe', 'gloria', 'glory', 'goal', 'goalie', 'god', 'goddard', 'godunov', 'goethe', 'going', 'gold', 'goldberg', 'golden', 'goldman', 'goldsmith', 'golf', 'gone', 'good', 'goodman', 'goodness', 'goodrich', 'google', 'gorbachev', 'gordon', 'gore', 'gospel', 'got', 'gotham', 'gothic', 'gottlob', 'goucher', 'gould', 'govern', 'governance', 'governing', 'government', 'governmental', 'governmentsponsored', 'govern

or', 'gq', 'grace', 'grad', 'grade', 'grader', 'graduate', 'graduated', 'graduating', 'graduation', 'graf', 'graffiti', 'graham', 'grain', 'gramercy', 'grammar', 'grammy', 'gramophone', 'grand', 'grandchild', 'granddaughter', 'grandfather', 'grandparent', 'grandson', 'grant', 'granted', 'granting', 'graph', 'graphic', 'grard', 'gravitation', 'gravitational', 'gravity', 'grawemeyer', 'gray', 'graz', 'great', 'greater', 'greatest', 'greatgrandfather', 'greatly', 'greece', 'greed', 'greek', 'green', 'greene', 'greenhouse', 'greensboro', 'greenwich', 'greer', 'greg', 'gregg', 'gregory', 'gresham', 'grew', 'grey', 'griffin', 'griffith', 'grisey', 'grnbaum', 'groningen', 'gross', 'ground', 'groundbreaking', 'group', 'grove', 'growing', 'grown', 'growth', 'gruber', 'gttingen', 'guangdong', 'guangzhou', 'guarantee', 'guard', 'guardian', 'guatemala', 'gubernatorial', 'guerrilla', 'guest', 'guggenheim', 'guidance', 'guide', 'guided', 'guideline', 'guild', 'guillaume', 'guinea', 'guitar', 'guitarist', 'gulf', 'gun', 'gunn', 'gunnar', 'guy', 'gwendolyn', 'gynecology', 'h', 'ha', 'habilitation', 'hackett', 'haddam', 'hadley', 'hagen', 'hague', 'hahn', 'haiti', 'haitian', 'half', 'hall', 'halo', 'hamburg', 'hamilton', 'hamiltonian', 'hampshire', 'hampton', 'han', 'hancock', 'hand', 'handbook', 'handel', 'handful', 'handling', 'handson', 'hanging', 'hank', 'hannah', 'hanoi', 'hansen', 'hanson', 'happen', 'happened', 'happiest', 'happiness', 'happy', 'harassment', 'harbor', 'hard', 'hardcore', 'hardware', 'harlem', 'harmonic', 'harmony', 'harness', 'harold', 'harper', 'harriman', 'harris', 'harrison', 'harry', 'hart', 'hartford', 'harvard', 'harvardsmithsonian', 'harvest', 'harvey', 'hat', 'hautes', 'haven', 'haverford', 'having', 'hawaii', 'hawk', 'hay', 'haynes', 'head', 'headed', 'heading', 'headliner', 'headquartered', 'headquarters', 'healing', 'health', 'healthcare', 'healthy', 'heard', 'hearing', 'hearst', 'heart', 'heat', 'heated', 'heather', 'heaven', 'heavily', 'heavy', 'hebrew', 'hed', 'hedge', 'heidelberg', 'height', 'heineman', 'heinrich', 'heir', 'held', 'helen', 'helium', 'hellenic', 'heller', 'helm', 'helmut', 'help', 'helped', 'helping', 'helsinki', 'hempel', 'henderson', 'hendry', 'henfield', 'henri', 'henry', 'henryk', 'hepp', 'herald', 'herb', 'herbert', 'herbie', 'heritage', 'herman', 'hermeneutics', 'hermitage', 'hermon', 'hero', 'herzegovina', 'herzog', 'heterogeneous', 'heuristic', 'heywood', 'hick', 'hid', 'hidden', 'hierarchy', 'higgins', 'high', 'higher', 'higherspin', 'highest', 'highland', 'highlight', 'highlighted', 'highly', 'highprofile', 'highspeed', 'hightech', 'hightemperature', 'highway', 'hilary', 'hilbert', 'hill', 'hillary', 'hillel', 'hillman', 'himalaya', 'hindex', 'hindi', 'hindu', 'hired', 'hiring', 'hiroshima', 'hispanic', 'histoire', 'historian', 'historic', 'historical', 'historically', 'history', 'historyborn', 'hit', 'hitler', 'hobby', 'hoboken', 'hoc', 'hochschule', 'hockey', 'hockney', 'hodder', 'hoffman', 'hold', 'holder', 'holding', 'hole', 'holiday', 'holland', 'holly', 'hollywood', 'holmes', 'holocaust', 'holt', 'holy', 'homage', 'home', 'homeland', 'homeschooled', 'hometown', 'homogeneous', 'homology', 'homosexual', 'homotopy', 'honeywell', 'hong', 'honolulu', 'honor', 'honorable', 'honorary', 'honored', 'honoree', 'honoring', 'honoris', 'honour', 'hood', 'hoosier', 'hoover', 'hope', 'hopewell', 'hopkins', 'horaiu', 'horizon', 'horn', 'horowitz', 'horror', 'horse', 'horton', 'hospital', 'host', 'hosted', 'hostile', 'hosting', 'hot', 'hotchkiss', 'hotel', 'hour', 'house', 'housed', 'household', 'housing', 'houston', 'howard', 'hoyt', 'hr', 'hubbell', 'hubble', 'hubert', 'hudson', 'huffington', 'huge', 'hugh', 'hughes', 'hugo', 'human', 'humane', 'humanism', 'humanitarian', 'humanity', 'humboldt', 'hume', 'hummer', 'humor', 'humphrey', 'hun', 'hungarian', 'hungary', 'hunger', 'hunter', 'huntington', 'hurricane', 'hurst', 'husband', 'hut', 'hutchinson', 'hutter', 'hybrid', 'hyde', 'hyderabad', 'hydrocarbon', 'hymn', 'hy

perelliptic', 'hypothesis', 'ia', 'iain', 'ian', 'ibm', 'ice', 'icm', 'icon', 'iconic', 'ictp', 'id', 'idaho', 'idea', 'ideal', 'identification', 'identified', 'identifies', 'identifying', 'identity', 'ideological', 'ideology', 'ieee', 'ignorance', 'igor', 'ihes', 'ii', 'iii', 'iit', 'il', 'ill', 'illegal', 'illinois', 'illness', 'illuminating', 'illusion', 'illustrate', 'illustrated', 'illustration', 'illustrator', 'ilya', 'im', 'image', 'imagination', 'imaginative', 'imagined', 'imaging', 'imbalance', 'immanuel', 'immediate', 'immediately', 'immigrant', 'immigrated', 'immigrating', 'immigration', 'immortality', 'impact', 'impasse', 'imperative', 'imperial', 'implement', 'implementation', 'implemented', 'implicated', 'implication', 'implied', 'implies', 'import', 'importance', 'important', 'importexport', 'impose', 'impossible', 'impression', 'impressive', 'imprint', 'improper', 'improve', 'improved', 'improvement', 'improving', 'improvisation', 'inaugural', 'inaugurated', 'incentive', 'inception', 'incident', 'include', 'included', 'includes', 'including', 'income', 'incoming', 'incorporated', 'incorporates', 'incorporating', 'increase', 'increased', 'increasing', 'increasingly', 'incumbent', 'independence', 'independent', 'independently', 'indepth', 'index', 'india', 'indian', 'indiana', 'indianapolis', 'indicate', 'indicated', 'indication', 'indie', 'individual', 'indochina', 'indonesia', 'indonesian', 'inducted', 'industrial', 'industry', 'inequality', 'infantry', 'infinite', 'infinitely', 'infinitesimal', 'infinity', 'inflation', 'inflationary', 'influence', 'influenced', 'influencing', 'influential', 'inform', 'informal', 'information', 'informed', 'informix', 'infrared', 'infrastructure', 'ingram', 'initial', 'initially', 'initiated', 'initiating', 'initiative', 'initiator', 'injured', 'injury', 'inn', 'inner', 'inning', 'innova', 'innovation', 'innovative', 'innovator', 'innsbruck', 'inquirer', 'inquiry', 'insead', 'insect', 'inside', 'insight', 'insisted', 'inspector', 'inspiration', 'inspired', 'inspiring', 'instability', 'installation', 'instance', 'instead', 'institut', 'institute', 'institutehe', 'institutein', 'institution', 'institutional', 'institutionalized', 'institutionborn', 'instituto', 'instruction', 'instructional', 'instructor', 'instrument', 'instrumental', 'insurance', 'integrable', 'integral', 'integrated', 'integration', 'integrity', 'intellectual', 'intelligence', 'intelligent', 'intended', 'intense', 'intensified', 'intensity', 'intensive', 'intention', 'interaction', 'interactive', 'interchange', 'intercollegiate', 'intercultural', 'interdisciplinary', 'interested', 'interesting', 'interface', 'interfacial', 'interfaith', 'intergovernmental', 'interim', 'interior', 'intern', 'internal', 'international', 'internationale', 'internationalin', 'internationally', 'internet', 'internship', 'interpersonal', 'interpretation', 'interpreted', 'interpreter', 'interrogation', 'interrupted', 'intersection', 'interstellar', 'intervene', 'intervention', 'interview', 'interviewed', 'interviewing', 'intimacy', 'intiman', 'intrigued', 'introduce', 'introduced', 'introducing', 'introduction', 'introductory', 'intuition', 'invariant', 'invasion', 'invented', 'inventing', 'invention', 'inventor', 'inverse', 'invested', 'investigate', 'investigated', 'investigates', 'investigating', 'investigation', 'investigative', 'investigator', 'investing', 'investment', 'investor', 'invitation', 'invitational', 'invited', 'involve', 'involved', 'involvement', 'involving', 'ion', 'iowa', 'ipcc', 'iphigenia', 'ipo', 'ira', 'iran', 'iranian', 'iraq', 'iraqi', 'ircam', 'ireland', 'irene', 'iris', 'irish', 'iron', 'ironically', 'irvine', 'irving', 'irwin', 'isaac', 'isabella', 'isabelle', 'isaiah', 'isbn', 'isi', 'islam', 'islamabad', 'islamic', 'islamist', 'island', 'islip', 'isnt', 'isolation', 'isotope', 'ispa', 'israel', 'israeli', 'israeliamerican', 'issue', 'issued', 'istanbul', 'italian', 'italy', 'italyhe', 'ithaca', 'iv', 'ives', 'ivy', 'iza', 'j',

'jack', 'jackson', 'jacob', 'jacques', 'jaffe', 'jahnsson', 'jakarta', 'jam', 'jamaica', 'james', 'jan', 'jane', 'janeiro', 'janet', 'january', 'japan', 'japanese', 'jared', 'jason', 'jay', 'jazz', 'jd', 'jean', 'jeanmarie', 'jee', 'jeff', 'jefferson', 'jeffrey', 'jenkins', 'jennifer', 'jenny', 'jensen', 'jeremiah', 'jeremy', 'jerome', 'jerry', 'jersey', 'jerusalem', 'jesse', 'jesuit', 'jesus', 'jew', 'jewish', 'jewishchristian', 'jewry', 'ji', 'jihad', 'jill', 'jim', 'jimmy', 'jinan', 'joan', 'joanna', 'job', 'joe', 'joel', 'johann', 'john', 'johnny', 'johnson', 'join', 'joined', 'joining', 'joint', 'jointly', 'jon', 'jonathan', 'jones', 'jordan', 'jos', 'jose', 'joseph', 'josh', 'joshua', 'josiah', 'journal', 'journalism', 'journalist', 'journey', 'joyce', 'jp', 'jr', 'jstor', 'juan', 'jubilee', 'judaism', 'judd', 'judge', 'judged', 'judgment', 'judicial', 'judiciary', 'judith', 'judy', 'juilliard', 'jules', 'julia', 'julian', 'julie', 'juliet', 'julius', 'july', 'jump', 'june', 'junior', 'junk', 'juris', 'jurisprudence', 'jurist', 'jury', 'just', 'justice', 'justification', 'justify', 'juvenile', 'k', 'kafka', 'kagan', 'kahn', 'kaiser', 'kanpur', 'kansa', 'kant', 'kantian', 'kappa', 'karen', 'karl', 'karlheinz', 'kate', 'katherine', 'kathleen', 'kathmandu', 'kathy', 'katie', 'katrina', 'katz', 'kauffman', 'kaufman', 'kavli', 'kay', 'kean', 'kearny', 'keeping', 'keith', 'keller', 'kelley', 'kellogg', 'kelly', 'ken', 'kennedy', 'kenneth', 'kent', 'kentucky', 'kenya', 'keogh', 'kept', 'kerr', 'kerrey', 'kerry', 'kevin', 'key', 'keyboard', 'keyboardist', 'keynote', 'khan', 'kid', 'kill', 'killed', 'killing', 'kilometer', 'kim', 'kimberly', 'kind', 'kindergarten', 'kinetic', 'kinetics', 'king', 'kingdom', 'kingston', 'kip', 'kippur', 'kirk', 'kiss', 'kissinger', 'kit', 'kitchen', 'klaus', 'klein', 'kleiner', 'km', 'knew', 'knicks', 'knight', 'knighted', 'knopf', 'knot', 'know', 'knowing', 'knowledge', 'known', 'knox', 'knuth', 'koch', 'kodak', 'kolkata', 'kong', 'konstanz', 'kor', 'korea', 'korean', 'korn', 'kramer', 'kripke', 'kris', 'kristin', 'krugman', 'ktheory', 'kuala', 'kumar', 'kung', 'kurt', 'kurtz', 'kuwait', 'kyoto', 'l', 'la', 'lab', 'label', 'labor', 'laboratory', 'labour', 'lack', 'lacked', 'lacrosse', 'laden', 'lady', 'laid', 'lakatos', 'lake', 'lakers', 'lama', 'lambda', 'lambert', 'lancaster', 'lance', 'land', 'landmark', 'landon', 'landscape', 'landslide', 'lane', 'lang', 'langevin', 'language', 'lanka', 'lao', 'laporte', 'laptop', 'large', 'largely', 'larger', 'largescale', 'largest', 'larry', 'lars', 'late', 'later', 'latest', 'latin', 'latino', 'latterday', 'lattice', 'laude', 'launch', 'launched', 'launching', 'laura', 'laurance', 'laureate', 'laurel', 'lauren', 'laurence', 'laurie', 'law', 'lawrence', 'lawrenceville', 'lawsuit', 'lawyer', 'lay', 'layer', 'lcd', 'lds', 'le', 'lea', 'leach', 'lead', 'leader', 'leadership', 'leading', 'leaf', 'league', 'lear', 'learn', 'learned', 'learning', 'leave', 'leaving', 'lebanon', 'lecture', 'lectured', 'lecturer', 'lectureship', 'lecturing', 'led', 'lee', 'leeds', 'left', 'leg', 'legacy', 'legal', 'legalization', 'legend', 'legendary', 'legion', 'legislation', 'legislative', 'legislator', 'legislature', 'legon', 'lehigh', 'lehmann', 'leibniz', 'leiden', 'leipzig', 'lending', 'length', 'lengthy', 'lenny', 'lens', 'leo', 'leon', 'leonard', 'leopoldina', 'lerner', 'leroy', 'lesbian', 'lesson', 'lester', 'let', 'letter', 'lettered', 'leukemia', 'lev', 'level', 'leverhulme', 'levin', 'levine', 'levy', 'lewis', 'lexecon', 'lexington', 'lgbt', 'li', 'liability', 'liaison', 'liberal', 'liberalism', 'liberation', 'libert', 'libertarian', 'liberty', 'librarian', 'library', 'libre', 'librettist', 'libretto', 'license', 'licensing', 'licentiate', 'licia', 'lie', 'lieb', 'liechtenstein', 'lieutenant', 'life', 'lifelong', 'lifestyle', 'lifetime', 'lift', 'light', 'like', 'likelihood', 'likely', 'lillian', 'lim', 'limb', 'limit', 'limitation', 'limited', 'lin', 'lincoln', 'lindback', 'lindsay', 'line', 'linear', 'lin

ebacker', 'linguist', 'linguistic', 'linguistics', 'link', 'linkage', 'linked', 'linux', 'lion', 'lionel', 'lippincott', 'lisa', 'list', 'listed', 'listener', 'listening', 'literacy', 'literal', 'literally', 'literary', 'literature', 'lithograph', 'litigation', 'little', 'liturgy', 'liu', 'live', 'lived', 'living', 'ljubljana', 'llb', 'llc', 'llm', 'llp', 'loan', 'lobbied', 'local', 'locally', 'locate', 'located', 'location', 'locke', 'lockwood', 'loeb', 'log', 'logic', 'london', 'londonbased', 'londonin', 'long', 'longer', 'longest', 'longing', 'longrange', 'longstanding', 'longterm', 'longtime', 'longy', 'look', 'looking', 'lookout', 'loop', 'loose', 'lord', 'lori', 'los', 'lose', 'losing', 'loss', 'lost', 'lot', 'lotus', 'lou', 'louis', 'louisa', 'louise', 'louisiana', 'louisville', 'louvain', 'love', 'loved', 'lover', 'low', 'lowdimensional', 'lowell', 'lower', 'lowry', 'loyola', 'lse', 'luc', 'lucas', 'luce', 'lucerne', 'lucien', 'luck', 'lucy', 'ludwig', 'lugar', 'luigi', 'luke', 'luminary', 'luminosity', 'luminous', 'luther', 'lutheran', 'lyman', 'lyndon', 'lynn', 'lynne', 'lyon', 'lyric', 'lyrical', 'lyricist', 'm', 'ma', 'mac', 'macarthur', 'macdowell', 'machine', 'machinery', 'mack', 'mackay', 'macmillan', 'macon', 'macquarie', 'macroeconomic', 'macroeconomics', 'mad', 'madhyamaka', 'madison', 'madness', 'madras', 'madrid', 'mae', 'magazine', 'magdalen', 'magic', 'magna', 'magnetic', 'magnetism', 'mahatma', 'mahayana', 'maier', 'mail', 'main', 'maine', 'mainly', 'mainstream', 'maintain', 'maintained', 'maintaining', 'maintains', 'maison', 'major', 'majored', 'majoring', 'majority', 'make', 'maker', 'making', 'malaria', 'malaya', 'malaysia', 'malaysian', 'malcolm', 'maldacena', 'male', 'mama', 'mammal', 'mammalian', 'man', 'managed', 'management', 'manager', 'manages', 'managing', 'manchester', 'mandate', 'mandatory', 'mandela', 'manhattan', 'manifested', 'manifold', 'manipulation', 'manitoba', 'manjul', 'manner', 'manon', 'mansfield', 'mansion', 'manual', 'manufacture', 'manufacturer', 'manufacturing', 'manuscript', 'manybody', 'map', 'maple', 'mapped', 'mapping', 'mar', 'marathon', 'marble', 'marc', 'marcel', 'march', 'marched', 'marcos', 'marcus', 'margaret', 'margin', 'maria', 'marie', 'marilyn', 'marina', 'marine', 'mariner', 'mario', 'marjorie', 'mark', 'marked', 'market', 'marketing', 'marriage', 'married', 'marshall', 'martha', 'martial', 'martin', 'martino', 'martinus', 'marty', 'marx', 'marxist', 'mary', 'maryland', 'mason', 'mass', 'massachusetts', 'massive', 'master', 'masterclasses', 'mastered', 'masterpiece', 'mastery', 'match', 'matched', 'mate', 'mater', 'material', 'materialist', 'materiality', 'maternal', 'math', 'mathematica', 'mathematical', 'mathematically', 'mathematician', 'mathematics', 'mathematisch', 'matrix', 'matt', 'matter', 'matterhorn', 'matthew', 'matthias', 'mature', 'maureen', 'maurice', 'mauricio', 'mauritania', 'mawr', 'max', 'maxplanck', 'maxwell', 'maya', 'maybe', 'mayor', 'mba', 'mccarter', 'mccarthy', 'mccloy', 'mcconnell', 'mccosh', 'mcdonnell', 'mcgill', 'mcgraw', 'mckenna', 'mcknight', 'mclaughlin', 'mclean', 'mcmaster', 'mcphee', 'md', 'mdiv', 'mead', 'meal', 'mean', 'meaning', 'measure', 'measurement', 'measuring', 'mechanic', 'mechanical', 'mechanism', 'medal', 'medalist', 'medias', 'medical', 'medicine', 'medicinein', 'medieval', 'meditation', 'mediterranean', 'medium', 'meet', 'meeting', 'meg', 'megan', 'melanie', 'melanogaster', 'melbourne', 'mellon', 'melrose', 'melvin', 'member', 'membership', 'memoir', 'memorable', 'memorial', 'memory', 'memphis', 'men', 'mendelssohn', 'mental', 'mentally', 'mention', 'mentioned', 'mentor', 'mentored', 'mentoring', 'mercer', 'merchant', 'meredith', 'merely', 'merged', 'merger', 'merit', 'meritorious', 'merkin', 'merrill', 'mershon', 'merton', 'mesa', 'message', 'met', 'metaethics', 'metal', 'metaphor', 'metaphysical', 'metaphysics', 'meteorologist', 'method', 'methodist', 'methodological', 'methodology', 'metric', 'metropolitan', 'mets', 'mexican', 'mexico',

'meyer', 'meyner', 'mfa', 'mia', 'miami', 'michael', 'michel', 'michelle', 'michigan', 'microbiology', 'microorganism', 'microscope', 'microsoft', 'microwave', 'mid', 'midatlantic', 'midcareer', 'middle', 'middlebury', 'middlesex', 'mids', 'midseason', 'midsummer', 'midtown', 'midwest', 'migration', 'mike', 'mikhail', 'milan', 'milanov', 'milbank', 'mile', 'milhaud', 'military', 'milky', 'mille', 'millennium', 'miller', 'million', 'milton', 'milwaukee', 'mina', 'mind', 'mineral', 'minimal', 'minimum', 'mining', 'minister', 'ministerial', 'ministry', 'minneapolis', 'minnesota', 'minor', 'minority', 'minute', 'minuteman', 'mira', 'miri', 'mirror', 'misconception', 'misconduct', 'mishel', 'miss', 'missed', 'missile', 'mission', 'missionary', 'mississippi', 'missouri', 'mistake', 'misunderstanding', 'mit', 'mitch', 'mitchell', 'mixed', 'mixing', 'ml', 'mla', 'mlb', 'mm', 'mn', 'mo', 'mobile', 'mod', 'modal', 'mode', 'model', 'modeled', 'modeling', 'modelling', 'modem', 'moderator', 'modern', 'moderne', 'modernism', 'modernist', 'modernity', 'modest', 'modification', 'modified', 'modular', 'module', 'modulus', 'molecular', 'molecule', 'moment', 'monaco', 'monash', 'monastery', 'mondale', 'monday', 'monde', 'monetary', 'money', 'monica', 'monitor', 'monitored', 'monitoring', 'monk', 'monkey', 'monmouth', 'monograph', 'monroe', 'montclair', 'montel', 'monterey', 'montgomery', 'month', 'monthly', 'montpellier', 'montral', 'montreal', 'monumental', 'moody', 'moon', 'moore', 'moot', 'moral', 'morality', 'morgan', 'mormon', 'morning', 'morocco', 'morpheus', 'morris', 'morrison', 'morristown', 'morse', 'mortgage', 'morton', 'moscow', 'moses', 'moshe', 'mosque', 'moss', 'mother', 'motion', 'motivate', 'motivated', 'motivation', 'motivational', 'motor', 'motorola', 'mount', 'mountain', 'mounted', 'mourning', 'mouse', 'moved', 'movement', 'movie', 'moving', 'mozart', 'mpa', 'mphil', 'mr', 'msc', 'msica', 'msnbc', 'msri', 'mt', 'mtv', 'multicultural', 'multidimensional', 'multidisciplinary', 'multilateral', 'multimedia', 'multimillion', 'multinational', 'multiple', 'multivariable', 'mumbai', 'munich', 'municipal', 'murder', 'murphy', 'murray', 'muscle', 'muse', 'museum', 'music', 'musica', 'musicae', 'musical', 'musician', 'musicological', 'musicologist', 'musicology', 'musique', 'muslim', 'musthave', 'mutant', 'mutation', 'mutual', 'mvp', 'mystery', 'mystical', 'myth', 'n', 'na', 'naacp', 'nacional', 'naked', 'nam', 'named', 'nancy', 'nanjing', 'nanos', 'nanotechnology', 'naomi', 'naples', 'narendra', 'narrative', 'narrow', 'narrowing', 'nasa', 'nathan', 'nathaniel', 'nation', 'national', 'nationale', 'nationalism', 'nationality', 'nationally', 'nationwide', 'native', 'nato', 'natural', 'naturalism', 'naturalistic', 'naturally', 'nature', 'naval', 'navy', 'nazi', 'nba', 'nbc', 'nbcs', 'nbody', 'nc', 'ncaa', 'ncr', 'nd', 'ne', 'nea', 'neal', 'near', 'nearly', 'nebraska', 'nebula', 'nec', 'necessary', 'necessity', 'need', 'needed', 'negative', 'negotiate', 'negotiating', 'negotiation', 'negotiator', 'neh', 'neighborhood', 'neil', 'nelson', 'neoclassical', 'nephew', 'nervous', 'net', 'netherlands', 'netscape', 'network', 'networking', 'neumann', 'neural', 'neurobiological', 'neurobiology', 'neurological', 'neurologist', 'neuron', 'neurophysiology', 'neuroscience', 'neutral', 'neutron', 'neuwirth', 'nevada', 'new', 'newark', 'newcastle', 'newly', 'newman', 'newport', 'news', 'newsday', 'newsgroup', 'newshour', 'newspaper', 'newsweek', 'newton', 'nfl', 'ngo', 'nhl', 'nica', 'nicaragua', 'nicaraguan', 'nice', 'nicholas', 'nichols', 'nick', 'nickname', 'nicknamed', 'nicole', 'niels', 'nielsen', 'nieman', 'nietzsche', 'nigeria', 'night', 'nightline', 'nightly', 'nih', 'nike', 'nikos', 'ninemonth', 'nineteen', 'nineteenth', 'nineteenthcentury', 'ninety', 'ninth', 'nist', 'nitze', 'nixon', 'nj', 'noam', 'nobel', 'noble', 'noether', 'nominated', 'nomination', 'nominee', 'non', 'nonexecutive', 'nonfiction', 'nongovernmental', 'nonlinear', 'nonnuclear', 'nono', 'nonpartisan', 'nonpro

fit', 'nonresident', 'nonstandard', 'nontrivial', 'noon', 'noonan', 'norfolk', 'normal', 'normale', 'normally', 'norman', 'normandy', 'normative', 'normativity', 'north', 'northeast', 'northeastern', 'northern', 'northfield', 'northwest', 'northwestern', 'norton', 'norway', 'norwegian', 'norwich', 'notable', 'notably', 'note', 'noted', 'noteworthy', 'notice', 'notion', 'notre', 'nova', 'novel', 'novelist', 'novella', 'november', 'npr', 'nrc', 'nsf', 'nuclear', 'nuffield', 'number', 'numbered', 'numerical', 'numerous', 'nuremberg', 'nurse', 'nutley', 'ny', 'nyc', 'nyse', 'nyu', 'nyus', 'o', 'oak', 'oakland', 'oakley', 'oates', 'obama', 'obe', 'oberlin', 'object', 'objective', 'objectivity', 'oboe', 'obrien', 'observation', 'observational', 'observatory', 'observed', 'observer', 'observing', 'observor', 'obsession', 'obtain', 'obtained', 'obtaining', 'oc', 'occasion', 'occasional', 'occasionally', 'occidental', 'occupation', 'occupational', 'occupied', 'occupy', 'occur', 'occurred', 'ocean', 'oceanography', 'oconnell', 'october', 'odd', 'odyssey', 'oecd', 'offbroadway', 'offense', 'offensive', 'offer', 'offered', 'offering', 'office', 'officer', 'official', 'officially', 'offseason', 'offshore', 'oh', 'ohio', 'oil', 'oklahoma', 'old', 'older', 'oldest', 'oligopoly', 'olin', 'olive', 'oliver', 'olson', 'olympiad', 'olympic', 'olympics', 'omaha', 'omitted', 'onassis', 'oncology', 'onegin', 'oneill', 'oneman', 'oneyear', 'ongoing', 'online', 'onscreen', 'onstage', 'ontario', 'ontology', 'onward', 'onwards', 'oped', 'opeds', 'open', 'opened', 'opening', 'openly', 'opera', 'operate', 'operates', 'operatic', 'operating', 'operation', 'operator', 'opinion', 'oppenheimer', 'opponent', 'opportunity', 'opposed', 'opposing', 'opposite', 'opposition', 'oprah', 'optic', 'optical', 'optimal', 'optimization', 'option', 'oracle', 'orange', 'orbis', 'orbit', 'orbital', 'orchestra', 'orchestral', 'orchestrating', 'ordained', 'order', 'ordering', 'ordinary', 'ordination', 'oregon', 'oreilly', 'organ', 'organic', 'organisation', 'organisational', 'organised', 'organism', 'organization', 'organizational', 'organize', 'organized', 'organizer', 'organizing', 'oriental', 'orientalism', 'orientation', 'origin', 'original', 'originally', 'originated', 'originating', 'originator', 'orlando', 'orleans', 'orley', 'ornithologist', 'orphan', 'orr', 'orsay', 'orthodox', 'orthodoxy', 'ortiz', 'orwell', 'osaka', 'osander', 'oslo', 'ostensibly', 'oster', 'ostriker', 'oswald', 'otago', 'ottawa', 'otto', 'ottoman', 'outbreak', 'outcome', 'outfielder', 'outflow', 'outgrowth', 'outlet', 'outline', 'outlook', 'output', 'outraged', 'outrageous', 'outreach', 'outside', 'outsider', 'outspoken', 'outstanding', 'overall', 'overcome', 'overdrive', 'oversaw', 'overseas', 'overseeing', 'oversees', 'overture', 'owe', 'owen', 'owing', 'owl', 'owned', 'owner', 'owning', 'owns', 'oxford', 'oxide', 'ozone', 'p', 'pa', 'pablo', 'pac', 'pace', 'pacer', 'pacific', 'package', 'packard', 'padic', 'padre', 'page', 'paid', 'paint', 'painted', 'painter', 'painting', 'pair', 'pakistan', 'palace', 'pale', 'palestine', 'palestinian', 'palm', 'palo', 'pam', 'pan', 'panama', 'panel', 'paper', 'paperback', 'parachutist', 'paradigm', 'paradise', 'paradox', 'parallel', 'parameter', 'parc', 'pardee', 'parent', 'paris', 'parish', 'parissud', 'park', 'parker', 'parliament', 'parliamentary', 'parson', 'partial', 'partially', 'participant', 'participate', 'participated', 'participating', 'participation', 'particle', 'particular', 'particularly', 'partisan', 'partitioning', 'partly', 'partner', 'partnered', 'partnering', 'partnership', 'parttime', 'party', 'partyin', 'pas', 'pasadena', 'pascal', 'passage', 'passaic', 'passed', 'passing', 'passion', 'passionate', 'past', 'pastor', 'pastoral', 'pastored', 'pat', 'patent', 'paternal', 'paterson', 'path', 'pathbreaking', 'pathogenesis', 'pathway', 'patient', 'patricia', 'patrick', 'patriot', 'patrol', 'patron', 'pattern', 'paul', 'pauline', 'paulo', 'pavilion', 'pawlenty', 'pay', 'payn

e', 'pb', 'pc', 'peabody', 'peace', 'peak', 'pearl', 'pedagogy', 'peer', 'peerreviewed', 'peg', 'peking', 'pen', 'penguin', 'peninsula', 'penn', 'pennsylvania', 'pension', 'pentagon', 'pentecostal', 'people', 'perceived', 'percent', 'percentage', 'perception', 'percussion', 'percussionist', 'percy', 'perfect', 'perform', 'performance', 'performancebased', 'performed', 'performer', 'performing', 'performs', 'perimeter', 'period', 'periodical', 'perkins', 'permanent', 'permission', 'permit', 'permitted', 'perpich', 'perry', 'persian', 'persistence', 'person', 'personal', 'personality', 'personally', 'personnel', 'perspective', 'persuade', 'persuaded', 'pertaining', 'peru', 'perugia', 'peruvian', 'pet', 'pete', 'peter', 'petersburg', 'peterson', 'petition', 'petra', 'petraeus', 'petroleum', 'pew', 'pfizer', 'ph', 'pharmaceutical', 'phase', 'phd', 'phelps', 'phenomenological', 'phenomenon', 'phi', 'phil', 'philadelphia', 'philanthropic', 'philanthropist', 'philanthropy', 'philharmonic', 'philip', 'philippe', 'philippine', 'phillip', 'phillips', 'philology', 'philosopher', 'philosophical', 'philosophy', 'phoenix', 'phone', 'photo', 'photograph', 'photographer', 'photographic', 'photography', 'photonics', 'physic', 'physical', 'physician', 'physicist', 'physicshe', 'physicsin', 'physikalische', 'physiology', 'physique', 'pianist', 'piano', 'picasso', 'pick', 'picked', 'picture', 'piece', 'pierre', 'pillar', 'pillow', 'pilot', 'pine', 'pinyin', 'pioneer', 'pioneered', 'pioneering', 'pisa', 'pitch', 'pitcher', 'pitching', 'pittsburgh', 'pivotal', 'place', 'placed', 'placement', 'placing', 'plainfield', 'plainsboro', 'plaintiff', 'plan', 'planck', 'plane', 'planet', 'planetary', 'planned', 'planner', 'planning', 'plant', 'plantation', 'plasma', 'plate', 'platform', 'platinum', 'plato', 'play', 'played', 'player', 'playful', 'playhouse', 'playing', 'playoff', 'playscript', 'playwright', 'playwriting', 'plaza', 'plc', 'pleasure', 'plenary', 'plurality', 'plus', 'pm', 'po', 'poe', 'poem', 'poet', 'poetic', 'poetics', 'poetry', 'poincar', 'point', 'pointed', 'pointing', 'poisoning', 'poland', 'police', 'policy', 'policyhe', 'policymakers', 'polish', 'political', 'politically', 'politicalmilitary', 'politician', 'politics', 'polk', 'poll', 'pollution', 'polo', 'polymer', 'polynomial', 'polyphony', 'polytechnic', 'polytechnique', 'polytopes', 'pomona', 'pompeu', 'pompidou', 'pontifical', 'poor', 'pop', 'pope', 'popular', 'popularizing', 'population', 'porphyrio', 'port', 'portauprince', 'porter', 'portfolio', 'portion', 'portland', 'portrait', 'portrayed', 'portugal', 'portuguese', 'pose', 'posed', 'posited', 'position', 'positive', 'possibility', 'possible', 'possibly', 'post', 'postcard', 'postcolonial', 'postconflict', 'postdoc', 'postdoctoral', 'posted', 'poster', 'postgraduate', 'posting', 'postmodern', 'postscript', 'postseason', 'postshe', 'postsoviet', 'postwar', 'potential', 'potentially', 'potomac', 'potter', 'pound', 'pour', 'poverty', 'power', 'powerful', 'pp', 'pr', 'practical', 'practice', 'practiced', 'practicein', 'practicing', 'practitioner', 'praeger', 'pragmatic', 'pragmatism', 'prague', 'prairie', 'praise', 'praised', 'preacher', 'preaching', 'precedent', 'precise', 'precisely', 'predecessor', 'predicted', 'prediction', 'predominantly', 'preface', 'preference', 'preliminary', 'prelude', 'premier', 'premiere', 'premiered', 'premodern', 'prentice', 'prep', 'preparatory', 'prepared', 'preparing', 'presbyterian', 'presbytery', 'preseason', 'presence', 'present', 'presentation', 'presentday', 'presented', 'presenting', 'presently', 'preserved', 'preserving', 'presidency', 'president', 'presidential', 'presiding', 'press', 'presshe', 'pressin', 'pressure', 'prestigious', 'presumed', 'pretoria', 'pretty', 'prevent', 'prevented', 'prevention', 'previous', 'previously', 'price', 'priest', 'primarily', 'primary', 'prime', 'primetime', 'prince', 'princess', 'princeton', 'principal', 'principally', 'principle', 'print', 'printing', 'prin

tmaker', 'prior', 'priori', 'priority', 'prism', 'prison', 'prisoner',
 'privacy', 'private', 'privately', 'privilege', 'privy', 'prix', 'priz
e', 'prizein', 'prizewinning', 'pro', 'proactive', 'probability', 'prob
ably', 'probe', 'problem', 'procedure', 'proceeding', 'process', 'proce
ssing', 'proclamation', 'prodigy', 'produce', 'produced', 'producer',
 'producing', 'product', 'production', 'productive', 'productivity', 'p
rof', 'profession', 'professional', 'professionalism', 'professionall
y', 'professor', 'professorial', 'professorship', 'profile', 'profile
d', 'profit', 'profound', 'profoundly', 'program', 'programme', 'progra
mmed', 'programmer', 'programming', 'progress', 'progressive', 'projec
t', 'projection', 'projectsin', 'prokofiev', 'proliferation', 'prolifi
c', 'prominence', 'prominent', 'promise', 'promised', 'promising', 'pro
montory', 'promote', 'promoted', 'promoter', 'promotes', 'promoting',
 'promotion', 'pronounced', 'pronunciation', 'proof', 'propaganda', 'pr
opagation', 'properly', 'property', 'prophecy', 'proponent', 'proportio
n', 'proposal', 'proposed', 'proposing', 'prosecuted', 'prosecution',
 'prosecutor', 'prospect', 'prostitution', 'protagonist', 'protect', 'p
rotecting', 'protection', 'protein', 'protest', 'protestant', 'protoco
l', 'prototype', 'prove', 'proved', 'proven', 'provide', 'provided', 'p
rovidence', 'provider', 'provides', 'providing', 'province', 'proving',
 'provisional', 'provocative', 'provoked', 'provost', 'prowess', 'pry',
 'psa', 'psalm', 'pseudonym', 'psychiatric', 'psychiatrist', 'psychiatr
y', 'psychoanalysis', 'psychoanalytic', 'psychological', 'psychologis
t', 'psychology', 'psychotherapist', 'public', 'publication', 'publicat
ionshe', 'publicity', 'publicized', 'publicly', 'publish', 'published',
 'publisher', 'publishes', 'publishing', 'puerto', 'pulitzer', 'pulse',
 'punctuated', 'punishment', 'pupil', 'purchase', 'purchased', 'purdu
e', 'pure', 'purpose', 'pursue', 'pursued', 'pursuing', 'pursuit', 'pus
h', 'pushed', 'pussycat', 'putnam', 'puzzle', 'qatar', 'qualification',
 'qualified', 'qualifying', 'qualitative', 'quality', 'quantitative', 'q
uantization', 'quantum', 'quarter', 'quarterback', 'quarterly', 'quarte
t', 'quasar', 'qubec', 'quebec', 'queen', 'queensland', 'queer', 'ques
t', 'question', 'questionable', 'questioning', 'quick', 'quickly', 'qui
et', 'quine', 'quite', 'quorum', 'quoted', 'r', 'rabbi', 'rabbinical',
 'rabin', 'race', 'raced', 'rachel', 'racial', 'racing', 'racist', 'rad
cliffe', 'radiation', 'radical', 'radio', 'radiotelevision', 'rafael',
 'ragdale', 'rai', 'railroad', 'rain', 'rainbow', 'rainer', 'raise', 'r
aised', 'raising', 'raleigh', 'ralph', 'ralston', 'ram', 'raman', 'rama
nujan', 'rambert', 'ramey', 'ran', 'rand', 'randall', 'random', 'rand
y', 'range', 'ranged', 'ranger', 'ranging', 'rank', 'ranked', 'rankin
g', 'raoul', 'rape', 'raphael', 'rapid', 'rapidly', 'rappaport', 'rar
e', 'raritan', 'rat', 'rate', 'rated', 'rating', 'ratio', 'rational',
 'rationale', 'rationality', 'rauch', 'raven', 'rawls', 'ray', 'raymon
d', 'rb', 'rca', 'rd', 'reach', 'reached', 'reaching', 'reaction', 'rea
d', 'readable', 'reader', 'reading', 'reagan', 'real', 'realism', 'real
ist', 'realistic', 'reality', 'realized', 'reallife', 'really', 'real
m', 'reappointed', 'reason', 'reasoning', 'rebecca', 'rebel', 'rebirt
h', 'rebound', 'rebuild', 'recall', 'receive', 'received', 'receiver',
 'receiving', 'recent', 'recently', 'reception', 'recherche', 'recipe',
 'recipient', 'recital', 'reckless', 'recognised', 'recognition', 'recog
nize', 'recognized', 'recognizes', 'recollection', 'recommendation', 'r
ecommended', 'reconciliation', 'reconstructed', 'reconstruction', 'reco
rd', 'recordbreaking', 'recorded', 'recording', 'recover', 'recreatio
n', 'recruited', 'recruitment', 'rector', 'red', 'redevelop', 'redevelo
pment', 'redgrave', 'rediscovered', 'redistricting', 'redskin', 'reduc
e', 'reduced', 'reducing', 'reduction', 'reed', 'reelected', 'reelectio
n', 'refer', 'referee', 'refereed', 'reference', 'referenced', 'referen

dum', 'referred', 'refers', 'refinement', 'reflected', 'reflection', 'r
eflects', 'reform', 'reformed', 'reforming', 'refugee', 'refusal', 'ref
used', 'regard', 'regarded', 'regarding', 'regardless', 'regatta', 'reg
eneration', 'regent', 'regime', 'regimehis', 'region', 'regional', 'reg
is', 'register', 'registered', 'regular', 'regularly', 'regulation', 'r
egulatory', 'rehearsed', 'reich', 'reinhard', 'reischauer', 'reissued',
'reject', 'rejected', 'rejecting', 'rejection', 'rejoined', 'relate',
 'related', 'relating', 'relation', 'relational', 'relationship', 'rela
tive', 'relatively', 'relativistic', 'relativity', 'release', 'release
d', 'relevant', 'reliable', 'reliance', 'relief', 'relies', 'relieved',
'religion', 'religious', 'remain', 'remainder', 'remained', 'remainin
g', 'remains', 'remarkable', 'remarked', 'remembered', 'remembrance',
 'remote', 'removal', 'removed', 'removing', 'ren', 'renaissance', 'ren
amed', 'renee', 'renegade', 'renewable', 'renewal', 'renewed', 'renorma
lization', 'renovation', 'renown', 'renowned', 'rensselaer', 'rent', 'r
ep', 'repair', 'reparation', 'repeatedly', 'repertoire', 'repertory',
 'replace', 'replaced', 'replacement', 'replacing', 'replication', 'rep
ly', 'report', 'reported', 'reportedly', 'reporter', 'reporting', 'repo
rtingin', 'represent', 'representation', 'representative', 'represente
d', 'representing', 'represents', 'reprinted', 'reproductive', 'republi
c', 'republican', 'reputation', 'request', 'requiem', 'require', 'requi
red', 'requiring', 'rer', 'rereleased', 'rescue', 'research', 'research
ed', 'researcher', 'researching', 'resembled', 'resentment', 'reserve',
'reside', 'residence', 'residency', 'resident', 'residential', 'reside
s', 'residing', 'resign', 'resignation', 'resigned', 'resistance', 'res
isting', 'resolution', 'resolve', 'resonance', 'resonant', 'resonate',
 'resonating', 'resort', 'resource', 'respect', 'respectable', 'respect
ed', 'respectively', 'respond', 'responded', 'responding', 'response',
 'responsibility', 'responsible', 'rest', 'restaurant', 'restless', 're
storation', 'restore', 'restored', 'restoring', 'restricted', 'restrict
ion', 'result', 'resulted', 'resulting', 'resumed', 'resurrection', 're
tailer', 'retained', 'rethinking', 'retina', 'retired', 'retirement',
 'retiring', 'retrospective', 'return', 'returned', 'returning', 'reuni
on', 'reunited', 'reuters', 'rev', 'reveal', 'revealed', 'revealing',
 'reveals', 'revelation', 'revenge', 'revenue', 'revered', 'reverend',
 'reverse', 'review', 'reviewed', 'reviewer', 'reviewing', 'revised',
 'revision', 'revisionist', 'revitalization', 'revolt', 'revolution',
 'revolutionary', 'revue', 'reynolds', 'rhetoric', 'rhetorical', 'rhod
e', 'rhodes', 'rhythm', 'rhythmic', 'riba', 'rica', 'rican', 'ricardo',
'ricci', 'rice', 'rich', 'richard', 'richards', 'richardson', 'richly',
'richmond', 'rick', 'rico', 'riddle', 'ride', 'rider', 'ridge', 'ridin
g', 'riemann', 'riemannian', 'rifkind', 'rift', 'right', 'righthanded',
'rigorous', 'riker', 'riley', 'rim', 'rinde', 'ring', 'rinpoche', 'rinp
oches', 'rio', 'rise', 'rising', 'risk', 'riskaccording', 'rite', 'ritu
al', 'rival', 'rivalry', 'river', 'riverside', 'rizzoli', 'rn', 'road',
'roanoke', 'rob', 'robb', 'robbins', 'robert', 'roberta', 'robertson',
 'robin', 'robinson', 'robot', 'robust', 'rochelle', 'rochester', 'roc
k', 'rockefeller', 'rodham', 'rodney', 'roger', 'rogers', 'roland', 'ro
le', 'roll', 'roller', 'rolling', 'rom', 'roman', 'romance', 'romania',
'romanian', 'romantic', 'romanticism', 'rome', 'romeo', 'romney', 'ro
n', 'ronald', 'rookie', 'room', 'rooney', 'roosevelt', 'root', 'roote
d', 'rosa', 'rose', 'rosemary', 'ross', 'roster', 'rotation', 'roth',
 'rothschild', 'rotterdam', 'rough', 'roughly', 'round', 'roundtable',
 'route', 'routledge', 'row', 'rowe', 'rowed', 'rower', 'rowing', 'ro
y', 'royal', 'royalty', 'rsi', 'rt', 'ru', 'rubin', 'ruby', 'rudolf',
 'rugby', 'rule', 'ruled', 'rumor', 'run', 'runnerup', 'running', 'runw
ay', 'rural', 'rush', 'russell', 'russia', 'russian', 'russianamerica

n', 'rutgers', 'rutgersnewark', 'ruth', 'rutherford', 'ryan', 's', 's
a', 'saban', 'sabbath', 'sabbatical', 'sachs', 'sacramento', 'sacred',
 'sadie', 'safe', 'safety', 'sage', 'said', 'saint', 'sais', 'salad',
 'sale', 'salem', 'salesman', 'salk', 'sally', 'salomon', 'salon', 'sal
t', 'salvador', 'sam', 'samesex', 'sample', 'samuel', 'san', 'sandber
g', 'sandbergs', 'sander', 'sandinista', 'sandra', 'sandy', 'sanford',
 'sang', 'santa', 'santos', 'sara', 'sarah', 'saratoga', 'sarcoma', 'sa
t', 'satellite', 'satirical', 'saul', 'saunders', 'save', 'saving', 'sa
w', 'saxophone', 'saxophonist', 'say', 'saying', 'sb', 'scalar', 'scal
e', 'scandal', 'scarlet', 'scarsdale', 'scattering', 'scenario', 'scen
e', 'schedl', 'schedule', 'scheduled', 'scheme', 'schenectady', 'schizo
phrenia', 'schlesinger', 'schmidt', 'schock', 'scholar', 'scholarinresi
dence', 'scholarly', 'scholarship', 'scholastic', 'school', 'schooled',
 'schoolhe', 'schoolin', 'schooling', 'schuster', 'schwartz', 'schwarz',
 'science', 'sciencehe', 'scienceshe', 'scienceshis', 'sciencetechnolog
y', 'scientific', 'scientifically', 'scientifiques', 'scientist', 'scie
ntisthe', 'scope', 'score', 'scored', 'scorer', 'scoring', 'scot', 'sco
tland', 'scott', 'scottish', 'scout', 'scratch', 'screen', 'screenwrite
r', 'scripps', 'script', 'scripted', 'sculpture', 'scuola', 'sea', 'sea
n', 'search', 'searching', 'season', 'seat', 'seattle', 'sebastian', 's
ecession', 'second', 'secondary', 'secondbest', 'secret', 'secretary',
 'secretarygeneral', 'section', 'sector', 'secular', 'secure', 'secure
d', 'security', 'sedimentary', 'seed', 'seeger', 'seeing', 'seek', 'see
king', 'seen', 'select', 'selected', 'selection', 'selects', 'self', 's
elfdetermination', 'selfknowledge', 'sell', 'seller', 'selling', 'seman
tic', 'semantics', 'semester', 'semiconductor', 'semifinal', 'semifinal
ist', 'seminal', 'seminar', 'seminary', 'seminaryhis', 'semiotics', 'se
miretired', 'senate', 'senator', 'senegal', 'senior', 'sense', 'sensiti
ve', 'sensitivity', 'sensor', 'sent', 'seoul', 'separate', 'separated',
 'separation', 'sept', 'september', 'sequel', 'sequence', 'sequencing',
 'serbia', 'serf', 'serge', 'sergeant', 'sergei', 'series', 'seriousl
y', 'sermon', 'servant', 'serve', 'served', 'server', 'service', 'servi
ng', 'session', 'set', 'seton', 'setting', 'settled', 'settlement', 'se
ttling', 'seung', 'seven', 'seventeenth', 'seventeenthcentury', 'sevent
h', 'seventy', 'sevenyear', 'severe', 'sewanee', 'sex', 'sexual', 'sexu
ality', 'seymour', 'sf', 'shadow', 'shakespeare', 'shandong', 'shane',
 'shanghai', 'shanti', 'shape', 'shaping', 'shapiro', 'share', 'share
d', 'sharing', 'sharon', 'sharp', 'shaw', 'sheaf', 'sheehan', 'sheffiel
d', 'sheila', 'shelby', 'sheldon', 'shell', 'shelley', 'shelter', 'she
n', 'shenandoah', 'shepherd', 'sherbrooke', 'shes', 'shield', 'shift',
 'shifted', 'shii', 'shiite', 'shine', 'ship', 'shipping', 'shirley',
 'shock', 'shoe', 'shop', 'shore', 'shorenstein', 'short', 'shorter',
 'shortlisted', 'shortly', 'shostakovich', 'shot', 'showcase', 'showe
d', 'showing', 'shown', 'shulman', 'shut', 'shutout', 'sibling', 'sieg
e', 'siegel', 'siegfried', 'sight', 'sigma', 'sigmod', 'sigmund', 'sig
n', 'signal', 'signature', 'signed', 'significance', 'significant', 'si
gnificantly', 'silence', 'silent', 'silk', 'silver', 'silverglate', 'si
lverstein', 'similar', 'simon', 'simple', 'simplified', 'simply', 'simu
lation', 'simultaneously', 'sin', 'sinfonietta', 'singapore', 'singer',
 'singing', 'single', 'singlea', 'singled', 'singularity', 'sioux', 'si
r', 'sister', 'site', 'sits', 'sitting', 'situated', 'situation', 'sixt
een', 'sixteenth', 'sixth', 'size', 'sized', 'sj', 'skeptical', 'skepti
cism', 'skill', 'skin', 'skip', 'sky', 'sl', 'slac', 'slam', 'slate',
 'slatecom', 'slavic', 'sleep', 'slightly', 'sloan', 'slovenia', 'slowe
d', 'smale', 'small', 'smaller', 'smart', 'smear', 'smith', 'smithsonia
n', 'smog', 'smoky', 'smu', 'snow', 'snyder', 'soas', 'socalled', 'socc
er', 'social', 'sociales', 'socialist', 'society', 'societyhis', 'socie

tyin', 'societyshe', 'socioeconomic', 'sociological', 'sociologist', 's
ociology', 'socit', 'socrates', 'sofia', 'soft', 'software', 'soho', 's
olar', 'sold', 'soldier', 'sole', 'solid', 'solidarity', 'solidstate',
 'soliton', 'solo', 'soloist', 'solomon', 'solution', 'solve', 'solve
d', 'solving', 'somerset', 'somewhat', 'son', 'sonata', 'song', 'songwr
iter', 'songwriting', 'sonia', 'soon', 'sophie', 'sophisticated', 'soph
ocles', 'sophomore', 'soprano', 'sorbonne', 'soros', 'sort', 'sought',
 'soughtafter', 'soul', 'sound', 'soundtrack', 'source', 'south', 'sout
hampton', 'southeast', 'southeastern', 'southern', 'southwest', 'southw
estern', 'sovereign', 'soviet', 'space', 'spaceshe', 'spacetime', 'spai
n', 'span', 'spanish', 'spanishspeaking', 'spanned', 'spanning', 'spari
ng', 'spark', 'speak', 'speaker', 'speaking', 'speaks', 'spear', 'spear
headed', 'special', 'specialised', 'specialist', 'specialization', 'spe
cialized', 'specializes', 'specializing', 'specialty', 'specie', 'speci
fic', 'specifically', 'specified', 'spectator', 'spectroscopy', 'spectr
um', 'speech', 'speed', 'spell', 'spencer', 'spend', 'spending', 'spen
t', 'sphere', 'spike', 'spill', 'spin', 'spiral', 'spirit', 'spiritua
l', 'spirituality', 'split', 'splitting', 'spogli', 'spoke', 'spoken',
 'spoleto', 'sponsor', 'sponsored', 'sponsorship', 'sport', 'sportsma
n', 'sportsmanship', 'spot', 'spread', 'spring', 'springer', 'springerv
erlag', 'spur', 'spy', 'squad', 'square', 'squibb', 'sr', 'sri', 'ssl',
 'st', 'stability', 'stabilization', 'stable', 'stadium', 'staff', 'stag
e', 'staged', 'stake', 'stakeholder', 'stamford', 'stan', 'stance', 'st
and', 'standard', 'standing', 'stanford', 'stanislav', 'stanley', 'sta
r', 'starkland', 'starledgers', 'starr', 'starred', 'starring', 'star
t', 'started', 'starting', 'startup', 'starwood', 'state', 'stated', 's
tatement', 'stateshis', 'statesin', 'statewide', 'stating', 'station',
 'stationed', 'statistic', 'statistical', 'status', 'staunch', 'stay',
 'stayed', 'steal', 'steering', 'stein', 'stellar', 'stem', 'stemming',
 'step', 'stephanie', 'stephen', 'stepping', 'stereotype', 'sterling',
 'stern', 'steve', 'steven', 'stevens', 'stevenson', 'stewardship', 'st
ewart', 'stint', 'stochastic', 'stock', 'stockhausen', 'stockholm', 'st
one', 'stony', 'stop', 'storage', 'store', 'storieshe', 'story', 'story
telling', 'straight', 'strain', 'strange', 'strangely', 'strasbourg',
 'strategic', 'strategist', 'strategy', 'stratford', 'stratified', 'str
aus', 'streak', 'stream', 'street', 'strengthen', 'strengthening', 'str
ess', 'stretch', 'strict', 'strike', 'striking', 'string', 'strip', 'st
rong', 'stronger', 'strongest', 'strongly', 'struck', 'structural', 'st
ructurally', 'structure', 'structured', 'struggle', 'stuart', 'studen
t', 'studi', 'studied', 'studieshe', 'studiesin', 'studio', 'study', 's
tudying', 'stumbling', 'stuttgart', 'stuyvesant', 'style', 'su', 'subco
mmittee', 'subject', 'submit', 'submitted', 'subsequent', 'subsequentl
y', 'subsidiary', 'subsidy', 'substance', 'substantial', 'substitute',
 'substituted', 'substitution', 'subtle', 'suburb', 'suburban', 'succee
d', 'succeeded', 'succeeding', 'success', 'successful', 'successfully',
 'successor', 'sudden', 'sue', 'suffer', 'suffered', 'suffering', 'suffi
cient', 'sufficiently', 'suggested', 'suggests', 'suicide', 'suit', 'su
ite', 'sullivan', 'sum', 'sumatra', 'summa', 'summarized', 'summarize
s', 'summer', 'summit', 'sun', 'sundance', 'sunday', 'sunset', 'suny',
 'super', 'supercomputer', 'supercomputing', 'superconductivity', 'supe
rconductors', 'supergravity', 'superieure', 'superintendent', 'superio
r', 'supernatural', 'supernova', 'superpower', 'supersonics', 'superstr
ing', 'supersymmetric', 'supersymmetry', 'supervised', 'supervises', 's
upervision', 'supervisor', 'supplement', 'supplied', 'supply', 'suppor
t', 'supported', 'supporter', 'supporting', 'supposedly', 'suppressor',
 'supreme', 'suprieure', 'surface', 'surfer', 'surfing', 'surfside', 'su
rgeon', 'surgery', 'surgical', 'surname', 'surprise', 'surprising', 'su

rrey', 'surrounding', 'surveillance', 'survey', 'survival', 'survivor', 'susan', 'suspect', 'suspended', 'suspension', 'suspicion', 'suspicious', 'sussex', 'sustainable', 'sustained', 'suzanne', 'swan', 'swarthmore', 'swartz', 'swarup', 'sweden', 'swedish', 'sweet', 'swim', 'swimmer', 'swing', 'swiss', 'switch', 'switched', 'switzerland', 'sworn', 'sxsw', 'sydney', 'syllabic', 'sylvia', 'symbol', 'symbolic', 'symmetry', 'symphony', 'symposium', 'symptom', 'syndicated', 'syndrome', 'synod', 'synthesis', 'synthesizer', 'synthetic', 'syracuse', 'systematic', 't', 'tab', 'table', 'taboo', 'tagging', 'taipei', 'taiwan', 'taiwanese', 'takashi', 'taken', 'taking', 'tale', 'talent', 'talk', 'talking', 'tall', 'tallied', 'tam', 'tampa', 'tampere', 'tan', 'tanglewood', 'tank', 'tanz', 'tao', 'tape', 'tapped', 'target', 'targeted', 'tarskis', 'task', 'tasked', 'taste', 'tata', 'tate', 'taubman', 'taught', 'tax', 'taxation', 'taxpayer', 'taylor', 'tbingen', 'teach', 'teacher', 'teacherhe', 'teaching', 'team', 'teamed', 'teammate', 'tech', 'technical', 'technician', 'technique', 'technological', 'technologist', 'technology', 'tectonics', 'ted', 'teenage', 'tehran', 'teitelboim', 'tel', 'telecommunication', 'telegraph', 'telescope', 'televised', 'television', 'tell', 'temperature', 'template', 'temple', 'templeton', 'temporary', 'tend', 'tendency', 'tendered', 'tennessee', 'tennis', 'tenor', 'tension', 'tenure', 'tenured', 'term', 'terminated', 'terminology', 'terrestrial', 'terrier', 'terror', 'terrorism', 'terrorist', 'terry', 'test', 'testament', 'tested', 'testified', 'testimony', 'testing', 'texan', 'texas', 'text', 'textbook', 'textual', 'th', 'thai', 'thailand', 'thanks', 'tharp', 'thcentury', 'theater', 'theatre', 'theatrical', 'theme', 'thengovernor', 'thenpresident', 'theodore', 'theologian', 'theological', 'theology', 'theorem', 'theoretical', 'theoretically', 'theorist', 'theory', 'theoryhe', 'theoryin', 'therapy', 'thermodynamic', 'thesis', 'thessaloniki', 'thing', 'think', 'thinker', 'thinking', 'thirteen', 'thirteenth', 'thirty', 'thirtyfive', 'thm', 'thom', 'thomas', 'thompson', 'thomson', 'thorne', 'thought', 'thousand', 'thread', 'threat', 'threedimensional', 'threepart', 'threetime', 'threeyear', 'thrive', 'throw', 'thse', 'thunder', 'thurgood', 'thurston', 'tibet', 'tibetan', 'ticket', 'tie', 'tiger', 'tight', 'till', 'tim', 'timber', 'time', 'timehe', 'timein', 'timeline', 'times', 'timesharing', 'timothy', 'tiny', 'tissue', 'title', 'titled', 'titus', 'tobias', 'tod', 'today', 'todd', 'togo', 'tokyo', 'told', 'toledo', 'tolerance', 'tom', 'tone', 'tongue', 'toni', 'tony', 'took', 'tool', 'topeka', 'topic', 'topography', 'topological', 'topologist', 'topology', 'topped', 'toronto', 'tort', 'torture', 'tossed', 'total', 'touch', 'touchdown', 'touched', 'tough', 'tour', 'toured', 'touring', 'tournament', 'tower', 'town', 'township', 'toxic', 'tr', 'trace', 'tracing', 'track', 'trade', 'traded', 'trademark', 'trading', 'tradition', 'traditional', 'traditionalist', 'traditionally', 'tragedy', 'tragic', 'trail', 'train', 'trained', 'trainer', 'training', 'transaction', 'transcription', 'transfer', 'transferred', 'transferring', 'transform', 'transformation', 'transformative', 'transformed', 'transforming', 'transit', 'transition', 'transitional', 'transitioned', 'translated', 'translation', 'translator', 'transmission', 'transnational', 'transport', 'transportation', 'trauma', 'travel', 'traveled', 'traveler', 'traveling', 'travelled', 'travelling', 'treasurer', 'treasury', 'treated', 'treating', 'treatise', 'treatment', 'treaty', 'tree', 'tremendous', 'trend', 'trent', 'trenton', 'trevor', 'trial', 'triangle', 'tribal', 'tribune', 'tribute', 'trick', 'tried', 'trieste', 'trigger', 'trilogy', 'trinity', 'trio', 'trip', 'triple', 'triplea', 'triumph', 'triumvirate', 'trombone', 'tropical', 'trouble', 'true', 'truly', 'truman', 'trumpet', 'trust', 'trustee', 'truth', 'try', 'trying', 'tsinghua', 'tu', 'tuck', 'tucker', 'tucson', 'tudes', 'tudo

r', 'tuft', 'tukey', 'tulane', 'tully', 'tulsa', 'tune', 'tunis', 'tur
f', 'turing', 'turkey', 'turkish', 'turn', 'turned', 'turner', 'turnin
g', 'turtle', 'tutelage', 'tutor', 'tuttle', 'tv', 'tweed', 'twelfth',
 'twentieth', 'twentiethcentury', 'twentyfirst', 'twentyfive', 'twentyf
our', 'twice', 'twig', 'twin', 'twitter', 'twodimensional', 'twotime',
 'twovolume', 'twoyear', 'twyla', 'tying', 'tyler', 'type', 'typical',
 'typically', 'typology', 'u', 'uc', 'uci', 'ucla', 'uclas', 'ucsc', 'u
ganda', 'uk', 'ukraine', 'ukrainian', 'ulrich', 'ultimate', 'ultimatel
y', 'uma', 'una', 'unanimous', 'unanimously', 'unc', 'uncertainty', 'un
cle', 'unconstitutional', 'uncovered', 'unctad', 'und', 'undefeated',
 'undergraduate', 'underlying', 'undermined', 'underneath', 'undersecre
tary', 'underserved', 'understand', 'understanding', 'understood', 'und
ertaken', 'undertaking', 'undertook', 'undoubtedly', 'unemployment', 'u
nesco', 'unexpected', 'unicef', 'unified', 'uniform', 'unilever', 'unio
n', 'unique', 'uniquely', 'unit', 'unitary', 'united', 'uniting', 'unit
y', 'univ', 'universal', 'universe', 'universidad', 'universit', 'unive
rsitat', 'universitt', 'university', 'universitya', 'universityaccordin
g', 'universityas', 'universityhe', 'universityhis', 'universityin', 'u
niversityshe', 'unix', 'unknown', 'unlike', 'unofficial', 'unopposed',
 'unprecedented', 'unpublished', 'unrelated', 'unreleased', 'unrest',
 'unseated', 'unstable', 'unsuccessful', 'unsuccessfully', 'unused', 'u
nusual', 'unveiled', 'upcoming', 'update', 'updated', 'upper', 'uppsal
a', 'uproar', 'upset', 'upstate', 'urban', 'urbanachampaign', 'urbanis
m', 'urged', 'usa', 'usafrica', 'usaid', 'usc', 'use', 'used', 'usefu
l', 'user', 'usgs', 'using', 'ussachevsky', 'ussr', 'usually', 'utah',
 'utilitarian', 'utility', 'utilize', 'utrecht', 'uw', 'v', 'vacant',
 'vacated', 'vacation', 'vacuum', 'valedictorian', 'validity', 'valle
y', 'valuable', 'value', 'van', 'vance', 'vancouver', 'vanderbilt', 'va
nity', 'variability', 'variable', 'variant', 'variation', 'variationa
l', 'variety', 'various', 'variously', 'varsity', 'vassar', 'vast', 've
blen', 'vector', 'vega', 'vehicle', 'velocity', 'venezuela', 'venice',
 'ventura', 'venture', 'ventured', 'venue', 'vera', 'verbal', 'verifie
d', 'verifying', 'vermont', 'vernacular', 'vernon', 'versailles', 'vers
ion', 'versus', 'vertebrate', 'vertical', 'veteran', 'veto', 'vi', 'vib
rant', 'vic', 'vicar', 'vice', 'vicechair', 'vicechairman', 'vicechance
llor', 'vicepresident', 'vicious', 'victim', 'victor', 'victoria', 'vic
torian', 'victory', 'video', 'vienna', 'vietnam', 'vietnamese', 'view',
 'viewer', 'viewpoint', 'vikas', 'villa', 'village', 'villanova', 'vince
nt', 'vintage', 'vinyl', 'viola', 'violation', 'violence', 'violent',
 'violin', 'violinist', 'virginia', 'virtual', 'virtuoso', 'virus', 'vi
salia', 'viscosity', 'vision', 'visionary', 'visit', 'visited', 'visiti
ng', 'visitor', 'visual', 'visualization', 'vladimir', 'vocabulary', 'v
ocal', 'vocalist', 'vodka', 'vogue', 'voice', 'vol', 'volume', 'volunta
ry', 'volunteer', 'volunteered', 'von', 'vote', 'voted', 'voter', 'voti
ng', 'vt', 'vulnerability', 'vulnerable', 'w', 'wa', 'wabctv', 'wage',
 'wagner', 'waiver', 'wake', 'walden', 'wale', 'walker', 'walking', 'wa
ll', 'wallach', 'wallpaper', 'walnut', 'walrasian', 'walt', 'walter',
 'wanda', 'wang', 'want', 'wanted', 'war', 'ward', 'wardlaw', 'warehous
e', 'warfare', 'warhol', 'warming', 'warner', 'warrant', 'warren', 'war
rior', 'warsaw', 'wartime', 'washington', 'wasnt', 'waste', 'watch', 'w
atchdog', 'watcher', 'water', 'waterloo', 'watermarking', 'waterston',
 'watkins', 'watson', 'watt', 'wave', 'way', 'wayne', 'weak', 'weaknes
s', 'wealth', 'wealthiest', 'weapon', 'wear', 'wearing', 'weather', 'we
aving', 'web', 'weber', 'website', 'webster', 'wedding', 'weed', 'wee
k', 'weekend', 'weekly', 'wei', 'weight', 'weil', 'weiler', 'weill', 'w
eimar', 'weinberg', 'weiner', 'weisman', 'weissberger', 'welch', 'wel
d', 'welfare', 'wellbeing', 'wellcome', 'wellesley', 'wellington', 'wel

lknown', 'welsh', 'wen', 'wendell', 'wendy', 'went', 'wentworth', 'wern
er', 'wesley', 'wesleyan', 'west', 'westchester', 'western', 'westfiel
d', 'westinghouse', 'westminster', 'westview', 'wharton', 'wheeler', 'w
hite', 'whitehead', 'whitman', 'whitney', 'wide', 'widely', 'widener',
 'wider', 'wideranging', 'widespread', 'wieland', 'wife', 'wightman',
 'wigner', 'wild', 'wilderness', 'wildlife', 'wiley', 'wilf', 'wilfs',
 'wilkes', 'willa', 'willard', 'willem', 'william', 'williams', 'willia
msburg', 'williamstown', 'willie', 'willis', 'willow', 'wilmer', 'wilmi
ngton', 'wilson', 'wimbledon', 'win', 'wind', 'window', 'wine', 'winer
y', 'winfrey', 'wing', 'winner', 'winning', 'winston', 'winstonsalem',
 'winter', 'wired', 'wireless', 'wiring', 'wisconsin', 'wisconsinmadiso
n', 'wisdom', 'wise', 'wiseman', 'wish', 'wished', 'wissenschaften', 'w
issenschaftskolleg', 'witch', 'withdrawal', 'withdrew', 'witness', 'wit
ten', 'wittgenstein', 'witty', 'wizard', 'wnbctv', 'wolf', 'wolff', 'wo
lfgang', 'wolfson', 'woman', 'won', 'wonder', 'wonderful', 'wong', 'woo
d', 'woodrow', 'woodward', 'word', 'wordplay', 'work', 'worked', 'worke
r', 'workforce', 'working', 'workingclass', 'workplace', 'workshop', 'w
orld', 'worldin', 'worldshe', 'worldviews', 'worldwide', 'worse', 'wors
hip', 'worst', 'worth', 'wounded', 'wrecking', 'wright', 'write', 'writ
er', 'writerinresidence', 'writes', 'writing', 'written', 'wrong', 'wro
te', 'wrought', 'ww', 'wyoming', 'x', 'xavier', 'xenakis', 'xerox', 'x
i', 'xiii', 'xml', 'xray', 'y', 'yaddo', 'yale', 'yalsa', 'yang', 'yang
mills', 'yankee', 'yard', 'yates', 'year', 'yearlong', 'yearly', 'yearo
ld', 'yearsin', 'yeast', 'yehoshua', 'yes', 'yeshiva', 'yield', 'yogi',
 'yom', 'yonkers', 'york', 'yorker', 'yorkin', 'yorktown', 'young', 'you
nger', 'youngest', 'youth', 'yoyo', 'yrj', 'yugoslavia', 'yumtong', 'yu
ri', 'z', 'zagreb', 'zakaria', 'zambia', 'zanzibar', 'zbigniew', 'zeala
nd', 'zen', 'zero', 'zeta', 'zhu', 'zimmer', 'zion', 'zionist', 'zone',
 'zoology', 'zrich', 'zurich']

There are 812 individuals and 8,034 unique words.

```
In [13]: # Most frequent 10 words
         X_mat = X.toarray()
         count = X_mat.sum(axis=0)
         top10_idx = count.argsort()[-10:].tolist()[::-1]
         print("Below are the 10 most frequent words, listed in descending orde
         r:")
         print(np.array(vocab)[top10_idx].tolist())
```

Below are the 10 most frequent words, listed in descending order:
['university', 'wa', 'ha', 'new', 'princeton', 'american', 'professor',
'school', 'born', 'science']

```
In [14]:  # Quantiles
          quantiles = [0, 0.25, 0.5, 0.75, 1]

          distribution = [np.quantile(count, q=quantile) for quantile in quantiles
          ]
          distribution = {"Quantile": [quantile * 100 for quantile in quantiles],
                          "Value": distribution}
          distribution = pd.DataFrame(distribution)
          print("The quantiles of the word counts are:")
          distribution
```

The quantiles of the word counts are:

Out[14]:

	Quantile	Value
0	0.0	2.0
1	25.0	2.0
2	50.0	5.0
3	75.0	11.0
4	100.0	3129.0

Question 2.2)

```
In [15]:  # TF-IDF
          transformer = TfidfTransformer()
          tfidf = transformer.fit_transform(X)
```

```
In [16]:  # Most frequent 10 words
          tfidf_mat = tfidf.toarray()
          tfidf_weight = tfidf_mat.sum(axis=0)
          top10_idx = tfidf_weight.argsort()[-10:].tolist()[::-1]
          print("Below are the 10 most frequent words after applying TF-IDF, liste
          d in descending order:")
          print(np.array(vocab)[top10_idx].tolist())
```

Below are the 10 most frequent words after applying TF-IDF, listed in d
escending order:
['university', 'wa', 'ha', 'new', 'professor', 'american', 'science',
'book', 'research', 'study']

Question 2.3)

```
In [17]:  # Remove stop words
          X_noStops, vocab_noStops = remove_stopwords(X, vocab)
          print("There are " + str(len(vocab_noStops)) + " words remaining.")
```

There are 7818 words remaining.

```
In [18]:   # Most frequent 10 words
           X_noStops_mat = X_noStops.toarray()
           noStops_count = X_noStops_mat.sum(axis=0)
           top10_idx = noStops_count.argsort()[-10:].tolist()[::-1]
           print("Below are the 10 most frequent words after removing stop words, l
           isted in descending order:")
           print(np.array(vocab_noStops)[top10_idx].tolist())
```

Below are the 10 most frequent words after removing stop words, listed
in descending order:
['law', 'team', 'physic', 'economics', 'philosophy', 'jersey', 'mathema
tics', 'columbia', 'economic', 'america']

Question 2.4)

```
In [19]:   # Text for Ben Bernanke
           bernanke = wikipedia[wikipedia.name=="Ben Bernanke"].text
           bernanke = np.array(bernanke)
           print(bernanke)
```

['ben shalom bernanke brnki brnangkee born december 13 1953 is an ameri
can economist at the brookings institution who served two terms as chai
rman of the federal reserve the central bank of the united states from
2006 to 2014 during his tenure as chairman bernanke oversaw the federal
reserves response to the late2000s financial crisisbefore becoming fede
ral reserve chairman bernanke was a tenured professor at princeton univ
ersity and chaired the department of economics there from 1996 to septe
mber 2002 when he went on public service leavefrom 2002 until 2005 he w
as a member of the board of governors of the federal reserve system pro
posed the bernanke doctrine and first discussed the great moderation th
e theory that traditional business cycles have declined in volatility i
n recent decades through structural changes that have occurred in the i
nternational economy particularly increases in the economic stability o
f developing nations diminishing the influence of macroeconomic monetar
y and fiscal policybernanke then served as chairman of president george
w bushs council of economic advisers before president bush nominated hi
m to succeed alan greenspan as chairman of the united states federal re
serve his first term began february 1 2006 bernanke was confirmed for a
second term as chairman on january 28 2010 after being renominated by p
resident barack obama his second term ended february 1 2014 when he was
succeeded by janet yellen']

```
In [20]:   # Words with 10 highest weights for Ben Bernanke
           bernanke_idx = wikipedia.index[wikipedia.name=="Ben Bernanke"].tolist()
           bernanke_tfidf = tfidf_mat[bernanke_idx, :]
           bernanke_weight = bernanke_tfidf.sum(axis=0)
           bernanke_top10_idx = bernanke_weight.argsort()[-10:].tolist()[::-1]
           bernanke_tfidf_vocab = np.array(vocab)[bernanke_top10_idx].tolist()
           print(bernanke_tfidf_vocab)
```

['bernanke', 'reserve', 'chairman', 'federal', 'term', 'bush', 'preside
nt', 'february', 'economic', 'janet']

```
In [21]:  # Most frequent 10 words for Ben Bernanke after removing stop words
          bernanke_noStops = X_noStops_mat[bernanke_idx, :]
          bernanke_count = bernanke_noStops.sum(axis=0)
          bernanke_top10_idx = bernanke_count.argsort()[-10:].tolist()[::-1]
          bernanke_nostops_vocab = np.array(vocab_noStops)[bernanke_top10_idx].tol
          ist()
          print(bernanke_nostops_vocab)

          ['chairman', 'federal', 'bernanke', 'reserve', 'economic', 'february',
          'second', 'bush', 'economy', 'alan']
```

```
In [22]:  # Similarity between the two methods
          i = list(set(bernanke_tfidf_vocab) & set(bernanke_nostops_vocab))
          u = list(set(bernanke_tfidf_vocab) | set(bernanke_nostops_vocab))
          j = len(i)/len(u)

          print("Intersect:", i)
          print("\nUnion:", u)
          print("\nJaccard similarity:", j)

          Intersect: ['chairman', 'february', 'federal', 'bernanke', 'bush', 'eco
          nomic', 'reserve']

          Union: ['second', 'february', 'bush', 'president', 'economic', 'econom
          y', 'janet', 'alan', 'chairman', 'federal', 'bernanke', 'term', 'reserv
          e']

          Jaccard similarity: 0.5384615384615384
```

The majority of the most important words obtained from the two methods are similar, with a Jaccard similarity index of 0.5385.

Question 2.5)

```
In [23]:  # Normalize
          tfidf_normalized = normalize(tfidf_mat, axis=1)
```

```
In [24]:  # Cluster using k-means
          np.random.seed(263)
          kmeans = KMeans(n_clusters=8)
          kmeans.fit(tfidf_normalized)
          y_pred = kmeans.fit_predict(tfidf)
```

```
In [25]:  # Cluster counts
          cluster_counts = [str(np.sum(kmeans.labels_==k)) for k in range(8)]
          print("Cluster counts: " + ",".join(cluster_counts))

          Cluster counts: 144,183,170,45,78,22,126,44
```

```python
In [26]: # find individuals assigned to i cluster
         words_clustered = pd.DataFrame({})
         quantiles_clustered = pd.DataFrame({}, index = [0, 25, 50, 75, 100])
         quantiles = [0, 0.25, 0.5, 0.75, 1]

         for i in range(8):
             index = [idx for idx, x in enumerate(y_pred) if x == i]
             cluster_mat = X_noStops_mat[index,:]
             count = cluster_mat.sum(axis = 0)
             top25_idx = count.argsort()[-25:].tolist()[::-1]
             top25_words = np.array(vocab_noStops)[top25_idx].tolist()
             words_clustered["Cluster " + str(i)] = top25_words

             distribution = [np.quantile(count[np.nonzero(count)], q=quantile) fo
         r quantile in quantiles]
             quantiles_clustered["Cluster " + str(i)] = distribution
```

`# Most frequent 25 words for each cluster`
`print("Below are the 25 most frequent words in each cluster, listed in d`
`escending order:")`
`words_clustered`

Below are the 25 most frequent words in each cluster, listed in descending order:

Out[27]:

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	
0	film	physic	philosophy	orchestra	law	architecture	economics	
1	writer	mathematics	church	festival	court	architect	economic	
2	play	computer	literature	composer	district	firm	affair	
3	theatre	mathematical	theological	symphony	jersey	architectural	foreign	
4	writing	engineering	language	jazz	governor	designer	global	
5	artist	theoretical	social	hall	senate	building	ambassador	
6	magazine	mathematician	religion	dance	legal	museum	council	cha
7	museum	quantum	theology	ensemble	firm	urban	relation	
8	poetry	laboratory	oxford	composition	democratic	practice	politics	
9	series	scientist	humanity	performed	republican	bridge	officer	
10	appeared	advanced	modern	london	attorney	founded	position	
11	jersey	paper	teaching	recording	justice	anne	war	p
12	began	contribution	seminary	conservatory	committee	planning	secretary	
13	child	problem	historian	contemporary	elected	columbia	government	
14	television	physicist	culture	chamber	chairman	interior	executive	
15	band	mit	chair	organ	yale	rose	deputy	
16	dance	model	jewish	performance	election	gould	economy	1
17	story	student	fellowship	premiered	judge	urbanism	european	
18	wrote	scientific	english	studied	commission	honor	committee	
19	festival	astronomy	st	concert	federal	traditional	social	
20	written	medal	yale	philharmonic	counsel	magazine	chairman	
21	gallery	elected	america	album	supreme	porphyrio	bank	
22	literature	position	academic	opera	representative	medieval	economist	
23	ballet	developed	student	electronic	lawyer	early	chief	
24	america	postdoctoral	written	record	property	modern	london	

The results from the *k*-means clustering make sense as the words within each cluster are all related to one another. For example, the 25 most frequently occurring words are all related to entertainment in cluster 0 and music in cluster 3. From the data frame above, we can more or less figure out the theme of each cluster.

```
In [28]:  # Quantiles of non-zero word counts
          print("The quantiles of the non-zero word counts in each cluster are:")
          quantiles_clustered
```

The quantiles of the non-zero word counts in each cluster are:

Out[28]:

	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
25	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
50	2.0	2.0	2.0	1.0	2.0	1.0	2.0	2.0
75	4.0	4.0	4.0	3.0	3.0	2.0	4.0	3.0
100	86.0	239.0	187.0	76.0	314.0	88.0	171.0	192.0

From the quantiles, we see that the distribution of the word frequency indeed follows Zipf's Law: most words do not appear very frequently, while some words dominate completely.