

In [4]:

```
pip install xlrd
```

Collecting xlrd

Downloading <https://files.pythonhosted.org/packages/b0/16/63576a1a001752e34bf8ea62e367997530dc553b689356b9879339cf45a4/xlrd-1.2.0-py2.py3-none-any.whl>
 (https://files.pythonhosted.org/packages/b0/16/63576a1a001752e34bf8ea62e367997530dc553b689356b9879339cf45a4/xlrd-1.2.0-py2.py3-none-any.whl) (103kB)

Installing collected packages: xlrd

Successfully installed xlrd-1.2.0

Note: you may need to restart the kernel to use updated packages.

In [7]:

```
import re
import csv
import matplotlib.pyplot as plt

text = "This $string is an $example"
[word for word in text.split() if word.startswith('$') or word.startswith('@)]]
```

Out[7]:

```
['$string', '$example']
```

In [1]:

```
import pandas as pd
import numpy as np

df=pd.read_excel("shuffled_gunlaw.xlsx")
df.columns=['i','date','Tweet','label']
X=df['Tweet']

df.head()
```

Out[1]:

	i	date	Tweet	label
0	0	2019-09-20	@SenSanders Bernie's promoting a 25% federal t...	for
1	1	2019-09-28	RT @McJovy Wonder how fetus lovers defend Khid...	against
2	2	2019-09-23	RT @chronovariance Texas mass shooting survivo...	for
3	3	2019-09-27	The next time you hear an elite or wealthy Dem...	for
4	4	2019-09-24	@OlofsdotterK @RoyaRahmani @NZAmbassadorUS @Ma...	for

In [2]:

```

processed_tweets=[]
for tweet in range(0, len(X)):
    #print (str(X[tweet]))
    #print ("\n\n")
    #[word for word in str(X[tweet]).split() if word.startswith('#') or word.startswith('@')
    processed_tweet=[]
    for word in str(X[tweet]).split():
        if word.startswith('#'):
            processed_tweet.append(word)
    processed_tweets.append(' '.join(processed_tweet))
print (len(processed_tweets))
print (processed_tweets)
df['processed_hashtags']=processed_tweets
df['processed_hashtags'].drop_duplicates()

```

13685

```

['#DrainTheDeepState #FridayFeeling #2020Election #DemocraticDebate #GunCo
ntrol', '', '#NotOneMore #Enough #BullyM...', '#guncontrol #guncontrol', '#G
unViolence #TheAtlanticFest #Ukraine', '#gunviolence', '#2A #2AShallNotBeI
nfringed #GunControl', '#GOPComplicitTraitors #Feels #GOPCorruption #GOPCo
mplicit #GOPCowards #GunControlNow #GunViolence #NRABloodMoney #nraisaterr
oristorganization #MoscowMitchHatesAmerica #MoscowMitchMassacre', '#GunCon
trol #txlege', '', '#libertarian #internetfreedom', '#MarchForOurLives',
'#Libertarian', '#ProtectOurDemocracy #GunViolence #ClimateChange', '#CPC1
9 #Elections2019 #GunControl #cdnpoli #canadaelection2019 #canada', '#GunC
ontrol #EndGunViolence', '#2A #GunControl #Constitution #2AShallNotBeInfri
nged #freedom', '#GunControl #2AShallNotBeInfringed', '#GunControl', '#bal
timore #trump #maga #dem...', '#Share #PlayingNow #Rock #GreenockMusicExplos
ion #GreenockMusicScene #DontGiveMeAGun #NeverAgain #GunControl #GunContro
lNow #Share #PlayingNow #Idwp #GunControlNow #USA #GunLaws #PrayForVegas',
'#MoscowMitch #GunControl', '#LiberalismIsAMentalDisorder #2ndAmendment',
'#2ndAmendment #SecondAmendment', '#gunviolence', '#autism #progressive #c
onservative #liberal #libertarian #green #independent #voluntaryist', '#Sa
turdaySatire #PeoplesVote #MarchForOurLives #StopBrexit', '#AssHat #2ndAme
ndment #American Patriots #DontGiveMeAGun #2ndAmendment', '#ChicagoGunViolence #gun

```

In [28]:

```
df['processed_hashtags'].unique()
```

Out[28]:

```

array(['#DrainTheDeepState #FridayFeeling #2020Election #DemocraticDebate #G
unControl',
      '', '#NotOneMore #Enough #BullyM...', ...,
      '#nswpol #Abortion #ProLife #ProChoice',
      '#ImpeachmentNow #golfer #ThursdayThoughts #journalist #boston #girls
whocode #MarchForOurLives',
      '#ExtremeRiskProtectionOrders #ERPO #RedFlag #GunViolence #GVP #RedFl
agLaw #GVRO'],
      dtype=object)

```

In [30]:

```

j=0
for i in (df['processed_hashtags'].unique()):
    with open ('uqe.csv','a', encoding="utf-8") as res:
        writer=csv.writer(res)
        s="{},{ }\n".format(i,j)
        res.write(s)
        print (s)
        j+=1

```

```

#DrainTheDeepState #FridayFeeling #2020Election #DemocraticDebate #GunCont
rol,0

```

```

,1

```

```

#NotOneMore #Enough #BullyM...,2

```

```

#guncontrol #guncontrol,3

```

```

#GunViolence #TheAtlanticFest #Ukraine,4

```

```

#gunviolence,5

```

```

#2A #2AShallNotBeInfringed #GunControl,6

```

```

#GOPComplicitTraitors #Feels #GOPCorruption #GOPComplicit #GOPCowards #Gun
ControlNow #GunViolence #NRABloodMoney #nraisaterroristorganization #Mosco
wMitchHatesAmerica #MoscowMitchMassacre,7

```

```

#G...

```

In [35]:

```
df4=pd.read_csv('uqe.csv')
df4.columns=['unique_hashtags','index']
df4
```

Out[35]:

	unique_hashtags	index
0	NaN	1
1	#NotOneMore #Enough #BullyM...	2
2	#guncontrol #guncontrol	3
3	#GunViolence #TheAtlanticFest #Ukraine	4
4	#gunviolence	5
5	#2A #2AShallNotBeInfringed #GunControl	6
6	#GOPComplicitTraitors #Feels #GOPCorruption #G...	7
7	#GunControl #txlege	8
8	#libertarian #internetfreedom	9
9	#MarchForOurLives	10
10	#Libertarian	11
11	#ProtectOurDemocracy #GunViolence #ClimateChange	12
12	#CPC19 #Elections2019 #GunControl #cdnpoli #ca...	13
13	#GunControl #EndGunViolence	14
14	#2A #GunControl #Constitution #2AShallNotBeInfr...	15
15	#GunControl #2AShallNotBeInfringed	16
16	#GunControl	17
17	#baltimore #trump #maga #dem...	18
18	#Share #PlayingNow #Rock #GreenockMusicExplosi...	19
19	#MoscowMitch #GunControl	20
20	#LiberalismIsAMentalDisorder #2ndAmendment	21
21	#2ndAmendment #SecondAmendment	22
22	#autism #progressive #conservative #liberal #l...	23
23	#SaturdaySatire #PeoplesVote #MarchForOurLives...	24
24	#AssHat #2ndAmendment #AmericanPatriots #Despi...	25
25	#Chicagogunviolence #gun #GunViolence #savingl...	26
26	#GunControl #2AShallNotBeInfringed #SelfDefens...	27
27	#ProGun #ProLife #BetoEsPuroPedo	28
28	#Ireland #40daysforlife #prolife #abortion	29
29	#NassauCounty #Jake4Na...	30
...
6491	#guncontrol #GunViolence #GunViolencePreventio...	6492
6492	#whistleblower #scam #gunviolence #gunreform #...	6493

	unique_hashtags	index
6493	#texas2a #ar15 #ar15life #ar15build #sbr #arpi...	6494
6494	#guncontrol #WarnerBros #Joker	6495
6495	#GunControl #BanAssaultWeapons #M4A #Education...	6496
6496	#gunban #guncontrol	6497
6497	#con #GlobalWarmingHoax #GunControl	6498
6498	#GunViolence #NationalVoterRegistrationDay #En...	6499
6499	#resistance #hashtag #Impeachment #ImpeachTrum...	6500
6500	#MoscowMitch	6501
6501	#GunControl #Txlege	6502
6502	#MarchForOurLives #Milwaukee	6503
6503	#Prolife #DefundPlannedParenthood	6504
6504	#CFR #GunControl	6505
6505	#Pulse #survivor #BrandonWolf #WESH #speech #C...	6506
6506	#ar15barrels #tacticalkinetics #ar15...	6507
6507	#USA #GunControl #Trump #NRA #StandUpComedy #C...	6508
6508	#GunReformNow #GunViolence #MarchForOurLives	6509
6509	#GunControl #Canada #Gaslighting #elxn43 #cdnpoli	6510
6510	#GunControl #2A #GunControlNow	6511
6511	#capit...	6512
6512	#2A #AR15 #RightToBearArms #2ndAmendment #GunC...	6513
6513	#Racism #GunControl	6514
6514	#libertarian #Marxists #Marxist #Marxist	6515
6515	#guns #ar15 #MassShooting	6516
6516	#gunsKillPeople	6517
6517	#Democrats #Socialism #Guncontrol	6518
6518	#nswpol #Abortion #ProLife #ProChoice	6519
6519	#ImpeachmentNow #golfer #ThursdayThoughts #jou...	6520
6520	#ExtremeRiskProtectionOrders #ERPO #RedFlag #G...	6521

6521 rows × 2 columns

In []:

In [36]:

```
process_tweets=[]
for tweet in range(0, len(X)):
    #print (str(X[tweet]))
    #print ("\n\n")
    #[word for word in str(X[tweet]).split() if word.startswith('#') or word.startswith('@')
    process_tweet=[]
    for word in str(X[tweet]).split():
        if word.startswith('@'):
            process_tweet.append(word)
    process_tweets.append(' '.join(process_tweet))
print (len(process_tweets))
df['Processed_mentions']=process_tweets
df
```

13685

In []:

In []:

In [37]:

#s Freq

```
from sklearn.feature_extraction.text import CountVectorizer
corpus=df['processed_hashtags'].astype('U')
vec = CountVectorizer().fit(corpus)
bag_of_words = vec.transform(corpus)
sum_words = bag_of_words.sum(axis=0)
words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
print (words_freq[:len(df)])
```

[('guncontrol', 3715), ('gunviolence', 1983), ('prolife', 1752), ('ar15', 1507), ('libertarian', 1202), ('2a', 953), ('maga', 790), ('2ndamendment', 661), ('guncontrolnow', 538), ('marchforourlives', 527), ('guns', 518), ('gunsense', 448), ('nra', 435), ('democrats', 337), ('trump', 335), ('usa', 335), ('2adefender', 330), ('liberty', 325), ('tcot', 280), ('tlot', 259), ('freedom', 250), ('gunreformnow', 243), ('colt', 227), ('trump2020', 218), ('resist', 211), ('2ashallnotbeinfringed', 210), ('climatechange', 209), ('gop', 196), ('politics', 192), ('gunrights', 189), ('kag', 181), ('conservative', 167), ('conservatives', 166), ('meme', 148), ('news', 148), ('teaparty', 146), ('abortion', 145), ('kag2020', 143), ('betoorourke', 142), ('gun', 141), ('conservativememes', 141), ('constitution', 140), ('firearms', 138), ('beto2020', 134), ('biggovsucks', 132), ('socialismsucks', 131), ('cdnpoli', 127), ('ethics', 127), ('secondamendment', 126), ('endgunviolence', 119), ('moscowmitch', 115), ('americafirst', 112), ('maga2020', 111), ('ak47', 109), ('government', 108), ('molonlabe', 106), ('prolifefeun', 105), ('walkaway', 102), ('impeachtrump', 102), ('poverty', 100), ('homelessness', 98), ('beto', 97), ('elxn43', 97), ('housing', 94), ('charity', 91), ('homeless', 90), ('america', 90), ('buildthewall', 89), ('wwg1wga', 88), ('progun', 87), ('prochoice', 87), ('abortionismurder', 86), ('antiracism', 82), ('idiot', 82), ('blacklivesmatter', 81), ('11

In [39]:

```
hash=[]  
freq=[]  
count=0  
for i,f in words_freq:  
    print (i,f)  
    count+=1  
    if count<15:  
        hash.append(i)  
        freq.append(f)
```

```
guncontrol 3715  
gunviolence 1983  
prolife 1752  
ar15 1507  
libertarian 1202  
2a 953  
maga 790  
2ndamendment 661  
guncontrolnow 538  
marchforourlives 527  
guns 518  
gunsense 448  
nra 435  
democrats 337  
trump 335  
usa 335  
2adefender 330  
liberty 325  
tcot 280  
lib 250
```

In [40]:

```
print (hash)
```

```
['guncontrol', 'gunviolence', 'prolife', 'ar15', 'libertarian', '2a', 'maga',  
'2ndamendment', 'guncontrolnow', 'marchforourlives', 'guns', 'gunsense',  
'nra', 'democrats']
```

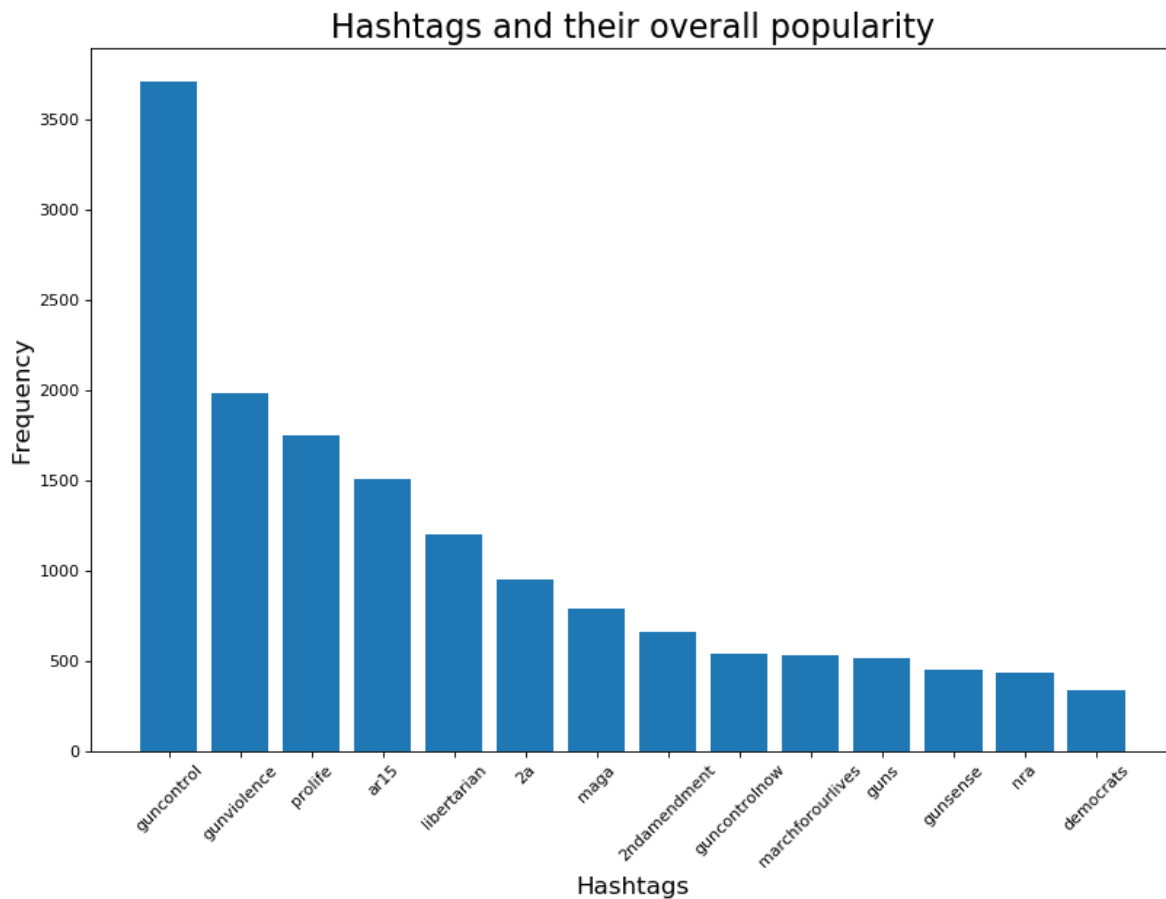
In [41]:

```
print (freq)
```

```
[3715, 1983, 1752, 1507, 1202, 953, 790, 661, 538, 527, 518, 448, 435, 337]
```


In [42]:

```
from matplotlib.pyplot import figure
figure(num=None, figsize=(12, 8), dpi=80, facecolor='w', edgecolor='k')
# this is for plotting purpose
#index = np.arange(len(x))
plt.bar(hash, freq)
plt.xlabel('Hashtags', fontsize=15)
plt.ylabel('Frequency', fontsize=15)
plt.xticks(fontsize=10, rotation=45)
plt.title('Hashtags and their overall popularity', fontsize=20)
plt.show()
```



In [46]:

```
df6=pd.read_csv('f_a2.csv')
df6.columns=['i','date','Tweet','label']

df6.head()
```

Out[46]:

	i	date	Tweet	label
0	2	2019-09-23	rt chronovarience texas mass shooting survivor...	for
1	3	2019-09-27	time hear elite wealthy democrat guncontrol re...	for
2	4	2019-09-24	olofsdotterk royarahmani nzambassadorus mars...	for
3	5	2019-09-25	arizona state representative jen longdon gunvi...	for
4	6	2019-09-20	kamalaharris lot senatemajldr senategop stup...	for

In [47]:

```

import spacy
nlp = spacy.load("en_core_web_sm")
ner={ 'PERSON':0, 'NORP':0, 'FAC':0, 'ORG':0, 'GPE':0, 'LOC':0, 'PRODUCT':0, 'EVENT':0, 'WORK_OF_ART':0, 'LAW':0, 'LANGUAGE':0, 'DATE':0, 'TIME':0, 'PERCENT':0, 'MONEY':0, 'QUANTITY':0, 'ORDINAL':0, 'CARDINAL':0}
PERSON=[]
NORP=[]
FAC=[]
ORG=[]
GPE=[]
LOC=[]
PRODUCT=[]
EVENT=[]
WORK_OF_ART=[]
LAW=[]
LANGUAGE=[]
DATE=[]
TIME=[]
PERCENT=[]
MONEY=[]
QUANTITY=[]
ORDINAL=[]
CARDINAL=[]
for txt in df6['Tweet']:
    ner={ 'PERSON':0, 'NORP':0, 'FAC':0, 'ORG':0, 'GPE':0, 'LOC':0, 'PRODUCT':0, 'EVENT':0, 'WORK_OF_ART':0, 'LAW':0, 'LANGUAGE':0, 'DATE':0, 'TIME':0, 'PERCENT':0, 'MONEY':0, 'QUANTITY':0, 'ORDINAL':0, 'CARDINAL':0}
    doc = nlp(txt)
    for ent in doc.ents:
        ner[ent.label_]=ner[ent.label_]+1
    PERSON.append(ner['PERSON'])
    NORP.append(ner['NORP'])
    FAC.append(ner['FAC'])
    ORG.append(ner['ORG'])
    GPE.append(ner['GPE'])
    LOC.append(ner['LOC'])
    PRODUCT.append(ner['PRODUCT'])
    EVENT.append(ner['EVENT'])
    WORK_OF_ART.append(ner['WORK_OF_ART'])
    LAW.append(ner['LAW'])
    LANGUAGE.append(ner['LANGUAGE'])
    DATE.append(ner['DATE'])
    TIME.append(ner['TIME'])
    PERCENT.append(ner['PERCENT'])
    MONEY.append(ner['MONEY'])
    QUANTITY.append(ner['QUANTITY'])
    ORDINAL.append(ner['ORDINAL'])
    CARDINAL.append(ner['CARDINAL'])
df6['PERSON']=PERSON
df6['NORP']=NORP
df6['FAC']=FAC
df6['ORG']=ORG
df6['GPE']=GPE
df6['LOC']=LOC
df6['PRODUCT']=PRODUCT
df6['EVENT']=EVENT
df6['WORK_OF_ART']=WORK_OF_ART
df6['LAW']=LAW
df6['LANGUAGE']=LANGUAGE
df6['DATE']=DATE
df6['TIME']=TIME
df6['PERCENT']=PERCENT
df6['MONEY']=MONEY

```

```
df6['QUANTITY']=QUANTITY
df6['ORDINAL']=ORDINAL
df6['CARDINAL']=CARDINAL
```

In [48]:

df6

Out[48]:

	i	date	Tweet	label	PERSON	NORP	FAC	ORG	GPE	LOC	...	WORK_
0	2	2019-09-23	rt chronovarience texas mass shooting survivor...	for	0	0	0	0	0	0	...	
1	3	2019-09-27	time hear elite wealthy democrat guncontrol re...	for	0	1	0	0	0	0	...	
2	4	2019-09-24	olofsdotterk royarahmani nzambassadorus mars...	for	0	0	0	0	0	0	...	
3	5	2019-09-25	arizona state representative jen longdon gunvi...	for	0	0	0	0	0	0	...	
4	6	2019-09-20	kamalaharris lot senatemajldr senategop stup...	for	0	0	0	0	0	0	...	
5	7	2019-09-26	ugh straight heart noncomplicitrtraitors feels	for	0	0	0	0	0	0	...	

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []:

In []: