In [2]:	from mutiplicitib.pupicity import import word count in more reported as plant from word-loud smarphrith in import mylorid as plant from word-loud smarphrith in important mylorid from many mylorid from mylorid f
<pre>In [6]: Out[6]:</pre>	Same
Out[7]:	['234 tweets with "loser" I feel sorry for Rosie \'s new partner in love whose parents are devastated at the thought of their daughter being with @Rosiea true loser. — Donald J. Trump (@realDonaldTrump) December 14, 2011 222.', 'tweets with "dumb" or "dummy" You must admit that Bryant Gumbel is one of the dumbest racists around - an arrogant dope with no talent. Failed at CBS etc-why still on TV? — Donal d J. Trump (@realDonaldTrump) August 21, 2013.', '204 tweets with "terrible" I loved beating these two terrible human beings. I would never recommend that anyone use her lawyer, he is a total loser! — Donald J. Trump (@realDonaldTrump) May 23, 2013.', '183 tweets with "stupid" @michellemalkin You were born stupid! — Donald J. Trump (@realDonaldTrump) March 22, 2013', '156 tweets with "weak" There is no longer a Bernie Sanders "political revolution." He is turning out to be a weak and somewhat pathetic figure, wants it all to end! — Donald J. Trump (@realDonaldTrump) July 24, 2016', '17 tweets with "dope" or "dopey" Dopey @Lord_Sugar I\'m worth \$8 billion and you\'re worth peanutswithout my show nobody would even know who you are. — Donald J. Trump (@realDonaldTrump) December 7, 2012', '17 the sets with "dishonest" A dishonest slob of a reporter, who doesn\'t understand my sarcasm when talking about him or his wife, wrote a foolish & boring Trump "hit" — Donald J. Trump (@realDonaldTrump) February 15, 2014 .']  # Joining the list into one string/text text = ''.join(tweets) text
Out[8]:	Heads with "Lose" I feel sorry for Royle N's new partner in lowe whose parents are devestated at the Unuguit of their daughter being with Residen a trump (BrealbonaldTrump) to be better at 1921 222, beets with "dumb" or "dummy" You must distil that were for the dumbers tracits arround -an arround dope with no tale not referented that anyone use her lawyer, he is a total loser! — Bonald J. Trump (BrealbonaldTrump) May 23, 2013. 130 tweets with "studie" given less that the properties of the studies
In [9]: Out[9]:	The analysis of the set item. They should, be weeded out!  **ROPECHALIDA ON CREATE ITEM. Translate(st.maketrans(", ", string.punctuation)) exith arguments (x, y, z) where 'x' and 'y'  ***ara replaced by characters in 'x'.  ***ara replaced by characters in 'x'.  ***ara replaced by characters in 'y'. '2'  ***ara replaced by ch
In [10]: In [11]: Out[11]: In [12]:	#Tokenization from nltk.tokenize import word_tokenize text_tokens = word_tokenize(no_punc_text) print(text_tokens[0:50])  ['234', 'tweets', 'with', 'loser', 'I', 'feel', 'sorry', 'for', 'Rosie', 's', 'new', 'partner', 'in', 'love', 'whose', 'parents', 'are', 'devastated', 'at', 'the', 'thought', 'of', 'their', 'daughter', 'being', 'with', 'Rosiea', 'true', 'loser', '-', 'Donald', 'J', 'Trump', 'realDonaldTrump', 'December', '14', '2011', '222', 'tweets', 'with', 'dumb', 'or', 'd ummy', 'You', 'must', 'admit', 'that', 'Bryant', 'Gumbel', 'is']  #Remove stopwords import nltk from nltk.corpus import stopwords nltk.download('punkt') nltk.download('stopwords')
In [13]:	<pre>my_stop_words = stopwords.words('english') my_stop_words.append('the') no_stop_tokens = [word for word in text_tokens if not word in my_stop_words] print(no_stop_tokens[0:a0])  ['234', 'tweets', 'loser', 'I', 'feel', 'sorry', 'Rosie', 'new', 'partner', 'love', 'whose', 'parents', 'devastated', 'thought', 'daughter', 'Rosiea', 'true', 'loser', '-', 'Donal d', 'J', 'Trump', 'realDonaldTrump', 'December', '14', '2011', '222', 'tweets', 'dumb', 'dummy', 'You', 'must', 'admit', 'Bryant', 'Gumbel', 'one', 'dumbest', 'racists', 'around', 'arrogant'] [nltk_data] Downloading package punkt to [nltk_data] C:\Users\acer\AppData\Roaming\nltk_data [nltk_data] Package punkt is already up-to-date! [nltk_data] Downloading package stopwords to [nltk_data] C:\Users\acer\AppData\Roaming\nltk_data [nltk_data] Package stopwords is already up-to-date!  ///////////////////////////////////</pre>
In [14]:  In [19]:	### State import PorterStemmer  In PorterStemmer()  In PorterStemm
In [23]:	2021-06-24 28:29:12.775556: I tensorliow/stream_executor/cuda/cudart_stub.oc:29] Ignore above cudart derror if you do not have a 6PU set up on your machine. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/platfora/objatinalityso_louder_co:30] Could not loud dynamic library cudart_alloudil; oliveror: cudart_file. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/platfora/objatinalityso_louder_co:30] Could not loud dynamic library cudart_alloudil; oliveror: cudart_file. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/platfora/objatinalityso_louder_co:30] Could not loud dynamic library cudart_alloudil. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/cuda/cudart_stub.cc.29] Ignore above cudart_derror if you do not have a 6PU set up on your machine. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/cuda/cudart_stub.cc.29] Ignore above cudart_derror if you do not have a 6PU set up on your machine. 2021-08-24 28:43:45.755056: W tensorliow/stream_executor/cuda/cudart_stub.cc.29] Ignore above cudart_derror if you do not have a 6PU set up on your machine. 2021-08-24 28:43:45.755026: W tensorliow/stream_executor. 2021-08-24 28:45:45.755026: W tensorliow/stream_executor. 2021-08-24 28:45:45:455026: W tensorliow/stream_executor. 2021-08-24 28:45:45:455026: W tensorliow/stream_executor. 2021-08-24 28:4
In [25]:	doc = nlp(' '.join(no_stop_tokens)) print(doc[0:40])  234 tweets loser I feel sorry Rosie new partner love whose parents devastated thought daughter Rosiea true loser — Donald J Trump realDonaldTrump December 14 2011 222 tweets dumb d ummy You must admit Bryant Gumbel one dumbest racists around arrogant  lemmas = [token.lemma_for token in doc] print(lemmas[0:25])  ['234', 'tweet', 'loser', 'I', 'feel', 'sorry', 'rosie', 'new', 'partner', 'love', 'whose', 'parent', 'devastate', 'think', 'daughter', 'Rosiea', 'true', 'loser', '-', 'Donald', 'J', 'Trump', 'realdonaldtrump', 'December', '14']  Feature Extraction  from sklearn.feature_extraction.text import CountVectorizer vectorizer = CountVectorizer() X = vectorizer.fit_transform(lemmas)
In [27]: In [28]:	print(vectorizer vocabulary.)  (*224.*24.*   thucut.*1 a88.*   thu
In [29]:  In [30]:  In [31]:	[0 0 0 0 0 0] [0 0 0 0 0 0] [print(X.toarray().shape)  (923, 440)  Let's see how can bigrams and trigrams can be included here  vectorizer_ngram_range = CountVectorizer(analyzer='word', ngram_range=(1,3), max_features = 100) bow_matrix_ngram =vectorizer_ngram_range.fit_transform(tweets)  print(vectorizer_ngram_range.get_feature_names()) print(bow_matrix_ngram_toarray()) print(bow_matrix_ngram_toarray()) ['10', '2012', '2013', '2016', 'about', 'all', 'always', 'am', 'an', 'and', 'are', 'at', 'august', 'back', 'bad', 'be', 'billion', 'boring', 'bryant', 'but', 'by', 'can', 'crazy', 'december', 'dishonest', 'don', 'donald', 'donald trump', 'donald trump', 'dopey', 'dumb', 'feel', 'for', 'get', 'has', 'have', 'he', 'if', 'in', 'is', 'isn', 'it', 'just', 'know', 'life', 'loser', 'low', 'may', 'me', 'moron', 'mo', 'no talent', 'not', 'of', 'of the', 'on', 'or', 'pathetic', 'people', 'realdonaldtrump august', 'realdonaldtrump december', 'stupid', 'talent', 'talent', 'tate, 'thet', 'thee', 'thes', 'time', 'trump' realdonaldtrump august', 'reump', 'trump realdonaldtrump august', 'trump realdonaldtrump september', 'tweets implying', 'tweets with', 'very', 'was', 'what', 'when', 'who', 'whose', 'with', 'with no', 'worth', 'would', 'you', 'you are'] [10 0 1 0 1 0] [10 0 1 0 1 0] [10 1 1 0]
In [32]:	Tried
In [33]:	# Import packages import matplotlib.pyplot as plt %matplotlib inline from wordcloud import Wordcloud, STOPWORDS # Define a function to plot word cloud def plot_cloud(wordcloud): # Set figure size plt.figure(figsize=(40, 30)) # Display image plt.imshow(wordcloud) # No axis details plt.axis("off");  # Generate wordcloud stopwords = STOPWORDS stopwords = STOPWORDS stopwords.add("will") wordcloud = Wordcloud(width = 3000, height = 2000, background_color='black', max_words=100,colormap='Set2',stopwords=stopwords).generate(text) # Plot
	Rosie  Whose  The part of the peanutt of the peanut
In [ ]:	

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)
import string # special operations on strings
import spacy # language models