Marijuana Legalization Rhetorical Analysis

Vivian Zhang, Madhav Soni, Nathalie Taing

Research Question

How has rhetoric about the legalization of marijuana changed over time?

Can we use the rhetorical patterns in articles about legalization to predict the legal status of marijuana at the time and region of publication?

Part 1: Data Extraction

Part 2: Data Cleaning & EDA

- Initial visualizations
- Preparding data for classification
- Visualizing data based on status of marijuana
- Word embedding EDA
- Morality and sentiment EDA

Part 3: Classification

- TF-IDF classification
- Document term matrix classification
- Doc2Vec classification
- Morality and sentiment analysis + classification

1: Data Extraction

“marijuanaNEAR/7(legalizor de-criminaliz)”

Extracted .txt file

New Law May Not Turn on 'Pot' Flow for Some Cancer Patients

Author: Anderson, Pat

Publication info: Los Angeles Times (1923-1995) ; Los Angeles, Calif. [Los Angeles, Calif]27 Dec 1979: sg2.

https://search.proquest.com/docview/162537071?accountid=14496

Abstract: In January, when marijuana becomes a legalized medicine, it will be a problem, as researchers see it, to get a large enough supply for the thousands of California patients who will be eligible for treatment.

Links: http://ucelinks.cdlib.org:8888/sfx_local?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Ahnplatimes&atitle=New+Law+May+Not+Turn+on+%27Pot%27+Flow+for+Some+Cancer+Patients&title=Los+Angeles+Times+%281923-1995%29&issn=04583035&date=1979-12-27&volume=&issue=&spage=SG2&au=Anderson%2C+Pat&isbn=&jtitle=Los+Angeles+Times+%281923-1995%29&btitle=&rft_id=info:eric/&rft_id=info:doi/

http://ucelinks.cdlib.org:8888/sfx_local?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Ahnplatimes&atitle=New+Law+May+Not+Turn+on+%27Pot%27+Flow+for+Some+Cancer+Patients&title=Los+Angeles+Times+%281923-1995%29&issn=04583035&date=1979-12-27&volume=&issue=&spage=SG2&au=Anderson%2C+Pat&isbn=&jtitle=Los+Angeles+Times+%281923-1995%29&btitle=&rft_id=info:eric/&rft_id=info:doi/

Title: New Law May Not Turn on 'Pot' Flow for Some Cancer Patients

Publication title: Los Angeles Times (1923-1995); Los Angeles, Calif.

Pages: sg2

Number of pages: 1

Publication year: 1979

Publication date: Dec 27, 1979

Section: San Gabriel Valley

Publisher: Tribune Publishing Company, LLC

Place of publication: Los Angeles, Calif.

Country of publication: United States, Los Angeles, Calif.

Publication subject: General Interest Periodicals--United States

ISSN: 04583035

Source type: Historical Newspapers

Language of publication: English

Document type: article

ProQuest document ID: 162537071

Document URL: https://search.proquest.com/docview/162537071?accountid=14496

Copyright: Copyright Times Mirror Company Dec 27, 1979

Last updated: 2011-12-31

Database: ProQuest Historical Newspapers: Los Angeles Times

Law May Not Turn On 'Pot' Flow for Some Cancer Patients

Author: Anderson, Pat

Publication info: Los Angeles Times (1923-1995) ; Los Angeles, Calif. [Los Angeles, Calif]09 Sep 1979: se8b.

https://search.proquest.com/docview/159011900?accountid=14496

Abstract: In January, when marijuana becomes a legalized medicine, it will be a problem, as researchers see it, to get a large enough supply for the thousands of California patients who will be eligible for treatment.

Links: http://ucelinks.cdlib.org:8888/sfx_local?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Ahnplatimes&atitle=Law+May+Not+Turn+On+%27Pot%27+Flow+for+Some+Cancer+Patients&title=Los+Angeles+Times+%281923-1995%29&issn=04583035&date=1979-09-09&volume=&issue=&spage=SE8B&au=Anderson%2C+Pat&isbn=&jtitle=Los+Angeles+Times+%281923-1995%29&btitle=&rft_id=info:eric/&rft_id=info:doi/

http://ucelinks.cdlib.org:8888/sfx_local?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Ahnplatimes&atitle=Law+May+Not+Turn+On+%27Pot%27+Flow+for+Some+Cancer+Patients&title=Los+Angeles+Times+%281923-1995%29&issn=04583035&date=1979-09-09&volume=&issue=&spage=SE8B&au=Anderson%2C+Pat&isbn=&jtitle=Los+Angeles+Times+%281923-1995%29&btitle=&rft_id=info:eric/&rft_id=info:doi/

Title: Law May Not Turn On 'Pot' Flow for Some Cancer Patients

	Publication	Location	Date	Title	Abstract	Full Text	Tags	Author
0	Winnipeg Free Press	NaN	Jan 11, 2012	Legalize pot but get over 'harmless' pipe dream	True, but anybody who likes to have a joint on...	How many of your friends or colleagues at work...	NaN	Marks, Don
1	Edmonton Journal	NaN	Feb 26, 2011	City murders far from random; Blue-collar city...	Each year Maclean's magazine publishes a rathe...	Depending on your method of statistical analys...	Marijuana; Abortion	Babiak, Todd
2	National Post	NaN	Nov 11, 2010	Full Pundit	In light of the government's decision to provi...	Every weekday on FullComment.com,the blog of t...	NaN	Anonymous
3	The Gazette	NaN	Aug 6, 1992	Dope-lobby head is a short-haired Republican: ...	There's this U.S. organization called NORML, t...	There's this U.S. organization called NORML,...	NaN	Camilli, Doug
4	The Windsor Star	NaN	Jan 6, 2004	GUEST COLUMN: Being 'dull' Canadians had some ...	There was a magazine cover in the distinguishe...	EDITOR'S NOTE: Charles Gordon is an Ottawa C...	NaN	Gordon, Charles

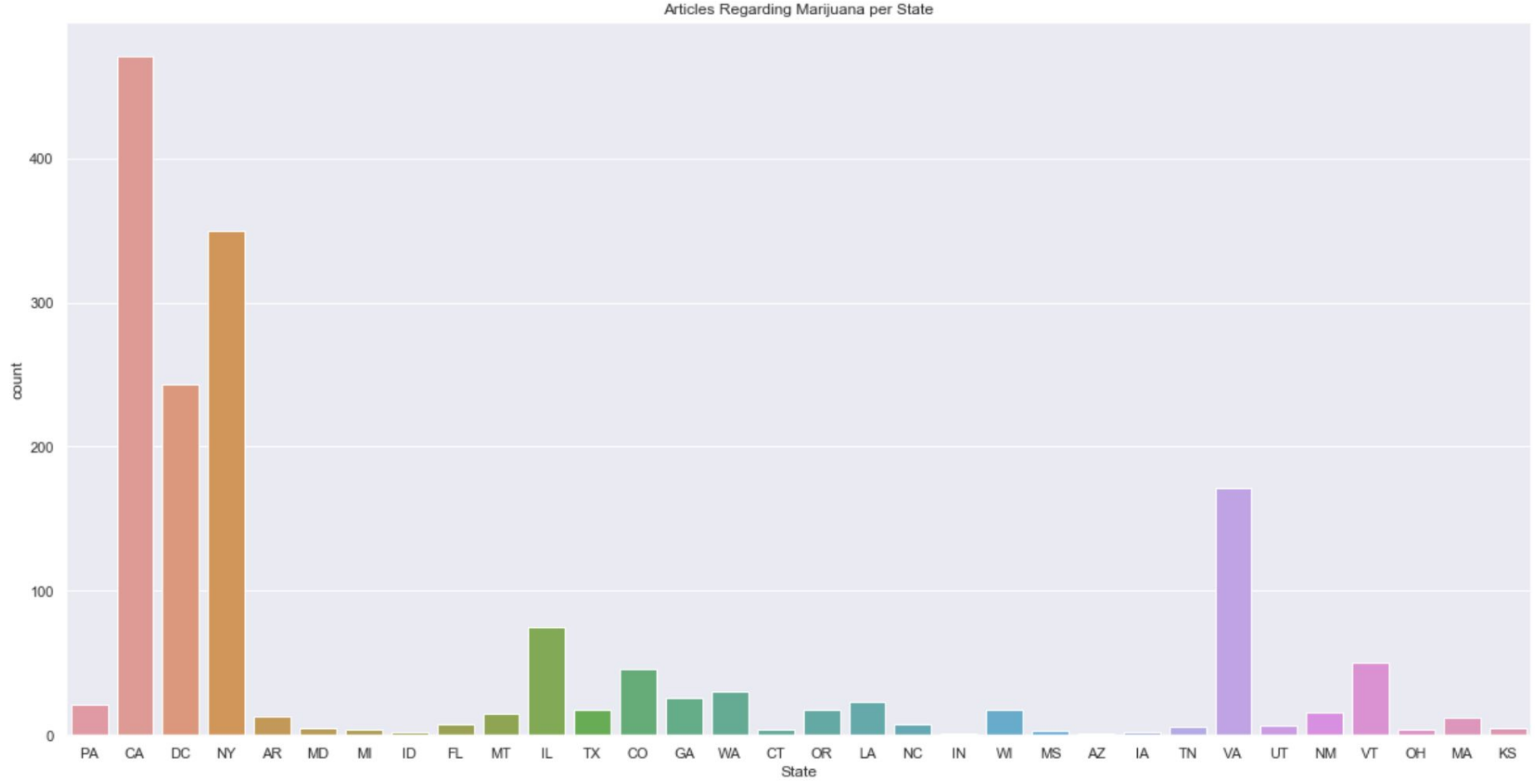
Compiled dataframe

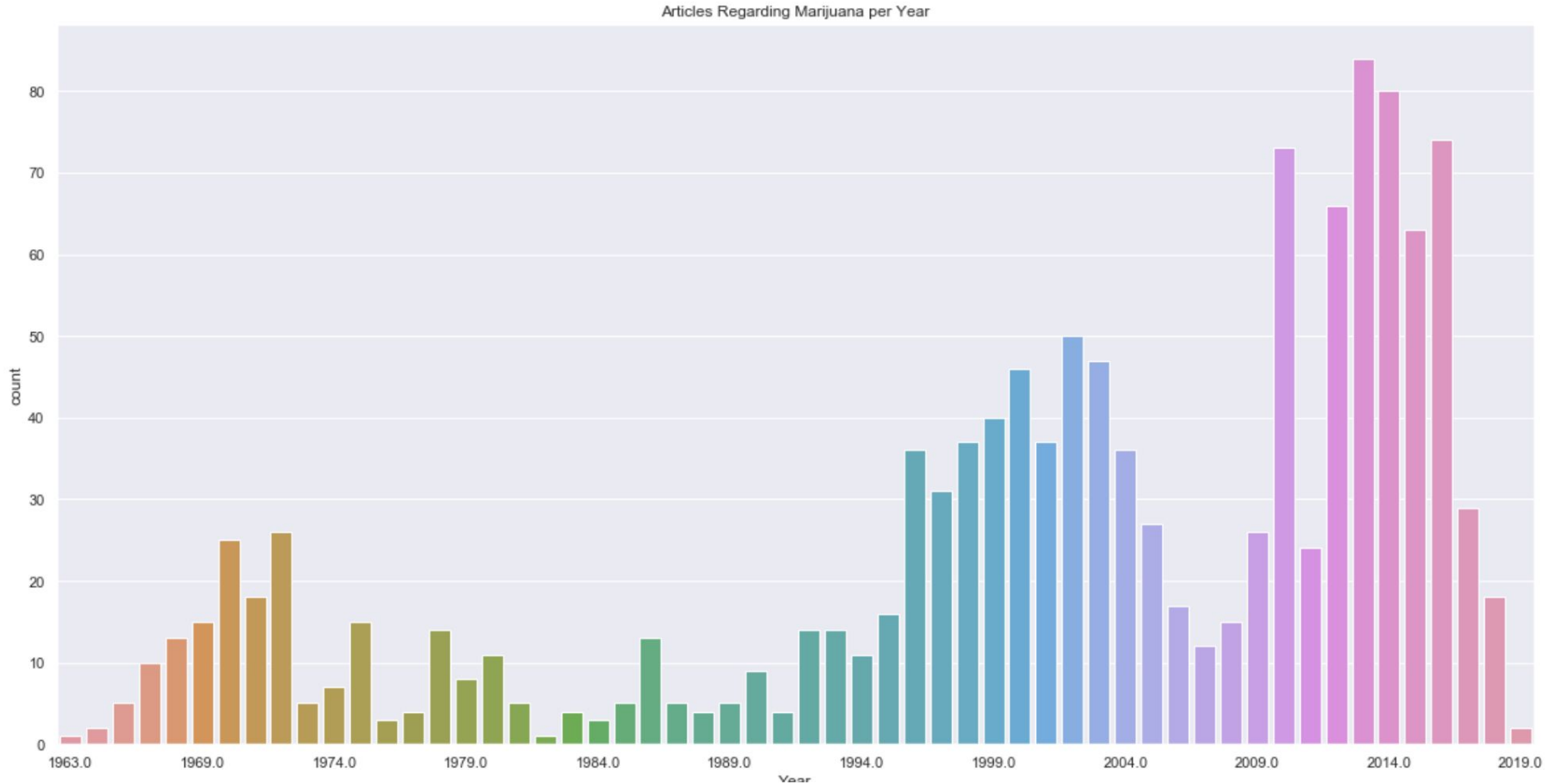
2: Data Cleaning & EDA

	Unnamed: 0	Publication	Location	Date	Title	Abstract	Full Text	Tags	Author
0	802	A. Magazine	Pennsylvania	May 31, 1997	A SOUL ON FIRE	The vigil for victims of anti-Asian violence w...	A soul on fire.SHE WAS CHINESE AMERICAN AND MI...	Asian Americans; Asians; Colleges & universiti...	Li, Kenneth
1	310	Ability	California	Feb 29, 2004	MONTEL WILLIAMS	[MONTEL WILLIAMS] began using marijuana to hel...	A threatening war cry screams out, breaking th...	Chronic illnesses; Developmental disabilities;...	NaN
2	1183	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	Mar 17-Mar 23, 2018	Women's March Leader Defends Herself for Atten...	According to CNN's Jake Tapper, Farrakhan call...	Tamika Mallory, Women's March organizer and co...	Womens rights movement; Jews; Marijuana; Activ...	Green, Micha
3	411	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	Apr 29-May 5, 2017	Getting Blacks a Seat at the D.C. Cannabis Table	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Medical marijuana; Drug legalization; Clinics;...	Green, Micha
4	504	Afro - American Red Star	Washington DC	Apr 29-May 5, 2017	Getting Blacks a Seat at the D.C. Cannabis Table	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Drug legalization; Clinics; Business ownership...	Green, Micha

Data cleaning

Initial Visualizations


Articles Regarding Marijuana per State
count
0
100
200
300
400
PA
CA
DC
NY
AR
MD
MI
ID
FL
MT
IL
TX
CO
GA
WA
CT
OR
LA
NC
IN
WI
MS
AZ
IA
TN
VA
UT
NM
VT
OH
MA
KS
State

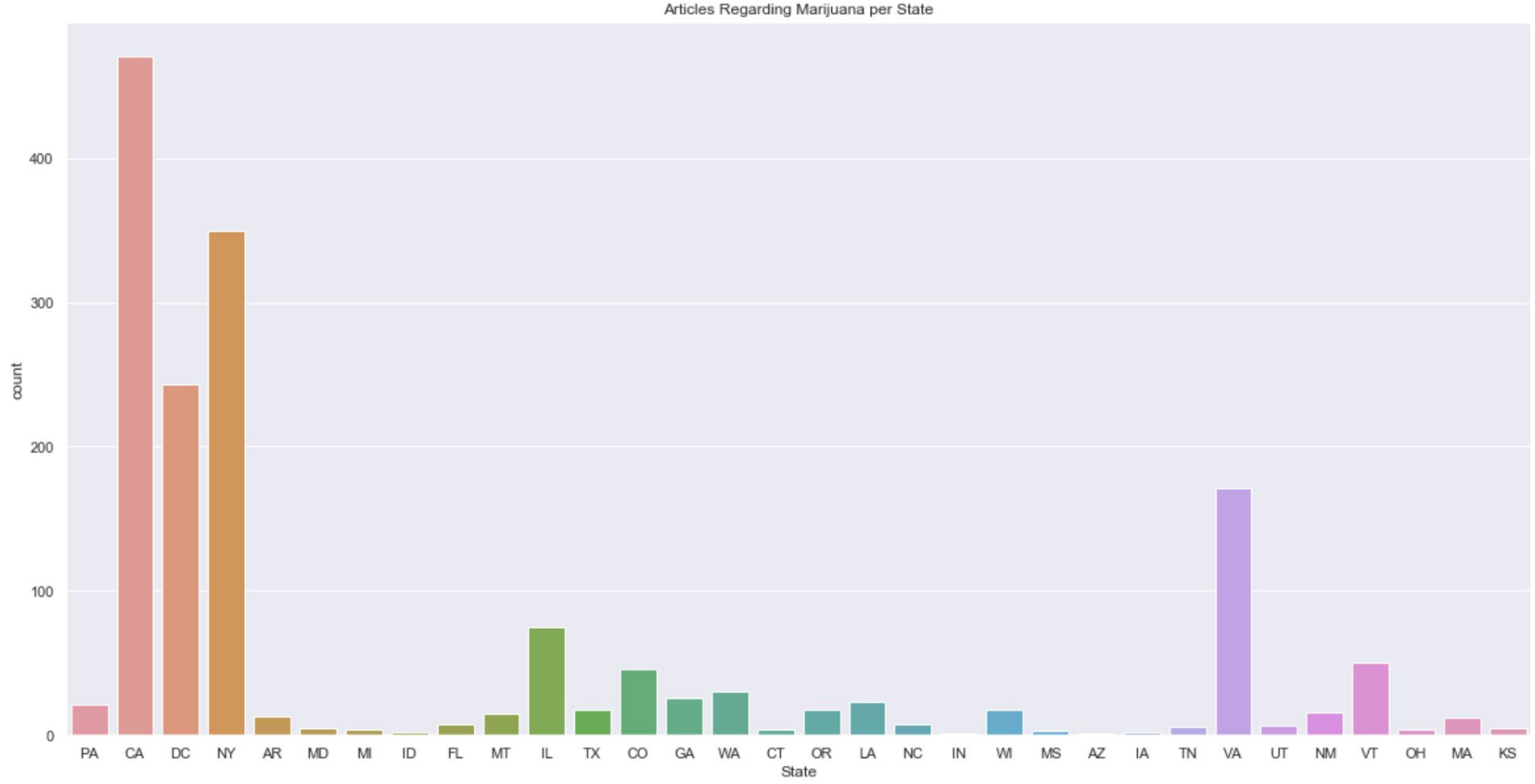

Articles Regarding Marijuana per Year
count
0
10
20
30
40
50
60
70
80
1963.0
1969.0
1974.0
1979.0
1984.0
1989.0
1994.0
1999.0
2004.0
2009.0
2014.0
2019.0
Year

```
[('Drug legalization', 126),
 ('Marijuana', 119),
 ('Medical marijuana', 74),
 ('Candidates', 41),
 ('Politics', 36),
 ('Political parties', 35),
 ('Clinics', 33),
 ('Drug policy', 33),
 ('Bills', 33),
 ('Political campaigns', 29),
 ('State laws', 26),
 ('Governors', 24),
 ('State elections', 22),
 ('Local elections', 21),
 ('Presidential elections', 20),
 ('Referendums', 20),
 ('Law enforcement', 17),
 ('Drug use', 17),
 ('Voter behavior', 16),
 ('Medical treatment', 16)]
```

Most commonly used tags

Preparing Data for Prediction

Labels: “illegal”, “rec”, “med”


Articles Regarding Marijuana per State
count
0
100
200
300
400
PA
CA
DC
NY
AR
MD
MI
ID
FL
MT
IL
TX
CO
GA
WA
CT
OR
LA
NC
IN
WI
MS
AZ
IA
TN
VA
UT
NM
VT
OH
MA
KS
State

California:

- Medical use legalized in 1996
- Recreational use legalized in 2016

Washington DC:

- Medical use legalized in 2010
- Recreational use legalized in 2014

Virginia:

Marijuana is illegal for all purposes in Virginia

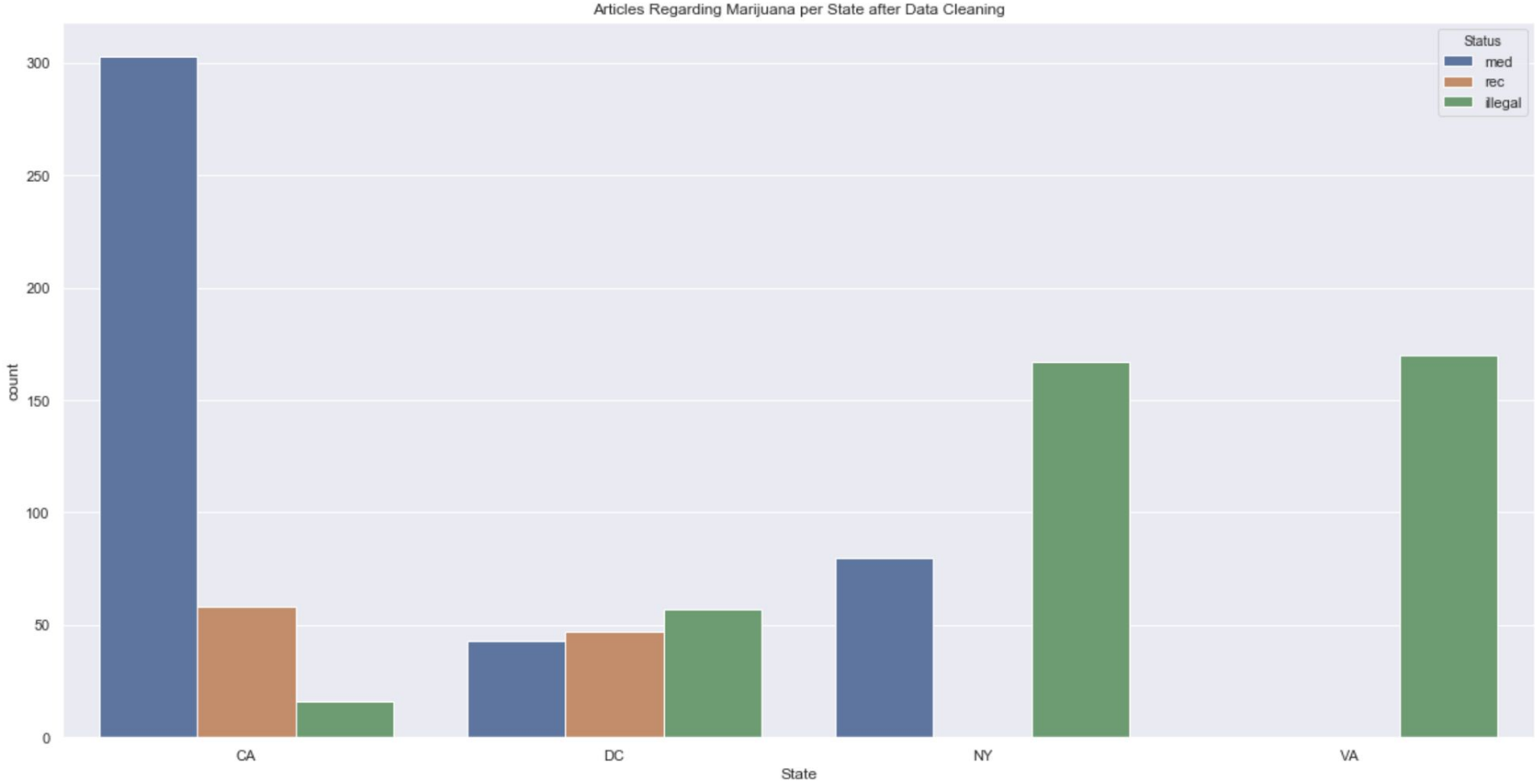
New York:

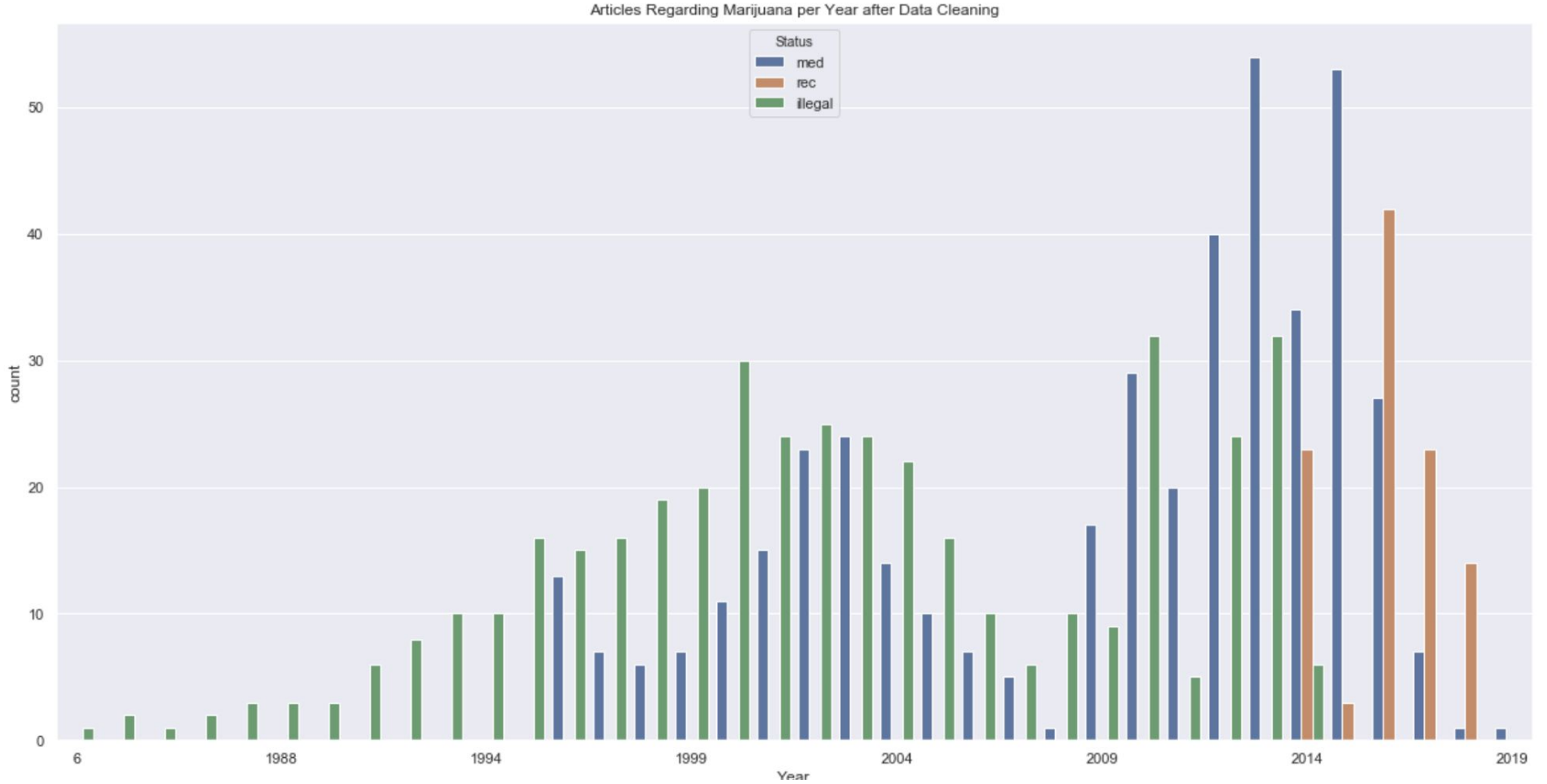
- Medical use legalized in 2014

Note that recreational use is still illegal in NY

	Publication	Location	Abstract	Full Text	Tags	Year	tokens	stemmed_tokens	Status
0	Ability	California	[MONTEL WILLIAMS] began using marijuana to hel...	A threatening war cry screams out, breaking th...	Chronic illnesses; Developmental disabilities;...	2004	[a, threatening, war, cry, screams, out, break...	[a, threaten, war, cri, scream, out, break, th...	med
1	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	According to CNN's Jake Tapper, Farrakhan call...	Tamika Mallory, Women's March organizer and co...	Womens rights movement; Jews; Marijuana; Activ...	2018	[tamika, mallory, women, s, march, organizer, ...	[tamika, mallori, women, s, march, organ, and,...	rec
2	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Medical marijuana; Drug legalization; Clinics;...	2017	[now, that, recreational, marijuana, use, is, ...	[now, that, recreat, marijuana, use, is, legal...	rec
3	Afro - American Red Star	Washington DC	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Drug legalization; Clinics; Business ownership...	2017	[now, that, recreational, marijuana, use, is, ...	[now, that, recreat, marijuana, use, is, legal...	rec
4	Afro - American Red Star	Washington DC	Leaders of a Beautiful Struggle (LBS), a Balti...	Leaders of a Beautiful Struggle (LBS), a Balti...	Public safety; Legislators; African Americans;...	2017	[leaders, of, a, beautiful, struggle, lbs, a, ...	[leader, of, a, beauti, struggl, lbs, a, balti...	rec

Dataframe with labels column


Articles Regarding Marijuana per State after Data Cleaning
Status
med
rec
illegal
count
300
250
200
150
100
50
0
CA
DC
NY
VA
State


Articles Regarding Marijuana per Year after Data Cleaning
Status
med
rec
illegal
count
0
10
20
30
40
50
6
1988
1994
1999
2004
2009
2014
2019
Year

Word2Vec EDA

```
#Word2Vec

textA = ''
textB = ''
textC = ''

count = 0
for index, row in illegal.iterrows():
    if count < 200:
        textA += row['Full Text'] + ' '
        count+=1
    else:
        break
count = 0
for index, row in medicinal.iterrows():
    if count < 200:
        textB += row['Full Text'] + ' '
        count+=1
    else:
        break
count = 0
for index, row in recreational.iterrows():
    if count < 70:
        textC += row['Full Text'] + ' '
        count+=1
    else:
        break
```

Word2Vec

```
modelA = gensim.models.Word2Vec(tokensA, size=300, window=5, \
                                min_count=5, sg=0, alpha=0.025, iter=5, batch_words=10000)
```

```
modelB = gensim.models.Word2Vec(tokensB, size=300, window=5, \
                                min_count=5, sg=0, alpha=0.025, iter=5, batch_words=10000)
```

```
modelC = gensim.models.Word2Vec(tokensC, size=300, window=5, \
                                min_count=5, sg=0, alpha=0.025, iter=5, batch_words=10000)
```

```
modelA.similarity('sales', 'medical')    0.9997554
modelB.similarity('sales', 'medical')    0.9738905
modelC.similarity('sales', 'medical')    0.9997512
```

Word2Vec

```
modelA.most_similar(positive=['marijuana', 'medical'], negative=['legalization'])
```

```
[('use', 0.9998553395271301),
 ('legalizing', 0.9998075366020203),
 ('legalize', 0.9997970461845398),
 ('legalized', 0.9997687935829163),
 ('medicinal', 0.9997315406799316),
 ('patients', 0.999721884727478),
 ('smoked', 0.999717116355896),
 ('legal', 0.9997045397758484),
 ('purposes', 0.999700665473938),
 ('possession', 0.9996973872184753)]
```

```
modelB.most_similar(positive=['marijuana', 'medical'], negative=['legalization'])
```

```
[('use', 0.9807037711143494),
 ('schutten', 0.979212760925293),
 ('tackle', 0.979049563407898),
 ('teen', 0.9727762937545776),
 ('chains', 0.9727222919464111),
 ('makeup', 0.9687608480453491),
 ('would', 0.9669689536094666),
 ('repeal', 0.9662026166915894),
 ('like', 0.9650996327400208),
 ('government', 0.9649485349655151)]
```

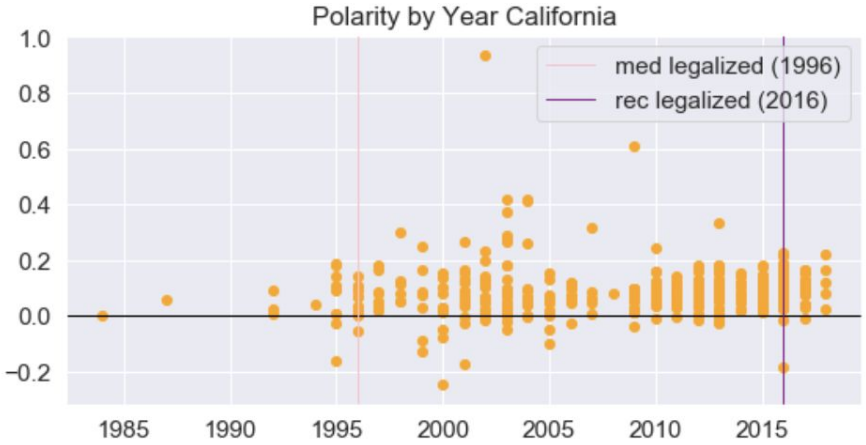
```
modelC.most_similar(positive=['marijuana', 'medical'], negative=['legalization'])
```

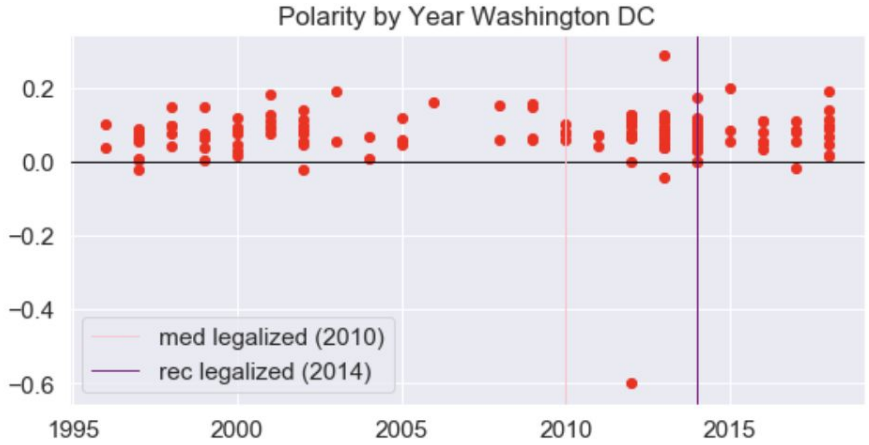
```
[('state', 0.9998083114624023),
 ('use', 0.9998024106025696),
 ('people', 0.9997979998588562),
 ('one', 0.9997966289520264),
 ('cannabis', 0.9997956156730652),
 ('first', 0.9997926950454712),
 ('said', 0.9997906684875488),
 ('drug', 0.9997871518135071),
 ('states', 0.999785840511322),
 ('would', 0.999781608581543)]
```

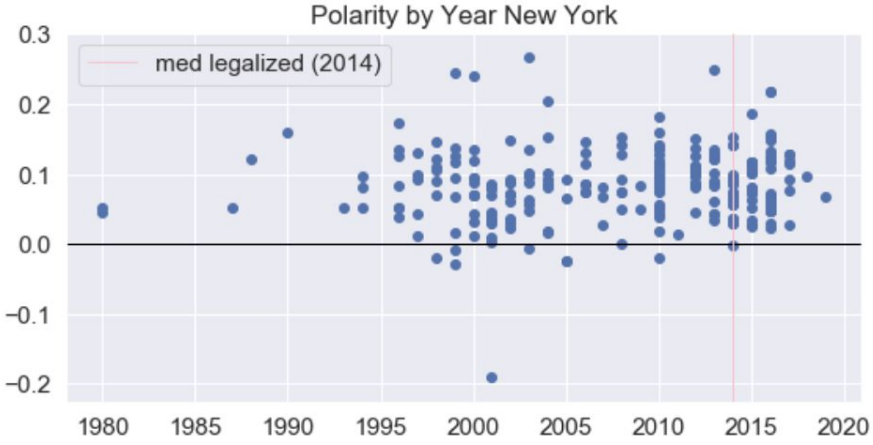
Morality & Sentiment Analysis: Polarity

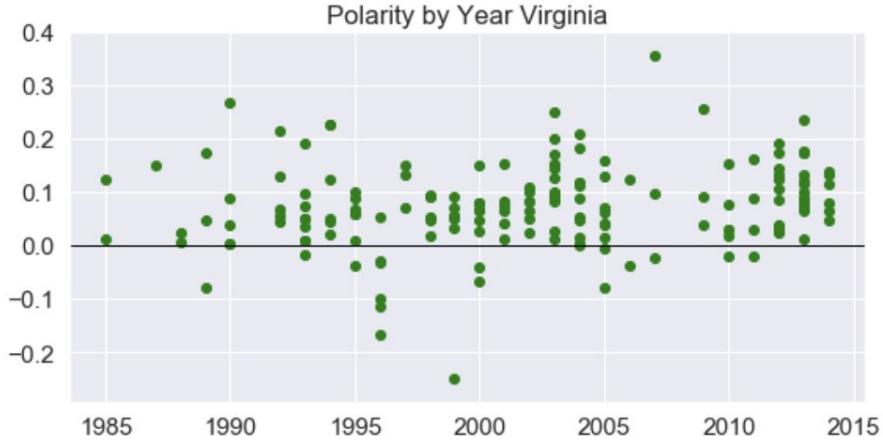
	Unnamed: 0	Publication	Location	Abstract	Full Text	Tags	Year	tokens	stemmed_tokens	Status	polarity
0	0	Ability	California	[MONTEL WILLIAMS] began using marijuana to hel...	A threatening war cry screams out, breaking th...	Chronic illnesses; Developmental disabilities;...	2004	['a', 'threatening', 'war', 'cry', 'screams', ...	['a', 'threaten', 'war', 'cri', 'scream', 'out...	med	0.093567
1	1	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	According to CNN's Jake Tapper, Farrakhan call...	Tamika Mallory, Women's March organizer and co...	Womens rights movement; Jews; Marijuana; Activ...	2018	['tamika', 'mallory', 'women', 's', 'march', '...	['tamika', 'mallori', 'women', 's', 'march', '...	rec	0.018572
2	2	Afro - American , 5 Star edition; Baltimore, M...	Washington DC	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Medical marijuana; Drug legalization; Clinics;...	2017	['now', 'that', 'recreational', 'marijuana', '...	['now', 'that', 'recreat', 'marijuana', 'use',...	rec	0.085300
3	3	Afro - American Red Star	Washington DC	Entrepreneurs hoping to benefit from the new l...	Now that recreational marijuana use is legal f...	Drug legalization; Clinics; Business ownership...	2017	['now', 'that', 'recreational', 'marijuana', '...	['now', 'that', 'recreat', 'marijuana', 'use',...	rec	0.085300
4	4	Afro - American Red Star	Washington DC	Leaders of a Beautiful Struggle (LBS), a Balti...	Leaders of a Beautiful Struggle (LBS), a Balti...	Public safety; Legislators; African Americans;...	2017	['leaders', 'of', 'a', 'beautiful', 'struggle'...	['leader', 'of', 'a', 'beauti', 'struggl', 'lb...	rec	0.108641

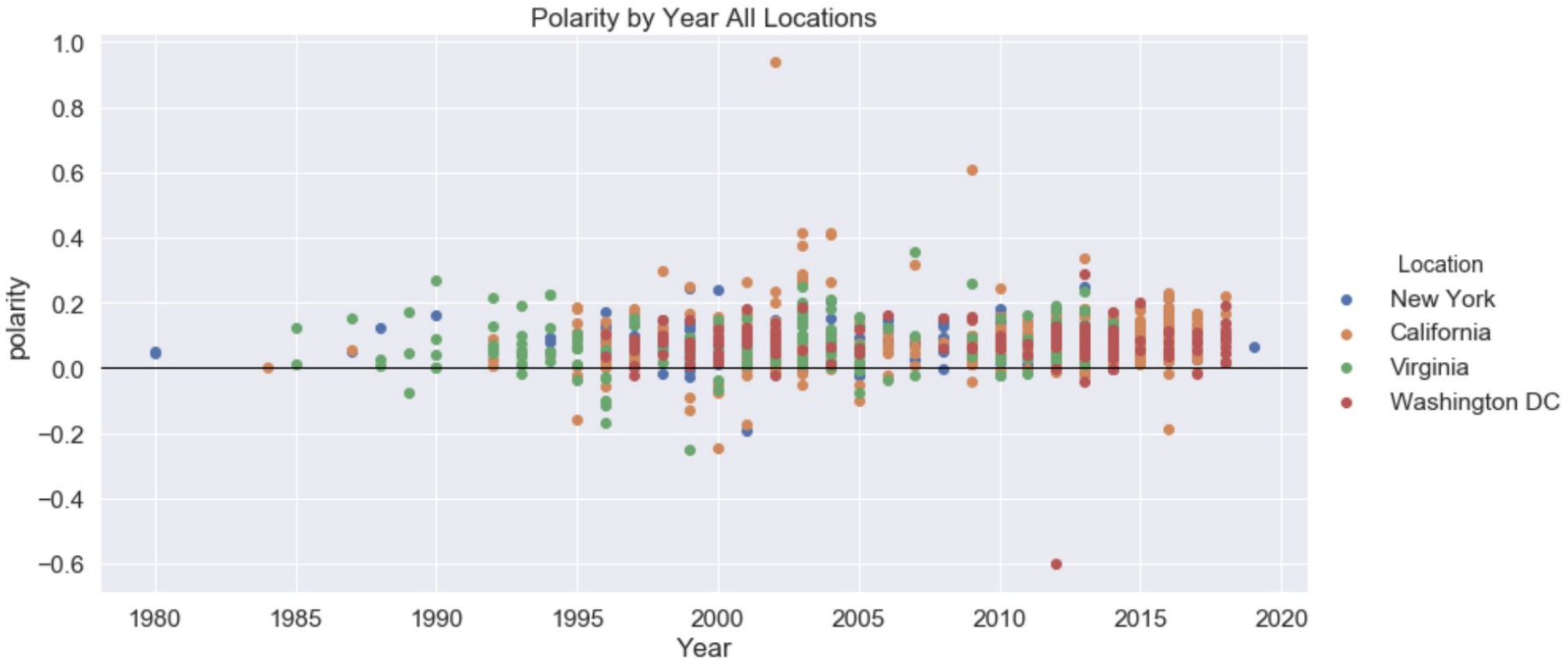
Dataframe with polarity scores

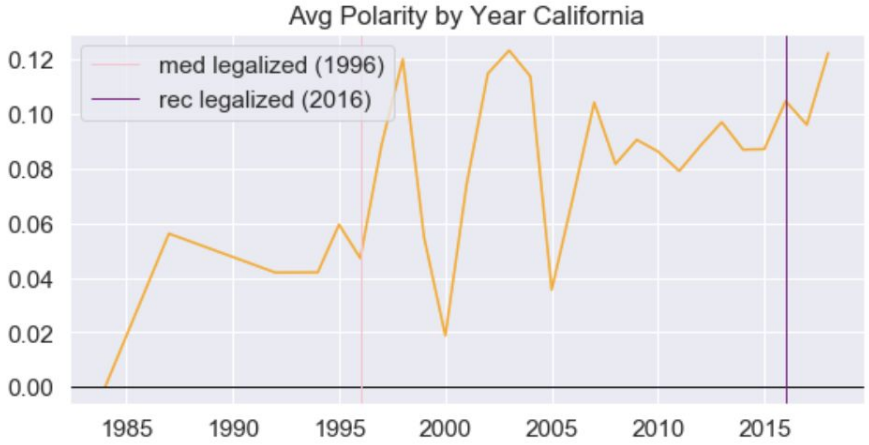

Polarity by Year California
med legalized (1996)
rec legalized (2016)
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
1985
1990
1995
2000
2005
2010
2015

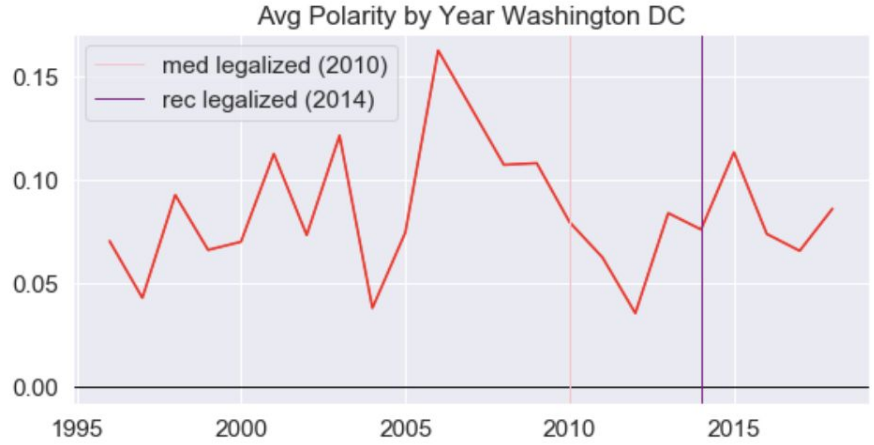

Polarity by Year Washington DC
med legalized (2010)
rec legalized (2014)
0.2
0.0
−0.2
−0.4
−0.6
1995
2000
2005
2010
2015

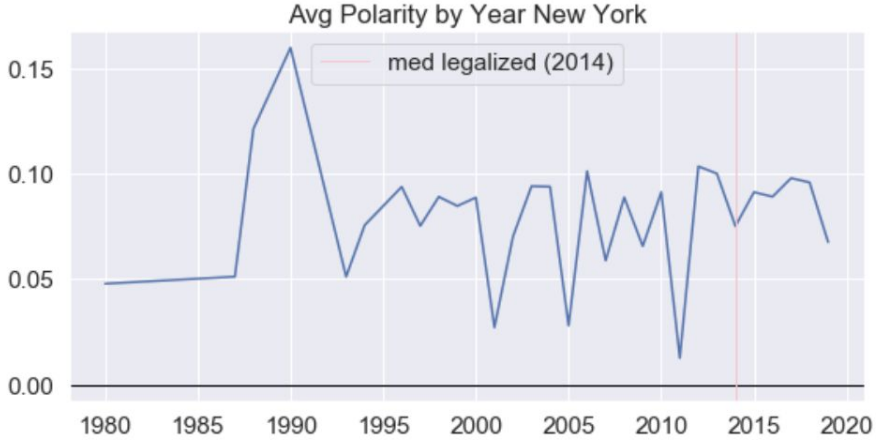

Polarity by Year New York
med legalized (2014)
0.3
0.2
0.1
0.0
−0.1
−0.2
1980
1985
1990
1995
2000
2005
2010
2015
2020

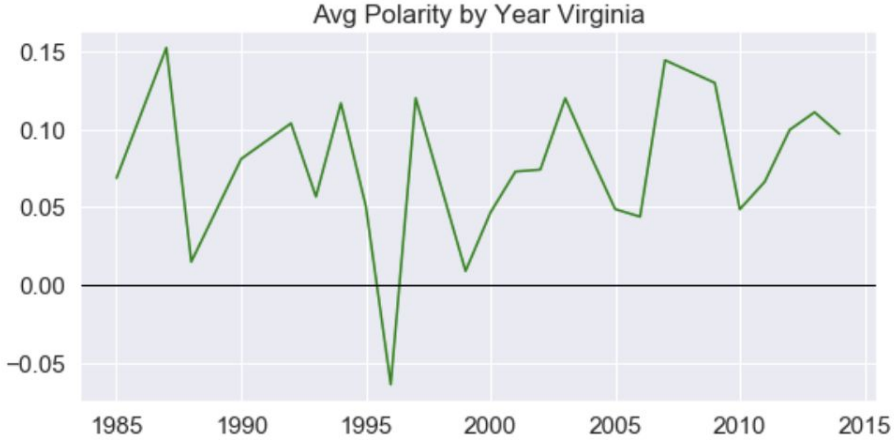

Polarity by Year Virginia
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
1985
1990
1995
2000
2005
2010
2015

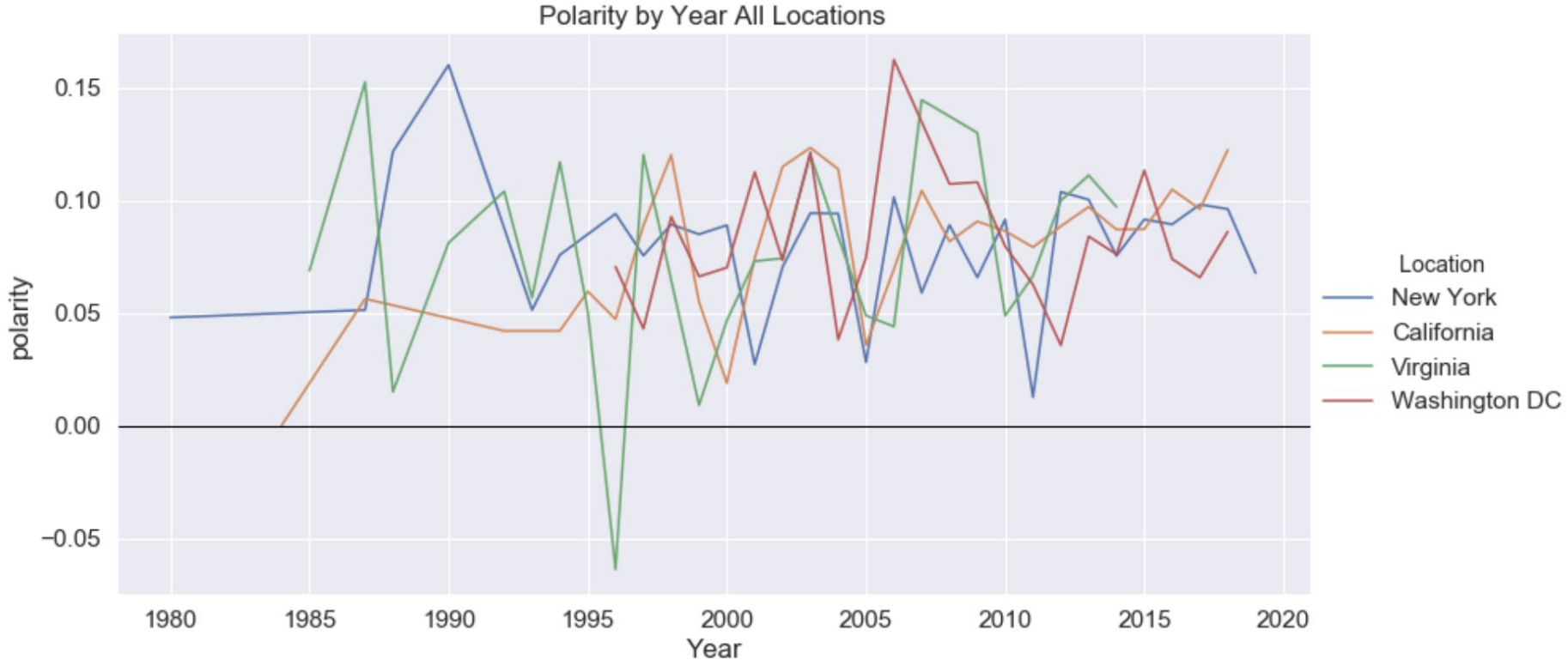

Polarity by Year All Locations
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
polarity
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Location
New York
California
Virginia
Washington DC

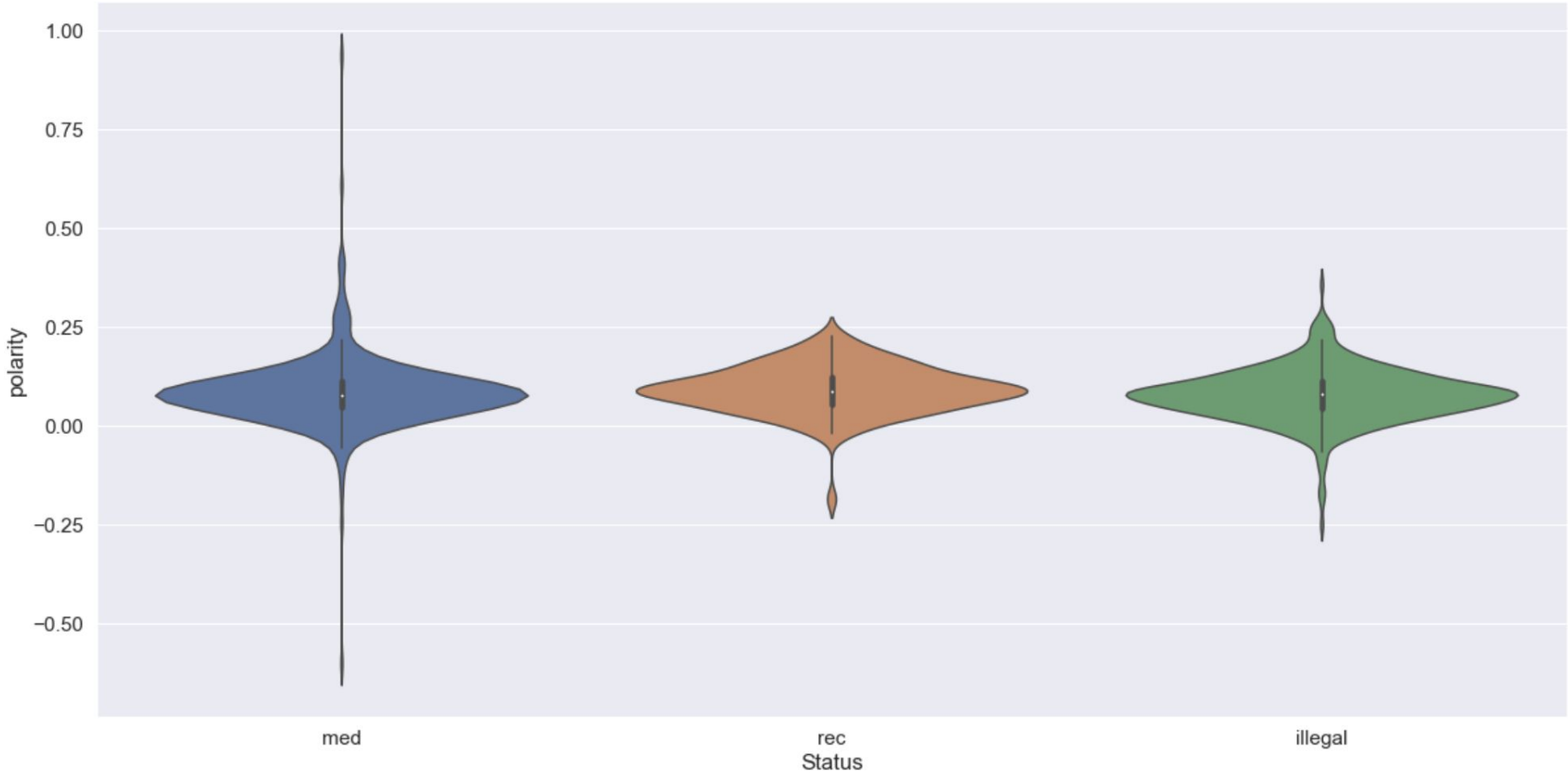

Avg Polarity by Year California
med legalized (1996)
rec legalized (2016)
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1985
1990
1995
2000
2005
2010
2015


Avg Polarity by Year Washington DC
med legalized (2010)
rec legalized (2014)
0.15
0.10
0.05
0.00
1995
2000
2005
2010
2015


Avg Polarity by Year New York
med legalized (2014)
0.15
0.10
0.05
0.00
1980
1985
1990
1995
2000
2005
2010
2015
2020


Avg Polarity by Year Virginia
0.15
0.10
0.05
0.00
−0.05
1985
1990
1995
2000
2005
2010
2015


Polarity by Year All Locations
0.15
0.10
0.05
0.00
−0.05
polarity
1980
1985
1990
1995
2000
2005
2010
2015
2020
Year
Location
New York
California
Virginia
Washington DC


1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
polarity
med
rec
illegal
Status

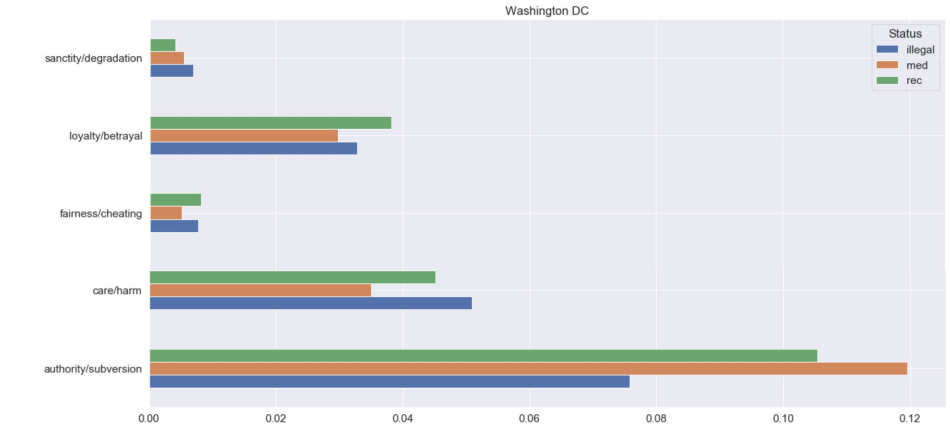
Morality & Sentiment Analysis: Moral Foundations Theory

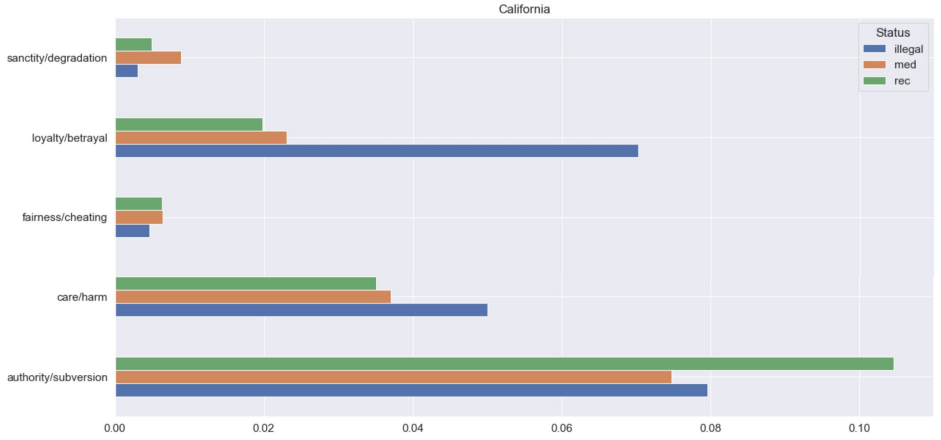
From haidt_dict.json

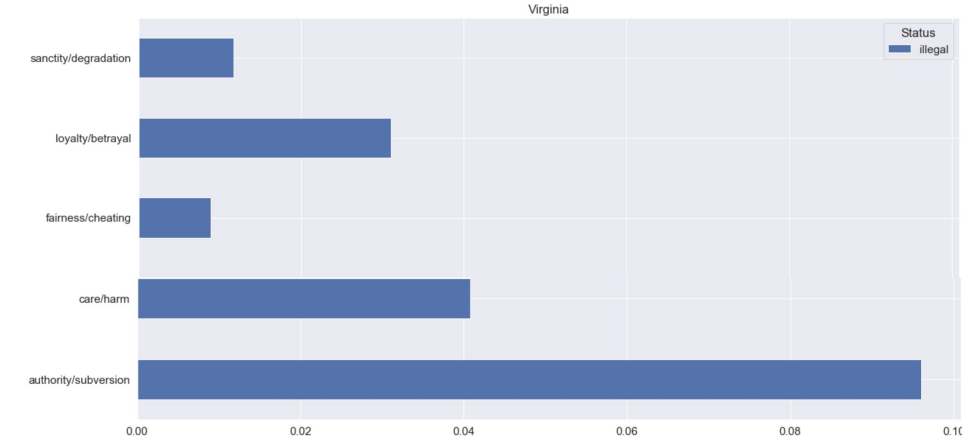
Moral Categories:
authority/subversion
care/harm
fairness/cheating
loyalty/betrayal
sanctity/degradation

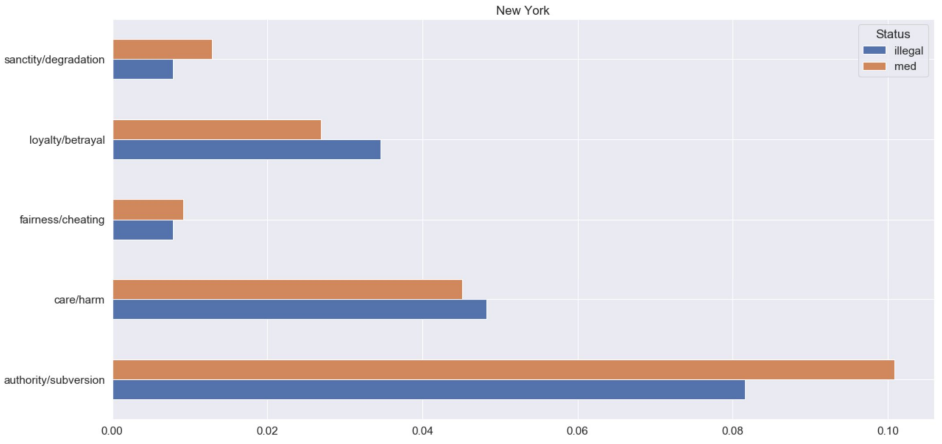
authority/subversion	care/harm	fairness/cheating	loyalty/betrayal	sanctity/degradation
0.047040	0.060480	0.000000	0.016800	0.016800
0.072451	0.090563	0.108676	0.090563	0.018113
0.143535	0.077288	0.000000	0.088329	0.000000
0.143535	0.077288	0.000000	0.088329	0.000000
0.077616	0.031046	0.000000	0.000000	0.000000

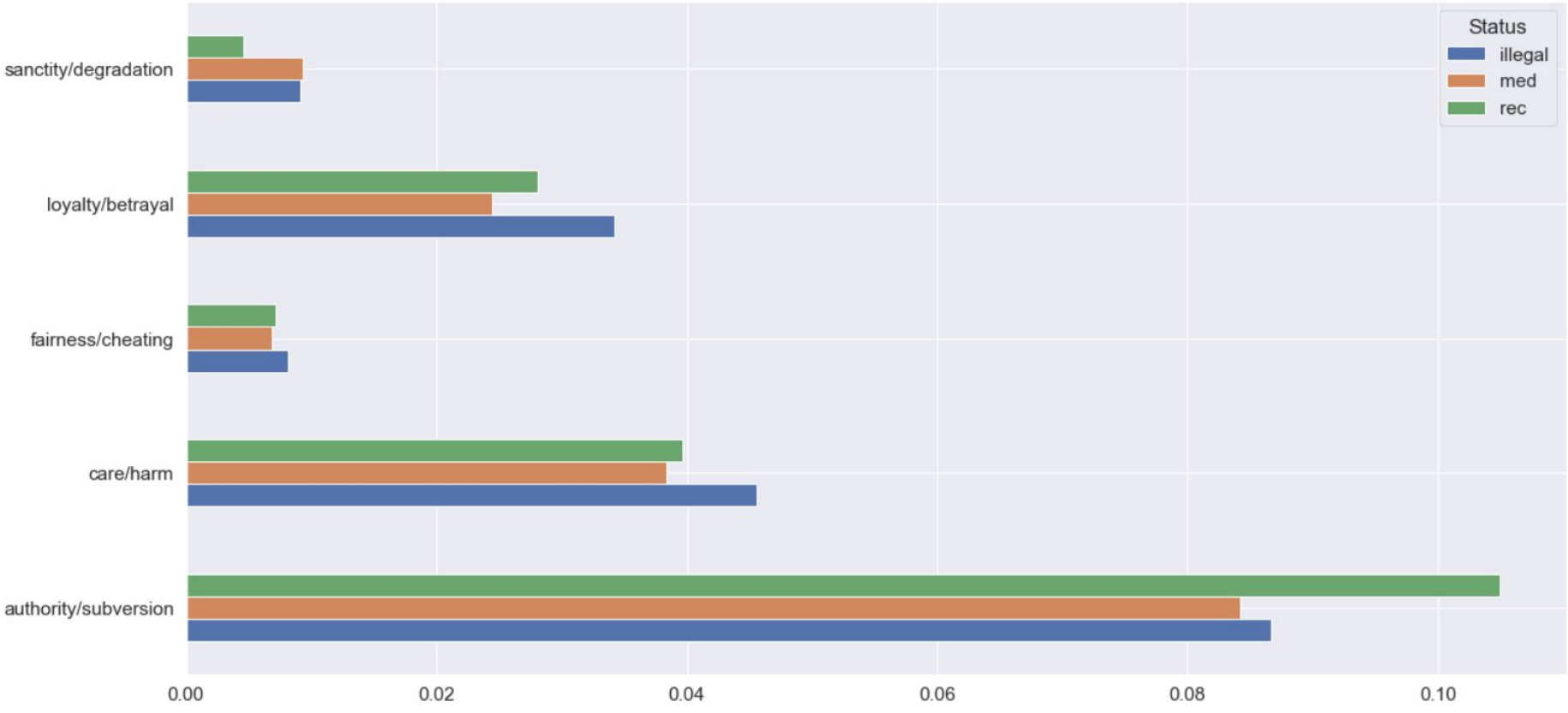
Moral category frequencies


Washington DC
Status
illegal
med
rec
sanctity/degradation
loyalty/betrayal
fairness/cheating
care/harm
authority/subversion
0.00
0.02
0.04
0.06
0.08
0.10
0.12


California
Status
illegal
med
rec
sanctity/degradation
loyalty/betrayal
fairness/cheating
care/harm
authority/subversion
0.00
0.02
0.04
0.06
0.08
0.10


Virginia
Status
illegal
sanctity/degradation
loyalty/betrayal
fairness/cheating
care/harm
authority/subversion
0.00
0.02
0.04
0.06
0.08
0.10


New York
Status
illegal
med
sanctity/degradation
loyalty/betrayal
fairness/cheating
care/harm
authority/subversion
0.00
0.02
0.04
0.06
0.08
0.10


Status
illegal
med
rec
sanctity/degradation
loyalty/betrayal
fairness/cheating
care/harm
authority/subversion
0.00
0.02
0.04
0.06
0.08
0.10

Combining Polarity & Moral Foundations Theory

Combination Metric:

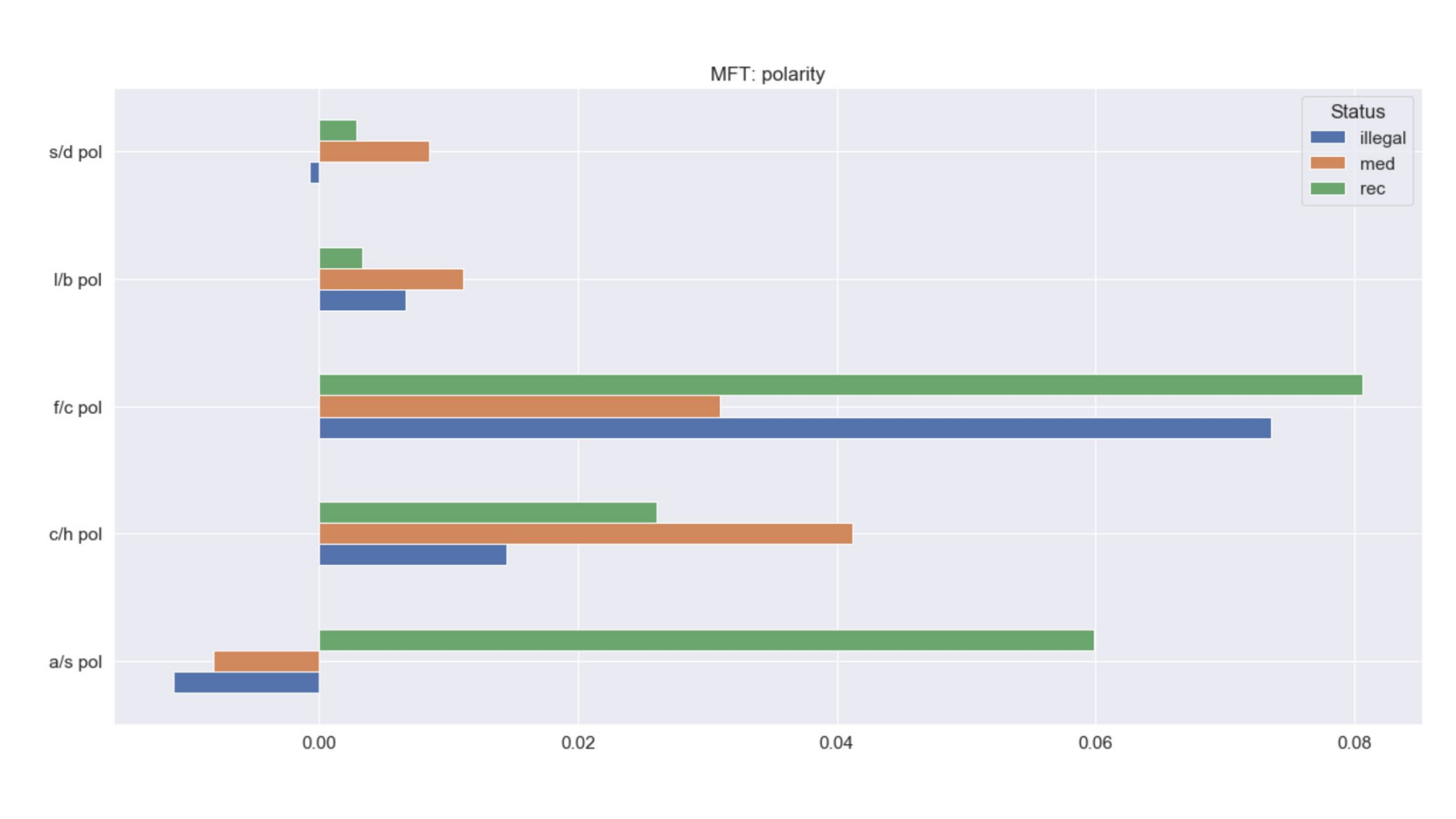
(frequency of moral category) * (polarity of moral category)

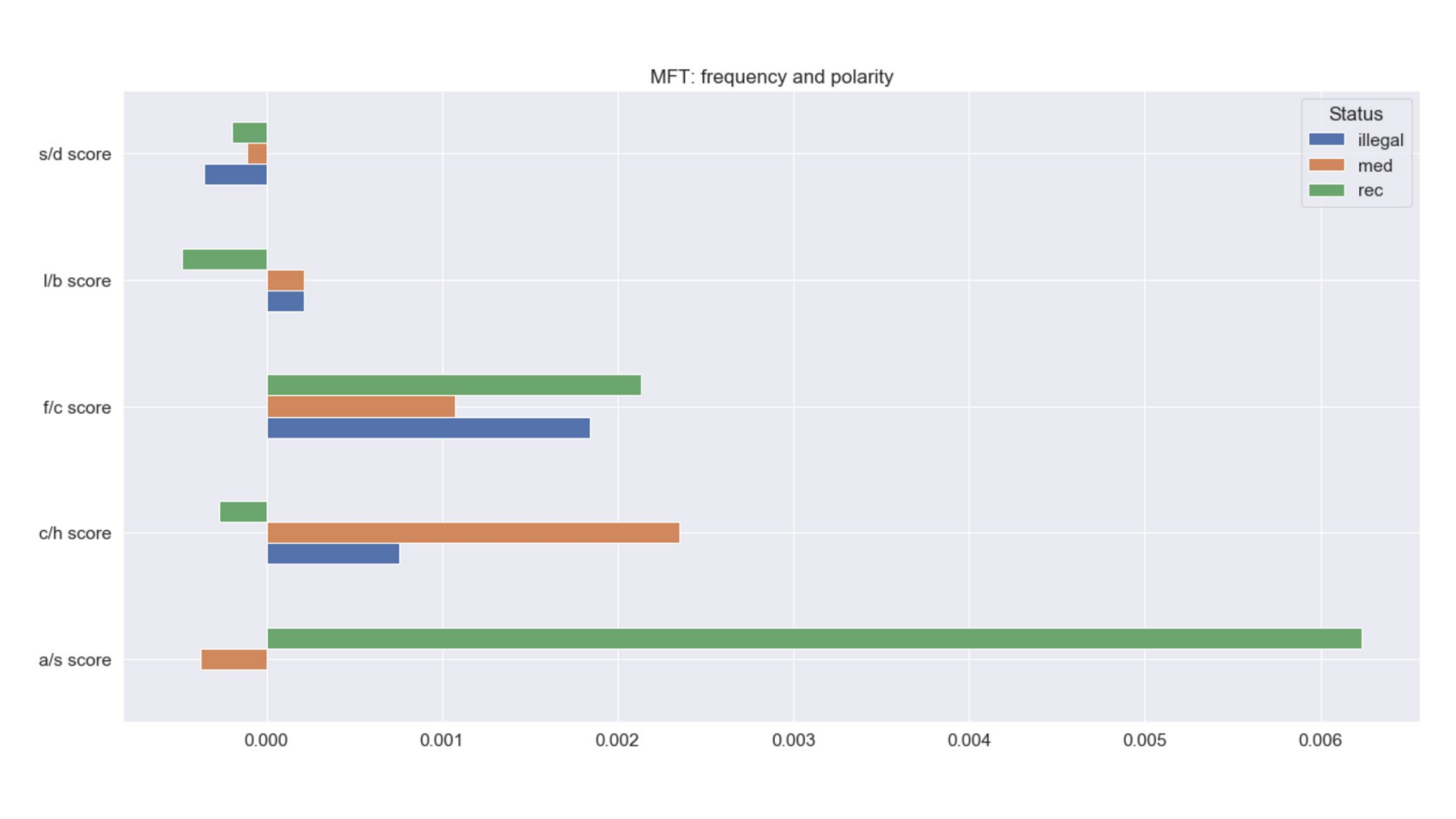
```
def keywords(text, foundation):
    result = []
    for i in range(len(text)):
        if i == 0:
            if text[i] in foundation:
                result.append(text[i])
                result.append(text[i+1])
            else:
                next
        elif i == max(range(len(text))):
            if text[i] in foundation:
                result.append(text[i-1])
                result.append(text[i])
            else:
                next
        else:
            if text[i] in foundation:
                result.append(text[i-1])
                result.append(text[i])
                result.append(text[i+1])
            else:
                next
    return result
```

Extracting terms relevant to moral categories + conext

	Status	polarity	authority/subversion	care/harm	fairness/cheating	loyalty/betrayal	sanctity/degradation	a/s pol	c/h pol	f/c pol	l/b pol	s/d pol
0	med	0.093567	0.047040	0.060480	0.000000	0.016800	0.016800	0.238889	0.0	0.000	0.111111	-0.190476
1	rec	0.018572	0.072451	0.090563	0.108676	0.090563	0.018113	0.000000	-0.2	0.175	0.000000	0.000000
2	rec	0.085300	0.143535	0.077288	0.000000	0.088329	0.000000	0.200000	0.0	0.000	0.000000	0.000000
3	rec	0.085300	0.143535	0.077288	0.000000	0.088329	0.000000	0.200000	0.0	0.000	0.000000	0.000000
4	rec	0.108641	0.077616	0.031046	0.000000	0.000000	0.000000	0.000000	0.0	0.000	-0.166667	0.000000

Dataframe with polarity, moral category frequency, and new combination metric


MFT: polarity
Status
illegal
med
rec
s/d pol
l/b pol
f/c pol
c/h pol
a/s pol
0.00
0.02
0.04
0.06
0.08

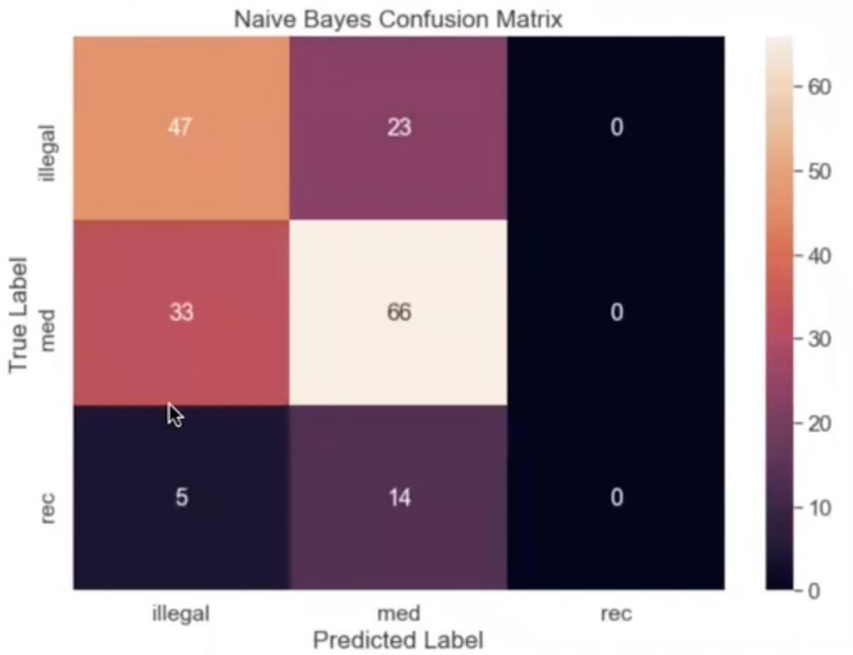

MFT: frequency and polarity
Status
illegal
med
rec
s/d score
l/b score
f/c score
c/h score
a/s score
0.000
0.001
0.002
0.003
0.004
0.005
0.006

3: Classification

TF-IDF - Classification

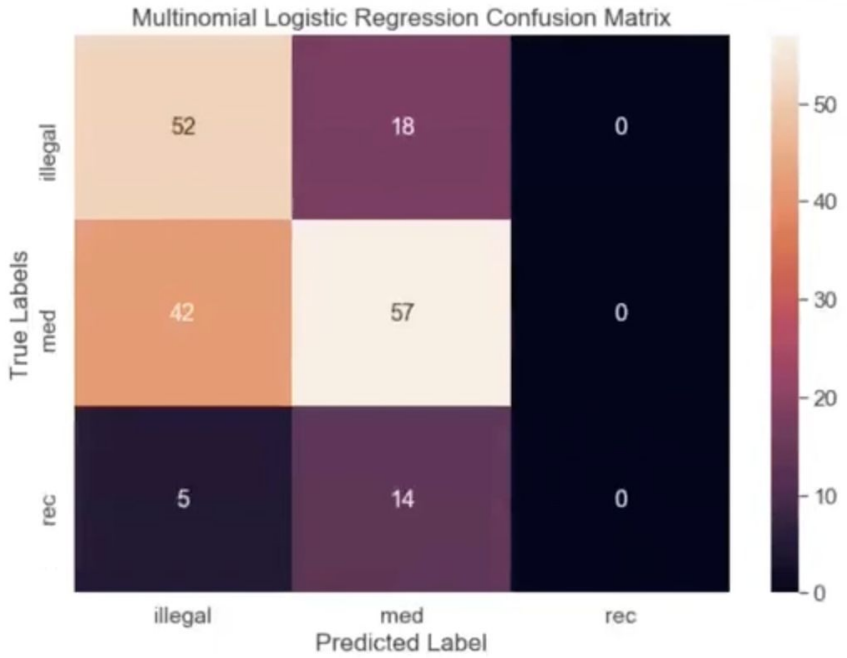
TF-IDF: naive bayes [training]

Score: ~.601



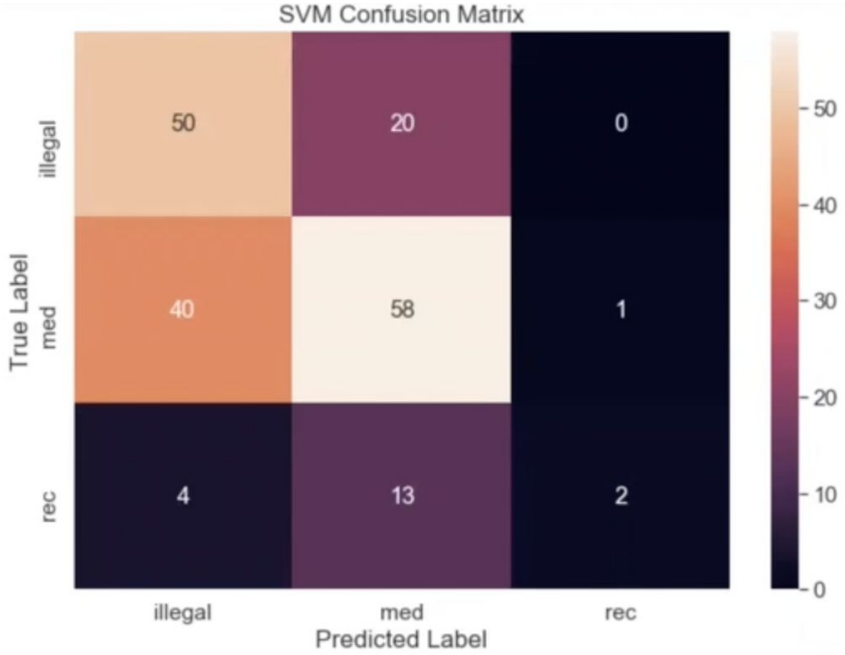
TF-IDF: multinomial logistic [training]

Score: ~.58



TF-IDF: SVM [training]

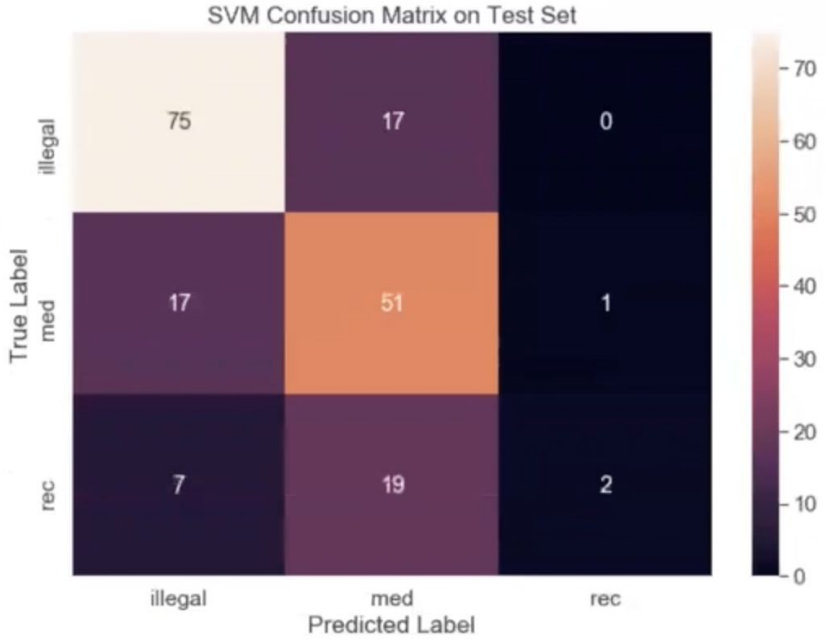
Score: ~.585



[BEST MODEL]

TF-IDF: SVM [test]

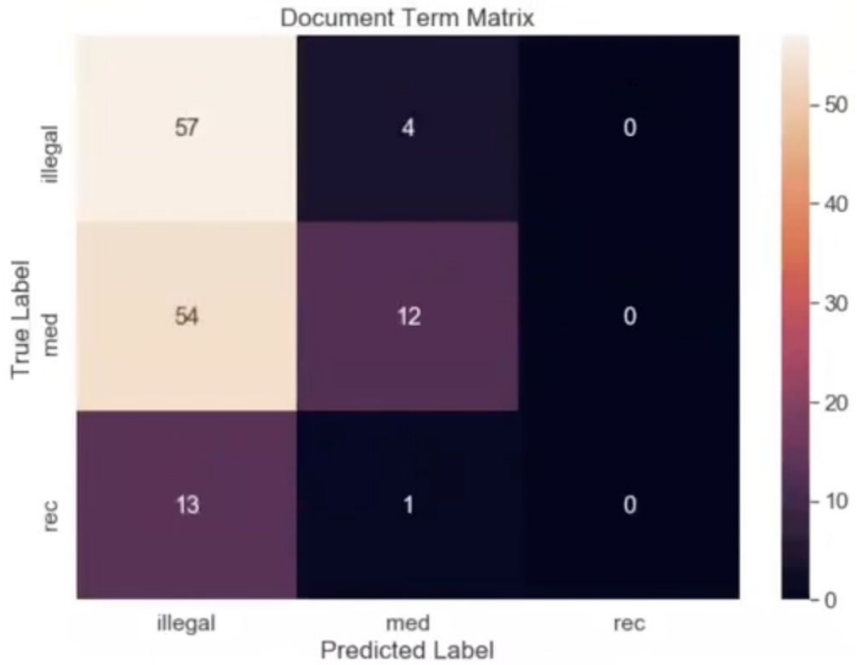
Score: ~.677



Document Term Matrix - Classification

Document Term Matrix: SVM [training]

Score: ~.496



Doc2Vec - Classification

Doc2Vec: logistic reg [training]

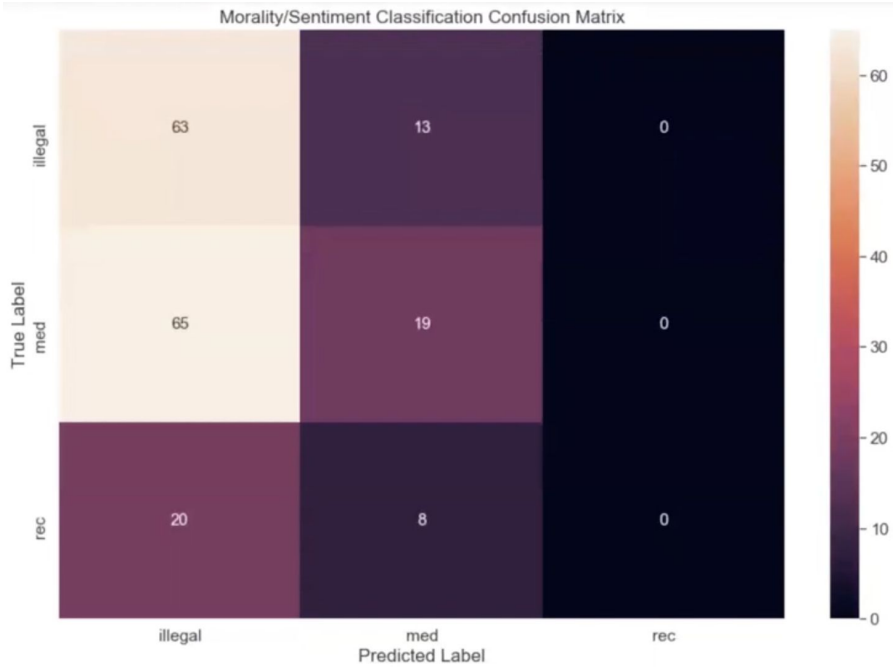
Testing accuracy: ~.561

Testing F1 score: ~.538

Morality & Sentiment - Classification

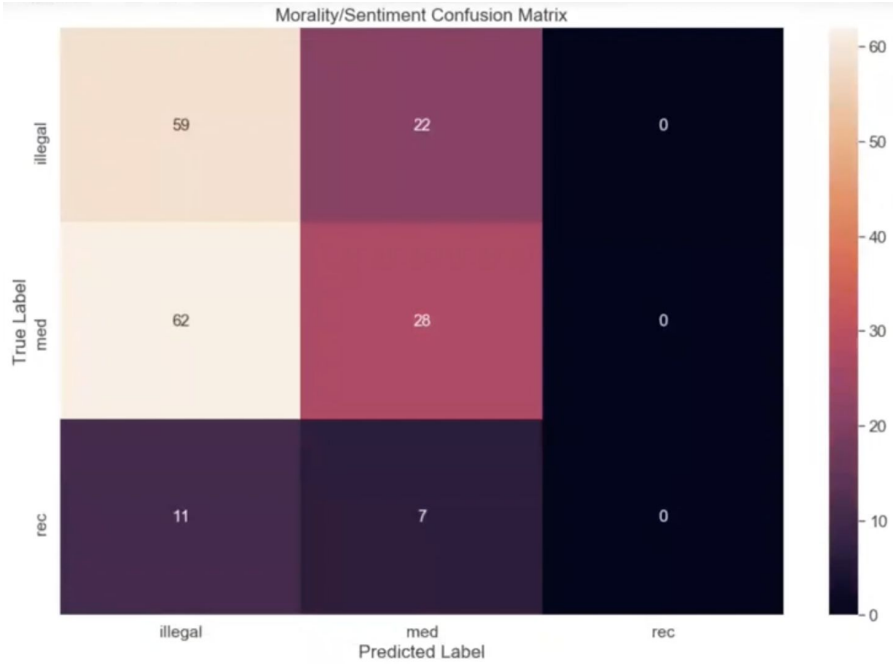
Morality: naive bayes [training]

Score: ~.436



Morality: naive bayes [test]

Score: ~.460



Conclusions

Our classification process also indicated rhetorical linkages between the text written in an article and the legal status of marijuana in the time/region it was written. TF-IDF was our best model.