

# THE WORLD BANK PRESENCE ON TWITTER

Judy S. Yang

February 22, 2016

General Assembly DAT-DC-10, Final Project

# OUTLINE

Intro/Descriptives

Analysis

- Token likelihoods
- Topic Modelling
- Naïve Bayes
- Sentiment Analysis

Conclusion

# QUESTIONS

1. Which accounts are the most popular?
2. Are there too many WB Twitter accounts?
3. How can the WB improve its Twitter presence?
4. What do people tweet about
  1. Internal WB events and publications
  2. WB findings
  3. International events



**DATA**

# SUMMARY DESCRIPTIVES

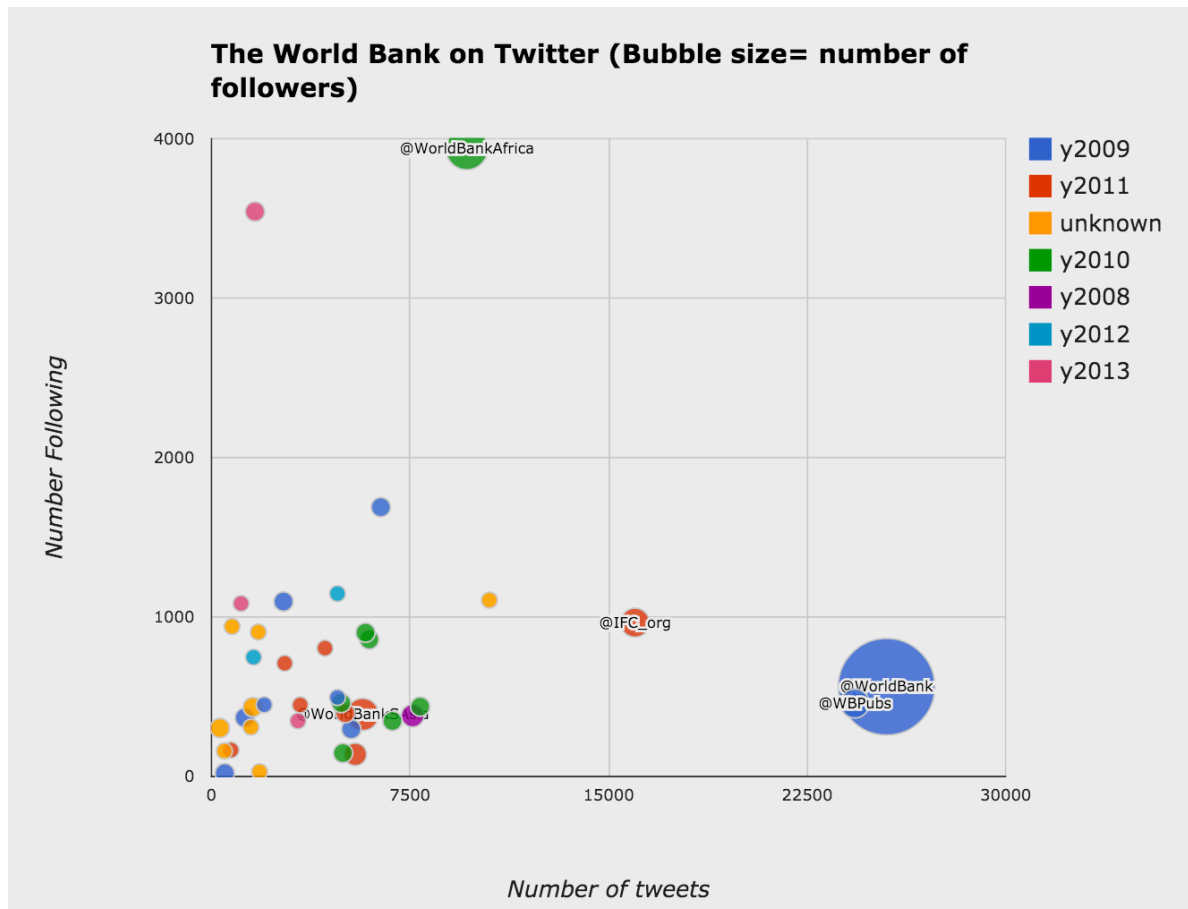
- 38 World Bank related Twitter accounts
- Total 93,733 tweets downloaded
  - Average of 336-3,200 tweets per account
  - Maximum tweets allowed to download is 3,200 per account
  - I downloaded my data in January/February 2016
- Keep only "own" tweets
  - By user, 6-47 percent of tweets are RTs from other users
  - 71,287 tweets remaining
  - 316 - 3,001 tweets per account
  - Data over the period March 2009 to February 2016

# AVAILABLE AND CREATED VARIABLES

API		Created
ID	User friends count	Dummy: has a hashtag (#)
text	User location	Dummy: is a retweet
created at		Dummy: Has a link
Retweet count		Dummy: Has an "@"
Favorite count		Dummy: user screen names
source		Dummy: year
Length		List of Hashtags (#)
user screen name		List of user tags (@)
user name		Clean text
user created at		Favorite/retweet count ratio
User description		
User followers count		

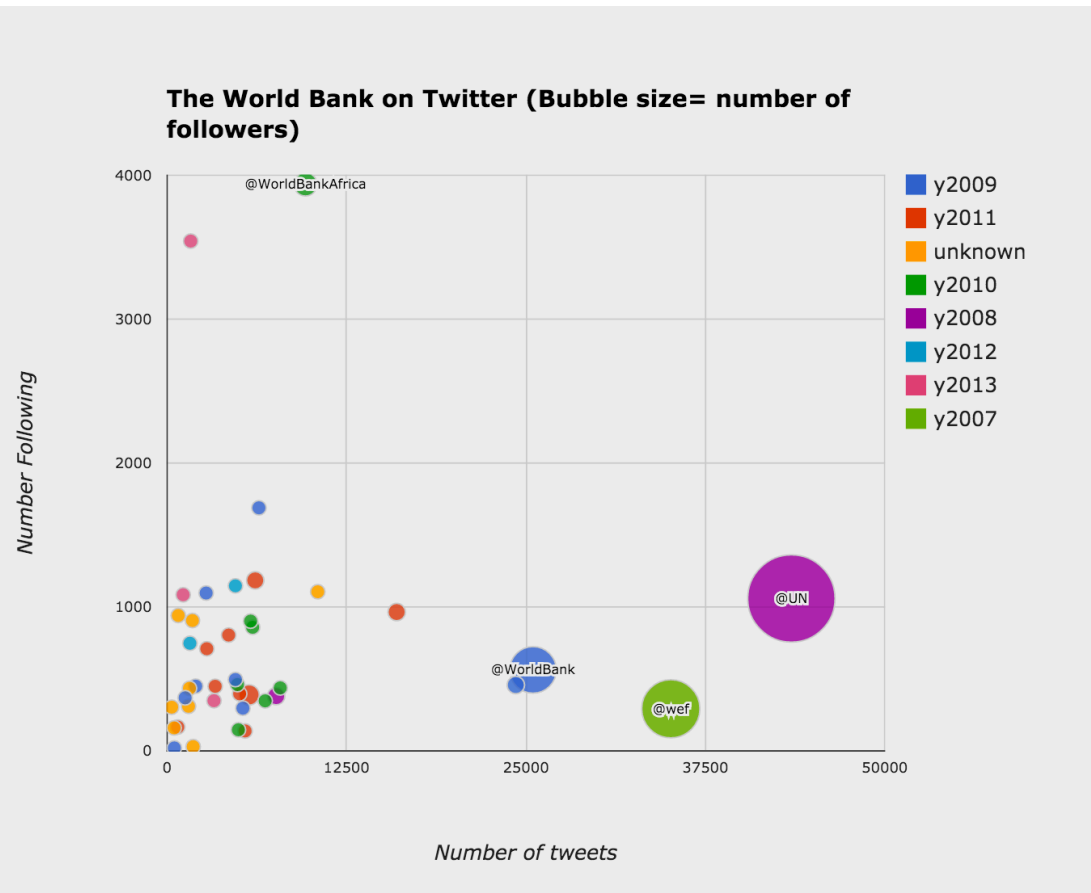
# WORLD BANK RELATED ACCOUNTS ON TWITTER: I FOUND 38

<a href="#">@WorldBank</a>	<a href="#">@WBG_Agriculture</a>	<a href="#">@WorldBank_IEG</a>	<a href="#">@WBG_Transport</a>
<a href="#">@WBPubs</a>	<a href="#">@WBG_Finance</a>	<a href="#">@WorldBankSAsia</a>	<a href="#">@WorldBankPhotos</a>
<a href="#">@IFC_org</a>	<a href="#">@WBG_LeadINLearn</a>	<a href="#">@wb_research</a>	<a href="#">@WBOpenFinances</a>
<a href="#">@WBG_Gender</a>	<a href="#">@WBG_PPP</a>	<a href="#">@WorldBankPSD</a>	<a href="#">@WorldBankECA</a>
<a href="#">@WorldBankAfrica</a>	<a href="#">@WorldBankVideos</a>	<a href="#">@WBG_Education</a>	<a href="#">@impacteval</a>
<a href="#">@WBG_Cities</a>	<a href="#">@MIGAWorldBank</a>	<a href="#">@wbg_gov</a>	<a href="#">@WBedutech</a>
<a href="#">@WBG_Climate</a>	<a href="#">@WorldBankIndia</a>	<a href="#">@WBG_Health</a>	<a href="#">@WBG_Dev4Peace</a>
<a href="#">@worldbankdata</a>	<a href="#">@WBG_SPLabor</a>	<a href="#">@WBG_Poverty</a>	<a href="#">@JimKim_WBG</a>
<a href="#">@WBG_Environment</a>	<a href="#">@WorldBankKenya</a>	<a href="#">@WorldBankSEI</a>	<a href="#">@WBG_Energy</a>
	<a href="#">@WBCaribbean</a>	<a href="#">@WorldBankWater</a>	



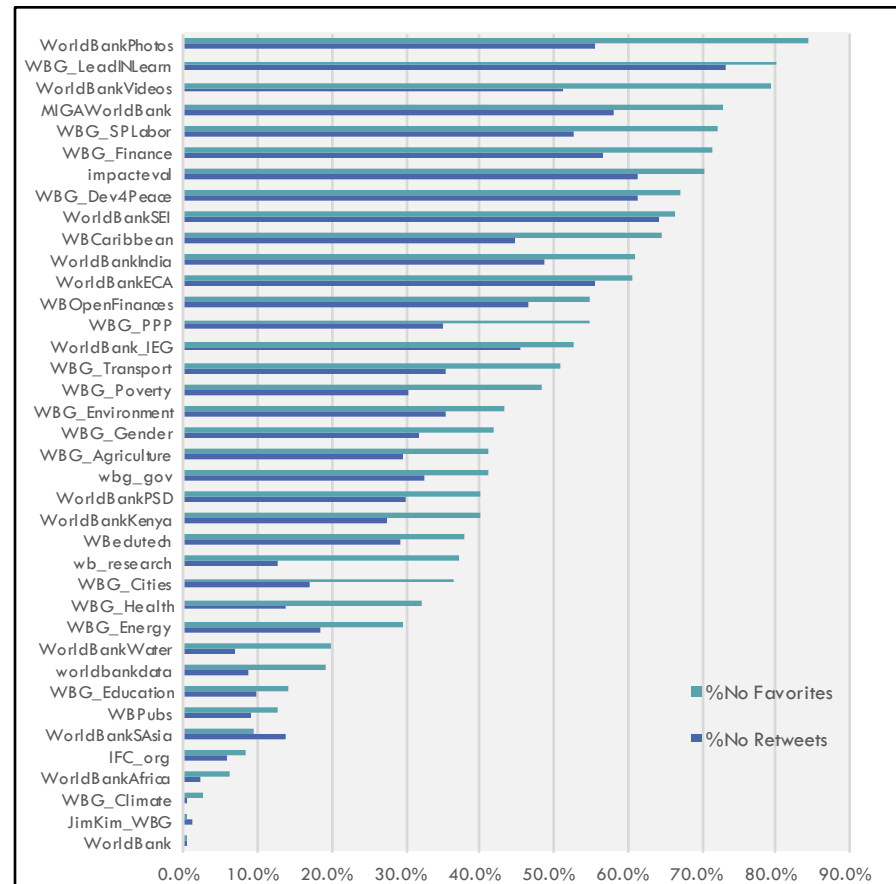
Source: Collected February 1, 2016





# PERCENT OF TWEETS WITHOUT FAVORITES OR RETWEETS

The World Bank's main account and the President's account have the most retweet and favorite activity.



# TOP @S

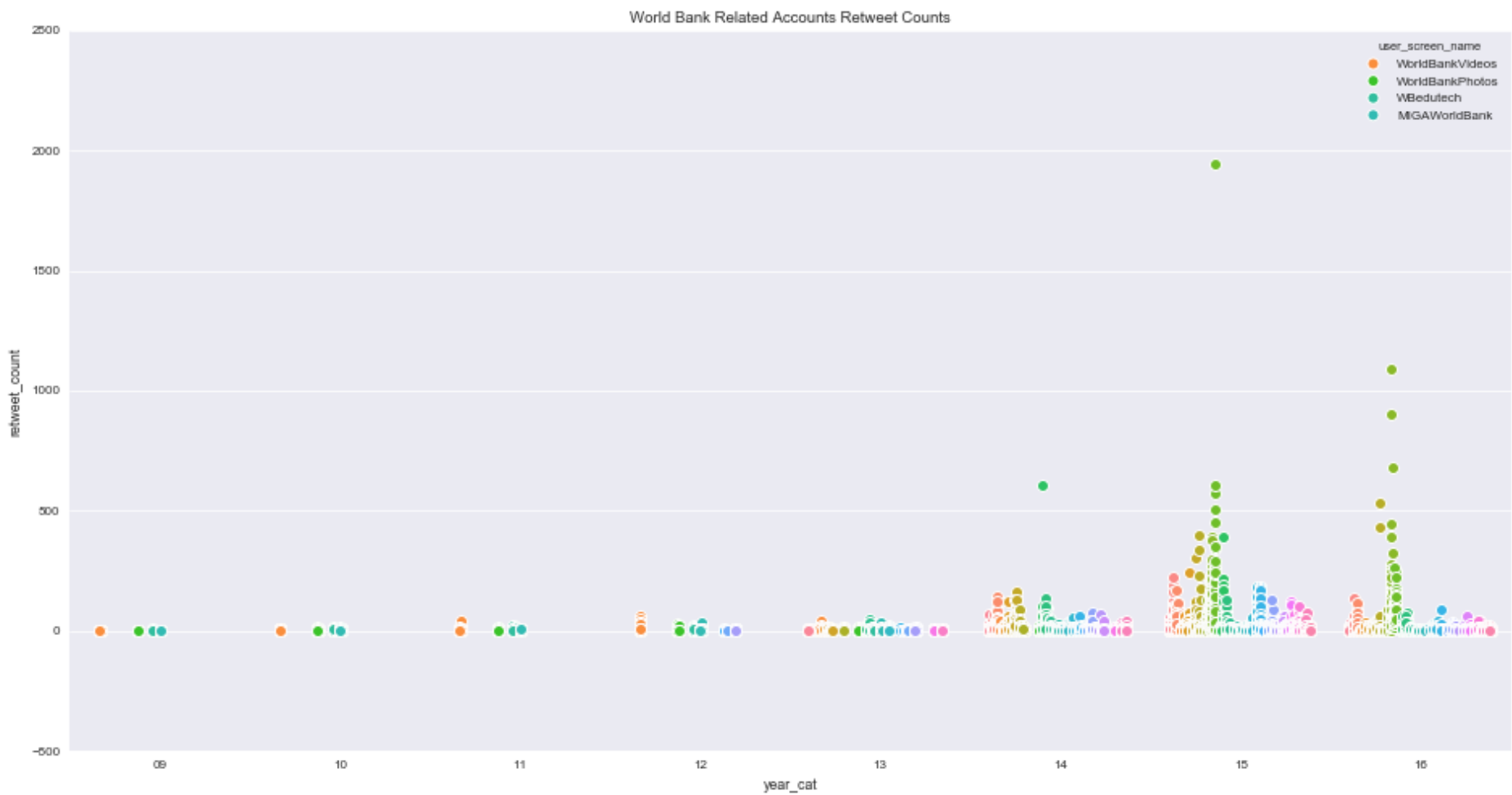
@user (71,287 tweets)		@user (retweets > 50, 1757 tweets)		@user (favorites > 50, 944 tweets)	
4,773	worldbank	110	jimkim_wbg	70	jimkim_wbg
796	jimkim_wbg	44	worldbank	14	worldbank
591	ifc_org	11	djokernole	9	djokernole
291	kaushikbasu	9	worldbanklive	6	mzveegh
274	rkyte365	8	un_women	5	malalafund
237	guardian	8	mzveegh	4	un_women
235	worldbankdata	6	malalafund	4	worldbanklive
226	worldbankafrica	5	cgap	4	thegef
210	who	5	worldbankrwanda	3	narendramodi
210	un	5	un	3	worldbankasia

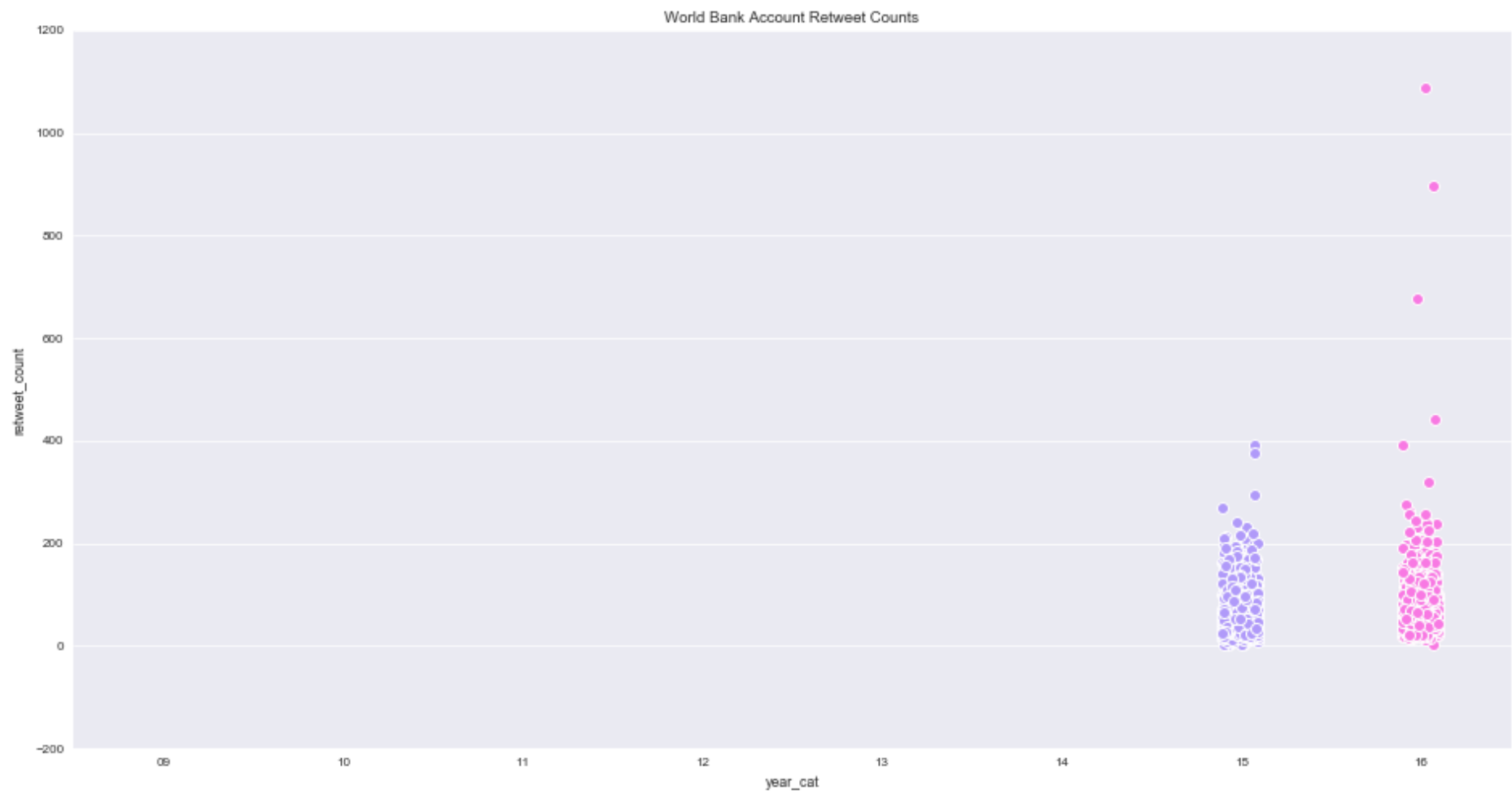
# TOP WORDS

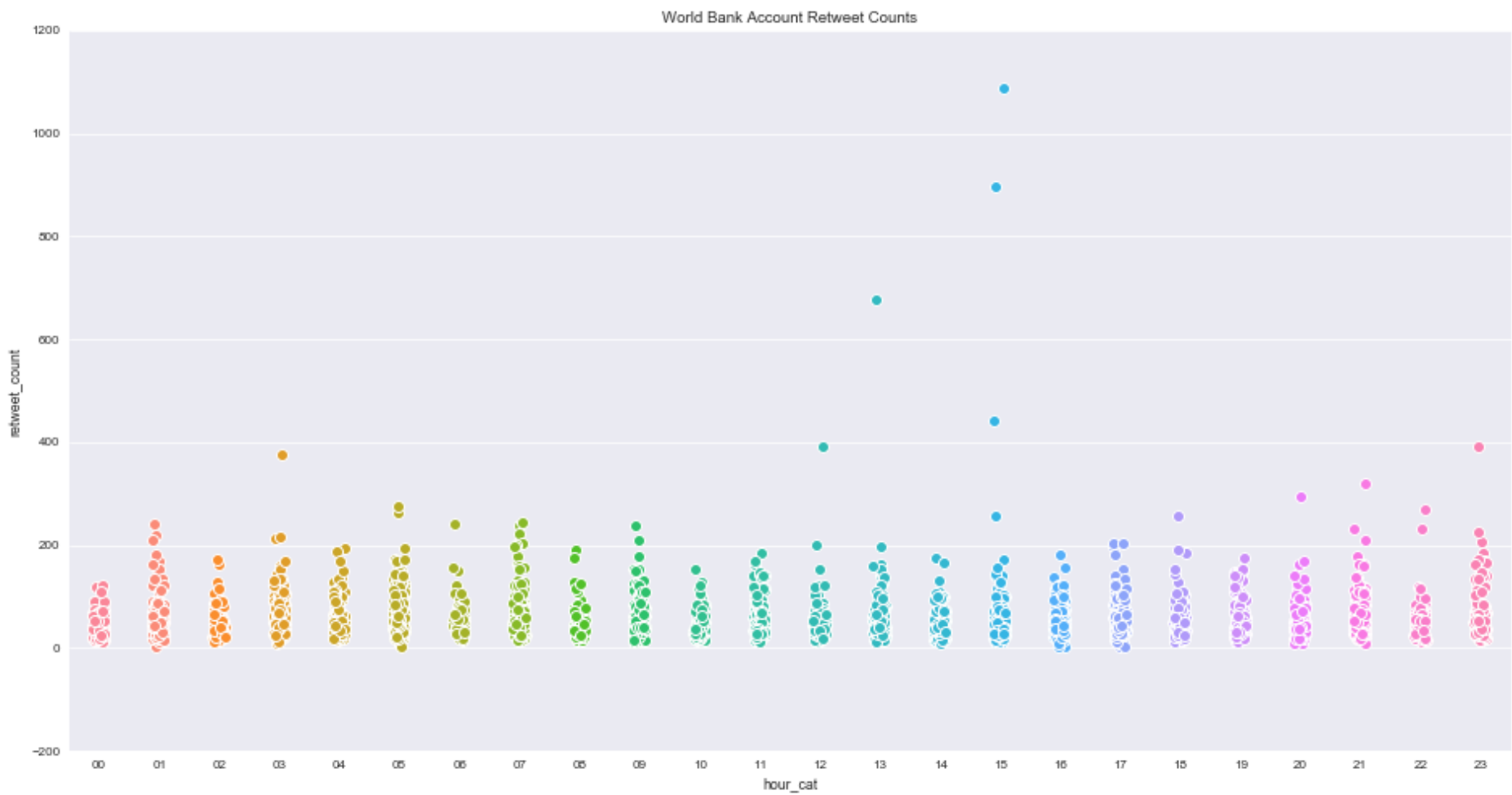
words		words (retweets > 50, 1757 tweets)		words (favorites > 50, 944 tweets)	
5068	new	214	poverty	129	poverty
3620	world	130	women	88	cop21
3383	countries	124	countries	77	wbg
3345	help	122	wbg	76	women
3312	development	119	cop21	67	countries
3308	women	115	africa	63	world
3188	climate	107	climatechange	61	endpoverty
3152	poverty	102	world	57	help
3139	blog	102	endpoverty	56	access
2920	africa	101	ppl	54	people

# TOP HASHTAGS

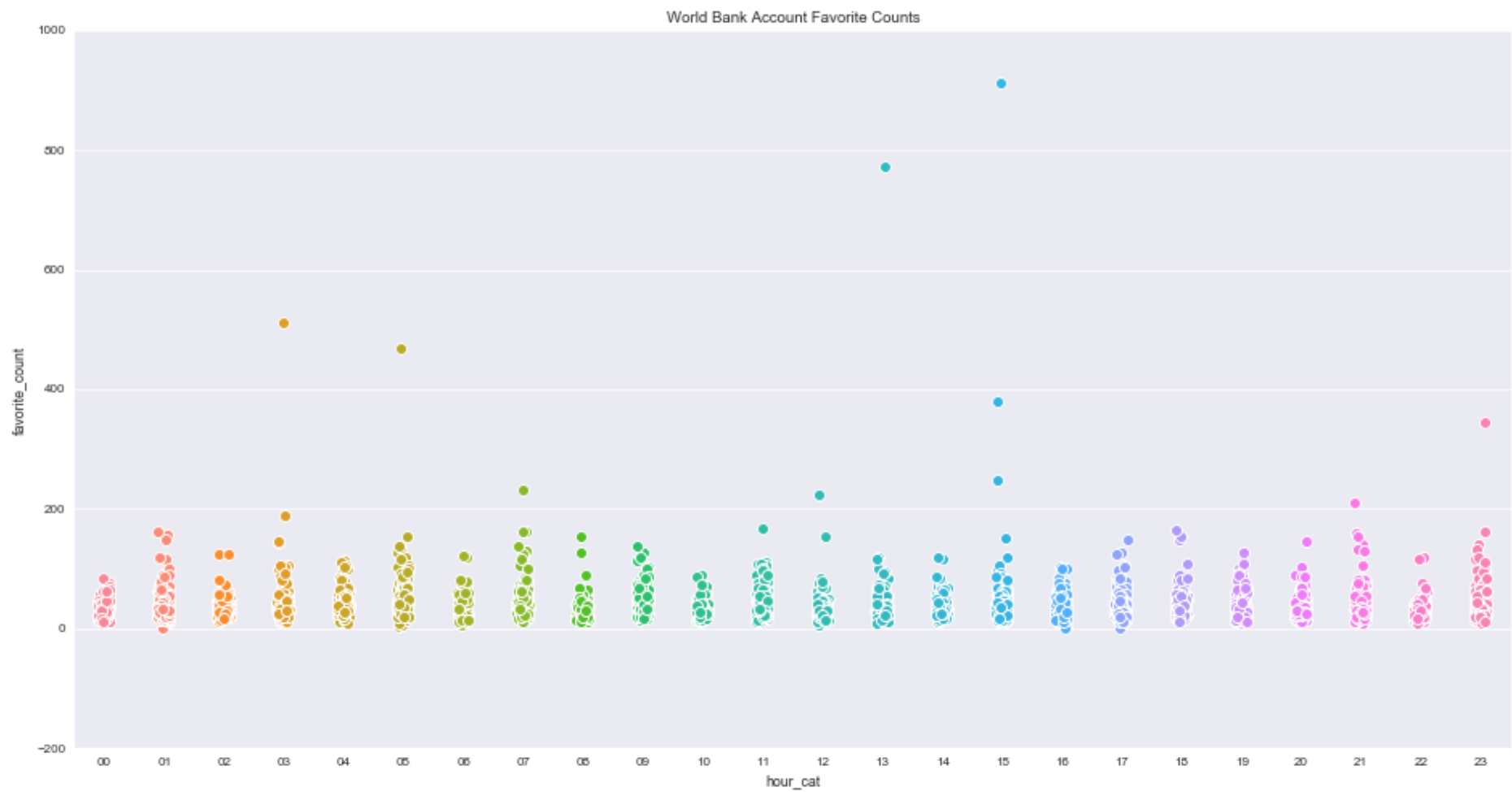
hashtags		hashtags (retweets > 50, 1757 tweets)		hashtags (favorites > 50, 944 tweets)	
2,314	africa	116	cop21	87	cop21
2,172	india	107	climatechange	61	endpoverty
1,689	climatechange	102	endpoverty	46	poverty
1,561	socent	97	africa	38	africa
1,465	opendata	80	poverty	36	climatechange
1,353	ppps	44	climate	24	education
1,278	endpoverty	35	education	20	climate
1,266	poverty	31	data	20	india
1,225	climate	29	india	20	wdr2016
1,224	cop21	28	wdr2016	16	solar



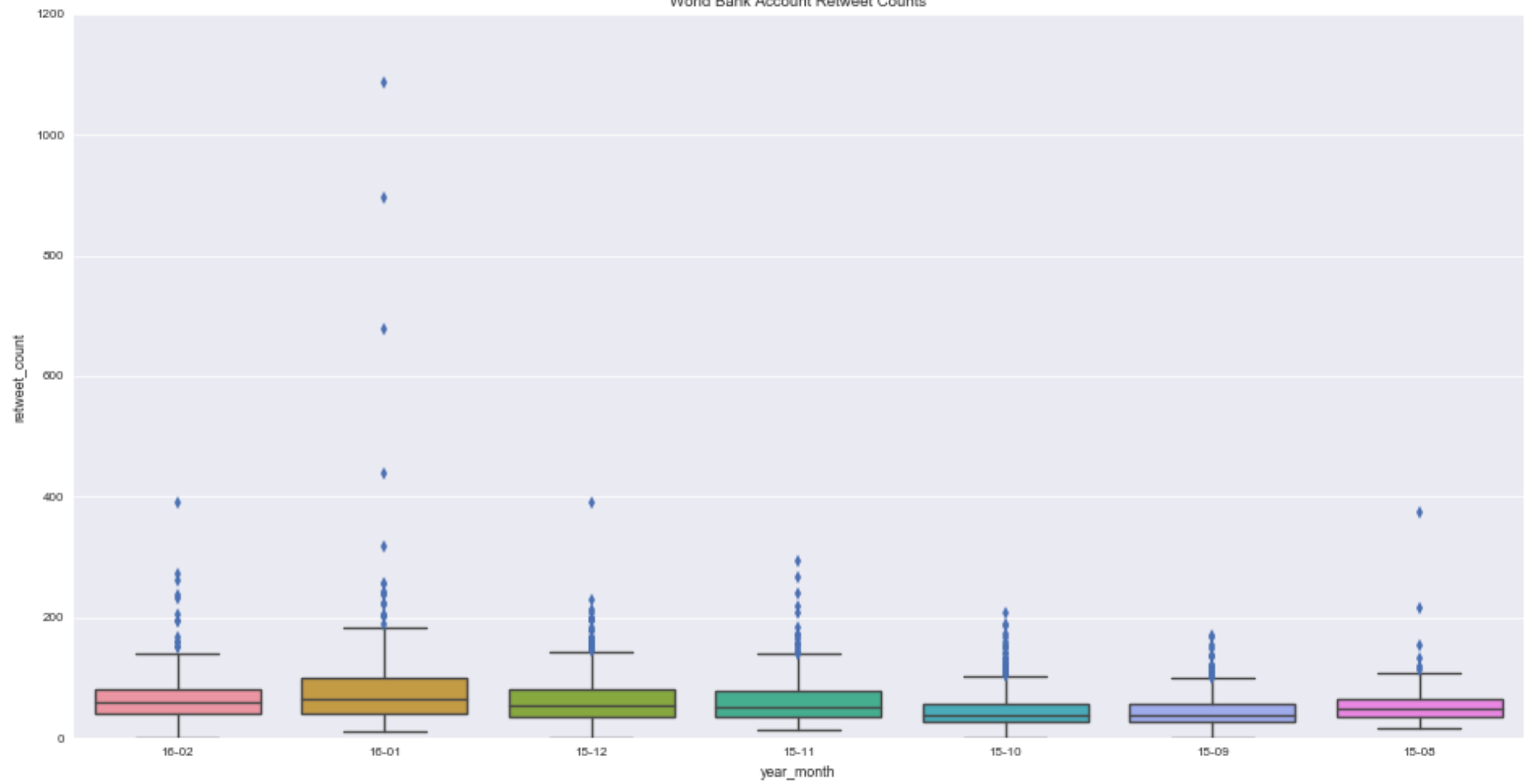








World Bank Account Retweet Counts





# ANALYSIS

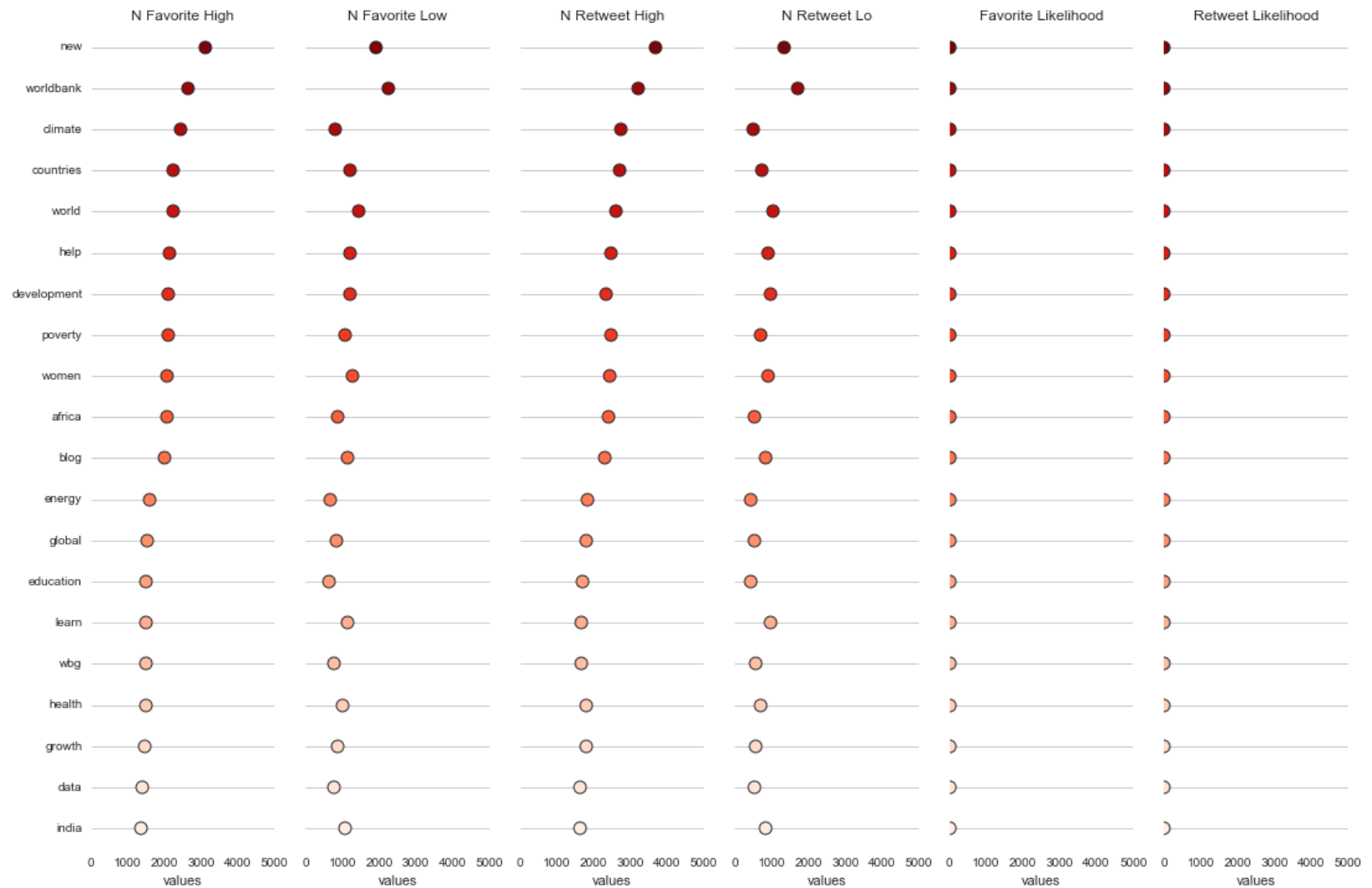
# ANALYSIS

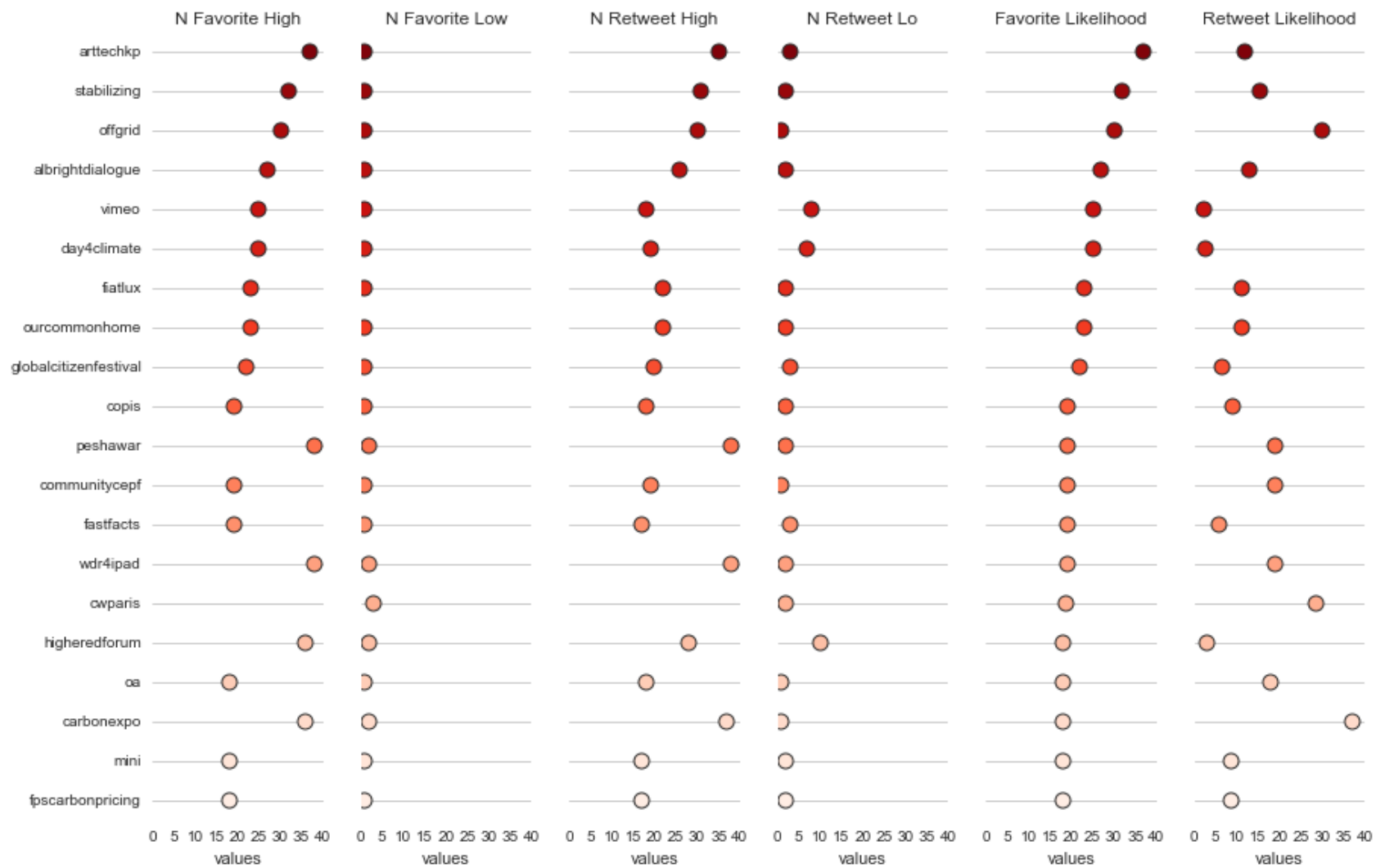
- Word counts and Token likelihood
  - What words are more or less common in favorite and RTs
- Topic Modelling
  - Which types of tweets get the most retweets or favorites?
- Predictions
  - Non-text (has @, has link, has #)
  - Text based
- Sentiment analysis
  - People retweet crisis but don't favorite crisis
  - Sentiment, who has the happiest account?
  - No relationship between sentiment and retweet or favorite counts

# TOKEN LIKELIHOODS:

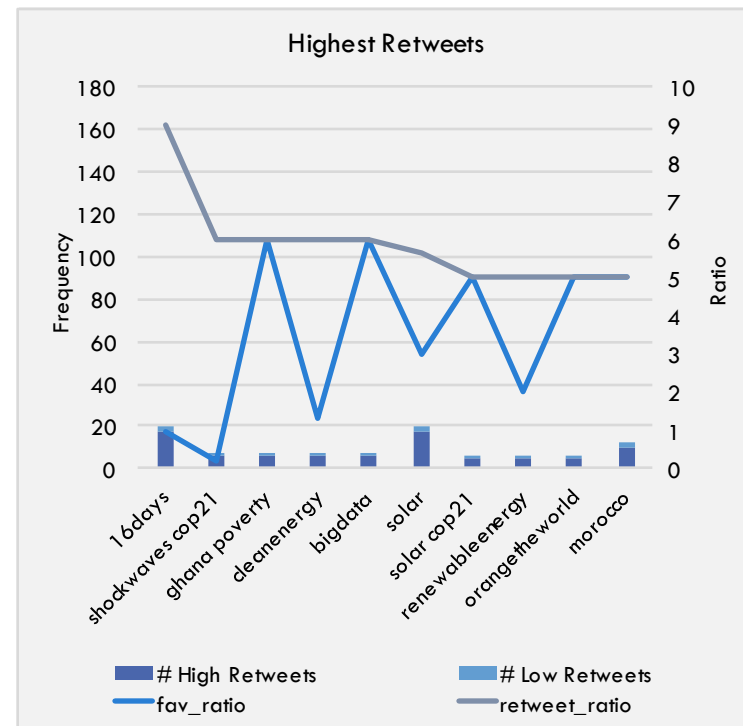
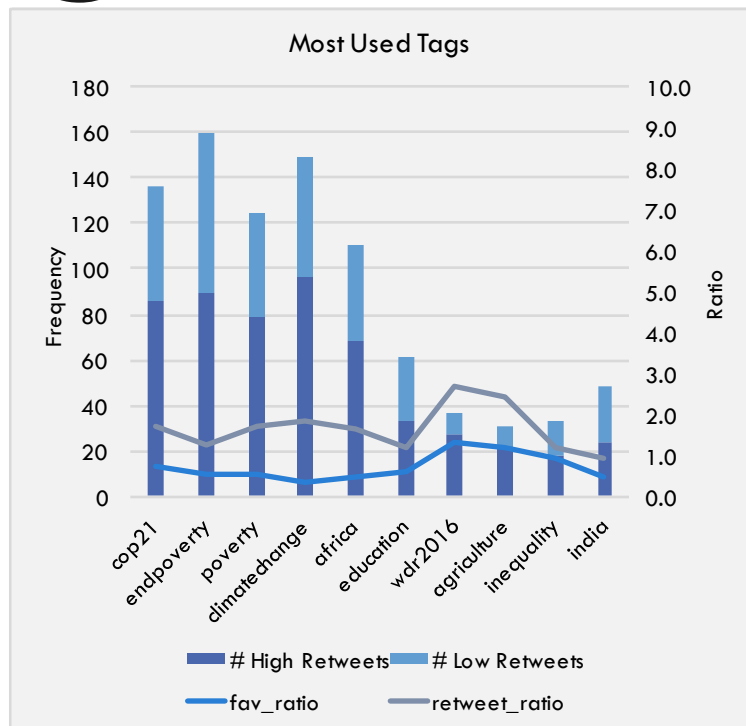
## WHICH WORDS ARE IN THE HIGH FAVORITE AND RETWEET COUNTS

1. Most common words are not revealing of popular tweets, they are overused (poverty, world bank)
2. Likelihood ratios are not as high as in the spam/ham example
3. @WorldBank:
  1. 713 tags
  2. 1,980 tag pairs
  3. 4,837 words





# @WORLDBANK





# TOPIC ANALYSIS

## 10 topics, all tweets, all WB accounts, based on tags

[u'0.030\*women + 0.026\*violence + 0.015\*conflict + 0.015\*16days + 0.011\*womenbizlaw + 0.010\*men + 0.010\*week + 0.010\*online + 0.009\*challenges + 0.009\*register',  
u'0.026\*financial + 0.023\*data + 0.021\*development + 0.012\*join + 0.010\*worldbank + 0.010\*close + 0.009\*opendata + 0.009\*new + 0.009\*global + 0.009\*inclusion',  
u'0.029\*climate + 0.026\*wbq + 0.023\*energy + 0.016\*agriculture + 0.016\*change + 0.013\*credit + 0.012\*ecourse + 0.011\*sign + 0.010\*economic + 0.010\*sustainable',  
u'0.019\*women + 0.016\*entrepreneurs + 0.013\*sdgs + 0.011\*achieve + 0.010\*genderequality + 0.009\*video + 0.008\*lives + 0.007\*like + 0.007\*reach + 0.007\*expand',  
u'0.030\*join + 0.026\*worldbank + 0.022\*poverty + 0.020\*webinar + 0.020\*live + 0.013\*growth + 0.012\*event + 0.012\*watch + 0.009\*discussion + 0.009\*est',  
u'0.025\*ecourse + 0.019\*learn + 0.018\*apply + 0.014\*foodsecurity + 0.013\*africa + 0.013\*wbq + 0.013\*agriculture + 0.012\*remittances + 0.012\*agricultural + 0.012\*register',  
u'0.025\*education + 0.024\*gender + 0.023\*women + 0.019\*get2equal + 0.014\*girls + 0.011\*countries + 0.011\*children + 0.010\*school + 0.010\*blog + 0.010\*youth',  
u'0.037\*ifc + 0.027\*food + 0.026\*org + 0.022\*climatechange + 0.013\*countries + 0.010\*developing + 0.009\*mobile + 0.009\*global + 0.009\*disasters + 0.008\*guardian',  
u'0.036\*financialinclusion + 0.021\*farmers + 0.020\*finaccess + 0.020\*women + 0.019\*access + 0.017\*worldbank + 0.016\*india + 0.015\*financial + 0.015\*services + 0.014\*finance',  
u'0.026\*new + 0.018\*sector + 0.014\*private + 0.014\*worldbank + 0.011\*jobs + 0.011\*urban + 0.011\*gender + 0.011\*help + 0.010\*learn + 0.010\*know']

## 10 topics, tweets with at least one retweet, all WB accounts, based on text\_clean

[u'0.027\*finaccess + 0.022\*private + 0.019\*ppps + 0.019\*new + 0.018\*worldbank + 0.018\*sector + 0.016\*public + 0.016\*data + 0.014\*remittances + 0.014\*financialinclusion',  
u'0.023\*violence + 0.021\*16days + 0.020\*worldbank + 0.013\*know + 0.012\*genderequality + 0.012\*corruption + 0.011\*live + 0.010\*strategy + 0.010\*wblive + 0.009\*want',  
u'0.037\*ifc + 0.025\*org + 0.021\*finance + 0.018\*sms + 0.016\*new + 0.016\*countries + 0.015\*credit + 0.011\*growth + 0.010\*low + 0.010\*worldbank',  
u'0.019\*worldbank + 0.013\*endenergypoverity + 0.012\*food + 0.012\*poverty + 0.012\*apply + 0.011\*youth + 0.010\*agriculture + 0.010\*read + 0.008\*school + 0.008\*gender',  
u'0.017\*education + 0.013\*learning + 0.012\*development + 0.011\*new + 0.011\*challenges + 0.009\*ag + 0.009\*women + 0.009\*conflict + 0.008\*financing + 0.008\*cash',  
u'0.033\*join + 0.014\*discussion + 0.013\*research + 0.013\*live + 0.012\*event + 0.011\*webinar + 0.010\*soil + 0.009\*nigeria + 0.008\*growing + 0.008\*development',  
u'0.049\*women + 0.019\*financialinclusion + 0.018\*gender + 0.018\*financial + 0.016\*wbq + 0.015\*help + 0.013\*get2equal + 0.012\*access + 0.010\*insurance + 0.009\*countries',  
u'0.029\*climate + 0.026\*agriculture + 0.022\*farmers + 0.016\*change + 0.014\*climatechange + 0.013\*foodsecurity + 0.011\*help + 0.011\*food + 0.009\*land + 0.008\*world',  
u'0.015\*new + 0.014\*education + 0.013\*women + 0.010\*africa + 0.009\*improve + 0.009\*quality + 0.009\*report + 0.009\*children + 0.008\*improving + 0.008\*access',  
u'0.040\*energy + 0.019\*india + 0.014\*solar + 0.014\*water + 0.014\*africa + 0.012\*entrepreneurs + 0.012\*mt + 0.011\*access + 0.011\*rural + 0.010\*power']

# PREDICTIONS

Logistic regression

$Y$  = Dummy has a retweet

$X_1$  = has@, has link, has#

$X_2$  = has@, has link, has#, dummy of WB accounts

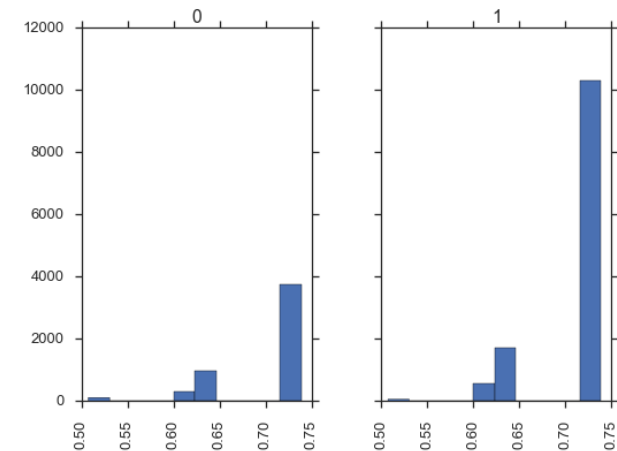
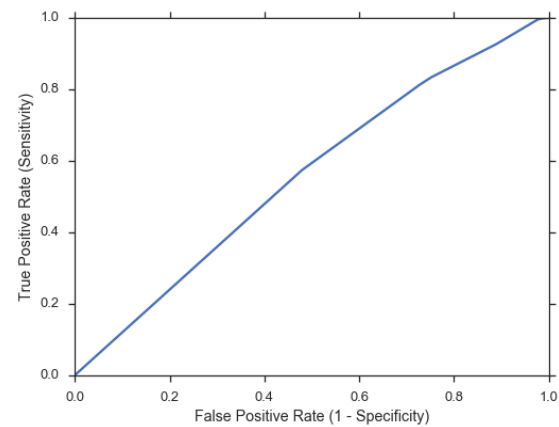
Y=has a retweet  
Cols=['has\_at', 'has\_ht',  
'has\_link']

Testing accuracy: 0.711311

Null accuracy: 0.711312

AUC: 0.557

C-V AUC: 0.547



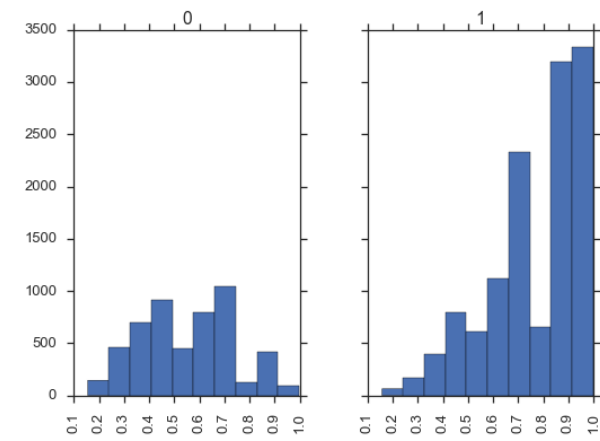
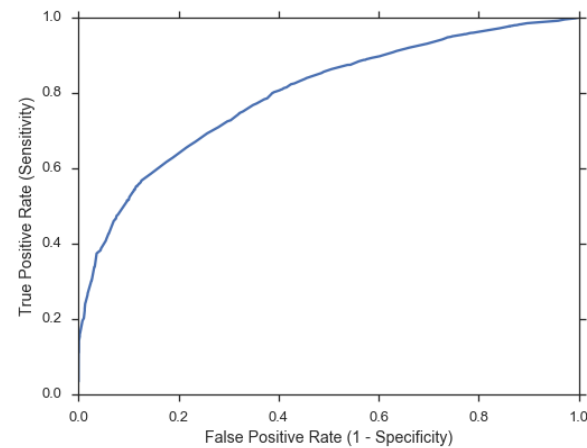
Y=has a retweet  
Cols=['has\_at', 'has\_ht',  
'has\_link', all user dummies]

Testing accuracy: 0.755

Null accuracy: 0.71131

AUC: 0.795

C-V AUC: 0.215



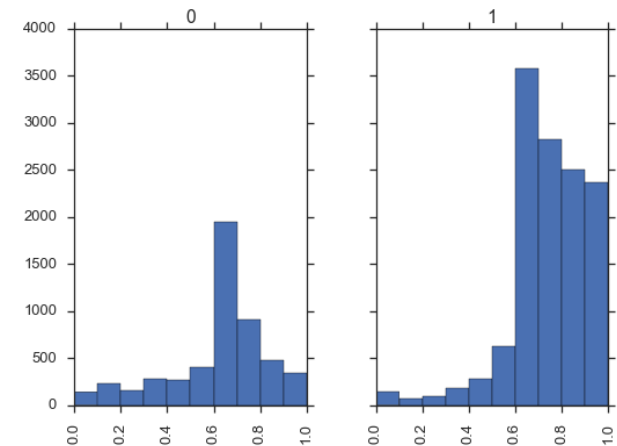
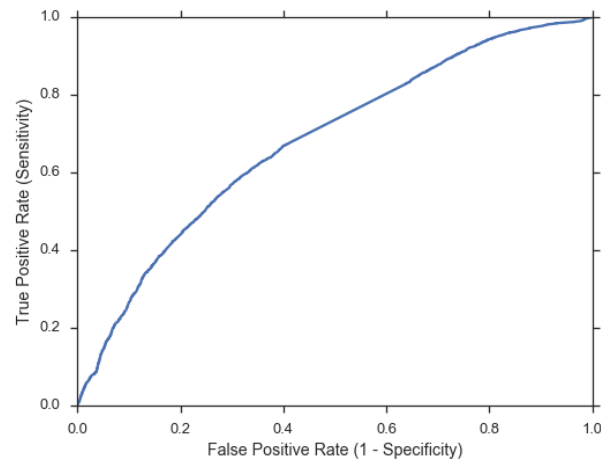
Y=has a retweet  
Cols=[document term matrix]

Testing accuracy: 0.727

Null accuracy: 0.711312

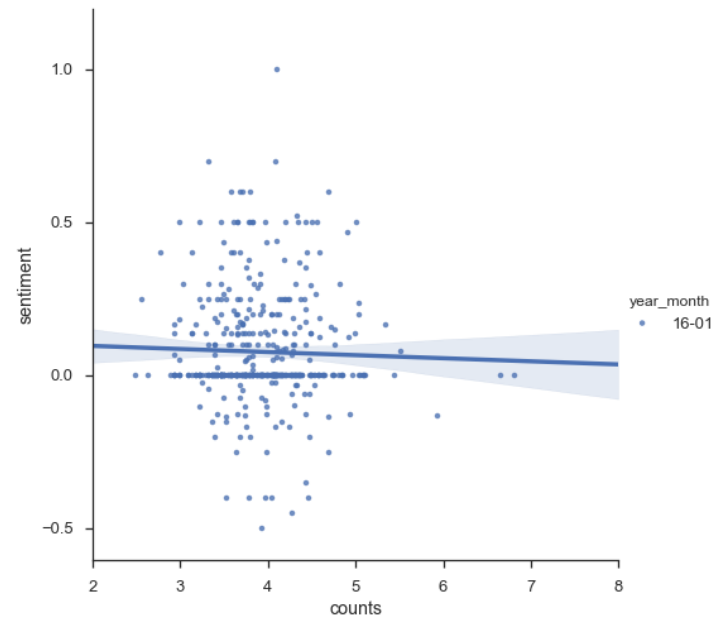
AUC: 0.679

C-V AUC:



- Columns of characteristics (@, #, link) is not predictive
- Words predict retweet incidence OK but not as well as user dummies
- Comparison of the main World Bank twitter account is not really comparable with others so focus only on the account @WorldBank for the text analysis

# SENTIMENT





# CONCLUSIONS

# RECOMMENDATIONS

1. Need to interact more broadly with people
2. Cross tag people and hashtags
3. Do live event tweeting
4. Sub-groups should always tag world bank, but then there is the issue of fatigue and over tweeting, does this bother users?





**THANK YOU!**

Judy: [judy.s.yang@gmail.com](mailto:judy.s.yang@gmail.com)