

THE WORLD BANK PRESENCE ON TWITTER

Judy S. Yang

February 24, 2016

General Assembly DAT-DC-10, Final Project

OUTLINE

Intro/Descriptives

Analysis

- Token likelihoods
- Topic Modeling
- Predictive models
- Sentiment Analysis

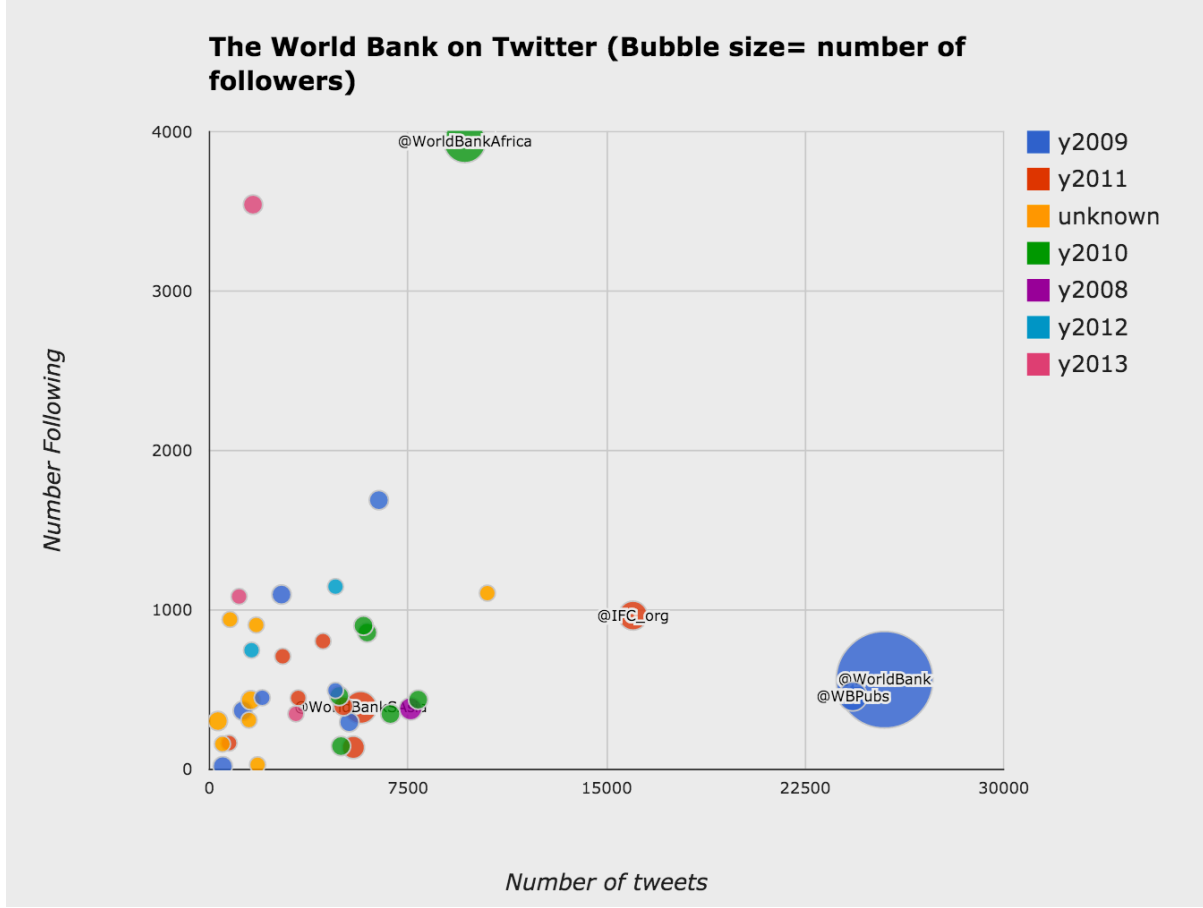
Conclusion

QUESTIONS

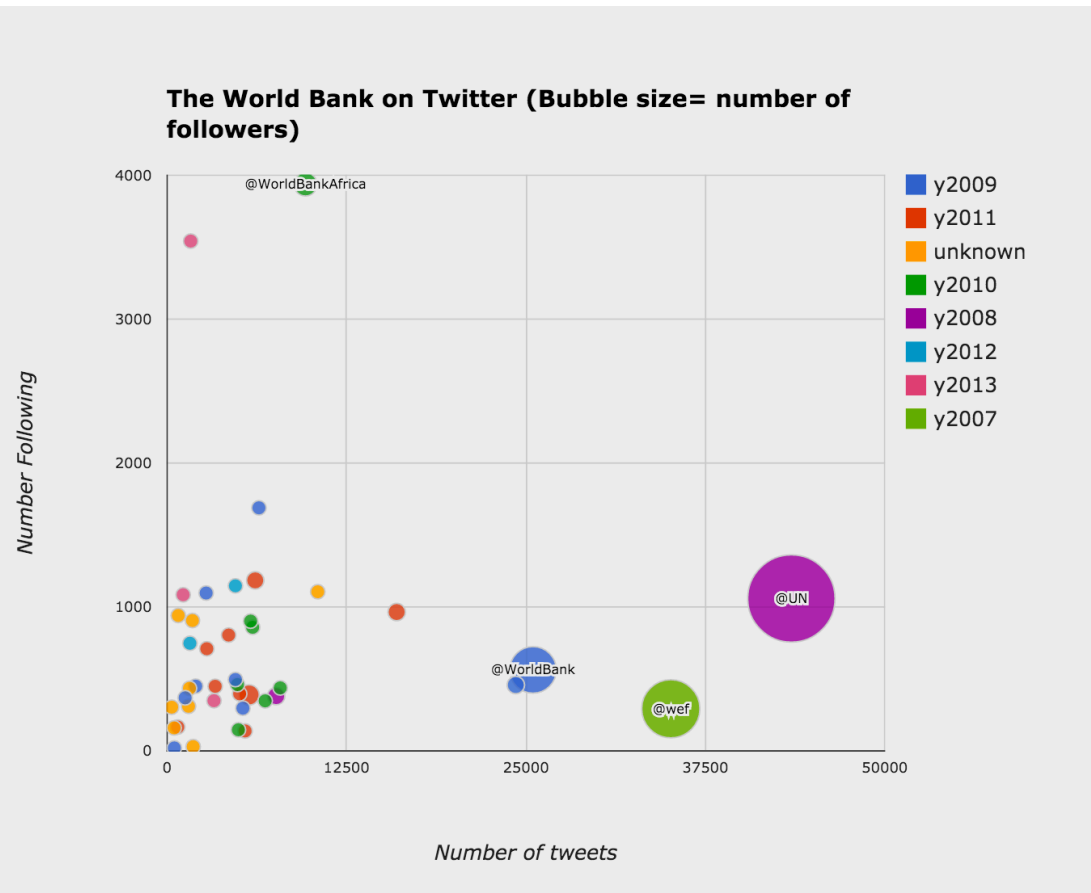
1. Which accounts are the most popular?
2. Are there too many WB Twitter accounts?
3. How can the WB improve its Twitter presence?
4. What do people tweet about

WORLD BANK RELATED ACCOUNTS ON TWITTER: I FOUND 38

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



Source: Collected February 1, 2016





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

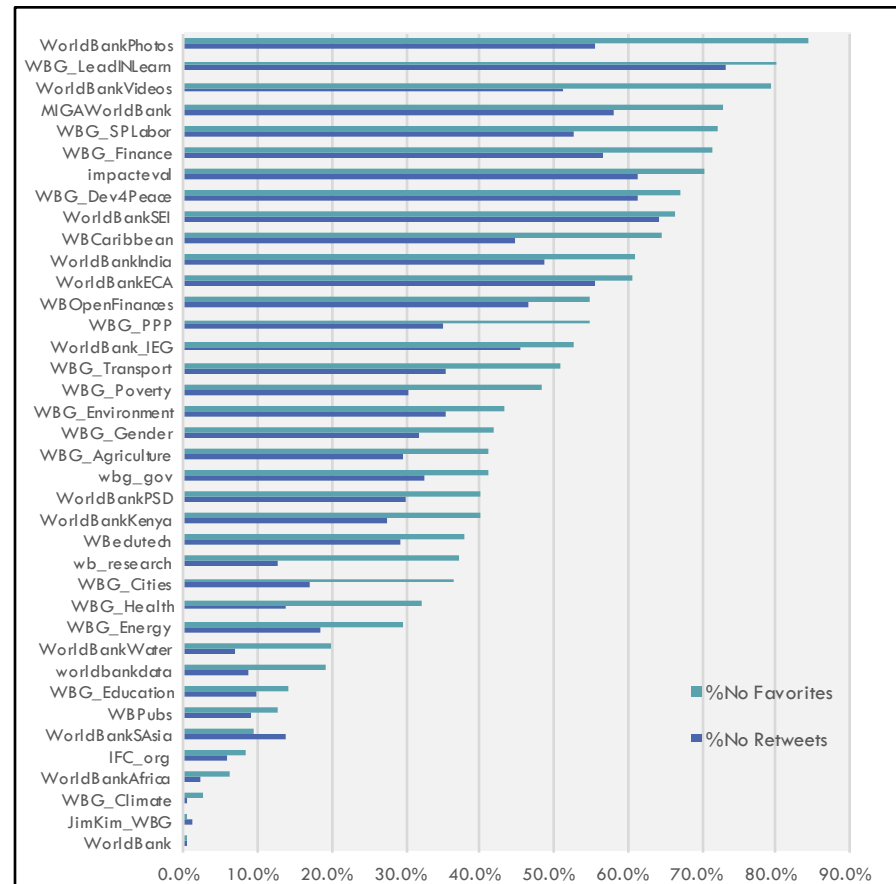
API		Created
ID	User friends count	Dummy: has a hashtag (#)
text	User location	Dummy: is a retweet
created at		Dummy: Has a link
Retweet count		Dummy: Has an "@"
Favorite count		Dummy: user screen names
source		Dummy: year
Length		List of Hashtags (#)
user screen name		List of user tags (@)
user name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

World Bank Related Accounts Retweet Counts



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



ANALYSIS

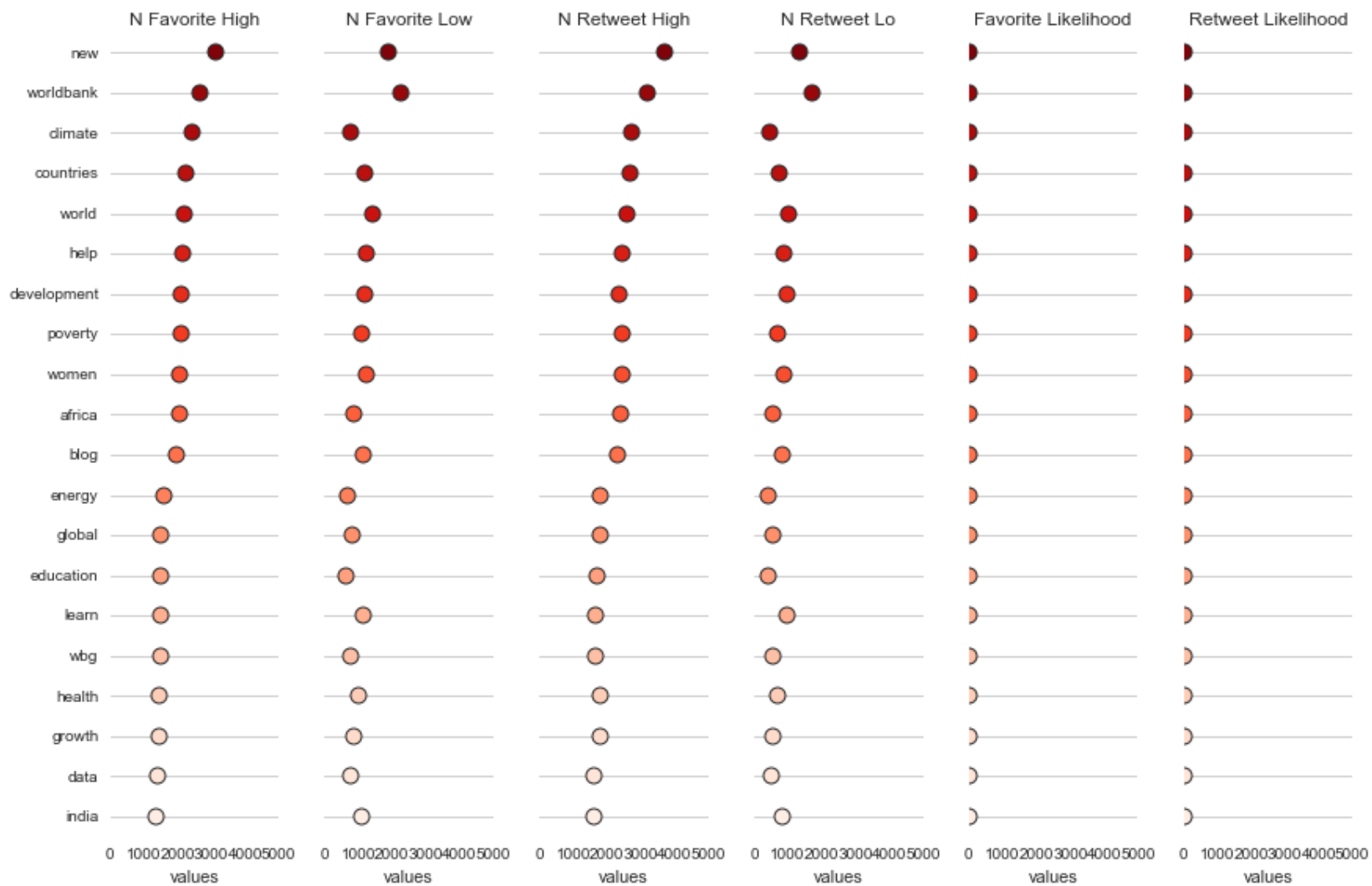
ANALYSIS

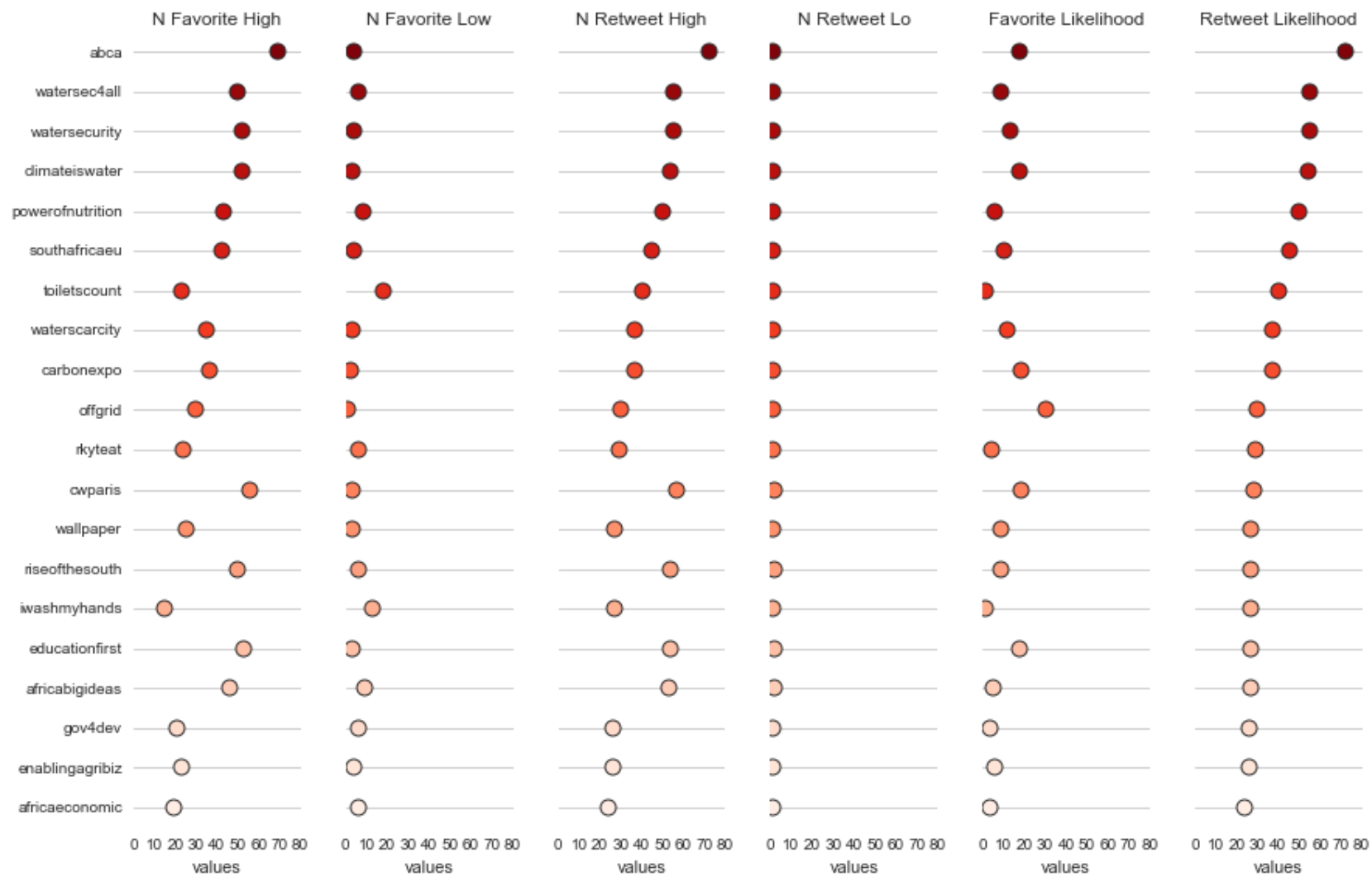
- Word counts and Token likelihood
 - What words are more or less common in favorite and RTs
- Topic Modelling
 - Which types of tweets get the most retweets or favorites?
- Predictions
 - Non-text (has @, has link, has #)
 - Text based
- Sentiment analysis

TOKEN LIKELIHOODS:

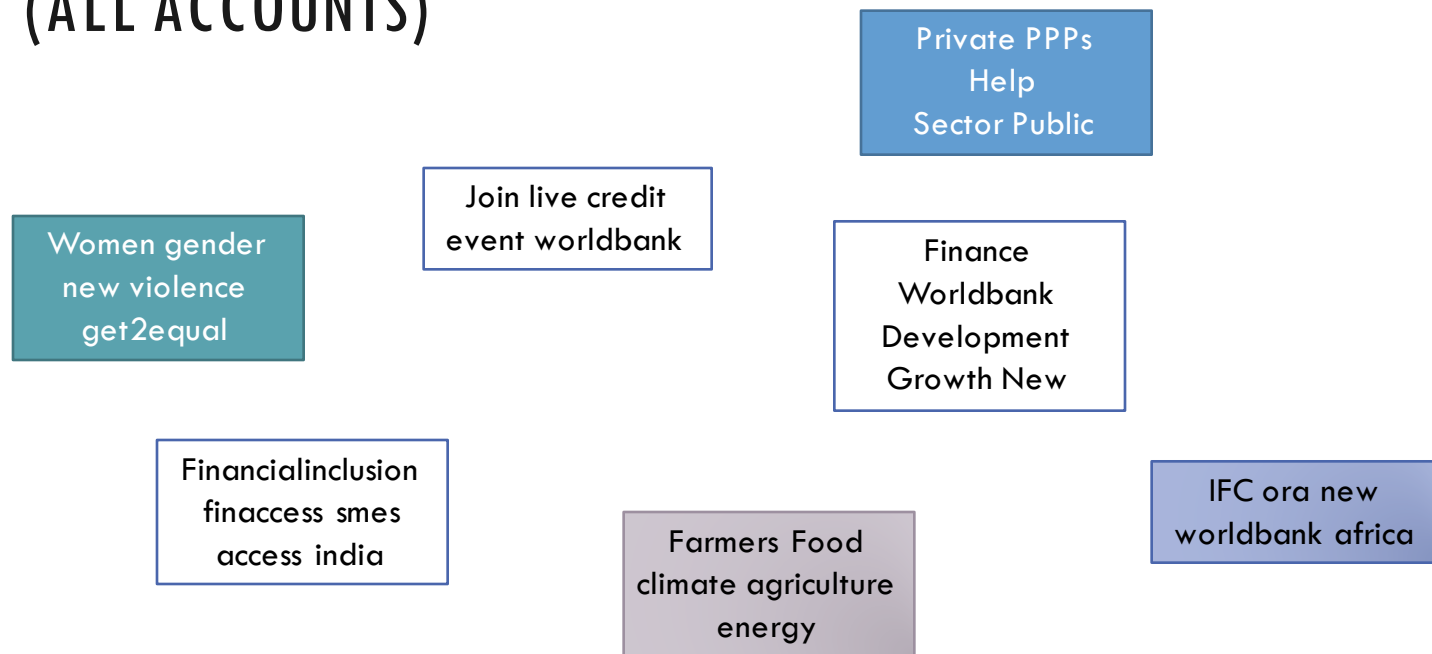
WHICH WORDS ARE IN THE HIGH FAVORITE AND RETWEET COUNTS

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. All accounts
 1. 5213 words
4. @WorldBank:
 1. 713 words





TOPIC MODELING (ALL ACCOUNTS)



PREDICTIONS

Logistic regression

Y= Dummy has a retweet

X1=has@, has link, has#

X2=has@, has link, has#, dummy of WB accounts

X3= document term matrix of tags

X4= document term matrix of cleaned text (note text includes tags)

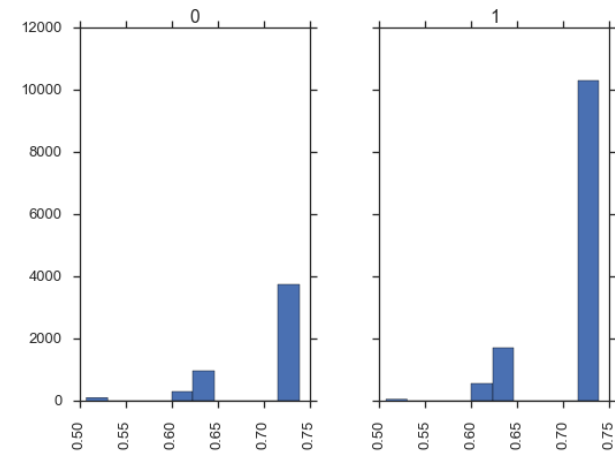
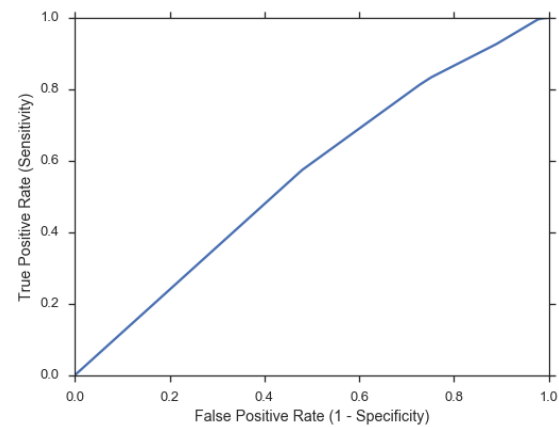
Y=has a retweet
Cols=['has_at', 'has_ht',
'has_link']

Testing accuracy: 0.711311

Null accuracy: 0.711312

AUC: 0.557

C-V AUC: 0.547



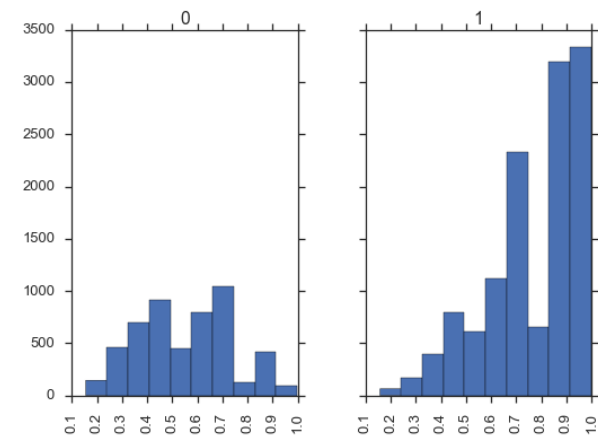
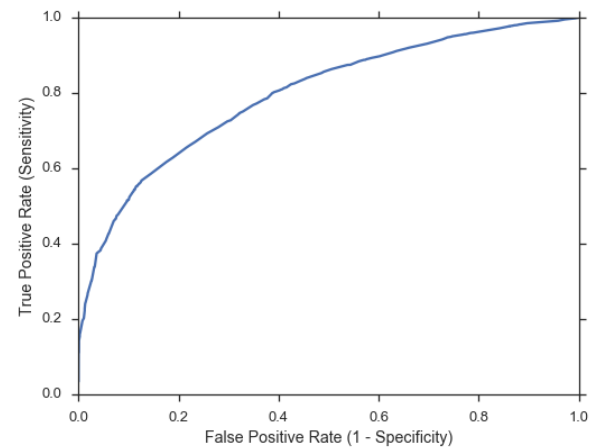
Y=has a retweet
Cols=['has_at', 'has_ht',
'has_link', all user dummies]

Testing accuracy: 0.755

Null accuracy: 0.71131

AUC: 0.795

C-V AUC: 0.215



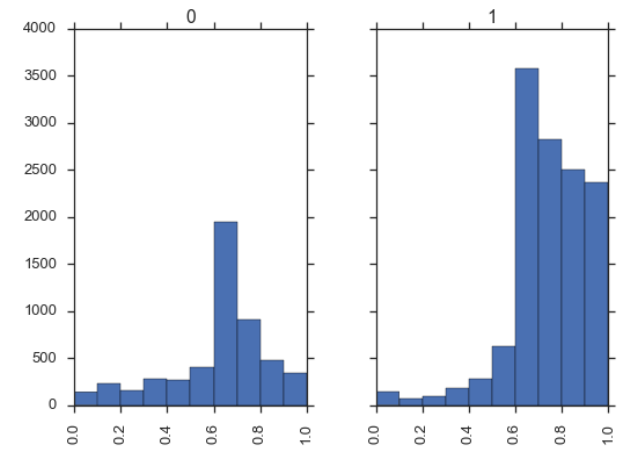
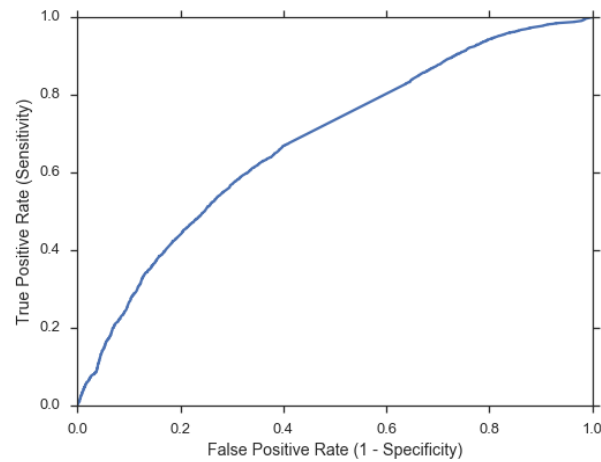
Y=has a retweet
Cols=[document term matrix
tags]

Testing accuracy: 0.727

Null accuracy: 0.711312

AUC: 0.679

C-V AUC:



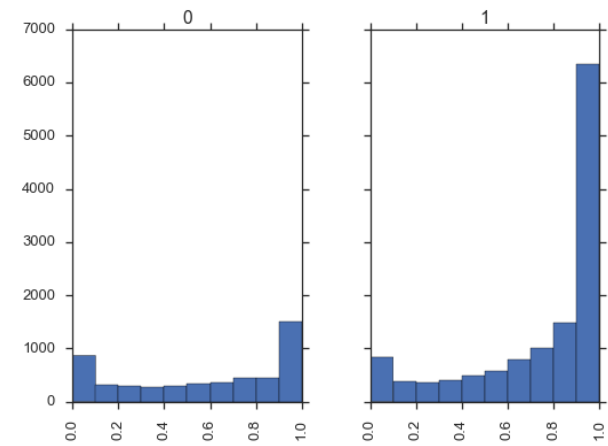
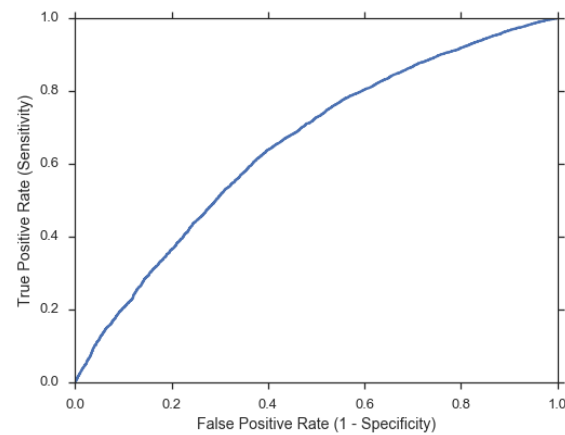
Y=has a retweet
Cols=[document term matrix
text]

Testing accuracy: 0.687

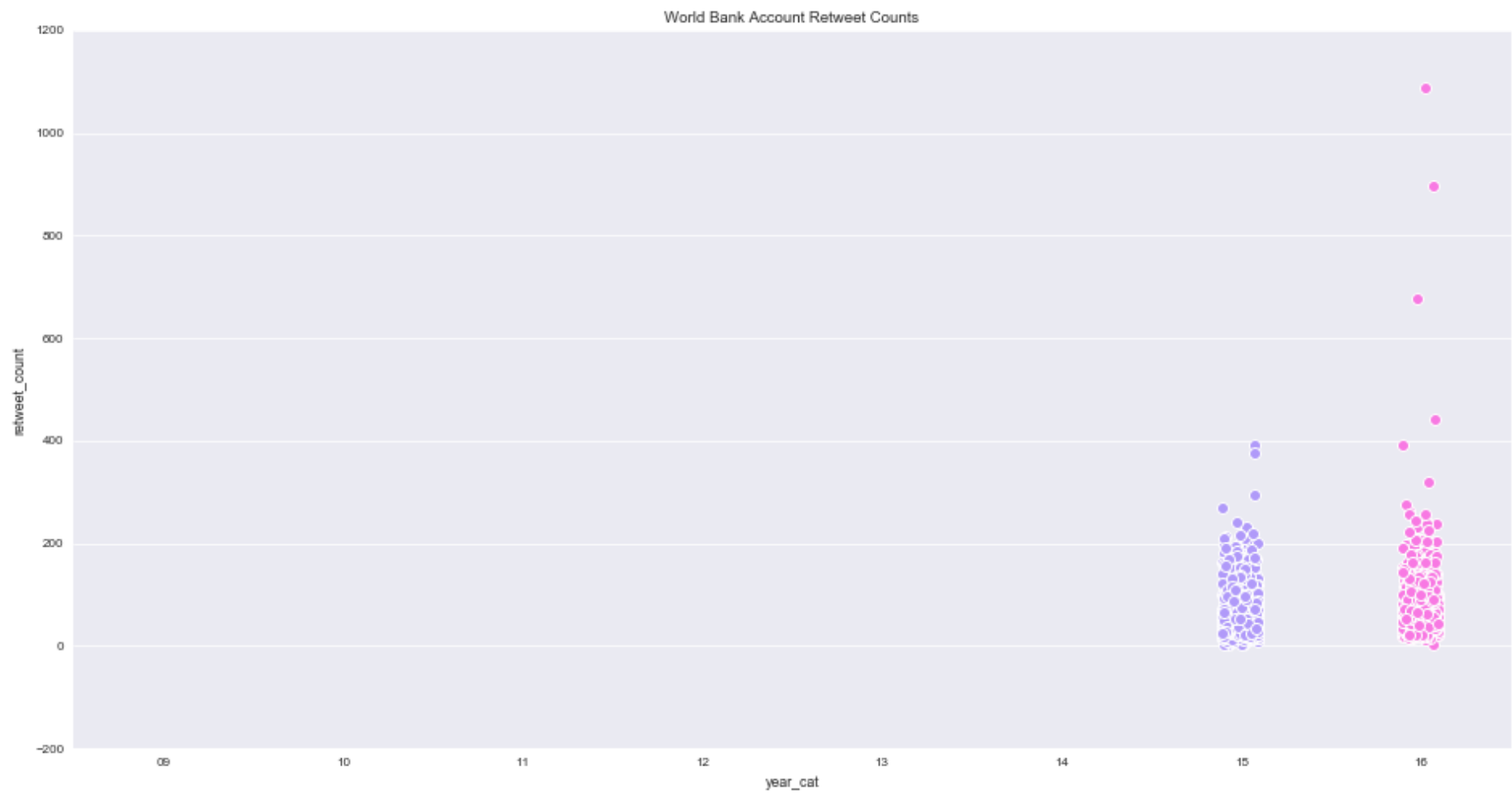
Null accuracy: 0.711312

AUC: 0.651

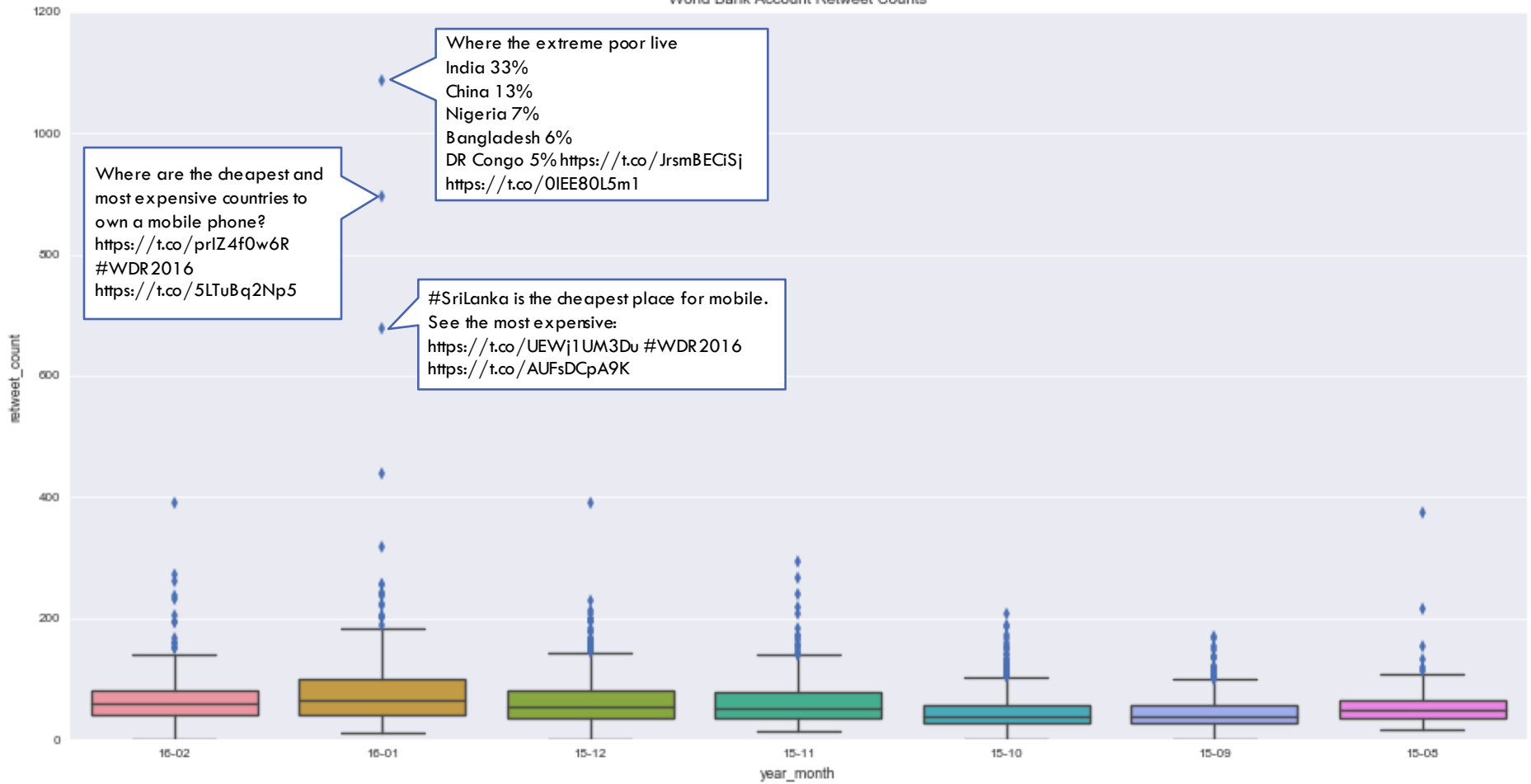
C-V AUC:



- 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
-
- → Look at the @WorldBank account only

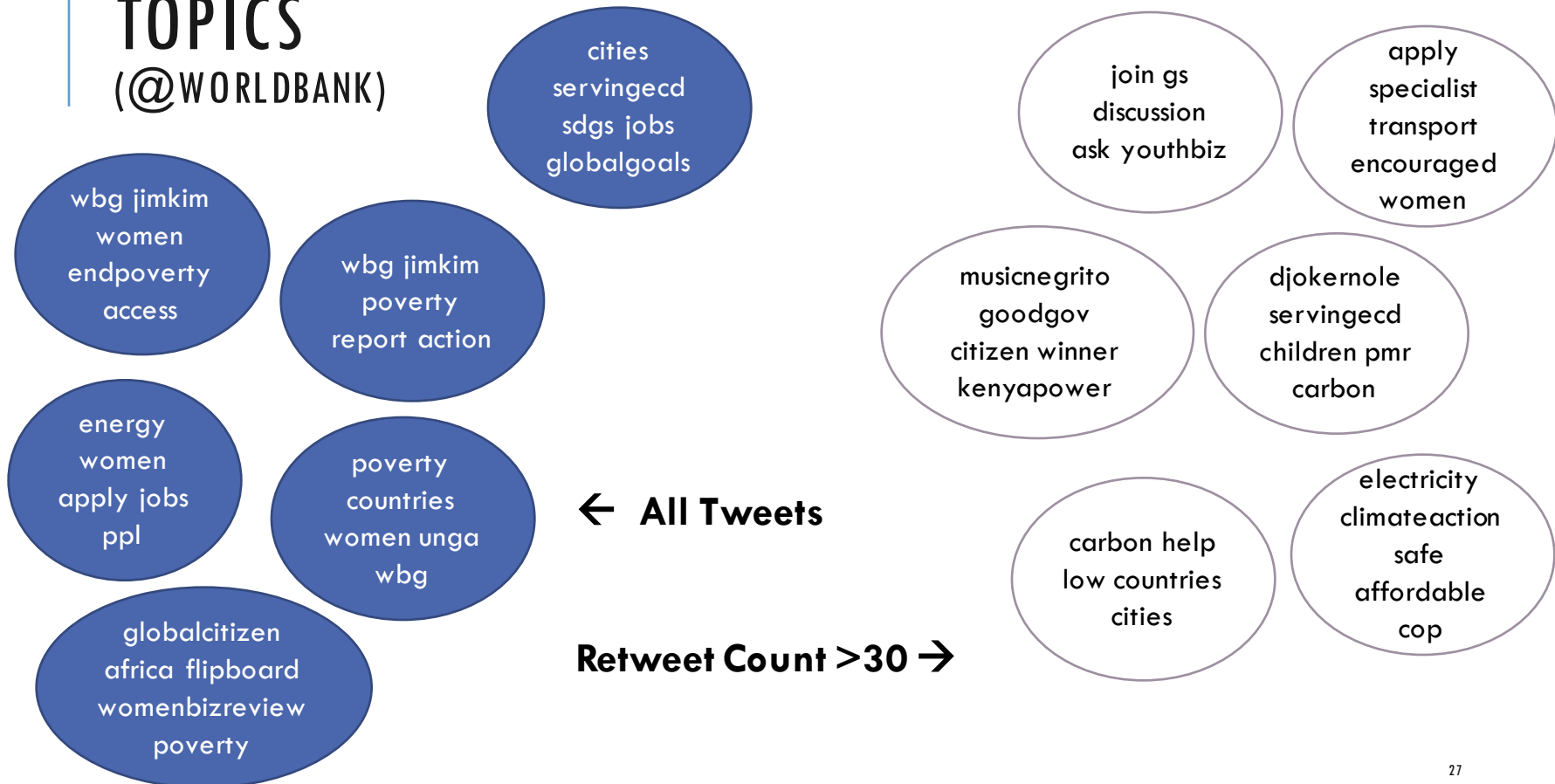


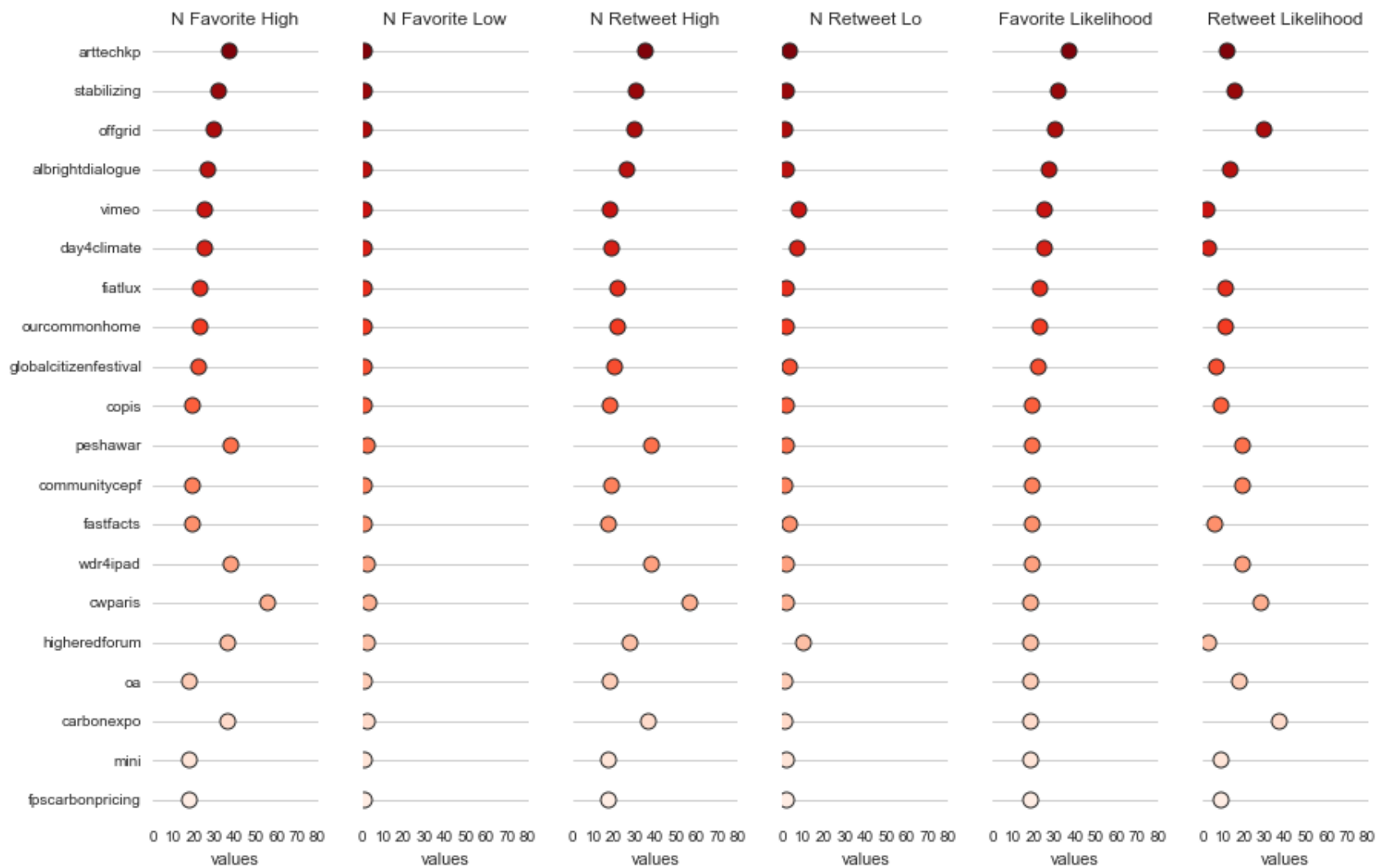
World Bank Account Retweet Counts

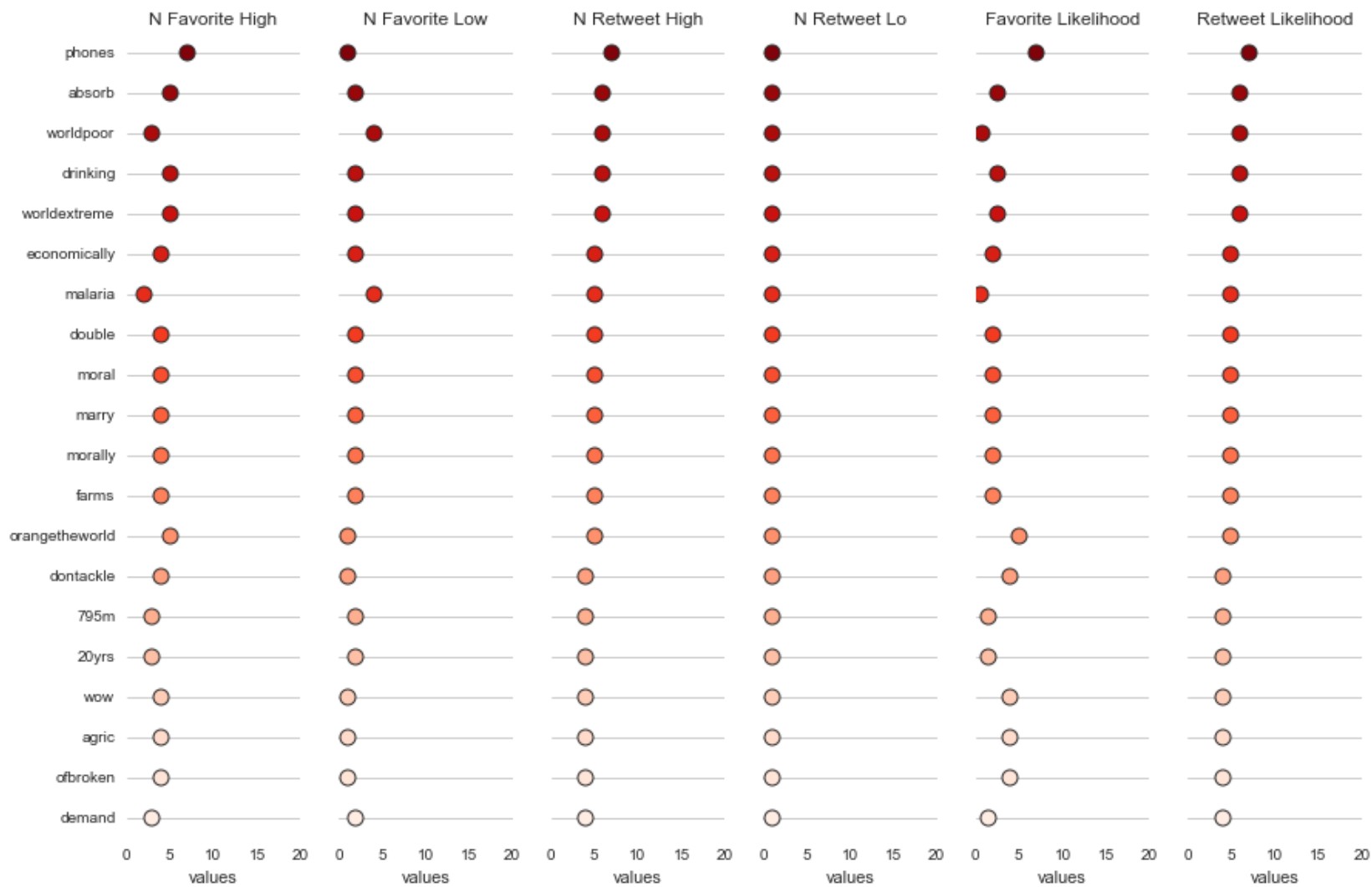


TOPICS

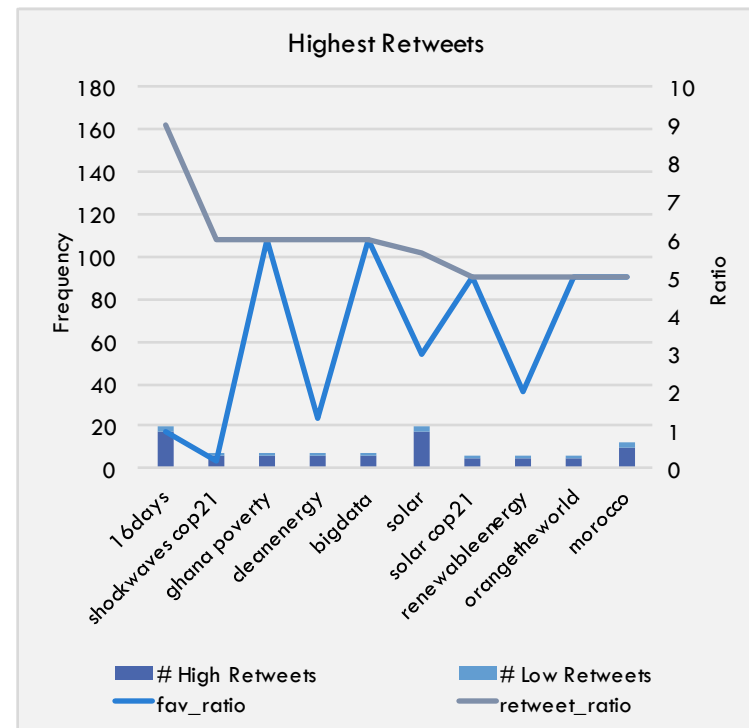
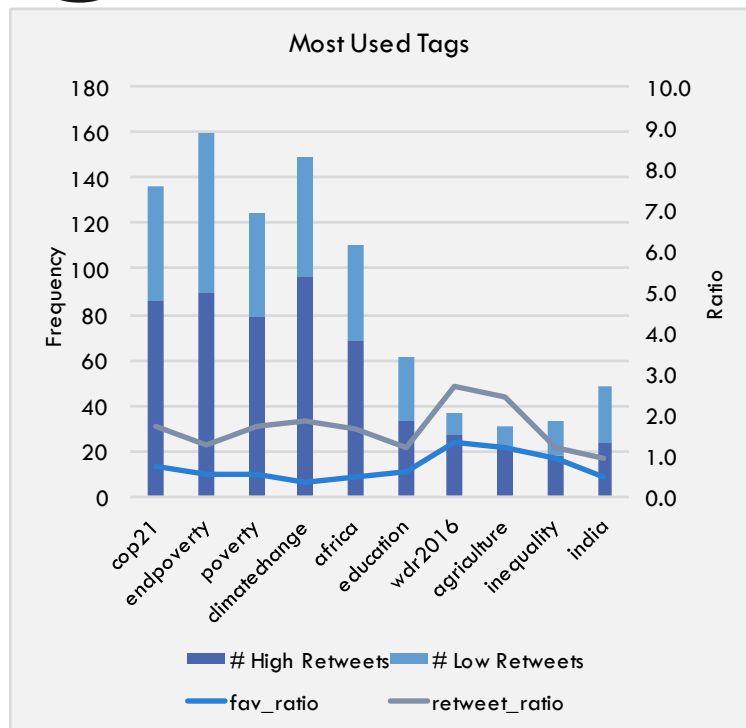
(@WORLDBANK)



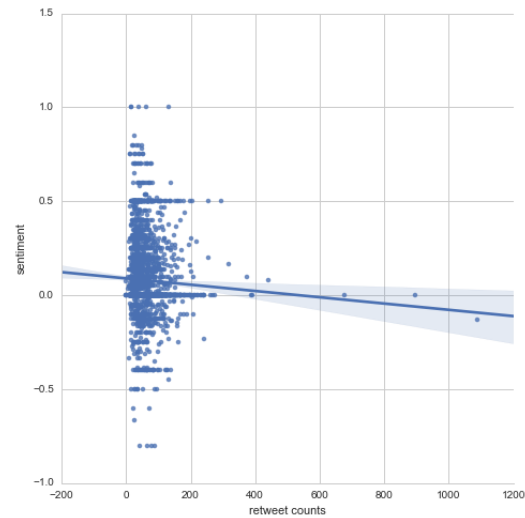
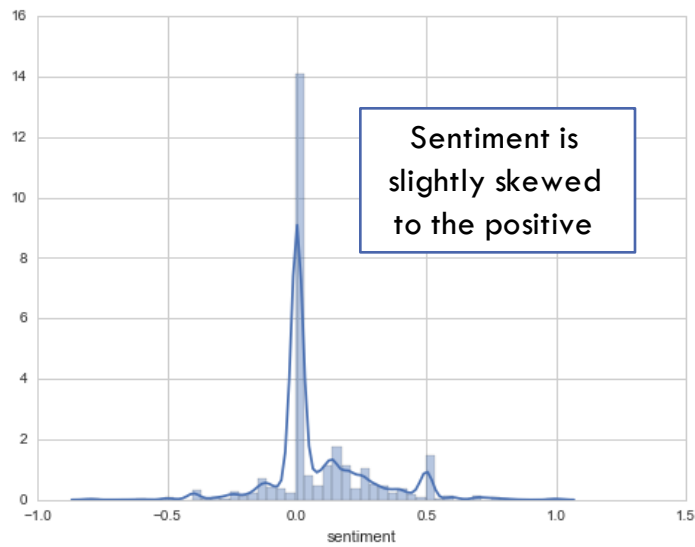




@WORLDBANK

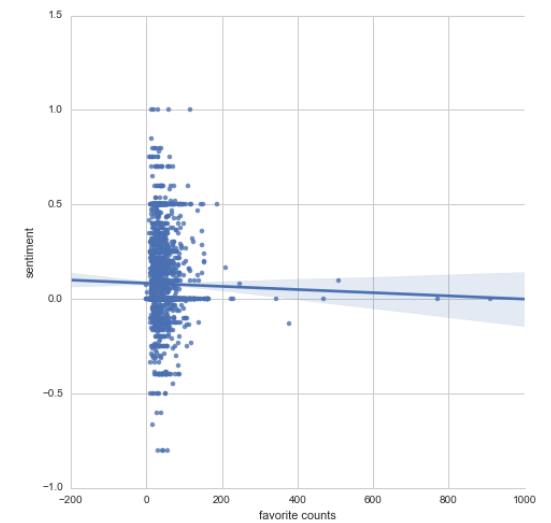


SENTIMENT (@WORLDBANK)



← By Retweet counts

By favorite counts →





CONCLUSIONS

CONCLUSION

1. Need to interact more broadly with people
2. Cross tag people and hashtags
3. What do people tweet about
 1. Internal WB events and publications
 2. WB findings
 3. International events



THANK YOU!

Judy: judy.s.yang@gmail.com