Diplomats on Twitter

2024-02-01

1. Project overview

Key themes: Data collection & wrangling, Mixed methodology, Sentiment analysis, Correspondence analysis, Hierarchical clustering. **Contact Bastian Struve** for any queries via "struve_11@hotmail.com".

This project analyses the content of tweets posted by Chinese diplomats on the platform X (Twitter) over a 5-year period. Dubbed 'wolf warrior diplomacy', journalists so far mostly just described a single trait: a high level of assertiveness. Empirical research is necessary to test that claim, and to add more insights into their communication behaviour.

The goal is to answer the following research questions: (1) what topics, locations, functions, and level of assertiveness are most prominent? (2) how do these features group into distinct profilees of communication behaviour? This project involves collecting the data from Twitter, computing a sentiment score for each tweet, creating a sample for manual coding, and then analysing the coding observations using multiple correspondence analysis and hierarchical clustering.

The data came from a list of 187 diplomats, who produced roughly 1 million tweets in the period from 1 January 2017 to 1 July 2022. Full details will be publicly available in a forthcoming publication by Sullivan, Struve, and Wang (2024). I focus here on original tweets and quotes, which total 271,787 tweets. Existing research leads us to hypothesise that the majority of tweets are highly assertive. In an exploratory effort, we add variables regarding the topic, location and communication function for further insights into the content and style of the diplomat's communications. So, the coding framework we developed has the four main variables: topics, location, function, lv of assertiveness (with 15, 17, 7, and 3 coding options, respectively).

2. Data collection and wrangling

2.1 Retrieve Twitter data

We created a list of 187 Chinese diplomats for this study through synergy of multiple sources. The data was collected following ethics approval and application to Twitter's developer account. Twitter users can change their handle, so we retrieved the user ID for the purpose of data collection.

```r
# Store Twitter authentication info as environment variable for data collection
# Retrieve ID for each Twitter handle using authentication
diplo_names <- readxl::read_xlsx("diplo_list_2507123.xlsx")
diplo_id <- get_user_id(diplo_names$handle, bearer_token=get_bearer(), all=TRUE, keep_na=FALSE)
diplo_list <- diplo_names %>% left_join(diplo_id)
```

Loop through each of the 187 users with the search query. This stores user and tweet-level information in a .json format which are then bound into a single tidyframe. Each tweet has its own row. The R package 'academictwitteR' allows full archive search and various search parameters. This was necessary given the large time frame and to limit the query to original tweets and quotes only (excluding retweets and replies). Then, I filtered for tweets written in English or Chinese for our analysis and for relevant metadata. This yields 271.787 tweets.

```r
# Define function with search query for data collection
get_original_and_quotes <- function(id){
```

```r
  academictwitteR:: get_all_tweets(
    users = diplo_id,
    start_tweets = "2017-01-01T00:00:00Z",
    end_tweets = "2022-07-01T00:00:00Z",
    is_retweet = FALSE,
    is_reply = FALSE,
    n=Inf,
    data_path = "tweets_orig_and_quotes",
    bind_tweets = FALSE)}

# Loop through each user with search query for data collection (producing .json files)
purrr::walk(diplo_list[['diplo_id']], get_original_and_quotes)
```

```r
# Bind json files into a single tidy dataframe
tweets_orig_and_quotes <- bind_tweets (data_path = "tweets_orig_and_quotes", output_format = "tidy")

# Filter for tweets in English and Chinese and relevant metadata
tweets_stripped <- tweets_orig_and_quotes %>%
  select("tweet_id", "user_username", "user_name", "created_at", "text", "lang") %>%
  filter(lang == 'en' | lang == 'zh')
```

```r
# The dataset includes 271,787 rows and 6 columns. Each tweet has its own row.
dim(tweets_stripped)
```

```
## [1] 271787      6
```

2.2 Add sentiment score for each tweet

To aid the manual coding, I computed a sentiment score for each tweet. This will help the manual coding of the level of assertiveness: There is an inherent subjectivity when measuring the assertiveness of a tweet, but with three experienced coders and the aid of the sentiment score this is reasonably reduced. I used the AFINN lexicon for the sentiment analysis: developed by Finn Årup Nielsen, words are not classified in a binary fashion but are assigned a numerical value ranging from -5 (the most negative) to +5 (the most positive).

Tokenise and remove stopwords. The unnest_tokens () function splits the tweet text into a one word per line format and put this in a column called word. I applied stopword lists for both English and Chinese words. The stopword packages I used are commonly used in text analytics. Small note though of the continuing discussion about challenges of word segmentation with Chinese characters (eg, Ma, Ganchev, and Weiss 2024: 'State-of-the-art Chinese Word Segmentation with Bi-LSTMs' or Huang et al. 2020: 'Towards Fast and Accurate Neural Chinese Word Segmentation with Multi-Criteria Learning').

```r
# Tokenise and remove stopwords
tweets_s_tidy <- tweets_s_cleaned %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words, by = "word") %>%
  anti_join(get_stopwords(language = "zh", source = "stopwords-iso"))
```

```
## Joining with `by = join_by(word)`
```

```r
head(tweets_s_tidy %>% select(tweet_id,user_name,created_at, lang, word))
```

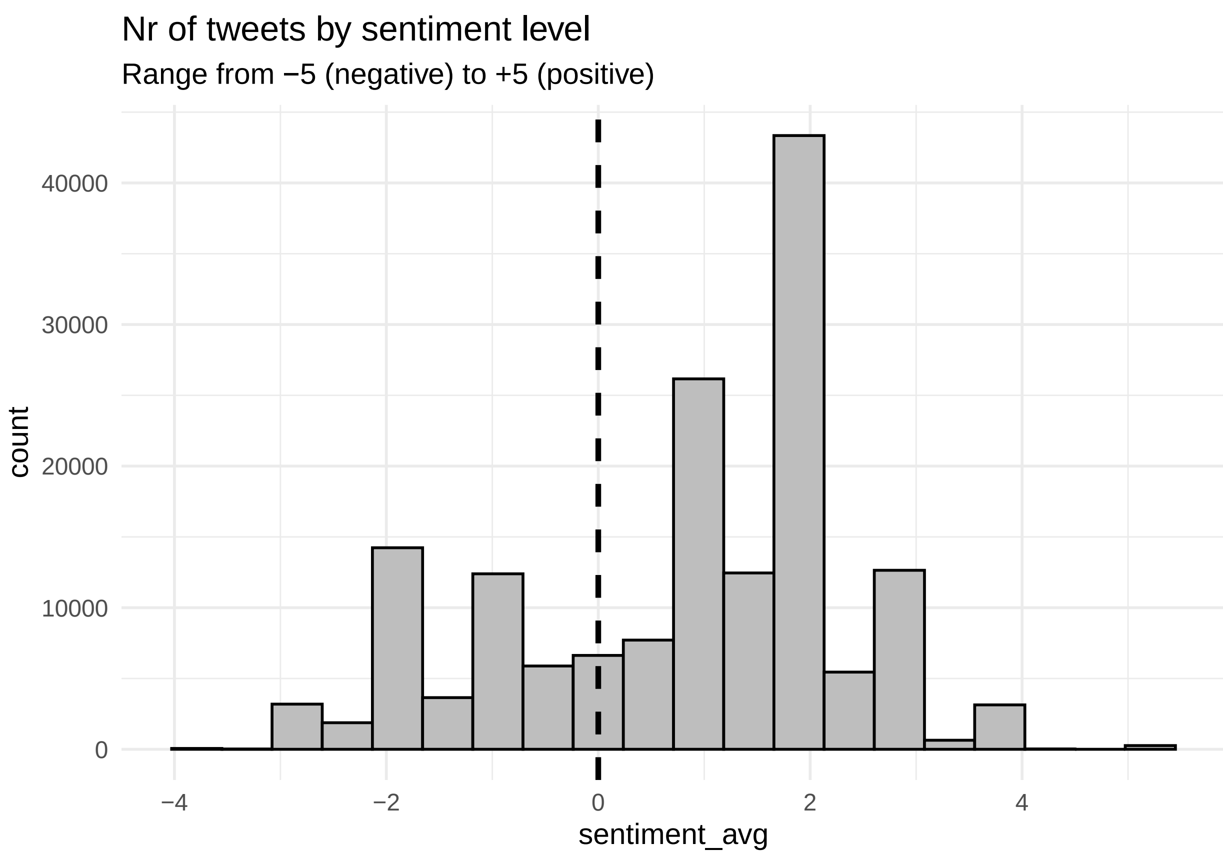
```
## # A tibble: 6 x 5
##   tweet_id            user_name          created_at               lang  word
##   <chr>               <chr>              <chr>                    <chr> <chr>
## 1 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    celebra~
```

```
## 2 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    anniver~
## 3 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    people's
## 4 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    republic
## 5 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    china
## 6 1045746418198155264 Lijian Zhao    2018-09-28T18:45:49.000Z en    huge
```

Assign a sentiment score to each token using the AFINN lexicon. Then, group these back to tweet level and compute an average and total sentiment score for each tweet. Use group_by() to group each tweet's words back together, filter out all the words without a sentiment value using filter(), and finally summarise(), mean() and sum() to produce an average and total sentiment value. Then use the left_join function to add the sentiment values to the data frame of tweets.

Here is a histogram that shows how frequently we had tweets with a certain sentiment score (range). There seem to be a skew towards the right side of the scale. Interesting - seems we were right in addressing the research gap: quite a few more tweets are positive.

```
diplo_final_tweets %>%
  ggplot(aes(x = sentiment_avg)) +
  geom_histogram(bins = 20, fill = 'grey', colour = 'black') +
  geom_vline(xintercept = 0, lwd = 1, lty = 'dashed') +
  theme_minimal() +
  labs(title = "Nr of tweets by sentiment level", subtitle = "Range from -5 (negative) to +5 (positive)
```




```
# Example: Show the 10 most positive tweets
diplo_final_tweets %>%
  select(user_username, text, sentiment_total) %>%
  arrange(desc(sentiment_total)) %>%
  head(10)
```

```
## # A tibble: 10 x 3
##    user_username   text                                            sentiment_total
##    <chr>           <chr>                                                     <dbl>
##  1 ChineseEmbinUK  "Congratulations to @NickyHarman_cn for winn~              27
##  2 China2ASEAN     "#OlympicSolidarity: China's weightlifting g~              27
##  3 ChineseEmbinTT  "What a wonderful display! May TT prosperous~              23
##  4 li_xiaosi       "Hubei&amp;Wuhan, top priority&amp;the decis~              22
##  5 AmbLiuQuan      "Glad to hear the situation of the CONVID 19~              20
##  6 ChineseZimbabwe "Congrats to Dinson Steel on winning of Zim ~              20
##  7 zhang_heqing    "Cool China-\"Woodball\" on ice:\"country ho~              20
##  8 zhang_heqing    "It's my great pleasure today to have 200000~              20
##  9 Chinaembmanila  "Congratulations to the three winners of #PH~              19
## 10 ChineseZimbabwe "#Happy2021\n2020 is a storm; but China and ~              19
```

2.3 Create final coding sample for manual coding

Now the data is ready for manual coding in Excel, following successful piloting of our coding framework. As a reminder, we are interested in the topic, location, function, and assertiveness variables. I used stratified random sampling to ensure that each diplomat is represented in the final coding sample. We coded 2646 tweets.

```r
# Create sample for coding
tweets_sample <- diplo_final_tweets %>%
  group_by(user_username) %>%
  slice_sample(prop = 0.01)

# Write file for coding
write.xlsx(tweets_sample, "tweets_sample_250723.xlsx", fileEncoding = "UTF-8") # UTF-8 for Chinese
```

3. Data analysis: Summary statistics

Now the manual coding is completed and ready for analysis. Here is an overview of how frequently each topic, location, function and lv of assertiveness was present in the sample data. These are coded as binary variables with 0 for absence and 1 for presence (a total of 42 variables). First a table overview, then some data visualization for better interpretation. The table overview helped spotting mistakes in the coding, which we corrected in the Excel file and then re-uploaded.

```r
# Load coding observations
tweets_coded <- readxl::read_xlsx("tweets_coded_250723.xlsx")

# Table overview with example variables
head(Hmisc:: describe(tweets_coded))
```

```
## tweets_coded
##
##  6  Variables      2646  Observations
## --------------------------------------------------------------------------------
## Diplomacy
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2      0.4      419   0.1584   0.2667
##
## --------------------------------------------------------------------------------
## Governance_and_PartyAffairs
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2    0.259      252  0.09524   0.1724
```

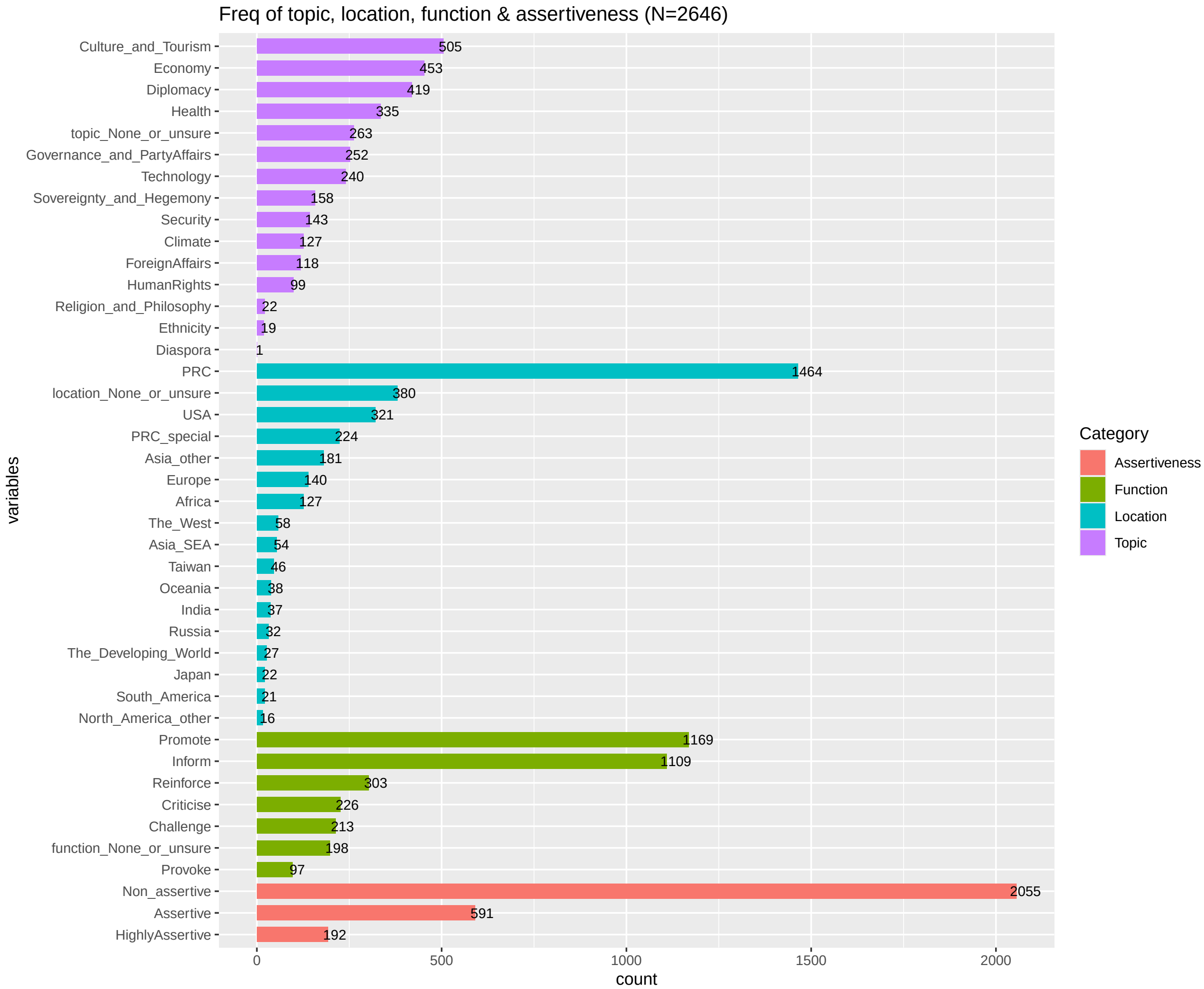
```
##
## --------------------------------------------------------------------------------
## HumanRights
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2    0.108       99  0.03741  0.07206
##
## --------------------------------------------------------------------------------
## Sovereignty_and_Hegemony
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2    0.168      158  0.05971   0.1123
##
## --------------------------------------------------------------------------------
## ForeignAffairs
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2    0.128      118   0.0446  0.08525
##
## --------------------------------------------------------------------------------
## Culture_and_Tourism
##        n  missing distinct     Info      Sum     Mean      Gmd
##     2646        0        2    0.463      505   0.1909    0.309
##
## --------------------------------------------------------------------------------
```

Here is some data visualization that is easier to interpret, with variables sorted by frequency of occurrence. Responding to the first research question, there are some take-aways here already. Tweets with a low level of assertiveness are by far most prominent. Regarding the communication function, tweets most prominently inform or promote, and have the PRC very strongly in focus. We can see that a wide range of topics covered, though with different degree of frequency. Altogether, this already paints a very clear picture regarding the research gap we wanted to address: a focus only on 'wolf warrior' traits does not account for what is most prominently present in the Twitter communications of the Chinese diplomats.

```r
freq_coded<- readxl:: read_xlsx("tweets_coded_freq_250723.xlsx")
```

```
## New names:
## * `` -> `...5`
```

```r
print(freq_plot <- freq_coded %>%
  ggplot(aes(x= fct_inorder(Variable), y = Present, fill = Category)) +
  geom_col(width = .70, position = position_dodge(width = .70))+
  geom_text(aes(label= Present), hjust = .2, size = 3, position = position_dodge(width = 0.7)) +
  coord_flip() + labs(x = "variables", y = "count", title = "Freq of topic, location, function & asserti
```



4. Data analysis: Correspondence analysis and hierarchical clustering

Next, I turn to exploratory efforts in order to answer the second research question. I seek to understand how the various observations we coded for relate to each other. This allows conclusions on any clusters of communication behaviour (which in our publication we use to discuss distinct commucniation profiles). Various options exist to analyse the interrelationships between variables. The focus here is on Multiple Correspondence Analysis (MCA): it is an extension of principal component analysis for when the variables to be analysed are categorical instead of quantitative.

4.1 Check data suitability for MCA

First test the suitability of the data and adjust the data structure to ready it for the subsequent MCA. As, there are various low correlations that would bias the findings if left untreated. THis is not surprising, for example, because of the low presence of various variables (see summary statistics, e.g. topic diaspora). Going through multiple iterations, several variables were removed, leaving 20 variables that are suitable for MCA. Various commonly used tests confirm as below, eg Bartlett and KMO tests.

```
# Remove variables with very marginal presence or correlations
tweets_coded_slim <- select (tweets_coded, -Religion_and_Philosophy, -Ethnicity, -Diaspora, -Oceania, -I
```

```r
psych:: cortest.bartlett(tweets_coded_slim)
```

```
## R was not square, finding R from data
```

```
## $chisq
## [1] 10017.16
##
## $p.value
## [1] 0
##
## $df
## [1] 190
```

```r
psych:: KMO(tweets_coded_slim)
```

```
## Kaiser-Meyer-Olkin factor adequacy
## Call: psych::KMO(r = tweets_coded_slim)
## Overall MSA =  0.71
## MSA for each item =
##              HumanRights Sovereignty_and_Hegemony           ForeignAffairs
##                     0.70                     0.76                     0.71
##       Culture_and_Tourism               Technology                  Economy
##                     0.62                     0.61                     0.52
##                 Security                      PRC              PRC_special
##                     0.67                     0.65                     0.48
##                      USA                   Taiwan                 Asia_SEA
##                     0.83                     0.65                     0.51
##                 The_West                   Inform                  Promote
##                     0.72                     0.58                     0.58
##                Challenge                Criticise                  Provoke
##                     0.69                     0.76                     0.69
##                Assertive          HighlyAssertive
##                     0.81                     0.81
```

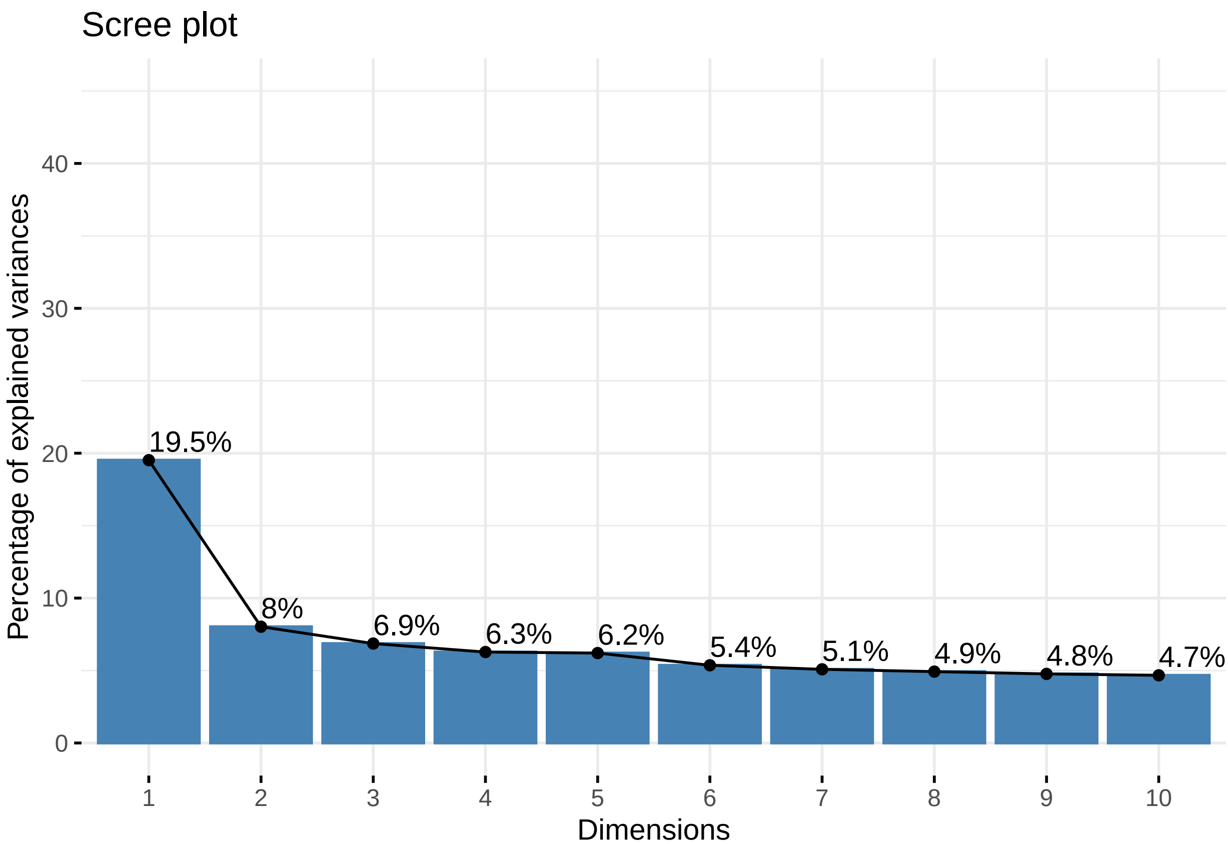
4.2 Run Multiple Correspondence analysis

Convert all binary variables into factors and then use the MCA() function from the FactoMineR package.

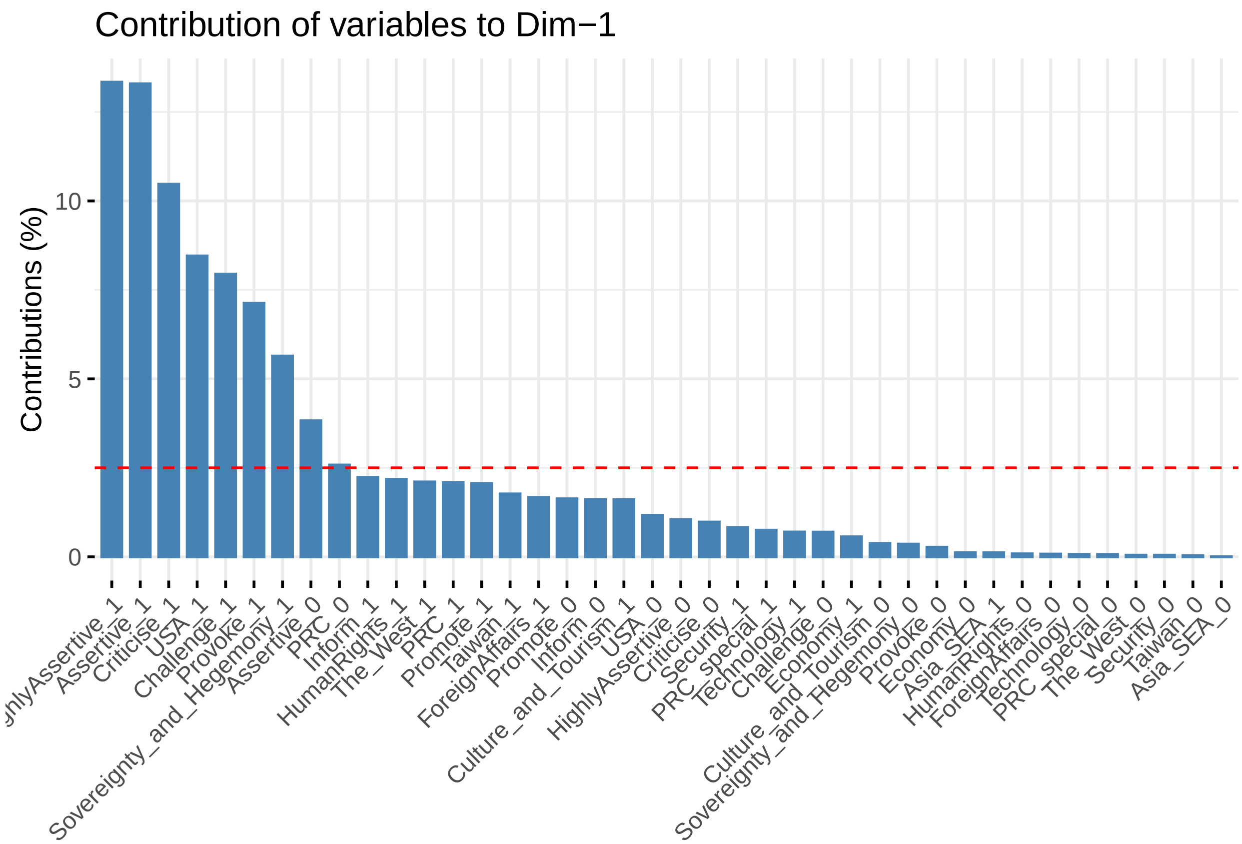
```r
# Convert all binary variables into factors and run MCA
df.2 <- tweets_coded_slim %>%
  mutate_if(~ is.numeric(.) && all(unique(.) %in% c(0, 1, NA)), factor)

res.mca <- MCA(df.2, ncp=6, graph=FALSE)

# Visualise eigenvalues and contribution in bar chart
fviz_screeplot(res.mca, addlabels = TRUE, ylim = c(0, 45))
```

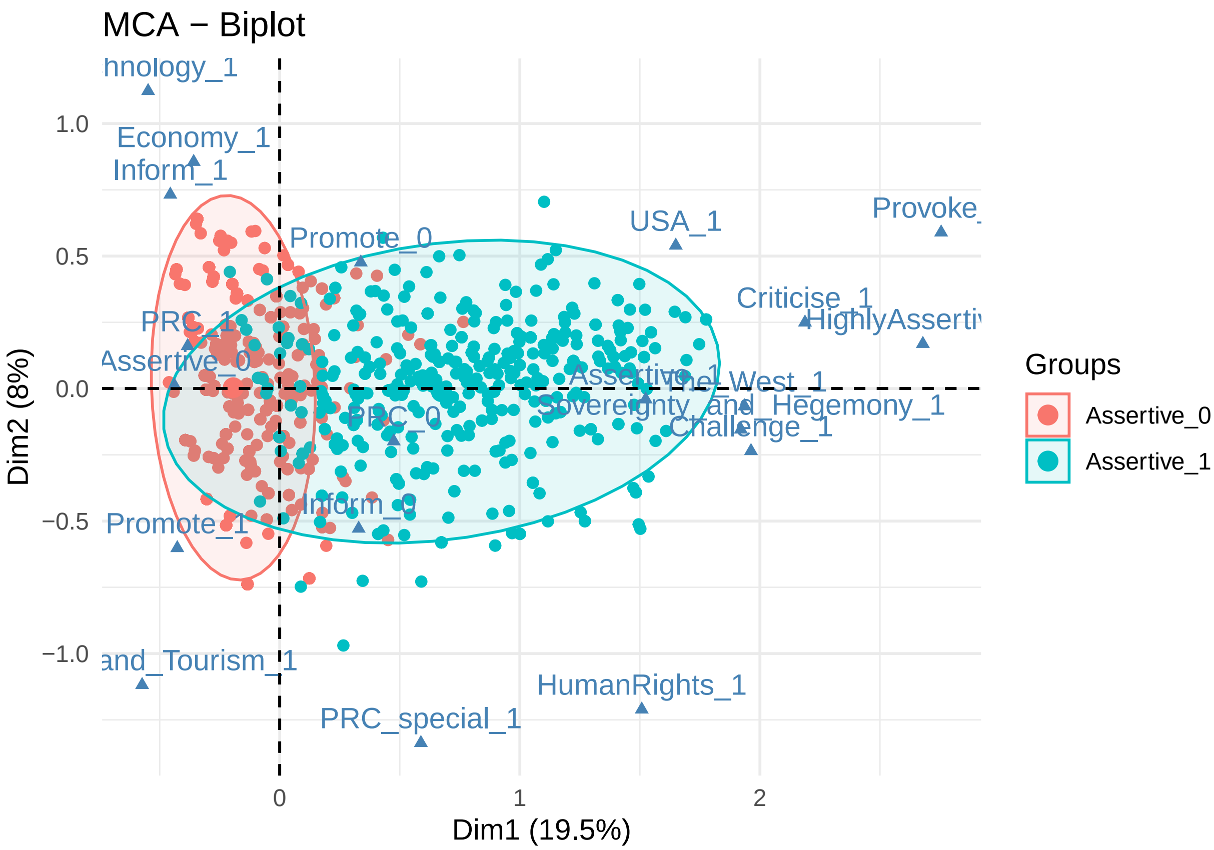



```r
# Visualise most defining variable categories for each dimension. Dimension 1 as example here
fviz_contrib(res.mca,  choice ="var", axes = 1)
```



Note to distinguish between variable categories: both the absence (variable name ending "_0") and presence ("_1") can be defining features of a dimension (nad ultimately help explaining data variability). I am using a more intuitive plot to interpret the findings.

```
grp <- res.mca$call$X$Assertive
fviz_mca_biplot(res.mca, axes = c(1,2), label="var", col.var ="steelblue", habillage=grp,
                addEllipses=TRUE, ellipse.level=0.95, select.var = list(contrib = 20)) + theme_minimal(
```



The output for dimensions 1 and 2 are represented as a factor map with the top 10 contributing variable categories. It shows the relationship between the variable categories for dimension 1 on the x-axis and dimension 2 on the y-axis. Added spheres for the point cloud of tweets further differentiates whether a tweet was assertive. While there is some overlap, the spheres in the factor map strongly signal that more and less assertive tweets focus on different topics, locations, or functions.

Variable categories with a similar profile are grouped together. Negatively associated variable categories are positioned on the opposite sides of the plot origin. Dimension 1 is defined by a high level of assertiveness, the functions criticise, challenge, and provoke, the locations the US and the West, and the topic of sovereignty and hegemony. Their proximity highlights the strong positive association between these variables. It is noteworthy that the PRC is located on the other end of the x-axis to this set of variables, signalling a negative correlation. In other words, when diplomats talk about China they do so in a way that is distinct from the defining characteristics of dimension 1. Dimension 2 is defined by the presence of informing and the topics of technology and economy, and a negative correlation with promoting and the topic of culture and tourism. More of this becomes clear through the clustering analysis.

4.3 Hierarchical clustering (HCPC)

The information produced by the MCA then served as input for the Hierarchical clustering upon principal components (HCPC) analysis.

```
options(width = 200)
res.hcpc <- HCPC (res.mca, nb.clust = 5 , metric = "euclidean", graph = FALSE)
res.hcpc$desc.var
```

```
##
## Link between the cluster variable and the categorical variables (chi-square test)
## =================================================================================
```

```
##                            p.value df
## PRC_special            0.000000e+00  4
## Assertive              9.186543e-313  4
## Economy                5.430659e-303  4
## Criticise              2.569382e-300  4
## HighlyAssertive        1.845146e-279  4
## Taiwan                 5.614155e-277  4
## Sovereignty_and_Hegemony 1.035559e-252  4
## USA                    4.521016e-192  4
## Provoke                1.159557e-145  4
## Challenge              1.392502e-142  4
## Technology             6.188096e-140  4
## PRC                    3.777355e-105  4
## HumanRights             1.693862e-99  4
## Inform                  4.456190e-81  4
## Culture_and_Tourism     1.010857e-66  4
## Promote                 5.376133e-50  4
## The_West                3.032088e-48  4
## ForeignAffairs          1.821389e-36  4
## Asia_SEA                1.038084e-30  4
## Security                3.172899e-23  4
##
## Description of each cluster by the categories
## =============================================
## $`1`
##                                                       Cla/Mod     Mod/Cla    Global       p.value
## Economy=Economy_1                                  88.3002208  68.1431005 17.120181 2.597632e-261  3
## Technology=Technology_1                            87.5000000  35.7751278  9.070295 5.498820e-117  2
## Culture_and_Tourism=Culture_and_Tourism_0          27.0434376  98.6371380 80.914588  4.138736e-49  1
## Inform=Inform_1                                    35.7980162  67.6320273 41.912320  3.279952e-46  1
## PRC=PRC_1                                          31.8989071  79.5570698 55.328798  2.405884e-43  1
## Asia_SEA=Asia_SEA_1                                87.0370370   8.0068143  2.040816  1.560753e-24  1
## Assertive=Assertive_0                              26.2287105  91.8228279 77.664399  5.450845e-24  1
## Criticise=Criticise_0                              24.1735537  99.6592845 91.458806  3.520086e-23
## USA=USA_0                                          24.7311828  97.9557070 87.868481  1.160215e-22
## HighlyAssertive=HighlyAssertive_0                  23.9201304 100.0000000 92.743764  1.500803e-22
## PRC_special=PRC_special_0                          24.0710157  99.3185690 91.534392  2.631364e-20
## Security=Security_0                                23.4119057  99.8296422 94.595616  3.870215e-15
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_0 23.4726688  99.4889267 94.028723  3.225242e-14
## Challenge=Challenge_0                              23.7566790  98.4667802 91.950113  3.958254e-14
## ForeignAffairs=ForeignAffairs_0                    23.2199367 100.0000000 95.540438  6.478550e-14
## Provoke=Provoke_0                                  23.0286387 100.0000000 96.334089  1.619334e-11
## HumanRights=HumanRights_0                          22.9289360  99.4889267 96.258503  4.899705e-08
## Taiwan=Taiwan_0                                    22.5769231 100.0000000 98.261527  8.709339e-06
## The_West=The_West_0                                22.5656878  99.4889267 97.808012  4.380586e-04
## Promote=Promote_1                                  24.8930710  49.5741056 44.179894  2.937608e-03
## Promote=Promote_0                                  20.0406229  50.4258944 55.820106  2.937608e-03  -
## The_West=The_West_1                                 5.1724138   0.5110733  2.191988  4.380586e-04  -
## Taiwan=Taiwan_1                                     0.0000000   0.0000000  1.738473  8.709339e-06  -
## HumanRights=HumanRights_1                           3.0303030   0.5110733  3.741497  4.899705e-08  -
## Provoke=Provoke_1                                   0.0000000   0.0000000  3.665911  1.619334e-11  -
## ForeignAffairs=ForeignAffairs_1                     0.0000000   0.0000000  4.459562  6.478550e-14  -
## Challenge=Challenge_1                               4.2253521   1.5332198  8.049887  3.958254e-14  -
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_1  1.8987342   0.5110733  5.971277  3.225242e-14  -
```

```
## Security=Security_1                                 0.6993007   0.1703578  5.404384  3.870215e-15  -
## PRC_special=PRC_special_1                         1.7857143   0.6814310  8.465608  2.631364e-20  -
## HighlyAssertive=HighlyAssertive_1                 0.0000000   0.0000000  7.256236  1.500803e-22  -
## USA=USA_1                                         3.7383178   2.0442930 12.131519  1.160215e-22  -
## Criticise=Criticise_1                             0.8849558   0.3407155  8.541194  3.520086e-23  -
## Assertive=Assertive_1                             8.1218274   8.1771721 22.335601  5.450845e-24 -
## Asia_SEA=Asia_SEA_0                              20.8333333  91.9931857 97.959184  1.560753e-24 -
## PRC=PRC_0                                        10.1522843  20.4429302 44.671202  2.405884e-43 -
## Inform=Inform_0                                  12.3617437  32.3679727 58.087680  3.279952e-46 -
## Culture_and_Tourism=Culture_and_Tourism_1         1.5841584   1.3628620 19.085412  4.138736e-49 -
## Technology=Technology_0                          15.6691604  64.2248722 90.929705 5.498820e-117 -2
## Economy=Economy_0                                 8.5271318  31.8568995 82.879819 2.597632e-261 -3
##
## $`2`
##                                                    Cla/Mod      Mod/Cla    Global       p.value
## Economy=Economy_0                                65.9370725  99.65541006 82.879819 1.106330e-169
## Assertive=Assertive_0                            65.2068127  92.35010338 77.664399  2.119142e-92
## PRC_special=PRC_special_0                        59.8678778  99.93108201 91.534392  5.182297e-81
## HighlyAssertive=HighlyAssertive_0                59.1279544 100.00000000 92.743764  5.635496e-71
## Criticise=Criticise_0                            59.5867769  99.37973811 91.458806  3.327188e-67
## Culture_and_Tourism=Culture_and_Tourism_1        86.9306931  30.25499655 19.085412  7.770183e-65
## Technology=Technology_0                          59.5594347  98.75947622 90.929705  5.646499e-60
## Challenge=Challenge_0                            59.1039868  99.10406616 91.950113  1.067782e-56
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_0 58.1591640  99.72432805 94.028723  6.366979e-50
## HumanRights=HumanRights_0                        56.9689831 100.00000000 96.258503  6.543078e-36
## Provoke=Provoke_0                                56.9242840 100.00000000 96.334089  3.528112e-35
## USA=USA_0                                        58.5806452  93.86629910 87.868481  1.202732e-25
## Taiwan=Taiwan_0                                  55.8076923 100.00000000 98.261527  8.121515e-17
## Asia_SEA=Asia_SEA_0                              55.9027778  99.86216402 97.959184  2.850372e-16
## Promote=Promote_1                                62.7031651  50.51688491 44.179894  4.227503e-13
## The_West=The_West_0                              55.7959815  99.51757409 97.808012  8.580162e-12
## PRC=PRC_1                                        59.9043716  60.44107512 55.328798  5.667619e-09
## PRC=PRC_0                                        48.5617597  39.55892488 44.671202  5.667619e-09
## The_West=The_West_1                              12.0689655   0.48242591  2.191988  8.580162e-12
## Promote=Promote_0                                48.6120515  49.48311509 55.820106  4.227503e-13
## Asia_SEA=Asia_SEA_1                               3.7037037   0.13783598  2.040816  2.850372e-16
## Taiwan=Taiwan_1                                   0.0000000   0.00000000  1.738473  8.121515e-17
## USA=USA_1                                        27.7258567   6.13370090 12.131519  1.202732e-25 -
## Provoke=Provoke_1                                 0.0000000   0.00000000  3.665911  3.528112e-35 -
## HumanRights=HumanRights_1                         0.0000000   0.00000000  3.741497  6.543078e-36 -
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_1  2.5316456   0.27567195  5.971277  6.366979e-50 -
## Challenge=Challenge_1                             6.1032864   0.89593384  8.049887  1.067782e-56 -
## Technology=Technology_1                           7.5000000   1.24052378  9.070295  5.646499e-60 -
## Culture_and_Tourism=Culture_and_Tourism_0        47.2676319  69.74500345 80.914588  7.770183e-65 -
## Criticise=Criticise_1                             3.9823009   0.62026189  8.541194  3.327188e-67 -
## HighlyAssertive=HighlyAssertive_1                 0.0000000   0.00000000  7.256236  5.635496e-71 -
## PRC_special=PRC_special_1                         0.4464286   0.06891799  8.465608  5.182297e-81 -
## Assertive=Assertive_1                            18.7817259   7.64989662 22.335601  2.119142e-92 -
## Economy=Economy_1                                 1.1037528   0.34458994 17.120181 1.106330e-169 -
##
## $`3`
##                                                   Cla/Mod    Mod/Cla    Global       p.value
## PRC_special=PRC_special_1                       82.589286  84.090909  8.465608 4.050550e-206  30
## PRC=PRC_0                                       16.920474  90.909091 44.671202  1.772019e-51  15
```

```
## HumanRights=HumanRights_1                           59.595960  26.818182  3.741497  4.558627e-41  13
## USA=USA_0                                             9.247312  97.727273 87.868481  5.076835e-08   5
## Economy=Economy_0                                     9.439124  94.090909 82.879819  3.381561e-07   5
## Challenge=Challenge_1                                18.779343  18.181818  8.049887  3.420747e-07   5
## Criticise=Criticise_0                                 8.925620  98.181818 91.458806  1.737822e-05   4
## Technology=Technology_0                               8.935993  97.727273 90.929705  2.957283e-05   4
## Provoke=Provoke_0                                     8.630836 100.000000 96.334089  1.871020e-04   3
## HighlyAssertive=HighlyAssertive_0                     8.761206  97.727273 92.743764  9.150943e-04   3
## Asia_SEA=Asia_SEA_0                                   8.487654 100.000000 97.959184  8.762890e-03   2
## Taiwan=Taiwan_0                                       8.461538 100.000000 98.261527  1.779310e-02   2
## Inform=Inform_0                                       9.368900  65.454545 58.087680  2.010683e-02   2
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_1 13.291139   9.545455  5.971277  2.844324e-02   2
## Assertive=Assertive_1                                10.490694  28.181818 22.335601  3.357784e-02   2
## ForeignAffairs=ForeignAffairs_0                       8.544304  98.181818 95.540438  3.396609e-02   2
## ForeignAffairs=ForeignAffairs_1                       3.389831   1.818182  4.459562  3.396609e-02  -2
## Assertive=Assertive_0                                 7.688564  71.818182 77.664399  3.357784e-02  -2
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_0  7.998392  90.454545 94.028723  2.844324e-02  -2
## Inform=Inform_1                                       6.853021  34.545455 41.912320  2.010683e-02  -2
## Taiwan=Taiwan_1                                       0.000000   0.000000  1.738473  1.779310e-02  -2
## Asia_SEA=Asia_SEA_1                                   0.000000   0.000000  2.040816  8.762890e-03  -2
## HighlyAssertive=HighlyAssertive_1                     2.604167   2.272727  7.256236  9.150943e-04  -3
## Provoke=Provoke_1                                     0.000000   0.000000  3.665911  1.871020e-04  -3
## Technology=Technology_1                               2.083333   2.272727  9.070295  2.957283e-05  -4
## Criticise=Criticise_1                                 1.769912   1.818182  8.541194  1.737822e-05  -4
## Challenge=Challenge_0                                 7.398274  81.818182 91.950113  3.420747e-07  -5
## Economy=Economy_1                                     2.869757   5.909091 17.120181  3.381561e-07  -5
## USA=USA_1                                             1.557632   2.272727 12.131519  5.076835e-08  -5
## HumanRights=HumanRights_0                             6.321162  73.181818 96.258503  4.558627e-41 -13
## PRC=PRC_1                                             1.366120   9.090909 55.328798  1.772019e-51 -15
## PRC_special=PRC_special_0                             1.445087  15.909091 91.534392 4.050550e-206 -30
##
## $`4`
##                                                        Cla/Mod    Mod/Cla    Global      p.value
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_1  48.7341772  82.795699  5.971277 6.509066e-87  19
## Taiwan=Taiwan_1                                     100.0000000  49.462366  1.738473 2.411610e-73  18
## Assertive=Assertive_1                                13.7055838  87.096774 22.335601 6.342628e-42  13
## Challenge=Challenge_1                                15.9624413  36.559140  8.049887 2.518954e-15   7
## Security=Security_1                                  15.3846154  23.655914  5.404384 1.440687e-09   6
## Inform=Inform_0                                       5.2049447  86.021505 58.087680 3.856138e-09   5
## Culture_and_Tourism=Culture_and_Tourism_0             4.2503503  97.849462 80.914588 5.803024e-07   4
## USA=USA_1                                             8.0996885  27.956989 12.131519 2.758021e-05   4
## Economy=Economy_0                                     4.1039672  96.774194 82.879819 3.196401e-05   4
## Criticise=Criticise_1                                 8.8495575  21.505376  8.541194 8.931125e-05   3
## HighlyAssertive=HighlyAssertive_1                     8.8541667  18.279570  7.256236 3.443784e-04   3
## Promote=Promote_0                                     4.6039269  73.118280 55.820106 5.089681e-04   3
## Technology=Technology_0                               3.8237739  98.924731 90.929705 1.425601e-03   3
## PRC=PRC_0                                             4.7377327  60.215054 44.671202 2.328513e-03   3
## Provoke=Provoke_1                                     8.2474227   8.602151  3.665911 2.517634e-02   2
## HumanRights=HumanRights_0                             3.6513545 100.000000 96.258503 2.703200e-02   2
## Asia_SEA=Asia_SEA_1                                   9.2592593   5.376344  2.040816 4.979440e-02   1
## Asia_SEA=Asia_SEA_0                                   3.3950617  94.623656 97.959184 4.979440e-02  -1
## HumanRights=HumanRights_1                             0.0000000   0.000000  3.741497 2.703200e-02  -2
## Provoke=Provoke_0                                     3.3346410  91.397849 96.334089 2.517634e-02  -2
## PRC=PRC_1                                             2.5273224  39.784946 55.328798 2.328513e-03  -3
```

```
## Technology=Technology_1                              0.4166667   1.075269  9.070295 1.425601e-03  -3
## Promote=Promote_1                                    2.1385800  26.881720 44.179894 5.089681e-04  -3
## HighlyAssertive=HighlyAssertive_0                    3.0969845  81.720430 92.743764 3.443784e-04  -3
## Criticise=Criticise_0                                3.0165289  78.494624 91.458806 8.931125e-05  -3
## Economy=Economy_1                                    0.6622517   3.225806 17.120181 3.196401e-05  -4
## USA=USA_0                                            2.8817204  72.043011 87.868481 2.758021e-05  -4
## Culture_and_Tourism=Culture_and_Tourism_1            0.3960396   2.150538 19.085412 5.803024e-07  -4
## Inform=Inform_1                                      1.1722272  13.978495 41.912320 3.856138e-09  -5
## Security=Security_0                                  2.8365961  76.344086 94.595616 1.440687e-09  -6
## Challenge=Challenge_0                                2.4249897  63.440860 91.950113 2.518954e-15  -7
## Assertive=Assertive_0                                0.5839416  12.903226 77.664399 6.342628e-42 -13
## Taiwan=Taiwan_0                                      1.8076923  50.537634 98.261527 2.411610e-73 -18
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_0  0.6430868  17.204301 94.028723 6.509066e-87 -19
##
## $`5`
##                                                        Cla/Mod    Mod/Cla    Global       p.value
## Assertive=Assertive_1                               48.9001692  97.966102 22.335601 1.697748e-207  30
## Criticise=Criticise_1                               84.5132743  64.745763  8.541194 9.061117e-175  28
## HighlyAssertive=HighlyAssertive_1                   88.5416667  57.627119  7.256236 3.606019e-159  20
## USA=USA_1                                           58.8785047  64.067797 12.131519 4.778403e-122  23
## Provoke=Provoke_1                                   91.7525773  30.169492  3.665911  1.907929e-80  18
## Challenge=Challenge_1                               54.9295775  39.661017  8.049887  2.332665e-63  10
## Inform=Inform_0                                     18.6727391  97.288136 58.087680  2.530486e-61  10
## Promote=Promote_0                                   19.1604604  95.932203 55.820106  6.195173e-61  10
## PRC=PRC_0                                           19.6277496  78.644068 44.671202  1.663669e-36  12
## The_West=The_West_1                                 72.4137931  14.237288  2.191988  1.046276e-28  11
## ForeignAffairs=ForeignAffairs_1                     46.6101695  18.644068  4.459562  1.538179e-23   9
## Culture_and_Tourism=Culture_and_Tourism_0           13.5450724  98.305085 80.914588  3.296526e-22   9
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_1 33.5443038  17.966102  5.971277  6.186504e-15   7
## HumanRights=HumanRights_1                           37.3737374  12.542373  3.741497  3.077308e-12   6
## Technology=Technology_0                             12.0116376  97.966102 90.929705  2.801118e-07   5
## Security=Security_1                                 23.0769231  11.186441  5.404384  2.879611e-05   4
## Economy=Economy_0                                   11.9927041  89.152542 82.879819  1.563689e-03   3
## Asia_SEA=Asia_SEA_0                                 11.3811728 100.000000 97.959184  1.577211e-03   3
## Taiwan=Taiwan_0                                     11.3461538 100.000000 98.261527  4.139130e-03   2
## Taiwan=Taiwan_1                                      0.0000000   0.000000  1.738473  4.139130e-03  -2
## Asia_SEA=Asia_SEA_1                                  0.0000000   0.000000  2.040816  1.577211e-03  -3
## Economy=Economy_1                                    7.0640177  10.847458 17.120181  1.563689e-03  -3
## Security=Security_0                                 10.4674391  88.813559 94.595616  2.879611e-05  -4
## Technology=Technology_1                              2.5000000   2.033898  9.070295  2.801118e-07  -5
## HumanRights=HumanRights_0                           10.1295642  87.457627 96.258503  3.077308e-12  -6
## Sovereignty_and_Hegemony=Sovereignty_and_Hegemony_0  9.7266881  82.033898 94.028723  6.186504e-15  -7
## Culture_and_Tourism=Culture_and_Tourism_1            0.9900990   1.694915 19.085412  3.296526e-22  -9
## ForeignAffairs=ForeignAffairs_0                      9.4936709  81.355932 95.540438  1.538179e-23  -9
## The_West=The_West_0                                  9.7758887  85.762712 97.808012  1.046276e-28 -11
## PRC=PRC_1                                            4.3032787  21.355932 55.328798  1.663669e-36 -12
## Promote=Promote_1                                    1.0265184   4.067797 44.179894  6.195173e-61 -10
## Inform=Inform_1                                      0.7213706   2.711864 41.912320  2.530486e-61 -10
## Challenge=Challenge_0                                7.3160707  60.338983 91.950113  2.332665e-63 -10
## Provoke=Provoke_0                                    8.0816006  69.830508 96.334089  1.907929e-80 -18
## USA=USA_0                                            4.5591398  35.932203 87.868481 4.778403e-122 -23
## HighlyAssertive=HighlyAssertive_0                    5.0937245  42.372881 92.743764 3.606019e-159 -20
## Criticise=Criticise_0                                4.2975207  35.254237 91.458806 9.061117e-175 -28
## Assertive=Assertive_0                                0.2919708   2.033898 77.664399 1.697748e-207 -30
```

Responding to the second research question, we found five distinct clusters of communication behaviour based on the associations between the variables we coded for. Full details will be in our publication, but here some insights already.

Cluster 1 refers to what we define as the informer profile. It is defined by the strong association between informing, economy and technology, and the absence of assertiveness. The summary statistics already highlighted that informing makes up a large part of what the diplomats do, with almost every second tweet doing so (N= 1109).

Cluster 2 refers to the promoter profile. In cluster 2 the absence of various variable categories can be seen, yet we also note that these were defining features of other clusters. Cluster 2 is again defined by the lack of assertiveness, as well as the presence and association of promoting, culture and tourism, and a focus on the PRC focus. The summary statistics shows how important promoting is to the diplomats, it is the most frequently use communication function and can found in almost every second tweet.

Cluster 3 describes the challenger profile in relation to PRC_special (e.g., Xinjiang, Tibet, Hong Kong). The range of defining variable categories is small, yet conclusive: there is a strong location focus on PRC_special, coupled with assertiveness, challenging, and the topics of human rights and sovereignty and hegemony. The summary statistics shows that this cluster receives much less attention to the previous ones, but is a defining feature of data variability. Filtering for tweets that focus on PRC_special shows some nuance in the communication behaviour: one third of the tweets are assertive, mainly covering human rights, sovereignty and hegemony, and governance and party affairs in roughly equal amounts. Yet, another feature are the non-assertive tweets, which focus on promoting the culture and tourism of the area.

Cluster 4 describes the challenger profile in relation to Taiwan. The range of defining variable categories is equally small, yet cluster 4 is clearly defined by the presence of various aspects. Besides the Taiwan focus, there seem to be a strong association between assertiveness, challenging, and the topics of sovereignty and hegemony and security. The summary statistics shows that Taiwan was rarely covered. Filtering for tweets on Taiwan highlights a very narrow field of communication focus, with two thirds focusing on the topic of sovereignty and hegemony, with diplomacy and security receiving some attention.

Cluster 5 is clearly distinguishable from the other clusters as the provoker profile. This cluster is uniquely defined by a high level of assertiveness and the provoke function variable. The functions criticise and challenge also appear in this cluster. Unsurprisingly to China politics researchers, the US and 'The West' are defining locations for this cluster. The topics of foreign affairs, sovereignty and hegemony, and security are strongly associated with it. The summary statistics demonstrates that in terms of frequency of occurrence, provoking and a high level of assertiveness were quite rare. Filtering for these two variables though produces a clear take-away in terms of target: roughly 2/3 of tweets that either provoke or are highly assertive focus on the US. Other locations that appear include special regions of the PRC, Taiwan and the West.

5. Conclusion

This study found that Chinese diplomats employ a range of distinct communication roles on Twitter. Informing and promoting were by far the primary communication functions. The functions of challenging, criticising, and provoking were marginally present, yet assertiveness and their use in a specific content context were their common denominator.

The findings provide evidence that on aggregate level Chinese diplomats are primarily using Twitter as a vehicle for informing and promoting. The informer profile is characterised by neutral language, a range of topics like the economy, technology, health or diplomatic activity, with focus on China itself or in relation to other geographies such as South East Asia. The promoter profile similarly covers a range of different topics, but does so especially with focus on topics related to culture and tourism (e.g., highlighting beautiful landscapes), and is with China and its administrative regions (e.g., Hong Kong, Xinjiang, Tibet) predominantly inward looking. Our findings echo what China researchers may already know about the leadership's priority to portray the "peaceful rise" of China.

Contrasting to this, we also found communication profiles linked to assertiveness and a specific content context. Assertiveness was marginally present, but a defining feature of how their communciation behaviour differs. With a focus on China's administrative regions and Taiwan, we found assertiveness and the challenger profile linked to topics of sovereignty and human rights among other. The provoker profile stands out in utilising assertiveness and even a high level of assertiveness, across the range of functions of criticising, challenging, and provoking, in the context of topics like foreign affairs, sovereignty and hegemony, and security. A high level of assertiveness is with 2/3 tweets mostly associated with the USA, yet we also found prominent the general reference to 'The West'.

Thus, we concluded based on our study on distinctive patterns in the content and style of Chinese diplomatic communications. We discuss our observations of the distinct communciaiton profiles more in the forthcoming publication. Be excited!