Smart cities in India: sketching public opinion trends on smart urban planning narratives

Capstone Project for Data Access and Regulation - Data Analytics for Politics Society and Complex Organizations, University of Milan

Carola Corti - 961514

February 2021

Abstract

This project aims at collecting information about the discourse above smart cities. I will focus on India's specific case, inspecting whether the debate related to smart cities' policies' implementation in such country develops positively or negatively.

The research will be composed of two main parts. In the first, I will provide sentiment and topic analyses processing a corpus of articles previews about India's smart cities development. In the second part, by interacting with the Twitter and Google Maps APIs, I will download via geographical selection the latest 2000 tweets written in India, containing the word "smart cities".

The analyses will be run in order to capture the overall sentiment among the texts to draw a better picture of how smart cities are perceived in its stakeholders' opinion and what can be addressed as a criticisms or advantage in the implementation of such urban planning decisions in the India's context. I will therefore take into account both the governmental and its citizens point of view, and, if the texts will include such topics, I will try to shed light on the role Covid-19 health emergency have played along with overcrowding, poverty and diseased conditions of India's context in relation to smart cities.

Introduction

Smart cities: definition and criticism

According to the definition provided by the European Commission, "smart cities" are places where traditional networks and services are made more efficient with the use of digital and telecommunication technologies for the benefit of its inhabitants and business. Smart metropolitan areas would do so by going beyond the use of information and communication technologies (ICT) in order to provide better resources while using less emissions. The implementation of policies aimed at providing efficient urban transport networks, upgraded water supplies, waste disposal facilities, electricity and heating technologies to let all goods reach each building and each societal stratus in the urban area, are at the heart of smart cities' goals. In other words, the underlining aim of adopting such policies would be ought to implement interactive and responsive city administration strategies, guaranteeing safer public spaces and taking into consideration the needs of the whole population.

Rob Kitchin, one of the leading scholars interested in the debate about smart cities, professor and researcher of human geography at the National Institute of Regional and Spatial Analysis at Maynooth University of Ireland, states that definition of Smart City has to take into account two important aspects. On the

one side, he defines as "smart" a urban context in which *everyware* technologies (Greenfield, 2006) are massively implemented. By using this term, he refers to the massively used ubiquitous computing practices, defined by human-machine interactions extended on several levels, capable of connecting physical and digital space through the continuous collection and management of the data citizens generate interacting with the surrounding material space. On the other hand, he sheds light on the importance of understanding the relationship that occurs between citizens and the technologies they interact with. The implementation of Information Communication Technologies does not in fact constitute, in itself, a necessary and sufficient condition for a city to be defined as really intelligent; it rather is the way in which datafied interactions are used and perceived by who are supposed to be its end-users, to define whether a city is "smart" or not.

While it is easy, in fact, to build city planning projects enthusiastically focusing on the implementation of highly advanced technologies in many fields, it is less immediate to inspect the urban fabrics taking into account both the physical urban environment (i.e. its elements, materials, form, scales, density and networks), and the psychological, socio-cultural, ecological, managerial and economic structures of metropolitan areas. The smart cities approaches have always tended, in fact, to focus on technological solutions to urban problems from the perspective of states and companies rather than critically taking into account people and communities. It is taking into account these considerations, that the debate around Smart City initiatives is more and more heated in developing countries where the ideal prerequisites to advance such policies are more likely to lack.

Smart cities in India

This research focuses on the most prominent example In the Global South: India. This country is in fact, on the one side, a key actor in implementing national-level smart city programs, while, on the other, it constantly faces the challenges caused by an increasingly growing digital divide among its population. According to a recent World Bank report, India is the country with the largest number of non-connected citizens: data show that with a total of 1.25 billion inhabitants, more than 2billion are "offline" citizens.

However, it is precisely in India that incredibly avant-garde realities can be observed from a technological point of view (just think of the Bengaluru district, considered the Indian Silicon Valley, bulwark of technological and innovative New India). The clearest example in the urban planning context is provided by the Smart Cities Mission (SCM), a project that was launched in 2015 as a urban renewal and retrofitting program by Narendra Modi, Prime Minister of the Government of India, with the aim to develop smart cities across the country, making them citizen friendly and sustainable. The plan initially included 100 cities, planning projects to be realized between 2019 and 2023. In spite of the innovation wave the Mission wanted to follow, although extremely advanced in the technological perspective, little consideration was given to its implementation's sociological consequences.

The model of citizenship that is usually offered in smart cities contexts expects, in fact, on the one hand citizens to become "*netizens*", and, on the other, to re-configure citizenship models by instrumentalizing technology and data, reinforcing, as an unintended consequence, the patterns of exclusion for marginalized groups within the urban society, even though citizen consultation and participation are often described as the heart of smart city proposals. It is indeed not immediate for all citizens to participate actively in smart cities projects, and, whether some can, other are excluded, leading to the exacerbation of existing urban historical, material and social inequalities (Katherine S.Willis, 2019, Dasgupta S. 2015).

For instance, the city of Chennai, Located on the Coromandel Coast of the Bay of Bengal, provides clear example of the possible implications such urban planning policies have generated for marginalized communities. In Chennai, the area-based development proposals were mainly focused around the assurance of electrical and water supply, the recycling of waste water, waste management, rain water harvesting and management at households and community levels and Wi-Fi connectivity in order to integrate public transports' utilities. In spite of what could appear at first glance, as Katherine S. Willis points out in her 2019 research on India's smart cities projects, the proposed solutions, although trying to lead to concrete changes in the life-quality within the city area, failed to recognize or give value to the existing urban informality, being rather driven and rationalized by technological deterministic market plans. Urban Informality appeared to be therefore

treated as a set of conditions and resources equivalent to "leaks" in a system that needed to be optimized and rationalized. As a consequence, many plans outlined in 2016 for the Smart City Chennai have failed to be realized, being even criticized through demonstrations of public protests in which banners containing writings such as "*We want bread and butter, not Smart City*" were exposed.

Covid-19 and urban settlements in India

With an estimated 90 percent of all reported COVID-19 cases, urban areas have became the epicenter of the pandemic. The virus have, in fact, highlighted the critical role local governments play as front-line responders within the health crisis context. In this perspective, in this last year, tackling inequalities and development deficits has became of primary importance when it comes to designing policies briefs aimed at preventing the virus spread. Coronavirus has been widening existing social, spacial and economic inequalities in cities both in the Global North and in the Global South. It has in fact became clear that vulnerability to COVID-19 depends on several conditions: where in a city a person lives and works, gender, age, income level, type of home, access to public services - such as health facilities- transportation and clean water.

Data collected by the United Nations and reported in the UN 2020 Policy Brief "*Covid-19 in an Urban World*" , suggest that the poorer neighborhoods and those with the largest households size were more likely to experience a high number of cases per capita. India, with its 35.2% of the country's whole population living in slums and informal settlements, constitutes a paradigmatic example, since many homes in such areas lack access to water and sanitation facilities, making, for example, safe and regular handwashing extremely difficult. It is in this context that the smart policies applied to urban centers are and were expected to be mainly aimed at stemming the contagion, putting in practice the technological advantage that characterize their strength.

Research question

As for the sake of my ultimate research goal to sketch a sort of public opinion analysis regarding this debate, I will focus on topic and sentiment analyses of the corpora I will collect from the voices into play within the discourse related to smart cities in India. I will take into account two main voices: the corporate stakeholder's one, represented by the Indian newspaper **Financial Express** articles, and those coming from a more democratic medium, **Twitter**, even if aware it still not provides a fully reliable representation of the slice of non-digitalized population public opinion. I will try to provide an answer to the following questions:

- How is sentiment (i.e. positive/negative etc.) mapped across the two main voices?
- How is the covid-19 pandemic addressed to in the smart cities discourse?

Analyzing corporate stakeholders' voice: *The Financial Express*

Collecting articles previews from the Financial Express web page

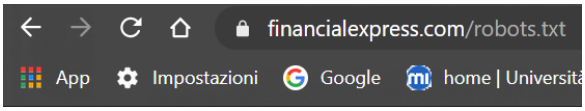
The Financial Express is an Indian English -language business newspaper owned by The Indian Express Group, which has been published since 1961. The newspaper specializes in Indian and international business and finance but its articles address different actuality topics as well. An entire session is dedicated to smart cities. I chose to analyze these short texts since this newspaper can provide great evidence of the discourse around smart cities in the perspective of its political and corporate stakeholders, as the name of the newspaper suggests.

The texts were imported in R following a series of web scraping operations using the R packages *Rvest*, *Httr* and *Stringr,* in order to download textual information to build a corpus by using the *Quanteda* package and,

eventually, analyze it performing text analytics. The aim, here, is to download the provided texts from pages 1 to 5 from the Financial Express website's smart cities-dedicated session. The following steps will report the sequence of code I ran in R to scrape the contents.

Step 1: */robots.txt* file inspection

To check whether the website's rules permitted to perform web scraping on that page, i checked the robots.txt file. As a result, the following lines of code are displayed:




```
User-agent: *
Disallow: /wp-admin/
Disallow: /fhw
Disallow: /travel
Disallow: /pharma
Disallow: /healthcare
Disallow: /adblocker
Disallow: /fepicturegallery.php?imageNo=/
Disallow: /moreHomeListing.php?pg/
Disallow: /fe_full/
Disallow: /*?s
Disallow: /search/
Disallow: /openinappbtn/*
```

The website shows it instructions to web robots by using the Robots Exclusion Protocol. The first line tells us this section applies to all robots; the other tell no robot can crawl the listed paths. I am therefore allowed to scrape the pages I need to collect my corpus since it doesn't belong to this list.

Step 2: Importing the data in R Studio via web scraping with *Rvest*

After having inspected the */robots.txt* file, I've selected the parts of the web page I needed to import in R using Selector Gadget to define the CSS tags corresponding to the elements. I then built a *for loop* in order to extract the main texts in each of the articles shown in the session, downloading it from the first to the last page of this session (pages 1 to 5).

```r
# Loading the required packages

library(stringr)
library(httr)
library(rvest)
library(dplyr)

# Politely asking for permission to scrape

url <-"https://www.financialexpress.com/about/smart-cities/ "

session <- RCurl::getURL(url,
                         useragent = str_c(R.version$platform,
                                           R.version$version.string,
                                           sep = ", "),
                         httpheader = c(From = "barosa.carola@gmail.com"))
```

Let's then download the titles of the articles:

```r
# Scraping the titles

page <- read_html(x = "https://www.financialexpress.com/about/smart-cities/")
nodes <- html_nodes(x = page, css = "h3 a")
knitr::kable(html_text(x = nodes))
```

x
Infrastructure sector crucial for India's economic growth, but these roadblocks need to be managed
Smart cities in India: Top 4 key points to know about Rajasthan's Udaipur
Smart cities mission: Rajasthan sprints ahead, set to start work in Udaipur
Britain offers to develop Pune, Indore, Amravati as smart cities
Smart Cities: Lucknow tops the list of 13 new names
'Expansion of smart cities list to evenly distribute benefits'
Smart cities gets 13 new names; Lucknow tops followed by Warangal, Dharamshala
Public-private partnership is way forward for urbanisation of India: Venkaiah Naidu
Smart Cities, First 20: Moving towards a concrete plan
India projected to add 300 million urban residents by 2050: UN
TT Nagar to be developed under Smart City Project, clarifies MP CM Shivraj Singh Chouhan
Of 100 Smart Cities, 17 figure in the list of WHO's most polluted
BJP member questions govt's smart cities project
Making citizens safe, cities smart
Private sector to play pivotal role in smart cities: Report
Urban Development Ministry asks smart cities to firm up bankable projects
Smart cities may help fight air pollution, say scientists
Delhi ranked 44th among world's 50 'future-ready cities'
The chosen six: Indian cities, European fix
10 things to expect from India's 20 smart cities
Local content requirements, financing challenges for smart cities project: US
Smart Cities list: Tripura 'shocked' over Agartala's exclusion
Over $150 bn investments required for smart cities: Study
Changing Cityscape
Govt unveils winners of 20 smart cities, all you need to know
Realtors see business opportunity from Smart City projects
Smart city winner full list: Check if your city makes it to the list of top 20

Downloading and storing the articles' links (from page 1 to 5) in a newly created directory in my project's folder:

```r
#downloading and getting the links

link_to_pages <- str_c("https://www.financialexpress.com/about/smart-cities/page/", 1:5)
dir.create("SmartCityArticles")

for(i in seq_along(link_to_pages)) {
  download.file(url = link_to_pages[i], destfile = here::here("SmartCityArticles", str_c("page", i, ".h
  Sys.sleep(1)
}

out <- vector(mode = "list", length = 5)

for(i in seq_along(link_to_pages)) {
```

```
  out[[i]]<-read_html(x = here::here("SmartCityArticles", str_c("page", i, ".html")))%>%
    html_nodes(css="h3 a")%>%
    html_attr("href")
}

knitr::kable(out [[1]])
```

x
https://www.financialexpress.com/industry/infrastructure-sector-crucial-for-indias-economic-growth-but-these-roadblocks-need-to-be-managed/2198533/
https://www.financialexpress.com/india-news/smart-city-list-in-india-top-5-key-points-to-know-about-rajasthans-udaipur-city/269102/
https://www.financialexpress.com/economy/smart-cities-mission-rajasthan-sprints-ahead-set-to-start-work-in-udaipur/268554/
https://www.financialexpress.com/economy/britain-offers-to-develop-pune-indore-amravati-as-smart-cities/267827/
https://www.financialexpress.com/economy/smart-cities-in-india-govt-unveils-list-of-13-new-smartcities-lucknow-tops/265061/
https://www.financialexpress.com/economy/expansion-of-smart-cities-list-to-evenly-distribute-benefits/264998/
https://www.financialexpress.com/economy/13-new-smart-cities-unveiled-by-govt-lucknow-bhagalpur-new-town-kolkata-faridabad-make-it-to-the-list/264182/
https://www.financialexpress.com/economy/public-private-partnership-is-way-forward-for-urbanisation-of-india-venkaiah-naidu/263037/
https://www.financialexpress.com/economy/smart-cities-first-20-moving-towards-a-concrete-plan/261702/
https://www.financialexpress.com/india-news/india-projected-to-add-300-million-urban-residents-by-2050-un/258690/
https://www.financialexpress.com/india-news/tt-nagar-to-be-developed-under-smart-city-project-clarifies-mp-cm-shivraj-singh-chouhan/256739/
https://www.financialexpress.com/economy/of-100-smart-cities-17-figure-in-the-list-of-whos-most-polluted/255137/
https://www.financialexpress.com/economy/bjp-member-questions-govts-smart-cities-project/252833/
https://www.financialexpress.com/economy/making-citizens-safe-cities-smart/242299/
https://www.financialexpress.com/economy/private-sector-to-play-pivotal-role-in-smart-cities-report/242239/
https://www.financialexpress.com/economy/urban-development-ministry-asks-smart-cities-to-firm-up-bankable-projects/241680/
https://www.financialexpress.com/world-news/smart-cities-may-help-fight-air-pollution-say-scientists/240097/
https://www.financialexpress.com/economy/delhi-ranked-44th-among-worlds-50-future-ready-cities/237411/
https://www.financialexpress.com/economy/smart-cities-the-chosen-six-indian-cities-european-fix/227656/
https://www.financialexpress.com/economy/10-things-to-expect-from-indias-20-smart-cities/210119/
https://www.financialexpress.com/economy/local-content-requirements-financing-challenges-for-smart-cities-project-us/209475/
https://www.financialexpress.com/india-news/smart-cities-list-tripura-shocked-over-agartalas-exclusion/207186/
https://www.financialexpress.com/economy/over-150-bn-investments-required-for-smart-cities-study/204780/
https://www.financialexpress.com/india-news/changing-cityscape/204305/
https://www.financialexpress.com/economy/smart-cities-list-winners-released-all-you-need-to-know/203918/

x
The infrastructural facilities such as roads, railways, metro rails, and so on are required to potentially increase the product
x
Smart cities in India: The smart city list in India is growing and various cities are moving quickly to drive home the first-r
x
The Smart Cities Mission is gathering momentum with 13 more cities selected for an urban overhaul. The new additions a
x
Britain will be collaborating with the state governments and the municipal authorities to develop Pune (Maharashtra), Inc
x
Lucknow topped the list of 13 smart cities announced by the Centre on Tuesday, taking the total number of such cities cho
x
Real estate industry experts today welcomed the addition of 13 cities to the list of smart cities, saying it will distribute th

x
https://www.financialexpress.com/economy/realtors-see-business-opportunity-from-smart-city-projects/203608/ https://www.financialexpress.com/economy/smart-cities-list-released-top-20-cities-bhubaneswar-jaipur-pune/203440/

Scraping the main text I'll need:

```
#scraping the articles previews texts

text<- rep(list(vector(mode="list", length = 27)), 5)

for(z in 1:5){
  text[[z]][1:27] <- read_html(link_to_pages[z]) %>%
    html_nodes("h4") %>%
    html_text()
}

knitr::kable(head((text [[1]])))
```

Scraping the date of publication for each article in each page:

```
# getting the publication dates

date <- rep(list(vector(mode="list", length = 27)), 5)

for(t in 1:5){
  date[[t]][1:27] <- read_html(link_to_pages[t]) %>%
    html_nodes(".minsago") %>%
    html_text()
}

knitr::kable((date)[[1]])
```

Creating a data frame in which to store the texts and the dates of publication:

x	x	x	x	x	x
Feb 20, 2021	May 30, 2016	May 30, 2016	May 27, 2016	May 25, 2016	May 24, 2016
x	x	x	x	x	x
May 24, 2016	May 23, 2016	May 21, 2016	May 18, 2016	May 16, 2016	May 14, 2016
x	x	x	x	x	x
May 11, 2016	Apr 25, 2016	Apr 24, 2016	Apr 22, 2016	Apr 20, 2016	Apr 15, 2016
x	x	x	x	x	x
Mar 19, 2016	Feb 11, 2016	Feb 10, 2016	Feb 04, 2016	Jan 31, 2016	Jan 30, 2016
x	x	x			
Jan 29, 2016	Jan 28, 2016	Jan 28, 2016			

```
texts <- unlist(text)
dates <- unlist(date)

articles_df <- data.frame(texts, dates, stringsAsFactors = FALSE)

head(articles_df)
```

```
##
## 1                                The infrastructural facilities such as roads, railways, metro rails
## 2 Smart cities in India: The smart city list in India is growing and various cities are moving quickl
## 3 The Smart Cities Mission is gathering momentum with 13 more cities selected for an urban overhaul.
## 4                                                                                Britain will be col
## 5                                                                  Lucknow topped the list of 13
## 6                 Real estate industry experts today welcomed the addition of 13 cities to the list
##          dates
## 1 Feb 20, 2021
## 2 May 30, 2016
## 3 May 30, 2016
## 4 May 27, 2016
## 5 May 25, 2016
## 6 May 24, 2016
```

Tidying the "*dates*" variable using the package *lubridate:*

```
library(lubridate)

articles_df$dates <- mdy(articles_df$dates)

head(articles_df$dates)
```

```
## [1] "2021-02-20" "2016-05-30" "2016-05-30" "2016-05-27" "2016-05-25"
## [6] "2016-05-24"
```

Step 3: Saving the texts in a .csv file

Saving the data frame in a .csv file called *articles_docs.csv.*

```
articles_docs <- write.csv(articles_df,"articles_docs.csv", row.names = TRUE)
```

Creating and analyzing the corpus

Corpus and the Document-feature matrix

Loading the required packages:

```
# Loading the required packages

library(quanteda)
library(quanteda)
library(ggplot2)
library(quanteda.textstats)
```

Reading the *articles_docs.csv* file to create the corpus:

```
x <- read.csv(file = "articles_docs.csv")
```

Creating the corpus:

```
articles_corpus <- corpus(
  x,
  docid_field = "doc_id",
  text_field = "texts",
  meta = list(),
  unique_docnames = TRUE
)

knitr::kable(head(summary(articles_corpus)))
```

Text	Types	Tokens	Sentences	X	dates
text1	29	32	1	1	2021-02-20
text2	34	42	2	2	2016-05-30
text3	36	41	3	3	2016-05-30
text4	22	28	1	4	2016-05-27
text5	26	32	1	5	2016-05-25
text6	30	36	1	6	2016-05-24

Creating the Document-features matrix by removing english stopwords, setting each text in lower case, removing punctuation and numbers.

```
# creating the dfm

myDfm <- dfm(articles_corpus, remove = stopwords("english"), tolower = TRUE,
             remove_punct = TRUE, remove_numbers=TRUE)

knitr::kable(head(myDfm)[,1:8])
```

doc_id	infrastructural	facilities	roads	railways	metro	rails	required	potentially
text1	1	1	1	1	1	1	1	1
text2	0	0	0	0	0	0	0	0
text3	0	0	0	0	0	0	0	0
text4	0	0	0	0	0	0	0	0
text5	0	0	0	0	0	0	0	0
text6	0	0	0	0	0	0	0	0

Most occurent words within the corpus

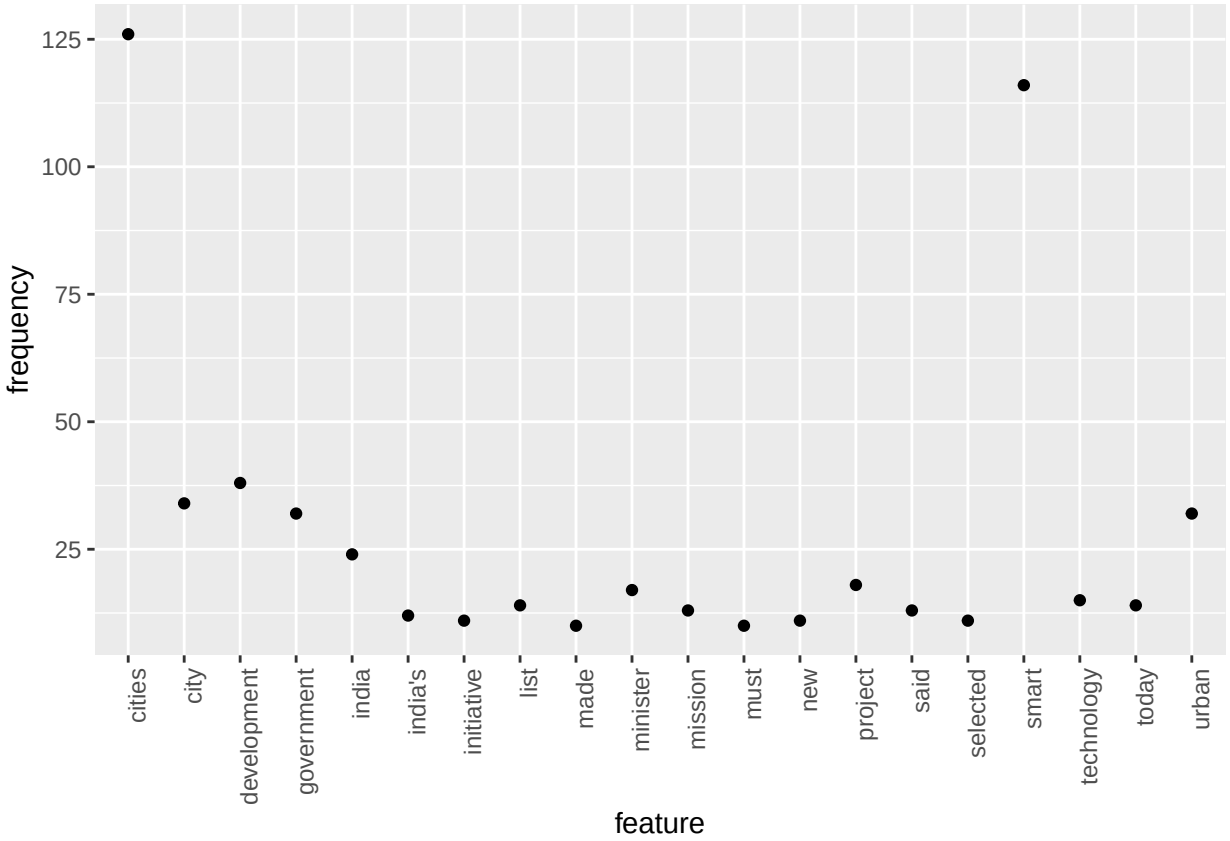
Top 10 words:

```
knitr::kable(topfeatures(myDfm , 10))
```

	x
cities	126
smart	116
development	38
city	34
urban	32
government	32
india	24
project	18
minister	17
technology	15

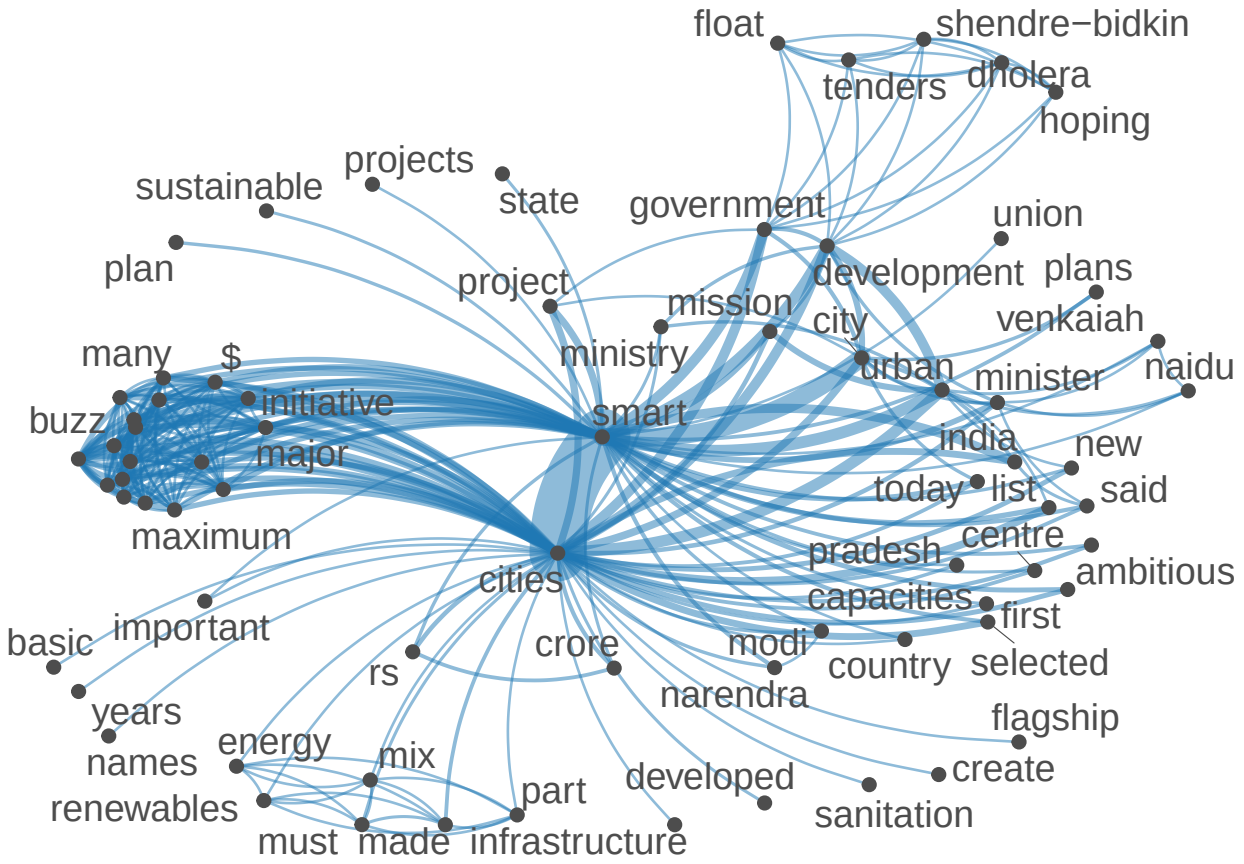
Top features frequency:

```
features_dfm <- textstat_frequency(myDfm, n = 20)

ggplot(features_dfm, aes(x = feature, y = frequency)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



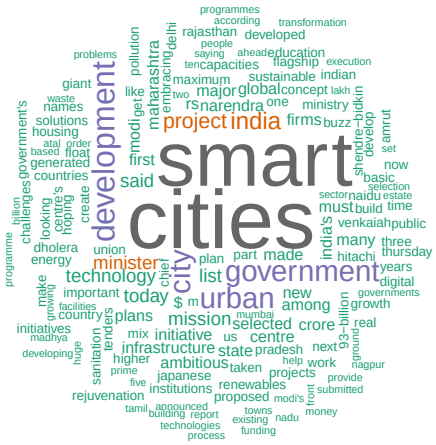
Top features co-occurrence:

```
top_feat <- names(topfeatures(myDfm, 100))
dfm_select <- dfm_select(myDfm, pattern = top_feat, selection = "keep")
set.seed(100)
textplot_network(dfm_select, min_freq = 0.9)
```



Displaying the features in a wordcloud:

```
set.seed(100)
textplot_wordcloud(myDfm , min.count = 6, random.order = FALSE,
                   rot.per = .25,
                   colors = RColorBrewer::brewer.pal(8,"Dark2"))
```



Applying the NRC sentiment dictionary to the corpus

```
library(syuzhet)

get_sentiment_dictionary(dictionary = 'nrc', language = "english")
nrc_vector <- get_sentiment(articles_corpus , method="nrc")
```

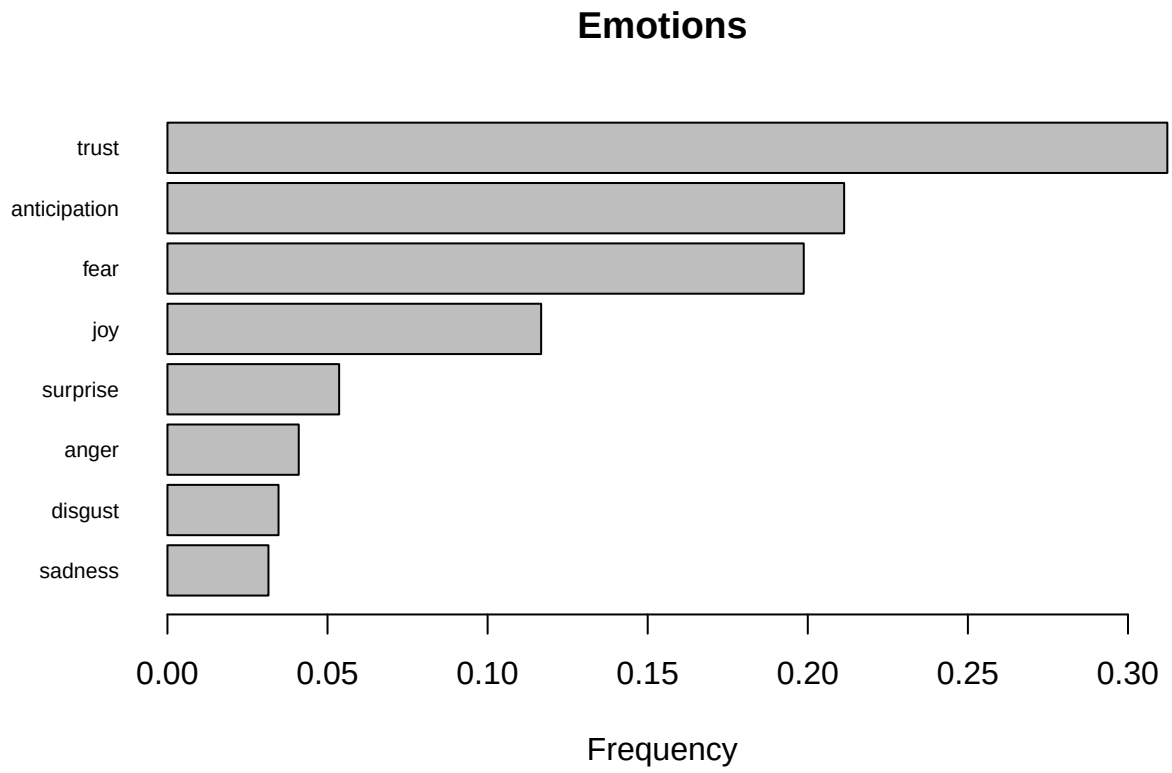
Applying the NRC vector to the corpus:

```
nrc_data_PR <- get_nrc_sentiment(articles_corpus, language = "english")

knitr::kable(head(nrc_data_PR))
```

	anger	anticipation	disgust	fear	joy	sadness	surprise	trust	negative	positive
text1	0	0	0	0	0	0	0	0	0	2
text2	0	0	0	0	0	0	0	0	0	2
text3	0	1	0	0	0	0	0	0	0	2
text4	0	1	0	0	0	0	0	0	0	1
text5	0	0	0	0	0	0	0	0	0	1
text6	0	0	0	0	1	0	0	1	0	3

Plotting the sentiment:

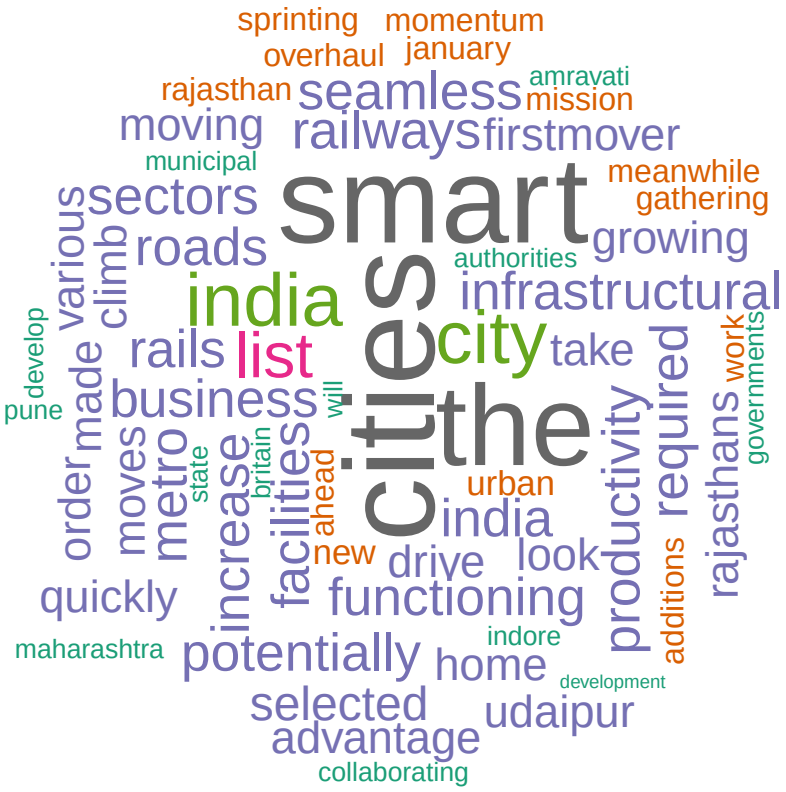
```
barplot(
  sort(colSums(prop.table(nrc_data_PR[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions", xlab="Frequency")
```



Let's now display in a wordcloud the most used words for the two extreme sentiment labels I have identified: the ones that express positivity and those that express negativity:

Positive words:

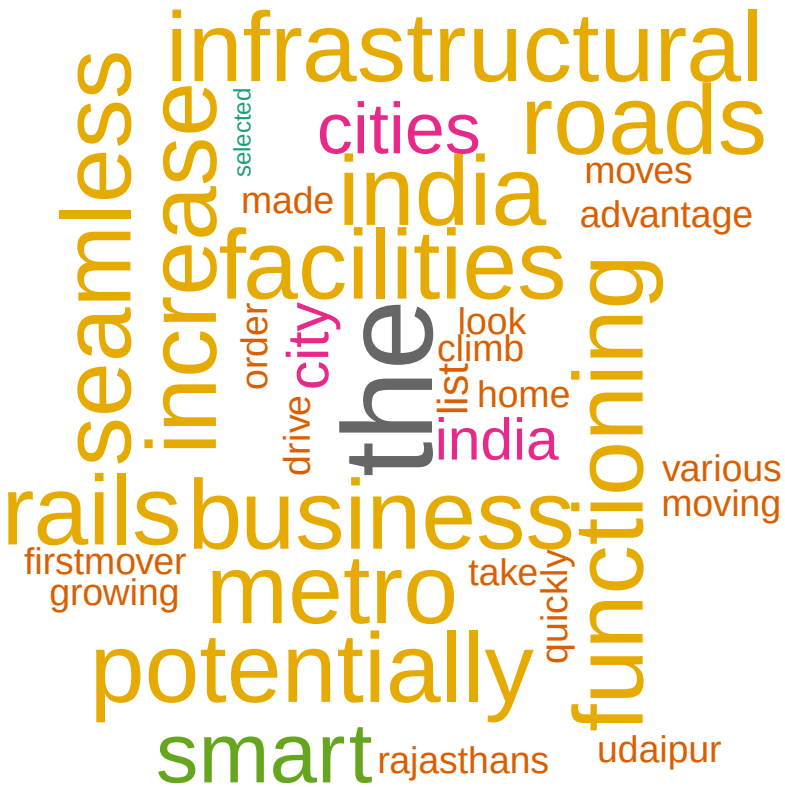
```
library(wordcloud)

set.seed(100)
wordcloud(articles_df$text[nrc_data_PR$positive] , random.order = FALSE,
          rot.per = .25,  min.count = 2,
          colors = RColorBrewer::brewer.pal(8,"Dark2"))
```



Negative words:

```
set.seed(100)
wordcloud(articles_df$text[nrc_data_PR$negative] , random.order = FALSE,
          rot.per = .25,  min.count = 2,
          colors = RColorBrewer::brewer.pal(8,"Dark2"))
```

x	x	x	x
covid	protest*	advant*	ineq*



No difference is observed between the displayed positive and negative words, even though the corpus was analyzed by applying the NRC dictionary, which is able to recognize difference in sentiment within a selected corpus. This lack could be addressed to two main possible reasons: the first provides an explanation lying in the short dimension of the texts, since they're article previews composed of just few sentences each; it is therefore maybe due to the reduced lexical variety of which the short synopses are made of that the NRC algorithm failed labeling correctly the sentiment across texts . The second, instead, addresses this result to the homogeneity of provenience of the articles in corpus: they're in fact all part of the same side of the debate. This may logically lead also to an homogeneity of lexicon and sentiment, since all the texts were written taking into account the same point of view within the debate. The most frequent words are in fact positive and optimist ones, letting us infer the discourse develops almost entirely around the perception of smart city as a positive and optimal solution for urban planning in Indian cities. Moreover, the articles were, for the vast majority, written in 2016: further evidence is therefore provided in favor of the greater presence of lexicon linked to a highly positive sentiment, since the *Smart Cities Mission* was started between those years and its first steps have been followed with great enthusiasm.

Looking for patterns in the use of words *Covid, Protests, Advantage*, and *Inequity*

```
dict <- dictionary(list(covid = c("covid"), prot = c("protest*"), adv = c("advant*"), ine = c("ineq*")))

knitr::kable(dict)
```

Word *covid:*

```
head(dfm_lookup(myDfm, dict)[,1])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 2 docvars.
##        features
## docs    covid
##   text1     0
##   text2     0
##   text3     0
##   text4     0
##   text5     0
##   text6     0
```

The word *covid* doesn't seem to occur in any of the texts. This is due to the fact that these articles were published before the pandemic. Just one document is datable to the post-spread period (the 2021 document) which also doesn't cite it.

Word *protest*:

```
head(dfm_lookup(myDfm, dict)[,2])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 2 docvars.
##        features
## docs    prot
##   text1    0
##   text2    0
##   text3    0
##   text4    0
##   text5    0
##   text6    0
```

The word *protest* is 99.3% sparse within the corpus. This means it occurs with a 0,7% frequency.

Word *advantage*:

```
head(dfm_lookup(myDfm, dict)[,3])
```

```
## Document-feature matrix of: 6 documents, 1 feature (83.3% sparse) and 2 docvars.
##        features
## docs    adv
##   text1   0
##   text2   1
##   text3   0
##   text4   0
##   text5   0
##   text6   0
```

The word *advantage* also occurs with a 0,7% frequency. It occurs, for example. in the second text.

Word *inequity:*

```
head(dfm_lookup(myDfm, dict)[,4])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 2 docvars.
##        features
## docs    ine
##   text1   0
##   text2   0
##   text3   0
##   text4   0
##   text5   0
##   text6   0
```

The word inequity is totally absent within the corpus.

Let's therefore display the contexts in which the words *protest* and *advantage* appear:

```
knitr::kable(head(kwic(articles_corpus, "protest*")))
```

docname	from	to	pre	keyword	post	pattern
text22	22	22	Front government has lodged a	protest	with the Centre on the	protest*

It appears in text 22:

```
texts(articles_corpus)[22]
```

```
##
## "Peeved over the exclusion of state capital Agartala from the Smart Cities list, Tripura's Left Fron
```

But it appears not to be related to a protest against smart cities, displaying, instead, a reverse position.

```
knitr::kable(head(kwic(articles_corpus, "advant*")))
```

docname	from	to	pre	keyword	post	pattern
text2	25	25	to drive home the first-mover	advantage	. Take a look at	advant*

It appears in text 2:

```
texts(articles_corpus)[2]
```

```
##
## "Smart cities in India: The smart city list in India is growing and various cities are moving quickl
```

Unexpectedly, the word *advantage* appears only one time (in text 2). It was in fact expected to be much more frequent on the one side because of its positive sentiment representation (according to the statistics shown above positive words are the vast majority), and because of its meaning, which is strongly intertwined with the progressive rhetoric used in most texts.

Analyzing citizens' voice: a look into TWITTER

Running the query via geolocalization

loading the required packages:

```
library(rtweet)
library(readtext)
library(quanteda)
library(ggplot2)
library(ggmap)
library(httpuv)
library(maps)
library(leaflet)
```

Running the query looking for tweets written in India containing the words "smart" and "city" :

```
# Retriving the tweets via geolocalization interacting with the Google Maps API

rt <- search_tweets( "smart city", n = 1000, lang = "en", include_rts = FALSE, geocode = lookup_coords(

# Adding latitude and longitude

rtll <- lat_lng(rt)

# Saving the results of the query in a .csv file

write_as_csv(rtll, "rtll.csv", prepend_ids = TRUE, na = "",
             fileEncoding = "UTF-8")
```

```
# Reading the file in which data are stored

sctwitter <- read.csv("rtll.csv", fileEncoding = "UTF-8")

knitr::kable(head(sctwitter)[,1:4])
```

user_id	status_id	created_at	screen_name
x991607477320826880	x1366136863464054784	2021-02-28 21:23:04	DeepakGehlod4
x1185398034	x1366116438340104196	2021-02-28 20:01:54	chai2kul
x1185398034	x1364191259074600963	2021-02-23 12:31:56	chai2kul
x1261286841950408705	x1366098760321835009	2021-02-28 18:51:40	TaufiqShahzeb
x3138585722	x1366098736493981696	2021-02-28 18:51:34	HubliHescom
x828520547466231810	x1366092057031376899	2021-02-28 18:25:01	Chandan1pandey

Displaying the first six tweets' texts:

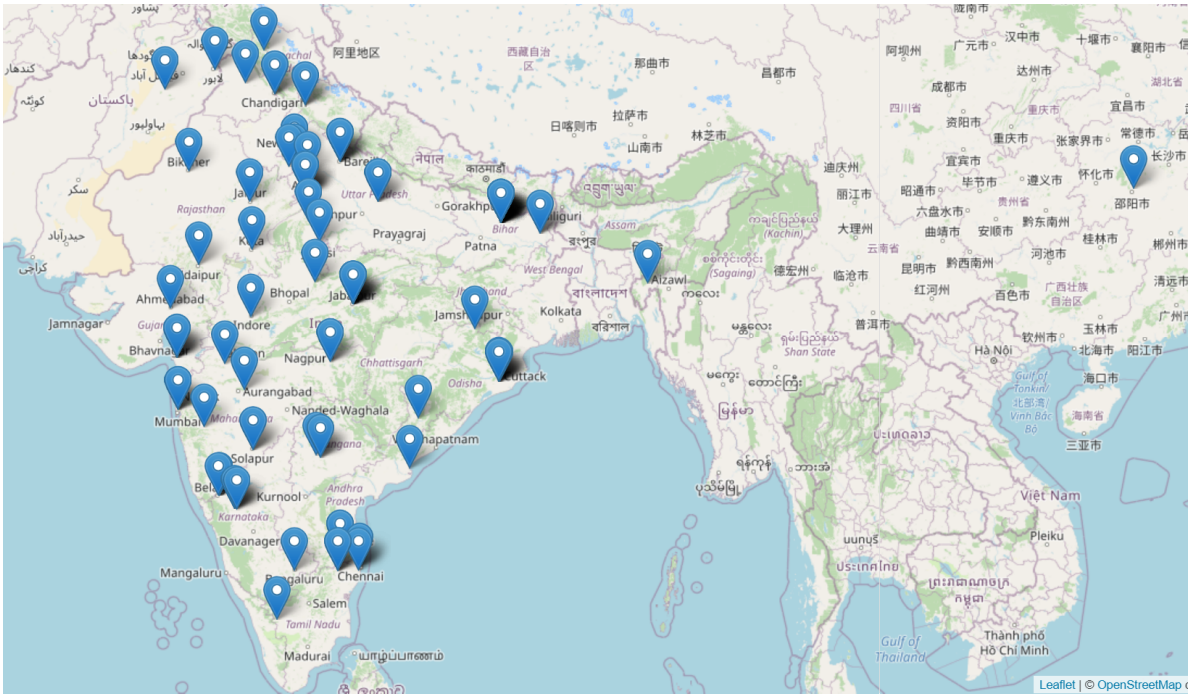
```
head(sctwitter)[,5]
```

```
## [1] "Just posted a photo @ Smart city indore https://t.co/J1x72IQa4u"
## [2] "@gurukumble @SarveshSonar24 CM laid foundation stone for this. Part of Smart City project."
```

```
## [3] "Sabarmati skywalk is gonna be a legendary project. Hope such projects are implemented under Smar
## [4] "@khanumarfa Good to see an alumnus back home.! Lihaaza, ab humara aligarh smart city hogya hai.
## [5] "@_ajay_acharya @Smart_Belagavi @allaboutbelgaum @BenakeAnil Dear Consumer, please contact your l
## [6] "Just posted a photo @ Silvassa Smart City https://t.co/N4GM93u4Pl"
```

Plotting the geocoded tweets using *Leaflet*:

```
m2 <- leaflet(sctwitter)
m2 <- addTiles(m2)
m2 <- addMarkers(m2, lng=sctwitter$lng, lat=sctwitter$lat, popup=sctwitter$text)
m2
```



Displaying the most frequent hashtags:

```
ht <- str_extract_all(sctwitter$text, '#[A-Za-z0-9_]+')
ht <- unlist(ht)
knitr::kable(head(sort(table(ht), decreasing = TRUE)))
```

ht	Freq
#SmartCity	20
#437DaysOfAmaravatiProtests	12
#SaveAmaravati	12
#VisakhaUkkuAndhrulaHakku	12
#modi_job_do	11
#LahoreSmartCity	10

Creating and analyzing the corpus

Corpus and Document-Feature Matrix

```
# Creating the corpus

Twittercorp <- corpus(sctwitter)

#Creating the Dfm

to_remove <- c("<", ">", "+", "-")

myDfmtwitt <- dfm(Twittercorp, remove = stopwords("english"), remove_punct = TRUE, remove_numbers=TRUE,
  dfm_remove(to_remove)


knitr::kable(head(myDfmtwitt)[,1:8])
```

```
## Warning: 'as.data.frame.dfm' is deprecated.
## Use 'convert(x, to = "data.frame")' instead.
## See help("Deprecated")
```

doc_id	just	posted	photo	smart	city	indore	@gurukumble	@sarveshsonar24
text1	1	1	1	1	1	1	0	0
text2	0	0	0	1	1	0	1	1
text3	0	0	0	1	1	0	0	0
text4	0	0	0	1	1	0	0	0
text5	0	0	0	0	1	0	0	0
text6	1	1	1	1	1	0	0	0

Most occurrent words within the corpus

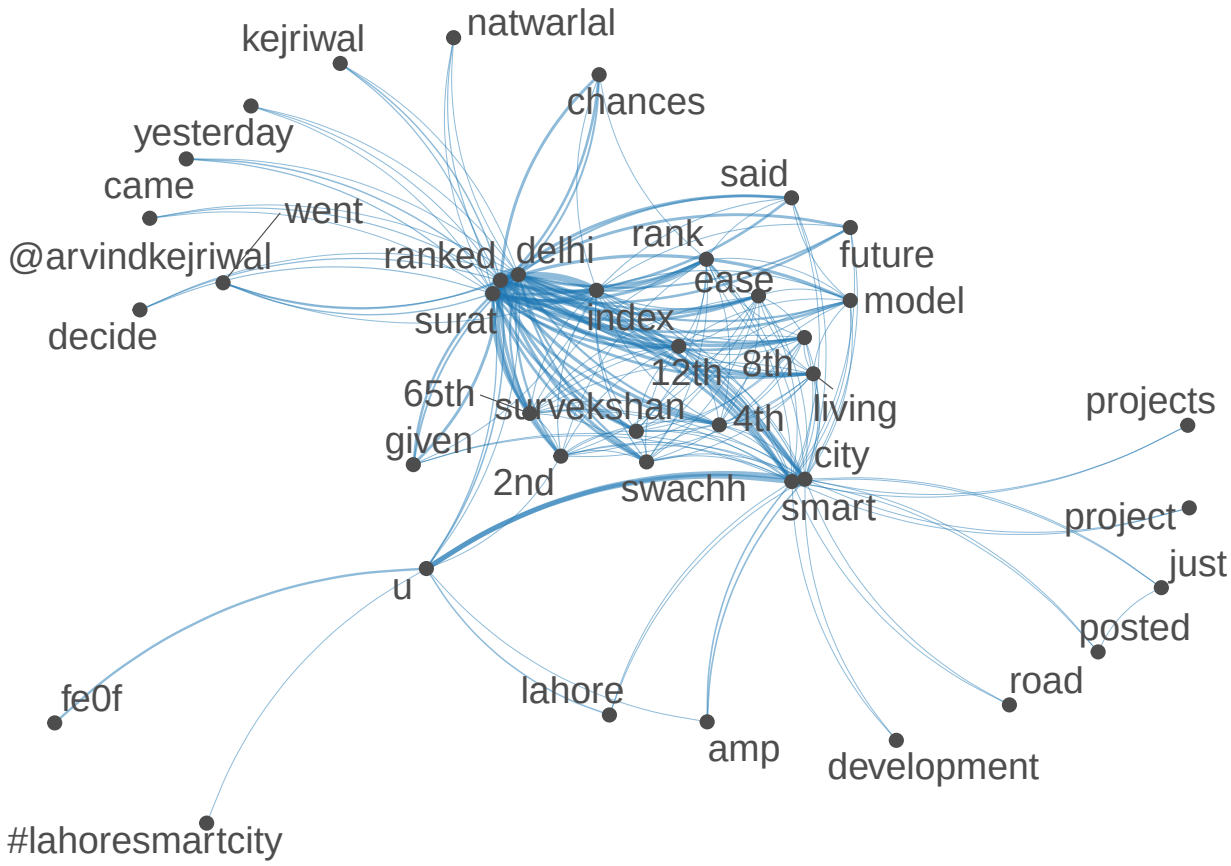
Top 10 words:

```
knitr::kable(topfeatures(myDfmtwitt , 10))
```

	x
city	889
smart	849
u	475
ranked	368
surat	323
delhi	256
amp	131
index	130
just	85
projects	83

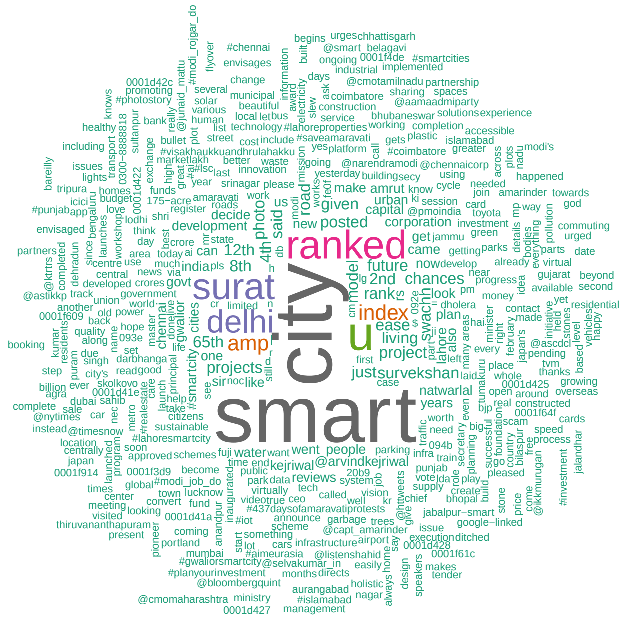
Displaying features in a network plot to show their co-occurrence:

```
top_feat <- names(topfeatures(myDfmtwitt, 100))
dfm_select <- dfm_select(myDfmtwitt , pattern = top_feat, selection = "keep")
set.seed(100)
textplot_network(dfm_select, min_freq = 0.9)
```



Displaying features' frequency in a wordcloud:

```
set.seed(100)
textplot_wordcloud(myDfmtwitt , min.count = 6, random.order = FALSE,
                   rot.per = .25,
                   colors = RColorBrewer::brewer.pal(8,"Dark2"))
```



Applying the NRC sentiment dictionary to the corpus

```
library(syuzhet)

get_sentiment_dictionary(dictionary = 'nrc', language = "english")
nrc_vector2 <- get_sentiment(Twittercorp , method="nrc")
```

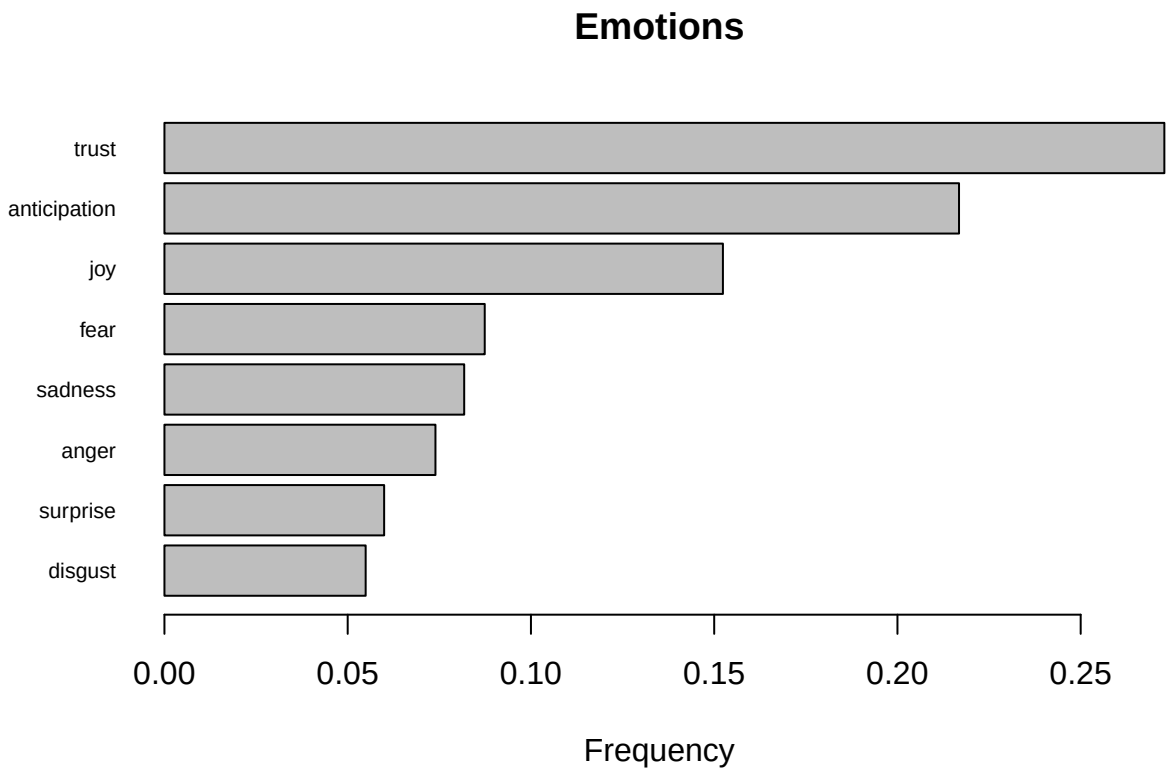
Applying the NRC vector to the corpus:

```
nrc_data_PR2 <- get_nrc_sentiment(Twittercorp, language = "english")
```

Plotting the sentiment:

```
 barplot(
  sort(colSums(prop.table(nrc_data_PR2[, 1:8]))),
  horiz = TRUE,
  cex.names = 0.7,
  las = 1,
  main = "Emotions", xlab="Frequency")
```

x	x	x	x
covid	protest*	advant*	ineq*



Even in the Twitter context, the vast majority of words were labeled as really positive, representing trust, anticipation and joyful sentiments over all. The overall sentiment frequency is really similar, to the one displayed in the previous session among the articles corpus.

Looking for patterns in the use of words *Covid, Protests, Advantage*, and *Inequity*

```
dict <- dictionary(list(covid = c("covid"), prot = c("protest*"), adv = c("advant*"), ine = c("ineq*")))

knitr::kable(dict)
```

Word *covid:*

```
head(dfm_lookup(myDfmtwitt, dict)[,1])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 91 docvars.
##        features
## docs    covid
##   text1     0
##   text2     0
##   text3     0
```

```
##   text4     0
##   text5     0
##   text6     0
```

Surprisingly, the word *covid* doesn't appear in any of the texts even though the collected tweets were written in the last days of February 2021.

Word *protest**:

```
head(dfm_lookup(myDfmtwitt, dict)[,2])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 91 docvars.
##        features
## docs    prot
##   text1    0
##   text2    0
##   text3    0
##   text4    0
##   text5    0
##   text6    0
```

*Word advant**:

```
head(dfm_lookup(myDfmtwitt, dict)[,3])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 91 docvars.
##        features
## docs    adv
##   text1   0
##   text2   0
##   text3   0
##   text4   0
##   text5   0
##   text6   0
```

Word *ineq**:

```
head(dfm_lookup(myDfmtwitt, dict)[,4])
```

```
## Document-feature matrix of: 6 documents, 1 feature (100.0% sparse) and 91 docvars.
##        features
## docs    ine
##   text1   0
##   text2   0
##   text3   0
##   text4   0
##   text5   0
##   text6   0
```

Surprisingly, no features among those I wanted to inspect seem to appear within this corpus.

Conclusions

No difference in sentiment or lexicon appear to occur between the two corpora I've analyzed. The aim of this research was to address these differences in order to sketch public opinion sentiment trends for both the clashing parts in the debate around smart cities in India; no inference can be made since the collected and analyzed data do not provide evidence to sustain my initial hypothesis in favor of the actual existence of two clashing voices in this context. Moreover, in spite of what was addressed by Katherine S.Willis (2019) about protests against the exclusion effects generated among the urban poor by smart cities policies, no words seem to be related to discontent or public protests against the implementation of such projects.

The results I obtained can instead be explained by the occurrence of the following phenomena:

- The impossibility, for the most disadvantaged societal strata, to efficiently
  express their opinion on the matter. It really difficult, in fact, for people leaving in the most diseased conditions, to express it in writing form and/or on social networks due to illiteracy, poverty and formal social exclusion from the public and political debate.
- My impossibility of analyzing tweets written in hindi, which I wasn't able to do since I don't know this language. The choice of analyzing only english tweets may have in fact affected the results since, even though english is India's second language, it is maybe not the most used language among those who, in my opinion and according to K.S.Willis research, can be "smart cities victims".

References

- Cardullo P, Di Feliciantonio, C., Kitchin, R. "The right to the smart city" Emerald Publishing, Wagon Lane
- Coletta C., Leighton E., Heaphy L., Kitchin, R. 2019 "Creating Smart Cities" Routledge, Oxon; New York
- Kitchin, R, 2014 "The real -time city? Big Data and smart urbanism" Geojournal 79 (1) 1-14
- United Nations, 2020 "Policy brief: Covid-19 in an Urban World"
- Financial Express Website: https://www.financialexpress.com/

Wordcount:

```
## [1] 3177
```