

Text Mining and Analysis - Some Examples

Kalidas Jana
Post-Doctoral Research Fellow, Economics/Data Science
Department of Economics
Fogelman College of Business and Economics
University of Memphis

DSAS Lab Seminar
February 21, 2020

Abstract

Using three articles from the magazine THE ECONOMIST, I discuss some steps of text mining and analysis. Particular aspects of text mining and analysis I look at are: text cleaning, term-document-matrix, world cloud, comparison cloud, sentiment analysis, and topic modeling.

Key Words: Text Cleaning; Term-Document-Matrix; World Cloud; Comparison Cloud; Sentiment Analysis; Topic Modeling.

- **Objective of the Presentation:** Discuss steps I followed to create some examples of text mining and analysis by selecting three recent articles relating to [COVID-19](#) from the magazine [THE ECONOMIST](#)

Step 1: Collected articles and read them in

- Collected 3 articles from 2020 Jan and Feb issues of [THE ECONOMIST](#)
- List of files =
 - [1] “Companies warn of economic crisis as China fights the coronavirus.txt”
 - [2] “How do you contain a global epidemic such as coronavirus.txt”
 - [3] “The race to produce a vaccine for the latest coronavirus.txt”
- Documents read:
 - [1] **“Viral injections: Companies warn of economic crisis as China fights the coronavirus.**

[Finance and economics Feb 6th 2020 edition Feb 4th 2020 SHANGHAI](#)

RARELY HAVE plans in China fallen apart so swiftly, so publicly. On January 12th the leaders of Hubei declared that the province’s GDP would grow by 7.5% this year. They also vowed to make the province a stronger link in high-tech supply chains. They made no mention of a mysterious new virus that was causing pneumonia and spreading fast through the cities and towns under their watch. But less than two weeks later its scale was too big to ignore. Under intense pressure to act, they placed the entire province under quarantine, hemming in 60m people and rendering their flashy economic targets almost certainly unreachable this year. Their focus instead shifted to stopping the illness and keeping people supplied with necessities. The lurch from confidence to anxiety has echoed throughout China. In the months leading up to the coronavirus outbreak, the stockmarket had rallied. Businesses were upbeat about their prospects this year, not least because China and America had finally reached a deal in their long-running trade war. But over the past two weeks, as the government has begun a full-scale fight against the epidemic, optimism has crumbled. Share prices in mainland China have fallen by 10% since January 20th. Factories and offices, already shut for the new-year holiday, were supposed to reopen in recent days but many have stayed shut. Most provinces have ordered them to remain idle until February 10th, if not longer. Poultry farmers have warned that their chickens might starve because roadblocks have snarled their feed supplies. Businesses have started dipping into their cash reserves. Restaurants and hotels have been hit especially hard because few people anywhere in China, not just Hubei, dare venture out. In one interview that was shared widely on social media before being censored, Jia Guolong, founder of Xibei, a popular restaurant chain, said that if the lockdown persisted for a few more months, vast numbers could lose their jobs. “Wouldn’t that be an economic crisis?” he asked. Analysts have rushed to cut their forecasts for economic growth. The consensus had previously been that GDP would expand by about 6% year-on-year in the first quarter. Now several think that 4%, the slowest since China started publishing quarterly figures in 1992, is more likely, with risks firmly tilted to the downside. As with past epidemics, there is sure to be a strong recovery when the virus is eventually contained. But there is much uncertainty

about when that might be. Three unknowns will dictate the recovery's timing: how long it takes to bring the virus under control; how long after that the government relaxes its heavy-handed restrictions on daily life; and how long after that people resume the whirl of activity that normally makes the Chinese economy so vibrant. For economic policy, this presents a challenge. Usually, the further into the future you peer, the greater the uncertainty. But China's officials can be reasonably confident in assuming that growth will return to its pre-virus trajectory next year. It is the next couple of months that are the black hole. In this environment flexible measures to help people and companies through a difficult patch are most sensible. These can be pared back when the rebound arrives. Getting them right, though, is not easy. It is worth noting what China is avoiding, so far at least. Some have speculated that officials will unleash a big stimulus, perhaps a vast new array of infrastructure projects, to get growth back up to speed. But it is too soon for that. The government does not want people on building sites or in factories at the moment. Much of its efforts are aimed at keeping them in their homes, in order to prevent the virus spreading. Moreover, there is a lag between unveiling infrastructure plans and breaking ground. The boost from projects announced today could start as the economy is gathering steam on its own, leading to overheating. Instead, China is using a combination of temporary cash support, market interventions and forbearance to get through the crisis. On February 3rd the central bank made headlines by injecting 1.2 trillion yuan (172 billion dollars) into the financial system (it bought treasury bonds from banks which promised to buy them back within 14 days). Banks will probably suffer from rising loan defaults in the coming weeks, and this gives them more cash to work with. When the repurchase agreements come due, the central bank could choose, in effect, to extend them if needed. Officials are meddling in the market (or, as they would say, managing it) out of concern that investors may be too pessimistic in the near term. Because many companies have pledged their equity as collateral for loans, they would need to sell assets as share prices fall, only adding to the downward pressure. So regulators, having already delayed the reopening of the stockmarket after the new-year holiday, told brokerages to bar clients from short-selling, according to Reuters. Chinese shares still dropped by 8% on February 3rd, their steepest one-day fall since 2015, but they were largely catching up with the Hong Kong market, which had been open the previous week. On February 4th, shares rose a little more than 1%, suggesting that the stabilisation tactics were working. Finally, officials have been advocating and orchestrating forbearance on various fronts. Shanghai was due to increase companies' social-security contributions on April 1st. The city has delayed that by three months, saving firms an estimated 10bn yuan. In Beijing, the municipal government has encouraged landlords to cut their commercial tenants' rents; in exchange, it will provide them with subsidies. And regulators have called on banks throughout the country to roll over loans to companies, such as small manufacturers, which would otherwise lack the cash buffers to survive the work stoppage. Even as the death toll continues to mount, some officials are already thinking about economic distortions that have arisen in the course of the battle against the epidemic. Hospitals have faced shortages of protective equipment such as masks, gowns and gloves. So the government has called on companies to increase production. Many, feeling a sense of duty, have heeded the call. But as Liu Shangxi, an adviser to the finance ministry, has noted, this means that medical-equipment firms will suffer from severe overcapacity after the crisis passes. The government should thus be

ready, he argues, to compensate them. Such proposals are a far cry from the bold growth-and-investment plans that Hubei's provincial leaders laid out less than a month ago. Yet the priority now is not to stimulate the economy or climb the technology ladder but to ensure that society remains stable as the quarantines and controls drag on. China's grim new reality is that everything, economic policy included, revolves around the question of how to beat the virus."

[2] **“Prepare for the worst: How do you contain a global epidemic such as coronavirus? Health officials and doctors can only do so much to stop the spread of diseases.**

[International Jan 30th 2020 edition The Economist](#)

“MADE IN CHINA” is a label the country's government would prefer to be associated with slick technology. Its trending export at present, however, is 2019-nCoV, the new coronavirus that struck in the Chinese city of Wuhan in December. The virus has now spread to at least 16 countries. As *The Economist* went to press, the World Health Organisation (WHO) and China had confirmed almost 7,800 infections and 170 deaths, almost all in China. When a new infectious disease begins to spread, decisions on how to stop it are based on patchy data that change by the hour. This is “the fog of war” phase, says David Heymann of the London School of Hygiene and Tropical Medicine. Health officials have to make decisions quickly and with uncertain information, says Jeremy Farrar of the Wellcome Trust, a charity. They must first determine the deadliness of new viruses. The first cases diagnosed are usually among the worst because those people are ill enough to go to hospital. Zika is a mosquito-borne virus that commonly causes nothing more than mild flu-like symptoms. But the first recorded cases were mostly mothers who contracted the infection during pregnancy and whose babies were born with brain damage as a result. As health officials start actively trying to identify infected people, milder cases are added to the total. As a result, early estimates tend to overstate the danger of new diseases. That is happening now with the Wuhan virus. At the end of January, reported deaths represent about 2% of confirmed infections. Around 20% of those reported to be infected become severely ill, suffering from pneumonia and respiratory failure. But modelling by Gabriel Leung and Joseph Wu at the University of Hong Kong suggests as of January 25th that the number of infections in Wuhan was closer to 44,000 (with a range of 20,000-78,000). Most of those infections will be mild, so the death rate for the virus could be as low as 0.1%-no deadlier than the common flu in America. Officials must then gauge how contagious a new virus is. As growing numbers arrive in hospitals, patterns emerge. If it turns out that most of the newly infected people are health-care workers and relatives of the sick, that would probably mean that the virus is transmitted through close rather than casual contact, so stemming its spread should be easier. Experts must next determine how it is passed from person to person. The common cold spreads through virus-laden droplets from coughs and sneezes that travel only a few metres. Influenza and measles are far more contagious because they ride on airborne particles-so a sneeze can infect an entire room. It is not yet clear how the Wuhan virus is transmitted. The WHO thinks that, like the one that causes Severe Acute Respiratory Syndrome (SARS), it hitches a ride on droplets. It is also unclear how often those who are infected but show no symptoms can spread the virus to others. Some infections in China and Germany seem to have been the result of this kind of transmission. If so, contagious people could be unknowingly infecting others for days. Both

SARS and Middle East Respiratory Syndrome (MERS), another lethal coronavirus, had superspreaders—patients with unusually high viral loads, who are exceptionally infectious. In South Korea in 2015 a patient with MERS infected 81 people during a 58-hour stay at a hospital emergency room. The threat of a global pandemic will put to the test the preparedness plans that countries and big cities have in place. These feature a worst-case scenario, usually the arrival from abroad of a hypothetical strain of influenza that is both very deadly and highly contagious—a rare combination of features that set apart the Spanish flu which swept the world in 1918, killing 20m-50m people. City authorities and hospitals routinely carry out drills to test their readiness for such a scenario. In some cases officials huddle in a conference room and plan what they would do. On January 24th New York City’s top officials held such a drill for the Wuhan virus. In other exercises, doctors and health officials don protective gear and get out on the streets to practise their response. To test its system, New York City routinely uses mystery patients who show up at hospitals pretending to have symptoms of notifiable diseases that doctors are supposed to report to public-health departments. When an outbreak starts to cross borders, as is now happening with the Wuhan virus, the knee-jerk reaction is to set up airport health-checks for passengers arriving from outbreak hotspots. But many health experts think such tests are a waste of time and money. In Canada, screening for SARS at airports in 2003 detected no instances of the disease; that year SARS killed 774 people, including 44 in Canada. The theatrics of airport checks suit politicians, who are anxious to be seen to be responding to worried citizens. It is more useful to the public to provide those arriving at airports with information explaining what to do if they develop symptoms, says Agoritsa Baka of the European Centre for Disease Prevention and Control. Efforts are better spent boosting infection-prevention measures at hospitals, says Ms Baka. Health workers are often among the first to be infected by a new virus, which they pick up from patients. They then pass it on to their families and other patients. In the global SARS outbreak in 2002-03 about a third of those infected were health workers. The best way to nip an outbreak in the bud is for disease detectives to locate those infected. The goal is to prevent them from passing the virus to others, by isolating them in hospitals and at home. Most countries ask people to quarantine themselves voluntarily. Some will demand court orders to enforce such rules if people rebel. When an outbreak grows from a few clusters of cases into an epidemic, cities may go further in their efforts to keep people apart. Japan and some European countries close schools for short periods if flu seasons look as though they will be particularly bad. In 2009 Mexico City shut down bars, cinemas, churches and football stadiums for 13 days to try to stop the spread of swine flu. Locking down large areas, as China is currently doing with the entire province of Hubei, with a population of nearly 60m, is untested in modern times. Such efforts can backfire. One lesson from the Ebola outbreaks in west Africa is that if those under quarantine are not cared for and do not feel that the suffering they are enduring for the common good is respected, they will try to evade the quarantine, says Jeremy Konyndyk of the Centre for Global Development, a think-tank in Washington. That makes matters worse because in an outbreak it is crucial to know who is infected, where they have been and where they are going. A heavy-handed attempt to quarantine West Point, a settlement of 70,000 people in Monrovia, Liberia’s capital, during the Ebola outbreak in 2014 was abandoned after residents responded with riots. By contrast, a similar but well-organised quarantine in Sierra Leone, in which traditional leaders were brought

on board first, did not meet resistance. If China's drastic measures help delay epidemics of the Wuhan virus in other countries by a few months, that could make a huge difference, says Dr Farrar. Hospitals in Europe and America will be better placed to handle a surge of infections in late spring, compared with February when they are overwhelmed by the peak in cases of the seasonal flu. Such a delay could also be crucial for testing a vaccine for the Wuhan virus. Several are already in the works in China, America and Australia. Dr Farrar reckons a vaccine could be ready for clinical trials in 6-12 months. When it is clear that an epidemic cannot be controlled, authorities go into mitigation mode. That involves setting up places to care for patients when hospitals overflow, and systems to identify which patients should be treated first when medical supplies run short. But few countries put such measures in place in advance. China is frantically building extra hospitals in Wuhan to accommodate the current surge in numbers of patients. Sometimes no preparation can suffice. It will not be possible to halt the spread of a new pandemic influenza virus, and it would be a waste of public-health resources and capacity to attempt to do so, admits Britain's flu pandemic preparedness strategy. At that point, officials are left hoping for the best."

[3] **“Run, don’t walk: The race to produce a vaccine for the latest coronavirus. Even if scientists are too late for this outbreak, their work will not be wasted. International Feb 8th 2020 edition The Economist**

IN RECENT WEEKS searches on Google for “contagion movie” have soared. In the film, a thriller from 2011, a virus spreads rapidly around the world, killing 26 m people. The plot follows the frantic efforts of scientists to produce a vaccine. Some 133 days after the first infection, they succeed. In the real world most recent vaccines have taken years to develop. Some have taken more than a decade. Others, such as a vaccine to stop HIV, the virus that causes AIDS, still elude scientists. But technological innovations and a more streamlined development process could dramatically shrink the time it takes to produce a vaccine against a new pathogen that has the potential to cause an epidemic. The new coronavirus that emerged in the Chinese city of Wuhan in December presents vaccine-makers with an urgent test. It has so far killed almost 600 people and infected more than 28,000. Scientists in China published the Wuhan virus's genetic sequence on January 12th, less than a week after they isolated the bug from a patient suffering from a mysterious respiratory infection. By late January, several groups around the world had started work on a vaccine using these genetic data. The first clinical tests on humans, for safety, could begin as early as April. With luck, a vaccine could be ready within a year. Next week the World Health Organisation (WHO) will convene a global meeting to set a research agenda. It will agree on rules, or protocols, for trials and work out which medical advances should be priorities. People have rushed to make new vaccines before. The west African Ebola outbreak of 2013-16 tested the world in many ways, but particularly in the need to speed up the delivery of new treatments. Organisations and institutions that normally work slowly, and at arm's length, came together to get the job done faster. Drug regulators from America and Europe, pharmaceutical firms, charities, experts and the WHO all worked closely to advance the trials and technologies needed. They succeeded. An outbreak of Ebola in 2018 in the Democratic Republic of Congo, which now appears to be on the wane, has been contained largely as a result of the wide availability of a vaccine. This process of scientific acceleration is under way again, this time “on steroids”, says Seth Berkley, the

boss of GAVI, a vaccine-finance agency. Even if a vaccine were ready within a year, it would be too late to stem the current epidemic in China. But it could help other countries. Fears are growing that the Wuhan virus will spread more widely and become an established seasonal disease around the world, like the common flu. China's extraordinary efforts to contain the virus, including quarantining over 50 m people, may stave off epidemics in other countries until next winter. It is too soon to tell how deadly the Wuhan virus is. But if it is at least as bad as seasonal flu, a vaccine for those most at risk will be vital. In 2017-18 more than 800,000 people were hospitalised and about 60,000 died in America alone as a result of influenza. The rush to develop a vaccine against the Wuhan virus has been led by the Coalition for Epidemic Preparedness Innovation (CEPI), a group set up in 2017 in the wake of the west African Ebola outbreak. CEPI's purpose is to forearm the world against future outbreaks of disease, without knowing what those diseases will be. Its aim is to have a vaccine against a previously unknown pathogen ready to test in humans within 16 weeks of its identification. To that end, some of the university research centres and biotechnology firms that it has funded have been working on "plug-and-play" vaccine design and manufacturing technologies that can be used for a number of pathogens. This allows the genetic sequence of a particular pathogen to be slotted into an existing molecular platform that forms the basis of the vaccine. In the past, laboratory work on a vaccine required stocks of the actual virus. It would be treated to make it harmless but still able to tickle the immune system into producing antibodies-proteins that fight off the wild virus if it attacks. Working with a deadly virus is tricky, naturally. It requires special containment facilities and exhaustive procedures to prevent it from escaping or infecting scientists. Gene sequencing has made this process quicker, safer and easier. Researchers can build synthetic versions of parts of viruses to work on vaccines without needing complete samples of the pathogens. Scientists have produced vaccines against other viruses, including Zika, Ebola and two other coronaviruses-SARS (Severe Acute Respiratory Syndrome) and MERS (Middle East Respiratory Syndrome)-using such technology. The vaccine research on these two cousins of the Wuhan virus has come in handy in recent weeks.

Going viral

Once a vaccine has been developed in a laboratory, it is sent to a factory where it is turned into a sterile vaccine mix. This is then put into vials and tested to ensure it is not contaminated before clinical trials in humans can be carried out. Many of these tests are done in petri-dishes; the process takes several months. Genetic sequencing can do the job much faster. By sequencing the DNA of everything in a vial of vaccine and examining the result, scientists can spot traces of viruses that should not be present. Vaccine research groups in Britain are in talks with the country's medicines regulator about an approval process for such alternative testing methods. The development of a vaccine can be speeded up if bottlenecks in the process are eliminated, says Sarah Gilbert. She leads a group at Oxford University which is working on a vaccine against the Wuhan virus. Her group has developed a template for vaccines that can be adapted quickly for new pathogens. The researchers can make the first small quantities of a new vaccine in just six to eight weeks. In the past the process would have taken up to a year. The other groups trying to come up with a vaccine for the Wuhan virus are using similar methods involving templates that have already been proven to work. Faster regulatory approval can also speed vaccines through clinical trials. Even as it started making the vaccine, Dr Gilbert's group began putting

together an application for clinical trials for it. The group plans to apply for an expedited ethical and regulatory review, which can be granted within days as it was for clinical trials of the Ebola vaccine conducted in Britain in 2014. Normally, the process takes about three months, says Dr Gilbert. Even if a vaccine is developed and approved, the rapid rise in cases of the Wuhan virus in China and its spread to other countries has created a new urgency: planning ahead for ways to make massive quantities of a vaccine quickly. There are not many factories that can mass-produce vaccines, so new vaccines often wait in a long queue. Aware of this problem, the American government has built dedicated manufacturing facilities that can produce vaccines rapidly for emergencies. Britain is doing something similar. When CEPI was planning its work, those involved were thinking about epidemics (outbreaks limited to one country), not pandemics (global epidemics), explains Richard Hatchett, the head of the group. Last week CEPI put out a call for vaccine candidates for the Wuhan virus that can be manufactured on a large scale with existing capacity. On February 3rd it brought on board as a partner GSK, a big drug firm, which has agreed to lend its highly effective adjuvant to a new vaccine. An adjuvant is a special ingredient that makes vaccines more efficient by boosting the immune response-which means that fewer doses of the vaccine or a lower concentration of its core ingredient is needed for vaccination. Even if a vaccine can be produced in sufficient quantities, getting it to the people who need it, regardless of where they live, can still be a problem. In theory, a vaccine for the Wuhan virus would go to those most at risk, such as health workers, the elderly and those with conditions that appear to make the virus more lethal, such as patients with immune deficiencies. The problem is that politics often intervenes during a pandemic, and governments that are the home to vaccine-making facilities can requisition some of it for their own use, citing national defence or security. This is a problem Mr Hatchett knows all too well; he worked at the White House on medical preparedness during a flu pandemic in 2009. The outbreak had a very low mortality rate, but exporting any vaccine before it was available to American citizens quickly became a vexed issue. Mr Hatchett is working with the WHO to try to ensure that the Wuhan virus vaccine is made at a number of different sites around the world including ones in small countries which would quickly be able to meet the needs of their entire populations. A jab in the dark The issues surrounding any potential vaccine make questions about medicines to treat those who have become gravely ill particularly acute. Licensed medicines to treat coronaviruses do not currently exist, but experimental drugs are in development, with some early data on their use. One that has been highlighted as promising is called remdesivir, which is made by Gilead, a drug firm. Two randomised controlled trials will start enrolling patients in mid-February. Remdesivir was developed to treat Ebola but in laboratory tests has been shown to be effective against a range of viruses. A combination of two drugs usually used in HIV treatment also looks promising and is already being tried on patients, says Vasee Moorthy who helps set research and development priorities at the WHO during epidemics. Randomised controlled trials-in which some people are given the drug being tested and some are given a placebo-are the gold standard of scientific evidence. These will probably go ahead in the coming weeks when it is clear which drugs seem most promising. Trials with hospitalised patients will probably involve a placebo arm. Everyone in the trial would receive intensive care but some would also be given the drug being tested. This is because no one yet knows whether the new drugs, which may have side effects, do more harm than good. The most gravely ill patients

may also be allowed to try untested drugs. Only so much preparation is ever possible in advance of a new disease. A drug or vaccine’s efficacy can only be tested during an outbreak. The urgency behind the search for treatments for the Wuhan virus is understandable. Such efforts were effective in the case of Ebola. People are willing to rush vaccines and drugs into use for a disease with a fatality rate around 70%, as Ebola’s was. The calculus is different for one that kills 2% (or less) of those infected. Should hasty decisions lead to products that are not completely safe, people’s faith in vaccines could be damaged. If so, the harm done to the world’s health could rival the worst feared of the Wuhan virus.”

≈≈≈ * ≈≈≈

Step 2: Cleaned up text, converted each document to character strings, and created a corpus

corpus =

[1] “viral injections companies warn economic crisis china fights coronavirus government steps support boost growth ensure social stability finance economics feb th edition feb th shanghai rarely plans china fallen apart swiftly publicly january th leaders hubei declared province gdp grow year also vowed make province stronger link high tech supply chains made mention mysterious new virus causing pneumonia spreading fast cities towns watch less two weeks later scale big ignore intense pressure act placed entire province quarantine hemming people rendering flashy economic targets almost certainly unreachable year focus instead shifted stopping illness keeping people supplied necessities lurch confidence anxiety echoed throughout china months leading coronavirus outbreak stockmarket rallied businesses upbeat prospects year least china america finally reached deal long running trade war past two weeks government begun full scale fight epidemic optimism crumbled share prices mainland china fallen since january th factories offices already shut new year holiday supposed reopen recent days many stayed shut provinces ordered remain idle february th longer poultry farmers warned chickens might starve roadblocks snarled feed supplies businesses started dipping cash reserves restaurants hotels hit especially hard people anywhere china just hubei dare venture one interview shared widely social media censored jia guolong founder xibei popular restaurant chain said lockdown persisted months vast numbers lose jobs wouldn economic crisis asked analysts rushed cut forecasts economic growth consensus previously gdp expand year year first quarter now several think slowest since china started publishing quarterly figures likely risks firmly tilted downside past epidemics sure strong recovery virus eventually contained much uncertainty might three unknowns will dictate recovery timing long takes bring virus control long government relaxes heavy handed restrictions daily life long people resume whirl activity normally makes chinese economy vibrant economic policy presents challenge usually future peer greater uncertainty china officials can reasonably confident assuming growth will return pre virus trajectory next year next couple months black hole environment flexible measures help people companies difficult patch sensible can pared back rebound arrives getting right though easy worth noting china avoiding far least speculated officials will unleash big stimulus perhaps vast new array infrastructure projects get growth back speed soon government want

people building sites factories moment much efforts aimed keeping homes order prevent virus spreading moreover lag unveiling infrastructure plans breaking ground boost projects announced today start economy gathering steam leading overheating instead china using combination temporary cash support market interventions forbearance get crisis february rd central bank made headlines injecting trn yuan bn financial system bought treasury bonds banks promised buy back within days banks will probably suffer rising loan defaults coming weeks gives cash work repurchase agreements come due central bank choose effect extend needed officials meddling market say managing concern investors may pessimistic near term many companies pledged equity collateral loans need sell assets share prices fall adding downward pressure regulators already delayed reopening stockmarket new year holiday told brokerages bar clients short selling according reuters chinese shares still dropped february rd steepest one day fall since largely catching hong kong market open previous week february th shares rose little suggesting stabilisation tactics working finally officials advocating orchestrating forbearance various fronts shanghai due increase companies social security contributions april st city delayed three months saving firms estimated bn yuan beijing municipal government encouraged landlords cut commercial tenants rents exchange will provide subsidies regulators called banks throughout country roll loans companies small manufacturers otherwise lack cash buffers survive work stoppage even death toll continues mount officials already thinking economic distortions arisen course battle epidemic hospitals faced shortages protective equipment masks gowns gloves government called companies increase production many feeling sense duty heeded call liu shangxi adviser finance ministry noted means medical equipment firms will suffer severe overcapacity crisis passes government thus ready argues compensate proposals far cry bold growth investment plans hubei provincial leaders laid less month ago yet priority now stimulate economy climb technology ladder ensure society remains stable quarantines controls drag china grim new reality everything economic policy included revolves around question beat virus”

[2] “prepare worst contain global epidemic coronavirus health officials doctors can much stop spread diseases international jan th edition made china label country government prefer associated slick technology trending export present however ncov new coronavirus struck chinese city wuhan december virus now spread least countries economist went press world health organisation china confirmed almost infections deaths almost china new infectious disease begins spread decisions stop based patchy data change hour fog war phase says david heyman london school hygiene tropical medicine health officials make decisions quickly uncertain information says jeremy farrar wellcome trust charity must first determine deadliness new viruses first cases diagnosed usually among worst people ill enough go hospital zika mosquito borne virus commonly causes nothing mild flu like symptoms first recorded cases mostly mothers contracted infection pregnancy whose babies born brain damage result health officials start actively trying identify infected people milder cases added total result early estimates tend overstate danger new diseases happening now wuhan virus end january reported deaths represent confirmed infections around reported infected become severely ill suffering pneumonia respiratory failure modelling gabriel leung joseph wu university hong kong suggests january th number infections wuhan closer range infections will mild death rate virus low deadlier common flu america officials must gauge contagious new virus growing numbers arrive hospitals patterns emerge turns newly infected people health care workers relatives sick probably mean virus transmitted close rather casual contact

stemming spread easier experts must next determine passed person person common cold spreads virus laden droplets coughs sneezes travel metres influenza measles far contagious ride airborne particles sneeze can infect entire room yet clear wuhan virus transmitted thinks like one causes severe acute respiratory syndrome sars hitches ride droplets also unclear often infected show symptoms can spread virus others infections china germany seem result kind transmission contagious people unknowingly infecting others days sars middle east respiratory syndrome mers another lethal coronavirus superspreaders patients unusually high viral loads exceptionally infectious south korea patient mers infected people hour stay hospital emergency room threat global pandemic will put test preparedness plans countries big cities place feature worst case scenario usually arrival abroad hypothetical strain influenza deadly highly contagious rare combination features set apart spanish flu swept world killing people city authorities hospitals routinely carry drills test readiness scenario cases officials huddle conference room plan january th new york city top officials held drill wuhan virus exercises doctors health officials don protective gear get streets practise response test system new york city routinely uses mystery patients show hospitals pretending symptoms notifiable diseases doctors supposed report public health departments outbreak starts cross borders now happening wuhan virus knee jerk reaction set airport health checks passengers arriving outbreak hotspots many health experts think tests waste time money canada screening sars airports detected instances disease year sars killed people including canada theatrics airport checks suit politicians anxious seen responding worried citizens useful public provide arriving airports information explaining develop symptoms says agoritsa baka european centre disease prevention control efforts better spent boosting infection prevention measures hospitals says ms baka health workers often among first infected new virus pick patients pass families patients global sars outbreak third infected health workers best way nip outbreak bud disease detectives locate infected goal prevent passing virus others isolating hospitals home countries ask people quarantine voluntarily will demand court orders enforce rules people rebel outbreak grows clusters cases epidemic cities may go efforts keep people apart japan european countries close schools short periods flu seasons look though will particularly bad mexico city shut bars cinemas churches football stadiums days try stop spread swine flu locking large areas china currently entire province hubei population nearly untested modern times efforts can backfire one lesson ebola outbreaks west africa quarantine cared feel suffering enduring common good respected will try evade quarantine says jeremy konyndyk centre global development think tank washington makes matters worse outbreak crucial know infected going heavy handed attempt quarantine west point settlement people monrovia liberia capital ebola outbreak abandoned residents responded riots contrast similar well organised quarantine sierra leone traditional leaders brought board first meet resistance china drastic measures help delay epidemics wuhan virus countries months make huge difference says dr farrar hospitals europe america will better placed handle surge infections late spring compared february overwhelmed peak cases seasonal flu delay also crucial testing vaccine wuhan virus several already works china america australia dr farrar reckons vaccine ready clinical trials months clear epidemic controlled authorities go mitigation mode involves setting places care patients hospitals overflow systems identify patients treated first medical supplies run short countries put measures place advance china frantically building extra hospitals wuhan accommodate current surge numbers patients sometimes preparation can suffice will possible halt spread new

pandemic influenza virus waste public health resources capacity attempt admits britain flu pandemic preparedness strategy point officials left hoping best”

[3] “run don walk race produce vaccine latest coronavirus even scientists late outbreak work will wasted international feb th edition economist recent weeks searches google contagion movie soared film thriller virus spreads rapidly around world killing people plot follows frantic efforts scientists produce vaccine days first infection succeed real world recent vaccines taken years develop taken decade others vaccine stop hiv virus causes aids still elude scientists technological innovations streamlined development process dramatically shrink time takes produce vaccine new pathogen potential cause epidemic new coronavirus emerged chinese city wuhan december presents vaccine makers urgent test far killed almost people infected scientists china published wuhan virus genetic sequence january th less week isolated bug patient suffering mysterious respiratory infection late january several groups around world started work vaccine using genetic data first clinical tests humans safety begin early april luck vaccine ready within year next week world health organisation will convene global meeting set research agenda will agree rules protocols trials work medical advances priorities people rushed make new vaccines west african ebola outbreak tested world many ways particularly need speed delivery new treatments organisations institutions normally work slowly arm length came together get job done faster drug regulators america europe pharmaceutical firms charities experts worked closely advance trials technologies needed succeeded outbreak ebola democratic republic congo now appears wane contained largely result wide availability vaccine process scientific acceleration way time steroids says seth berkley boss gavi vaccine finance agency even vaccine ready within year late stem current epidemic china help countries fears growing wuhan virus will spread widely become established seasonal disease around world like common flu china extraordinary efforts contain virus including quarantining people may stave epidemics countries next winter soon tell deadly wuhan virus least bad seasonal flu vaccine risk will vital people hospitalised died america alone result influenza rush develop vaccine wuhan virus led coalition epidemic preparedness innovation cepi group set wake west african ebola outbreak cepi purpose forearm world future outbreaks disease without knowing diseases will aim vaccine previously unknown pathogen ready test humans within weeks identification end university research centres biotechnology firms funded working plug play vaccine design manufacturing technologies can used number pathogens allows genetic sequence particular pathogen slotted existing molecular platform forms basis vaccine past laboratory work vaccine required stocks actual virus treated make harmless still able tickle immune system producing antibodies proteins fight wild virus attacks working deadly virus tricky naturally requires special containment facilities exhaustive procedures prevent escaping infecting scientists gene sequencing made process quicker safer easier researchers can build synthetic versions parts viruses work vaccines without needing complete samples pathogens scientists produced vaccines viruses including zika ebola two coronaviruses sars severe acute respiratory syndrome mers middle east respiratory syndrome using technology vaccine research two cousins wuhan virus come handy recent weeks going viral vaccine developed laboratory sent factory turned sterile vaccine mix put vials tested ensure contaminated clinical trials humans can carried many tests done petri dishes process takes several months genetic sequencing can job much faster sequencing dna everything vial vaccine examining result scientists can spot traces viruses present vaccine research groups britain talks coun-

try medicines regulator approval process alternative testing methods development vaccine can speeded bottlenecks process eliminated says sarah gilbert leads group oxford university working vaccine wuhan virus group developed template vaccines can adapted quickly new pathogens researchers can make first small quantities new vaccine just six eight weeks past process taken year groups trying come vaccine wuhan virus using similar methods involving templates already proven work faster regulatory approval can also speed vaccines clinical trials even started making vaccine dr gilbert group began putting together application clinical trials group plans apply expedited ethical regulatory review can granted within days clinical trials ebola vaccine conducted britain normally process takes three months says dr gilbert even vaccine developed approved rapid rise cases wuhan virus china spread countries created new urgency planning ahead ways make massive quantities vaccine quickly many factories can mass produce vaccines new vaccines often wait long queue aware problem american government built dedicated manufacturing facilities can produce vaccines rapidly emergencies britain something similar cepi planning work involved thinking epidemics outbreaks limited one country pandemics global epidemics explains richard hatchett head group last week cepi put call vaccine candidates wuhan virus can manufactured large scale existing capacity february rd brought board partner gsk big drug firm agreed lend highly effective adjuvant new vaccine adjuvant special ingredient makes vaccines efficient boosting immune response means fewer doses vaccine lower concentration core ingredient needed vaccination even vaccine can produced sufficient quantities getting people need regardless live can still problem theory vaccine wuhan virus go risk health workers elderly conditions appear make virus lethal patients immune deficiencies problem politics often intervenes pandemic governments home vaccine making facilities can requisition use citing national defence security problem mr hatchett knows well worked white house medical preparedness flu pandemic outbreak low mortality rate exporting vaccine available american citizens quickly became vexed issue mr hatchett working try ensure wuhan virus vaccine made number different sites around world including ones small countries quickly able meet needs entire populations jab dark issues surrounding potential vaccine make questions medicines treat become gravely ill particularly acute licensed medicines treat coronaviruses currently exist experimental drugs development early data use one highlighted promising called remdesivir made gilead drug firm two randomised controlled trials will start enrolling patients mid february remdesivir developed treat ebola laboratory tests shown effective range viruses combination two drugs usually used hiv treatment also looks promising already tried patients says vasee moorthy helps set research development priorities epidemics randomised controlled trials people given drug tested given placebo gold standard scientific evidence will probably go ahead coming weeks clear drugs seem promising trials hospitalised patients will probably involve placebo arm everyone trial receive intensive care also given drug tested one yet knows whether new drugs may side effects harm good gravely ill patients may also allowed try untested drugs much preparation ever possible advance new disease drug vaccine efficacy can tested outbreak urgency behind search treatments wuhan virus understandable efforts effective case ebola people willing rush vaccines drugs use disease fatality rate around ebola calculus different one kills less infected hasty decisions lead products completely safe people faith vaccines damaged harm done world health rival worst feared wuhan virus”

≈≈≈ * ≈≈≈

Step 3: Created Word Cloud using the above corpus

Word Cloud



≈≈≈ * ≈≈≈

Step 4: Constructed Term-Document-Matrix to be used to create Comparison Cloud
(shown below) in Step 5

tdm = TermDocumentMatrix (terms: 1256, documents: 3)
Non-/sparse entries: 1564/2204
Sparsity : 58%

Term Document Matrix =

Terms	Docs		
	Warn	How	Race
according	1	0	0
act	1	0	0
activity	1	0	0
adding	1	0	0
adviser	1	0	0
advocating	1	0	0
ago	1	0	0
agreements	1	0	0
aimed	1	0	0
almost	1	2	1
already	3	1	2
also	1	2	4
america	1	3	2
analysts	1	0	0
announced	1	0	0
anxiety	1	0	0
anywhere	1	0	0
apart	1	2	0
april	1	0	1
argues	1	0	0
arisen	1	0	0
around	1	1	5
array	1	0	0
arrives	1	0	0
asked	1	0	0
assets	1	0	0
assuming	1	0	0
avoiding	1	0	0
back	3	0	0

bank	2	0	0
banks	3	0	0
bar	1	0	0
battle	1	0	0
beat	1	0	0
begun	1	0	0
beijing	1	0	0
big	2	1	1
black	1	0	0
bold	1	0	0
bonds	1	0	0
boost	2	0	0
bought	1	0	0
breaking	1	0	0
bring	1	0	0
brokerages	1	0	0
buffers	1	0	0
building	1	1	0
businesses	2	0	0
buy	1	0	0
call	1	0	1
called	2	0	1
can	2	5	17
cash	4	0	0
catching	1	0	0
causing	1	0	0
censored	1	0	0
central	2	0	0
certainly	1	0	0
chain	1	0	0
chains	1	0	0
challenge	1	0	0
chickens	1	0	0
china	11	8	4
chinese	2	1	1
choose	1	0	0
cities	1	2	0
city	1	5	1
clients	1	0	0
climb	1	0	0
collateral	1	0	0
combination	1	1	1
come	1	0	2
coming	1	0	1
commercial	1	0	0

companies	6	0	0
compensate	1	0	0
concern	1	0	0
confidence	1	0	0
confident	1	0	0
consensus	1	0	0
contained	1	0	1
continues	1	0	0
contributions	1	0	0
control	1	1	0
controls	1	0	0
coronavirus	2	3	2
country	1	1	2
couple	1	0	0
course	1	0	0
crisis	4	0	0
crumbled	1	0	0
cry	1	0	0
cut	2	0	0
daily	1	0	0
dare	1	0	0
day	1	0	0
days	2	2	2
deal	1	0	0
death	1	1	0
declared	1	0	0
defaults	1	0	0
delayed	2	0	0
dictate	1	0	0
difficult	1	0	0
dipping	1	0	0
distortions	1	0	0
downside	1	0	0
downward	1	0	0
drag	1	0	0
dropped	1	0	0
due	2	0	0
duty	1	0	0
easy	1	0	0
echoed	1	0	0
economic	7	0	0
economics	1	0	0
economy	3	0	0
edition	1	1	1
effect	1	0	0

efforts	1	3	3
encouraged	1	0	0
ensure	2	0	2
entire	1	2	1
environment	1	0	0
epidemic	2	3	3
epidemics	1	1	4
equipment	2	0	0
equity	1	0	0
especially	1	0	0
estimated	1	0	0
even	1	0	5
eventually	1	0	0
everything	1	0	1
exchange	1	0	0
expand	1	0	0
extend	1	0	0
faced	1	0	0
factories	2	0	1
fall	2	0	0
fallen	2	0	0
far	2	1	1
farmers	1	0	0
fast	1	0	0
feb	2	0	1
february	4	1	2
feed	1	0	0
feeling	1	0	0
fight	1	0	1
fights	1	0	0
figures	1	0	0
finally	2	0	0
finance	2	0	1
financial	1	0	0
firmly	1	0	0
firms	2	0	2
first	1	6	3
flashy	1	0	0
flexible	1	0	0
focus	1	0	0
forbearance	2	0	0
forecasts	1	0	0
founder	1	0	0
fronts	1	0	0
full	1	0	0

future	1	0	1
gathering	1	0	0
gdp	2	0	0
get	2	1	1
getting	1	0	1
gives	1	0	0
gloves	1	0	0
government	7	1	1
gowns	1	0	0
greater	1	0	0
grim	1	0	0
ground	1	0	0
grow	1	0	0
growth	5	0	0
guolong	1	0	0
handed	1	1	0
hard	1	0	0
headlines	1	0	0
heavy	1	1	0
heeded	1	0	0
help	1	1	1
hemming	1	0	0
high	1	1	0
hit	1	0	0
hole	1	0	0
holiday	2	0	0
homes	1	0	0
hong	1	1	0
hospitals	1	8	0
hotels	1	0	0
hubei	3	1	0
idle	1	0	0
ignore	1	0	0
illness	1	0	0
included	1	0	0
increase	2	0	0
infrastructure	2	0	0
injecting	1	0	0
injections	1	0	0
instead	2	0	0
intense	1	0	0
interventions	1	0	0
interview	1	0	0
investment	1	0	0
investors	1	0	0

january	2	3	2
jia	1	0	0
jobs	1	0	0
just	1	0	1
keeping	2	0	0
kong	1	1	0
lack	1	0	0
ladder	1	0	0
lag	1	0	0
laid	1	0	0
landlords	1	0	0
largely	1	0	1
later	1	0	0
leaders	2	1	0
leading	2	0	0
least	2	1	1
less	2	0	2
life	1	0	0
likely	1	0	0
link	1	0	0
little	1	0	0
liu	1	0	0
loan	1	0	0
loans	2	0	0
lockdown	1	0	0
long	4	0	1
longer	1	0	0
lose	1	0	0
lurch	1	0	0
made	2	1	3
mainland	1	0	0
make	1	2	6
makes	1	1	1
managing	1	0	0
manufacturers	1	0	0
many	3	1	3
market	3	0	0
masks	1	0	0
may	1	1	3
means	1	0	1
measures	1	3	0
meddling	1	0	0
media	1	0	0
medical	1	1	2
mention	1	0	0

might	2	0	0
ministry	1	0	0
moment	1	0	0
month	1	0	0
months	4	2	2
moreover	1	0	0
mount	1	0	0
much	2	1	2
municipal	1	0	0
mysterious	1	0	1
near	1	0	0
necessities	1	0	0
need	1	0	2
needed	1	0	2
new	5	9	11
next	2	1	2
normally	1	0	2
noted	1	0	0
noting	1	0	0
now	2	3	1
numbers	1	2	0
offices	1	0	0
officials	5	8	0
one	2	2	4
open	1	0	0
optimism	1	0	0
orchestrating	1	0	0
order	1	0	0
ordered	1	0	0
otherwise	1	0	0
outbreak	1	7	6
overcapacity	1	0	0
overheating	1	0	0
pared	1	0	0
passes	1	0	0
past	2	0	2
patch	1	0	0
peer	1	0	0
people	6	11	9
perhaps	1	0	0
persisted	1	0	0
pessimistic	1	0	0
placed	1	1	0
plans	3	1	1
pledged	1	0	0

pneumonia	1	1	0
policy	2	0	0
popular	1	0	0
poultry	1	0	0
pre	1	0	0
presents	1	0	1
pressure	2	0	0
prevent	1	1	1
previous	1	0	0
previously	1	0	1
prices	2	0	0
priority	1	0	0
probably	1	1	2
production	1	0	0
projects	2	0	0
promised	1	0	0
proposals	1	0	0
prospects	1	0	0
protective	1	1	0
provide	1	1	0
province	3	1	0
provinces	1	0	0
provincial	1	0	0
publicly	1	0	0
publishing	1	0	0
quarantine	1	5	0
quarantines	1	0	0
quarter	1	0	0
quarterly	1	0	0
question	1	0	0
rallied	1	0	0
rarely	1	0	0
reached	1	0	0
ready	1	1	3
reality	1	0	0
reasonably	1	0	0
rebound	1	0	0
recent	1	0	3
recovery	2	0	0
regulators	2	0	1
relaxes	1	0	0
remain	1	0	0
remains	1	0	0
rendering	1	0	0
rents	1	0	0

reopen	1	0	0
reopening	1	0	0
repurchase	1	0	0
reserves	1	0	0
restaurant	1	0	0
restaurants	1	0	0
restrictions	1	0	0
resume	1	0	0
return	1	0	0
reuters	1	0	0
revolves	1	0	0
right	1	0	0
rising	1	0	0
risks	1	0	0
roadblocks	1	0	0
roll	1	0	0
rose	1	0	0
running	1	0	0
rushed	1	0	1
said	1	0	0
saving	1	0	0
say	1	0	0
scale	2	0	1
security	1	0	1
sell	1	0	0
selling	1	0	0
sense	1	0	0
sensible	1	0	0
several	1	1	2
severe	1	1	1
shanghai	2	0	0
shangxi	1	0	0
share	2	0	0
shared	1	0	0
shares	2	0	0
shifted	1	0	0
short	1	2	0
shortages	1	0	0
shut	2	1	0
since	3	0	0
sites	1	0	1
slowest	1	0	0
small	1	0	2
snarled	1	0	0
social	3	0	0

society	1	0	0
soon	1	0	1
speculated	1	0	0
speed	1	0	2
spreading	2	0	0
stabilisation	1	0	0
stability	1	0	0
stable	1	0	0
start	1	1	1
started	2	0	2
starve	1	0	0
stayed	1	0	0
steam	1	0	0
steepest	1	0	0
steps	1	0	0
still	1	0	3
stimulate	1	0	0
stimulus	1	0	0
stockmarket	2	0	0
stoppage	1	0	0
stopping	1	0	0
strong	1	0	0
stronger	1	0	0
subsidies	1	0	0
suffer	2	0	0
suggesting	1	0	0
supplied	1	0	0
supplies	1	1	0
supply	1	0	0
support	2	0	0
supposed	1	1	0
sure	1	0	0
survive	1	0	0
swiftly	1	0	0
system	1	1	1
tactics	1	0	0
takes	1	0	3
targets	1	0	0
tech	1	0	0
technology	1	1	1
temporary	1	0	0
tenants	1	0	0
term	1	0	0
think	1	2	0
thinking	1	0	1

though	1	1	0
three	2	0	1
throughout	2	0	0
thus	1	0	0
tilted	1	0	0
timing	1	0	0
today	1	0	0
told	1	0	0
toll	1	0	0
towns	1	0	0
trade	1	0	0
trajectory	1	0	0
treasury	1	0	0
trn	1	0	0
two	2	0	4
uncertainty	2	0	0
unknowns	1	0	0
unleash	1	0	0
unreachable	1	0	0
unveiling	1	0	0
upbeat	1	0	0
using	1	0	3
usually	1	2	1
various	1	0	0
vast	2	0	0
venture	1	0	0
vibrant	1	0	0
viral	1	1	1
virus	6	16	20
vowed	1	0	0
want	1	0	0
war	1	1	0
warn	1	0	0
warned	1	0	0
watch	1	0	0
week	1	0	3
weeks	3	0	5
whirl	1	0	0
widely	1	0	1
will	6	7	9
within	1	0	4
work	2	0	8
working	1	0	4
worth	1	0	0
wouldn	1	0	0

xibei	1	0	0
year	8	1	3
yet	1	1	1
yuan	2	0	0
abandoned	0	1	0
abroad	0	1	0
accommodate	0	1	0
actively	0	1	0
acute	0	1	2
added	0	1	0
admits	0	1	0
advance	0	1	2
africa	0	1	0
agoritsa	0	1	0
airborne	0	1	0
airport	0	2	0
airports	0	2	0
among	0	2	0
another	0	1	0
anxious	0	1	0
areas	0	1	0
arrival	0	1	0
arrive	0	1	0
arriving	0	2	0
ask	0	1	0
associated	0	1	0
attempt	0	2	0
australia	0	1	0
authorities	0	2	0
babies	0	1	0
backfire	0	1	0
bad	0	1	1
baka	0	2	0
bars	0	1	0
based	0	1	0
become	0	1	2
begins	0	1	0
best	0	2	0
better	0	2	0
board	0	1	1
boosting	0	1	1
borders	0	1	0
born	0	1	0
borne	0	1	0
brain	0	1	0

britain	0	1	3
brought	0	1	1
bud	0	1	0
canada	0	2	0
capacity	0	1	1
capital	0	1	0
care	0	2	1
cared	0	1	0
carry	0	1	0
case	0	1	1
cases	0	6	1
casual	0	1	0
causes	0	2	1
centre	0	2	0
change	0	1	0
charity	0	1	0
checks	0	2	0
churches	0	1	0
cinemas	0	1	0
citizens	0	1	1
clear	0	2	1
clinical	0	1	5
close	0	2	0
closer	0	1	0
clusters	0	1	0
cold	0	1	0
common	0	3	1
commonly	0	1	0
compared	0	1	0
conference	0	1	0
confirmed	0	2	0
contact	0	1	0
contagious	0	4	0
contain	0	1	1
contracted	0	1	0
contrast	0	1	0
controlled	0	1	2
coughs	0	1	0
countries	0	6	4
court	0	1	0
cross	0	1	0
crucial	0	2	0
current	0	1	1
currently	0	1	1
damage	0	1	0

danger	0	1	0
data	0	1	2
david	0	1	0
deadlier	0	1	0
deadliness	0	1	0
deadly	0	1	2
deaths	0	2	0
december	0	1	1
decisions	0	2	1
delay	0	2	0
demand	0	1	0
departments	0	1	0
detected	0	1	0
detectives	0	1	0
determine	0	2	0
develop	0	1	2
development	0	1	4
diagnosed	0	1	0
difference	0	1	0
disease	0	4	4
diseases	0	3	1
doctors	0	3	0
don	0	1	1
drastic	0	1	0
drill	0	1	0
drills	0	1	0
droplets	0	2	0
early	0	1	2
easier	0	1	1
east	0	1	1
ebola	0	2	8
economist	0	1	1
emerge	0	1	0
emergency	0	1	0
end	0	1	1
enduring	0	1	0
enforce	0	1	0
enough	0	1	0
estimates	0	1	0
europe	0	1	1
european	0	2	0
evade	0	1	0
exceptionally	0	1	0
exercises	0	1	0
experts	0	2	1

explaining	0	1	0
export	0	1	0
extra	0	1	0
failure	0	1	0
families	0	1	0
farrar	0	3	0
feature	0	1	0
features	0	1	0
feel	0	1	0
flu	0	7	3
fog	0	1	0
football	0	1	0
frantically	0	1	0
gabriel	0	1	0
gauge	0	1	0
gear	0	1	0
germany	0	1	0
global	0	4	2
goal	0	1	0
going	0	1	1
good	0	1	1
growing	0	1	1
grows	0	1	0
halt	0	1	0
handle	0	1	0
happening	0	2	0
health	0	12	3
held	0	1	0
heymann	0	1	0
highly	0	1	1
hitches	0	1	0
home	0	1	1
hoping	0	1	0
hospital	0	2	0
hotspots	0	1	0
hour	0	2	0
however	0	1	0
huddle	0	1	0
huge	0	1	0
hygiene	0	1	0
hypothetical	0	1	0
identify	0	2	0
ill	0	2	2
including	0	1	3
infect	0	1	0

infected	0	9	2
infecting	0	1	1
infection	0	2	2
infections	0	6	0
infectious	0	2	0
influenza	0	3	1
information	0	2	0
instances	0	1	0
international	0	1	1
involves	0	1	0
isolating	0	1	0
jan	0	1	0
japan	0	1	0
jeremy	0	2	0
jerk	0	1	0
joseph	0	1	0
keep	0	1	0
killed	0	1	1
killling	0	1	1
kind	0	1	0
knee	0	1	0
know	0	1	0
konyndyk	0	1	0
korea	0	1	0
label	0	1	0
laden	0	1	0
large	0	1	1
late	0	1	3
left	0	1	0
leone	0	1	0
lesson	0	1	0
lethal	0	1	1
leung	0	1	0
liberia	0	1	0
like	0	2	1
loads	0	1	0
locate	0	1	0
locking	0	1	0
london	0	1	0
look	0	1	0
low	0	1	1
matters	0	1	0
mean	0	1	0
measles	0	1	0
medicine	0	1	0

meet	0	1	1
mers	0	2	1
metres	0	1	0
mexico	0	1	0
middle	0	1	1
mild	0	2	0
milder	0	1	0
mitigation	0	1	0
mode	0	1	0
modelling	0	1	0
modern	0	1	0
money	0	1	0
monrovia	0	1	0
mosquito	0	1	0
mostly	0	1	0
mothers	0	1	0
must	0	3	0
mystery	0	1	0
ncov	0	1	0
nearly	0	1	0
newly	0	1	0
nip	0	1	0
nothing	0	1	0
notifiable	0	1	0
number	0	1	2
often	0	2	2
orders	0	1	0
organisation	0	1	1
organised	0	1	0
others	0	3	1
outbreaks	0	1	2
overflow	0	1	0
overstate	0	1	0
overwhelmed	0	1	0
pandemic	0	3	2
particles	0	1	0
particularly	0	1	2
pass	0	1	0
passed	0	1	0
passengers	0	1	0
passing	0	1	0
patchy	0	1	0
patient	0	1	1
patients	0	7	5
patterns	0	1	0

peak	0	1	0
periods	0	1	0
person	0	2	0
phase	0	1	0
pick	0	1	0
place	0	2	0
places	0	1	0
plan	0	1	0
point	0	2	0
politicians	0	1	0
population	0	1	0
possible	0	1	1
practise	0	1	0
prefer	0	1	0
pregnancy	0	1	0
preparation	0	1	1
prepare	0	1	0
preparedness	0	2	2
present	0	1	1
press	0	1	0
pretending	0	1	0
prevention	0	2	0
public	0	3	0
put	0	2	2
quickly	0	1	4
range	0	1	1
rare	0	1	0
rate	0	1	2
rather	0	1	0
reaction	0	1	0
readiness	0	1	0
rebel	0	1	0
reckons	0	1	0
recorded	0	1	0
relatives	0	1	0
report	0	1	0
reported	0	2	0
represent	0	1	0
residents	0	1	0
resistance	0	1	0
resources	0	1	0
respected	0	1	0
respiratory	0	3	3
responded	0	1	0
responding	0	1	0

response	0	1	1
result	0	3	3
ride	0	2	0
riots	0	1	0
room	0	3	0
routinely	0	2	0
rules	0	1	1
run	0	1	1
sars	0	5	1
says	0	6	4
scenario	0	2	0
school	0	1	0
schools	0	1	0
screening	0	1	0
seasonal	0	1	2
seasons	0	1	0
seem	0	1	1
seen	0	1	0
set	0	2	3
setting	0	1	0
settlement	0	1	0
severely	0	1	0
show	0	2	0
sick	0	1	0
sierra	0	1	0
similar	0	1	2
slick	0	1	0
sneeze	0	1	0
sneezes	0	1	0
sometimes	0	1	0
south	0	1	0
spanish	0	1	0
spent	0	1	0
spread	0	7	2
spreads	0	1	1
spring	0	1	0
stadiums	0	1	0
starts	0	1	0
stay	0	1	0
stemming	0	1	0
stop	0	3	1
strain	0	1	0
strategy	0	1	0
streets	0	1	0
struck	0	1	0

suffering	0	2	1
suffice	0	1	0
suggests	0	1	0
suit	0	1	0
superspreaders	0	1	0
surge	0	2	0
swept	0	1	0
swine	0	1	0
symptoms	0	4	0
syndrome	0	2	2
systems	0	1	0
tank	0	1	0
tend	0	1	0
test	0	3	2
testing	0	1	1
tests	0	1	3
theatrics	0	1	0
thinks	0	1	0
third	0	1	0
threat	0	1	0
time	0	1	2
times	0	1	0
top	0	1	0
total	0	1	0
traditional	0	1	0
transmission	0	1	0
transmitted	0	2	0
travel	0	1	0
treated	0	1	1
trending	0	1	0
trials	0	1	9
tropical	0	1	0
trust	0	1	0
try	0	2	2
trying	0	1	1
turns	0	1	0
uncertain	0	1	0
unclear	0	1	0
university	0	1	2
unknowingly	0	1	0
untested	0	1	1
unusually	0	1	0
useful	0	1	0
uses	0	1	0
vaccine	0	2	39

viruses	0	1	4
voluntarily	0	1	0
washington	0	1	0
waste	0	2	0
way	0	1	1
well	0	1	1
wellcome	0	1	0
went	0	1	0
west	0	2	2
whose	0	1	0
workers	0	3	1
works	0	1	0
world	0	2	9
worried	0	1	0
worse	0	1	0
worst	0	3	1
wuhan	0	9	14
york	0	2	0
zika	0	1	1
able	0	0	2
acceleration	0	0	1
actual	0	0	1
adapted	0	0	1
adjuvant	0	0	2
advances	0	0	1
african	0	0	2
agency	0	0	1
agenda	0	0	1
agree	0	0	1
agreed	0	0	1
ahead	0	0	2
aids	0	0	1
aim	0	0	1
allowed	0	0	1
allows	0	0	1
alone	0	0	1
alternative	0	0	1
american	0	0	2
antibodies	0	0	1
appear	0	0	1
appears	0	0	1
application	0	0	1
apply	0	0	1
approval	0	0	2
approved	0	0	1

arm	0	0	2
attacks	0	0	1
availability	0	0	1
available	0	0	1
aware	0	0	1
basis	0	0	1
became	0	0	1
began	0	0	1
begin	0	0	1
behind	0	0	1
berkeley	0	0	1
biotechnology	0	0	1
boss	0	0	1
bottlenecks	0	0	1
bug	0	0	1
build	0	0	1
built	0	0	1
calculus	0	0	1
came	0	0	1
candidates	0	0	1
carried	0	0	1
cause	0	0	1
centres	0	0	1
cepi	0	0	4
charities	0	0	1
citing	0	0	1
closely	0	0	1
coalition	0	0	1
complete	0	0	1
completely	0	0	1
concentration	0	0	1
conditions	0	0	1
conducted	0	0	1
congo	0	0	1
contagion	0	0	1
containment	0	0	1
contaminated	0	0	1
convene	0	0	1
core	0	0	1
coronaviruses	0	0	2
cousins	0	0	1
created	0	0	1
damaged	0	0	1
dark	0	0	1
decade	0	0	1

dedicated	0	0	1
defence	0	0	1
deficiencies	0	0	1
delivery	0	0	1
democratic	0	0	1
design	0	0	1
developed	0	0	4
died	0	0	1
different	0	0	2
dishes	0	0	1
dna	0	0	1
done	0	0	3
doses	0	0	1
dramatically	0	0	1
drug	0	0	6
drugs	0	0	6
effective	0	0	3
effects	0	0	1
efficacy	0	0	1
efficient	0	0	1
eight	0	0	1
elderly	0	0	1
eliminated	0	0	1
elude	0	0	1
emerged	0	0	1
emergencies	0	0	1
enrolling	0	0	1
escaping	0	0	1
established	0	0	1
ethical	0	0	1
ever	0	0	1
everyone	0	0	1
evidence	0	0	1
examining	0	0	1
exhaustive	0	0	1
exist	0	0	1
existing	0	0	2
expedited	0	0	1
experimental	0	0	1
explains	0	0	1
exporting	0	0	1
extraordinary	0	0	1
facilities	0	0	3
factory	0	0	1
faith	0	0	1

faster	0	0	3
fatality	0	0	1
feared	0	0	1
fears	0	0	1
fewer	0	0	1
film	0	0	1
firm	0	0	2
follows	0	0	1
forearm	0	0	1
forms	0	0	1
frantic	0	0	1
funded	0	0	1
gavi	0	0	1
gene	0	0	1
genetic	0	0	4
gilbert	0	0	3
gilead	0	0	1
given	0	0	3
gold	0	0	1
google	0	0	1
governments	0	0	1
granted	0	0	1
gravely	0	0	2
group	0	0	6
groups	0	0	3
gsk	0	0	1
handy	0	0	1
harm	0	0	2
harmless	0	0	1
hasty	0	0	1
hatchett	0	0	3
head	0	0	1
helps	0	0	1
highlighted	0	0	1
hiv	0	0	2
hospitalised	0	0	2
house	0	0	1
humans	0	0	3
identification	0	0	1
immune	0	0	3
ingredient	0	0	2
innovation	0	0	1
innovations	0	0	1
institutions	0	0	1
intensive	0	0	1

intervenes	0	0	1
involve	0	0	1
involved	0	0	1
involving	0	0	1
isolated	0	0	1
issue	0	0	1
issues	0	0	1
jab	0	0	1
job	0	0	2
kills	0	0	1
knowing	0	0	1
knows	0	0	2
laboratory	0	0	3
last	0	0	1
latest	0	0	1
lead	0	0	1
leads	0	0	1
led	0	0	1
lend	0	0	1
length	0	0	1
licensed	0	0	1
limited	0	0	1
live	0	0	1
looks	0	0	1
lower	0	0	1
luck	0	0	1
makers	0	0	1
making	0	0	2
manufactured	0	0	1
manufacturing	0	0	2
mass	0	0	1
massive	0	0	1
medicines	0	0	3
meeting	0	0	1
methods	0	0	2
mid	0	0	1
mix	0	0	1
molecular	0	0	1
moorthy	0	0	1
mortality	0	0	1
movie	0	0	1
national	0	0	1
naturally	0	0	1
needing	0	0	1
needs	0	0	1

ones	0	0	1
organisations	0	0	1
oxford	0	0	1
pandemics	0	0	1
particular	0	0	1
partner	0	0	1
parts	0	0	1
pathogen	0	0	3
pathogens	0	0	3
petri	0	0	1
pharmaceutical	0	0	1
placebo	0	0	2
planning	0	0	2
platform	0	0	1
play	0	0	1
plot	0	0	1
plug	0	0	1
politics	0	0	1
populations	0	0	1
potential	0	0	2
priorities	0	0	2
problem	0	0	4
procedures	0	0	1
process	0	0	8
produce	0	0	5
produced	0	0	2
producing	0	0	1
products	0	0	1
promising	0	0	3
proteins	0	0	1
protocols	0	0	1
proven	0	0	1
published	0	0	1
purpose	0	0	1
putting	0	0	1
quantities	0	0	3
quarantining	0	0	1
questions	0	0	1
queue	0	0	1
quicker	0	0	1
race	0	0	1
randomised	0	0	2
rapid	0	0	1
rapidly	0	0	2
real	0	0	1

receive	0	0	1
regardless	0	0	1
regulator	0	0	1
regulatory	0	0	2
remdesivir	0	0	2
republic	0	0	1
required	0	0	1
requires	0	0	1
requisition	0	0	1
research	0	0	5
researchers	0	0	2
review	0	0	1
richard	0	0	1
rise	0	0	1
risk	0	0	2
rival	0	0	1
rush	0	0	2
safe	0	0	1
safer	0	0	1
safety	0	0	1
samples	0	0	1
sarah	0	0	1
scientific	0	0	2
scientists	0	0	7
search	0	0	1
searches	0	0	1
sent	0	0	1
sequence	0	0	2
sequencing	0	0	3
seth	0	0	1
shown	0	0	1
shrink	0	0	1
side	0	0	1
six	0	0	1
slotted	0	0	1
slowly	0	0	1
soared	0	0	1
something	0	0	1
special	0	0	2
speeded	0	0	1
spot	0	0	1
standard	0	0	1
stave	0	0	1
stem	0	0	1
sterile	0	0	1

steroids	0	0	1
stocks	0	0	1
streamlined	0	0	1
succeed	0	0	1
succeeded	0	0	1
sufficient	0	0	1
surrounding	0	0	1
synthetic	0	0	1
taken	0	0	3
talks	0	0	1
technological	0	0	1
technologies	0	0	2
tell	0	0	1
template	0	0	1
templates	0	0	1
tested	0	0	5
theory	0	0	1
thriller	0	0	1
tickle	0	0	1
together	0	0	2
traces	0	0	1
treat	0	0	3
treatment	0	0	1
treatments	0	0	2
trial	0	0	1
tricky	0	0	1
tried	0	0	1
turned	0	0	1
understandable	0	0	1
unknown	0	0	1
urgency	0	0	2
urgent	0	0	1
use	0	0	3
used	0	0	2
vaccination	0	0	1
vaccines	0	0	12
vasee	0	0	1
versions	0	0	1
vexed	0	0	1
vial	0	0	1
vials	0	0	1
vital	0	0	1
wait	0	0	1
wake	0	0	1
walk	0	0	1

wane	0	0	1
wasted	0	0	1
ways	0	0	2
whether	0	0	1
white	0	0	1
wide	0	0	1
wild	0	0	1
willing	0	0	1
winter	0	0	1
without	0	0	2
worked	0	0	2
years	0	0	1

≈≈≈ * ≈≈≈

Step 5: Created Comparison Cloud (shown below) using the above Term Document Matrix

Step 6: Did Sentiment Analysis

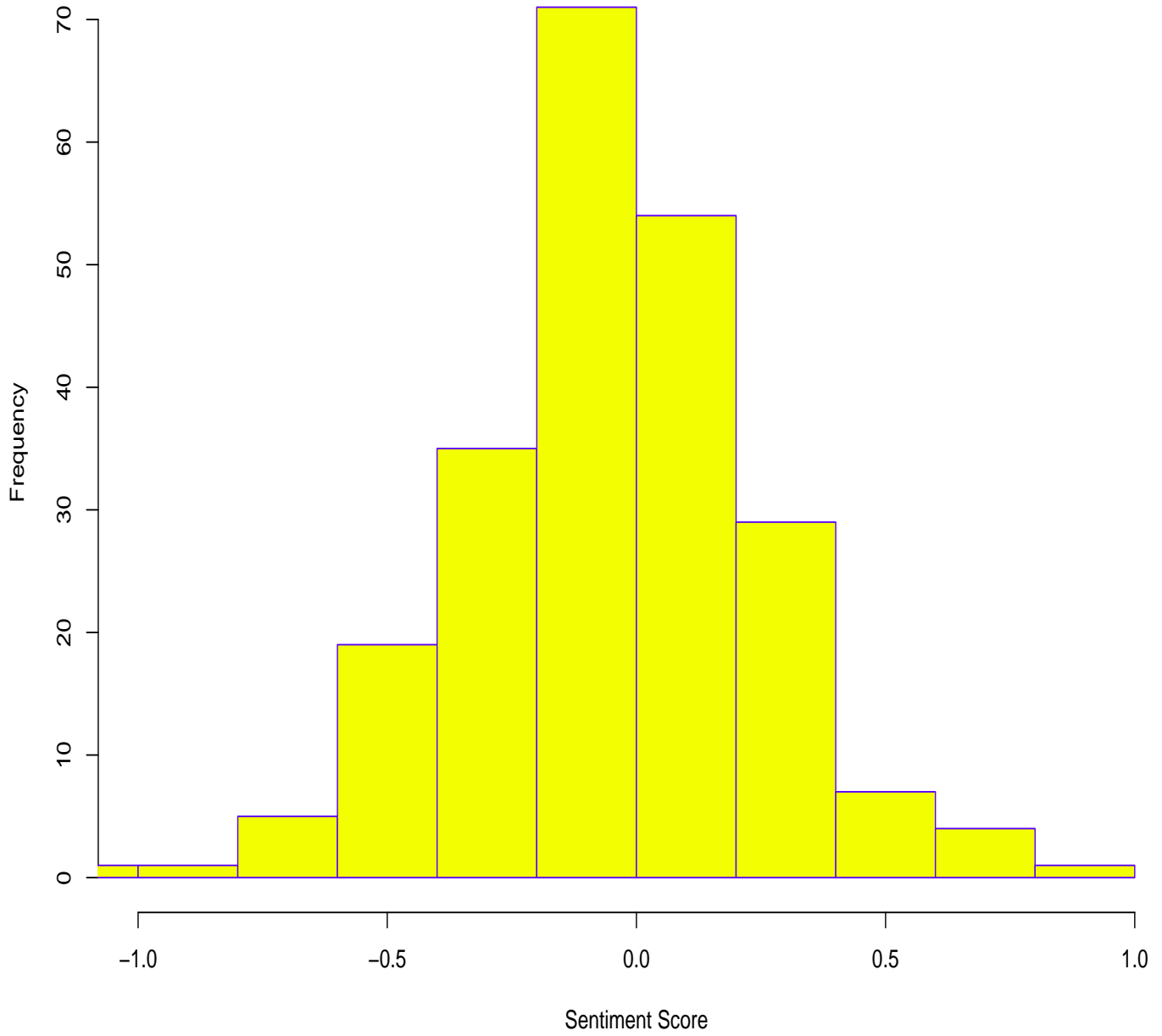
Sentiment1 =

	element_id	sentence_id	word_count	sentiment
1:	1	1	12	-0.5773503
2:	1	2	11	0.1507557
3:	1	3	17	0.1940285
4:	1	4	15	0.1290994
5:	1	5	24	-0.1837117
—				
225:	3	89	9	0.2666667
226:	3	90	21	-0.1527525
227:	3	91	13	-0.4160251
228:	3	92	18	-0.0942809
229:	3	93	18	-0.4714045

Sentiment2 =

	element_id	word_count	sd	ave_sentiment
1:	1	1141	0.2606340	-0.003142519
2:	2	1451	0.3418249	-0.161881214
3:	3	1875	0.3241119	0.006319311

Histogram of Sentiment Score



≈≈≈ * ≈≈≈

Finally, **Step 7**: Implemented LDA for Topic Modeling

- **Rather advanced** relative to aspects of text mining and analysis we looked at thus far!
- Useful for efficient analysis of large volumes of text.
- Most celebrated topic modeling technique: Latent Dirichlet Allocation (LDA).

It is a Bayesian inference model that

- (i) allocates each document's words to few topics;
- (ii) in each topic, assigns high probability to few terms.

That is, LDA associates each document with a probability distribution over topics, where topics are probability distributions over words, as shown below

Construct document-term-matrix:

dtm = DocumentTermMatrix (documents: 3, terms: 1256)
Non-/sparse entries: 1564/2204
Sparsity: 58%

Top 15 terms for each of the 3 topics

ldaOut = A LDA_Gibbs topic model with 3 topics.

terms.ldaOut =

	Topic 1	Topic 2	Topic 3
[1,]	"people"	"china"	"vaccine"
[2,]	"new"	"year"	"virus"
[3,]	"health"	"will"	"can"
[4,]	"officials"	"government"	"wuhan"
[5,]	"outbreak"	"months"	"will"
[6,]	"patients"	"economic"	"vaccines"
[7,]	"infected"	"companies"	"world"
[8,]	"first"	"february"	"ebola"
[9,]	"flu"	"growth"	"trials"

[10,]	“hospitals”	“long”	“make”
[11,]	“spread”	“plans”	“new”
[12,]	“one”	“already”	“weeks”
[13,]	“countries”	“cash”	“process”
[14,]	“disease”	“crisis”	“also”
[15,]	“says”	“days”	“around”

- `three_docs_lda` = A LDA_VEM topic model with 3 topics.
- LDA assigned topics the following probability distributions over words:

```
three_docs_topics =
# A tibble: 3,768 x 3
  topic term      beta
  <dbl> <dbl> <dbl>
1     1 according 4.11e-29
2     2 according 2.89e-35
3     3 according 1.54e- 3
4     1 act       3.54e-29
5     2 act       1.25e-35
6     3 act       1.54e- 3
7     1 activity 4.56e-29
8     2 activity 1.29e-35
9     3 activity 1.54e- 3
10    1 adding  3.90e-29
# ... with 3,758 more rows
```

⇒

Topic 1:

$4.11 \times 10^{-27}\%$ according

$3.54 \times 10^{-27}\%$ act

$4.56 \times 10^{-27}\%$ activity

⋮

Topic 2:

$2.89 \times 10^{-33}\%$ according

$1.25 \times 10^{-33}\%$ act

$1.29 \times 10^{-33}\%$ activity

⋮

Topic 3:

$1.54 \times 10^{-1}\%$ according

$1.54 \times 10^{-1}\%$ act

$1.54 \times 10^{-1}\%$ activity

⋮

⇒ The term “according” has

$4.11 \times 10^{-27}\%$ probability of being generated from Topic 1,

$2.89 \times 10^{-33}\%$ probability of being generated from Topic 2,
 $1.54 \times 10^{-1}\%$ probability of being generated from Topic 3,

⋮

- LDA associated each document with following probability distributions over topics:

three.docs.lda.gamma =

A tibble: 9 x 3

	document	topic	gamma
1	1	1	0.0000598
2	2	1	1.00
3	3	1	0.0000377
4	1	2	0.0000598
5	2	2	0.0000493
6	3	2	1.00
7	1	3	1.00
8	2	3	0.0000493
9	3	3	0.0000377

⇒

Document 1:

0.00598% topic 1

0.00598% topic 2

100% topic 3

Document 2:

100% topic 1

0.00493% topic 2

0.00493% topic 3

Document 3:

0.00377% topic 1

100% topic 2

0.00377% topic 3

≈≈≈ * ≈≈≈

In next analysis, I will attempt alternative implementation of LDA using Mallet as in the paper entitled “Applying topic modeling to identify the multifactorial attributes of drug-induced liver injury” by Dr. Bowman, Dr. George, and other co-authors.

≈≈≈ * ≈≈≈

Thank You!