|  |
| --- |
| > library(NLP)  > library(tm)  > library(RColorBrewer)  > library(wordcloud)  > data(crude) # 20 new articles from Reuter- 21578 data set  > help(crude) # To know abour crude data  > # information about the first file in crude data  > # crude[[1]]: first article  > # Data structure  > str(crude[[1]]) #summary structure, information of the document  List of 2  $ content: chr "Diamond Shamrock Corp said that\neffective today it had cut its contract prices for crude oil by\n1.50 dlrs a b"| \_\_truncated\_\_  $ meta :List of 15  ..$ author : chr(0)  ..$ datetimestamp: POSIXlt[1:1], format: "1987-02-26 17:00:56"  ..$ description : chr ""  ..$ heading : chr "DIAMOND SHAMROCK (DIA) CUTS CRUDE PRICES"  ..$ id : chr "127"  ..$ language : chr "en"  ..$ origin : chr "Reuters-21578 XML"  ..$ topics : chr "YES"  ..$ lewissplit : chr "TRAIN"  ..$ cgisplit : chr "TRAINING-SET"  ..$ oldid : chr "5670"  ..$ places : chr "usa"  ..$ people : chr(0)  ..$ orgs : chr(0)  ..$ exchanges : chr(0)  ..- attr(\*, "class")= chr "TextDocumentMeta"  - attr(\*, "class")= chr [1:2] "PlainTextDocument" "TextDocument"  > content(crude[[1]]) #content of the first article.content(crude[[n]]) to see other articles  [1] "Diamond Shamrock Corp said that\neffective today it had cut its contract prices for crude oil by\n1.50 dlrs a barrel.\n The reduction brings its posted price for West Texas\nIntermediate to 16.00 dlrs a barrel, the copany said.\n \"The price reduction today was made in the light of falling\noil product prices and a weak crude oil market,\" a company\nspokeswoman said.\n Diamond is the latest in a line of U.S. oil companies that\nhave cut its contract, or posted, prices over the last two days\nciting weak oil markets.\n Reuter"  > content(crude[[2]]) # The second article  [1] "OPEC may be forced to meet before a\nscheduled June session to readdress its production cutting\nagreement if the organization wants to halt the current slide\nin oil prices, oil industry analysts said.\n \"The movement to higher oil prices was never to be as easy\nas OPEC thought. They may need an emergency meeting to sort out\nthe problems,\" said Daniel Yergin, director of Cambridge Energy\nResearch Associates, CERA.\n Analysts and oil industry sources said the problem OPEC\nfaces is excess oil supply in world oil markets.\n \"OPEC's problem is not a price problem but a production\nissue and must be addressed in that way,\" said Paul Mlotok, oil\nanalyst with Salomon Brothers Inc.\n He said the market's earlier optimism about OPEC and its\nability to keep production under control have given way to a\npessimistic outlook that the organization must address soon if\nit wishes to regain the initiative in oil prices.\n But some other analysts were uncertain that even an\nemergency meeting would address the problem of OPEC production\nabove the 15.8 mln bpd quota set last December.\n \"OPEC has to learn that in a buyers market you cannot have\ndeemed quotas, fixed prices and set differentials,\" said the\nregional manager for one of the major oil companies who spoke\non condition that he not be named. \"The market is now trying to\nteach them that lesson again,\" he added.\n David T. Mizrahi, editor of Mideast reports, expects OPEC\nto meet before June, although not immediately. However, he is\nnot optimistic that OPEC can address its principal problems.\n \"They will not meet now as they try to take advantage of the\nwinter demand to sell their oil, but in late March and April\nwhen demand slackens,\" Mizrahi said.\n But Mizrahi said that OPEC is unlikely to do anything more\nthan reiterate its agreement to keep output at 15.8 mln bpd.\"\n Analysts said that the next two months will be critical for\nOPEC's ability to hold together prices and output.\n \"OPEC must hold to its pact for the next six to eight weeks\nsince buyers will come back into the market then,\" said Dillard\nSpriggs of Petroleum Analysis Ltd in New York.\n But Bijan Moussavar-Rahmani of Harvard University's Energy\nand Environment Policy Center said that the demand for OPEC oil\nhas been rising through the first quarter and this may have\nprompted excesses in its production.\n \"Demand for their (OPEC) oil is clearly above 15.8 mln bpd\nand is probably closer to 17 mln bpd or higher now so what we\nare seeing characterized as cheating is OPEC meeting this\ndemand through current production,\" he told Reuters in a\ntelephone interview.\n Reuter"  > meta(crude[[1]]) #information of meta. The type of elements  author : character(0)  datetimestamp: 1987-02-26 17:00:56  description :  heading : DIAMOND SHAMROCK (DIA) CUTS CRUDE PRICES  id : 127  language : en  origin : Reuters-21578 XML  topics : YES  lewissplit : TRAIN  cgisplit : TRAINING-SET  oldid : 5670  places : usa  people : character(0)  orgs : character(0)  exchanges : character(0)  > lapply(crude, content) #information of the file's content(file's name,content)  $`127`  [1] "Diamond Shamrock Corp said that\neffective today it had cut its contract prices for crude oil by\n1.50 dlrs a barrel.\n The reduction brings its posted price for West Texas\nIntermediate to 16.00 dlrs a barrel, the copany said.\n \"The price reduction today was made in the light of falling\noil product prices and a weak crude oil market,\" a company\nspokeswoman said.\n Diamond is the latest in a line of U.S. oil companies that\nhave cut its contract, or posted, prices over the last two days\nciting weak oil markets.\n Reuter"  $`144`  [1] "OPEC may be forced to meet before a\nscheduled June session to readdress its production cutting\nagreement if the organization wants to halt the current slide\nin oil prices, oil industry analysts said.\n \"The movement to higher oil prices was never to be as easy\nas OPEC thought. They may need an emergency meeting to sort out\nthe problems,\" said Daniel Yergin, director of Cambridge Energy\nResearch Associates, CERA.\n Analysts and oil industry sources said the problem OPEC\nfaces is excess oil supply in world oil markets.\n \"OPEC's problem is not a price problem but a production\nissue and must be addressed in that way,\" said Paul Mlotok, oil\nanalyst with Salomon Brothers Inc.\n He said the market's earlier optimism about OPEC and its\nability to keep production under control have given way to a\npessimistic outlook that the organization must address soon if\nit wishes to regain the initiative in oil prices.\n But some other analysts were uncertain that even an\nemergency meeting would address the problem of OPEC production\nabove the 15.8 mln bpd quota set last December.\n \"OPEC has to learn that in a buyers market you cannot have\ndeemed quotas, fixed prices and set differentials,\" said the\nregional manager for one of the major oil companies who spoke\non condition that he not be named. \"The market is now trying to\nteach them that lesson again,\" he added.\n David T. Mizrahi, editor of Mideast reports, expects OPEC\nto meet before June, although not immediately. However, he is\nnot optimistic that OPEC can address its principal problems.\n \"They will not meet now as they try to take advantage of the\nwinter demand to sell their oil, but in late March and April\nwhen demand slackens,\" Mizrahi said.\n But Mizrahi said that OPEC is unlikely to do anything more\nthan reiterate its agreement to keep output at 15.8 mln bpd.\"\n Analysts said that the next two months will be critical for\nOPEC's ability to hold together prices and output.\n \"OPEC must hold to its pact for the next six to eight weeks\nsince buyers will come back into the market then,\" said Dillard\nSpriggs of Petroleum Analysis Ltd in New York.\n But Bijan Moussavar-Rahmani of Harvard University's Energy\nand Environment Policy Center said that the demand for OPEC oil\nhas been rising through the first quarter and this may have\nprompted excesses in its production.\n \"Demand for their (OPEC) oil is clearly above 15.8 mln bpd\nand is probably closer to 17 mln bpd or higher now so what we\nare seeing characterized as cheating is OPEC meeting this\ndemand through current production,\" he told Reuters in a\ntelephone interview.\n Reuter"  $`191`  [1] "Texaco Canada said it lowered the\ncontract price it will pay for crude oil 64 Canadian cts a\nbarrel, effective today.\n The decrease brings the company's posted price for the\nbenchmark grade, Edmonton/Swann Hills Light Sweet, to 22.26\nCanadian dlrs a bbl.\n Texaco Canada last changed its crude oil postings on Feb\n19.\n Reuter"  $`194`  [1] "Marathon Petroleum Co said it reduced\nthe contract price it will pay for all grades of crude oil one\ndlr a barrel, effective today.\n The decrease brings Marathon's posted price for both West\nTexas Intermediate and West Texas Sour to 16.50 dlrs a bbl. The\nSouth Louisiana Sweet grade of crude was reduced to 16.85 dlrs\na bbl.\n The company last changed its crude postings on Jan 12.\n Reuter"  $`211`  [1] "Houston Oil Trust said that independent\npetroleum engineers completed an annual study that estimates\nthe trust's future net revenues from total proved reserves at\n88 mln dlrs and its discounted present value of the reserves at\n64 mln dlrs.\n Based on the estimate, the trust said there may be no money\navailable for cash distributions to unitholders for the\nremainder of the year.\n It said the estimates reflect a decrease of about 44 pct in\nnet reserve revenues and 39 pct in discounted present value\ncompared with the study made in 1985.\n Reuter"  $`236`  [1] "Kuwait\"s Oil Minister, in remarks\npublished today, said there were no plans for an emergency OPEC\nmeeting to review oil policies after recent weakness in world\noil prices.\n Sheikh Ali al-Khalifa al-Sabah was quoted by the local\ndaily al-Qabas as saying: \"None of the OPEC members has asked\nfor such a meeting.\"\n He denied Kuwait was pumping above its quota of 948,000\nbarrels of crude daily (bpd) set under self-imposed production\nlimits of the 13-nation organisation.\n Traders and analysts in international oil markets estimate\nOPEC is producing up to one mln bpd above a ceiling of 15.8 mln\nbpd agreed in Geneva last December.\n They named Kuwait and the United Arab Emirates, along with\nthe much smaller producer Ecuador, among those producing above\nquota. Kuwait, they said, was pumping 1.2 mln bpd.\n \"This rumour is baseless. It is based on reports which said\nKuwait has the ability to exceed its share. They suppose that\nbecause Kuwait has the ability, it will do so,\" the minister\nsaid.\n Sheikh Ali has said before that Kuwait had the ability to\nproduce up to 4.0 mln bpd.\n \"If we can sell more than our quota at official prices,\nwhile some countries are suffering difficulties marketing their\nshare, it means we in Kuwait are unusually clever,\" he said.\n He was referring apparently to the Gulf state of qatar,\nwhich industry sources said was selling less than 180,000 bpd\nof its 285,000 bpd quota, because buyers were resisting\nofficial prices restored by OPEC last month pegged to a marker\nof 18 dlrs per barrel.\n Prices in New York last week dropped to their lowest levels\nthis year and almost three dollars below a three-month high of\n19 dollars a barrel.\n Sheikh Ali also delivered \"a challenge to any international\noil company that declared Kuwait sold below official prices.\"\n Because it was charging its official price, of 16.67 dlrs a\nbarrel, it had lost custom, he said but did not elaborate.\n However, Kuwait had guaranteed markets for its oil because\nof its local and international refining facilities and its own\ndistribution network abroad, he added.\n He reaffirmed that the planned meeting March 7 of OPEC\"s\ndifferentials committee has been postponed until the start of\nApril at the request of certain of the body\"s members.\n Ecuador\"s deputy energy minister Fernando Santos Alvite said\nlast Wednesday his debt-burdened country wanted OPEC to assign\na lower official price for its crude, and was to seek this at\ntalks this month of opec\"s pricing committee.\n Referring to pressure by oil companies on OPEC members, in\napparent reference to difficulties faced by Qatar, he said: \"We\nexpected such pressure. It will continue through March and\nApril.\" But he expected the situation would later improve.\n REUTER"  $`237`  [1] "Indonesia appears to be nearing a\npolitical crossroads over measures to deregulate its protected\neconomy, the U.S. Embassy says in a new report.\n To counter falling oil revenues, the government has\nlaunched a series of measures over the past nine months to\nboost exports outside the oil sector and attract new\ninvestment.\n Indonesia, the only Asian member of OPEC and a leading\nprimary commodity producer, has been severely hit by last year\"s\nfall in world oil prices, which forced it to devalue its\ncurrency by 31 pct in September.\n But the U.S. Embassy report says President Suharto\"s\ngovernment appears to be divided over what direction to lead\nthe economy.\n \"(It) appears to be nearing a crossroads with regard to\nderegulation, both as it pertains to investments and imports,\"\nthe report says. It primarily assesses Indonesia\"s agricultural\nsector, but also reviews the country\"s general economic\nperformance.\n It says that while many government officials and advisers\nare recommending further relaxation, \"there are equally strong\npressures being exerted to halt all such moves.\"\n \"This group strongly favours an import substitution economy,\"\nthe report says.\n Indonesia\"s economic changes have been welcomed by the World\nBank and international bankers as steps in the right direction,\nthough they say crucial areas of the economy like plastics and\nsteel remain highly protected, and virtual monopolies.\n Three sets of measures have been announced since last May,\nwhich broadened areas for foreign investment, reduced trade\nrestrictions and liberalised imports.\n The report says Indonesia\"s economic growth in calendar 1986\nwas probably about zero, and the economy may even have\ncontracted a bit. \"This is the lowest rate of growth since the\nmid-1960s,\" the report notes.\n Indonesia, the largest country in South-East Asia with a\npopulation of 168 million, is facing general elections in\nApril.\n But the report hold out little hope for swift improvement\nin the economic outlook. \"For 1987 early indications point to a\nslightly positive growth rate not exceeding one pct. Economic\nactivity continues to suffer due to the sharp fall in export\nearnings from the petroleum industry.\"\n \"Growth in the non-oil sector is low because of weak\ndomestic demand coupled with excessive plant capacity, real\ndeclines in construction and trade, and a reduced level of\ngrowth in agriculture,\" the report states.\n Bankers say continuation of present economic reforms is\ncrucial for the government to get the international lending its\nneeds.\n A new World Bank loan of 300 mln dlrs last month in balance\nof payments support was given partly to help the government\nmaintain the momentum of reform, the Bank said.\n REUTER"  $`242`  [1] "Saudi riyal interbank deposits were\nsteady at yesterday's higher levels in a quiet market.\n Traders said they were reluctant to take out new positions\namidst uncertainty over whether OPEC will succeed in halting\nthe current decline in oil prices.\n Oil industry sources said yesterday several Gulf Arab\nproducers had had difficulty selling oil at official OPEC\nprices but Kuwait has said there are no plans for an emergency\nmeeting of the 13-member organisation.\n A traditional Sunday lull in trading due to the European\nweekend also contributed to the lack of market activity.\n Spot-next and one-week rates were put at 6-1/4, 5-3/4 pct\nafter quotes ranging between seven, six yesterday.\n One, three, and six-month deposits were quoted unchanged at\n6-5/8, 3/8, 7-1/8, 6-7/8 and 7-3/8, 1/8 pct respectively.\n The spot riyal was quietly firmer at 3.7495/98 to the\ndollar after quotes of 3.7500/03 yesterday.\n REUTER"  $`246`  [1] "The Gulf oil state of Qatar, recovering\nslightly from last year's decline in world oil prices,\nannounced its first budget since early 1985 and projected a\ndeficit of 5.472 billion riyals.\n The deficit compared with a shortfall of 7.3 billion riyals\nin the last published budget for 1985/86.\n In a statement outlining the budget for the fiscal year\n1987/88 beginning today, Finance and Petroleum Minister Sheikh\nAbdul-Aziz bin Khalifa al-Thani said the government expected to\nspend 12.217 billion riyals in the period.\n Projected expenditure in the 1985/86 budget had been 15.6\nbillion riyals.\n Sheikh Abdul-Aziz said government revenue would be about\n6.745 billion riyals, down by about 30 pct on the 1985/86\nprojected revenue of 9.7 billion.\n The government failed to publish a 1986/87 budget due to\nuncertainty surrounding oil revenues.\n Sheikh Abdul-Aziz said that during that year the government\ndecided to limit recurrent expenditure each month to\none-twelfth of the previous fiscal year's allocations minus 15\npct.\n He urged heads of government departments and public\ninstitutions to help the government rationalise expenditure. He\ndid not say how the 1987/88 budget shortfall would be covered.\n Sheikh Abdul-Aziz said plans to limit expenditure in\n1986/87 had been taken in order to relieve the burden placed on\nthe country's foreign reserves.\n He added in 1987/88 some 2.766 billion riyals had been\nallocated for major projects including housing and public\nbuildings, social services, health, education, transport and\ncommunications, electricity and water, industry and\nagriculture.\n No figure was revealed for expenditure on defence and\nsecurity. There was also no projection for oil revenue.\n Qatar, an OPEC member, has an output ceiling of 285,000\nbarrels per day.\n Sheikh Abdul-Aziz said: \"Our expectations of positive signs\nregarding (oil) price trends, foremost among them OPEC's\ndetermination to shoulder its responsibilites and protect its\nwealth, have helped us make reasonable estimates for the coming\nyear's revenue on the basis of our assigned quota.\"\n REUTER"  $`248`  [1] "Saudi Arabian Oil Minister Hisham Nazer\nreiterated the kingdom's commitment to last December's OPEC\naccord to boost world oil prices and stabilise the market, the\nofficial Saudi Press Agency SPA said.\n Asked by the agency about the recent fall in free market\noil prices, Nazer said Saudi Arabia \"is fully adhering by the\n... Accord and it will never sell its oil at prices below the\npronounced prices under any circumstance.\"\n Nazer, quoted by SPA, said recent pressure on free market\nprices \"may be because of the end of the (northern hemisphere)\nwinter season and the glut in the market.\"\n Saudi Arabia was a main architect of the December accord,\nunder which OPEC agreed to lower its total output ceiling by\n7.25 pct to 15.8 mln barrels per day (bpd) and return to fixed\nprices of around 18 dlrs a barrel.\n The agreement followed a year of turmoil on oil markets,\nwhich saw prices slump briefly to under 10 dlrs a barrel in\nmid-1986 from about 30 dlrs in late 1985. Free market prices\nare currently just over 16 dlrs.\n Nazer was quoted by the SPA as saying Saudi Arabia's\nadherence to the accord was shown clearly in the oil market.\n He said contacts among members of OPEC showed they all\nwanted to stick to the accord.\n In Jamaica, OPEC President Rilwanu Lukman, who is also\nNigerian Oil Minister, said the group planned to stick with the\npricing agreement.\n \"We are aware of the negative forces trying to manipulate\nthe operations of the market, but we are satisfied that the\nfundamentals exist for stable market conditions,\" he said.\n Kuwait's Oil Minister, Sheikh Ali al-Khalifa al-Sabah, said\nin remarks published in the emirate's daily Al-Qabas there were\nno plans for an emergency OPEC meeting to review prices.\n Traders and analysts in international oil markets estimate\nOPEC is producing up to one mln bpd above the 15.8 mln ceiling.\n They named Kuwait and the United Arab Emirates, along with\nthe much smaller producer Ecuador, among those producing above\nquota. Sheikh Ali denied that Kuwait was over-producing.\n REUTER"  $`273`  [1] "Saudi crude oil output last month fell\nto an average of 3.5 mln barrels per day (bpd) from 3.8 mln bpd\nin January, Gulf oil sources said.\n They said exports from the Ras Tanurah and Ju'aymah\nterminals in the Gulf fell to an average 1.9 mln bpd last month\nfrom 2.2 mln in January because of lower liftings by some\ncustomers.\n But the drop was much smaller than expected after Gulf\nexports rallied in the fourth week of February to 2.5 mln bpd\nfrom 1.2 mln in the third week, the sources said.\n The production figures include neutral zone output but not\nsales from floating storage, which are generally considered\npart of a country's output for Opec purposes.\n Saudi Arabia has an Opec quota of 4.133 mln bpd under a\nproduction restraint scheme approved by the 13-nation group\nlast December to back new official oil prices averaging 18 dlrs\na barrel.\n The sources said the two-fold jump in exports last week\nappeared to be the result of buyers rushing to lift February\nentitlements before the month-end.\n Last week's high export levels appeared to show continued\nsupport for official Opec prices from Saudi Arabia's main crude\ncustomers, the four ex-partners of Aramco, the sources said.\n The four -- Exxon Corp <XON>, Mobil Corp <MOB>, Texaco Inc\n<TX> and Chevron Corp <CHV> -- signed a long-term agreement\nlast month to buy Saudi crude for 17.52 dlrs a barrel.\n However the sources said the real test of Saudi Arabia's\nability to sell crude at official prices in a weak market will\ncome this month, when demand for petroleum products\ntraditionally tapers off. Spot prices have fallen in recent\nweeks to more than one dlr below Opec levels.\n Saudi Arabian oil minister Hisham Nazer yesterday\nreiterated the kingdom's commitment to the December OPEC accord\nand said it would never sell below official prices.\n The sources said total Saudi refinery throughput fell\nslightly in February to an average 1.1 mln bpd from 1.2 mln in\nJanuary because of cuts at the Yanbu and Jubail export\nrefineries.\n They put crude oil exports through Yanbu at 100,000 bpd\nlast month, compared to zero in January, while throughput at\nBahrain's refinery and neutral zone production remained steady\nat around 200,000 bpd each.\n REUTER"  $`349`  [1] "Deputy oil ministers from six Gulf\nArab states will meet in Bahrain today to discuss coordination\nof crude oil marketing, the official Emirates news agency WAM\nreported.\n WAM said the officials would be discussing implementation\nof last Sunday's agreement in Doha by Gulf Cooperation Council\n(GCC) oil ministers to help each other market their crude oil.\n Four of the GCC states - Saudi Arabia, the United Arab\nEmirates (UAE), Kuwait and Qatar - are members of the\nOrganiaation of Petroleum Exporting Countries (OPEC) and some\nface stiff buyer resistance to official OPEC prices.\n Reuter"  $`352`  [1] "Saudi Arabian Oil Minister Hisham Nazer\nreiterated the kingdom's commitment to last December's OPEC\naccord to boost world oil prices and stabilize the market, the\nofficial Saudi Press Agency SPA said.\n Asked by the agency about the recent fall in free market\noil prices, Nazer said Saudi Arabia \"is fully adhering by the\n... accord and it will never sell its oil at prices below the\npronounced prices under any circumstance.\"\n Saudi Arabia was a main architect of December pact under\nwhich OPEC agreed to cut its total oil output ceiling by 7.25\npct and return to fixed prices of around 18 dollars a barrel.\n Reuter"  $`353`  [1] "Kuwait's oil minister said in a newspaper\ninterview that there were no plans for an emergency OPEC\nmeeting after the recent weakness in world oil prices.\n Sheikh Ali al-Khalifa al-Sabah was quoted by the local\ndaily al-Qabas as saying that \"none of the OPEC members has\nasked for such a meeting.\"\n He also denied that Kuwait was pumping above its OPEC quota\nof 948,000 barrels of crude daily (bpd).\n Crude oil prices fell sharply last week as international\noil traders and analysts estimated the 13-nation OPEC was\npumping up to one million bpd over its self-imposed limits.\n Reuter"  $`368`  [1] "The port of Philadelphia was closed\nwhen a Cypriot oil tanker, Seapride II, ran aground after\nhitting a 200-foot tower supporting power lines across the\nriver, a Coast Guard spokesman said.\n He said there was no oil spill but the ship is lodged on\nrocks opposite the Hope Creek nuclear power plant in New\nJersey.\n He said the port would be closed until today when they\nhoped to refloat the ship on the high tide.\n After delivering oil to a refinery in Paulsboro, New\nJersey, the ship apparently lost its steering and hit the power\ntransmission line carrying power from the nuclear plant to the\nstate of Delaware.\n Reuter"  $`489`  [1] "A study group said the United States\nshould increase its strategic petroleum reserve to one mln\nbarrels as one way to deal with the present and future impact\nof low oil prices on the domestic oil industry.\n U.S. policy now is to raise the strategic reserve to 750\nmln barrels, from its present 500 mln, to help protect the\neconomy from an overseas embargo or a sharp price rise.\n The Aspen Institute for Humanistic Studies, a private\ngroup, also called for new research for oil exploration and\ndevelopment techniques.\n It predicted prices would remain at about 15-18 dlrs a\nbarrel for several years and then rise to the mid 20s, with\nimports at about 30 pct of U.S. consumption.\n It said instead that such moves as increasing oil reserves\nand more exploration and development research would help to\nguard against or mitigate the risks of increased imports.\n Reuter"  $`502`  [1] "A study group said the United States\nshould increase its strategic petroleum reserve to one mln\nbarrels as one way to deal with the present and future impact\nof low oil prices on the domestic oil industry.\n U.S. policy now is to raise the strategic reserve to 750\nmln barrels, from its present 500 mln, to help protect the\neconomy from an overseas embargo or a sharp price rise.\n The Aspen Institute for Humanistic Studies, a private\ngroup, also called for new research for oil exploration and\ndevelopment techniques.\n It predicted prices would remain at about 15-18 dlrs a\nbarrel for several years and then rise to the mid 20s, with\nimports at about 30 pct of U.S. consumption.\n The study cited two basic policy paths for the nation: to\nprotect the U.S. industry through an import fee or other such\ndevice or to accept the full economic benefits of cheap oil.\n But the group did not strongly back either option, saying\nthere were benefits and drawbacks to both.\n It said instead that such moves as increasing oil reserves\nand more exploration and development research would help to\nguard against or mitigate the risks of increased imports.\n Reuter"  $`543`  [1] "Unocal Corp's Union Oil Co said it\nlowered its posted prices for crude oil one to 1.50 dlrs a\nbarrel in the eastern region of the U.S., effective Feb 26.\n Union said a 1.50 dlrs cut brings its posted price for the\nU.S. benchmark grade, West Texas Intermediate, to 16 dlrs.\nLouisiana Sweet also was lowered 1.50 dlrs to 16.35 dlrs, the\ncompany said.\n No changes were made in Union's posted prices for West\nCoast grades of crude oil, the company said.\n Reuter"  $`704`  [1] "The New York Mercantile Exchange set\nApril one for the debut of a new procedure in the energy\ncomplex that will increase the use of energy futures worldwide.\n On April one, NYMEX will allow oil traders that do not\nhold a futures position to initiate, after the exchange closes,\na transaction that can subsequently be hedged in the futures\nmarket, according to an exchange spokeswoman.\n \"This will change the way oil is transacted in the real\nworld,\" said said Thomas McKiernan, McKiernan and Co chairman.\n Foreign traders will be able to hedge trades against NYMEX\nprices before the exchange opens and negotiate prices at a\ndifferential to NYMEX prices, McKiernan explained.\n The expanded program \"will serve the industry because the\noil market does not close when NYMEX does,\" said Frank Capozza,\nsecretary of Century Resources Inc.\n The rule change, which has already taken effect for\nplatinum futures on NYMEX, is expected to increase the open\ninterest and liquidity in U.S. energy futures, according to\ntraders and analysts.\n Currently, at least one trader in this transaction, called\nan exchange for physical or EFP, must hold a futures position\nbefore entering into the transaction.\n Under the new arrangement, neither party has to hold a\nfutures position before entering into an EFP and one or both\nparties can offset their cash transaction with a futures\ncontract the next day, according to exchange officials.\n When NYMEX announced its proposed rule change in December,\nNYMEX President Rosemary McFadden, said, \"Expansion of the EFP\nprovision will add to globalization of the energy markets by\nproviding for, in effect, 24-hour trading.\"\n The Commodity Futures Trading Commission approved the rule\nchange in February, according to a CFTC spokeswoman.\n Reuter"  $`708`  [1] "Argentine crude oil production was\ndown 10.8 pct in January 1987 to 12.32 mln barrels, from 13.81\nmln barrels in January 1986, Yacimientos Petroliferos Fiscales\nsaid.\n January 1987 natural gas output totalled 1.15 billion cubic\nmetrers, 3.6 pct higher than 1.11 billion cubic metres produced\nin January 1986, Yacimientos Petroliferos Fiscales added.\n Reuter"  > # inspect function  > inspect(crude[1:3]) #see data in detail(including Metadata,Content) of 1 to 3 articles  <<VCorpus>>  Metadata: corpus specific: 0, document level (indexed): 0  Content: documents: 3  $`reut-00001.xml`  <<PlainTextDocument>>  Metadata: 15  Content: chars: 527  $`reut-00002.xml`  <<PlainTextDocument>>  Metadata: 15  Content: chars: 2634  $`reut-00004.xml`  <<PlainTextDocument>>  Metadata: 15  Content: chars: 330  > inspect(crude[[1]]) #see data in detail(including Metadata,Content) of article 1  <<PlainTextDocument>>  Metadata: 15  Content: chars: 527  Diamond Shamrock Corp said that  effective today it had cut its contract prices for crude oil by  1.50 dlrs a barrel.  The reduction brings its posted price for West Texas  Intermediate to 16.00 dlrs a barrel, the copany said.  "The price reduction today was made in the light of falling  oil product prices and a weak crude oil market," a company  spokeswoman said.  Diamond is the latest in a line of U.S. oil companies that  have cut its contract, or posted, prices over the last two days  citing weak oil markets.  Reuter  > # 1. remove punctuation in documnet -->use tm\_map function.  > crude<-tm\_map(crude, removePunctuation) #overwrite  > content(crude[[1]]) #punctuations of contents of article1 are removed  [1] "Diamond Shamrock Corp said that\neffective today it had cut its contract prices for crude oil by\n150 dlrs a barrel\n The reduction brings its posted price for West Texas\nIntermediate to 1600 dlrs a barrel the copany said\n The price reduction today was made in the light of falling\noil product prices and a weak crude oil market a company\nspokeswoman said\n Diamond is the latest in a line of US oil companies that\nhave cut its contract or posted prices over the last two days\nciting weak oil markets\n Reuter"  > # 2. remove numbers  > crude<-tm\_map(crude, removeNumbers)  > content(crude[[1]]) #numbers of contents of article1 are removed  [1] "Diamond Shamrock Corp said that\neffective today it had cut its contract prices for crude oil by\n dlrs a barrel\n The reduction brings its posted price for West Texas\nIntermediate to dlrs a barrel the copany said\n The price reduction today was made in the light of falling\noil product prices and a weak crude oil market a company\nspokeswoman said\n Diamond is the latest in a line of US oil companies that\nhave cut its contract or posted prices over the last two days\nciting weak oil markets\n Reuter"  > # 3. remove stopwords  > crude<-tm\_map(crude, function(x) removeWords(x,stopwords())) #functions(x)-->user made function  > content(crude[[1]]) #stopwords of contents of article1 are removed  [1] "Diamond Shamrock Corp said \neffective today cut contract prices crude oil \n dlrs barrel\n The reduction brings posted price West Texas\nIntermediate dlrs barrel copany said\n The price reduction today made light falling\noil product prices weak crude oil market company\nspokeswoman said\n Diamond latest line US oil companies \n cut contract posted prices last two days\nciting weak oil markets\n Reuter"  > stopwords() #meaningless contents of the whole article  [1] "i" "me" "my" "myself" "we" "our" "ours"  [8] "ourselves" "you" "your" "yours" "yourself" "yourselves" "he"  [15] "him" "his" "himself" "she" "her" "hers" "herself"  [22] "it" "its" "itself" "they" "them" "their" "theirs"  [29] "themselves" "what" "which" "who" "whom" "this" "that"  [36] "these" "those" "am" "is" "are" "was" "were"  [43] "be" "been" "being" "have" "has" "had" "having"  [50] "do" "does" "did" "doing" "would" "should" "could"  [57] "ought" "i'm" "you're" "he's" "she's" "it's" "we're"  [64] "they're" "i've" "you've" "we've" "they've" "i'd" "you'd"  [71] "he'd" "she'd" "we'd" "they'd" "i'll" "you'll" "he'll"  [78] "she'll" "we'll" "they'll" "isn't" "aren't" "wasn't" "weren't"  [85] "hasn't" "haven't" "hadn't" "doesn't" "don't" "didn't" "won't"  [92] "wouldn't" "shan't" "shouldn't" "can't" "cannot" "couldn't" "mustn't"  [99] "let's" "that's" "who's" "what's" "here's" "there's" "when's"  [106] "where's" "why's" "how's" "a" "an" "the" "and"  [113] "but" "if" "or" "because" "as" "until" "while"  [120] "of" "at" "by" "for" "with" "about" "against"  [127] "between" "into" "through" "during" "before" "after" "above"  [134] "below" "to" "from" "up" "down" "in" "out"  [141] "on" "off" "over" "under" "again" "further" "then"  [148] "once" "here" "there" "when" "where" "why" "how"  [155] "all" "any" "both" "each" "few" "more" "most"  [162] "other" "some" "such" "no" "nor" "not" "only"  [169] "own" "same" "so" "than" "too" "very"  > # 4. contruct term-doucument matrix  > tdm<-TermDocumentMatrix(crude) #How many times this term appears in one document  > inspect(tdm) # term documents can be appeared in matrix form  <<TermDocumentMatrix (terms: 962, documents: 20)>>  Non-/sparse entries: 1710/17530  Sparsity : 91%  Maximal term length: 16  Weighting : term frequency (tf)  Sample :  Docs  Terms 144 236 237 242 246 248 273 489 502 704  bpd 4 7 0 0 0 2 8 0 0 0  crude 0 2 0 0 0 0 5 0 0 0  dlrs 0 2 1 0 0 4 2 1 1 0  last 1 4 3 0 2 1 7 0 0 0  mln 4 4 1 0 0 3 9 3 3 0  oil 12 7 3 3 5 9 5 4 5 3  opec 13 6 1 2 1 6 5 0 0 0  prices 5 5 1 2 1 9 5 2 2 3  said 11 10 1 3 5 7 8 2 2 4  the 2 0 1 1 3 1 4 1 2 4  > 962\*20 #term \* documents  [1] 19240  > # 5. read tdm as a matrix  > m<-as.matrix(tdm) #convert tdm into matrix. shows what terms appear in documents  > head(m)  Docs  Terms 127 144 191 194 211 236 237 242 246 248 273 349 352 353 368 489 502 543 704 708  abdulaziz 0 0 0 0 0 0 0 0 5 0 0 0 0 0 0 0 0 0 0 0  ability 0 2 0 0 0 3 0 0 0 0 1 0 0 0 0 0 0 0 0 0  able 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0  abroad 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0  accept 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0  accord 0 0 0 0 0 0 0 0 0 5 1 0 2 0 0 0 0 0 0 0  > dim(m) # 962rows and 20columns. out of 962 words, there are 20 frequency.  [1] 962 20  > # 6. sorting in high frequency to low frequency  > v<-sort(rowSums(m), decreasing=TRUE) # Count from the highest from the lowest  > v[1:10] #Top 10 frequency from the highest to the lowest  oil said prices opec mln the last bpd dlrs crude  85 73 48 42 31 26 24 23 23 21  > # 7. match with freq and word names  > d<-data.frame(word=names(v), freq=v) #make the dataframe  > head(d) #Top 6 frequencies  word freq  oil oil 85  said said 73  prices prices 48  opec opec 42  mln mln 31  the the 26  > tail(d) #6 lowest frequencies  word freq  whether whether 1  wishes wishes 1  worldwide worldwide 1  xon xon 1  yergin yergin 1  yesterdays yesterdays 1  > # 7. plot a word cloud  > wordcloud(d$word, d$freq) #row-->word, column-->frequency  > # 7-2. color plot with frequent words plotted first  > pal <- brewer.pal(9,"BuGn") #from RcolorBrewer  > head(pal) #codenames of the colors  [1] "#F7FCFD" "#E5F5F9" "#CCECE6" "#99D8C9" "#66C2A4" "#41AE76"  > pal <- pal[-(1:4)]  > wordcloud(d$word,d$freq,c(8,.3),2,,FALSE,,.15,pal)  There were 50 or more warnings (use warnings() to see the first 50)  > library(RCurl)  > library(XML)  > t = readLines('https://en.wikipedia.org/wiki/Data\_science')  > d = htmlParse(t, asText = TRUE) #parse the html to text file. Mandatory process  > clean\_doc = xpathSApply(d,"//p", xmlValue) # change the data type for R programming.  > library(tm) #tm library for text data mining  > library(SnowballC) #SnowballC is used for word stemming  > doc = Corpus(VectorSource(clean\_doc)) #Corpus is a kind of the chunk  > inspect(doc)  <<SimpleCorpus>>  Metadata: corpus specific: 1, document level (indexed): 0  Content: documents: 15  [1] \n  [2] Data science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structural and unstructured data.[1][2] Data science is related to data mining, deep learning and big data.\n  [3] Data science is a "concept to unify statistics, data analysis, machine learning, domain knowledge and their related methods" in order to "understand and analyze actual phenomena" with data.[3] It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, domain knowledge and information science. Turing award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.[4][5]  [4] Data science is an interdisciplinary field focused on extracting knowledge from data sets, which are typically large (see big data).[6] The field encompasses analysis, preparing data for analysis, and presenting findings to inform high-level decisions in an organization. As such, it incorporates skills from computer science, mathematics, statistics, information visualization, graphic design, and business.[7][8] Statistician Nathan Yau, drawing on Ben Fry, also links data science to human-computer interaction: users should be able to intuitively control and explore data.[9][10] In 2015, the American Statistical Association identified database management, statistics and machine learning, and distributed and parallel systems as the three emerging foundational professional communities.[11]  [5] Many statisticians, including Nate Silver, have argued that data science is not a new field, but rather another name for statistics.[12] Others argue that data science is distinct from statistics because it focuses on problems and techniques unique to digital data.[13]Vasant Dhar writes that statistics emphasizes quantitative data and description. In contrast, data science deals with quantitative and qualitative data (e.g. images) and emphasizes prediction and action.[14]Andrew Gelman of Columbia University and data scientist Vincent Granville have described statistics as a nonessential part of data science.[15][16]\nStanford professor David Donoho writes that data science is not distinguished from statistics by the size of datasets or use of computing, and that many graduate programs misleadingly advertise their analytics and statistics training as the essence of a data science program. He describes data science as an applied field growing out of traditional statistics.[17] \nIn summary, data science can be therefore described as an applied branch of statistics.\n  [6] In 1962, John Tukey described a field he called “data analysis,” which resembles modern data science.[17] Later, attendees at a 1992 statistics symposium at the University of Montpellier II acknowledged the emergence of a new discipline focused on data of various origins and forms, combining established concepts and principles of statistics and data analysis with computing.[18][19]  [7] The term “data science” has been traced back to 1974, when Peter Naur proposed it as an alternative name for computer science.[20] In 1996, the International Federation of Classification Societies became the first conference to specifically feature data science as a topic.[20] However, the definition was still in flux. In 1997, C.F. Jeff Wu suggested that statistics should be renamed data science. He reasoned that a new name would help statistics shed inaccurate stereotypes, such as being synonymous with accounting, or limited to describing data.[21] In 1998, Chikio Hayashi argued for data science as a new, interdisciplinary concept, with three aspects: data design, collection, and analysis.[22]  [8] During the 1990s, popular terms for the process of finding patterns in datasets (which were increasingly large) included “knowledge discovery” and “data mining.”[23][20]  [9] The modern conception of data science as an independent discipline is sometimes attributed to William S. Cleveland.[24] In a 2001 paper, he advocated an expansion of statistics beyond theory into technical areas; because this would significantly change the field, it warranted a new name.[23] "Data science" became more widely used in the next few years: in 2002, the Committee on Data for Science and Technology launched Data Science Journal. In 2003, Columbia University launched The Journal of Data Science.[23] In 2014, the American Statistical Association's Section on Statistical Learning and Data Mining changed its name to the Section on Statistical Learning and Data Science, reflecting the ascendant popularity of data science.[25]  [10] The professional title of “data scientist” has been attributed to DJ Patil and Jeff Hammerbacher in 2008.[26] Though it was used by the National Science Board in their 2005 report, "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century," it referred broadly to any key role in managing a digital data collection.[27]  [11] There is still no consensus on the definition of data science and it is considered by some to be a buzzword.[28]  [12] Data science is a growing field. A career as a data scientist is ranked at the third best job in America for 2020 by Glassdoor, and was ranked the number one best job from 2016-2019.[29]  [13] In order to become a data scientist, there is a significant amount of education and experience required. The first step in becoming a data scientist is to earn a bachelor's degree, typically in a field related to computing or mathematics.[30] Coding bootcamps are also available and can be used as an alternate pre-qualification to supplement a bachelor's degree in another field.[30] Most data scientists also complete a master’s degree or a PhD in data science.[30] Once these qualifications are met, the next step to becoming a data scientist is to apply for an entry-level job in the field.[30] Some data scientists may later choose to specialize in a sub-field of data science.[30]  [14] Big data is very quickly becoming a vital tool for businesses and companies of all sizes.[32] The availability and interpretation of big data has altered the business models of old industries and enabled the creation of new ones.[32] Data-driven businesses are worth $1.2 trillion collectively in 2020, an increase from $333 billion in the year 2015.[33] Data scientists are responsible for breaking down big data into usable information and creating software and algorithms that help companies and organizations determine optimal operations.[33] As big data continues to have a major impact on the world, data science does as well due to the close relationship between the two.[33]  [15] There are a variety of different technologies and techniques that are used for data science which depend on the application. More recently, full-featured, end-to-end platforms have been developed and heavily used for data science and machine learning.\n  > #tm\_map(): The function to preprocess  > doc = tm\_map(doc, content\_transformer(tolower)) #content transform from capital letter to lower letter  Warning message:  In tm\_map.SimpleCorpus(doc, content\_transformer(tolower)) :  transformation drops documents  > doc = tm\_map(doc, removeNumbers) #remove numbers  Warning message:  In tm\_map.SimpleCorpus(doc, removeNumbers) : transformation drops documents  > doc = tm\_map(doc, removeWords, stopwords('english')) #remove unEnlgish words  Warning message:  In tm\_map.SimpleCorpus(doc, removeWords, stopwords("english")) :  transformation drops documents  > doc = tm\_map(doc, removePunctuation) # remove punctuation  Warning message:  In tm\_map.SimpleCorpus(doc, removePunctuation) :  transformation drops documents  > doc = tm\_map(doc, stripWhitespace) #remove space from words and sentences  Warning message:  In tm\_map.SimpleCorpus(doc, stripWhitespace) :  transformation drops documents  > dtm = DocumentTermMatrix(doc) #make preprocessed documents to matrix  > dim(dtm) # 15 rows and 370 columns  [1] 15 376  > inspect(dtm) # summarize the information of the matrix  <<DocumentTermMatrix (documents: 15, terms: 376)>>  Non-/sparse entries: 500/5140  Sparsity : 91%  Maximal term length: 17  Weighting : term frequency (tf)  Sample :  Terms  Docs analysis big data field learning name new science scientist statistics  10 0 0 2 0 0 0 0 1 0 0  13 0 0 7 3 0 0 0 2 3 0  14 0 4 6 0 0 0 1 1 0 0  2 0 1 5 1 1 0 0 2 0 0  3 1 0 5 0 1 0 0 6 0 2  4 2 1 6 2 1 0 0 3 0 2  5 0 0 11 2 0 1 1 8 1 8  6 1 0 3 1 0 0 1 1 0 2  7 1 0 5 0 0 2 2 4 0 2  9 0 0 8 1 2 2 1 7 0 1  > # <<DocumentTermMatrix (documents: 16, terms: 392)>>  > # out of 16 documents we extract 392 words  > # Non-/sparse entries: 526/5730  > # out of 6272, 5744 are sparsity(92%)  > 16\*392  [1] 6272  > library(wordcloud) #use wordcloud to visualize the result  > m = as.matrix(dtm) #convert result to document term matrix  > v = sort(colSums(m), decreasing = TRUE) # Order from highest to lowest one  > d = data.frame(word = names(v), freq = v) #row-->word,column-->freq  > wordcloud(words = d$word, freq = d$freq,  + min.freq = 1, max.words = 100, random.order = FALSE,  + rot.per = 0.35)  > library(RColorBrewer) #Color Palette  > pal = brewer.pal(11,"Spectral")  > # Controlling counting in wordcloud function  > wordcloud(words = d$word, freq = d$freq,  + min.freq = 1, max.words = 50,  + random.order = FALSE, rot.per = 0.50, colors = pal)  > # adjusting font --> ex.)family = "mono", font = 2  > wordcloud(words = d$word, freq = d$freq,  + min.freq = 1, max.words = 50,  + random.order = FALSE, rot.per = 0.50, colors = pal,  + family = "mono", font = 2)  > library(wordcloud2)  > wordcloud2(d)  > # need to select words becuase wordcloud2 cannot designate how many counts we can control  > d1 = d[1:200, ] #select up to 200 words  > wordcloud2(d1, shape = 'star') #option of the wordcloud. The background will be a star shape  > wordcloud2(d1, minRotation = pi/4, maxRotation = pi/4,  + rotateRatio = 1.0) #select the range,angle,rotation of the shape of the words in the wordcloud  > findFreqTerms(dtm, lowfreq = 10) #select top 10 frequent words  [1] "data" "field" "science" "statistics"  > findAssocs(dtm, terms = 'science', corlimit = 0.7)  $science  data columbia statistics  0.81 0.79 0.74  > findAssocs(dtm, terms = 'data', corlimit = 0.8)  $data  science  0.81  > findAssocs(dtm, terms = 'data', corlimit = 0.6)  $data  science statistics columbia field another can  0.81 0.68 0.68 0.63 0.62 0.62  > barplot(d[1:10, ]$freq, las = 2, names.arg = d[1:10,]$word,  + col = 'lightblue',  + main = 'Top10 frequent words', ylab = 'freq') |
|  |
| |  | | --- | | > | |