

By Gerard Utoware

Overview:

R was chosen due to the ability to analyse, clean and present data all in one platform. It is commonly used throughout research, and it allows ideas and data to be easily understood regardless of data literacy. R also allowed data from excel to be transferred to it which meant there was no hassle and no loss of data while being transferred. If python were to be used for example, it would take several more steps to transfer data, which could lead to lost data.

Based on the data gathered and analysed, the consulting industry appears to be the most fruitful and well rated out of the companies in the dataset. Companies from the consulting industry generally tended to have ratings closer to 5 compared to companies from finance, accounting, or energy.

Zurich, a city well known and established due to its financial relevance, has not had any new companies build their headquarters in its city in over 150 years. This may be part of a larger shift away from Zurich and its relevance in finance. On the other hand, it may be an untapped area to explore for new headquarters for companies.

Woodridge and Chicago, both found in Illinois, are both crucial to different sectors in this dataset. Entering this dataset, this was unexpected as there are several larger cities in the USA that are in larger markets which would have influenced ratings. Woodridge is very dominant in the IT Services sector and has a large market share of said sector.

Q4

The first screenshot shows the RStudio interface with a data frame named 'BusinessAnalyst' loaded. The data frame has 4,092 entries and 17 columns. The columns are: Index, Job Title, Salary Estimate, Job Description, and Rating. The data is displayed in a table view, showing the first 7 rows. The console shows the command `BusinessAnalyst[c('Job Title', 'Rating', 'Location', 'Industry')]` being executed, which filters the data to show only these four columns. The resulting table is displayed in the console, showing the first 29 rows of the filtered data.

The second screenshot shows the same RStudio interface, but with the data frame 'BusinessAnalyst' filtered to show only the first 250 rows. The console shows the command `BusinessAnalyst[c('Job Title', 'Rating', 'Location', 'Industry')]` being executed, which filters the data to show only these four columns. The resulting table is displayed in the console, showing the first 250 rows of the filtered data.

The two screenshots here depict the code and resulting table to display 4 relevant columns. With over 4000 entries, not every entry was displayed.

Q5

```

1 BusinessAnalyst %>%
2   select('Job Title','Sector','Industry','Headquarters','Rating') %>%
3   arrange(desc(Rating)) %>% print(n=20)
4   unique('Company Name', keep_all = FALSE)
5

```

Job Title	Sector	Industry	Headquarters	Rating
1 Business Analyst	-1	-1	San Diego, CA	5
2 Business Analyst	Business Services	Consulting	Yorktown Heights, NY	5
3 Business Process Analyst	Business Services	Staffing & Outsourcing	Far Hills, NJ	5
4 Business Analyst	Business Services	Staffing & Outsourcing	New York, NY	5
5 Business Analyst	-1	-1	Jersey City, NJ	5
6 Business Analyst - Commercial Lending	Business Services	Staffing & Outsourcing	Coconut Creek, FL	5
7 Business Systems Analyst	Information Technology	IT Services	Piscataway, NJ	5
8 Business Analyst	-1	-1	Shepton Mallet, United Kingdom	5
9 Business Analyst	Information Technology	IT Services	Woodridge, IL	5
10 System/Business Analyst	Business Services	Consulting	Yorktown Heights, NY	5
11 Business Analyst - Technical	Business Services	Staffing & Outsourcing	Far Hills, NJ	5
12 Business Development (Institutional Investors)	Information Technology	Enterprise Software & Network Solutions	San Mateo, CA	5
13 Business Intelligence Analyst	Business Services	Staffing & Outsourcing	Secaucus, NJ	5
14 Business Intelligence Analyst	Business Services	Staffing & Outsourcing	New York, NY	5
15 Technical Business Analyst	Information Technology	Enterprise Software & Network Solutions	Berkeley Lake, GA	5
16 Entry Level Business Analyst	Information Technology	IT Services	Woodridge, IL	5
17 Business Data Analyst	-1	-1	West Roxbury, MA	5
18 Business Operations Analyst	Transportation & Logistics	Logistics & Supply Chain	Los Angeles, CA	5
19 Business Analyst- Trainee -Entry Level	Information Technology	IT Services	Woodridge, IL	5
20 Urgent - Entry Level Business Analyst	Information Technology	IT Services	Woodridge, IL	5

The image depicts the line of code used in order to display the top 20 industries, sectors and headquarters by Rating where there are no outright duplicates, only unique values.

Q6a

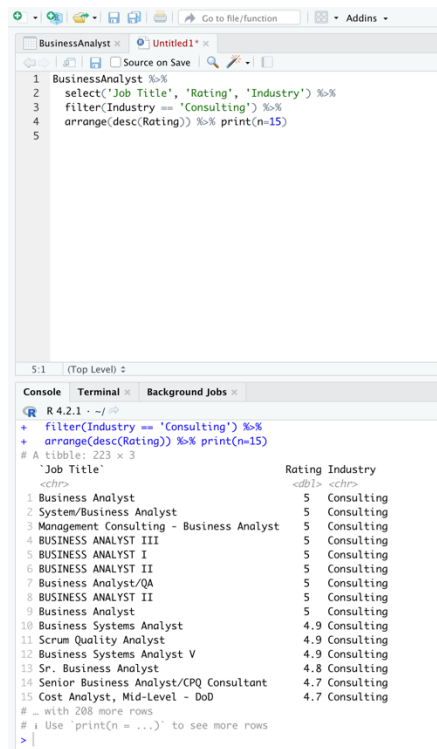
```

> BusinessAnalyst %>%
+   select('Job Title', 'Rating') %>%
+   arrange(desc(Rating)) %>% print(n=15)
# A tibble: 4,092 x 2
#   Job Title          Rating
#   <chr>             <dbl>
1 Business Analyst    5
2 Business Analyst    5
3 Business Process Analyst 5
4 Business Analyst    5
5 Business Analyst    5
6 Business Analyst - Commercial Lending 5
7 Business Systems Analyst 5
8 Business Analyst    5
9 Business Analyst    5
10 System/Business Analyst 5
11 Business Analyst - Technical 5
12 Business Development (Institutional Investors) 5
13 Business Intelligence Analyst 5
14 Business Intelligence Analyst 5
15 Technical Business Analyst 5
# ... with 4,077 more rows
# Use `print(n = ...)` to see more rows

```

Screenshot shows the line of code used to call the top 15 jobs based on rating after loading the dplyr package. All 15 jobs shown here had a perfect rating of 5.

By Gerard Utoware



The screenshot shows the RStudio IDE. The script editor at the top contains the following R code:

```
1 BusinessAnalyst %>%
2   select('Job Title', 'Rating', 'Industry') %>%
3   filter(Industry == 'Consulting') %>%
4   arrange(desc(Rating)) %>% print(n=15)
5
```

The console at the bottom shows the output of the code:

```
R 4.2.1 ~ /
+ filter(Industry == 'Consulting') %>%
+ arrange(desc(Rating)) %>% print(n=15)
# A tibble: 223 x 3
#   Job Title          Rating Industry
#   <chr>              <dbl> <chr>
1 Business Analyst      5 Consulting
2 System/Business Analyst 5 Consulting
3 Management Consulting - Business Analyst 5 Consulting
4 BUSINESS ANALYST III 5 Consulting
5 BUSINESS ANALYST I 5 Consulting
6 BUSINESS ANALYST II 5 Consulting
7 Business Analyst/QA 5 Consulting
8 BUSINESS ANALYST II 5 Consulting
9 Business Analyst 5 Consulting
10 Business Systems Analyst 4.9 Consulting
11 Scrum Quality Analyst 4.9 Consulting
12 Business Systems Analyst V 4.9 Consulting
13 Sr. Business Analyst 4.8 Consulting
14 Senior Business Analyst/CPQ Consultant 4.7 Consulting
15 Cost Analyst, Mid-level - DoD 4.7 Consulting
# ... with 208 more rows
# Use 'print(n = ...)' to see more rows
>
```

Image shows the line of code used to call the top 15 jobs based on rating in the consulting industry after loading the dplyr package. Most of the jobs listed here appear to involve some form of Business analytics.

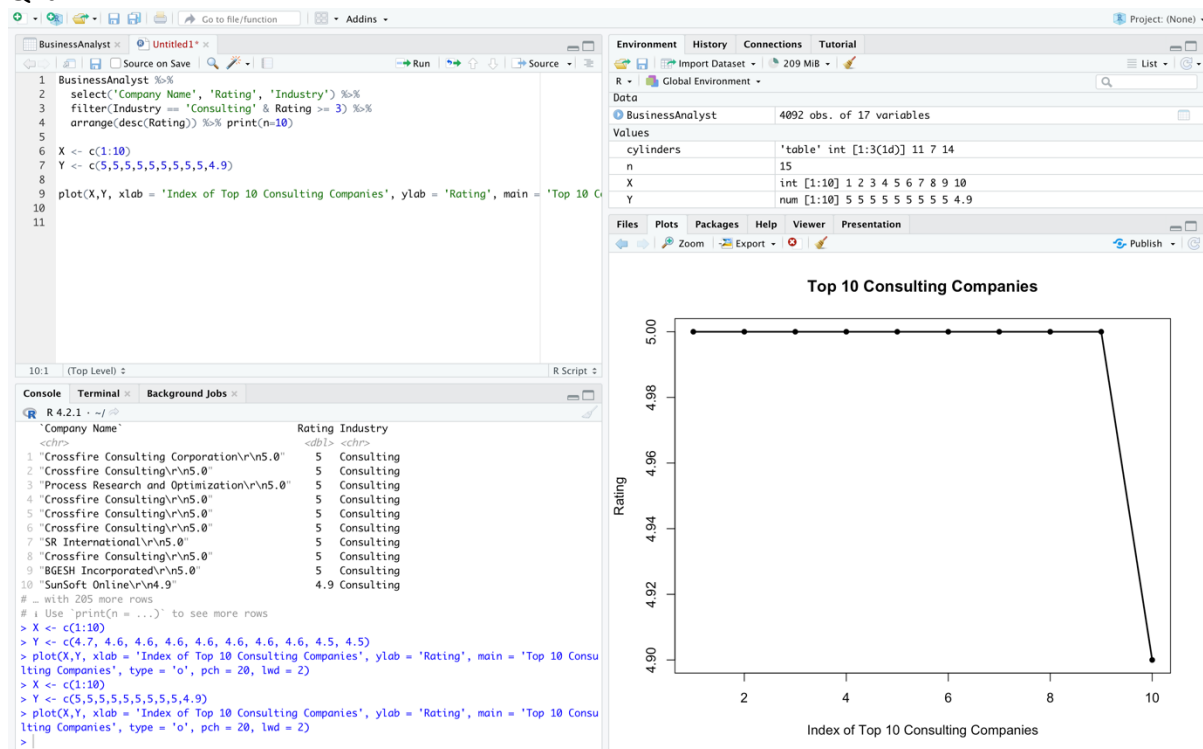
By Gerard Utware

```
BusinessAnalyst %>%
  select('Job Title', 'Rating') %>%
  arrange(Rating) %>% print(n=15)
4 |
```

```
R 4.2.1 ~ ./
10 System/Business Analyst 5
11 Business Analyst - Technical 5
12 Business Development (Institutional Investors) 5
13 Business Intelligence Analyst 5
14 Business Intelligence Analyst 5
15 Technical Business Analyst 5
# _ with 4,077 more rows
# i Use 'print(n = ...)' to see more rows
> BusinessAnalyst %>%
+ select('Job Title', 'Rating') %>%
+ arrange(Rating) %>% print(n=15)
# A tibble: 4,092 x 2
# Job Title Rating
<chr> <dbl>
1 Business Analyst -1
2 Business Analyst -1
3 Business Analyst -1
4 Business Analyst -1
5 Business Analyst -1
6 Business Analyst -1
7 Business Analyst -1
8 Business Analyst -1
9 Senior Data Analyst (E-commerce, Marketing) -1
10 Business Analyst -1
11 Business Analyst -1
12 Business Analyst -1
13 Business Systems Analyst -1
14 Business Analyst -1
15 Business Analyst -1
# _ with 4,077 more rows
# i Use 'print(n = ...)' to see more rows
> |
```

Image shows the line of code used to call the bottom 15 jobs based on rating after loading the dplyr package.

Q7a



Screenshot shows the line of code used to call the top 10 companies based on ratings more than 3 and in the Consulting industry after loading the dplyr package. The consulting had the highest average rating amongst itself, energy, and the accounting industries with only one company in the top 10 not having a rating of 5.

7b

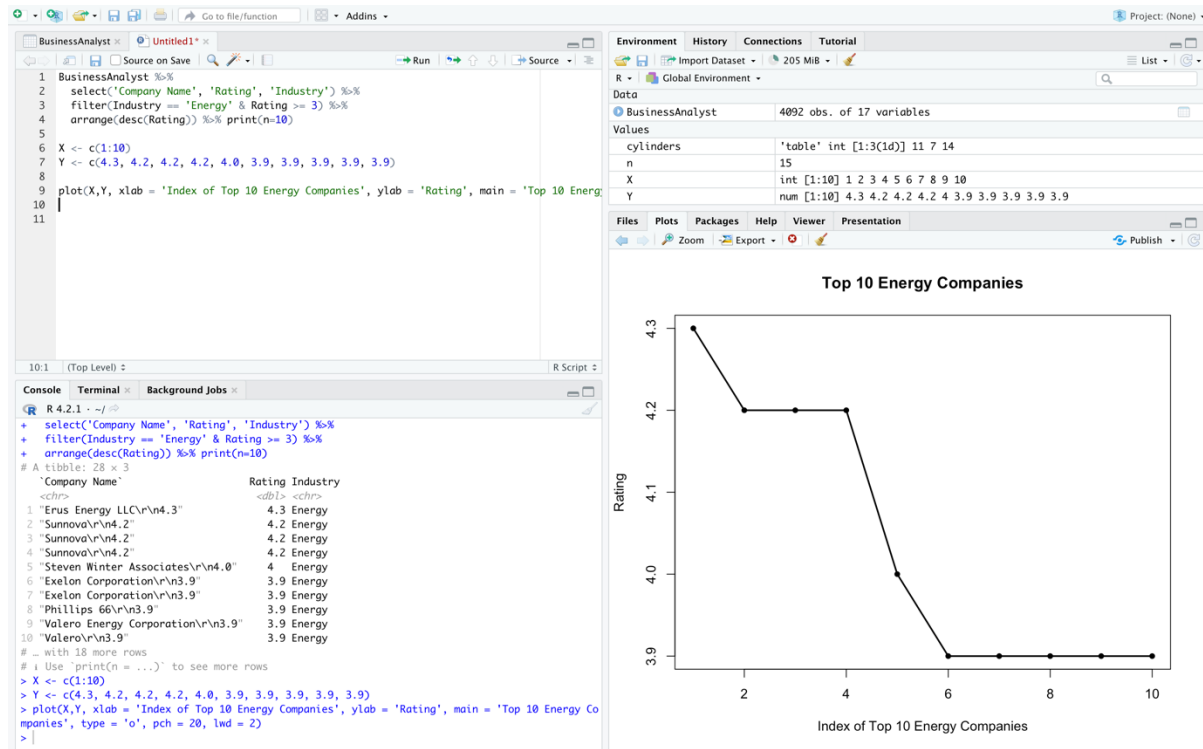


Image depicting the line of code used to call the top 10 companies based on ratings more than 3 and in the Energy industry after loading the dplyr package.

7c

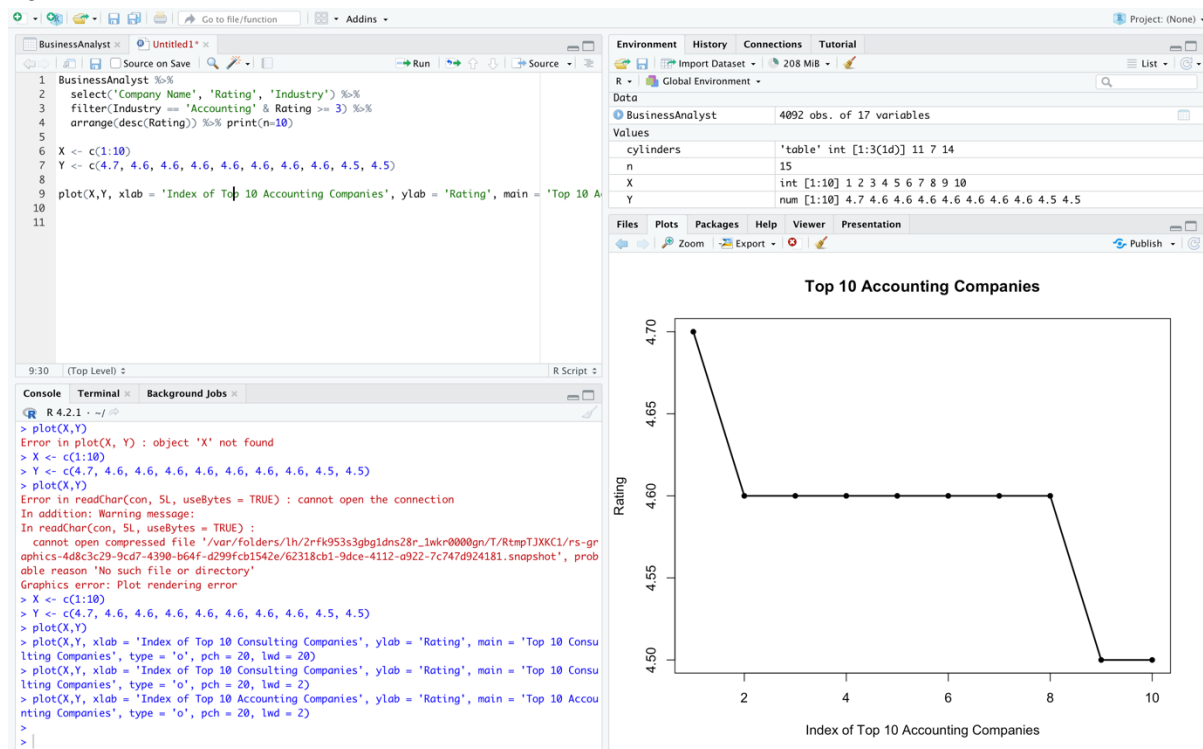


Image depicting the line of code used to call the top 10 companies based on ratings more than 3 and in the accounting industry after loading the dplyr package.

By Gerard Utware

Q8: Extra Summaries

The screenshot shows the RStudio interface with a script editor containing the following R code:

```
1 BusinessAnalyst %>%
2   select('Company Name', 'Rating', 'Industry', 'Headquarters', 'Sector') %>%
3   filter(Sector == 'Finance') %>%
4   arrange(desc(Rating)) %>% print(n=5)
```

The console output shows an error: "Error in 'filter()': Problem while computing '...1 = Sector == 'Finance''. Caused by error in 'maskEval_all_filter()': object 'Sector' not found." The user then modifies the code to select the top 5 finance companies by rating, and the console displays the following tibble:

Company Name	Rating	Industry	Headquarters	Sector
"CreditNinja\r\n5.0"	5	Lending	Chicago, IL	Finance
"Trivest\r\n5.0"	5	Venture Capital & Private Equity	Coral Gables, FL	Finance
"EP Wealth Advisors\r\n4.8"	4.8	Investment Banking & Asset Management	Torrance, CA	Finance
"Braviant Holdings\r\n4.8"	4.8	Lending	Chicago, IL	Finance
"KGS\r\n4.7"	4.7	Investment Banking & Asset Management	New York, NY	Finance

Image depicting the line of code used to call the top 5 Finance companies based on rating. Notably, two of these companies are in Chicago.

The screenshot shows the RStudio interface with a script editor containing the following R code:

```
1 BusinessAnalyst %>%
2   select('Company Name', 'Rating', 'Industry', 'Headquarters') %>%
3   filter(Headquarters == 'New York, NY' & Rating >= 3) %>%
4   arrange(desc(Rating)) %>% print(n=10)
```

The console output shows the following tibble:

Company Name	Rating	Industry	Headquarters
"Creative Data Resources\r\n5.0"	5	Staffing & Outsourcing	New York, NY
"ingenium.agency\r\n5.0"	5	Staffing & Outsourcing	New York, NY
"Wellth Inc.\r\n5.0"	5	Enterprise Software & Network Solutions	New York, NY
"VISTRADA\r\n5.0"	5	Advertising & Marketing	New York, NY
"ingenium.agency\r\n5.0"	5	Staffing & Outsourcing	New York, NY
"ingenium.agency\r\n5.0"	5	Staffing & Outsourcing	New York, NY
"ingenium.agency\r\n5.0"	5	Staffing & Outsourcing	New York, NY
"Teachers Pay Teachers\r\n4.9"	4.9	Internet	New York, NY
"UTC Associates, Inc.\r\n4.7"	4.7	Accounting	New York, NY
"KGS\r\n4.7"	4.7	Investment Banking & Asset Management	New York, NY

Image depicting the line of code used to call the top 10 companies with a rating of over 3 and their headquarters being located in New York. With the Staffing & Outsourcing industry being the most dominant here.

By Gerard Utware

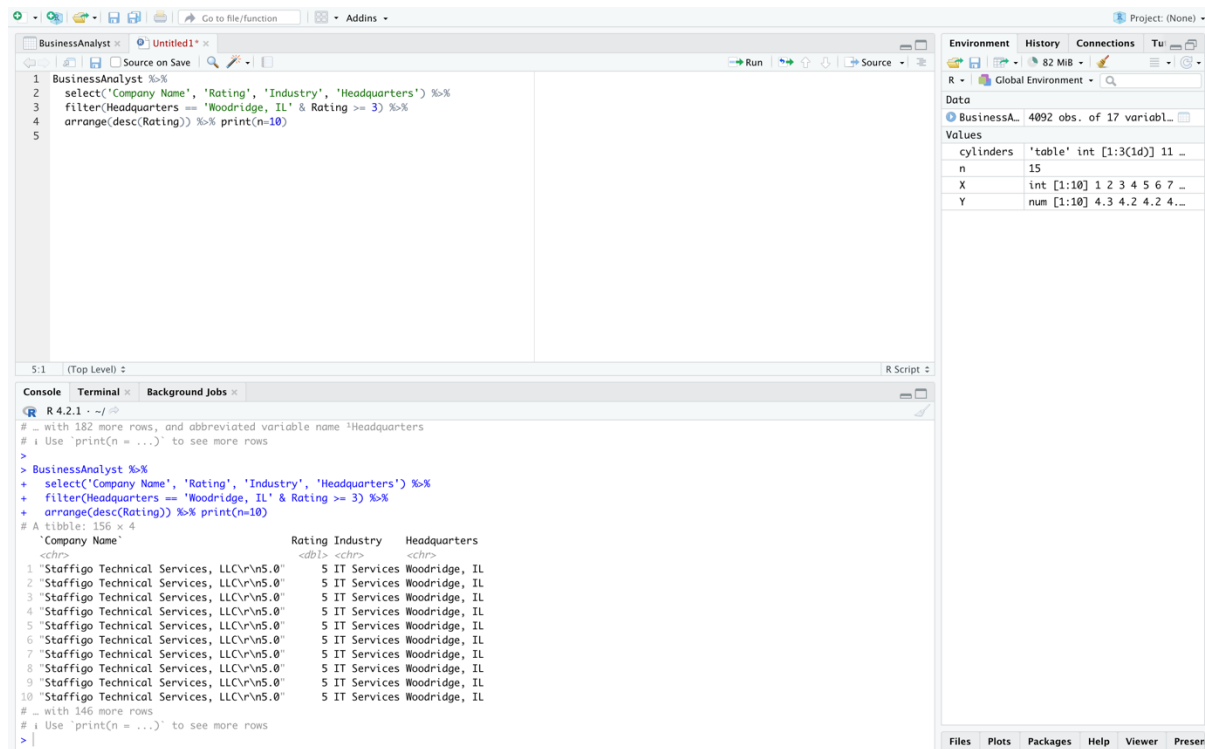


Image depicting the line of code used to call the top 10 companies in Woodridge, IL. Woodridge appears to be dominated by IT Services.

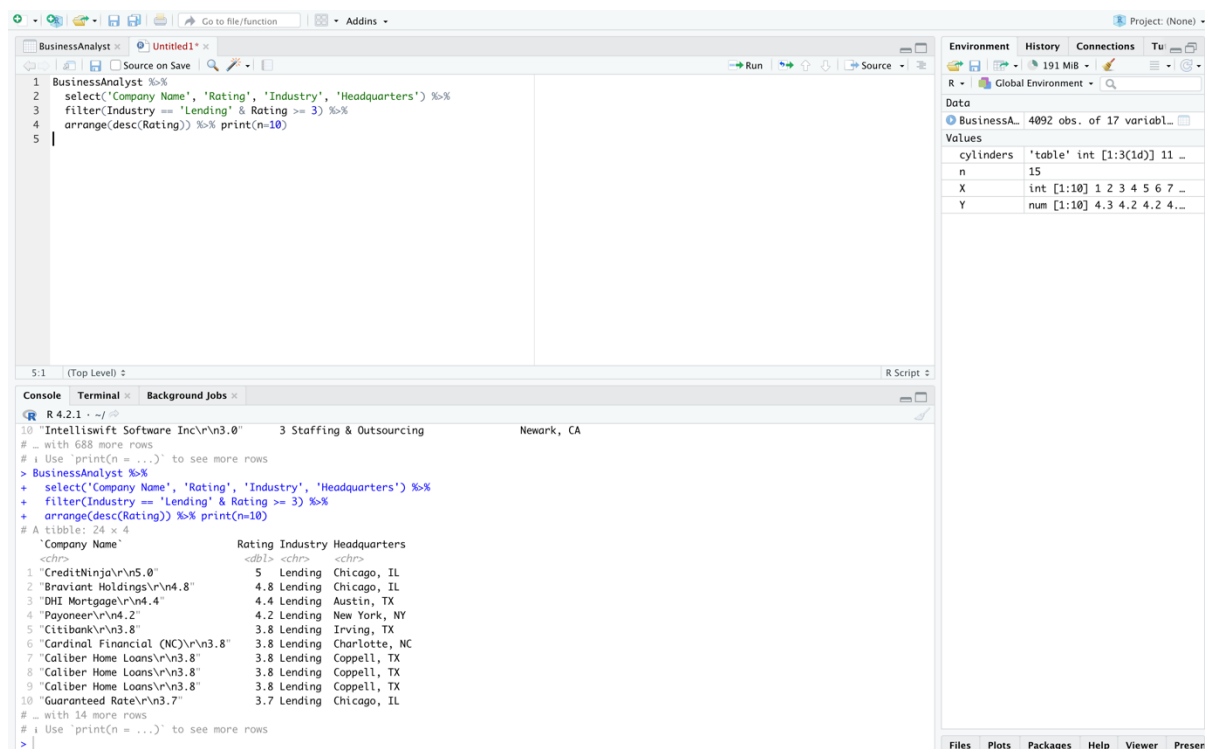


Image depicting the line of code used to call the top 10 Lending industry companies. There is a big variance in ratings here between 5 and 3.7.

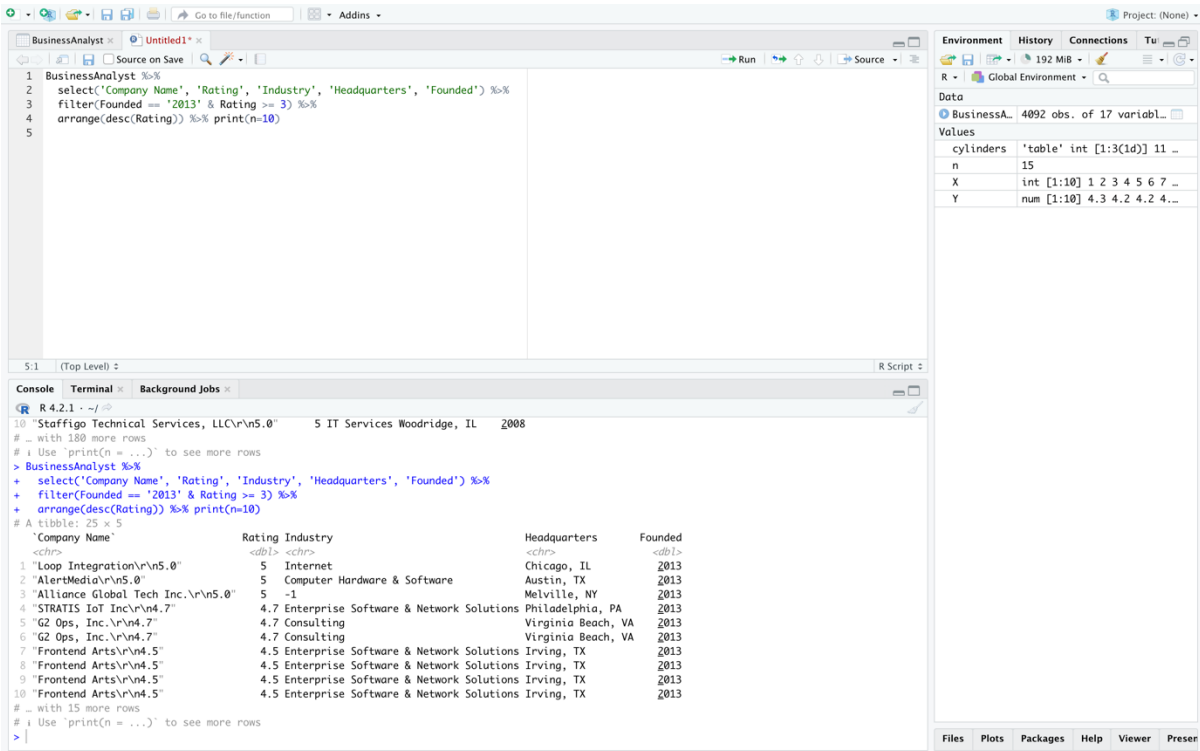


Image depicting the line of code used to call the top 10 companies founded in 2013.

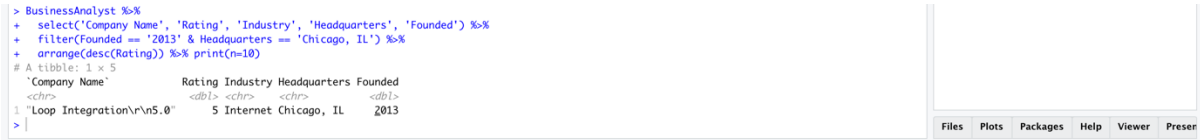


Image depicting the line of code used to call the top companies founded in 2013 and in Chicago. Loop Integration was the only company to meet these criteria.

By Gerard Utware

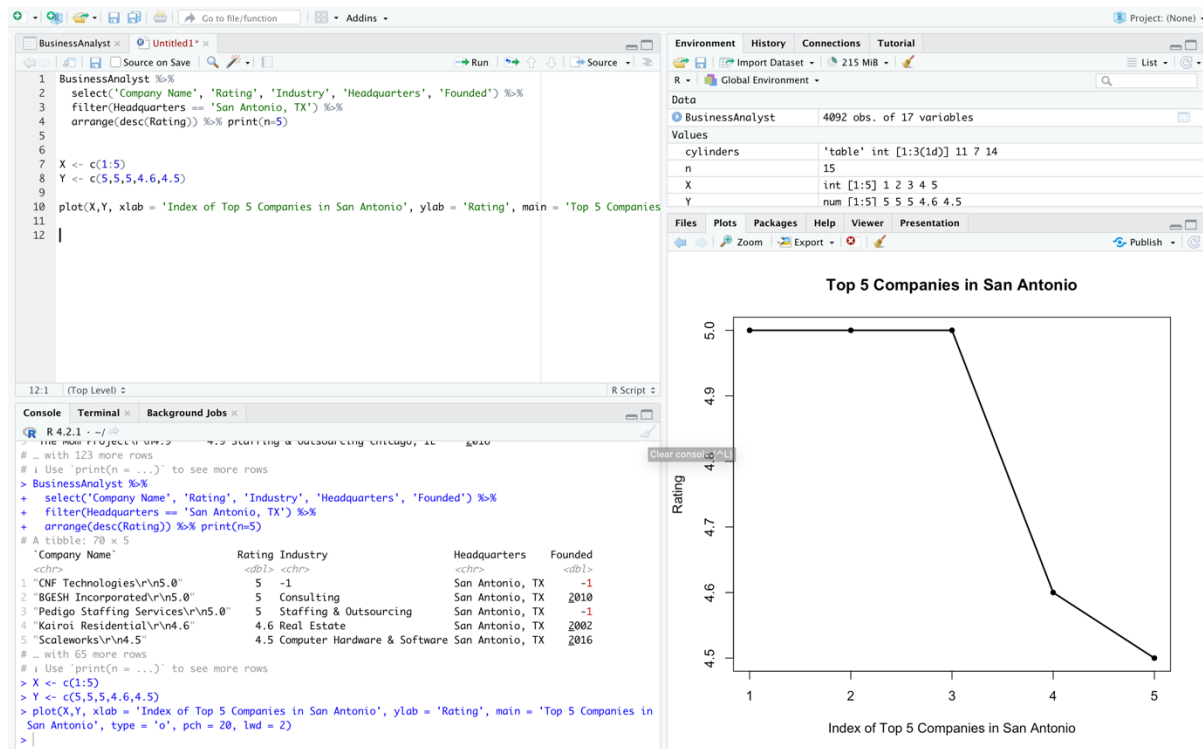


Image depicting the line of code used to call the line graph displaying the ratings of the top 5 companies that have their headquarters in San Antonio. Two companies have their founding year unknown.

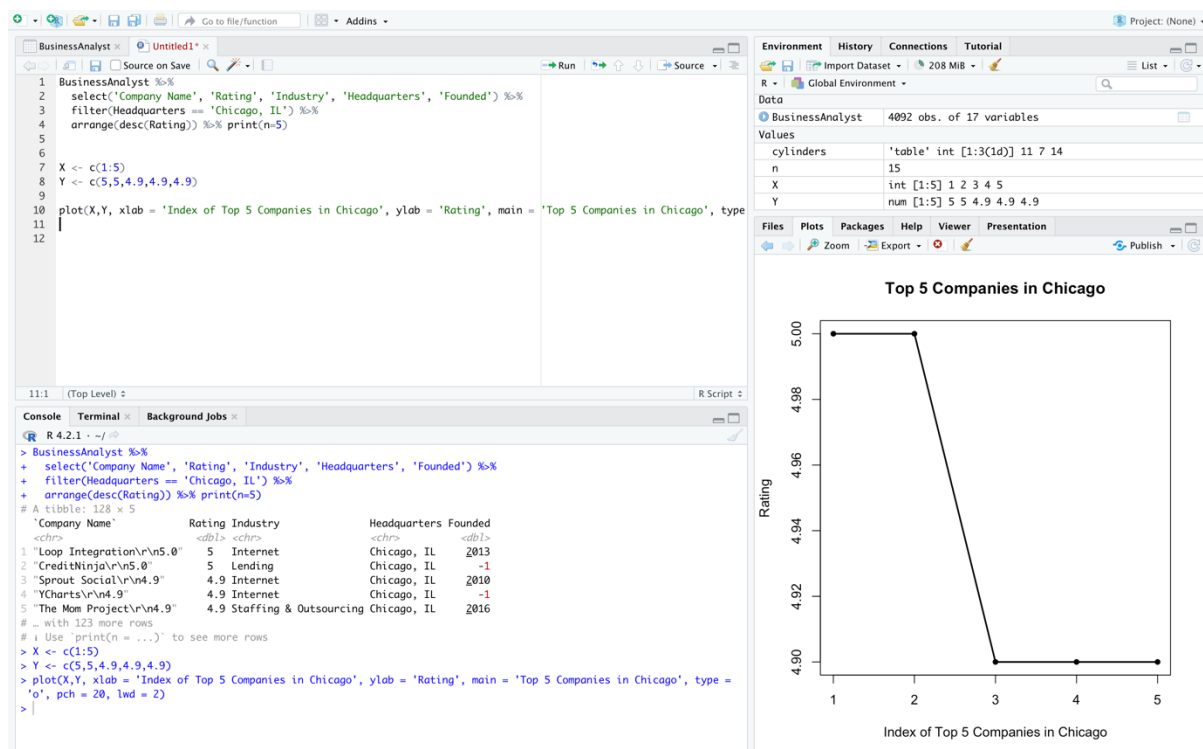


Image depicting the line of code used to call the line graph displaying the ratings of the top 5 companies that have their headquarters in Chicago. Two companies have their founding year unknown.

By Gerard Utoware

```
> BusinessAnalyst %>%
+   select('Company Name', 'Rating', 'Industry', 'Headquarters', 'Founded') %>%
+   filter(Headquarters == 'Philadelphia, PA') %>%
+   arrange(desc(Founded)) %>% print(n=5)
# A tibble: 45 x 5
  `Company Name`      Rating Industry      Headquarters      Founded
  <chr>              <dbl> <chr>              <chr>              <dbl>
1 "STRATIS IoT Inc\r\n4.7" 4.7 Enterprise Software & Network Solutions Philadelphia, PA    2013
2 "IntegriChain\r\n3.2"   3.2 IT Services      Philadelphia, PA    2005
3 "Seer Interactive\r\n3.9" 3.9 Internet         Philadelphia, PA    2002
4 "Seer Interactive\r\n3.9" 3.9 Internet         Philadelphia, PA    2002
5 "FreedomPay\r\n3.6"      3.6 Enterprise Software & Network Solutions Philadelphia, PA    2000
# ... with 40 more rows
# i Use `print(n = ...)` to see more rows
```

Image depicting the line of code used to call the 5 most recent companies in the Philadelphia area.

```
> BusinessAnalyst %>%
+   select('Company Name', 'Rating', 'Industry', 'Headquarters', 'Founded') %>%
+   filter(Headquarters == 'Zurich, Switzerland') %>%
+   arrange(desc(Founded)) %>% print(n=5)
# A tibble: 4 x 5
  `Company Name`      Rating Industry      Headquarters      Founded
  <chr>              <dbl> <chr>              <chr>              <dbl>
1 "Swiss Re\r\n3.8"      3.8 Insurance Agencies & Brokerages Zurich, Switzerland 1863
2 "UBS\r\n3.6"          3.6 Investment Banking & Asset Management Zurich, Switzerland 1862
3 "Chubb\r\n3.3"        3.3 Insurance Carriers Zurich, Switzerland 1792
4 "Chubb\r\n3.3"        3.3 Insurance Carriers Zurich, Switzerland 1792
> |
```

Image depicting the line of code used to call the most recent companies in Zurich. No new companies with their headquarters here in over 150 years according to the dataset.