**Task 13: Data Analysis using RStudio**

The original data was submitted in Excel format, and it consisted of 4,092 records and 17 fields.

**Task 1: Data preparation**

The first task was to open the file and check for duplicates and errors.

* On inspection, 400 rows were found to be misplaced and this were aligned appropriately.
* A total of 10,070 data points were recorded as -1 and these were all replaced with “Nothing”
* There were 30 duplicate records and hence 15 records were deleted.
* General formatting and preparation were done to the resultant 4,077 records.
* The ‘cleaned’ dataset (4,077 records) were loaded to RStudio.

**Task 2: Data exploration**

Data exploration (otherwise known as exploratory data analysis) gives an idea about the structure of the dataset like number of continuous or categorical variables and number of observations (rows).

For this dataset, I have carried out some basic examinations of the variables, and because most of the variables are categorical, not a lot of measures have been generated.

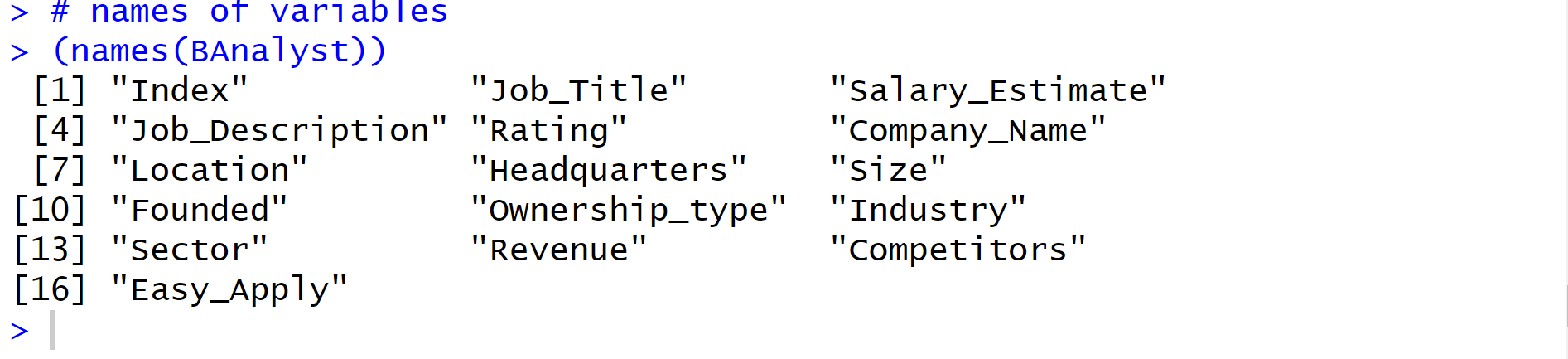
* **Dimension of the dataset (number of records and fields)**

> # dimension of the data

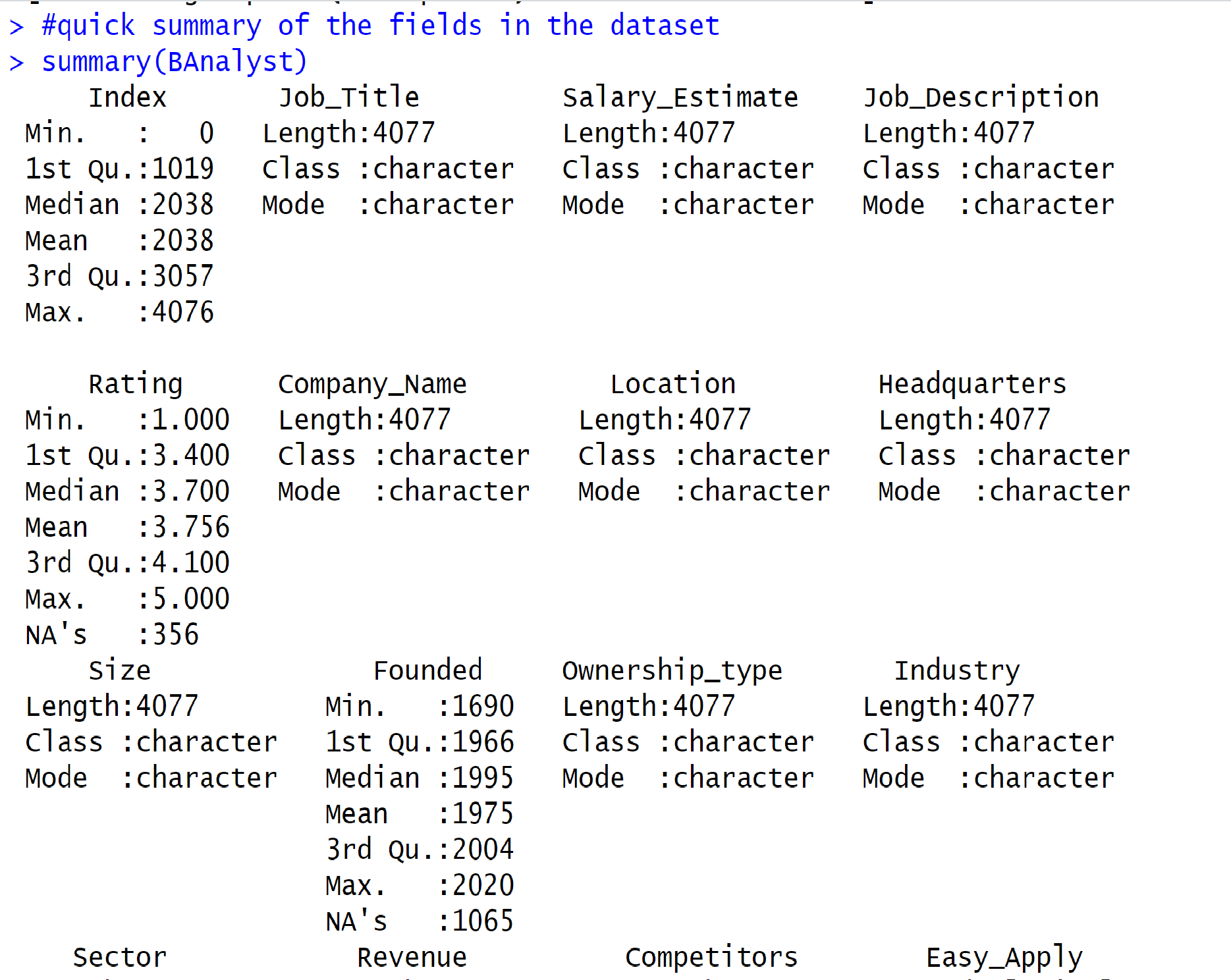
> dim(BAnalyst)

[1] 4077 16

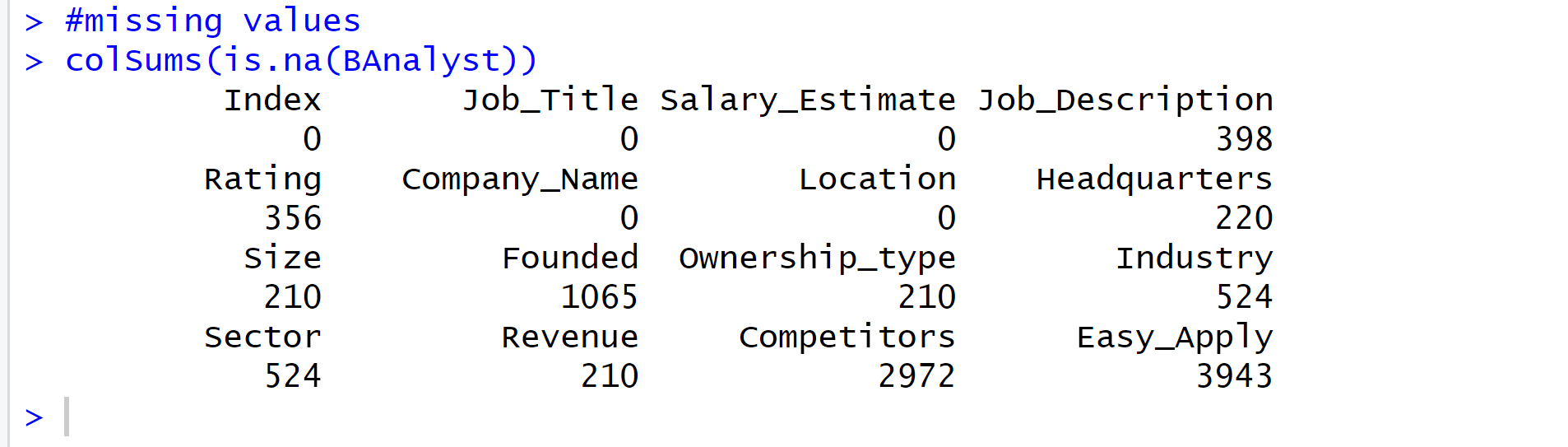
* **Names of the variables**

****

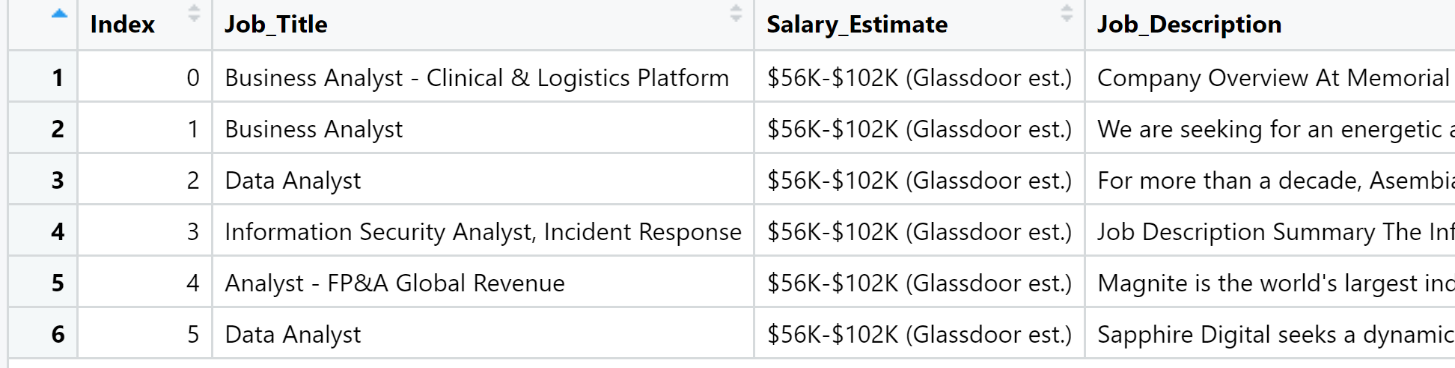
* **Quick Summary of the fields in the dataset**



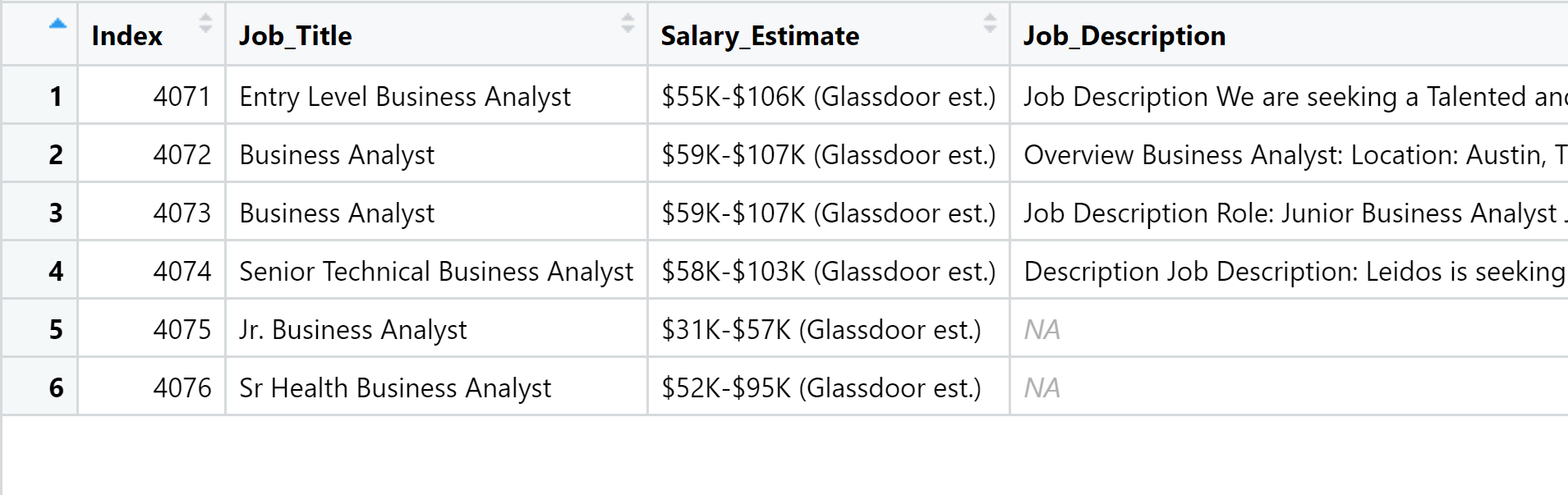
* **Number of missing values in each field in the dataset**

****

* **Head – display of the first 6 rows of the dataset**

****

* **Tail – display of the last rows of the dataset**

****

* **Histogram of the Rating**

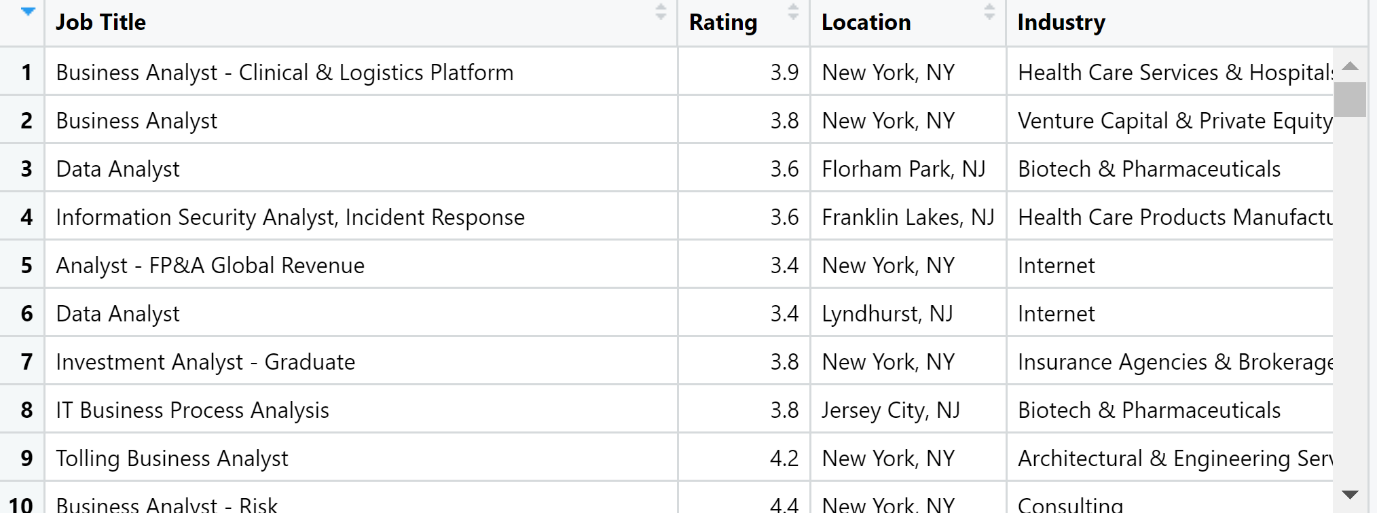
This shows the distribution of the Ratings in the dataset. As can be seen, the most frequent ratings for the companies a rating of 3 and above. Rating is the only numerical variable in the dataset so no histogram was produced for any other variable.

**A graph with a bar graph

Description automatically generated**

**Task 4: To display four columns: Job Title, Rating, Location, Industry**

The 4 columns are displayed below:



**Task 5: List of selected Top 20 variables:**

**5a. Top 20 Industries:**

|  | **Industry** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| **1** | IT Services | 647 |
| **2** | Staffing & Outsourcing | 510 |
| **3** | Computer Hardware & Software | 245 |
| **4** | Consulting | 234 |
| **5** | Health Care Services & Hospitals | 167 |
| **6** | Insurance Carriers | 162 |
| **7** | Investment Banking & Asset Management | 157 |
| **8** | Enterprise Software & Network Solutions | 127 |
| **9** | Banks & Credit Unions | 110 |
| **10** | Internet | 79 |
| **11** | Accounting | 78 |
| **12** | Advertising & Marketing | 77 |
| **13** | Aerospace & Defense | 71 |
| **14** | Biotech & Pharmaceuticals | 66 |
| **15** | Federal Agencies | 56 |
| **16** | Colleges & Universities | 53 |
| **17** | Lending | 41 |
| **18** | Energy | 37 |
| **19** | Architectural & Engineering Services | 29 |
| **20** | Financial Transaction Processing | 29 |

The table above shows the top 20 Industries in the dataset. As can be seen, IT Services was the most industry in demand, followed by Staffing and Outsourcing.

**5b. Top 20 Sectors:**

|  | **Sector** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| 1 | Information Technology | 1098 |
| 2 | Business Services | 894 |
| 3 | Finance | 362 |
| 4 | Insurance | 177 |
| 5 | Health Care | 167 |
| 6 | Manufacturing | 132 |
| 7 | Government | 85 |
| 8 | Accounting & Legal | 82 |
| 9 | Education | 72 |
| 10 | Aerospace & Defense | 71 |
| 11 | Retail | 71 |
| 12 | Biotech & Pharmaceuticals | 66 |
| 13 | Oil, Gas, Energy & Utilities | 63 |
| 14 | Media | 38 |
| 15 | Transportation & Logistics | 35 |
| 16 | Non-Profit | 29 |
| 17 | Consumer Services | 25 |
| 18 | Real Estate | 22 |
| 19 | Construction, Repair & Maintenance | 21 |
| 20 | Telecommunications | 21 |

The table above shows the top 20 Sectors. Information Technology and Business Services were the top two sectors in demand, and this is rightly aligned with the Industry demand.

|  | **Headquarter** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| 1 | New York, NY | 249 |
| 2 | San Diego, CA | 192 |
| 3 | Woodridge, IL | 167 |
| 4 | Chicago, IL | 135 |
| 5 | Houston, TX | 123 |
| 6 | Jacksonville, FL | 91 |
| 7 | San Antonio, TX | 70 |
| 8 | Dallas, TX | 67 |
| 9 | Los Angeles, CA | 67 |
| 10 | Austin, TX | 64 |
| 11 | Irving, TX | 62 |
| 12 | Atlanta, GA | 58 |
| 13 | Tampa, FL | 54 |
| 14 | San Jose, CA | 48 |
| 15 | Philadelphia, PA | 47 |
| 16 | San Francisco, CA | 47 |
| 17 | London, United Kingdom | 42 |
| 18 | Phoenix, AZ | 42 |
| 19 | Princeton, NJ | 42 |
| 20 | Santa Clara, CA | 40 |

The table above shows that most of the companies are based in New York, followed by San Diego, CA.

This is not unexpected, since New York is also the headquarters of United Nations.

**Task 6: Jobs based on Rating**

**6a. Top 15 jobs based on Rating:**

|  | **Job Title** | **Rating** |
| --- | --- | --- |
|  |  |  |
| **1** | Business Analyst | 5 |
| **2** | Business Process Analyst | 5 |
| **3** | Business Analyst - Commercial Lending | 5 |
| **4** | Business Systems Analyst | 5 |
| **5** | System/Business Analyst | 5 |
| **6** | Business Analyst - Technical | 5 |
| **7** | Business Development (Institutional Investors) | 5 |
| **8** | Business Intelligence Analyst | 5 |
| **9** | Technical Business Analyst | 5 |
| **10** | Entry Level Business Analyst | 5 |
| **11** | Business Data Analyst | 5 |
| **12** | Business Operations Analyst | 5 |
| **13** | Business Analyst Trainee | 5 |
| **14** | Data Analyst | 5 |
| **15** | Management Consulting - Business Analyst | 5 |

From the above table, Business (related) Analysts is the most demanded job, based on the rating.

**6b. Top 15 jobs based on Rating under ‘Consulting’ industry:**

|  | **Job Title** | **Industry** | **Rating** |
| --- | --- | --- | --- |
|  |  |  |  |
| **1** | System/Business Analyst | Consulting | 5.0 |
| **2** | Management Consulting - Business Analyst | Consulting | 5.0 |
| **3** | Business Analyst III | Consulting | 5.0 |
| **4** | Business Analyst II | Consulting | 5.0 |
| **5** | Business Analyst/QA | Consulting | 5.0 |
| **6** | Scrum Quality Analyst | Consulting | 4.9 |
| **7** | Business Systems Analyst V | Consulting | 4.9 |
| **8** | Senior Business Analyst/CPQ Consultant | Consulting | 4.7 |
| **9** | Cost Analyst | Consulting | 4.7 |
| **10** | Program Management Analyst II | Consulting | 4.7 |
| **11** | Configuration Management Analyst | Consulting | 4.7 |
| **12** | Technical Lead / Business Analyst | Consulting | 4.6 |
| **13** | Business Systems Analyst - Salesforce | Consulting | 4.6 |
| **14** | Commerce Business Analyst | Consulting | 4.6 |
| **15** | IT Applications - Business Analyst- Reporting | Consulting | 4.6 |

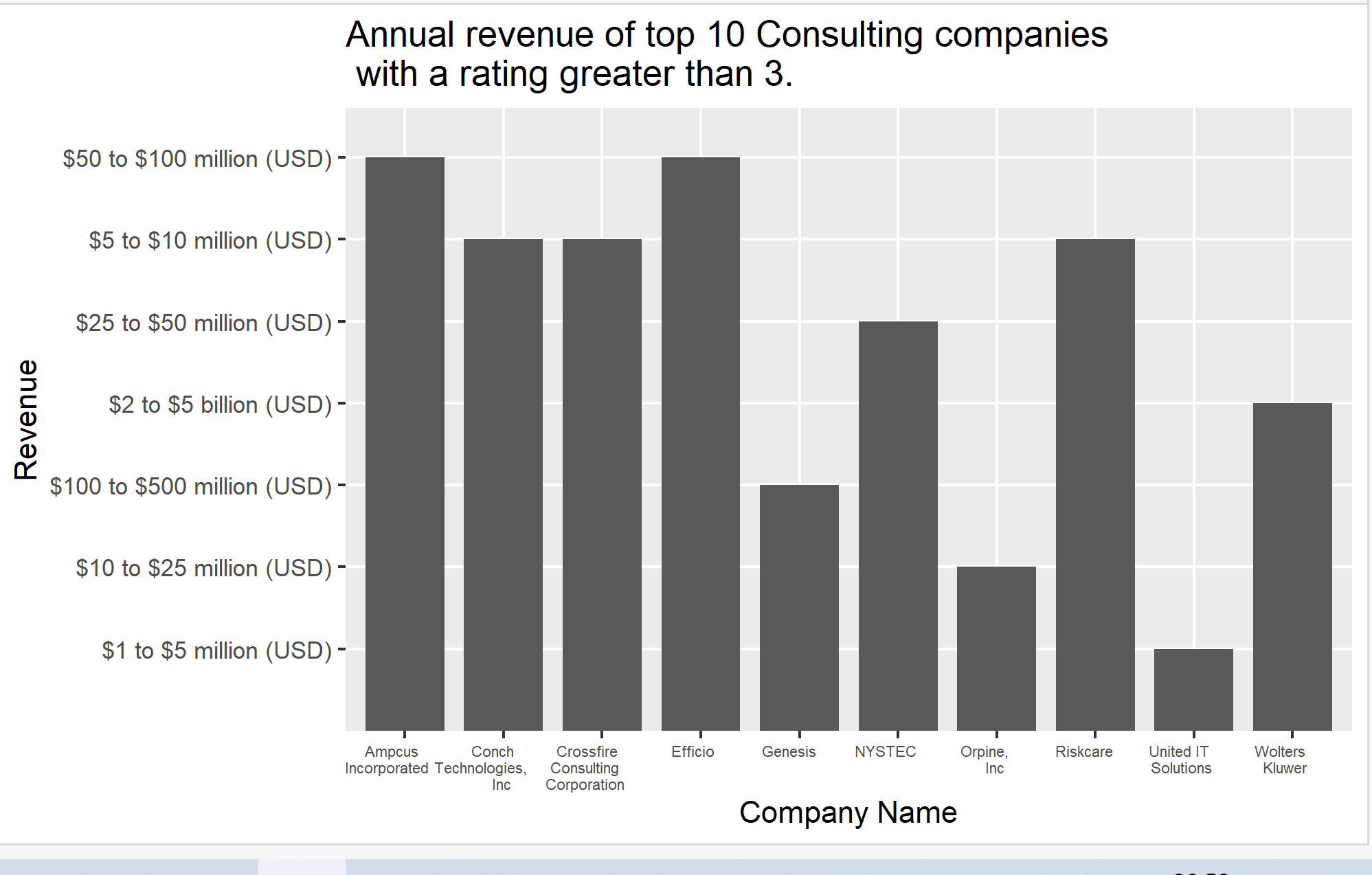
**6c: Bottom 15 jobs based on Rating:**

|  | **Job Title** | **Rating** |
| --- | --- | --- |
|  |  |  |
| **1** | Data Analyst III (Healthcare Analytics) | 2.3 |
| **2** | Procure to Pay Analyst | 2.3 |
| **3** | BI Analyst/ETL Developer | 2.2 |
| **4** | Applications Delivery Analyst | 2.1 |
| **5** | Sr Business System Analyst | 2.1 |
| **6** | Business Analyst (Volunteer) | 2.0 |
| **7** | Business Intelligence Analyst/ Credit Risk Analyst | 2.0 |
| **8** | SAP HCM - Business Analyst | 2.0 |
| **9** | Business Analyst, Senior | 2.0 |
| **10** | Program Analyst - All Levels | 2.0 |
| **11** | Claims Configuration Business Process Analyst II | 2.0 |
| **12** | Data Analyst MDM | 1.0 |
| **13** | Finance Analyst, Business Operations | 1.0 |
| **14** | Entry Level Oracle ERP Analysts | 1.0 |
| **15** | Hiring a Business Analyst | 1.0 |

From the above table, other related Analysts is the least demanded job, based on the rating.

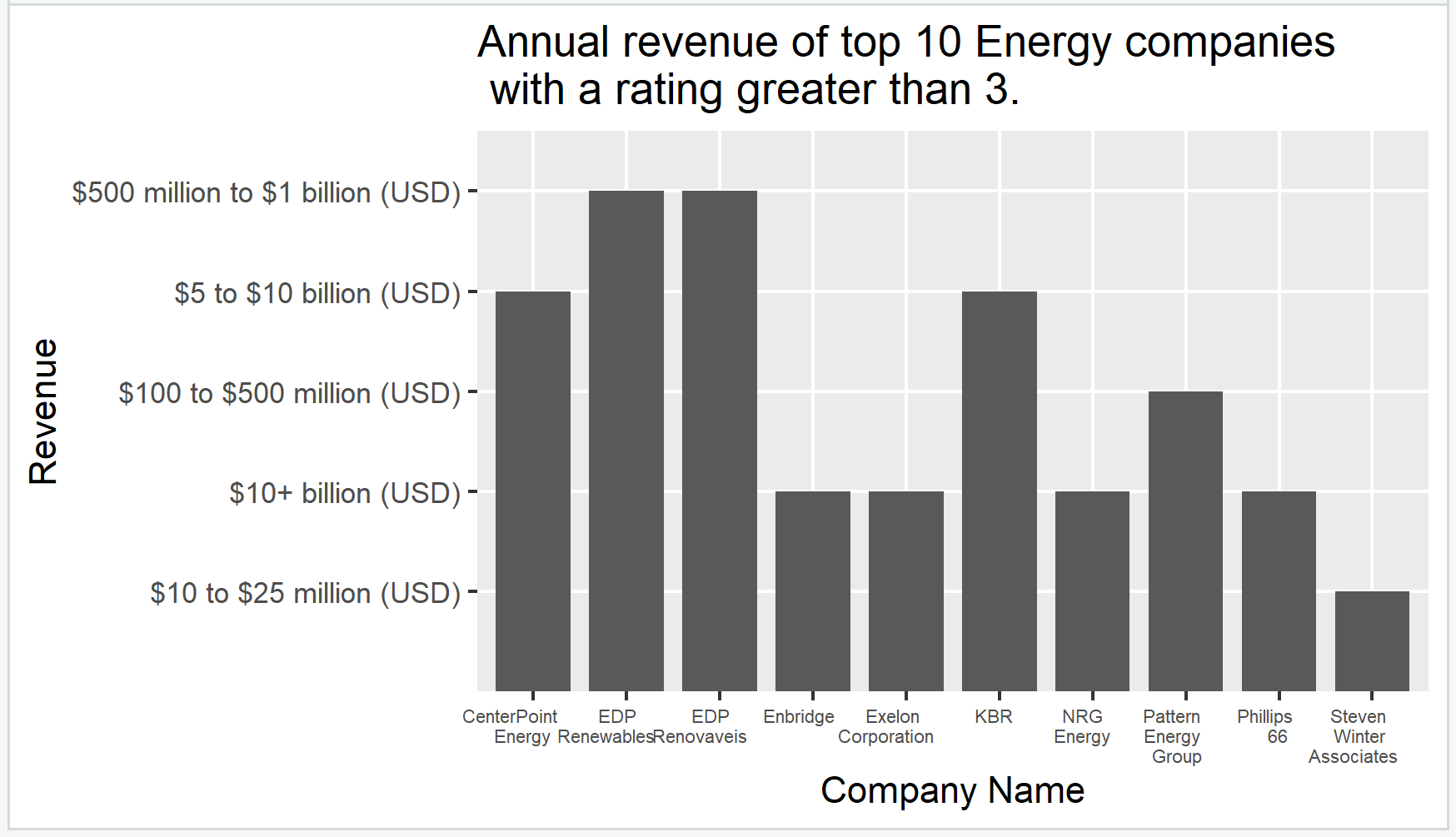
**Task 7:**

**7a. Top 10 Consulting Industry Companies with a rating greater than 3 (by annual Revenue):**

****

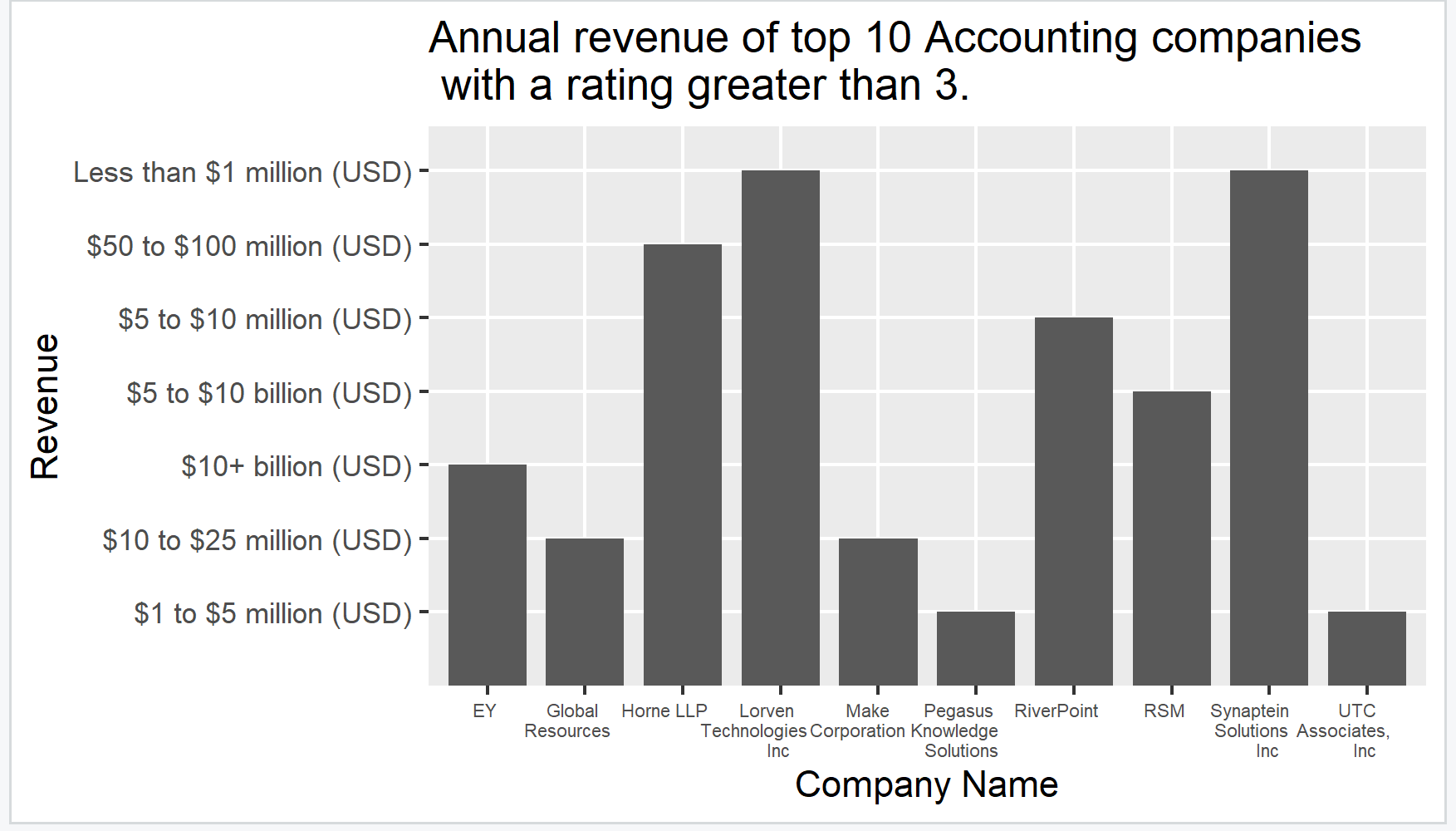
The graph above shows annual revenue of the top 10 Consulting companies which have a rating greater than 3.

**7b. Top 10 Energy Industry Companies with a rating greater than 3 (by annual Revenue):**

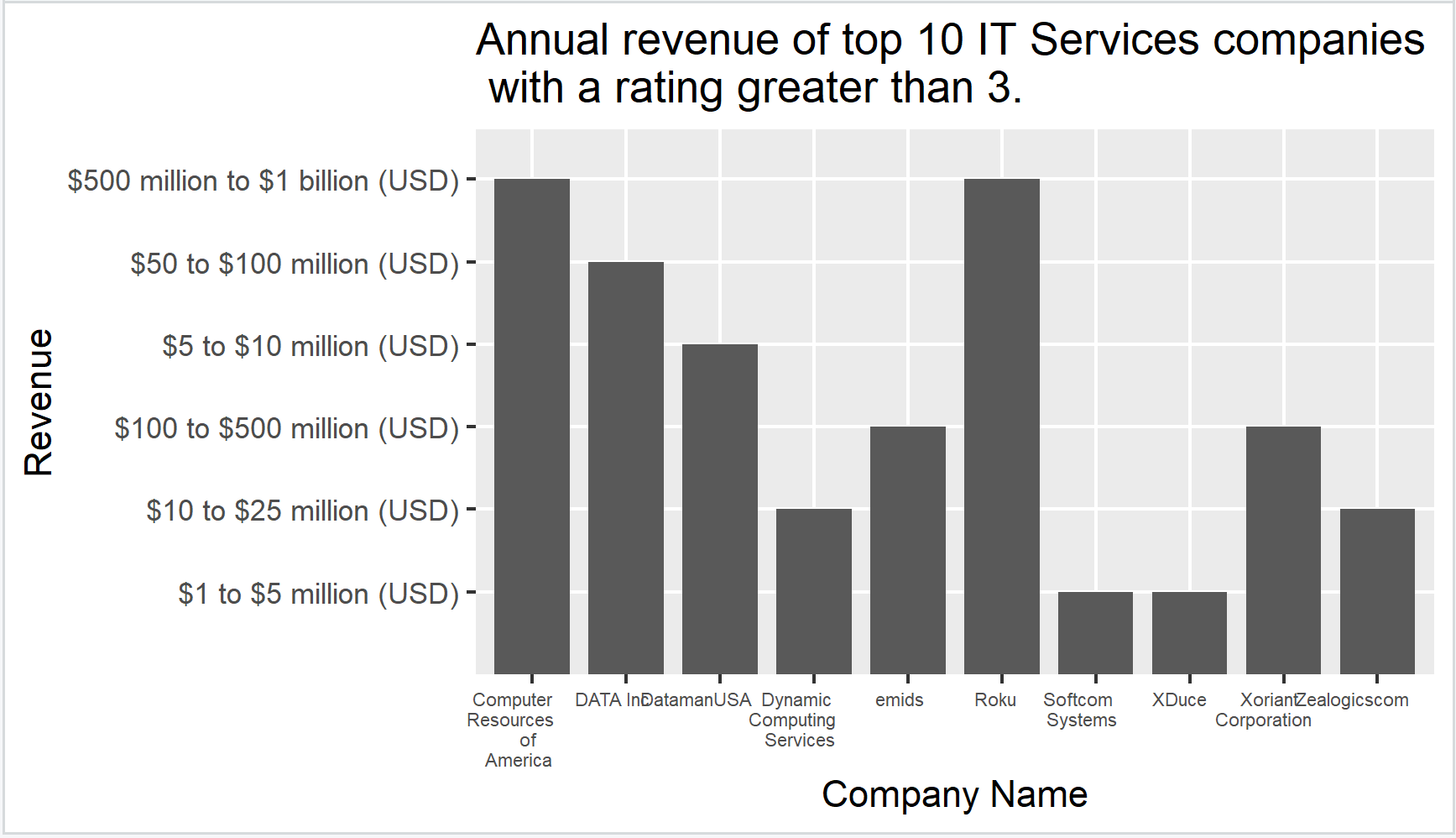
****

The graph above shows annual revenue of the top 10 Energy companies which have a rating greater than 3.

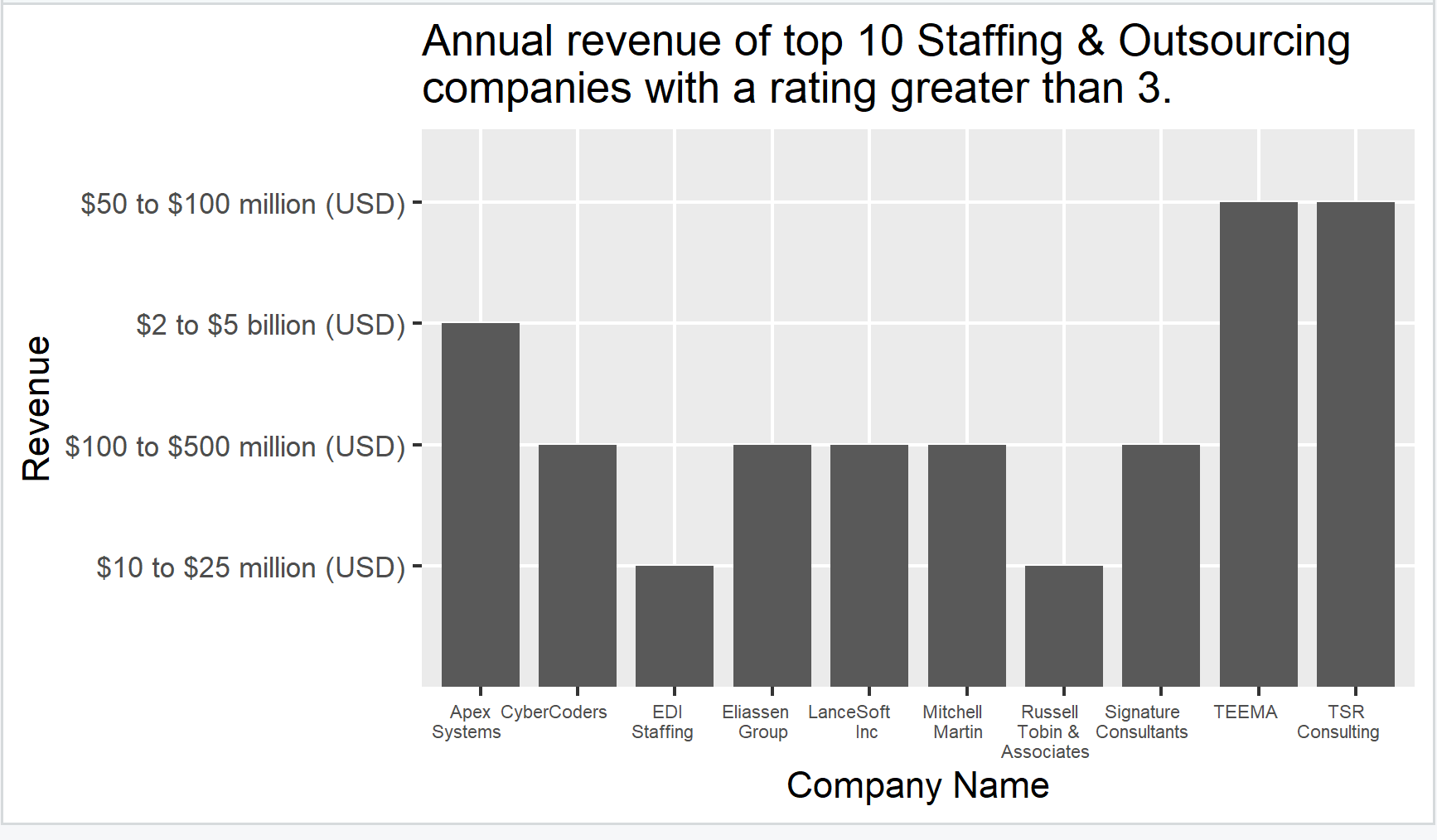
**7c. Top 10 Accounting Industry Companies with a rating greater than 3 (by annual Revenue):**



**8a. Top 10 IT Services Industry Companies with a rating greater than 3 (by annual Revenue):**

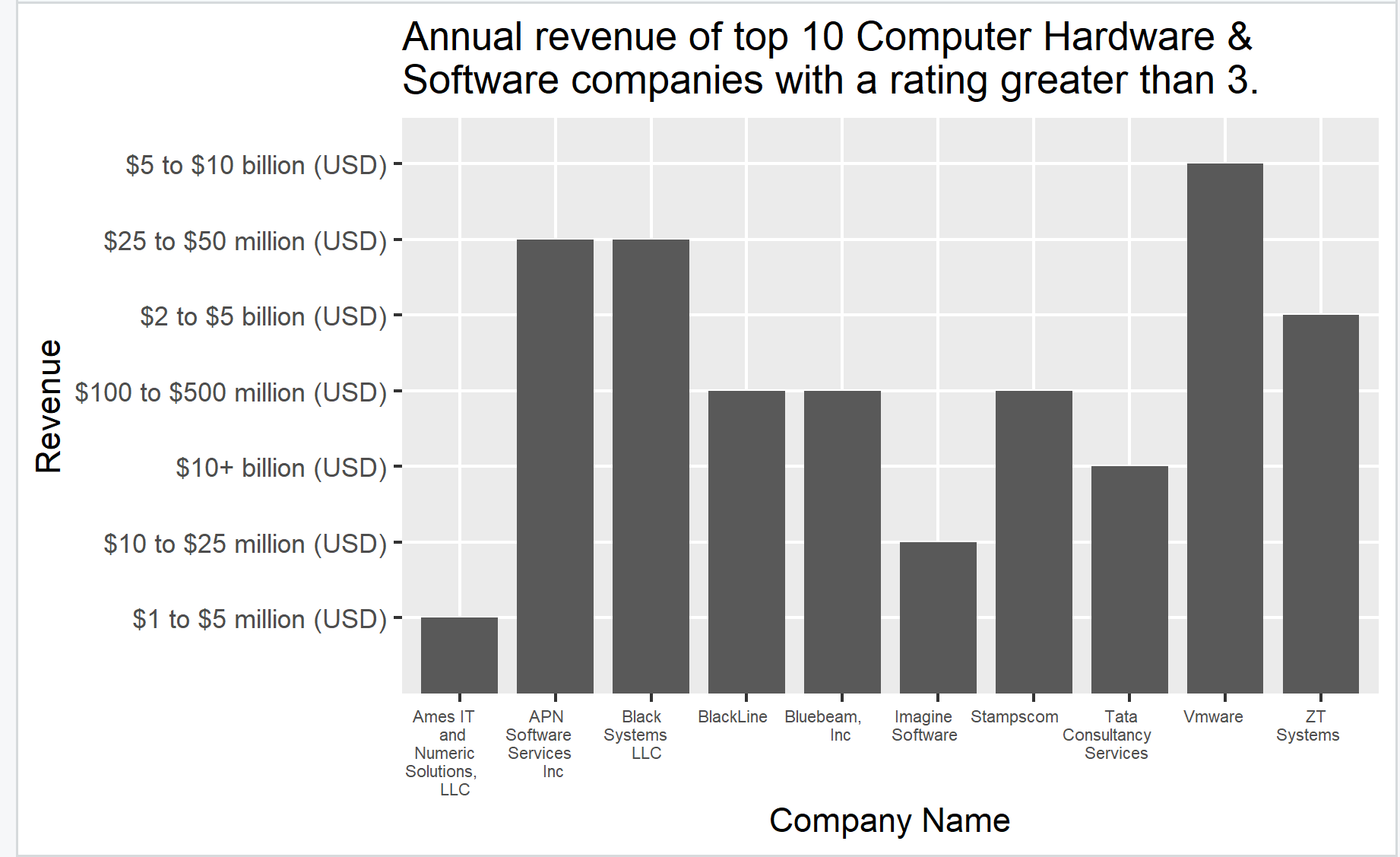


**8b. Top 10 Staffing & Outsourcing Industry Companies with a rating greater than 3 (by annual Revenue):**

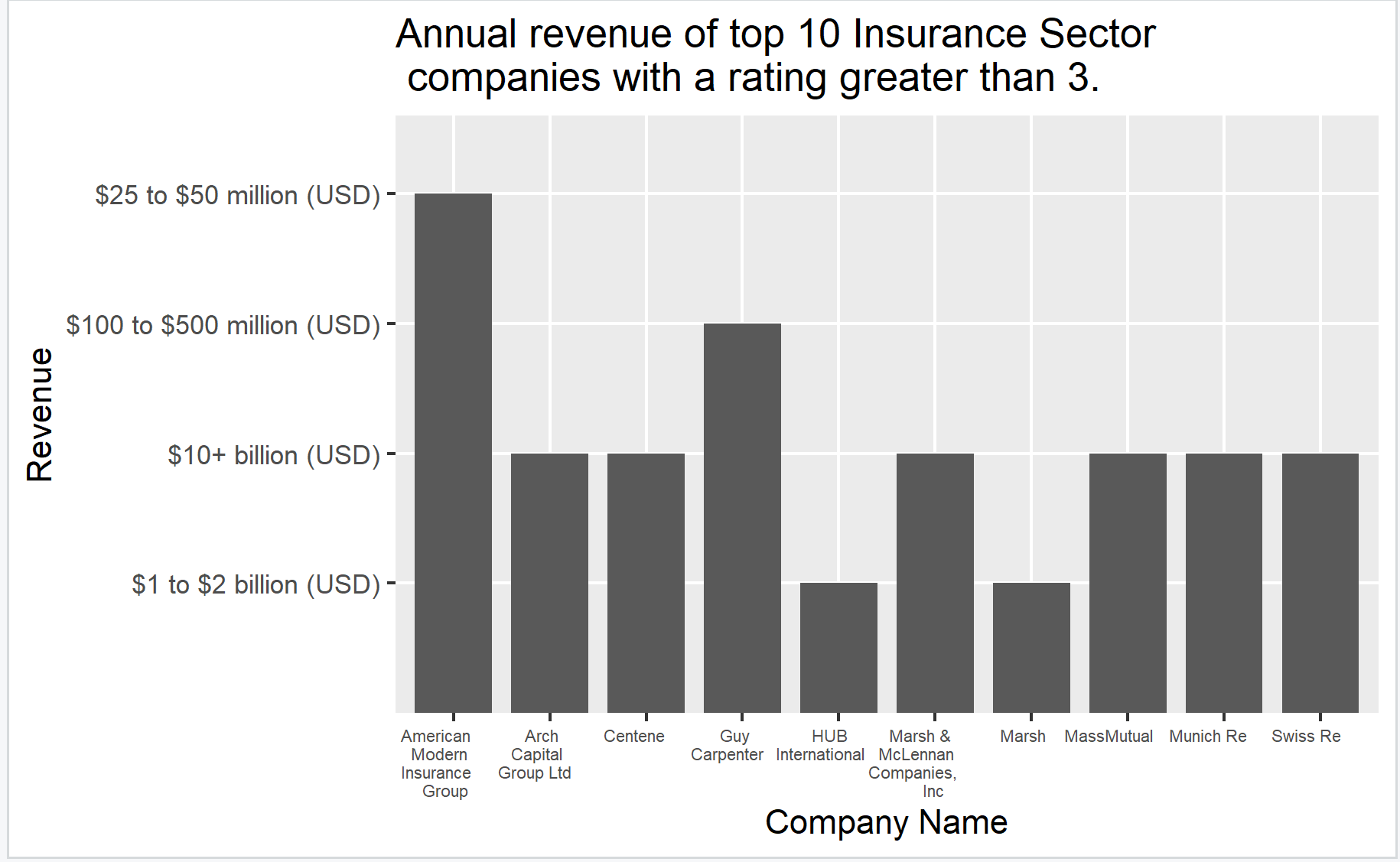


**8c. Top 10 Computer Hardware & Software Industry Companies with a rating greater than 3**

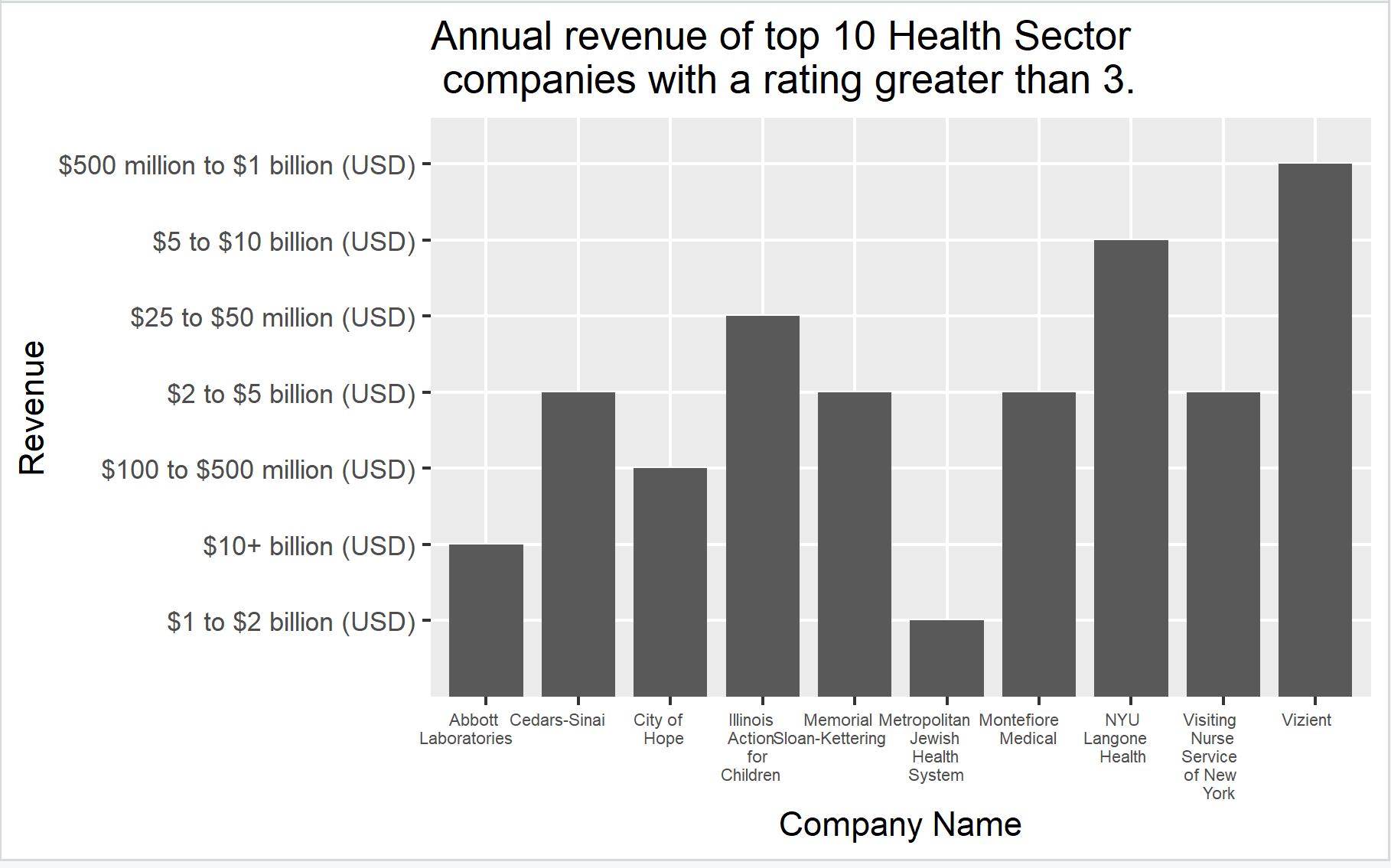
**(by annual Revenue):**



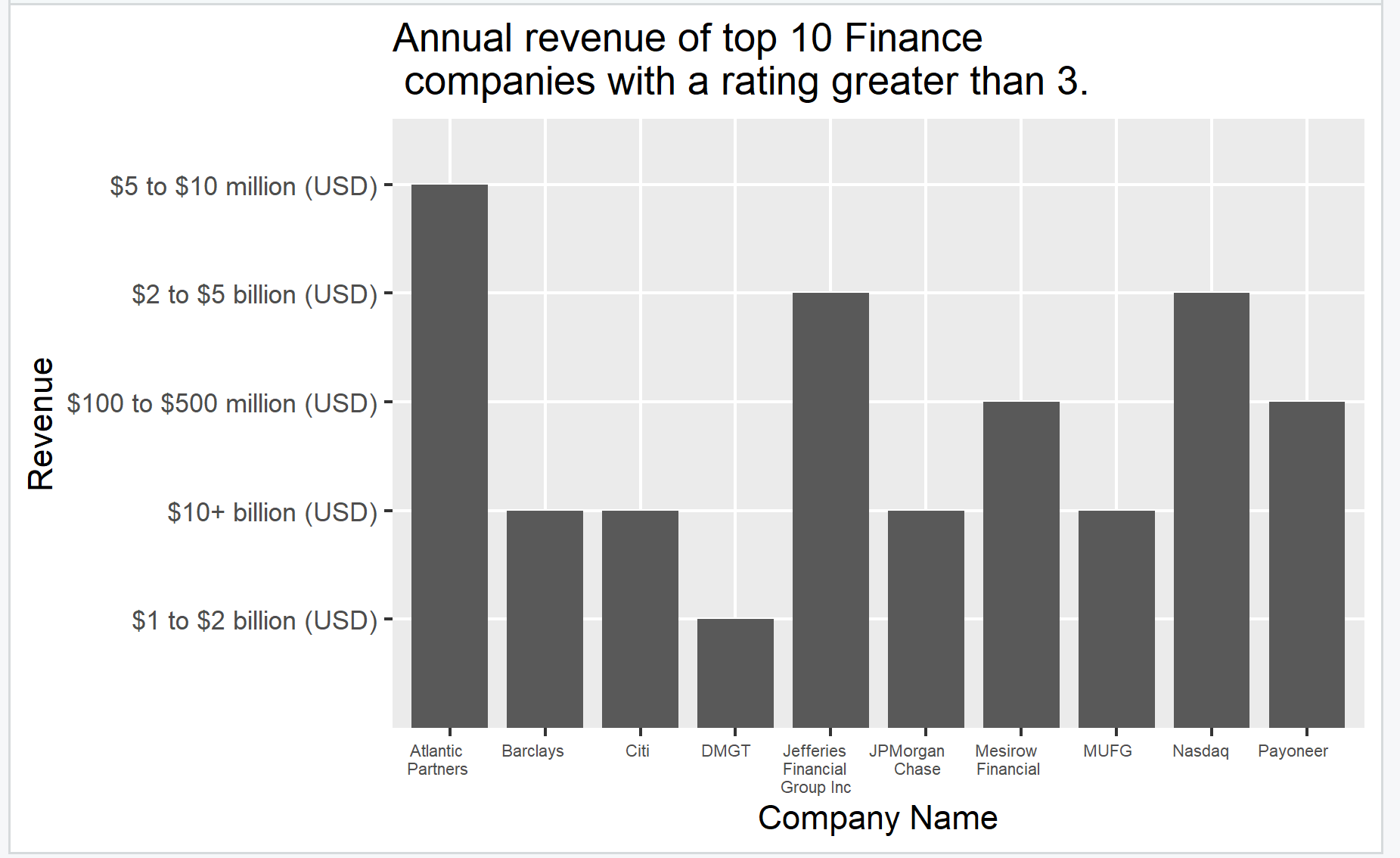
**8d. Top 10 Insurance Sector Companies with a rating greater than 3 (by annual Revenue):**



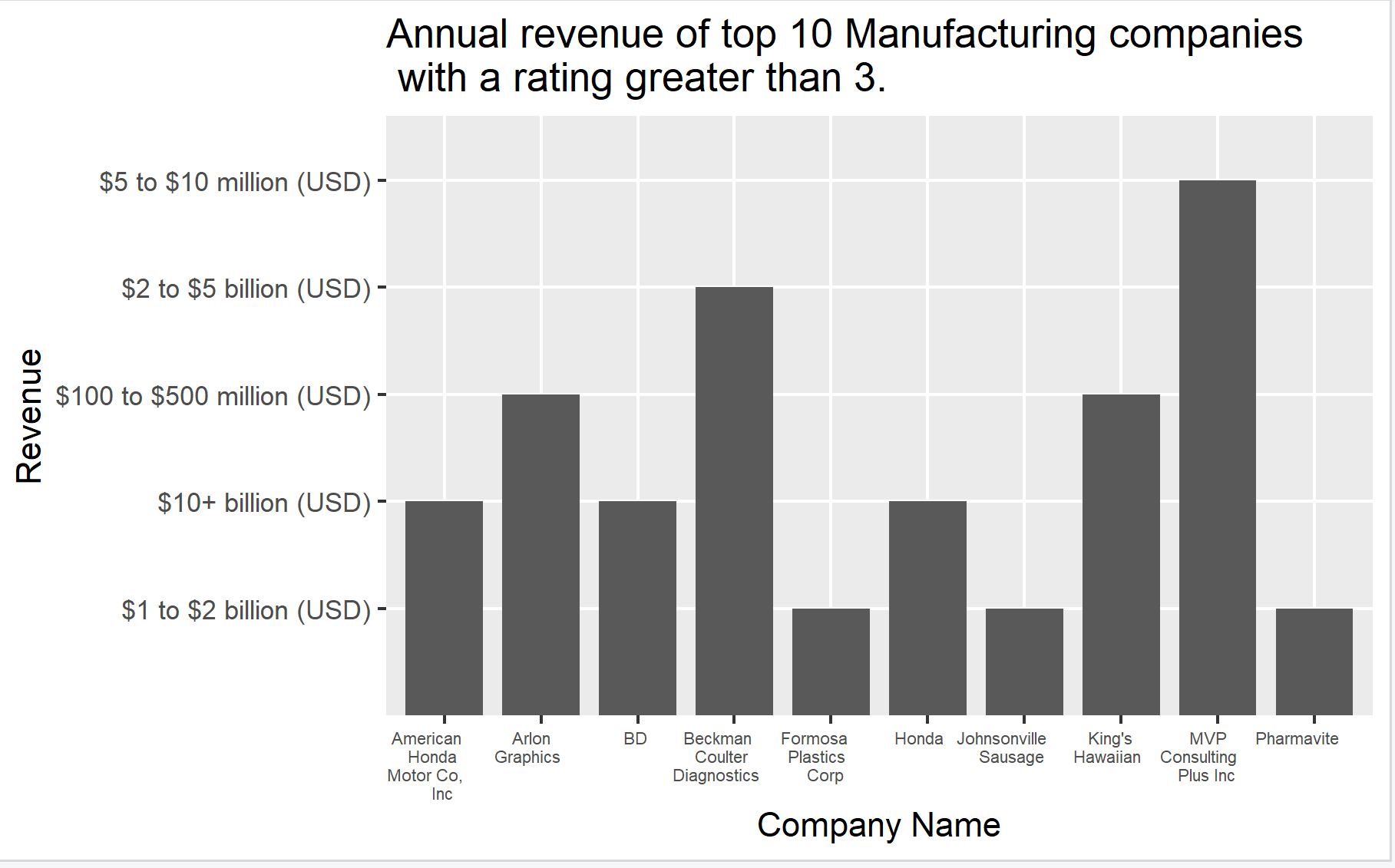
**8e. Top 10 Health Care Sector Companies with a rating greater than 3 (by annual Revenue):**

****

**8f. Top 10 Finance Sector Companies with a rating greater than 3 (by annual Revenue):**

****

**8g. Top 10 Manufacturing Sector Companies with a rating greater than 3 (by annual Revenue):**

****

**8i: Top 10 Locations by number of Companies**

|  | **Location (City)** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| **1** | Jacksonville, FL | 390 |
| **2** | San Diego, CA | 345 |
| **3** | Houston, TX | 340 |
| **4** | Austin, TX | 328 |
| **5** | San Antonio, TX | 324 |
| **6** | Chicago, IL | 268 |
| **7** | New York, NY | 247 |
| **8** | Phoenix, AZ | 195 |
| **9** | Philadelphia, PA | 178 |
| **10** | Dallas, TX | 176 |

**8j: Top 5 Employee Size by number of Companies**

|  | **Employee Size** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| **1** | 10000+ employees | 845 |
| **2** | 51 to 200 employees | 772 |
| **3** | 1001 to 5000 employees | 720 |
| **4** | 1 to 50 employees | 504 |
| **5** | 201 to 500 employees | 403 |

**8k: Top 5 Ownership Type by number of Companies**

|  | **Type** | **Number of Companies** |
| --- | --- | --- |
|  |  |  |
| **1** | Company - Private | 2247 |
| **2** | Company - Public | 1042 |
| **3** | Subsidiary or Business Segment | 185 |
| **4** | Nonprofit Organization | 122 |
| **5** | Government | 93 |

**Conclusion (some Insights)**

From the dataset, the following insights or conclusion can made:

1. IT Services was the top industry in demand.
2. New York was the centre of business and hence most of the companies are headquartered there. UN is headquartered in New York and that may also be a factor.
3. The top job in demand was Business related Analysis.
4. Jacksonville, FL was the top location of the companies.
5. The top companies are all located in the US.
6. Companies employing 10,000 and above employees were the most common, thus illustrating the size of those conglomerates.
7. The top two ownership types were private companies and private companies.