Wordcloud

Jawaid Hakim

2022-10-17

Contents

1 Load library

```
library(DBI)
library(RMySQL)
library(RMariaDB)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(knitr)
library(wordcloud)
```

```
## Loading required package: RColorBrewer
```

```
library(RColorBrewer)
```

2 Data

Setup database connection parameters.

```
# connection params
user <- 'guest'
pw <- 'guestpass'
hostname <- 'cunyspsds.c5iiratvieki.us-east-1.rds.amazonaws.com'
```

Connect to database and list tables.

```
projectDb <- dbConnect(MariaDB(), user='guest', password=pw,
                       dbname='Project3', host=hostname)

# preview the tables
dbListTables(projectDb)
```

```
## [1] "EDUCATION"           "EDUCATION_IN_DEMAND" "SKILL"
## [4] "SKILL_IN_DEMAND"     "SOURCE"
```

Load skills data.

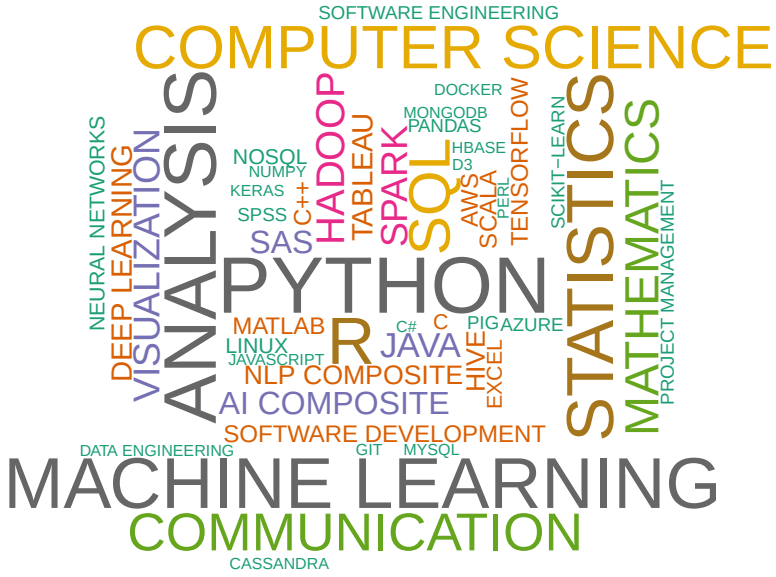
```
qry <- "SELECT * FROM SKILL_IN_DEMAND;"

# store the results as a dataframe
rs <- dbSendQuery(projectDb, qry)

skills <- dbFetch(rs)

dbClearResult(rs) # clear the result
```

Summarize skill counts.

```
skills_summary <- skills %>%
                    group_by(SKILL_KEYWORD) %>%
                    summarise(TOTAL = sum(COUNT))
```

Generate skills wordcloud.

```
set.seed(1234)
wordcloud::wordcloud(words = skills_summary$SKILL_KEYWORD,
                     freq = skills_summary$TOTAL,
                     min.freq = 100,
                     max.words = 50,
                     random.order = FALSE,
                     random.color = FALSE,
                     rot.per = 0.25,
                     colors = brewer.pal(8, "Dark2"),
                     scale = c(2.5, 0.40))
```


SOFTWARE ENGINEERING
COMPUTER SCIENCE
DOCKER
MONGODB
PANDAS
HBASE
D3
NOSQL
NUMPY
KERAS
SPSS
SAS
C++
HADOOP
TABLEAU
SPARK
SQL
AWS
SCALA
PERL
TENSORFLOW
SCIKIT-LEARN
STATISTICS
MATHEMATICS
PROJECT MANAGEMENT
NEURAL NETWORKS
DEEP LEARNING
VISUALIZATION
ANALYSIS
PYTHON
MATLAB
R
C#
C
PIG
AZURE
LINUX
JAVA
JAVASCRIPT
HIVE
EXCEL
NLP COMPOSITE
AI COMPOSITE
SOFTWARE DEVELOPMENT
DATA ENGINEERING
GIT
MYSQL
MACHINE LEARNING
COMMUNICATION
CASSANDRA