Explore and Mine Data

Code

Analytical Query I

Top five journals with the most articles published in them for the time period.

Hide

```
journal_article_count <- function(){
  dbGetQuery(mysqlconnection, 'SELECT Journal_Id,
                                      Title,
                                      Publish_year,
                                      SUM(Article_count)
                               FROM PracticumII.Fact_Journal
                               WHERE Publish_year = 1976
                               GROUP BY 1, 2, 3
                               ORDER BY 4 DESC
                               LIMIT 5')
}
journal_article_count()
```

Journal_Id <chr>	Title <chr>
444	The Journal of pharmacy and pharmacology
11	Biochimica et biophysica acta
119	The Journal of biological chemistry
48	Comparative biochemistry and physiology. A, Comparative physiology
567	Annales de l'anesthesiologie francaise

5 rows | 1-3 of 4 columns

With the fact table, if someone is interested in knowing the top journal by article count during a period of time, they can just change the filter in the where clause. This can be done by someone with limited SQL knowledge. This information is better represented by a table, since table contains more information about journal title, and the time period.

Analytical Query II

Number of articles per journal per year broken down by quarter.

Hide

```
articles_per_journal <- function(){
  dbGetQuery(mysqlconnection, 'SELECT Journal_Id,
                                      Title,
                                      Publish_year,
                                      IFNULL(SUM(CASE WHEN Publish_Quarter = "Q1" THEN A
rticle_count END), 0) as "Q1",
                                      IFNULL(SUM(CASE WHEN Publish_Quarter = "Q2" THEN A
rticle_count END), 0) as "Q2",
                                      IFNULL(SUM(CASE WHEN Publish_Quarter = "Q3" THEN A
rticle_count END), 0) as "Q3",
                                      IFNULL(SUM(CASE WHEN Publish_Quarter = "Q4" THEN A
rticle_count END), 0) as "Q4"
                               FROM PracticumII.Fact_Journal
                               GROUP BY 1, 2, 3')
}
articles_per_journal <- articles_per_journal()
articles_per_journal
```

Journal_Id <chr>
1
1
1
1
10
10
10
10
100
100

1-10 of 4,730 rows | 1-1 of 7 columns Previous **1** 2 3 4 5 6 … 100 Next

Hide

```
library(data.table)
library(ggplot2)
plot_articles_per_journal <- function(journal_title, year) {
  df <- articles_per_journal[articles_per_journal$Title == journal_title
                             & articles_per_journal$Publish_year == year, c("Q1", "Q2",
"Q3", "Q4")]
  df_t <- transpose(df)
  df_t$Quarter <- c("Q1", "Q2", "Q3", "Q4")
  p <-ggplot(df_t, aes(x=Quarter, y=V1)) + geom_bar(stat="identity")
  p + labs(title=journal_title,
        x ="Quarter", y = "Count")
}
plot_articles_per_journal("Biochemical medicine", 1975)
```