Practicum II - Dylan Horgan

```
library("XML")

path <- "/Users/dyhorgangmail.com/sqlite/"
xmlFile <- "pubmed_sample.xml"
xmlDoc <- xmlParse(file = paste(path, xmlFile, sep="/"), validate=T)
```

```
library(RMySQL)
```

```
## Loading required package: DBI
```

```
db_user<-'admin'
db_password<-'qwertyuiop'
db_name<-'PracticumIIDb'
db_host<-'practicumiidb.c53dkg2tawlv.us-east-1.rds.amazonaws.com'
db_port<-3306

dbcon <- dbConnect(MySQL(), user=db_user, password=db_password, name=db_name, host=db_host, port=db_por
```

```
DROP DATABASE IF EXISTS practicumIIdb;
```

```
CREATE DATABASE practicumIIdb;
```

```
USE practicumIIdb;
```

```
DROP TABLE IF EXISTS Articles
```

```
DROP TABLE IF EXISTS Journals
```

```
DROP TABLE IF EXISTS History
```

```
DROP TABLE IF EXISTS Authors
```

```
DROP TABLE IF EXISTS AuthorArticlePairs
```

```
DROP TABLE IF EXISTS JournalArticlePairs
```

```
CREATE TABLE Articles(
ArticleTitle TEXT,
Pagination TEXT,
ArticleDate TEXT,
PMID VARCHAR(10) PRIMARY KEY
)
```

```
CREATE TABLE Journals(
ISSN VARCHAR(10) PRIMARY KEY,
Issue TEXT,
Title TEXT,
Volume TEXT,
PubDate TEXT
)
```

```
CREATE TABLE Authors(
LastName TEXT,
FirstName TEXT,
Initials TEXT,
Affiliation TEXT,
AuthorId VARCHAR(10) PRIMARY KEY
)
```

```
CREATE TABLE AuthorArticlePairs(
PMID VARCHAR(10),
AuthorId VARCHAR(10),
PRIMARY KEY(PMID, AuthorId),
FOREIGN KEY (PMID) REFERENCES Articles(PMID),
FOREIGN KEY (AuthorId) REFERENCES Authors(AuthorID)
)
```

```
CREATE TABLE JournalArticlePairs(
ISSN VARCHAR(10),
PMID VARCHAR(10),
PRIMARY KEY(PMID, ISSN),
FOREIGN KEY (PMID) REFERENCES Articles(PMID),
FOREIGN KEY (ISSN) REFERENCES Journals(ISSN)
)
```

```
CREATE TABLE History(
HistoryId VARCHAR(10) PRIMARY KEY,
PMID VARCHAR(10),
YEAR TEXT,
MONTH TEXT,
Day TEXT,
Status TEXT
)
```

```
xpathEx <- "//PubmedArticle/MedlineCitation/Article/ArticleTitle"
xpathEx2 <- "//PubmedArticle/MedlineCitation/Article/Pagination"
xpathEx3 <- "//PubmedArticle/MedlineCitation/Article/ArticleDate"
articleIdPath <- "//PubmedArticle/MedlineCitation/PMID"

x <- xpathSApply(xmlDoc, xpathEx)
x2 <- xpathSApply(xmlDoc, xpathEx2)
x3 <- xpathSApply(xmlDoc, xpathEx3)
articleIdPointers <- xpathSApply(xmlDoc, articleIdPath)

articleTitleArray <- xmlValue(x)
```

```
paginationArray <- xmlValue(x2)
dateArray <- xmlValue(x3)
articleIdArray <- xmlValue(articleIdPointers)
for (i in 1:length(dateArray)){
  date <- dateArray[[i]]
  year <- substr(date,0,4)
  month <- substr(date,5,6)
  day <- substr(date,7,8)
  dateArray[[i]] <- paste(year,month,day,sep="-")
}

articleInfoArray <- paste(articleIdArray, articleTitleArray, paginationArray, dateArray, sep="#")

historyInfoArray <- c()
for(id in articleIdArray){
  xpathEx14 <- paste("//PubmedArticle[MedlineCitation/PMID=",id,"]/PubmedData/History/PubMedPubDate/Yea
  monthPath <- paste("//PubmedArticle[MedlineCitation/PMID=",id,"]/PubmedData/History/PubMedPubDate/Mon
  dayPath <- paste("//PubmedArticle[MedlineCitation/PMID=",id,"]/PubmedData/History/PubMedPubDate/Day",
  statusPath <- paste("//PubmedArticle[MedlineCitation/PMID=",id,"]/PubmedData/History/PubMedPubDate/@Pu
  x14 <- xpathSApply(xmlDoc, xpathEx14)
  monthPointers <- xpathSApply(xmlDoc, monthPath)
  dayPointers <- xpathSApply(xmlDoc, dayPath)
  statuses <- xpathSApply(xmlDoc, statusPath)

  years <- xmlValue(x14)
  months <- xmlValue(monthPointers)
  days <- xmlValue(dayPointers)

  dates <- paste(years, months, days, sep="#")

  datesString <- toString(dates)

  statusesString <- toString(statuses)
  History <- paste(id, datesString, statusesString, sep="/")
  historyInfoArray <- append(historyInfoArray,History)
}


historyColumns <- c("HistoryId", "PMID", "Year", "Month", "Day", "Status")
articleColumns <- c("PMID", "ArticleTitle", "Pagination", "ArticleDate")

articleFrame <- data.frame(matrix(nrow=0, ncol=length(articleColumns)))
historyFrame <- data.frame(matrix(nrow=0, ncol=length(historyColumns)))

colnames(articleFrame) <- articleColumns
colnames(historyFrame) <- historyColumns
HistoryId = 1
for(rowString in historyInfoArray){
  rowStringArray <- unlist(strsplit(rowString, split="/"))
  PMID <- rowStringArray[[1]]
  Dates <- rowStringArray[[2]]
  Statuses <- rowStringArray[[3]]
  DatesArray <- unlist(strsplit(Dates, split=","))
```

```
  StatusArray <- unlist(strsplit(Statuses, split=","))
  for(dateIndex in 1:length(DatesArray)){
    datestring <- DatesArray[[dateIndex]]

    dateElements <- unlist(strsplit(datestring, split="#"))

    Year <- dateElements[[1]]

    Month <- dateElements[[2]]

    Day <- dateElements[[3]]

    Status <- StatusArray[[dateIndex]]

    historyFrame[nrow(historyFrame)+1,] <- c(as.character(HistoryId), PMID, Year, Month, Day, Status)

    HistoryId <- HistoryId + 1
  }
}

for(row in articleInfoArray){
  articleStringArray <- unlist(strsplit(row, split="#"))
  PMID <- articleStringArray[[1]]
  ArticleTitle <- articleStringArray[[2]]
  Pagination <- articleStringArray[[3]]
  ArticleDate <- articleStringArray[[4]]
  articleFrame[nrow(articleFrame)+1,] <- c(PMID,ArticleTitle, Pagination, ArticleDate)
}

print(articleFrame)
```

```
##        PMID
## 1  23874253
## 2  23194934
## 3  23091119
## 4  23080348
## 5  23068970
## 6  23008025
## 7  22677144
## 8  22634871
## 9  22434554
## 10 22426260
## 11 22305625
## 12 22235286
## 13 22189576
## 14 22115762
## 15 22024892
## 16 22020835
## 17 21851550
## 18 21778465
## 19 21301391
##
## 1                                        Regional anesthesia for children undergoing orthopedic ambu
```

```
## 2        Demographics and perioperative outcome in patients with depression and anxiety undergoing to
## 3                        Cerebrovascular reserve and stroke risk in patients with carotid stenosis or
## 4           Comparative perioperative outcomes associated with neuraxial versus general anesthesia fo
## 5                                        Vagus nerve stimulation vs. corpus callosotomy in the treat
## 6                                                    Have bilateral total knee arthroplasties
## 7                        The metabolic syndrome in patients undergoing knee and hip arthroplasty: tren
## 8                       Utilization of critical care services among patients undergoing total hip and
## 9                                                                          Visualization of the
## 10 FDG-PET assessment of rectal cancer response to neoadjuvant chemoradiotherapy is not associated wi
## 11                                                                  Factors influencing unexpected
## 12                                                 Intra- and inter-tumor heterogeneity of BRAF(V60
## 13           Beyond repeated-measures analysis of variance: advanced statistical methods for the anal
## 14                              In-hospital patient falls after total joint arthroplasty: incidence, den
## 15                                                  Metabolic syndrome and lumbar spine fusion
## 16                                Impact of race on survival in patients with clinically nonmetastati
## 17      Decision curve analysis assessing the clinical benefit of NMP22 in the detection of bladder c
## 18                     Mortality of patients with respiratory insufficiency and adult respiratory dist
## 19                                                          Comparative safety of simultaneous
##      Pagination ArticleDate
## 1         133-6  2012-06-20
## 2        149-57  2012-11-27
## 3       2884-91  2012-10-12
## 4        638-44  2012-09-25
## 5           3-8  2012-06-05
## 6         17-25  2012-03-21
## 7  1743-1749.e1  2012-02-04
## 8        107-16  2012-01-03
## 9        959-64  2011-11-23
## 10       378-86  2011-10-21
## 11        89-95  2011-08-18
## 12       e29336  2011-07-21
## 13       99-105  2012-06-20
## 14     823-8.e1  2012-11-27
## 15       989-95  2012-10-12
## 16      3145-52  2012-09-25
## 17       685-90  2012-06-05
## 18       306-11  2012-03-21
## 19       247-55  2012-02-04
```

```
print(historyFrame)
```

```
##    HistoryId     PMID  Year Month Day        Status
## 1          1 23874253  2012     1  15      received
## 2          2 23874253  2012     4  16      accepted
## 3          3 23874253  2012     6  20      epublish
## 4          4 23874253  2013     7  23        entrez
## 5          5 23874253  2013     7  23        pubmed
## 6          6 23874253  2013     7  23       medline
## 7          7 23194934  2012     7  16      received
## 8          8 23194934  2012     8  17       revised
## 9          9 23194934  2012     8  20      accepted
## 10        10 23194934  2012    11  27  aheadofprint
## 11        11 23194934  2012    12   1        entrez
## 12        12 23194934  2012    12   1        pubmed
```

```
## 13       13 23194934  2014     1  15       medline
## 14       14 23091119  2012    10  24        entrez
## 15       15 23091119  2012    10  24        pubmed
## 16       16 23091119  2013     1   4       medline
## 17       17 23080348  2012    10  20        entrez
## 18       18 23080348  2012    10  20        pubmed
## 19       19 23080348  2013     4   9       medline
## 20       20 23068970  2012     4   9      received
## 21       21 23068970  2012     9  18       revised
## 22       22 23068970  2012     9  22      accepted
## 23       23 23068970  2012    10  12 aheadofprint
## 24       24 23068970  2012    10  17        entrez
## 25       25 23068970  2012    10  17        pubmed
## 26       26 23068970  2013     7   3       medline
## 27       27 23008025  2012     2  13      received
## 28       28 23008025  2012     9   7      accepted
## 29       29 23008025  2012     9  25 aheadofprint
## 30       30 23008025  2012     9  26        entrez
## 31       31 23008025  2012     9  26        pubmed
## 32       32 23008025  2013     5  29       medline
## 33       33 22677144  2011     8  17      received
## 34       34 22677144  2012     4  11      accepted
## 35       35 22677144  2012     6   5 aheadofprint
## 36       36 22677144  2012     6   9        entrez
## 37       37 22677144  2012     6   9        pubmed
## 38       38 22677144  2013     5  17       medline
## 39       39 22634871  2012     5  29        entrez
## 40       40 22634871  2012     5  29        pubmed
## 41       41 22634871  2013     7   9       medline
## 42       42 22434554  2012     1  25      received
## 43       43 22434554  2012     2   1      accepted
## 44       44 22434554  2012     2   1       revised
## 45       45 22434554  2012     3  21 aheadofprint
## 46       46 22434554  2012     3  22        entrez
## 47       47 22434554  2012     3  22        pubmed
## 48       48 22434554  2013     1   8       medline
## 49       49 22426260  2012     3  20        entrez
## 50       50 22426260  2012     3  20        pubmed
## 51       51 22426260  2012     5   5       medline
## 52       52 22305625  2011     6  17      received
## 53       53 22305625  2011     9  13       revised
## 54       54 22305625  2011    10  12      accepted
## 55       55 22305625  2012     2   4 aheadofprint
## 56       56 22305625  2012     2   7        entrez
## 57       57 22305625  2012     2   7        pubmed
## 58       58 22305625  2012     7  26       medline
## 59       59 22235286  2011     8   1      received
## 60       60 22235286  2011    11  25      accepted
## 61       61 22235286  2012     1   3      epublish
## 62       62 22235286  2012     1  12        entrez
## 63       63 22235286  2012     1  12        pubmed
## 64       64 22235286  2012     5  15       medline
## 65       65 22189576  2011    12  23        entrez
## 66       66 22189576  2011    12  23        pubmed
```

```
## 67        67 22189576  2012     7  31       medline
## 68        68 22115762  2011     2  15      received
## 69        69 22115762  2011    10   7      accepted
## 70        70 22115762  2011    11  23  aheadofprint
## 71        71 22115762  2011    11  26        entrez
## 72        72 22115762  2011    11  26        pubmed
## 73        73 22115762  2013     1   3       medline
## 74        74 22024892  2011    10  26        entrez
## 75        75 22024892  2011    10  26        pubmed
## 76        76 22024892  2012     9   8       medline
## 77        77 22020835  2011     6  27      received
## 78        78 22020835  2011     8  26       revised
## 79        79 22020835  2011     9  19      accepted
## 80        80 22020835  2011    10  21  aheadofprint
## 81        81 22020835  2011    10  25        entrez
## 82        82 22020835  2011    10  25        pubmed
## 83        83 22020835  2012     8  16       medline
## 84        84 21851550  2011     8  18  aheadofprint
## 85        85 21851550  2011     8  20        entrez
## 86        86 21851550  2011     8  20        pubmed
## 87        87 21851550  2012     4  24       medline
## 88        88 21778465  2011     7  21  aheadofprint
## 89        89 21778465  2011     7  23        entrez
## 90        90 21778465  2011     7  23        pubmed
## 91        91 21778465  2013     1  18       medline
## 92        92 21301391  2011     2   9        entrez
## 93        93 21301391  2011     2   9        pubmed
## 94        94 21301391  2012    10  16       medline
```

```
dbWriteTable(dbcon, "History", historyFrame, append = TRUE, row.names=FALSE)
```

```
## [1] TRUE
```

```
dbWriteTable(dbcon, "Articles", articleFrame, append = T, row.names=FALSE)
```

```
## [1] TRUE
```

```
ISSNPath <- "//PubmedArticle/MedlineCitation/Article/Journal/ISSN"

ISSNPointers <- xpathSApply(xmlDoc, ISSNPath)

ISSNArray <- xmlValue(ISSNPointers)

distinctISSNArray = c()
for (ISSNstring in ISSNArray){
  if(!(ISSNstring %in% distinctISSNArray)){
    distinctISSNArray <- append(distinctISSNArray, ISSNstring)
  }
}

volumeArray <- c()
issueArray <- c()
```

```
titleArray <- c()
yearArray <- c()
pubdateArray <- c()
pmidArray <- c()
#j <- 0
for (issn in distinctISSNArray){
  xpathEx4 <- paste("//PubmedArticle/MedlineCitation/Article/Journal[ISSN='",issn,"']/JournalIssue/Volu
  xpathEx5 <- paste("//PubmedArticle/MedlineCitation/Article/Journal[ISSN='",issn,"']/JournalIssue/Issu
  xpathEx6 <- paste("//PubmedArticle/MedlineCitation/Article/Journal[ISSN='",issn,"']/Title[1]",sep="")
  xpathEx7 <- paste("//PubmedArticle/MedlineCitation/Article/Journal[ISSN='",issn,"']/JournalIssue/PubDa
  xpathEx8 <- paste("//PubmedArticle/MedlineCitation/Article/Journal[ISSN='",issn,"']/JournalIssue/PubDa
  pmidPath <- paste("//PubmedArticle/MedlineCitation[Article/Journal[ISSN='",issn,"']]/PMID",sep="")
  x4 <- xpathSApply(xmlDoc, xpathEx4)
  x5 <- xpathSApply(xmlDoc, xpathEx5)
  x6 <- xpathSApply(xmlDoc, xpathEx6)
  x7 <- xpathSApply(xmlDoc, xpathEx7)
  x8 <- xpathSApply(xmlDoc, xpathEx8)
  pmidPointers <- xpathSApply(xmlDoc, pmidPath)
  volumeArray <- append(volumeArray, xmlValue(x4[1]))
  issueArray <- append(issueArray, xmlValue(x5[1]))
  titleArray <- append(titleArray,xmlValue(x6[1]))
  yearArray<- append(yearArray, xmlValue(x7[1]))
  fulldate<- paste(xmlValue(x8[1]),xmlValue(x7[1]), sep="/")
  pubdateArray<- append(pubdateArray, fulldate)
  pmidArray<-append(pmidArray, xmlValue(pmidPointers[1]))
}

journalColumns <- c("ISSN", "volume","issue", "title", "pubdate")
journalArticlePairsColumns <-c("PMID", "ISSN")

journalFrame <- data.frame(matrix(nrow=0, ncol=length(journalColumns)))
journalArticleFrame <- data.frame(matrix(nrow=0, ncol=length(journalArticlePairsColumns)))
colnames(journalFrame) <- journalColumns
colnames(journalArticleFrame) <- journalArticlePairsColumns
for(i in 1:length(distinctISSNArray)){
  ISSN <- distinctISSNArray[[i]]
  Volume <- volumeArray[[i]]
  Issue <- issueArray[[i]]
  Title <- titleArray[[i]]
  PubDate <- pubdateArray[[i]]
  journalFrame[nrow(journalFrame)+1,] <- c(ISSN,Volume, Issue, Title, PubDate)
}

for(j in 1:length(ISSNArray)){
  ISSN <- ISSNArray[[j]]
  pmidPath <- paste("//PubmedArticle/MedlineCitation[Article/Journal[ISSN='",ISSN,"']]/PMID",sep="")
  pmidPointers <- xpathSApply(xmlDoc, pmidPath)
  for(pointer in pmidPointers){
    pmid <- xmlValue(pointer)
    journalArticleFrame[nrow(journalArticleFrame)+1,] <- c(pmid, ISSN)
  }
}
```

```
print(journalFrame)
```

```
##         ISSN volume issue
## 1  1556-3316      8     2
## 2  1545-7206     54     2
## 3  1524-4628     43    11
## 4  1532-8651     37     6
## 5  1532-2688     22     1
## 6  1528-1132    471     1
## 7  1532-8406     27    10
## 8  1528-1175    117     1
## 9  1432-1998     42     8
## 10 1530-0358     55     4
## 11 1873-4529     24     2
## 12 1932-6203      7     1
## 13 1528-1159     37    11
## 14 1097-0142    118    12
## 15 1464-410X    109     5
## 16 1525-1489     27     5
##                                                                        title
## 1  HSS journal : the musculoskeletal journal of Hospital for Special Surgery
## 2                                                                Psychosomatics
## 3                                 Stroke; a journal of cerebral circulation
## 4                                     Regional anesthesia and pain medicine
## 5                  Seizure : the journal of the British Epilepsy Association
## 6                                 Clinical orthopaedics and related research
## 7                                            The Journal of arthroplasty
## 8                                                           Anesthesiology
## 9                                                       Pediatric radiology
## 10                                        Diseases of the colon and rectum
## 11                                          Journal of clinical anesthesia
## 12                                                                  PloS one
## 13                                                                     Spine
## 14                                                                    Cancer
## 15                                                        BJU international
## 16                                      Journal of intensive care medicine
##     pubdate
## 1  Jul/2012
## 2     NA/NA
## 3  Nov/2012
## 4     NA/NA
## 5  Jan/2013
## 6  Jan/2013
## 7  Dec/2012
## 8  Jul/2012
## 9  Aug/2012
## 10 Apr/2012
## 11 Mar/2012
## 12  NA/2012
## 13 May/2012
## 14 Jun/2012
## 15 Mar/2012
## 16    NA/NA
```

```
print(journalArticleFrame)
```

```
##        PMID      ISSN
## 1  23874253 1556-3316
## 2  23194934 1545-7206
## 3  23091119 1524-4628
## 4  23080348 1532-8651
## 5  22189576 1532-8651
## 6  23068970 1532-2688
## 7  23008025 1528-1132
## 8  22677144 1532-8406
## 9  22115762 1532-8406
## 10 22634871 1528-1175
## 11 22434554 1432-1998
## 12 22426260 1530-0358
## 13 22305625 1873-4529
## 14 22235286 1932-6203
## 15 23080348 1532-8651
## 16 22189576 1532-8651
## 17 22677144 1532-8406
## 18 22115762 1532-8406
## 19 22024892 1528-1159
## 20 21301391 1528-1159
## 21 22020835 1097-0142
## 22 21851550 1464-410X
## 23 21778465 1525-1489
## 24 22024892 1528-1159
## 25 21301391 1528-1159
```

```
dbWriteTable(dbcon, "Journals", journalFrame, append=TRUE, row.names=FALSE)
```

```
## [1] TRUE
```

```
dbWriteTable(dbcon, "JournalArticlePairs", journalArticleFrame, append=TRUE, row.names=FALSE)
```

```
## [1] TRUE
```

```
xpathEx9 <- "//PubmedArticle/MedlineCitation/Article/AuthorList/Author/LastName"
xpathEx10 <- "//PubmedArticle/MedlineCitation/Article/AuthorList/Author/ForeName"
xpathEx11 <- "//PubmedArticle/MedlineCitation/Article/AuthorList/Author/Initials"
xpathEx13 <- "//PubmedArticle/MedlineCitation/Article/AuthorList/Author"

x9 <- xpathSApply(xmlDoc, xpathEx9)
x10 <- xpathSApply(xmlDoc, xpathEx10)
x11 <- xpathSApply(xmlDoc, xpathEx11)
x13 <- xpathSApply(xmlDoc, xpathEx13)

lastauthorArray <- xmlValue(x9)
firstauthorArray <- xmlValue(x10)
initialsArray <- xmlValue(x11)
fullAuthors <- xmlValue(x13)
```

```
authorArray <- paste(firstauthorArray, lastauthorArray, initialsArray, sep="/")

distinctAuthorArray <- c()

for(y in 1:length(authorArray)){

  author <- authorArray[[y]]

  if(!(author %in% distinctAuthorArray)){
    distinctAuthorArray <- append(distinctAuthorArray, author)
  }
}
finalAuthorList <- c()
for(d in 1:length(distinctAuthorArray)){
  authorElement <- distinctAuthorArray[[d]]
  wordArray <- unlist(strsplit(authorElement, "/"))
  firstName <- wordArray[[1]]
  xpathEx12 <- paste("//PubmedArticle/MedlineCitation/Article/AuthorList/Author[ForeName='",firstName,"
  x12 <- xpathSApply(xmlDoc, xpathEx12)
  affiliation <- xmlValue(x12[1])
  if(length(x12) > 0){
    str <- paste(authorElement, affiliation, sep="/")
    finalAuthorList <- append(finalAuthorList, str)
  }else{
    finalAuthorList <- append(finalAuthorList, distinctAuthorArray[[d]])
  }
}

authorColumns <- c("AuthorId","LastName", "FirstName", "Initials", "Affiliation")
authorArticlePairsColumns <-c("AuthorId", "PMID")

authorFrame <- data.frame(matrix(nrow=0, ncol=length(authorColumns)))
authorArticleFrame <- data.frame(matrix(nrow=0, ncol=length(authorArticlePairsColumns)))
colnames(authorFrame) <-  authorColumns
colnames(authorArticleFrame) <- authorArticlePairsColumns

pmids <- list()

for(author in finalAuthorList){
  authorElements <- unlist(strsplit(author, split="/"))
  firstName <- authorElements[[1]]
  lastName <- authorElements[[2]]

  creditsPath <- paste("//PubmedArticle/MedlineCitation[Article/AuthorList/Author/ForeName[.='",firstNa

  creditsPointers <- xpathSApply(xmlDoc, creditsPath)

  pmids[[author]] <- c()
  for(pointer in creditsPointers){
    pmids[[author]] <- append(pmids[[author]],xmlValue(pointer))
  }
}
```

```
AuthorId <- 1
for(row in finalAuthorList){
  rowElements <- unlist(strsplit(row, split="/"))
  FirstName <- rowElements[[1]]
  LastName <- rowElements[[2]]
  Initials <- rowElements[[3]]
  Affiliation <- "NULL"
  if(length(rowElements) > 3){
    Affiliation <- rowElements[[4]]
  }
  authorFrame[nrow(authorFrame)+1,] <- c(as.character(AuthorId), LastName,FirstName, Initials, Affiliati
  AuthorId <- AuthorId + 1
}

AuthorId <- 1
for(authorListing in names(pmids)){
  authorElements <- unlist(strsplit(authorListing, split="/"))
  for(pmid in pmids[[authorListing]]){
    authorArticleFrame[nrow(authorArticleFrame)+1,] <- c(as.character(AuthorId), pmid)
  }
  AuthorId <- AuthorId+1
}

print(authorFrame)
```

```
##    AuthorId            LastName          FirstName Initials
## 1         1                Kuo             Cassie        C
## 2         2            Edwards             Alison        A
## 3         3           Mazumdar              Madhu        M
## 4         4          Memtsoudis          Stavros G       SG
## 5         5           Stundner            Ottokar        O
## 6         6           Kirksey             Meghan        M
## 7         7              Chiu             Ya Lin       YL
## 8         8          Poultsides            Lazaros        L
## 9         9              Gerner              Peter        P
## 10       10               Gupta               Ajay        A
## 11       11              Chazen             J Levi       JL
## 12       12             Hartman               Maya        M
## 13       13             Delgado              Diana        D
## 14       14             Anumula             Nikesh        N
## 15       15                Shao              Huibo        H
## 16       16              Segal             Alan Z       AZ
## 17       17              Kamel             Hooman        H
## 18       18             Leifer               Dana        D
## 19       19             Sanelli             Pina C       PC
## 20       20                Chiu             Ya-Lin       YL
## 21       21                 Sun             Xuming        X
## 22       22           Fleischut              Peter        P
## 23       23             Fritsch            Gerhard        G
## 24       24             Lancman              Guido        G
## 25       25                Virk            Michael        M
## 26       26          Greenfield          Jeffrey P       JP
## 27       27           Weinstein             Steven        S
```

```
## 28       28             Schwartz         Theodore H       TH
## 29       29             Mantilla           Carlos B       CB
## 30       30              Parvizi               Javad        J
## 31       31 Gonzalez Della Valle          Alejandro        A
## 32       32                   Ma                Yan        Y
## 33       33                Nurok            Michael        M
## 34       34             Pastores          Stephen M       SM
## 35       35          Kovanlikaya               Arzu        A
## 36       36            Rosenbaum              Daniel        D
## 37       37              Dunning             Allison        A
## 38       38                Brill            Paula W       PW
## 39       39                 Ruby          Jeannine A       JA
## 40       40              Leibold              Tobias        T
## 41       41              Akhurst          Timothy J       TJ
## 42       42                 Shia               Jinru        J
## 43       43                Saltz          Leonard B       LB
## 44       44               Riedel             Elyn R       ER
## 45       45               Larson           Steven M       SM
## 46       46              Guillem             José G       JG
## 47       47            Swamidoss           Cephas P       CP
## 48       48              Edwards           Alison M       AM
## 49       49              Liguori          Gregory A       GA
## 50       50            Yancovitz              Molly        M
## 51       51            Litterman               Adam        A
## 52       52                 Yoon             Joanne        J
## 53       53                   Ng              Elise        E
## 54       54              Shapiro          Richard L       RL
## 55       55               Berman          Russell S       RS
## 56       56              Pavlick             Anna C       AC
## 57       57           Darvishian             Farbod        F
## 58       58             Christos               Paul        P
## 59       59                Osman               Iman        I
## 60       60              Polsky              David        D
## 61       61                   Dy      Christopher J       CJ
## 62       62          Della Valle Alejandro Gonzalez       AG
## 63       63            Pumberger            Matthias        M
## 64       64              Girardi         Federico P       FP
## 65       65           Koscuiszka             Michael        M
## 66       66              Hatcher               David        D
## 67       67             Christos             Paul J       PJ
## 68       68                 Rose               Amy E       AE
## 69       69            Greenwald            Holly S       HS
## 70       70                 Chiu             Ya-lin       YL
## 71       71               Taneja             Samir S       SS
## 72       72                  Lee               Peng        P
## 73       73             Barbieri      Christopher E       CE
## 74       74                  Cha            Eugene K       EK
## 75       75            Chromecki            Thomas F       TF
## 76       76                Lotan               Yair        Y
## 77       77               Svatek           Robert S       RS
## 78       78               Scherr          Douglas S       DS
## 79       79          Karakiewicz            Pierre I       PI
## 80       80                  Sun             Maxine        M
## 81       81              Shariat         Shahrokh F       SF
```

```
## 82       82          Bombardieri         Anna Maria       AM
## 83       83                 Walz         J Matthias       JM
## 84       84              Passias            Peter G       PG
##
## 1                                                                                  Department of Anesthesiolo
## 2
## 3
## 4       Department of Anesthesiology, Hospital for Special Surgery, Weill Medical College of Cornell
## 5                                                    Department of Anesthesiology, Hospital for Sp
## 6
## 7
## 8
## 9
## 10 Department of Radiology, Weill Cornell Medical School, New York-Presbyterian Hospital, 525 East 68
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24                                                                Department of Neurological Surg
## 25                                                                      Department of U
## 26
## 27
## 28
## 29
## 30
## 31                                                                   Department of Orthoped
## 32                                                 Research Division, Hospital for Special Surgery, Wei
## 33                                                                      Department of U
## 34
## 35               Department of Radiology, Weill Cornell Medical College, New York Presbyterian Hospi
## 36
## 37
## 38
## 39                                                                                        Dep
## 40
## 41
## 42
## 43
## 44
## 45
## 46
## 47
## 48
## 49
## 50                                                          Department of Dermatology, New York Un
```

```
## 51
## 52
## 53
## 54
## 55
## 56
## 57
## 58
## 59
## 60
## 61
## 62
## 63
## 64
## 65                                                                                                  Department of U
## 66
## 67
## 68
## 69
## 70
## 71
## 72
## 73                                                                  Department of Urology, Weill Medical College o
## 74
## 75
## 76
## 77
## 78
## 79
## 80
## 81
## 82
## 83
## 84                                                                                                    Division of
```

```
print(authorArticleFrame)
```

```
##     AuthorId     PMID
## 1          1 23874253
## 2          2 23874253
## 3          3 23874253
## 4          3 23194934
## 5          3 23091119
## 6          3 23080348
## 7          3 23068970
## 8          3 23008025
## 9          3 22677144
## 10         3 22634871
## 11         3 22434554
## 12         3 22426260
## 13         3 22305625
## 14         3 22235286
## 15         3 22189576
## 16         3 22115762
```

```
## 17          3 22024892
## 18          3 22020835
## 19          3 21851550
## 20          3 21778465
## 21          3 21301391
## 22          4 23874253
## 23          4 23194934
## 24          4 23080348
## 25          4 23008025
## 26          4 22677144
## 27          4 22634871
## 28          4 22305625
## 29          4 22189576
## 30          4 22115762
## 31          4 22024892
## 32          4 21778465
## 33          4 21301391
## 34          5 23194934
## 35          5 23080348
## 36          5 23008025
## 37          5 22634871
## 38          6 23194934
## 39          6 22024892
## 40          7 23194934
## 41          7 22677144
## 42          7 22024892
## 43          7 21778465
## 44          7 21301391
## 45          8 23194934
## 46          8 23080348
## 47          9 23194934
## 48          9 23080348
## 49         10 23091119
## 50         11 23091119
## 51         12 23091119
## 52         13 23091119
## 53         14 23091119
## 54         15 23091119
## 55         15 23068970
## 56         16 23091119
## 57         17 23091119
## 58         18 23091119
## 59         19 23091119
## 60         20 23080348
## 61         20 22634871
## 62         20 22115762
## 63         21 23080348
## 64         21 22634871
## 65         22 23080348
## 66         23 23080348
## 67         24 23068970
## 68         25 23068970
## 69         26 23068970
## 70         27 23068970
```

```
## 71        28 23068970
## 72        29 23008025
## 73        30 23008025
## 74        31 22677144
## 75        32 22677144
## 76        32 22305625
## 77        32 22189576
## 78        32 22115762
## 79        32 22024892
## 80        32 21778465
## 81        32 21301391
## 82        33 22634871
## 83        34 22634871
## 84        35 22434554
## 85        36 22434554
## 86        37 22434554
## 87        37 21851550
## 88        38 22434554
## 89        39 22426260
## 90        40 22426260
## 91        41 22426260
## 92        42 22426260
## 93        43 22426260
## 94        44 22426260
## 95        45 22426260
## 96        46 22426260
## 97        47 22305625
## 98        48 22305625
## 99        49 22305625
## 100       50 22235286
## 101       51 22235286
## 102       52 22235286
## 103       53 22235286
## 104       54 22235286
## 105       55 22235286
## 106       56 22235286
## 107       57 22235286
## 108       58 22235286
## 109       59 22235286
## 110       59 22020835
## 111       60 22235286
## 112       61 22115762
## 113       62 22115762
## 114       63 22024892
## 115       64 22024892
## 116       64 21301391
## 117       65 22020835
## 118       66 22020835
## 119       67 22020835
## 120       68 22020835
## 121       69 22020835
## 122       70 22020835
## 123       71 22020835
## 124       72 22020835
```

```
## 125        73 21851550
## 126        74 21851550
## 127        75 21851550
## 128        76 21851550
## 129        77 21851550
## 130        78 21851550
## 131        79 21851550
## 132        80 21851550
## 133        81 21851550
## 134        82 21778465
## 135        83 21778465
## 136        84 21301391
```

```
dbWriteTable(dbcon, "Authors", authorFrame, append=TRUE, row.names=FALSE)
```

```
## [1] TRUE
```

```
dbWriteTable(dbcon, "AuthorArticlePairs", authorArticleFrame, append=TRUE, row.names=FALSE)
```

```
## [1] TRUE
```

```
SELECT *
FROM History
```

Table 1: Displaying records 1 - 10

HistoryId	PMID	YEAR	MONTH	Day	Status
1	23874253	2012	1	15	received
10	23194934	2012	11	27	aheadofprint
11	23194934	2012	12	1	entrez
12	23194934	2012	12	1	pubmed
13	23194934	2014	1	15	medline
14	23091119	2012	10	24	entrez
15	23091119	2012	10	24	pubmed
16	23091119	2013	1	4	medline
17	23080348	2012	10	20	entrez
18	23080348	2012	10	20	pubmed

```
SELECT * FROM Authors ORDER BY LastName
```

Table 2: Displaying records 1 - 10

LastName	FirstName	Initials	Affiliation	AuthorId
Akhurst	Timothy J	TJ	NULL	41
Anumula	Nikesh	N	NULL	14
Barbieri	Christopher E	CE	Department of Urology, Weill Medical College of Cornell University, New York-Presbyterian Hospital, New York, NY 10065, USA.	73

LastName	FirstName	Initials	Affiliation	AuthorId
Berman	Russell S	RS	NULL	55
Bombardieri	Anna Maria	AM	NULL	82
Brill	Paula W	PW	NULL	38
Cha	Eugene K	EK	NULL	74
Chazen	J Levi	JL	NULL	11
Chiu	Ya-Lin	YL	NULL	20
Chiu	Ya Lin	YL	NULL	7

```
SELECT * FROM Journals
```

Table 3: Displaying records 1 - 10

ISSN	Issue	Title	Volume	PubDate
1097-0142	12	Cancer	118	Jun/2012
1432-1998	8	Pediatric radiology	42	Aug/2012
1464-410X	5	BJU international	109	Mar/2012
1524-4628	11	Stroke; a journal of cerebral circulation	43	Nov/2012
1525-1489	5	Journal of intensive care medicine	27	NA/NA
1528-1132	1	Clinical orthopaedics and related research	471	Jan/2013
1528-1159	11	Spine	37	May/2012
1528-1175	1	Anesthesiology	117	Jul/2012
1530-0358	4	Diseases of the colon and rectum	55	Apr/2012
1532-2688	1	Seizure : the journal of the British Epilepsy Association	22	Jan/2013

```
SELECT * FROM Articles
```

Table 4: Displaying records 1 - 10

ArticleTitle	Pagination	ArticleDate	PMID
Comparative safety of simultaneous and staged anterior and posterior spinal surgery.	247-55	2012-02-04	21301391
Mortality of patients with respiratory insufficiency and adult respiratory distress syndrome after surgery: the obesity paradox.	306-11	2012-03-21	21778465
Decision curve analysis assessing the clinical benefit of NMP22 in the detection of bladder cancer: secondary analysis of a prospective trial.	685-90	2012-06-05	21851550
Impact of race on survival in patients with clinically nonmetastatic prostate cancer who deferred primary treatment.	3145-52	2012-09-25	22020835
Metabolic syndrome and lumbar spine fusion surgery: epidemiology and perioperative outcomes.	989-95	2012-10-12	22024892
In-hospital patient falls after total joint arthroplasty: incidence, demographics, and risk factors in the United States.	823-8.e1	2012-11-27	22115762
Beyond repeated-measures analysis of variance: advanced statistical methods for the analysis of longitudinal data in anesthesia research.	99-105	2012-06-20	22189576

ArticleTitle	Pagination	ArticleDate	PMID
Intra- and inter-tumor heterogeneity of BRAF(V600E))mutations in primary and metastatic melanoma.	e29336	2011-07-21	22235286
Factors influencing unexpected disposition after orthopedic ambulatory surgery.	89-95	2011-08-18	22305625
FDG-PET assessment of rectal cancer response to neoadjuvant chemoradiotherapy is not associated with long-term prognosis: a prospective evaluation.	378-86	2011-10-21	22426260

```
SELECT * FROM AuthorArticlePairs
```

Table 5: Displaying records 1 - 10

PMID	AuthorId
23874253	1
23091119	10
23091119	11
23091119	12
23091119	13
23091119	14
23068970	15
23091119	15
23091119	16
23091119	17

```
SELECT * FROM JournalArticlePairs
```

Table 6: Displaying records 1 - 10

ISSN	PMID
1097-0142	22020835
1432-1998	22434554
1464-410X	21851550
1524-4628	23091119
1525-1489	21778465
1528-1132	23008025
1528-1159	21301391
1528-1159	22024892
1528-1175	22634871
1530-0358	22426260

```
CREATE SCHEMA IF NOT EXISTS starschema;
```

```
USE starschema;
```

```
DROP TABLE IF EXISTS starschema.author
```

```
DROP TABLE IF EXISTS starschema.article

DROP TABLE IF EXISTS starschema.journal

DROP TABLE IF EXISTS starschema.history

DROP TABLE IF EXISTS starschema.AuthorArticlePairs

DROP TABLE IF EXISTS starschema.JournalArticlePairs

DROP TABLE IF EXISTS starschema.publishedArticlesAuthorTimeDimensions

DROP TABLE IF EXISTS starschema.publishedArticlesJournalTimeDimensions

DROP TABLE IF EXISTS starschema.publishedArticlesAuthorFact

DROP TABLE IF EXISTS starschema.publishedArticlesJournalFact

DROP TABLE IF EXISTS starschema.publishedArticlesHistoryDimensions

DROP TABLE IF EXISTS starschema.historySubmit

DROP TABLE IF EXISTS starschema.historyPublish

CREATE TABLE starschema.author
  AS SELECT LastName, FirstName, Initials, Affiliation FROM practicumIIdb.Authors;

CREATE TABLE starschema.article
  AS SELECT PMID, ArticleTitle, Pagination, Date(ArticleDate) as "ArticleDate" FROM practicumIIdb.Artic

CREATE TABLE starschema.journal
 AS SELECT ISSN, volume,issue, title, pubdate FROM practicumIIdb.Journals;

CREATE TABLE starschema.history
  AS SELECT HistoryId, PMID, Year, Month, Day, Status FROM practicumIIdb.History;

CREATE TABLE starschema.AuthorArticlePairs
  AS SELECT AuthorId, PMID FROM practicumIIdb.AuthorArticlePairs;

CREATE TABLE starschema.JournalArticlePairs
  AS SELECT PMID, ISSN FROM practicumIIdb.JournalArticlePairs;

CREATE TABLE starschema.publishedArticlesAuthorTimeDimensions (
  dimensionId int NOT NULL AUTO_INCREMENT PRIMARY KEY,
  when_published TEXT,
  quarter INT,
  year_num INT,
  AuthorId TEXT,
  PMID TEXT
)
```

```
INSERT INTO starschema.publishedArticlesAuthorTimeDimensions(when_published,quarter,year_num,AuthorId, I
  SELECT ArticleDate as when_published, quarter(ArticleDate) as quarter, year(ArticleDate) as year_num,
  FROM starschema.article
  INNER JOIN starschema.AuthorArticlePairs ON starschema.AuthorArticlePairs.PMID = starschema.article.PI
```

```
CREATE TABLE starschema.publishedArticlesJournalTimeDimensions(
  journalDimensionId int NOT NULL AUTO_INCREMENT PRIMARY KEY,
  when_published TEXT,
  quarter INT,
  year_num INT,
  ISSN TEXT,
  PMID TEXT
)
```

```
INSERT INTO starschema.publishedArticlesJournalTimeDimensions(when_published,quarter,year_num,ISSN, PMI
  SELECT ArticleDate as when_published, quarter(ArticleDate) as quarter, year(ArticleDate) as year_num,
  FROM starschema.article
  INNER JOIN starschema.JournalArticlePairs ON starschema.JournalArticlePairs.PMID = starschema.article
```

```
CREATE TABLE starschema.publishedArticlesHistoryDimensions(
  PMID VARCHAR(10) PRIMARY KEY,
  submitTime DATE,
  publishTime DATE,
  totalTime INT,
  quarter INT
)
```

```
CREATE TABLE starschema.historySubmit(
HistoryId TEXT,
PMID TEXT,
submitTime TEXT,
Status TEXT
)
```

```
CREATE TABLE starschema.historyPublish(
HistoryId TEXT,
PMID TEXT,
publishTime TEXT,
Status TEXT
)
```

```
INSERT INTO starschema.historySubmit(HistoryId, PMID, submitTime, Status)
  SELECT HistoryId, PMID, CONCAT(Year,"-",Month,"-",Day) as "submitTime", Status FROM starschema.histor
```

```
INSERT INTO starschema.historyPublish(HistoryId, PMID, publishTime, Status)
  SELECT HistoryId, PMID, CONCAT(Year,"-",Month,"-",Day) as "publishTime", Status FROM starschema.histo
```

```
INSERT INTO starschema.publishedArticlesHistoryDimensions(PMID, submitTime, publishTime, totalTime, qua
  SELECT DISTINCT starschema.historySubmit.PMID, submitTime, publishTime, DATEDIFF(publishTime,submitTi
  FROM starschema.historySubmit
  INNER JOIN starschema.historyPublish ON starschema.historyPublish.PMID=starschema.historySubmit.PMID
  INNER JOIN starschema.publishedArticlesAuthorTimeDimensions ON starschema.publishedArticlesAuthorTimeI
```

```
CREATE TABLE starschema.publishedArticlesAuthorFact (
  AuthorId VARCHAR(50),
  year_num INT,
  articlesPublished INT,
  CONSTRAINT PRIMARY KEY(AuthorId, year_num)
)
```

```
CREATE TABLE starschema.publishedArticlesJournalFact (
  ISSN VARCHAR(10),
  quarter INT,
  articlesPublished int,
  CONSTRAINT PRIMARY KEY(ISSN, quarter)
)
```

```
INSERT INTO starschema.publishedArticlesAuthorFact(AuthorId, year_num, articlesPublished)
  SELECT AuthorId, year_num, count(PMID) as articlesPublished
  FROM starschema.publishedArticlesAuthorTimeDimensions
  GROUP BY year_num, AuthorId
  ORDER BY AuthorId
```

```
INSERT INTO starschema.publishedArticlesJournalFact(ISSN, quarter, articlesPublished)
  SELECT ISSN, quarter, count(PMID) as articlesPublished
  FROM starschema.publishedArticlesJournalTimeDimensions
  GROUP BY quarter, ISSN
  ORDER BY ISSN
```

```
SELECT * FROM starschema.publishedArticlesJournalTimeDimensions ORDER BY ISSN
```

Table 7: Displaying records 1 - 10

journalDimensionId	when_published	quarter	year_num	ISSN	PMID
1	2012-09-25	3	2012	1097-0142	22020835
2	2011-11-23	4	2011	1432-1998	22434554
3	2012-06-05	2	2012	1464-410X	21851550
4	2012-10-12	4	2012	1524-4628	23091119
5	2012-03-21	1	2012	1525-1489	21778465
6	2012-03-21	1	2012	1528-1132	23008025
7	2012-02-04	1	2012	1528-1159	21301391
8	2012-10-12	4	2012	1528-1159	22024892
9	2012-01-03	1	2012	1528-1175	22634871
10	2011-10-21	4	2011	1530-0358	22426260

```
SELECT * FROM publishedArticlesAuthorTimeDimensions
```

Table 8: Displaying records 1 - 10

dimensionId	when_published	quarter	year_num	AuthorId	PMID
1	2012-06-20	2	2012	1	23874253
2	2012-10-12	4	2012	10	23091119

dimensionId	when_published	quarter	year_num	AuthorId	PMID
3	2012-10-12	4	2012	11	23091119
4	2012-10-12	4	2012	12	23091119
5	2012-10-12	4	2012	13	23091119
6	2012-10-12	4	2012	14	23091119
7	2012-06-05	2	2012	15	23068970
8	2012-10-12	4	2012	15	23091119
9	2012-10-12	4	2012	16	23091119
10	2012-10-12	4	2012	17	23091119

```
SELECT * FROM publishedArticlesAuthorFact
```

Table 9: Displaying records 1 - 10

AuthorId	year_num	articlesPublished
1	2012	1
10	2012	1
11	2012	1
12	2012	1
13	2012	1
14	2012	1
15	2012	2
16	2012	1
17	2012	1
18	2012	1

```
SELECT * FROM publishedArticlesJournalFact
```

Table 10: Displaying records 1 - 10

ISSN	quarter	articlesPublished
1097-0142	3	1
1432-1998	4	1
1464-410X	2	1
1524-4628	4	1
1525-1489	1	1
1528-1132	1	1
1528-1159	1	1
1528-1159	4	1
1528-1175	1	1
1530-0358	4	1

```
SELECT * FROM starschema.history
```

Table 11: Displaying records 1 - 10

HistoryId	PMID	Year	Month	Day	Status
1	23874253	2012	1	15	received
10	23194934	2012	11	27	aheadofprint
11	23194934	2012	12	1	entrez
12	23194934	2012	12	1	pubmed
13	23194934	2014	1	15	medline
14	23091119	2012	10	24	entrez
15	23091119	2012	10	24	pubmed
16	23091119	2013	1	4	medline
17	23080348	2012	10	20	entrez
18	23080348	2012	10	20	pubmed

```
SELECT * FROM publishedArticlesHistoryDimensions
```

Table 12: Displaying records 1 - 10

PMID	submitTime	publishTime	totalTime	quarter
22020835	2011-06-27	2012-08-16	416	3
22115762	2011-02-15	2013-01-03	688	4
22235286	2011-08-01	2012-05-15	288	3
22305625	2011-06-17	2012-07-26	405	3
22434554	2012-01-25	2013-01-08	349	4
22677144	2011-08-17	2013-05-17	639	1
23008025	2012-02-13	2013-05-29	471	1
23068970	2012-04-09	2013-07-03	450	2
23194934	2012-07-16	2014-01-15	548	4
23874253	2012-01-15	2013-07-23	555	2

```
SELECT * FROM starschema.historyPublish
```

Table 13: Displaying records 1 - 10

HistoryId	PMID	publishTime	Status
13	23194934	2014-1-15	medline
16	23091119	2013-1-4	medline
19	23080348	2013-4-9	medline
26	23068970	2013-7-3	medline
32	23008025	2013-5-29	medline
38	22677144	2013-5-17	medline
41	22634871	2013-7-9	medline
48	22434554	2013-1-8	medline
51	22426260	2012-5-5	medline
58	22305625	2012-7-26	medline

```
SELECT * FROM historySubmit
```

Table 14: Displaying records 1 - 10

HistoryId	PMID	submitTime	Status
1	23874253	2012-1-15	received
20	23068970	2012-4-9	received
27	23008025	2012-2-13	received
33	22677144	2011-8-17	received
42	22434554	2012-1-25	received
52	22305625	2011-6-17	received
59	22235286	2011-8-1	received
68	22115762	2011-2-15	received
7	23194934	2012-7-16	received
77	22020835	2011-6-27	received

```
library(sqldf)
```

```
## Loading required package: gsubfn
```

```
## Loading required package: proto
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 6): Library not loaded: /opt/X1
##   Referenced from: /Library/Frameworks/R.framework/Versions/4.1/Resources/modules/R_X11.so
##   Reason: image not found
```

```
## Could not load tcltk.  Will use slower R code instead.
```

```
## Loading required package: RSQLite
```

```
##
## Attaching package: 'RSQLite'
```

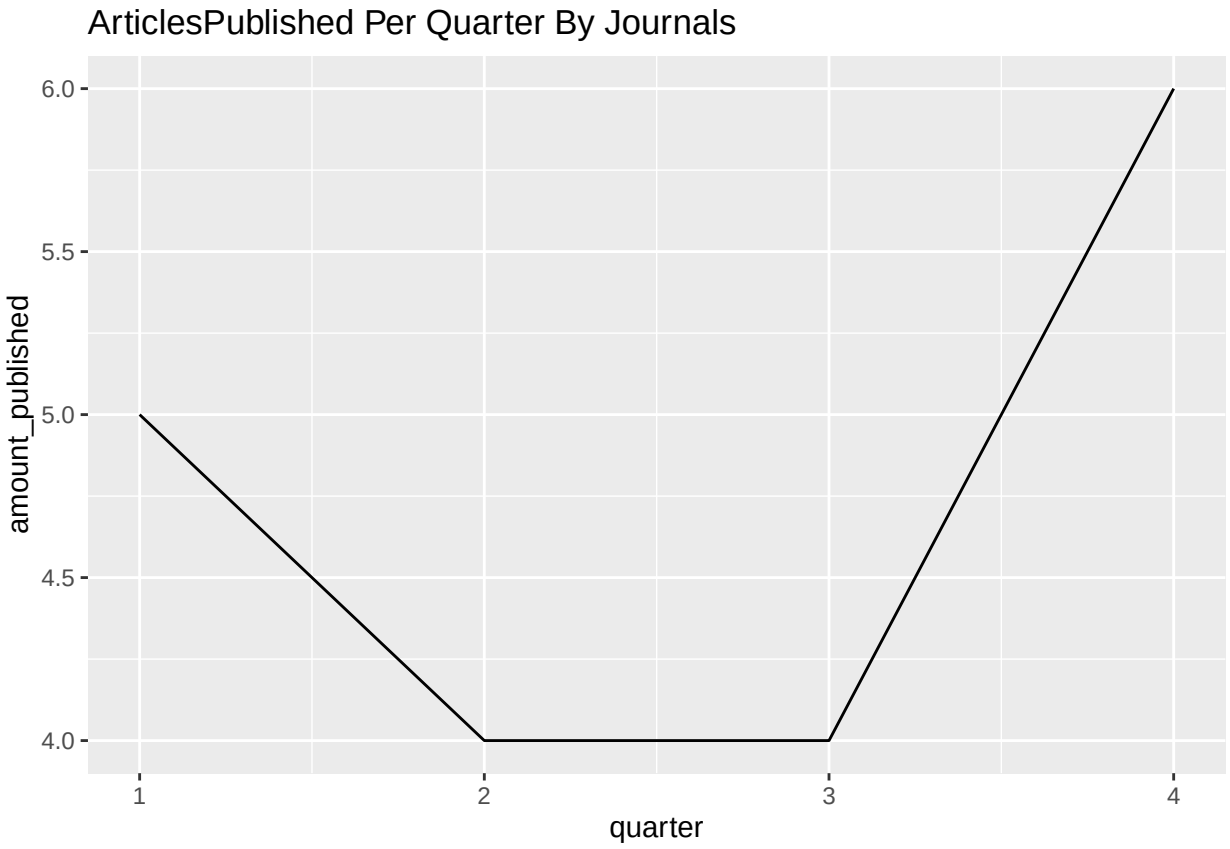
```
## The following object is masked from 'package:RMySQL':
##
##     isIdCurrent
```

```
## sqldf will default to using MySQL
```

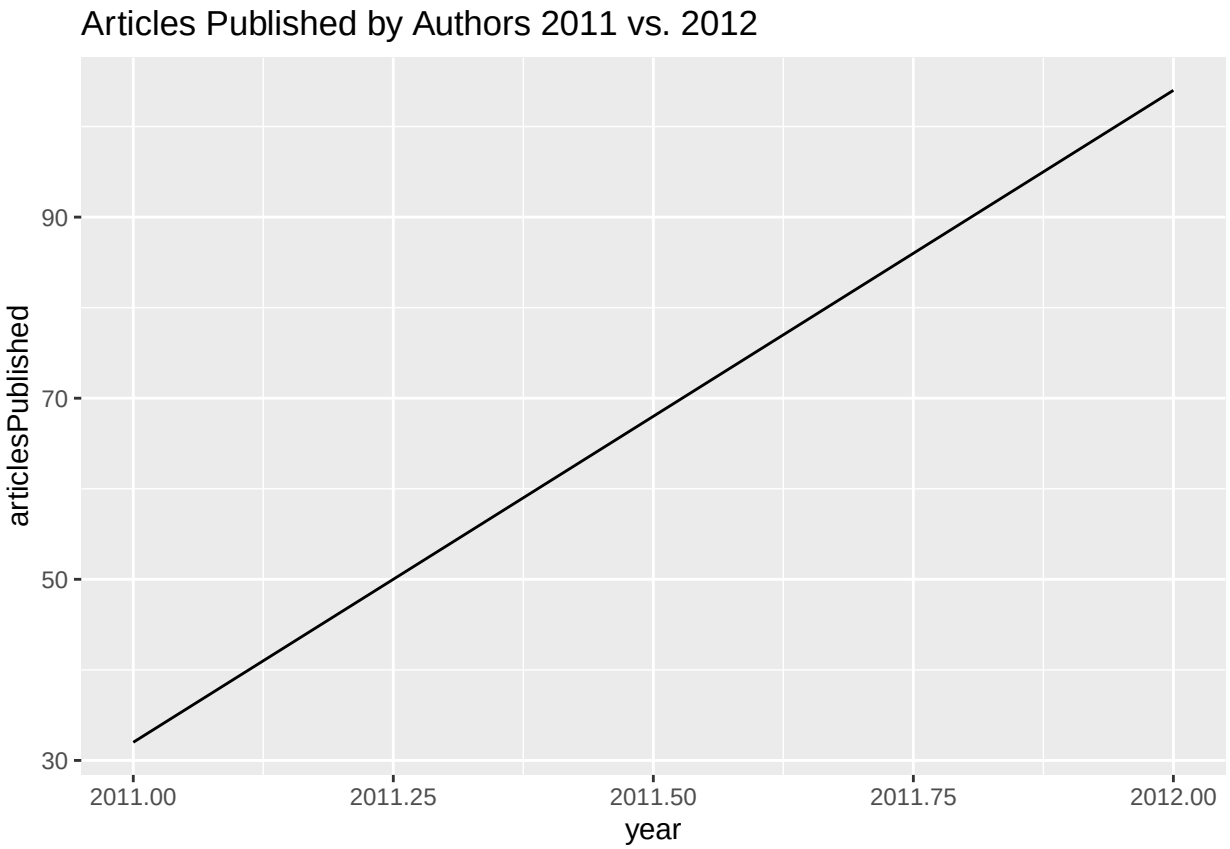
```
library(ggplot2)
df <- dbReadTable(dbcon, "publishedArticlesJournalFact")
options(sqldf.driver = "SQLite")
res <- sqldf("SELECT SUM(articlesPublished) as amount_published, quarter FROM df GROUP BY quarter ")
print(res)
```

```
##   amount_published quarter
## 1                5       1
## 2                4       2
## 3                4       3
## 4                6       4
```

```
quarter<-res$quarter
amount_published<-res$amount_published
#plot(num,type="p",col="red",xlab="years?",ylab="articlesPublished's?")
ggplot(res, aes(x=quarter, y=amount_published)) +
  geom_line() + ggtitle("ArticlesPublished Per Quarter By Journals")
```

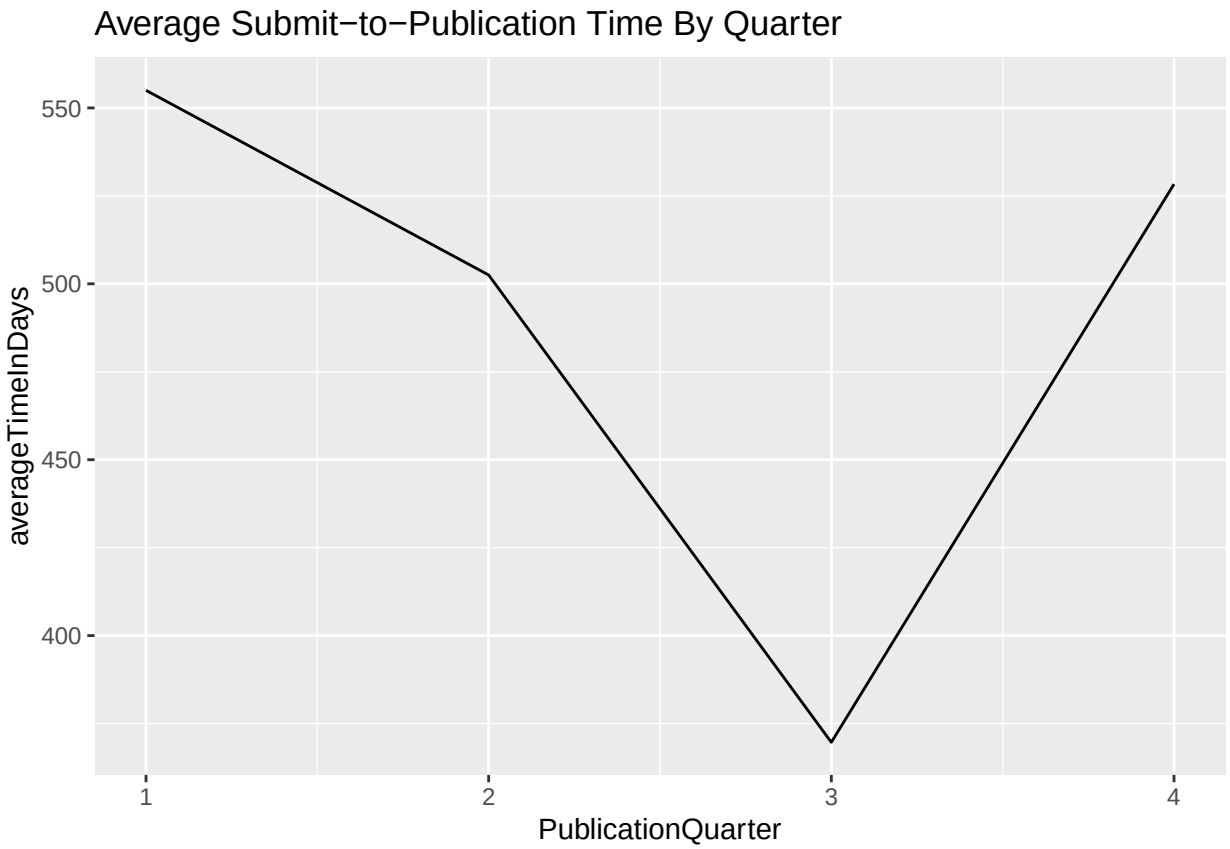



```
df2 <- dbReadTable(dbcon, "publishedArticlesAuthorFact")
res2 <- sqldf("SELECT SUM(articlesPublished) as articlesPublished, year_num FROM df2 GROUP BY year_num"
year <- res2$year_num
articlesPublished <- res2$articlesPublished
ggplot(res2, aes(x=year, y=articlesPublished)) +
  geom_line() + ggtitle("Articles Published by Authors 2011 vs. 2012")
```




```
df3 <- dbReadTable(dbcon, "publishedArticlesHistoryDimensions")
res3 <- sqldf("SELECT AVG(totalTime) as averageTime, quarter FROM df3 GROUP BY quarter")
averageTimeInDays <- res3$averageTime
PublicationQuarter <- res3$quarter

ggplot(res3, aes(x=PublicationQuarter, y=averageTimeInDays)) +
  geom_line() + ggtitle("Average Submit-to-Publication Time By Quarter")
```



I did have to make some changes to my starschema to make the last graph. I had to make a new table with history info bec
needed info about when certain parts of the publishing process occurred. I couldn't just use the normal History table becau
I need the pubdate and the submit date in the same row of my table.

```
dbDisconnect(dbcon)
```

```
## [1] TRUE
```