

week_9_assignment

Adejare Windokun

Saturday, October 25, 2014

Part 1: mongoimport Code to import files

```
C:079>mongoimport --host localhost --db unitedstates --collection states --type csv --
headerline --file statedata.csv connected to: localhost 2014-10-25T12:10:49.253-0400
check 9 51 2014-10-25T12:10:49.263-0400 imported 50 objects
```

```
C:079>mongoimport --host localhost --db unitedstates --collection states --type tsv --
headerline --file districtdata.txt connected to: localhost 2014-10-25T12:11:09.303-0400
imported 1 objects
```

```
C:079>mongoimport --host localhost --db unitedstates --collection states --type tsv --
headerline --file inhabitedterritorydata.txt connected to: localhost 2014-10-
25T12:11:18.595-0400 imported 5 objects
```

```
if (!require(rmongodb)) install.packages('rmongodb')
## Loading required package: rmongodb

library(rmongodb)
if (!require(jsonlite)) install.packages('jsonlite')
## Loading required package: jsonlite
##
## Attaching package: 'jsonlite'
##
## The following object is masked from 'package:utils':
##
##     View

library(jsonlite)
if (!require(plyr)) install.packages('plyr')
## Loading required package: plyr

library(plyr)
# install rmongodbHelper package from GitHub
# install.packages("devtools")
if (!require(devtools)) install.packages('devtools')

## Loading required package: devtools
## WARNING: Rtools is required to build R packages, but is not currently
installed.
##
```

```
## Please download and install Rtools 3.1 from http://cran.r-project.org/bin/windows/Rtools/ and then run find_rtools().
```

```
library(devtools)
```

```
if (!require(rmongodbHelper))  
devtools::install_github("joyofdata/rmongodbHelper")
```

```
## Loading required package: rmongodbHelper
```

```
library(rmongodbHelper)
```

```
# Example of the rmongodbHelper
```

```
# json_qry <-
```

```
# '{  
#   "$or": [  
#     {"a":1},  
#     {"a":3}  
#   ]  
# }'  
#
```

```
# bson <- rmongodbhelper::json_to_bson(json_qry)
```

```
connect to the mongodb database
```

```
mongo = mongo.create()  
mongo.is.connected(mongo)
```

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

```
see all the records
```

```
#No query criteria
```

```
q = "{}"  
if (mongo.is.connected(mongo)) {  
  cursor <- mongo.find(mongo, "unitedstates.states", query = q, limit =  
10L)  
  while (mongo.cursor.next(cursor))  
    print(mongo.cursor.value(cursor))  
  mongo.cursor.destroy(cursor)  
}
```

```
## _id : 7      544bce1347e597fde9f8aec0  
## territory : 2    American Samoa  
## abbr : 2      AS  
## capital : 2     Pago Pago  
## acquisition_year : 16    1900  
## territorial_status : 2    "Unincorporated, unorganized"  
## population : 16      55519  
## area_sq_miles : 16    581  
## land_area_sq_miles : 16      76  
## house_seats : 16      1
```

```

## _id : 7      544bce1447e597fde9f8aec1
## territory : 2    Guam
## abbr : 2      GU
## capital : 2     Hagatna
## acquisition_year : 16    1899
## territorial_status : 2    "Unincorporated, organized"
## population : 16      159358
## area_sq_miles : 16    571
## land_area_sq_miles : 16    210
## house_seats : 16      1
## _id : 7      544bce1447e597fde9f8aec2
## territory : 2    Northern Mariana Islands
## abbr : 2      MP
## capital : 2     Saipan
## acquisition_year : 16    1986
## territorial_status : 2    "Unincorporated, organized"
## population : 16      53883
## area_sq_miles : 16    1976
## land_area_sq_miles : 16    182
## house_seats : 16      1
## _id : 7      544bce1447e597fde9f8aec3
## territory : 2    Puerto Rico
## abbr : 2      PR
## capital : 2     San Juan
## acquisition_year : 16    1899
## territorial_status : 2    "Unincorporated, organized"
## population : 16      3725789
## area_sq_miles : 16    5325
## land_area_sq_miles : 16    3424
## house_seats : 16      1
## _id : 7      544bce1447e597fde9f8aec4
## territory : 2    U.S. Virgin Islands
## abbr : 2      VI
## capital : 2     Charlotte Amalie
## acquisition_year : 16    1917
## territorial_status : 2    "Unincorporated, organized"
## population : 16      106405
## area_sq_miles : 16    733
## land_area_sq_miles : 16    134
## house_seats : 16      1
## _id : 7      544bce3147e597fde9f8aec5
## federal_district : 2    "Washington, D.C."
## abbr : 2      DC
## establishment_date : 2    7/16/1790
## population : 16      646449
## area_sq_miles : 16    68
## land_area_sq_miles : 16    61
## house_seats : 16      1
## _id : 7      544bce4747e597fde9f8aec6
## state : 2      Alabama

```

```
## abbr : 2      AL
## capital_city : 2      Montgomery
## largest_city : 2      Birmingham
## population : 16      4833722
## area_sq_miles : 16      52420
## land_area_sq_miles : 16      50645
## house_seats : 16      7
## statehood_date : 2      12/14/1819
## _id : 7      544bce4747e597fde9f8aec7
## state : 2      Alaska
## abbr : 2      AK
## capital_city : 2      Juneau
## largest_city : 2      Anchorage
## population : 16      735132
## area_sq_miles : 16      665384
## land_area_sq_miles : 16      570641
## house_seats : 16      1
## statehood_date : 2      1/3/1959
## _id : 7      544bce4747e597fde9f8aec8
## state : 2      Arizona
## abbr : 2      AZ
## capital_city : 2      Phoenix
## largest_city : 2      Phoenix
## population : 16      6626624
## area_sq_miles : 16      113990
## land_area_sq_miles : 16      113594
## house_seats : 16      9
## statehood_date : 2      2/14/1912
## _id : 7      544bce4747e597fde9f8aec9
## state : 2      Arkansas
## abbr : 2      AR
## capital_city : 2      Little Rock
## largest_city : 2      Little Rock
## population : 16      2959373
## area_sq_miles : 16      53179
## land_area_sq_miles : 16      52035
## house_seats : 16      4
## statehood_date : 2      6/15/1836

## [1] FALSE
```

Part 2: rmongodb Query to select out the states { state: { \$exists: true } }

```
json_qry <-
  '{
    "state":
      {"$exists":true}
  }'
```

```

bson <- rmongodbHelper::json_to_bson(json_qry)

states = data.frame()
cursor <- mongo.find(mongo, "unitedstates.states", bson)
i = 1
while(mongo.cursor.next(cursor))
{
  tmp = mongo.bson.to.list(mongo.cursor.value(cursor))
  tmp.df = as.data.frame(t(unlist(tmp)), stringAsFactors = FALSE)
  tmp.df = subset(tmp.df, select = c(-1))
  states = rbind.fill(states, tmp.df)
  i = i + 1
}
head(states)

##      state abbr capital_city largest_city population area_sq_miles
## 1  Alabama  AL  Montgomery   Birmingham   4833722      52420
## 2  Alaska   AK    Juneau     Anchorage    735132      665384
## 3  Arizona  AZ    Phoenix    Phoenix     6626624     113990
## 4  Arkansas AR  Little Rock  Little Rock  2959373      53179
## 5 California CA  Sacramento Los Angeles  38332521     163695
## 6  Colorado CO    Denver     Denver     5268367     104094
##  land_area_sq_miles house_seats statehood_date
## 1          50645          7    12/14/1819
## 2          570641          1     1/3/1959
## 3          113594          9     2/14/1912
## 4           52035          4     6/15/1836
## 5          155779         53     9/9/1850
## 6          103642          7     8/1/1876

```

For the districts Query to select out the districts { federal_district: { \$exists: true } }

```

json_qry <-
'{'
  "federal_district":
    {"$exists":true}
}'

bson <- rmongodbHelper::json_to_bson(json_qry)

districts = data.frame()
cursor <- mongo.find(mongo, "unitedstates.states", bson)
i = 1
while(mongo.cursor.next(cursor))
{
  tmp = mongo.bson.to.list(mongo.cursor.value(cursor))
  tmp.df = as.data.frame(t(unlist(tmp)), stringAsFactors = FALSE)
  tmp.df = subset(tmp.df, select = c(-1))
}

```

```

    districts = rbind.fill(districts, tmp.df)
i = i + 1
}
head(districts)

##      federal_district abbr establishment_date population area_sq_miles
## 1 "Washington, D.C."  DC              7/16/1790      646449          68
##   land_area_sq_miles house_seats
## 1                  61          1

```

For territories Query to select out the territory { territory: { \$exists: true } }

```

json_qry <-
'{'
  "territory":
    {"$exists":true}
}'

bson <- rmongodbHelper::json_to_bson(json_qry)

territories = data.frame()
cursor <- mongo.find(mongo, "unitedstates.states", bson)
i = 1
while(mongo.cursor.next(cursor))
{
  tmp = mongo.bson.to.list(mongo.cursor.value(cursor))
  tmp.df = as.data.frame(t(unlist(tmp)), stringAsFactors = FALSE)
  tmp.df = subset(tmp.df, select = c(-1))
  territories = rbind.fill(territories, tmp.df)
i = i + 1
}
head(territories)

##           territory abbr      capital acquisition_year
## 1      American Samoa  AS      Pago Pago           1900
## 2                Guam  GU      Hagatna             1899
## 3 Northern Mariana Islands  MP      Saipan             1986
## 4        Puerto Rico  PR      San Juan             1899
## 5    U.S. Virgin Islands  VI Charlotte Amalie          1917
##           territorial_status population area_sq_miles
## 1 "Unincorporated, unorganized"      55519          581
## 2 "Unincorporated, organized"      159358          571
## 3 "Unincorporated, organized"      53883          1976
## 4 "Unincorporated, organized"     3725789          5325
## 5 "Unincorporated, organized"     106405           733
##   land_area_sq_miles house_seats
## 1                76          1
## 2               210          1
## 3               182          1

```

## 4	3424	1
## 5	134	1