Midterm Project

Ang Li

2017/11/13

```
#install.packages("RMySQL")  # you may have to install first
library(RMySQL)
```

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

```
mydb = dbConnect(MySQL(), user='mssp', password='mssp2017', dbname='yelp_db', host='45.63.90.29')
```

```
dbListTables(mydb)
```

```
##  [1] "attribute"    "business"     "category"     "checkin"
##  [5] "elite_years"  "friend"       "hours"        "photo"
##  [9] "review"       "tip"          "user"         "user_id_1000"
```

```
#This will return a list of the tables in our connection.
```

```
dbListFields(mydb, 'business') # eg: have a look what variables are in 'user' table
```

```
##  [1] "id"           "name"         "neighborhood" "address"
##  [5] "city"         "state"        "postal_code"  "latitude"
##  [9] "longitude"    "stars"        "review_count" "is_open"
```

```
#This will return a list of the fields(columns) in table 'user'.

#dbListFields(mydb, 'friend')
```

```
business.sql = dbSendQuery(mydb, "
                           select *
                            from business
                            " )  # still in mysql
business = fetch(business.sql, n = -1)   # fetch back to R
```

```
#head(business)
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(data.table)
```

```
##
## Attaching package: 'data.table'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
```

```
library(lme4)
```

```
## Loading required package: Matrix
```

```
library(arm)
```

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
##
## arm (Version 1.9-3, built: 2016-11-21)
```

```
## Working directory is /Users/ALLEN/Documents/684/Midterm Project
```

```
business %>%
  summarise(count_all = n_distinct(state))
```

```
##   count_all
## 1        51
```

```
fit<-lmer(stars~review_count+is_open+(1|state), data=business)
summary(fit)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: stars ~ review_count + is_open + (1 | state)
##    Data: business
##
## REML criterion at convergence: 435067.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.9116 -0.6679  0.1783  0.7782  1.7299
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  state    (Intercept) 0.01212  0.1101
##  Residual             0.94089  0.9700
## Number of obs: 156639, groups:  state, 51
##
## Fixed effects:
##               Estimate Std. Error t value
## (Intercept)  3.555e+00  2.808e-02  126.58
## review_count 1.816e-04  2.564e-05    7.08
## is_open      1.473e-01  6.772e-03   21.75
##
## Correlation of Fixed Effects:
##             (Intr) rvw_cn
## review_cont -0.010
## is_open     -0.208 -0.036
```