Record Linkage

Yue Xiong*

Chapter 04 Record Linkage

Not sure whether we have similar packages in R.

```r
# install.packages("DBI")
library("RSQLite")
```

```
## Warning: package 'RSQLite' was built under R version 4.0.5
```

```r
library("DBI")
```

```
## Warning: package 'DBI' was built under R version 4.0.5
```

```r
library('dbplyr')
```

```
## Warning: package 'dbplyr' was built under R version 4.0.5
```

```r
library('dplyr')
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

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

```
## The following objects are masked from 'package:dbplyr':
##
##     ident, sql
```

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

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

Now we can create the connection to the corresponding mammals database.

---

*LMU, yue.xiong@stat.uni-muenchen.de

```
database_path = "F:/hiwi_work_notebook/bdss-notebooks/R_Notebooks/data_raw/portal_mammals.sqlite"
conn = dbConnect(SQLite(), database_path)
src_dbi(conn)
```

```
## src:  sqlite 3.36.0 [F:\hiwi_work_notebook\bdss-notebooks\R_Notebooks\data_raw\portal_mammals.sqlite]
## tbls: plots, species, surveys
```

Then we can fetch all query results into a normal r dataframe using `dbGetQuery` provided by the DBI package.

```
mammals_list = dbGetQuery(conn, "select * from surveys")
mammals_df = data.frame(mammals_list)
nrow(mammals_df)
```

```
## [1] 35549
```