

intro to taxizedb

Scott Chamberlain

install

```
devtools::install_github("ropenscilabs/taxizedb")
```

```
library("taxizedb")  
library("dplyr")
```

Download SQL DBs

- MySQL - <http://dev.mysql.com/downloads/installer/>
- PostgreSQL - <http://www.postgresql.org/download>

start your SQL DBs

Remember to start your PostgreSQL database for ITIS and ThePlantList and your MySQL database for COL

Download and load DBs

ITIS

```
x <- db_download_itis()  
db_load_itis(x)
```

The Plant List (TPL)

```
x <- db_download_tpl()  
db_load_tpl(x)
```

Cataloge of Life (COL)

```
x <- db_download_itis()  
db_load_itis(x)
```

connect to the DBs

ITIS

```
src <- src_itis()
```

TPL

```
src <- src_tpl()
```

COL

```
src <- src_col()
```

query with SQL syntax

```
sql_collect(src, "select * from hierarchy limit 5")
#> Source: local data frame [5 x 5]
#>
#>      hierarchy_string    tsn parent_tsn level childrencount
#>      <chr>    <int>      <int> <int>      <int>
#> 1          202422 202422         0     0          145306
#> 2      202422-846491 846491      202422     1           2666
#> 3      202422-846491-660046 660046      846491     2           2654
#> 4      202422-846491-660046-846497 846497      660046     3             7
#> 5 202422-846491-660046-846497-846508 846508      846497     4             6
```

```
# or pipe the src to sql_collect
src %>% sql_collect("select * from hierarchy limit 5")
#> Source: local data frame [5 x 5]
#>
#>      hierarchy_string    tsn parent_tsn level childrencount
#>      <chr>    <int>      <int> <int>      <int>
#> 1          202422 202422         0     0          145306
#> 2      202422-846491 846491      202422     1           2666
#> 3      202422-846491-660046 660046      846491     2           2654
#> 4      202422-846491-660046-846497 846497      660046     3             7
#> 5 202422-846491-660046-846497-846508 846508      846497     4             6
```

use dplyr verbs

get a tbl

```
hiers <- src %>% tbl("hierarchy")
```

select certain fields

```
hiers %>% select(tsn, level)
#> Source:   query [?? x 2]
#> Database: postgres 9.4.5 [sacmac@localhost:5432/ITIS]
#>
#>      tsn level
#>      <int> <int>
#> 1  202422     0
#> 2  846491     1
#> 3  660046     2
#> 4  846497     3
```

```
#> 5 846508 4
#> 6 846553 5
#> 7 954935 6
#> 8 5549 7
#> 9 5550 8
#> 10 954936 6
#> .. ...
```