

Exploratory Data Analysis with R

Microsoft Power BI - Part II

Xuemao Zhang
East Stroudsburg University

November 30, 2022

Outline

- Introduction
- Connect PowerBI to a database
- Visualizing mammals data

Introduction

- We'll connect the mammals database in PowerBI and create several dashboards

```
ALTER TABLE surveys  
DROP COLUMN IF EXISTS tail_length;  
  
DROP TABLE if exists surveys_weight;
```

- We first update the database Database1 by combining the three columns month, day and year in table surveys

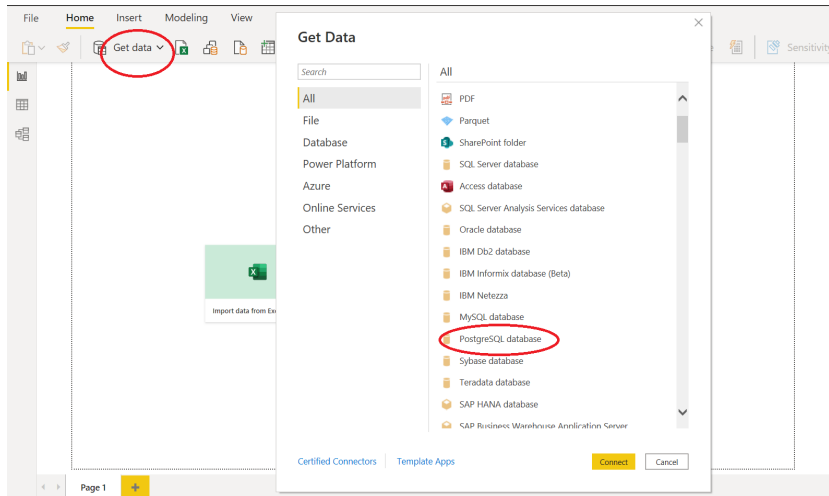
Introduction

```
/*Add a column: time*/  
ALTER TABLE surveys  
ADD COLUMN time date;  
  
/*update the column value*/  
UPDATE surveys  
SET time = make_date(year, month, day);  
  
/*remove the NULL rows in weight and hindfoot_length*/  
SELECT * INTO surveys_weight  
FROM surveys  
WHERE weight is NOT NULL AND hindfoot_length is NOT NULL;  
  
SELECT *  
FROM surveys_weight limit 10;
```

- **Remark:** If we do not conduct the SQL queries above, we can do this in PowerBI using Power Query.

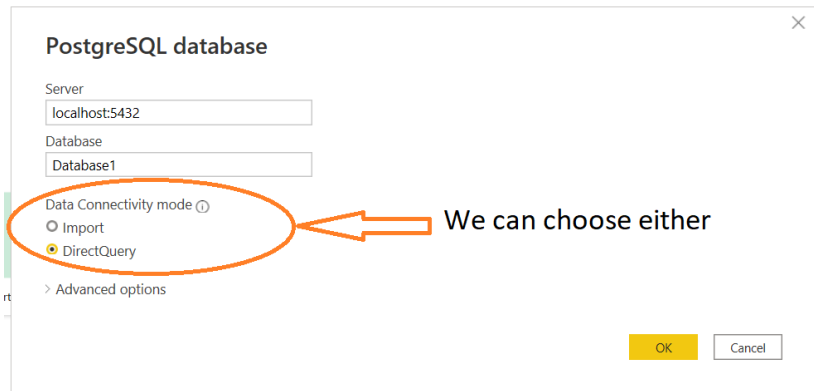
Connect PowerBI to a database

- We'll connect a postgresSQL database
 - ▶ You can read [Connecting PostgreSQL to PowerBI](https://help.scalegrid.io/docs/connecting-postgresql-to-powerbi)
<https://help.scalegrid.io/docs/connecting-postgresql-to-powerbi>



Connect PowerBI to a database

- Enter Server address and the name of your database
 - ▶ Then choose either connectivity mode



PostgreSQL database

Server
localhost:5432

Database
Database1

Data Connectivity mode (D)

☐ Import

☒ DirectQuery

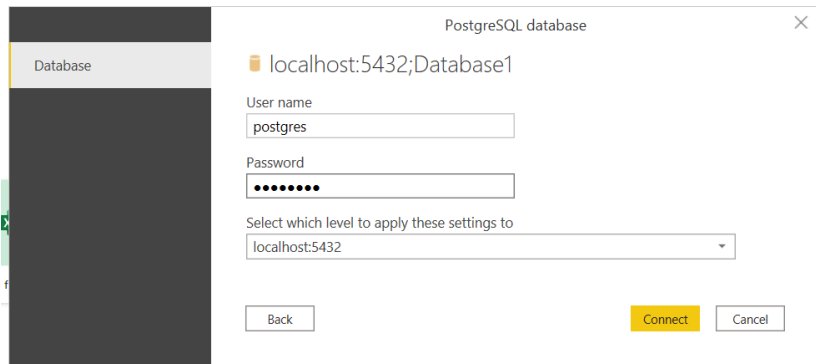
> Advanced options

OK Cancel

We can choose either

Connect PowerBI to a database

- Enter User name and password if asked



The screenshot shows a 'PostgreSQL database' connection dialog box. On the left is a dark sidebar with a 'Database' tab highlighted. The main area of the dialog displays the connection string 'localhost:5432;Database1' with a folder icon. Below this are fields for 'User name' (containing 'postgres') and 'Password' (masked with dots). A dropdown menu labeled 'Select which level to apply these settings to' is set to 'localhost:5432'. At the bottom are three buttons: 'Back', 'Connect' (highlighted in yellow), and 'Cancel'.

PostgreSQL database

Database

localhost:5432;Database1

User name
postgres

Password
●●●●●●●●

Select which level to apply these settings to
localhost:5432

Back Connect Cancel

Connect PowerBI to a database

- Choose the tables to be loaded

The screenshot shows the PowerBI Navigator window. On the left, under 'Display Options', the tables 'public.plots', 'public.species', 'public.surveys', and 'public.surveys_weight' are listed with checkboxes. The 'public.surveys_weight' table is selected and highlighted. On the right, a preview of the 'public.surveys_weight' table is shown with columns: record_id, month, day, year, plot_id, species_id, and sex. The data rows show survey records from 1977.

Navigator

Display Options ▾

- localhost:5432: Database1 [4]
 - ☒ public.plots
 - ☒ public.species
 - ☒ public.surveys
 - ☒ public.surveys_weight

public.surveys_weight

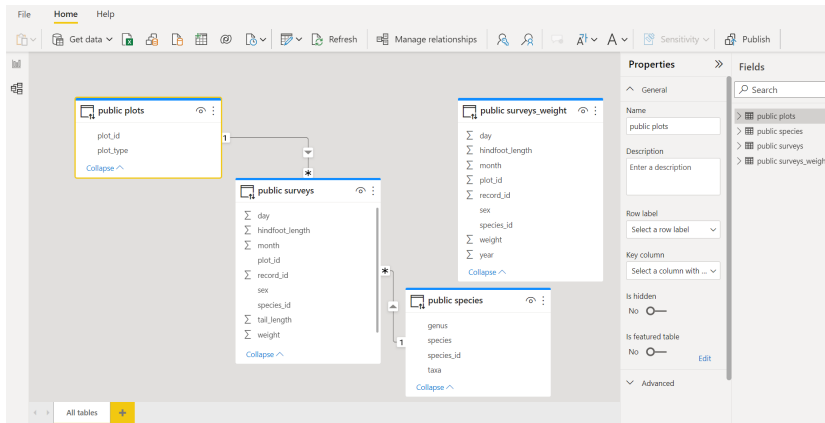
record_id	month	day	year	plot_id	species_id	sex
63	8	19	1977	3	DM	M
64	8	19	1977	7	DM	M
65	8	19	1977	4	DM	F
66	8	19	1977	4	DM	F
67	8	19	1977	7	DM	M
68	8	19	1977	8	DO	F
69	8	19	1977	2	PF	M
70	8	19	1977	3	OX	F
71	8	19	1977	7	DM	F
74	8	19	1977	8	PF	M
75	8	19	1977	8	DM	F
78	8	19	1977	1	PF	M
79	8	19	1977	7	DM	F
81	8	19	1977	4	PF	F
82	8	19	1977	4	DM	F
83	8	20	1977	6	DM	F
84	8	20	1977	19	DM	F
85	8	20	1977	23	DM	F
86	8	20	1977	18	DM	F
87	8	20	1977	5	PF	F
88	8	20	1977	18	DM	F
89	8	20	1977	12	PP	F
90	8	20	1977	18	DM	M

Select Related Tables

Load Transform Data Cancel

Connect PowerBI to a database

- We won't see the data in PowerBI if the connectivity mode is DirectQuery
- We can create/update the data model in PowerBI. The following is the model inherited from the database



Visualizing mammals data

- **Scatter Plot**

Visualizing mammals data

- Time Series

Visualizing mammals data

- **Box plots** by species_id with R

The following code to create a dataframe and remove duplicated rows is al

```
# dataset <- data.frame(hindfoot_length, plot_id, sex, species_id, weight)
# dataset <- unique(dataset)
```

Paste or type your script code here:

```
library(ggplot2)
library(gridExtra)
p1=ggplot(data = dataset, aes(x = species_id, y = weight,
                             fill=species_id)) +
  geom_boxplot()+
  labs(title='Box plot of weight by species');
p2=ggplot(data = dataset, aes(x = species_id, y = hindfoot_length,
                             fill=species_id)) +
  geom_boxplot()+
  labs(title='Box plot of hindfoot_length by species');
grid.arrange( p1, p2, ncol=1)
```

License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).