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`

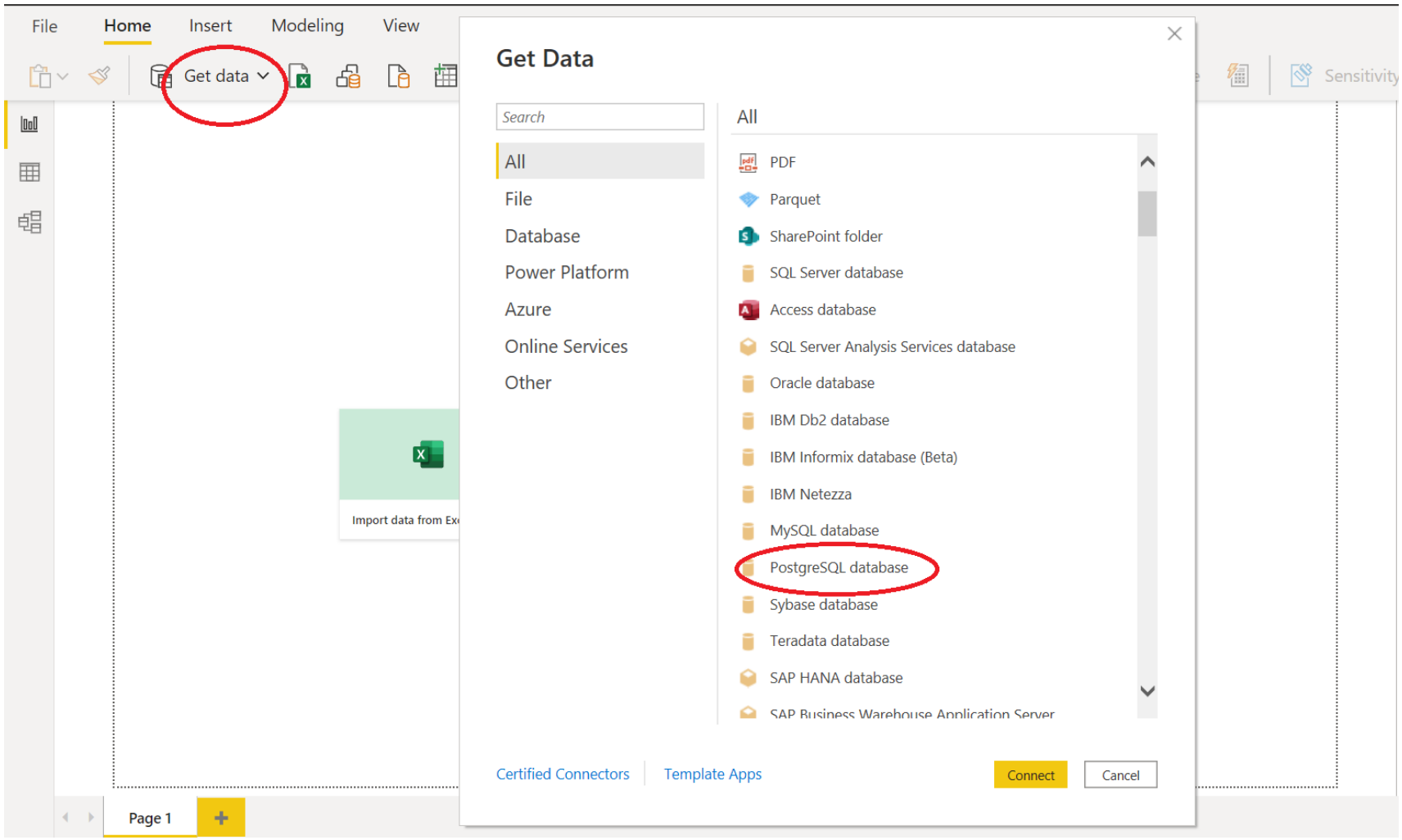
Introduction

```
/*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.

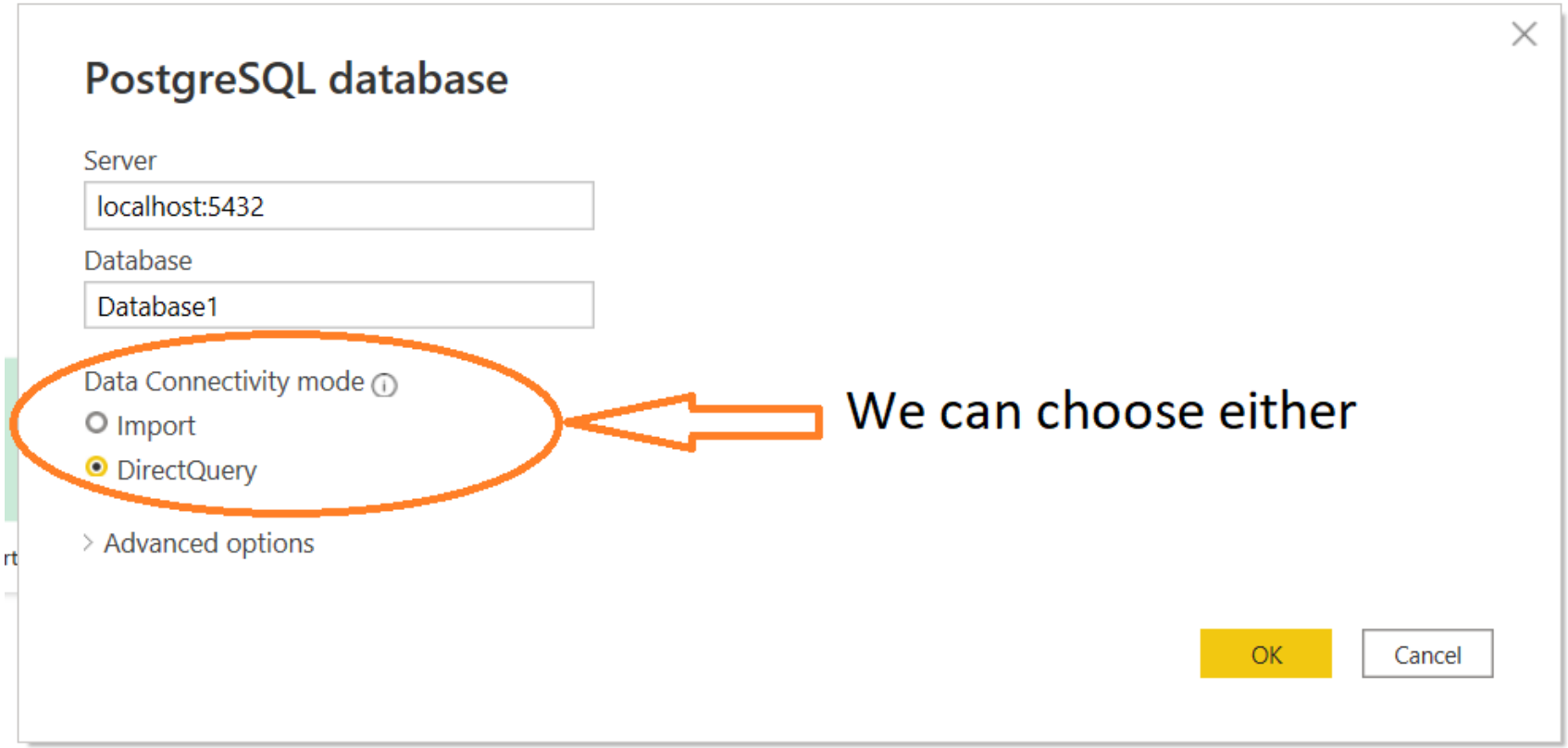
Connect PowerBI to a database

- We'll connect a postgreSQL database
  - You can read Connecting PostgreSQL to PowerBI https://help.scalegrid.io/docs/connecting-postgresql-to-powerbi



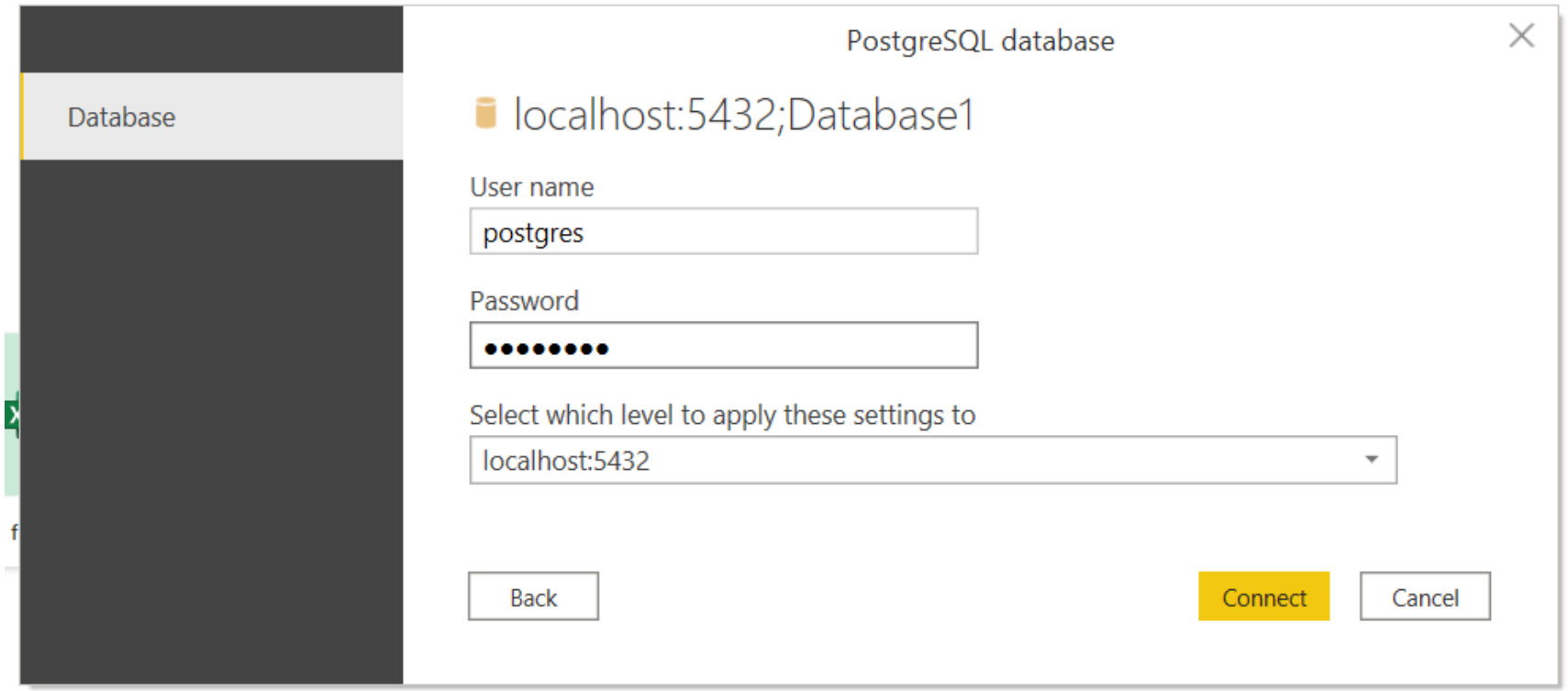
Connect PowerBI to a database

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



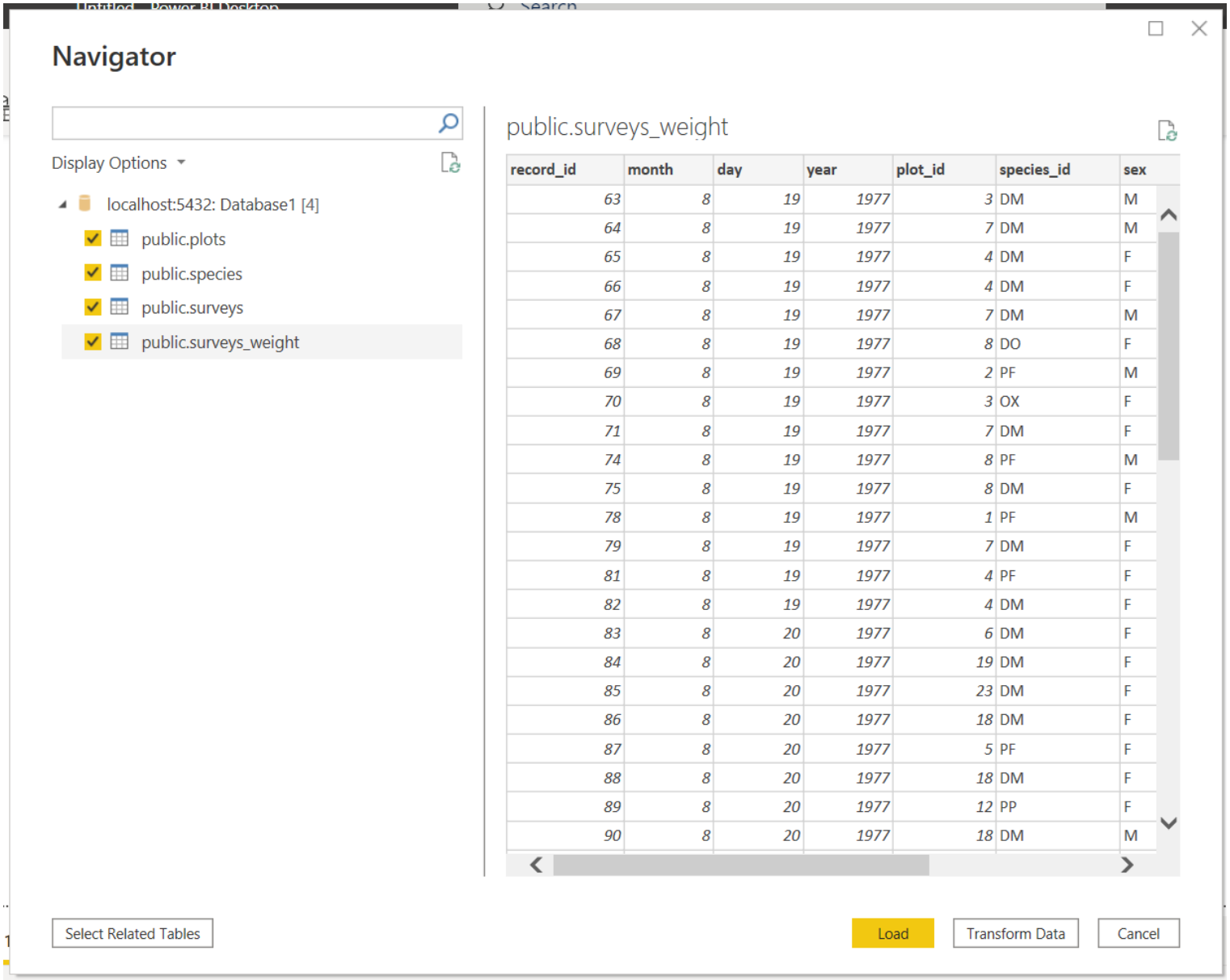
Connect PowerBI to a database

- Enter User name and password if asked



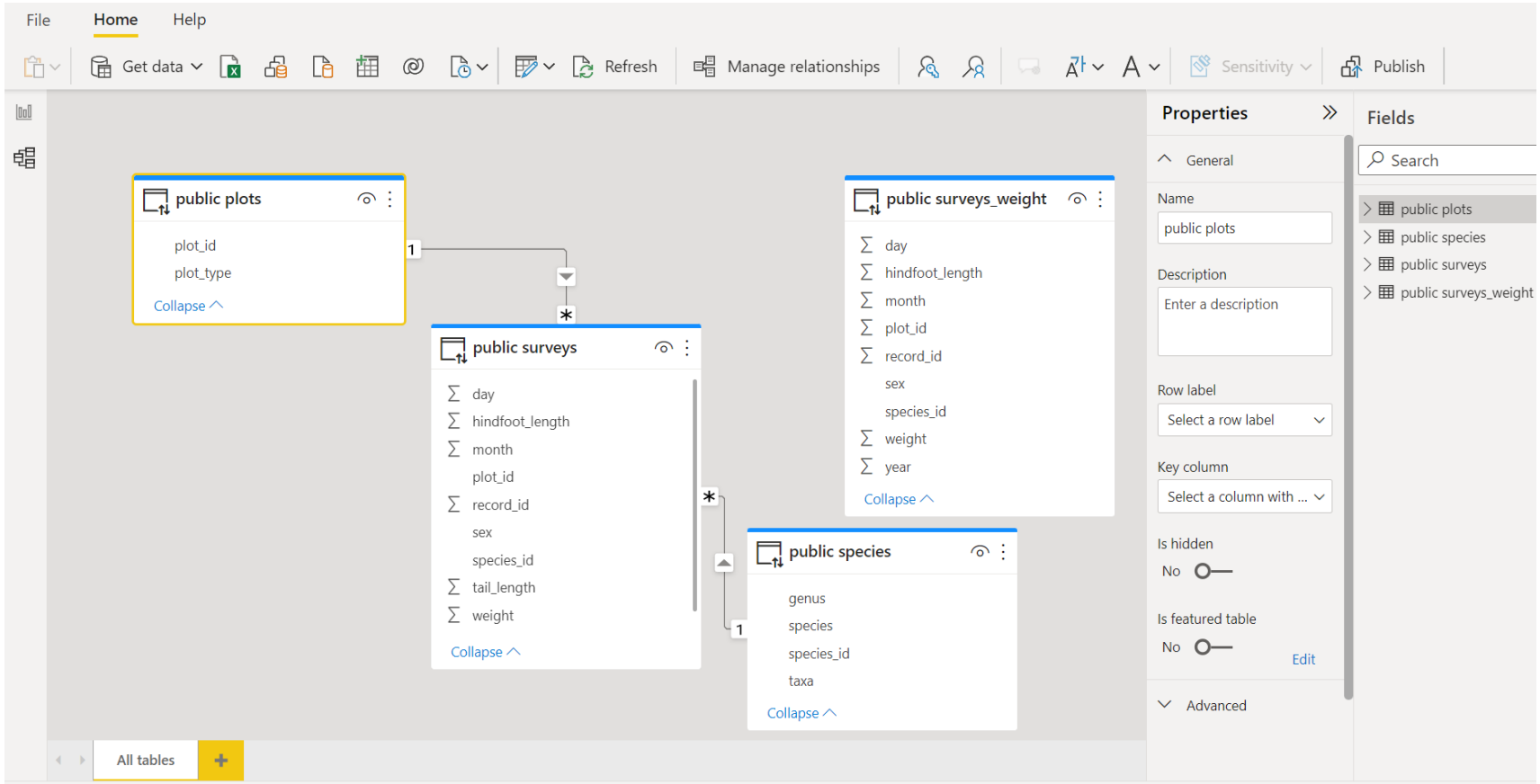
Connect PowerBI to a database

- Choose the tables to be loaded



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