[180 min] DO: Practicum III / Mine a Database

Due Aug 15 by 9:29am **Points** 100 **Submitting** a file upload **File Types** pdf, rmd, and zip
Available until Aug 17 at 9:29am

This assignment was locked Aug 17 at 9:29am.

Format

Method: Individual
Materials: RStudio or RStudio.cloud; SQLite

Instructions

Do the following activities in an R Notebook:

1. Inspect the Plant Catalog XML (**plants.xml**).
2. Load the Plant Catalog XML directly into a dataframe using *xmlToDataFrame*.
3. Create a new column retail that is a numeric column and has the data from the price column; note that the price column is text and contains a leading '$' and needs to be parsed properly.
4. Remove the original price column (the one that is text) from the dataframe.
5. Update all prices: increase them by 4.25%.
6. Create a histogram of the number of plants by price. Adorn the chart -- do some research on what you can do.
7. Using *sqldf* create a SQL query that finds the common names of all plants that cost less than $8 and grow in Sunny light.
8. Using either *sqldf* or dataframe functions, how many plants grow in full shade?
9. Write the dataframe to a new table in a new SQLite database using the *dbWriteTable* function.

In SQLite (or SQLiteOnline):

1. Inspect the database created in (9) above. Is the table there? What is its name? Are the columns correct?
2. Write a SQL UPDATE statement that decreases all prices by 0.75%.

In R Studio:

1. Connect to the SQLite database from above (if you used SQLiteOnline, you need to save the database file locally and upload to RStudio.cloud if you use the cloud version of R Studio).
2. Build and execute a SQL query that finds the number of plants by light.

In R Studio:

1. Parse the Plant Catalog XML and write an XPath query that finds all plants that grow in full shade or in full sun.
2. Put the result from the previous XPath query into a dataframe and then calculate the 10% trimmed mean of the prices.

Submission

Submit the Rmd file, a knitted HTML or PDF of your R Notebook (if HTML, zip the file), and screen shot of the database inspected in SQLite (online or desktop).

Practicum III Rubric

Criteria	Ratings			Pts
Load the Plant Catalog XML directly into a dataframe using xmlToDataFrame.	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Create a new column retail that is a numeric column and has the data from the price column; note that the price column is text and contains a leading '$' and needs to be parsed properly.	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Remove the original price column (the one that is text) from the dataframe.	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Update all prices: increase them by 4.25%.	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Create a histogram of the number of plants by price.	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Using sqldf create a SQL query that finds the common names of all plants that cost less than $8 and grow in Sunny light.	**10.0 pts** **Full Marks**	**5.0 pts** **Acceptable**	**0.0 pts** **No Marks**	10.0 pts
Using either sqldf or dataframe functions, how many plants grow in full shade?	**5.0 pts** **Full Marks**	**3.0 pts** **Adequate**	**0.0 pts** **No Marks**	5.0 pts
Write the dataframe to a new table in a new SQLite database using the dbWriteTable function.	**5.0 pts** **Full Marks**	**0.0 pts** **No Marks**		5.0 pts
Write a SQL UPDATE statement that decreases all prices by 0.75%. All work is shown.	**5.0 pts** **Full Marks**	**3.0 pts** **Acceptable**	**0.0 pts** **No Marks**	5.0 pts
Connect to the SQLite database from R	**5.0 pts** **Full Marks**	**0.0 pts** **No Marks**		5.0 pts
Build and execute a SQL query in R that finds the number of plants by light.	**15.0 pts** **Full Marks**	**10.0 pts** **Adequate**	**5.0 pts** **Acceptable**	**0.0 pts** **No Marks**	15.0 pts
Parse the Plant Catalog XML and write an XPath query that finds all plants that grow in full shade or in full sun.	**20.0 pts** **Full Marks**	**15.0 pts** **Good**	**8.0 pts** **Adequate**	**4.0 pts** **Poor but has some merit**	**0.0 pts** **No Marks**	20.0 pts
Put the result from the previous XPath query into a dataframe.	**5.0 pts** **Full Marks**	**0.0 pts** **No Marks**				5.0 pts
Calculate the 10% trimmed mean of the prices.	**5.0 pts** **Full Marks**	**0.0 pts** **No Marks**				5.0 pts
					Total Points: 100.0	