

Branch: master ▾

Find file

Copy path

## 5YrMonitoringRpt / Project.md



marybecker Update Project.md

7ab96bb 2 days ago

[1 contributor](#)

Raw

Blame

History



85 lines (59 sloc) 5.73 KB

### Project Description:

Quite awhile ago now, the Monitoring group put together a [5 year summary of our inland water data for the years 2006 – 2010](#) . It contains information on the types of data we collect and a summary of the range of conditions we saw in CT inland waters over that period of time. To date, we have not updated this report or completed another report like this for 2011 – 2015. My goal is to create a pared down web-based report that can be automatically updated in the future as we collect more information. We will also only focused on our ambient water collection in wadable rivers and streams. This will include our water chemistry data (e.g. Total Phosphorus, Chloride, Turbidity etc.), biological data (e.g. bugs, fish, diatoms) and water temperature data. In addition, some of the calculations we will put together will help inform some of our QA process going forward.

We will be mostly be using R and SQL and then maybe a little HTML/CSS/Javascript to complete the project.

To start we will focus on the water chemistry data. I have compiled and cleaned up data needed for this portion of the project. Sample collection data can be found in the [data folder: monrpt.db](#) file. This is a sqlite database which can be queried in R (see examples in '5YrMonRpt\_Summary.R') or by using an SQLite database manager. I suggest [SQLite studio github link for latest release](#) . SQLite studio is a good place to start playing around with some SQL queries and getting familiar with the data. [Below](#) is a description of the tables and fields in the tables. You should have a better idea for the data and fields below after reading through the [2006 – 2010 report](#) . If you are not familiar with our drainage basin system (i.e. watersheds) you can find some additional information on our CT eco page maintained by UCONN Clear: [Major Basins Subregional Basins](#) .

In addition to the database, there is also a csv file in the data folder called [site\\_env.csv](#) . This contains landscape information for each of the sites such as drainage area, total upstream stream miles, percent impervious cover etc. Included [below](#) is a description of some of relevant fields we will use for the project.

See [below](#) for some ideas to get started. We first want to get gain familiarity with the data and get comfortable with SQL, R and some programming foundations.

## Getting Started

---

1. Read through the [2006 – 2010 report](#)
2. Clone [5YrMonitoringRpt GitHub repository](#) to a local repository where you will be working.
3. Practice some select SQL queries using SQLite in a SQLite database manager or R.
4. Go through the examples in the '[5YrMonRpt\\_Summary.R](#)' script .
5. Try writing some code to answer the questions at the bottom of the script.
6. As you work, push changes to the repo frequently! By doing so I can peak in to help as needed and collaborate.
7. Post your progress, changes, comments, questions, thoughts or anything else project related on the slack channel.

## Description of data in DB Tables

## chemdata

- sta\_seq – unique site identifier (pk)
- lab\_accession – accession number from the laboratory (pk)
- collect\_date – data sample was collected
- chemparameter - the chemical parameter that was collected (pk)
- value – the chemical parameter value
- uom – the chemical parameter value unit of measure
- relt\_det\_cond – the result detection condition (less than the minimum detection limit)
- method\_det\_limit – the method detection limit (not available for all data use the mdl table)
- activity\_type – activity type
- activity\_name – activity name
- station\_type – the type of station
- duplicate – 0 or 1 (no, yes) notes if it is a duplicate sample (sample taken for QA purposes) (pk)

## sites

- sta\_seq – unique site identifier (pk)
- name – name of waterbody
- descript – description of waterbody location
- town – town waterbody is located in
- adb\_seq – assessment database segment number
- ylat – site latitude
- xlong – site longitude
- sbasn – subregional basin number

## basin

- sbasn – subregional basin number (pk)
- subregion – subregional basin name
- rbasn - regional basin number
- regional – regional basin name
- mbasn – major basin number
- major – major basin name

## mdl

- chemparameter - the chemical parameter that was collected (pk)
- MDL – the minimum detection limit for a parameter

## **Landscape database (relevant fields we will use in this project below)**

- sta\_seq – unique site identifier (pk)
- name – name of waterbody
- ylat – site latitude
- xlong – site longitude
- Strahler – strahler stream order
- StrMi – total upstream stream miles
- SqMi – total upstream drainage basin in miles
- IC\_Avg – Percent impervious cover in upstream drainage basin
- SD\_Pct - Percent stratified drift