APCD Import Instructions

# Requirements

* RStudio (<https://www.rstudio.com/>)
* Rtools (<https://cran.r-project.org/bin/windows/Rtools/>)
* ODBC Driver for SQL Server (<https://www.microsoft.com/en-us/download/details.aspx?id=56567>)
* Some sort of SQL Server software (ex: Microsoft SQL Server Management Studio, MSSMS)
  + Permissions: SELECT, UPDATE, INSERT, CREATE, DROP, ALTER, EXEC
* User DSN set up for the SQL Server you are using (ODBC)
* System or User Path variable set up for BCP and Rtools functions
* Schemas final, ref, and stage set up properly on the SQL Server

# Instructions

1. Download and install RStudio
2. Determine the version of R you have installed
   1. Open RStudio
   2. Tools
   3. Global Options
   4. R version is listed in the General – Basic tab
3. Download and install the version or Rtools for your version of R
4. Add directories for BCP and Rtools to Path variable
   1. Open System Properties
   2. Advanced tab
   3. Environment Variables
   4. Under System variables, find the Rtools variable and note the directory listed in the value
   5. Edit the Path variable
   6. New > Rtools directory\usr\bin
   7. New > C:\Windows\System32
5. Create User DSN for the SQL Server
   1. Open ODBC Data Source Administrator
   2. Add
   3. Select ODBC Driver for SQL Server
   4. Give the ODBC a name and point to your SQL Server
   5. Select the authentication method (confirm with your IT department)
   6. Set the default database
   7. Set any other settings (confirm with your IT department)
   8. Test Data Source
6. Log into <https://mft.kingcounty.gov/APDEDE-WA-APCD/WA-APCD/documentation>
7. Download the contents of the documentation folder and save to the directory you plan to run your R scripts from. You could create a new project in RStudio and then copy these files into that project’s directory.
8. Update the apcd\_import\_config.yaml file
   1. table\_file\_path – name of the APCD\_tables\_########.xlsx file (this will update for each data set)
   2. ftp\_keyring – name this whatever you want or leave it as the default
   3. db\_name
   4. server\_path
   5. odbc\_name
   6. base\_dir – the directory you plan to store all of the raw GZ files
   7. ref\_dir, stage\_dir, final\_dir – subdirectories in the base\_dir, you can leave these as the default
9. Open apcd\_import.R with RStudio
   1. Step 1
      1. Loads the config file and functions needed to import
      2. Defines directory variables
      3. Create the ETL log table if one does not exist
   2. Step 2
      1. Creates the keyring (login credentials) for the FTP if it does not exist
      2. Gets a list of files from the FTP
      3. Creates ETL log entries for each new file
   3. Step 3
      1. Select schemas to download
      2. Select tables from the selected schemas to download
      3. Downloads files to the defined directories
   4. Step 4
      1. Select schemas to import
      2. Select tables from the selected schemas to import
      3. Imports tables
         1. Extracts data
         2. Counts rows in file
         3. Deletes old archive tables
         4. Archives old tables
         5. Creates new table
         6. Loads data to new table
         7. Counts row in table
         8. Deletes the extracted data
         9. Records everything to the ETL log