Instructions for Calculating LCMRLs using RStudio

# Introduction

This document describes the procedure for calculating the Lowest Concentration Minimum Reporting Level (LCMRL) using the free software RStudio. The instructions for the calculations are divided into three parts: 1.) downloading required software, 2.) testing and using the LCMRL program within RStudio, and 3.) the procedure for collecting LCMRL data.

# Definitions

## R

Formally called The R Project for Statistical Computing, R is a free software environment for statistical computing and graphics.

## RStudio

RStudio is an integrated development environment (IDE) for R. It is also free software. Although LCMRLs can be calculated in R, RStudio makes downloading the required statistical function packages more user friendly.

## Working Directory

The directory where the LCMRL source code, LCMRL input files, and LCMRL output files reside.

## Input File for LCMRL Data

The input file is a Microsoft Excel comma separated values (.csv) file containing the results of the LRB and LFB samples for a partial or completed LCMRL procedure. These data must be arranged in a specific format within the .csv file that is compatible with the LCMRL program ([Sect. 3.8](#_Creating_the_Input)).

## Laboratory Fortified Blank (LFB)

The laboratory fortified blank (LFB) is an aliquot of reagent water, containing method preservatives, to which known quantities of the method analytes are added. The concentration of the analytes in an LFB is called a “spiking level” in the terminology of the LCMRL procedure ([Sect. 5](#_Procedure_for_Collecting)).

## Laboratory Reagent Blank (LRB)

The laboratory reagent blank (LRB) is an aliquot of reagent water containing the method preservatives. LRBs are used to collect data for the required “zero spiking level” in the LCMRL procedure ([Sect. 5](#_Procedure_for_Collecting)).

## Lowest Concentration Minimum Reporting Level (LCMRL)

The single-laboratory LCMRL is the lowest spiking concentration such that the probability of spike recovery in the 50% to 150% range is at least 99%.1,2

## LCMRL Scripts

Scripts are command lines to execute the code for calculating LCMRLs and constructing the LCMRL graphs. These scripts output the LCMRL results as a Microsoft Excel .csv file in tabular format to the working directory and the LCMRL graphs as a PDF to the working directory.

## Source File

The source file contains the code, or program, for calculating LCMRLs in RStudio. The source file, 100518 MRL.LCMRL.Stats.r, is available from [EPA’s LCMRL web page](https://www.epa.gov/dwanalyticalmethods/lowest-concentration-minimum-reporting-level-lcmrl-calculator).

## Packages

Packages are shared code available from the R Project website for performing statistical functions in R. The LCMRL program requires two packages: *lattice* and *car*.

## Test Data

The results of a completed LCMRL procedure arranged in a properly formatted input file (.csv) that is available from [EPA’s LCMRL web page](https://www.epa.gov/dwanalyticalmethods/lowest-concentration-minimum-reporting-level-lcmrl-calculator). The test data file is used to verify proper download and operation of the LCMRL program within RStudio.

# Creating the RStudio Computing Environment

The program, or code, for calculating LCMRLs runs within the RStudio statistical computing environment. Two free software downloads, R and RStudio, and two free packages, *lattice* and *car*, are required from non-EPA sources. EPA provides the instruction file (this document), source file ([Sect. 2.9](#_Source_File)) and test data file ([Sect 2.11](#_Test_Data)).

## R Download

Go to <https://cran.mtu.edu/>. This will take you to a webpage titled *The Comprehensive R Archive Network.*

In the framed text on this page, click on **Download R for Windows**. This will take you to a webpage titled *R for Windows*. Find and select the link, **Install R for the First Time**.

On the next webpage, click on the link, **Download R 4.0.0 for Windows**. An executable file, R-4.0.0-Win.exe (current version; accessed June 2020), will download to your computer.

Open this file and follow the steps in the install wizard. Install only the 64-bit version. R will install at the default location, C:\PROGRAM Files\R\R-4.0.0.

## RStudio Download

Go to the *Download RStudio* webpage, <https://rstudio.com/products/rstudio/download/>. Follow the links for the RStudio Desktop open source license. Select the Windows operating system, currently Windows 10/8/7, to download the executable file, RStudio-1.2.5042.exe (current version; accessed June 2020).

Open this file and follow the steps in the install wizard. RStudio will install at the default location, C:\PROGRAM Files\RStudio. Create a RStudio desktop icon or pin RStudio to the taskbar.

## Create the Working Directory

### Default Working Directory

The default working directory built into the source file is C:\Data. Create this folder as instructed in in this section prior to continuing.

Using Windows File Explorer, open the operating system (OS) directory (C:\). Click on the **New Folder** button in the ribbon at the top of the window. A folder will appear with the name “New Folder”. Rename this folder “Data”, matching case and with just these four characters. Figure 1. shows the correct location, or path, for the working directory.

Download this help file, the LCMRL source code (100518 MRL.LCMRL.Stats.r) ([Sect. 3.5](#_Downloading_LCMRL_Source)), and test data (R LCMRL Test.csv) ([Sect. 3.6](#_Downloading_the_Test)) into this directory.

### Changing the Working Directory

If you wish to change the working directory, to a USB drive for example, create the “Data” folder at that location; however, avoid the Windows Programs folder. On some systems, an administrative privileges setting may block the LCMRL script from writing to that location.

From the **Session** drop down menu in RStudio, select **Set Working Directory** > **Chose Directory**. Browse to the location of the new working directory, click on the “Data” folder, then click **Open**. The command line in the RStudio Console will echo the command, for example, “setwd (‘F:/Data’)”. The program will reset to the default directory each time RStudio is opened. Therefore, if C:\Data is not used as the working directory, follow the steps in this section to set the working directory prior to every session.

## Download the Required Packages

Start RStudio. It is not necessary to start R. The RStudio console will appear with a message (version number and disclaimer) followed by a command prompt indicated by “>” (Figure 2). Under the **Tools** dropdown menu, select **Install Packages**. In the **Install Packages** dialog box verify that the **Install From** drop-down list is set to “Repository CRAN”. In the **Packages** text-input box, type *lattice*. Important: The **Install dependencies** check box must be selected. Click the **Install** button. Ignore the warning message that Rtools is required and should be downloaded. R will return this message indicating a successful install: “package ‘lattice’ successfully unpacked and MD5 sums checked”.

Repeat the instructions above for the package, *car*. Ignore the warning message that Rtools is required and should be downloaded. After a long list of actions, R will return this message indicating a successful install: “package ‘car’ successfully unpacked and MD5 sums checked”.

## Downloading LCMRL Source Code

Download the LCMRL source code file, 100518 MRL.LCMRL.Stats.r, from [EPA’s LCMRL web page](https://www.epa.gov/dwanalyticalmethods/lowest-concentration-minimum-reporting-level-lcmrl-calculator) and save it in the working directory, for example, C:\Data.

## Downloading the Test Data

Download the input file for the test data, R LCMRL Test.csv, from [EPA’s LCMRL web page](https://www.epa.gov/dwanalyticalmethods/lowest-concentration-minimum-reporting-level-lcmrl-calculator) and save it in the working directory, for example, C:\Data.

## Verify Operation

Follow the steps in [Section 4](#_Running_the_LCMRL) to calculate LCMRLs for the five analytes in the test data input file. After processing is complete, open the working directory, C:\Data, and find the file named “LCMRL.values.R LCMRL Test.csv. Verify that the LCMRLs in Column B and the messages in Column G are identical to those listed in Table 1. Ignore the information in the other columns. Two of the analytes, Analyte 2 and Analyte 4, should return error messages stating that an additional spiking level is needed to determine a valid LCMRL. These represent the two possible error messages for an incomplete LCMRL determination.

For Analyte 2, an estimated LCMRL of 1.3 ng/L is reported with the message, “Lower spiking level needed to bracket LCMRL”. [See Section 5.2.3](#_When_LCMRL_Estimate_1) for more information on this circumstance. For Analyte 4, the LCMRL is reported as “0” with the message, “LCMRL is above highest spiking level”. [See Section 5.2.4](#_When_LCMRL_Estimate) for more information on this circumstance.

Follow the steps in [Section 4.6.1](#_Graphs_for_Data) to generate graphs for the test data. Open the working directory, C:\Data, and find the file named “R LCMRL Test.LCMRL.Graphs.pdf”. Two graphs for each analyte should appear: the QC Interval Coverage Plot and the LCMRL Plot. Both graphs for Analyte 2, should display this error message: “LCMRL is Below Lowest Non-Zero SL”. For Analyte 4, no graphs will appear in the PDF output file.

## Creating the Input File for LCMRL Data

The input files are expected to be Microsoft Excel .csv files with a header in the first row. If the file is composed of multi-analyte data, then all analytes should be analyzed using the same method: either including non-negative values or not. The following rules apply:

Header First Row: "Analyte", "Lab", "Spike", "Result", "Dilution Factor", "Units"

Column Field Description

A Analyte name (alphanumeric; no comas should appear in the name)

B Lab name (alphanumeric; no comas should appear in the name)

C Spiking level (numeric); enter “0” for laboratory reagent blanks (LRBs)

D Measurement (numeric)

E Dilution factor (numeric). Program at this time only expects dilution factors of 1.

F Units of measurement (alphanumeric with no comas)

The method analytes and concentration levels can be inserted in any order. EPA recommends using the test file provided, R LCMRL Test.csv, as a template. Save the input files to the working directory, for example, C:\Data.

# Running the LCMRL Program within RStudio

Start RStudio. It is not necessary to start R. The RStudio console will appear with a message (version number and disclaimer) followed by a command prompt indicated by “>” (Figure 2). The following instructions use the test data, R LCMRL Test.csv, as an example. The steps required to calculate LCMRLs for your laboratory’s input file are identical.

## Load the Source File into the RStudio Interface

The source code must be loaded into RStudio prior to each session. From the **Code** drop-down menu, select **Source File**. Browse to the working directory, for example, C:\Data. Open the file, 100518 MRL.LCMRL.Stats.r. When the source file is loaded, RStudio will echo the load command and then return the message: “Loading required package: carData” (Figure 3).

## Set the Working Directory

The default working directory is C:\Data. No user actions are required if this location is used. Otherwise change the directory as instructed in [Section 3.3](#_Set_the_Working). Ensure that the source file, 100518 MRL.LCMRL.Stats.r, the test file, R LCMRL Test.csv, and the user’s data input files ([Sect. 3.8](#_Creating_the_Input)) reside in the working directory.

## Running LCMRL Scripts – Overview

For LCMRL data consisting of only positive values, three command lines are required to calculate LCMRLs and generate the graphs, each followed by a carriage return:

> fh.labdata <- "R LCMRL Test.csv "

> LCMRL.Values(fh.labdata,rnnr=1)

> LCMRL.Graphs(fh.labdata,rnnr=1)

For LCMRL data sets that include negative values, three command lines are required to calculate LCMRLs and generate the graphs, each followed by a carriage return:

> fh.labdata <- "R LCMRL Test.csv "

> LCMRL.Values(fh.labdata,rnnr=0)

> LCMRL.Graphs(fh.labdata,rnnr=0)

The command prompt “>” is not part of the script.

Experienced users will be able to complete the LCMRL calculations using this summary. Expanded instructions follow.

Note: The scripts must be entered into the RStudio console with the exact syntax shown above including the keyboard spaces. It is recommended that the user open the data directory and help file windows on the desktop with the RStudio console to facilitate copying these commands and the name of the input file rather than typing them. An example desktop would appear similar to Figure 4.

## Calculating LCMRLs for Data with only Positive Values

Copy the following command into the RStudio console inserting the name of the user input file in place of “R LCMRL Test.csv” followed by a carriage return:

> fh.labdata <- "R LCMRL Test.csv"

A fresh command prompt will appear on the line below. Copy the following LCMRL script followed by a carriage return to calculate LCMRLs:

> LCMRL .Values(fh.labdata,rnnr=1)

RStudio will return a single-line message as the LCMRL for each analyte is completed, for example, “For: Analyte 1--EPA-TSC”.

If RStudio reports an error for one of the analytes, processing will stop at that point. The most likely cause is an incorrect entry or unallowed syntax in the input file. Verify that the input file is correctly formatted as described in [Section 3.8](#_Creating_the_Input) and return to the beginning of this step.

## Calculating LCMRLs for Data with Negative Values

Copy the following command into the RStudio console inserting the name of the user input file in place of “R LCMRL Test.csv” followed by a carriage return:

> fh.labdata <- "R LCMRL Test.csv"

A fresh command prompt will appear on the line below. Copy the following LCMRL script followed by a carriage return to calculate LCMRLs:

> LCMRL .Values(fh.labdata,rnnr=0)

RStudio will return a single-line message as the LCMRL for each analyte is completed, for example, “For: Analyte 1--EPA-TSC”.

If RStudio reports an error for one of the analytes, processing will stop at that point. The most likely cause is an incorrect entry or unallowed syntax in the input file. Verify that the input file is correctly formatted as described in [Section 3.8](#_Creating_the_Input) and return to the beginning of this step.

### LCMRL Output File

If no errors occur, the program writes a new Microdoft Excel .csv file to the working directory with the analytes and their LCMRL values listed in the order they were entered into the input file. The script uses the name of the input file preceded by “LCMRL.Values.” to name the output file. If the name of the user input file is “XYZ LCMRL Data.csv”, the LCMRL result file name would be “LCMRL.values.XYZ LCMRL Data.csv”.

## Generating LCMRL Graphs

### Graphs for Data with only Positive Values

After R completes the processing for calculating LCMRLs, a fresh command line will appear. Copy the following script followed by a carriage return to generate the LCMRL graphs:

> LCMRL.Graphs(fh.labdata,rnnr=1)

RStudio will return a two-line message as each plot is constructed, for example, “For Analyte 3” and “Detection Values for Labs”.

### Graphs for Data with Negative Values

After R completes the processing for calculating LCMRLs, a fresh command line will appear. Copy the following script followed by a carriage return to generate the LCMRL graphs:

> LCMRL.Graphs(fh.labdata,rnnr=0)

RStudio will return a two-line message as each plot is constructed, for example, “Detection Values for Labs” and “For Analyte 3”.

### Output File for LCMRL Graphs

The program writes a PDF to the working directory with two graphs for each analyte. The script uses the name of the input file appended with “.LCMRL.Graphs” to name the output file. If the name of the user input file is “XYZ LCMRL Data.csv”, the PDF name would be “XYZ LCMRL Data.LCMRL.Graphs.csv”.

# Procedure for Collecting LCMRL Data

The LCMRL Procedure requires a minimum of four LFBs at each of seven concentrations, or “spiking levels”. These LFBs, plus a minimum of four LRBs, or “zero spiking level”, are processed through the entire method procedure. All method specified steps, such as sample extraction and sample preservation, must be included.

The following subsections provide detailed instruction for conducting the LCMRL study. An overview is provided here.

Calibrate the analytical instrument. Suggested calibration ranges for the method analytes and concentrations for the internal standards and surrogates can be found in the EPA method.

Start by selecting five spiking levels that will be used to estimate the LCMRL for each analyte. [Section 5.2](#_LFB_Concentrations_Must) including subsections discuss criteria useful for determining appropriate concentrations. Make sure each LFB concentration you select is bracketed by calibration standards ([Sect. 5.1.2](#_Encompass_LFB_Spiking)). After you complete five of the seven LFB levels, calculate LCMRLs for the method analytes.

Select at least two more LFB levels. See [Section 5.4](#_Estimate_LCMRL_after) for guidelines on appropriate LFB concentrations. As you complete these additional levels, update the LCMRL input file with the additional data and calculate revised LCMRLs. For each analyte, the LCMRL program will display a message in the output file indicating whether a valid LCMRL has been determined or if additional spiking levels are required. Follow the recommendations in [Section 5.2.3](#_When_LCMRL_Estimate_1) (lower level LFB needed) and [Section 5.2.4](#_When_LCMRL_Estimate) (higher level LFB needed) to add additional spiking levels until valid LCMRLs are determined for each analyte.

## LCMRL Calibration Range

### Use a Typical Range

The concentration of the highest calibration standard should be consistent with routine analysis. The instrument calibration range should be typical of expected routine sample analysis. The use of an instrument calibration that has an unreasonably narrow range may yield a lower LCMRL, but the LCMRL will not reliably reflect performance during routine analysis. The calibration range must not exceed the response that begins to saturate the detector.

### Range Must Encompass LFB Spiking Levels

The concentration of the LFBs must lie within the range of the calibration. Spiking levels are invalid if the concentration exceeds the calibration range at the low or high end.

## LFB Concentrations Must Bracket LCMRL

The LCMRL must be bracketed by at least one LFB concentration level that is at, or below, the LCMRL and by at least one LFB level that is at, or above, the LCMRL. If this is not the case, then at least one more level of LFBs must be processed to include the LCMRL within the concentration range of the LFBs. An estimate for a low-level spiking concentration might be where the mean of replicates is close to the accuracy extremes of 50 and 150% or where reproducibility begins to break down. When interferences are present, one might use a low-level estimate of twice the equivalent concentration of the analyte peak as found in the method blank. The accuracy and precision for data lower than the LCMRL may be of poor quality, but it is necessary to find at what concentration data quality begins to break down.

### Range of LFB Concentrations

It is recommended that the range of LFB concentrations be within about a factor of twenty of the LCMRL. Spiking levels that far exceed the LCMRL lose significance when estimating an LCMRL. More than seven levels of LFBs may be needed to observe this recommendation for multi-analyte methods.

### Counting LFB Levels

For each analyte, three of the four analyses in each spiking level must be reported as non-zero results or that level does not count toward the required seven LFB levels. LRBs do not count as one of the required seven spiking levels.

### When LCMRL Estimate is Less than the Lowest LFB Concentration

When the calculated LCMRL is lower than the lowest LFB spiking level, an estimated LCMRL result will appear in the output file with a message: “Lower spiking level needed to bracket the LCMRL.” This means that the LCMRL program needs a lower LFB level to calculate a valid LCMRL. Fortify this additional LFB below the estimated LCMRL while retaining as much signal to noise as possible. Practically speaking, at least one spiking level must fail LCMRL QC limits to determine an LCMRL.

If the lower spiking level is below the current calibration range, a calibration standard must be added at, or below, the concentration of the added LFB. Run a full calibration with the new standard included. The additional low-level calibration standard may fail the method QC limits for calibration, but this calibration level should be allowed so that the LCMRL calculator can determine an LCMRL. Completed LFB levels do not have to be re-analyzed.

### When LCMRL Estimate is Greater than the Highest LFB Concentration

When the calculator determines that an LCMRL estimate is higher than the highest LFB spiking level, an LCMRL result will not appear in the output file. Instead, this message will appear: “LCMRL is above highest spiking level”. This means that when the calculator processed the LFB results, the LCMRL QC probability limits were greater than 50 to 150%, even at the highest level. If an LCMRL is to be determined, a higher LFB level must be processed. This assumes that accuracy and precision will improve at higher spiking levels. For example, if the method cannot extract at least a recovery of 50% at any concentration, an LCMRL defined as “between 50 and 150% recovery” cannot be determined. At least one spiking level needs to pass LCMRL QC probability limits to determine an acceptable LCMRL.

If the higher spiking level exceeds the current calibration range, a calibration standard must be added at, or above, the concentration of the added LFB. Run a full calibration with the new standard included. Completed LFB levels do not have to be re-analyzed.

### Running Out of Response

If the calculated LCMRL is below the lowest LFB concentration and data cannot be obtained below the LCMRL, then the laboratory should set the LCMRL the lowest LFB concentration.

## Include Laboratory Reagent Blanks in Data Set

At least four LRBs must be included in the dataset. If the LRB does not have a response, report “0”.

## Estimate LCMRL after Five Spiking Levels

It is recommended to calculate the LCMRL after the first five spiking levels have been processed so that the next two LFB concentrations can be selected based on the estimated LCMRL. A lower or higher concentration may be needed to ensure the LCMRL is bracketed by LFBs. If the estimated LCMRL is already bracketed, select additional concentrations within the existing range, or expand the range at either end keeping the LFB concentrations as close as possible to the estimated LCMRL.

## Outliers

Extreme data observations, or outliers, can occur. An outlier may represent the actual lack of reproducibility of the method at that concentration. Or, an outlier may result from changed conditions that represent a different population of data. Or, an outlier may due to analyst error. If the reason for an outlier is known, such as a double spike, that data should not be used. Otherwise, the data should be used. Extreme outliers are down-weighted by the LCMRL calculator.

## Formatting Conventions for Data

### Numerical Results

At least three significant figures for each non-zero result should be reported for LFB and LRB samples and entered into the LCMRL input file. Otherwise, the variance may be censored, resulting in an artificially low LCMRL. Because data systems may truncate the number of digits at the low end of the calibration curve, this requirement may need to be addressed by selecting a lower set of units for the LCMRL analyses. Note that numerical results ending in zero are truncated in .csv files. For example, a result of 4.00 ng/L would appear in the .csv input file truncated as “4”. This is normal and does not affect the calculation.

### Report No Response as Zero

Instrument software may not report a result when there is not a peak to integrate. Report such an analysis as “0”. In addition, when a software system reports “below calibration,” enter “0”.

### Negative Values

At low-level concentration, some instrument software systems will report negative numbers. Enter negative values into the input file and calculate LCMRLs using the script for data sets that include negative numbers ([Sect. 4.5](#_Calculating_LCMRLs_for)).

## No Time Requirement

The LCMRL procedure does not have a time requirement for LFBs to be processed and analyzed. The LFBs can be processed all at once or over time.

## Confidence Level and Data Quality Interval

A reporting level is defined in terms of level of confidence and data quality. The 99% confidence level that is used for the prediction interval is considered conservative, but for drinking water surveys the quality of the data is important. The data quality interval that was chosen for the LCMRL, 50 to 150%, was based upon the judgment of experienced analysts at OGWDW.

# References

1. US EPA. *Statistical Protocol for the Determination of the Single-Laboratory Lowest Concentration Minimum Reporting Level (LCMRL) and Validation of Laboratory Performance at or Below the Minimum Reporting Level (MRL)*; EPA 815-R-05-006; Office of Water: Cincinnati, OH, November 2004.
2. US EPA. *Technical Basis for the Lowest Concentration Minimum Reporting Level (LCMRL) Calculator*; EPA 815-R-11-001; Office of Water: Cincinnati, OH, December 2010.

Tables

| Analyte | LFB Spiking Levels | LCMRL | Message Table . LCMRL Results for the Test Data, ng/L |
| --- | --- | --- | --- |
| Analyte 1 | 1.0, 2.0, 4.0, 6.0, 10, 14, 20 | 1.6 | Valid LCMRL |
| Analyte 2 | 2.0, 4.0, 6.0, 10, 14, 20 | 1.3 | Lower spiking level needed to bracket LCMRL |
| Analyte 3 | 1.0, 2.0, 4.0, 6.0, 10, 14, 20 | 3.4 | Valid LCMRL |
| Analyte 4 | 1.0, 2.0, 4.0, 6.0, 10 | 0 | LCMRL is above highest spiking level |
| Analyte 5 | 4.0, 6.0, 10, 14, 20, 41, 82 | 16 | Valid LCMRL |

# Figures

Figure . Proper Path for the Default Directory

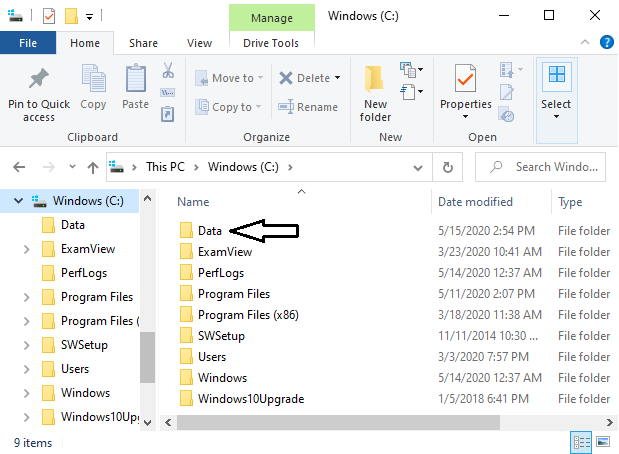


Figure . RStudio Console

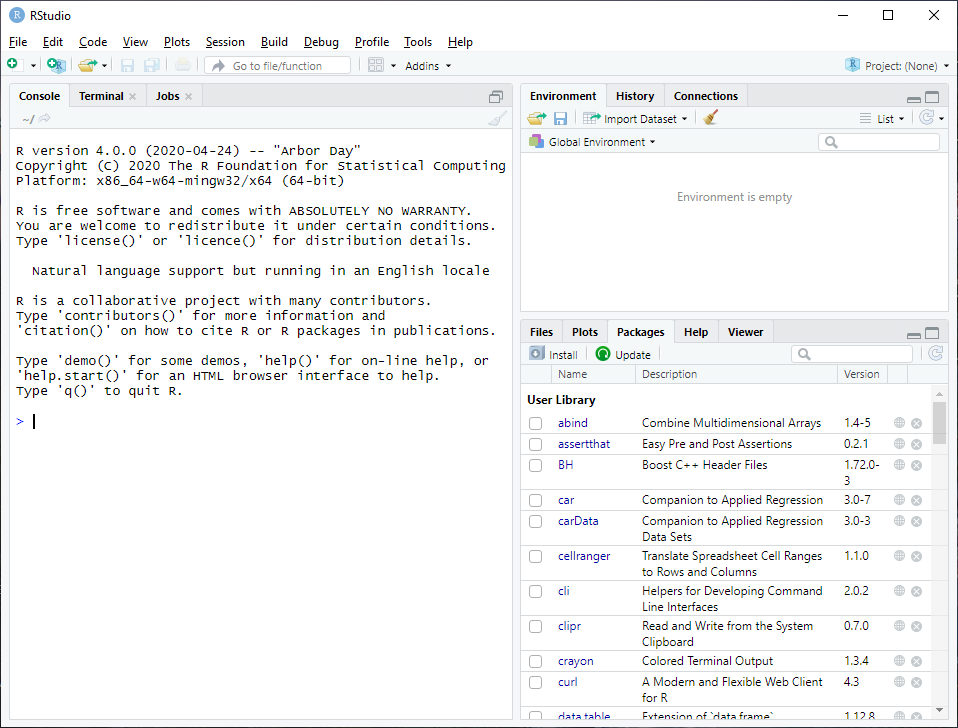


Figure . RStudio Console after Successfully Loading the Source File

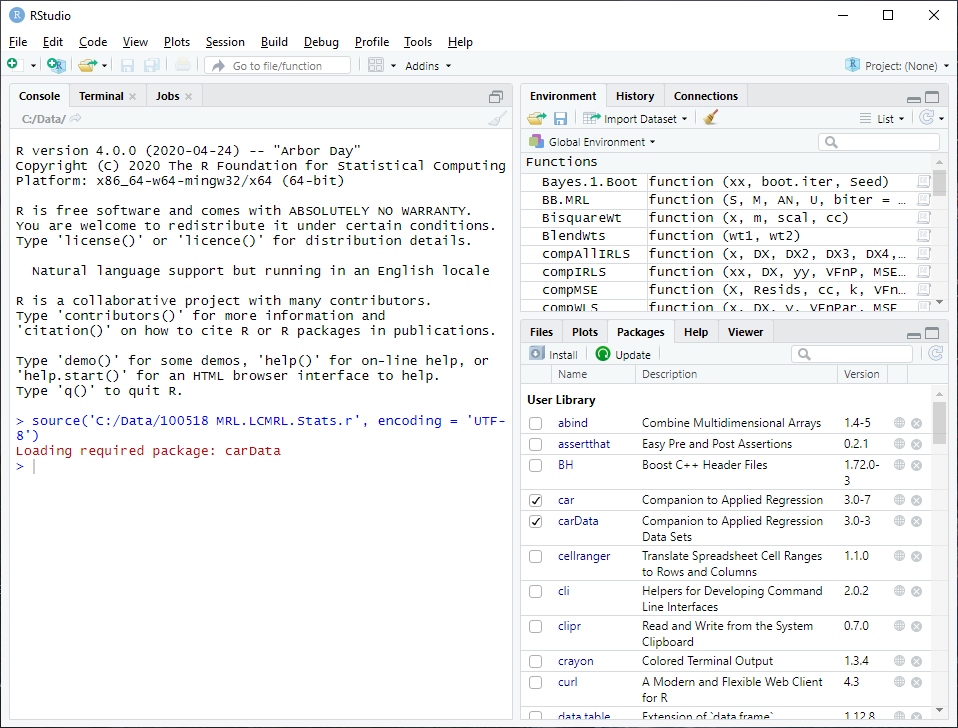


Figure . Recommended Desktop Environment for Calculating LCMRLs

