

Installing the EdSurvey Package on a Restricted-Use Data Computer

Thomas Fink

October 2020

This user guide provides guidance to someone trying to set up the `EdSurvey` package on a computer that must be an isolated stand-alone computer with no Internet, or network connectivity, for working with restricted-use data (RUD). When a computer has no Internet/network connection, it is known as being “air-gapped,” which makes setting up `EdSurvey` more complex.

The `EdSurvey` package focuses on mainly National Center for Education Statistics (NCES) data, and all applicable laws must be strictly followed. If you are planning to use NCES RUD, the [NCES RUD Manual website](#) is a great resource for referencing NCES rules and regulations.

Typically, R packages are downloaded from the CRAN repository using an Internet connection, then installed to the computer. Without an Internet connection, it becomes difficult to install R packages, especially one like `EdSurvey` with additional package requirements. This guide will walk through a simple and straightforward method to help users in this scenario. This method requires the ability to have an additional computer connected to the Internet and separate from the isolated nonnetworked stand-alone computer. This guide uses screenshots and filepaths relating to a Microsoft Windows environment, but it is applicable for other operating systems as well. Because of the differences in operating system filepath syntax, you will need to adapt them to your specific environment. Both computers should use the same computing environment and the same versions of R and `RStudio` to reduce any chance of errors.

Preliminary Setup and Configuration

Initial setup will involve having a second computer to which you are able to connect to the Internet to be able to successfully download and install the dependent/imported packages required for using `EdSurvey`. Going forward, this his guide will refer to the second computer (which is connected to the Internet) as the network computer and the air-gapped isolated computer as the RUD computer. You also will need some nonnetworked way to transfer files from the network computer to the RUD computer. The most common way is to use a USB hard drive, or flash/thumb drive, for transferring files between computers. An older alternative method would involve burning the contents to a DVD/CD.

If you don't have R or `RStudio` installed, download the program install files from the following R Project and `RStudio` sites. If you already have R and `RStudio` installed on the network and/or RUD machine, you can use these to update your installations to the latest versions (recommended). Use the same installer files on both machines to ensure that both computers are running the same software versions. Install the base R files first, followed by the `RStudio` software.

- The R Program installer file from the [R Project site](#)

- The RStudio installer file from the [RStudio site](#)

Using the USB drive, or alternative method, run the installation setup files on both machines following their setup instructions. Once you have the same versions of software running on both the network and RUD machines, you can begin the package setup process.

Package Setup

With both the networked and RUD computers set up in the same R and RStudio environment, the networked computer can then be used to download the EdSurvey R package, along with all the required R packages. You can do this through the package manager within RStudio, as well as the R `install.packages` function.

```
install.packages("EdSurvey")
```

Performing the package installation with an Internet connection will allow it to properly install all the necessary dependent R packages without the hassle of doing so manually. For example, EdSurvey requires the lme4 package to be available, which has its own required packages, so having this completed automatically is recommended.

If you plan to use packages other than EdSurvey on the RUD machine, proceed with installing them at this time using the same method (on the network computer). Before proceeding, ensure that EdSurvey is working correctly by loading the package with the `library` function or through the RStudio menu options (Figure 1). If installed correctly, EdSurvey will print package information (while loading) to the console. If an error appears, the package was not correctly installed, which must be fixed before proceeding.

```
library(EdSurvey)
```

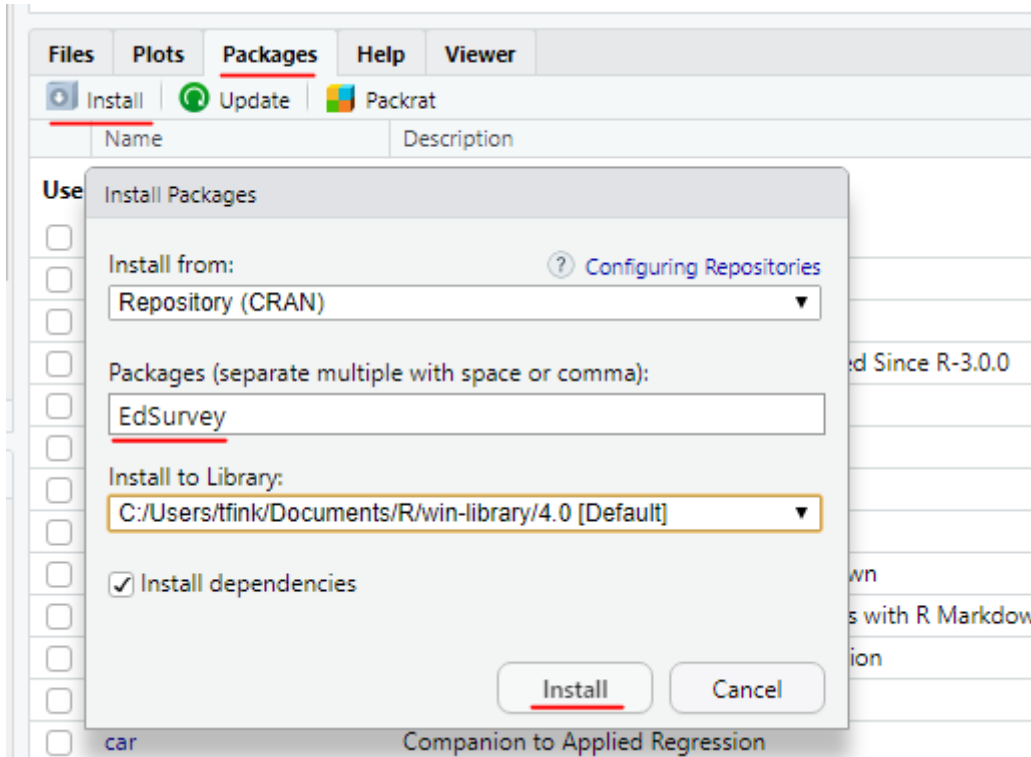


Figure 1: Using RStudio package install menu

Migrating Package Files Between Environments

With EdSurvey and the required packages installed on the networked computer, the next step is to migrate this environment to the RUD computer. This will make the RUD computer *mirror* the networked computer's R package environment. The first step will be to determine where the packages are installed on the networked computer. The easiest way to determine the folder paths is to use the `.libPaths()` R function typed into the networked computer's R console to have the folder path(s) display for you. Typically, two or more folder paths will be displayed. One path generally points to the **base** R packages that are included by default with the R installation files. The other path(s) typically is where all the **downloaded** R package files are stored. If only the **base** folder is listed, you may not have installed any packages yet.

```
.libPaths()

> .libPaths()
[1] "C:/Users/tfink/Documents/R/win-library/4.0" "C:/Program Files/R/R-4.0.0/library"
> |
    ↳ path to "downloaded" R package files      ↳ path to "base" R package files
```

Figure 2: Output results: `.libPaths()` function

Once the **downloaded** R package folder(s) locations are identified, copy the entire folder structure onto a USB drive (or alternate method) that will be used to transfer the files from the networked computer to the RUD computer. Make sure to keep the folder structure as-is (i.e., not modified) and copy all the contents and subfolders for transfer (at the root level) of the folder path (Figure 3). Once the contents are fully

copied to your transfer device, switch over to working on the RUD computer.

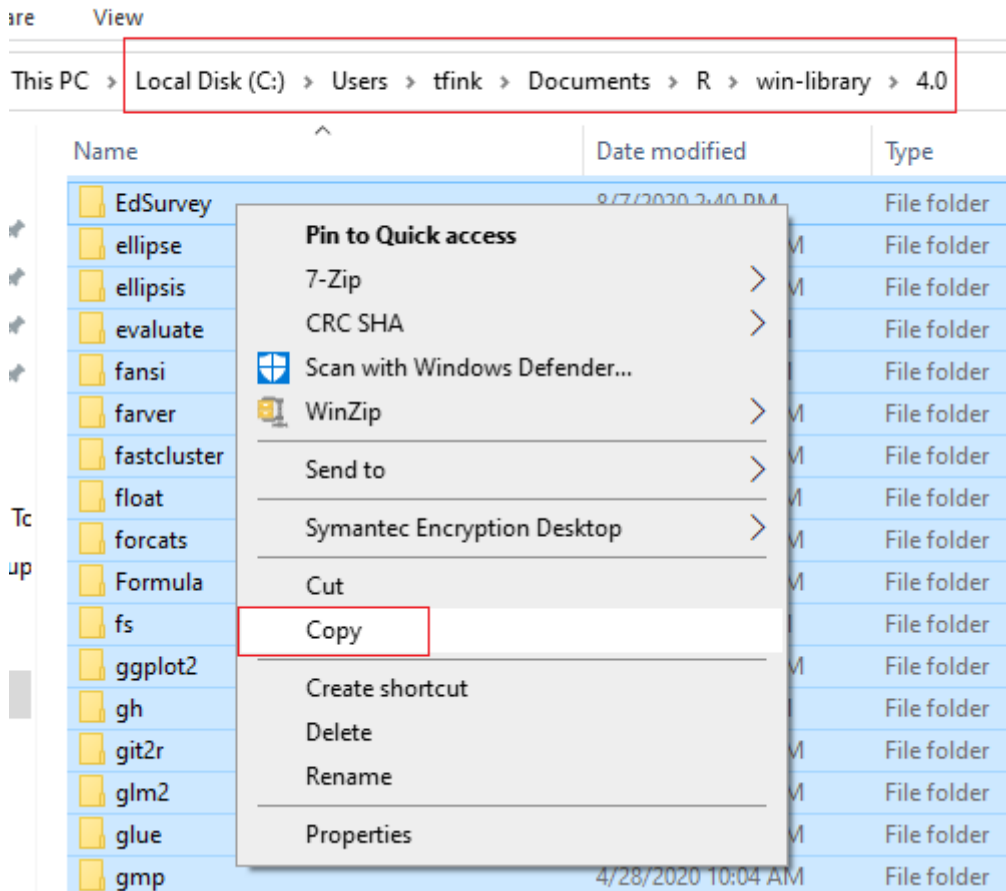


Figure 3: Copying folder contents in Windows Explorer

On the RUD computer, you will perform a very similar process. Begin by identifying the folder location to which you will copy the package files from the transfer device. Within R, use the same `.libPaths()` function call in the R console to have it display the folder paths in use on the RUD computer. Remember the folder path(s) and then exit `RStudio`. Generally, this should be nearly the same, or identical, folder path(s) that you used on the networked computer. Using the transfer device, move/paste all of the folder contents that copied from the networked computer into the package folder path on the *RUD* computer. If you have any existing folders in the package root path on the RUD computer, you may run into instances where folders/files already exist. In those cases, your operating system may ask you if you want to overwrite the files. You **will want to overwrite any existing files** on the RUD computer with the files you copied from the networked computer to ensure that all files are the proper version. Once all files have been successfully copied, relaunch `RStudio` and inspect the installed packages list. The RUD computer now has all of the packages and dependencies that you have on the networked computer.

Wrap-Up and Future Updates

Going forward, you can repeat this process when you need to update packages or to keep the two computers in the exact same package environment. This process will work for any R package and helps the user focus on their code without having to worry about package management. Be sure to always follow your specific RUD license, as the license ensures proper data security.