How to Install and Update the EPIDEMIA Forecasting System

Dawn Nekorchuk, Michael Wimberly, and EPIDEMIA Team Members
Department of Geography and Environmental Sustainability, University of Oklahoma
dawn.nekorchuk@ou.edu; mcwimberly@ou.edu

Updated June 23, 2020

Contents

1 Initial Set-up

1.1 Accounts

1.1.1 Google Earth Engine

If you are only running with the given environmental data, you do not need the Google Earth Engine account.

Environmental data can be obtained from Google Earth Engine (GEE). GEE is a cloud-based platform for hosting satellite imagery. GEE also provides tools to process these remote sensing images and other geospatial data sets. Instead of downloading the raw satellite files and processing them on your own computer, which requires significant internet bandwidth and processing power, these steps are done in the cloud. And at the end, we only need to download the summarized output.

1. Request a GEE account: sign up at https://earthengine.google.com/. If you do not already have a Google Account, it will prompt you to make one. Your Google account will also contain a Google Drive account, which is where the GEE data will be downloaded to.

2. Wait for the account confirmation email in your gmail account.

Optional: You can access your drive online at https://www.google.com/drive/. Or you can download a desktop app https://www.google.com/drive/download/ and use it like a folder on your computer.

1.1.2 Github

Github (https://github.com/) is a web host for Git repositories. Git is a tool for version control. Version control helps keep track of changes while developing computer code. Github can also be used to share code with other users, like we are for the epidemiar-demo project and the epidemiar package.

1.2 Software

1.2.1 R

R (https://www.r-project.org/) is a statistical programming language that runs all of our analyses and produces reports and documentation, including this document. It is free and has a wide variety of packages built by users all around the world to do many different statistical analyses.

It is easiest to download R from CRAN: https://cloud.r-project.org/. Click on the link for "Download R for Windows." Choose "base," then "Download R for Windows." Run this file and install R on your system. Use the default settings for the installer.

1.2.2 RStudio

RStudio (https://www.rstudio.com/) is an integrated development environment (IDE) and graphical user interface (GUI) that makes R easier to use. Download RStudio Desktop Open Source License (the free version). Choose the appropriate installer for your operating system (likely Windows Vista/7/8/10) from https://www.rstudio.com/products/rstudio/download/. Install it using the default settings in the installer.

1.2.3 MiKTeX

MiKTeX (https://miktex.org/) is an implementation of LaTeX, which is what allows us to automatically produce PDF reports of our forecasting results. Navigate to the site, and click on the How To link for your operating system (https://miktex.org/howto/install-miktex).

Click on the download page link (https://miktex.org/download) and then download the basic Installer. Run the installer, but make one change to the default settings. **The option for "Install missing packages on-the-fly" should be changed to "Always".** You will initially install an incomplete copy of MiKTeX, and it will be updated automatically whenever you run the code. This means that the first time you run the code to make a pdf, you will need an internet connection to acquire the missing MiKTeX packages.

After MiKTeX is installed, **restart your computer**. This will help RStudio recognize that MiKTeX is installed.

1.3 R project & packages

1.3.1 R packages

An R package is a collection of functions, data, and documentation that extends the capabilities of base R. Many packages are used in our code for malaria forecasting, and will need to be installed. The easiest way to do this is to have an internet connection the first time you run the `run_epidemiar_demo.R` script in the `epidemiar-demo` project (installed later). In the first section, the script will try to load needed libraries, and if they are missing, it will attempt to download and install them for you. To get a head start on this, in RStudio console, run the command below and that will install a good number of the packages that we use in the `tidyverse` set of packages.

```
install.packages("tidyverse")
```

Instead of typing the command to install packages, you also use RStudio menus. Inside of RStudio, click on the 'Packages' tab in the Files/Plots/Packages/Help pane, and click the 'Install' button. In the pop-up window, type the name of the package (tidyverse).
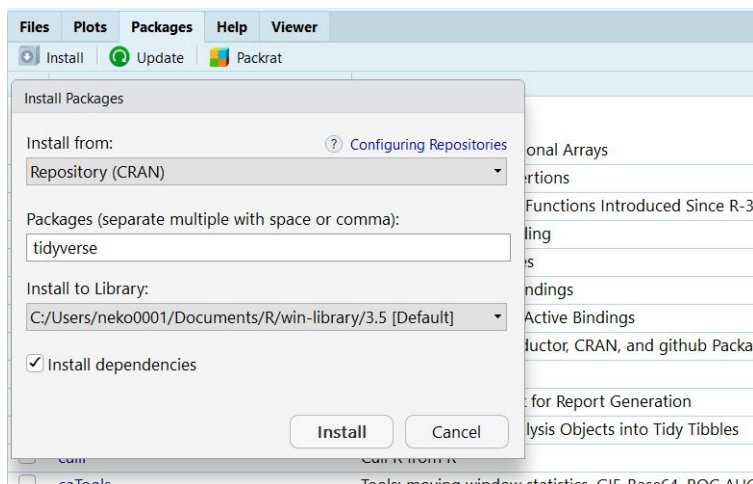
1.3.2 epidemiar package
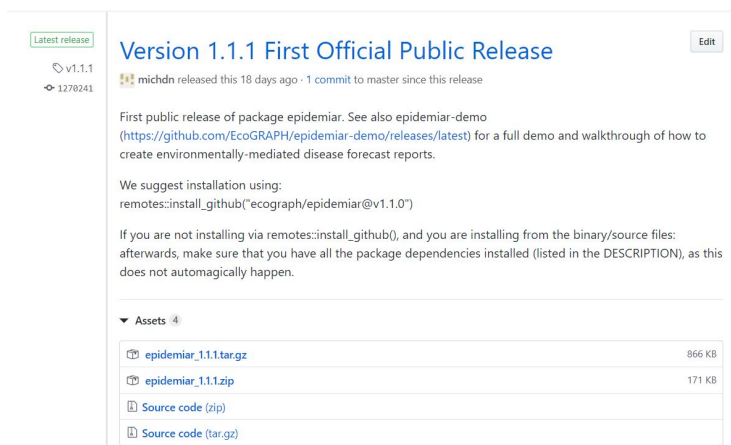
The EPIDEMIA team has developed a R package called `epidemiar` for modeling malaria and creating the forecasting report.

1. We suggest installing using `remotes::install_github()` function as this will automatically install dependent packages (the modified build_opts should allow vignettes to be installed also). Go to https://github.com/EcoGRAPH/epidemiar/releases/latest, and copy the suggest command for installation. The command will look similar to the one below, which would install version "v.3.1.0". (But make sure to get the latest version!) Note: You may need to install `devtools` and `remotes` packages first, if you have not done so already.

```
remotes::install_github("ecograph/epidemiar@v3.1.0", build = TRUE,
                        build_opts = c("--no-resave-data", "--no-manual"))
```

Alternatively, we can download using the binaries provided with the latest release.

1. If you follow this method, first, you will want to make sure that you have all the package dependencies installed (listed in the Description: https://github.com/EcoGRAPH/epidemiar/blob/master/DESCRIPTION).

2. Navigate to https://github.com/EcoGRAPH/epidemiar-demo/releases/latest, and download the latest .zip or .gz.tar file.

3. In RStudio, click on the "Packages" tab in the Files/Plots/Packages/Help pane, and click the "Install" button.

4. Under "Install From", select "Package Archive File (.zip, .tar.gz)". Navigate to and select the zip file, and click "Install".
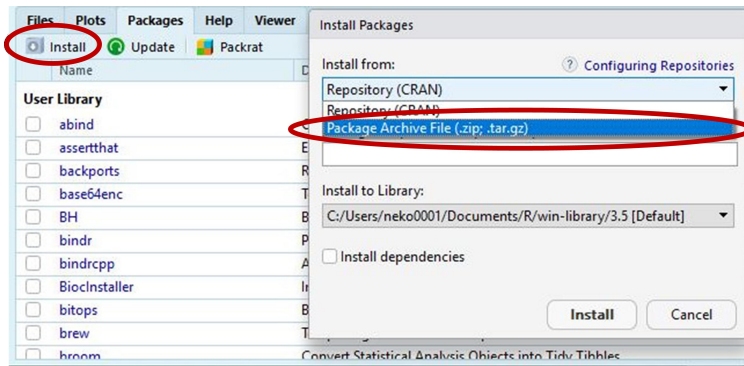
1.3.3 epidemiar-demo project

The demo project can be downloaded from Github directly.

1. Navigate to https://github.com/EcoGRAPH/epidemiar-demo/releases/latest. Click "Source code (zip)" to download the zipped folder of the latest update. (Note: below picture is of an earlier release, but the concept is the same.)
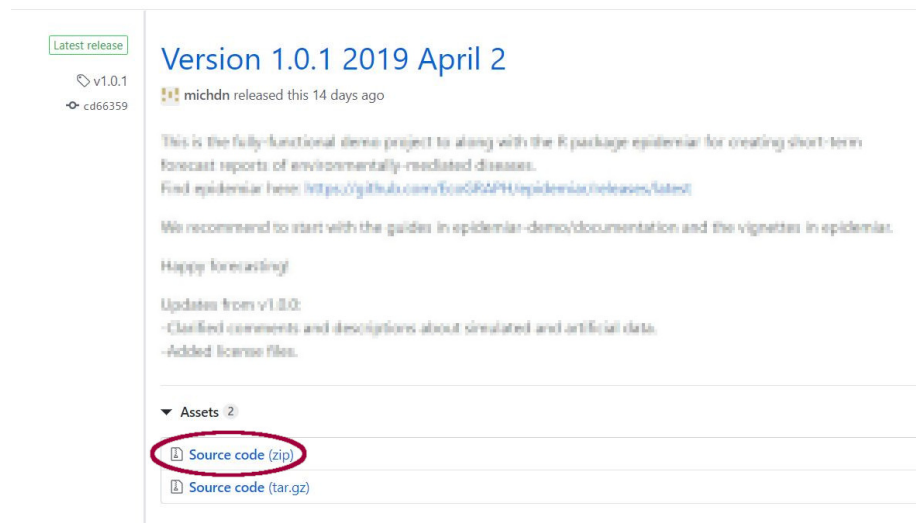
2. Unzip the file to your Desktop or other preferred directory. You must have full read/write permissions to the directory, so do not use a system folder like Program Files. Unzipping the file will create a folder in the directory you choose. The folder name will have the version number in the file name, e.g. `epidemiar-demo-2.0.0`. This will help you keep track of which version of the project you are using.

1.4 Google Earth Engine (GEE) scripts

In the demo project, we show three different ways of interacting with GEE to acquire environmental data.

1.4.1 Option A) GEE Code Editor

1. In the epidemiar-demo-version# folder (created in this step), there is a subfolder named `GEE`. Inside this subfolder will be a text file with the latest Amhara summary GEE script (EPI-DEMIA_GEE_script_v{X.X}.txt.

2. Navigate to https://code.earthengine.google.com/.

3. In the "New Script" center section of the page, copy and paste the text of the script.
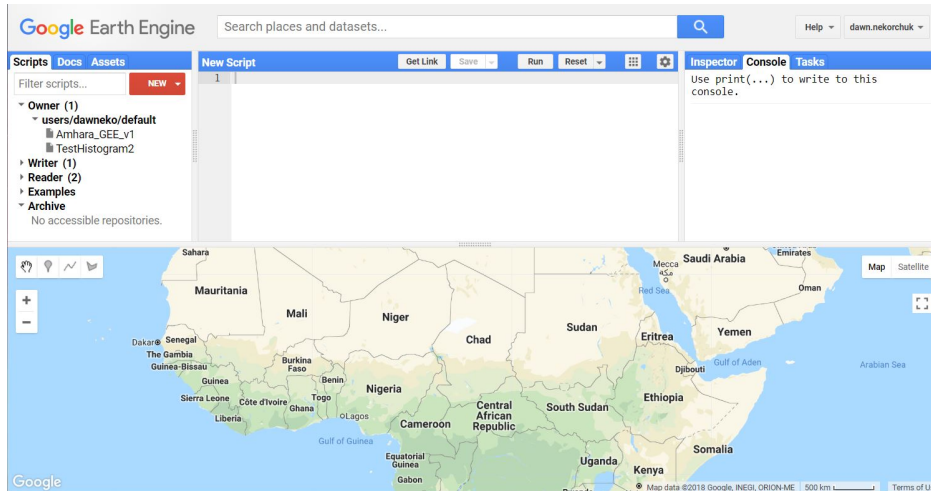
3. Click on save, and name it the same as the text file.

1.4.2   Option B) GEE App

We have turned the GEE code in an app that does not require a GEE account, but it is limited to short date ranges for data download: https://dawneko.users.earthengine.app/view/epidemiar-demo

1.4.3   Option C) Python connection

This option is from a different project, so the data is not compatible with this demo project, but it is here as a demonstration of how you can link R to python to GEE, and therefore can request GEE data from a command in R rather than going to an external application. A GEE account is required, as well as addition python software installed.

1. See the installation directions in https://github.com/EcoGRAPH/epidemia-gee/releases/latest for how to set up your system.

2. Look at the script in this project: /GEE/EPIDEMIA_GEE_python.R.

2 Updating

Once everything is installed, there are only three pieces of software that could need updating when we make a new release: the epidemiar-demo project, the epidemiar package, or the GEE script. Updating these will be very similar to the initial install.

2.1 epidemiar package

Updating the `epidemiar` package is exactly like installing, so please see that section above.

2.2 epidemiar-demo project

Updating the `epidemiar-demo` project is the same as installing, with a few additional steps at the end. After getting the latest version, if you have gathered any new or different data (`/data`, `/data_epidemiological`, or `/data_environmental`), you will need to copy those file(s) over into the appropriate folder(s).

2.3 GEE script

If you are only running with the given environmental data, you do not need to run the Google Earth Engine script.

Updates to the GEE script will be included in the `epidemiar-demo` project in the folder `GEE`. The new version will need to be copied into your Google Earth Engine account.

1. Update the `epidemiar-demo` project with the instructions above. The new GEE script will be in epidemiar-demo/GEE folder, named like `EPIDEMIA_GEE_script_v3.1.txt`. The version number will change with new updates.

2. Navigate to https://code.earthengine.google.com/.

3. On the left hand section, click the red "New" button, and select "File". Name the file the same as the new GEE script text file.
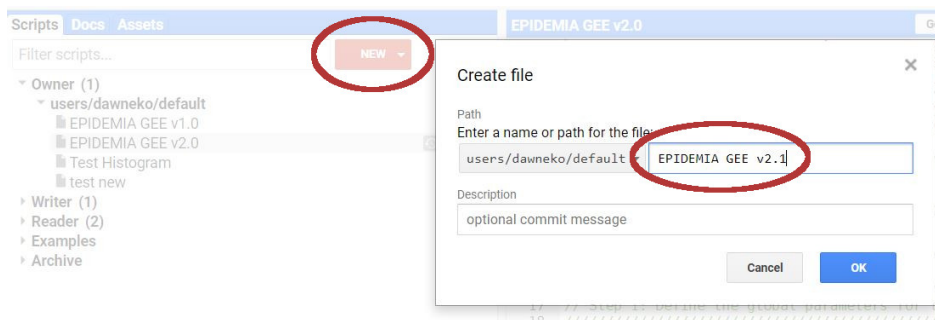
3. In the left section, click on the file name that you just created.

4. In the now blank center section of the page, copy and paste the text of the new GEE script.
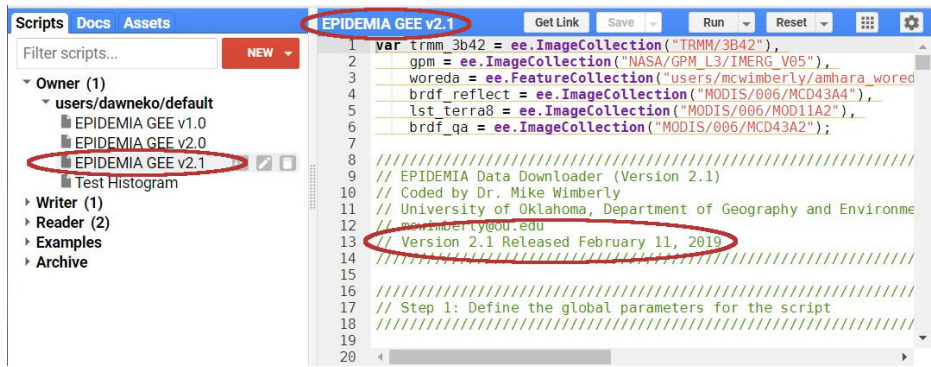
5. Use the new version when you want to collect more environmental data from Google Earth Engine.