

Laptop Prep for “Hands-on: Data Wrangling for Machine Learning”

Overview

Laptop preparation for class consists of five steps, with detailed instructions below:

1. Download the .ZIP from GitHub
2. Installation of the R programming language.
3. Installation of the RStudio IDE.
4. Unzipping the “DataWranglingForMLStudentFiles.zip” file archive to your local hard drive.
5. Installation of R packages.

NOTE – Administrator permission may be required to complete laptop prep. Also, often it is necessary to disable anti-virus software to allow for the installation of R packages. As such, disabling any anti-virus is recommended before laptop prep. Lastly, if you have older R and RStudio already installed, upgrading to the latest versions is highly recommended.

The GitHub repository with all required course files is located here:

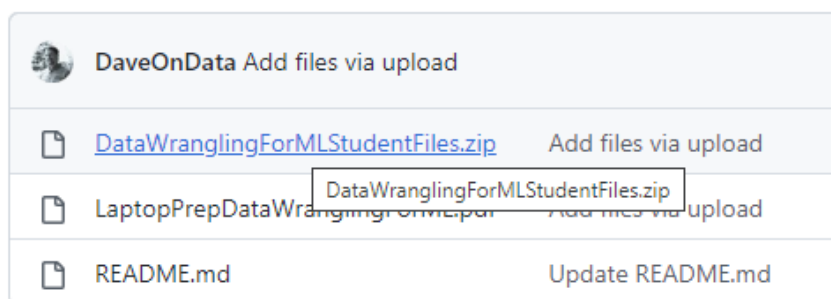
- <https://github.com/DaveOnData/TDWDataWranglingForML>

Hardware Requirements

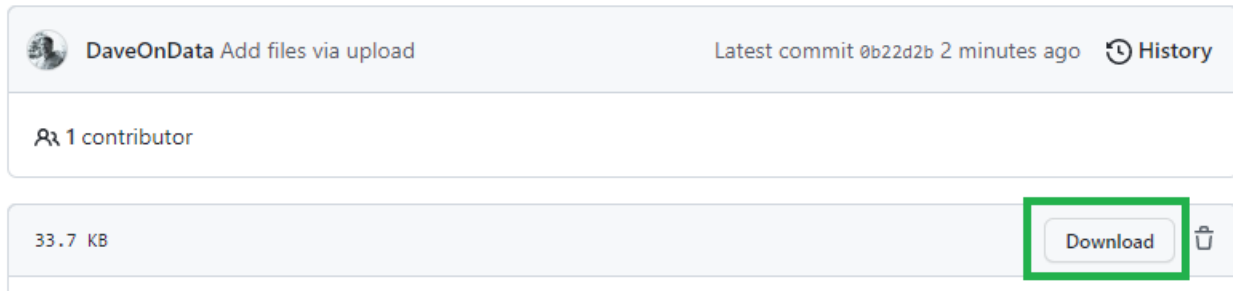
1. Windows or Mac OS X preferred (instructors have no experience with Linux).
2. 4GB of RAM, 8GB of RAM preferred.
3. 500MB of free drive space.
4. WiFi capability.

Step 1 - Download the .ZIP file from GitHub

1. Within the GitHub repository page, click on the link for the “DataWranglingForMLStudentFiles.zip” file:



2. On the next page, click the “Download” button:



3. Move the .ZIP file from your Downloads folder to a well-known location on your laptop.

Step 2 - R Installation

1. Open your browser and navigate to: <https://cran.rstudio.com/>
2. Select the R installer applicable to your laptop:

Download and Install R
Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

3. Download the applicable installer for your laptop (Windows shown below):
Subdirectories:
[base](#) Binaries for base distribution. This is what you want to [install R for the first time](#).
4. Run the R installer, accept all the default installer settings, and install R.

Step 3 - RStudio Installation

1. Open your browser and navigate to:
<https://www.rstudio.com/products/rstudio/download/#download>
2. Scroll down the webpage to select and download the installer applicable to your laptop:

All Installers and Tarballs

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use [an older version of RStudio](#).

Linux users may need to import [Posit's public code-signing key](#) prior to installation, depending on the operating system's security policy.

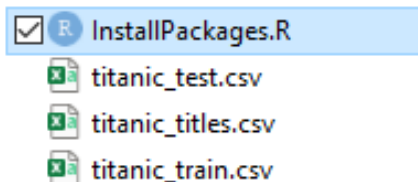
OS	Download	Size	SHA-256
Windows 10/11	RSTUDIO-2022.12.0-353.EXE ↓	202.76MB	FD8EA4B4
macOS 10.15+	RSTUDIO-2022.12.0-353.DMG ↓	365.70MB	FD4BEBB5

3. Run the RStudio installer, accept all the default installer settings, and install RStudio.

Step 4 – Unzipping Class Files

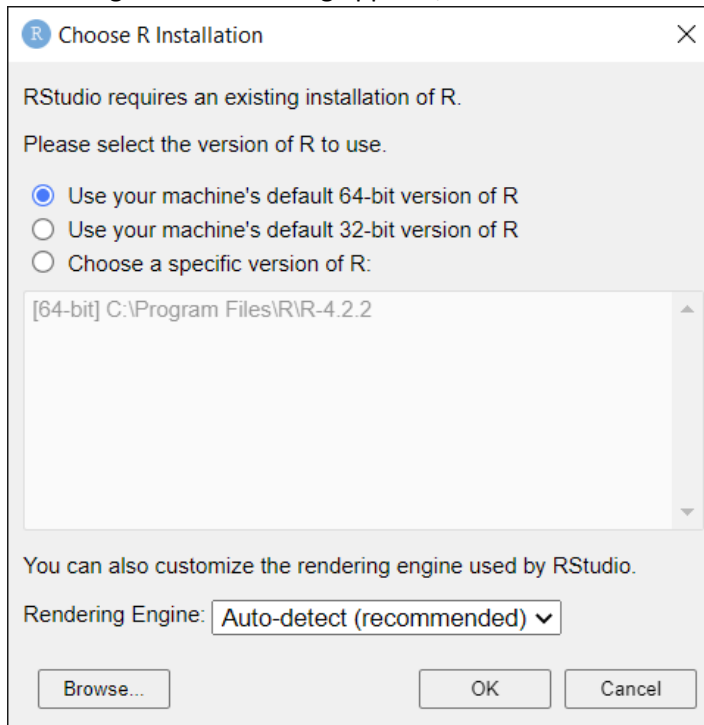
1. Unzip the DataWranglingForMLStudentFiles.zip file to a well-known location on your laptop's hard drive.
2. You should see the following files in the chosen file folder:

☐ Name

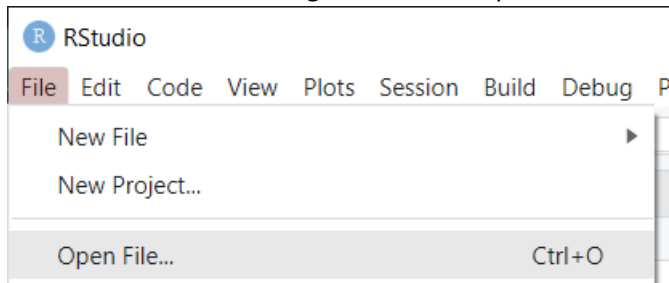


Step 5 - R Packages Installation

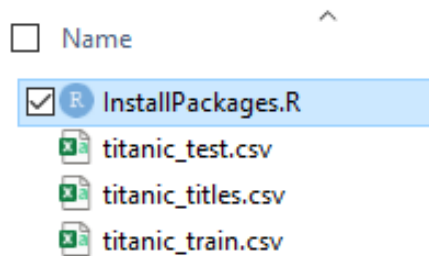
1. Open RStudio.
2. If a dialog like the following appears, click the “OK” button:



3. From within RStudio navigate to File -> Open File:



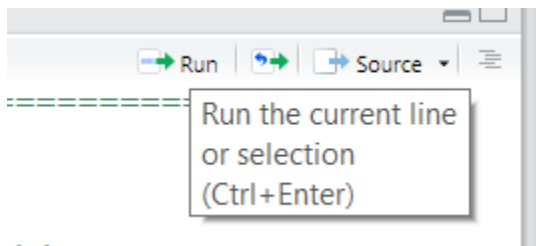
4. Using the file dialog, navigate to the “InstallPackages.R” file you unzipped and open in RStudio:



5. Using your mouse, highlight the code:

```
14 # Install packages required for the course.
15 install.packages(c("skimr", "tidyverse", "nycflights13"),
16                  dependencies = TRUE)
17
18 library(skimr)
19 library(tidyverse)
20 library(nycflights13)
```

6. Within the RStudio IDE, run the R code by clicking the “Run” button:



7. Running the R code will download and install many packages to your laptop. This process can take several minutes, depending on your Internet connection and laptop speed. **NOTE** – To install the required packages, your laptop will need to be able to download files from <https://cran.rstudio.com>.

8. If you see the following warning, it is safe to ignore:

A screenshot of the RStudio console window. The window has tabs for "Console", "Terminal", and "Background Jobs". The "Console" tab is active. The console shows the R prompt ">" followed by the code from step 5. Below the code, a warning message is displayed in red text: "WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:". The console title bar shows "R 4.2.2" and the file path "C:/Users/david/AppData/Local/Temp/Temp2_DataWranglingForMLStudentFiles.zip/".

```
> # Install packages required for the course.
> install.packages(c("skimr", "tidyverse", "nycflights13"),
+                  dependencies = TRUE)
WARNING: Rtools is required to build R packages but is not currently installed.
Please download and install the appropriate version of Rtools before proceeding:
```

9. The following indicates that the required packages are being downloaded:

```
Console Terminal Background Jobs
R 4.2.2 ~
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/glue_1.6.2.zip'
Content type 'application/zip' length 162267 bytes (158 KB)
downloaded 158 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/knitr_1.41.zip'
Content type 'application/zip' length 1493289 bytes (1.4 MB)
downloaded 1.4 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/rmarkdown_2.19.zip'
Content type 'application/zip' length 3720458 bytes (3.5 MB)
downloaded 3.5 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.2/testthat_3.1.6.zip'
Content type 'application/zip' length 2103856 bytes (2.0 MB)
downloaded 2.0 MB
```

10. The following indicates that the installation was successful:

```
Console Terminal Background Jobs
R 4.2.2 C:/Users/david/Dropbox/TDWI/2023_02_February/LasVegas/DataWranglingForMachineLearning/StudentPrep/
> library(skimr)
> library(tidyverse)
— Attaching packages — tidyverse 1.3.2 —
✓ ggplot2 3.4.0      ✓ purrr 1.0.0
✓ tibble 3.1.8       ✓ dplyr 1.0.10
✓ tidyr 1.2.1        ✓ stringr 1.5.0
✓ readr 2.1.3        ✓ forcats 0.5.2
— Conflicts — tidyverse_conflicts() —
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag() masks stats::lag()
> library(nycflights13)
> |
```

Congratulations! You are now ready for the class!