Exercises

Setting up

Julian Flowers

2024-04-23

Table of Contents

## Introduction

These exercises accompany the two talks on data for the Saudi Public Health Authority Health Intelligence team. They are designed to illustrate a reproducible approach to data analysis in public health analysis and utilise a number of public health open datasets.

The analyses are conducted using the R programming language and Tableau public online software.

## Getting started

To set up you need to install the following software:

* R software ([Windows](https://cran.r-project.org/bin/windows/base/)) ([Mac](https://cran.r-project.org/bin/macosx/)). The version should be R 4.3.3 Angel Food Cake
* RStudio ([Windows](https://download1.rstudio.org/electron/windows/RStudio-2023.12.1-402.exe)) ([Mac](https://download1.rstudio.org/electron/macos/RStudio-2023.12.1-402.dmg)). The latest version is 2023.12.1-402 Ocean Storm
* Rtools (if you are using windows OS) ([Rtools](https://cran.r-project.org/bin/windows/Rtools/rtools40-x86_64.exe))

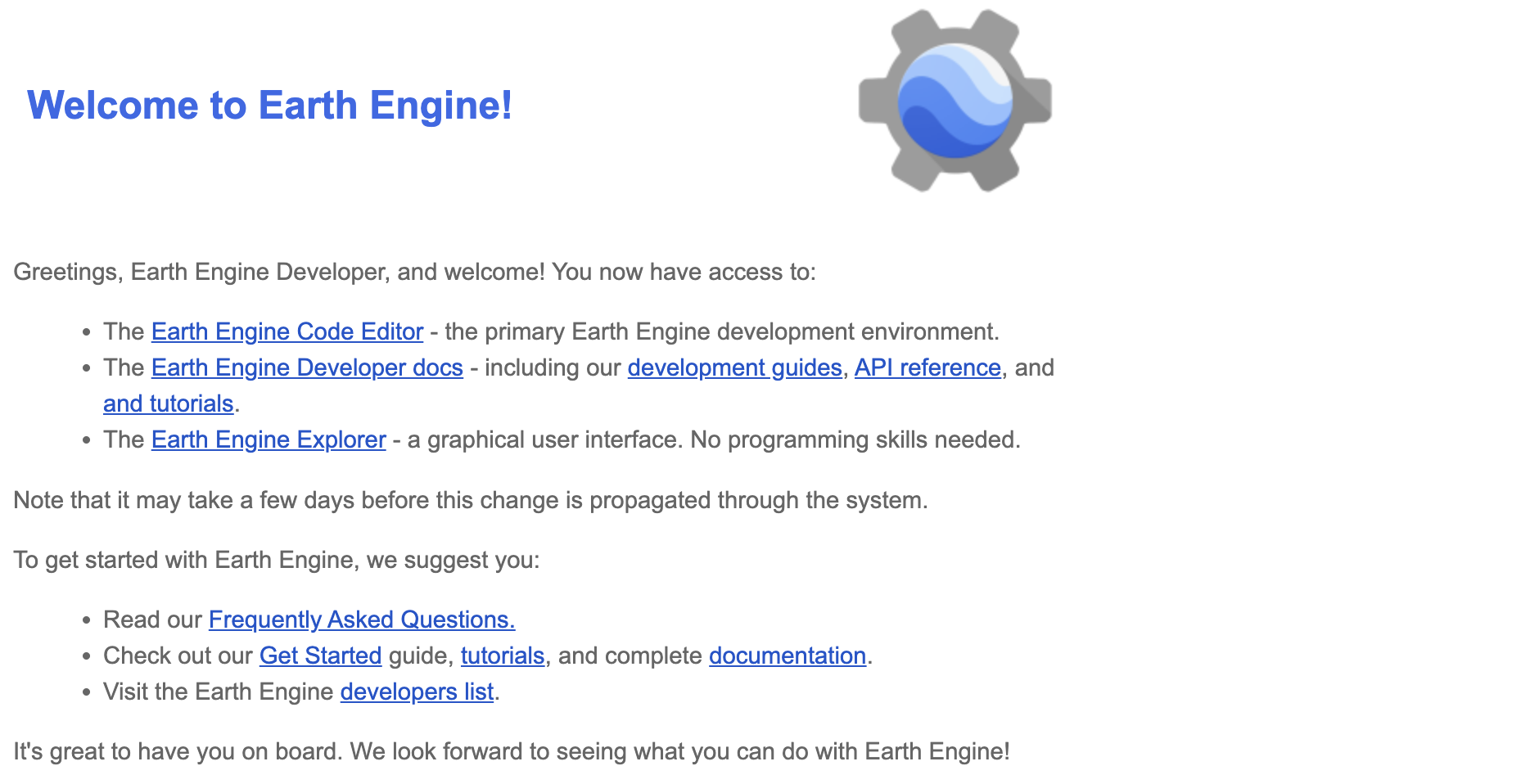
and to create a Tableau Public online account [here](https://public.tableau.com/public/apis/auth/signup).

### Advanced

To follow the advanced exercise on air pollution (using satellite imagery data) you will need to establish a Google Earth Engine (GEE) account as follows:

1. You will need a Google account and sign-up
2. Complete [this form](https://signup.earthengine.google.com/#!/no_redirect) to register a non-commercial use of GEE.

If successful you will receive a notification:



## Using R

Having download and installed R to your computer you will need to install some add on R packages.

To do this run the following code. When asked if you want to load the needs package reply Yes (option 2)

install.packages("pak", repos = "http://cran.us.r-project.org") ## installs packges to install and load other packages

##   
## The downloaded binary packages are in  
## /var/folders/bk/jrqs03tx5mq9s28mhml5xzhm0000gn/T//Rtmpp2bkpK/downloaded\_packages

install.packages("needs", repos = "http://cran.us.r-project.org")

##   
## The downloaded binary packages are in  
## /var/folders/bk/jrqs03tx5mq9s28mhml5xzhm0000gn/T//Rtmpp2bkpK/downloaded\_packages

library("pak")  
library("needs") ## load the needs and pak package  
  
pak(c("tidyverse", "sf", "raster", "stars")) ## this installs packages for manipulating and plotting data, and for spatial analysis  
  
library(tidyverse)  
library(sf)  
library(raster)  
library(stars)

## Getting data for the exercises

Create a directory to download these datasets to (e.g. ~\spha\_data\_exercise)

#### Exercise 1: Replicate John Snow’s 1854 cholera investigation

Download and unzip [this file](https://github.com/julianflowers/spha/blob/main/john-snow-data-master.zip) to you exercise directory.

[File](https://github.com/julianflowers/spha/blob/main/output/john-snow.Rmd)

#### Exercise 2: Diabetes (Public Health Outcome Framework)

[Data](https://github.com/julianflowers/spha/blob/main/dm_model_data.csv?raw=TRUE)

[Codebook](https://github.com/julianflowers/spha/blob/main/dmmod_codebook.csv?raw=TRUE)

[File](https://github.com/julianflowers/spha/blob/main/output/diabetes_model.Rmd)

Exercise 2a: Undertake unsupervised analysis (clustering) of English public health outcome framework diabetes data - explore patterns of diabetes care.

Exercise 2b: Undertake supervised analysis (modelling) of English public health outcome framework diabetes data - does the process of diabetes care improve diabetes outcomes?

#### Exercise 3: Explore and analyse burden of disease and attributable risks in Saudi Arabia

[Data](https://raw.githubusercontent.com/julianflowers/spha/main/IHME-GBD_2019_DATA-a5616352-1.csv?raw=TRUE)

[File](https://github.com/julianflowers/spha/blob/main/output/gbd-saudi.Rmd)

#### Exercise 4: (Optional) Gapminder open data

[Data](https://github.com/julianflowers/spha/blob/main/gapminder.zip)

File

#### Exercise 5: Advanced. Explore air quality and pollution in Saudi Arabia using satellite data