README

Here, we used the example dataset to illustrate how to compute global ePRS and the variance of PRS (computed by using global ancestry proportion).

The primary R function for computing global ePRS and variance of PRS is in “compute\_global\_ePRS.R”

The example code is in “example.R”

There are three example datasets used for illustration: gaProp\_example.csv, sumstat\_example.csv, and ancestry\_specific\_AF\_by\_GAFA\_example.csv

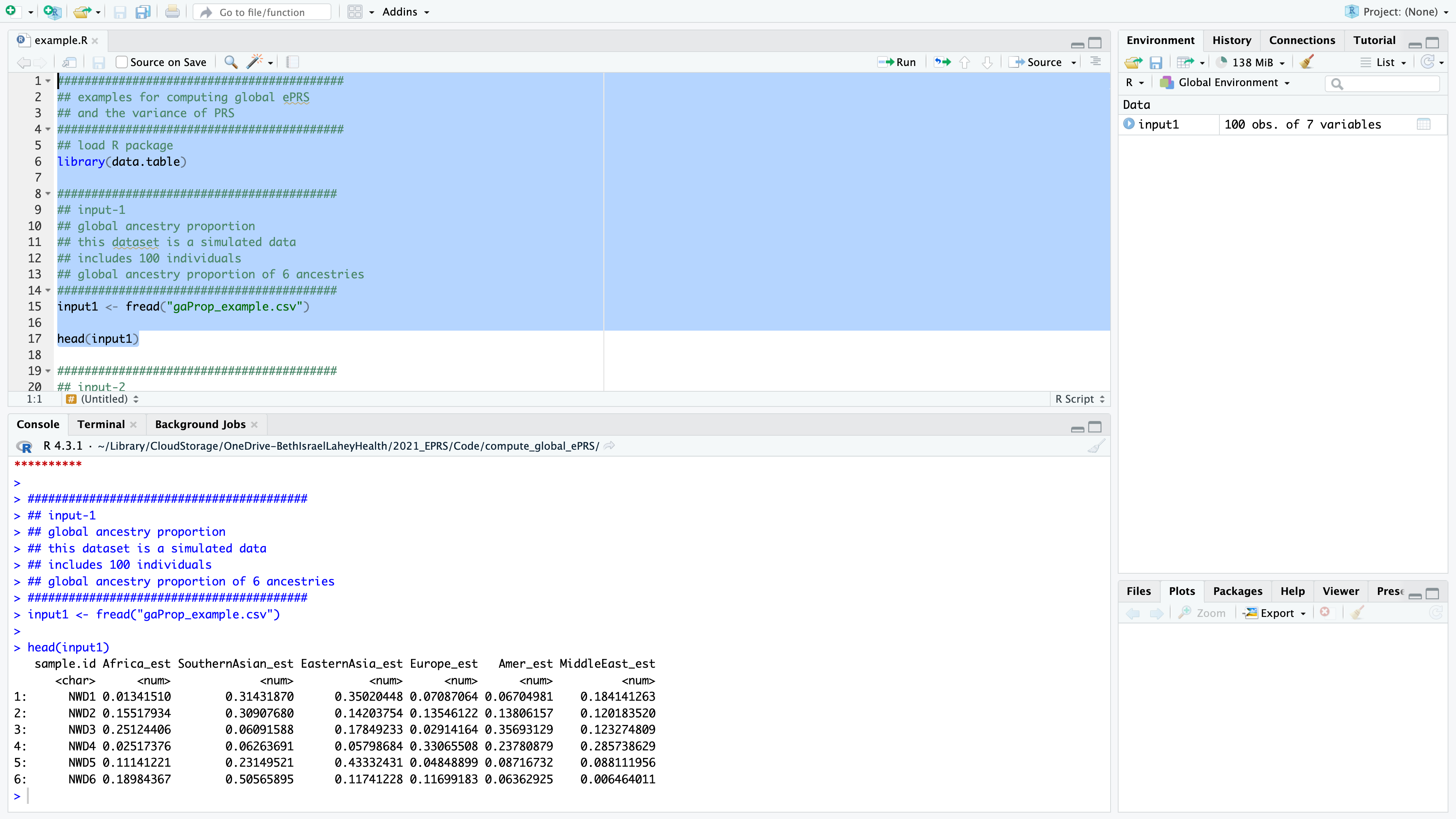
(note: these three files are the three inputs of the “compute\_global\_ePRS.R”)

Before using the “compute\_global\_ePRS.R” function, the user needs to prepare three input files, and make sure the column names of the input files should be the same as the example datasets

First input file: global ancestry proportion file (example dataset: gaProp\_example.csv)

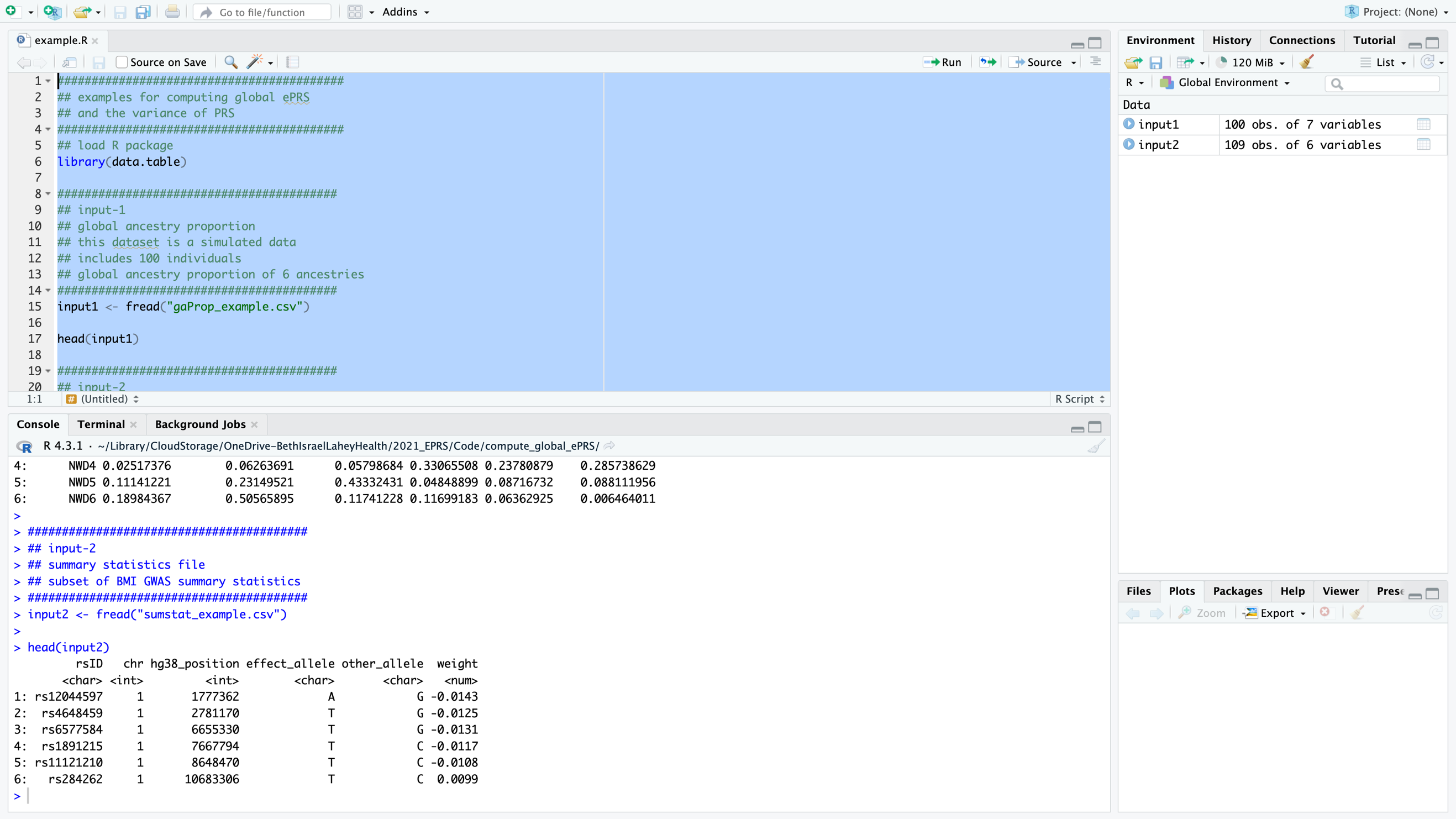
The following is how this file looks: the row represents each individual, and the first column represents the sample ID information.

From the 2nd to the 7th columns represent the global ancestry proportion of each individual from 6 ancestries.



Second input file: summary statistics file (example dataset: sumstat\_example.csv)

The following is how this file looks: the row represents each variant, and the column represents different information for the variant. The position information of each variant is built on hg38 reference genome. The last column “weight” means the effect size of effect allele.



Third input file: ancestry-specific allele frequency file (example dataset: ancestry\_specific\_AF\_by\_GAFA\_example.csv)

The following is how this file looks: the row represents each variant, and the column represents different information about each variant. The ancestry-specific allele frequencies computed using TOPMed individuals are listed from the 6th to 11th columns. These values represent the allele frequency of the “Alternative” allele (ALT, 5th column of this example dataset).

