# Folders

You need to have the folders:

* datasets
* input\_data
* metadata
* program\_costs
* Rbooks
* scripts
* working\_outputs

## Rbooks

This folder has several R notebooks. These need to be run in the order of the number prefix. As they create datasets which the subsequent notebooks use.

### The SS file

The SS which stands for screen scrapping, contains some python and takes a long time to run. Thus, the output has been included in the dataset folder: fees\_durations\_SAC.csv

Should not need to be run often.

## Input Data

This folder contains the data which has been output from the Datalab or obtained from Stats NZ via custom request. I can see only one output from custom extract. What are other two files, please elaborate , did you get it out from IDI ?

The t3 files the occupation by field of study data which we could not get from the custom request.

## Scripts

Contains R scripts which contain functions which can be shared between R notebooks.

## Program costs

Contains a file to do with SAC funding mapped to NZSCED codes, only required by the screen scrapping notebook. (where this comes from? Is it something given? Please elaborate)

See: [GO Obtaining Annual Fees, Durations and SAC funding](https://universitiesnewzealand.sharepoint.com/Shared/Shared%20Documents/Shared/Graduate%20Outcomes%20tool/Census%20based/2018_process/GO%20Obtaining%20Annual%20Fees,%20Durations%20and%20SAC%20funding.docx?web=1) > Obtaining SAC Funding

“A table constructed during the 2013 Census analysis is used to map NZSCED narrow fields of study and NZQF levels with the SAC funding codes.”

## Metadata

These are mostly lookup files downloaded from Stats NZ.

## Working outputs

Some csv or excel files are or were written to this file to have a look at the output.

## Datasets

A folder that seemed like a good idea at the time. Originally it was going to be used like we use the SAS datasets folders, but then I changed my mind and put most of that into the SQL database.

# Database

This process needs an SQL database to be setup, the connection is created in the connect\_to\_GO\_analysis\_database function which is defined in the unz\_common script.