2. Importing data into iNZight

Key iNZight skills addressed:

- Loading data into iNZight from various sources

Side note: File Formats and CSV files

Each program has its own way of formatting and saving data in a file. A common data format you may have experienced would be in Microsoft Excel (its file names have extensions .xls or .xlsx indicating the data was saved in that particular form); another is Google Sheets (which has extension .gsheet).

The data format of one file may not always be translated properly if you open them in another program (e.g., if you attempt to open an .xlsx file in Google Sheet).

To address this, people often save data in a more universal comma-separated values (CSV) file format. Files using this format have the extension .csv.

This file format converts the data into plain text (rather than saving with the unique features of the program), where each data point (number, response, etc) is separated by a comma.
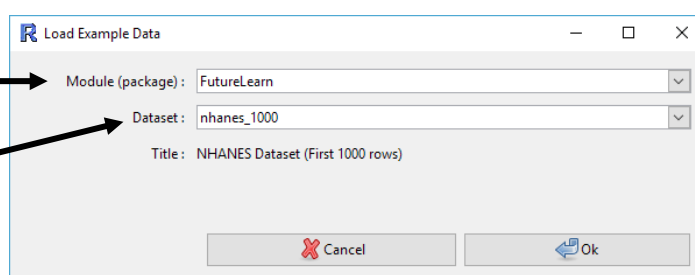
This format ensures that no matter what program you use, you will be able to read in the data correctly.

Such a file can be opened and viewed in a simple text editor, e.g., Notepad, or any other program. Some form of 'import' feature within the 'open file' process of the program will usually facilitate this (enabling you to specify that the file is a .csv file).

INSTRUCTIONS

To load an example data file stored inside iNZight:

- Open iNZight and go to **File > Example data …**

- Under **Module (package)** select **FutureLearn**.
- Under **Dataset** you will see many available datasets.
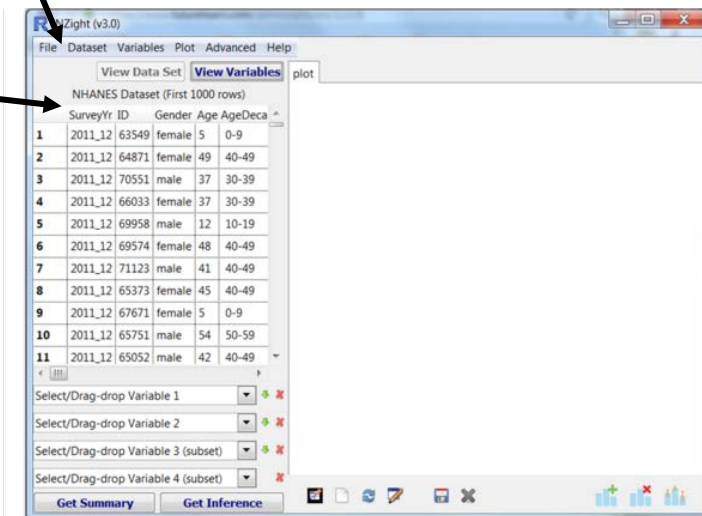- Select the data file **nhanes_1000** and click **OK**

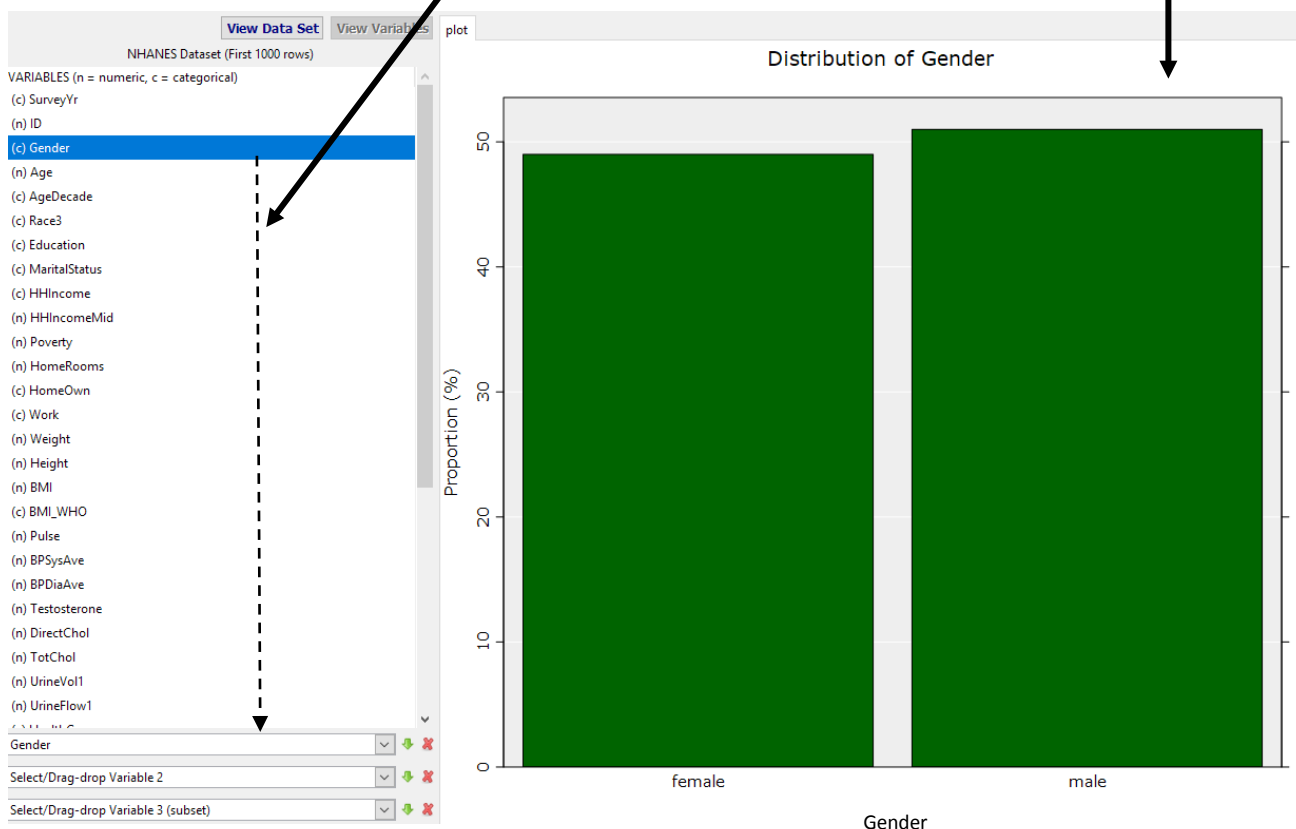The data will appear in the left hand side of your window.

The top row shows the names of each variable.

Each row represents a record (e.g., individual) and their responses (data) for each variable appear in columns – this forms the dataset.

You can view the dataset or just variable names by clicking on the top buttons, **View Data Set** and **View Variables**.

You now have data, so let's explore. Drag a variable name (e.g., Gender or Age) into the **Select/Drag-drop Variable 1** rectangle at the bottom left and see what happens. This graph appears.
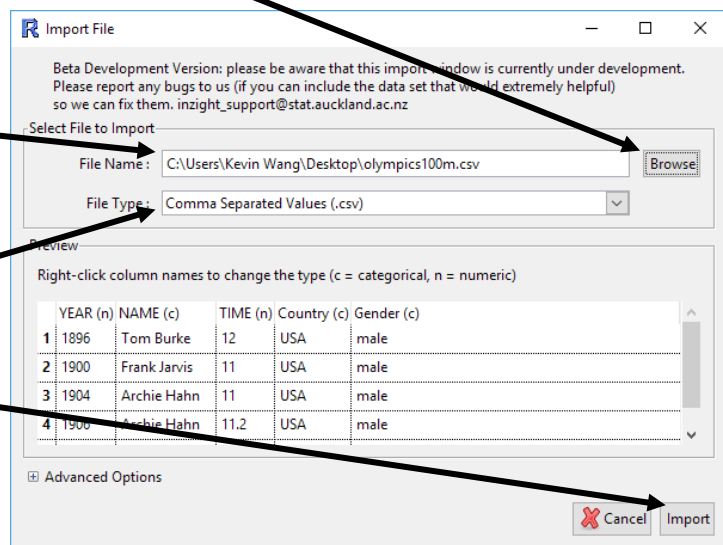

Practice
Load the **gapminder_2008** data set into iNZight. Change the size of your windows to see more or less of the data. Click **View Variables** to become familiar with the variables.
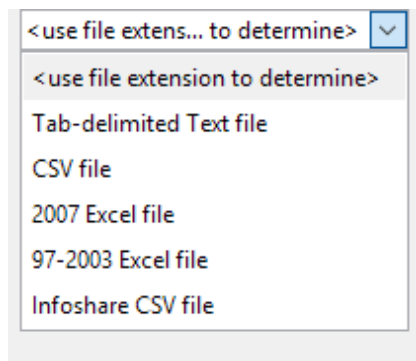
Import a data file from outside iNZight

Download the file _olympics100m.csv_ from
https://www.stat.auckland.ac.nz/~wild/data/FutureLearn/olympics100m.csv and place it on your computer's **Desktop** (you may choose any location, but we refer to the file being on your Desktop in the instructions that follow).

Then select:

- **File > Import data (Beta)…** and click **Browse**

- Select **Desktop** and then select **olympics100m.csv**. The csv file should then appear here.

- The File Type is chosen automatically using the filename's extension, e.g., .csv, .xlxs, .sav, .dta

- Then click **Import**

NOTE: Other files may be imported. E.g.,

Troubleshooting

- ***My data won't import***
  - ➢ Try opening your data file in Excel (or similar) and save it in **Text (tab-delimited)** or CSV format. When you go to import the data you should see a **.txt** or **.csv** extension on the end of its name.

- ***My data won't load***
  - ➢ Check your data set for spaces. If you are using csv files, ensure that your:
    - ▪ column names do not have commas in them;
    - ▪ numerical variables do not have "O" s instead of "0"s.

  - ➢ Try loading just the necessary rows and columns of your data by cutting and pasting them into another spreadsheet and saving as a new file.

  - ➢ Use the *"Input Data (Beta)"* option in the file menu to open your file.

- ***Still having problems?***
  - ➢ Read iNZight's <u>FAQ page on importing data</u>. See also the article called <u>Using your own data</u>.