Getting Research Data Into Jigsaw

Derek Norcott

September 12, 2015

Abstract

The Gauquelin data is presented on the C.U.R.A. website (`http://cura.free.fr/`), but the form in which it is presented may seem inaccessible to some. It has come to my attention that there are some astrologers and astrology researchers who would like to import this type of data into a statistical analysis program called *Jigsaw*[1], which allows for researching astrological hypotheses. I shall take the reader step-by-step through the process of converting such data, making it ready for import into *Jigsaw*.

1 Preparation

The first thing to do is have your tools ready to hand before getting started, so that the flow of action is made easier. Below, I set out what is required before you set forth; and please don't skimp on backing up your data, as there's bound to be false steps and errors along the way.

Although the data manipulation and conversion can be done on any platform, the final import file needs to be ready for use on MS Windows, as *Jigsaw* is currently only available on that platform. So for those of you who wish to use the Mac OS X or Linux platforms to do the all the editing and data manipulation, remember when doing the final edit of the import (.dat) file, that it must use Windows-style line-endings throughout, or Jigsaw will generate an error when you go to import it.

1.1 File Extension Visibility

The first thing to do is ensure that the file-type extension is visible on your files, so that you can tell which files have the '.dat' extension, which have the '.txt' extensions, *etc.*

[1] Jigsaw by B.Brady & Esoteric Technologies.

By default on MS Windows, the file extension is hidden. To correct this, simply go to *Start / Control Panel / Appearances and Personalization*, then *Folder Options*. Select the *View* tab then under *Advanced Settings* clear the *Hide extensions for known file types* check box, and then click *OK*.

Other operating systems like Apple Mac's OS X and Linux show the file-type by default.

1.2 Text Editor

A good text editor is always useful to have at hand. This will be needed to create a plain text file containing the data that is to be cut & pasted. In MS Windows, I'd recommend using Notepad++ (`https://notepad-plus-plus.org/`); you *can* use the default *Notepad* (found under *Start / Accessories*), but be aware that it may truncate large files.

In Linux, there are plenty to choose from—I'd personally recommend *Geany*. Mac OS X also has many text editors to choose from.

1.3 Spreadsheet

The spreadsheet is where the messy task of data conversion takes place. For MS Windows, there's *MS Excel* from the *MS Office suite*.

However, there are free cross-platform spreadsheets that are equally capable of performing the tasks. I'd personally recommend *Calc* from the *LibreOffice Suite*, or from the *OpenOffice Suite*.

1.4 Jigsaw

Jigsaw is the research program that you'll be using to import the data. It doesn't matter which version you use. The import file-type extension for its files is '.dat': a simple text file that uses this extension instead of '.txt'.

1.5 Work Folder

Decide where you are going to do your work. I suggest creating a folder on your desktop or perhaps under the *Documents* folder. Maybe call it `\JigsawImport\` as an example.

1.6 Backup

I highly recommend having a memory stick set aside to back up your work as you proceed, as there's nothing worse than making changes to large data-sets and then having to start over again after you've lost previous files or crucial changes.

You'll need to make a backup of the saved text file from which you've created a cut & paste of the data; a backup of the spreadsheet file that you've created after importing the text file into it; then a backup of other spreadsheet files after you've made major modifications to the columns of text; and finally, a backup of the '.dat' file you've exported from the spreadsheet after final manipulation.

2 Understanding Jigsaw's Import Format

In order for Jigsaw to be able to import data, the import file has to follow its strict import structure. To see what this is, venture into the *Jigsaw* folder. This may be `C:\Jigsaw\userdata\` or `C:\Program Files\Jigsaw\userdata\`. Then find any '.dat' file—perhaps 'Actors.dat'—and make a copy of it, saving it to your work folder. Note: The '.nat' file is for purposes of rectification, which we don't require, so can be safely ignored.

Load up your text editor of choice, and open the (.dat) file.

Here are the first five rows of the *Actors.dat* file as an example:

```
"Ada Alan",1936,1,28,5,7,0,0,3,"EST",5,"New York NY",73.95,40.75,-1,0,""
"Albert Eddie",1908,4,22,11,30,0,0,3,"CST",6,"Rock Island, ILL",91,41.5,-1,0,""
"Aldredge Tom",1928,2,28,12,15,0,0,3,"EST",5,"Dayton, OH",84.2,39.75,-1,0,""
"Ameche Don",1908,5,31,19,0,0,0,3,"CST",6,"Kenosha, WI",87.81667,42.58333,-1,0,""
"Andrews Julie",1935,10,1,6,0,0,0,3,"BDT",-1,"Walton on Thames, UK",.1,51.36666,-1,0,""
```

This tells you that:

- it's a comma-separated file;
- the data starts directly on the first line, *i.e.* there's no heading line;
- there are 17 columns for each row.

Following inspection, it can be deduced that the columns, required to be the following strict order displayed in table 1; columns 8, 9, 16, 17 don't appear to change, so we shall not worry about what they actually represent for now.

Table 1: DAT file: **Actors.dat**

Column 01	full name of person (in double-quotes)
Column 02	year (yyyy)
Column 03	month (m/mm)
Column 04	day (d/dd)
Column 05	hours (h/hh)
Column 06	minutes (m/mm)
Column 07	seconds (s/ss) (assumed)
Column 08	[0] (leave as is for now)
Column 09	[3] (leave as is for now)
Column 10	timezone (in double-quotes)
Column 11	hours (*minus*=subtracted)
Column 12	full place name (in double-quotes)
Column 13	Longitude (as a decimal; *minus*=East)
Column 14	Latitude (as a decimal; *minus*=South)
Column 15	Daylight Saving (*minus* 1hr) / War (*minus* 2hrs)
Column 16	[0] (leave as is for now)
Column 17	[“”] (empty double-quotes) (leave as is for now)

3 Extracting the Data

We now need to extract the data from the web page. We'll be using *2088 Sports Champions* (`http://cura.free.fr/gauq/902gdA1.html`) as our example data source.

Open up your text editor and create a new blank page.

Then scroll down the web page ignoring the *Professions* and *Sports Index*—we will return to the *Notice* section later.

Ignore the 'heading row' as we just want the pure data. Highlight all of the data from under the heading to the end row.

Then ***copy*** (either select **Ctrl+C** from keyboard or select **Edit | Copy** from menu using the mouse).

Then place the cursor on the new blank page in your text editor (by left-clicking the mouse), then ***paste*** (either select **Ctrl+V** from keyboard or select **Edit | Paste** from menu using the mouse).)

There should be around 2087 rows of data (not sure where the 2088th row went!). Save the file as `902gdA1.txt` or whatever you want to call it.

4 Importing into the Spreadsheet

We now need to manipulate the data, move the columns around, *etc.* This is facilitated by using a spreadsheet.

Load up the spreadsheet program of your choice. Then select **Open** or **Import** (depending on your flavour of spreadsheet). You'll be confronted with a box giving you 'Separator options'. Just select **Separate by** and **comma**, un-checking any other options like space, tab, semicolon; also uncheck any requests for including quoted text. The standard/general default format should be used. Select **OK** button.

Note: in *LibreOffice Calc*, I also had to specify **tab** to get it to work. You may have to experiment a bit to get your spreadsheet to import it correctly. For instance, using *LibreOffice Calc*, the *standard* format interpreted the 'E' for East in the longitude column as being numerical notation, so I had to change the format of the longitude column to *Text* before hitting **OK** to get it to import correctly.

Once you have imported it correctly, save it as a spreadsheet (in MS Excel it should save as `902gdA1.xls`; in LibreOffice Calc as `902gdA1.ods`). Make another copy saving it as `modify.xls/modify.ods`.

Close the spreadsheet program and have a cup of well-deserved tea. This a good time to remind you to make backups.

5 Format Synchronisation

Load up the spreadsheet program and open up your copy spreadsheet file you called `modify.xls/modify.ods`. Go back to the web page and you'll find the section titled *Notice*; this tells you what each column in your `Modify` spreadsheet file represents. You need to get the columns in the same sequence of 17 columns, and the same format, as the import file.

According to the *Notice* section the columns are arranged as described in table 2

5.1 Build the Import Format

Click on a new tab at the bottom of the spreadsheet to create a new sheet (**Sheet2**); this will be for building your new sheet that corresponds with table 1.

Click the previous tab again for sheet **902gdA1** showing the columns in table 2. The first column full of 'C's are of no use to us, so ignore it.

Table 2: Spreadsheet: **902gdA1**

Column 01 (A)	Profession (to be deleted)
Column 02 (B)	Index Number (effectively, the full name)
Column 03 (C)	Country code
Column 04 (D)	day (d/dd)
Column 05 (E)	month (m/mm)
Column 06 (F)	year (yyyy)
Column 07 (G)	hours (h/hh)
Column 08 (H)	minutes (m/mm)
Column 09 (I)	seconds (s/ss)
Column 10 (J)	timezone (hours)
Column 12 (K)	Latitude (nnNSnn)
Column 13 (L)	Longitude (nnWEnn)
Column 14 (M)	District code
Column 15 (N)	City

In the following paragraphs we will build the new sheet by copying columns—highlighting a column and selecting **Ctrl-C**—from the old sheet **902gdA1**, then pasting **Ctrl-V** into the new sheet (**Sheet2**). Each paragraph will progress sequentially, building new columns in sheet (**Sheet2**), from the old columns in sheet **902gdA1**.

New column 01 (A) for quoted full name: Copied from old column 2 (B) full of sequential index numbers. If you want to replace these with their respective names (and you have the patience!), these are supplied on another web page; however, for purposes of this tutorial, we shan't bother.

Important! Highlight the column, right-click to bring up a context menu and select **Format Cells**, change the default *General* to *Text* and select **OK**—this will ensure that when we save the file in CSV format later, it will place double-quotes around the cell contents, as this new column is expecting quoted text in the form of a full name.

New column 02 (B) for year: Copy from old column 06 (F).

New column 03 (C) for month: Copy from old column 05 (E).

New column 04 (D) for day: Copy from old column 04 (D).

New column 05 (E) for hours: Copy from old column 07 (G).

New column 06 (F) for minutes: Copy from old column 08 (H).

New column 07 (G) for seconds: Copy from old column 09 (I).

New column 08 (H) for 0: Create a new column, in the first cell of that column add *0*, then copy this cell to populate the rest of the cells in that column by highlighting that cell and selecting **Ctrl+C**, highlighting all the cells down to the last (you'll see where by looking at the bottom of the adjacent cell)—you'll notice the spreadsheet populates the range with incremented numbers—select **Ctrl+V** which will replace the entered sequential numbers with *0*s.

New column 09 (I) for 3: Create a new column, in the first cell of that column add *3*, then copy this cell to populate the rest of the cells in that column by highlighting that cell and selecting **Ctrl+C**, highlighting all the cells down to the last (you'll see where by looking at the bottom of the adjacent cell)—you'll notice the spreadsheet populates the range with incremented numbers—select **Ctrl+V** which will replace the entered sequential numbers with *3*s.

New column 10 (J) for quoted timezone: Create a new column, in the first cell of that column add *CET*, representing 'Central European Time', then copy this cell to populate the rest of the cells in that column, by highlighting that cell, pull the cell boundary box down the column to the last (you'll see where by looking at the bottom of the adjacent cell) row, and you'll notice the spreadsheet populates the range with *CET*.

Important! Highlight the column, right-click to bring up a context menu and select **Format Cells**, change the default *General* to *Text* and select **OK**—this will ensure that when we save the file in CSV format later it will place double-quotes around the cell contents, as this new column is expecting quoted text in the form of a time zone.

New column 11 (K) for zone hours: Create a new column, in the first cell of that column add *0*, then copy this cell to populate the rest of the cells in that column by highlighting that cell and selecting **Ctrl+C**, highlighting all the cells down to the last (you'll see where by looking at the bottom of the adjacent cell)—you'll notice the spreadsheet populates the range with incremented numbers—select **Ctrl+V** which will replace the entered sequential numbers with *0*s.

New column 12 (L) for quoted full place name: Highlight old column 3 (C), then do an **Edit | Find & Replace**, put in **F** to be found and replaced with **France**, then select **OK**. Do the same sort of thing for the other country codes specified under the heading *Repartition by countries (vol. 1)*.

Then go to **Sheet2**, column L row 1 (cell: **L1**) and enter the following formula into this cell (note: this is all one line):

LibreOffice Calc:

```
=$902gdA1.M1 & ", " & $902gdA1.N1 & ", " & $902gdA1.C1
```

MS Excel:

```
=902gdA1!M1 & ", " & 902gdA1!N1 & ", " & 902gdA1!C1
```

and hit **Enter** key.

Then drag the cell box to the bottom of the column: this will apply the formula relatively to all the cells in the column. What this does is effectively merge the three columns from the old sheet to produce the new *place name* column.

Then convert the cells in the column from *formula* cells to *value* cells. In *LibreOffice Calc*, this is done by highlighting all of the cells in the column, then select **Tools | Cell Contents . . . | Formula to Value**.

MS Excel will do something similar.

Important! Highlight the column, right-click to bring up a context menu and select **Format Cells**, change the default *General* to *Text* and select **OK**—this will ensure that when we save the file in CSV format later, it will place double-quotes around the cell contents, as this new column is expecting quoted text in the form of a full place name.

New column 13 (M) for decimal longitude:

Stage One: The figures in the column *must have* two numbers before 'W' or 'E', so to ensure this is so create a temporary column in **Sheet2**, go to column S row 1 ; then (cell: **S1**) and enter the following formula into this cell (note: this is all one line):

LibreOffice Calc:

```
=IF(IFERROR(FIND(OR("E"; "W"), $902gdA1.L1, 2), 0), "", "0")
& $902gdA1.L1
```

MS Excel:

```
=IF(IFERROR(FIND(OR("E"; "W"), 902gdA1!L1, 2), 0), "", "0")
& 902gdA1!L1
```

and hit **Enter** key.

Then drag the cell box to the bottom of the column: this will apply the formula relatively to all the cells in the column. What this does is effectively converts degrees and minutes into a decimal to produce a temporary *longitude* column.

Then convert the cells in the column from *formula* cells to *value* cells. In *LibreOffice Calc*, this is done by highlighting all of the cells in the column, then select **Tools | Cell Contents . . . | Formula to Value**.

MS Excel will do something similar.

Stage Two: Now that the format of the longitude column is standardised, we need to convert it to decimal format, with a minus representing East. Go to **Sheet2**, column M row 1; then (cell: **M1**) and enter the following formula into this cell (note: this is all one line):

LibreOffice Calc:

```
=IF(IFERROR(FIND("E", S1), 0), -1, +1)
* VALUE(LEFT(S1, 2)) + VALUE(RIGHT(S1, 2)) / 60
```

MS Excel:

```
=IF(IFERROR(FIND("E", S1), 0), -1, +1)
* VALUE(LEFT(S1, 2)) + VALUE(RIGHT(S1, 2)) / 60
```

and hit **Enter** key.

Then drag the cell box to the bottom of the column: this will apply the formula relatively to all the cells in the column. What this does is effectively converts degrees and minutes into a decimal to produce the new *longitude* column.

Then convert the cells in the column from *formula* cells to *value* cells. In *LibreOffice Calc*, this is done by highlighting all of the cells in the column, then select **Tools | Cell Contents . . . | Formula to Value**.

MS Excel will do something similar.

Stage Three: Highlight column M, right-click to bring up a context menu and select **Format Cells**, change the format to *Number* if it isn't already, change the *Format Code* to **00.00**, and select **OK**. This will ensure that the number is in the correct import format.

Tidy up by deleting the temporary column S.

New column 14 (N) for decimal latitude:

Stage One: The figures in the column *must have* two numbers before 'N' or 'S', so to ensure this is so create a temporary column in **Sheet2**, go to column S row 1 ; then (cell: **S1**) and enter the following formula into this cell (note: this is all one line):

LibreOffice Calc:
```
=IF(IFERROR(FIND(OR("S"; "N"), $902gdA1.K1, 2), 0), "", "0")
& $902gdA1.K1
```

MS Excel:
```
=IF(IFERROR(FIND(OR("S"; "N"), 902gdA1!K1, 2), 0), "", "0")
& 902gdA1!K1
```

and hit **Enter** key.

Then drag the cell box to the bottom of the column: this will apply the formula relatively to all the cells in the column. What this does is effectively converts degrees and minutes into a decimal to produce a temporary *latitude* column.

Then convert the cells in the column from *formula* cells to *value* cells. In *LibreOffice Calc*, this is done by highlighting all of the cells in the column, then select **Tools | Cell Contents . . . | Formula to Value**.

MS Excel will do something similar.

Stage Two: Now that the format of the latitude column is standardised, we need to convert it to decimal format, with a minus representing South. Go to **Sheet2**, column N row 1; then (cell: **N1**) and enter the following formula into this cell (note: this is all one line):

LibreOffice Calc:

```
=IF(IFERROR(FIND("S", S1), 0), -1, +1)
* VALUE(LEFT(S1, 2)) + VALUE(RIGHT(S1, 2)) / 60
```

MS Excel:

```
=IF(IFERROR(FIND("S", S1), 0), -1, +1)
* VALUE(LEFT(S1, 2)) + VALUE(RIGHT(S1, 2)) / 60
```

and hit **Enter** key.

Then drag the cell box to the bottom of the column: this will apply the formula relatively to all the cells in the column. What this does is effectively converts degrees and minutes into a decimal to produce the new *latitude* column.

Then convert the cells in the column from *formula* cells to *value* cells. In *LibreOffice Calc*, this is done by highlighting all of the cells in the column, then select **Tools | Cell Contents . . . | Formula to Value**.

MS Excel will do something similar.

Stage Three: Highlight column N, right-click to bring up a context menu and select **Format Cells**, change the format to *Number* if it isn't already, change the *Format Code* to **00.00**, and select **OK**. This will ensure that the number is in the correct import format.

Tidy up by deleting the temporary column S.

New column 15 (O) for daylight saving hours: Copy from old column 10 (J).

This is because from inspection it seems that this is the column where Gauquelin has used '-1' to determine whether it requires removing the Daylight Saving hour to convert it to GMT.

New column 16 (P) for 0: Create a new column, in the first cell of that column add *0*, then copy this cell to populate the rest of the cells in that column, by highlighting that cell and selecting **Ctrl+C**, highlighting all the cells down to the last (you'll see where by looking at the bottom of the adjacent cell)—you'll notice the spreadsheet populates the range with incremented numbers—select **Ctrl+V** which will replace the entered sequential numbers with *0*s.

New column 17 (Q) for " ": Create a new column, in the first cell of that column add a space, then copy this cell to populate the rest of the cells in that column by highlighting that cell and selecting **Ctrl+C**, highlighting all the cells down to the last (you'll see where by looking at the bottom of the adjacent cell)—the spreadsheet will populate the range with a single space.

Important! Highlight the column, right-click to bring up a context menu and select **Format Cells**, change the default *General* to *Text* and select **OK**—this will ensure that when we save the file in CSV format later, it will place double-quotes around the cell contents, as this new column is expecting quoted 'null string'.

5.2 Import the File

Whatever you do, make sure you save it as `sports.xls` or `sports.ods`. Keep the spreadsheet open, with **Sheet2** active.

Then **File | Save a Copy ...** a box will appear where you can change the format of the file to save as *Text CSV (.csv)*. Then enter the name to save it as `sports.csv` to your work directory; another box should appear with options ...

> In the case of *LibreOffice Calc*, ensure that the **Quote All Text Cells** and **Save Content As Shown** boxes are checked. The comma should be the **Field Delimiter** and double-quotes, the **Text Delimiter** by default. MS Excel should have something similar.

Then save the file as `sports.csv` to your work directory. Close the spreadsheet application.

Use Windows Explorer to locate your working directory, and rename `sports.csv` as `sports.dat`. Then find open it using your text editor. Check it over and compare the format to `Actors.dat`. Assuming everything is okay, backup your files. Then copy `sports.dat` into the `Userdata` folder under the `Jigsaw` folder.

Load up *Jigsaw* and open the `sports.dat` file. If this has worked, it will whizz through importing your data and coming up as a success. If not, it's back to the drawing board and see where the import file is failing, comparing it with the import requirements.

If it hasn't, go back through everything to find where you have gone wrong.

About the Author

Derek Norcott has studied and researched astrology since 1969. He is self-educated and has a large collection of books on astrology. He has attended the Astrological Lodge regularly for about twelve years, has presented a seminar at the Lodge, and has previously written articles for the Quarterly. Derek is a past member of the Astrological Association. He is also a web developer and typesetter. He keeps a diary and urges other astrologers to do the same, as it is one of the best methods for testing astrology to one's personal satisfaction.