
Subject: READ_CSV() gotcha
Posted by [wlandsman](#) on Fri, 03 Apr 2015 17:44:57 GMT
[View Forum Message](#) <> [Reply to Message](#)

The documentation for READ_CSV() describes how a column is stored as Double, rather than an integer type, if it has a decimal point or an exponent.

What it doesn't say is that READ_CSV() only looks at the first 100 rows, so that if these first 100 numbers are compatible with an integer (no decimal points or exponents) then the entire column is read as an integer. In my case, the data becomes floating point after about 2000 rows (with values between 0 and 1), so these are all truncated to zero.

There doesn't seem to be an easy fix, e.g. a way to force a column to be read as Double, so I ended up writing a specialized reader. --Wayne

Subject: Re: READ_CSV() gotcha
Posted by [penteado](#) on Wed, 08 Apr 2015 23:36:51 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello Wayne,

It seems this has come in a bit late, but I have an altered version of read_csv(), which provides a bunch of additional options, including allowing the user to choose how many rows to use for testing (select 0, to use all rows), it makes the resulting structure get field names based on the csv header, and has more elaborate options to control the testing for column types:

http://ppenteado.net/idl/pp_lib/doc/read_csv_pp.html

I generally use it with the /transp keyword, so that the result is a structure array, with one element per csv row, which I find more useful.

It also has a companion, write_csv_pp, which is a wrapper to write_csv, to create a csv with a table header with a structure's field names.

http://ppenteado.net/idl/pp_lib/doc

On Friday, April 3, 2015 at 2:45:00 PM UTC-3, wlandsman wrote:

> The documentation for READ_CSV() describes how a column is stored as Double, rather than an integer type, if it has a decimal point or an exponent.

>

> What it doesn't say is that READ_CSV() only looks at the first 100 rows, so that if these first 100 numbers are compatible with an integer (no decimal points or exponents) then the entire column is read as an integer. In my case, the data becomes floating point after about 2000 rows (with values between 0 and 1), so these are all truncated to zero.

>

> There doesn't seem to be an easy fix, e.g. a way to force a column to be read as Double, so I ended up writing a specialized reader. --Wayne
