Title

stata.com

html2docx — Convert an HTML file to a Word (.docx) document

Description

`html2docx` converts an HTML file to a Word (`.docx`) document. The HTML file can be either a file on the local disk or a URL on a remote website.

Quick start

Convert HTML file `myfile.html` to a Word document saved as `myfile.docx`

```
html2docx myfile
```

Same as above, but save the Word document as `mydoc.docx`

```
html2docx myfile, saving(mydoc)
```

Same as above, and overwrite the existing `mydoc.docx`

```
html2docx myfile, saving(mydoc) replace
```

Syntax

`html2docx` *srcfile* [, *options*]

srcfile is an HTML file, either a local file or a URL. If *srcfile* is specified without an extension, `.html` is assumed. If *srcfile* contains embedded spaces or other special characters, enclose it in double quotes.

options	Description
`saving(`*targetfile*`)`	specify the target Word (`.docx`) document to be saved
`replace`	replace the target Word (`.docx`) document if it already exists
`nomsg`	suppress message with link to *targetfile*
`base(`*string*`)`	specify the base directory or base URL for relative links in *srcfile*

Options

`saving(`*targetfile*`)` specifies the target Word (`.docx`) document file to be saved. If *targetfile* is specified without an extension, `.docx` is assumed. If *targetfile* contains embedded spaces or other special characters, enclose it in double quotes. If `saving()` is not specified, the target filename is constructed using the source filename (*srcfile*) with the `.docx` extension. `saving()` is required if the *srcfile* is a URL.

`replace` specifies that the target Word (`.docx`) document be replaced if it already exists.

`nomsg` suppresses the message that contains a link to the target file.

`base(`*string*`)` specifies the base directory or the base URL for the relative links in the *srcfile*.

Remarks and examples

stata.com

`html2docx` converts HTML files to Word (`.docx`) documents. It attempts to preserve the styles of various HTML elements in the `.docx` file. However, for some HTML elements, there is no direct translation for a `.docx` file. For instance, an apostrophe in an HTML file may be replaced with another character in the `.docx` file. Thus, your target Word document may require some cleaning after the `html2docx` conversion.

`html2docx` expects a valid HTML file—one that contains essential HTML elements such as `<!DOCTYPE html>`, `<html>`, `<head>`, and `<body>`. If the HTML file is not valid, `html2docx` will go through a tidying process to attempt to make it valid. `html2docx` will produce an error message if this tidying process fails. You may check whether an HTML file is valid by using the W3C online Markup Validation Service at https://validator.w3.org/#validate_by_upload+with_options.

If you are working with a Markdown-formatted text file, you can convert this file directly to a Word document by specifying the `docx` option with `markdown`; see [RPT] **markdown**. Similarly, you can use `dyndoc` to convert a text file with Stata commands and Markdown-formatted text to a Word document with Stata output; see [RPT] **dyndoc**.

▷ Example 1: Converting an HTML file to a Word document

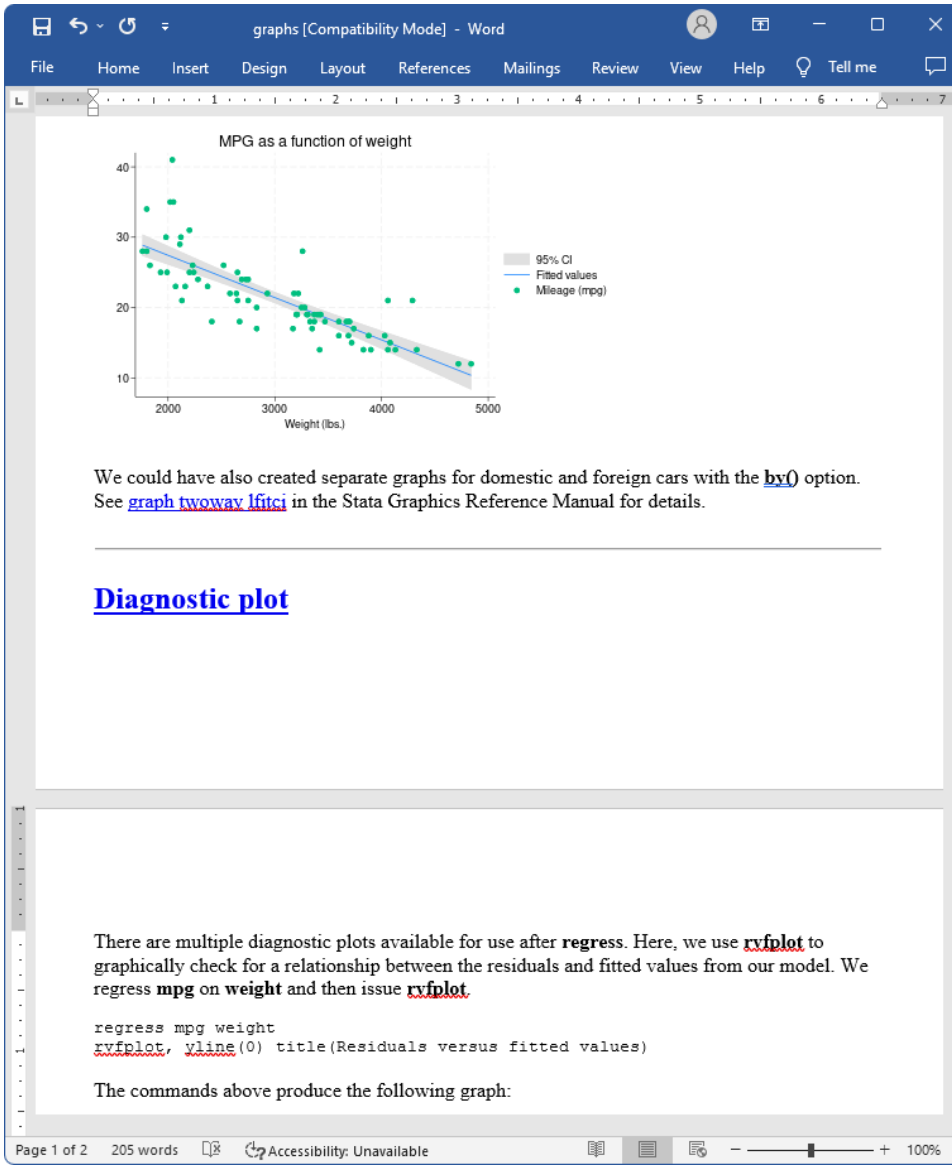
We have an HTML file, `graphs.html`, that includes some Stata graphs. You can copy this file to your current working directory by typing

```
. copy https://www.stata-press.com/data/r18/reporting/graphs.html .
```

To convert `graphs.html` to a Word document, we type

```
. html2docx graphs.html
```

The file is saved as `graphs.docx`. Here is a portion of this file:



We display the second page of the document to show both text and images found in the Word document. You can see the whole file at https://www.stata-press.com/data/r18/reporting/graphs.docx.

Also see

[RPT] **docx2pdf** — Convert a Word (.docx) document to a PDF file

[RPT] **dyndoc** — Convert dynamic Markdown document to HTML or Word (.docx) document

[RPT] **markdown** — Convert Markdown document to HTML file or Word (.docx) document

[RPT] **putdocx intro** — Introduction to generating Office Open XML (.docx) files