Table 1 – continued from previous page

Dependency	Minimum Version	Notes
gcsfs	0.2.2	Google Cloud Storage access
html5lib		HTML parser for read_html (see *note*)
lxml	3.8.0	HTML parser for read_html (see *note*)
matplotlib	2.2.2	Visualization
numba	0.46.0	Alternative execution engine for rolling operations
openpyxl	2.5.7	Reading / writing for xlsx files
pandas-gbq	0.8.0	Google Big Query access
psycopg2		PostgreSQL engine for sqlalchemy
pyarrow	0.12.0	Parquet, ORC (requires 0.13.0), and feather reading / writing
pymysql	0.7.11	MySQL engine for sqlalchemy
pyreadstat		SPSS files (.sav) reading
pytables	3.4.2	HDF5 reading / writing
pyxlsb	1.0.6	Reading for xlsb files
qtpy		Clipboard I/O
s3fs	0.3.0	Amazon S3 access
tabulate	0.8.3	Printing in Markdown-friendly format (see tabulate)
xarray	0.8.2	pandas-like API for N-dimensional data
xclip		Clipboard I/O on linux
xlrd	1.1.0	Excel reading
xlwt	1.2.0	Excel writing
xsel		Clipboard I/O on linux
zlib		Compression for HDF5

Optional dependencies for parsing HTML

One of the following combinations of libraries is needed to use the top-level `read_html()` function:

Changed in version 0.23.0.

- BeautifulSoup4 and html5lib

- BeautifulSoup4 and lxml

- BeautifulSoup4 and html5lib and lxml

- Only lxml, although see *HTML Table Parsing* for reasons as to why you should probably **not** take this approach.

> **Warning:**
>
> - if you install BeautifulSoup4 you must install either lxml or html5lib or both. `read_html()` will **not** work with *only* BeautifulSoup4 installed.
>
> - You are highly encouraged to read *HTML Table Parsing gotchas*. It explains issues surrounding the installation and usage of the above three libraries.