Importing/exporting data

- **Import and export data from Excel .xls and .xlsx files**
- **Import and export CSV and delimited data**
- Copy/paste data from spreadsheets
- Input data in spreadsheet editor
- Read from and write to SQL sources with ODBC (see below)
- Import and export fixed-format data using a dictionary
- Import and export any type of text data
- Unicode (UTF-8) support, including conversion from/to extended ASCII
- Import EBCDIC data and convert EBCDIC to ASCII
- Import and export data in the format required by the FDA for NDA submittals
- Import and export SAS Transport XPORT files
- **Import Federal Reserve Economic Data** New
- Import from Haver Analytics databases
- **Import data from Wharton Research Data Services (WRDS) via ODBC**
- Import and export dBase files New
- High-level import/export of full Excel worksheets
- **Low-level cell-by-cell access to write results to and read data from Excel, including graphs, formulas, date formats, currency formats, bold, italics, and more**

Watch **Import FRED (Import Federal Reserve Economic Data)**.

Watch **Import Excel® data into Stata**.
Watch **Copy/Paste from Excel® into Stata**.

ODBC support

- Import data from any ODBC data source, such as Oracle, SQL Server, Access, Excel, MySQL, and DB2
- Export data to new or existing ODBC tables
- Execute custom SQL commands individually or in batches

Automatic memory management

- Up to 1.5 TB of RAM supported
- Up to 120,000 variables in Stata/MP; up to 32,767 variables in Stata/SE Updated
- **20 billion or more observations in Stata/MP**
- Up to 2.1 billion observations (Stata/SE and Stata/IC)

Sorting

- Ascending or descending sorts
- Multiple-key sorts
- Numeric and string sorts
- Locale-aware Unicode string sorting and comparison

Combining datasets

- Merge datasets
  - By key variables
  - By observations
- Join datasets
- Outer join
- Append datasets
- Append time series

Special datasets

- **Longitudinal data/panel data**
- **Survival/duration data**
- **Time series**
- **Survey data**
- **Multiple imputations**

Dynamic document generation New

- **Create HTML files with embedded Stata code, output, and graphs**
- **Markdown**

Watch **Automatic production of web pages from dynamic Markdown documents**.

- Customize ODBC connection strings
- Support for ODBC
- Support for VARCHARs/CLOBs and BLOBs
- Support for Unicode

Built-in spreadsheet editor
- Clipboard Preview Tool lets you control how data will be pasted
- Manage variables with the Variables Tool
- For Windows 📄, Mac 📄, and Unix 📄

Properties window
- Manage variables
- Manage dataset properties
- For Windows 📄, Mac 📄, and Unix 📄

Variables Manager
- Change storage types, names, and formats
- Add and edit value labels
- Attach notes to variables
- Filter variables
- For Windows 📄, Mac 📄, and Unix 📄

Watch **Label variables**.
Watch **Label the values of categorical variables**.
Watch **Change the display format of a variable**.
Watch **Add notes to a variable**.

Functions 📄
- Statistical functions 📄 Updated
- Mathematical functions 📄
- Trigonometric functions 📄
- String functions 📄
- Unicode functions 📄
- Regular expressions 📄
- Date and time functions 📄
- Time-series functions 📄
- Random-number functions 📄
  - 18 functions 📄 Updated
  - Stream random numbers 📄 New
- Matrix functions 📄 📄
- Programming functions 📄

Watch **Round a continuous variable**.

Data reorganization
- Row−column transposition 📄
- Data reshaping 📄
- Stacking of variables 📄
- Collapsing into means, totals, etc. 📄

Watch **Reshape data from wide format to long format**.
Watch **Reshape data from long format to wide format**.

Creation of Word, Excel, and PDF files
- **High-level creation of Word documents containing Stata results and graphs** 📄 New
- Low-level programmatic access for fine-control creation of Word documents 📄 Updated
- **High-level creation of PDF files containing Stata results and graphs** 📄 New
- Low-level programmatic access for fine-control creation of PDF files 📄 Updated
- High level import/export of full Excel worksheets 📄
- Low-level cell-by-cell access to write results to and read data from Excel, including graphs, formulas, date formats, currency formats, bold, italics, and more. 📄 📄

Watch **Create Word® documents from within Stata**.
Watch **Create PDF reports from within Stata**.

Image output
- Save graphs as PDFs 📄
- Save graphs to EPS or TIFF files for publication 📄 📄 📄
- Save graphs to PNG or SVG files for the web 📄 📄 Updated

Utilities
- Count number of observations that satisify specified conditions 📄
- Formatted and unformatted disk I/O 📄 📄
- Zip-file support 📄
- Unicode conversion from/to extended ASCII 📄 📄
- Custom filters to manipulate text files 📄

Variable management
- Generation of new variables 📄 📄
- Replacement of existing variables 📄
- Renaming variables 📄 📄
- Encoding and decoding string variables 📄
- Reordering variables in dataset 📄

Watch **Create a new variable that is calculated from other variables**.
Watch **Convert categorical string variables to labeled numeric variables**.
Watch **Create a categorical variable from a continuous variable**.
Watch **Convert a string variable to a numeric variable**.
Watch **Identify and replace unusual data values**.

Dataset utilities
- Flexible description of variables, labels, and types 📄

Unicode support

- UTF-8
- Translation of extended ASCII to UTF-8
- Unicode-aware string functions
- Locale-based sorting and string comparison

Watch **Unicode in Stata**.

Labels

- Dataset labels
- Variable labels
- Value labels (e.g., male and female for 0 and 1)
- Ability to switch between multiple sets of data, variable, and value labels
- Missing-value labels
- Support for multiple languages, including Unicode support

Notes

- Extensive notes can be attached to a dataset

Data snapshots

- Allow multiple levels of undo to modified datasets

- List values of variables
- Data signatures to verify the integrity of datasets
- Codebooks for variables
- Value-label reports
- Duplicates and missing values tables
- Compress (make dataset as small as possible without loss of accuracy)

Watch **Optimize the storage of variables**.
Watch **Identify and remove duplicate observations**.

Variable types

- Numeric storage types
  - Byte
  - Integer (int)
  - Long
  - Float
  - Double
- String (including Unicode, very long strings and BLOBs)
- Dates and times
- **Business calendars**

Watch **Create a date variable from a date stored as a string**.

Long string support

- Up to 2 billion character long strings
- Coalescing of duplicate values to save memory
- Binary 'strings' (BLOBs)
- Import and export entire files into long strings/BLOBs
- Unicode (UTF-8) strings

Stored results

- Save results to disk for later use
- Store estimation results in memory
- Create tables to compare results