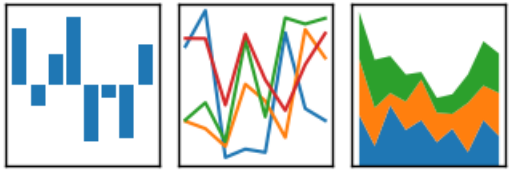
Fork me on GitHub

pandas

$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$



Python Data Analysis Library

pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

pandas is a NumFOCUS sponsored project. This will help ensure the success of development of *pandas* as a world-class open-source project, and makes it possible to donate to the project.

A Fiscally Sponsored Project of
NUMFOCUS
OPEN CODE = BETTER SCIENCE

v0.23.4 Final (August 3, 2018)

This is a minor bug-fix release in the 0.23.x series and includes some regression fixes, bug fixes, and performance improvements. We recommend that all users upgrade to this version.

The release can be installed with conda from conda-forge or the default channel:

```
conda install pandas
```

Or via PyPI:

```
python3 -m pip install --upgrade pandas
```

See the full whatsnew for a list of all the changes.

VERSIONS

Release
0.23.4 - August 2018
download // docs // pdf

Development
0.24.0 - 2018
github // docs

Previous Releases
0.23.3 - download // docs // pdf
0.23.2 - download // docs // pdf
0.23.1 - download // docs // pdf
0.23.0 - download // docs // pdf
0.22.0 - download // docs // pdf
0.21.1 - download // docs // pdf
0.21.0 - download // docs // pdf
0.20.3 - download // docs // pdf
0.19.2 - download // docs // pdf
0.18.1 - download // docs // pdf
0.17.1 - download // docs // pdf
0.16.2 - download // docs // pdf
0.15.2 - download // docs // pdf
0.14.1 - download // docs // pdf
0.13.1 - download // docs // pdf
0.12.0 - download // docs // pdf

ABOUT PANDAS

What's New
Getting Started
Issue Tracker

v0.23.0 Final (May 15, 2018)

This is a major release from 0.22.0 and includes a number of API changes, new features, enhancements, and performance improvements along with a large number of bug fixes.

Highlights include:

- Round-trippable JSON format with 'table' orient.
- Instantiation from dicts respects order for Python 3.6+.
- Dependent column arguments for assign.
- Merging / sorting on a combination of columns and index levels.
- Extending Pandas with custom types.
- Excluding unobserved categories from groupby.

The release candidate can be installed with conda from our development channel (builds for osx-64, linux-64 and win-64 for Python 2.7, Python 3.5, and Python 3.6 are all available):

```
conda install pandas
```

or conda forge:

```
conda install -c conda-forge pandas
```

Or via PyPI:

```
python3 -m pip install --upgrade pandas==0.23.0
```

See the full whatsnew for a list of all the changes.

Best way to Install

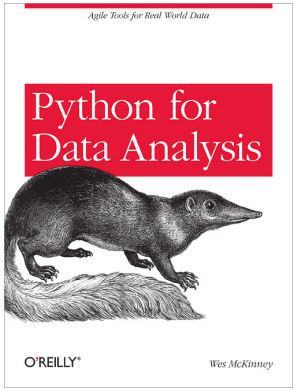
The best way to get pandas is via conda

```
conda install pandas
```

License
Contributing Guidelines

Fork me on GitHub

GET THE BOOK



RELATED TOOLS

SciPy
NumPy
StatsModels
scikit-learn
Jupyter
matplotlib

Packages are available for all supported python versions on Windows, Linux, and MacOS.

Wheels are also uploaded to PyPI and can be installed with

```
pip install pandas
```

Quick vignette

10-minute tour of pandas from Wes McKinney on Vimeo.

What problem does *pandas* solve?

Python has long been great for data munging and preparation, but less so for data analysis and modeling. *pandas* helps fill this gap, enabling you to carry out your entire data analysis workflow in Python without having to switch to a more domain specific language like R.

Combined with the excellent IPython toolkit and other libraries, the environment for doing data analysis in Python excels in performance, productivity, and the ability to collaborate.

pandas does not implement significant modeling functionality outside of linear and panel regression; for this, look to statsmodels and scikit-learn. More work is still

needed to make Python a first class statistical modeling environment, but we are well on our way toward that goal.

What do our users have to say?

Roni Israelov, PhD
Portfolio Manager
AQR Capital Management

> "*pandas* allows us to focus more on research and less on programming. We have found *pandas* easy to learn, easy to use, and easy to maintain. The bottom line is that it has increased our productivity."

David Himrod
Director of Optimization & Analytics
AppNexus

> "*pandas* is the perfect tool for bridging the gap between rapid iterations of ad-hoc analysis and production quality code. If you want one tool to be used across a multi-disciplined organization of engineers, mathematicians and analysts, look no further."

Olivier Pomel
CEO
Datadog

> "We use *pandas* to process time series data on our production servers. The simplicity and elegance of its API, and its high level of performance for high-volume datasets, made it a perfect choice for us."

Library Highlights

- A fast and efficient **DataFrame** object for data manipulation with integrated indexing;
- Tools for **reading and writing data** between in-memory data structures and different formats: CSV and text files, Microsoft Excel, SQL databases, and the fast HDF5 format;

Fork me on GitHub

Fork me on GitHub

- Intelligent **data alignment** and integrated handling of **missing data**: gain automatic label-based alignment in computations and easily manipulate messy data into an orderly form;
- Flexible **reshaping** and pivoting of data sets;
- Intelligent label-based **slicing**, **fancy indexing**, and **subsetting** of large data sets;
- Columns can be inserted and deleted from data structures for **size mutability**;
- Aggregating or transforming data with a powerful **group by** engine allowing split-apply-combine operations on data sets;
- High performance **merging and joining** of data sets;
- **Hierarchical axis indexing** provides an intuitive way of working with high-dimensional data in a lower-dimensional data structure;
- **Time series**-functionality: date range generation and frequency conversion, moving window statistics, moving window linear regressions, date shifting and lagging. Even create domain-specific time offsets and join time series without losing data;
- Highly **optimized for performance**, with critical code paths written in Cython or C.
- Python with *pandas* is in use in a wide variety of **academic and commercial** domains, including Finance, Neuroscience, Economics, Statistics, Advertising, Web Analytics, and more.