



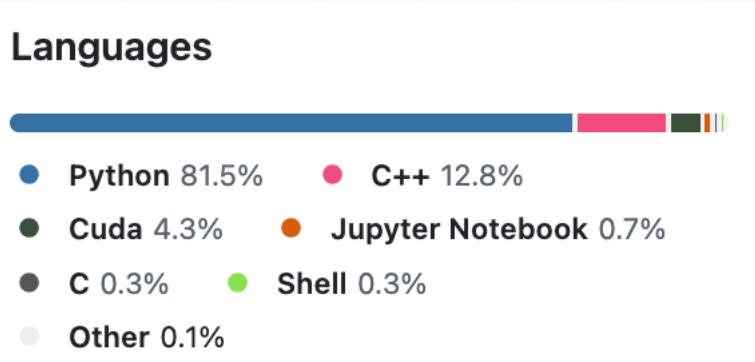
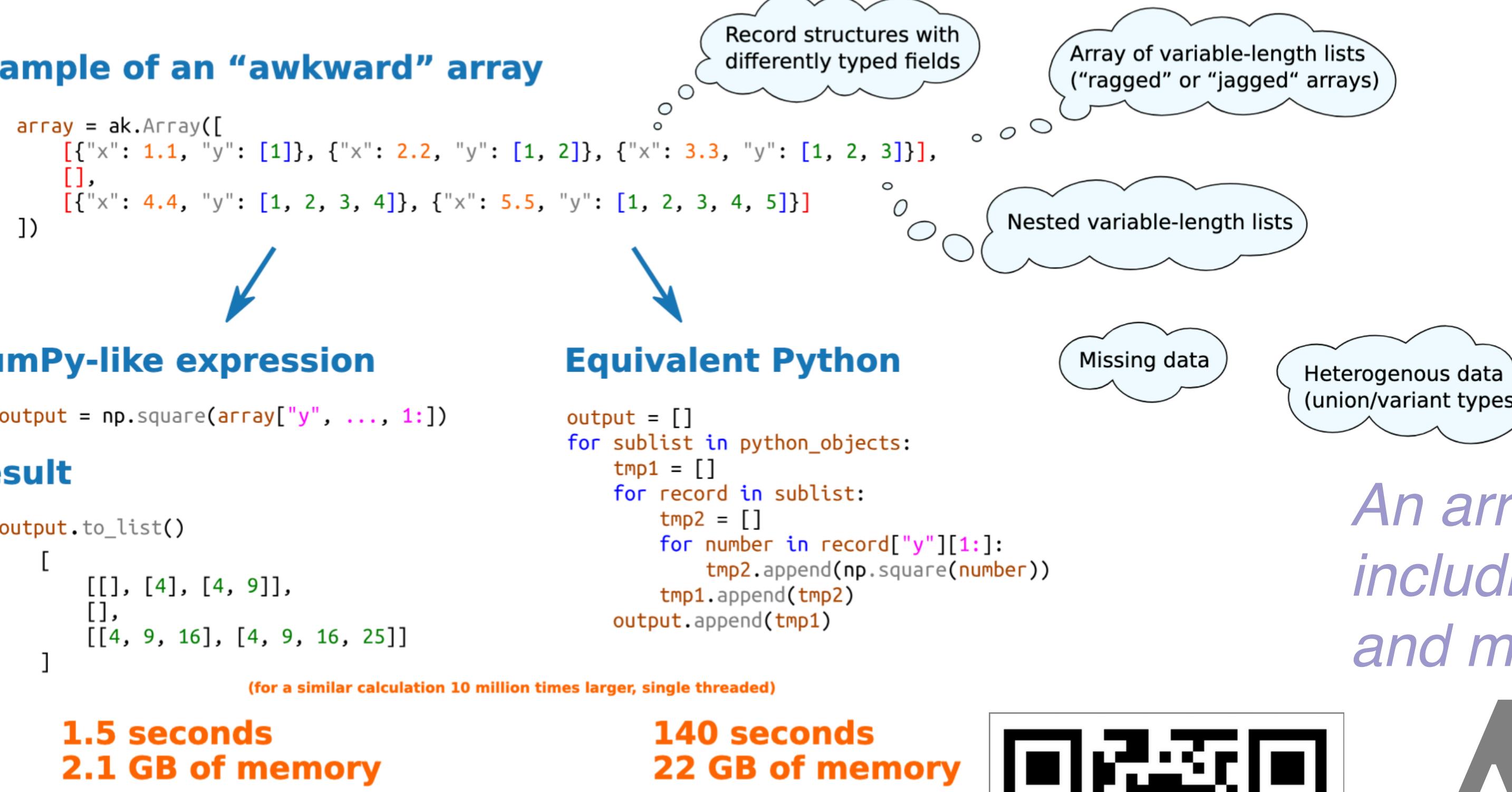
Awkward Arrays: Accelerating scientific data analysis on irregularly shaped data

Framework, OAC-2103945

PI: J. Pivarski (P. Elmer)

What Awkward Array does

Example of an “awkward” array

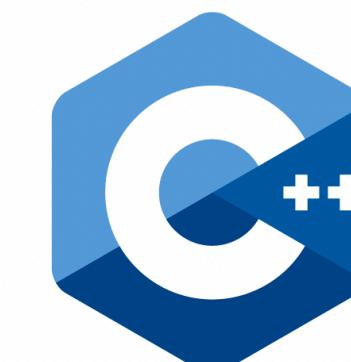


Awkward Array Library Integrations

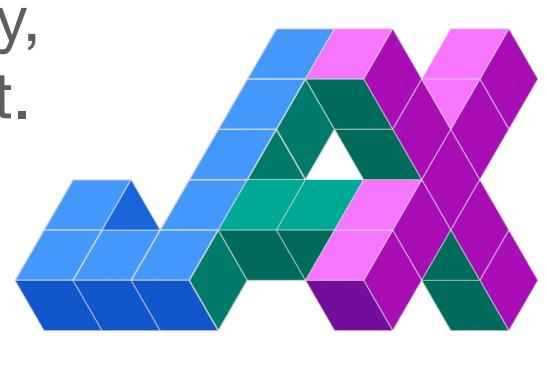
Awkward Array is a part of larger scientific Python ecosystem. It must work well with other libraries.



JIT-compiled as a C++ iterator in ROOT's RDataFrame workflow.



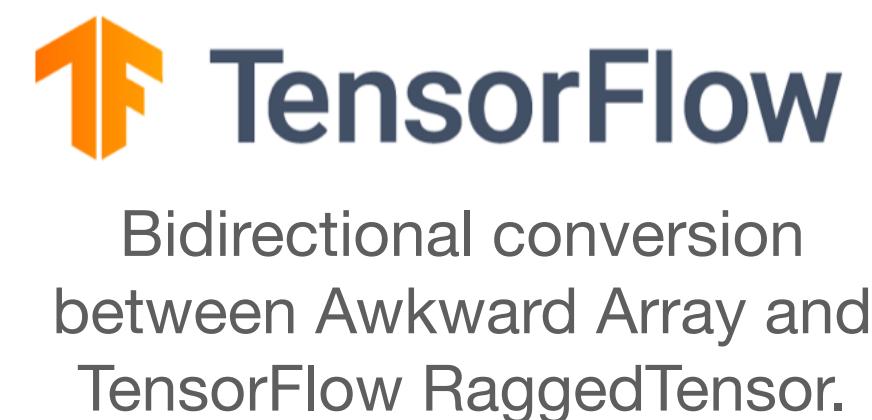
JIT-compile Awkward Arrays in C++ + with cppyy. Header-only library to build Awkward Arrays in C++ and transfer them to Python.



Efficiently iterate over or build Awkward Arrays in code JIT-compiled with Numba, on CPUs and GPUs



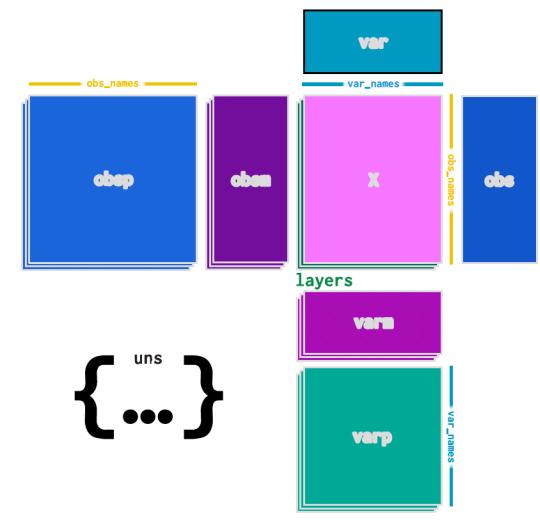
Awkward Arrays on GPUs, backed by CuPy.



Bidirectional conversion between Awkward Array and TensorFlow RaggedTensor.



dask-awkward is a high-level collection type for Dask, alongside Array and DataFrame.



Store Awkward Arrays in the Tiled database and slice them in the cloud.



Integrated into AnnData for single-cell genomics.



Read any Kaitai Struct format into Awkward Arrays as DataFrame columns.

Awkward Releases Highlights (Aug 2024 - now)

- Named axis support for ak.Array - making Awkward Array's multidimensional irregular data manipulation readable and robust
- Virtual Arrays support - “read what you need”
- Enhanced error reporting and performance improvements
- Improved and more complete GPU support - CUDA backend
- Major progress on segmented reducers: argmin, argmax, count_nonzero
- Expanded Tensor framework support (TensorFlow, PyTorch, JAX, cuDF)
- Continuous compatibility maintenance (Python 3.13, Numpy 2.3, etc.)
- Concurrents composable parallelism (currently 3-5x faster than serial)
- Extensible columnar memory model (nested columnar arrays)

An array library for **nested, variable-sized data**, including arbitrary-length lists, records, mixed types, and missing data, using **NumPy-like idioms**.

Awkward Array

Arrays are **dynamically typed**, but operations on them are **compiled and fast**.

Coincides with NumPy when arrays are regular; generalizes when they're not.

Adoption

Awkward Array was developed in the context of particle physics

"Non-HEP" use-cases include

- neutrinos: P-ONE, CIEMAT-Neutrino, Harvard-Neutrino, LEGEND, DUNE, ICARUS, IceCube, nEXO, GraphNeT
- statistics: errors in annotated data (UKPLab/nessie), Bayesian analysis (bat/batty), cmpatino/optimal-observables multivariate shapelets (bianchimario/MARS), text embedding (taylorai/mlx_embedding_models)
- climate science: Lagrangian ocean probes (Cloud-Drift/clouddrift), cloud microphysics (yoctoyotta1024/CLEO)
- biology: single-cell genetics (scverse), genotyping (kage-genotyper/kage), astrocytes (janreising/astroCAST)
- astronomy: active galactic nuclei (Zstone19/TempMap), dark matter direct detection (XENONnT/fuse)
- defense: identifying missile launches through ionic disturbances in GPS (tylerni7/missile-tid)
- other: simulation of fire (silvxlabs/DripTorch), food orders (EASS-HIT-PART-A-2022-CLASS-II/food-ordering), analyzing lastFM data (delannoy/lastfm-explorer), software (WayScience/software-landscape-analysis), polygon meshes (dewloosh/PolyMesh), CAD (j8sr0230/codelink, AxisVM/pyaxisvm), adarshchbs/pdf2md

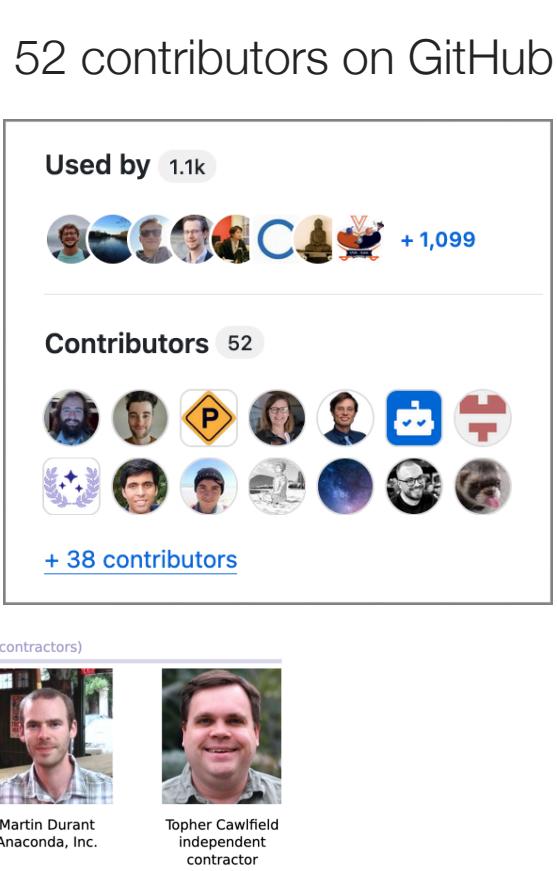
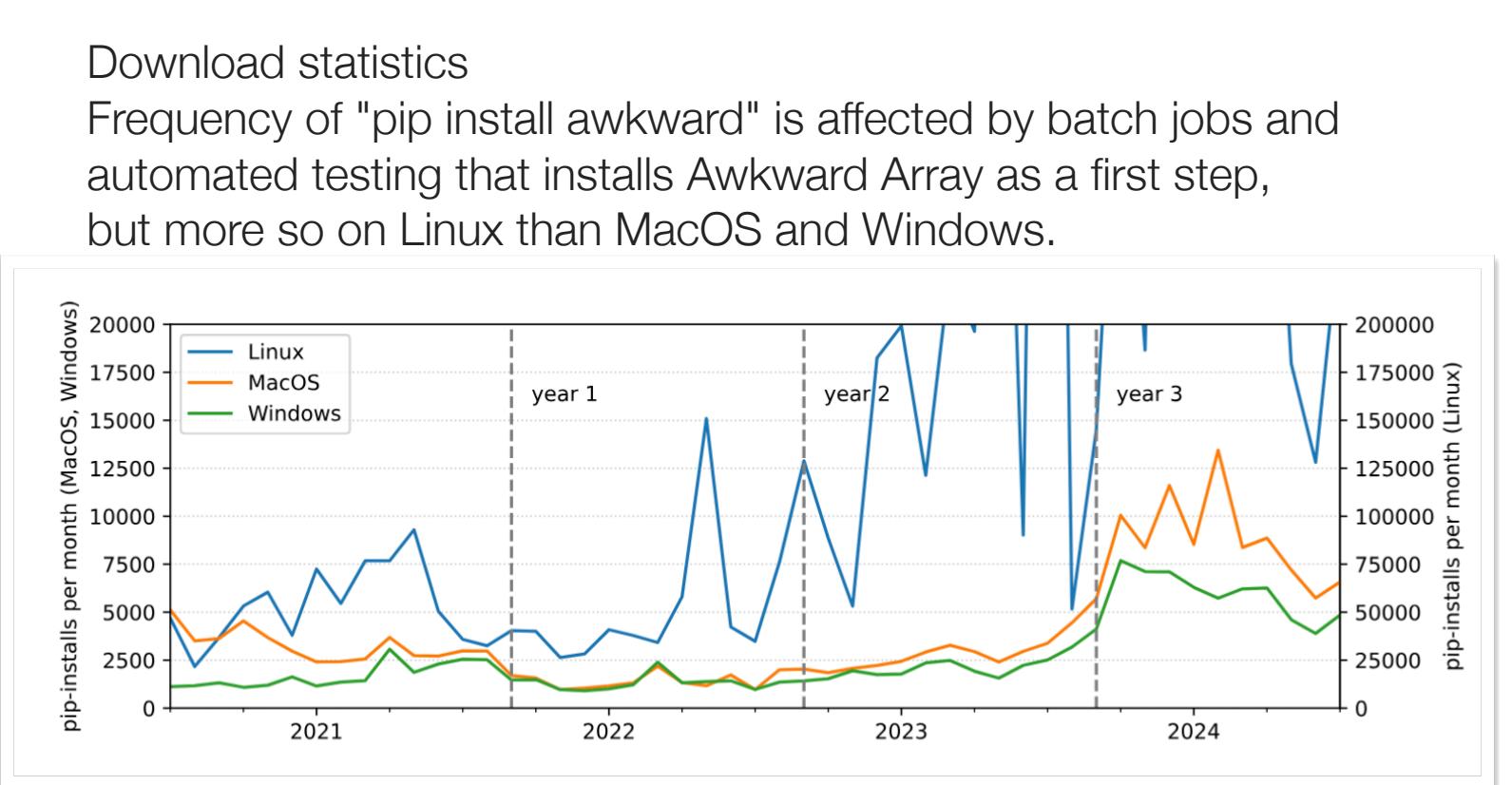
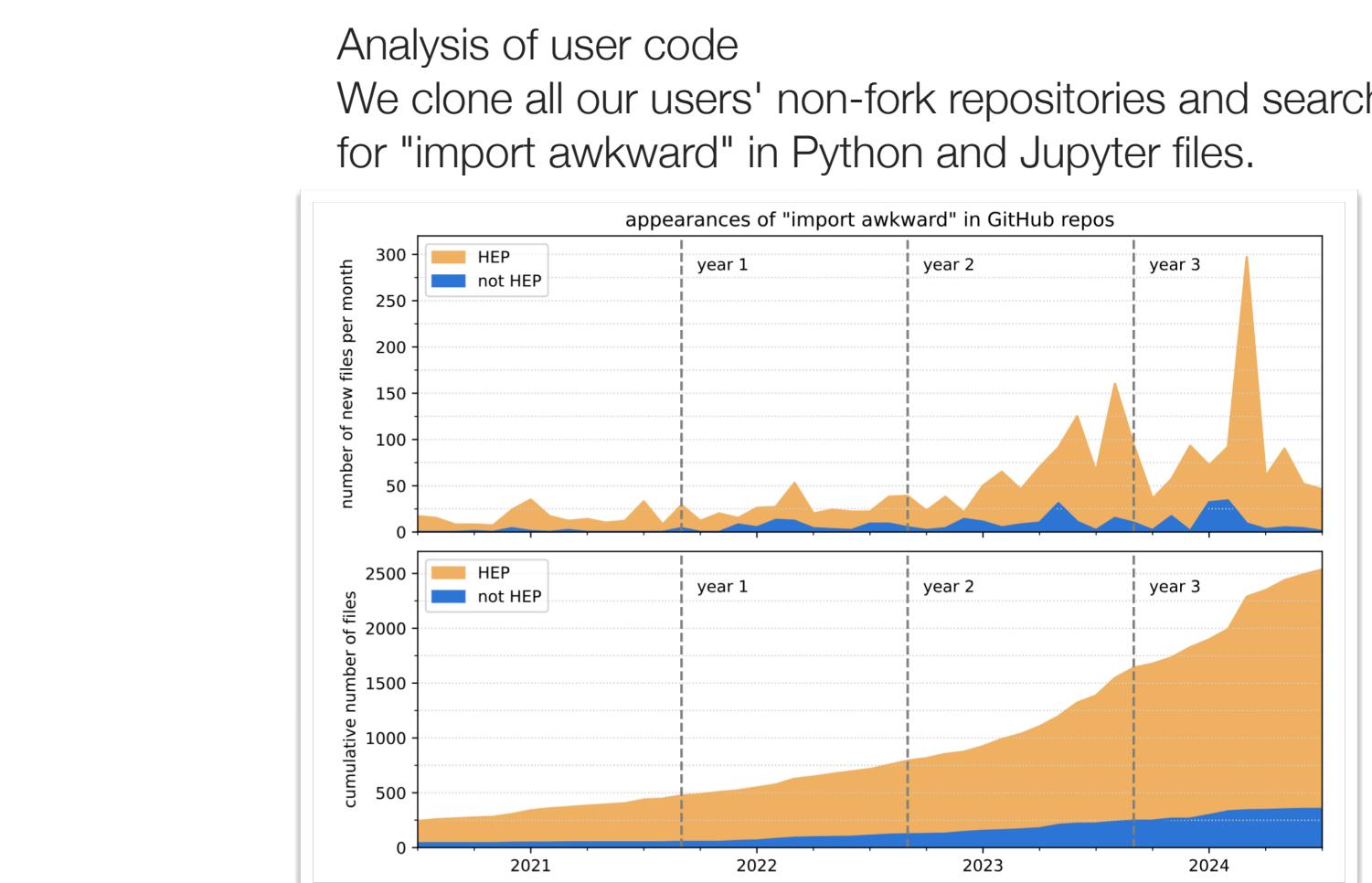


Documentation

The documentation website provides detailed information on various aspects of Awkward Array, including data types, API reference, and contributor guides.

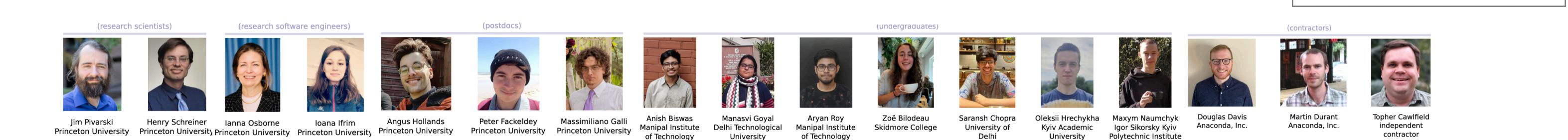
The API reference page details the interface for working with Awkward Array, including functions like ak.sum() and ak.flatten().

This page introduces the basic concepts of Awkward Array, such as dynamic typing and nested lists, with examples and code snippets.



Team

Regular contributors to Awkward Array during the time of this grant, September 2021 – present.



Training, Education and Outreach

The most recent events that include teaching columnar data analysis with Awkward Array in interactive Jupyter notebooks.

Recent events include SciPy 2024, CODAS-HEP, HSF-India HEP Software Workshops, Computational and Data Science Training for High Energy Physics, and USCMS/IRIS-HEP Analysis Software Training 2025.

Self-study Tutorials

Tutorials cover topics like Jagged, ragged, Awkward Arrays, and Scikit-HEP Tutorial, providing step-by-step guides for learning the software.

This section provides a brief introduction to the key concepts of Awkward Array, including its dynamic typing and nested list support.

A table showing the most frequently used functions in Awkward Array, such as ak.sum() and ak.flatten(), along with their calling patterns and descriptions.

Analysis of user code
We clone all our users' non-fork repositories and search for "import awkward" in GitHub repos.

