0.0.1 DataHaskell

Report by:	Marco Zocca
Participants:	Nikita Tchayka, Mahdi Dibaiee, John Vial, Stefan Dresselhaus, and many others
Status:	Ongoing

The DataHaskell community was initiated in September 2016 as a gathering place for scientific computing, machine learning and data science practitioners and Haskell programmers; we observe a growing interest in using functional composition, domain-specific languages and type inference for implementing robust and reusable data processing pipelines.

DataHaskell revolves around a Gitter chatroom [1] and a GitHub organization [2]. The community is slowly but steadily growing; new interested people join the chatroom discussion (which now counts more than 230 unique users) every few days, and 20 or so are active on an average week.

We have a documentation site [3] that serves both as a knowledge base of related Haskell packages and frameworks and to coordinate development, along with a package benchmarking repository [4].

After an informal survey we concluded that large part of our userbase seems to be lacking most

○ an IDE for exploratory data analysis,

○ a generic 'data-frame' for fast import and manipulation of heterogeneous tabular data,

○ a native numerical back-end.

The notebook-IDE situation is improving, thanks for example to an updated 'iHaskell' and the native 'Haskell.do' [5] editor. Dataframes are a more subtle topic, that require domain-specific optimizations, and we are also actively working on that after adopting 'analyze' [6].

Some of us met in person at ZuriHac (Zurich) and ICFP 2017 (Oxford). An informal DataHaskell workshop took place on Saturday 9, 2017 at ICFP, during which a series (7!) of lightning talks showed various applications of the current state of the library ecosystem, e.g. rendering mathematics-heavy code on Hackage, high-performance numerical computing, probabilistic EDSLs and notebook usage for exploratory data analysis.

The workshop was well-received and we are evaluating various options for upcoming meetings, e.g. to be hosted at either functional programming or data science conferences.

We cherish the open and multidisciplinary nature of our community, and welcome all new users and contributions.

Further reading

datahaskell.org

---

[1] https://gitter.im/dataHaskell/Lobby
[2] https://github.com/DataHaskell
[3] http://www.datahaskell.org/docs/
[4] https://github.com/DataHaskell/numeric-libs-benchmarks
[5] http://haskell.do
[6] https://github.com/DataHaskell/analyze