



Intro to Pandas

PyLadies DC

February 21, 2015

Clara Bennett

@csojinb





Tyler Jorgensen

[doesn't use Twitter]

What is Pandas?

- Tool for manipulating and analyzing tabular datasets and timeseries
- Functional-style (like statistical language R)
- Optimized for holistic operations on data

What is Pandas good for?

- Data exploration
- Data munging (cleanup and transformation)
- Extract - Transform - Load



How do you use it?

Tutorial

Integration with other data tools

Pandas data structures based on numpy arrays, so they play well with other Python data libraries, including:

- numpy
- scipy
- scikit-learn
- matplotlib



How do you integrate it?

Case study

Other resources

- 10-minute Tour of Pandas
- Greg Reda on Pandas
- David Rojas Pandas lessons

We're hiring!