Kung-fu Pandas Two: 2 Pandas 2 Furious

Pandas, Rivanna, NAD

Alan Wang

BII-SDAD

June 16, 2023

BIOCOMPLEXITY INSTITUTE

Contents

Format

1. Each team will rock paper scissors, and the winners present first
2. Each team will have 10 minutes to present their findings on their merge, join, or concat. Preferably, how it is used, what issues they ran into, what are the pro's and cons

Creating Python Tools: add.py

```
python add.py -h
python add.py -a 1 -b 2.0
python add.py -a 1 -b random
```

Brainstorming Session

1. What are some operations for computation that you use every day and would be useful as a tool?
2. How do you collaborate with others to improve on a tool?
3. What are ways to write a bad tool?

Some design guides

1. Keep it Simple Stupid (KISS), when possible, write like you will have amnesia in a day
2. Don't Repeat Yourself (DRY) (if you copy-paste, write a function)
3. Test often, test early, test quickly
4. Technical Debt (hint for next week)

What is Rivanna

1. How to access
2. Unix tutorial
3. Custom Jupyter Kernels

Jupyter Notebooks

1. You will often see people prepare Jupyter Notebooks to run asynchronously, jumping up and down cells
2. I think this is BAD and an anti-pattern, a type of technical debt

Geographies

Geographic Entities and Concepts

Exercise

- Each 4 teams are assigned a county. Fork the national address database

Exercise

- Each 4 teams are assigned a county. Fork the national address database
- Each team will need to do a spatial join, and identify which census blocks are not present using 2020 shapefiles

Exercise

- Each 4 teams are assigned a county. Fork the national address database
- Each team will need to do a spatial join, and identify which census blocks are not present using 2020 shapefiles
- Each team will update csv.xz files for all county fips in the region. Each csv will minimally need the following columns: state, county, longitude, latitude, address, and 15 character fips code. Fips can have empty address cells

Exercise

- Each 4 teams are assigned a county. Fork the national address database
- Each team will need to do a spatial join, and identify which census blocks are not present using 2020 shapefiles
- Each team will update csv.xz files for all county fips in the region. Each csv will minimally need the following columns: state, county, longitude, latitude, address, and 15 character fips code. Fips can have empty address cells
- Submit a pull request to the NAD repository by next Friday when we meet