

Pandas 1

Introduction to Pandas

Alan Wang

BII-SDAD

June 9, 2023



BIOCOMPLEXITY INSTITUTE

Contents

- 1 Pandas Background
- 2 How to use Pandas?
- 3 Exercise

What is Pandas

w3schools:

Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

What is Pandas

Alan:

Pandas is a glorified excel sheet

Install pandas

```
# change directory to the project environment  
. .venv/bin/activate # Activate the virtual  
↪ environment  
pip install pandas
```

Open File

```
import pandas as pd
df =
    ↪ pd.read_csv('https://raw.githubusercontent.com/uva-bi
    ↪ dtypes={'fips':object})
# df
```

Filter Data

```
df[df['fips']== '01']
```

Save Data

```
df.to_csv(<filename>, index=False)
# df.to_pickle(<filename>) # saves the object, but
↪ not readable on all computers
```


Exercise

- First, everyone creates a github branch of their own

Exercise

- First, everyone creates a github branch of their own
- Then Add two columns about animals and an interesting fact about yourself. For the animal, enter if you were an animal, which one would you self-identify as?

Exercise

- First, everyone creates a github branch of their own
- Then Add two columns about animals and an interesting fact about yourself. For the animal, enter if you were an animal, which one would you self-identify as?
- Push your data onto your branch as a csv

Exercise

- First, everyone creates a github branch of their own
- Then Add two columns about animals and an interesting fact about yourself. For the animal, enter if you were an animal, which one would you self-identify as?
- Push your data onto your branch as a csv
- As groups of 3 create a team branch and merge your individual csv's to the team branch (Team 1, Team 2, Team 3). Team 1 can only use **merge**, Team 2 can only use **join**, and Team 3 can only use **concat**

Exercise

- First, everyone creates a github branch of their own
- Then Add two columns about animals and an interesting fact about yourself. For the animal, enter if you were an animal, which one would you self-identify as?
- Push your data onto your branch as a csv
- As groups of 3 create a team branch and merge your individual csv's to the team branch (Team 1, Team 2, Team 3). Team 1 can only use **merge**, Team 2 can only use **join**, and Team 3 can only use **concat**
- Next Friday, each team will have a completed branch and present your findings on how the function was used. Save your code in the code/ folder