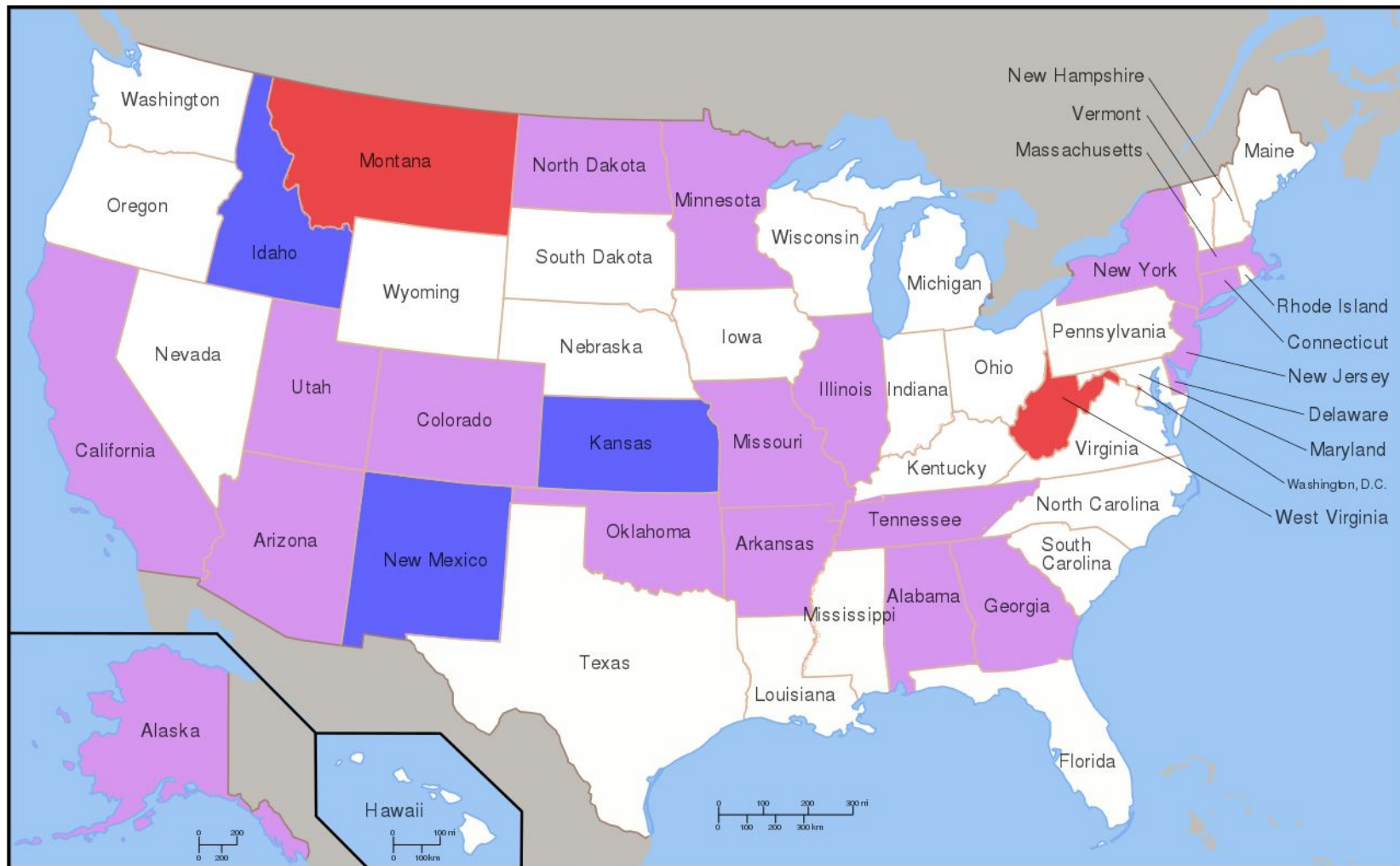


CSCI 285 Scientific Computing



Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)



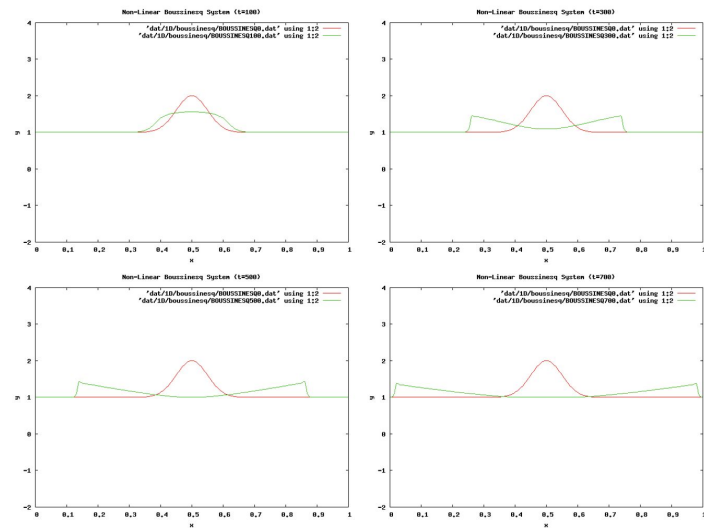
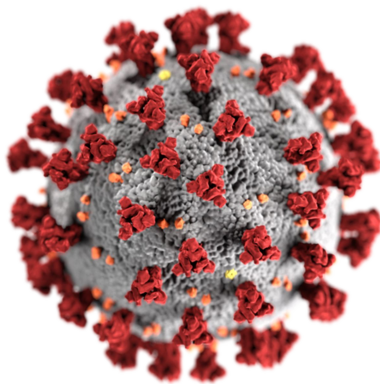
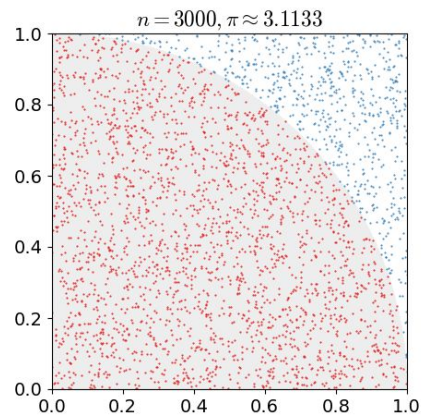
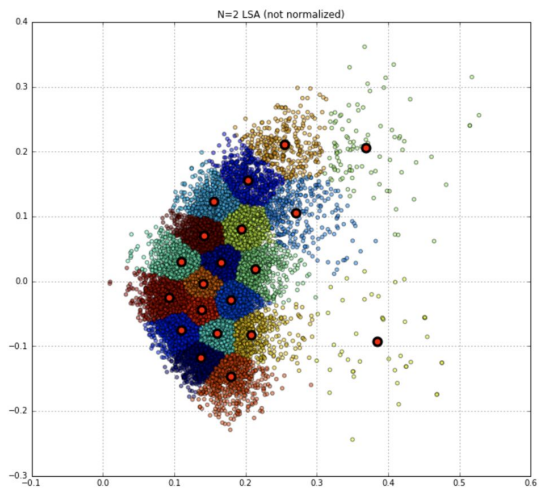


Figure 4: Initial gaussian wave with zero velocity. $\mu = 0.001$, $\varepsilon = 1$, $L = 1$, $N = 256$. $ch\# = 0.1$, $\sigma = 0.05$. Nonlinear shallow water.

Why did you sign up?

CSCI 285 Learning Goals

Module #1: Data Analysis

- Analyze & visualize data sets from a variety of sources.
- Learn several analysis techniques include EDA, clustering, and regression.

Module #2: Modeling

- Model and solve system dynamics problems.
- Construct a Monte-Carlo simulation model.
- Develop agent-based models for complex simulations.

Module #3: Numerical Techniques

- Approximate the roots of continuous functions.
- Understand the strengths and limitations of numerical techniques.

Write idiomatic python and use scientific python libraries.

CSCI 285 Course Overview

<https://hendrix-cs.github.io/csci285/index.html>

Policies

- Attendance
- Check ins / Office Hours (TBD)
- Late Work

Coursework / In-class

- Lecture (36%)
- Labs (27%)
- Exams / Module Review (20%)
- Final Project (17%)

More Info

- Course Calendar / Class Notes / Project Timeline
- W2 Requirement
- Grading scale (TBD)
- Prerequisites: MATH 130 & CSCI 150
- Teams - comms / submitting assignments

Commitments

- Active Participation
- Constructive Feedback
- Academic Integrity
- Learning Accommodation
- Physical & Mental Health

Module #1: Data Analysis

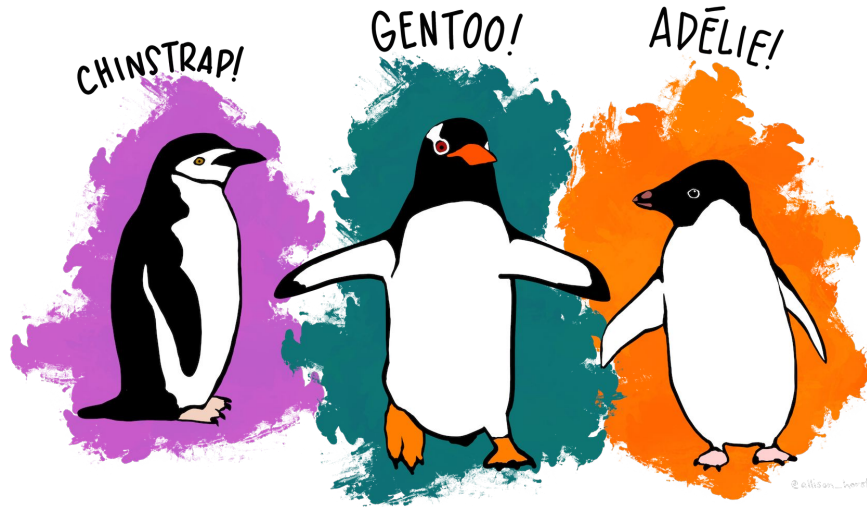
pandas is an open-source python library built for data manipulation and analysis. It is part of the standard library for many teams of data scientists and engineers. pandas introduces new types that have special syntax for data manipulation that are not shared with python's builtin types (e.g. list, dict). Some of the new syntax can look jarring at first, but is *lingua franca* for many data researchers.

Getting Started with pandas

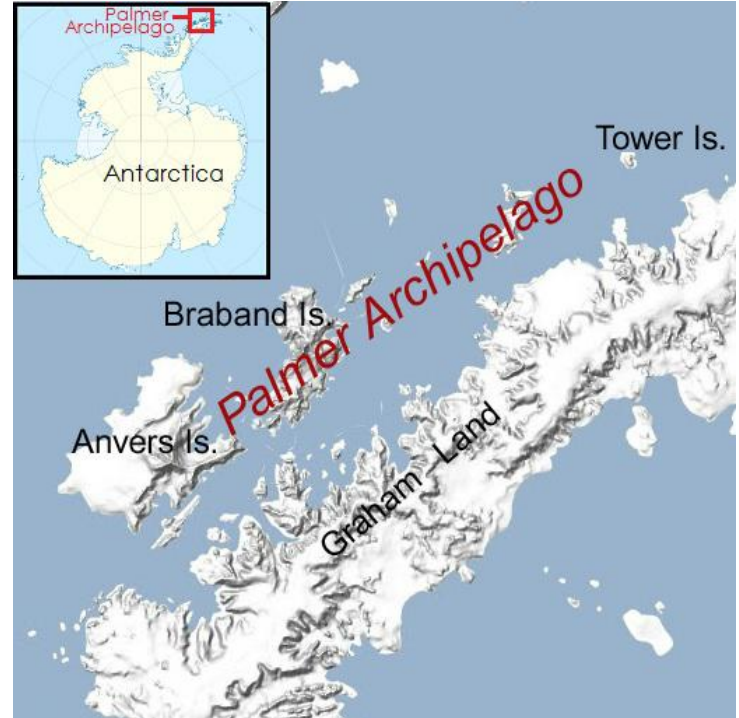
- https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html
- <https://chrisalbon.com/>
- <https://www.datacamp.com/courses/data-manipulation-with-pandas>



Palmer Penguins!



Artwork by @allison_horst".



Getting Started - Development Environment

1. Visit <https://www.anaconda.com/>
2. Download the open source distribution.
3. Follow the Anaconda3 installer instructions.
4. Launch Anaconda-Navigator (Mac, Windows, Linux)
5. Create new environments, launch processes, surf learning resources, etc.

(Alternatively, check out [miniconda](#) if you prefer a more lightweight approach)



Getting Started - pandas

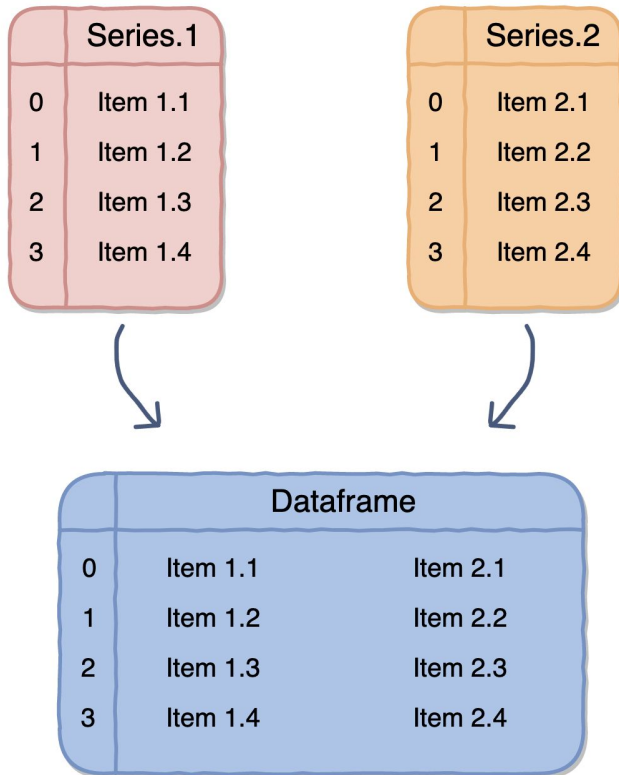
pd.Series



	Series.1
0	Item 1.1
1	Item 1.2
2	Item 1.3
3	Item 1.4

	Series.2
0	Item 2.1
1	Item 2.2
2	Item 2.3
3	Item 2.4

pd.DataFrame



Components of a DataFrame



The Columns, Index, and Data

Columns

Index

	trip_id	usertype	gender	starttime	stoptime	tripduration	from_station_name	latitude_start	longitude_start	dpcapacity_start	to_station_name	latitude
0	7147	Subscriber	Male	2013-06-28 19:01:00	2013-06-28 19:17:00	993	Lake Shore Dr & Monroe St	41.8811	-87.617	11	Michigan Ave & Oak St	41.8811
1	7524	Subscriber	Male	2013-06-28 22:53:00	2013-06-28 23:03:00	623	Clinton St & Washington Blvd	41.8834	-87.6412	31	Wells St & Walton St	41.8834
2	10927	Subscriber	Male	2013-06-30 14:43:00	2013-06-30 15:01:00	1040	Sheffield Ave & Kingsbury St	41.9096	-87.6535	15	Dearborn St & Monroe St	41.9096
3	12907	Subscriber	Male	2013-07-01 10:05:00	2013-07-01 10:16:00	667	Carpenter St & Huron St	41.8946	-87.6534	19	Clark St & Randolph St	41.8946
4	13168	Subscriber	Male	2013-07-01 11:16:00	2013-07-01 11:18:00	130	Damen Ave & Pierce Ave	41.9094	-87.6777	19	Damen Ave & Pierce Ave	41.9094

Data

Description

- Columns - label each column
- Index - label each row
- Data - actual values in DataFrame

Alternative Names

- Columns - column names/labels, column index
- Index - index names/labels, row names/labels
- Data - values

Axis Number

- Columns: 1
- Index: 0

10 minutes to pandas



- https://pandas.pydata.org/pandas-docs/stable/user_guide/10min.html
- <https://chrisalbon.com/>
- <https://www.datacamp.com/courses/data-manipulation-with-pandas>

Trout!