# Intro to Python for Data Analysis

August 12, 2021

Brooke Luetgert, PhD

THE UNIVERSITY OF CHICAGO
Office of Research and National Laboratories
Research Computing Center

# Contact Information

**Brooke Luetgert, PhD.** Computational Scientist at Research and Computing Center (RCC)

Email: **luetgert@uchicago.edu**
Office: TAAC 2, 5607 South Drexel
Telephone: (773)-834-5313

RCC Help Desk: Reg. 216, Mon-Fri 9 AM-5 PM

Materials on GitHub- use search bar, enter **user:luetgert**
Our folder is **luetgert/DataAnalysisCOVID**

# Plan for Today

We will explore Python and several libraries for basic data analysis and visualization.

1. Download Anaconda
2. Look at Jupyter Notebook and Markdown
3. Build some basic Python syntax and introduce libraries
4. Import COVID data from .csv into a data frame with Pandas
5. Plot data with Matplotlib and export

# Why Python?

## Web integration

- Python integrates exceptionally well into web APIs and interactive interfaces

## Resolves 2-language problem

- Python can be used for both research and production, eliminating need for Java and C

## OS independent, large user base, extensive libraries

# Essential Libraries

**NumPy**- tools for fast reading and writing array-based datasets, linear algebra, transformations and random number generator

**pandas**- adds flexible data manipulation capabilities of spreadsheets and relational databases with indexing, reshaping, aggregation, time-series and merging functionalities

**matplotlib**-most popular library for plots and 2D visualizations

**Jupyter**-web-based code notebooks allowing integration of Markdown and HTML for the creation of rich docs with code and text. Also useful for visualization and de-bugging.