

# Programming and Database Fundamentals for Data Scientists

## Introduction

Varun Chandola

School of Engineering and Applied Sciences  
State University of New York at Buffalo  
Buffalo, NY, USA  
chandola@buffalo.edu



# Outline

Preliminaries

Introduction

Need for Programming in Data Science

Why Python?

Why Databases and SQL?

# Before we begin

- ▶ Class webpage -  
<https://cse.buffalo.edu/~chandola/eas503ubcourse/>
- ▶ UBLearns - Automatically enrolled
- ▶ Local setup
  - ▶ Python 3 (Anaconda python)
  - ▶ MySql Database
  - ▶ Python MySql bindings (pymysql)
  - ▶ iPython/jupyter Notebooks

# Starting from the beginning

- ▶ What is a *computer*?

# Starting from the beginning

- ▶ What is a *computer*?
  - ▶ Machine that **stores** and **manipulates** **information** under the control of a **changeable program**.
- ▶ What is a *computer program*?

# Starting from the beginning

- ▶ What is a *computer*?
  - ▶ Machine that **stores** and **manipulates information** under the control of a **changeable program**.
- ▶ What is a *computer program*?
  - ▶ A **detailed, step-by-step** set of instructions telling the computer **exactly** what to do.

# What is Computer Science?

- ▶ Computer Science  $\neq$  Study of Computers!!!

# What is Computer Science?

- ▶ Computer Science  $\neq$  Study of Computers!!!
- ▶ **How can we make computers to do what we want.**
- ▶ **What can be computed?**



# The Holy Trinity of Computer Science

## Design

- ▶ Write algorithms to solve problems

## Analyze

- ▶ Is a problem *solvable* (efficiently)?

## Experiment

- ▶ Implement algorithms and assess performance

# Why Programming in Data Science?

## Data Science Pillars

- ▶ Math
  - ▶ Statistics
  - ▶ **Programming**
- 
- ▶ Store and access data
  - ▶ Manipulate data
  - ▶ Get insights from data
  - ▶ Report/visualize results

# Why Python?

- ▶ “Good enough” for data science
- ▶ Quick and intuitive coding
- ▶ Thriving ecosystem - abundance of libraries (e.g., Pandas)
- ▶ Huge adoption among data scientists



# Why Databases and SQL?

- ▶ Relational databases – that is where most data lies.
- ▶ SQL - language to communicate with the database

# References