Discussion + Python Intro

INFO 370

Learning Objectives

Critique academic articles -- the first reading assignment -- (first half of class)

Discuss programming for data science, and how we'll do it in this class

Learn the basics of **using Python** in a Jupyter Notebook

Start Notebook Set 1 (due Friday night)

Discussion

Discussion Structure / Expectations

Structure

- Break up into 8 discussion groups by topic/paper
- Spend ~15 minutes **discussing** issues in your group, **drawing** on the board
- Spend ~5 minutes interacting with the class: focus on what you find interesting

Expectations

- Participate
- Give others your attention
- Remember: this will be as interactive and interesting as you make it

Programming for Data Science

Do you need to write code to ***do*** data science?

Programming for Data Science

Allows you to complete a project in a single environment

Enables you to connect analysis environments

Not restricted by tool limitations

Writing Code to Work with Data

Why

Customizable

Repeatable

Transparent

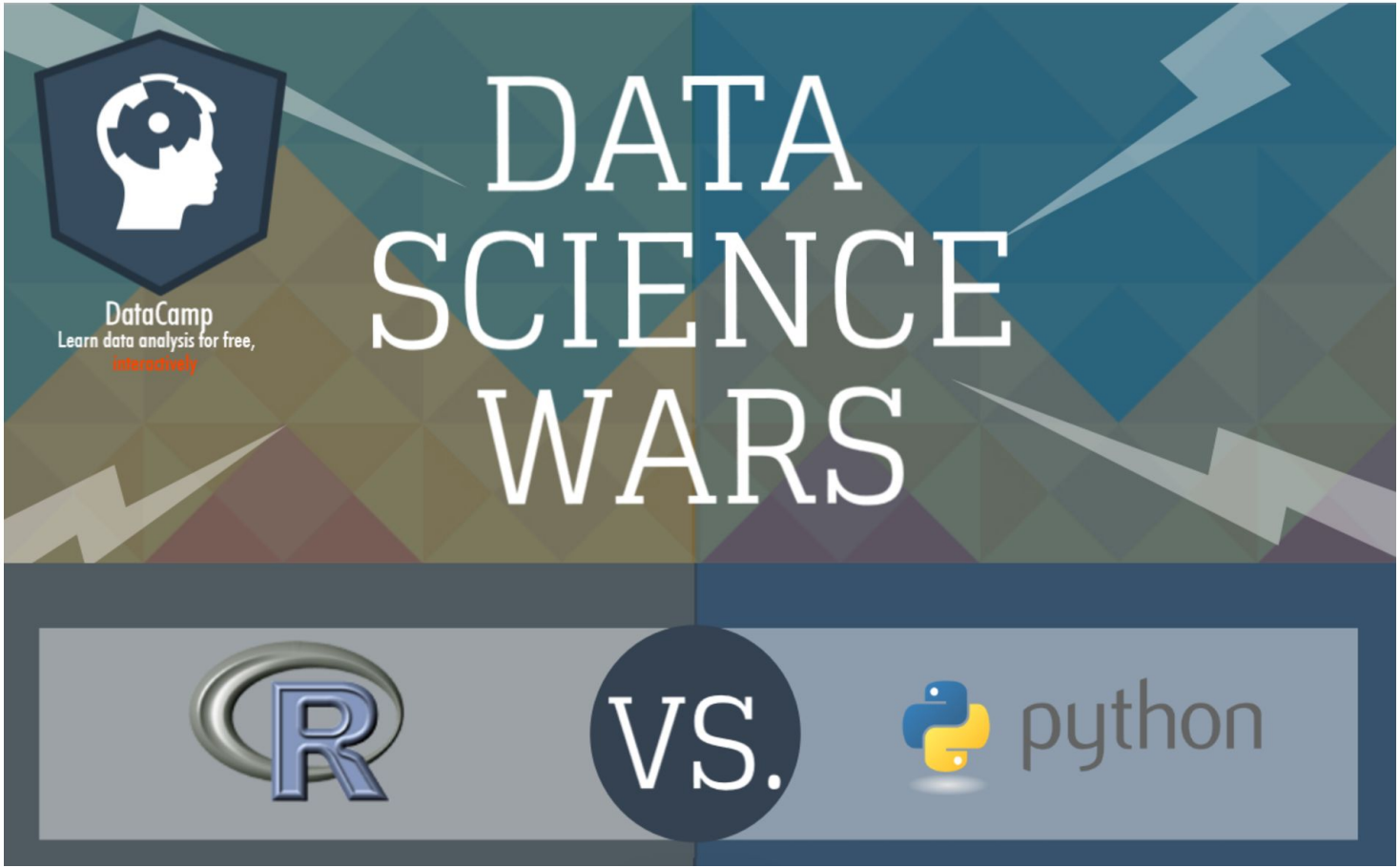
Scalable

Why Not

Time consuming

Error prone

Sometimes less clear



R v.s. Python (link)

Introduction to Python

Overview

"An interpreted high-level programming language for general-purpose programming"

Can implement object-oriented or functional approaches

Variety of applications in software engineering and data science

In This Course

Aim is a practical understanding of Python:

- Understand basics of the language (syntax)
- Write clear, well formatted code
- Able to identify, interpret, and implement relevant approaches

We'll use Python 3

Work in **Jupyter Notebooks**

Getting Started

In this exercise, you'll get familiar with the following basics of python:

- Creating variables (lists, dictionaries)
- Writing functions
- List compressions

Basic variables

Variables in python are loosely typed, so you can quickly create numeric, character, or boolean values. Variable names should be expressive, and be written as lower-case words that are underscore separated (i.e., `variable_name`)

In [1]:
```
# Create a numeric variable `hours_in_a_year` equal to the number of hours in a year
```

In [2]:
```
# Create a numeric variable `seconds_in_a_day` equal to the number of seconds in a day
```

In [3]:
```
# Create a boolean variable `more_seconds_than_hours` that is true
# if there are more seconds in a day than hours in a year
```

notebook-set-1

Upcoming...

Notebook Set 1 due ***Friday night***

Assignment 1 due before class ***next Tuesday***

Thursday: web-scraping in Python

FYI: I will be offline for the holiday weekend beginning Friday morning