DS-GA 1007 Programming for Data Science

Lecture 2

Agenda

- Review
- Lesson
- Demo

Reminders

- Materials
  - Linked to Week in NYU Classes
- Assignments
  - Lab 1 and Lab 2 collected week 3
  - Homework 1
    - Available under Resources in NYU Classes
    - Upload notebook to Gradescope

Review

Review

- Set-Up
  - Launch in Browser
  - Files and Running Files

Review

- Set-Up
  - Launch in Browser
  - Files and Running Files
- Cells
  - Type
  - Mode

Review

- Set-Up
  - Launch in Browser
  - Files and Running Files
- Cells
  - Type
  - Mode
- Getting Help

Review

- Set-Up
  - Launch in Browser
  - Files and Running Files
- Cells
  - Type
  - Mode
- Getting Help
- Exporting

Review

- Section 1
  - Jupyter and JupyterHub
    - Cells

Review

- Section 1
  - Jupyter and JupyterHub
    - Cells
  - Discussion 1
    - Variables
    - Types
    - Accessing data in Row/Column format

Review

- Section 1
  - Jupyter and JupyterHub
    - Cells
  - Discussion 1
    - Variables
    - Types
    - Accessing data in Row/Column format
  - Lab 1
    - Import Package
    - Arranging data in Row/Column format
    - Properties

Review

- How can we process tabular data files?

Review

- How can we process tabular data files?
- How can we do the same operations on many different values?

Review

- How can we process tabular data files?
- How can we do the same operations on many different values?
- How can we do different operations based on data values?

Lesson

- Code
  - Instructions for computer
- Text/Code
  - Instructions for developers
- Text
  - Instructions for users

Lesson

- Text/Code
  - Help Organize the Coding

Lesson

- Text/Code
  - Help Organize the Coding
  - Code Review Easier

Lesson

- Text/Code
  - Help Organize the Coding
  - Code Review Easier
  - Iterative Approach to Coding

Lesson

- Text/Code
  - Help Organize the Coding
  - Code Review Easier
  - Iterative Approach to Coding
  - Helps with Revisions

Lesson

- Text/Code
  - Help Organize the Coding
  - Code Review Easier
  - Iterative Approach to Coding
  - Helps with Revisions
  - Shorter Text

Lesson

```
Get number of quizzes as a parameter

1.          Initialize "sum" and "count" variables to 0
2.          while count < number of quizzes
             2.1        get quiz grade
             2.2        add quiz grade to "sum"
             2.3        increment count
3.          compute average of sum over number of quizzes
4.          return average
```

Lesson

```
Get a list of quiz grades as a parameter


1.          Initialize "sum" variable to 0
2.          Go through each quiz grade in the list
             2.1        add quiz grade to "sum"
3.          compute average of sum over number of quizzes
4.          return average
```

Lesson

```
#  Get a list of quiz grades as a parameter

def compute_quiz_average(quiz_grade_list):
    # Initialize "sum" variable to 0
    sum = 0
    # Go through each quiz grade in the list
    for qgrade in quiz_grade_list:
        # add quiz grade to "sum"
        sum += qgrade
    # compute average of sum over number of quizzes
    num_quizzes = len(quiz_grade_list)
    average = float(sum)/num_quizzes
    # return average
    return average
```

Demo

- How can I store many values together?

```
files = ['large-file-01.csv',
         'myscript.py',
         'large-file-02.csv',
         'small-01.csv',
         'small-02.csv']
```

- Write a script to split into groups

```
large_files = ['large-file-01.csv', 'large-file-02.csv']
small_files = ['small-01.csv', 'small-02.csv']
other_files = ['myscript.py']
```