CS 105 TA Lecture Notes
October 13, 2014

Goals for the rest of the semester

- How do we process data using Excel? What are the limitations?
- Then, we'll go back to JavaScript for some deeper analysis
- Finally, how can we combine Excel and JavaScript?

Data and File Extensions

All data is made up of 0s and 1s.

- Numbers, letters (ASCII), pixels (red, green, blue), images

File extensions, which are the part of the file that appears *after the dot,* tell us how to interpret the 0s and 1s. They give us rules by which we read 0s and 1s from the disk and process that data.

- .js file: JavaScript code
- .html file: HyperText Markup Language
- .xlsx file: Excel
- .xml file: eXtensible Markup Language
- .csv file: Comma Separated Values
- .fasta: DNA Data
- .png, .gif, or .jpeg file: image

Some file types do not make sense outside of their intended context.

- For example, try opening an Excel (.xlsx) file in Notepad++ -- it's gibberish!
- .xlsx is a binary format – it contains a markup that contains an encoding that can only be read by the Excel program

---

APIs and Data Formats

How do we get the data out? We'll use APIs – Application Programming Interfaces.

- Allows a programmer (like you) to access functionality or data from someone else.
- For example, Google Graphs API is an API created by Google that creates bar graphs and charts from given data.
- Another example is d3: we can create really detailed infographs using the d3 API.
- We will be using data-based APIs, which give us data in different formats.

There are a few different data formats.

- XML: very similar to HTML
  - consists of three types of tags (start, end, and empty-element)
  - all start tags must have an end tag
  - example: Yahoo! Currency API gives us currency conversions in an XML format
- CSV: Comma Separated Values
  - related values are listed on one line, separated by commas

    - can be opened by a plaintext editor and is human-readable
    - Excel will read CSV files directly
    - example: Yahoo! Finance API gives us stock prices of specific stock symbols
- Application specific data
    - many applications use their own data formats, such as FASTA file format for DNA Data

---

Microsoft Excel: a spreadsheet application

Spreadsheets provide users with a 2-dimensional layout of **cells** that are organized into **rows** and **columns.**

- Rows: labeled numerically (1, 2, 3, …)
- Columns: labeled alphabetically (A, B, C, …)

Excel is one of many spreadsheet tools. Others include Google Docs, OpenOffice, and Numbers for Mac. However, Excel is the *most widely used* spreadsheet tool.

The **only** format accepted for this class will be Excel spreadsheets (.xlsx).

Each of the cells in a spreadsheet is a variable. We can perform computations and functions with the variables in a spreadsheet.

- Row 2, column C is variable `C2`.
- In cell D2, we can type `=C2 + C3`, which means cell/variable D2 gets the value of variable `C2` plus the value of variable `C3`.
- In cell E2, we can type `=D2 – C2`, which means cell/variable `E2` gets the value of variable `D2` minus the value of variable `C2`.
- We can find the sum of values of cells `C2` through `C6` by using the function `=SUM(C2, C3, C4, C5, C6)`.
- We can find the average of values of cells `C2` through `C6` by using the function `=AVERAGE(C2, C3, C4, C5, C6)`.

Data Types and Excel Formulas

Excel has five basic data types.

- Numbers
- Strings: don't need quotes in Excel strings – Excel recognizes any non-number as a string
- Dates: can ask for one day after a specific date; is treated differently from a regular number
- Currencies: $ processed as a US dollar
- Booleans: true or false

If the first character in your input is an equals sign (=), **the value of the cell is read as a program**.

- Can reference other cells' value by entering their cell column and row identification (e.g., `=C3`)
- `=SUM()` can take in a series of numbers or a range of numbers.

- `=SUM(C2, C3, C4, C5, C6)`
- `=SUM(C2:C6)` takes sum over the first cell in our range (`C2`) to the last cell in our range (`C6`)