Penny Math

- Penny Math is a simple formula
  - A (or a) costs 1 penny
  - B (or b) costs 2 pennies
  - …
  - Z (or z) costs 26 pennies
  - Everything else is FREE
- Thus
  - “Sergey” costs 19+5+18+7+5+25=79 cents

Dollar Words

- “Dollar words” are words that cost EXACTLY $1 (100 cents)
- Who wants to search around dreaming up words and testing their cost???
- But what if I had a dictionary file?

What is a File?

- A file is a collection of data that is stored on secondary storage like a disk or a thumb drive.
- Accessing a file means establishing a connection between the file and the program and moving data between the two.

Two Types of Files

- Files come in two general types:
  - Text files: files where control characters such as “\n” are translated. These are generally human readable
  - Binary files: all the information is taken directly without translation. Not readable and contains non-readable info.

File Objects or Stream

- When opening a file, you create a file object or file stream that is a connection between the file information on disk and the program.
- The stream contains a “buffer” of the information from the file, and provides the information to the program

Buffering

- Reading from a disk is very slow. Thus the computer will read a lot of data from a file in the hope that, if you need the data in the future, it will be “buffered” in the file object.
- This means that the file object contains a copy of information from the file called a cache (pronounced “cash”).

Making a File Object

fileObject = open(“myFile.txt”, “r”)

- myFile is the file object. It contains the buffer of information. The open function creates the connection between the disk file and the file object. The first quoted string is the file name on disk, the second is the mode to open it (here,“r” means to read).

Where is the Disk File?

- When opened, the name of the file can come in one of two forms:
  - “file.txt” assumes the file name is file.txt, and it is located in the current program directory.
  - “c:\bill\file.txt” is the fully qualified file name and includes the directory information.

Different Modes

- “r” is to read as a text file.
- “w” is to write as a text file. Wipes the contents of the file if there is any, creates file otherwise.
- “a” is append, adds to the end of an existing file.
- “b” is a modifier, indicating a binary file. No character translation is done.
- “+” is a modifier, indicating both read and write. With “r”, file must exist. With “w”, makes or truncates the file, with “a” appended to the file.

Be Careful with Write Modes

- Be careful if you open a file with the ‘w’ mode. It sets an existing file’s contents to be empty, destroying any existing data.
- The ‘a’ mode is nicer, allowing you to write to the end of an existing file without changing the existing contents.

Text Files use Strings

- If you are interacting with text files (which is all we will do for this semester), remember that everything is a string:
  - everything read is a string
  - if you write to a file, you can only write a string

Getting File Contents

- Once you have a file object:
- fileObject.read()
  - Reads the entire contents of the file as a string and returns it. It can take an optional argument integer to limit the read to N bytes, that is fileObject.read(N).
- fileObject.readline()
  - Delivers the next line as a string.

More File Contents

- fileObject.readLines()
  - Returns a single list of all the lines from the file.
- for line in fileObject:
  - Iterator to go through the lines of a file.

Closing

- When done, you close the file. Closing is important because the information in the fileObject buffer is “flushed” out of the buffer and into the file on disk, making sure that no information is lost.

```
fileObject.close()
```

Writing

- Once opened, you can write to a file (if the mode is appropriate):
- fileObject.write(s)
  - writes the string s to the file
- fileObject.writelines(list)
  - write a list of strings (one at a time) to the file

Back to Penny Math

- Let's look at how I can use a dictionary file to compute dollar words…

Making and using a File Object

- fileObject = open(“myFile.txt”, “r”)
  - Opens the file “myFile.txt” in read-only mode
- fileObject.read()
  - Reads the entire contents of the file as a string and returns it.
- fileObject.readline()
  - Delivers the next line as a string.

Making and using a File Object

- fileObject.readLines()
  - Returns a single list of all the lines from the file.
- for line in fileObject:
  - Iterator to go through the lines of a file.
- fileObject.write(s)
  - writes the string s to the file
- fileObject.writelines(list)
  - write a list of strings (one at a time) to the file

Making and using a File Object

- fileObject.close()
  - When done, you close the file.

Spreadsheets

- The spreadsheet is a very popular, and powerful, application for manipulating data.
- Its popularity means there are many companies that provide their own version of the spreadsheet.
- It would be nice if those different versions could share their data…

CSV, Basic Sharing

- A basic approach to share data is the comma separated value (CSV) format:
  - it is a text format, accessible to all apps
  - each line (even if blank) is a row
  - in each row, each value is separated from the others by a comma (even if it is blank)

Reading CSV data

```
for row in file:
    varA, varB, varC = row.split(",")
```

- Split will split up the line by the argument (in this example, it is a comma
- Split will save the first string value into the first variable on the left side of the comma
  - E.g. first string gets saved in variable “last”, second gets saved in variable “height”
- Each line must have the exact same number of variables