Lecture 11 - Input Output

Data for Homework 5

Information for homework 5 is in this and the next lecture. You will need to download:

Data for Genome

YouTube

Input Output - Files - https://youtu.be/dL2oKyi5riw
Input from the terminal - https://youtu.be/gZUsJZYcCXc

From Amazon S3 - for download (same as youtube videos)

Input Output - Files
Input from the terminal

First: The Problem

Lots of times the process is much simpler to do than you might expect. For example with Cystic Fibrosis (CF), You have a single letter mutation that causes the disease.

Overview of how genetics works.

Video: Stephie.

One of the first Machine Leering AI Recommended compounds.

Drug under the brand: Trikafta.

Now the problem is that if you don't have the mutation and you take the drug it kills you. If you do have the mutation and you take it it cures you.

The Question --- Dose Stephie have the mutation.

We are looking for

```
ACTGGCTGGCCTAGTGACTTCCAGCTGCACAGCTATCGACCCAGGGCTGGACAGCCCCTGCCTGGC
```

v.s. a healthy person:

```
ACTGGCTGGCCTAGTGACTTCCAGCTGCACAGCTATCGTCCCAGGGCTGGACAGCCCCTGCCTGGC
```

2nd Video: of Stephie.

Unix and DOS File System

It's important to know that if you are using Windows - this applies but Windows is going away. It has become Windows/Linux and to get games to run on Linux there is now a Windows subsystem written by Microsoft that runs on Linux. Microsoft Azure runs 3x more Linux that Windows and that ratio is increasing in favor of Linux

Input Output

1. Open a file
2. Read in lines one at a time
3. Process each line

A little more detail:

- Read A File one line at a time
  - Set some values like the line number
  - while not at the end of the file
    - if we find a string
      - report it
      - exit loop
- if we never found the string
  - report that we never found it