CS 133 - Introduction to Computational and Data Science

Instructor: Renzhi Cao
Computer Science Department
Pacific Lutheran University
Fall 2017

Previous class

- I put slides on course website: cs.plu.edu/133
- Apply CS Account
- Finish the survey

Review - Problem-Solving

A. Understand the Problem
- Do you understand all the words & terms that are being used?
- What are you being asked to find or show?
- Is there enough information to solve the problem?
- Can you draw a picture that might help?

B. Come Up With a Plan
- Guess and check, make a list, or draw a picture.
- Look for a pattern, or find a key equation.
- Try solving a simplified version of the problem.
- Work backwards.

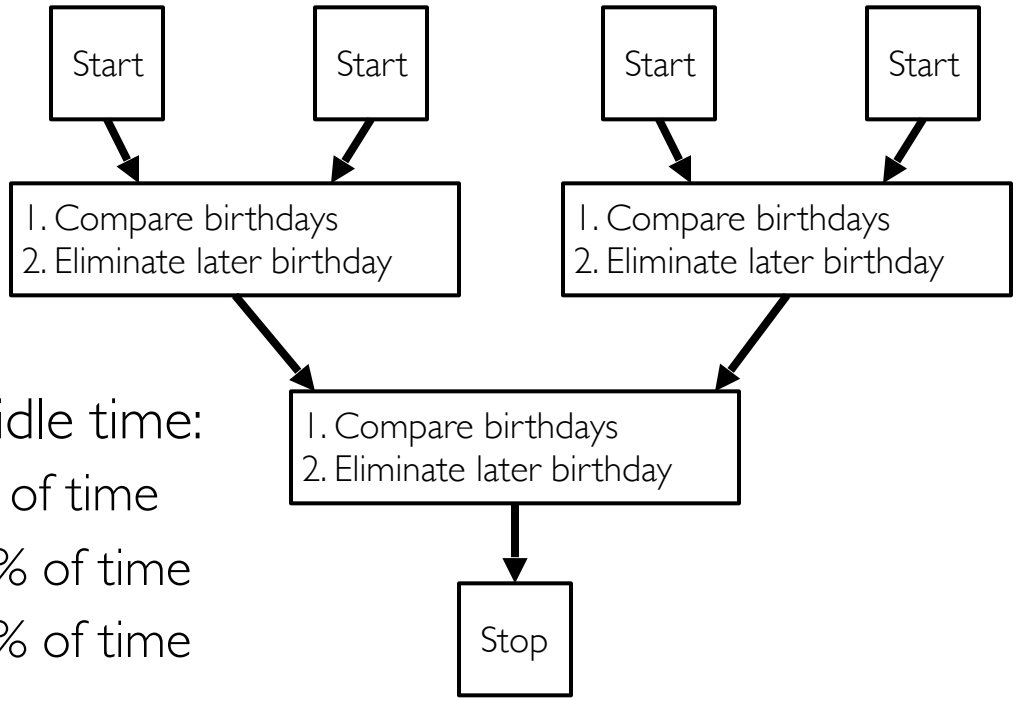
C. Carry Out the Plan
- Be aware that you may run into roadblocks or dead-ends!
- Check to see if your results make sense.
- Don't be afraid to start over!

D. Make Your Solution Computer-Friendly
- Imagine you are writing to a student not in this class.
- Keep things brief… but make sure that you don't leave anything out.
- Write a step-by-step list of instructions… like writing a recipe.

Review - Problem solving

Finding the earliest birthday - method 2

- Simultaneous events mean fewer steps:
  - 4 people – 2 steps
  - 16 people – 4 steps
  - 32 people – 5 steps



- Fewer steps mean less idle time:
  - 4 people – idle ≤ 50% of time
  - 16 people – idle ≤ 75% of time
  - 32 people – idle ≤ 80% of time

Conclusion #1: Computers can't see the "big picture" – only the immediate task at hand.
Conclusion #2: Not all programs are equal – some are faster or more flexible than others.

Review - Problem-Solving

Some Practice Questions

Here are a few problems to think about. Use the strategies from the previous slide, and write down at least three facts or observations that you think are important when it comes to solving the problem. We'll discuss the pros and cons of each fact/observation before trying to solve the problems.

1. Same birthday. You and your classmates want to know if there are students sharing the same birthday. You have everyone's birthday date (Month and Day), how do you quickly find it out?

2. Pizza Prices. You're trying to decide what size pizza to order, and have the choice of a 12" pizza for $13 or a 14" pizza for $16. Which one gives you the most pizza per dollar?

3. Finding the Day of the Week. What day of the week is 23 December 2017? What about 23 December 2087?

Review - Problem-Solving

Video related to numbers

http://www.ted.com/talks/arthur_benjamin_does_mathemagic#t-898833

Data science

What comes to mind when I say the word “DATA”?

Data presence in our daily life

- Websites track user's clicks
- Smart phones are tracking your location, searches, patterns
- Smart watches
- Smart cars
- Amazon collects purchase habits
- Databases
- Government
- Sports

What can we do with all of this data?

What is Data Science?

Book defines a data scientist as: “Data scientist is someone who knows more statistics than a computer scientist and more computer science than a statistician”

Better definition for data scientist: individual that extracts insights from unorganized data.

Facebook: https://www.facebook.com/notes/facebook-data-science/nfl-fans-on-facebook/10151298370823859

Target: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?_r=0

Government: http://www.marketplace.org/2014/08/22/tech/beyond-ad-clicks-using-big-data-social-good

First problem with data

- You know the salaries of 10 people and the number of years that they have worked for the company. What can we learn from this data?

Salary	Years of Experience
83000	8.7
88000	8.1
48000	0.7
76000	6
69000	6.5
76000	7.5
60000	2.5
83000	10
48000	1.9
63000	4.2

Second Problem

Assume a list of users:

ID	Name
1	Hero
2	Dunn
3	Sue
4	Chi
5	Thor
6	Clive
7	Hicks
8	Devin
9	Kate
10	Klein

Problem cont…

- Assume a list of users:

ID	Name
1	Hero
2	Dunn
3	Sue
4	Chi
5	Thor
6	Clive
7	Hicks
8	Devin
9	Kate
10	Klein

- We know something about their friendships

Friendships
Hero-Dunn
Hero-Sue
Dunn-Sue
Dunn-Chi
Sue- Chi
Chi – Thor
Thor – Clive
Clive – Hicks
Clive – Devin
Hicks – Kate
Devin – Klein
Kate - Klein

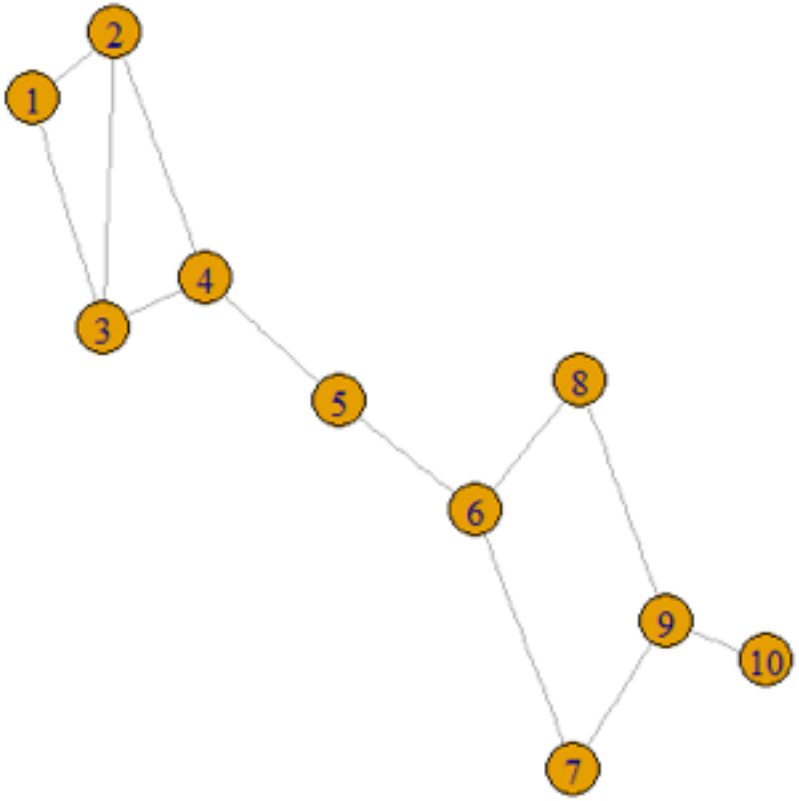
Problem cont…

- Assume a list of users:

ID	Name
1	Hero
2	Dunn
3	Sue
4	Chi
5	Thor
6	Clive
7	Hicks
8	Devin
9	Kate
10	Klein

- Hard to read. Let's fix it

Friendships
1 - 2
1 - 3
2 - 3
2 - 4
3 - 4
4 - 5
5 - 6
6 - 7
6 - 8
7 - 9
8 - 9
9 - 10


1
2
3
4
5
6
7
8
9
10

Let's analyze our graph

- What can we learn by looking at it?
  - What is the average number of friends per person?
  - Who is the most popular person?
  - Who is the most important person in the network?

A little taste of R

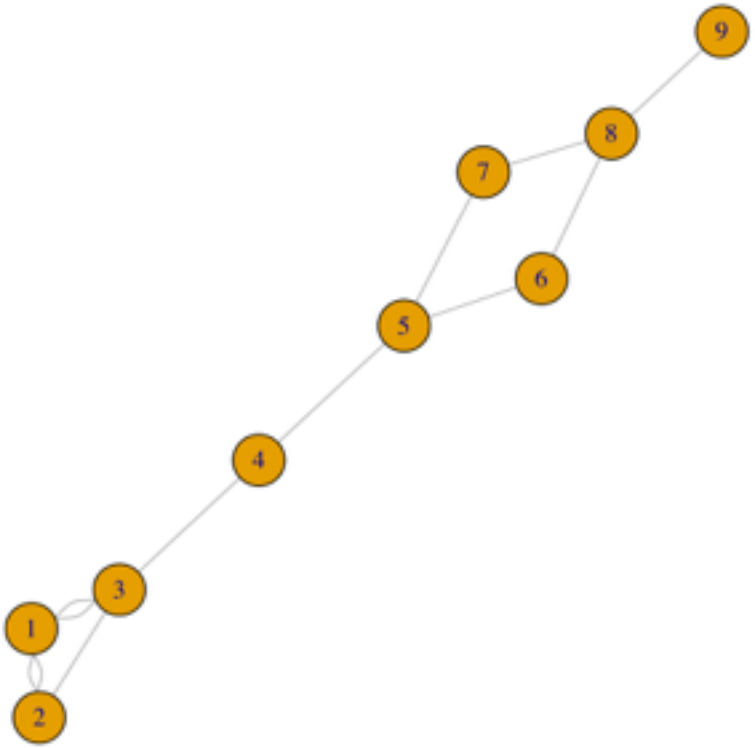
We will cover R in the future in much more detail, but this is a taste of the things you can do.

Open R "as administrator"

```
> install.packages("igraph")
> library(igraph)
> graph.non <- graph(c(1,2, 1,3, 1,2, 1,3, 2,3, 3,4, 4,5, 5,6, 5,7, 6,8, 7,8, 8,9),directed=FALSE)
➢ plot(graph.non)
➢ tkplot(graph.non,layout=layout.kamada.kawai)
```

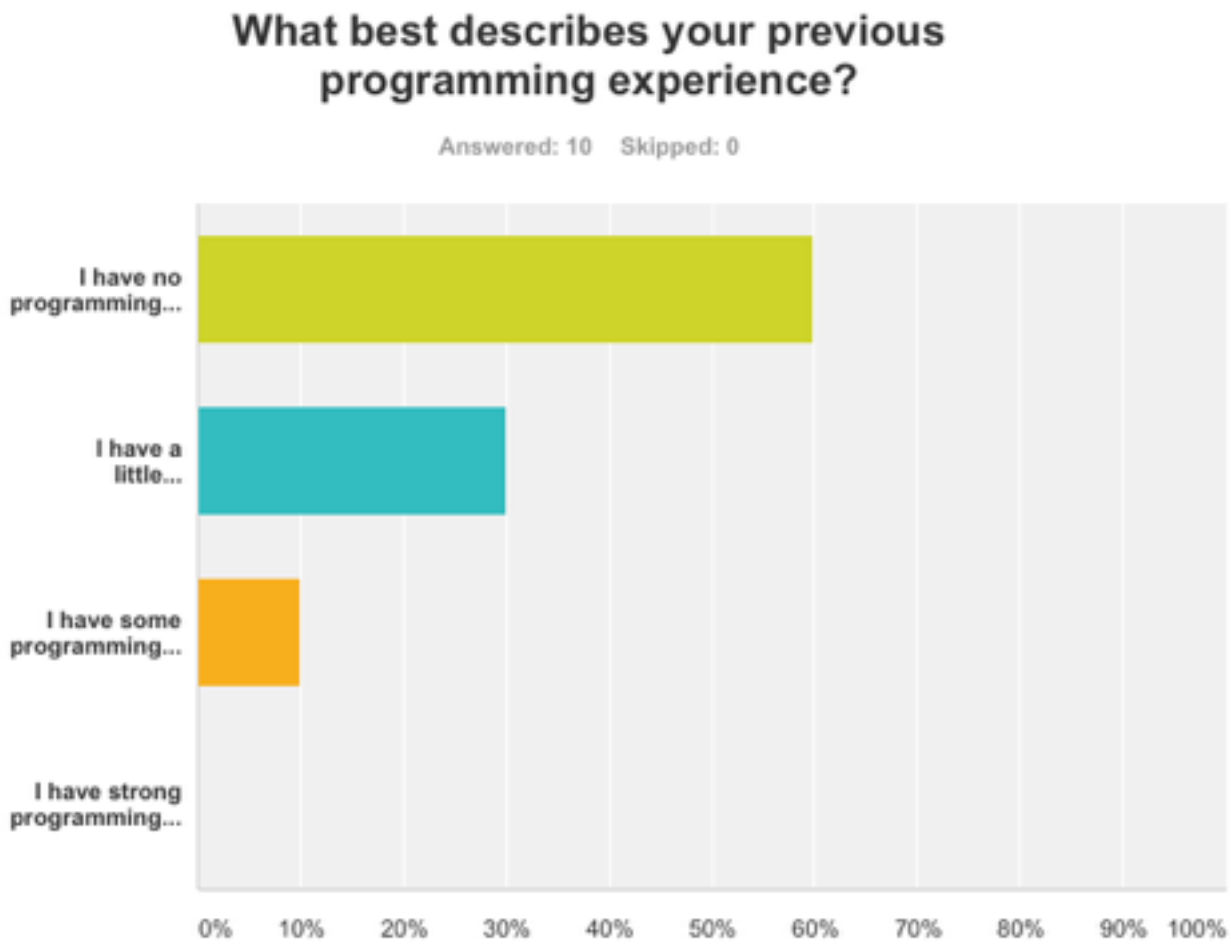
Disclaimer: Don't worry if this looks too complex. It will all make sense at the end of the semester!

A little taste of R



Data presence in our daily life

Let's start for the programming part

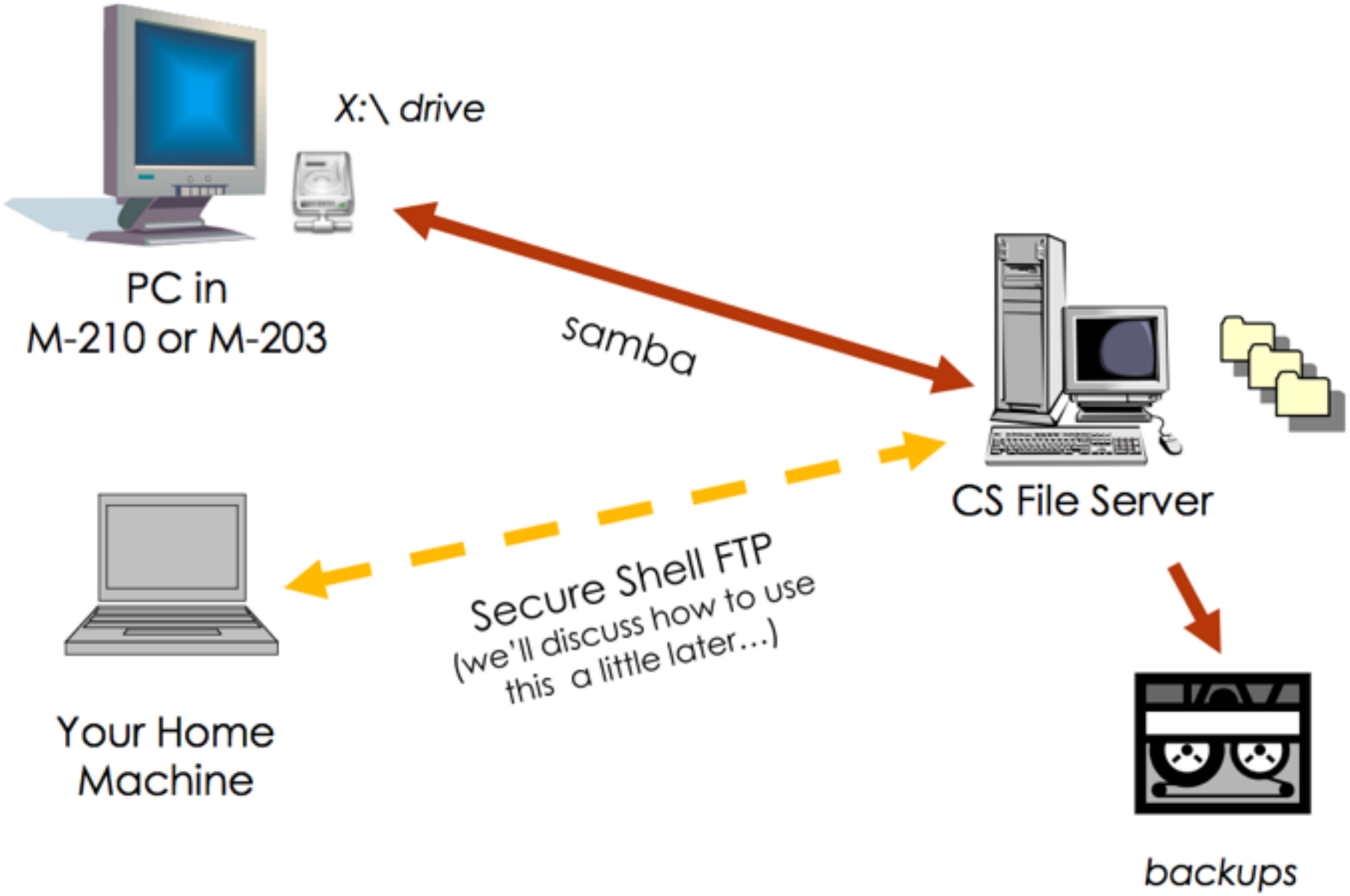


Data presence in our daily life

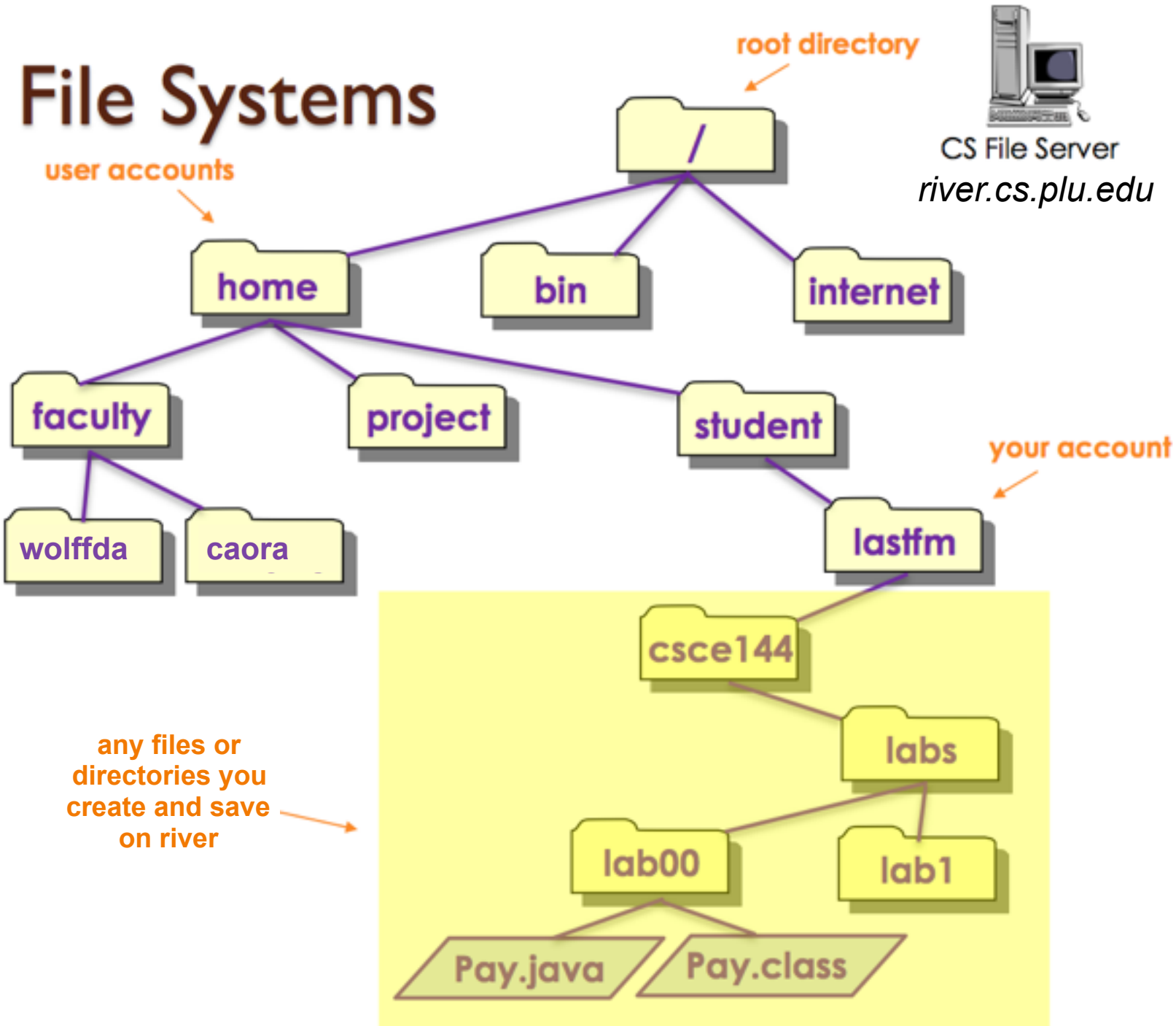
We are going to learn today:

1. **Navigate drives and directories from both Graphical interface and command prompt**
2. **Understanding File Systems and department file server**
3. **Practice using Atom editor**
4. **Write your first Python code!**

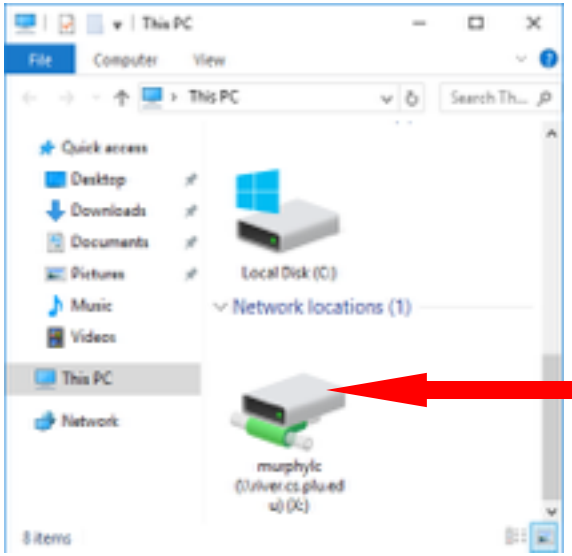
Navigating Drives & Directories…



File Systems



When you logon to the CSCI lab machines in Morken 203 or 210 using your* epass *and password the PC's "X" drive is automatically mapped to your* river *account



your account
on river

`userid`

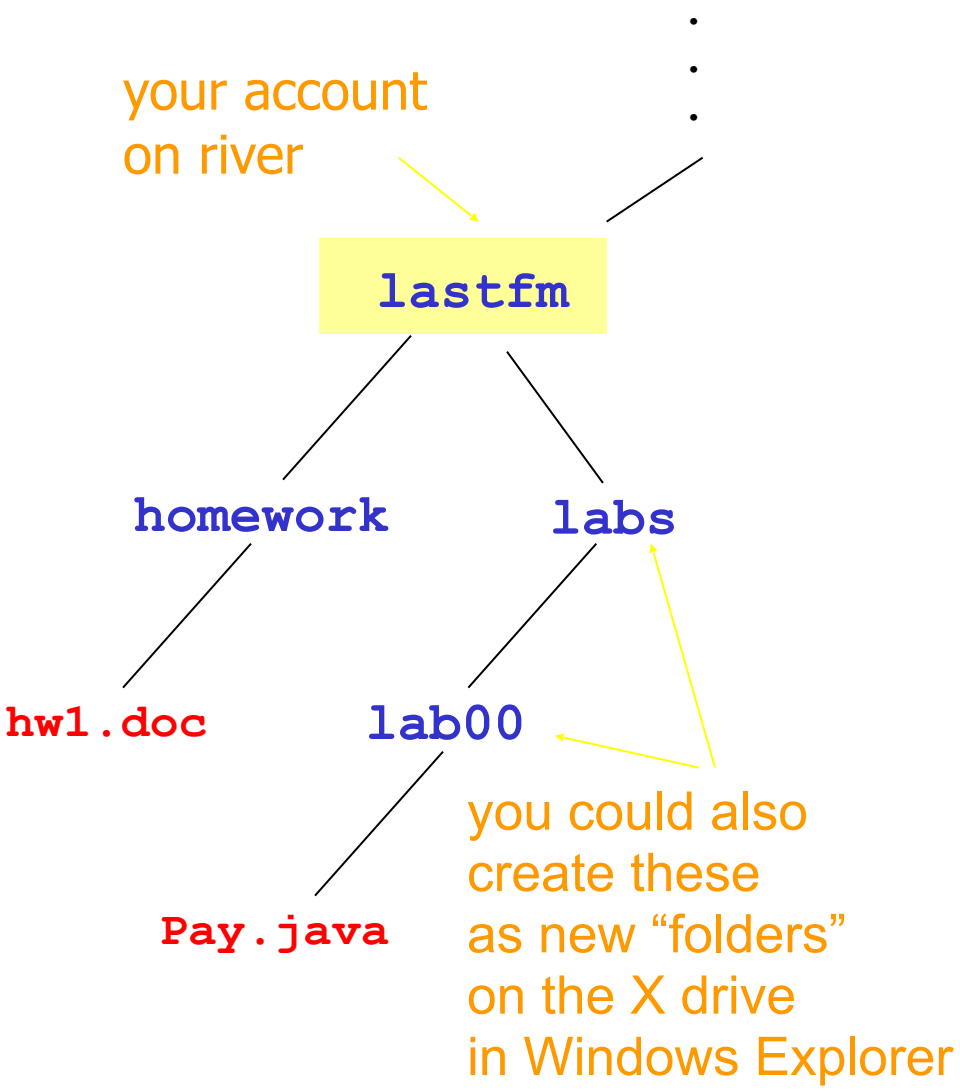
***Any files or directories (folders) you create and save to the "X" drive are saved in your account (directory) on* river**

If from the DOS prompt you type:

```
x:\> mkdir homework

x:\> mkdir labs

x:\> cd labs

x:\labs> mkdir lab00
```

on the PC you create your homework assignment in Word and save it in the homework folder on X drive

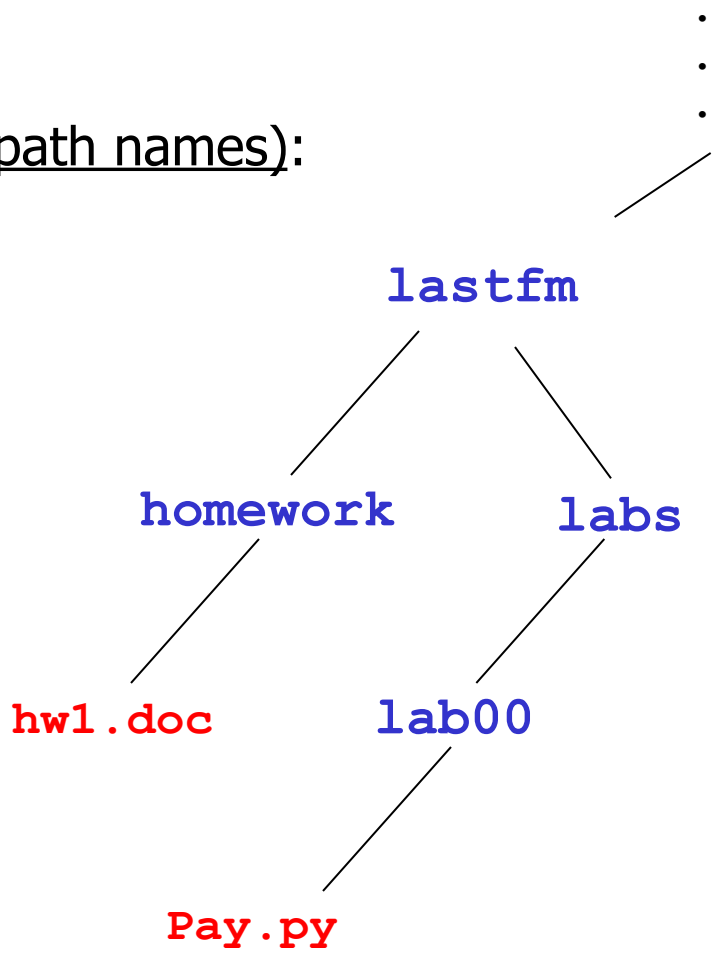
on the PC you use Atom to create your python program source file and save it in the lab00 folder on X



Path Names

Files may be referred to by their
full path names (also called absolute path names):

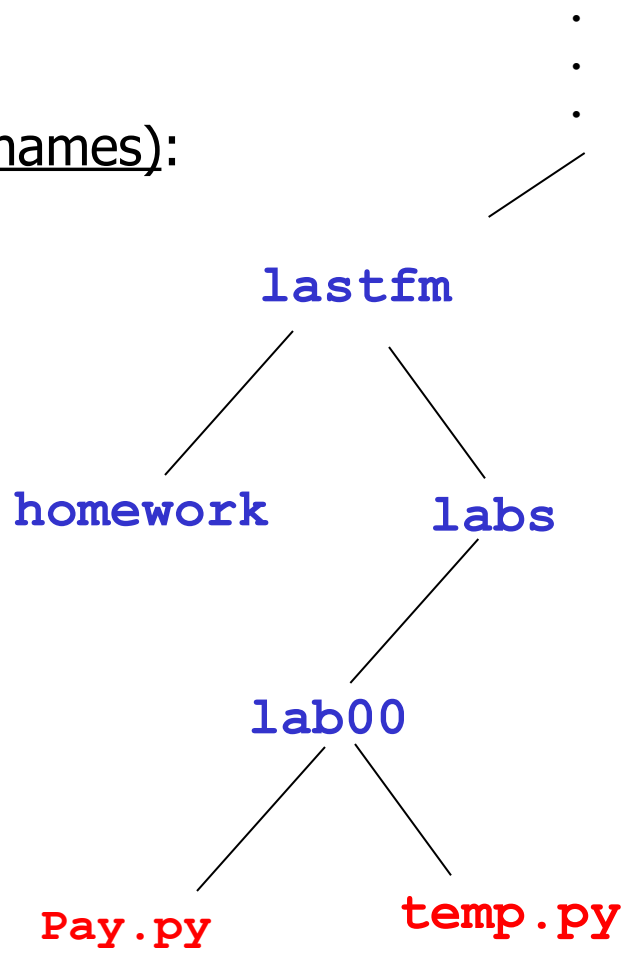
```
x:\> del X:\homework\hw1.doc
```



Path Names

Files may be referred to by their
full path names (also called absolute path names):

```
x:\> del X:\homework\hw1.doc
```



Or files may be referred to by their
relative path names:

```
x:\> cd labs
x:\labs>cd lab00
x:\labs\lab00>copy Pay.py temp.py
```

Data presence in our daily life

Read the handout and understand Filesystems, command line. Leave the last page for now.

Data presence in our daily life

Learn how to use Atom

Data presence in our daily life

Learn how to use Atom

1. How does Python looks like?
2. How to run Python code?
3. Your first python program. (I will give a simple demo, today we are going to try it, next class we will go through this again to make sure you understand it).

Common Errors…

Data presence in our daily life

Second handout about pay.py