

What is programming?

phpm 672/677

Hye-Chung Kum
Population Informatics Research Group
<http://pinformatics.org/>

License:
Data Science in the Health Domain by Hye-Chung Kum is licensed under a
[Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Course URL:
<http://pinformatics.tamhsc.edu/phpm672>



Programming (Thur)

- Watch (lab)
- Do (assignment)
- Sometimes lectures feed into this session
- Other times it doesn't (data science lectures)



Lab Session (Optional)

- Tutor: Zhang Hao
- Fri afternoon?
- Office hours: Computer Lab
 - Thur: ? (2h)
 - Fri: ? (2h)
 - Mon: ? (2h)



Programming: Wikipedia

- **Computer programming** (often shortened to **programming**) is the **comprehensive process that leads from an original formulation of a computing problem to executable programs**. It involves activities such as analysis, understanding, and generically solving such problems resulting in an algorithm, verification of requirements of the algorithm including its correctness and its resource consumption, implementation (or coding) of the algorithm in a target programming language, testing, debugging, and maintaining the source code, implementation of the build system and management of derived artefacts such as machine code of computer programs. **The algorithm is often only represented in human-parseable form and reasoned about using logic**. Source code is written in one or more programming languages (such as C++, C#, Java, Python, Smalltalk, JavaScript, etc.). **The purpose of programming is to find a sequence of instructions that will automate performing a specific task or solve a given problem**. The process of programming thus often requires expertise in many different subjects, including knowledge of the application domain, specialized algorithms and formal logic.



Programming: Wikipedia (cont)

- There is an on-going debate on the extent to which **the writing of programs is an art form, a craft, or an engineering discipline**. In general, good programming is considered to be the measured application of all three, with the **goal of producing an efficient and evolvable software solution (the criteria for "efficient" and "evolvable" vary considerably)**. The discipline differs from many other technical professions in that programmers, in general, do not need to be licensed or pass any standardized (or governmentally regulated) certification tests in order to call themselves "programmers" or even "software engineers." Because the discipline covers many areas, which may or may not include critical applications, it is debatable whether licensing is required for the profession as a whole. In most cases, the discipline is self-governed by the entities which require the programming, and sometimes very strict environments are defined (e.g. United States Air Force use of AdaCore and security clearance). However, representing oneself as a "professional software engineer" without a license from an accredited institution is illegal in many parts of the world.
- Another on-going debate is the extent to which **the programming language used in writing computer programs affects the form that the final program takes**. This debate is analogous to that surrounding the Sapir-Whorf hypothesis in linguistics and cognitive science, which postulates that **a particular spoken language's nature influences the habitual thought of its speakers**. Different language patterns yield different patterns of thought. This idea challenges the possibility of representing the world perfectly with language, because it acknowledges that **the mechanisms of any language condition the thoughts of its speaker community**.



Programming for us

- Code/language = SAS
- Two ways to tell the computer what to do
 - Programming: Reusable Code
 - point & click
- Software Engineering
 - Project management
 - Build a process (sequences of codes/data) that can be efficiently managed
 - Tractable, replicable, agile (quick to modify)
 - Need to be able to pick up after 6 months
- Very disciplined activity
 - Must keep to rules



Programming: Hye-Chung Kum

- Take INPUT and turn it into OUTPUT
 - OUTPUT : Know what you want/goal
 - INPUT : figure out what you have to work with
- Change what you have (INPUT) to what you need (OUTPUT)
 - Break up the problems into small subproblems
 - Intermediate results (scrap paper)
 - **Use functions** to calculate the intermediary results



Working environment

- Computer lab
 - Must save out files to/from dropbox or USB
 - Recommend running directly from C:/
 - Recommend using the same computer
 - No guarantee on files
- On your laptop
 - Install before next week
- On your desktop
 - If home PC: Install before next week
 - If office: figure out software before next week
 - Figure out how to remote into from the lab
- Virtual Lab



Hello World

- SAS
 - How to start
 - Saving out log, lst
- Text editors: window into communicating with computers
- If time permits, go over code in the assignment



Resources

- ucla ats: <http://www.ats.ucla.edu/stat/sas>
- SAS online manual
- Population informatics website

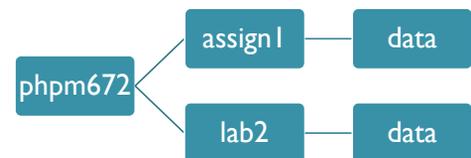


Programming

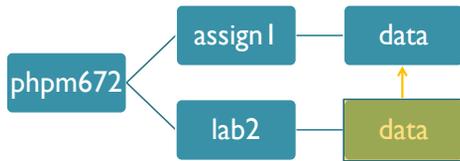
- Know your computer
 - Do you know exactly where everything is?
 - Be the boss, stay in charge
- File management:
 - logical organization that makes sense to you
 - More than one logical organization
 - Short cuts
 - Symbolic links (ln -s)
 - Not too deep, but enough to keep organized
- File extensions
 - Tells your computer what the files are for



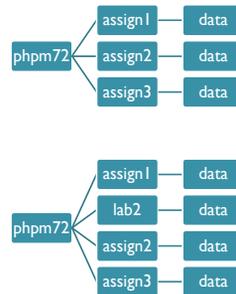
Logical structure



Physical structure



Logical structure



Remember

- Read
 - Read recommended modules in the assignment
- Practice
 - Try using SAS
 - Try examples in the modules you are reading
- Assignment #1
 - Due 11:59pm on Jan 23 (Tuesday)

Assignment 1

- **Decide your work environment for the class (MOST IMPORTANT)**
- You need to first decide your mode of work for this class. The overhead of working "temporarily" is too high for this class. So you need to decide which computer you will be using for this class for the full semester. The following is a list of requirements for the computer you choose to use.
 - It must have SAS installed. All students can install SAS on your personal computers for free, using the SAS University Edition (https://www.sas.com/en_us/software/university-edition.html). But sometimes you will need help getting that installed. So if this is what you plan to do, make sure to get SAS installed ASAP. If you are using computer owned by the university (e.g., your desktop in the office) you will have to figure out who will pay for the license. There is no support for this.
 - The VDAL is another alternative. Using VDAL you are able to access SAS using your personal or university computer as long as you have internet access.
 - It is best if there is hard disk space for you to store your work for the whole semester. If not, such as using the lab computers, you will have to purchase a USB external harddrive to store class material.
 - You need to be able to access it during class. That is, it is a lab computer, or you have your laptop with you, or you remote desktop into the computer from the lab.
 - If you do not have better options, you will have to use the lab computers. This means you will buy the USB to store class material.
- If at all possible, we recommend that you decide and setup, then use that computing environment to do this assignment. This will make it possible for you to focus on learning the rest of the semester.

Assignment 1

```

*This program imports, describes, and displays the dataset;
* EXAMPLE: libname data 'C:\phpm672\assign1';
libname data 'LOCATION OF FOLDER'; *you have to type in location;

*you have to type in location and filename;
* EXAMPLE: datafile= C:\phpm672\assign1\flu.csv;
proc import out= data.flu
  datafile= "LOCATION AND NAME OF FILE"
  dbms=csv replace;
  getnames=yes;
  datarow=2;

proc contents data=flu;
proc print data=flu(obs=10);
  
```

4 things to submit

- **Submission.** Submit on E-campus by 11:59pm the day before the class they are due.
 - Commented code (SAS: **InameN.sas**, where N indicates the assignment number and Iname is your last name)
 - Output from your code (SAS: **InameN.log** & **InameN.lst** or **InameN.html**)
 - **Readme.txt** file answers to questions