



Larry Gray, PhD

Data Scientist at KPMG;  
Data Science Faculty at Georgetown University  
Core Contributor for Yellowbrick.

LinkedIn: <https://linkedin.com/in/larry-gray-phd>

Github: [github.com/lwgray](https://github.com/lwgray)

Email: [lwgray@gmail.com](mailto:lwgray@gmail.com)

# PYTHON BASICS FOR DATA ANALYSIS

About the Instructor

Dr. Larry Gray  
Maryland Institute College of Art  
October, 18, 2020

See syllabus in Google folder

# AGENDA SESSION 1



Installing  
Software



Python  
Overview



What is  
computing?



Explore the  
Shell



Interactivate  
Python

# INSTALLING SOFTWARE

Google Folder containing all materials: Use link below to access them

<https://drive.google.com/drive/folders/1hMj9foSt9w-QdxQ4K4VBVV3YhWSk72PJ?usp=sharing>

First let's download and install the Anaconda distribution of **Python 3.7** for your operating system at

<https://www.anaconda.com/distribution>

**Directions in Google folder! Windows users, please pay attention to the configuration screens!**

And the text editor Sublime Text from

<https://www.sublimetext.com>

**Directions (though it is quite easy) are on in Google folder!**

**THEN A PEP TALK**

You can do it!!!

# HTTP://HACKERTYPER.NET

```
struct group_info init_groups = { .usage = ATOMIC_INIT(2) };

struct group_info *groups_alloc(int gidsetsize){
    struct group_info *group_info;
    int nblocks;
    int i;

    nblocks = (gidsetsize + NGROUPS_PER_BLOCK - 1) / NGROUPS_PER_BLOCK;
    /* Make sure we always allocate at least one indirect block pointer */
    nblocks = nblocks ? : 1;
    group_info = kmalloc(sizeof(*group_info) + nblocks*sizeof(gid_t *), GFP_USER);
    if (!group_info)
        return NULL;

    group_info->ngroups = gidsetsize;
    group_info->nblocks = nblocks;
    atomic_set(&group_info->usage, 1);

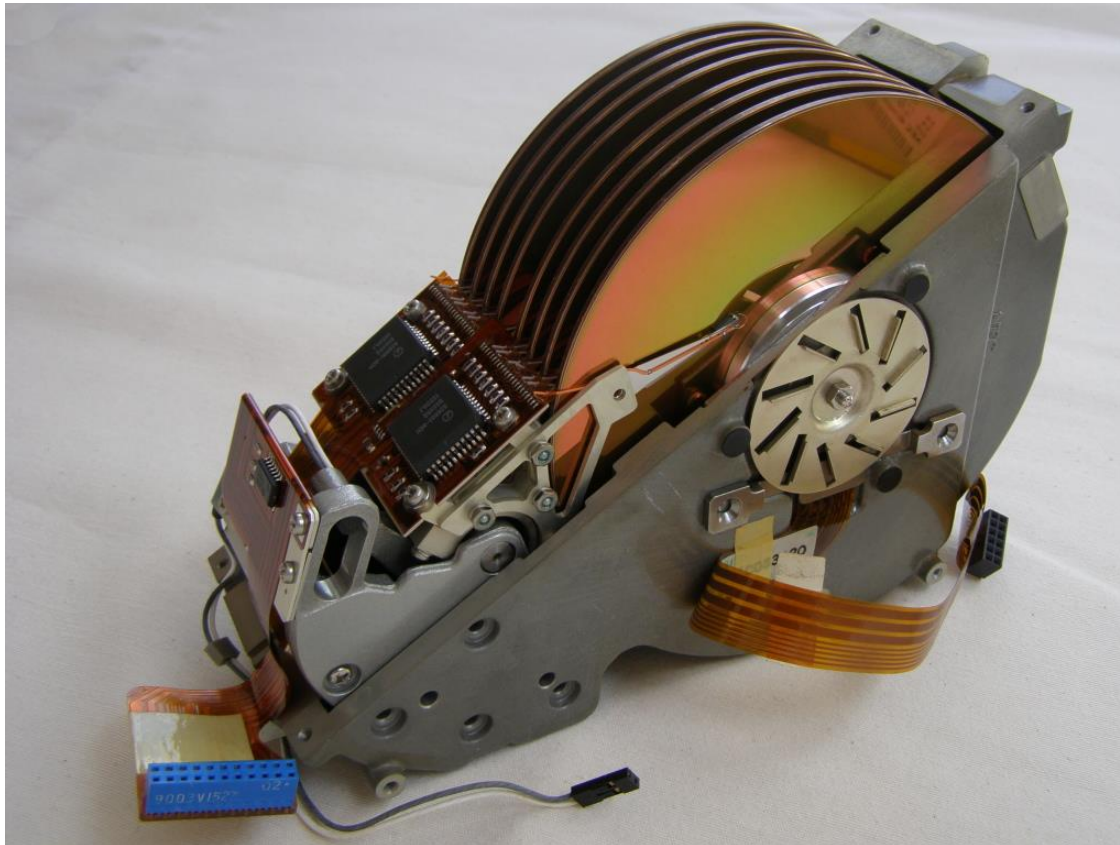
    if (gidsetsize <= NGROUPS_SMALL)
        group_info->blocks[0] = group_info->small_block;
    else {
        for (i = 0; i < nblocks; i++) {
            gid_t *b;
            b = (void *)__ge
```

**WHAT IS COMPUTING?**

# BACK TO BASICS



# DISK: NEC D5662 HARD DISK



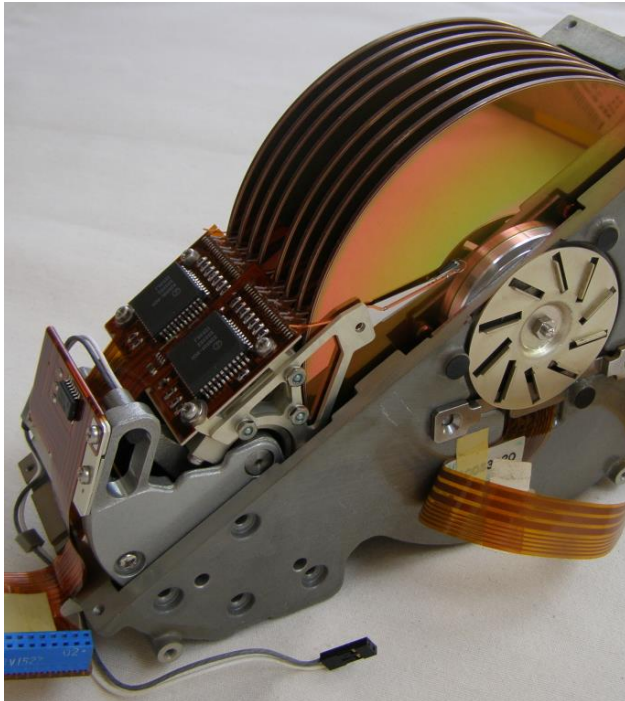


# PROCESSOR: INTEL CORE I5 2.66 GHZ

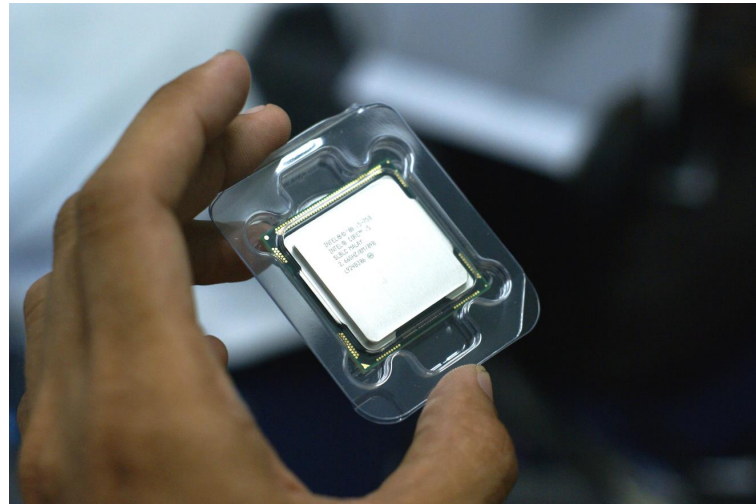


# COMPUTING: SO EASY RIGHT?

**Disk = Data**

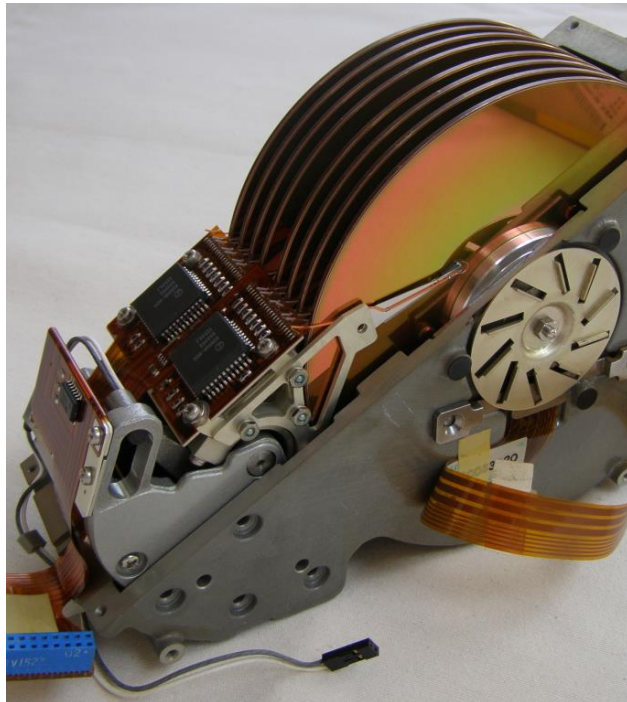


**Processor = Instructions**



# COMPUTING: SO EASY RIGHT?

**Disk = Data**  
**= Data Types/Classes**

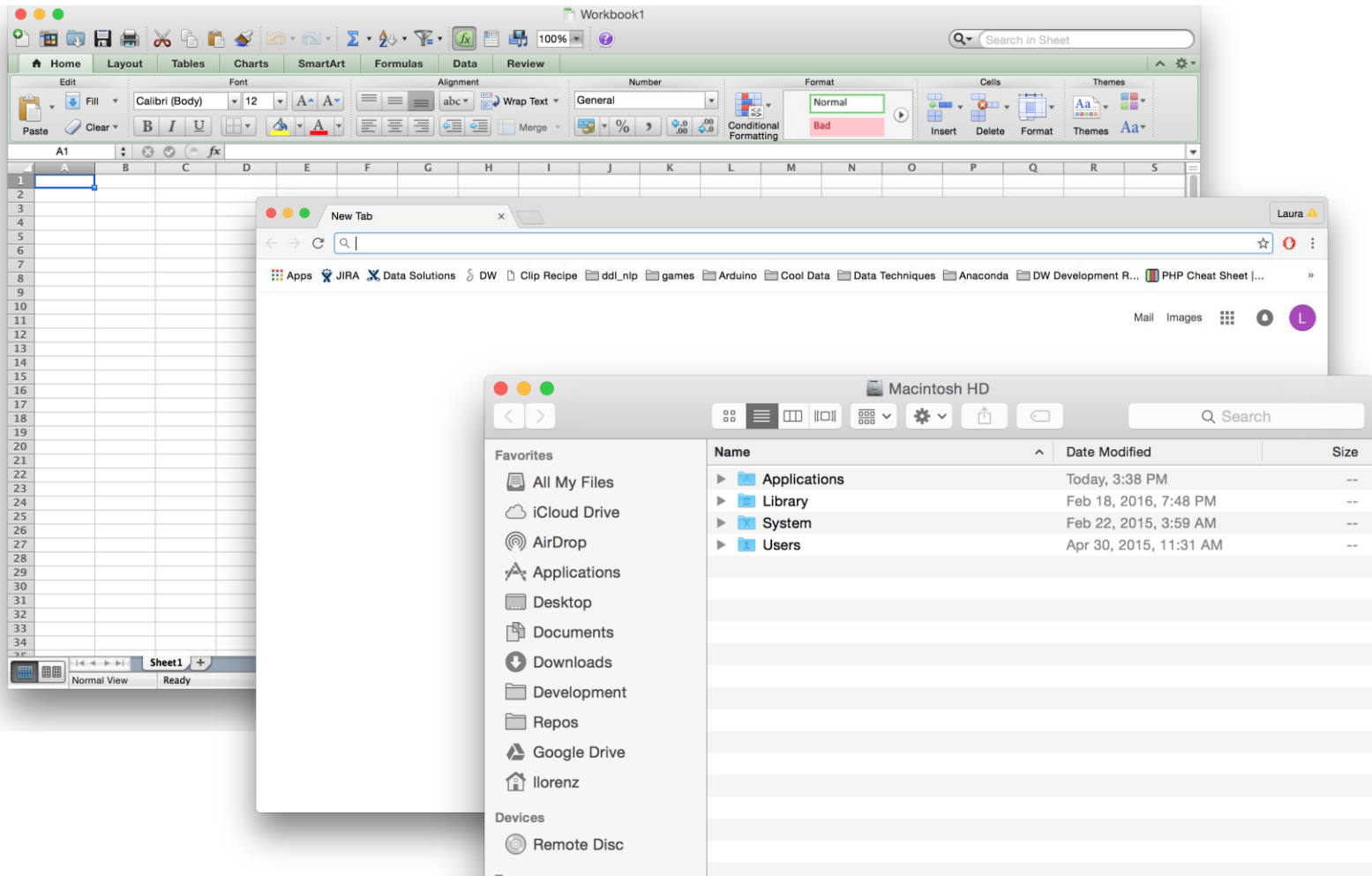


**Processor = Instructions**  
**= Methods/Functions**



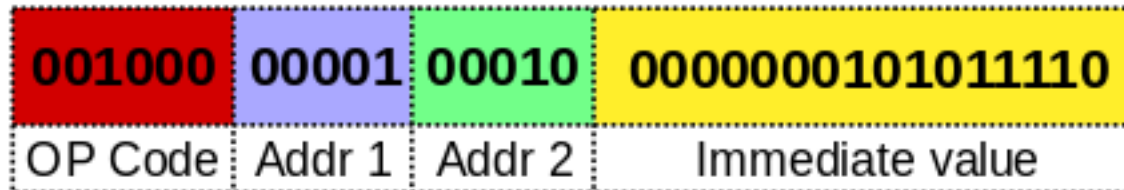
**TALKING TO YOUR COMPUTER**

# GUIs!



# PROGRAMMING: INSTRUCTION SETS/ASSEMBLY

## MIPS32 Add Immediate Instruction



Equivalent mnemonic:      **addi** \$r1, \$r2, 350

# PROGRAMMING: C

---

```
//C hello world example  
#include <stdio.h>  
  
int main()  
{  
    printf("Hello world\n");  
    return 0;  
}
```

# PROGRAMMING: R

```
summary <- function(x) {  
  funs <- c(mean, median, sd, mad, IQR)  
  lapply(funs, function(f) f(x, na.rm = TRUE))  
}
```



# PROGRAMMING: PYTHON

```
class MRWordCount(MRJob):

    def mapper(self, _, line):
        # self.set_status('mapper')
        self.increment_counter('mapper_group', 'items_mapped', 1)
        for word in line.split():
            yield word, 1

    def reducer(self, word, counts):
        # self.set_status('reducer')
        self.increment_counter('reducer_group', 'items_reduced', 1)
        yield word, sum(counts)
```

# EXPLORING THE SHELL

# SHELL/COMMAND LINE/TERMINAL

```
Fall 2016 — bash — 80x24
~ $ cd Documents/Georgetown/Python\ Basics\ for\ Data\ Analysis/Fall\ 2016/
Fall 2016 $ pwd
/Users/llorenz/Documents/Georgetown/Python Basics for Data Analysis/Fall 2016
Fall 2016 $ ls
PythonBasicsforDataAnalysis_Day1.pptx
PythonForDataAnalysis_Syllabus_Fall2016.docx
PythonForDataAnalysis_Syllabus_Fall2016.pdf
Seminar 1.ipynb
Fall 2016 $
```

```
Windows PowerShell
PS C:\Users\llorenz> pwd
Path
----
C:\Users\llorenz

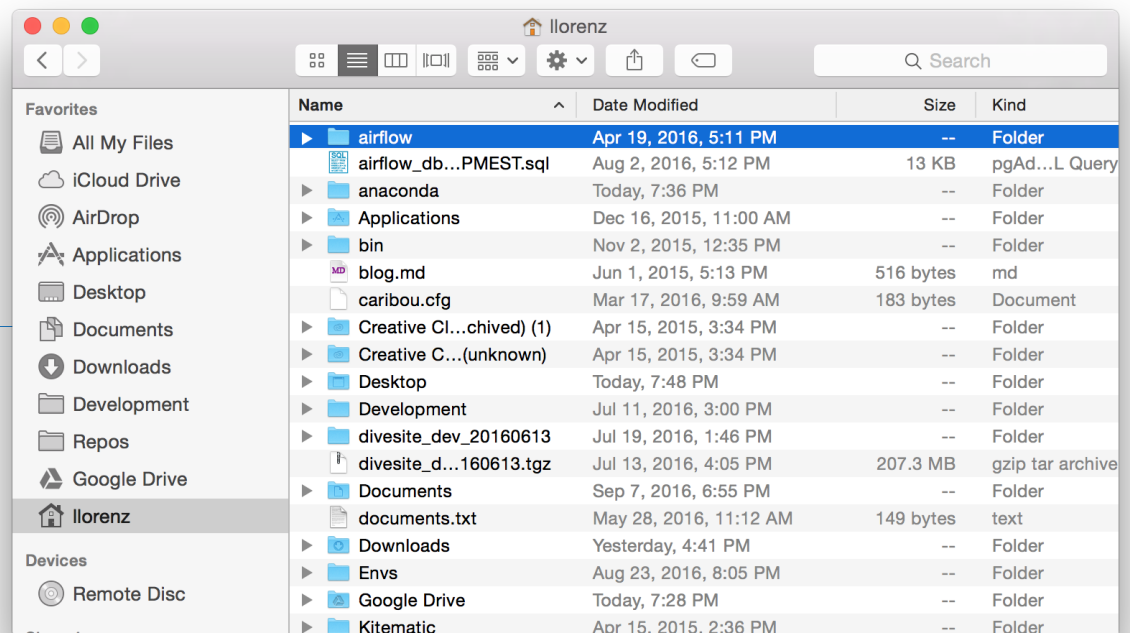
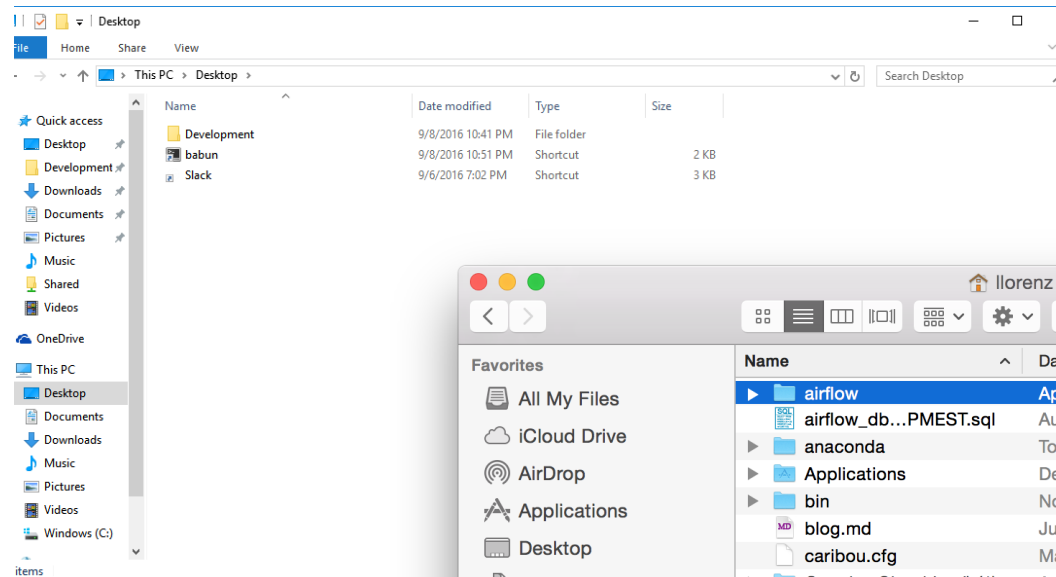
PS C:\Users\llorenz> ls

Directory: C:\Users\llorenz

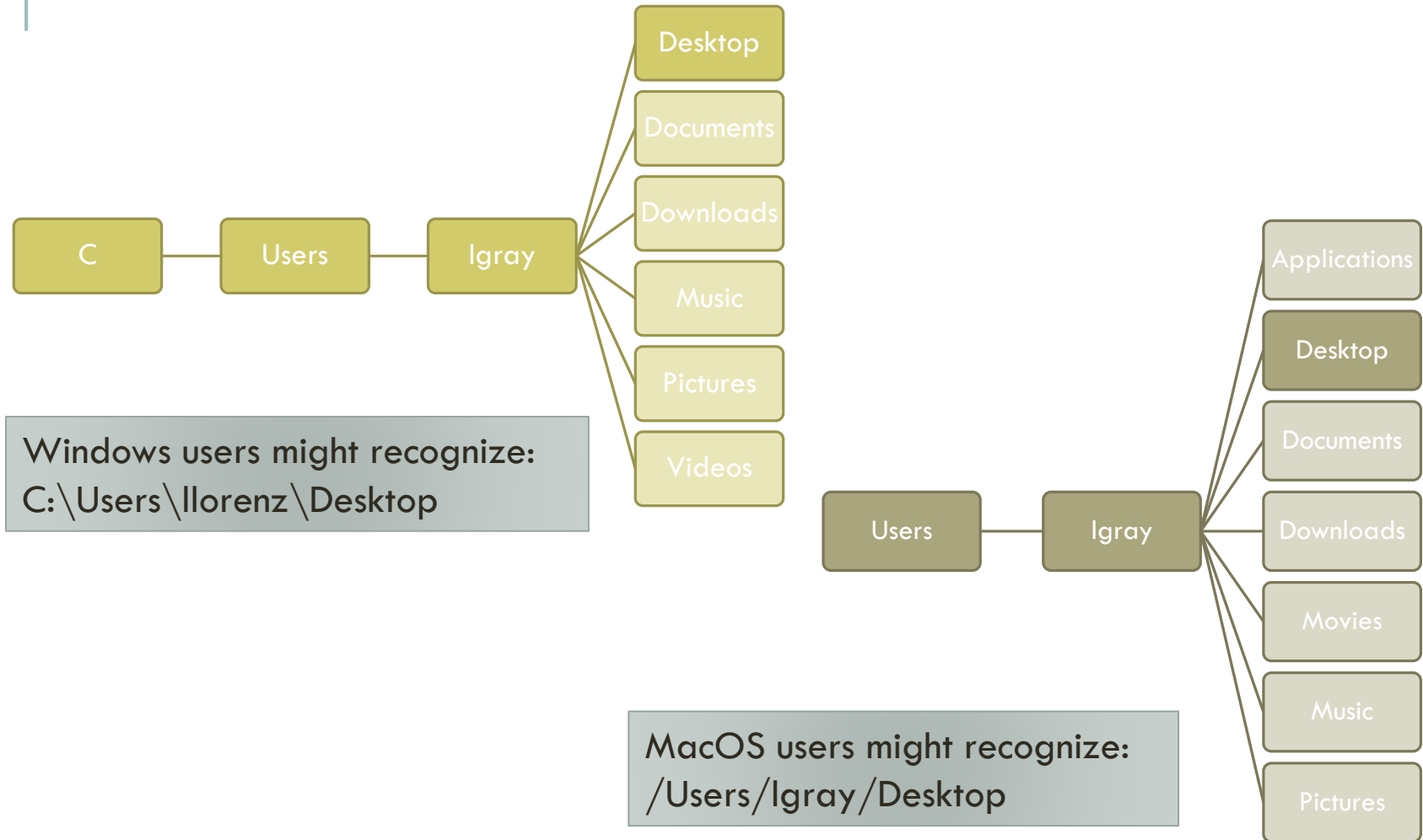
Mode                LastWriteTime         Length Name
----                -
d-----          7/12/2017  8:57 PM             Anaconda3
d-r-----          7/12/2017  8:37 PM             Contacts
d-r-----          7/12/2017  8:37 PM             Desktop
d-r-----          7/12/2017  8:57 PM             Documents
d-r-----          7/12/2017  8:47 PM             Downloads
d-r-----          7/12/2017  8:37 PM             Favorites
d-r-----          7/12/2017  8:44 PM             Links
d-r-----          7/12/2017  8:37 PM             Music
d-r-----          7/12/2017  8:44 PM             OneDrive
d-r-----          7/12/2017  8:38 PM             Pictures
d-r-----          7/12/2017  8:37 PM             Saved Games
d-r-----          7/12/2017  8:37 PM             Searches
d-r-----          7/12/2017  8:37 PM             Videos
-a-----          7/12/2017  9:32 PM              8 test.txt

PS C:\Users\llorenz>
```

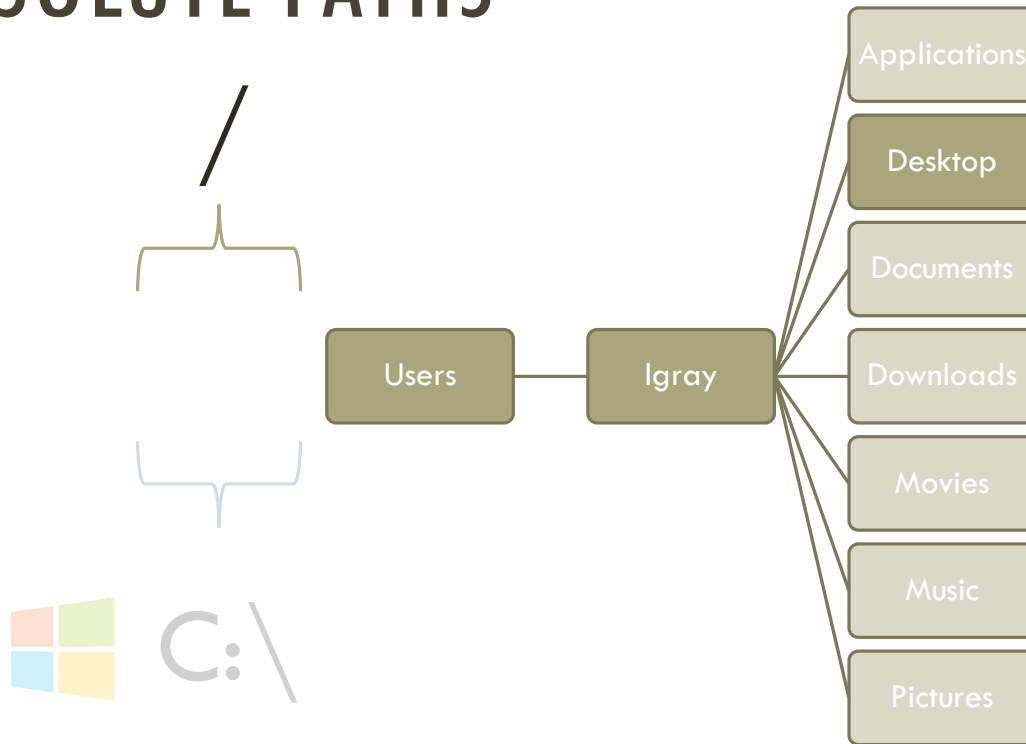
# FILE SYSTEM BASICS



# FILE SYSTEM BASICS



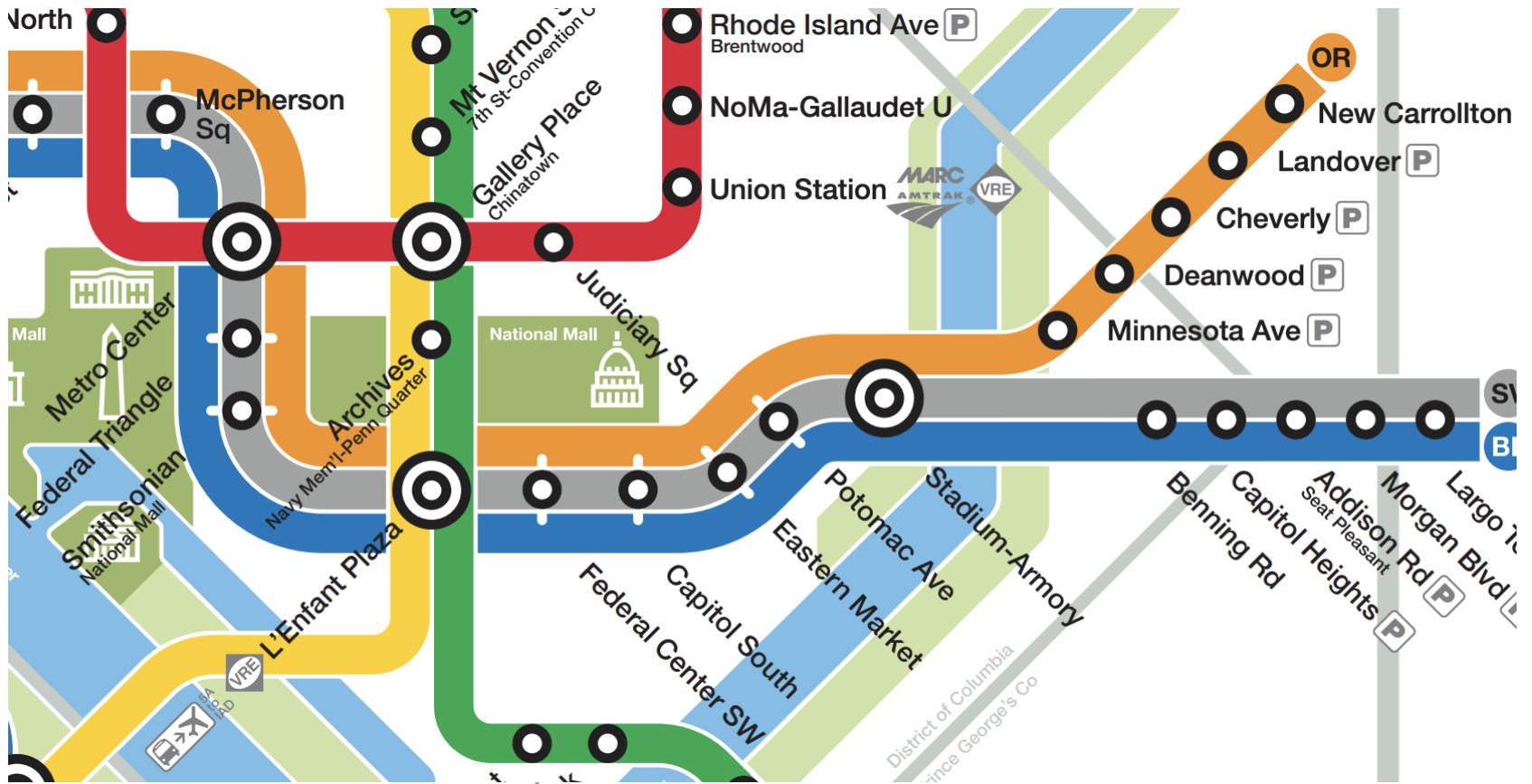
# ABSOLUTE PATHS



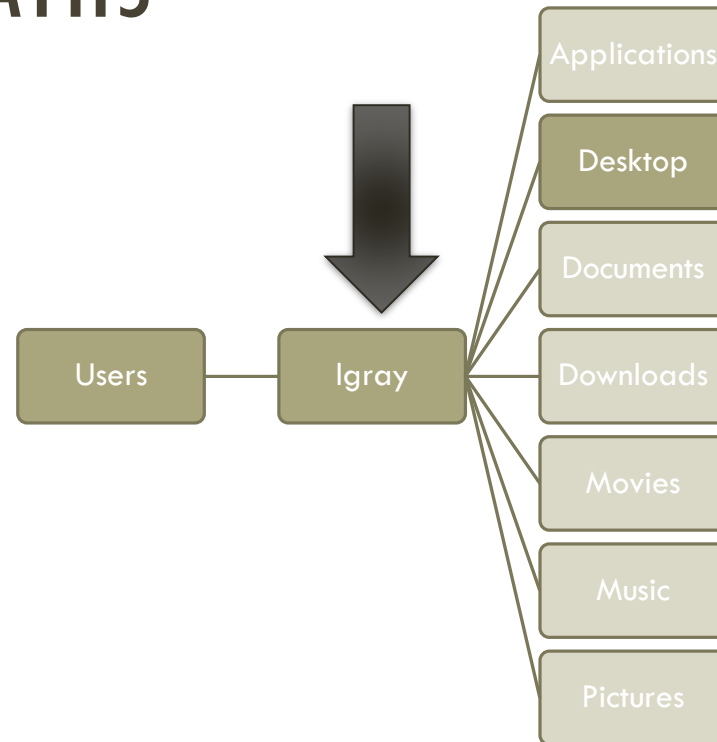
/Users/Igray/Desktop



C:\Users\Igray\Desktop

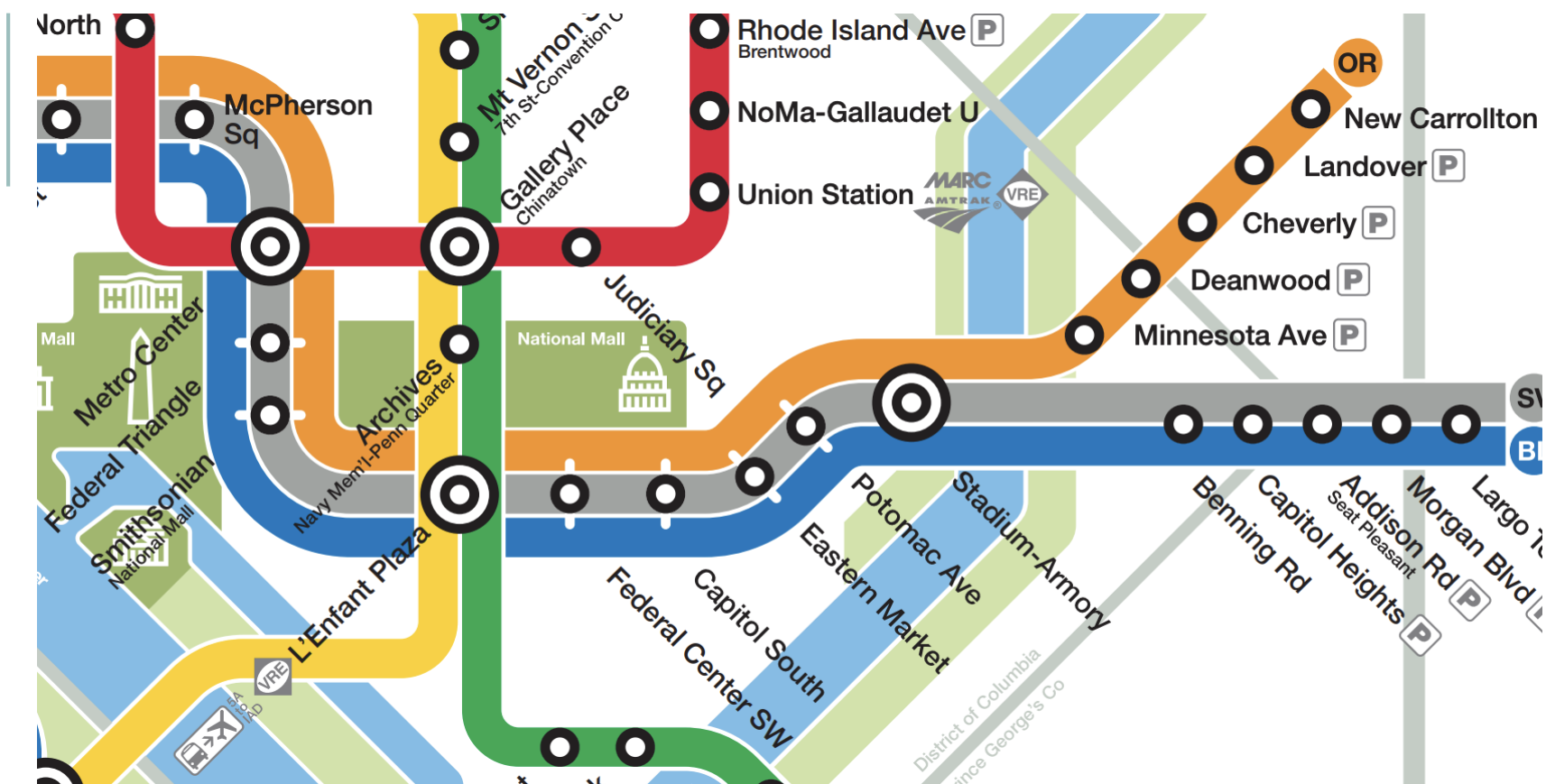


# RELATIVE PATHS

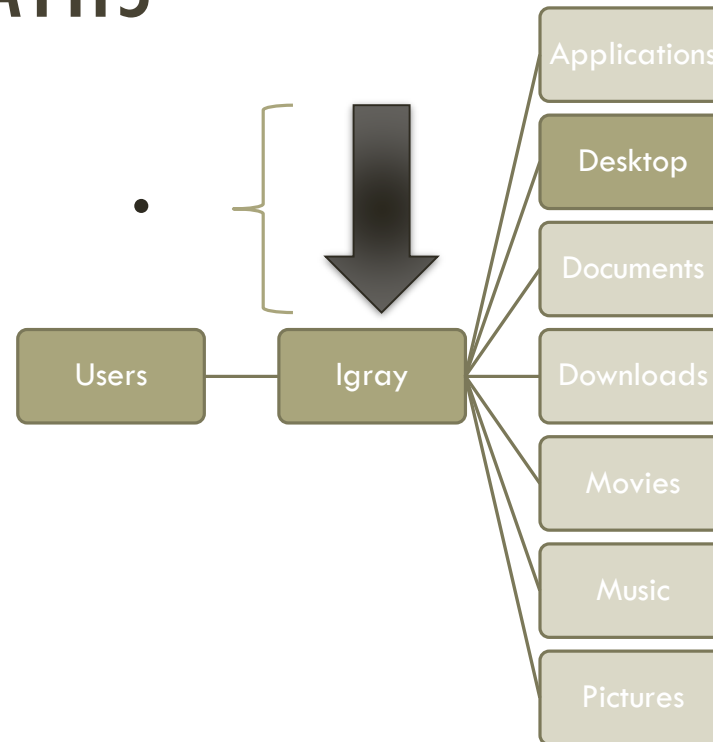


Desktop





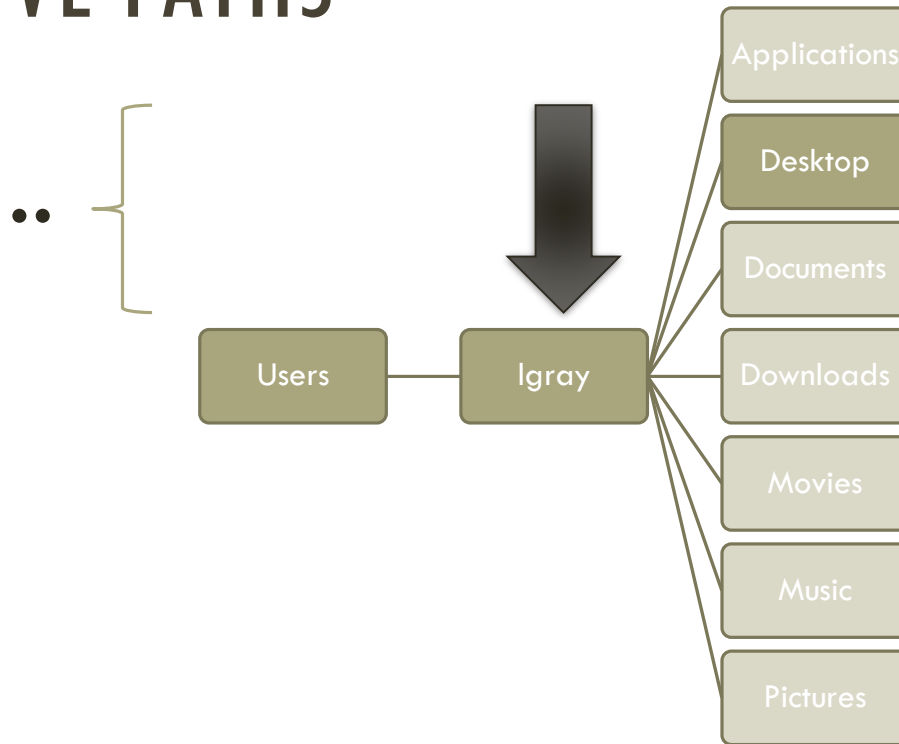
# RELATIVE PATHS



Desktop  
./Desktop  
.\Desktop



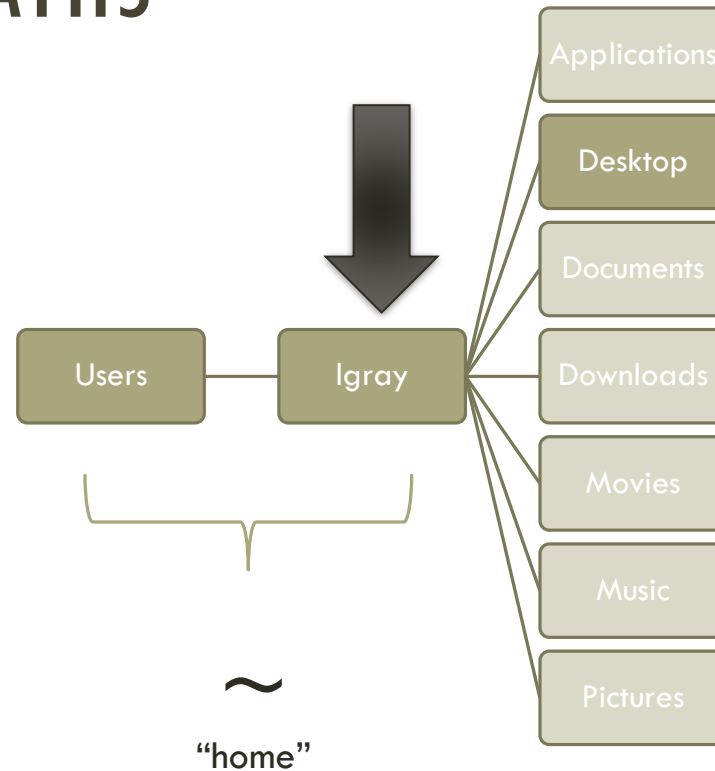
# RELATIVE PATHS



Desktop  
./Desktop  
../Igray/Desktop



# RELATIVE PATHS



~ is a concept specific to your terminal program.

**Terminal default:**

`/Users/Igray`

**Powershell default:**

`C:\Users\Igray`

Desktop  
./Desktop  
~/Desktop

# SHELL/COMMAND LINE/TERMINAL

```
Fall 2016 — bash — 80x24
~ $ cd Documents/Georgetown/Python\ Basics\ for\ Data\ Analysis/Fall\ 2016/
Fall 2016 $ pwd
/Users/llorenz/Documents/Georgetown/Python Basics for Data Analysis/Fall 2016
Fall 2016 $ ls
PythonBasicsforDataAnalysis_Day1.pptx
PythonForDataAnalysis_Syllabus_Fall2016.docx
PythonForDataAnalysis_Syllabus_Fall2016.pdf
Seminar 1.ipynb
Fall 2016 $
```

```
Windows PowerShell
PS C:\Users\llorenz> pwd
Path
----
C:\Users\llorenz

PS C:\Users\llorenz> ls

Directory: C:\Users\llorenz

Mode                LastWriteTime         Length Name
----                -
d-----          7/12/2017  8:57 PM              Anaconda3
d-r---          7/12/2017  8:37 PM              Contacts
d-r---          7/12/2017  8:37 PM              Desktop
d-r---          7/12/2017  8:57 PM              Documents
d-r---          7/12/2017  8:47 PM              Downloads
d-r---          7/12/2017  8:37 PM              Favorites
d-r---          7/12/2017  8:44 PM              Links
d-r---          7/12/2017  8:37 PM              Music
d-r---          7/12/2017  8:44 PM              OneDrive
d-r---          7/12/2017  8:38 PM              Pictures
d-r---          7/12/2017  8:37 PM              Saved Games
d-r---          7/12/2017  8:37 PM              Searches
d-r---          7/12/2017  8:37 PM              Videos
-a----          7/12/2017  9:32 PM              8 test.txt

PS C:\Users\llorenz>
```

# SHELL/COMMAND LINE / TERMINAL

```
Fall 2016 — bash
~ $ cd Documents/Georgetown/Python\ Basics\ for
Fall 2016 $ pwd
/Users/llorenz/Documents/Georgetown/Python Basics
Fall 2016 $ ls
PythonBasicsforDataAnalysis_Day1.pptx
PythonForDataAnalysis_Syllabus_Fall2016.docx
PythonForDataAnalysis_Syllabus_Fall2016.pdf
Seminar 1.ipynb
Fall 2016 $
```

```
bash
~ $ clear
~ $ pwd
/home/Laura
~ $ cd anaconda3/
{ anaconda3 } $ pwd
/home/Laura/anaconda3
{ anaconda3 } $ ls
api-ms-win-core-console-l1-1-0.dll      api-ms-win-crt-time-l1-1-0.dll
api-ms-win-core-datetime-l1-1-0.dll    api-ms-win-crt-utility-l1-1-0.dll
api-ms-win-core-debug-l1-1-0.dll       concrt140.dll
api-ms-win-core-errorhandling-l1-1-0.dll  conda-meta
api-ms-win-core-file-l1-1-0.dll         cwp.py
api-ms-win-core-file-l1-2-0.dll         DLLs
api-ms-win-core-file-l2-1-0.dll         Doc
api-ms-win-core-handle-l1-1-0.dll       envs
api-ms-win-core-heap-l1-1-0.dll         etc
api-ms-win-core-interlocked-l1-1-0.dll  include
api-ms-win-core-libraryloader-l1-1-0.dll info
api-ms-win-core-localization-l1-2-0.dll lib
```

```
Windows PowerShell
PS C:\Users\llorenz> pwd
Path
----
C:\Users\llorenz
```

```
Command Prompt
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\Laura>dir
Volume in drive C is Windows
Volume Serial Number is FA4F-4AC2

Directory of C:\Users\Laura

09/08/2016  11:44 PM  <DIR>          .
09/08/2016  11:44 PM  <DIR>          ..
09/08/2016  10:50 PM  <DIR>          .babun
09/08/2016  10:19 PM  <DIR>          Contacts
09/08/2016  10:51 PM  <DIR>          Desktop
09/08/2016  11:23 PM  <DIR>          Documents
09/08/2016  10:43 PM  <DIR>          Downloads
09/08/2016  10:19 PM  <DIR>          Favorites
09/08/2016  10:19 PM  <DIR>          Links
09/08/2016  10:19 PM  <DIR>          Music
09/08/2016  10:22 PM  <DIR>          OneDrive
09/08/2016  10:19 PM  <DIR>          Pictures
06/11/2016  12:09 PM  <DIR>          Roaming
09/08/2016  10:19 PM  <DIR>          Saved Games
09/08/2016  10:19 PM  <DIR>          Searches
09/08/2016  10:19 PM  <DIR>          Videos
               0 File(s)            0 bytes
              16 Dir(s)  80,485,457,920 bytes free

C:\Users\Laura>cd Documents
```

```
Name
----
Anaconda3
Contacts
Desktop
Documents
Downloads
Favorites
Links
Music
OneDrive
Pictures
Saved Games
Searches
Videos
test.txt
```

# SHELL/COMMAND LINE / TERMINAL

```
Fall 2016 — bash
~ $ cd Documents/Georgetown/Python\ Basics\ for Fall 2016
Fall 2016 $ pwd
/Users/llorenz/Documents/Georgetown/Python Basics for Fall 2016
Fall 2016 $ ls
PythonBasicsforDataAnalysis_Syllabus_Fall2016.pdf
PythonForDataAnalysis_Seminar 1.ipynb
Fall 2016 $
```

bash

```
bash
~ $ clear
~ $ pwd
/home/Laura
~ $ cd anaconda3/
{ anaconda3 } $ pwd
/home/Laura/anaconda3
{ anaconda3 } $ ls
api-ms-win-core-console-l1-1-0.dll
api-ms-win-core-datetime-l1-1-0.dll
api-ms-win-core-debug-l1-1-0.dll
api-ms-win-core-errorhandling-l1-1-0.dll
api-ms-win-core-file-l1-1-0.dll
api-ms-win-core-file-l1-2-0.dll
api-ms-win-core-file-l2-1-0.dll
api-ms-win-core-handle-l1-1-0.dll
api-ms-win-core-heap-l1-1-0.dll
api-ms-win-core-interlocked-l1-1-0.dll
api-ms-win-core-libraryloader-l1-1-0.dll
api-ms-win-core-localization-l1-2-0.dll
api-ms-win-core-memory-l1-1-0.dll
api-ms-win-core-nls-l1-1-0.dll
api-ms-win-core-ole32-l1-1-0.dll
api-ms-win-core-oleaut-l1-1-0.dll
api-ms-win-core-processmail-l1-1-0.dll
api-ms-win-core-profile-l1-1-0.dll
api-ms-win-core-shell-l1-1-0.dll
api-ms-win-core-string-l1-1-0.dll
api-ms-win-core-synch-l1-1-0.dll
api-ms-win-core-sysinfo-l1-1-0.dll
api-ms-win-core-time-l1-1-0.dll
api-ms-win-core-util-l1-1-0.dll
api-ms-win-crt-heap-l1-1-0.dll
api-ms-win-crt-runtime-l1-1-0.dll
api-ms-win-crt-utility-l1-1-0.dll
concrtdll.dll
conda-meta
etc
include
info
lib
```

bash

```
Windows PowerShell
PS C:\Users\llorenz> pwd
C:\Users\llorenz
Path
----
C:\Users\llorenz
```

PowerShell

```
Command Prompt
Microsoft Windows [Version 10.0.14393]
(c) 2016 Microsoft Corporation. All rights reserved.

C:\Users\Laura>dir
Volume in drive C is Windows
Volume Serial Number is FA4F-4AC2

Directory of C:\Users\Laura

09/08/2016  11:44 PM  <DIR>          .
09/08/2016  11:44 PM  <DIR>          ..
09/08/2016  10:50 PM  <DIR>          .babun
09/08/2016  10:19 PM  <DIR>          Contacts
09/08/2016  10:51 PM  <DIR>          Desktop
09/08/2016  11:23 PM  <DIR>          Documents
09/08/2016  10:43 PM  <DIR>          Downloads
09/08/2016  10:19 PM  <DIR>          Favorites
09/08/2016  10:19 PM  <DIR>          Links
09/08/2016  10:19 PM  <DIR>          Music
09/08/2016  10:22 PM  <DIR>          OneDrive
09/08/2016  10:19 PM  <DIR>          Pictures
06/11/2016  12:09 PM  <DIR>          Roaming
09/08/2016  10:19 PM  <DIR>          Saved Games
09/08/2016  10:19 PM  <DIR>          Searches
09/08/2016  10:19 PM  <DIR>          Videos
               0 File(s)            0 bytes
              16 Dir(s)  80,485,457,920 bytes free

C:\Users\Laura>cd Documents
```

DOS

```
Name
----
Anaconda3
Contacts
Desktop
Documents
Downloads
Favorites
Links
Music
OneDrive
Pictures
Saved Games
Searches
Videos
test.txt
```

# SHELL/COMMAND LINE/TERMINAL

```
Fall 2016 — bash — 80x24
~ $ cd Documents/Georgetown/Python\ Basics\ for\ Data\ Analysis/Fall\ 2016/
Fall 2016 $ pwd
/Users/llorenz/Documents/Georgetown/Python Basics for Data Analysis/Fall 2016
Fall 2016 $ ls
PythonBasicsforDataAnalysis_Day1.pptx
PythonForDataAnalysis_Syllabus_Fall2016.docx
PythonForDataAnalysis_Syllabus_Fall2016.pdf
Seminar 1.ipynb
Fall 2016 $
```

```
Windows PowerShell
PS C:\Users\llorenz> pwd
Path
----
C:\Users\llorenz

PS C:\Users\llorenz> ls

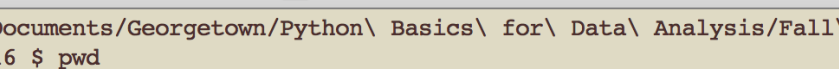
Directory: C:\Users\llorenz

Mode                LastWriteTime         Length Name
----                -
d-----          7/12/2017   8:57 PM             Anaconda3
d-r---          7/12/2017   8:37 PM             Contacts
d-r---          7/12/2017   8:37 PM             Desktop
d-r---          7/12/2017   8:57 PM             Documents
d-r---          7/12/2017   8:47 PM             Downloads
d-r---          7/12/2017   8:37 PM             Favorites
d-r---          7/12/2017   8:44 PM             Links
d-r---          7/12/2017   8:37 PM             Music
d-r---          7/12/2017   8:44 PM             OneDrive
d-r---          7/12/2017   8:38 PM             Pictures
d-r---          7/12/2017   8:37 PM             Saved Games
d-r---          7/12/2017   8:37 PM             Searches
d-r---          7/12/2017   8:37 PM             Videos
-a----          7/12/2017   9:32 PM              8 test.txt

PS C:\Users\llorenz>
```



# SHELL/COMMAND LINE/TERMINAL



Fall 2016 — bash — 80x24

```
~ $ cd Documents/Georgetown/Python\ Basics\ for\ Data\ Analysis/Fall\ 2016/  
Fall 2016 $ pwd  
/Users/llorenz/Documents/Georgetown/Python Basics for Data Analysis/Fall 2016  
Fall 2016 $  
PythonBasic  
PythonForDa  
PythonForDa  
Seminar 1.i  
Fall 2016 $
```

Mac Users:  
~/Applications/Utilities/Terminal

## Mac Users:

~/Applications/Utilities/Terminal

A screenshot of a Windows PowerShell terminal window. The title bar shows a blue icon and the text "Windows PowerShell". The command prompt shows "PS C:\Users\lllorenz> pwd" followed by "Path" and "C:\Users\lllorenz". Then, the user enters "ls", and the output shows the directory path "Directory: C:\Users\lllorenz". Below this, a table-like listing of files and folders is shown with columns for Mode, LastWriteTime, Length, and Name. A semi-transparent grey box is overlaid on the right side of the terminal, containing the text "Windows Users Start > Windows PowerShell".

```
Windows PowerShell
PS C:\Users\lllorenz> pwd
Path
C:\Users\lllorenz

PS C:\Users\lllorenz> ls

Directory: C:\Users\lllorenz


Mode                LastWriteTime         Length Name
----                -
d-----          7/12/2017   8:57 PM           Anaconda3
d-r-----          7/12/2017             Videos
d-r-----          7/12/2017             test.txt
```

Windows Users  
Start > Windows  
PowerShell

```
PS C:\Users\lllorenz>
```

## Windows Users:

Start > Windows PowerShell > Windows PowerShell

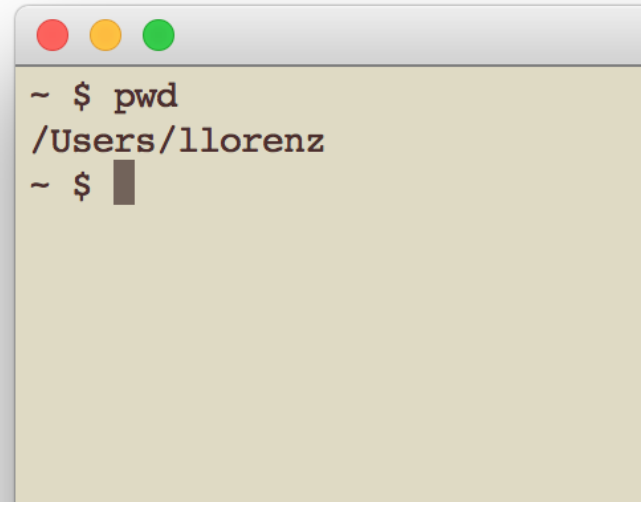
# SHELL/COMMAND LINE/TERMINAL

## pwd

“present working directory”

or,

“WHERE AM I RIGHT NOW?”



```
~ $ pwd
/Users/llorenz
~ $
```

A terminal window with a light beige background and a grey title bar with three colored window control buttons (red, yellow, green). The terminal shows the command 'pwd' being executed, resulting in the output '/Users/llorenz'. The prompt '~ \$' is shown before and after the command.

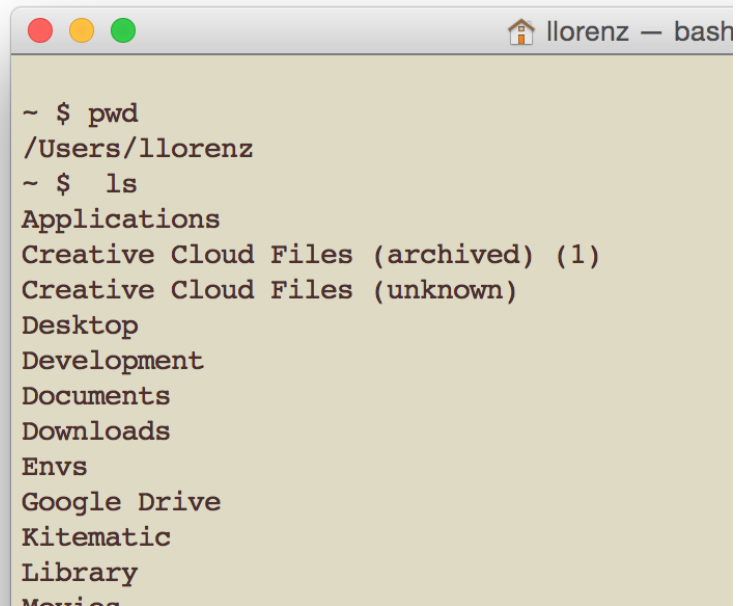
# SHELL/COMMAND LINE/TERMINAL

# ls

“list directory contents”

or,

“what is even in here??”

A screenshot of a macOS terminal window. The title bar shows a home icon, the name 'llorenz', and the shell 'bash'. The terminal has a light beige background. It shows the user running 'pwd' and 'ls'. The output of 'ls' is a list of directories and files in the user's home directory.

```
~ $ pwd
/Users/llorenz
~ $ ls
Applications
Creative Cloud Files (archived) (1)
Creative Cloud Files (unknown)
Desktop
Development
Documents
Downloads
Envs
Google Drive
Kitematic
Library
Movies
```

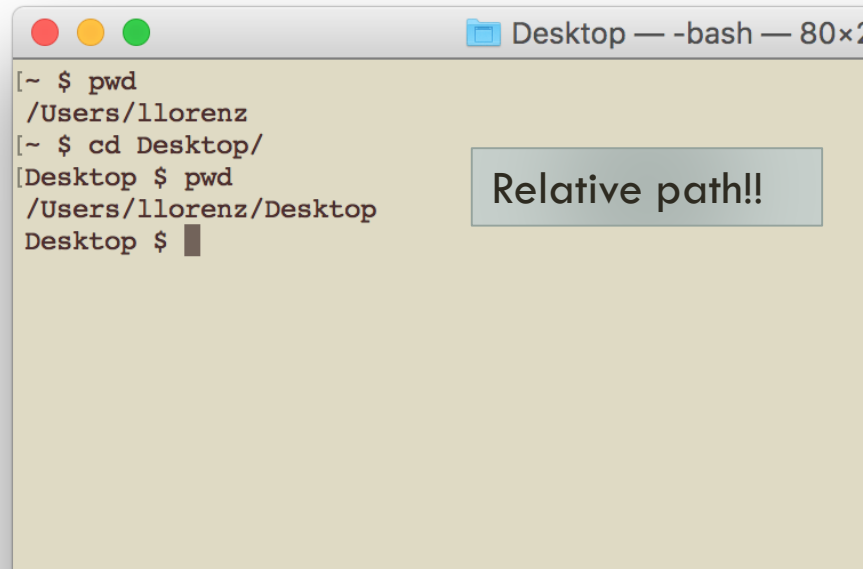
# SHELL/COMMAND LINE/TERMINAL

cd

“change directory”

or,

“get me somewhere else!”

A screenshot of a macOS terminal window titled "Desktop — -bash — 80x20". The terminal shows a sequence of commands and their outputs: first, "pwd" returns "/Users/llorenz"; then, "cd Desktop/" is executed; finally, "pwd" returns "/Users/llorenz/Desktop". The prompt changes from "~" to "Desktop \$". A grey callout box with the text "Relative path!!" points to the "Desktop/" argument in the "cd" command.

```
Desktop — -bash — 80x20
[~ $ pwd
/Users/llorenz
[~ $ cd Desktop/
[Desktop $ pwd
/Users/llorenz/Desktop
Desktop $
```

Relative path!!

# SHELL/COMMAND LINE/TERMINAL

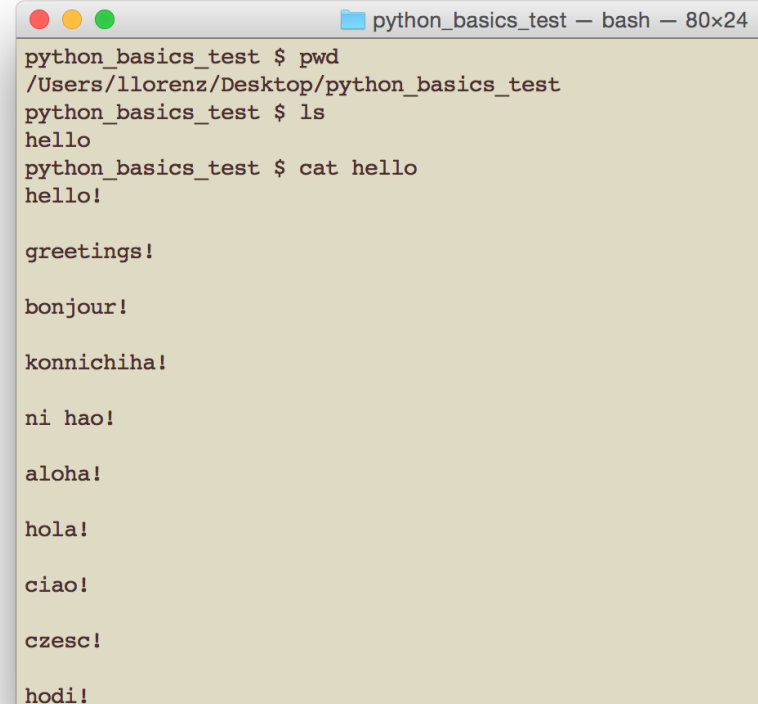
Bonus! You can 'redirect' output into another file with `>`, e.g.  
`cat hello > hello2`

## cat

“concatenate and print files”

or,

“print out all the contents of  
this/these file/files to my terminal  
screen”



```
python_basics_test — bash — 80x24
python_basics_test $ pwd
/Users/llorenz/Desktop/python_basics_test
python_basics_test $ ls
hello
python_basics_test $ cat hello
hello!

greetings!

bonjour!

konnichiha!

ni hao!

aloha!

hola!

ciao!

czesc!

hodi!
```

# SHELL/COMMAND LINE/TERMINAL

## head

“display first lines of a file”



cat filename -first 10

```
python_basics_test $ pwd
/Users/llorenz/Desktop/python_basics_test
python_basics_test $ ls
hello
python_basics_test $ head hello
hello!

greetings!

bonjour!

konnichiha!

ni hao!

python_basics_test $
```

# SHELL/COMMAND LINE/TERMINAL

## tail

“display the last part of a file”



cat filename -last 10

```
python_basics_test $ pwd
/Users/llorenz/Desktop/python_basics_test
python_basics_test $ ls
hello
python_basics_test $ tail hello

czesc!

hodi!

namaste!

marhabaan!

zdraveite!
python_basics_test $
```



# CAT/HEAD/TAIL EXAMPLE

```
Windows PowerShell
PS C:\Users\llorenz> cat .\hello
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
PS C:\Users\llorenz> cat .\hello -first 10
1
2
3
4
5
6
7
8
9
10
PS C:\Users\llorenz> cat .\hello -last 10
6
7
8
9
10
11
12
13
14
15
PS C:\Users\llorenz> _
```



# EXPLORING PYTHON

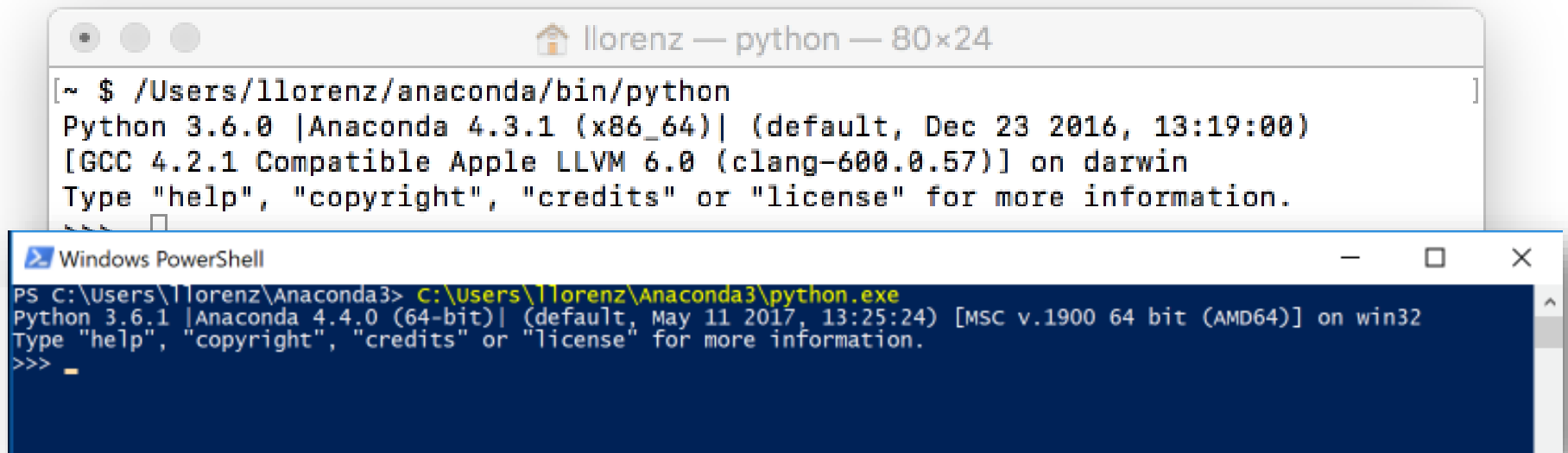
# RUNNING PYTHON INTERACTIVELY FROM THE COMMAND LINE

If you installed on MacOS, or on Windows having selected to add Anaconda to your PATH, just type `python`

Otherwise...

# RUNNING PYTHON INTERACTIVELY FROM THE COMMAND LINE

If you installed on MacOS, or on Windows having selected to add Anaconda to your PATH, just type python



The image shows two terminal windows. The top window is a macOS terminal titled 'llorenz — python — 80x24'. It shows the command `/Users/llorenz/anaconda/bin/python` being executed, resulting in the output: `Python 3.6.0 |Anaconda 4.3.1 (x86_64)| (default, Dec 23 2016, 13:19:00) [GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin`. The bottom window is a Windows PowerShell window titled 'Windows PowerShell'. It shows the command `C:\Users\llorenz\Anaconda3\python.exe` being executed, resulting in the output: `Python 3.6.1 |Anaconda 4.4.0 (64-bit)| (default, May 11 2017, 13:25:24) [MSC v.1900 64 bit (AMD64)] on win32`.

```
[~ $ /Users/llorenz/anaconda/bin/python
Python 3.6.0 |Anaconda 4.3.1 (x86_64)| (default, Dec 23 2016, 13:19:00)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> ]
```

```
Windows PowerShell
PS C:\Users\llorenz\Anaconda3> C:\Users\llorenz\Anaconda3\python.exe
Python 3.6.1 |Anaconda 4.4.0 (64-bit)| (default, May 11 2017, 13:25:24) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> _
```

Remember to use the path where YOU installed Anaconda!  
It may be different from mine!