

DataScience for Development and Social Change, 2015

Tools

What you need to get started

Your Toolset

- ❖ Terminal window
- ❖ Text editor (Sublime Text)
- ❖ Version control (Git) and repository (Github)
- ❖ Visualisation tools (Tableau Public, QGIS)
- ❖ Languages (Python, R, Javascript, D3)

Terminal Window

- ❖ Mac:
 - ❖ applications > utilities > terminal
- ❖ Windows:
 - ❖ Taskbar Start Button > Command Prompt
- ❖ Linux:
 - ❖ applications > accessories > terminal

Terminal: basic commands

- ❖ **pwd**: print current directory name (Windows: **dir**)
- ❖ **cd xxx**: change to directory xxx
- ❖ **cd ~**: go to your 'home' directory
- ❖ **ls**: list files in directory (Windows: **dir**)
- ❖ **ls -al**: list *all* files in directory (including hidden ones)
- ❖ **mkdir xxx**: create a directory called xxx

Terminal: create a work directory

- ❖ Coders often put their code into a folder called “workspace”

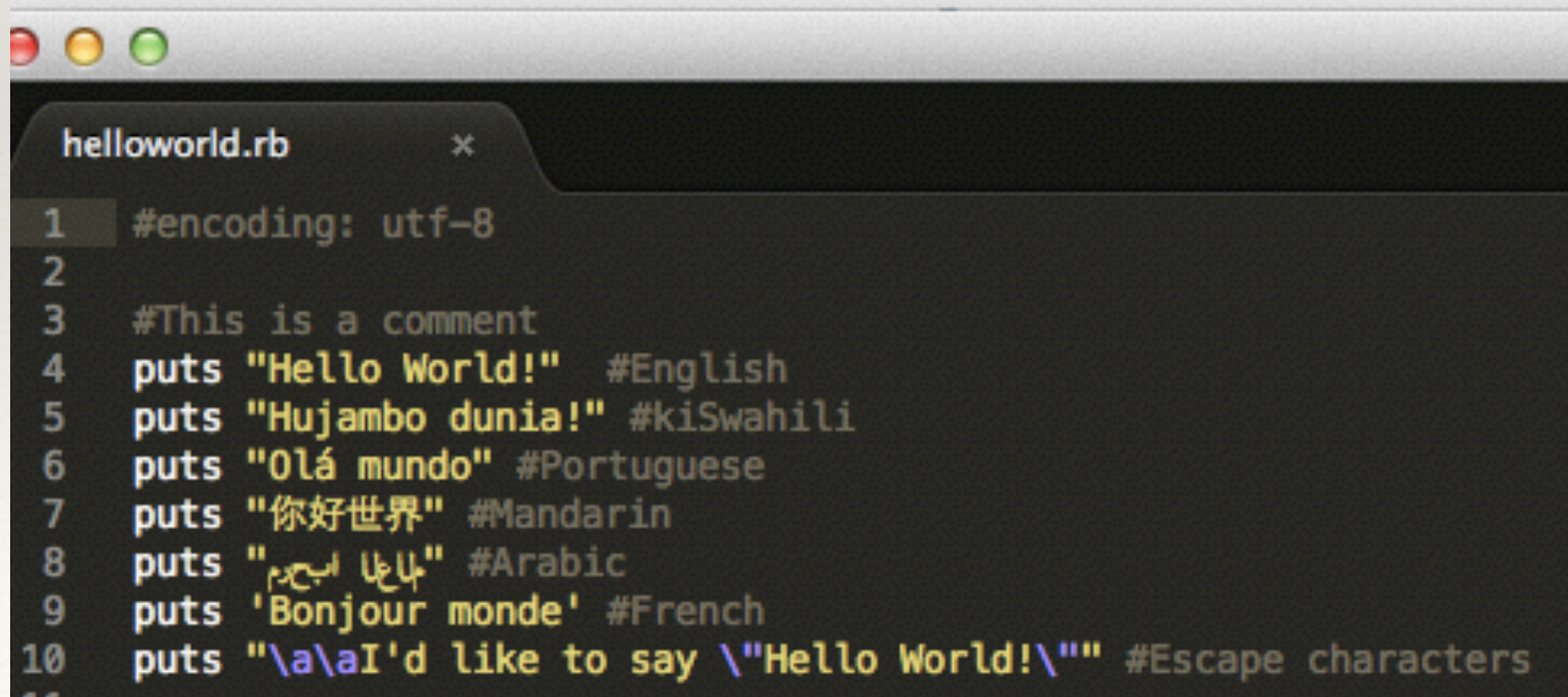
cd ~

mkdir workspace

- ❖ It doesn't matter where you put your code, but make sure you can find it again!
- ❖ e.g. I work in Python and Php on a Mac, so all my files are in /Users/sara/Sites

Text Editor

- ❖ You need a text editor that highlights your code
- ❖ We're using SublimeText in these lectures:
- ❖ <http://www.sublimetext.com>



```
helloworld.rb *  
1 #encoding: utf-8  
2  
3 #This is a comment  
4 puts "Hello World!" #English  
5 puts "Hujambo dunia!" #kiSwahili  
6 puts "Olá mundo" #Portuguese  
7 puts "你好世界" #Mandarin  
8 puts "مرحبا بالعالم" #Arabic  
9 puts 'Bonjour monde' #French  
10 puts "\a\aI'd like to say \"Hello World!\"" #Escape characters
```

Git and Github

❖ Git

- ❖ Version Control System (VCS)
- ❖ Logs changes to files
- ❖ Can “roll back” to a previous version of a file or project

❖ Github

- ❖ Remote repository
- ❖ Online versions of your project
- ❖ Share code with other people
- ❖ Log changes to files (what, who, when)

Git/Github: getting started

- ❖ Get a github account
 - ❖ <https://github.com/>
- ❖ Install git
 - ❖ <http://git-scm.com/downloads>
 - ❖ *not* the GUI version (although that could be useful to you later)
- ❖ **OSX Snow leopard:** install git version 1.7.5 from <https://code.google.com/p/git-osx-installer/downloads/list>

<http://git-scm.com/book/en/Getting-Started-Installing-Git>

Github - getting your course notes

- ❖ In the terminal window:

- ❖ cd to the directory that you want your notes in, then type:

- `git clone https://github.com/bodacea/datasciencefordevelopment`**

- ❖ You should see a new directory appear, containing the files from the repo

- ❖ If the online course notes change, you can get the new files from the terminal window:

- ❖ cd into the new directory, then type:

- `git pull`**

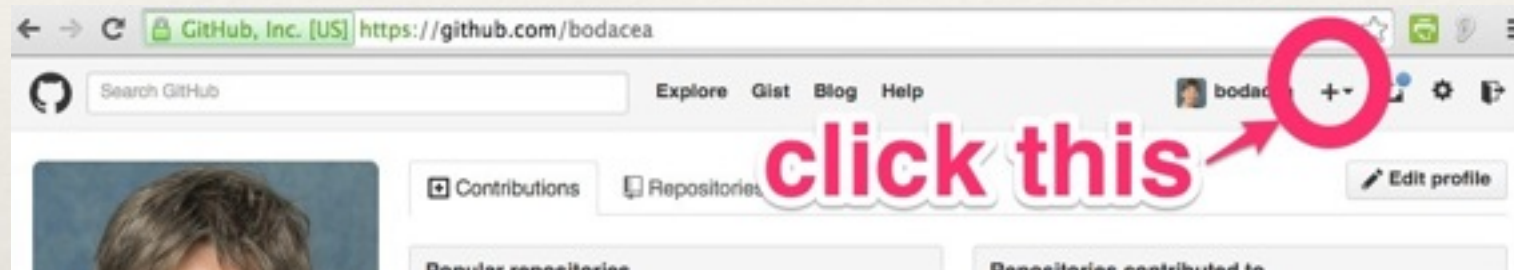
Github – some repos to check out

- ❖ Human development:
 - ❖ **UNOCHA**: <https://github.com/OCHA-DAP>
 - ❖ **UNHCR**: <https://github.com/unhcr>
 - ❖ **US State Dept Humanitarian Info Unit**: <https://github.com/state-hiu>
 - ❖ **Humanitarian OSM**: <https://github.com/hotosm>
 - ❖ **Ushahidi** <https://github.com/ushahidi>
 - ❖ **Sahana** <https://github.com/sahana>
 - ❖ **OpenStreetMap website** <https://github.com/openstreetmap/openstreetmap-website>

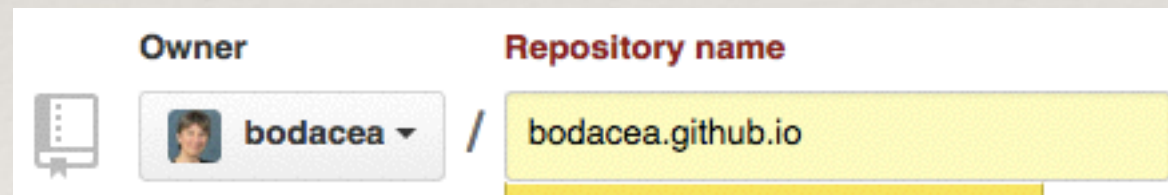
Your D3 portfolio: GitHub.io

- ❖ Github side:

- ❖ go to <https://github.com/yourusername>. Click the “+”.



- * Select “new repository”. Create a repository called yourusername.github.io



- * Click “create repository”

(Details: <https://pages.github.com/>)

Your D3 portfolio: GitHub.io

- * In the terminal window, cd to the directory where you want this repository to live:
 - * `git clone https://github.com/yourusername/yourusername.github.io`
 - * `cd yourusername.github.io`
 - * `echo "Hello World" > index.html`
 - * `git add --all`
 - * `git commit -m "Initial commit"`
 - * `git push -u origin master`
- * Go to <https://github.com/yourusername/yourusername.github.io>
 - * You'll see the file you created (index.html) there
- * Go to <https://bodacea.github.io/>
 - * You'll see "Hello World". Github.io has opened the index.html file for you. You'll use the same mechanism to run D3 files later on.

Installing Tableau Public

- ❖ Go to <http://public.tableau.com/>
- ❖ Click “Download The App”
- ❖ Start Tableau Public

Mac OSX10.6 users: we know Tableau doesn't support you. We're working on it.

Installing R

- ❖ Go to <http://www.r-project.org>
- ❖ Click “download R”

Installing Python

- ❖ Go to <https://www.python.org/downloads/>
- ❖ Click “download Python 2.7.x”
- ❖ (Using Python 2.7 not Python 3.x because libraries!)

- ❖ **Windows users:**
 - ❖ **Use the 32-bit version** of Python, not the 64-bit one (because libraries)
 - ❖ **Set your Path environment variable** after you install Python:
 - ❖ From your desktop, open any folder. Look at the list on the left of the folder: right-click any folder, then click “properties” then “advanced” then “environment variables”.
 - ❖ Find “path” in the list of variables. If you installed python into C:/Python27 (the default setting), double-click on “path” and add this to the path string: “;C:/Python27;C:/Python27/Scripts”.
 - ❖ Open up a new terminal window. Type “python” in it: Python should start. Type “exit()” to get out of python. In the terminal window, type “pip install xlrd”... pip should start installing.

Using Python pre-2.7.9? You might also need pip: <http://pip.readthedocs.org/en/latest/installing.html#install-pip>

Adding Python Libraries

- ❖ Libraries contain reusable code.
- ❖ To add a library, type “pip install libraryname” (without the “s”) in the terminal window... whilst you’re connected to the Internet.
- ❖ Add these libraries:
 - ❖ xlrd
 - ❖ xlwt
 - ❖ requests
 - ❖ BeautifulSoup
 - ❖ scrapy
 - ❖ ipython
 - ❖ numpy
 - ❖ scipy
 - ❖ matplotlib
 - ❖ pandas
 - ❖ nltk
 - ❖ shapely

Installing QGIS

- ❖ Go to <http://www.qgis.org>
- ❖ Windows users: click “download now”
- ❖ Mac users: this is a little more complicated...

Installing D3

- ❖ Get D3:
 - ❖ Go to <http://d3js.org>
 - ❖ Click on “d3.zip”
- ❖ Check that you have javascript enabled in your browser:
<http://www.enable-javascript.com>

Done?

- ❖ Start looking at datasets and visualizations
- ❖ Start thinking about what you'd like to build

If all the installs fail: <http://learn.adicu.com/setup/>