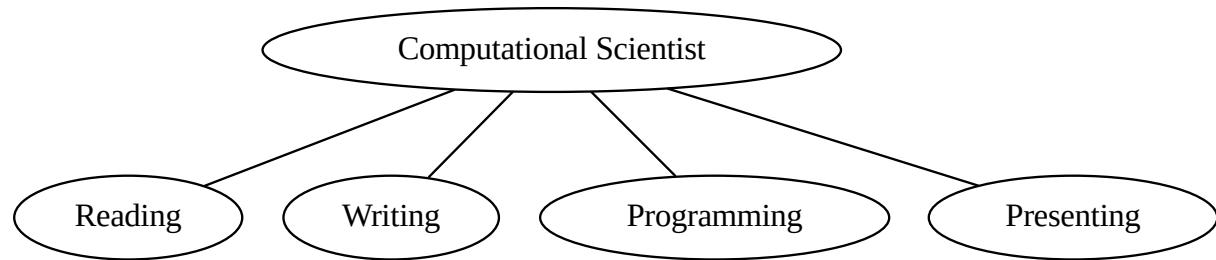


Tools of the Computational Scientist



Omri Har-Shemesh PhD Candidate Computational Science Lab

[https://github.com/omrihar/tools_computational_science_2016]



Reading



Reading

- Library (<http://uba.uva.nl/en> (<http://uba.uva.nl/en>))
- Reference Managers:
 - Mendeley (<http://www.mendeley.com>)
 - Zotero (<http://www.zotero.org>)
 - ReadCube (<http://www.readcube.com>)
 - Paperpile (<http://www.paperpile.com>)
 - ...
- Google Scholar (<http://scholar.google.com>)
- Web of science (https://apps.webofknowledge.com/WOS_GeneralSearch_input.cgi?product=WOS&search_mode=GeneralSearch&SID=P2pH5das4FHHHYqVLVM)

Writing



Writing

- Text Editor
 - Vim
 - Emacs
 - Sublime Text 3
 - Atom.io
 - PyCharm
 - ...
- Jupyter Notebook!

Writing

- Publishing
 - Markdown + Pandoc
 - LaTeX
 - Overleaf (<https://www.overleaf.com> (<https://www.overleaf.com>))
 - Authorea (<https://www.authorea.com> (<https://www.authorea.com>))
 - shareLaTeX (<https://www.sharelatex.com> (<https://www.sharelatex.com>))
 - Google Docs
 - Microsoft Word
 - ...

```
In [4]: IFrame("http://commonmark.org/help/", "100%", 600)
```

Out[4]:

Markdown is a simple way to format text that looks great on any device. It doesn't do anything fancy like change the font size, color, or type – just the essentials, using keyboard symbols you already know.

TRY OUR 10 MINUTE MARKDOWN TUTORIAL

Type	... to Get
Italic	<i>Italic</i>
Bold	Bold
# Heading 1	<h1>Heading 1</h1>
## Heading 2	<h2>Heading 2</h2>
[Link](http://a.com)	Link
! [Image](http://url/a.png)	

```
In [5]: IFrame("http://pandoc.org/", "100%", 600)
```

```
Out[5]:
```

Pandoc a universal document converter

About pandoc

If you need to convert files from one markup format into another, pandoc is your swiss-army knife. Pandoc can convert documents in [markdown](#) (<http://daringfireball.net/projects/markdown/>), [reStructuredText](#) (<http://docutils.sourceforge.net/docs/ref/rst/introduction.html>), [textile](#) (<http://redcloth.org/textile>), [HTML](#) (<http://www.w3.org/TR/html40/>), [DocBook](#) (<http://www.docbook.org/>), [LaTeX](#) (<http://www.latex-project.org/>), [MediaWiki markup](#) (<http://www.mediawiki.org/wiki/Help:Formatting>), [TWiki markup](#) (<http://twiki.org/cgi-bin/view/TWiki/TextFormattingRules>), [OPML](#) (<http://dev.opml.org/spec2.html>), [Emacs Org-Mode](#) (<http://orgmode.org>), [Txt2Tags](#) (<http://txt2tags.org/>), Microsoft Word [docx](#) (<http://www.microsoft.com/interop/openup/openxml/default.aspx>), LibreOffice [ODT](#) (<http://en.wikipedia.org/wiki/OpenDocument>), [EPUB](#) (<http://en.wikipedia.org/wiki/EPUB>), or [Haddock markup](#) (<http://www.haskell.org/haddock/doc/html/ch03s08.html>) to

- HTML formats: XHTML, HTML5, and HTML slide shows using [Slidy](#) (<http://www.w3.org/Talks/Tools/Slidy>), [reveal.js](#) (<http://lab.hakim.se/reveal-js/>), [Slideous](#) (<http://goessner.net/articles/slideoous/>), [S5](#) (<http://meyerweb.com/eric/tools/s5/>), or [DZSlides](#) (<http://paulrouget.com/dzslides/>).

Writing

- Publishing
 - Markdown + Pandoc
 - LaTeX
 - Overleaf (<https://www.overleaf.com> (<https://www.overleaf.com>))
 - Authorea (<https://www.authorea.com> (<https://www.authorea.com>))
 - shareLaTeX (<https://www.sharelatex.com> (<https://www.sharelatex.com>))
 - Google Docs
 - Microsoft Word
 - ...

```
In [6]: IFrame("http://www.nature.com/news/scientific-writing-the-online-cooperative-1.16039", "100%", 600)
```

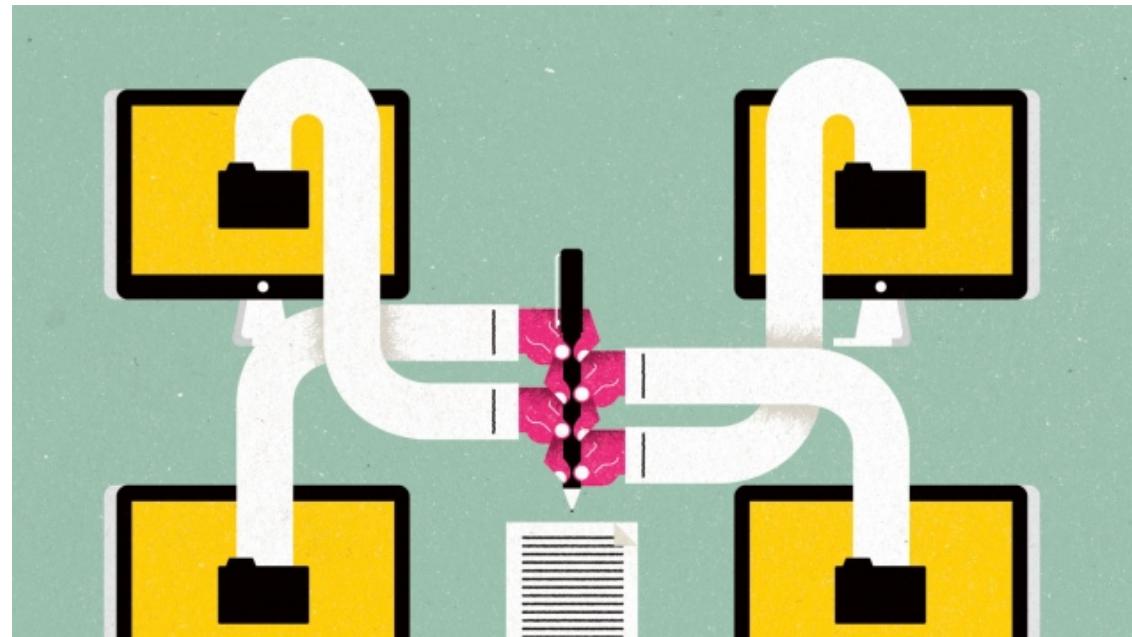
Out[6]: [NATURE | TOOLBOX](#)

Scientific writing: the online cooperative

Collaborative browser-based tools aim to change the way researchers write and publish their papers.

Jeffrey M. Perkel

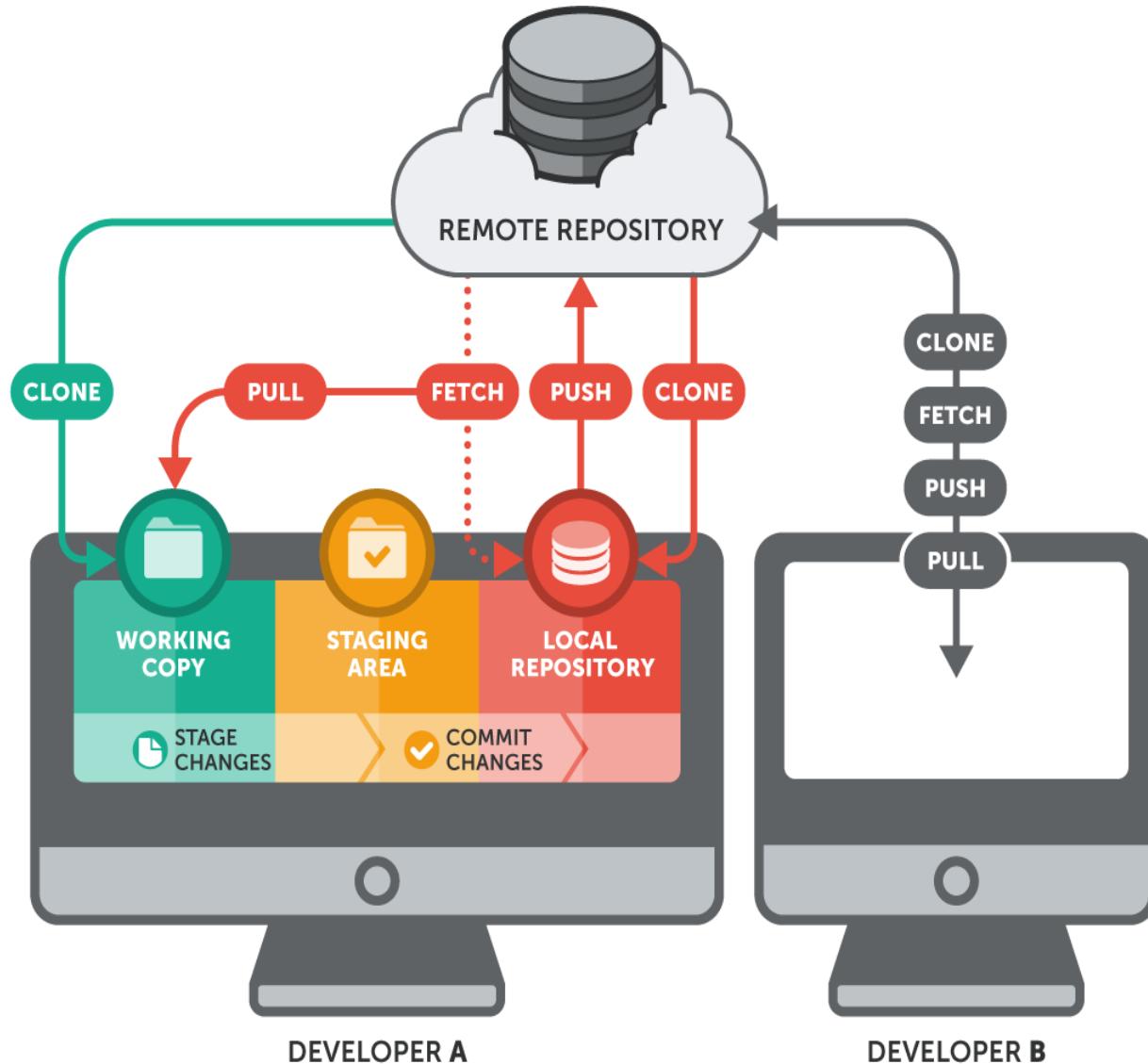
01 October 2014 Clarified: 06 October 2014



Programming



First of all - Version Control (Git)!



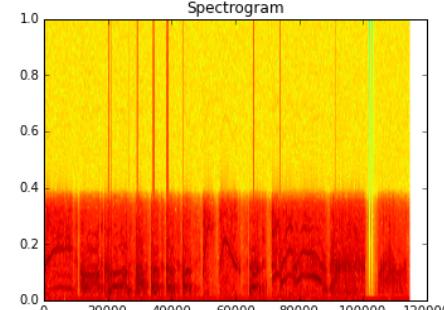
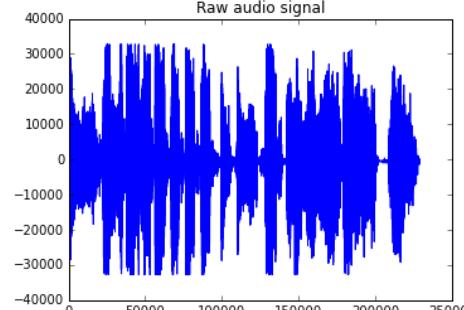
Jupyter Notebook

Jupyter spectrogram Last Checkpoint: an hour ago (autosaved) 

File Edit View Insert Cell Kernel Help

Cell Toolbar: None

```
In [1]: from scipy.io import wavfile  
rate, x = wavfile.read('test_mono.wav')  
  
In [2]: import matplotlib.pyplot as plt  
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))  
ax1.plot(x); ax1.set_title('Raw audio signal')  
ax2.specgram(x); ax2.set_title('Spectrogram')  
plt.show()
```



```
In [7]: import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns

# Obtain the Anscombe dataset
df = sns.load_dataset("anscombe")

df.head()
```

Out[7]:

	dataset	x	y
0	I	10.0	8.04
1	I	8.0	6.95
2	I	13.0	7.58
3	I	9.0	8.81
4	I	11.0	8.33

```
| dataset |
```

```
In [8]: df.groupby(df.dataset).describe(percentiles=[]).loc[["I", "II"]]
```

Out[8]:

		x	y
dataset			
I	count	11.000000	11.000000
	mean	9.000000	7.500909
	std	3.316625	2.031568
	min	4.000000	4.260000
	50%	9.000000	7.580000
	max	14.000000	10.840000
II	count	11.000000	11.000000
	mean	9.000000	7.500909
	std	3.316625	2.031657
	min	4.000000	3.100000
	50%	9.000000	8.140000
	max	14.000000	9.260000

dataset			
---------	--	--	--

```
In [9]: df.groupby(df.dataset).describe(percentiles=[]).loc[["III", "IV"]]
```

Out[9]:

		x	y
dataset			
III	count	11.000000	11.000000
	mean	9.000000	7.500000
	std	3.316625	2.030424
	min	4.000000	5.390000
	50%	9.000000	7.110000
	max	14.000000	12.740000
IV	count	11.000000	11.000000
	mean	9.000000	7.500909
	std	3.316625	2.030579
	min	8.000000	5.250000
	50%	8.000000	7.040000
	max	19.000000	12.500000

dataset | | |

```
In [10]: df.groupby(df.dataset).agg(['mean', 'var'])  
df.groupby(df.dataset).corr()
```

Out[10]:

		x	y
dataset			
I	x	1.000000	0.816421
	y	0.816421	1.000000
II	x	1.000000	0.816237
	y	0.816237	1.000000
III	x	1.000000	0.816287
	y	0.816287	1.000000
IV	x	1.000000	0.816521
	y	0.816521	1.000000


```
In [11]: # Show the results of a linear regression within each dataset
sns.lmplot(x="x", y="y", col="dataset", hue="dataset", data=df,
            col_wrap=2, ci=None, palette="muted", size=4,
            scatter_kws={"s": 50, "alpha": 1})
plt.show()
```

```
In [12]: # Let's build a dashboard!

from IPython.display import display, Javascript
import ipywidgets as widgets

L = widgets.Label("Hello World")
display(L)
```

```
In [13]: L.value = "Howdy World"
```

```
In [14]: from bqplot import Figure, Scatter, LinearScale, Axis

x = df[df.dataset == "I"].x
y = df[df.dataset == "I"].y
x_sc = LinearScale()
y_sc = LinearScale()
ax_x = Axis(label="", scale=x_sc)
ax_y = Axis(label="", scale=y_sc, orientation='vertical', side='left')

scatter = Scatter(x=x, y=y, scales={'x': x_sc, 'y': y_sc})

fig = Figure(marks=[scatter], axes=[ax_x, ax_y], title='Dataset
I',
            animation_duration=200)
```

```
In [15]: display(fig)
```

```
In [16]: scatter.y = range(11)
```

```
In [17]: # Let's add an option to select displayed dataset...
datasets = df.dataset.unique()
datasets
```

```
Out[17]: array(['I', 'II', 'III', 'IV'], dtype=object)
```

```
In [18]: # Create the select widget
select_dataset = widgets.Dropdown(options=list(datasets))
display(select_dataset)
```

```
In [19]: # Code to run when a selection is made
def _on_select(obj):
    selected = obj['new']
    scatter.x = df[df.dataset == selected].x
    scatter.y = df[df.dataset == selected].y
    fig.title = "Dataset %s" % selected

select_dataset.observe(_on_select, 'value')
```

```
In [20]: # Putting it all together
from ipywidgets import HBox, VBox
display(VBox([fig, HBox([widgets.Label("Select Dataset: "), select_dataset])))
```

Presenting

- Powerpoint
- Prezi
- LaTeX (Beamer class, for example)
- Jupyter Notebook! (These slides, for example)

Take home messages

1. Use a reference manager!
2. Use a version control system (git)!
3. Learn Markdown.
4. Python Python Python.
5. Jupyter Notebook

Thank you very much for your attention!