

http://ipython.org/static/IPyheader.png

<u>IPython</u> in action creating reproducible and publishable interactive work.

What is this?

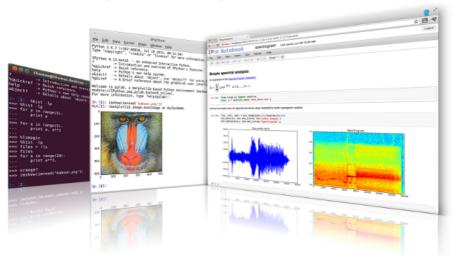
This repo contains the complete <u>talk</u> I intend to deliver (have delivered) at <u>PyConZA2013</u>. It contains all the files needed to build a final publishable PDF document from an interactive notebook and even adds a custom front page.

The Complete Talk GitHub Website can be accessed here

Background

IPython had become a popular choice for doing interactive scientific work. It extends the standard Python interpreter and adds many useful new futures. There is really no need to use the standard Python interpreter anymore. In addition to this IPython offers a web based Notebook that makes interactive work much easier, and have been used to write repeatable scientific papers and more recently a book has been written using this platform, the online Notebook Viewer and GitHub. The development of this material and tool chain to compile the notebook to a publishable PDF, has inspired me to maybe even try and turn this into a complete (free) book. Let's see what happens.

Combining the most common scientific packages with IPython makes it a formidable tool and serious competition to R. (R is still awesome!)



http://ipython.org/static/ipy0.13.png

As a matter of fact you can run R in the notebook session, embed YouTube Videos, Images and lots more but let me not get ahead of myself....

The science stack consists of (but not limited to):

package	description	
<u>pandas</u>	dataframe implementation (based on numpy)	

scipy	efficient numerical routines	
sympy	symbolic mathematics	
<u>matplotlib</u>	python standard plotting package	
sci-kit learn	machine learning and well documented!	

Talk contents

The talk will aim to introduce these tools and explore some practical interactive examples. Once completed it will be shown how easy it is to publish your work to various formats. Some of the topics covered in the talk are listed below:

item	description	
ipython	quick intro to ipython and the notebook	
setup	set up your environment / get the talk files	
notebook basics	navigate the notebook	
notebook magic's	special notebook commands that can be very useful	
getting input	as from IPython 1.00 getting input from sdtin is possible	
local files	how to link to local files in the notebook directory	
plotting	how to create beautiful inline plots	
symbolic math	quick demo of sympy model	
pandas	quick intro to pandas dataframe	
typesetting	include markdown, Latex via MathJax	
loading code	how to load a remote .py code file	
gist	paste some of your work to gist for sharing	
js	some javascript examples	
customising	loading a customer css and custom matplotlib config file	
git cell	add code to a special cell that would commit to git	
output formats	tput formats how to publish your work to html, pdf or jeveal.js presentation	

Get the processed presentation files here:

format	description
IPython notebook	.ipynb file to run in browser
IPython html notebook	converted to HTML and served online
IPython pdf notebook	converted to PDF for download (to be added, needs pandoc)
IPython pdf book	converted to pdf and a front-page stitched to it)
Ipython reveal.js presentation	converted to a reveal.js presentation and served online

Online IPython NBveiwer	view on the ipython notebook viewer
-------------------------	-------------------------------------

Dependencies

I was given the challenge to develop all of this on a Windows machine as some of my sponsors want to demonstrate that this stuff can not only be done on GNU/Linux/OSX. So all the tool chains are Windows based. If you know Linux, then you are the type of person that would easily port this. That being said the Windows GitHub client is refreshing. I have also added a MacBook Air to my arsenal and have been porting the toolchain to Mac aswell and it seems to be working fine.

package	description		
IPython	To use NBConvert you need V1.00. If you only want to use the interactive notebook then v0.13 will be ok.		
pandoc	The document converter used by IPythonr		
MikeTex	If you want to do a TEX to PDF transform. I had so many issues with the TEX to PDF conversion by NBConvert, so settled for wkhtmltopdf(below) to convert HTML to PDF rather. (Convert notebook to HTML with NBconvert and then from HTML to PDF with wkhtmltopdf		
wkhtmltopdf Convert HTML to PDF (i could only install this on windows			
wkpdf	I couldn't get wkhtmltopdf to work on os x so i installed wkpdf for handling the HTML to PDF conversion on my Mac. It's a Ruby Gem install and painless.		
pdftk	Can be used to combine PDF's. In this case add a frontpage to the generated IPython notebook PDF. Only available for Windows.		
ImageMagick for compressing the PDF. Still experimenting with this.(have not got this working yet so not needed)GhostScript	needed by ImageMagick(not needed as PDF compression is not functional yet)		
anaconda	install anaconda from Continuum Analytics. Almost all the Python packages are included and it has a virtual environment manager via it's console application `conda'		

How to run the Interactive Notebook

Navigate to the src directory and run from the command line:

python ipython notebook

If everything works your browser should open and you can select the notebook and start experimenting!

PDF, HTML, Slideshow Build Script

There is a build script in the src directory. It is an IPython file. You can basically build shell scripts this way. To use the power of IPython commands save the file with the .ipy extension and call it with IPython. Even the magic's work. To build the document use ipython builddocs.ipy You will have to change the paths to the software however. Currently I can use the build script on Windows and on my Mac but it is a bit of a hack.

Cross Platform Output Rendering

I have tested the HTML outputs on my Galaxy S3 and S4, IPAD and Nexus7. They render very well. Even the downloaded PDF was easily readable on the NEXUS 7 in landscape mode. In conclusion the produces work is really very well packaged and easily consumed on most platforms. This is not bad, and all done with open source software.

Some interesting links

- A book written with IPython Notebook
- Notebook Viewer
- Anaconda Installing almost everything you need

About the presenter

- I am an Electrical Engineer and is currently working for a <u>consulting firm</u> where I manage the Business Analytics and Quantitative Decision Support Services division.
- I use python in my day to day work as a practical alternative to the limitations of EXCEL in using large data sets.
- LinkedIn
- I am also a co-founder at House4Hack

The IPython notebook

The IPython notebook is part of the IPython project. The IPython project is one of the packacges making up the python scientific stack called SciPi. SciPy (pronounced "Sigh Pie") is a Python-based ecosystem of open-source software for mathematics, science, and engineering. In particular, these are some of the core packages:



NumPy Base N-dimensional array package



SciPy library Fundamental library for scientific computing



Matplotlib Comprehensive 2D Plotting



IPython Enhanced Interactive Console



SympySymbolic mathematics



pandas Data structures & analysis

SciPy

Quick IPython introdution

IPython provides a rich architecture for interactive computing with:

- Powerful interactive shells (terminal and Qt-based).
- A browser-based notebook with support for code, text, mathematical expressions, inline
 plots and other rich media.
- Support for interactive data visualization and use of GUI toolkits.
- Flexible, embeddable interpreters to load into your own projects.
- Easy to use, high performance tools for parallel computing.

The main reasons I have been using it includes:

- A superior shell
- Plotting is possible in the QT console or the Notebook
- the magic functions makes life easier (magics gets called with a %, use %-tab to see them all)
- I also use it as a replacement shell for Windows Shell or Terminal
- Code Completion
- GNU Readline based editing and command history

Some helpfull commands

The four most helpful commands, as well as their brief description, is shown to you in a banner, every time you start IPython:

command	description	
?	Introduction and overview of IPython's features.	
%quickref	Quick reference.	
help	Python's own help system.	
object?	Details about 'object', use 'object??' for extra details.	

Some imports and settings

The following code cells make sure that plotting is enabled and also loads a customised matplotlib confirguration file that spices up the inline plots. The custom matplotlib file has been taken from the Bayesian Methods for Hackers Project

```
In [8]: # makes sure inline plotting is enabled
%pylab --no-import-all inline
```

Populating the interactive namespace from numpy and matplotlib

```
In [2]: #loads a customer marplotlib configuration file

def CustomPlot():
    import json
    s = json.load( open("static/matplotlibrc.json") )
    matplotlib.rcParams.update(s)
    figsize(18, 6)
```

Changing the notebook layout

The code cell below is an example of how you should not be chaning the layout and css of the notebook. From IPython V1.00 it is possible to include custom css by creating IPython profiles. Since this file needs to be distributable I have opted for the hack below as used by the Bayesian Methods for Hackers Team

```
In [53]: from IPython.core.display import HTML
    def css_styling():
        styles = open("static/custom.css", "r").read()
        return HTML(styles)
        css_styling()
Out[53]:
```

Notebook basics

The IPython Notebook is a web-based interactive computational environment where you can combine code execution, text, mathematics, plots and rich media into a single document.

- Code Completion
- Help
- Docstrings
- Markdown cells
- Running a Code cell (Shift+Enter)
- Setting a cell to be included in the presentation

Run the contents of a cell

- SHIFT+ENTER will run the contents of a cell and move to the next one
- CTRL+ENTER run the cell in place and don't move to the next cell. (best for presenting)
- CTRL-m h show keyboard shorcuts

```
In [56]: # press shift-enter to run code
    print "Hallo Pycon"
Hallo Pycon
```

Save the notebook

CTRL-S will save the notebook

Lets get some help

The %quickref commmand can be used to obtain a bit more information

```
In [5]: #IPython -- An enhanced Interactive Python - Quick Reference Card 
%quickref # now press shift-ender
```

Code completion and introspection

The cell below defines a function with a bit of a long name. By using the ? command the docstring can we viewed. ?? will open up the source code. The autocomplete function is also demostrated, and for fun the function is called and the output displayed

```
In [40]: # lets degine a function with a long name.
         def long_silly_dummy_name(a, b):
            This is the docstring for dummy.
            It takes two arguments a and b
            It returns the sum of a and b
             No error checking is done!
             return a+b
In [41]: # lets get the docstring or some help
         long_silly_dummy_name?
In [42]: long_silly_dummy_name??
In [67]: #press tab to autocplete
         long_silly_dummy_name
Out[67]: <function __main__.long_silly_dummy_name>
 In [8]: # press shift-enter to run
         long_silly_dummy_name(5,6)
 Out[8]: 11
```

Setting up the notebook to enable a slideshow view

You need to activate the Cell Toolbar in the Toolbar above. Here you can set if this cell should be compiled as a slide or not. The options are given below:

- slide
- sub slide
- fragment
- skip

Using markdown

You can set the contents type of a cell in the toolbar above. When Markdown is selected you can enter markdown in a cell and it's contents will be rendered as HTML. The markdown syntax can by <u>found here</u>

This is heading 1

This is heading 2

This is heading 5

Beautiful is better than ugly. Explicit is better than implicit. Simple is better than complex. Complex is better than complicated.

Notebook magics

IPython has a set of predefined 'magic functions' that you can call with a command line style syntax. There are two kinds of magics, line-oriented and cell-oriented. Line magics are prefixed with the % character and work much like OS command-line calls: they get as an argument the rest of the line, where arguments are passed without parentheses or quotes. Cell magics are prefixed with a double %%, and they are functions that get as an argument not only the rest of the line, but also the lines below it in a separate argument.

Timeit magic

The timeit magic can be used to evaluate the average time your loop or piece of code is taking to complete it's run.

Know when the kernel is busy

Have a look at the top right hand side of the notebook and run the code cell above again. This shows that the kernel is busy running the current cell.

User input

In the snippet below it the raw_input() function is used to read some user input to a variable raw and printed to stdout.

```
In [13]: from IPython.display import HTML
    raw = raw_input("enter your input here >>> ")
    print "Hallo, ",raw
    enter your input here >>> world!
    Hallo, world!
```

How to link to the filesystem

```
In [25]: from IPython.display import FileLink, FileLinks
         FileLinks('.', notebook_display_formatter=True)
Out[25]: ./
           .DS Store
           builddocs.ipy
           pycon13_ipython.ipynb
           README.md
         ./.ipynb_checkpoints/
           pycon13_ipython-checkpoint.ipynb
           CapeTown_2009_Temperatures.csv
           READEME.md
         ./output/
           .DS_Store
           pycon13_ipython.html
           pycon13_ipython.slides.html
           pycon13_ipython_complete.pdf
           pycon13_ipython_pdf.pdf
         ./static/
           .DS_Store
           custom.css
           frontpage.docx
           frontpage.pdf
           ip.png
           ip2.png
           matplotlibrc.json
           python-vs-java.jpg
           scistack.png
```

Running shell commands

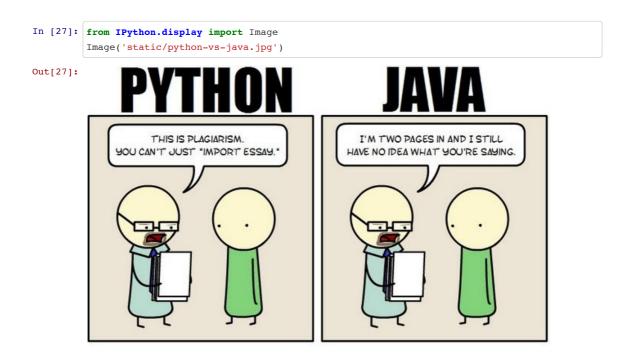
I now use ipython as my default shell scripting language. lets put the contents of the current directory into a list. by using the ! before a command indicates that you want to run a system command.

```
In [17]: filelist = !ls  #read the current directory into variable
for x,i in enumerate(filelist):
    print '#',x, '--->', i

# 0 ---> README.md
# 1 ---> builddocs.ipy
# 2 ---> data
# 3 ---> output
# 4 ---> pycon13_ipython.ipynb
# 5 ---> static
```

Embedding Images

Image released under CC BY-NC-ND 2.5 IN) by Rhul Singh



Adding YouYube videos

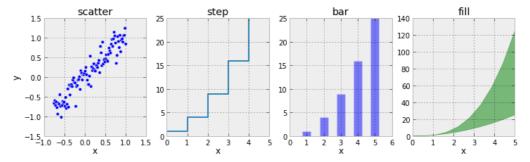
I am making the video small as it does not embed into the final output pdf.

```
In [28]: from IPython.display import YouTubeVideo
    YouTubeVideo('iwVvqwLDsJo', width=200, height=200)
Out[28]:
```

Plotting with Matplotlib

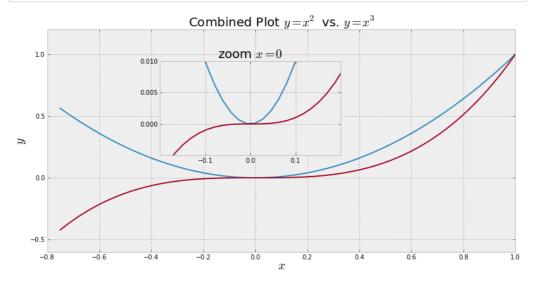
matplotlib is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. matplotlib can be used in python scripts, the python and ipython shell, web application servers, and six graphical user interface toolkits.

```
In [6]: CustomPlot()
        from numpy import *
        #generate some data
        n = array([0,1,2,3,4,5])
           = np.linspace(-0.75, 1., 100)
          = linspace(0, 5, 10)
        fig, axes = plt.subplots(1, 4, figsize=(12,3))
        axes[0].scatter(xx, xx + 0.25*randn(len(xx)))
        axes[0].set_title('scatter')
        axes[1].step(n, n**2, lw=2)
        axes[1].set_title('step')
        axes[2].bar(n, n**2, align="center", width=0.5, alpha=0.5)
        axes[2].set_title('bar')
        axes[3].fill_between(x, x**2, x**3, color="green", alpha=0.5);
        axes[3].set_title('fill')
        for i in range(4):
            axes[i].set_xlabel('x')
        axes[0].set_ylabel('y')
        show()
```



Combined plots

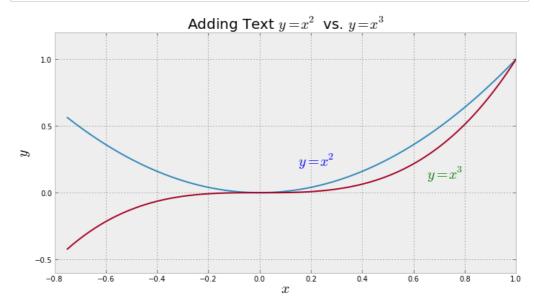
```
In [7]: CustomPlot()
        font size = 20
        figsize(11.5, 6)
        fig, ax = plt.subplots()
        ax.plot(xx, xx**2, xx, xx**3)
        ax.set title(r"Combined Plot $y=x^2$ vs. $y=x^3$", fontsize = font size)
        ax.set xlabel(r'$x$', fontsize = font size)
        ax.set ylabel(r'$y$', fontsize = font size)
        fig.tight_layout()
        # inset
        inset_ax = fig.add_axes([0.29, 0.45, 0.35, 0.35]) # X, Y, width, height
        inset_ax.plot(xx, xx**2, xx, xx**3)
        inset_ax.set_title(r'zoom $x=0$',fontsize=font_size)
        # set axis range
        inset_ax.set_xlim(-.2, .2)
        inset_ax.set_ylim(-.005, .01)
        # set axis tick locations
        inset_ax.set_yticks([0, 0.005, 0.01])
        inset_ax.set_xticks([-0.1,0,.1]);
        show()
```



Adding text to a plot

```
In [13]: CustomPlot()
  figsize(11.5, 6)
  font_size = 20
  fig, ax = plt.subplots()
```

```
ax.plot(xx, xx**2, xx, xx**3)
ax.set_xlabel(r'$x$', fontsize = font_size)
ax.set_ylabel(r'$y$', fontsize = font_size)
ax.set_title(r"Adding Text $y=x^2$ vs. $y=x^3$", fontsize = font_size)
ax.text(0.15, 0.2, r"$y=x^2$", fontsize=font_size, color="blue")
ax.text(0.65, 0.1, r"$y=x^3$", fontsize=font_size, color="green");
```



Symbolic math using SymPy

SymPy is a Python library for symbolic mathematics. It aims to become a full-featured computer algebra system (CAS) while keeping the code as simple as possible in order to be comprehensible and easily extensible. SymPy is written entirely in Python and does not require any external libraries.

```
In [15]: a = 1/x + (x*\sin(x) - 1)/x

Out[15]: \frac{x\sin(x) - 1}{x} + \frac{1}{x}

In [16]: \sin(x)
```

Data analysis using the Pandas library

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive

The two primary data structures of pandas, Series (1-dimensional) and DataFrame (2-dimensional), handle the vast majority of typical use cases in finance, statistics, social science, and many areas of engineering.

For R users, DataFrame provides everything that R's data.frame provides and much more. pandas is built on top of NumPy and is intended to integrate well within a scientific computing environment with many other 3rd party libraries.

```
In [9]: from pandas import DataFrame, read_csv

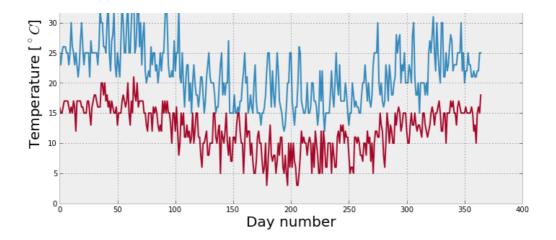
Cape_Weather = DataFrame( read_csv('data/CapeTown_2009_Temperatures.csv' ))
Cape_Weather.head()
```

Out[9]:

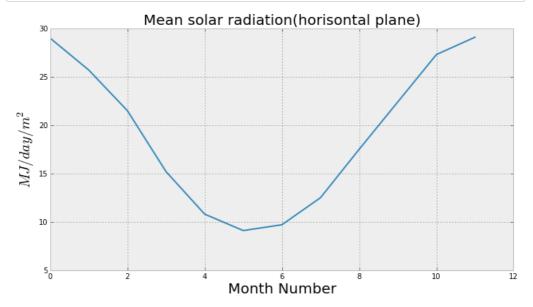
	high	low	radiation
0	25	16	29.0
1	23	15	25.7
2	25	15	21.5
3	26	16	15.2
4	26	17	10.8

```
In [10]: CustomPlot()
    figsize(11.5, 6)
    font_size = 20
    title('Cape Town temparature(2009)',fontsize = font_size)
    xlabel('Day number',fontsize = font_size)
    ylabel(r'Temperature [$^\circ C$] ',fontsize = font_size)
    Cape_Weather.high.plot()
    Cape_Weather.low.plot()
    show()
```





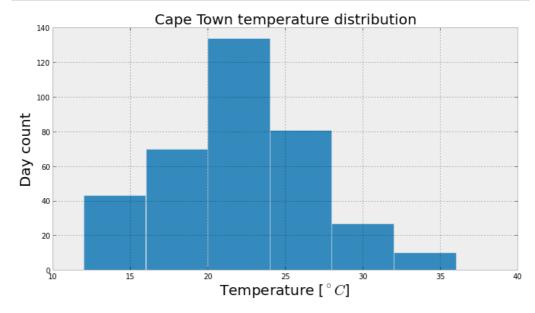
```
In [11]: CustomPlot()
    figsize(11.5, 6)
    font_size = 20
    title( 'Mean solar radiation(horisontal plane)', fontsize=font_size)
    xlabel('Month Number', fontsize = font_size)
    ylabel(r'$MJ / day / m^2$',fontsize = font_size)
    Cape_Weather.radiation.plot()
    show()
```



```
radiation dtype: int64
```

```
In [46]: # Basic descriptive statistics
         print Cape_Weather['high'].describe()
         count
                 365.000000
                 21.545205
         mean
         std
                   4.764943
         min
                  12.000000
         25%
                  18.000000
         50%
                  21.000000
         75%
                  25.000000
                   36.000000
         dtype: float64
```

```
In [12]: CustomPlot()
    figsize(11.5, 6)
    font_size = 20
    title('Cape Town temperature distribution', fontsize=font_size)
    ylabel('Day count',fontsize = font_size)
    xlabel(r'Temperature [$^\circ C $] ',fontsize = font_size)
    Cape_Weather['high'].hist(bins=6)
    show()
```



Typesetting

Latex

Latex is rendered using the mathjax javascript library

```
In [43]: from IPython.display import Math
          Math(r'F(k) = \int_{-\infty}^{\infty} f(x) e^{2\pi i k} dx')
                                                  F(k) = \int_{-\infty}^{\infty} f(x)e^{2\pi ik} dx
Out[43]:
In [44]: from IPython.display import Latex
          Latex(r"""\begin{eqnarray}
          & = \frac{4\pi}{c}\operatorname{mathbf}{j}} \
          \label{eq:continuous} $$ \aligned \ensuremath{E} \ & = 4 \pi \rho \\
          t} & = \sqrt{\{\mathbb{0}\}} \
          \end{eqnarray}""")
Out[44]:
                                                  \nabla \times \vec{\mathbf{B}} - \frac{1}{c} \frac{\partial \vec{\mathbf{E}}}{\partial t} = \frac{4\pi}{c} \vec{\mathbf{j}}
                                                 \nabla \times \vec{\mathbf{E}} + \frac{1}{c} \frac{\partial \vec{\mathbf{B}}}{\partial t} = \vec{\mathbf{0}}
\nabla \cdot \vec{\mathbf{B}} = 0
```

Using the Python Debugger - pdb

```
In [46]: %pdb on
         Automatic pdb calling has been turned ON
In [47]: foo = 1
         bar = 'a'
         print foo+bar
                                                      Traceback (most recent call last)
          <ipython-input-47-08464a413a31> in <module>()
               1 \text{ foo} = 1
               2 bar = 'a'
          ---> 3 print foo+bar
         TypeError: unsupported operand type(s) for +: 'int' and 'str'
         > <ipython-input-47-08464a413a31>(3)<module>()
                1 \text{ foo} = 1
                2 bar = 'a'
          ---> 3 print foo+bar
         ipdb> q
```

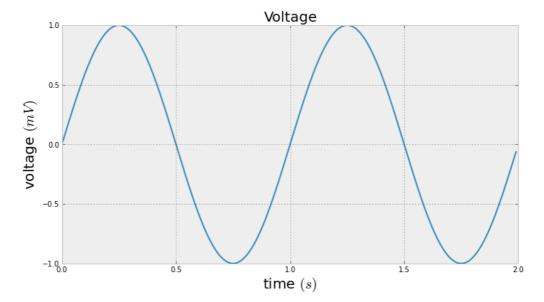
Loading Code Snippets

```
In [13]: %load http://pastebin.com/raw.php?i=mGiV1FwY
```

```
In [14]: CustomPlot()
font_size = 20
figsize(11.5, 6)

t = arange(0.0, 2.0, 0.01)
s = sin(2*pi*t)
plot(t, s)

xlabel(r'time $(s)$', fontsize=font_size)
ylabel('voltage $(mV)$', fontsize=font_size)
title('Voltage', fontsize=font_size)
grid(True)
```



It's in a browser, can it do Javascript?

source

```
javascript = """
         <script type="text/Javascript">
             function set_value(){
                 var var_name = document.getElementById('var_name').value;
                 var var_value = document.getElementById('var_value').value;
                 var command = var_name + " = '" + var_value + "'";
                 console.log("Executing Command: " + command);
                 var kernel = IPython.notebook.kernel;
                 kernel.execute(command);
         </script>
         HTML(input_form + javascript)
Out[66]:
             Variable Name: var
             Variable Value: val
              Set Value
In [54]: PyCon
Out[54]: Roks
```

Saving a Gist

It is possible to save spesific lines of code to a GitHub gist. This is achieved with the pastebin magic as demonstrated below.

```
In [56]: %pastebin "cell one" 0-10
Out[56]: u'https://gist.github.com/6564883'
```

Connect to this kernel remotely

Using the %connect_info magic you can obtain the connection info to connect to this workbook from another ipython console or qtconsole using :

```
ipython qtconsole --existing
```

```
In [58]: %connect_info
{
```

Publishing your Work

Newly added in the 1.0 release of IPython is the nbconvert tool, which allows you to convert an .ipynb notebook document file into various static formats.

Currently, nbconvert is provided as a command line tool, run as a script using IPython. A direct export capability from within the IPython Notebook web app is planned.

The command-line syntax to run the nbconvert script is: MORE OPTIONS

```
ipython nbconvert --to FORMAT notebook.ipynb
```

This page is converted and published to the following formats using this tool:

- HTML
- PDF (the PDF is created using wkhtml2pdf that takes the html file as an input)
- LATEX
- Reveal.js slideshow

Building(exporting) from within the notebook

You can even call the build script from the notebook. The script will convert this page to an html and slide file. It will also compile to PDF and stitch a front page to it. Some of the last text in the building process wont appear as this notebook is being updated as it is being compile. Maybe not the best idea but saved a lot of time...

File links to exported content

The links below can be used to verify the output from the convertion process. This saved me a lot of time as I could just click below and have a look at the files without exiting the notebook.

Links to some interesting notebooks

The following notebooks showcase multiple aspects of IPython, from its basic use to more advanced scenarios. They introduce you to the use of the Notebook and also cover aspects of IPython that are available in other clients, such as the cell magics for multi-language integration or our extended display protocol.

For beginners, we recommend that you start with the 5-part series that introduces the system, and later read others as the topics interest you.

Once you are familiar with the notebook system, we encourage you to visit our gallery where you will find many more examples that cover areas from basic Python programming to advanced topics in scientific computing.

- Animations Using clear output
- Cell Magics
- <u>Custom Display Logic</u>
- Cython Magics
- Data Publication API
- Frontend-Kernel Model
- Octave Magic
- Part 1 Running Code
- Part 2 Basic Output
- Part 3 Pylab and Matplotlib
- Part 4 Markdown Cells
- Part 5 Rich Display System
- Progress Bars
- R Magics
- Script Magics
- SymPy Examples
- <u>Trapezoid Rule</u>
- Typesetting Math Using MathJax

Sources / References

Since this talk focussed on the life cycle of the analysis to publication many of the code examples were taken from their respective websites. If I have not given credit at any point please let me know and I will make sure that the work is updated

- 1. SciPy
- 2. Fernando Pérez, Brian E. Granger, IPython: A System for Interactive Scientific Computing, omputing in Science and Engineering, vol. 9, no. 3, pp. 21-29, May/June 2007, doi:10.1109/MCSE.2007.53. URL: http://ipython.org3
- 3. Hunter, J. D.Matplotlib: A 2D graphics environment
- 4. Sympy
- 5. Bayesian Methods for Hackers, the use of the custom css and also the custom matplotlib skin
- 6. Custom Stylesheets and here

Embedding the final presentation into the notebook!

can also use normal HTML in a cell and using the iframe tag the slideshow was embedded to a cell below. Since this document has not been build yet...we are editing it now, the slideshow below is linked to the previous saved version of this notebook. So if we did not make to many changes it should be pretty close to being the same thing.

The build script generates a slideshow version of this notebook and saves it in the output directory. You