Wissenschaftliches Publizieren mit Python

Pycon DE 2013, Köln

Yves Hilpisch
Continuum Analytics Europe GmbH
yves@continuum.io
@dyjh

CONTINUUM ANALYTICS

Agenda

- Fortschritt und Sprache
- Python Ecosystem
- Python Tools
- Case Study
- Open, Collaborative Research

Fortschritt und Sprache

Fortschritt im Allgemeinen

- **Effizienz**: Etwas Gegebenes besser machen

  100m schneller laufen.

- **Effektivität**: Mehr machen mit dem Gegebenen

  Den Speer weiter werfen.

- **Qualität**: Neues machen, schaffen

  Alleine um die Welt segeln
  (1895–1898 Joshua Slocum, erste Alleinumseglung der Erde.)

Wissenschaftlicher Fortschritt

„Scientific progress is the idea that science increases its ***problem-solving ability*** *through the application of the scientific method.“*

http://en.wikipedia.org/wiki/Scientific_progress

Wissenschaftliche Methode

*„The scientific method is a body of techniques [**and tools**] for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge.To be termed scientific, a method of inquiry must be based on empirical and measurable evidence subject to specific principles of reasoning.“*

http://en.wikipedia.org/wiki/Scientific_method

Die Rolle der Sprache

"It is argued that language plays an active role in the development of scientific ideas. ...

And is mathematics somehow more or less than a language? ...

A particular characteristic of mathematics which appears in one aspect to differentiate it from language is its appeal to visual thinking."

Alan Ford and F. David Peat (1988): "The Role of Language in Science."

Sprachen in der Wissenschaft

- **Englisch** – geschriebene/gesprochene Sprache
- **Mathematik** – Symbolsprache, Logik
- **Code** – „Pseudo-Code“, Computersprache

Code + Struktur + Highlighting/Coloring

Python Ecosystem

Python Ecosystem

- **Python**: Basissystem bzw. Interpreter
- **Scientific Stack**: NumPy, SciPy, pandas, Cython, matplotlib, scikit-learn
- **Tools**: IPython, Spyder, Sphinx, Latex

Python spielt vierfache Rolle

- **Python**: Programmiersprache zur Erarbeitung von Forschungsergebnissen
- **Kommunikation**: Python zur Kommunikation von z.B. Algorithmen
- **Dokumentation**: Python zur Dokumentation von Algorithmen und Dartstellung von Ergebnissen
- **Publikation**: Tools zur Publikation

Python Tools

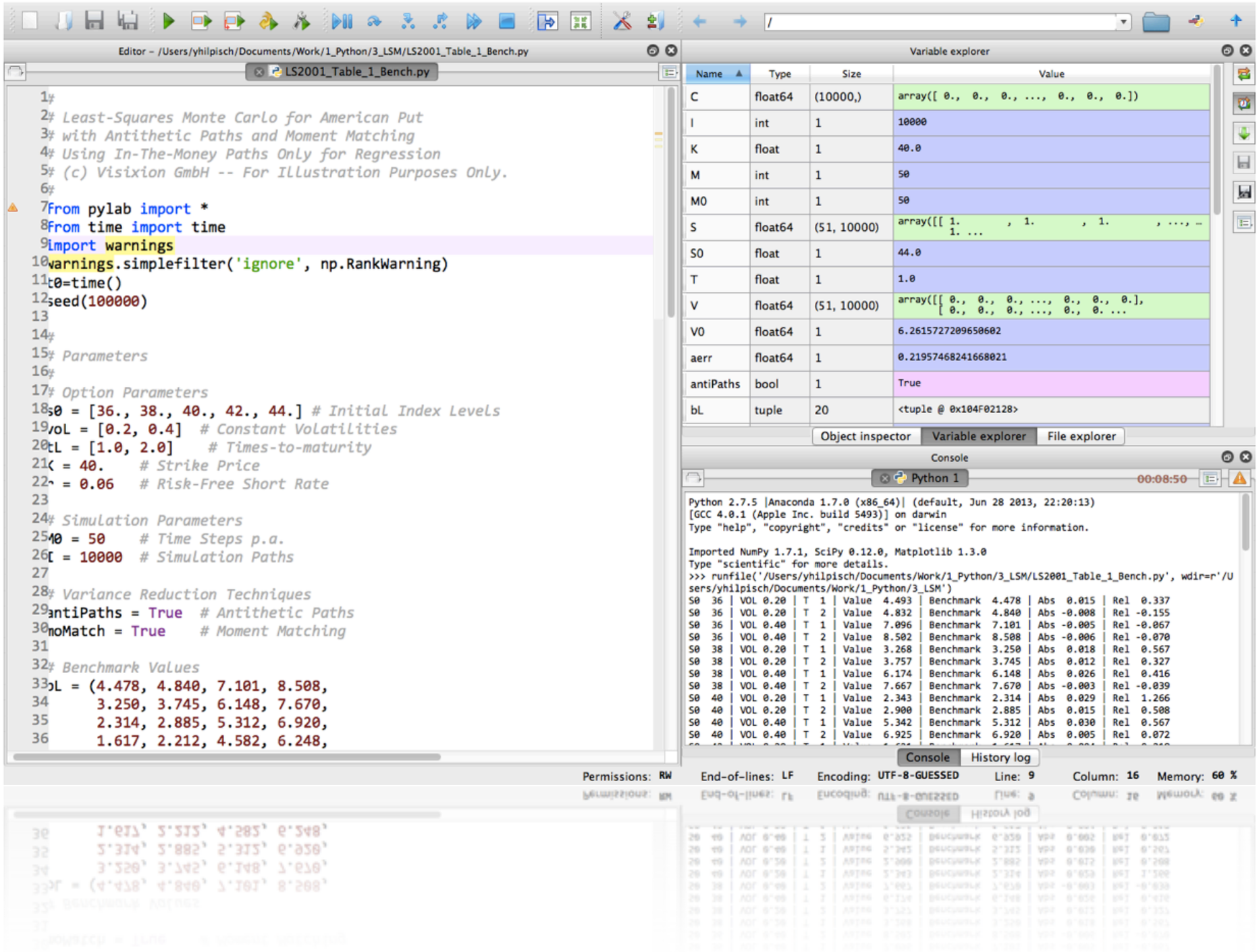
Python Tools

- **Spyder**: Integrierte Entwicklungsumgebung
- **IPython**: Pythons Killer App, insbesondere die Notebook Variante
- **Sphinx**: Dokumentationsumgebung und flexibles Autorensystem
- **PythonTEX**: Python für den Latex-Compiler

Spyder

- Code Editor mit Highlighting
- Code Checking, z.B. für PEP8
- Debugging
- Object und Variable Inspector
- Python CLI
- IPython Integration

Spyder


Editor – /Users/yhilpisch/Documents/Work/1_Python/3_LSM/LS2001_Table_1_Bench.py
LS2001_Table_1_Bench.py
# Least-Squares Monte Carlo for American Put
# with Antithetic Paths and Moment Matching
# Using In-The-Money Paths Only for Regression
# (c) Visixion GmbH -- For Illustration Purposes Only.
from pylab import *
from time import time
import warnings
warnings.simplefilter('ignore', np.RankWarning)
t0=time()
seed(100000)
# Parameters
# Option Parameters
s0 = [36., 38., 40., 42., 44.] # Initial Index Levels
vol = [0.2, 0.4] # Constant Volatilities
tL = [1.0, 2.0] # Times-to-maturity
K = 40. # Strike Price
r = 0.06 # Risk-Free Short Rate
# Simulation Parameters
M0 = 50 # Time Steps p.a.
I = 10000 # Simulation Paths
# Variance Reduction Techniques
antiPaths = True # Antithetic Paths
momatch = True # Moment Matching
# Benchmark Values
bL = (4.478, 4.840, 7.101, 8.508,
3.250, 3.745, 6.148, 7.670,
2.314, 2.885, 5.312, 6.920,
1.617, 2.212, 4.582, 6.248,
Variable explorer
Name
Type
Size
Value
C
float64
(10000,)
array([0., 0., 0., ..., 0., 0., 0.])
I
int
1
10000
K
float
1
40.0
M
int
1
50
M0
int
1
50
S
float64
(51, 10000)
array([[1. , 1. , 1. , ...,
1. ...
S0
float
1
44.0
T
float
1
1.0
V
float64
(51, 10000)
array([[0., 0., 0., ..., 0., 0., 0.],
[0., 0., 0., ..., 0., 0. ...
V0
float64
1
6.2615727209650602
aerr
float64
1
0.21957468241668021
antiPaths
bool
1
True
bL
tuple
20
<tuple @ 0x104F02128>
Object inspector
Variable explorer
File explorer
Console
Python 1
00:08:50
Python 2.7.5 |Anaconda 1.7.0 (x86_64)| (default, Jun 28 2013, 22:20:13)
[GCC 4.0.1 (Apple Inc. build 5493)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
Imported NumPy 1.7.1, SciPy 0.12.0, Matplotlib 1.3.0
Type "scientific" for more details.
>>> runfile('/Users/yhilpisch/Documents/Work/1_Python/3_LSM/LS2001_Table_1_Bench.py', wdir=r'/U
sers/yhilpisch/Documents/Work/1_Python/3_LSM')
S0 36 | VOL 0.20 | T 1 | Value 4.493 | Benchmark 4.478 | Abs 0.015 | Rel 0.337
S0 36 | VOL 0.20 | T 2 | Value 4.832 | Benchmark 4.840 | Abs -0.008 | Rel -0.155
S0 36 | VOL 0.40 | T 1 | Value 7.096 | Benchmark 7.101 | Abs -0.005 | Rel -0.067
S0 36 | VOL 0.40 | T 2 | Value 8.502 | Benchmark 8.508 | Abs -0.006 | Rel -0.070
S0 38 | VOL 0.20 | T 1 | Value 3.268 | Benchmark 3.250 | Abs 0.018 | Rel 0.567
S0 38 | VOL 0.20 | T 2 | Value 3.757 | Benchmark 3.745 | Abs 0.012 | Rel 0.327
S0 38 | VOL 0.40 | T 1 | Value 6.174 | Benchmark 6.148 | Abs 0.026 | Rel 0.416
S0 38 | VOL 0.40 | T 2 | Value 7.667 | Benchmark 7.670 | Abs -0.003 | Rel -0.039
S0 40 | VOL 0.20 | T 1 | Value 2.343 | Benchmark 2.314 | Abs 0.029 | Rel 1.266
S0 40 | VOL 0.20 | T 2 | Value 2.900 | Benchmark 2.885 | Abs 0.015 | Rel 0.508
S0 40 | VOL 0.40 | T 1 | Value 5.342 | Benchmark 5.312 | Abs 0.030 | Rel 0.567
S0 40 | VOL 0.40 | T 2 | Value 6.925 | Benchmark 6.920 | Abs 0.005 | Rel 0.072
Console
History log
Permissions: RW
End-of-lines: LF
Encoding: UTF-8-GUESSED
Line: 9
Column: 16
Memory: 60 %

IPython

- Interaktive Python Shell
- Vielzahl von „%magic“ Funktionen
- Command Line Historie
- Shell, QT version und Notebook (Browser)
- Parallele Programmausführung (ipcluster)
- viele Konvertierungsmöglichkeiten

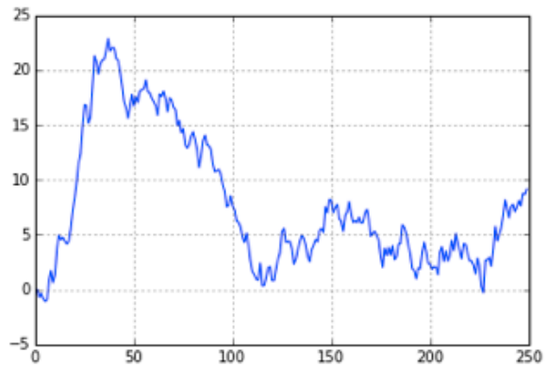
IPython Shell

bash | python | bash | python | python | python

```
In [3]: exit
Yvess-MacBook-Air:Sphinx yhilpisch$ ipython --pylab
Python 2.7.5 |Anaconda 1.7.0 (x86_64)| (default, Jun 28 2013, 22:20:13)
Type "copyright", "credits" or "license" for more information.

IPython 1.0.0 -- An enhanced Interactive Python.
?         -> 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.
Using matplotlib backend: MacOSX

In [1]: a = linspace(0, 100, 33)

In [2]: a
Out[2]:
array([   0.   ,    3.125,    6.25 ,    9.375,   12.5  ,   15.625,
         18.75 ,   21.875,   25.   ,   28.125,   31.25 ,   34.375,
         37.5  ,   40.625,   43.75 ,   46.875,   50.   ,   53.125,
         56.25 ,   59.375,   62.5  ,   65.625,   68.75 ,   71.875,
         75.   ,   78.125,   81.25 ,   84.375,   87.5  ,   90.625,
         93.75 ,   96.875,  100.   ])

In [3]: a ** 2
Out[3]:
array([  0.00000000e+00,   9.76562500e+00,   3.90625000e+01,
         8.78906250e+01,   1.56250000e+02,   2.44140625e+02,
         3.51562500e+02,   4.78515625e+02,   6.25000000e+02,
         7.91015625e+02,   9.76562500e+02,   1.18164062e+03,
         1.40625000e+03,   1.65039062e+03,   1.91406250e+03,
         2.19726562e+03,   2.50000000e+03,   2.82226562e+03,
         3.16406250e+03,   3.52539062e+03,   3.90625000e+03,
         4.30664062e+03,   4.72656250e+03,   5.16601562e+03,
         5.62500000e+03,   6.10351562e+03,   6.60156250e+03,
         7.11914062e+03,   7.65625000e+03,   8.21289062e+03,
         8.78906250e+03,   9.38476562e+03,   1.00000000e+04])

In [4]: a[3:10]
Out[4]: array([  9.375,  12.5  ,  15.625,  18.75 ,  21.875,  25.   ,  28.125])

In [5]:
```

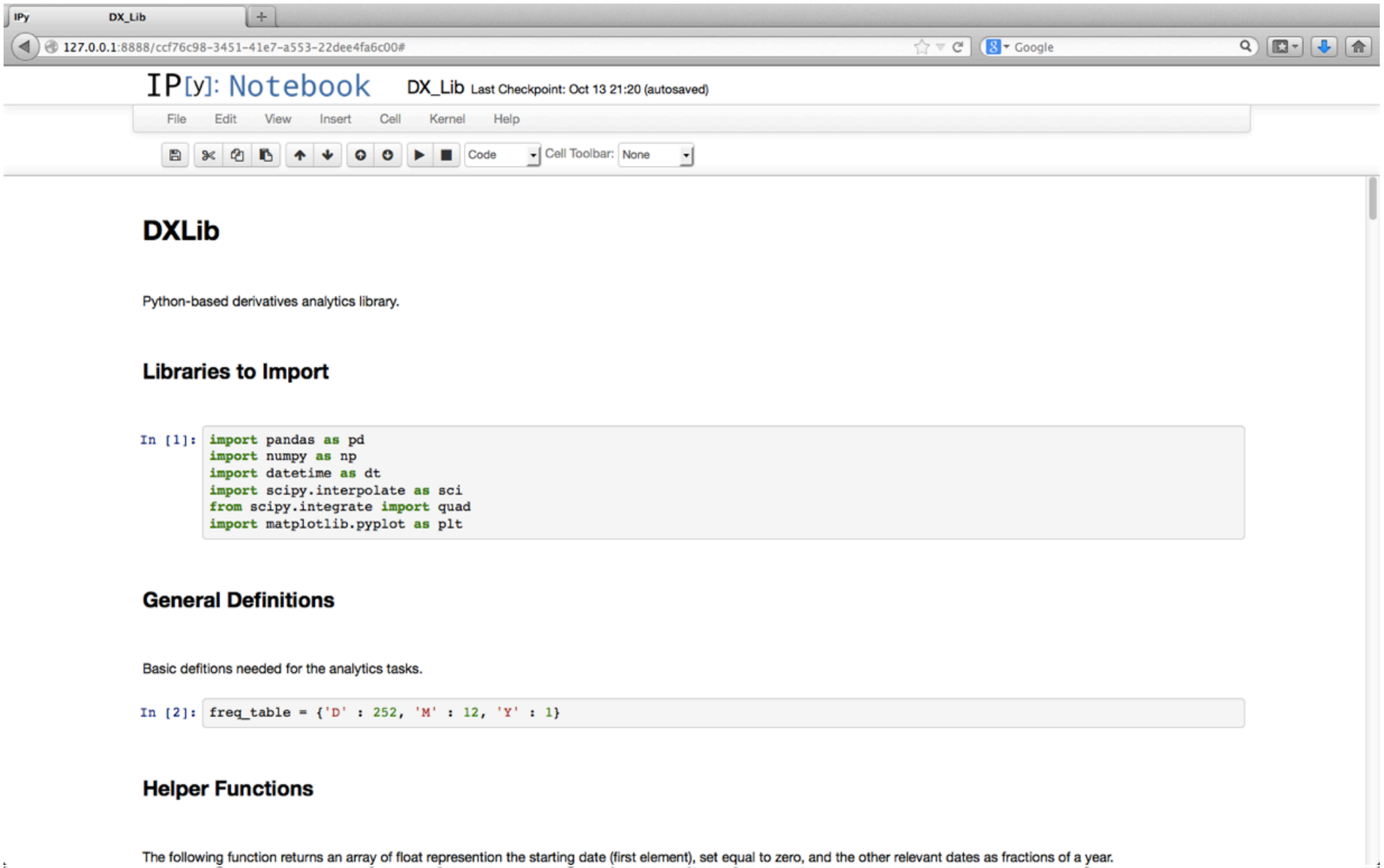
IPython QT Console

```
Python 2.7.5 |Anaconda 1.7.0 (x86_64)| (default, Jun 28 2013, 22:20:13)
Type "copyright", "credits" or "license" for more information.

IPython 1.0.0 -- An enhanced Interactive Python.
?         -> 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.
%guiref   -> A brief reference about the graphical user interface.

In [1]: a = standard_normal(250)

In [2]: c = cumsum(a)

In [3]: figure()
   ...: plot(c)
   ...: grid(True)
```




```
In [4]: figure()
   ...: hist(a)
   ...: grid(True)
```

IPython Notebook



Sphinx

- Umgebung zur Dokumentation von Python-Projekten und vielem mehr ...
- Syntax basiert i.W. auf reStructuredText (.rst)
- Source Code comilierbar in verschiedene Formate (HTML, Latex/PDF)
- Gute Erweiterbarkeit (extensions) und Anpassbarkeit (CSS)
- Integration mit IPython (über extension)

PythonTEX – Source Code

\section{Plots with matplotlib}

```
\section{Plots with matplotlib}

We can create plots with matplotlib, perfectly matching the plot fonts with
the document fonts. No more searching for the code that created a figure!

You may want to use matplotlib's PGF backend when creating plots.

\begin{pylabblock}
rc('text', usetex=True)
rc('font', family='serif')
rc('font', size=10.0)
rc('legend', fontsize=10.0)
rc('font', weight='normal')
x = linspace(0, 10)
figure(figsize=(4, 2.5))
plot(x, sin(x), label='$\sin(x)$')
xlabel(r'$x\mathrm{-axis}$')
ylabel(r'$y\mathrm{-axis}$')
legend(loc='lower right')
savefig('myplot.pdf', bbox_inches='tight')
\end{pylabblock}

\begin{center}
\includegraphics{myplot.pdf}
\end{center}
```

PythonTEX – Output

```
phi = Symbol(r'\phi')
h = Integral(exp(-phi**2), (phi, 0, oo))
```

$$\int_0^{\infty} e^{-\phi^2}\, d\phi = \frac{1}{2}\sqrt{\pi}$$

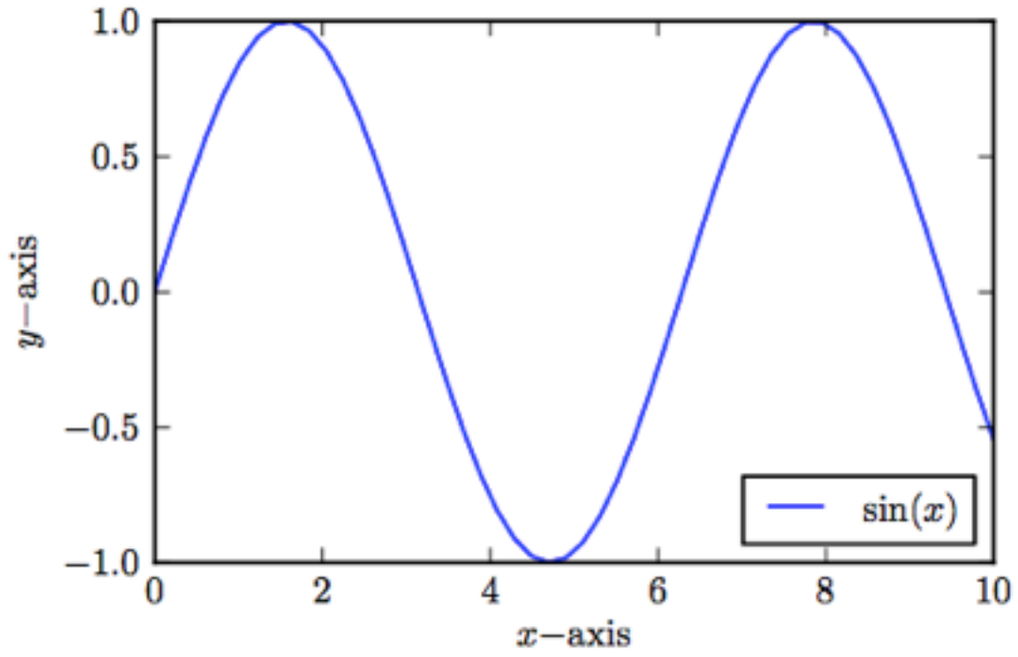
5 Plots with matplotlib

We can create plots with matplotlib, perfectly matching the plot fonts with the document fonts. No more searching for the code that created a figure!

You may want to use matplotlib's PGF backend when creating plots.

```
rc('text', usetex=True)
rc('font', family='serif')
rc('font', size=10.0)
rc('legend', fontsize=10.0)
rc('font', weight='normal')
x = linspace(0, 10)
figure(figsize=(4, 2.5))
plot(x, sin(x), label='$\sin(x)$')
xlabel(r'$x\mathrm{-axis}$')
ylabel(r'$y\mathrm{-axis}$')
legend(loc='lower right')
savefig('myplot.pdf', bbox_inches='tight')
```

PythonTEX – Output



6 Basic pylab interaction

```
from scipy.integrate import quad
myintegral = quad(lambda x: e**-x**2, 0, inf)[0]
```

$$\int_0^\infty e^{-x^2}\,dx = 0.886226925453$$

Case Study

Case Study

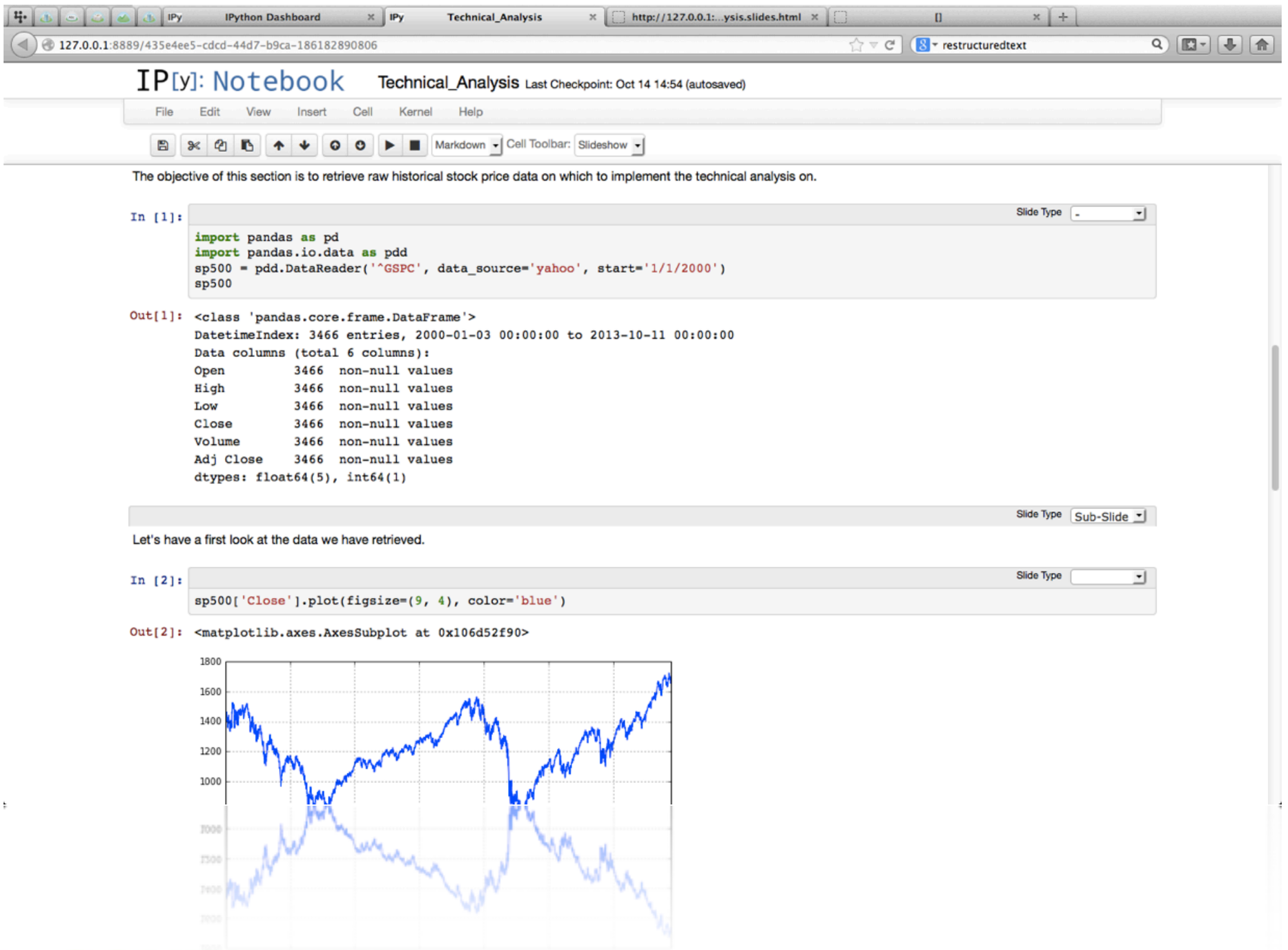
Einfaches Beispiel aus dem Finanzbereich, das

- Text
- Struktur
- Code
- Graphiken
- Formeln

aufweist.

IPython Notebook Umsetzung

Das IPython Notebook (ipynb)



CONTINUUM ANALYTICS

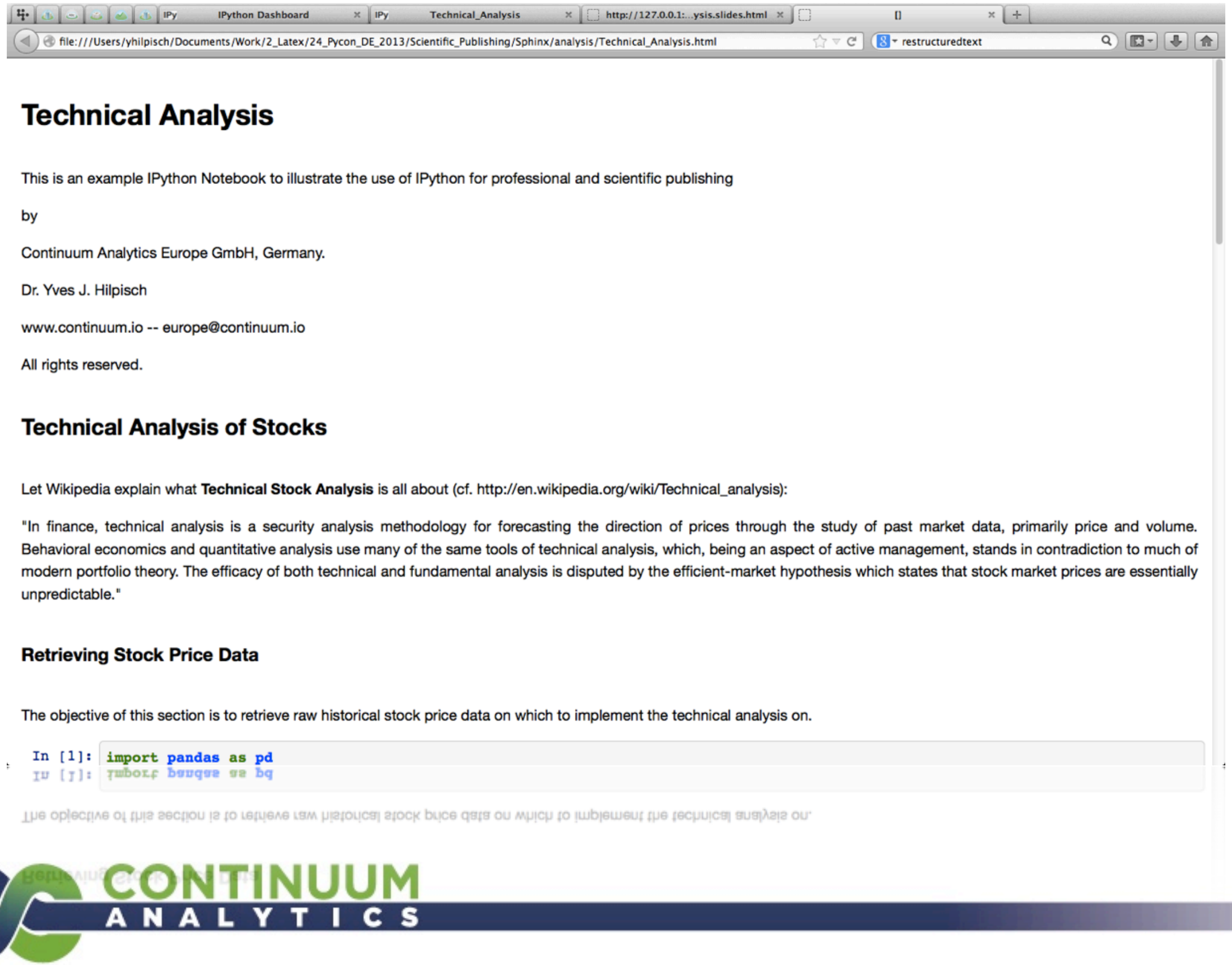
IPython Notebook Markdown

Conclusions

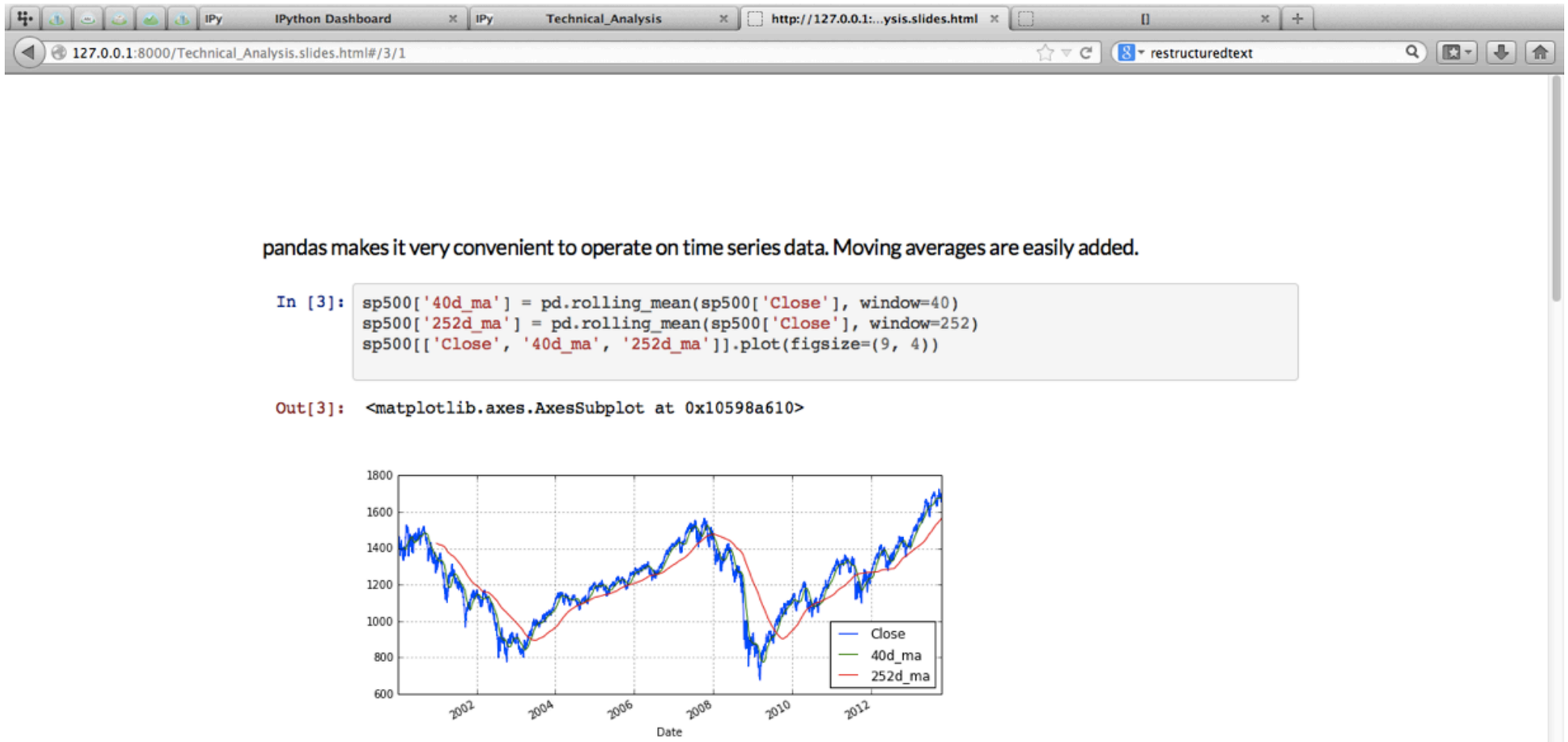
Slide Type: Sub-Slide

```
IPython is an excellent envrionment to not only document Python projects and code. It is rather a great
tool for interactively develeoping professional and scientific publications. It allows to generate output
in such standard formats as:

 * **HTML**, i.e. a Web page
 * **PDF**, i.e. via Latex
 * **ePub**, i.e. for electronic books

Since Python code and its results are already included in the Notebook, it saves a lot of the typical
copy & paste tasks.

In addition, it is also quite easy to include Latex code/formulas directly into the document.

$dS_t = r S_t dt + \sigma S_t dZ_t$

IPython seems the right choice when it comes to documents that include

  * English
  * Mathematics
  * Python
```

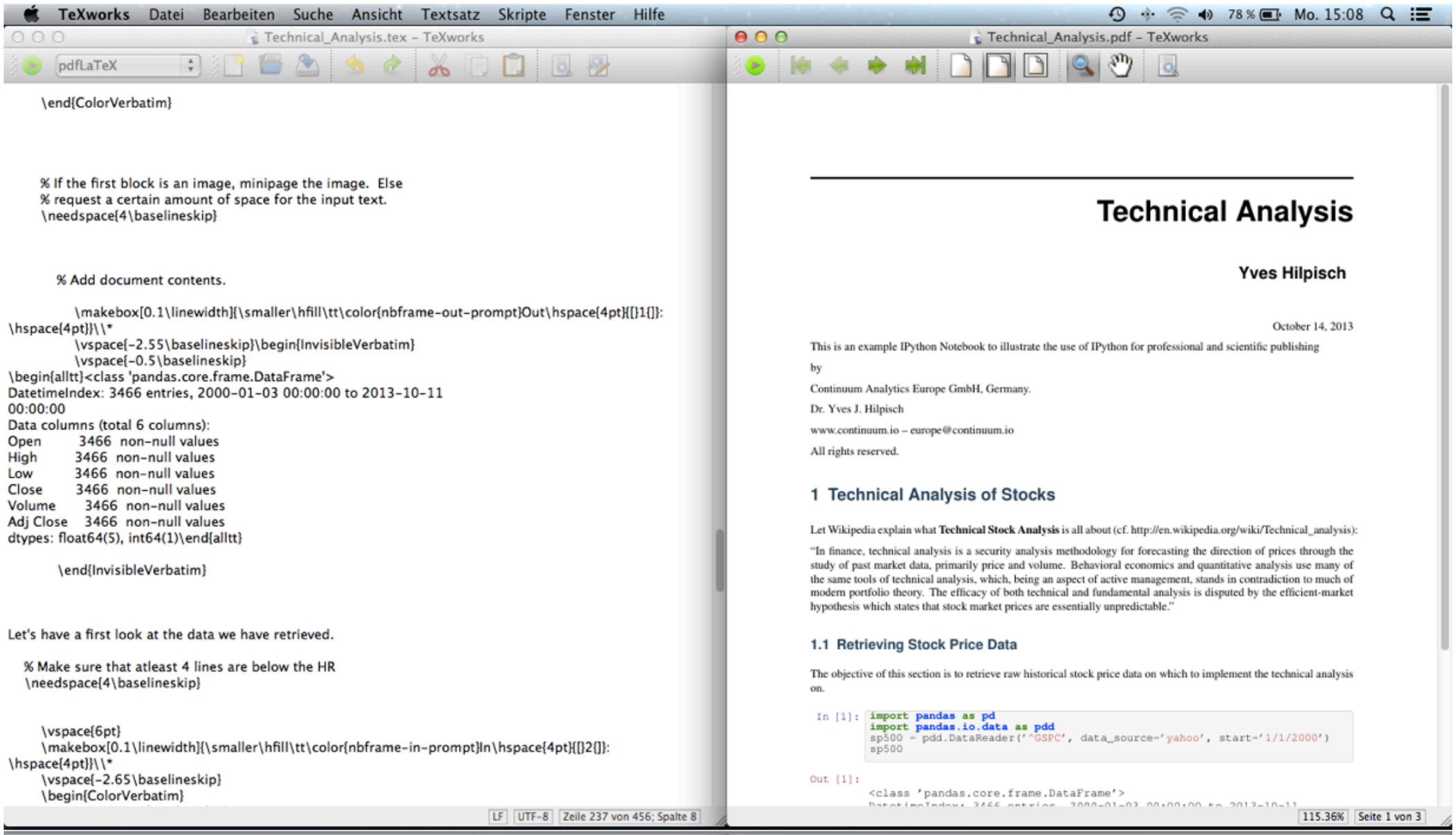
IPython HTML output im Browser



IPython HTML5 Slides (mit reveal.js)



IPython Latex und PDF Output



CONTINUUM ANALYTICS

Sphinx Umsetzung

Sphinx Source Code

technical_analysis.rst | conclusion.rst | index.rst

```
23
24  ------------------------------
25   Retrieving Stock Price Data
26  ------------------------------
27
28  The objective of this section is to retrieve raw historical stock price data on which to implement the technical
    analysis on.
29
30  .. ipython:: python
31
32      import pandas as pd
33
34      import pandas.io.data as pdd
35
36      sp500 = pdd.DataReader('^GSPC', data_source='yahoo', start='1/1/2000')
37
38      sp500
39
40
41  Let's have a first look at the data we have retrieved.
42
43  .. ipython:: python
44
45      # @savefig sp500.png
46      sp500['Close'].plot(figsize=(9, 4), color='blue')
47
48
49  :num:`Figure #figure1` shows the resulting figure.
50
51  .. _figure1:
52
53  .. figure:: sp500.png
54     :scale: 70 %
55     :align: center
56
57     The S&P 500 index since the beginning of 2000.
58
59
```

Line 43, Column 11 | Spaces: 2 | reStructuredText

Sphinx Source Code (Analytics)

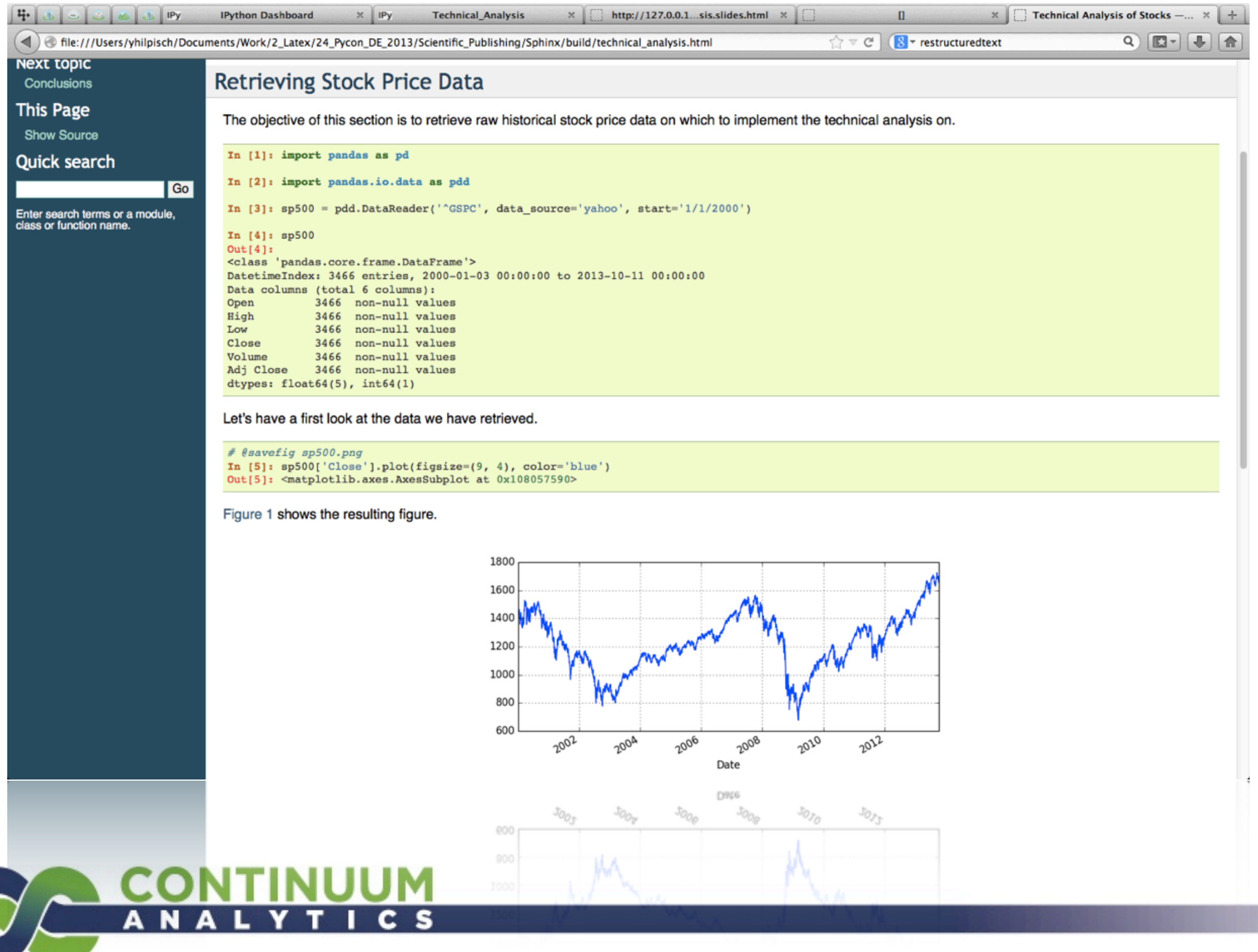
```
29
30  .. ipython:: python
31
32      import pandas as pd
33
34      import pandas.io.data as pdd
35
36      sp500 = pdd.DataReader('^GSPC', data_source='yahoo', start='1/1/2000')
37
38      sp500
39
```

Sphinx Source Code (Graphics)

```
43  .. ipython:: python
44
45      # @savefig sp500.png
46      sp500['Close'].plot(figsize=(9, 4), color='blue')
47
48
49  :num:`Figure #figure1` shows the resulting figure.
50
51  .. _figure1:
52
53  .. figure:: sp500.png
54     :scale: 70 %
55     :align: center
56
57     The S&P 500 index since the beginning of 2000.
58
59
```

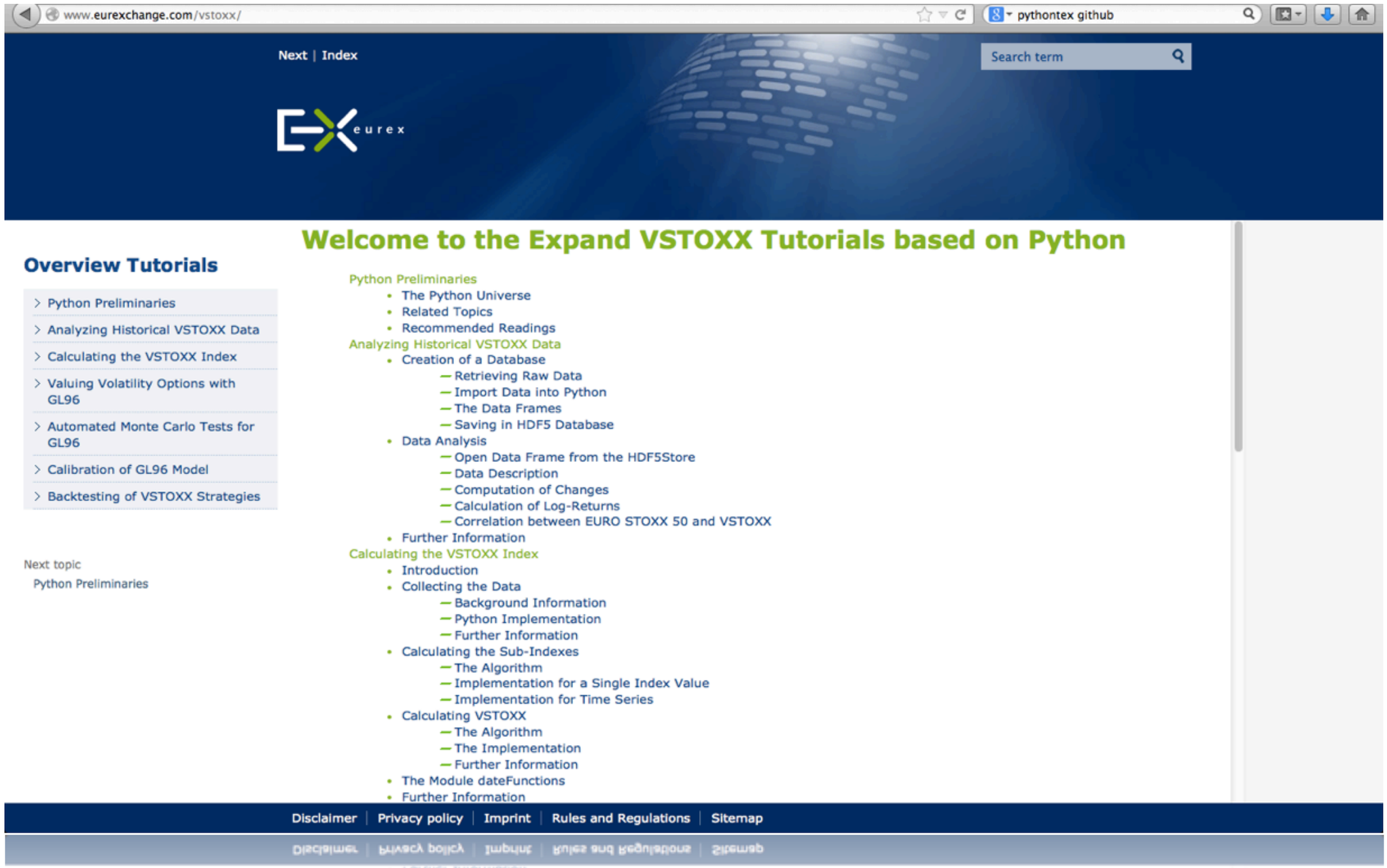
Line 43, Column 11

Sphinx HTML Output

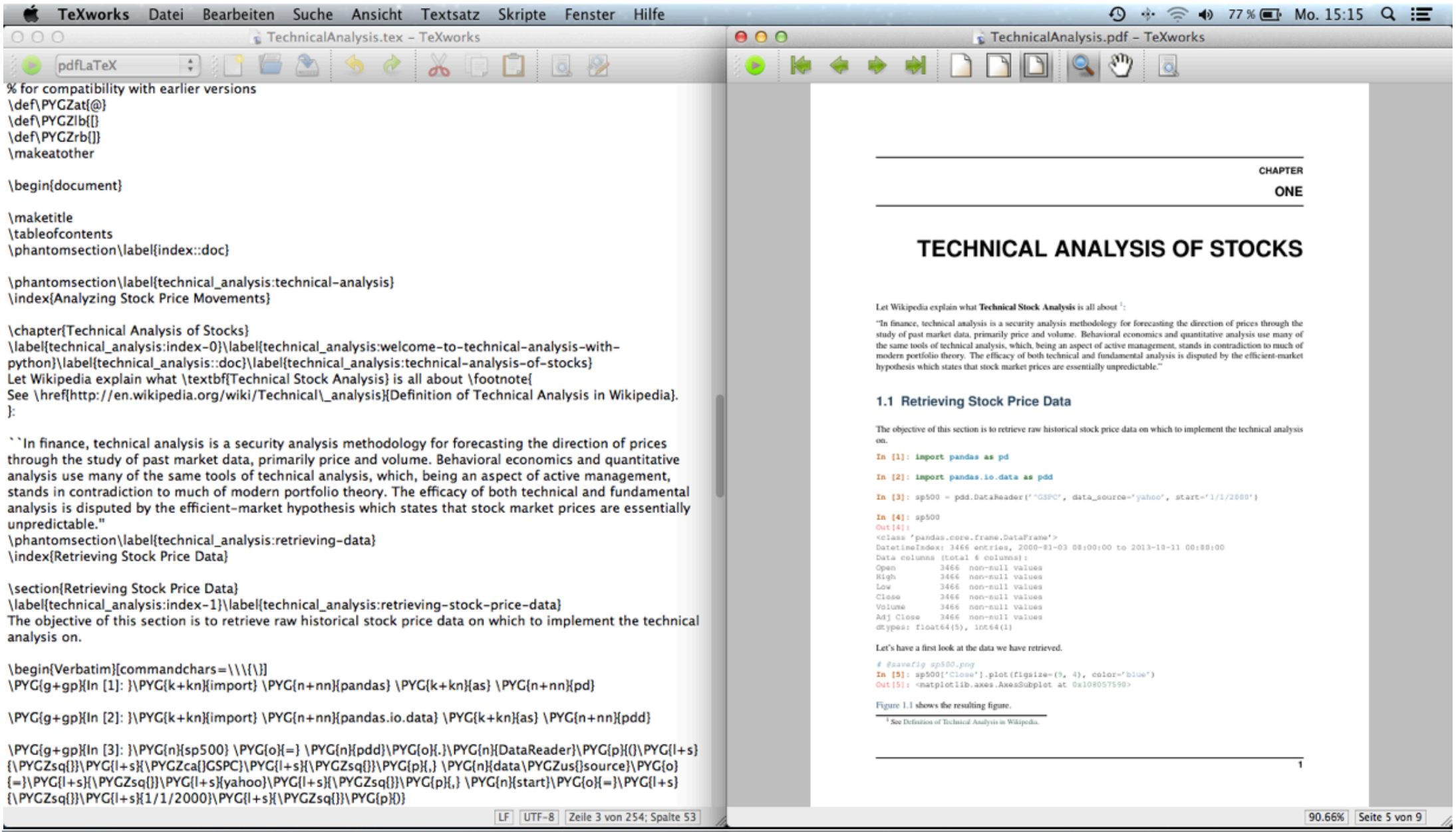


CONTINUUM ANALYTICS

Sphinx Customized HTML Site



Sphinx Latex und PDF Output



Open, Collaborative Research

CONTINUUM ANALYTICS

Open Research in Physik mit IPython

Das Forschungspapier auf arxiv.org

Submitted for publication in ApJ Letters
Preprint typeset using LaTeX style emulateapj v. 12/16/11

http://arxiv.org/abs/1303.2690

THE KINEMATICS OF THE LOCAL GROUP IN A COSMOLOGICAL CONTEXT

J. E. Forero-Romero[1], Y. Hoffman[2], S. Bustamante[3], S. Gottlöber[4], G. Yepes[5]

(Dated: March 10, 2013)
Submitted for publication in ApJ Letters

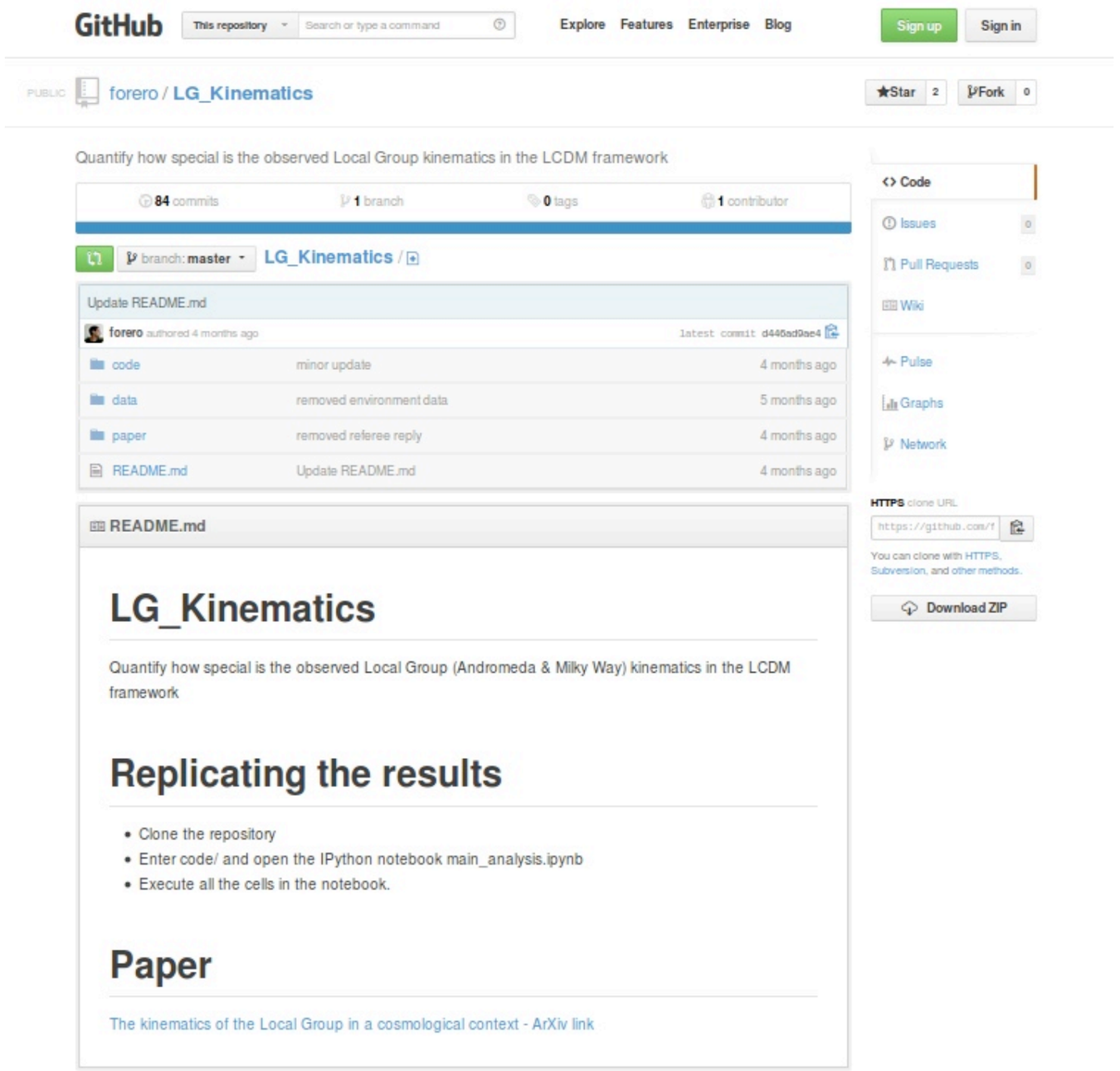
ABSTRACT

Recent observations constrained the tangential velocity of M31 with respect to the Milky Way (MW) to be $v_{\rm M31,tan} < 34.4$ km s^{-1}and the radial velocity to be in the range $v_{\rm M31,rad} = -109 \pm 4.4$km s^{-1}(van der Marel et al. 2012). In this study we use a large volume high resolution N-body cosmological simulation (Bolshoi) together with three constrained simulations to statistically study this kinematics in the context of the ΛCDM. The comparison of the ensembles of simulated pairs with the observed LG at the 1-σ level in the uncertainties has been done with respect to the radial and tangential velocities, the reduced orbital energy ($e_{\rm tot}$), angular momentum ($l_{\rm orb}$) and the dimensionless spin parameter, λ. Our main results are: (i) the preferred radial and tangential velocities for pairs in ΛCDM are $v_{\rm r} = -80 \pm 20$km s^{-1}, $v_{\rm t} = 50 \pm 10$km s^{-1}, (ii) pairs around that region are 3 to 13 times more common than pairs within the observational values, (iii) 15% to 24% of LG-like pairs in ΛCDM have energy and angular momentum consistent with observations while (iv) 9% to 13% of pairs in the same sample show similar values in the inferred dimensionless spin parameter. It follows that within current observational uncertainties the quasi-conserved quantities that characterize the orbit of the LG, i.e. $e_{\rm tot}$, $l_{\rm orb}$ and λ, do not challenge the standard ΛCDM model, but the model is in tension with regard to the actual values of the radial and tangential velocities. This might hint to a problem of the ΛCDM model to reproduce the observed LG.

Subject headings: galaxies: kinematics and dynamics, Local Group, methods:numerical

CONTINUUM ANALYTICS

Das Github Repository



https://github.com/forero/LG_Kinematics

Das IPython Notebook

IP[y]: Notebook main_analysis Last saved: Jul 02 09:55

File Edit View Insert Cell Kernel Help

Markdown Cell Toolbar: None

The Kinematics of the Local Group in a Cosmological Context

This notebook containts all data analysis for the ApJ Letter *The Kinematics of the Local Group in a Cosmological Context* by Jaime E. Forero-Romero, Yehuda Hoffman, Sebastian Bustamante, Stefan Gottloeber and Gustavo Yepes.

Follow through the notebook to generate the figures and numbers used in the paper . You can also generate results that were mentioned but not explicitly reported. For instance, the results for pairs obtained for a Friend-of-Friends halo finder.

In [1]:
```
%pylab inline
%load_ext autoreload
%autoreload 2
from make2DHistogram import *
from generateInfo import *
```

In [2]:
```
# Global paths and constants
data_path = "../data/"

G_GRAV = 4.54E-48 #units of Mpc^3 Msun^-1 s^-2
KM_TO_MPC = 3.2E-20
HUBBLE = 0.70
E_UNITS = 1.0E-36
L_UNITS = 1.0
```

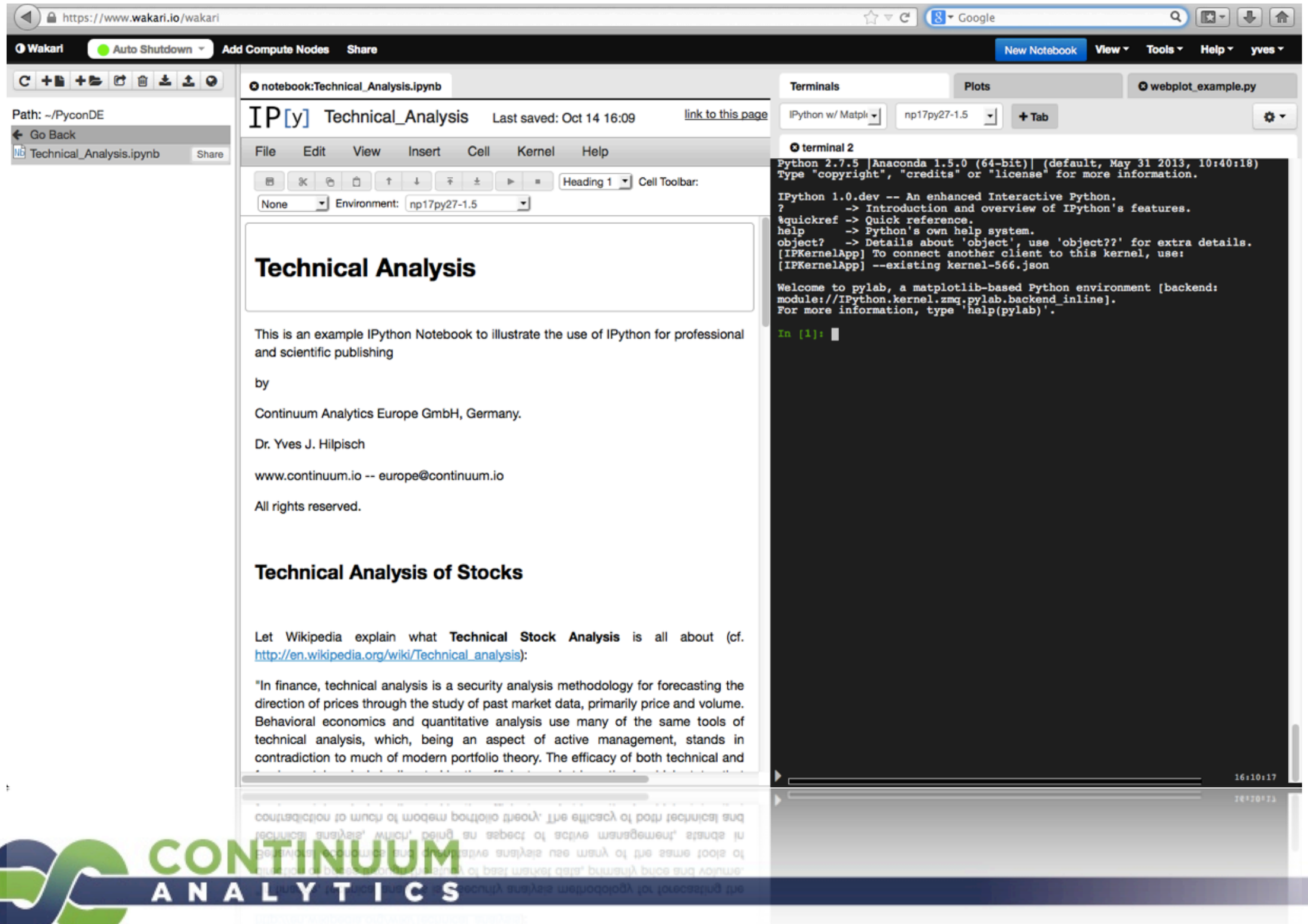
Collaborative Research mit Wakari.io

Wakari.io

Wakari ist eine vollständige Python Umgebung

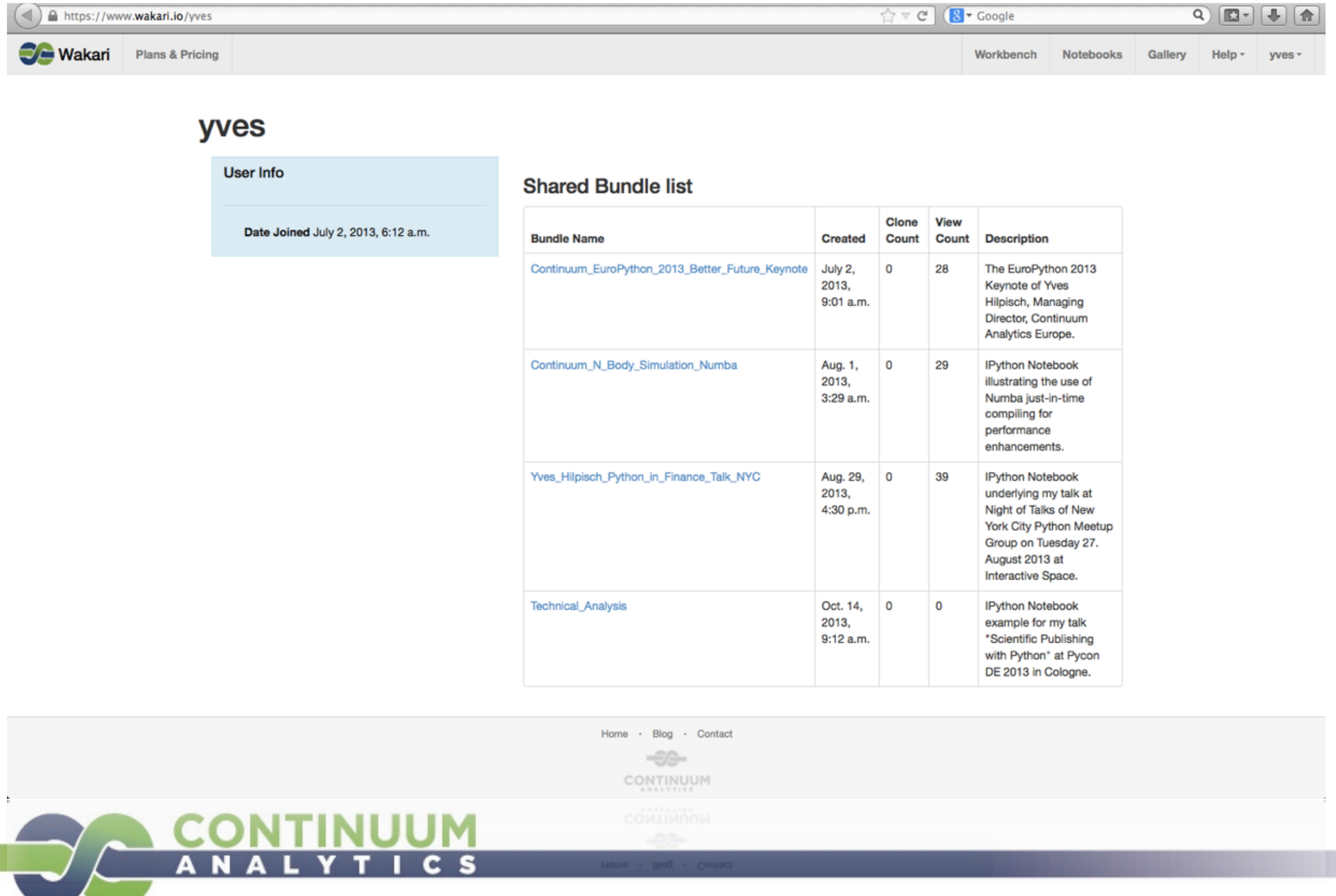
- Linux Server inkl. Shell & File Manager
- Cloud- & Browser-basiert
- Python, Anaconda, IPython
- Packaging & Sharing von IPython Notebooks
- Skalierbar (Zuschalten von Nodes)
- Wakari Enterprise ab November 2013

Wakari.io als Python Umgebung



CONTINUUM ANALYTICS

Teilen von ganzen IPYNB Projekten



CONTINUUM ANALYTICS

Continuum Analytics Europe GmbH
Rathausstrasse 75-79
66333 Voelklingen
Germany

www.continuum.io
europe@continuum.io

Dr. Yves J. Hilpisch | @dyjh

CONTINUUM
ANALYTICS