

# Setting up a data mining machine

## Installation of a Python & data mining toolkits

Peter Bonanati

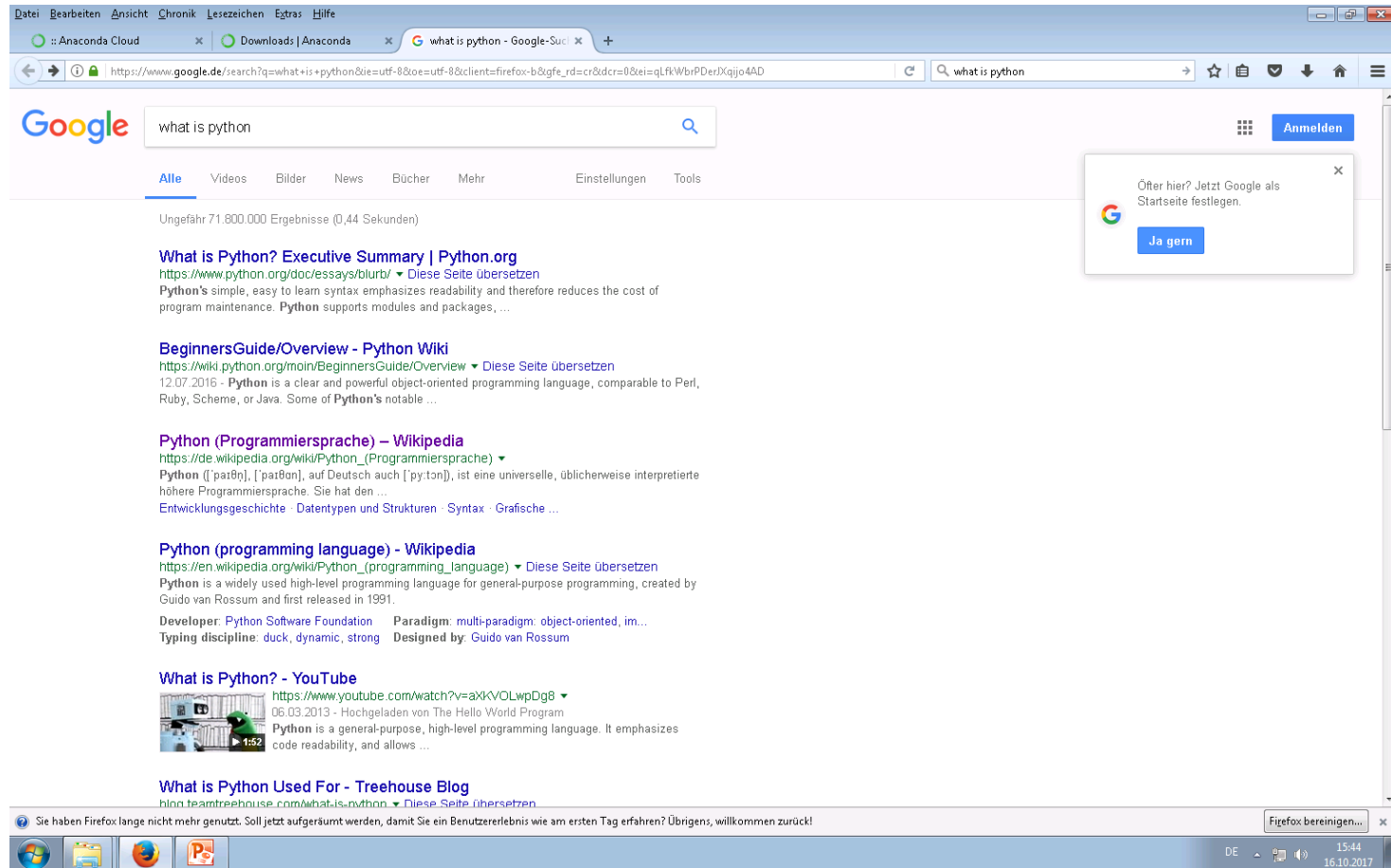


## Data mining?

- Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.



# What is Python?





## What is Python – and how I see it...

- Many things I often do can be done by a computer very well.  
Python helps me to translate my wishes into computer language.
- This task will work nice and is quit simple:
  - My Whish:
    - Open the resistance measurement data and draw me the resistance vs. time.
  - Python Code:
    - `Flow_data = Pd.read_csv('Flow.txt')`
    - `Flow_data['Resistance'].plot()`

This is already data mining.



## What is Python – and how I see it...

- Other tasks are more complicated will most likely fail:
  - Whish:
    - Check if the experiment was running correctly.
  - Code:
    - Many things to check and not so simple to implement

!!!!BUT!!!!

- Using Python will help me to build smaller tools to get along with most tasks
- See the programming rather as a tool to build new tools helping you for your mining.
- It will help you to skip (boring) repetitive work and concentrate on what distinguishes you from a computer:
  - your intelligence



# What is Python

- Python is a way to express your commands in a language the computer will understand. See it just as a language like French you will need to talk in France to get your glass of wine.
- “Python has a design philosophy that emphasizes code readability [...], and a syntax that allows programmers to express concepts in fewer lines of code [...]<sup>1</sup>. The language provides constructs intended to enable writing clear programs on both a small and large scale.<sup>[25]</sup>” Wikipedia



## What is Python – What makes it special

- I don't know why but many people are using Python as a language to build many, many tools for lots of use cases.
- Additionally most of those tools are shared and anybody can use it.
- Such tools are nothing more than a list of precise commands for a PC, telling "him" what to do step by step.  
Once this step-by-step recipe is ready, it can be easily applied to similar use cases.
- Those step-by-step commands can be seen as cookbooks and are generally called **Libraries**



## What is Python

- So if Python is just a language, where do I get those ‘step-by-step cookbook’ from?
  - The most useful ones come in as a set of multiple libraries. Since some of the libraries are depending on each other one has to take care that we have all needed libraries together.
  - This task is a bit complicated for humans, but a typical task for a computer.
  - That’s why we use ANACONDA, which checks the dependencies and bundles a well working package for us with most of the tools we will ever need.






**Let's start!**

Installing Anaconda



# ANACONDA



## ANACONDA DISTRIBUTION STARTER GUIDE

See full documentation for Anaconda Distribution  
[docs.anaconda.com/anaconda/](https://docs.anaconda.com/anaconda/)

### BEFORE YOU START

**Why do I need Anaconda Distribution?**  
Installing Python in a terminal is no joy. Many scientific packages require a specific version of Python to run, and it's difficult to keep them from interacting with each other. It is even harder to keep them updated. Anaconda Distribution makes getting and maintaining these packages quick and easy.

**What is Anaconda Distribution?**  
It is an open source, easy-to-install high performance Python and R distribution, with the conda package and environment manager and collection of 1,000+ open source packages with free community support.

**Then what is Miniconda?**  
It's Anaconda Distribution without the collection of 1,000+ open source packages. With Miniconda you install only the packages you want with the conda command.  
`conda install PACKAGENAME`  
**Example:** `conda install anaconda-navigator`

### GET IT

**Will it work on my machine?**  
Yes, Anaconda Distribution is available for Windows, macOS or Linux x86 or POWER8, 32- or 64-bit, 3GB HD available. Miniconda is the same but needs only 400 MB HD.

**Quick install it**  
[docs.anaconda.com/anaconda/install](https://docs.anaconda.com/anaconda/install)


**Get your conda cheat sheet**  
[conda.io/docs/using/cheatsheet.html](https://conda.io/docs/using/cheatsheet.html)

**Take the test drive**  
[conda.io/docs/test-drive.html](https://conda.io/docs/test-drive.html)

### NOW PLAY WITH THE WORLD'S MOST AWESOME SCIENTIFIC PACKAGES

Included in Anaconda 4.4+, or get with `"conda install PACKAGENAME"`

<b>1. NumPy</b> <a href="https://numpy.org">numpy.org</a> N-dimensional array for numerical computation	<b>7. Scikit-Learn</b> <a href="https://scikit-learn.org/stable">scikit-learn.org/stable</a> Python modules for machine learning and data mining
<b>2. SciPy</b> <a href="https://scipy.org">scipy.org</a> Scientific computing library for Python	<b>8. NLTK</b> <a href="https://nltk.org">nltk.org</a> Natural language toolkit
<b>3. Matplotlib</b> <a href="https://matplotlib.org">matplotlib.org</a> 2D Plotting library for Python	<b>9. Jupyter Notebook</b> <a href="https://jupyter.org">jupyter.org</a> Web app that allows you to create and share documents that contain live code, equations, visualizations and explanatory text
<b>4. Pandas</b> <a href="https://pandas.pydata.org">pandas.pydata.org</a> Powerful Python data structures and data analysis toolkit	<b>10. R essentials</b> <a href="https://conda.pydata.org/docs/with-conda.html">conda.pydata.org/docs/with-conda.html</a> R with 80+ of the most used R packages for data science <code>"conda install r-essentials"</code> R package list <a href="https://docs.anaconda.com/anaconda/language-pkg-docs">docs.anaconda.com/anaconda/language-pkg-docs</a>
<b>5. Seaborn</b> <a href="https://seaborn.pydata.org/">seaborn.pydata.org/</a> Statistical graphics library for Python	
<b>6. Bokeh</b> <a href="https://bokeh.pydata.org">bokeh.pydata.org</a> Interactive web visualization library	



CONTINUED ON BACK →



## GET IT

Will it work on my machine? Yes, Anaconda Distribution is available for Windows, macOS or Linux x86 or POWER8, 32- or 64-bit, 3GB HD available. Miniconda is the same but needs only 400 MB HD.

Quick install it [docs.anaconda.com/anaconda/install](https://docs.anaconda.com/anaconda/install)

Get your conda cheat sheet [conda.io/docs/using/cheatsheet.html](https://conda.io/docs/using/cheatsheet.html)

Take the test drive [conda.io/docs/test-drive.html](https://conda.io/docs/test-drive.html)

## NOW PLAY WITH THE WORLD'S MOST AWESOME SCIENTIFIC PACKAGES

Included in Anaconda 4.4+, or get with `"conda install PACKAGENAME"`

### 1. NumPy

[numpy.org](https://numpy.org)

N-dimensional array for numerical computation

### 2. SciPy

[scipy.org](https://scipy.org)

Scientific computing library for Python

### 3. Matplotlib

[matplotlib.org](https://matplotlib.org)

2D Plotting library for Python

### 4. Pandas

[pandas.pydata.org](https://pandas.pydata.org)

Powerful Python data structures and data analysis toolkit

### 5. Seaborn

[seaborn.pydata.org/](https://seaborn.pydata.org/)

Statistical graphics library for Python

### 6. Bokeh

[bokeh.pydata.org](https://bokeh.pydata.org)

Interactive web visualization library

### 7. Scikit-Learn

[scikit-learn.org/stable](https://scikit-learn.org/stable)

Python modules for machine learning and data mining

### 8. NLTK

[nltk.org](https://nltk.org)

Natural language toolkit

### 9. Jupyter Notebook

[jupyter.org](https://jupyter.org)

Web app that allows you to create and share documents that contain live code, equations, visualizations and explanatory text

### 10. R essentials

[conda.pydata.org/docs/r-with-conda.html](https://conda.pydata.org/docs/r-with-conda.html)

R with 80+ of the most used R packages for data science

`"conda install r-essentials"`

R package list

[docs.anaconda.com/anaconda/rlanguage-pkg-docs](https://docs.anaconda.com/anaconda/rlanguage-pkg-docs)



CONTINUED ON BACK →



File Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

anaconda - Google-Suche

https://www.google.de/search?q=anaconda&ie=utf-8&oe=utf-8&client=firefox-b&gfe\_rd=cr&dcr=0&eis=ELbkWY3FLOLDXPlm8gE#q=anaconda

Google anaconda

Anmelden

Alle Suche Videos Bilder News Shopping Mehr Einstellungen Tools

Ungefähr 43.100.000 Ergebnisse (0,83 Sekunden)

**Anaconda Cloud**  
<https://anaconda.org/> [Diese Seite übersetzen](#)  
 Where packages, notebooks, projects and environments are shared. Powerful collaboration and package management for open source and private projects.  
[Conda](#) · [Forgot Username](#) · [Anaconda ...](#) · [Login](#) · [Anaconda Cloud](#) · [Pricing](#)


**Python :: Anaconda Cloud**  
<https://anaconda.org/anaconda/python> [Diese Seite übersetzen](#)  
 Description. Python is a widely used high-level, general-purpose, interpreted, ...

**Große Anakonda – Wikipedia**  
[https://de.wikipedia.org/wiki/Große\\_Anakonda](https://de.wikipedia.org/wiki/Große_Anakonda)  
 Die Große Anakonda (Eunectes murinus) ist eine Schlangenart aus der Familie der Boas .... Rivas, M. C. Muñoz, G. M. Burghardt, J. B. Thorbjarnarson: Sexual size dimorphism and mating system of the Green **Anakonda** (Eunectes murinus).  
[Beschreibung](#) · [Verhalten](#) · [Jagdweise und Nahrung](#) · [Fortpflanzung](#)

**Anakonda (Film) – Wikipedia**  
[https://de.wikipedia.org/wiki/Anakonda\\_\(Film\)](https://de.wikipedia.org/wiki/Anakonda_(Film))  
**Anakonda** (Alternativtitel: **Anakonda** - It Will Take Your Breath Away) ist ein US-amerikanischer Spielfilm aus dem Jahr 1997.  
**Altersfreigabe:** FSK 16 **Regie:** Luis Llosa  
**Länge:** ca. 89 Minuten **Musik:** Ice Cube, Randy Edelman

**Anakonda (Python distribution) - Wikipedia**  
[https://en.wikipedia.org/wiki/Anakonda\\_\(Python\\_distribution\)](https://en.wikipedia.org/wiki/Anakonda_(Python_distribution)) [Diese Seite übersetzen](#)  
**Anakonda** is a freemium open source distribution of the Python and R programming languages for large-scale data processing, predictive analytics, and ...  
**License:** [New BSD License](#) **Developer(s):** Continuum analytics

**Anakonda (Mantel) - Wikipedia**






 Mehr Bilder

**Große Anakonda**  
 Schlange

Die Große Anakonda ist eine Schlangenart aus der Familie der Boas und wird dort in die Unterfamilie der Boaschlangen gestellt. Sie ist eine der größten Schlangen der Welt. Große Anakondas bewohnen die Tropen Südamerikas. [Wikipedia](#)

**Wissenschaftlicher Name:** Eunectes murinus  
**Höhere Klassifizierung:** [Anakondas](#)  
**Rang:** Art  
**Länge:** Weibchen: 4,6 m (Erwachsener), Männlich: 3 m (Erwachsener)

**Andere suchten auch nach** Über 15 weitere ansehen

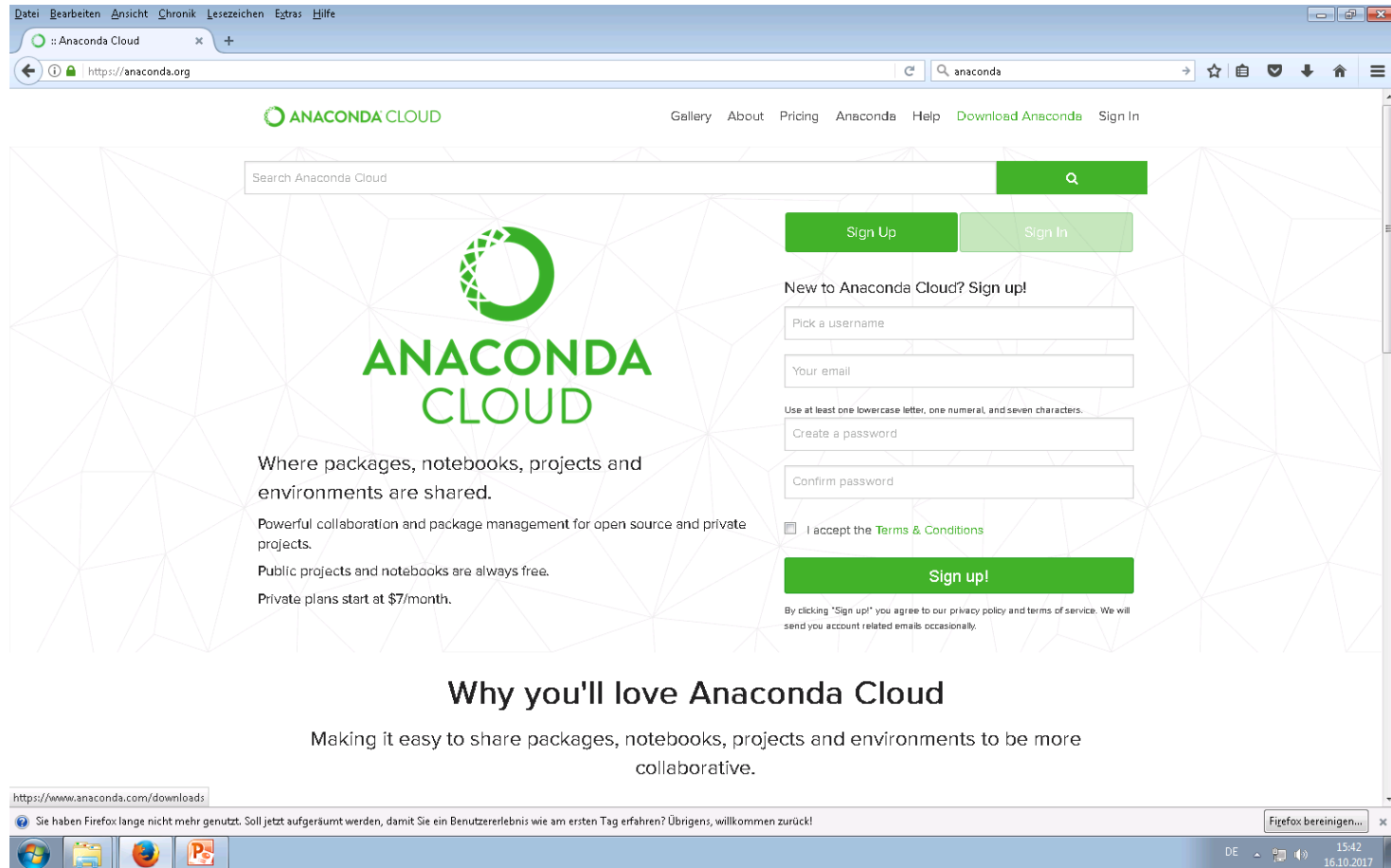
 Schlange  Anakondas  Pythons  Boas  Eigentliche Pythons

Feedback

Sie haben Firefox lange nicht mehr genutzt. Soll jetzt aufgeräumt werden, damit Sie ein Benutzererlebnis wie am ersten Tag erfahren? Übrigens, willkommen zurück!

Firefox bereinigen...




DE 15:41 16.10.2017





Download Anaconda Distribution

Version 5.0.0 | Release Date: September 26, 2017

Download For:   

High-Performance Distribution

Easily install 1,000+ [data science packages](#)

Package Management

Manage packages, dependencies and environments with [conda](#)

Portal to Data Science

Uncover insights in your data and create interactive visualizations

Windows macOS Linux

Anaconda 5.0.0 For Windows Installer

Python 3.6 version \*

[Download](#)

64-Bit Graphical Installer (53.5 MB) ⓘ  
32-Bit Graphical Installer (43.6 MB)

Python 2.7 version \*

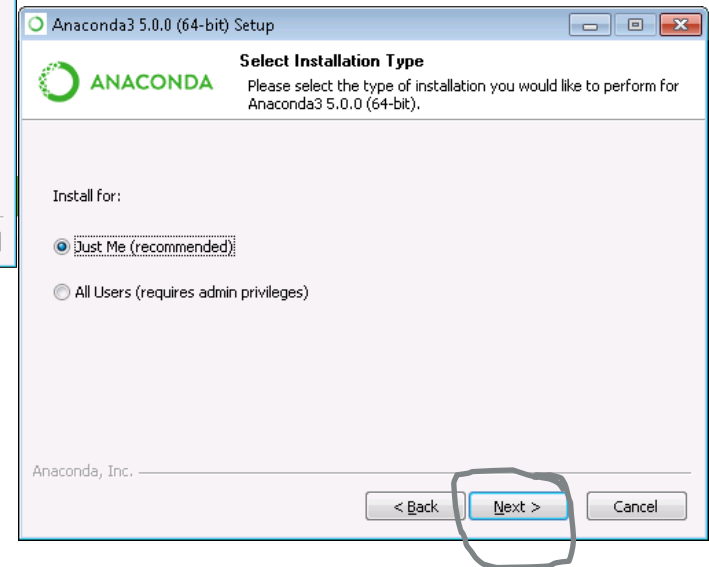
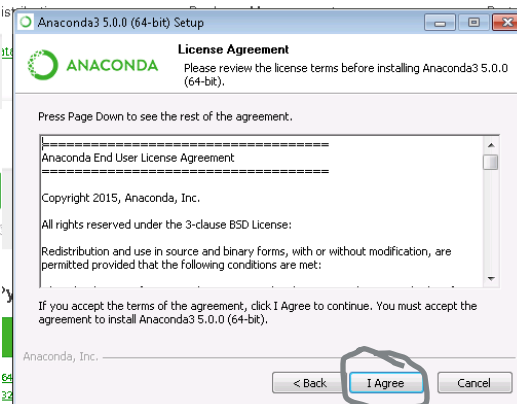
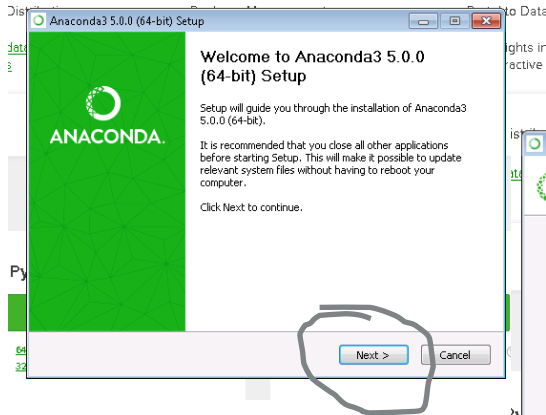
[Download](#)

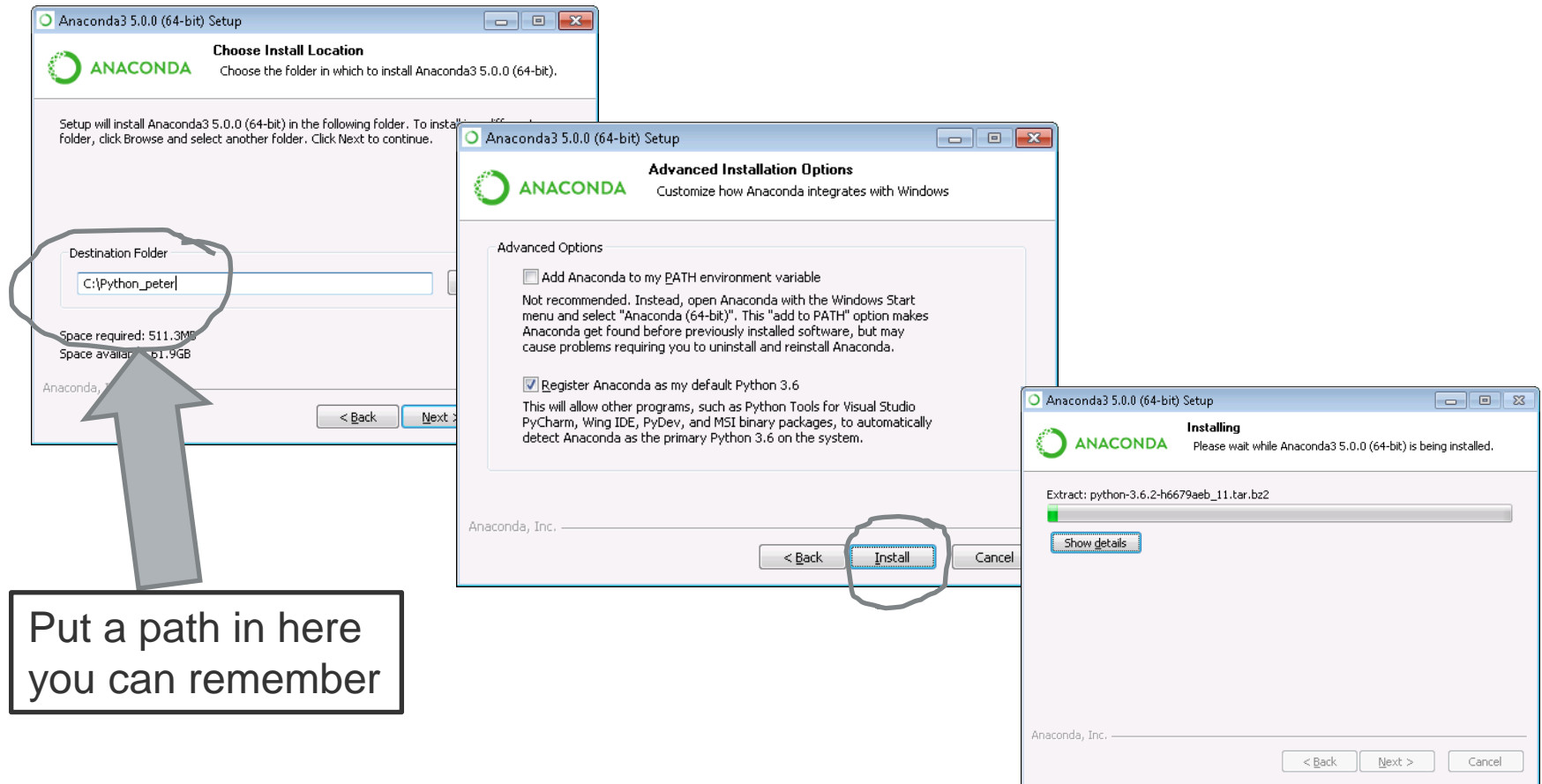
64-Bit Graphical Installer (52.2 MB) ⓘ  
32-Bit Graphical Installer (42.1 MB)

Sie haben Firefox lange nicht mehr genutzt. Soll jetzt aufgeräumt werden, damit Sie ein Benutzererlebnis wie am ersten Tag erfahren? Übrigens, willkommen zurück!

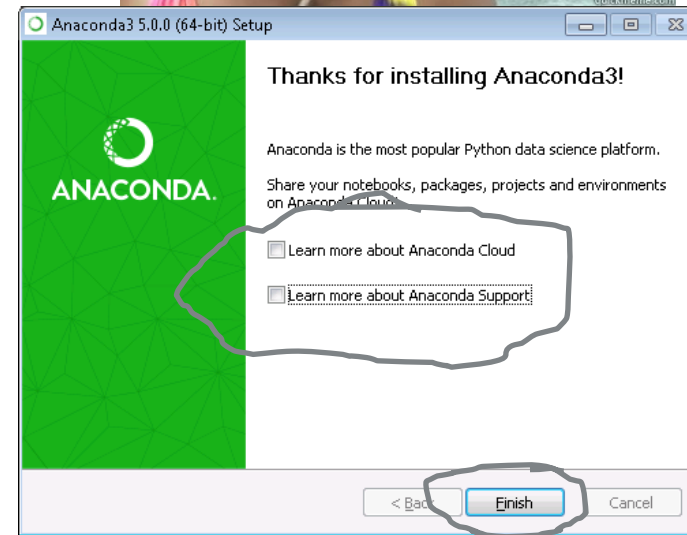
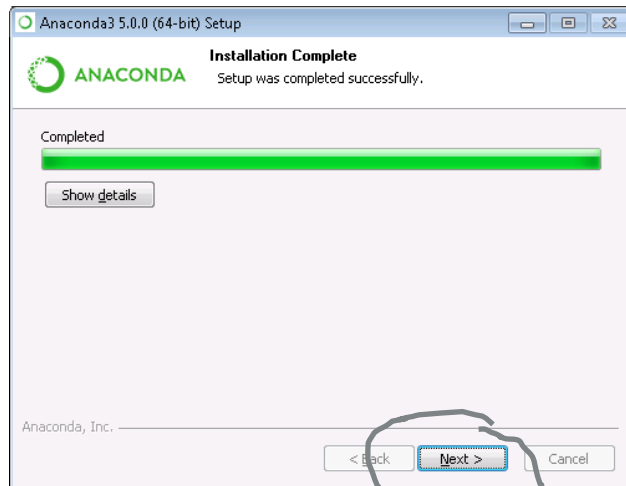
Firefox bereinigen...

15:42 16.10.2017














## ANACONDA NAVIGATOR CHEAT SHEET

See full documentation for Anaconda Navigator  
[docs.anaconda.com/anaconda/navigator/](https://docs.anaconda.com/anaconda/navigator/)

**Before you Start**

**What is Anaconda Navigator?**

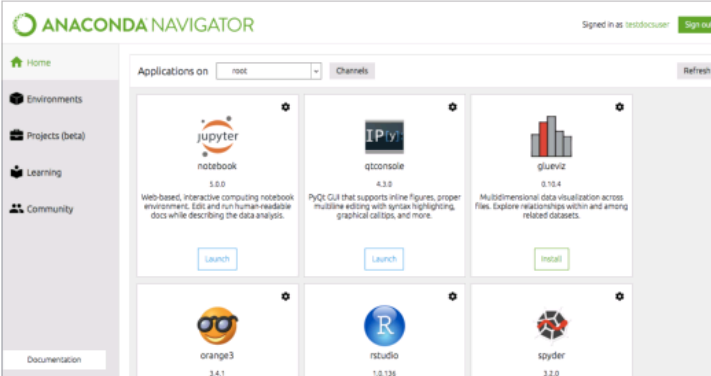
**Get It**

**Will it work on my machine?**

**Follow the graphical install instructions**

**Open Anaconda Navigator**

**NOW PLAY WITH THE WORLD'S MOST AWESOME SCIENTIFIC PACKAGES**

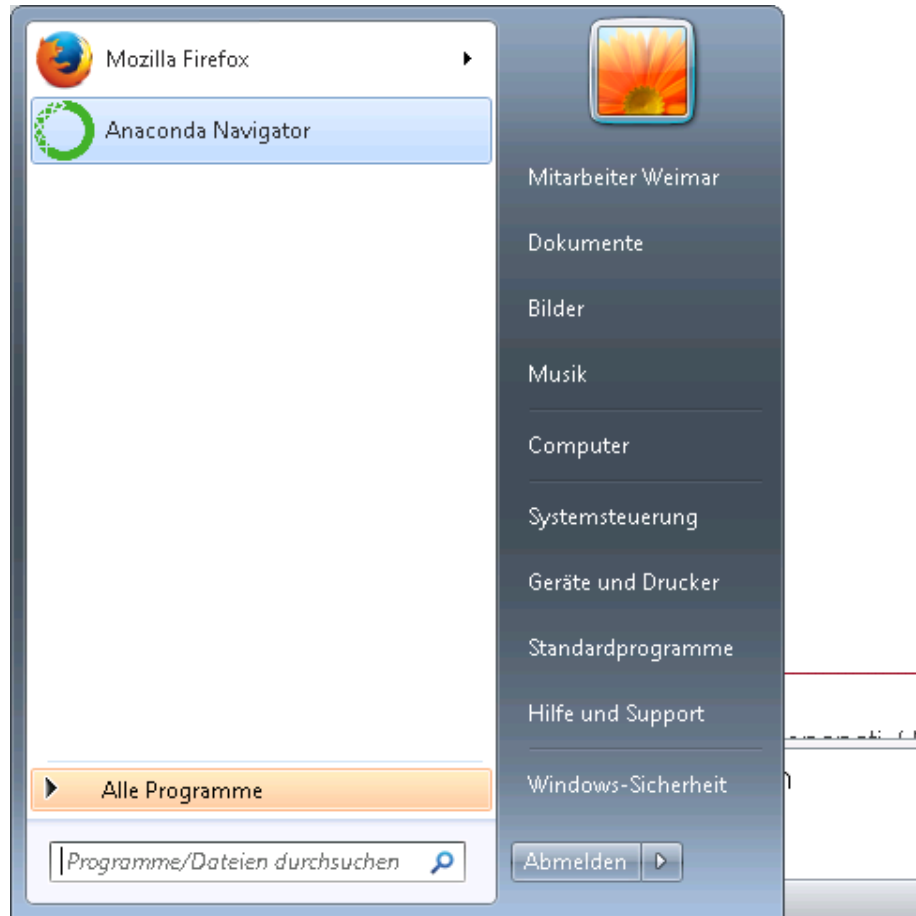


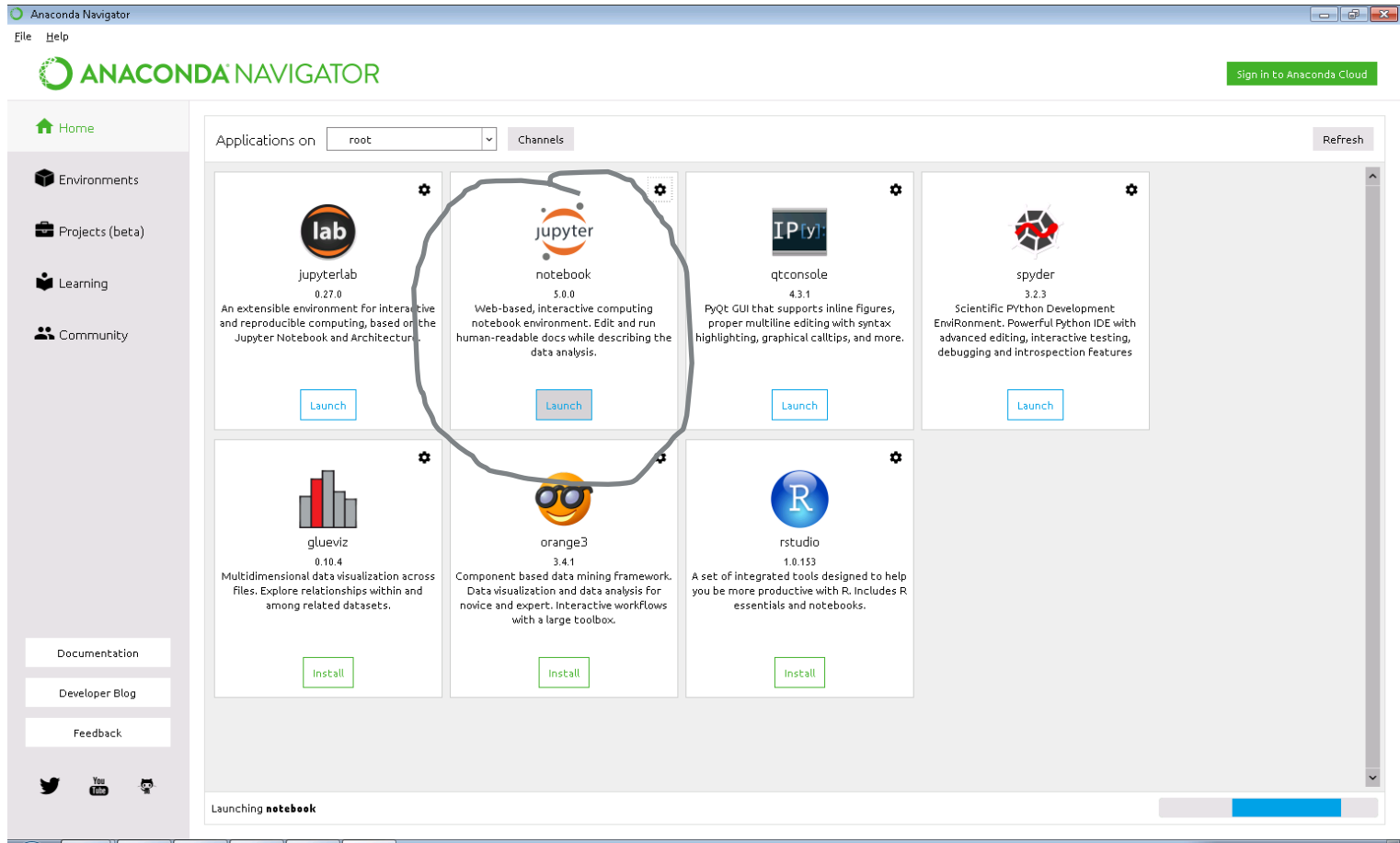
**MORE RESOURCES**

Free email group support	<a href="http://bit.ly/anaconda-community">http://bit.ly/anaconda-community</a>
Paid support	<a href="https://anaconda.com/anaconda-support">anaconda.com/anaconda-support</a>
Training	<a href="https://anaconda.com/training">anaconda.com/training</a>
Consulting	<a href="https://anaconda.com/anaconda-consulting">anaconda.com/anaconda-consulting</a>

Follow us on Twitter [@anacondainc](https://twitter.com/anacondainc) and join the [#AnacondaCrew](https://twitter.com/AnacondaCrew)

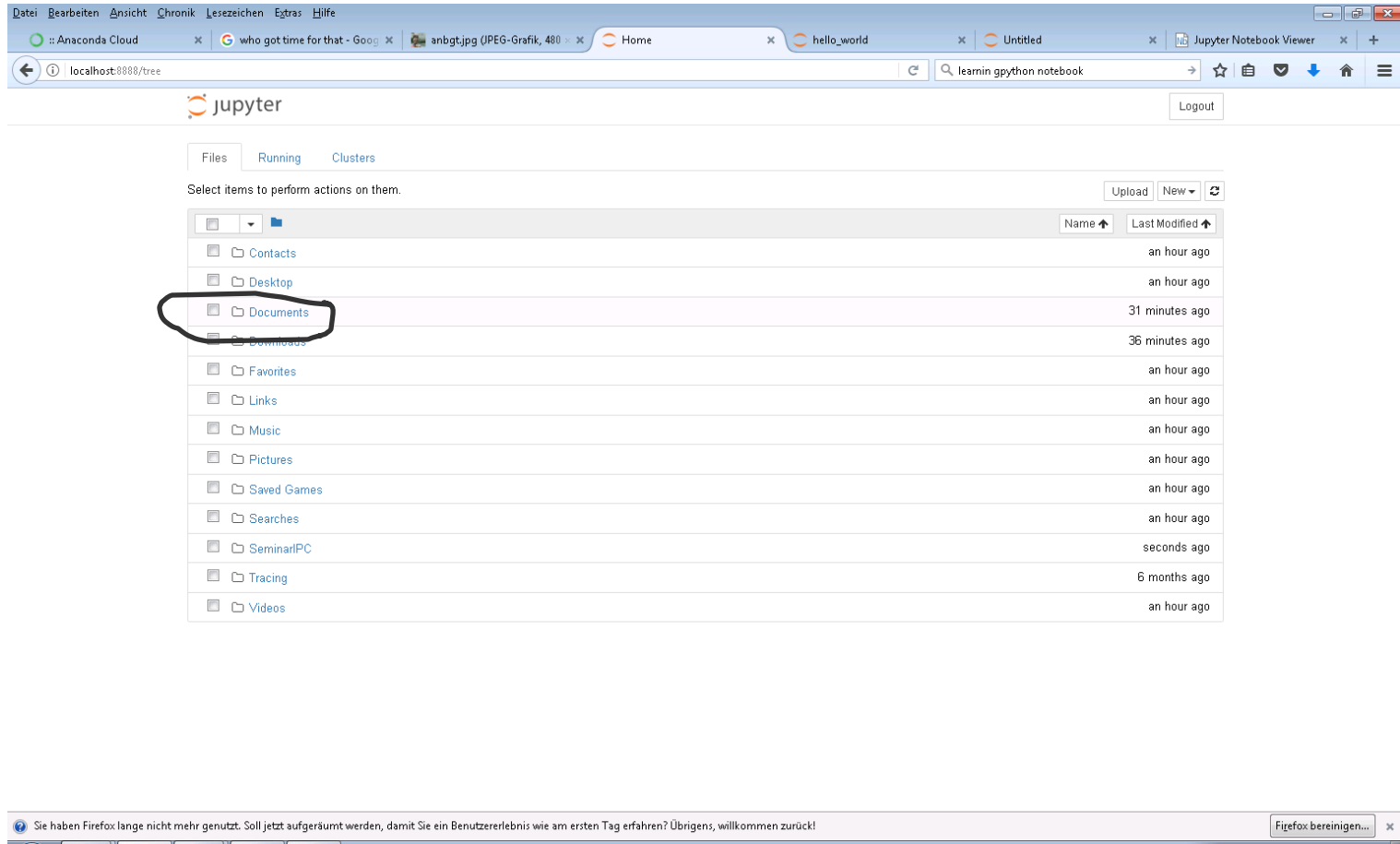
Connect with talented, like-minded data scientists and developers while contributing to the open source movement. Visit [anaconda.com/community](https://anaconda.com/community).





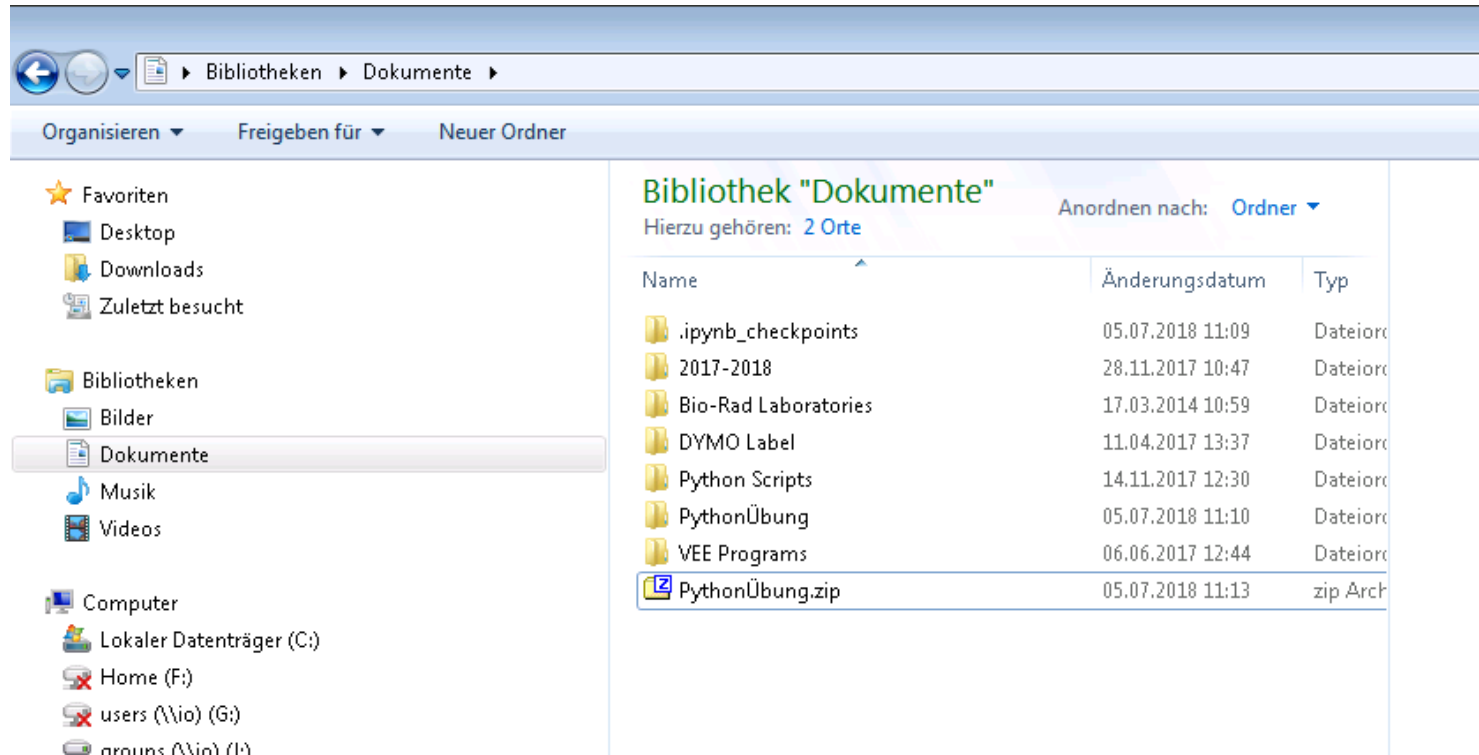


# The Browser will show a representation of your “Home-Folder”



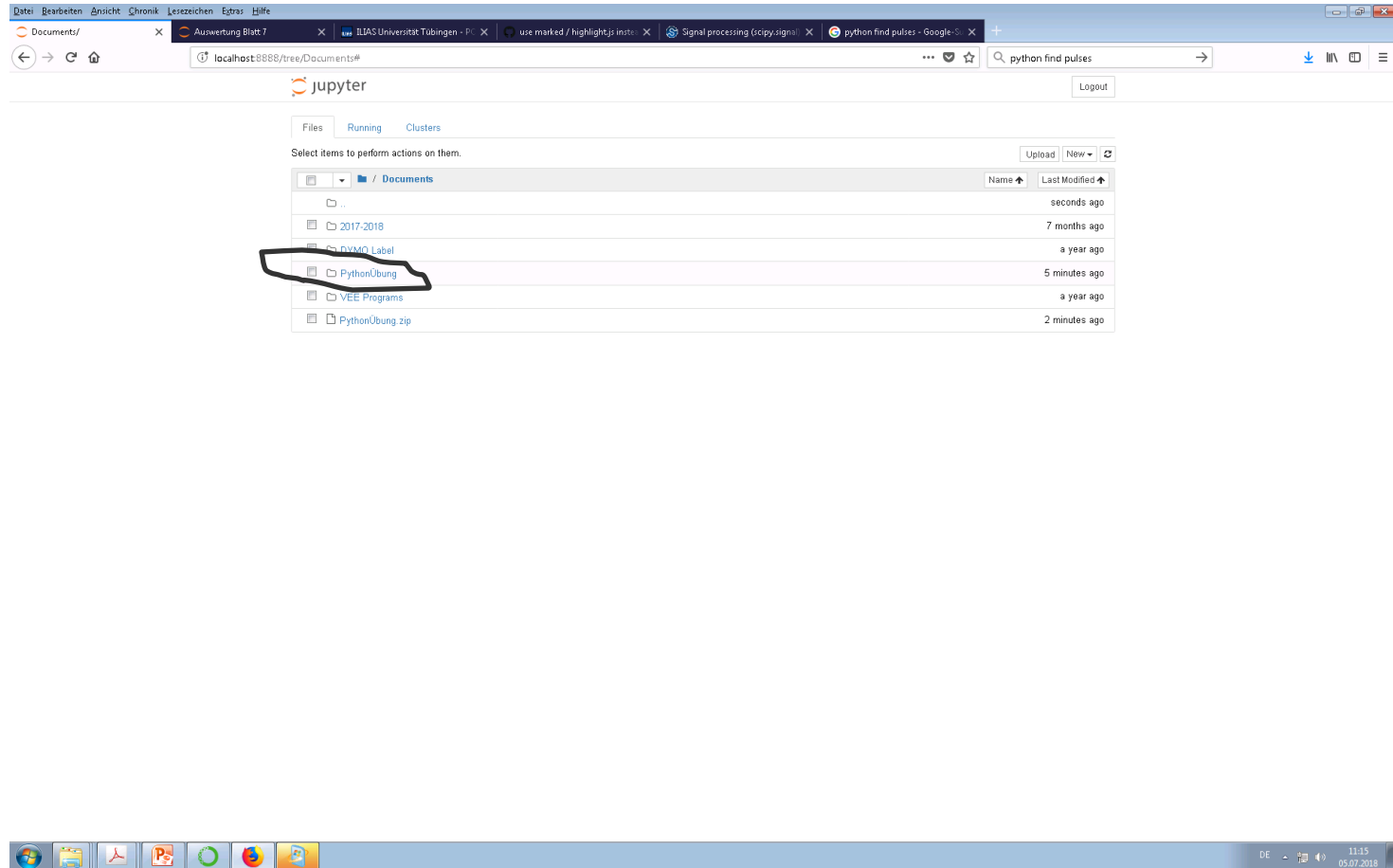


# Find and extract the file “PythonÜbungen.zip” in your file browser



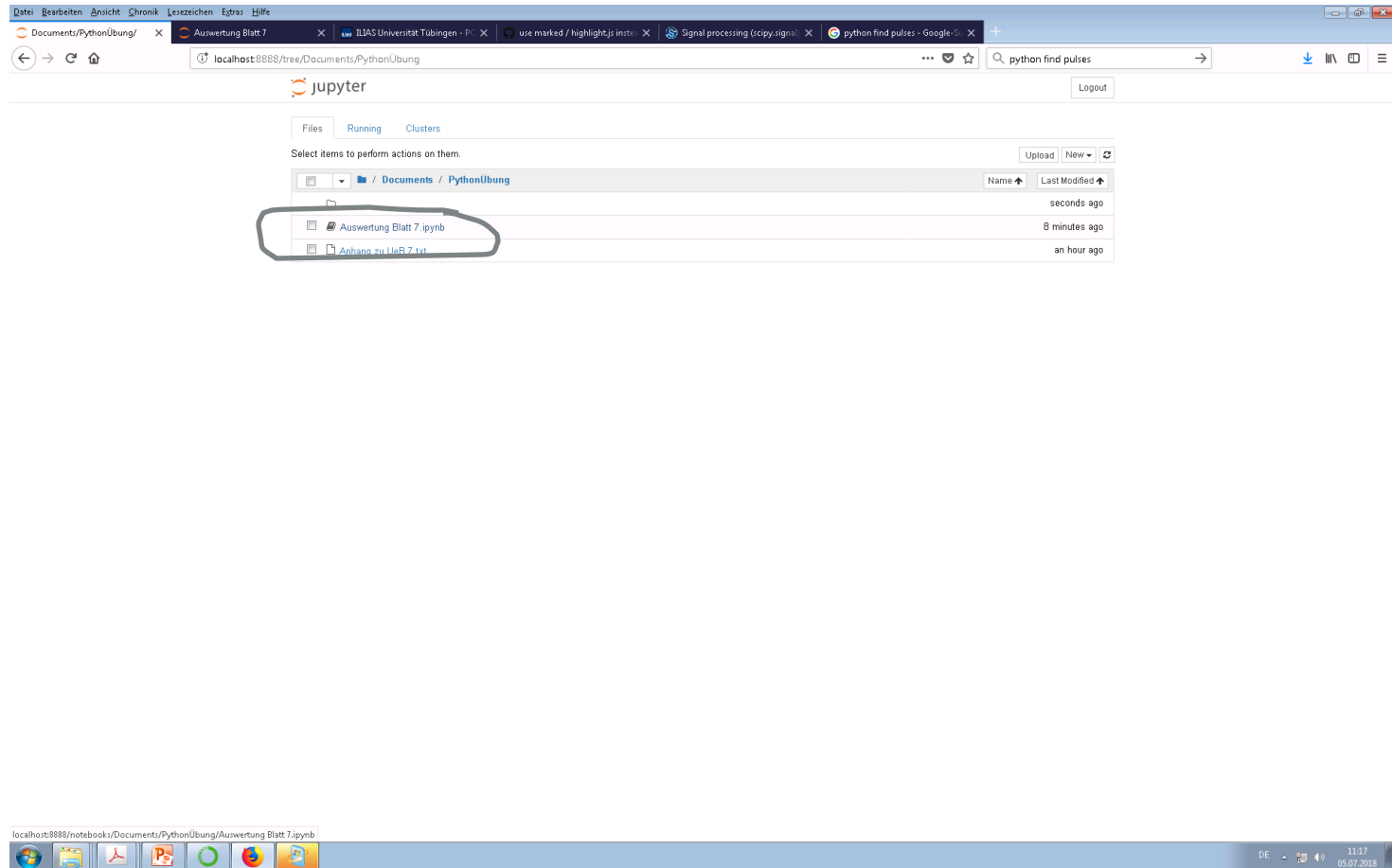


# Int the web browser navigate to the extracted files.





**Click on the file with the “ipynb” ending to open it.**



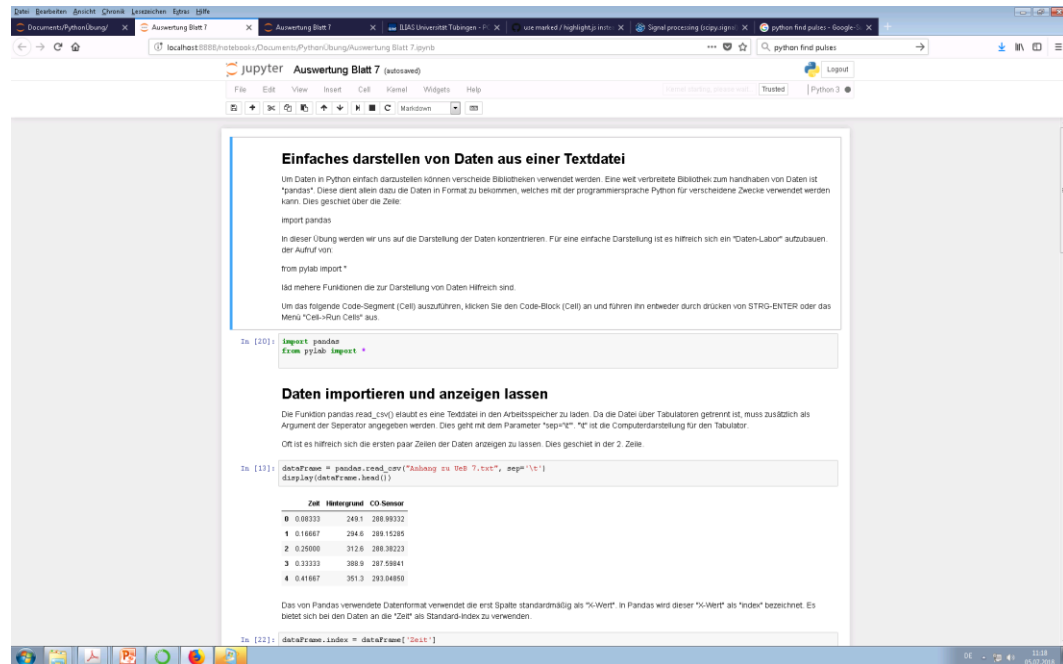
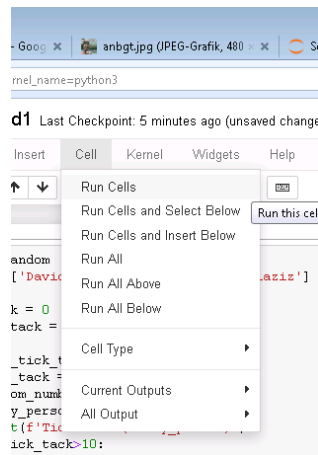




# The program/script is structured in different cells.

Besides the code you will also find explanations and results in the notebook.  
You can execute the code by (“Cell->Run Cells”) or by pushing “CTRL-ENTER” on your keyboard.

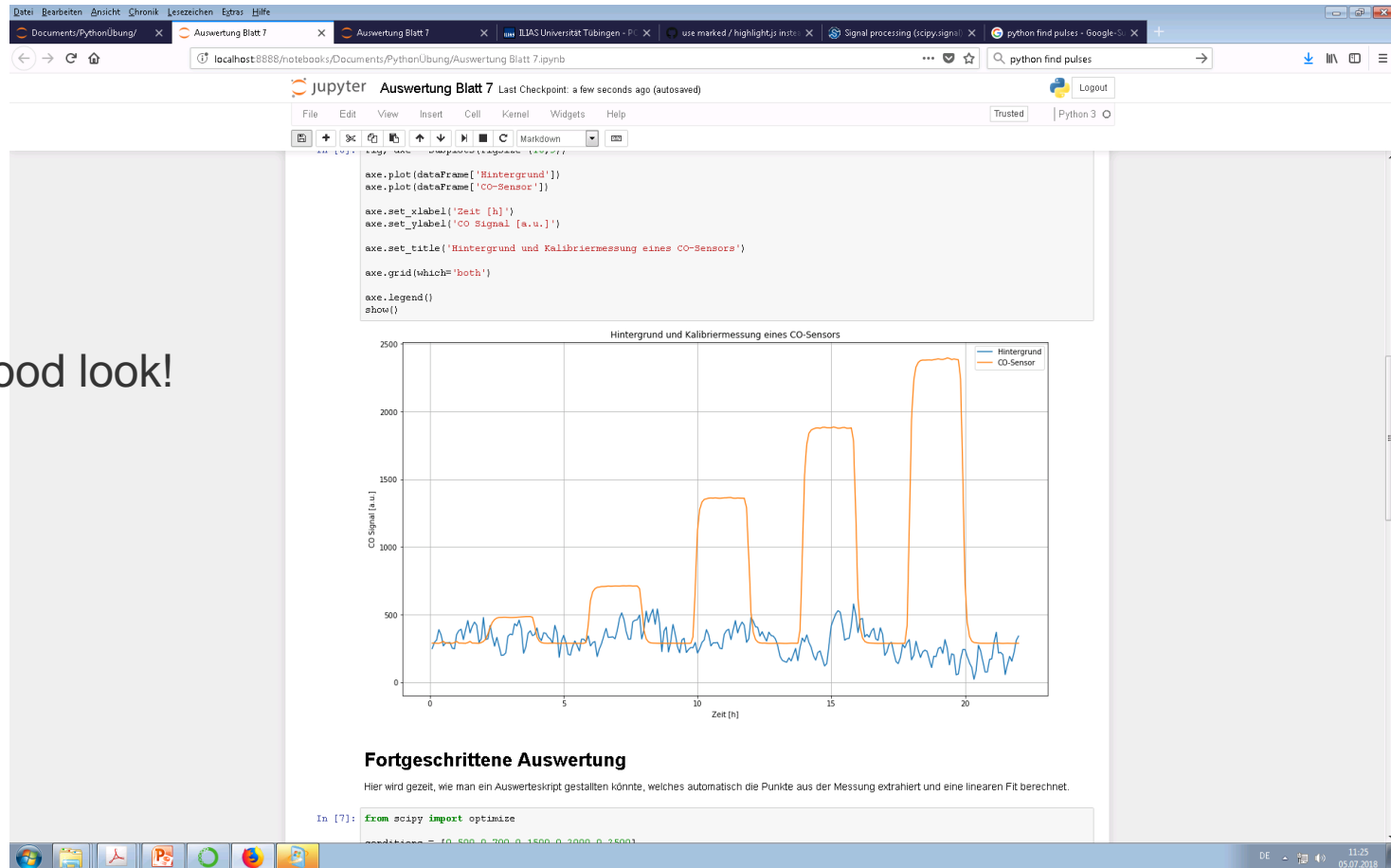
Run some cells and see what is happening.





# Try to get the following plot on your own. You may also want to play a bit around with the code.

Good look!



## 27 | Peter Bonanati / Numerisches Modell





# Thank you.

Contact: Peter Bonanati

**Institute of Physical and Theoretical  
Chemistry / AG Weimar**

Auf der Morgenstelle 15, 72076 Tübingen

Phone: +49 7071 29-77633

Fax: +49 7071 29-5960

[Peter.Bonanati@ipc.uni-tuebingen.de](mailto:Peter.Bonanati@ipc.uni-tuebingen.de)