EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

FACULTY OF
SCIENCE

Setting up a data mining machine

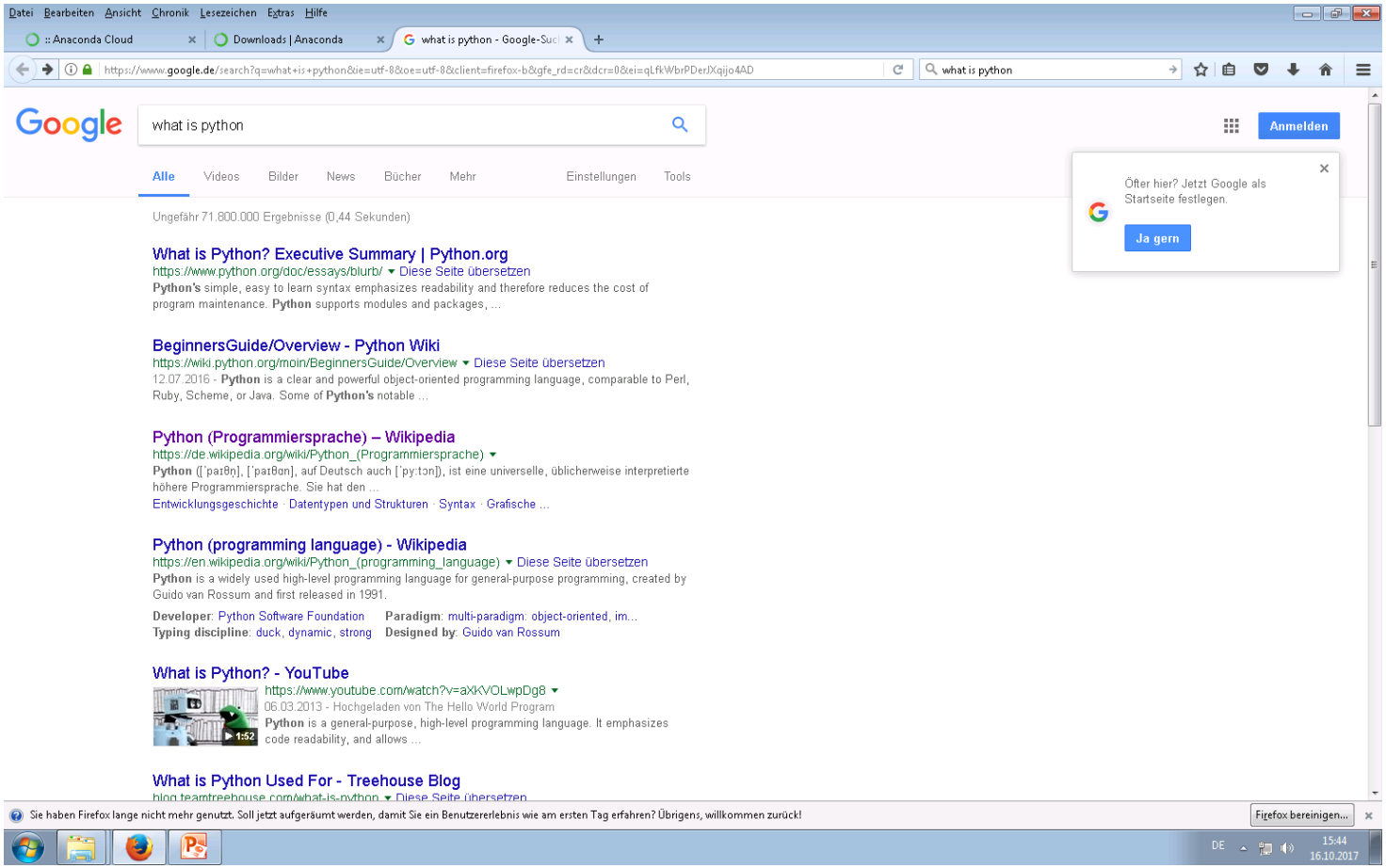
Installation of a Python & data mining toolkits

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 whishes 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 […]. The language provides constructs intended to enable writing clear programs on both a small and large scale.[25]” 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 that 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 us cases.
- Those step-by-step commands can be seen as cookbooks and are generally called **Libraries**

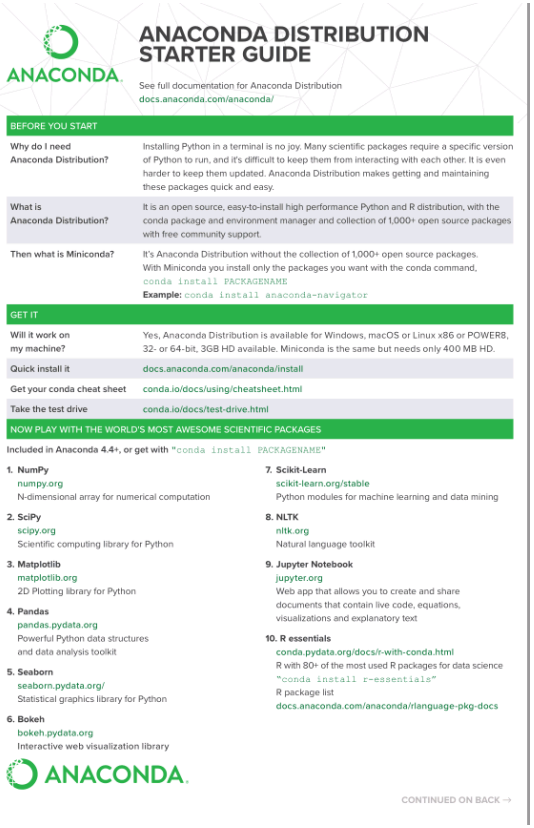
What is Python

- 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 decencies and bundles a well working package for us with most of the tools we will ever need.

Let’s start!

Installing Anaconda

ANACONDA



ANACONDA.

ANACONDA DISTRIBUTION STARTER GUIDE

See full documentation for Anaconda Distribution
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
Get your conda cheat sheet	conda.io/docs/using/cheatsheet.html
Take the test drive	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
   N-dimensional array for numerical computation
2. **SciPy**
   scipy.org
   Scientific computing library for Python
3. **Matplotlib**
   matplotlib.org
   2D Plotting library for Python
4. **Pandas**
   pandas.pydata.org
   Powerful Python data structures and data analysis toolkit
5. **Seaborn**
   seaborn.pydata.org/
   Statistical graphics library for Python
6. **Bokeh**
   bokeh.pydata.org
   Interactive web visualization library
7. **Scikit-Learn**
   scikit-learn.org/stable
   Python modules for machine learning and data mining
8. **NLTK**
   nltk.org
   Natural language toolkit
9. **Jupyter Notebook**
   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
    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

ANACONDA.

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
Get your conda cheat sheet	conda.io/docs/using/cheatsheet.html
Take the test drive	conda.io/docs/test-drive.html

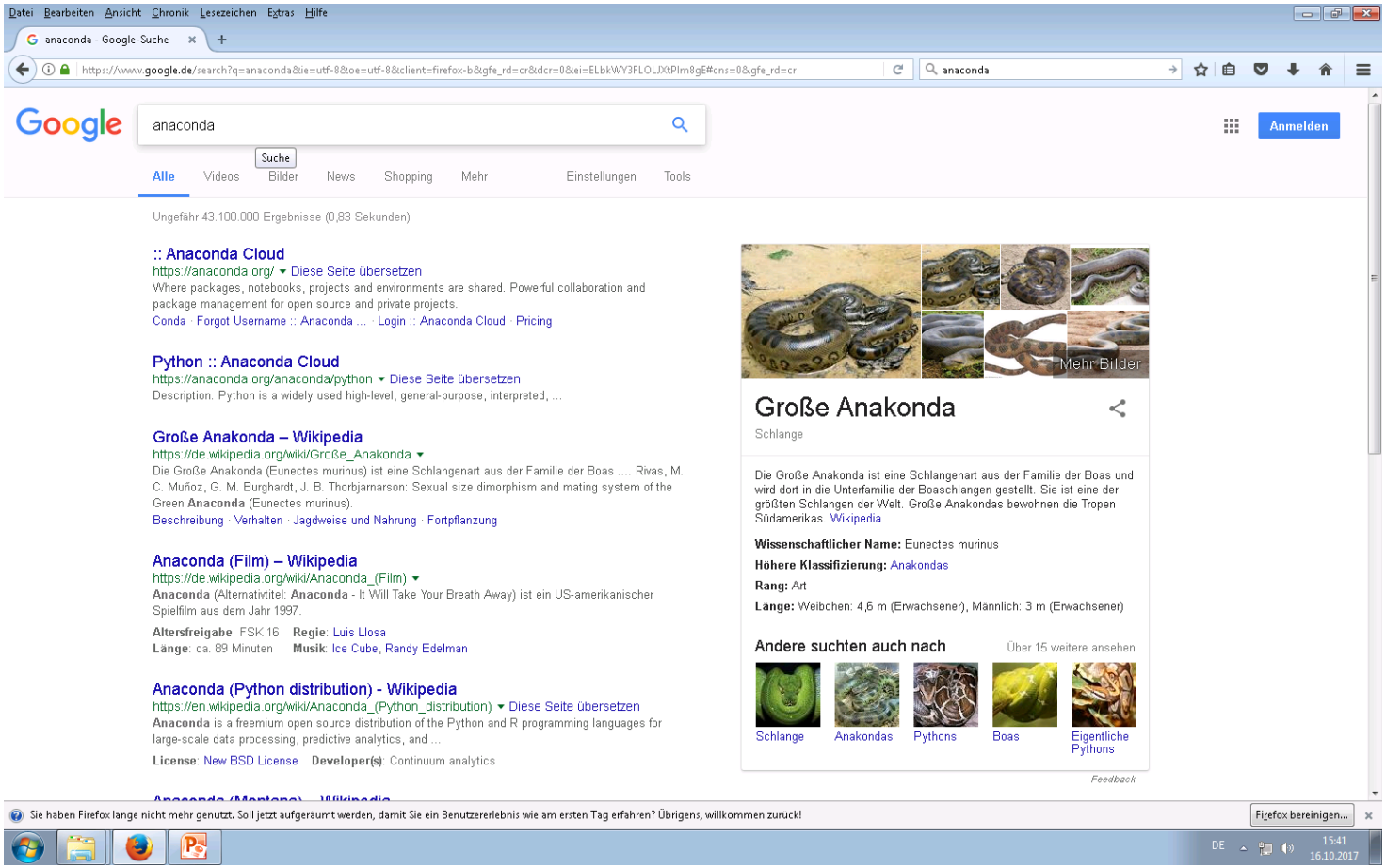
NOW PLAY WITH THE WORLD'S MOST AWESOME SCIENTIFIC PACKAGES

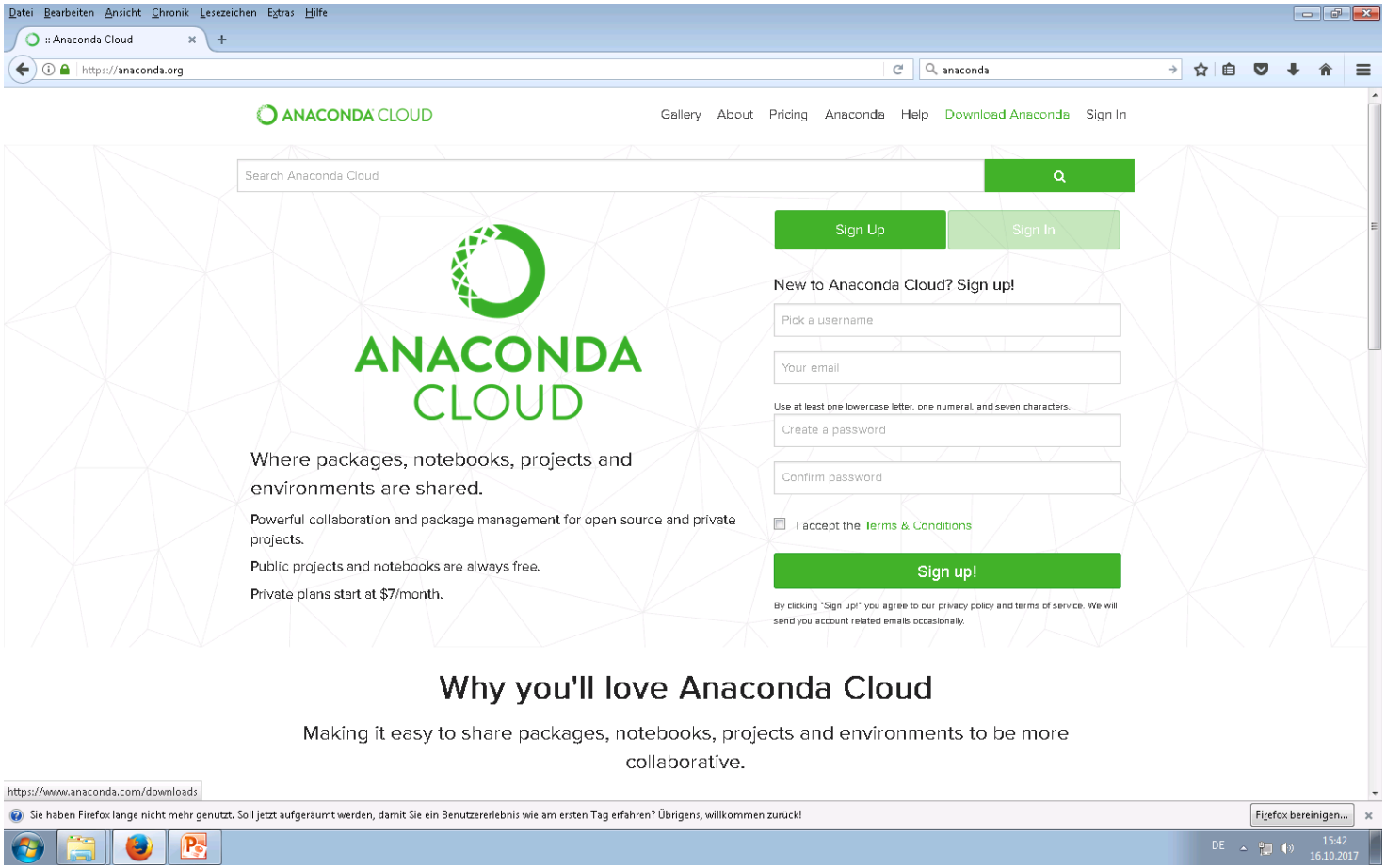
Included in Anaconda 4.4+, or get with `"conda install PACKAGENAME"`

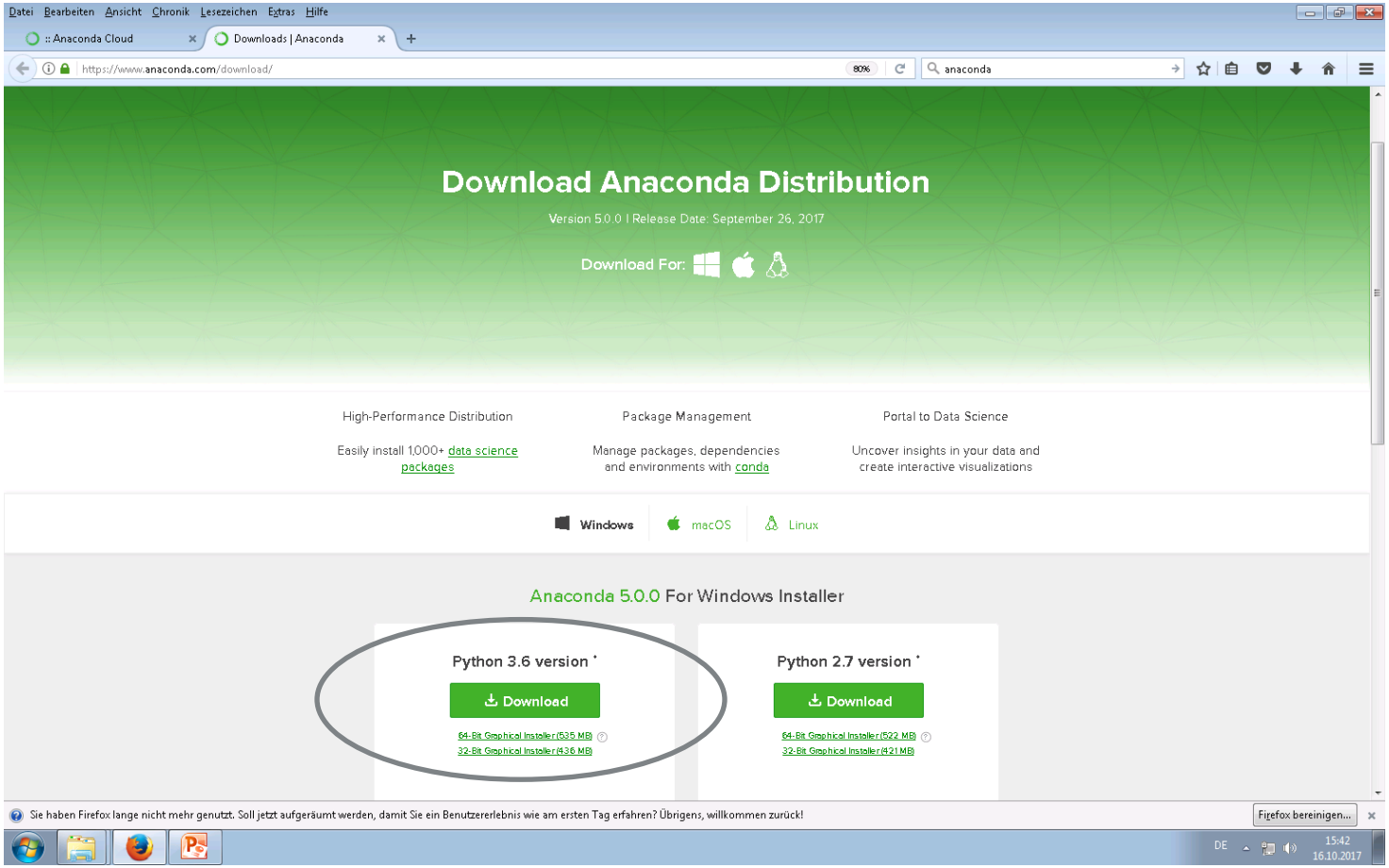
1. **NumPy**
   numpy.org
   N-dimensional array for numerical computation
2. **SciPy**
   scipy.org
   Scientific computing library for Python
3. **Matplotlib**
   matplotlib.org
   2D Plotting library for Python
4. **Pandas**
   pandas.pydata.org
   Powerful Python data structures and data analysis toolkit
5. **Seaborn**
   seaborn.pydata.org/
   Statistical graphics library for Python
6. **Bokeh**
   bokeh.pydata.org
   Interactive web visualization library
7. **Scikit-Learn**
   scikit-learn.org/stable
   Python modules for machine learning and data mining
8. **NLTK**
   nltk.org
   Natural language toolkit
9. **Jupyter Notebook**
   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
    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

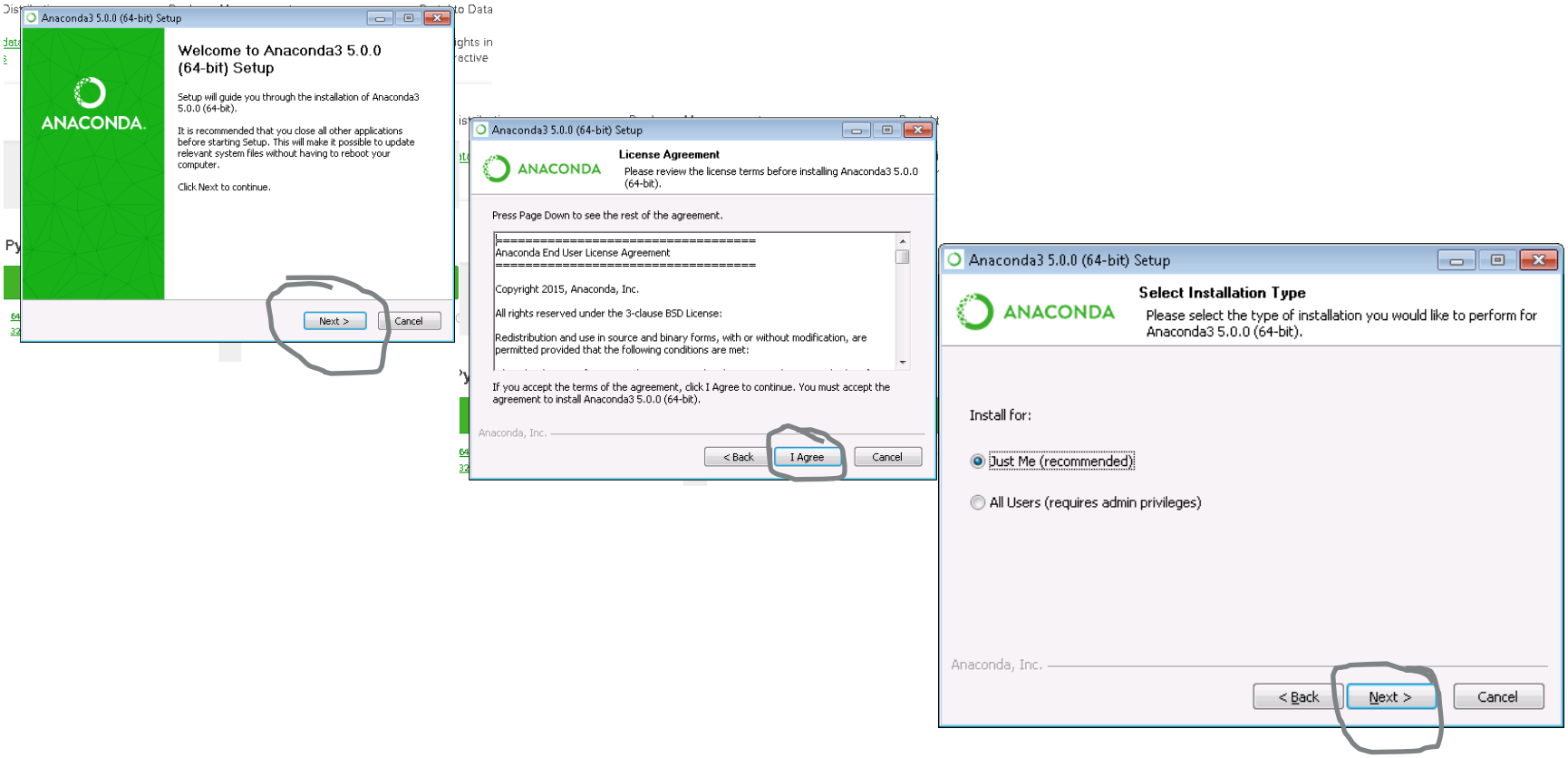
ANACONDA.

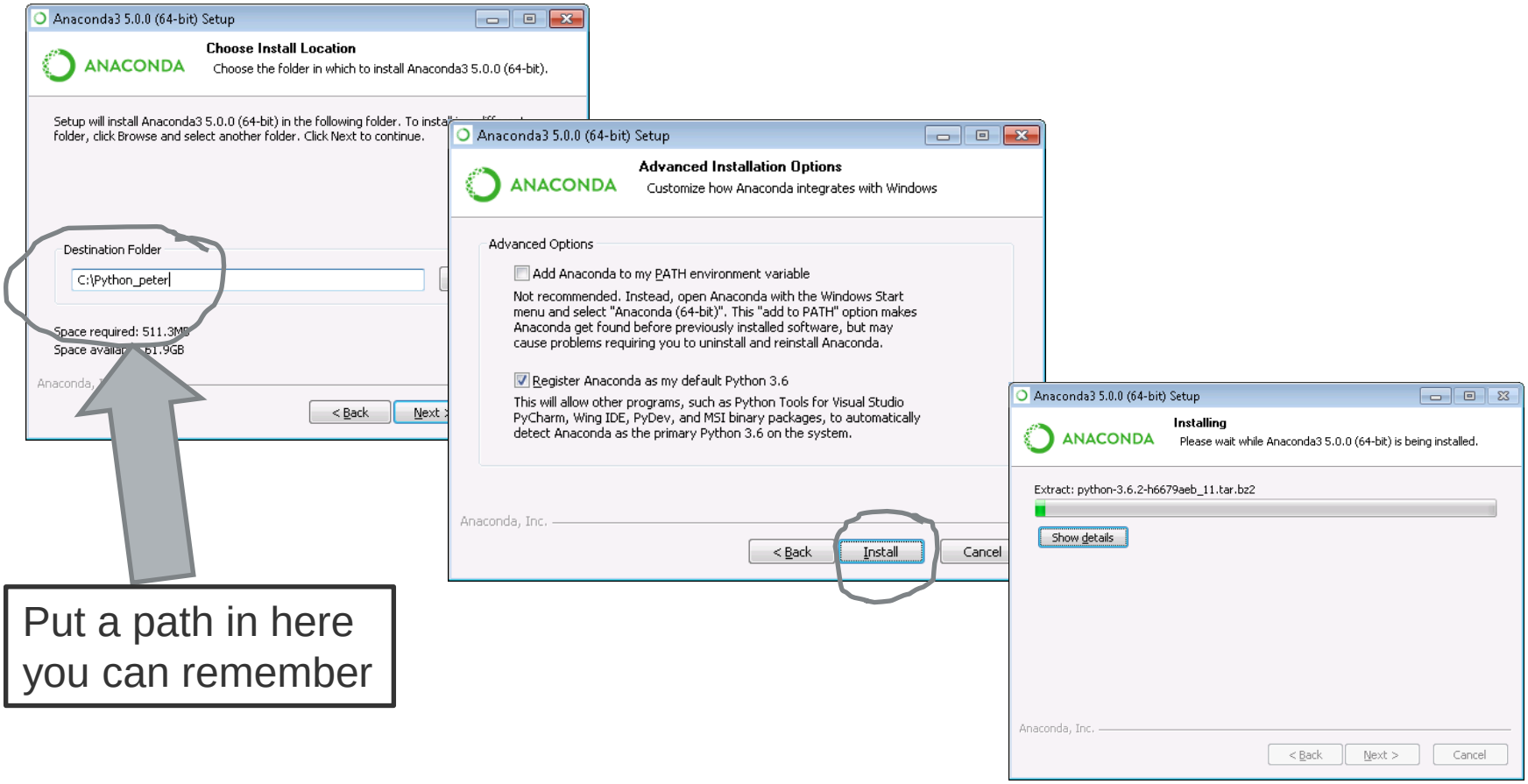
CONTINUED ON BACK →

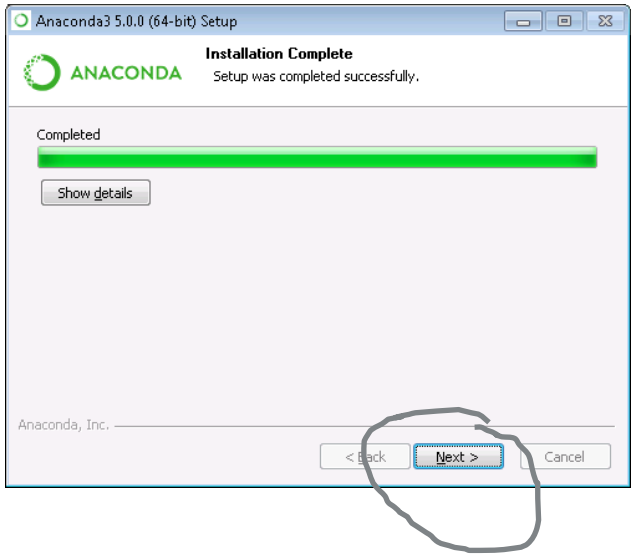


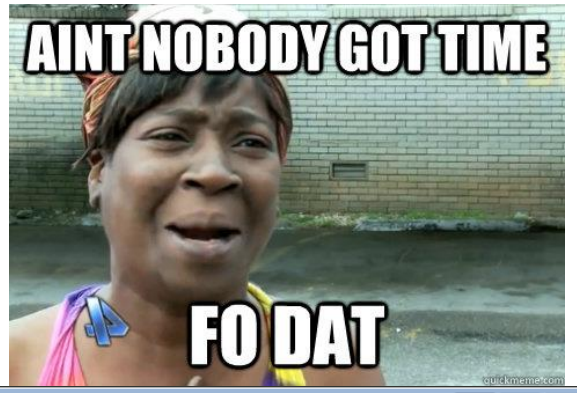


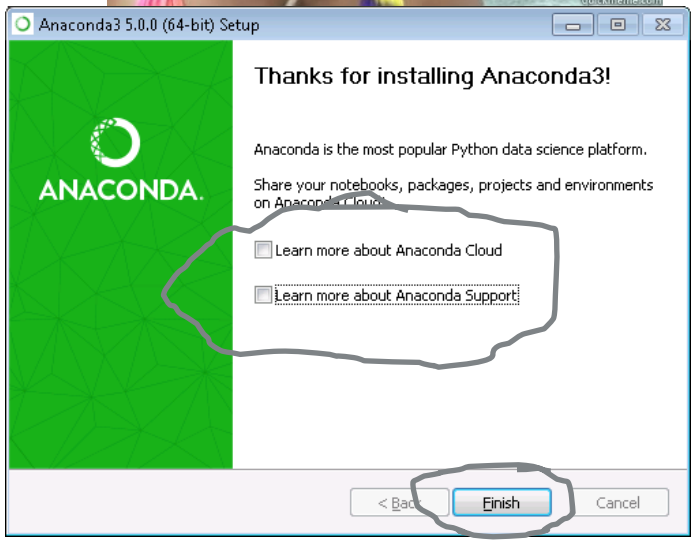












ANACONDA NAVIGATOR CHEAT SHEET

ANACONDA.

See full documentation for Anaconda Navigator
docs.anaconda.com/anaconda/navigator/

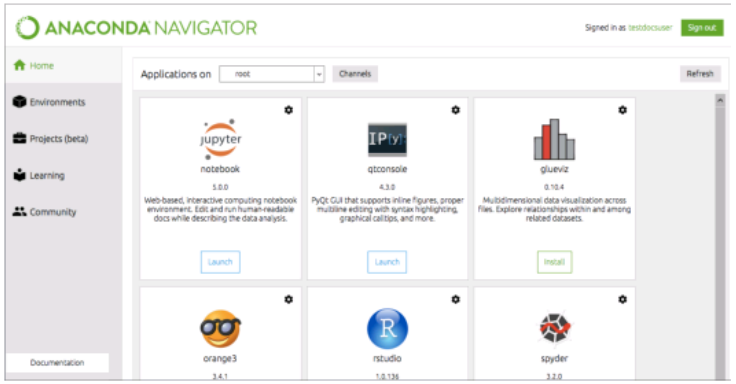
Before you Start

What is Anaconda Navigator?	Anaconda Navigator is an easy way to use graphical Python programs without having to use command line commands.

Get It

Will it work on my machine?	Anaconda Navigator is available for Windows, macOS or Linux, 32- or 64-bit, 3GB HD available. Navigator is automatically installed when you install Anaconda Distribution.
Follow the graphical install instructions	docs.anaconda.com/anaconda/install
Open Anaconda Navigator	After install, look on your desktop or programs menu for Anaconda Navigator and click it.

NOW PLAY WITH THE WORLD'S MOST AWESOME SCIENTIFIC PACKAGES



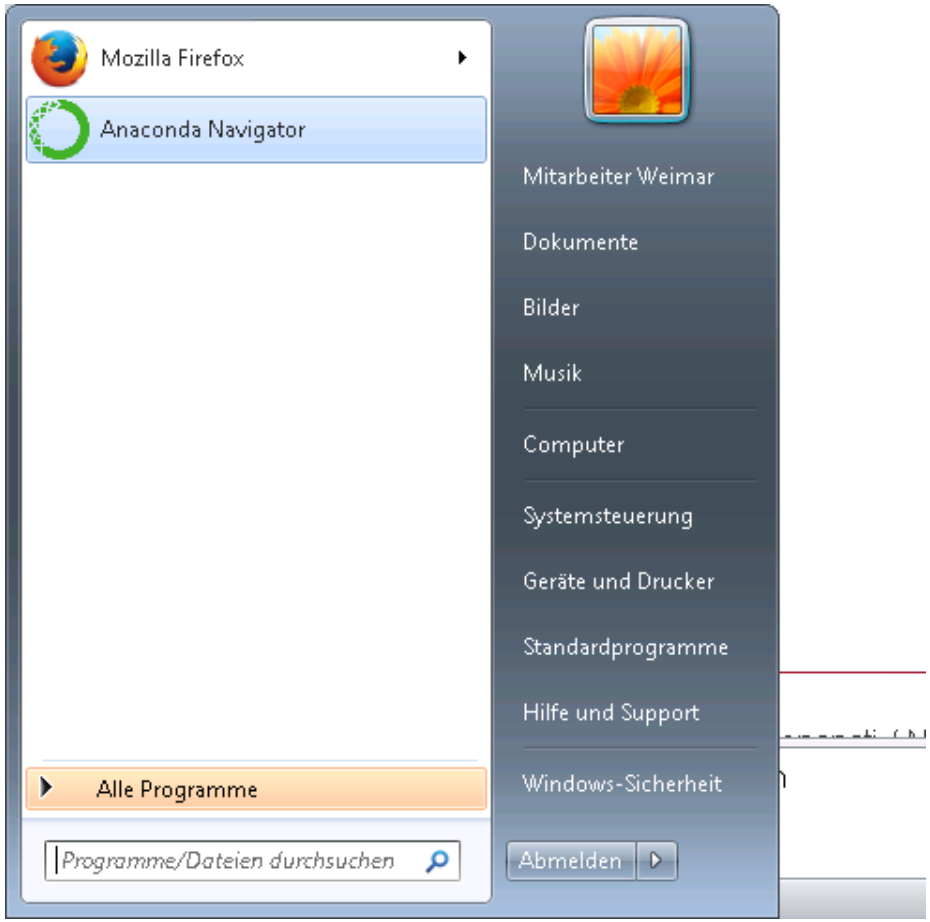
MORE RESOURCES

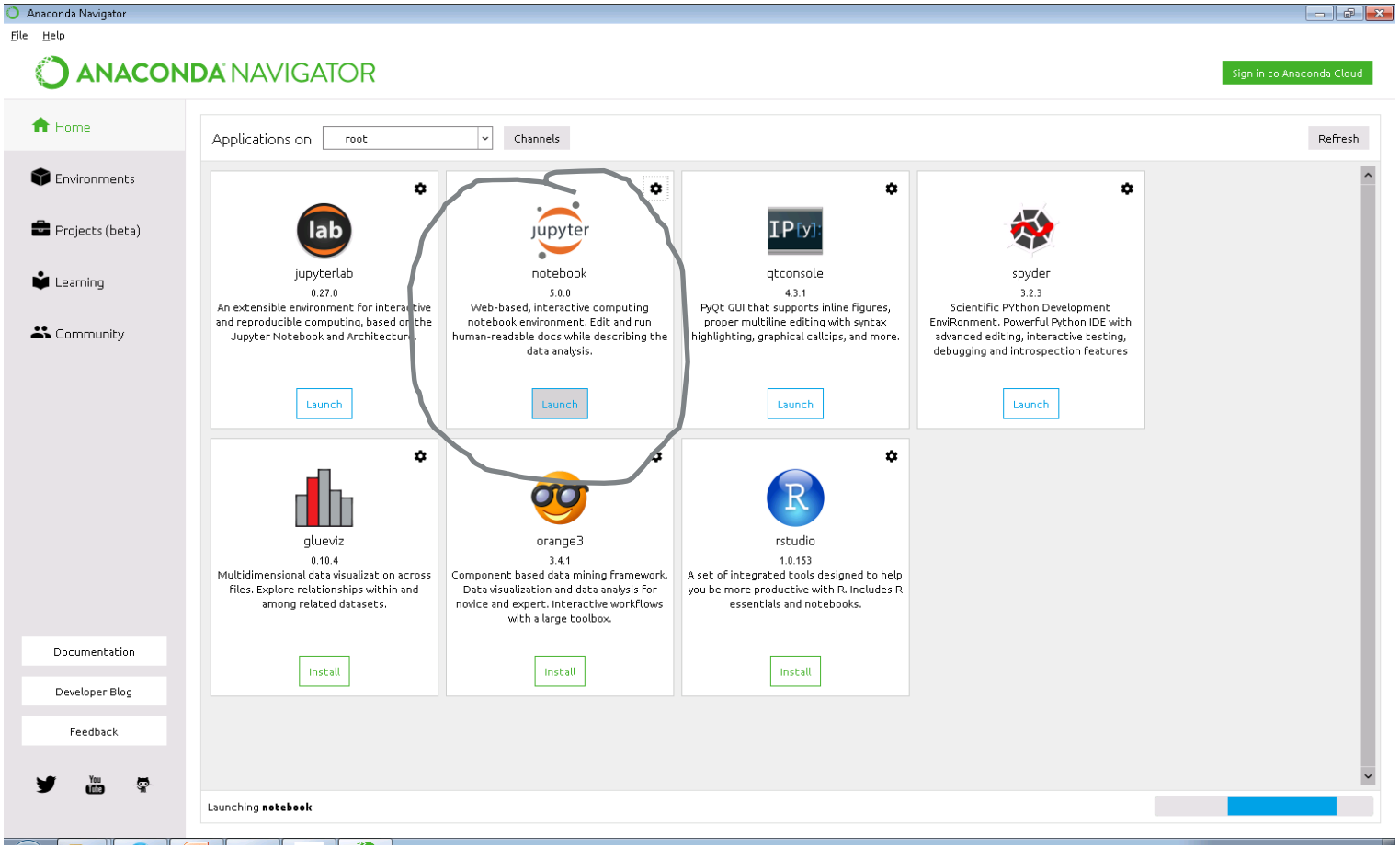
Free email group support	http://bit.ly/anaconda-community
Paid support	anaconda.com/anaconda-support
Training	anaconda.com/training
Consulting	anaconda.com/anaconda-consulting

Follow us on Twitter @anacondainc and join the #AnacondaCrew!

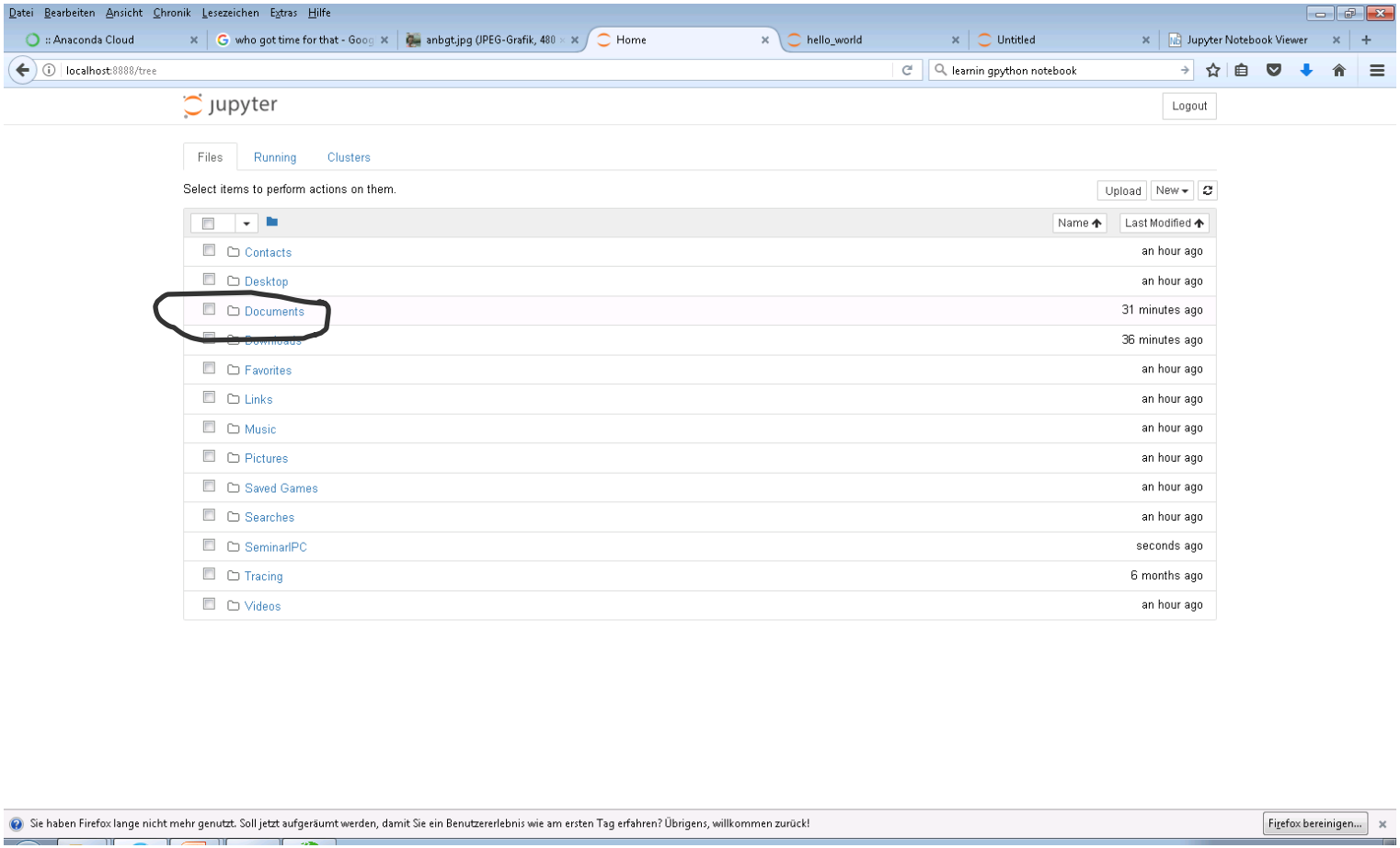
Connect with talented, like-minded data scientists and developers while contributing to the open source movement. Visit anaconda.com/community.

ANACONDA. anaconda.com · info@anaconda.com · 512-776-1066

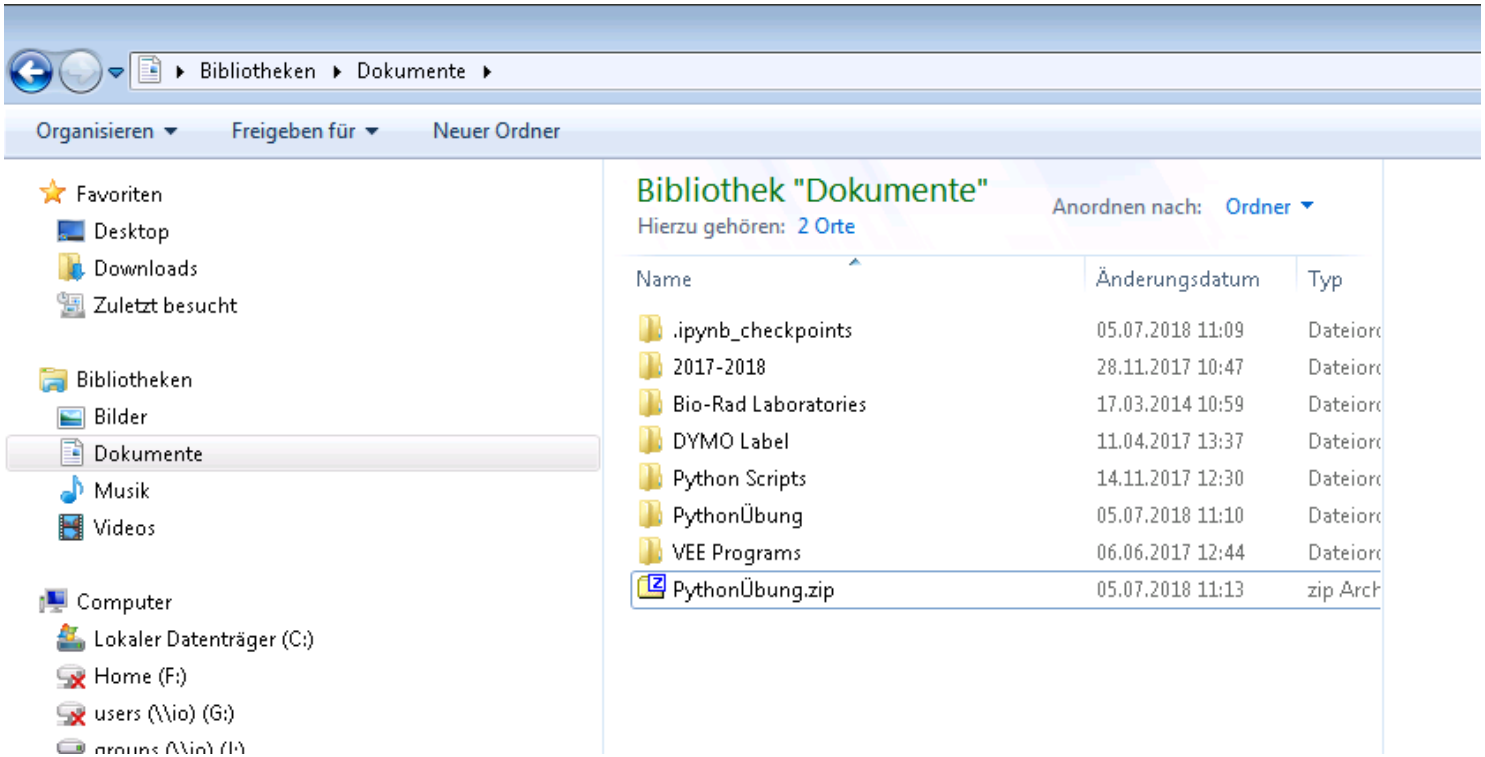




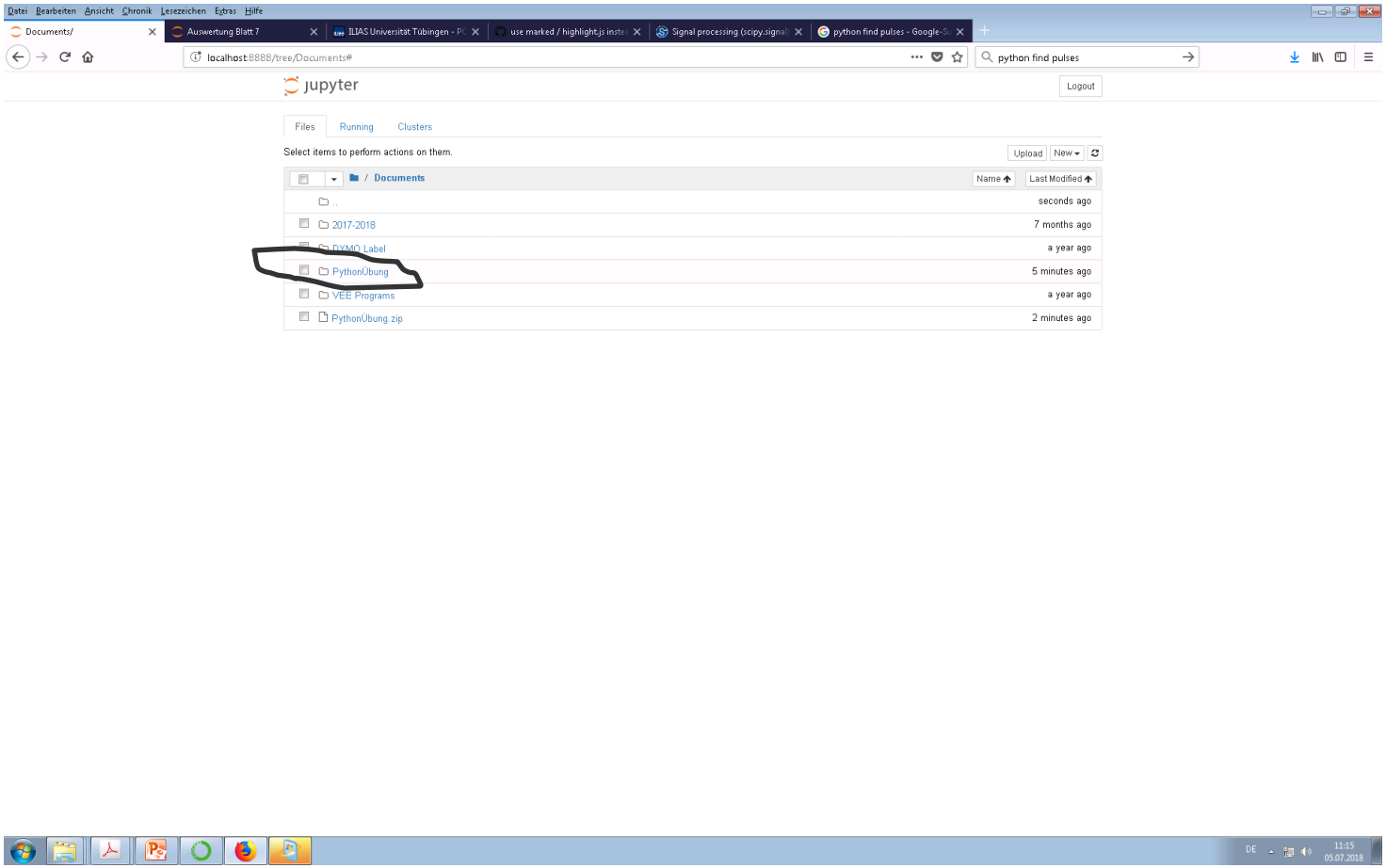
The Browser will show a representation of your “Home-Folder”



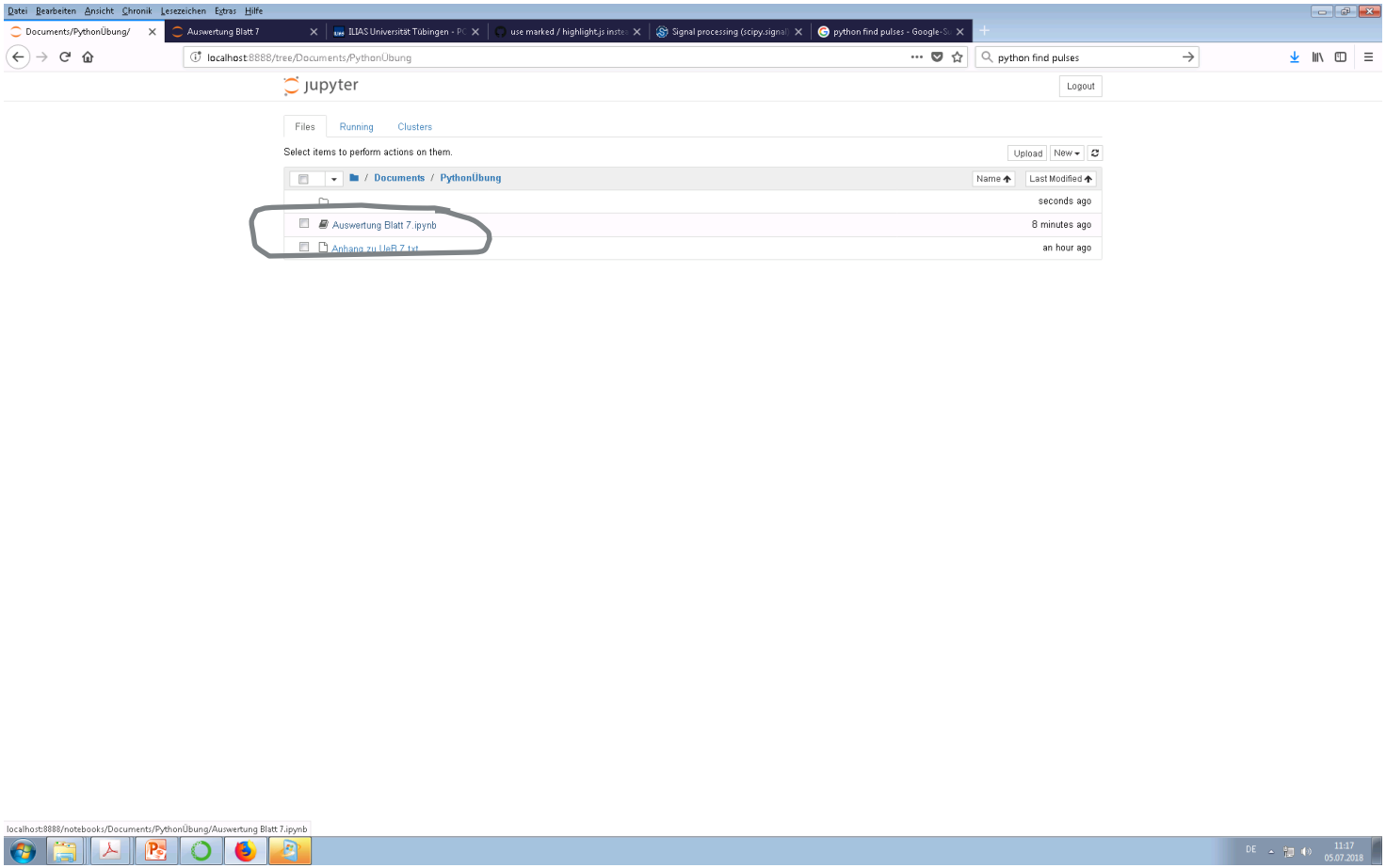
Find and extract the file “PythonÜbungen.zip” in your file browser



Int the web browser navigate to the extracted files.

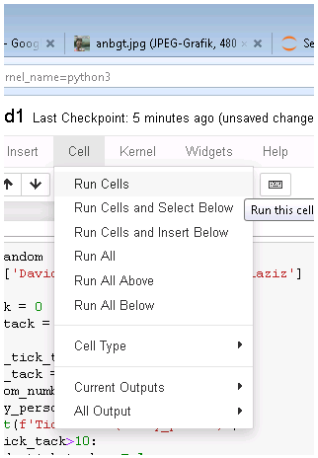


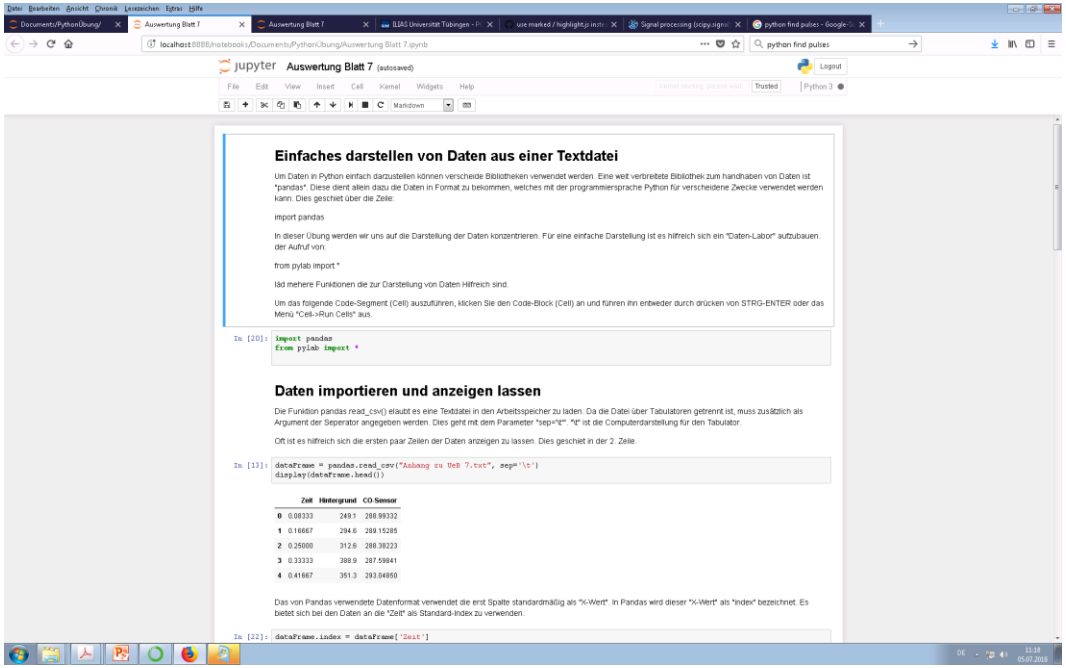
Click on the file with the “ipynb” ending to open it.



The program/script is structured in different cells.

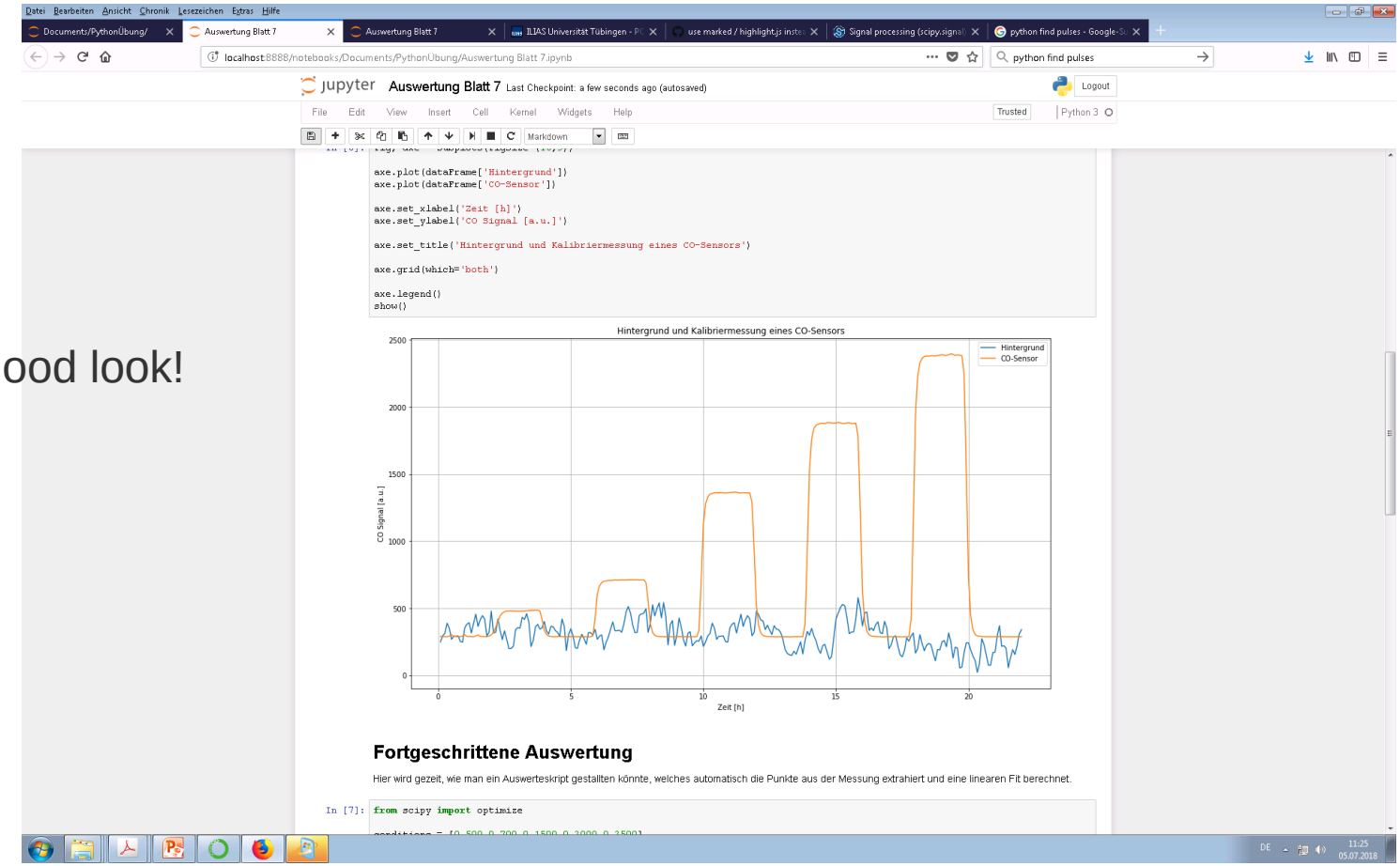
Besides the code you will also find explanations and results in the notebook.
You can execute the code by (“Cell->Run Cells”) of by pushing “CTRL-ENTER” on your keyboard.
Run some cells and see what is happening.



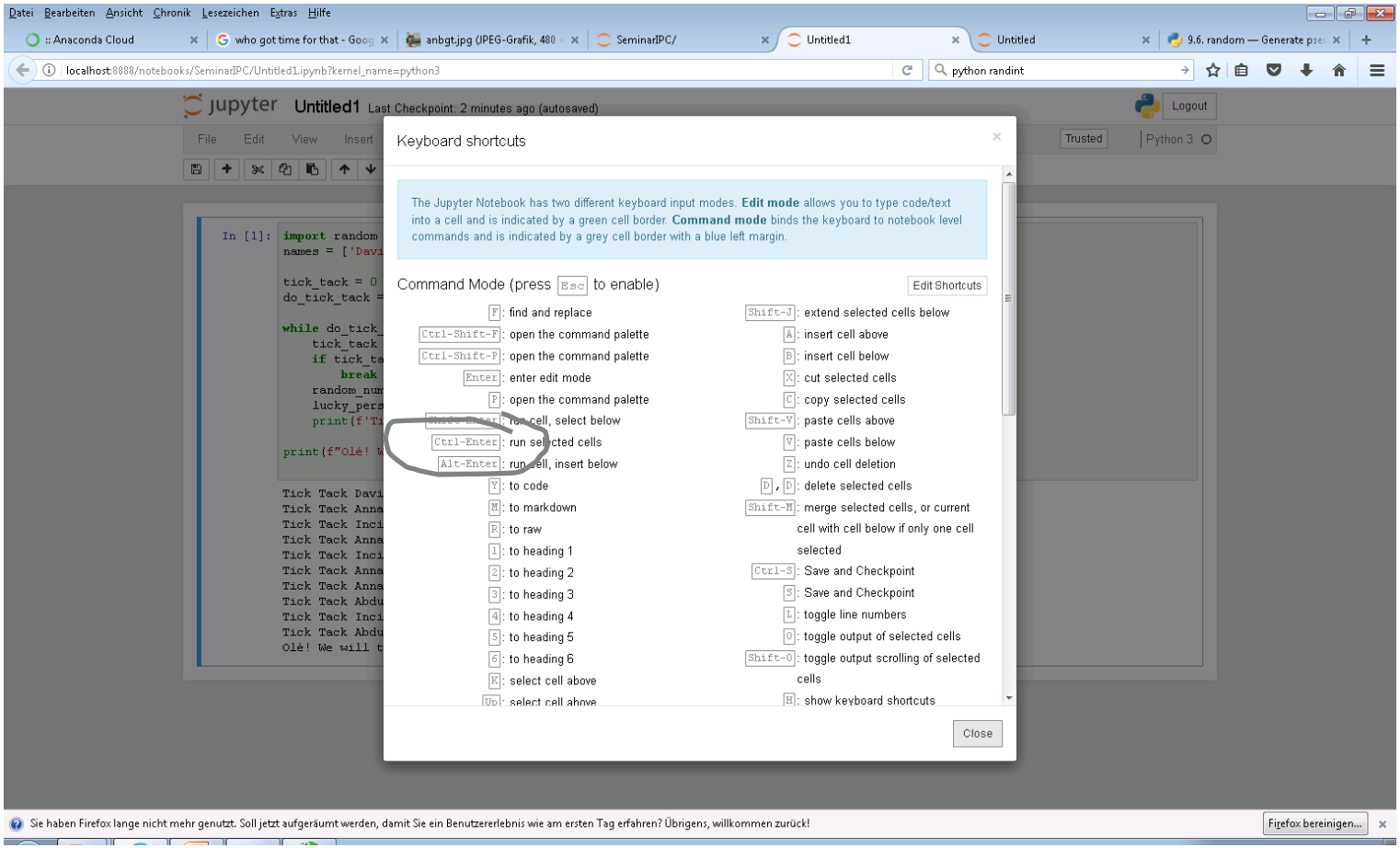


Try to get the following plot on your own. You may also want to play a bit around with the code.

Good look!



Um Shortcuts nutzen bietet sich die Hilfsfunktion an: Drücke “ESC”, und dann “h”.



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