technische universität dortmund

forschungs-daten-management

Tag der Forschungsdaten 2023

Einführung in die Programmierung mit Python

Dr. Bernd Zey
Abteilung Forschungsdatenmanagement
fdm@tu-dortmund.de
https://mailman.tu-dortmund.de/mailman/listinfo/fdm

Überblick

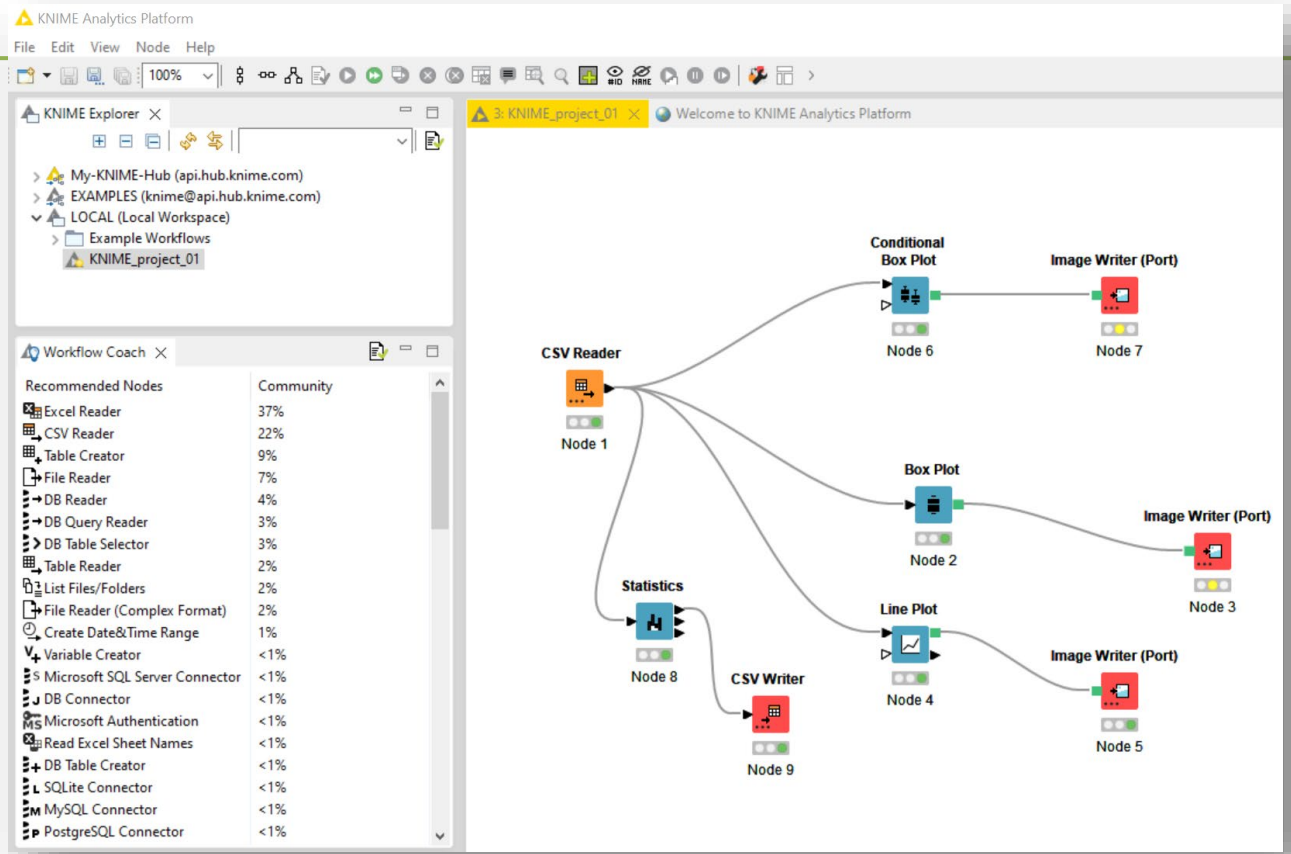
- Worum geht es heute hier?
  - Kurz-Einstieg in Python
  - Anwendungsfall
    - Tabellarische Daten laden und auswerten
    - Ergebnisse grafisch darstellen
    - Ergebnisse als Datei speichern
  - Keine Python-/Programmierkenntnisse erforderlich
- Motivation
  - Warum Python?
  - Entwicklungsumgebungen (IDE)
- Ausblick, Take-Home
- Live-Coding
  - JupyterLab

Warum Python?

- Standard in Data Science
- Zahlreiche Module für die statistische Auswertung und Visualisierung (Open Source)
  - Pandas, NumPy, …
  - Matplotlib, Seaborn, …
- Einsteigerfreundliche Programmiersprache
  - Allgemeine Programmiersprache, Plattform-unabhängig
  - Viele (wissenschaftliche) Open Source-Projekte
  - Einfache Syntax
  - Objektorientierung, Prozedural, …
  - Schnell umfangreiche Programme möglich
  - Einfaches Datei-Handling
  - Viele Module, u.a. für Machine Learning (TensorFlow, Keras, …)
  - Online-Hilfe

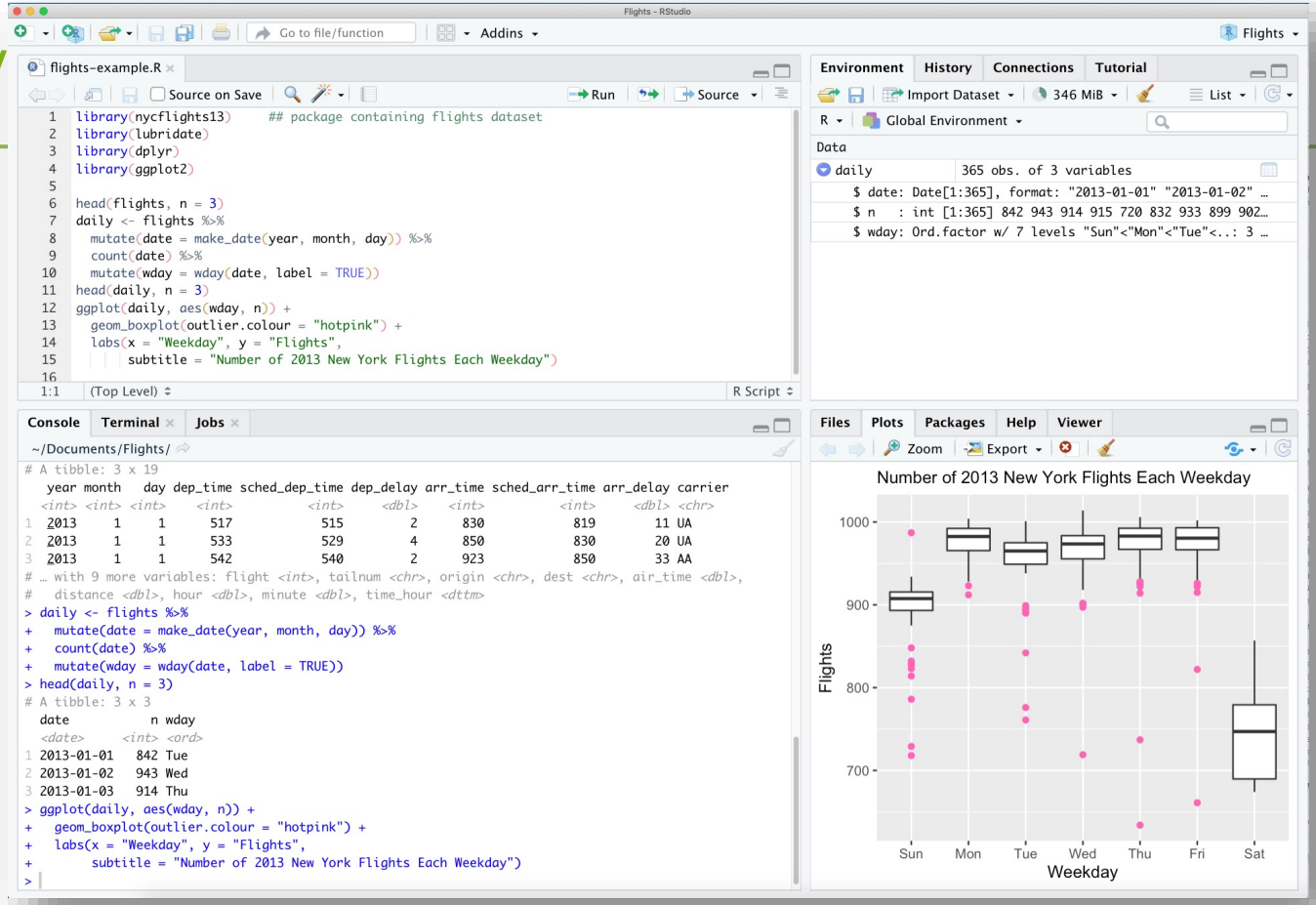
Warum Python? Was ist mit Tool XY?

Alternativen
- KNIME



Warum Py

Alternativen
- KNIME
- R

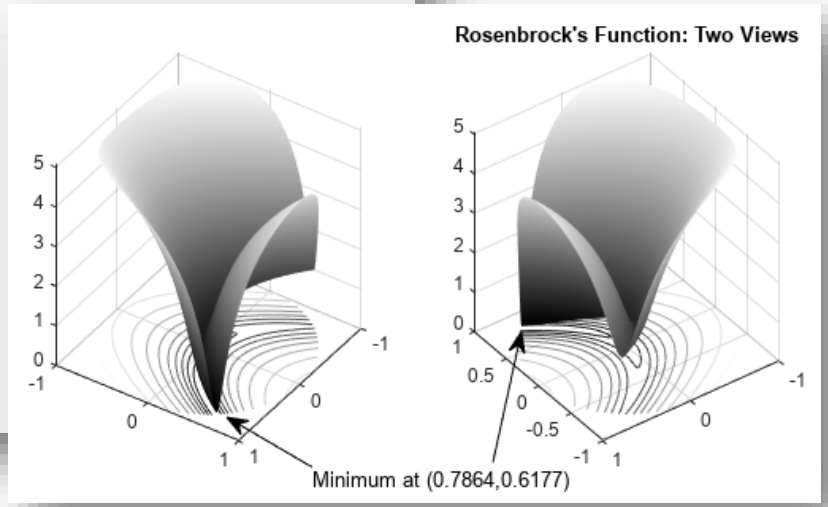


Screenshot von commons.wikimedia.org/

Warum Python? Was ist mit Tool XY?

Alternativen
- KNIME
- R
- Matlab (Octave)

```
rosenbrock = @(x)100*(x(:,2) - x(:,1).^2).^2 + (1 - x(:,1)).^2; % Vectorized function

figure1 = figure('Position',[1 200 600 300]);
colormap('gray');
axis square;
R = 0:.002:1;
TH = 2*pi*(0:.002:1);
X = R'*cos(TH);
Y = R'*sin(TH);
Z = log(1 + rosenbrock([X(:),Y(:)]));
Z = reshape(Z,size(X));

% Create subplot
subplot1 = subplot(1,2,1,'Parent',figure1);
view([124 34]);
grid('on');
hold on;

% Create surface
surf(X,Y,Z,'Parent',subplot1,'LineStyle','none');

% Create contour
contour(X,Y,Z,'Parent',subplot1);
```



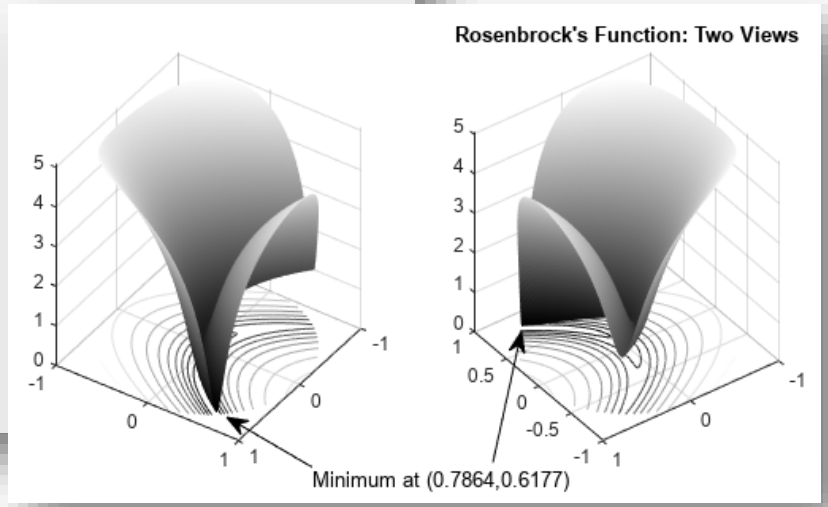
Screenshots von mathworks.com

Warum Python? Was ist mit Tool XY?

Alternativen

- KNIME
- R
- Matlab (Octave)
- Excel
- SPSS (PSPP)
- Stata
- Andere Programmier-sprachen (C++, Java, ...)
- ...

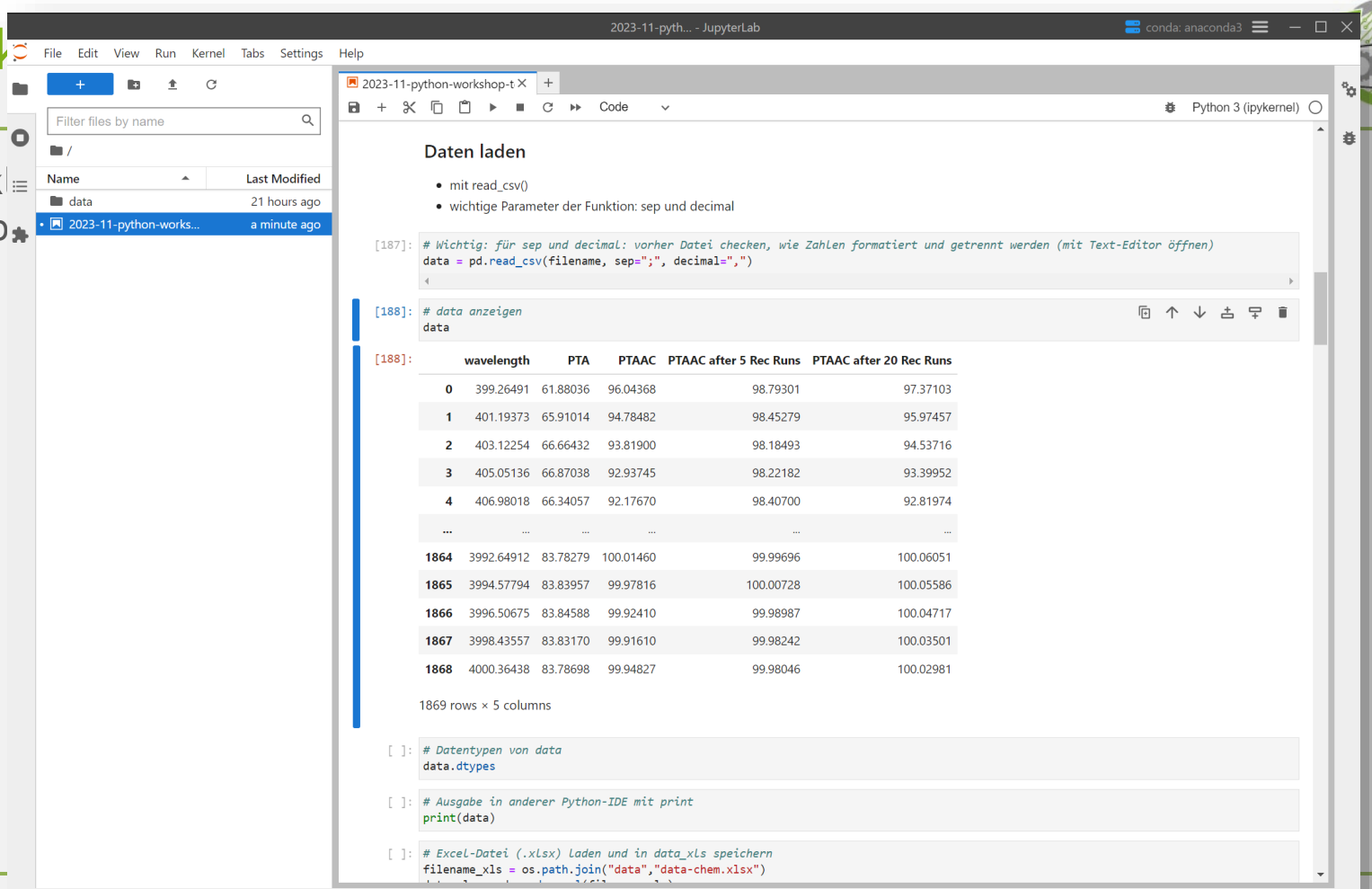
```
rosenbrock = @(x)100*(x(:,2) - x(:,1).^2).^2 + (1 - x(:,1)).^2; % Vectorized function

figure1 = figure('Position',[1 200 600 300]);
colormap('gray');
axis square;
R = 0:.002:1;
TH = 2*pi*(0:.002:1);
X = R'*cos(TH);
Y = R'*sin(TH);
Z = log(1 + rosenbrock([X(:),Y(:)]));
Z = reshape(Z,size(X));

% Create subplot
subplot1 = subplot(1,2,1,'Parent',figure1);
view([124 34]);
grid('on');
hold on;

% Create surface
surf(X,Y,Z,'Parent',subplot1,'LineStyle','none');

% Create contour
contour(X,Y,Z,'Parent',subplot1);
```



Screenshots von mathworks.com

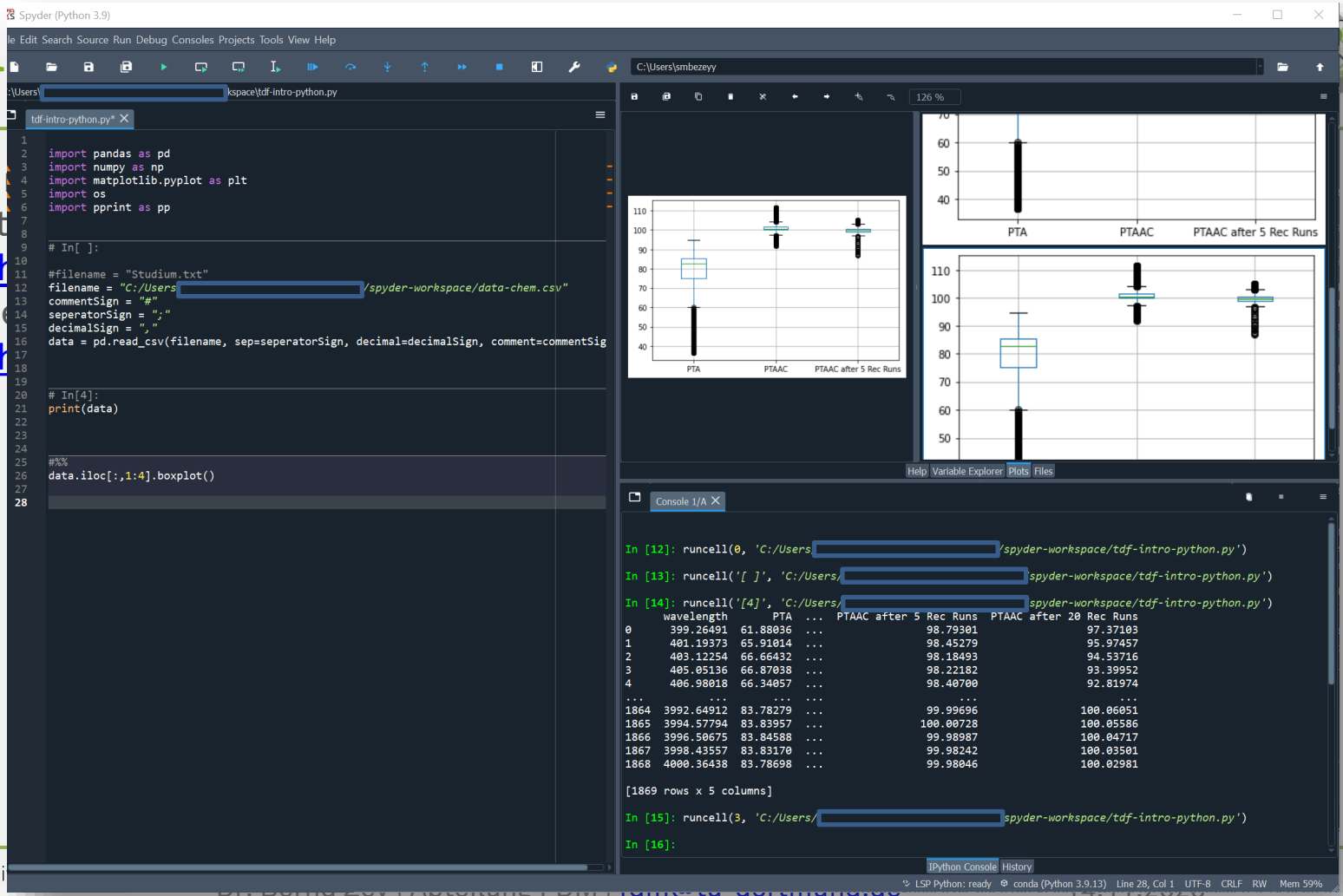
Entwick

- Entwick
  - Jup



Entwickl

- Entwicklu
  - Jupyt
    - h
  - Spyd
    - h

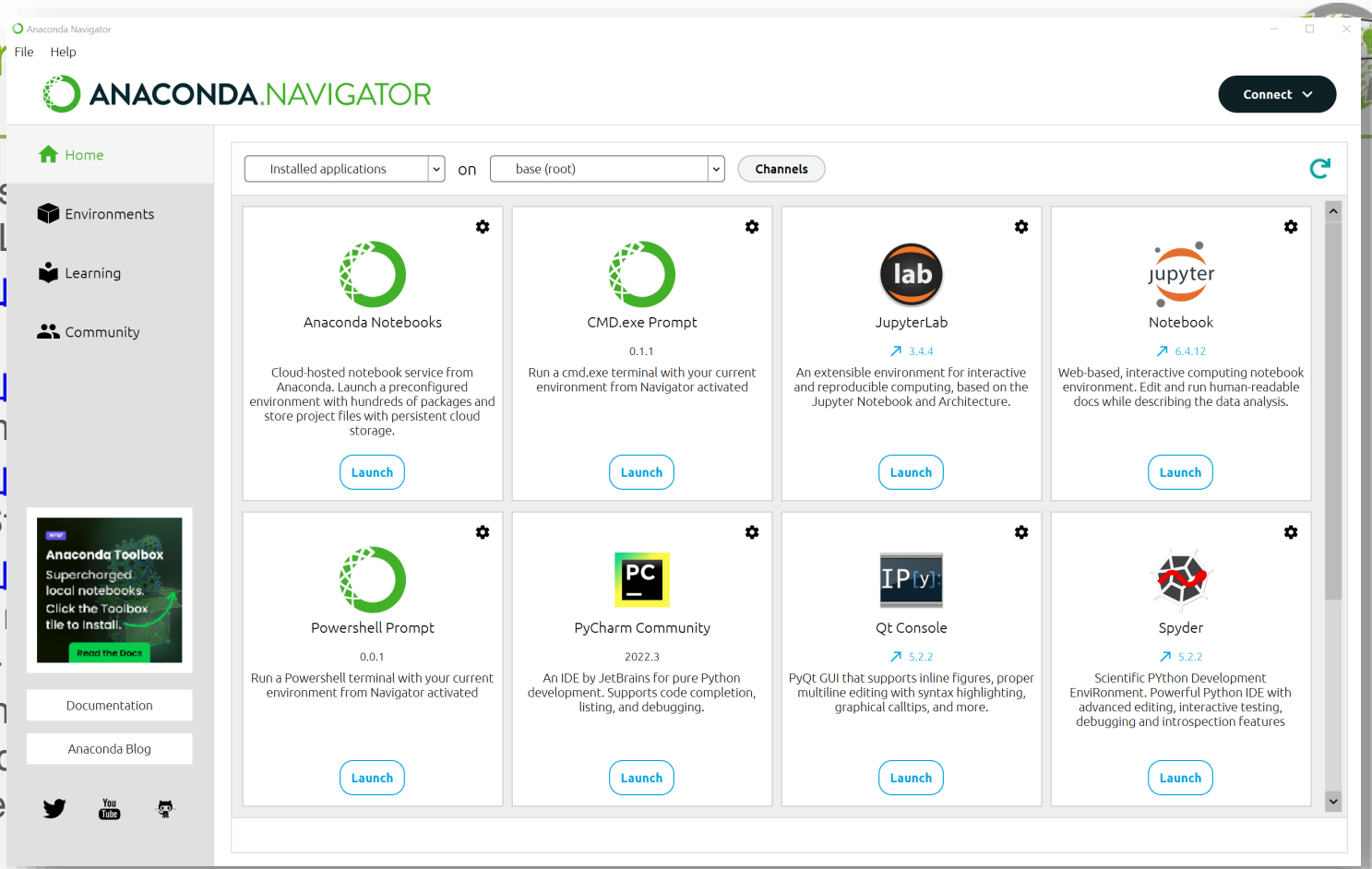


Entwicklungsumgebungen

- Entwicklungsumgebung – IDE (Integrated Development Environment)
  - JupyterLab
    - https://jupyter.org/
  - Spyder
    - https://www.spyder-ide.org/
  - PyCharm
    - https://www.jetbrains.com/pycharm/
  - Visual Studio Code
    - https://code.visualstudio.com/
- Mein Tipp zum Einstieg: Anaconda
  - Python-Distribution
  - Liefert mehrere IDEs
  - Anaconda Navigator
  - Installiert Python

Entwicklur

- Entwicklungs
  - Jupyterl
    - http
  - Spyder
    - http
  - PyCharm
    - http
  - Visual S
    - http
- Mein Tipp zu
  - Python-
  - Liefert n
  - Anaconc
  - Installie



Workshop

- Vorbereiteter Code und Live Coding
- JupyterLab
  - Abschnittsweise Ausführung
  - Markdown-Dokumentation und -Beschreibung
  - Direkte Ausgabe der Diagramme
- Code verfügbar auf Homepage
  - https://fdm.tu-dortmund.de/informationen/tag-der-forschungsdaten-2023-einfuehrung-in-python/
  - JupyterLab-Notebook
  - Lizenz: MIT
- Verwendete Daten basieren auf:
  - L. Hombach, A. K. Beine, “Carbon Supported Polyoxometalates as Recyclable Solid Acid Catalysts in Aqueous Reactions.”, 2023, doi: 10.22000/1106

Logo von https://radar4chem.radar-service.eu/radar/de/home

Ausblick, Take-Home

- Python *einfache* Programmiersprache
  → Automatisierung
- Kostenlos
- Mächtige Module (heute: Pandas) für Data Science
- Online gute Hilfe (Diskussionen, Code-Schnipsel)
  - z.B. https://stackoverflow.com
  - Websuche → „python pandas *thema*“
- Zum Einstieg: Anaconda
  https://www.anaconda.com/download
  - Entwicklungsumgebung JupyterLab
    (in Anaconda enthalten)