

Tag 7

Inhaltsverzeichnis

- Referenzen
- Was ist ein Jupyter Notebook?
- Mein erstes Notebook Beispiel
- Daten mit dem Notebook lesen und anzeigen
- Übungen
- BYOQ

Referenzen

Viele davon auf Englisch

- Offizielle Dokumentation: <https://jupyter-notebook.readthedocs.io/en/stable/>
- Miniconda Installationsanleitung: [hier](#)
- Eine gute Einführung von Jupyter Notebook:
<https://www.dataquest.io/blog/jupyter-notebook-tutorial/>
- Daten mit Plotly visualisieren: <https://plotly.com/python/>
- Markdown: <https://de.wikipedia.org/wiki/Markdown>
- Pandas **10 Min. Guide** [hier](#)
 - Siehe *Notebook_ITP_DM0D7_Pandas10MinGuide_V1.0.ipynb*
- Blick ins Cheatsheets – Verzeichnis zu werfen, lohnt sich
 - And the Medium's Python Cheatsheet, → [siehe hier](#)

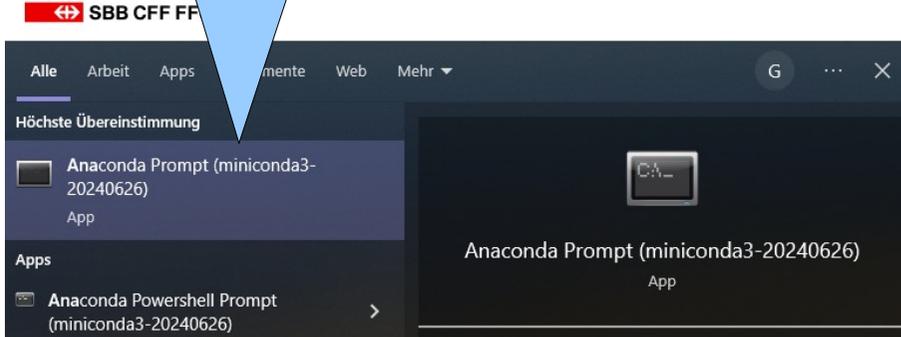
Hat jemand Miniconda noch nicht installiert?

Was ist ein Jupyter Notebook? Prinzip

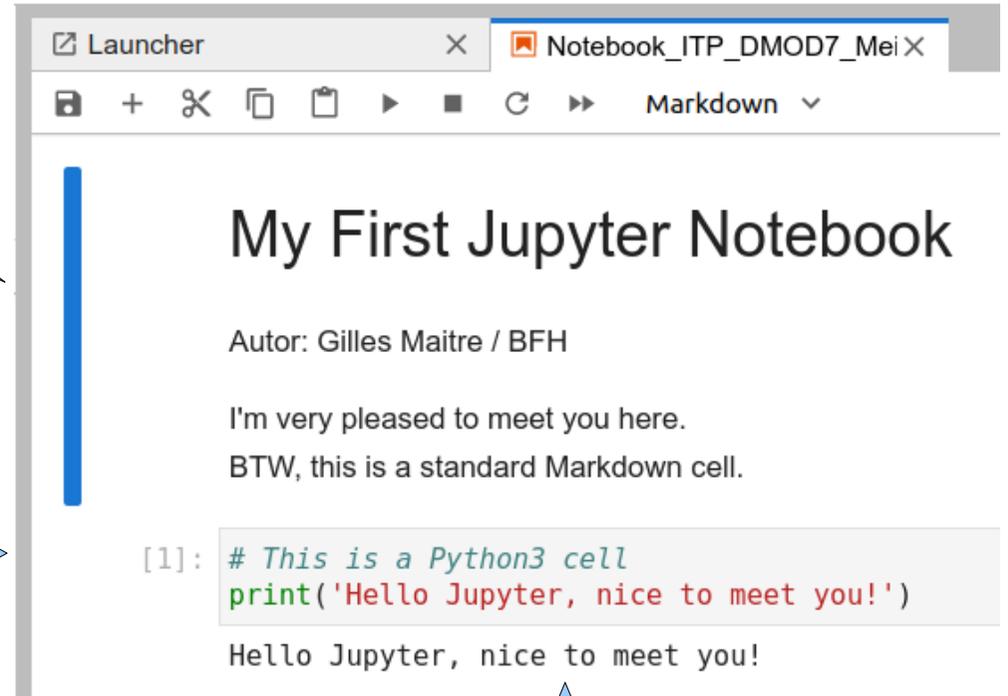
1

Auf Windows "Anaconda Prompt" ("Terminal" auf Mac / Linux) starten, und "jupyter lab" eintippen.

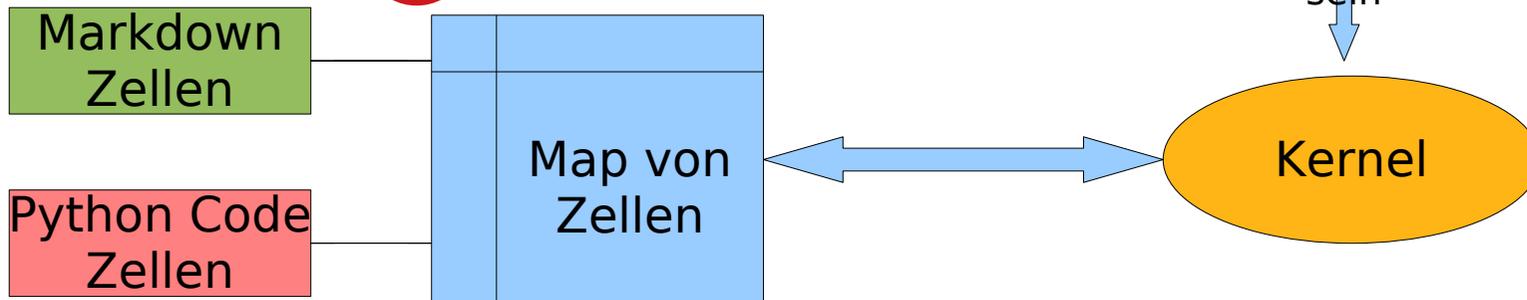
Interaktive Ausführung im Default-Browser



2



4



Was ist ein Jupyter Notebook? Mein erstes Beispiel (1)

File/Shut Down
to exit Kernel

Menu "Edit"
to edit cells

The screenshot shows the Jupyter Notebook interface. On the left is the file browser, and on the right is the notebook editor.

File Browser: The file browser shows a directory structure with a search bar and a list of files. The file `Notebook_ITP_DMOD7_MyFirstJupyterNotebook_V1.0.ipynb` is selected and highlighted in blue. A red circle with the number '1' is around the file name column header.

Name	Last Modified
DIAMOND-TeamRadar-V1.0.ipynb	3 months ago
DIAMOND-TeamRadar-V1.1.ipynb	3 months ago
DIAMOND-TeamRadar-V1.2.ipynb	3 months ago
Notebook_ITP_DMOD7_Bio-LWBetriebe-V2.2.ipynb	a year ago
Notebook_ITP_DMOD7_Bio-und-Konv-LWBetriebe-V2.2.ipynb	a year ago
Notebook_ITP_DMOD7_GroupAllocator.ipynb	8 months ago
Notebook_ITP_DMOD7_MyFirstJupyterNotebook_V1.0.ipynb	7 days ago
Notebook_ITP_DMOD7_Pandas10MinGuide_V1.0.ipynb	6 months ago

Notebook Editor: The notebook editor shows the title `My First Jupyter Notebook` and the author `Author: Gilles Maitre / BFH`. A red circle with the number '2' is around the author text. Below the title is a Markdown cell containing the text: `I'm very pleased to meet you here.` and `BTW, this is a standard Markdown cell.` Below that is a Python code cell with the code: `[]: # This is a Python3 cell` and `print('Hello Jupyter, nice to meet you!')` Below the code cell is the text `Jupyter knows 2 working modes:`

Click to
show/hide directories

Right-click
to "CRUD" Notebooks

Was ist ein Jupyter Notebook? Mein erstes Beispiel (2)

Save with CTRL-s

The screenshot shows a Jupyter Notebook window with the following content:

- Tab: Launcher Notebook_ITP_DM0D7_Mei X
- Toolbar: +, ✂, 📄, 📄, ▶, ■, ↺, ▶▶, Markdown ▾
- Title: My First Jupyter Notebook
- Author: Autor: Gilles Maitre / BFH
- Text: I'm very pleased to meet you here.
BTW, this is a standard Markdown cell.
- Code Cell: [1]: # This is a Python3 cell
print('Hello Jupyter, nice to meet you!')
- Output: Hello Jupyter, nice to meet you!

A Markdown cell

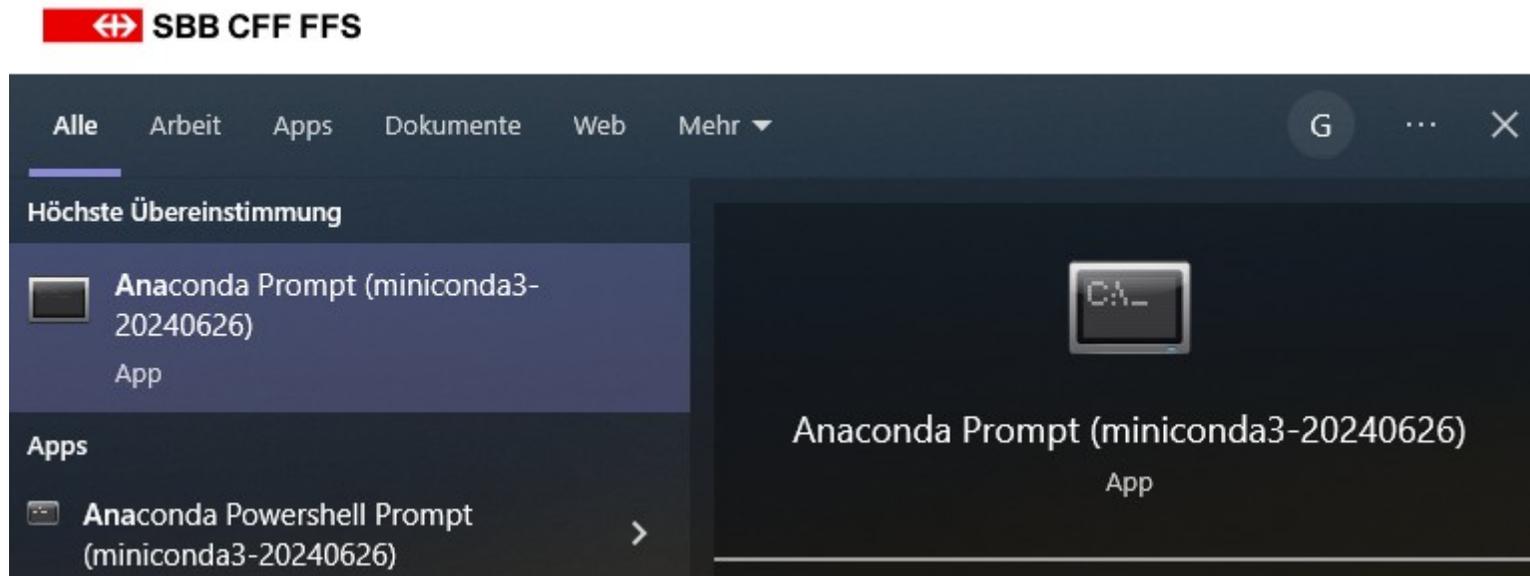
A Python code cell

Map-Index

Output, after having run the cell ;-)

jupyter lab starten

Anaconda Prompt / Terminal



- Auf Windows "Anaconda Prompt" ("Terminal" auf Mac / Linux) starten, und "jupyter lab" eintippen.
- Mit dem Befehl "cd" kann man in die Verzeichnisstruktur navigieren.
- Mit "cd .." geht man sich ein Verzeichnis nach oben.
- Die Verzeichnispfade kann man aus dem Windows Explorer kopieren und mit rechtem Mausklick einfügen.
- Beispiel zeigen...

Was ist ein Jupyter Notebook?

Notebook- und Daten-Verzeichnis

Es gibt ein Verzeichnis für die Notebooks und ein für die Daten.

- 1) Die Zip-Datei mit dem Name *CAS-ITP_DMODO_NotebooksAndData_V*.zip* aus Moodle herunterladen.
- 2) Diese Zip-Datei entpacken und die beiden Verzeichnisse in Ihr CAS-ITP Verzeichnis kopieren (am besten unter *C:\Users\<<IhrUsername>\CAS-ITP_DMODO_NotebooksAndData*).
- 3) Jetzt sollen Sie in diesem Verzeichnis beiden Unterverzeichnisse haben:
 - 1) Notebooks
 - 2) Data

Falls Sie die Zip-Datei anderswo speichern wollen, sollen Sie:

- Anaconda Prompt starten.
- Darin folgenden Befehl eintippen: "*cd <Verzeichnis, wo Sie die Zip-Datei entzippt haben>*".
- Dann "jupyter lab" eintippen.

Was ist ein Jupyter Notebook? Mein erstes Beispiel (3)

Alle starten
jupyter lab und laden das Notebook
"Notebook_ITP_DMOD7_MyFirstJupyterNotebook_V..."

Was ist ein Jupyter Notebook?

Vergleich Skript vs Notebook-Entwicklung

Skript/Programm (zB. Spyder)

- **Lineare** Entwicklung
"Edit/(Compile)/Debug"-Zyklus
- Debugger/Tracing gehört dazu
- Ideal für "**produktive**" Entwicklung
- Nicht ideal für "**prospektive**" Entwicklung

Notebook (zB. Jupyter)

- **Interaktive / Inkrementelle** Entwicklung
- Kein Debugger
=> Arbeit mit den Zellen
- Nicht ideal für "**produktive**" Entwicklung
- Ideal für "**prospektive**" Entwicklung

Anwendungsbeispiel bei der SBB-Infrastruktur:

- 1) Data-Scientists/Bahn-Ingenieure entwickeln Prototypen mit Notebooks
- 2) Software-Ingenieure *integrieren/weiterentwickeln* diese für TEST/PROD-Systeme

Was ist ein Jupyter Notebook?

Spickzettel

- Hier der Link auf den Spickzettel
 - <SHIFT>-<RETURN> Aktuelle Zelle(n) ausführen
 - <CTRL>-<RETURN> Aktuelle Zelle(n) ausführen
 - a oder b Code-Zelle "**a**bove" / "**b**elow" addieren (in Command-Modus)
 - d d aktuelle Zelle löschen (in Command-Modus)
 - z Gelöschte Zelle wieder beleben (in Command-Modus)
 - <CTRL>-s Notebook ins .checkpoint Verzeichnis speichern
 - <TAB> "Code completion", je nach Kontext
 - ? Help anzeigen (Funktion oder Objekt)
- Wichtigste Pandas *DataFrame*-Befehle
 - df # Am Schluss einer Zelle: DataFrame anzeigen
 - df.shape # DataFrame-Dimensionen (Zeilen x Spalten)
 - df.head() # Erste 5 Zeilen anzeigen
 - df.info() # DataFrame-Datentypen anzeigen
 - df.columns # DataFrame-Spalten (ohne "()") anzeigen
 - df.index # DataFrame-Index (ohne "()") anzeigen
 - df.describe() # Einfache Statistik durchführen

Was ist ein Jupyter Notebook?

jupyter notebook vs. jupyter lab

jupyter notebook

- "Altes" Werkzeug für die Arbeit mit Notebooks
- Sehr gut bekannt und dokumentiert
- Ende der Entwicklung absehbar
- Rudimentär, weit weg einer Entwicklungsumgebung, wie Spyder

jupyter lab

- Künftiges Werkzeug für die Arbeit mit Jupyter Notebooks
- 100% kompatibel mit den Jupyter Notebooks
- Weniger bekannt und dokumentiert
- In der Entwicklung, Überraschungen nicht ausgeschlossen ;-)
- "Dark-mode"-Theme vorhanden ;-)
 - Siehe "Setting / Themes"

Im Unterricht
brauchen wir
jupyter lab

Pandas in a Nutshell

Was ist Pandas?

pandas is a Python package (...) [which] (...) aims to be the fundamental high-level building block for doing **practical, real-world data analysis in Python**.

pandas is well suited for many different kinds of data:

In one word:
a fantastic *data engineering* tool

- Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
- Ordered and unordered (not necessarily fixed-frequency) time series data.
- Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels

- Basic data structures: Series (1-dimensional) and DataFrame (2-dimensional)
- Handling of missing data
- Columns and rows CRUD-ing
- Reshaping, pivoting data
- Merging, joining data
- Powerful groupBy (split-apply-combine)
- IO tools to load data from CSV, JSON, Excel, HDF5

Pandas in a Nutshell

Was sind DataFrames?

DataFrame object

Label index
(country code)

Column names

	Country	Popu	Percent
IT	Italy	61	0.83
ES	Spain	46	0.63
GR	Greece	11	0.15
FR	France	65	0.88
PO	Portugal	10	0.14

Data
(different type in each column)

Pandas in a Nutshell

Typen von DataFrames

Excel

- Not really a DataFrame tool, but...
- **Local** computations on **one** CPU (?)

Python / Pandas

- Classical / well known
- **Local** computations on **one or more** CPUs
- Used by Anaconda / Data Scientists

Python / Spark (PySpark)

- **Distributed** computations on **a cluster of** CPUs (BigData)
- Used by Spark / Hadoop / Data Scientists

Pandas in a Nutshell

Ein Reference Notebook

- Object creation
- Viewing data
- Selection
- Handling missing data
- Operations / Applying
- Merging / Joining
- Grouping
- Reshaping (stacking and pivoting)
- Time series
- Data frames
- Reading CSV, Excel and JSON files

Das alles habe ich in einen Notebook gepackt

Analyse Bio-LWBetriebe 1990-2020

Suchen nach Daten

- Ein erster Versuch: [OpenData](#) und dann [Wiki Quickstart](#)
 - [Hier sieht man, welche Daten](#) verfügbar sind.
 - Und spezifisch [für die Schweiz](#). Gut zu wissen!
- Ein anderer Weg: Das [Bundesamt für Statistik](#)
 - Datenüberblick [hier](#)
 - Ich suche nach "landwirtschaft betriebe"
 - Unter [Landwirtschaftliche Betriebe und Beschäftigte nach Kanton](#) finden wir etwas
 - Die Daten kann man [hier](#) herunterladen, aber man muss sagen, welche =>
 - Folgende Felder wähle ich:
 - Beschäftigte Total, alle Kantone (ohne Schweiz), "Bio-Betriebe" als Betriebssystem, für die Jahre 1990, 2000, 2010, 2020
 - Ergebnisse speichern als: *Textdatei (Komma-getrennt) ohne Kopfzeile*
 - Die Datei jetzt lokal mit Excel öffnen und als Kodierung: "ISO-8859-1" wählen

Laden Sie diese Daten
jetzt herunter
und mit Excel öffnen

Analyse Bio-LWBetriebe 1990-2020

Analyse im Notebook / Überblick

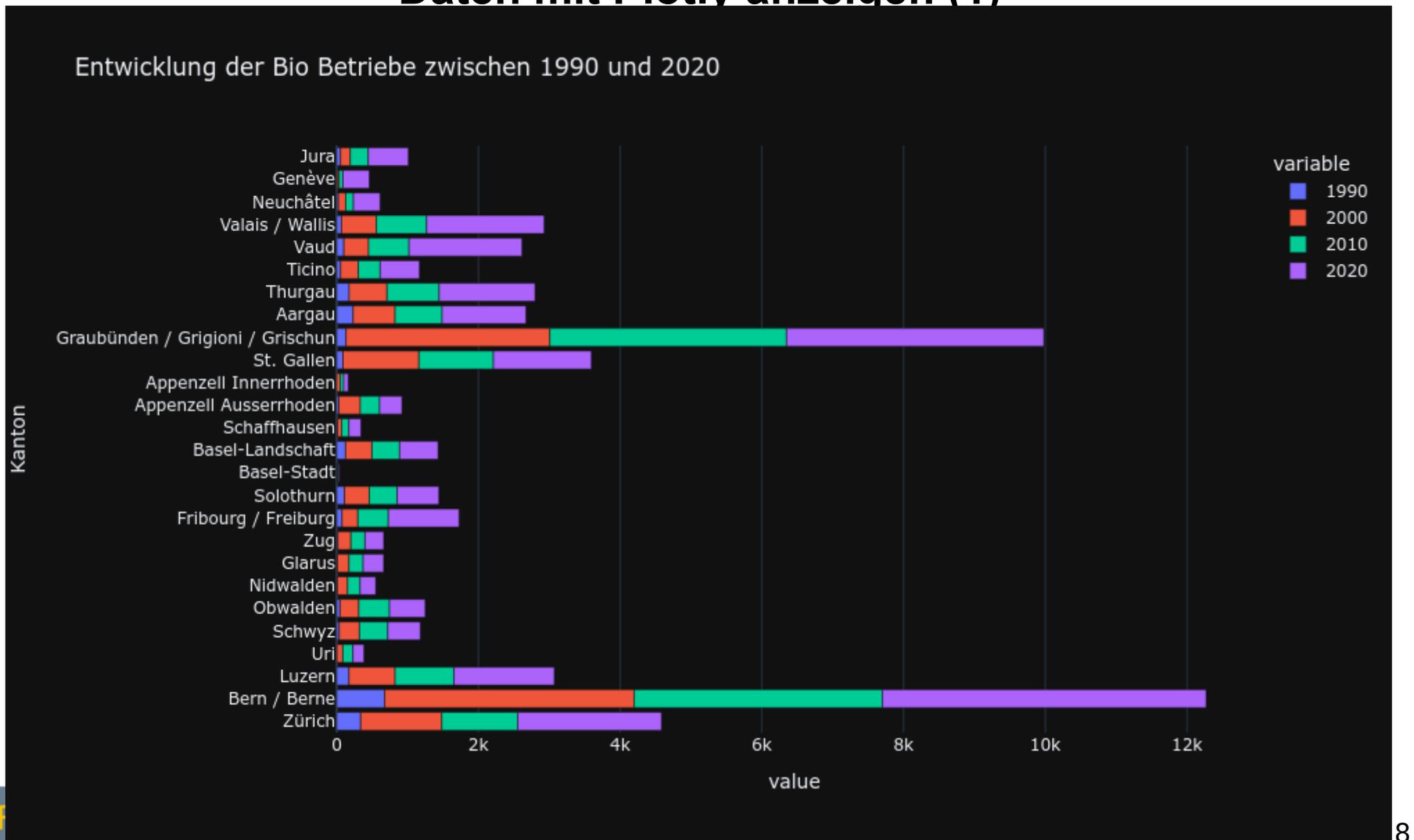
Folgende Operationen führen wir jetzt im Notebook aus:

- 1) Daten in ein Pandas *Data-Frame* laden und anzeigen
- 2) Ein paar Spalten löschen
- 3) Einfache Statistik ausführen, mit *describe()*
- 4) Die Daten grafisch mit Plotly darstellen
- 5) Entwicklung zwischen 1990 und 2020 in % anzeigen

Jetzt ins Notebook umsteigen...

Analyse Bio-LWBetriebe 1990-2020

Daten mit Plotly anzeigen (1)



Analyse Bio-LWBetriebe 1990-2019

Daten mit Plotly anzeigen (2)

Plotly ist die *grosse Revolution*, um Daten zu visualisieren:

- 1) Arbeitet Hand in Hand mit Pandas *Data-Frames*.
- 2) Haufen von "Traces". Siehe [hier](#).
- 3) Einfache Scatter-, Bar- und Line-Plots.
- 4) *Diese Diagramme sind interaktiv... und farbig*
Tolle Beispiele [hier](#). (jemand farbenblind?)
- 5) Die Diagramme kann man als PNG (statisch) oder HTML (mit Animation) speichern.

Analyse Bio-LWBetriebe 1990-2020

Analyse im Notebook, Entwicklung in %

Jetzt noch die Entwicklung zwischen 1990 und 2020 in %

Dafür addieren wir eine neue Spalte `diff1990_2020`, die diese %-Zahlen enthält. Und sortieren das DataFrame nach dieser Spalte, aufsteigend.

Der Kanton Appenzell Innerrhoden hat eine unendliche Entwicklung hinter sich ;-)

```
biobetriebe_df['diff1990_2020'] = biobetriebe_df['2020']*100/biobetriebe_df['1990'] - 100
biobetriebe_df.sort_values(by = 'diff1990_2020')
```

	Kanton	2020	2010	2000	1990	diff1990_2020
11	Basel-Stadt	15	11	8	7	114.285714
12	Basel-Landschaft	543	391	372	125	334.400000
18	Aargau	1184	664	591	232	410.344828
10	Solothurn	589	394	348	111	430.630631
0	Zürich	2028	1073	1144	338	500.000000
1	Bern / Berne	4569	3504	3520	677	574.889217
19	Thurgau	1353	736	534	176	668.750000
2	Luzern	1416	832	649	174	713.793103
20	Ticino	555	309	249	55	909.090909
25	Jura	566	254	138	54	948.148148
14	Appenzell Ausserrhoden	315	273	303	30	950.000000
5	Obwalden	504	433	265	47	972.340426

Analyse Bio-LWBetriebe 1990-2020

Fazit

- 1) Wir haben mit wenig Python-Code eine einfache Analyse der Daten durchgeführt.
- 2) Mit Excel hätten wir das auch geschafft.
- 3) Wird die Analyse komplexer, ist IMHO die Jupyter-Lösung wesentlich stärker.
- 4) Ihre Meinung?

Notebooks vergleichen/teilen

Wie geht das?

- 1) Jupyter Notebooks sind dafür nicht gut geeignet.
- 2) Pragmatische Lösung dazu:
 - 1) Das Notebook als Python-Skript (Menü "File") herunterladen.
 - 2) Mit einem Dateivergleichsprogramm (z.B. kdiff3) die Python-Skripts vergleichen.
- 3) Um Notebooks zu verteilen, geht es *am einfachsten*, die *.ipynb-Dateien* herumzuschicken.

Übungen

Analyse der konventionellen Betriebe

1) Führen Sie selber eine Analyse der *konventionellen* Landwirtschaftsbetriebe am Ende des Notebooks aus.

Tipp: Machen Sie **eine Kopie** meiner originalen Notebooks. Das können Sie direkt im Jupyter File Browser machen, mit dem Befehl *Duplicate*.