

# Data Mining I: Introduction to Python



# Installation



- Install Anaconda (Python Distribution)
  - <https://www.anaconda.com/distribution/>
  - Use Python 3.x



Anaconda Installers		
Windows 	MacOS 	Linux 
Python 3.8 64-Bit Graphical Installer (466 MB) 32-Bit Graphical Installer (397 MB)	Python 3.8 64-Bit Graphical Installer (462 MB) 64-Bit Command Line Installer (454 MB)	Python 3.8 64-Bit (x86) Installer (550 MB) 64-Bit (Power8 and Power9) Installer (290 MB)

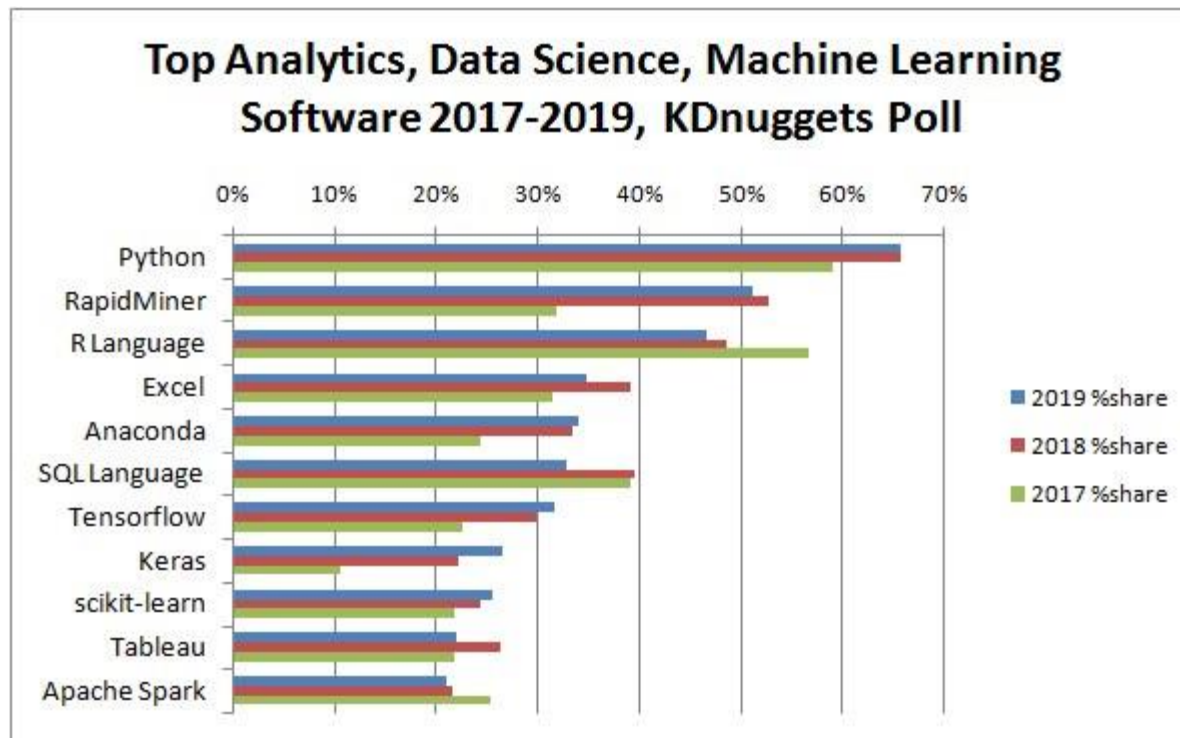
- If you don't have at least 3 GB disc space
  - Option 1 (better): Get a bigger disc!
  - Option 2: install miniconda
    - <https://docs.conda.io/en/latest/miniconda.html>
- Alternative: Use [Google Colab](#)

# Python

- Started in 1989 by Guido van Rossum
  - The name is a tribute to the British comedy group Monty Python
- High-level, general-purpose programming language
  - Multi-paradigm: functional, imperative, object-oriented, reflective
- Design goals
  - Be extensible, simple, and readable

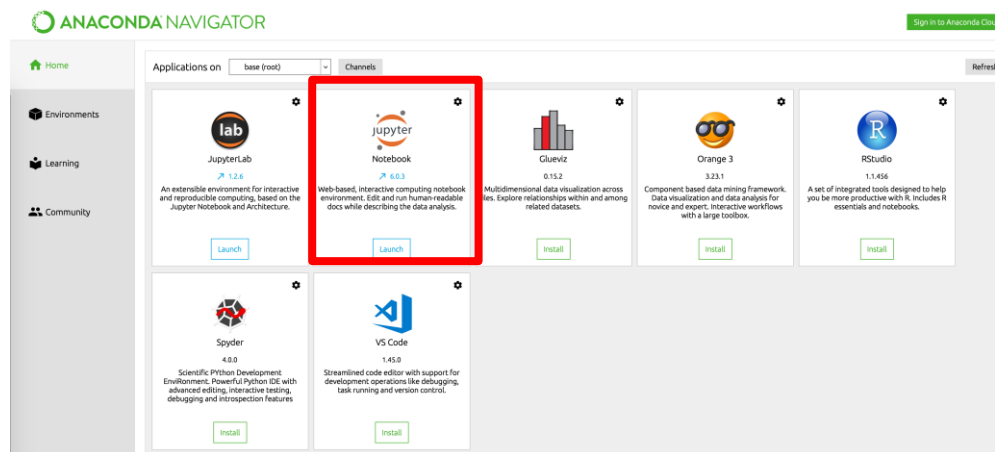


# Popularity



# Jupyter notebooks

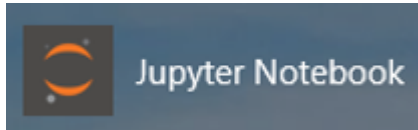
- In the exercises, we will use Jupyter Notebooks
- The start procedure depends on your operating system
  - in general: run the **Anaconda Navigator** and click on **jupyter Notebook**



- The Jupyter Notebook App can access only files within its start-up folder (including any sub-folder)
  - default is your home folder
    - Windows: usually `C:\Users\{username}` Linux: `/home/{username}`

# Jupyter notebooks

- Start Jupyter - Option 1 (Windows)



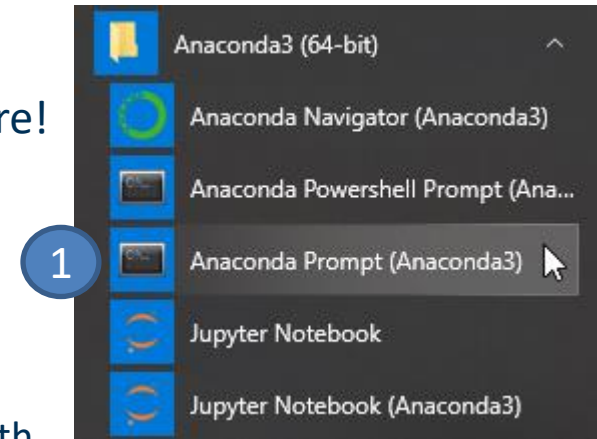
- Click on the *Jupyter Notebook* icon in the start menu
- To change this folder:
  - Copy the Jupyter Notebook launcher from the menu to the desktop.
  - Right click on the new launcher and change the Target field, change %USERPROFILE% to the full path of the folder which will contain all the notebooks.
  - Use the *Jupyter Notebook* desktop launcher to start the notebook



# Jupyter notebooks

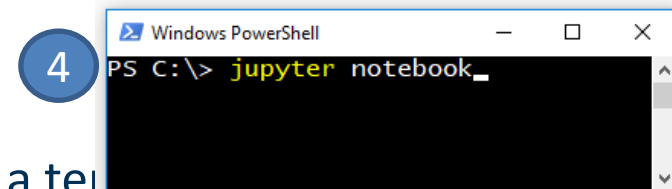
- Start Jupyter – Option 2 (Linux and Windows)

- Run „jupyter notebook“ in command line
  - Navigate to folder that you want to access before!
  - Or (Windows):
    1. Open the Anaconda Prompt
    2. Copy the path to your folder
    3. Write “cd ” and then make a **right** mouse click into the terminal folder to insert the copied path
    4. then type “jupyter notebook” + enter



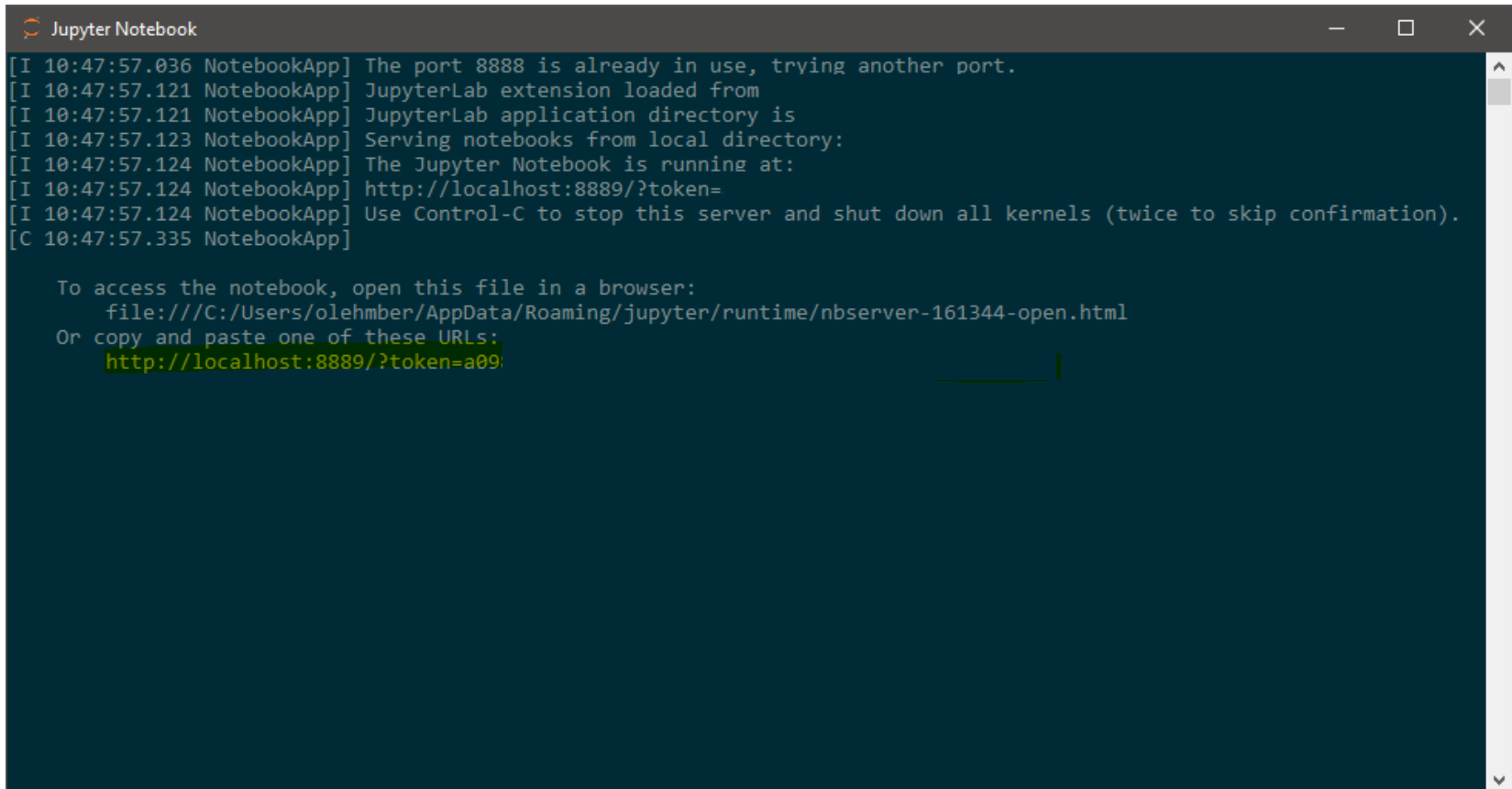
- Start Jupyter – Option 3 (Mac OS)

- Click on spotlight, type “terminal” to open a terminal
- Enter the startup folder by typing “cd /some\_folder\_name”.
- Type “jupyter notebook” to launch the Jupyter Notebook App



# Jupyter notebooks

- A local server is started
- Open the URL on screen in your browser, if not already opened



```
Jupyter Notebook
[I 10:47:57.036 NotebookApp] The port 8888 is already in use, trying another port.
[I 10:47:57.121 NotebookApp] JupyterLab extension loaded from
[I 10:47:57.121 NotebookApp] JupyterLab application directory is
[I 10:47:57.123 NotebookApp] Serving notebooks from local directory:
[I 10:47:57.124 NotebookApp] The Jupyter Notebook is running at:
[I 10:47:57.124 NotebookApp] http://localhost:8889/?token=
[I 10:47:57.124 NotebookApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 10:47:57.335 NotebookApp]

To access the notebook, open this file in a browser:
    file:///C:/Users/olehmber/AppData/Roaming/jupyter/runtime/nbserver-161344-open.html
Or copy and paste one of these URLs:
    http://localhost:8889/?token=a09
```



# Jupyter Home Screen

- Startscreen in browser
  - like a file explorer

upload files to  
notebook folder

Create new  
files/notebooks



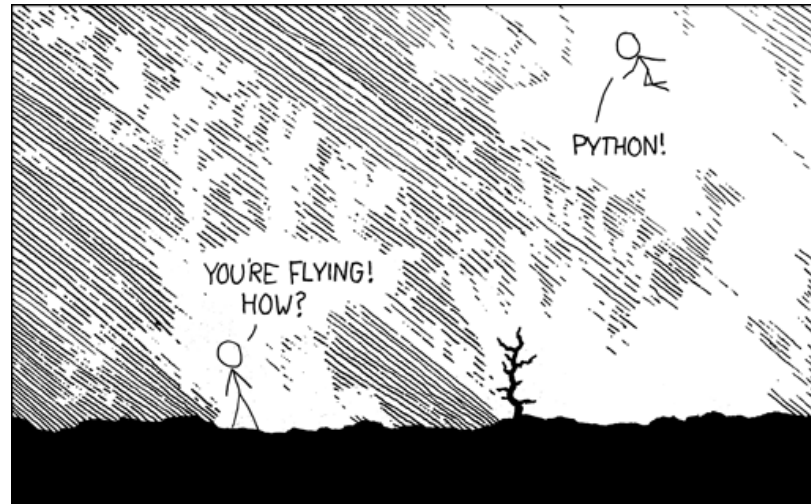
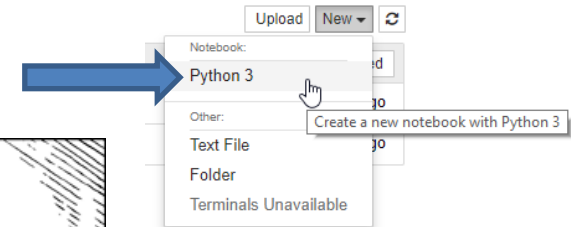
The screenshot shows the Jupyter Home Screen interface. At the top left is the Jupyter logo. To the right is a 'Logout' button. Below the logo are three tabs: 'Files' (selected), 'Running', and 'Clusters'. A message says 'Select items to perform actions on them.' Below this is a file list table. The table has columns for 'Name' and 'Last Modified'. It contains two rows: 'Untitled.ipynb' (with a green notebook icon) and 'untitled.txt' (with a red text file icon). The 'Untitled.ipynb' row shows 'Running' in green text and 'seconds ago'. The 'untitled.txt' row shows 'seconds ago'. To the right of the file list are buttons for 'Upload', 'New' (with a dropdown arrow), and a refresh icon. Annotations with blue arrows point to these elements: 'upload files to notebook folder' points to the 'Upload' button; 'Create new files/notebooks' points to the 'New' button; and 'file explorer' points to the 'untitled.txt' file entry.

	Name	Last Modified
<input type="checkbox"/>	Untitled.ipynb	Running seconds ago
<input type="checkbox"/>	untitled.txt	seconds ago

file explorer

# Now try it out

- Click in browser „New“ -> „Python 3“



<https://xkcd.com/353/>

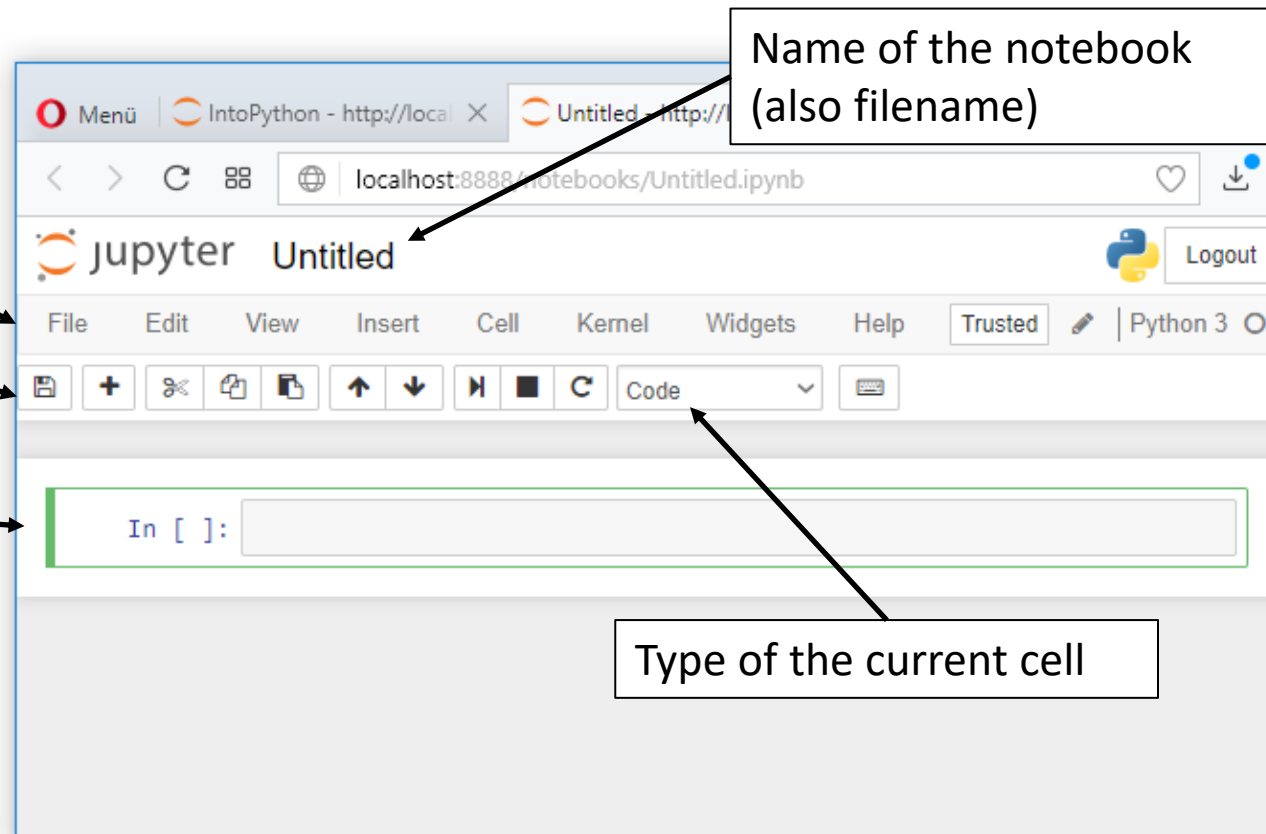
# Jupyter Notebook

- Every notebook is composed of cells
  - Cells contain a specific type of content
  - markdown cells (for documentation and structure)
  - code cells

Menu Bar

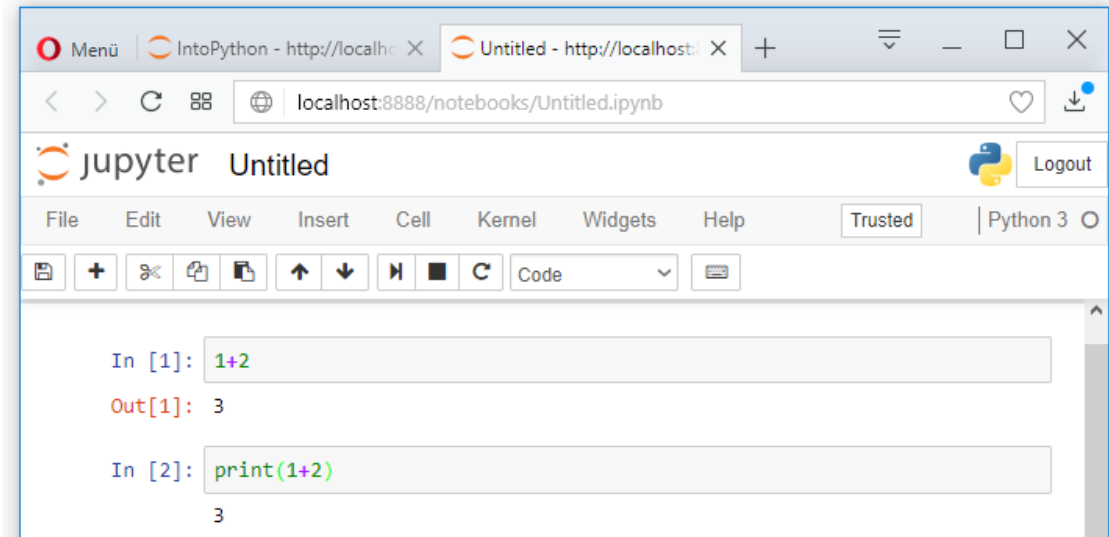
Shortcuts

Cell



# Jupyter Cells

- Code cell:
  - You can type python code (because you created a python notebook)
    - Hit „Ctrl + Enter“ to run the code
    - Hit „Shift + Enter“ to run it and create a new cell
    - Try it and type  $1 + 2$
  - The output is shown below the cell



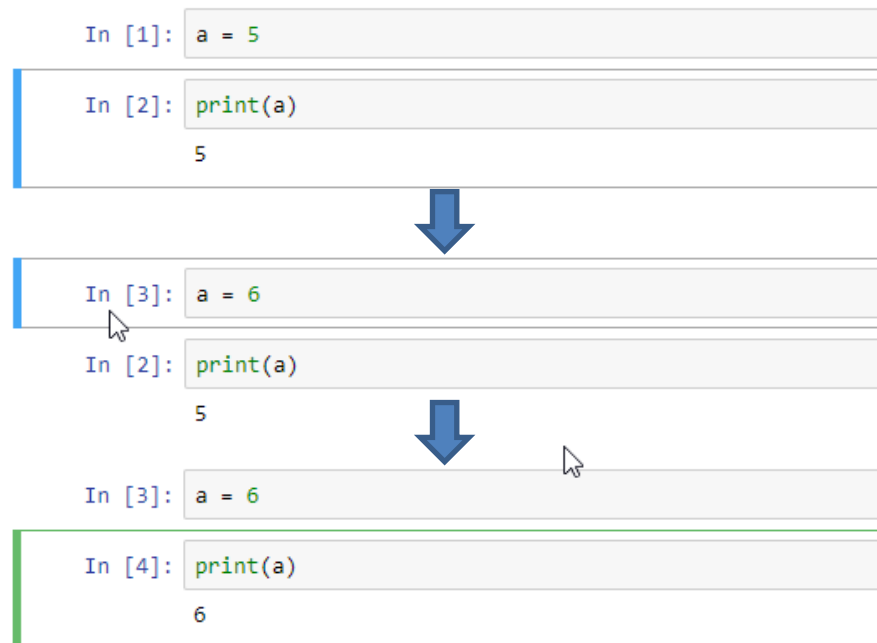
The screenshot shows a web browser window with two tabs: 'Menü' and 'Untitled - http://localhost:8888/notebooks/Untitled.ipynb'. The address bar shows 'localhost:8888/notebooks/Untitled.ipynb'. The Jupyter interface includes a 'Logout' button and a menu bar with 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. Below the menu bar is a toolbar with icons for file operations and execution. The main area displays two code cells. The first cell, labeled 'In [1]:', contains the code '1+2' and its output 'Out[1]: 3'. The second cell, labeled 'In [2]:', contains the code 'print(1+2)' and its output '3'.

```
In [1]: 1+2
Out[1]: 3

In [2]: print(1+2)
3
```

# Jupyter Cells

- Each „code cell“ can be reevaluated (indicated by a number)
  - All previous results / variables are stored (like in R workspace)



Change code  
and run again  
[1] -> [3]

run next cell  
again [2] -> [4]

# Jupyter Cells

- Autocomplete by pressing <tab> when writing

```
In [2]: my_very_long_variable_with_hundreds_of_characters = 5  
my_second_very_long_variable_with_hundreds_of_characters = 6
```

```
In [ ]: my|  
my_second_very_long_variable_with_hundreds_of_characters  
my_very_long_variable_with_hundreds_of_characters
```

- Signature of function by pressing <shift>+<tab>

```
corr(datar, color_grades=5)
```

Signature:

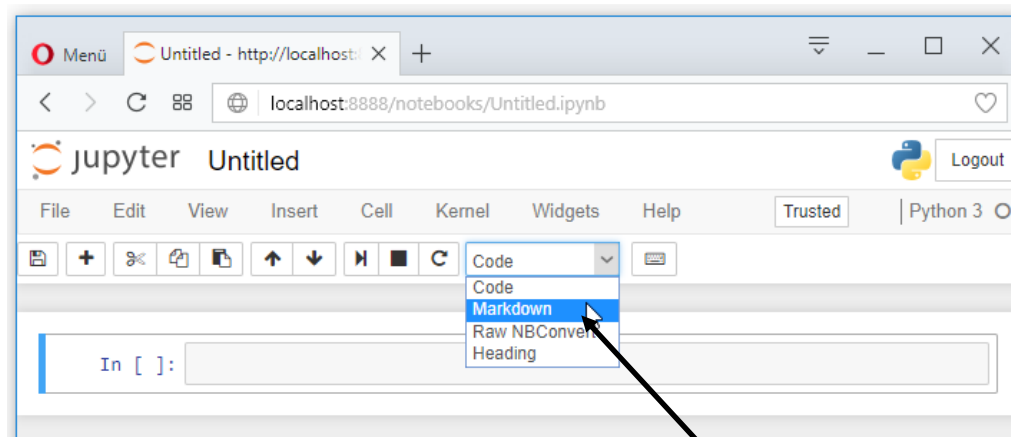
```
corr(  
    ['data', "corr_method='spearman'", 'annot=False', 'mask=True', 'line_width=1', "line_  
    color='black'", 'color_grades=5', 'auto_sizing=True', "palette='default'", "style='asteti
```

```
5 #from astetik import corr
```



# Jupyter Cells

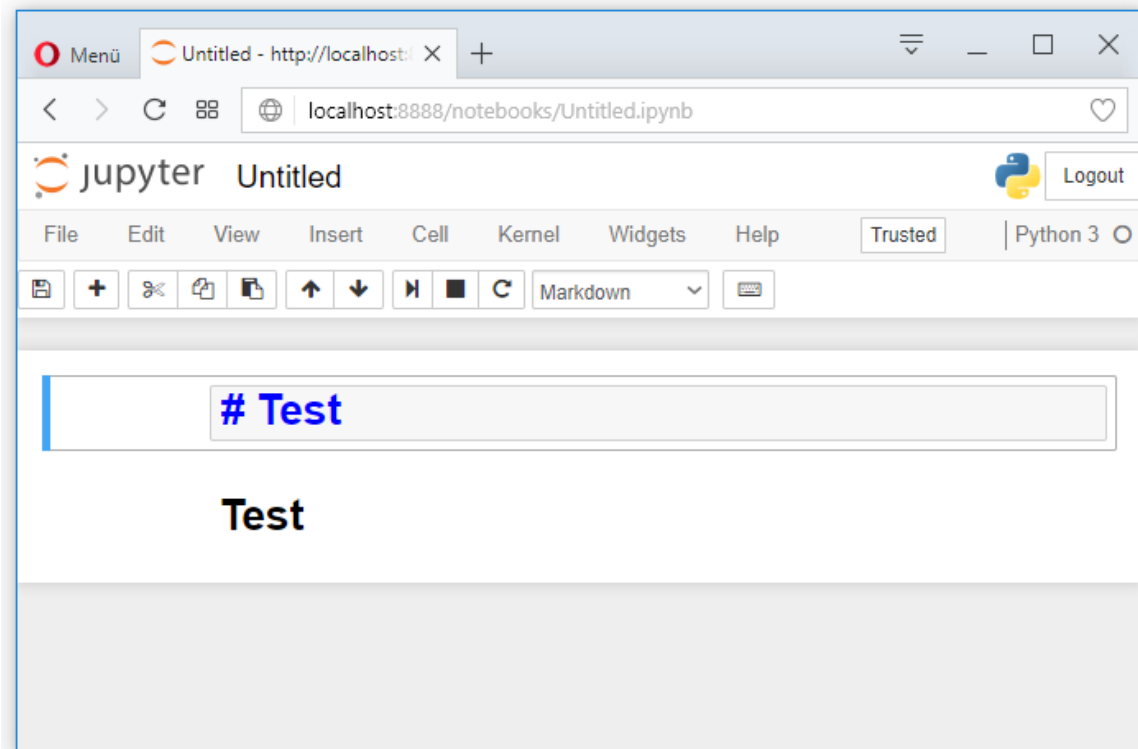
- What makes a notebook a notebook?
  - Markdown cells let you add documentation and notes
  - Create a new cell („Insert->Insert Cell Below“)
  - Change the type to Markdown



Type of the current cell

# Jupyter Cells

- What makes a notebook a notebook?
  - Type „# Test“ which creates a heading (add more „#“ for smaller headline)
    - Whitespace after #
  - Evaluate the cell and see the result



# Jupyter Cells - Markdown

- Different possibilities to structure

- Header

```
# H1  
## H2  
### H3
```

- Unordered List (use "\*", "+", or "-" in front)

```
- Item  
- Item
```

- Ordered list

```
1. Item one  
2. Item two
```

- Links

```
[link to google](https://www.google.com)
```

- Image

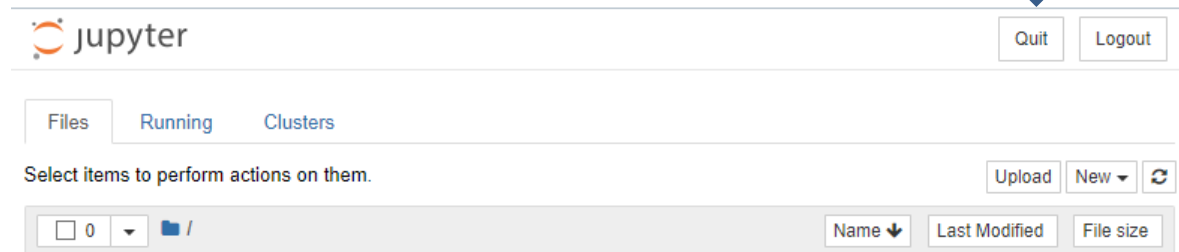
```
![Mannheim Image](https://www.uni-mannheim.de/1/00_UM_Dachmarke_DE_RGB.jpg)
```

- Quote

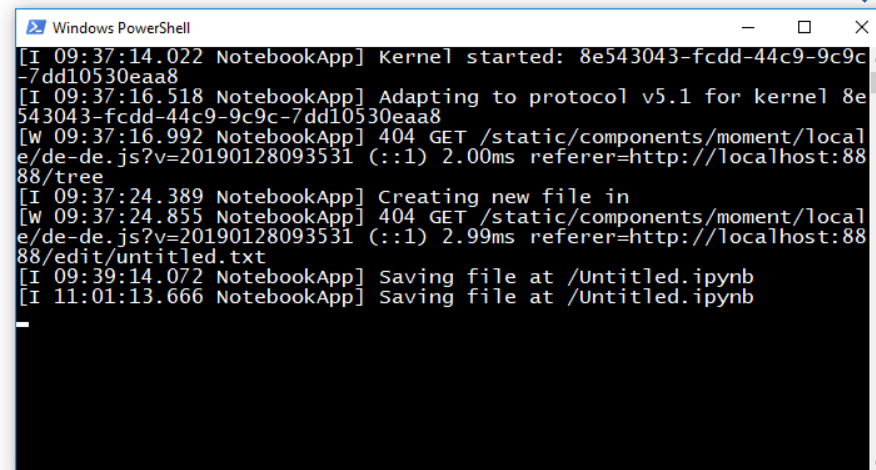
```
> This is a quotation
```

# Shut down Jupyter

- Closing the browser (or the tab) will not close the Jupyter server
- Option 1: click on **Quit** in the jupyter homepage



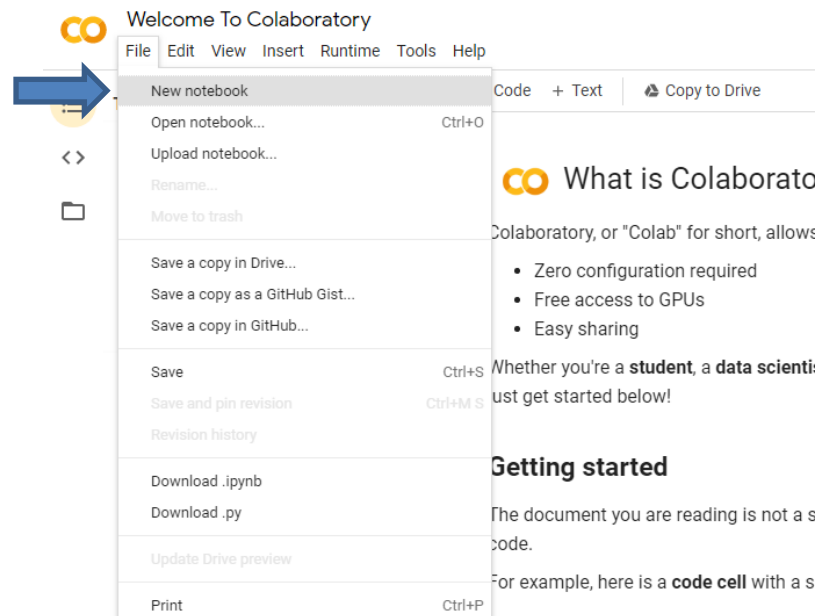
- Option 2: close the associated terminal or press “Ctrl” + “C”



# Google Colab

- Runs in the cloud on Google Servers and free to use with Google account
- Uses jupyter notebooks with a modified interface, which are saved in your google drive

Create new notebook

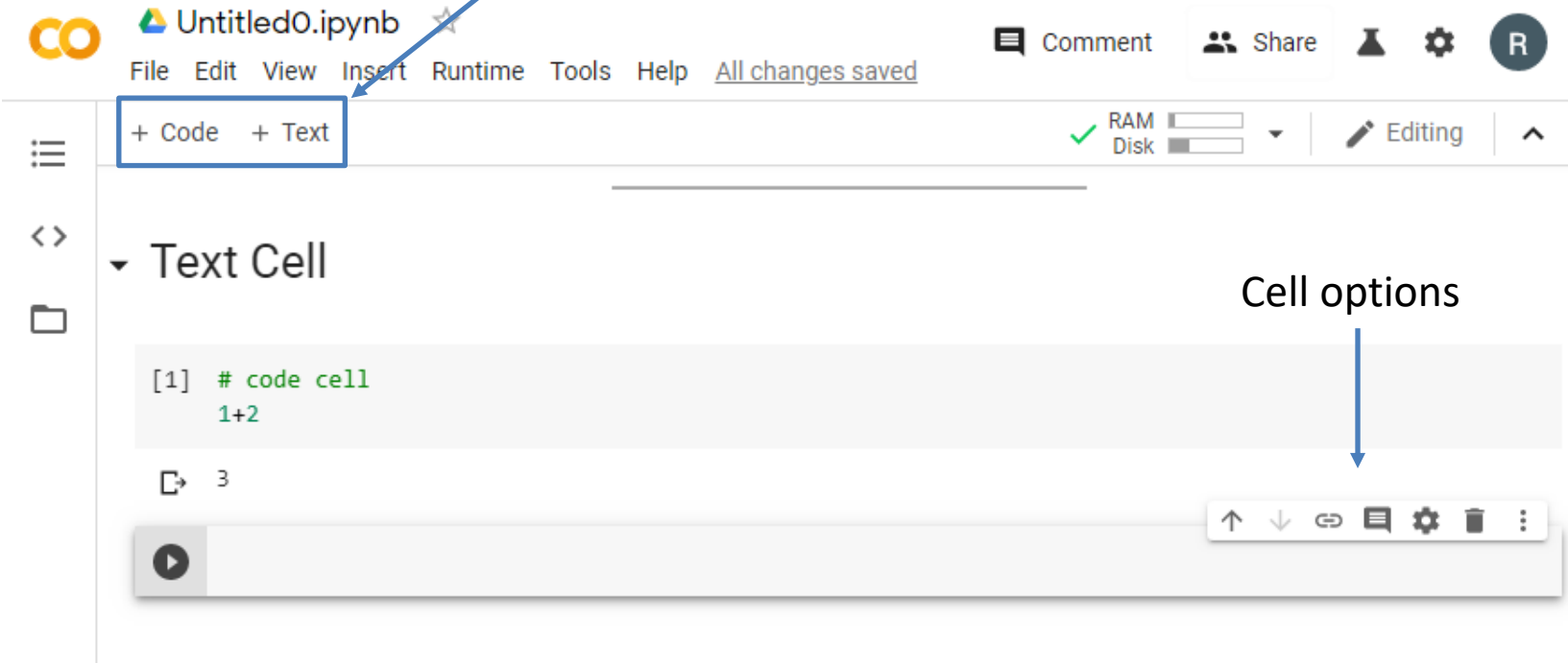


# Google Colab Layout

- Code and Text cells
- Shortcuts like Shift+Enter work the same way

Change name of notebook

Insert code or text cell

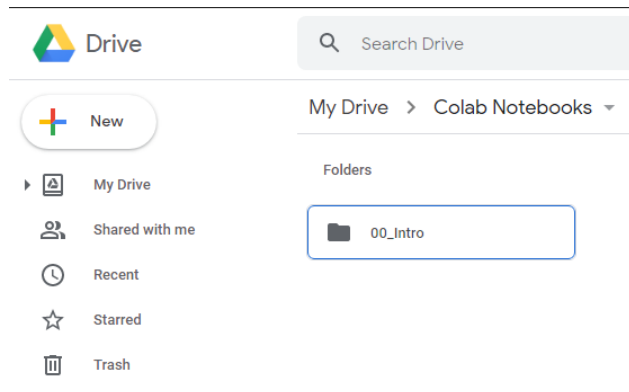


The screenshot shows the Google Colab interface. At the top, the notebook is titled 'Untitled0.ipynb'. Below the title bar, there is a menu with 'File', 'Edit', 'View', 'Insert', 'Runtime', 'Tools', and 'Help'. A blue arrow points from the text 'Change name of notebook' to the notebook title. Another blue arrow points from the text 'Insert code or text cell' to the '+ Code' and '+ Text' buttons in the left sidebar. The main workspace contains a 'Text Cell' with the text '[1] # code cell' and '1+2'. A blue arrow points from the text 'Cell options' to the cell options menu at the bottom right of the cell, which includes icons for up, down, link, comment, settings, delete, and a more options menu.

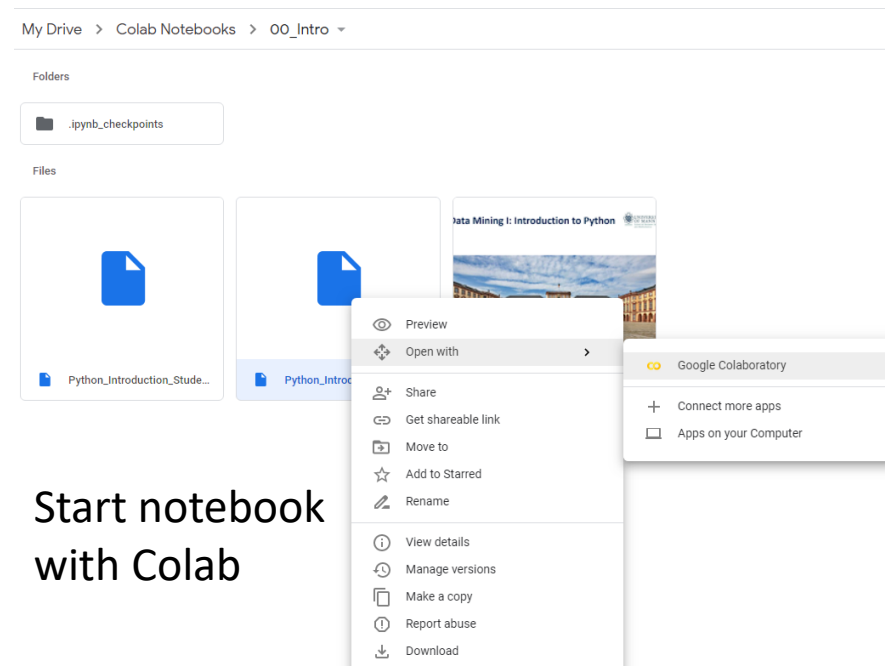


# Google Colab

- Saves notebooks to folder “Colab Notebooks” on your Google Drive by default
- To facilitate exercises and have stuff “just work”: upload/unzip exercise material in this folder



File Location



Start notebook  
with Colab