

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)

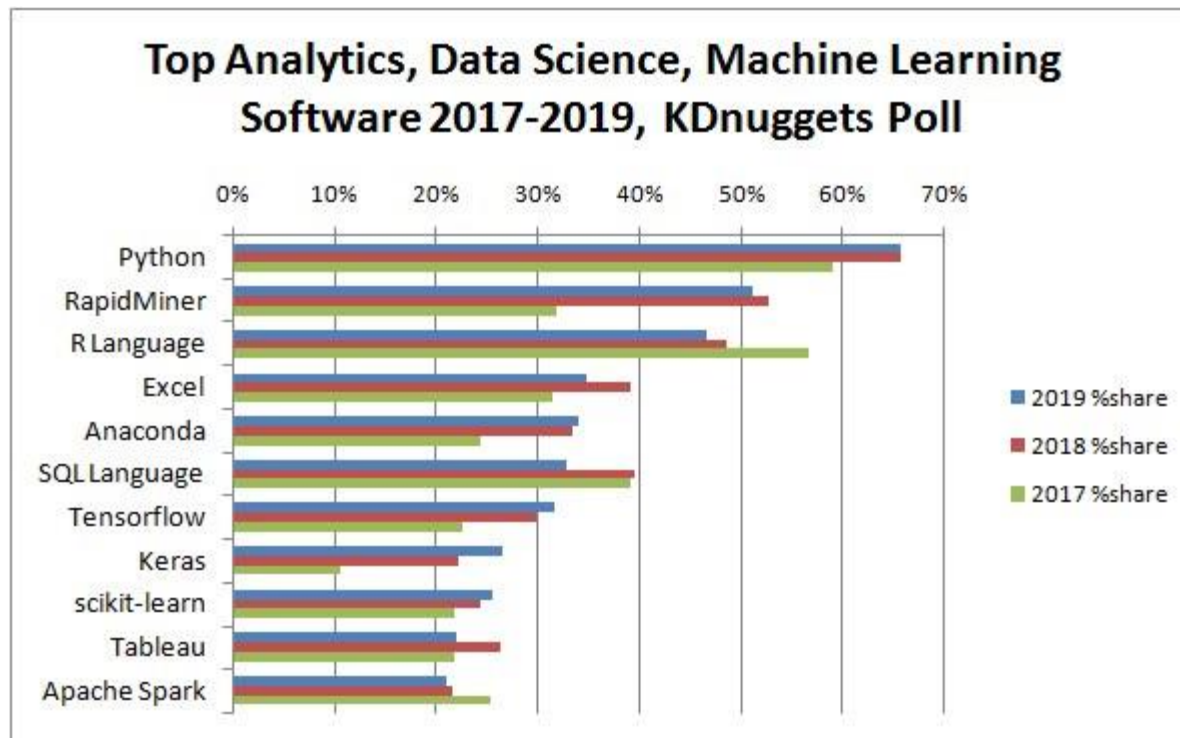
- If you don't have at least 3 GB disc space
 - Option 1 (better): Get a bigger disc!
 - Option 2: install miniconda
 - <https://docs.conda.io/en/latest/miniconda.html>
- Alternative: Use [Google Colab](#)

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

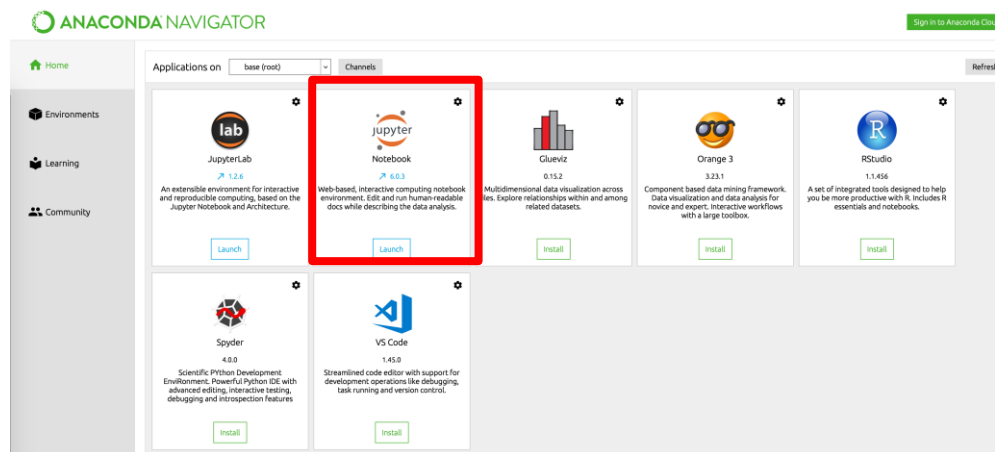


Popularity



Jupyter notebooks

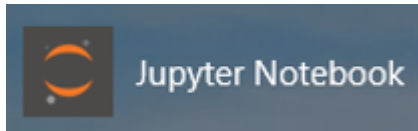
- In the exercises, we will use Jupyter Notebooks
- The start procedure depends on your operating system
 - in general: run the **Anaconda Navigator** and click on **jupyter Notebook**



- 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

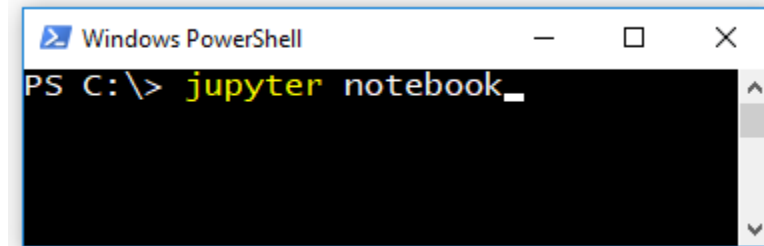
- Start Jupyter - Option 1 (Windows)



- 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

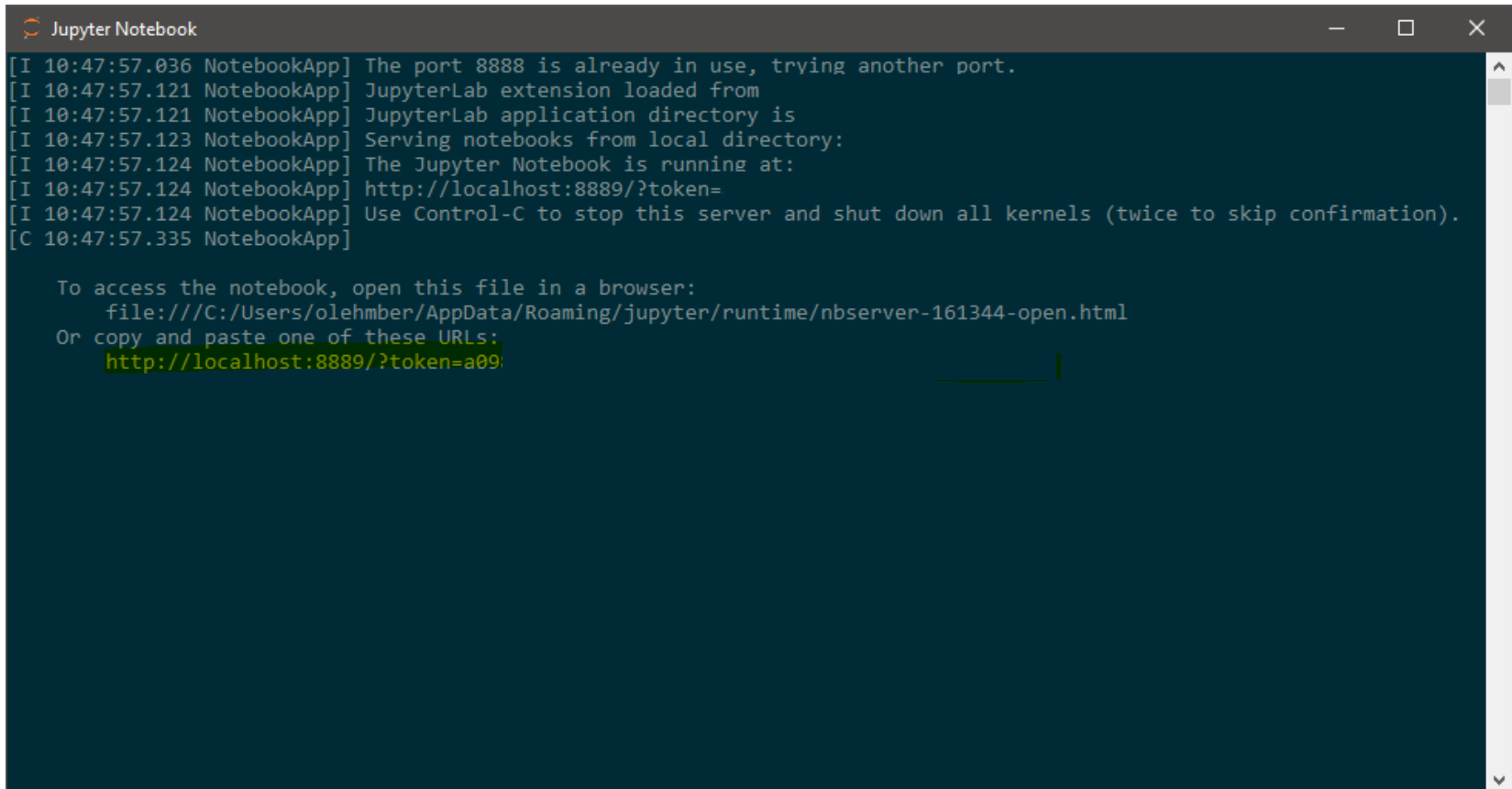
- Start Jupyter – Option 2 (Linux and Windows)
 - Run „jupyter notebook“ in command line
 - Navigate to the folder that you want to access before!
 - Or (Windows): “Shift+Right Click” in the corresponding folder and then "open command window/power shell here“



- Start Jupyter – Option 3 (Mac OS)
 - Click on spotlight, type “terminal” to open a terminal window
 - 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



jupyter

Logout

Files Running Clusters

Select items to perform actions on them.

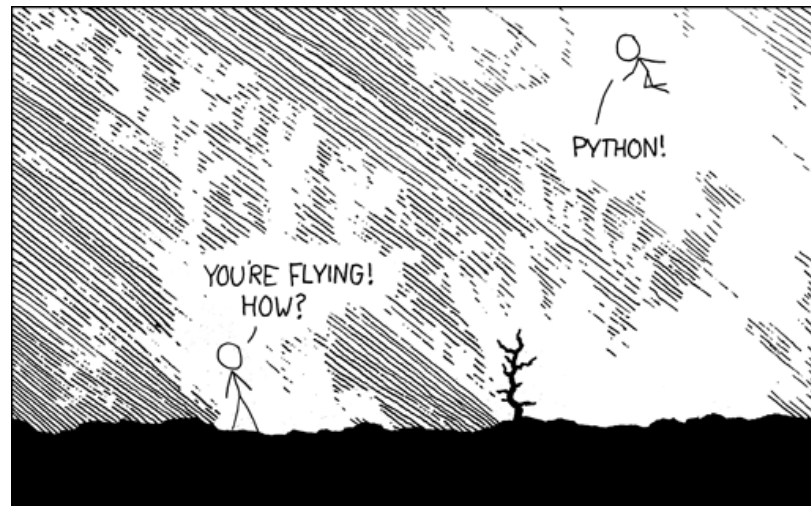
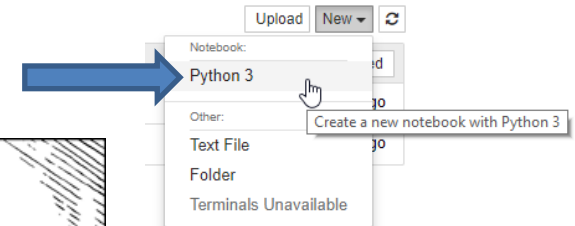
Upload New Refresh

	Name ↓	Last Modified
<input type="checkbox"/>	0	
<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/>

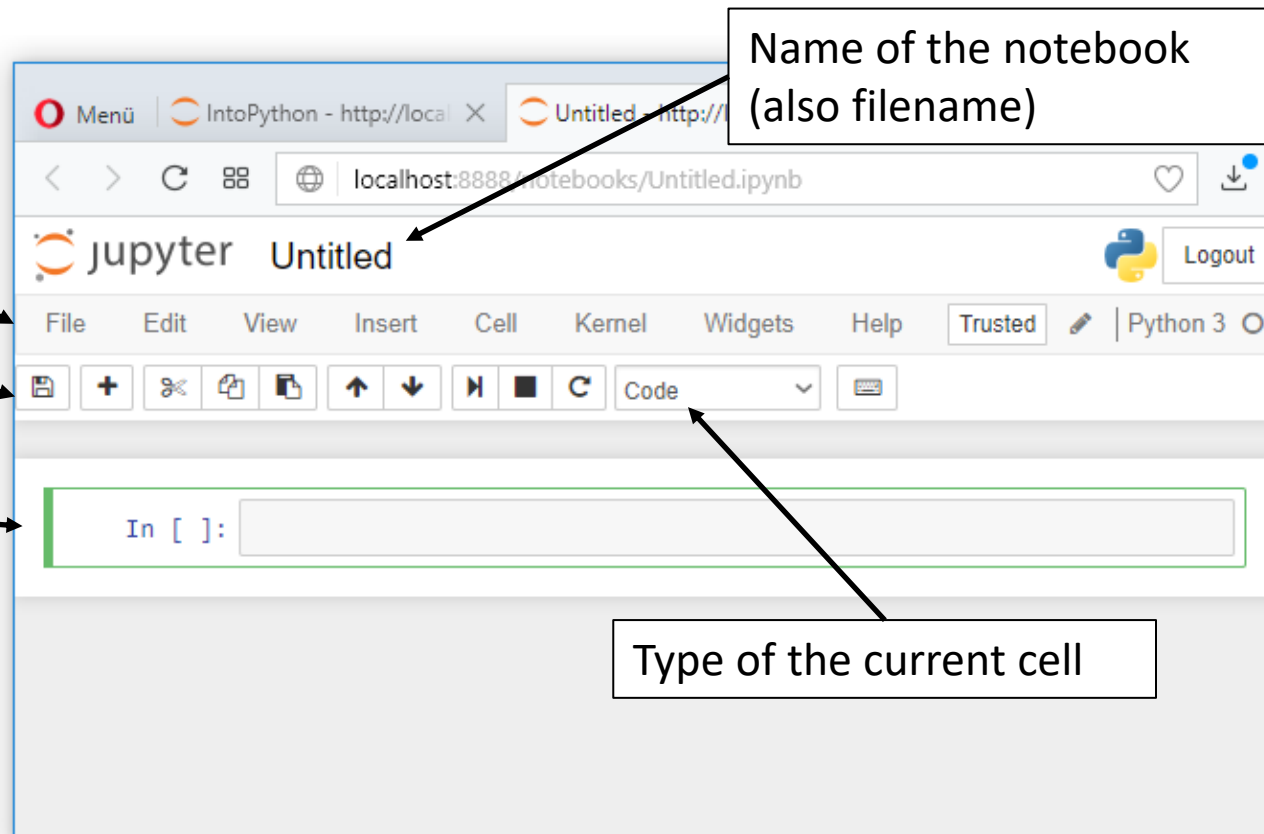
Jupyter Notebook

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

Menu Bar

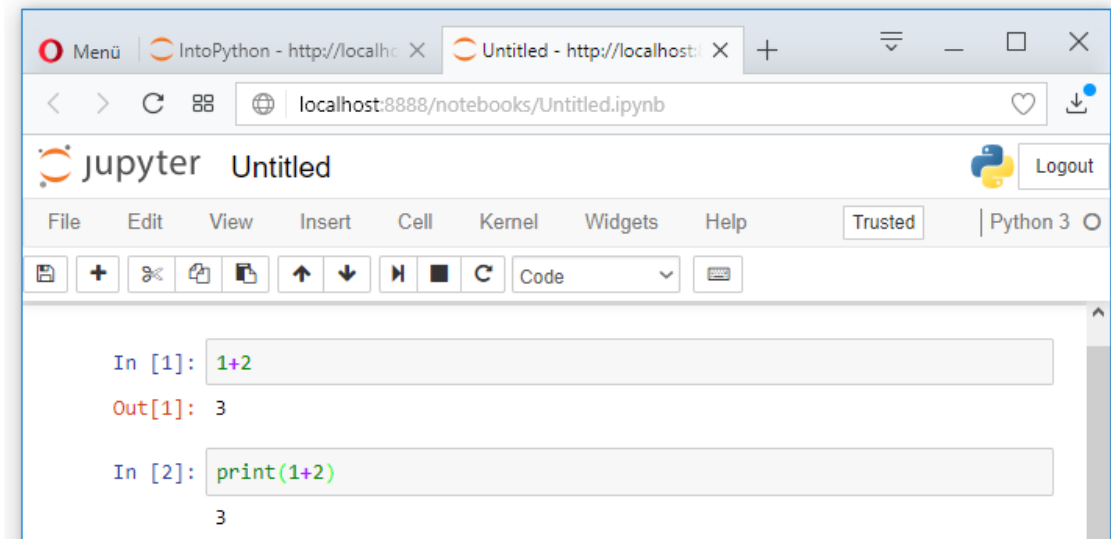
Shortcuts

Cell



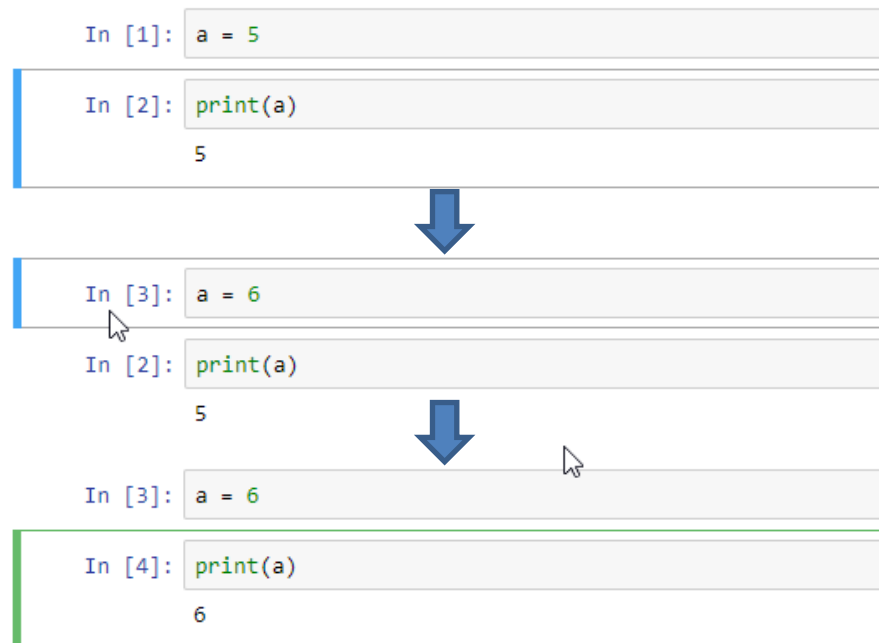
Jupyter Cells

- 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



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>

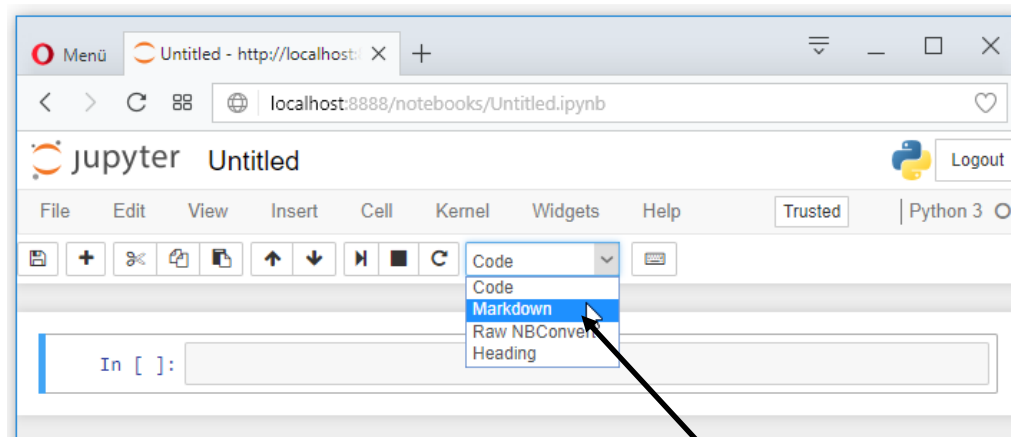
```
corr(datar, color_grades=5)
```

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

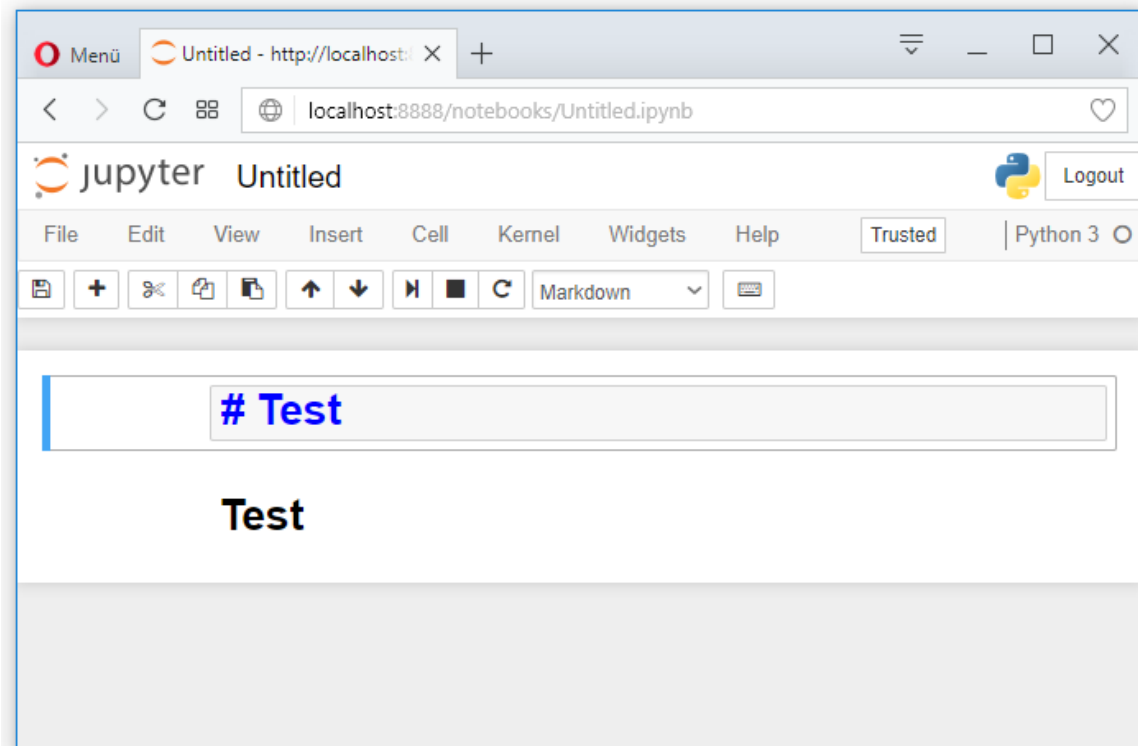
- What makes a notebook a notebook?
 - Markdown cells let you add documentation and notes
 - Create a new cell („Insert->Insert Cell Below“)
 - Change the type to Markdown



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

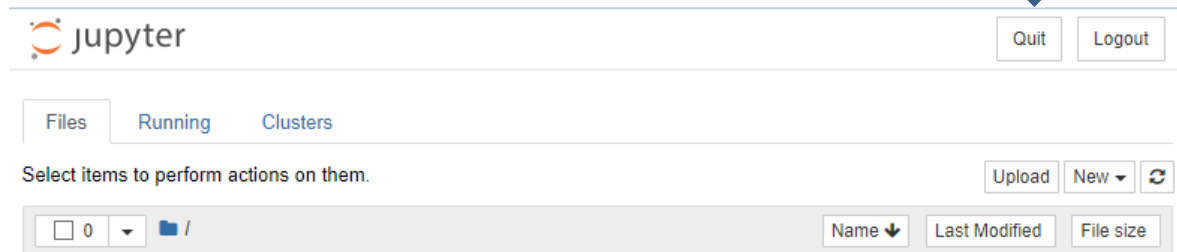
```
![Mannheim Image](https://www.uni-mannheim.de/1/00_UM_Dachmarke_DE_RGB.jpg)
```

- Quote

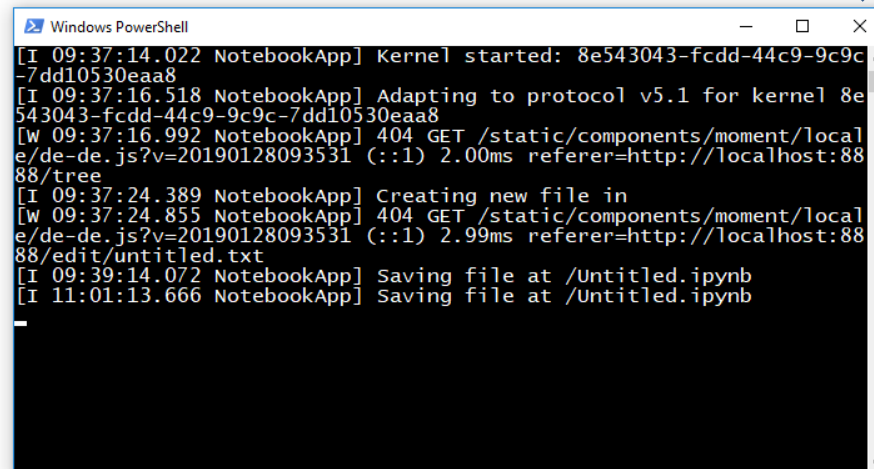
```
> This is a quotation
```

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”

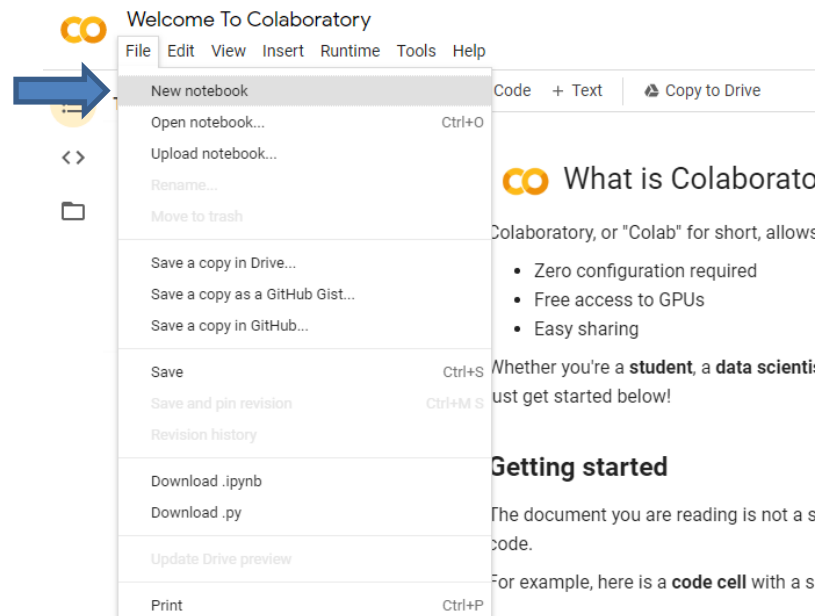


```
Windows PowerShell
[I 09:37:14.022 NotebookApp] Kernel started: 8e543043-fcdd-44c9-9c9c-7dd10530eaa8
[I 09:37:16.518 NotebookApp] Adapting to protocol v5.1 for kernel 8e543043-fcdd-44c9-9c9c-7dd10530eaa8
[W 09:37:16.992 NotebookApp] 404 GET /static/components/moment/locale/de-de.js?v=20190128093531 (::1) 2.00ms referer=http://localhost:8888/tree
[I 09:37:24.389 NotebookApp] Creating new file in
[W 09:37:24.855 NotebookApp] 404 GET /static/components/moment/locale/de-de.js?v=20190128093531 (::1) 2.99ms referer=http://localhost:8888/edit/untitled.txt
[I 09:39:14.072 NotebookApp] Saving file at /Untitled.ipynb
[I 11:01:13.666 NotebookApp] Saving file at /Untitled.ipynb
```

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

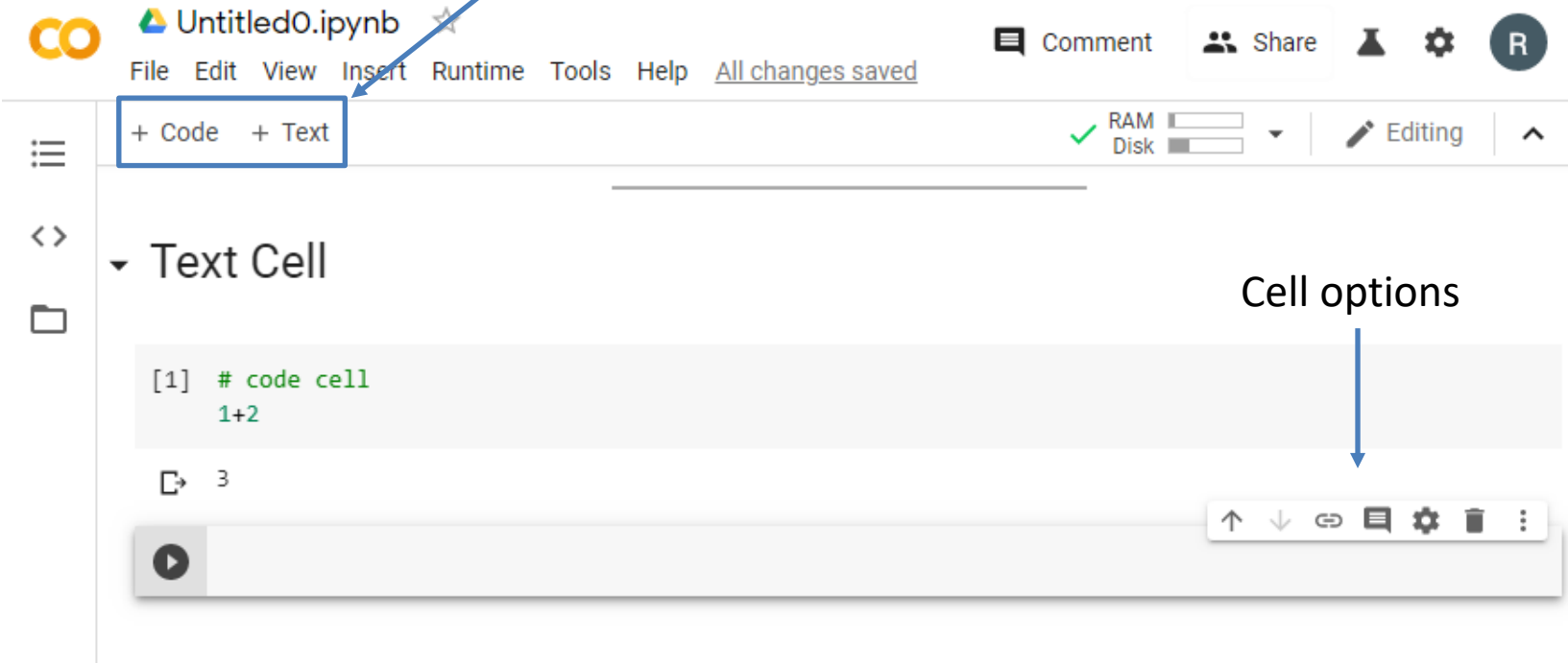


Google Colab Layout

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

Change name of notebook

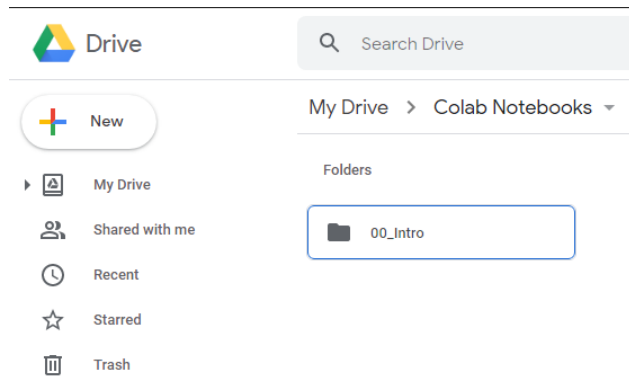
Insert code or text cell



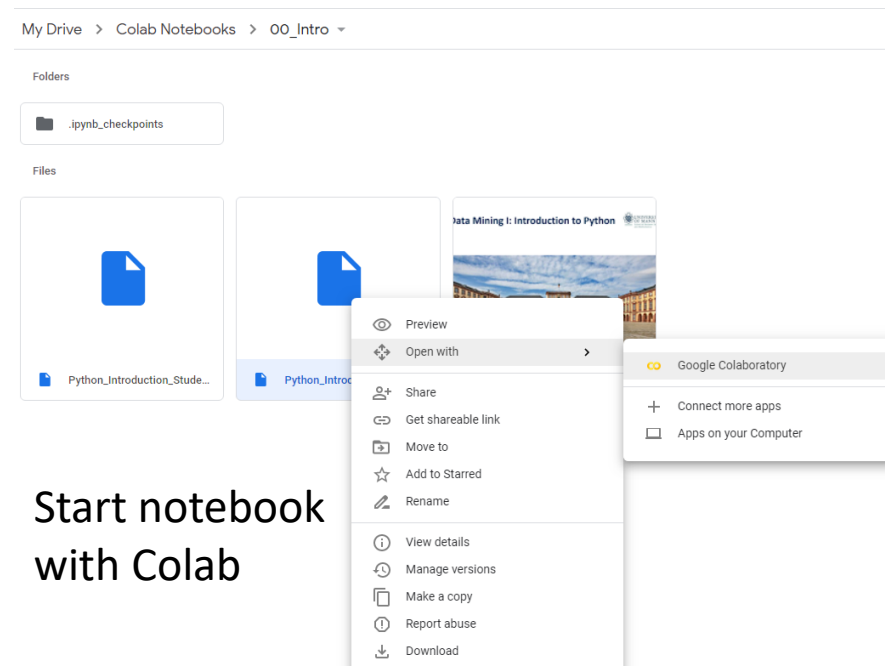
The screenshot displays the Google Colab interface. At the top, the notebook is titled 'Untitled0.ipynb'. Below the title bar, there is a menu with options: File, Edit, View, Insert, Runtime, Tools, and Help. A blue arrow points from the text 'Change name of notebook' to the notebook title. Another blue arrow points from the text 'Insert code or text cell' to the '+ Code' and '+ Text' buttons in the left sidebar. The main workspace shows a 'Text Cell' containing the code '[1] # code cell' and '1+2'. A blue arrow points from the text 'Cell options' to the options menu (three dots) at the bottom right of the cell. The interface also includes a 'Comment' button, a 'Share' button, and a 'RAM Disk' indicator.

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