

Google Colab

Tips for working with data and collaborators

Dan Keemahill

USA TODAY data team

dkeemahill@gannett.com

What is Colab?

Python notebooks in Google's cloud

Helpful for machine learning: see [training](#) on Fast.AI

What is a Python notebook?

Files that run Python code and display text and images

Helpful for reproducible and shareable data analysis

[FirstPythonNotebook.org](https://www.firstpythonnotebook.org)

What is a Python notebook?

Notebooks are not typically development environments

pythonanywhere.com

Google Colab is by... Google

Many other notebook options:

[Jupyter](#), [RStudio](#), [Observable](#), [Azure notebooks](#)...

Things I like about Colab

My biases: I already use Python and Google tools

Easy to install tools in Colab

GPU for speed and computational power with big datasets

Demo: Parsing PDFs

This [notebook](#) installs pdfplumber and ImageMagick, downloads a PDF and parses it to a Pandas dataframe, all in the browser!

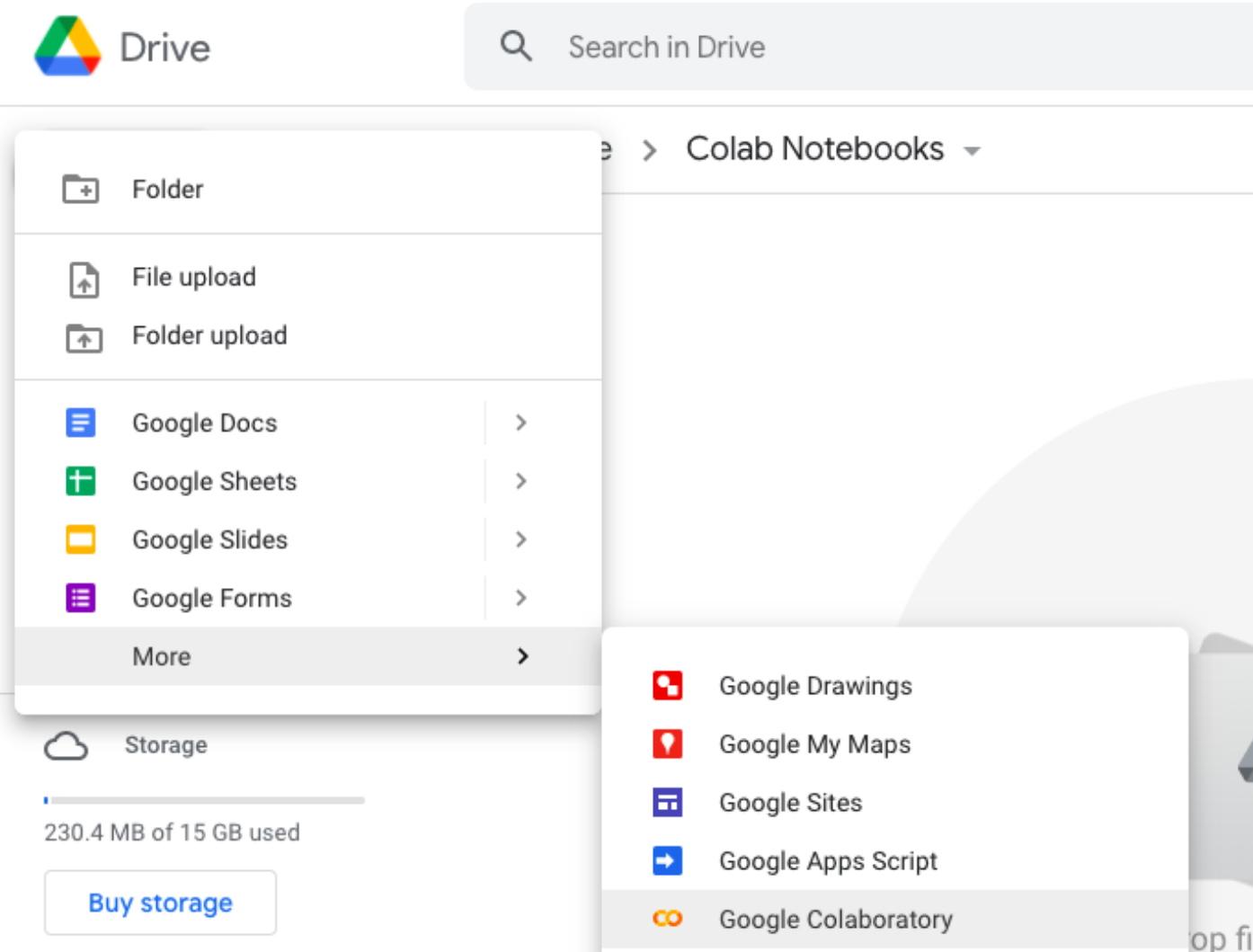
Tricky parts of Colab

Accessing Google resources

Sharing with collaborators

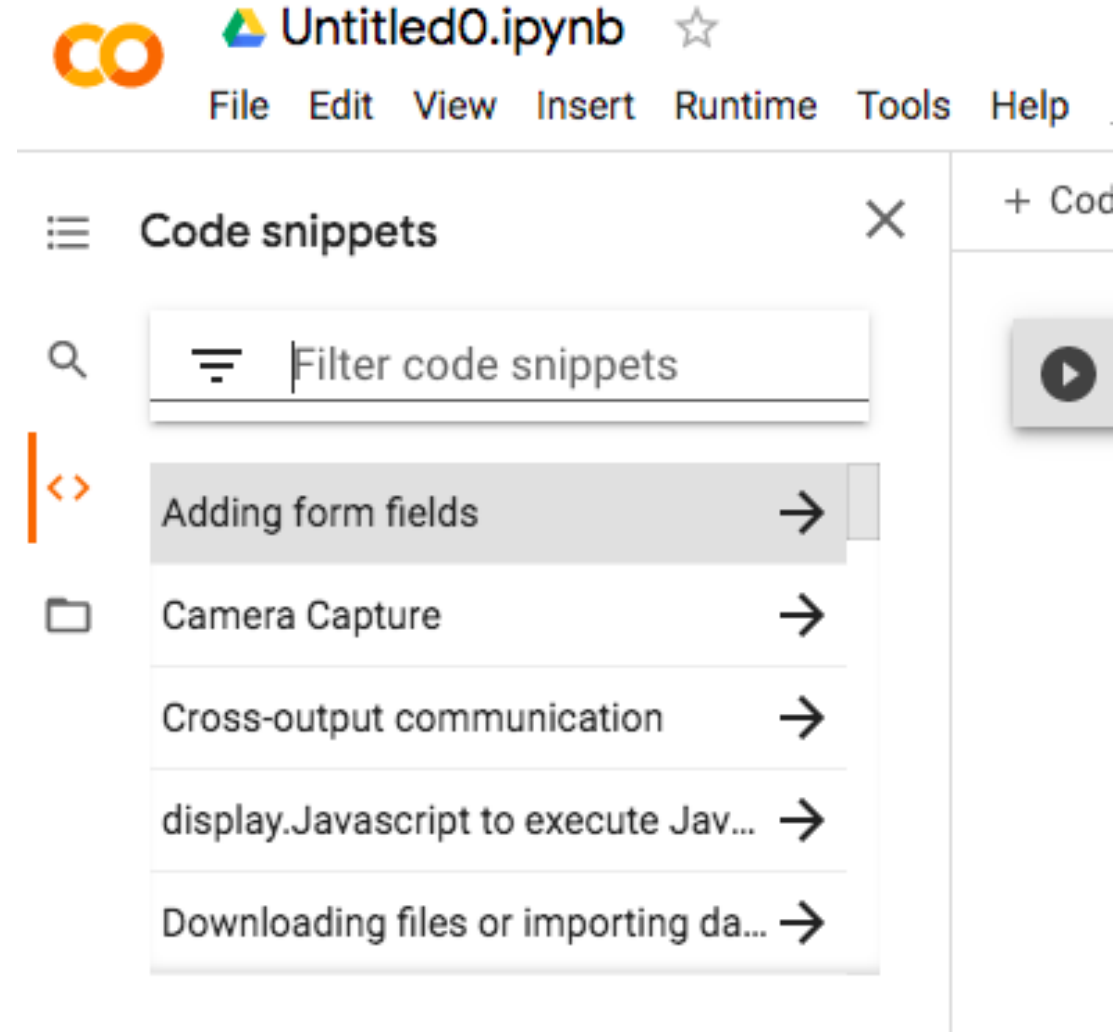
Managing resources

Start a Colab notebook



or go to colab.research.google.com
and click **New notebook**

Code snippets



The image shows a Jupyter Notebook interface. At the top, there is a logo with the letters 'CO' in orange and a star icon. The title of the notebook is 'Untitled0.ipynb'. Below the title is a menu bar with the following items: File, Edit, View, Insert, Runtime, Tools, and Help. On the left side, there is a sidebar titled 'Code snippets' with a close button (X) and a search icon. Below the title is a search input field with the placeholder text 'Filter code snippets'. Below the search field is a list of code snippets, each with a right-pointing arrow:

- Adding form fields →
- Camera Capture →
- Cross-output communication →
- display.Javascript to execute Jav... →
- Downloading files or importing da... →

On the right side of the sidebar, there is a button with a play icon and the text '+ Cod'.

Demo: Accessing your Google Drive

Several techniques described here:

<https://colab.research.google.com/notebooks/io.ipynb>

Sharing notebooks

Beware which parts of your notebook are unique to you, especially when "Mounting Google Drive"

I often "Save a Copy" before editing someone else's notebook

Playground Mode – run code without saving output

Managing resources

There are limits, but they can be vague

<https://research.google.com/colaboratory/faq.html#resource-limits>

Disk space, memory, timeouts, number of notebooks being run at the same time

Memory issues: be careful with for loops

Always be learning!

Join [USA TODAY Network Data](#) in Microsoft Teams

Visit training.usatodaynetwork.com