

# CS501

## Introductory Computation for Scientists



Dr. Ash Pahwa

---

### Lesson 2.3

### Google's Colab



# Outline

---

- Colab
  - What is Colab?
  - Cost of Colab
  - Benefits and Drawbacks of Colab
  - How to Use Colab
  - Data IO in Colab



# Google's Colab

---



# What is Colab?

---

- Colaboratory or “Colab” is a Google research project created to help disseminate machine learning education and research
- It's a Jupyter notebook environment that requires no setup to use and runs entirely in the cloud



# When did Google Colab come out? 2017

---

- **Google** first started working with the Jupyter Development Team in 2014 to **release** an early version of the tool
- Google Colab is identical in Jupyter Notebook
- Google have released Colaboratory: a web IDE for python, to enable Machine Learning with storage on the cloud
- This internal tool had a quiet public release in late 2017 and is set to make a huge difference in the world of machine learning, artificial intelligence and data science work



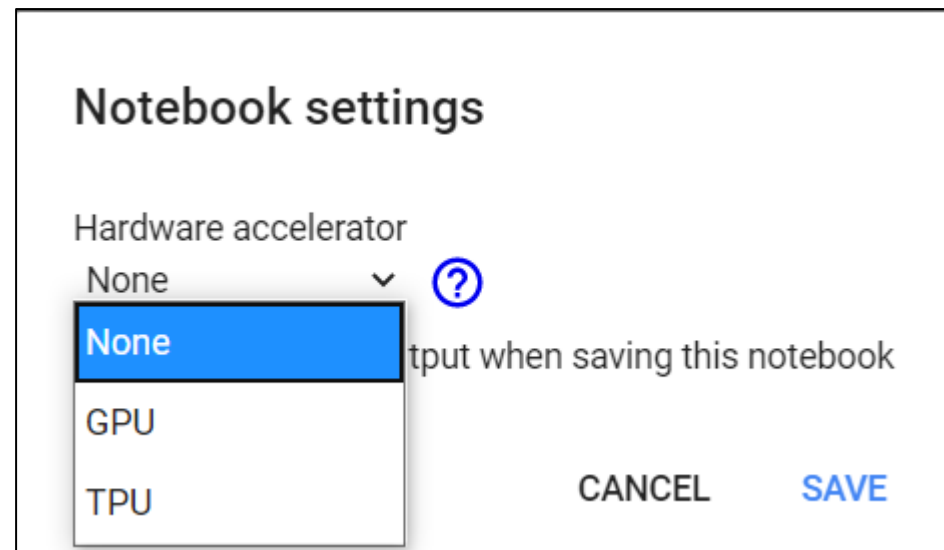
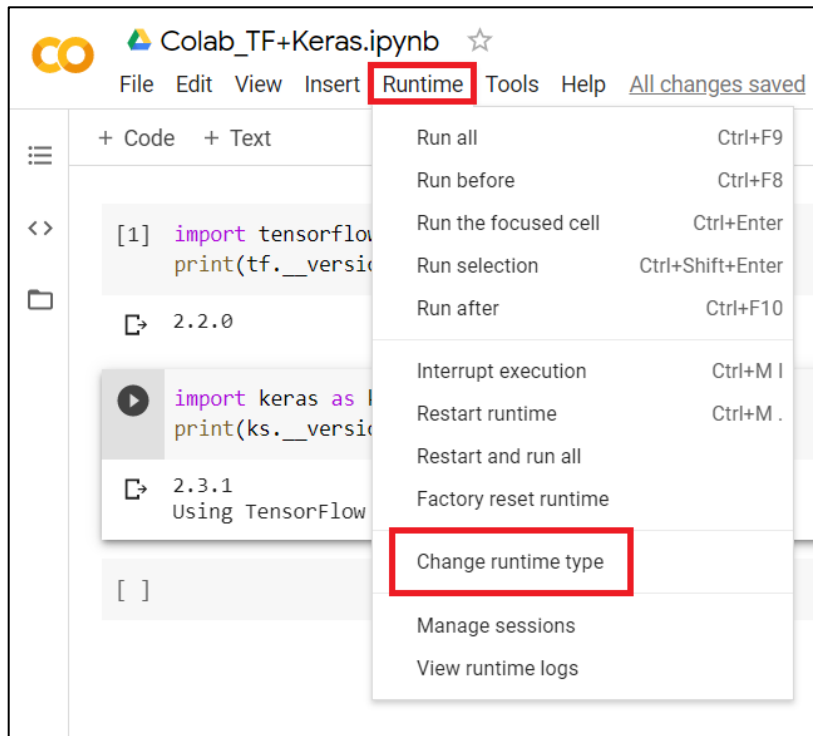
# Cost of Using Colab

---

- Google Colab is a **free** to use Jupyter notebook
- Allows you to use **free** Tesla K80 GPU
- RAM of 12 GB
- You can use Colab
  - 24 hours continuously without GPU
  - 12 hours with a GPU

# GPU Access

- Runtime > Change runtime type
- Select GPU as Hardware accelerator





# Benefits and Drawbacks of Google's Colab

---





# Benefits of Colab

---

- **Free** virtual machines for you to use with about 12GB RAM and 50GB hard drive space, with common dependencies such as numpy, pandas, and even TensorFlow pre-installed.
- **Free** GPU access
- Supports Python 2 and Python 3
- There is integration with Google Drive, you can share, and control permissions and you'll be able to see other collaborators work instantly
- You can import an existing Jupyter/IPython notebooks



# Drawbacks of Colab

---

- All Colab notebooks must be stored in Google Drive — so you'll need to log into a Google account before you can access the tool
- Long-running background computations may be stopped
- You'll need to install all specific libraries which do not come with standard python (and you'll need to repeat this with every session)
- Google Storage is used with your current session, so if you have downloaded a file and want to use it later, you'd better save it before closing the session
- It can be difficult (and potentially costly) to work with bigger datasets as you must download and store them in Google drive (only **15GB** is free in Google Drive)

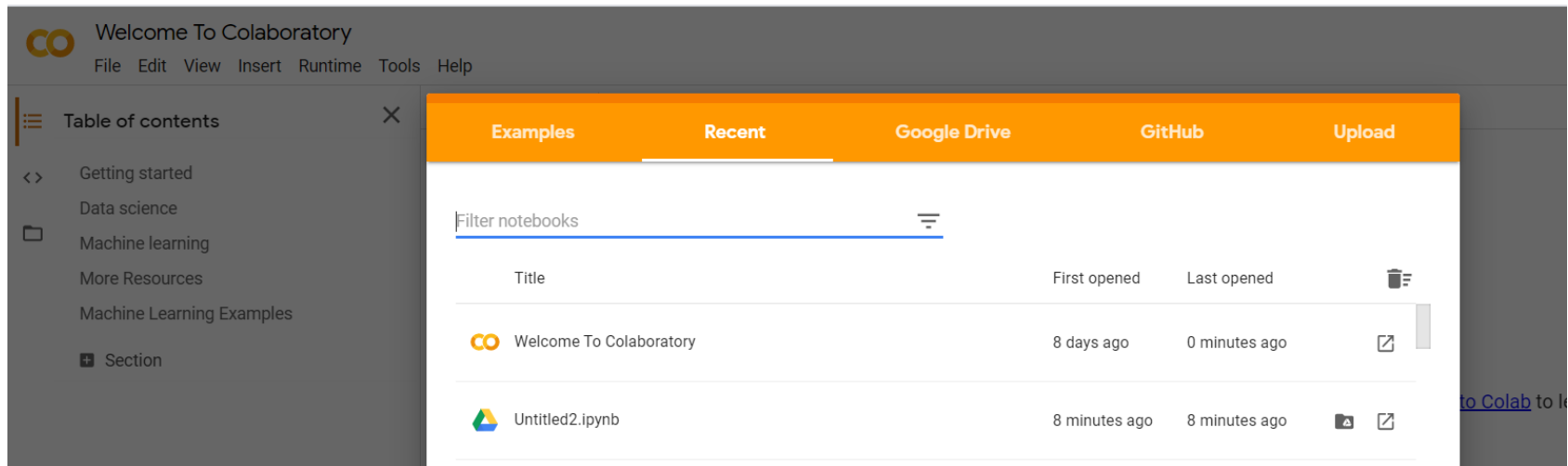


# How do I use Google Colab?





---

# How do I use Google Colab?

- To start working with **Colab** you first need to log in to your **google** account, then go to this link
- <https://colab.research.google.com>.
  - **EXAMPLES**: Contain a number of Jupyter notebooks of various examples.
  - **RECENT**: Jupyter notebook you have recently worked with.
  - **GOOGLE DRIVE**: Jupyter notebook in your **google** drive.



The screenshot displays the Google Colaboratory web interface. On the left is a sidebar with a 'Table of contents' panel showing links like 'Getting started', 'Data science', 'Machine learning', and 'Machine Learning Examples'. The main area has a top navigation bar with tabs: 'Examples', 'Recent' (selected), 'Google Drive', 'GitHub', and 'Upload'. Below the 'Recent' tab is a search bar labeled 'Filter notebooks'. A table lists recent notebooks:

Title	First opened	Last opened	
 Welcome To Colaboratory	8 days ago	0 minutes ago	
 Untitled2.ipynb	8 minutes ago	8 minutes ago	

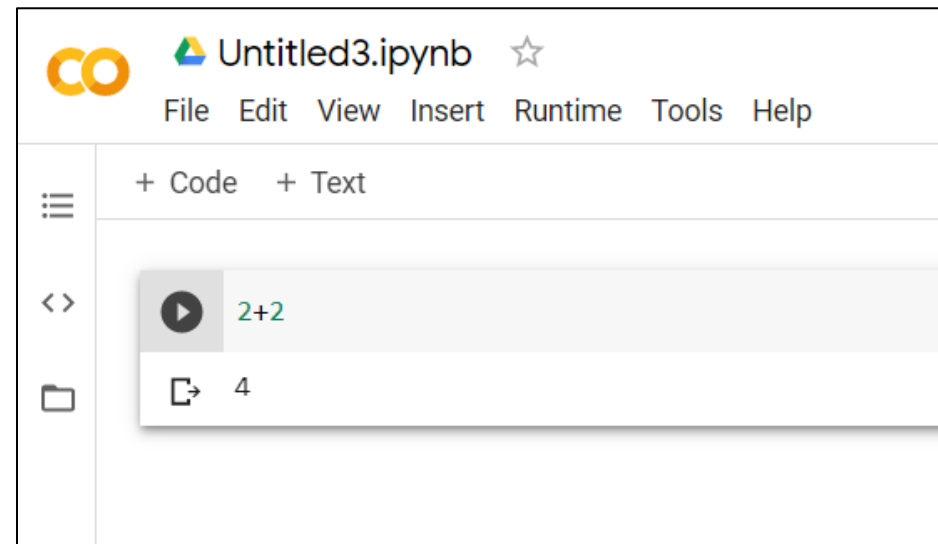
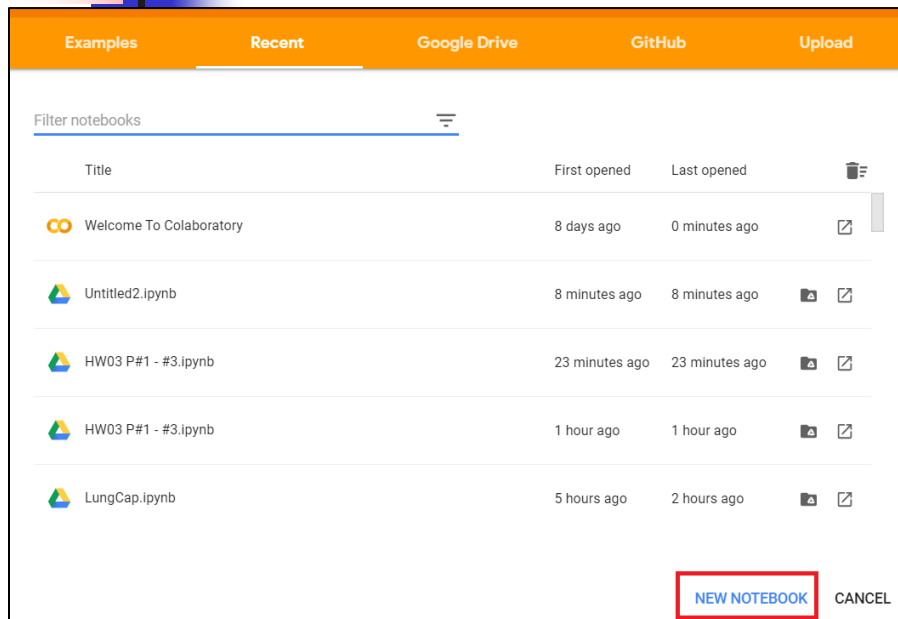


# Select Jupyter File

---

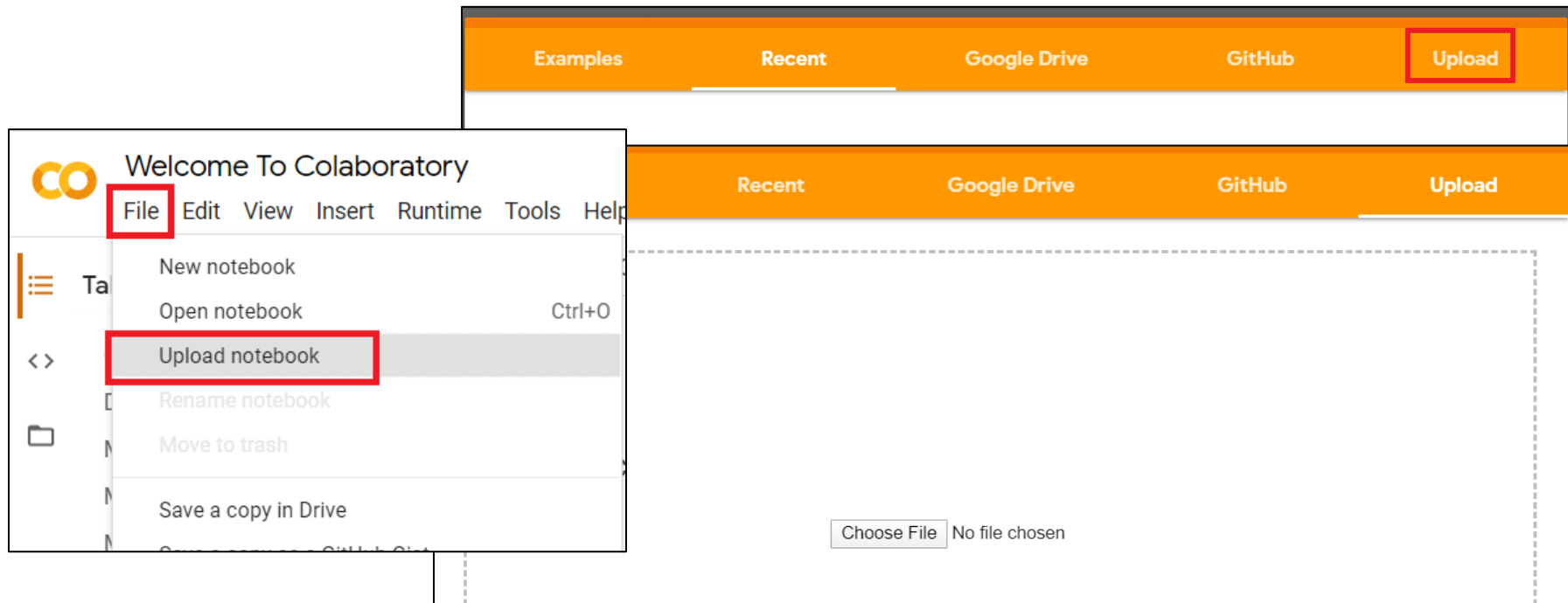
1. Start a New Jupyter File
2. Select a Jupyter file from your Computer
3. Select a Jupyter file from Google Drive

# 1. Start a New Jupyter File Select: New Notebook

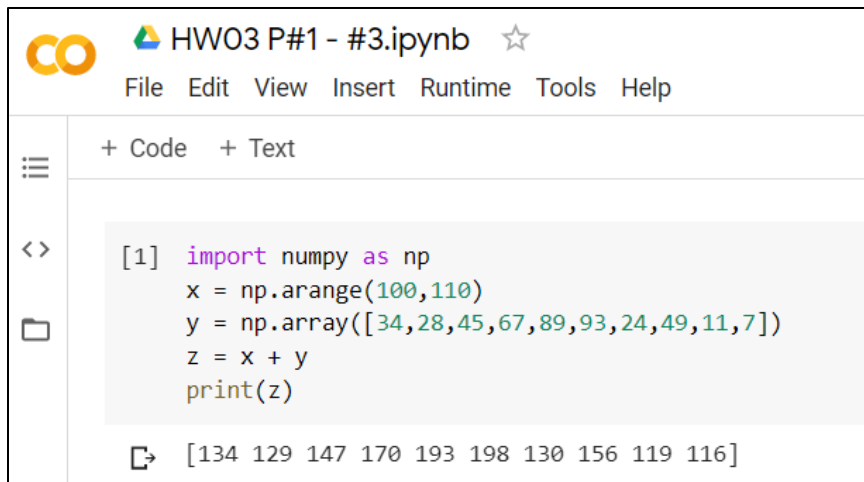


## 2. Select a Jupyter file from your Computer

- Upload: Select a file from your computer



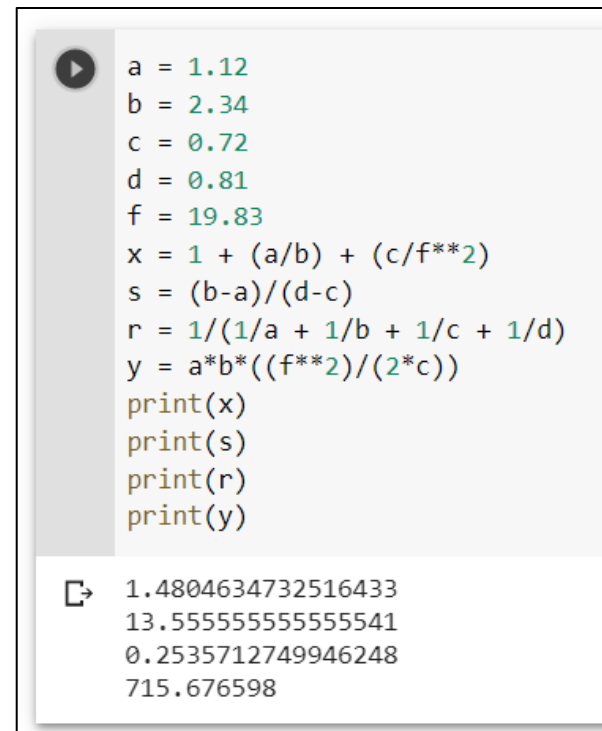
# Run Python Code in Colab



The image shows a Google Colab notebook interface. At the top, there's a logo and the title "HW03 P#1 - #3.ipynb" with a star icon. Below the title is a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". Under the menu bar, there are tabs for "+ Code" and "+ Text". The main area contains a code cell with the following Python code:

```
[1] import numpy as np
x = np.arange(100,110)
y = np.array([34,28,45,67,89,93,24,49,11,7])
z = x + y
print(z)
```

Below the code cell, the output is displayed as a list of numbers: [134 129 147 170 193 198 130 156 119 116].



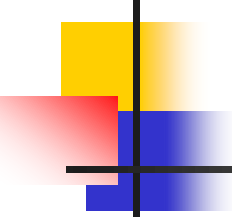
The image shows a Google Colab notebook interface. At the top, there's a play button icon. Below it, the code cell contains the following Python code:










```
a = 1.12
b = 2.34
c = 0.72
d = 0.81
f = 19.83
x = 1 + (a/b) + (c/f**2)
s = (b-a)/(d-c)
r = 1/(1/a + 1/b + 1/c + 1/d)
y = a*b*((f**2)/(2*c))
print(x)
print(s)
print(r)
print(y)
```

Below the code cell, the output is displayed as a list of numbers: 1.4804634732516433, 13.555555555555541, 0.2535712749946248, 715.676598.



### 3. Select a Jupyter file from Google Drive



Examples					Recent		Google Drive		GitHub		Upload	
Filter notebooks												
Title					Owner		Last modified		Last opened			
 Untitled3.ipynb					Ash Pahwa		9 minutes ago		9 minutes ago		 	
 Untitled2.ipynb					Ash Pahwa		19 minutes ago		19 minutes ago		 	
 HW03 P#1 - #3.ipynb					Ash Pahwa		34 minutes ago		34 minutes ago		 	



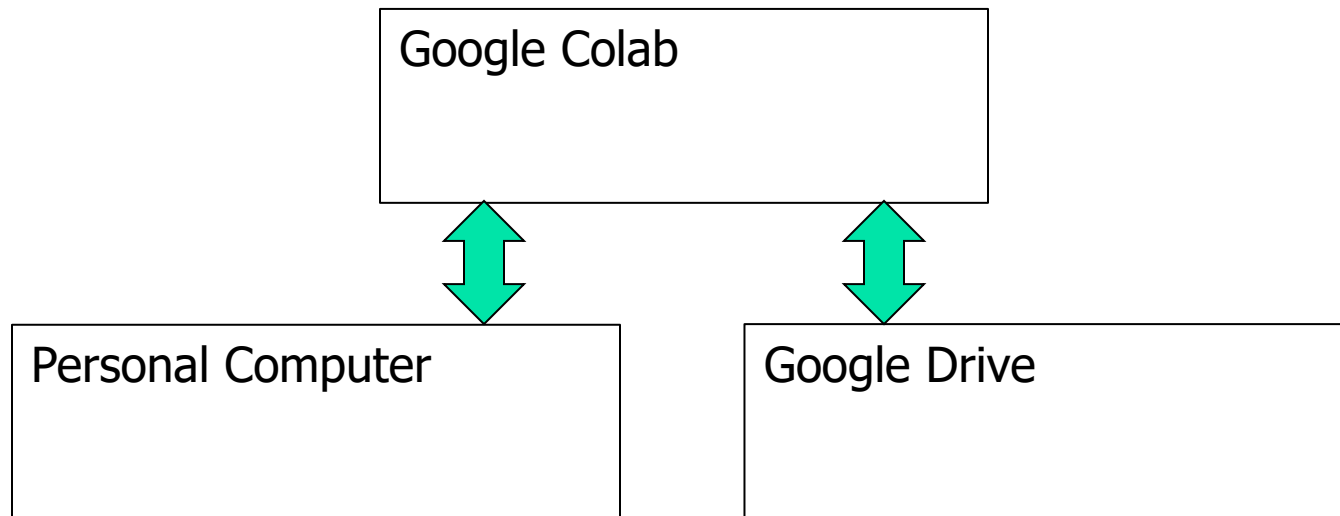
# Data IO in Colab

---

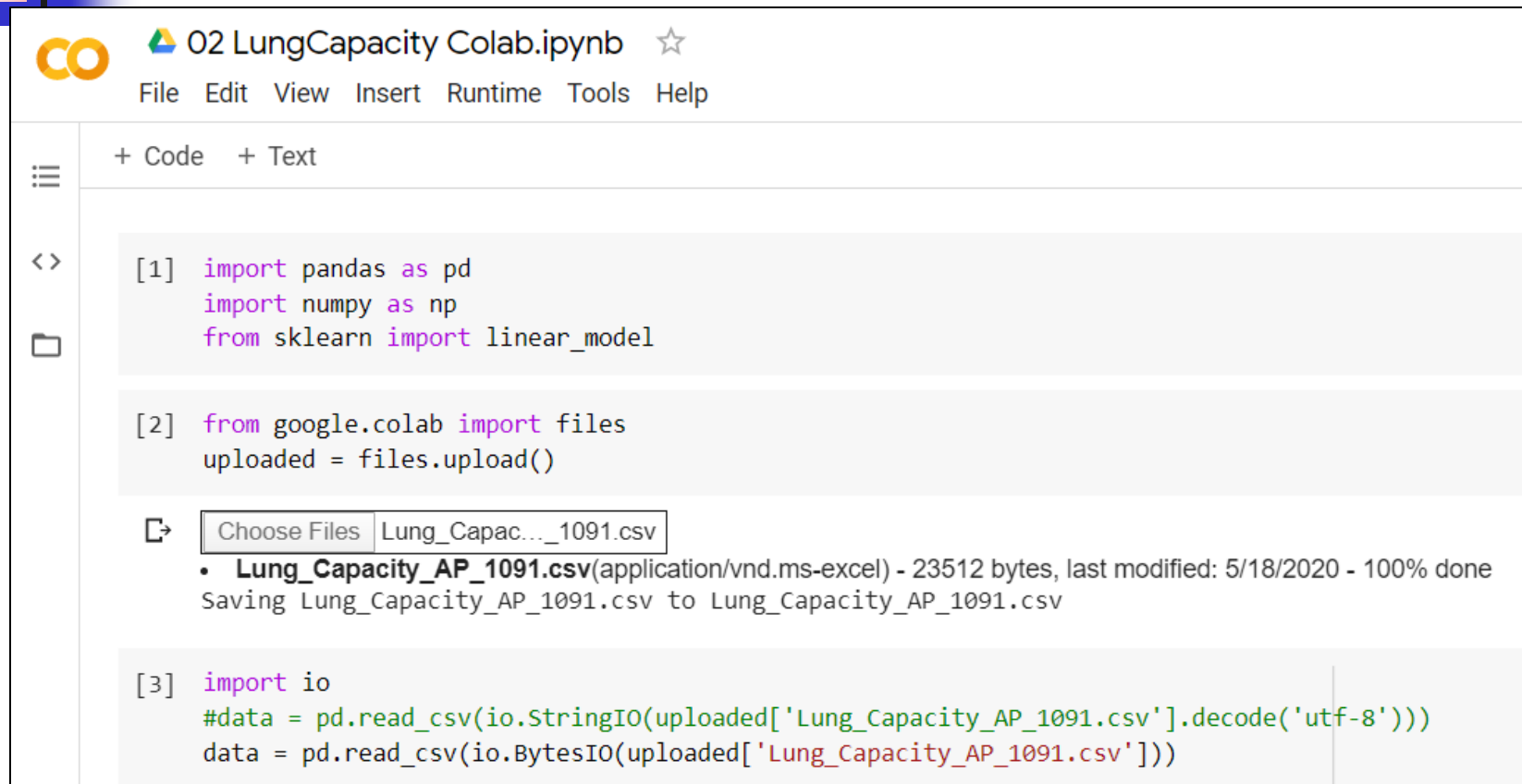
# Data (I/O) File in Colab

## Load Data, Save Data

- From Personal Computer
- From Google Drive



# Loading Data From Personal Computer



The image shows a Google Colab notebook titled "02 LungCapacity Colab.ipynb". The interface includes a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". On the left, there are icons for a menu, a code editor, and a file explorer. The main area contains three code cells. The first cell imports pandas as pd, numpy as np, and sklearn's linear\_model. The second cell imports files from google.colab and calls upload(). Below this, a file selection dialog shows "Lung\_Capac...\_1091.csv" selected. A message indicates the file was uploaded successfully (23512 bytes, 100% done) and is being saved as "Lung\_Capacity\_AP\_1091.csv". The third cell imports io and uses it to read the uploaded CSV file into a pandas DataFrame.

```
[1] import pandas as pd
import numpy as np
from sklearn import linear_model

[2] from google.colab import files
uploaded = files.upload()

Choose Files Lung_Capac..._1091.csv
• Lung_Capacity_AP_1091.csv(application/vnd.ms-excel) - 23512 bytes, last modified: 5/18/2020 - 100% done
Saving Lung_Capacity_AP_1091.csv to Lung_Capacity_AP_1091.csv

[3] import io
#data = pd.read_csv(io.StringIO(uploaded['Lung_Capacity_AP_1091.csv'].decode('utf-8')))
data = pd.read_csv(io.BytesIO(uploaded['Lung_Capacity_AP_1091.csv']))
```

# Loading Data From Personal Computer

```
[5] import io
     data = pd.read_csv(io.StringIO(uploaded['Lung_Capacity_AP_1091.csv'].decode('utf-8')))
```

```
▶ data = pd.read_csv("Lung_Capacity_AP_1091.csv")

print(data.head())
print(data.tail())
data.shape
```

```
↗
```

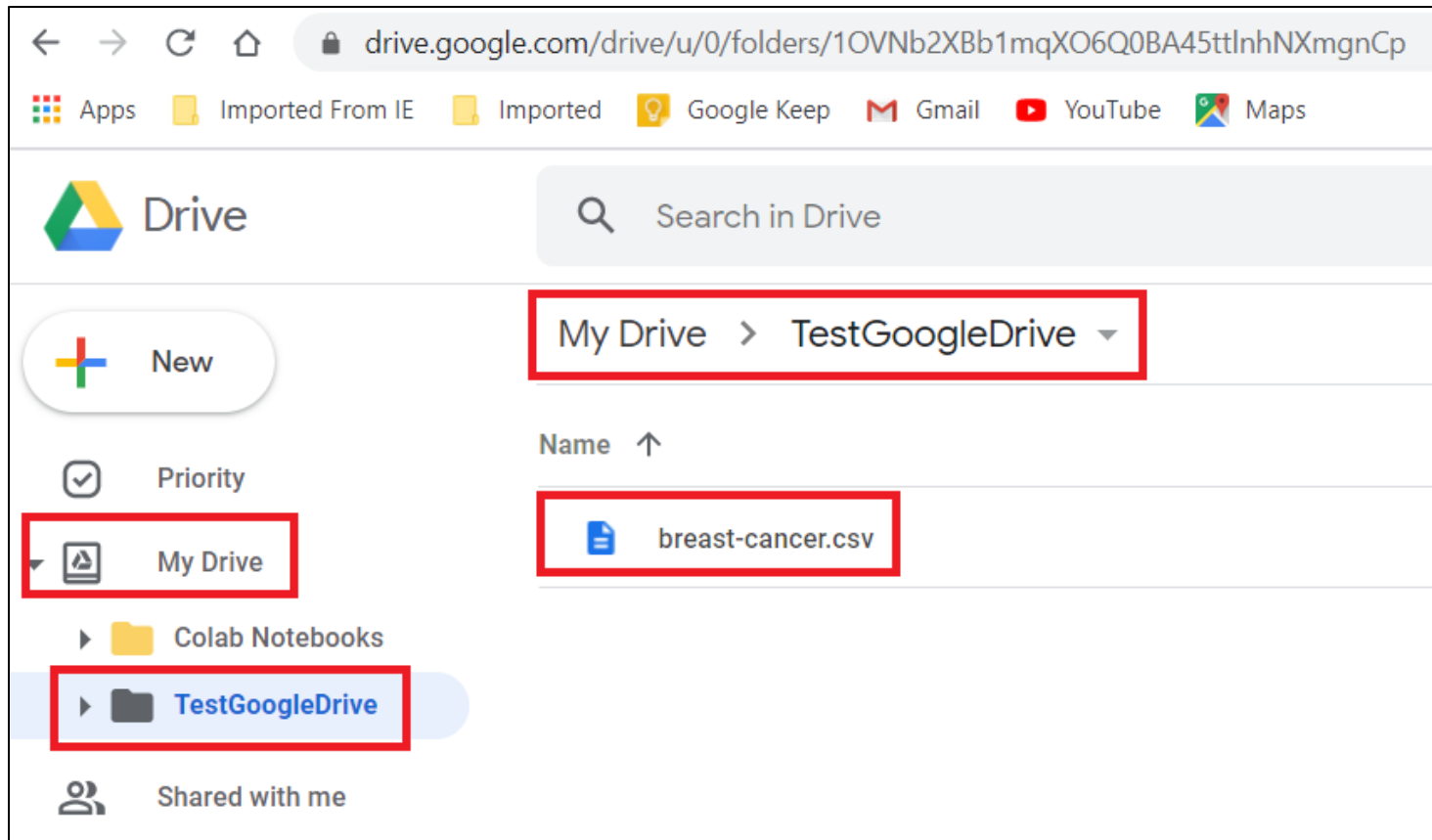
	Age	LungCap	Height	Gender	Smoke
0	9	3.124	57.0	female	no
1	8	3.172	67.5	female	no
2	7	3.160	54.5	female	no
3	9	2.674	53.0	male	no
4	9	3.685	57.0	male	no

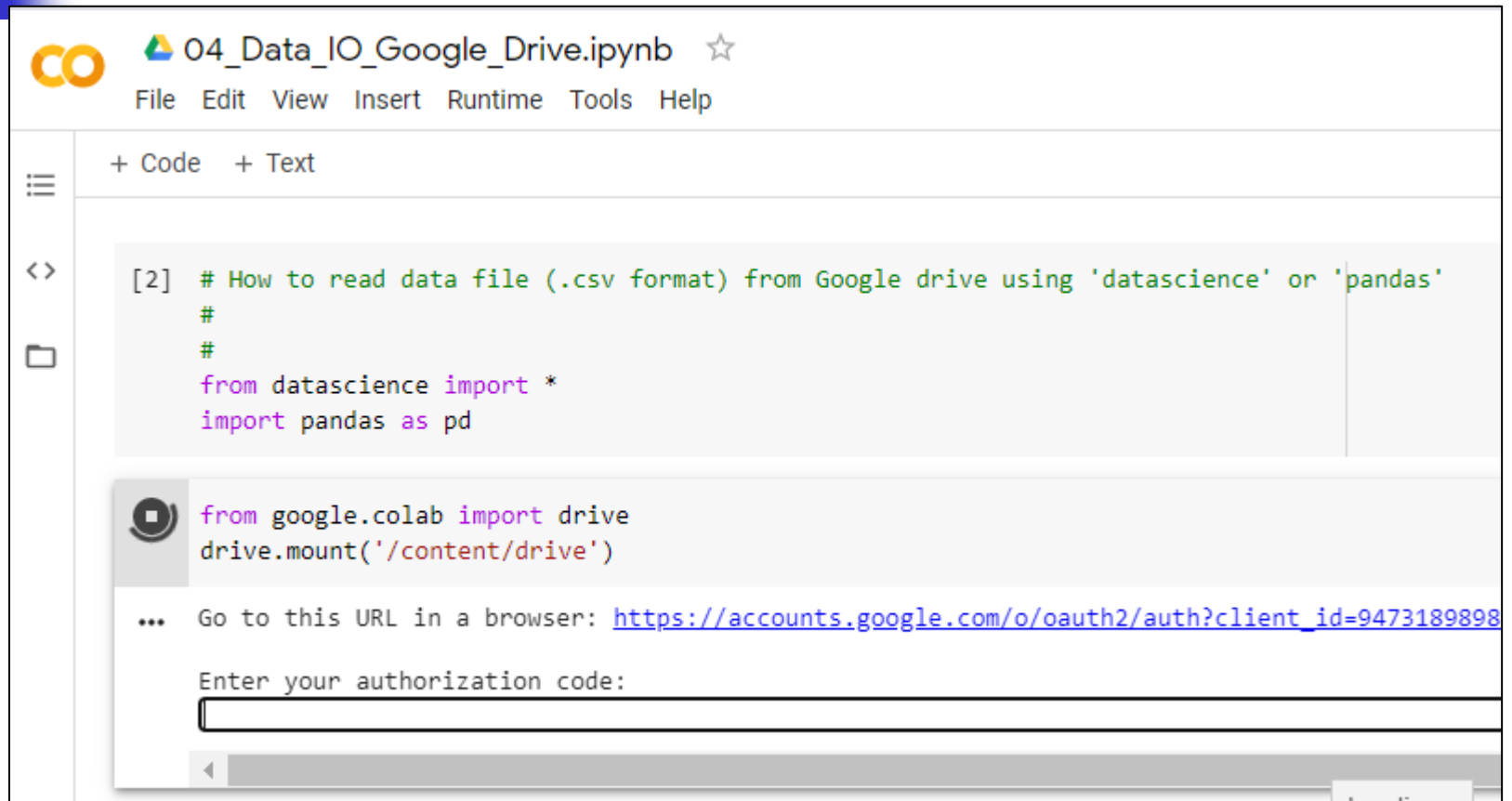
	Age	LungCap	Height	Gender	Smoke
1086	12	8.557	65.5	female	no
1087	13	10.696	70.5	male	no
1088	10	6.310	62.0	male	no
1089	12	8.023	65.5	female	no
1090	10	7.270	65.0	male	no

(1091, 5)

# Loading Data From Google Drive



# Loading Data From Google Drive



The image shows a Google Colab notebook titled "04\_Data\_IO\_Google\_Drive.ipynb". The interface includes a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", and "Help". Below the menu bar, there are buttons for "+ Code" and "+ Text". The notebook contains two code cells. The first cell is a comment: "[2] # How to read data file (.csv format) from Google drive using 'datascience' or 'pandas'". The second cell contains the following code: 

```
from datascience import *  
import pandas as pd  
  
from google.colab import drive  
drive.mount('/content/drive')
```

 Below the code cells, there is a message: "... Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=9473189898](https://accounts.google.com/o/oauth2/auth?client_id=9473189898)". Below the URL, there is a prompt: "Enter your authorization code:" followed by a text input field.


```
co 04_Data_IO_Google_Drive.ipynb ☆  
File Edit View Insert Runtime Tools Help  
  
+ Code + Text  
  
[2] # How to read data file (.csv format) from Google drive using 'datascience' or 'pandas'  
#  
#  
from datascience import *  
import pandas as pd  
  
from google.colab import drive  
drive.mount('/content/drive')  
  
... Go to this URL in a browser: https://accounts.google.com/o/oauth2/auth?client\_id=9473189898  
  
Enter your authorization code:  

```









# Give Permission to Access Google Drive

## Copy the Authorization Code

**Google Drive File Stream** wants to access your Google Account

 ash@ashpahwa.com

This will allow **Google Drive File Stream** to:


-  See, edit, create, and delete all of your Google Drive files 
-  View the photos, videos and albums in your Google Photos 
-  View Google people information such as profiles and contacts 
-  See, edit, create, and delete any of your Google Drive documents 

**Make sure you trust Google Drive File Stream**

You may be sharing sensitive info with this site or app. Learn about how Google Drive File Stream will handle your data by reviewing its [terms of service](#) and [privacy policies](#). You can always see or remove access in your [Google Account](#).

[Learn about the risks](#)


[Cancel](#) [Allow](#)



Sign in

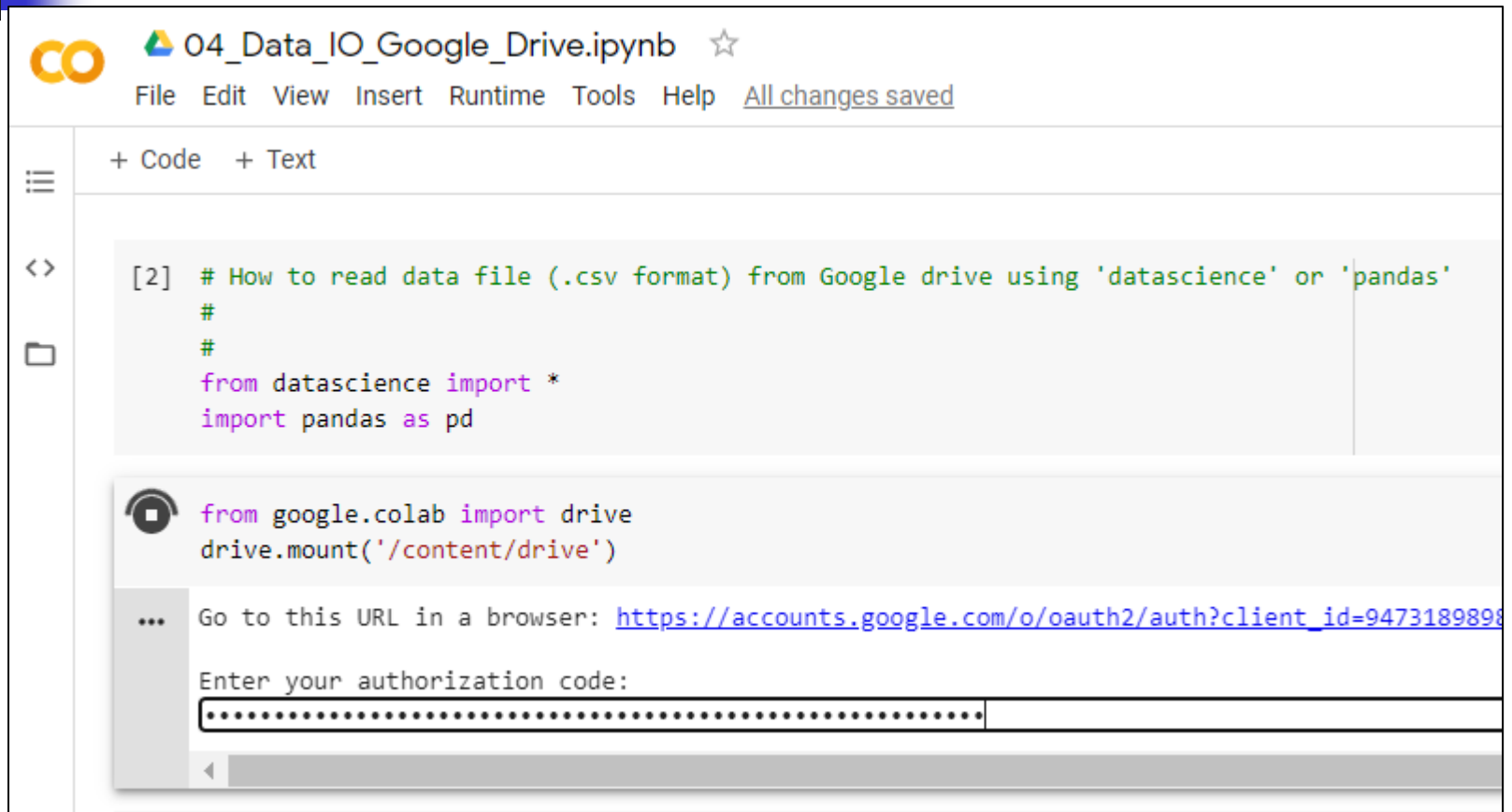
Please copy this code, switch to your application and paste it there:

**4/1wEiqEXKF-  
GiA838x9PVFmdygyqYp6tebocAeekB23sw2i2x2Sxc8Qs**





# Paste the Authorization Code

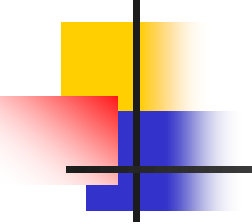


The image shows a Google Colab notebook titled "04\_Data\_IO\_Google\_Drive.ipynb". The notebook has a menu bar with "File", "Edit", "View", "Insert", "Runtime", "Tools", "Help", and "All changes saved". Below the menu bar, there are tabs for "+ Code" and "+ Text". The notebook contains two code cells. The first cell is a comment: "[2] # How to read data file (.csv format) from Google drive using 'datascience' or 'pandas'". The second cell contains the following code:

```
from datascience import *
import pandas as pd
```

Below the code cells, there is a prompt to enter an authorization code. The prompt says: "Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=9473189898](https://accounts.google.com/o/oauth2/auth?client_id=9473189898)". Below the URL, there is a text input field with the label "Enter your authorization code:". The input field is currently empty.

# Read File From Google Drive



04\_Data\_IO\_Google\_Drive.ipynb ☆

File Edit View Insert Runtime Tools Help

+ Code + Text

```
[3] from google.colab import drive
drive.mount('/content/drive')
```

Go to this URL in a browser: [https://accounts.google.com/o/oauth2/auth?client\\_id=947318989803-6bn6qk8qdgf4r](https://accounts.google.com/o/oauth2/auth?client_id=947318989803-6bn6qk8qdgf4r)

Enter your authorization code:  
.....  
Mounted at /content/drive

```
patients = Table.read_table('/content/drive/My Drive/TestGoogleDrive/breast-cancer.csv').drop('ID')
patients.show(3)
```

Clump Thickness	Uniformity of Cell Size	Uniformity of Cell Shape	Marginal Adhesion	Single Epithelial C
5	1	1	1	
5	4	4	5	
3	1	1	1	

... (680 rows omitted)

```
[ ] data = pd.read_csv('/content/drive/My Drive/TestGoogleDrive/breast-cancer.csv')
data.head
```

ID	Clump Thickness	...	Mitoses	Class
0 1000025	5	...	1	0
1 1002945	5	...	1	0
2 1015425	3	...	1	0
3 1016277	6	...	1	0
4 1017023	4	...	1	0
...	...	...	...	...
678 776715	3	...	1	0
679 841769	2	...	1	0



# Summary

---

- Colab
  - What is Colab?
  - Cost of Colab
  - Benefits and Drawbacks of Colab
  - How to Use Colab
  - Data IO in Colab