Msila University

Bioinformatics

Welcome

Dr. Khodja

What is Bioinformatics

- Bioinformatics = Bio + Informatics
- Informatics = Information + Automation
  - **Informatics** applies the principles of information science to solve problems using data
  - Informatics describe the science of automating information interactions
- Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data
- Bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret the biological data.

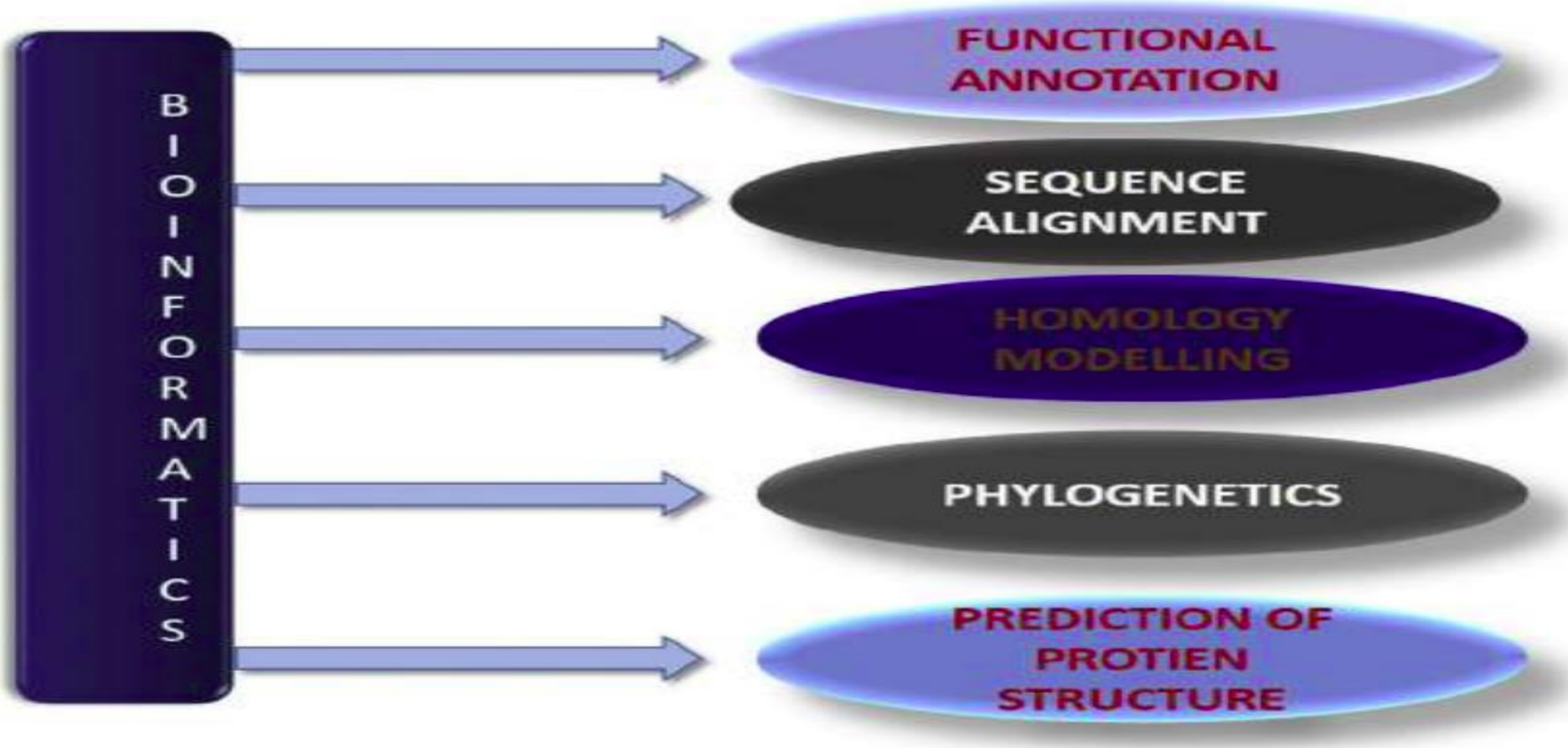


FIGURE 1.1

Bioinformatics— A crosstalk between different Sciences.

Common Activities in Bioinformatics

- **Genomics** refers to the sequencing and analysis of genomic entities, including genes and transcripts, in an organism.
  - A genome is the complete set of DNA sequences that codes for the hereditary material that is passed on from generation to generation.
  - These DNA sequences include all of the genes (the functional and physical unit of heredity passed from parent to offspring) and transcripts (the RNA copies that are the initial step in decoding the genetic information) included within the genome.
- **Proteomics** refers to the analysis of the complete set of proteins or proteome.

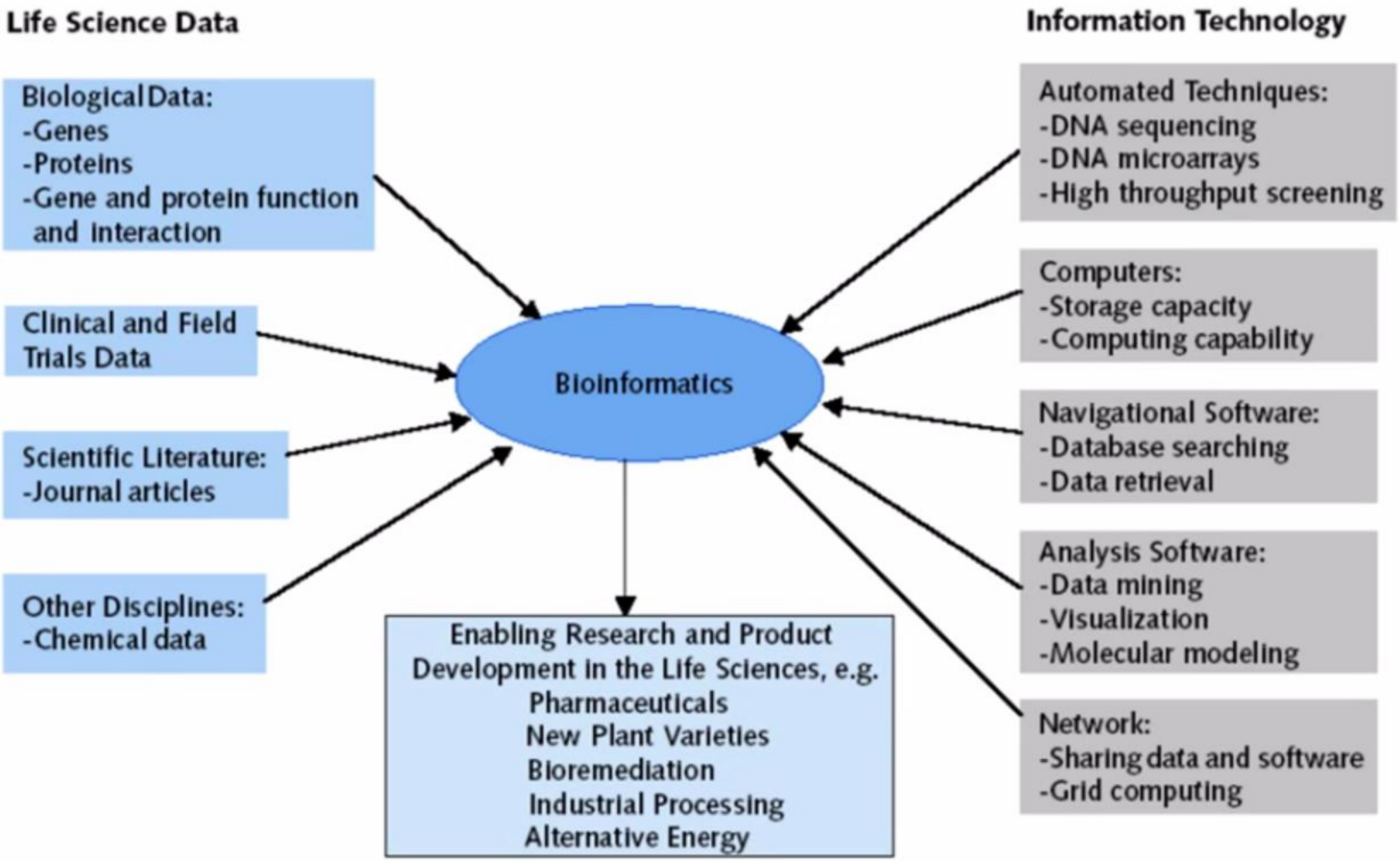
Common Activities in Bioinformatics

- **Metabolomics** is the large-scale study of small molecules , commonly known as metabolites, within cells, biofluids, tissues or organisms.
- **Transcriptomics** is the study of the transcriptome.
  - A transcriptome is the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using high-throughput methods, such as microarray analysis

Usefulness of Bioinformatics

- To understand the organizational principles within nucleic acid and protein sequences, called proteomics
- Used in genomics and proteomics
- For the identification of candidates genes and single nucleotide polymorphisms (SNPs).
- Useful for understanding the genetic basis of disease, unique adaptations, desirable properties (esp. in agricultural species), or differences between populations.
- Bioinformatics provides the tools for applying scientific method to large-scale data

– Bioinformatics Discipline



Life Science Data

Biological Data:
-Genes
-Proteins
-Gene and protein function and interaction

Clinical and Field Trials Data

Scientific Literature:
-Journal articles

Other Disciplines:
-Chemical data

Bioinformatics

Enabling Research and Product Development in the Life Sciences, e.g.
Pharmaceuticals
New Plant Varieties
Bioremediation
Industrial Processing
Alternative Energy

Information Technology

Automated Techniques:
-DNA sequencing
-DNA microarrays
-High throughput screening

Computers:
-Storage capacity
-Computing capability

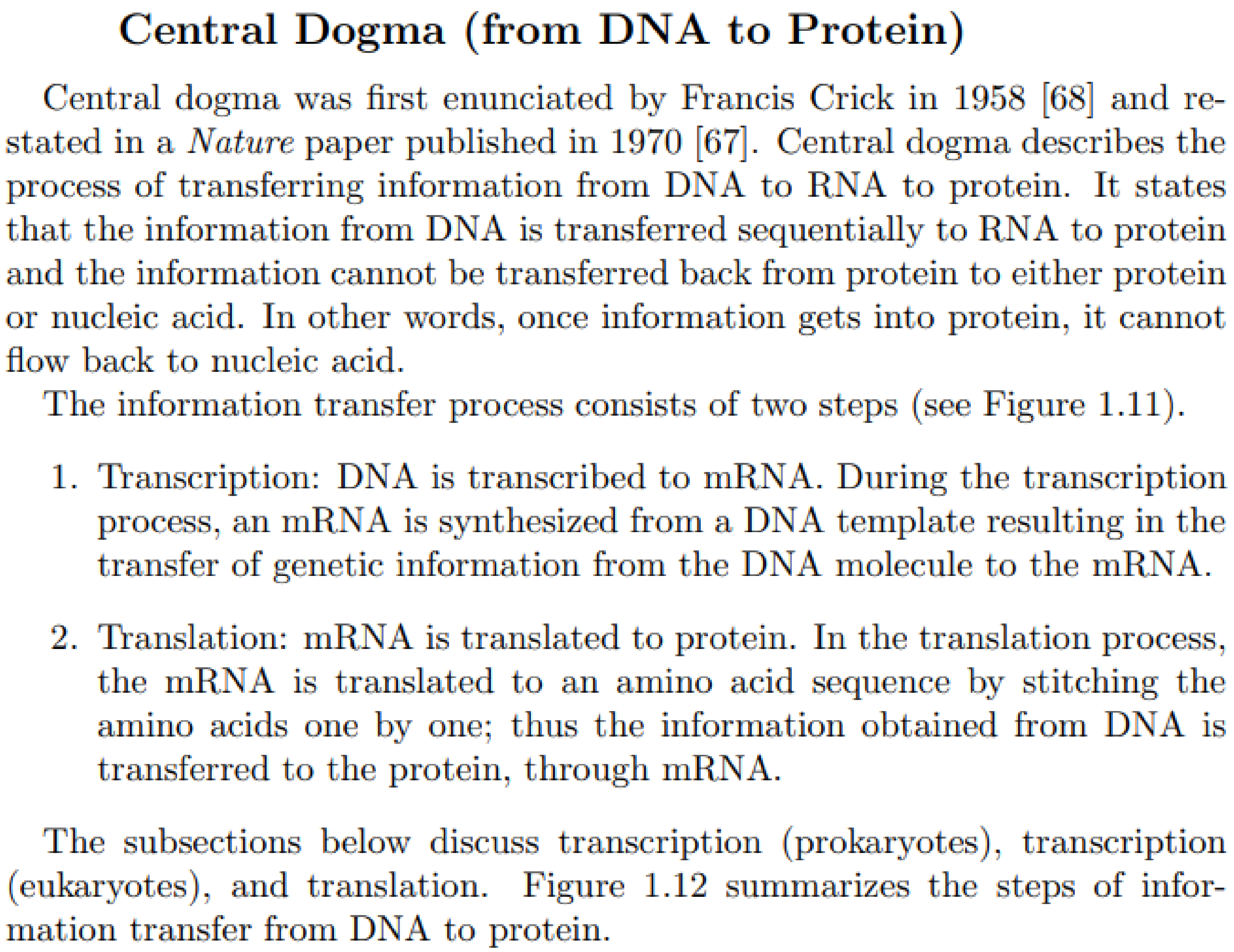
Navigational Software:
-Database searching
-Data retrieval

Analysis Software:
-Data mining
-Visualization
-Molecular modeling

Network:
-Sharing data and software
-Grid computing

- Since the word Bioinformatics was coined in 1991, more and more topics have been integrated into the loosely definition of Bioinformatics. Just like the word biotechnology, Bioinformatics encompasses disciplines that may or may not be directly related to Bioinformatics itself.

Central Dogma & Protein Structures



Central Dogma (from DNA to Protein)

Central dogma was first enunciated by Francis Crick in 1958 [68] and restated in a *Nature* paper published in 1970 [67]. Central dogma describes the process of transferring information from DNA to RNA to protein. It states that the information from DNA is transferred sequentially to RNA to protein and the information cannot be transferred back from protein to either protein or nucleic acid. In other words, once information gets into protein, it cannot flow back to nucleic acid.

The information transfer process consists of two steps (see Figure 1.11).

1. Transcription: DNA is transcribed to mRNA. During the transcription process, an mRNA is synthesized from a DNA template resulting in the transfer of genetic information from the DNA molecule to the mRNA.

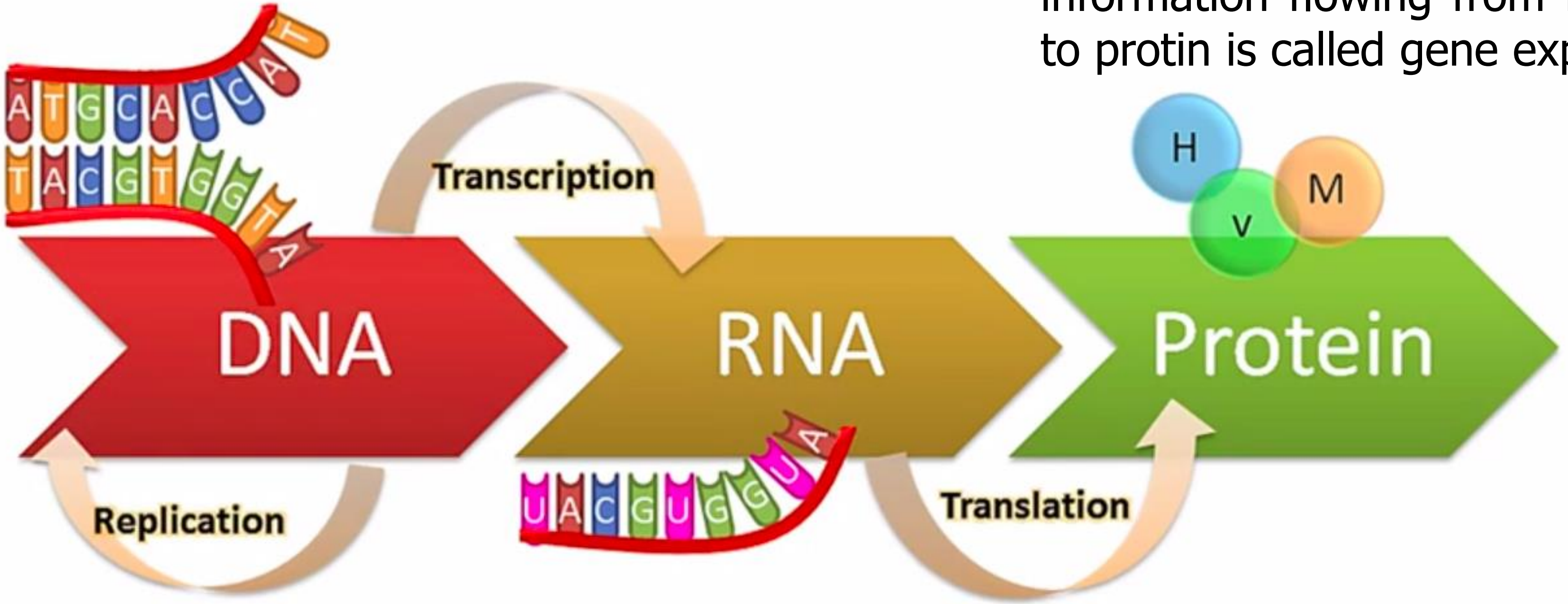
2. Translation: mRNA is translated to protein. In the translation process, the mRNA is translated to an amino acid sequence by stitching the amino acids one by one; thus the information obtained from DNA is transferred to the protein, through mRNA.

The subsections below discuss transcription (prokaryotes), transcription (eukaryotes), and translation. Figure 1.12 summarizes the steps of information transfer from DNA to protein.

– Central Dogma

Let's begin the central dogma of biology describes as just that it provides the basic framework for how genetic information flows from a DNA sequence to a protein powder inside cells.

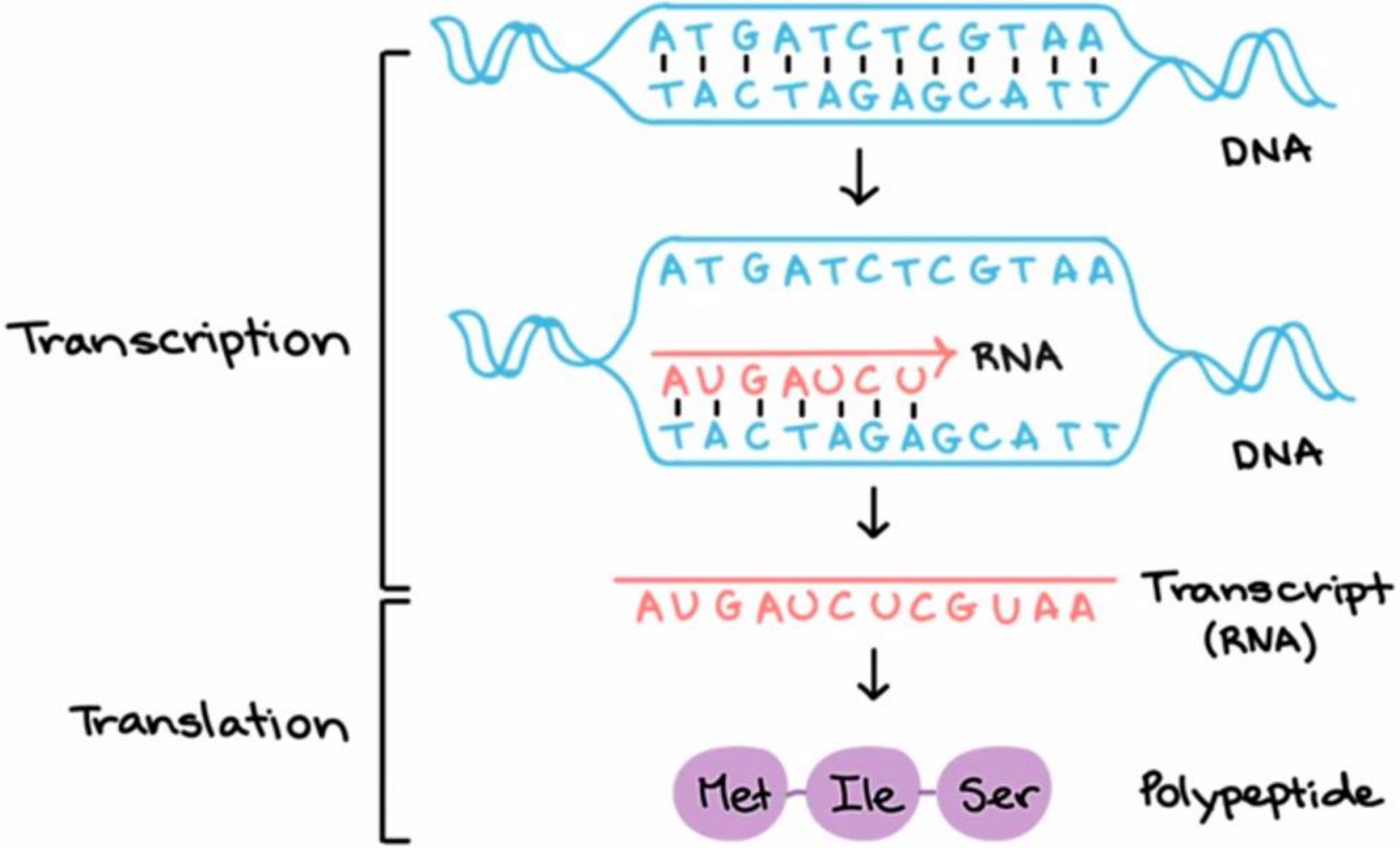
This process of genetics information flowing from DNA RNA to protin is called gene expression.



The central dogma of molecular biology deals with the detailed transfer of genetic information from DNA – RNA – Proteins in a sequential manner.

- The central dogma of biology describes just that.
- It provides the basic framework for how genetic information flows from a DNA sequence to a protein product inside cells.
- This process of genetics information flowing from DNA to RNA to protein is called gene expression.

– Central Dogma - Replicate



DNA sequence information is converted to RNA in a process called transcription, which is the first step of gene expression, where a segment of DNA is copied into RNA polymerase.

- First, the double-stranded DNA molecule is partially 'unzipped' and an enzyme called RNA polymerase literally copies the gene's nucleotides one by one into an RNA molecule.
- Like DNA, RNA is made of a particular sequence of nucleotides.
- Unlike DNA, RNA has only a single strand, and is a more fragile and temporary molecule inside the cell.
- Very importantly, RNA is small and can easily exit the nucleus and go to the cytoplasm, where proteins are made.
- The sequence encoded in the RNA molecule is decoded and converted into an amino acid sequence in a process called translation.

TRANSCRIPTION OF A DNA STRAND

Transcription of a DNA sequence

DNA = CGCGTAGCTAGCTGTAC

RNA polymerase

RNA = GCGCAUCGAUCGACAUG

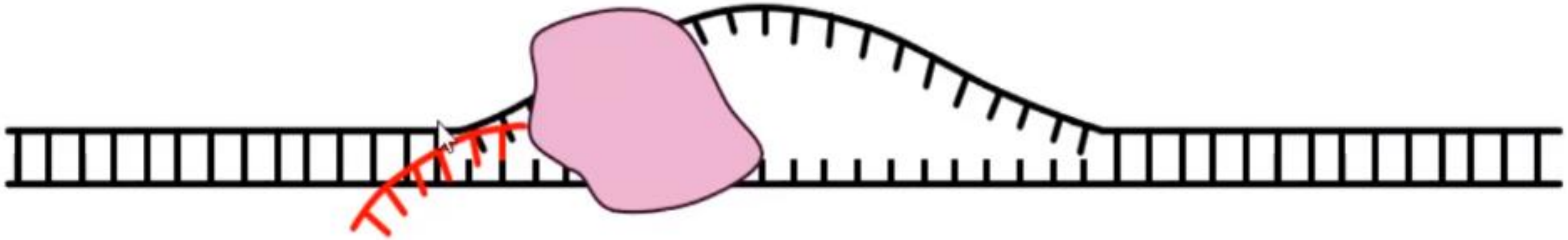
MAJOR DIFFERENCE

DNA			RNA	
A	→	Adenine ←	A	
C	→	Cytosine ←	C	
G	→	Guanine ←	G	
T	→	Thymine	U	→ Uracyl

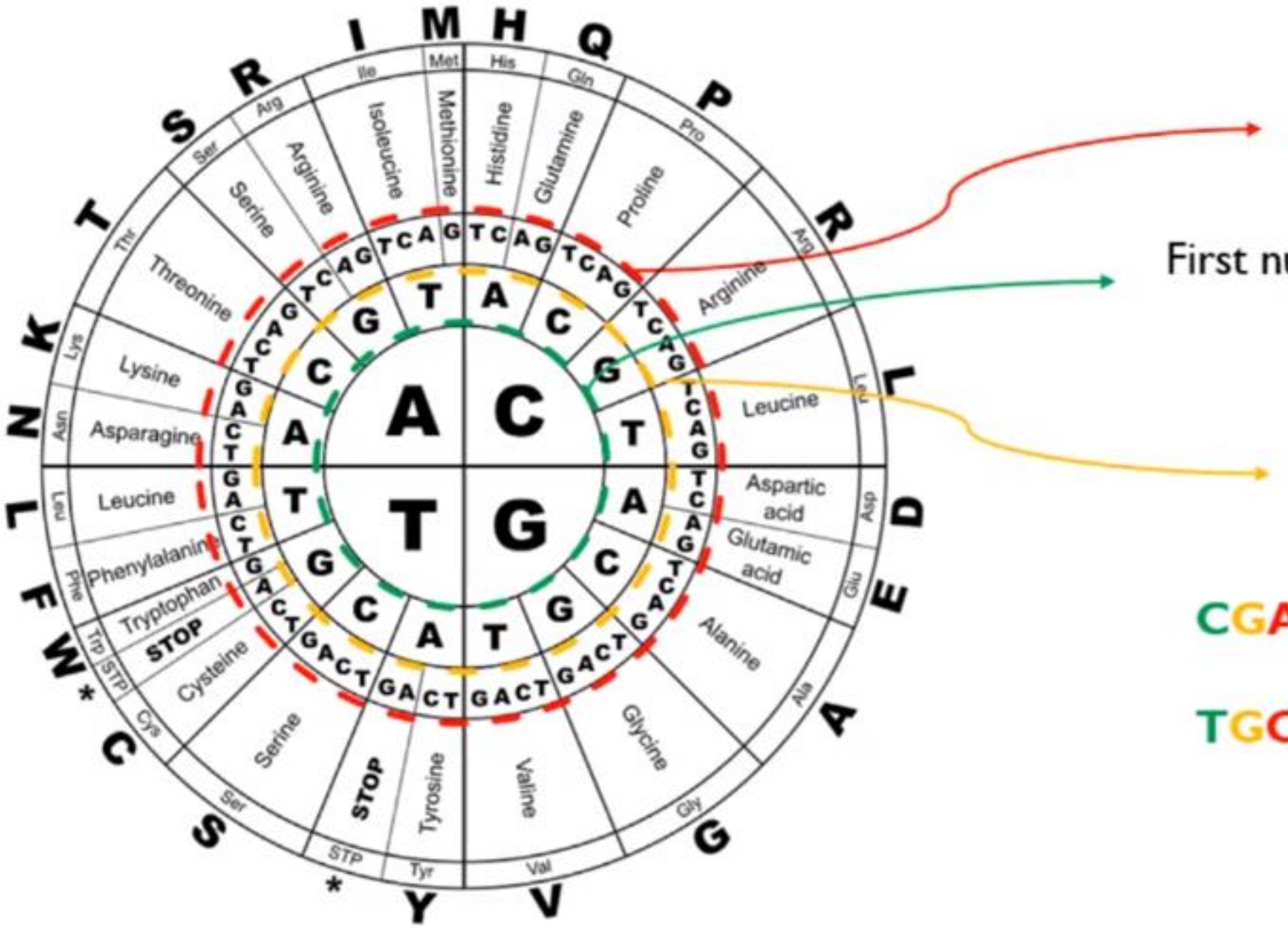
TRANSCRIPTION OF A DNA STRAND

Transcription of a DNA sequence

DNA
RNA
RNA polymerase



GENETIC CODES – FROM GENES TO PROTEIN



Third nucleotide

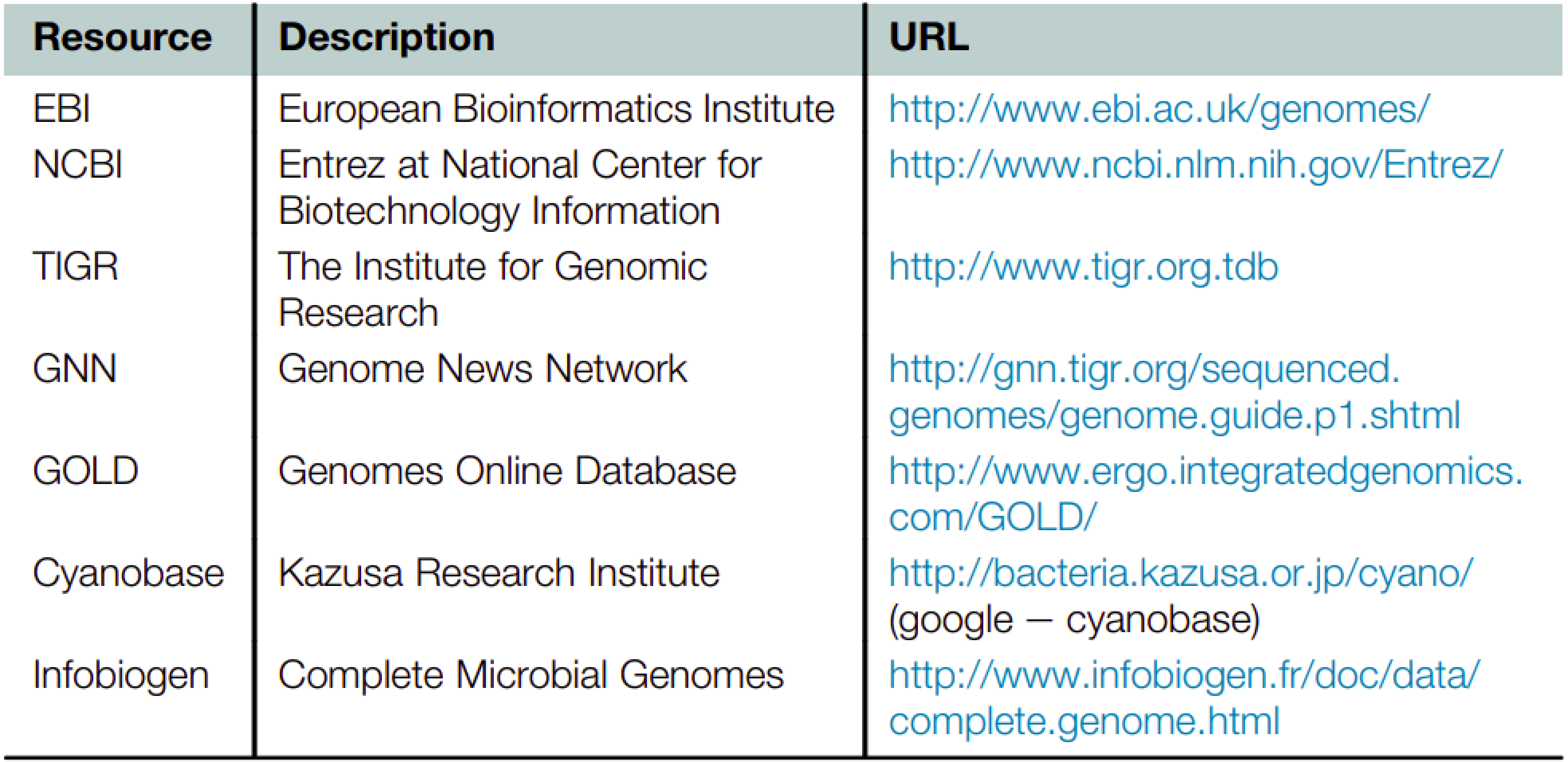
First nucleotide

Second nucleotide

CGA → Arginine

TGC → Cysteine

Table 2.1 Web-based resources for completed genome projects.

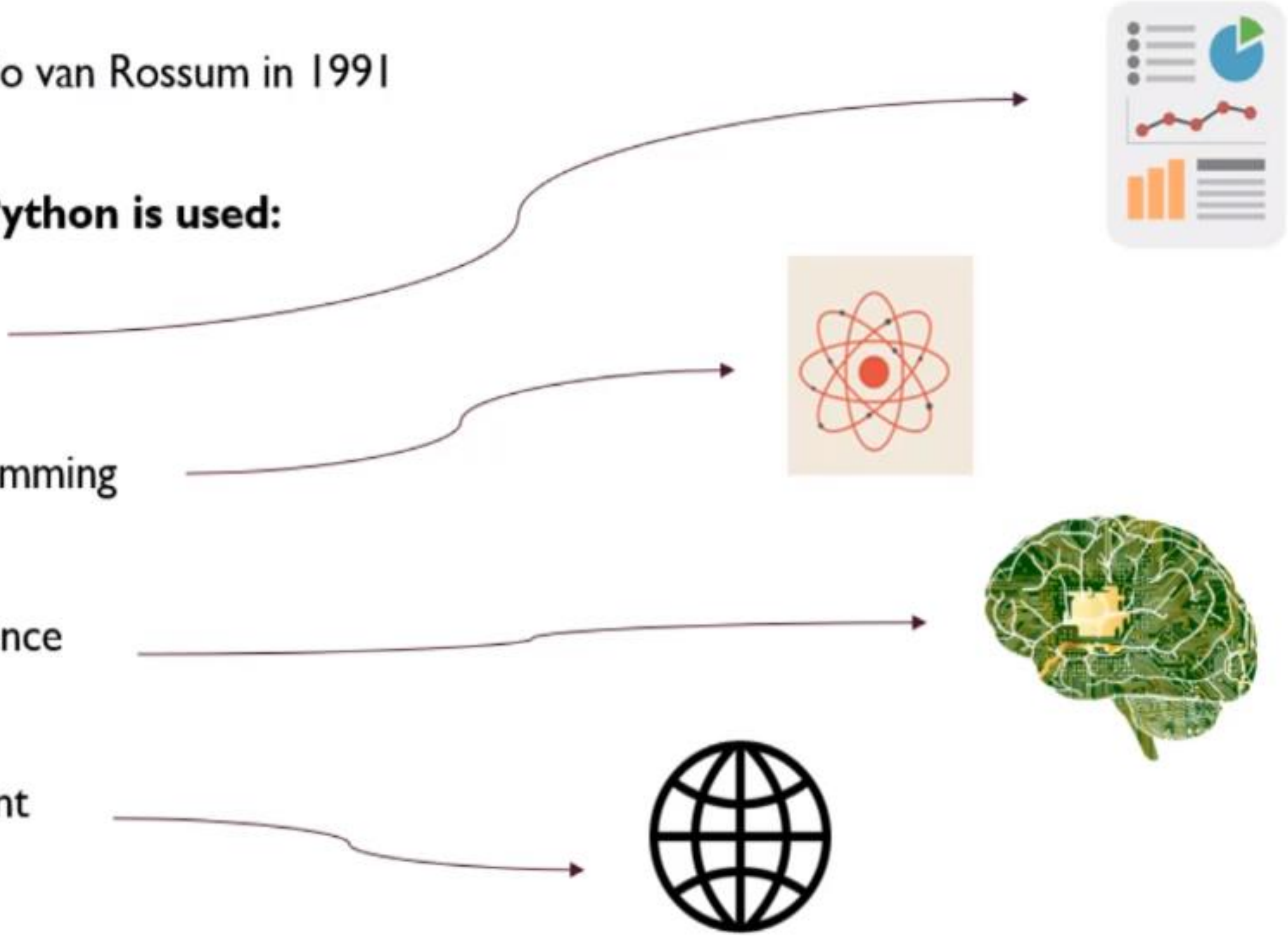


Resource	Description	URL
EBI	European Bioinformatics Institute	http://www.ebi.ac.uk/genomes/
NCBI	Entrez at National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/Entrez/
TIGR	The Institute for Genomic Research	http://www.tigr.org.tdb
GNN	Genome News Network	http://gnn.tigr.org/sequenced.genomes/genome.guide.p1.shtml
GOLD	Genomes Online Database	http://www.ergo.integratedgenomics.com/GOLD/
Cyanobase	Kazusa Research Institute	http://bacteria.kazusa.or.jp/cyano/ (google – cyanobase)
Infobiogen	Complete Microbial Genomes	http://www.infobiogen.fr/doc/data/complete.genome.html

WHAT IS PYTHON USED FOR?

Created by Guido van Rossum in 1991

Areas where Python is used:

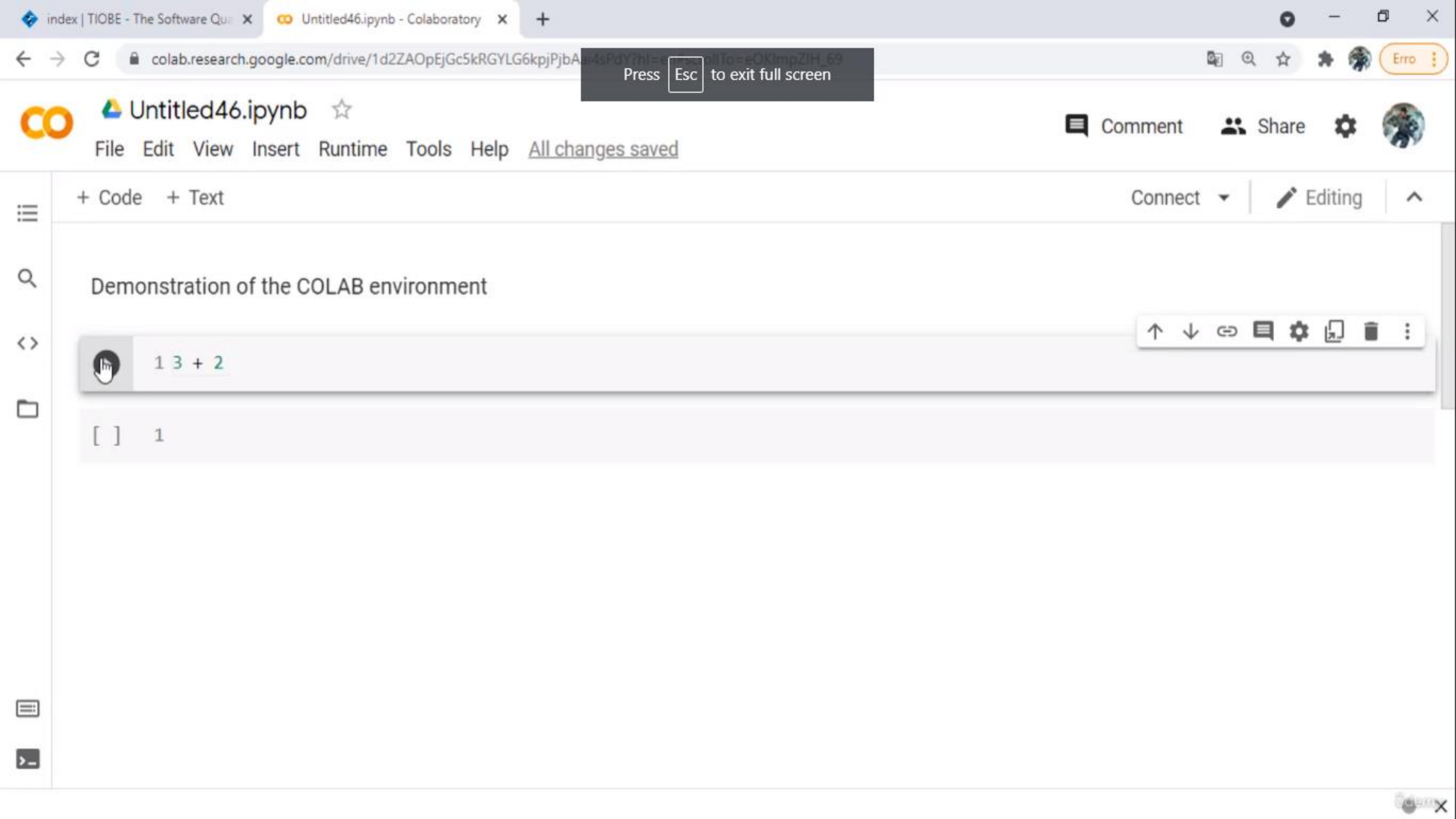


Data Science

Scientific programming

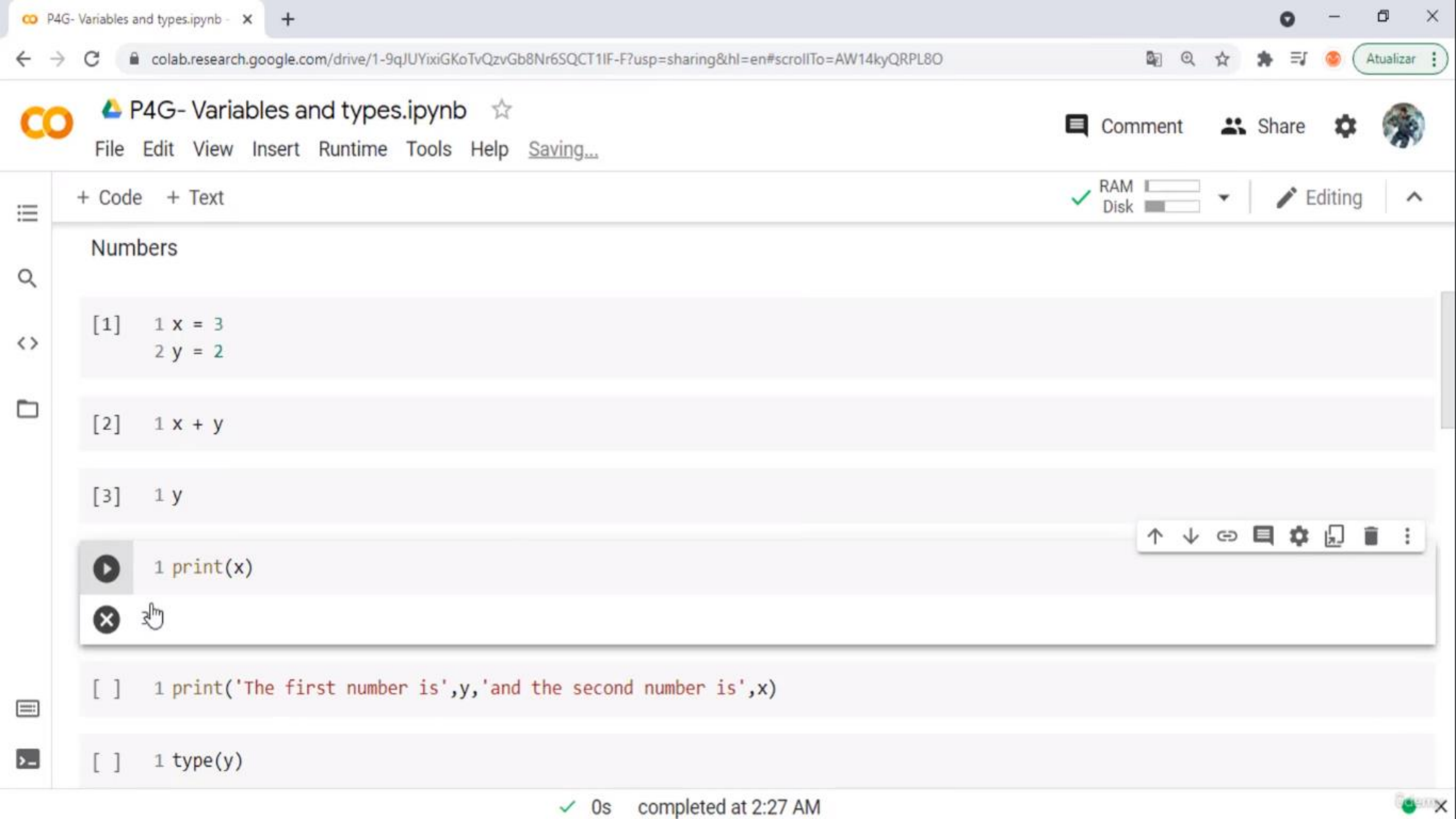
Artificial inteligence

Web development



Demonstration of the COLAB environment

```
3 + 2
```



Numbers

```
[1]  x = 3
     y = 2
```

```
[2]  x + y
```

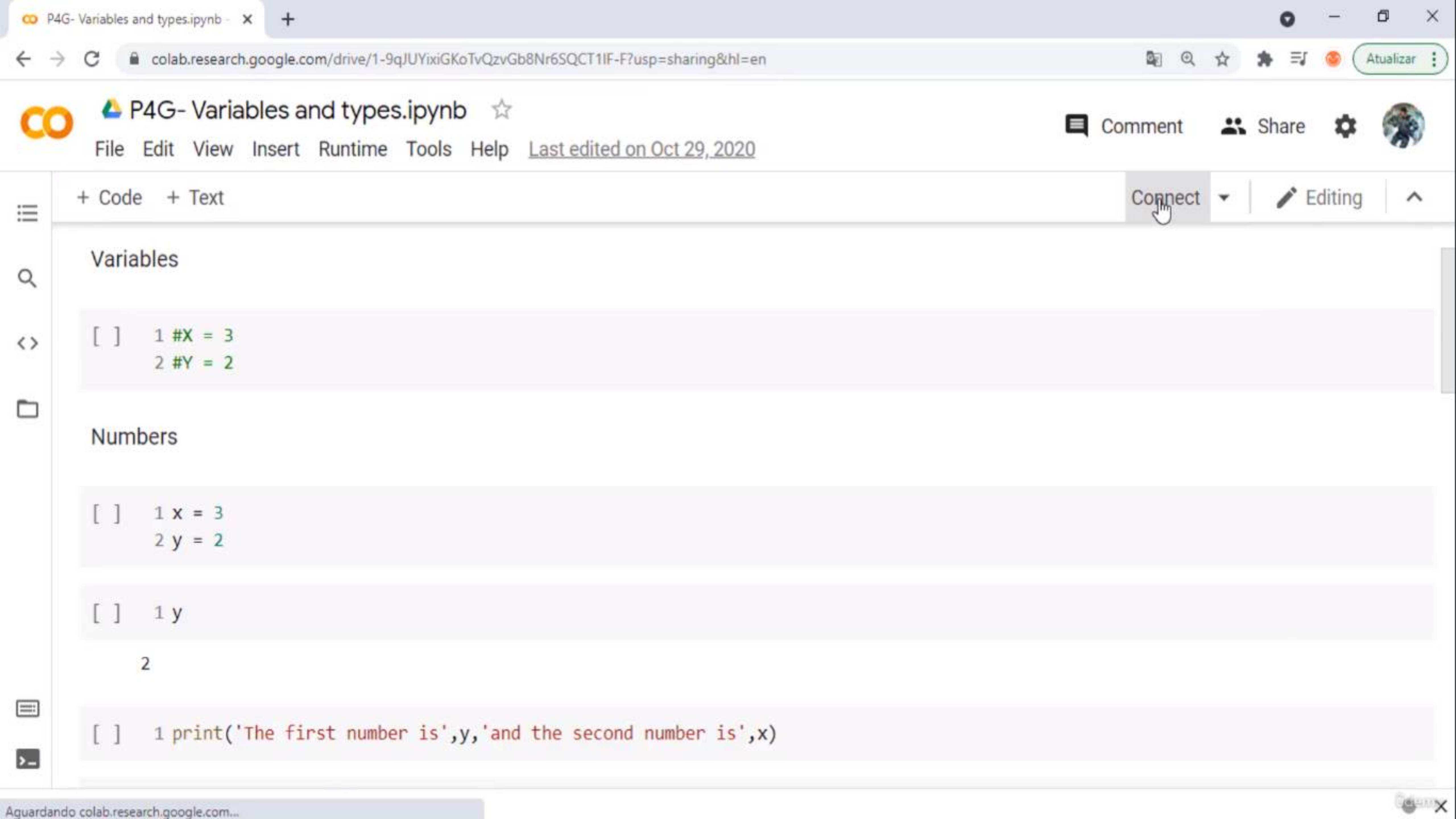
```
[3]  y
```

```
     print(x)
```

```
[ ]  print('The first number is',y,'and the second number is',x)
```

```
[ ]  type(y)
```

✓ 0s completed at 2:27 AM



Variables

```
#X = 3
#Y = 2
```

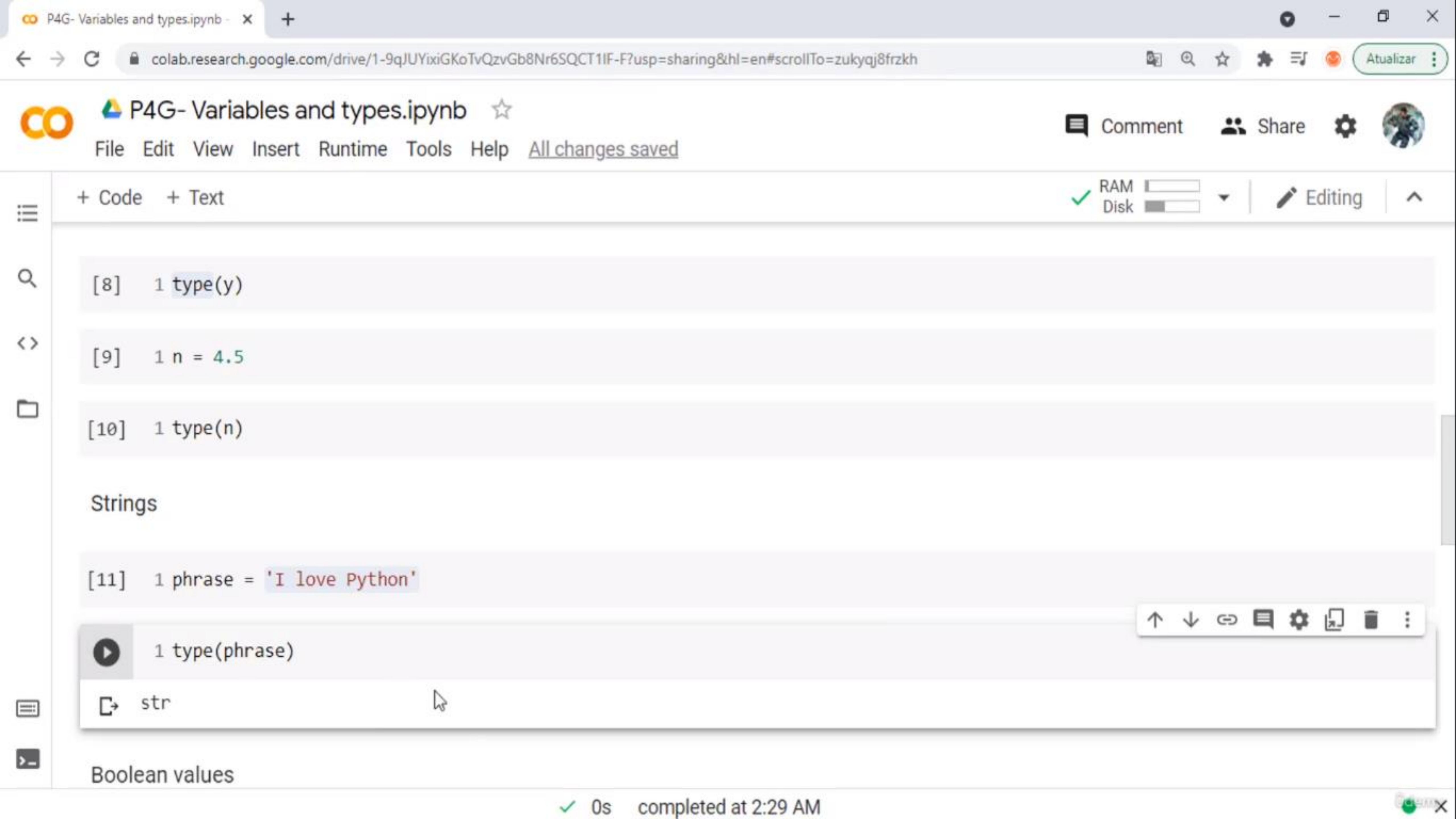
Numbers

```
x = 3
y = 2
```

```
y
```

```
2
```

```
print('The first number is',y,'and the second number is',x)
```



```
[8]  type(y)
```

```
[9]  n = 4.5
```

```
[10] type(n)
```

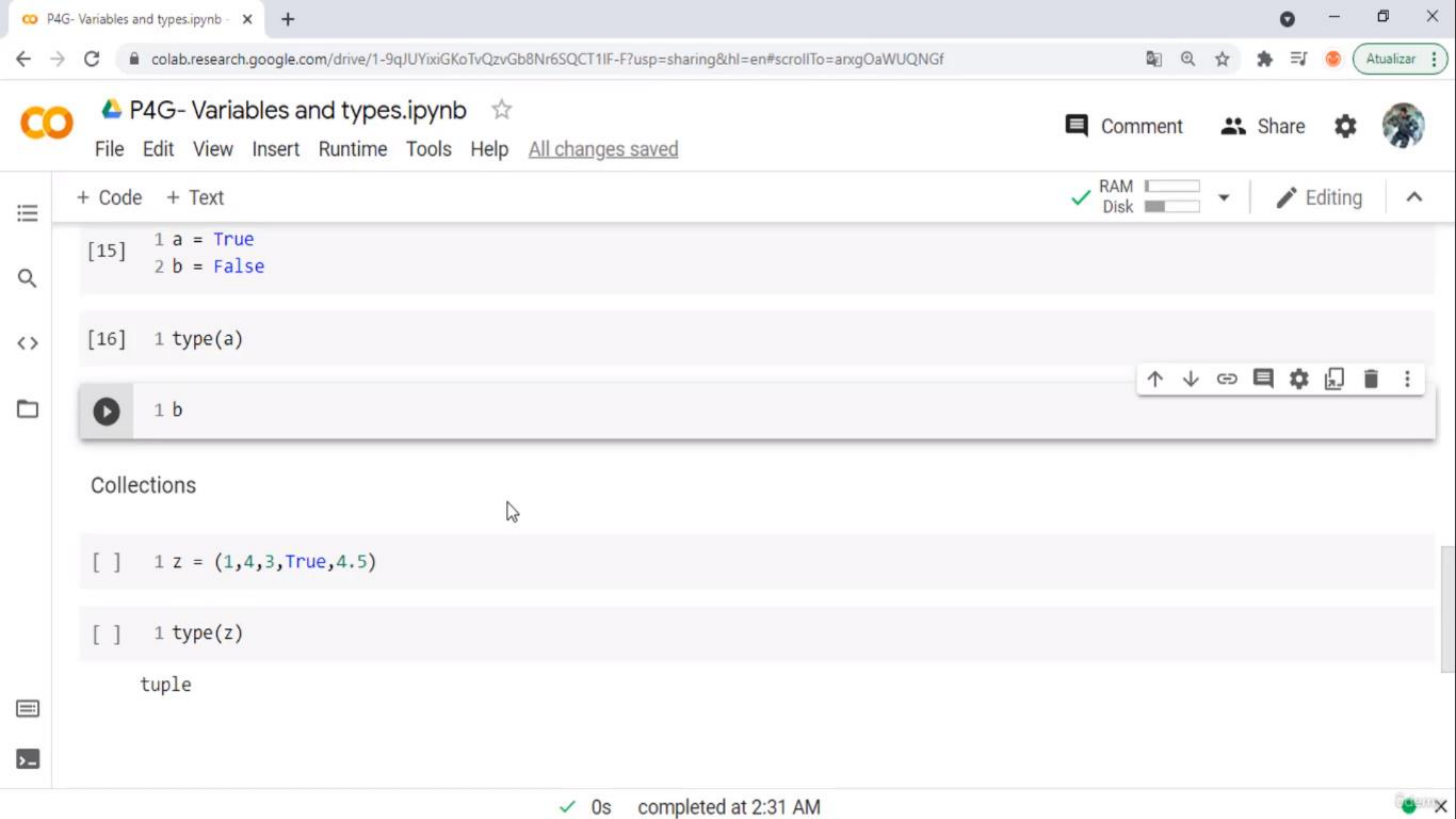
Strings

```
[11] phrase = 'I love Python'
```

```
type(phrase)
```

```
str
```

Boolean values



[15]

```
a = True
b = False
```

[16]

```
type(a)
```

```
b
```

Collections

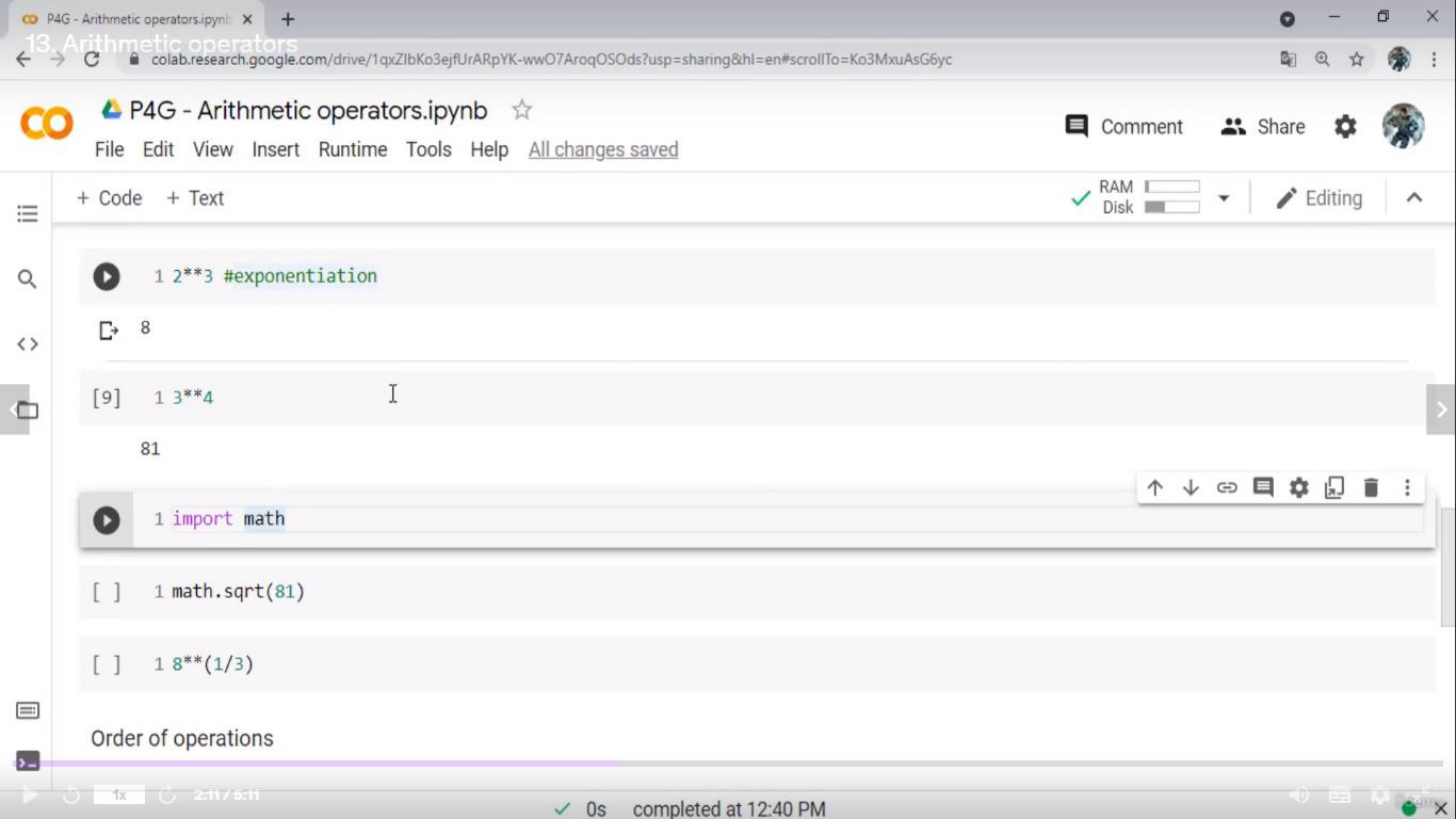
[ ]

```
z = (1,4,3,True,4.5)
```

[ ]

```
type(z)
```

```
tuple
```



```
2**3 #exponentiation
```

```
8
```

[9]

```
3**4
```

```
81
```

```
import math
```

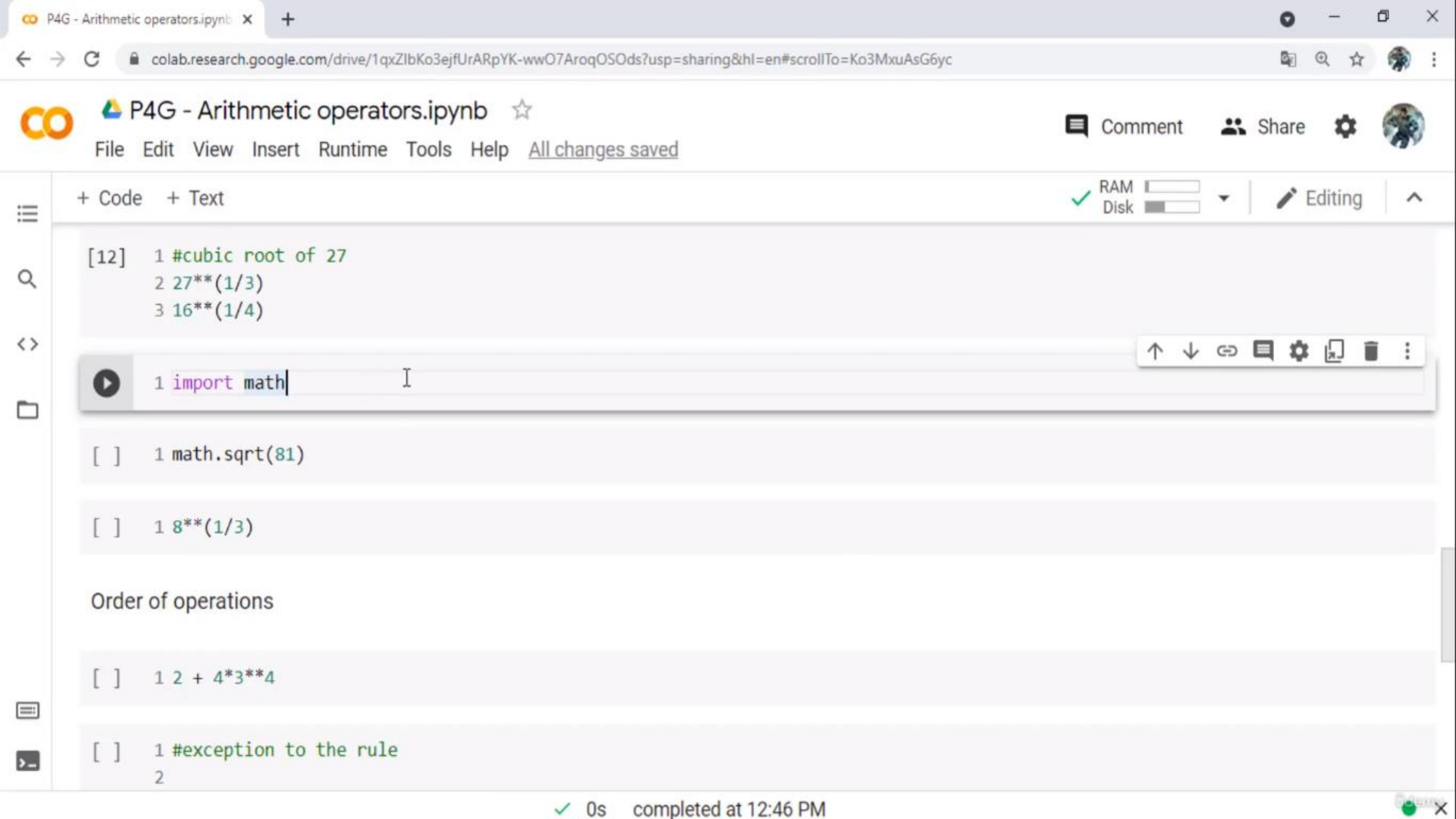
[ ]

```
math.sqrt(81)
```

[ ]

```
8**(1/3)
```

Order of operations



[12]

```
#cubic root of 27
27**(1/3)
16**(1/4)
```

```
import math
```

[ ]

```
math.sqrt(81)
```

[ ]

```
8**(1/3)
```

Order of operations

[ ]

```
2 + 4*3**4
```

[ ]

```
#exception to the rule

```

✓ 0s completed at 12:46 PM