



GCE SOLUTIONS

Derive Value from Excellence...

R vs. Python, Techniques and Challenges for the SAS Programmer

By Murali Neela

PhUSE US Connect, Baltimore, MD, February 24, 2019



Table of Contents

- How to use R
 - IDE
 - Downloading
 - Functions
- How to use python
- R vs. python
- R vs. SAS
- python vs. SAS
- Conclusion





Advantages of using R

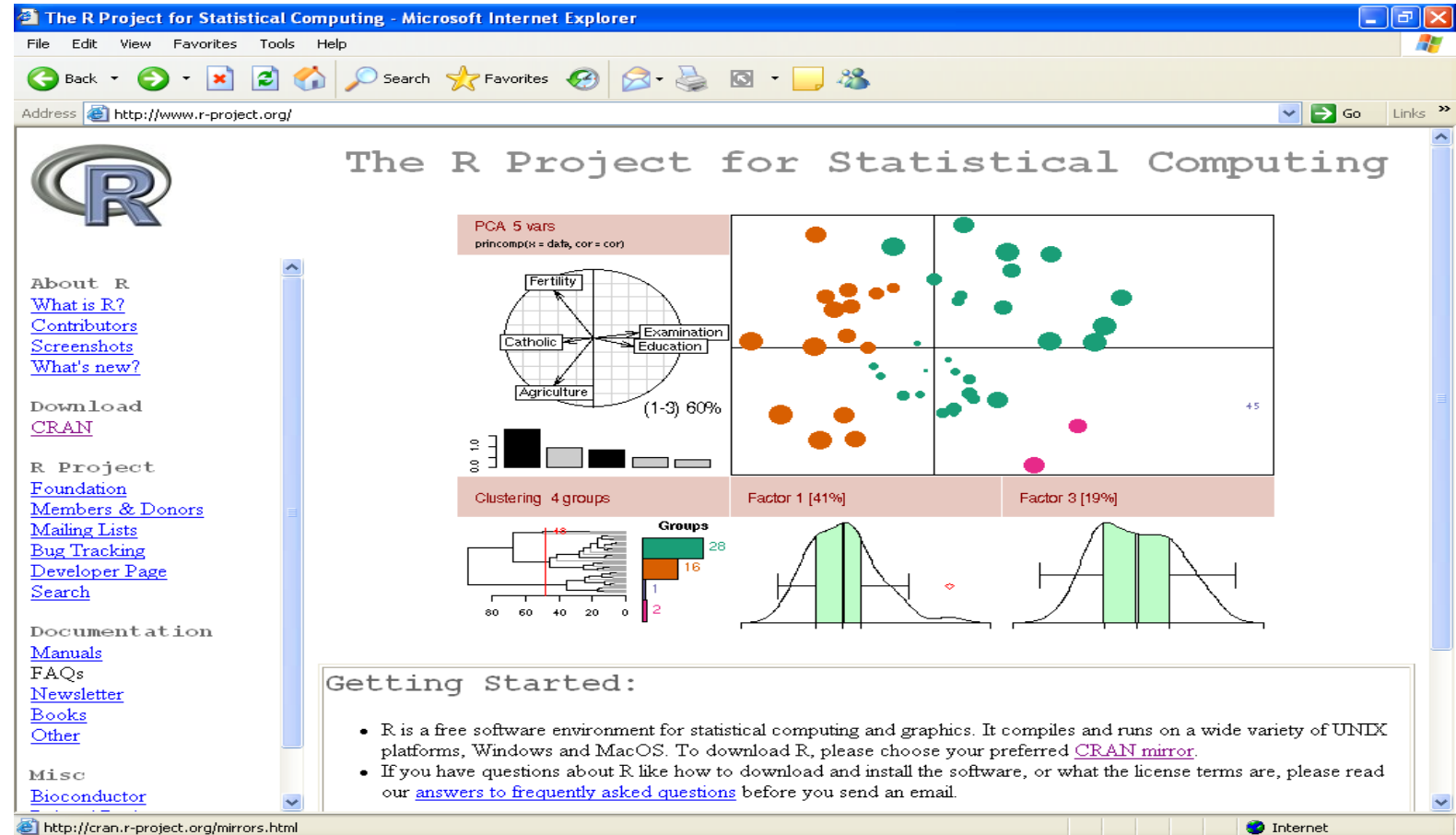
- R is free open source software for statistical computing and graphics
- R consists of core software and enhanced by using software packages.
- Large catalog for data analysis
- GitHub interface
- Offers great flexibility for analysis
- R makes it is easy to think while doing your analysis
- Exceptional data visualization tools





R Downloading – 1

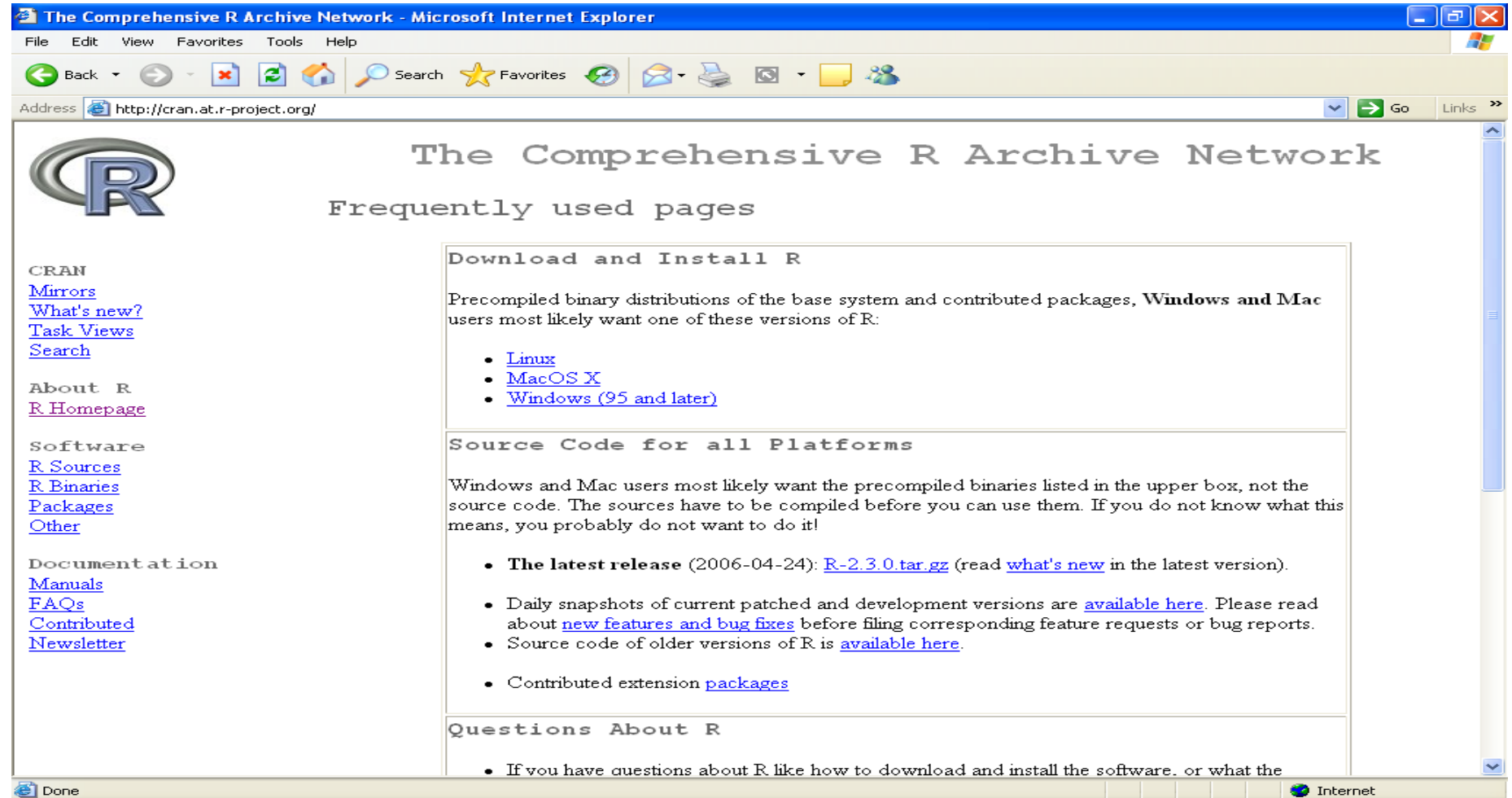
<https://cran.r-project.org/bin/windows/base/>





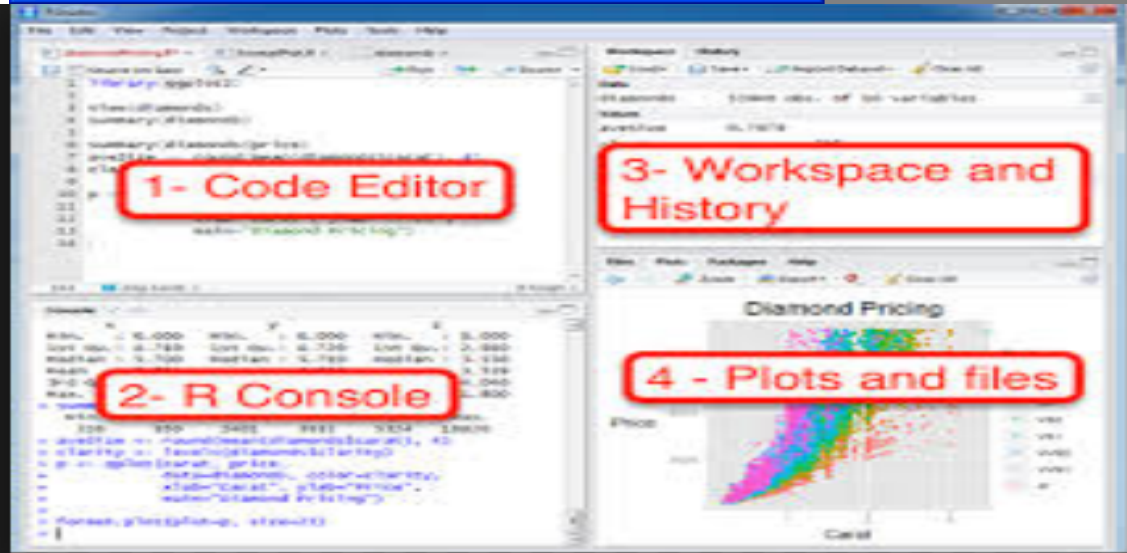
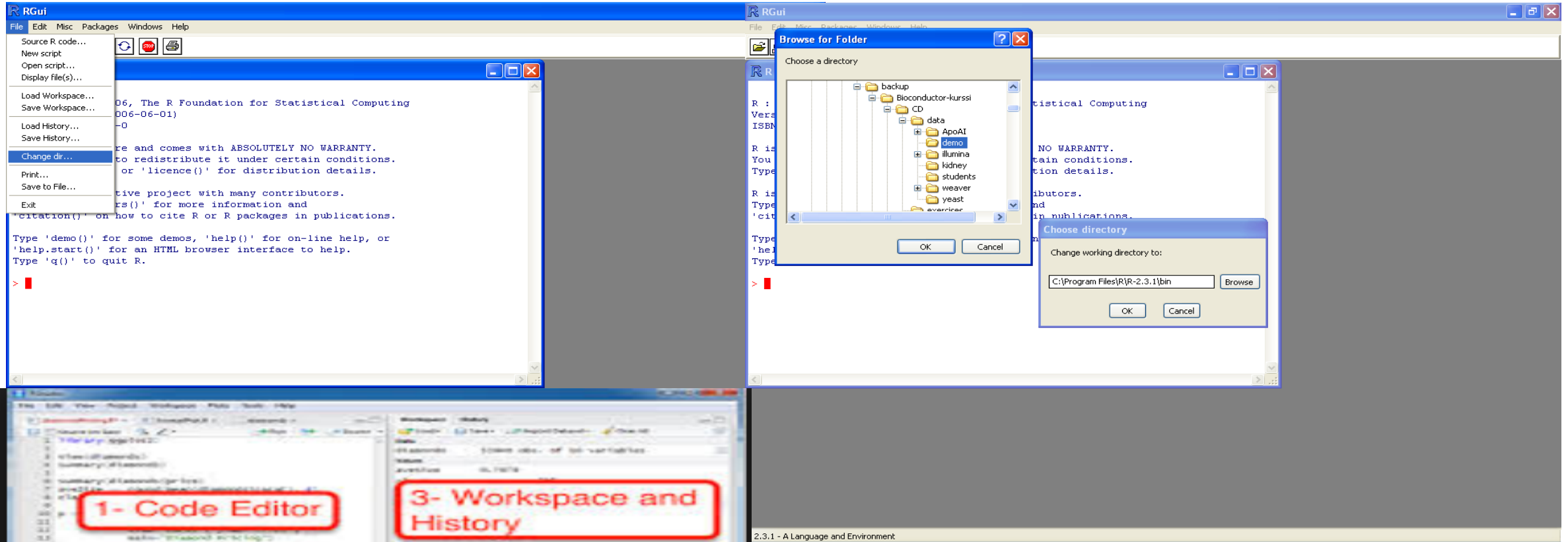
R Downloading - 2

<https://cran.r-project.org/>





R Environment





R as a calculator

- `log2(32)`

- `[1] 5`

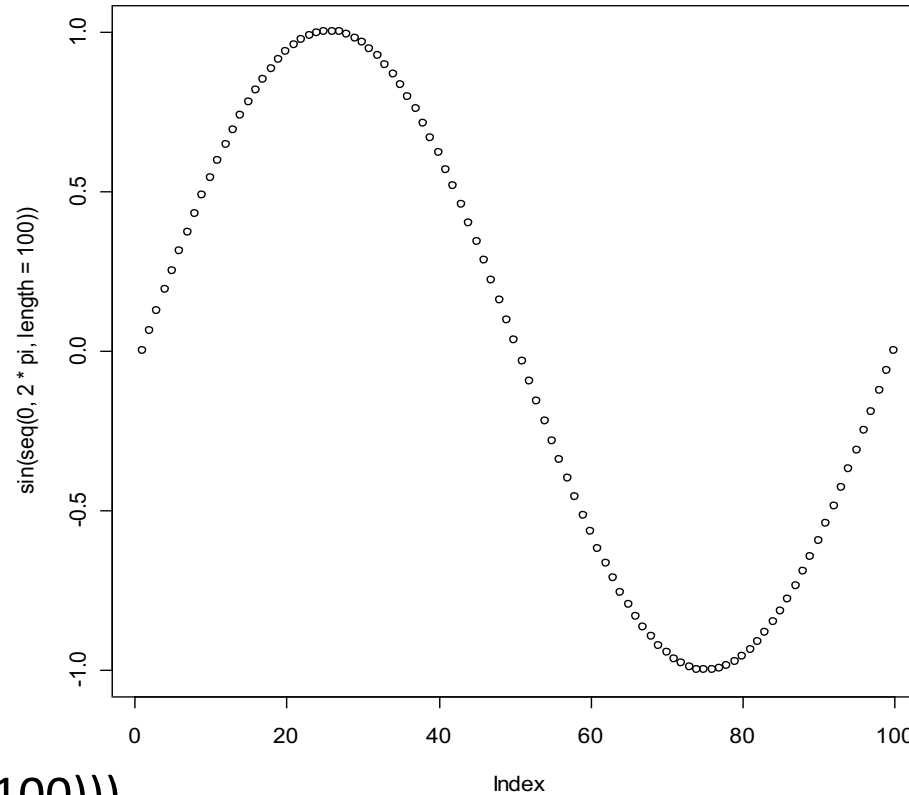
- `sqrt(2)`

- `[1] 1.414214`

- `seq(0, 5, length=6)`

- `[1] 0 1 2 3 4 5`

- `plot(sin(seq(0, 2*pi, length=100)))`





R Statistical Functions

- Descriptive Statistics
- Statistical Modeling
Regressions: Linear and Logistic
Probit, Tobit Models
Time Series
- Multivariate Functions
- Inbuilt Packages, contributed packages





R Descriptive Statistics

- Has functions for all common statistics
- `Summary()` gives lowest, mean, median, first, third quartiles, highest for numeric variables
- `Stem()` gives stem-leaf plots
- `Table()` gives tabulation of categorical variables





R Synopsis of Operators

Operator	Usually means	In Formula means
+ or -	add or subtract	add or remove terms
*	multiplication	main effect and interactions
/	division	main effect and nesting
:	sequence	interaction only
^	exponentiation	limiting interaction depths
%in%	no specific	nesting only





python Introduction

- python was developed by Guido van Rossum
- Open source general-purpose language.
- python for the purpose of doing mathematical calculations
- Use python for data preparation, data munging especially for unstructured data like web, images, text etc.
- Great flexibility and ability to extract information from free text, websites, and social media sites
- Good with mining images and prepare data for analysis





How to Use Python

- **code** or **source code**: The sequence of instructions in a program.
- **syntax**: The set of legal structures and commands that can be used in a particular programming language.
- **output**: The messages printed to the user by a program.
- **console**: The text box onto which output is printed.
 - Some source code editors pop up the console as an external window, and others contain their own console window.

The screenshot shows two windows from the Python 2.5.1 environment. The main window is titled 'Python Shell' and contains the following text:

```
Python 2.5.1 (r251:54863, Apr 18 2006)
win32
Type "copyright", "credits" or "license()" for more

>>>
*****
Personal firewall software may
makes to its subprocess using
interface. This connection is
interface and no data is sent
*****

IDLE 1.2.1
>>> ===== RESTART =====
>>>
Hello, world!
4
>>>
```

Overlaid on top of the shell is a smaller window titled 'example.py - C:/Docum...'. It contains the following Python code:

```
File Edit Format Run Options
Windows Help

print "Hello, world!"
print 2 + 2

Ln: 2 Col: 11
```

The screenshot shows a Windows command prompt window titled 'C:\WINDOWS\system32\cmd.exe'. It displays the output of the Python program:

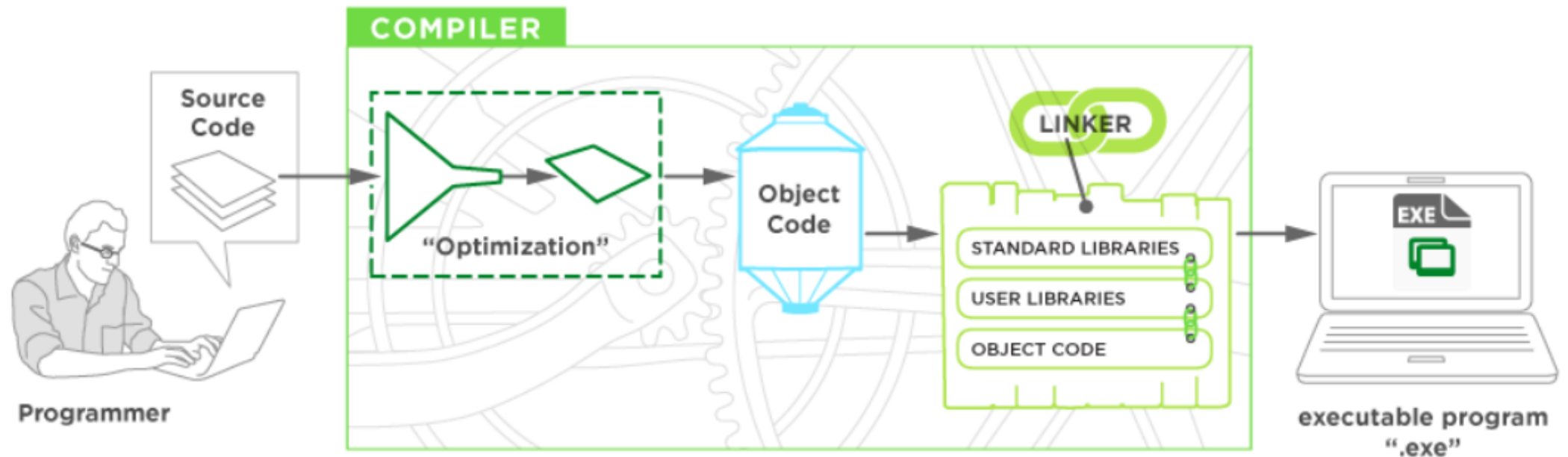
```
Hello, world!
Press any key to continue
```





R is Compiled

How does code written by a programmer become machine-readable code a computer can understand? Compilers optimize source code, create intermediary object code, then link that code to make it readable by a computer's CPU. The result is an executable file a CPU can easily process and run.

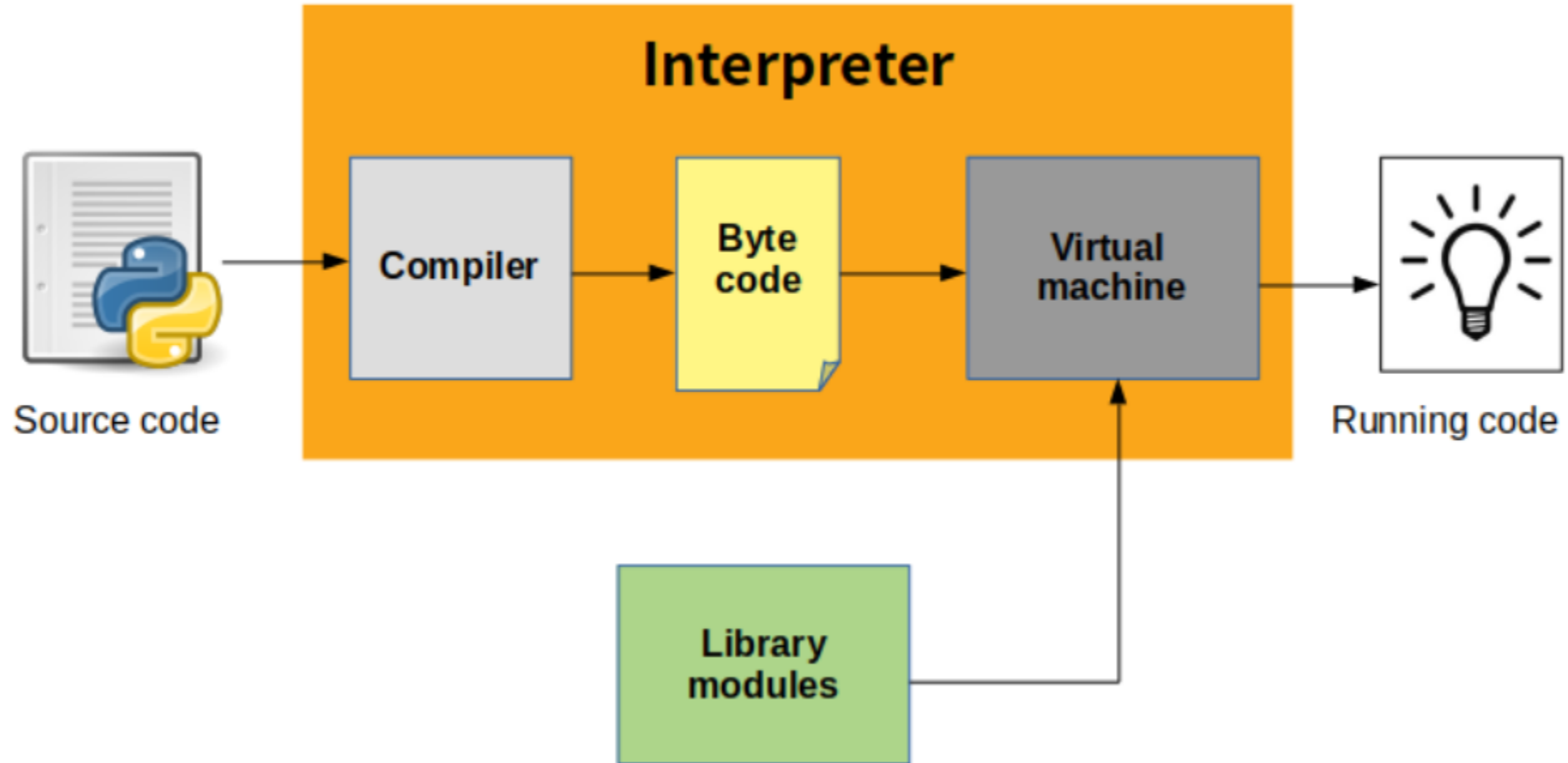


Compiled languages are converted into machine-readable code **prior to execution**, compared with interpreted languages, which are compiled to machine code **at the time of execution**.





python is Interpreted





Compiling vs. Interpreting

- A **compiled language** is a programming **language** whose implementations are typically compilers (translators that generate machine code from source code),
- Interpreters (step-by-step executors of source code, where no pre-runtime translation takes place).
- Compiled languages are executed once and used many times. Interpreters are always executed every time they are used.





R vs. python

R	python
Analysis, data visualization, and modeling	Data preparation, data munging especially for unstructured data like web, images, text etc
Objective is Data analysis and statistics	Deployment and production
Primary Users are Scholar and R&D	Primary Users are Programmers and developers.
Easy to use available library	Easy to construct new models from scratch.
Exceptional data visualization tools	Can handle large volume of data better than R



R vs. SAS

R	SAS
Environment contains only one programming language	In the SAS System multiple “sub-languages” are used (SAS/BASE with its data step, Macro language, IML etc.).
An open source counterpart and can be easily downloaded	SAS is a commercial tool and expensive
Debugging can be cumbersome as error messages are not easily comprehensible	Easy to debug code,(easily comprehensible error messages)
Incorporates newer technologies quicker as packages get added on by programmers all over the world	New statistical and machine learning techniques implemented only in new version rollouts
Advanced graphical capabilities. Supports various professional graphics templates	Graphical capabilities are limited as compared to R.



Python vs. SAS

Python	SAS
Interpreted	Environment is Display manager
IDE is Spyder(one of many)	Environment is Display manager
Series=array-like object with an index, Data frames=rows and columns with two indices	Dataset structures are SAS dataset
Easy to construct new models from scratch.	Programming structure is DATA and PROC steps
Single statements or function calls	Exceptional data visualization tools





Conclusion

- Why use SAS, R or Python?
 1. The SAS language is a computer programming language used for statistical analysis.
 2. R programming language is used by data scientists to extract or data mine information from a large data set or surveys.
 3. Python is a high-level, interpreted and general-purpose dynamic programming language that focuses on code readability.
- SAS is outdated or behind technology?
 1. SAS is not outdated by any means, Yes R is gaining way more popularity but no fortune 500 company can do away with SAS in a blink. And to counter the market by R
- R and Python are new technology?
 1. Computer Science. Python is the most common coding language I typically see required in data science roles, along with Java, Perl, or C/C++. Python a great programming language for data scientists.





Q & A
Thank you

Murali Neela

murali.neela@gcesolutions.com

101,1st Floor Abhi's Ganga Plot No 15 Shilpi Valley
Enclave Gafoornagar Madhapur Hyderabad, India



**GCE
SOLUTIONS**
Derive Value from Excellence...