AN INTRODUCTION TO R

MOHAMMAD NAYEEM HASAN

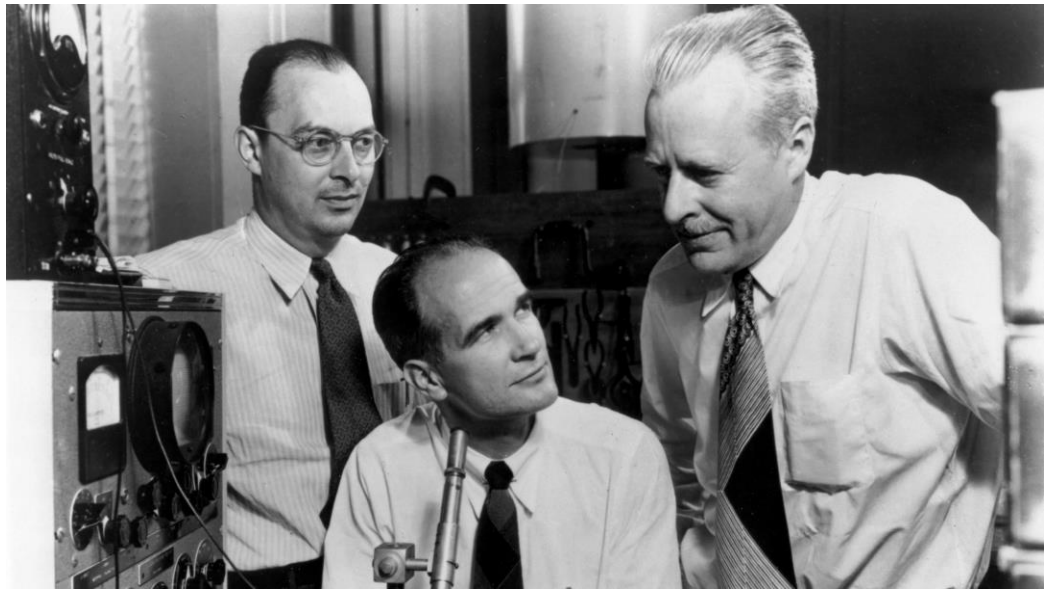
TABLE OF CONTENTS

HISTORY

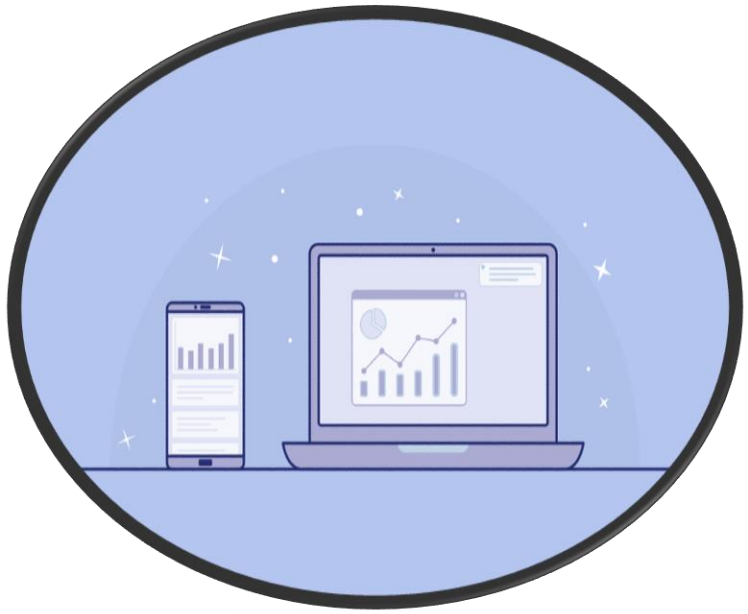
R is a programming language it was an implementation over S language. R was first designed by Ross Ihaka and Robert Gentleman at the University of Auckland in 1993.

It was stable and released on October 31st, 2014, 4 months ago, by R Development Core Team under GNU General Public License.

BELL LABS



Origin in the Bell Labs in the 1970s



INTRODUCTION

R

Is a programming language and software environment for statistical computing and graphics. R source code is written in C, Fortran, and R

THE R LANGUAGE

Is widely used among statisticians and data miners for developing statistical tools and data analysis.

IT COMPILES AND RUN

On a wide variety of UNIX platforms, Windows and Mac OS. R can be downloaded and installed from CRAN (Comprehensive R Archive Network) website.

THE COPYRIGHT

The primary source code for R is held by the R foundation and is published under the GNU General Public License version 2.0.

R BASICS

Why R?

- Free and open-source, Large community users, Latest cutting-edge technology, independent platform.
- The most extensive modeling resources in scientific research.
- The fine publishing quality graphs, have a robust visualization library.
- Easy to develop your own models.
- R is freely available under the GNU General Public License.
- Go to language for statistics and data science
- Used almost every industry and gateway to a lucrative career.

APPLICATIONS

THE APPLICATIONS OF R

- Statistical Computing
- Machine Learning
- Data Science

FACEBOOK AND TWITTER

- Facebook, for behavior analysis related to status updates and profile pictures.
- Twitter, for data visualization and semantic clustering.

GOOGLE AND MICROSOFT

- Google, for advertising effectiveness and economic forecasting.
- Microsoft, Acquired revolution R company, and use it for a variety of purposes.

The R language is cross-platform interoperable and fully portable which means the R program that you write on one platform can be carried out to other platforms and run there.

R PACKAGES

- A Package is a collection of R functions with comprehensive documents.
- A package includes R functions, Data examples, Help files, Namespace, and a description.
- The default installation is kept at a minimum.
- The function of R could be extended by loading R packages.

FEATURES

Open Source:

The source code of the R program and the extensions could be examined line by line

Integrating with other programming languages:

R is an interpreting language, that can be rather slow but could integrate with highly efficient languages such as C, C++, or Fortran.

OS Independence:

UNIX, Linux, Windows, MacOS, FreeBSD.

Fast calculation:

R can be used to perform complex mathematical and statistical calculations on data objects of a wide variety.

COMPARISON WITH OTHERS

Parameter	R Programming	Python
Objective	Data analysis and Statistical Modelling	Data Science, Web Development, etc.
Workability	Consists of many easy to use packages for statisticians	Can easily perform matrix computation as well optimization
IDE	Rstudio, R GUI	Spyder, IPython, Jupiter Notebook
Used by	Statisticians, Analyst and Data Scientist	Developer, Data engineers and Data Scientist
Suitable for	People with no prior experience in programming	Newbies to experienced IT professionals
Essential Packages and library	Ggplot2, tidyverse, caret	Numpy, pandas, scipy, TensorFlow.

MERITS

- R is the most comprehensive statistical analysis package available. It incorporates all of the standard statistical tests, models, and analysis, as well as providing a comprehensive language for managing and manipulating data.
- R is a programming language and environment developed for statistical analysis by practicing statisticians and researchers.
- The graphical capacity of R is outstanding, providing a fully programmable graphics language that surpasses most other statistical and graphical packages.
- R is free and open source software, allowing anyone to use, importantly to modify it.
- R has over 4800 packages available from multiple repositories specializing in topics like econometrics, data mining, spatial analysis, and bioinformatics.
- Online help and discussion, solution of big data.

DEMERITS

R is slow:

R is an interpreting language and is not very fast. Could be 1/40 of C. However, no steeper than for other statistical languages.

Limitation of Memory:

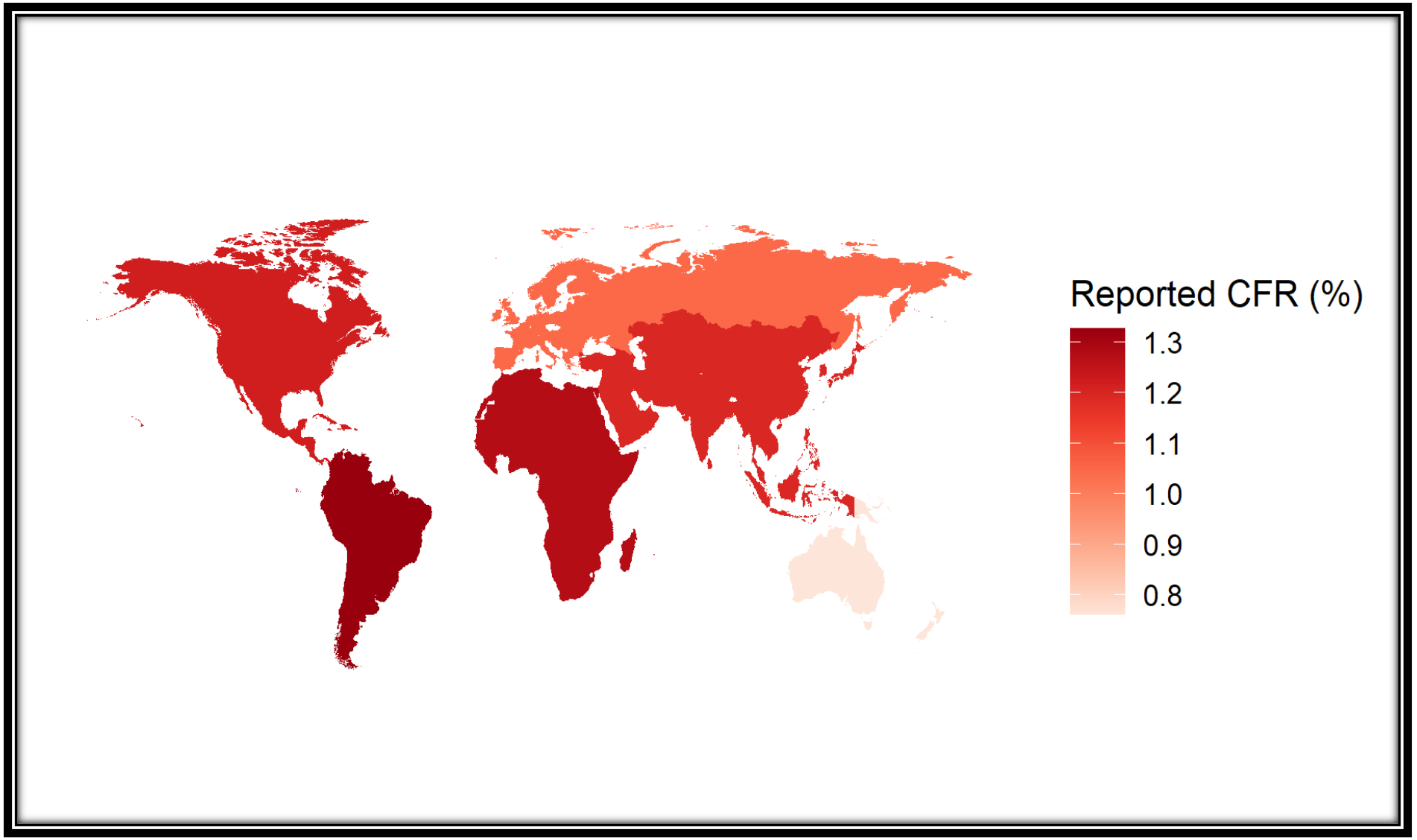
All the objects are in memory. So, speed and efficiency are probably the biggest challenges in R.

R is hard to learn:

One has to memorize the commands/functions and understand the logic of programming. Fluency in R requires great time and energy.

“I'M NOT A GREAT PROGRAMMER; I'M JUST A GOOD PROGRAMMER WITH GREAT HABITS”

KENT BECK



THANK YOU!

MOHAMMAD NAYEEM HASAN

NHASAN.COM | NAYEEM5847@GMAIL.COM