



DATA ANALYTICS

Unit 1:The R Programming Environment

Mamatha.H.R

Department of Computer Science and Engineering

DATA ANALYTICS

Unit 1:The R Programming Environment

Mamatha H R

Department of Computer Science and Engineering

- R is a free, open-source programming language and software environment for **statistical computing, bioinformatics, visualization and general computing.**
- R is an **interpreted language.**
- **Ross Ihaka & Robert Gentleman in 1993**
- Extensive catalog of statistical and graphical methods
- Environment for statistical analysis, graphics representation and reporting

- **Most widely used data analysis software**
Used by 2M+ data scientists, statisticians and analysts
- **Most powerful statistical programming language**
Flexible, extensible and comprehensive for productivity
- **Create beautiful and unique data visualizations**
As seen in New York Times, Twitter and Flowing Data
- **Thriving open-source community**
Leading edge of analytics research

DATA ANALYTICS

What is R used for?

- Statistical Computing
- Data analysis
- Machine learning algorithm
- Visualization
- Scientific research



DATA ANALYTICS

Why use R for data analytics, statistical computing and graphics?

- R is open source and free!
- R is popular – and increasing in popularity
- R runs on all platforms
- Thousands of R packages.
- R is being used by the biggest tech giants
- R is interactive language and expressive syntax
- R has easy-to-use interface.



Applications of R Programming in Real World

R – allows to collect data in real-time, perform statistical and predictive analysis, create visualizations and communicate actionable results to stakeholders

More than 9100 packages of statistical function.

R's expressive syntax allows to quickly import, clean and analyze data from various data sources.

R also has charting capabilities, which means you can plot your data and create interesting visualizations from any dataset.

DATA ANALYTICS

Applications of R Programming in Real World

R - used in predictive analytics and machine learning.

Packages for ML tasks like linear and non-linear regression, decision trees, linear and non-linear classification and many more.

R has been used primarily in academics and research and is great for exploratory data analysis.

In recent years, enterprise usage has rapidly expanded. Statisticians, engineers, and scientists without computer programming skills.

It's popular in academia, finance, pharmaceuticals, media, and marketing



DATA ANALYTICS

Applications of R Programming in Real World

R has a rich ecosystem of cutting-edge interface packages available to communicate between open-source languages.

This allows users to string their workflows together, which is especially useful for data analysis.



DATA ANALYTICS

Data Handling Capabilities

- R is great for data analysis because of its huge number of packages, readily usable tests, and the advantage of using formulas.
- It can handle basic data analysis without needing to install packages.
- Big datasets require the use of packages such as data.table and dplyr.



DATA ANALYTICS

Popular Libraries and Packages

- dplyr, tidyr and data.table to easily manipulate data
- stringr to manipulate strings
- zoo to work with regular and irregular time series
- ggplot2 to visualize data
- caret for machine learning



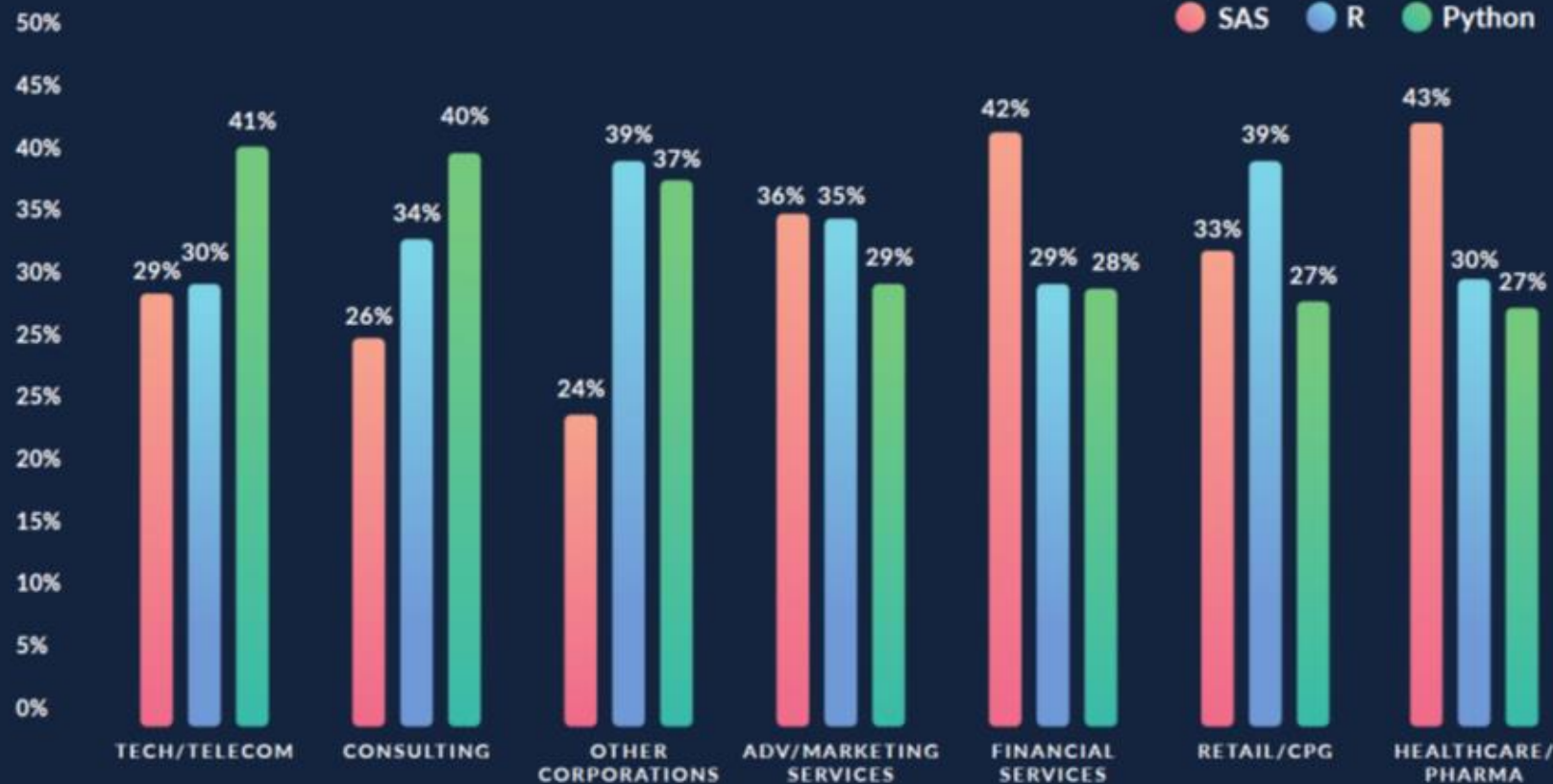
- R is the highest paid IT skill
- R most-used data science language after SQL
- R is used by 70% of data miners
- R is #8 of all programming languages(ref. TIOBE Index 2020)
- R growing faster than any other data science language
- R is the #1 Google Search for Advanced Analytics software
- R has more than 2 million users worldwide

- Python – Popular general purpose language
- SAS (Statistical Analysis System)
- SPSS – Software package for statistical analysis

DATA ANALYTICS

Preference by Industry

SAS, R or Python preference by industry



- The **S** language has been developed since the late 1970s by **John Chambers** and colleagues at Bell Labs as a **language for programming with data**.
 - S language combines ideas from a variety sources (awk, lisp, APL,) and provides an environment for quantitative computations and visualization.
 - Provides an explicit and consistent structure for manipulating, analyzing statistically, and visualizing data.
- **S-Plus** is a commercialization of the Bell Labs framework. It is "S" plus "graphics".

- R - Open source statistical environment/platform developed by Robert Gentleman and Ross Ihaka (University of Auckland, NZ) during the 1990s.
- R is currently maintained by the R core-development team, a hard-working, international team of volunteer developers.
- The primary R system is available from the Comprehensive R Archive Network, also known as CRAN.
- CRAN also hosts many add-on packages that can be used to extend the functionality of R. Over 6,789 packages are available on CRAN that have been developed by users and programmers around the world.

Finding out the latest version of R:

- To find out what is the latest version of R, you can look at the CRAN (Comprehensive R Network) website, <http://cran.r-project.org/>.

Installing R on Windows:

- To install R on your Windows computer, follow the below steps:
- Go to <http://ftp.heanet.ie/mirrors/cran.r-project.org>. or <https://cran.rstudio.com/bin/windows/base/>
- Under “Download and Install R”, click on the “Windows” link.
- Download the required .exe file. You should see a link saying something like “Download R 4.0.2 for Windows” (or R X.X.X, where X.X.X gives the version of R, eg. R 4.0.2). Click on that link.
- After downloading double-click on the R-4.0.2-win.exe to run it.
- You will be asked what language to install it in – choose English.

DATA ANALYTICS

Download and Install RStudio on Windows

<https://www.rstudio.com/products/rstudio/download/#download/>

and click on [DOWNLOAD RSTUDIO DESKTOP.](#)

List of IDEs

- RStudio
- StatET for R (eclipse based)
- R-Brain IDE (RIDE)
- IntelliJ IDEA
- R Tool for Visual Studio



R getwd() Function

- Working directory is the directory where R finds all R file for reading and writing.
- getwd() function returns an absolute filepath representing the current working directory of the R process.
 - **getwd()**
 - **"C:/Users/*****/Documents"**

R setwd() Function

- setwd(dir) is used to set the working directory to dir.

setwd("e:/folder/")

R dir() Function

- dir() function lists all the files in a directory.

dir()

R ls() Function

- ls() is a function in R that lists all the object in the working environment. **ls()**
- It can be used in scenario where you want to clean the environment before running code. Below command will remove all the object from R environment.

rm(list = ls())

- To get general help just type the below command

help.start()

- To access documentation for the standard lm (linear model) function, for example, enter the command.

help(lm) / help("lm") / ?lm /?"lm"

- To see the list of pre-loaded data (datasets), type the function,

data()

- The primary location for obtaining R packages is [CRAN](https://cran.r-project.org/).
- Information about the available packages on CRAN with the *available.packages()* function.

```
a <- available.packages()
```

- Packages can be installed with the *install.packages()* function in R.

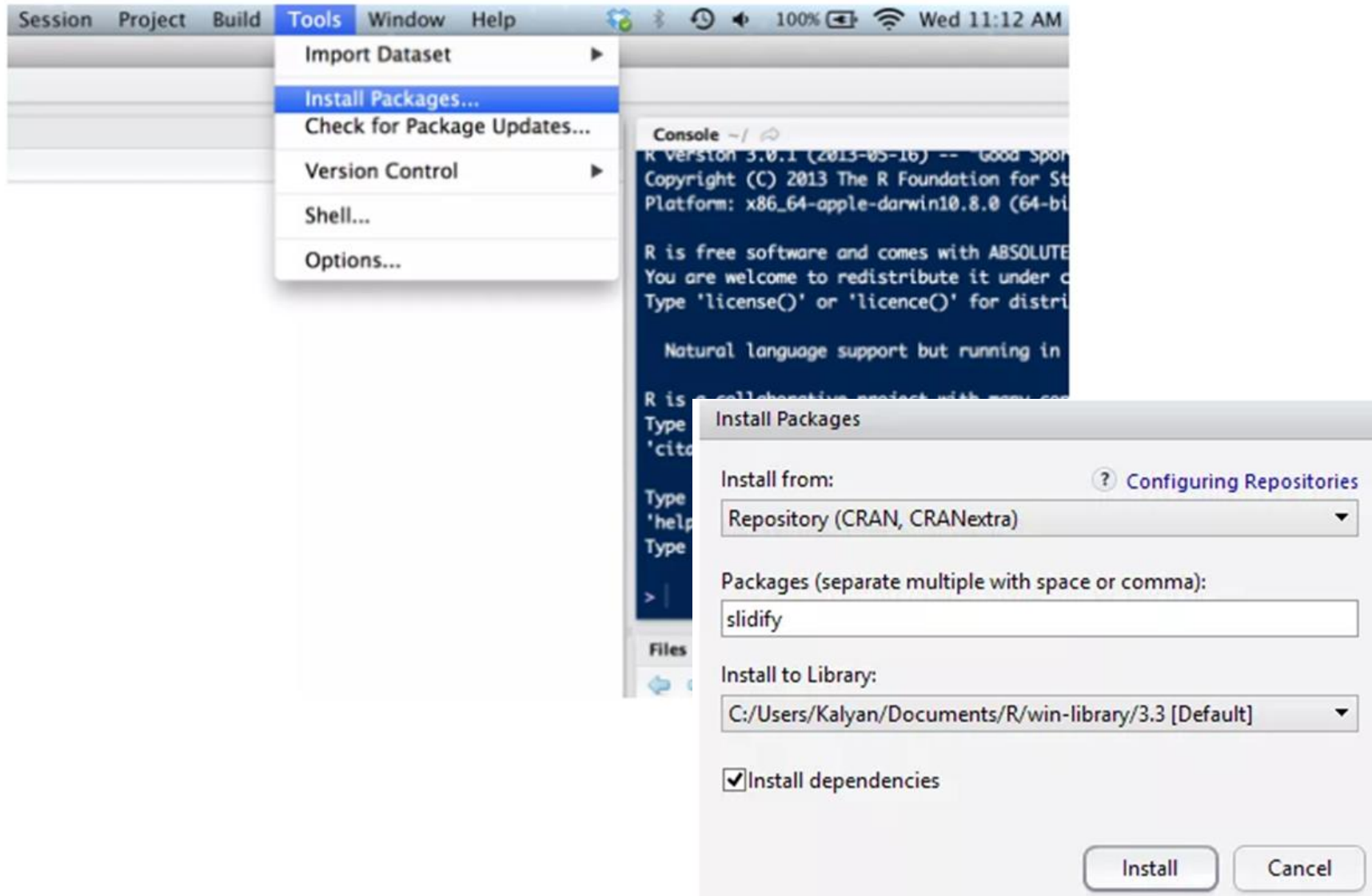
```
install.packages("ggplot2")
```

- Install multiple R packages at once with a single call to *install.packages()*. Place the names of the R packages in a character vector.

```
install.packages(c("caret", "ggplot2",  
"dplyr"))
```

DATA ANALYTICS

Installing an R Package in RStudio



The screenshot shows the RStudio interface. The 'Tools' menu is open, and 'Install Packages...' is selected. The 'Install Packages' dialog box is open, showing the following options:

- Install from:** Repository (CRAN, CRANextra) (with a link to 'Configuring Repositories')
- Packages (separate multiple with space or comma):** slidify
- Install to Library:** C:/Users/Kalyan/Documents/R/win-library/3.3 [Default]
- ☒ **Install dependencies**

Buttons at the bottom: Install, Cancel

- Installing a package does not make it immediately available to you in R; you must load the package. The `library()` function is used to load packages into R.

`library(ggplot2)`

- After loading a package, the functions exported by that package will be attached to the top of the `search()` list (after the workspace).

**`library(ggplot2)`
`search()`**

- To save your workspace to a file, you may type

`save.image()` or use **Save Workspace...** in the **File** menu

- The default workspace file is called **`.RData`**

DATA ANALYTICS

Exercise

Install R and R Studio



<https://www.tutorialspoint.com/>



THANK YOU

Dr.Mamatha H R

Professor,Department of Computer Science

mamathahr@pes.edu

+91 80 2672 1983 Extn 834