Fakulti Sains Komputer dan Matematik

UNIVERSITI TEKNOLOGI MARA

Premier Digital Tech Institution

INTRODUCTION TO R PROGRAMMING

BUILDING A KNOWLEDGE FOUNDATION FOR SUCCESS IN THE BUSINESS WORLD

Dr Mohammad Nasir Abdullah
PhD(Statistics), MSc (Medical Statistics), BSc(hons) (Statistics), Diploma in Statistics, Data Sciences Specialist, Graduate Statistician
Senior Lecturer,
Department of Statistics,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
Tapah Campus, Tapah, Perak.
nasir916@uitm.edu.my

Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

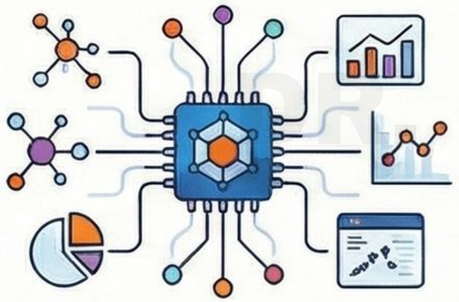
Premier Digital Tech Institution

posit

Watch this!

WHAT IS R?

A Language for Statistical Computing



A free software environment for statistical modeling, analysis, and graphical representation.

An Open-Source Implementation of S

Created as a GNU project similar to the S language from Bell Labs.

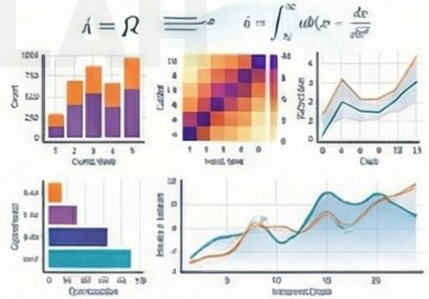
Core Strengths

Wide Variety of Statistical Techniques

Includes modeling, classical tests, time-series analysis, classification, clustering, and more.

Publication-Quality Graphics

Easily create well-designed plots with full user control and mathematical symbols.

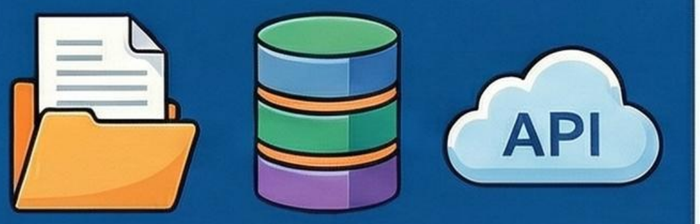


Free & Cross-Platform

Available on UNIX-like systems (including Linux), Windows, and MacOS.

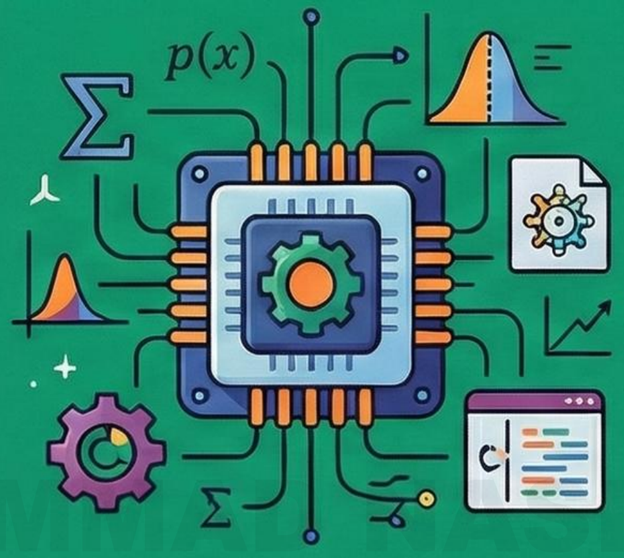
Universiti Teknologi MARA
Fakulti Sains Komputer dan Matematik
Premier Digital Tech Institution
posit

WHAT R DOES?



Comprehensive Data Handling

Read, clean, and transform data from files, databases, and APIs.

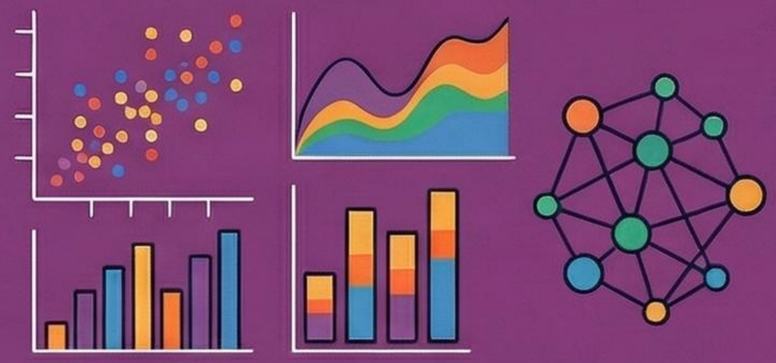


Powerful Statistical Analysis

Perform everything from hypothesis testing to machine learning and predictive modeling.

A True Programming Language

Includes matrix algebra, loops, branching, and object-oriented features.



High-Quality Data Visualization

Create publication-ready piots and build interactive dashboards using Shiny.

Advanced Capabilities & Automation

Automated Reporting

Generate reproducible reports and dynamic documents in PDF, HTML, or Word.

Specialized Field Applications

Widely used in bioinformatics, finance, social science research, and AI.

Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

Premier Digital Tech Institution

posit

WHAT R DOES NOT?

BUILT-IN LIMITATIONS

- **Lacks a native Graphical User Interface (GUI).**
- **Is not a database.**
- **The language interpreter can be slow.**
- **Has no built-in spreadsheet view.**

WORKAROUNDS & IDEAL USE CASES

- **Connects to External GUI Toolkits.**
  It can connect to GUI toolkits like Java and TclTk.
- **Utilizes Database Connectors.**
  It provides connectors for most database management systems (DBMSs).
- **Integrates Faster C/C++ Code.**
  It allows you to call your own faster C/C++ code.
- **Direct Connection to MS Office.**
  It can connect and export data with Excel and other Office tools.

WHEN TO CHOOSE ANOTHER TOOL

- **R is Analysis-Centric, Not Application-Centric.**
  It is purpose-built for data analysis, not general software development.
- **R is not ideal for:**
  Mobile app development, game development, or low-level systems programming.
- **No Commercial Support**
  R is open-source and relies on community support.

Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

Premier Digital Tech Institution

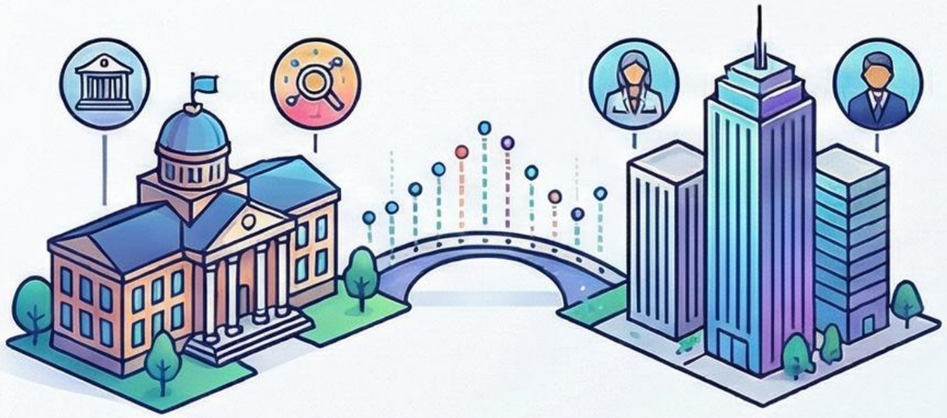
posit

WHY CHOOSE R ?

ACCESSIBLE & WIDELY ADOPTED

Completely Free & Open Source

R has no licensing costs, making powerful data analysis tools accessible to everyone.



Trusted by Industry & Academia

Used by professional researchers, analysts, and data scientists around the world.

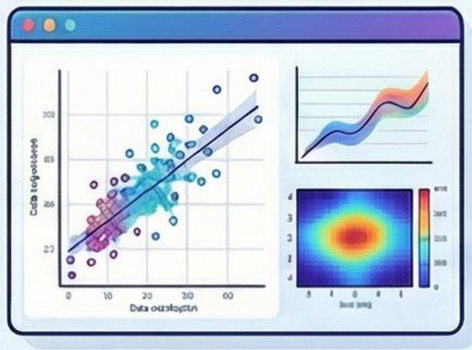
POWERFUL TECHNICAL CAPABILITIES

20,000+ Packages in Ecosystem

The CRAN repository provides a package for almost any statistical method imaginable.

Designed Specifically for Data

Core statistical methods are built-in, requiring less code for complex analysis.



Elite Visualization & Reporting

Create publication-quality graphics and combine code with reports for reproducible research.

ADVANTAGES USING R

- FAST AND OPEN SOURCE
- INTEGRATES WITH PYTHON, SQL, SAS, SPSS, EXCEL
- EXCELLENT GRAPHIS AND VISUALIZATION
- STRONG STATISTICAL CAPABILITIES
- LARGE AND ACTIVE COMMUNITY
- HIGHLY ESTENSIBLE WITH PACKAGES
- IDEAL FOR RESEARCH AND ANALYTICS

DISADVANTAGES USING R

- STEEPER LEARNING CURVE FOR BEGINNERS
- MEMORY-BASED (LOADS DATA INTO RAM)
- SLOWER FOR VERY LARGE DATASETS (COMPARED TO LOW-LEVEL LENGUAGES)
- NOT SUITABLE FOR GENERAL SOFTWARE DEVELOPMENT

Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

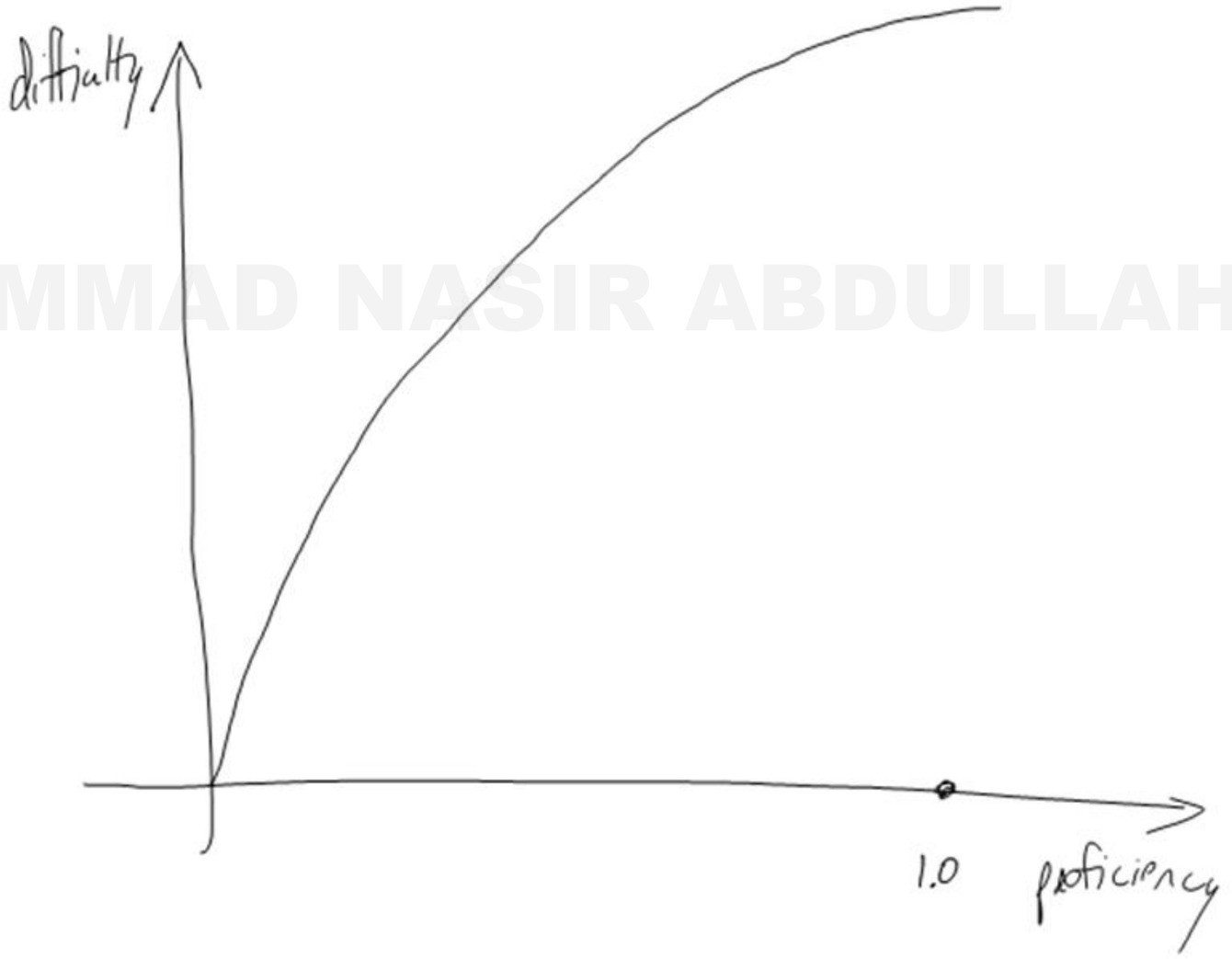
Premier Digital Tech Institution

posit

R HAS A STEEP LEARNING CURVE

Steeper for those that knew SAS or other software before.

- Basic commands
- Interface
- Data types
- Packages
- Procedures



R HISTORY

- R is a comprehensive statistical and graphical programming language and is a dialect of the S language.
  - 1988 – S2: RA Becker, JM Chambers, A Wilks.
  - 1992 – S3: JM Chambers, TJ Hastie.
  - 1998 – S4: JM Chambers
- R: Initially written by Ross Ihaka and Robert Gentleman at Department of Statistics of University of Auckland, New Zealand during 1990s.
  - He retired as an associate professor of statistics at the University of Auckland
- Since 1997: International "R-Core" team of 15 people with access to common CVS archive.

Ross Ihaka

Robert Gentleman

INSTALLING R

http://cran.r-project.org
http://posit.co

1: Install R

RStudio requires R 3.6.0+. Choose a version of R that matches your computer's operating system.

R is not a Posit product. By clicking on the link below to download and install R, you are leaving the Posit website. Posit disclaims any obligations and all liability with respect to R and the R website.

DOWNLOAD AND INSTALL R

LATEST VERSION : 4.5.2

2: Install RStudio

DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS

Size: 296.74 MB | SHA-256: 439D3200 | Version: 2025.09.2+418 | Released: 2025-10-29

R PACKAGES

http://cran.r-project.org

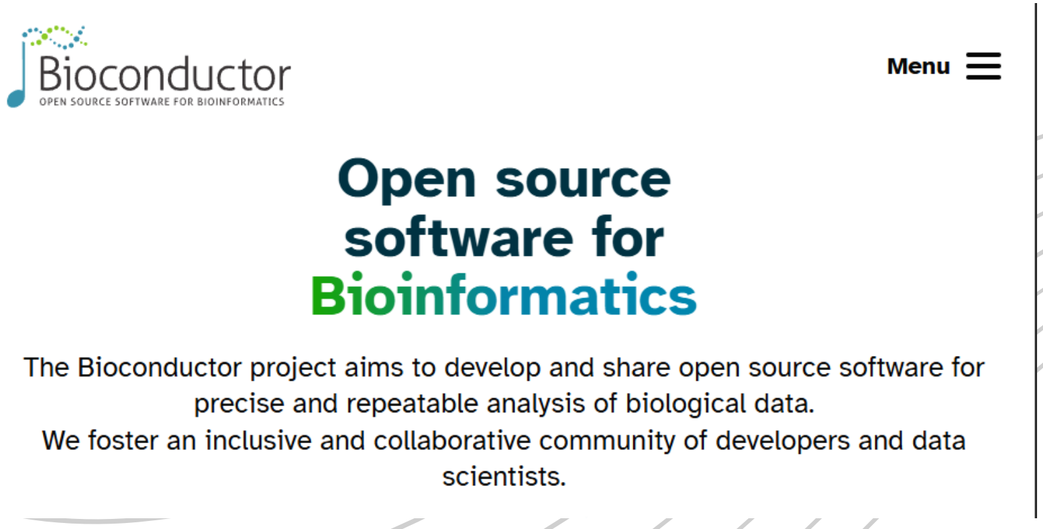
- The most important single innovation in R is the package system, which provides a cross-platform system for distributing and testing code and data.

- The comprehensive R Archive Network (http://cran.r-project.org) distributes public packages, but packages are also useful for interval distribution.

https://www.bioconductor.org/

A PARTICULAR R STRENGTH - GENETICS

- Bioconductor is a suite of additional function and some 300 packages dedicated to analysis, visualization, and management of genetic data.
- Much more functionality than software released by Affy or Illumina.

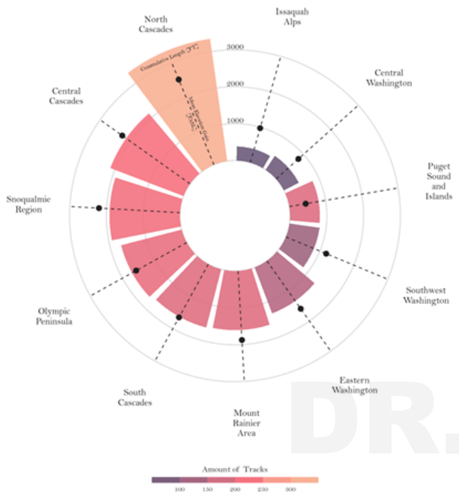


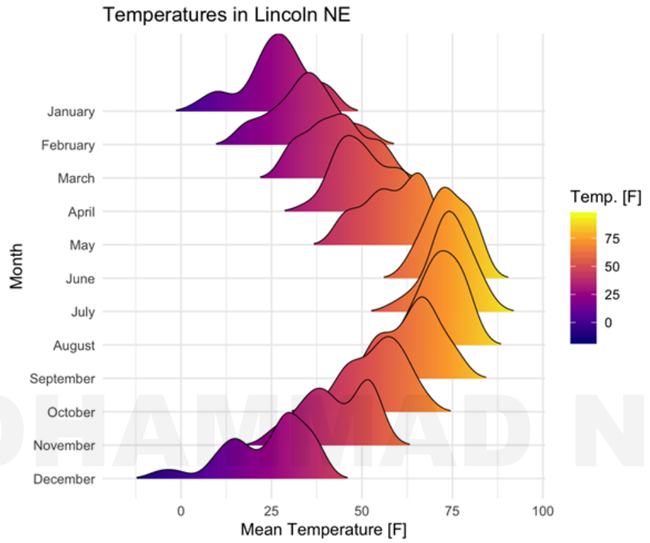
Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

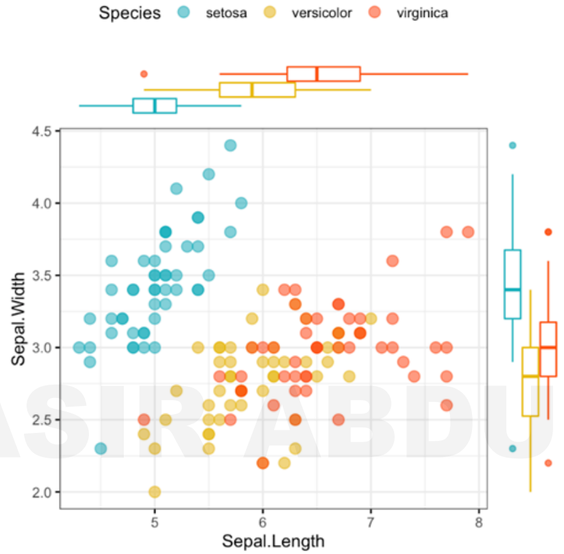
Premier Digital Tech Institution

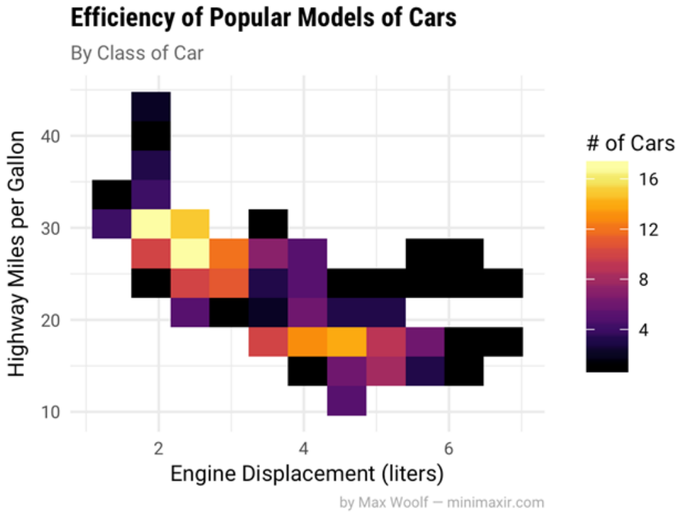
posit

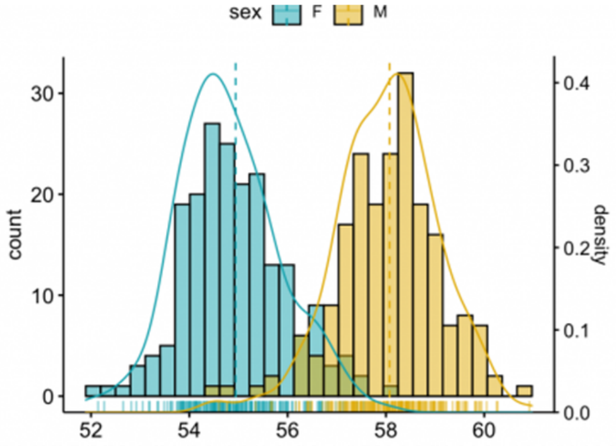
THE POWER OF R GRAPHICS

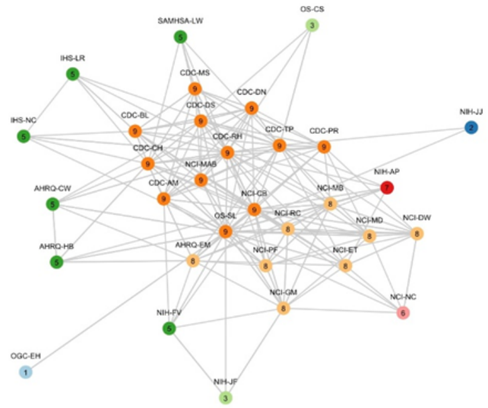












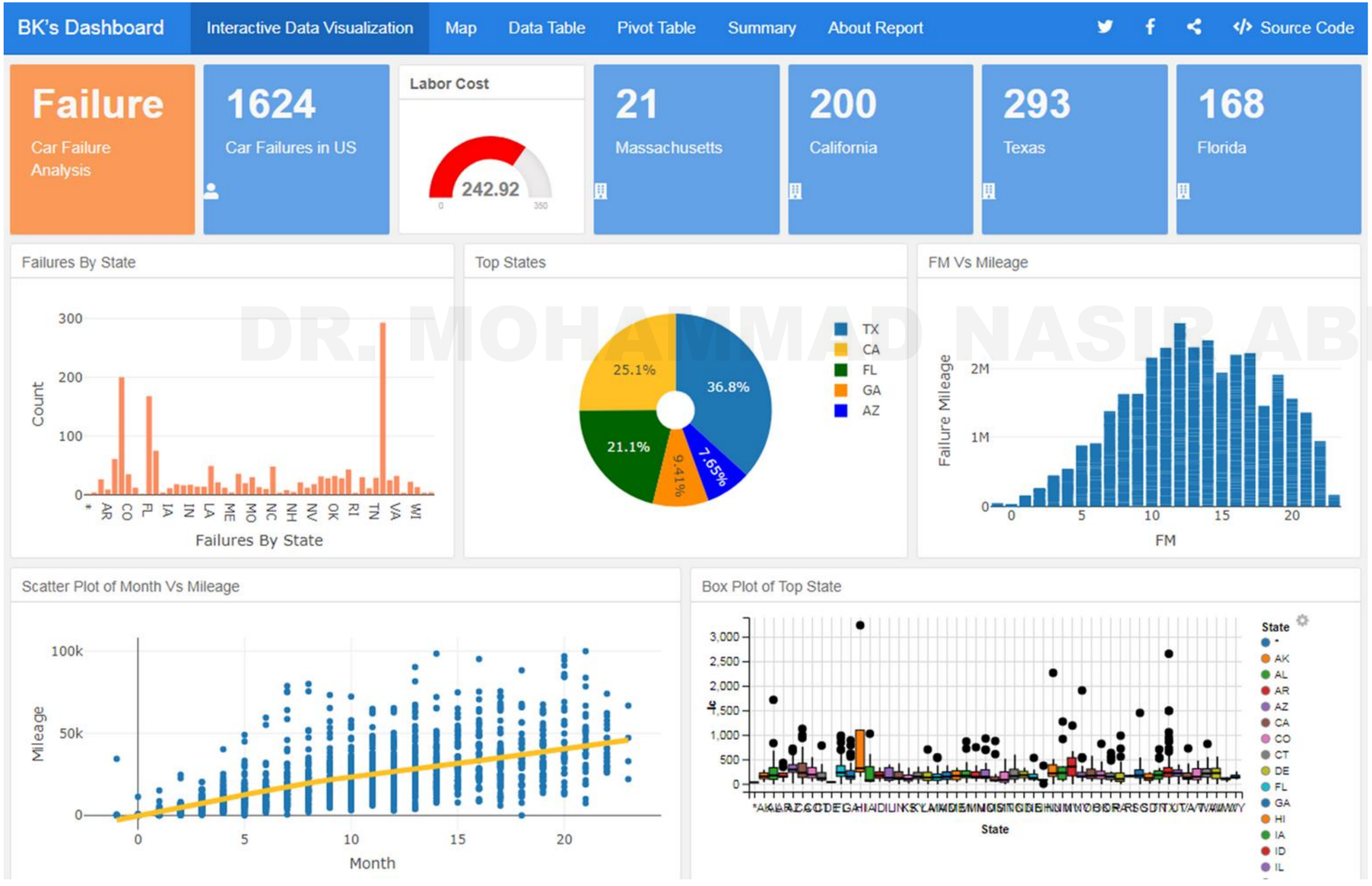
- For example: Create beautiful statistical graphics with ggplot2.
- R has a very powerful graphics system.
- With low-level tools allowing customization of every detail.

Universiti Teknologi MARA | Fakulti Sains Komputer dan Matematik

Premier Digital Tech Institution

posit

THE POWER OF R GRAPHICS



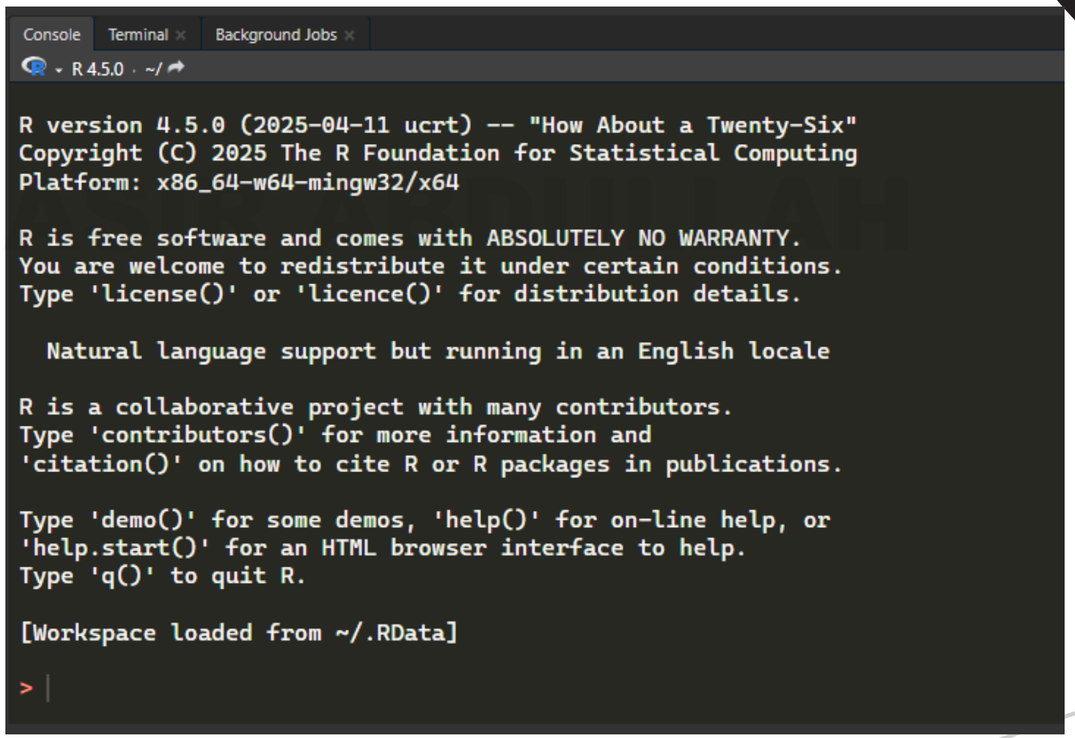
- Dashboard

USEFUL WEB LINKS

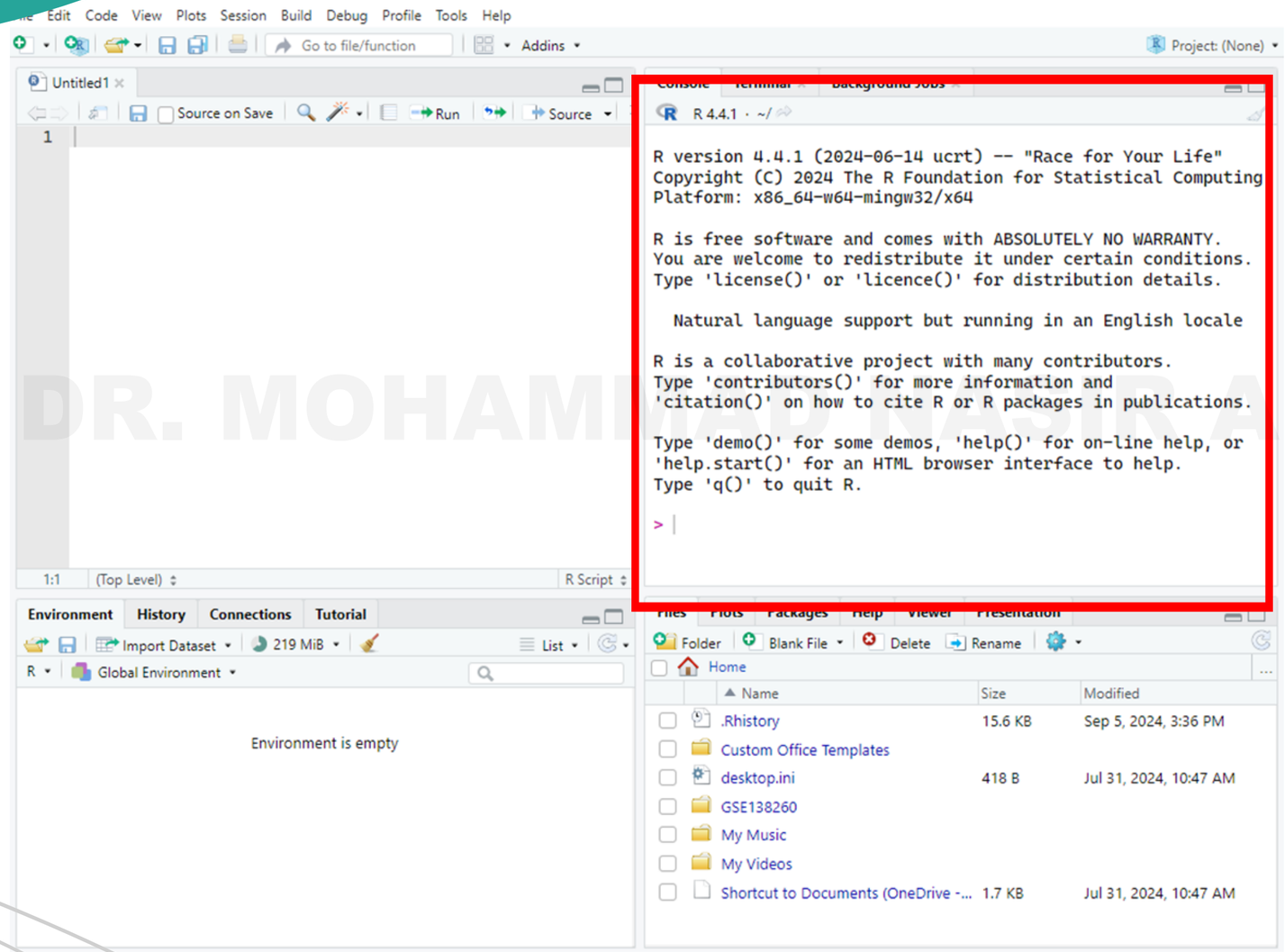
- Datacamp - http://www.datacamp.com
- UCLA Institute for Digital Research and Education - http://www.ats.ucla.edu/stat/r/
- R Reference card - http://cran.r-project.org/doc/contrib/Short-refcard.pdf
- Quick R: http://statmethods.net

WORKING WITH R

- The R console “interprets” whatever you type
  - Calculator
  - Creating variables
  - Applying functions

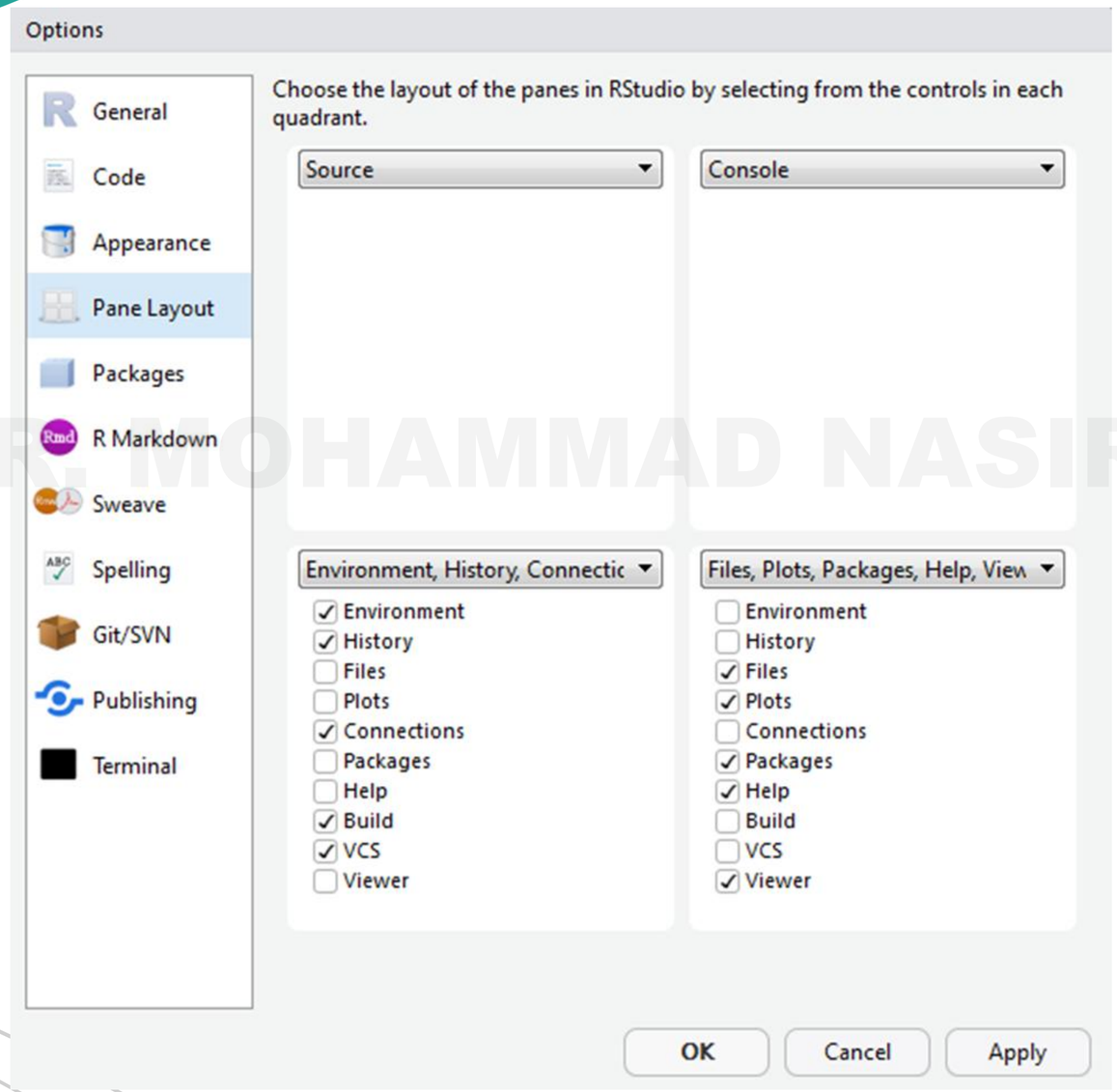


- Easier working with R
  - Syntax highlighting, code completion, and smart indentation.
  - Easily manage multiple working directories and projects.
- More information
  - Workspace browser and data viewer.
  - Plot history, zooming and flexible image, and PDF export.
  - Integrated R help and documentation.
  - Searchable command history.



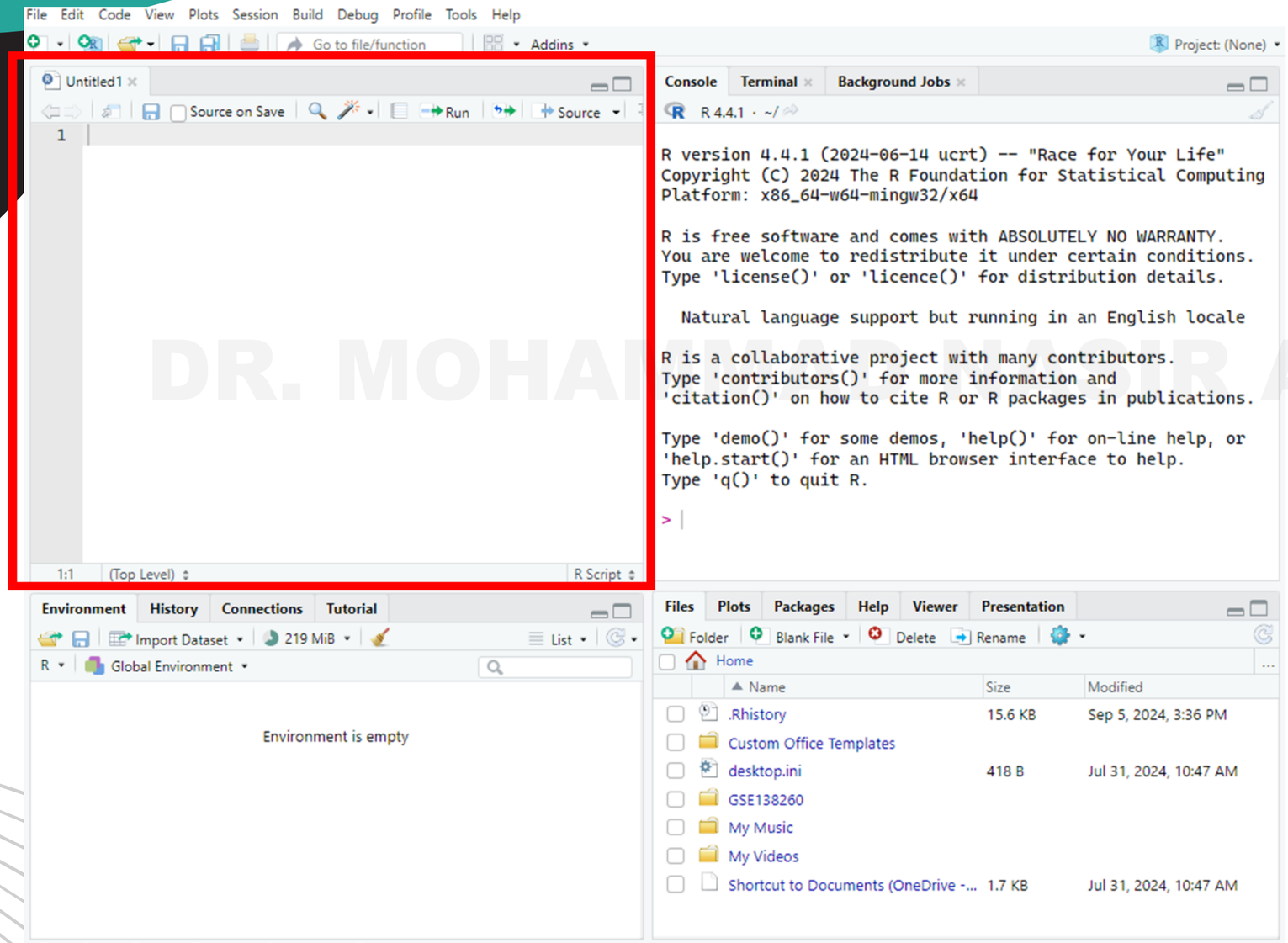
- Where code is executed (where things happen)
- You can type here for things interactively.
- Code is not saved on your disk

RSTUDIO LAYOUT



•Rstudio -> Tools -> Global Options -> Pane Layout

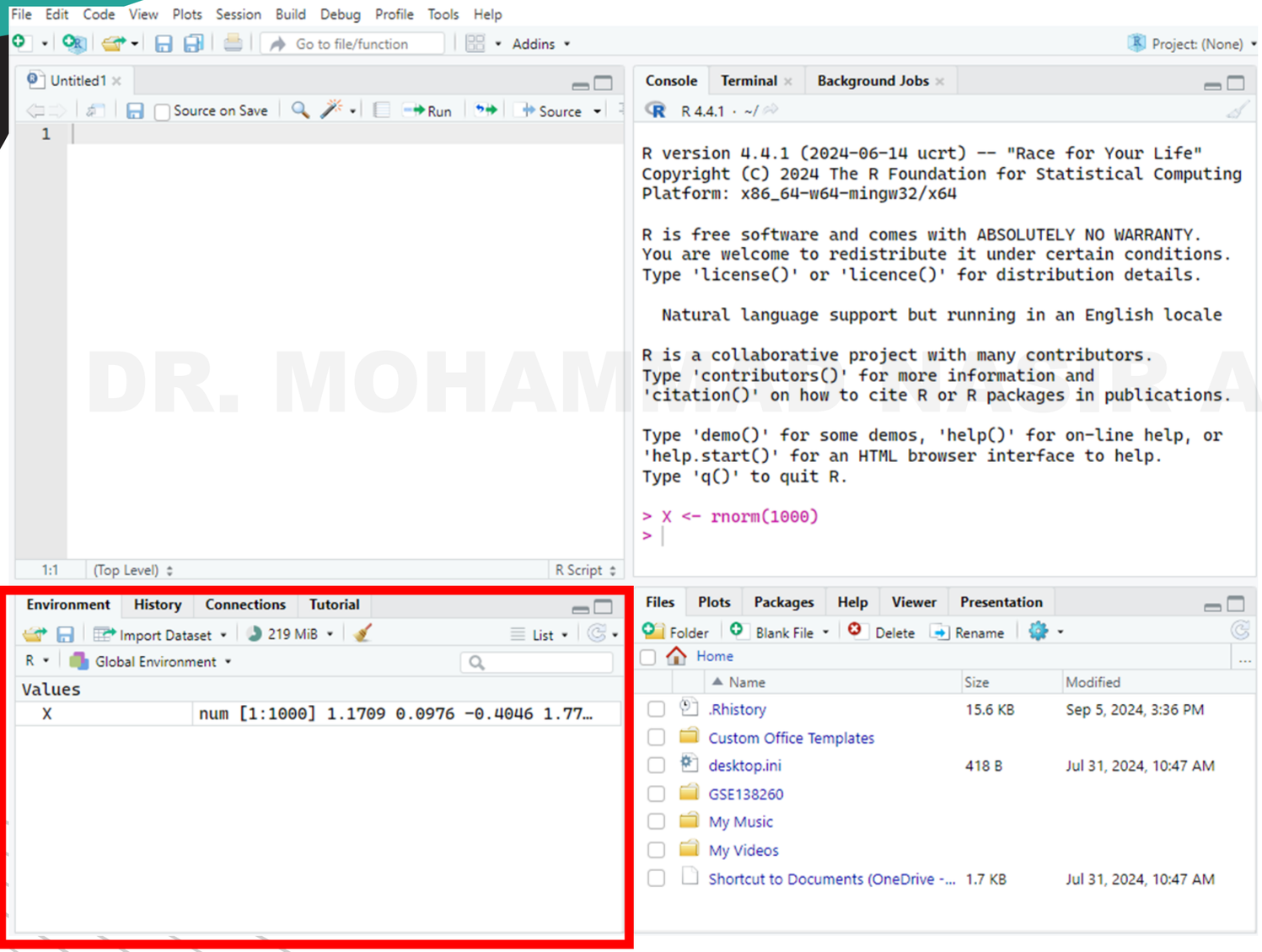
SOURCE/EDITOR



- Where files open to
- Have R code and comments in them
- Can highlight and press (CMD+Enter (Mac)) or Ctrl+Enter (Windows)) to run the code

- In **Demo.R** file (we call a script), code is saved on your disk

WORKSPACE/ENVIRONMENT



- Tells you what objects are in R.
- What exists in memory/ what is loaded?/ what did I read in?

History

- Shows previous commands. Good to look at for debugging, but don't rely on it a scripts. Make a script!
- Also type the "up" key in the console to scroll through previous commands.

OTHER PANES

- Files
  - Shows the files on your computer of the directory you are working in
- Viewer
  - Can view data or R objects
- Help
  - Shows documentation of R function
- Plots
  - Pretty graph
- Packages
  - List of R packages that are loaded in memory

THANK YOU


stringr
www.rstudio.com
ggplot2
www.rstudio.com
dplyr
dplyr.tidyverse.org
tidyr
www.rstudio.com
readr
www.rstudio.com
purrr
www.rstudio.com
tibble
www.rstudio.com
forcats
www.rstudio.com