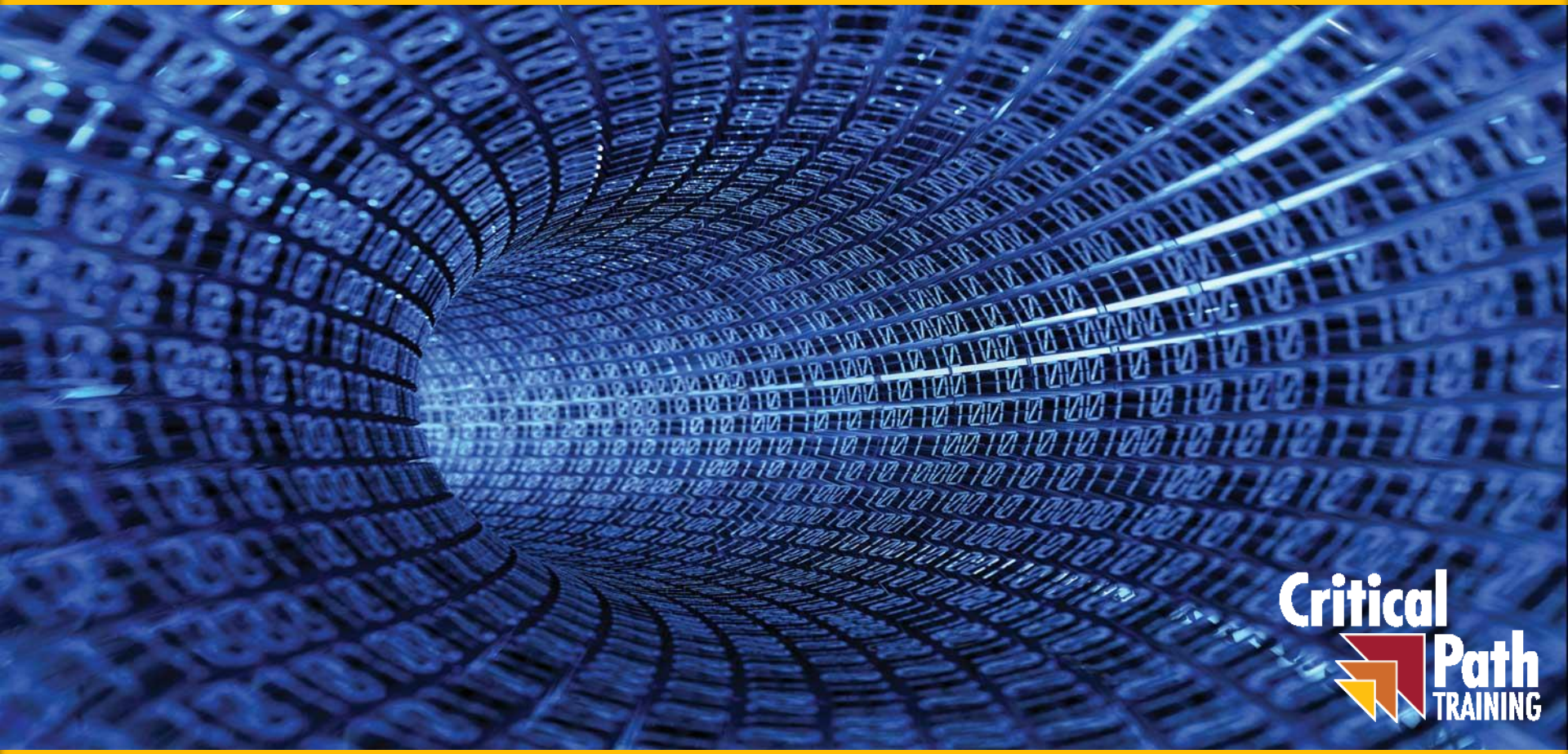
Introduction to R with Power BI



Agenda

- Understanding R as an Analytics Platform
- Installing Microsoft R Open and RStudio
- Writing R Code in RStudio
- Integrating R with Power BI Desktop

What is R?

- What is R?
  - Platform for statistics, data analysis and visualization
  - Free, cross-platform, open source software
  - Programming language + Runtime layer + Libraries
  - R code distributed and versioned using packages
  - Flourishing ecosystem of R package authors
- Why do you need it?
  - Analyzing data and generating statistics
  - Creating rich graphs and charts
  - Fitting statistical models for predictive analysis

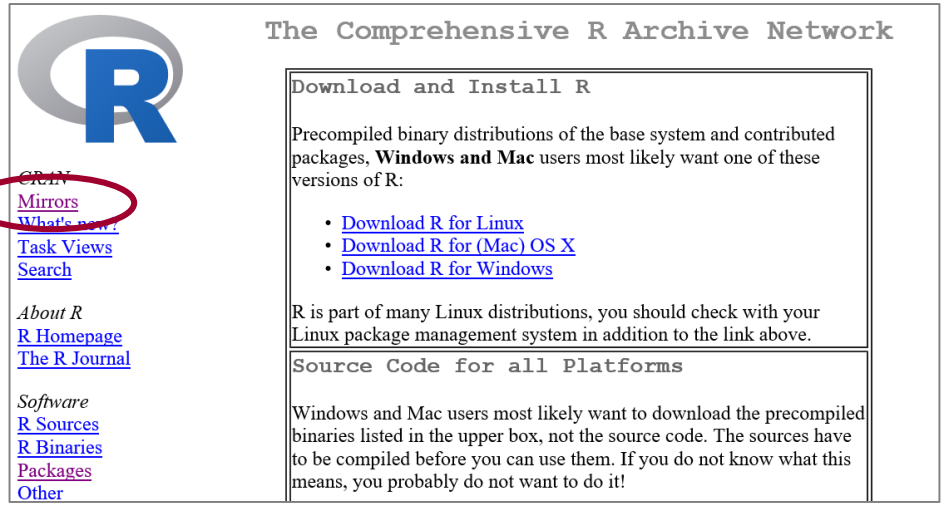
R Packages

- Package is versioned redistributable unit of code
  - Package contains functions, data and compiled code
  - R is installed with a default set of packages
  - Other packages can be downloaded and installed

- Examples of available domain-specific packages
  - Packages to download and unpack data in zip archive
  - Packages to create fancy charts and graphs
  - Packages to optimize financial portfolios
  - Packages predict component failure times
  - Packages to analyze genomic sequences

CRAN

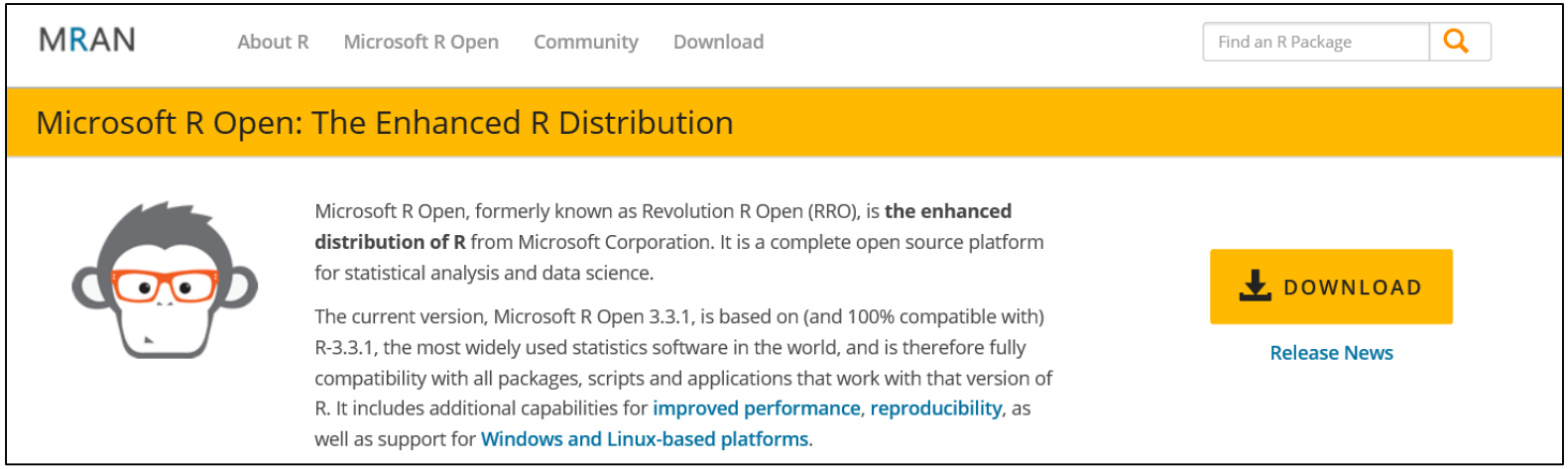
- The Comprehensive R Archive Network
  - Public archive with over 8,000 downloadable packages
  - http://cran.us.r-project.org/



Thailand		
	http://mirrors.psu.ac.th/pub/cran/	Prince of Songkla University, Hatyai
Turkey		
	http://cran.pau.edu.tr/	Pamukkale University, Denizli
	http://cran.ncc.metu.edu.tr/	Middle East Technical University Northern Cyprus Campus, Mersin
UK		
	https://www.stats.bris.ac.uk/R/	University of Bristol
	http://www.stats.bris.ac.uk/R/	University of Bristol
	https://mirrors.ebi.ac.uk/CRAN/	EMBL-EBI (European Bioinformatics Institute)
	http://mirrors.ebi.ac.uk/CRAN/	EMBL-EBI (European Bioinformatics Institute)
	http://cran.ma.imperial.ac.uk/	Imperial College London
	http://mirror.mdx.ac.uk/R/	Middlesex University London
	http://star-www.st-andrews.ac.uk/cran/	St Andrews University
USA		
	https://cran.cnr.berkeley.edu/	University of California, Berkeley, CA
	http://cran.cnr.berkeley.edu/	University of California, Berkeley, CA
	http://cran.stat.ucla.edu/	University of California, Los Angeles, CA
	http://mirror.las.iastate.edu/CRAN/	Iowa State University, Ames, IA

Microsoft R Open

- What is Microsoft R Open?
  - An enhanced distribution of R from Microsoft
  - Improved performance and multithreading
  - Reproducibility through package versioning stability
  - Free, cross-platform, open source software
  - Available at https://mran.microsoft.com/open/



Stages of R Awareness

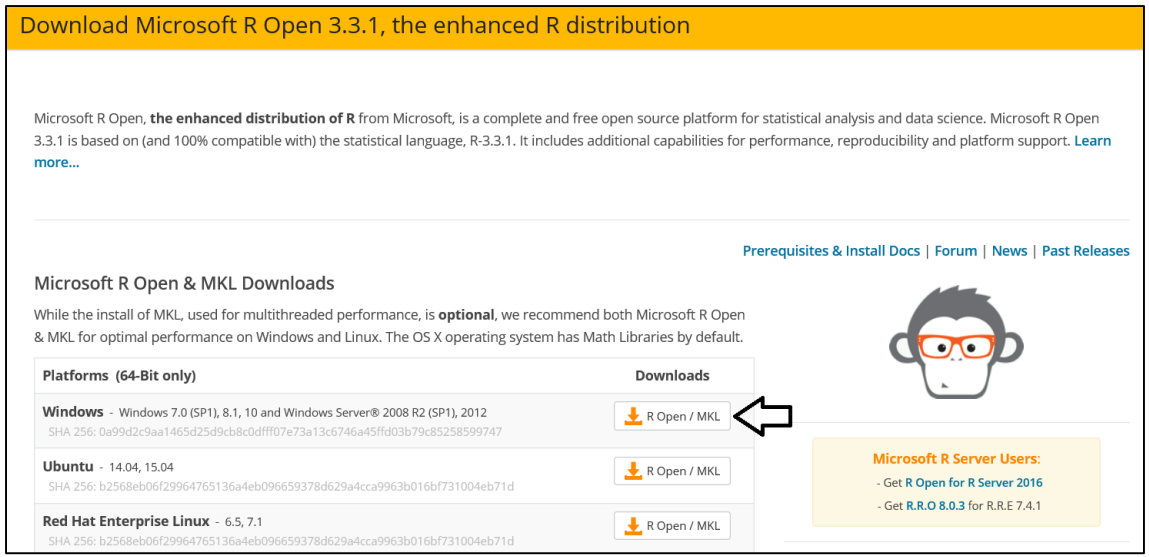
- Stage 1: Standing Up
  - Installing the environment and playing with data
- Stage 2: Walking
  - Writing & testing R code and creating graphs and charts
- Stage 3: Jogging
  - Crunching numbers to generate advanced statistics
- Stage 4: Running
  - Creating a domain-specific predictive model
- Stage 5: Sprinting
  - Distributing your predictive model as a CRAN package

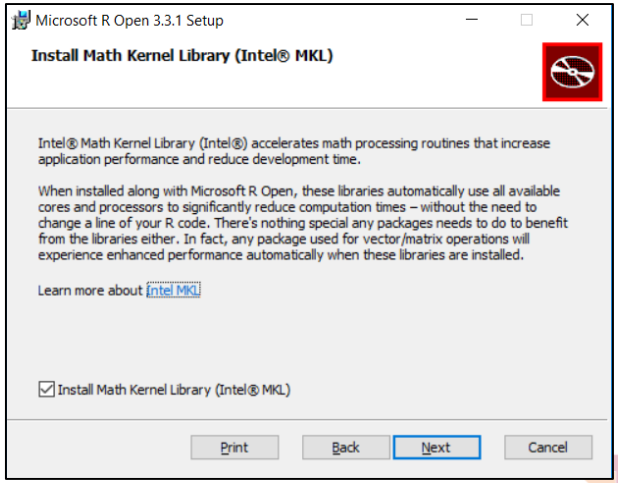
Agenda

- ✓ Understanding R as an Analytics Platform
- ➢ Installing Microsoft R Open and RStudio
- Writing R Code in RStudio
- Integrating R with Power BI Desktop

Install Microsoft R Open

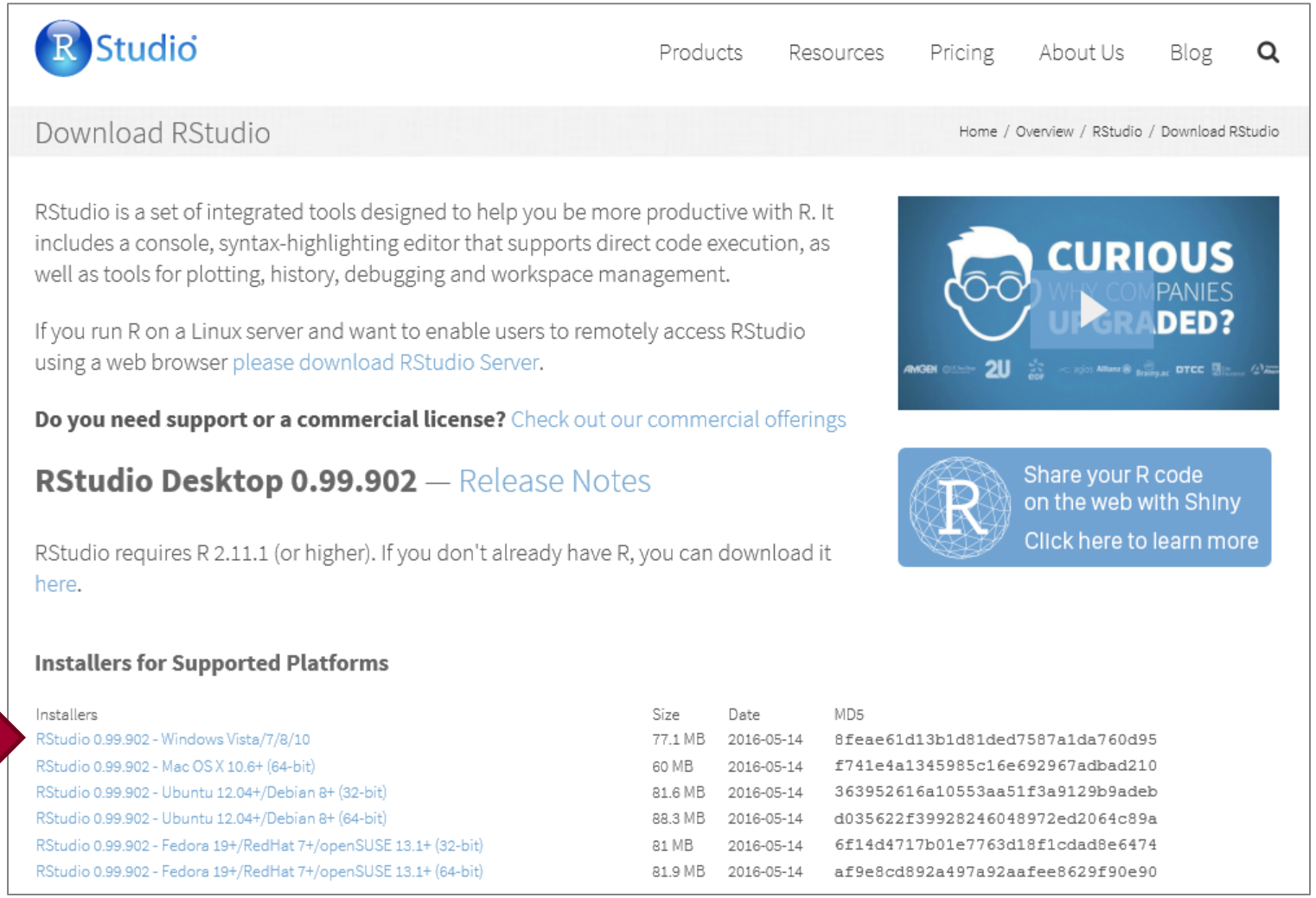
- https://mran.microsoft.com/download/



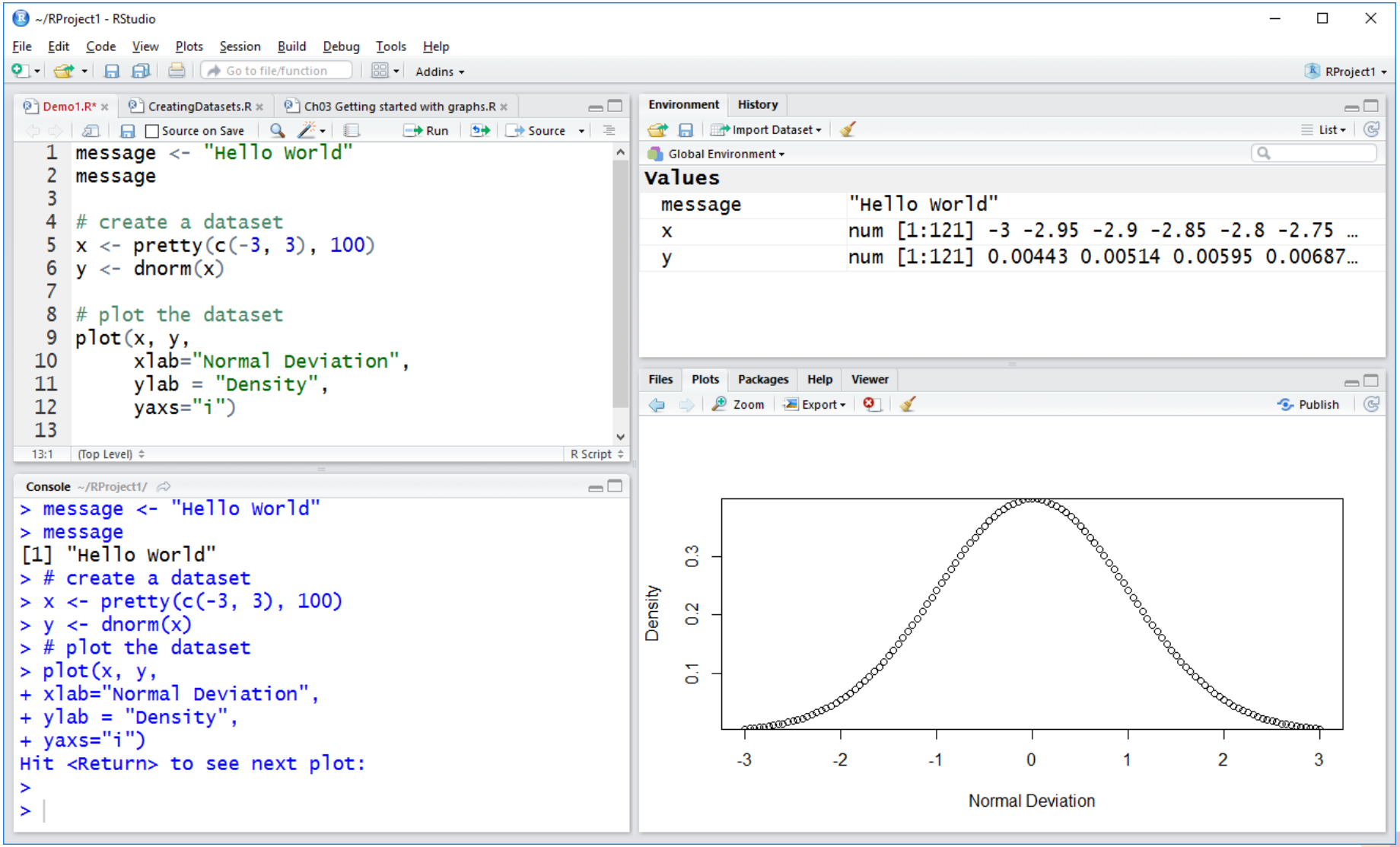


Installing R Studio

- https://www.rstudio.com/products/rstudio/download/



The RStudio IDE



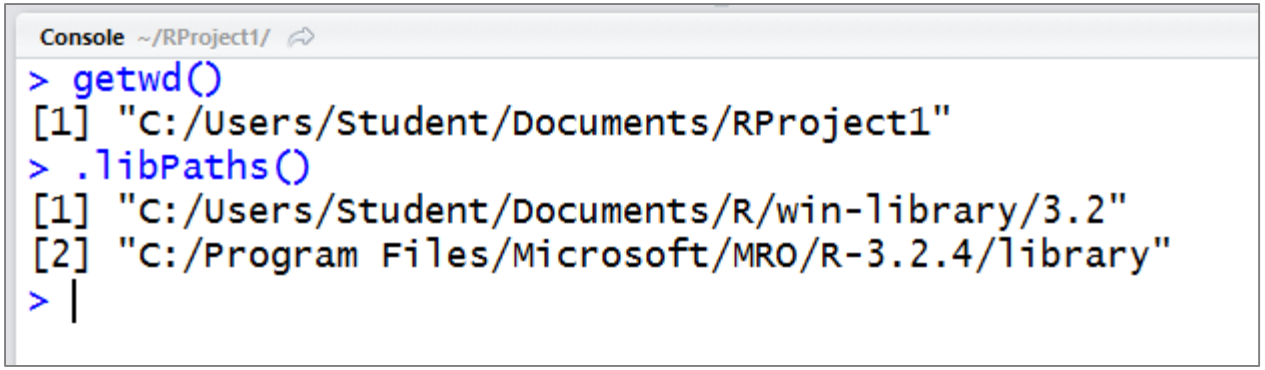
Agenda

- ✓ Understanding R as an Analytics Platform
- ✓ Installing Microsoft R Open and RStudio
- ➢ Writing R Code in RStudio
- Integrating R with Power BI Desktop

R Projects and Workspaces

- R projects based on folder structure
  - Data and scripts added to current working directory

- Each R project defines a workspace
  - Workspace tracks set of user-defined objects
  - Workspace defines set of loaded packages
  - Workspace data saved/loaded using .RData files



Writing and Testing R Code in Scripts

01_GettingStarted.R

Source on Save

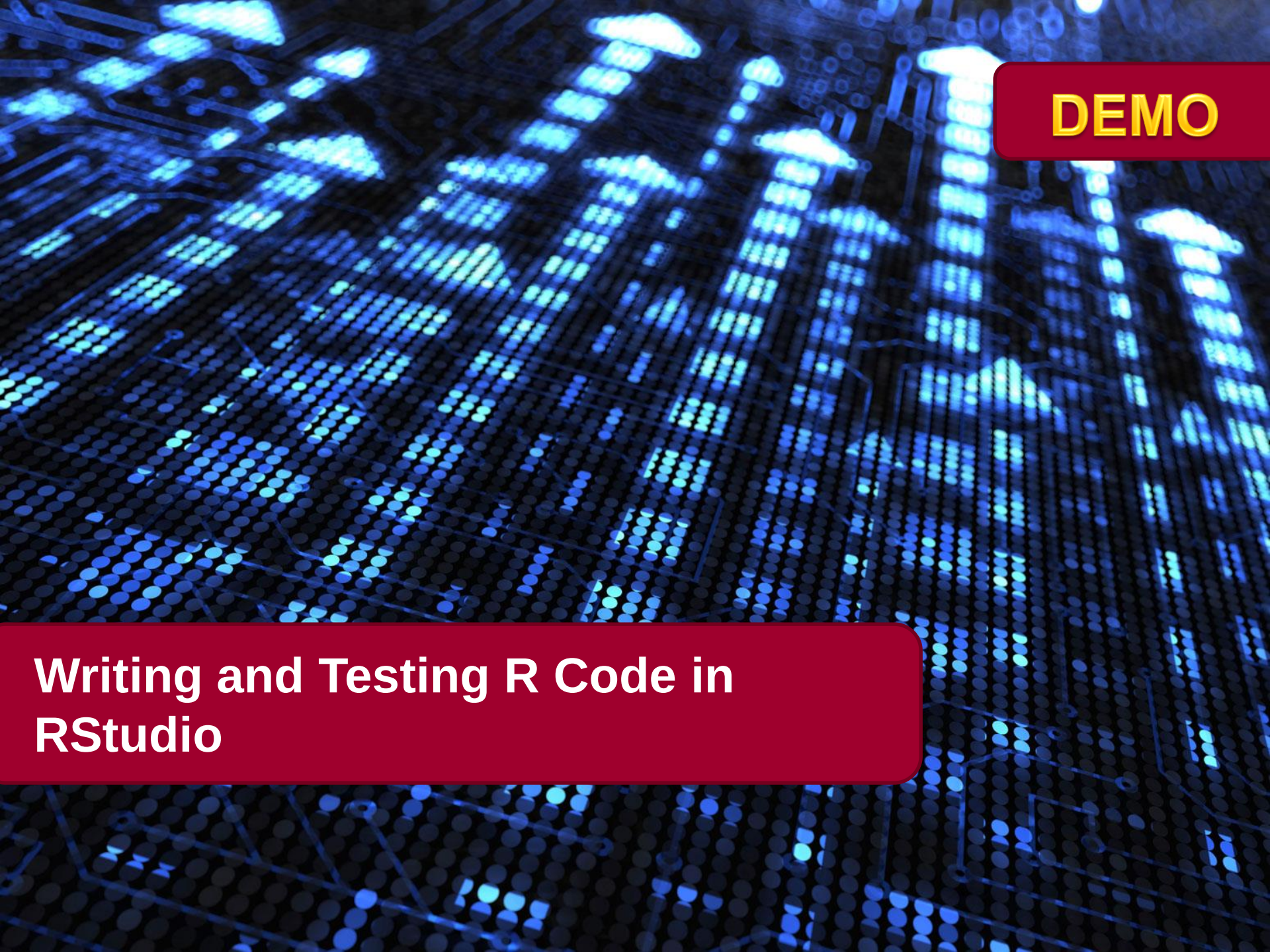
```r
# use <- for variable assignment
message <- "Hello World"

print(message)

# create vector using the c function
vector1 <- c(2, 4, 6,  8)

# create vectors using sequence
vector2 <- 1:10
vector3 = letters[1:5]
vector4 = LETTERS[24:26]
vector6 = 2^(1:8)

# create vector with electin years
election.years <- seq(from = 1996, to = 2016, by = 4)

# enumerate through election years using for loop
for (year in election.years){
    print(paste(year, "is an election year"))
}

# remove all objects from workspace
rm(list=objects())
```

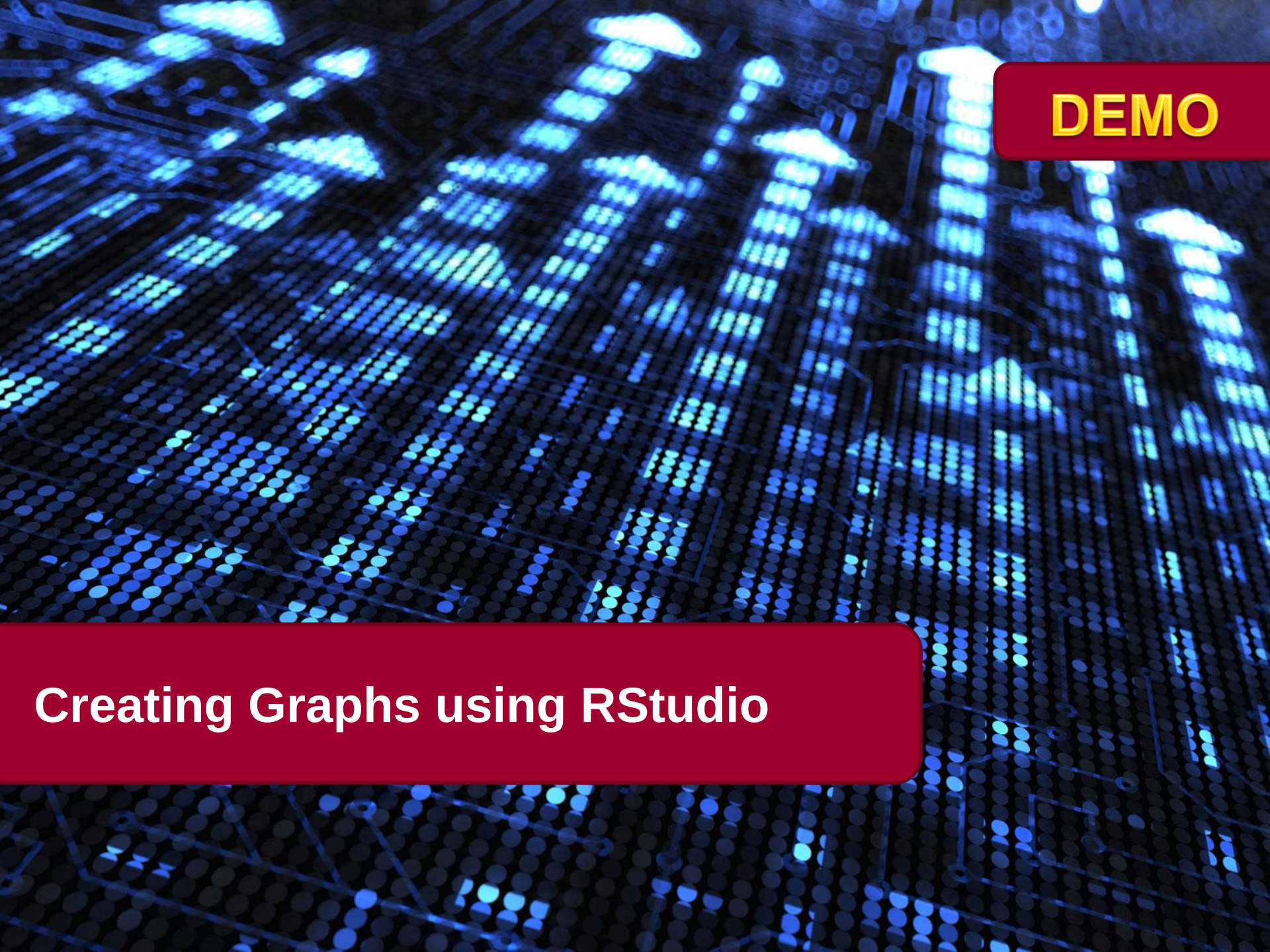
R Objects

- In R, variables represent named objects
- Object names can contain
  - Letters
  - Numbers
  - Underscores (_)
  - Dots (.)

Essential Data Structures in R

- Vector
  - One-dimensional, single-mode array
- Matrix
  - Two-dimensional, single-mode array
- Array
  - N-dimensional, single-mode array
- List
  - Ordered collection of multi-mode objects
- Data frame
  - Two-dimensional, multi-mode array
- Factor
  - Integer-backed list of categorical values


DEMO
Writing and Testing R Code in RStudio


DEMO
Creating Graphs using RStudio

Agenda

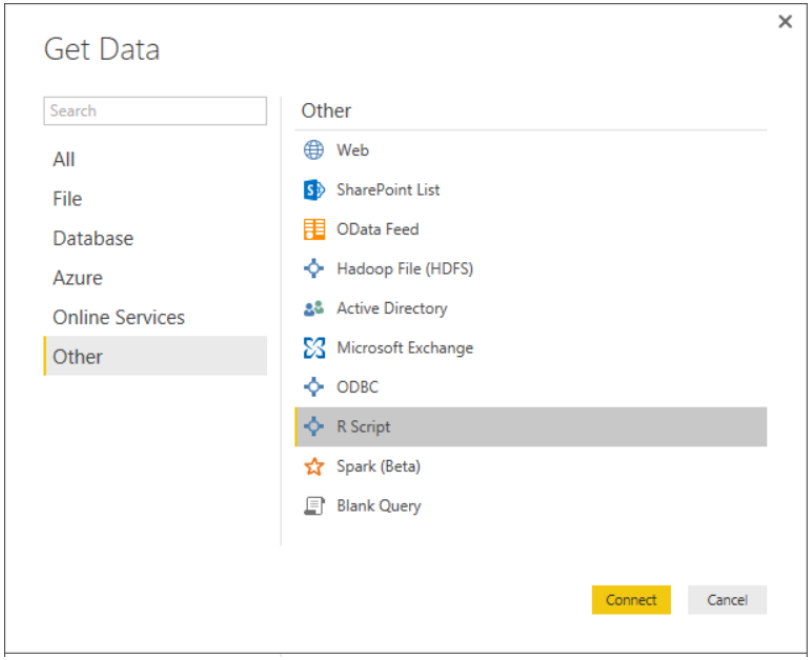
- ✓ Understanding R as an Analytics Platform
- ✓ Installing Microsoft R Open and RStudio
- ✓ Writing R Code in RStudio
- ➢ Integrating R with Power BI Desktop

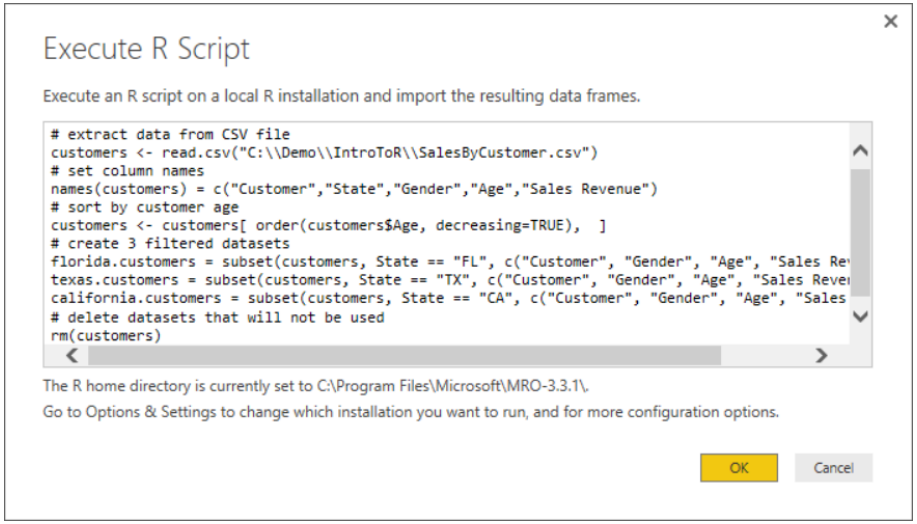
Where Can You Use R Code in PBIDT?

- As a data source to a query
  - You can use R code to import and reshape data
- Within a Query Applied Step
  - You can use R code to add transforms to a query
- Inside an R Visual in a Power BI Report
  - You can use R code to creates charts from your data

Using R Code as a Query Data Source

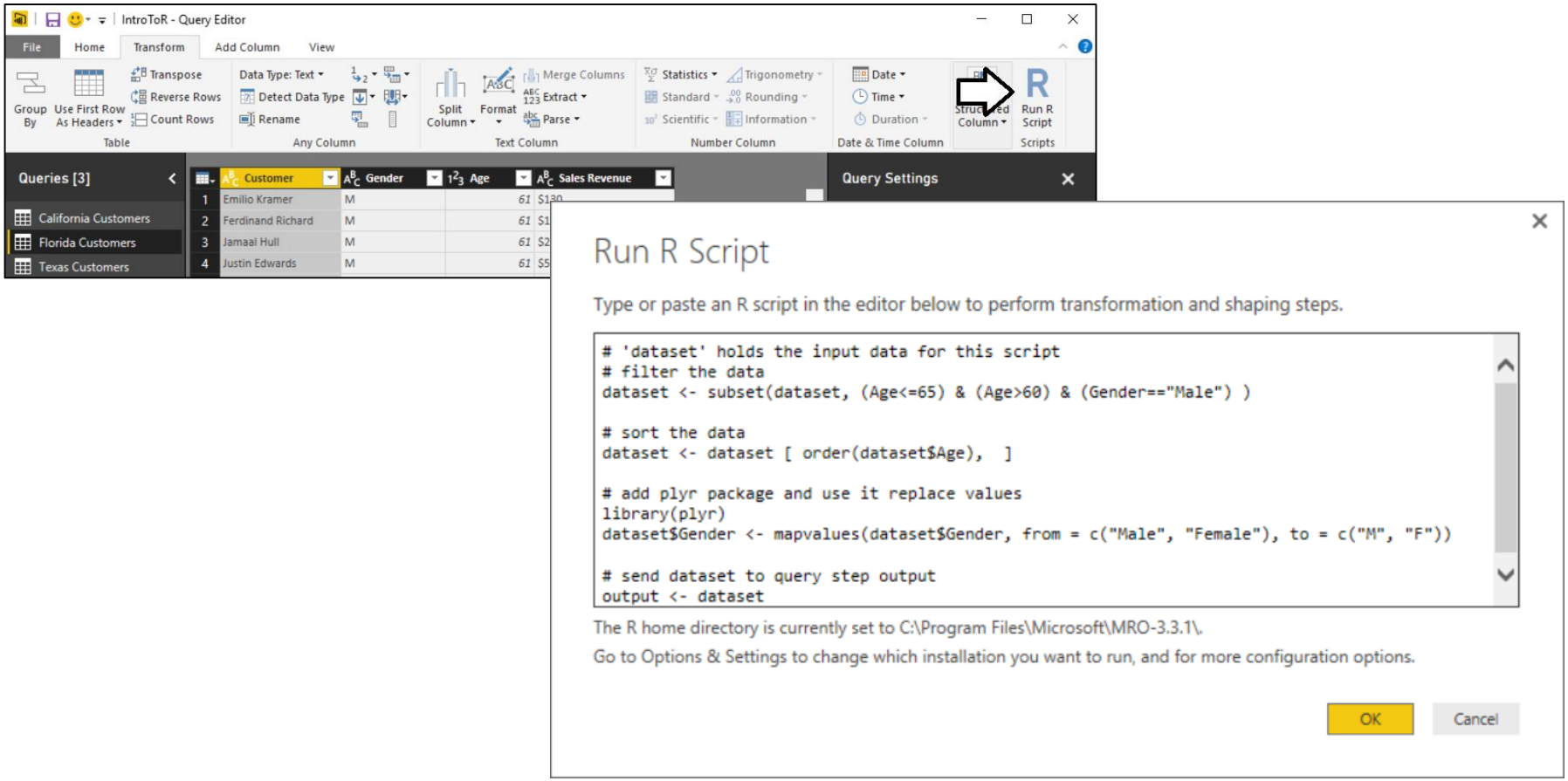
- Create new query based on R script
  - Copy and paste code from RStudio into PBIDT



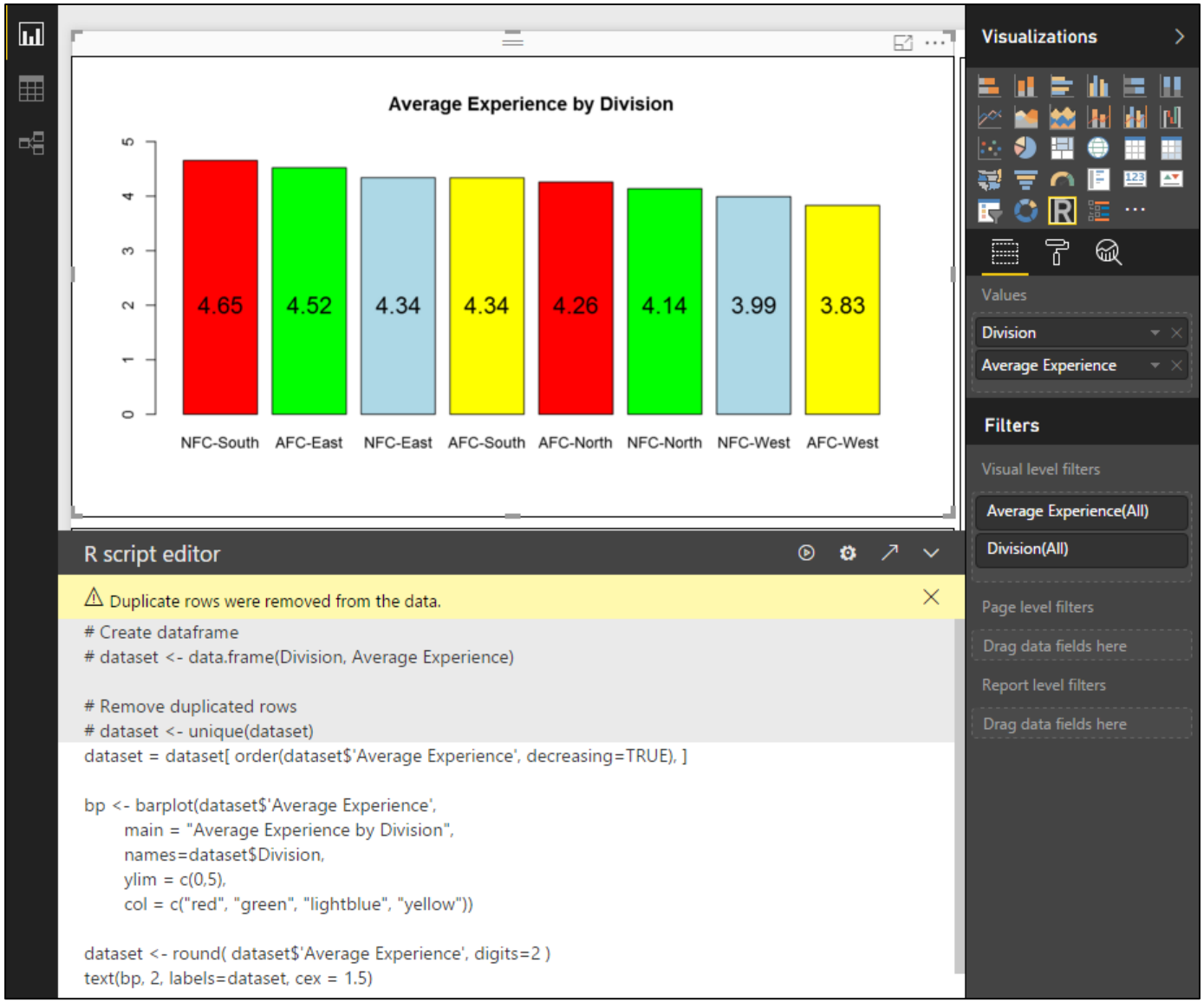


Using R Code as an Applied Query Step

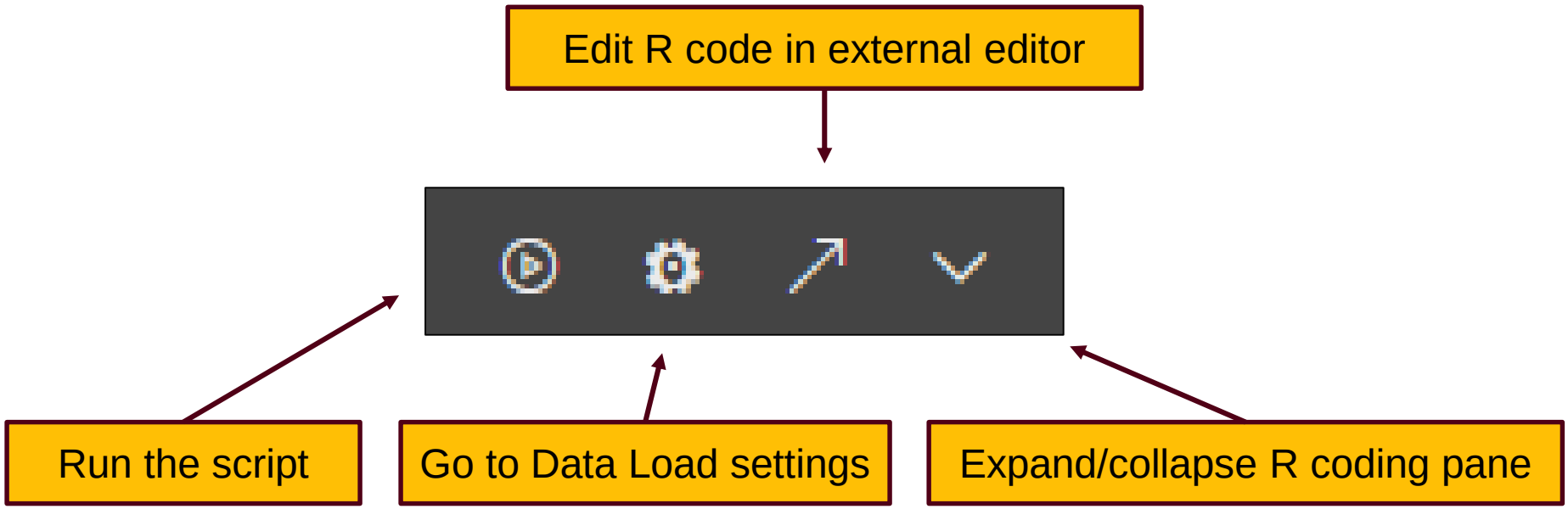
- Add new Run R Script step to query
  - Use R code and R packages to transform data

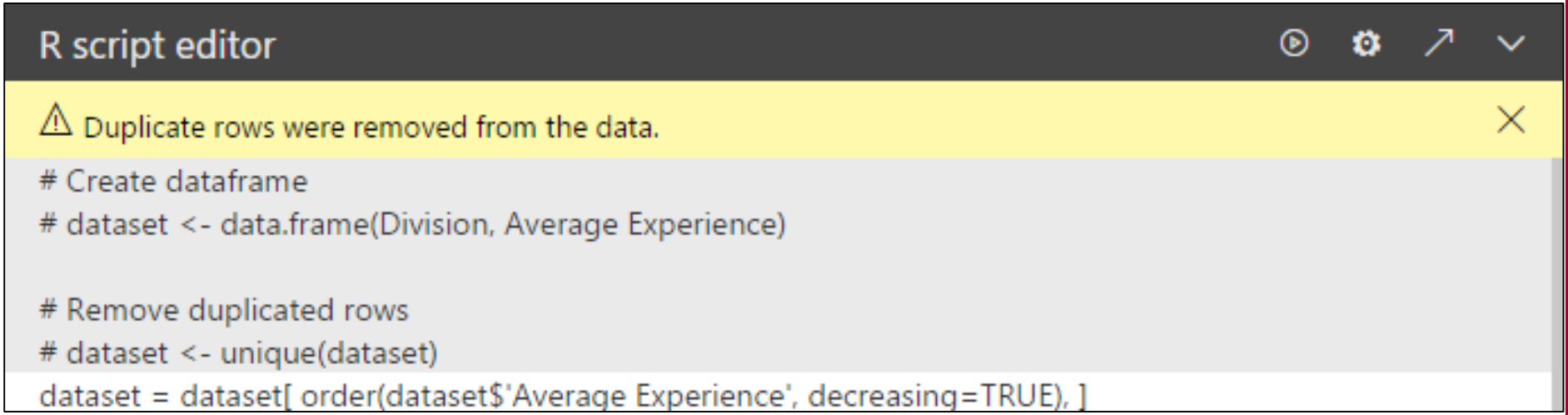


R Visuals in Power BI



New R Features in September Release





R Integration Limitations with Power BI

- Power BI Desktop R Limitations
  - Only data frames are imported
  - Complex columns and Vector columns are not imported
  - Values that are N/A are translated to NULL values
  - Prompting for user input halts script
  - R visual data for plotting is limited to 150,000 rows
  - R visual calculation times out with error after 5 minutes
  - R visual is not interactive – no highlighting support
  - Plots can only be displayed to R default display device

Summary

- ✓ Understanding R as an Analytics Platform
- ✓ Installing Microsoft R Open and RStudio
- ✓ Writing R Code in RStudio
- ✓ Integrating R with Power BI Desktop