Overview & Introduction

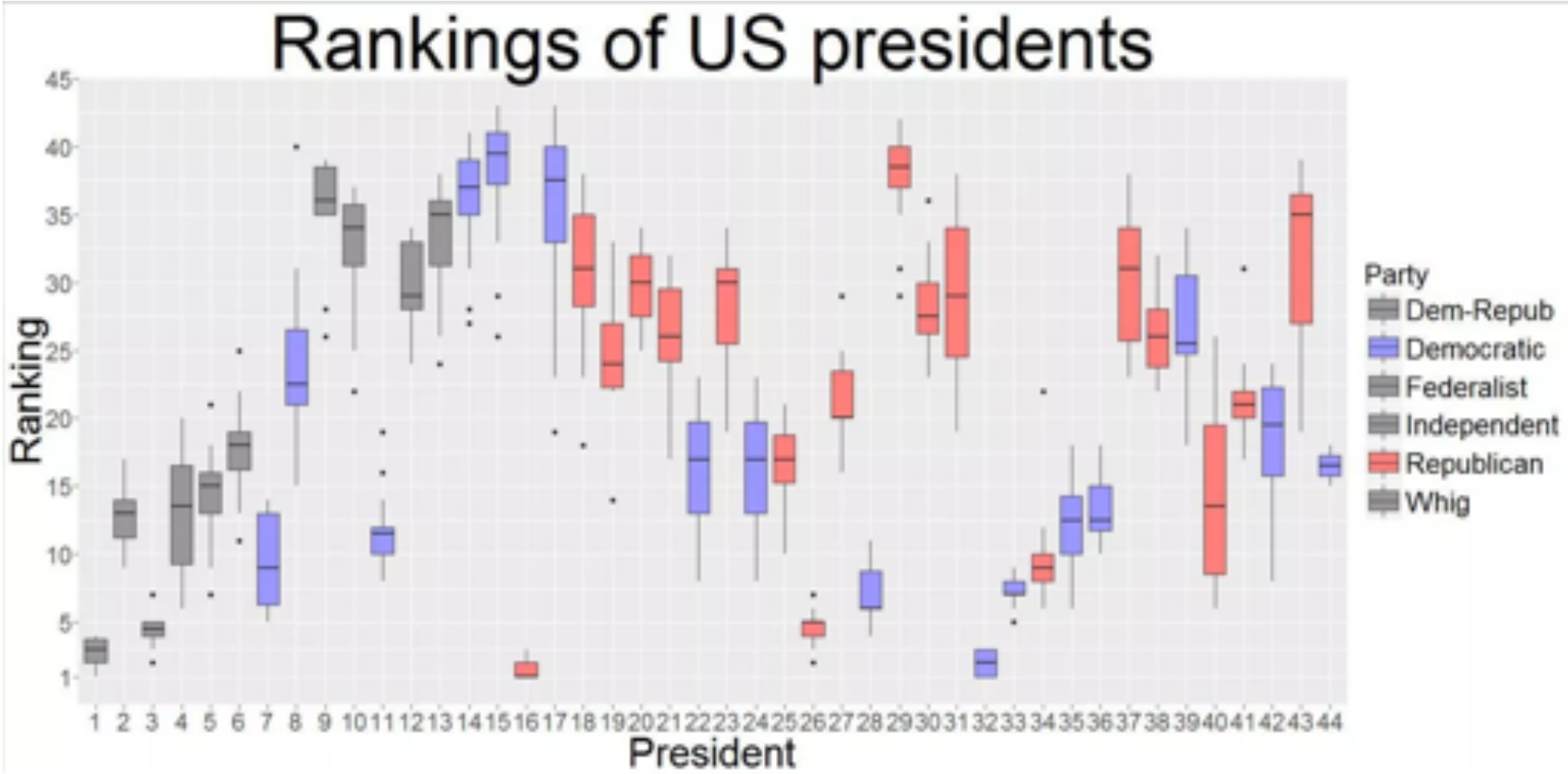
Business Analytics

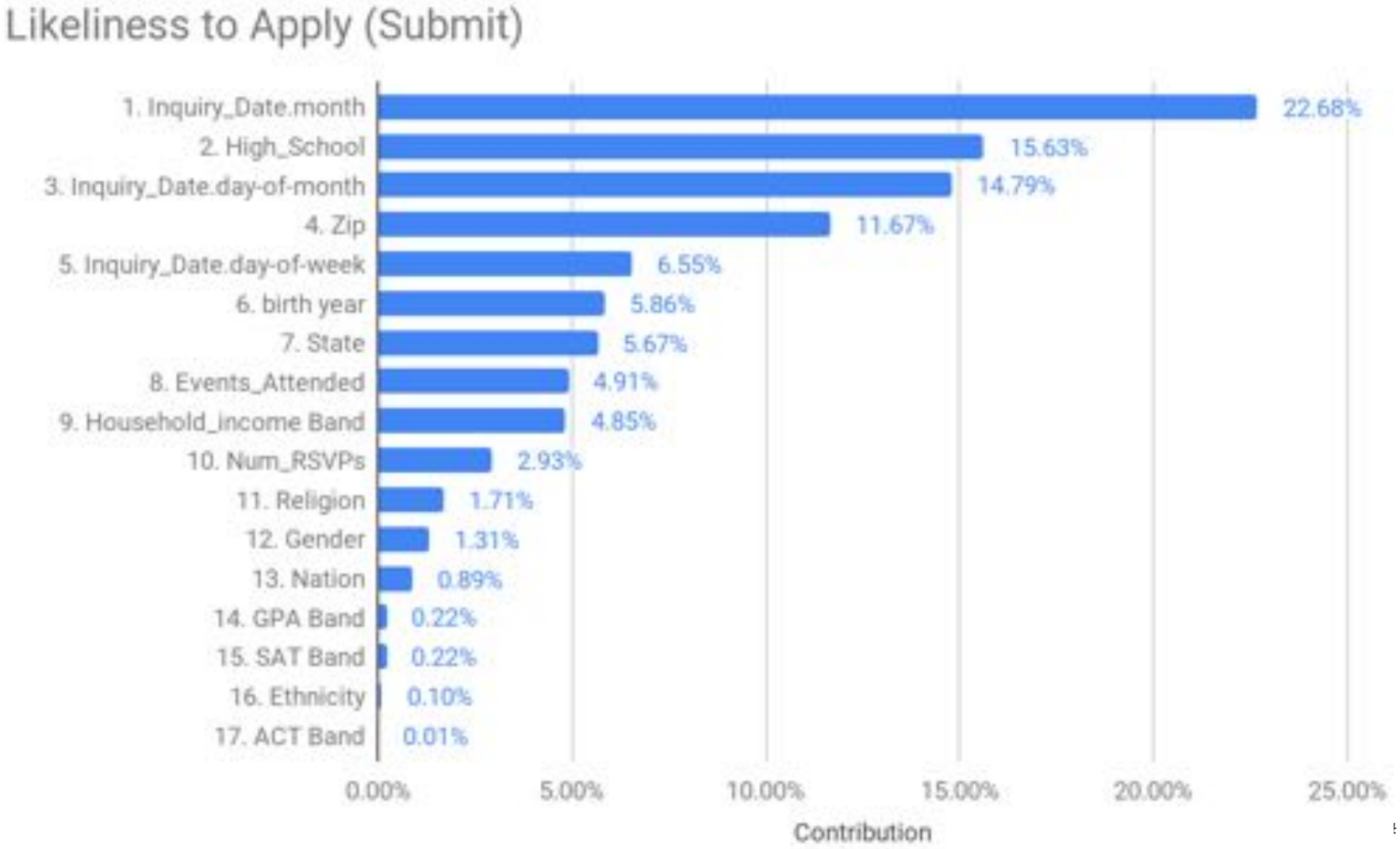
Meet the Instructor

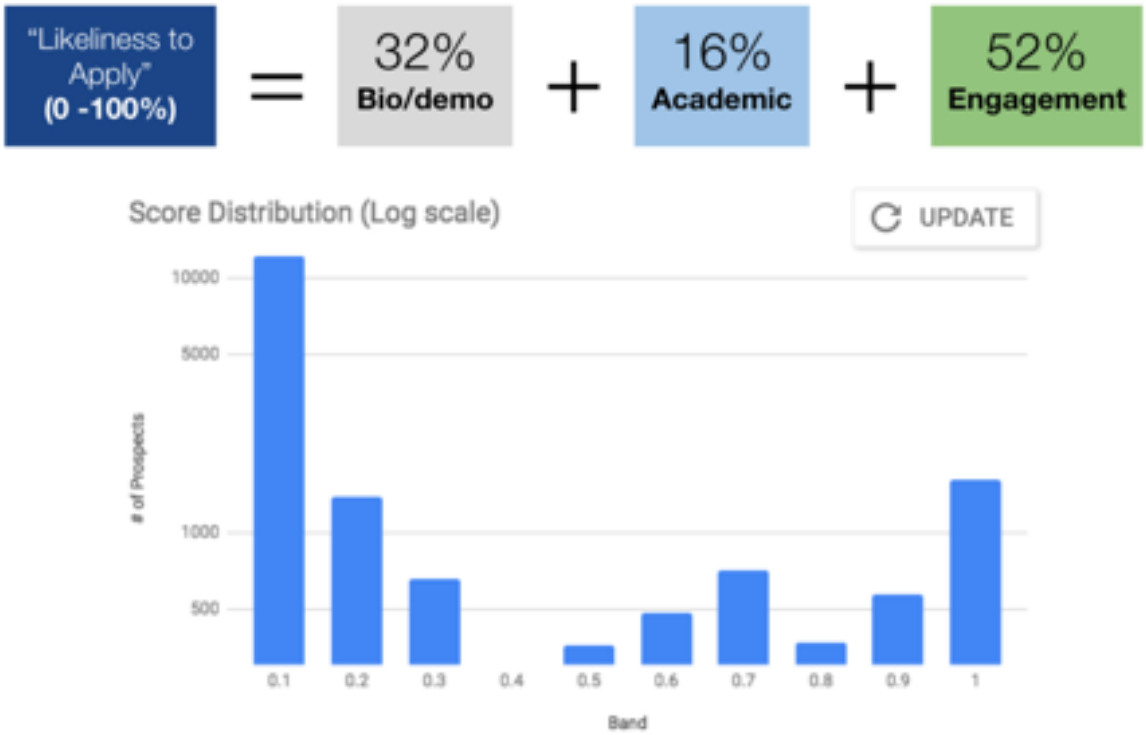
Prof. JeanCarlo Bonilla

Chief Data Scientist & Head of Analytics
Element451
LinkedIn | jb3379@nyu.edu


BUSINESS
ANALYTICS
DATA
BA
PLANNING
PREDICTIVE
FACT-BASED
SPACE
OPTIMIZE
DECISIONS
INTELLIGENCE
MODELING
ACROSS
SKILLS
FOCUSES
DECISION
CONSUMER
OLAP
MAKING
ALSO
COMPETE
BASED
ENTERPRISE
USE
DRIVE
METHODS
HAPPEN
STORAGE
PERFORMANCE
ANSWER
PAST
TYPE
CUSTOMER
QUALITY
TOOLS
STATISTICAL
QUERYING
REPORTING
QUESTIONS
SUFFICIENT
OPTIMIZATION
MAY
REFERS
CHARACTERISTICS
PRACTICES
CONTINUOUS
CAPABILITIES
WORDS
CONSIDERED
APPLICATIONS
EXPLORATION
ORGANIZATION
DIFFERENT
TECHNIQUES
NUMBER
COMPLEX
NECESSARY
TOOL
ITERATIVE
SALES
LAST
UNITS
AFTER-THE-FACT
SPECIFIC
ADVOCATE
MUST
PROCESSES
EXAMINING
DEVELOPING
VOLUMES
HAPPENED
NOW
SENIOR
MEANS
CONTRAST
NEEDED
PROBLEM
THIS
USED
USING
APPEAL
FUNCTIONS
EXTENSIVE
INSIGHT
METRICS
FOCUSED
INTEGRATING


Rankings of US presidents
Ranking
President
Party
Dem-Repub
Democratic
Federalist
Independent
Republican
Whig


Likeliness to Apply (Submit)
1. Inquiry_Date.month
2. High_School
3. Inquiry_Date.day-of-month
4. Zip
5. Inquiry_Date.day-of-week
6. birth year
7. State
8. Events_Attended
9. Household_income Band
10. Num_RSVPs
11. Religion
12. Gender
13. Nation
14. GPA Band
15. SAT Band
16. Ethnicity
17. ACT Band
22.68%
15.63%
14.79%
11.67%
6.55%
5.86%
5.67%
4.91%
4.85%
2.93%
1.71%
1.31%
0.89%
0.22%
0.22%
0.10%
0.01%
0.00%
5.00%
10.00%
15.00%
20.00%
25.00%
Contribution



"Likeliness to Apply"	Prospect → APP
High probability (6-10)	18%
Medium probability (3-5)	10%
Low probability (0-2)	72%

Course Description

Business analytics is a set of **data analysis and modeling techniques** for understanding business situations and **improving business decisions**. This course provides an **introduction** to business analytics concepts, methods and tools with concrete examples **from industry applications**.

Throughout the course, we explore the **challenges** that can arise in implementing analytical approaches within an organization. The course emphasizes that business analytics is not a theoretical discipline: these techniques are only interesting and important to the extent that they can be used to **provide real insights and improve the speed, reliability, and quality of decisions.**

- In the first part of the course, we will focus on descriptive analytics and exploratory data analysis concepts with a refresher on basic probability and statistics.
- In the second part, we will cover principles, techniques, and techniques for spatial data, time series, and text as data.
- The final part of the course will introduce a project that links business impact and modern data analytics techniques for managerial decision making in functional areas, including finance, marketing, and operations.

Pre-Requisites: Probability, Statistics, Linear Algebra, & Excel

Required Materials

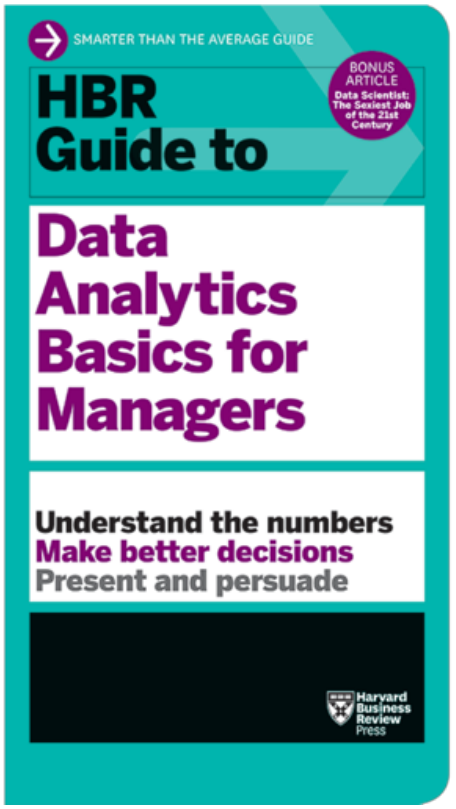
Required Textbook

HBR Guide to Data Analytics for Basic Managers. 2018

Suggested Textbook for R

R for Data Analysis in easy steps - R Programming essentials. Mike McGrath, 2018

R: Data Mining and Business Analytics with R, Johannes Ledolter, 1st Edition



Letter Grade	100% Scale	Grade Point Value
A	100-95	4.0
A-	94-90	3.7
B+	89-85	3.3
B	84-80	3.0
B-	79-75	2.7
C+	74-70	2.3
C	69-65	2.0
F	64-0	0

Lette	
A	
A-	
B+	
B	
B-	
C+	
C	
F	

NO GRADE NEGOTIATION IN THIS COURSE!

Lette	
A	
A-	
B+	
B	
B-	
C+	
C	
F	

I've designed this course for a B!

Statement of Academic Integrity

Students are expected to follow standards of excellence set forth by New York University. Such standards include respect, honesty, and responsibility. This class does not tolerate violations to academic integrity including:

- Plagiarism
- Cheating on an exam
- Submitting your own work toward requirements in more than one course without prior approval from the instructor
- Collaborating with other students for work expected to be completed individually
- Giving your work to another student to submit as his/her own
- Purchasing or using solutions or work online or from a commercial firm and presenting it as your own work

Important Tools

NYUClasses: all announcements and class-related documents (supplemental and suggested readings, discussion questions, etc.) will be posted there. http://classes.nyu.edu/

SLACK: class chat room at nyuba.slack.com for all coordination and messaging.

***Github*.** Lecture, data, and code will be posted in the following github repository https://github.com/jcbonilla/BusinessAnalytics

Grading Policy

- **Weekly Assignments - 30%**
  - Mostly data analysis and programming assignments. Some assignments will include theoretical aspects to make sure students understand the important mathematical concepts in data analytics.
- **Readings in Analytics & Class participation- 10%**
- **Projects – 45%**
  - This is the capstone experience of the course where students will form groups consisting of between 3 and 4 people depending upon the size of class. Teams will build a project using a publicly accessible datasets. They will motivate the business problem, do enough explanatory analysis and generate data driven strategic insight. For each project, teams will provide a brief presentation on the project scope, data and methods utilized, findings, and business implications.
  - This courses builds around 3 unique projects:
    - Project 1: Descriptive Analytics - 10%
    - Project 2: Predictive Analytics - 15%
    - Project 3: Final project - 20%
- **Exam – 15%**
  - Two exams on covering theory and applications
    - Midterm Exam: 5%
    - Final Exam: 10%

Readings in Analytics

- Weekly reading analysis and team presentation on HBR cases
- 3 students assigned on a weekly basis to a chapter.
- 1 student is selected randomly to present at the end of class
- Presentations. I want a critical review and debrief of the article
  - 3 slides (3-5min)
  - Think about the problem (what is the article trying to solve?), the solution (How is analytics solving the problem). Does it make sense do you?
  - Be critical and provide you own opinion of the article

Course Schedule & Grading

See NYUClasses

Spring 2019 Course Schedule*

Module	Dates	Topics
Overview & Context	*Week 1* ***Jan 29***	Competing on Analytics in today's Business Landscape. The data lifecycle. Analytics vs. Intelligence vs. Data Science.
Descriptive Analytics	*Week 2 - 5* ***Feb 5 - 26***	Using R scripts, we will review concepts in descriptive statistics such as measures of central tendency, measures of dispersion, and measures of association between two variables. In this module we will cover the principles of hypothesis testing, correlations analysis, and statistical significance. In addition, this will be the first session on effective data visualization In addition, this module covers fundamentals of exploratory data analytics and Tableau. We will cover the approach and philosophy of Exploratory Data Analysis (EDA) for understanding business situations and improving business decisions
	Week 6 ***March 5***	**Project 1 - Dashboard** Project presentations. Teams will have 10 to showcase a project using descriptive analytics
Exam	*Week 6* ***March 5***	**Midterm Exam**

Predictive Analytics	*Week 7-10* ***Mar 12- Apr 9***	This module covers the basics of regression analysis. Our emphasis will be on applications and interpretation of the results for making real life business/policy decisions and the mathematical and statistical properties of the techniques used to produce these results. In order to provide a broad intuition of the concepts and methods, we will use data and examples from marketing decision making such as segmentation, estimating market potential and forecasting demand, etc. We will continue covering regression analysis including non-linear transformations and dummy variables. In addition to advanced regression, we will cover prediction and classification trees. Finally, we will learn the principles of model selection, model performance, and operationalizing models.
	Week 11 ***Apr 16***	**Project 2 - Prediction** Project presentations. Teams will have 10min to showcase a project using predictive analytics

Special Topics in Analytics **Text Analytics ETL, MLaaS**	*Week 12-13* ***Apr23 - 30***	Using text as data presents an opportunity to move words into document-term-matrices that can be analyzed via statistical analysis. In this lesson we will cover text processing, analysis, and interpretations of summary statistics of a corpus In addition, we will cover ETL and MLaaS applications
Final Project	*Week 14-15* ***May 7- 14***	**Project Consultations** In-class prep to review scope and preliminary findings of the final project **Project 2 - Prediction** Project presentations. Teams will have 15min to showcase a project using predictive analytics
Exams	*Week 15* ***May 14***	**Final Exam**

What is Business Analytics?

Business Analytics Definition

Business analytics refers to the application of data analysis and modeling techniques for understanding business situations and improving business decisions.

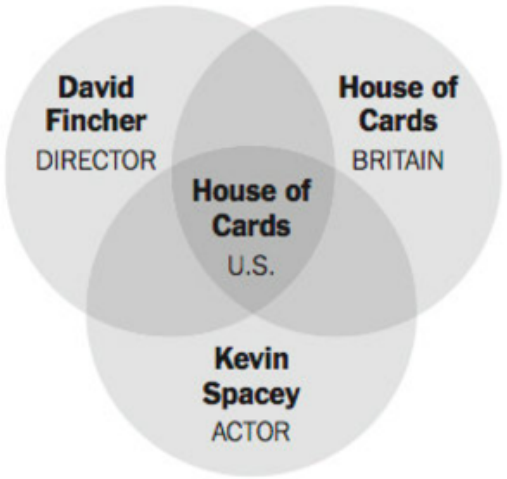
IMPLICATIONS:

- data → past business performance
- methods → statistics + mathematics + computational methods
- business decisions → actionable insight

What analytics examples can you think of?

Circles of Proven Success

Netflix determined that the overlap of these three areas would make "House of Cards" a successful entry into original programming.



THE NEW YORK TIMES


BUSINESS
PLATINUM
Capital One
4000 1234 5678 9010
4000
VALID DATES
00/00 - 00/00/00
SMALL BUSINESS, INC.
LEE M CARDHOLDER
VISA

guests who viewed this item ultimately bought

$199.99

reg: $299.99

Nikon Coolpix L840 16.1MP Digtial Camera...

Nikon

spend $25, get free shipping

(11)

$249.99

Canon PowerShot SX400 IS Digital Camera ...

Canon

spend $25, get free shipping

(5)

see low price

reg: $219.99

Sony DSCH300/B 20MP Digital Camera w...

Sony

spend $25, get free shipping

(25)

$129.99

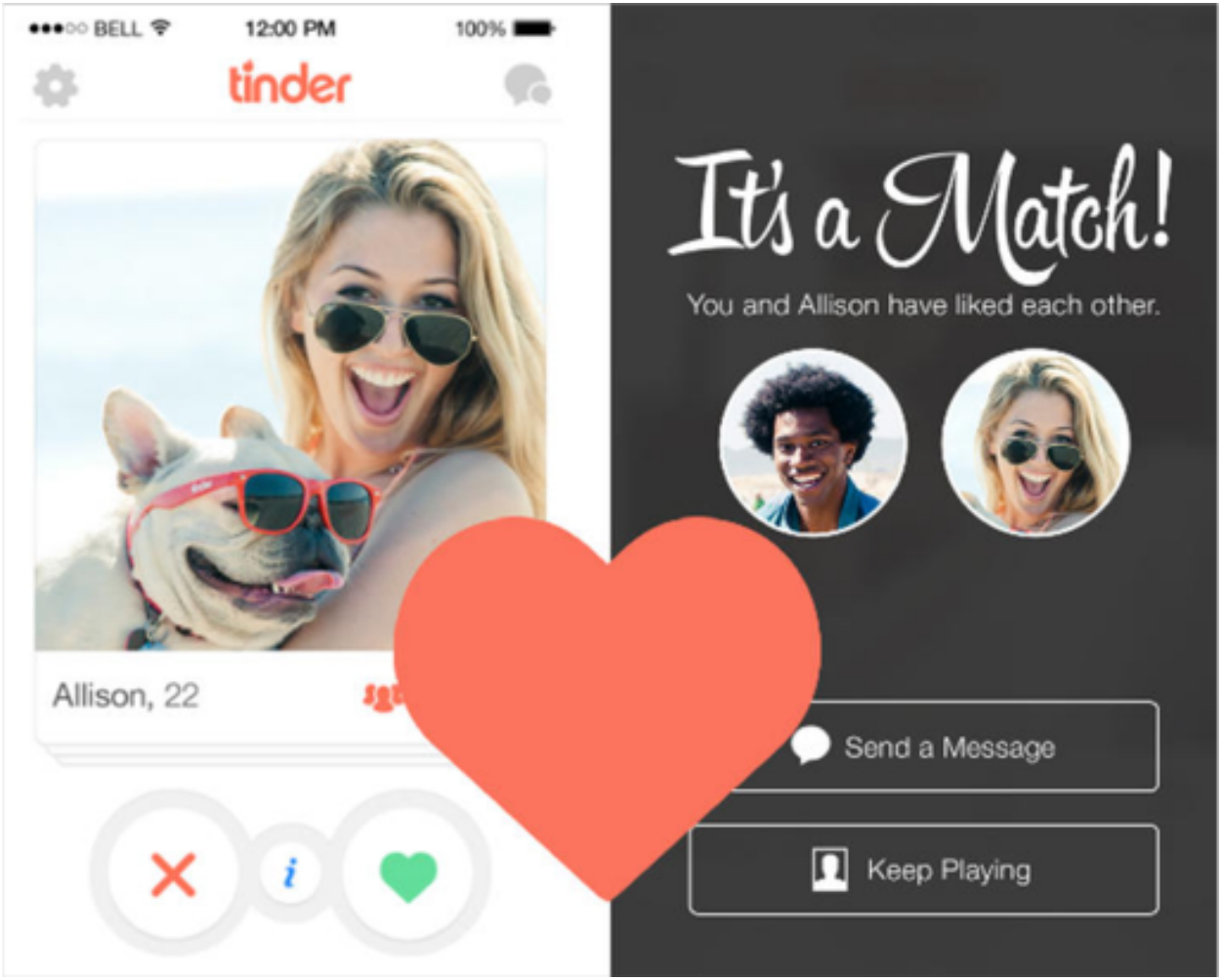
List: $139.99

Sony Cybershot DSCW830 20.1MP Digital Ca...

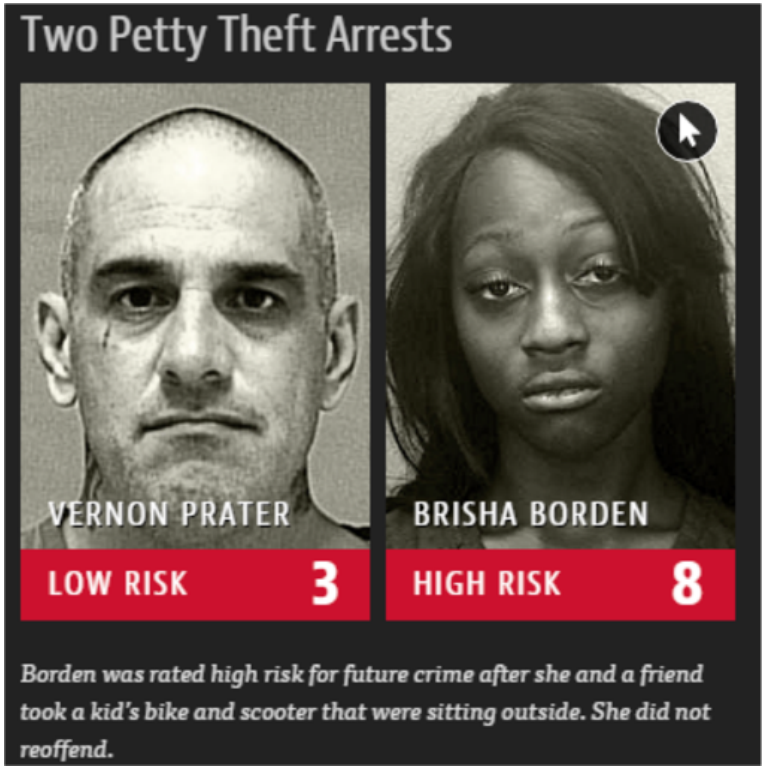
Sony

spend $25, get free shipping

(139)


BELL
12:00 PM
100%
tinder
Allison, 22
It's a Match!
You and Allison have liked each other.
Send a Message
Keep Playing

Analytics can also do harm!



Types of Analytics

Descriptive
Analytics

Descriptive Analytics:
Encompasses the set of techniques that describe what has happened in the past whether that is one minute or one year ago. The vast majority of the statistics we use like sums, averages, percent changes fall into this category.

About 35% of companies surveyed say they do this consistently.

Use Descriptive statistics when you need to understand at an aggregate level what is going on in your company, and when you want to summarize and describe different aspects of your business.

Types of Analytics

Descriptive Analytics

Predictive Analytics

Predictive Analytics:
This type of analytics are about understanding the future and providing estimates about the likelihood of a future outcome. They combine historical data found in ERP, CRM, HR and POS systems to identify patterns and apply models and algorithms to capture relationships between various data sets.

However, less than 1% of companies surveyed have tried this yet.

Use Predictive analysis any time you need to know something about the future or fill in the information that you do not have.

.

Types of Analytics

Prescriptive
Analytics

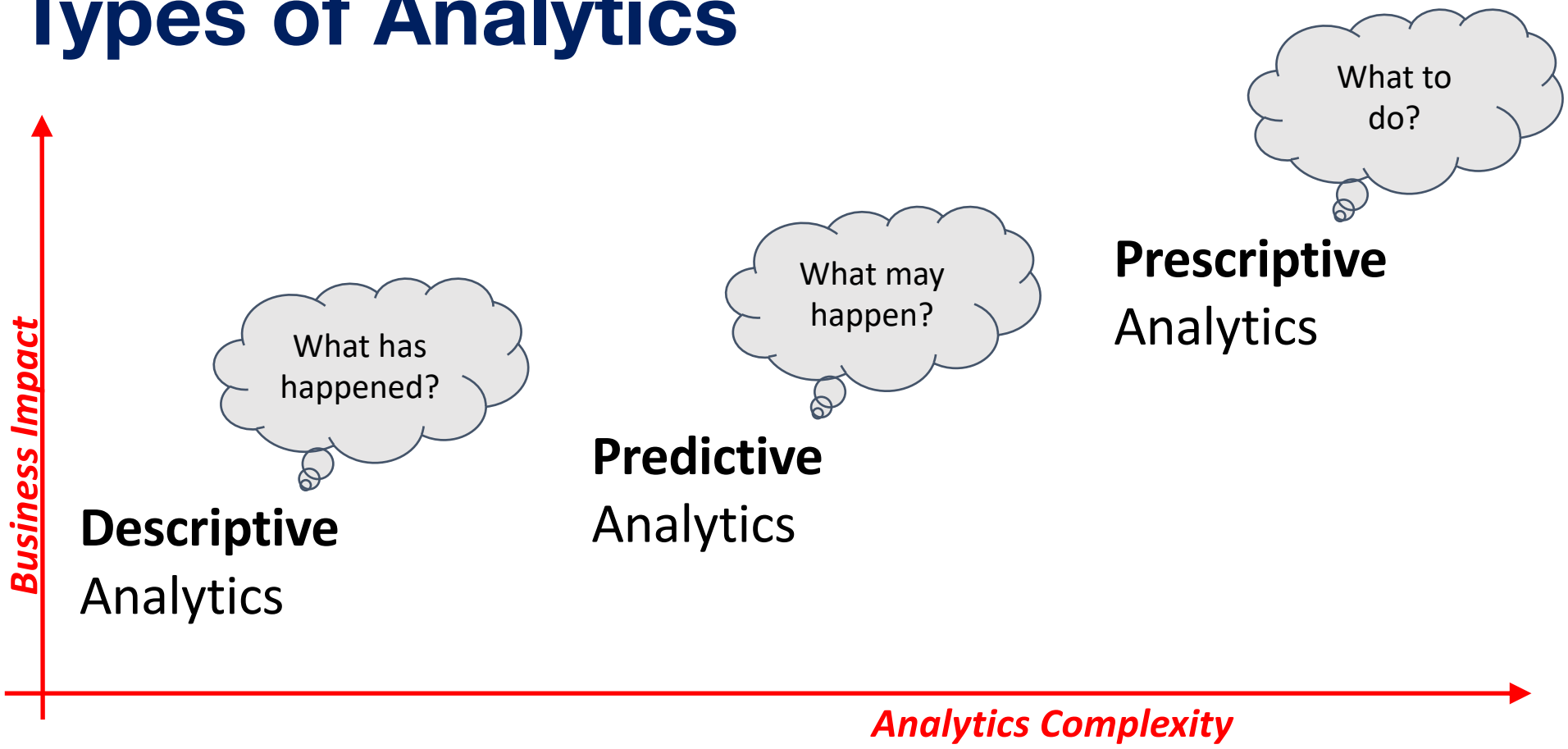
Prescriptive Analytics:
Prescriptive analytics attempt to quantify the effect of future decisions in order to advise on possible outcomes before the decisions are actually made. Prescriptive analytics are relatively complex to administer, and most companies are not yet using them in their daily course of business.

Use prescriptive statistics anytime you need to provide users with advice on what action to take.

Predictive
Analytics

Descriptive
Analytics

Types of Analytics



Analytics vs BI vs Data Science

Business Intelligence (BI): tools and systems to gather, store, access and analyze an organization's raw data. BI actually presents the insights determined by Analytics in reports, dashboards, or interactive visualizations

- querying and reporting tools, dashboards
- traditionally used to determine trends in historical data

Data Science: involves using automated methods to extract knowledge or insights from structured or unstructured data.

- employs techniques and theories drawn from mathematics, statistics, information science, machine learning, AI, and others.

For purposes of this course, *Analytics* is a catch-all term that encompasses both BI and Data Science as well data mining/modeling, and forecasting

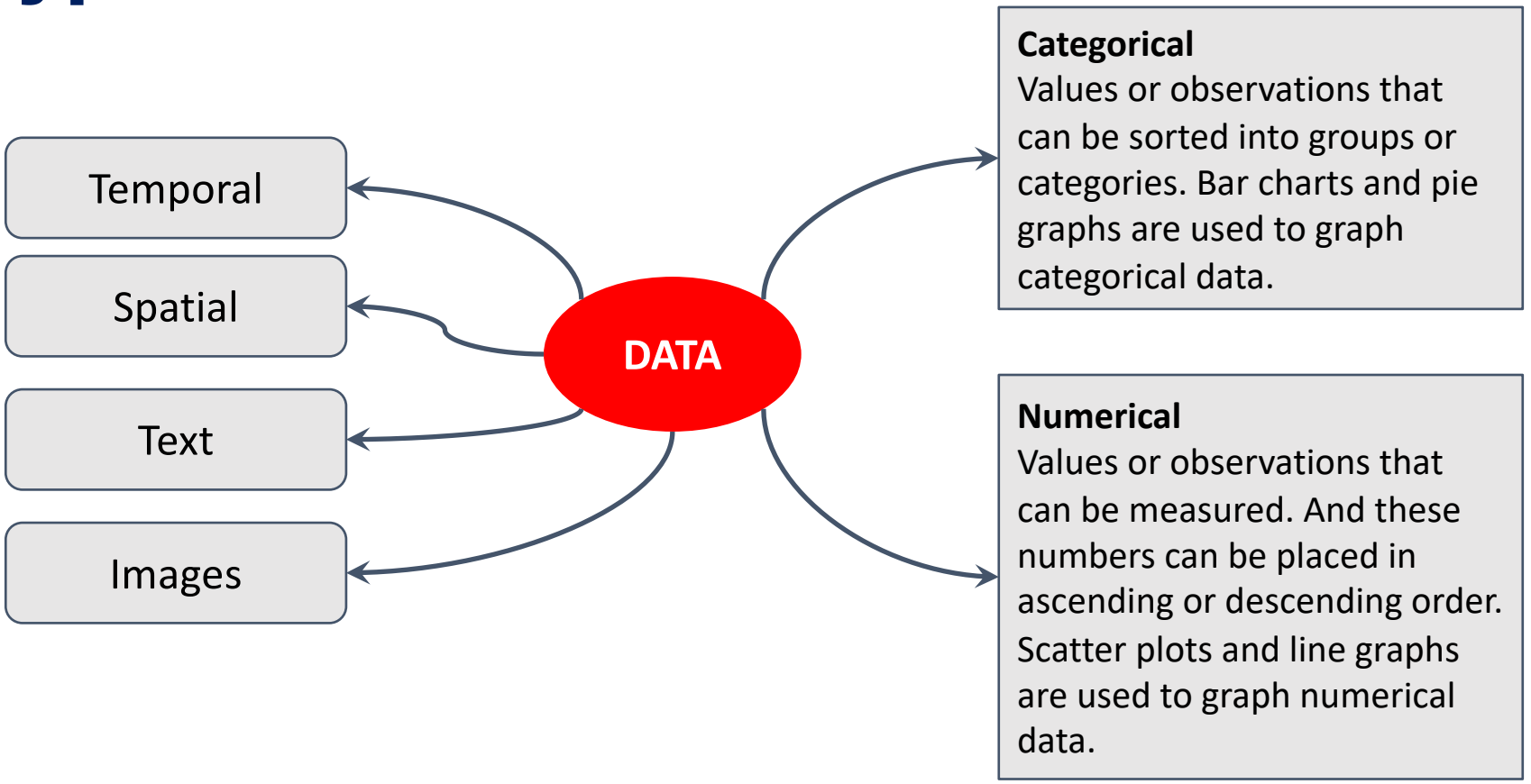
Business Analytics Definition

Business analytics refers to the application of data analysis and modeling techniques for understanding business situations and improving business decisions.

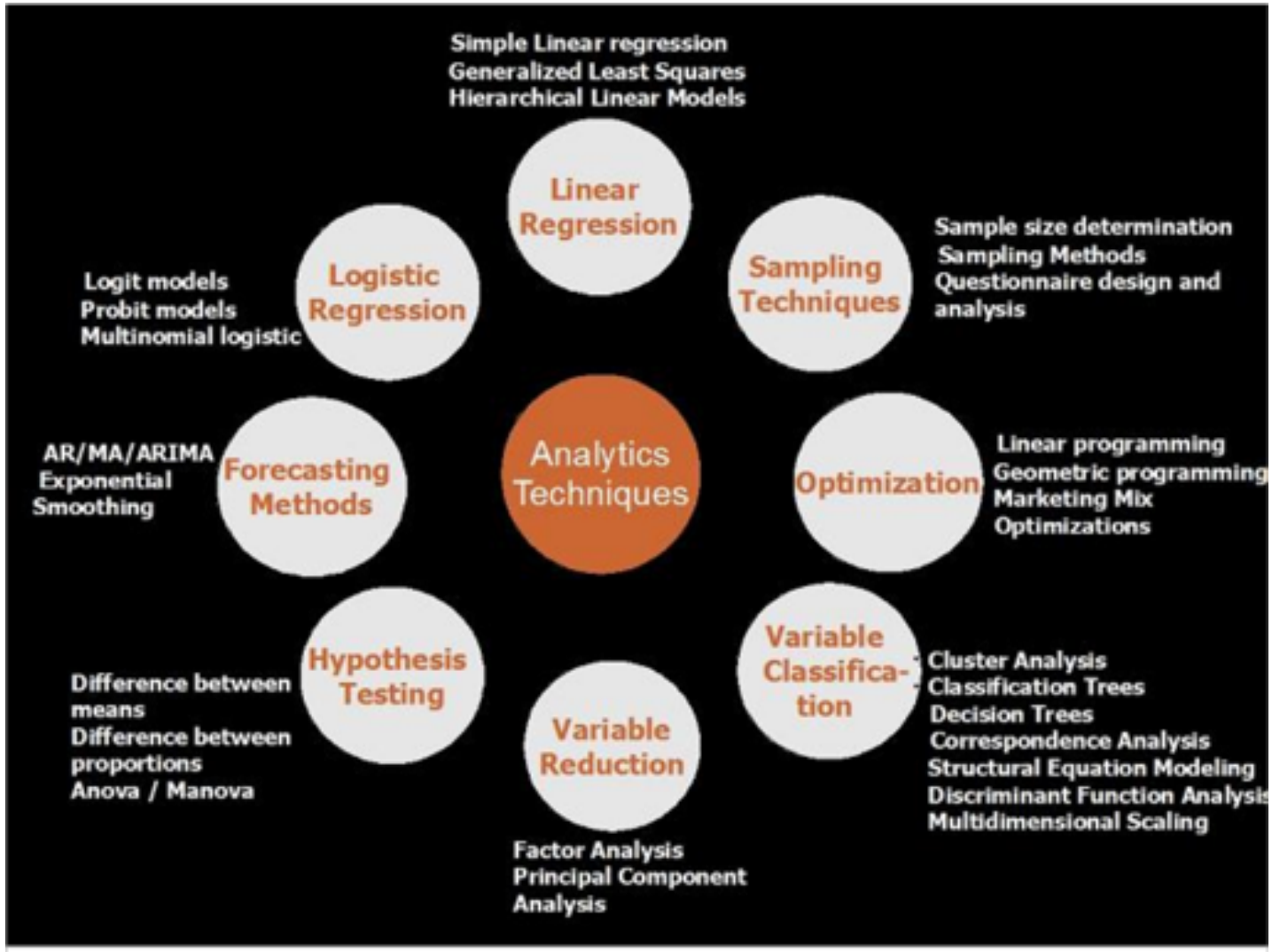
IMPLICATIONS:

- data → past business performance
- methods → statistics + mathematics + computational methods
- business decisions → actionable insight

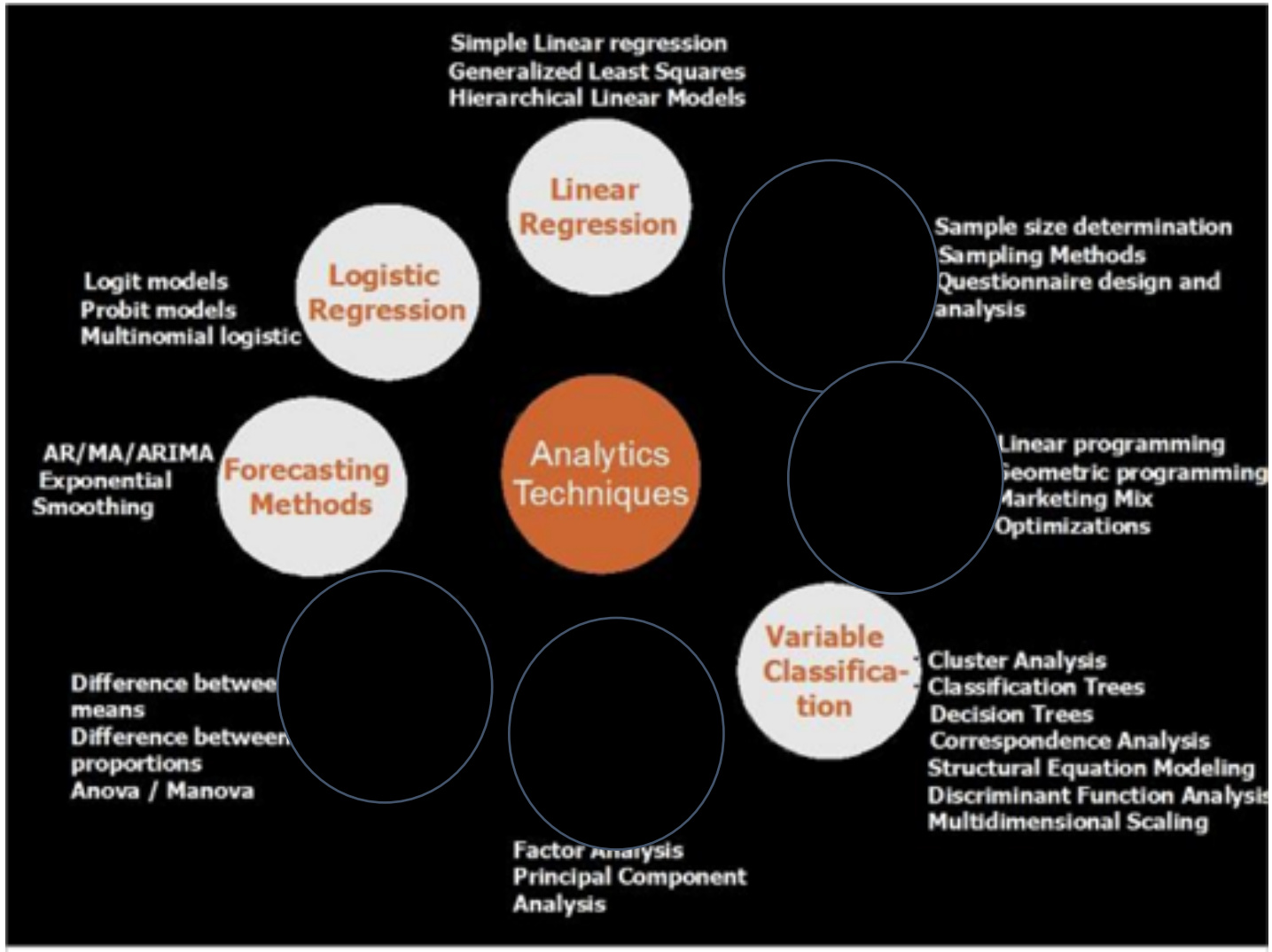
Types of data



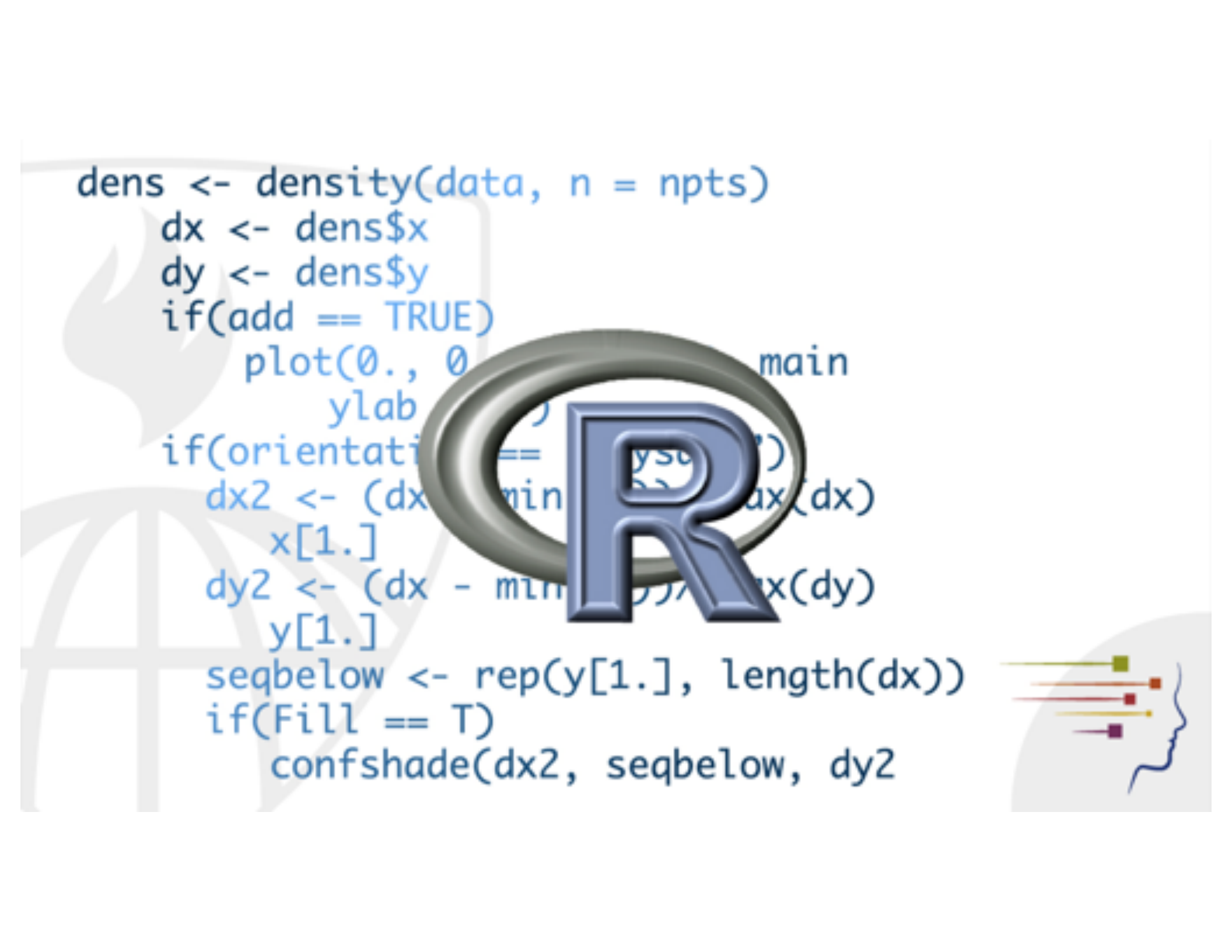
Methods


Simple Linear regression
Generalized Least Squares
Hierarchical Linear Models
Linear Regression
Logistic Regression
Logit models
Probit models
Multinomial logistic
Sampling Techniques
Sample size determination
Sampling Methods
Questionnaire design and analysis
AR/MA/ARIMA
Exponential Smoothing
Forecasting Methods
Analytics Techniques
Optimization
Linear programming
Geometric programming
Marketing Mix Optimizations
Hypothesis Testing
Difference between means
Difference between proportions
Anova / Manova
Variable Reduction
Factor Analysis
Principal Component Analysis
Variable Classifica-tion
Cluster Analysis
Classification Trees
Decision Trees
Correspondence Analysis
Structural Equation Modeling
Discriminant Function Analysis
Multidimensional Scaling

Methods


Simple Linear regression
Generalized Least Squares
Hierarchical Linear Models
Linear Regression
Logistic Regression
Logit models
Probit models
Multinomial logistic
Sample size determination
Sampling Methods
Questionnaire design and analysis
AR/MA/ARIMA
Exponential Smoothing
Forecasting Methods
Analytics Techniques
Linear programming
Geometric programming
Marketing Mix
Optimizations
Difference betwee
means
Difference between
proportions
Anova / Manova
Variable Classifica-
tion
Cluster Analysis
Classification Trees
Decision Trees
Correspondence Analysis
Structural Equation Modeling
Discriminant Function Analysis
Multidimensional Scaling
Factor Analysis
Principal Component
Analysis


planning
outcome
stage
time
strategy
prioritization
approach
model
evaluation
objective
decision
step
goal
cost
estimating
value
perspective
assumption
analysis
process
making
opportunity
choice
resources
criteria
problem
scenario
alternative
action
risk
brainstorming
optimization


dens <- density(data, n = npts)
dx <- dens$x
dy <- dens$y
if(add == TRUE)
plot(0., 0
main
ylab
if(orientati
==
dx2 <- (dx
min
ax(dx)
x[1.]
dy2 <- (dx - min
x(dy)
y[1.]
seqbelow <- rep(y[1.], length(dx))
if(Fill == T)
confshade(dx2, seqbelow, dy2

Introduction to R

- R is a language and environment for statistical computing and graphics. It was developed at Bell Laboratories (formerly AT&T, now Lucent Technologies) by John Chambers and colleagues. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, …) and graphical techniques, and is highly extensible.
- You will need to install **Rstudio** an Integrated development environment (IDE). RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

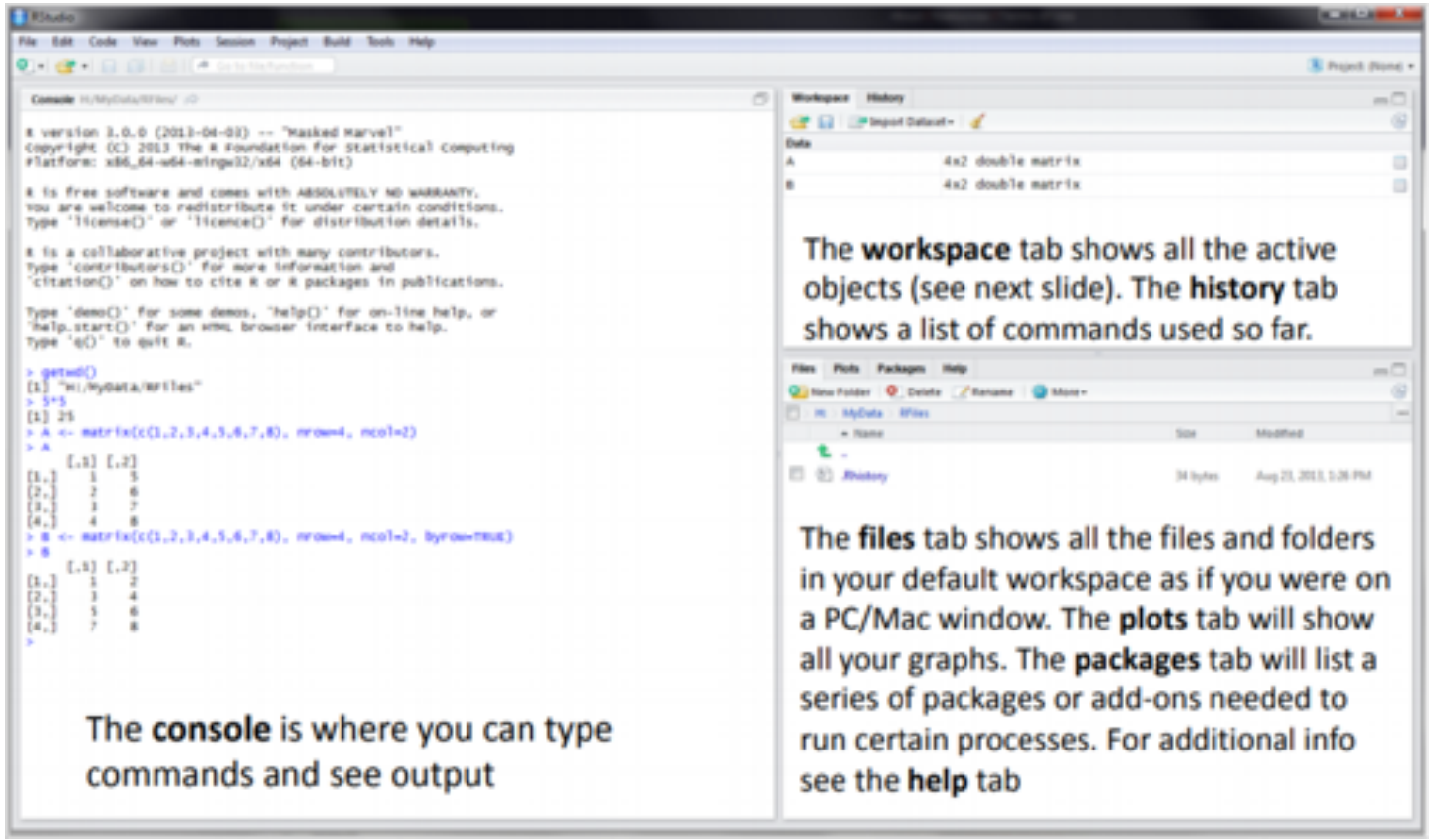
Compiled C vs Interpreted R

- C requires a complete program to run
  - -Program is translated into machine code
  - -Can then be executed repeatedly
- R can run interactively
  - -Statements converted to machine instructions as they are encountered
  - -This is much more flexible, but also slower

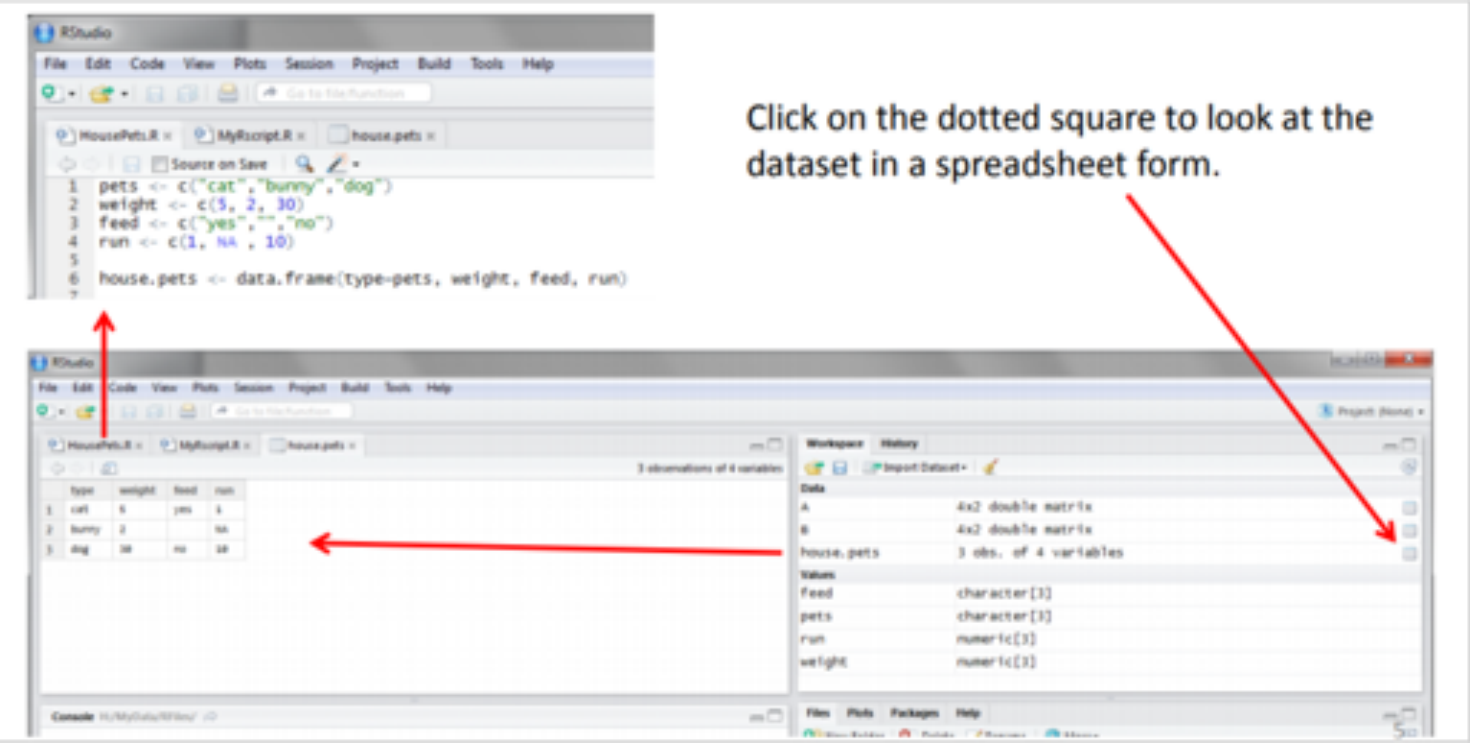
Installing RStudio

- Go to www.rstudio.com and click on the “Download Rstudio” button
- Click on the “Download Rstudio Desktop”
- Click on the version recommended for your system

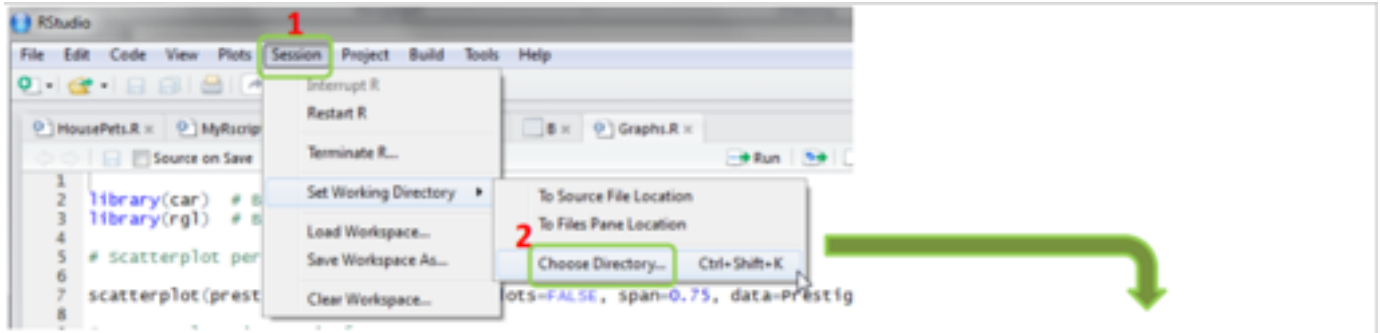
Introduction to RStudio



Introduction to RStudio

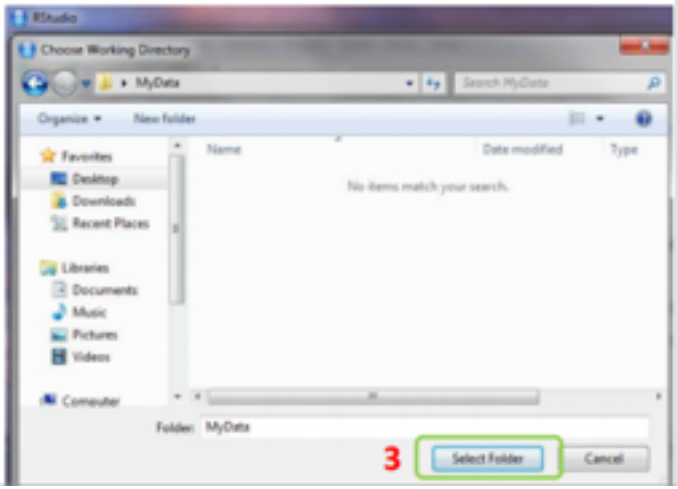


Introduction to RStudio

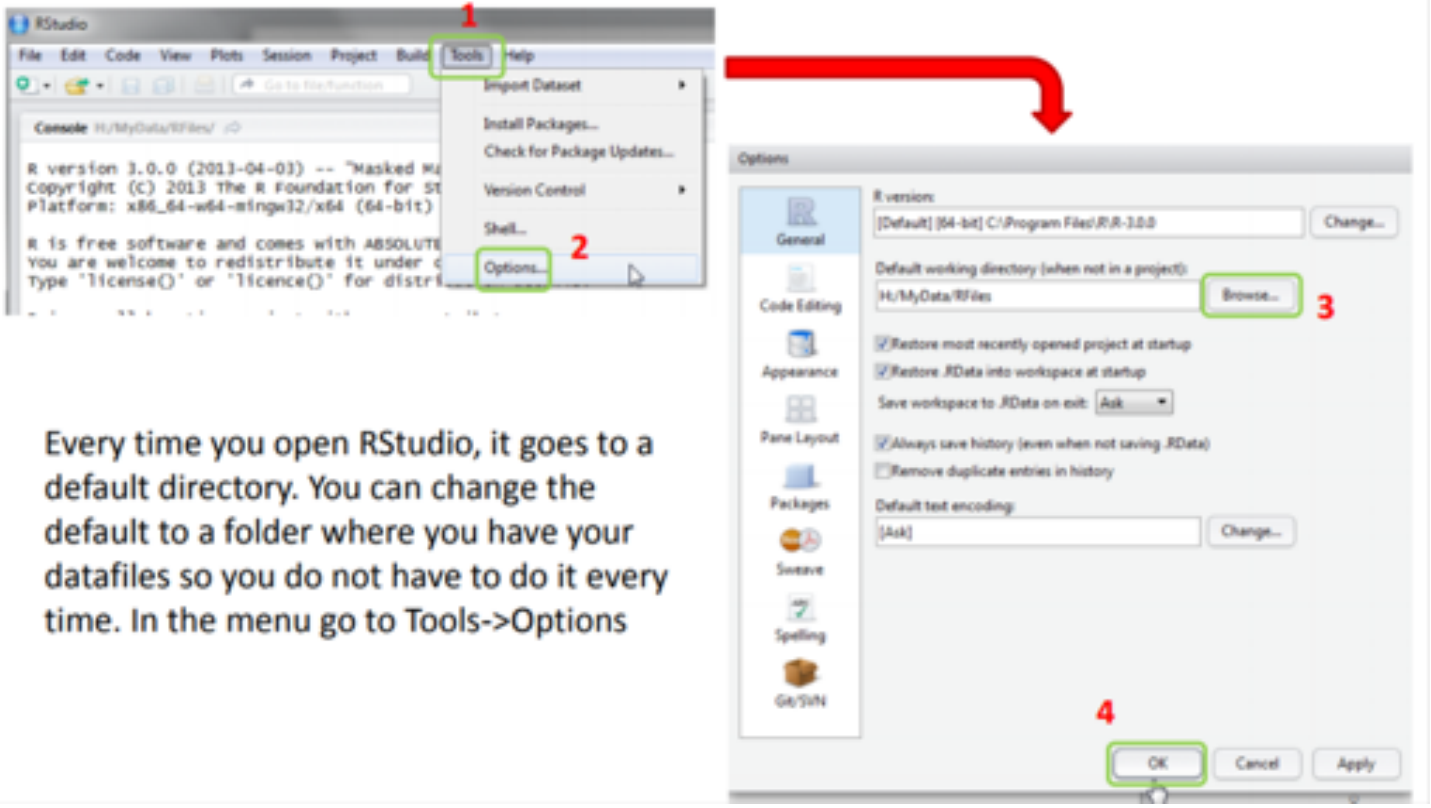


If you have different projects you can change the working directory for that session, see above. Or you can type:

```
# Shows the working directory (wd)

getwd()

# Changes the wd

setwd("C:/myfolder/data")
```



Introduction to RStudio



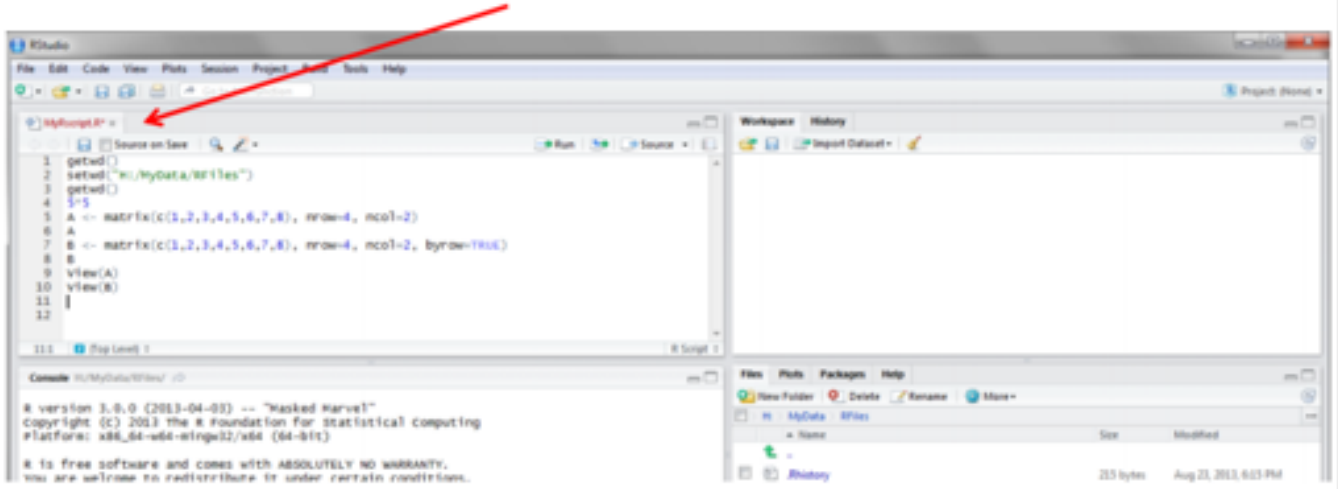
Every time you open RStudio, it goes to a default directory. You can change the default to a folder where you have your datafiles so you do not have to do it every time. In the menu go to Tools->Options

Introduction to RStudio

The usual Rstudio screen has four windows:

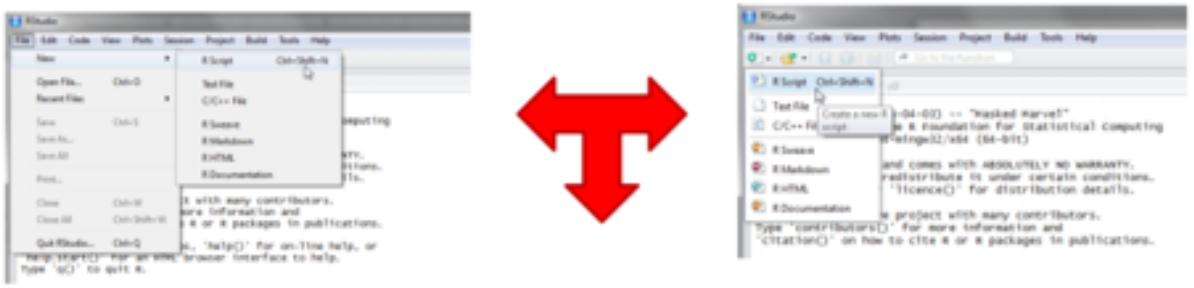
1. Console.
2. Workspace and history.
3. Files, plots, packages and help.
4. The R script(s) and data view.

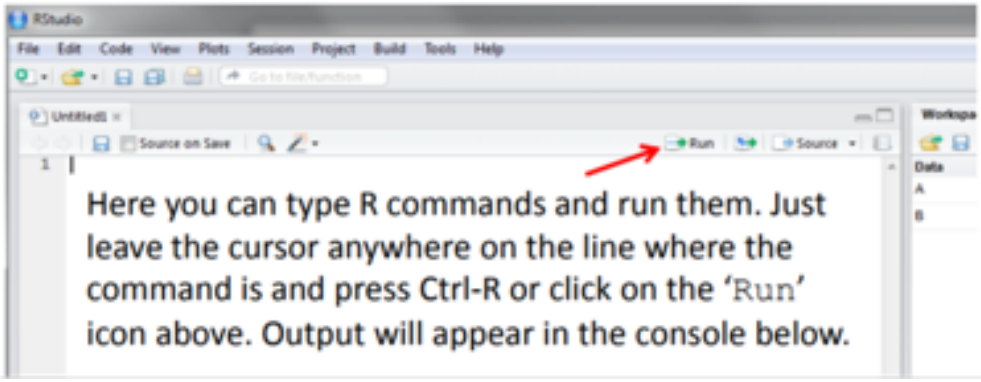
The R script is where you keep a record of your work. For Stata users this would be like the do-file, for SPSS users is like the syntax and for SAS users the SAS program.



Introduction to RStudio

To create a new R script you can either go to `File -> New -> R Script`, or click on the icon with the "+" sign and select "R Script", or simply press Ctrl+Shift+N. Make sure to save the script.





R- Objects and Arithmetic

R stores information and operates on objects. The simplest objects are:

- Scalars
- Vectors
- Matrices
- Lists
- Dataframes for example.

An important feature of R is that it will do different things on different types of objects. For example, type:

```
> 4+6
```

The result should be

```
[1] 10
```

So, R does scalar arithmetic returning the scalar value 10. (In actual fact, R returns a vector of length 1 - hence the [1] denoting first element of the vector.

We can assign objects values for subsequent use. For example:

```
x<-6 y<-4 z<-x+y
```

R- Objects and Arithmetic

- Logical Values
  - 3<4
  - 2+6 == 5
  - T and F are shorthand for TRUE and FALSE
- Functions
  - sum(), sqrt(), mean(), min()
  - c()
  - seq(), rep()
  - help()
  - example()
- Vectors
  - R can perform maths on vectors same as it does on variables.

R- Objects and Arithmetic

Data Types

- There are several R data types that are of frequent occurrence in routine R calculations
  - numeric, integer, logical, character
- *class( )* is used to determine the datatype of given object.
- You can also convert/transform classes as such
  - as.numeric, as.Date, as.Character

Loading Data (usually CSV files)

read.csv( ) is used for this purpose

Methods 1: loading data from .csv file

Methods 2: loading data from a working directory

Methods 3: from GitHub (via ”url”)

R- Objects and Arithmetic

Working with Data Frames

- Once we load the CSV file it time to explore the data. The structure this file takes in R is called data frame. It's similar to matrices
- We use *name( ). class( ), dim( ), str( ), tail( ) , head( )to* get the very basic information about the loaded file. these functions are used for getting an overall idea of data frames content.
  - head(), tail() Returns the first or last parts of data frame
  - str() Compactly display the internal structure of an R object.

- Various arithmetic formulas can directly be applied on the file such as *summary( )*

R as a Calculator

```
> 1 + 1     # Simple Arithmetic
[1] 2

> 2 + 3 * 4  # Operator precedence
[1] 14

> 3 ^ 2     # Exponentiation
[1] 9

> exp(1)   # Basic mathematical functions are available
[1] 2.718282

> sqrt(10)
[1] 3.162278

> pi  # The constant pi is predefined
[1] 3.141593

> 2*pi*6378   # Circumference of earth at equator (in km)
[1] 40074.16
```

R as a Smart Calculator

```
> x <- 1  # Can define variables
> y <- 3  # using "<-" operator to set values
> z <- 4
> x * y * z
[1] 12

> X * Y * Z   # Variable names are case sensitive
Error: Object "X" not found

> This.Year <- 2004  # Variable names can include period
> This.Year
[1] 2004
```

Defining Vectors

```
> rep(1,10)          # repeats the number 1, 10 times
[1] 1 1 1 1 1 1 1 1 1 1

> seq(2,6)           # sequence of integers between 2 and 6
[1] 2 3 4 5 6        # equivalent to 2:6

> seq(4,20,by=4)  # Every 4th integer between 4 and 20
[1] 4 8 12 16 20

> x <- c(2,0,0,4)   # Creates vector with elements 2,0,0,4
> y <- c(1,9,9,9)
> x + y              # Sums elements of two vectors
[1] 3 9 9 13

> x * 4              # Multiplies elements
[1] 8 0 0 16

> sqrt(x)            # Function applies to each element
[1] 1.41 0.00 0.00 2.00       # Returns vector
```

Accessing Vector Elements

```
> x <- c(2,0,0,4)

> x[1]     # Select the first element, equivalent to x[c(1)]
[1] 2

> x[-1]    # Exclude the first element
[1] 0 0 4

> x[1] <- 3 ; x
[1] 3 0 0 4

> y < 9    # Compares each element, returns result as vector
[1] TRUE FALSE FALSE FALSE

> y[4] = 1
> y < 9
[1] TRUE FALSE FALSE TRUE
```

Loading file

```
#Methods 1: loading data from .csv file:

data<- read.csv("~/Google Drive/_NYU GDrive/Teaching/Business Analytics/BA Data/zagat.csv",
header=TRUE,
                stringsAsFactors=FALSE)   #direct method

#Methods 2: loading data from a working directory
getwd()   # display active directory
setwd("~/Google Drive/_NYU GDrive/Teaching/Business Analytics/BA Data")
data<- read.csv("zagat.csv", header=TRUE,stringsAsFactors=FALSE)

#Methods 3: from GitHub
url<-"https://raw.githubusercontent.com/jcbonilla/BusinessAnalytics/master/BAData/zagat.CSV"
data<-read.csv(url, header=TRUE,stringsAsFactors=FALSE)
```

Exploring File

```
>names(data)
>dim(data)
>class(data)
>data
>data[1:4]              # brackets [ ] allow indexing, columns 1-4
>data[1:10,1:3]         # displays 10 rows and 3 columns
>data$Price             # displays values for column "Price"
>data$Price[1:10]
>price<-data$Price      # dollar symbol $ is used to invoke a vector in a matrix
```

Subset

subset() Return subsets of vectors, matrices or data frames which meet conditions.

example:

>subset(parent dataset, condition)

>subset(class_results, class_results$maths > 50)

>subset(class_results, maths > 50)

>subset(class_results, maths > 50 & physics < 70)

R Prep: