Drawing and Data Visualization

Dr. Abu Nowshed Chy

Department of Computer Science and Engineering

University of Chittagong

Dec 15, 2024

Faculty Profile

Data Visualization

Data visualization is using visual elements and makes it easier to analyze data, find patterns in data, and report insights gleaned from data.

The Need to Visualize Data

Use a picture. It's worth a thousand words.

-Tess Flanders

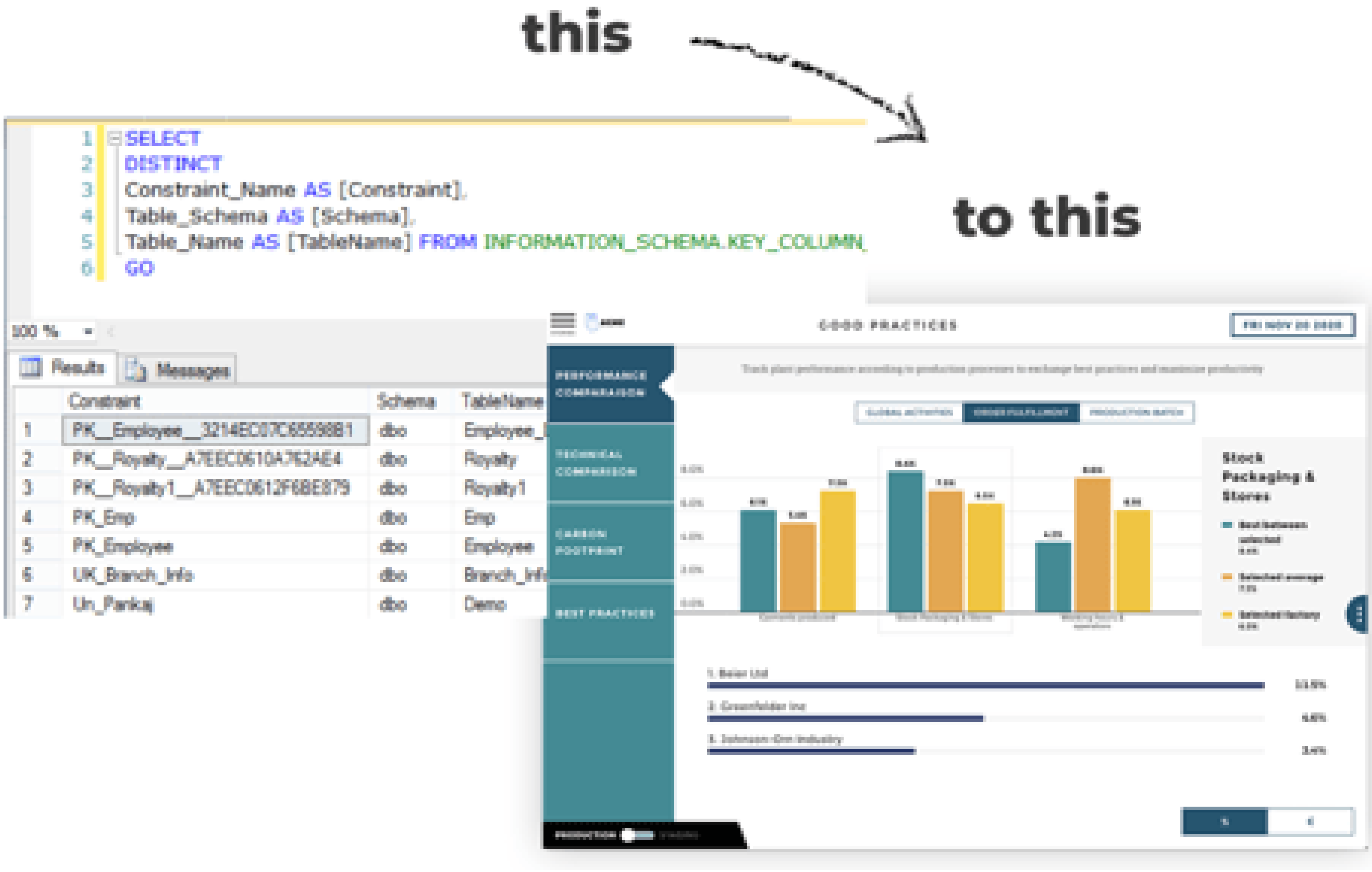
90 percent of the information transmitted to the brain is visual.

-MIT News

The human brain processes images 60,000x faster than text.

-The UM/3M Study

The Need to Visualize Data



```
SELECT
DISTINCT
Constraint_Name AS [Constraint],
Table_Schema AS [Schema],
Table_Name AS [TableName] FROM INFORMATION_SCHEMA.KEY_COLUMN_
GO
```

	Constraint	Schema	TableName
1	PK__Employee__3214EC07C65598B1	dbo	Employee_
2	PK__Royalty__A7EEC0610A762AE4	dbo	Royalty
3	PK__Royalty1__A7EEC0612F6BE879	dbo	Royalty1
4	PK_Emp	dbo	Emp
5	PK_Employee	dbo	Employee
6	UK_Branch_Info	dbo	Branch_Inf
7	Un_Pankaj	dbo	Demo

How to Select the Right Chart?

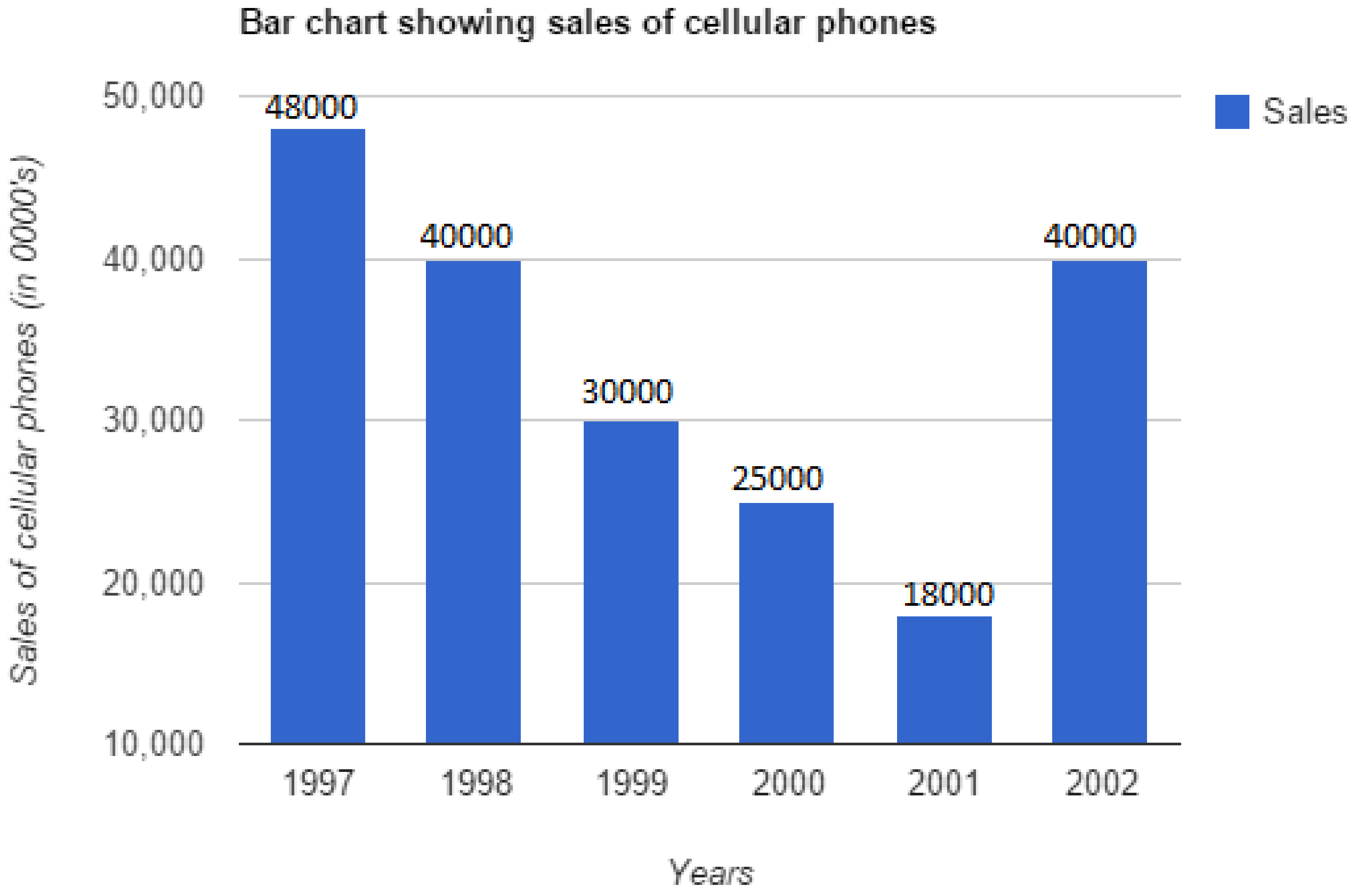
1. **Want to compare values?**

Charts are perfect for comparing one or many value sets, and they can easily show the low and high values in the data sets.

Use these charts to show comparisons:

- Column/bar
- Circular area
- Line
- Scatter plot
- Bullet

How to Select the Right Chart?



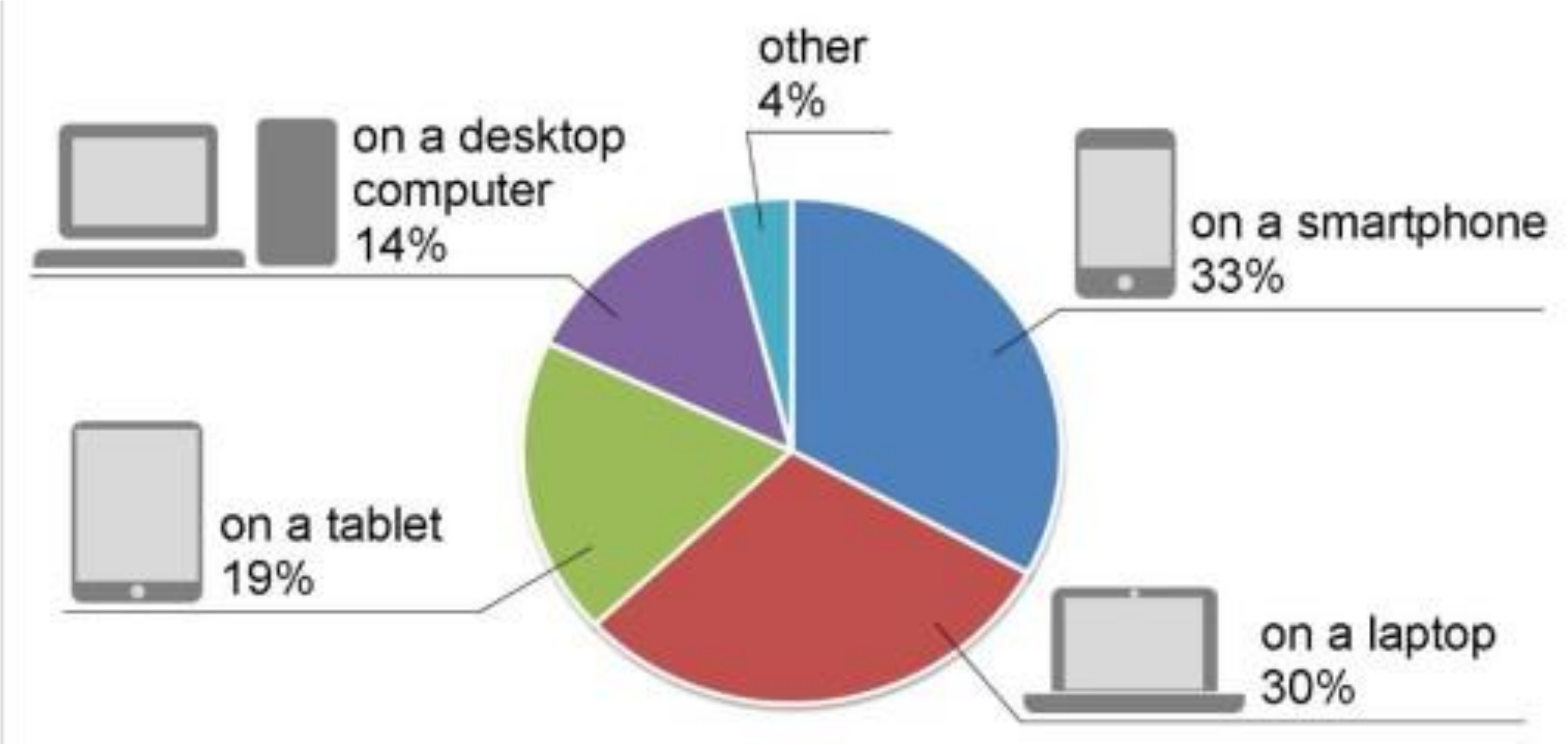
How to Select the Right Chart?

2. Want to show the composition of something? To show how individual parts make up the whole of something (such as the device used by visitors to your website, or total sales broken down by sales rep)

Use these charts to show composition:

- Pie
- Stacked bar
- Stacked column
- Area

How to Select the Right Chart?



How to Select the Right Chart?

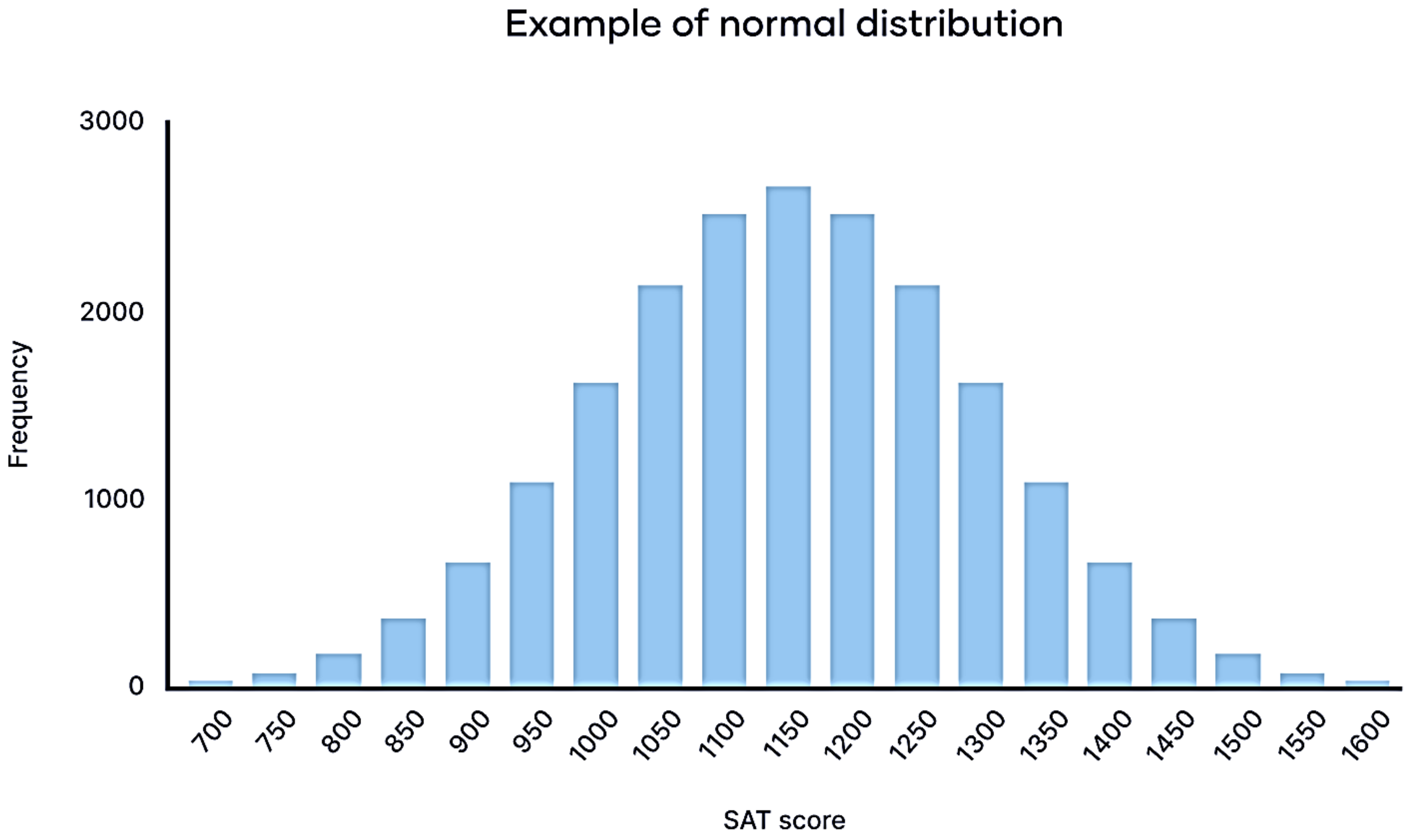
3. Want to understand the distribution of your data?

Distribution charts help you to understand outliers, the normal tendency, and the range of information in your values.

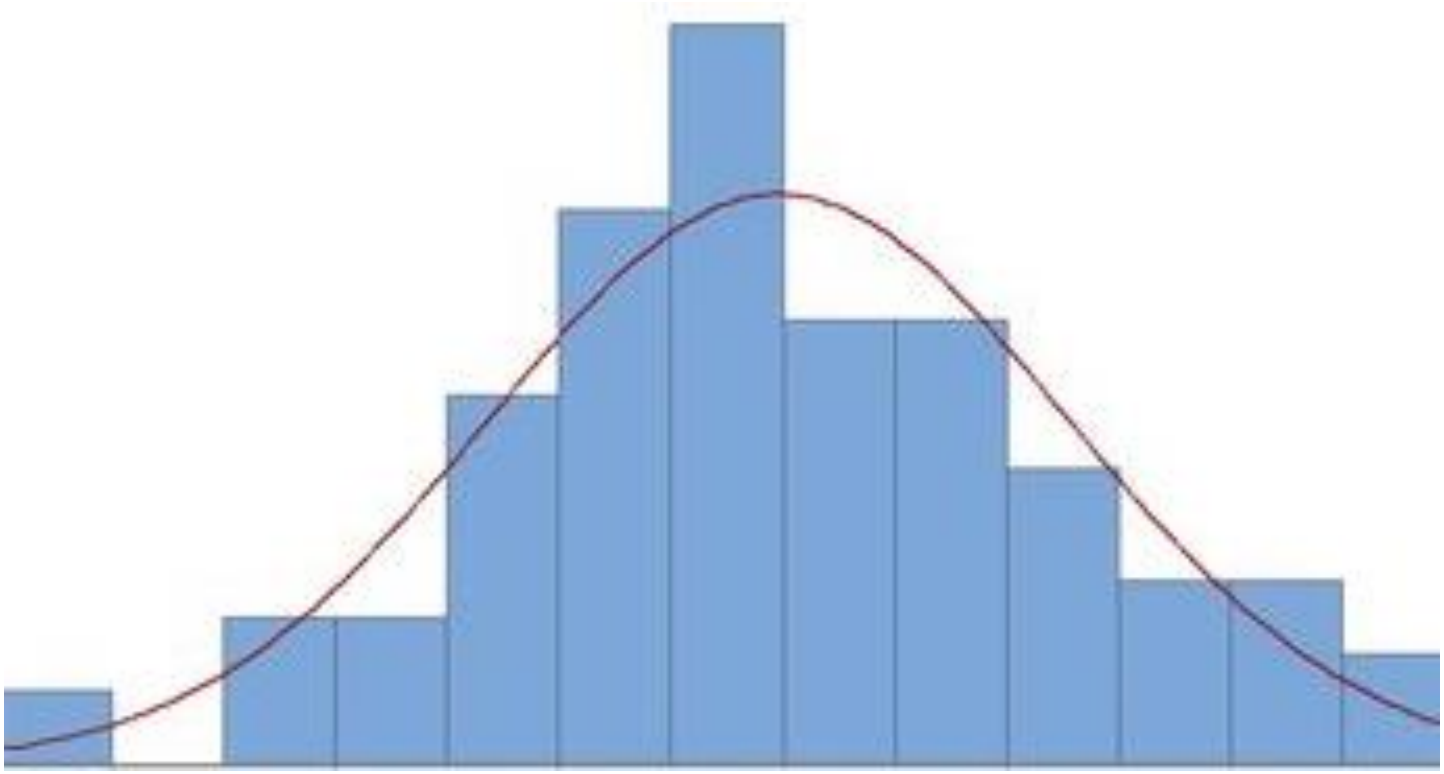
Use these charts to show distribution:

- Scatter plot
- Line
- Column
- Bar

How to Select the Right Chart?



How to Select the Right Chart?



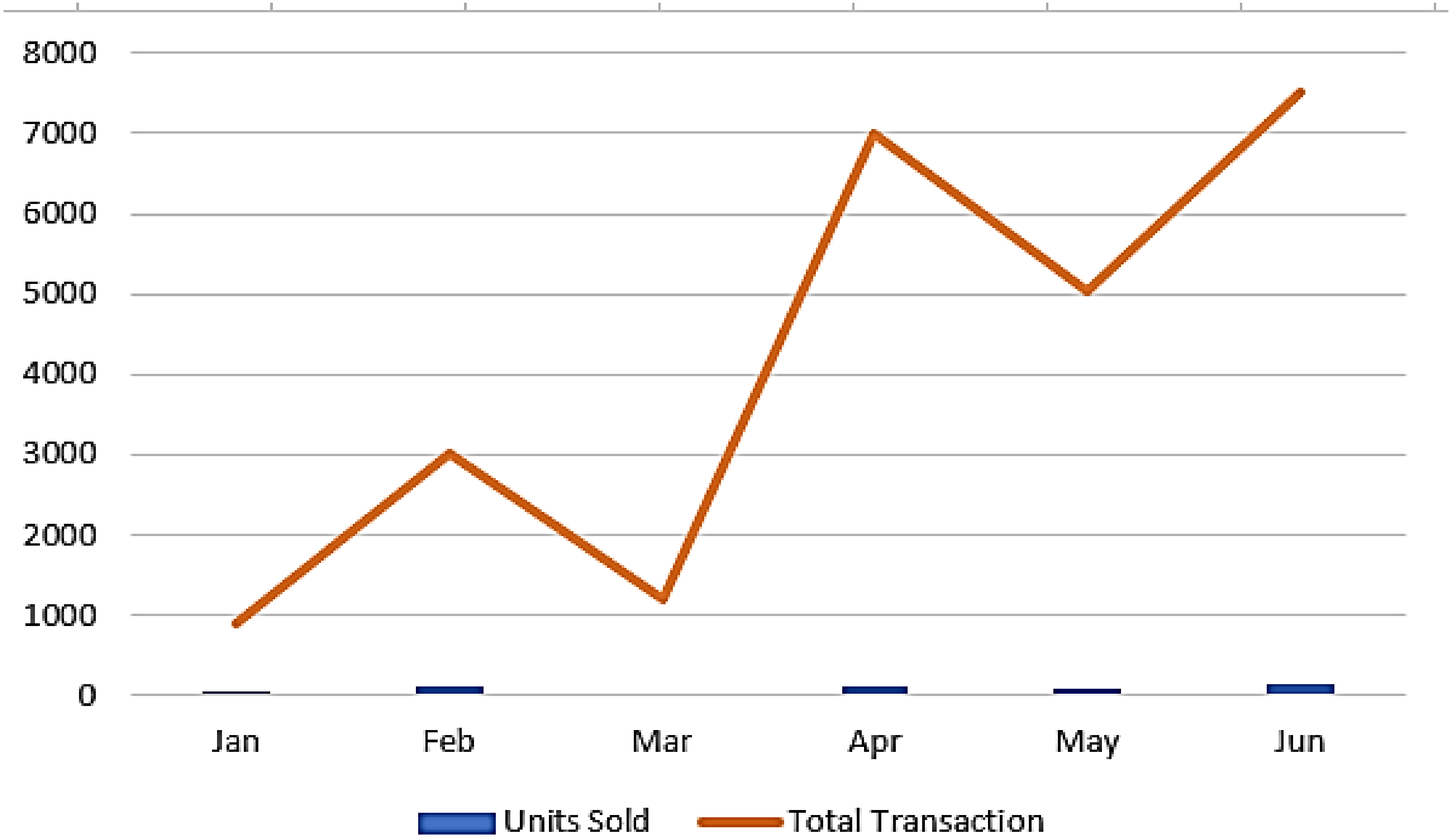
How to Select the Right Chart?

4. Interested in analyzing trends in your data set?
If you want more information about how a data set performed during a specific period, there are specific chart types that do this extremely well.

Use these charts to analyze trends:

- Line
- Dual-axis line
- Column

How to Select the Right Chart?



How to Select the Right Chart?

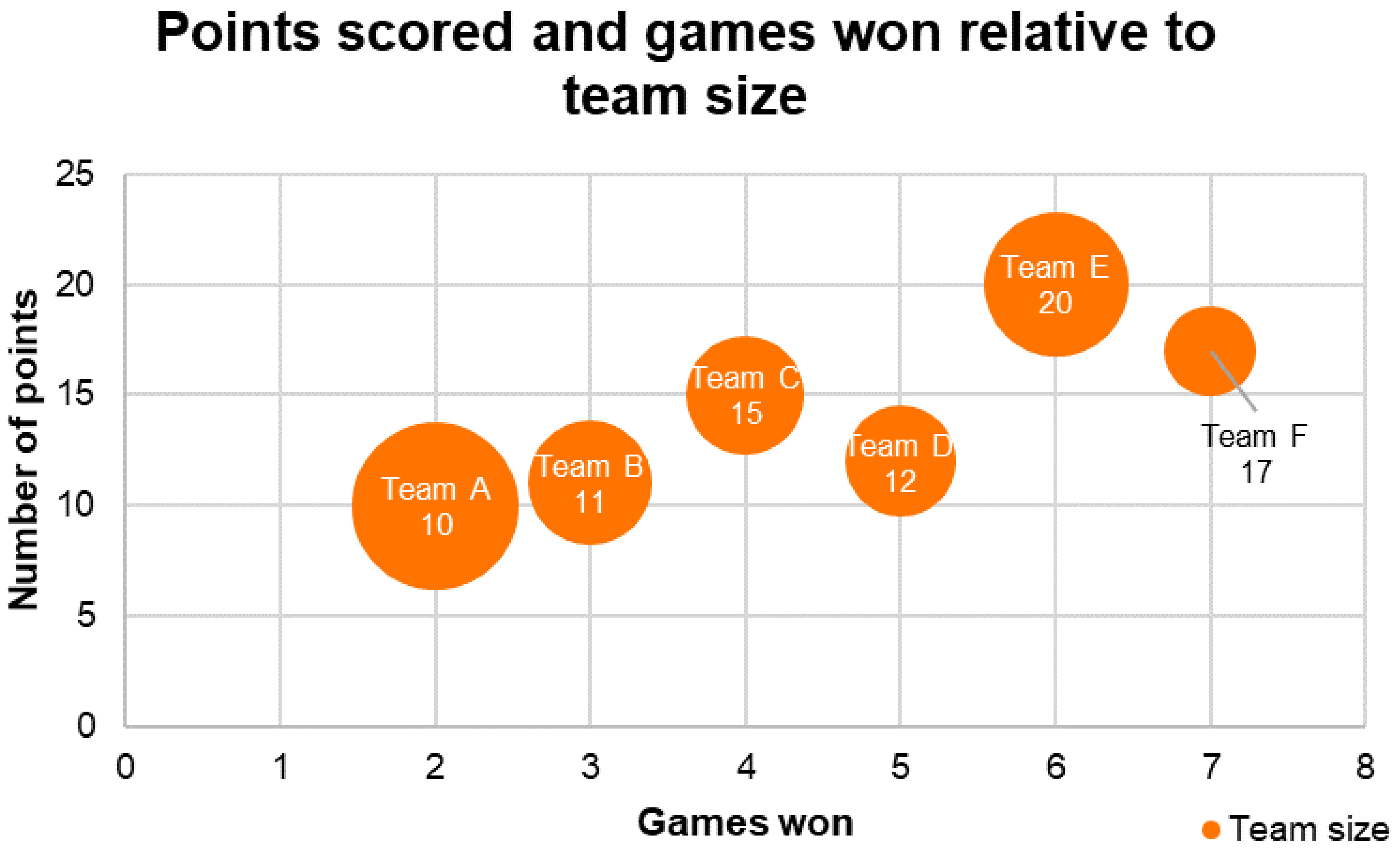
5. Want to better understand the relationships among value sets?

Relationship charts are designed to show how one variable relates to one or many different variables. You could show how something positively affects (or has no effect, or negatively affects) another variable.

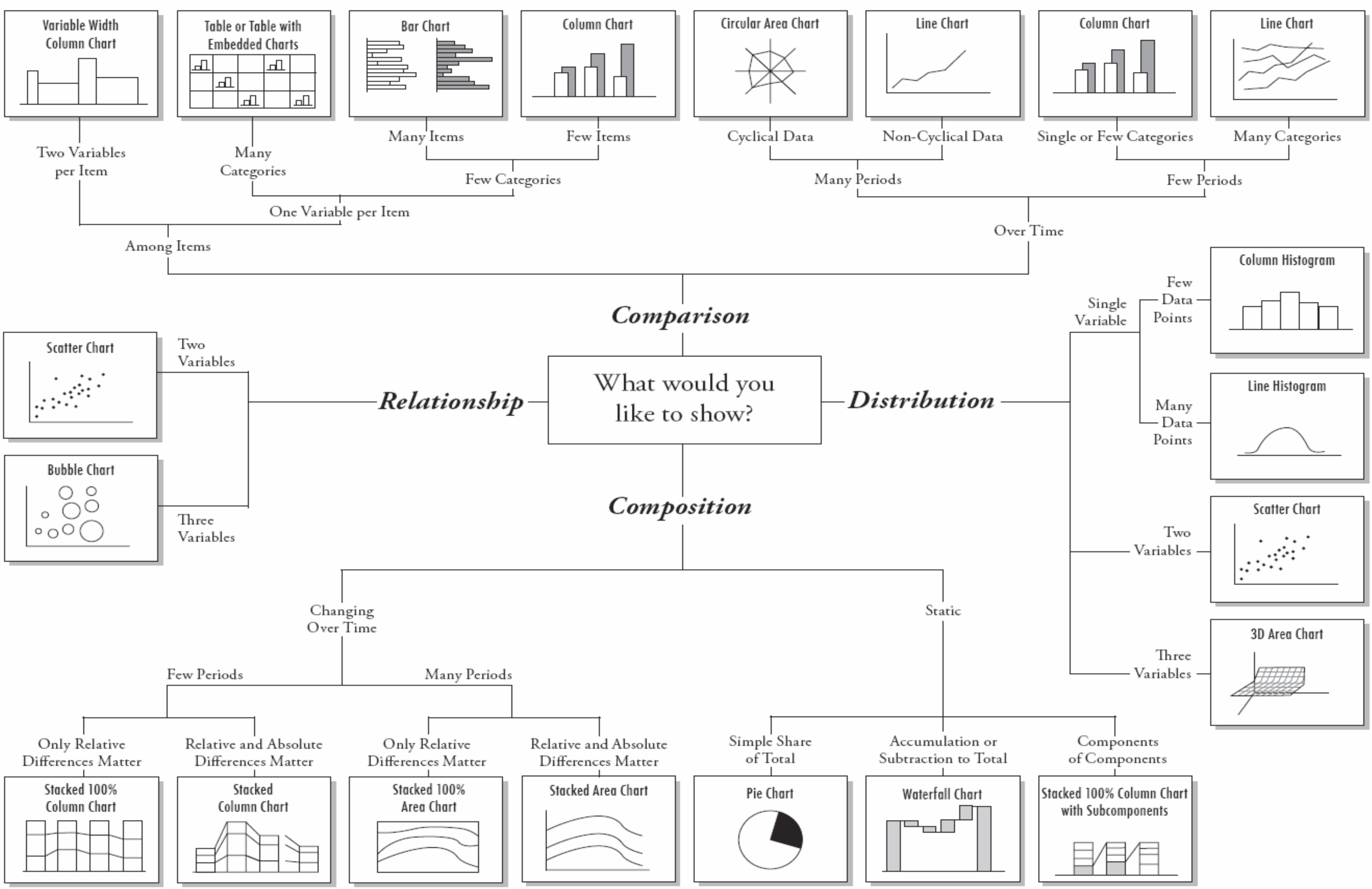
Use these charts to show relationships:

- Scatter plot
- Bubble
- Line

How to Select the Right Chart?



How to Select the Right Chart?



What would you like to show?

Comparison

Among Items

Two Variables per Item

Variable Width Column Chart

One Variable per Item

Many Categories

Table or Table with Embedded Charts

Few Categories

Many Items

Bar Chart

Few Items

Column Chart

Over Time

Many Periods

Cyclical Data

Circular Area Chart

Non-Cyclical Data

Line Chart

Few Periods

Single or Few Categories

Column Chart

Many Categories

Line Chart

Relationship

Two Variables

Scatter Chart

Three Variables

Bubble Chart

Distribution

Single Variable

Few Data Points

Column Histogram

Many Data Points

Line Histogram

Two Variables

Scatter Chart

Three Variables

3D Area Chart

Composition

Changing Over Time

Few Periods

Only Relative Differences Matter

Stacked 100% Column Chart

Relative and Absolute Differences Matter

Stacked Column Chart

Many Periods

Only Relative Differences Matter

Stacked 100% Area Chart

Relative and Absolute Differences Matter

Stacked Area Chart

Static

Simple Share of Total

Pie Chart

Accumulation or Subtraction to Total

Waterfall Chart

Components of Components

Stacked 100% Column Chart with Subcomponents

- Marketinggrowmatics.com

This **IIT Graduate Quit his INTEL Job** in the US, **returned to India** and **bought 20 cows** and Started a dairy farm Now, his dairy Brand Earns **Rs 44 Crore** revenue

Impact of Dedication

Impact of Dedication

Impact of Dedication

R Programming

R is a popular programming language used for statistical computing and graphical presentation.

Why Use R?

- Great resource for data analysis, data visualization, and data science
- It provides many statistical techniques
- It is easy to draw graphs in R
- It is open-source and free
- It has a large community support
- It has many packages (libraries of functions) that can be used to solve different problems

R How did R get its name as ‘R’?

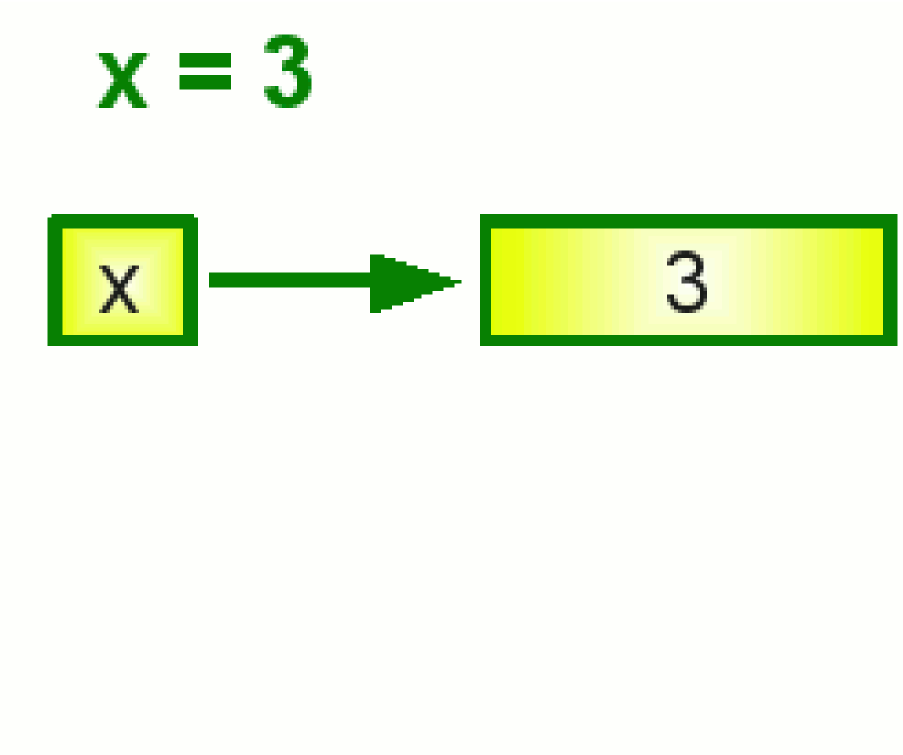
R's developers, statisticians Ross Ihaka and Robert Gentlemen, named their language after the first initial in their names.

Package Installation in R

```
install.packages(“readxl”)

install.packages(“dplyr”)

install.packages("caret”)

install.packages('e1071')

install.packages("Metrics")
```

Variable in R

A variable is a location in memory that holds one or more values. It has a label or name to identify it and its values can be changed



Variable in R

=

```
text <- "Hello World"

text <- 5

text <- 5.5

text <- ‘c’
```

Data Types of Variables in R

In R, variables do not need to be declared with any particular type

❖numeric - (10.5, 55, 787)

❖integer - (1L, 55L, 100L, where "L" declares it an integer)

❖complex - (9 + 3i, where "i" is the imaginary part)

❖character (a.k.a. string) - ("k", "R is exciting", "11.5")

❖logical (a.k.a. boolean) - (TRUE or FALSE)

Rules for Variable Name

- A variable name must start with a letter and can combine letters, digits, period(.) and underscore (_).
- A variable name cannot start with a number or underscore (_)
- Variable names are case-sensitive (age, Age and AGE are three different variables)
- Reserved words cannot be used as variables (TRUE, FALSE, NULL, if...)

Multiple Variables

```
# Assign the same value to multiple variables

var1 <- var2 <- var3 <- "Orange"
```

Print the Variable in R

```
text <- "Hello World“
print(text)
```

Comments in R

A comment is basically a text note that gives an explanation about the source code.

```
# This is a comment
```

Built-in Math Functions

```
max(5, 10, 15)
min(5, 10, 15)
abs(-4.7)

sqrt(16)
ceiling(1.4)
floor(1.4)
```

https://www.datacamp.com/doc/r/functions

Strings in R

To find the number of characters in a string, use the nchar( ) function

```
str <- "Hello World!"

nchar(str)
```

Check a String

Use the grepl( ) function to check if a character or a sequence of characters are present in a string

grep → global regular expression print

```
str <- "Hello World!"

grepl("H", str)
grepl("Hello", str)
grepl("X", str)
```

Concatenate Strings

You can also concatenate, or join, two or more string elements, by using the paste( ) function. To combine both text and a variable, R uses comma (,):

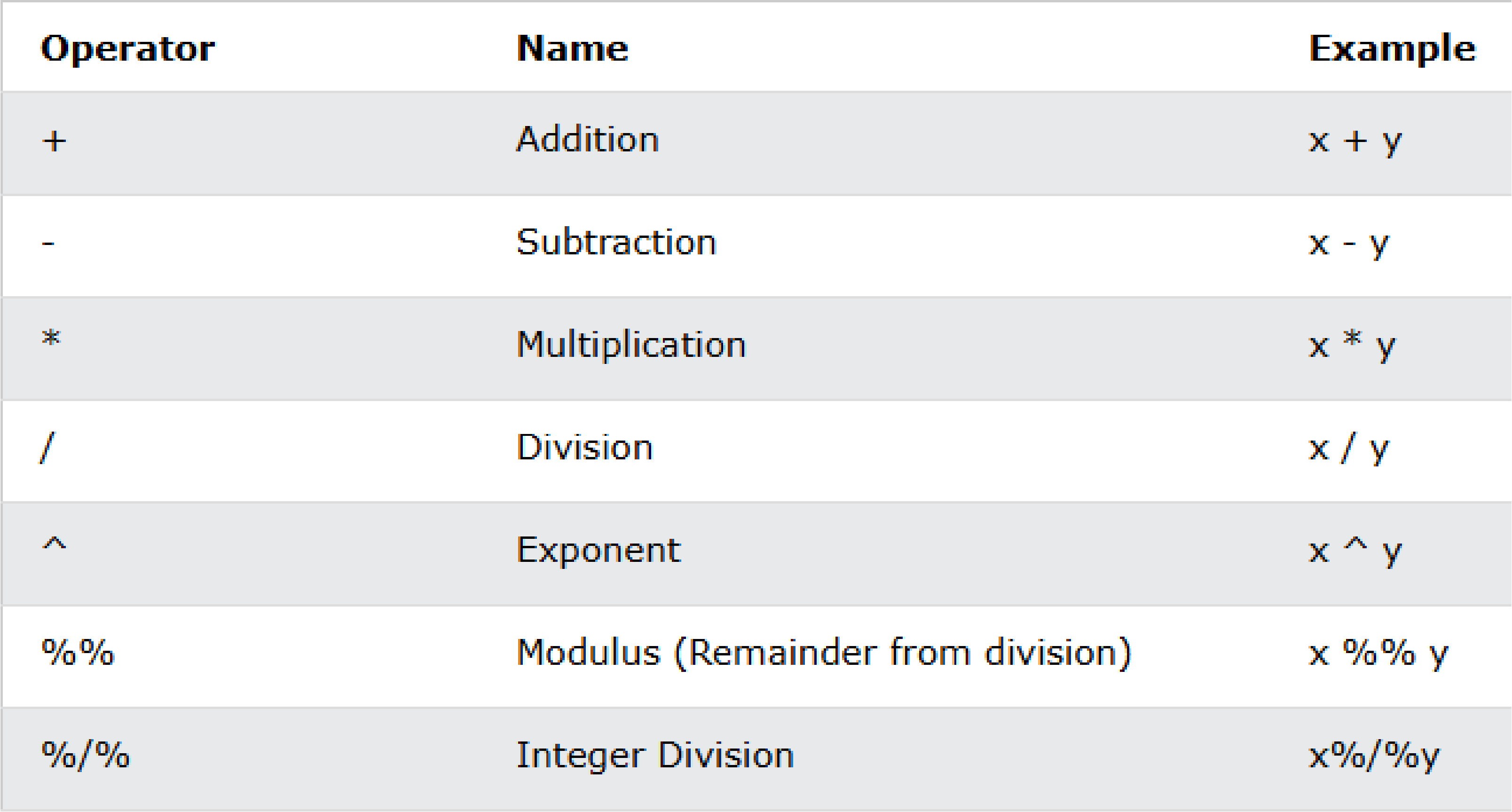
```
text <- "awesome"
paste("R is", text)
```

```
text1 <- "R is"
text2 <- "awesome"
paste(text1, text2)
```

Escape Characters

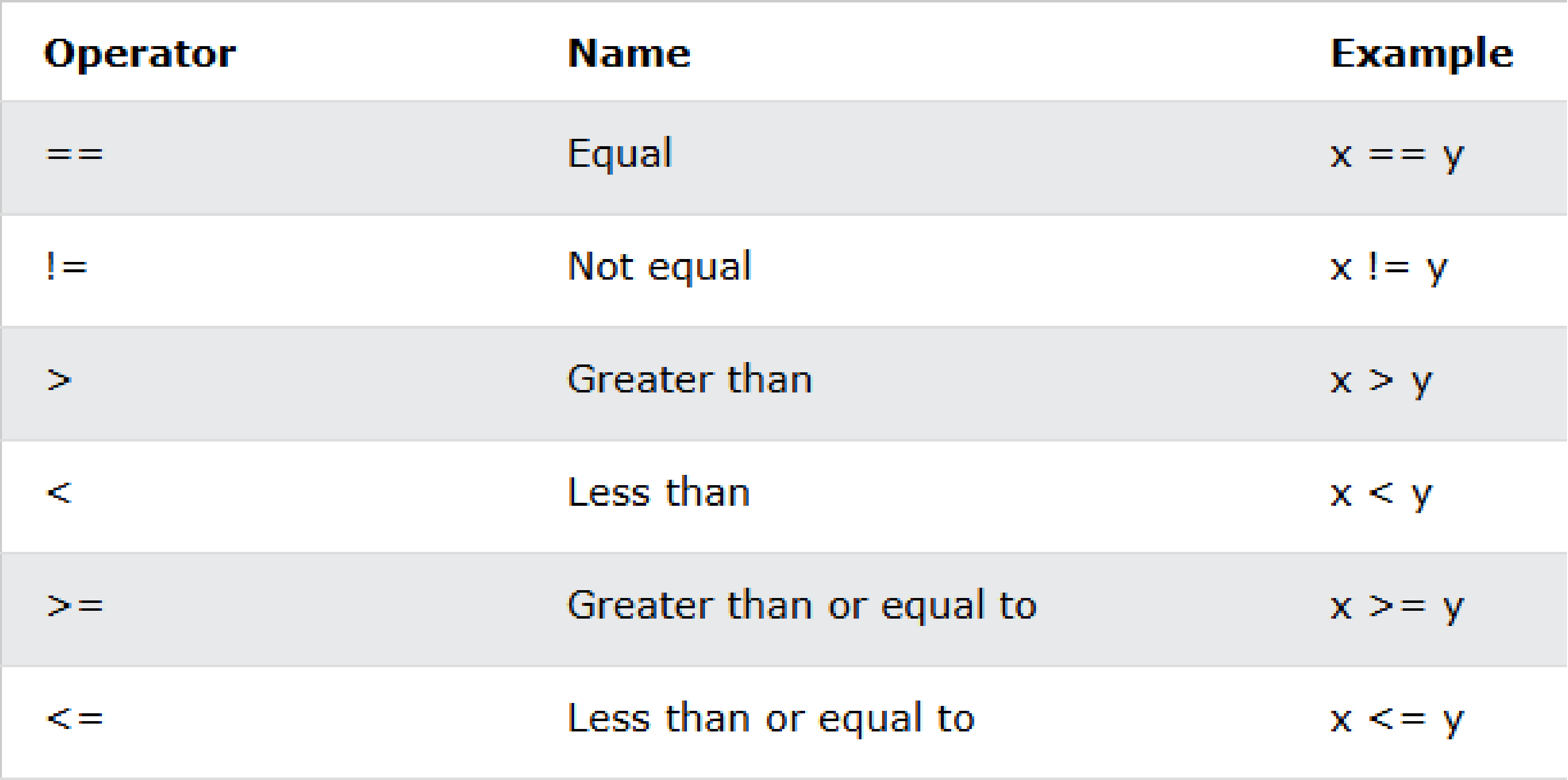
```r
str <- "We are the so-called \"Vikings\",
        from the north."
```

R Arithmetic Operators



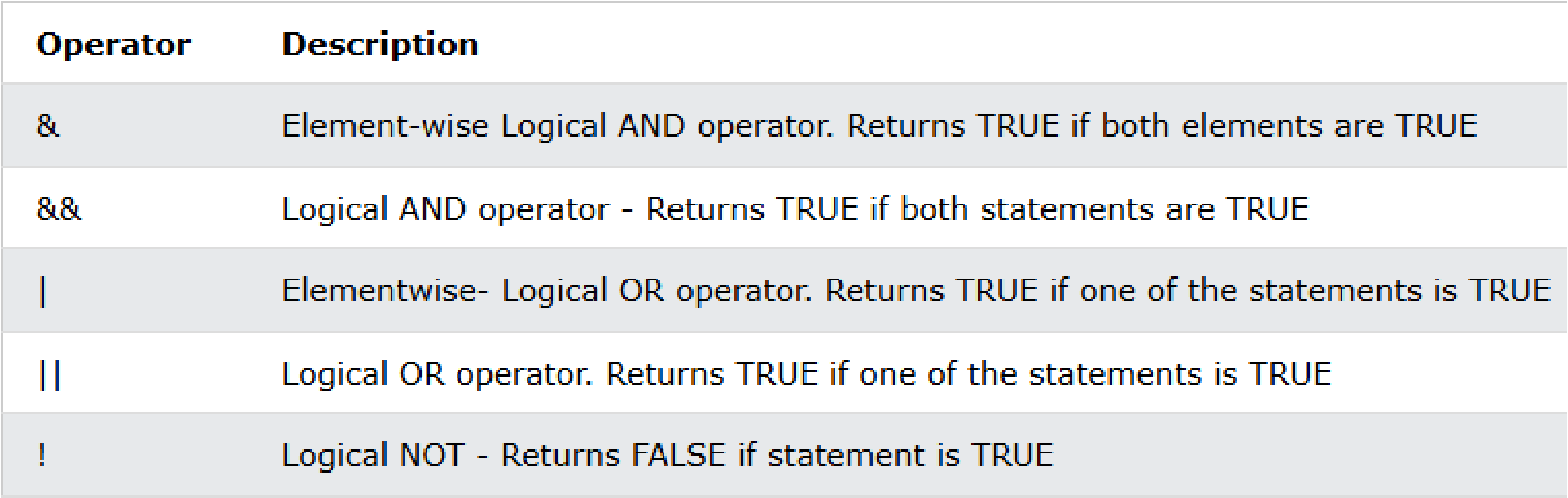
Operator	Name	Example
+	Addition	x + y
-	Subtraction	x - y
*	Multiplication	x * y
/	Division	x / y
^	Exponent	x ^ y
%%	Modulus (Remainder from division)	x %% y
%/%	Integer Division	x%/%y

R Comparison Operators



Operator	Name	Example
==	Equal	x == y
!=	Not equal	x != y
>	Greater than	x > y
<	Less than	x < y
>=	Greater than or equal to	x >= y
<=	Less than or equal to	x <= y

R Logical Operators



Operator	Description
&	Element-wise Logical AND operator. Returns TRUE if both elements are TRUE
&&	Logical AND operator - Returns TRUE if both statements are TRUE
|	Elementwise- Logical OR operator. Returns TRUE if one of the statements is TRUE
||	Logical OR operator. Returns TRUE if one of the statements is TRUE
!	Logical NOT - Returns FALSE if statement is TRUE

Conditions and If Statements

```
a <- 200
b <- 33

if (b > a) {
  print("b is greater than a")
  } else if (a == b) {
  print("a and b are equal")
  } else {
  print("a is greater than b")
}
```

For Loop

```
for (x in 1:10) {
  print(x)
}
```

```
dice <- c(1, 2, 3, 4, 5, 6)

for (x in dice) {
  print(x)
}
```

R Function

```
my_function <- function(fname) {
  paste(fname, "Griffin")
}

my_function("Peter")
my_function("Lois")
my_function("Stewie")
```

R Function

```
my_function <- function(fname, lname) {
  paste(fname, lname)
}

my_function("Peter", "Griffin")
```

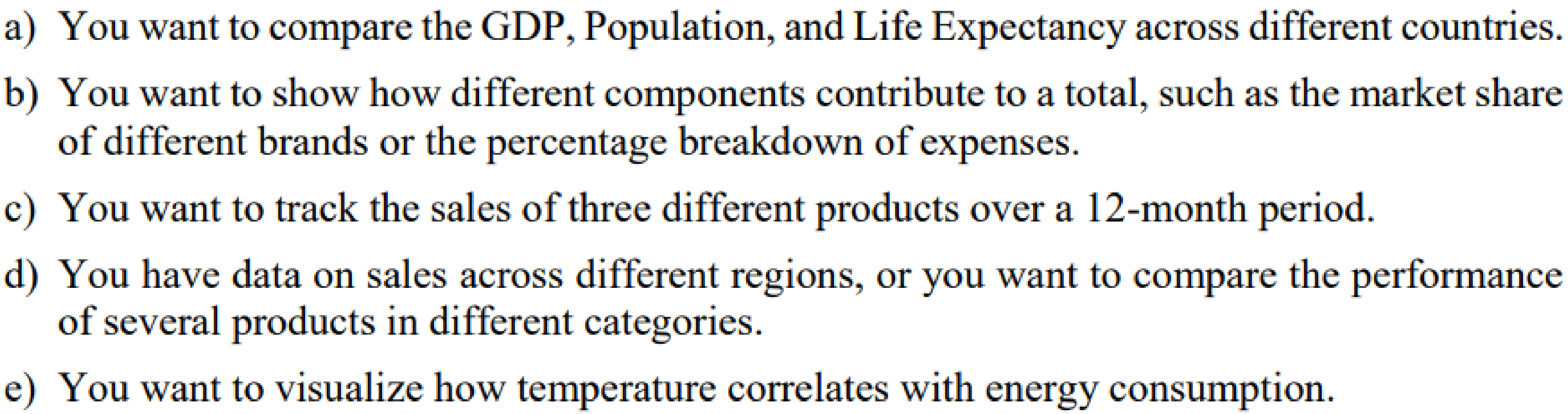
Package Installation in R

```
install.packages(“readxl”)

install.packages(“dplyr”)

install.packages("caret”)

install.packages('e1071')

install.packages("Metrics")
```

‘R’ Script

How to save ‘R’ script in a directory/drive folder and run it from the console!?

Sample Chart Related Question



a) You want to compare the GDP, Population, and Life Expectancy across different countries.

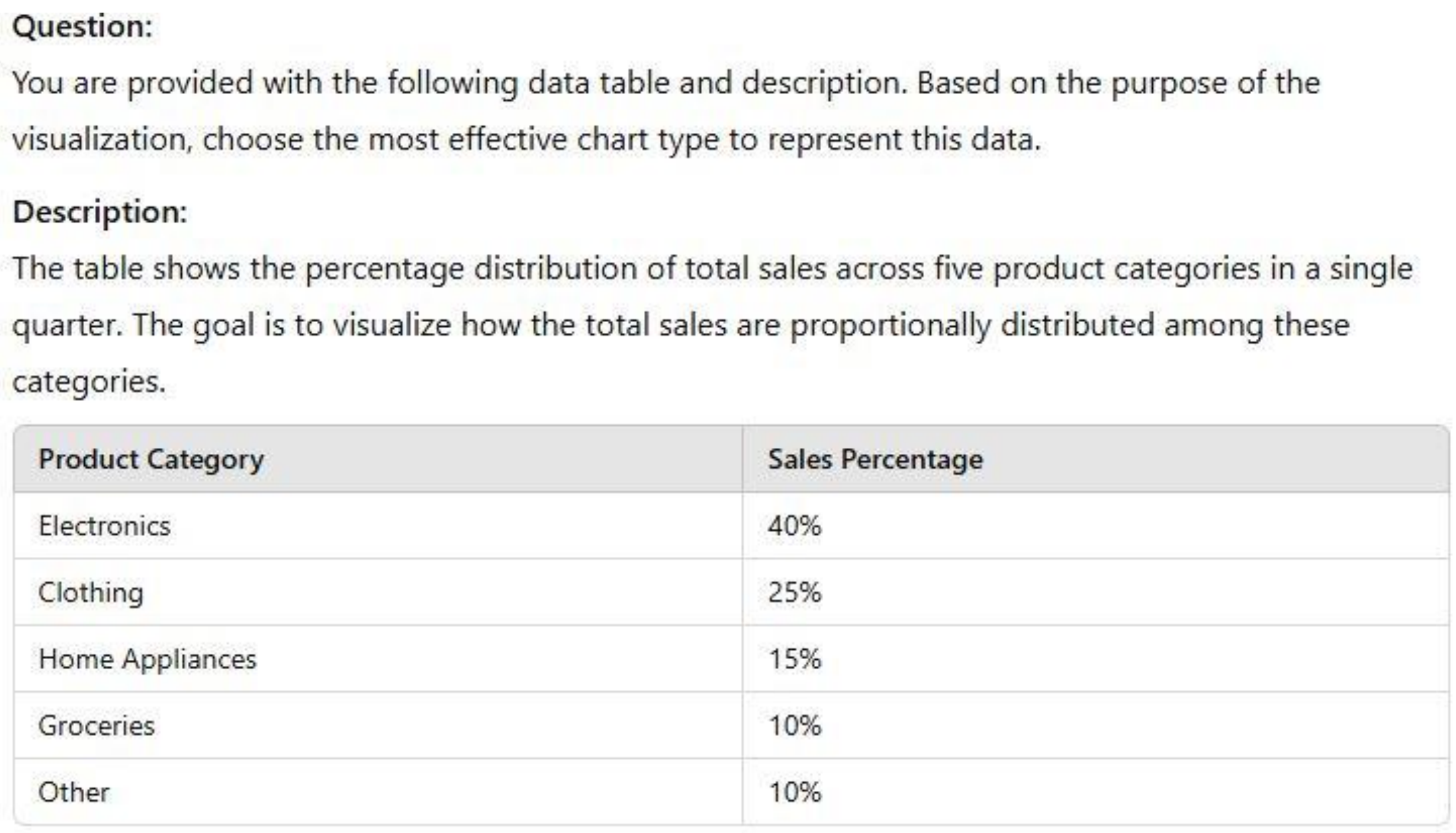
b) You want to show how different components contribute to a total, such as the market share of different brands or the percentage breakdown of expenses.

c) You want to track the sales of three different products over a 12-month period.

d) You have data on sales across different regions, or you want to compare the performance of several products in different categories.

e) You want to visualize how temperature correlates with energy consumption.

Sample Questions for Final Lab



Question:

You are provided with the following data table and description. Based on the purpose of the visualization, choose the most effective chart type to represent this data.

Description:

The table shows the percentage distribution of total sales across five product categories in a single quarter. The goal is to visualize how the total sales are proportionally distributed among these categories.

Product Category	Sales Percentage
Electronics	40%
Clothing	25%
Home Appliances	15%
Groceries	10%
Other	10%

Sample Questions for Final Lab

Evaluation Criteria:

- Correct chart selection
- Add proper labels and legends
- Proper color and shading
- Number of parameters used

Thank You!