

Week 02

Understanding and Applying Data in Middle East & Africa

Middle East and African Studies, GSIAS, HUFS : DooYoung Wicks

Qualitative and Quantitative issues

Quantitative

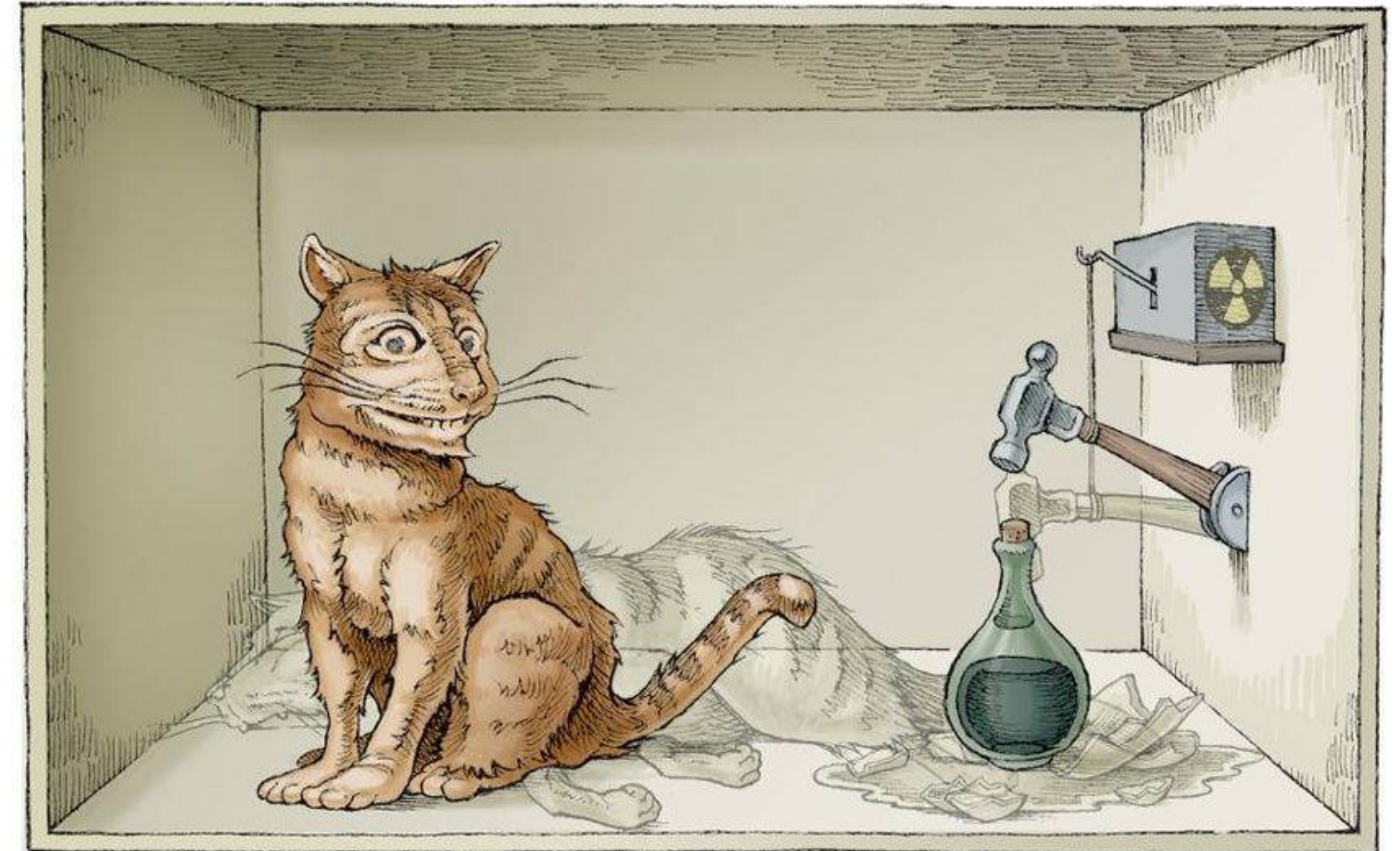
- Research knows the research target
- The research should be fully designed
- Availability of the data, e.g.) economic and political indicators, survey data or other equivalent numerical data.
- Is it possible to transform its data into numerical?
 - Is it EMPIRICAL?
 - Testing the Hypothesis.

Qualitative

- Describing “**the big picture**”
- May only roughly know the research question and is in the preliminary stage.
- **However, if the researcher is confident to give a completely detailed description.**
- Data may be pictures, community, and literature, which can be interpreted with participation.
- It can be more rich information than Quantitative.

Observers' paradox

- I need to go into community A to conduct the research
- But by going into Community A, How can I be sure I am not influencing or altering the behaviour of Community A by simply being present?
- But, I cannot gather the information if I do not go.



How to select a good research theme?

1. Select the right research question approach.
 - Empirical questions
 - Normative questions
 - Analytical questions
2. Presenting the criteria for a good research theme and determining.
 - Clarity
 - Testability
 - Theoretical significance
 - Practical relevance
 - Originality

Types of the research questions

No research question, No beginning

- It is a fundamental component of the research design
- Defines overall directions and quality of study

- Analysing the impact of oil revenue on economic development and political stability in Angola.
- Analyzing the role of small and medium enterprises in Gulf countries and their impact on job creation and economic growth.
- Analyzing the trends in Chinese investment and trade in East African countries.
- How can governments in Northern Africa encourage foreign investment while ensuring that local economies benefit from it?
- What are the most effective measures for promoting women's rights and gender equality in Libya
- What are the implications of foreign investment in Kenya and Ethiopia on their economic growth and political stability?
- How has the role of women in the labour market evolved in the Gulf countries and MENA, and what impact has this had on economic growth and development?

Types of the research questions

Empirical, Normative and Analytical

- Identify the problem
- Select a right solution

Empirical questions

Normative questions

Analytical questions

Empirical Questions

No research question, No beginning

- It is a fundamental component of the research design
- Defines overall directions and quality of study

Q1: Analysing the impact of oil revenue on economic development and political stability in Angola.

→ It can evaluate the correlation between the oil revenues (e.g., OPEC) and political stability (e.g., [Worldwide Governance Indicators \(WGI\)](#))

Q1: Analyzing the role of small and medium enterprises in Gulf countries and their impact on job creation and economic growth.

→ It can evaluate the correlation between the [MSME Economic Indicators](#) and Global Jobs Indicators Database (e.g., World Bank: JOIN)

Q1: Analyzing the Chinese investment and trade trends in East African countries.

→ It can evaluate the correlation between the Chinese FDI Flows and Global trade (e.g., World Bank: JOIN)

Normative questions

What it should be: ought-isms

- Desirable issues and what is not
- Judgement depends on the values of the researcher.

Q1: How can governments in Northern Africa encourage foreign investment while ensuring that local economies benefit from it?

→ An evaluation of what is important is required.

Q2: What are the most effective measures for promoting women's rights and gender equality in Libya?

→ Supporting and encouraging women to participate in political activity. Ensure women's access to education and employment opportunities.

Analytical questions

Comparing, contrasting and evaluating

- Known as formal theory
- Multiple approach

Q1: What are the implications of foreign investment in Kenya and Ethiopia on their economic growth and political stability?

Q2: How has the role of women in the labour market evolved in the Gulf countries and MENA, and what impact has this had on economic growth and development?

→ They usually involve comparing, contrasting, evaluating, or interpreting different aspects of a topic

How to find 'Good Research'

1. Good Research Theme | Alan Monroe (2000), Essentials of political research

- **Clarity**
- **Testability**
- **Theoretical significance**
- **Practical relevance**
- **Originality**

What is the “Good Theory”

It does not refer to a specific, well-known theory in political economy research.

- **Based on Empirical Evidence:** Uses real-world data for testing and improvement.
- **Explanatory Power:** Explains and predicts economic and political phenomena.
- **Verifiable:** Must be falsifiable to be scientifically sound.
- **Simple and Concise:** Uses the principle of parsimony, or Occam's Razor.
- **Allows Generalization:** Applies to various situations and contexts, beyond specific boundaries.

The Research Design Process

Can ensure a systematic approach

- It could help a robust and reliable understanding of complex phenomena.

- **Puzzle Identification:**

Find a problem or question for further exploration.

- **Theoretical Framework Development:**

Create a set of assumptions or propositions to explain the puzzle.

- **Working Hypotheses Derivation:**

Develop specific working hypotheses from the theory.

- **Data Collection for Hypothesis Testing:**

Gather relevant data to test the working hypotheses.

- **Hypothesis Testing and Validation:**

Use data to verify the validity of working hypotheses and the robustness of the theoretical framework.

The Research Design Process

Let's Apply with the African context

- **Example: Political Economy of Africa and the Resource Curse**

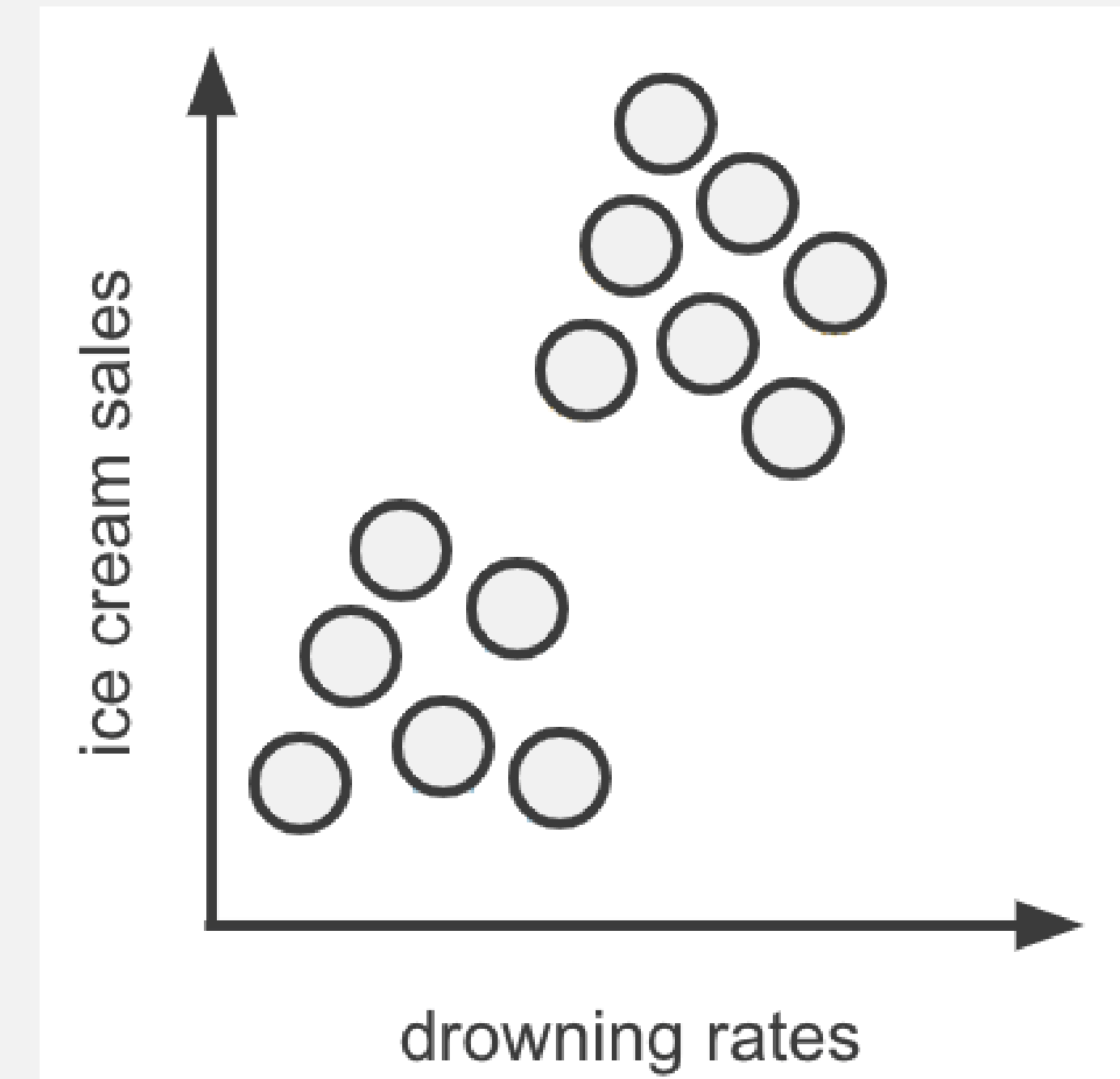
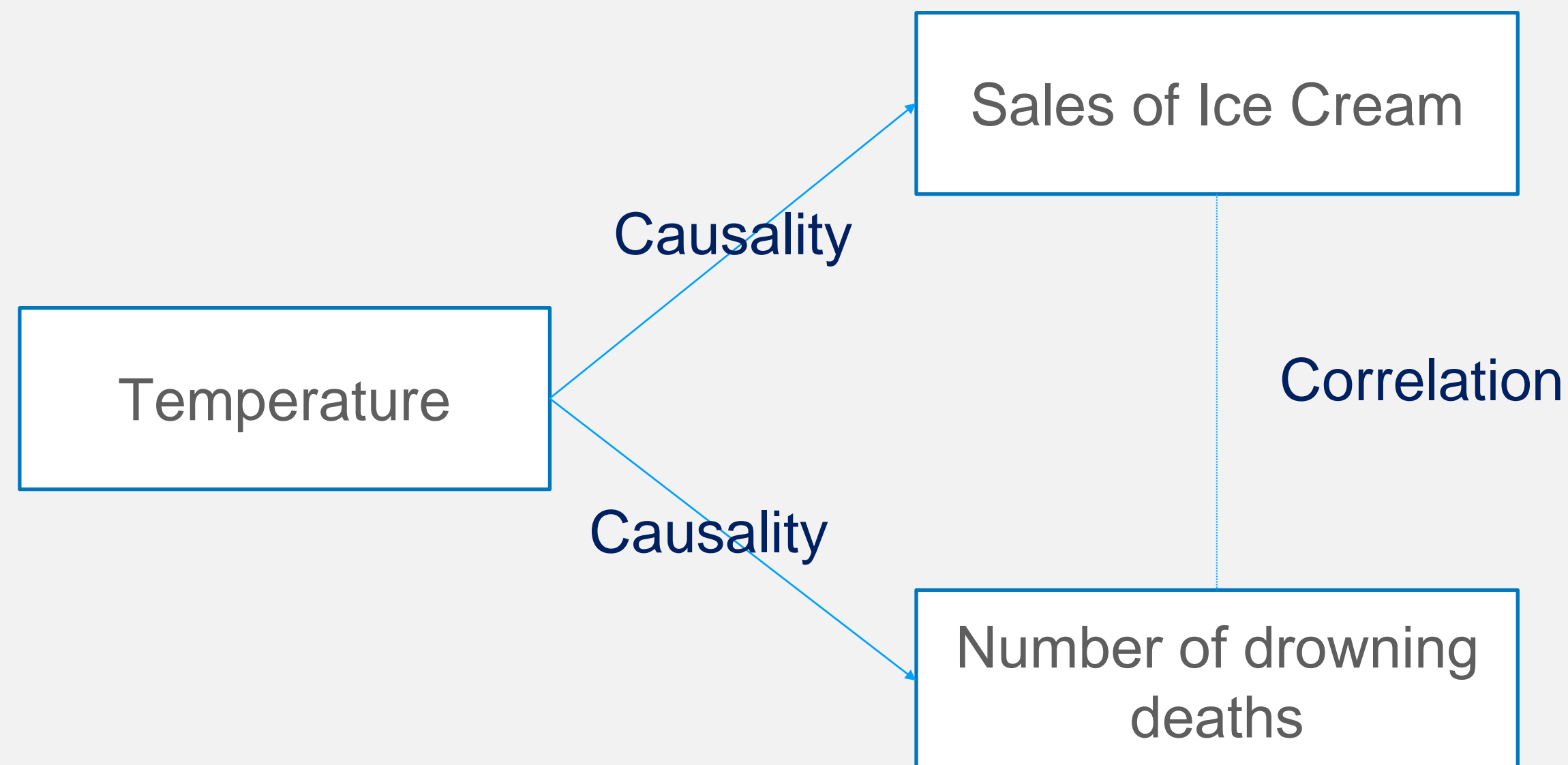
- **Puzzle:** Paradox of natural resource wealth leading to adverse outcomes.
- **Theoretical Framework:** Resource curse theory - abundant resources leading to negative consequences.
- **Working Hypotheses:** Countries with higher resource dependence experience poor governance and corruption.
- **Data Collection:** Gather data on resource dependence, governance quality, and corruption levels from sources like the World Bank, IMF, and Transparency International.
- **Hypothesis Testing and Validation:** Use statistical techniques to test the relationship between resource dependence and governance quality, supporting or refuting the resource curse theory in Africa.

The Three Criteria for Establishing Causation

- 1.The cause precedes the effect.
- 2.The cause-and-effect covary.
- 3.No significant factors are other than the cause of change.

- Start Asking Your Data 'Why?'
- Avoiding Spurious Correlations by Understanding Common Causes

Causality and Correlation



The data suggests a correlation between the variables. Does this imply causation?

The Three Criteria for Establishing Causation

- **Start Asking Your Data 'Why?'**
- **Avoiding Spurious Correlations by Understanding Common Causes**

- **Empirical Analysis:** The role of logical consistency in theories
Understand the difference between correlated variables and causal relationships.
- **Correlation vs Causation Example:**
Temperature, ice cream sales, and drowning deaths.
- **Importance in Empirical Economic Analysis:**
Determine whether relationships are causal or merely correlational.
- **Three Conditions for Establishing Causation:**
 - a. Cause Precedes the Effect
 - b. Cause-and-Effect Covary
 - c. No Significant Factors Other Than the Cause of Change
- **Applying Conditions to Example:**
Assess the relationship between temperature and ice cream sales using the three criteria.

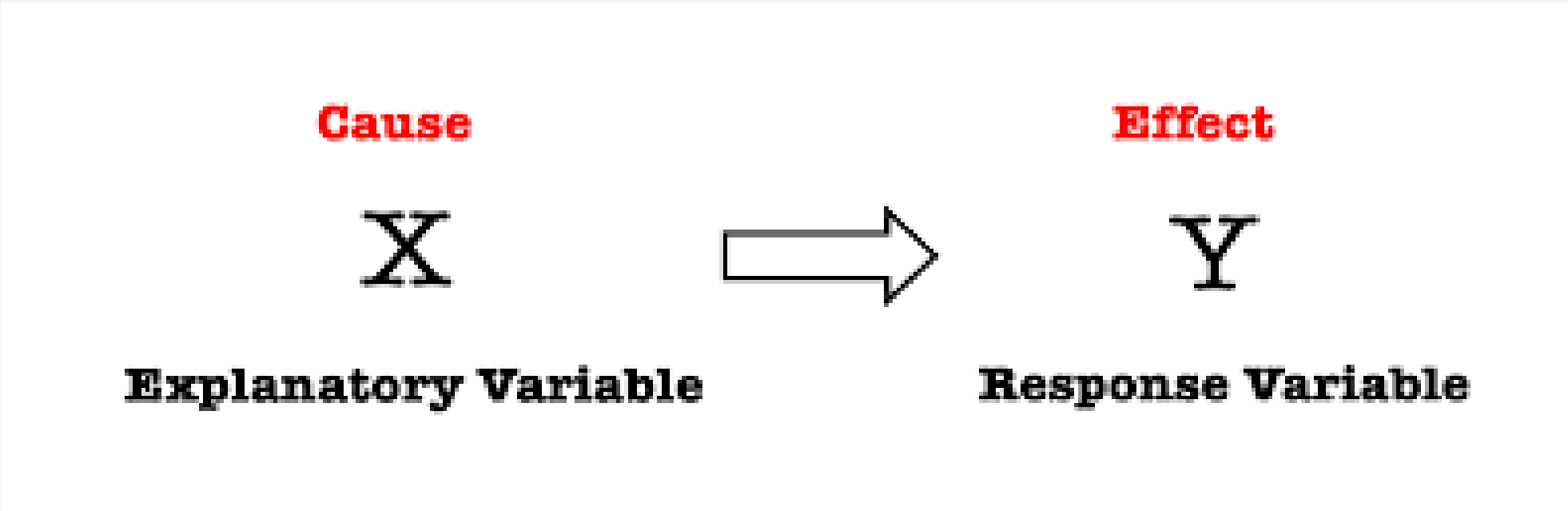
What is the Theory

"general discourse on cause and effect"

- It is a bundle of multiple assumptions
- a theory relies on multiple "ifs."

X = Explanatory variable (Independent Variable)

Y = Response variable (Dependent Variable)



x	Independent Variable	Cause	Predictor Variable		Explanatory Variable
y	Dependent Variable	Effect	Outcome Variable	Response Variable	Criterion Variable

Criteria for a Good Theory

"Accurate, Testable, and Comprehensive: The Essence of a Good Political Economy Theory"

- **Accurate & Testable Insights**
- **Broad Scope for Understanding**

Falsifiability: Embrace the possibility of error to evolve and improve theories.

Observable Implications: Theories should have numerous testable predictions for higher falsifiability.

Specificity: Aim for precise predictions to increase observable implications and falsifiability.

Simplicity: Keep theories simple for broader applications, easier understanding, and higher falsifiability.

Hypotheses and the Process of Hypothesis Testing

Specificity, Testability, Falsifiability, Predictive ability

- Known as formal theory
- Multiple approach

Defining Working Hypotheses

- Definition: narrative concerning specific variables derived from a theory
- Articulated as "If this theory is accurate, then... should transpire."
- More detailed and specific than a theory
- Addresses observable predictions stemming from the theory

Examples

Ethnic and Religious Conflicts

- Working Hypothesis: Ethnic and religious conflicts in African countries
- Hypothesis Possesses: Higher ethnic and religious diversity correlates with increased conflict likelihood

Corruption and Governance

- Working Hypothesis: Corruption challenges in African countries
- Hypothesis Possesses: Improved governance and anti-corruption measures increase trust and economic development

Democratic Transitions and Consolidation

- Working Hypothesis: Democratic transitions and consolidation challenges in African countries
- Hypothesis Possesses: Strong civil society organizations and independent media positively influence democratic consolidation

Regional Integration and Cooperation

- Working Hypothesis: Regional integration and cooperation in Africa
- Hypothesis Possesses: Increased regional cooperation and integration improve economic development, peace, and security

Examples

Resource Management and Distribution

- Working Hypothesis: Resource management and distribution issues in Africa
- Hypothesis Possesses: Transparent resource management policies lead to more equitable distribution, reducing the "resource curse"

Terrorism and Insurgency

- Working Hypothesis: Terrorism and insurgency in Africa
- Hypothesis Possesses: Addressing root causes decreases the prevalence and impact of extremist groups

Refugee Crises and Displacement

- Working Hypothesis: Refugee crises and displacement in Africa
- Hypothesis Possesses: Enhanced regional cooperation and burden-sharing improve living conditions and promote stability

Selection of Unit of Analysis

It is the entity that you are analyzing

- Individuals, Groups, Artifacts (News Media), Geographical Units (Town, State), Social interactions (divorce, arrest)
- For different analyses in the same study, you may have different units of analysis

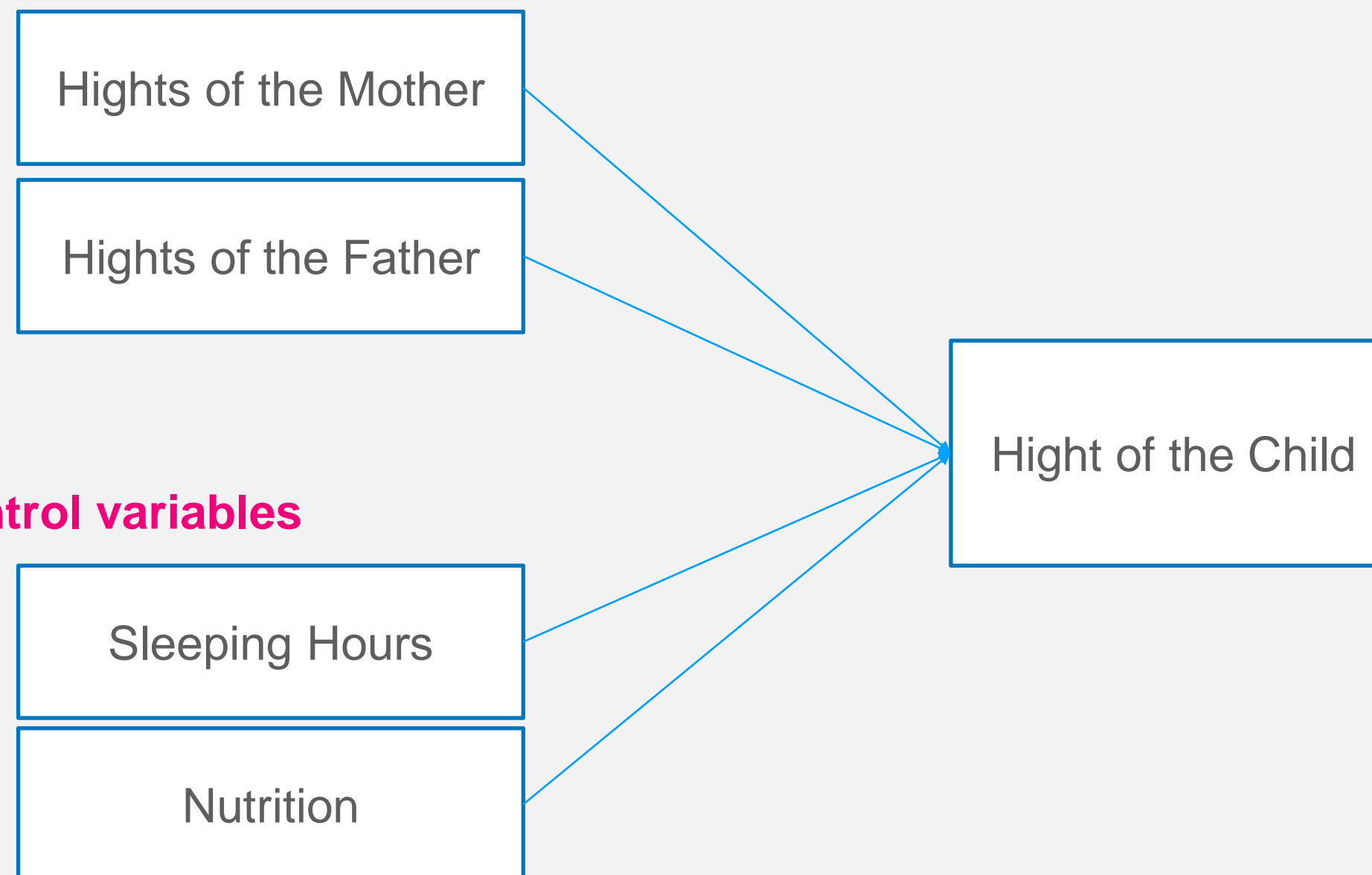


Control Variables

Collect your variables much as possible and deduct

- **Explanatory variables:** Variables used in the working hypotheses derived from the theory.
- **Control variables:** Variables other than the main explanatory variables believed to affect the dependent variable.

Main explanatory variables



R Programming Language in Political Economy

R is the most intuitive programming Language for Social Science

- **Versatility:** R is suitable for diverse tasks in political economy, including data manipulation, statistical modelling, and visualisation.
- **Open-source:** Free and accessible, R democratises the field and encourages research collaboration.
- **Comprehensive library:** R offers a vast collection of packages tailored for various applications relevant to political economy.
- **Community support:** An active community of users provides assistance and shares knowledge, making it easier for researchers to find resources.
- **Reproducible research:** R facilitates reproducible research, ensuring the integrity and reliability of findings in political economy.
- **Cross-disciplinary collaboration:** R's popularity across various fields enables collaboration and knowledge sharing among researchers in the interdisciplinary field of political economy.

Why do we analyze data in the social sciences?

The Importance of Data Analysis in Social Sciences

- **Descriptive Analysis:** Summarizing data characteristics to identify trends and patterns.
- **Explanatory Analysis:** Uncovering underlying causes and mechanisms of observed phenomena.
- **Predictive Analysis:** Forecasting future events and outcomes for better decision-making.
- **Hypothesis Testing:** Validating theories and contributing to the development of new ideas.
- **Evaluating Policies and Interventions:** Assessing the effectiveness and providing evidence-based recommendations.
- **Generalizability:** Concluding applicable to broader populations and contexts.
- **Enhancing Decision-Making:** Informing policymakers and organisations with evidence-based insights.
- **Advancing Knowledge:** Contributing to understanding human behaviour, social structures, and institutions.

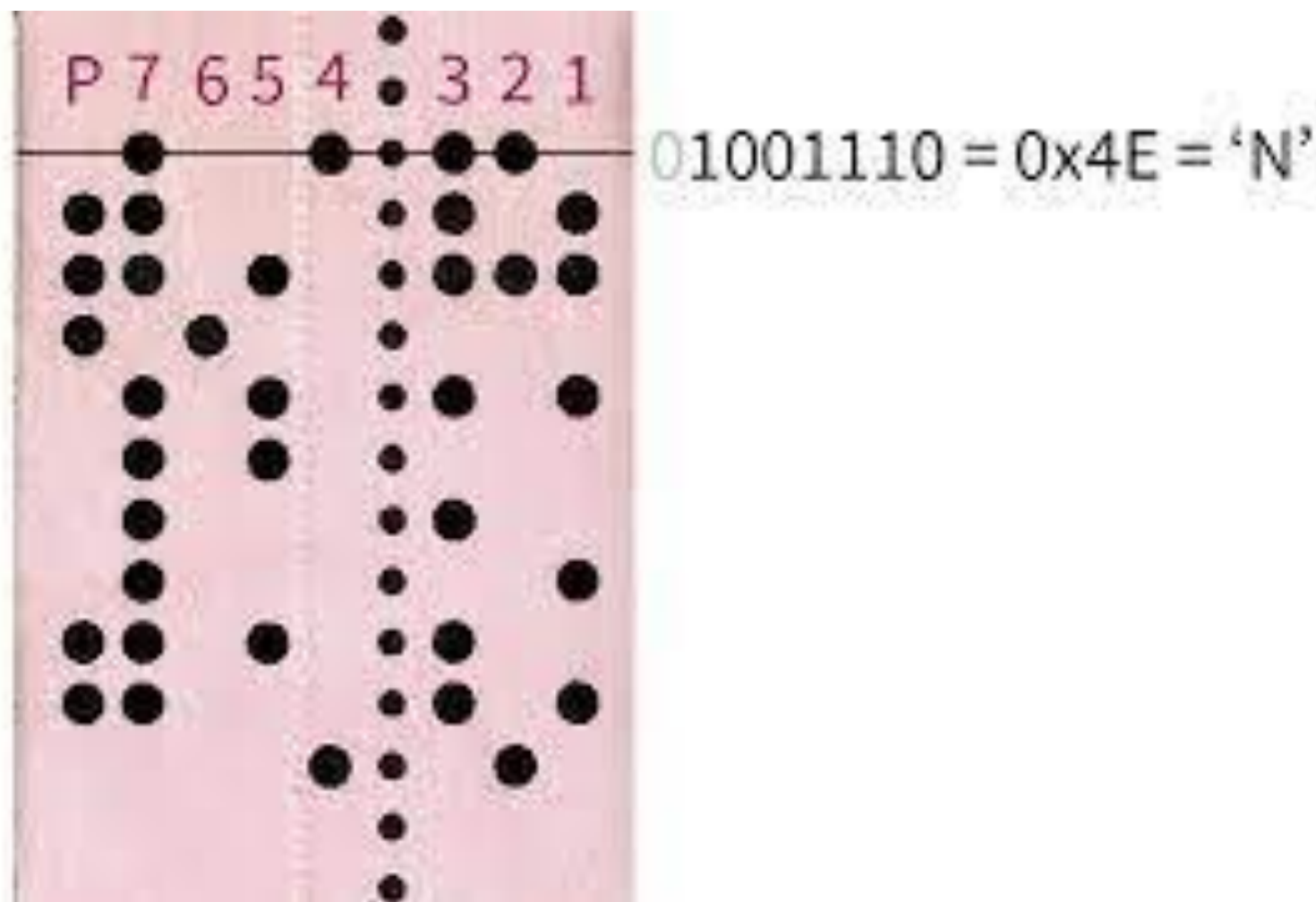
The Value of Data Analysis Skills in Social Sciences and Beyond

The Importance of Data Analysis in Social Sciences

- **Essential skills for social scientists:** Data analysis enables researchers to measure population characteristics, make predictions, and evaluate cause-and-effect relationships, answering critical questions in the field.
- **Applicable to various professions:** Even for non-social scientists, understanding data analysis and distinguishing between well-conducted and poorly-conducted quantitative studies is valuable.
- **Marketable skills in high demand:** With advancements in computing power and data availability, the need for data analysts who can inform decision-making in the public and private sectors has grown significantly.
- **Enhancing everyday decision-making:** Analytical skills can improve personal choices, such as voting or optimising productivity.
- **Building critical thinking abilities:** Learning the strengths and limitations of quantitative methods helps resist faulty inferences from data.
- **Becoming informed consumers of research:** In the era of big data, all individuals can benefit from being knowledgeable about quantitative research, regardless of whether they conduct research themselves.

Learning to code

The Importance of Data Analysis in Social Sciences



- 1. Purpose of coding:** Writing and running code enable data analysis by providing instructions for a computer to execute.
- 2. Programming languages:** Code is written in a specific programming language, with R being a popular choice among data analysts.
- 3. Accessibility of coding:** Learning to code is less daunting than it seems and can be enjoyable for many people.
- 4. Evolution of coding:** Unlike the 1940s when only highly trained mathematicians could code, today, anyone with access to a computer, some spare time, and patience can learn how to code.
- 5. R programming language:** R is a widely used language for data analysis, offering powerful tools and a strong community of users.

Getting Ready

Install R and Rstudio (IDE)



To download and install R, visit <https://cran.r-project.org/>, select the link corresponding to your operating system, and then follow the on-screen instructions.

Download R For Windows: <https://cran.r-project.org/bin/windows/>
Download R For MacOS: <https://cran.r-project.org/bin/macosx/>



To download and install Rstudio, visit <https://posit.co/products/open-source/rstudio/>, select the link corresponding to your operating system, and then follow the on-screen instructions.

Download Rstudio <https://posit.co/download/rstudio-desktop/>

Familiarize yourself with Rstudio.

Click "Run" to send your code to the console

The image shows the RStudio desktop environment. The top toolbar contains buttons for 'Run', 'Source', and 'Environment'. A pink circle highlights the 'Run' button, with an arrow pointing to it from the text 'Click "Run" to send your code to the console'. The interface is divided into four main panes, each with a white callout box explaining its function:

- Source**
 - This is where you write your code
 - Your code will not be evaluated
 - Until you "Run" them to the console
- Console**
 - This is where your code from the Source is evaluated by R
 - You can also use the console to perform quick calculations that you don't need to save
- Environment / History**
 - Here you can see what objects are in your working space (Environment) or view your command history (History)
- Files/ Plots / Packages/ Help**
 - Here you can see directories, view plots, see your packages, and access R Help.

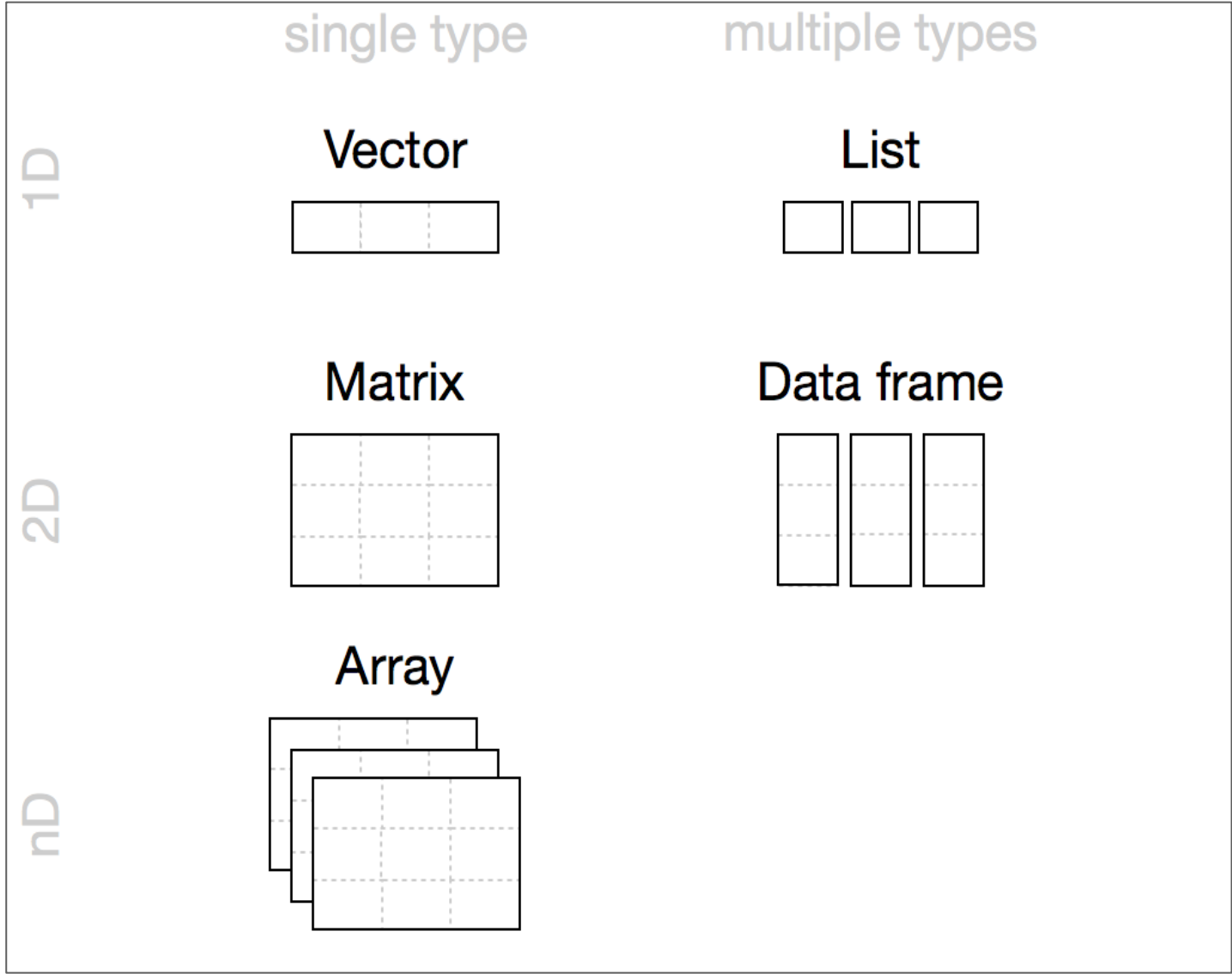
The bottom-left pane shows the R console output, including the R version (4.1.1) and some Korean text. The bottom-right pane shows the 'Files' tab, displaying a directory structure.

Introduction to R

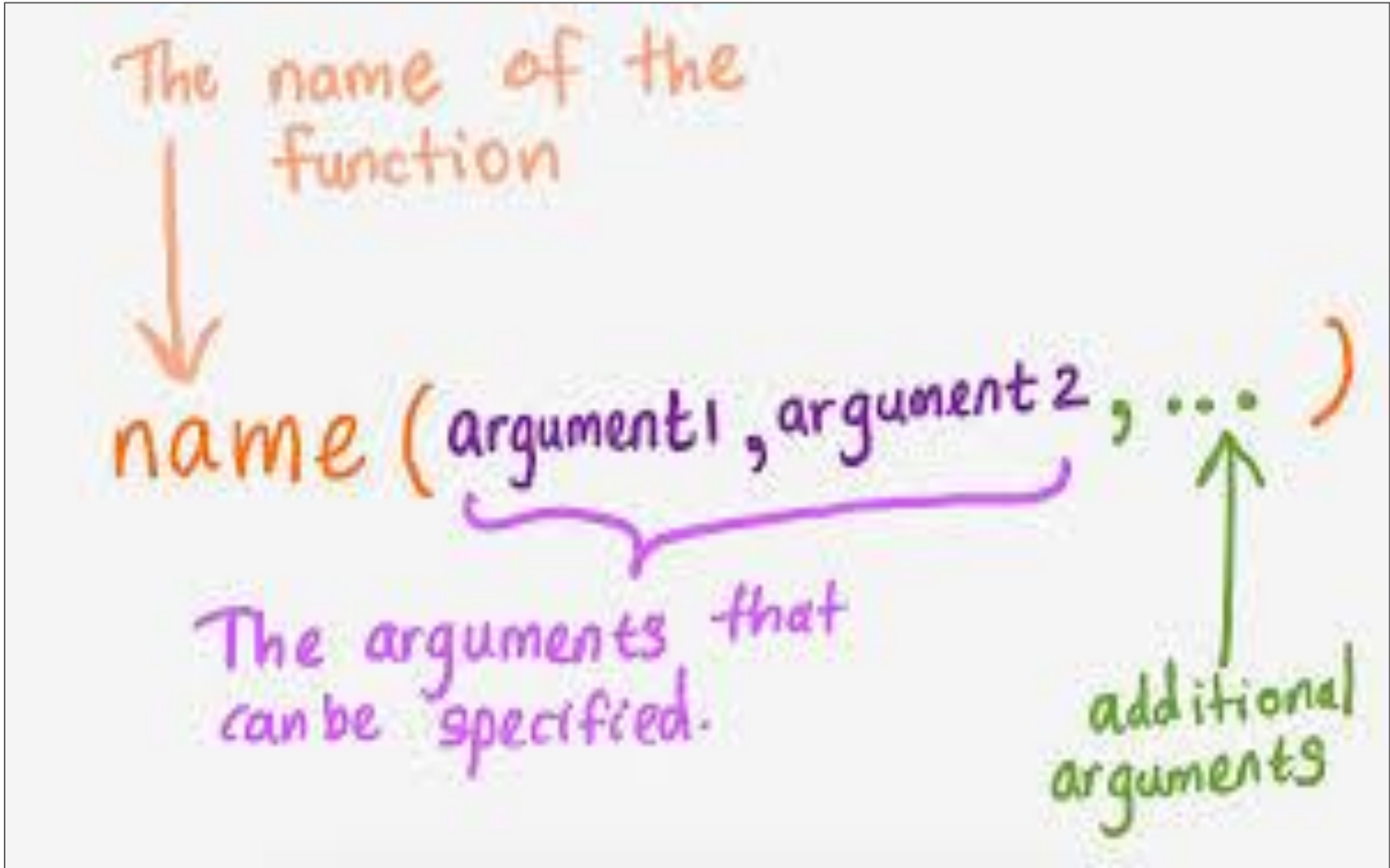
Calculations, Object Creation, and use functions



Calculations



Objects



Functions

Calculation

Use it as the calculator



Calculations

```
#We can use R as a calculator. R is capable of performing  
# summation (+), subtraction (-), multiplication (*), and division (/)  
# in addition to more complex mathematical operations.  
# For instance, the code to ask R to add 1 and 3 is: 1+3.
```

```
1 + 6
```



```
[1] 7
```

Calculation

Use it as the calculator

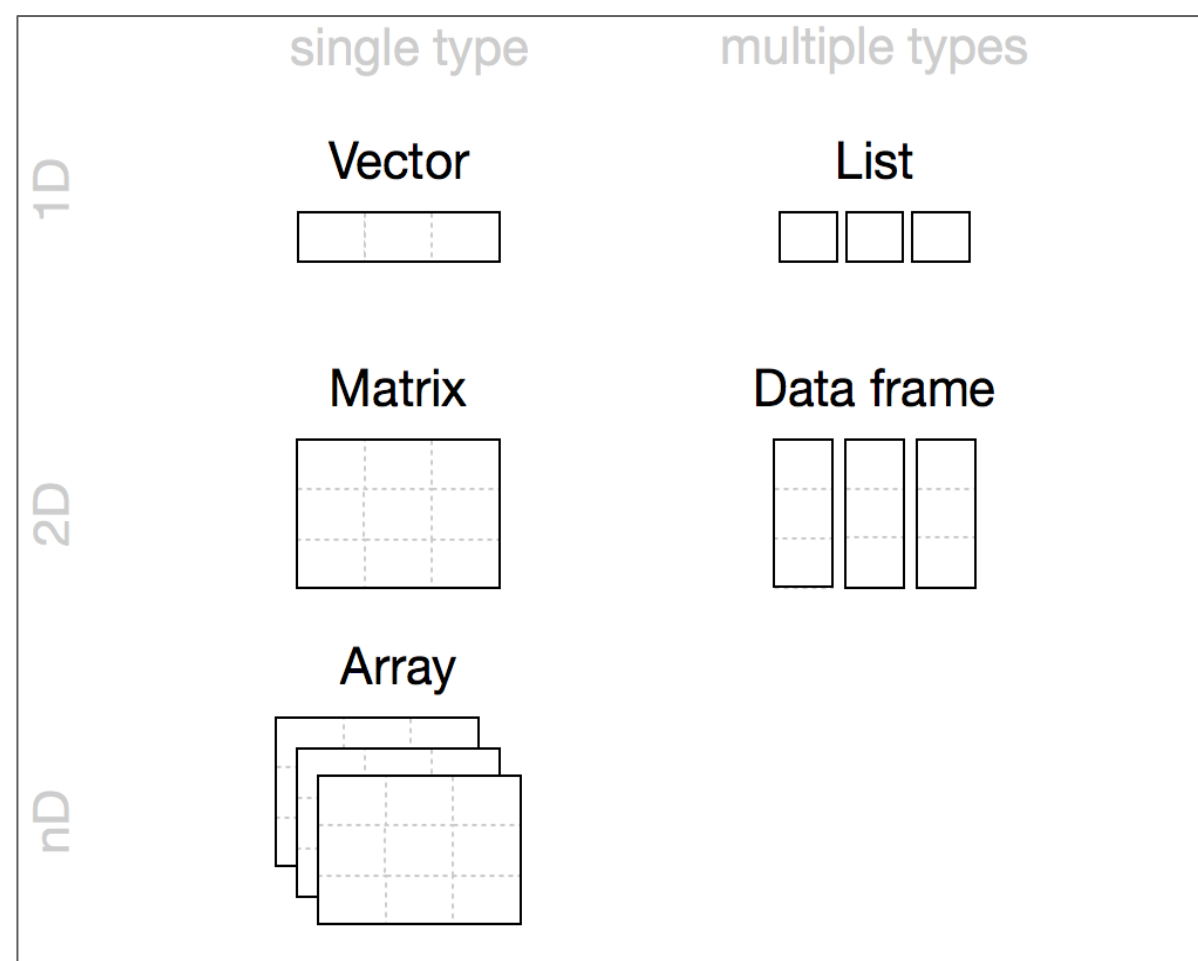


Calculations

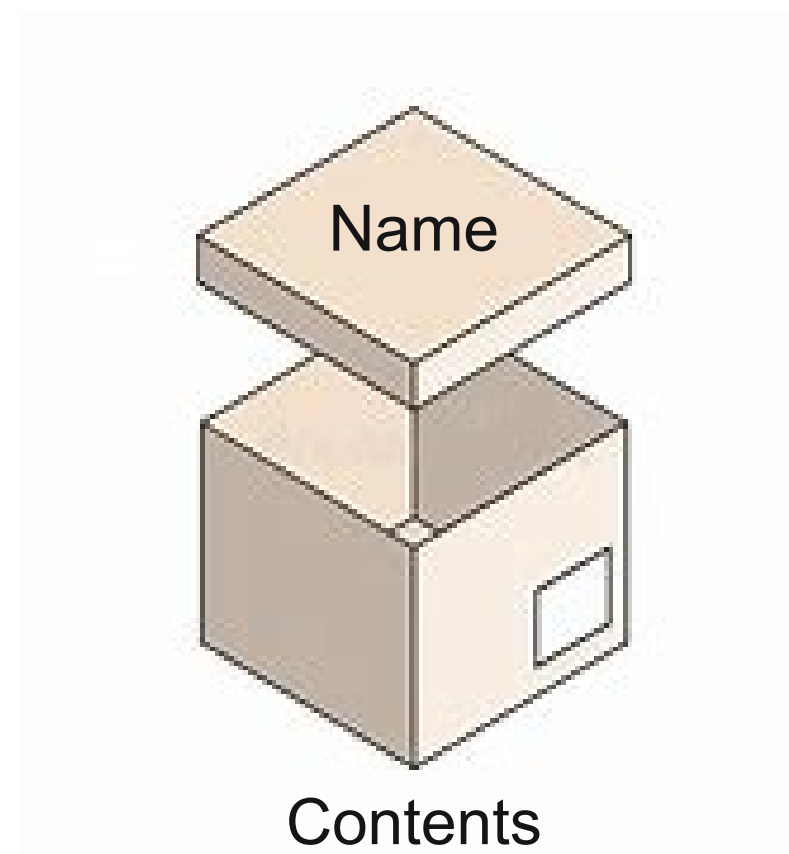
1 + 2: Summation
3 - 4: Subtraction
5 * 6: Multiplication
7 / 8: Division
2^3 : Square
sqrt(4) : Square root
log(2) : Natural logarithm

Creating Objects

We need to load and store datasets



Objects



```
# Creating Objects: to store data as an object  
# in R, we write the code with the following format  
# object.name <- object.contents
```

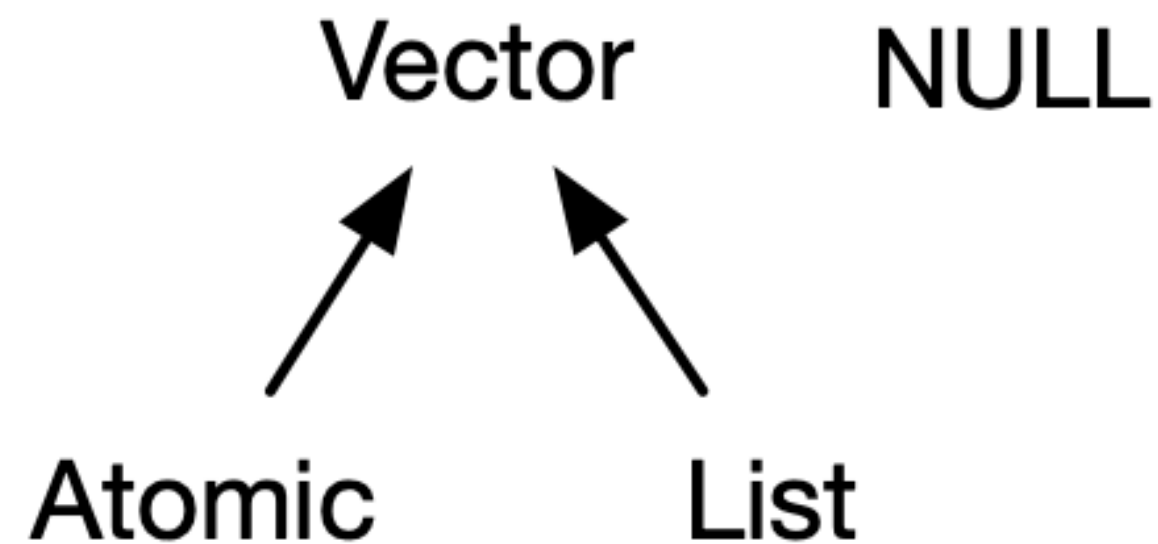
```
# object.name is the name we want to give the object  
# <- is the assignment operator, which creates an object by  
# assigning contents to a name  
# object.contents is the data we want to store in the object
```

```
seven <- 1 + 6  
seven
```



Creating Vectors

A vector is a basic data structure



Here is an example where more than one result is displayed: in R, a vector can be created with the function
For example, the vector (1, 3, 5) can be constructed by `c(1,3,5)`.
Adding 1 to each element of this vector can be calculated with the command: `vector1 + 1`

```
vector1 <- c(1,3,5)
vector.add.1 <- vector1 + 1
vector1
vector.add.1
```



```
> vector1
[1] 1 3 5
> vector.add.1
[1] 2 4 6
```

Creating Vectors

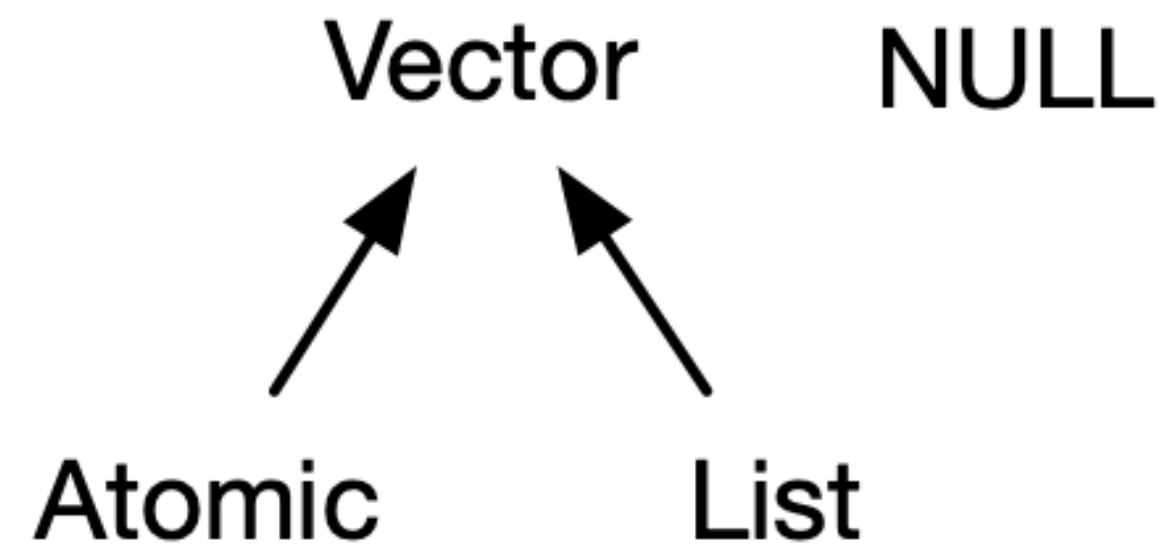
Seq() generates the sequence of numbers

The seq() is a built-in R method that generates the sequence of
numbers. The basic syntax of the seq() function is seq(from=1,
to=1, by=1, length.out=NULL, along.with=NULL), where “from” is # the
starting value of the sequence, “to” is the end value of the # # sequence,
and “by” is the value to increment by. The default is 1.

```
seq.num <- seq(0, 5)  
seq.num
```

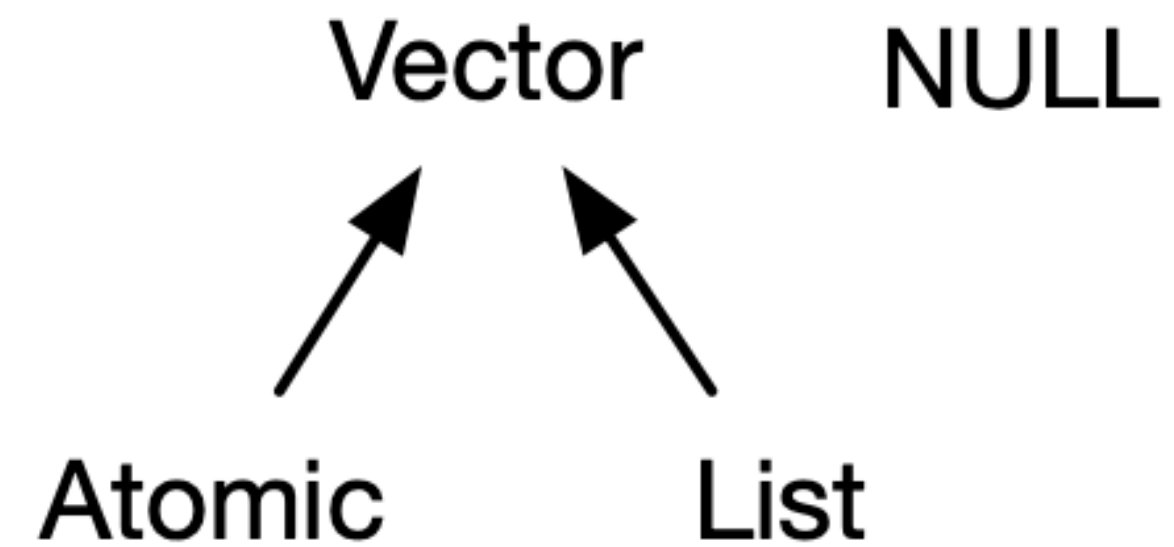


```
[1] 0 1 2 3 4 5
```



Creating Vectors

`Seq()` generates the sequence of numbers



```
# Easy way to calculate the mean and median in R.
```

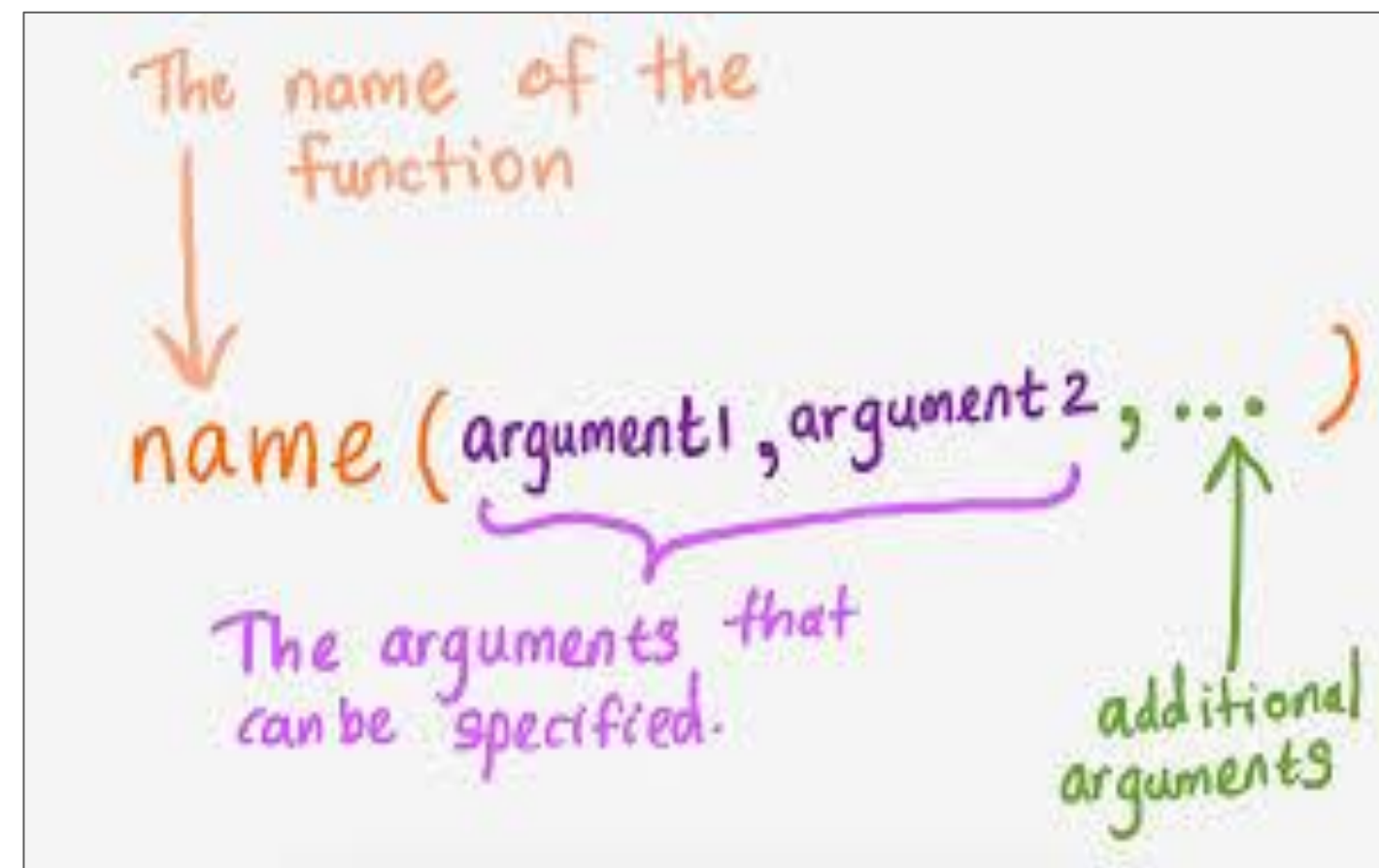
```
hello <- "Hello World"  
Hello
```



```
[1] Hello World
```

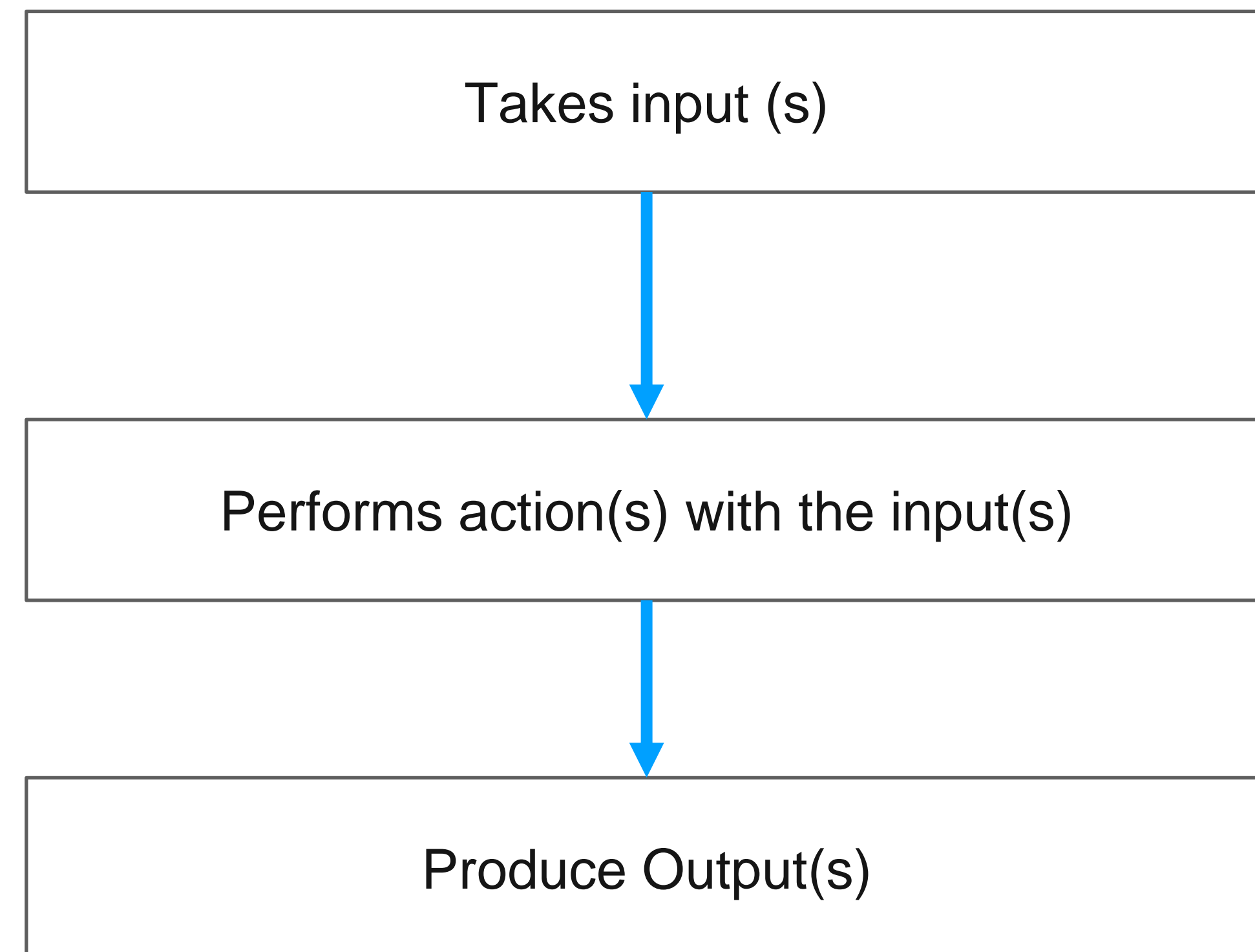
Using Functions

interact with data



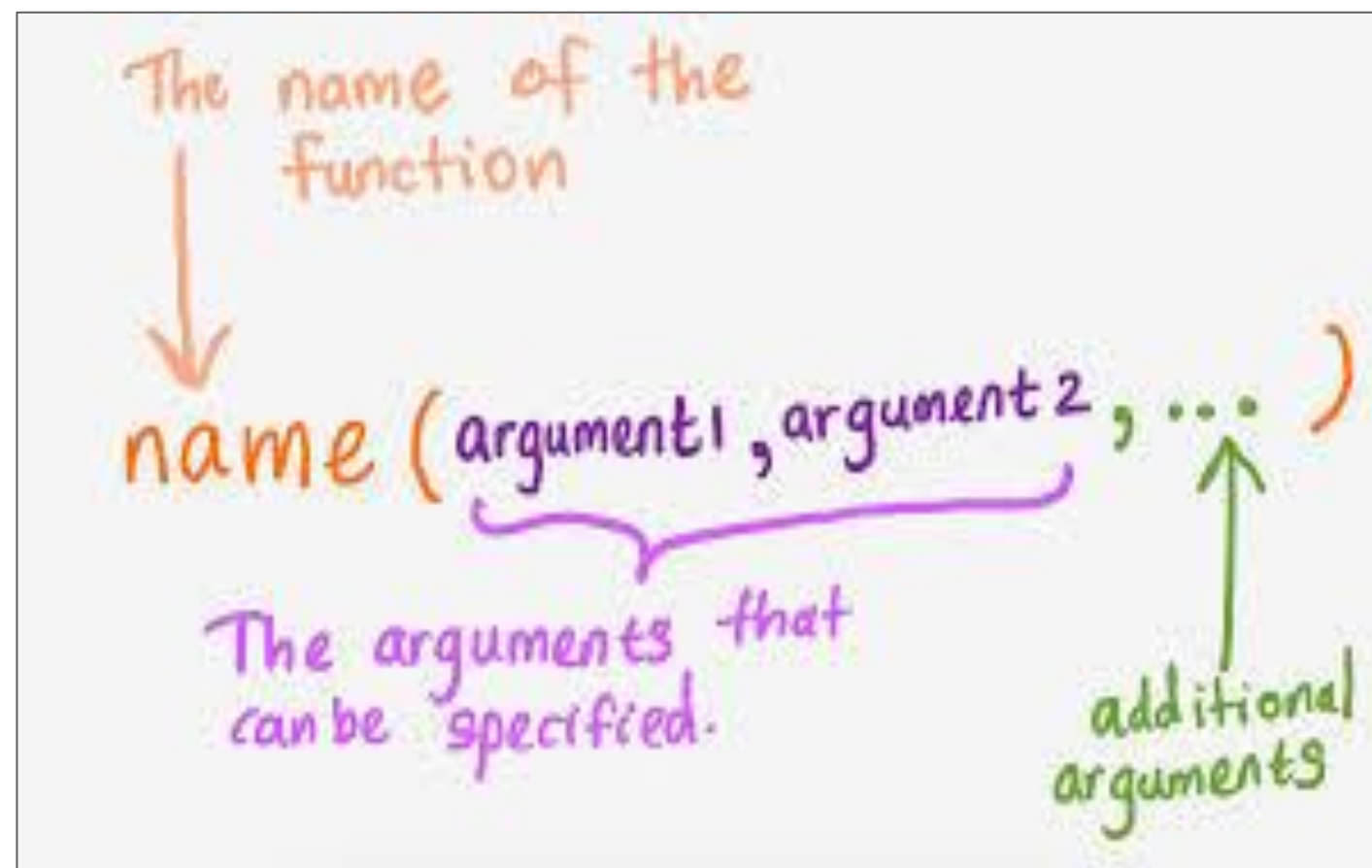
Functions

An R Functions



Using Functions

interact with data



Functions

Using R functions

To use a function in R, we typically write code in one of these two formats:

- (a) `function_name(required_argument)`
- (b) `function_name(required_argument, optional_argument_name = optional_argument)`

where:

function_name is the name of the function; for example, “mean” is the name of the function that computes the mean of a set of values

Required_argument is the argument the function requires, such as the values we want to calculate the mean of; we typically do not include the names of required arguments; we enter the required arguments first, and if there is more than one, we enter them in the order expected by R

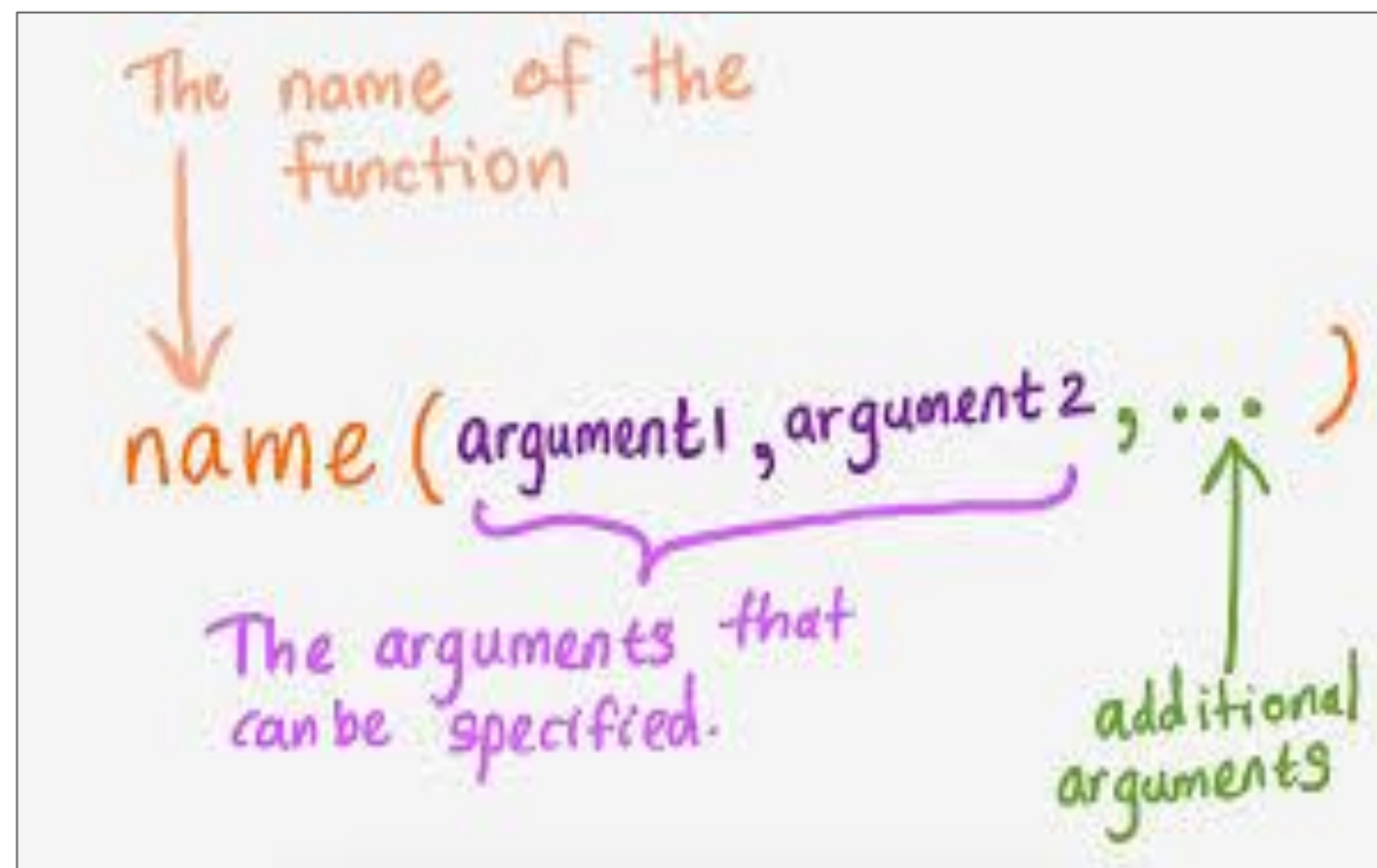
, is a comma, which we use to separate different arguments

optional_argument_name is the name of the optional argument we want to use, such as the argument that enables us to eliminate missing values before calculating a mean

optional_argument is what we set the optional argument to be.

Using Functions

interact with data



Functions



Create a vector

```
age <- c(23, 27, 51, 35, 41)
```

getting means from age

```
age.mean <- mean(age)
```

getting median from age

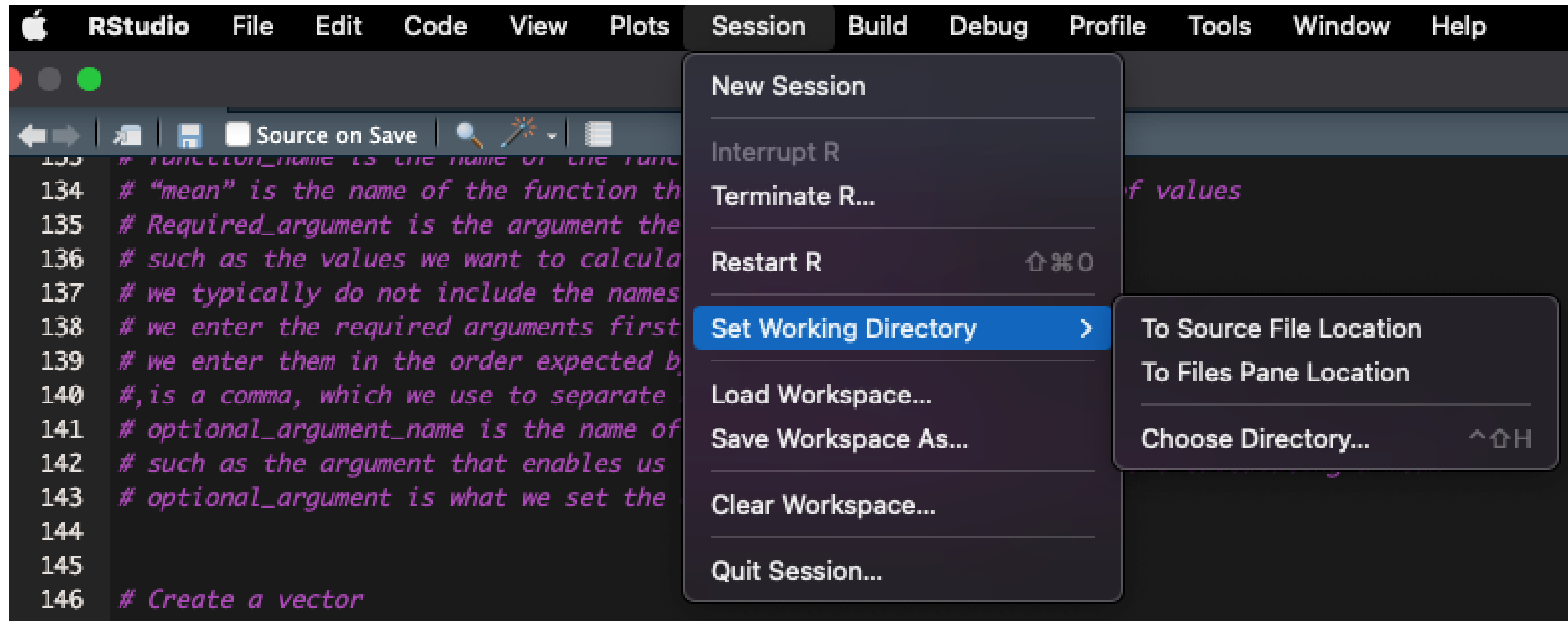
```
age.median <- median(age)
```

```
age.mean  
age.median
```



Set the Working Directory

interact with data



Session > Set Working Directory > To Source File Location.

set the working directory via
using the function setwd()

example of setwd() for Mac
setwd("~/Desktop/DSS")

example for Windows
setwd("C:/user/Desktop/DSS")

You can confirm the working #
directory with
getwd()

Installing libraries and load

Using Libraries

```
# prepare the library (install)
```

```
install.packages(readxl)
```

```
install.packages(tidyverse)
```

```
#loading the library
```

```
library(readxl)
```

```
library(tidyverse)
```

```
#alternative method
```

```
if(!require('tidyverse'))install.packages('tidyverse')
```


Load the dataset

Choose a File Interactively

You can use the `file.choose()` function in R to bring up a file explorer window that allows # you to choose a file path to work with interactively.

To use this function, type the following into your R console

```
file.choose()
```

Load a excel file

```
SourceFile.01 <- read_excel(file.choose())
```

Load a CSV file

```
SourceFile.02 <- read_csv(file.choose())
```

Load a TXT file

```
SourceFile.03 <- readtext(fileToLoad)
```

Summary

Good Theory

- **Empirical Analysis Process:** Building theories, deriving hypotheses, and verifying with data; correlation does not equal causation.
- **Good Theory Characteristics:** Falsifiable, predictive, specific, and straightforward.
- **Hypothesis Types:** Theoretical and operational; operationalisation of variables.
- **Unit of Analysis:** Consistency is needed; it affects data and analysis.
- **Explanatory Variables:** Primary explanatory and control variables.
- **Refuting Theories:** Counter-theory, counter-hypothesis, and statistical analysis.
- **Proper Operationalization:** Accurate measurement of concepts is essential.

Summary

Good Research Question

- Good research questions are necessary to complete a good research paper.
- Research questions are questions about the social phenomenon under study.
- The research question is central to the research design, sets the overall direction of the research and determines the quality of the research.
- Research questions can be divided into three categories according to the type of question: empirical, normative and analytical.
- Normative questions are so-called 'ought' questions, which ask what is desirable and what is not.
- It is possible to extract and examine the core issues as empirical problems from the many important issues raised by the 'ought theory'.
- Analytical problems examine the validity of propositions that are more abstract than the facts as they actually occur.
- When considering an empirical problem, the researcher does not make his or her own value judgements but 'uses observed information' to make a judgement.
- A 'good research topic' satisfies five conditions: clarity, verifiability, theoretical importance, practicality and originality.
- Eight elements make up a research paper: introduction, previous research, presentation of theory, presentation of hypothesis, presentation of counter-hypothesis, explanation of working hypothesis, presentation of evidence, and conclusion.

Next Week

Correlation and Causation