Designing Social Science Research

Methodological process and Research questions

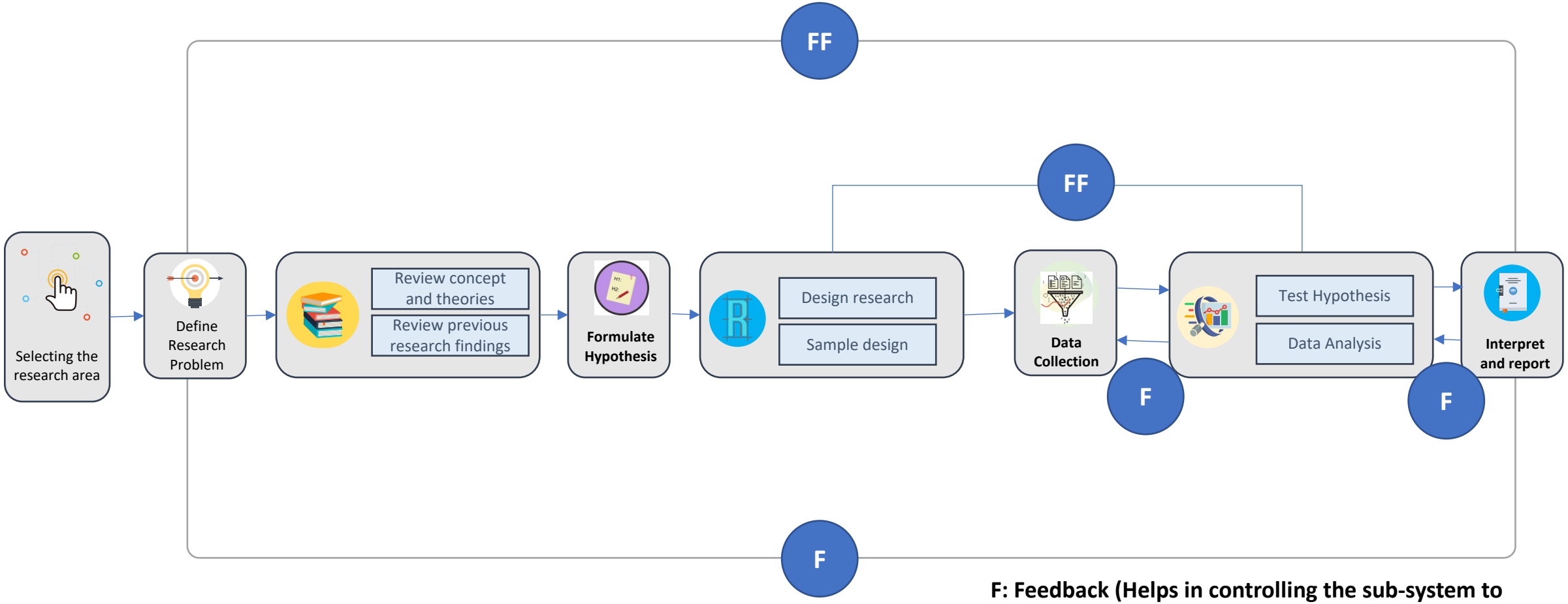
Week 03

Week3: Research Designing Social Science

For example, studies in the social sciences focus on methodological reflections on empirical research and conce ptual integration of chosen methodologies.

- Methodological process
- Research questions
- Source Data: Selection and Procurement
- Pre-processing Preparations
- Analysis

Methodological Process of the Research



F: Feedback (Helps in controlling the sub-system to which it is transmitted)
FF: Feedforward (Serves vital function of providing criteria for evaluation)

Research Problems

Three things you should be aware of.

- Your research should either fix a problem or expand our understanding.
- Your answer should be new one.
- The research should be logical and timely

Research Problems and Hypothesis

Example 1: The perspective difference between media types.

Importance: The differences between online and print media potentially lead to a generation gap in perceptions based on news media content. It is, therefore, important to know more about the content of these news media.

Literature Review:

Research Questions: Could media reporting provide different aspects of immigration between popular online and print media?

RQ1: How do aspects of interactivity, convergence, and immediacy affect frame usage, and (how) does online news differ from print news on the immigration crisis?

Hypothesis:
H_0: Expect no differences between online and offline news media.
H_1: Online news applies stronger frames than its print counterparts.

Research Problems and Hypothesis

Example 2: Effect of public policy on Internet use

Importance: Public ICT policy planning and its interventions appear to be essential for technological adaptation and the spread of technology. But little or no real-world research has been done to determine how these policy plans will work.

Literature Review:

To some extent, the impact of the regulatory policy has been studied (Hargittai, 1999; Guille n & Sua rez, 2001; Hawkins & Hawkins, 2003).

Research Questions: Does public policy planning mirror or predict the spread of technologies?

RQ1: Is it possible to know how well public plans will work between the planning and target times?

Hypothesis:
H_0: _____
H_1: _____

Research Problems and Hypothesis

Example 3: Rentier and Sovereign Sustainability

Importance: Effective state subsidy management is one of the most significant predictors of rentier states' citizen satisfaction and anti-government control.

Literature Review: Numerous studies indicate that state subsidies and the provision of foreign domestic labour contributed to the reduction of government discontent.

Research Questions: Does government subsidy influence levels of discontent with the government?

RQ1: Does government subsidy affect dissatisfaction with the government?

Hypothesis:

H_0: _____

H_1: _____

Research Problems

HOW TO DO

Step 1: Select a topic that interests you.

(for example: leadership, social framework, digital exclusion, leapfrogging, distribution, and smart city...)

Step 2: Narrow it down further until the literature becomes manageable.

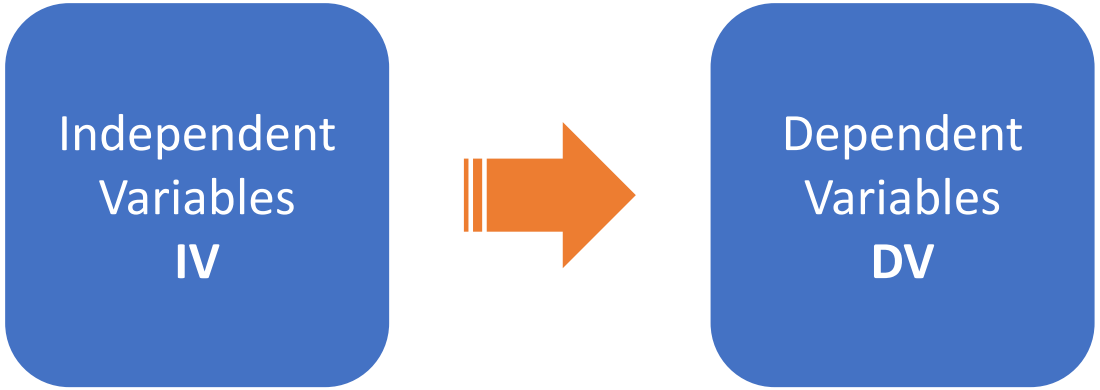
(e.g., the impact of policy planning and technical adaptation, cause of A and B..., causal relationship between A and B...)

Step 3: Read the literature and formulate an important, novel, and timely research question.

Research Problems

Goal; Formulate Your research Question…

- In terms of **cause and effect**
- Using **Clear and measurable** concepts



Bad Research Questions

- **What should news organisations do?**
- How do the news media frame the political arena?
- Does factor X lead to news media being conservative (Y) ?

- **What should business do**
- What increases sales of business?

- Does strategy X lead to more sales than strategy Y?

Types of Variables

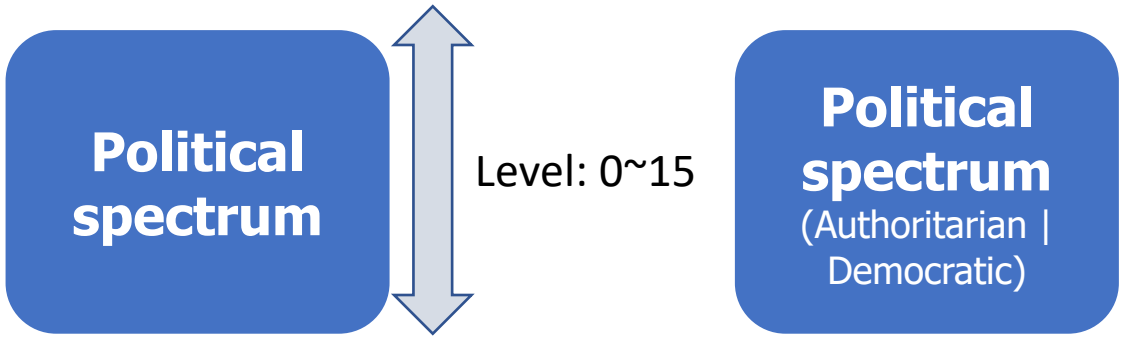
Variable: Something that varies/has several level

Categorical variable: something that varies without any ordering of the level

- Gender (male, female)
- Occupation (teacher, researcher, student, policeman, ….)
- Nationality (Korean, British, Japanese, American, ….)

Numerical variables: something that varies from low to high

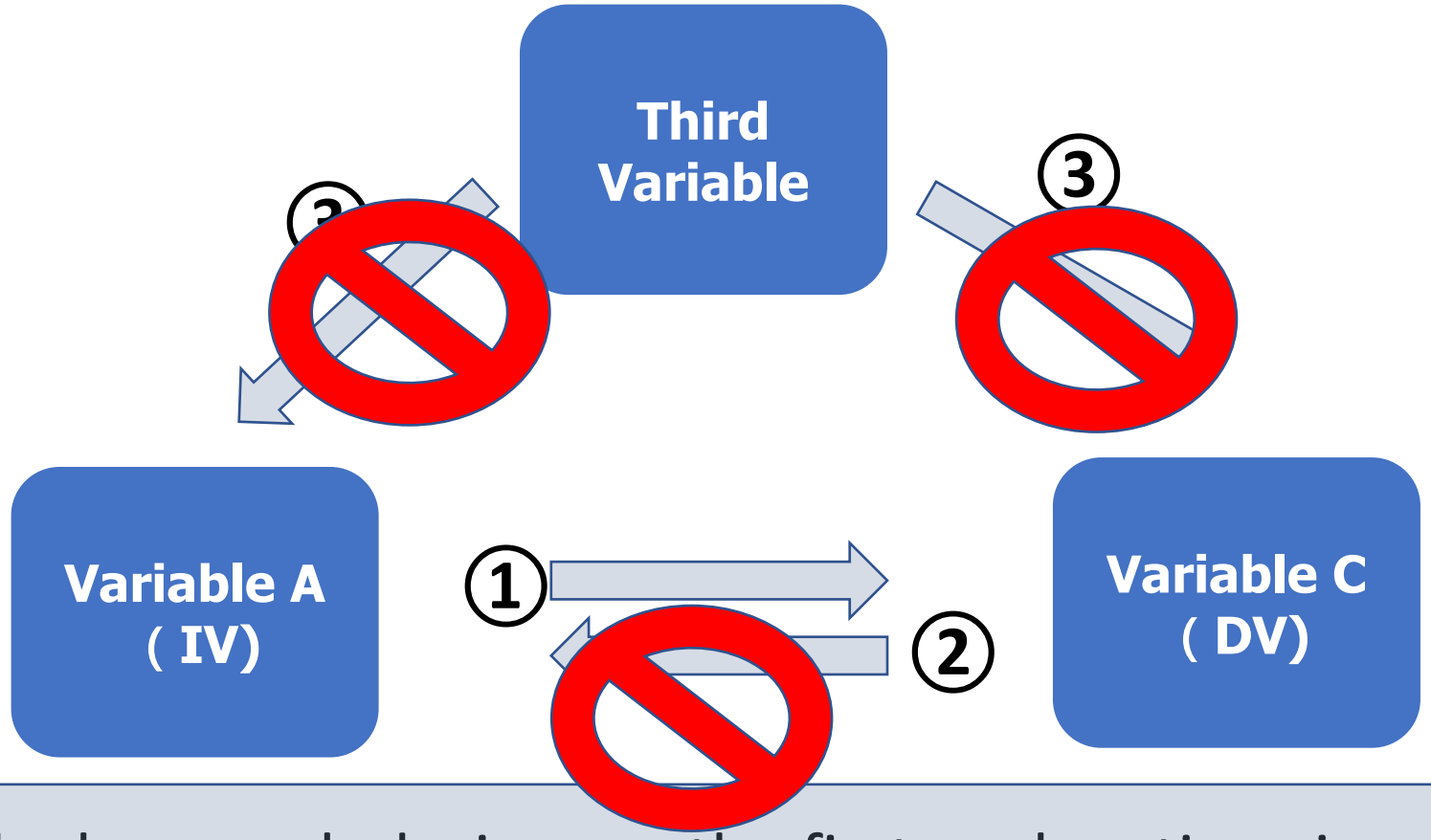
- Number of friends (1, 2, 3, 4, …)
- Amount of sales (1, 2, 3, 4,….)
- Body weight (1, 2, 3, 4,….)



Types of Variables

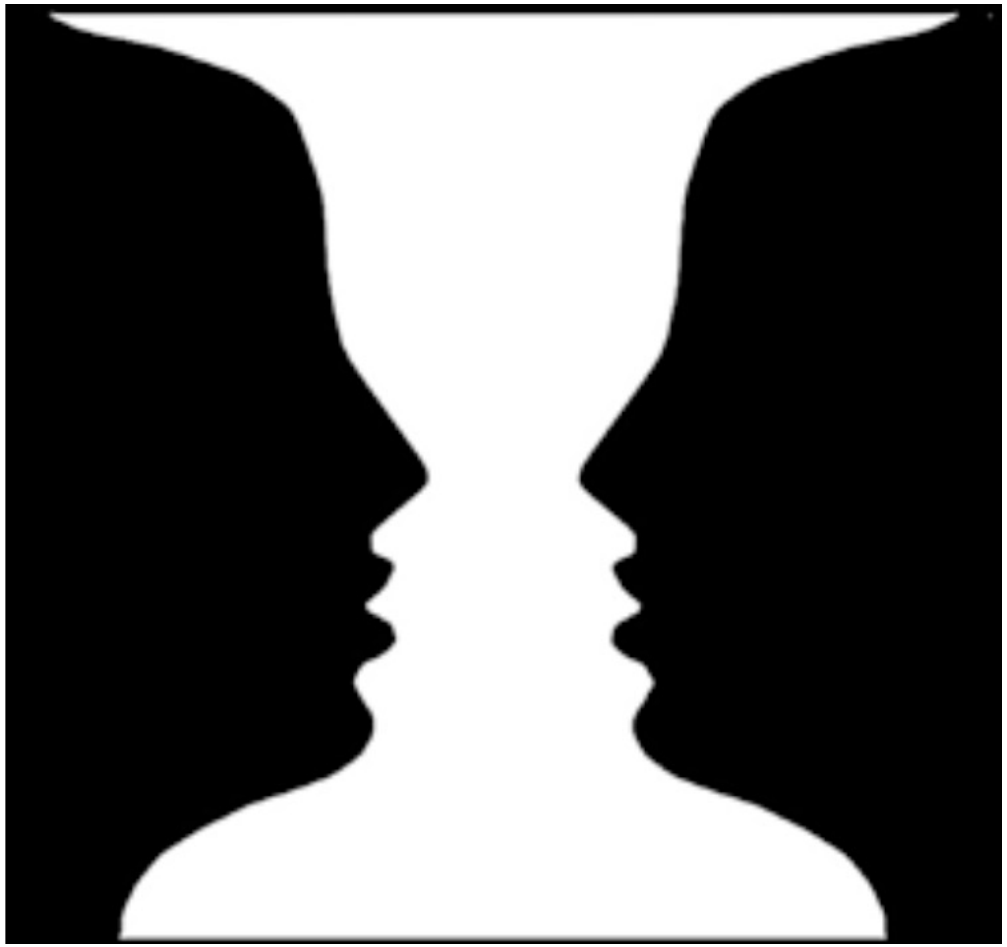
There are THREE main explanation for a statistical relationship

1. **The IV influenced the DV**
2. **The DV influenced the IV**
3. **A third variable influenced the IV and the DV**

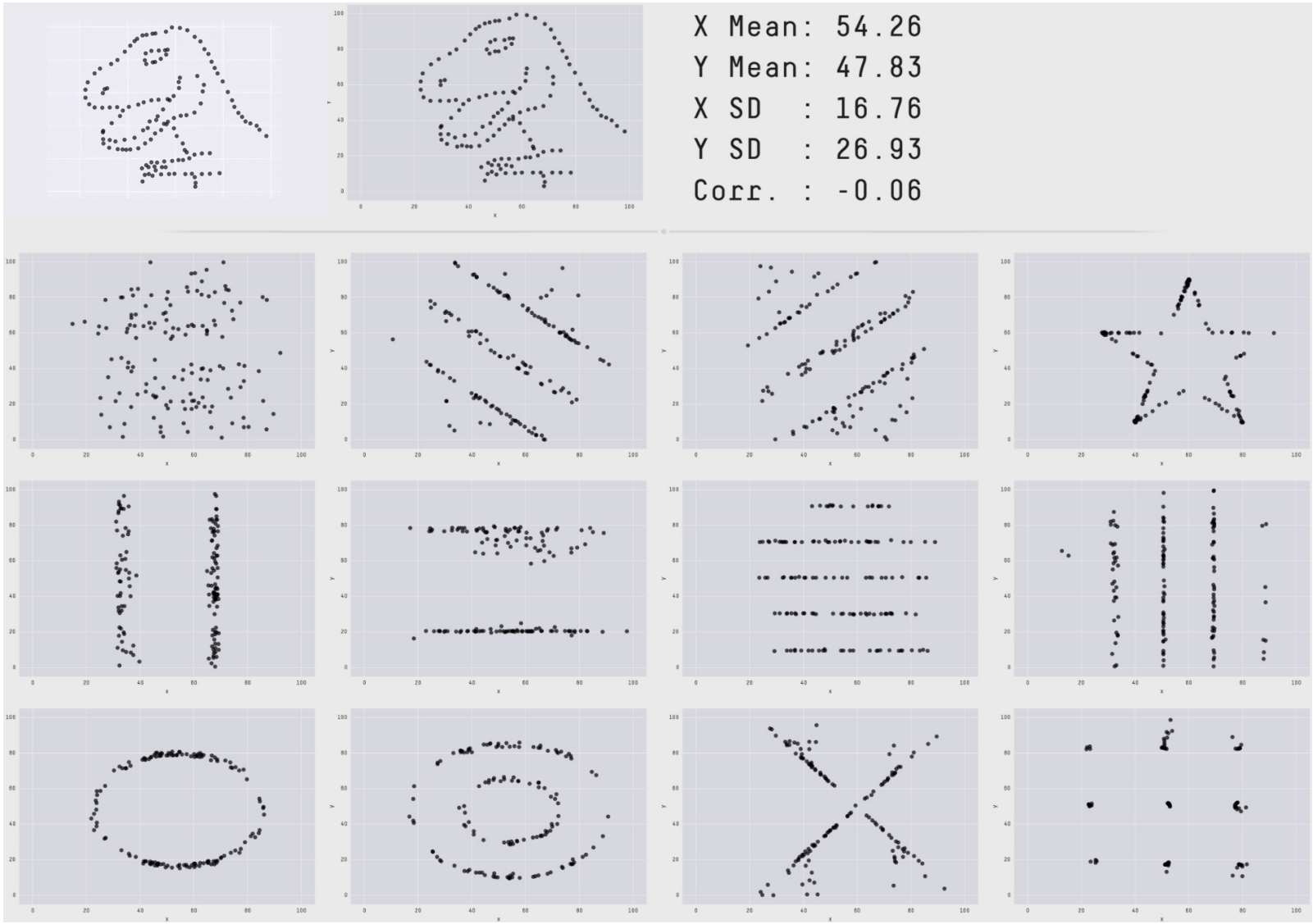


In and ideal research design, on the first explanations is possible

What this look like



These have same descriptive statistics



https://www.autodesk.com/research/publications/same-stats-different-graphs

R Programming

Methodological process and Research questions

Week 03

```r
#Load Library
library(dplyr)
library(readxl)

# The lecture recommends using a google style R guide
# https://web.stanford.edu/class/cs109l/unrestricted/resources/google-style.html



# means comment
# integer
# Use L if you are not going to be adding then or you're never going to be creating, arithmetic operations, Just Put L after the number.

x <- 2L

# typeof describes the types of  x have and in order to check the type all you  have to do is type in "typeof ()”

typeof(x)
```

```
# double

y <- 2.5
Typeof(y)

# complex

z <- 3 + 2i

typeof(z)

#character
a <- "h"
typeof(a)
```

```
#logical
q1 <- TRUE # or T
typeof(q1)
q2 <- FALSE # or F
typeof(q2)

# Using Variables
# How to save: Just Click Control + S (Windows), for MAC Command + S

A <- 10
B <- 5
C <- A + B
C
print(C)

# Multiple selections, use your mouse to highlight.
```

```r
# variable 01
var1 <- 2.5
# variable 02
var2 <- 4

results <- var1 / var2
results

# sqrt is a function and the brackets
# indicate that the values inside are
# being passed on to this function.

answer <- sqrt(var2)
answer
```

```r
# paste(): Takes multiple elements from the multiple vectors and concatenates them into a single element.
# paste0(): The paste0() function has space as its default separator and limits your opportunities in the output as well.

greeting <- "Hello"
name <- "Bob"
message <- paste (greeting, name)
message

# > paste(1,'two',3,'four',5,'six')
# [1] "1 two 3 four 5 six"
# > paste0('df',1:5,collapse = '_')
# [1] "df1_df2_df3_df4_df5"
```

```
# Logical Operators
# TRUE T
# FALSE F

4 < 5
10 > 100
4 == 5
```

```
# Logical Expression

# == equal to
# != not equal to
# <
# >
# <=
# >=
# ! not
# | or
# &
# isTRUE(x)
```

```
results <- 4 < 5
results
typeof(results)



results2 <- !(5 > 1)
results2

results | results2
results & results2

isTRUE (results)
```

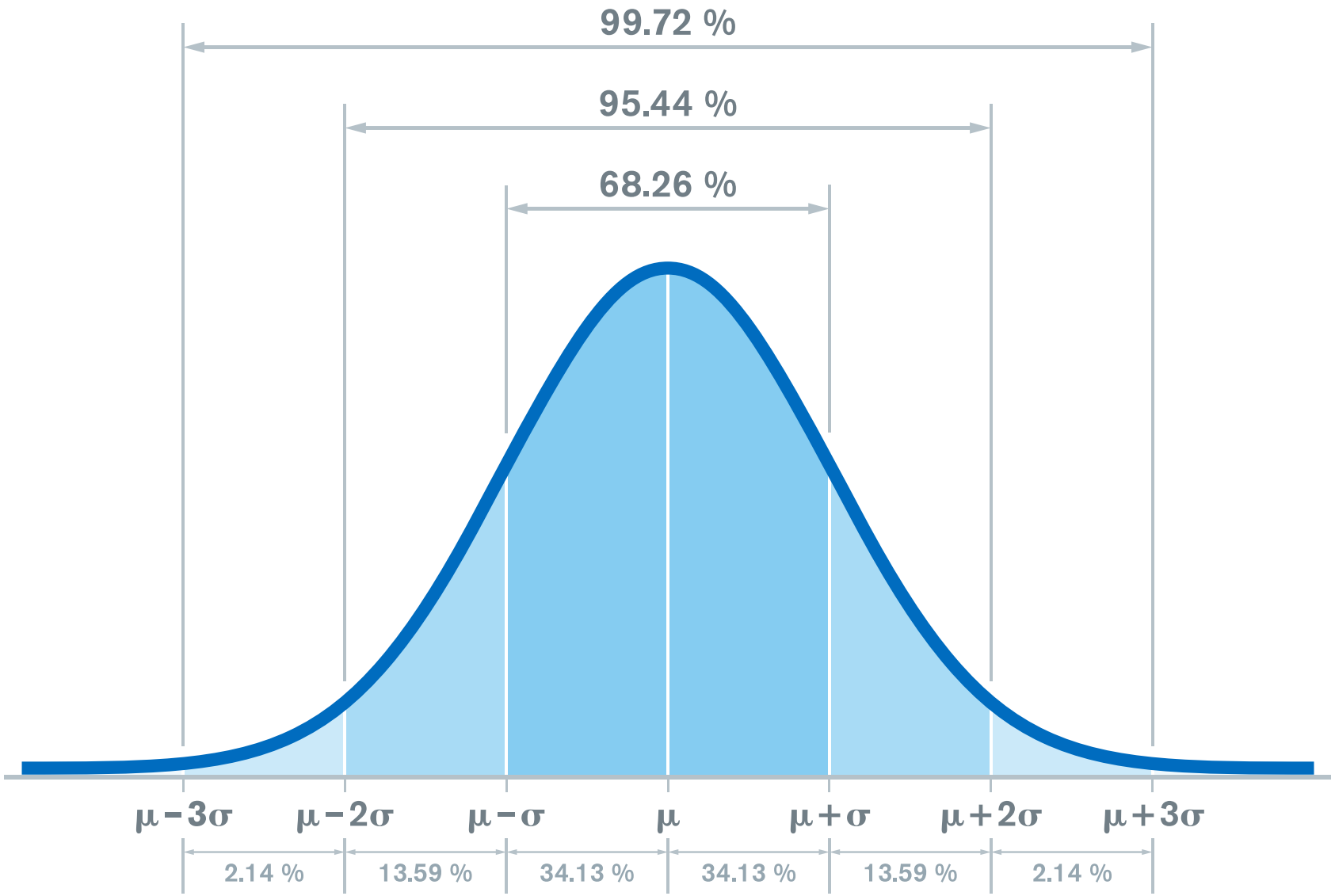
```
# CREATING LOOP

4 < 5
10 > 100
4 == 5

while(FALSE){

  print("Hello")

}
>
```

```
while(FALSE){

  print("Hello")

}

[1] "Hello"
[1] "Hello"
[1] "Hello"
[1] "Hello"
[1] "Hello"
[1] "Hello"
[1] "Hello"
[1] "Hello"
>

# to stop press ESC
```

```
#  You to perform certain actions such as loop
# for iterating certain operations
counter <- 1
while(counter < 7 ) {
  print(counter)
  counter <- counter + 1

}
[1] 1
[1] 2
[1] 3
[1] 4
[1] 5
[1] 6
```

```
# For Loop

for(i in 1:5){
  print("Hello World")

}
[1] "Hello World"
[1] "Hello World"
[1] "Hello World"
[1] "Hello World"
[1] "Hello World"
>
```

Normal Distribution



```r
# IF Statement
# ---- -2 ---- -1 ---- 0 ---- 1 ----- 2 -----
# rnorm: Vector of normally distributed
random numbers.

rnorm(1)

# 2 nested statement

rm(answer)
x <- rnorm(1)
if(x > 1){
  answer  <- "It is Greater than ONE"
} else {
  answer <- "It is Less than 1"
}
```

```r
# 3 nested statement

rm(answer)
x <- rnorm(1)
if(x > 1){
  answer  <- "It is Greater than ONE"
} else {
  if (x >= -1){
    answer <- "Between -1 and 1"
  } else {
    answer <- "Less than -1"
  }
  }
```

```r
# 3 nested statement

rm(answer)
x <- rnorm(1)
if(x > 1){
  answer  <- "It is Greater than ONE"
} else {
  if (x >= -1){
    answer <- "Between -1 and 1"
  } else {
    answer <- "Less than -1"
  }
  }
```

```r
# Chain Statement

rm(answer)
x <- rnorm(1)
if(x > 1){
  answer  <- "It is Greater than ONE"
} else if (x >= -1) {
    answer <- "Between -1 and 1"
  } else {
    answer <- "Less than -1"
  }
  }
```

```
# Manipulating Data Frame

hmnrghts<-read.table(file.choose(),
                     header=TRUE, na="NA")
head(hmnrghts)

afganistan <- hmnrghts[hmnrghts$country=="afganistan",]
afganistan <- hmnrghts[hmnrghts$country=="afganistan", c("country","democ")]

aletter <- hmnrghts[hmnrghts$country==c("afganistan", "albania"),]
```

```
# Test the law of large number for N random normally distributed numbers with
# mean = 0, Standard Divination (stdev) = 1

# Create an R script that will count how many of these numbers
# fall between -1 and 1
# and divide by the total equation of N

# You know that E(X) = 68.2%
# Check that Mean (XN) -> E(X) as your return script while increasing N

N <-  _                      #Number of degrees of freedom (Sample Size)
counter <- _                 #Reset the counter
for (i in rnorm (N)){        #interate over vector numbers
  if (_ _ _ & _ _ _) {       #Check where integrated variable falls
    counter <- _ _ _         #increase counter if the condition is met
  }
}
answer <- counter / N     #Calculate hit-ratio
print(answer)             #print answer in console
```

```r
# Reference answer


N <-  100                          #Number of degrees of freedom (Sample Size)
counter <- 0                       #Reset the counter
for (i in rnorm (N)){              #iterate over vector numbers
  if (i >  - 1 &  i < 1) {         #Check where the iterated variable falls
    counter <- counter + 1         #increase counter if the condition is met
  }
}
answer <- counter / N     #Calculate hit-ratio
print(answer)             #print answer in console
```

Next Week

Please bring your laptop.

Week 4: The Basics and Practice (Causality)
Examples of data mining that can be used in the social sciences in Africa and Middle Eastern fields is

- Logical values and operations and simple conditional statements with factors
- Causal effects and the counterfactual
- Descriptive statistics
- Exercises