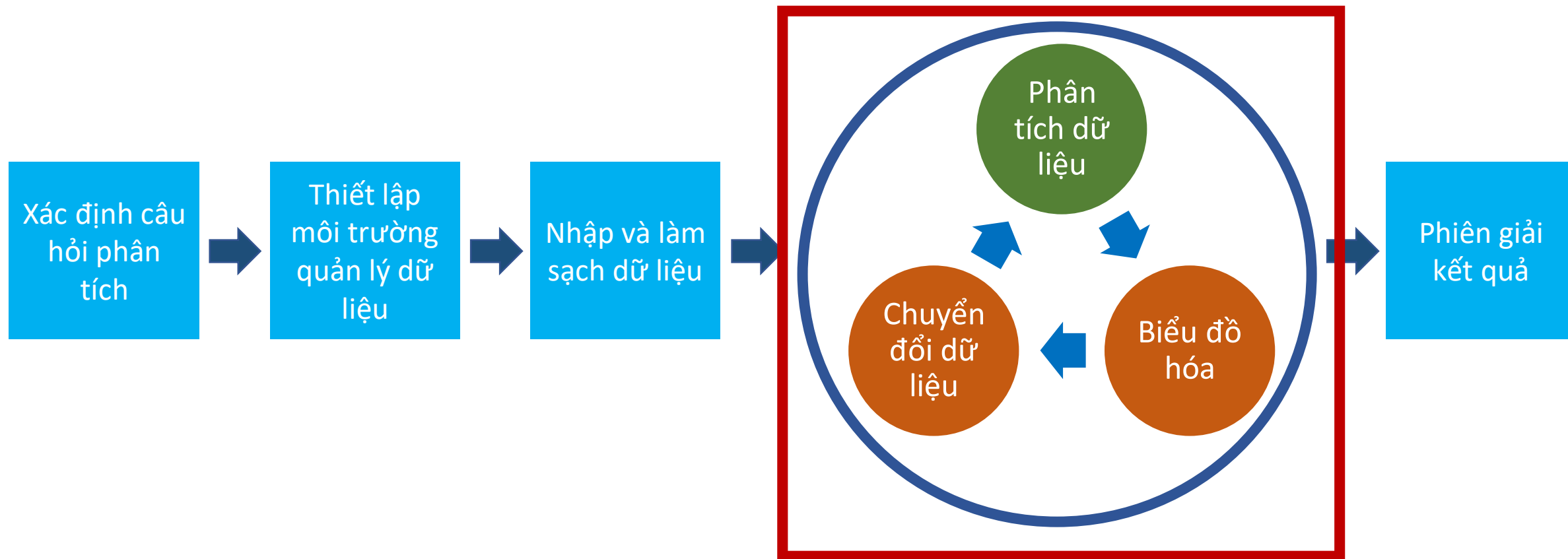


# Bài 3: Xử lý và phân tích mô tả với phần mềm R

# R Quy trình phân tích dữ liệu



Biểu đồ 1. Quy trình phân tích dữ liệu trong R

# R Nội dung bài học

---

- **Chuyển đổi dữ liệu với R**
  - Các dạng câu lệnh
  - Các tác vụ chính và câu lệnh tương ứng
- **Phân tích mô tả tự động**
  - Định tính x định tính
  - Định tính x định lượng
  - Định lượng x định lượng

## R B4. Chuyển đổi dữ liệu – Một số lệnh cơ bản

### **library(tidyverse)**

#### **Lệnh độc lập**

- mutate()
- filter()
- select()
- arrange()

#### **Lệnh phụ thuộc**

- case\_when()
- ifelse()
- mean()
- min()
- max()



## B4. Chuyển đổi dữ liệu - Tác vụ chính chuyển đổi dữ liệu

- Chuyển đổi biến
  - Tạo biến mới: mutate **[ex1]**
  - Mã hóa biến, phân nhóm: mutate(), case\_when(), ifelse() **[ex2]**
  - Xóa biến: select()
- Lọc dữ liệu:
  - Lọc cột: select() **[ex3]**
  - Lọc hàng: filter() **[ex4]**
- Đổi tên biến: rename(), select()
- Sắp xếp dữ liệu: arrange()
- Kết hợp biến số: merge(), rbind(), cbind().

## B4. Chuyển đổi dữ liệu - Tác vụ chính chuyển đổi dữ liệu

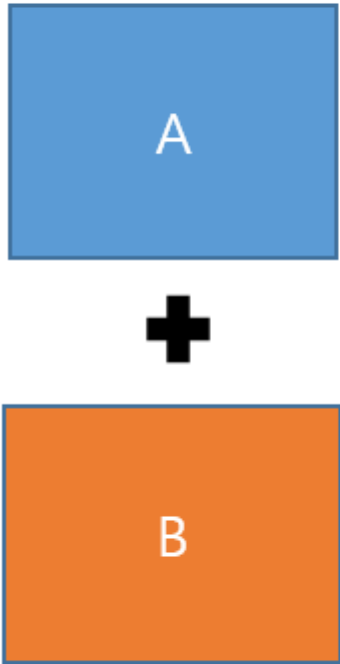
### Lọc hàng (filter)

Kí hiệu	Giải thích	Ví dụ
>	Lớn hơn	<code>filter(x&gt;3)</code>
<	Bé hơn	<code>filter(x&lt;3)</code>
<=	Bé hơn bằng	<code>filter(x&lt;=3)</code>
>=	Lớn hơn bằng	<code>filter(x&gt;=3)</code>
==	Bằng	<code>filter(x==3)</code>
!=	Không bằng	<code>filter(x!=3)</code>
%in%	Bao gồm	<code>filter(x %in% c(1,3,5))</code>

# R B4. Chuyển đổi dữ liệu - Tác vụ chính chuyển đổi dữ liệu

## Kết hợp dữ liệu

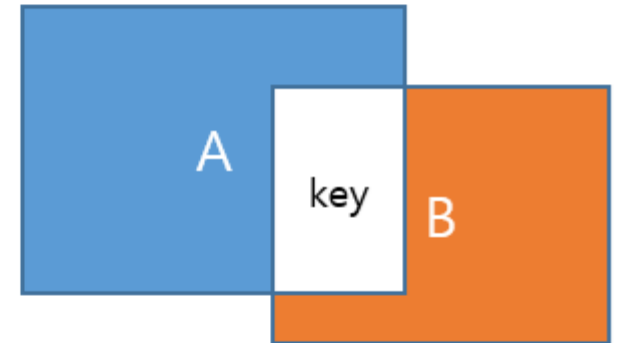
`rbind(A,B)`



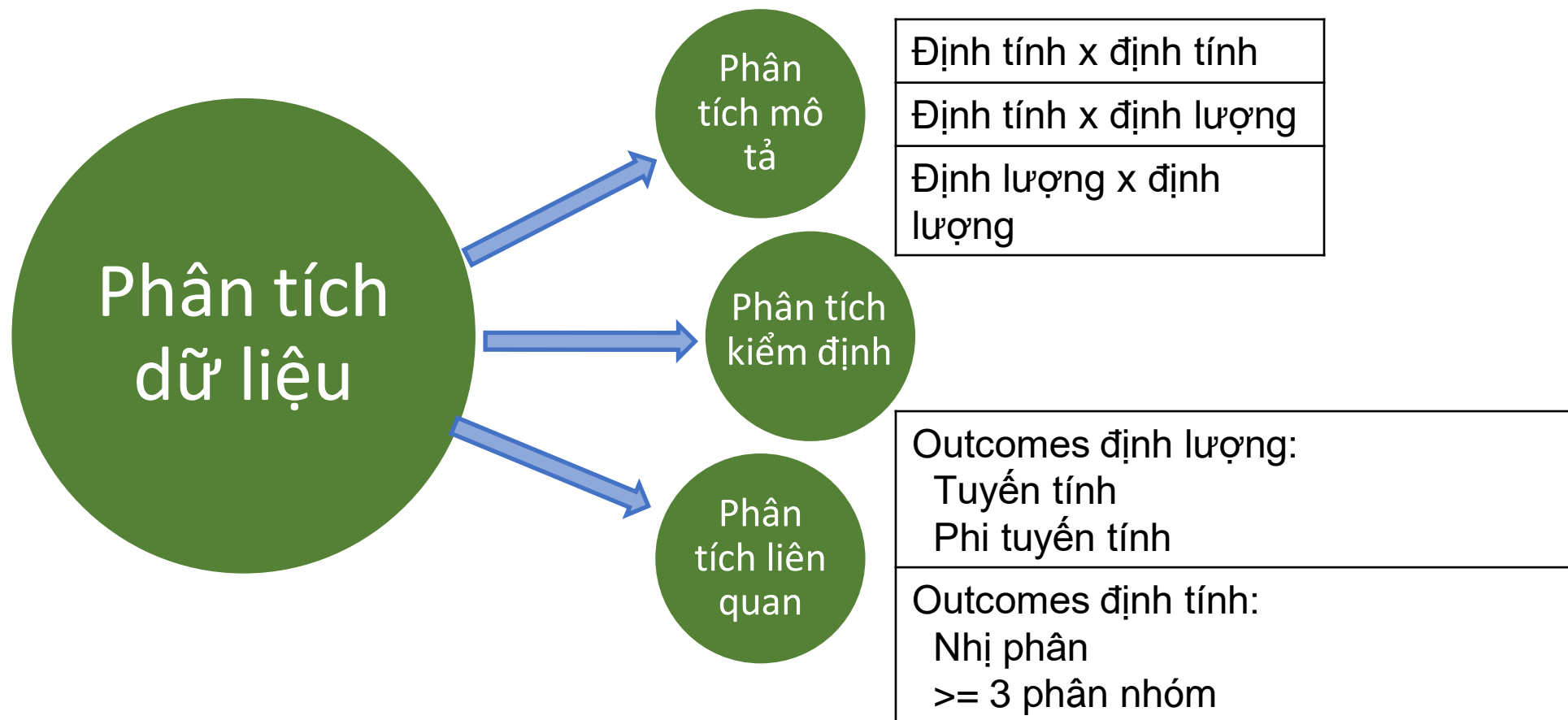
`cbind(A,B)`



`merge(A,B,by='key')`



# R B5. Phân tích dữ liệu





## R B5. Phân tích dữ liệu

### Tạo bảng mô tả

Quy tắc xây dựng bảng mô tả:

- B1. Xác định số lượng biến số và loại biến số dùng để tạo bảng
- B2. Dựa vào sự kết hợp giữa các biến số để xác định bảng biểu và đồ thị.

Nhóm biến số	Loại bảng	Biểu đồ
Định tính x định tính	n, %	Biểu đồ cột chồng (stacked bars) hay biểu đồ cột theo cụm (Clustered bars)
Định tính x định lượng	n, mean, SD, median, (Q1,Q3)	Biểu đồ cột, Biểu đồ đường
Định lượng x định lượng	correlation efficient	Biểu đồ tán xạ, Histogram

# R Lợi ích của tạo bảng mô tả tự động

---

- Tự động cập nhật kết quả với dữ liệu mới
- Tiện lợi cho việc format và copy vào words (Kết quả được xuất và lưu trữ dưới nhiều định dạng như excel, csv)

# R

## B5. Phân tích dữ liệu

### Tạo bảng mô tả

Nhóm biến số	Câu lệnh	Mẫu cấu trúc câu lệnh
Định tính (x1) x định tính (x2)	Package: janitor => tabyl()	<pre>p_load(janitor) tb1 &lt;- df %&gt;%   tabyl(x1,x2) %&gt;%   adorn_percentages() %&gt;%   adorn_pct_formatting(digits = 1) %&gt;%   adorn_ns(position = "front")</pre>
Định tính (x1) x định lượng (y1)	select() group_by() summarize()	<pre>tb2 &lt;- df %&gt;%   group_by(x1) %&gt;%   summarize(n=n(),             mean = mean(y1),             sd = sd(y1),             median = median(y1),             q1 = quantile(y1, probs = 0.25),             q3 = quantile(y1, probs = 0.75))</pre>
Định lượng (y1) x định lượng (y2)	cor()	<pre>cor(y1,y2, method = "pearson") #spearman</pre>

## R B5. Phân tích dữ liệu

### Tạo bảng mô tả

- Lưu dữ liệu/ xuất dữ liệu bảng mô tả

*Ví dụ:*

Tạo bảng mô tả định tính x định lượng

```
Table1 <- df %>% select(x,y) %>%  
  group_by(x) %>%  
  summarize(n=n(), mean = mean(y), sd = sd(y))
```

```
export(table1, here("Results", "table1.xlsx"))
```

**Table1** có thể xuất dưới các định dạng sau

- Excel(.xlsx)
- CSV

- Làm sạch và chuyển đổi dữ liệu trước khi phân tích đóng vai trò vô cùng quan trọng để nâng cao hiệu suất của quá trình phân tích.
- Xây dựng bảng mô tả tự động sẽ giúp tiết kiệm thời gian đối với các tác vụ phân tích lặp lại và bộ số liệu được cập nhật theo thời gian.

# R Thực hành

---

Sử dụng file code3.Rmd trong file code để tiến hành thực tập:  
Lưu ý: cần copy file code3.Rmd vào trong thư mục Rproject của mỗi học viên trước khi thực tập.