

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**THU THẬP VÀ TIỀN XỬ LÝ DỮ LIỆU**  
**LAB2**

**CODE BOOK**  
**DATASET DIABETES PATIENT**  
**RECORDS**

**Họ và tên : Lưu Quang Tiến Hoàng**

**MSSV : 20521342**

**Lớp : DS103.M21**

**Tên bộ dữ liệu:**

*Diabetes patient records.*

**Nguồn thu thập và cách thu thập:**

Từ 2 nguồn là:

Máy đo tự động: Có đồng hồ bấm giờ tự động.

Thu thập bằng tay: Sử dụng bản ghi giấy.

**Kích thước dữ liệu:****File raw data:**

Số thuộc tính: 4

Kích thước: Một folder 681KB gồm tập hợp các file: **data-01, data-02,...**

**File tidy data:**

Số thuộc tính: 4

Tidydata được lưu thành Tidy\_data.csv

Kích thước: 859KB

Số dòng dữ liệu: 29330

**Thông tin các thuộc tính:**

**Date:** Ngày thu thập, định dạng: MM-DD-YYYY

**Time:** Giờ thu thập, định dạng: XX:YY (24 giờ).

**Code:** Mã code theo danh sách sau:

33 = Regular insulin dose

34 = NPH insulin dose

35 = UltraLente insulin dose

48 = Unspecified blood glucose measurement

57 = Unspecified blood glucose measurement

58 = Pre-breakfast blood glucose measurement

59 = Post-breakfast blood glucose measurement

60 = Pre-lunch blood glucose measurement  
61 = Post-lunch blood glucose measurement  
62 = Pre-supper blood glucose measurement  
63 = Post-supper blood glucose measurement  
64 = Pre-snack blood glucose measurement  
65 = Hypoglycemic symptoms  
66 = Typical meal ingestion  
67 = More-than-usual meal ingestion  
68 = Less-than-usual meal ingestion  
69 = Typical exercise activity  
70 = More-than-usual exercise activity  
71 = Less-than-usual exercise activity  
72 = Unspecified special event

**Value:** Giá trị thu thập được.

**Thông tin tác giả:**

*kahn@informatics.WUSTL.EDU (Internet) or 70333,34 (CompuServe)*

**Instruction list:**

```
rm(list=ls())  
myfile <- list.files(path="dataset/", pattern="data-")  
k = TRUE  
for (f in myfile)  
{  
  if (k==TRUE)  
  {  
    file <- read.csv(paste("dataset/", f, sep=""), sep="\t", header = FALSE)  
    k = FALSE  
  }  
}
```

```
else
{
  file <- rbind(file, read.csv(paste("dataset/",f,sep=""),sep="\t",header = FALSE))
}
}
dataset <- file
variables <- c("Date", "Time", "Code", "Value")
colnames(dataset) <- variables
write.csv(dataset, file = "tidy_data.csv",row.names = FALSE)
```

**--- HÉT ---**