Alex Gibbons and Lauren Kawecki

CMPT 363 – Data Mining

Dr. Ankur Agrawal

Due: February 15, 2024

Assignment 4

Code:

library(RMySQL)

height\_data <- read.table("C:\\Users\\Kelly\\Desktop\\MCSP24\\CMPT363 - Data Mining\\tab\_sep\_data.txt",header=TRUE,sep="\t")

weight\_data <- read.table("C:\\Users\\Kelly\\Desktop\\MCSP24\\CMPT363 - Data Mining\\comma\_sep\_data.txt",header=TRUE,sep=",")

mean\_height <- mean(height\_data$height, na.rm=TRUE)

for(i in 1:nrow(height\_data)) {

if(is.na(height\_data[i,2])) { height\_data[i,2] <- mean\_height } }

mean\_weight <- mean(weight\_data$weight, na.rm=TRUE)

for(i in 1:nrow(weight\_data)) {

if(is.na(weight\_data[i,2])) { weight\_data[i,2] <- mean\_weight } }

student\_measurements <- merge(x=weight\_data,y=height\_data, by="id", all=TRUE)

mydb <- dbConnect(MySQL(),user="root",password="root",host="localhost")

dbSendQuery(mydb, "DROP DATABASE IF EXISTS student\_data;")

dbSendQuery(mydb, "CREATE DATABASE student\_data;")

dbSendQuery(mydb, "USE student\_data;")

dbSendQuery(mydb, "CREATE TABLE measurements(id INT, weight INT, height INT);")

query <- "INSERT INTO measurements VALUES"

query <- paste0(query, paste(sprintf("('%f','%f','%f')",student\_measurements$id,student\_measurements$weight,student\_measurements$height),collapse=","))

dbSendQuery(mydb, query)

query\_results <- fetch(dbSendQuery(mydb, "SELECT id,weight,height FROM measurements;"))

print(query\_results)

Results:

A blue screen with orange text

Description automatically generated