##################################################################

# ESTRELA 2024年10月号「研究ノート」記事スクリプト

###########################

# 社会生活基本調査のSSDSE-Dシリーズのcsvファイルを読み込む。

# 　実行文１行目の“SSDSE-D-2023.csv“を”SSDSE-D-2021.csv”に変えれば、

# 別途ダウンロードしたSSDSE-D-2021.csvを分析することもできます。

ssdse\_data\_path <- “SSDSE-D-2023.csv“

ssdse <- read.csv(ssdse\_data\_path, fileEncoding="Shift-JIS")

ssdse <- ssdse[-1, ] # 項目名の行を削除

ssdse <- type.convert(ssdse, as.is=TRUE) # 数値項目の型を再定義

library(dplyr)

##########################################################

# データの前処理： 男女種別を示す列が、”1\_男”である行だけを抜き出し、

# Prefecture列が”全国”である行を削除。その後、Prefecture列を

# 表側に設定してから、行動種類（20種類）の列(MG01からMG20)だけを残す。

colnames(ssdse)[1] <- "Male\_Female"

ssdse\_small <- ssdse |> filter(Male\_Female == "1\_男")|> filter(Prefecture != "全国")

rownames(ssdse\_small) <- ssdse\_small$Prefecture

ssdse\_small <- ssdse\_small |> select(num\_range("MG", 1:20, width=2))

# データの前処理：生活時間（１日１４４０分）を比率化し、生活時間項目毎に標準化

ssdse\_small.p <- (ssdse\_small / 1440 )\* 100

ssdse\_small.std <- ssdse\_small.p |> scale()

###############################

# ward法階層的クラスター分析を実施する

ssdse\_small.dist <- sqrt(1/2)\*dist(ssdse\_small.std)

methname <- "ward.D2"

ssdse\_small.hclust <- hclust(ssdse\_small.dist, method=methname)

####################

# デンドログラムを描画する

cluster\_num.p <- 3

plot(ssdse\_small.hclust)