

# 8. Úkol na předmět Mnohorozměrná analýza

Jaromír Macoun

Duben 2022

## 1 Úvod

Zabýváme se v tomto úkolu daty *athletic*, které obsahují záznamy o rekordních časech v atletice. Přesněji řečeno data obsahují národní rekordy 55 zemí v 8 běžeckých disciplínách. Disciplíny to jsou následující: běh na 100, 200, 400, 800, 1500, 5000, 10000 metrů a maratonský běh.

## 2 Clusterová analýza

### 2.1

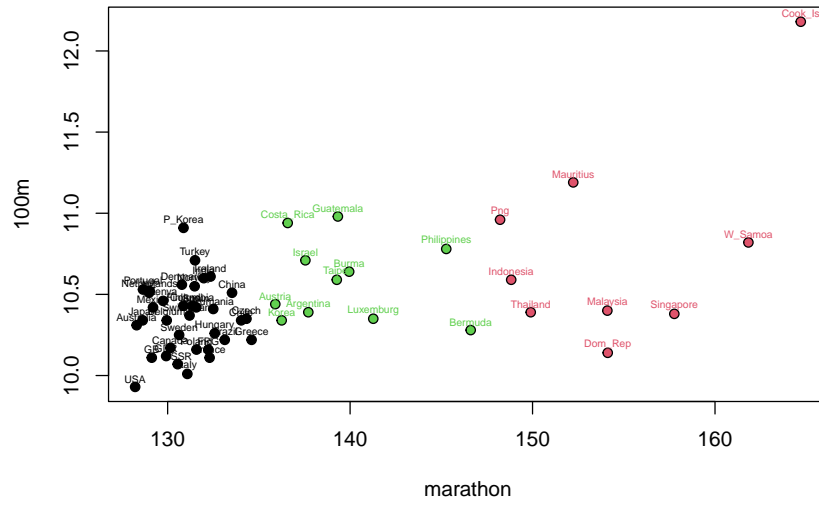
Po provedení K means algoritmu nastiňují získané výsledky následující obrázky č. 1, 2 a 3. Jednotlivé clustery jsou rozlišené černou zelenou a červenou barvou. Jak je vidět z obrázku č. 1, tak jednotlivé skupiny nejsou rozlišitelné vzhledem běhu na sto metrů, naopak co se týče času na maratonu, tak zde zřetelné tři skupiny jsou vidět. Další obrázky pouze potvrzují, že clustery víceméně se nejvíce liší v časech na dlouhých tratích. Jsou seřazeny od nejlepších časů (černá barva) na maratonu, 10km. Zelená je prostřední skupina a červená je skupina států, které mají nejhůřší časy na dlouhých tratích.

### 2.2

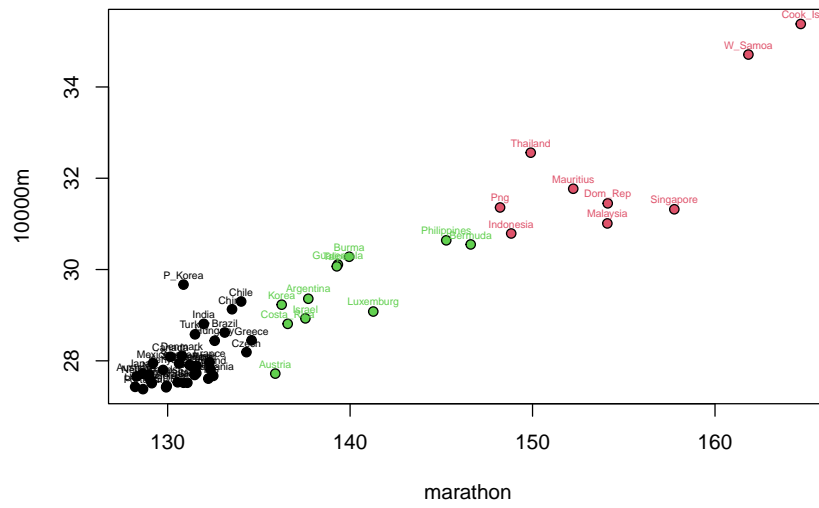
Po použití hierarchické metody dostáváme výsledek, který je shrnut v obrázku č. 4. Data nám to rozdělilo na jednu početnou skupinu a dvě relativně malé. V těchto malých skupinách se víceméně celá červená skupina z předešlé analýzy. Kromě toho obsahuje asi dvě země ze zelené skupiny. Tyto dvě země alespoň z předešlých obrázků se zdají velmi blízké červené skupině.

Tu menší skupinu z těchto malých clusterů tvoří dvojice W Samoa a Cook islands. Opět můžeme vidět z obrázků z předešlé analýzy, že se v některých obrázcích od červené skupiny velmi vzdalují.

Dále je zajímavé, že tyto dvě malé skupiny jsou tvořeny víceméně krom Dominikánské republiky ostrovními, poloostrovními státy z asiie a oceánie.



Obrázek 1



Obrázek 2

