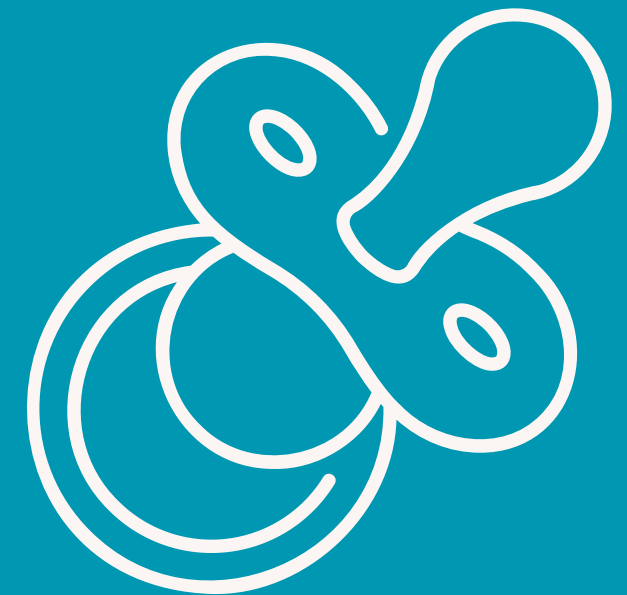


# ANALÝZA SOCIOEKONOMICKÝCH A ZDRAVOTNÍCH FAKTORŮ

Jejich vliv na úmrtnost novorozenců

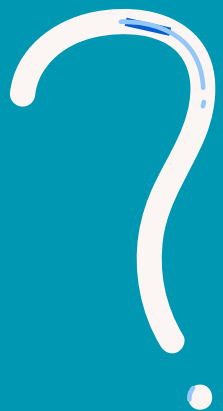


# ZÁKLADNÍ INFORMACE

Vzaly jsme data pro všechny země světa – po procesu čištění dat zbylo  
175 zemí

Úmrtnost novorozenců je počet úmrtí dětí mezi 0–11 měsíci na 1000  
narozených dětí

Veškerá data jsou z roku 2021



# VÝZKUMNÉ OTÁZKY

**Základní otázka:** Jaký socioekonomický či zdravotní faktor významně ovlivňuje úmrtnost novorozenců v zemích světa?

## **Další výzkumné otázky:**

Jak jsou rozložení hodnot jednotlivých faktorů?

Jaké jsou trendy nebo vztahy mezi faktory a úmrtností novorozenců?

Jak se faktory mění v závislosti na geografické poloze zemí?

Jak se geograficky liší vysoká, střední a nízká úmrtnost novorozenců a jaké socioekonomické faktory mohou být spojeny s těmito rozdíly?

# VYBRANÉ FAKTORY

Geographic area (jednotlivé země)

HDP

Life Expectancy Male/Female (očekávaná délka života)

Total fertility rate (porodnost)

Net Migration rate (migrace)

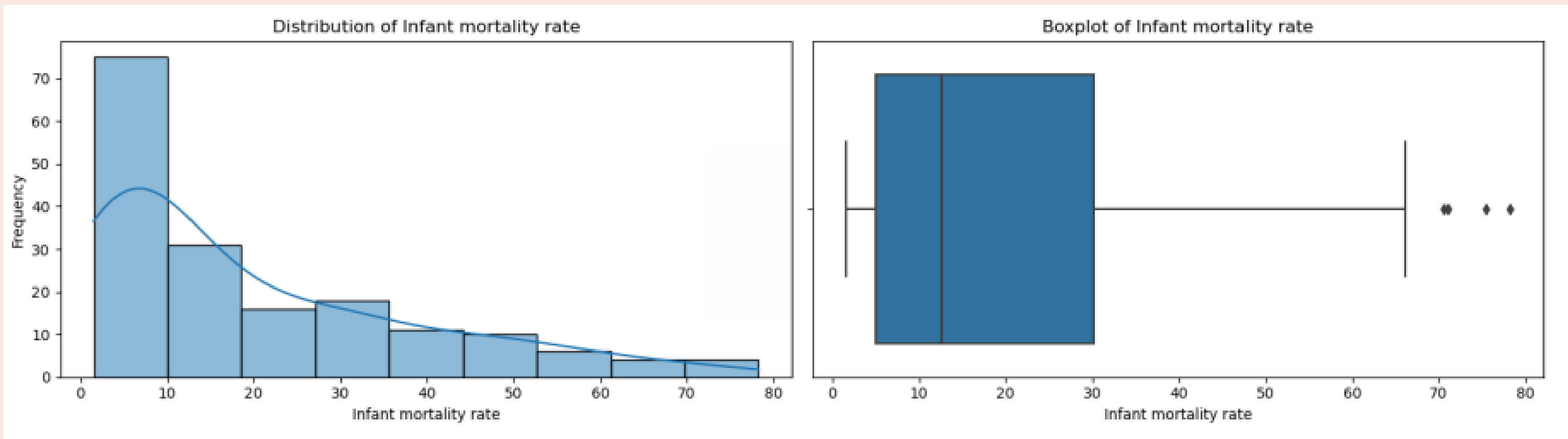
Share of Urban Population (urbanizace)

Access to electricity (přístup k elektřině)

**Infant Mortality (úmrtnost novorozenců)**

# JAK JSOU ROZLOŽENÍ HODNOT JEDNOTLIVÝCH FAKTORŮ?

Distribuce a boxplot pro naši hlavní proměnnou: **Úmrtnost novorozenců**

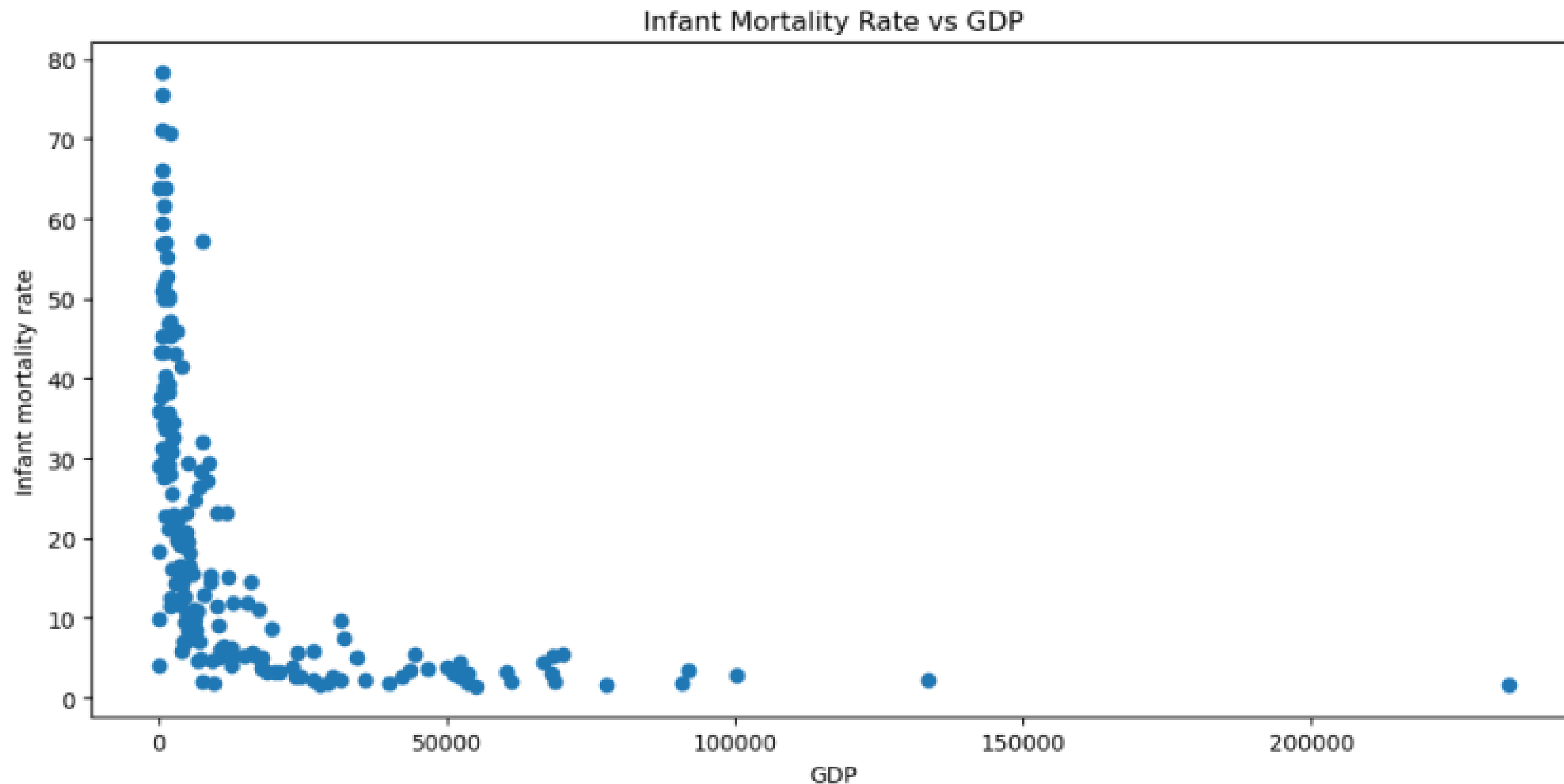


Většina zemí má spíše nižší úmrtnost novorozenců, existují země s výrazně vyšší hodnotou

Jsou zde outlieři s výrazně vyšší úmrtností novorozenců

# JAKÉ JSOU TRENDY NEBO VZTAHY MEZI FAKTORY A ÚMRTNOSTÍ NOVOROZENCŮ?

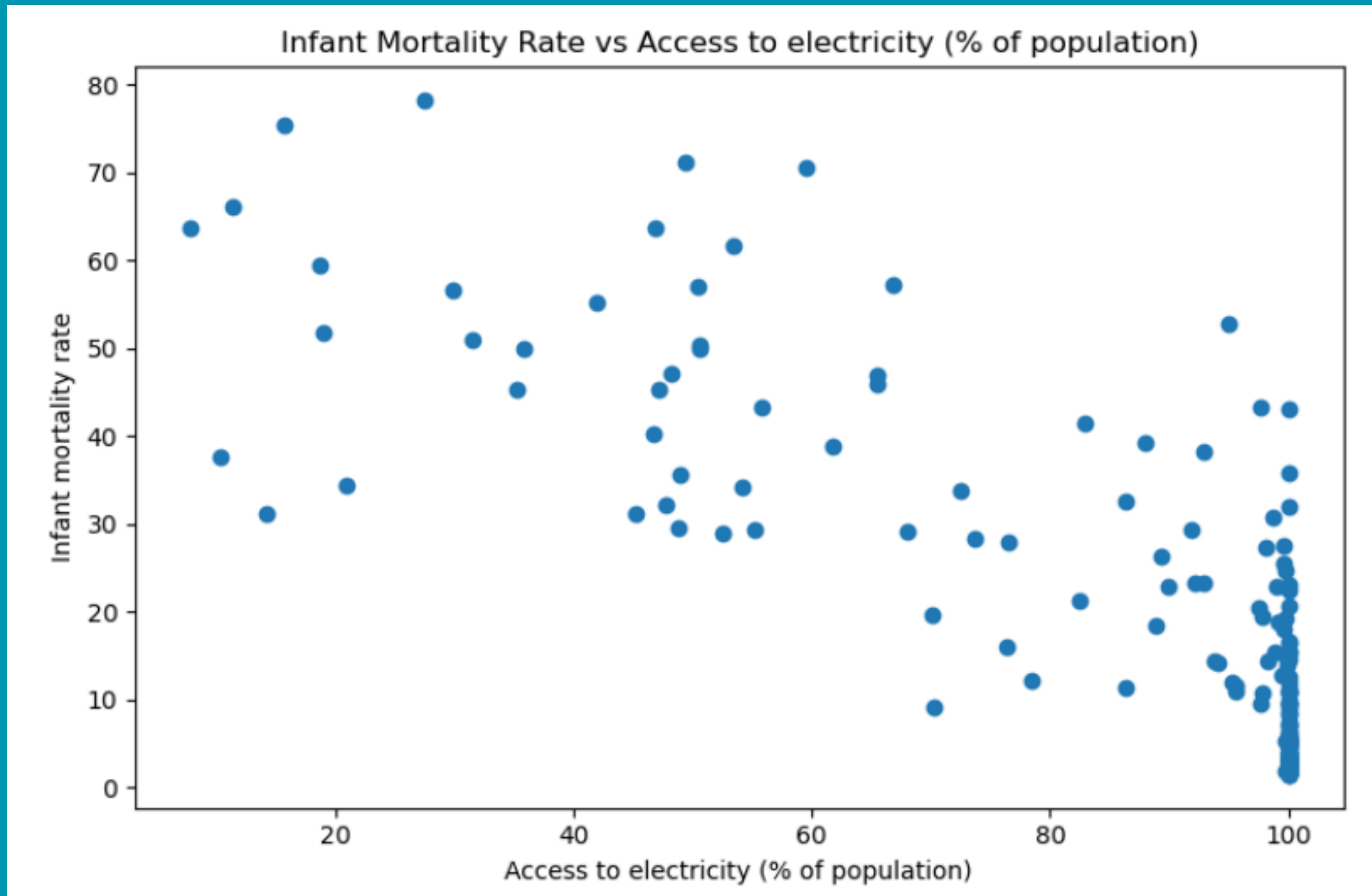
Scatter ploty pro nejvýraznější vztahy mezi faktory a úmrtností novorozenců



↙  
Vyšší HDP →  
nižší úmrtnost  
novorozenců

# JAKÉ JSOU TRENDY NEBO VZTAHY MEZI FAKTORY A ÚMRTNOSTÍ NOVOROZENCŮ?

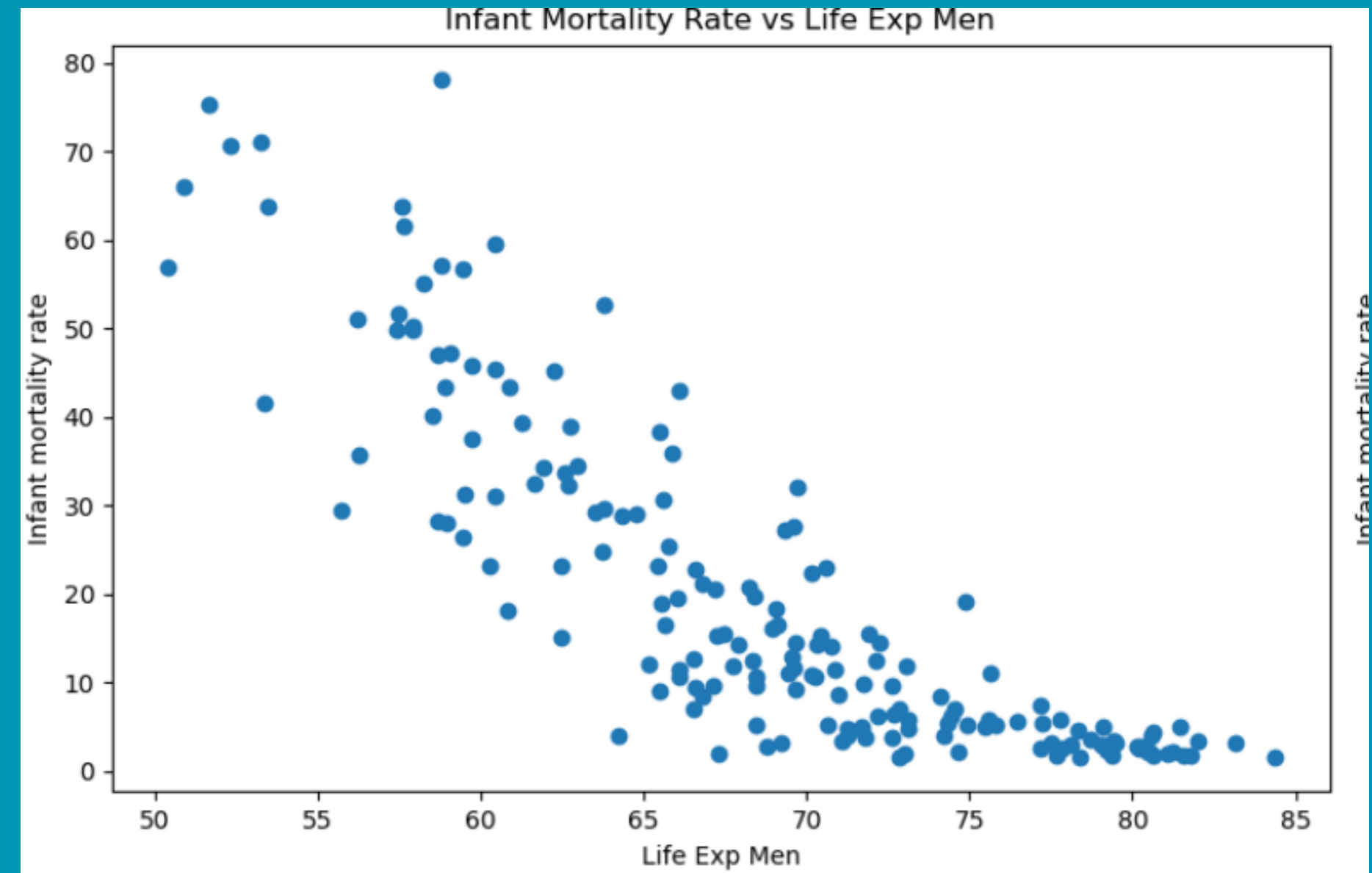
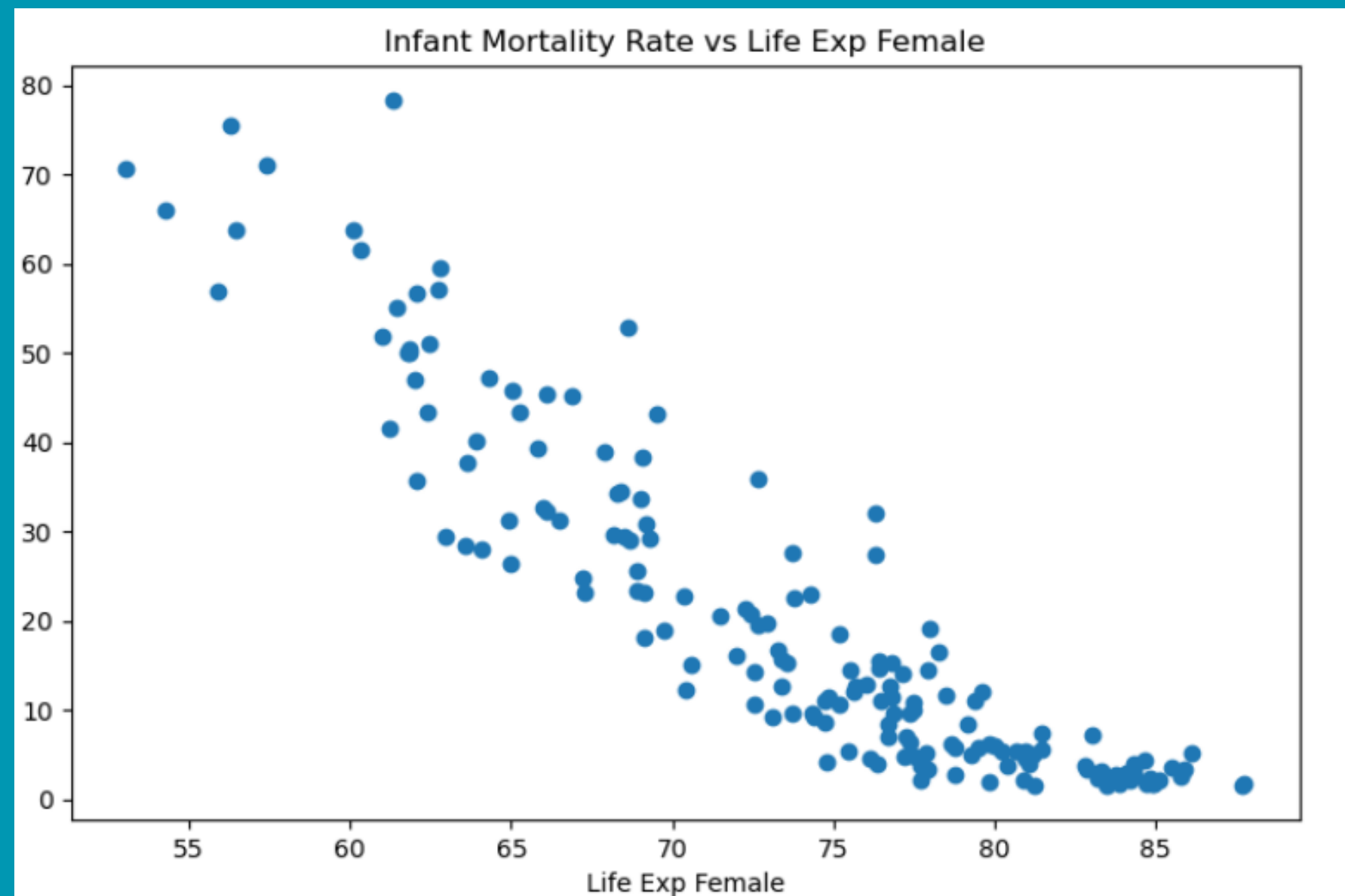
Scatter ploty pro nejvýraznější vztahy mezi faktory a úmrtností novorozenců



Vyšší přístup k  
elektríně → nižší  
úmrtnost  
novorozenců

# JAKÉ JSOU TRENDY NEBO VZTAHY MEZI FAKTORY A ÚMRTNOSTÍ NOVOROZENCŮ?

Scatter plots pro nejvýraznější vztahy mezi faktory a úmrtností novorozenců

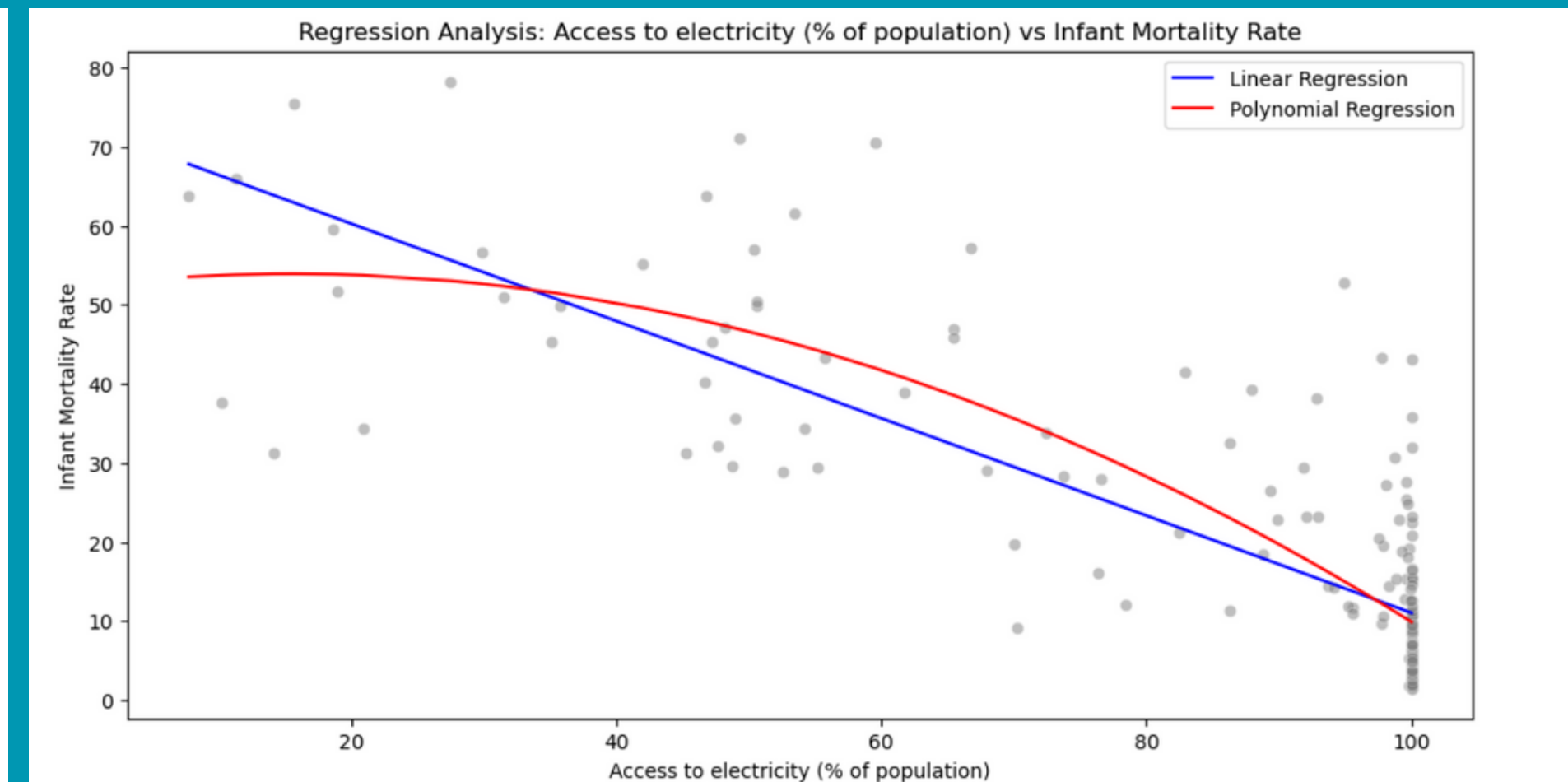
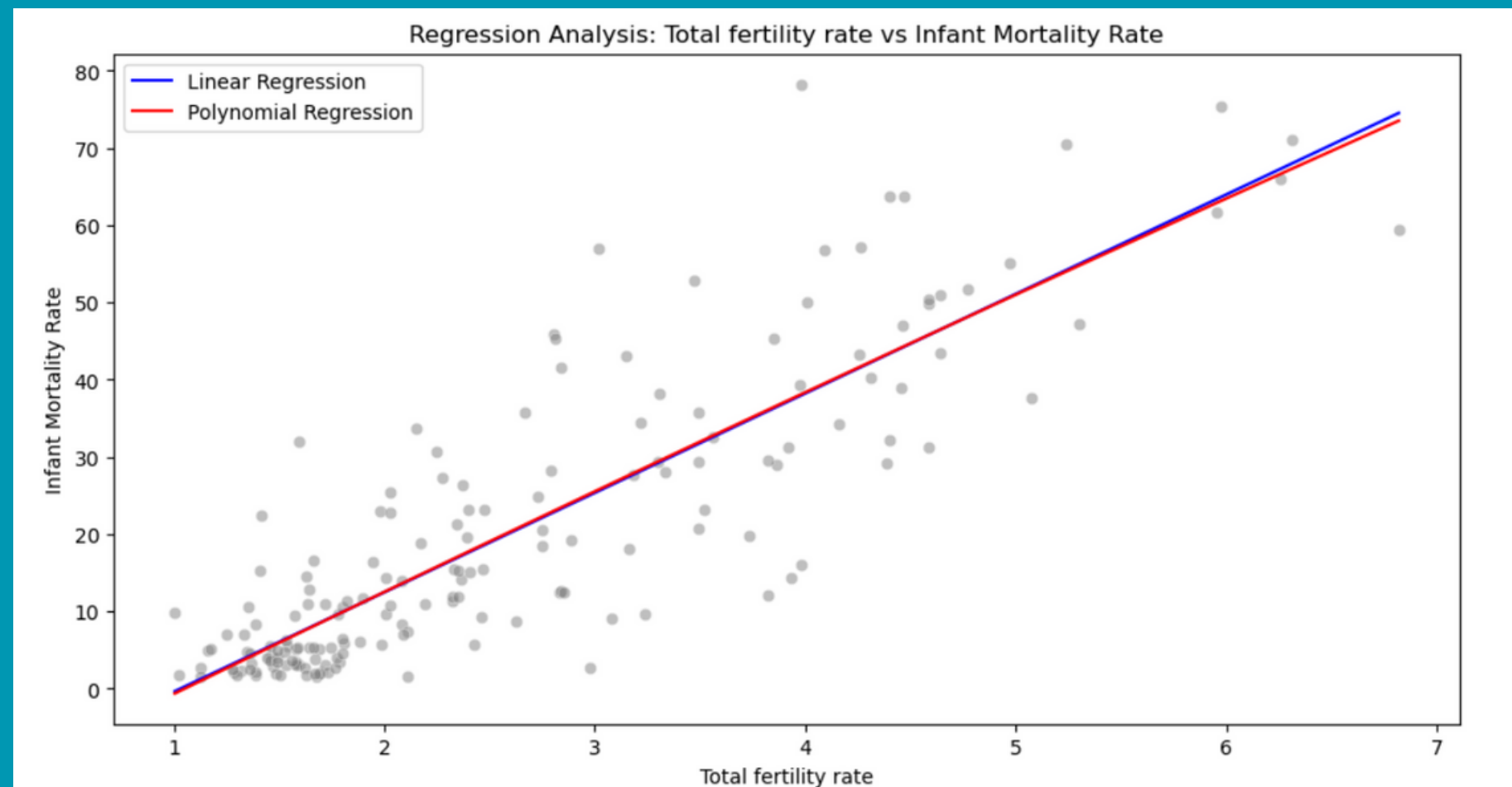


Vyšší očekávaný věk dožití → nižší úmrtnost novorozenců



# JAKÉ JSOU TRENDY NEBO VZTAHY MEZI FAKTORY A ÚMRTNOSTÍ NOVOROZENCŮ?

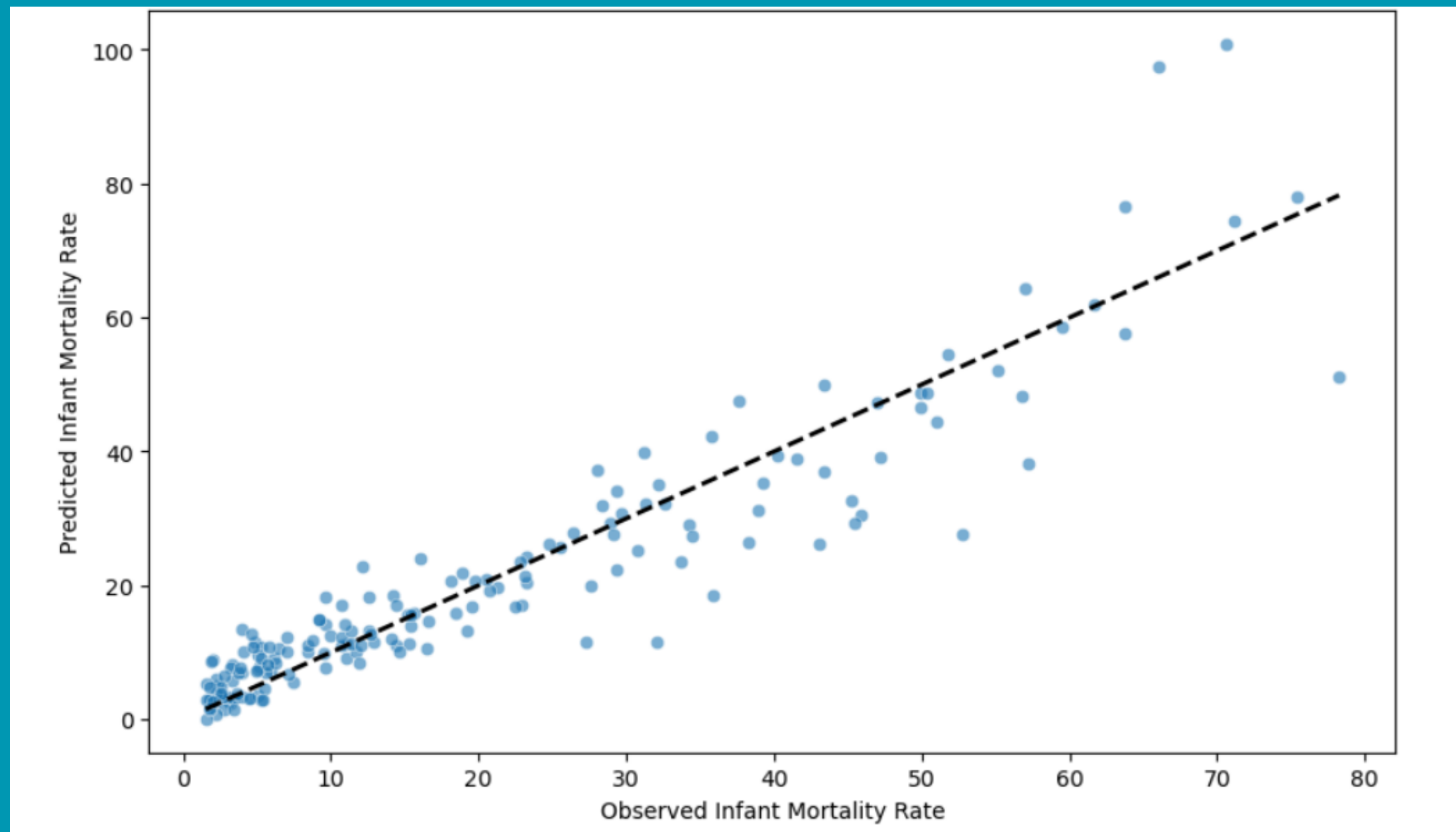
## Lineární a polynominální regrese



Lepší výsledky jsme dostaly pro polynominální reg. u všech faktorů  
→ naznačuje nelineární vztah mezi proměnnými

# JAKÉ JSOU TRENDY NEBO VZTAHY MEZI FAKTORY A ÚMRTNOSTÍ NOVOROZENCŮ?

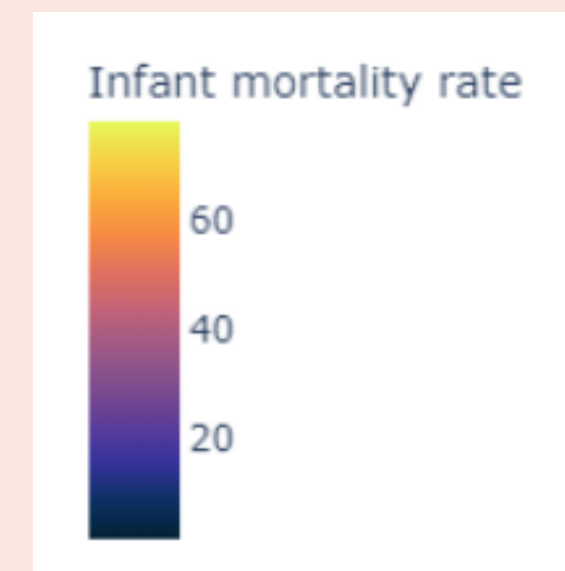
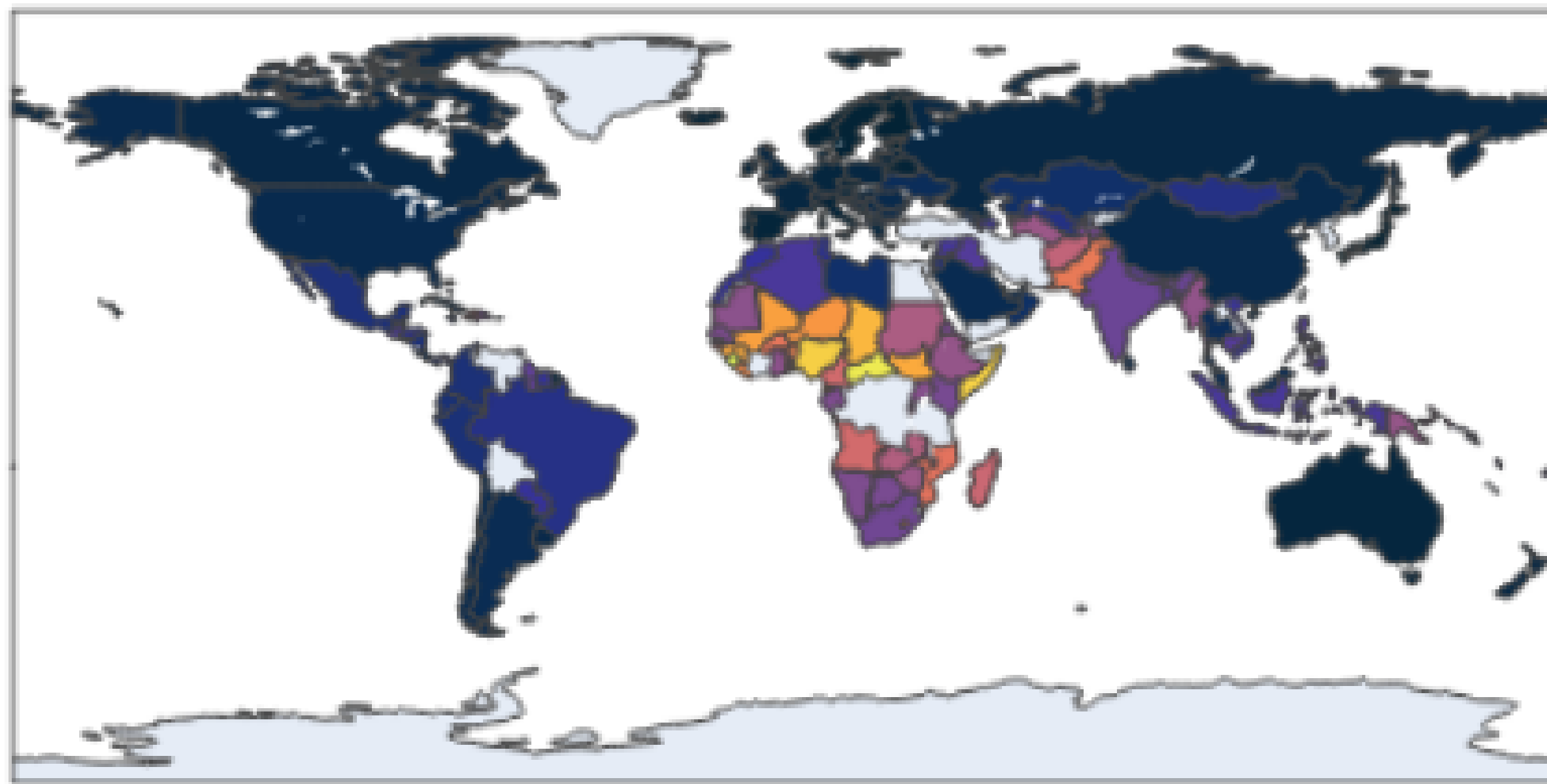
model Poisson. GLM



Nejlepší model zahrnoval  
proměnné HDP, očekávané  
dožití a porodnost  
→ nejdůležitější faktor

# JAK SE FAKTORY MĚNÍ V ZÁVISLOSTI NA GEOGRAFICKÉ POLOZE ZEMÍ?

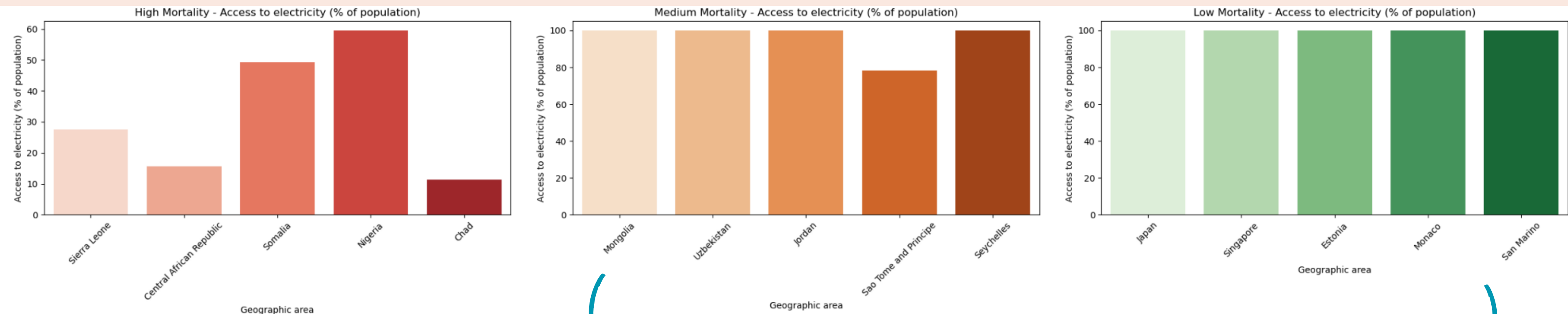
Geografické zobrazení pro naši hlavní proměnnou: **Úmrtnost novorozenců**



→ Výšší úmrtnost  
novorozenců  
především v  
Afrických  
zemích

# JAK SE FAKTORY MĚNÍ V ZÁVISLOSTI NA GEOGRAFICKÉ POLOZE ZEMÍ?

Sloupcové grafy pro nízkou, středí a vysokou úmrtnost novorozenců v zemích pro některé faktory  
**Faktor: přístup k elektřině**

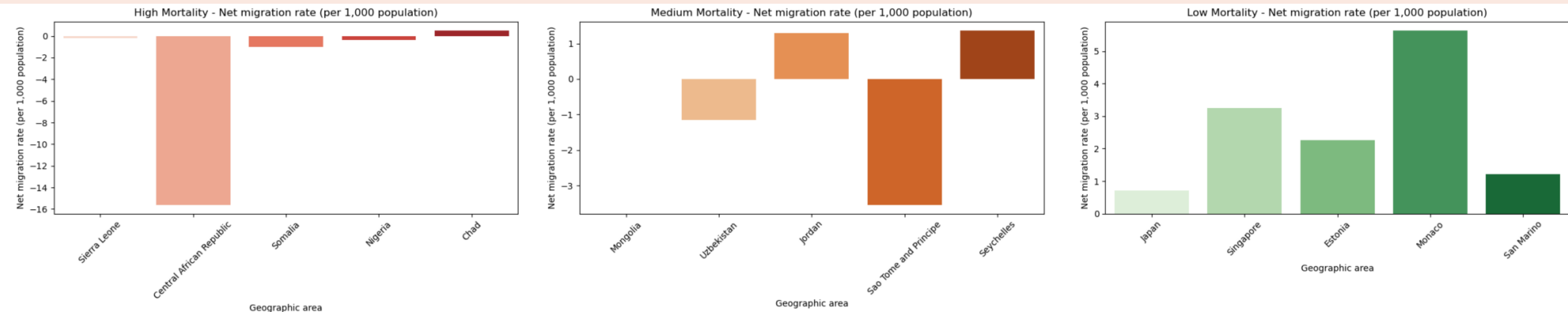


Téměř všude 100 % přístup k elektřině →  
ale v zelených zemích výrazně nižší  
úmrtnost → v oranžových zemích mají na  
úmrtnost vliv spíše jiné faktory

# JAK SE FAKTORY MĚNÍ V ZÁVISLOSTI NA GEOGRAFICKÉ POLOZE ZEMÍ?

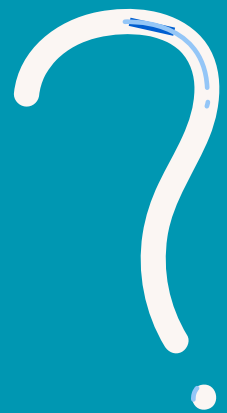
Sloupcové grafy pro nízkou, středí a vysokou úmrtnost novorozenců v zemích pro některé faktory

## Faktor: Míra migrace



V zemích odkud se lidé stěhují pryč  
(záporná migrace) je vyšší úmrtnost  
novorozenců → naznačuje obecně  
špatnou kvalitu života



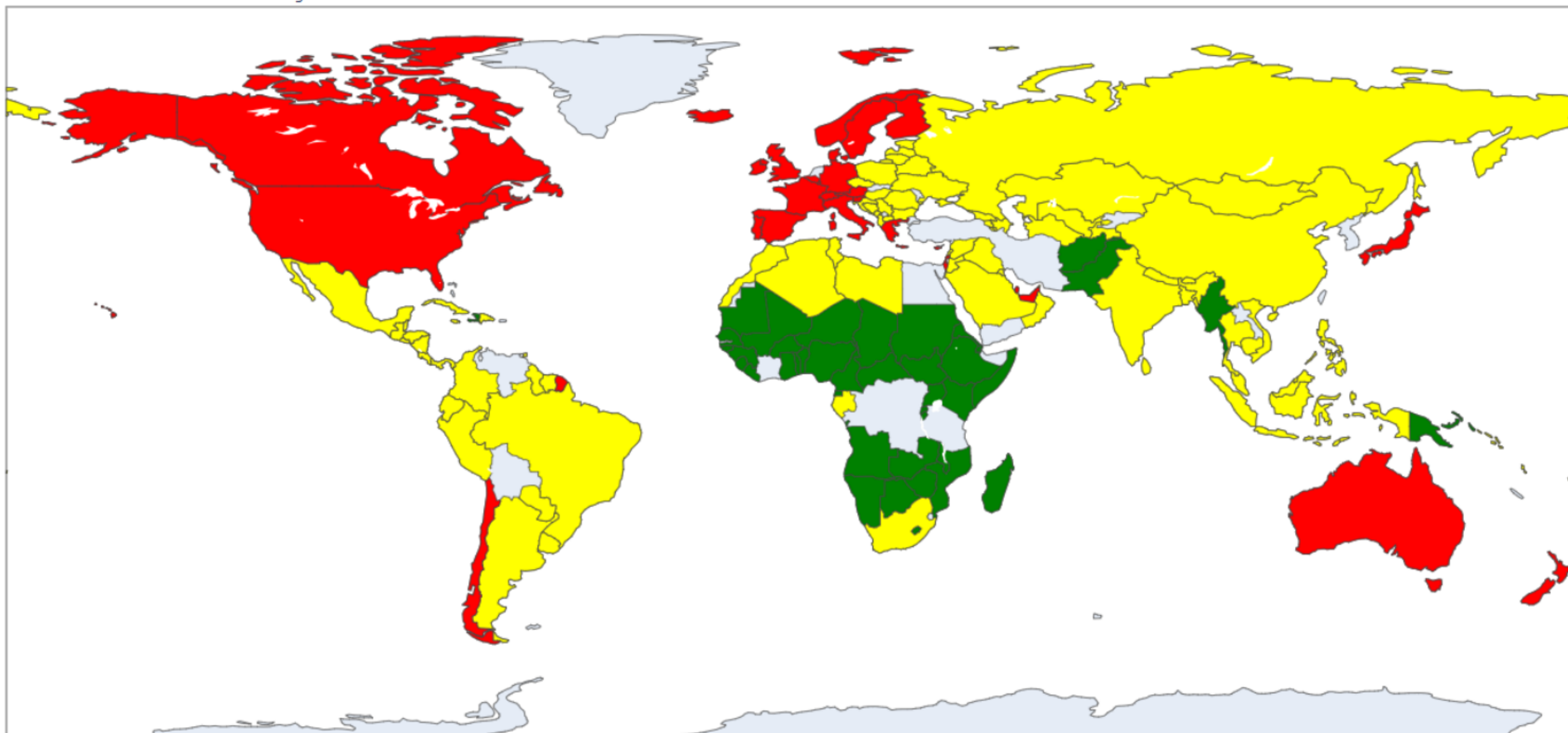


# JAK SE GEOGRAFICKY LIŠÍ VYSOKÁ, STŘEDNÍ A NÍZKÁ ÚMRTNOST NOVOROZENCŮ A JAKÉ SOCIOEKONOMICKÉ FAKTORY MOHOU BÝT SPOJENY S TĚMITO ROZDÍLY?

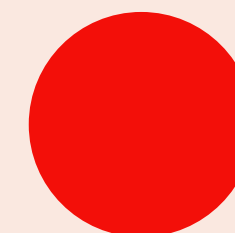
Kategorizace zemí na základě jejich socioekonomických faktorů → CLUSTERING

Metoda: K-means → 3 clustery

K-means Clustering



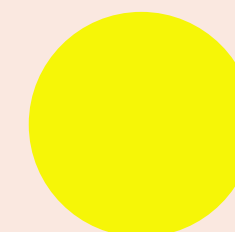
vysoké HDP, přístup k



elektríně

nízká úmrtnost  
novorozenců

střední HDP, ne vždy



přístup k elektríně  
střední úmrtnost  
novorozenců

nízké HDP, ne přístup k



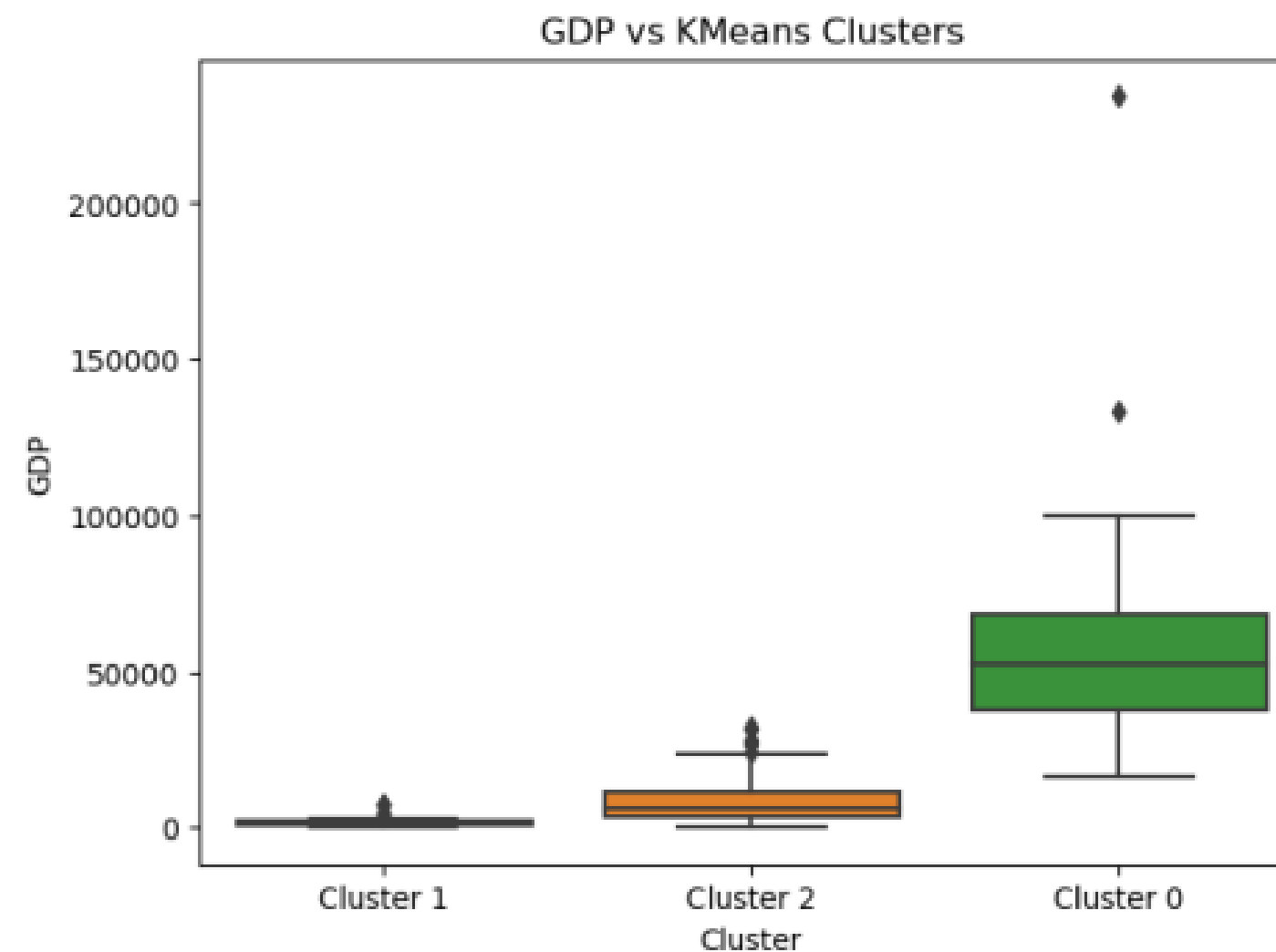
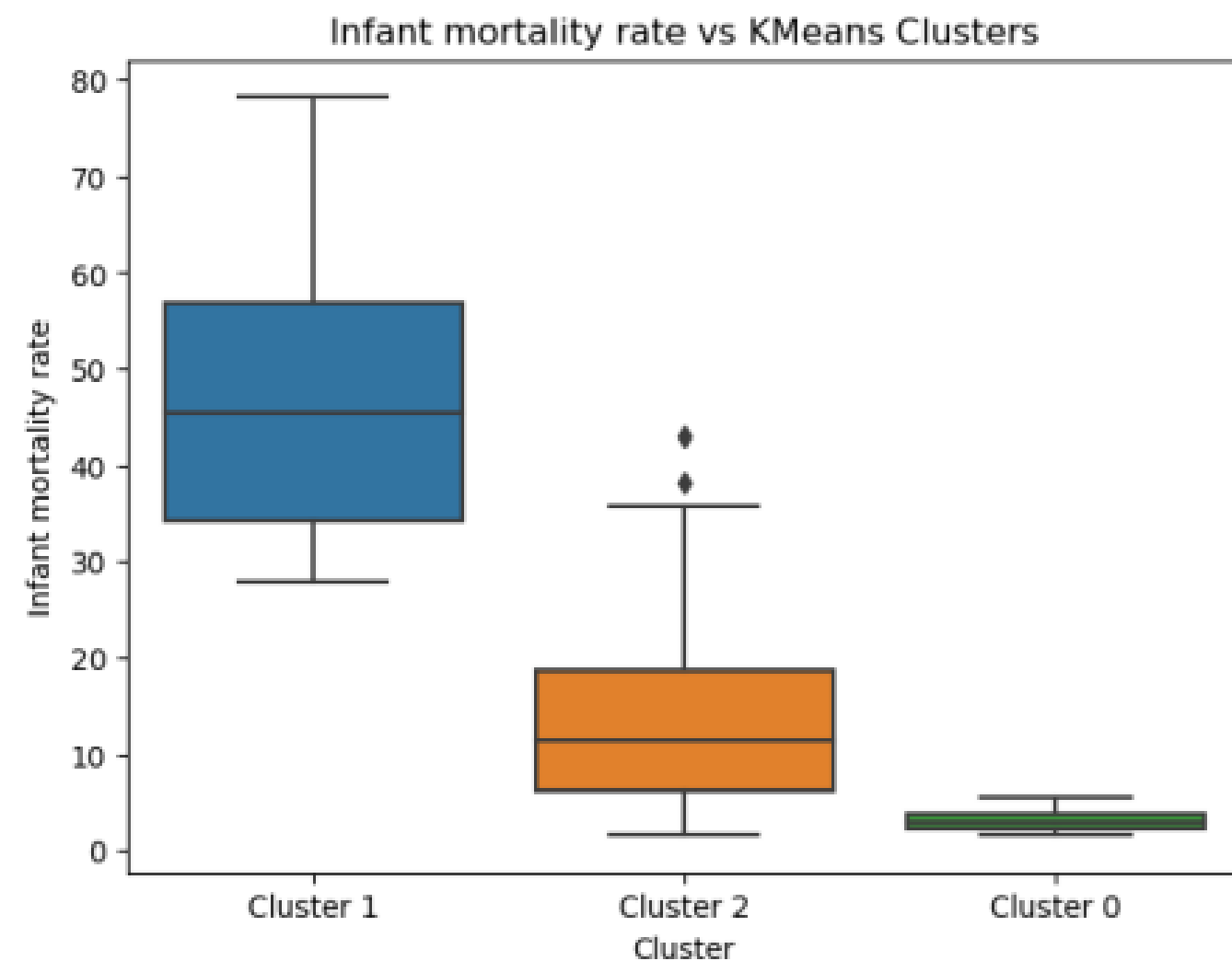
elektríně

vysoká úmrtnost  
novorozenců

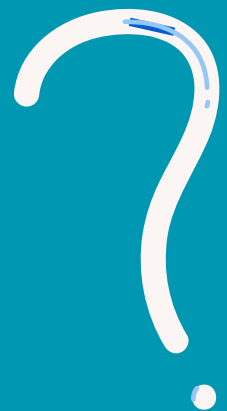
# JAK SE GEOGRAFICKY LIŠÍ VYSOKÁ, STŘEDNÍ A NÍZKÁ ÚMRTNOST NOVOROZENCŮ A JAKÉ SOCIOEKONOMICKÉ FAKTORY MOHOU BÝT SPOJENY S TĚMITO ROZDÍLY?

Kategorizace zemí na základě jejich socioekonomických faktorů → CLUSTERING

Metoda: K-means → Vizuální zobrazení



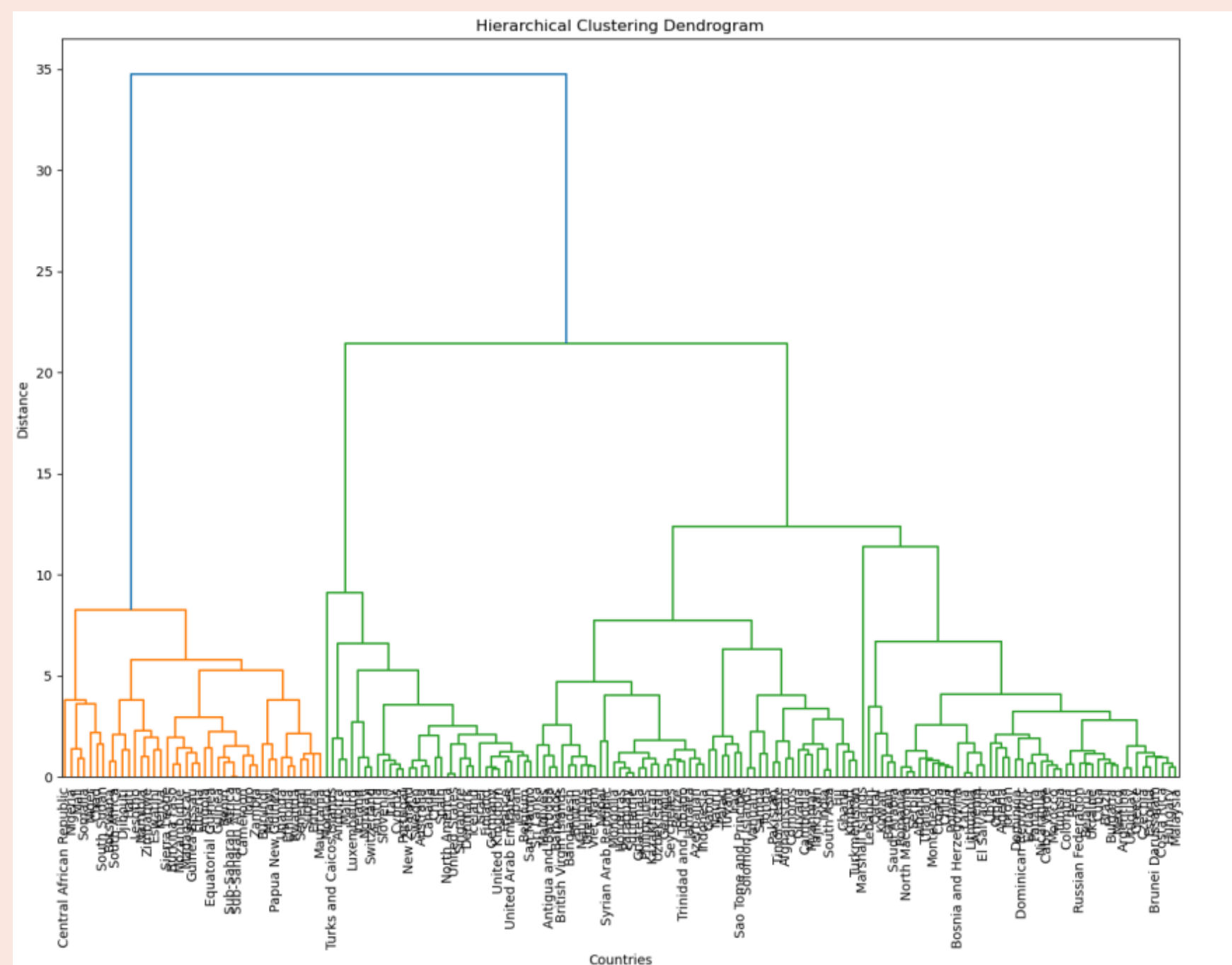
V clusteru s vysokým  
HDP → nejnižší  
úmrtnost novorozenců



# JAK SE GEOGRAFICKY LIŠÍ VYSOKÁ, STŘEDNÍ A NÍZKÁ ÚMRTNOST NOVOROZENCŮ A JAKÉ SOCIOEKONOMICKÉ FAKTORY MOHOU BÝT SPOJENY S TĚMITO ROZDÍLY?

Kategorizace zemí na základě jejich socioekonomických faktorů → CLUSTERING

## Metoda: Hierarchické – Dendogram



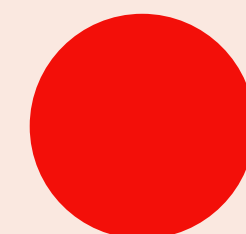
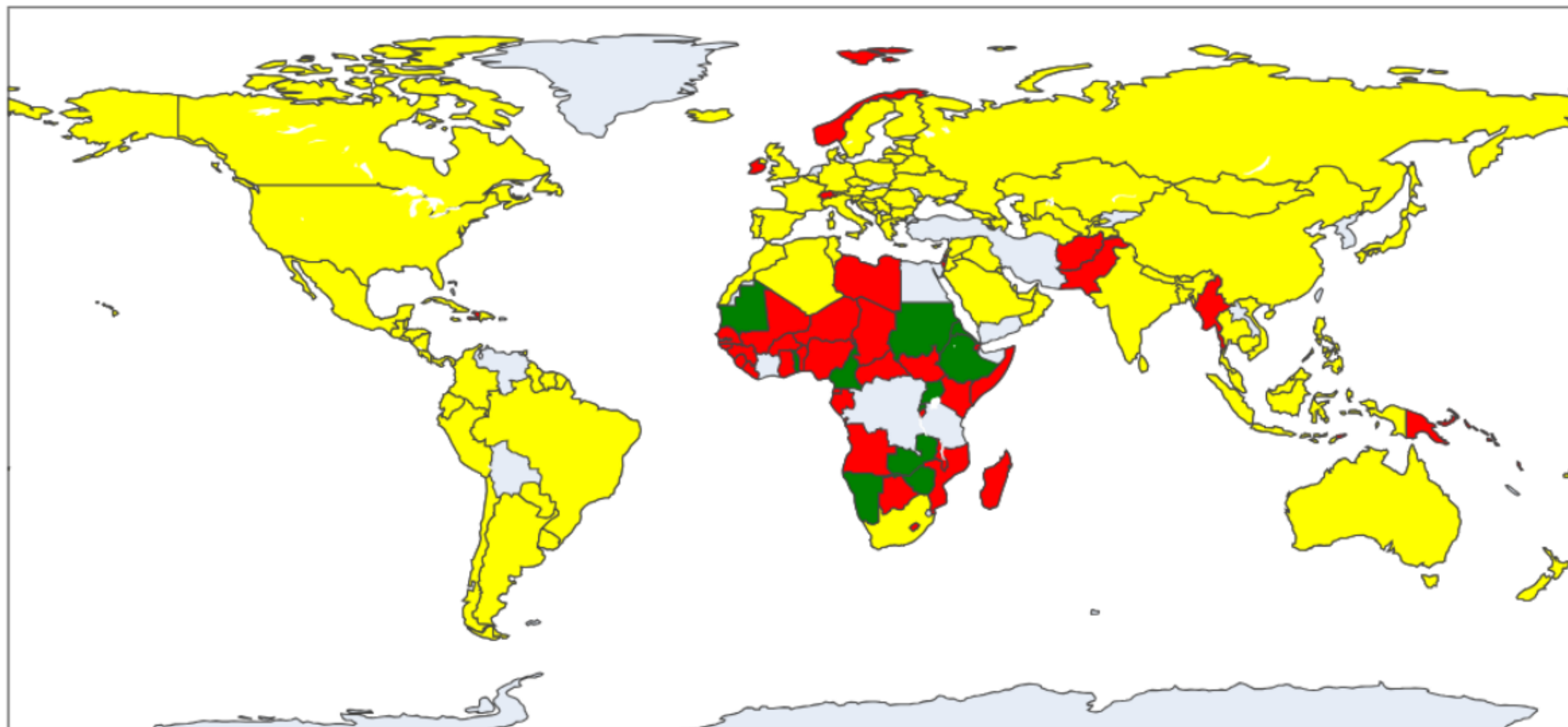




# JAK SE GEOGRAFICKY LIŠÍ VYSOKÁ, STŘEDNÍ A NÍZKÁ ÚMRTNOST NOVOROZENCŮ A JAKÉ SOCIOEKONOMICKÉ FAKTORY MOHOU BÝT SPOJENY S TĚMITO ROZDÍLY?

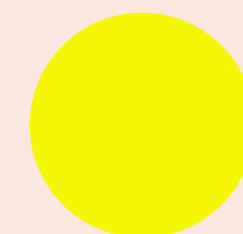
Kategorizace zemí na základě jejich socioekonomických faktorů → CLUSTERING

**Metoda: DBSCAN**

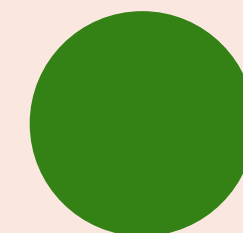


**Outliers/noise**

země s atypickými  
socioekonomickými a  
zdravotními  
charakteristikami ve  
srovnání s ostatními



**Cluster 2**

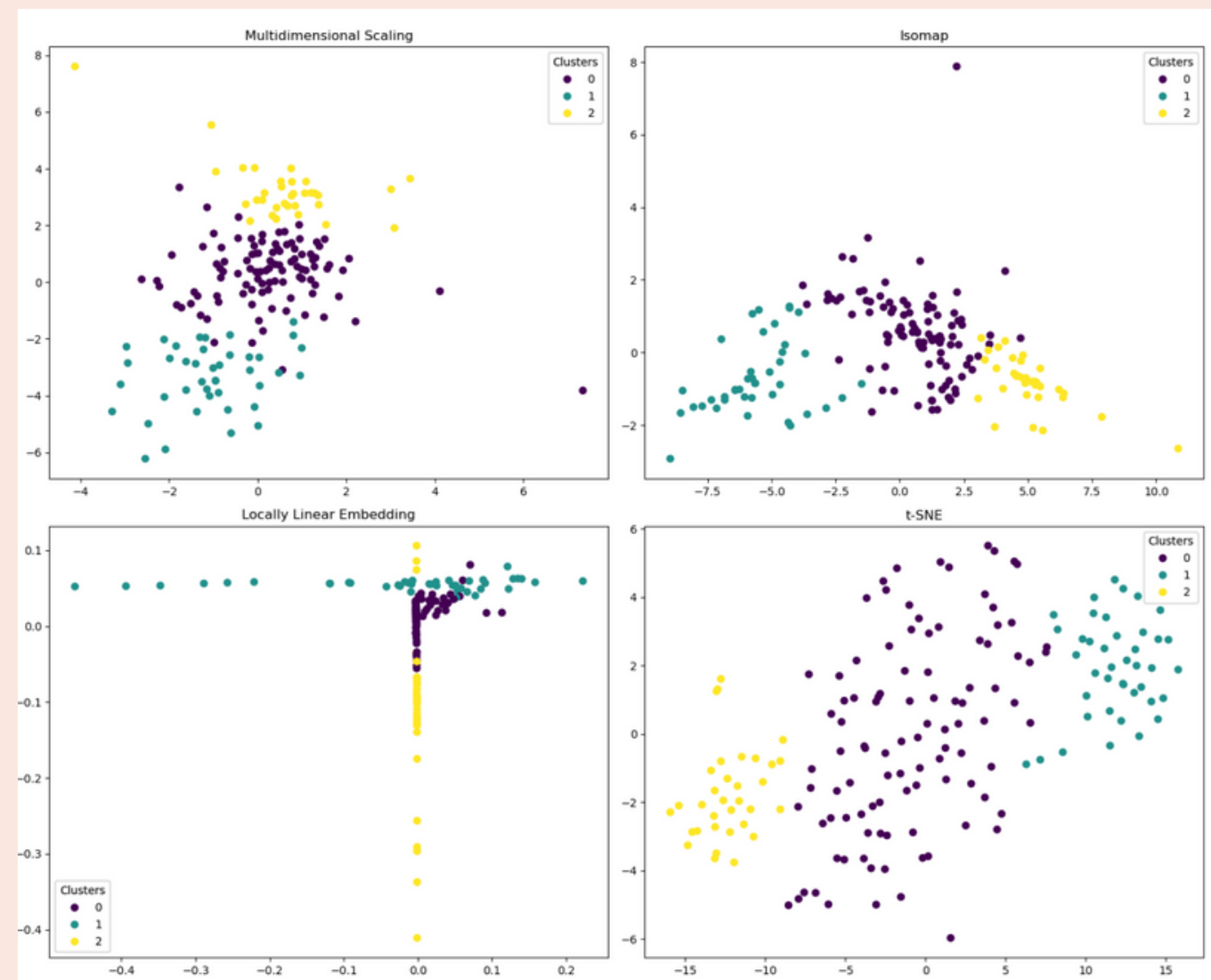
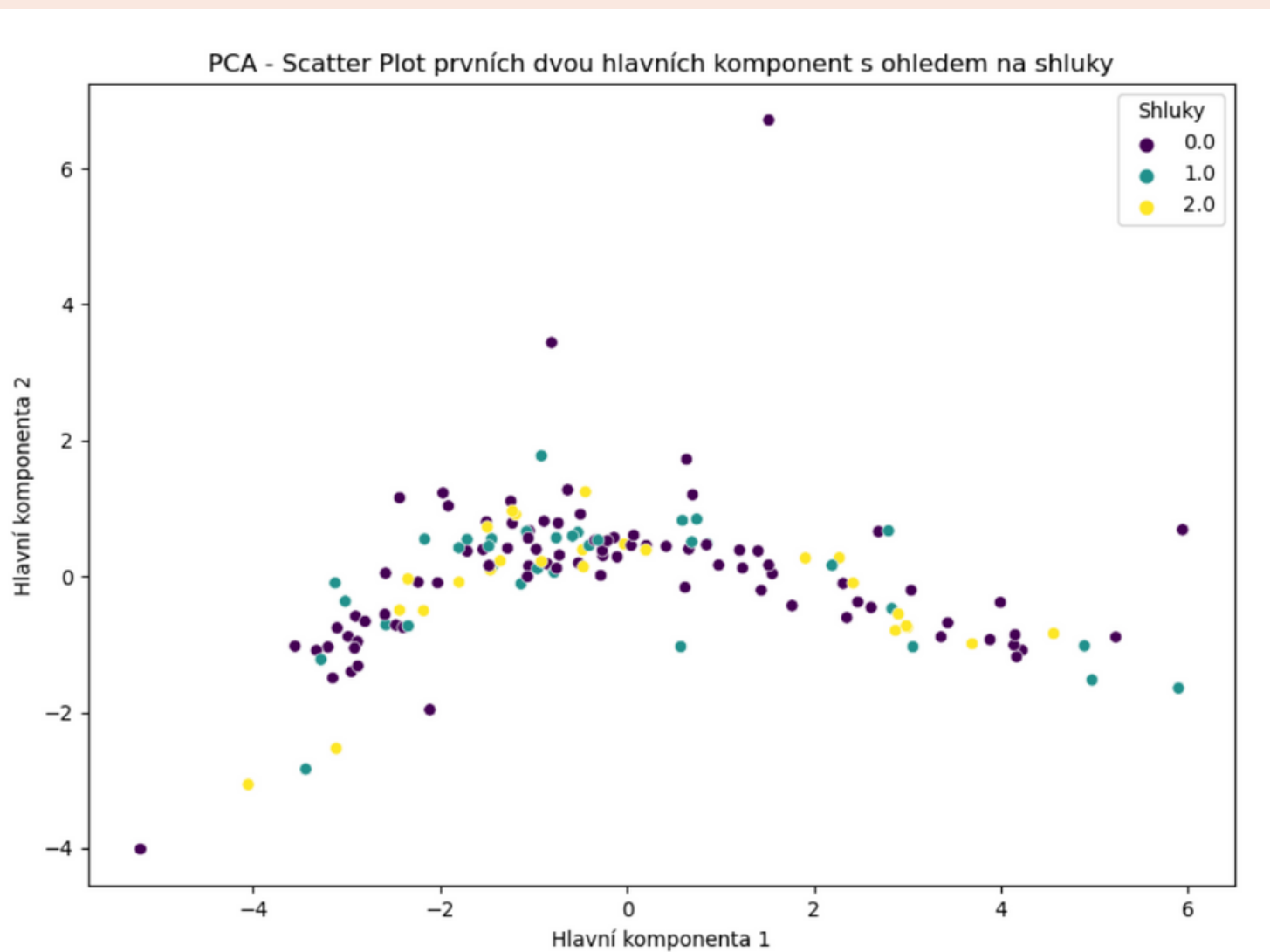


**Cluster 1**

# JAK SE GEOGRAFICKY LIŠÍ VYSOKÁ, STŘEDNÍ A NÍZKÁ ÚMRTNOST NOVOROZENCŮ A JAKÉ SOCIOEKONOMICKÉ FAKTORY MOHOU BÝT SPOJENY S TĚMITO ROZDÍLY?

Kategorizace zemí na základě jejich socioekonomických faktorů

## Redukce dimenze



Metody PCA a Locally Linear Embedding nerozdělily jednotlivé země tak dobře, t-SNE může být nejlepší volbou

# ZÁVĚR

Hlavní výzkumná otázka: Jaký socioekonomický či zdravotní faktor významně ovlivňuje úmrtnost novorozenců v zemích světa?

Výsledky ukázaly, faktory nejvíce ovlivňující úmrtnost novorozenců v zemích světa jsou:

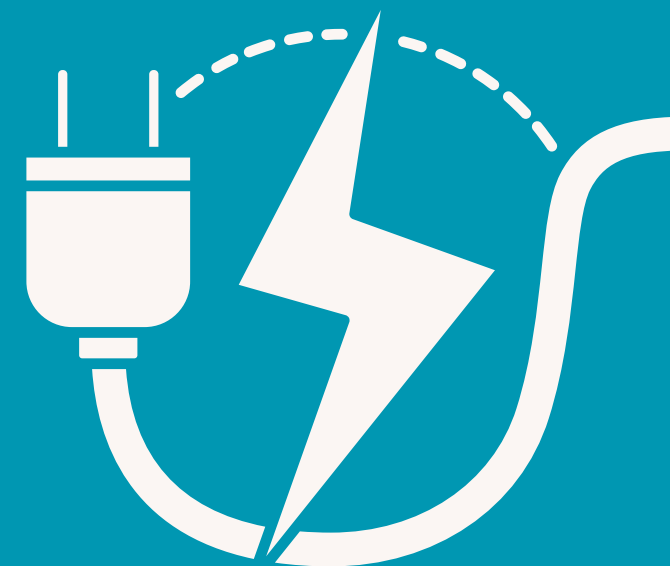
OČEKÁVANÁ DÉLKA  
ŽIVOTA



HDP



PŘÍSTUP K  
ELEKTŘINĚ



# ZÁVĚR

Výsledky jednotlivých použitých metod

**Lineární vs polynominální  
modely**

pro polynominální lepší  
výsledky (vyšší R squared  
pro všechny faktory)

**GLM Poissonův model**  
nejnižší AIC pro kombinaci  
faktorů: HDP, očekávaný  
věk dožití, porodnost,  
urbanizace

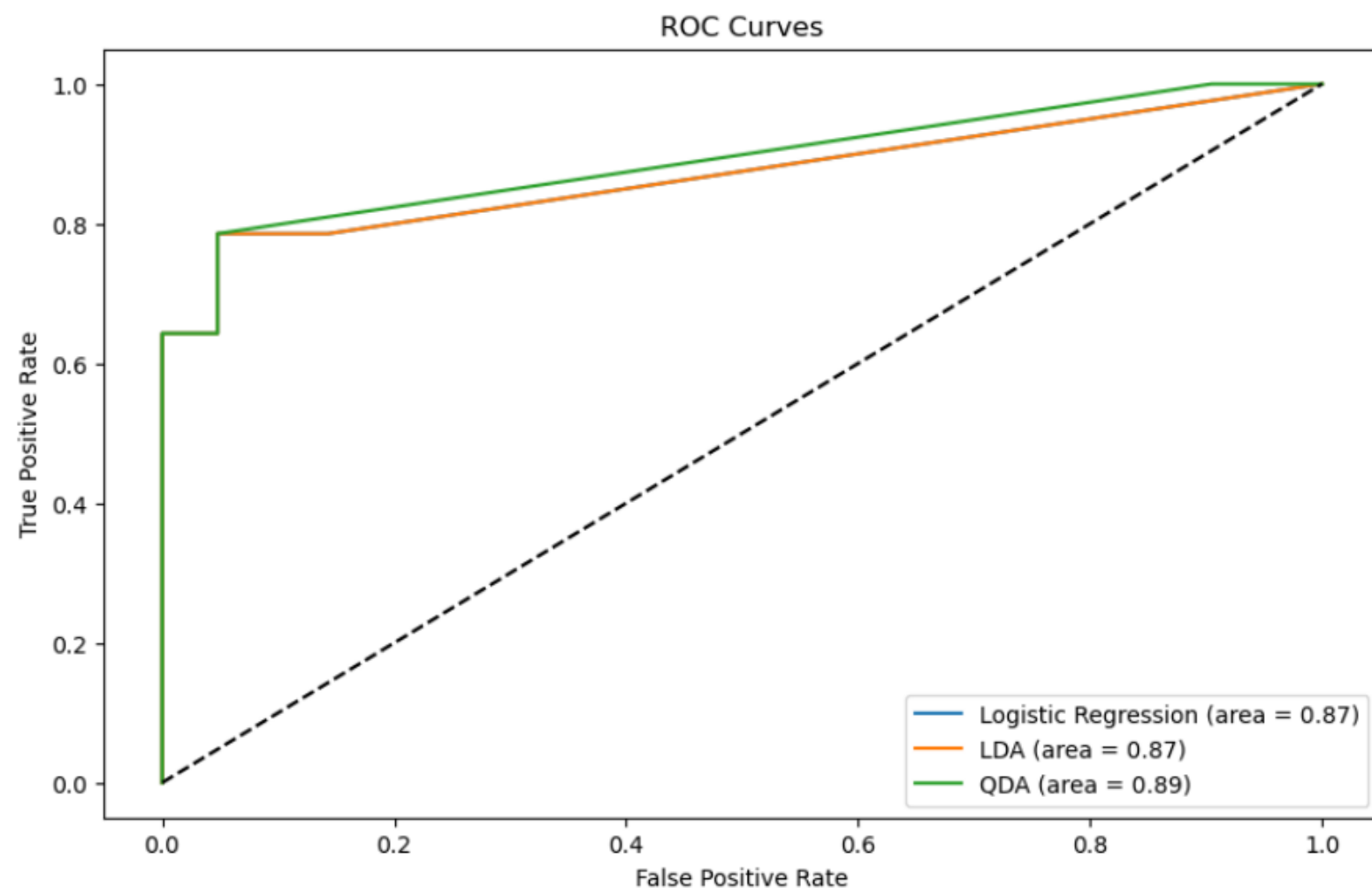
# ZÁVĚR

Výsledky jednotlivých použitých metod

## Binární klasifikace

Log. reg. – nejvyšší přesnost

QDA – nejvyšší ROC-AUC



## Clustering

K-Means: identifikovala 3 shluky

GMM: 3 shluk

Hierarchické shlukování: 3  
shluky

DBSCAN: Identifikoval 2 hlavní  
shluky a několik zemí jako šum

Nejnižší BIC – K-Means

DĚKUJEME ZA  
POZORNOST!