

# Egy járvány terhének mérése – módszerek és modellek

Ferenci Tamás  
ferenci.tamas@nik.uni-obuda.hu

2021. szeptember 14.

# Tartalomjegyzék

- 1 Elméleti keret
- 2 Eredmények
- 3 Kérdések és problémák

# A többlethalózás fogalma

- **Definíció:** adott időszak tényleges halálozásának, és a korábbi időszakok adataiból előrejelzett halálozásának a különbözete
- **Logika:** ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- **Előny:** teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a haláloki besorolástól)
- **Hátrány:**

# A többlethalálozás fogalma

- Definíció: adott időszak tényleges halálozásának, és a korábbi időszakok adataiból előrejelzett halálozásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a haláloki besorolástól)
- Hátrány:

# A többlethalózás fogalma

- Definíció: adott időszak tényleges halálzásának, és a korábbi időszakok adataiból előrejelzett halálzásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a halálki besorolástól)
- Hátrány:
  - Függs az előrejelzés jóságától

# A többlethalálozás fogalma

- Definíció: adott időszak tényleges halálozásának, és a korábbi időszakok adataiból előrejelzett halálozásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a haláloki besorolástól)
- Hátrány:
  - Függ az előrejelzés jóságától
  - Nettó mutató, nem különíthetők el az indirekt hatások (amik ráadásul lehetnek pozitívak is és negatívak is!)
  - Nagyon lassú mutató

# A többlethalózás fogalma

- Definíció: adott időszak tényleges halálozásának, és a korábbi időszakok adataiból előrejelzett halálozásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a haláloki besorolástól)
- Hátrány:
  - Függ az előrejelzés jóságától
  - Nettó mutató, nem különíthetők el az indirekt hatások (amik ráadásul lehetnek pozitívak is és negatívak is!)
  - Nagyon lassú mutató

# A többlethalálozás fogalma

- Definíció: adott időszak tényleges halálozásának, és a korábbi időszakok adataiból előrejelzett halálozásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a haláloki besorolástól)
- Hátrány:
  - Függ az előrejelzés jóságától
  - Nettó mutató, nem különíthetők el az indirekt hatások (amik ráadásul lehetnek pozitívak is és negatívak is!)
  - Nagyon lassú mutató



# A többlethalózás fogalma

- Definíció: adott időszak tényleges halálzásának, és a korábbi időszakok adataiból előrejelzett halálzásának a különbözete
- Logika: ha most valamilyen mortalitást befolyásoló esemény zajlik (ami a korábbi időszakban még nem volt jelen) akkor a fenti adat annak hatását fogja mérni
- Előny: teljesen független a mortalitás regisztrálásának jóságától (jelen kontextusban leginkább ez két dolgot jelent: független a teszteléstől és független a halálki besorolástól)
- Hátrány:
  - Függ az előrejelzés jóságától
  - Nettó mutató, nem különíthetők el az indirekt hatások (amik ráadásul lehetnek pozitívak is és negatívak is!)
  - Nagyon lassú mutató

# Az előrejelzés módszertana

A naiv módszerek (és egyúttal a problémák illusztrálása!):

- Legutolsó év
- Legutolsó néhány év átlaga
- Legutolsó néhány év adatainak meghosszabbítása

# Az előrejelzés módszertana

A naiv módszerek (és egyúttal a problémák illusztrálása!):

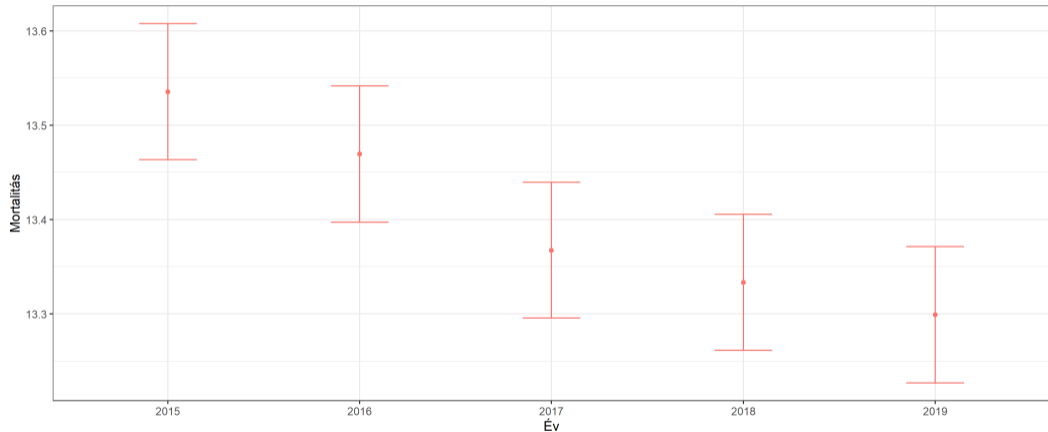
- Legutolsó év
- Legutolsó néhány év átlaga
- Legutolsó néhány év adatainak meghosszabbítása

# Az előrejelzés módszertana

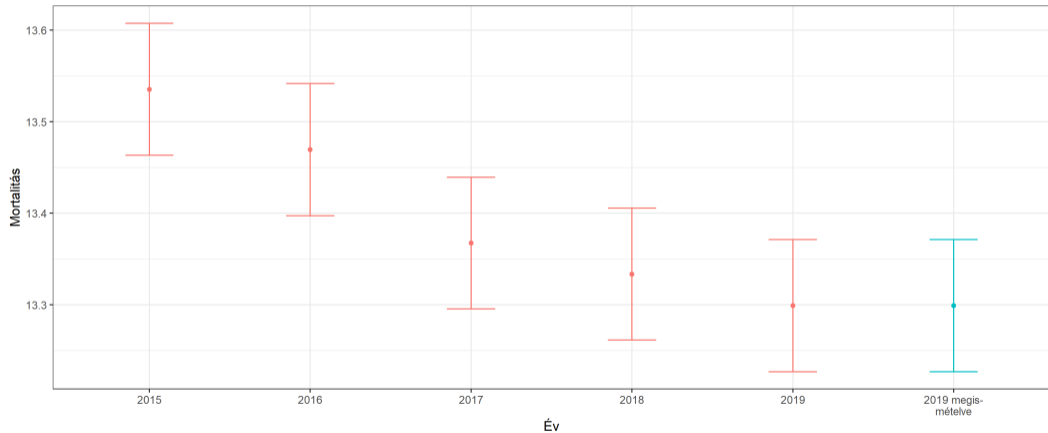
A naiv módszerek (és egyúttal a problémák illusztrálása!):

- Legutolsó év
- Legutolsó néhány év átlaga
- Legutolsó néhány év adatainak meghosszabbítása

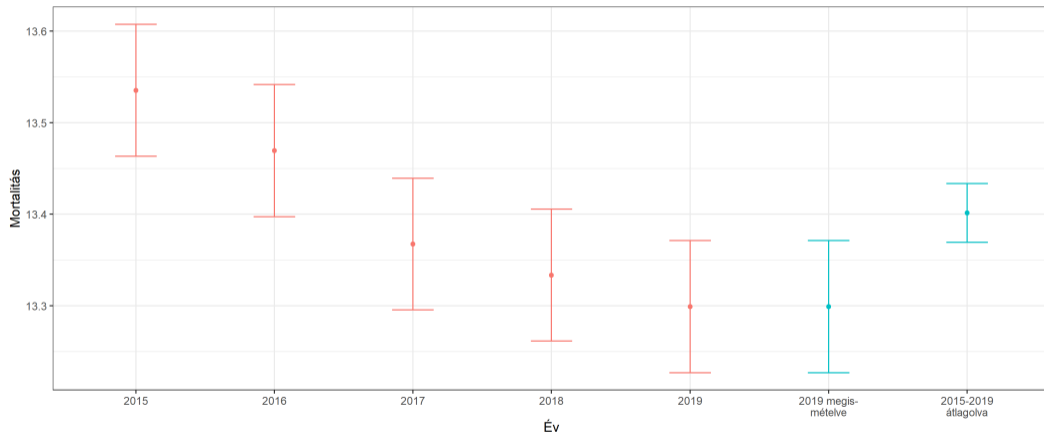
# Példa-ország



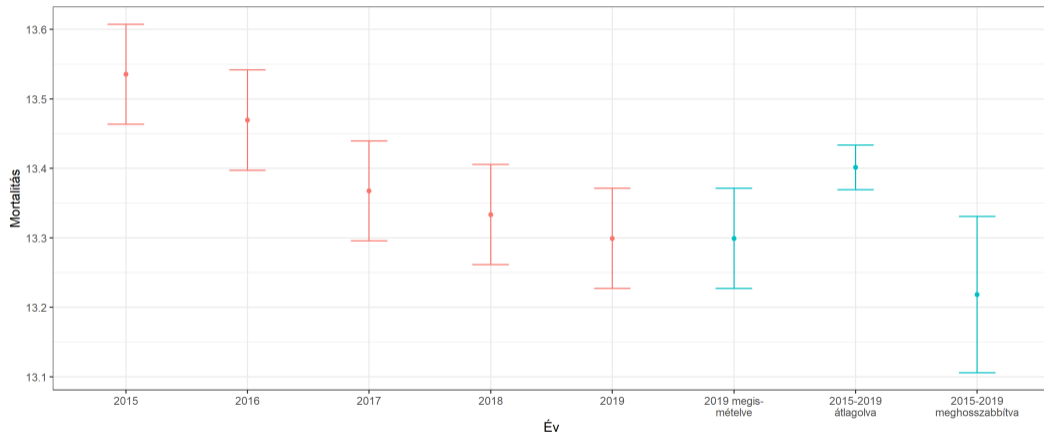
# Legutolsó év



# Legutolsó néhány év átlaga



# Legutolsó néhány év adatainak meghosszabbítása





# Problémák

- **Összbenyomás: bias–variance trade-off**
  - Az utolsó megoldás a jó, de lehetne okosabban extrapolálni
  - További probléma a szezonális (nagyon markáns)

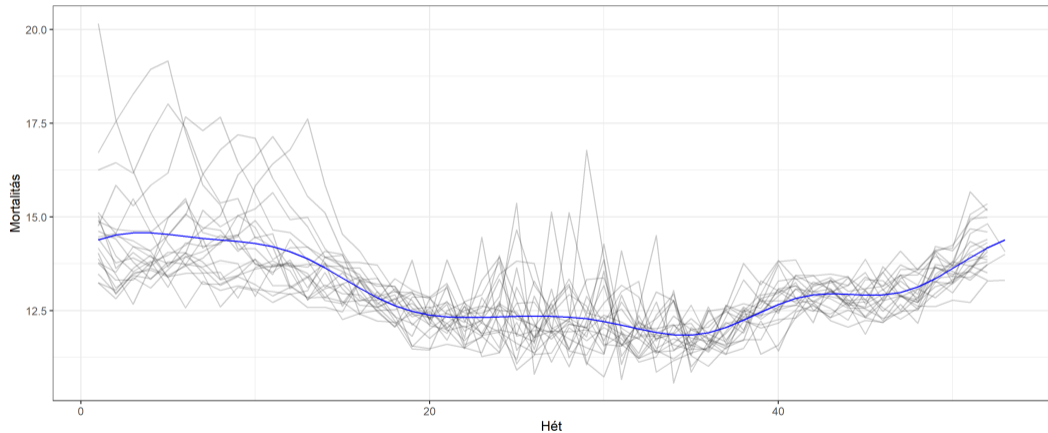
# Problémák

- Összbenyomás: bias–variance trade-off
- Az utolsó megoldás a jó, de lehetne okosabban extrapolálni
- További probléma a szezonális (nagyon markáns)

# Problémák

- Összbenyomás: bias–variance trade-off
- Az utolsó megoldás a jó, de lehetne okosabban extrapolálni
- További probléma a szezonális (nagyon markáns)

# Szezonális



# Előrejelzés kifinomultabb módszerekkel

Acosta és Irizzary (2020):

- Kell egy valószínűségi modell:

$$Y_t | \varepsilon_t \sim \text{Poi}(\mu_t [1 + f(t)] \varepsilon_t),$$

ahol  $\varepsilon_t$  nem feltétlenül fehérzaj, lehet autokorrelált, az adatok autokorreláltságának elszámolására (a napi adat erősen az, a heti nem feltétlenül), legyen többváltozós normális AR( $p$ ) szerinti kovarianciamátrixszal és

$$\mu_t = N_t \exp[\alpha(t) + s(t) + w(t)],$$

ahol  $\alpha(t)$  a hosszú távú – lassan változó – trend,  $s(t)$  az éven belüli mintázat (szezonálisitás),  $w(t)$  pedig a hét napja hatás (ha napi adatunk van), és  $N_t$  a háttérpopuláció

- $f(t)$  lesz a keresett többlet (szorzóként, hiszen log-link mellett multiplikatív az egész modell)
- $f(t)$  lehet például a háttérpopuláció relatív változása, azaz  $f(t) = N_t - N_{t-1}$  (ha a háttérpopuláció nem változik, akkor  $f(t) = 0$ )
- $f(t)$  lehet például a háttérpopuláció relatív változása, azaz  $f(t) = N_t - N_{t-1}$  (ha a háttérpopuláció nem változik, akkor  $f(t) = 0$ )

# Előrejelzés kifinomultabb módszerekkel

Acosta és Irizzary (2020):

- Kell egy valószínűségi modell:

$$Y_t | \varepsilon_t \sim \text{Poi}(\mu_t [1 + f(t)] \varepsilon_t),$$

ahol  $\varepsilon_t$  nem feltétlenül fehérzaj, lehet autokorrelált, az adatok autokorreláltságának elszámolására (a napi adat erősen az, a heti nem feltétlenül), legyen többváltozós normális  $\text{AR}(p)$  szerinti kovarianciamátrixszal és

$$\mu_t = N_t \exp[\alpha(t) + s(t) + w(t)],$$

ahol  $\alpha(t)$  a hosszú távú – lassan változó – trend,  $s(t)$  az éven belüli mintázat (szezonális),  $w(t)$  pedig a hét napja hatás (ha napi adatunk van), és  $N_t$  a háttérpopuláció

- $f(t)$  lesz a keresett többlet (szorzóként, hiszen log-link mellett multiplikatív az egész modell)
- A becslése cseles, ML, de óvatosan kell eljárni ( $\varepsilon_t$  is elég általános, és  $f(t)$  is nézhet ki furcsán, pl. szakadása van)

## Előrejelzés kifinomultabb módszerekkel

Acosta és Irizarry (2020):

- Kell egy valószínűségi modell:

$$Y_t | \varepsilon_t \sim \text{Poi}(\mu_t [1 + f(t)] \varepsilon_t),$$

ahol  $\varepsilon_t$  nem feltétlenül fehérzaj, lehet autokorrelált, az adatok autokorreláltságának elszámolására (a napi adat erősen az, a heti nem feltétlenül), legyen többváltozós normális  $\text{AR}(p)$  szerinti kovarianciamátrixszal és

$$\mu_t = N_t \exp[\alpha(t) + s(t) + w(t)],$$

ahol  $\alpha(t)$  a hosszú távú – lassan változó – trend,  $s(t)$  az éven belüli mintázat (szezonális),  $w(t)$  pedig a hét napja hatás (ha napi adatunk van), és  $N_t$  a háttérpopuláció

- $f(t)$  lesz a keresett többlet (szorzóként, hiszen log-link mellett multiplikatív az egész modell)
- A becslése cseles, ML, de óvatosan kell eljárni ( $\varepsilon_t$  is elég általános, és  $f(t)$  is nézhet ki furcsán, pl. szakadása van)

## Előrejelzés kifinomultabb módszerekkel

Acosta és Irizarry (2020):

- Kell egy valószínűségi modell:

$$Y_t | \varepsilon_t \sim \text{Poi}(\mu_t [1 + f(t)] \varepsilon_t),$$

ahol  $\varepsilon_t$  nem feltétlenül fehérzaj, lehet autokorrelált, az adatok autokorreláltságának elszámolására (a napi adat erősen az, a heti nem feltétlenül), legyen többváltozós normális  $\text{AR}(p)$  szerinti kovarianciamátrixszal és

$$\mu_t = N_t \exp[\alpha(t) + s(t) + w(t)],$$

ahol  $\alpha(t)$  a hosszú távú – lassan változó – trend,  $s(t)$  az éven belüli mintázat (szezonális),  $w(t)$  pedig a hét napja hatás (ha napi adatunk van), és  $N_t$  a háttérpopuláció

- $f(t)$  lesz a keresett többlet (szorzóként, hiszen log-link mellett multiplikatív az egész modell)
- A becslése cseles, ML, de óvatosan kell eljárni ( $\varepsilon_t$  is elég általános, és  $f(t)$  is nézhet ki furcsán, pl. szakadása van)



# Adatok megjelenítése

- Abszolút szám: egyértelmű, de országok között nem összevethető
- Relatív mutató: országok között is összevethető
  - A várt adatokra vetítés (ezzel egy becsült értékre vetítünk, ami nem jó, de ami jó, hogy eleve benne van a korösszetétel, társbetegségek, környezeti tényezők stb. stb. hatása)
  - Lényegében vetítés (figyelnél meg a vetítés határait)
- Milyen pontosságot sugallunk az adatok megadásakor

# Adatok megjelenítése

- Abszolút szám: egyértelmű, de országok között nem összevethető
- Relatív mutató: országok között is összevethető
  - A várt adatokra vetítés (ezzel egy becsült értékre vetítünk, ami nem jó, de ami jó, hogy eleve benne van a korösszetétel, társbetegségek, környezeti tényezők stb. stb. hatása)
  - Lélekszámra vetítés (függeni fog a nyers halandóságtól)
- Milyen pontosságot sugallunk az adatok megadásakor

# Adatok megjelenítése

- Abszolút szám: egyértelmű, de országok között nem összevethető
- Relatív mutató: országok között is összevethető
  - A várt adatokra vetítés (ezzel egy becsült értékre vetítünk, ami nem jó, de ami jó, hogy eleve benne van a korösszetétel, társbetegségek, környezeti tényezők stb. stb. hatása)
    - Lélekszámra vetítés (függeni fog a nyers halandóságtól)
  - Milyen pontosságot sugallunk az adatok megadásakor

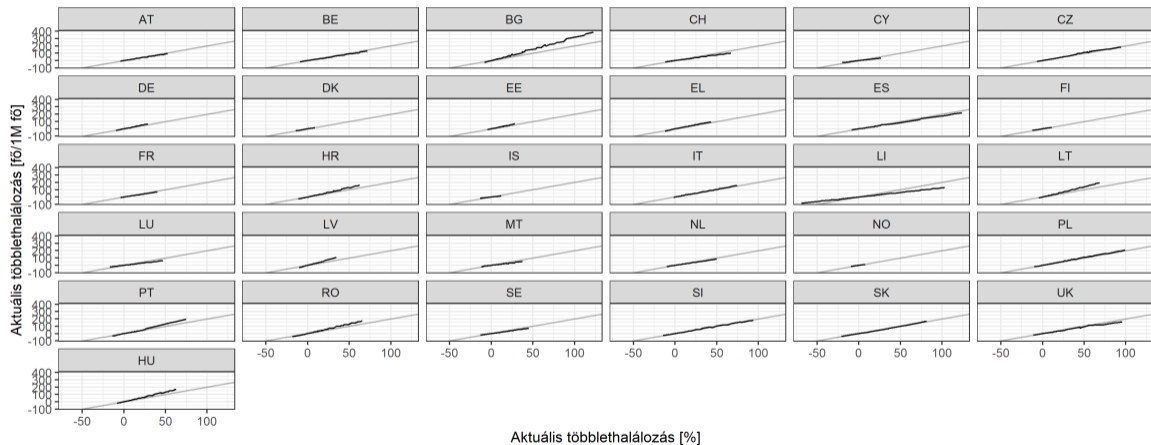
# Adatok megjelenítése

- Abszolút szám: egyértelmű, de országok között nem összevethető
- Relatív mutató: országok között is összevethető
  - A várt adatokra vetítés (ezzel egy becsült értékre vetítünk, ami nem jó, de ami jó, hogy eleve benne van a korösszetétel, társbetegségek, környezeti tényezők stb. stb. hatása)
  - Lélekszámra vetítés (függeni fog a nyers halandóságtól)
- Milyen pontosságot sugallunk az adatok megadásakor

# Adatok megjelenítése

- Abszolút szám: egyértelmű, de országok között nem összevethető
- Relatív mutató: országok között is összevethető
  - A várt adatokra vetítés (ezzel egy becsült értékre vetítünk, ami nem jó, de ami jó, hogy eleve benne van a korösszetétel, társbetegségek, környezeti tényezők stb. stb. hatása)
  - Lélekszámra vetítés (függeni fog a nyers halandóságtól)
- Milyen pontosságot sugallunk az adatok megadásakor

# Relatív mutató kérdései

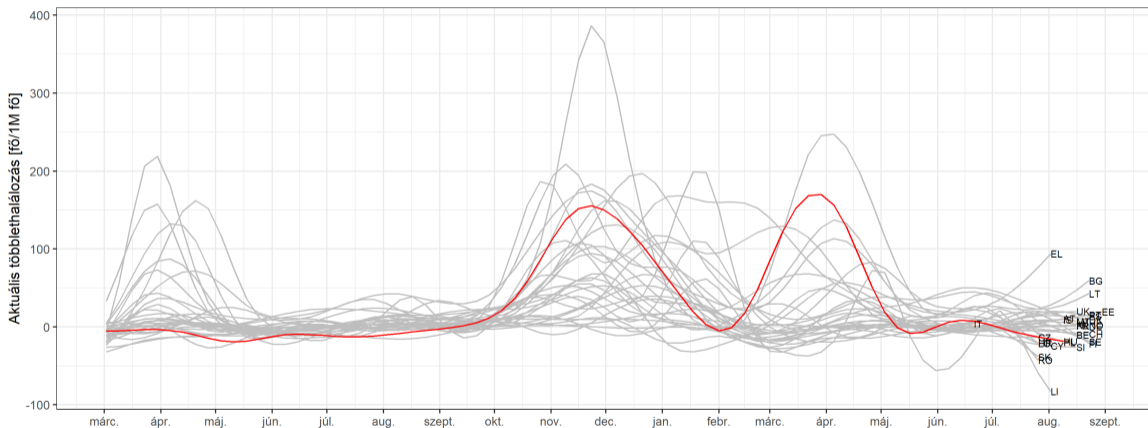


Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Tartalomjegyzék

- 1 Elméleti keret
- 2 Eredmények**
- 3 Kérdések és problémák

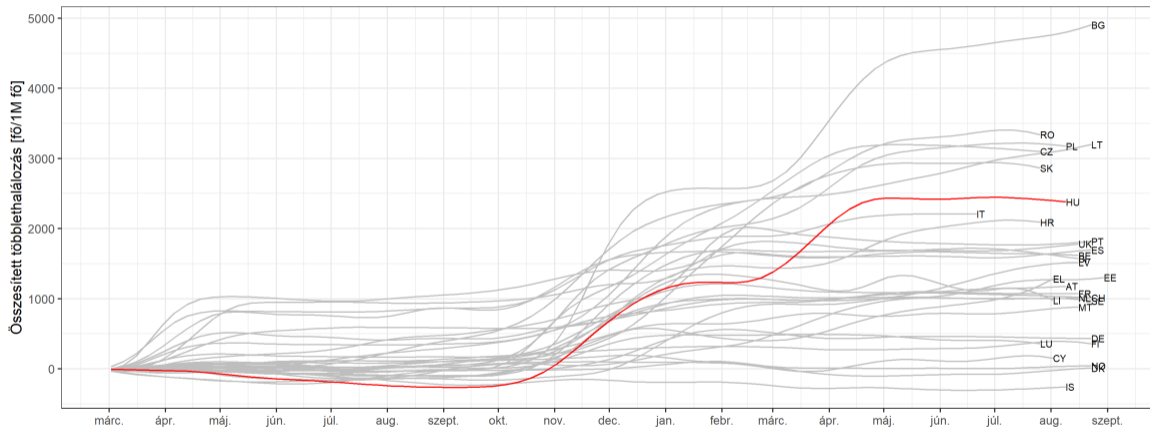
# Aktuális helyzet



Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

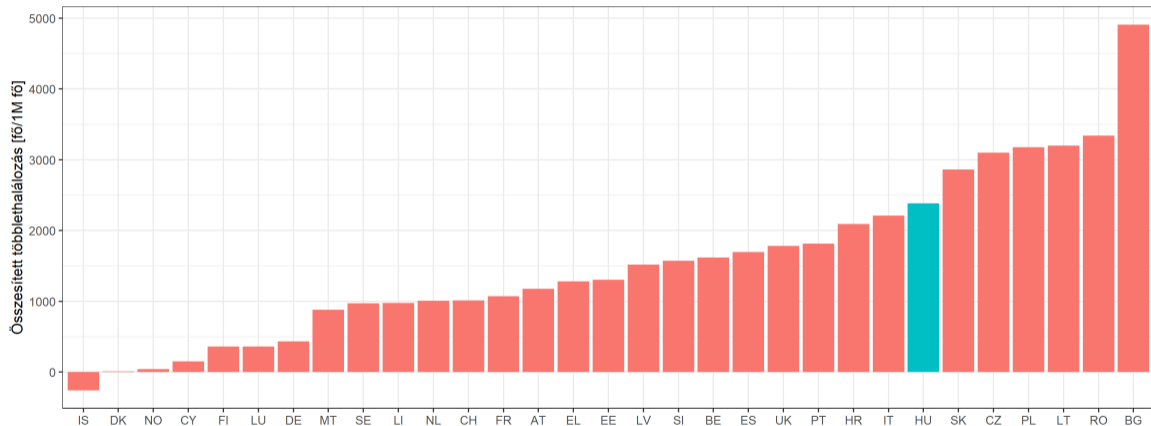


# Kumulált helyzet



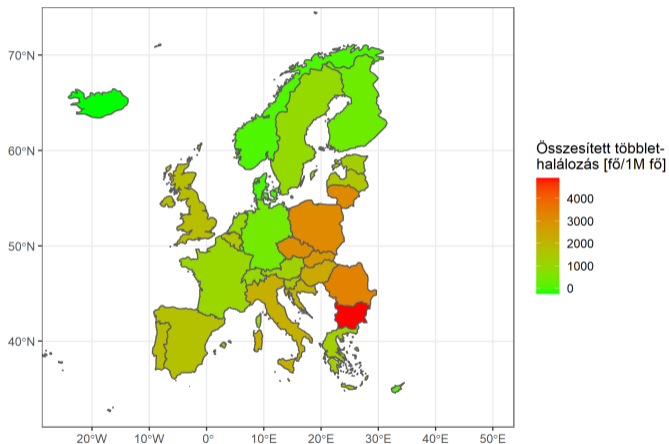
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Kumulált helyzet



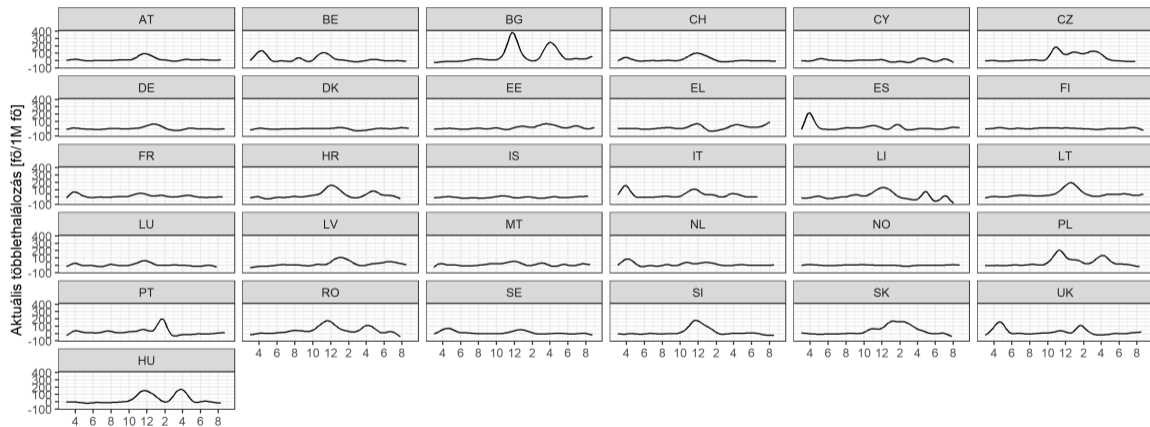
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Kumulált helyzet



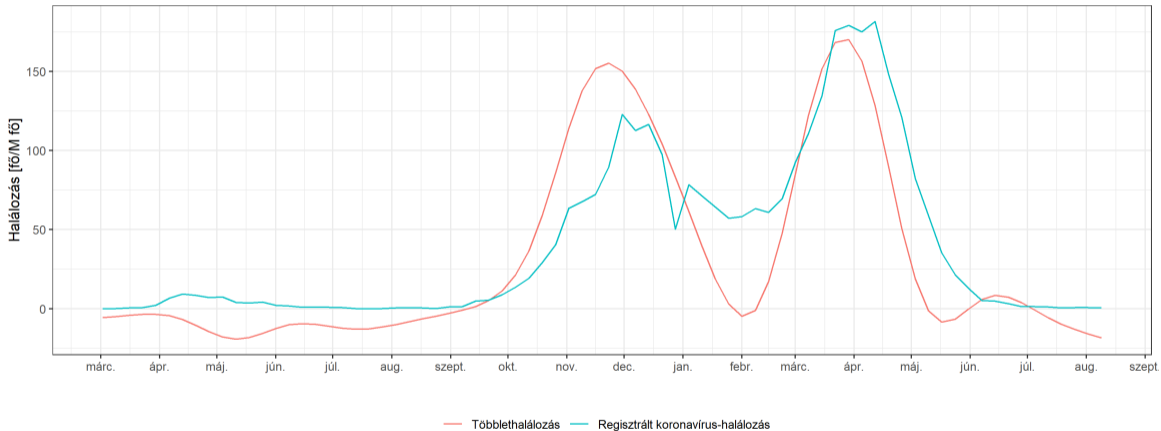
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Országoként külön-külön



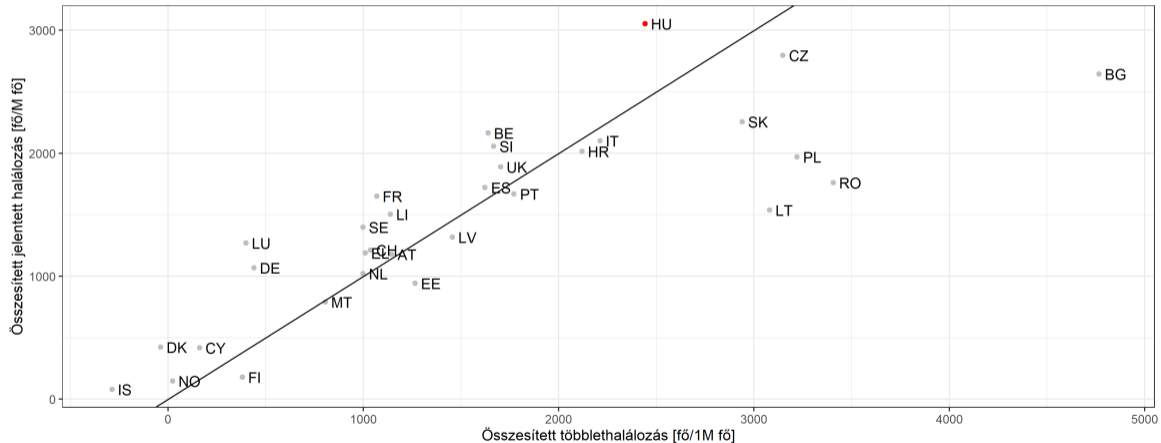
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Jelentett vs. többlet halálozás



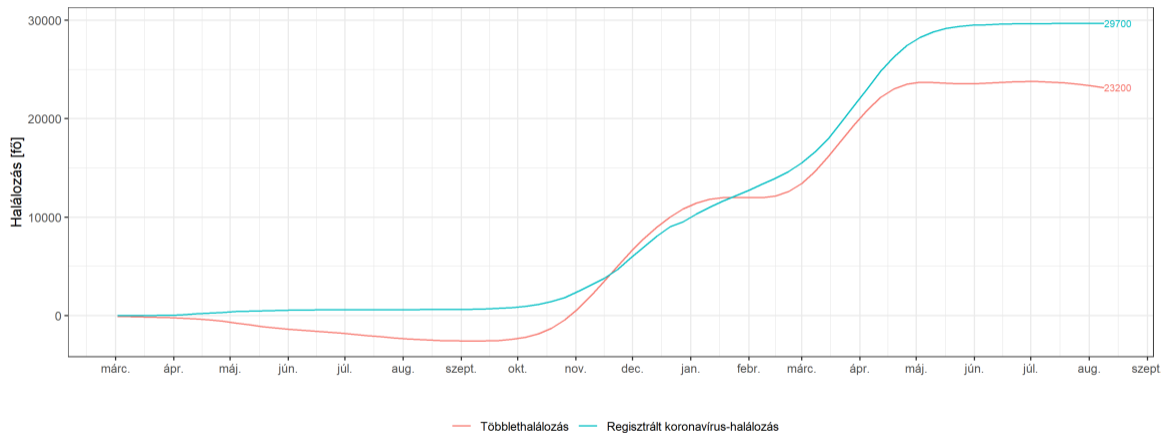
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Jelentett vs. többlet halálozás



Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Magyar helyzet



Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

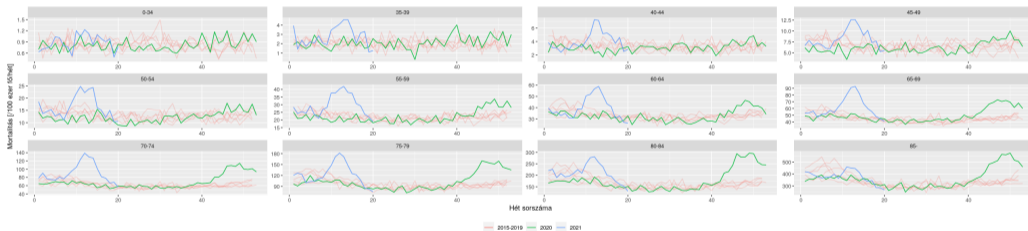
# Tartalomjegyzék

- 1 Elméleti keret
- 2 Eredmények
- 3 Kérdések és problémák



# Könnyebben kezelhető problémák

- Életkor- és kor-specifikus lebontás (hátha mások a tendenciák az egyes életkori vagy nemi csoportokban!)

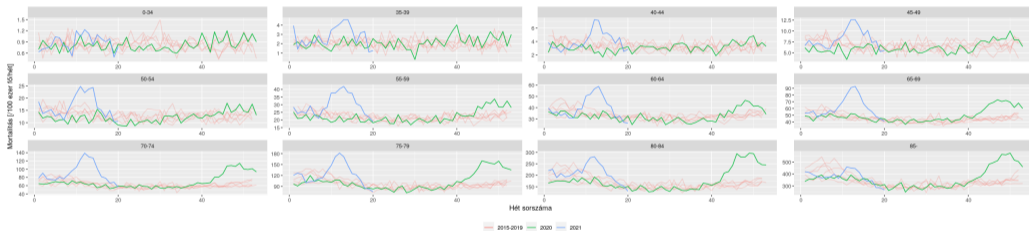


Ferenc Tamás, <https://research.physon.uni-obuda.hu>  
Adatok forrása: KSH

- Standardizálás? (kell egyáltalán?)
- Területi adatok használata

# Könnyebben kezelhető problémák

- Életkor- és kor-specifikus lebontás (hátha mások a tendenciák az egyes életkori vagy nemi csoportokban!)

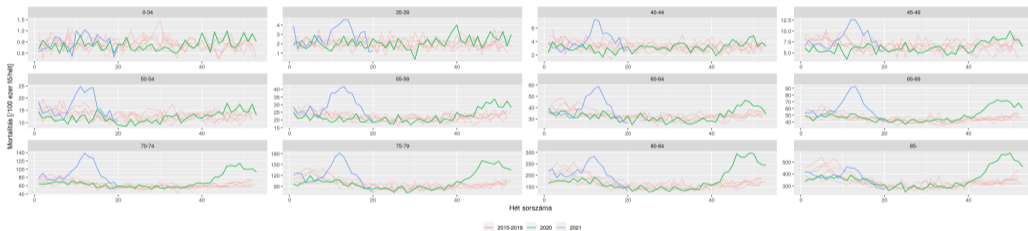


Ferenc Tamás, <https://research.physon.uni-obuda.hu>  
Adatok forrása: KSH

- Standardizálás? (kell egyáltalán?)
- Területi adatok használata

# Könnyebben kezelhető problémák

- Életkor- és kor-specifikus lebontás (hátha mások a tendenciák az egyes életkori vagy nemi csoportokban!)



Ferenc Tamás, <https://research.physon.uni-obuda.hu/>  
Adatok forrása: KGI

- Standardizálás? (kell egyáltalán?)
- Területi adatok használata

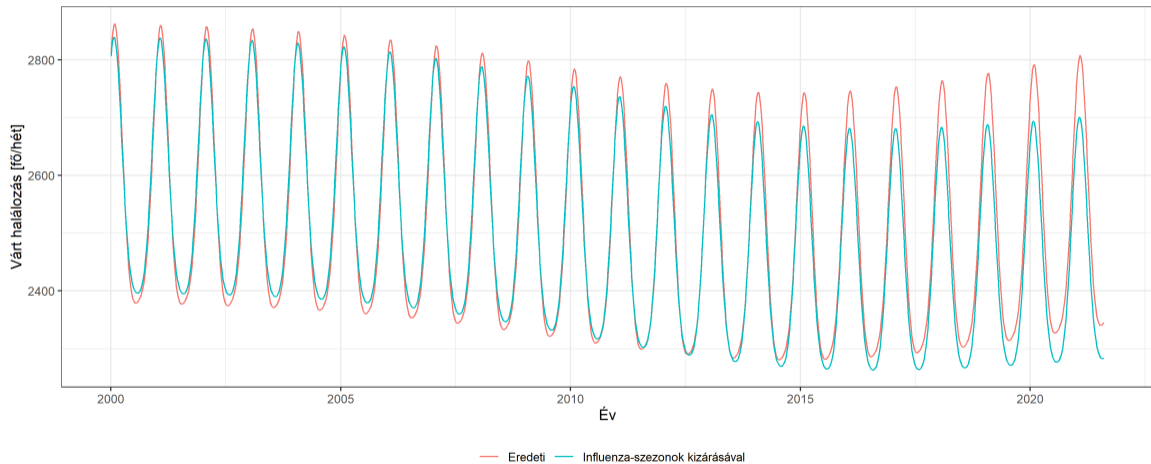
## Nehéz feladat: bruttósítás

- Aknamező: mit vonunk ki? miért pont azt? a többit miért nem?
- Kivonjuk-e a hőhullámot? Az azerbajdzsáni háborút? és... az influenzát?

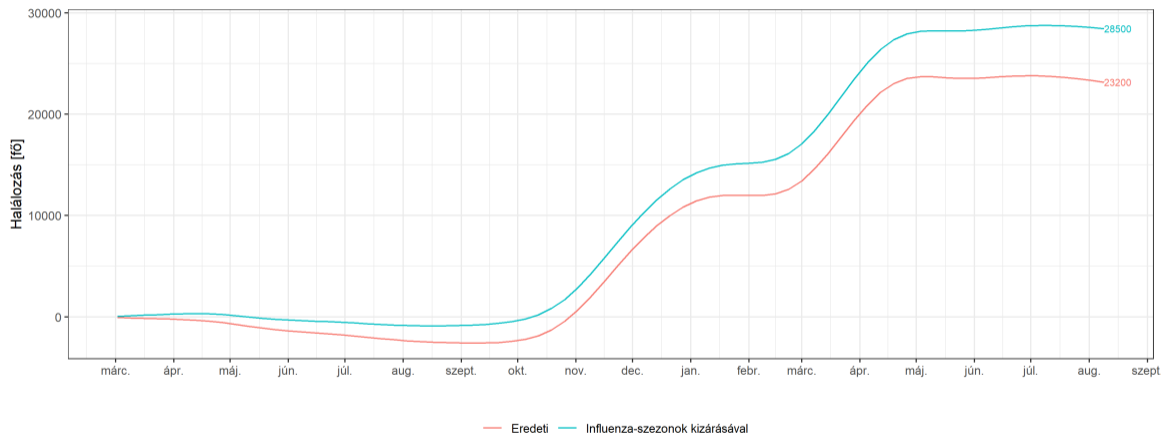
## Nehéz feladat: bruttósítás

- Aknamező: mit vonunk ki? miért pont azt? a többit miért nem?
- Kivonjuk-e a hőhullámot? Az azerbajdzsáni háborút? és... az influenzát?

# Magyar adatok influenzával és a „nélkül”

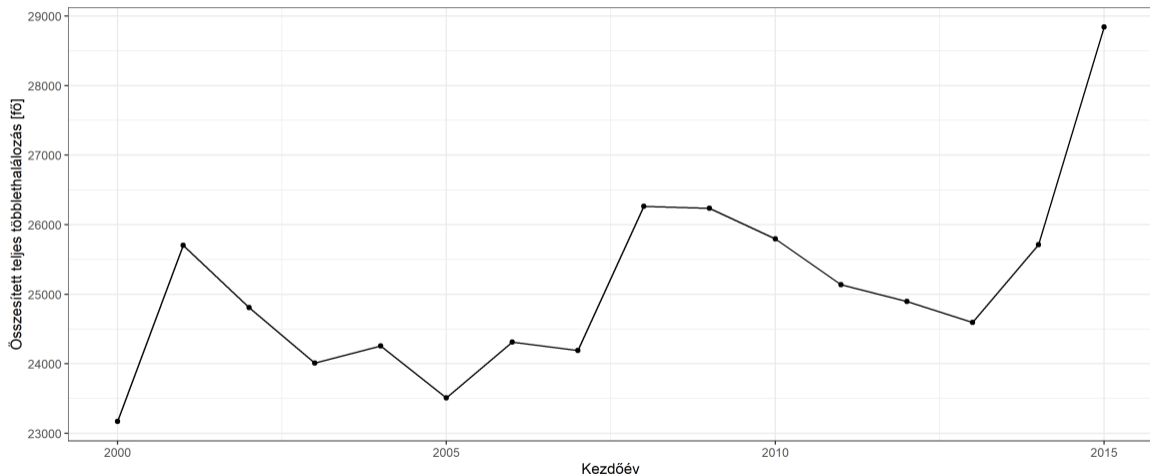


# Magyar adatok influenza „nélkül”



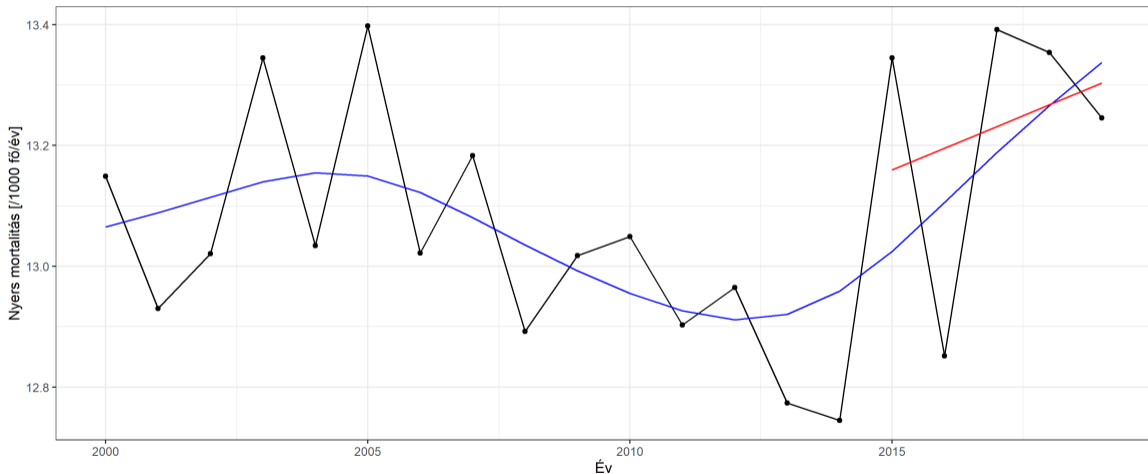
Ferenci Tamás, <https://github.com/tamas-ferenci/ExcessMortEUR/>  
 Adatok forrása: Eurostat és STMF, lekérdezés dátuma: 2021. 09. 13.

# Érzékenységvizsgálat: kezdőév hatása

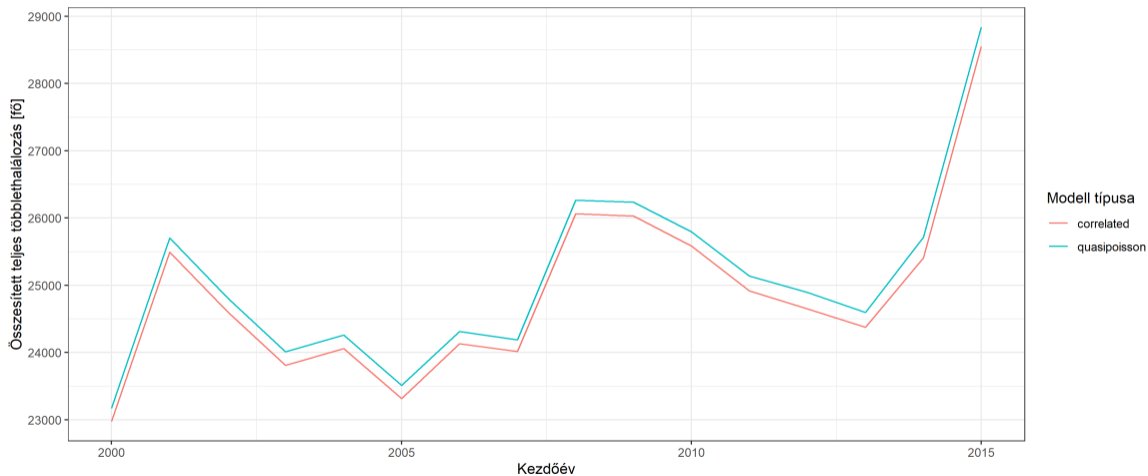




# Érzékenysgvizsgálat: magyarázat a kezdőév hatására



# Érzékenységvizsgálat: korrelációs struktúra



## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
  - Például közeli területi egységek, vagy életkorcsoportok
  - Ha elég nagyok, akkor semmit
  - De ha kicsik? (kis terület, szűk életkori tartomány)
  - Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
  - („Borrowing strength” gondolat)
  - Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
  - Ha elég nagyok, akkor semmit
  - De ha kicsik? (kis terület, szűk életkori tartomány)
  - Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
  - („Borrowing strength” gondolat)
  - Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
- Ha elég nagyok, akkor semmit
- De ha kicsik? (kis terület, szűk életkori tartomány)
- Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
- („Borrowing strength” gondolat)
- Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
- Ha elég nagyok, akkor semmit
- De ha kicsik? (kis terület, szűk életkori tartomány)
- Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
- („Borrowing strength” gondolat)
- Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
- Ha elég nagyok, akkor semmit
- De ha kicsik? (kis terület, szűk életkori tartomány)
- Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
- („Borrowing strength” gondolat)
- Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
- Ha elég nagyok, akkor semmit
- De ha kicsik? (kis terület, szűk életkori tartomány)
- Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
- („Borrowing strength” gondolat)
- Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!



## „Borrowing strength” becslés

- Mit tegyünk akkor, ha egymáshoz hasonló mintázatot mutató csoportokban kell meghatározni az előrejelzést?
- Például közeli területi egységek, vagy életkorcsoportok
- Ha elég nagyok, akkor semmit
- De ha kicsik? (kis terület, szűk életkori tartomány)
- Jó lenne kihasználni, hogy valószínűleg közeli lesz az eredmény
- („Borrowing strength” gondolat)
- Talán a spline-ok használata tűnik a legjobbnak, nyitott kérdés!

## Köszönöm a figyelmet!

E mutatók valós idejű nyomonkövetése:

<https://research.physcon.uni-obuda.hu/COVID19MagyarEpi>

Többlethalálzási adatok európai összevetésben:

<https://github.com/tamas-ferenci/ExcessMortEUR>