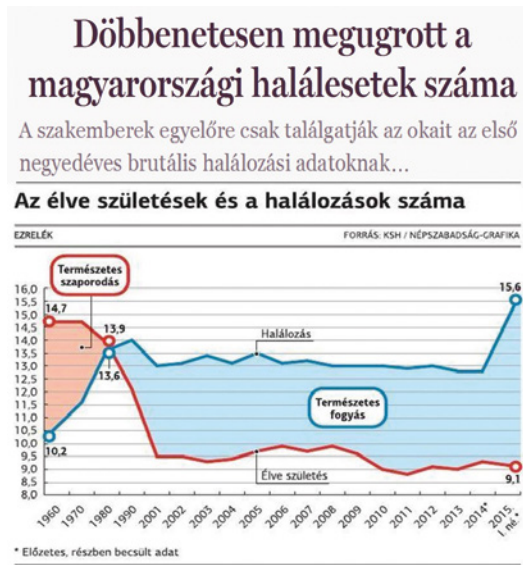


## A STATISZTIKA ÉS AZ ADATELEMZÉS HASZNOSSÁGA

DOI 10.35402/kek.2026.1.23

Amikor kommunikáció és média vagy nemzetközi tanulmányok alapszakon kell bevezető statisztika kurzust tartanom, akkor mindig egy nagy kihívás az első szemináriumon ráirányítani – a kvantitatív témák iránt alapvetően nem érdeklődő – hallgatók figyelmét a statisztika és adatelemzés hasznosságára a saját – elsősorban kvalitatív – szakterületükben. Erre a kihívásra többek között egy olyan bevezető példával reagálok, amiben az alább látható, eléggé erős címmel rendelkező 2015-ös Népszabadság cikkben található adatvizualizációt elemezzük<sup>1</sup>.



Az újságcikkben szereplő idősoros ábra az adatok grafikus ábrázolásának legalapvetőbb gyerekbetegségeit hordozza: a tengelyek nem korrekt kezelést. A függőleges tengely nem nulláról indul, így egy egységnyi változás jelentősebbnek tűnik az ábrán. A vízszintes tengely kezelése pedig egyáltalán nem konzisztens: van, hogy egy egységnyi távolság rajta 10 évet, van, hogy 1 évet jelöl. Ezen kívül az utolsó adatpont egy negyedéves adat, amely éves átlagokkal kerül összevetésre. Ebből adódóan könnyen lehetséges, hogy a „döbbenetesen megugró 2015-ös halálozási szám” egy természetes szezonális ingadozás eredménye, amely a 2015. évi teljes átlagot

nézve egyáltalán nem lesz kiugró, mivel a maradék három negyedév kisimítja azt. Amennyiben a KSH adatai<sup>2</sup> alapján elkészítjük a magyar halálozások számát bemutató idősoros vizualizációt a szakmai szabályok követésével, azaz korrekt és konzisztens tengelykezeléssel, megbizonyosodhatunk róla, hogy a sejtésünk a szezonális ingadozásáról beigazolódik.



A halálozások számának 2015-2020 közötti, negyedéves bontású alakulása egyértelműen megmutatja, hogy az első negyedévre mindig egy éves átlaghoz képest magasabb halálozás szám jellemző, amelynek mértékét alapvetően a szezonális influenzajárvány befolyásolja. A 2015-ös influenzajárvány egy relatíve súlyos influenzajárványnak számított, de semmivel nem súlyosabb az adatok alapján, mint a 2017-es vagy 2019-es. Érdemi változást a szezonális tendenciákban csak a koronavírus járvány második hulláma hoz 2020 végén.

Ez az egyszerű kis példa tapasztalataim alapján kellően megragadja a kommunikáció és média vagy nemzetközi tanulmányok szakos hallgatóság figyelmét is, és felhívja a figyelmet arra, hogy még ezeken a szakterületeken is elengedhetetlen az adatkezelés és adatvizualizáció szakmai alapelveinek ismerete. Ugyanis, ezek következetes alkalmazásával elkerülhető, hogy pontatlan, vagy éppen feleslegesen pánikot keltő cikkek, tanulmányok készüljenek közérdeklődésre számot tartó kérdésekben.

Sajnálatos módon a bevezető statisztika tárgyának tananyagában jellemzően csak egy szemináriumnyi idő jut az adatvizualizáció témakörének, így az alapok tisztázásán túl nem igazán jut lehetőség a klasszikus egyetemi oktatás kereteiben elmélyedni a témában olyan kvalitatív jellegű szakokon, mint például a kommunikáció és média. Nagyon szívesen javasoltam volna a hallgatóimnak eddig is olyan könyvet, amely az adatvizualizációt részletesebben tárgyalja. Azonban, eddig magyar nyelven

<sup>1</sup> A cikk teljes terjedelemben [itt](#) elérhető. (letöltve: 2018. 09. 02.)

<sup>2</sup> [https://www.ksh.hu/stadat\\_files/nep/hu/nep0065.html](https://www.ksh.hu/stadat_files/nep/hu/nep0065.html) (letöltve: 2025. 06. 18.)

olyan könyvet nem találtam, amely jól egyensúlyozott volna a szemléletes, érdekesítő példák és az adatvizualizáció szakmai szabályainak korrekt bemutatása között. Jellemzően az eddig általam ismert könyvek vagy az egyik, vagy a másik irányba „dőlték el”. Vagy nagyon látványos és komplex példákat mutattak az olvasó lenyűgözésére, ám az adatvizualizáció szakmai elveiről és szabályairól szinte semmit nem írtak. Ezeket szoktam „National Geographic-stílusú könyvként” emlegetni. Érdekeségeket meg lehet belőle tudni a területről, de átfogó, szakmailag is korrekt képet nem kap belőlük az olvasó. A legtöbb ilyen jellegű könyv alcímében is utal arra, hogy a téma megközelítése vizuális kommunikáció/pszichológia nézőpontú. A másik véglet a nagyon száraz tankönyv jelleget öltő munkák csoportja, ahol az adatvizualizáció szabályait kifogástalan szakmai profizmussal mutatták be, ám a példák nem voltak annyira látványosak, hogy egy kvantitatív terület iránt alapvetően nem kvantitatív szempontból érdeklődő olvasó érdeklődését is fenntartsák. Ennek a megközelítésnek a „csúcsa” Leland Wilkinson *The Grammar of Graphics* című munkája, amely az adatvizualizációt matematikai formalizmussal közelíti meg. E két szempont közti egyensúlyt képviseli Bátorfy Artilla: *Adatvizualizáció* című könyve, amely hiánypótlónak bizonyul a téma bevezető szintű hazai irodalmában.

Bátorfy munkája a szemléletes és látványos példák, valamint a vizualizációs szerkesztés szabályainak korrekt bemutatása közötti harmóniát teljes terjedelmében megőrzi. A könyv első fejezete az adatvizualizáció elméleti alapjainak tárgyalását foglalja magában, amely történeti fejlődését is bemutatja. Ezen a ponton külön érdemes kiemelni, hogy a témát nem csak kvantitatív nézőpontból közelíti meg, hanem annak (képző)művészeti hátterét és szempontjait is bemutatja. Ilyen szempontból például statisztikus kollégáknak is bátran tudom ajánlani a könyv forgatását, mivel számos olyan ábrázolási vagy színkezelési szemponttal találkoztam a könyvben, amelyek az adatvizualizáció statisztika és/vagy adattudomány keretrendszeren belül tárgyaló munkákban szinte egyáltalán nem kapnak hangsúlyt. Ebből adódóan statisztikusként is új ismeretekkel gazdagodtam a könyvből az adatvizualizáció művészeti aspektusairól.

Az elméleti keretek bemutatása után két fejezetben foglalkozik a szerző az adatvizualizáció rendszerével és módszertanával, azaz az adatok grafikus ábrázolásának technikai lehetőségeivel és azok megfelelő alkalmazásainak szabályrendszerével. E két

fejezet kapcsán kiemelném a könyv azon egyedi nézőpontját, hogy ezen elsősorban száraznak tűnő témák szemléltetéséhez kiváló érzékkel adagolja a gyakorlati példákat. A szerző kiemelkedően odafigyel arra, hogy minden diagramtípushoz, minden alkalmazási módszerhez (vagy éppen kerülendő megoldáshoz) mutasson egyszerű és összetett gyakorlati példákat is. A példák kimondottan üde színfoltja, hogy nem csak a leghíresebb nemzetközi megoldásokat mutatja be, hanem saját gyakorlati tapasztalataiból is merítve korszerű, magyar vonatkozású vizualizációkkal szemlélteti az elméleti megfontolások gyakorlati alkalmazását.

A könyv további nagy erényének tartom, hogy támpontokat ad az adatvizualizációra alkalmas szoftveres megoldások tekintetében is. A szerző gyakorlati adatvizualizációs szakemberhez méltó módon egy saját készítésű vizualizációval szemlélteti saját munkafolyamatát, és annak egyes állomásain alkalmazott szoftveres megoldásokat. Így a kezdő olvasó könnyen megtalálja, hogy a saját érdeklődési szintjének gyakorlati műveléséhez milyen szoftverek kezelését szükséges elsajátítania. A könyv egy konkrét adatvizualizációs szoftver működését sem mutatja be, de egy ilyen bevezető jellegű átfogó és rendszerező munkának ez nem is feladata. A könyvben megadott támpontok bőven elégségesek ahhoz, hogy az érdeklődő el tudjon indulni a saját vizualizációs készítésének útján.

Hiányérzetem a könyvben az adatvizualizáció rendszere és módszere kapcsán volt, lényegében két ponton. Mindkét hiányérzet nagy valószínűséggel a statisztikus szemüvegemen keresztül olvasatból fakad, és a laikus érdeklődők számára ettől még a könyv nem közvetít téves vagy érdemben hiányos képet az adatvizualizáció rendszeréről és módszereiről. Ugyanakkor, mindkét hiányosságot „kihagyott ziccernek” érzem. Ki is fejtem miért.

Az első hiányérzetem, hogy a szerző látványosan kerül a megfigyelés és változó (vagy ismérv) fogalmak bevezetését és használatát. Ennek kapcsán pedig nem beszél az adatszerkezetek (keresztmetszet, idősor, panel) és a változók mérési skáláinak (nominális, ordinális, intervallum, arány) fogalmairól sem. A döntés egyrészt teljesen érthető, mivel ezek elsősorban már a klasszikus statisztika és adatelemzés területének fogalmai, és ha a könyv célja, hogy akár képzőművészettel foglalkozók figyelmét is az adatvizualizáció területére irányítsa, akkor nem biztos, hogy szerencsés lett volna ennyire „kemény” fogalmakkal élni. A munka egyik kiemelkedő eredménye, hogy ezen fogalmak nélkül is szakszerűen

mutatja be az adatvizualizáció rendszerét és módszereit. Implicite végigvonul a könyvön az a gondolat, hogy a megfigyelések szerkezet és a változók mérési skálája nagymértékben meghatározza a belőlük készíthető adatvizualizációk azon körét, amelyek információértékkel bírnak és következtetések levonására jól felhasználhatók. Ugyanakkor a „kemény” fogalmak használatával még egységesebb rendszerbe lehetett volna foglalni a bemutatott adatvizualizációs megoldásokat, és megmutatni a rendszert a néhány esetben a könyv alapján némileg ad hoc-nak tűnő vizualizációs megoldások mögött. Ez nagyban segítené a kezdő adatvizualizáció készítőket abban, hogy olyan gyakori hibákat kövessenek el, mint hogy ordinális változók közti kapcsolatokat pontdiagramon próbáljanak meg szemléltetni (az értékek közti távolság mérhetőségének helytelen feltételével) ahelyett, hogy azok együttes diszkrét gyakorisági eloszlásukat vizsgálnák.

A másik lényegi hiányérzetem az adatminőségi kérdések bemutatásának hiánya. A „selejtől selejt” (angolban garbage in, garbage out) kifejezés számos kontextusban értelmezhető, így az adatvizualizáció területén is. A rossz minőségű (hiányos, anomáliákkal szennyezett stb.) adatforrásokból az adatvizualizáció – könyvben is bemutatott – szakmai szabályait hibátlanul és következetesen alkalmazva is olyan vizualizációkat kapunk eredményül, melyek félrevezetőek, hibásak lesznek. Az adatforrás minősége egy olyan korlátozó tényező,

amellyel minden adatvizualizációval foglalkozó szakembernek tisztában kell lennie, és amennyire lehetséges az adatforrás anomáliáit azonosítani kell, majd okait feltárni és kezelni. Ebben a folyamatban az adatvizualizáció egyszerű eloszlásokat jellemző diagramjai (pl. hisztogram, doboz ábra) rengeteg támogatást tudnak nyújtani. Ennek legfrissebb gyakorlati példája Tátrai Annamária és Gábos András tanulmánya<sup>3</sup>, amelyben egyszerű hisztogramok segítségével azonosítják a KSH által szolgáltatott EU-SILC jövedelem adatok anomáliáit. Sajnáltam, hogy a könyv ilyen jellegű adatminőséget ellenőrző adatvizualizációs alkalmazásokkal nem foglalkozik.

Mіндеzen hiányosságok ellenére bátran ajánlom Bátorfy Attila munkáját az adatvizualizációval ismerkedni kívánó, alapvetően humán terület felől érkező érdeklődőknek, mivel könnyen emészthető, szemléletes és mégis szakmailag teljes mértékben helytálló bevezetést kapnak az adatvizualizáció rendszerébe. A „keményebb” fogalmakkal való megismerkedés pedig bőven ráérhet a tanulási folyamat következő fázisában. Ezen kívül a könyv érdekes és hasznos lehet gyakorló adattudósoknak, vagy adatelemzéssel foglalkozó egyéb „kvantitatív” szakembereknek is. Személy szerint az adatvizualizációk művészeti és esztétikai kérdéseiről magam is sokat tanultam könyvből.

Bátorfy Attila: Adatvizualizáció - Bevezetés az adatok grafikus ábrázolásának elméletébe és gyakorlatába, ELTE, Eötvös Kiadó Kft. 2024, 216.old. ISBN, 9789633123973.

<sup>3</sup> <https://www.valaszonline.hu/2025/04/01/eu-silc-magyarorszag-i-jovedelemadok-kutato-i-elemzes/> (letöltve: 2025. 06. 20.)