

FACSARJ

% EGY KIS INFORMÁCIÓT AZ ADATOKBÓL!

FACSARJ EGY KIS INFORMÁCIÓT AZ ADATOKBÓL!

A Budapesti Corvinus Egyetem Közgazdaságtudományi Karának középiskolás adatelemző versenye, az Adatelemző Központ szervezésében.

ADATOK GRAFIKUS MEGJELENÍTÉSE

Bevezetés

A legtöbb adatelemzési folyamat első lépése, hogy a rendelkezésünkre álló adatot ábrázoljuk. Egy-egy jól elkészített ábra segíthet az adataink megértésében, a főbb jellemzők és összefüggések feltárásában, érdekes mintázatok azonosításában, esetleges adathibák megtalálásában, megfigyeléseink kategorizálásában. Ugyanakkor a legtöbb elemzési folyamatnak nem csupán az első, hanem az utolsó lépését is az adatok grafikus megjelenítése adja. Elemzési folyamatunk végtermékét, az eredményeinket leírhatjuk szövegesen, bemutathatjuk táblázatokban, ugyanakkor grafikusan is megmutathatjuk őket. Sőt, gyakran, például egy kollégának, osztálytársaknak, vizsgabizottságnak, vagy zsűrinek bemutatott szóbeli prezentáció esetén az ábrák alkalmazása lehet a legcélravezetőbb eszköz, hiszen egy jól elkészített ábra sokkal inkább megragadja a figyelmet, mint egy táblázat, vagy egy hosszú folyószöveg.

Versen्यünk első körében tehát ezzel a roppant izgalmas és hasznos témakörrel, az adatok grafikus megjelenítésével foglalkozunk. A forduló célja, hogy a kiadott adatok felhasználásával olyan ábrát, vagy ábrákat készítsetek, amelyek könnyen értelmezhetőek, megfelelnek mindenféle fontos követelménynek és válaszolnak azon kérdésekre, amelyeket felteszünk. A feladatkiírás 2. pontjában bemutatjuk az ábrakészítés néhány nagyon fontos szabályát, majd a 3. pontban röviden illusztráljuk, mi a különbség egy rossz és egy jó ábra között. A 4. pontban pedig a feladatkiírást tesszük közzé, az általatok megoldandó feladatot mutatjuk be.

Az ábrakészítés esszenciája

Egy ábra elkészítése négy fő lépésből áll. Első lépésként meg kell határoznunk azt a világos **üzenetet**, amelyet az ábra segítségével közvetíteni akarunk. Ha nincs üzenet, akkor ábrát készíteni is felesleges. Az üzenet legyen jól definiált, kellőképpen körülhatárolt! Szeretném megmutatni, hogy a magyar GDP az elmúlt húsz évben trendszerűen növekedett. Szeretném megmutatni, hogy az áramfogyasztás mennyiségében napon belüli szezonális van. Szeretném megmutatni, hogy azon országokban, ahol magasabb az egy főre eső csokoládéfogyasztás, több a Nobel-díjasok száma (ha

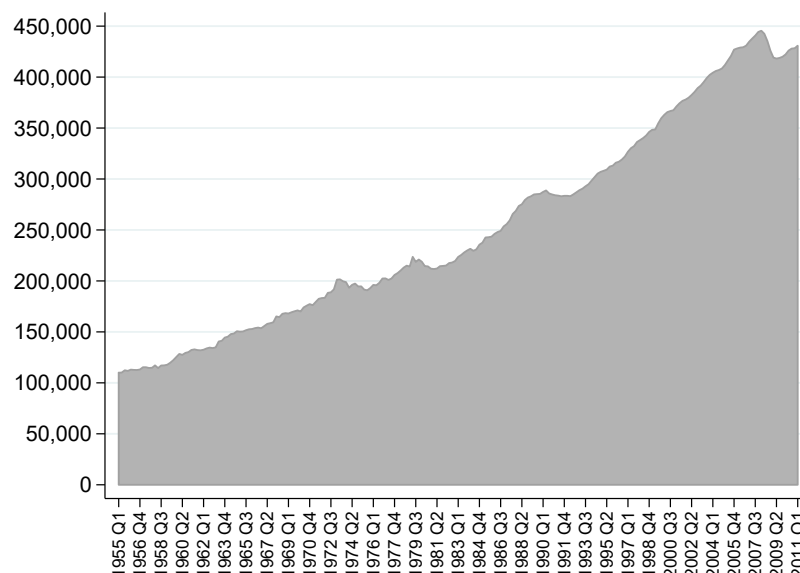
kellően sok csokit eszem, lehet, hogy Nobel-díjas leszek? – ezt a kérdéskört is érintjük majd egy későbbi fordulóban). Ezek mind jó üzenetek.

Második lépésként kategorizálnom kell az üzenetemet. Azt szeretném megmutatni, hogy egy csoport egy része mekkora arányt képvisel az egészből (például, az osztálytársaim között mennyi balkezes van)? Valaminek az időbeli alakulását szeretném megmutatni (például, hogyan változott a testsúlyom az elmúlt 15 évben)? Esetleg valamilyen dolgok sorrendjét akarom ábrázolni (a középiskolai rangsor első húsz helyezettje), vagy valamilyen dolgok közötti kapcsolatra (cigarettafogyasztás és tüdőrák kialakulásának valószínűsége) szeretnék rávilágítani? Netalán valami egészen mást akarok? Mielőtt ábrát készítek, pontosan meg kell határoznom, hogy milyen típusú **összehasonlítást** szeretnék végezni.

Harmadik lépésként ki kell választanom az első két lépéshez **legjobban illeszkedő ábrát**. A különböző jelenségekre különböző típusú vizuális eszközök segítségével tudom felhívni a figyelmet – erről majd még később lesz szó. Végül pedig, az utolsó lépésben, vissza kell térnem az ábrakészítési folyamat legelejére és fel kell tennem a kérdést: **Az elkészült ábra közvetíti azt az üzenetet, amit az elején megfogalmaztam?** Ha igen, akkor kész van az ábrám.

Rosszból jó, avagy egy ábra evolúciója

Ebben a részben egy rövid példával illusztráljuk, hogy miben különbözik egy jó és egy rossz ábra. A makroökonómia egyik legfontosabb mutatószáma a bruttó hazai termék, rövidebb nevén GDP. A GDP a gazdaságok teljesítményét méri és nagyon nagy érdeklődés övezi. Közgazdászok elemzik, a különböző médiumokban nap, mint nap idézik, nem nagyon létezik olyan ember, aki ne hallott már volna a GDP-ről. A GDP módszertanával, elemzésével, történetével rengeteg tanulmány és könyv foglalkozik. Például, a brit Diane Coyle GDP: A brief but affectionate history (magyarul: GDP – Egy rövid, de meghatározó jelentőségű történet) című könyve. A könyv egy nagyon olvasható, izgalmas írás, amely a GDP fejlődéstörténetével foglalkozik, egészen a kialakulásától a mai, modern gazdaságokban elfoglalt szerepéig. A könyv egyelőre csak angol nyelven kapható, de talán egyszer anyanyelvünkön is megjelenik! A könyv belső borítóján szerepel egy ábra, ezt az ábrát használjuk most „elrettentő példaként”. A könyv egyébiránt nagyon jó, tanulságos, széles olvasótáborra szólít meg, csak ezt a fránya ábrát sikerült elszúrni, de nagyon! Íme, a könyvben szereplő ábra (nevezzük őt 1. ábrának).

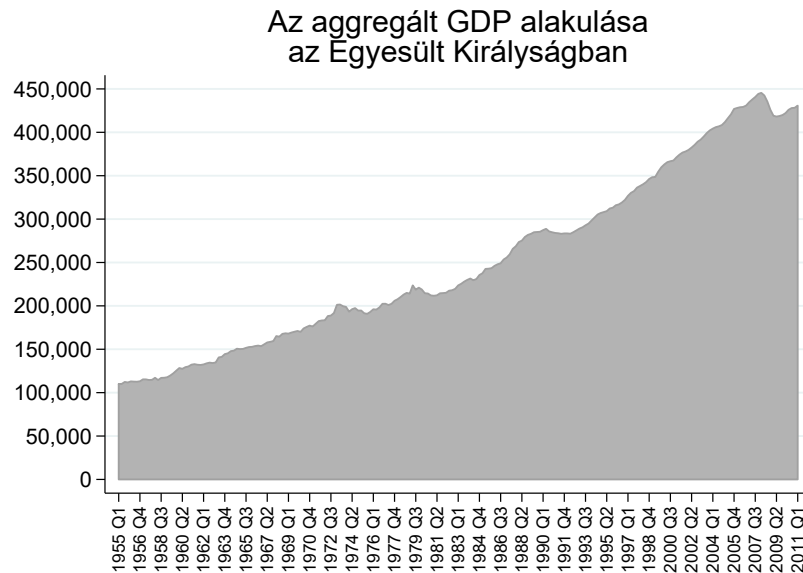


1. ábra: Az eredeti ábra

Egy ábrával kapcsolatos legfontosabb szabály (abból a százezzerből, amit figyelembe kell vennünk), hogy azt az adott szövegkörnyezetből kiemelve, önállóan is tudnunk kell azt értelmezni. Sajnos már ez az elvárás sem teljesül! Mi van ezen az ábrán? Ha az olvasó tudja, hogy melyik könyvből van az ábra, akkor sejtheti, hogy ez bizony GDP-t ábrázol, de többet nem. Melyik ország? Milyen típusú

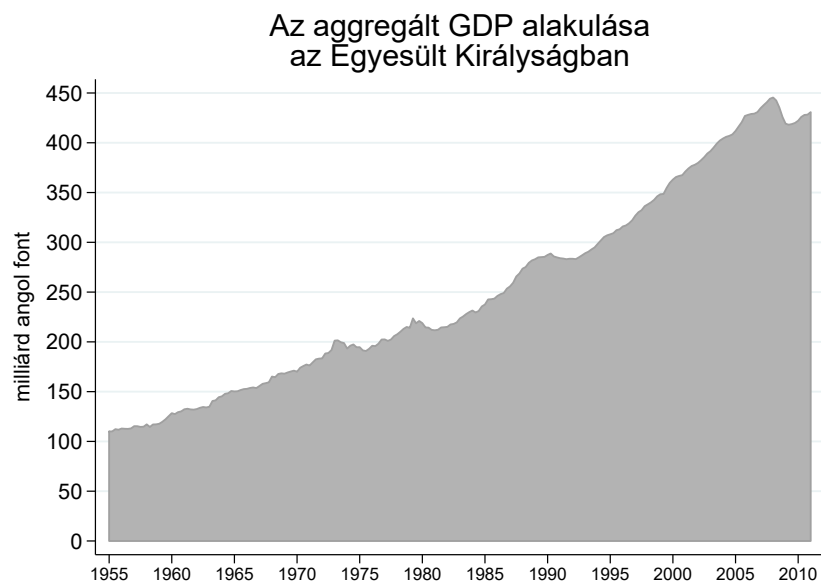
GDP? Mi az ábra üzenete?

Némi nyomozás után az ember rájöhet, hogy ez bizony az Egyesült Királyság GDP-je! Egészen pontosan aggregált GDP-je, azaz azt mutatja meg, hogy az országban összesen mennyi terméket és szolgáltatást állítanak elő adott időszakban, nem pedig azt, hogy ebből egy emberre átlagosan mennyi jut (utóbbit nevezzük egy főre eső GDP-nek). Az ábra tehát aggregált GDP adatokat mutat és az Egyesült Királyságra vonatkozik. Nem pedig egy főre eső magyar GDP-t ábrázoltunk. Ez az információ nagyon fontos! Legyen hát ez a cím! Egy ábra címétől azt várjuk, hogy rövid és tömör legyen, és (legalább nagyvonalakban – erre majd még visszatérünk) leírja, hogy mit látunk az adott ábrán (2. ábra)!



2. ábra: Az eredeti ábra, címmel bővítve

Az ábrára tekintve most már legalább azt tudjuk, milyen mutatót ábrázoltunk és melyik országra vonatkozik, azonban annak mértékegységét és azt, hogy minek a függvényében ábrázoltuk, még mindig nem tudjuk. Mi van a függőleges tengelyen? Milliárd forint? Milliárd euró? Angol font? Amerikai dollár? Mi van a vízszintes tengelyen? Sejthetjük, hogy negyedévek (a Q betű az angol quarter szó jelölése), de miért úgy, ahogy? Egyszer első negyedév (Q1), máskor negyedik (Q4), megint máskor pedig harmadik, vagy második. Tegyük tehát rendbe a tengelyeket! Az eredményt a 3. ábra mutatja.



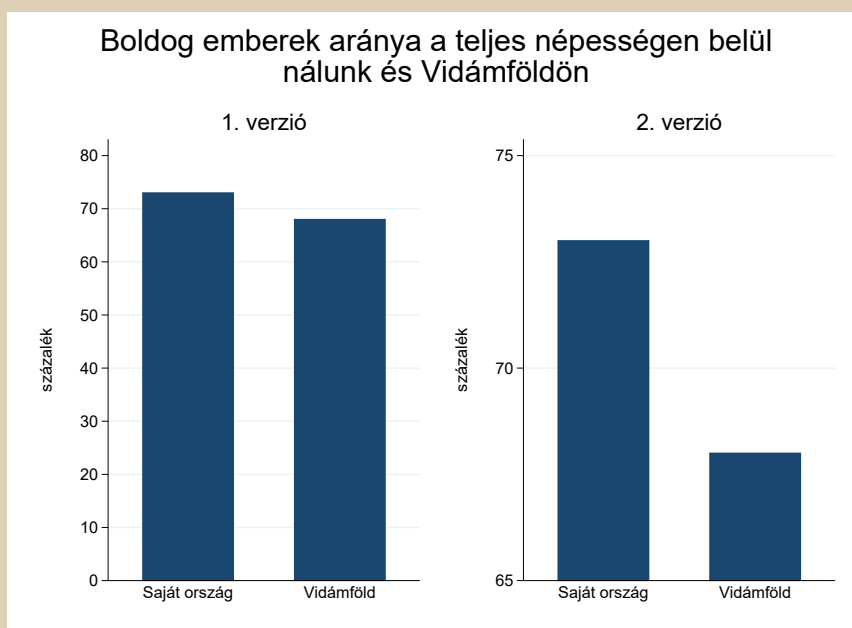
3. ábra: Az eredeti ábra, címmel bővítve és a tengelyek rendbetétele után

A vízszintes tengely az átalakítás után sokkal áttekinthetőbb, csupán minden ötödik év szerepel és ide tengelycímet sem szükséges kiírunk, adja magát, hogy évekről van szó. Esetleg, ha biztosra akarunk menni, akkor betehetünk egy alcímet, amelyben megjelenítjük, hogy „1955 és 2011 között”.

A függőleges tengelyen pedig a következő módosításokat végeztük. Egyrészt, mindent átszámoltunk milliárd fontra (az eredeti millió fontról), így néhány nullát megspórolunk, és könnyebben olvashatóak lesznek az értékek. Másrészt pedig tengelycímként felírtuk, hogy az adatokat milliárd angol fontban mérjük. Így már ránézésre tudjuk, hogy mi a mértékegység és eszünkbe sem jut az ábránkat közvetlenül egy magyar forintban mért GDP időssorral, vagy éppen egy Németországról készült, euró-adatokat megmutató hasonló ábrával összevetni.

Intermezzo: Tengelyek szerepe az ábrákon

Tegyük egy rövid kitérőt a tengelyekkel kapcsolatban! A tengelyek kezdőértékének jó megválasztása egy fontos lépése az ábrázolásnak. Különböző kezdőértékek egészen más jelenséget sugallhatnak, ábrakészítésnél mindig járjunk el etikusan (ne „csaljunk” annak érdekében, hogy a valóságostól eltérő, ám számunkra kedvezőbb eredményeket mutassunk)! Rövid illusztrációként tekintsük az alábbi kitalált történetet példaként! Tegyük fel, hogy született egy nemzetközi felmérés, amiben egy nemzetek feletti szervezet azt mérte, hogy az adott ország teljes népességének mekkora százaléka él boldogságban. A mi országunkat és régi nagy politikai ellenfelünket, Vidámföldét is felmérték, és a mi számaink kedvezőbbek: nálunk a népesség 73%-a boldog, Vidámföldén pedig csak a 68%-a. Országunk kormányzata természetesen nagyon büszke az eredményre és szeretné a polgárok felé is kommunikálni ezt, ezért szórólapokat készít, amelyen a 68%-ot és a 73%-ot ábrázolja. A kommunikációs tanácsadók kétféle szórólapverziót javasolnak, amelyeket az alábbi ábra mutat. Az 1. verziót egy szakmailag korrektül eljáró statisztikus javasolja, míg a 2. verziót egy kevésbé korrekt szakember. Mind a két ábra ugyanazt az eredményt mutatja be (68% vs. 73%), de vegyük észre, hogy a felületes vizsgáló számára mennyire más üzenetet közvetít! Az 1. verzió azt mutatja, hogy az országaink polgárainak kicsit nagyobb része boldog, mint Vidámföldén. A 2. verzió pedig azt sugallja, hogy a mi országunkban sokkal, de sokkal nagyobb a boldogok aránya. Mindig ügyeljünk arra, hogy az ábrákat ne felületesen szemléljük, ne hagyjuk magunkat félrevezetni! Ha pedig mi kerülünk olyan helyzetbe, hogy ábrát kell készítenünk, járjunk el etikusan! (Az etikátlan ábrázolásnál már csak az az elítélhetőbb, ha közvetlenül az adatokat hamisítjuk meg.)



Visszatérve a GDP-s ábránkhöz, a tengelyek most már rendben vannak, menjünk tovább, tekintsük a következő 4. ábrát! Az ábrázolt adatok nem saját gyűjtés, nem saját számítás, egyszerűen egy adatszolgáltatótól, jelen esetben a brit statisztikai hivataltól, az ONS-től kölcsönöztük őket. Mivel

nem saját adatról van szó, fel kell tüntetni a forrást (lehetőleg minél pontosabban, hogy ha valaki ellenőrizni szeretné a használt adatokat, azt könnyen megtehesse) – ezt mutatja a 4. ábra. Ha saját adatokkal dolgozunk, akkor is jeleznünk kell azt, például oly módon, hogy forrásként azt írjuk: „Az adatok a szerző saját számításai”, vagy „Az adatok a szerző saját gyűjtése.”!



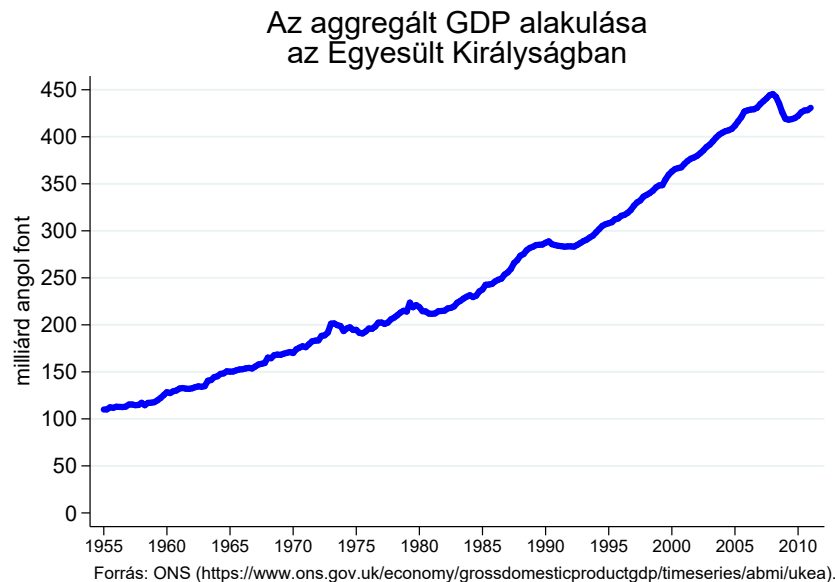
4. ábra: Az eredeti ábra, címmel bővítve, a tengelyek rendbetétele után, forrásmegjelölés hozzáadásával

Szintén egy nagyon fontos – az egyik legfontosabb – mozzanat az, hogy az adatainkat milyen formában jelenítjük meg. Számos ábrafélét ismerünk: oszlop, kör, pont, terület, vonal stb., illetve ezeknek a millió féle változatait. A típus kiválasztása nem önkényes, különböző szabályokat, szakmai konszenzusokat kell figyelembe venni: nem mindegy, hogy milyen típusú adathoz milyen típusú diagramot használunk. Néhány példa (inkább amolyan általános irányelvek, mintsem mindenkor alkalmazandó törvények):

- Valami időbeli alakulását jellemzően vonaldiagramon mutatjuk meg. Például, égetett szeszesitalok kiskereskedelmi forgalma az elmúlt 20 évben.
- Rész-egész viszonyt a legegyszerűbb kördiagram segítségével megmutatni. Például, az osztályodban mennyi fiú és mennyi lány diák van.
- Két változó egymáshoz fűződő viszonyát ábrázolhatjuk pontdiagram segítségével. Például, milyen kapcsolat van a pénzállomány változása és az infláció között.
- Esetleg, kombinálva két előző pontot: szeretnénk megmutatni a rész-egész viszony időbeli alakulását? Használjunk terület-diagramot! Például, hogyan alakult a magyar GDP felhasználása az elmúlt 15 évben.

Az általunk használt terület-diagram azt sugallja, hogy valamiféle dolgot felhalmozunk és az idő múlásával egyre több és több lesz belőle. Ha ismerjük a GDP fogalmát, akkor tudhatjuk, hogy az nem igazán egy ilyen jelenséget ír le, azt mutatja meg, hogy adott időszakban mennyi terméket és szolgáltatást hozunk létre. (Ha valaki szeretné ezt a problémát megérteni, akkor olvasson a folyamat- és állományjellegű, angolul flow és stock változók közötti különbségekről!) Sokkal jobb lenne (sőt, erről kicsit többről, szakmai konszenzusról van szó) a GDP alakulását egy más típusú ábra, például vonaldiagram segítségével megjeleníteni (és ha már így járunk el, kicsit tegyük is színesebbé a végeredményt)! Ezt mutatja az 5. ábra.

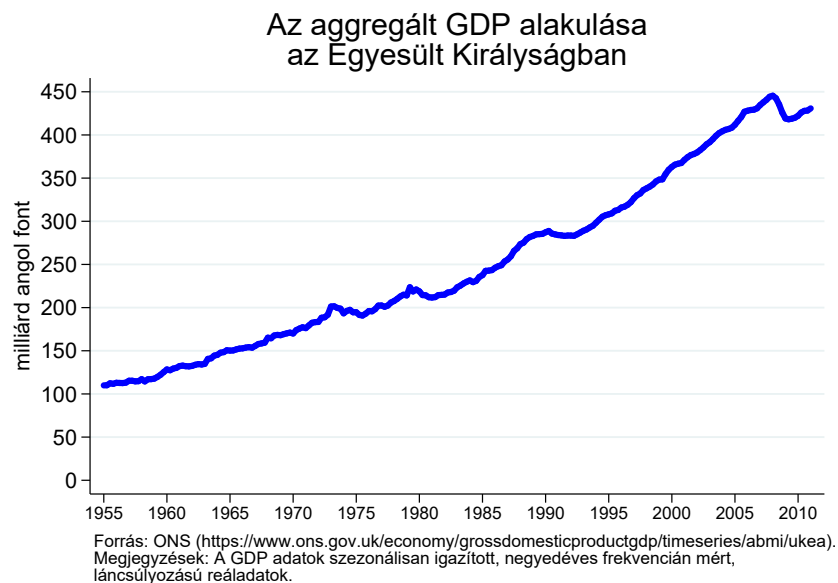




5. ábra: Az eredeti ábra, címmel bővítve, a tengelyek rendbetétele után, forrásmegjelölés hozzáadásával, a tartalomhoz jobban illeszkedő ábratípussal

Végül pedig, a szakmai igényesség jegyében, minden olyan egyéb tartalmi, módszertani információt, amelyet relevánsnak tartunk, érdemes az ábra alatt, megjegyzésként elhelyezni. A végeredményt mutatja a 6. ábra.

- Tekintsük tehát a végeredményünket! Anélkül, hogy szöveggörnyezetet adnánk az ábránknak, az alábbi dolgokat tudjuk meg róla:
- Milyen adatot ábrázol milyen dimenzióban? (A brit aggregált GDP időbeli alakulását ragadja meg.)
- Milyen mértékegységben mérünk? (Milliárd angol fontban.)
- Honnan származik az ábrázolt adat? (Az ONS-től.)
- Az ábrázolt GDP pontosan milyen GDP? (Aggregált, negyedéves frekvenciájú, szezonálisan tisztított és inflációszűrt.)



6. ábra: Végeredmény: az eredeti ábra, címmel bővítve, a tengelyek rendbetétele után, forrásmegjelölés hozzáadásával, a tartalomhoz jobban illeszkedő ábratípussal és módszertani megjegyzésekkel ellátva

Az ábránkat pedig tovább bővíthetnénk, alakíthatnánk annak függvényében, hogy mit szeretnénk vele megmutatni, mi az üzenetünk (azon túl, hogy egy GDP-vel foglalkozó könyv elején mi más is lehetne, mint GDP). A GDP-vel kapcsolatban a makroökonómusok két fontos jelenséget szoktak vizsgálni:

1. Miért növekszik az hosszú távon (miért van az, hogy ma nagyobb a GDP, mint 10-20-30-40 éve volt)?
2. Miért vannak ingadozások a GDP alakulásában (miért váltják egymást rosszabbul és jobban teljesítő időszakok)?

Hasznos gyakorlatként gondolkozzatok el azon, hogy ezen kérdéseket (üzeneteket) hogyan jeleníthetnénk meg az elkészült ábránkon!

Végül, néhány további írásos anyag és honlap, amelyek segíthetnek az első fordulóra leadott pályázat elkészítésében:

- Hunyadi László (2002): Grafikus ábrázolás a statisztikában. Statisztikai Szemle 80 (1): 22-52.
- Schwabish, JA (2014): An Economist's Guide to Visualizing Data. Journal of Economic Perspectives 28 (1): 209-234.
- A www.gapminder.org honlapon belül a „Gapminder Tools” menüpont alatt nagyon sok, érdekes adatot jeleníthetünk meg nagyon sokféleképp.

Megjegyzés

Van néhány kifejezés, amit a szövegben használtunk, de mivel nem a hétköznapi kifejezéstár része, nem biztos, hogy tudod, mit takarnak. És ugye pont azért veszel részt a játékban, mert tanulni szeretnél, így elmagyarázzuk.

A főszövegben használt ábra a megtermelt bruttó hazai termék (ez a GDP) volumenének alakulását mutatja az Egyesült Királyságban. Miután az aggregált GDP mennyiségi alakulásának alapvető tendenciáit szeretnék érzékeltetni, ezért:

- az infláció hatását kiszűrik, un. változatlan áron veszik számba a GDP nagyságát, erre utal a **reáladat** kifejezés.
- A változatlan árat időről időre revideálják, mert az árárányok is változnak. A változatlan árak időszakos cseréje a **változó súlyozás**, vagy **láncolás**.
- Nem a negyedéves tényleges mennyiségi adatokat mutatják be, mert éven belül a gazdasági idősorok közül számos tartalmaz **szezonalitást**, pl. a GDP általában a IV. negyedévben kiugróan magas az évvégi hajrá következtében. A szezonális hatását elsimítják, erre alkalmas statisztikai módszerekkel, erre utal a **szezonális kiigazítás**.

VERSENYFELADAT

Most pedig nézzük, milyen feladatot kell nektek megoldani! A versenyfeladat nem egy tipikus közgazdaságtani jelenséghez kapcsolódik, ám egy olyan terület, amely elemezhető a közgazdaságtan eszközeivel, és van makrogazdasági relevanciája. Ez a téma pedig nem más, mint a bűnözés. A feladatok megoldásához két adatbázist kaptok. A feladathoz mellékelt Excel fájl **1. adatbázis** nevű fülén néhány amerikai állam bűnözési rátája (százezer főre jutó bejelentett bűnesetek száma) található az 1960 és 2014 közötti időszakra. Az adatbázisban DC adatai, valamint az amerikai átlag (USA Total) is elérhető. A másik adatbázis (**2. adatbázis** nevű fül) a kiválasztott amerikai államok jogvégrehajtó tisztviselőinek (nevezzük most őket rendőrnek, bár annál tágabb kategóriát fed le, pl. a kisegítő személyzetet is tartalmazza) számát és az adott állam népességét tartalmazza a 2014-es évre! Ezen két változóból számoljatok százezer főre jutó rendőrszámot és az elemzés során azt használjátok! Miért fontos ez a transzformáció? Írjátok le néhány mondatban az ábrák bemutatása előtt! Sőt, még jobb, ha ábra segítségével is bemutatjátok, hogy miért fontos!

- Hogyan alakult a vizsgált időszakban (1960-2014) az átlagos bűnözési ráta az Amerikai Egyesült Államokban?
- Hogyan viszonyulnak a kiválasztott államok rátái az átlagos rátához? Együtt mozognak azzal? Közel vannak hozzá? Vannak jól elkülöníthető államcsoportok? Van-e bármilyen érdekes, meglepő, az átlagtól nagyon különböző jelenség?
- Milyen kapcsolat van a rendőrszám és a bűnözési ráta között? Van-e olyan állam, amely nagyon kilóg a sorból? Ha van, és nélküle hasonlítjuk össze a két változó közötti kapcsolatot, akkor nagyon más eredményre jutunk? (Később majd foglalkozunk a témával részletesen, de már most megéri elgondolkodni: ha van kapcsolat, az oksági kapcsolatot is jelent egyben?)

Válaszolatok a fenti kérdésekre szövegesen is, illetve ábra/ábrák segítségével is! Az ábrát/ábrákat úgy készítsétek el, hogy a lehető legkevesebb legyen belőlük: a feltett kérdésekre a lehető legkevesebb számú ábra adjon választ, azaz próbáljátok az üzeneteket kisszámú ábrába tömöríteni! A leadandó munka egy 1-3 oldalas, írásos, word dokumentum (12-es TNR betűk, másfeles sortáv, normál margók) plusz az ábra/ábrák, amelyet/amelyeket bármilyen szoftver felhasználásával elkészíthettek. Az 1-3 oldalas dokumentumban egyrészt válaszolatok a feltett kérdésekre, másrészt pedig azt is írjátok le, hogy az adott kérdésre a választ az ábrákon/ábráitokon mely elem adja meg. Ha jól van elkészítve az ábra, akkor ránézésre is tudnunk kell, hogy mit miért csináltatok, melyik elem/szín/elrendezés milyen üzenetet közvetít, de most kivételesen írjátok is azt le!

A feladathoz használjátok a mellékelt excel fájlt: [facsarj_adat1.xls](#)

HÁT AKKOR FACSARJ EGY KIS INFORMÁCIÓT AZ ADATOKBÓL!!!

HATÁRIDŐ: 2018. DECEMBER 22.

Ha további kérdéseid vannak, lépj velünk kapcsolatba a facebookon: BCEdata, vagy kkar.corvinus oldalakon, de akár emailt is írhat a data@uni-corvinus.hu címre.

FACSARJ

 EGY KIS INFORMÁCIÓT AZ ADATOKBÓL!