

# Dinamikus geometriai szoftverrel megvalósított antropomorf adatvizualizáció és alkalmazása

Bölcskei Attila<sup>1</sup>, Budai László<sup>2</sup>, Keresztes Éva Réka<sup>2</sup>, Talata István<sup>1</sup>

<sup>1</sup>*füiskolai tanár, <sup>2</sup>egyetemi docens*

<sup>1, 2</sup>Budapesti Gazdasági Egyetem, Külkereskedelmi Kar,  
Társadalomtudományi Módszertan Tanszék

E-mail: bolcskei.attila@uni-bge.hu, budai.laszlo@uni-bge.hu,  
keresztes.eva@uni-bge.hu, talata.istvan@uni-bge.hu

DOI: [10.29180/978-615-6342-90-4\\_3](https://doi.org/10.29180/978-615-6342-90-4_3)

**Összefoglalás:** Az adatvizualizáció egyik nagy problémája összetartozó, nagy mennyiségű adat egyszerre való megjelenítése. A tanulmányban az alkalmas eljárások egyikét, a közelmúltban 50 éves Chernoff-arcokat mutatjuk be, új megközelítésben. A módszer alapötlete az, hogy a multidimenzionális minta változói, mint paraméterek egy képregény figura arcának különböző részeit vezérlik, lehetővé téve, hogy a felhasználó a rajzok segítségével például összefüggéseket fedezzen fel az egyes adatsorok között. Bemutatjuk a Chernoff-arcok új, GeoGebra dinamikus geometriai szoftverrel készített implementációját, amely segítségével az arcot 21 különböző paraméter egyidejű megjelenítésére tettük alkalmassá. Az élvezetes vizualizációt Budapestet és a 19 vármegyét érintő, komplex KSH adatsorokon teszteltük, melyek eredményét szintén bemutatjuk.

**Kulcsszavak:** adatvizualizáció, multidimenzionális minták, statisztikai elemzés, Chernoff-arcok, regionális gazdaságfejlesztés

**Abstract:** One of the big problems with data visualisation is to present large amounts of data at the same time. In this paper a new approach of one of the suitable methods will be presented. These are the so-called Chernoff faces, that recently turned to 50 years old. The basic idea of the method is that the variables of the multidimensional pattern as parameters control different parts of a cartoon character's face, allowing the user to discover, for example, relationships between the individual data sets based on the drawings. We present a new implementation of the Chernoff faces using GeoGebra dynamic geometry software, which has been used to make the face capable of simultaneously representing 21 different parameters. We have tested the vivid visualization on complex HCSO datasets for Budapest and the 19 counties, the results of which are also presented.

**Keywords:** data visualisation, multidimensional samples, statistical analysis, Chernoff faces, regional economic development

## 1. Multidimenzionális adatvizualizáció: cél és eszközök

Érzékszerveink közül a látás szerve kiemelt jelentőségű, ugyanis az emberek a környezetből származó információknak 60, de akár 80%-át is látás útján gyűjtik be. Az emberek döntő többsége így tehát a világot elsősorban vizuálisan közelíti meg és értelmezi. A vizualitás jelentősége napjainkban még

inkább felértékelődik, hiszen lépten nyomon tapasztalható, hogy az információk átadásában a szövegeket különféle vizuális tartalmak (videók, képek, ikonok, infografikák, stb.) váltják fel.

Ebben a dolgozatban a vizualizáció tág fogalomkörét az információvizualizációra szűkítjük le. Ez azt jelenti, hogy azzal a céllal láttatunk, hogy elsősorban információt közöljünk vele, és ne csak szórakoztassunk általa. Az információ gazdasági, üzleti, műszaki, stb. célú vizualizációja azt követeli meg, hogy az adatokat rendszerben ábrázoljuk.

Az adatvizualizáció egy elterjedt megközelítés [1] szerint nem más, mint leképezés az adat és annak vizuális megjelenítése között, pontosabban leképezés az agy két modalitása, nevezetesen a matematikai és a vizuális között. Manovich szerint az adatvizualizáció célja az, hogy nagy mennyiségű adatot mások számára áttekinthetővé tegyünk. Ennek nagy szerepe van a statisztikusok munkájában, amennyiben az adatgyűjtés után, az adatfeldolgozási fázisban az adatokat áttekinthetővé téve mintázatokat, összefüggéseket mutathat a változók között, s ennek nyomán választható ki az adekvát matematikai-statisztikai módszer.

Míg az elméleti matematikusok általában ódzkodnak a szemléletes magyarázatoktól és elvont megoldásra, absztrakt következtetési lánc felállítására törekednek, addig az alkalmazott matematika és a matematika oktatása mindig is felhasználta a szemléltetést. A számítógépek megjelenése előtt még a praktikus (mérnöki, gazdasági) számítások elvégzéséhez is alkalmaztak ábrákat, pl. nomogramokat. Ezek szerkesztése külön tanulmányokat követelt, míg az általuk elérhető pontosság csekély volt. Ezzel együtt a mai napig a szemléltető (nem bizonyító) ábrázolás szerepe jelentős maradt, a matematika oktatásban elég akár a függvényábrázolásra, míg a tudományokban a statisztikai adatsorok különféle grafikonokon való megjelenítésére utalni.

Összefoglalóan végül is azt mondhatjuk, hogy az adatvizualizáció olyan módszer, amely számok (adatok) közötti kapcsolatokat mutat be vizuálisan, abból a célból, hogy például folyamatok trendjeit illusztráljuk, mintázatokat ismerjünk fel, megtaláljunk kiugró értékeket, stb.

Minderre azért is van növekvő igény, mivel az adatok óriási és egyre fokozódó ütemben termelődnek (Big Data jelenség). Becslések szerint 2025-ben nagyjából egy hét alatt annyi adat keletkezik majd, mint 2013-ban összesen (4,3 zettabájt), holott már ez is több mint 850-szerese volt a valaha írt összes könyvben tárolt adatmennyiségnek (5 exabájt) [2].

Az adatvizualizáció fontos eszköze az adatbázisokban történő tudásfeltárással (KDD = Knowledge discovery in databases) kapcsolatos feladatoknak. Manapság könnyen gyűjthetünk és tárolhatunk nagy mennyiségű adatot, de mindig kérdés, hogy abból milyen információkat tudunk kinyerni? A KDD mindig felhasználta az adatvizualizációban rejlő lehetőségeket. Azonban annak eldöntése, hogy egy adott adathalmazt hogyan ábrázoljunk a lehető leginformatívabb módon, egy intuitív feladat marad [3].

A vizuális megjelenítés előnyeit az alábbiakban foglalhatjuk össze: segíti a főbb következtetések megjegyzését; javítja az elemző azon képességét, hogy kapcsolatokat vegyen észre; a jelenségek könnyebben bemutathatók, elmesélhetőek, ha ábrázolva vannak; illetve végezetül lehetővé teszi bonyolult számítások áttekintését.

A [4] publikáció által leírt taxonómia szerint az adatvizualizációs technikák osztályai főbb vonalakban az alábbiak: geometrikus módszerek, ikonos technikák, pixelorientált technikák, hierarchikus, grafikus illetve hibrid módszerek.

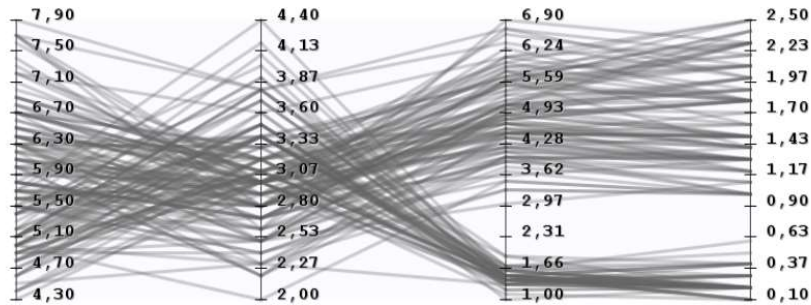
Szokás ezenkívül megkülönböztetni standard és nem-standard ábrázolási módszereket; előbbieken értve a leggyakrabban és legrégebben használt eljárásokat: a kör-, és oszlopdiagrammot, a hisztogramot, az empirikus eloszlásfüggvényt és gyakorisági poligont. Nem-standard ábrázolási eljárásokkal több változót is figyelembe vevő, vagy a változásokat jobban kiemelő vizualizációkra is lehetőség nyílik.

A következőekben, az értelmezési keretet szűkítve, arra a nem-standard problémára keressük a választ, hogy miként lehet a  $k$ -dimenziós tér pontjait úgy megjeleníteni, hogy a változók közötti, feltételezett komplex összefüggések is megjelenjenek. A vizualizáció céljaul azt tűzzük ki, hogy a valószínűsíthető kapcsolatokat feltárjuk, megsejtsük melyeket azután a megfelelő statisztikai módszerekkel igazolni is lehet. A  $k$ -dimenziós vektorok, amelyekkel dolgozunk olyan mintáknak felelnek meg, amelyek egy adott alany (jelenség, időpont, stb.)  $k$  db különböző ismérv szerinti értékeit tükrözik. Például autógyárak különböző típusú termékeinek nagy számú műszaki és kereskedelmi paramétereit tekinthetjük a különböző ismérveknek. Így minden autó típushoz egy nagy adatsor tartozhat, ezek lennének az ábrázolandó  $k$ -dimenziós vektorok. A gyakorlat rengeteg egyéb példát szolgáltat ilyen típusú adatfelvételekre.

A fenti felosztás szerinti geometrikus illetve ikonos megoldásokra hozunk néhány példát.

Egyszerű geometrikus módszer amikor  $k$  darab párhuzamos számegyenes jellemzi a  $k$  darab változót és az egyes adatfelvételekhez tartozó szám-k-ások egy-egy töröttvonalaként kerülnek ábrázolásra, melyek csúcspontjai éppen az adott változón felvett értékek (Parallel Coordinates Plot = PCP módszer). Érdekes ehhez a felvehető értékek tartományát minden változón 0 és 1 közé normálni, vagy legalábbis azonos hosszúságúvá nyújtani.

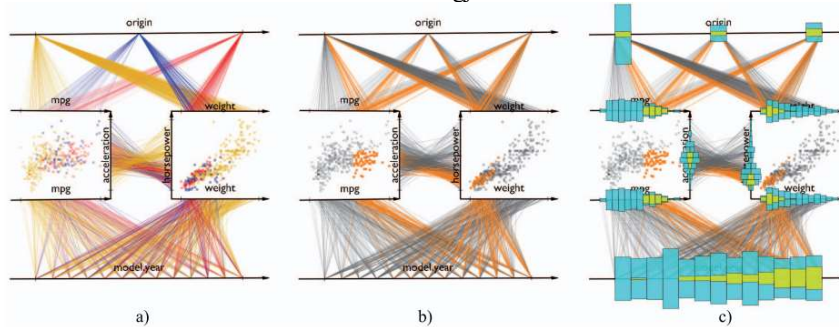
1. ábra: Példa 4-dimenziós vektorok töröttvonalaként való ábrázolására a PCP módszerrel



*Forrás: [3], Fig.2*

Ennek továbbfejlesztett változata az [5] publikációban leírt Flexible Linked Axes (FLA = Összekapcsolt flexibilis tengelyek) módszer, amelynek nevében a flexibilitás a tengelyek igen sokféle (részben párhuzamos, de akár metsző) elhelyezési módjára utal. A tengelyek mindegyike itt is egy ismérvet jelenít meg és egy tartományt ábrázol. A kapcsolatokat szintén szálakkal, (akár színes) törött vonalakkal lehet megjeleníteni a vásznon (képernyőn). Mindez végső fokon a radardiagram illetve a PCP közös általánosításának is tekinthető, melybe akár egyéb adatvizualizációk, mint pl. pontfelhő diagramok, hisztogramok, stb. is beépítésre kerülhetnek.

2. ábra: Példa FLA módszerrel megjelenített adatsorokra

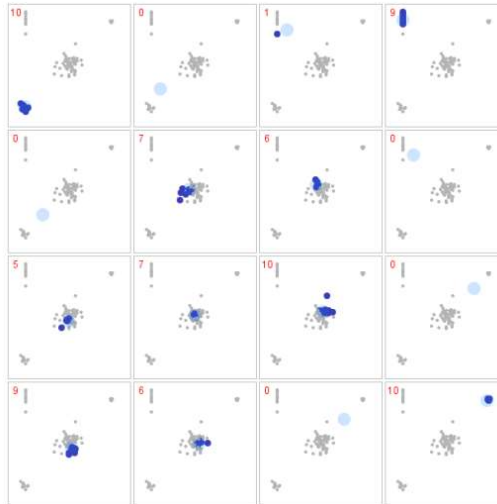


*Forrás: [5], Fig.4*

A [3] tanulmányban ikonalapú és geometriai alapú vizualizációs technikák kombinációját javasolják, amelyet egy önszervező térkép támogat (Self-

organizing Grid = SOM)). Ez a módszer magas-dimenziós adatok alacsonyabb dimenziós rácsba való leképezésével, majd az adatok topológiát és főbb metrikát megőrző összetömörítésével teszi lehetővé akár igen nagy adattömegek feldolgozását is.

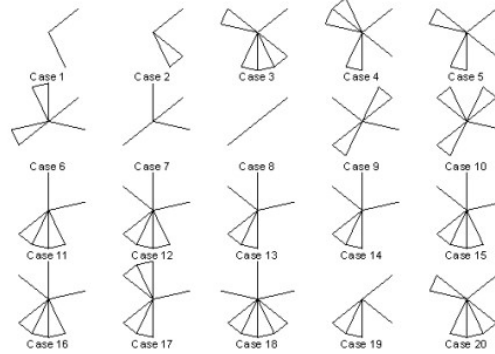
3. ábra: Példa SOM technika alkalmazására



Forrás: [3], Fig.1

Tisztán ikonos ábrázolási technika a glifák (glyph) felhasználásán alapul. Maga a fogalom igen sokrétű és a szakirodalomban meglehetősen kuszán használt [6]. Alapvetően olyan szimbólumot jelölünk vele, amely nem verbálisan közvetít információt (pl. útjelző tábla). Adatvizualizációban a glifa az egyes, adatokat leíró képi elemek összességét jelenti. Multi-dimenzionális ábrázolásban egyik leggyakrabban használt típusa az ún. "whisker plot" vagy bajusz ábrázolás (nem azonos a "box-and-whisker plot"-tal), ahol minden változót egy központi pontból kiinduló vonalszakasz ábrázol. A vonalszakasz hossza jelzi a megfelelő változó értékét. Ennek egyik változata, az ún. "csillag plot", ekkor a szomszédos vonalszakaszok végei össze is vannak kötve.

4. ábra: Példa bináris adatsoron képezett csillag plotra (glifára)



Forrás: [7], Fig.3

Ugyancsak a tisztán ikonos ábrázolásra érdekes példa lehet még a Korhonen által kidolgozott harmonikus házak módszere [8]. Ebben a szerző nem egyszerűen szemléltetni kíván nagy mennyiségű adatot, hanem döntési alternatívákat szeretne vizuálisan megjeleníteni ezen nagy mennyiségű feltétel esetére és a döntést igyekszik a szimmetria és harmónia segítségével megkönnyíteni. Ezt szembe állítja a később részletesen taglalt Chernoff-arcokkal, melyek a döntést nem segítik, mivel nem úgy vannak feltétlenül kialakítva, hogy a „vonzóbb” arc mutasson egy jobb döntést. A harmonikus házak egy ideális, szimmetrikus házból indulnak ki, mely meghatározó sarokpontjait a változók egy paraméter-dobozban áthelyezhetik, ezzel rontva el a harmóniát. A kiindulási ház komplexségét a változók számához lehet alakítani, amivel a paraméterek száma úgyszólván korlátlanra tehető. A módszer sikerességét jelzi, hogy adatsorokat így feldolgozva az alanyok 98%-ban meg tudták mondani, hogy mely vizualizációk mutatnak be olyan céget, amely csődbe ment. Ugyanez pusztán az adatok alapján nem sikerült a résztvevőknek [9].

Megjegyezzük még, hogy egy további lehetséges módszer lehet  $k$  változó hatásának egyetlen ábrán való ábrázolására az Andrews által [10] javasolt technika, amely Fourier-sor alakot használ a megjelenítésre, ahol a megfelelő együtthatókat a tetszőleges méretű  $k$ -dimenziós vektor  $x_1, x_2, x_3, \dots$  koordinátái adják az alábbiak szerint.

$$\mathbf{f}_x(t) = x_1/\sqrt{2} + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \dots; t \in (-\pi, \pi)$$

## 2. Részletesen a Chernoff-arcok módszeréről

Az egyik legszórakoztatóbb és legprovokatívabb próbálkozás arra, hogy nagy mennyiségű változó hatását egyszerre érzékeltessük, Hermann Chernoff-tól származik [11]. A hamarosan részletesen ismertetett eljárás lényege az, hogy a  $k$ -dimenziós vektor egy képregény figura arcának különböző sajátosságait (az arc formája; a szem, szemöldök, száj és orr formája, elhelyezkedése) vezérli. Ezáltal a jellegzetességek kiválóan szemléltethetővé válnak. A kiinduló feltevés az volt, hogy az emberi elme apró különbségeket is képes detektálni az emberi arcon, így lehetőség van finom különbségek észlelésére és arra is, hogy egy arcon a változást észrevegyük (ez például idősor elemzésnél lehet hasznos). Nem jelent gondot az emberi elme számára az, ha a rajzolt arc nem egészen realisztikus, hiszen a rajzfilmek, animációk figuráinak érzelmeit is könnyen át tudjuk élni. Az arcok ezen kívül könnyebben megjegyezhetők: az érzelmek, a jellegzetességek könnyebben idézhetők fel a későbbiekben, akár sok idő elteltével is.

Az arcokkal történő adatvizualizáció lehetővé teszi például, hogy klaszter analízist végezzünk, ilyenkor az összetartozó adatmintákat csoportosítjuk. Viszont az említett idősor elemzésnél, amikor az időben egymást követő  $k$ -dimenziós adatsorokat arcokként jelenítjük meg, az arc karakterének változásával azonosíthatók a folyamatban előforduló töréspontok. Könnyen lehet kiugró értékeket, mással kapcsolatba nem hozható adatsorokat is azonosítani: az arc karaktere ilyenkor minden másétől jelentősen el fog térni.

Ha a vizualizáció kínálta előnyöket nézzük, akkor biztosan állítható, hogy a Chernoff-arcok segítik a főbb következtetések megjegyzését és nagyban segítenek az elemzőnek abban, hogy kapcsolatokat azonosítson.

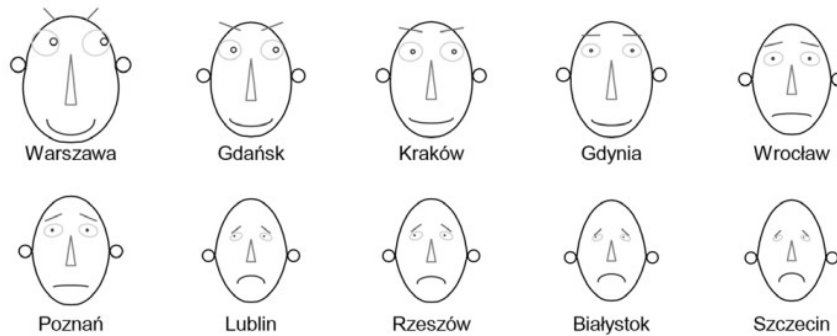
A módszer éppen a közelmúltban volt 50 éves, de érdekessége miatt még nem szorul ráncfelvarrásra; a mai napig aktuális és hasznos eszköz az adatelemzésben, erre utal a [11] publikáció majdnem 2400 hivatkozása is (2024. januári adat, forrás <https://scholar.google.com>). Alkalmazására hozunk fel néhány aktuális, illusztrált példát a gazdasági, üzleti tudományok és oktatás vonatkozásában.

A [12] dolgozat például a Chernoff-arcokat borok brendjével kapcsolatos social media elemzésére használja fel. A közösségi médiában való megjelenési adatokat előbb Social Mention eszköz segítségével összegyűjtötték, feldolgozták, és az elemzést annak Chernoff-arcokkal való bemutatásával kötötték össze. Végeredményben a cikkben hat bormárkát hasonlítottak össze

hét ismérvet véve figyelembe. A következtetések levonását az ideális (100%-os) arccal való összehasonlítás segítette.

Egy 2022-es eredményt mutat be a [13] tanulmány. Ebben Lengyelország tizenhét városa vonatkozásában a lakások többdimenziós összehasonlító elemzését mutatják be a szerzők 2017 első negyedétől 2021 első negyedévéig, a gazdasági biztonság fenntartása szempontjából. A kutatás többdimenziós összehasonlító elemzéseket használt klaszter felbontás céljából, melyhez felhasználták a Chernoff-arcok módszerét is. Ez utóbbi lehetővé tette, hogy idősorosan vizsgálják a lengyelországi egyes városok 1 m<sup>2</sup> -es lakóingatlan árainak hasonlóságait és különbségeit. A tekintett paraméter az ingatlanok négyzetméter ára volt, mégpedig 2017 első negyedétől kezdődően, negyedévente egészen 2021 első negyedévéig, ami összesen 17 paramétert jelentett.

5. ábra: Példa Chernoff-arc ábrázolásra ingatlanpiaci elemzés céljából

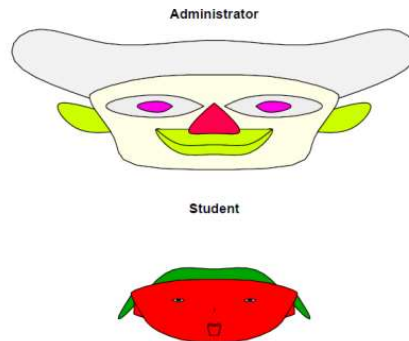


*Forrás: [13], Fig.5, részlet*

Ugyancsak 2022-es publikáció a [14]. Ebben a szerzők Chernoff-arcok segítségével modellezik egy manilai (Fülöp-szigeteki) tanárképző intézmény hallgatóinak, oktatóinak és adminisztratív személyzetének (összesen több, mint 300 alany) válaszait az eredményalapú tanárképzési tanterv (OBTEC) trimeszter rendszerű bevezetésére. Egy felmérés során a résztvevők 15 kérdést kaptak, melyre 1-6 skálán adtak választ. Ezzel a 15 paraméterrel készültek el a vizualizációk, melyekhez az R nyelv statisztikai programcsomagját használták. Az arcokat a szerzők nagyon hasznos segítségnek tartják a begyűjtött vélemények elemzéséhez és a problémák azonosításához.



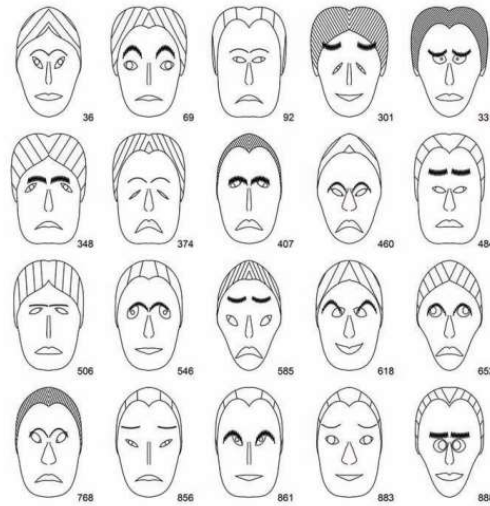
6. ábra: 100%-os és 0%-os arcok a [14] publikációból



*Forrás: [14], Fig.1*

Amint látható, a különböző publikációkban feltűnő arcok nagy különbözőséget mutatnak, a felhasznált szoftvertől és módszertől függően. A felhasználóknak nagy segítséget nyújthat az ábrázolásban Raciborski [15] fejlesztése. Az újragondolt Chernoff-arcokat, akár 51 változó megjelenítésére teszi alkalmassá és implementációját a Stata statisztikai-ökonometriai programcsomagba ágyazza be, mely a chernoff parancs segítségével hívható meg. Az arcok itt a legrealisztikusabbak, a haj formája és a száj, illetve szemöldök vastagsága is állítható, de füleket nem találunk.

7. ábra: Sokrétűen variálhatóak az arcok a Stata programcsomaggal

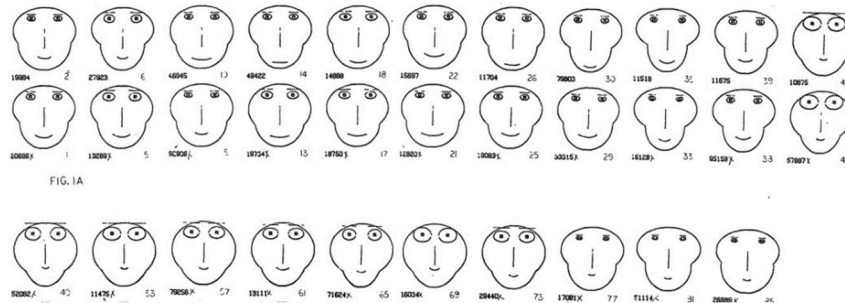


*Forrás: [15], Fig.1*

Megjegyezzük még, hogy Chernoff módszere akár hegesztés minőségellenőrzésében [16], vagy állattenyésztési adatok feldolgozásában [17] is sikerrel alkalmazható.

Hermann Chernoff eredeti munkájában [11] 18 olyan paramétert vezetett be, amelyek segítségével az arc szerkezete változtatható. Ez azt is jelenti, hogy minden általa előállított arc megfeleltethető egy 18 dimenziós folytonos eseménytér egy mintájának. Chernoff paraméterei a fenti cikk függelékéből olvashatók is. Ezek alapján megállapítható, hogy mely paraméter arc mely területét vezérli. Leszögezzük, hogy Chernoff (és a későbbiekben mi is ezt fogjuk tenni) csakis tengelyesen szimmetrikus arcokat készített, azaz az arc bal és jobb oldala egymás tükörképe. Emiatt a paraméterek zöme csak az arc egyik felének geometriáját határozza meg, a másik oldal tükrözéssel adódik. Az első 5 paraméter nála az arc formáját befolyásolja a következő módon: az arcot két egymást metsző ellipszisből építi fel, majd ezek burkát tekinti kontúrként. A két ellipszisével így folytonosan, de nem érintőfolytonosan csatlakozik egymáshoz. Egy középpontot, mint origót rögzítve, a (mondjuk jobb oldali) csatlakozási pont, ahol a két ív metszi egymást két (polárkoordinátákkal leírt) paramétert kap. Az arc magassága képezi a harmadik paramétert, míg a 4. és 5. paraméter a felső illetve alsó ellipszis excentricitása. A következőekben az arc egyes részeit paraméterezi fel a következőképpen. Legyen az orr egy origó középpontú függőleges vonal. Ennek hossza a 6. paraméter. A körív alakú száj kialakítása 3 paramétert kap, úgymint a száj középpontjának y koordinátája (7. paraméter), a száj görbülete (a kör sugarát jellemző adat, 8. paraméter) és a száj ívének hossza (9. par.). A szem jelentősége nyilván kiemelt, ezért annak formája 5 paramétert érdemel, míg a pupilla helyzete további egy paraméter az eredeti munkában. Részletezve: a szemek legyenek kis ellipszisek. Középpontjuk leírásához 2 paraméterre van szükség; azután a 12. paraméter az ellipszis nagyengelyének ferdeségét írja le, míg a 13. az ellipszis excentricitását, a 14. paraméter pedig nagyengelyének hosszát határozza meg. A pupilla (mely egy kis, állandó sugarú körként jelenik meg az ábrákon) helyzetét írja le a 15. paraméter. Chernoff az arcon már csak a szemöldököt jeleníti meg, melyhez 3 paramétert rendel hozzá. A szemöldököt mindig pontosan a szem fölé helyezi, így 16. paraméterként a szemtől való távolságot tekinti, majd utolsó előttinek a ferdeséget, utolsó paraméternek pedig a szemöldök hosszát adja meg. Példaként bemutatjuk a [11] cikk egyik adatvizualizációját, az első Chernoff-arcokat:

8. ábra: Példa az eredeti Chernoff-arcokra



Forrás: [11], 362. oldal (részletek az ottani 1/A és 1/B ábrákból)

Mint fentebb említettük, a Chernoff-arcok és paraméterezésük az idők során variálódott, például az arc újabb elemeinek megkonstruálásával.

Feladatunknak éreztük, hogy a dinamikus geometriai szoftverek adta lehetőségeket kihasználva újra gondoljuk a Chernoff-arcok felépítését és lehetővé tegyük, hogy a felhasználó egyrészt a csúszkákat változtatva könnyen beállíthasson adatvizualizációkat az adatsoraihoz, másrészt magukat a változtatásokat valós időben nyomon tudja követni, a látványban örömet lelje. Ilyen irányú törekvés ismereteink szerint ezidáig nem történt a témában.

### 3. Chernoff-arcok konstruálása dinamikus geometriai szoftver segítségével

Az alábbiakban ismertetjük azt a paraméterezést, amelyet mi választottunk az arcok megszerkesztéséhez. Alapelvként azt mondhatjuk, hogy célként egy, az eredetinel realisztikusabb, részletesebb, nagyobb felhasználói élményt nyújtó, ugyanakkor a képregények báját megőrző konstrukció kidolgozását fogalmaztuk meg. Eszközként a szabadon elérhető GeoGebra dinamikus geometriai szoftvert használtuk [18], mely számos geometriai probléma szemléltetésében már bizonyított [19, 22].

Megoldásunkban az arc formáját szintén tükörszimmetrikusnak tekintettük, melyet azonban ellipszisek helyett, a nagyobb fokú alakíthatóságot lehetővé tévő köbös spline görbékkel valósítottunk meg. Ehhez 7 paramétert vezetünk be az alábbi módon. Először is rögzítsünk egy koordináta-rendszer kezdőpontot. Az arc magasságát erre a kezdőpontra szimmetrikus, rögzített szakasznak tekintjük, szemben Chernoff megoldásával. Nála az arc magassága paraméterként szerepel, ám a minták összehasonlíthatósága miatt végül minden arcot egyforma magasságúra kicsinyített/nagyított. Ezt elkerülendő tehát legyen az arc függőleges dimenziója egy állandó. Ugyancsak eltértünk az

eredeti megoldástól abban, hogy az arc felső és alsó felét éppen közösen illesztettük össze, azaz éppen az  $x$  tengelyen. Az arc felső felén Bezier görbét alakítottunk ki oly módon, hogy az azt meghatározó 4 pontrollpont közül az első az arc legfelső, ezek szerint tehát rögzített pontja, a második kontrollpontot ugyanilyen magasan vesszük fel: ezzel azt biztosítjuk, hogy az arc bal és jobb oldala érintőfolytonosan csatlakozzon, azaz ne legyen törés az ívekben. Ennél a pontnál tehát egy paraméterre van szükség csupán. A harmadik kontrollpont szabadon változtatható (bizonyos finom megkötésekkel, melyeket nem részletezünk itt), így ennek két paramétere van, végül a negyedik kontrollpont, mely arc szélességét adja, továbbá ahol az arc felső illetve alsó fele csatlakozik, éppen az  $x$  tengelyen van, azaz egyetlen paraméterrel leírható. Ez eddig összesen 4 paraméter, melyhez az alsó arc részt leíró újabb 3 paraméter szükséges, mégpedig azért nem 4, hiszen az arc szélessége meg kell egyezzen, a két spline-nak egy kontrollpontja tehát közös. Megjegyezzük, hogy az arc alsó és felső fele nem fog érintőfolytonosan csatlakozni egymáshoz (Chernoffnál sem volt így.).

Konstrukciónkat az orr leírásával folytatjuk: ennek középpontja a szimmetriatengelyen van, a középpont és legalsó pont közötti felső negyedelőpontban. Egy paraméterrel (8. paraméter) írjuk le a hosszát, mely a középpontra szimmetrikusan mérődik. Formája nem egyszerű szakasz, hanem egy hosszabb függőleges, és egy ezzel 45 fokot bezáró, negyedolyan hosszú szakasz együttese.

A száj leírásához három paramétert használtunk mi is: 9. paraméterünk a száj középpontjának (mely a szimmetriatengelyen, azaz az  $y$  tengelyen van) függőleges koordinátája, a 10. és 11. paraméter pedig a körív egyik végpontjának relatív helyzetét írja le a középponthez képest. A körívet ezek után a középpontból és a körív most megkapott egyik végpontjából, illetve ennek tükörképéből lehet megszerkeszteni.

A szemek szerkesztése ismét némileg eltér az eredetitől. Abból a közismert megfigyelésből indultunk ki, hogy a szemek meglehetősen pontosan az arc legfelső és legalsó pontja között közösen helyezkednek el. Ez az arcok realisztikus szerkesztésének egyik arany szabálya, az arc egyik alaparánya. Mivel arra törekedtünk, hogy a megkonstruált arcok minél kevésbé legyenek torzak, a szemek függőleges helyzetét a vízszintes középvonalon, azaz az  $x$  tengelyen rögzítettük (Chernoffnál a szemek függőlegesen emelhetők voltak). Így mi négy paramétert használunk: a 12. paraméter az ellipszisnek tekintett szem középpontjának  $x$  koordinátáját adja, a 13. és 14. paraméter az egyik fókuszpont relatív helyzetét írja le ehhez képest, míg a 15. paraméterrel az ellipszis nagytengelyének hosszát adjuk meg. Nem merülünk itt el azokban a

részletekben, melyek a fenti paraméterek megfelelő beállításaihoz vezettek, csak utalunk rá, hogy itt gondolni kellett arra, hogy a szemek lehetőleg ne fedjék át egymást és lehetőleg bármilyen extrém beállításánál sem lógnak le az arcról.

A szem leírásához tartozik még, hogy a pupillát egy pontnak feleltettük meg a GeoGebrában, mely az ellipszis két fókuszpontja között helyezkedik el a 16. paraméter értékének megfelelően.

A szemöldök leírása hasonló az eredetihez: középpontja a szem ellipszisének középpontja felett van, a 17. paraméter szerinti magasságban, a 18. paraméter szerinti hosszal, a 19. paraméter által leírt ferdeséggel ( $m = -\frac{1}{2}$  és  $\frac{1}{2}$  között változhat).

Az arcot további elemekkel szerettük volna kiegészíteni a realisztikusabb hatás kedvéért és azért is, hogy még több változó megjelenítésére legyen alkalmas.

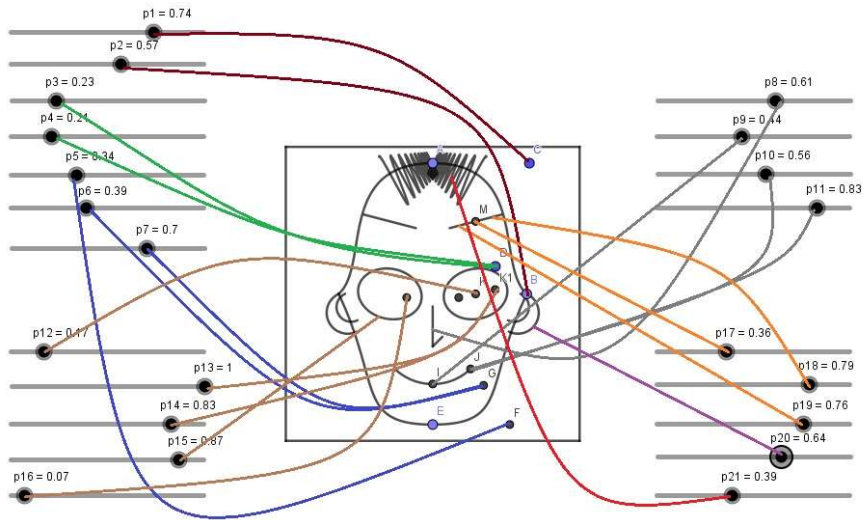
Így füleket alkotunk az arc két oldalára, melyek 2 félkörből állnak. A meghatározó, nagyobb félkör végpontja a felső és alsó arc részlet csatlakozó pontja. A fül félköre úgy került elhelyezésre, hogy az azt leíró átmérő az alsó arcfelet leíró spline görbe érintőjével essék egybe (így a fül viszonylag harmonikusan tud csatlakozni az archoz). A 20. paraméter ennek a félkörnek a sugarát írja le. Az esztétikusabb megjelenítés érdekében egy második félkört is megrajzoltunk, mely az elsőből származik elforgatással és kicsinyítéssel. Úgy éreztük, hogy haj nélkül az arc nem eléggé plasztikus. A haj leírását az

$$f(x) = x \sin\left(\frac{1}{x}\right) \quad (2)$$

függvény egy affin transzformáltja segítségével oldottuk meg, megfelelő paraméterezéssel. A megrajzolendő görbe hosszát a 21. paraméter befolyásolja.

A paraméterek és az arc meghatározó adatai közötti kapcsolatokat szemléletesen az alábbi ábrán lehet követni.

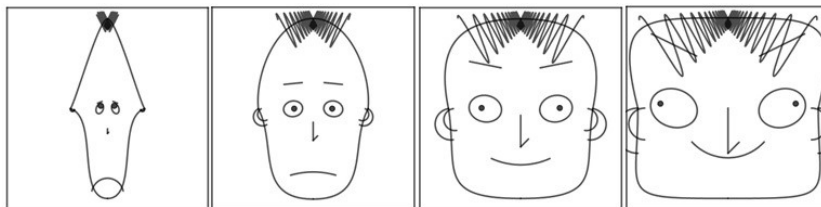
9. ábra: A 21 paraméter kapcsolata az arc meghatározó adataival



Forrás: saját szerkesztés

Illusztrációképpen bemutatunk egy ábrát, ahol a paramétereket mind egyforma értékekre állítottuk be, mégpedig sorban 10, 40, 70 majd végül 100%-ra.

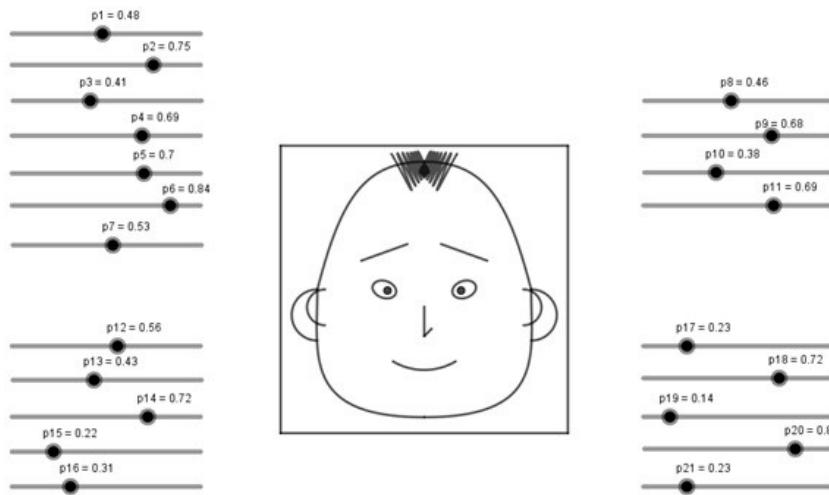
10. ábra: Arcok, ahol az összes paraméter értéke azonos, mégpedig rendre 10, 40, 70 és végül 100%



Forrás: saját szerkesztés

Hangsúlyozzuk, hogy a 21 paraméter mindegyike a másiktól függetlenül hangolható, így az arcok úgyszólván végtelen változatosságban állíthatók elő. Ennek bemutatására szolgál az alábbi ábra, ahol egy tetszőleges paramétersorozat által szolgáltatott arc látható.

11. ábra: Random paraméterekkel megjelenített Chernoff-arc, a GeoGebra felülettel együtt



*Forrás: saját szerkesztés*

#### 4. Esettanulmány komplex KSH adatsorok alapján

A fenti Chernoff-arcok alkalmazására mutatunk be egy példát. Az általunk választott adatelemzési feladatnál egyrészt kíváncsiak voltunk arra, hogy valós minta paraméterei mellett mennyire használható a kifejlesztett adatvizualizációs módszer, másfelől szerettük volna lehetőség szerint teljes körűen kihasználni az arcok paraméterezési lehetőségeit, mind a 21 paraméter bevonásával (Chernoff eredeti cikkében a 18 paraméteres arcokhoz csak olyan vizualizációkat mutat be, ahol 6 illetve 12 paramétert változtat, a többit rögzítetten hagyja.).

Az adatvizualizáció céljára kiszemelt probléma Magyarország megyéinek és a fővárosnak különböző, gazdasági, kereskedelmi, ipari, mezőgazdasági, turisztikai, fogyasztási, sőt oktatási, szociális mérőszámok alapján való összehasonlítása és a lehetséges kapcsolatok feltérképezése. Mindezek alapjául a Központi Statisztikai Hivatal STADAT összefoglaló tábláit (<https://www.ksh.hu/stadat>) használtuk 2023. szeptember 6.-i lekérdezéssel. A kimentett táblázatok Területi vonatkozásúak, Budapestre és a vármegyékre, továbbá minden esetben a 2021-es évre vonatkoznak, mivel egyes táblázatokban a 2022-es adatok még nem voltak elérhetőek. Kiválasztásunkban alapvető szerepet játszott az, hogy a tekintetbe vett információk minél szélesebb körűek legyenek, az adott térséget minél tágabb spektrumban mutassák be.

A 21 ismérv, amelyet legyűjtöttünk és az ábrákon szerepeltetünk az alábbiak:

1. A kiskereskedelmi üzletek száma (tízezer lakosra jutó üzletek száma)
2. Nettó átlagkereset (Ft)
3. A kiskereskedelmi forgalom (millió Ft)
4. A gazdasági szervezetek beruházásainak teljesítményértéke, negyedévente kumulált (millió Ft)
5. Külföldi közvetlentőke-befektetések Magyarországon (nettó tartozás állomány, milliárd Ft)
6. Vállalkozásdemográfia (működő vállalkozások száma, db)
7. Lakásállomány (100 lakásra jutó lakos), január 1.
8. Helyhez kötött internetes előfizetések száma (vezetékes és nem vezetékes összesen), december 31. (db)
9. Az ipari termelés értéke telephely szerint (Millió Ft)
10. Az építőipari termelés értéke (millió Ft)
11. Az építőipari termelés értéke a kivitelezés helye szerint (millió Ft)
12. Fajlagos lakásépítési költség, összesítve (ezer forint/m<sup>2</sup>)
13. Szántó földforgalmi ára (Ft/hektár)
14. Gabonafélék, Betakarított összes termés (tonna)
15. Gyümölcsstermelés, Betakarított összes termés (tonna)
16. Sertésállomány, december 1. (ezer db)
17. Egy hektárra felhasznált öntözővíz mennyisége (m<sup>3</sup>/ha)
18. A nyilvántartott álláskeresők száma, december (fő)
19. Az iskoláskorú népesség (fő)
20. Személygépkocsik száma, december 31. (db)
21. A kereskedelmi szálláshelyeken eltöltött vendégéjszakák összesen (db)

A fenti felsorolás egyúttal a paraméterezés sorrendjét is jelenti. Ezzel kapcsolatosan néhány megjegyzést kell tenni. Az egyik az, hogy igyekeztünk az összetartozó adatokat az arc egymással összetartozó részein megjeleníteni. Így például a 12-16. paraméterek, melyek mind mezőgazdasági vonatkozásúak a szem és pupilla 5 paramétereként jelennek meg. Ez alól az elv alól egy kivétel van, a 2. paraméter, melynek kiválasztásához az a megfontolás vezetett, hogy ez az ismérv nem vesz fel extrém alacsony értéket, amellyel így elkerülhető az, hogy az arcok túlságosan soványak legyenek.

A KSH oldaláról legyűjtött nyers adatokat át kellett alakítani az ábrázoláshoz, hiszen a paraméterek 0 és 100% közötti értékeket vehetnek fel. Ehhez minden oszlopban, minden paraméter esetén meghatároztuk annak maximális értékét, amelyet 100%-nak tekintettünk. A többi értéket ennek százalékában fejeztük ki és ezzel a paraméterrel dolgoztunk a továbbiakban. Megjegyezzük, hogy az említett skálázástól el lehetne térni. Például megtehetnénk, hogy egy adott ismérv szerint a legkisebb előforduló értéket tekintjük nullának, a



legnagyobbat pedig 100-nak és ez alapján rendelnénk hozzá a nyers adatokhoz a paramétereket. Ezzel a különbségek elnyújtottabban képeződnek le. Más megközelítés lenne, ha egy abszolút skálához viszonyítanánk az adatokat, ekkor a lehetséges paraméterértékek maximuma 100 alatt maradna. Ez utóbbinak előnye lehet, hogy olyan újabb adatsorok is megjeleníthetők lennének a későbbiekben, melyek értékei a korábbiaknál magasabbak. (A mi módszerünkben ilyenkor 100%-nál magasabb érték lépne fel, ami ábrázolhatatlan lenne.) A választandó technika tehát mérlegelés tárgyát képezi. Mi azért döntöttünk éppen emellett, mivel egyrészt minden adatsor rendelkezésre állt, azaz magasabb értékre nem számítunk, másrészt ezzel a skálázással az adatok kellően, de nem túlzott mértékben különböztetődtek meg. A számított paramétereket az alábbi táblázat tartalmazza.

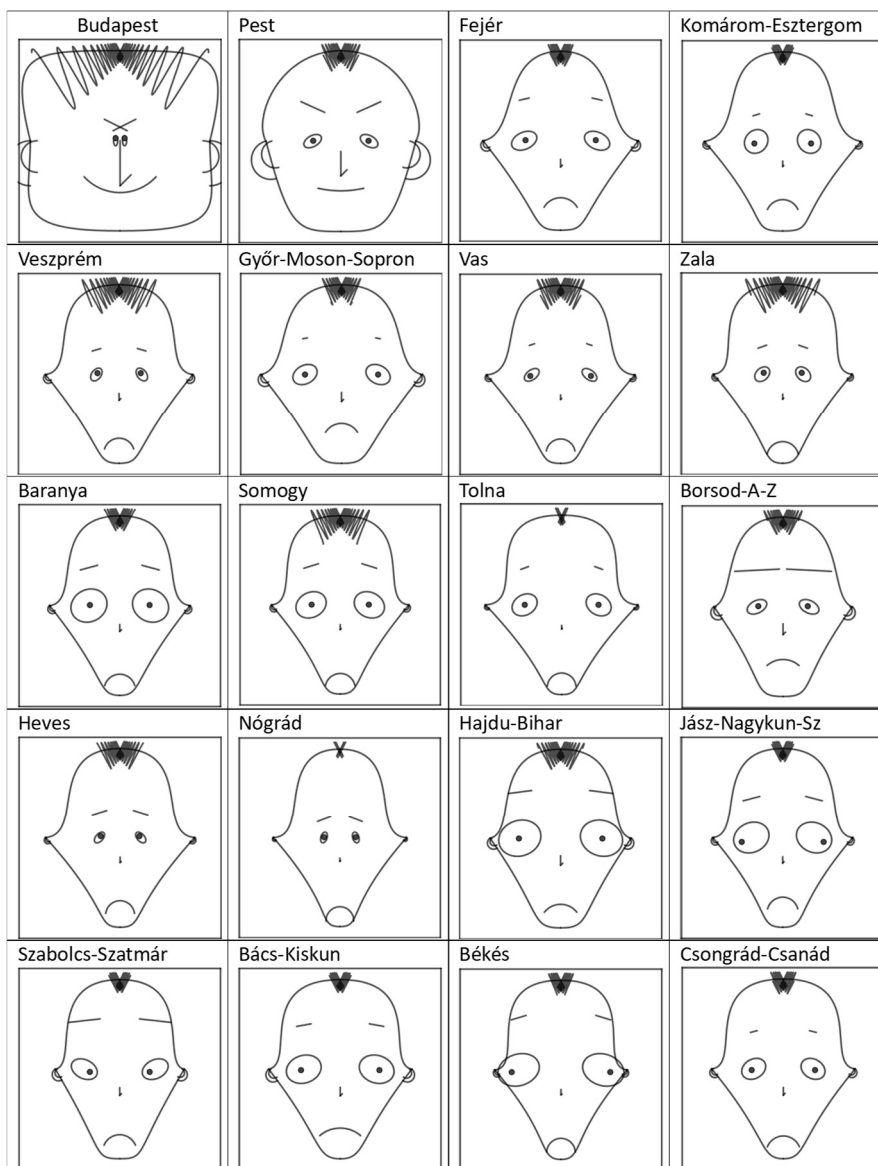
1. táblázat: Az ismérvek százalékra átdolgozott paraméter értékei

Megyék	p1	p2	p3	p4	p5	p6	p7	p8	p9	p10	p11	
	arc formája							orr	száj			
Budapest	100%	100%	100%	100%	100%	100%	71%	100%	90%	100%	100%	
Pest	68%	74%	73%	26%	22%	57%	100%	54%	100%	41%	58%	
Fejér	70%	80%	21%	7%	10%	15%	90%	17%	65%	14%	20%	
Komárom-Esztergom	77%	81%	14%	9%	15%	10%	89%	12%	60%	8%	19%	
Veszprém	90%	73%	17%	5%	4%	13%	85%	15%	32%	5%	16%	
Győr-Moson-Sopron	83%	85%	26%	8%	11%	18%	89%	20%	98%	11%	22%	
Vas	79%	74%	12%	3%	4%	9%	85%	10%	29%	4%	12%	
Zala	93%	65%	14%	3%	1%	10%	78%	11%	16%	9%	11%	
Baranya	83%	67%	16%	4%	1%	13%	81%	15%	14%	7%	14%	
Somogy	90%	65%	15%	3%	2%	10%	79%	11%	12%	6%	12%	
Tolna	99%	77%	9%	3%	1%	7%	82%	8%	16%	4%	9%	
Borsod-Abaúj-Zemplén	72%	65%	25%	10%	12%	16%	85%	24%	85%	11%	31%	
Heves	84%	74%	13%	4%	3%	9%	82%	12%	38%	4%	10%	
Nógrád	74%	61%	7%	1%	1%	5%	80%	7%	9%	2%	3%	
Hajdú-Bihar	81%	67%	22%	7%	10%	19%	85%	21%	25%	12%	25%	
Jász-Nagykun-Szolnok	81%	66%	16%	4%	5%	10%	80%	14%	54%	7%	13%	
Szabolcs-Szatmár-Bereg	85%	57%	20%	5%	6%	16%	94%	17%	27%	10%	17%	
Bács-Kiskun	97%	69%	23%	7%	4%	19%	78%	19%	57%	31%	26%	
Békés	92%	60%	13%	3%	1%	10%	74%	12%	14%	4%	7%	
Csongrád-Csanád	95%	69%	19%	4%	2%	16%	77%	17%	21%	11%	16%	
	p12	p13	p14	p15	p16	p17	p18	p19	p20	p21		
	szem					szemöldök			fül	haj		
Budapest	0%	0%	0%	0%	21%	88%	51%	100%	100%	100%		
Pest	63%	51%	8%	14%	46%	83%	53%	95%	86%	22%		
Fejér	83%	70%	3%	29%	43%	91%	22%	27%	26%	13%		
Komárom-Esztergom	70%	31%	1%	68%	73%	86%	14%	19%	19%	10%		
Veszprém	62%	25%	1%	14%	34%	87%	19%	21%	22%	42%		
Győr-Moson-Sopron	88%	59%	3%	38%	50%	82%	10%	31%	30%	21%		
Vas	75%	42%	4%	12%	63%	83%	12%	15%	17%	28%		
Zala	48%	32%	5%	24%	33%	86%	19%	15%	17%	41%		
Baranya	72%	62%	1%	97%	44%	87%	38%	23%	20%	14%		
Somogy	64%	64%	2%	64%	47%	89%	33%	19%	18%	34%		
Tolna	99%	55%	1%	47%	49%	95%	17%	13%	13%	4%		
Borsod-Abaúj-Zemplén	54%	56%	8%	18%	41%	88%	100%	45%	32%	19%		
Heves	50%	26%	6%	8%	32%	87%	32%	19%	16%	23%		
Nógrád	48%	10%	1%	6%	47%	84%	29%	12%	10%	3%		
Hajdú-Bihar	99%	84%	6%	100%	54%	91%	52%	37%	27%	24%		
Jász-Nagykun-Szolnok	67%	71%	1%	80%	100%	79%	37%	24%	19%	10%		
Szabolcs-Szatmár-Bereg	75%	67%	100%	36%	19%	87%	70%	39%	28%	10%		
Bács-Kiskun	81%	79%	10%	68%	63%	83%	32%	32%	32%	10%		
Békés	100%	100%	1%	73%	76%	91%	33%	20%	18%	12%		
Csongrád-Csanád	72%	53%	3%	44%	64%	100%	15%	25%	22%	13%		

Forrás: saját szerkesztés a <https://www.ksh.hu/stadat> oldal adatai alapján

A fenti adatok alapján elkészítettük a főváros és a 19 vármegye Chernoff-arc adatvizualizációját, melyet az alábbi ábrán foglaltunk össze.

12. ábra: KSH adatsorok alapján készített Chernoff-arc adatvizualizáció a 19 vármegyéhez és Budapesthez



*Forrás: saját szerkesztés*

## 5. Regionális gazdaságfejlesztési konzekvenciák – elemzés a grafikák alapján

Megállapítható, hogy a táblázatos formában elrendezett ábrák viszonylagos változatosságot mutatnak, karakteresek, így az adatvizualizációt a szemléletesség szempontjából sikeresnek értékeltük.

A kapott Chernoff-arcokkal kapcsolatban azonban kíváncsiak voltunk arra is, hogy mennyire alkalmasak arra, hogy általuk összefüggéseket fedezzünk fel az adatsorok között.

Ebből a célból a következő kísérletet végeztünk el. A 12. ábrán látható arcokat külön-külön kinyomtattuk és pusztán számokkal láttuk el. A 20 képecskéből álló gyűjteményt azután 6 embernek megmutattuk és arra kértük őket, hogy “a rajzokat csoportosítsák családonként, nem tudva, hogy hány családot kell kialakítani”. Az utasítás pusztán ennyi volt. Hangsúlyozzuk, hogy a kísérletben részt vevők sem a kutatás háttérébe, sem az ábrázolt arcok és régiók közötti kapcsolatba, úgyszólván tehát semmilyen részletbe nem voltak beavatva, mindössze a számokkal ellátott képek csoportosítása volt a feladatuk.

A kapott megoldások között volt olyan, amely mindössze 4 családba rendezte a képeket, de tipikusan 6-7 családdá szervezték őket a tesztalanyok. A 20 ábra közötti kapcsolatokat egy gyakorisági mátrixban gyűjtöttük össze, amely azt mutatja, hogy a megfelelő oszlopa és sora szerinti képeket hány személy gondolta összetartozónak. A számozás a fenti táblázatot követte: 1: Budapest, 2: Pest, 3: Fejér, ..., 9: Baranya, ..., 20: Csongrád-Csanád. Ha a k. és n. ábrák között m db kapcsolatot találtak, akkor az a mátrixban csak egyszer szerepel (csakis  $k \leq n$  esetére). A színezett átlóban akkor látunk értéket, ha az alany a szóban forgó arcot egyedül hagyta, senkivel sem hozta kapcsolatba.

2. táblázat: Kapcsolatokat mutató gyakorisági mátrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Σ	
1		4	2																			2
2			3				1					1			1							5
3				2	5				2	1		4			1	1	2					19
4					1	2			1	1	2	1			2	2	1	3	2			22
5							3	5			1		4	3	1							18
6									1	1		5			2	1	2					20
7								3					5	2		1				1		15
8									1	2	1		4	4							1	22
9										5	2	1		1	2	1		1	3	4		25
10											2	1		1	2	1		1	3	3		25
11														3	1	1		2	2	4		20
12															2	1	3					19
13																						16
14																				1	1	19
15																2	1	3	1	1	1	22
16																	2	3	4	1		21
17																		1	2	1		14
18																			1	2	2	19
19																					3	24
20																						23

Forrás: saját szerkesztés

A mátrixban a legalább 4 gyakoriságú kapcsolatokat színezéssel tüntettük fel. Érdemesnek látjuk, hogy a feltárt kapcsolatokat több szinten közelítsük meg.

Megállapítjuk, hogy a lehetséges legnagyobb fokú kapcsolat, azaz amikor mind a 6 alany ugyanazt a rokonságot fedezte volna fel, nem áll elő.

Ugyanakkor 5-ös szintű kapcsolatokat (amikor 5 tesztalany is egy családba sorolta a képeket) véve tekintetbe négy összefüggés is mutatkozik: Fejér, Győr-Moson-Sopron és Borsod-Abaúj-Zemplén vármegye hármasa; ezen felül pedig még 3 párt láttak alanyaink: Veszprém és Zala; Vas és Heves, továbbá Baranya és Somogy vármegyék vonatkozásában.

Még további, 4-es szintű, gyengébb kapcsolatokat is figyelembe véve, a fenti 4 család hárommá egyesül. Ugyanakkor megjelenik, mint önálló entitás Budapest is, illetve egy további új pár. Így tehát, egyéb gyengítéseket már nem megengedve, végülis öt halmazt kapunk, melyeket egyenként, a kísérlet nyomán, erősen összefüggőeknek vélünk.

Ezek: Budapest önmagában;

Fejér, Győr-Moson-Sopron és Borsod-Abaúj-Zemplén vármegyék;

Veszprém, Vas, Zala, Heves és Nógrád vármegyék;

Baranya, Somogy, Tolna és Csongrád-Csanád vármegyék;

Jász-Nagykun-Szolnok és Békés vármegyék.

Megjegyzendő, hogy Budapest azért nem jelentkezik korábban, magasabb szinten, mivel őt és Pest vármegyét egyesek összekapcsolták (amint az a táblázatból látható is).

Másfelől, ha a többiektől eltérő, kiugró értékre fókuszálunk a reprezentációkban, akkor a kapcsolatokat feltáró táblázat utolsó oszlopa alapján, ami egy elem többi elemmel való összes kapcsolatának számát mutatja, elmondható, hogy Budapest és Pest megye szerepe kivételes. Budapest csakis Pest vármegyével szerepel egy kalapban, de ez is szórványos. Budapest tehát a reprezentáció alapján kiugró szereppel bír. Kicsit kisebb mértékben, de hasonló Pest vármegyéről is elmondható.

Bízunk benne, hogy fenti eredmények új szempontot hozhatnak a regionális gazdaságfejlesztés területére.

## 6. Összefoglalás

A bemutatott, félszáz éves Chernoff-arcok módszerére jelen publikáció egy új, dinamikus geometriai eljárást kínál. Maga az alapelv éppen az ellenkezője a képek digitalizálásának, amely ezekből számokat generál abból a célból, hogy mondjuk a betáplált arcok megkülönböztethetők legyenek. Chernoff-arcok segítségével éppen számokból készítünk képeket, arcokat, hogy a számsorokban rejlő különbségek láthatóvá, könnyebben értelmezhetővé váljanak. Megjegyezhető, hogy az arcok akár további paraméterekkel is felszerelhetők lennének. Őket például buborék diagrammal ötvözve a síkbeli helyzet által további 2 paramétert nyerhetünk, s az arc mérete pedig még egy további paramétert jelenthet.

Elmondjuk, hogy a Chernoff-arcok módszerét számos kritika is érte az arc egyes elemeinek szerepe szempontjából. Például [20] szerint a szem és szemöldök szerepe fontosabb az orrnál vagy szájánál. Vizsgálataink alapján a szemméret és a szemöldök ferdeség a legpontosabb jellemzők. Eredményeink alapján úgy vélik, hogy a Chernoff-arcok nem feltétlenül rendelkeznek jelentős előnnyel más megjelenítési technikákkal szemben a többdimenziós információ megjelenítésében. Ugyanakkor érdemes megemlíteni, hogy a szerzők által használt Java program például az arc formáját nem is variálja, pusztán a szem (konstans pupillával), szemöldök, orr és száj kerül megjelenítésre egy kör alakú arcban. Úgy véljük, hogy nagyobb variabilitást megengedve fenti kritika finomítható.

Az oktatást érintő érdekes és figyelemre méltó eredmény található a [21] tanulmányban. Ebben a szerzők standard és nem-standard adatvizualizációkat mutattak dizájn szakos és nem szakmabeli egyetemi hallgatóknak. A kutatásból az derült ki, hogy a laikusok jobban bíznak és értik a standard technikákat, míg a nagyobb vizuális ismeretekkel rendelkező, dizájn szakos hallgatókat jobban vonzzák a nem-standard módszerek. A vizsgálatban a nem-standard oldalon infografikák szerepeltek, Chernoff-arc vizualizációk azonban nem. Úgy tűnik, hogy az egyetemi képzésben nagyobb szerepet kellene szánni a nem-standard módszereknek is, hogy azok érthetővé váljanak a hallgatók, a jövő szakemberei számára.

Mindazok számára, akik szerény matematikai/statisztikai ismereteik miatt nem tudnak kimutatásokat vagy ezekhez kapcsolódó grafikonokat készíteni, hasznos segédeszköz lehet a Chernoff-arcok módszere. Ezen felül, a többi adatvizualizációhoz képest a Chernoff-arcok a legszórakoztatóbbak, úgy hisszük a többi vizualizációhoz képest az arcok elemzésével szívesebben töltenek időt az emberek.

## Irodalomjegyzék

- [1] Manovich, L. (2020). *Cultural analytics*. MIT Press; doi: <https://doi.org/10.7551/mitpress/11214.001.0001>
- [2] Arbia, E. (2023). *The regulation of digital platforms: the case of pagoPA*. (Doctoral thesis);
- [3] Morais, A. M. M., Quiles, M. G., & Santos, R. D. (2014). Icon and geometric data visualization with a self-organizing map grid. In *Computational Science and Its Applications–ICCSA 2014: 14th International Conference, Guimarães, Portugal, June 30–July 3, 2014, Proceedings, Part VI 14* (pp. 562-575). Springer International Publishing; doi: [https://doi.org/10.1007/978-3-319-09153-2\\_42](https://doi.org/10.1007/978-3-319-09153-2_42)
- [4] Keim, D. A., & Kriegel, H. P. (1996). Visualization techniques for mining large databases: A comparison. *IEEE Transactions on knowledge and data engineering*, 8(6), 923-938.; doi: <http://dx.doi.org/10.1109/69.553159>
- [5] Claessen, J. H., & Van Wijk, J. J. (2011). Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2310-2316. doi: <http://dx.doi.org/10.1109/TVCG.2011.201>
- [6] Borgo, R., Kehrer, J., Chung, D. H., Maguire, E., Laramée, R. S., Hauser, H., ... & Chen, M. (2013, May). Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics (state of the art reports)* (pp. 39-63);
- [7] Lee, M. D., Reilly, R. E., & Butavicius, M. E. (2003, January). An empirical evaluation of Chernoff faces, star glyphs, and spatial visualizations for binary data. In *Proceedings of the Asia-Pacific symposium on Information visualisation-Volume 24* (pp. 1-10);
- [8] Korhonen, P. (1991). Using harmonious houses for visual pairwise comparison of multiple criteria alternatives. *Decision Support Systems*, 7(1), 47-54.; doi: [http://dx.doi.org/10.1016/0167-9236\(91\)90076-N](http://dx.doi.org/10.1016/0167-9236(91)90076-N)
- [9] Dombi, J. (2012). *Intelligens rendszerek, Elektronikus jegyzet*, Szegedi Tudományegyetem;
- [10] Andrews, D. F. (1972). Plots of high-dimensional data. *Biometrics*, 125-136.; doi: <http://dx.doi.org/10.2307/2528964>
- [11] Chernoff, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American statistical Association*, 68(342), 361-368.; doi: <http://dx.doi.org/10.1080/01621459.1973.10482434>
- [12] Pitt, L., Mills, A. J., Chan, A., Menguc, B., & Plangger, K. (2011, June). Using Chernoff faces to portray social media wine brand images. In *6th AWBR International Conference*;
- [13] Kozicki, B., Stajniak, M., Magniszewski, M., Lorek, M., & Mitkow, S. Z. (2022). Multidimensional Analysis of Real Estate Prices in Seventeen Cities in Poland in Terms of Economic Security. *Journal of Security and Sustainability Issues*, 12(1); doi: <http://dx.doi.org/10.47459/jssi.2022.12.3>
- [14] Lopez-Conde, R. C., Nalipay, J. N., Ancho, I. V., Abulon, E. L. R., Rungduin, T. T., Montealegre, M. A. C., & Madronero, J. A. (2022). Utilizing Chernoff Faces in Modeling Responses in the Evaluation of Trimester Scheme Impementation. *Dalat University Journal of Science*, 125-138.;
- [15] Raciborski, R. (2009). Graphical representation of multivariate data using Chernoff faces. *The Stata Journal*, 9(3), 374-387.; doi: <http://dx.doi.org/10.1177/1536867X0900900302>
- [16] Zhang, H., Hou, Y., Zhao, J., Wang, L., Xi, T., & Li, Y. (2017). Automatic welding quality classification for the spot welding based on the Hopfield associative memory neural network and

- Chernoff face description of the electrode displacement signal features. *Mechanical Systems and Signal Processing*, 85, 1035-1043.; doi: <http://dx.doi.org/10.1016/j.ymsp.2016.06.036>
- [17] Teke, E. Ç., Koşkan, Ö., & Köknaroglu, H. (2020). Chernoff faces application in livestock. *Ziraat Fakültesi Dergisi*, 15(1), 64-71.;
- [18] GeoGebra, <https://www.geogebra.org/>;
- [19] Talata, I., Bölcskei, A., Budai, L., Keresztes, É. R. (2023). Lineáris programozási feladatok vizualizációja GeoGebrával. In: Ország, Adrienn; Baják, Szabolcs (szerk.) I. Csernyák László konferencia közleményei Budapest, Magyarország : Budapesti Gazdasági Egyetem (BGE) (2023) pp. 24-39.; doi: [http://dx.doi.org/10.29180/978-615-6342-61-4\\_3](http://dx.doi.org/10.29180/978-615-6342-61-4_3)
- [20] Ebert, D., Morris, C., & Rheingans, P. (1999). An Experimental Analysis of the Effectiveness of Features in Chernoff Faces. University of Maryland Baltimore County;
- [21] Quispel, A., & Maes, A. (2014). Would you prefer pie or cupcakes? Preferences for data visualization designs of professionals and laypeople in graphic design. *Journal of Visual Languages & Computing*, 25(2), 107-116. doi: <http://dx.doi.org/10.1016/j.jvlc.2013.11.007>
- [22] Kocsó, E.; Cserné Pekkel, M. (2020). Using of Dynamic Animations to Illustrate Mathematical Theorems *TRANSACTIONS ON IT AND ENGINEERING EDUCATION* 3 : 1 pp. 1-16. , 16 p. <http://jtite.eu/article/view/23/18>