**Frauenanteil im Hochschulbereich –**

**eine komplexe Analyse von internationalen Daten**

**unter dem Gesichtspunkt der Regionalität und von Zeitreihen**

(*Annual ratio of women in tertiary education worldwide –*

*a complex analysis of international data concerning regional and time aspects*)

László Pitlik (jun) – PhD László Pitlik

ELTE TTK / SZIE MY-X Forschungsgruppe

**Die vorliegende Studie hat das Ziel, das zahlenmäßige Verhältnis von Frauen und Männern an Hochschulen (jährlich und landbezogen) so darzustellen – basierend auf den Statistiken von UNESCO (1970-2015), dass daraus die Trends vor und nach der Krise, bzw. in regionalen Untergliederungen sichtbar werden.**

**Die Analyse ist eine Art Fortsetzung eines Projektes der My-X Forschungsgruppe, welches Thema bereits in der III. Tagung der Multikulturalität (2017) vorgetragen wurde – mit dem Schwerpunkt: Visualisierung. Die erstellten Animationen sind inzwischen einen festen Bestandsteil des Beitrages für die EDEN Konferenz in Jönköping geworden. Die Motivation für die Bearbeitung des Themas lässt sich auf die andauernden Reformversuche des Bildungswesens weltweit zurückführen, um sehen zu können, welche Schlussfolgerungen in Raum und Zeit aus den Änderungen in der Vergangenheit für die Zukunft mit passenden Werkzeugen für Benchmarking und Visualisierungseffekten abgeleitet werden können. Als Ergebnis werden regionale Entwicklungstrends gefiltert – und daran gemessen die Prozesse vor allem in Mittel- und Ost-Europa bewertet – sowohl vor als auch nach der Krise. Das verwendete Visualisierungsprogramm ist GAPMINDER, wo man in 6 Dimensionen Zusammenhänge graphisch darstellen kann. Basierend auf den Animationseffekten werden auch die ungarischen Daten mit westlichen und östlichen Verhaltensmustern verglichen. Die bisherigen Ergebnisse hatten einen globalen Charakter, während hier und jetzt die Krise und die Regionen in den Mittelpunkt gestellt werden. Die Ergebnisse der Forschungsaktivitäten sollten passende fachliche und visuelle Impulse für die datengesteuerte Politikgestaltung liefern können.**

***Schlüsselwörter: Bildung, Krise, Regionen, Zeitreihen, Analyse***

### **1. Einführung – Vorgeschichte – Prolog**

Das Problem, charakteristische Differenzen in Zeitreihen zu identifizieren, ist eine Art universelles Problem (vgl. GPS – *general problem solving*). Das kontextunabhängige Wissen wird in solchen Fällen kontextualisiert: hier und jetzt ist der Kontext nichts anderes, als der Anteil von Frauen an Hochschulen weltweit anhand von Zeitreihen. Die adaptierte Frage lautet: Welche Region/Ländergruppe scheint durch die Krise ab 2008 am negativsten betroffen zu sein? Die gleichen Daten von UNESCO wurden bereits verarbeitet (Pitlik 2017) – mit dem Ziel: das Visualisierungsniveau zu maximieren.

Das Motto der Studie wurde von Knuth (1992) übernommen: „*Science is what we understand well enough to explain to a computer. Art is everything else we do.*“ Anders (durch die Autoren seit Jahren) formuliert: Wissen/Wissenschaft ist, was in Quellcode zu transformieren ist, jede andere menschliche Aktivität ist Kunst (vgl. [http://miau.gau.hu/miau2009/index\_tki.php3?\_filterText0=\*knuth](http://miau.gau.hu/miau2009/index_tki.php3?_filterText0=*knuth), ill. <http://miau.gau.hu/miau/208/20151120.pptx> – Pitlik 2015).

Parallel dazu schrieb Enyedi in 2001: „*Arra vonatkozóan, hogy ilyennek képzeltem-e az ezredforduló magyar társadalmát, rövid a válaszom: nem ilyennek. A világ és Kelet-Közép-Európa folyamataiból ez nem volt levezethető. Ilyesmi elképzelésem volt (az utólagos rekonstrukciók persze mindig pontatlanok): a kelet-közép-európai társadalmak – az 1970-es évek elején bizonyítható – gazdasági felzárkózásuk eredményeként közeledni fognak a nyugat-európaihoz (amiben a nyugati szociális piacgazdaság is közrejátszik).*“

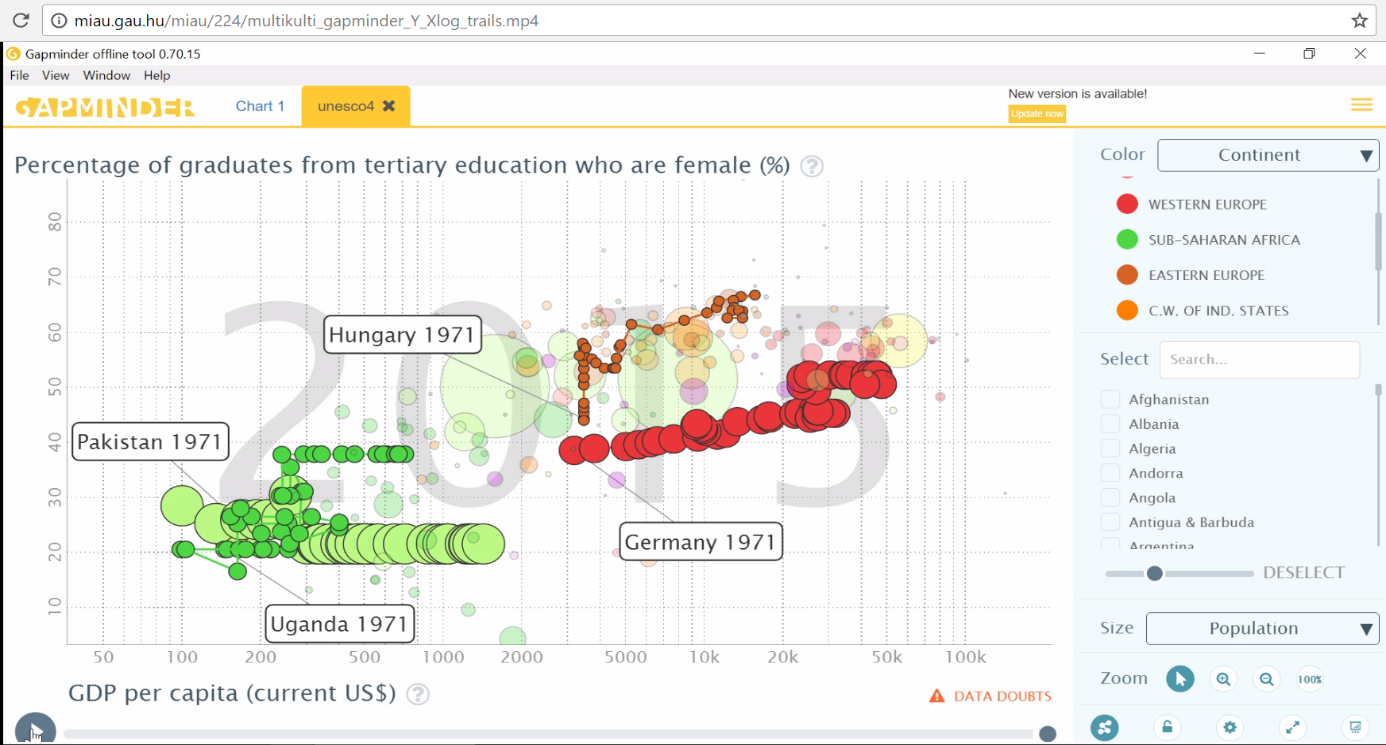
Wie es dem Zitat von Enyedi zu entnehmen ist: Wissenschaftliche Aussagen sollten ohne jegliche Rekonstruktion mit den passenden Fakten (auf elementaren, messbaren Ebenen) vergleichbar sein. Menschliche Begriffe (wie „Aufholen“ = „*felzárkózás”)* sind nur dann interpretierbar, wenn diese als „beweisbar“ (= „bizonyítható“) anzuerkennen sind (vgl. <http://real.mtak.hu/39021/1/Studia_Mundi_Vol_2_No_3_14_Varga.pdf> – über nachträgliche Kontrolle des „Aufholens“ – Varga 2014).

In dieser Studie werden die elementaren Schichten der Problemstellung gehandhabt – und gleichzeitig werden alle Aussagen aus den Fakten abgeleitet – so, dass eine Art Automatisierung jeder Zeit möglich sein wird.

### **2. Theoretischer Hintergrund**

Da diese Studie als eine direkte Folge der Studie über die Multikulturalität und über deren Datenvisualisierungsmöglichkeiten (siehe Abbildung 1 – wo länderspezifische Zeitreihen schwer zu analysieren sind) anzusehen (Pitlik 2017) ist, beschränkt sich dieses Kapitel auf die verwendete Methodologie – und zwar auf die kurze, generelle Darstellung des Verfahrens, wo man in Zeitreihen nach den bestmöglichen Schnittpunkten sucht – basierend auf Künstliche Intelligenz Algorithmen (wie Ketten aus Ähnlichkeitsanalysen).

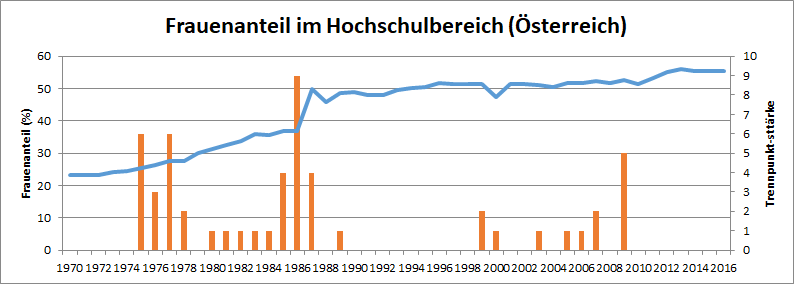
### **Abbildung 1: Beispiel für die Rosling-Animationen**



**Quelle: Eigene Darstellung**

Nimmt man eine zufällig ausgewählte der Abbildungen (siehe Abbildung1) des Vortrages zu dieser Studie (Pitlik – Pitlik 2017), dann wird es sofort ersichtlich, dass den besten Trennpunkt mit menschlichen Augen zu finden, eine scheinbar einfache Herausforderung zu sein scheint. Die orangenfarbenen Balken der Trennpunktstärke (Y2-Achse – rechts) haben einen maximalen Wert gerade dort (etwa in der Mitte der 80-er Jahre – siehe X-Achse), wo das menschliche Auge spontan die grobsten Unterschiede im Verlauf der Zeitreihe ohne nachzudenken entdecken kann. Eine solche elementare Herausforderung kann man bereits für ein KI-Verfahren weitergeben: man muss lediglich fähig sein, den Begriff „Anders-Sein“ zu definieren und zu quantifizieren. Dies ist möglich z. B. mit Ähnlichkeitsanalysen (Pitlik 2015b). In technischen Anwendungen reichte es sogar die künstlich optimierte Begriffsbildung durch klassische Statistiken zu ersetzen (Pitlik L – Pitlik M. 2016/2017).

### **Abbildung 2: Beispiel für die Suche nach Trennungspunkten**



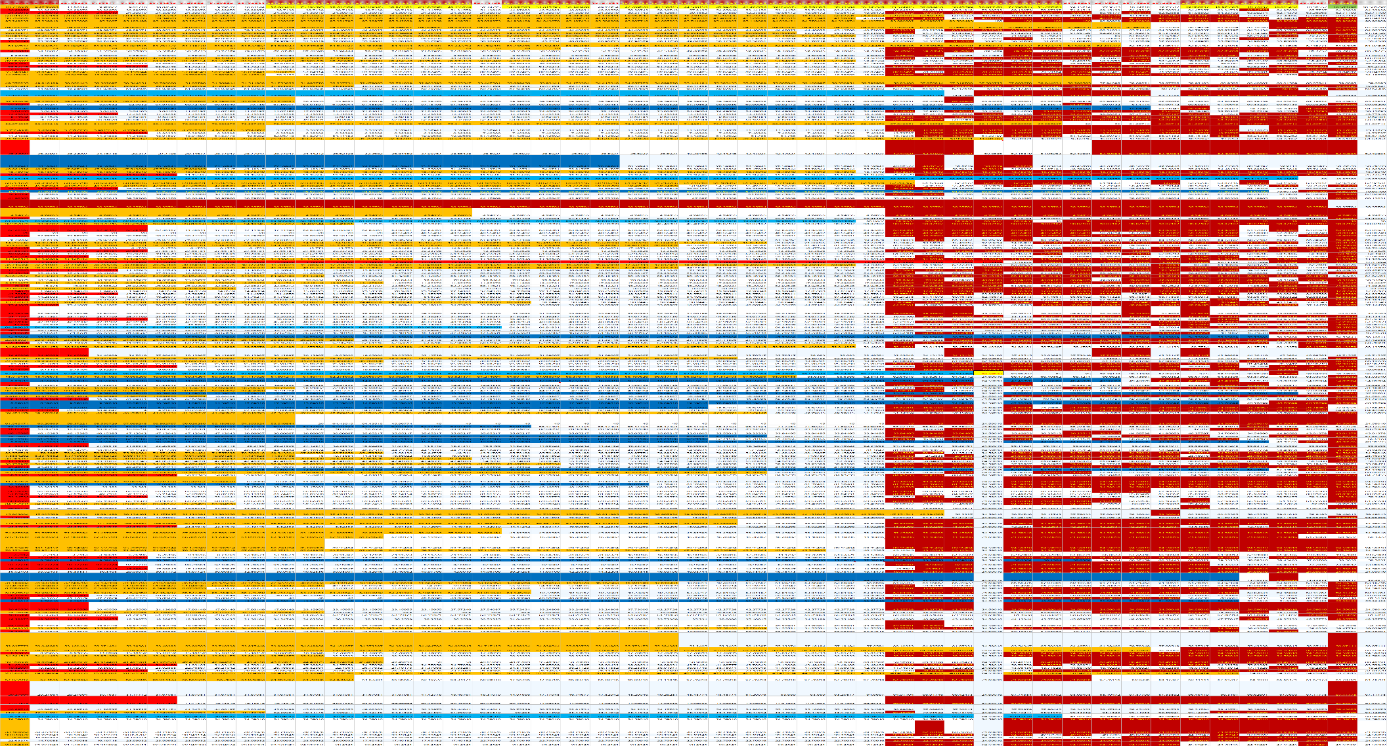
**Quelle: Eigene Darstellung**

### **3. Daten und Methoden**

Die für diese Studie verarbeiteten Daten stammen aus einer UNESCO Datenbasis (<http://miau.gau.hu/miau/229/enyedi2/unesco_enyedi_v3.xlsx>).

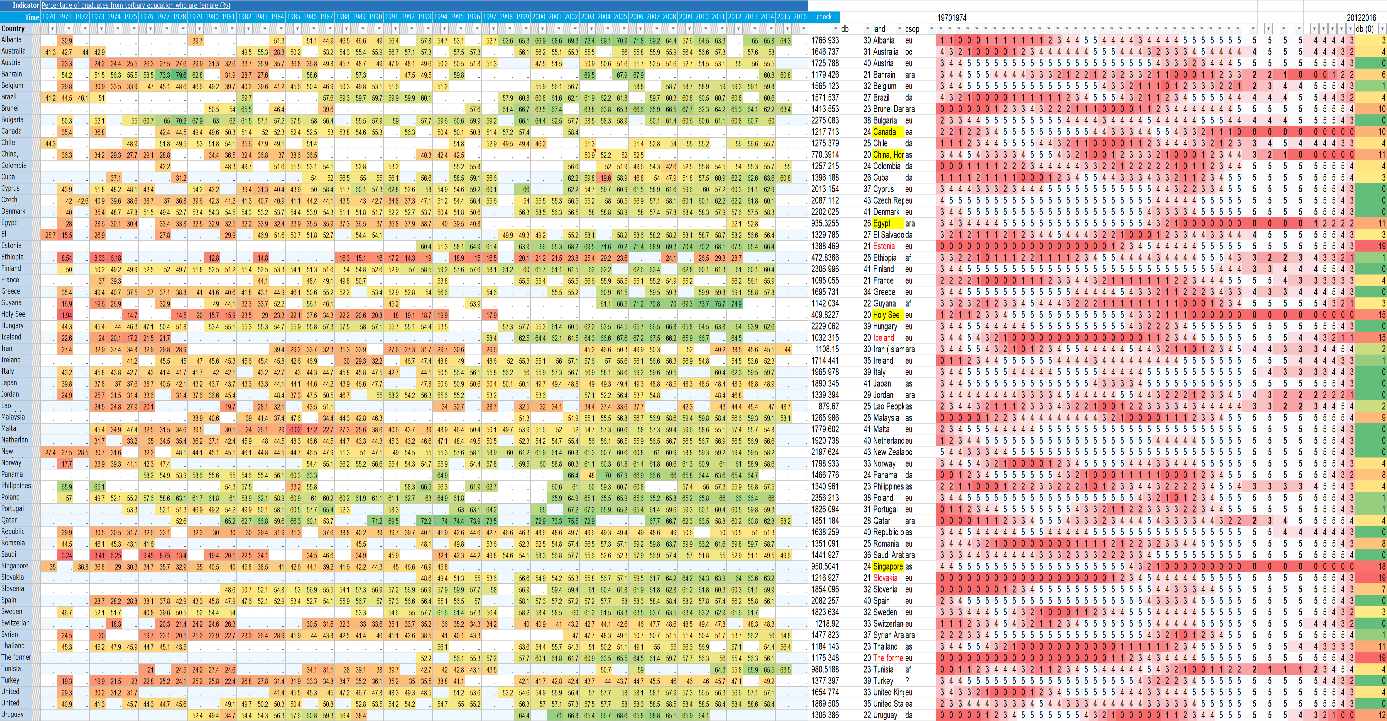
Die Qualität der Daten lässt sich nicht loben: d.h. es ist kaum zu akzeptieren, dass eine Weltorganisation (finanziert letztendlich aus Steuern) von den Mitgliedsstaaten nur diese Art „Qualität und Quantität“ erwarten darf und nur solche an die Forschung (welche als eine triviale, wiederum gemeinnützige Aktivität ist) weitergeben kann.

**Abbildung 3: Beispiel Datenqualität I.**



**Quelle: Eigene Darstellung**

### **Abbildung 4: Beispiel Datenqualität II.**



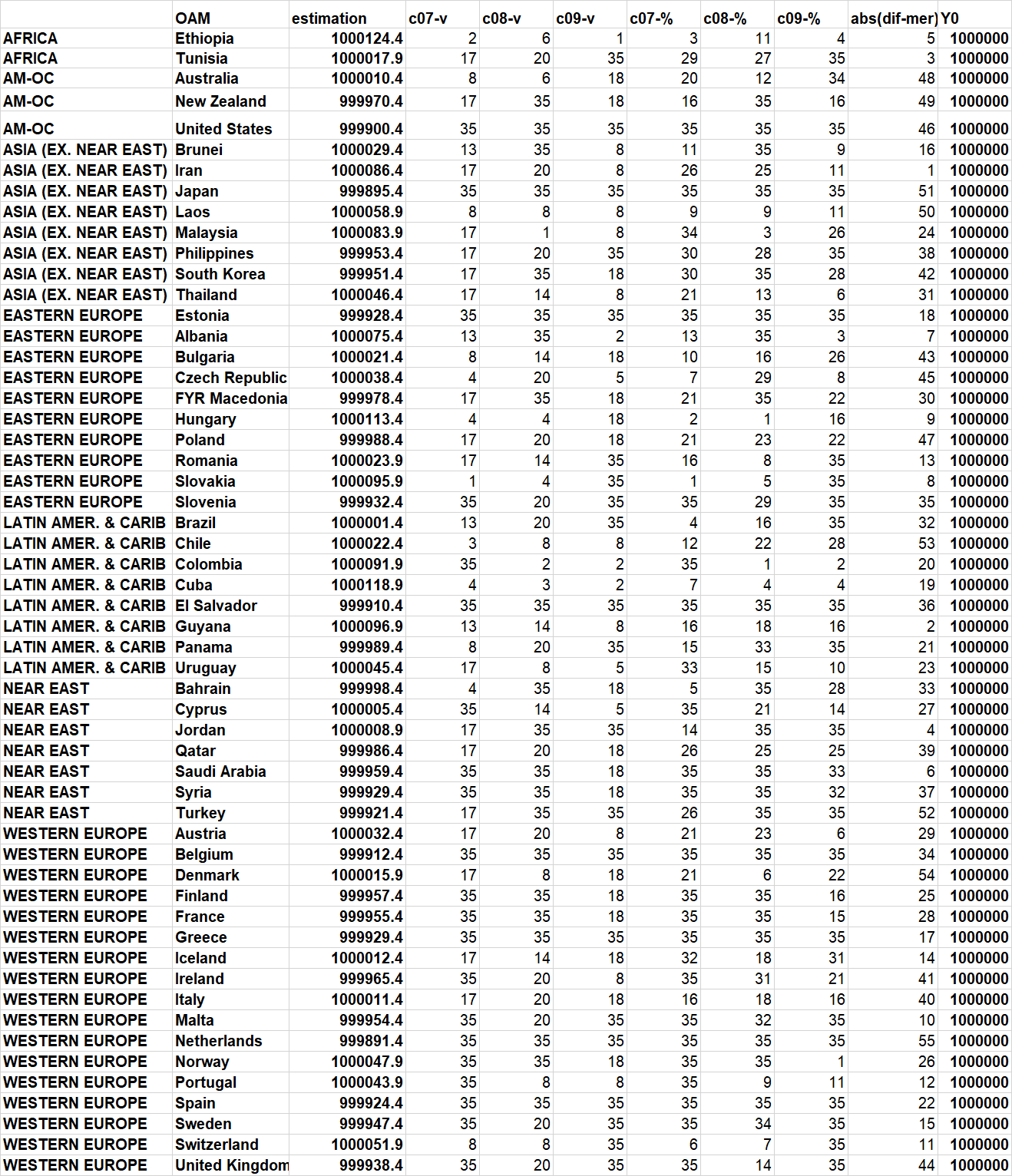
**Quelle: Eigene Darstellung**

Die Abbildungen 3 und 4 zeigen die Daten zunächst für 241 Länder und Ländergruppen so, dass nur die schwarz-weißen Zellen als gelierte Zahlen zu interpretieren sind. Es wurden aber lediglich die Länder verarbeitet, wo in der potenziellen Zeitspanne (1970-2016 – 47 Jahre) mindestens 20 Werte vorlagen. So blieben nur 55 Objekte übrig.

Abbildung 4 zeigt in der linken Seite eine Art Schnell-Interpretationshilfe: die grünlichen Werte sind die höheren Anteile, während die rötlichen Zellen die niedrigeren Frauenanteile beinhalten. Die rechte Seite zeigt parallel eine Art Wärmekamera-Aufnahme, wo die rötlichen Zellen darauf hinweisen, dass 5-Jahres Intervallen keine oder nur wenige Daten beinhalten.

Was die verwendete Methode betrifft: es wurde die vereinfachte (online) Trennpunktanalysator-Lösung (<http://miau.gau.hu/myx-free/tools/cutting_robot/cutting3.php>) miteinbezogen. Die Mindestlänge für die Berechnung der Zeitreihen-Charakteristika wurde am Anfang und am Ende der Zeitreihen der Länder mit dem Wert 6 versehen.

## Abbildung 5: Objekt-Attribute-Matrix (OAM) für Ähnlichkeitsanalyse



**Quelle: Eigene Darstellung**

Die Länder (Objekte) wurden mit Attributen versehen (s. Abbildung 5) wie folgt:

## c07-v: Trennungsstärke im Jahr 2007

## c08-v: Trennungsstärke im Jahr 2008

## c09-v: Trennungsstärke im Jahr 2009

## c07-%: relative Trennungsstärke im Jahr 2007 (100% ergibt sich durch die maximale Trennungsstärke entlang der Gesamtzeit)

## c08-%: relative Trennungsstärke im Jahr 2008 (100% ergibt sich durch die maximale Trennungsstärke entlang der Gesamtzeit)

## c09-%: relative Trennungsstärke im Jahr 2009 (100% ergibt sich durch die maximale Trennungsstärke entlang der Gesamtzeit)

## ABS(DIF-MER): Absoluter Differenzwert zwischen Trendsteigungsparameter vor (2003-2007) und nach der Krise (2009-2013)

## Y0: Modellkonstant für eine anti-diskriminative Modellierung

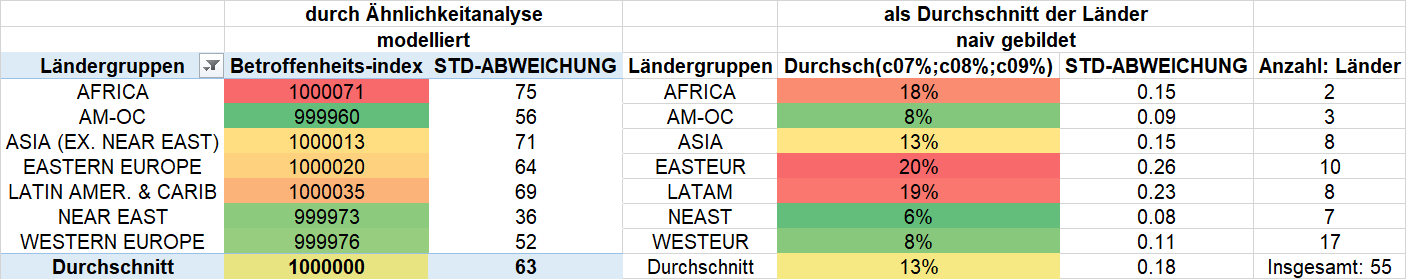
Basierend auf OAM wurde eine Anti-Diskriminanz-Analyse durchgeführt um prüfen zu können, ob eventuell jedes Land auf eine andere Art und Weise, jedoch die gleiche Betroffenheit hatte? Die Anti-Diskriminanz-Analyse sucht nach einer solchen Treppenfunktion, wonach die unterschiedlichen Betroffenheitsschichten nach Attributen und deren Rangzahlen die perfekteste Annäherung an das Gleichheitsprinzip (dargestellt durch eine konstante Zahl als abhängige Variable von 1.000.000) ermöglicht.

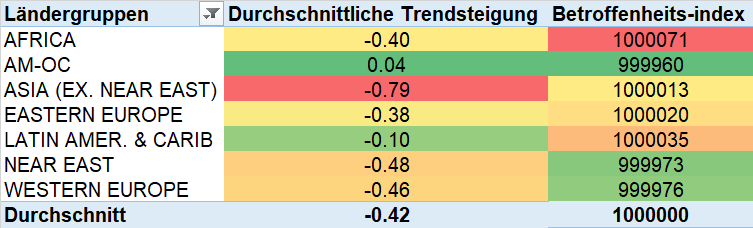
### **4. Ergebnisse**

Die Abbildung 6 gibt eine deutliche Antwort auf die elementare Frage der Studie: Welche Region/Ländergruppe wurde am meisten durch die Krise betroffen? Wie es an der linken Seite der Abbildung6 zu sehen ist (basierend auf einer optimierungsorientierten Lösung), die relative Betroffenheitsrangfolge lässt sich wie folgt darstellen: Afrika>Latein-Amerika>Osteuropa (inkl. Ungarn).

Die rechte Seite der Abbildung6 führt eine naive Lösung vor, wo die Betroffenheitsgründe eine Art zufällige Gewichtstruktur haben, und zwar die durchschnittlichen relativen Trennungsstärken als Bewertungsskalen akzeptiert werden. Außerdem: die linke Seite der Abbildung6 stellt eine Lösung dar, wo optimierte Gewichte im Hintergrund aller Attribute mit Ähnlichkeitsanalyse abgeleitet wurden. Die naive und die optimierte Bewertung der durchschnittlichen Betroffenheit der Ländergruppen laufen grundsätzlich parallel. Die Abbildung8 (oder die nicht absoluten Trendsteigungsdifferenzen) macht es jedoch erst möglich über die Art der Betroffenheit zu sprechen. Rötliche Markierungen weisen auf eine höhere Betroffenheit (Empfindlichkeit) hin. Die gelblichen Markierungen symbolisieren eine Art Norm, und die grünen Markierungen stellen ein relativ niedriges Niveau an Betroffenheit dar. Der untere Teil der Abbildung6 lässt sich wie folgt interpretieren: Afrika ist stark empfindlich und die Krise hat eine Art Verschlechterung der Situation hervorgerufen. Ozeanien ist rel. nicht empfindlich und die Krise brachte sogar eine Verbesserung. Asien ist mittelmäßig empfindlich, aber die Verschlechterung der Trends erreichte hier eine Art Höhepunkt. Ost- und Mittel-Europa liegt soweit in der Mitte sowohl in der Empfindlichkeit gegenüber der Krise als auch was die durchschnittliche Verschlechterungsquote betrifft. Latein-Amerika ist relativ empfindlich mit einer niedrigen Verschlechterung. Naher Osten und Westeuropa sind quasi nicht empfindlich, aber eine massive Verschlechterung ist wahrzunehmen.

## Abbildung 6: Ergebnisse



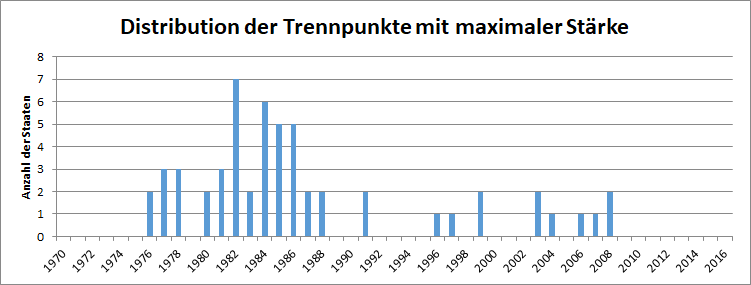


**Quelle: Eigene Darstellung**

Die Abbildung 7 demonstrierte eine Art nichtgeplantes Ergebnis, indem man aus den Trennungsstärken in den ganzen Zeitreihen auch das ableiten kann, wo (noch präziser wann) die Charakteristika der Zeitreihen am häufigsten auf primäre Trendunterschiede hinweisen. Wie es zu sehen ist, ist etwa in der Mitte der 80-er Jahre eine weltweite Wende zu vermuten. Die sog. Krise (2008) hatte viel weniger Wirkung auf die Anteile der Frauen an den Hochschulen hinterlassen, als die Zeiten davor.

Die optimierte und die naive Lösungen zeigen keine störenden Unterschiede in diesem Fall – u.a. eben deswegen konnte sich dieser naive Lösungsansatz so lange in der Geschichte halten.

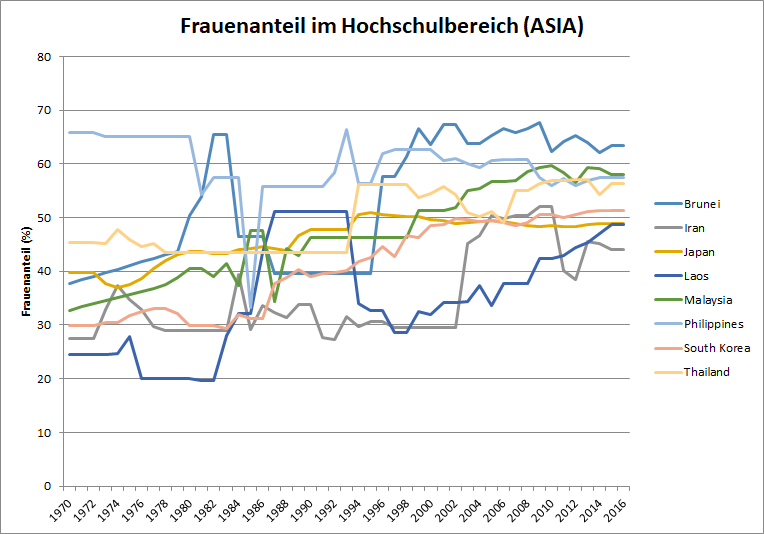
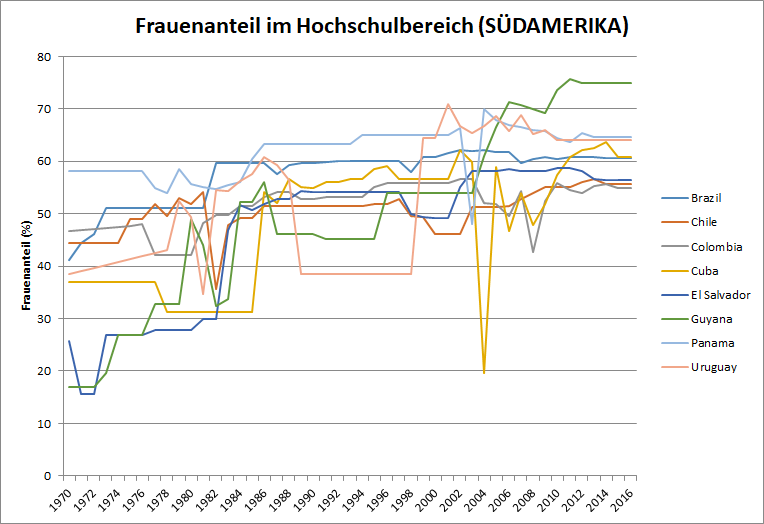
**Abbildung 7: Krise als Nebeneffekt**

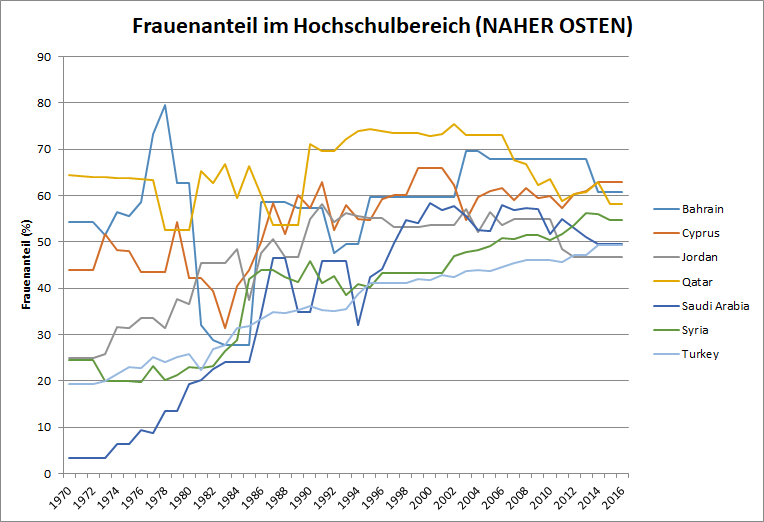
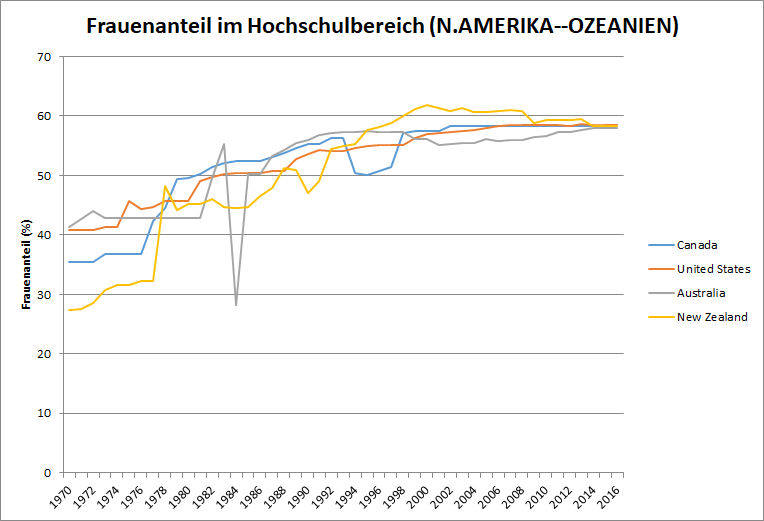


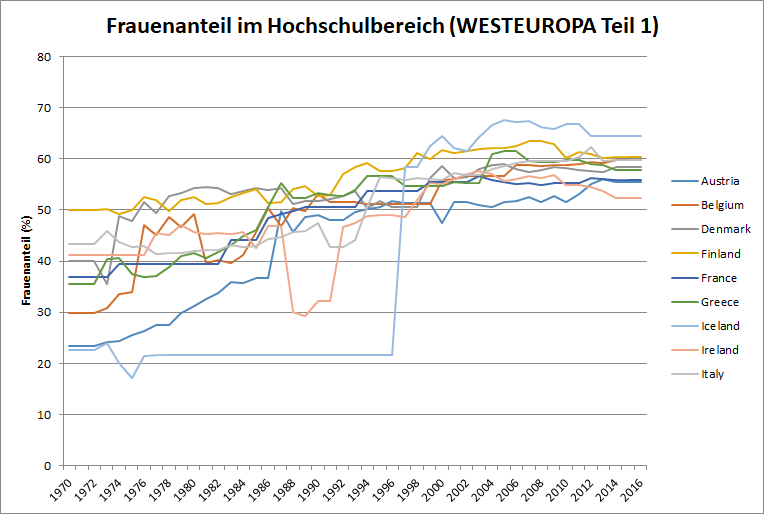
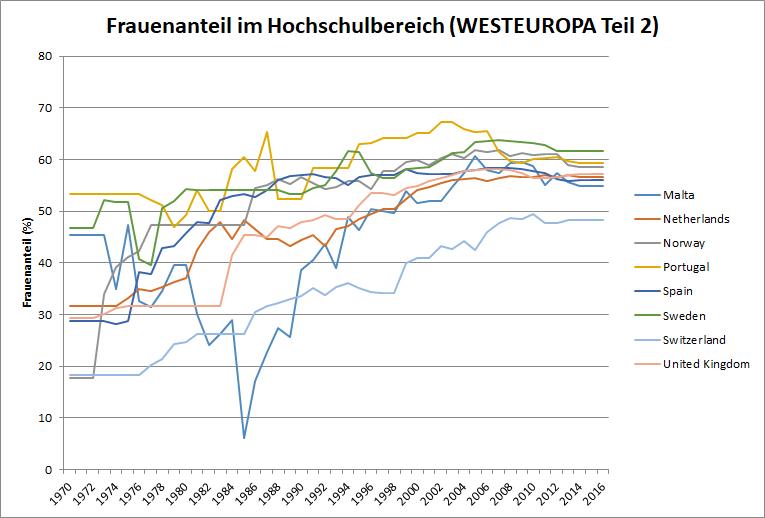
**Quelle: Eigene Darstellung**

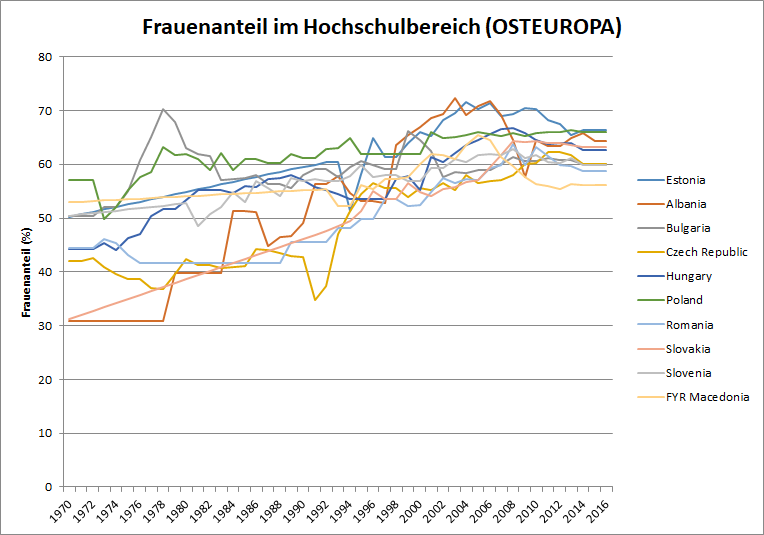
All diese Information herauszufiltern, wäre wahrscheinlich für das durchschnittliche menschliche Auge nicht möglich gewesen. Einen Test kann jeder Leser aufgrund der Abbildung8 selber durchführen. Abbildung 8 zeigt die Zeitreihen der Ländergruppen (im Fall von Westeuropa gibt es zwei Graphiken in der dritten Reihe).

**Abbildung 8: Primäre Daten der Ländergruppen**



**Quelle: Eigene Darstellung**

**5. Konklusionen**

Über sog. signifikante Unterschiede (in diesem Fall entlang der Zeit) kann man anhand von Trennungsstärken überzeugend sprechen. Ähnlichkeiten spielen bei der künstlichen Begriffsbildung (wie Betroffenheit) eine zentrale Rolle. Ein Turing-Test wäre angebracht um zu messen, wie hoch der Anteil der zufällig gewählten Menschen wäre, die die Lösung anhand der Trennungsstärken und der anti-diskriminative Modellierung als gleich rational betrachten würden, wie eine Schulbuchlösung (vgl. naive Bewertung).

Zum Schluss sollte noch auf eine direkt didaktische Art und Weise aufmerksam gemacht werden, dass all diese Analysen selbstverständlich dann notwendig und sinnvoll sind, wenn man annimmt, dass die Analyseergebnisse z.B. für Robot-Politiker weitergesendet werden, damit sie sofort und fundiert wissen: wo brannte es am ehesten. Außerdem können die Problem-Identifikationen auch für die Medien interessant sein. Statt fake-news zu produzieren, könnten in Zukunft Robot-Nachrichtendienste erstellt werden: Robuste Nachrichteninhalte zur Entscheidungsunterstützung lassen sich also durch Big-Data-Lösungen (automatisierbar) ableiten. Das Konzept Industrie 4.0 bedeutet also für die Forschung, für die Medien und schließlich für die Politik wichtige Impulse im Sinne von KNUTH (1992 – wo über KI noch kaum die Rede sein konnte!

### **VERWENDETE LITERATUR**

Enyedi, Gy. (2001): Településformáló folyamatok. – Educatio, 10. 2001. 4. 663–669.

Knuth, D. E. (1992): Literate Programming, Stanford LCSI – durch SYI, 2007: Egyben az egész.

Pitlik, L. (2015a): Tudomány-reform, MIAÚ, <http://miau.gau.hu/miau/208/20151120.pptx>

Pitlik, L. (2015b): Automatikus idősor-szegmentálás hasonlóságelemzéssel, MIAÚ, <http://miau.gau.hu/miau/200/szakaszolas.doc>

Pitlik, L. (2017): Rosling-vizualizáció multikulturális jelenségek esetén, <http://miau.gau.hu/miau/225/multikulti_3_pitlik_2017.docx>

Pitlik, L. – Pitlik. L. (2017): Frauenanteil im Hochschulbereich (eine komplexe Analyse von internationalen Daten unter dem Gesichtspunkt der Regionalität und von Zeitreihen), PPT, MIAÚ, <http://miau.gau.hu/miau/229/enyedi2/Unesco_Enyedi_pL_PL.ppt>

Pitlik, L. – Pitlik. M. (2016/2017): Online Cutting Services, MIAÚ, <http://miau.gau.hu/miau2009/index.php3?x=e0&string=cutt>

Varga, Z. (2014): Integrálódik-e az EU? <http://real.mtak.hu/39021/1/Studia_Mundi_Vol_2_No_3_14_Varga.pdf>