Version vom 16. Januar 2024

Anzahl Datenpunkte vergleichen und Chi-Werte berechnen

Inhalt

1.	Nuller rausfiltern	. 2
2.	Die Anzahl des einen Layers zum anderen hinzufügen (via Join)	. 2
3.	Normalisierung	. 3
4.	Chi-Wert berechnen	. 5
5.	Gitternetz nach Chi-Werten einfärben	. 6

Anzahl Datenpunkte vergleichen und Chi-Werte¹ berechnen

Wenn wir verschiedene Punktesets zählen (z. B. Ortsnamen pro Hexagon und Wald-Ortsnamen pro Hexagon), können wir die Dichten vergleichen. Z. B. können wir den Chi-Wert (<u>Signed Chi value</u>) berechnen, um eine erwartete Verteilung (basierend auf allen Ortsnamen) mit der aktuellen Verteilung (Wald-Ortsnamen) zu vergleichen. Danach können wir sehen, wo die Wald-Ortsnamen über- und wo sie untervertreten sind im Vergleich zu allen Ortsnamen in demselben Hexagon. Dafür bringen wir die Anzahl Wald-Ortsnamen und die Anzahl aller Ortsnamen pro Hexagon in die gleiche Attributtabelle und berechnen den Chi-Wert in einem neuen Feld in der Attributtabelle.

Wir brauchen also zwei Mal einen Layer mit Anzahl Punkten für den Vergleich. Falls ihr erst einen Layer mit Counts habt, müsst ihr noch einen zweiten erstellen mit der Anzahl Wald-Ortsnamen resp. der Anzahl Ortsname pro Gitterzelle (siehe dazu Abschnitt **Error! Reference source not found.**). Ihr müsst die Zellen für die Berechnung der Chi-Werte nicht anders einfärben. Wichtig ist, dass auf dem gleichen Grid gezählt wird und dann jeweils ein neuer Layer erstellt wird.

1. Nuller rausfiltern

Für die Berechnung des Chi-Wertes wählen wir nur diejenigen Hexagone aus, in denen die Anzahl Ortsnamen grösser als 0 ist, weil sonst die Formel nicht funktioniert (da Null im Nenner, siehe Formel weiter unten). Wir öffnen die **Attributtabelle** des Layers mit der Anzahl aller Ortsnamen (Rechtsklick auf den Layer \rightarrow Open Attribute Table) und filtern nach NUMPOINTS (die Spalte, in der wir die Punkte aufsummieren) > 0. Wir klicken auf **Select Features**, worauf die ausgewählten Hexagone gelb markiert werden.

Q count_alle_Ortsnamen_per_grid	cell — Features T	otal: 153, Filtered: 153, Selected: 106	—	\Box \times
/ 🐹 🗟 🕄 🐂 🖮 🗠 🔯 🔯	ء 📃 🗧 🗣	🕇 🛎 🐥 🔎 i 🕼 🕼 🗶 🛗 i 🔍 📾		
$\epsilon_{ m p}$ Expression	•			
□ 1	▲ id			Exclude Field
2	left			Exclude Field
3	top			Exclude Field
	right			Evelude Field
L 5	rignu			
7	bottom			Exclude Field
8	Anz_alle_O	0		Greater than (>)
9				
□ 10	Reset Form	Elach Features Show in Table Zoom to Features	act Features	Filter Features
🔍 🔍 🕨 🕪 1 / 153 🂡 💠 🎉				The readines +
🛅 Show All Features 🖕				

Abbildung 1: Alle Zeilen, d.h. Zellen, mit 0 Ortsnamen werden rausgefiltert (Quelle: Screenshot QGIS).

Wir speichern den Layer neu ab mit einem Rechtsklick \rightarrow Export \rightarrow Save Selected Features As...

Dasselbe machen wir für den Layer mit den Wald-Ortsnamen.

2. Die Anzahl des einen Layers zum anderen hinzufügen (via Join)

Wir brauchen die Anzahl aller Ortsnamen pro Gitterzelle in der gleichen Attributtabelle wie die Wald-Ortsnamen sind. Um sie anzuhängen, machen wir einen Join: Ein Rechtsklick auf den Layer der Wald-

¹ J. Wood, J. Dykes, A. Slingsby and K. Clarke, "Interactive Visual Exploration of a Large Spatio-temporal Dataset: Reflections on a Geovisualization Mashup.," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 6, pp. 1176-1183, Nov.-Dec. 2007, doi: 10.1109/TVCG.2007.70570.

Ortsnamen \rightarrow Properties. Dann klicken wir in der Auswahl links auf **Joins.** Wir klicken auf das grüne **Plus**-Symbol, um einen neuen Join zu machen. Im neuen Fenster wählen wir folgendes:

Male die id.

möchten.

Join Layer: Layer mit allen Ortsnamen

Join und Target field: Wir wählen hier beide

Ortsnamen in unseren Wald-Layer übertragen

Ortsnamen-Layer ein neues Feld hinzugefügt, das die Anzahl aller Ortsnamen für eine

Joined fields: Das Feld, das wir von allen

Stimmt die id überein, wird im Wald-

Wir klicken auf **OK**. Dann auf **Apply** und nochmals auf **OK**. Der Join ist nun bei den

bestimmte Gitterzelle enthält.

Properties aufgeführt.

🔇 Add Vector Join	×
Join layer	Count_alle_Ortsnamen_per_gridcell
Join field	1.2 id
Target field	1.2 id
✔ Cache join layer in men	nory
Create attribute index o	n join field
Dynamic form	
Editable join layer	
▼ ✓ Joined fields	
☐ id ☐ left ☐ top ☐ right ☐ bottom ✔ Anz_alle_O	
Custom field <u>n</u> ame	prefix

Abbildung 2: Wir joinen die Spalte der Anzahl aller Ortsnamen zur Attributtabelle mit den Wald-Ortsnamen (Quelle: Screenshot QGIS).

3. Normalisierung

Da die beiden Verteilungen eine unterschiedliche Anzahl haben, müssen wir die Formel für den Chi-Wert (siehe weiter unten) normalisieren, indem wir herausfinden, wie viele Werte wir bei beiden Layern total haben. Dazu klicken wir oben auf **Vector** → **Analysis Tools** → **Basic Statistics for Fields**.



Abbildung 3: Die Summe schlagen wir in der Statistik nach (Quelle: Screenshot QGIS).

Darauf öffnet sich das folgende Fenster:

Q Basic Statistics for Fields	×
Parameters Log Input layer Count_alle_Ortsnamen_per_gridcell [EPSG:2056] Count_alle_Ortsnamen_per_gridcell [EPSG:2	 Basic statistics for fields This algorithm generates basic statistics from the analysis of a values in a field in the attribute table of a vector layer. Numeric, date, time and string fields are supported. The statistics returned will depend on the field type. Statistics are generated as an HTML file.
Statistics [optional] [Save to temporary file]	Cancel
Advanced Run as Batch Process	Run Close Help

Abbildung 4: Eingaben für die Statistik (Quelle: Screenshot QGIS).

Wir geben den entsprechenden **Input Layer** an (derjenige mit den Counts von allen Ortsnamen) und das **Feld**, von dem wir die Statistik möchten. Danach klicken wir auf **Run**.

Im Log-File, das herauskommt, interessiert uns nur die Summe:

🔇 Basic Statis	tics for	Fields										\times
Parameters	Log						•	Basic sta	ntist	ics f	or fie	lds
Execution comp Results: {'COUNT': 1 'CV': 1.081 'EMPTY': 0, 'FILLED': 1 'FIRSTQUART 'IQR': 3654 'MAJORITY': 'MAX': 8773 'MEAN': 212 'MEDIAN': 1 'MIN': 0.0, 'MINORITY': 'OUTPUT_HTML 'MIN': 0.0, 'MINORITY': 'OUTPUT_HTML 'RANGE': 87' 'STD_DEV': ' 'STD_DEV': 1 'SUM': 3252 'THIRDQUART 'UNIQUE': 1 Loading resultin Algorithm 'Basic HTML output ha Open the result	leted in (53, 886291: 0, 0, 0, 5,9150: 558.0, 3.0, L_FILE 73.0, 2299.9 65.0, ILE': : 03} g layers s dialog t	0.07 seconds 976641, 0.0, 326797385, ': 'C:/User /ac3a342b2d ntml', 98331763282 3654.0, s for fields' fini- enerated by thi to check it.	s/ksoland/App 34415a86f15b8 , shed s algorithm.	Data/Local/1 ae915e2ac/	Cemp/			This algorithm from the analy the attribute ta Numeric, date, supported. The statistics r field type. Statistics are g	gener sis of a ble of time a eturne enera	ates ba a values a vecto and strii ed will d ted as a	sic statist s in a field r layer. ng fields epend on an HTML 1	tics d in are n the file.
[
				0%		_		1			Can	icel
Advanced 🔻 🖡	Run as Ba	tch Process				Cha	ing	ge Parameters	Cl	ose	He	lp

Abbildung 5: Statistik-Output, der sich auch speichern lässt bei Bedarf (Klick auf die Diskette oder mit Copy-Paste) (Quelle: Screenshot QGIS).

Wir schreiben uns diese auf, im Beispiel sind es 325'265. Dasselbe machen wir für den Layer mit der Anzahl Wald-Ortsnamen. Dort ist die Summe 9127.

4. Chi-Wert berechnen

Um wen Chi-Wert zu berechnen, öffnen wir die Attributtabelle des Layer mit der Anzahl Wald-Ortsnamen pro Gitterzelle. Wir wählen den **Field Calculator** aus. Das Icon befindet sich in der

Attributtabelle oben auf der rechten Seite:

Q count_wald_points_groesser0 — Field Calculator		×
Only update 0 selected features		
✓ Create a new field Update existin	ng field	
Create virtual field		
Output field name chi		
Output field type 1.2 Decimal number (real)		*
Output field length 10 Precision 3		
Expression Function Editor		
	Q Search Show Values	group field
<pre>(((325265 / 9127) * "NUMPOINTS") - "count_alle_Ortsnamen_groesser0_Anz_alle_O") / ("count_alle_Ortsnamen_groesser0_Anz_alle_O" ^ 0.5) = + - // * ^ () "n" Feature 100</pre>	 Color Conditionals Conversions Date and Time Fields and Values feature geometry id NULL 1.2 id 1.2 left 1.2 right 1.2 bottom 1.2 NUMPOINTS 	Double-click to add field name to expression string. Right-Click on field name to open context menu sample value loading options. Notes Values Search All Unique 10 Samples
Provinu -21 142217606150716	1.2 count alle Ortsnamen groesser0 A	
You are editing information on this layer but the layer is currently	y not in edit mode. If you click OK, edit mode will automatically	be turned on.

Abbildung 6: Field Calculator Eingaben, um den Chi-Wert zu berechnen (Quelle: Screenshot QGIS).

Wir erstellen ein neues Feld für den Chi-Wert (das Häkchen ist standardmässig so gesetzt, dass ein neues Feld erstellt wird). Wir geben dem Feld einen Namen bei **Output field name** (z. B. Chi) und setzen den **Output field type** auf **1.2 Decimal number (real).** Unten links im Feld **Expression** können wir nun die Formel für den Chi-Wert eingeben. Diese lautet:

Chi = (beobachtete Häufigkeit – erwartete Häufigkeit) / $\sqrt{}$ erwartete Häufigkeit,

(Quelle: https://ieeexplore.ieee.org/abstract/document/4376138)

wobei die beobachtete Häufigkeit die Anzahl ist, die uns interessiert ist (z. B. die Wald-Ortsnamen), und die erwartete Häufigkeit ist die zugrundeliegende Anzahl (z. B. alle Ortsnamen). Die beobachtete Häufigkeit normalisieren wir noch mit den beiden Summen, die wir grad vorher rausgesucht haben. Die finale Formel sieht dann so aus:

(((325265 / 9127) * Anz_Wald_Ortsnamen) - Anz_alle_Ortsnamen) /

\sqrt{Anz} alle_Ortsnamen

Die Zahlen tippen wir direkt ins weisse Feld. Die mathematischen Operatoren wählen wir unter dem weissen Feld aus. Die Attribute (z. B. NUMPOINTS) können wir in der Liste in der Mitte unter **Fields** and **Values** auswählen, indem wir sie zweimal anklicken.

Nachdem wir OK geklickt haben, sehen wir in der Attributtabelle die neue Spalte mit den Chi-

Werten. Mit einem Klick auf die kleine Diskette mit dem roten Stift (wir unsere neuen Daten in der Tabelle speichern.

5. Gitternetz nach Chi-Werten einfärben

Ein negativer Chi-Wert bedeutet, dass die Wald-Ortsnamen in dieser Gitterzelle untervertreten sind, während ein positiver Werte auf eine Übervertretung hinweist. Wir können das nun noch farblich darstellen, indem wir die Zellen aufgrund der Chi-Werte einfärben: Rechtsklick auf den Layer \rightarrow **Properties** \rightarrow **Symbology**. Wir wählen die gelb markierten Optionen:

Q	Layer Properties	s — count_wald_points_groesser0 — Symbology	\times
Q		Graduated	Ŧ
i	Information	Value 1.2 chi	3 -
с С	Source	Symbol	•
~	Symbology	Legend format %1 - %2	🕙 🗘 🗸 Trim
abc	Labels	Color ramp	•
abc	Masks	Classes Histogram	
Ŷ	3D View	Symbol Values Legend	
۹.	Diagrams	-55.75 - 27.88 -56 - 28	
	Fields	 ✓ -27.88 - 0.00 - 28 - 0 ✓ -0.00 - 27.88 0 - 28 → 27.09 - 57.58 0 - 28 	
-8	Attributes Form	✓ 27.88 - 55.75 - 182.87 56 - 183	
•	Joins	Mode Equal Interval	asses 6
đ	Auxiliary	▼ ✓ Symmetric Classification	
٩	Actions	Create class astride symmetry value	•
—	Display	Classify 🖶 📼 Delete All	Advanced 🔻
*	Rendering	✓ Link class boundaries	
٩	Temporal	▼ Layer Rendering	
3	Variables	Opacity	35.7 % 🖾 🜲
1	Elevation	Layer Feature Blending mode Normal	•
2	Metadata	Draw effects	•
	Dependencies	Control feature rendering order	₹↓
÷	Legend	▼ Style ▼ OK Cancel Apply	Help

Abbildung 7: Symbology für die Chi-Werte. Die Null wird so gesetzt, dass sie farblich zwischen grün und orange liegt (Quelle: Screenshot QGIS).

Die symmetrische Klassifizierung ist nur beim Mode **Equal Interval** und einer geraden Anzahl Klassen möglich.

QGIS Tutorial: Neues Projekt erstellen

UZH/DSI GIS Hub



Die Karte, die daraus resultiert, mit der Legende zu den Chi-Werten:



Abbildung 8: Chi-Werte, die anzeigen, ob die Wald-Ortsnamen über- (rötliche Farbtöne) oder untervertreten (blau-grüne Farbtöne) sind im Vergleich zu allen Ortsnamen in der Schweiz (Quelle: Screenshot QGIS).