

QGIS Tutorial

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 (Signed Chi value) 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 → 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.

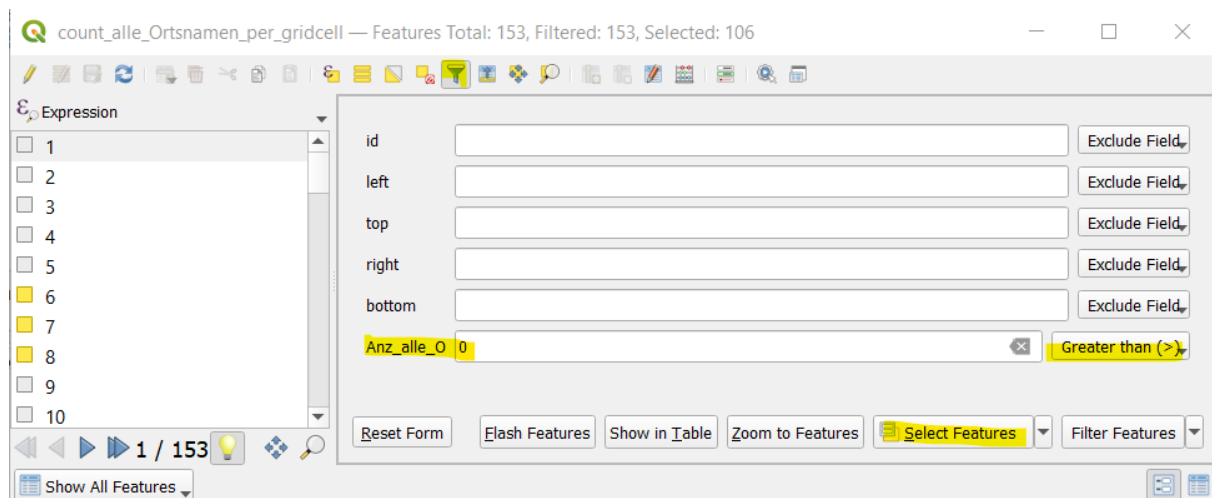


Abbildung 1: Alle Zeilen, d.h. Zellen, mit 0 Ortsnamen werden rausgefiltert (Quelle: Screenshot QGIS).

Wir speichern den Layer neu ab mit einem Rechtsklick → **Export** → **Save Selected Features As...**

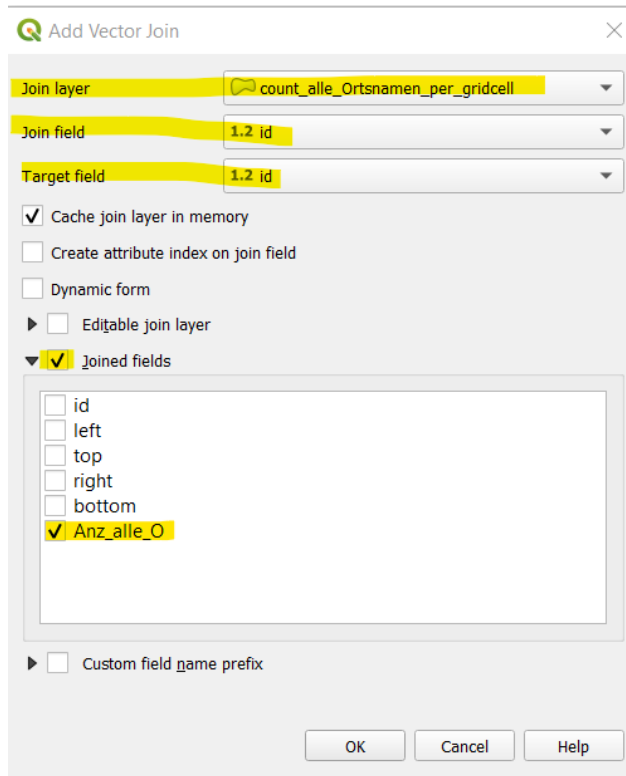
Dasselbe machen wir für den Layer mit den Wald-Ortsnamen.

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 → 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:



Join Layer: Layer mit allen Ortsnamen

Join und Target field: Wir wählen hier beide Male die id.

Joined fields: Das Feld, das wir von allen Ortsnamen in unseren Wald-Layer übertragen möchten.

Stimmt die id überein, wird im Wald-Ortsnamen-Layer ein neues Feld hinzugefügt, das die Anzahl aller Ortsnamen für eine bestimmte Gitterzelle enthält.

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

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**.

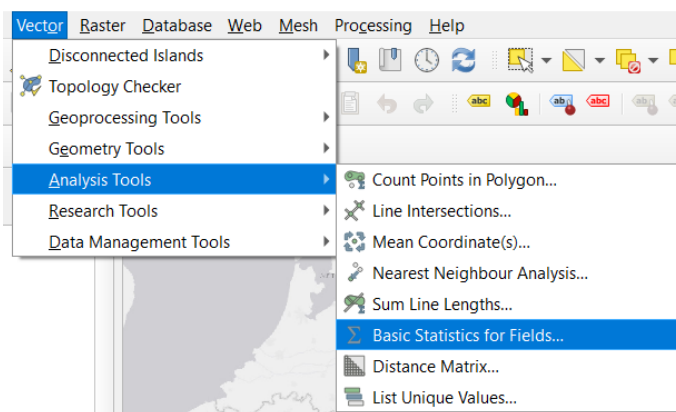


Abbildung 3: Die Summe schlagen wir in der Statistik nach (Quelle: Screenshot QGIS).

Darauf öffnet sich das folgende Fenster:

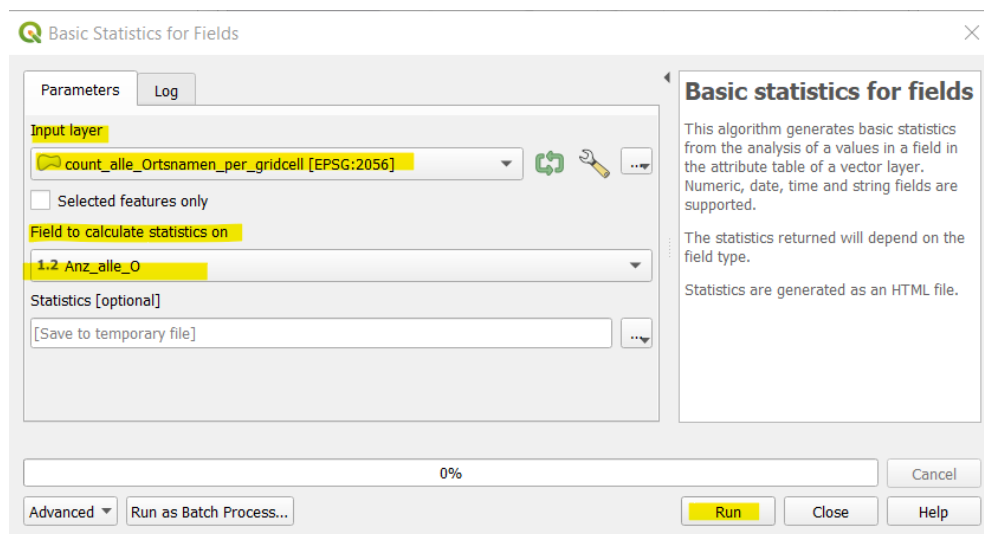


Abbildung 4: Eingaben für die Statistik (Quelle: Screenshot QGIS).

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:

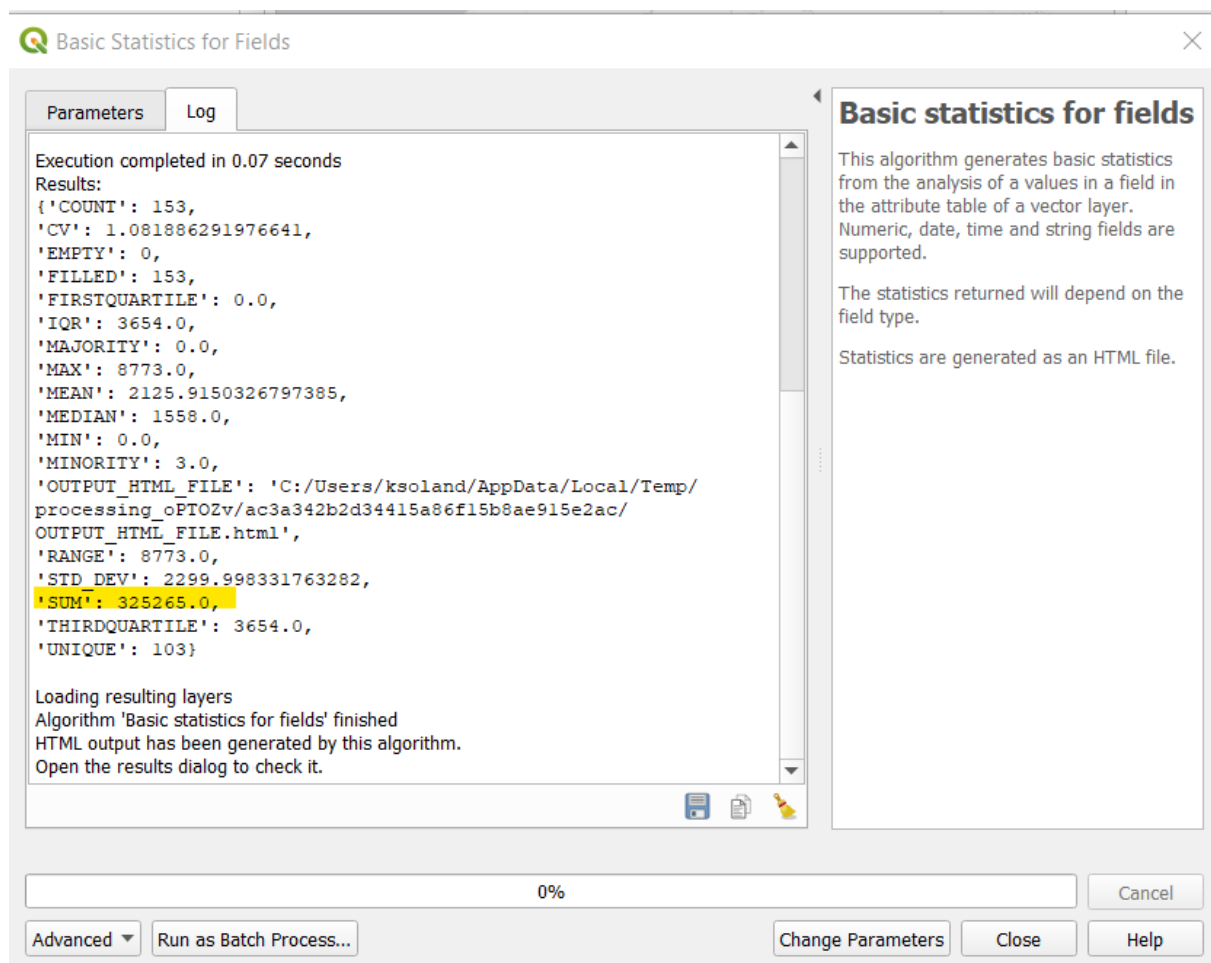



Abbildung 5: Statistik-Output, der sich auch speichern lässt bei Bedarf (Klick auf die Diskette oder mit Copy-Paste) (Quelle: Screenshot QGIS).

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: 

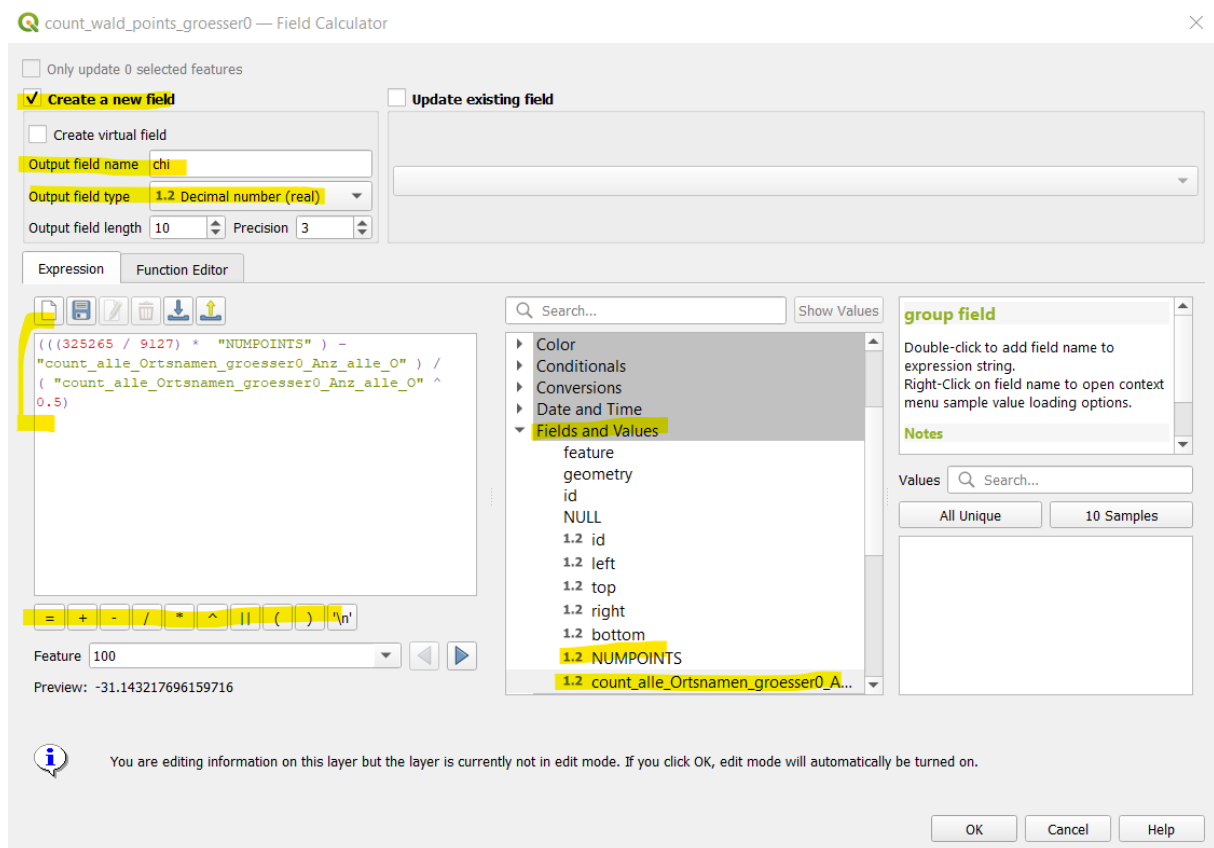


Abbildung 6: Field Calculator Eingaben, um den Chi-Wert zu berechnen (Quelle: Screenshot QGIS).

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:

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


(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:

$$\frac{(((325265 / 9127) * \text{Anz_Wald_Ortsnamen}) - \text{Anz_alle_Ortsnamen})}{\sqrt{\text{Anz_alle_Ortsnamen}}}$$

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 ( 3. Icon von links) können 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 → **Properties** → **Symbology**. Wir wählen die gelb markierten Optionen:

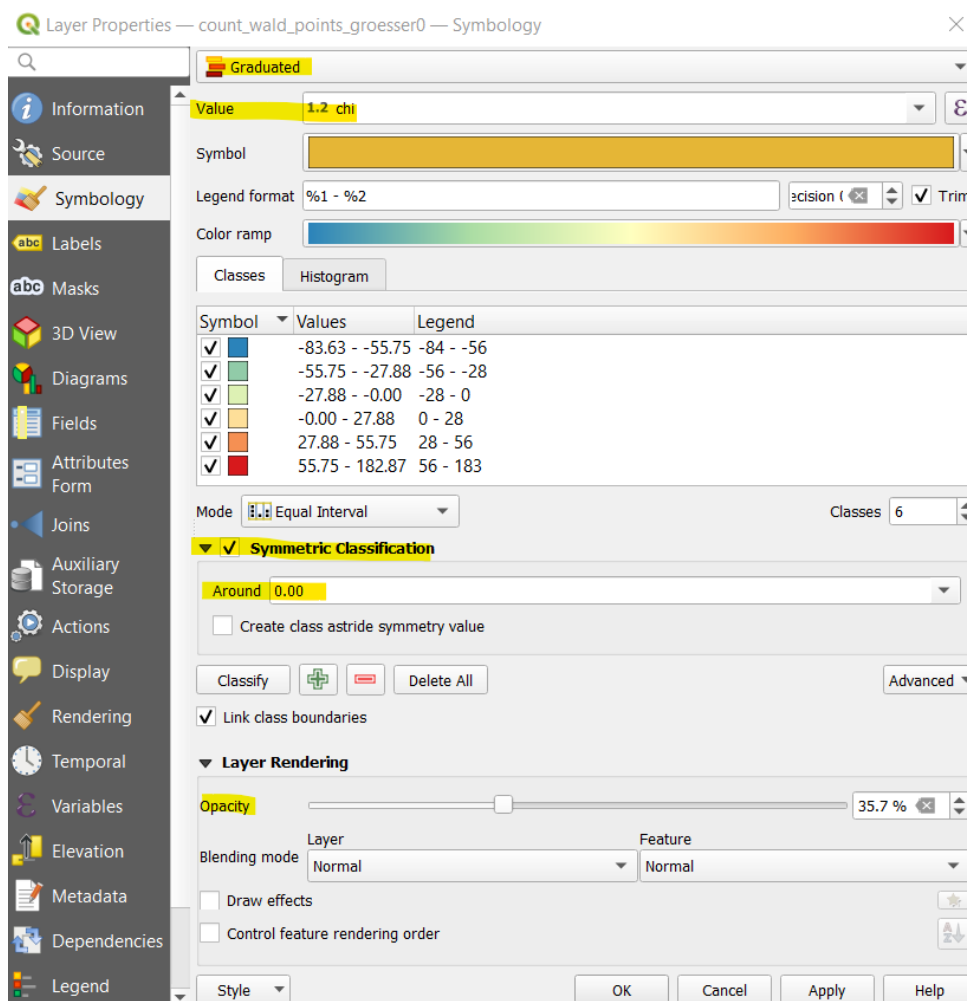
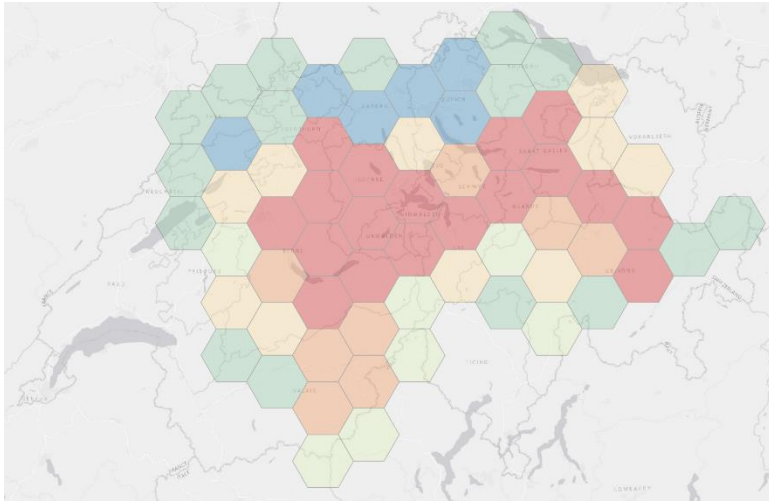


Abbildung 7: Symbology für die Chi-Werte. Die Null wird so gesetzt, dass sie farblich zwischen grün und orange liegt (Quelle: Screenshot QGIS).

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



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

- 84 - -56
- 56 - -28
- 28 - 0
- 0 - 28
- 28 - 56
- 56 - 183

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).