

OpenSchoolMaps: Joining tables

OpenSchoolMaps.ch — Freie Lernmaterialien zu freien Geodaten und Karten

Einleitung

Ziele

Die Ziele dieses Arbeitsblattes sind:

- Verstehen, was das Verknüpfen ("Joinen") von zwei Tabellen ist und wie das mit QGIS geht.
- Die Beziehungen zwischen zwei Tabellen wie **one-to-one**, **one-to-many** und **many-to-many** verstehen.
- Einfache numerische Klassifizierungsalgorithmen kennen und wissen, wie und wann man diese im Desktop-Programm QGIS einsetzen kann.

Voraussetzungen

Um das Arbeitsblatt durcharbeiten zu können, brauch man folgende Dinge:

- Software: QGIS (getestet mit QGIS 3.4, 64-bit).
- Internetzugang zum Herunterladen von Daten und zur eventuellen Installation von QGIS ([download](#)).
- Daten: [Joining Tables Aufgabenblatt Daten](#).

In den Daten sollten folgende Dateien zu finden sein:

- `World_Countries.gpkg`
- `CO2_Readings.xlsx`
- `Worldbank_Countries.xlsx`
- `RenewableEnergy_Percentages.csv` und `RenewableEnergy_Percentages.dbf`



Wenn du bisher noch keine Erfahrung mit QGIS hast, kannst du dir unsere Arbeitsblätter auf [OpenSchoolMaps](#) zum Thema "Einführung in QGIS and in Geoinformationssysteme (GIS)" anschauen.

Einführung

Strukturierte Daten werden in einer zusammengesetzten Datenstruktur modelliert. Strukturierte

Daten sind oft in Tabellen organisiert, d.h. in tabellarischer Form. In einem tabellarischen Datenmodell gibt es **Columns** (auch bekannt als: Spalten, Felder, Attribute, Zellen) und **Rows** (auch bekannt als: Datensätze, Zeilen, Einträge). Diese Spalten haben auch einen sogenannten Datentyp, der den Inhalt der Spalte näher beschreibt.

Eine Adressliste ist ein gängiges Beispiel für *strukturierte Daten* in tabellarischer Form, eine Tabelle, die aus Spalten wie z.B. der Postleitzahl besteht, die vom Typ Ganzzahl ist (zumindest in einigen Ländern).

Im Gegensatz dazu gibt es *unstrukturierte Daten*, z.B. ein Fliertextdokument, das nicht wie eine Tabelle organisiert ist (nur eine lange Zeichenkette).

Zwei Tabellen können miteinander verknüpft werden, wenn beide mindestens eine Spalte enthalten, welche ähnliche Daten enthalten. Diese Spalten werden auch als *Key* bezeichnet, die Beziehungen zwischen verschiedenen Tabellen herstellen können. Bekannte Beispiele für *Keys* sind Postleitzahlen oder Autokennzeichen, da diese bereits in ihrer realen Form als eindeutige Identifikatoren dienen. Die Tabellen können weiterhin getrennt voneinander geändert und angezeigt werden. Man stellt sich das als einen Vertrag vor, der zwischen zwei Tabellen geschlossen wird. Eine Spalte in jeder der beiden Tabellen gibt genau an, auf welche Zeile sie verweist. Wenn Daten aus einer Tabelle abgerufen werden, liefert die verbundene Tabelle ihre Daten gemäss diesem Vertrag.



Keys sind Bezeichner, die auf bestimmte Einträge in einer Tabelle verweisen. Aus diesem Grund werden sie zur Verknüpfung von Tabellen verwendet. Der am häufigsten verwendete *Key* ist wahrscheinlich eine *id*, die jedem Eintrag eine eindeutige Nummer gibt.

Dann gibt es noch den Sonderfall, dass der Link über ein besonders aktiviertes Attribut verläuft. In diesem Fall handelt es sich um Geodaten. Geodaten sind Daten, die mindestens ein räumliches Attribut haben. Ein solches räumliches Attribut enthält in der Regel einen Punkt, eine Linie oder ein Polygon, es kann aber auch eine Postleitzahl, ein geografischer Name, ein Geohash-Code oder ein anderer geografischer Code sein.

Klassifikation

Die Klassifizierung ist ein wichtiges Hilfsmittel für das Verständnis frisch gewonnener Daten. Informationen allein reichen nicht aus, um Daten zu interpretieren. Deshalb klassifizieren wir Daten.

Datentypen

Ein Datentyp bestimmt den zulässigen Wertebereich eines Attributs. Beispiele für Datentypen sind **text** oder **integer**. Und ein typischer Wertebereich einer vierstelligen Ganzzahl - wie etwa eine Schweizer Postleitzahl, die von **1000** bis **9999** reicht.

Die folgenden Datentypen sind typischerweise für die Verknüpfung geeignet, aber natürlich gibt es noch mehr Möglichkeiten:

- **Ganzzahl (Integer)**. Ganzzahlen werden oft zur Verknüpfung von Tabellen verwendet. Sie können zum Beispiel als ID verwendet werden.

- **String**. Auch Texte können zur Verknüpfung von Tabellen verwendet werden. Sie können zum Beispiel einen Namen als Schlüsselwert verwenden.
- **Date** (z.B. Datum ISO-8601 wie 2022-04-01). Ein Tagesdatum bietet sich auch selbst als Schlüssel für die Verknüpfung an. Auf diese Weise können Sie Daten nach Zeit aufschlüsseln.

Zwei Tabellen verknüpfen ("joinen")

Bei der Verknüpfung von zwei Tabellen will man Daten aus einer Tabelle mit passenden Daten aus einer anderen Tabelle verbinden. Zu diesem Zweck definiert man eine Verbindung zwischen den beiden Tabellen, die aus jeweils einer Spalte der beiden Tabellen besteht. Diese gleiche Spalte muss nicht den gleichen Namen haben, aber sie muss den gleichen Datentyp haben. Durch die Verknüpfung können neue Erkenntnisse über die ursprünglichen Daten gewonnen werden.

Im folgenden Beispiel können die beiden Tabellen verknüpft bzw. verbunden werden, weil sie das gemeinsame Attribut **country** haben. Wir können diese Spalte als Primärschlüssel verwenden. In diesem Fall ist der Datentyp des Schlüssels, der zur Verknüpfung der beiden Tabellen verwendet wird, ein **string**.

Beziehungen

Country	Prefix	Country	Population
Switzerland	+41	Switzerland	8'637'000
France	+33	France	67'390'000
Germany	+49	Germany	83'240'000

Abbildung 1. Diese beiden Tabellen können über **country** verbunden werden, da dies ein gemeinsames Attribut der beiden Tabellen ist.

Zwischen zwei Tabellen können die Daten in einer der folgenden besonderen Beziehungen zueinander stehen:

- **one-to-one** (1:1)
- **one-to-many** (1:n)
- **many-to-many** (n:m)

One-to-One und **One-to-Many** werden später in der Übung verwendet werden.

One-to-one (1:1)

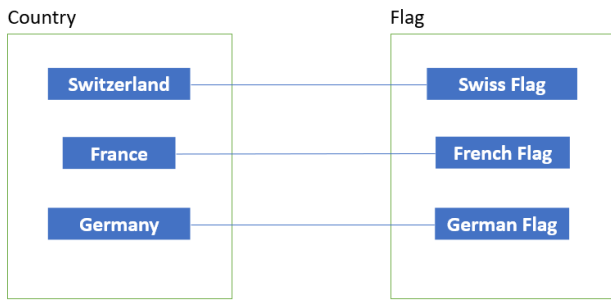


Abbildung 2. Jedes **Land** zeigt auf genau eine **Flagge** und umgekehrt.

One-to-one bedeutet, dass jede Zeile in der ersten Tabelle mit einer einzigen entsprechenden Zeile in der zweiten Tabelle verknüpft ist. Zum Beispiel hat jedes Land eine Flagge, die es repräsentiert. Im Gegenzug ist jede Flagge mit einem bestimmten Land verknüpft.

One-to-many (1:n)

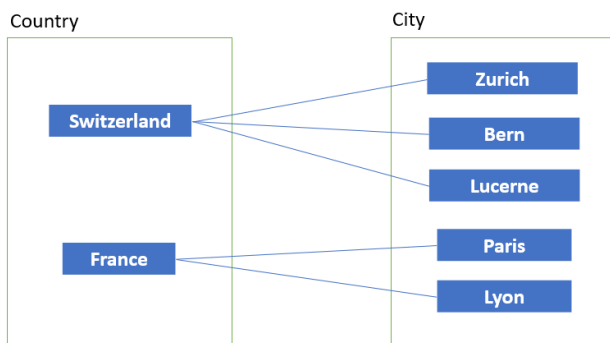


Abbildung 3. Ein **Land** kann auf mehrere **Städte** verweisen, aber eine **Stadt** kann nur auf ein **Land** verweisen.

Zum Beispiel kann jedes Land mehrere Landessprachen haben (z.B. die Schweiz). Umgekehrt kann eine Sprache die Landessprache mehrerer Länder sein (z.B. Französisch).



Es ist wichtig zu wissen, dass zwei gleichnamige Städte nicht als "dieselbe" Stadt gelten.

Many-to-many (n:m)

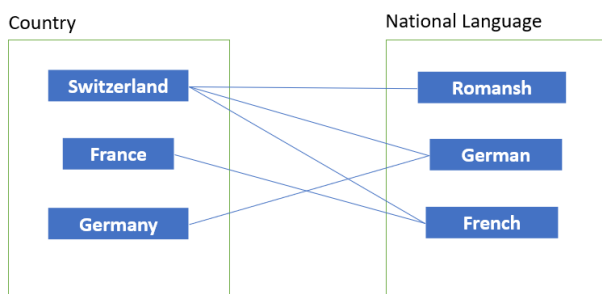


Abbildung 4. Ein **Land** kann auf mehr als eine **Nationalsprache** verweisen. Eine **Nationalsprache** kann auch auf mehrere **Länder** verweisen.

Terminologie

Im Folgenden werden einige Begriffe zum besseren Verständnis noch einmal erklärt:

- **Tabelle** - eine Tabelle ist eine Sammlung von zusammenhängenden Informationen, die in einem Format mit Spalten und Zeilen gespeichert werden. Tabellen sind Grundstrukturen von Datenbanken.
- **Attribut** - eine Komponente einer Tabelle, meist ist es das Feld, das ein bestimmtes Merkmal des Datensatzes beschreibt.
- **Record** - Zeile der Datentabelle.
- **Join Operation** - Prozess, bei dem die Felder einer Tabelle mit denen einer anderen Tabelle auf der Grundlage eines gemeinsamen Attributs verbunden werden.
- **Datenklassifizierung** - Prozess der Gruppierung von Datenreihen in sinnvolle Kategorien, um die Interpretation zu verbessern.

Übungen

In diesem Teil geht es darum, das Verbinden von Tabellen zu üben. Das Ziel ist es, weitere Attribute an eine bestehende Tabelle in der Raumdaten gespeichert sind, anzuhängen. Die neu erstellte Ebene kann gespeichert werden, wodurch eine neue Datei entsteht, oder sie kann verworfen werden, wenn sie nicht zufriedenstellend ist oder nicht weiter gebraucht wird.

Übung 1.1 - Informationssammlung über Join

In dieser Übung werden wir den Prozess der Verknüpfung von Tabellen durchgehen, um so viel Wissen wie möglich aus numerischen Mengen zu sammeln.

Folgende Daten werden für diese Übung benötigt:

- [World_Countries.gpkg](#)
- [CO2_Lesungen_Welt.xlsx](#)
- [Weltbank_Länder.xlsx](#)

Hier bestehen die numerischen Daten aus CO2-Emissionswerten pro Land, die in einer MS Excel-Tabelle gespeichert sind. Der Kontext wird durch eine andere Datei bereitgestellt, die eine Liste der Bevölkerung jedes Landes enthält. Das Projekt, mit dem wir arbeiten, hat eine Liste von Länder der Welt. Jede dieser Dateien enthält nur eine kleine Information. Nach der Zusammenführung dieser Datensätze liefern sie angereicherte und interpretierbare Daten. In diesem Fall wären das die CO2-Emissionen und die Bevölkerung aller Länder in einem räumlichen Format.

Die Bevölkerungszahl eines Landes kann eine Erklärung für seine hohen bzw. niedrigen CO2-Emissionen sein. Wenn wir mit der Erstellung dieser neuen, grösseren Tabelle fertig sind, gehen wir zum zweiten Teil dieser Übung über: der Datenklassifizierung. Dort werden wir die Daten in relevante Kategorien gruppieren, um die Interpretation noch weiter zu verbessern.

Heruntergeladene Daten können bereits eine räumliche Komponente enthalten, wie z.B. die Bevölkerungszahl in verschiedenen Städten, aber um die Daten in QGIS zu verwenden, müssen sie mit dem bereits vorhandenen Datensatz verbunden werden. An dieser Stelle bietet sich das Verbinden von Tabellen an, da beide Datendateien miteinander verbunden werden können, um mehr Informationen zu erhalten.

1. Öffne QGIS und wähle im Projekt **Project > New**, um ein neues QGIS-Projekt zu erstellen.
2. Füge die Datei **World_Countries.gpkg** zum Projekt hinzu, indem du sie von ihrem Speicherort auf den Ebenenbereich auf der linken Seite des QGIS Editors ziehst.

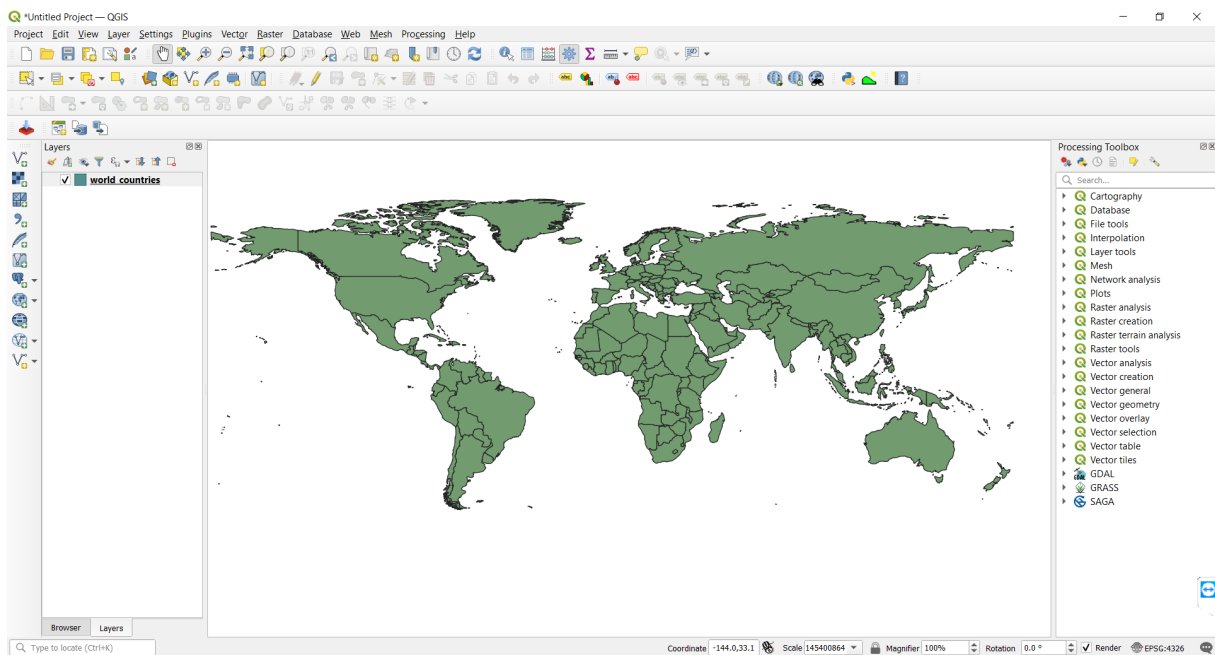


Abbildung 5. So sollte alles aussehen, nachdem **World_Countries.gpkg** importiert wurde.

3. Um die Attributtabelle der soeben importierten Ebene zu prüfen, klicke mit der rechten Maustaste darauf und wähle: **open attribute table**

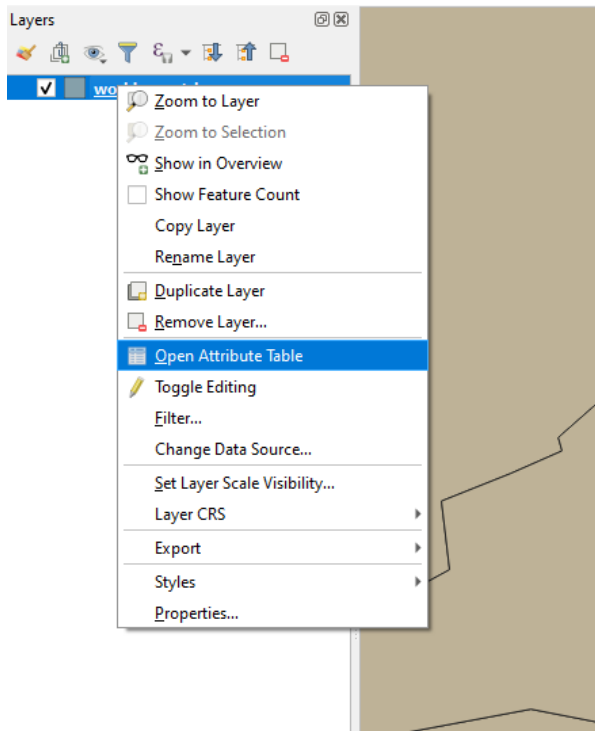


Abbildung 6. Kontextmenü zum Öffnen der Attributtabelle

Du solltest nun die Daten aus **World_Countries.gpkg** sehen können. Es sollte drei Spalten geben: **REGION**, **NAME_12** & **C_NAME**.

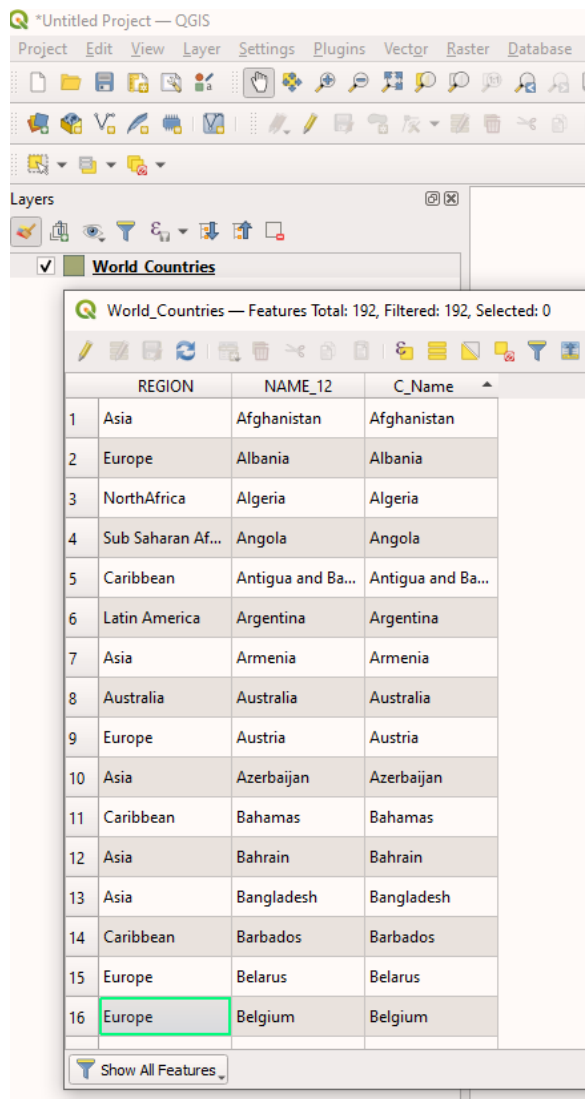


Abbildung 7. Attributtabelle, die die Daten aus `World_Countries.gpkg` anzeigt

4. Schliesse die Attributtabelle.
5. Jetzt können wir zusätzliche Daten zu den Ländern hinzufügen. Klicke auf: **Layers** > **Datenquellenmanager**

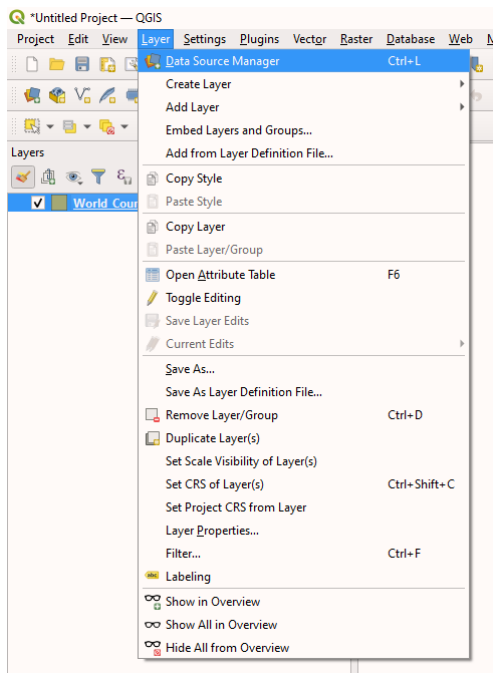


Abbildung 8. Kontextmenü zum Öffnen des **Data Source Manager**.

- Wähle im **Data source manager** die Option **Browser** in der Seitenleiste, um deine Ordnerstruktur anzuzeigen. Suche den Ordner, in dem du **CO2_Readings_World.xlsx** gespeichert hast:

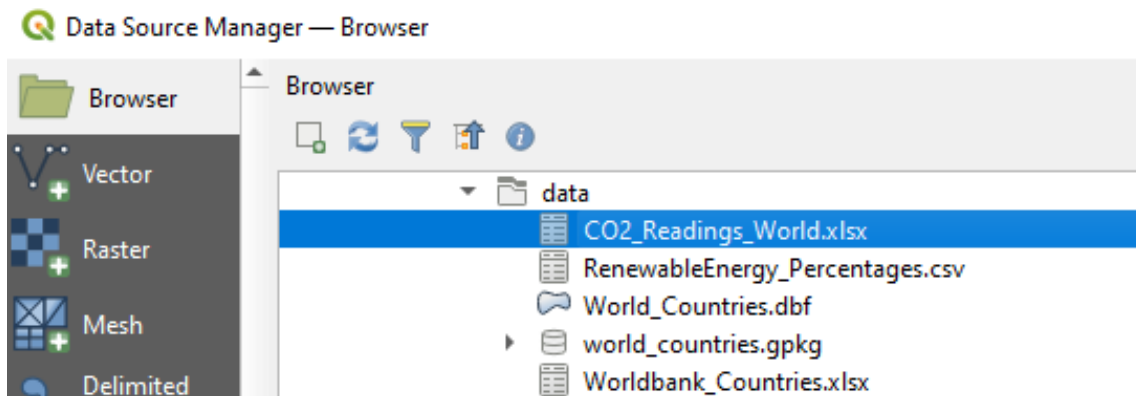


Abbildung 9. QGIS's Data Source Manager

Doppelklicke auf die Datei, um sie zu den Ebenen in deinem Projekt hinzuzufügen. Danach kannst du den Data Source Manager schließen.

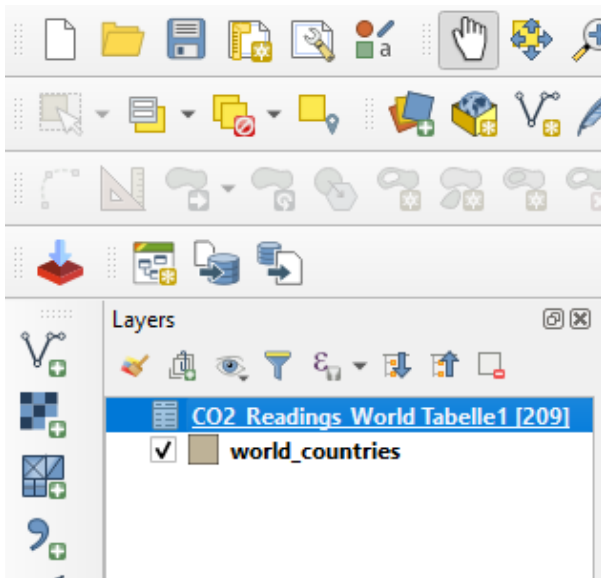


Abbildung 10. Die Datei `CO2_Readings_World.xlsx` wurde zur Erstellung einer neuen Ebene verwendet.

7. Klicke mit der rechten Maustaste auf die neu hinzugefügte Ebene und öffne deine Attribute. Erkunde die Daten, um dich mit ihnen vertraut zu machen.
8. Schliesse die Attributtabelle.
9. Klicke mit der rechten Maustaste auf das Fenster **World_Countries layer > Properties > Layer Properties**:

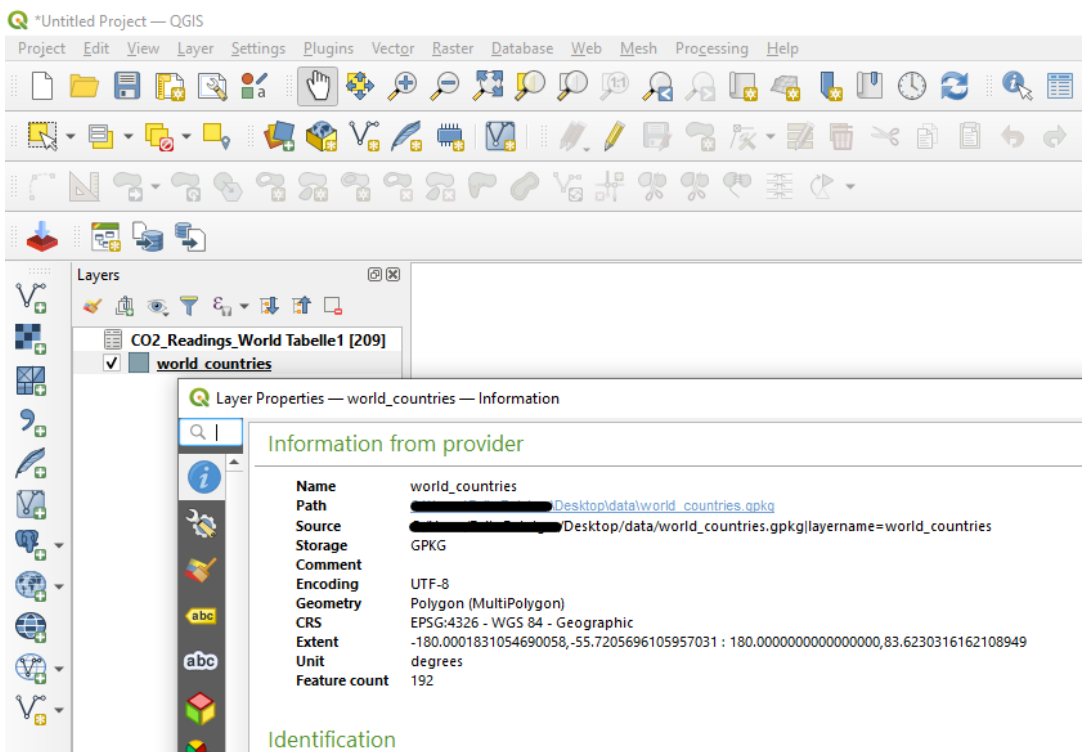


Abbildung 11. Fenster Ebeneneigenschaften.

10. Um die Joins für die Tabelle zu konfigurieren, wähle in der Seitenleiste **Joins**.



Abbildung 12. joins in der Seitenleiste der Ebeneneigenschaften

11. Klicke auf das grüne Symbol [+] in der unteren linken Ecke des Fensters "Join". Dadurch wird das Fenster **Add Vector Join** geöffnet.
12. Wähle das Attribut **C_Name** im Menü **Join verbinden** und **Target Field**. Aktiviere das Kontrollkästchen **Custom field name prefix** und füge den Typ "joined" in das Textfeld ein.

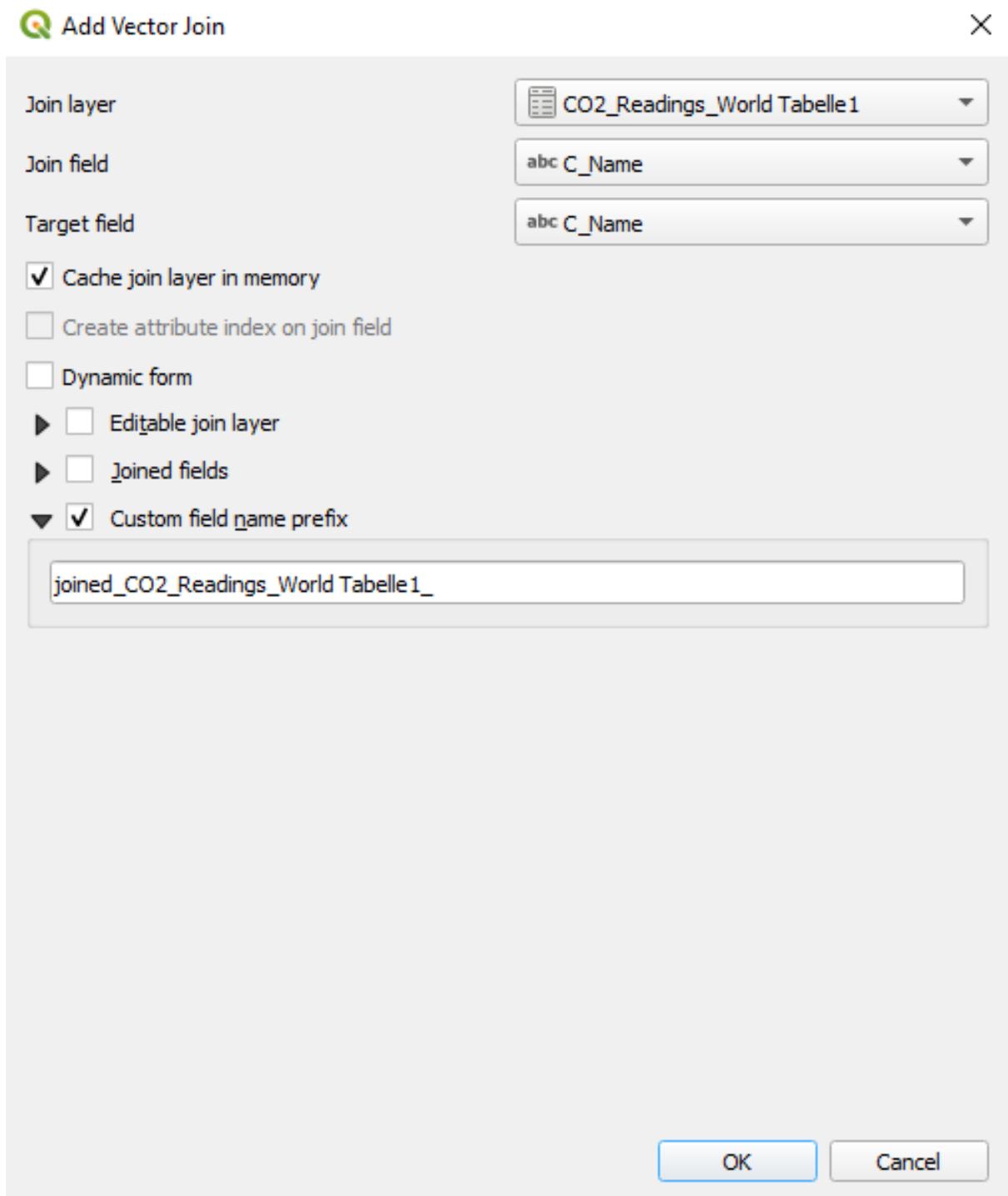


Abbildung 13. Alle Attribute wurden ausgewählt und der neue Name wurde vergeben.

Klicke auf [**OK**].

13. Erweitere die Liste, indem du auf das schwarze Dreieck neben **Ebene verbinden** klickst, um die Zusammenfassung der Operation zu sehen.

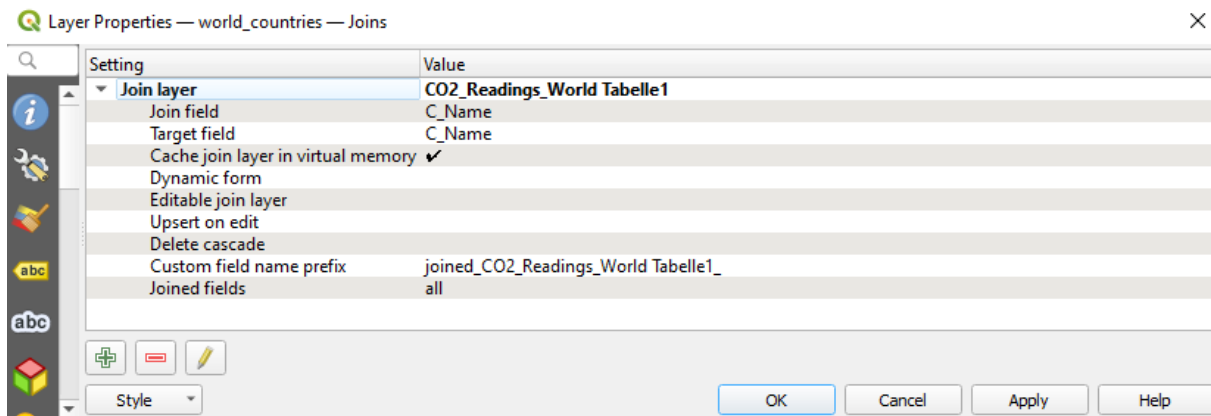


Abbildung 14. Die Zusammenfassung der Operation "Tabelle verbinden".

Klicke auf [OK].

14. Öffne erneut die Attributtabelle der Ebene World_Countries.

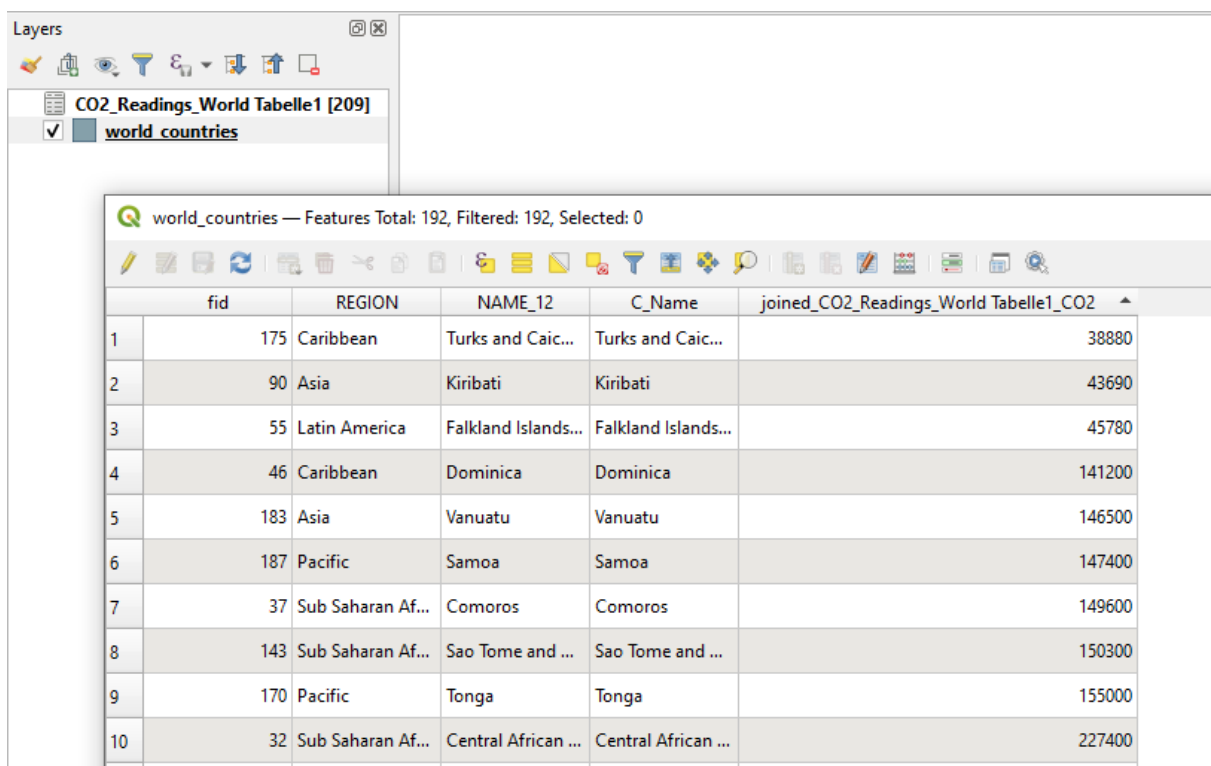


Abbildung 15. Neue verbundene Attributtabelle der Ebene World_Countries

15. Wiederhole die Schritte 5 - 13 und füge eine weitere Tabellenkalkulationsdatei: **Worldbank_Countries.xlsx** zum Projekt hinzu und verbinde die Tabellen unter Verwendung des Feldnamens als Verbindung.

16. Die neuen Tabellen werden vorübergehend im Projekt gespeichert. Um sie auf der Festplatte zu speichern, klicke auf **World_Countries layer > Export > Save Features as ...**

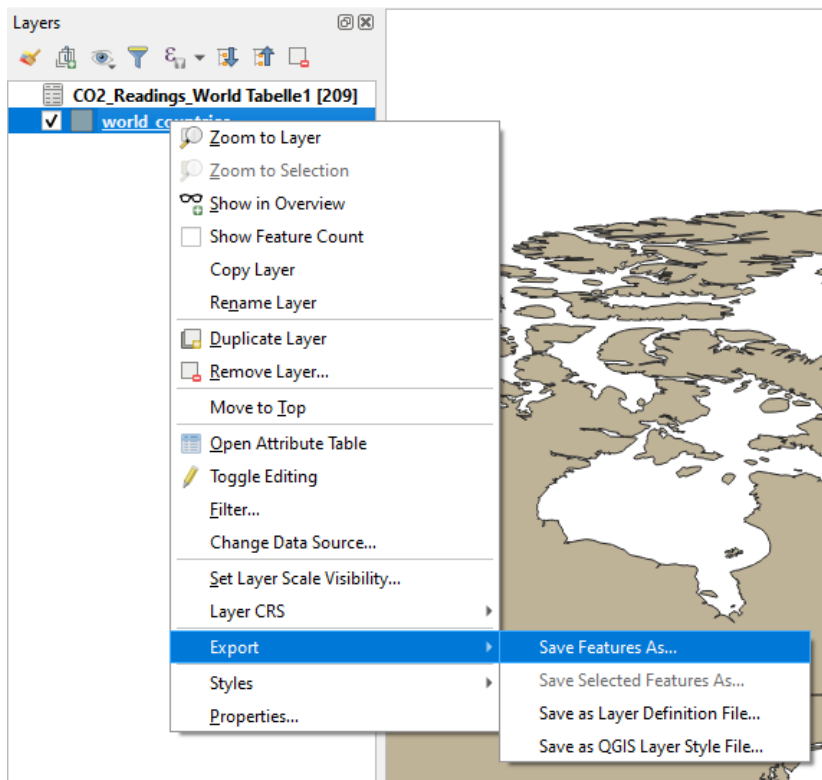


Abbildung 16. Wie wird die neue Ebene der verbundenen Tabellen gespeichert?

Wähle **GeoPackage** als Format.

Setze den Dateinamen auf `world_countries_co2`.

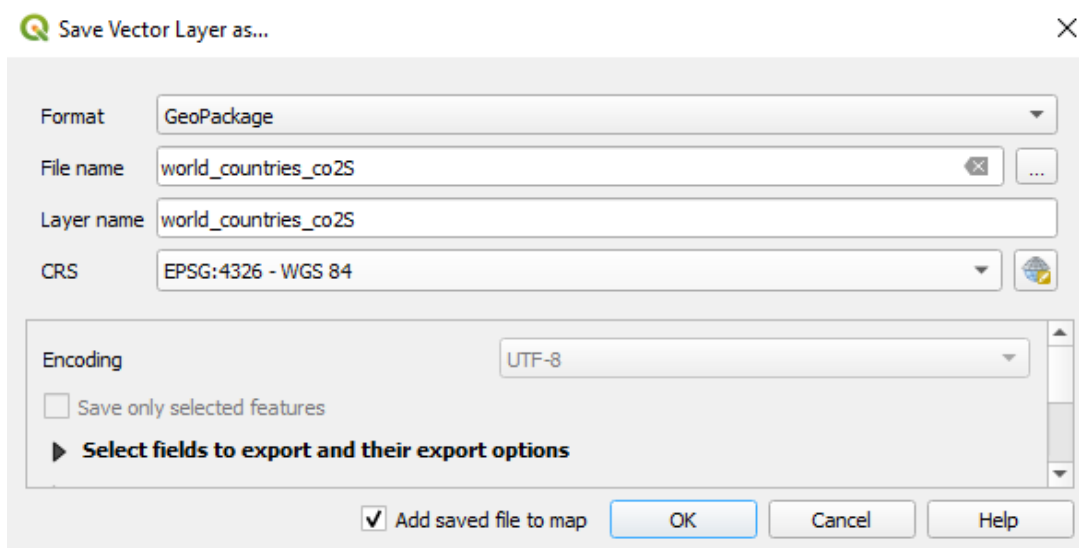


Abbildung 17. Einstellungen zum Speichern des neuen GeoPackages.

Überprüfe, ob **Add saved file to map** angekreuzt ist. Wenn das Häkchen gesetzt ist, kannst du fortfahren, indem du auf **[OK]** klickst. Die Meldung auf dem grünen Feld: "Layer export: Successful" sollte oberhalb der Karte im Hauptfenster zu sehen sein.

17. Entferne die ursprünglichen Ebenen *World_Countries*, *CO2_Readings_World* und *Worldbank_Countries*. Klicke auf jede Ebene und wähle: **[Remove layer]**.

18. Du kannst das Projekt nun in deinem Ordner speichern.

Übung 1.2 - Weitere Übung des Zusammenfügens

Jetzt kannst du selbst versuchen, Tabellen zu verbinden:

Verbinde die Tabellen aus der Originaltabelle in der Datei `World_Countries.gpkg`, aber verwende dieses Mal `CO2_Readings_World.xlsx`. Versuche, die Anleitung nicht als Leitfaden zu verwenden.

Übungen (Bonus-Teil)

Es mag zwar einfach sein, Daten zu erhalten, aber es ist ziemlich schwierig, ihre Bedeutung zu interpretieren. Deshalb werden wir in diesem Kapitel unsere nun verarbeiteten Daten klassifizieren und visualisieren. Lass uns also einen Blick darauf werfen, wie man genau das macht; wie man zusätzliche Informationen aus Tabellendatensätzen herausholen kann.

Übung 2.1 - Daten mit Farbe klassifizieren

In dieser Übung werden wir zusammenhängende Daten sichtbar machen, indem wir die Daten mit Farbe visualisieren.

Die folgenden Daten werden für diese Übung benötigt:

- `World_Countries.gpkg`
- `CO2_Lesungen_Welt.xlsx`
- `Weltbank_Länder.xlsx`

1. Öffne das Projekt aus Übung 1.2 in QGIS:

Nun kannst du entscheiden, welche Option du verwenden möchtest:

- **Project > Open**
- **Project > Open From** oder
- **Project > Open Recent** (Dies ist die bequemste Option, wenn du in letzter Zeit an deinem Projekt gearbeitet hast).

2. Öffne das Bedienfeld für die Ebenengestaltung und wähle als Ziel die Ebene **Verbundene Tabellen**.

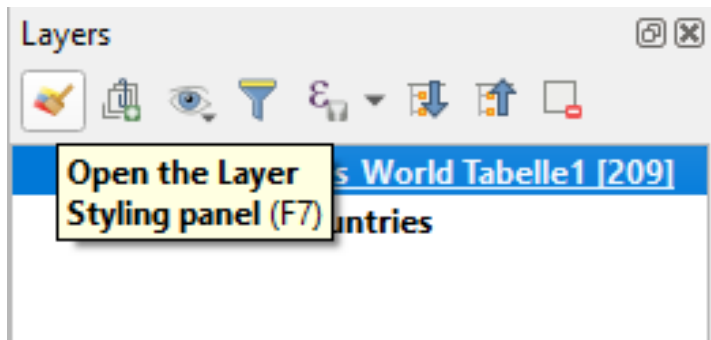


Abbildung 18. Wo findest du das Layer Styling Panel?

3. Wechsel zum Bedienfeld für die Ebenengestaltung und lege fest:

- Graduated Renderer
- Symbolwert auf `joined_CO2`
- Legendenformat als Standard belassen
- Farbrampe in eine beliebige Farbe ändern
- Modus auf Gleiche Anzahl (Quantil)

Wenn du fertig bist, klicke auf **[classify]** in der unteren linken Ecke des Fensters. Je nach der von dir gewählten Farbe sollte das Ergebnis in etwa so aussehen wie in der folgenden Abbildung:

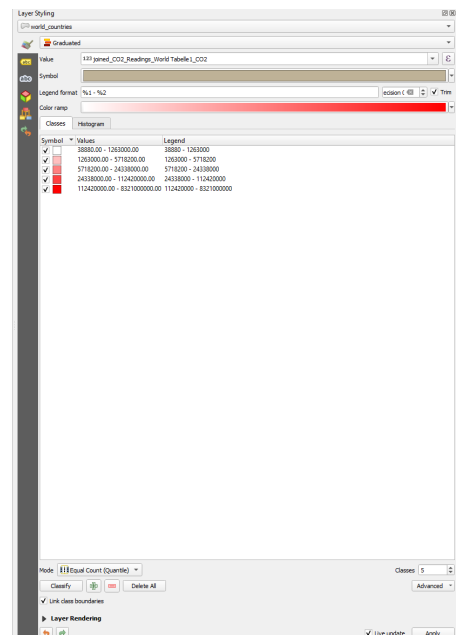
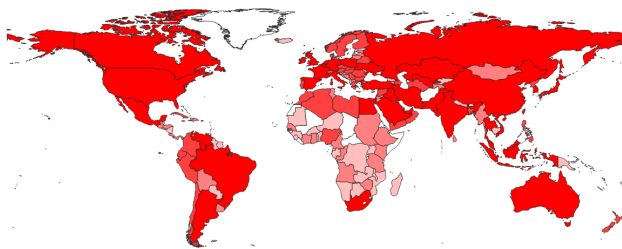


Abbildung 19. Result of applying the Equal Count classifier to the CO2 value in the joined table

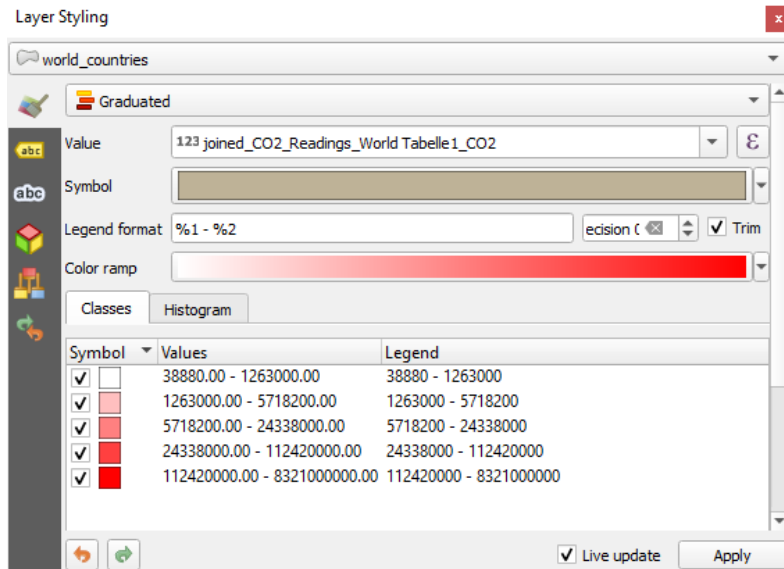


Abbildung 20. Classification settings in detail

4. Wiederhole Schritt 3 für den Wert der Weltbevölkerung. Vergiss nicht, zu einer anderen Farbe zu wechseln, sonst kannst du sie nicht mehr von der CO2-Schicht unterscheiden.

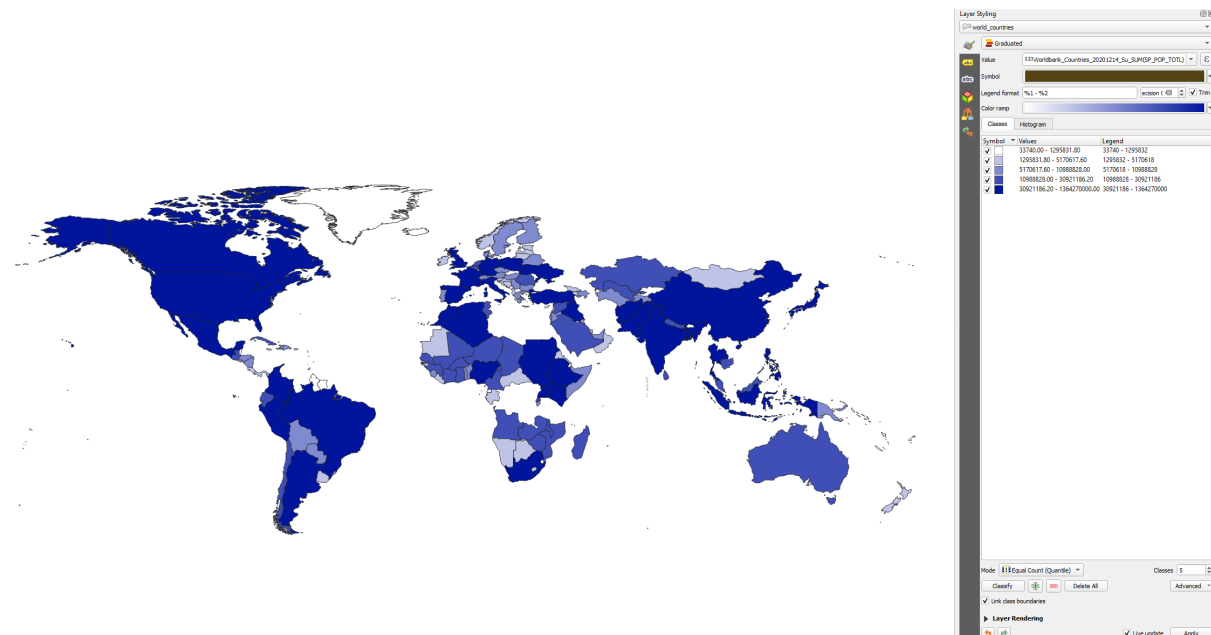


Abbildung 21. Result of applying the Equal Count classifier to the world population value in the joined table

Einige Informationen über die Datenmodi oder Klassifikatoren für fortgeschrittene Benutzer:

- **Equal count (Quantil):**

Stichprobenquantile werden entsprechend den gegebenen Wahrscheinlichkeiten erstellt. Die kleinste Beobachtung wird durch 0 und die grösste durch 1 dargestellt.

- **Equal Interval:**

Unterteilt den Bereich der Attributwerte in gleich grosse Klassen. Die Anzahl der Klassen wird vom Benutzer festgelegt. Die Klassifizierungsmethode des gleichen Intervalls eignet sich am besten für kontinuierliche Datensätze wie Niederschlag, Temperatur oder Bevölkerung.

- **Logarithmic Scale:**

Eine Methode, um numerische Daten über einen sehr grossen Wertebereich kompakt darzustellen - typischerweise sind die grössten Zahlen in den Daten Hunderte oder sogar Tausende Male grösser als die kleinsten Zahlen.

- **Natural Breaks (Jenks):**

Jenks ist eine Art "optimales" Klassifizierungsschema, das Klassenunterbrechungen findet, die (für eine gegebene Anzahl von Klassen) die Varianz innerhalb der Klasse minimieren und den Unterschied zwischen den Klassen maximieren.

- **Pretty Breaks:**

Rundet jeden Trennungspunkt auf oder ab. Anstatt eines Bruchpunkts bei 599,85 wird er also bei 600,00 gesetzt

- **Standard Deviation:**

Ermittelt den Mittelwert und platziert dann die Klassenunterbrechungen über und unter dem Mittelwert in Intervallen von entweder 0,25, 0,5 oder 1 Standardabweichung, bis alle Datenwerte in Klassen enthalten sind. Werte, die mehr als drei Standardabweichungen vom Mittelwert entfernt sind, werden in zwei Klassen zusammengefasst: grösser als und kleiner als die genannten Standardabweichungen.

5. Du kannst auch die Anzahl der Klassen ändern. Findest du die Zahl, die deinem Verständnis der Daten am besten entspricht.

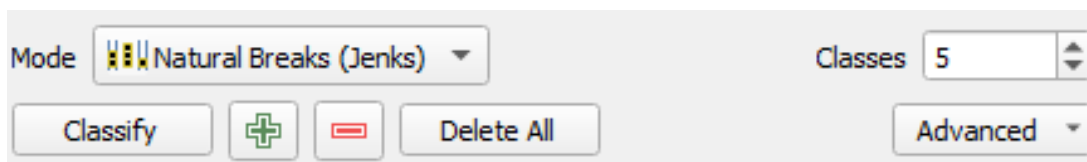


Abbildung 22. Die Anzahl an Klassen kann unten rechts gefunden werden.

6. Um die Unterschiede zwischen den Daten in verschiedenen Klassen hervorzuheben, ist es manchmal nützlich, verschiedene Farben zu verwenden. Die Option, die uns dies ermöglicht, heisst RdBu-Farbrampe:

Klicke auf:

Color ramp color > All Color Ramps > RdBu

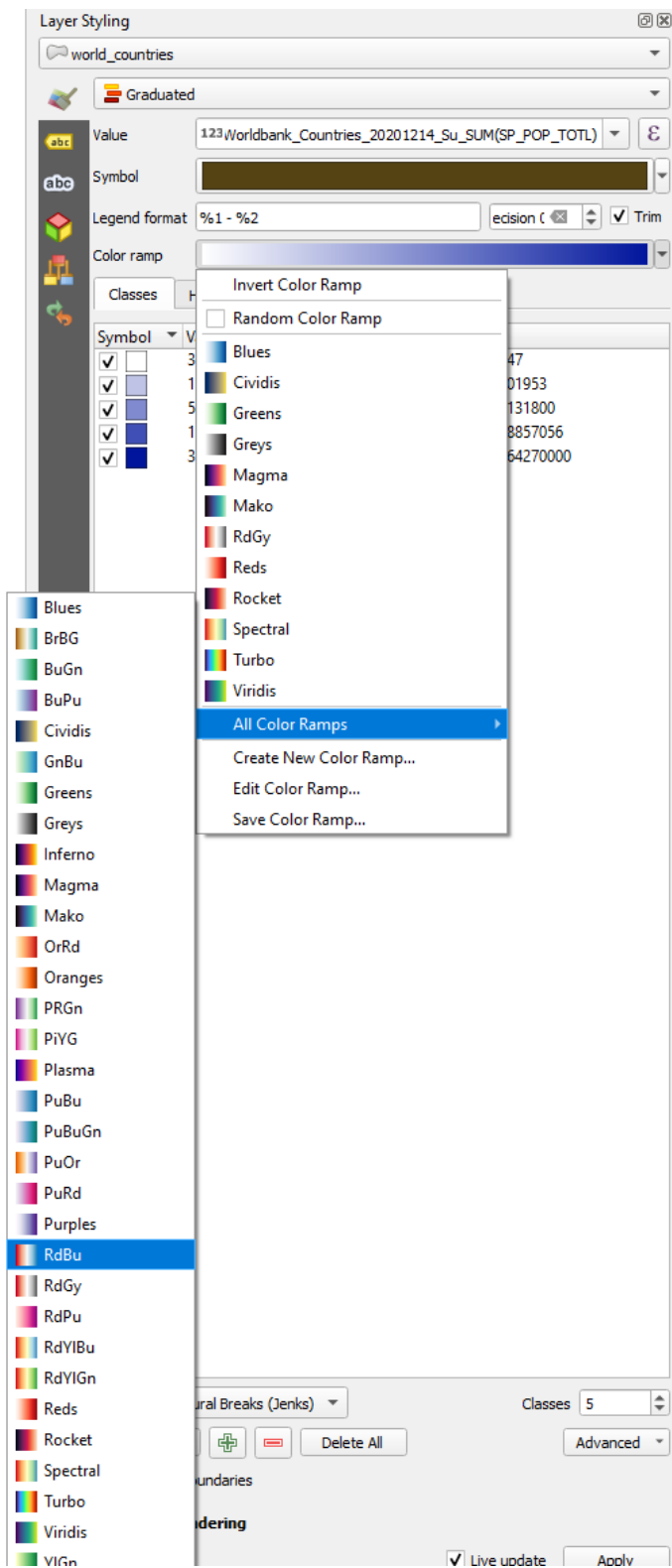


Abbildung 23. RdBu Option

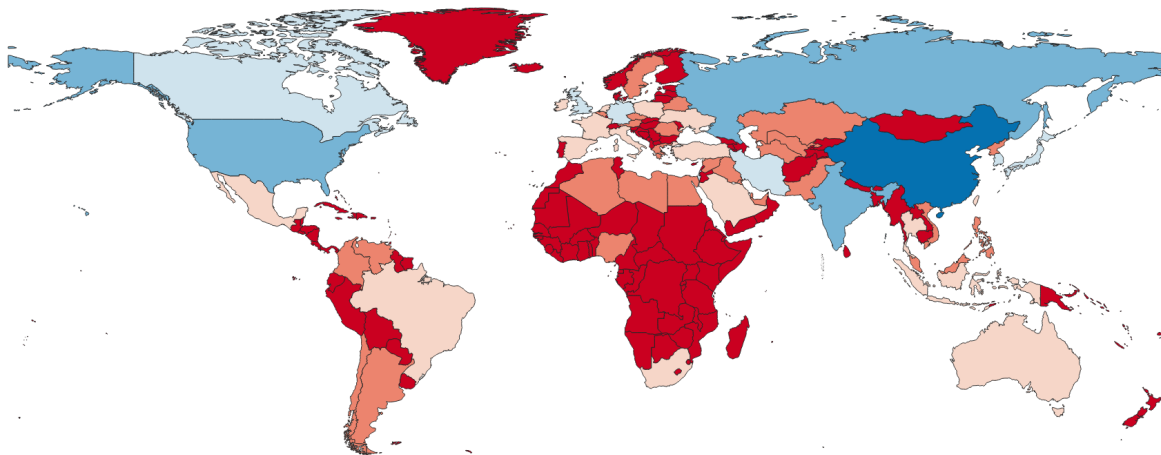


Abbildung 24. RdBu for 6 classes and Natural Breaks (Jenks) classification mode

[image] | [joining_tables/qgis_colourful_values.png](#)

Abbildung 25. Default Legende

Gebe in die Legende ein, um die Zeichenfolgen durch Wörter zu ersetzen:

[image] | [joining_tables/qgis_adjusted_values.png](#)

Abbildung 26. Modifizierte Legende

Übung 2.2 - Zusätzliche Daten verbinden

Im Datenordner finden Sie eine Datei: [RenewableEnergy_Percentage.csv](#)

Verwende diese Datei für zusätzliche Daten, die du in deine Tabelle einfügen musst (wiederhole die Schritte aus Teil 1 dieses Tutorials). Führe die Klassifizierung durch. Beschreibe die Ergebnisse und erläutere die Verwendung der Anzahl der Klassen und Klassifizierungsmodi, die du gewählt hast. Die Verknüpfung zusätzlicher Daten wird manchmal auch als *Data Enrichment* bezeichnet.

Fazit und Ausblick

Zusammenfassend lässt sich sagen, dass dieses Tutorial darauf ausgerichtet war, zu erklären, wie man tabellarische Daten verbindet. Das Ziel der Übungen in QGIS war es, die Methoden kennenzulernen, die es dem Benutzer ermöglichen, Datentabellen zu verbinden und so neue Tabellen mit erweitertem Kontext zu erstellen. Die grundlegende Einführung in die verschiedenen Beziehungen zwischen den Daten, wie **one-to-one**, **one-to-many** und **many-to-many**, kann weiter studiert werden, um das Wissen über Datensätze zu vertiefen. Der Teil, der sich mit der Datenanalyse befasst, vermittelt die Grundlagen der Datenklassifizierung. Die zusätzlichen Informationen über die Klassifizierungsalgorithmen geben einen komfortablen Vorsprung, der von den Schülern individuell

weiter ausgebaut werden kann. Dieses Tutorial dient als Anleitung zum Importieren von Daten in eine QGIS-Umgebung, zum Zusammenführen in eine neue Tabelle, zum Klassifizieren und schliesslich zum Analysieren der Daten.

Die **Many-to-many-Beziehung** wurde in den Übungen hier nicht behandelt. Es ist jedoch möglich, **Many-to-many-Beziehungen** in **QGIS** zu verwenden. Wenn du mehr darüber erfahren möchtest, kannst du das in: [QGIS Doc](#).

Wenn du mehr über Joins in Datenbanken und deren Verwendung erfahren möchtest, findest du weitere Informationen unter dem Stichwort **SQL JOIN**.

Noch Fragen? Wende dich an [OpenStreetMap Schweiz](#) oder [Stefan Keller](#)!



Frei verwendbar unter [CC0 1.0](#)