FIT3179 Data Visualisation

Week 8 Studio Activity: Create Maps with Vega-Lite

Table of Contents

1. Overview of Spatial Data

We will briefly introduce you to GeoJSON, TopoJSON and shapefile as we will use those file types in this tutorial.

1.1 GeoJSON format

One of the most popular spatial data formats is GeoJSON. GeoJSON is a JSON text file containing specialised attributes for geospatial data storage. The GeoJSON standard format is specified under IETF ‘s (Internet Engineering Task Force) RFC 7946. The GeoJSON format consists of several geometry objects (figure 1).

- **Point**: a single geospatial position.
- **LineString:** is an array of two or more positions.
- **Polygon:** an array of linear ring coordinate arrays.
- **MultiPoint:** an array of positions.
- **MultiLineString**: an array of LineString coordinate arrays.
- **MultiPolygon**: an array of Polygon coordinate arrays.
- **GeometryCollections**: collection of geometry objects.

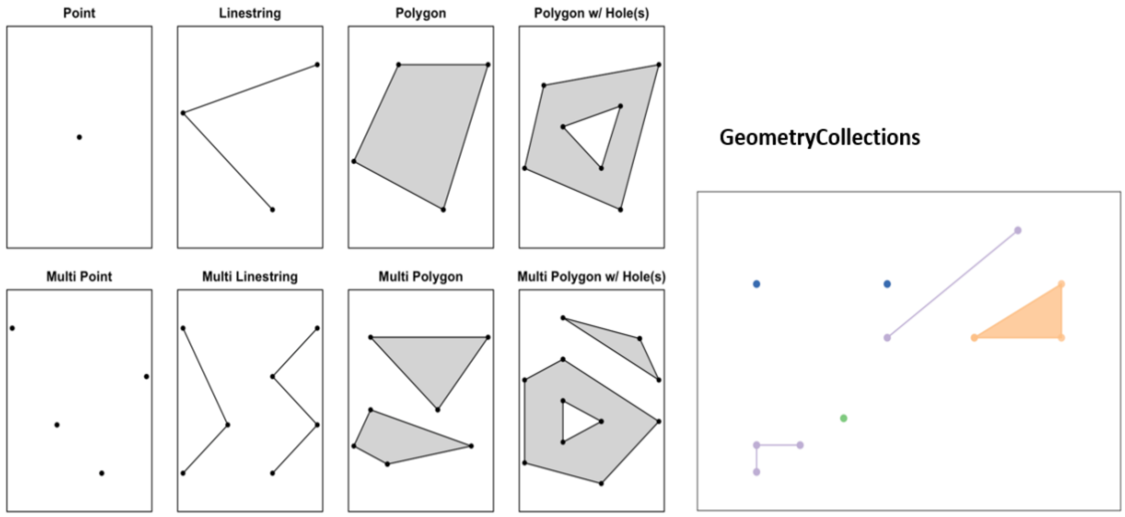


Figure 1. Geometry objects. Source.

1.2 TopoJSON

TopoJSON is an extension of GeoJSON that encodes topology. Rather than representing geometries discreetly, geometries in TopoJSON files are stitched together from shared line segments called arcs.

The main benefit of using TopoJSON file is that it improves shape simplification by avoiding artifacts that would be caused by simplifying shapes independently. It also enables applications like map coloring and selective meshing and makes the format more compact by removing redundant coordinates.

TopoJSON is the only spatial file format that is directly supported in Vega-Lite v5.

1.3 Shapefile

Shapefile is another spatial data format. This data format is developed and regulated by Esri, the leading company in Geographic Information System (GIS) software. ArcGIS (a widely used GIS) is one of their products. A Shapefile is actually a group of files stored in a single directory. These files are **.shp, .shx,** and **.dbf**.

- .**shp**: contains geometry objects.
- .**shx**: contains index.
- .**dbf**: contains attribute of dBase format.

The Shapefile format also has other files such as **.prj, .cpg,** or **.qpg**. Make sure to always keep all files that make up a Shapefile dataset in the same directory.

2. Obtaining Spatial Data

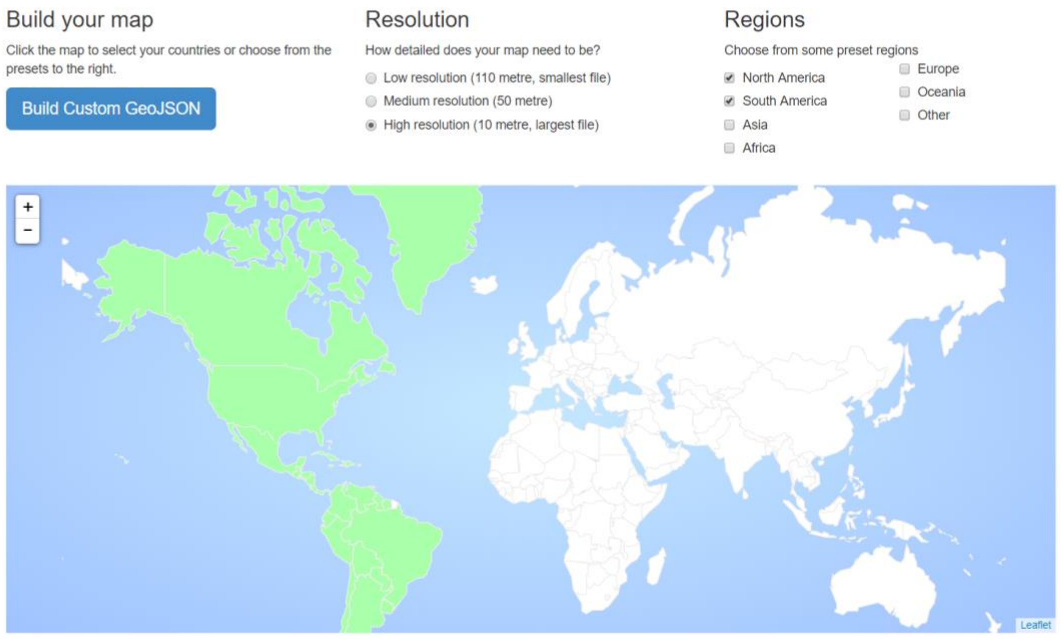
As Vega-Lite supports TopoJSON, the first step is obtaining data in TopoJSON format. While it is sometimes not easy to get a TopoJSON file directly, there is a useful website (https://mapshaper.org/) which can convert between shapefile, GeoJSON, and TopoJSON file formats.

In this subsection, we will illustrate how to obtain a GeoJSON file and a Shapefile file for a global map, and then convert them to the TopoJSON file that we need for Vega-Lite.

2.1 From GeoJSON to TopoJSON

As a starting point, we will extract a GeoJSON file of a specific area using https://geojson-maps.ash.ms.

Step 1. Go to the website. Select high resolution ("110 metre, low resolution") and check North and South America. (Note: This website uses data from https://www.naturalearthdata.com, which provides additional data layers. "110 metre" is an apparent misunderstanding of the authors of this website, because the data is optimised for a map at a scale of 1:110 million.)



Build your map

Click the map to select your countries or choose from the presets to the right.

Resolution

How detailed does your map need to be?

○ Low resolution (110 metre, smallest file)
○ Medium resolution (50 metre)
● High resolution (10 metre, largest file)

Regions

Choose from some preset regions

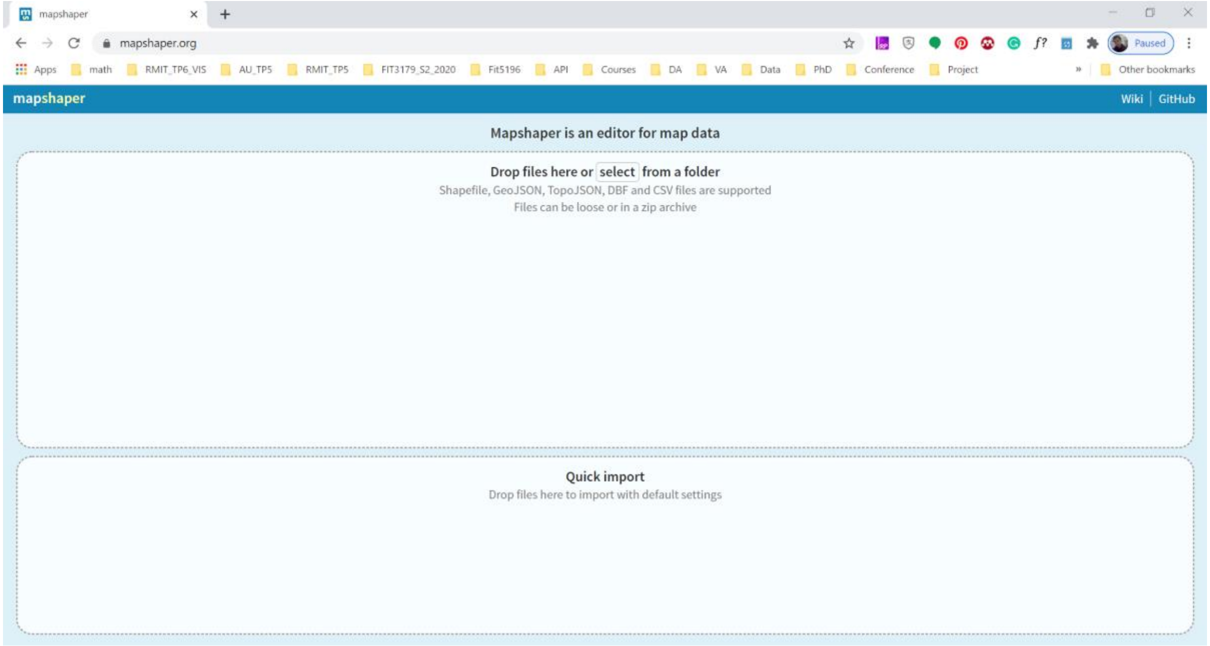
☑ North America
☑ South America
☐ Asia
☐ Africa
☐ Europe
☐ Oceania
☐ Other

Step 2. Click Build Custom GeoJSON. It will generate and download a **custom.geo.json** file.

Step 3. To make it easy to remember, let’s rename that file to **america.geo.json**.

Step 4. You now have the GeoJSON file of the two American continents.

Step 5. Next, open the website: https://mapshaper.org/



Mapshaper is an editor for map data

Drop files here or select from a folder

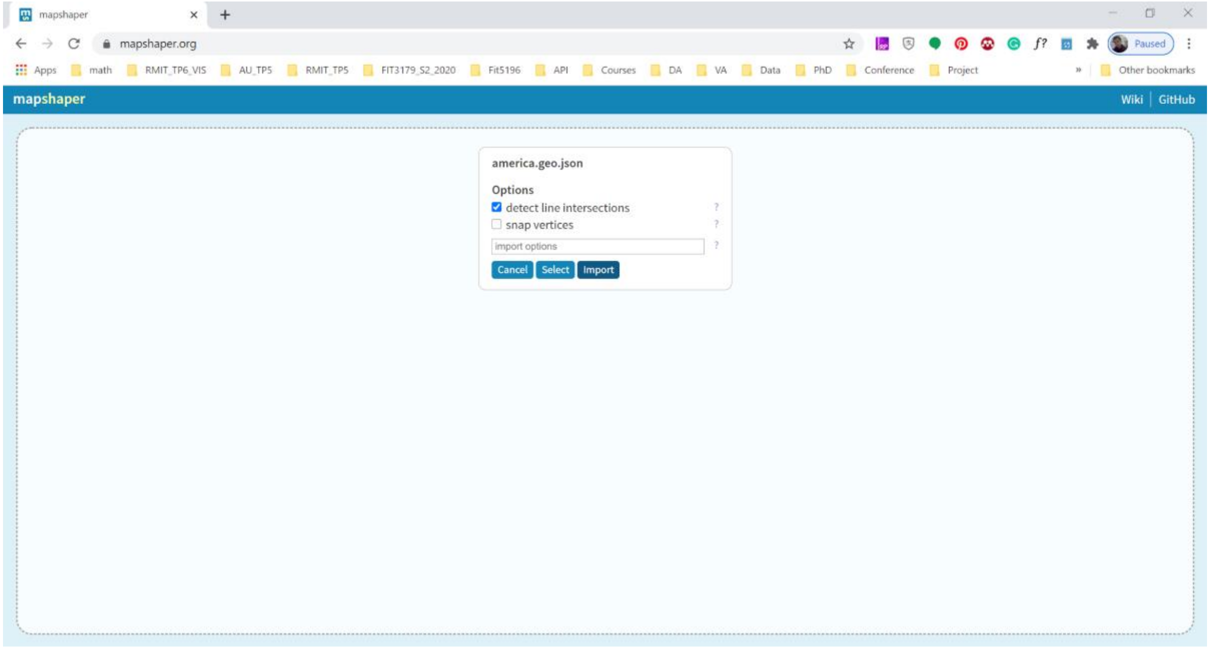
Shapefile, GeoJSON, TopoJSON, DBF and CSV files are supported

Files can be loose or in a zip archive

Quick import

Drop files here to import with default settings

Step 6. Click “Select”. Then choose the “**america.geo.json**.” file that you just saved in Step 3. There will be a window popping up. Just click “Import” to continue.



america.geo.json

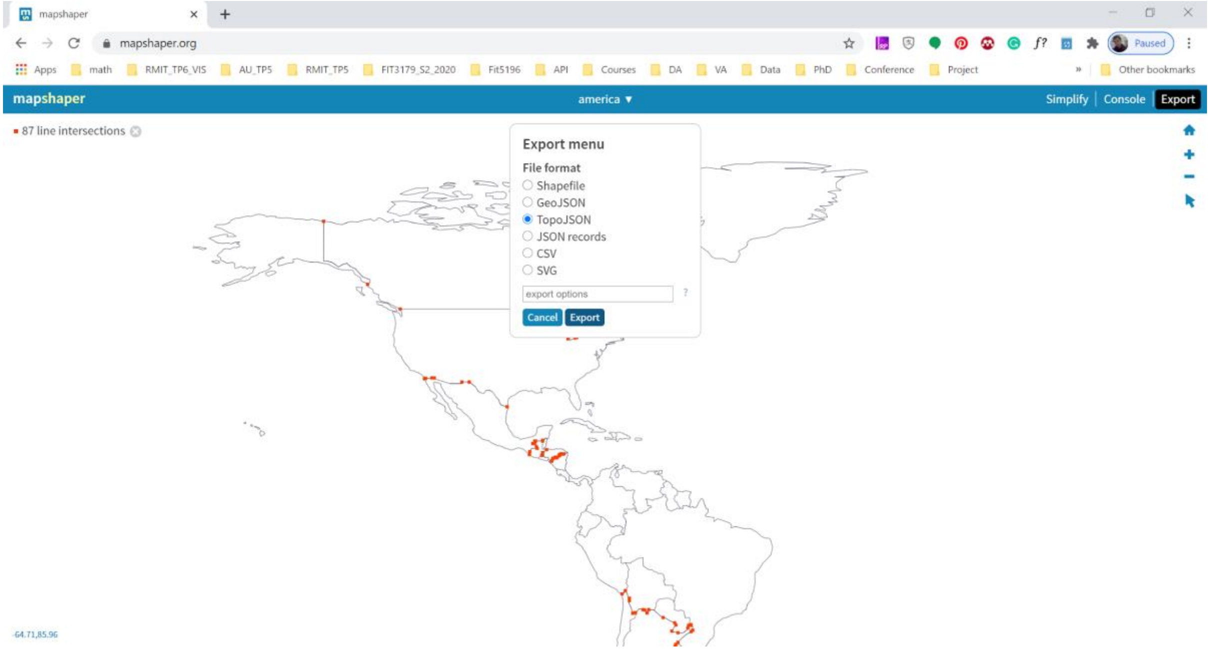
Options

☑ detect line intersections

□ snap vertices

import options

Step 7. The map of the Americas will be shown on the page. Now, click "Export" on the top-right corner, and select "TopoJSON" > "Export" to get the TopoJSON file.



87 line intersections

Export menu

File format

○ Shapefile

○ GeoJSON

● TopoJSON

○ JSON records

○ CSV

○ SVG

export options

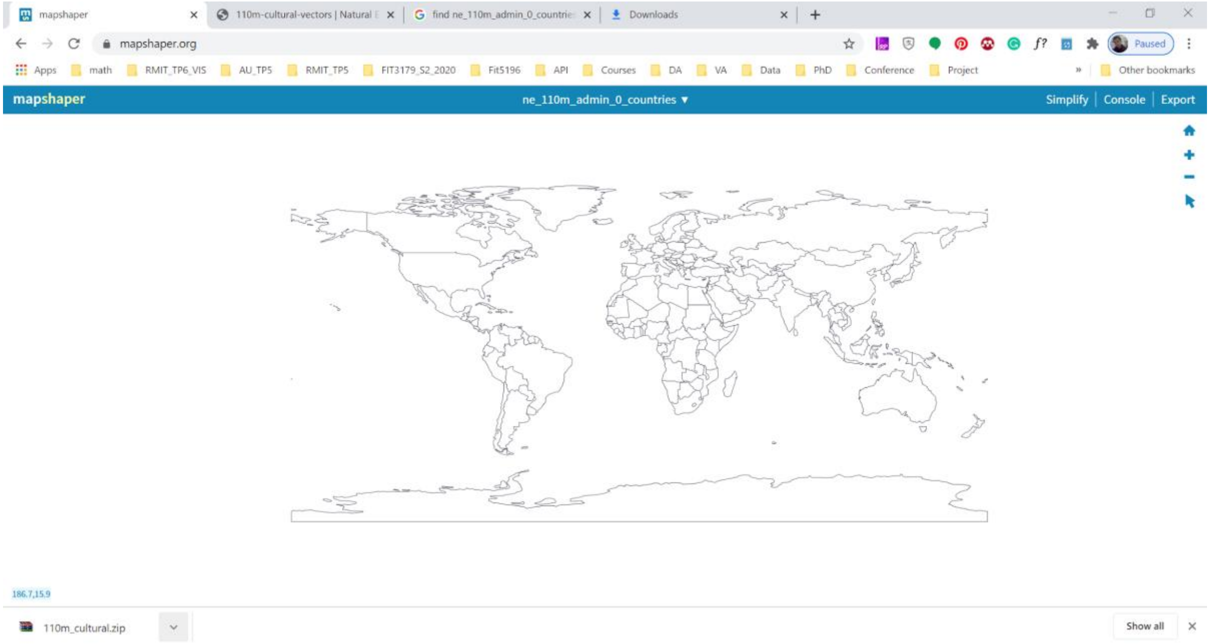
2.2 From Shapefile to TopoJSON

We can also download the shapefile of the global map directly from https://www.naturalearthdata.com , and then convert the shapefile to TopoJSON.

Step 1. Go to https://www.naturalearthdata.com/downloads/, under “Small scale data, 1:110m” > select “Cultural” > click “Download all 110m cultural themes (1.3MB)”. Then the file will be downloaded as “110m_cultural.zip”. This zip file contains all the required files for the shapefile format. You do not need to unzip the file.

Note: You can also download other files from https://www.naturalearthdata.com/downloads/ and check their differences.

Step 2. Now open https://mapshaper.org/, click “Select” > find “110m_cultural.zip”. Repeat Steps 6 and 7 in Section 2.1 to get your TopoJSON file.



Note: There are many useful online websites providing various types of spatial data. You can just google your required spatial regions (e.g., global, Australia, America) + file types (e.g., shapefile, GeoJSON, TopoJSON). If the file type is not TopoJSON, you can then convert the file to TopoJSON in https://mapshaper.org/ (or using some other tools that you find, e.g., QGIS, geopandas package in python, etc.). Here are some useful websites that provide spatial data:

- Global scale: https://www.naturalearthdata.com
- Australia Spatial Data: https://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.001
- Victoria Suburbs: https://data.gov.au/dataset/ds-dga-af33dd8c-0534-4e18-9245-fc64440f742e/details

3. Proportional Symbol Maps with Vega-Lite

In this section, we will build a simple proportional symbol map in Vega-Lite using an earthquake dataset. The dataset contains the earthquakes of magnitude 4 or larger between 4 Oct 2020 and 10 Oct 2020. The data is downloaded from https://earthquake.usgs.gov/earthquakes/search/. The downloaded data is available as CSV table here:
https://github.com/KaneSec/vega_lite/blob/main/2_symbol_map/data/earthquake.csv

The main attributes that we use include:

- latitude
- longitude
- mag
- depth

We will also use the TopoJSON data (from Section 1.2.2) to create a base map for our proportional symbol map. The TopoJSON data is available from here:
https://github.com/KaneSec/vega_lite/tree/main/2_symbol_map/js

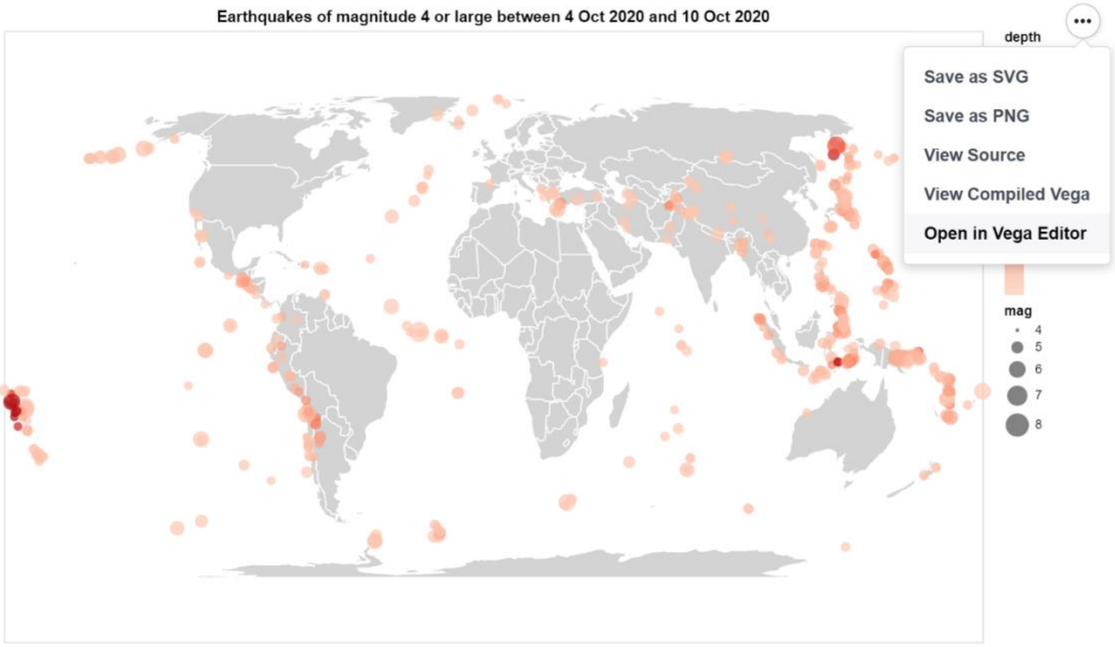
3.1 Overview

The final proportional symbol map will look like:
https://kanesec.github.io/vega_lite/2_symbol_map/

The example GitHub repository is available here:
https://github.com/KaneSec/vega_lite/tree/main/2_symbol_map

3.2 Understand the Vega-Lite JSON file

While we walk through the process of creating this map, you can see the final map here https://kanesec.github.io/vega_lite/2_symbol_map/. From the visualisation page, click “Open in Vega Editor” (as shown below) to edit the JSON code.



Here is the JSON specification. We explain it line by line below.

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "width": 800,
  "height": 500,
  "title": "Earthquakes of magnitude 4 or large between 4 Oct 2020 and 10 Oct 2020",
  "projection": {"type": "equalEarth"},
  "layer": [
    {
      "data": {
        "url":
"https://raw.githubusercontent.com/KaneSec/vega_lite/main/2_symbol_map/js/ne_110m_admin_0_countries.topojson"
,
        "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
      },
      "mark": {"type": "geoshape", "fill": "lightgray", "stroke": "white"}
    },
    {
      "data": {
        "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/2_symbol_map/data/earthquake.csv"
      },
      "mark": {"type": "circle", "tooltip": {"content": "data"}},
      "encoding": {
        "longitude": {"field": "longitude", "type": "quantitative"},
        "latitude": {"field": "latitude", "type": "quantitative"},
        "size": {
         "field": "mag",
         "type": "quantitative",
         "scale": {"domain": [4, 8]}
        },
        "color": {
         "field": "depth",
         "type": "quantitative",
         "scale": {"scheme": "reds"}
        }
      }
    }
  ]
}
```

[Please view the line numbers in the Vega Editor]

Lines 2-5: We specify the Vega-Lite version, the width and height of the map, as well as the title of the visualisation.

```
"$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "width": 800,
  "height": 600,
  "title": "Earthquakes of magnitude 4 or large between 4 Oct 2020 and 10 Oc
t 2020",
```

Line 6: We set up the map projection as "`equalEarth`". Vega-Lite supports a few different cartographic projections including the Equirectangular projection, the orthographic projection, etc. Please check this page for all the supported projections: https://vega.github.io/vega-lite/docs/projection.html. Change the projection type (in Vega Editor) to a better projection and see the differences.

```
"projection": {"type": "equalEarth"},
```

Lines 7–42 define the two layers of the map. The first layer consists of the outline of countries (defined in lines 8–14), and the second layer consists of proportional symbols (defined in lines 15–34). Let’s have a look at the map layer first. Please paste the code below to see the map layer.

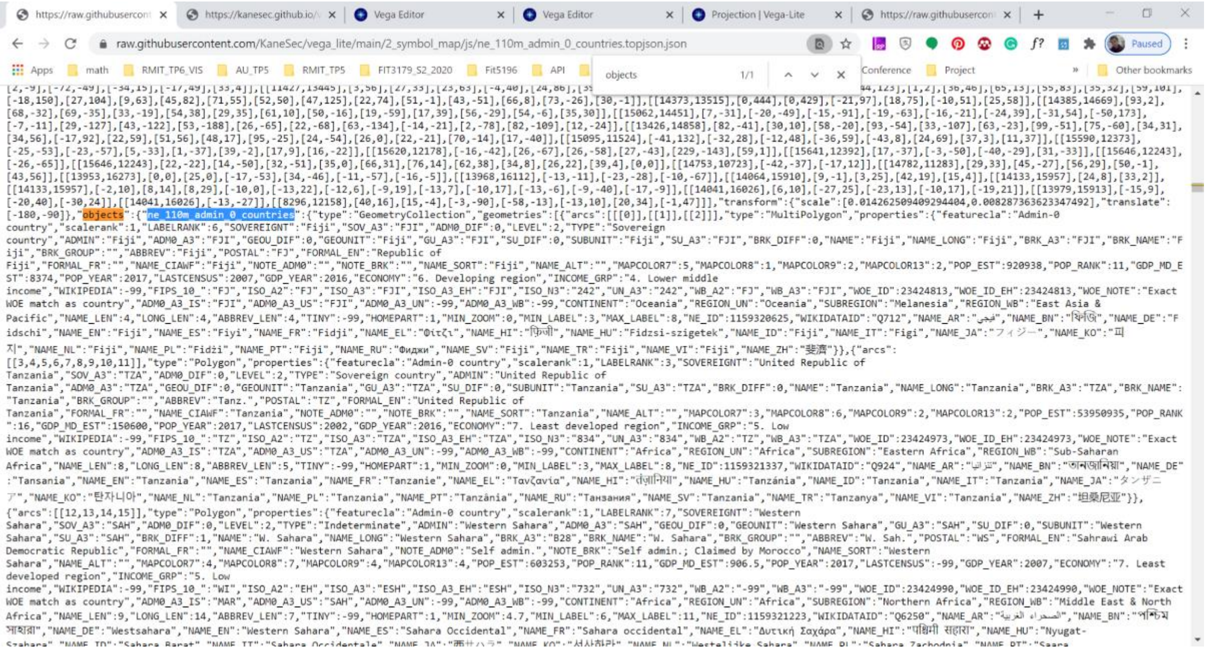
```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "width": 800,
  "height": 600,
  "title": "Earthquakes of magnitude 4 or large between 4 Oct 2020 and 10 Oct 2020",
  "projection": {"type": "equalEarth"},
  "layer": [
    {
      "data": {
        "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/2_symbol_map/js/ne_110m_admin_0_countries.topojson",
        "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
      },
      "mark": {"type": "geoshape", "fill": "lightgray", "stroke": "white"}
    }]
}
```

This first layer defines our base map:

- url: this is the TopoJSON file (that we included in our GitHub repository) – please check the GitHub repository in Section 1.2.1 for the project layout.
- format > type: topojson
- format > feature: "ne_110m_admin_0_countries" (this is the name of the TopoJSON object set to convert to a GeoJSON feature collection, e.g., where it contains the boundary information of each country). In general, you will need to understand your TopoJSON file to find such information.

  Tips: 1. to quickly identify this name in your TopoJSON file, just search “objects” in the TopoJSON file, the key under that could be used as your feature name).

  2. Try to copy the TopoJSON content to an online json formatter (e.g., [link]). Then check the formatted json file to understand the structure of TopoJSON file.


https://raw.githubuserconte
https://kanesec.github.io/
Vega Editor
Vega Editor
Projection | Vega-Lite
https://raw.githubuserconte
raw.githubusercontent.com/KaneSec/vega_lite/main/2_symbol_map/js/ne_110m_admin_0_countries.topjson.json
Paused
Apps
math
RMIT_TP6_VIS
AU_TP5
RMIT_TP5
FIT3179_S2_2020
Fit5196
API
objects
1/1
Conference
Project
Other bookmarks

```
[2,-9],[-72,-49],[-34,15],[-17,49],[33,4]],[[11427,13445],[3,56],[27,33],[23,63],[-4,40],[24,86],[3 44,123],[1,2],[36,46],[65,13],[55,83],[35,32],[59,101],
[-18,150],[27,104],[9,63],[45,82],[71,55],[52,50],[47,125],[22,74],[51,-1],[43,-51],[66,8],[73,-26],[30,-1]],[[14373,13515],[0,444],[0,429],[-21,97],[18,75],[-10,51],[25,58]],[[14385,14669],[93,2],
[68,-32],[69,-35],[33,-19],[54,38],[29,35],[61,10],[50,-16],[19,-59],[17,39],[56,-29],[54,-6],[35,30]],[[15062,14451],[7,-31],[-20,-49],[-15,-91],[-19,-63],[-16,-21],[-24,39],[-31,54],[-50,173],
[-7,-11],[29,-127],[43,-122],[53,-188],[26,-65],[22,-68],[63,-134],[-14,-21],[2,-78],[82,-109],[12,-24]],[[13426,14858],[82,-41],[30,10],[58,-20],[93,-54],[33,-107],[63,-23],[99,-51],[75,-60],[34,31],
[34,56],[-17,92],[22,59],[51,56],[48,17],[95,-25],[24,-54],[26,0],[22,-21],[70,-14],[17,-40]],[[15095,11524],[-41,132],[-32,28],[-12,48],[-36,59],[-43,8],[24,69],[37,3],[11,37]],[[15590,12373],
[-25,-53],[-23,-57],[5,-33],[1,-37],[39,-2],[17,9],[16,-22]],[[15620,12178],[-16,-42],[26,-67],[26,-58],[27,-43],[229,-143],[59,1]],[[15641,12392],[17,-37],[-3,-50],[-40,-29],[31,-33]],[[15646,12243],
[-26,-65]],[[15646,12243],[22,-22],[14,-50],[32,-51],[35,0],[66,31],[76,14],[62,38],[34,8],[26,22],[39,4],[0,0]],[[14753,10723],[-42,-37],[-17,12]],[[14782,11283],[29,33],[45,-27],[56,29],[50,-1],
[43,56]],[[13953,16273],[0,0],[25,0],[-17,-53],[34,-46],[-11,-57],[-16,-5]],[[13968,16112],[-13,-11],[-23,-28],[-10,-67]],[[14064,15910],[9,-1],[3,25],[42,19],[15,4]],[[14133,15957],[24,8],[33,2]],
[[14133,15957],[-2,10],[8,14],[8,29],[-10,0],[-13,22],[-12,6],[-9,19],[-13,7],[-10,17],[-13,-6],[-9,-40],[-17,-9]],[[14041,16026],[6,10],[-27,25],[-23,13],[-10,17],[-19,21]],[[13979,15913],[-15,9],
[-20,40],[-30,24]],[[14041,16026],[-13,-27]],[[8296,12158],[40,16],[15,-4],[-3,-90],[-58,-13],[-13,10],[20,34],[-1,47]]],"transform":{"scale":[0.014262509409294404,0.008287363623347492],"translate":
[-180,-90]},"objects":{"ne_110m_admin_0_countries":{"type":"GeometryCollection","geometries":[{"arcs":[[[0]],[[1]],[[2]]],"type":"MultiPolygon","properties":{"featurecla":"Admin-0
country","scalerank":1,"LABELRANK":6,"SOVEREIGNT":"Fiji","SOV_A3":"FJI","ADM0_DIF":0,"LEVEL":2,"TYPE":"Sovereign
country","ADMIN":"Fiji","ADM0_A3":"FJI","GEOU_DIF":0,"GEOUNIT":"Fiji","GU_A3":"FJI","SU_DIF":0,"SUBUNIT":"Fiji","SU_A3":"FJI","BRK_DIFF":0,"NAME":"Fiji","NAME_LONG":"Fiji","BRK_A3":"FJI","BRK_NAME":"F
iji","BRK_GROUP":"","ABBREV":"Fiji","POSTAL":"FJ","FORMAL_EN":"Republic of
Fiji","FORMAL_FR":"","NAME_CIAWF":"Fiji","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"Fiji","NAME_ALT":"","MAPCOLOR7":5,"MAPCOLOR8":1,"MAPCOLOR9":2,"MAPCOLOR13":2,"POP_EST":920938,"POP_RANK":11,"GDP_MD_E
ST":8374,"POP_YEAR":2017,"LASTCENSUS":2007,"GDP_YEAR":2016,"ECONOMY":"6. Developing region","INCOME_GRP":"4. Lower middle
income","WIKIPEDIA":-99,"FIPS_10_":"FJ","ISO_A2":"FJ","ISO_A3":"FJI","ISO_A3_EH":"FJI","ISO_N3":"242","UN_A3":"242","WB_A2":"FJ","WB_A3":"FJI","WOE_ID":23424813,"WOE_ID_EH":23424813,"WOE_NOTE":"Exact
WOE match as country","ADM0_A3_IS":"FJI","ADM0_A3_US":"FJI","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Oceania","REGION_UN":"Oceania","SUBREGION":"Melanesia","REGION_WB":"East Asia &
Pacific","NAME_LEN":4,"LONG_LEN":4,"ABBREV_LEN":4,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":0,"MIN_LABEL":3,"MAX_LABEL":8,"NE_ID":1159320625,"WIKIDATAID":"Q712","NAME_AR":"فيجي","NAME_BN":"ফিজি","NAME_DE":"F
idschi","NAME_EN":"Fiji","NAME_ES":"Fiyi","NAME_FR":"Fidji","NAME_EL":"Φίτζι","NAME_HI":"फ़िजी","NAME_HU":"Fidzsi-szigetek","NAME_ID":"Fiji","NAME_IT":"Figi","NAME_JA":"フィジー","NAME_KO":"피
지","NAME_NL":"Fiji","NAME_PL":"Fidżi","NAME_PT":"Fiji","NAME_RU":"Фиджи","NAME_SV":"Fiji","NAME_TR":"Fiji","NAME_VI":"Fiji","NAME_ZH":"斐濟"}},{"arcs":
[[3,4,5,6,7,8,9,10,11]],"type":"Polygon","properties":{"featurecla":"Admin-0 country","scalerank":1,"LABELRANK":3,"SOVEREIGNT":"United Republic of
Tanzania","SOV_A3":"TZA","ADM0_DIF":0,"LEVEL":2,"TYPE":"Sovereign country","ADMIN":"United Republic of
Tanzania","ADM0_A3":"TZA","GEOU_DIF":0,"GEOUNIT":"Tanzania","GU_A3":"TZA","SU_DIF":0,"SUBUNIT":"Tanzania","SU_A3":"TZA","BRK_DIFF":0,"NAME":"Tanzania","NAME_LONG":"Tanzania","BRK_A3":"TZA","BRK_NAME":
"Tanzania","BRK_GROUP":"","ABBREV":"Tanz.","POSTAL":"TZ","FORMAL_EN":"United Republic of
Tanzania","FORMAL_FR":"","NAME_CIAWF":"Tanzania","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"Tanzania","NAME_ALT":"","MAPCOLOR7":3,"MAPCOLOR8":6,"MAPCOLOR9":2,"MAPCOLOR13":2,"POP_EST":53950935,"POP_RANK
":16,"GDP_MD_EST":150600,"POP_YEAR":2017,"LASTCENSUS":2002,"GDP_YEAR":2016,"ECONOMY":"7. Least developed region","INCOME_GRP":"5. Low
income","WIKIPEDIA":-99,"FIPS_10_":"TZ","ISO_A2":"TZ","ISO_A3":"TZA","ISO_A3_EH":"TZA","ISO_N3":"834","UN_A3":"834","WB_A2":"TZ","WB_A3":"TZA","WOE_ID":23424973,"WOE_ID_EH":23424973,"WOE_NOTE":"Exact
WOE match as country","ADM0_A3_IS":"TZA","ADM0_A3_US":"TZA","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Africa","REGION_UN":"Africa","SUBREGION":"Eastern Africa","REGION_WB":"Sub-Saharan
Africa","NAME_LEN":8,"LONG_LEN":8,"ABBREV_LEN":5,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":0,"MIN_LABEL":3,"MAX_LABEL":8,"NE_ID":1159321337,"WIKIDATAID":"Q924","NAME_AR":"تنزانيا","NAME_BN":"তানজানিয়া","NAME_DE"
:"Tansania","NAME_EN":"Tanzania","NAME_ES":"Tanzania","NAME_FR":"Tanzanie","NAME_EL":"Τανζανία","NAME_HI":"तंज़ानिया","NAME_HU":"Tanzánia","NAME_ID":"Tanzania","NAME_IT":"Tanzania","NAME_JA":"タンザニ
ア","NAME_KO":"탄자니아","NAME_NL":"Tanzania","NAME_PL":"Tanzania","NAME_PT":"Tanzânia","NAME_RU":"Танзания","NAME_SV":"Tanzania","NAME_TR":"Tanzanya","NAME_VI":"Tanzania","NAME_ZH":"坦桑尼亚"}},
{"arcs":[[12,13,14,15]],"type":"Polygon","properties":{"featurecla":"Admin-0 country","scalerank":1,"LABELRANK":7,"SOVEREIGNT":"Western
Sahara","SOV_A3":"SAH","ADM0_DIF":0,"LEVEL":2,"TYPE":"Indeterminate","ADMIN":"Western Sahara","ADM0_A3":"SAH","GEOU_DIF":0,"GEOUNIT":"Western Sahara","GU_A3":"SAH","SU_DIF":0,"SUBUNIT":"Western
Sahara","SU_A3":"SAH","BRK_DIFF":1,"NAME":"W. Sahara","NAME_LONG":"Western Sahara","BRK_A3":"B28","BRK_NAME":"W. Sahara","BRK_GROUP":"","ABBREV":"W. Sah.","POSTAL":"WS","FORMAL_EN":"Sahrawi Arab
Democratic Republic","FORMAL_FR":"","NAME_CIAWF":"Western Sahara","NOTE_ADM0":"Self admin.","NOTE_BRK":"Self admin.; Claimed by Morocco","NAME_SORT":"Western
Sahara","NAME_ALT":"","MAPCOLOR7":4,"MAPCOLOR8":7,"MAPCOLOR9":4,"MAPCOLOR13":4,"POP_EST":603253,"POP_RANK":11,"GDP_MD_EST":906.5,"POP_YEAR":2017,"LASTCENSUS":-99,"GDP_YEAR":2007,"ECONOMY":"7. Least
developed region","INCOME_GRP":"5. Low
income","WIKIPEDIA":-99,"FIPS_10_":"WI","ISO_A2":"EH","ISO_A3":"ESH","ISO_A3_EH":"ESH","ISO_N3":"732","UN_A3":"732","WB_A2":"-99","WB_A3":"-99","WOE_ID":23424990,"WOE_ID_EH":23424990,"WOE_NOTE":"Exact
WOE match as country","ADM0_A3_IS":"MAR","ADM0_A3_US":"SAH","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Africa","REGION_UN":"Africa","SUBREGION":"Northern Africa","REGION_WB":"Middle East & North
Africa","NAME_LEN":9,"LONG_LEN":14,"ABBREV_LEN":7,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":4.7,"MIN_LABEL":6,"MAX_LABEL":11,"NE_ID":1159321223,"WIKIDATAID":"Q6250","NAME_AR":"الصحراء الغربية","NAME_BN":"পশ্চিম
সাহারা","NAME_DE":"Westsahara","NAME_EN":"Western Sahara","NAME_ES":"Sahara Occidental","NAME_FR":"Sahara occidental","NAME_EL":"Δυτική Σαχάρα","NAME_HI":"पश्चिमी सहारा","NAME_HU":"Nyugat-
Szahara","NAME_ID":"Sahara Barat","NAME_IT":"Sahara Occidentale","NAME_JA":"西サハラ","NAME_KO":"서사하라","NAME_NL":"Westelijke Sahara","NAME_PL":"Sahara Zachodnia","NAME_PT":"Saara
```

- Mark: this defines the encoding of our map:
    - "type": "geoshape",
    - "fill": "lightgray" (this defines the fill colour of each country)
    - "stroke": "white" (this is the outline colour of each country)

Lines 15–34 define our second layer – earthquake data presented as proportional symbols.

```
{
  "data": {
   "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/2_symbol_map/data/earthquake.csv"
  },
  "mark": {"type": "circle", "tooltip": {"content": "data"}},
  "encoding": {
   "longitude": {"field": "longitude", "type": "quantitative"},
   "latitude": {"field": "latitude", "type": "quantitative"},
   "size": {
     "field": "mag",
     "type": "quantitative",
     "scale": {"domain": [4, 8]}
   },
   "color": {
     "field": "depth",
     "type": "quantitative",
      "scale": {"scheme": "reds"}
     }
    }
   }
```

- “data”: the earthquake data (CSV file). Vega-Lite also supports values separated by tabs and custom delimiters, please see here: https://vega.github.io/vega-lite/docs/data.html
- “mark”: this defines the mark as “circle”, and add tooltips which shows the details of the data. For more custom-built tooltips, please check https://vega.github.io/vega-lite/docs/tooltip.html.
- Encodings:
    - “longitude”: this is the longitude data key required by Vega-Lite to position the circles.
        - field: “longitude” (this is the column name in our CSV data)
    - “latitude”: similar to “longitude”.
    - “size”:
        - Field: the attribute name in our csv file – we select “mag” which is the magnitude of the earthquake
        - Type: the attribute type is quantitative.
        - Scale – domain: we can define the magnitude range that we would like to display on the map.
    - “colour”
        - Field: the attribute name in our csv file – we select “depth” which is the earthquake depth
        - Type: the attribute type is quantitative.
        - Scale – scheme: provide a set of named colour palettes for both discrete and continuous colour encodings. Please check here for more details: https://vega.github.io/vega/docs/schemes/#reference

Exercise:
Try to modify the encodings and check the corresponding results.

3.3 The HTML document and other examples

Please check the example GitHub repository (in Section 3.1) to understand the HTML document. We defined a div called “symbol_map” in our index.html file.

- Styles.css: we control the size and location of the div.
- Data folder: this includes all of our data files.
- JS folder: this includes all of our JavaScript and JSON files. (Some programmers might prefer to put JSON files in the data folder).

If you have any difficulty understanding those files, please check the HTML, CSS and JavaScript tutorial in the previous lab activities.

For more examples of proportional symbol maps or dot maps, please check the following:

- One Dot per Airport in the U.S. Overlayed on Geoshape: https://vega.github.io/vega-lite/examples/geo_layer.html
- One Dot per Zipcode in the U.S.: https://vega.github.io/vega-lite/examples/geo_circle.html

4. Choropleth Maps with Vega-Lite

In this section, we build a choropleth map in Vega-Lite using a COVID-19 dataset. The dataset contains the COVID-19 statistical data of different countries on 10 Oct 2020. The data is available here: https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/data/covid_10_10_2020.csv

The main attributes that we use include:

- Country
- Active, which contains the active cases of each country on 10 Oct 2020
- Confirmed
- Population

We will also use the TopoJSON data (from Section 2.2) to create a global map. The TopoJSON data is available here: https://github.com/KaneSec/vega_lite/blob/main/3_choropleth_map/js/ne_110m_admin_0_countries.topojson

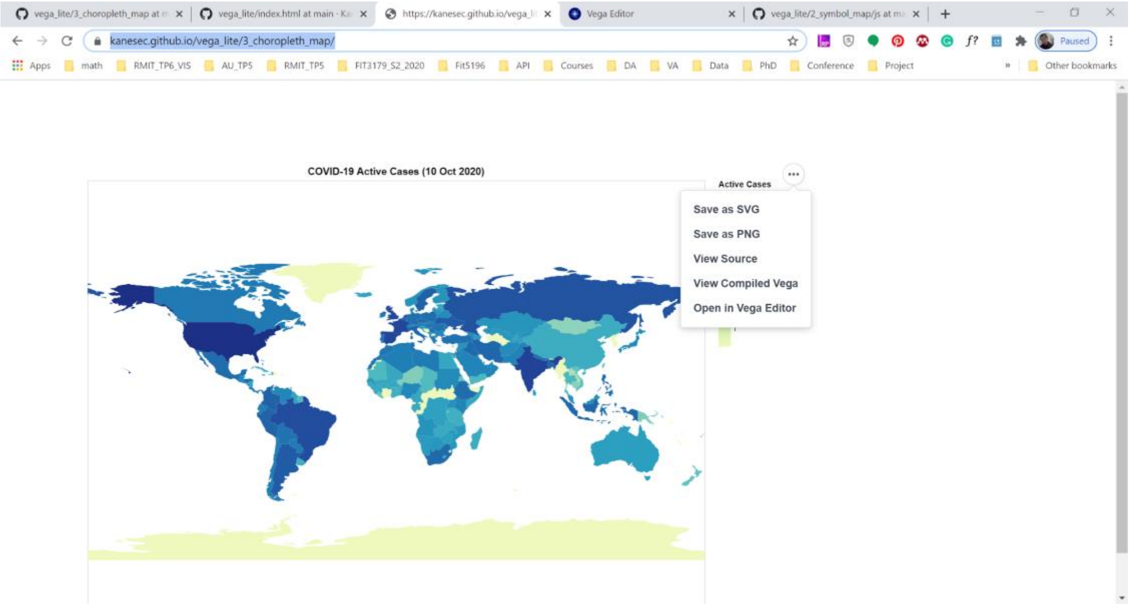
4.1 Overview

The final Choropleth map will look like this: https://kanesec.github.io/vega_lite/3_choropleth_map/

The example GitHub repository is available here: https://github.com/KaneSec/vega_lite/tree/main/3_choropleth_map

4.2 Understand the Vega-Lite JSON file

Here, we walk through the process of creating this map. On the visualisation page at https://kanesec.github.io/vega_lite/3_choropleth_map/, click “Open in Vega Editor” (as shown below).

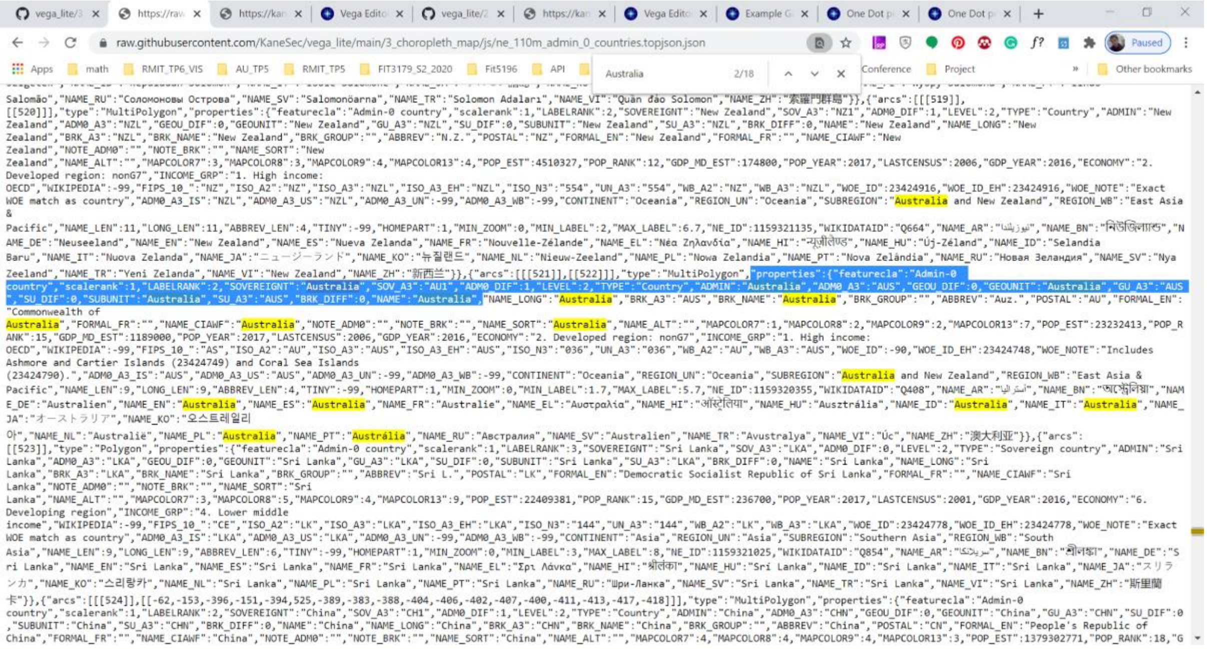

vega_lite/3_choropleth_map at m
vega_lite/index.html at main · Ka
https://kanesec.github.io/vega_li
Vega Editor
vega_lite/2_symbol_map/js at ma
kanesec.github.io/vega_lite/3_choropleth_map/
Paused
Apps
math
RMIT_TP6_VIS
AU_TP5
RMIT_TP5
FIT3179_S2_2020
Fit5196
API
Courses
DA
VA
Data
PhD
Conference
Project
Other bookmarks
COVID-19 Active Cases (10 Oct 2020)
Active Cases
Save as SVG
Save as PNG
View Source
View Compiled Vega
Open in Vega Editor

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "COVID-19 Active Cases (10 Oct 2020)",
  "width": 800,
  "height": 600,
  "projection": {"type": "equirectangular"},
  "data": {
    "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/js/ne_110m_admin_0_countries.topojson",
    "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
  },
  "transform": [
    {
      "lookup": "properties.NAME",
      "from": {
        "data": {
          "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/data/covid_10_10_2020.csv"
        },
        "key": "Country",
        "fields": ["Active"]
      }
    },
    {"calculate": "datum.Active + 0.1", "as": "Active Cases"}
  ],
  "mark": {"type": "geoshape"},
  "encoding": {
    "color": {
      "field": "Active Cases",
      "type": "quantitative",
      "scale": {"type": "log"}
    },
    "tooltip": [
      {"field": "properties.NAME", "type": "nominal", "title": "Country"},
      {"field": "Active", "type": "quantitative"}
    ]
  }
}
```

Lines 2–9: this is the similar as our symbol map (where we defined a base map layer), except that we used the `equirectangular` projection instead of `equalEarth`.

Lines 10–22: This is where we define our data to map. We will map the “country” column from our CSV file to the “country name” defined in the TopoJSON file.

- "lookup": "properties.NAME": this is the key where we stored the country name in our TopoJSON file. To find the name of this data key in your own TopoJSON file, you can search for a data example in your TopoJSON file (e.g., if you are plotting data based on countries, search for a country name – “Australia”; if you are plotting suburbs in Victoria, search for a suburb name – “Caulfield”). In our example, the country name is stored in the dictionary – “properties” > with the key “NAME”.



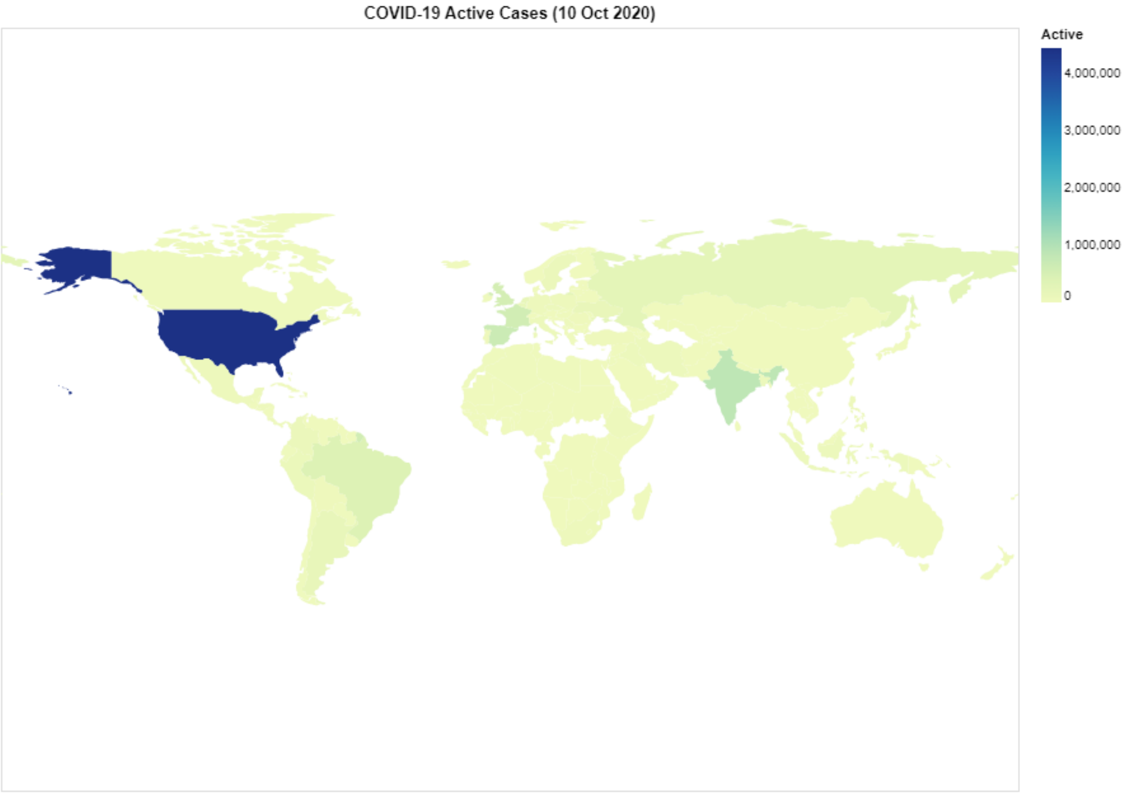
```
Salomão","NAME_RU":"Соломоновы Острова","NAME_SV":"Salomonöarna","NAME_TR":"Solomon Adaları","NAME_VI":"Quần đảo Solomon","NAME_ZH":"索羅門群島"}},{"arcs":[[[519]],
[[520]]],"type":"MultiPolygon","properties":{"featurecla":"Admin-0 country","scalerank":1,"LABELRANK":2,"SOVEREIGNT":"New Zealand","SOV_A3":"NZ1","ADM0_DIF":1,"LEVEL":2,"TYPE":"Country","ADMIN":"New
Zealand","ADM0_A3":"NZL","GEOU_DIF":0,"GEOUNIT":"New Zealand","GU_A3":"NZL","SU_DIF":0,"SUBUNIT":"New Zealand","SU_A3":"NZL","BRK_DIFF":0,"NAME":"New Zealand","NAME_LONG":"New
Zealand","BRK_A3":"NZL","BRK_NAME":"New Zealand","BRK_GROUP":"","ABBREV":"N.Z.","POSTAL":"NZ","FORMAL_EN":"New Zealand","FORMAL_FR":"","NAME_CIAWF":"New
Zealand","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"New
Zealand","NAME_ALT":"","MAPCOLOR7":3,"MAPCOLOR8":3,"MAPCOLOR9":4,"MAPCOLOR13":4,"POP_EST":4510327,"POP_RANK":12,"GDP_MD_EST":174800,"POP_YEAR":2017,"LASTCENSUS":2006,"GDP_YEAR":2016,"ECONOMY":"2.
Developed region: nonG7","INCOME_GRP":"1. High income:
OECD","WIKIPEDIA":-99,"FIPS_10_":"NZ","ISO_A2":"NZ","ISO_A3":"NZL","ISO_A3_EH":"NZL","ISO_N3":"554","UN_A3":"554","WB_A2":"NZ","WB_A3":"NZL","WOE_ID":23424916,"WOE_ID_EH":23424916,"WOE_NOTE":"Exact
WOE match as country","ADM0_A3_IS":"NZL","ADM0_A3_US":"NZL","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Oceania","REGION_UN":"Oceania","SUBREGION":"Australia and New Zealand","REGION_WB":"East Asia
&
Pacific","NAME_LEN":11,"LONG_LEN":11,"ABBREV_LEN":4,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":0,"MIN_LABEL":2,"MAX_LABEL":6.7,"NE_ID":1159321135,"WIKIDATAID":"Q664","NAME_AR":"نيوزيلندا","NAME_BN":"নিউজিল্যান্ড","N
AME_DE":"Neuseeland","NAME_EN":"New Zealand","NAME_ES":"Nueva Zelanda","NAME_FR":"Nouvelle-Zélande","NAME_EL":"Νέα Ζηλανδία","NAME_HI":"न्यूज़ीलैण्ड","NAME_HU":"Új-Zéland","NAME_ID":"Selandia
Baru","NAME_IT":"Nuova Zelanda","NAME_JA":"ニュージーランド","NAME_KO":"뉴질랜드","NAME_NL":"Nieuw-Zeeland","NAME_PL":"Nowa Zelandia","NAME_PT":"Nova Zelândia","NAME_RU":"Новая Зеландия","NAME_SV":"Nya
Zeeland","NAME_TR":"Yeni Zelanda","NAME_VI":"New Zealand","NAME_ZH":"新西兰"}},{"arcs":[[[521]],[[522]]],"type":"MultiPolygon","properties":{"featurecla":"Admin-0
country","scalerank":1,"LABELRANK":2,"SOVEREIGNT":"Australia","SOV_A3":"AU1","ADM0_DIF":1,"LEVEL":2,"TYPE":"Country","ADMIN":"Australia","ADM0_A3":"AUS","GEOU_DIF":0,"GEOUNIT":"Australia","GU_A3":"AUS
","SU_DIF":0,"SUBUNIT":"Australia","SU_A3":"AUS","BRK_DIFF":0,"NAME":"Australia","NAME_LONG":"Australia","BRK_A3":"AUS","BRK_NAME":"Australia","BRK_GROUP":"","ABBREV":"Auz.","POSTAL":"AU","FORMAL_EN":
"Commonwealth of
Australia","FORMAL_FR":"","NAME_CIAWF":"Australia","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"Australia","NAME_ALT":"","MAPCOLOR7":1,"MAPCOLOR8":2,"MAPCOLOR9":2,"MAPCOLOR13":7,"POP_EST":23232413,"POP_R
ANK":15,"GDP_MD_EST":1189000,"POP_YEAR":2017,"LASTCENSUS":2006,"GDP_YEAR":2016,"ECONOMY":"2. Developed region: nonG7","INCOME_GRP":"1. High income:
OECD","WIKIPEDIA":-99,"FIPS_10_":"AS","ISO_A2":"AU","ISO_A3":"AUS","ISO_A3_EH":"AUS","ISO_N3":"036","UN_A3":"036","WB_A2":"AU","WB_A3":"AUS","WOE_ID":-90,"WOE_ID_EH":23424748,"WOE_NOTE":"Includes
Ashmore and Cartier Islands (23424749) and Coral Sea Islands
(23424790).","ADM0_A3_IS":"AUS","ADM0_A3_US":"AUS","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Oceania","REGION_UN":"Oceania","SUBREGION":"Australia and New Zealand","REGION_WB":"East Asia &
Pacific","NAME_LEN":9,"LONG_LEN":9,"ABBREV_LEN":4,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":0,"MIN_LABEL":1.7,"MAX_LABEL":5.7,"NE_ID":1159320355,"WIKIDATAID":"Q408","NAME_AR":"أستراليا","NAME_BN":"অস্ট্রেলিয়া","NAM
E_DE":"Australien","NAME_EN":"Australia","NAME_ES":"Australia","NAME_FR":"Australie","NAME_EL":"Αυστραλία","NAME_HI":"ऑस्ट्रेलिया","NAME_HU":"Ausztrália","NAME_ID":"Australia","NAME_IT":"Australia","NAME_
JA":"オーストラリア","NAME_KO":"오스트레일리
아","NAME_NL":"Australië","NAME_PL":"Australia","NAME_PT":"Austrália","NAME_RU":"Австралия","NAME_SV":"Australien","NAME_TR":"Avustralya","NAME_VI":"Úc","NAME_ZH":"澳大利亚"}},{"arcs":
[[523]],"type":"Polygon","properties":{"featurecla":"Admin-0 country","scalerank":1,"LABELRANK":3,"SOVEREIGNT":"Sri Lanka","SOV_A3":"LKA","ADM0_DIF":0,"LEVEL":2,"TYPE":"Sovereign country","ADMIN":"Sri
Lanka","ADM0_A3":"LKA","GEOU_DIF":0,"GEOUNIT":"Sri Lanka","GU_A3":"LKA","SU_DIF":0,"SUBUNIT":"Sri Lanka","SU_A3":"LKA","BRK_DIFF":0,"NAME":"Sri Lanka","NAME_LONG":"Sri
Lanka","BRK_A3":"LKA","BRK_NAME":"Sri Lanka","BRK_GROUP":"","ABBREV":"Sri L.","POSTAL":"LK","FORMAL_EN":"Democratic Socialist Republic of Sri Lanka","FORMAL_FR":"","NAME_CIAWF":"Sri
Lanka","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"Sri
Lanka","NAME_ALT":"","MAPCOLOR7":3,"MAPCOLOR8":5,"MAPCOLOR9":4,"MAPCOLOR13":9,"POP_EST":22409381,"POP_RANK":15,"GDP_MD_EST":236700,"POP_YEAR":2017,"LASTCENSUS":2001,"GDP_YEAR":2016,"ECONOMY":"6.
Developing region","INCOME_GRP":"4. Lower middle
income","WIKIPEDIA":-99,"FIPS_10_":"CE","ISO_A2":"LK","ISO_A3":"LKA","ISO_A3_EH":"LKA","ISO_N3":"144","UN_A3":"144","WB_A2":"LK","WB_A3":"LKA","WOE_ID":23424778,"WOE_ID_EH":23424778,"WOE_NOTE":"Exact
WOE match as country","ADM0_A3_IS":"LKA","ADM0_A3_US":"LKA","ADM0_A3_UN":-99,"ADM0_A3_WB":-99,"CONTINENT":"Asia","REGION_UN":"Asia","SUBREGION":"Southern Asia","REGION_WB":"South
Asia","NAME_LEN":9,"LONG_LEN":9,"ABBREV_LEN":6,"TINY":-99,"HOMEPART":1,"MIN_ZOOM":0,"MIN_LABEL":3,"MAX_LABEL":8,"NE_ID":1159321025,"WIKIDATAID":"Q854","NAME_AR":"سريلانكا","NAME_BN":"শ্রীলঙ্কা","NAME_DE":"S
ri Lanka","NAME_EN":"Sri Lanka","NAME_ES":"Sri Lanka","NAME_FR":"Sri Lanka","NAME_EL":"Σρι Λάνκα","NAME_HI":"श्रीलंका","NAME_HU":"Srí Lanka","NAME_ID":"Sri Lanka","NAME_IT":"Sri Lanka","NAME_JA":"スリラ
ンカ","NAME_KO":"스리랑카","NAME_NL":"Sri Lanka","NAME_PL":"Sri Lanka","NAME_PT":"Sri Lanka","NAME_RU":"Шри-Ланка","NAME_SV":"Sri Lanka","NAME_TR":"Sri Lanka","NAME_VI":"Sri Lanka","NAME_ZH":"斯里蘭
卡"}},{"arcs":[[[524]],[[-62,-153,-396,-151,-394,525,-389,-383,-388,-404,-406,-402,-407,-400,-411,-413,-417,-418]]],"type":"MultiPolygon","properties":{"featurecla":"Admin-0
country","scalerank":1,"LABELRANK":2,"SOVEREIGNT":"China","SOV_A3":"CH1","ADM0_DIF":1,"LEVEL":2,"TYPE":"Country","ADMIN":"China","ADM0_A3":"CHN","GEOU_DIF":0,"GEOUNIT":"China","GU_A3":"CHN","SU_DIF":0
,"SUBUNIT":"China","SU_A3":"CHN","BRK_DIFF":0,"NAME":"China","NAME_LONG":"China","BRK_A3":"CHN","BRK_NAME":"China","BRK_GROUP":"","ABBREV":"China","POSTAL":"CN","FORMAL_EN":"People's Republic of
China","FORMAL_FR":"","NAME_CIAWF":"China","NOTE_ADM0":"","NOTE_BRK":"","NAME_SORT":"China","NAME_ALT":"","MAPCOLOR7":4,"MAPCOLOR8":4,"MAPCOLOR9":4,"MAPCOLOR13":3,"POP_EST":1379302771,"POP_RANK":18,"G
```

- "from": this is related to the CSV data to be plotted
  - "data": "url": The location of our CSV file.
  - "key": "Country": This is the column name in our CSV file that is going to map the country name in our TopoJSON file.
  - "fields": ["Active"]: This contains all the column names (from the CSV file) that we use in our visualisation mappings. Here we only included the active cases.

Log and Linear scales:

While the remaining lines are easy to understand, please note that we defined a log scale for the colour mapping. The codes and visualisation without the log scale are shown below. Please check the visualisation and think about it – why is the log scale better than a linear scale in this case? Also, why do we include “datum.Active + 0.1” in line 24 of the original codes? Please remove this “+0.1” part and re-check the visualisation output.

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "COVID-19 Active Cases (10 Oct 2020)",
  "width": 800,
  "height": 600,
  "projection": {"type": "equirectangular"},
  "data": {
    "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/js/ne_110m_admin_0_countries.topojson",
    "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
  },
  "transform": [
    {
      "lookup": "properties.NAME",
      "from": {
        "data": {
          "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/data/covid_10_10_2020.csv"
        },
        "key": "Country",
        "fields": ["Active"]
      }
    }
  ],
  "mark": {"type": "geoshape"},
  "encoding": {
    "color": {
      "field": "Active",
      "type": "quantitative"
    },
    "tooltip": [
      {"field": "properties.NAME", "type": "nominal", "title": "Country"},
      {"field": "Active", "type": "quantitative"}
    ]
  }
}
```



Besides the log scale, Vega-Lite also supports linear, ordinal, band, and point scales. Please check here for more details: https://vega.github.io/vega-lite/docs/scale.html

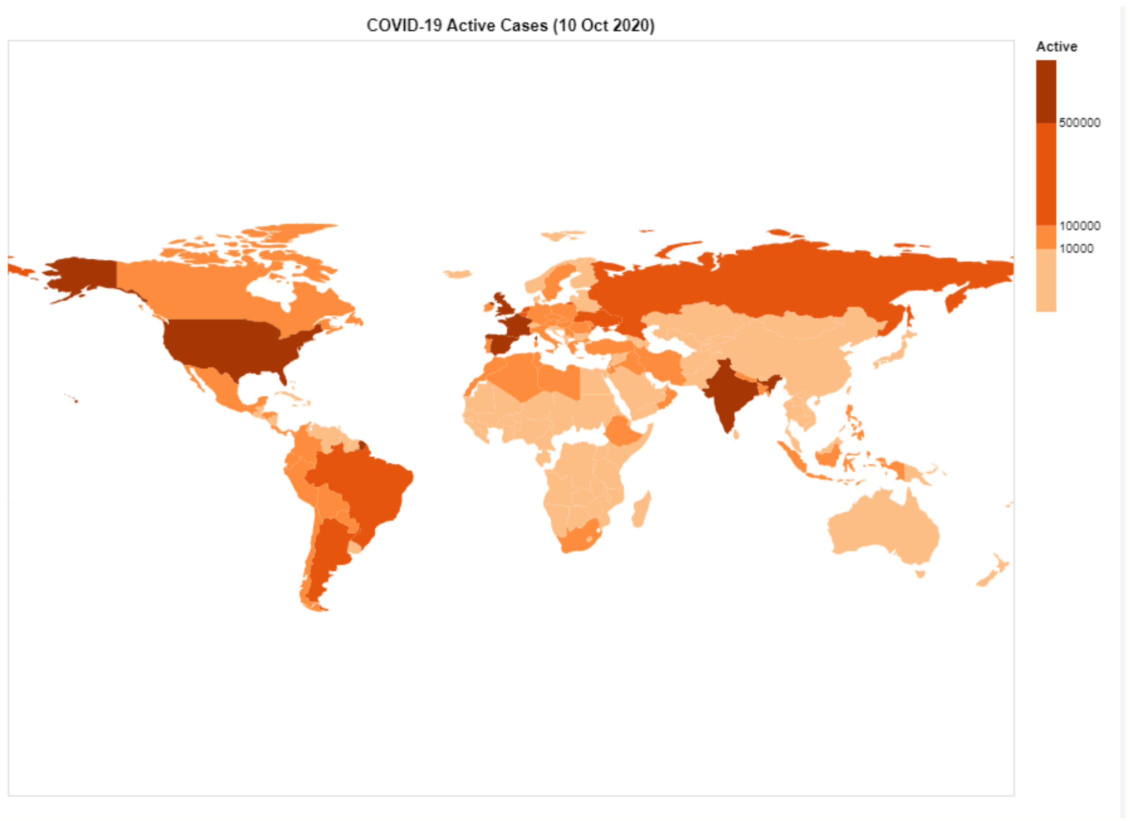
Customised classification and colour scale

We can also customise the classification and the colours. This can be implemented with scale in the colour encoding part:

```
      "scale": {
        "type": "threshold",
        "domain": [10000, 100000, 500000],
        "range": ["#fdbe85", "#fd8d3c", "#e6550d", "#a63603"]
      }
```

Based on three thresholds defined in the “domain”, we divide the number of cases into four ranges [0, 10000], [10000, 100000], [100000, 500000], [500000, infinity]. “range” defines the four colours mapped with these four value ranges.

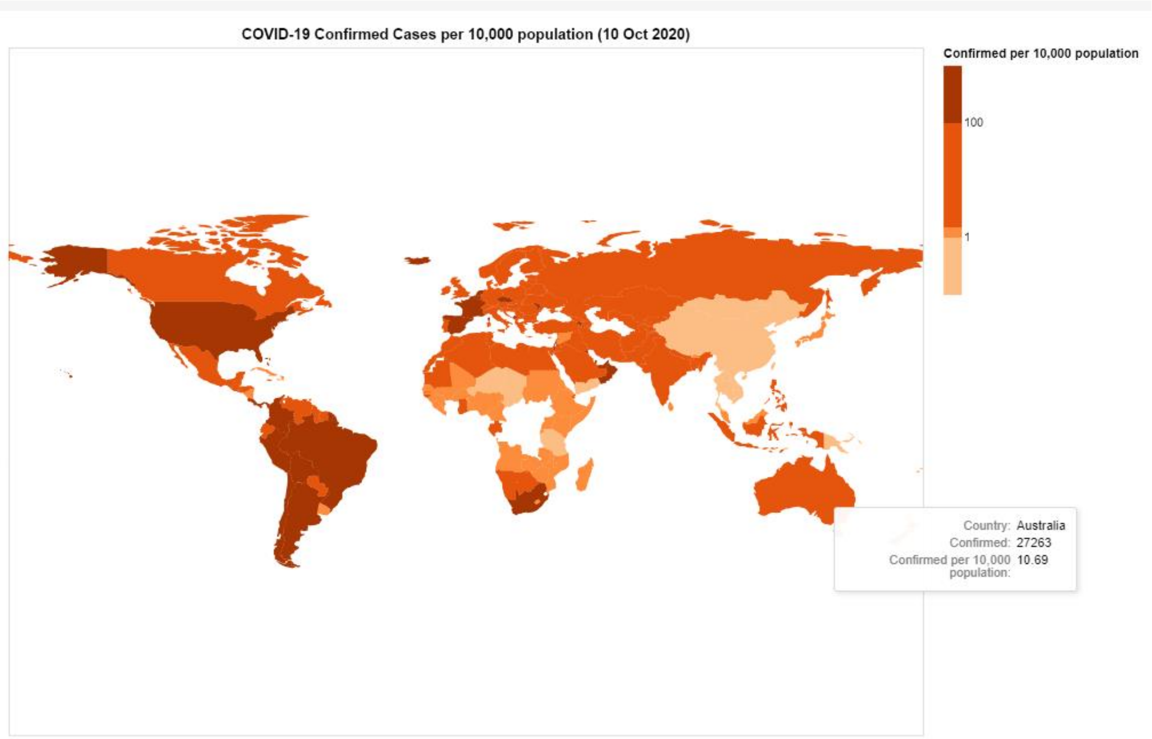
Please check the visualisation result and the complete Vega-Lite code below:



```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "COVID-19 Active Cases (10 Oct 2020)",
  "width": 800,
  "height": 600,
  "projection": {"type": "equirectangular"},
  "data": {
    "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth
_map/js/ne_110m_admin_0_countries.topojson",
    "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
  },
  "transform": [
    {
      "lookup": "properties.NAME",
      "from": {
        "data": {
          "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_chor
opleth_map/data/covid_10_10_2020.csv"
        },
        "key": "Country",
        "fields": ["Active"]
      }
    }
  ],
  "mark": {"type": "geoshape"},
  "encoding": {
    "color": {
      "field": "Active",
      "type": "quantitative",
      "scale": {
        "type": "threshold",
        "domain": [10000, 100000, 500000],
        "range": ["#fdbe85", "#fd8d3c", "#e6550d", "#a63603"]
      }
    },
    "tooltip": [
      {"field": "properties.NAME", "type": "nominal", "title": "Country"},
      {"field": "Active", "type": "quantitative"}
    ]
  }
}
```

Normalise cases by population

The choropleth maps created so far are clearly misleading, because they do not normalise data. We need to normalise the numbers based on the population of each country. In Vega-Lite, we can simply do a calculation when loading the data, and then plot the choropleth map based on this calculation. The following example presents the total confirmed cases per 10,000 population of each country. The colour is mapped based on four ranges [0,1], [1, 10], [10,100] and [100, indefinity]. The result and the code are shown below.


COVID-19 Confirmed Cases per 10,000 population (10 Oct 2020)
Confirmed per 10,000 population
100
1
Country: Australia
Confirmed: 27263
Confirmed per 10,000 population: 10.69

```
{
  "$schema": "https://vega.github.io/schema/vega-lite/v5.json",
  "title": "COVID-19 Active Cases (10 Oct 2020)",
  "width": 800,
  "height": 600,
  "projection": {"type": "equirectangular"},
  "data": {
    "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/js/ne_110m_admin_0_countries.topojson",
    "format": {"type": "topojson", "feature": "ne_110m_admin_0_countries"}
  },
  "transform": [
    {
      "lookup": "properties.NAME",
      "from": {
        "data": {
          "url": "https://raw.githubusercontent.com/KaneSec/vega_lite/main/3_choropleth_map/data/covid_10_10_2020.csv"
        },
        "key": "Country",
        "fields": ["Active", "Confirmed", "Deaths", "Population"]
      }
    },
    {"calculate": "datum.Confirmed/datum.Population * 10000", "as": "Confirmed per 10,000 population"}
  ],
  "mark": {"type": "geoshape"},
  "encoding": {
    "color": {
      "field": "Confirmed per 10,000 population",
      "type": "quantitative",
      "scale": {
        "type": "threshold",
        "domain": [1, 10, 100],
        "range": ["#fdbe85", "#fd8d3c", "#e6550d", "#a63603"]
      }
    },
    "tooltip": [
      {"field": "properties.NAME", "type": "nominal", "title": "Country"},
      {"field": "Confirmed", "type": "quantitative"},
      {"field": "Confirmed per 10,000 population", "type": "quantitative", "format": ".2f"}
    ]
  }
}
```

Question:
Did you notice that there are some countries (e.g., Congo) shown as blank on the map? Discuss with the tutor and your peers:

- What could be the cause of this?
- What is the best way to handle this?

4.3 The HTML document and other examples

Please check the example GitHub repository (in Section 4.1) to understand the HTML document. We defined a div called “choropleth_map” in our index.html file.

- Styles.css: we control the size and location of the div.
- Data folder: this includes all of our data files.
- JS folder: this includes all of our JavaScript and JSON files. (Some programmers like to put JSON files in the data folder).

For more examples of choropleth maps, please check the following:

- Choropleth of Unemployment Rate per County: https://vega.github.io/vega-lite/examples/geo_choropleth.html
- Other examples: https://vega.github.io/vega-lite/examples/#maps-geographic-displays