An Phríomh-Oifig Staidrimh
Central Statistics Office

Mapping a Table of Data with Esri Shapefiles in R

Tanzania

An introductory tutorial to mapping opensource data with two shapefiles using the "*ggplot2*" and "tmap" packages in R

Kevin McCormack

Dr. Mary Smyth

Sinead Phelan

October 2018

1. Introduction

In this tutorial we discuss how to join Tanzanian GDP data, in CSV format, with two Esri Shapefiles, all opensource, and construct the map in Figure 1, within the R environment.

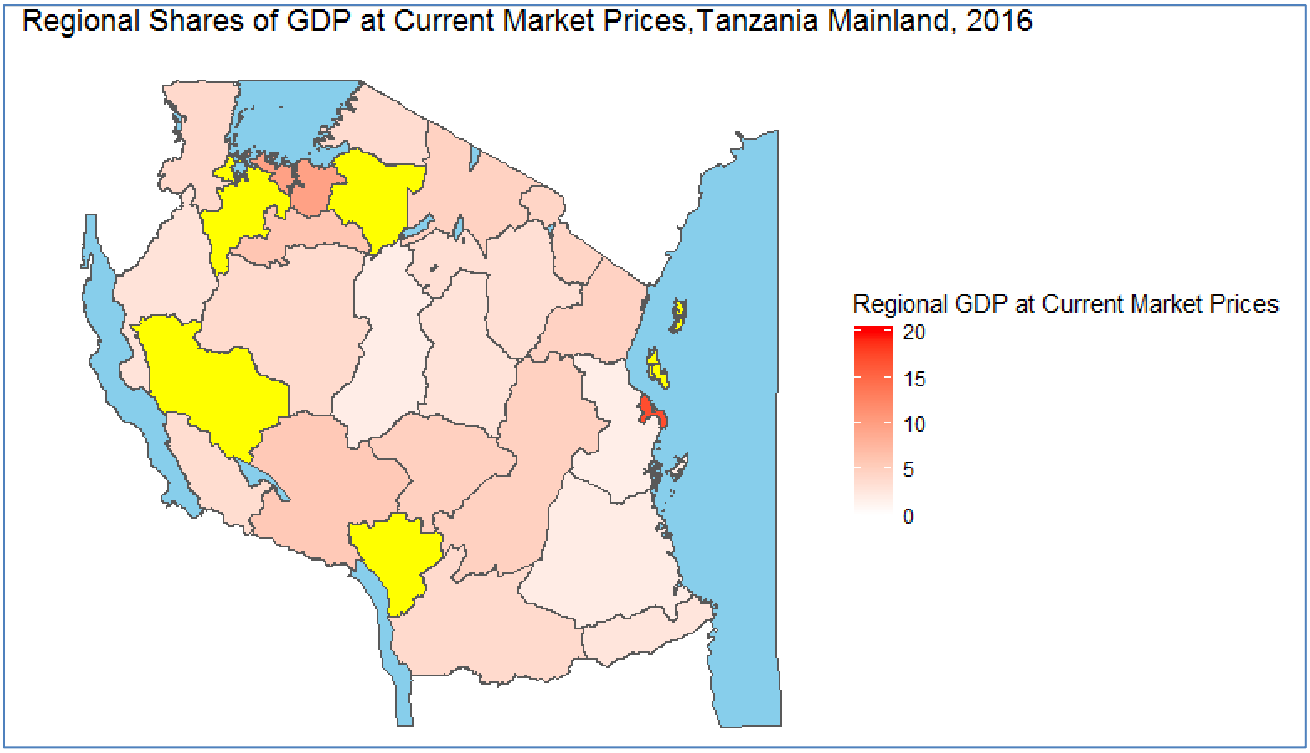

Regional Shares of GDP at Current Market Prices,Tanzania Mainland, 2016
Regional GDP at Current Market Prices
20
15
10
5
0

Figure 1: Geospatial presentation of GDP data.

The following steps are involved in the construction of this map:

- creating a CSV file with GDP data extracted from an official Tanzanian publication,
- create an R data-frame from this data,
- add a geospatial reference to this data-frame,
- download two Esri shapefiles, regions and water bodies, from the Tanzanian National Bureau of Statistics' (TNBS) website,
- create geospatial data-frames from the Esri shapefiles,
- join the region geospatial data-frame to the GDP data-frame, and
- plot the data using the "*ggplot2"* R package, and overlay the region and water bodies geospatial data frames.

2. Data Source - Tanzanian National Bureau of Statistics

The Tanzanian National Bureau of Statistics (TNBS) has been established as an autonomous public office by the Statistics Act, 2015 and has the mandate to provide official statistics to the Government, business community and the public at large. The Act also gives NBS the mandate to play the role as a co-coordinating agency, within the National Statistical System (NSS) to ensure that quality official statistics is produced. Before the enactment of the Statistics Act of 2015, the NBS was one of the Government Executive Agencies which was established on the 26th March,1999 under the Executive Agencies Act, 1997.

3. Data

For this tutorial we will be using the data extracted from "*Table 21: Regional Shares of GDP at Current Market Prices***,** of the TNBS publication "*National Accounts of Tanzania Mainland 2008 – 2016*". (See Section 10)

http://www.nbs.go.tz/nbs/takwimu/na/National_Accounts_Statistics_of_Tanzania%20_Mainland_2016.pdf

The data was manually extracted from the above PDF file to the following CSV file

"*Tanzania RegionCodes GDP Pct.csv*"

Note that the **Region** column is referenced as "*Region_Num*" to allow for linking to the *Regions* Esri shapefile later in this tutorial. Data is not available for all regions. However, in order for the map to include the shape of all regions in Tanzania, references for the missing regions are added to the end of the data table.

4. R-studio

R is a language and environment for statistical computing and graphics and is rapidly becoming the leading programming language in statistics and data analytics.

It is recommended to use R-studio, which, provides popular open source and enterprise-ready professional software for the R statistical computing environment.

R-studio can be downloaded here: https://www.rstudio.com/

5. Reading the GDP data table using R

5.1 Setting the working directory

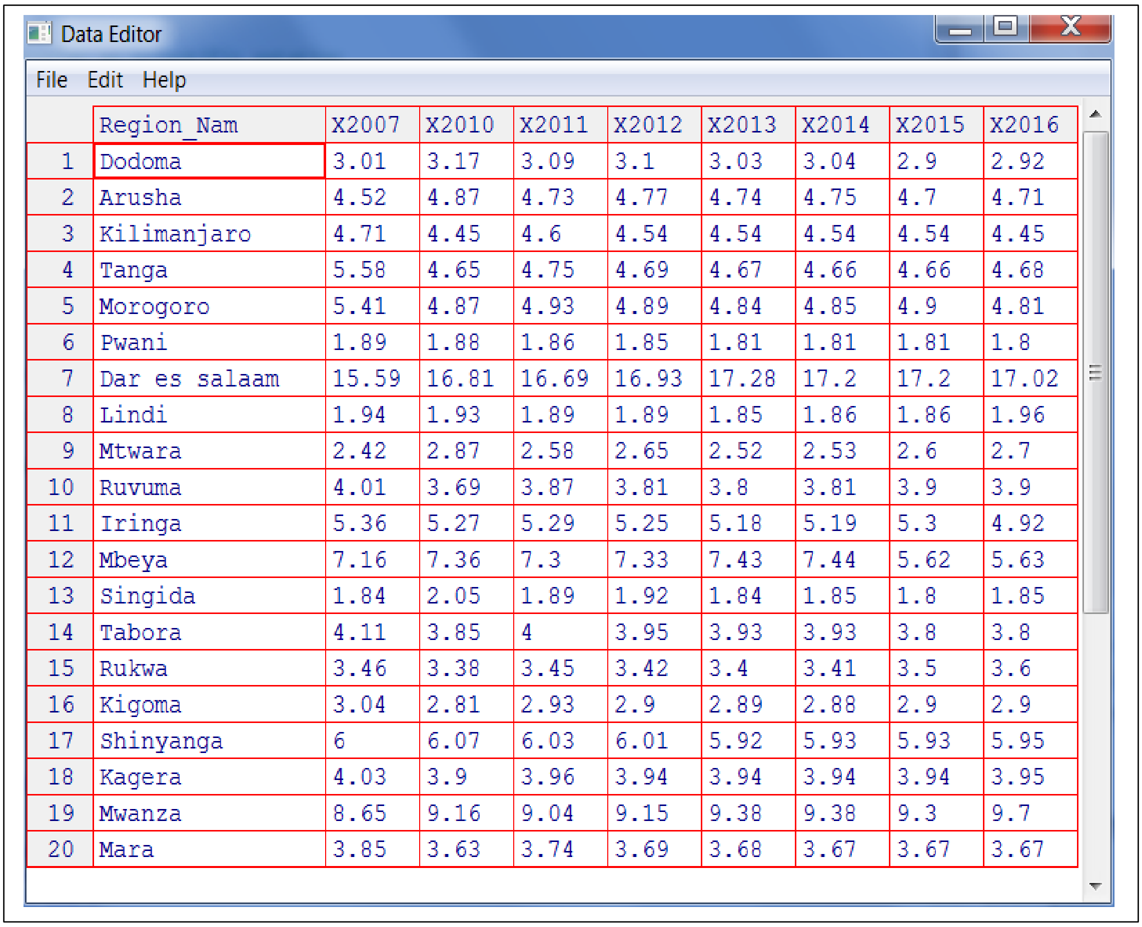
Firstly, set the working directory, for example:

setwd("C:/My Documents/R/SDG/Tanzania/")

5.2 Reading the CSV file

Next the CSV file is read in and an R data-frame is created, which is referenced as “*region*”.

region <- read.csv("Tanzania RegionCodes GDP Pct.csv")



	Region_Nam	X2007	X2010	X2011	X2012	X2013	X2014	X2015	X2016
1	Dodoma	3.01	3.17	3.09	3.1	3.03	3.04	2.9	2.92
2	Arusha	4.52	4.87	4.73	4.77	4.74	4.75	4.7	4.71
3	Kilimanjaro	4.71	4.45	4.6	4.54	4.54	4.54	4.54	4.45
4	Tanga	5.58	4.65	4.75	4.69	4.67	4.66	4.66	4.68
5	Morogoro	5.41	4.87	4.93	4.89	4.84	4.85	4.9	4.81
6	Pwani	1.89	1.88	1.86	1.85	1.81	1.81	1.81	1.8
7	Dar es salaam	15.59	16.81	16.69	16.93	17.28	17.2	17.2	17.02
8	Lindi	1.94	1.93	1.89	1.89	1.85	1.86	1.86	1.96
9	Mtwara	2.42	2.87	2.58	2.65	2.52	2.53	2.6	2.7
10	Ruvuma	4.01	3.69	3.87	3.81	3.8	3.81	3.9	3.9
11	Iringa	5.36	5.27	5.29	5.25	5.18	5.19	5.3	4.92
12	Mbeya	7.16	7.36	7.3	7.33	7.43	7.44	5.62	5.63
13	Singida	1.84	2.05	1.89	1.92	1.84	1.85	1.8	1.85
14	Tabora	4.11	3.85	4	3.95	3.93	3.93	3.8	3.8
15	Rukwa	3.46	3.38	3.45	3.42	3.4	3.41	3.5	3.6
16	Kigoma	3.04	2.81	2.93	2.9	2.89	2.88	2.9	2.9
17	Shinyanga	6	6.07	6.03	6.01	5.92	5.93	5.93	5.95
18	Kagera	4.03	3.9	3.96	3.94	3.94	3.94	3.94	3.95
19	Mwanza	8.65	9.16	9.04	9.15	9.38	9.38	9.3	9.7
20	Mara	3.85	3.63	3.74	3.69	3.68	3.67	3.67	3.67

Figure 2: “*region”* data-frame

6. Esri Shapefiles - Tanzania

The Esri shape files used in this tutorial are called *Regions.shp* and *Water_body.shp.*

They were downloaded to the working directory from the TNBS website.

http://www.nbs.go.tz/nbstz/index.php/english/statistics-by-subject/population-and-housing-census/258-2012-phc-shapefiles-level-one-and-two

7. Creating an R data-frame from an Esri Shapefile.

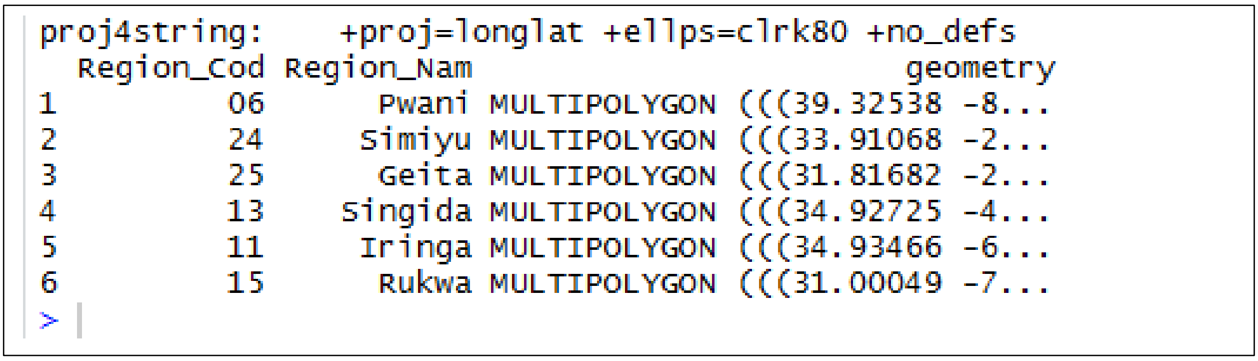
To create an R data-frame from an Esri Shapefile one first needs to load the following library:

library(sf)

7.1 Regions shapefile

Using the *st_read( )* function the Esri *Regions* shapefile is read in as a data frame, "*TanzmyDF*".

TanzmyDF <- st_read("C:/My Documents/R/SDG/Tanzania/Regions.shp", stringsAsFactors = FALSE)



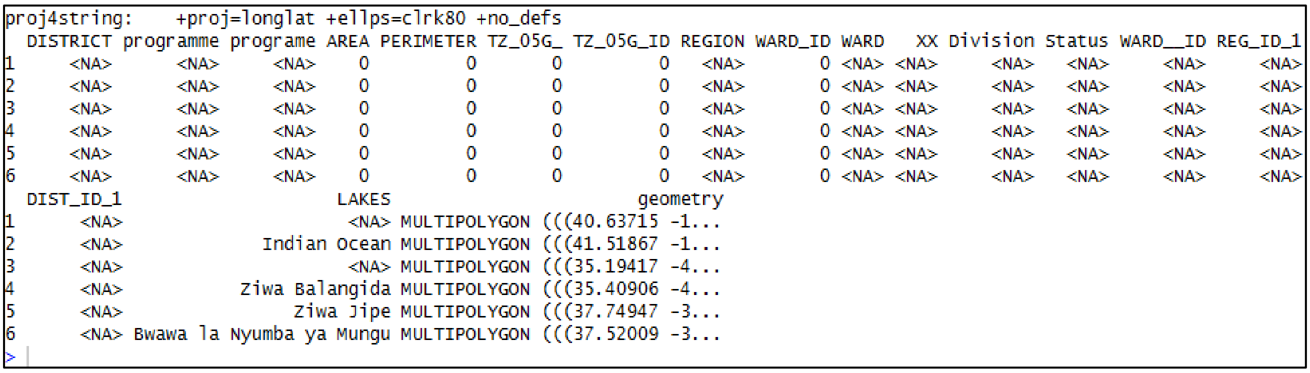
```
proj4string:    +proj=longlat +ellps=clrk80 +no_defs
  Region_Cod Region_Nam                       geometry
1         06      Pwani MULTIPOLYGON (((39.32538 -8...
2         24     Simiyu MULTIPOLYGON (((33.91068 -2...
3         25      Geita MULTIPOLYGON (((31.81682 -2...
4         13    Singida MULTIPOLYGON (((34.92725 -4...
5         11     Iringa MULTIPOLYGON (((34.93466 -6...
6         15      Rukwa MULTIPOLYGON (((31.00049 -7...
>
```

Figure 3: TanzmyDF dataframe.

7.2 Water Bodies shapefile

Using the *st_read( )* function the Esri *water_body* shapefile is read in as a data frame, "*TanzmyDFW*".

TanzmyDFW <- st_read("C:/My Documents/R/SDG/Tanzania/Water_body.shp", stringsAsFactors = FALSE)



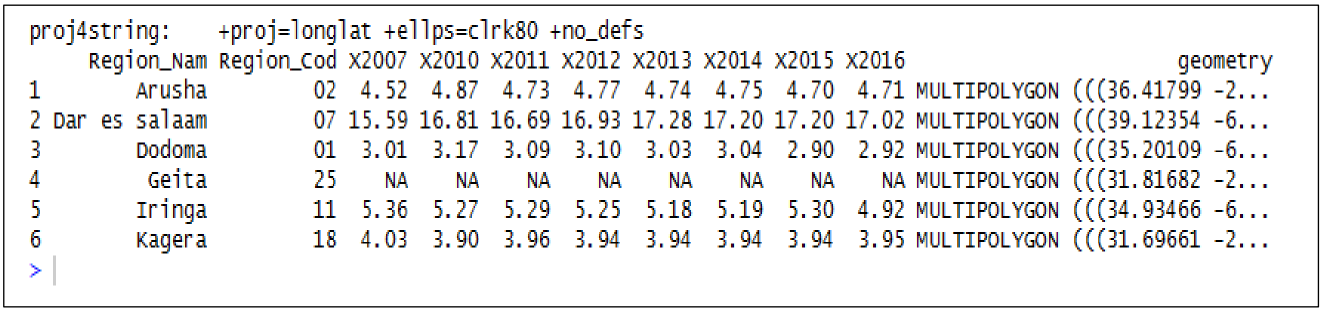
```
proj4string:    +proj=longlat +ellps=clrk80 +no_defs
  DISTRICT programme programe AREA PERIMETER TZ_05G_ TZ_05G_ID REGION WARD_ID WARD   XX Division Status WARD__ID REG_ID_1
1     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
2     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
3     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
4     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
5     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
6     <NA>      <NA>     <NA>    0         0       0         0   <NA>       0 <NA> <NA>     <NA>   <NA>     <NA>     <NA>
  DIST_ID_1                    LAKES                       geometry
1      <NA>                     <NA> MULTIPOLYGON (((40.63715 -1...
2      <NA>             Indian Ocean MULTIPOLYGON (((41.51867 -1...
3      <NA>                     <NA> MULTIPOLYGON (((35.19417 -4...
4      <NA>           Ziwa Balangida MULTIPOLYGON (((35.40906 -4...
5      <NA>                Ziwa Jipe MULTIPOLYGON (((37.74947 -3...
6      <NA> Bwawa la Nyumba ya Mungu MULTIPOLYGON (((37.52009 -3...
> 
```

Figure 4: TanzmyDFW dataframe.

8. Merging the geospatial and GDP data frames, TanzmyDF and region

The merge( ) function is use to join geospatial and GDP data to create a data-frame titled *TanzReg*, using the *'Region_Nam' reference.*

TanzReg <- merge(TanzmyDF, region, by='Region_Nam')



```
proj4string:    +proj=longlat +ellps=clrk80 +no_defs
     Region_Nam Region_Cod X2007 X2010 X2011 X2012 X2013 X2014 X2015 X2016                       geometry
1        Arusha         02  4.52  4.87  4.73  4.77  4.74  4.75  4.70  4.71 MULTIPOLYGON (((36.41799 -2...
2 Dar es salaam         07 15.59 16.81 16.69 16.93 17.28 17.20 17.20 17.02 MULTIPOLYGON (((39.12354 -6...
3        Dodoma         01  3.01  3.17  3.09  3.10  3.03  3.04  2.90  2.92 MULTIPOLYGON (((35.20109 -6...
4         Geita         25    NA    NA    NA    NA    NA    NA    NA    NA MULTIPOLYGON (((31.81682 -2...
5        Iringa         11  5.36  5.27  5.29  5.25  5.18  5.19  5.30  4.92 MULTIPOLYGON (((34.93466 -6...
6        Kagera         18  4.03  3.90  3.96  3.94  3.94  3.94  3.94  3.95 MULTIPOLYGON (((31.69661 -2...
> 
```

Figure 5: *TanzReg* data-frame

9. Mapping the GDP data.

There are quite a number of R libraries that can be use in mapping geospatial data and in this tutorial the *ggplot( )* function from the *ggplot2* library is used. First load the library:

library(ggplot2)

The plot is build up using this function followed by a *"+"* to build up the layers.

The R code below creates a map referenced as "*Tanzania*" as follows:

- selects the *TanzReg* data to be mapped - *ggplot,*
- identifies the year 2016 as the fill - *geom_sf*
- provides the colours & limits for the scale fill and the name of the scale – *scale_fill_gradient2*, and where there is no data, the region is coloured yellow, "*na.value*".
- provides a title - *ggtitle*
- removes the default axis titles and axis ticks – *theme.*

```
(Tanzania <- ggplot(TanzReg) + # input data
  geom_sf(aes(fill=TanzReg$`2016`)) +
      scale_fill_gradient2 (low = "white", high = "red", # colours
                    limits = c(0, 20), #limts
                    na.value = "yellow", # colour when there are not values
                    name = "Regional GDP at Current Market Prices") + # legend options
  ggtitle("Regional Shares of GDP at Current Market Prices,Tanzania Mainland, 2016") +
  theme(plot.title = element_text(hjust = 0.1)) +
  theme(axis.text = element_blank(), # change the theme options
       axis.title = element_blank(), # remove axis titles
       axis.ticks = element_blank() ))# remove axis ticks

Tanzania + geom_sf(data=TanzmyDFW, fill ="skyblue") + blank()
```

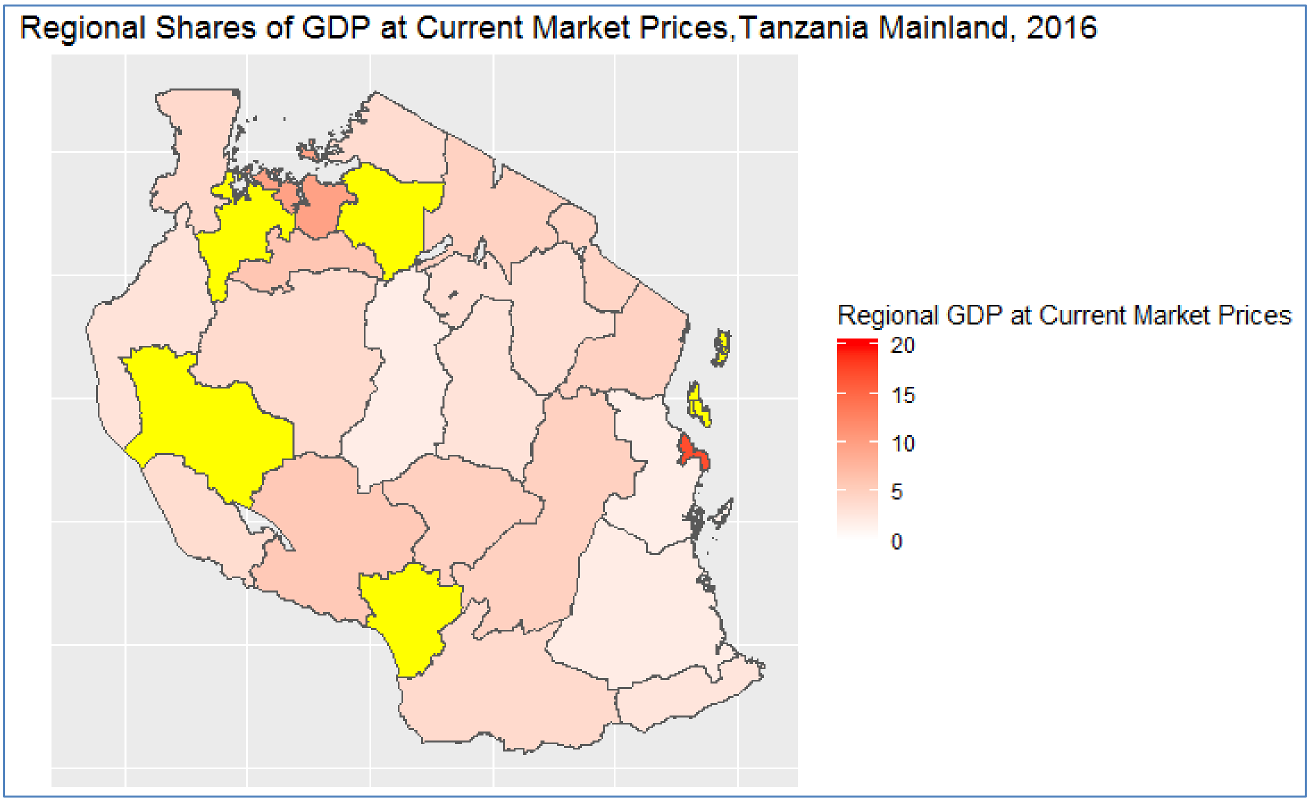

Regional Shares of GDP at Current Market Prices,Tanzania Mainland, 2016
Regional GDP at Current Market Prices
20
15
10
5
0

Figure 6: Regional GDP, 2016 – no water bodies

9.1 Overlaying the water bodies map

In order to improve the look of the map, we will overlay the water bodies data-frame and add a blank background, using the following code.

```
  Tanzania # original map +
 geom_sf(data=TanzmyDFW, fill ="skyblue") +  # overlay the water-bodies
blank()  # blank background
```

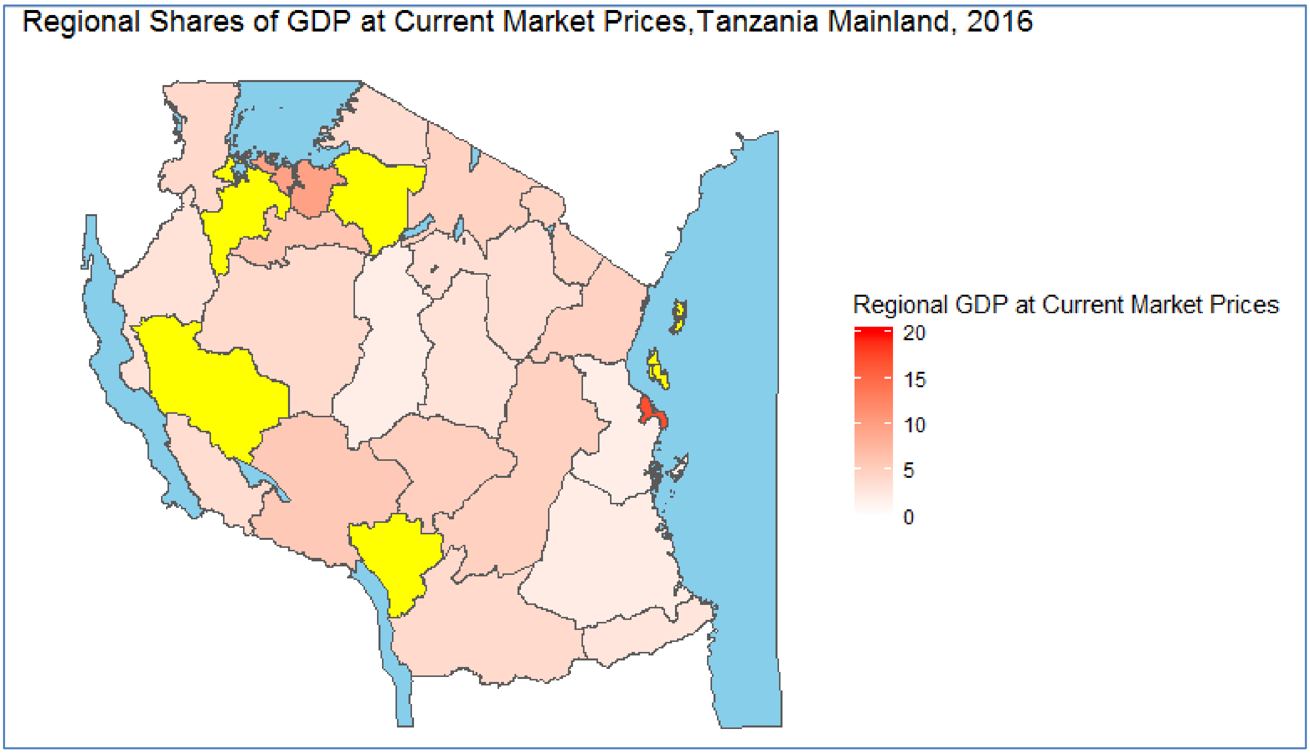

Regional Shares of GDP at Current Market Prices,Tanzania Mainland, 2016
Regional GDP at Current Market Prices
20
15
10
5
0

Figure 7: Regional GDP, 2016 – including water bodies and blank background

10. GDP data

Table 21: Regional Shares of GDP at Current Market Prices, Tanzania Mainland, 2007 – 2016

Region_Nam	2007	2010	2011	2012	2013	2014	2015	2016
Dodoma	3.01	3.17	3.09	3.1	3.03	3.04	2.9	2.92
Arusha	4.52	4.87	4.73	4.77	4.74	4.75	4.7	4.71
Kilimanjaro	4.71	4.45	4.6	4.54	4.54	4.54	4.54	4.45
Tanga	5.58	4.65	4.75	4.69	4.67	4.66	4.66	4.68
Morogoro	5.41	4.87	4.93	4.89	4.84	4.85	4.9	4.81
Pwani	1.89	1.88	1.86	1.85	1.81	1.81	1.81	1.8
Dar es salaam	15.59	16.81	16.69	16.93	17.28	17.2	17.2	17.02
Lindi	1.94	1.93	1.89	1.89	1.85	1.86	1.86	1.96
Mtwara	2.42	2.87	2.58	2.65	2.52	2.53	2.6	2.7
Ruvuma	4.01	3.69	3.87	3.81	3.8	3.81	3.9	3.9
Iringa	5.36	5.27	5.29	5.25	5.18	5.19	5.3	4.92
Mbeya	7.16	7.36	7.3	7.33	7.43	7.44	5.62	5.63
Singida	1.84	2.05	1.89	1.92	1.84	1.85	1.8	1.85
Tabora	4.11	3.85	4	3.95	3.93	3.93	3.8	3.8
Rukwa	3.46	3.38	3.45	3.42	3.4	3.41	3.5	3.6
Kigoma	3.04	2.81	2.93	2.9	2.89	2.88	2.9	2.9
Shinyanga	6	6.07	6.03	6.01	5.92	5.93	5.93	5.95
Kagera	4.03	3.9	3.96	3.94	3.94	3.94	3.94	3.95
Mwanza	8.65	9.16	9.04	9.15	9.38	9.38	9.3	9.7
Mara	3.85	3.63	3.74	3.69	3.68	3.67	3.67	3.67
Manyara	3.43	3.32	3.38	3.34	3.32	3.33	3.33	3.34
Songwe							1.82	1.82
Simiyu								
Geita								
Katavi								
Njombe								
Kusini Pemba								
Kaskazini Unguja								
Kusini Unguja								
Mjini Magharibi								
Kaskazini Pemba								

Note that the set of regions with no data are included to ensure the map is complete.

11. Thematic Maps - tmap

Thematic maps are geographical maps in which spatial data distributions are visualized. This package offers a flexible, layer-based, and easy to use approach to create thematic maps, such as choropleths[1].

With the ***tmap*** package, thematic maps can be generated with great flexibility. The syntax for creating plots is similar to that of *ggplot2,* but tailored to maps. The initial command specifies the shape object and data input (*tm_shape*()) and is followed by the map layer (e.g*., tm_polygons*()). Layers can be stacked similar to ggplot2 using the **+** symbol. In addition, attribute elements can be added to the map and maps can be faceted similar to using facets in ggplot2.

With ***tmap*** there is not a lot of data preparation that needs to happen before mapping. With very little code you can create a simple map.

Furthermore, **tmap** has a unique capability to generate static and interactive maps using the same code via tmap_mode().

11.1 Static Plotting or Interactive Viewing - tmap_mode

The global option tmap.mode determines the whether thematic maps *are plotted in the graphics device*, or shown as an *interactive leaflet map*. The function tmap_mode is a wrapper to set this global option.

There are two options for tmap_mode***, plot*** or ***view,*** both of which will be applied to the *TanzReg* dataframe created in *Section 8*.

"plot"
Thematic maps are shown in the graphics device. This is the default mode, and supports all tmap's features and extensive layout settings tm_layout

"view"
Thematic maps are viewed interactively in the web browser or RStudio's Viewer pane. Maps are fully interactive with tiles from OpenStreetMap or other map providers (see tm_tiles). e. This mode

[1] Choropleth: Areal regions, such as countries or municipalities, are filled with colours that represent a variable which is either a density or a ratio. The usage of class intervals encourages the readability of the data values.

generates a leaflet widget, which can also be directly obtained with tmap_leaflet. With RMarkdown, it is possible to publish it to an HTML page.

11.2 Static Plot

First load the tmap library.

library (tmap)   # for static and interactive maps

Next set the mode to *plot*, this is the default mode.

tmap_mode("plot")

The plot is build up using this function followed by a *"+"* to build up the layers.

The R code below creates a map referenced as "*Tanzania2016*" as follows:

- selects the *TanzReg* data to be mapped – *tm_shape*
- identifies the year 2016 as the fill - *tm_polygons*
- the colours & limits for the scale fill and the name of the scale are automatically selected, and where there is no data, the region is coloured grey, "*na.value*".
- provides a title - *title*
- a main title is provided, including *position and size – tm_layout*
- removes the default axis titles and axis ticks – *theme.*

```
(Tanzania2016 <- tm_shape(TanzReg) +
   tm_polygons(col = "X2016",
            title="Regional GDP at \nCurrent Market Prices", title.size = 1) +
   tm_layout(main.title = "Regional Shares of GDP at Current Market Prices, Tanzania Mainland, 
2016" ,
                              main.title.position = c("top", "center") ,
                              main.title.size = 1.25))
```

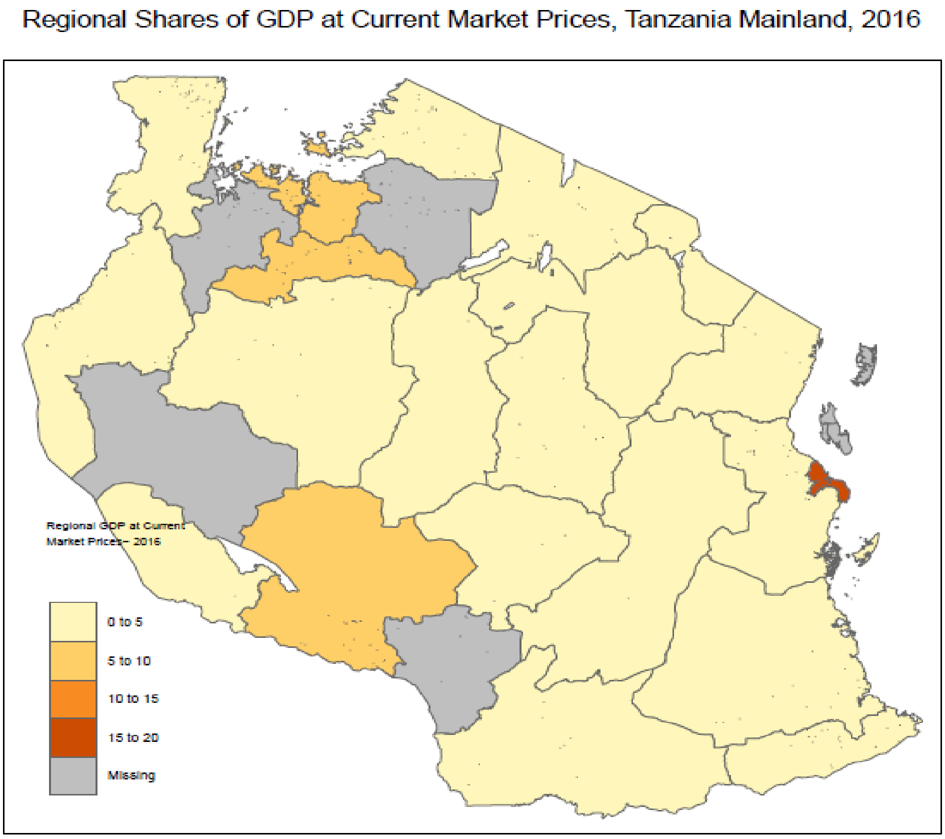

Regional Shares of GDP at Current Market Prices, Tanzania Mainland, 2016
Regional GDP at Current Market Prices– 2016
0 to 5
5 to 10
10 to 15
15 to 20
Missing

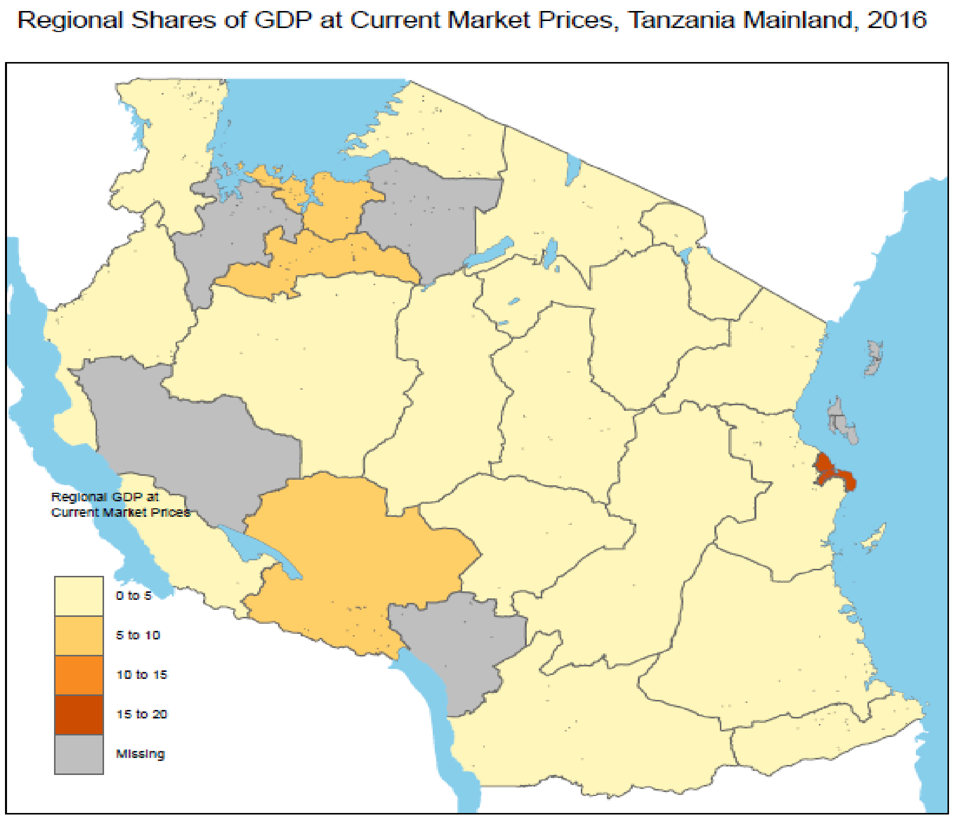
Figure 8: Regional GDP, 2016 – no water bodies

11.2.1 Overlaying the water bodies map

In order to improve the look of the map, we will overlay the water bodies data-frame and add a blank background, using the following code. (See Fogure 9)

```
(TanzaniaB <- tm_shape(TanzmyDFW) + tm_fill(col = "lightblue") )

(TanzaniaA2016 <- Tanzania2016 + TanzaniaB)
```



Regional Shares of GDP at Current Market Prices, Tanzania Mainland, 2016
Regional GDP at
Current Market Prices
0 to 5
5 to 10
10 to 15
15 to 20
Missing

Figure 9: Regional GDP, 2016 – including water bodies

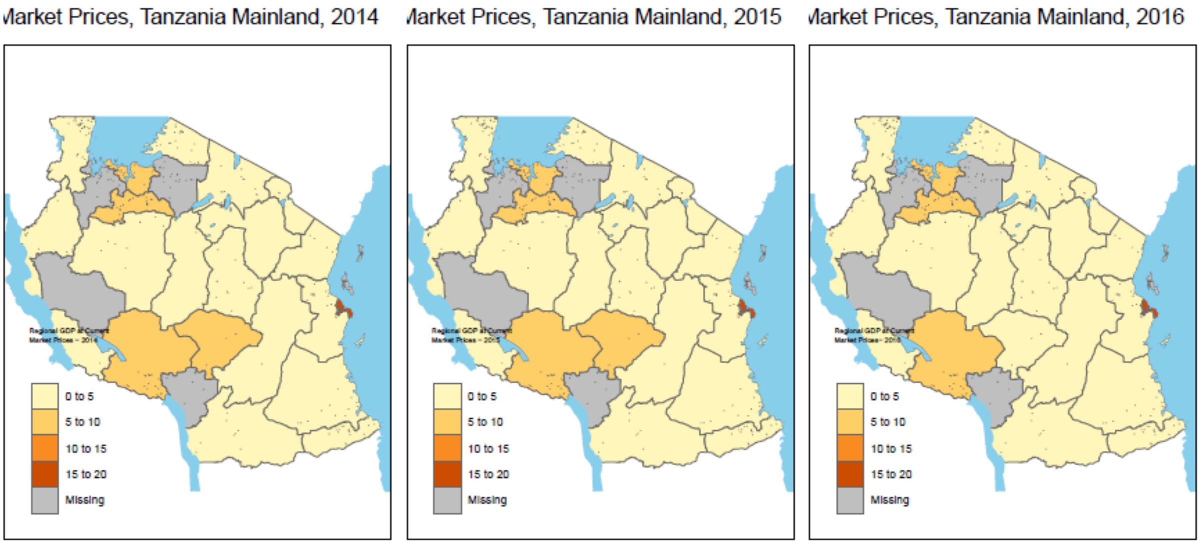
11.3 Plotting multiple maps – tmap_arrange()

Multiple maps can be arranged in a single metaplot with tmap_arrange(), which can be used to visualise a time series.

For example, if the script in *Section 11.2* is run for the years 2014 and 2015, (i.e. *col="X2014"* and *col=" X2015*" in the tm_polygons statement). There are now have three maps of a time series that can be visualised.

The following code arranges the three maps in a single metaplot, where the output for 2014 and 2015 are *TansaniaA2014* and *Tanzania2015A* respectively. (See *Figure 10)*

```
tmap_arrange(TanzaniaA2014, TanzaniaA2015, TanzaniaA2016 )
```



Market Prices, Tanzania Mainland, 2014
Market Prices, Tanzania Mainland, 2015
Market Prices, Tanzania Mainland, 2016
Regional GDP at Current Market Prices - 2014
Regional GDP at Current Market Prices - 2015
Regional GDP at Current Market Prices - 2016
0 to 5
5 to 10
10 to 15
15 to 20
Missing

Figure 10: Regional GDP, 2014-16

11.4 Interactive Maps

While static can enhance geographic datasets, interactive maps can take them to a new level. Interactivity can take many forms, the most common and useful of which is the ability to *pan around* and *zoom* into any part of a geographic dataset overlaid on a ***'web map'*** to show context.

It should be noted that the map is always projected according to the Web Mercator projection. Although this projection is the de facto standard for interactive web-based mapping, it lacks the equal-area property, which is important for many thematic maps, especially choropleths.

A unique feature of ***tmap,*** previously mentioned, is its ability to create static and interactive maps using the same code. Maps can be viewed interactively at any point by switching to view mode, using the command tmap_mode("view"). This is demonstrated in the code below, which creates an interactive map of Tanzania on the tmap object Tanzania2014, created in Section 11.3 and illustrated in Figure 11.

```
tmap_mode("view")

TanzaniaA2014 +
tm_legend(outside = TRUE) +
tm_view(view.legend.position = c("left", "bottom", legend.title.size = .5,
legend.text.size = .5))
```

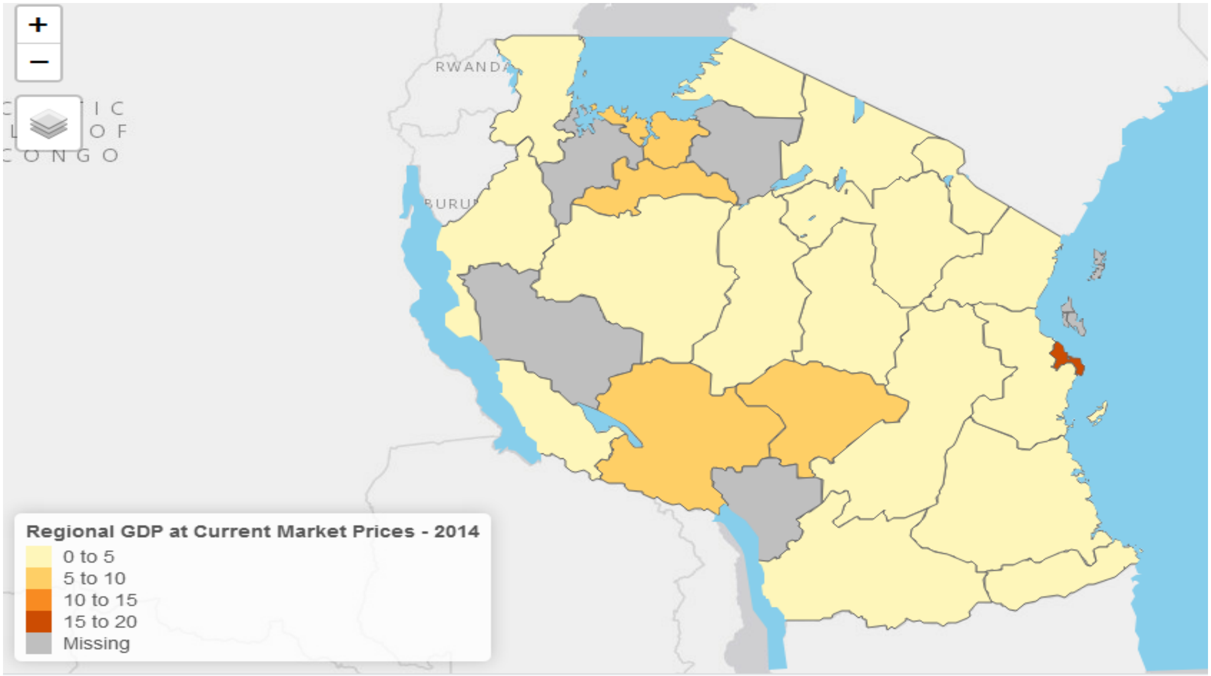


Figure 11: Regional GDP, 2016 – interactive map overlaid on Esri.WorldGreyCanvas, web map.

11.4.1 Zoom on an Interactive Map

One can zoom in and out of an interactive map using the “+” and “-“ buttons on the top left hand corner of the map. (See *Figure 12*).

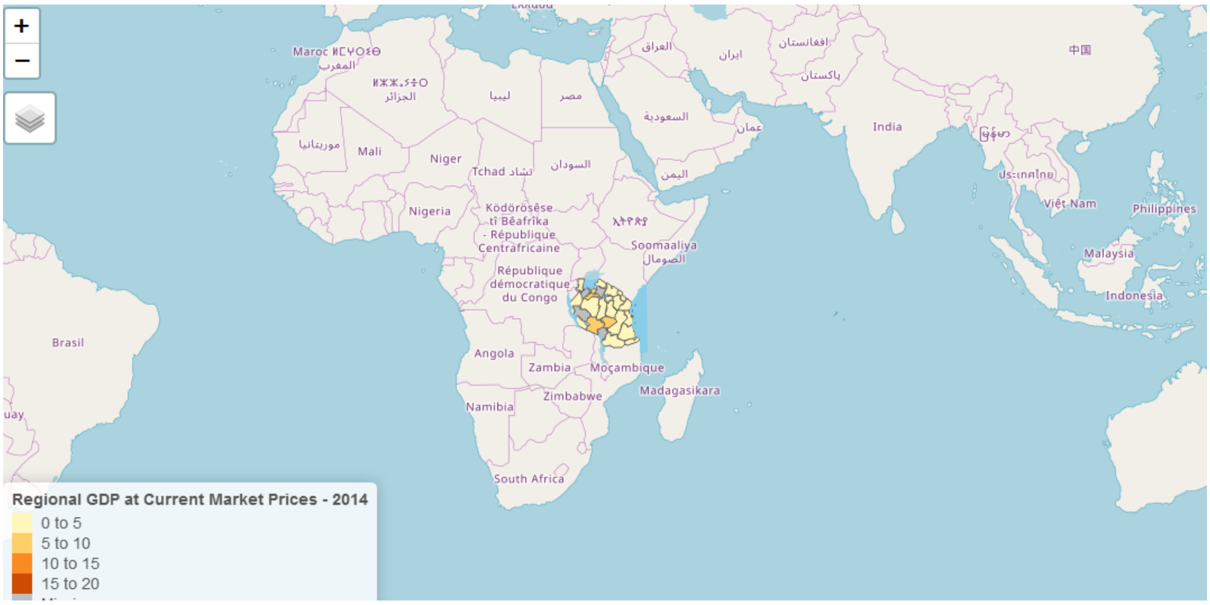


Figure 12: Regional GDP, 2016 – interactive map, Figure 11, zoomed out .

11.4.2 Changing the Web Map

The web map upon which the interactive map is layer can be changed using the layer button on the top left hand corner of the map. There are two other options, Open Street Map, *(Figure 13*), and Esri.WorldTopoMap, (*Figure 14*).

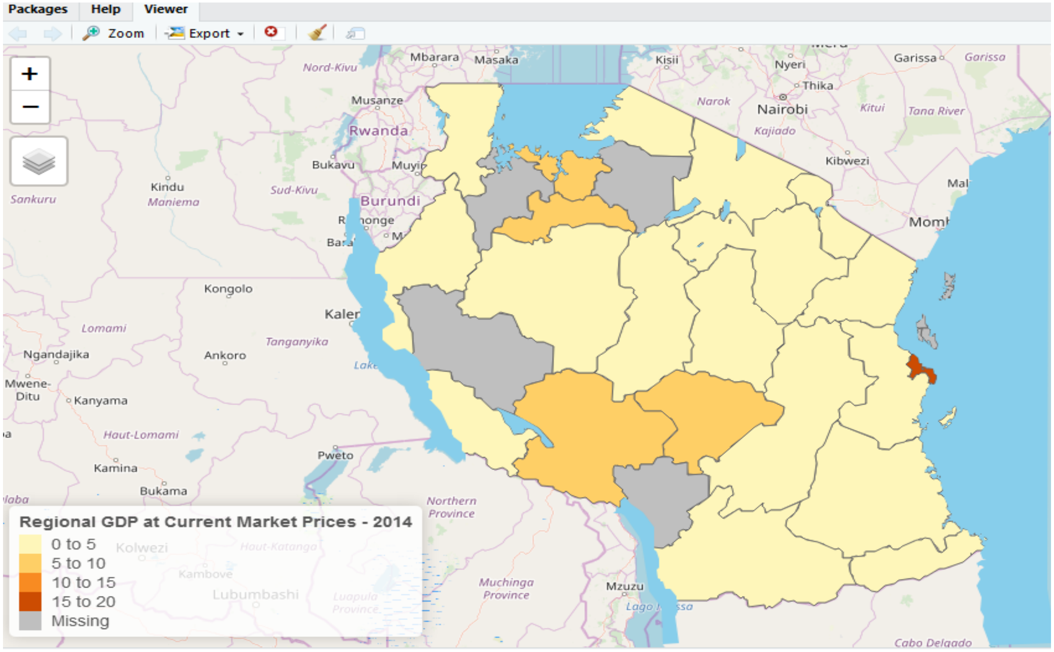


Figure 13: Regional GDP, 2016 – interactive map overlaid on Open Street Map web map.

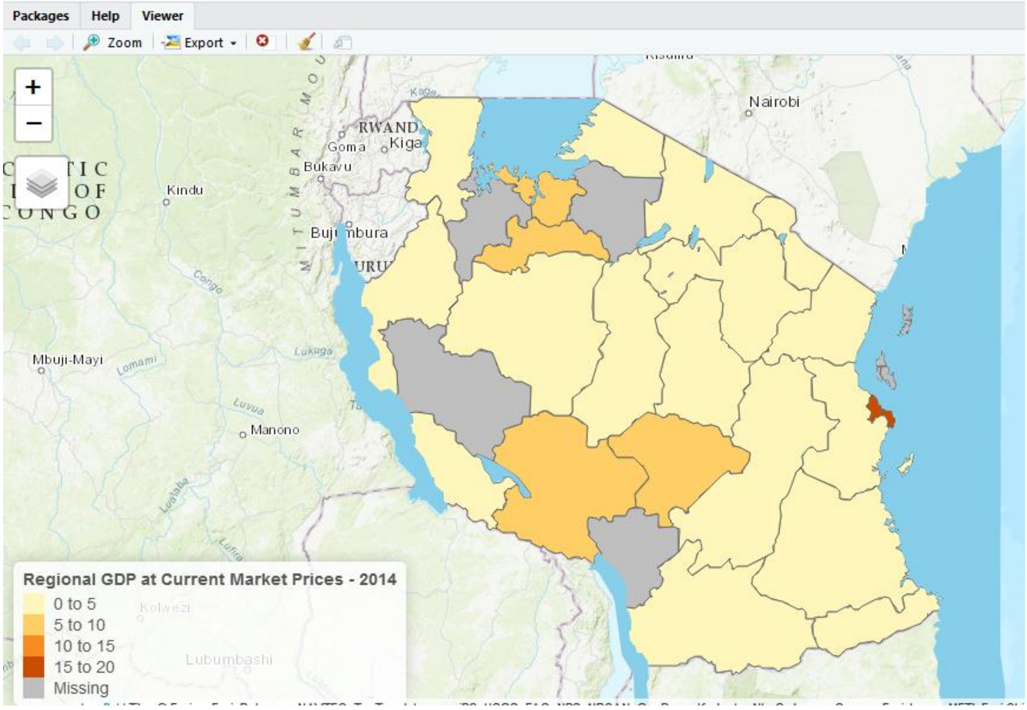


Figure 14: Regional GDP, 2016 – interactive map overlaid on Esri.WorldTopoMap web map.