Evolution of countries after the colonization process (Colonial Heritage)

Carlos Branco
71003 - G47
Lisboa, Portugal
carlos.m.branco@tecnico.ulisboa.pt

1. INTRODUCTION

With the arrival of the climate change movement other underlying problems have recently reemerged, such as the latent economic growth of third world countries. But how easily obtainable and understandable is this information, if accurate at all? Prior to this, is also another period of mankind were heavy exploitation of resources, mainly from the African and South American continents following the fallout of the pink map division occurred. How far behind are these countries? How have they evolved since their colonizers left and are they improving significantly over the last forty years?

With this work we set to develop a visual tool in order to allow the user to better grasp how these countries have evolved via direct manipulation of the data in an intuitive way. Furthermore, we will also be looking at how this tool can be used to answer not only the previously mentioned questions but others, regarding the analysis of key performance indicators associated with each country. No similar solution or comparison tool was found, thus the need for creation, not only to help lay people interpreting and evaluating their hypothesis but also ease the work for anthropologists in their research.

In addition, we will be looking at how these mechanisms were implemented and thought out, how problems surrounding the data and the design were overcome and the shortcomings of some solutions.

With the most recent data from the United Nations Food and Agricultural Department, and key indicators from the World Bank Group we set to to answer the following questions:

- Does Angola compare favorably with Niger in common indexes such as gross domestic product, life expectancy at birth and human development index?
- How did Syria's economic indicators evolve after the break out of the war in 2011?
- Do Argentina's recessions coincide with Zambia's from 1993 on?
- Is Angola further from Namibia than Libya?
- What can be said about the average gross domestic product per capita between 2000 and 2015 for these countries? How does gdp growth of these countries compare in the same time span? Can any global event be identified?

Given the aforementioned, we expected to gain further knowledge on world economics and country relations with this project by visually analysing patterns underlying the data.

2. RELATED WORK

To the present date no similar work was found. Despite a lot of sources offering engines to view raw data only very little provided simplified and easily understandable visualizations of said data and when such happened a very coarse and predefined visualization took place, thus our need to group multiple of these idioms in a single interface and deliver a more flexible and intuitive tool to the user.

One example of such work used as a "model" can be found at:

https://globaledge.msu.edu/comparator/home/results?year=&field=BN-CAB-XOKA-CD&field=FP-CPI-TOTL-ZG&country=79&country=205

Nonetheless, we opted to design a bottom up custom solution for our dashboard using design heuristics previously attained in other courses not relying on any implemented similar solution.

3. DATA

The data used as foundation for this work was gathered from the following websites in an attempt to bring together a high quality, high fidelity dataset reflecting the current economic situation around the world:

- Colony ownership from: https://dataverse.harvard.edu/file.xhtml?persistentId=doi:10.790/DVN/T9SDEW/FYTVWA&version=2.0
- Detailed matrix for exchanges between countries: http://www.fao.org/faostat/en/#data/TM
- Life expectancy data from: https://data.worldbank.org/indicator/SP.DYN.LE00.FE.IN?view=chart
- Human Development indexes http://hdr.undp.org/en/content/human-development-index-hdi
- Agricultural production indexes: http://www.fao.org/faostat/en/#data/QV
- Livestock Production: http://www.fao.org/faostat/en/#data/QL
- Country metadata: https://meta.wikimedia.org/wiki/List_of_countries_by_regional_classification
- Gross domestic product and gdp growth: https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG?end=2018&start=1960&view=chart

Several problems had to be handled before the data could be used, hence a cleaning step prior to aggregation took place. First, the longest period of time of the shortests was chosen as the time window under inspection, in order to get a consistent time window for viewing. Despite lowering the completeness of the data, a compromise had to be made so as to guarantee that most of the countries are consistently comparable for a given time span.

Then, names had to conciliated since countries changed their designations and boundaries through time, as well as conciliating different nomenclatures from the sources (remove/merge instances). In addition, the data was aggregated into two sources, namely 4_threenum.csv and q2.csv, representing the exchange of goods between countries and their international indicator values.

This process required not only pivoting data matrices but also merging similar instances and filtering out unwanted/outdated data in order to get one row per year per country with all the attributes for said year. This process was also done for the migration data, which later was not considered while developing the final prototype and solution.

Another problem found was the volume of said data to filter, which totaled over 4GB at the beginning (including the migration data for countries from 1990 onwards which is not mentioned in this report as it was also later discarded). To do this a set of python scripts and Pentaho transformations were created to unify files, filter similar records and remove aggregated data (files also included regional and continental data). At the end of said transformation pipeline, the two files previously mentioned were created, occupying only 470Kb in total. Furthermore, attributes had to be filtered from some of these files as there were over 500 attributes in total (both numerical and categorical) that were later reduced to less than 15 for usability issues.

After selection and creation of the intermediate representation format, by crossing information contained in the indicators new aggregation measures were created representing the average of each country's colonizerers for each indicator, culminating in the final dataset. In other words, after reducing one of the datasets to aprox 10 indicators, these were then expanded to 20 (each of which with the colonizer average, created for a more general overview of said colonizer when comparing, if required).

These transformations had to be revisited later after due to technical details of drawing the areas and maps using d3js (given a topojson file some of the country names mismatched, to the current date not all countries are correctly encoded such as congo having one representation in the dataset while there are actually two countries named congo, an error made during the transformation of data that was only later discovered). This was done to better fit the data for handling, loading and manipulating with the library used. Some attributes were dropped, most were renamed for better used experience and some names had to be further looked into to unify the name (St and Saint and Snt. , Cape and Cabo, and others).

Missing values in the dataset were also imputed with the value -0 to allow calculating averages and plotting while being clearly distinguished from normal instances when inspected.

4. VISUALIZATION

To solve this problem a dashboard was created from which the user is able to select not only countries but also indicators to compare.

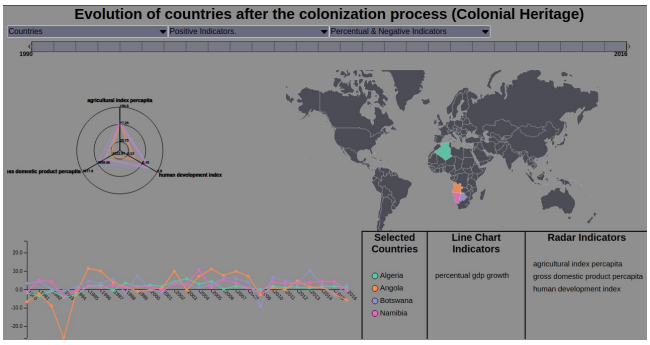


Figure 1. Dashboard overview encoding the idioms and selection tools. At the top of the screen three dropdown boxes for filtering and selecting attributes and countries can be found. In the visualization area an interactive world map, radar chart and connected scatter plot can be found, upon which information will be represented respectively. Finally, on the bottom right, an indicator section be found to quickly inform the user of the attributes under inspection for each idiom.

Selection combo boxes and time selection were placed at the top of the screen in order to maximize the useful area for the visualization idioms while allowing the user to select countries and time periods to further explore. Despite the solutions found looking quite simple they achieve their purpose without bringing further complexity and clutter while being intuitive. In these areas we allow the user to choose from multiple pre-selected indicators, as well as computed aggregations. Selecting these attributes via their checkboxes will cause the data to be added to the idioms, and posteriorly plotted for visualizing. Furthermore, the time span can also be adjusted by grabbing any of the handles in the timeline and shortening their inspection period. This bar, or rather time interval, can also be slid to the side in order to keep the time span but changing the years, or individual cells can be clicked to select a single year. All these steps were thought out in order to provide the user with the most ways in which they can explore data at different granularities, in this case the time dimension and combinations of indicators.

Figure 2. Detail of an open drop down menu for selection. Furthermore, the timeline can also be seen with a selected time interval between 2000 and 2016.

Following the selection menu, four distinct areas were created. Given the inherent rectangular shape of the mercator projection, the world map was placed in the top right corner. Not only this map is meant to represent what countries are selected but also to allow the user to infer their distance and quickly identify their relative position in the world and to each other since the mercator projection does not significantly distort vertical distances. Furthermore this map allows for direct selection of countries. The channels used for highlighting country selection were always the same and provided by d3 and the marks are the approximate shapes of the countries and seen in most mercator projections.

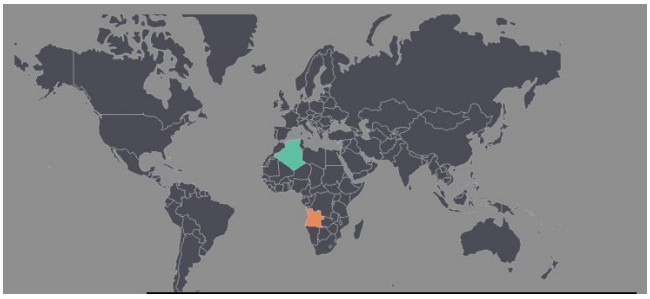


Figure 3. Detail of the world map with selected countries highlighted with their respective categorical color.

In addition, and since different types of numerical attributes are being used (numeric diverging vs positive) two further idioms were created.

First, a radar chart (Fig. 3) for viewing non negative non percentual indicators was placed on the top left corner (in order to maintain coherence and context for the magnitude of values displayed while maximizing the available space as circles comprising the radar chart will inherently force a square shape for the bound data) followed by a connected scatterplot (Fig. 4) which will be holding negative and percentual indicators together. Both of which were encoded using the same channels but different marks, with lines abstracting the trend between two data points, circles for their exact value at the year of measurement and a unique color for each country.

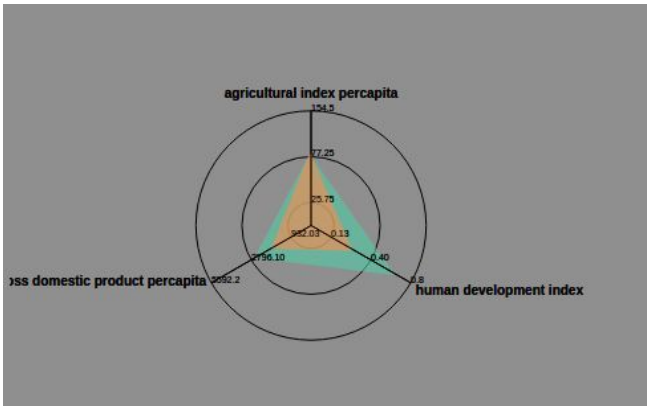


Figure 4. Radar Chart overview upon which positive gross indicators will be plotted. The colors represent the average of each indicator for a given time window. Mouse hovering

causes the elements to pop a tip upon which the country same under selection is displayed. (Not fully implemented)

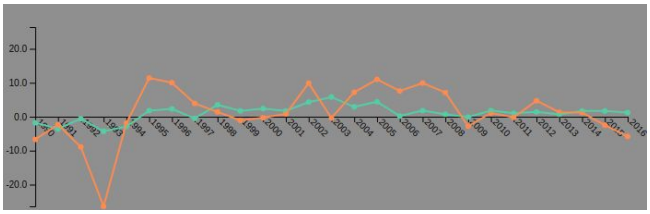


Figure 5. Connected scatterplot for negative and percentual indicators. These are grouped under the same visualization in order to keep comparison accurate (context) by not allowing to compare percentual/possibly negative with others (positive and concrete valued indicators). Similarly to the radar chart hovering causes a tooltip to pop with more specific information. (Fully implemented)

Finally, on the bottom right corner a legend was created to allow the user to quickly identify what are the selected countries, indicators under observation for each visualization and to keep him informed of the application state at any time, being dynamically managed by d3.

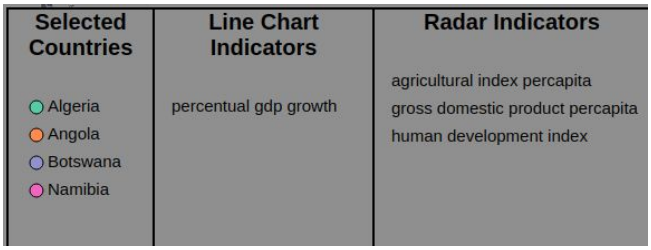


Figure 6. Legend: transitioning, data updating and highlighting are fully implemented for the country selection. Similar work should be done for the indicators

The categorical coloring for the countries was done via the d3js categorical color scale which creates a sequence of hues that optimize the task of quickly distinguishing countries by color, an identified problem in infovis, leaving the choice of marks for the group to decide (implemented as previously explained in this chapter).

Alternatively a small multiples bubble chart was also considered, however given the lack of perception by humans when comparing areas (encoding attributes), this solution was left out.

4.2 Rationale

First, we started by designing a first sketch of the prototype (Fig. 2) in which the screen area was split into five distinct areas, one for selection and three for data representation. With the selection on the left, a world map/network graph of selected countries on the top right. At the left of it a selection area with menus would withhold the indicators and allow to control de visualizations. On the bottom a line chart for percentual/negative indicators, followed by a radar chart for non negative/percentual country indicators and finally a small multiples area upon which bubble charts and scatter plots with migration/trade data would be plotted.

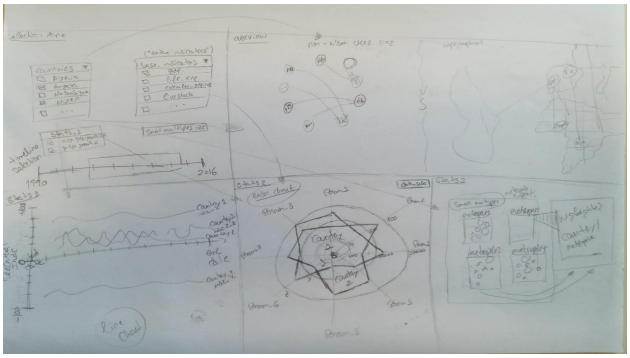


Figure 7. Initial sketches comprising selection and visualization areas. Multiple cases can be seen for some idioms as they were under discussion of which would better suit the overall solution.

In the end, the mercator map was chosen due to its natural resemblance with the countries thus allowing the user to easily recognize the domain being represented with no added effort rather than interpreting.

Then, the line chart was chosen to hold the datapoints in the dataset that were either percentual or negative. By having the 0 in the middle of the magnitude of they Y scale, expressing the range for the attributes. With this in mind, we can not only analyse the general trend but also quickly distinguish positive from negative indicators. Time was attributed to the X axis also due to its inherent representation of a line. Diverging attributes will be represented in this graph as they would lose meaning if compared with others of different magnitudes.

The radar chart was thought to allow comparison of all positive indicators with the 0 in the center with attributes being radially distributed. Multiple circles encoding different levels of magnitude were drawn, upon which is added the value at that distance. The scale is drawn to contain the maximum of each indicator for the data under inspection. The resulting areas represent the average for the time period selected, thus this allows us to see which country is higher or lower placed (relative comparison) based on each ones average with the tips indicating the average of the attributed in that axis.

Channel wise, colors came from the D3JS library categorical color scale function designed for such functions.

The attribute chosen to be passed to this function, thus being encoded by the color channel was the country name, with lines (chosen marks) representing the trend between time points in the case of the connected scatterplot.

Given that at any moment the user could select any number of countries and indicators some restrictions had to be imposed on the flexibility of tool more precisely limiting the number of selected countries. To avoid cluttering of the visualizations, a hard limit of 6 countries was imposed, number after which no new country will be inserted unless another is removed, while notifying the user of said restriction. Indicators on the other hand are unbounded in selection quantity for both radar chart and connected scatterplot.

After some feedback and planning the layout was changed in order to make better use of the space available as follows:

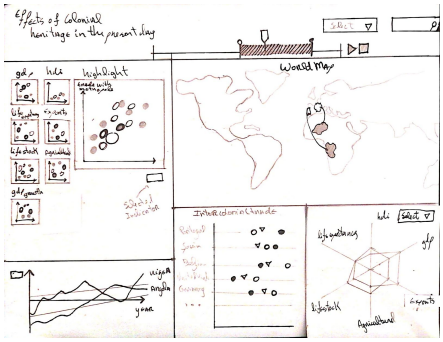


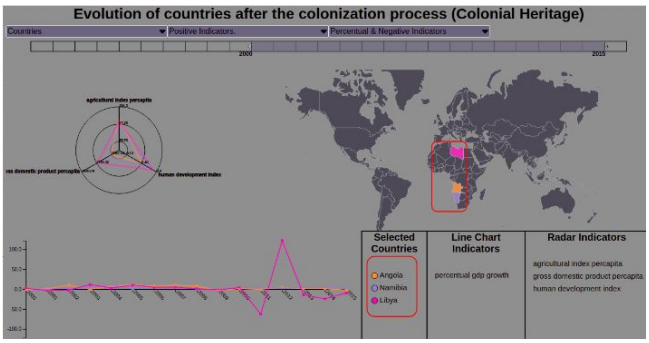
Figure 8.Reviewed sketch of the interface.

This version was later revised to what is the final delivery, with added behaviours not represented such as tooltip interactivity. Highlighting however was already planned at this early stage and is fully implemented. Additionally a legend area was placed on the bottom right as the interface might have lacked some visual explanation of what is being represented.

4.3 Demonstrate the Potential

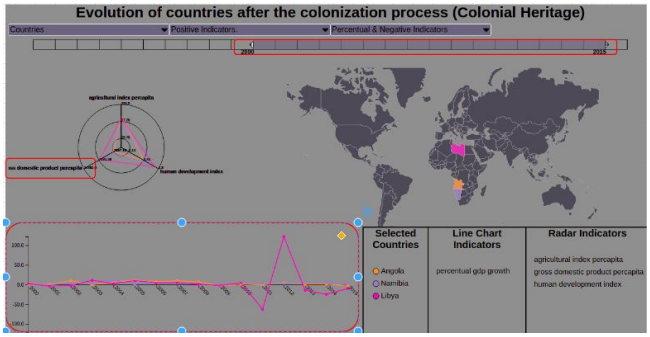
In this chapter we will be demonstrating the potential of our solution by answering some of the previous set questions.

Question 1: Is Angola further from Namibia than Libya?



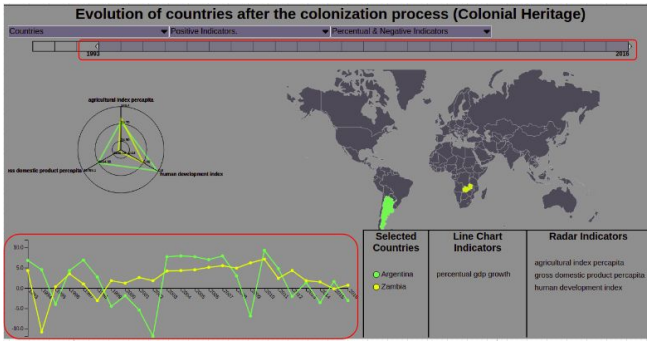
Answer: No.

Question 2: What can be said about the average gross domestic product per capita between 2000 and 2015 for these countries? How does gdp growth of these countries compare in the same time span? Can any global event be identified?



Answer: Average gross domestic product per capita is higher for Lybia than Angola and Namibia. With very similar growths (represented in the linear) and abrupt change in gdp growth can be found for Lybia with no impact on the remaining countries. This change corresponds to the 2011 Lybian civil war.

Question 3: Do Argentina's recessions coincide with Zambia's from 1993 on?



Answer: Apart from the plunge in both countries domestic product growth in 1994 and 1995, Argentina has suffered

much more recessions, with higher (and harsher) shrinkage than Zambia which has only suffered milder recessions.

IMPLEMENTATION DETAILS

Implementation of the visualization was done using plain Javascript, CSS and HTML along with the D3JS library which contains multiple functions that allow the user to draw freely on an svg element. Given this, one of the main problems encountered while implementing the solution was getting a data structure which encodes the positions of countries and their shapes in order to plot them on the map chart. Furthermore, and given that d3js can be seen as a set of function calls, throughout development it became clearer that the lack of prior knowledge on how this library and svg elements work together and their properties would not be an easy task to overcome, associated to the lack of documentation (multiple versions, not cross supported and poorly documented) and examples to practice for the syntax and semantics of the latest version.

Nesting of data also had to be done in order to properly plot the connected scatterplot, for which very little information was available (how to associate, update and transition nested data and their parent DOMs).

Cross visualization interactivity was achieved by associating function calls to DOM elements that when triggered (either by hovering or clicking) will cause all elements relevant elements to the action to be notified change appropriately.

Whilst it proved relatively easy to find similar visualization idioms to the ones implemented it was much harder to adapt some of them to our data (namely due to insertion of nesting). As a consequence, the enter update pattern provided by d3js to update data was significantly harder than expected as one idiom not only uses circles (acting like a scatter plot since data is non continuous) but also lines. These had to be conciliated not only with the data structure but also with one another in order to achieve proper functionality.

Designing the tooltips themselves also proved to be quite hard due to their positioning and styling obfuscating some of the content displayed and sometimes exceeding the margins of the website (something that was duely fixed).

Styling also proved to be quite hard as a lot of information is available but no generalizable heuristics are agreed among the scientific community. Given this, the palletone website color schemes were used (one preset style was chosen), as advised, since it's premade color schemes try to accompany the state of the art investigation on the subject.

Upon change of attributes, elements are notified, updated, transitioned and finally removed accordingly. This is done via the update pattern in the d3js library which also takes care of changing the associated DOMs and tailored functions to locate and filter them.

Finally, the timeline selects time windows by using brushing, either via handle dragging or single cell clicking, in order to notify other elements about the time window of events under inspection, and transition them accordingly. Highlighting is placed between the handles and the limit dates are written beneath each handle so as to reflect the year at each "slider tick", thus informing the users of the intervals being selected and which ones data will not be displayed for.

CONCLUSION & FUTURE WORK

In summary we can state that most of the objectives were achieved (despite being one idiom shorter than what was requested and having one incomplete idiom, namely the radar chart for which data transitioning was not implemented due to the lack of time). If it were the case for the project to start over, more care would have been taken in gathering the data (more completeness, less problems during implementation, knowledge of the names of the countries that could be represented in the map prior to name unification).

Furthermore the interface could be revamped using bootstrap for a more aesthetic and pleasing look, the menus could be changed in order to allow not only multi combo box selecting but also allow the user to search by typing.

Zooming and panning could also be added to the map in order to allow better contextualization as well as indicating the computed distance between pairs of countries.

In addition. the radar chart would be greatly improved by implementing not only axis transitioning but also by doing small adjustments in some of the labels (as some overlap with others, thus not making them legible). These changes to the radar chart would greatly help the user managing context changes (similarly to the connected scatterplot).

Tooltips with a question mark on the corner of each element could also be added in order to provide help and explanation of each idiom in case any doubt of its functionality were to rise in the user.

Given enough time more idioms and aggregation metrics could be included in order to improve the overall quality of the answer provided. An example of said inclusion would be adding information on the population such as average school level attained per country which could be further aggregated into regions that could be interactively explored in a bar chart that would support semantic zoom.