JBI100 Final Report: The PACT: Broaden Your Horizons

Catherine Kmiec Polina Pelster Hanna Shylova Thorvald Rovers

ABSTRACT

This paper presents the visualization tool PACT, designed to enhance any highly skilled migrant's ability to choose countries to immigrate to. The visualization is our main contribution, a visualization application that uses the 2024-2025 CIA Global Database and other sources to enhance the data. There is a multi-component design featuring an interactive world explorer, comparison views, and a global patterns analyzer. The tool facilitates an in-depth exploration into countries and serves as a decision-support system to help users select a country they might be interested in.

Key words: Visualization, Interactive Tools, International Migration, Highly Skilled Migrants, Performance Analysis, Comparative Analysis, Parallel Coordinate Plot, Multivariate Data.

1 INTRODUCTION

Choosing a country to move to is a difficult task for many since relevant information is scattered, inconsistent, and hard to compare. To make this important decision, individuals have to go on a scavenger hunt across the internet to collect tidbits of information and compare conditions, demographics, safety, living standards, and visa information. Comparing countries manually like this is slow and often unreliable.

For this project, the main goal of the visualization tool we have created is to compile the available data for each country and provide an interactive visualization tool that allows for user exploration and analysis. The visualization is built for users who identify as highly skilled migrants. The domain problem addressed within the visualization is the difficulty of finding reliable and complete information in one source regarding multiple countries, being able to compare multiple attributes between countries, and comparing across multiple highly-skilled migrant visas. This domain problem is relevant in the current world context where international mobility and immigration decisions are common, and where users need a clear overview to make informed decisions about future immigration. Visualization is the right tool in this case as comparing dozens of numerical values across multiple countries is cognitively demanding ([5, p. 6]). Along with the cognitive load associated with numerical figures, simple summary statistics for each country would not be sufficient since there is no one "perfect" country for every human. Visualization therefore fills in the gap that summary statistics and other methods leave behind, and allows for user interaction that is crucial to satisfy the need of each user. Visualization empowers users to further exploration outside of the typical scope they might have originally had, identify patterns by regions, and analyze previous beliefs they might have had about different countries, which would otherwise be difficult to detect by just numerical values.

To create this visualization, we have decided to use the 2024-2025 CIA country Kaggle dataset. This dataset provides a comprehensive collection of country-level attributes regarding demographics, economics, energy, geography, governmental, transportation, and communications data. To further enhance this dataset, we chose to compile manually data regarding highly-skilled visas in the most prevalent countries. Along with this manually compiled data, we created derived parameters as discussed in the following section. The data we have chosen is relevant for our target users as it includes a wide range of attributes for all countries allowing for proper exploration, analysis, and includes relevant targeted information.

2 DATA ANALYSIS

2.1 Domain Data Specification

As mentioned previously, the visualization tool uses the 2024-2025 CIA Global Statistical Dataset as the primary source of information. The dataset is composed of different similar datasets for all domains that have been mentioned previously.

Missing values throughout this dataset were treated mostly as unknowns rather than substituted with artificial values. We treated them in most cases as such since creating artificial values for attributes is near impossible, as each country in each region has the potential to be vastly different. However, if an attribute was deemed as essential by our team, the values were filled in manually from authoritative public sources to preserve the accuracy of the data. Due to the lengthy nature of the list of sources, the sources are attached in a separate column of each dataset where values were filled in. To determine whether or not an attribute was essential to our end goal and, therefore, needed to be filled in, a checklist was devised. Firstly, if an attribute was overtly important to explain the overall "health" of a country, like GDP per capita, it was filled in. Secondly, if the attribute value was shown anywhere on the visualization, it was filled in. Some attributes that were irrelevant to both the visualization and the derived parameters, such as internet country code were deleted. A similar process to the filling in of data was followed during the attribute deletion. If an attribute was not used in either the derived parameters nor the final visualization as a "raw" attribute, it was deleted. A full list of the kept attributes is available within the data abstraction section that follows.

To further user understanding of a variety of attributes and consolidate important metrics, derived parameters were created, referred to from this point onward as "indices". We first imported the Global Terrorism Index [3] and then the Global Peace Index [2] from the Institute for Economics and Peace. We then imported the Crime Index [4] from the organization of Global Organized Crime. Then, the Safety Index was created through:

$$\text{SI}_{risk} = 100 \cdot \Big(0.40 \cdot \text{crime}_{safe} + 0.40 \cdot \text{terror}_{safe} + 0.20 \cdot \text{peace}_{safe}\Big)$$

For the Health Index, we imported the Life Expectancy Index from the World Population [8], then imported the Healthcare Index from Numbeo [6]. We then combined it with the Infant Mortality Rate from the CIA database and the Health Index equation is:

$$\text{HI} = 100 \cdot \big(0.30 \cdot \text{Healthcare}_{good} + 0.40 \cdot \text{LifeExp}_{good} + 0.30 \cdot \text{InfMort}_{good}\big)$$

Next, the Economic Stability Index was calculated through first importing inflation rate from the World Bank Group [7], and then combined with the GDP per capita, unemployment rate, and GDP growth rate from the CIA dataset. The equation for the Economic Stability Index ended up being:

$$\text{StabilityBase} = \frac{\text{GDPpc}_{norm} + (1 - \text{Unemp}_{norm}) + (1 - \text{Infla}_{norm})}{3}$$

$$\text{EconomicStability} = \text{StabilityBase} \cdot (0.5 + 0.5 \cdot \text{GDPgrowth}_{norm})$$

Up next, the Demographic Pressure Index was created using population growth rate and infant mortality rate from the CIA Dataset. The equation is as follows:

$$DPI = 100 \cdot (0.45 \cdot PopGrowth_{norm} + 0.45 \cdot InfMortality_{norm}$$
$$+0.10 \cdot (PopGrowth_{norm} \cdot InfMortality_{norm}))$$

The Future Earning Potential Index was calculated through GDP per capita, GDP growth rate, and the normalized unemployment rate from the CIA dataset. Then, we combined the Economic Stability Index that was created previously. The equation is as follows:

$$\text{EarningPotential}_{future} = 100 \cdot \sqrt{\text{GDPpc}_{norm} \cdot (1 - \text{Unemp}_{norm})}$$
$$\cdot \sqrt{(\text{EconomicStability} \cdot (0.6 + 0.4 \cdot \text{GDPgrowthrate}_{norm}))}$$

The last derived parameter is the Good Country Index. The Good Country Index uses the Health Index, the Safety Index, the Economic Stability Index, and the Earning Potential Index. The equation is as follows:

$$QualityofLife = 0.55 \cdot Safety + 0.45 \cdot Health$$
$$Opportunity = 0.55 \cdot EconomicStability + 0.45 \cdot EPI_{future}$$
$$GoodCountryIndex = 0.60 \cdot QoL + 0.40 \cdot Opportunity$$

As discussed briefly in the previous section, we manually compiled a separate dataset filled with information regarding highly skilled migrant visas. We discovered during compilation that many countries have a variety of approaches to highly-skilled migrants, whether that be through specific visas or pathways, all were recorded. During the creation of this dataset, we chose to only find visas for the selected areas. To find the names and details of each visa from each country, only government websites were used. After finding each visa detail, we set out to find data regarding specific acceptance counts for each visa. However, since most of the visas available outside of the USA are merit-based and are given case-by-case, large-scale fresh data is not available, therefore we chose to only focus on visa names and pathways, along with some basic information regarding requirements. Since this was a manually-created dataset, no preprocessing or missing value handling occurred.

2.2 Data Abstraction

The key attribute within our dataset is country name, as this is the only defining and distinct attribute, and it acts as the linking element across all the datasets. The types of attributes present within the dataset are both categorical and ordered. Attributes such as country name or type of government in a country are categorical, as they have no intrinsic meaning in the order they are listed. In contrast, attributes such as total population or suffrage age are quantitative sequential ordered attributes. The following is a list of all attributes the final visualization uses or incorporates within the CIA dataset:

- **Categorical:** Country and Capital
- **Ordered quantitative sequential:** Telephone fixed subscriptions total, mobile cellular subscriptions total, internet users total, total population, population growth rate, birth rate, death rate, net migration rate, median age, infant mortality rate, total fertility rate, total literacy rate, male literacy rate, female literacy rate, youth unemployment rate, GDP per capita, unemployment rate, population below poverty line percent, electricity access percent, and suffrage age.

Along with the CIA dataset, as mentioned previously, indices were both imported along with some metrics and then further used to create new indices, along with some attributes from the CIA dataset. All indices and metrics imported or derived became new attributes as part of the dataset. The key attribute remains the country name. For the sake of clarity within the report, they have been separated to be sorted as follows:

- **Ordered quantitative sequential:** Global Terrorism Index, Global Peace Index, Crime Index, Safety Index, Life Expectancy Index, Healthcare Index, Health Index, Inflation rate, Economic Stability Index, Demographic Pressure Index, Future Earning Potential Index, and Good Country Index

The last addition to the dataset was the manually created visa and pathways dataset, with the key attribute remaining as country name. The attributes have been separated for in the interest of clarity are as follows:

- **Categorical:** Visa name, temporary/permanent, minimum education level, sponsor necessary
- **Ordered quantitative sequential:** Cost

The database contains no inherent network nor hierarchical structure. Therefore, according to Munzner's taxonomy [5], each dataset is represented as a multivariate table. Each row of the multivariate table corresponds to a country name and each column corresponds to an attribute describing that country. Furthermore, the data abstraction is a single item type described by many attributes, both in categorical and ordinal nature, forming a multivariate tabular dataset intended for comparison and exploratory analysis.

3 Task Analysis

3.1 Domain Specific Tasks

While approaching the visualization, a user could potentially have a multitude of sets of tasks and questions they would like to answer. Due to the ability of the visualization to support exploratory analysis, it delivers the perfect answers to the questions.

- **T1: Identify countries meeting criteria**
  - **T1.1 Identify countries that meet multiple criteria simultaneously** A user would like to find a country that satisfies multiple criteria they believe to be the most important. Users may ask a question such as "what countries have low unemployment, a high good country index, and low population density?"
  - **T1.2 Identify countries that meet criteria within constrained subsets** A user would like to find a country that satisfies multiple criteria within a particular region. Users could ask questions such as "which countries within the EU have low unemployment and high earning potential?"
- **T2: Compare countries**

- **T2.1: Compare two countries on a single criteria** A user has two countries they would like to compare based on a single criteria. They may have a question such as ”does Sweden or Norway have higher GDP?”
- **T2.2: Compare two countries on multiple criteria** A user may have two countries they would like to compare based on multiple criteria. They may have a question such as ”What are the differences between Poland and the USA across the good country index, the GDP, and the birth rate?”
- **T2.3: Compare multiple countries on multiple criteria** A user may have multiple countries they have picked and they would like to compare them based on multiple criteria. User could ask a question such as ”What are the differences between Andorra, Spain, and France across the healthcare quality index, safety index, and the good country index?

- **T3: Find a specific value for a country** A user could have an interest in finding a specific value of a country. For example, a user could ask the question ”what is Japan’s literacy rate?”
- **T4: Understand regional differences and patterns**
  - **T4.1: Identify which region a country belongs to** A user might have a fundamental misunderstanding of geography. A user would like, therefore, to find which region a country belongs to. Therefore, this user could have a question like ”is Russia in Eastern Europe or North America?”
  - **T4.2: Compare regions across multiple attributes** A user would like to be able to compare different regions on multiple attributes. This user might have a question like ”How does Eastern Europe compare with North America on GDP and cost of living?”
  - **T4.3: Identify patterns or commonalities within a specific region** A user might want to find patterns within a specific region. This user would then ask the question ”What are the common patterns or commonalities within Eastern Europe across different attributes?”

3.2 Task Abstraction

Following Tamara Munzner’s abstraction framework [5], the domain tasks can be generalized into domain-independent visualization tasks.

- **T1** Browsing the dataset to identify specific features within it.
- **T2** Looking up the targets to compare many attributes.
- **T3** Looking up the target to identify one attribute.
- **T4**
  - **T4.1** Looking up the target to identify spatial data.
  - **T4.2** Looking up the targets to compare many attributes
  - **T4.3** Exploring the dataset to summarize trends or distributions

4 Final Solution: The PACT: Broaden Your Horizons

The PACT is a visualization tool created to ease finding a country as a highly-skilled migrant to move to. The following subsections explain each of the 4 components of the app and the design justifications.

4.1 Design of PACT

4.1.1 C1 Geographical Exploration: The Map

Based on the expressiveness and effectiveness principles of visual encoding, since the dataset follows a categorical key attribute of country name, the visual channel best equipped is spatial region. With the principles of visual encoding and the inherent knowledge of a world map from a user point of view, an interactive map with support of zooming and panning became the first and central component of PACT.

As the mouse hovers over a country, the region highlights to a distinct hue, providing immediate visual feedback. Clicking on a country triggers a popup as shown on figure 1 [1], which adheres to the visual information seeking mantra. Within the popup there are values gathered regarding the indices available, what the indices mean, a radar chart, and bar chart. The values of the indices are displayed to allow for easy user comparison and understanding by clicking the link, as shown on figure 2 [2]. The radar chart allows for a rapid assessment of the country’s attributes, while the bar chart displays population indicators to help users understand the profile of the country.

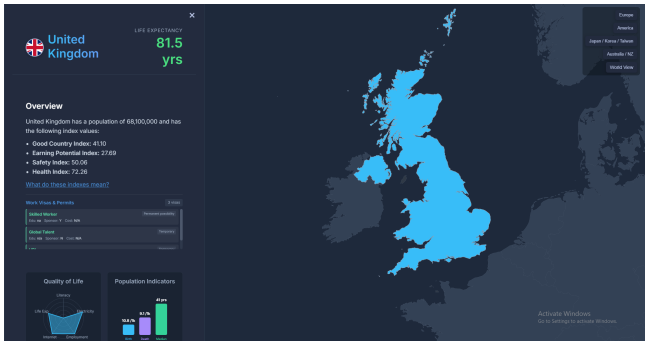

Figure 1: C1 Popup modal when clicking on a country

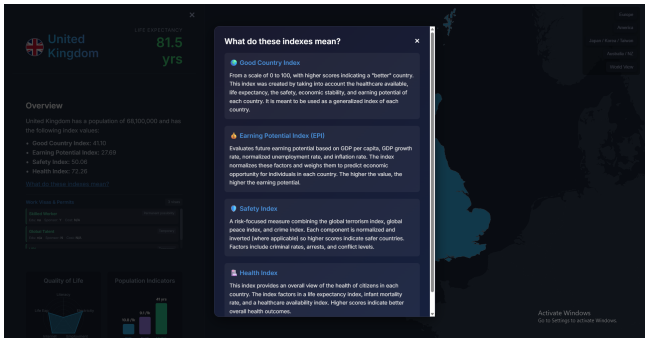

Figure 2: C1 Explanation of index values

This component of PACT fits into T3 and gives the user a starting point and initial understanding for the other components.

4.1.2 C2 Comparison and Selection

Once users clicks on the next icon on the left-hand menu, users are prompted to select as many or as few countries and attributes they would like to compare. Once the selection has been made, a bar graph along with a radar graph appears, as shown on figure 3 [3]. This is a coordinated multiple view, as when a user selects one value on one chart, the other chart is updated as well. These graphs were selected as both provide direct comparison between each attribute and country chosen.

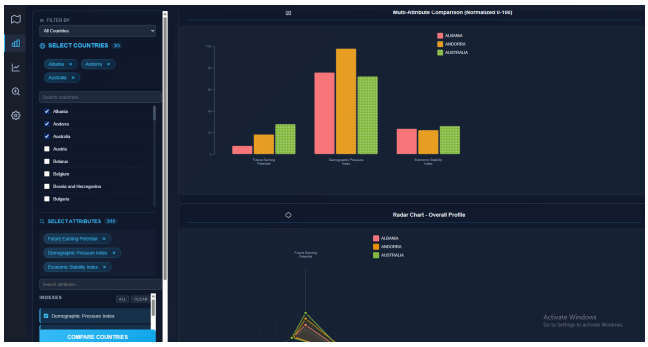

Figure 3: C2 Barchart and Radarchart

After users click on the subsection within this component, users will be prompted to manipulate sliders to assign personal importance to the attributes given. Once they have made a selection, they will be recommended the top 3 or 5 countries, as shown in figure 4 [4]. With the combination of bar graph and radar graph as in the other subsection, it will be easy to compare all countries.

These two subsections within C2 fit well within T1 and T2, as countries are identified based on criteria and countries are compared based on attributes.

Figure 4: C2 Top countries

4.1.3 C3 Index Map

The third icon contains a chloropleth map that is hover-able to show exact value. In the top right corner, as shown in the visualization, different indices can be selected to be viewed on the chloropleth map. This component would fit best under T2, as the user is able to compare countries regarding the indices.

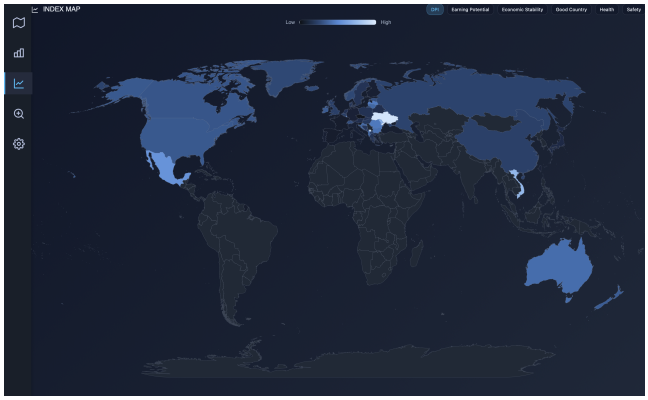

Figure 5: C3 Chloropleth map

4.1.4 C4 Global Patterns Explorer: PCP Greatness

The fourth icon on the left hand of PACT leads the user to a Global Patterns Explorer. Upon entering the icon, the user will find a PCP and scatter plot with a multitude of attributes as shown in figure 6[6]. The two graphs are coordinated to allow for selecting many different attributes to highlight. There are different attributes that the legend could be colored by to further support interaction. The PCP and scatter plot were chosen to allow for best comparison between attributes and country.

This component fits best within the task abstraction of T4.

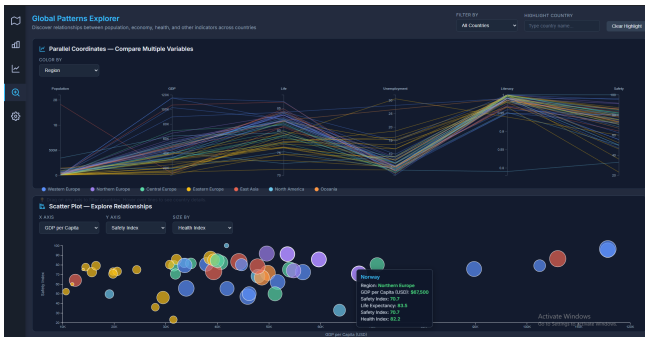

Figure 6: C4 Global patterns Explorer

4.2 Features of the solution

4.2.1 Interaction

This tool focuses heavily on user interaction and basing most of the design on interaction principles.

- **C1:** Zooming and panning across the map, selecting a country, selecting to find more information about indices.
- **C2:** Selecting countries, selecting attributes, and selecting on the slider importance of attributes.
- **C3:** Hovering over countries and selecting which indices to view.
- **C4:** Selecting filter of countries, selecting color based on which attributes, selecting x and y -axes on scatterplot, selecting how to size the scatterplot dots.

4.2.2 Coordinated multiple views

- **C2:** Selection coordination between the bar graph and radar graph
- **C4:** Selection coordination between the PCP, scatterplot, and legend.

4.2.3 Advanced multivariate visualization

- **C3:** Utilizes a chloropleth map to show index values.
- **C4:** Utilizes a parallel coordinate plot and scatterplot to allow for user exploration.

4.2.4 Navigation strategies

PACT is designed as 4 different pages, called components in the report, which are always available to click on the left-hand side. Within C1, zooming and panning is used to navigate. In C2, C3, and C4, only clicking is necessary to navigate within the components.

5 Implementation

The PACT application is based on NeutralinoJS. It uses standard web languages such as HTML, CSS and JS. The visualizations utilize D3 and GeoJSON [1].

6 Use Cases

6.1 Use Case 1: Finding the top scoring countries

6.1.1 Context

In this scenario, a family member wants to focus on finding the three top scoring countries to move to regarding education, economic stability and income level per person. The family member thus utilizes the tool to find potential countries to move to.

6.1.2 Procedure

A family member opens the application and navigates towards the data page. Here, the family member must click on ”choose importance of attributes”. On this page, the family member can then search for education, stability and income level per person to increase their weights. After, they also change the amount of countries from five to three. Then, they can press the ”recommend best countries” to view the graphs.

6.1.3 Observations

The barchart and radar plot then show the top three scoring countries.

- On the barchart the total score per country is displayed. With an arrangement from top-bottom.
- On the radar chart the weighted values are shown per country and a weighted average of all the countries is also shown.

6.1.4 Impact Analysis

As we can see from the figures [7,8] the best countries are Monaco, Luxembourg and Ireland respectively. This gives the family member a good indication to which countries they should migrate to.

Figure 7: Use Case1 Finding best countries bar chart

Figure 8: Use Case1 Finding best countries radar chart

6.1.5 Conclusion

This use case highlights the usability of the tool to quickly find the top scoring countries on multiple weighted attributes. The use of a linked bar and radar chart allow for efficient comparing between the countries. And as such shows the effectiveness of the tool. The tool proved to be highly essential in finding the countries in such a quick and easy manner.

6.2 Use Case 2:

6.2.1 Context

A young highly skilled migrant has narrowed down their selection of countries to move to based on job offers. The following countries are in their interest: Germany, Netherlands Canada and Australia. They want to perform a detailed comparison between these countries and a plenitude of quality of life factors, these being healthcare, safety, economic stability, environmental impact and connectivity.

6.2.2 Procedure

The young person opens the tool and navigates themselves to the data page. They stay in the default view and select their four countries. They then search for and select eight to ten different attributes from the different categories according to the requirements that they set.

6.2.3 Observations

The generated graphs then show the following,

- A parallel coordinate plot displaying all chosen attributes on the vertical axis, and the countries being different lines on the graph.
- The radar chart shows the overall profile of each country over all of the selected attributes.
- As the charts are linked, clicking on a line on the parallel coordinate plot will also highlight the selected country on the radar chart and vice versa.

6.2.4 Impact Analysis

From the figures[9,10] it is clear that the Netherlands performs slightly higher than the other countries. With Germany performing the worst in these attributes. The young highly skilled migrant could thus make the conclusion that they should move to the Netherlands.

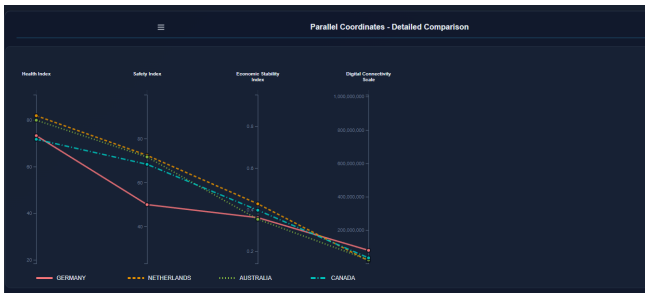

Figure 9: Use Case 2 parallel coordinate plot displaying the different attributes and countries

Figure 10: Use Case 2 Radar-chart displaying the different attributes and countries

6.2.5 Conclusion

This use case demonstrates the tool's effectiveness in detailed multi-attribute comparison of known alternatives. The tool thus proved essential in making an informed decision by revealing subtle patterns and to identify trade-offs clearly.

6.3 Evaluation

Both use cases confirm the usability of the PACT application The PACT application was evaluated against the initial tasks (T1-T4). It effectively reduced the "scavenger hunt" effect. The use of the coordinated multiple views (C2 and C4) allow users to transition between high-level filtering and attribute analysis without losing context. With Case 2 highlighting the tools ability to reveal patterns.

7 Conclusion and Future Work

By tailoring the interface of the PACT application to the specific needs of highly skilled migrants. We address the common issues related with multi-criteria attribute selection. Through the integration of a geographic chart, a weighted recommendation engine and advanced multivariate plots such as the parallel coordinate plot, PACT ensures that users can perform their own personalized analysis of different countries.

In the future, we could add a plethora of new features. These include but are not limited to a real-time currency exchange between countries' currencies, real-time tracking of inflation, mobile optimization (e.g., creating a mobile app), and community insights such as reviews or testimonials from people living in the selected country.

References

[1] A. Bonnasseau. Global terrorism index.
[2] I. for Economics and Peace. Global peace index.
[3] I. for Economics and Peace. Global terror index.
[4] GlobalOrganized. Criminal index.
[5] T. Munzner. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, 2014.
[6] Numbeo. Healthcare index.
[7] WorldBankGroup. Global inflation.
[8] WorldPopulationReview. Life expectancy index.