Visualization of Economies & Consumption Patterns

Proposal by Red Anchor

Tan Kei Rong Benjamin | Sean Chua Kian Shun | Zoey Teo Kai Ying

9/2/2015

Contents

1. Sponsor Background Introduction

Arisaig Partners (Asia) Pte Ltd is an independent investment management company established since 1996. With Singapore as their head office, they have expanded their presence into Hong Kong, Mumbai, United Kingdom and many more. Since establishment, Arisaig investment focuses exclusively on consumer sector.

2. Business Problems & Motivation

Yearly, Arisaig Partners holds a Consumer Symposium for potential clients around the globe. The symposium helps their clients understand the consumption trends around the world so they know where to invest.

However, year after year they have been using the similar presentation format - Powerpoint snapshots with excel charts. Using only tools meant for static visualization they faced problems such as:
1. Cluttered visualization display
2. Visualizations are static and cannot show the impact of changing variables in forecasts
3. Disjointed visualization that are unable to provide overview and comparison
4. Lack of dynamic visualization to showcase how variables affect each other
5. Clients are unable to interact with their charts and figures

Arisaig Partners would like to increase the effectiveness of their presentation by allowing their clients to interact with the visualizations.

3. Project Objectives

The aim of this project is to allow investors to understand the areas where they can focus their investment in, and understand the demographic and economic factors that drive the trends in consumption patterns.

The clienteles are highly educated individuals from large institutes, and our aim is to provide a dashboard that allows them to understand the trends without much guidance from Arisaig Partners investment analysts.

The objectives of our analytics project are:
1. To visualize the overall consumption pattern across the world and regions
2. Allow users to understand the impact of macro environment factors (i.e. economic, demographic) on consumption patterns
3. Allow users to interact with the visualizations to derive their own findings
4. Allow settings of visualizations to be changed dynamically to see how changes in factors affect the forecasted values (e.g. scale factor)
5. To allow clients to view the visualizations easily on their mobile device (i.e. tablets)

4. Review of Previous Work

Research has been undertaken to:
1) Understand what factors are of interest in understanding consumption patterns
2) Determine what kind of visualizations are most ideal in displaying the change in trends over time

(1) Factors of Interest in Understanding Consumption Patterns

An article by Minakshi Trivedi [1] discussed analysis performed on changes in consumer behavior patterns of healthy food. The article mentioned that both aggregated and disaggregated data should be taken into account when looking into trends. The article quoted an example that the baby food market (aggregated data) in 2006 saw a fairly flat sales growth from 2005 to 2006 and only a 3.1% increase in 2007. However, when the data is disaggregated, it can be seen that in the same time period, the growth of organic baby food sales saw a 16.4% jump followed by a 21.6% jump in 2007.The increase in popularity did not spill over to potato chips or other snack foods, and was thus unobservable at the aggregate level. Therefore, for our project, we should not only look at overall consumption patterns, but also consumption patterns in smaller product categories.

The same article also found that spatial relationships (spatial correlation, spatial causality or spatial interaction) have a significant importance in the consumption patterns. From their findings, they observed that areas that are in the same cluster produce similar test results. What this means for us is that countries that are geographically close to each other may display similar trends in consumption patterns because "near things are more related than distant things". There is likely to be more interaction between countries that are geographically close to each other, and hence they are more likely to influence one another. This is an interesting point and we will observe if there is indeed such a pattern shown in our visualizations.

The articles also discussed some demographic factors that were significant to consumption behavior. Interestingly, marriage rate was a significant factor in the consumption pattern of beer. Income and population density were also significant factors. Locations with higher income are associated with higher consumption, and for population density, as density increases, healthy consumption decreases. These findings might aid us in interpreting our findings in our project.

Worldwide trends have seen the changing attitude towards dietary habits, with respect to basic staples to packaged and fast food. [2] Changes to agricultural practices have increased food capacity and reduced seasonal dependence, resulting in considerable changed to food consumption patterns in developing countries.

A rise of almost 400kcal a day was observed globally (with some exceptions of negative growth in developing nations), and the shift of available consumption of calories to meat, sugar and vegetable oils. Also, urbanization has affected food consumption by changes in dietary behavior to favor the fast-food industry by providing quick access to cheap take-away meals. The major consequences from a nutrition perspective of urbanization are a profound shift towards higher food energy, more fats and oils and more animal protein from meat and dairy foods.

(2) Visualizations Displaying Change in Trends Over Time

In this research area, we have two goals, (1) visualization for changes over time and (2) previous visualization work for consumption pattern over time.

In the research for visualization over time, one of the most notable visualization is by Hans Rosling in "The best stats you've ever seen". The representation made by Hans Rosling combined the use of visualization together with animations to bring forth the impact. Some of the graphs that were used are:
1. Stacked graph
2. Bubble graph

Through the use of animation, the research is able simplify the visualization to avoid clustering and overload the viewers. Additionally through the use of differentiation methods such as color coding, size differentiation, the research allow the user to focus easily on what that matters to them. The tools used in Hans Rosling research showed that it is not just enough to showcase the performance on yearly basis, it is also necessary to show the path of change i.e. from one point to another. This is because it allows for the use of other simpler techniques to derive information such as Line graph.

As for previous visualization work, one research is CME Group research on "Global Consumption, Production and Trade Patterns". In this research, CME group made its case through the use of overlaying graphs one over another. Specifically, the research used bar charts over world map visualization.

MEAT CONSUMPTION BY SPECIES, SELECTED COUNTRIES
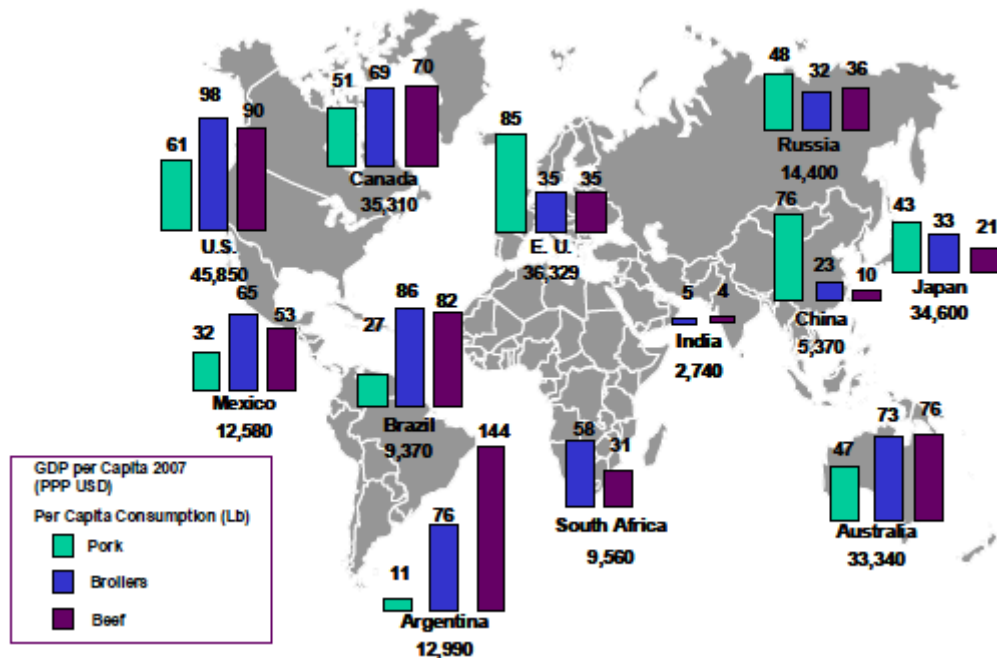
There could be many criticisms for the visualization made in this article e.g. bad use of the world map. However, this research gives us the insight that such visualization could be used for displaying additional information for the major players around the world.

5. Scope of Project

The visualization will allow users to see an overview of the consumption patterns around the world, and then drill down to each country and investigate the macro environment factors that affect the consumption pattern.

Some indicators and areas of focus that we have selected preliminarily are:
- Consumption patterns
  - Modern Retail
  - Fast Food
  - Beer
  - Oral Care
  - Packaged Food
  - E-Commerce
- Demographic factors
  - Birth rate per women
  - Average age of country/region
  - Income
  - Household size & Household Income
- Economic factors
  - GDP, GDP growth rate
  - Debt to GDP ratios, Savings rate
  - Capital Market Size

Additionally, the project has to also be sure to satisfy & produce the following deliverables:
1. Provision of use case scenarios
2. Dynamic and scalable solution that allows end user to modify specific fields (e.g. GDP growth rate predictions)

6. Possible Extension of Scope of Project

With further collaboration with the presenters from Arisaig Partners to better understand their presentation flow for the Consumer Symposium, the deliverable could be made "tablet-friendly". This improves the mobility of the investment analysts on the ground as well as delivering a better user-experience for the clienteles as they could surf it using their personal devices.

7. Timeline & Schedule

Project Timeline

Project tasks are estimated based on an average of 6 hours a week per team membe

Total project hours estimated: 252

Legend:
B - Benjamin
S - Sean
Z - Zoey

Tasks	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Requirements Gathering & Scoping														
Confirm project topic	B,S,Z													
Gather requirements from client		B,S,Z												
Brainstorm ideas for visualization		B,S,Z	B,S,Z											
Research & Preparation														
Research on finance and investment management			B											
Explore tools			Z											
Proposal														
Create proposal			S											
Create wiki page			Z											
Milestone: Proposal Submission (23 January 2015)														
Data Cleaning														
Data cleaning				B,S,Z										
Data exploration				S										
Prototype														
Create prototype				B										
Review & revise prototype with client				B,S,Z										
Development														
Develop basic structure of webpage				S	S									
Visualization 1					Z	Z								
Visualization 2					B	B								
Midterm Presentation														
Prepare midterm presentation slides							Z							
Create midterm report							S							
Update wiki							B							
Milestone: Midterm Presentation														
Visualization 3							S	S						
Visualization 4							Z	Z						
Visualization 5								B	B					
Integration									S					
Testing & Iteration														
Review system with client									B,S,Z			B,S,Z		
Iterate through system based on feedback									B,S,Z	B,S,Z	B,S,Z	B,S,Z	B,S,Z	B,S,Z
Final Presentation														
Prepare final presentation slides													B,S,Z	
Update wiki													Z	
Finalize system													B,S,Z	B,S,Z
Milestone: Final Presentation														
Research Paper														
Set up structure of research paper					Z									
Update research paper (iterative)					B,S,Z		B,S,Z		B,S,Z		B,S,Z	B,S,Z	B,S,Z	B,S,Z
Milestone: Research Paper Submission														
Poster														
Create Poster													Z	
Milestone: Poster Day														

8. Limitations & Assumptions

Some constraints posed by the sponsor (Arisaig Partners) are:
1. The application should be on a web-based platform
2. Data visualization tools used should be open-source.

9. Methodology for Chosen Solutions

Data Compilation

We have identified the following sets of data to be useful in helping understand the consumption patterns:

Demographic
- Birth rate per woman
- Average age of country/region
- Income
- Household size
- Household Income
- Age composition
- Level of education
- Population size

Economic
- GDP
- GDP per capita
- GDP growth rate
- Debt to GDP ratios
- Savings rate
- Capital Market Size

We are retrieving data from Euromonitor, World Bank, United Nations, CEIC database as they provide consumption data, economic and demographic data for each country.

Data Preparation

The data that we obtain are for each country, and the data spans across decades. The first data that we have starts from 1977, and we have data until 2030 (estimates from Euromonitor).

The first step we took in preparing the data was to understand our dataset. We look at each of the categories of data we have and decide which ones are important to us and which are the ones we don't need. There is a large number of missing data from the years before 2000. Using a dataset with large amounts of missing data will render our analysis inaccurate. Hence, we decided to focus on data after year 2000. There is still a large amount of data for us to use (30 years of data) so it is still sufficient for us to perform analysis and visualization on it.

Next, as we are concerned in patterns across regions, we classify the countries according to regions. We use the regions defined by World Bank so as to ensure consistency. The GDP of the countries are also given in their local currency, so we have to convert all of them into a common currency (USD) so that we can perform comparisons.

Technology

In selecting the technology to be used for our dashboard, we took into consideration Arisaig's specifications for a free and open source technology. We shortlisted the following tools because they are widely used and have many examples and sites to get resources from while carrying out our project. They are: Processing.js, Flot, D3.js.

Processing.js

Processing.js allows us to make visualizations without installing plugins. Processing.js is coded in the Processing language, which we are not familiar with, and the code is rendered into HTML Canvas objects. HTML Canvas loads large amount of data fast, but it is harder to modify as special programming languages are needed to create the canvas, i.e. Processing.

Flot

Flot is an open source jQuery plug-in designed for drawing diagrams in Canvas. Flot has defined charts and allows users to draw visualizations directly, unlike Processing.js and D3.js which creates shapes to make up visualizations. This makes creating simple charts easier, but flexibility to customize the visualizations is also limited. However, Flot only provides bar, line, area and point charts, which does not fulfill our project requirements of interactive and dynamic visualization.

D3.js

D3.js is a JavaScript library for developing visualizations on the web. D3.js is coded in Javascript and use SVG objects for visualization, which allows for more flexibility. SVG objects are also scalable and supports visualization on mobile devices. It is convenient as a JavaScript library can run on all modern browsers without users having to install additional software.

We have decided to create our dashboard using D3.js. In contrast to other JavaScript libraries, D3 is immensely flexible and allows great control over the final visual result, and enables dynamic and interactive graphical forms. Some challenges we might face with using D3.js is the steep learning curve, however, we have experience in developing visualizations with D3.js and the impact of this risk will be reduced.

Visualizations

Our dashboard will comprise of the following visualizations:

Sunburst chart

A sunburst chart provides a quick overview on categorical consumption data. It allows our viewers to understand and compare proportions at a glance. Sunburst charts also show patterns across hierarchical regions and countries at a glance. We will place 2 sunburst charts of data from different years side by side so users can make comparisons on the change in patterns over the years.
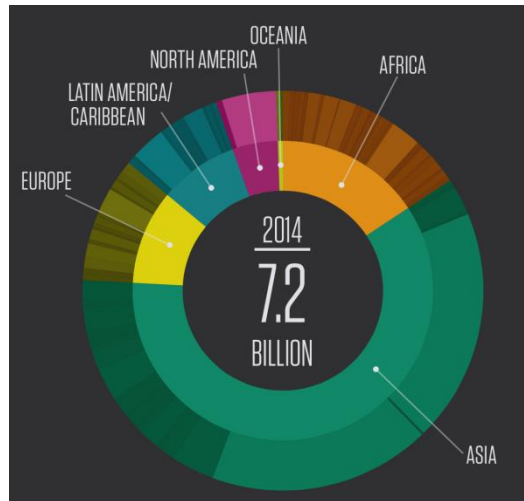
Time series line chart

This chart will be used in conjunction with the sunburst chart. The sunburst chart provides an overview whereas the time series line chart allows the user to narrow down on a particular region or country. The line chart will highlight the trends in absolute numbers and allow users to dynamically adjust the forecasted portion according to their own assumptions.
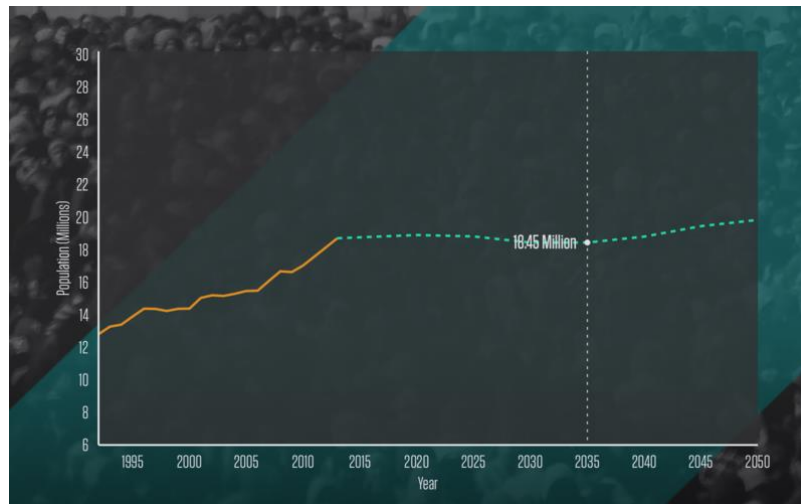
Treemap

A treemap can convey our hierarchical data with 2 additional attributes via color and size, allowing us to dissect the relationship between the two. The size of each block represents the GDP, and we can allow the user to fill in the colors with other economic indicators - savings rate, capital market size to grasp patterns in the data.
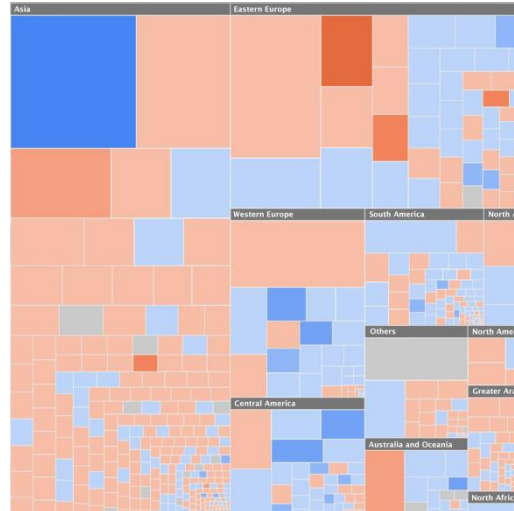
Motion charts

The motion charts will be animated according to time dimension and emphasize the growth of our 2 macro indicators relative to each other, and divulge patterns in their correlation

Heatmap

The use of heatmap is to allow for easy comparison of the countries and regions performance based on the selected variable. The benefits of using heatmap are:
1. Insights on the number of countries/regions are contributing
2. Insights on the intensity (degree) of contribution
3. Insights on any patterns of influence (e.g. possible hub responsible in any region)

Horizon charts

Not particularly intuitive, but are a space efficient way of analyzing the multiple time series data sets in our macro indicators. Horizon charts reduce space use by reducing charts to horizontal bands, making them better suited to mobile interfaces.

Box plot

Can be used to analyses larger geographical regional patterns by grouping our observed countries into regions.

10. Storyboard
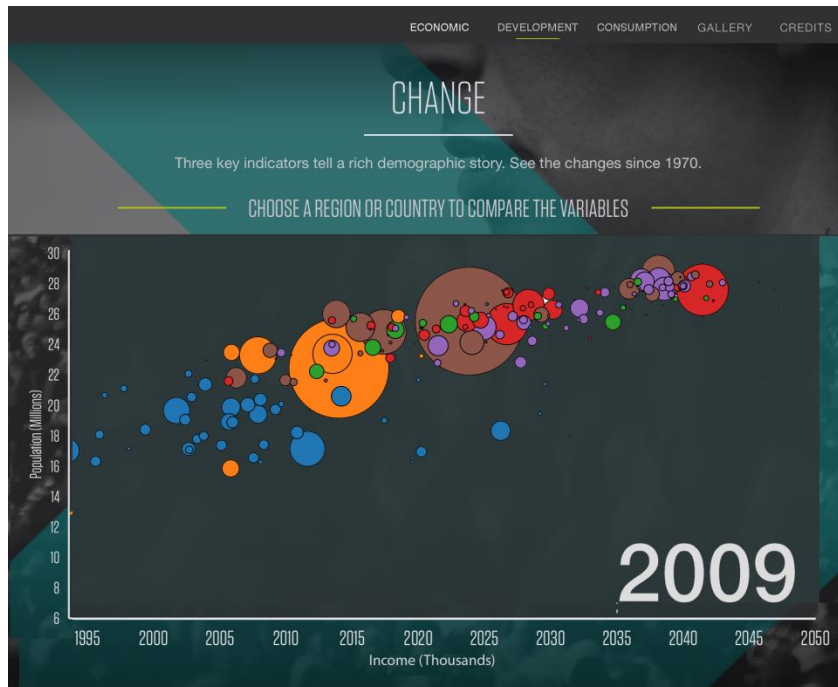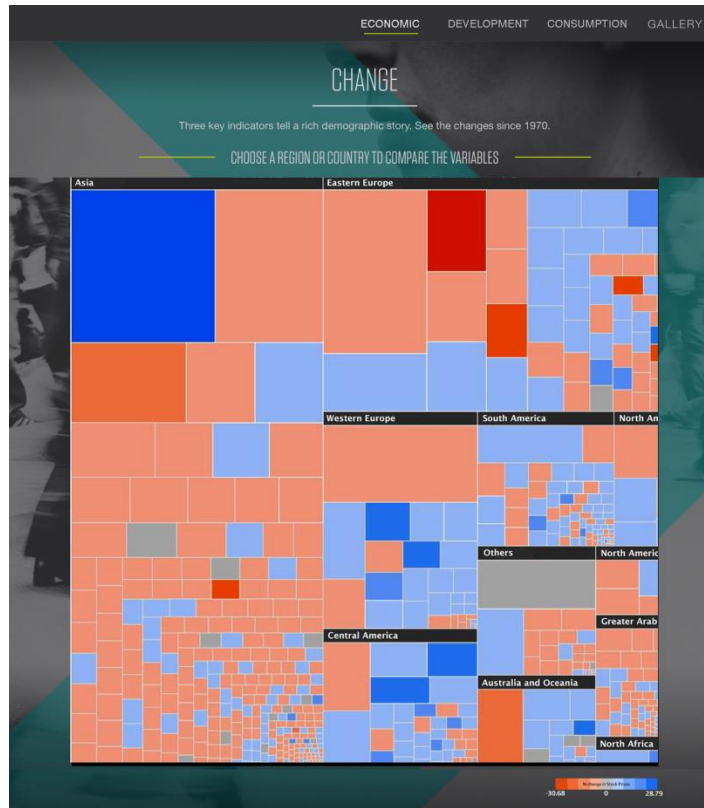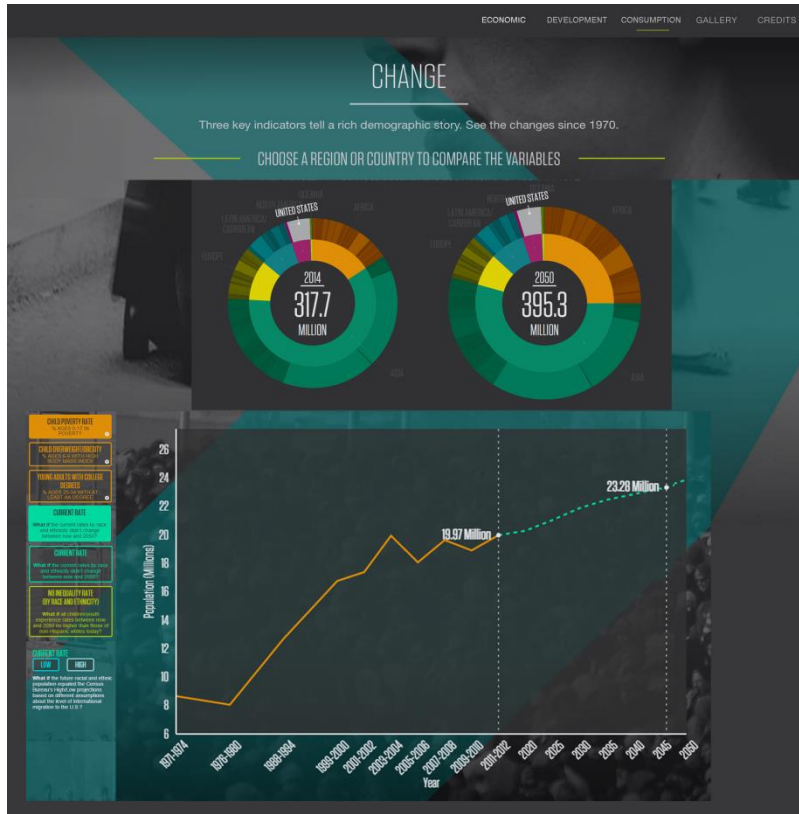
11. References

[1] Trivedi, M. (2011). Regional and Categorical Patterns in Consumer Behavior: Revealing Trends. *Journal of Retailing* , 18-30.

[2]Kearney, J. (2010) Food consumption trends and drivers. *The Royal Society Publishing*

12. Revision History

Version	Changes	Date	Author
1	• Document created	23 January 2015	Sean Chua
2	• Added in other technology considered within Methodology section	23 January 2015	Zoey Teo
3	• Added in literature review section • Added in data compilation and data preparation into the methodology section • Combined business problems with motivation	9 February 2015	Zoey Teo

4	• Added in updates on literature review • Added in heatmap as a visualization technique to use	9 February 2015	Benjamin Tan