Decade of Development?

Design Document for CS 7450 Information Visualization Project

A visualization system to help reflect over the past decade while raising questions about world development and its effects

By Abhinaya Shetty [Shetty_abhinaya@gatech.edu]

& Nitish Mhalgi [nitishmhalgi@gatech.edu]

System Overview

Motivation and Background:

This project aims at visualizing the development of countries and its impact on the environment. We are trying to see if having a higher GDP actually symbolizes a developing world.

Development of humans has always been a pressing issue throughout the history of mankind. At a point of time when the gap between developed and under-developed countries is massive, a lot of organizations are trying to make an effort to bridge this gap. But we must remember that development in GDP has effects on the population strength and CO2 emissions of the countries as well. Thus, world development in terms of GDP may not really mean development in the true sense of the word.

There is a need for tools for analysis of development patterns and key differences between countries over time. We know that pictures speak louder than words i.e. it is always easier to understand something if we can see it. Though a lot of data is available on this subject, it is not easy to look at this data and understand everything necessary to make decisions for the betterment of the world at large. This motivates us to take up the challenge of using the data that is out there to create effective visualizations, so as to facilitate the understanding of world development.

Prior Work:

World Development and its effects have been a pressing issue in the modern world, and there has been a lot of work already to study this area and to create awareness about the ill-effects it can possibly have on our environments. The most popular of them all perhaps, is the Gapminder system, which creates animated visualizations about the changes in the environment over a period of time, comparing different factors.

The World Bank regularly releases datasets regarding world development, and there have been systems in this area focusing on one particular aspect, or with one particular view, including eAtlas, which creates heat-maps projected on world-maps on different parameters of development.

Our goal in this system is to create a system that analyses effects and inter-connections between different aspects of development, and thus try to judge how the economic development compares with environmental development.

Why use visualizations?

When we talk about world development, we are talking about a large number of factors, all for more than 200 countries. If we try to view such a huge volume of data in its original format or on a normal graph, it would be very difficult to draw any conclusions.

Creating visualizations makes it much simpler to understand and compare the values of different factors like GDP, Food Production Index, etc. for different countries. While Machine Learning may provide an alternative to find clusters and patterns present in the data, we would still have to create visualizations of these patterns to make sense of them. This is the reason we concluded creating a visualization system was apt for this problem.

Users, Tasks and Data

Users of this system:

This project is chiefly aimed at the **environmental researchers, as well as the research scientists in the field of economics and world development**. It can also be used by **journalists reporting on stories based on environmental and development issues.** The use of our system for each of these users can be explained as follows:

Researchers:

- Goal: Study of population of a country, its technological advancement, or even a relation between its population and pollution level may be helped.
- Our Tool: Explore CO2 trends on a screen to see how energy use/production and population affects the CO2 levels in a country. Visualize data about a country's working-age population, technology use, etc. on the Global view screen
- Use: Could find valuable insights on the causes of alarming levels of CO2 or see the variation of levels of CO2 over time. Country comparison feature could be used to confirm some of their theories.

Journalists:

- Goal: Need to back-up a development story with relevant research reports and effective visualisations.
- Our Tool: Visualise data on the Global View page or see temporal data on the cause-and-effect graphs in different sections.
- Use: Provides visual proof to back up their proposed arguments

User Tasks and Interactions:

Tasks

When using this system, the user will first see the home-screen, which is the overview of the data that's being visualized. The project mainly consists of three modules, one dealing with the CO_2 emissions, one

with the effects of population and the last one with the global trends in factors like GDP, Energy Use, and technological advancements. The home-screen connects all of these 3 together, by a visualization plotting the GDP data, along with the CO_2 emissions and population data. In this way, we have tried to follow Shneiderman's taxonomy, of **giving an overview first and then giving the details by zooming in on demand**.

The user can try to **find correlations** among the different attributes of data being visualized, like finding if there's any correlation between urbanization of a country and its CO_2 emissions.

This system can also be used **to compare** the data of different countries, in order to find out the relative progress of the countries. Comparing against a standard benchmark, or comparing the countries which are closer geographically, can help in understanding the data better.

The user can also **detect anomalies or outliers** in the data from these visualizations, which will help him/her find out if some country is doing exceptionally well on the development front, or is doing a really bad job of controlling its emissions, and if that country's is worth following or should be avoided.

With the gradient-maps provided for the Global development goals section, the user can find the geographical **distribution** of factors like technological advancements, or efficient use/production of energy. It can help him/her gauge if having a highly developed high-income country in a region aids or hampers the growth of other countries in that region.

From all the visualizations provided in this system, the user can try to **make inferences** on the development of countries, and whether this development can maintain the ecological balance, while focusing on economic growth.

Interactions

User interactions are critical to this visualization system for effective use. The system offers several options for visualizations which the user can access by clicking on the buttons or menus provided. He can also hover over data points to find out more information about those points. The design discussion in this document lists all the interactions each particular module offers.

The user can navigate through the system, by clicking on tabs provided to switch pages. He can alter the parameters being visualized by selecting from the drop-down menus provided. For example, he can see a

Scatterplot of CO_2 emissions versus Energy use of each country, on navigating to the emissions module. He can then choose from the drop-down menu to plot Energy Production data or urban population data instead of the energy use data in the scatterplot. He can also hover over a point representing a country in the scatterplot, to find more data about that country.

Data:

Source and processing:

We will be using a part of the dataset related to world development indicators released by the World Bank. It includes data about development and progress of various countries over time based on several factors, out of which we will focus on the following:

Parameter Name	Data Format	Processing Required
GDP	In US Dollars	Divide by 1 billion to account for the spread of data
Rural Population	% of total population & Absolute values	Normalise absolute value to millions to represent better on graphs
Urban Population	% of total population & Absolute values	Normalise absolute value to millions to represent better on graphs
Total population	% of total population & Absolute values	Normalise absolute value to millions to represent better on graphs
Rural Population Growth	Annual % increase	None
Urban Population Growth	Annual % increase	None
Total population Growth	Annual % increase	None
CO2 Emissions	Metric tons per capita	None
Food Production Index	Rating out of 100	None
Energy production	kt of oil equivalent	None

Energy use	kt of oil equivalent	None
No. of internet users	Per 100 people	None
No. of telephone users	Per 100 people	None
Food Production Index	Rating out of 100	None
Depth of the food deficit	kilocalories per person per day	None

We have data corresponding to the above parameters for various countries over time. We will be considering data for the last decade [2001-2010], for this project.

The data is provided in Excel sheets. While a large portion of the data is clean and ready to be used, some data will require processing as shown above. We will be ignoring countries with missing values for this visualisation. No type conversion is required as the data is numeric.

In addition to the above data we will be using region and income group related data available at:

http://librarians.acm.org/sites/default/files/World%20Bank%20List%20Economies%20July%202010.pdf

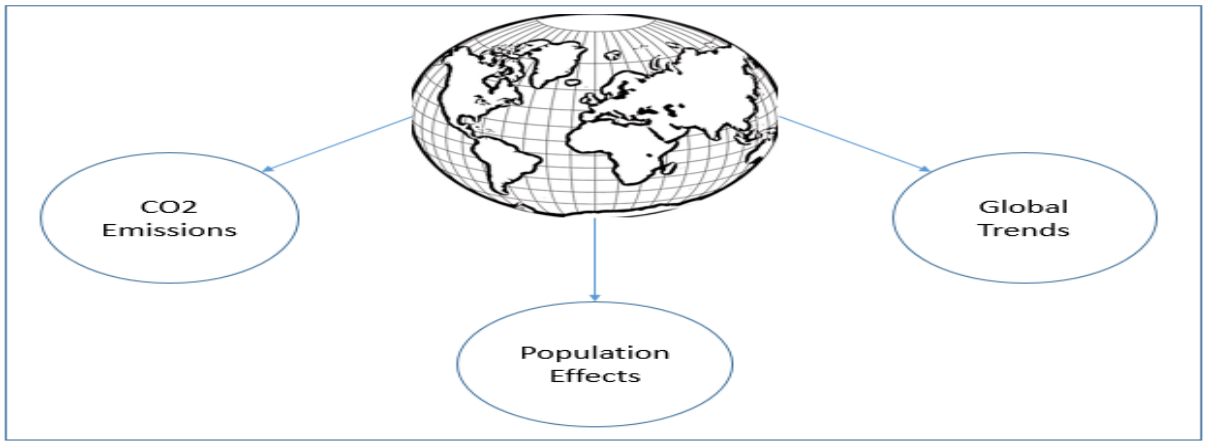
This will be used in our visualisations, to identify the correspondence between countries and regions and also determine the class a country belongs to. We will do this by a simple join on the 2 data sets.

Usage: The above data will be visualised on 4 different screens to help analyse some important side-effects of development i.e. population growth, undernourishment, CO2 emissions, energy usage, etc. In addition we will also look at the trends in growth of technology, population, CO2 emissions, etc. in each country over time. The exact mapping of data and screens can be found in the design section.

System Designs

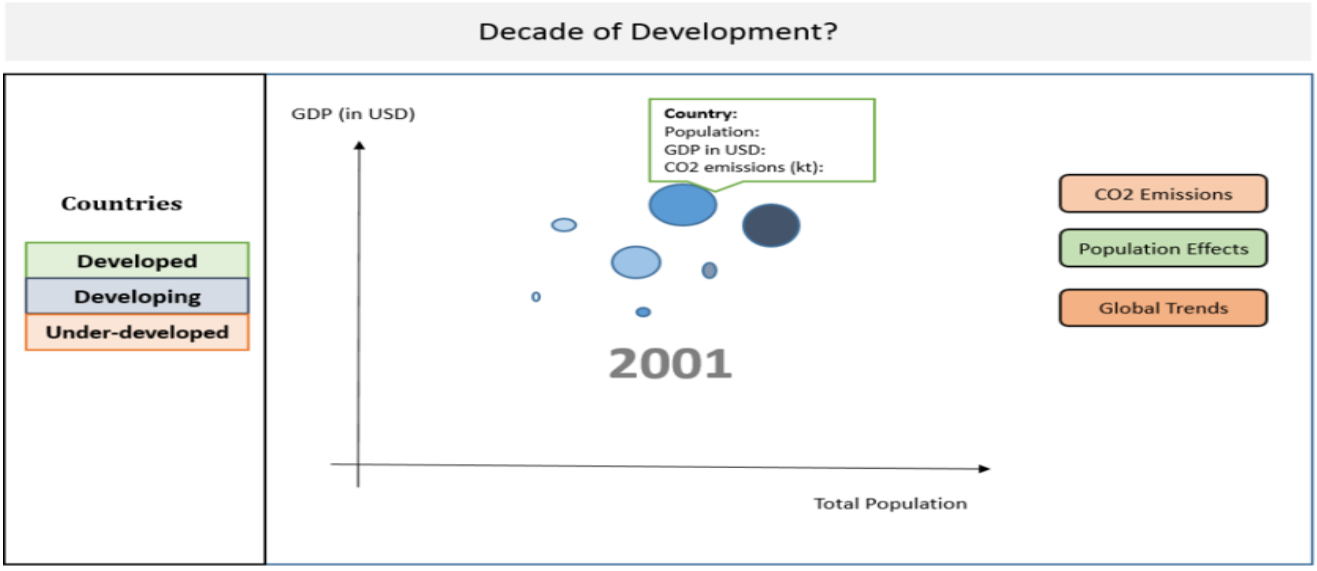
1. Homepage:

Initial Design:



We wanted to provide the user with an integrated view of the ideas being visualized. For this we came up with the above sketch. However, we soon realized that this wasn't really a visualization, as it didn't really convey any information. It would simply increase the number of clicks a user would need to reach an informative screen. It would surely be better if the default screen had some valuable information, as first impressions must be good. We thus decided to show the effects of a rising GDP on our key parameters i.e. CO2 emissions and population, over time. We decided to do this in an animated view which would show the changes in CO2 level based on GDP and population over a decade in a continuous loop, inspired by Hans Rosling's visualization on Ted Talks.

The newer version thus looked as follows:



This design highlights the primary aspects of our data i.e. GDP, population and CO2 emissions. It shows the variation in all 3 factors over time without using a 3-D graph, by encoding CO2 as the size of circles on the scatterplot. Thus, now one can pre-attentively see how each country's CO2 emissions varies according to the other 2 factors. This view is animated to show the increase/decrease in CO2 levels/ population over time to help analyze if a country is really developing. Also there is a provision to view only developed/ developing/ underdeveloped countries on the plot. This was done as we felt that showing 214 bubbles on the screen wasn't a good idea. To link this page to the next pages we have provided buttons to the right.

User Interaction:

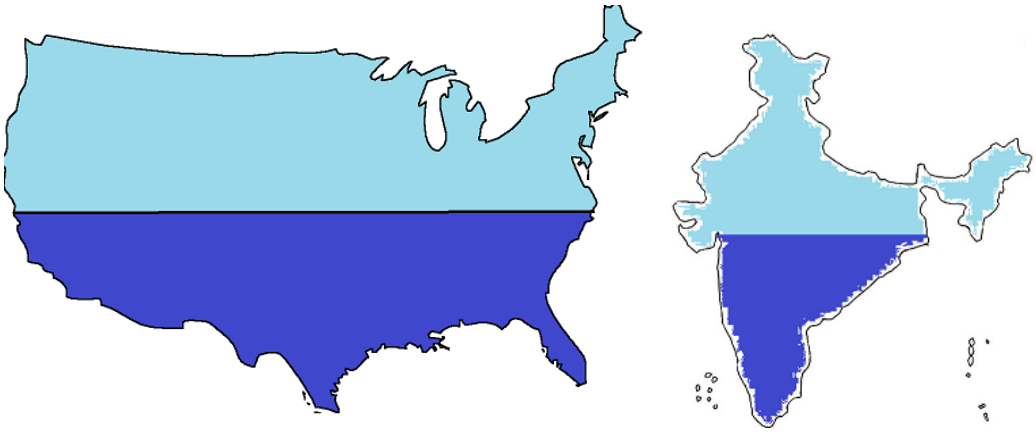
Users are provided with an overview of 2 important side-effects of development here. They can hover over the bubbles on the screen to see the actual details corresponding to each country. Since the animation plays in a loop users can see how this development and its effects have changed over time for better or for worse. Users can easily recognize if they are viewing developed / developing or underdeveloped countries by seeing the color of the graph and they can change the view using the buttons on the left. They may also wish to see a more detailed view of the aspect [CO2 emissions, population] important to their study. For this they can use the corresponding buttons provided on the right to take them to a page having additional information.

Feedback Received:

Though the idea suits this situation, it is very similar to that used in GapMinder world. As we hadn't visited this site earlier, we were encouraged to have a look at it and think about how we could make ours different from the one already available.

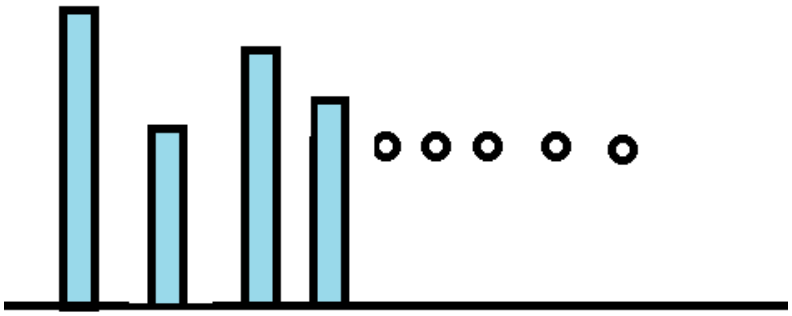
2. Population effects

Initial Idea:



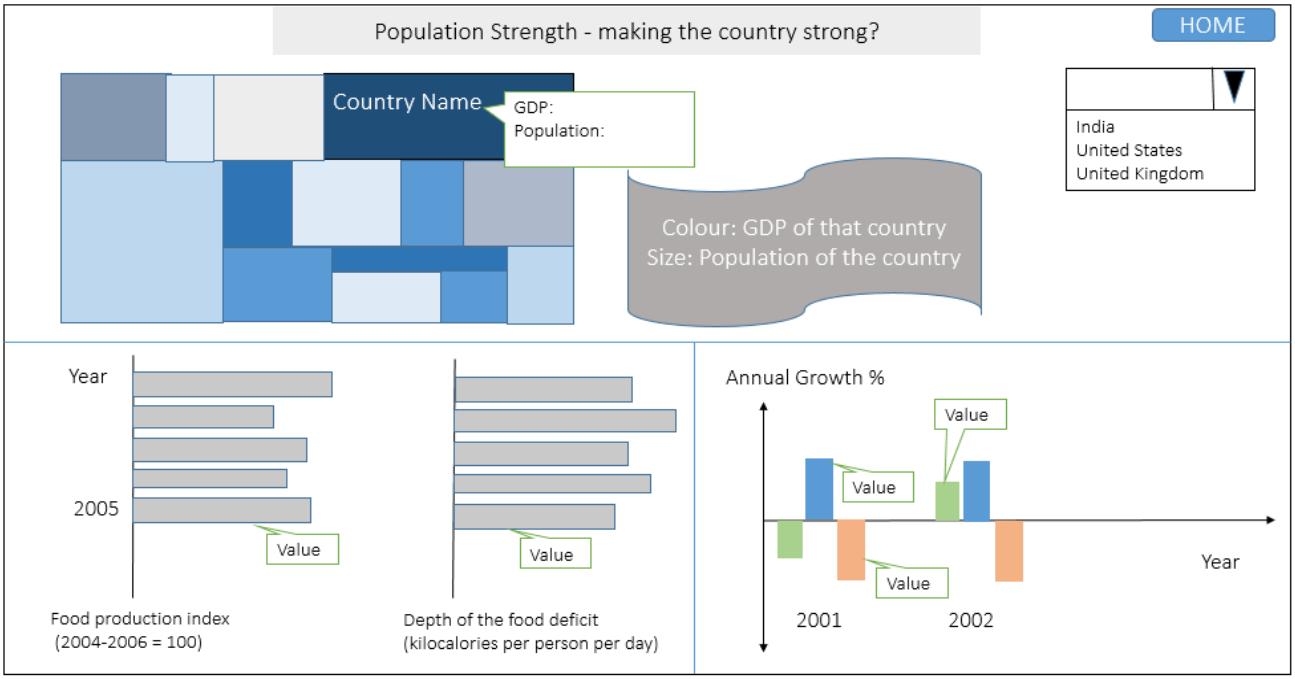
We had planned to show the population related details of each country by filling their maps in corresponding proportions as shown above. However, we then realized that it would be difficult for a user to judge what percentage of the map is actually filled while trying to quantify the information. It could also lead to misrepresentation of data because a 36% in reality may not look the same on images of varied shapes such as country maps. Hence, we dropped this idea.

We didn't want to show details related to population on a normal bar graph, as below since that would cause a lot of clutter on the screen considering that we have 214 countries to display.



We felt something that encoded area could better symbolize population and hence decided to go with a treemap for this view.

The newer version thus looked as follows:



***Tree Map**: Hierarchy: Continent - Income Group, Color: GDP of the country, Size of rectangle: Population of the country*
***Crosstabs**: Vertical Axis: timeline, Size of the bars: Value of the parameter denoted*
***Population Growth Bar-graph**: X: Timeline, Height of bars: Group of 3 bars for growth in Rural, Urban and Total Population respectively, Colors of these bars are just to differentiate 3 bars, and they don't carry the same meaning in other visualizations*

The main intent of this page is to show in detail how population is affected by GDP. This page shows a treemap in which the population of each country is encoded as the area of the blocks. This will help users pre-attentively identify if a country has a larger population.

The map is based on the following hierarchy:

Continent -> Income Group -> Country -> Population [Size] and GDP [Color]

We also provide a deviation graph on the screen so that the actual population growth of a country over time can be viewed. It includes additional details for rural and urban population. Each type of population measure [rural, urban and total] will be displayed in a different color.

Additionally the crosstabs on the screen will help visualise the trend of depth of food deficit when the food production index is high in a particular country over time. They are arranged vertically so that data can be easily compared along the time scale too. Side-by-side representation makes comparison between the two easy. These bars will be in the same colour.

User Interaction:

Users reach this page on clicking the population button on the home screen. There is a provision to go back to the home screen by clicking the home button at the top right corner. This button is consistent across the remaining pages. Through this view users can pre-attentively see which country in which continent has a larger population and what is its corresponding GDP. They can also see how the food deficit in the country is reducing/increasing over time. This would help identify potential areas of concern.

The user can also choose to go into more detail by choosing a particular country to look at, from a drop-down menu. When this is done the corresponding country in the treemap will be highlighted and the bar graph and crosstab will be updated to show details corresponding to this new country. This would help if a user wishes to study a particular country in more detail. Hovering over the different sections in the graph would display the details as shown in the figure.

Users will be able to observe the trends in the growth/decline in rural and urban population of the country, as compared to each other as well as total population growth. This can help in answering questions about whether a country is able to cope up with the rate of its urbanization and population growth.

Feedback received:

A suggestion was made to link the page to the previous by keeping some part of the data same.

3. Carbon Dioxide Emissions:

We wanted to analyse the emissions of countries, based on different aspects of their growth such as:

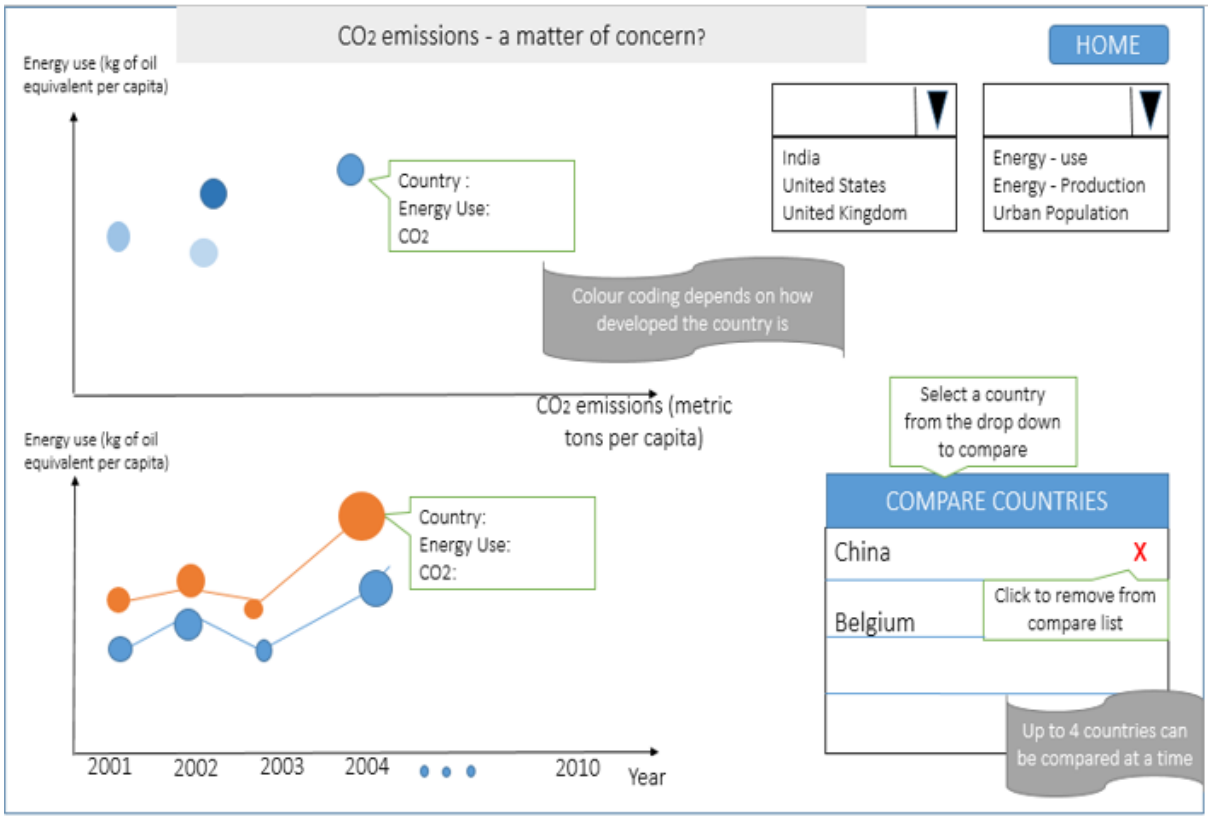
- GDP
- Energy use and Production
- Urban Population

Initial Idea:

We first considered creating bar-charts to show comparisons of CO_2 emissions against different parameters as discussed above. But the main flaws in this design were:

Stacking 214 bars on one screen doesn't create a visualization which is easily understandable to the user. Even dividing the countries into sections like developed countries, developing countries and under-developed countries was not sufficient to overcome this.

Thus, we decided to use a Scatterplot to represent this data, showing trends present in it across countries which helps in understanding the data better.



Scatterplot: X: CO2 emissions, Y: Drop-down menu, Color of bubble: Income of the country,

Line graph: Size of bubble: CO2 emissions, X: Time-period, Y: Drop-down from Scatterplot, Color: Income of the country

This also allows plotting three different parameters together, something not possible with the bar-charts. Our design thus has CO2 emissions along X-axis, Energy use along the Y-axis, and GDP represented by the size of the circle for each country on the Scatterplot.

Main advantages of this design choice are:

- Y-axis of both graphs can be modified in order to update the plots
- Allows both spatial and temporal analysis.
- Since comparison between countries is a key factor in understanding development, we have a provision to select up to 4 countries whose data can be compared.

User Interactions:

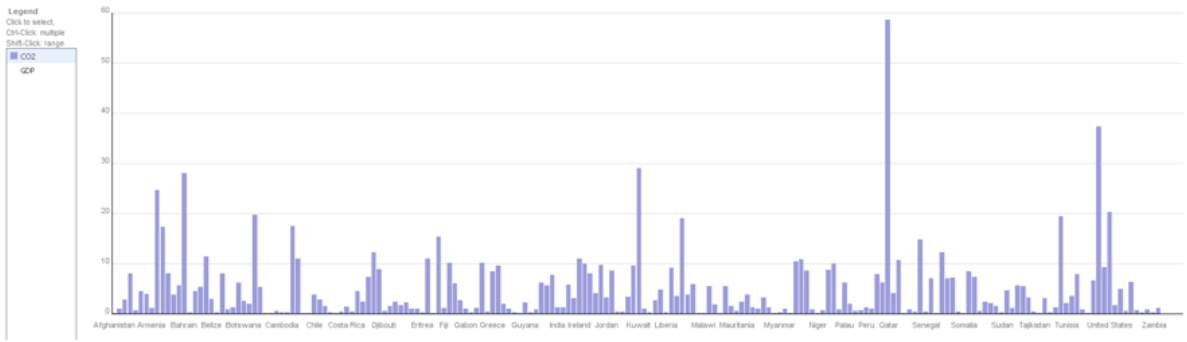
Here, users can observe the scatterplot of actual emissions against the energy use of a country. One can choose to observe the comparisons of emissions against any factor selected from the drop-down menu.

One can observe the changes in emission levels for one or more countries over the last decade 2001-2010, by selecting names from the drop-down menu. The corresponding line-graph shows changing values of the emissions from that country over time. Using this feature a comparison across countries can be seen. Hovering over a bubble in both graphs will give the user information about the country, including its name, energy use and the emissions it produces.

4. Envisaging Global Trends:

Initial design:

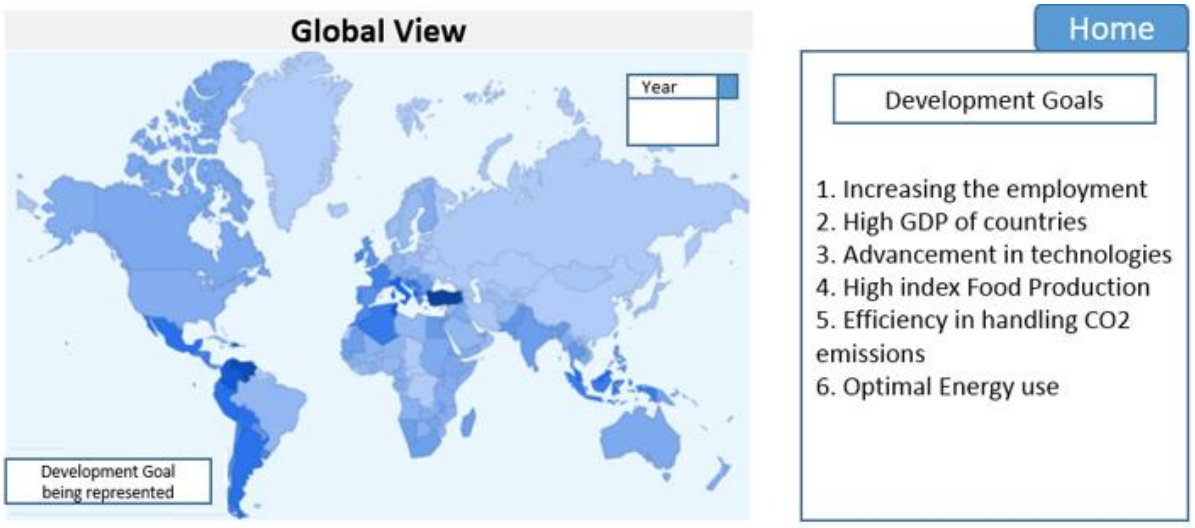
The idea of this module is to give the user an idea about the geographical distributions of different parameters, like GDP of countries, their efficiency in handling CO_2 emissions, and their energy usage.



Possible option considered, Height: Value of CO2 emissions.

We thought of creating bar graphs, but we felt it was too simplistic, which could be done easily on a tool like Tableau, and would not prove much effective in visualizing a trend either.

We decided to visualize this with the help of World maps encoded with color gradients to show these trends.



Color Gradient is used to represent the value of a particular parameter, e.g.: GDP, of each country on the map

This module aims at allowing the user to look at a world map and observe how far any region has reached, in attaining a particular development goal.

User Interactions:

On this page users can observe the geographical distribution of various parameters by observing the intensity of colors on the world map.

The user can select a particular goal he wants to look at, from a menu provided alongside the visualization. He/she will then be presented with a World-map, color-coded with a gradient, to represent the value of that particular parameter for every country. The user can hover over any country, to look at the country name and the absolute value of that parameter. He/she will also be provided with a drop-down selection, to choose which year's data he/she wants to look at.

Design intent

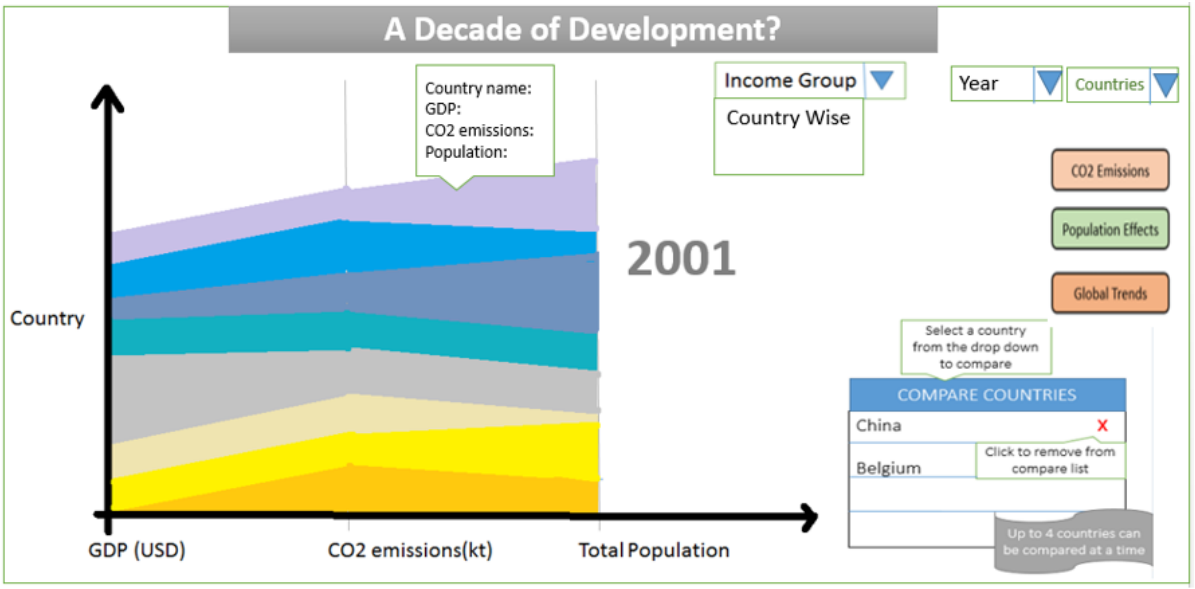
While the feedback from our poster presentation noted the clear description of our design decisions, it also asked us to elaborate on how our work differs from an existing system, Gapminder. It also encouraged us to talk in more detail, about the tasks for the users of this system.

How is our project different from Gapminder?

Gapminder is primarily concerned with showing changes in various factors related to world development over time. Our system is primarily aimed at finding out the relations between a country's development and its effect on environment, i.e. the GDP of the country against extent of its urbanization, and the harmful CO_2 emissions it produces. It aims to check if a country's economic progress has adverse effects on the environment.

Final Design:

The feedback from the poster session brought one major problem in our design to light, our homepage was similar to an existing design. We thus, needed something better for showing the information on our homescreen. As there were many designs showing scatter plots already available on the internet we decided to go for something different. We will now be using a categorised stack graph instead as shown below:



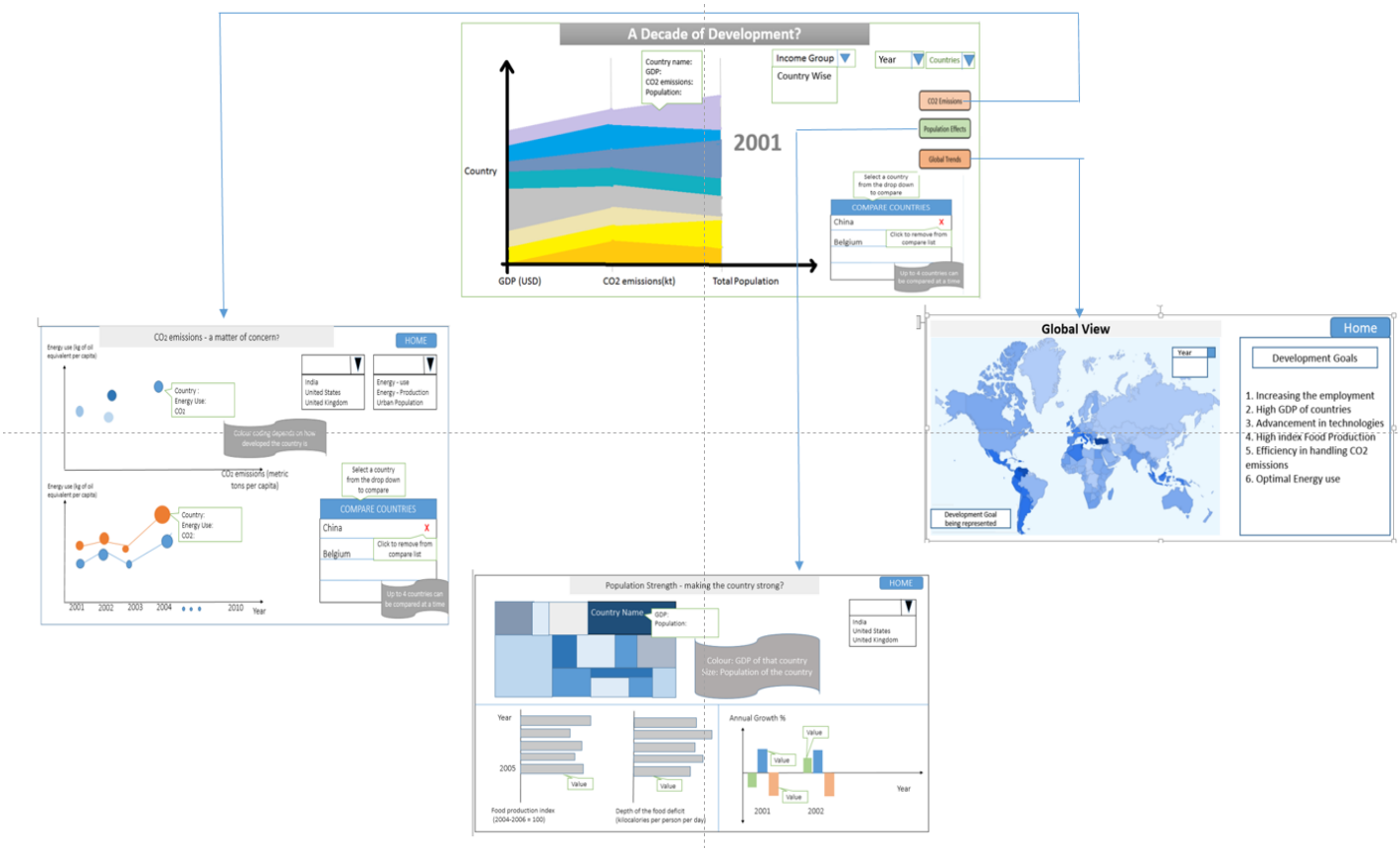
Stack Colours: Will be a gradient of the income group of the country

We will be showing the 3 important parameters on the 3 parallel axis as shown above. The data will be displayed year wise and country wise. As we have income group information in addition to country names we will be using that to show average values per income group too. There is also a compare option available to compare data for different countries.

User interaction:

Users are provided with an overview of 2 important side-effects of development here. They can hover over the areas on the screen to see the actual details corresponding to each country. Users can see how this development and its effects have changed over time for better or for worse by changing the year from the dropdown provided. An option to view average values over an income group as opposed to details for each country has been provided on the right. Selecting countries on the right adds them to the compare box on the bottom right corner of the screen. Up to 4 countries can be compared at a time. These countries will be highlighted on the graph to help visualize their differences. To see a more detailed view of the aspect [CO2 emissions, population] important to their study users can use the corresponding buttons provided on the right to take them to a page having additional information.

The rest of our intended design follows from our preferred choices made for the poster session. Each of those design choices and details about them have already been discussed in the design discussions section. The overview of our design can thus be represented as follows.



Final System: A Decade Of Development

Implementation:

We are planning to implement this system using the D3. So, we will be making a collection of web-pages to create these visualizations. The data will be contained in files in Comma Separated Values (CSV) format. We will be using color scales from color brewer as we need a wide range of colors.

As we are only two group members in this project, we are planning to divide the wok equally. Thus, each of us will work on half of the total different screens for the project, while performing testing on each other's work. The integration of modules into the complete system can then be done jointly.