

Theory Assignment: Analysis of Gapminder

Davide Pedranz

d.pedranz@student.rug.nl

Student Number: S3543757

November 17, 2017

CONTENTS

1	Application Domain and Visualization Solution	2
2	Problem Description	2
2.1	Questions	2
2.2	Users	4
2.3	Purpose	4
3	Data Description	4
3.1	Data Attributes	4
3.2	Data Model	5
4	Visual Design Study	6
4.1	Type of visualization	6
4.2	Visual encoding	6
4.3	Improvements	20
5	Description of Findings	22
5.1	Life Expectancy	23
5.2	Income Trend	25
5.3	World Situation	25
6	Conclusions	26

1 APPLICATION DOMAIN AND VISUALIZATION SOLUTION

The visualization solution analyzed in this assignment is Gapminder¹. Gapminder is a tool that allows to explore a dataset with about 500 indicators measured around the entire world over different time periods. Each observation in the dataset represents the value measured for one of the indicators for a specific country in the world in a given year. Gapminder provides data for all (or almost all countries) in the world. Countries are organized in groups by geographical region.

Indicators are organized in hierarchical categories (economy, education, environment, health, population, etc.) and each category contains one or more dimensions. Example of indicators available in Gapminder are:

- child mortality rate;
- income per person;
- CO2 per capita.

It is also possible to load and visualize additional indicators provided by the user.

The tool allows to visualize and compare the dimensions in the dataset using the following type of charts:

- “bubbles” (bubble charts);
- “maps”;
- “mountains” (area charts);
- “rankings” (bar charts);
- “lines” (multivariable line charts).

From the large number of available indicators, I have selected the followings for this assignment:

- time (expressed in year);
- geographical location (country, continent);
- average income per person (GDP /capita, PPP\$ inflation-adjusted);
- average life expectancy (years);
- total population size.

Gapminder is available in 2 versions: an online² and a desktop³ one. Both versions seems to be similar to each other, but in this assignment I will focus on the latter.

2 PROBLEM DESCRIPTION

2.1 QUESTIONS

Gapminder tries to answer to many questions. We start from low-level concrete questions, then try to use them to understand the real problem that the tool aims to solve.

¹<https://www.gapminder.org>

²<https://www.gapminder.org/tools/>

³<https://www.gapminder.org/tools-offline/>

CONCRETE QUESTIONS

- Where are the countries with the largest population located?
- Which continent has the largest population? And the smallest?
- What is the life expectation of the world population?
- Which countries have the highest life expectation? Which have the lowest?
- How did the life expectation of people living in Europe change from 1800 to today?
- Was there any drop in the average life expectation during the world wars? In which countries?
- Which countries have the highest income? Which have the lowest income?
- What is the average income of the population in Africa?
- Did the average income of Asia increased from 1800 to today?
- Is there any country with high income but low life expectation? What about the opposite?
- Are the average incomes of small countries above or under the world average?
- Where are the countries with high income and life expectation located?
- Do small countries have a longer life expectation?

HIGHER-LEVEL QUESTIONS

- What is the population distribution in the world?
- Is the size of the population changing over time? Where is it changing? Where is it not?
- What is the global trend of the life expectation? Is it different for different continents?
- Do countries closed to each other have similar life expectations?
- What is the global trend of the average income?
- Do countries closed to each other have similar incomes?
- Is there a correlation between life expectation and income today?
- Did the correlation between life expectation and income change over time?
- Does the size of a population influence the life expectation?
- Is there a correlation between population size and average income?
- Are there outliers countries in the different continents (in terms of life expectation and income)? How are they different from the other countries?

SUMMARY

- What is the world situation in terms of population, average income, life expectation?
- How did the world situation change over the last 2 centuries?
- Can you predict the trend for the next years?

The central element in Gapminder is time, which is present in all visualizations. The tool visualizes changes and trends for the selected indicators for different countries over time. Even though it is possible to build visualizations where time is not used, Gapminder is not designed for that. We will discuss in greater detail how time is visualized in Gapminder in Section 4.

2.2 USERS

Each indicator in Gapminder has time and geographical information. Indicators range among very different categories and the tool makes trivial to visualize them together and compare the situation of different countries from different points of view. The tool is very easy to download, install (there is even the online version) and use. The offered visualizations are simple and relatively easy to understand even for people without any previous knowledge about the data.

The range of users for this tool is wide. In general, we can say that the tool is designed for the grand public, for everybody which is curious in the exploration of social/political/historical/economical trends for different countries or continents in the world. The tool is also perfect for teachers who want to present the world situation (and its evolution over time) using concrete data and effective visualizations (Gapminder's website has an entire section dedicated to teachers⁴).

Each indicator in Gapminder provides a detailed description, the sources, the confidence about the measures, and describes how data are collected, aggregated and normalized. However, this information is not explicit in the tool, but is only available from the Gapminder website⁵. Gapminder is not designed for professional economists, analysts and researchers, since they usually need to analyze this kind of information in their work.

2.3 PURPOSE

The purpose of the visualization is to present a picture of the global world situation from many different points of view. The tool allows to easily compare many different indicators, analyze the situation of different countries in the world, study the historical evolution of the global world situation and predict trends for the next years.

On the one side, Gapminder can be used to confirm the known. The tool can be used to both confirm or discard hypothesis about the world current and past situation. For example, one could think that wars influence the life expectation of a country. Gapminder can be used to check the life expectations during the first and second world wars of countries in different continents and compare their situations: the visualization will confirm that the countries directly involved in the wars had a remarkable drop in life expectancy. The tool allows to answer both generic question (e.g. the distribution of values for a certain indicator over the world or correlations between different indicators) and specific questions (e.g. countries with the highest value for a given indicator).

On the other side, the tool can be used to discover the unknown. The user can select and combine different indicators and analyze their evolution over time. During this process, interesting patterns can be visualized. One can notice, for example, that there is a sudden drop in the population life expectation in both Ukraine and Russia in 1933 and investigate to what it is due⁶. One can then change the visualized indicators and further explore interesting aspects.

3 DATA DESCRIPTION

3.1 DATA ATTRIBUTES

Each observation in Gapminder is the value of an indicator measured for a specific country in the world in a given year. The country can be seen as an independent variable, since it is not influenced

⁴<http://www.gapminder.org/for-teachers/>

⁵<https://www.gapminder.org/data/>

⁶<https://en.wikipedia.org/wiki/Holodomor>

by the indicators (e.g. the average income does not determine what it the country). All measured indicators are dependent variables, since they refer to a specific country. In addition, indicators change over time. The dataset can thus be seen as a 4-dimensional dataset, where the first dimension (country) does not change over time, while the other dimensions (income, life expectation and population size) are time dependent. The geographical location of countries (i.e. the region of the world to which they belong) should not change over time, so this does not represent an extra dimension (there is a functional dependence between the country and the geographical location as a region of the world).

GEOGRAPHICAL LOCATION The geographical location is expressed in terms of countries, grouped into world regions. Since it does not exists a natural ordering of countries, the attribute type is categorical. Gapminder has data about 244 different countries in the world, grouped in 4 regions (which approximate the world's division into continents).

AVERAGE INCOME The average income is expressed as the gross domestic product per person, adjusted for differences in purchasing power (measured in international dollars). The attribute is clearly quantitative. Values range from 300 to about 150.000.

LIFE EXPECTATION The life expectation is expressed in years. The attribute is an average and thus a continuous values. Values range from 1 to about 85.

POPULATION SIZE The population size is expressed in number of inhabitants. This is clearly a natural number, so the population size attribute type is discrete: the normal equality, ordering and sum defined in \mathbb{N} apply also for the natural numbers. Although it is possible to compare ratios between population sizes, we can not say that the attribute type is continuous (quantitative). The population size ranges from 100 to about 1.400.000.000 inhabitants.

TIME Time is measured in years and is clearly a discrete attribute. It is possible to state if 2 given years are the same of not by simply comparing their number. The “less than” $<$ relationship of the set of natural numbers \mathbb{N} defines a total order between years. It is also possible to sum or subtract a natural number for a given year and the result is still a year. The dataset contains points for the time range 1800 to 2015, for a total of 215 years.

3.2 DATA MODEL

3.2.1 INTERNAL REPRESENTATION

Each indicator in Gapminder is internally represented as a table: the x-axis contains the time in years, the y-axis the country, and the data in each cell the value measured for the indicator in the given year and country.

Not all indicators in Gapminder have the same size. Most of them are defined for a small time period (few years) and have missing data for some country and / or some particular year. Some indicator are not measured for some countries. Missing data corresponds to empty cells in some of these tables.

3.2.2 ALTERNATIVE REPRESENTATION

An alternative representation of the data is a set of tables (one for each year) `dataset_<year>`, where `<year>` is replaced by the year to which the table refers. Each table has 4 columns: `country`, `income`,

life_expectation and population_size. The column country is the primary key. Each row in the dataset represents the situation of a country, measured in terms for the 3 indicators described above for the year specified in the table's name. In other words, each table represents the projection of the entire dataset for a single year. The same representation would also apply to a bigger subset of indicators: we would simply add an extra column for each considered indicator.

Since all tables have the same structure, they can be combined in a cube: the first dimensions represents the countries, the second dimension contains all indicators measured for the countries and the last dimension present time. The projection of the cube along the third dimension (i.e. filtering the data for a specific year) results in the tables described in the previous paragraph. The projection along the second dimension (i.e. selecting a single indicator) gives the representation described in Section 3.2.1. The projection along the first dimension (i.e. fixing a country) represents the evolution over time of all indicators for a country and can be seen as a table with the indicators on the x-axis and time in the y-axis.

Additionally, countries are grouped into regions. This can be represented with an extra table countries with 2 columns: country (primary key) and region. The column country of each table dataset_<year> is a foreign key for countries.country.

4 VISUAL DESIGN STUDY

4.1 TYPE OF VISUALIZATION

Gapminder is an Information Visualization (InfoVis) for the following reasons:

- The domain of the data is discrete: Gapminder visualizes indicators measured over different countries in the world. The space of countries is clearly discrete.
- Data attributes have different types: some are categorical (e.g. countries), some are discrete (e.g. time in years), some are quantitative (e.g. average income).
- The dimensionality of data values is high. We consider only 3 indicators for each country, but the complete dataset of Gapminder has over 500 indicators.
- All dimensions (except the geographical location) are time dependent, i.e. indicators are measured and vary over time.
- The visualization is interactive and has the scope to look for correlations between different indicators about the world situation.

Gapminder is clearly not a Scientific Visualization (SciVis). SciVis data have a spatial domain, numerical values for the data attributes and a low dimension. None of these properties apply to the dataset of Gapminder.

Gapminder is also clearly not a Infographic, since it is interactive and allows to explore different indicators. Infographics are static, visualize summaries of the data and have the goal to present some content to a large audience.

4.2 VISUAL ENCODING

Gapminder offers 5 different types of chart, called respectively “bubbles”, “maps”, “mountains”, “rankings” and “lines”. Since each chart uses the visual variables in different ways, we will analyze each of them separately.

For each chart, the screen is divided in different sectors: a navigation bar in the top, the chart in the center, a slider at the bottom and a side bar on the right.

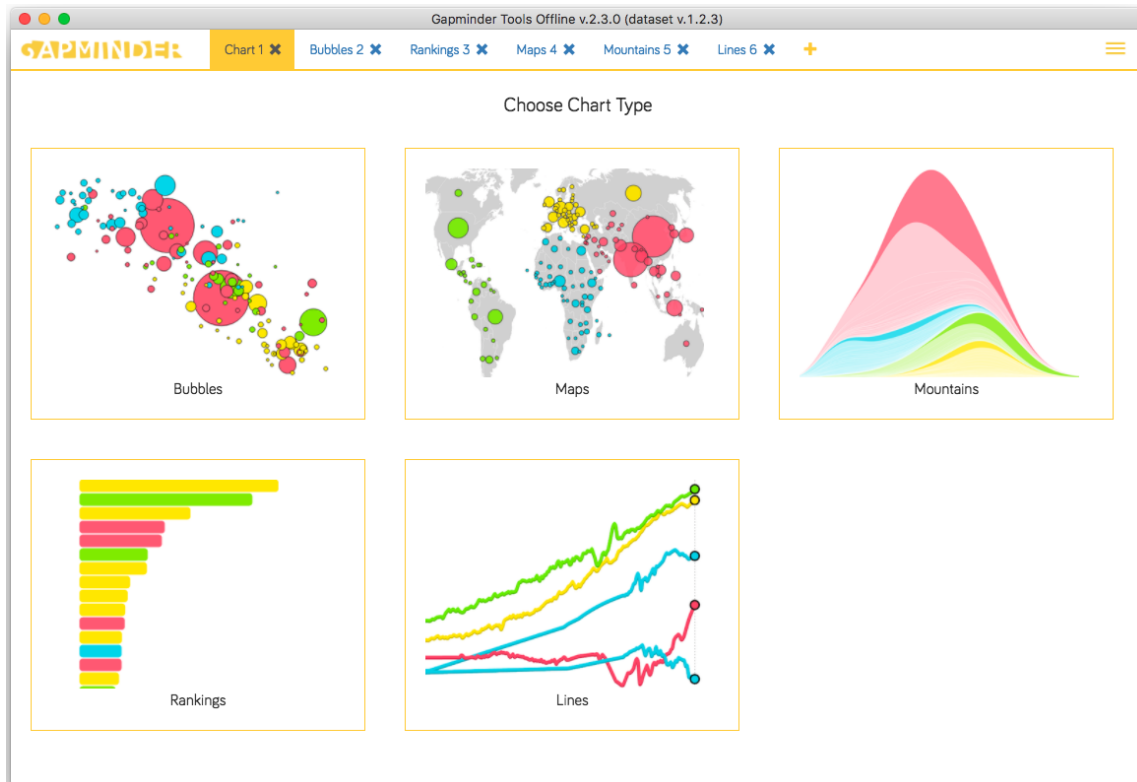


Figure 4.1: Gapminder allows to work on multiple visualization at the same time. When creating a new chart, the user is asked to select the wanted type among the 5 available.

NAVIGATION Gapminder allows the user to create and use multiple charts at the same time: each chart is visualized as a tab in the navigation bar. The navigation bar is located in the top part of the windows and is always visible. The user can change the chart currently visualized by clicking the corresponding tab in the navigation bar. To create a new chart, the user can use the “plus” button at the end of the navigation bar. When the user clicks the button, a new tab is created and displayed (Figure 4.1): the user can choose a type of chart by clicking the correspondent button.

RIGHT PANEL The panel on the right contains different controls (note that the panel is hidden during the selection of the chart type); an example is shown in Figure 4.2. We will describe the controls from the top to the bottom. At the very top, a menu allows the user to change the variable encoded as color. Immediately below, Gapminder shows a colormap for the chosen variable. Next we find a list of checkbox, one for each country. This is a filter: it allows the user to select or deselect the countries to show. At the very bottom there are some buttons that allows to control the zoom level, change the settings or activate the presentation mode. Depending on the type of chart, Gapminder shows additional buttons in this area; we will discuss this aspect in the sections about each chart’s type.

CHART In the center-left part of the screen there is the current chart (Figure 4.2). The chart occupies most of the area available on the screen.

BOTTOM SLIDER In the bottom part of the screen we find a “play” button and a time slider (Figure 4.2). The button is used to start and stop the animation. The slider allows to manually change the visualized year.

4.2.1 BUBBLES

Gapminder’s “bubbles” are bubble charts with animations. We consider the particular instance of bubble chart with the income on the x-axis, the life expectation on the y-axis and the population size encoded as the bubbles size; this is the default visualization offered by Gapminder when creating this type of chart.

The visual variables being used are: position, size, color, text labels, animations and transparency.

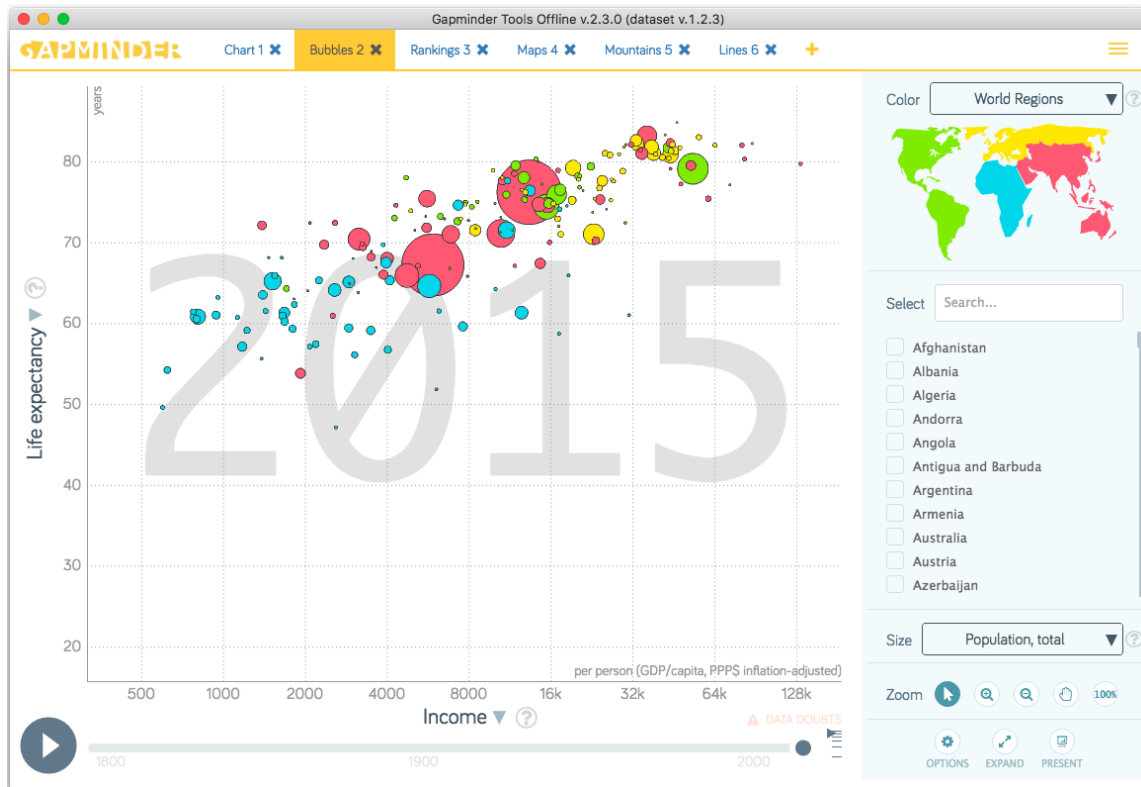


Figure 4.2: Chart of type “bubbles” in Gapminder.

POSITION Position is used to encode the value of 2 indicators: the position along the x-axis encodes the first indicator (in our case the income), the position along the y-axis the second one (in our case the life expectation). Each point in the chart corresponds to a pair of income and life expectation. By default, Gapminder limits the range of values showed on each axis between the smallest and the biggest value for the attribute displayed on that axis (plus some additional margin to make sure points do not overlap with the axes). It is possible, if needed, to manually change the ranges using the “Options” menu at the bottom of the right side panel.

Each bubble in the chart represents a country. Position is a very powerful visual variable: it is associative, selective, ordinal and quantitative. The user will thus:

- Distinguish relatively easily different points (i.e. different countries).
- Perceive each point independently of the other visual variables used. This allows to use other variables (eg. color or size) to encode additional dimensions of the data.
- Be able to easily order the points. This allows to easily tell which countries have lowest and highest values for one indicator and order the countries based on some indicator.
- Be able to easily compute ratios between the values of the indicators encoded by each point.

This allows to compare the situation of pairs of different countries from the point of view of the 2 indicator.

However, there are some limitations with this encoding. For example, countries with the same income and life expectation will be shown in the same position: it would be impossible for the user to distinguish them. Also, bubbles with a big size tend to overlap with the others (since they take a lot of space on the screen). Gapminder tries to reduce this problem in 2 ways: it shows smaller bubbles above bigger bubbles and it interactively shows the name of the countries in a label when the user moves the mouse on a bubble.

Overall, the encoding is good: it uses a single visual variable (position) to represent 2 dimensions of the dataset. Although it is possible, it is unlikely that 2 or more countries have exactly the same income and life expectation. Gapminder takes some measures to reduce the problem of overlapping bubbles. It also allows to encode, with some limitations, other attributes in size and color.

Position is also used in the colormap for the world's regions, as explained in the colors paragraph.

SIZE Size is used to encode the third indicator, in our case the population size. In particular, the value is encoded as the area of the bubble.

Size is selective, ordinal and quantitative. The user can thus:

- Easily perceive countries with different populations as different from each other.
- Easily order countries by population size.
- Easily compute ratios between the size of the populations of different countries.

However, the tool does not visualize a legend for the sizes of the bubble. This makes difficult to perform the inverse mapping from area of the bubble to size of the population by just looking at the bubble chart. The value is visualized on the right side of the screen only if the user moves the mouse over a bubble.

The encoding of the population sizes is overall clear, but it can be improved. On the one hand, the user is able to distinguish and compare the population of different countries. On the other hand, it is difficult to understand the real absolute values of the populations.

COLOR Color can be used to encode any of the available dimensions of the dataset. In our case, the tool uses the color to encode the world region to which each country belongs. Gapminder shows a colormap in the top right corner of the windows: the colormap is visualized a small stylized planisphere where each continent is filled with the color that corresponds to its region. The tool differentiate 4 world regions: americas, africa, europe and russia, asia and oceania.

The colormap is categorical and uses different color hues for each different value. The visualization uses the following 4 colors: green, blue, yellow, red. The choice of the color is good, since they are all distinct. On the other hand, this choice does not take into account colorblind people that might have difficulties to distinguish green and red. There is no way to change the colormap or force a different mapping.

Each bubble has a black border: this makes it easier to identify a bubble and distinguish it from the other ones. The visualization uses a white background, with a big gray text label in the center (to visualize the year, see the animation paragraph). This is a good design choice, since both white and grey are neutral color that do not interfere with the visualization and create a high contrast with the colors used for the bubbles.

The encoding is clear: the 4 chosen colors are very different from each other, so it is very easy to distinguish point that corresponds to countries in different regions. The inverse mapping from color

to region of the world is also pretty straightforward, since there are only 4 possible regions.

The color of a bubble together with the colormap allows to match each country with a specific location on the planisphere. In other words, this allows the user to approximately locate the geographical position of each county.

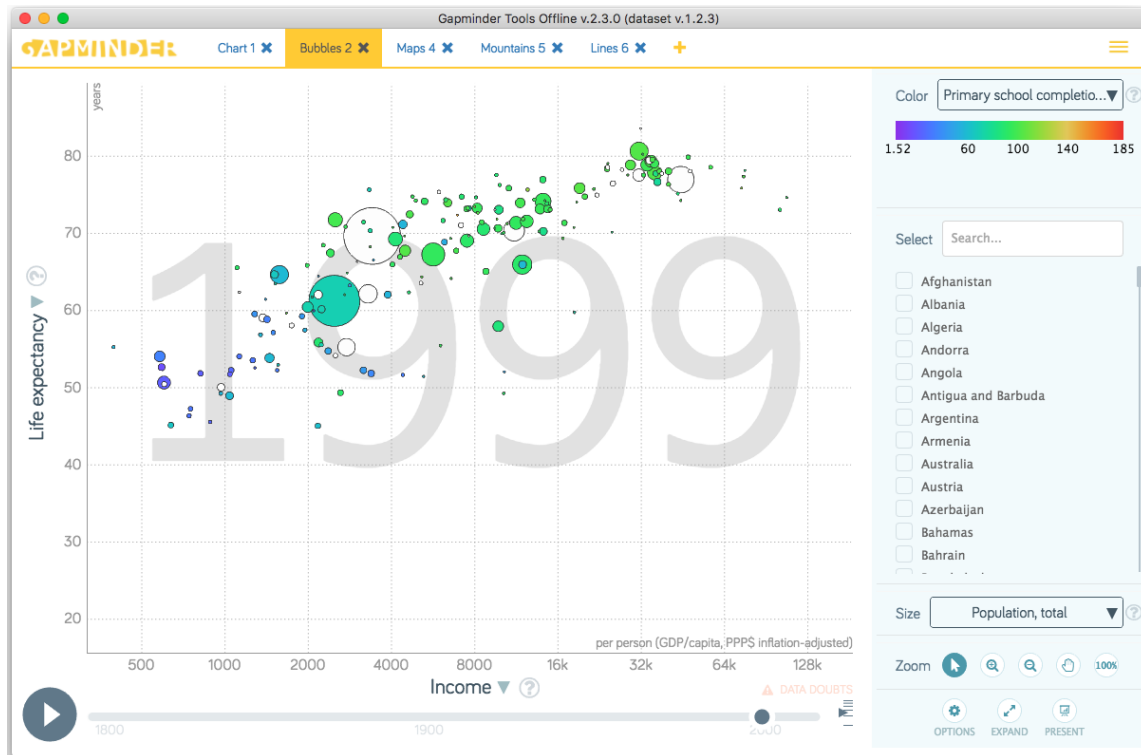


Figure 4.3: Use of color to encode a continuous attribute in a chart of type “bubbles” in Gapminder.

Gapminder allows to change the attribute encoded in color. Figure 4.3 shows an example where color is used to encode the percentage of children that complete the primary school. Gapminder automatically detects that the type of attribute is continuous and changes the colormap to a continuous rainbow colormap, ranging from violet to represent the minimum value to red that to represent the maximum. It is possible to choose between the linear or a logarithmic scale using the dedicated menu; by default the tool uses a linear scale.

This encoding is not optimal: rainbow colormaps have been proven to be confusing (due to the non-obvious ordering of colors) and misleading (since it creates bands of colors with almost constant hues and sharp transitions in between) [1]. On the other hand, the rainbow has an advantage: the banding effect implicitly defines groups of countries with a similar value for the indicator encoded in color. This allows the user to easily separate countries into clusters, where each cluster corresponds to a band.

In case the attribute mapped to color is missing, Gapminder shows a white bubble with a black border. This is a good choice, since white is a neutral color that does not interfere with the other colors.

TEXT LABELS Text labels are used to show the name of the indicator visualized on the axes, the measure units and the value for the ticks. The measure unit is shown above to the right for the x-axis. Below the axis there are the values for the ticks (about 10). Below them in the center there is the name of the indicator. The y-axis is similar. All labels are gray (a neutral color) not to attract the attention; this is a good design choice, since it allows the user to focus on the visualization.

Labels are used to visualize additional information about the bubbles. When the user selects a par-

ticular bubble (either by moving the mouse over it, clicking it or selecting the checkbox of the corresponding country in the right panel), a text label is shown to the top right of the bubble. The text label contains the name of the country, followed by the year.

Additionally, 2 dashed lines parallel to the axis are displayed when the mouse pointer is over a bubble. The lines start to from the bubble and intercept respectively the x and y-axis. A new text label with the value for the indicators of the country that corresponds to the bubble is shown at the intersection with the axis. This is a good design choice, since it allows the user to easily obtain additional information about interesting countries without polluting the visualization with the details about all the countries.

ANIMATION Animation is used to show encode time. Opposite to x-axis, y-axis, color and size, it is not possible to change the dimension of the dataset encoded with the animation.

The animation is started by pressing the “play” button in the bottom right corner of the windows. When the animation is running, bubbles change position, size and color according to the new value of the corresponding indicators. The slider on the bottom and the label with the year in the background change respectively position and value to reflect the time to which the indicators refers. The “play” button changes icon to “pause”: the user can stop the animation by pressing it again. A small control to the right of the slider allows to change the speed of the animation.

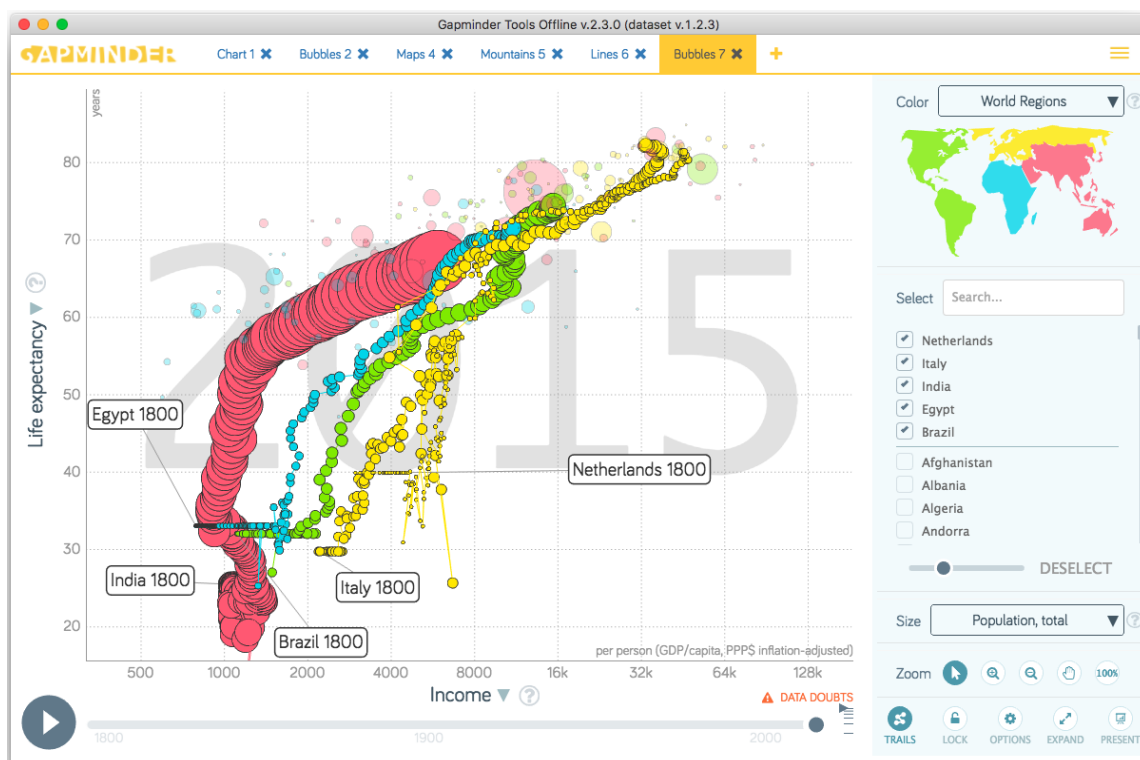


Figure 4.4: Animation in a chart of type “bubbles” in Gapminder.

The user can also select one or more countries before running the animation. In this case, Gapminder shows multiple bubbles for each selected countries, in particular one for each year the animation is played. Successive bubbles are connected by a line of the same color of the bubble (if they do not partially overlap already), defining a trail. Note that it is possible to disable the trails functionality using the button “Trails” in the right panel. The unselected countries are “hidden” by reducing their level of transparency (see the next paragraph).

This feature allows the user to visualize trends easily and effectively. One can first play the anima-

tion without selecting any particular country to can an idea about the general trend, and then select a group of countries of interest: Gapminder will visualize the evolution over time for the selected countries (Figure 4.4). This also makes easy to compare the evolution of the situation of some countries during time.

It is interesting to notice that trails are showed by default when selecting one or more countries and can be disabled if needed, but it is not possible to activate them if no countries is selected. This is a good design choice, since showing the trails for all countries in the world would likely create too much occlusion to give any real advantage.

Animations are very powerful, but present some drawbacks as well. In the example shown in Figure 4.4, the problem of overlapping described in the position paragraph is heightened. This is particularly true for bubbles with a big size (India in our example); an expert user can solve the problem by disabling the trails functionality or reduce it by changing the scale size for the bubbles using the “Options” menu: this allows to change the tradeoff between the occlusion and expressiveness added by the animation. Nevertheless, animations are really effective to visualize changes and trends in the indicators over time.

TRANSPARENCY Transparency does not encode any attribute of the dataset, but it is used to temporarily hide elements from the visualization. When the user selects some bubbles, the unselected ones are hidden by reducing their level of transparency. A slider is shown in the right panel and allows to change the transparency level.

This is a very good design choice, since it solves the problem of too many / overlapping bubbles and allows the user to focus on some particularly interesting country (or group of countries). The same argument is valid during animations.

4.2.2 MAPS

Gapminder’s “maps” visualize the value of an indicator as bubbles centered in each country on a simplified world map displayed in the background. Figure 4.5 shows a chart of type “maps” that visualizes the population size of the different countries. The visual variables used are position, size, color, text labels, animation and transparency. The encodings for size, text labels and transparency are the same as for the type “bubbles”, so we skip them here.

POSITION Position is used to encode the geographical location of a country: each bubble is centered in the country to which it refers. This allows to easily compare the situation of countries which are located closed to each other. The encoding is very clear, but it has the same problem of overlapping of bubbles presented in the Section 4.2.1: the problem is particular present for europe and africa, since there are many countries in a relatively small area.

COLOR By default, color is used to encode the geographical region of the countries, like for the “bubbles” chart. The encoding is clear, but it does not make much sense here, since the same information is encoded already in position. This overloading wastes an important visualization variable that can be used to represent another dimension of the dataset. Gapminder allows to change the dimension encoded in color using the corresponding menu. The same discussion about color in Section 4.2.1 applies here as well.

ANIMATION Animation is used to show encode time. Like for the “bubbles” chart, the animation is controlled using the “play” / “pause” button and the slider in the bottom. During the animation, the

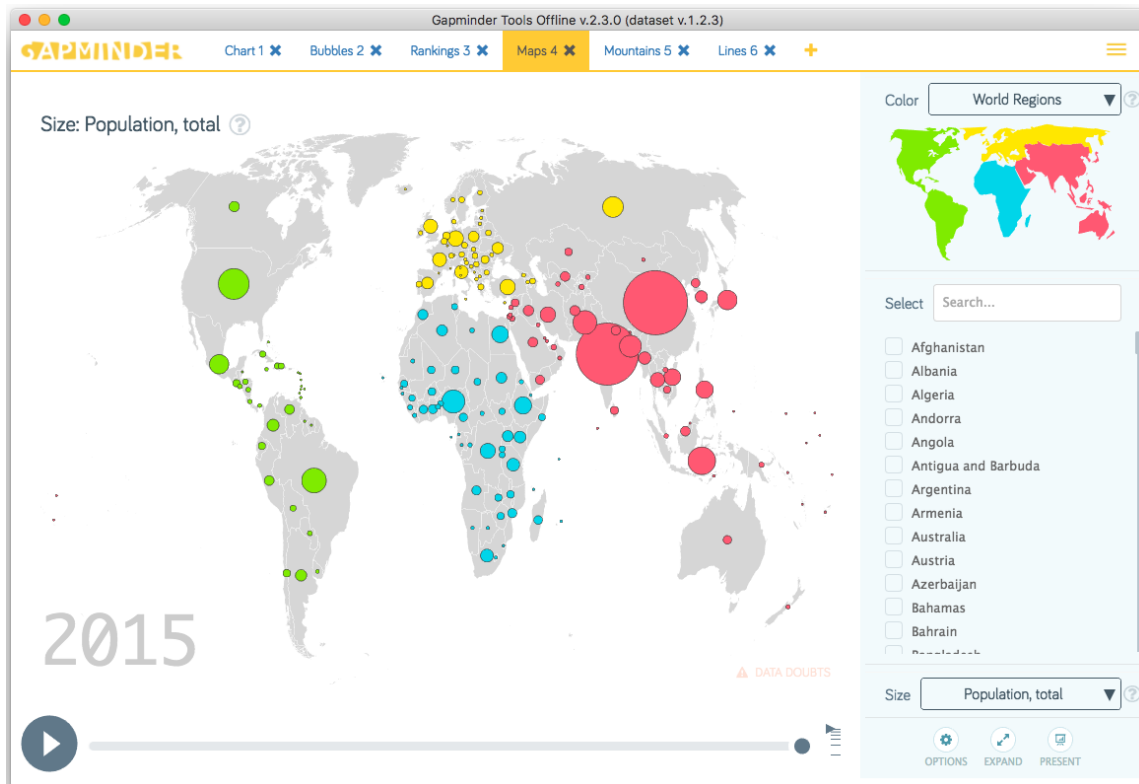


Figure 4.5: Chart of type “maps” showing the population size of the different countries in Gapminder.

bubbles only changes size and color, but not the position, since it encode the country (which is the independent variable). For the same reason, selected countries do not have any particular behaviour during the animation, as opposite to the “bubbles” chart.

4.2.3 MOUNTAINS

Gapminder’s “mountains” are area charts with animations. This chart shows the distribution of number of people by income (measured as \$ / day). Opposite to the other chart’s types, Gapminder does not allow to change the indicator displayed. Gapminder allows to choose to stack all areas together, stack them by color (this option is enables only if the color mapping is categorical) or not stack them at all. Figure 4.6 shows a screenshot of the “mountains” chart with stacked areas that shows the distribution of number of people by income.

Each area display a continuous distribution of number of people by income, but the underling data used by Gapminder is inherently discrete (see Section 3.2.1). Gapminder probably combines the information about the population size and the average income per people to compute the area for each country, then it displays it as a Gaussian distribution centered on the average income per person. It is not clear how this is done: for example, which value is used as standard deviation to define the Gaussian distributions? Another clear problem is the skewness: it is unlikely that all countries have exactly a symmetric distribution of number of people by income. This representation may be misleading when single countries are examined in isolation. The problem is less evident if the areas are stacked together.

The visual variables used are position, size, color, text labels, animations and transparency.

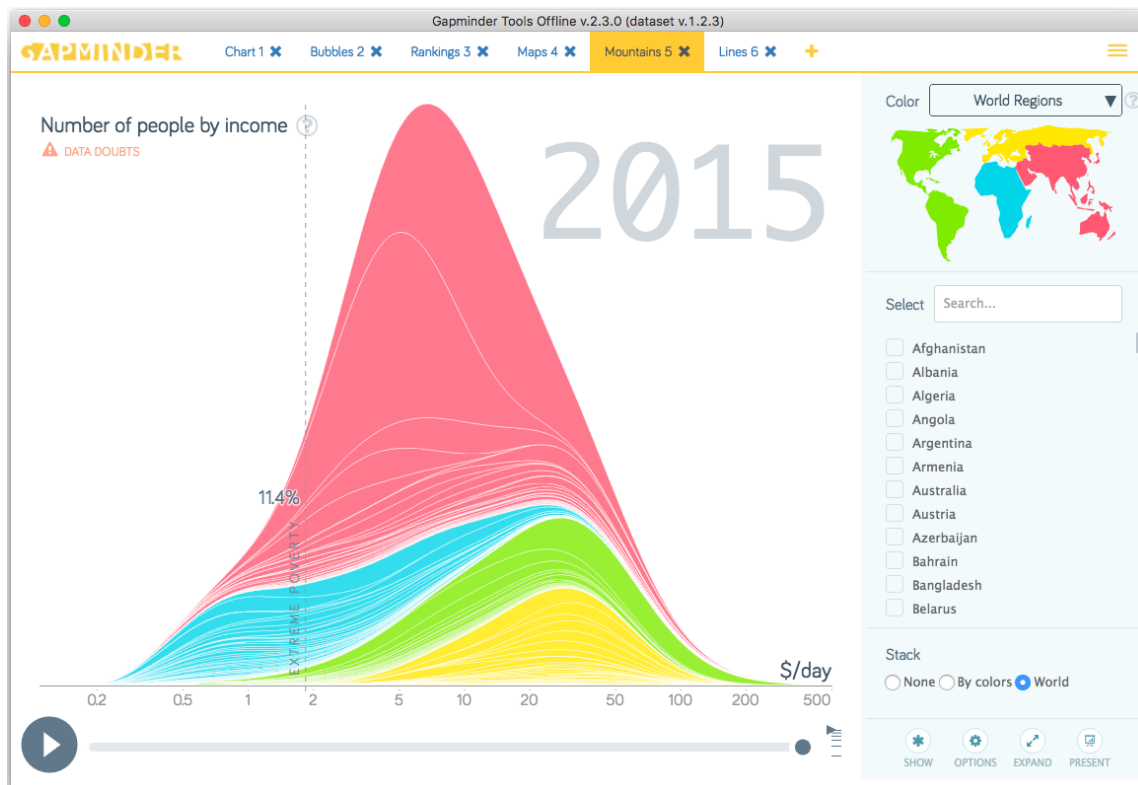


Figure 4.6: Chart of type “mountains” with stacked areas in Gapminder.

POSITION Along the x-axis, position represents the daily income of the people (on a logarithmic scale). Along the y-axis, position assumes different meanings depending if the areas are stacked or not. If the areas are not stacked, the position along the y-axis indicates the number of people with some given income (determined by the position along the x-axis). If the areas are stacked, the meaning is more complicated: it is still related to the number of people by income, but the value is measured starting from the point more in the bottom in the area for the fixed x-coordinate. In other words, the meaning of each point in the stacked version depends on the shape of the area where it is contained.

Assuming basic knowledge of logarithms, the encoding along the x-axis is clear and easy to invert: the user can simply use the ticks and labels along the x-axis to figure out the daily income value represented by a point. However, Gapminder does not explicitly state that the scale is logarithmic, which may be confusing for the general public.

For the non-stacked version, the mapping along the y-axis is also quite clear: the distance from the x-axis is proportional to the number of people with some income. This allows the user to relatively easily compare the number of people with a certain income in different countries and compute approximate ratios. For example, one could say that the number of people with an income between 5 and 10 \$/day in China and India in 2015 is about the same. However, it is difficult (or almost impossible) to read the absolute values, i.e. how many people have a certain income in one country, since there is no explicit scale for the y-axis.

For the stacked version, the encoding along the less y-axis intuitive, but allows to easily compute aggregations: the distance from the x-axis of each point represents the sum of people with the given income in all countries stacked together. From Figure 4.6 it is possible to say that most people in the world have an income between 5 and 10 \$/day in 2015.

The non-stacked version has overlapping problems: countries with similar income's distributions

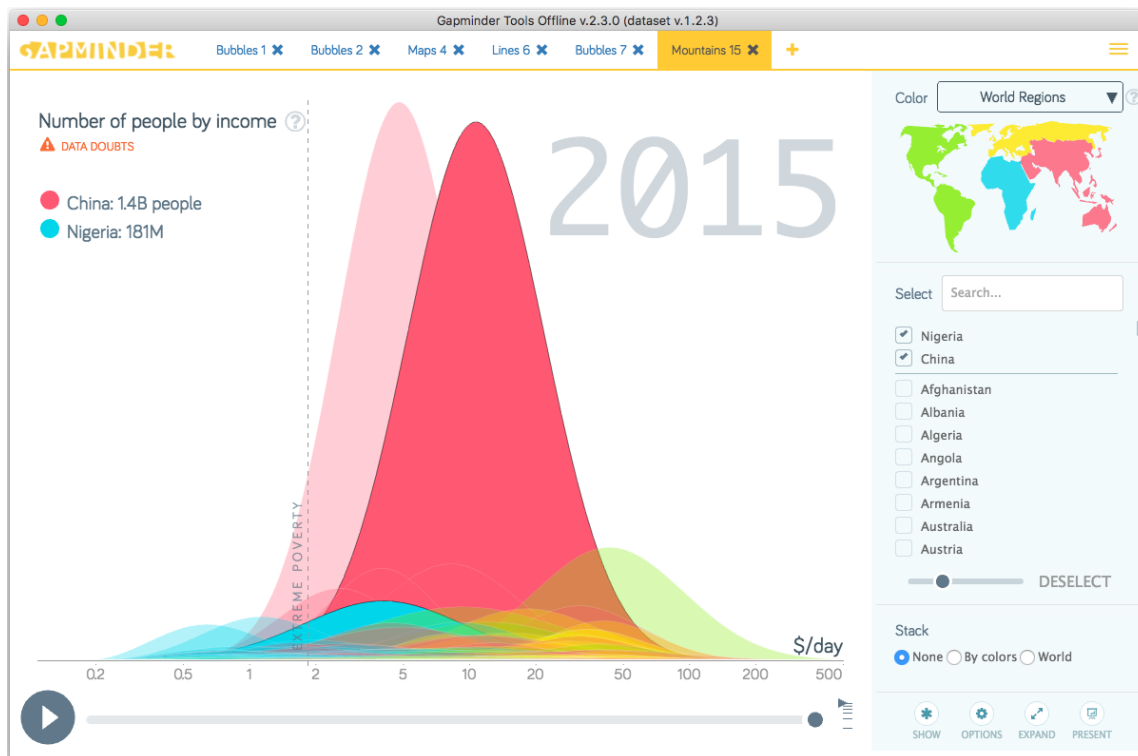


Figure 4.7: Chart of type “mountains” with non-stacked areas in Gapminder.

result in overlapping areas with a similar shape. The problem is only partially solved by allowing the user to select only some countries (using the menu in the right panel). Figure 4.7 shows an example of non-stacked areas in Gapminder: we have selected China and Nigeria to highlight them, but there are still clear overlapping problems, especially for Nigeria.

SIZE The size of an area encodes the size of the population of the corresponding country, both in the stacked and non-stacked versions. The encoding allows to approximately compare the sizes of population of different countries. However, this encoding is not clear for 2 reasons:

- The scale logarithmic along the x-axis deforms the areas.
- There is no legend for the size of the areas.

COLOR The encoding of color is the same of the one for “bubbles” charts discussed in Section 4.2.1. This encoding has a problem in the non-stacked areas version: when some countries are selected, the others are partially hidden by changing their level of transparency. Overlapping areas with different colors and a non-zero level of transparency create new colors which does not belong to the colormap. This problem is evident with categorical color mapping: in Figure 4.7, the overlapping of red area that represents China with the big green area on the right generates a big orange area.

TEXT LABELS Labels are used for different scopes:

- The title on the top-left part of the window gives a description of the visualization.
- The big text label with the year in the top-right part of the window display the year in which the population income is measured.
- The numbers below the x-axis indicate the daily income for points with the corresponding x-

coordinate. This is fundamental for the inverse mapping from the position along the x-axis to the absolute value of daily income.

- A label above the x-axis indicates the measure unit used for the x-axis.
- When the user selects some countries, some text labels are displayed under the title (see Figure 4.7): they provide a color legend for the selected countries and indicate the population size.
- Finally, labels are displayed when the user moves the mouse on a particular area to indicate the name of the corresponding country.

All labels use neutral color (white, gray or black) not to attract the attention. This is a good design choice, since it allows the user to focus on the visualization.

ANIMATION Animation is used to show encode time. Like for the “bubbles” chart, the animation is controlled using the “play” / “pause” button and the slider in the bottom. During the animation, position, size, color and text labels are updated to reflect the data of the visualized year.

Animation allows the user to understand how the distribution of incomes in different countries has changed in time. If the user is interested in particular countries, it can select and highlight them.

If no country is selected during the animation for the stacked version, countries in the same region are combined together. This allows the user to understand the changes and trends for different regions in the world. However, this encoding gets really difficult to understand if the user changes the attribute encoded in color. Figure 4.8 shows an example where color encodes the average income per person: the user has no way to understand the meaning of different areas, since there are neither a legend nor labels to indicate it.

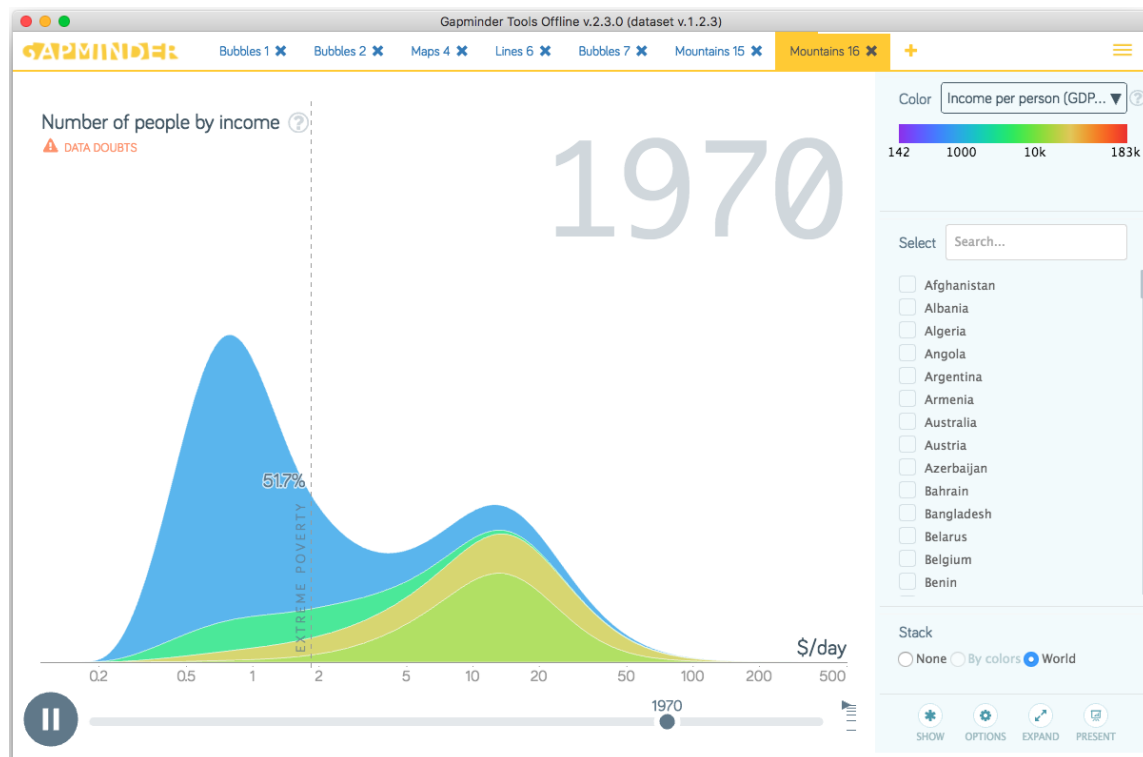


Figure 4.8: Chart of type “mountains” with stacked areas in Gapminder. The screenshot is taken during an animation.

TRANSPARENCY The encoding of color is the same as the “bubbles” charts discussed in Section 4.2.1. We have discussed the problems of combining transparency and color in the color paragraph.

4.2.4 RANKINGS

Gapminder’s “rankings” are bar charts with animations. Each bar corresponds to a country. The bar charts is oriented horizontally. The visual variables used are position, size, color, text labels, animations and transparency. The encodings for color and transparency are the same as for the “bubbles”, so we skip them here.

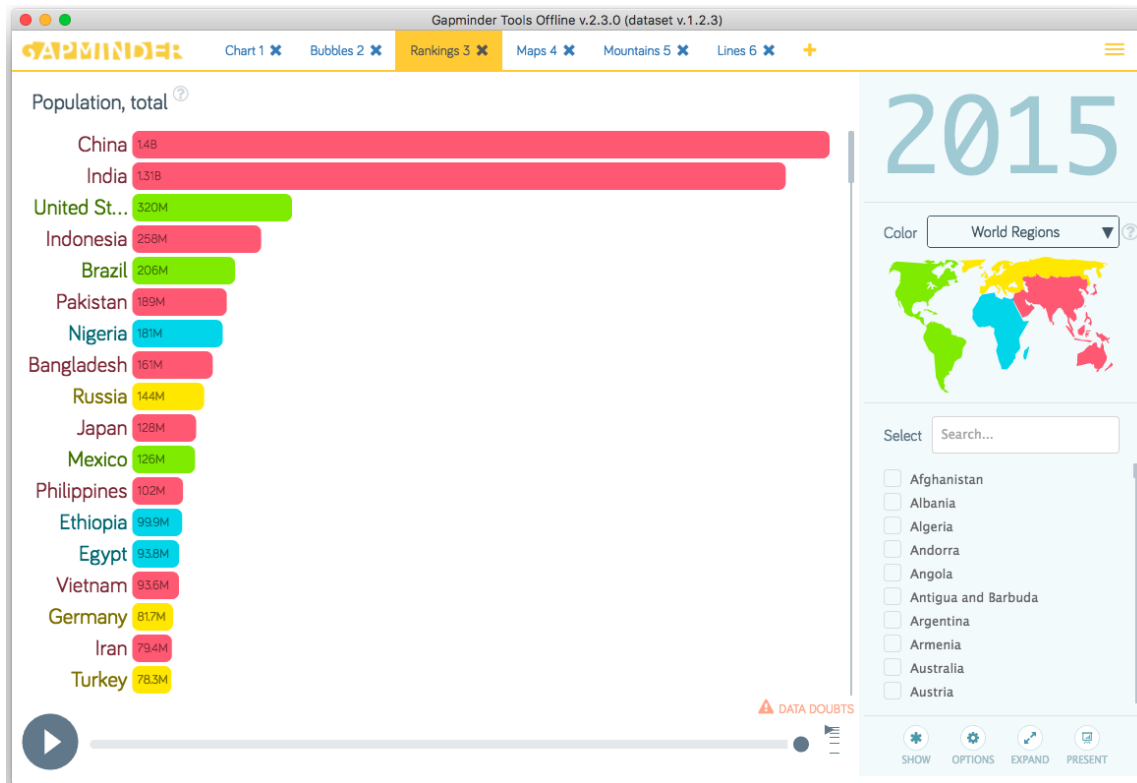


Figure 4.9: Chart of type “rankings” in Gapminder.

POSITION Position is used to encode the ranking of countries by the value of the indicator encoded in size. Countries are ordered for decreasing values: countries with higher values are located at the top, the ones with lower values at the bottom. It is not possible to change the ordering, i.e. put the countries with lower values at the top. Countries without data for the indicator are put to the bottom.

The encoding allows the user to easily find the top countries and the ones with the smaller values for the displayed indicator, although this usually requires to scroll down with the mouse (since a normal size screen is not long enough to display all bars). It also allows to compare pairs of countries in the ranking for the indicator. In this sense, the encoding is good.

SIZE Size is used to encode an indicator, such as the size of the population. The length of each bar is linearly proportional to the value of the indicator. It is possible to change the scale to logarithmic by clicking on the title of the chart, above the first bar.

The user is able to:

- Order countries by the value of the indicator.

- Compute ratios between the value of the indicator for different countries. In the example of Figure 4.9, the user can for example estimate the population of China to 3 to 4 times bigger than the one of the United States.

The encoding is good for ordering and computing rates. However, it is not easy to extract the absolute value of a bar, since there is no length legend. This information is only encoded in text labels inside each bar, which are not always easy to read.

TEXT LABELS Text labels does not encode directly any dimension, but are fundamental for different aims:

- The title of the chart indicates which indicator is encoded in the length of the bar.
- The labels before each bar indicates which country the bar refers to.
- The labels inside each bar report the value of the indicator measured in the corresponding country.
- A big text label in the top part of the right panel indicates in which year are measured the displayed data.

The labels with the country names and those inside the bars have the same color hue of bar to which they refer (with a slightly higher saturation). On the one hand, this makes easier to associate them with the correct bar. On the other hand, it is particularly difficult to read the labels inside red and blue bars due to the low contrast.

ANIMATION Animation is used to show time. During the animation, position, size, label and labels of the bars are updated to reflect the displayed year. The user can select one or more countries to highlight them: in this case, Gapminder will also automatically scroll the view if the selected bars exits the current one.

The animation allows to track the evolution of the selected countries over time. However, it does not allow to have a visualize the global trend, since there are too many variations (also, not all bars fit in the size of a the monitor). It is also difficult to compare 2 or more countries, since small changes in the measured value can cause significant changes in the ranking and / or a swap in their relative order: tracking all these variations over successive visualizations is hard. For this kind of task, the chart of type “bubbles” should be preferred.

4.2.5 LINES

Gapminder’s “lines” are multivariable line charts with animations. They show the variation of one indicator over time: the x-axis represents time (in years), while the y-axis a given indicator. Figure 4.10 shows a Gapminder “lines” chart visualizing the average income of different countries over time. By default, only some countries are shown; the user can change the selected countries using the menu in the right panel.

The visual variables used are position, color, text labels and animation. The encoding for the color is the same as for the “bubbles”, so we omit it here.

POSITION Position is used to encode the pair time (along the x-axis) and the selected indicator, for example income per person in Figure 4.10. The encoding is very clear and effective: each point in the chart has a one-to-one mapping with a the pair time and income. The encoding is easy to reverse thanks to the dashed gray lines in the background, the ticks and the labels on the axes.

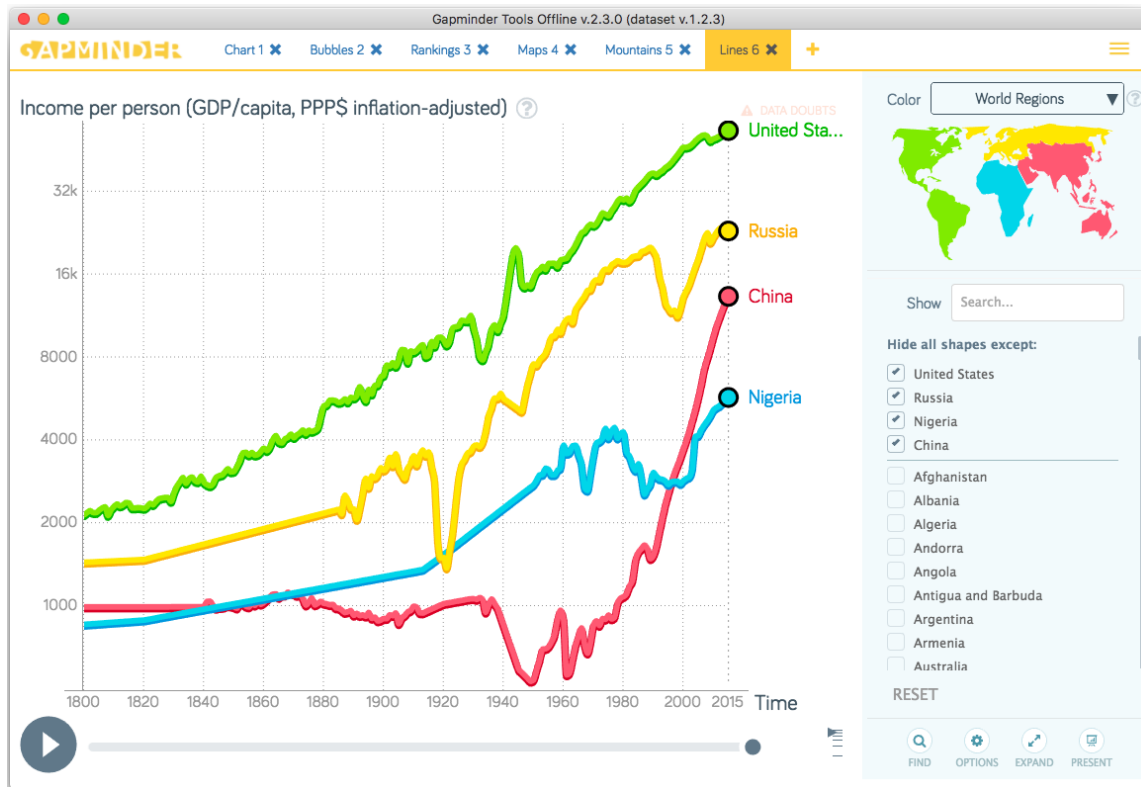


Figure 4.10: Chart of type “lines” in Gapminder.

TEXT LABELS Text labels are used for:

- Describe the dimensions and measure units represented by the axes.
- Provided a legend for the plotted line (each line has a label with the name of the country closed to the end of the line).

The use of labels is fundamental for the inverse mapping of position and color.

The labels with the name of countries are of the same color of the line that corresponds to them. This is a good design choice since it uses color to make the association line-label easier to read.

ANIMATION Animation encodes time. In this type of chart, time is already encoded in the x-axis. The animation does not show any new information, but simply make the drawing of the line interactive: when the user starts the animation, lines are removed; at each step of the animation, lines are drawn from the first year to the current year in the animation. Figure 4.11 shows an example of line chart when the animation reaches year 1935.

The animation does not encode any additional information, so it is not really useful to display the data. On the other hand, the animation does not have negative effects on the visualization, since the user can simply not use it. We believe that the animation was added here to maintain a design similar to the other types of charts.

TRANSPARENCY Transparency is not used in this type of chart: only the selected countries are visualized, the others are not drawn. This is a good design choice, since it effectively solves the problem of occlusion typical of multivariable line charts. By default, only 4 countries are selected, so that the occlusion is minimum.



Figure 4.11: Chart of type “lines” in Gapminder. The screenshot is taken during an animation.

4.3 IMPROVEMENTS

Most problems found in the visualizations are common to all types of charts available in Gapminder. We will first discuss the common ones, then we will focus on specific problems.

COLOR The default dimension encoded in color in Gapminder is the region of the world. There are 4 regions and Gapminder uses 4 very different colors: blue, green, yellow and red. This choice is perfect for normal people. However, about 10% of the male population is affected by color blindness [1] and have difficulties to distinguish red from green. We suggest to include an option to change the colormap to a color blindness friendly one. The tool ColorBrewer⁷ provides many choices⁸ for a blindness friendly categorical colormap.

A similar problem is the choice of the rainbow colormap for continuous data attributes. The rainbow colormap has been proven to be confusing and misleading [1] in many papers in the literature. We suggest to use a two-hue as the default for continuous values, since it allows a natural ordering of colors and it is thus easier to invert (heat colormap would also work, but would create ambiguities in case of missing values). The rainbow colormap has the only advantage to artificially define groups because of the banding effects: this may actually be desired for some visualizations, if the user is conscious of it. There should be an option to switch between the two-hue and the rainbow colormap. At best, every color map should also have a short description of its use cases (i.e. for which task it is better than the other).

TEXT LABELS In all chart types, Gapminder shows a label with the name of the country when the user moves the mouse on the corresponding bubble, area, bar or line. At the same time, the values

⁷<http://colorbrewer2.org>

⁸<http://colorbrewer2.org/#type=diverging&scheme=BrBG&n=4>

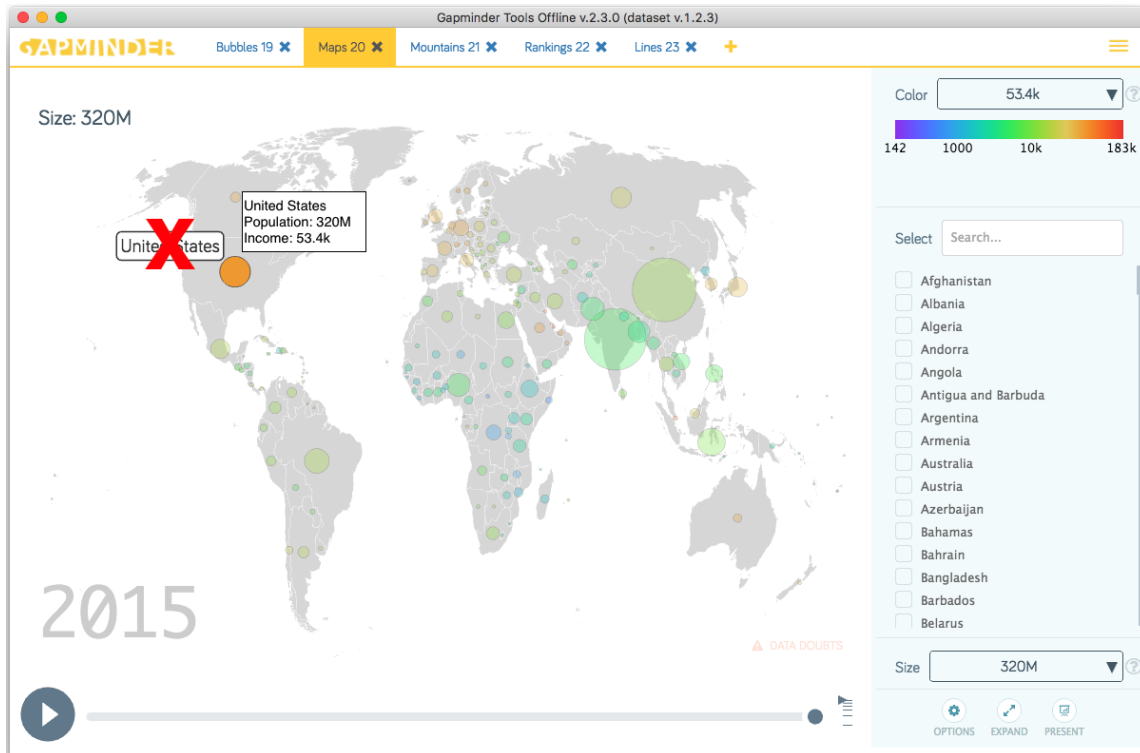


Figure 4.12: Proposal of improvement in the use of labels in a chart of type “map” in Gapminder. We propose to replace the existing label (marked with the red “x”) with a label with additional information in addition to the country name.

of attribute encoded in color in showed close to the colormap, the value of the attribute encoded in size is shown in the title of the visualization etc. We propose to enrich the existing labels with information about the attributes for that country. Figure 4.12 shows a sketch for this improvement: we added information about the population size (encoded in size) and average income (encoded in color).

BUBBLES’ SIZE In the chart of types “bubbles” and “map”, countries are represented as bubbles and the size is used to encode a given dimension of the dataset. The encoding is good since it allows to easily compute orderings, ranking and ratios, but it misses a legend: the user can not invert the mapping and read the absolute values of the display data. We propose a trivial solution to solve the problem: add a legend for the size, something similar to the one shown in Figure 4.13.

BUBBLES’ OVERLAPPING A way to partially solve the problem of overlapping bubbles described in Section 4.2.1 is to use a shaded profile: the center of each bubble has a higher luminance, which decreases with the distance from the center. When bubbles of the same color overlap, the shading helps seeing the overlaps better and thus distinguishing different bubbles, without taking any extra space. This technique is inspired to cushion treemaps, where shading is used to convey the hierarchy organization of rectangles closed to each other.

MOUNTAINS’ COLOR & TRANSPARENCY The combination of overlapping areas, use of color and transparency generates new colors in the charts of type “mountains”. This problem is evident when color encodes a categorical attribute, for example in Figure 4.7. We believe that the main cause of problems is the use of transparency to highlight the selected countries (of better, to hide the unse-

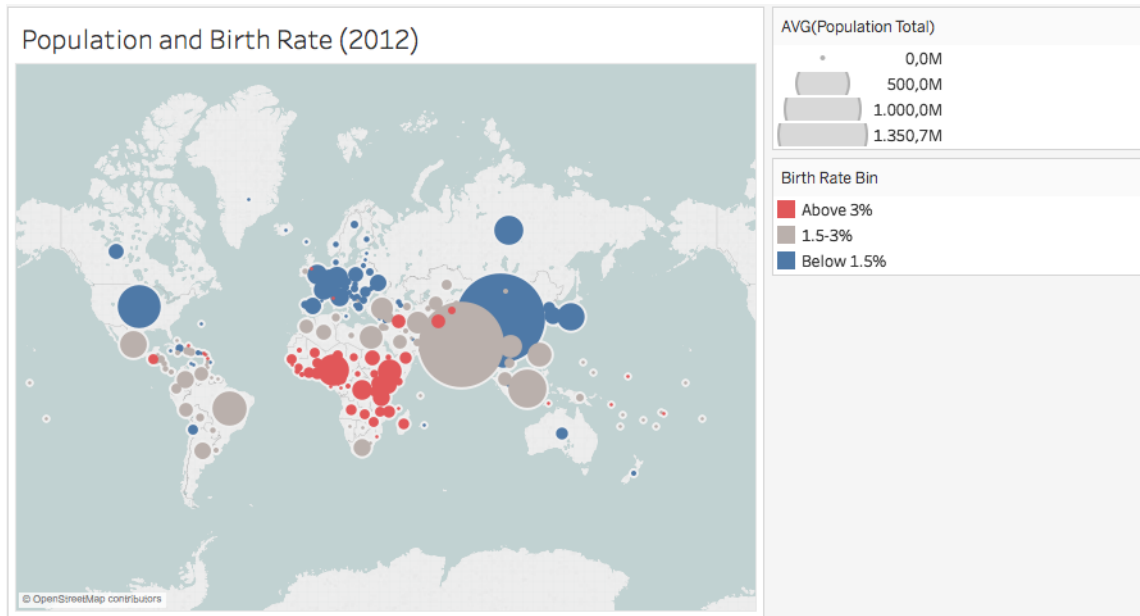


Figure 4.13: Example of size legend in a visualization in Tableau. The panel on the right show the legend for the size of the bubbles used in the visualization. The example is taken from the “World Indicators” sample workbook in Tableau Desktop 10.4.

lected ones). We propose to use the same approach as the charts of type “lines”, where the unselected countries are totally hidden. In this way we can avoid the use of transparency and avoid the creation of addition colors not defined in the colormap.

MOUNTAINS’ Y-AXIS In Section 4.2.3 we have described the problem of performing the reverse mapping for the y-axis (number of people) in the charts of type “mountains”. To solve the problem, we propose to draw the y-axis and add some ticks and labels to indicate the corresponding values. In other words, we propose to create a legend for the y-axis. Figure 4.14 shows a sketch of this idea.

RANKINGS’ LABELS The charts of type “rankings” have labels to indicate the value encoded in the bars’ length (Section 4.2.4). The labels are needed for the reverse mapping from the length of the bars to the value of the attribute encoded. Unfortunately, the labels use a color very similar to the one of the bars, which makes them difficult to read. We suggest to choose colors that create a bigger contrast, in order to make the labels easy to read. In addition to this, the x-axis could be drawn explicitly (like the y-axis in the “mountains” chart, see the previous paragraph).

5 DESCRIPTION OF FINDINGS

Gapminder allows to answer many questions about the global world situation and its evolution over time. We will try to give an answer to some of them:

- What is the life expectation of the world population?
- What is the global trend of the average income?
- What is the world situation in terms of population, average income, life expectation?

The other questions listed in Section 2 can be answered using visualizations similar to the ones described in this section. Most questions are actually contained in the more general one: *What is the*

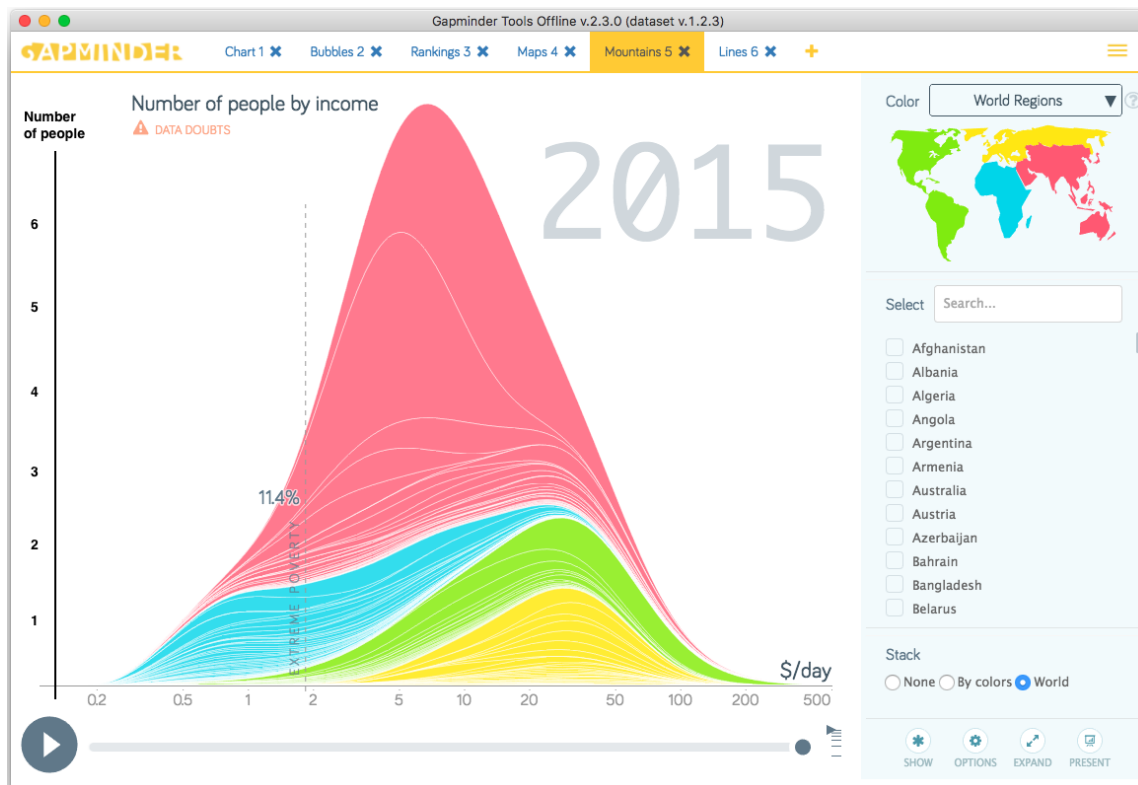


Figure 4.14: Sketch for an improvement in the charts of type “mountains” in Gapminder. We propose add a legend for the y-axis: we drawn the y-axis and put some text labels to make the scale explicit. Please note that the number used for the labels do not represent the data but have the only purpose to present the concept.

world situation in terms of population, average income, life expectancy?.

5.1 LIFE EXPECTANCY

We can answer this question by using a chart of type “maps” and changing both the indicator shown as color and size to life expectancy. We also reduce the size scale to reduce the overlapping of bubbles. Each bubble represents a country and the position of the bubble its geographical location. The resulting visualization is shown in Figure 5.1. We use the most recent available data of 2015.

We can immediately notice groups of bubbles with different colors.

- Bubbles in Europe are red or dark orange, which indicates an average life expectancy between 75 (orange) and 85 (red) years. Countries in east Europe tend to have a slightly lower life expectancy than countries in the center and west.
- Countries in Asia tend to have shorter life expectancy, between 65 and 75 years. There are however some exceptions: Japan, South Korea and Singapore have higher life expectations (respectively 83, 81 and 83 years), while Afghanistan has a much shorter one, only 54 years.
- New Zealand and Australia have high life expectations, while the small countries in the Pacific Ocean have quite much lower ones (between 60 and 65).
- Bubbles in Africa range from green to yellow / light orange. This indicates life expectations between 50 and 65 years. The situation is slightly better in the very north: life expectancy here is between 70 and 75 years.

- Countries in America go from orange to red. Life expectation is quite high in the north (about 80 years) and in the south (about 75 years) and lower in the center (from 70 to 75 years).



Figure 5.1: Life expectancy in 2005 for different countries of the world.

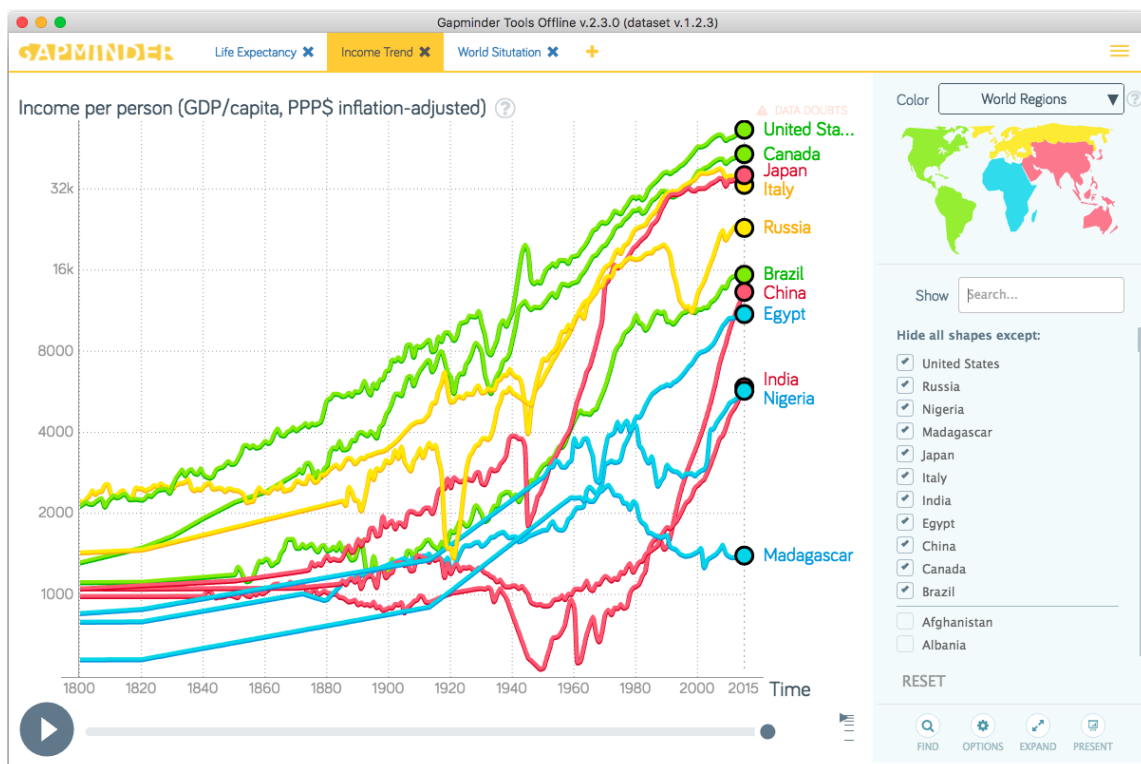


Figure 5.2: Trend of the average income per person for different countries of the world.

5.2 INCOME TREND

We can answer this question by using a chart of type “lines”. The default visualization already shows the income per person for countries in different regions, we only add some countries to the visualized one to have a better overview of the global situation. The visualization is shown in Figure 5.2.

In general, income tends to grow in all regions in the world. Countries in Europe, America and rich countries in Asia had an almost constant growth from 1850 to 2015, with only some small drops around 1920 and 1945 (probably due to the World Wars). Countries in Asia such as China and India did not have a significant growth until the second half of the 20th century: after that moment, the growth is very fast, much faster than Europe or America. Countries in africa show different situations: some of them are growing (eg. Egypt or Nigeria), others had a growth in the first half of the 20th century, followed by a decline. The situation is quite complex and would require some more complex analysis which is out of the scope of this document.

5.3 WORLD SITUATION

We can answer this question by using a chart of type “bubbles”. We use the default visualization offered by the tool, since it already presents income, life expectancy and population size. The visualization is shown in Figure 5.3.

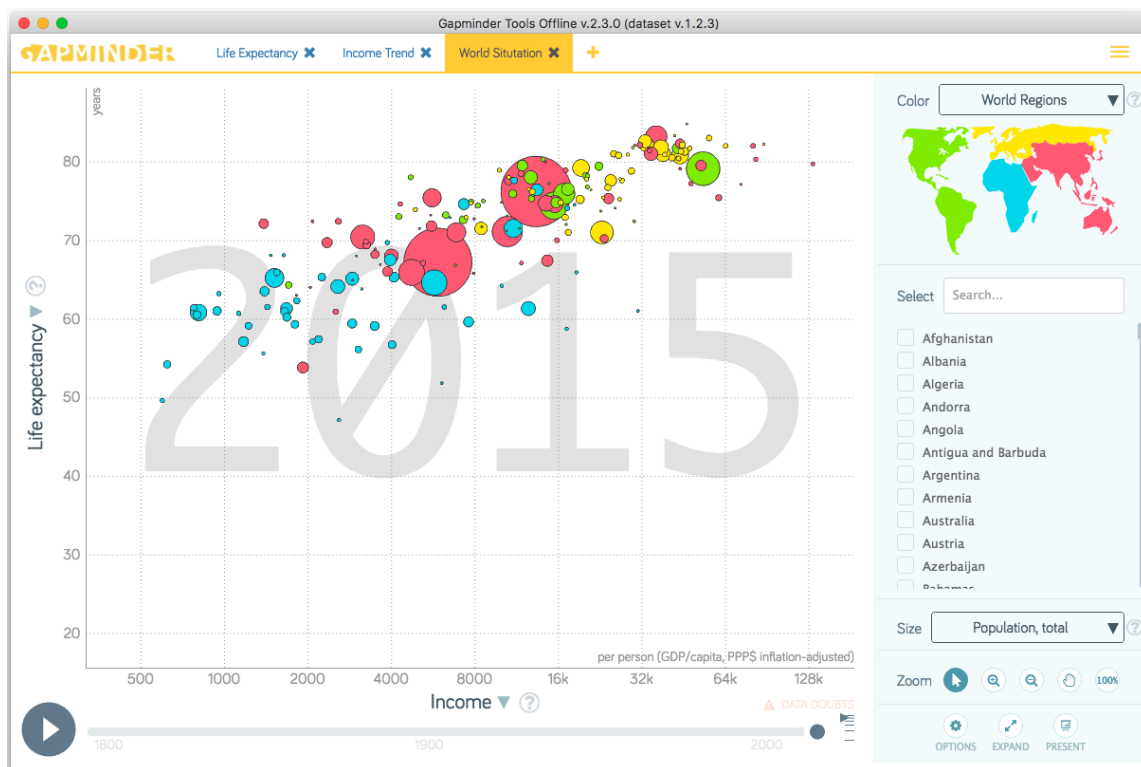


Figure 5.3: World situation from the point of view of income, life expectation and population size.

Overall, we can notice interesting patterns.

- Countries tend to be aligned on one line: this indicates a strong correlation between income and life expectation. In general, countries with higher incomes tend to have longer life expectations.
- Yellow and green bubbles are concentrated in the right part, while the blue ones are on the left. Countries in Europe and America and rich countries in Asia (Japan, South Korea etc.) have

both high incomes and life expectations, while countries in Africa have lower incomes and life expectancy. The situation of Asia is somewhere in between, since most of the bubbles are in the center of the chart.

- Most of the world population is concentrated in Asia, in particular in China and India. European countries tend to have small populations, while countries in America and Africa have slightly larger ones.

6 CONCLUSIONS

Gapminder allows to analyze a wide variety of indicators about the situation of all countries in the world over the last 2 centuries. The tool is though for the large public and does not require any particular knowledge about the showed data or statistic. The visualizations are quite simple and use animations to show changes of the indicators over time. Overall, the tool allows to give general answers to a variety of questions, from simple ones that involve only few countries, to more complex ones about the global world situation. The tool is suited to both confirm the known, i.e. use the visualizations to confirm some hypothesis, or to discover the unknown, letting the user to explore a big variety of different indicators.

REFERENCES

- [1] D. Borland and R. M. T. Ii, "Rainbow Color Map (Still) Considered Harmful," *IEEE Computer Graphics and Applications*, vol. 27, no. 2, pp. 14–17, March 2007.