

Chapter 8

Seeing Dynamic Data Visualizations in Action: Gapminder Tools



Peter Kovacs , Klara Kazar, and Eva Kuruczleki

Abstract Citizens need sophisticated ways of thinking in order to understand complex, real social and economic phenomena and to interpret relationships among social and business data correctly. Huge amounts of data, data sources and visualization tools provide an opportunity to illustrate complex relations within real data. At the same time, the misuse of these tools can lead to misinterpretations and in turn perhaps to poor decision making. The Gapminder Foundation has developed several visualization tools (using bubble charts, trend lines, ranks, maps, etc.) in order to facilitate the exploration of complex relationships between variables, to rank countries, and to explore things such as time series, age distributions and income distributions. The goals of this chapter are to: show these different tools in action in classes exploring poverty and the ageing society; to describe their advantages and disadvantages; and to compare.

Keywords Gapminder · Ageing Society · Poverty · Gapminder offline · Nonlinear relationship · Population pyramids

8.1 Introduction

When we analyze civic, social, or economic data, data visualization has a unique role to play in exploring relationships and trends. With the help of visualization, we can tell a story, teach about, and demonstrate complex phenomena relatively simply. Elaboration of civic problems contributes to an understanding of these real problems and can improve civic engagement whether in statistical or non-statistical classes. There is a wealth of online and offline databases, tools and apps which can support the analysis and visualization of real data, that can be used alone, or in combination with others (for examples, see Chaps. 5 and 6). Gapminder is a free, online

P. Kovacs (✉) · K. Kazar · E. Kuruczleki
University of Szeged, Szeged, Hungary
e-mail: kovacs.peter@eco.u-szeged.hu; kazar.klara@eco.u-szeged.hu;
kuruczleki.eva@eco.u-szeged.hu

visualization package. After a short summary of Gapminder tools, this chapter shows Gapminder in action, embedded in lesson plans developed by the ProCivicStat project. The first two applications use Gapminder's Bubble chart; in the third application we used Gapminder tools (to show trends, ranks, age profiles) in combination with other providers' population pyramids.

8.2 About Gapminder

The Gapminder Foundation was established in Sweden by Hans Rosling, Ola Rosling and Anna Rosling Rönnlund in 2005. Gapminder's mission is "Fighting devastating ignorance with fact-based worldviews everyone can understand". They identify misconceptions about facts, compare what people believe to what the real data show and develop course materials, visualization tools, and courses, which are available freely. Their famous book is *Factfulness*, which was published in 2018. The Gapminder webpage (www.gapminder.org) offers a collection of tools for visualizing data. As Hans Rosling said at ICOTS8 in 2010, "by serendipity Gapminder Foundation found that the beauty of a statistical database can be unveiled to a broad audience if the data set can be animated in newly developed software." On the Gapminder YouTube channel,¹ several videos are available on various topics, which can be useful in classes to illustrate a social issue. For instance, Don't Panic, The truth about population with Hans Rosling, which also was a 1-h length TV program on BBC, presents Rosling and the Gapminder philosophy.

Gapminder has a menu of over 500 indicators that can be explored via online visualization tools. From the website, all of the data are downloadable in csv or xlsx formats; metadata is provided, along with links to the original data source and provider.

There are several visualization tools at Gapminder website: Bubbles, Income Mountains, Maps, Ranks, Trends, Ages, and Dollar Street:

- The best-known tool is the **Bubble Chart**, which is a dynamic scatter plot. We can visualize five variables at the same time with the help of the two axes, the color and the size of bubbles, and a timeline. The axis can be changed to show linear or logarithmic scales. This chart is described in detail below in Sect. 8.3.
- Gapminder **Trends** is a tool to present simple line charts, used to visualize time series for the countries of interest, on the variable chosen.
- On Gapminder **Maps** sizes of bubbles represent values of a selected indicator at the country level on a map. Time is on the x-axis, and changes over time can be animated.
- Gapminder **Ranks** is a tool to prepare a horizontal bar chart if we want to show the order of countries by a selected variable. We can see not only the ranks of the

¹<https://www.youtube.com/c/gapminder/featured>

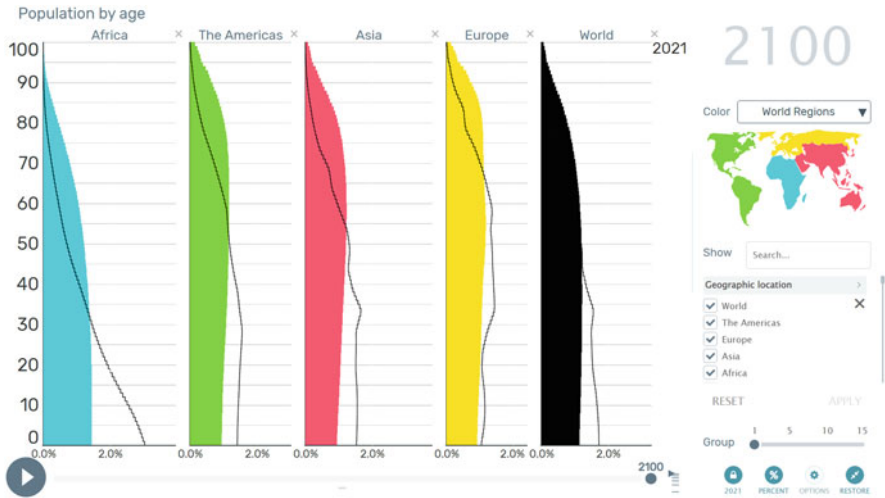


Fig. 8.1 Gapminder ages: populations’ age distribution in different world regions, in 2020, represented by the black line, and projected to 2100 (colored graph). Showing a free visualization from Gapminder.org, CC-by license

countries but also the absolute value of their data, represented by the length of the relevant stack. This tool also contains a timeline, which gives an opportunity to compare rank orders between different years.

- **Income Mountain** shows the distribution of people by income in different countries, regions, and at the world level. With the help of the timeline, changes in the distributions can be animated.
- **Ages** is illustrated in Fig. 8.1. It can show the age-distribution of selected countries, or groups in different years, which gives the opportunity to compare distributions of groups, or temporal changes of the age-structures. A forecast for the future until 2100 is built into this tool. The horizontal axis can present either raw frequencies or relative frequencies of the different age groups. It is possible to lock a pyramid of a year, which is represented by a black line as a reference line, and compare the different years’ age distributions to the reference year.
- **Dollar Street** is a rather different sort of visualization. It is based on a large collection of photographs in different countries of the domestic circumstances of families at different levels of wealth. Users can see the tools used for cleaning teeth, or sleeping accommodation, or toilet facilities for people at different income levels around the world. Hence, the pictures are not exact statistical data, but rather are illustrations of living conditions worldwide at different wealth levels. The measurement of wealth levels is available in FAQ on the Dollar Street website. This tool is based on a non-representative sample of images but can be used to stimulate student discussions about poverty, relative poverty, and the transitions most likely to move people out of poverty.

There is also a facility for using Gapminder offline. This enables not only the use of Gapminder without the Internet, but it also offers the opportunity to use Bubble Charts, Lines and Ranks with our own data. It is possible to import data in .xls and .

Step 2: Choose how your data is arranged:

Time goes down

Column 1: entities, Column 2: time values, Column 3 and on: indicators (see example)

	A	B	C	D	E	F	G	H
1	Country	Year	Region	Population	Child mortality	Child Survival	GDP per capita	Gini
2	Afghanistan	1800	Asia	3280000	468.58	531.42	603	30.51
3	Albania	1800	Europe	410445	375.2	624.8	667	41.24
4	Algeria	1800	Africa	2503218	460.21	539.79	716	57.71
5	Angola	1800	Africa	1567028	485.68	514.32	618	48.93
6	Antigua and Barbudi	1800	America	37000	473.6	526.4	757	40
7	Argentina	1800	America	534000	402.19	597.81	1507	47.1
8	Armenia	1800	Europe	413326	371.2	628.8	514	29.12
9	Australia	1800	Asia	351014	390.99	609.01	815	41.31
10	Austria	1800	Europe	3205587	387.32	612.68	1848	53.35
11	Azerbaijan	1800	Europe	879960	384	616	775	40
12	Bahamas	1800	America	27350	471.04	528.96	1445	40
13	Bahrain	1800	Asia	64474	439.79	560.21	1235	40
14	Bangladesh	1800	Asia	19227358	507.88	492.12	876	41.83
15	Barbados	1800	America	81729	469.76	530.24	913	40
16	Belarus	1800	Europe	2355081	365.67	634.33	608	57.85
17	Belgium	1800	Europe	3138137	322.4	677.6	2412	62.44
18	Belize	1800	America	25526	467.2	532.8	579	40
19	Benin	1800	Africa	636559	430.46	569.54	597	47.95
20	Bhutan	1800	Asia	89989	463.36	536.64	629	40
21	Bolivia	1800	America	1100000	404.69	595.31	854	40

Time goes right

Column 1: entities, Column 2: indicators, Column 3 and on: time values (see example)

	A	B	C	D	E	F	G	H	I	J	K
1	country	indicator	2005	2006	2007	2008	2009	2010	2011	2012	2013
2	United States	CO2 emissions (metric ton)	19.61	19.12	19.24	18.49	17.19	17.48	17.02	16.29	16.39
3	Uzbekistan	CO2 emissions (metric ton)	4.31	4.36	4.30	4.36	3.86	3.65	3.87	3.89	3.41
4	St. Vincent and the Grenadines	CO2 emissions (metric ton)	2.02	2.02	2.08	2.08	2.85	2.01	1.81	2.31	1.91
5	Vietnam	CO2 emissions (metric ton)	1.19	1.23	1.25	1.39	1.57	1.69	1.84	1.78	1.70
6	West Bank and Gaza	CO2 emissions (metric ton)	0.83	0.67	0.67	0.57	0.56	0.53	0.57	0.54	0.58
7	World	CO2 emissions (metric ton)	4.52	4.62	4.66	4.74	4.63	4.84	4.98	5.00	5.00
8	Samoa	CO2 emissions (metric ton)	0.94	0.97	1.03	1.04	1.07	1.10	1.25	1.34	1.25
9	Yemen	CO2 emissions (metric ton)	0.98	0.99	0.97	1.00	1.07	0.99	0.81	0.74	0.99
10	South Africa	CO2 emissions (metric ton)	8.82	9.35	9.62	10.09	10.06	9.34	9.22	9.02	8.86
11	Congo	CO2 emissions (metric ton)	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04	0.04
12	Zambia	CO2 emissions (metric ton)	0.19	0.17	0.14	0.17	0.19	0.19	0.20	0.24	0.25
13	United States	Renewable energy consumption (metric ton of oil equivalent)	1.86	1.93	1.35	2.32	2.21	2.64	2.24	2.37	
14	Uzbekistan	Renewable energy consumption (metric ton of oil equivalent)	5.81	5.59	5.45	5.34	5.36	5.24	5.13		
15	St. Vincent and the Grenadines	Renewable energy consumption (metric ton of oil equivalent)	13.69	12.52	14.45	15.70	12.32	12.68	10.32	12.36	11.20
16	Vietnam	Renewable energy consumption (metric ton of oil equivalent)	67.52	73.40	43.56	47.35	39.68	39.14	34.18		
17	West Bank and Gaza	Renewable energy consumption (metric ton of oil equivalent)	17.04	17.12	17.08	17.34	17.97	17.78	17.87	18.12	
18	World	Renewable energy consumption (metric ton of oil equivalent)	29.62	29.20	28.88	27.66	26.14	26.46	23.15		
19	Samoa	Renewable energy consumption (metric ton of oil equivalent)	0.91	0.89	0.93	0.89	0.86	0.86	0.84	0.99	1.00
20	Yemen	Renewable energy consumption (metric ton of oil equivalent)	17.00	15.92	16.29	16.65	16.88	17.11	16.93		
21	South Africa	Renewable energy consumption (metric ton of oil equivalent)	97.88	97.42	97.34	97.19	96.98	97.04	96.84	96.26	95.96

Fig. 8.2 Importable data structures to Gapminder. Based on free material from Gapminder.org, CC-by license

csv formats as well. Data can be arranged in two ways. In one structure time goes down and in the other time goes right. In the import window, several examples are available on how to handle data structures, use the correct forms of time values, etc. (Fig. 8.2). Results can be saved or exported—for instance as a picture.

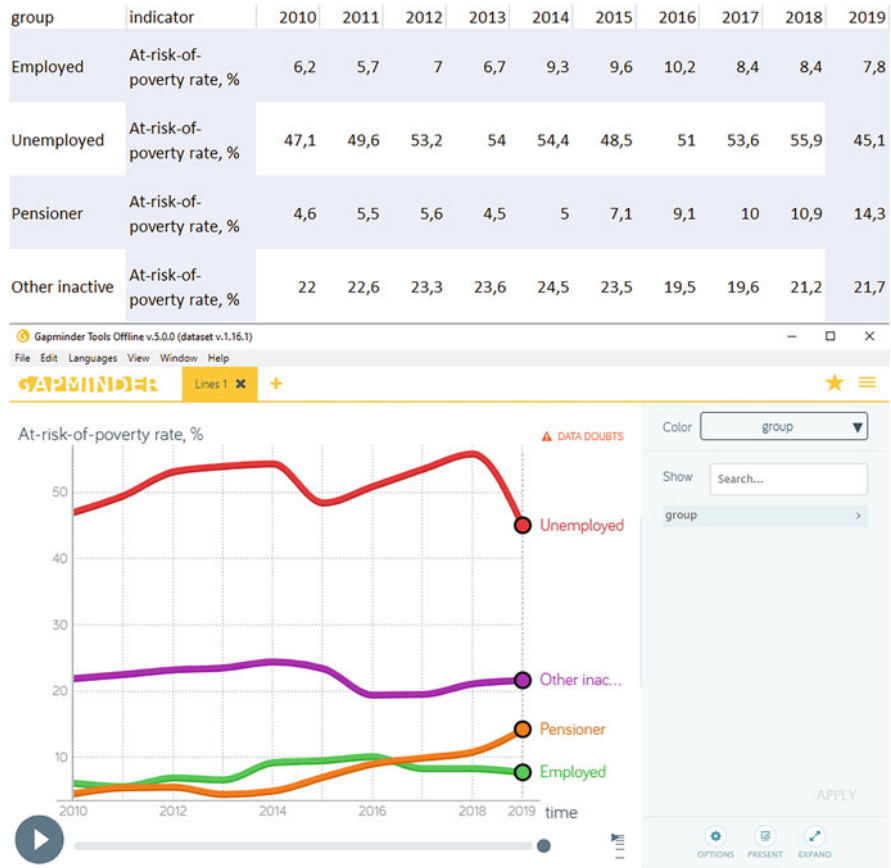


Fig. 8.3 Imported data on at-risk-of-poverty rate by the most frequent activity. Based on free material from Gapminder.org, CC-by license

The Gapminder offline tool also can visualize group data also without geo data. For instance, Fig. 8.3 shows the data structure (yearly data) and the graphical representation of the Hungarian at-risk-of-poverty rate values (half-yearly) amongst the most frequent activity groups between 2010 and 2019.

Why Is It Worth Using Gapminder?

Official statistical databases and associated visualizations, and Gapminder have different advantages and disadvantages. We recommended using them together (see for example in Chap. 10). From the point of view of teaching, we must consider the needs of students when we use databases. Based on generation theories in education (Eckleberry-Hunt & Tucciarone, 2011; Reilly, 2012), people in the same generation have common characteristics from the point of view of expectations of teaching and learning (e.g., in relation to electronic devices, and the importance of team working). Our students are members of Generations Y, Z or Alpha (Ramadlani & Wibisono, 2017). Common traits of these generations are that during lessons these

young people have a need for technology, team working, dealing with entertaining tasks and receiving information visually. Therefore, using static databases and charts are not exciting enough for them—however, those “old fashioned” tools are important and necessary parts of teaching, collecting, and using data.

If we take into account some databases (see more in Chap. 6), we can say that the Eurostat Database² includes a wide range of indicators; data are downloadable, and metadata (i.e. detailed definitions of variables) are available. However, the Eurostat Database offers only graphs and maps, and there are no dynamic (animated) features embedded in these graphs and maps. This lack of interactivity makes it more difficult to provide enjoyable tasks for students. (As noted earlier, if we download data from those databases, then we can embed them into other interactive tools; however, this requires more steps.)

A different part of the Eurostat website hosts the Visualization Tools section.³ It includes some spectacular and dynamic elements (such as “You in the EU” or “Quality of Life”); therefore, they are more attractive and enjoyable for students. Most of the visualization tools include the source of data (which directs the visitor to the related Eurostat Database). However, the number of available indicators included in visualizations is modest compared to the Eurostat Databases.

OECD.Stats⁴ offers a wide range of downloadable indicators with directly available metadata. However, OECD.Stats offers only graphs, and its dynamic feature shows only temporal changes.

A different part of the OECD website contains the Data Visualization Lab⁵ that includes dynamic charts. We can access downloadable data and metadata via more clicks (there is no direct option for downloading data or reading data); however, the number of available data sets is modest compared to OECD.Stats.

Gapminder is a spectacular dynamic visualization tool, and it is easy to find delightful and stimulating tasks for students. It is possible to download data and to access the definitions of the indicators. There is only one drawback: the number of available indicators is modest compared to Eurostat Databases or to OECD.Stats. Table 8.1 offers a comparison of the databases discussed so far.

Based on the comparison of positive and negative features of the databases examined, it may be concluded that Gapminder has the greatest potential from the point of view of teaching Generation Y, Z or Alpha. There are good reasons to make use of other databases (e.g., Eurostat Database, OECD.Stat) as well; however, the usage of Gapminder can make lectures particularly engaging and entertaining for students. Summarizing the main advantages of using Gapminder in classes we can mention the following:

- Gapminder is a free, browser-based, open-source web application.
- Gapminder was designed to be accessible to non-technical audiences; it has a user-friendly interface.

²<https://ec.europa.eu/eurostat/data/database>

³<https://ec.europa.eu/eurostat/help/first-visit/tools>

⁴<https://stats.oecd.org/>

⁵<https://www.oecd.org/dac/financing-sustainable-development/datavisualisations/>

Table 8.1 Comparison of statistical databases (after Kazár, 2016)

Database	Database (Database and tables by themes)	Potential for downloading data	Number of available indicators	Availability of metadata	Potential for enjoyable tasks	Potential for applying visualization tools
Eurostat	Database (Database and tables by themes) Visualization tools (e.g. You in the EU, Quality of life)	Yes	Wide	Yes	Lower	Only graphs and maps
OECD	OECD.Stat OECD Data Visualization Lab	Yes Not directly	Narrower Wide	Through Eurostat Database Yes	Higher Lower	Some, dynamic; less options Only graphs
Gapminder		Yes	Narrower	Yes	Higher	Dynamic; few options Dynamic; more options



Fig. 8.4 Displaying multiple dimensions on the Bubble Chart. Showing a free visualization from Gapminder.org, CC-by license

- It is obviously ‘educational’, and its dynamic visualizations (often revealing spectacular counter-intuitive insights) are very attractive to students with high expectations of the properties of visual displays.
- It is easy to use for students with no background in statistical analysis, and encourages questioning, investigation, and exploration of statistical concepts.
- The data sets are downloadable for further analysis.
- There is an option for importing data in the offline version.

There are, however, some disadvantages:

- Gapminder is limited in terms of the kinds of graphical representations and analyses it offers compared to commercial statistical software packages; however, the restricted menu avoids students being overwhelmed with choices between options they do not understand.
- Gapminder presents a narrower range of indicators than (say) Eurostat or OECD.
- It is not possible to download the visualizations created as video files.

8.3 The Gapminder Bubble Chart in Action

This section introduces the use of the Gapminder Bubble Chart in class. The Bubble Chart is a dynamic visualization tool that can be used when we want to visualize temporal changes in relationships between two variables (indicators) country by country (Fig. 8.4).

How Might the Gapminder Bubble Chart Be Used in Teaching?

From the point of view of teaching, we can use the Bubble Chart for a variety of purposes:

- Visualizing a wide variety of relationships: Gapminder has a long list of indicators on several themes. If we discuss a topic (e.g. social, economic or environmental issues) in a lesson, we can visualize some selected indicators in a spectacular way (e.g. for examining *poverty*, we can visualize and describe the relationship between the *income per capita* and *life expectancy*).
- Visualizing correlation: if we want to make the concept of bivariate correlation or regression more understandable for the students, a scatter plot is a useful tool. However, the Gapminder Bubble Chart is much more than a simple scatter plot; it presents a great deal of data in an accessible way. Instead of a simple scatter of same-size, monochrome dots, there are bubbles with different sizes (here representing population) and colors (indicating region). Temporal changes can be explored using the sliding timeline. If we consider a Civic Statistics topic (e.g. *poverty*), we can visualize a positive correlation between the *income per capita* (log) and *life expectancy*; moreover we can see how this relationship has changed over time.
- Making comparisons: in Fig. 8.4 the bubbles represent countries, so it is possible to make territorial comparisons with the help of a Bubble Chart. There is a color option that helps us to make comparisons among world regions as well as between individual countries. For instance, on the topic of *poverty*, we can highlight that the *income per capita* was lower in African countries than in European countries in 2018.
- Comparing groups: again, with the help of color option mentioned above it is possible to create groups based on other grouping variables (e.g. religion or OECD membership). If we highlight the topic of *poverty* again, we can see that in OECD countries the *income per capita* and *life expectancy* were also higher in 2018 than in the rest of the countries.

Configuring the Gapminder Bubble Chart

There are built-in visualizations (e.g. Wealth and Health of Nations) in the Bubble Chart; however, users can change all the main options when creating the content of the visualization:

- Axes: users can select two indicators, one for the horizontal and one for the vertical axes. The indicators are grouped in topics and subtopics, but they are also listed in the Data menu. It is possible to display the selected indicator on a logarithmic scale (a log scale can make it easier to see trends, and a log scale expands the scale at low values and compresses the scale at high values. Students need to be careful to read the scales of variables that are displayed. To illustrate: the opening visualization (Wealth and Health of Nations) displays income on a log scale—this can be seen (and changed) when one clicks on the indicator (Fig. 8.5).
- Size of bubbles: each bubble represents a country, and the size of a bubble shows the size of the population of the country as a default setting. However, it is possible to represent other indicators using the size of the bubble; or using no indicators (i.e making each bubble the same size).

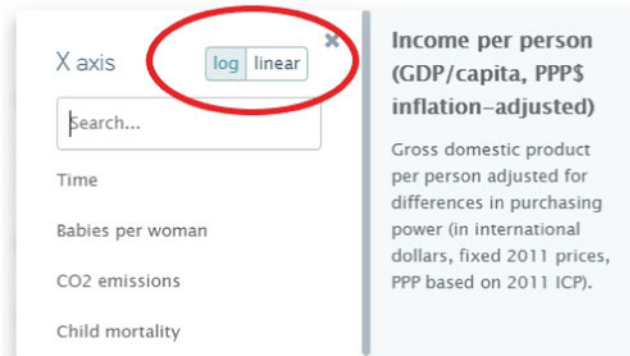


Fig. 8.5 Rescaling the x -axis. Based on free material from Gapminder.org, CC-by license

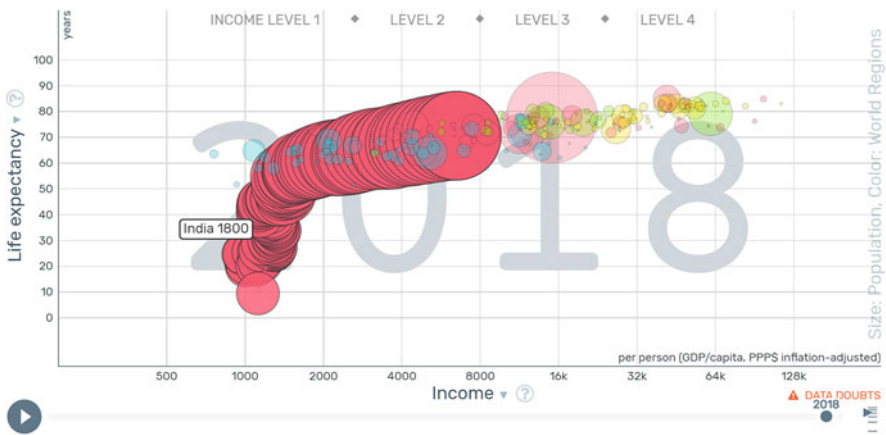


Fig. 8.6 Tracing the trajectory of a country over time. Showing a free visualization from Gapminder.org, CC-by license

- Color of bubbles: users can select the color based on several grouping variables (e.g. country properties such as *region*, *landlocked*, *G77* and *OECD countries*, *income*, or *main religion*) or based on some other metric indicator (also listed in topics). It is also possible to use no coloring in the visualization.
- Selecting a country: users can select one or just a few countries in a visualization. Selecting the *trail* button causes Gapminder to draw the path of the selected country over time (see Fig. 8.6).
- Time: can show the temporal change of the selected indicators. One can set the speed of visualization; and can pause the visualization at any time (e.g., when it shows outliers or sharp increases/decreases); one can pause the display in class and ask students to make forecasts.

Besides the content, there are other options connected to the appearance of the visualization. The opacity of the bubbles, the size of labels, the look of visualization and the position of the options menu can all be set. One can zoom in and zoom out, use a presentation mode (which enlarges the labels), or expand the visualization to full screen.

Besides the options available on the Bubble chart, it is worth mentioning other options in Gapminder that are connected to the usage of any visualization tools in Gapminder:

- **Data:** there are two main reasons why the data menu is important. First, metadata for each indicator are available. Gapminder uses other databases (e.g. OECD data), so we can find the source of the indicator in the data menu, as well. In addition, we can download data from this menu. This is important if we want to do further calculations or analysis with data visualized in Gapminder.
- **Offline usage:** it is possible to use Gapminder offline; versions can be downloaded for Windows, for Mac and for Linux. Using the offline version, students can use Gapminder without an internet connection.
- The offline version can also be used for visualizing your own data sets.
- Tutorial videos are available explaining the usage, and for the data structure specification.⁶
- **Sharing charts:** to show a visualization you have created, you can create a link and share it via several platforms (e.g. Facebook, Twitter, e-mail); or the link can be embedded into a presentation.

How Can I Use the Gapminder Bubble Chart in My Teaching?

Several teaching resources (e.g. videos, lesson plans, quizzes, answers, presentations) can be found on the Gapminder Website. Examples of how we have used Gapminder in our undergraduate classes can be found in ProCivicStat lesson plans via the CivicStatMap⁷ (see more description about CivicStatMap in Chap. 6). These lesson plans set out tasks for students, and detailed plans for teachers (which include additional information such as some background to the theme is discussed, what statistical topics are included, level of the lesson plan, prerequisite knowledge, and possible answers). We have used these tasks with both BA and Master level students in the University of Szeged Faculty of Economics and Business Administration since 2017. In the following, we share some experiences concerning the applications of parts of the lesson plans.

Application 1

The lesson plan called “**Introduction to the usage of Gapminder—Level 1**” shows how the Gapminder Bubble Chart works via a built-in visualization. With the help of the questions and tasks of this lesson plan, students learn the functions of the Bubble Chart. This lesson plan is valuable if we want to explore or introduce how

⁶<https://vizabi.org/tutorials/2017/04/03/show-your-data/>

⁷https://rstudio.up.pt/shiny/users/pcs/CivicStatMap_English/

Gapminder works; therefore, the usage of it is important when students use Gapminder for the first time. In the frame of required preliminary knowledge, it is worth ensuring that students know about different chart types (e.g. line charts, bar charts) and about different functions (e.g. linear and log functions). You can access the lesson plan with sample activities solutions via CivicStatMap here.⁸

We introduced these tasks with a variety of undergraduate students (BA level) in the University of Szeged Faculty of Economics and Business Administration as part of an introductory statistics course in 2017. The topic of the given class was the use of official statistical databases. We wanted to collect their first impressions and experiences; therefore, we surveyed the students' opinions about the class via open-ended questions (and collected qualitative data). Based on our experiences and the students' opinions about the application, we can highlight the following:

- It was difficult for the students to answer questions because they found it hard to interpret what they saw. With the help of some additional questions (e.g. “What are the axes?”; “What do the bubbles indicate?”), students were able to better understand the meaning of the visualization. (See more detailed questions in the teacher version of the lesson plan here.⁹)
- If students are not familiar with the usage of Gapminder, it is difficult for them to understand how the platform works (technical details) and to understand the message from the specific topic (content) together. It is important to understand how to handle the platform (which highlights the importance of this lesson plan), then they can focus on the message of the applied topic.
- In conventional classes, it can be difficult to grab a student's attention; however, during the Gapminder application they paid great attention to the tasks. Moreover, students said they liked the application and in particular the dynamic visualization tools in Gapminder.
- Based on our experiences, Gapminder works very well in class. However, BA students needed some extra questions and discussions to understand the meaning of the visualization; open-ended questions were too challenging at the outset.

Application 2

The lesson plan (Application 1) described above focused on technical aspects of the Bubble Chart. However, another lesson plan (called **Examining poverty and income inequalities with the help of Gapminder—Level 2**¹⁰ shows how to use Bubble Chart to examine a specific topic and how to explore specific statistical methods (e.g. correlation or log transformation) via visualization. In terms of prerequisite knowledge for this lesson plan, it is worth having information about:

⁸https://rstudio.up.pt/shiny/users/pcs/civicstatmap/5.301_TV_gapminder_level1_EN.pdf

⁹https://rstudio.up.pt/shiny/users/pcs/civicstatmap/5.301_TV_gapminder_level1_EN.pdf

¹⁰https://rstudio.up.pt/shiny/users/pcs/civicstatmap/5.302_TV_gapminder_level2_poverty_EN.pdf

- the usage of the Gapminder Bubble Chart (see Application 1).
- poverty and income inequality indicators.
- log/linear transformations, the coefficient of correlation.

Sample activities of this lesson plan can be seen below:

Introduction

Watch the Wealth and Health of Nations in Gapminder World and answer the following questions!

1. *What is being explored in this display? (What are the indicators?)*
2. *What can you state about the changes in time?*
3. *Are there any regional differences? Which countries are rich?*
4. *Are big (large population) countries rich?*
5. *What is the relation between life expectancy and income per person?*

(Students can answer these questions by making presentations, or they can have an oral discussion about the questions; you can collect ideas for other sorts of assessment tasks from Chaps. 14–16).

Examine the relationship between extreme poverty (economy—poverty—ratio of people below 3.20\$ a day) and food supply (health—nutrition)!

6. *What can you see in the visualization? Create an analysis in a few sentences!*
7. *Use the urban poverty indicator instead of (% urban people below national urban poverty line) instead of extreme poverty (ratio of people below 3.20\$ a day)! What can you state about India’s position in 2000 compared to Task 6)?*

We have used these tasks with Master level students in the University of Szeged Faculty of Economics and Business Administration since 2018, in the context of a multivariate statistics course. The topic of the given lesson was correlation and regression analysis; and the aim of this lesson was to understand the concepts of correlation and regression analysis by using them for analyzing indicators from the topic of *poverty*. As a sample activity from this lesson, we watched the visualization called “Wealth and Health of Nations” in Gapminder, and addressed Task 1–4 above (as “warming up” questions). Then Task 5 (What is the relation between life expectancy and income per person?) to explore the concept of correlation; and we discussed the situation from 2020 (see Fig. 8.7).

There is a positive linear relationship between *life expectancy* and the (*log*) *income per capita* (there is a nonlinear relationship between *life expectancy* and *income per capita*; the *income per capita* is expressed in a log scale in Gapminder as a default setting). This is true over time; however the relationship seemed to be stronger sometimes (e.g. in 2020). If a country has higher (*log*) *income per capita*, there is higher *life expectancy* (and vice versa).

You can access the entire list of tasks with solutions via CivicStatMap.¹¹

¹¹ https://rstudio.up.pt/shiny/users/pcs/civicstatmap/5.302_TV_gapminder_level2_poverty_EN.pdf



Fig. 8.7 Visualizing positive correlation in Gapminder. Showing a free visualization from [Gapminder.org](https://www.gapminder.org), CC-by license

With the help of the questions and tasks of this lesson plan, students gained knowledge via:

- Understanding statistical concepts (e.g. correlation, log transformation) better with the help of poverty indicators.
- Understanding *poverty* better by getting to know the situation of several countries via the visualizations (e.g. Task 3, Task 4, Task 6).
- Understanding that *poverty* is a multi-dimensional concept. Students use several indicators (e.g. *income per person*, *extreme poverty*, or *urban poverty*) during the tasks for describing poverty; and poverty can show different pictures depending on the selected indicators. Therefore, students can see that there is no “one and the best” indicator for describing poverty. Students need to realize that they have to describe poverty using a variety of indicators if they are to have a better picture of (or a more nuanced approach to) the concept.

We wanted to collect the students’ first impressions and experiences; therefore, we interviewed the students’ after the class in open ended questions (and also collected qualitative data). Based on our experiences and the students’ opinions about the application, we can highlight the following:

- Initially, students required time to recognize what they see; however, it was easier for them (MA students) to understand what they saw than BA students. We needed fewer additional questions to facilitate understanding around the meaning of the visualization.
- Students were able to recognize the positive correlation between some variables; however, they needed some explanation of possible problems interpreting the visualization because the *GDP/capita* in the “Wealth and Health” visualization is shown on a log scale.
- Similar to BA students, MA students also liked using Gapminder; this visualization tool was able to grab their attention.

- Based on our experiences, Gapminder helped students to visualize correlation and regression. However, MA students also needed some probing questions to encourage them to explore all of the options of menus in Gapminder.

Summarizing all of our experiences of using the Gapminder Bubble Chart, we note that Gapminder is a useful tool to grab students' attention; and it can help to understand and visualize statistical methods (e.g. correlation and regression) better. However, students watching the Gapminder Bubble Chart for the first time can find it difficult to understand what they see. However, if sufficient time is allocated for introducing the Bubble Chart, this problem is easily overcome.

Bubble charts can only visualize relationships between variables. If more analysis is needed, for instance, via regression models, or calculating correlations between variables, we need to combine Gapminder with other software, for instance with Excel. An example of this is the analysis of the relationship between *income* and *life expectancy*. To conduct appropriate analyses, we need to download the data from Gapminder. Data tables can be downloaded from the Facts/Data menu, which is situated under the Facts menu, or by clicking on the question mark on either axis. This gives access to details such as the source of the data and about the used indicator. The original value of the indicator can be downloaded independently of the fact that a linear or logarithmic axis is used in the visualisation. After downloading the data, we have separate Excel files for the different variables. To analyze the relationship, it is required to merge the tables. We must pay attention to this step. Different tables do not contain each country or territory, so the number of rows can be different in the case of the different tables even though the data structures are the same.

After preparing a merged Excel file we can examine the original questions. We suggest preparing a Scatter plot in Excel and selecting and fitting a regression curve on the data and showing the model equation and the predictive power of the model with the help of the fit via a trendline option (Fig. 8.8).

Throughout the book we have emphasised the importance of showing and discussing nonlinear relations on societal issues (see more in Chap. 4). With this example we can describe the nonlinear relationship between the variables. The logarithmic model has the best fit to the data ($R^2 = 0.7034$); the correlation index (square root of R^2) is 0.84 so shows a strong relationship between *life expectancy* and *income/capita*. This exercise also can show that the linear model has lower predictive power ($R^2 = 0.4655$).

As an interesting supplement, this application can highlight dangers encountered when interpreting visualizations. If we ask which country had the largest GDP/capita in 2020, the answer is Qatar according to the Excel sheet. But, if we use the online Bubble Chart tool, the answer could be Luxemburg (Fig. 8.7). The correct answer is Qatar. The basis of this misperception is that, when variables are selected in Gapminder, we think that we can see all the countries, but this is definitely not true. In this example the visible part of axis X does not show the maximum value of the variable. If we modify the zoom to 100%, the data of Qatar will be perceptible on the chart.

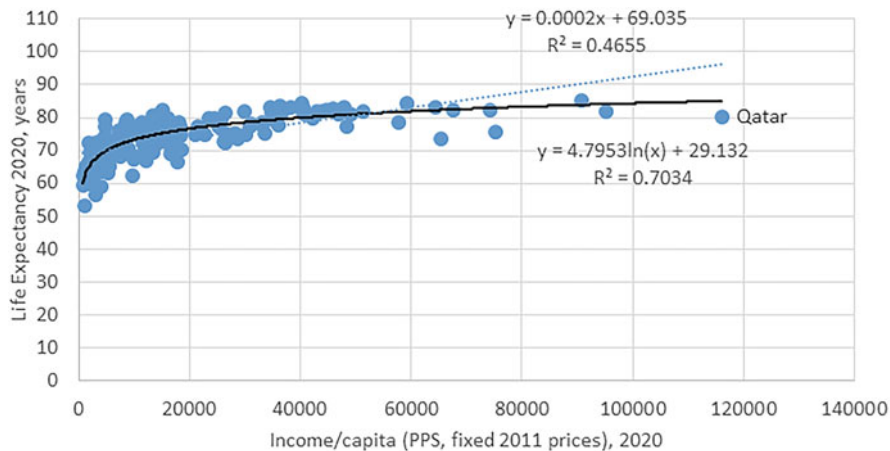


Fig. 8.8 Additional regression analysis to Gapminder in Excel. Based on free material from [Gapminder.org](https://gapminder.org), CC-by license

8.4 Other Gapminder Tools in Action: Ageing Society

The Ageing Society is a societal challenge with social and economic impacts in the developed world. There are several online visualization tools, population pyramids, and datasets that can be used to explore ageing societies from a statistical perspective, for instance via population pyramids, census data, and the websites of National Statistical Agencies. These resources are used in the ProCivicStat Ageing Society module.¹² The first part of this module concentrates on the definition of age distributions, and the interpretation of population pyramids with the help of visualization tools. The second part focuses on the challenges of an ageing society *per se*. The exploration of ageing societies can start with a warm up question, for instance, *which country has the oldest population?* We think that asking this question as a starting point, is in line with the Gapminder's logic: firstly, we collect information on how students view this phenomenon, and explore any misconceptions that emerge.

After that in a group discussion, we can focus on what the term 'ageing society' means, exploring the likely causes, and consequences. We need indicators to analyze the ageing of a society, so we need to understand how official statistics can measure and examine the ageing society. We can focus on different indicators such as the *Old age dependency ratio*, the *Ageing index*, the *Growth rate of population*, the *Median age*, the *ratio of elderly people*, *population growth*, *fertility rate*, *mortality rate*, *life expectancy*, etc. For instance, the old age dependency ratio compares the number of elderly people (aged 65 and over) to the number of people of working age (15–64 years old). The *ageing index* is the ratio of the *number of elderly people (65–X years)* and the *child population (0–14 years)*.

¹²https://rstudio.up.pt/shiny/users/pcs/civicstatmap/5.303_TV_AgeingSociety_EN.pdf

After the discussion phase, we suggest turning to statistical databases and visualization tools to analyze the phenomenon. Ageing is a process, so we need time series or data from different years to describe this phenomenon, and understand what data are showing. The selection of the data source and visualization tools is based on our goals. We also can use National Statistical Agencies' websites and reports if we want to analyze national data only, without Gapminder. We suggest starting at the national level and compare different years. For instance, asking students to collect data to calculate the *ageing index*, and *old age dependency ratio* for: the most recent year for which data are available; 50 years ago; and in projections for 2100; It means data collection, calculation and interpretation of ratios and visualisation via population pyramids. With the help of this exercise, we can estimate (for instance) how big is the proportion of elderly people whose needs have to be supported by economically active people in 50 years time. Based on this result we can provoke newer questions. What kind of social and economic challenges does this result lead to?—for instance, challenges in the labour market, the sustainability of pension- and health care systems. What kind of political decisions should we make to solve these issues, to stop the ageing trend, etc.

In the next phase, if we want to do international comparisons or international forecasts, we can use (for instance) online population pyramids or international data sources. Gapminder also can be used to analyze ageing patterns in different countries over time; Fig. 8.9 provides some illustrative visualizations and can answer the warmup question if we analyze it based on the *old age dependency ratio*. Students can be set a variety of interesting challenges, for instance:

1. Analyze the order of countries by old age dependency ratio. Which countries occupy the first three places? What is the rank of your own country?
2. What was the order 50 years ago?
3. Visualize time series for the last 50 years in the case of the countries in the first three positions.
4. What answers can we give to the previous two questions if we analyze the ratio of population years 65 and above?
5. Compare your country's age distribution to the continent's age distribution.
6. Compare the different continents' age distributions to the world's age distribution.
7. How can you interpret your results?
8. What kind of thoughts are implied by the results?

We ask the students to document their work. They prepare different charts, which can illustrate answers to the questions with the help of these questions. They interpret and compare data and present their results orally. This activity can show the territorial differences and similarities in ageing and temporal changes in the values of the indicators. We can ask other questions about the phenomenon: is there any relationship between social and economic differences and differences in ageing? Or between temporal changes in ageing indicators and the changes in the social and economic situation?

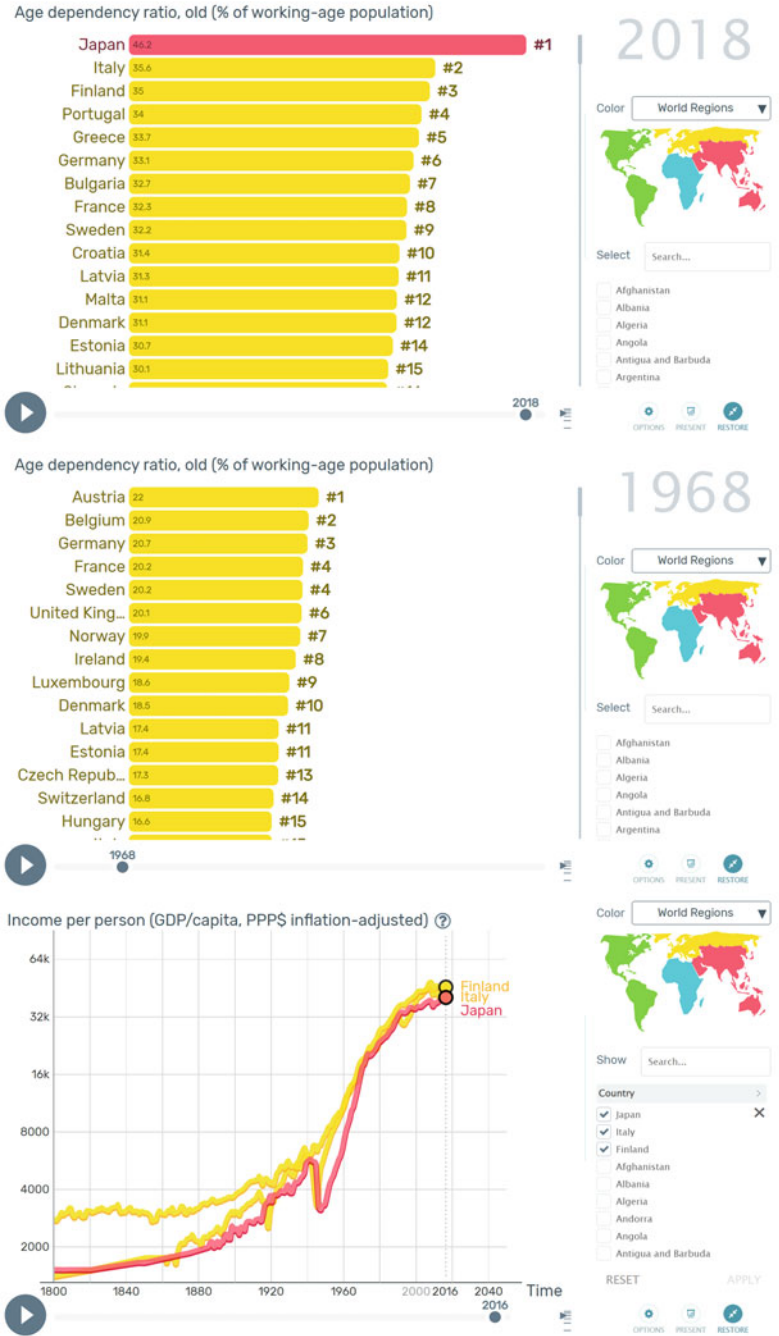


Fig. 8.9 Gapminder in action on the Ageing Society. Showing a free visualization from Gapminder.org, CC-by license

Engaging in these activities should result in several important take-home messages for students from different points of view: data sources, definition and methodology associated with the indicators, interpretations, software functionalities, and the likely robustness of projection methodology. For example:

- Data sources do not necessarily have the indicators we might want, or the indicators can be found in different menus. For example, In Gapminder, the proportion of the population years 60 and above are available in the group of population indicators. Other indicators are situated in More data/wdi/Health/Population/Dynamics and More data/wdi/Health/Population/Structure groups. We must be careful about the indicators, and knowledgeable about sources of data. When we prepare a lesson plan, we suggest selecting indicators which we want to use in classes in advance, and after that, search for a visualization tool or database, which contains the selected indicator.
- ‘The answer’ can depend on the choice (and definition) of the indicator used. For example, if we compare the ranks by the different variables in the same year, Japan is situated in the first place in both cases in 2018, but we can see different orders for other countries.
- An improvement in the rankings does not necessarily mean that the underlying situation has improved. For instance, 50 years ago Hungary was situated in 15th place, improving to 21st place in 2018. So now more countries have a higher old-age dependency ratio than Hungary. Nevertheless, the Hungarian old-age dependency ratio increased from 16.6 to 28.8% during this period.
- The different charts allow us to see if something is a local or a global problem; here, it is clear that ageing is a problem across the developed world.
- The combination of different Gapminder tools can highlight different aspects of a complex phenomenon.
- Software tools have different merits, and students can develop an aesthetic for data visualization (here, it is clear that Gapminder is not a good choice if our purpose is to create a dynamic population pyramid).
- comparing distributions between groups is not a straightforward process (here we illustrated the use of relative frequencies).
- Modelling is a complex process. Different developers built population projections into their visualization tools, based on different models, and assumptions. It is essential to understand the models, and the assumptions being made. Comparing projections from different models (say for different countries) is deeply problematic and should be avoided.

8.5 Conclusion

Gapminder offers several visualization tools, which can help to visualize complex problems, trends, and relationships between variables from different points of view. These user-friendly apps are engaging for students and encourage exploration and problem-based learning. Gapminder has some limitations regarding the available

data and tools available for statistical analysis. However, lessons with Gapminder provide a strong base for further statistical education and a springboard to other analytical and visualization tools, to which we can import Gapminder data and do more data analysis.

Gapminder can be used not only from primary school education to tertiary education, but also in the case of adult education to discover the World, to analyze social and economic problems with the help of data and visualization, and to develop statistical thinking and reasoning. The combined use of the different Gapminder applications for instance (Bubble chart and Ranks) can help to illustrate complex problems from different points of views.

References

- Eckleberry-Hunt, J., & Tucciarone, J. (2011). The challenges and opportunities of teaching “generation Y”. *Journal of Graduate Medical Education*, 3(4), 458–461.
- Kazár, K. (2016). Using statistical datasets for describing poverty and income inequality. In J. Engel (Ed.), *Promoting understanding of statistics about society proceedings of the Roundtable conference of the International Association of Statistics Education (IASE)*. International Statistical Institute. <http://iase-web.org/documents/papers/rt2016/Kazar.pdf>
- Ramadlani, A. K., & Wibisono, M. (2017). Visual literacy and character for alpha generation. In A. M. Al-Ma’youf et al. (Eds.), *Proceedings international seminar on language, education and culture* (pp. 1–7). University Press. Accessed October 26, 2019, from <http://sastra.um.ac.id/wp-content/uploads/2017/12/0-cover-depan.pdf>
- Reilly, P. (2012). Understanding and teaching generation Y. *English Teaching Forum*, 50(1), 2–11.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

