MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

DATA VISUALIZATION AND VISUAL ANALYTICS

Guidelines
to laboratory and independent work
for Master's (second) degree
students of speciality 122 "Computer Science"

Kharkiv
S. Kuznets KhNUE
2020

UDC 044.032.6(07.034)
D24

Compiled by: V. Derykhovska
V. Shlykova

Затверджено на засіданні кафедри статистики і економічного прогнозування.
Протокол № 3 від 16.09.2019 р.

Самостійне електронне текстове мережеве видання

Data visualization and visual analytics [Electronic resource] :
D24 guidelines to laboratory and independent work for Master's (second) degree students of speciality 122 "Computer Science" / compiled by V. Derykhovska, V. Shlykova. – Kharkiv : S. Kuznets KhNUE, 2020. – 102 p. (English)

Tasks for laboratory and independent work on the academic discipline and guidelines to them are given to help the students gain theoretical and practical skills in the use of the modern analytical tools and infographics for the analysis of complex mass social and economic phenomena and processes.

For Master's (second) degree students of speciality 122 "Computer Science".

UDC 044.032.6(07.034)

Introduction

Visual analysis is a fast-growing, promising area that combines the benefits of graphical visualization and the power of analytical computing when working with large masses of digital information. Data visualization allows you to identify patterns, trends, and correlations that may otherwise remain unnoticed in traditional reports or spreadsheets.

The laboratory and independent work is aimed at the development of students' skills in the effective use of the modern analytical tools and infographics for the analysis of complex mass social and economic phenomena and processes. Students will learn to make informed decisions based on the preprocessing of data and its interactive visualization.

Studying this discipline enables the student:

to choose the most effective means of data visualization for solving specific business tasks;

to use infographic tools and choose analytical methods for processing the information in a changing information environment;

to navigate the information space to collect the necessary data;

to prepare information for processing by modern methods of business analytics;

to interpret and present the results of analytical research;

to prepare a report on the results of the conducted analytical research.

The purpose of teaching this academic discipline is the expansion and deepening of theoretical knowledge and professional competences in the visual analysis of business processes and making effective solutions through the use of analytical methods and means of data visualization.

The object of the discipline is business processes and phenomena.

The subject of the discipline is the theoretical and practical aspects of economic interpretation and interactive visualization of the results of the analysis of business processes on the basis of the use of modern analytical methods and tools of infographics.

Content module 1
The theoretical basis for data visualization

Topic 1. Visual information in the information society

Laboratory work 1.1
The formation of information space for visualization. An overview of information sources

The purpose of the work is to acquire the skills in finding information in accordance with the purpose of the study.

The task is to get acquainted with domestic and foreign sources of statistical information and to form the information space on the chosen research topic.

Guidelines

Currently, there is a wide range of distribution channels for statistical information, which is largely due to the needs and requests of different categories of users. As a consequence, users express a desire to receive information in various ways and means: on removable media, from the channels of communication, in the on-line mode. It is important, however, to satisfy all their requirements both from the point of view of the way data is presented, and from the point of view of the content, time of presentation of data, formats and prices.

All users of statistical information can be divided into the following main groups:

1. Government bodies:
the President;
the Council of Ministers;
the Ministry of Economics;
the National Bank;
the Ministry of Finance;
the sectorial ministries and departments.
2. Regions:
executive authorities in the regions.

3. Science and research:

research institutes;

higher educational institutions (universities, institutes);

libraries (the Central State Library and scientific libraries).

4. Financial institutions and other enterprises, associations and organizations:

banks;

exchange brokers;

insurance companies;

unions;

trade associations;

companies.

5. International and interstate organizations:

the Eurostat;

the International Monetary Fund, the World Bank, the UNECE, the ILO, the UNESCO and others;

the Interstate Economic Committee, the Executive Committee of the Commonwealth of Independent States and others (for the CIS countries);

embassies.

6. The public:

Parliament;

mass media (television, radio and press);

social organizations;

educational institutions (eg schools);

public libraries;

citizens.

The main data distribution channels are:

1) publication;

2) on-line services;

3) mass media and other forms of dissemination.

The Internet is the fastest, most accessible and convenient means of dissemination of statistical information.

The main sources of statistical information which are authoritative and disseminate current, accurate information in accordance with the purpose of using it and its profile are presented in Table 1.1.

Table 1.1

On-line statistical services

Statistical services of Ukraine	
The State Statistical Service of Ukraine	*http://www.ukrstat.gov.ua*
Head of statistics department in Kharkiv region	*http://kh.ukrstat.gov.ua*
The only public web portal of open data	*https://data.gov.ua*
Statistical services of other states and international organizations	
The United Nations Statistics Division	*https://unstats.un.org*
The Statistical Office of the European Economic Commission	*http://w3.unece.org/PXWeb/en*
The Statistics Department of the International Labor Organization (ILO)	*http://www.ilo.org/global/statistics-and-databases/lang-en/index.htm*
The International Monetary Fund	*http://www.imf.org/en/Data*
The United Nations Development Program	*http://www.undp.org*
The United Nations Educational, Scientific and Cultural Organization (UNESCO)	*https://en.unesco.org*
The UNESCO Institute for Statistics	*http://uis.unesco.org*
The United Nations Children's Fund (UNICEF)	*https://www.unicef.org*
The World Bank	*https://data.worldbank.org*
The World Health Organization (WHO)	*http://www.who.int/gho/en/*
The World Trade Organization (WTO)	*https://www.wto.org/english/res_e/statis_e/*
The Eurostat	*http://ec.europa.eu/eurostat/data/database*
The International Statistical Institute	*https://www.isi–web.org*
The Interstate Statistical Committee of the Commonwealth of Independent States (CIS Statistical Committee)	*http://cisstat.org*
The Organization for Economic Cooperation and Development (OECD)	*http://www.oecd.org*
The Eurasian Economic Commission	*http://www.eurasiancommission.org*

The task for independent work: visit all the websites of statistical services and provide statistical information on the chosen topic. Statistical data should be presented in the tabular form in the worksheets of Microsoft Excel. Each table should contain the name, the headlines (columns' name), and the units of measure.

Laboratory work 1.2
Visualizing the data in MS Excel

The purpose of the work is to acquire skills in the visualization of statistical information by means of MS Excel.

The task is to learn how to graphically represent statistical data by using the Chart Wizard of MS Excel. Also discussed are ways to create, configure, and format diagrams, methods of working with sparklines, application of the rules of conditional formatting to extract data that meets the specified data.

Guidelines

MS Excel is a multifunctional application that is used in any field of activity. Various MS Excel tools for calculation analysis allow you to process a large amount of data and visually present the results.

1. One way to visualize the data is a diagram. MS Excel allows you to build a large number of different diagrams. A diagram is a way to visualize the information given in the form of a table of numbers. The graphical format of the diagram simplifies the understanding of large amounts of information and connections between different series of data.

The diagram also provides an overview of the situation making it possible to present the data and find important trends.

Diagrams are created on the basis of data contained in working sheets. Excel charts are dynamic which means that if the user changes the data on the working sheets, on which the diagram was built, they will automatically update the diagram.

The diagram in Excel includes many objects, each of which can be selected and changed separately. When you move the mouse pointer over the chart, a tooltip appears next to it, indicating the name of the object. Fig. 1.1 shows an example of a diagram with all objects.

The presence of all these objects is a prerequisite for correct visualization of data.

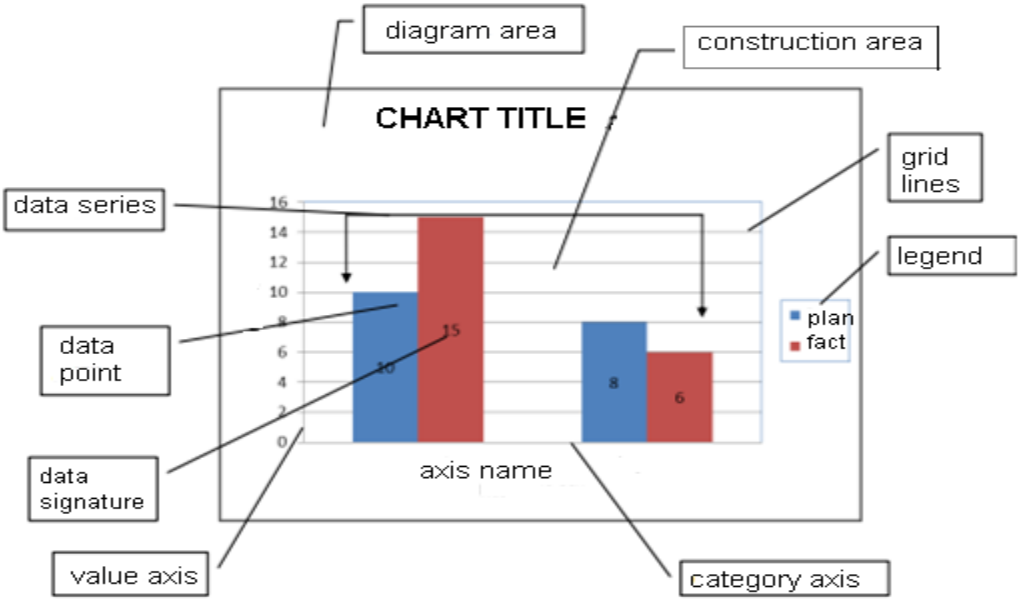


Fig. 1.1. **An example of a diagram with all objects**

There are 11 standard chart types offered by Excel that allow you to effectively present data. For each type of chart, there are several subtypes. Many diagrams have a volumetric format. In most volumetric formats, diagrams do not add new information to the presentation of data on a flat chart, but make more impression when preparing reports or slides for the report. Fig. 1.2 shows which chart types can be built in Excel 2010.

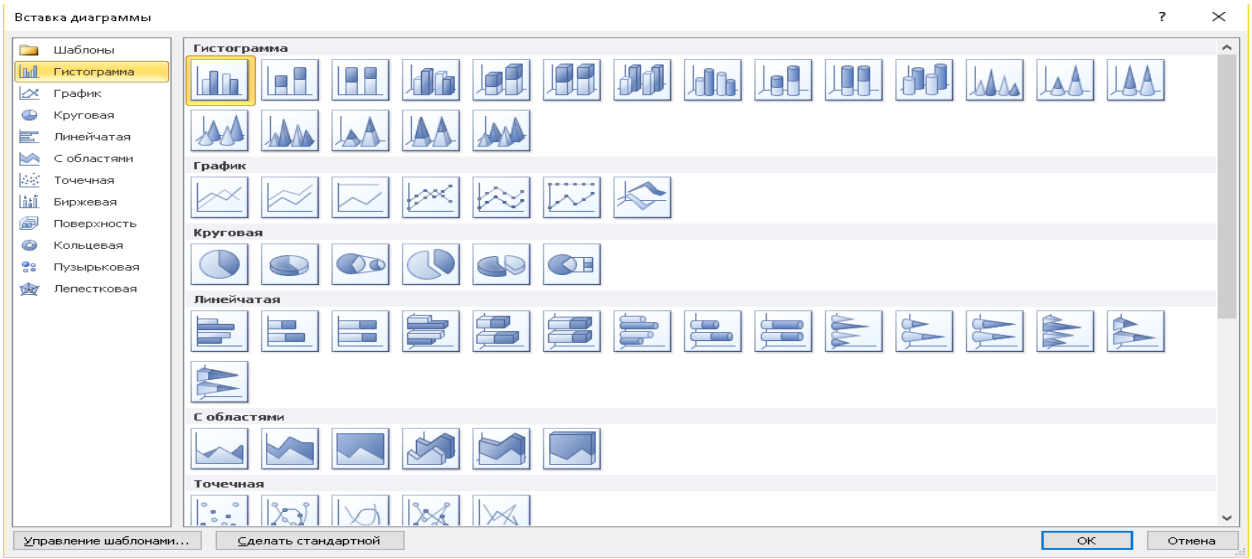


Fig. 1.2. **The window for selecting the standard diagram type**

The standard types of diagrams, their purpose and limitations for use, as well as their subtypes are presented in Table 1.2

Table 1.2

The standard chart types and their subtypes

The name of the diagram type	The purpose	The subtypes
1	2	3
Histograms	are used to compare individual quantities or their changes over a period of time. In the histogram, each data point is displayed as a separate vertical column whose height displays the value	Histograms with grouping and volumetric histograms with grouping. Histograms with accumulation and volumetric histograms with accumulation. Normalized histograms with accumulation and volume normalized accumulated histograms. 3D bar graphs. Cylinder, pyramid and cone histograms
Graphs	allow you to display changes in continuous quantities over time. On the graphs, category data is evenly distributed along the horizontal axis, and all values are evenly distributed along the vertical axis	Graphs and graphs with markers. Stacked graphs and accumulated graphs with markers. Normalized schedules with accumulation and normalized schedules with accumulation with markers. Volumetric charts
Pie charts	Data in one column or row of a sheet can be represented as a pie chart. A pie chart reflects the contribution to the total amount of each. Data points that are output as percentages of the entire circle. Pie-chart sectors can be pushed out of the general circle to emphasize the data points they display. **Pie charts are recommended if:** 1) only one row of data is required to be displayed; 2) the values in the data series are non-negative	Pie charts and 3D pie charts. Cut pie charts and three-dimensional cut pie charts. Secondary pie charts and secondary histograms

Table 1.2 (continuation)

1	2	3
Bar charts	Bar charts resemble histograms used for comparison of individual elements. The data in such a chart are displayed left to right and bottom to top	Grouped bar charts and grouped volumetric bar charts. Stacked bar charts and volumetric line charts with accumulation. Normalized bar charts with accumulation and volumetric normalized bar charts with accumulation. Ruled cylinder, pyramid and cone charts
Diagrams with areas	are used in the same situations as charts, to display the change of data over a certain period of time, and allow you to trace the change in the sum of the values of all series of data and the contribution of each series to this amount. **It is not recommended to use more than five rows of data in a diagram with areas**, otherwise it will be cluttered with information	Diagrams with areas and volumetric diagrams with areas. Diagrams with areas with accumulation of volumetric diagrams with areas with accumulation. Normalized diagrams with areas with accumulation and volume normalized diagrams with areas with accumulation
Dot charts or point diagrams	are usually used to display data corresponding to continuous time intervals. From most other types of diagrams they differ in that they do not have an axis of categories. The axes display the values. Point diagrams are often used to illustrate the dependencies between two variables. As a rule, this type of diagram is used in the field of scientific and engineering research	Point diagrams with markers. Point diagrams with smooth lines and dot diagrams with smooth lines and markers. Point diagrams with straight lines and point diagrams with straight lines and markers
Stock charts	are most often used to illustrate changes in stock prices. However, they can also be used to derive scientific data. For example, with the help of a stock chart you can imagine daily or yearly temperature fluctuations	Charts (highest rate, lowest rate, closing rate). Opening-maximum-minimum-closing. Volume-maximum-minimum-closing

Table 1.2 (the end)

1	2	3
	To create an exchange chart, you must correctly organize the data. Depending on the type, three to five series of data are required to construct such a diagram	Volume-discovery-maximum-minimum-closing
Surface diagrams	display two or more series of data in the form of a surface. Unlike other types of diagrams, in the surface chart, colors are not identified by rows of data, but by values. The number of colors used for this is determined by the value of the main division axis of values. One color corresponds to one basic division of the axis. A surface diagram is useful if it is required to find an optimal combination of data from two sets – as on a topographic map, related to the same ranges, while highlighting colors and hatching. Surface charts can be used to illustrate categories and data sets that are numeric values	Surface 3D diagrams. Wire surface diagrams. Contour diagrams. Wire contour diagrams
Ring diagrams	are similar to circular ones, except that, in the center, they have a hole and they can output more than one row of data. **In the ring diagram, it is not possible to identify series of data directly**	Ring diagrams. Fragmented ring diagrams
Flap charts (petal diagrams)	In the pie chart, a separate axis is defined for each category. The axes are directed outward from the center of the diagram. The value of each data point is marked on the corresponding axis	Petal diagrams and petal diagrams with markers. Flap charts with areas
Bubble charts	A bubble chart can be thought of as a standard point diagram, in which additional values are represented by bubbles of different sizes. Like in the point diagram, in the bubble diagram, both axes are axes of values, there are no axes of categories in it	Bubble or volume bubble diagrams

2. Building a basic diagram. To build a diagram, you need to take the following steps:

1. Select the data you want to use to build a diagram. It is desirable for the selected data to include header lines and columns.

2. On the *Insert* tab in the *Diagrams* group, select the diagram type and subtype. To view all available chart types, click the group button that opens the *Insert Diagram* dialog box to select the chart type (Fig. 1.3).

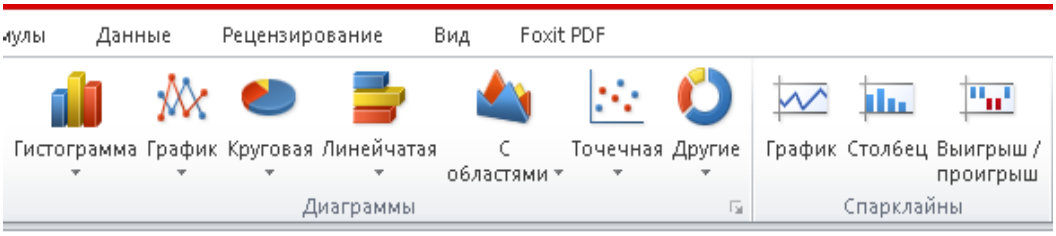


Fig. 1.3. **Diagram groups**

By default, a diagram is added to the worksheet as an embedded diagram.

3. Using the context tab commands "Working with charts", change the appearance of the diagram, its structure, adding or deleting the elements of the diagram.

3. Editing a diagram.

All editing or formatting operations are performed with the selected chart or with its selected elements. By creating a chart, you can change any of its elements. There are a number of different ways to modify the base diagram created in Excel.

For most flat charts, you can change the type of the entire chart, to give it a completely different look, or choose another type of chart for any row of data to convert a diagram into a mixed (combined) one. For bubble and volume charts you can only change the type of the entire chart. To change the type of the entire chart, click on the chart to display the *Diagrams* tab.

1. In the *Designer* tab in the *Type* group, click the *Change Chart Type*.

2. In the *Change Chart Type* dialog box, select the desired type of diagram (Fig. 1.4).

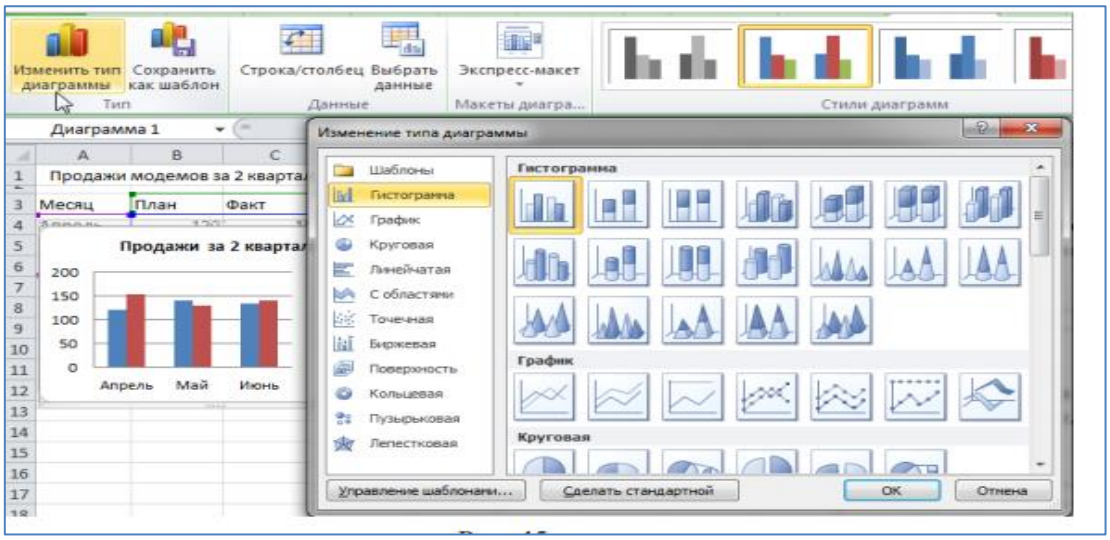


Fig. 1.4. **Changing the chart type**

You can change the type of the chart using the context menu (Fig. 1.5)

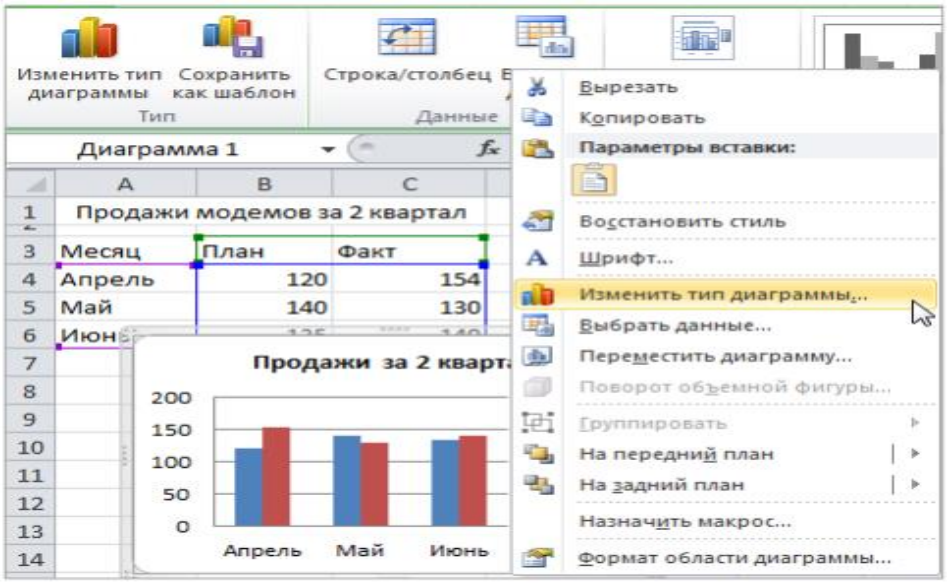


Fig. 1.5. **The *Change Chart Type* dialog box**

4. Combined diagrams.

A combined diagram is a diagram that consists of several data series and uses different types of charts simultaneously. In a combined diagram, one type, for example, a histogram, can be used, but the second axis of values is contained.

The easiest way to create a composite diagram is to change the chart type for one row.

To change the chart type for a row of data, select it.

1. On the *Designer* tab in the *Type* group, click the *Change Chart Type*. In the *Change Chart Type* dialog box, change the type for a number of data.

2. Double-click the highlighted row or right-click mouse to call the *Data Format* dialog box, go to the tab *Parameters of the row* and set the switch on the auxiliary axis (Fig. 1.6).

Fig. 1.6. **Changing the type of diagram for a series of data**

You can change the range of data represented in the diagram using the ribbon commands or the context menu (Fig. 1.7).

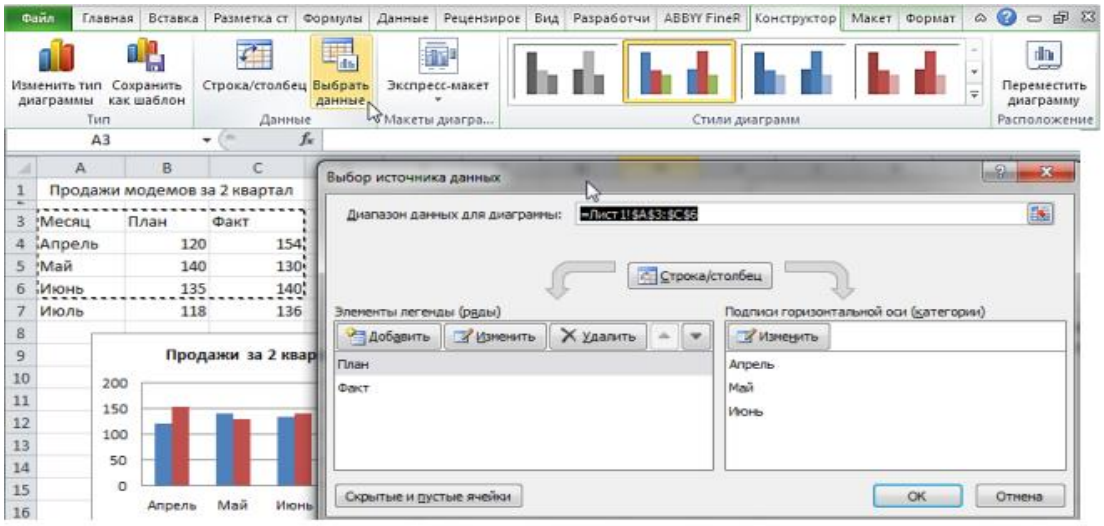


Fig. 1.7. **The context menu of the data source selection**

You can also change the name of a row and its sequence by using the *Data Selection* context menu (Fig. 1.8).

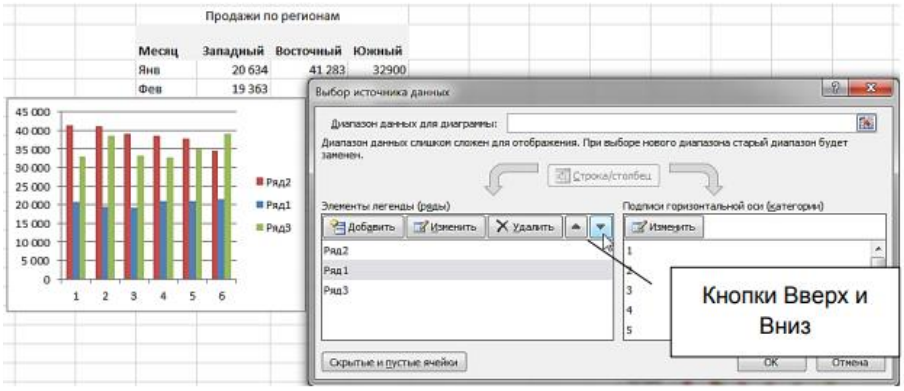


Fig. 1.8. **Work with the menu**

5. Adding and removing the chart elements.

Regardless of the selected layout of the chart, you can add and delete individual elements of it. To do this, use the elements of the *Layout / Work tabs* tab (Fig. 1.9).

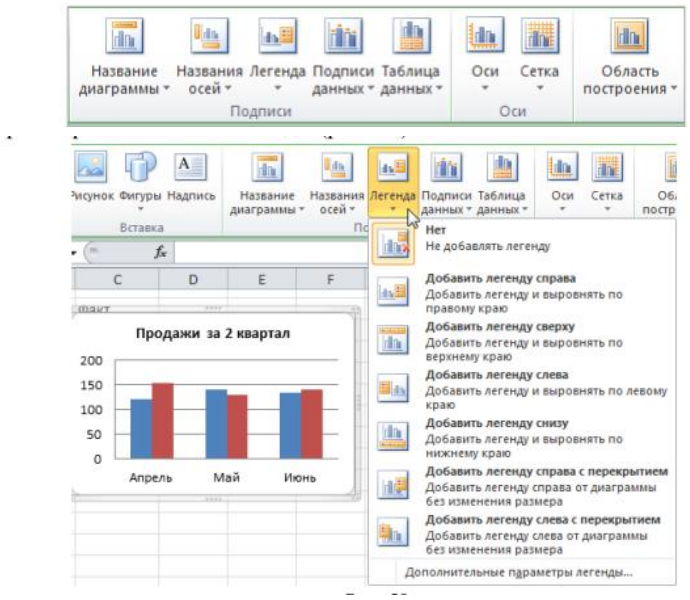


Fig. 1.9. **Changing of the diagram layout**

6. Application of styles.

Applying the styles for drawing a chart:

1. Highlight the chart.

2. On the *Designer* tab in the *Style sheets* group, select the style that you want to use (Fig. 1.10). You can also change the style of the figure of the selected element of the diagram.

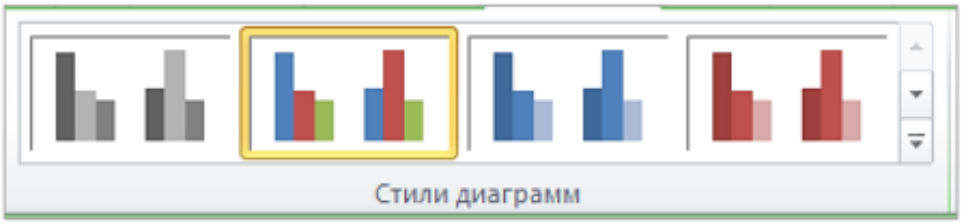


Fig. 1.10. **The *Style sheets* group**

7. Visualizing the data with a sparkline.

MS Excel has new tools for visualizing data: the diagram of the sparklines that allow you to create small diagrams inside a cell for graphical representation of the data range.

When working with large amounts of data it can be difficult to understand them. Using small diagrams called sparklines (infolines or infocells) and displayed next to the data, you can visually show the meaning of these values, as well as simplify the search for patterns and trends in them.

Sparklines are small diagrams that are placed inside individual cells on a sheet. Due to their size, the sparklines make it possible and very clearly visualize regularities in large data sets. Sparklines allow you to pay attention to such phenomena as economic cycles or seasonal growth or decline of any indicators, as well as highlight the maximum and minimum values in different colors. Sparklines are most effective when they are close to the data they represent.

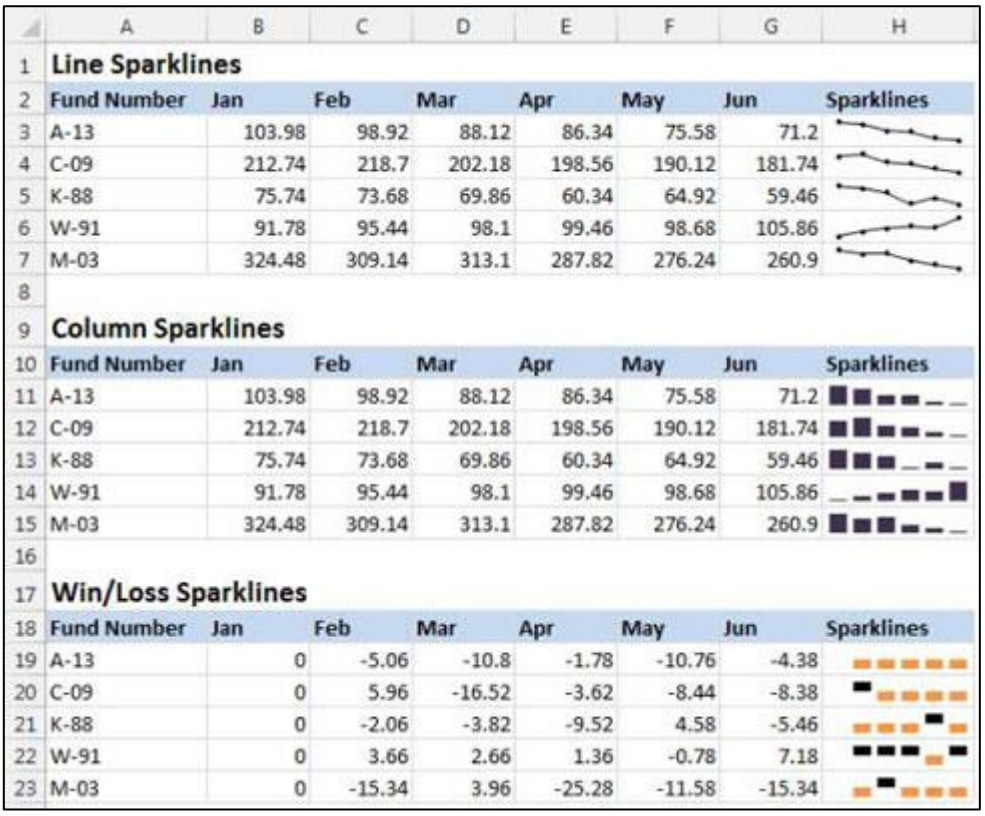
In MS Excel, three types of sparklines are created: a graph, a histogram (column), and a win/loss diagram (Fig. 1.11).

Fig. 1.11. **Types of sparklines**

Sparkline line (graphics). In the graphs, the line goes from left to right and displays changes in the data in the form of trends (Fig. 1.12). The points with the minimum values can be selected in a different color. Sparklines only allow you to quickly determine what is happening and do not give an answer to the question why it happens.

Sparkline column (histograms). Sparkline histograms are rows of vertical columns, the larger the value, the higher the column (Fig. 1.12). To simplify the perception, the maximum and minimum values can be highlighted in a different color.

Sparklines of win/loss. In sparklines of this type, the winning is indicated by a rectangle at the top of the cell, and the loss is indicated by a rectangle at the bottom (Fig. 1.12). This type of sparkline shows the positive values displayed by the rectangles at the top – these are the wins, the negative values correspond to the rectangles at the bottom – these are the losses. Losses are highlighted in red, so you can immediately see which values were unprofitable.

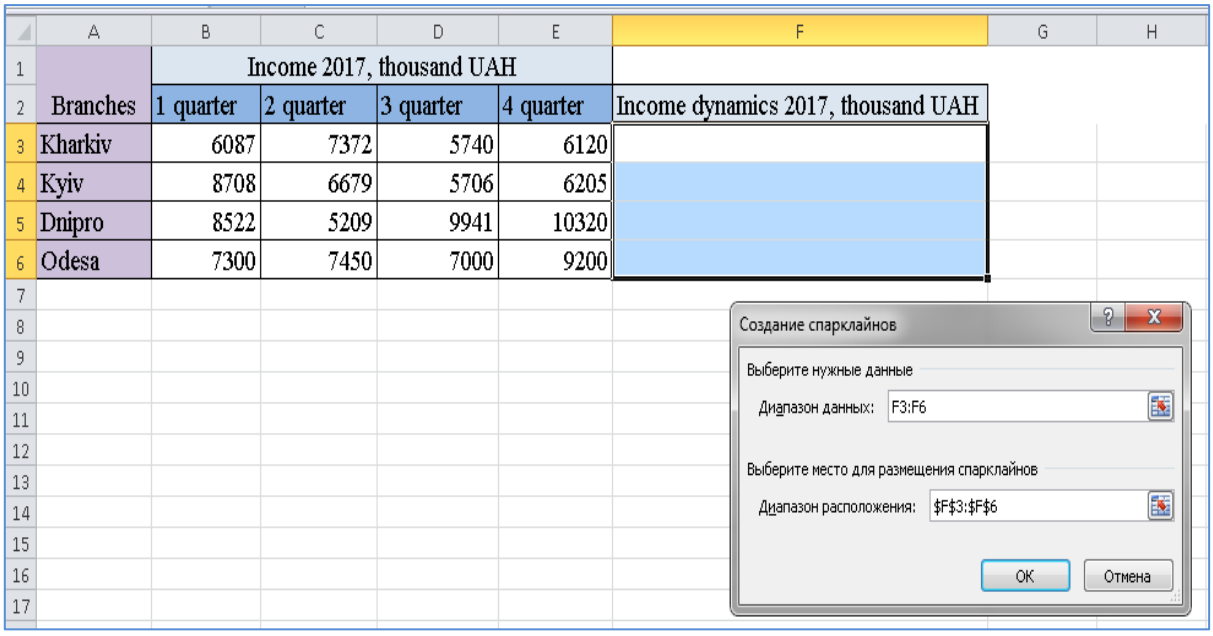


	A	B	C	D	E	F	G	H
1	**Line Sparklines**							
2	**Fund Number**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Sparklines**
3	A-13	103.98	98.92	88.12	86.34	75.58	71.2	
4	C-09	212.74	218.7	202.18	198.56	190.12	181.74	
5	K-88	75.74	73.68	69.86	60.34	64.92	59.46	
6	W-91	91.78	95.44	98.1	99.46	98.68	105.86	
7	M-03	324.48	309.14	313.1	287.82	276.24	260.9	
8								
9	**Column Sparklines**							
10	**Fund Number**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Sparklines**
11	A-13	103.98	98.92	88.12	86.34	75.58	71.2	
12	C-09	212.74	218.7	202.18	198.56	190.12	181.74	
13	K-88	75.74	73.68	69.86	60.34	64.92	59.46	
14	W-91	91.78	95.44	98.1	99.46	98.68	105.86	
15	M-03	324.48	309.14	313.1	287.82	276.24	260.9	
16								
17	**Win/Loss Sparklines**							
18	**Fund Number**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Sparklines**
19	A-13	0	-5.06	-10.8	-1.78	-10.76	-4.38	
20	C-09	0	5.96	-16.52	-3.62	-8.44	-8.38	
21	K-88	0	-2.06	-3.82	-9.52	4.58	-5.46	
22	W-91	0	3.66	2.66	1.36	-0.78	7.18	
23	M-03	0	-15.34	3.96	-25.28	-11.58	-15.34	

Fig. 1.12. **The types of sparklines**

7.1. Creating the sparklines.

To create sparklines, you need to select the data range that you want to analyze. On the *Insert* tab in the *Sparkline* group, select the sparkline type, and then specify the location for the sparklines (Fig. 1.13).



	A	B	C	D	E	F	G	H
1		Income 2017, thousand UAH						
2	Branches	1 quarter	2 quarter	3 quarter	4 quarter	Income dynamics 2017, thousand UAH		
3	Kharkiv	6087	7372	5740	6120			
4	Kyiv	8708	6679	5706	6205			
5	Dnipro	8522	5209	9941	10320			
6	Odesa	7300	7450	7000	9200			

Fig. 1.13. **The data range for sparklines and the space for sparklines**

Typically, sparklines are created based on a continuous series of cell values of one row or one column. The results of the construction of a sparkline are shown in Fig. 1.14.

	A	B	C	D	E	F
1		Income 2017, thousand UAH				
2	Branches	1 quarter	2 quarter	3 quarter	4 quarter	Income dynamics 2017, thousand UAH
3	Kharkiv	6087	7372	5740	6120	
4	Kyiv	8708	6679	5706	6205	
5	Dnipro	8522	5209	9941	10320	
6	Odesa	7300	7450	7000	9200	
7						

Fig. 1.14. **The sparkline line**

7.2. Setting up the sparklines.

To work with sparklines, use the contextual tab *Designer* which usually appears automatically when you select the cell in which the sparkline (infocurve) is located (Fig. 1.15).

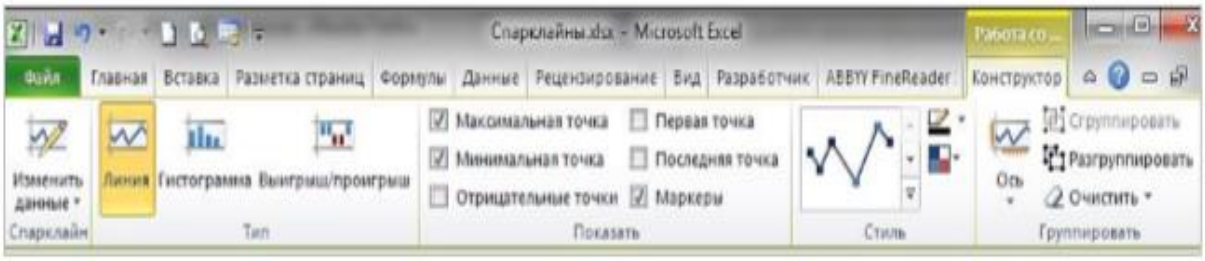


Fig. 1.15. **The *Designer* tab**

After creation, you can change the type of sparkline. To do this, in the *Type* group of the contextual tab *Designer*, click the button with another type of sparkline (Fig. 1.16).

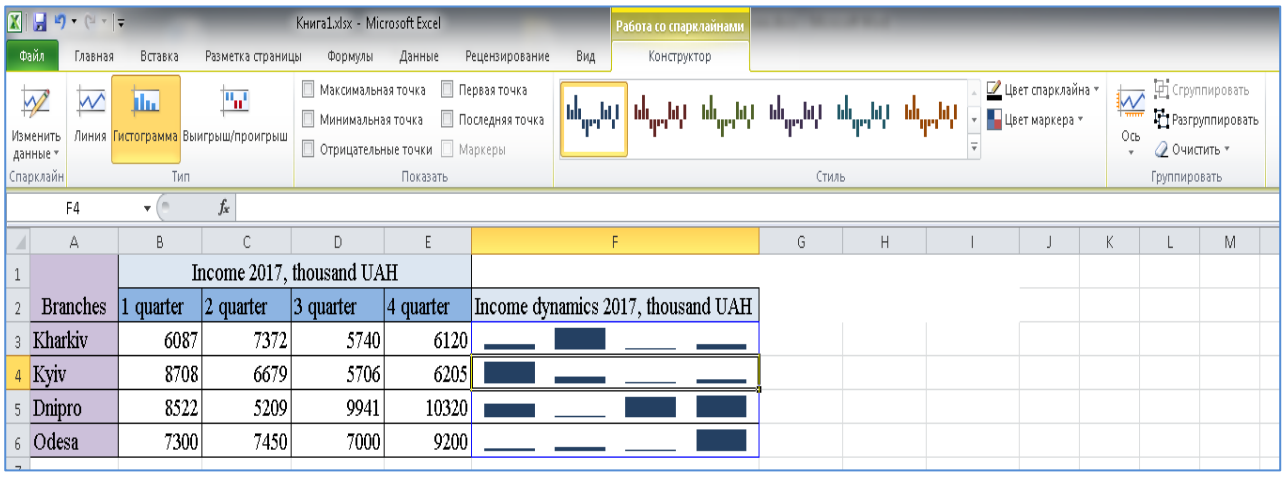


Fig. 1.16. **Changing the colour of a sparkline**

Sparklines located in adjacent cells are grouped by default. When you select one cell with a sparkline, the entire range is highlighted. In this case, all configuration and design actions will be performed with all sparklines in the range, even if only one cell is selected. To be able to work with each sparkline individually in the group, use the contextual tab *Constructor* and click *Ungroup* (Fig. 1.17). To remove the sparkline, select the cell or several cells with sparklines and in the group *Grouping* of the contextual tab *Constructor*, click the *Clear* button (Fig. 1.17).

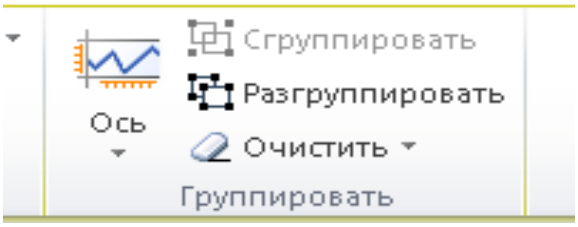


Fig. 1.17. **The *Constructor* tab**

7.3. Controlling the display of data points.

To select specific data in the *Design* tab, choose the *Display* group which controls the display of points on the sparkline. On the sparkline, you can select individual or all markers (values) of the data, specifying their visibility (Fig. 1.18). To enable all values, select the *Markers* check box.

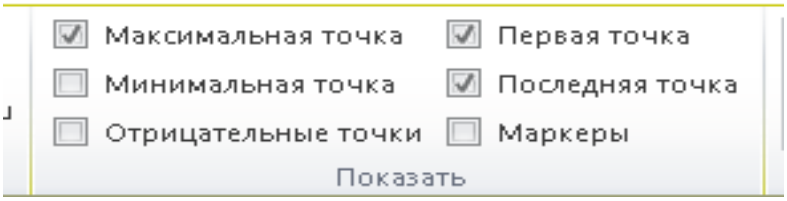


Fig. 1.18. **The *Markers* check box**

7.4. Display and settings of axis parameters.

By default, sparklines of all types do not have any axes. If there are negative values, you can display the horizontal axis. To do this, select the cell with the sparkline. In the *Constructor* contextual tab, in the group *Group*, click the *Axis* button and in the resulting menu, select *Show axis* (Fig. 1.19).

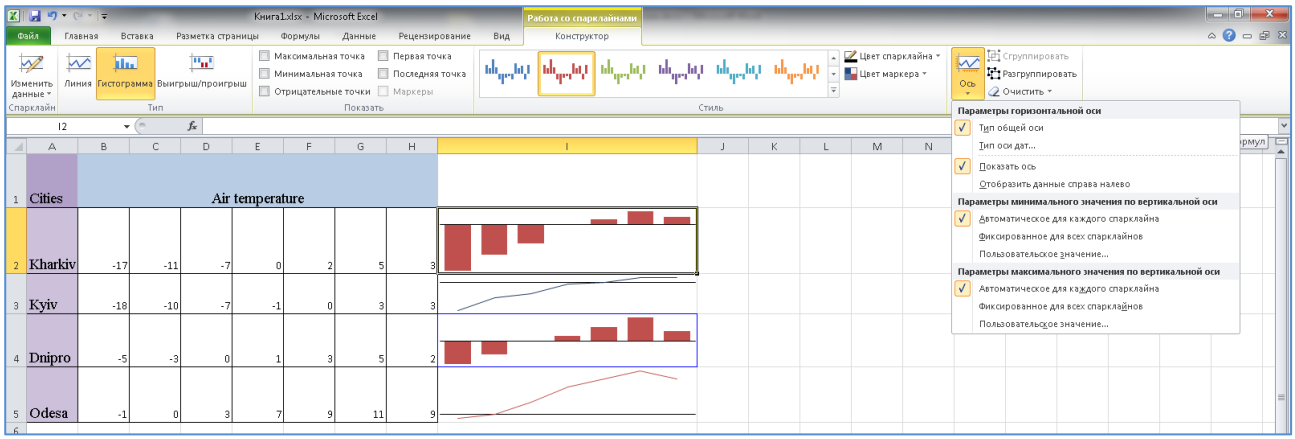


Fig. 1.19. **The display and settings of axis parameters**

7.5. Changing the style or format of the sparkline.

To change the style or format of the sparkline, a collection of styles is created on the *Constructor* tab (Fig. 1.20).

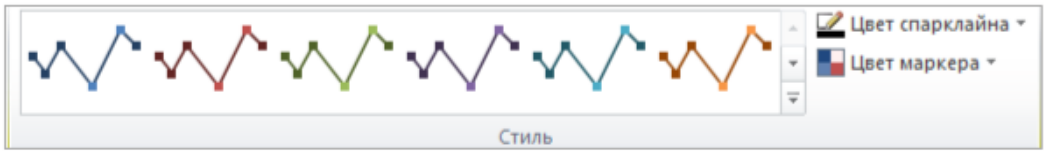


Fig. 1.20. **Changing the style or format of the sparkline**

On the *Constructor* tab, in the *Style* group, you can change the color, the thickness of the line on the graph, and control the color of the marker (Fig. 1.21).

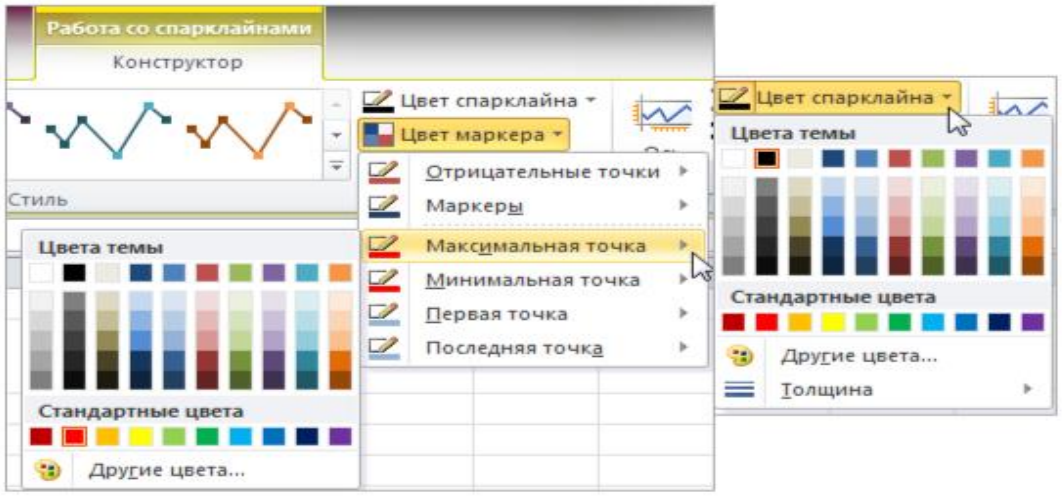


Fig. 1.21. **Changing the color of the marker**

Attention! Colors of lines (columns) can be changed regardless of the selected style. For the sparkline type *Histogram* or *Win/Loss*, it is not possible to change the thickness of the contour of the columns.

8. Visualizing the data with conditional formatting.

Conditional formatting provides visibility in the study and analysis of data, changing the format of the data representation in cells depending on the values. Conditional formatting can be applied to a range of cells, a Microsoft Excel spreadsheet, or a PivotTable report. Fig. 1.22 shows a worksheet with six bands, each of which is applied in a separate type of conditional formatting.

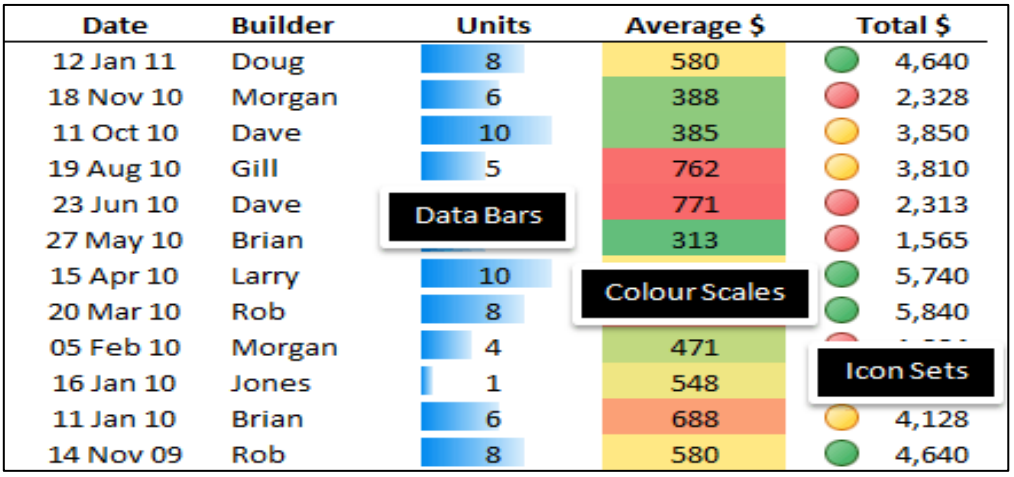


Fig. 1.22. **The types of conditional formatting**

8.1. Creating conditional formatting.

To apply conditional formatting to a cell or a range of cells, you must select one of the formatting rules on the *Home* tab in the *Styles* group in the *Conditional formatting* drop-down list (Fig. 1.23).

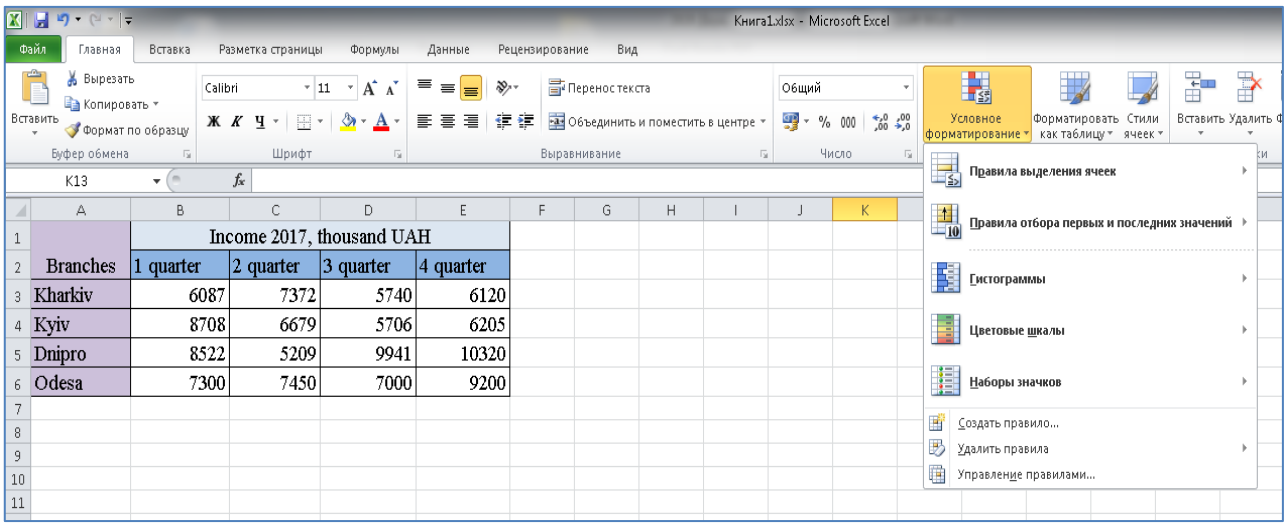


Fig. 1.23. **The *Conditional formatting* tab**

Here is a list of the rule groups for conditional formatting.

1. The rules for selection of cells. The rules of this group are based on comparing the contents of the cell with a given value. Among these rules there

are also rules that search for the occurrence of a given text line in the contents of the cell and check for duplicate values.

2. The rules for selecting the first and last values. The rules of this group distinguish the first or last 10 values, the first or last 10 % of the values, as well as values that are above or below the average value.

3. Histograms. The application of these rules results in the mapping directly into the cells of the horizontal columns, the value of which is proportional to the value contained in the cell.

4. Color scales. In accordance with these rules, the background of cells of the selected range will be a gradient fill, with the background hue of each cell depending on the value in this cell.

5. Sets of icons. The application of these rules results in the display in each cell of a graphic icon corresponding to the value in this cell.

6. Create a rule. This command allows you to create your own conditional formatting rules, including the rules based on logical formulas.

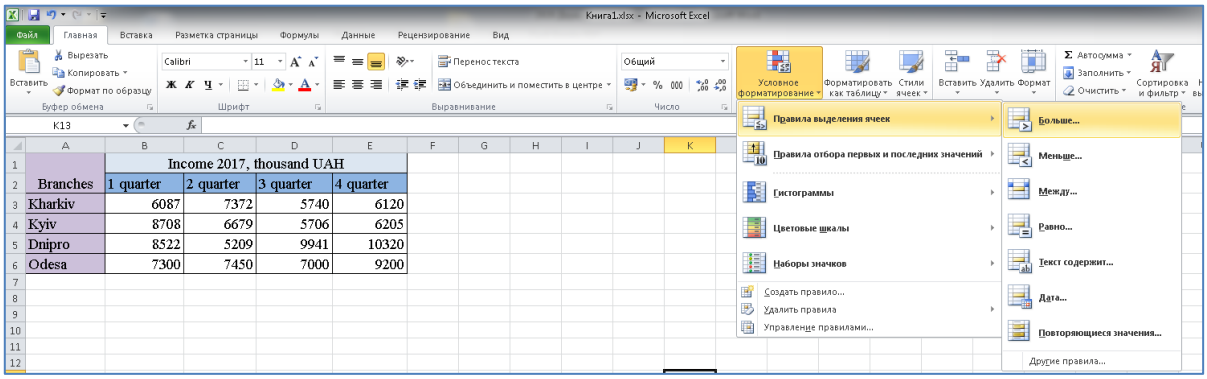
7. Delete the rules. This command deletes the conditional formatting rules from the selected range of cells or from the entire worksheet.

8. Managing the rules. This command opens the *Conditional Formatting Rules Manager* dialog box, where you can create new rules, edit the existing rules, or delete the selected rules.

8.2. Selecting the values.

With the help of conditional formatting, you can select values that correspond to any condition (more, less, between, equal, etc.). Select the cell or range of cells.

On the *Home* tab, in the *Styles* group, from the *Conditional formatting* drop-down list, select the *Cell selection rules* command, and then select the condition. The dialog box for setting the condition parameters opens (Fig. 1.24).



	A	B	C	D	E
1		Income 2017, thousand UAH			
2	Branches	1 quarter	2 quarter	3 quarter	4 quarter
3	Kharkiv	6087	7372	5740	6120
4	Kyiv	8708	6679	5706	6205
5	Dnipro	8522	5209	9941	10320
6	Odesa	7300	7450	7000	9200

Fig. 1.24. **Creating the conditional formatting**

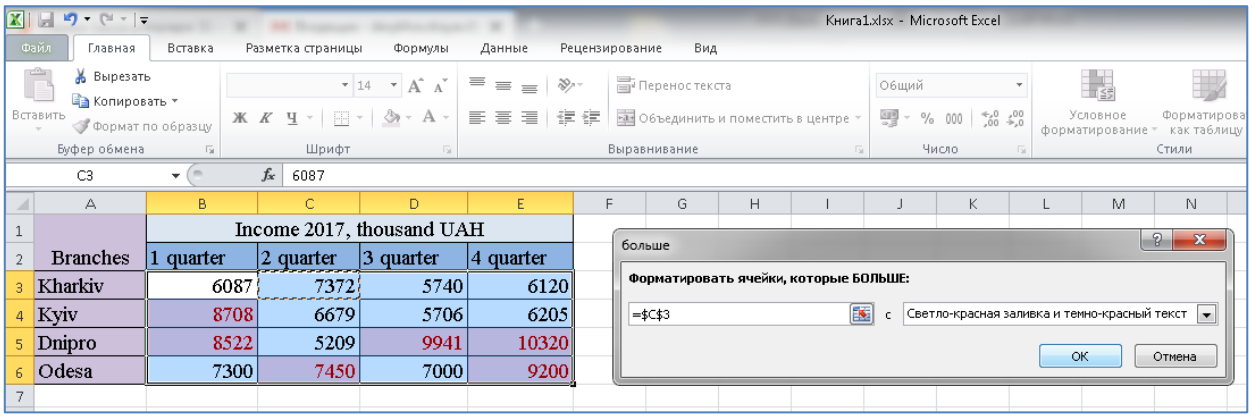


Fig. 1.24. (The end)

The name and the content of the settings window for the condition parameters depend on the selected condition.

8.3. Formatting based on the selection of values.

Using conditional formatting, you can find the maximum and minimum values in a range of cells based on the specified threshold value. Select the cell or range of cells. On the *Home* tab, in the *Styles* group, from the *Conditional formatting* drop-down list, select the *Rules* for selecting the first and last values, and then select the condition. The dialog box for setting the condition parameters opens (Fig. 1.25).

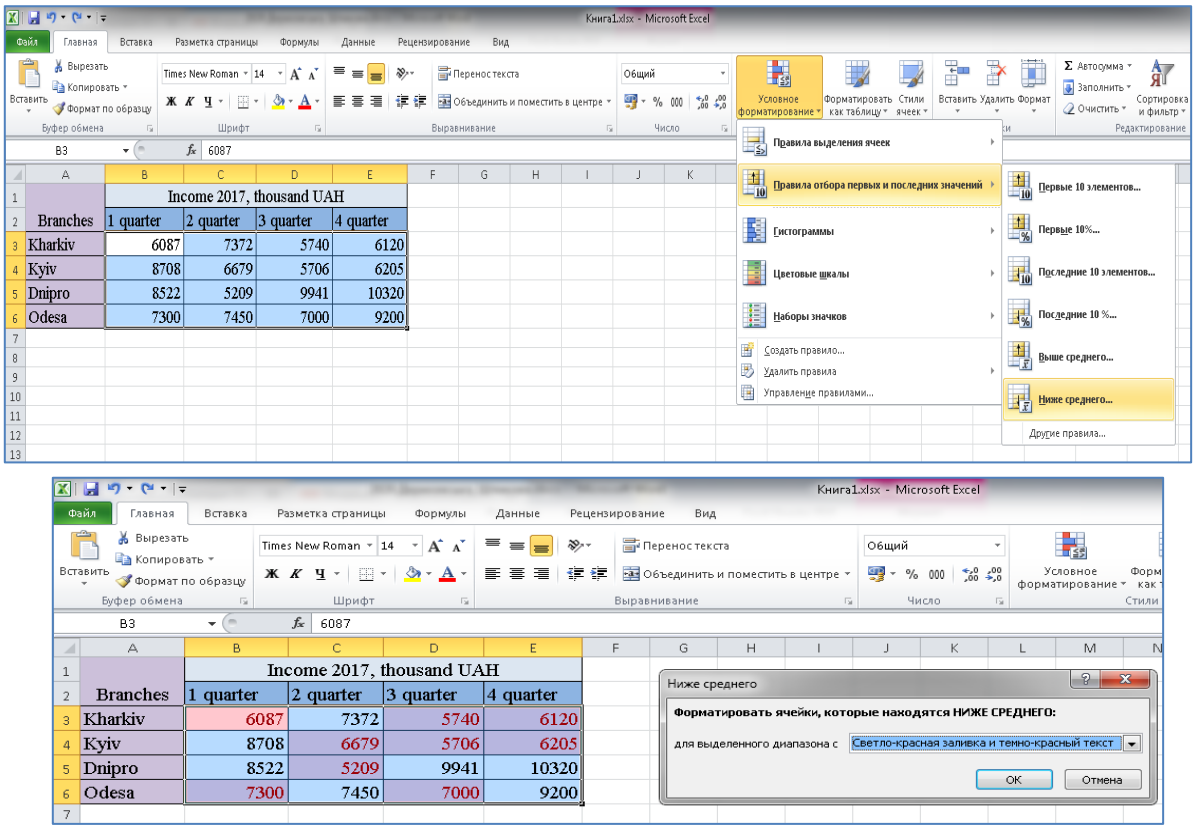


Fig. 1.25. **The box for setting the condition parameters**

8.4. Formatting using the histogram, color bar, icon set.

1. Histograms help to examine the value in a cell relative to other cells. The length of the histogram corresponds to the value in the cell. The longer it is, the greater the value. Histograms are optimal for determining the main indicators, especially in large amounts of data (Fig. 1.26).

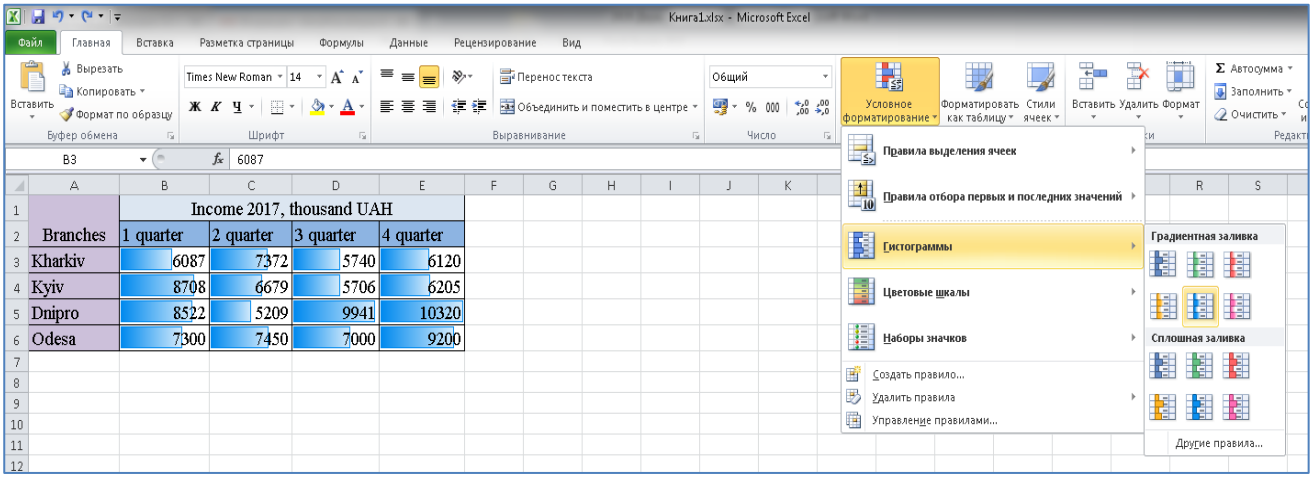


Fig. 1.26. **The result of creating the conditional formatting**

2. Color scales are visual elements that help to understand the distribution and dispersion of data. A three-color scale helps to compare the range of cells by using a gradation of three colors. The brightness level of the color corresponds to high, medium or low values. For example, in the red-yellow-green scale, you can specify that cells with high values will be green, cells with medium values will be yellow, and cells with low values will be red (Fig. 1.27).

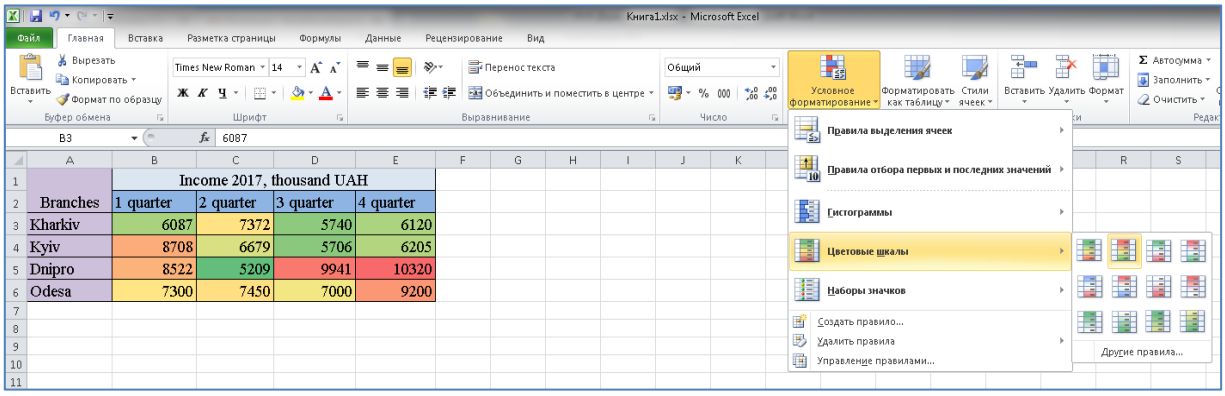
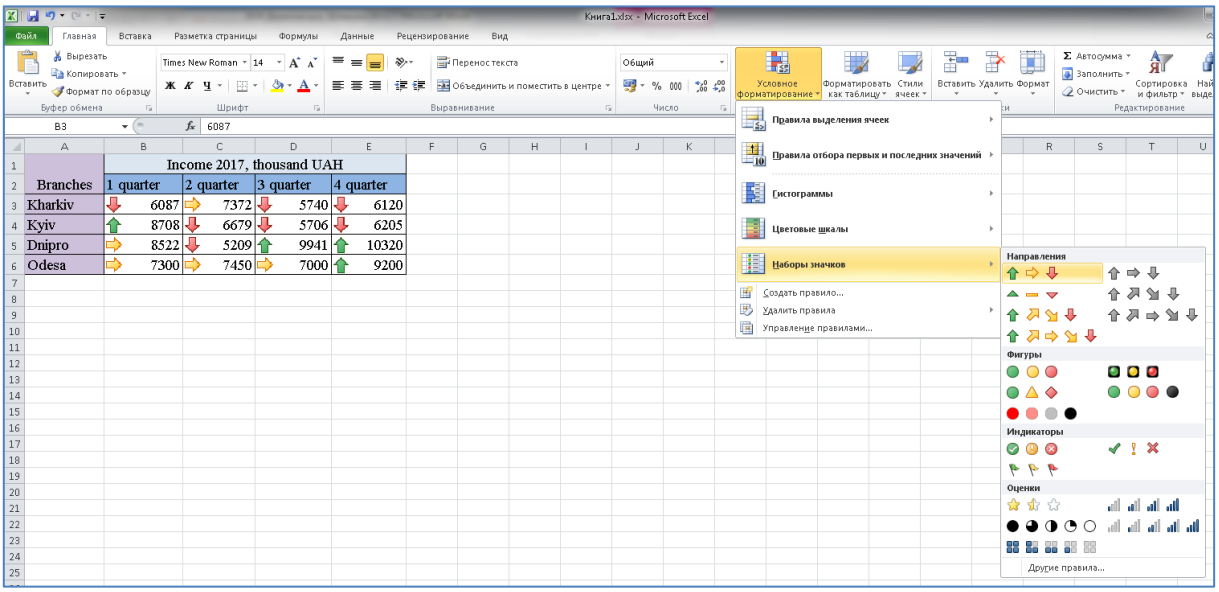


Fig. 1.27. **Changing the color**

3. The icon set is used to annotate and classify data in three to five categories, separated by a threshold value. Each icon corresponds to a range

of values. For example, in the set of icons 3, the red up arrow corresponds to high values, the yellow arrow pointing to the side corresponds to the average values, and the green down arrow corresponds to the low values (Fig. 1.28).



	A	B	C	D	E
1		Income 2017, thousand UAH			
2	Branches	1 quarter	2 quarter	3 quarter	4 quarter
3	Kharkiv	6087	7372	5740	6120
4	Kyiv	8708	6679	5706	6205
5	Dnipro	8522	5209	9941	10320
6	Odesa	7300	7450	7000	9200

Fig. 1.28. **The icon set**

8.5. Advanced formatting.

Conditional formatting rules can be created, modified and deleted. Select the cell or range of cells. On the *Home* tab, in the *Styles* group, from the *Conditional formatting* drop-down list, click *Manage rules*.

In the *Conditional Formatting Rules Manager* dialog box (Fig. 1.29), do one of the following:

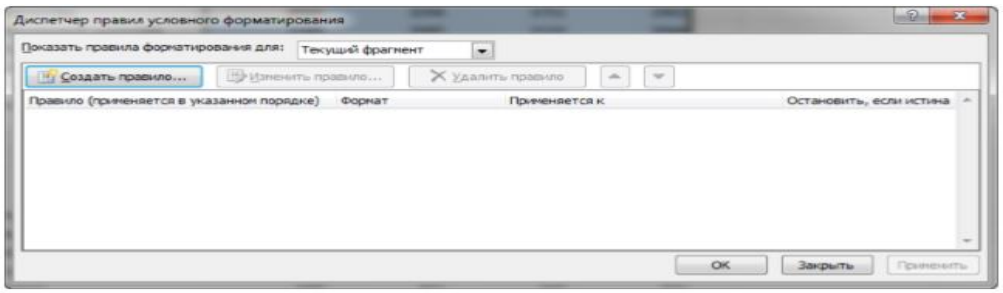


Fig. 1.29. **The *Conditional Formatting Rules Manager* dialog box**

To add a conditional formatting rule, click the *New rule* button. The *Create Formatting Rule* dialog box appears (Fig. 1.30), in which you modify the formatting options.

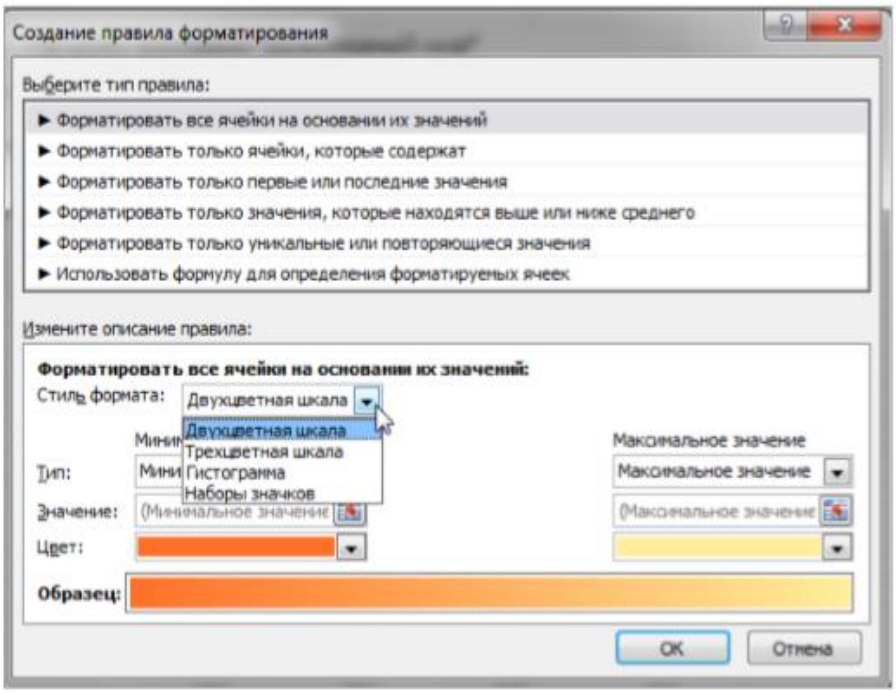


Fig. 1.30. **The *Create Formatting Rule* dialog box**

To change the *Conditional Formatting Rule*, select the rule, and then click *Edit Rule*. In the *Edit Formatting Rule* dialog box, you can select the type of another rule, change the condition and formatting parameters (Fig. 1.31).

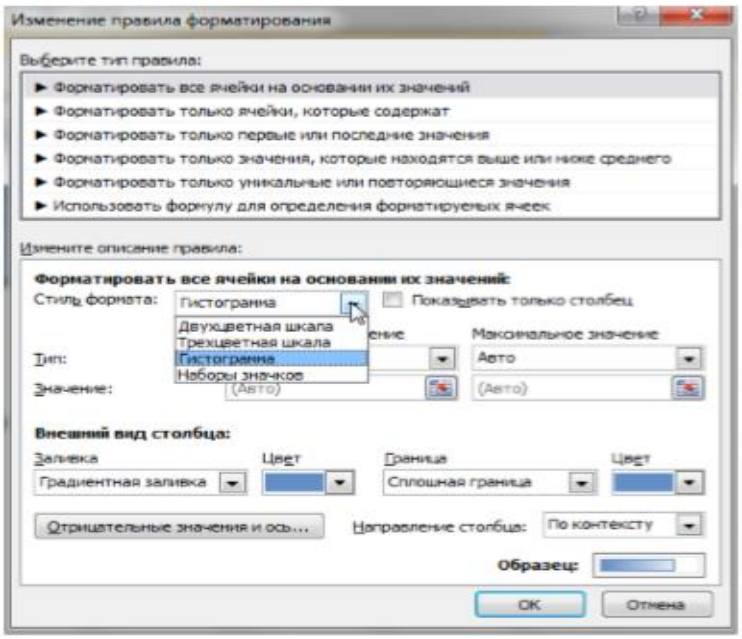


Fig. 1.31. **The *Edit Formatting Rule* dialog box**

If there are several rules for the same range, you can change the order in which they are applied by clicking the *Up* and *Down* buttons in the *Conditional Formatting Rules Manager* dialog box.

Tasks for independent work

Task 1. Create a table on Sheet 1, similar to Table 1.3, and enter your original data into it.

Table 1.3

Years	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
Average cost per month per household, UAH	3073.3	3458	3592.1	3820.3	4048.9	4952	5720.4
Food and non-alcoholic beverages	1585.8	1774	1799.6	1914	2101.4	2629.5	2848.8
Alcoholic beverages, tobacco products	104.49	117.57	125.72	133.71	137.66	163.42	165.89
Non-food products and services	1072.6	1224.1	1336.3	1398.2	1469.8	1807.5	2316.8
Nonconsumptive total costs	310.4	342.34	330.47	374.39	340.11	351.59	388.99

Copy the table to Sheet 1.

On Sheet 1 below the table, build a basic chart-like Graph (view – a graph with markers).

If necessary (to display all the inscriptions in the diagram completely and without distortion), change the size of the diagram.

Change the type of diagram to the *Histogram* for the "Total cost per month for one household" series (the view is a chart with accumulation).

Set the colors: for the histogram choose gradient, for graph lines: food and non-alcoholic beverages – green; alcoholic beverages, tobacco products – yellow;

non-food products and services – red; nonconsumptive aggregate expenses – serene.

Insert the chart name.

Format the values and categories axes (*Y* and *X*): change the alignment of the signatures on the category axis.

Compare the diagram you have constructed with the one shown in Fig. 1.32. If there is a discrepancy between them, make the necessary changes to your chart.

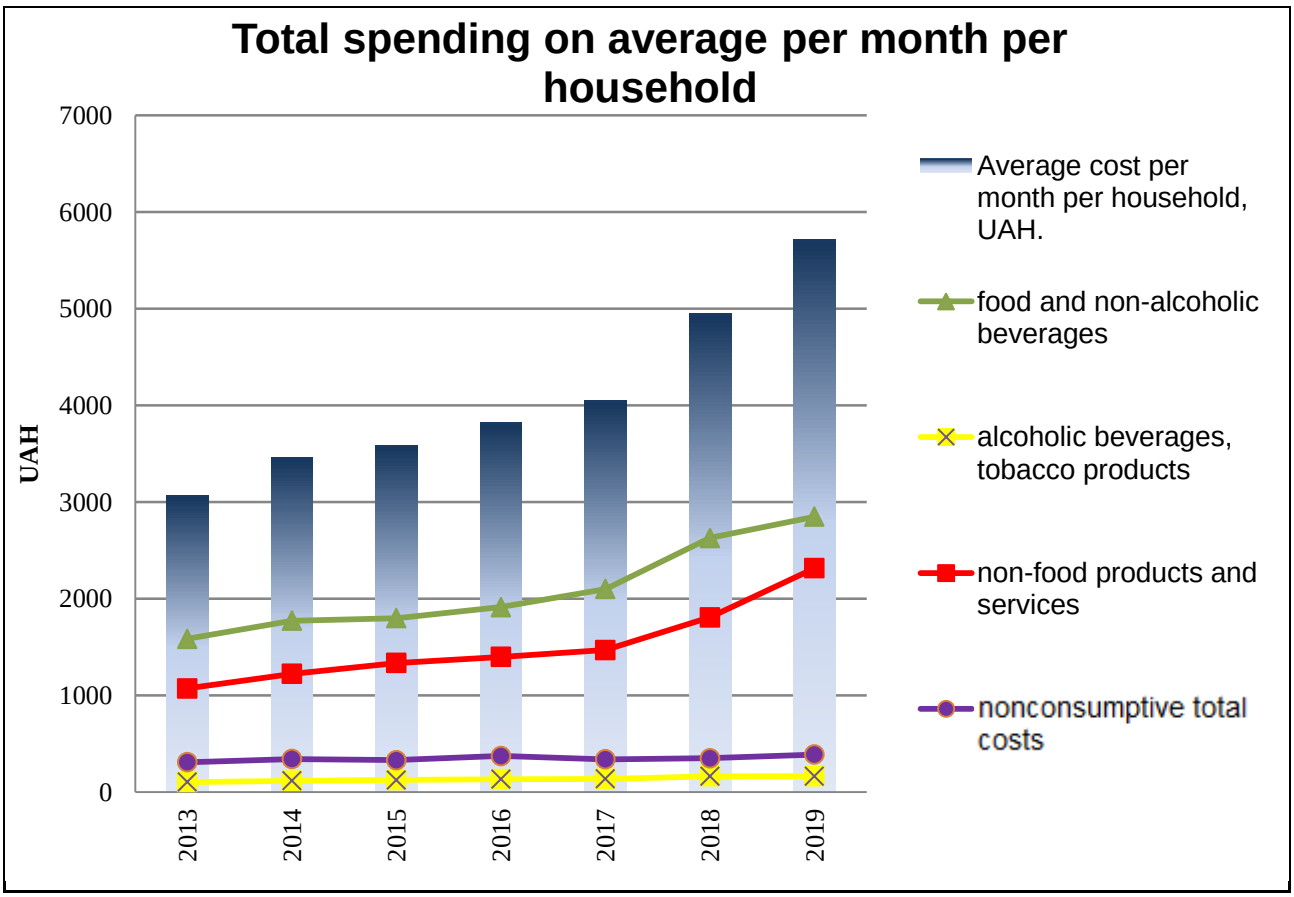


Fig. 1.32. **Total spending on average per month per household**

Task 2. Using the initial data of Table 1.3, create a volume circular diagram that shows the average monthly expenditure structure of one family in 2019.

Any color scheme can be chosen for the sectors of the circular diagram.

The obligatory elements of this diagram should be: the name, the subscription of the sectors with the indication of their percentage share.

The *Uncompromising* total costs sector should be located separately from other sectors.

To check the correctness of the work, compare the constructed diagram with Fig. 1.33.

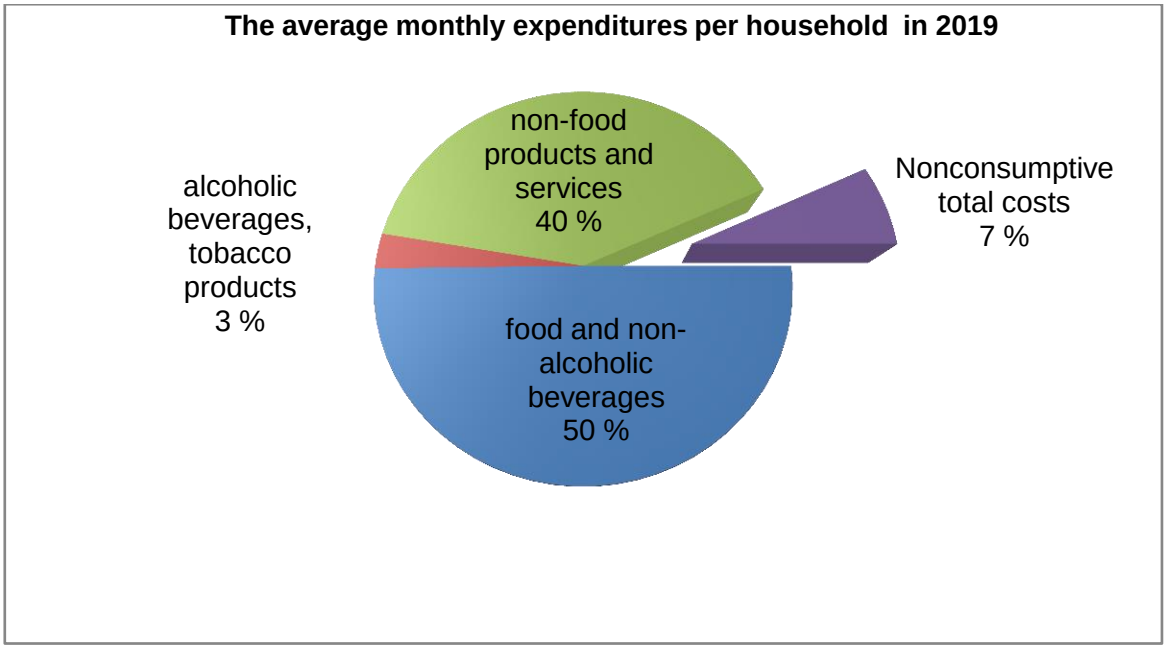


Fig. 1.33. **The average monthly expenditure structure of one family in 2019**

Task 3. Add the "Trend" column to Table 1.3 and build the following types of sparklines in this column: for food and non-alcoholic beverages – the *Sparkline Line*; for alcoholic beverages, tobacco products – the *Sparkline Column*; for non-food products and services – the *Sparkline Graph* and for nonconsumptive total costs – the *Sparkline Win/Loss*.

Change the row height and column width with sparklines to visualize the trends. Please, format the sparklines: mark the minimum and maximum values in the graph; on the histogram, select the minimum value in color.

Task 4. On the new sheet, according to Table 1.3 apply:

1) conditional formatting using histograms;

2) conditional formatting using the icon set.

The results of the assignments should be presented in the form of a report on the laboratory work.

The list of questions for independent work

1. What is information?
2. What is the classification of modern information?
3. What are the essential requirements for the presentation of digital information?
4. What place does information take in modern society?
5. What is the knowledge economy?
6. Describe the features of the knowledge economy and its place in the information society.
7. Disclose the essence of the main stages of data visualization.
8. What is grammar graphics?
9. Describe the relevance of data visualization for the knowledge economy.
10. What approaches to the interpretation of the term "visualization" are used?

Test questions for self-assessment on topic 1

1. What is the main task of visualization:
a) encoding the data using visual aids;
b) reducing the time and effort of the reader and the user to perceive information;
c) transformation of the dataset?

2. What is the first stage of visualizing the data:
a) refining;
b) acquiring;
c) interaction?

3. What elements are basic geometry in the grammar of graphics:
a) a point, a line;
b) an axis, coordinates;
c) gridlines?

4. What are the basic principles that support the reader's acquisition of information quickly and easily:
a) reducing the clutter;
b) integrating the text and the graph;

c) showing the data;
d) all of the above?

5. Pre-attentive processing allows the reader:
a) to perceive things serially;
b) to perceive multiple basic visual elements simultaneously.

6. What should be guided when choosing a graph type:
a) only features of the source information;
b) only how graphic objects are perceived;
c) the features of the source information and how graphic objects are perceived?

7. Which type of diagrams can be used to represent the structure of the phenomenon:
a) a scatter chart, sparklines;
b) a pie chart, a stacked bar chart;
c) a cartogram, a heat map?

8. What are the main causes of the distortion of information in the construction of bar charts:
a) the beginning of the vertical axis is not from zero;
b) non-compliance with the scale;
c) disconnected legend with the graph?

9. Which legend is more accessible to perception:
a) off to the right;
b) directly on the chart, or at the end of a line;
c) below the graph?

10. What technique can be used to visualize thinking:
a) a heat map;
b) a scatter plot;
c) a mind map?

Topic 2. Modern data visualization tools. Infodesign

Laboratory work 2
Visualization of textual information. A mental map (mind map)

The purpose of the work is to acquire the skills in building mental maps.

The task is to get acquainted with software products for building mental maps and to create a mental map of the chosen research topic.

Guidelines

The mental map is a technique for visualizing thinking with which you can better and more efficiently process some information.

A mental map is a way of fixing the thinking process, most similar to how thoughts and ideas are born and developed in our brain.

Mental maps have many different names (Fig. 2.1).

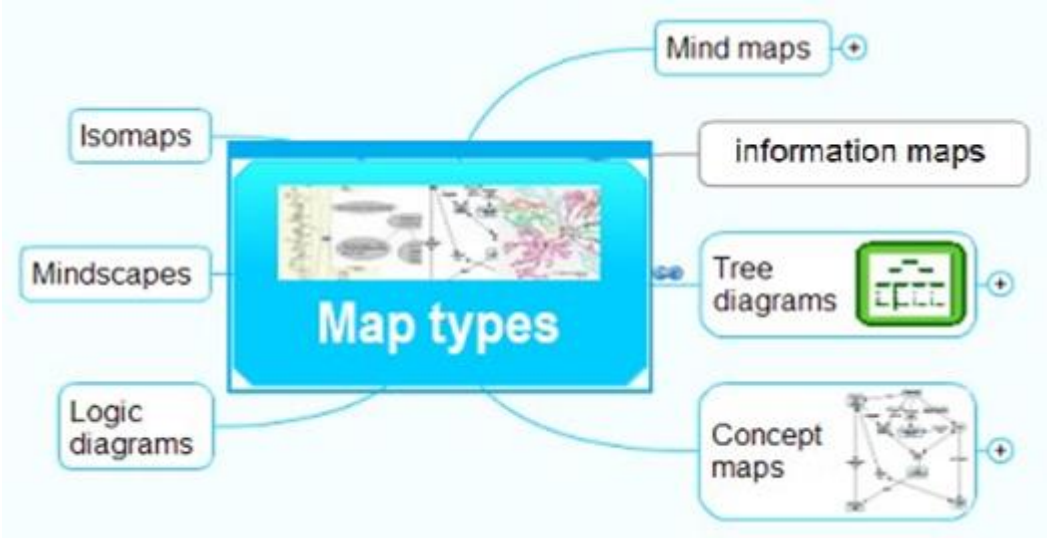


Fig. 2.1. **The mental map types**

A mind map is a diagram used to visually organize information. A mind map is hierarchical and shows relationships between the pieces of the whole.

The mindmapping technology is widely used for many purposes, including those shown in Fig. 2.2.

Fig. 2.2. **The functions of mind maps**

- *Fixing the information*. Mental maps are a convenient form of data recording, allowing you even with a large volume, to present it in a capacious form.
- *Memorizing the information*. When you record something in a convenient form, it is automatically fixed in the memory.
- *Easy access to information*. Data recorded in the form of mental maps is easy to remember, even at first sight.
- *Analysis of information*. A mental map is constructed in such a way that it allows you to see previously unnoticed connections between its parts, small unaccounted details, which can be very valuable in making decisions. Also, with the help of a mental map, you can view all the information in general, in a complex way, which generally contributes to understanding the topic and the data.

Mental maps help you easily use information, which makes them very useful in many areas of our lives: learning, work, personal life.

Why apply mental maps?

The main reasons are as follows.

The first reason lies in the very construction of the map – it is depicted in a radial (non-linear) form, the key image is in the center, and further branches diverge from it. This simplifies the flow of information to the brain – after all,

we also see the world around us as a whole – the central image and details around it (Fig. 2.3).

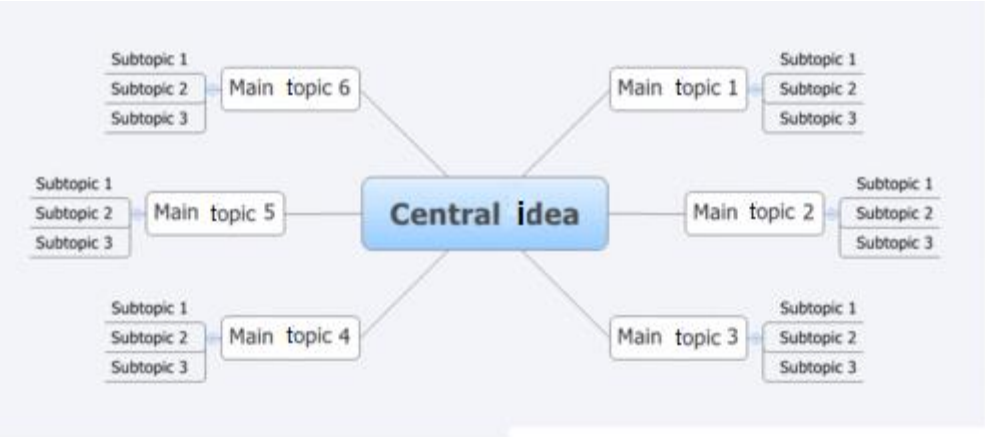


Fig. 2.3. **The structure of a mind map**

The second reason also lies in the processing of information by our brain – visualization as a whole improves perception. Words with pictures are remembered 6 times better than just words (Fig. 2.4).

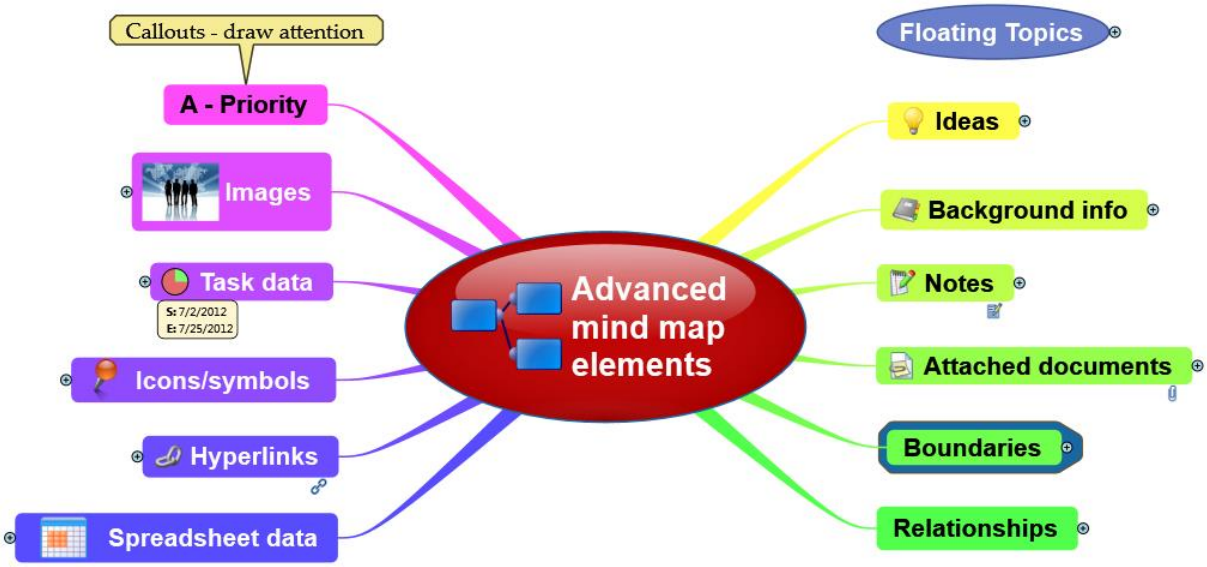


Fig. 2.4. **The main elements of a mind map**

When constructing a mental map, various abilities of our thinking are activated. When composing branches and keywords, we use hierarchies, for images – visualizations and associative thinking, in general, space-shaped thinking is used. All this activates the memory and allows you to remember both the data structure and their important aspects, so the use of mental maps improves the memorization of information by about 32 %.

Since we use different thinking abilities to construct mental maps, such as creativity, logic, imagination, they all develop and improve in the process.

Since mental maps display the picture as a whole, it allows you to establish all the relationships between the objects, even if they were initially not so obvious. In turn, this leads to the emergence of a new perspective of the information, as well as new ideas and thoughts. The structure and logic of the data become more "transparent", easy to understand and remember.

Another undeniable argument in favor of using mental maps is the fact that this technique is very easy to learn, and helps to memorize information at once, in large amounts and very effectively (Fig. 2.5).

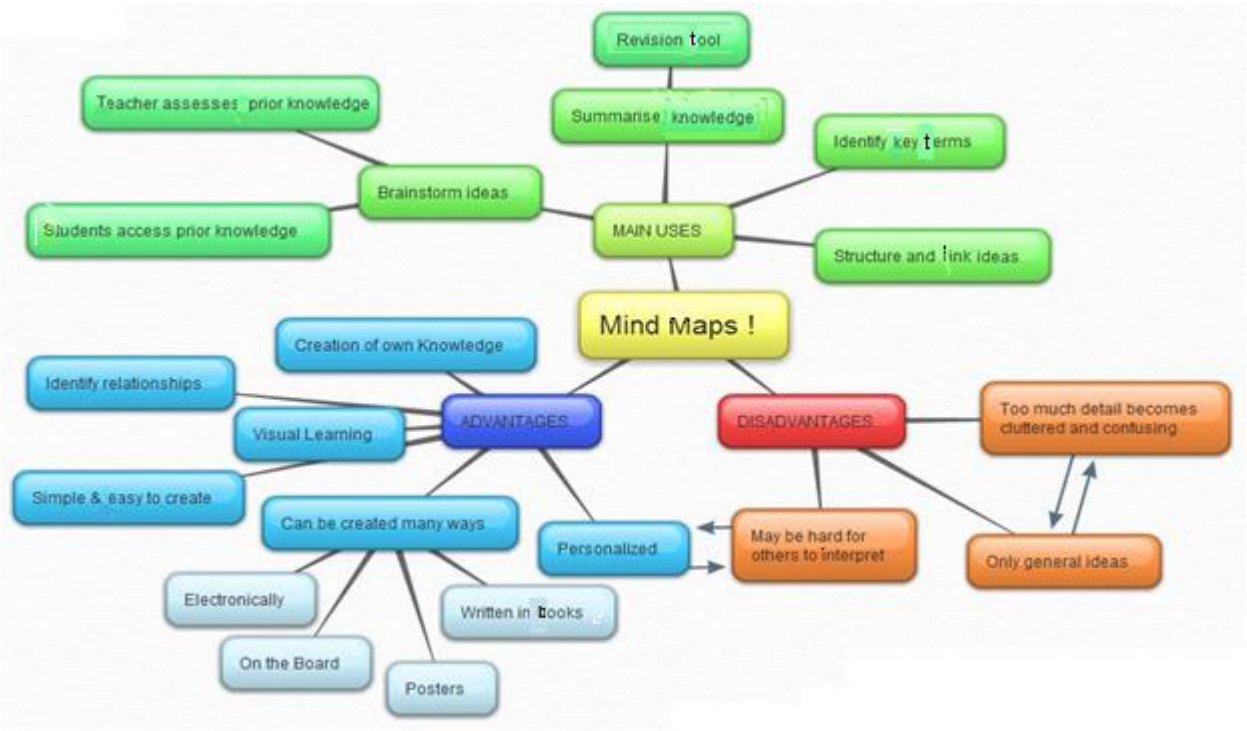
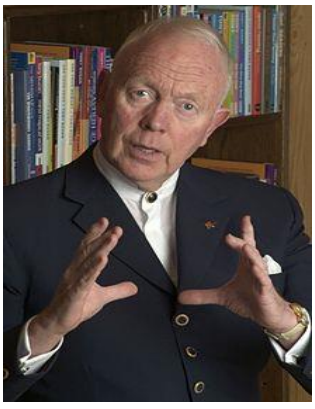


Fig. 2.5. **The main purposes of using a mind map**

Who invented mental maps?



Analogues of different connection diagrams have been used since quite early times, but the creator of the modern methodology of mental maps, such as we know it now, is Tony Buzan, an English psychologist, the author and co-author of more than 80 books, the most famous of which are "Superthinking" and "Use both sides of your brain". The first book, describing the technique of mental maps, "Use your head", was published in 1974.

Are there any programs for creating mental maps?

There are a large number of both online editors and programs/applications for building mental maps:

✓ Coggle is a simple and easy-to-learn online mental map editor [18].

✓ XMind is a serious software for building both mental maps and other various diagrams [26].

✓ Bubbl.us is an online editor and application [16].

✓ Mindmeister is a very nice online map editor with a lot of color themes, pictures, symbols and other features [24].

✓ Mapul is another representative of online editors [23].

How to create a mind map using the mindmeister application?

1. Any map begins with the main theme, an idea located in the center. Sometimes this can be called the essence of reflection. For example, let's make a mind map to learn the theme "Mental maps" (Fig. 2.6).

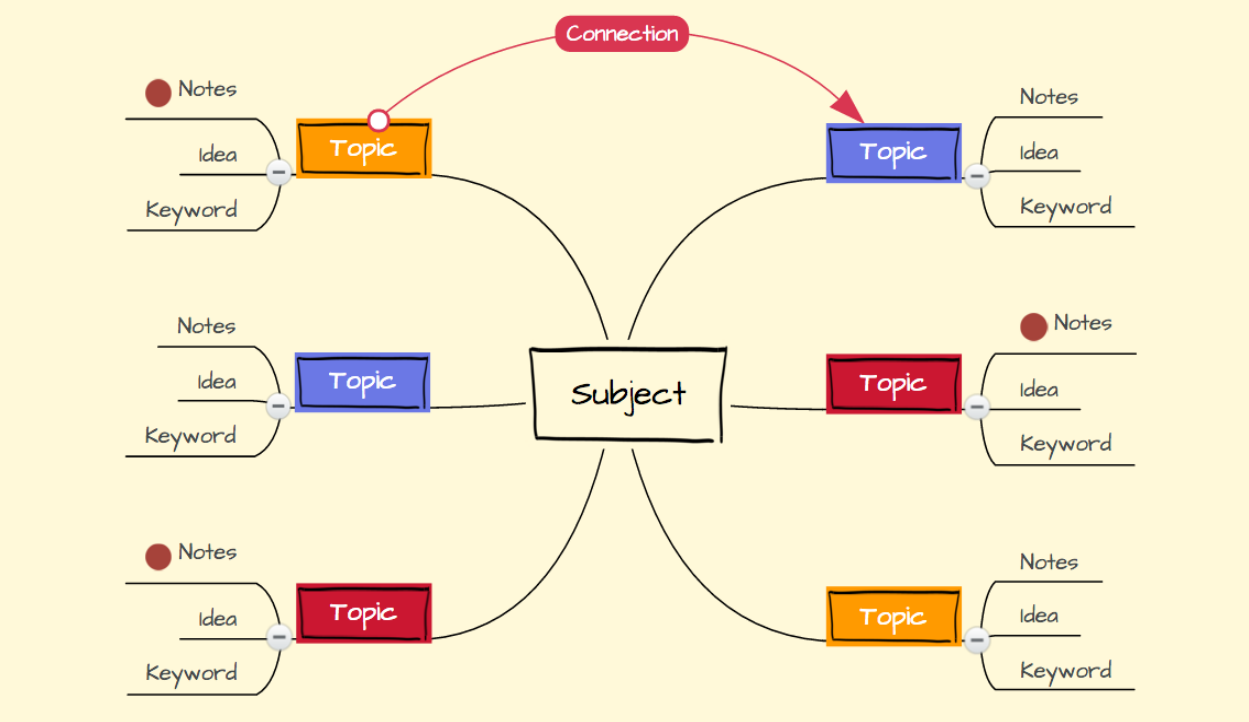


Fig. 2.6. **The composition of a mind map**

For this purpose, in the center, we place the key idea: to learn the theme "Mental maps" (Fig. 2.7).

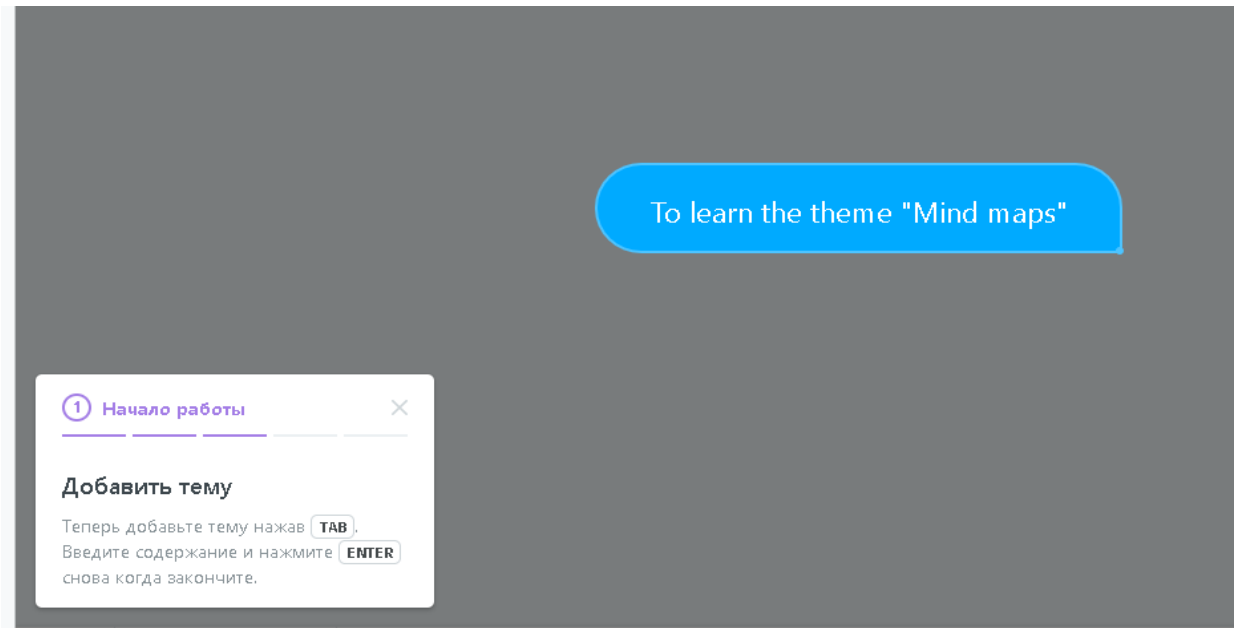


Fig. 2.7. **Getting started with a mental map**

2. Now add the first level themes. To do this, select the object from which the topic will go, and then press the tab key. Here you will see the key categories, steps and chapters related to the central question. In the process of brainstorming, you will begin to understand which ones are all the more important and deserve to go to the first level, and which are not so significant and can be attributed to some more global theme. Thus, the map hierarchy begins to be formed, which can be changed in the process (Fig. 2.8).

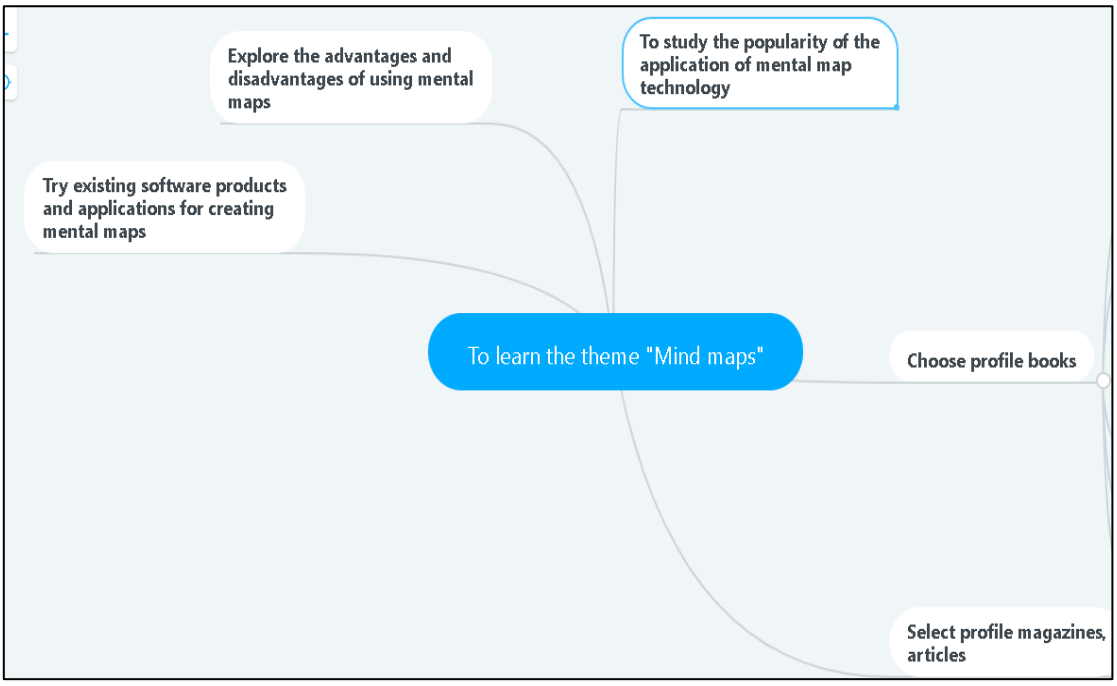


Fig. 2.8. **Adding a level to the mind map**

The first level themes have short names, after all, in fact, these are categories that help to build thoughts. Sometimes, if the categories are too abstract, you can attach images to them. This creates associations with content categories and allows you to quickly distribute ideas between them. According to generally accepted concepts, the image is always more powerful than the text name, but sometimes the pictures are used to a greater extent for presentations.

3. At the third stage, the themes of the second level are sub-themes. Thus a third level of the mental map is created. Subtopics are necessary for the concretization of ideas. This level should be more organized and thoughtful (Fig. 2.9).

Fig. 2.9. **Adding sublevels to the mind map**

If necessary, at this level, you can resort to descriptive phrases. Exit for the fourth level is undesirable, as this reduces the convenience of reading the map as a whole. If further concretization is required, it makes sense to go to the attached notes. If you need a detailed description, then, of course, there must be a note. If the map turns out to be too large, it is preferable to divide it into "floating" topics or create a new map, the reference to which is in the main map. Applications like MindMeister allow you to do this.

4. For applications, support of additional tools is preferable: notes, leaders (callouts) and connections (links).

❖ Notes. This is a great way to remind yourself about the end result of an idea. You can easily create a map of thoughts, which in a week will become incomprehensible even to you. Usually, notes are attached to topics as shortcuts with pop-up text. To see the contents, you need to hover the mouse or click on the note (Fig. 2.10).

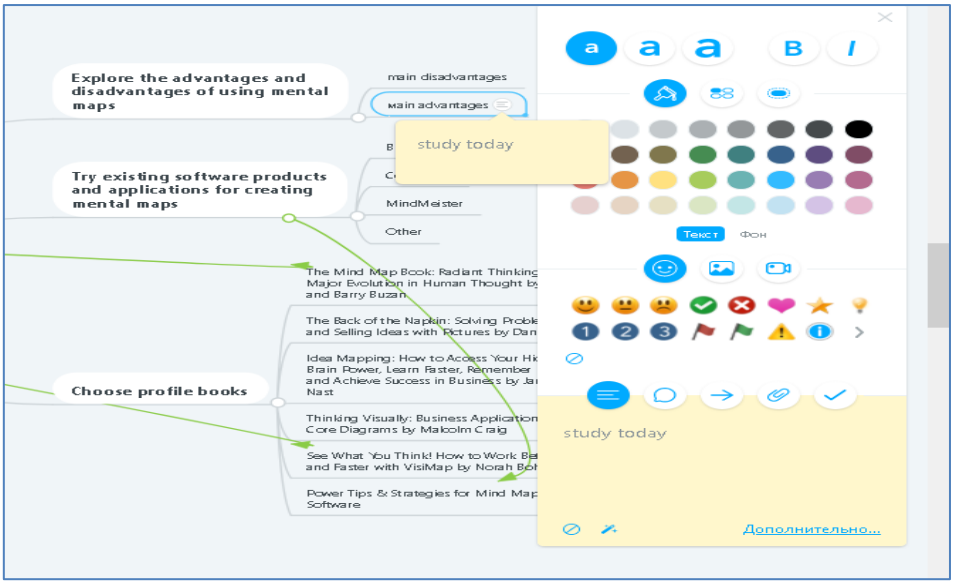


Fig. 2.10. **Adding notes**

❖ Leaders. These are short notes, which are usually highlighted in color, and their contents can be seen right away. Leaders can be used as pointers or explanations in cases where creating another level of the topic does not make sense (Fig. 2.11).

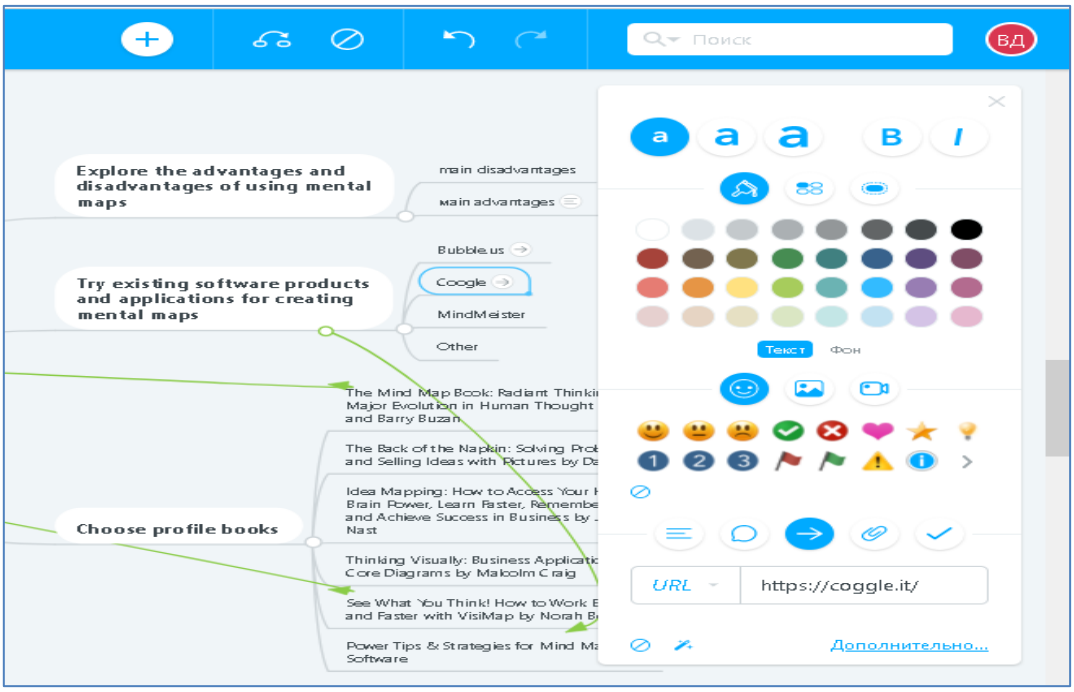


Fig. 2.11. **Adding links**

❖ Connections. These are pointers between the elements of the map. They can be used simply as reminders to oneself or associate different ideas, notes and callouts at different levels. So the map becomes more logical in the event that the connections between two or more elements of the map are not obvious or original. In addition, links make it possible to avoid duplication of topics (Fig. 2.12).

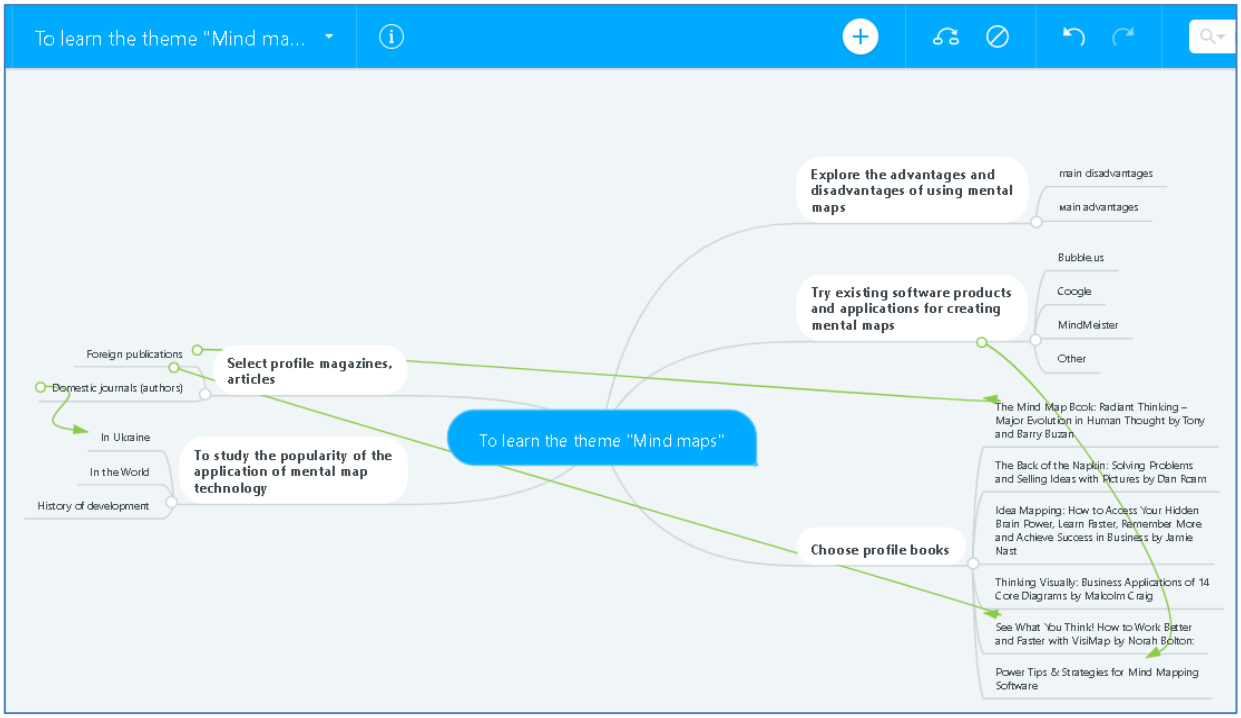


Fig. 2.12. **Adding connections between the levels and themes of the mental map**

5. Formatting mental maps. To format the mental map in general, use the left side panel and the top panel (Fig. 2.13).

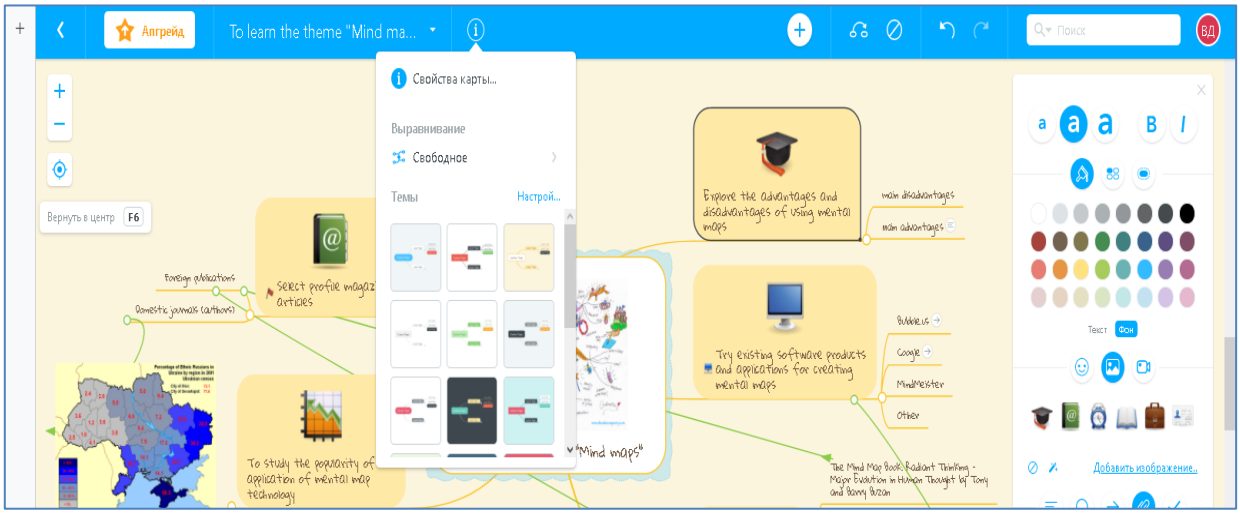


Fig. 2.13. **Special panels for formatting mental maps**

The +/− buttons increase or decrease the size of the map.

The button in the form of a target allows you to place the map in the center of the sheet.

The I button is required to select the properties of the map (the skin and alignment method).

The right sidebar allows you to change the font size, background color, add icons, pictures/videos, notes, links, upload files and much more (Fig. 2.14).

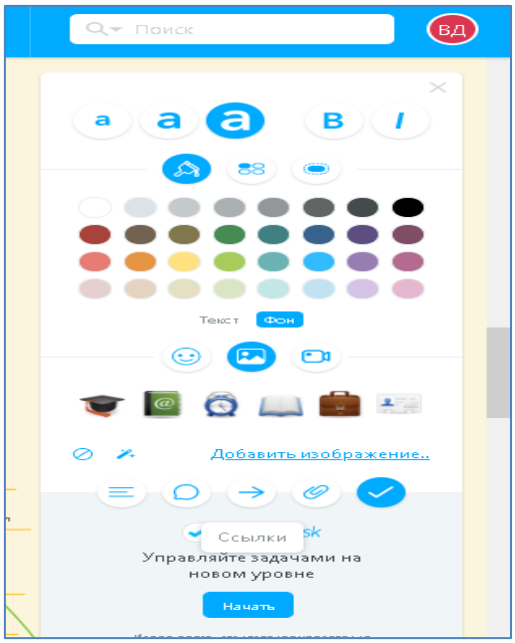


Fig. 2.14. **Formatting mental maps**

MindMeister is a program both for mindmapping and creating presentations (Fig. 2.15). Our online version allows you in seconds to turn a mental map into a presentation and play it right inside the map editor. Slideshows can also be exported, broadcast on the Internet or inserted.

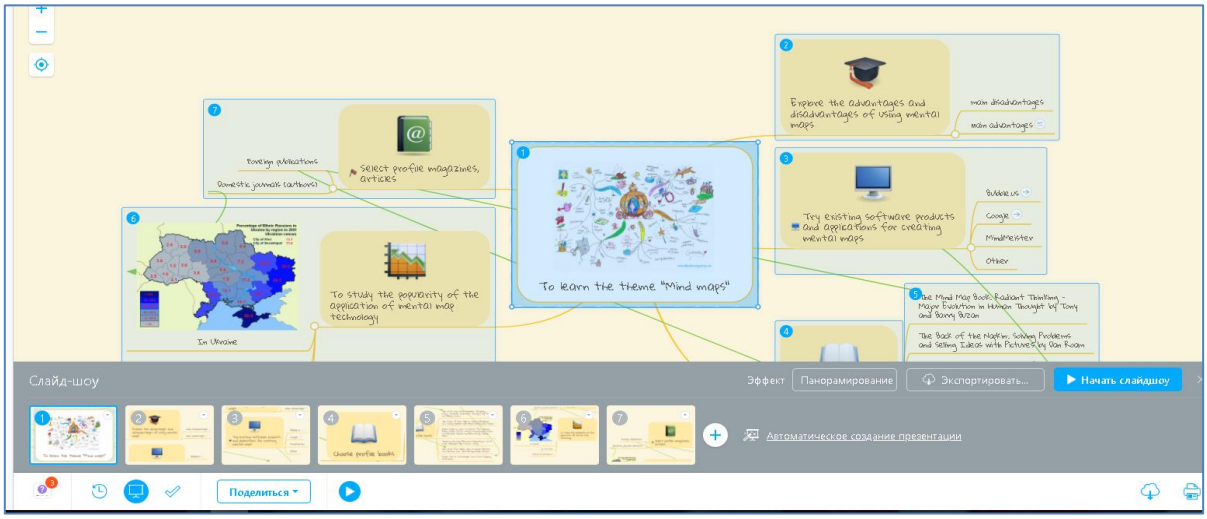


Fig. 2.15. **The presentation mode**

As a result of the use of special formatting panels, we receive a ready-made mental map (Fig. 2.16).

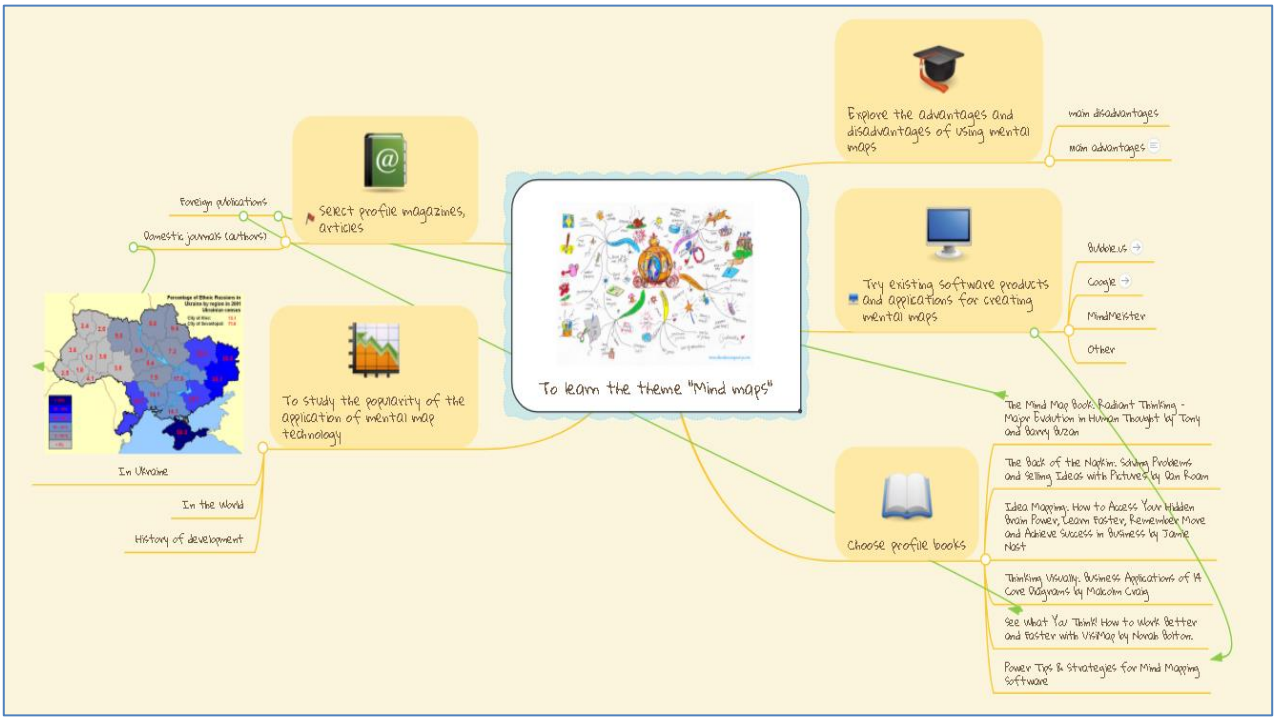


Fig. 2.16. **The results of constructing a mental map**

Advice. Do not forget about the basic elements of the correct construction of mental maps (Fig. 2.17).

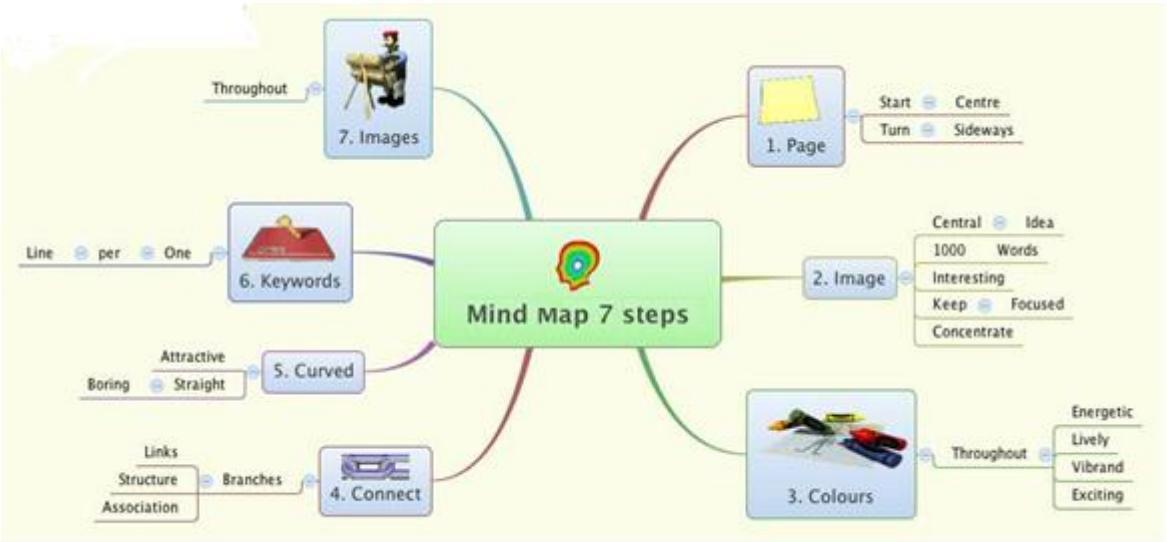


Fig. 2.17. **Some advice on constructing a mental map**

Tasks for independent work

Task 1. Create a mental map on the theme "Programming languages" using the application Bubbl.us and Coggle.it.

Task 2. Create a mental map for the key concept of your research topic.

The results of the assignments should be presented in the form of a report on the laboratory work.

The list of questions for independent work

1. What is visualization of information?
2. What is the object and subject of visual analytics?
3. In which areas is visualization of information used?
4. What are the basic stages of formation of information space for visualization?
5. Name the modern methods of visual analysis.
6. List the tools of each of the methods of business analytics.
7. What is info design?
8. What are the practical benefits and limitations of using the info design?
9. Describe the modern toolbars of business analytics.
10. Make a comparative analysis of two or three of the current tools for data visualization.

Test questions for self-assessment on topic 2

1. What data visualization tools do not require writing a code:
a) Plotly;
b) Power BI;
c) Highcharts;
d) Qlikview?

2. The visualization library for Python is:
a) dygraphs;
b) ggplot;
c) Bokeh.

3. The visualization library for R is:
a) ggplot2;
b) Bokeh;
c) D3.

4. The visualization library for R ggplot2 supports:

a) 3-dementional graphics;

b) only 2-dementional graphics.

5. What kind of attributes are used as quantitative:

a) added marks and orientation;

b) shape;

c) length and 2-dimensional location?

6. Nonquantitative perception is:

a) enclosure;

b) size;

c) intensity.

7. One of the laws of perception relating to grouping is:

a) orientation;

b) proximity;

c) blur.

8. In what spheres is information design used:

a) only in applied linguistics;

b) only in applied ergonomics;

c) only in applied psychology;

d) in all spheres named above and others?

9. The function of typography is:

a) to identify, investigate and offer solutions to language-related real-life problems;

b) to make written language legible, readable, and appealing;

c) to provide selection of pre-attentive visual attributes.

10. What are the recommendations if you need to use a large number of colors in the visualization?

a) Use a wide range in both hue and brightness.

b) Use natural patterns of color.

c) Use a static set of colors.

Content module 2
Using modern methods of analytical data processing

Topic 3. Planning, collecting and preparing data for visualization. The methods for collecting analytical information

Laboratory work 3
Infographics: cartography (mapping)

The purpose of the work is to learn the skills in visualization of data with the help of mapping tools.

The task is to get acquainted with software products for cartography and to create a map of the chosen research topic.

Guidelines

A map is one of the most common methods for visualization of data. Since school we have been used to maps that indicate the geographical location of countries and continents, landscapes of various territories. The main purpose of maps is to clearly explain any socio-economic, political, cultural and demographic processes and manifestations.

Maps consist of objects, comparing which we can understand the structure of phenomena and processes, the differences between them and understand the overall picture, in a word, explain the world.

Modern journalism, science and research actively use infographic tools, moreover, all world organizations and funds use cartography as one of the interactive tools for visualizing data in the preparation of reports.

In support of this fact, you are offered a selection of 7 most interesting visual maps prepared by different sources.

In 2013, the World Economic Forum presented the results of a survey of residents of 140 countries (Fig. 3.1). Respondents expressed their attitude to foreign visitors – so the organization managed to draw a map of the most and least friendly states. The most hospitable countries were Iceland, New Zealand and Morocco. However, see for yourself.

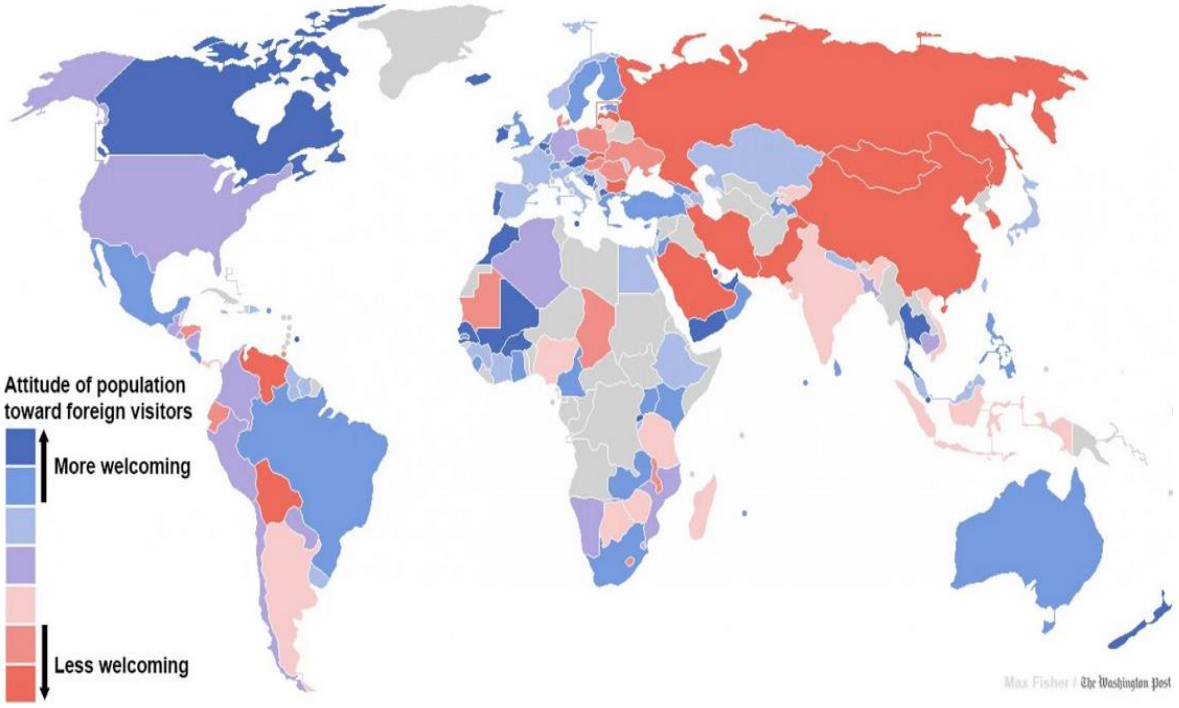


Fig. 3.1. **The most hospitable countries**

The map in Fig. 3.2 is a reminder that linguistic and written traditions do not always coincide with state borders.

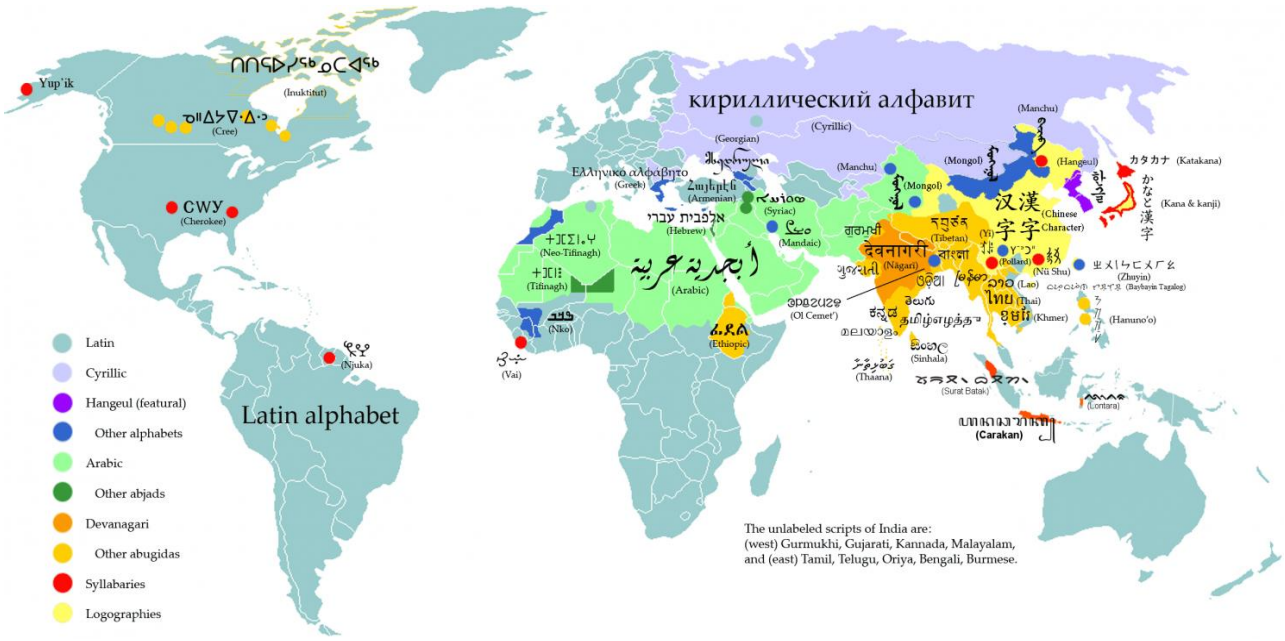


Fig. 3.2. **The basic written systems of the world**

This visualization also reflects many historical processes: European colonialism, the Arab conquests of the 7th century, the Russian expansion of the 19th and 20th centuries, and the (still ongoing) unification of India and China.

Who knows what love is?

Between 2006 and 2007, the Gallup company (formerly the American Institute of Public Opinion) surveyed the residents in 136 countries. The respondents were asked to answer the question: "Did you feel being loved yesterday during the day?" So, it was possible to map the countries where people are most and least loved (Fig. 3.3).

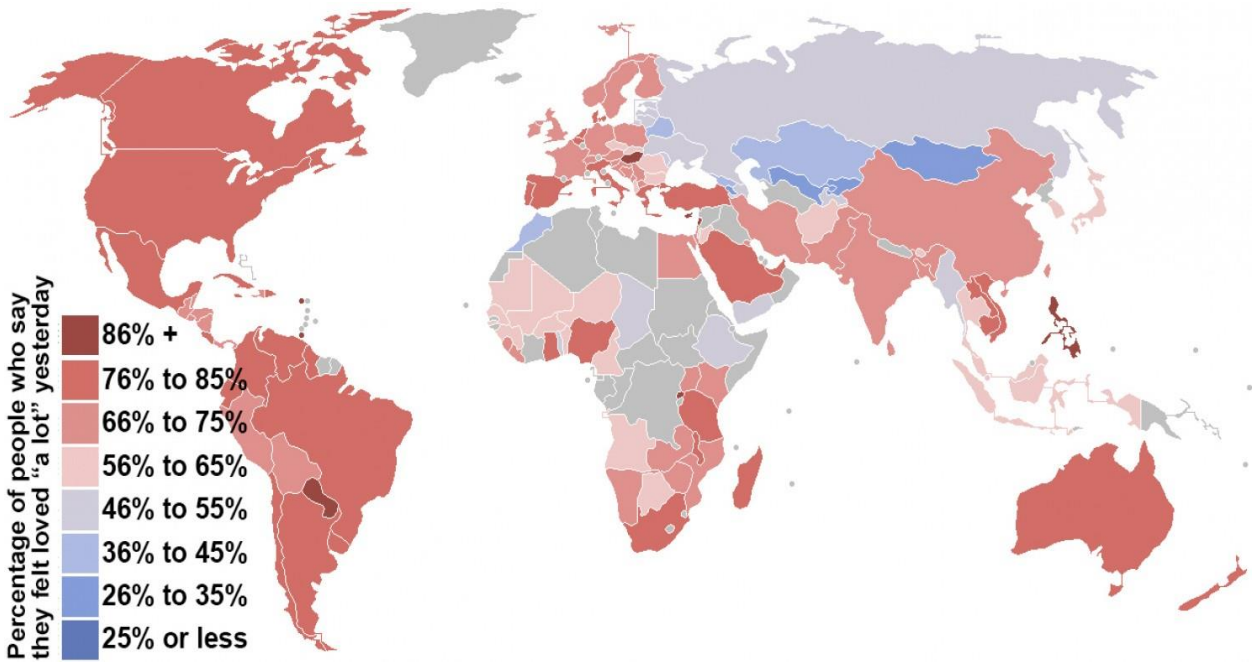


Fig. 3.3. **Who knows what love is?**

Nobody in the world smokes so much as residents of Eastern European countries. This is evidenced by the results of recent studies. In these countries, for each person in the year, there are about 2000 smoked cigarettes (Fig. 3.4).

The most tobacco-dependent people are Serbs. Russians are also in the top five "leaders". But the Romanians made an exception among the peoples of Eastern Europe.

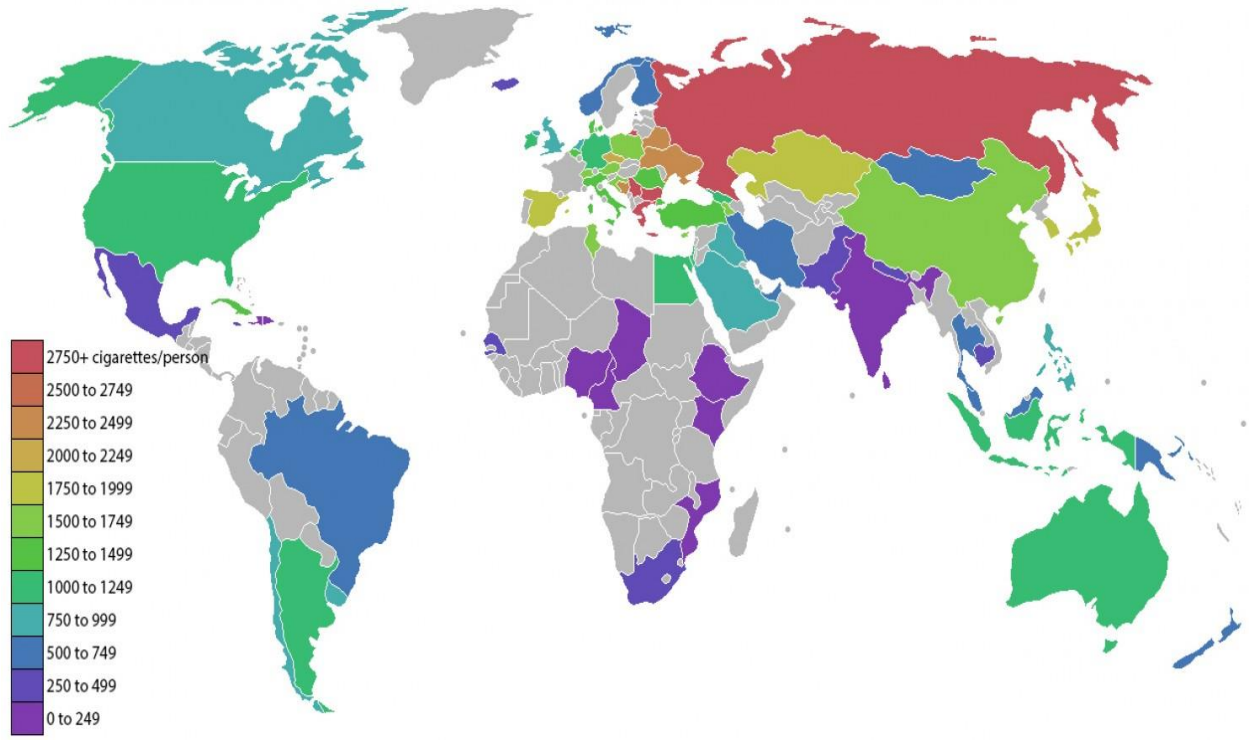


Fig. 3.4. **The tobacco-dependent peoples**

Few people will be surprised at the fact that atheism is common in Japan, China or European countries. However, for you, for sure, it will be a surprise that this is not uncommon in some Arab states, like Saudi Arabia, where it is considered a crime and is prosecuted by law (Fig. 3.5).

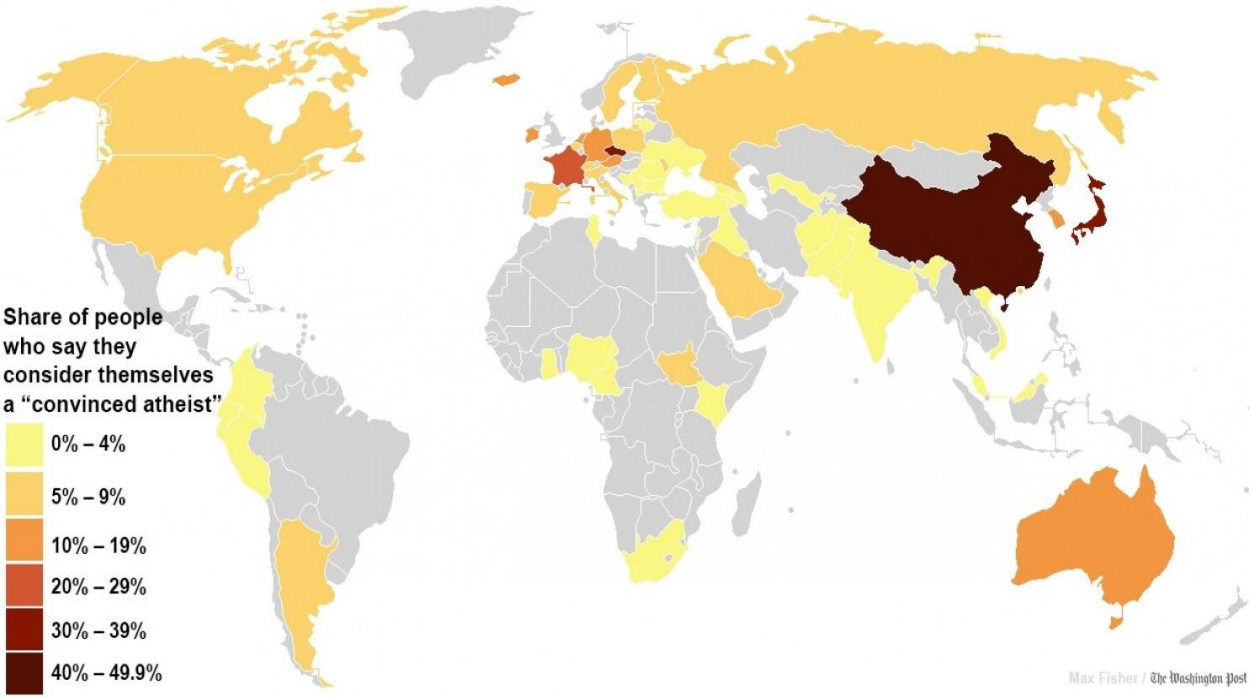


Fig. 3.5. **Where do atheists live?**

Recently, the London-based non-governmental organization "Save the Children" conducted a large-scale study, setting the conditions for motherhood in virtually every country in the world (Fig. 3.6).

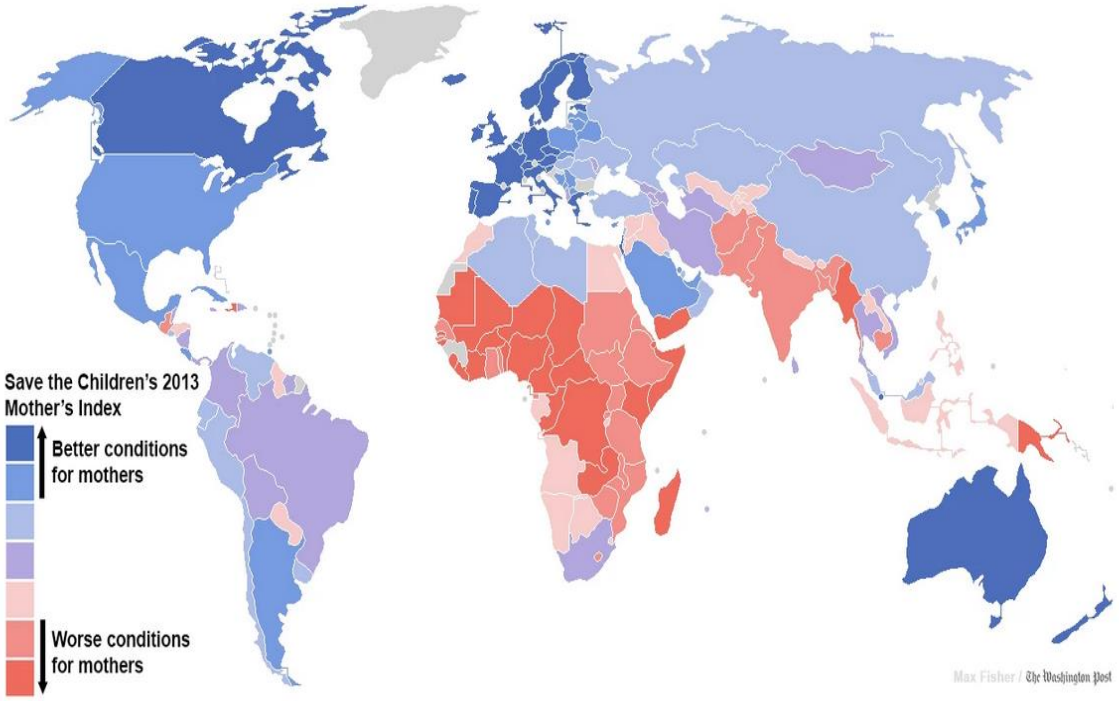


Fig. 3.6. **The best countries for motherhood**

Specialists had to analyze huge amounts of data collected from UN reports and other sources. On their basis, a map was drawn: on it, the countries where young mothers feel most comfortable are blue, the states where being a mother is not so easy are red.

One of the most common available mapping tools is the application Datawrapper [19].

Datawrapper is a service for visualization of data. You can merge Datawrapper with data of any formats (Excel, CSV, PDF). The created charts and maps that can be built into any website.

Datawrapper makes graphics and maps interactive: move the cursor on the line to see the exact values or compare different datasets with the buttons or timeline. The service can adapt fonts, colors and distance to your style.

Let's look at the step-by-step process of creating a map using this on-line application.

The first task is to prepare the data. As an example, let's take the data of the regional statistics of Ukraine, namely the region-based level of incomes of the population in 2016 (Table 3.1).

Table 3.1

The input data

Regions of Ukraine	Income of the population, million UAH
Vinnytsia	53 623
Volyn	29 879
Dnipro	140 077
Donetsk	83 053
Zhytomyr	39 324
Zakarpattia	31 980
Zaporizhzhia	73 605
Ivano-Frankivsk	41 946
Kyiv	67 102
Kropyvnytskyi	30 367
Luhansk	27 168
Lviv	86 105
Mykolaiv	38 823
Odesa	90 534
Poltava	52 339
Rivne	34 791
Sumy	38 480
Ternopil	28 500
Kharkiv	99 945
Kherson	33 479
Khmelnytskyi	42 610
Cherkasy	38 312
Chernivtsi	24 593
Chernihiv	33 149
The city of Kyiv	266 907

The Datawrapper application offers the creation of a map in 4 steps.

So, in the first step we have to choose the type of map (Fig. 3.7).

- *Choropleth maps* (maps that use differences in shading, coloring, or the placing of symbols within predefined areas to indicate the average values of a property or quantity in those areas).
- *Symbol maps* which show individual points to highlight a certain geographic distribution.
- *Locator maps* (maps of this kind add markers to a map to show where something is located or happened).

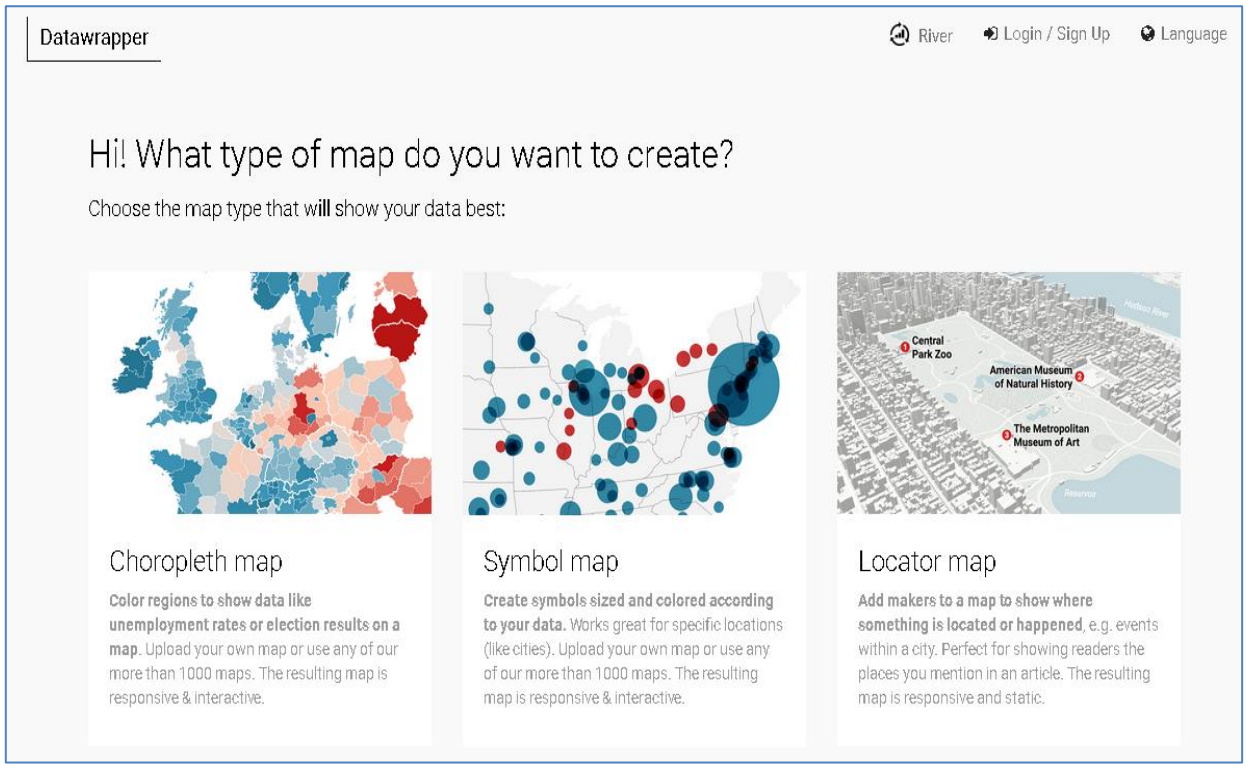


Fig. 3.7. **The types of maps**

Let's try all of them. Let's start with the first type – the Choropleth map. Click on the appropriate map type (Fig. 3.8).

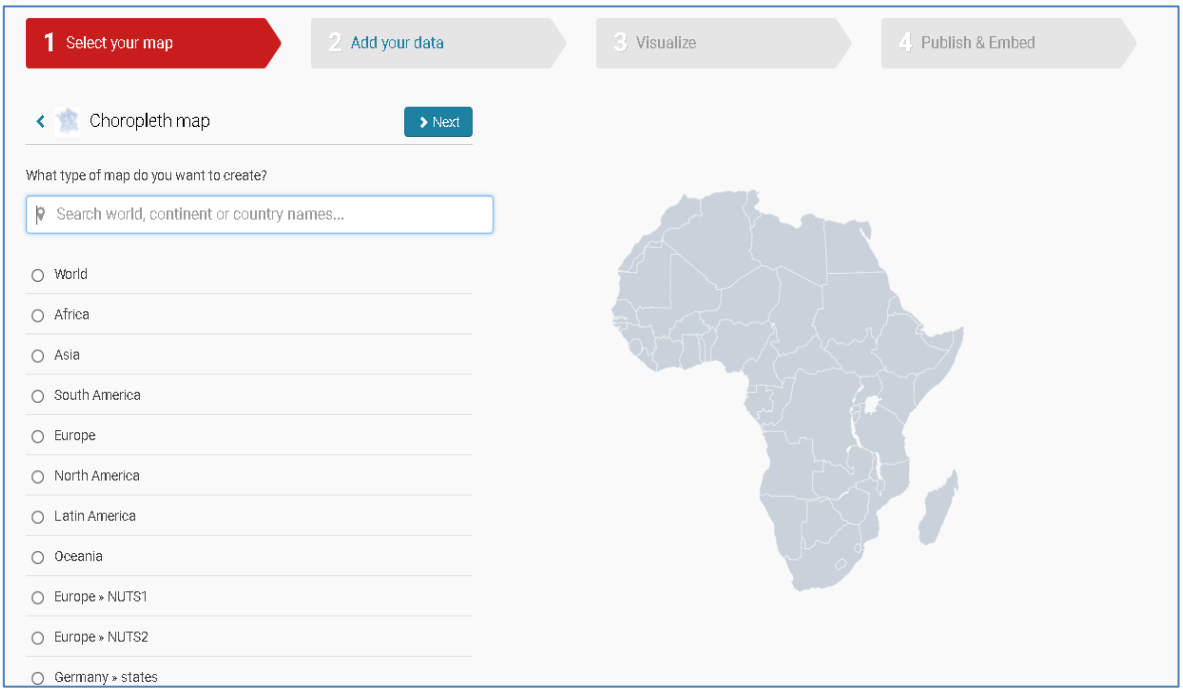


Fig. 3.8. **The Choropleth map**

Then we must choose a geographical continent, a specific country or all countries of the world. So, we choose Ukraine (Fig. 3.9).

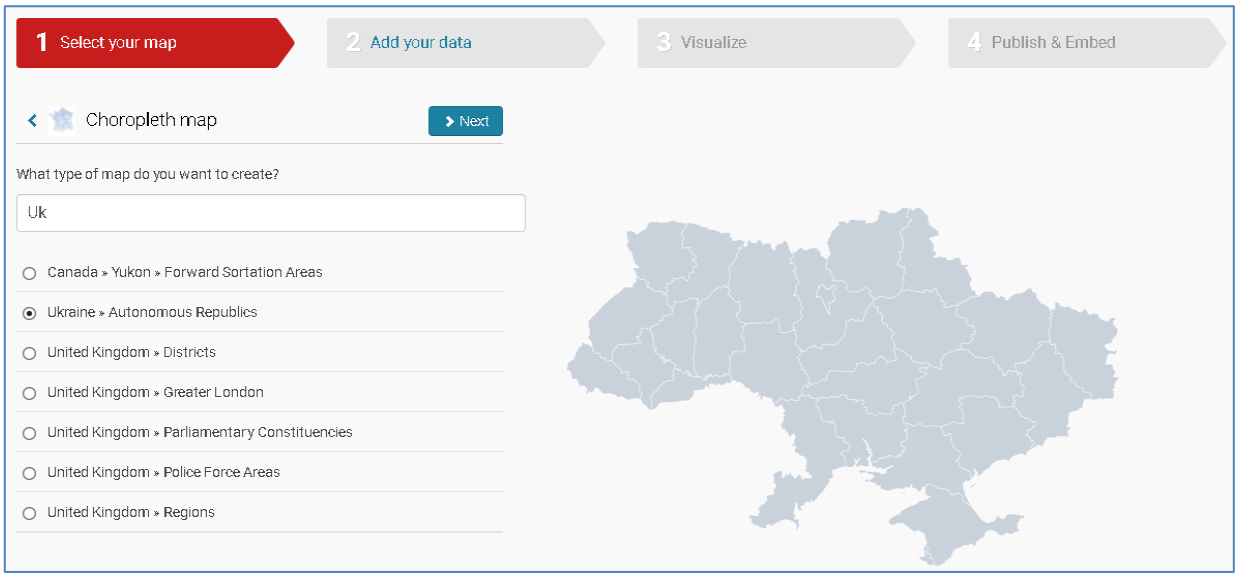


Fig. 3.9. **Choosing a country**

In the right side of the window we see the geographical map of Ukraine, divided into regions. We confirm your choice by clicking the *Next* button.

The second step. Now it's time to add data to your map.

Fill the table below with the values you want to visualize (Fig. 3.10). You can add additional columns to the table by right-clicking. You can also import your dataset automatically by clicking on the *Import* button below the table.

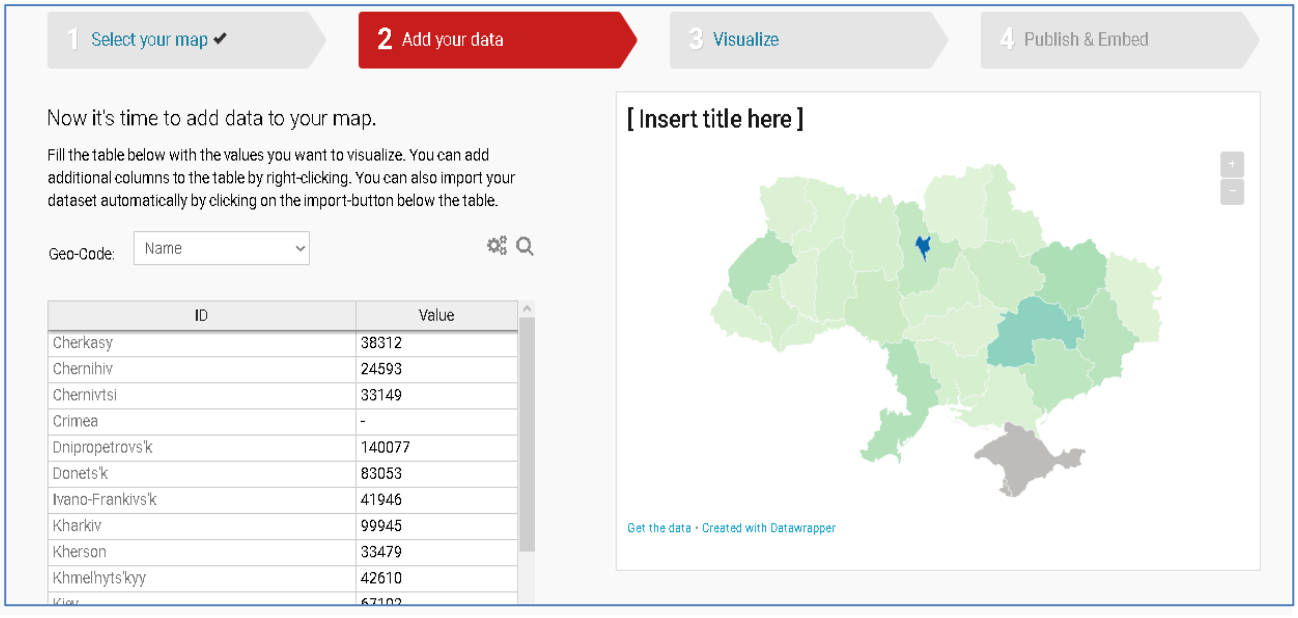


Fig. 3.10. **Adding the input data**

Then press the button *Proceed.*

In the third step, we perform the visualization of the map. Fill the necessary information in three basic tabs:

✓ Refine (specify) the color gamut, customize the tooltips, map labels, map key (legend) (Fig. 3.11).

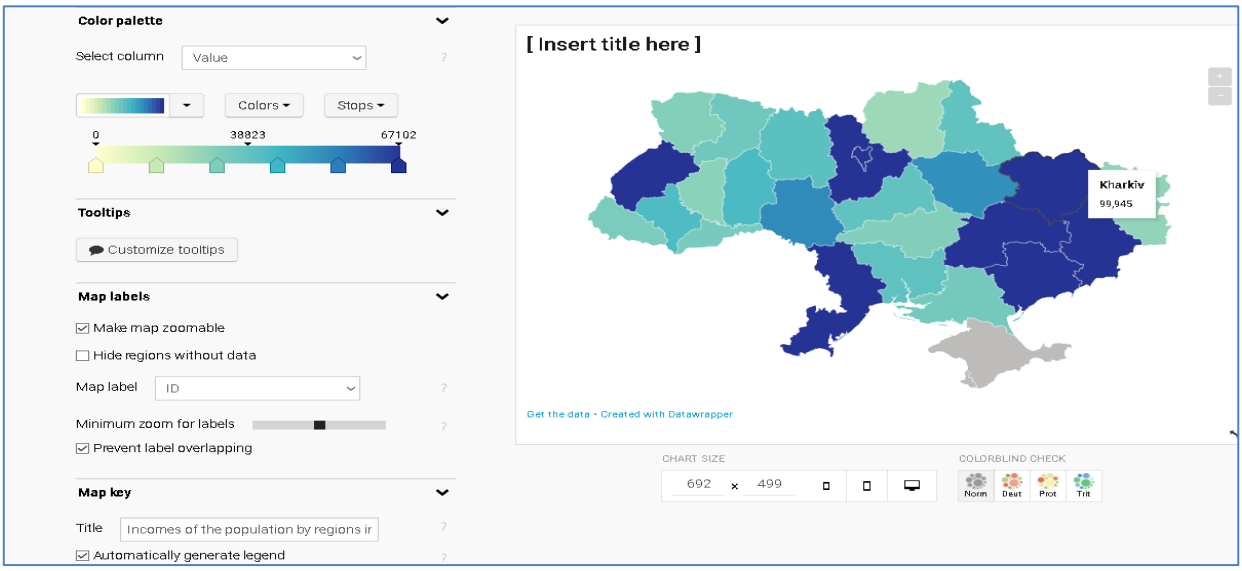


Fig. 3.11. **Choosing the design characteristics**

✓ Annotate (give the title name, description, note, data source, link to the data source, byline, and highlight the most important elements (optional)) (Fig. 3.12).

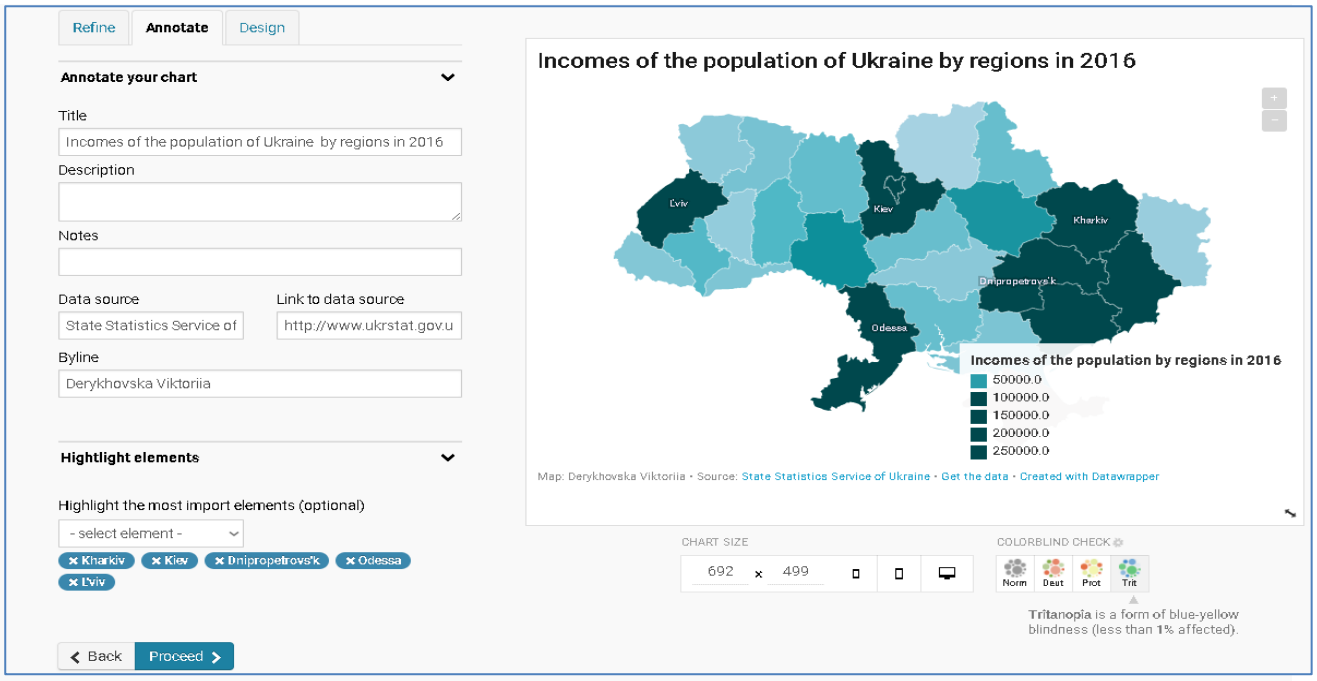


Fig. 3.12. **Filling the annotation**

✓ Design (*Select layout*, *Output locale* (defines decimal and thousand separators as well as translation of month and weekday names), *Social Sharing*) (Fig. 3.13).

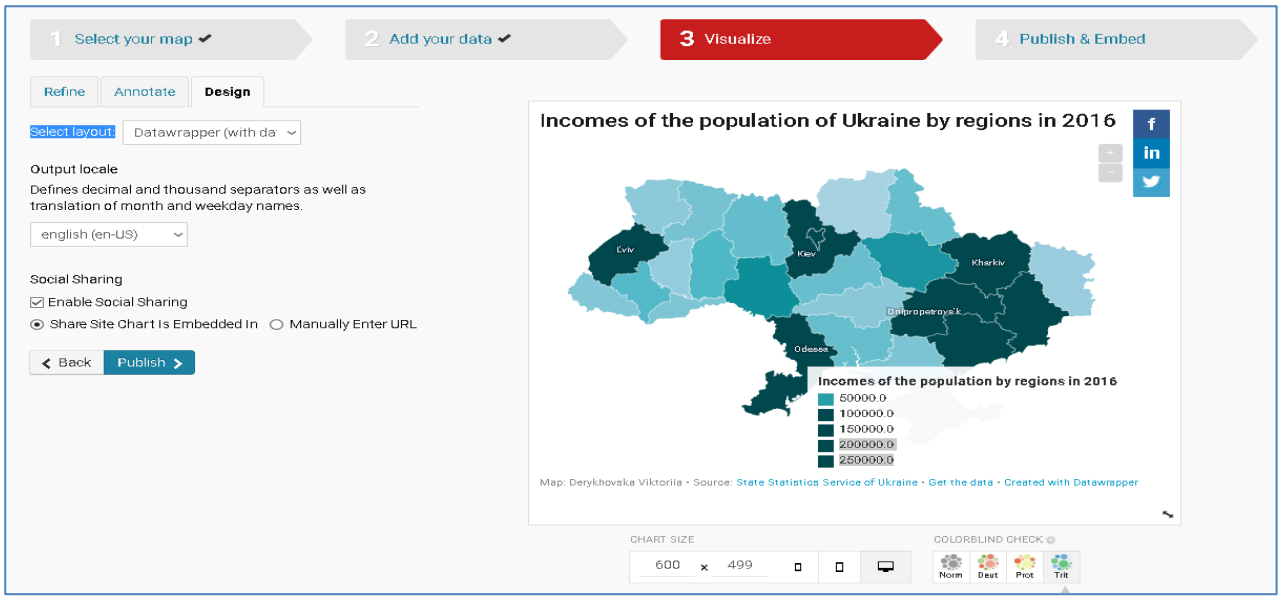


Fig. 3.13. **Choosing the design layout**

In the last, fourth step, we store the results of the cartography (Fig. 3.14). To share or embed your map you need to publish it. It will only be visible to people who know the map URL. Also you can *Publish, Export* or *Duplicate* the map.

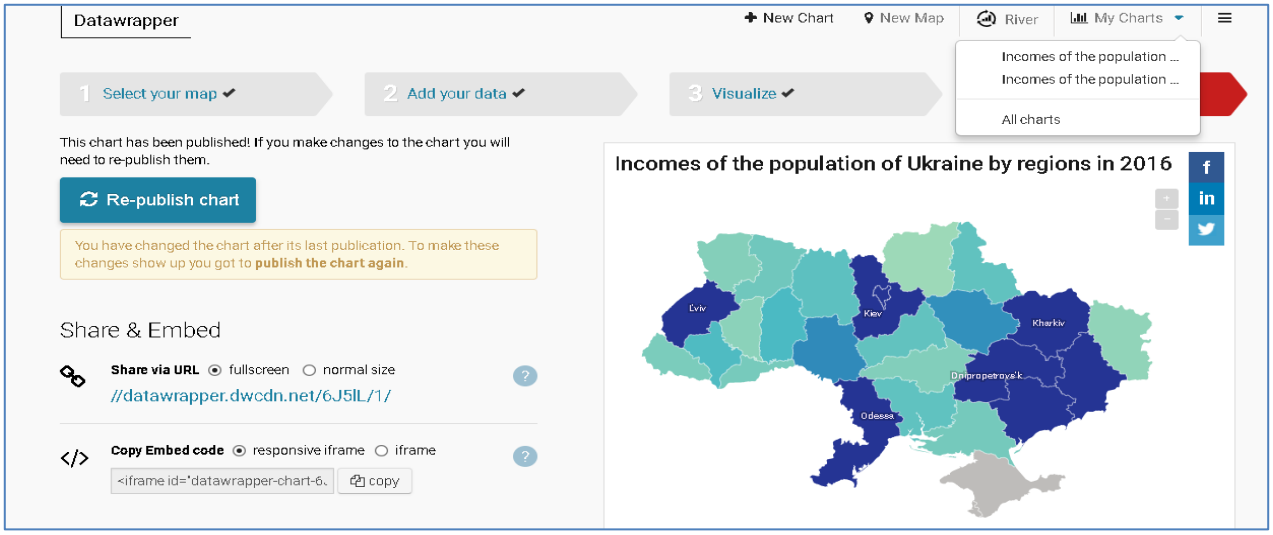


Fig. 3.14. **Choosing the way of sharing your map**

The best and easiest way to use a *Datawrapper* map is to embed it directly into your website or CMS. You can also upgrade your account to download your map as an image or PDF file to use it in different contexts.

Now the finished map of the distribution of the population of the regions of Ukraine in terms of income for 2016 is available by the reference //datawrapper.dwcdn.net/6J5lL/1/ (Fig. 3.15).

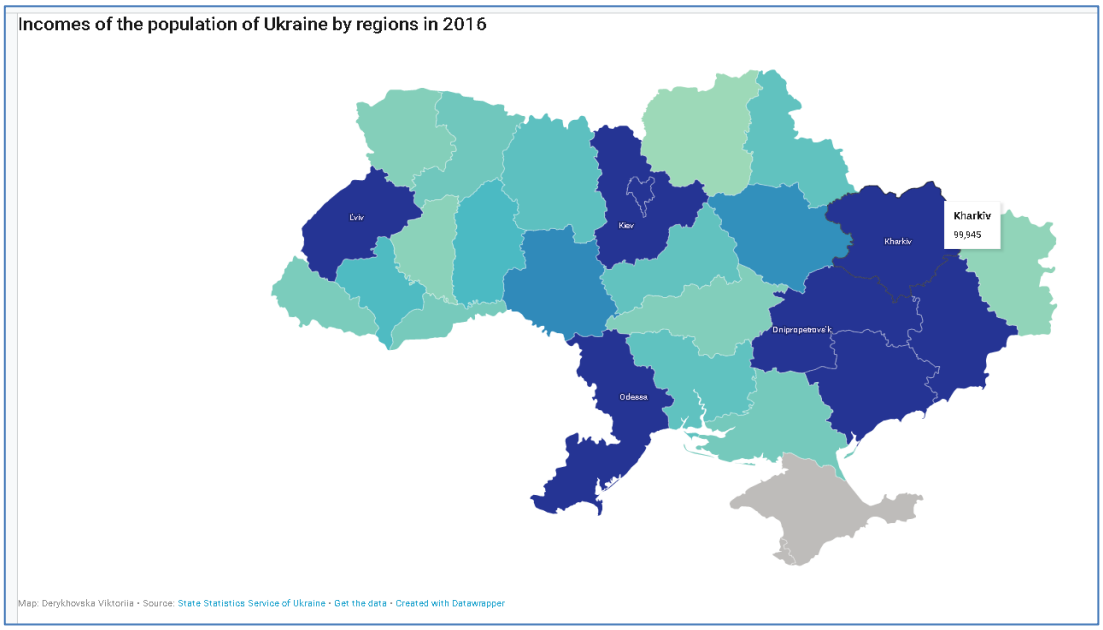


Fig. 3.15. **The result of cartography**

The map constructed with the help of the second type *Symbol map* will have the following form https://datawrapper.dwcdn.net/1x1Ex/1/ (Fig. 3.16).

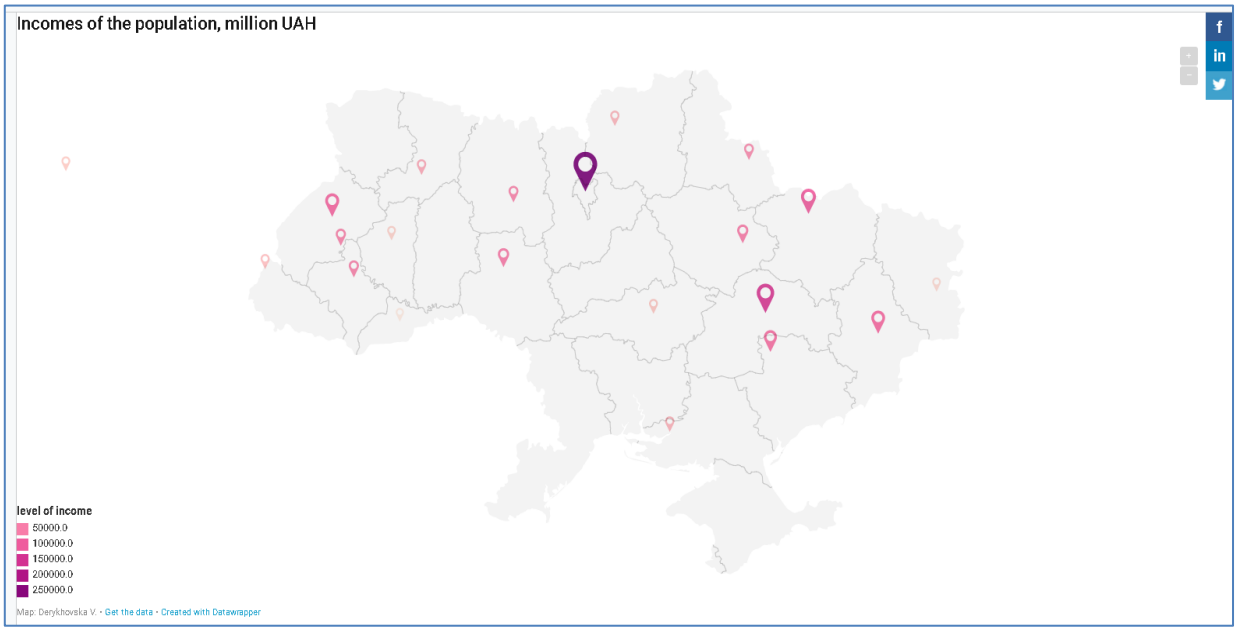


Fig. 3.16. **The symbol map of incomes of the population of Ukraine**

We can also build locator maps using the same algorithm as for the other two types:

To create a locator map, go to *datawrapper.de* and click on the *New Map* and then the *Locator Map*. This will open the *locator map* creation pipeline. Four steps will lead you to your result: 1. Add markers. 2. Design map. 3. Annotate. 4. Publish.

The editor steps follow the three basic elements that a locator map is made of: the symbols and texts, which we call "markers"; the map where the markers are added to; and finally the annotations on top and below the map (Fig. 3.17).

Fig. 3.17. **The basic elements of the locator map**

*Step 1. Add markers to the map (*Fig. 3.18 – 3.19).

Markers are the symbols, lines, areas or/and the text that you add on top of the map. The position of every marker is fixed to a map position: if you move the map view, the markers will move along with it. You have a wide range of options for customizing the markers.

While you're adding markers you will also start to adjust the map view. You can pan and zoom the view, but also rotate and even tilt the view angle. This is something you can do in all three steps.

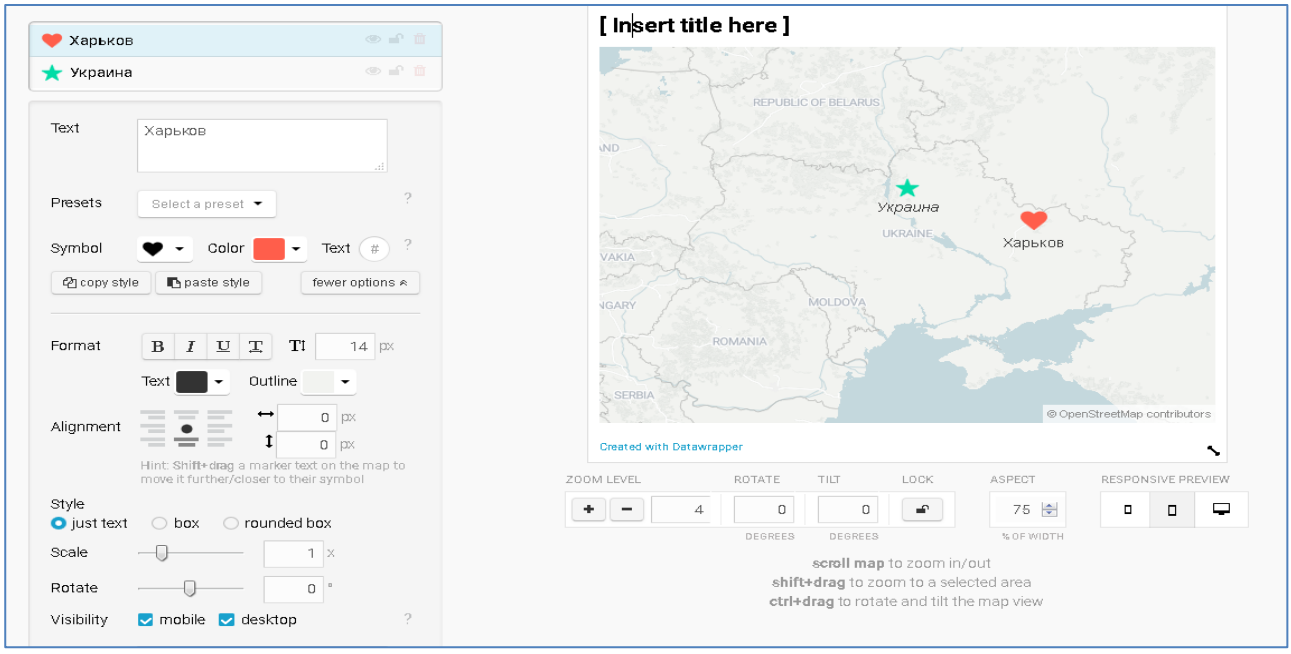


Fig. 3.18. **Adding markers to the map**

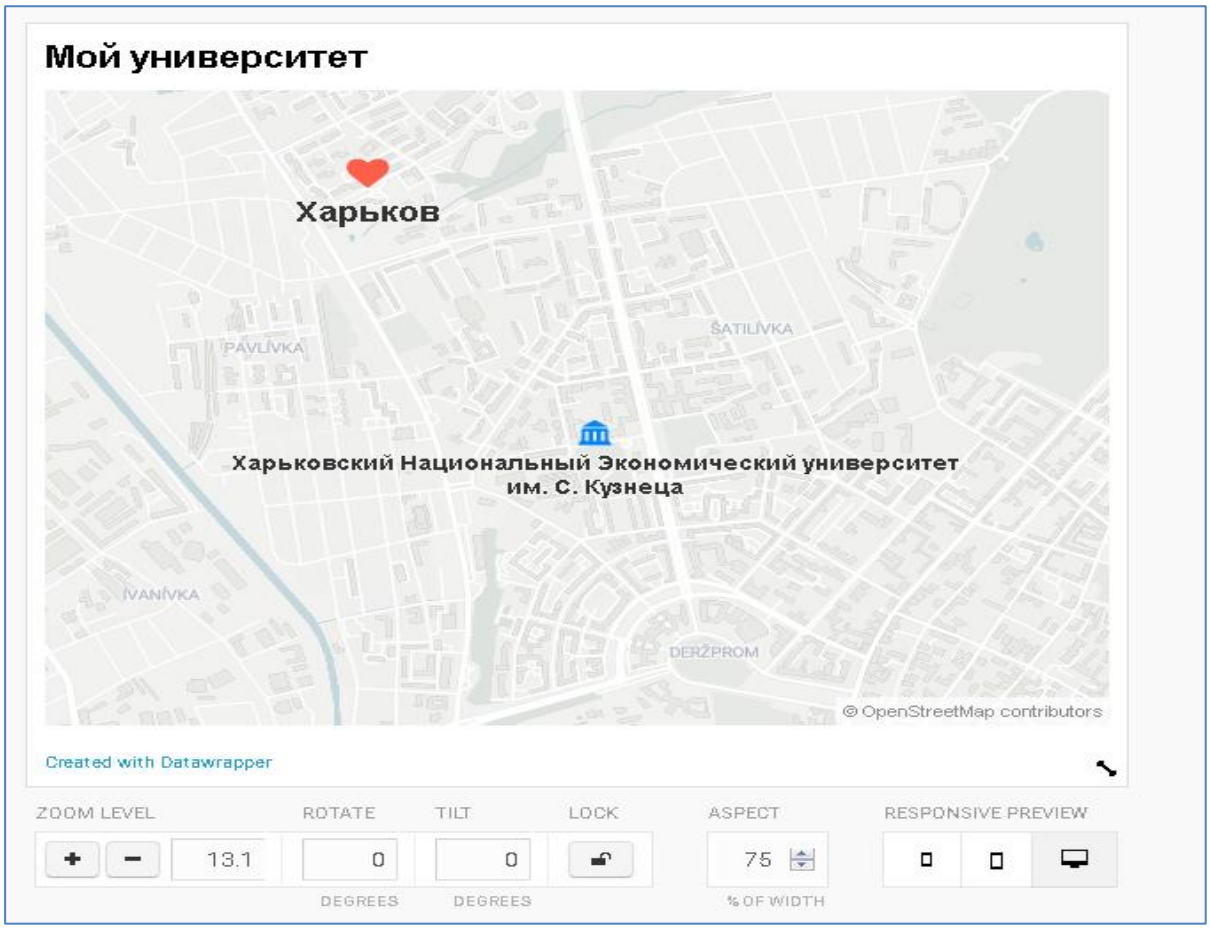


Fig. 3.19. **The result of adding markers to the map**

*Step 2. Design the map and add extras (*Fig. 3.20).

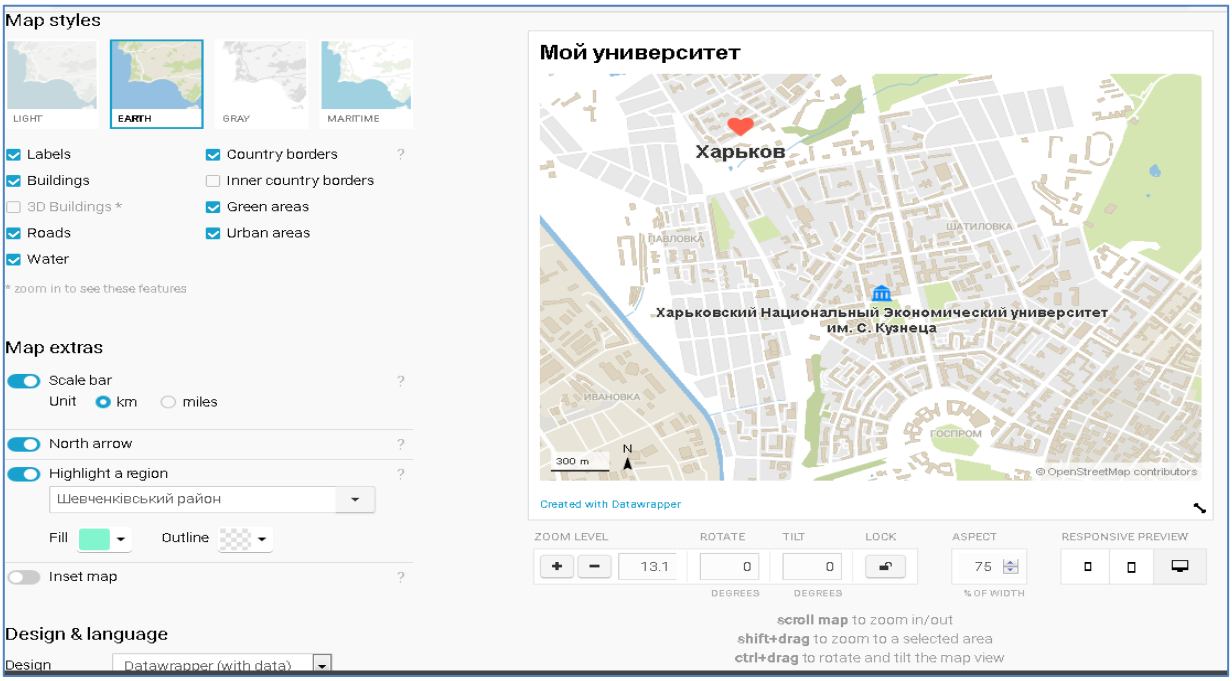


Fig. 3.20. **Choosing the map styles**

In this step, you can change how the map under your markers looks like. We offer four different map styles and you can toggle individual layers on or off (e.g. roads or water).

We also offer four extras that you can add to your map: a scale bar, a north arrow, an inset map, and a highlight region.

*Step 3. Annotate the map and add a map key (*Fig. 3.21).

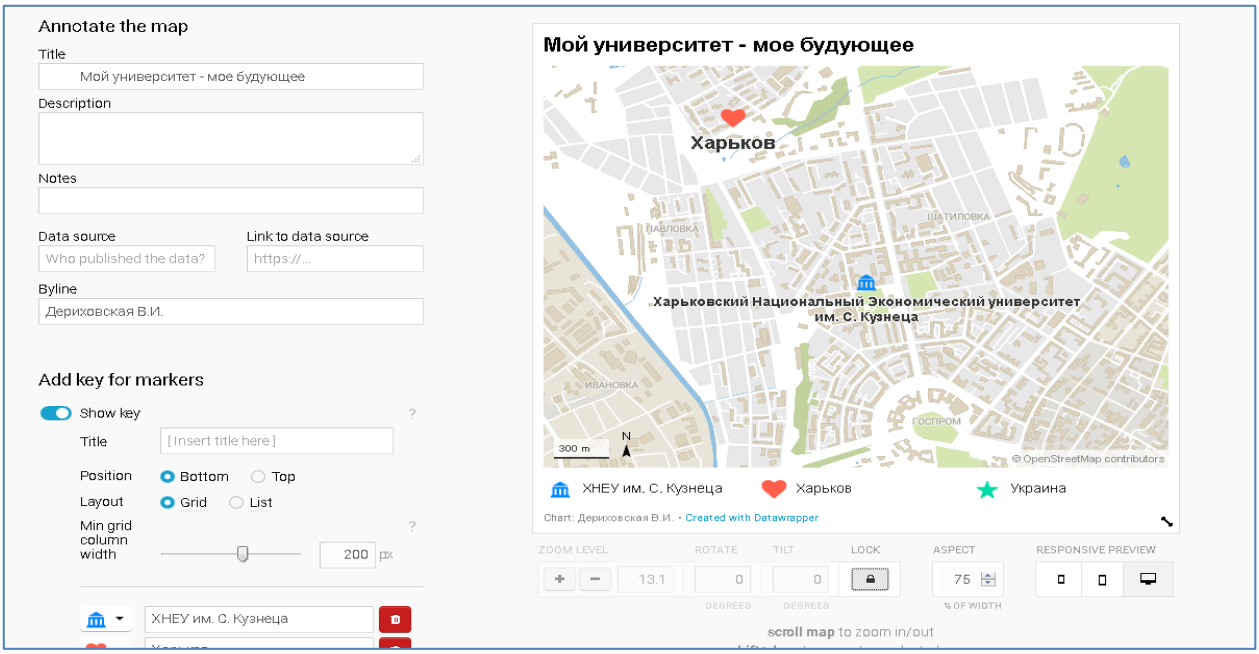


Fig. 3.21. **Filling the annotation of the map**

Finally, you can give your map a title and description, add notes below the map, give credit to a source, and set your byline. If you want to add more text to the places you marked on the map, you can use the special map key option.

The last step is to publish your locator map. You can embed your map in your website for free with *Datawrapper.*

Tasks for independent work

According to the selected research topic, build three maps, using all types (*Choropleth, Symbol* and *Locator Maps*) of the cartography from *Datawrapper*.

The results of the assignments should be presented in the form of a report on the laboratory work.

The list of questions for independent work

1. Describe the current toolbase of business analytics.
2. How is the source of information reviewed?
3. Name the criteria for choosing the sources of information.
4. Describe the processes of planning, collecting and preparing data for visualization.
5. What are the main mistakes of a business analyst when rendering data?
6. Name the basic methods of processing the digital information.
7. Describe the analytical methods of data processing.
8. Describe the expert methods of data processing.
9. Compare the advantages and disadvantages of using analytical and expert methods for processing the digital information.
10. Name the main reasons that lead to inaccurate expert assessments.

Test questions for self-assessment on topic 3

1. Reporting is one of the data collection methods:

a) yes;
b) no.

2. What type of data collection methods is more appropriate for more complex questions:

a) interviews;
b) questionnaires?

3. What type of interview needs to be interpreted and analysed even during the interview:

a) open-ended;
b) structured?

4. Such method of data collection as registry implies:

a) only new records;
b) new records and indication that a particular record is inactive or some record changes in operations.

5. Questionnaires refer to the forms filled in:

a) by respondents alone;
b) by the enumerator, who poses questions directly.

6. Socio-economic data can be obtained through questionnaires and, in all cases, variables obtained are:

a) an opinion;
b) a direct measurement.

7. In practice, observers:

a) only make direct observations;
b) might be involved in data processing and analysis;
c) do not conduct interviews and surveys using questionnaires.

8. The household data can be collected by:

a) reporting;
b) registration;
c) questionnaires and interviews.

9. Trade data should be collected from such sources as:

a) support industry;
b) consumers;
c) government-related agencies and institutions.

10. What resource includes the database "Fisheries and aquaculture":

a) http://unctad.org;
b) http://www.fao.org;
c) http://www.oecd.org?

Topic 4. The basics of business analytics. The basic methods for processing the digital information

Laboratory work 4

The purpose is to get acquainted with the basics of business analysis, get practical knowledge about the possibilities of using the analytical and expert methods of data fragmentation.

The task is to learn how to apply analytical methods of processing the digital information in accordance with available data and to master the basics of visualization of digital information.

Guidelines

In research, analytical methods are often used, with the help of which one can establish a mathematical relationship between the parameters of the subject being studied. These methods provide deep and comprehensive study of the processes under investigation, establish accurate quantitative relationships between arguments and functions, and help deeply analyze the phenomena being investigated. In practice, analytical methods of research are presented by a wide range of analytical tools, including correlation and regression analysis, factor and cluster analysis, discriminatory analysis, etc. In more detail, you get acquainted with the above methods in the course "Statistical Thinking about the Science of Data".

As part of this lab work we will dwell on other methods of data processing.

4.1. Data grouping.

Grouping is partitioning (dividing) data sets into groups according to a certain sign (feature, characteristics).

Let's consider the method of grouping based on the following example: we have data on foreign trade of Ukraine (exports) with EU countries in 2018 (Table 4.1).

Table 4.1

The input data

Country name	Export, thousand dollars USA
1	2
Austria	361 323,9
Belgium	251 541,9
Bulgaria	418 193,3
The United Kingdom	317 792,1
Greece	159 123,2
Denmark	155 453,9
Estonia	98 048,2
Ireland	45 483,9

Table 4.1 (the end)

1	2
Spain	1 004 547,4
Italy	1 929 575,6
Cyprus	53 481,4
Latvia	138 155,3
Lithuania	258 222,5
Luxemburg	5 087,0
Malta	7 746,0
The Netherlands	995 322,6
Germany	1 423 735,2
Poland	2 200 010,1
Portugal	228 126,0
Romania	716 981,4
Slovakia	471 362,6
Slovenia	16 758,2
Hungary	1 053 084,2
Finland	62 355,7
France	453 674,3
Croatia	39 065,9
The Czech Republic	560 756,1
Sweden	71 275,4

Grouping of statistical data should be performed in the mode *Histogram*, *Data Analysis* add-in in the package MS Excel. The mode *Histogram* is used to calculate the frequency of contact data in the specified boundaries of intervals, and to build the histogram of the interval variation distribution range (Fig. 4.1).

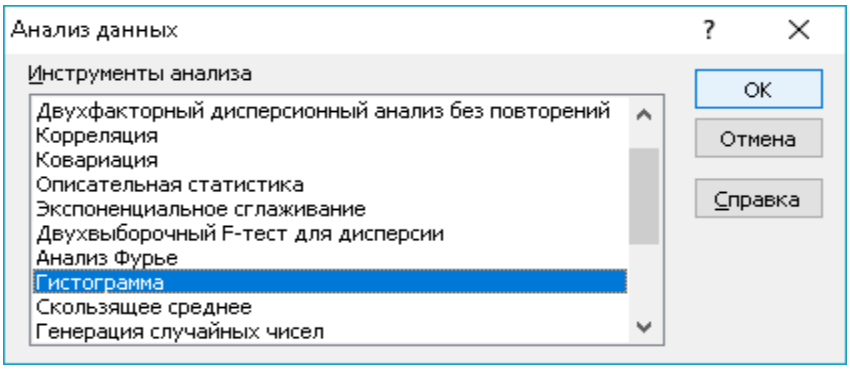


Fig. 4.1. **Choosing the mode *Histogram* in the package MS Excel**

The dialog box of this mode assumes the following parameters:

1. *The input range*, i.e. the location of the input data on your worksheet (only numeric data).

2. *The bin range* (an optional parameter), i.e. entering a link to the cells containing a set of limits that define the intervals (pocket). These values must be entered in the ascending order. In Microsoft Excel the number of intervals is calculated based on the data. The limits of intervals have a strict lower limit and a non-strict upper limit. If the bin range was not entered, then a set of intervals evenly distributed between the minimum and maximum values of the data will be generated automatically.

3. *Labels*, which are set in an active state if the first row (column) in the input range contains titles. If there are no titles, the checkbox should be deactivated. In this case, standard names for the output range will be automatically created.

4. *Output range* / *New worksheet* / *New Workbook* activates the field where the link to the upper-left cell of the output range should be entered. The output range is automatically detected and a message appears in case of the possible imposition of the original range to the output range.

5. *A pareto* (a sorted histogram) is set in an active state to represent the data in the descending order of frequency. If it is unchecked, the data in the output range will be listed in the order of intervals.

6. *Cumulative percentage* is set in an active state to calculate the percentage of accumulated frequencies and include the graph of accumulated frequencies in the histogram. Table 4.2 explains the calculation of the cumulated frequencies. Cumulative percentage is calculated based on the frequency (see Table 4.2, the "Frequencies" column). Each value of the cumulative frequency is divided by the maximum accumulated value. The result is relative frequency expressed in fractions of a unit. After the transformation to a percentage format the final result is obtained.

7. The *Chart output* is set in the active state for automatic building of an embedded chart on a worksheet containing the input range.

We need to construct a histogram and accumulated frequencies in the histogram, also visualize the results of grouping. In order to solve the task, the *Histogram* mode could be used. The parameters in the dialog box *Histogram* are shown in Fig. 4.2.

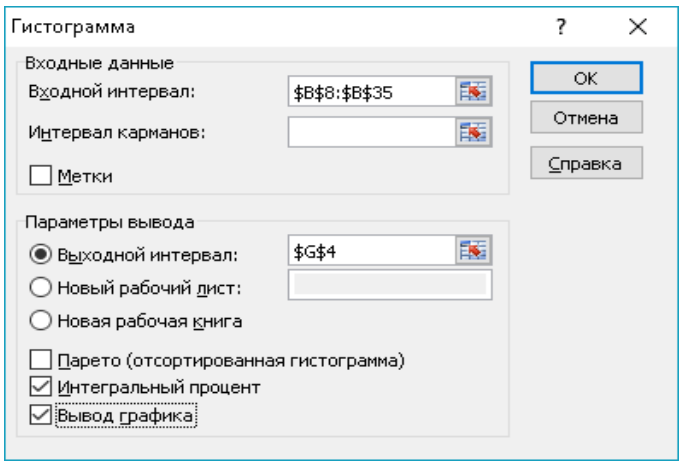


Fig. 4.2. **Defining the parameters in the *Histogram* mode**

The frequencies and the cumulative percentage are presented in Table 4.2.

Table 4.2

The frequencies and the cumulative percentage

Bin	Frequencies	Cumulative %
5 087.04213	1	3.70 %
444 071.6503	16	62.96 %
883 056.2584	4	77.78 %
1 322 040.867	3	88.89 %
1 761 025.475	1	92.59 %
More	2	100.00 %

So, we can see uneven export-based distribution: Ukraine exports from 5 087.04 thou to 444 071.65 thou worth of products from 16 countries. This means that Ukraine has the largest export volume of up to 45 thousand US dollars. The histogram of the constructed and accumulated frequencies is shown in Fig. 4.3.

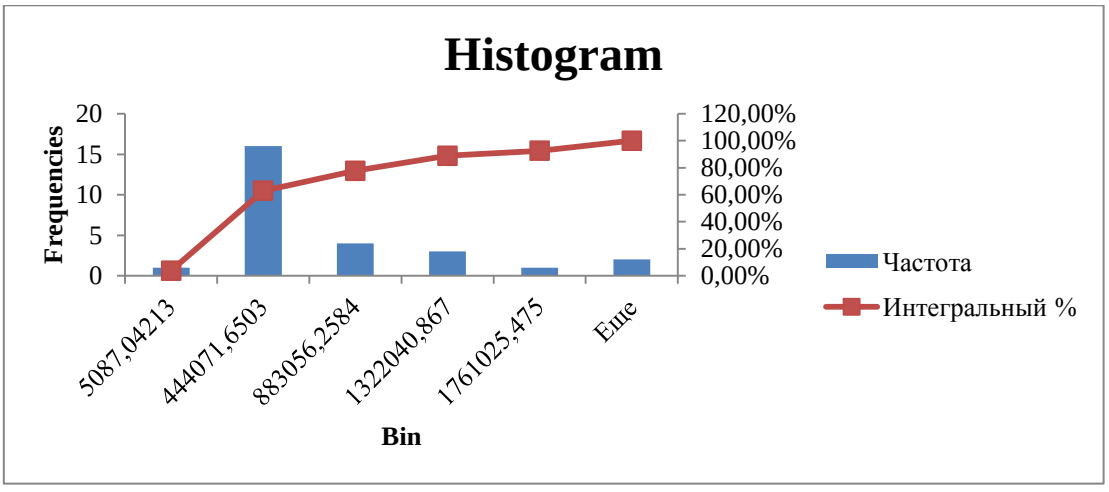


Fig. 4.3. **Visualization of the analysis results**

In conclusion, we can say that all the countries of the European Union, from which Ukraine exports its products, can be divided into 5 groups. At the same time, the maximum volume of exports falls to Poland and Italy, and the minimum – to Luxembourg. On the whole, exports are at a low level, which means that the Ukrainian government should reform its foreign economic policy and strengthen promotion (marketing), improve the quality and competitiveness of the product inventory in the markets of Europe.

4.2. Time series analysis.

Time series apply to studying of change of the phenomena in time.

The analysis of speed and intensity of the phenomenon in time is carried out by means of statistical indicators received as a result of comparison of levels. The compared level is called actual, and the level to which the comparison is made is basic. There are chain and basic indicators of the change of levels of a time series, and an average characteristic of the rows of dynamics.

If the base of comparison is the previous period, it is called a variable, and the received characteristics are called a chain. If the base of comparison is the initial level, it is called a constant, and the received characteristics are called basic.

The absolute increase shows how much the studied indicator changed in comparison with the previous (4.1) or basic period of time (4.2):

$$\Delta_B = Y_i - Y_0, \tag{4.1}$$

$$\Delta_C = Y_i - Y_{i-1}, \tag{4.2}$$

where y_0 is the base level of the time series;

y_i is the actual level of the time series;

y_{i-1} is the previous level of the row.

The growth rate shows by how many times the studied indicator changed in comparison with the previous period of time (4.3) or with the basic period of time (4.4):

$$Gr_C = \frac{y_i}{y_{i-1}} \tag{4.3}$$

$$Gr_B = \frac{y_i}{y_0} \tag{4.4}$$

The rate of increase shows the percentage of change of the studied indicator in comparison with the previous period of time or with the basic period of time (4.5):

$$R_i = Gr - 100 \quad (4.5)$$

The absolute value of one percent of an increase A % shows how many absolute units are contained in 1 % of the increase (4.6):

$$A\%_t^C = \frac{\Delta_t^C}{R_{i_t}^C} = \frac{y_t - y_{t-1}}{\frac{y_t - y_{t-1}}{y_{t-1}} \times 100} = \frac{y_{t-1}}{100} \quad (4.6)$$

Let's consider an example. We have some data on the amount of investment in the sector on a yearly basis (Table 4.3).

Table 4.3

The volume of investments in the sector

Years	2010	2013	2014	2015	2016	2017
The volume of investments in the sector, thou UAH	365.00	374.00	381.00	396.00	405.00	380.00

It is necessary to calculate all the parameters of the dynamic series (the base of comparison is 2010), the average annual growth rate and increase rate for the following periods: 1) 2010 – 2013; 2) 2013 – 2017 and 3) 2010 – 2017.

Let's create a file "Time series analysis" (Fig. 4.4). In column A and column B of sheet 1, the source data columns will be ("Years" and "The volume of investments into the sector, thou UAH" respectively).

	Years	The volume of investments in the sector, thou UAH	Absolute increase		Growth rate (Gr)		Rate of increase		A 1 %, thousand
			Base	Chain	Base	Chain	Base	Chain	
3	2010	365							
4	2013	374							
5	2014	381							
6	2015	396							
7	2016	405							
8	2017	380							

Fig. 4.4. **Creating the table with input data**

1. In order to define the absolute increase:

- enter the formula = B4 – $B$3 in the cell C4. Apply this formula to the entire column;
- enter the formula = B5 – B4 in the cell D5 (for 2013 it is impossible to calculate the absolute increase in the chain manner, because there is no previous period, i.e. 2012) and also apply the formula to the entire column.

The results are presented in Fig. 4.5.

	A	B	C	D
1	Years	The volume of investments in the sector, thousand UAH	(Δ y), thousand UAH	
2			Base	Chain
3	2010	365	-	-
4	2013	374	9	-
5	2014	381	16	7
6	2015	396	31	15
7	2016	405	40	9
8	2017	380	15	-25
9				

Fig. 4.5. **The results of the absolute growth rates calculations**

2. In order to define the growth rate:

- enter the formula = B4/$B$3 × 100 in the cell E4. Apply this formula to the entire column;
- enter the formula = B5/B4 × 100 in the cell F5 (for 2013, the growth rate as an absolute increase, cannot be calculated by the chain method, because there is no previous period, i.e. 2012) and also apply the formula to the entire column.

3. To define the rate of increase (1 % increase rate):

- enter the formula = E4 – 100 in the cell G4. Apply this formula to the entire column;
- enter the formula = F5 – 100 in the cell H5 and also apply the formula to the entire column.

4. To define the absolute value of 1 % of the increase:

enter the formula = B4/100 in the cell I5. Apply this formula to the entire column.

The calculation results are given in Fig. 4.6.

	A	B	C	D	E	F	G	H	I
1	Years	The volume of investments in the sector, thousand UAH	(Δ y), thousand UAH		Growth rate (G_t), %		Rate of increase (R_j),%		A 1 %, thousand UAH
2			Base	Chain	Base	Chain	Base	Chain	
3	2010	365	-	-	-	-	-	-	-
4	2013	374	9	-	102,47	-	2,47	-	-
5	2014	381	16	7	104,38	101,87	4,38	1,87	3,74
6	2015	396	31	15	108,49	103,94	8,49	3,94	3,81
7	2016	405	40	9	110,96	102,27	10,96	2,27	3,96
8	2017	380	15	-25	104,11	93,83	4,11	-6,17	4,05
9									

Fig. 4.6. **The results of the growth rate and increase rate calculations**

5. Let's define the average dynamics parameters.

5.1. The average level of the series is determined in the following way. Since we have a number of missed levels, it is expedient to calculate the average level of series for the period of 2013 – 2017, i.e. to select the built-in aggregate function *AVERAGE* in the cell B10.

5.2. To calculate the average absolute growth rate, enter the formula = (B8 – B3)/8 in the cell B11.

5.3. The average growth rate is calculated as follows:

1. For 2010 – 2013 one should enter the formula = (B4/B3)^(1/3) in the cell B14.

2. For 2013 – 2017:

the basic method is to enter the formula = (B8/B4)^(1/4) in the cell B16;

the chain method is to enter the formula = ((F5/100) × (F6/100) × × (F7/100) × (F8/100))^(1/4) in the cell B17.

3. For 2010 – 2017:

the basic method is to enter the formula = (B8/B3)^(1/7) in the cell B19;

the chain method of the weighted geometric mean formula is to enter the formula = ((B14)^3 × (B16)^4)^(1/7) in the cell B20.

5.4. The average increase rate is calculated in the following way:

1. For 2010 – 2013 one should enter the formula C14 = (B14 – 1) × 100 into the cell.

2. For 2013 – 2017 one should enter the formula = (B16 – 1) × 100 into the cell C16.

3. For 2010 – 2017 one should enter the formula = (B20 – 1) × 100 into the cell C20.

The results of the calculations of the average values are shown in Fig. 4.7.

10	The average level of the series	387,2	
11	The average absolute growth	1,875	
12		The average growth rate	The average rate of increase
13	1. 2010 - 2013		
14	basic method	1,0082	0,0082
15	2. 2013 - 2017		
16	basic method	1,0040	0,0040
17	chain method	1,4006	0,4006
18	3. 2010 - 2017		
19	basic method	1,0058	0,0058
20	chain method of weight geometric mean formula	1,0058	0,0058

Fig. 4.7. **The results of determining the average dynamics**

Thus, during the period of 2010 – 2016 a steady growth in investment was observed, but in 2017 the volume of investments decreased in comparison with 2016 by 25 thousand UAH (compared to 2010 it increased by 15 thousand UAH). The growth rate in 2017, calculated by the basic method was 104.11 %, i.e. the increase in the output was 4.11 %. The chain increase rate showed a decrease in the volume of production in comparison with 2016 by 6.17 %. 1 % growth rate in 2017 was equal to 405 UAH. In the period of 2010 – 2017, the investment increased annually by an average of 1.875 UAH, i.e. 0.58 %. The average investment in the period of 2013 – 2017 was 387.2 thousand UAH. In the period from 2010 to 2013 the investment was growing on average by 0.82 %, from 2013 to 2017 – by 0.4 %.

4.3. Analysis of development trends (extrapolation).

In theory and practice, in the process of forecasting financial indicators, quite often the extrapolation methodology is used, in which, conclusions about the value of predictive indicators in future periods are made on the basis of studying their dynamics in the previous periods. A necessary element in this case is the construction and analysis of the so-called series of dynamics, which classifies the values of time indices in the context of individual periods and describes the dynamics of their development. We emphasize that the analysis

of a number of dynamics of a separate indicator, for example, proceeds from the sale of products, is purely descriptive and does not explain the reasons for certain changes in the trend.

Extrapolation methods are used for the relatively stable development of the enterprise (or individual indicators of its activity) or in the presence of seasonal or cyclical fluctuations with a clearly defined trend. A trend is a long-term tendency of changes in economic indicators in economic forecasting. If the development of indicators of financial and economic activity of the enterprise in previous periods is characterized by significant instability and significant fluctuations of financial indicators, then their extrapolation for future periods will be impossible, and therefore, it is inappropriate to use such methods.

Perform extrapolation in MS Excel:

We have data on the gross domestic product (GDP) for 2006 – 2017 (Table 4.4) of a particular country. It is necessary to determine the forecast value of the GDP in 2018 by means of analytical alignment.

Table 4.4

The GDP volume for 2006 – 2017

Years	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
The GDP, billion UAH	190	210	200	220	240	230	220	240	260	260	270	280

On page 1 of the Excel workbook, the column A – B will be columns of output data (Fig. 4.8).

	A	B
1	Years	The GDP, billion UAH
2	2006	190
3	2007	210
4	2008	200
5	2009	220
6	2010	240
7	2011	230
8	2012	220
9	2013	240
10	2014	260
11	2015	260
12	2016	270
13	2017	280
14		

Fig. 4.8. **The input data in the Excel workbook**

To construct a trend line, you need to build a graph of the output, select a time series and select the *Add Trend Line* command from the context menu (Fig. 4.9).

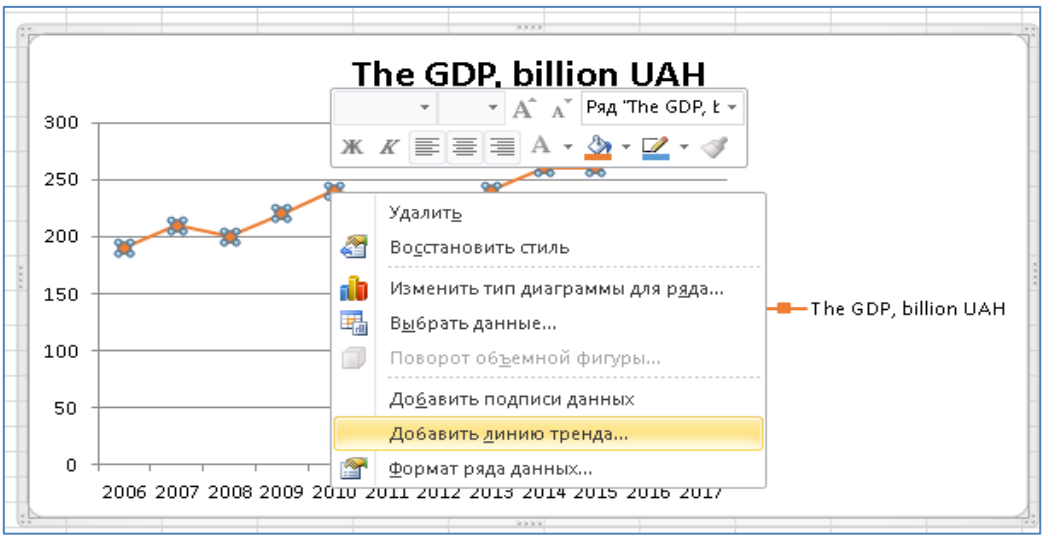


Fig. 4.9. **Visualization of the input data and the command *Add Trend Line***

After choosing the command, a dialog will open, in which we choose the type and parameters of the trend construction (Fig. 4.10).

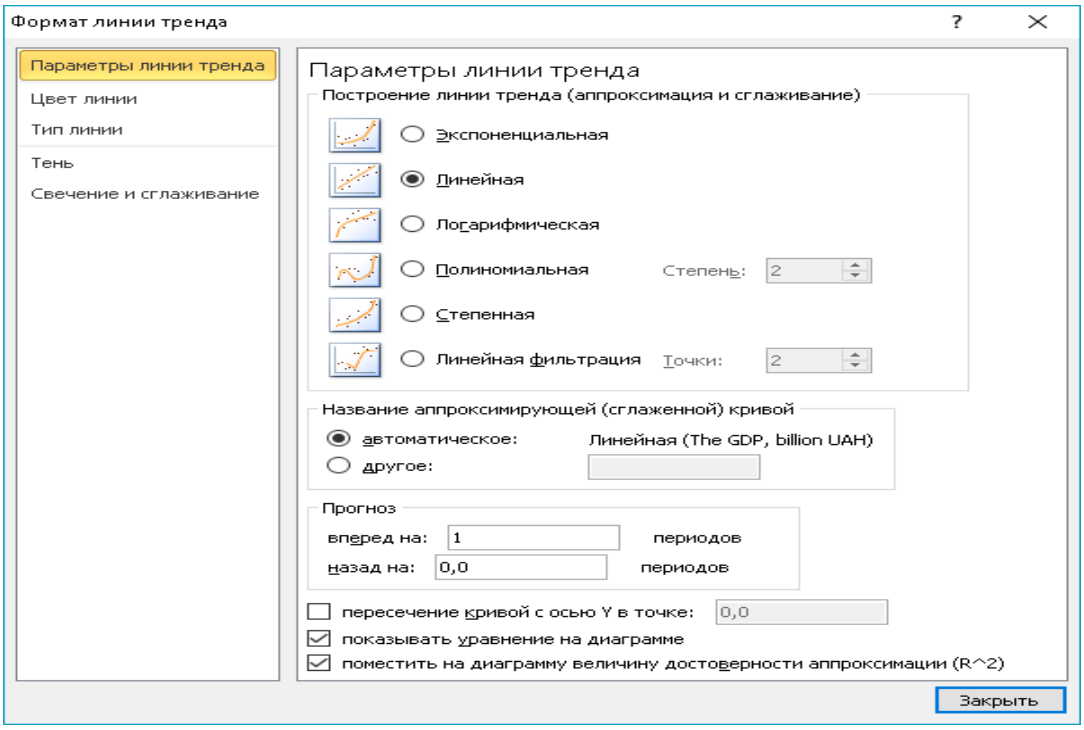


Fig. 4.10. **Choosing the parameters of the *Trend Line***

The result of constructing a linear trend is shown in Fig. 4.11.

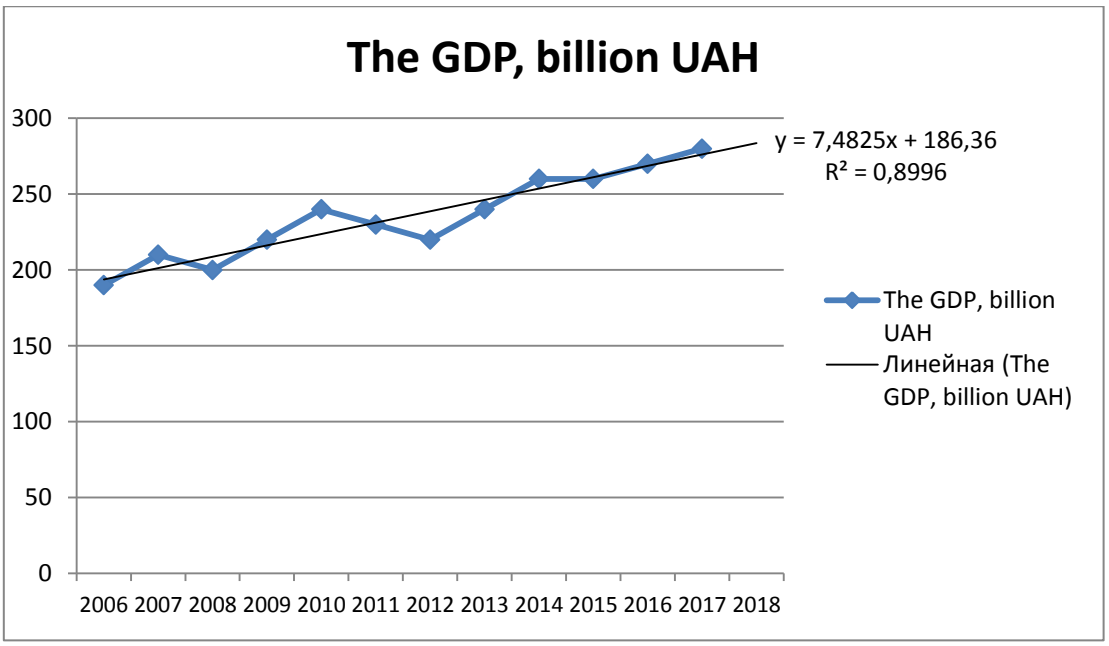


Fig. 4.11. **The result of adding the *Trend Line***

Thus, the equation of the straight line has the form:

$Y = 186.36 + 7.44825 \times t$.

The GDP in 2017 can be calculated by means of:

1. Point forecast, using the equation of the straight line, where *t* is a conditional indicator of the time at which the forecast will be constructed. That is, the volume of the GDP in 2017 is expected at the level:

$Y = 186.36 + 7.44825 \times t$;

$t = 13$;

$Y = 283.63$ billion UAH.

2. Using the standard *ПРЕДСКАЗ* (*FORECAST*) function. After selecting this function, a dialog opens in which we set the parameters (Fig. 4.12, 4.13).

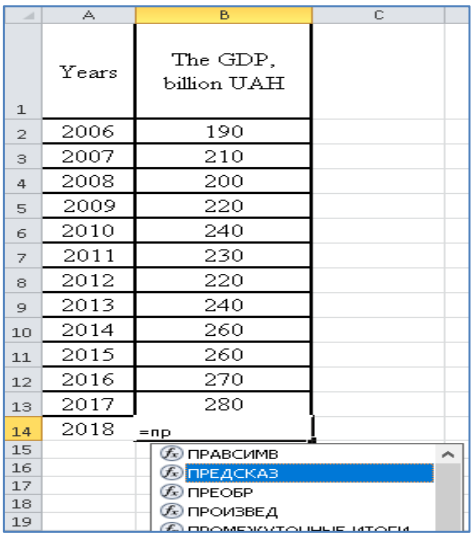


Fig. 4.12. **Choosing the function *ПРЕДСКАЗ***

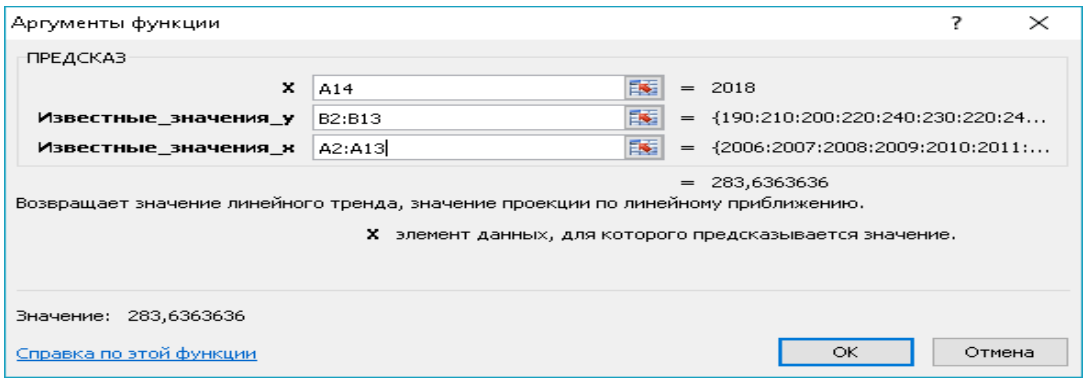
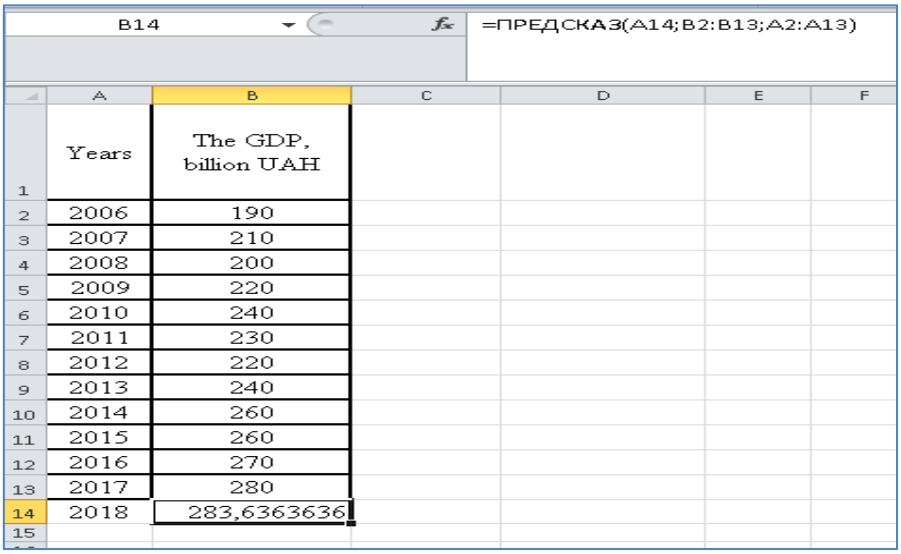


Fig. 4.13. **Choosing the parameters of the function *ПРЕДСКАЗ***

The results of the predicted value are shown in Fig. 4.14.



	A	B
1	Years	The GDP, billion UAH
2	2006	190
3	2007	210
4	2008	200
5	2009	220
6	2010	240
7	2011	230
8	2012	220
9	2013	240
10	2014	260
11	2015	260
12	2016	270
13	2017	280
14	2018	283,6363636
15		

Fig. 4.14. **The window of the results**

3. Using the standard *ТЕНДЕНЦИЯ (TREND)* function. After selecting this function, a dialog opens, in which we set the parameters (Fig. 4.15, 4.16).

Fig. 4.15. **Choosing the function *ТЕНДЕНЦИЯ***

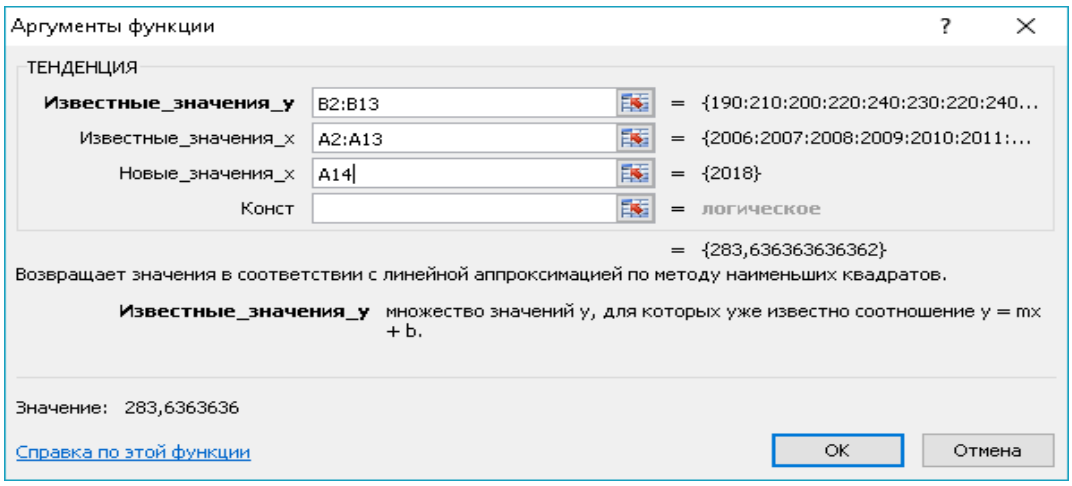
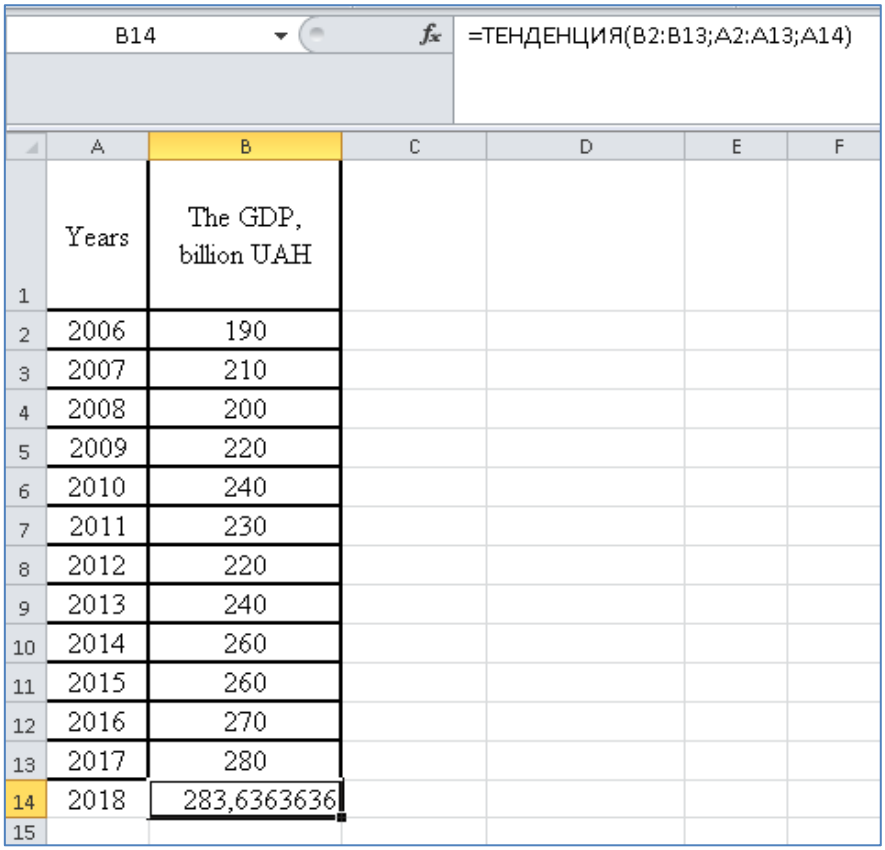


Fig. 4.16. **Choosing the parameters of the function *ТЕНДЕНЦИЯ***

The results of the predicted value using the *TREND* function are shown in Fig. 4.17.



	A	B	C	D	E	F
1	Years	The GDP, billion UAH				
2	2006	190				
3	2007	210				
4	2008	200				
5	2009	220				
6	2010	240				
7	2011	230				
8	2012	220				
9	2013	240				
10	2014	260				
11	2015	260				
12	2016	270				
13	2017	280				
14	2018	283,6363636				
15						

Fig. 4.17. **The window of the results**

As you can see, the use of three different methods produced the same results to be obtained graphically in Fig. 4.18.

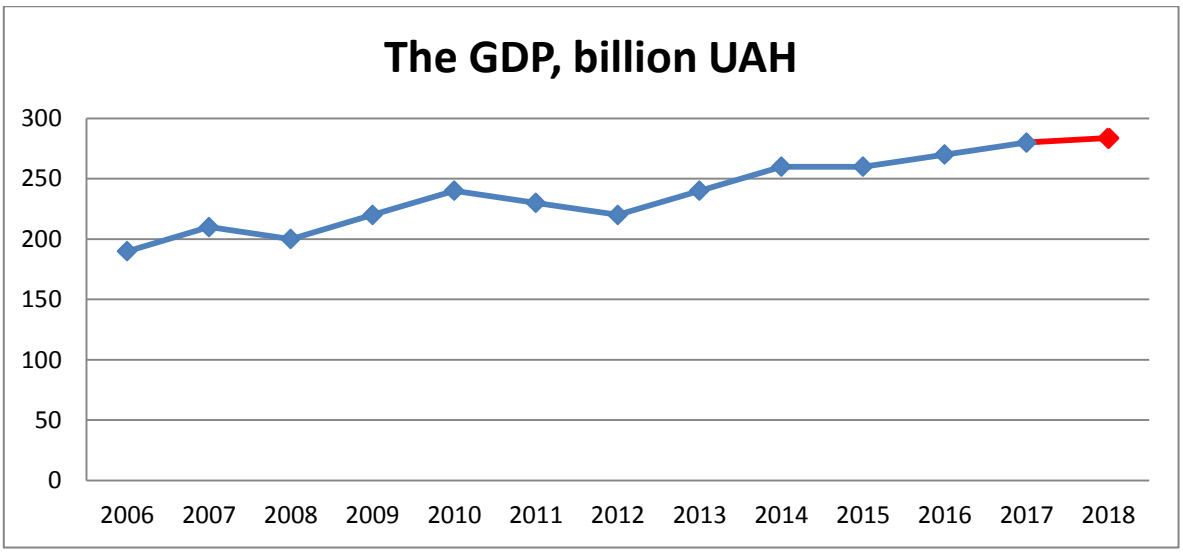


Fig. 4.18. **Visualization of the analysis results**

In conclusion, it is expedient to note that the GDP of this country has a positive tendency to increase.

4.4. The expert methods and the integral assessment method (the developmental level method).

Expert assessment methods are part of a broad field of decision theory, and expert evaluation is the procedure for obtaining an assessment of a problem based on the opinions of experts with a view to the subsequent decision (choice). In cases of extreme complexity of the problem, its novelty, lack of available information, impossibility of mathematical formalization of the decision process, one must turn to the recommendations of competent experts who are well aware of the problem. Their solution to the problem, the argumentation, the formation of quantitative estimates, the processing of the latter by formal methods have been called the method of expert assessments.

There are two groups of peer reviews:

1. Individual assessments based on the use of the opinions of individual experts, independent of each other.

2. Collective assessments based on the use of the collective opinion of experts.

The methods for measuring objects are as follows.

Ranking is the arrangement of objects in order of increase or decrease of any inherent property. Ranking allows you to select the most significant of the studied set of factors.

A paired comparison is the establishment of a preference for objects when all possible pairs are compared. Here it is not necessary, as in the ranking, to arrange all objects, it is necessary, in each of the pairs, to identify a more significant object or establish their equality.

Immediate evaluation. It is often desirable not only to order (rank the objects of analysis), but also to determine how much one factor is more significant than others. In this case, the range of the characteristics of the object is divided into separate intervals, each of which is assigned a certain score, for example, from 0 to 10. That is why the method of direct evaluation is sometimes also called the point method.

Ranking of economic objects by different methods has been carried out in the following example. It is necessary to evaluate and analyze the level of economic security of Ukraine. The objects of the research are 15 countries of the world. The input data of the study are shown in Table 4.5. It is evident from the table, that for 15 countries, the values of the selected five energy security indicators are diverse: the share of the dominant fuel resource in the consumption of fuel and energy resources, % (X1); the energy intensity of the GDP, kg of conditional fuel / UAH (X2); the volume of coal production, million tons (X3); the share of own sources in the balance of fuel and energy resources of the state, % (X4) and the degree of supply of fuel and energy resources (X5).

Table 4.5

The significance of the energy security indicators

Countries	X1	X2	X3	X4	X5
Austria	0.513	8.14	28.086	0.658	1.23
Belgium	0.441	11.184	46.61	0.423	1.2
Bulgaria	0.391	67.683	473.8	0.728	0.629
Finland	0.66	13.741	0	0.645	1.181
France	0.598	7.579	6.851	0.746	0.569
Germany	0.426	9.359	2 922.118	0.672	1
Italy	0.422	9.181	4.466	0.527	1.333
Poland	0.447	27.106	6 933.373	1.058	1.03
Spain	0.494	10.292	4 295.369	0.618	0.924
Sweden	0.828	8.865	0	0.909	0.948
Switzerland	0.695	4.622	0	0.755	0.952
The UK	0.39	7.828	1 005.633	1.954	2.16
Belorussia	0.603	80.723	0	0.521	1.343
Russia	0.562	121.363	3 960.709	2.9	2.942
Ukraine	0,541	179.236	2 740.338	0.676	1.077

Determine the ranking of countries based on the indicator of energy intensity of the GDP (X2) and build their ranking.

Ranking is a list of any objects (such as companies, countries, people, etc.) that can be arranged by any of the available scorecards. Unlike the rating, this is not a fixed form or method of calculation, and a database to get all the interesting options for ranking the original list. A distinctive feature of ranking is the objectivity, the independence of the result, and the ability to reverse the source list for an interesting ranking.

Differentiation of the values of the matrix of observations must be divided into stimulants and destimulants. The basis for such a distribution is the nature of the impact of each of them on the object being analyzed. Signs that positively affect the overall level of development of the object are called stimulants, and signs that inhibit development are called destimulants.

To do this, sorting the indicators by the X2 column is carried out as a sign of growth. Proceeding from the fact that the selected indicator is a non-stimulator, the ranked position 1 is given to the country with the lowest value of the indicator (Fig. 4.19).

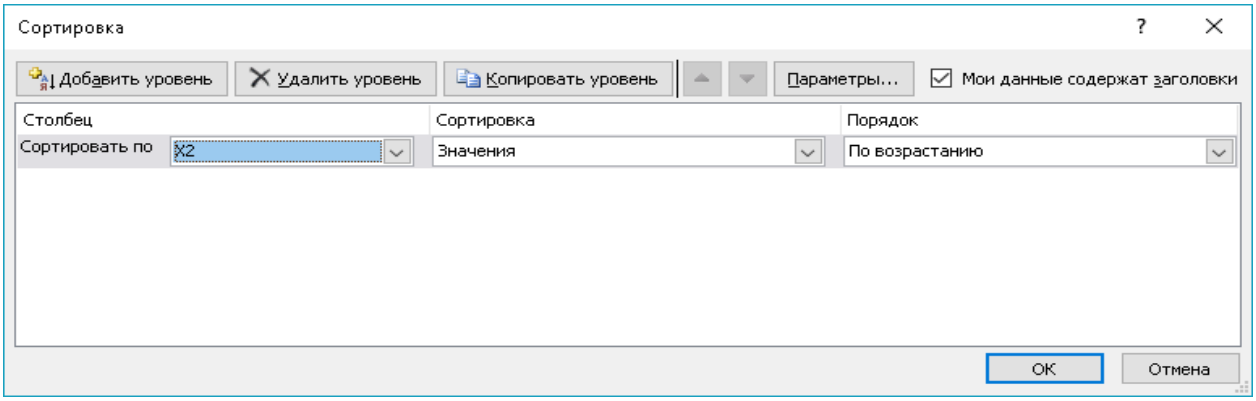


Fig. 4.19. **Sorting the indicators by the X2 column**

The result of the sorting are presented in Table 4.6.

Table 4.6

The ranking of countries based on the indicator of energy intensity of the GDP

No.	Countries	X1	X2	X3	X4	X5
1	2	3	4	5	6	7
1	Switzerland	0.695	**4.622**	0	0.755	0.952
2	France	0.598	**7.579**	6.851	0.746	0.569
3	The UK	0.39	**7.828**	1005.633	1.954	2.16

Table 4.6 (the end)

1	2	3	4	5	6	7
4	Austria	0.513	**8.14**	28.086	0.658	1.23
5	Sweden	0.828	**8.865**	0	0.909	0.948
6	Italy	0.422	**9.181**	4.466	0.527	1.333
7	Germany	0.426	**9.359**	2922.118	0.672	1
8	Spain	0.494	**10.292**	4295.369	0.618	0.924
9	Belgium	0.441	**11.184**	46.61	0.423	1.2
10	Finland	0.66	**13.741**	0	0.645	1.181
11	Poland	0.447	**27.106**	6933.373	1.058	1.03
12	Bulgaria	0.391	**67.683**	473.8	0.728	0.629
13	Belorussia	0.603	**80.723**	0	0.521	1.343
14	Russia	0.562	**121.363**	3960.709	2.9	2.942
15	Ukraine	0.541	**179.236**	2740.338	0.676	1.077

The taxonomy method (the general indicator of the level of development).

The first stage of the taxonomy method is standardization of signs. Standardization of the matrix of the output data is carried out according to the formulas 4.7 – 4.9:

$$z_{tj} = \frac{x_{ij} - \overline{x_j}}{s_j}, \qquad (4.7)$$

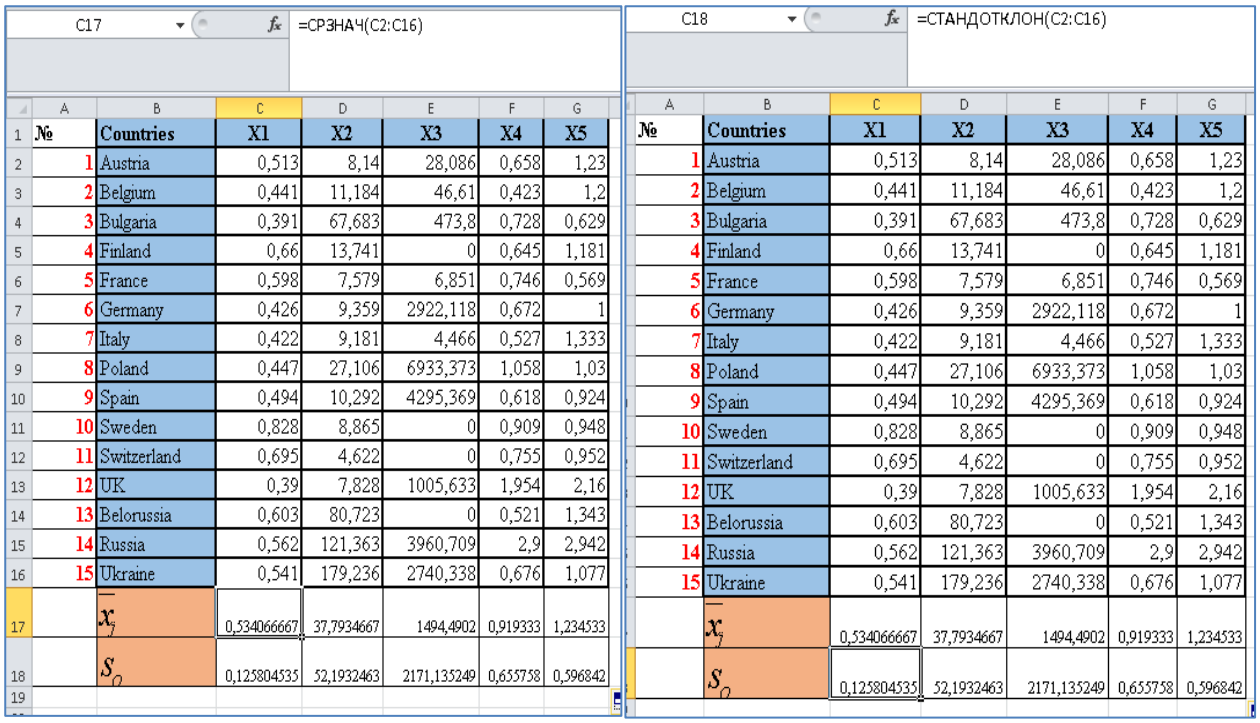
$$\overline{x_j} = \frac{1}{m}\sum_{t=1}^{m} x_{ij}, \qquad (4.8)$$

$$\text{and } s_j = \left[\frac{1}{m}\sum_{i=1}^{m}\left(x_{ij} - \overline{x_j}\right)^2\right]^{1/2}, \qquad (4.9)$$

where $\overline{x_j}$ is the arithmetic average value of j;

S_j is standard deviation of the indicator j.

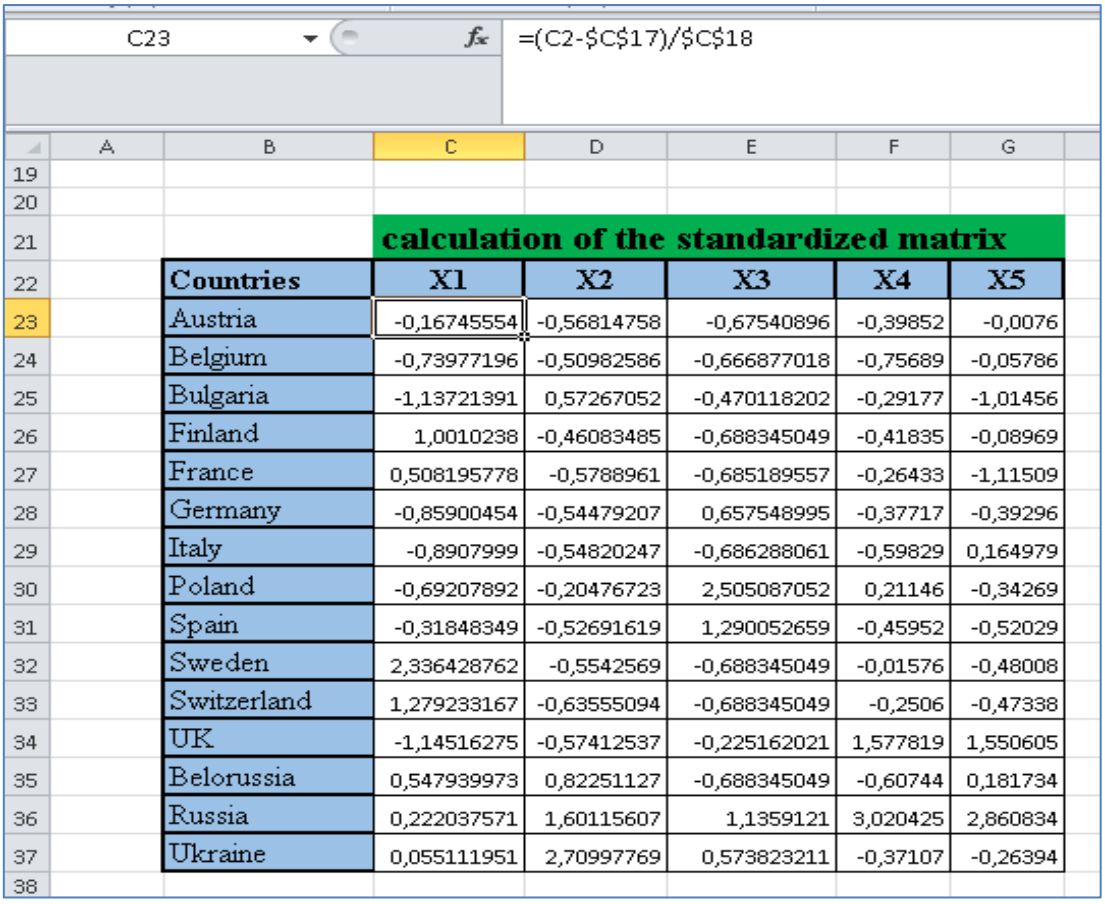
Preliminary calculations are carried with the use of the standard functions of the *СРЗНАЧ (AVERAGE)* and *СТАНДАРТОТКЛ (STDEV)*, as shown in the Fig. 4.20.



№	Countries	X1	X2	X3	X4	X5
1	Austria	0,513	8,14	28,086	0,658	1,23
2	Belgium	0,441	11,184	46,61	0,423	1,2
3	Bulgaria	0,391	67,683	473,8	0,728	0,629
4	Finland	0,66	13,741	0	0,645	1,181
5	France	0,598	7,579	6,851	0,746	0,569
6	Germany	0,426	9,359	2922,118	0,672	1
7	Italy	0,422	9,181	4,466	0,527	1,333
8	Poland	0,447	27,106	6933,373	1,058	1,03
9	Spain	0,494	10,292	4295,369	0,618	0,924
10	Sweden	0,828	8,865	0	0,909	0,948
11	Switzerland	0,695	4,622	0	0,755	0,952
12	UK	0,39	7,828	1005,633	1,954	2,16
13	Belorussia	0,603	80,723	0	0,521	1,343
14	Russia	0,562	121,363	3960,709	2,9	2,942
15	Ukraine	0,541	179,236	2740,338	0,676	1,077
	$\overline{x_j}$	0,534066667	37,7934667	1494,4902	0,919333	1,234533
	S_o	0,125804535	52,1932463	2171,135249	0,655758	0,596842

Fig. 4.20. **Finding of the functions *AVERAGE* and *STDEV***

After the calculation of the mean values and the standard deviation for each of the factors, a standardized matrix is calculated (Fig. 4.21).



calculation of the standardized matrix

Countries	X1	X2	X3	X4	X5
Austria	-0,16745554	-0,56814758	-0,67540896	-0,39852	-0,0076
Belgium	-0,73977196	-0,50982586	-0,666877018	-0,75689	-0,05786
Bulgaria	-1,13721391	0,57267052	-0,470118202	-0,29177	-1,01456
Finland	1,0010238	-0,46083485	-0,688345049	-0,41835	-0,08969
France	0,508195778	-0,5788961	-0,685189557	-0,26433	-1,11509
Germany	-0,85900454	-0,54479207	0,657548995	-0,37717	-0,39296
Italy	-0,8907999	-0,54820247	-0,686288061	-0,59829	0,164979
Poland	-0,69207892	-0,20476723	2,505087052	0,21146	-0,34269
Spain	-0,31848349	-0,52691619	1,290052659	-0,45952	-0,52029
Sweden	2,336428762	-0,5542569	-0,688345049	-0,01576	-0,48008
Switzerland	1,279233167	-0,63555094	-0,688345049	-0,2506	-0,47338
UK	-1,14516275	-0,57412537	-0,225162021	1,577819	1,550605
Belorussia	0,547939973	0,82251127	-0,688345049	-0,60744	0,181734
Russia	0,222037571	1,60115607	1,1359121	3,020425	2,860834
Ukraine	0,055111951	2,70997769	0,573823211	-0,37107	-0,26394

Fig. 4.21. **The standardized matrix**

1. Building a developmental model (etalon) involves the differentiation of a subset of stimulants and destimulants. Among the following indicators, the stimulants are: the share of own sources in the balance of fuel and energy resources of the state (X4), the volume of coal production (X3), the degree of supply of fuel and energy resources (X5). The destimulants include: the share of the dominant fuel resource in the consumption of fuel and energy resources (X1), energy intensity of the GDP (X2). The division of indicators into two groups is carried out according to their economic content.

The developmental standard has the following form (4.10):

$$Z_0 = [z_{0i},...,z_{on}], \tag{4.10}$$

where for stimulants $z_{0i} = \max_i z_{is}$, if $s \in I$;

for destimulants $z_{0i} = \min_t z_{is}$, if $s \notin I$ $(s = 1,...,n)$;

I is a set of stimulants.

According to the division of indicators into stimulants and destimulants with the use of the standard functions MAX and MIN, the value of the developmental model was calculated (Fig. 4.22).

	calculation of the standardized matrix				
Countries	**X1**	**X2**	**X3**	**X4**	**X5**
Austria	-0,16745554	-0,56814758	-0,67540896	-0,39852	-0,0076
Belgium	-0,73977196	-0,50982586	-0,666877018	-0,75689	-0,05786
Bulgaria	-1,13721391	0,57267052	-0,470118202	-0,29177	-1,01456
Finland	1,0010238	-0,46083485	-0,688345049	-0,41835	-0,08969
France	0,508195778	-0,5788961	-0,685189557	-0,26433	-1,11509
Germany	-0,85900454	-0,54479207	0,657548995	-0,37717	-0,39296
Italy	-0,8907999	-0,54820247	-0,686288061	-0,59829	0,164979
Poland	-0,69207892	-0,20476723	2,505087052	0,21146	-0,34269
Spain	-0,31848349	-0,52691619	1,290052659	-0,45952	-0,52029
Sweden	2,336428762	-0,5542569	-0,688345049	-0,01576	-0,48008
Switzerland	1,279233167	-0,63555094	-0,688345049	-0,2506	-0,47338
UK	-1,14516275	-0,57412537	-0,225162021	1,577819	1,550605
Belorussia	0,547939973	0,82251127	-0,688345049	-0,60744	0,181734
Russia	0,222037571	1,60115607	1,1359121	3,020425	2,860834
Ukraine	0,055111951	2,70997769	0,573823211	-0,37107	-0,26394
Etalon Z0	-1,14516275	-0,63555094	2,505087052	3,020425	2,860834

Fig. 4.22. **Defining the etalon**

2. Calculation of distances between the observation points and the point-etalon Z_0 is carried out according to the following formula (4.11):

$$c_{i0} = \left[\sum_{s=1}^{n} \left(z_{ij} - z_{0j} \right)^2 \right]^{1/2}, i = 1, 2, ..., m \qquad (4.11)$$

where C_{io} is the distance between the observation points and the point-etalon z_0;

z_{ij} is the value of the observation points;

z_{oj} is the point-etalon.

The calculation allows us to form a matrix of distances (Fig. 4.23).

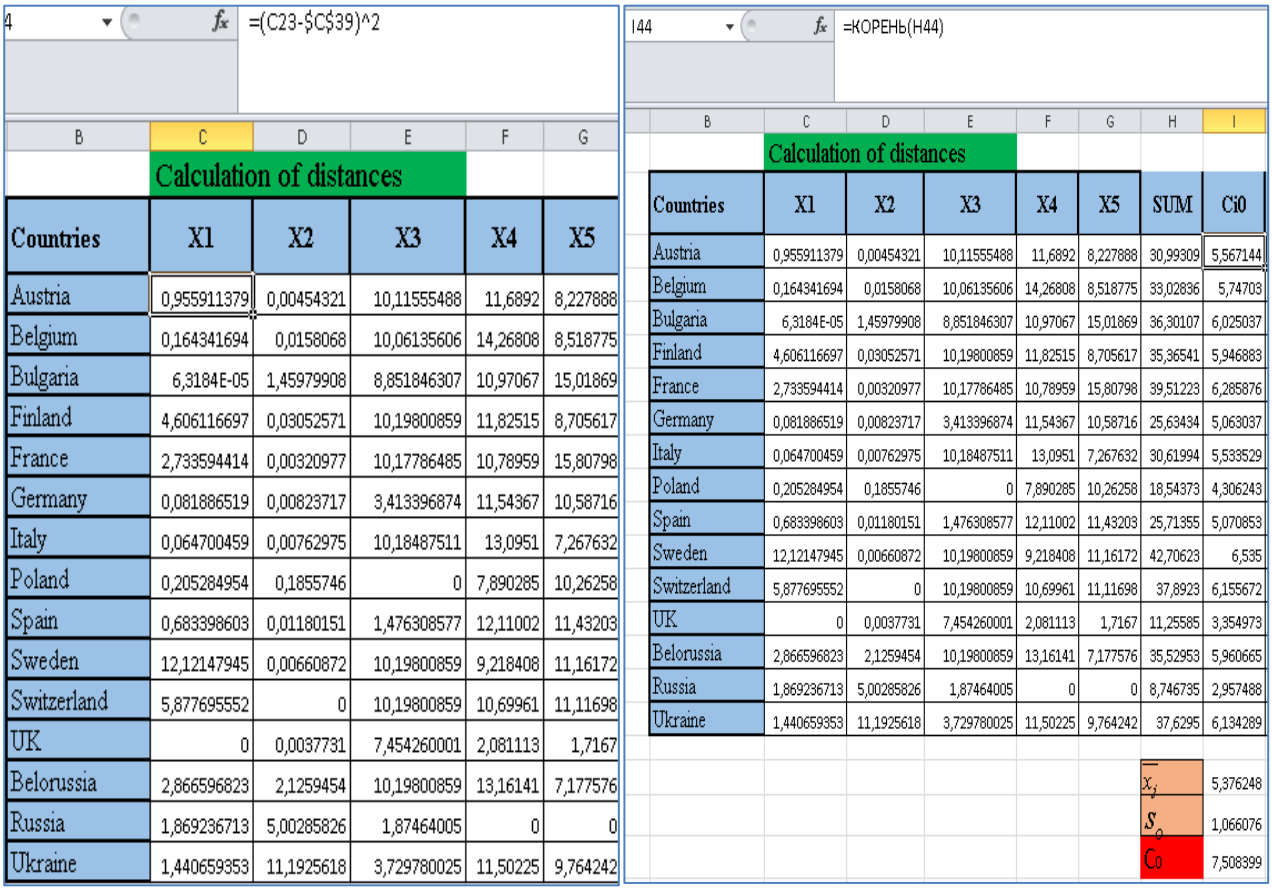


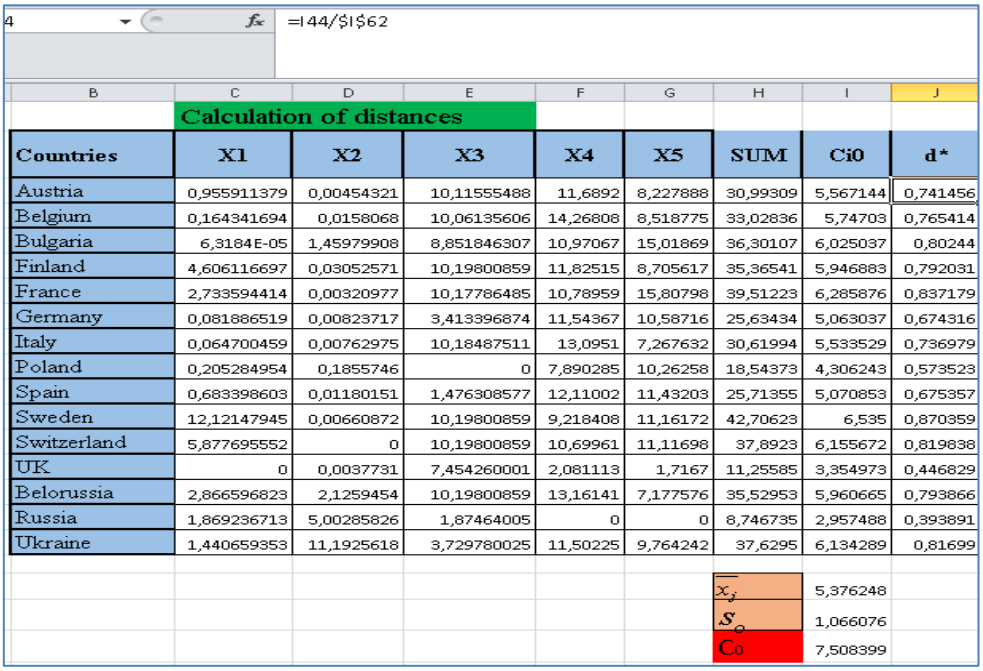
Fig. 4.23. **Defining the basic parameters of the matrix of distances**

3. Calculation of the taxonomic level of development (4.12 – 4.13):

$$d^* = (d_1^*, d_2^*, ..., d_m^*), \; d_i^* = \frac{c_{i0}}{c_0}, \qquad (4.12)$$

$$c_0 = \overline{c_0} + 2 \times S_0, \quad \overline{c_0} = \frac{\sum_{i=1}^{m} c_{i0}}{m}, \quad S_0 = \sqrt{\frac{\sum_{i=1}^{m} \left(c_{i0} - \overline{c_0} \right)^2}{m}}. \qquad (4.13)$$

Fig. 4.24 shows the calculations for the vector of the values of d*.



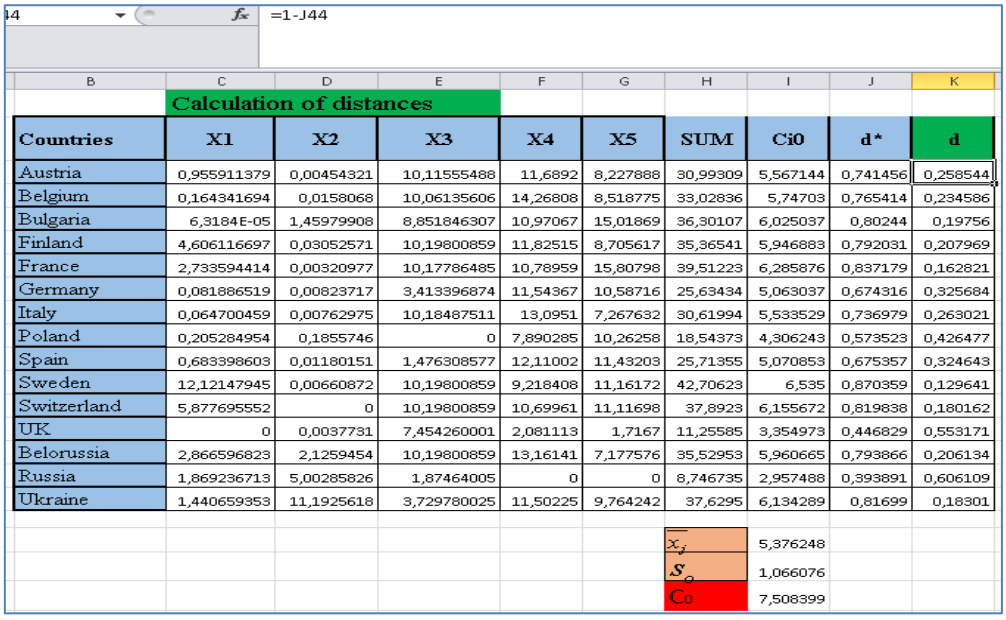
=I44/$I$62

	B	C	D	E	F	G	H	I	J
		Calculation of distances							
	Countries	X1	X2	X3	X4	X5	SUM	Ci0	d*
	Austria	0,955911379	0,00454321	10,11555488	11,6892	8,227888	30,99309	5,567144	0,741456
	Belgium	0,164341694	0,0158068	10,06135606	14,26808	8,518775	33,02836	5,74703	0,765414
	Bulgaria	6,3184E-05	1,45979908	8,851846307	10,97067	15,01869	36,30107	6,025037	0,80244
	Finland	4,606116697	0,03052571	10,19800859	11,82515	8,705617	35,36541	5,946883	0,792031
	France	2,733594414	0,00320977	10,17786485	10,78959	15,80798	39,51223	6,285876	0,837179
	Germany	0,081886519	0,00823717	3,413396874	11,54367	10,58716	25,63434	5,063037	0,674316
	Italy	0,064700459	0,00762975	10,18487511	13,0951	7,267632	30,61994	5,533529	0,736979
	Poland	0,205284954	0,1855746	0	7,890285	10,26258	18,54373	4,306243	0,573523
	Spain	0,683398603	0,01180151	1,476308577	12,11002	11,43203	25,71355	5,070853	0,675357
	Sweden	12,12147945	0,00660872	10,19800859	9,218408	11,16172	42,70623	6,535	0,870359
	Switzerland	5,877695552	0	10,19800859	10,69961	11,11698	37,8923	6,155672	0,819838
	UK	0	0,0037731	7,454260001	2,081113	1,7167	11,25585	3,354973	0,446829
	Belorussia	2,866596823	2,1259454	10,19800859	13,16141	7,177576	35,52953	5,960665	0,793866
	Russia	1,869236713	5,00285826	1,87464005	0	0	8,746735	2,957488	0,393891
	Ukraine	1,440659353	11,1925618	3,729780025	11,50225	9,764242	37,6295	6,134289	0,81699
							$\overline{x_j}$	5,376248	
							S_o	1,066076	
							Co	7,508399	

Fig. 4.24. **Defining the parameter d***

In order for the indicator d* to take high values at higher values of stimulants and low values at low values of stimulants, it is converted to the form (Fig. 4.25) (4.14):

$$d_i = 1 - \frac{c_{io}}{c_o}. \qquad (4.14)$$



=1-J44

	B	C	D	E	F	G	H	I	J	K
		Calculation of distances								
	Countries	X1	X2	X3	X4	X5	SUM	Ci0	d*	d
	Austria	0,955911379	0,00454321	10,11555488	11,6892	8,227888	30,99309	5,567144	0,741456	0,258544
	Belgium	0,164341694	0,0158068	10,06135606	14,26808	8,518775	33,02836	5,74703	0,765414	0,234586
	Bulgaria	6,3184E-05	1,45979908	8,851846307	10,97067	15,01869	36,30107	6,025037	0,80244	0,19756
	Finland	4,606116697	0,03052571	10,19800859	11,82515	8,705617	35,36541	5,946883	0,792031	0,207969
	France	2,733594414	0,00320977	10,17786485	10,78959	15,80798	39,51223	6,285876	0,837179	0,162821
	Germany	0,081886519	0,00823717	3,413396874	11,54367	10,58716	25,63434	5,063037	0,674316	0,325684
	Italy	0,064700459	0,00762975	10,18487511	13,0951	7,267632	30,61994	5,533529	0,736979	0,263021
	Poland	0,205284954	0,1855746	0	7,890285	10,26258	18,54373	4,306243	0,573523	0,426477
	Spain	0,683398603	0,01180151	1,476308577	12,11002	11,43203	25,71355	5,070853	0,675357	0,324643
	Sweden	12,12147945	0,00660872	10,19800859	9,218408	11,16172	42,70623	6,535	0,870359	0,129641
	Switzerland	5,877695552	0	10,19800859	10,69961	11,11698	37,8923	6,155672	0,819838	0,180162
	UK	0	0,0037731	7,454260001	2,081113	1,7167	11,25585	3,354973	0,446829	0,553171
	Belorussia	2,866596823	2,1259454	10,19800859	13,16141	7,177576	35,52953	5,960665	0,793866	0,206134
	Russia	1,869236713	5,00285826	1,87464005	0	0	8,746735	2,957488	0,393891	0,606109
	Ukraine	1,440659353	11,1925618	3,729780025	11,50225	9,764242	37,6295	6,134289	0,81699	0,18301
							$\overline{x_j}$	5,376248		
							S_o	1,066076		
							Co	7,508399		

Fig. 4.25. **Defining the taxonomic level of development**

The results of the calculation are presented in Table 4.7

Table 4.7

The value of the general indicator of energy security

Countries	d
Russia	0.606109
The UK	0.553171
Poland	0.426477
Germany	0.325684
Spain	0.324643
Italy	0.263021
Austria	0.258544
Belgium	0.234586
Finland	0.207969
Belorussia	0.206134
Bulgaria	0.19756
Ukraine	0.18301
Switzerland	0.180162
France	0.162821
Sweden	0.129641

As can be seen from Table 4.7, Ukraine's indicator of energy security is not the worst among the selected countries, but it is an outsider. Countries with lower energy security do not extract coal, or extract it in small quantities. That is, it can be concluded that Ukraine does not use its resources efficiently with sufficiently high own reserves of energy resources, which is confirmed by the fact that Ukraine has the worst indicator of energy intensity of the GDP among the European countries.

Application of the additive method when the total rating of the object is expressed by the total sum of all elements of the line. It is possible to use weight coefficients for each column, which characterizes the significance of the ranking on this basis. The final rating number is calculated based on formula 4.15:

$$R_i = \sum_{i=1}^{n} \alpha_i \cdot R_{ij}, \qquad (4.15)$$

where R_i is the final rating;

R_{ij} is the individual rating of the objects of observation;

α_i is the weighted coefficient determined by the expert method.

For each of the indicators it is necessary to set the rating (Table 4.8). The procedure is appropriate to the calculations performed in Fig. 4.6.

Table 4.8

The rating score for each of the indicators of the energy security

No.	Countries	X1	X2	X3	X4	X5
1	Austria	8	4	9	10	5
2	Belgium	5	3	8	15	6
3	Belorussia	12	13	12	14	3
4	Bulgaria	2	12	7	7	14
5	Finland	13	10	13	11	7
6	France	11	2	10	6	15
7	Germany	4	7	4	9	10
8	Italy	3	6	11	13	4
9	Poland	6	11	1	3	9
10	Russia	10	14	3	1	1
11	Spain	7	8	2	12	13
12	Sweden	15	5	14	4	12
13	Switzerland	14	1	15	5	11
14	The UK	1	3	6	2	2
15	Ukraine	9	15	5	8	8

The next step is to determine the weighting coefficients that are presented by the expert method. To determine the weighting factors, the ranking method is used. This method implies that each expert is asked to place the signs in the order of preference. One denotes the most important sign, number two follows it in the importance, etc. An expert survey is conducted based on the significance of each of the indicators. Five experts must distribute 15 points among 5 indicators. The obtained data on the significance of the feature for assessing the level of energy security are summarized in Table 4.9.

Table 4.9

The expert assessments

	Expert $_1$	Expert $_2$	Expert $_3$	Expert $_4$	Expert $_5$
X1	1	2	1.5	1	2
X2	2.5	2	1.5	2.5	1
X3	2.5	2	3	2.5	3
X4	4	5	4.5	4.5	4
X5	5	4	4.5	4.5	5

Before summarizing the views of experts, we will check them for consistency, which is a prerequisite for the quality of the evaluations. The consistency of expert opinions will be checked using the coefficient of rank correlation of Kendall, Spearman and the coefficient of concordation in the Statistica 8.0 package.

In order to calculate the correlation coefficient of Kendall and Spearman, it is necessary to present the data in the format shown in Fig. 4.26.

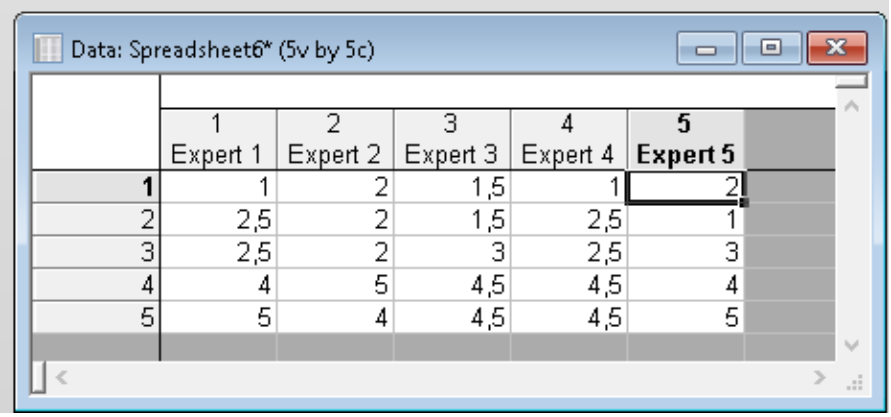


Fig. 4.26. **The input data in the package Statistica 8.0**

In the *Nonparametric Statistics* module, choose the *Correlation* menu, which allows you to calculate the rank correlation coefficients. The figures show the procedure for calculating the Spearman rank correlation coefficient (Fig. 4.27 – 4.28).

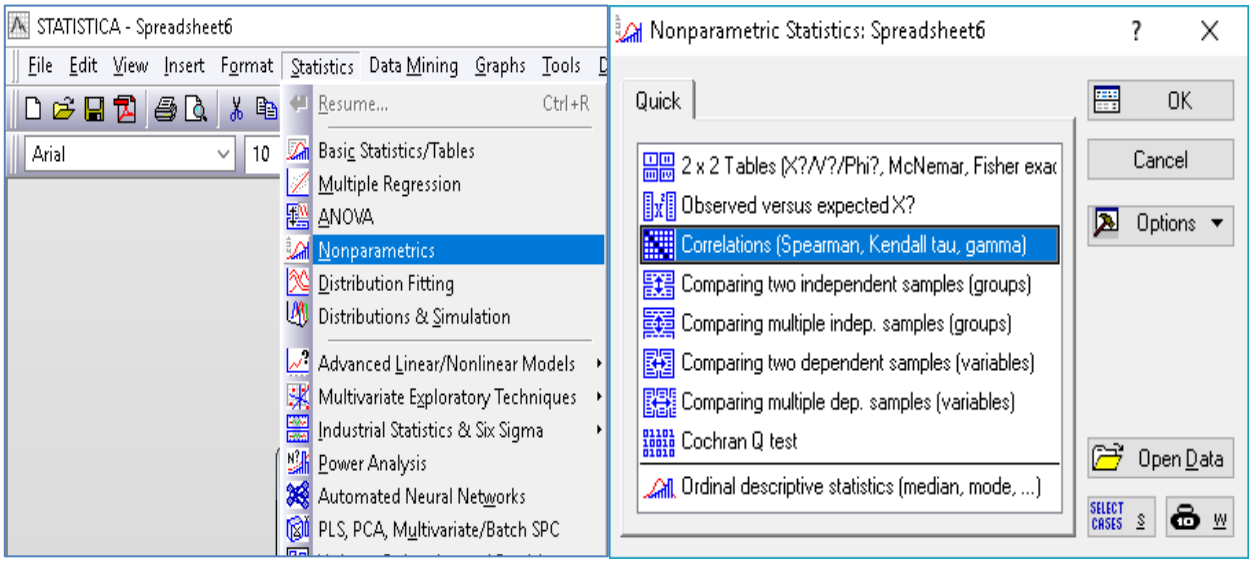


Fig. 4.27. **Choosing the parameters in Statistica 8.0**

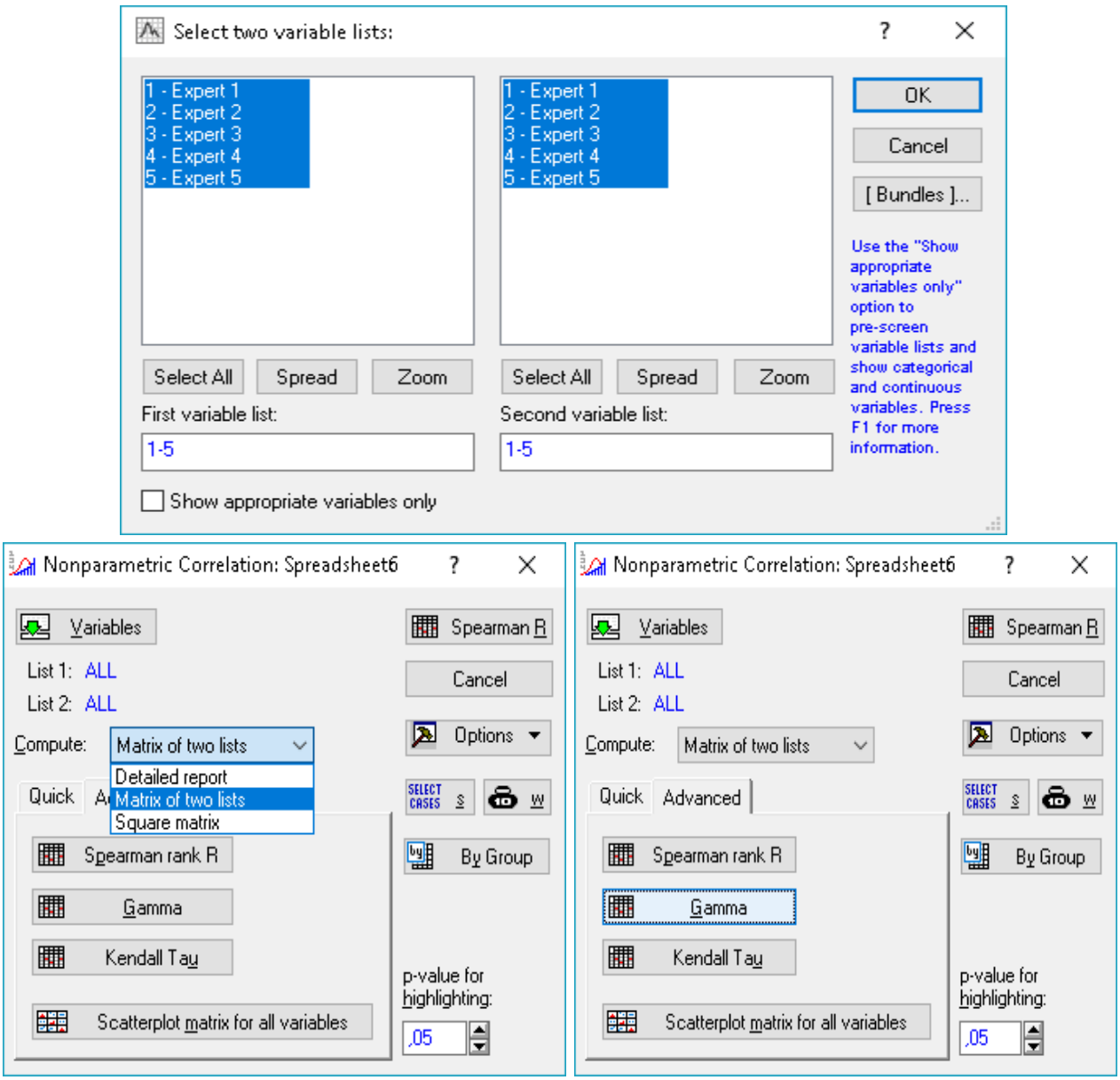


Fig. 4.28. **Choosing the parameters of the nonparametric correlation**

The matrix of the Kendall rank correlation coefficient is shown in Fig. 4.29.

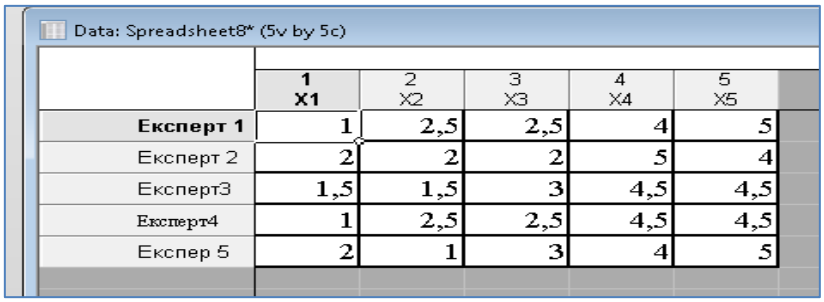
Gamma Correlations (Spreadsheet6)
MD pairwise deleted
Marked correlations are significant at p <,05000

Variable	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5
Expert 1	1,000000	0,714286	1,000000	1,000000	0,777778
Expert 2	0,714286	1,000000	1,000000	1,000000	0,714286
Expert 3	1,000000	1,000000	1,000000	1,000000	1,000000
Expert 4	1,000000	1,000000	1,000000	1,000000	0,750000
Expert 5	0,777778	0,714286	1,000000	0,750000	1,000000

Fig. 4.29. **The matrix of the Kendall rank correlation coefficient**

The figure shows the Spearman rank correlation matrix from which it is evident that most experts have a coherent opinion about the research issues.

To calculate the coefficient of concordation, you must present the data as shown in the Fig. 4.30.



	1 X1	2 X2	3 X3	4 X4	5 X5
Експерт 1	1	2,5	2,5	4	5
Експерт 2	2	2	2	5	4
Експерт3	1,5	1,5	3	4,5	4,5
Експерт4	1	2,5	2,5	4,5	4,5
Експер 5	2	1	3	4	5

Fig. 4.30. **Transformation of the input data**

Use the *Nonparametric Statistics* module to select the *Comparing multiple dependent samples* menu. Sample and proceed to Friedman's rank dispersion analysis and calculate the Kendall concordance coefficient (Fig. 4.31).

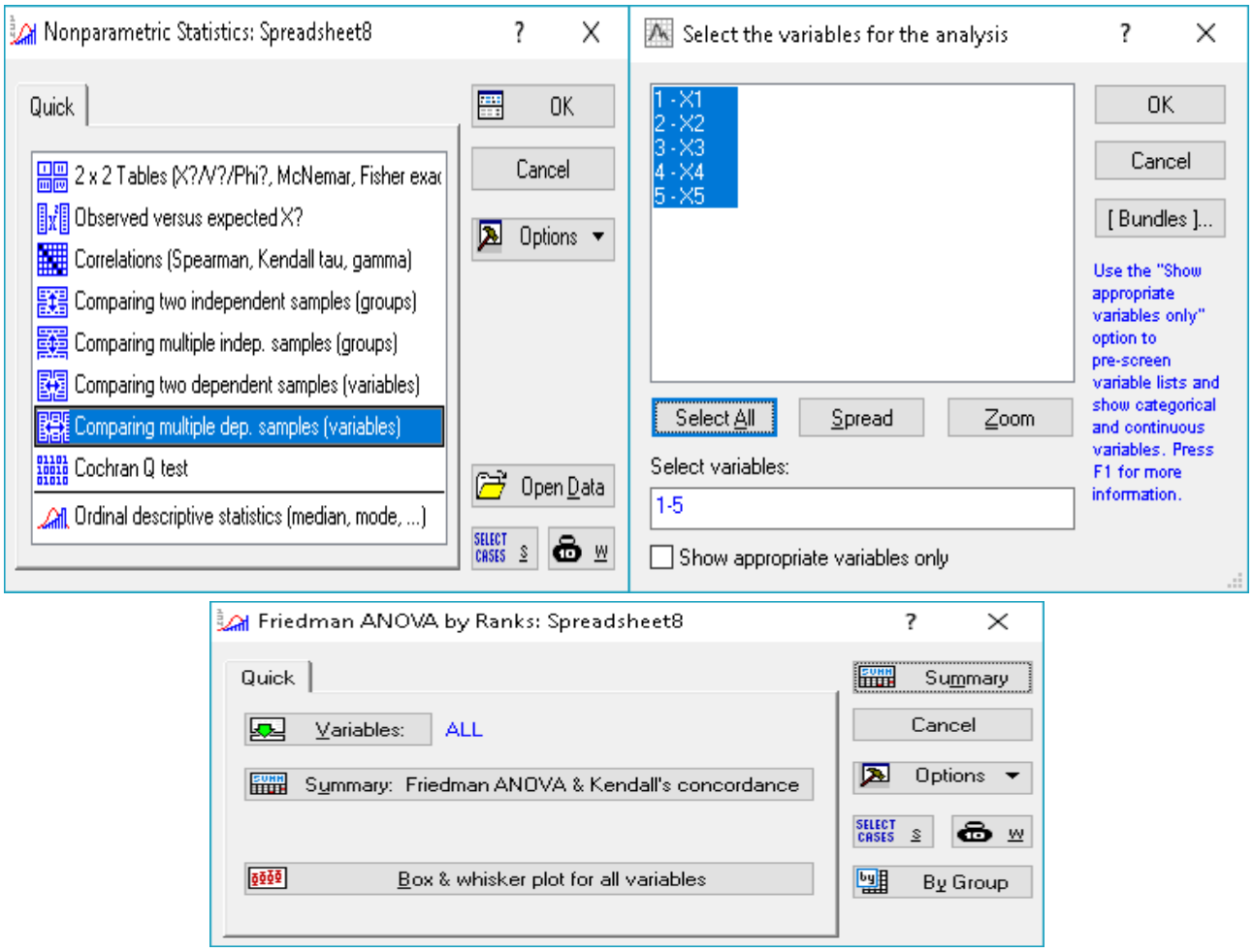
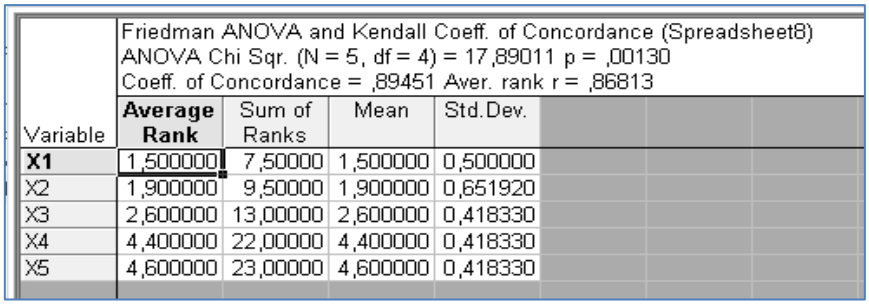


Fig. 4.31. **Choosing the parameters**

Fig. 4.32 shows the results of calculating the Kendall concordance coefficient and assessing its significance based on the criterion X2. With a probability of 0.13 %, the coefficient of concordance is significant, that is, experts have a coherent opinion.



Friedman ANOVA and Kendall Coeff. of Concordance (Spreadsheet8)
ANOVA Chi Sqr. (N = 5, df = 4) = 17,89011 p = ,00130
Coeff. of Concordance = ,89451 Aver. rank r = ,86813

Variable	Average Rank	Sum of Ranks	Mean	Std.Dev.
X1	1,500000	7,50000	1,500000	0,500000
X2	1,900000	9,50000	1,900000	0,651920
X3	2,600000	13,00000	2,600000	0,418330
X4	4,400000	22,00000	4,400000	0,418330
X5	4,600000	23,00000	4,600000	0,418330

Fig. 4.32. **The results of calculation of the Kendall concordance coefficient**

Since the expert opinion is agreed, the results of the examination can be used for further calculations.

With the help of the methods of mathematical statistics it is necessary to get a general opinion of experts. The average rank is determined, the mean statistical significance of the S_j *j*-th characteristic is calculated as follows (4.16):

$$S_j = \frac{\sum_{j=1}^{m} \alpha_{ij}}{m_{kj}}, \qquad (4.16)$$

where m_{kj} is the number of evaluators of the *j*-th feature ($m_k \le m$);

i is the number of expert; $i = 1, ..., m$;

j is the sign number; $j = 1, 2, ..., n$.

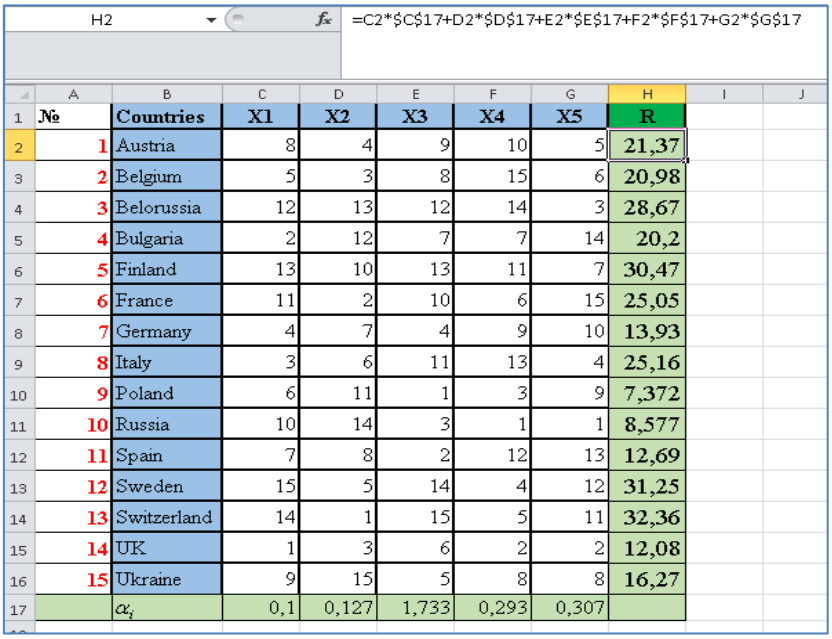
The calculation results are shown in Table 4.10.

Table 4.10

Calculation of the average rank

	E_1	E_2	E_3	E_4	E_5	S (average value)	α_i (the proportion of the average rank)
X1	1	2	1.5	1	2	1.5	1.5/15 =0.1
X2	2.5	2	1.5	2.5	1	1.9	0.1267
X3	2.5	2	3	2.5	3	2.6	1.733
X4	4	5	4.5	4.5	4	4.4	0.293
X5	5	4	4.5	4.5	5	4.6	0.307
Sum						15	

Using formula 4.15 for calculation of the overall rating in Fig. 4.33, the rating is calculated for each of the countries.



№	Countries	X1	X2	X3	X4	X5	R
1	Austria	8	4	9	10	5	21,37
2	Belgium	5	3	8	15	6	20,98
3	Belorussia	12	13	12	14	3	28,67
4	Bulgaria	2	12	7	7	14	20,2
5	Finland	13	10	13	11	7	30,47
6	France	11	2	10	6	15	25,05
7	Germany	4	7	4	9	10	13,93
8	Italy	3	6	11	13	4	25,16
9	Poland	6	11	1	3	9	7,372
10	Russia	10	14	3	1	1	8,577
11	Spain	7	8	2	12	13	12,69
12	Sweden	15	5	14	4	12	31,25
13	Switzerland	14	1	15	5	11	32,36
14	UK	1	3	6	2	2	12,08
15	Ukraine	9	15	5	8	8	16,27
	α_i	0,1	0,127	1,733	0,293	0,307	

Fig. 4.33. **Estimation of the world rankings based on the level of energy security using expert assessments**

We construct a ranking of countries based on the level of energy security (Table 4.11).

Table 4.11

The ranking of countries based on the level of energy security

No.	Countries	R
1	Poland	7.372
2	Russia	8.577
3	UK	12.08
4	Spain	12.69
5	Germany	13.93
6	Ukraine	16.27
7	Bulgaria	20.2
8	Belgium	20.98
9	Austria	21.37
10	France	25.05
11	Italy	25.16
12	Belorussia	28.67
13	Finland	30.47
14	Sweden	31.25
15	Switzerland	32.36

The analysis of the calculation of the ranking by the three methods has proved that the introduction of the weighting factors corresponding to the preferences of experts makes it possible to overcome the shortcomings of the first two methods and get a fairly adequate assessment of the situation.

Tasks for independent work

According to the selected research topic, use all the appropriate analytical methods to analyze the past, current and future situation.

The results of the assignments should be presented in the form of a report on the laboratory work.

The list of questions for independent work

1. Name the main analytical methods of data processing.
2. What expert data processing methods do you know?
3. Name the basic methods of processing the digital information.
4. Describe the analytical methods of data processing.
5. Describe the expert methods of data processing.
6. Compare the advantages and disadvantages of using the analytical and expert methods for processing the digital information.
7. What are the main causes of inaccurate expert assessments?
8. What software products are used to visualize digital information?
9. List the shortcomings (limitations) in the application of expert methods.
10. What are the features of the combination of analytical and expert methods?

Test questions for self-assessment on topic 4

1. Which module of *Data Analysis* in MS Excel is used for grouping of statistical data:
   a) *Histogram*;
   b) *Regression*;
   c) *Descriptive statistics*?

2. The main indicators of time series analysis are:
   a) the absolute increase, the growth rate, the rate of increase;
   b) frequency, cumulative frequency.

3. How is the weight coefficient determined with the additive ranking method:

a) by using Spearman's rank correlation coefficient;
b) by using the average value of the indicator;
c) by using the expert method?

4. The Kendall's rank correlation coefficient is used to:

a) standardize the values;
b) determine the relationship between two rank variables.

5. Which module of Statistics 8.0 package is used for determining Kendall's and Spearman's rank correlation:

a) *Multiple Regression*;
b) *Nonparametrics*?

6. The correlation coefficient (r) can take values in the range:

a) $-1 \le r \le 1$;
b) $-1 \le r \le 0$;
c) $0 \le r \le 1$.

7. The value of the correlation coefficient $r = 1$ indicates:

a) the strongest possible agreement;
b) the strongest possible disagreement.

8. For standardization of the initial data the following formula can be used:

a) $z_{tj} = \frac{x_{ij} - \overline{x_j}}{s_j}$;

b) $R_i = \sum_{i=1}^{n} \alpha_i \times R_{ij}$;

c) $S_j = \frac{\sum_{j=1}^{m} \alpha_{ij}}{m_{kj}}$.

9. What built-in features of Excel can be used to make a forecast:

a) trend, prediction, trend line;

b) trend line, trend, standardization;

c) prediction, trend, extrapolation?

10. What coefficient is used to determine the consistency of expert opinions:

a) Kendall's coefficient;

b) Pearson's coefficient;

c) The coefficient of Spearman?

Topic 5. Economic interpretation of visual analysis of data. Preparation of reports

Laboratory work 5

The purpose is to get acquainted with the rules of the preparation of the analytical report, to get knowledge about the structure and main elements of the analytical report.

The task is to make an analytical report on the study of the chosen industry (sphere) of the economy or another research theme.

Guidelines

An analytical report is a document describing a thorough investigation of a particular issue. Firstly, it is a report on the work done to study and summarize the novelty experience, and secondly, it is an analytical report that involves primarily an analysis of the materials obtained as a result of the work performed rather than a list of what was done.

An analytical report goes beyond just focusing on the result. This kind of report presents the results, analyzes the results, and provides a conclusion based on the results. The purpose of an analytical report is to suggest an option, an action plan, and a procedure.

Analytical reports consist of technical business writing that typically communicates a solution to a problem. The reports contain well-documented

research, and they synthesize a plethora of information to draw informed conclusions. Although technical, they are crafted to be easy to access and understand, and they often include visualization, analysis, recommendations and action plans.

So let's take a closer look at what an analytical report is for. So, the functions of the analytical report are:

1. Visibility.
2. Analysis.
3. Innovativeness.

1. Visibility. Since the analytical report contains a huge number of graphs, charts and diagrams, one can judge its visibility. All information is in full view which allows you to make the right decision faster.

2. Analysis. It turns out a little tautological function, but it is thanks to analytical reports that the analysis of the performance of a particular specialist or enterprise is carried out. With the worker everything is clear.

If he improperly copes with his functions, he can be kicked out. But this happens rarely. As a rule, mistakes are made in his activities and in the organization as a whole, which are later corrected.

3. Innovativeness. Yes, innovation is appreciated in any company. And it is no wonder, since the introduction of something new in a large number of cases involves certain breakthroughs. Naturally, this does not always happen. Innovations should be introduced in the areas they are needed. Only in this case they make sense. And the analytical report allows you to clearly demonstrate what and how it works.

The terms of the use of the analytical report are as follows:

a) use in the process of collecting the material about the experience of different sources and methods of research, which allows you to obtain diverse material that requires the establishment of links and dependencies;

b) insufficient evidence of the productivity of the generalized experience, which requires further in-depth study;

c) insufficient scientific validity and explanation of the experience, according to the purposes of the research. The analytical report allows preserving the material's materiality and its openness for further research.

Formatting the report. While analytical reports may vary slightly based on the need and audience, they often share common elements: a title page, a table of contents, an introduction, a methodology section, body sections, conclusions and recommendations, a bibliography, and an appendices section.

Each section is noted by a heading, and subheadings are used when necessary. Most analytical reports follow scientific style.

An analytical report is written in strict scientific language. Tables, graphs, schemes are used in it, which is not only a means of presenting and describing the experience, but also a means of analyzing it and demonstrating the effectiveness of the results.

In the analytical report, the main thing is to choose the type of analysis of the material collected in the study of the experiment. It can be comparative, problematic, critical, correlational, systemic and other types of analysis.

Gathering the information. Writing analytical reports goes beyond descriptive writing toward synthesis and critique. To achieve this end, begin by collecting data and gathering resources. The research process should produce sufficient information so that you feel comfortable with all aspects of the concept: the pros and cons, counter arguments, and proposals. Quality research will enable you to analyze the information and put forth quality recommendations and solutions to problems. Research should be compiled from high-quality, industry-relevant sources.

Organizing, analyzing, synthesizing. After collecting the appropriate research, synthesize the information. Synthesis involves critiquing a source's argument, validity or methodology based on your own research and findings in an effort to present new information, draw conclusions, or present findings. Report information accurately and in context. Synthesized research should result in clear and logical findings, with recommendations if called for. In this stage, you should also determine how to organize and present the information: chronologically, geographically, spatially, categorically, or by importance or comparison.

Conclusions and recommendations. Once the data has been collected, the information has been processed, and conclusions have been drawn, the composition process begins. While the bulk of the report will present and analyze your findings, most reports focus on one of three elements: conclusions, an argument's logic, or recommendations.

Guidelines on the structure of the report

A single analytical report structure includes:

- a title page,
- a table of contents,

- an introduction,
- a methodology section,
- body sections,
- conclusions and recommendations,
- a bibliography,
- an appendices section.

Each section should contain a headline, a subhead and text. All pages should be numbered. Seems like this is not much, but let us take a closer look at each point.

Title page. Here you should put topic details of your research and all authors' names. It is a front page, a so-called "face" of your work.

Table of content. Arrange report's structure marking section names together with page numbers where they start.

Introduction. Here is where real work starts. Here authors have to state urgency of the research, reasons for choosing their topics, the main aims, the methods they used for investigations. Do not forget to analyze all sources that were used for your report.

Methodology section. Describe the methods you have used for getting the results. This section should not be too large, but it is good to mention the origins of all methods applied for the study.

Body sections. This is the report itself. It should include several chapters. Each should be titled separately. The structure of the whole department should be logical and each subsection must be a continuation of the previous one.

Mind your grammar and watch the sentence structure. It is wise to use short constructions to avoid bulkiness. Short messages are easier to understand and help avoid confusion with data.

Conclusions and recommendations. Here the authors should put a final statement explaining the result of their work. Point the conclusions you have arrived to.

Bibliography. This section should include all sources that were used for the report. Names of books, articles, other paper works, publishing years, authors' names, web addresses if permitted.

Appendix. All attachments should go to the appendix. These are the materials you referred to in your body: all calculations, graphs, diagrams, large information blocks that clarify all research stages described in the paper.

How to write an effective analytical report

The success of your analytical report is a combination of accurate content discussion, proper formatting, strong and appropriate recommendations and overall document development. That being said, here is how you can write an effective analytical report:

- Gather the information that you need for your analytical report to be credible. The data that you have collected can greatly affect the effectiveness of the discussion that you will have.
- Identify the resources that you can look into if you need to further list down details which are not identified in your initial information collection.
- Assess the information that you have collected. If you want to have a thorough synthesis of this information, you have to use metrics that can identify the validity of the sources' claims, arguments, suggestions, and other findings.
- Properly report the items that you have observed. Your analytical report must be based on the actual context of your gathered information. You have to be accurate when doing this as wrong information can affect the entire document.
- Present the information in a way that it will be best used by the people to whom the report is for. You can either base the organization of content depending on the importance of the discussion, the outline example that you would like to achieve or the comparison that you would like to highlight.
- Do not forget about data visualization. This is the most effective way to present information and present the results of the analysis.
- Come up with an appealing and strong recommendation with the help of the conclusions that you have drawn. Your report should end in a way that stakeholders can be influenced to decide or look into the matter as to how you would like them to.

Tasks for independent work

According to the selected research topic, prepare an analytical report which must consist of a title page, a table of contents, an introduction, a methodology section, body sections, conclusions and recommendations, a bibliography and an appendices section.

The results of the assignments should be presented in the form of an analytical report on the chosen research topic.

The list of questions for independent work

1. Name the most common ways to visualize data.
2. Name the basic methods and methods of processing the text and digital information.
3. What is the essence and main stages of visualization of digital information?
4. List the properties of digital information.
5. Name the features and difficulties that an analyst encounters when interpreting the results of a visual analysis of data.
6. What are the features of economic interpretation of the results?
7. What is an analytical report?
8. What are the stages of the preparation of the analytical report?
9. What major mistakes are made during the preparation of the analytical report?
10. What visual effects may improve the visibility of the perception of the analytical report?

Test questions for self-assessment on topic 5

1. The economic interpretation of the research results is provided in:
   a) the conclusion;
   b) the introduction;
   c) the main part of the report.

2. Infographics should not be used in:
   a) conclusions;
   b) subdivisions;
   c) appendices of the report.

3. The form and method of presentation of the analytical report depends on:
   a) direct users;
   b) availability of special devices;
   c) limited presentation time.

4. What is an analytical report:
   a) a document that describes a detailed study of a particular issue (problem);
   b) a report on the work done on the study and synthesis of innovative experience;

c) a report that involves above all an analysis of the materials obtained as a result of the work performed rather than just a list of what was done;

d) all answers mentioned?

5. What is the nature of the presentation of the material of an analytical report:

a) scientific;

b) journalistic;

c) arbitrary?

6. What kind of analytical report is a portfolio report:

a) a report of a political and practical nature;

b) a corporate report;

c) a monitoring and evaluation report;

d) a report of the environment scan?

7. What kind of analytical report is a progress report:

a) a report of political and practical nature;

b) a corporate report;

c) a monitoring and evaluation report;

d) a report of the environment scan?

8. The functions of the analytical report include:

a) monitoring, processing and presentation;

b) visibility, analysis and innovation;

c) analysis, synthesis and visibility.

9. How many obligatory elements are included in the structure of the analytical report:

a) 8:

b) 7;

c) 6?

10. The final date of the preparation of a report is indicated in

a) the conclusions;

b) the cover sheet

c) the last section of the report;

d) the last appendices sheet.

References

Main

1. Когнитивная бизнес-аналитика : учебник / под науч. ред. д-ра техн. наук, проф. Н. М. Абдикеева. – Москва : ИНФРА-М, 2011. – 511 с.

2. Марець О. Р. Представлення статистичної інформації за допомогою графічного методу / О. Р. Марець, О. М. Вільчинська // International Scientific Journal. – 2015. – № 9. – С. 118–125.

3. Основы бизнес-анализа : учебное пособие / В. И. Бариленко, В. В. Бердников, Р. П. Булыга [и др.] ; под ред. В. И. Бариленко. – Москва : КНОРУС, 2016. – 272 с.

4. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям / Н. Б. Паклин, В. И. Орешков. – Питер : б. и., 2013. – 706 с.

5. Blazquez D. Big Data sources and methods for social and economic analyses / D. Blazquez, J. Domenech // Technological Forecasting and Social Change. – 2018. – Vol. 130. – P. 99–113.

6. Few S. Information Dashboard Design: The Effective Visual Communication of Data / S. Few. – Sabastopol, California : O'Reilly Media, 2006. – 206 p.

7. Few S. Show Me the Numbers: Designing Tables and Graphs to Enlighten / S. Few. – Oakland, California : Analytics Press, 2004. – 263 p.

8. Schwabish J. A. An Economist's Guide to Visualizing Data / J. Schwabish // The Journal of Economic Perspectives. – 2014 – Vol. 28, No. 1, Winter. – P. 209–234.

9. Tufte E. R. Visual Display of Quantitative Information / E. R. Tufte. – 2nd ed. – Cheshire, Connecticut : Graphics Press, 2001. – 191 p.

10. Ware C. Visual Thinking for Design / C. Ware. – Burlington : Morgan Kaufmann, 2008. – 183 p.

Additional

11. Comparative Benchmarking of European and US Research Collaboration and Researcher Mobility / A report prepared in collaboration between Science Europe and Elsevier's SciVal Analytics. – S. l. : s. n., 2013. – 48 p.

12. Thomas James J. Illuminating the Path: The R&D Agenda for Visual Analytics / J. James Thomas, Kristin A. Cook. – S. l. : Visualization and Analytics Center, 2005. – 30 p.

Information resources

13. Газін А. Візуалізація даних як навичка / А. Газін [Електронний ресурс]. – Режим доступу : https://aiukraine.com/wp–content/uploads/2017/10/1_4–Gazin.pdf.

14. Офіційний сайт департаменту статистики Організації Об'єднаних Націй. – Режим доступу : http://unstats.un.org/unsd/default.htm.

15. Офіційний сайт державної служби статистики України. – Режим доступу : http://www.ukrstat.gov.ua.

16. Bubbl [Electronic resource]. – Access mode : https://bubbl.us.

17. Cele N. Report Writing and Data Visualization [Electronic resource] / N. Cele, H. Visser // Institutional Research Institute 2015. – Access mode : http://www.saair-web.co.za/wp-content/uploads/2015/08/5-NCHV-Report-writing-and-data-visualization.pdf.

18. Coggle [Electronic resource]. – Access mode : https://coggle.it.

19. Datawrapper [Electronic resource]. – Access mode : https://www.datawrapper.de.

20. How to design better data visualizations? [Electronic resource]. – Access mode : https://www.creativebloq.com/how-design-better-data-visualisations-8134175.

21. How to Make Mind Maps. Visualize Your Ideas for Better Brainstorming [Electronic resource]. – Access mode : https://zapier.com/blog/mind-mapping-tutorial.

22. Introduction to R. Grammar of Graphics [Electronic resource]. – Access mode : https://ramnathv.github.io/pycon2014–r/visualize/ggplot2.html.

23. Mapul mind mapping [Electronic resource]. – Access mode : https://www.mapul.com.

24. Mindmeister [Electronic resource]. – Access mode : https://www.mindmeister.com.

25. The Analysis Factor [Electronic resource]. – Access mode : http://www.theanalysisfactor.com.

26. XMind [Electronic resource]. – Access mode : https://www.xmind.net.

Content

НАВЧАЛЬНЕ ВИДАННЯ

ВІЗУАЛІЗАЦІЯ ДАНИХ ТА ВІЗУАЛЬНА АНАЛІТИКА

Методичні рекомендації до лабораторних та самостійних робіт

для студентів спеціальності 122 "Комп'ютерні науки" другого (магістерського) рівня

(англ. мовою)

Самостійне електронне текстове мережеве видання

Укладачі: **Дериховська** Вікторія Ігорівна
Шликова Вікторія Олександрівна

Відповідальний за видання *О. В. Раєвнєва*

Редактор *З. В. Зобова*

Коректор *З. В. Зобова*

Подано завдання до лабораторних та самостійних робіт із навчальної дисципліни та методичні рекомендації для їх виконання, що допоможуть студентам одержати теоретичні та практичні навички використання сучасних аналітичних засобів та інфографіки для аналізу складних масових соціально-економічних явищ і процесів.

Рекомендовано для студентів спеціальності 122 "Комп'ютерні науки" другого (магістерського) рівня.

План 2020 р. Поз. № 105 ЕВ. Обсяг 102 с.

Видавець і виготовлювач – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Науки, 9-А

Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.