

# АНАЛИЗ ДАННЫХ

## Тема 1. Данные в экономике, их визуализация и предварительная обработка

Доцент Соловьев А.И., кандидат технических наук, доцент

- ▶ Лекция 1.3 Визуализация качественных признаков в Microsoft Excel.  
Сводные таблицы и сводные диаграммы в Microsoft Excel.  
Таблицы сопряженности и парадокс Симпсона.  
Иерархия признаков в Microsoft Excel.

# В предыдущей лекции

## Инструменты описательной статистики в Microsoft Excel.

Описательная статистика

Измерение центра распределения.

Измерение разброса данных.

Описательная статистика в надстройке «Анализ данных» Microsoft Excel.

Диаграммы размаха в Microsoft Excel.

# 1. Визуализация качественных признаков в Microsoft Excel

В предыдущих лекциях мы рассмотрели визуализацию количественных признаков.

► Качественные данные так же могут быть визуализированы.

► При большом количестве данных важно понять с чего начать их анализ. В этом случае следует создать сводную таблицу или построить сводную диаграмму.

## 2. Сводные таблицы и сводные диаграммы в Microsoft Excel

**Сводная таблица** – это инструмент обработки данных, упрощающий их обобщение и визуализацию.

Если имеются исходные данные в форме таблицы, столбцы которой содержат однотипные или малоинформативные значения, то

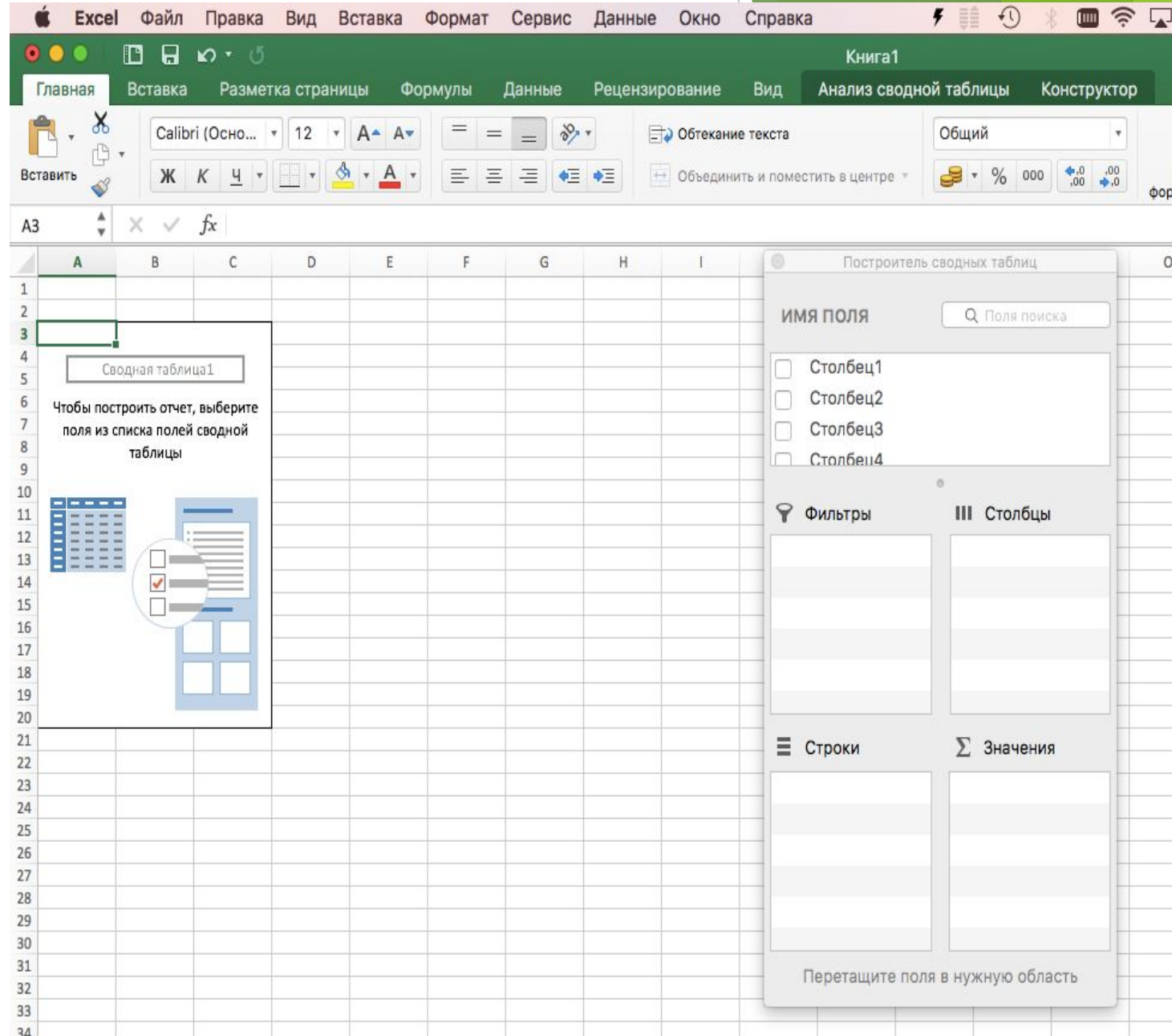
**сводная таблица** – это таблица, связанная с исходной таблицей и обобщающая исходные данные путем различных вычислительных процедур, например, вычисления сумм, средних, долей и т. п.

В EXCEL предусмотрено построение сводной таблицы:

- Сводная таблица создается при помощи нескольких действий с помощью «Построителя сводной таблиц», потом ее можно быстро настроить, чтобы выбрать наиболее информативное представление результатов.
- Построитель сводной таблицы находится во вкладке «ВСТАВКА»

## 2. Сводные таблицы и сводные диаграммы в Microsoft Excel

### ► Построитель сводной таблицы



The screenshot displays the Microsoft Excel interface with the PivotTable Wizard dialog box open. The dialog box is titled "Построитель сводных таблиц" and contains the following sections:

- ИМЯ ПОЛЯ:** A search box labeled "Поля поиска".
- Columns:** A list of fields with checkboxes: "Столбец1", "Столбец2", "Столбец3", and "Столбец4".
- Фильтры:** A section for selecting filter fields.
- Столбцы:** A section for selecting column fields.
- Строки:** A section for selecting row fields.
- Значения:** A section for selecting value fields.

At the bottom of the dialog box, there is a prompt: "Перетащите поля в нужную область".

In the background, a PivotTable is visible in the worksheet. The PivotTable is titled "Сводная таблица1" and contains the following text:

Чтобы построить отчет, выберите поля из списка полей сводной таблицы

The PivotTable is currently empty, with only the header row visible.

## 2. Сводные таблицы и сводные диаграммы в Microsoft Excel

- ▶ На основе сводной таблицы создается *сводная диаграмма*, с автоматическим обновлением при изменении данных.

Например:

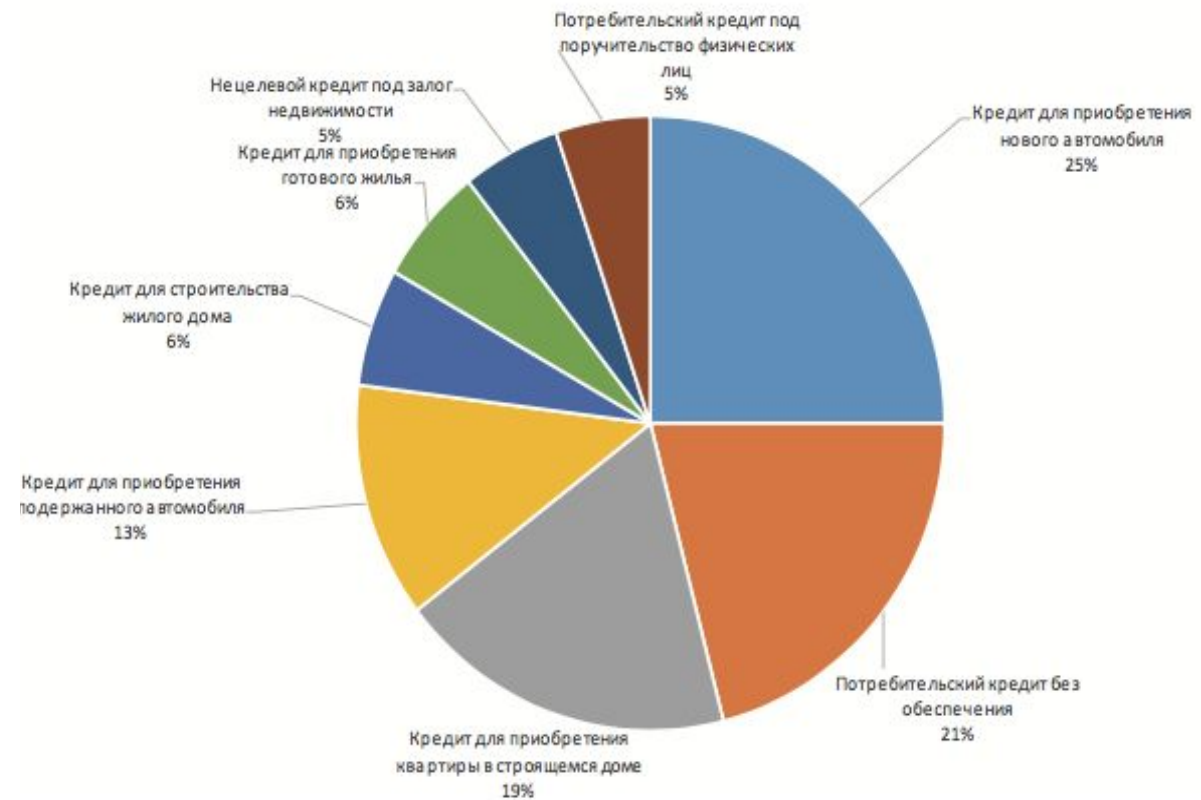
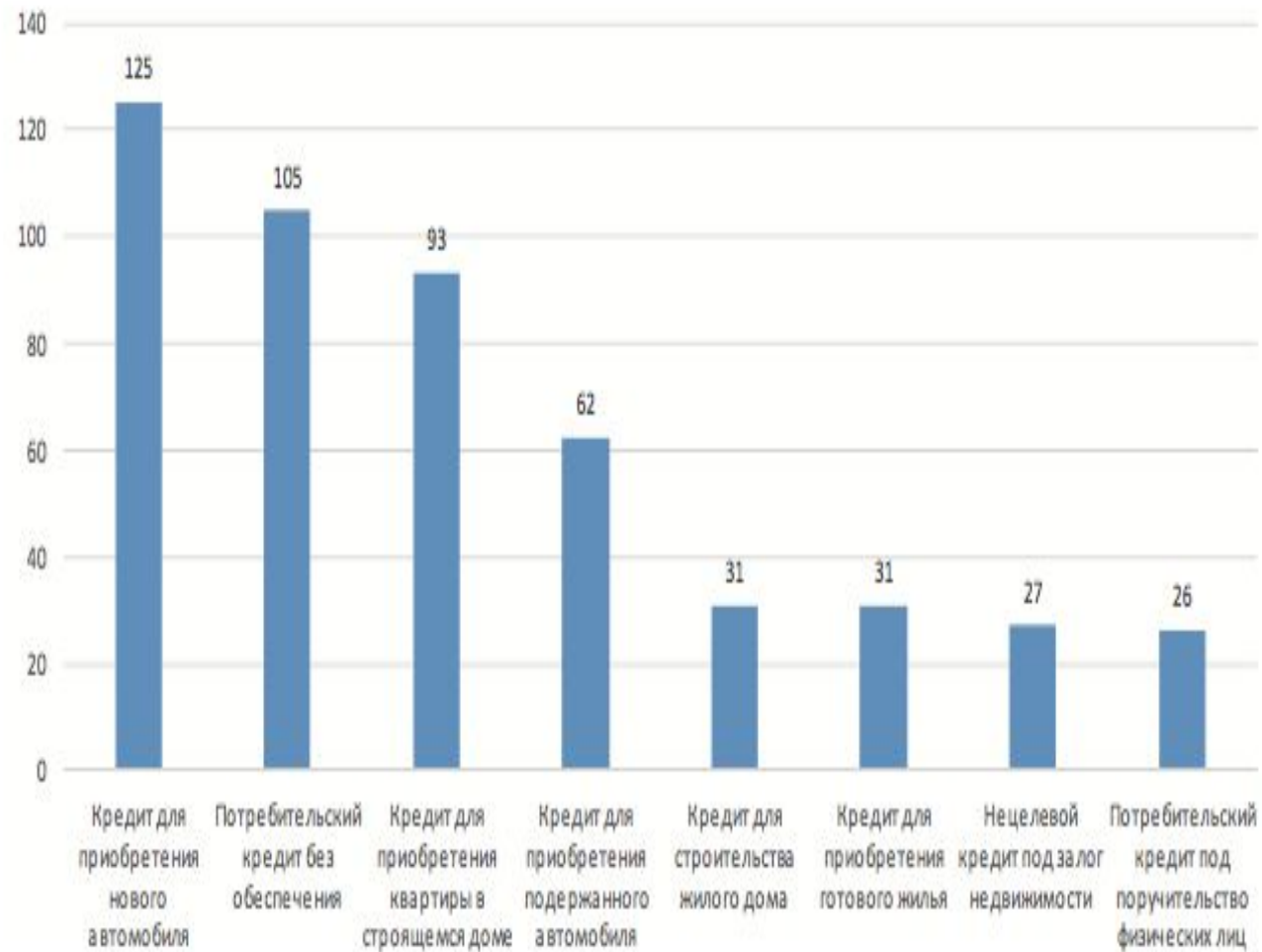
	A	B	C	D
1	Номер договора	Вид кредита	Названия строк	Количество по полю Вид кредита
2	1	Кредит для приобретения подержанного автомобиля	Кредит для приобретения нового автомобиля	125
3	2	Потребительский кредит без обеспечения	Потребительский кредит без обеспечения	105
4	3	Кредит для строительства жилого дома	Кредит для приобретения квартиры в строящемся доме	93
5	4	Кредит для приобретения подержанного автомобиля	Кредит для приобретения подержанного автомобиля	62
6	5	Потребительский кредит без обеспечения	Кредит для строительства жилого дома	31
7	6	Кредит для приобретения нового автомобиля	Кредит для приобретения готового жилья	31
8	7	Кредит для приобретения подержанного автомобиля	Нецелевой кредит под залог недвижимости	27
9	8	Кредит для приобретения нового автомобиля	Потребительский кредит под поручительство физических лиц	26
10	9	Кредит для строительства жилого дома	<b>Общий итог</b>	<b>500</b>
...	...	...		

Таблица служит основой для построения сводной диаграммы.



## 2. Сводные таблицы и сводные диаграммы в Microsoft Excel

- ▶ На основе сводной таблицы создается *сводная диаграмма*, с автоматическим обновлением при изменении данных.
- ▶ Воспользуемся диаграммами типа «ГИСТОГРАММА» и «КРУГОВАЯ»



### 3. Таблицы сопряженности и парадокс Симпсона

- ▶ Для анализа взаимосвязей между качественными признаками бывает полезно использовать *таблицы сопряженности*.
- ▶ *Таблица сопряженности* – это таблица, строки которой соответствуют значениям одного признака, столбцы соответствуют значениям другого признака, а в каждой ячейке указывается, сколько раз в наборе данных встречается соответствующая комбинация значений двух признаков.
- ▶ Для построения *таблицы сопряженности* проще всего воспользоваться инструментарием создания сводных таблиц.



### 3. Таблицы сопряженности и парадокс Симпсона

#### Таблица сопряженности.

	A	B	C
1	<b>Банк</b>	<b>Вид кредита</b>	<b>Статус</b>
2	A-банк	Автокредит	Погашен
3	A-банк	Автокредит	Погашен
4	A-банк	Автокредит	Погашен
5	A-банк	Автокредит	Не погашен
6	A-банк	Автокредит	Погашен
7	A-банк	Автокредит	Погашен
8	A-банк	Автокредит	Погашен
9	A-банк	Автокредит	Погашен
10	A-банк	Автокредит	Погашен
11	A-банк	Автокредит	Погашен
12	A-банк	Автокредит	Погашен
13	A-банк	Автокредит	Погашен
14	A-банк	Автокредит	Погашен
15	A-банк	Автокредит	Погашен
16	A-банк	Автокредит	Погашен
17	A-банк	Автокредит	Погашен
18	A-банк	Автокредит	Погашен
19	A-банк	Автокредит	Погашен
...	...	...	...
402	B-банк	Потребительский кредит	Погашен
403	B-банк	Потребительский кредит	Не погашен
404	B-банк	Потребительский кредит	Погашен

а)

	A	B	C	D	E	F	G
1	<b>Банк</b>	<b>Вид кредита</b>	<b>Статус</b>	<b>Количество по полю Статус</b>	<b>Названия столбцов</b>		
2	A-банк	Автокредит	Погашен	<b>Названия строк</b>	<b>Не погашен</b>	<b>Погашен</b>	<b>Общий итог</b>
3	A-банк	Автокредит	Погашен	A-банк	35	165	200
4	A-банк	Автокредит	Погашен	Автокредит	3	37	40
5	A-банк	Автокредит	Не погашен	Потребительский кредит	32	128	160
6	A-банк	Автокредит	Погашен	B-банк	34	169	203
7	A-банк	Автокредит	Погашен	Автокредит	27	154	181
8	A-банк	Автокредит	Погашен	Потребительский кредит	7	15	22
9	A-банк	Автокредит	Погашен	<b>Общий итог</b>	<b>69</b>	<b>334</b>	<b>403</b>
10	A-банк	Автокредит	Погашен				
11	A-банк	Автокредит	Погашен	<b>Количество по полю Статус</b>	<b>Названия столбцов</b>		
12	A-банк	Автокредит	Погашен	<b>Названия строк</b>	<b>Не погашен</b>	<b>Погашен</b>	<b>Общий итог</b>
13	A-банк	Автокредит	Погашен	A-банк	17,50%	82,50%	100,00%
14	A-банк	Автокредит	Погашен	Автокредит	7,50%	92,50%	100,00%
15	A-банк	Автокредит	Погашен	Потребительский кредит	20,00%	80,00%	100,00%
16	A-банк	Автокредит	Погашен	B-банк	16,75%	83,25%	100,00%
17	A-банк	Автокредит	Погашен	Автокредит	14,92%	85,08%	100,00%
18	A-банк	Автокредит	Погашен	Потребительский кредит	31,82%	68,18%	100,00%
19	A-банк	Автокредит	Погашен	<b>Общий итог</b>	<b>17,12%</b>	<b>82,88%</b>	<b>100,00%</b>
...	...	...	...				
402	B-банк	Потребительский кредит	Погашен				
403	B-банк	Потребительский кредит	Не погашен				
404	B-банк	Потребительский кредит	Погашен				

### 3. Таблицы сопряженности и парадокс Симпсона

#### Парадокс Симпсона.

- ▶ «Здравый смысл» нас иногда подводит:
- ▶ Предположим и автокредиты, и потребительские кредиты в банке-А возвращаются чаще, чем в банке-Б, то и в целом кредиты в банке-А возвращаются чаще, чем в банке-Б. Но, если посмотреть не на относительные, а на абсолютные значения, то можно заметить, что в банке-А выдается больше потребительских кредитов, тогда как банк-Б больше специализируется на автокредитах.
- ▶ Причина парадокса Симпсона заключается в некорректном усреднении. Интуитивно предполагается, что при усреднении найденных зависимостей соотношение количества выданных автокредитов и потребительских кредитов в двух банка одинаково, поэтому можно просто вычислить средние проценты невозврата кредитов в двух банках. Однако это не так.
- ▶ Задача об обслуживании клиентов 2 операторами Call-центра.

## 4. Иерархия признаков в Microsoft Excel.

В данных имеется определенная *иерархия признаков*, когда все множество объектов разделяется на какие либо категории, каждая категория, в свою очередь, делится на подкатегории и т. д.

Распознать такую *иерархию признаков* можно с помощью визуализации данных.

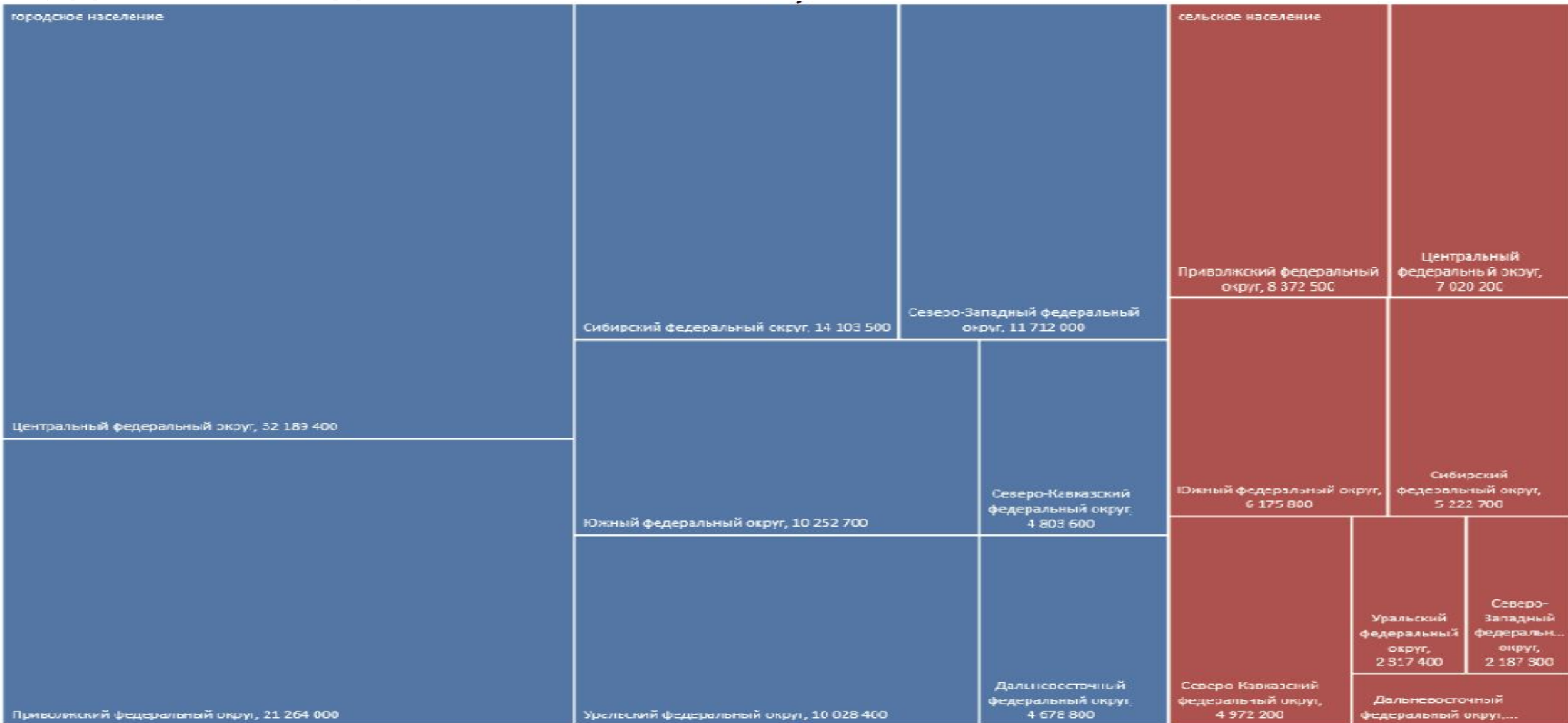
### А) ГИСТОГРАММА С НАКОПЛЕНИЕМ



## 4. Иерархия признаков в Microsoft Excel.

Распознать такую *иерархию признаков* можно с помощью визуализации данных.

### Б) ДИАГРАММА «ДЕРЕВО»

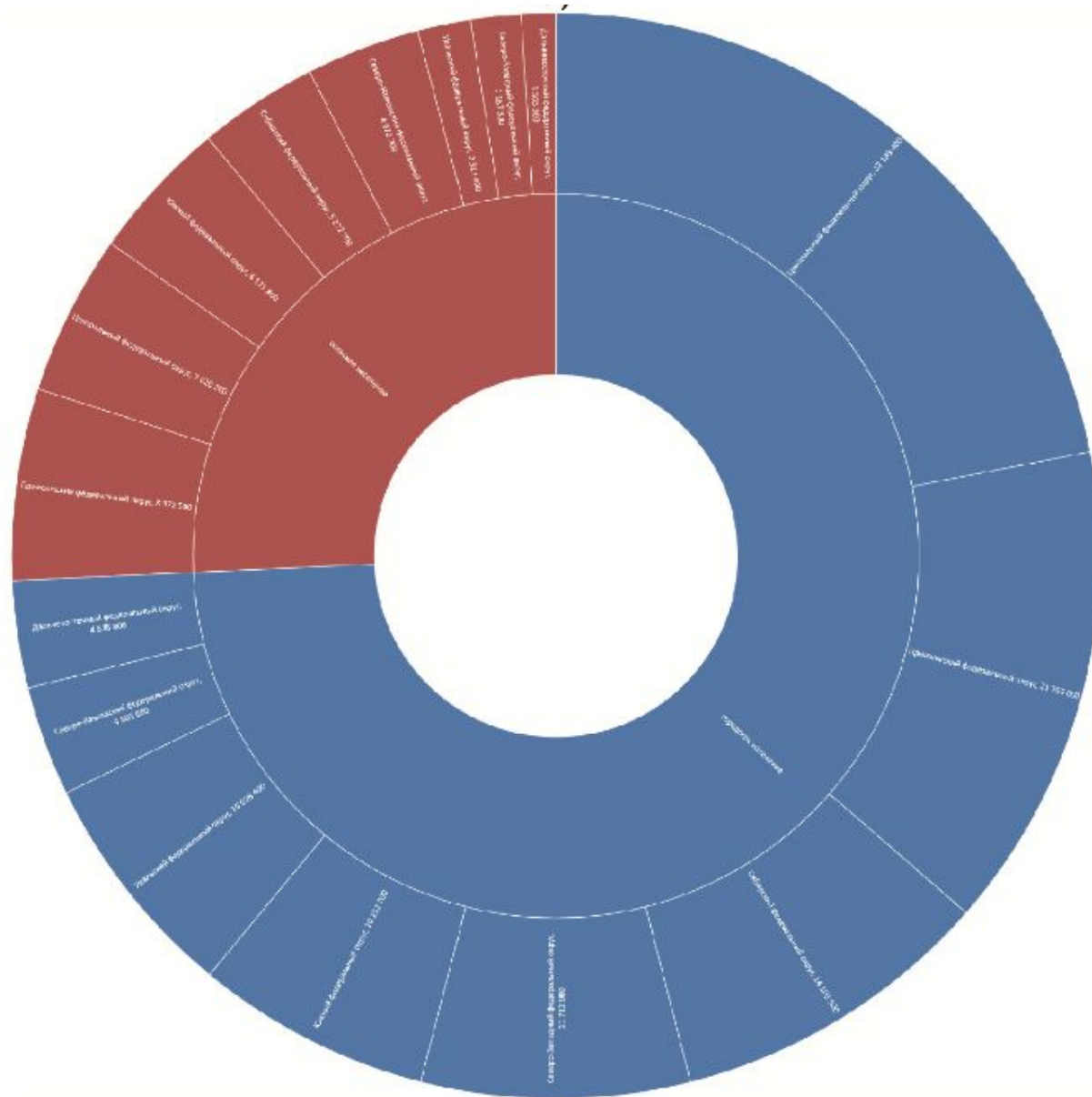




## 4. Иерархия признаков в Microsoft Excel.

Распознать такую *иерархию признаков* можно с помощью визуализации данных.

### Б) ДИАГРАММА «СОЛНЕЧНЫЕ ЛУЧИ»



## 5. Семинар 1.3

1. **СВОДНАЯ ТАБЛИЦА** и инструменты EXCEL для построения сводной таблицы:
2. **Построитель Сводной таблицы.**
3. **СВОДНАЯ ДИАГРАММА** и инструменты EXCEL для построения Сводной диаграммы.
4. **Способы визуализации данных СВОДНОЙ ТАБЛИЦЫ и СВОДНОЙ ДИАГРАММЫ.**

**Задание: решение задач по плану преподавателя.**