Финансовый университет при Правительстве РФ
Департамент анализа данных, принятия решений и финансовых технологий

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

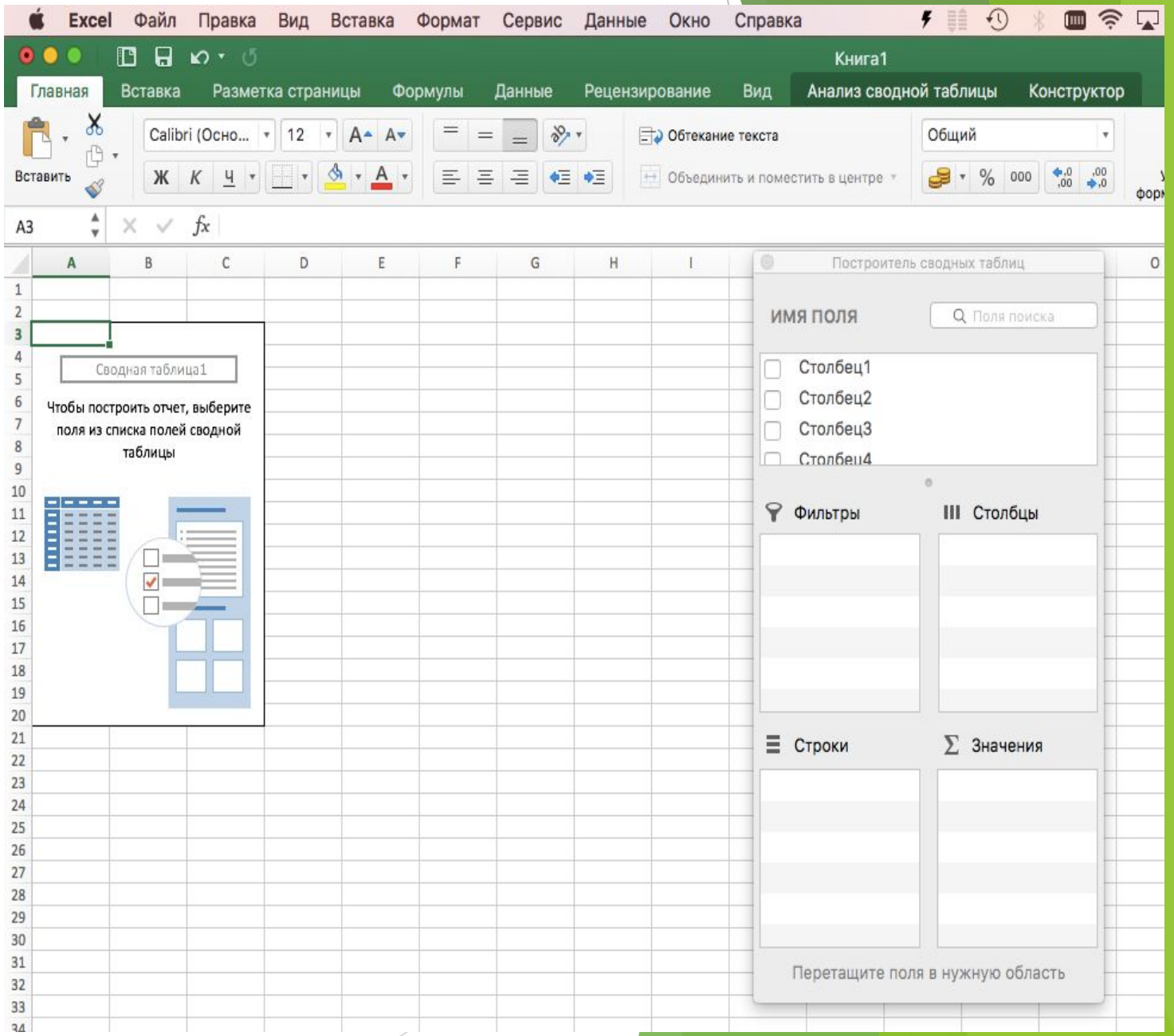
Сводная таблица — это инструмент обработки данных, упрощающий их обобщение и визуализацию.
Если имеются исходные данные в форме таблицы, столбцы которой содержат однотипные или малоинформативные значения, то
сводная таблица — это таблица, связанная с исходной таблицей и обобщающая исходные данные путем различных вычислительных процедур, например, вычисления сумм, средних, долей и т. п.

В **EXEL** 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



2. Сводные таблицы и сводные диаграммы в Microsoft Excel

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

Например:

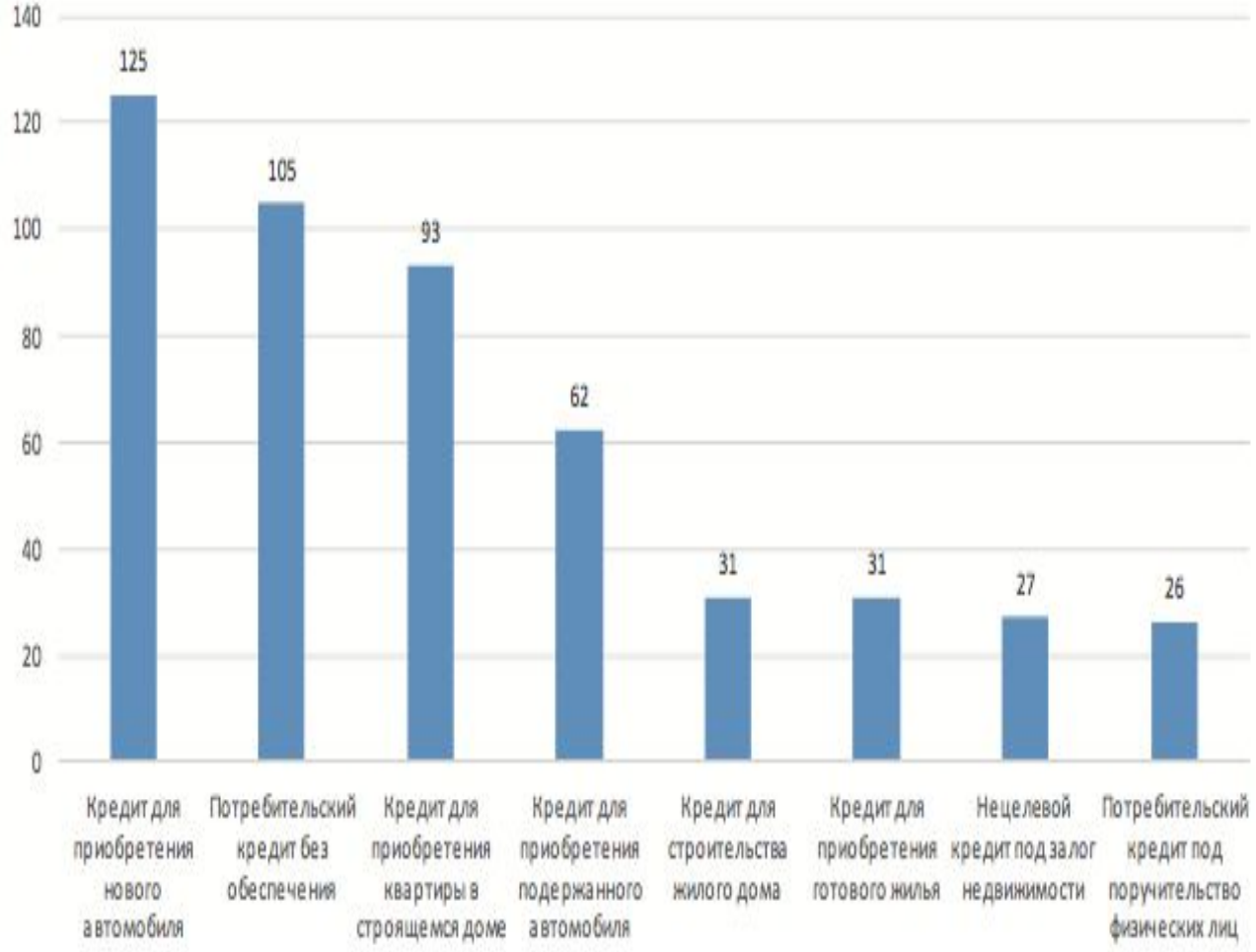
	A	B
1	**Номер договора**	**Вид кредита**
2	1	Кредит для приобретения подержанного автомобиля
3	2	Потребительский кредит без обеспечения
4	3	Кредит для строительства жилого дома
5	4	Кредит для приобретения подержанного автомобиля
6	5	Потребительский кредит без обеспечения
7	6	Кредит для приобретения нового автомобиля
8	7	Кредит для приобретения подержанного автомобиля
9	8	Кредит для приобретения нового автомобиля
10	9	Кредит для строительства жилого дома
...	...	...

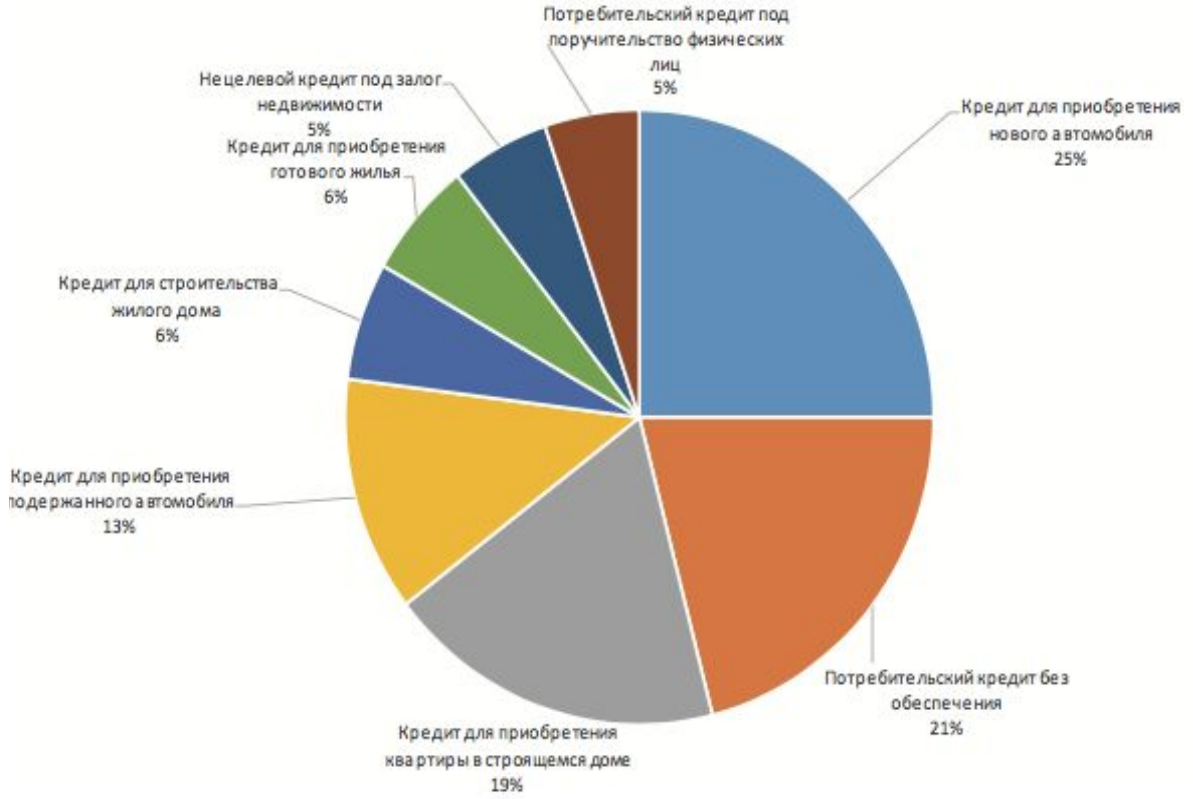
C	D
Названия строк	**Количество по полю Вид кредита**
Кредит для приобретения нового автомобиля	125
Потребительский кредит без обеспечения	105
Кредит для приобретения квартиры в строящемся доме	93
Кредит для приобретения подержанного автомобиля	62
Кредит для строительства жилого дома	31
Кредит для приобретения готового жилья	31
Нецелевой кредит под залог недвижимости	27
Потребительский кредит под поручительство физических лиц	26
Общий итог	**500**

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	Банк	Вид кредита	Статус
2	А-банк	Автокредит	Погашен
3	А-банк	Автокредит	Погашен
4	А-банк	Автокредит	Погашен
5	А-банк	Автокредит	Не погашен
6	А-банк	Автокредит	Погашен
7	А-банк	Автокредит	Погашен
8	А-банк	Автокредит	Погашен
9	А-банк	Автокредит	Погашен
10	А-банк	Автокредит	Погашен
11	А-банк	Автокредит	Погашен
12	А-банк	Автокредит	Погашен
13	А-банк	Автокредит	Погашен
14	А-банк	Автокредит	Погашен
15	А-банк	Автокредит	Погашен
16	А-банк	Автокредит	Погашен
17	А-банк	Автокредит	Погашен
18	А-банк	Автокредит	Погашен
19	А-банк	Автокредит	Погашен
...	...	...	...
402	Б-банк	Потребительский кредит	Погашен
403	Б-банк	Потребительский кредит	Не погашен
404	Б-банк	Потребительский кредит	Погашен

$a)$

	A	B	C	D	E	F	G
1	Банк	Вид кредита	Статус	Количество по полю Статус	Названия столбцов		
2	А-банк	Автокредит	Погашен	Названия строк	Не погашен	Погашен	Общий итог
3	А-банк	Автокредит	Погашен	**А-банк**	**35**	**165**	**200**
4	А-банк	Автокредит	Погашен	Автокредит	3	37	40
5	А-банк	Автокредит	Не погашен	Потребительский кредит	32	128	160
6	А-банк	Автокредит	Погашен	**Б-банк**	**34**	**169**	**203**
7	А-банк	Автокредит	Погашен	Автокредит	27	154	181
8	А-банк	Автокредит	Погашен	Потребительский кредит	7	15	22
9	А-банк	Автокредит	Погашен	**Общий итог**	**69**	**334**	**403**
10	А-банк	Автокредит	Погашен				
11	А-банк	Автокредит	Погашен	Количество по полю Статус	Названия столбцов		
12	А-банк	Автокредит	Погашен	Названия строк	Не погашен	Погашен	Общий итог
13	А-банк	Автокредит	Погашен	**А-банк**	**17,50%**	**82,50%**	**100,00%**
14	А-банк	Автокредит	Погашен	Автокредит	7,50%	92,50%	100,00%
15	А-банк	Автокредит	Погашен	Потребительский кредит	20,00%	80,00%	100,00%
16	А-банк	Автокредит	Погашен	**Б-банк**	**16,75%**	**83,25%**	**100,00%**
17	А-банк	Автокредит	Погашен	Автокредит	14,92%	85,08%	100,00%
18	А-банк	Автокредит	Погашен	Потребительский кредит	31,82%	68,18%	100,00%
19	А-банк	Автокредит	Погашен	**Общий итог**	**17,12%**	**82,88%**	**100,00%**
...	...	...	...				
402	Б-банк	Потребительский кредит	Погашен				
403	Б-банк	Потребительский кредит	Не погашен				
404	Б-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 разделается на какие либо категории, каждая категория, в свою очередь, делится на подкатегории и т. д.

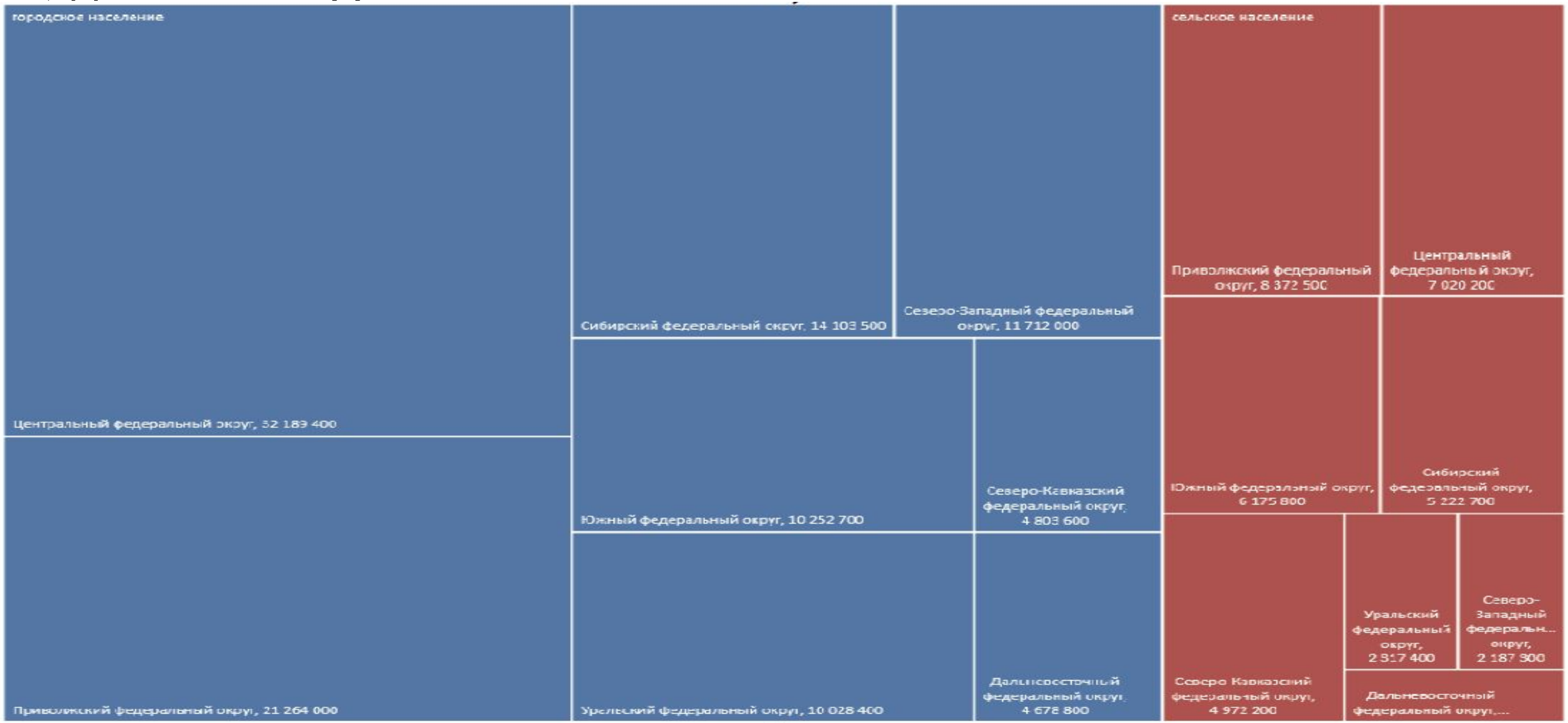
Распознать такую ***иерархию признаков*** можно с помощью визуализации данных.

А) ГИСТОГРАММА С НАКОПЛЕНИЕМ

4. Иерархия признаков в Microsoft Excel.

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

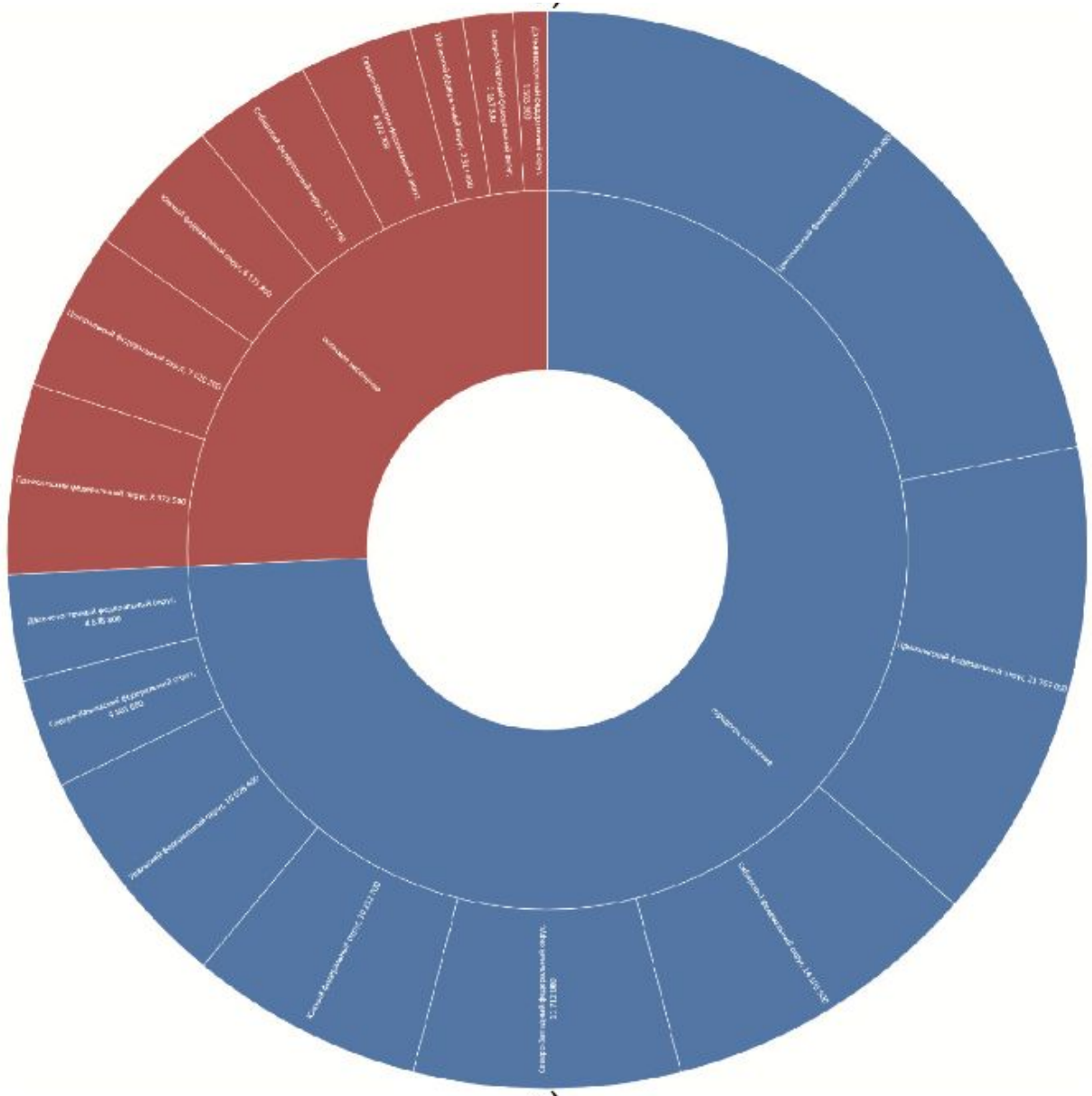
Б) ДИАГРАММА «**ДЕРЕВО**»



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

Задание: решение задач по плану преподавателя.