

«Информационные технологии в экономике»

установочные лекции

25.02.2021

старший преподаватель Суязова Светлана Андреевна

к.э.н., доцент Крамаренко Инна Владимировна

Данные — это совокупность сведений, зафиксированных на определенном носителе в форме, пригодной для постоянного хранения, передачи и обработки.

Информация — это результат преобразования и анализа данных. Информация появляется в результате обработки данных при решении конкретных задач.

ФЗ РФ от 27 июля 2006 г. N 149-ФЗ «Об информации, информационных технологиях и о защите информации» (последнее изменение 18.12.2018 г.)

Знания — это зафиксированная и проверенная практикой обработанная информация, которая использовалась и может многократно использоваться для принятия решений.

ДАННЫЕ, ИНФОРМАЦИЯ, ЗНАНИЯ – ПИРАМИДА DIKW

#datasciencewiz

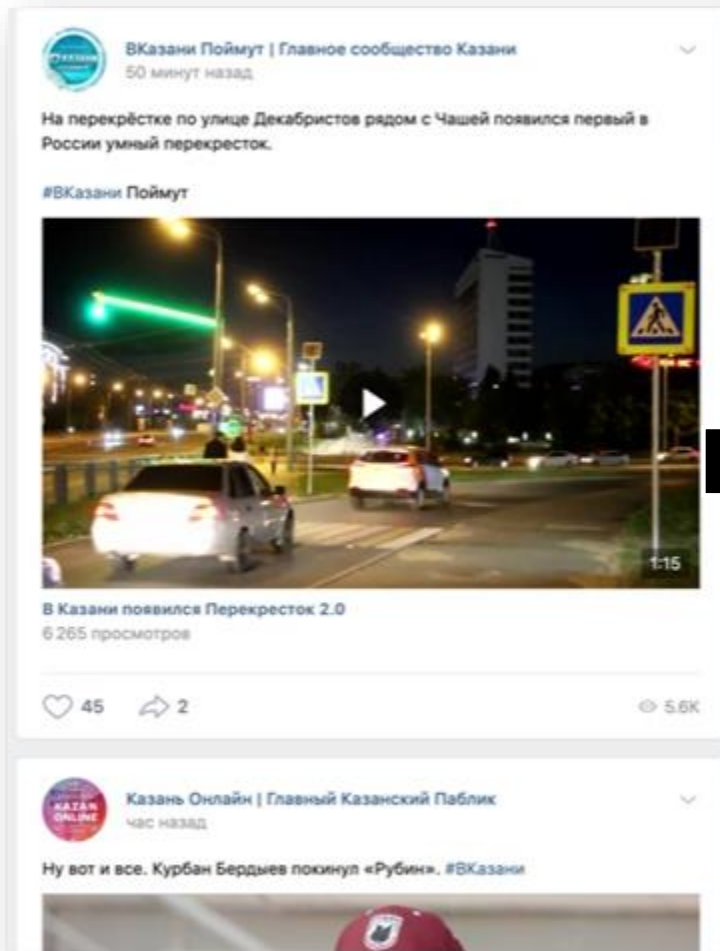


Пример цикла обработки данных

Неструктурированный
поток информации

Сбор данных в
заданную структуру
(таблицы)

Визуализация и анализ
данных



Дата сбора данных (пн или пт) в формате ДД/ММ/ГГГГ	Группа (наименование паблика)	Ссылка на пост
22/02/2021	Nachos History	https://vk.com/wall-156929481_133380
22/02/2021	Nachos History	https://vk.com/wall-156929481_133392
22/02/2021	Nachos History	https://vk.com/wall-156929481_133396
22/02/2021	Nachos History	https://vk.com/wall-156929481_133400
22/02/2021	Nachos History	https://vk.com/wall-156929481_133402
22/02/2021	Nachos History	https://vk.com/wall-156929481_133411
22/02/2021	Nachos History	https://vk.com/wall-156929481_133419
22/02/2021	Nachos History	https://vk.com/wall-156929481_133434
22/02/2021	Альбом	https://vk.com/wall-107232016_2602
22/02/2021	Альбом	https://vk.com/wall-107232016_2602
22/02/2021	Альбом	https://vk.com/wall-107232016_2602
22/02/2021	Альбом	https://vk.com/wall-107232016_26028
22/02/2021	Альбом	https://vk.com/wall-107232016_26030



Теория по дисциплине

Работа с данными

- И.В.Крамаренко Рабочая версия учебного пособия по дисциплине «Информационные технологии в экономике»
- Н.Б.Паклин, В.И.Орешков Бизнес-аналитика: от данных к знаниям
- М.А. Абросимова Информационные технологии в государственном и муниципальном управлении

Работа в Excel

- И.В.Крамаренко, С.А.Суязова Рабочие версии методических указаний к выполнению контрольных заданий
- М.Александр, Р.Куслейка Excel 2019. Библия пользователя

Практическое содержание дисциплины

Работа с данными

- Обработка данных в различных шкалах
- Визуализация данных
- Описательные статистики

Работа в Excel

- Диаграммы
- Надстройка «Анализ данных»
- Сводные таблицы

Контрольные задания

- Контрольное задание №1: простая и стандартизованная ранжировки.

Данные: стандартные варианты. Максимальный балл: 10

- Контрольное задание №2: визуализация и анализ данных на примере статистики по соцсети «ВКонтакте».

Данные: необходимо собрать самостоятельно. Максимальный балл: 25

- Контрольное задание №3: Описательные статистики.

Данные: стандартные варианты. Максимальный балл: 10

Контрольное задание №1

Цель задания

На примере данных по оценке качества работы министров РФ (статистика ВЦИОМ):

1. Оценить, как изменился рейтинг между двумя периодами
2. Построить совершенную ранжировку
3. Применить условное форматирование, чтобы отследить изменение рейтинга

Работа в Excel

- Извлечение данных функцией СМЕЩ()
- Условное форматирование, автофильтр
- Функции РАНГ.РВ(), РАНГ.СР() и ручной расчёт совершенных рангов

Пример к заданию №1

Работа с данными в порядковой шкале

- Рейтинги, балльные оценки
- Разрешённые сравнения позиций: равно-не равно, больше-меньше
- Приводятся к одному знаменателю с помощью *совершенной ранжировки n объектов*:
 1. Лучшая оценка имеет ранг 1, худшая – ранг n .
 2. Ранги располагаются по убыванию качества оценок.
 3. Группа из одинаковых оценок получает средний ранг группы.
 4. Проверка: $\sum_{i=1}^n r_i = n(n + 1)/2$

Пример к заданию №1

Средние баллы оценки работы пяти министров (выдержка из данных к заданию №1)

	М1	М2	М3	М4	М5
Баллы, III кв. 2012	3,15	3,90	3,36	3,13	3,13
Баллы, I кв. 2013	3,03	3,94	3,40	3,06	3,06

Проверка согласованности с помощью рангового коэффициента Спирмена ρ_S , где d_i^2 – квадраты разностей совершенных рангов.

Интерпретация – [шкала Чеддока](#)

$$\rho_S = 1 - \left(\frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \right)$$

Контрольное задание №3

Формируем общее представление о выборке

- Расчёт описательных статистик:
 1. Меры центра распределения: среднее, медиана, мода
 2. Меры разброса: дисперсия, стандартное отклонение, размах, коэффициент вариации
 3. Показатели формы распределения: коэффициенты асимметрии и эксцесса
- Построение гистограммы
- Как интерпретировать гистограмму, коробчатую диаграмму

Пример к заданию №3

Цели примера:

- Посмотреть, какие описательные статистики подходят для показателей в различных шкалах измерения
- Построить гистограмму и коробчатую диаграмму

Работа в Excel:

- Процедура «Анализ данных»
- Расчёт описательных статистик с помощью формул
- Гистограмма и коробчатая диаграмма

Непрерывная количественная шкала (интервалов, отношений)

Примеры данных в непрерывной шкале: стоимостные показатели

Допустимые операции: все арифметические

Что измеряем	Как измеряем
Характеристики центра распределения	Среднее (арифметическое), медиана, мода
Разброс относительно центра распределения	Дисперсия, стандартное отклонение, размах вариации
Форма распределения	Коэффициент асимметрии, Коэффициент эксцесса

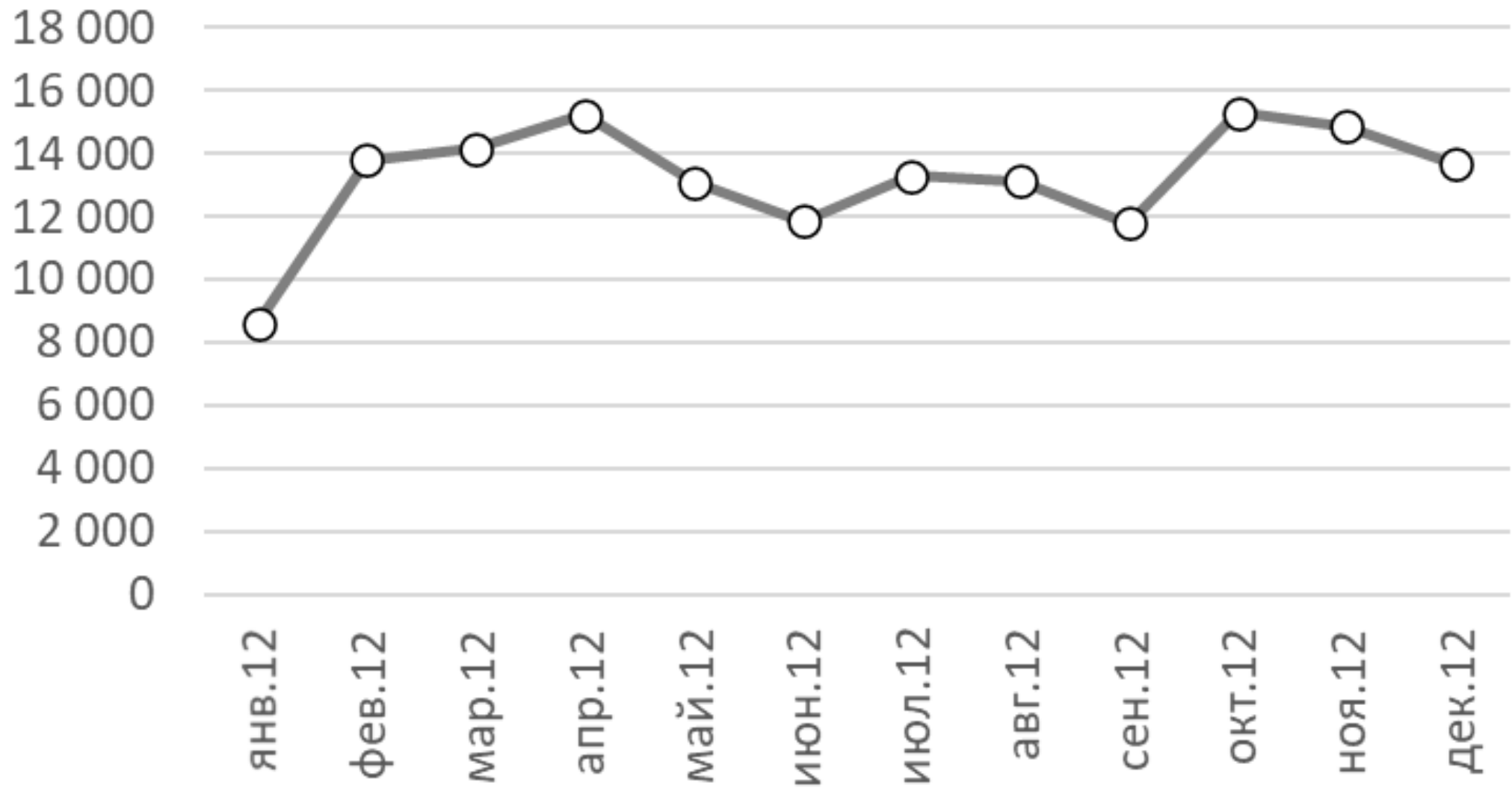
Исследуем непрерывный показатель

1. Рассчитать описательные статистики для показателя **X** – величины вновь созданных организаций в 2012 году.
2. Постройте график динамики показателя **X**.
3. Постройте гистограмму показателя **X**.
4. Постройте усиковую диаграмму показателя **X** («ящик с усами»).

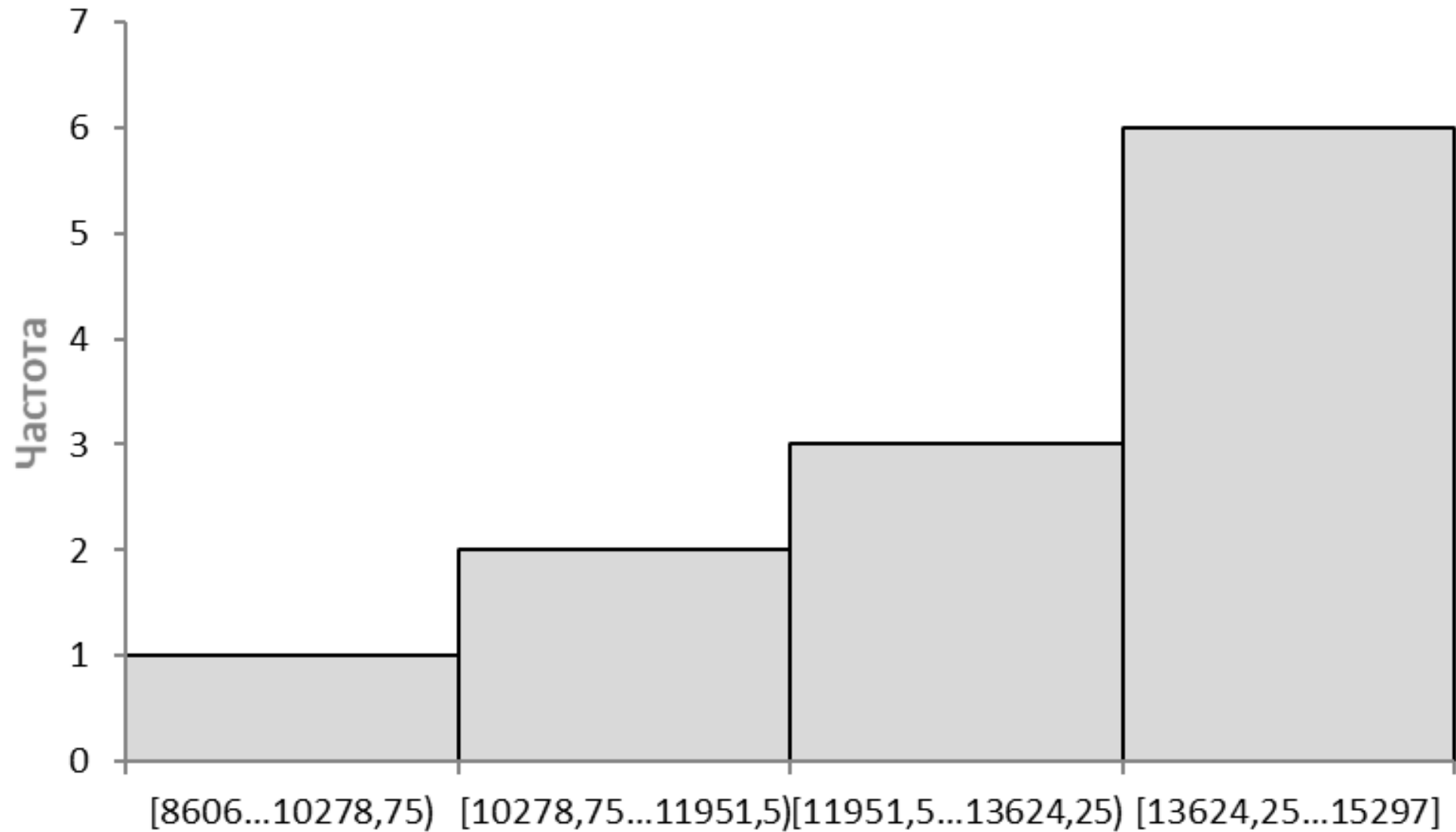
месяц	X
январ.12	8 606
февр.12	13 802
мар.12	14 160
апр.12	15 235
май.12	13 097
июн.12	11 854
июл.12	13 293
авг.12	13 125
сен.12	11 800
окт.12	15 297
ноя.12	14 860
дек.12	13 662

месяц	X	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$((x_i - \bar{x})/s)^3$	$((x_i - \bar{x})/s)^4$
янв.12	8 606	-4 627	21 405 273,34	-15,62	39,05
фев.12	13 802	569	324 235,34	0,03	0,01
мар.12	14 160	927	860 101,67	0,13	0,06
апр.12	15 235	2 002	4 009 672,51	1,27	1,37
май.12	13 097	-136	18 382,84	0,00	0,00
июн.12	11 854	-1 379	1 900 492,01	-0,41	0,31
июл.12	13 293	60	3 650,17	0,00	0,00
авг.12	13 125	-108	11 574,17	0,00	0,00
сен.12	11 800	-1 433	2 052 295,01	-0,46	0,36
окт.12	15 297	2 064	4 261 816,17	1,39	1,55
ноя.12	14 860	1 627	2 648 485,01	0,68	0,60
дек.12	13 662	429	184 398,67	0,01	0,00
сумма			37 680 376,92	-13,00	43,31

График динамики X



Гистограмма X



Порядковая шкала

Данные в порядковой шкале: рейтинги (необязательно числовые)

Допустимая операция: упорядочивание по возрастанию и убыванию

Что измеряем	Как измеряем
Характеристики центра распределения	Медиана, мода
Разброс относительно центра распределения	размах вариации, межквартильный интервал
Форма распределения	Максимум, минимум, квартили

Исследуем порядковый показатель

Рассчитать описательные статистики для показателя **X** – бальной оценки разнообразия контента паблика в социальной сети.

№ подписчика	X
1	8
2	7
3	6
4	8
5	7
6	10
7	9
8	10
9	5
10	7

Номинальная шкала

Примеры данных в номинальной шкале: качественные показатели

Допустимые операции: сравнение равно/не равно

Что измеряем	Как измеряем
Характеристики центра распределения	Мода
Разброс относительно центра распределения	нет
Форма распределения	нет

Исследуем показатель в номинальной шкале

Рассчитать описательные статистики для показателя **X** – теги, приписанные постам паблика «Вконтакте» за январь.

№ п/п	Тег	№ п/п	Тег
1	имнорм	9	реклама
2	реклама	10	имнорм
3	неестественное	11	имнорм
4	имнорм	12	фанатское
5	фанатское	13	неестественное
6	фанатское	14	реклама
7	неестественное	15	имнорм
8	неестественное	16	реклама

Контрольное задание №2

Анализ статистики по группам соцсети на примере «ВКонтакте»

Сбор данных: 3 паблика, 2 раза в неделю (понедельник, пятница)

- *Таблица 1: посты в пабликах.* Единица наблюдения – пост в паблике. Каждый пост фиксируется в данных только один раз. Фиксируем значения показателей по посту: количество лайков, репостов, просмотров, относится ли к рекламе, приложено ли видео, аудио, опрос, и т.д.
- *Таблица 2: численность подписчиков в пабликах.* Единица наблюдения – паблик. На каждую дату наблюдения фиксируем численность подписчиков.

Контрольное задание №2

Анализ статистики по группам соцсети на примере «ВКонтакте»

Анализ данных

- *Консолидация и срезы.* Строим сводные таблицы, чтобы рассчитать характеристики, например: максимальное количество лайков/репостов/просмотров на дату в группе; максимальное количество комментариев
- *Визуализация.* Строим графики: чтобы показать динамику показателей (численность подписчиков); чтобы показать доли постов (с рекламой и без неё); чтобы сравнить активность пользователей (будни и выходные)

Контрольное задание №2

Анализ статистики по группам соцсети на примере «ВКонтакте»

Представление результатов

- Максимум 10 слайдов
- Презентация со структурой: (1) кратко о пабликах – (2) гипотезы, или на какие вопросы хотим ответить с помощью данных – (3...9) слайды с визуализацией статистики – (10) выводы
- Формат слайда с визуализацией: заголовок-гипотеза + график или таблица на основе данных, подкрепляющие гипотезу

Примеры к заданию №2

Цели примера 2.1

- Очистить данные на примере небольшой выборки
- Сформировать сводную таблицу на основе выборки

Цель примера 2.2


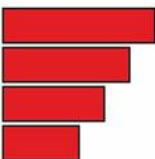
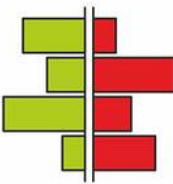
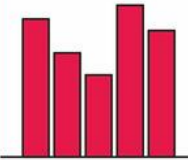
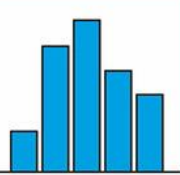
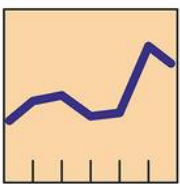
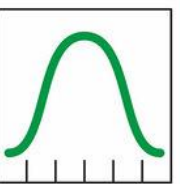

- Построить диаграммы различных типов, подходящие к данным

Работа в Excel

- Работа с датами и форматами чисел
- Сводные таблицы
- Построение диаграмм

ОСНОВНЫЕ ТИПЫ ДИАГРАММ

ТИПЫ СРАВНЕНИЯ

	ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
КРУГОВАЯ					
ЛИНЕЙЧАТАЯ					
ГИСТОГРАММА					
ГРАФИК					
ТОЧЕЧНАЯ					

Источник: Джин Желязны,
«Говори на языке
диаграмм»

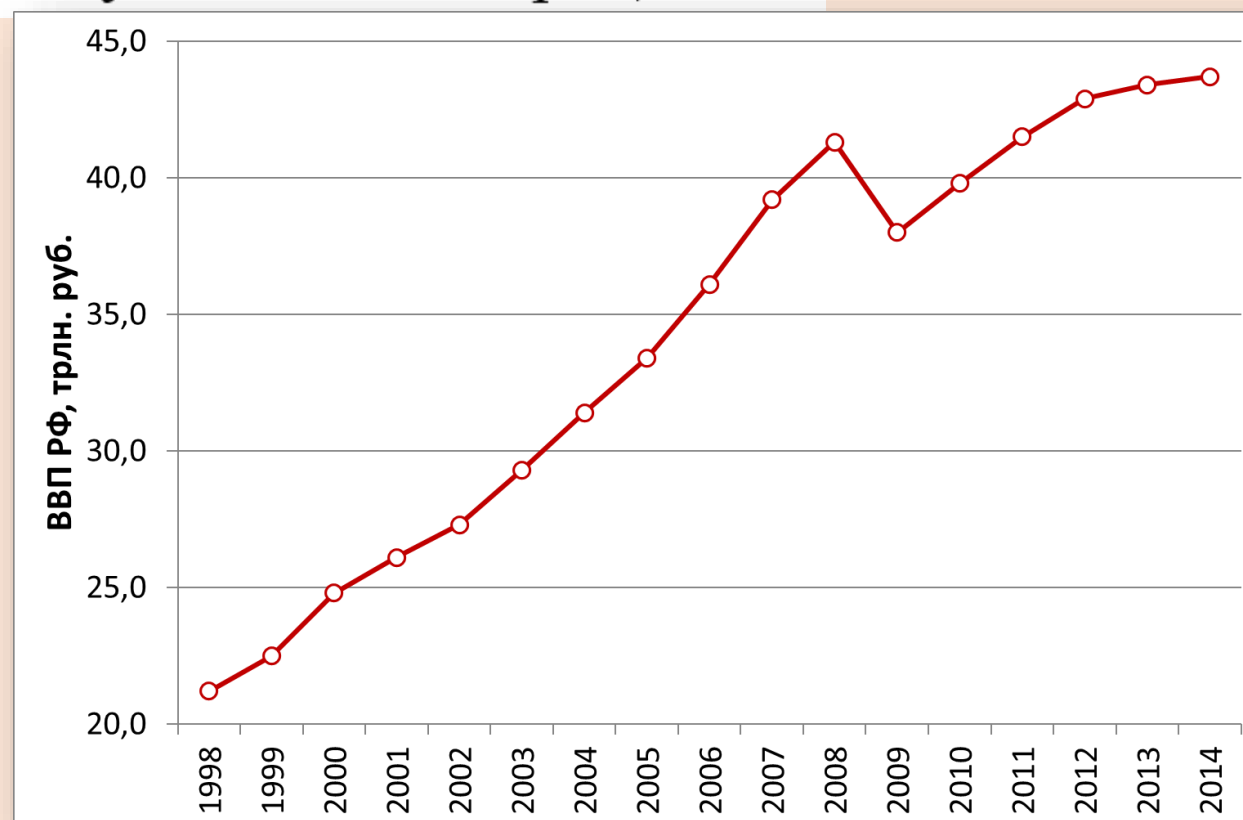
Ключевые правила построения диаграммы

- Понятное название
- Крупный шрифт
- Меньше нулей
- Подписи осей (включая единицы измерения)
- Легенда нужна только если рядов больше одного
- Время по горизонтали
- Не пытайтесь визуализировать сразу всё на одной диаграмме

Варианты представления данных

Временной ряд (time series data) - собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного) исследуемого процесса.

- *Пример:* ВВП РФ в рыночных ценах с 1998 по 2014
- *Тип сравнения:* временное
- *Вид диаграммы:* график, столбчатая диаграмма



Варианты представления данных

Кросс-секционные, перекрёстные данные (Cross-Sectional data)

Множество объектов (респонденты, домашние хозяйства, регионы, фирмы и т.п.) в один момент времени, несколько показателей (признаков).

- *Пример:*
среднемесячная номинальная з/п работников по регионам УФО в 2019
- *Тип сравнения:*
позиционное,
частотное
- *Вид диаграммы:*
линейчатая
диаграмма,
гистограмма

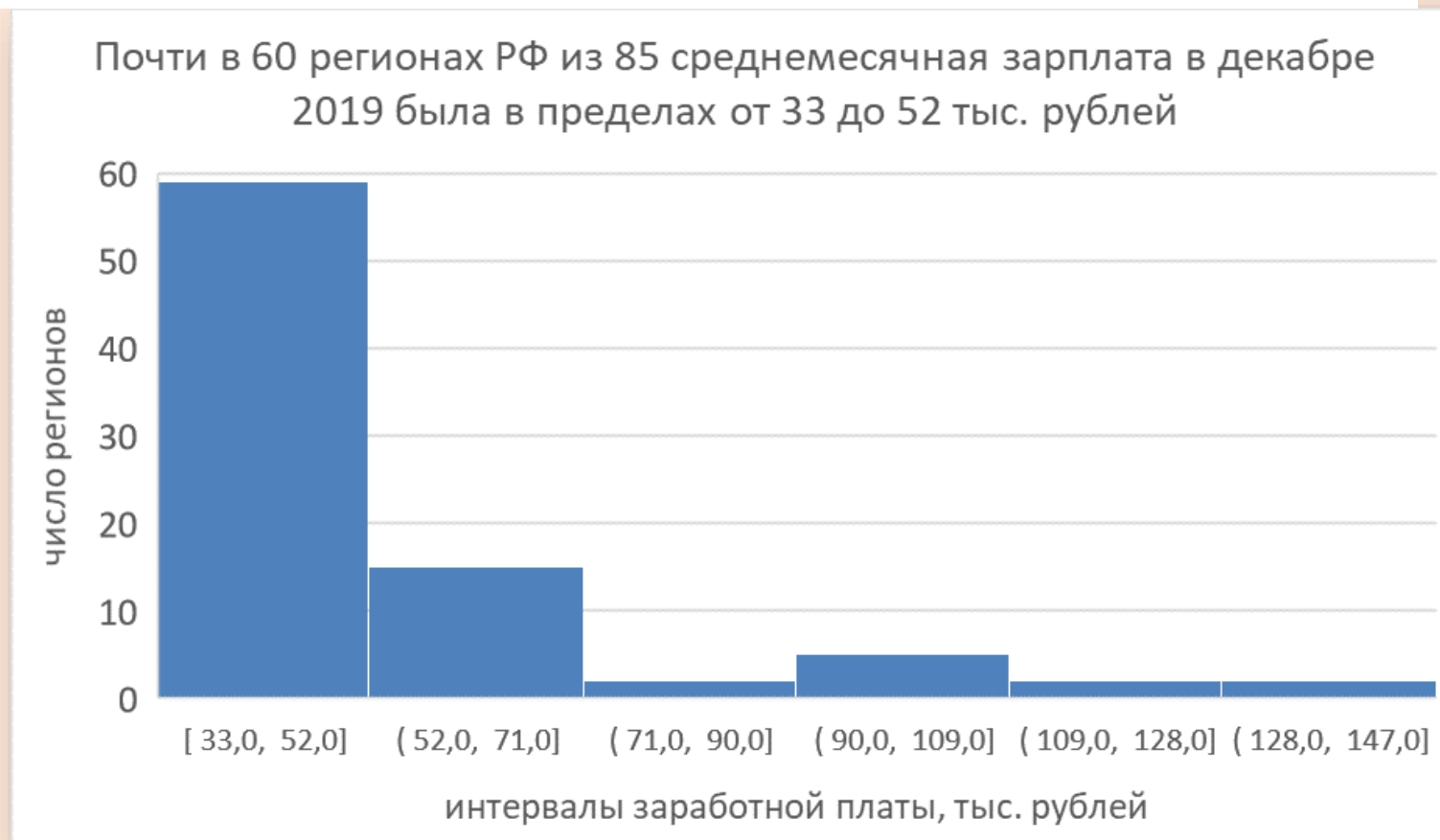


Варианты представления данных

Кросс-секционные, перекрёстные данные (Cross-Sectional data)

Множество объектов (респонденты, домашние хозяйства, регионы, фирмы и т.п.) в один момент времени, несколько показателей (признаков).

- *Пример:*
среднемесячная
номинальная з/п
работников по
регионам УФО в 2019
- *Тип сравнения:*
позиционное,
частотное
- *Вид диаграммы:*
линейчатая
диаграмма,
гистограмма

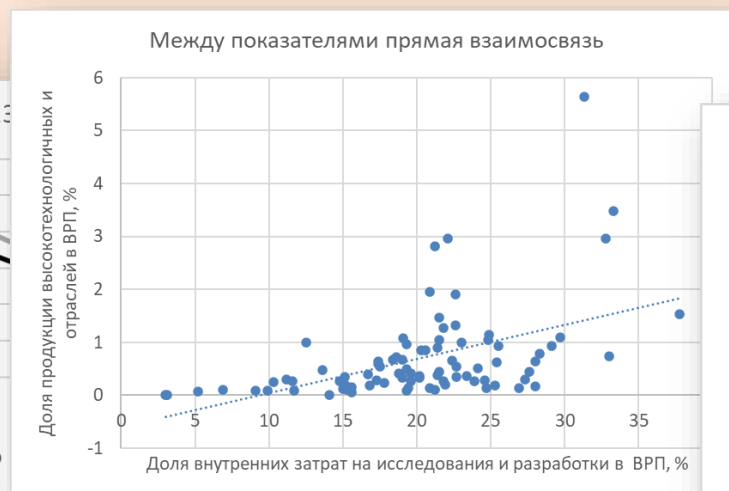
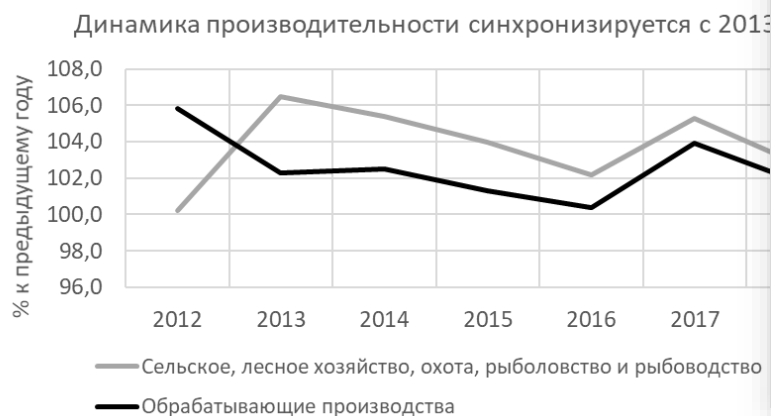


Варианты представления данных

Панельные данные (Panel data)

Множество объектов (респонденты, домашние хозяйства, регионы, фирмы и т.п.) наблюдаются в последовательные моменты времени. Каждый объект характеризуется несколькими показателями (признаками).

- *Пример: ВРП и производительность труда по регионам РФ за 2012 – 2018*



Варианты представления данных

- В одной таблице может быть несколько вариантов представления данных
- Главное не пытаться нарисовать всё это на одной диаграмме

panel data (green arrow)

time-series data (red arrow)

cross-section data (blue arrow)

	1998г.	1999г.	2000г.	2001г.	2002г.	2003г.
Валовой региональный продукт по субъектам Российской Федерации (валовая добавленная стоимость в основных ценах)	2 251 977,5	3 827 375,5	5 753 671,6	7 170 968,2	8 744 219,2	10 742 423,3
Центральный федеральный округ	634 372,0	1 190 894,4	1 841 498,9	2 243 525,0	2 878 664,5	3 577 142,5
Белгородская область	18 245,5	32 060,6	42 074,5	49 941,8	62 404,4	76 054,5
Брянская область	11 051,3	16 809,4	24 650,5	30 110,3	37 374,1	43 700,3
Владимирская область	14 936,9	24 481,0	33 017,7	42 075,4	50 359,9	61 818,6
Воронежская область	22 381,9	36 278,8	49 523,9	60 014,6	83 001,1	100 143,3
Ивановская область	8 278,4	11 743,7	16 900,0	22 175,9	26 981,3	33 214,6
Калужская область	10 097,3	16 009,8	23 903,3	31 860,0	37 283,1	48 792,7
Костромская область	8 478,9	13 363,6	16 662,2	22 221,5	25 952,9	29 692,1
Курская область						
Липецкая область						
Московская область						

Численность экономически активного населения по субъектам Российской Федерации (по данным выборочных обследований населения по месту жительства)

	2000	2001	2002	2003
Российская Федерация	72769,9	71546,6	72357,1	72391,4
Центральный федеральный округ	19530,8	19375,2	19597,2	19641,2
Белгородская область	740,6	748,3	720,2	719,5
Брянская область	664,6	651,5	636,1	650,7
Владимирская область	797,9	785,4	808,0	795,7
Воронежская область	1171,3	1106,1	1140,1	1113,1
Ивановская область	591,9	572,1	567,5	565,5
Калужская область	561,4	542,5	520,7	539,9
Костромская область	380,2	373,7	370,5	362,2
Курская область	636,1	608,7	615,7	583,4
Липецкая область	601,1	598,4	607,2	603,0
Московская область	3612,4	3504,5	3505,3	3617,8

Диаграммы для долевого показателей

- Показатели должны суммироваться (абсолютные, не усреднённые значения)
- Категории в сумме должны давать 100%
- Секторов (долей) на графике не больше 7-8
- *Пример:* Оборот организаций по РФ в 2018, млн. руб., по 18 видам экон. деятельности. Сортируем >> оставляем 7 видов с наибольшим оборотом, прочее объединяем в сектор «остальные»



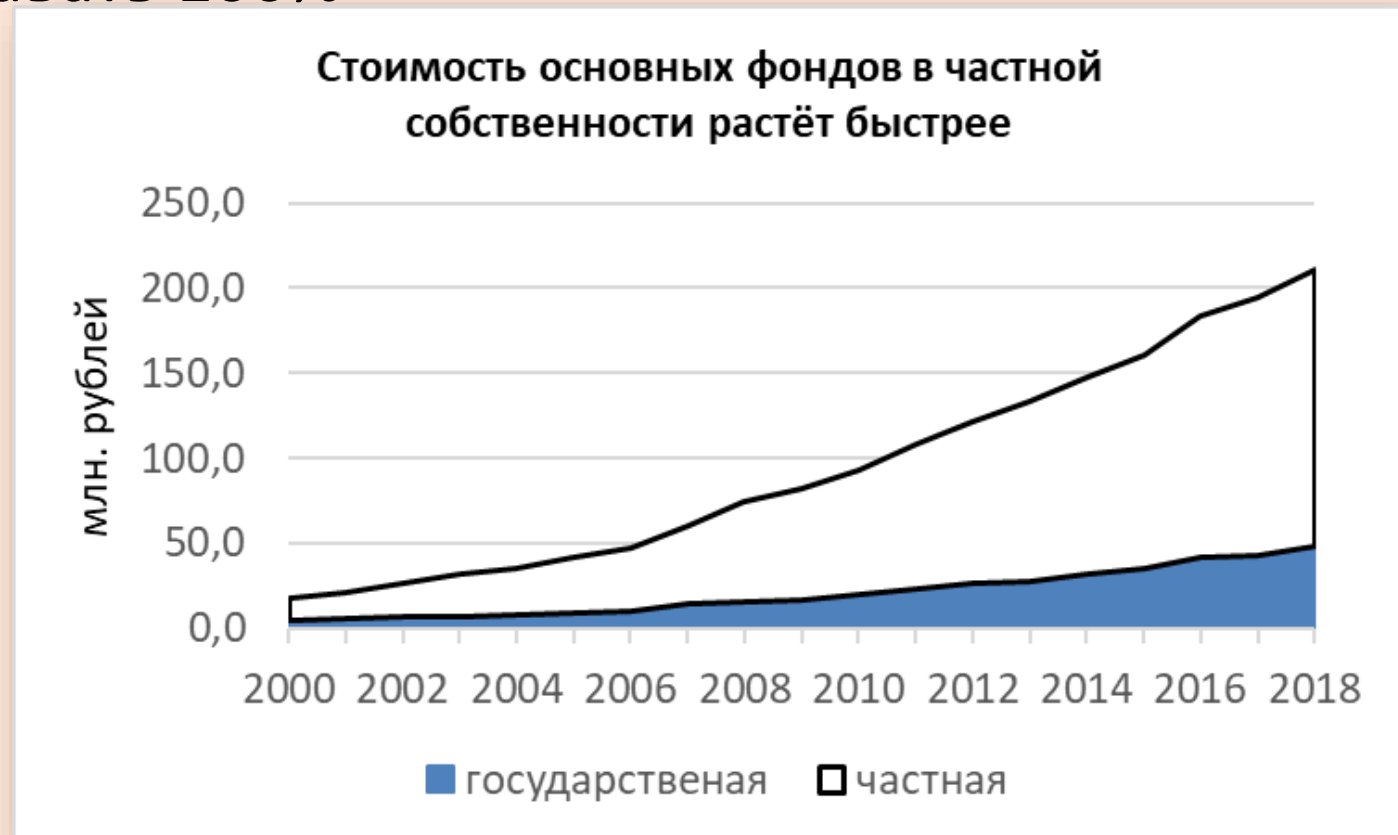
Диаграммы для долевого показателей

- Показатели должны суммироваться (абсолютные, не усреднённые значения)
- Категории в сумме должны давать 100%
- Варианты со временем: долевая или с накоплением
- *Пример:* Доля сельского и городского населения РФ с 2001 по 2019. Сравнение временное и компонентное, показатель изменяется медленно, поэтому промежуточные годы убираем



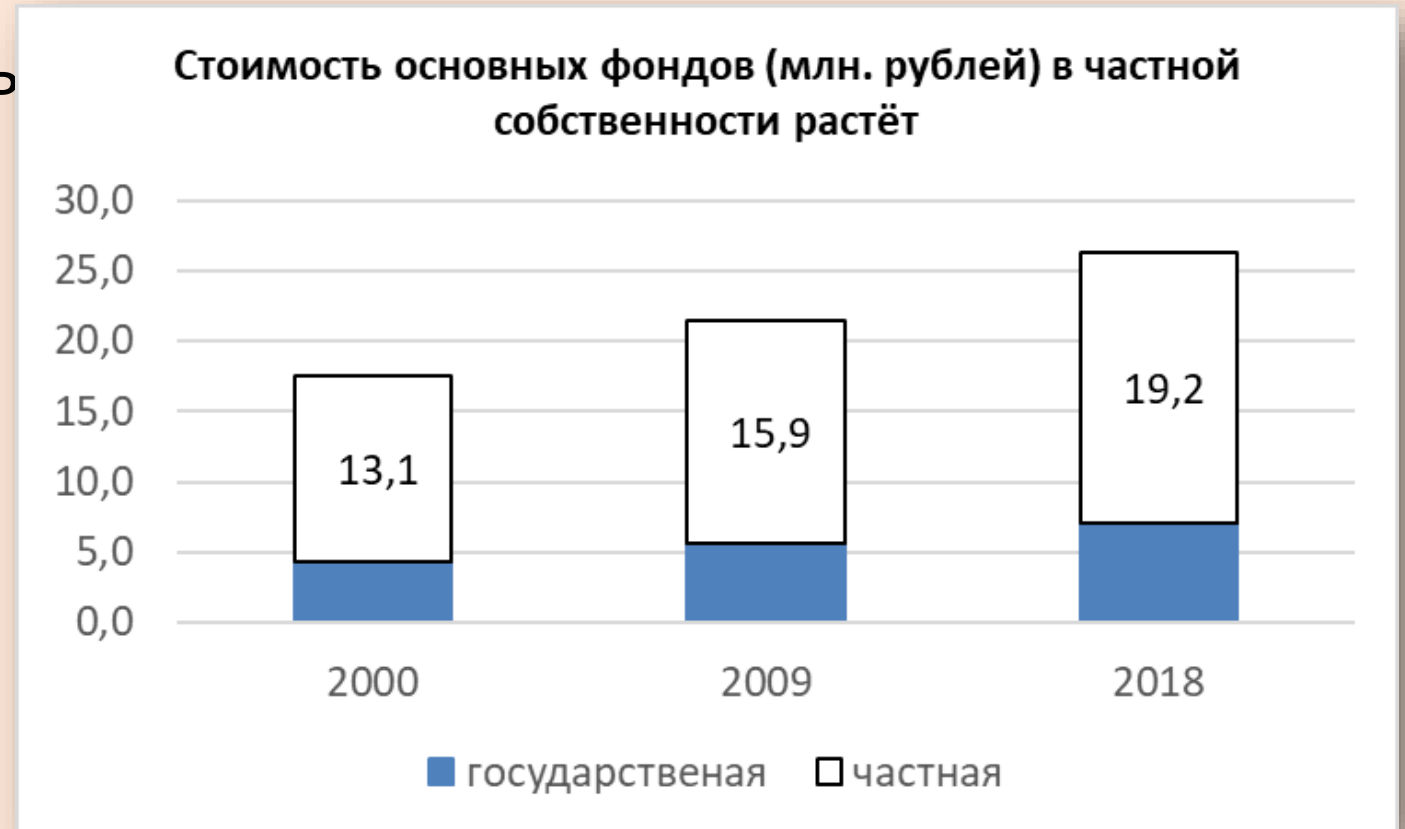
Диаграммы для долевого показателей

- Показатели должны суммироваться (абсолютные, не усреднённые значения)
- Категории в сумме должны давать 100%
- *Пример:* Стоимость основных фондов в РФ по формам собственности, 2000 - 2018. Сравнение временное и компонентное

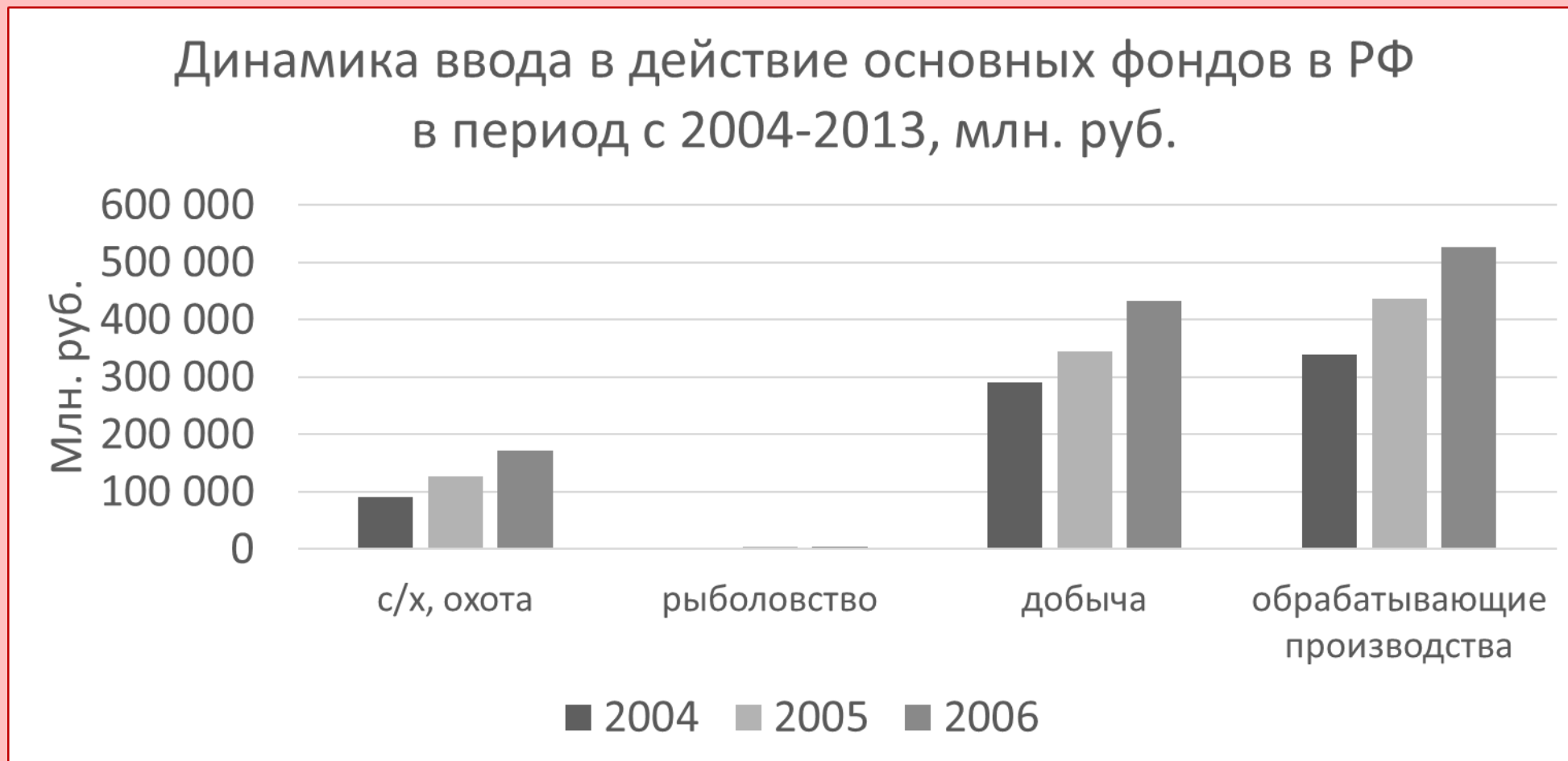


Диаграммы для долевого показателей

- Показатели должны суммироваться (абсолютные, не усреднённые значения)
- Категории в сумме должны давать 100%
- *Пример:* Стоимость основных фондов в РФ по формам собственности, 2000 - 2018. Сравнение временное и компонентное



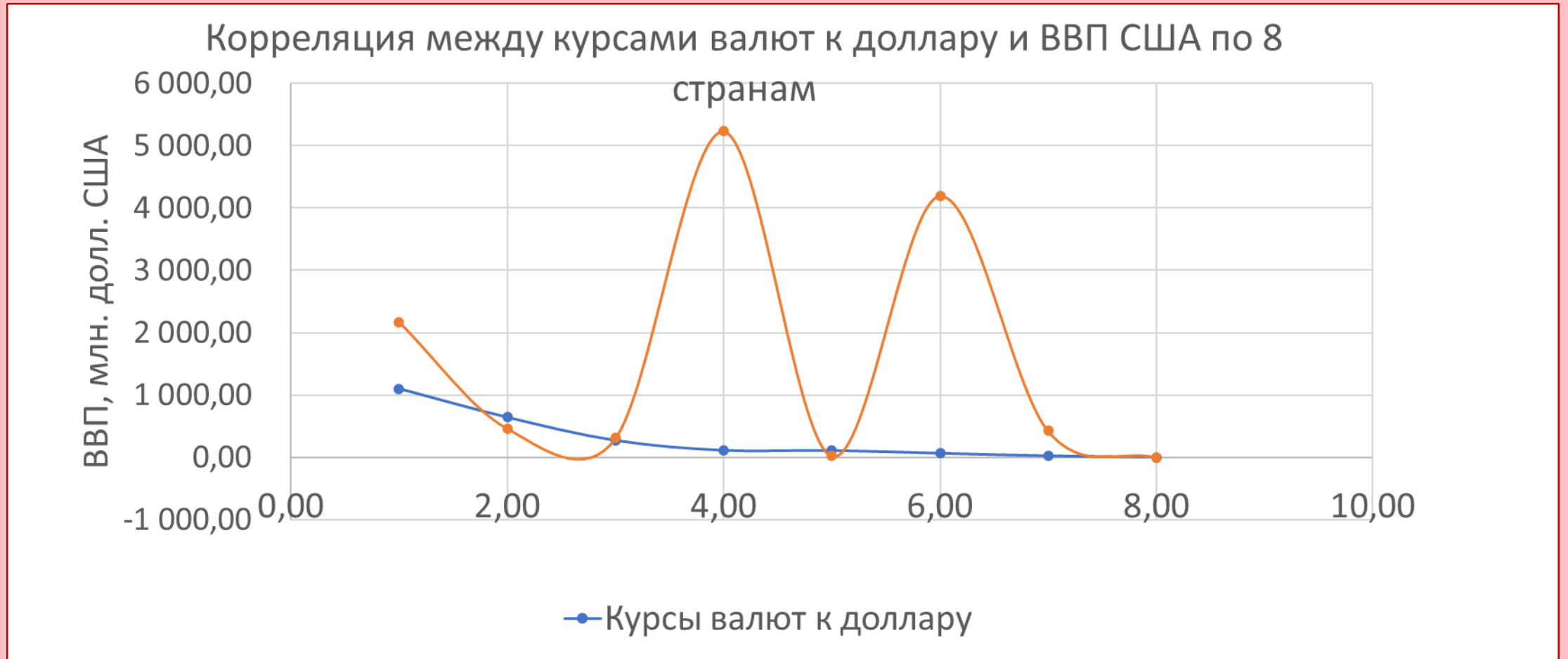
Антипримеры диаграмм



Антипримеры диаграмм



Антипримеры диаграмм



Спасибо за внимание

@Суязова Светлана Андреевна

Материалы выложены в файлах канала «Заочное отделение»
в Teams; код присоединения к команде:

pdk9dco