**Способы графического представления данных.**

**Гистограмма** — (histogram) График распределения частот,

который строится при помощи прямоугольников, чья площадь

пропорциональна частоте нахождения данной величины в

интервале, на котором построен данный прямоугольник.

****

**Boxplot**

В R (и не только) для построения диаграмм размахов служит функция boxplot().

Таким образом, в отличие от других статистических программ, в R при построении

диаграмм размахов используются устойчивые (робастные) оценки центральной

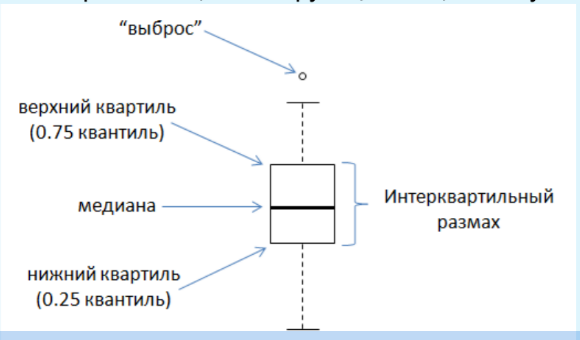
тенденции (медиана) и разброса (интерквартильный размах, ИКР). Верхний "ус"

простирается от верхней границы "ящика" до наибольшего выборочного значения,

находящегося в пределах расстояния 1.5 х ИКР от этой границы. Аналогично,

нижний "ус" простирается от нижней границы "ящика" до наименьшего выборочного

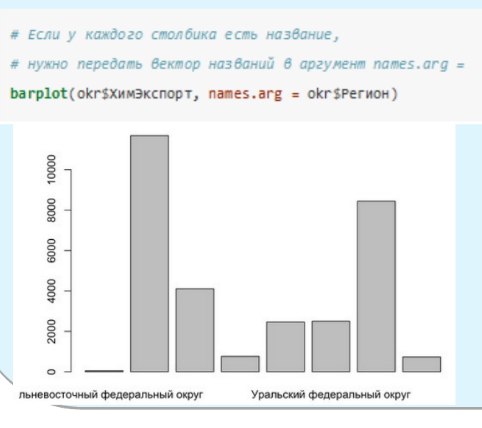
значения, находящегося в пределах расстояния 1.5 х ИКР от этой границы.

****

**Столбчатые графики** — barplot — отображают вектор числовых данных в виде

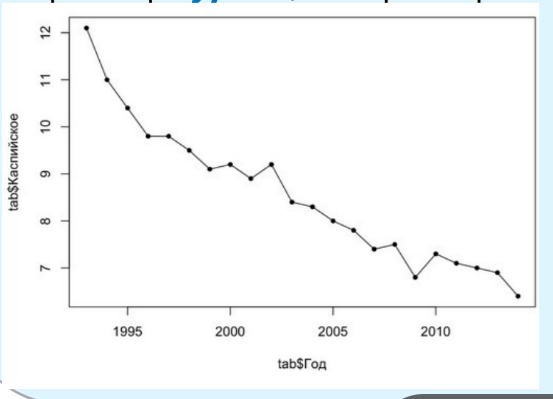
столбиков. Это простейший вид графика (наряду с dotchart), который используется

для сравнения абсолютных величин.



**Линейные графики**

По умолчанию функция plot() рисует именно точки



**Круговые диаграммы (англ. piechart) строятся с помощью функции pie()**

в каждую метку можно добавить процент данного округа в общей массе.

****

**Примеры использования гистограммы для обработки фотографий и оценки качества изделий.**

Анализ гистограммы яркости стал привычным делом во время

редактирования фотографий в графическом редакторе. Гистограмма

яркости помогает оценить общее качество фотографии на

основе математической интерпретации тонального диапазона

изображения в виде специальной диаграммы.

Т. е. гистограмма яркости представляет собой график, который

показывает, какие оттенки присутствуют в изображении. Диапазон

оттенков изображения представлен в виде последовательности

вертикальных линий, расположенных слева направо от самого темного

до самого светлого. Высота каждой линии показывает, сколько в

изображении пикселей соответствующего оттенка.

Гистограмма является вспомогательной функцией для оценки

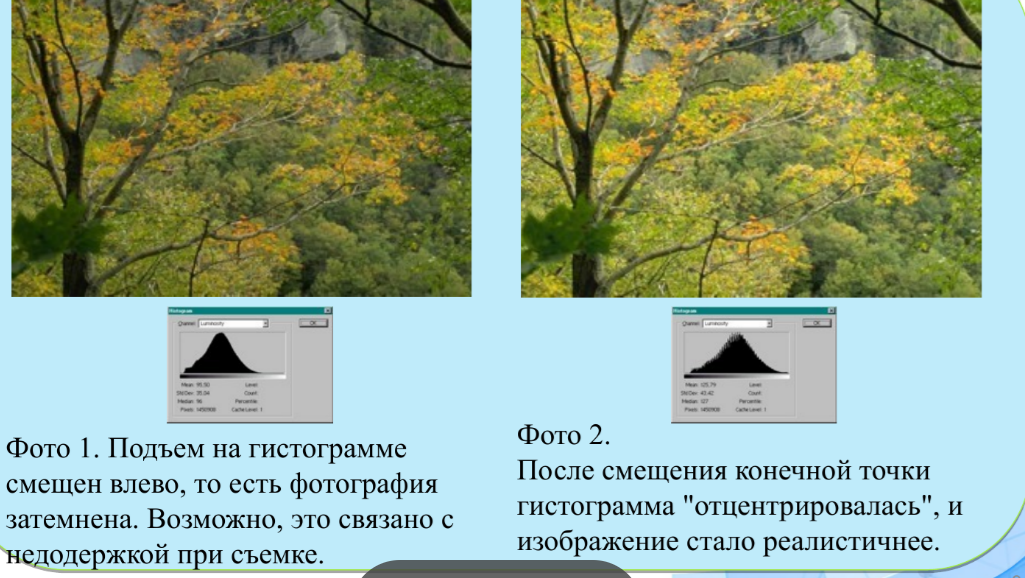
качества фотографии.

Слева на графике отображаются темные участки снимка -

тени. Справа - светлые участки. Если подъемы расположены

примерно симметрично, ближе к центру гистограммы, то

изображение хорошо сбалансировано.

****

**Контроль качества**

Пользуясь шириной интервалов как основанием, строят прямоугольники,

высота каждого из которых равна частоте попадания результатов

наблюдений в соответствующий интервал.

**Виды столбчатых диаграмм и их интерпретация.**

Столбчатые диаграммы - это графический способ представления данных в виде столбцов, где высота каждого столбца пропорциональна значению, которое он представляет. График в виде столбчатой диаграммы предназначен для сравнения ряда дискретных категорий: по одной её оси расположены сравниваемые категории, по второй — измеряемая величина.

Существует несколько видов столбчатых диаграмм, таких как вертикальные столбчатые диаграммы(гистограммы), гистограммы с накоплением (гистограмма, которая показывает накопленную сумму. Такая диаграмма используется для визуализации вклада каждой составляющей в общий результат),гистограмма с группировкой (это вариант гистограммы с накоплением, в которой сегменты расположены рядом друг с другом)и горизонтальные столбчатые диаграммы(линейчатая диаграмма),линейчатая диаграмма с накоплением и группировкой(аналогично гистограмме, только горизонтально), группированные и сложенные столбчатые диаграммы и другие.

Вертикальные столбчатые диаграммы используются для сравнения значений между различными категориями. Горизонтальные столбчатые диаграммы используются для сравнения значений внутри одной категории.

Группированные столбчатые диаграммы используются для сравнения значений между двумя или более группами. Сложенные столбчатые диаграммы используются для сравнения значений между двумя или более категориями внутри одной группы.

**Boxplot и его интерпретация, связь этого графика с другими элементами анализа.**

Boxplot - это график, который показывает распределение данных. Он состоит из коробки (поэтому он и называется box plot), усиков и точек. Коробка показывает интерквартильный размах распределения, то есть соответственно 25% ( Q1 ) и 75% ( Q3 ) перцентили. Усы показывают диапазон данных, не являющихся выбросами. Точки представляют выбросы.

Boxplot может быть использован для сравнения распределений между различными наборами данных или для отображения распределения одного набора данных.

Boxplot может быть использован вместе с другими графиками для анализа данных. Например, его можно использовать вместе с гистограммой для лучшего понимания распределения данных.

**Для чего нужны гипотезы в анализе данных, какие существуют приемы работы с гипотезами?**

Гипотеза – предположение о свойстве какого-либо параметра

Тестирование гипотезы (hypothesis testing) –процедура, в которой мы решаем, принять гипотезу или отвергнуть (reject).

Гипотеза формулируется о свойствах генеральной совокупностиОпровергнуть гипотезу в принципе легче, чем подтвердить

Формулируем ДВЕ взаимоисключающие гипотезы:

H0 (нулевая гипотеза, hypothesis) – её мы собираемся опровергать

H1 (альтернативная гипотеза, alternative hypothesis) – её мы примем,если удастся отвергнуть H0 .

Итак, нужен критерий, который позволит нам решить,отвергнуть ли Н0

Решение о том, принять или отвергнуть гипотезу принимается на основе статистики критерия (test statistic)– характеристики выборки.

В большинстве практических ситуаций принято считать, что нулевая гипотеза Н0 соответствует состоянию «по умолчанию»

Соответственно, альтернативная гипотеза Н1 обозначает противоположную ситуацию, которая обычно трактуется как менее вероятная, неординарная, требующая какой-либо реакции.

Максимальный шанс (вероятность) допустить ошибку 1-го рода обозначается α (альфа).Нулевую гипотезу отвергают, если наше значение p ниже уровня значимости, т. е., если p < α.

Мощность — это вероятность отклонения нулевой гипотезы,когда она ложна.

**Общая схема тестирования гипотезы:**

Формулируем Н0 и Н1. Строим распределения такие, как будто Н0 верна:

2 Устанавливаем условия, при которых мы отвергнем Н0 –

определяем:уровень значимости;

Считаем параметр выборки и статистику критерия для реальной

выборки, сравниваем их с критическими значениями.

4 Интерпретируем результаты:

Можем ли мы отвергнуть Н0? Т.е., достоверны ли результаты

статистически?Если да, достоверны ли они ПРАКТИЧЕСКИ?