УДК 004

Графическое представление квартета Ансамба на языке программирования Python

Кизянов Антон Олегович
Приамурский государственный университет имени Шолом-Алейхема
Студент

Аннотация
В данной статье будет рассказано, что такое квартет Ансамба и написана программа для его визуализации.
Ключевые слова: Python, matplotlib

Graphical representation of the Quartet Ensemble in the Python programming language

Kizyanov Anton Olegovich
Sholom-Aleichem Priamursky State University
student

Abstract
This article will tell you what Ansamba quartet is and write a program for its visualization.
Keywords: Python, matplotlib

Квартет Ансамба — четыре набора числовых данных, у которых простые статистические свойства идентичны, но их графики существенно отличаются. Каждый набор состоит из 11 пар чисел. Квартет был составлен в 1973 году английским математиком Ф. Дж. Энскомбом для иллюстрации важности применения графиков для статистического анализа, и влияния выбросов значений на свойства всего набора данных.

Цель исследования – демонстрация на примере квартета Ансамба важности представления данных в разных формах.

Ранее этим вопросом интересовались А.В. Петрухин, А.С. Стешенко развивали тему «Компьютерная визуализация биржевых данных о динамике фондового рынка» [1] в которой рассказывается про типы формализованных отображений информации, получаемой с биржи посредством торговых терминалов. Показаны проблемы, возникающие при построении специализированных модулей биржевой визуализации. Предлагаются методики построения подсистем визуализации с использованием библиотек OxyPlot и matplotlib, упрощающие процесс построения соответствующих программных приложений. Е.С.Могирева с темой «Визуализация информации: наглядное отображение количественной информации» [2], а подробнее про извлечение из данных первоначальную полезную

информацию с помощью их визуализации. П.И. Балк, А.С. Долгаль опубликовали статью «Сплайн-сглаживание экспериментальных данных при нулевом медианном значении помех» [3], в которой рассказали про алгоритм вычисления оценок параметров сглаживающего кубического сплайна, минимизирующих оценку математического ожидания потерь.

Ансамбский квартет - классический пример, иллюстрирующий, почему важно визуализировать данные. Квартет состоит из четырех наборов данных со сходными статистическими свойствами. Каждый набор данных имеет ряд значений x и зависимых значений y. Однако, если построить эти наборы данных, они будут выглядят разными по сравнению друг с другом.

Сначала нужно импортировать все нужные библиотеки.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib as mpl
from dautil import report
from dautil import plotting
import numpy as np
from tabulate import tabulate
```

Определить функцию, чтобы вычислить среднее значение, дисперсию и корреляцию x и y внутри набора данных.

```
df = sns.load_dataset("anscombe")
    agg = df.groupby('dataset')\
              .agg([np.mean, np.var])\
              .transpose()
    groups = df.groupby('dataset')
    corr = [g.corr()['x'][1] for _, g in groups]
    builder = report.DFBuilder(agg.columns)
    builder.row(corr)
    fits = [np.polyfit(g['x'], g['y'], 1) for _, g in groups]
    builder.row([f[0] for f in fits])
    builder.row([f[1] for f in fits])
    bottom = builder.build(['corr', 'slope', 'intercept'])
    return df, pd.concat((agg, bottom))
```

Следующая функция возвращает строку, которая является частично Markdown, частично реструктурированным текстом, и частично HTML, потому что Markdown официально не поддерживает таблицы:

```
def generate(table):
    writer = report.RSTWriter()
    writer.h1('Anscombe Statistics')
    writer.add(tabulate(table, tablefmt='html', floatfmt='.3f'))
    return writer.rst
```

Графики будут строиться благодаря функции lmplot():

```
def plot(df):
    sns.set(style="ticks")
    g = sns.lmplot(x="x", y="y", col="dataset",
         hue="dataset", data=df,
         col_wrap=2, ci=None, palette="muted", size=4,
         scatter_kws={"s": 50, "alpha": 1})
    plotting.embellish(g.fig.axes)
```

Отобразить таблицу квартета можно с помощью кода ниже:

```
df, table = aggregate()
from IPython.display import display_markdown
display_markdown(generate(table), raw=True)
```

Таблица квартета Ансамба изображена на рисунке 1.

Anscombe Statistics

('x', 'mean')	9.000	9.000	9.000	9.000
('x', 'var')	11.000	11.000	11.000	11.000
('y', 'mean')	7.501	7.501	7.500	7.501
('y', 'var')	4.127	4.128	4.123	4.123
corr	0.816	0.816	0.816	0.817
slope	0.500	0.500	0.500	0.500
intercept	3.000	3.001	3.002	3.002

Рис. 1. Таблица квартета Ансамба

На рисунке 2 изображено четыре графика построенных по данным с рисунка 1.

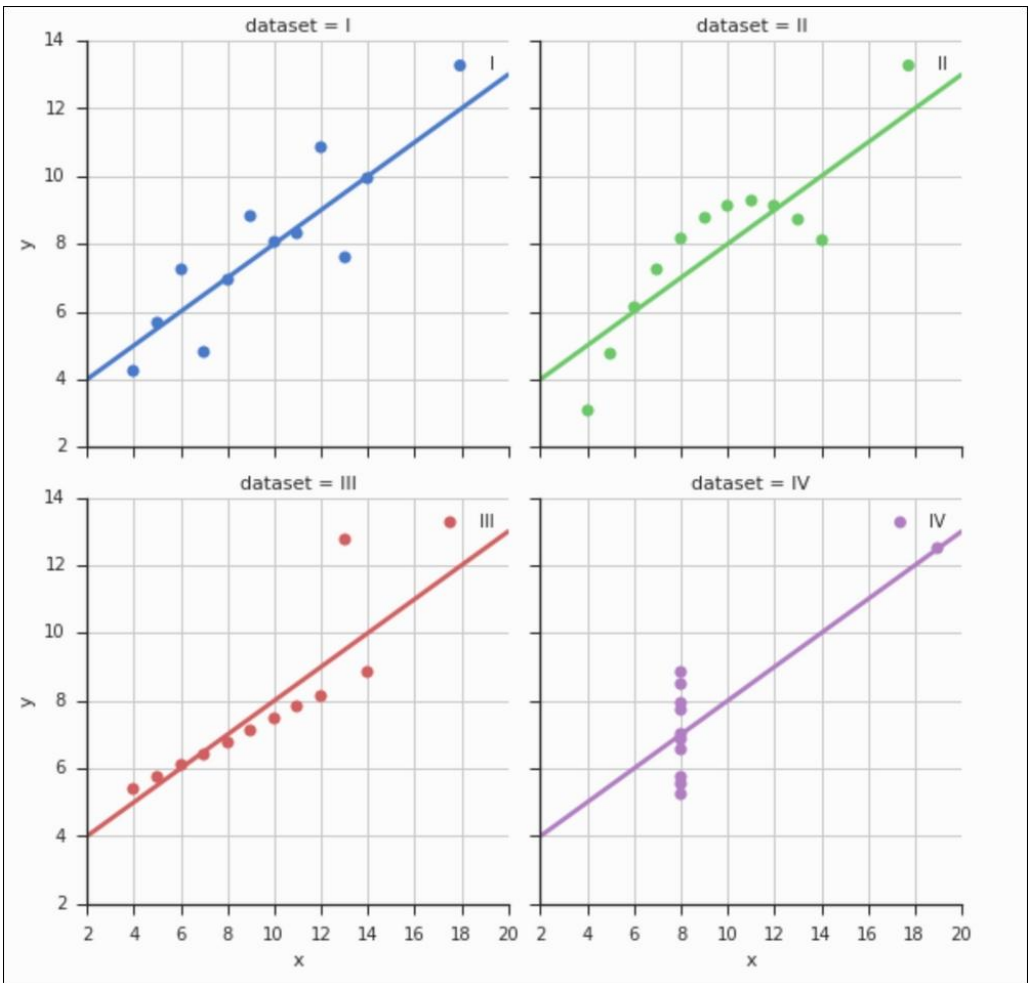

Рис. 2. Графики по четырем столбцам квартета Ансамба

Вывод

Таким образом, можно сделать вывод, что не всегда стоит доверять данным представленным в виде чисел, иногда необходимо использовать и другие способы представления, например графический.

Библиографический список

1. Петрухин А.В., Стешенко А.С. Компьютерная визуализация биржевых данных о динамике фондового рынка // Известия Волгоградского государственного технического университета. 2015. С. 124-129. URL: https://elibrary.ru/item.asp?id=24334292 (Дата обращения: 11.07.2018)
2. Могирева Е.С. Визуализация информации: наглядное отображение количественной информации // Образовательные ресурсы и технологии. 2014. С. 18-22. URL: https://elibrary.ru/item.asp?id=22315293 (Дата обращения: 11.07.2018)
3. Балк П.И., Долгаль А.С. Сплайн-сглаживание экспериментальных данных при нулевом медианном значении помех // Известия Волгоградского государственного технического университета. 2017. С. 138-156. URL: https://elibrary.ru/item.asp?id=24334292 (Дата обращения: 11.07.2018)