

УДК 00.004.896

АНАЛИТИКА ВОЗМОЖНОСТЕЙ ВИЗУАЛИЗАЦИИ ДАННЫХ В РАЗНООБРАЗНЫХ ТЕМАХ ОФОРМЛЕНИЯ НА ОСНОВЕ БИБЛИОТЕК MATPLOTLIB И SEABORN

Пылов П. А., студент группы ИТб-162, IV курс

Протодьяконов А.В.,¹ к.т.н., доцент

Кузбасский государственный технический университет имени Т.Ф. Горбачева
г. Кемерово

Изображение простых, но в то же время красивых графиков помогает предоставить необходимую информацию максимально понятно и визуально более располагает аудиторию (визуализация также играет ключевую роль в донесении количественной информации). В такой ситуации необходимо таким образом представить данные, который привлечёт внимание зрителя.

Библиотека matplotlib² обладает широкими возможностями настройки, но сложна для быстрого понимания. Необходимо выставлять правильные настройки для достижения визуально приятного графика. Библиотека Seaborn поставляется с несколькими настроенными темами и высокоуровневым интерфейсом для управления внешним видом фигур в matplotlib (рисунок 1)

```
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

Рисунок 1. Добавление библиотек в проект решения.

Определим простую функцию для построения некоторых смешанных синусоидальных волн, которые могут помочь увидеть различные стилистические параметры. Эти параметры можно настроить функцией, пример настройки приведён на рисунке 2

¹ Научный руководитель

² <https://matplotlib.org/3.1.1/index.html#>

```
def sinplot(flip=1):  
    x = np.linspace(0, 14, 100)  
    for i in range(1, 7):  
        plt.plot(x, np.sin(x + i * .5) * (7 - i) * flip)
```

Рисунок 2. Функция для построения графиков.

График по умолчанию в matplotlib на рисунке 3

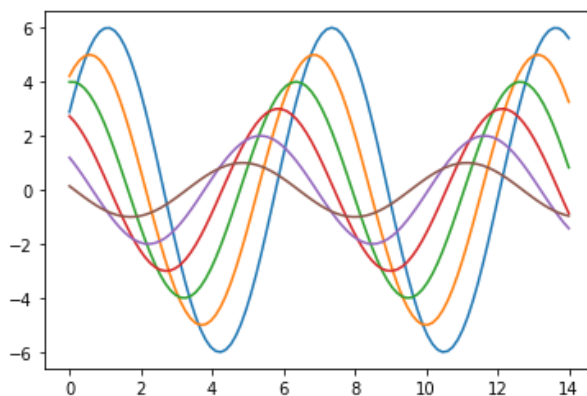


Рисунок 3. График синусоидальных зависимостей.

Вызываем функцию `set()` для переключения на библиотеку `seaborn` по умолчанию в языке программирования (рисунок 4)

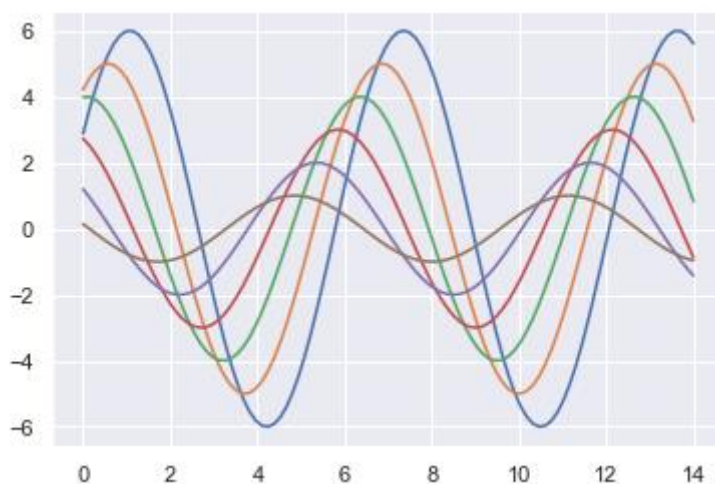


Рисунок 4. Переключение в графики `seaborn`.

`Seaborn` действует с параметрами `matplotlib` разделением на две независимые группы. Первая группа устанавливает эстетический стиль графика,

а вторая масштабирует различные элементы фигуры, чтобы ее можно было легко включить в различные контексты.

Интерфейсом для управления этими параметрами являются две пары функций. Для управления стилем используются функции `axes_style()` и `set_style()`. Для масштабирования графика используются функции `plotting_context()` и `set_context()`. В обоих случаях первая функция возвращает словарь параметров, а вторая устанавливает значения по умолчанию для `matplotlib`.

Библиотекой предусмотрено пять предустановленных тем: темная тема с координатной сеткой, белая тема с сеткой, темная тема, белая и тема в виде формы скрипки. Каждая из них подходит для различных целей. Тема по умолчанию – темная координатная сетка. Как упомянуто выше, координатная сетка превращает график в справочник для количественной информации, а белый цвет помогает не допустить противоречий в координатной сетке с линиями, которые представляют данные (рисунок 5)

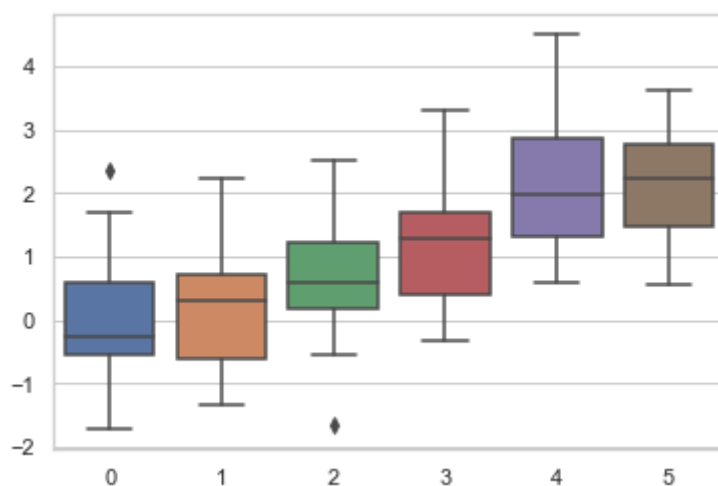


Рисунок 5. Представление данных в виде диаграммы ящика на координатной сетке.

Для многих графиков, где точные числовые значения не играют важную роль (иначе можно сформулировать, что график доносит лишь визуальную первичную оценку данных, а не прецизионную аналитическую информацию) сетка является менее необходимой, её можно убрать (рисунок 6)

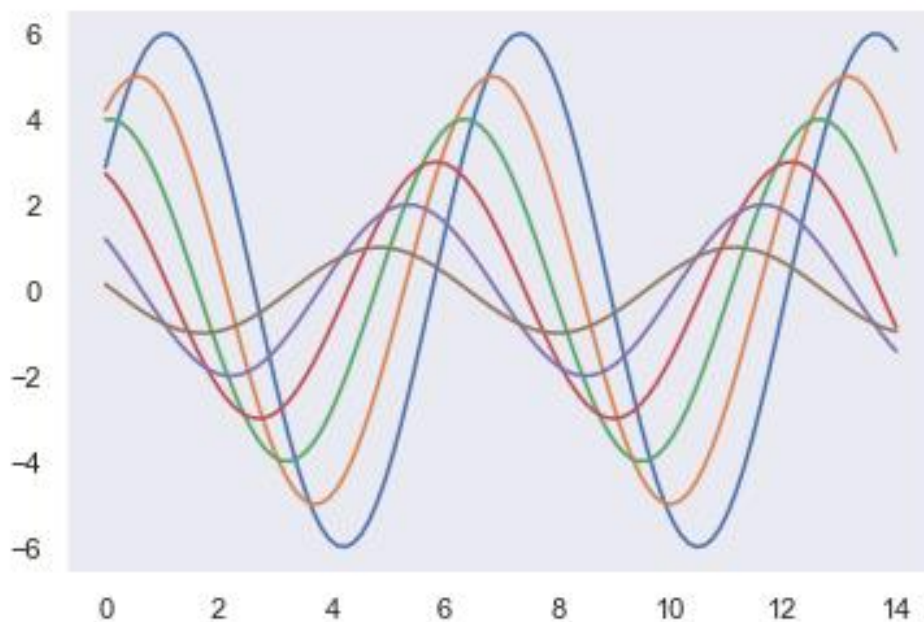


Рисунок 6. График без координатной сетки.

Для придания дополнительной структуры графикам, можно использовать аргумент «ticks» в `set_style()` (пример работы приведён на рисунках 7-8)

```
sns.set_style("ticks")  
sinplot()
```

Рисунок 7. Использование «ticks».

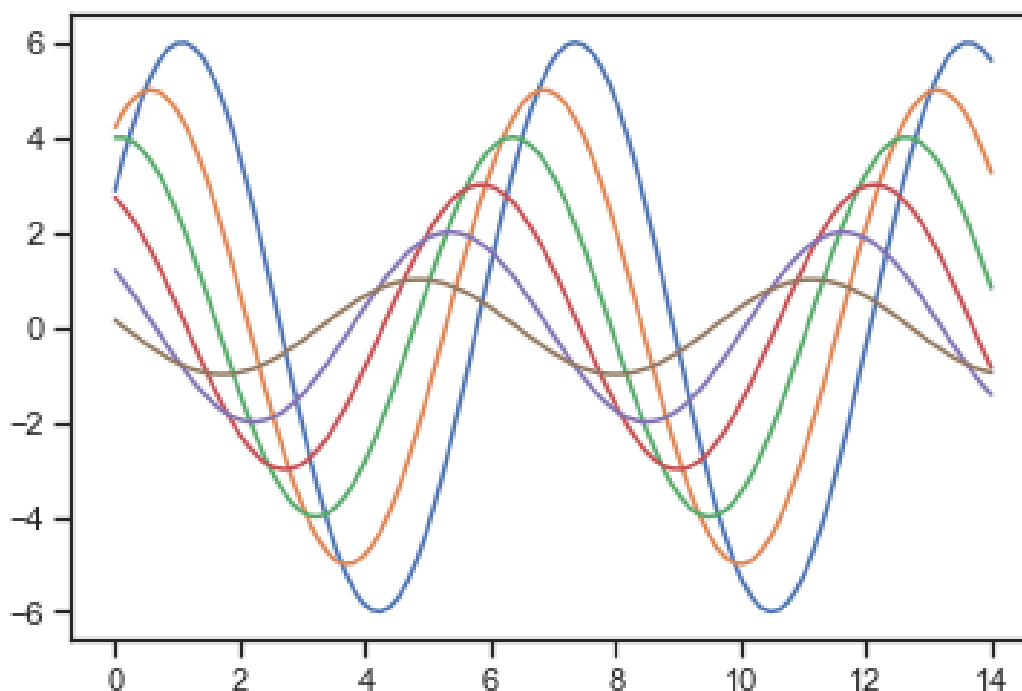


Рисунок 8. Дополнительная структура графика.

Применение различных параметров позволяет гибко подстраивать визуализацию данных, создавать более приятный и читабельный вид. Знание параметров и основных методов библиотеки существенно облегчает задачу аналитики данных, позволяя за короткие промежутки времени представлять и анализировать большие объёмы данных. Такие приёмы экономят время и ресурсы Data–Science – специалистов, позволяя сконцентрироваться на других этапах своей деятельности.

Список литературы:

1. M. Narasimha Murty, V. Susheela Devy. Introduction to pattern recognition and machine learning, IISc press, New Jersey – London, 2015
2. S. Theodoridis, A. Pikrakis. Introduction to Pattern Recognition: A MATLAB Approach, Elsevier press, 2010
3. Christopher Bishop. Pattern Recognition and Machine Learning, Plenum press, New York – London, 1971
4. A. Geron. Hands-on Machine learning with Scikit-Learn, Keras, and TensorFlow, OREILLY Sebastopol, California – USA, 2019
5. P. Bruce, A. Bruce. Practical statistics for data scientists, OREILLY Sebastopol, California – USA, 2017