



Инструменты для визуализации данных



Олег Булыгин



Олег Булыгин

- Преподаватель на курсах “Основы языка программирования Python”, “Продвинутый Python”, “Python для анализа данных” в Нетологии
- Начальник бюро планирования и управления в АО “НПО автоматики”

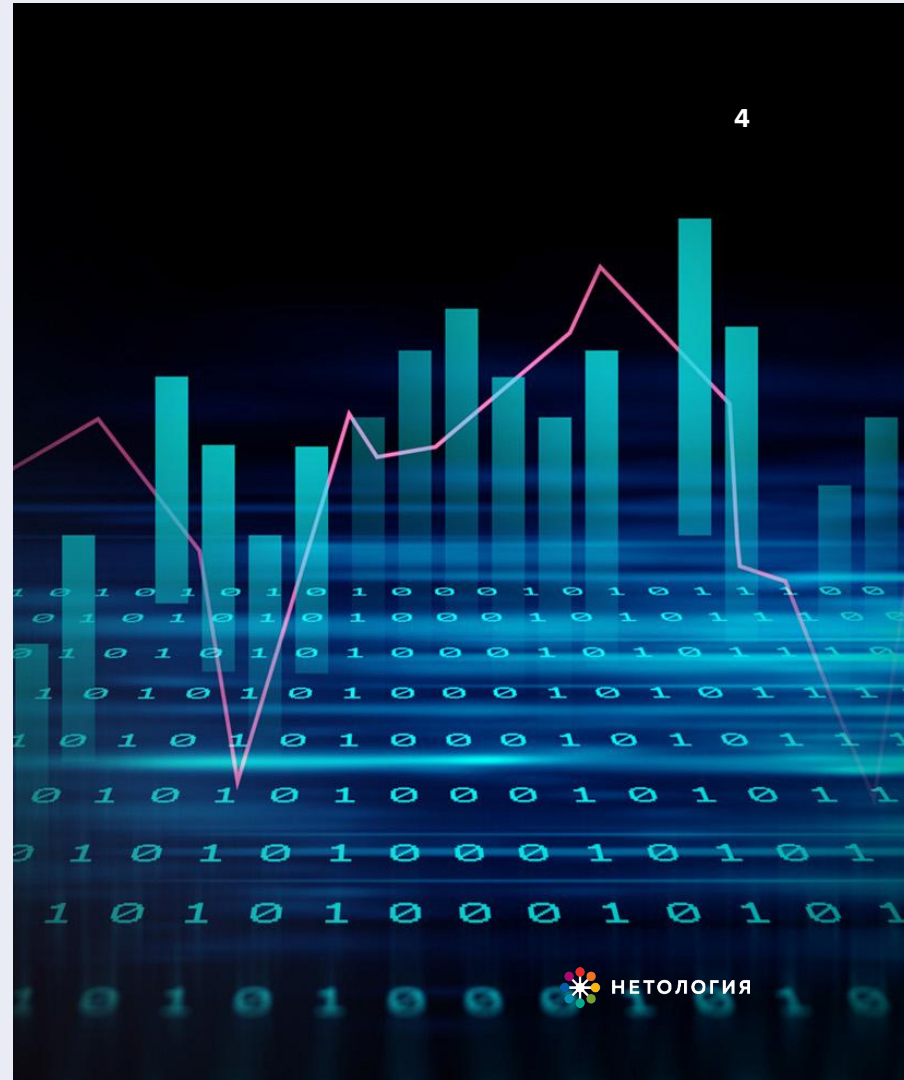
О чём мы поговорим сегодня

3

1. Зачем нужна визуализация
2. Matplotlib
3. Seaborn
4. Plotly

Что такое визуализация данных?

Визуализация данных — это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению.





**А нужна ли
визуализация вообще?**

Пример выборок

6

Все статистики четырех выборок одинаковы:

- $\text{mean } x = 9$
- $\text{mean } y = 11.5$
- выборочная дисперсия $x = 11$
- выборочная дисперсия $y = 4.125$
- корреляция между x и $y = 0.816$

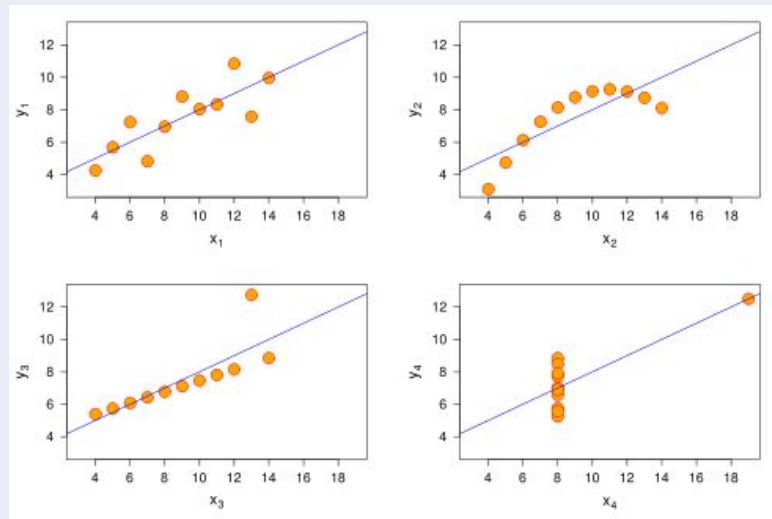
Квартет Энскомба							
I		II		III		IV	
x	y	x	y	x	y	x	y
10,0	8,04	10,0	9,14	10,0	7,46	8,0	6,58
8,0	6,95	8,0	8,14	8,0	6,77	8,0	5,76
13,0	7,58	13,0	8,74	13,0	12,74	8,0	7,71
9,0	8,81	9,0	8,77	9,0	7,11	8,0	8,84
11,0	8,33	11,0	9,26	11,0	7,81	8,0	8,47
14,0	9,96	14,0	8,10	14,0	8,84	8,0	7,04
6,0	7,24	6,0	6,13	6,0	6,08	8,0	5,25
4,0	4,26	4,0	3,10	4,0	5,39	19,0	12,50
12,0	10,84	12,0	9,13	12,0	8,15	8,0	5,56
7,0	4,82	7,0	7,26	7,0	6,42	8,0	7,91
5,0	5,68	5,0	4,74	5,0	5,73	8,0	6,89

Квартет Энksomба

7

Пример был придуман **Фрэнсисом Энksomбом** в 1973 году и демонстрирует:





































































- важность визуализация для анализа данных;
- влияние выбросов на статистические показатели.




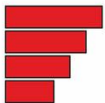
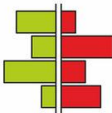
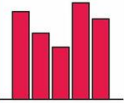
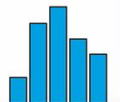
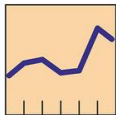
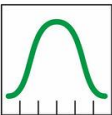



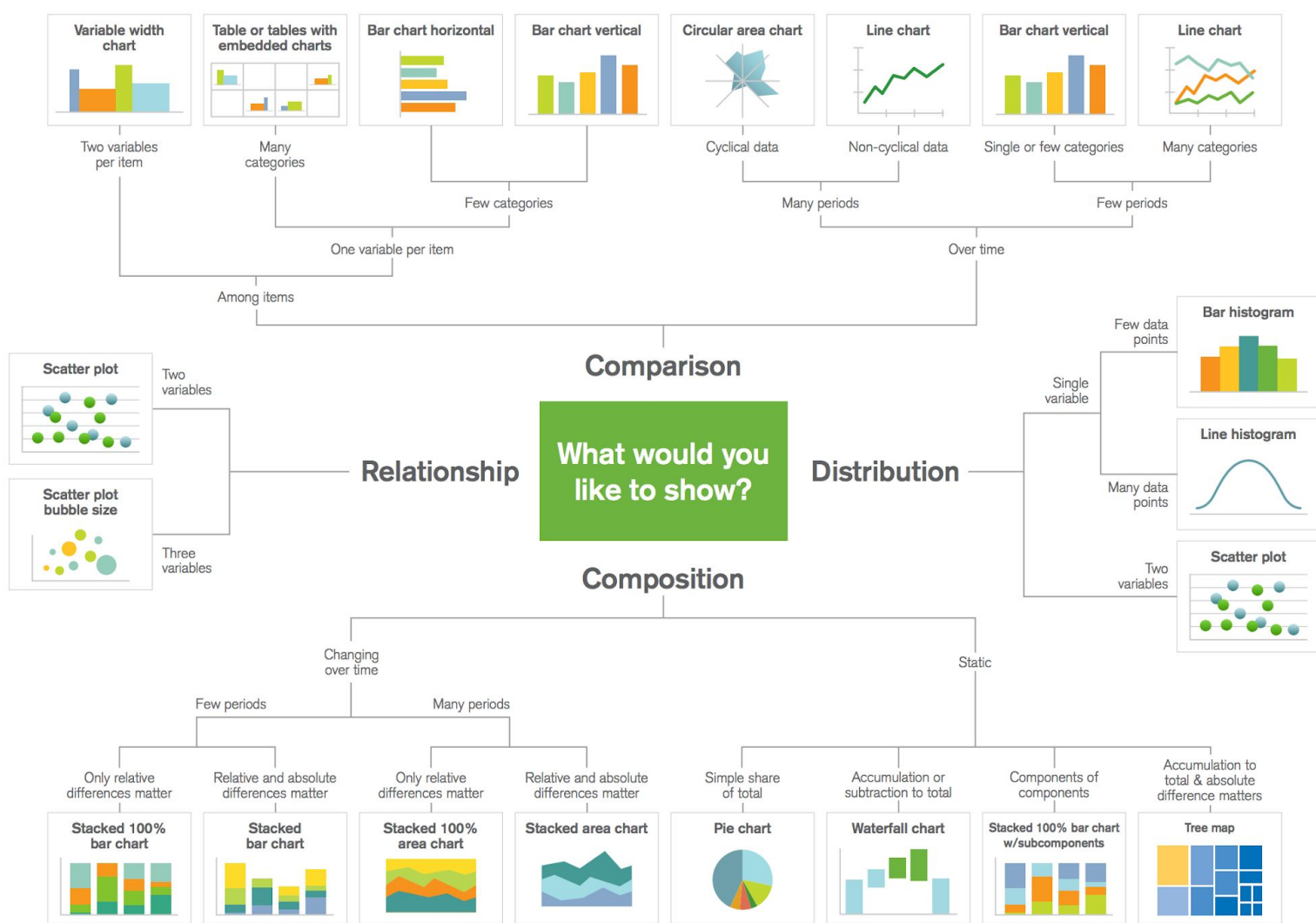
**Какие типы графиков вы
знаете?**

A PERIODIC TABLE OF VISUALIZATION METHODS

<div></div> <div>C</div> <div>continuum</div>		<div></div> <div>Data Visualization</div> <div>Visual representations of quantitative data in schematic form (either with or without axes)</div>										<div></div> <div>Strategy Visualization</div> <div>The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.</div>										<div></div> <div>G</div> <div>graphic facilitation</div>	
<div></div> <div>Tb</div> <div>table</div>	<div></div> <div>Ca</div> <div>cartesian coordinates</div>	<div></div> <div>Information Visualization</div> <div>The use of interactive visual representations of data to amplify cognition. This means that the data is transformed into an image, it is mapped to screen space. The image can be changed by users as they proceed working with it</div>										<div></div> <div>Metaphor Visualization</div> <div>Visual Metaphors position information graphically to organize and structure information. They also convey an insight about the represented information through the key characteristics of the metaphor that is employed</div>										<div></div> <div>Ct</div> <div>cartoon</div>	
<div></div> <div>Pi</div> <div>pie chart</div>	<div></div> <div>L</div> <div>line chart</div>	<div></div> <div>Concept Visualization</div> <div>Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.</div>										<div></div> <div>Compound Visualization</div> <div>The complementary use of different graphic representation formats in one single schema or frame</div>										<div></div> <div>Ri</div> <div>rich picture</div>	
<div></div> <div>B</div> <div>bar chart</div>	<div></div> <div>Ac</div> <div>area chart</div>	<div></div> <div>R</div> <div>radar chart cobweb</div>	<div></div> <div>Pa</div> <div>parallel coordinates</div>	<div></div> <div>Hy</div> <div>hyperbolic tree</div>	<div></div> <div>Cy</div> <div>cycle diagram</div>	<div></div> <div>T</div> <div>timeline</div>	<div></div> <div>Ve</div> <div>venn diagram</div>	<div></div> <div>Mi</div> <div>mindmap</div>	<div></div> <div>Sq</div> <div>square of oppositions</div>	<div></div> <div>Cc</div> <div>concentric circles</div>	<div></div> <div>Ar</div> <div>argument slide</div>	<div></div> <div>Sw</div> <div>swim lane diagram</div>	<div></div> <div>Gc</div> <div>gant chart</div>	<div></div> <div>Pm</div> <div>perspectives diagram</div>	<div></div> <div>D</div> <div>dilemma diagram</div>	<div></div> <div>Pr</div> <div>parameter ruler</div>	<div></div> <div>Kn</div> <div>knowledge map</div>						
<div></div> <div>Hi</div> <div>histogram</div>	<div></div> <div>Sc</div> <div>scatterplot</div>	<div></div> <div>Sa</div> <div>sankey diagram</div>	<div></div> <div>In</div> <div>information lense</div>	<div></div> <div>E</div> <div>entity relationship diagram</div>	<div></div> <div>Pt</div> <div>petri net</div>	<div></div> <div>Fl</div> <div>flow chart</div>	<div></div> <div>Cl</div> <div>clustering</div>	<div></div> <div>Lc</div> <div>layer chart</div>	<div></div> <div>Py</div> <div>minto pyramid technique</div>	<div></div> <div>Ce</div> <div>cause-effect chains</div>	<div></div> <div>Tl</div> <div>toulmin map</div>	<div></div> <div>Dt</div> <div>decision tree</div>	<div></div> <div>Cp</div> <div>cpm critical path method</div>	<div></div> <div>Cf</div> <div>concept fan</div>	<div></div> <div>Co</div> <div>concept map</div>	<div></div> <div>Ic</div> <div>iceberg</div>	<div></div> <div>Lm</div> <div>learning map</div>						
<div></div> <div>Tk</div> <div>tukey box plot</div>	<div></div> <div>Sp</div> <div>spectrogram</div>	<div></div> <div>Da</div> <div>data map</div>	<div></div> <div>Tp</div> <div>treemap</div>	<div></div> <div>Cn</div> <div>cone tree</div>	<div></div> <div>Sy</div> <div>system dyn./simulation</div>	<div></div> <div>Df</div> <div>data flow diagram</div>	<div></div> <div>Se</div> <div>semantic network</div>	<div></div> <div>So</div> <div>soft system modeling</div>	<div></div> <div>Sn</div> <div>synergy map</div>	<div></div> <div>Fo</div> <div>force field diagram</div>	<div></div> <div>Ib</div> <div>ibis argumentation map</div>	<div></div> <div>Pr</div> <div>process event chains</div>	<div></div> <div>Pe</div> <div>pert chart</div>	<div></div> <div>Ev</div> <div>evocative knowledge map</div>	<div></div> <div>V</div> <div>Vee diagram</div>	<div></div> <div>Hh</div> <div>heaven 'n' hell chart</div>	<div></div> <div>I</div> <div>informal</div>						

Как выбрать тип диаграммы?

		ТИПЫ СРАВНЕНИЯ				
		ПОКОМПОНЕНТНОЕ	ПОЗИЦИОННОЕ	ВРЕМЕННОЕ	ЧАСТОТНОЕ	КОРРЕЛЯЦИОННОЕ
ОСНОВНЫЕ ТИПЫ ДИАГРАММ	КРУГОВАЯ					
	ЛИНЕЙЧАТАЯ					
	ГИСТОГРАММА					
	ГРАФИК					
	ТОЧЕЧНАЯ					



Python библиотеки для визуализации

12

- matplotlib
- seaborn
- plotly
- ggplot
- bokeh
- pygal
- и т.д.

matplotlib

plotly

Seaborn

ggplot2



bokeh

Pygal

<https://catalog.data.gov/dataset/baby-names-from-social-security-card-applications-national-level-data>

<https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings>

Библиотека matplotlib – первая библиотека на python для визуализация. Очень гибкая и сравнительно монструозная.

Модуль pyplot позволяет пользователю сосредоточиться на выборе готовых решений и настройке базовых параметров рисунка.

Стандарт вызова pyplot в python:

import matplotlib.pyplot as plt

- *plt.plot(x, y)* – построение линейного графика;
- *plt.scatter(x, y)* – построение точечной диаграммы;
- *plt.bar(x, y)* – построение столбчатой диаграммы;
- *plt.hist(x)* – построение гистограммы;
- *plt.pie(x)* – построение круговой диаграммы;
- *plt.boxplot(x)* – построение boxplot (“ящик с усами”).

После непосредственно построения графика его можно отобразить при помощи `plt.show()`

Некоторые настройки графиков

16

plt.xlabel('Текст') и *plt.ylabel('Текст')* позволят подписать оси;

plt.title('Текст') – заголовок графика;

plt.grid() – добавляет сетку;

plt.text(x, y, 'Текст') – добавляет текст в нужной позиции на график.

Сохранение и изменение размеров графиков

17

При помощи функции ***plt.savefig('путь сохранения/название файла')*** можно сохранить график.

Изменить размер графика можно, применив функцию ***plt.figure(figsize=(x, y))*** перед созданием графика (размер указывается в дюймах).

Метод `.plot()`

18

Метод дата фрейма `.plot()` позволяет строить графики с некоторыми заданными первоначальными настройками (на самом деле “под капотом” используется библиотека *matplotlib*). Таким образом упрощается быстрое построение графиков для дата фреймов.

Метод `.plot()` как правило сам корректно определяет оси, учитывая индексы дата фреймов.

Аргумент `kind`, позволяет задать тип графика.

Seaborn

19

Библиотека seaborn – это библиотека визуализации python на основе matplotlib. Ее основная цель – предоставить краткий, высокоуровневый интерфейс для составления статистической графики.

Позволяет построить достаточно сложные графики за пару строк кода, имеет симпатичные default стили, однако для изменения мелочей потребуется лезть в дебри matplotlib.

Стандарт вызова pyplot в python:

import seaborn as sns

Библиотека plotly – изначально online-платформа, где можно создавать и публиковать свои графики. У нее есть offline-mode, который позволяет использовать ее без регистрации и публикации данных и графиков на сервер plotly.

Построена на *plotly.js*, которая, в свою очередь, базируется на *d3.js*.
Позволяет строить динамические графики.

Стандарт вызова pyplot в python:

```
import plotly  
from plotly.offline import init_notebook_mode, iplot  
import plotly.graph_objs as go  
init_notebook_mode(connected=True)
```

<http://pandas.pydata.org/pandas-docs/stable/visualization.html>

http://matplotlib.org/users/pyplot_tutorial.html

<https://github.com/d3/d3/wiki/Gallery>

<http://datavizcatalogue.com/>

<http://www.storytellingwithdata.com/>



Инструменты для визуализации данных

Вопросы?

Олег Булыгин

Соцсеть
fb.com/obulygin91

Почта
obulygin91@ya.ru