

# Инструменты для визуализации данных

Олег Булыгин

Нетология





---

## Булыгин Олег

- Преподаватель курсов по Python в Нетологии
- Начальник бюро планирования и управления в АО “НПО автоматики”



[o.bulygin@netology.ru](mailto:o.bulygin@netology.ru)



<https://www.facebook.com/obulygin91>



---

# О чем мы поговорим сегодня

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 выборки

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

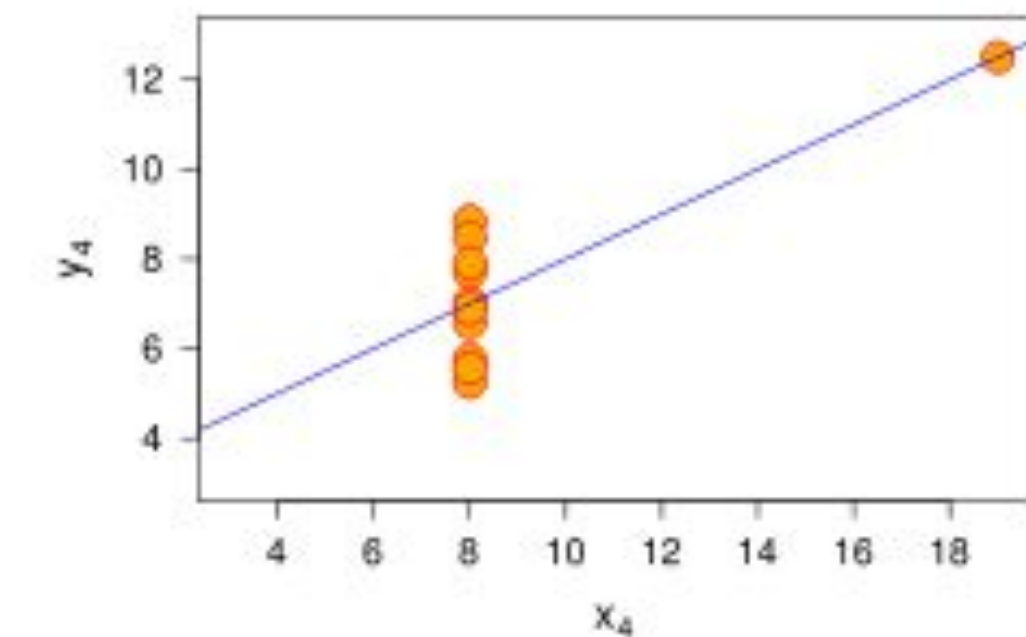
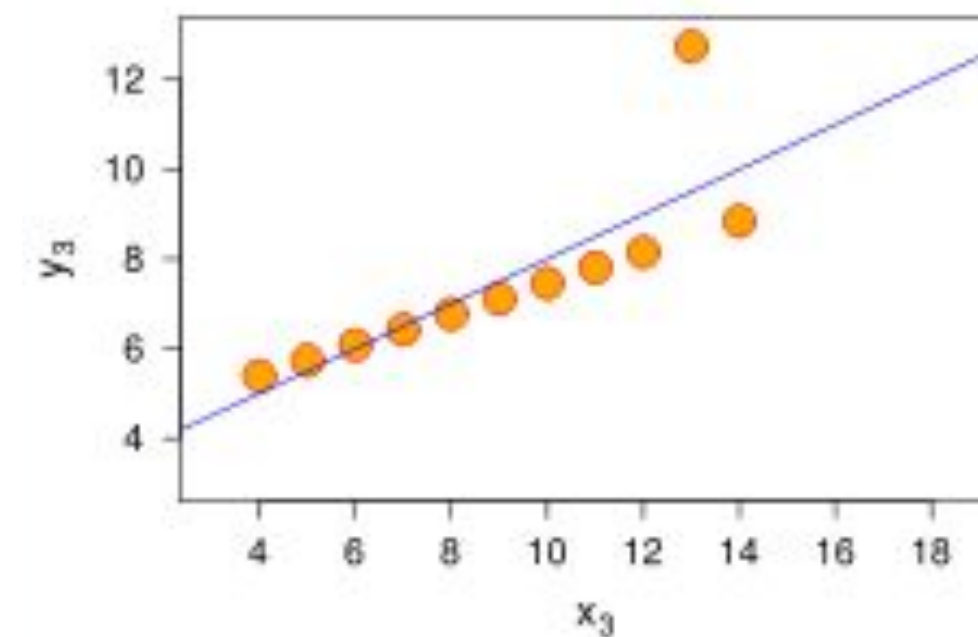
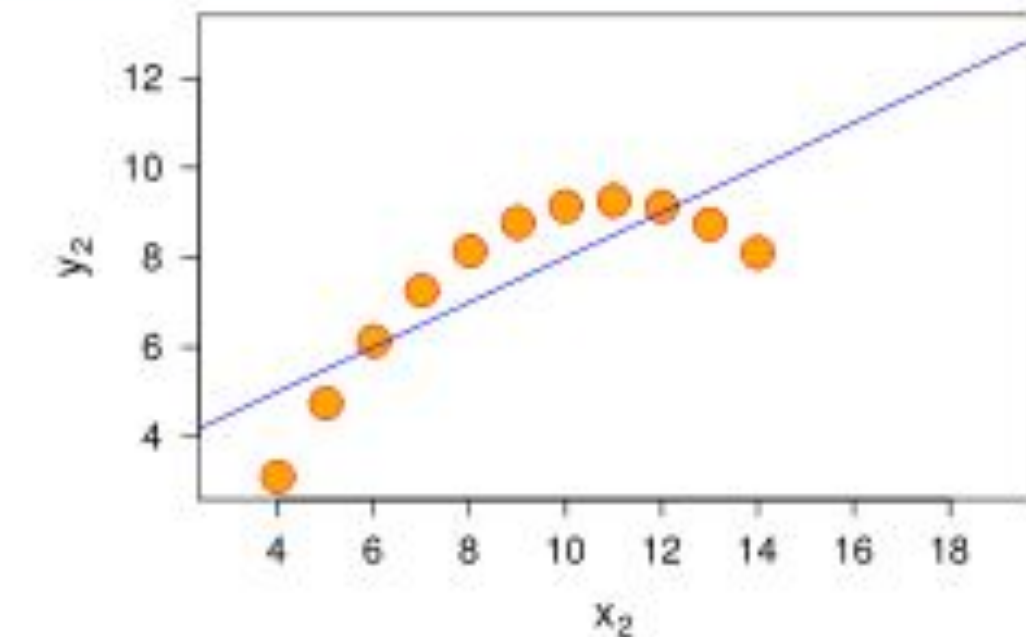
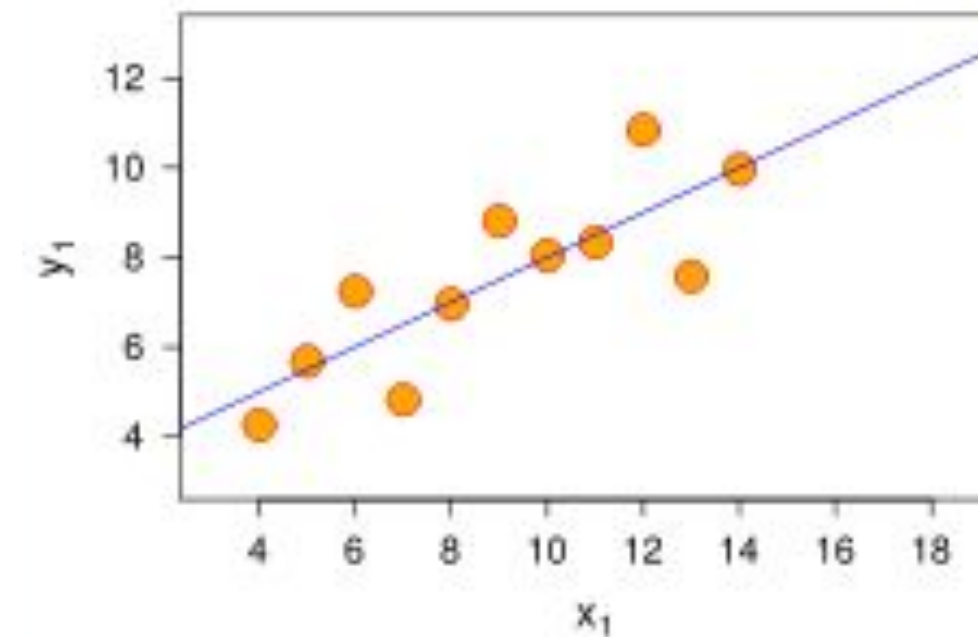
| I    |       | II   |      | III  |       | IV   |       |
|------|-------|------|------|------|-------|------|-------|
| x    | y     | x    | y    | x    | y     | x    | y     |
| 10,0 | 8,04  | 10,0 | 9,14 | 10,0 | 7,46  | 8,0  | 6,58  |
| 8,0  | 6,95  | 8,0  | 8,14 | 8,0  | 6,77  | 8,0  | 5,76  |
| 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

Пример был придуман  
Фрэнсисом Энksomбом в 1973  
году и демонстрирует:

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




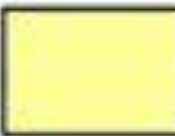
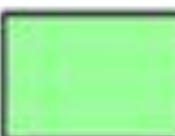

















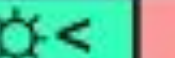

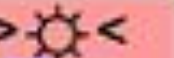











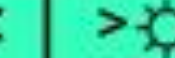

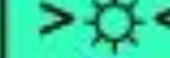








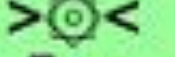








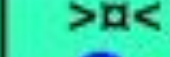






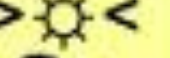


















Какие типы графиков вы  
знаете?



# A PERIODIC TABLE OF VISUALIZATION METHODS


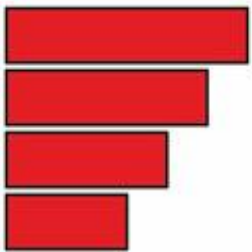
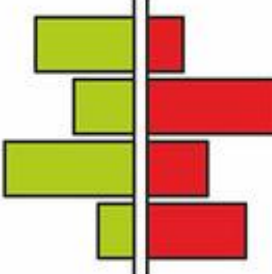

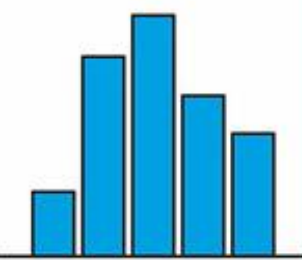
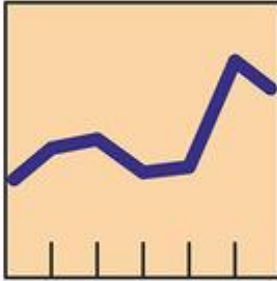
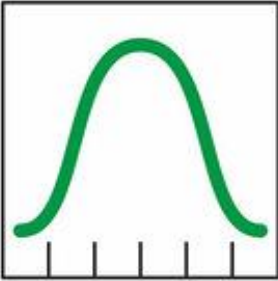



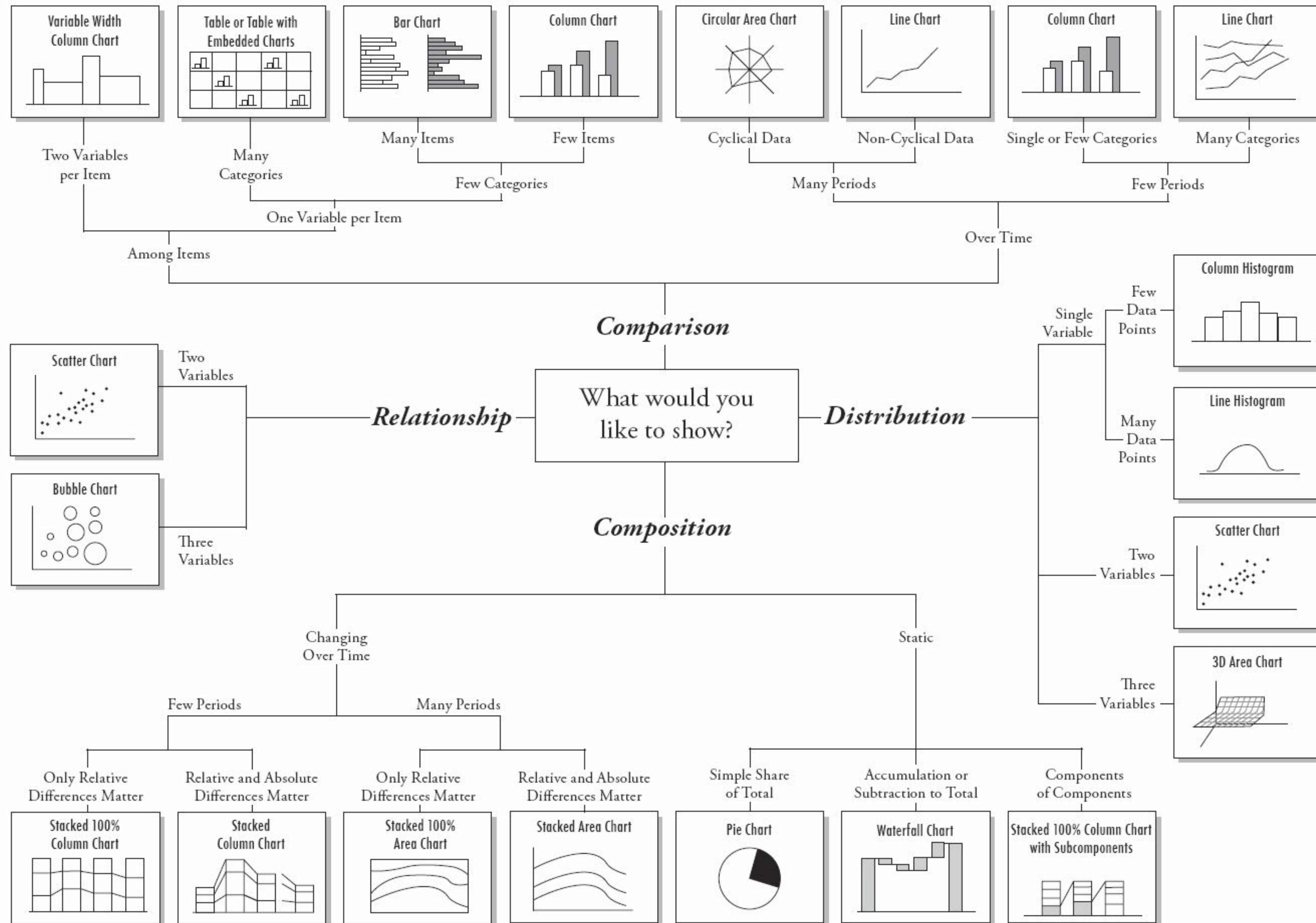
|   |  |   |  |   |  |  |   |   |  |  |   |   |   |  |  |  |   |   |   |   |   |
|---|--|---|--|---|--|--|---|---|--|--|---|---|---|--|--|--|---|---|---|---|---|
| <div><br/><b>C</b><br/>continuum</div>         | <div><div><div><b>Data Visualization</b><br/><i>Visual representations of quantitative data in schematic form (either with or without axes)</i></div><div><b>Information Visualization</b><br/><i>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</i></div><div><b>Concept Visualization</b><br/><i>Methods to elaborate (mostly) qualitative concepts, ideas, plans, and analyses.</i></div></div><div><div><b>Strategy Visualization</b><br/><i>The systematic use of complementary visual representations in the analysis, development, formulation, communication, and implementation of strategies in organizations.</i></div><div><b>Metaphor Visualization</b><br/><i>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</i></div><div><b>Compound Visualization</b><br/><i>The complementary use of different graphic representation formats in one single schema or frame</i></div></div></div> |   |  |   |  |  |   |   |  |  |   |   |   |  | <div><br/><b>G</b><br/>graphic facilitation</div> |  |   |   |   |   |   |
| <div><br/><b>Tb</b><br/>table</div>            | <div><br/><b>Ca</b><br/>cartesian coordinates</div>   |   |  |   |  |  |   |   |  |  |   |   |   |  |  | <div><br/><b>Me</b><br/>meeting trace</div>           | <div><br/><b>Mm</b><br/>metro map</div>      | <div><br/><b>Tm</b><br/>temple</div>           | <div><br/><b>St</b><br/>story template</div> | <div><br/><b>Tr</b><br/>tree</div>   | <div><br/><b>Ct</b><br/>cartoon</div>      |
| <div><br/><b>Pi</b><br/>pie chart</div>        | <div><br/><b>L</b><br/>line chart</div>   |   |  |   |  |  |   |   |  |  |   |   |   |  |  | <div><br/><b>Co</b><br/>communication diagram</div>   | <div><br/><b>Fp</b><br/>flight plan</div>    | <div><br/><b>Cs</b><br/>concept skeleton</div> | <div><br/><b>Br</b><br/>bridge</div>         | <div><br/><b>Fu</b><br/>funnel</div> | <div><br/><b>Ri</b><br/>rich picture</div> |
| <div><br/><b>B</b><br/>bar chart</div>        | <div><br/><b>Ac</b><br/>area chart</div>   | <div><br/><b>R</b><br/>radar chart<br/>cobweb</div> | <div><br/><b>Pa</b><br/>parallel coordinates</div> | <div><br/><b>Hy</b><br/>hyperbolic tree</div>             | <div><br/><b>Cy</b><br/>cycle diagram</div>           | <div><br/><b>T</b><br/>timeline</div>            | <div><br/><b>Ve</b><br/>venn diagram</div>      | <div><br/><b>Mi</b><br/>mindmap</div>               | <div><br/><b>Sq</b><br/>square of oppositions</div>    | <div><br/><b>Cc</b><br/>concentric circles</div>   | <div><br/><b>Ar</b><br/>argument slide</div>          | <div><br/><b>Sw</b><br/>swim lane diagram</div>     | <div><br/><b>Gc</b><br/>gantt chart</div>               | <div><br/><b>Pm</b><br/>perspectives diagram</div>     | <div><br/><b>D</b><br/>dilemma diagram</div>     | <div><br/><b>Pr</b><br/>parameter ruler</div>        | <div><br/><b>Kn</b><br/>knowledge map</div> |   |   |   |   |
| <div><br/><b>Hi</b><br/>histogram</div>      | <div><br/><b>Sc</b><br/>scatterplot</div>   | <div><br/><b>Sa</b><br/>sankey diagram</div>       | <div><br/><b>In</b><br/>information lense</div>   | <div><br/><b>E</b><br/>entity relationship diagram</div> | <div><br/><b>Pt</b><br/>petri net</div>              | <div><br/><b>Fl</b><br/>flow chart</div>        | <div><br/><b>Cl</b><br/>clustering</div>       | <div><br/><b>Lc</b><br/>layer chart</div>          | <div><br/><b>Py</b><br/>minto pyramid technique</div> | <div><br/><b>Ce</b><br/>cause-effect chains</div> | <div><br/><b>Tl</b><br/>toulmin map</div>            | <div><br/><b>Dt</b><br/>decision tree</div>        | <div><br/><b>Cp</b><br/>cpm critical path method</div> | <div><br/><b>Cf</b><br/>concept fan</div>             | <div><br/><b>Co</b><br/>concept map</div>       | <div><br/><b>Ic</b><br/>iceberg</div>               | <div><br/><b>Lm</b><br/>learning map</div> |   |   |   |   |
| <div><br/><b>Tk</b><br/>tukey box plot</div> | <div><br/><b>Sp</b><br/>spectrogram</div>   | <div><br/><b>Da</b><br/>data map</div>             | <div><br/><b>Tp</b><br/>treemap</div>             | <div><br/><b>Cn</b><br/>cone tree</div>                  | <div><br/><b>Sy</b><br/>system dyn./simulation</div> | <div><br/><b>Df</b><br/>data flow diagram</div> | <div><br/><b>Se</b><br/>semantic network</div> | <div><br/><b>So</b><br/>soft system modeling</div> | <div><br/><b>Sn</b><br/>synergy map</div>             | <div><br/><b>Fo</b><br/>force field diagram</div> | <div><br/><b>Ib</b><br/>ibis argumentation map</div> | <div><br/><b>Pr</b><br/>process event chains</div> | <div><br/><b>Pe</b><br/>pert chart</div>               | <div><br/><b>Ev</b><br/>evocative knowledge map</div> | <div><br/><b>V</b><br/>Vee diagram</div>        | <div><br/><b>Hh</b><br/>heaven 'n' hell chart</div> | <div><br/><b>I</b><br/>infomural</div>     |   |   |   |   |



# Как выбрать тип диаграммы?



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







# Python библиотеки для визуализации

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

**matplotlib**



bokeh

Pygal



# Файлы для практики

<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>





# Matplotlib

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*



# Модуль Pyplot

- *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

Функции *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

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()`

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

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*



# Plotly

Библиотека plotly – изначально online-платформа, где можно создавать и публиковать свои графики. У библиотеки 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)*



## Полезные ссылки

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

[http://matplotlib.org/users/pyplot\\_tutorial.html](http://matplotlib.org/users/pyplot_tutorial.html)

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

<http://datavizcatalogue.com/>

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



# Спасибо за внимание!

Олег Булыгин

Нетология



<https://www.facebook.com/obulygin91>



[o.bulygin@netology.ru](mailto:o.bulygin@netology.ru)



<https://vk.com/obulygin91>



<https://www.linkedin.com/in/obulygin>