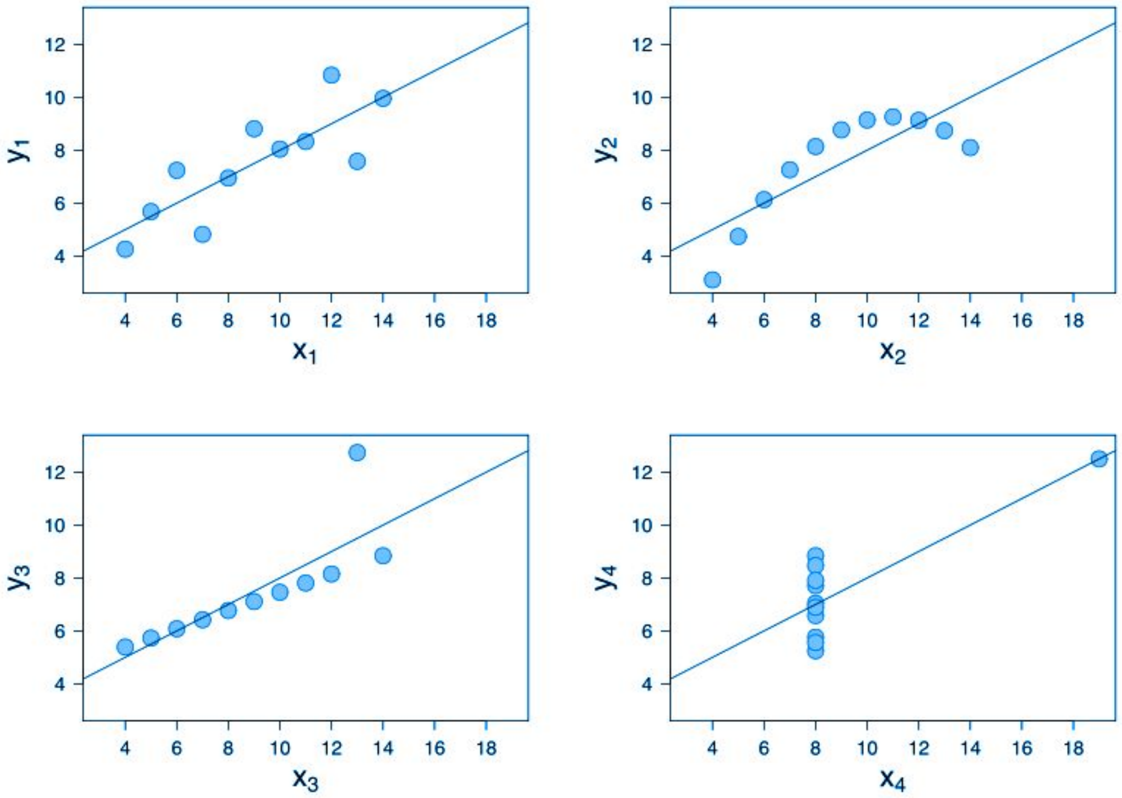
Wizualizacja danych z R

Mateusz Staniak
2 XII 2019

Dlaczego

	I		II		III		IV	
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

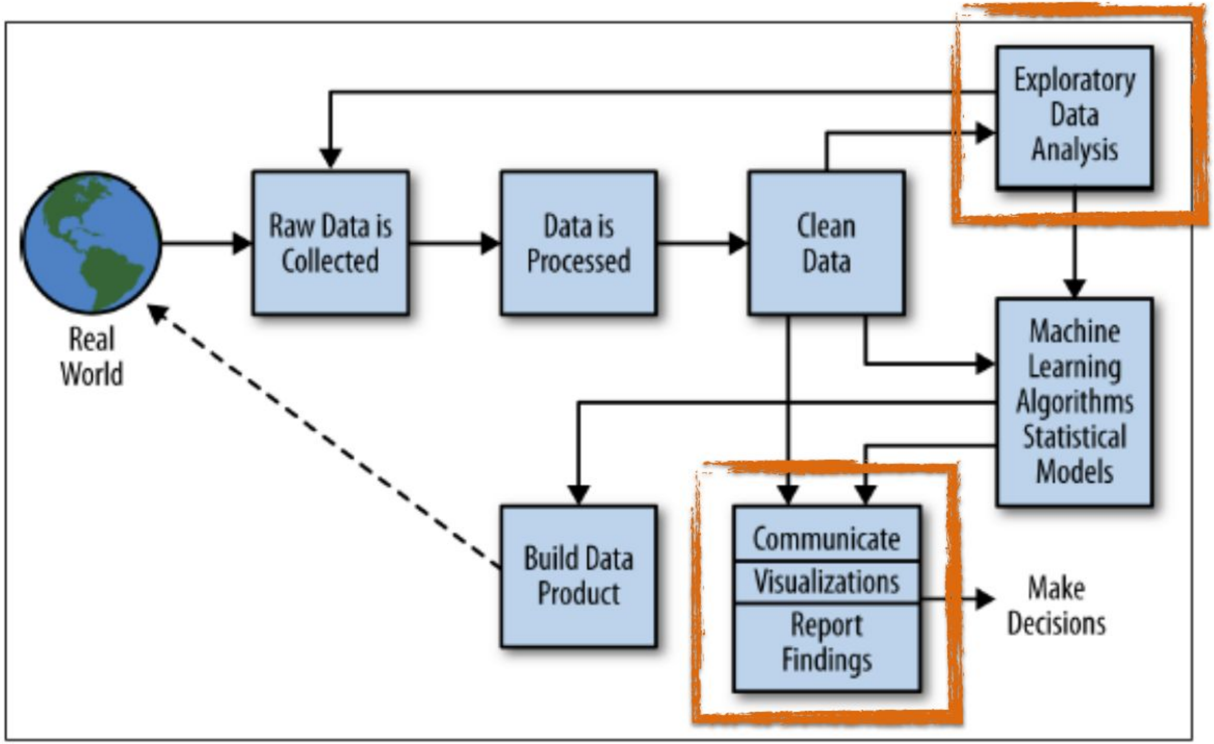
Kwartet Anscombe’a z http://businessq-software.com/2016/04/19/power-data-visualization-anscombes-story/



Kwartet Anscombe’a z http://businessq-software.com/2016/04/19/power-data-visualization-anscombes-story/

Nasze oczy są doskonałym narzędziem do znajdowania wzorców.

```
> install.packages('datasauRus')
> library(datasauRus)
```

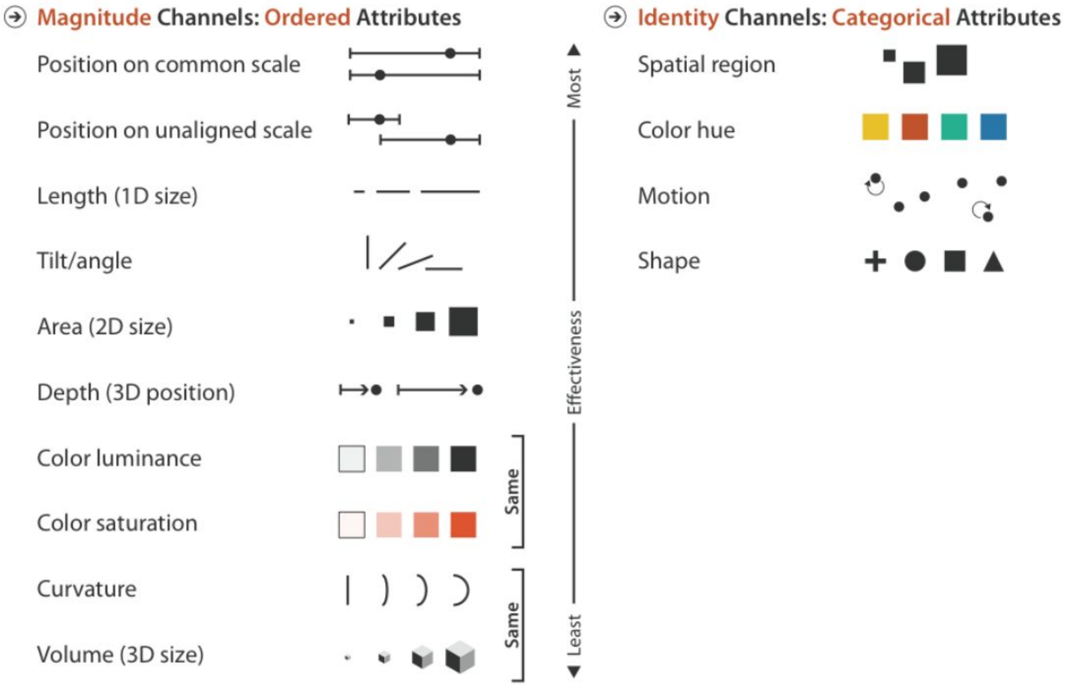


Doing Data Science: Straight Talk from the Frontline,
by Cathy O'Neil and Rachel Schutt

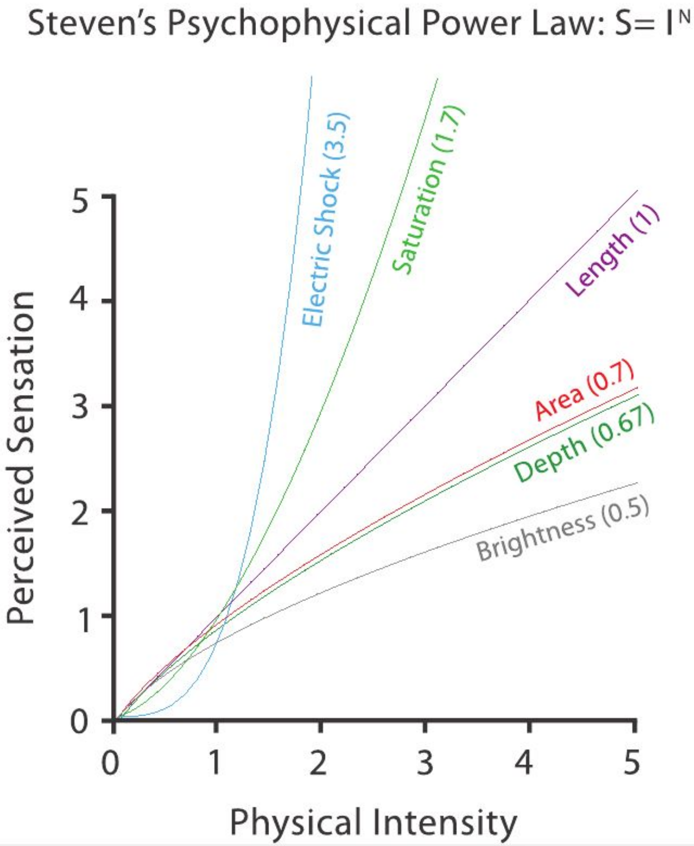
https://www.youtube.com/watch?v=jqbAwVvNse0&list=PL2u38g_AG4MEmRQIKov9Q4iReyiUICz6t

The Vis Vocabulary

How can we effectively encode data using graphical marks?



https://www.youtube.com/watch?v=jqbAwVvNse0&list=PL2u38g_AG4MEmRQIKov9Q4iReyiUICz6t



S = sensation
I = intensity

https://www.youtube.com/watch?v=jqbAwVvNse0&list=PL2u38g_AG4MEmRQIKov9Q4iReyiUICz6t

Wizualizacja

- Dzięki wykresom można szybko zaobserwować fakty, które trudno wywnioskować z samych statystyk opisowych lub surowych danych.
- Te same dane można przedstawić na różne sposoby: różne metody w różnym stopniu ułatwiają zrozumienie danych.
- W większości wizualizacji należy się trzymać podstawowych zasad przedstawiania danych (za chwilę).
- Skutki ignorowania tych zasad: http://smarterpoland.pl/index.php/2018/12/najgorszy-wykres-2018/

Gramatyka wizualizacji

A Layered Grammar of Graphics

Hadley WICKHAM

A grammar of graphics is a tool that enables us to concisely describe the components of a graphic. Such a grammar allows us to move beyond named graphics (e.g., the "scatterplot") and gain insight into the deep structure that underlies statistical graphics. This article builds on Wilkinson, Anand, and Grossman (2005), describing extensions and refinements developed while building an open source implementation of the grammar of graphics for R, `ggplot2`.

The topics in this article include an introduction to the grammar by working through the process of creating a plot, and discussing the components that we need. The grammar is then presented formally and compared to Wilkinson's grammar, highlighting the hierarchy of defaults, and the implications of embedding a graphical grammar into a programming language. The power of the grammar is illustrated with a selection of examples that explore different components and their interactions, in more detail. The article concludes by discussing some perceptual issues, and thinking about how we can build on the grammar to learn how to create graphical "poems."

Supplemental materials are available online.

Key Words: Grammar of graphics; Statistical graphics.

A	B	C	D
2	3	4	a
1	2	1	a
4	5	15	b
9	10	80	b

x	y	Shape
2	4	a
1	1	a
4	15	b
9	80	b

$$\text{floor}\left(\frac{x - \min(x)}{\text{range}(x)} * \text{screen width}\right)$$

Kroki budowy prostego wykresu

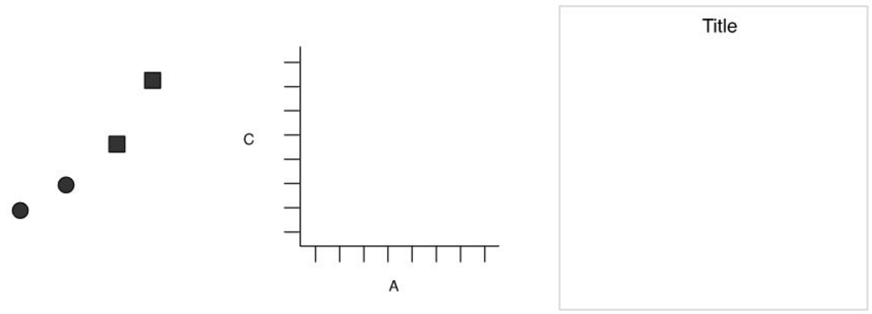


Figure 1. Graphics objects produced by (from left to right): geometric objects, scales and coordinate system, plot annotations.

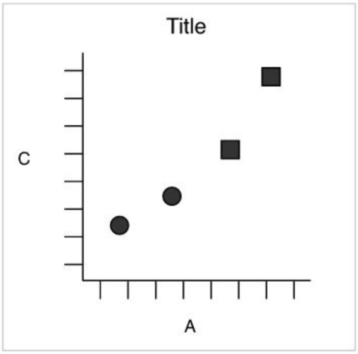


Figure 2. The final graphic, produced by combining the pieces in Figure 1.

Together, the data, mappings, statistical transformation, and geometric object form a layer. A plot may have multiple layers, for example, when we overlay a scatterplot with a smoothed line.

To be precise, the layered grammar defines the components of a plot as:

- a default dataset and set of mappings from variables to aesthetics,
- one or more layers, with each layer having one geometric object, one statistical transformation, one position adjustment, and optionally, one dataset and set of aesthetic mappings,
- one scale for each aesthetic mapping used,
- a coordinate system,
- the facet specification.

3.1.3 Geometric Object

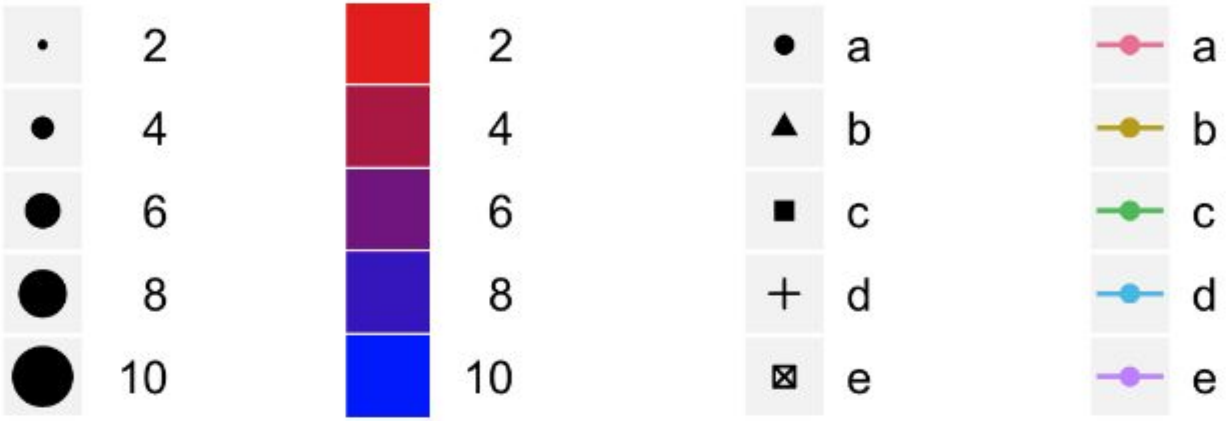
Geometric objects, or **geom**s for short, control the type of plot that you create. For example, using a point geom will create a scatterplot, whereas using a line geom will create a line plot. We can classify geoms by their dimensionality:

- 0d: point, text,
- 1d: path, line (ordered path),
- 2d: polygon, interval.

3.2 Scales

A **scale** controls the mapping from data to aesthetic attributes, and so we need one scale for each aesthetic property used in a layer. Scales are common across layers to ensure a consistent mapping from data to aesthetics. The legends associated with some scales are illustrated in Figure 7.

A scale is a function, and its inverse, along with a set of parameters. For example, the color gradient scale maps a segment of the real line to a path through a color space. The



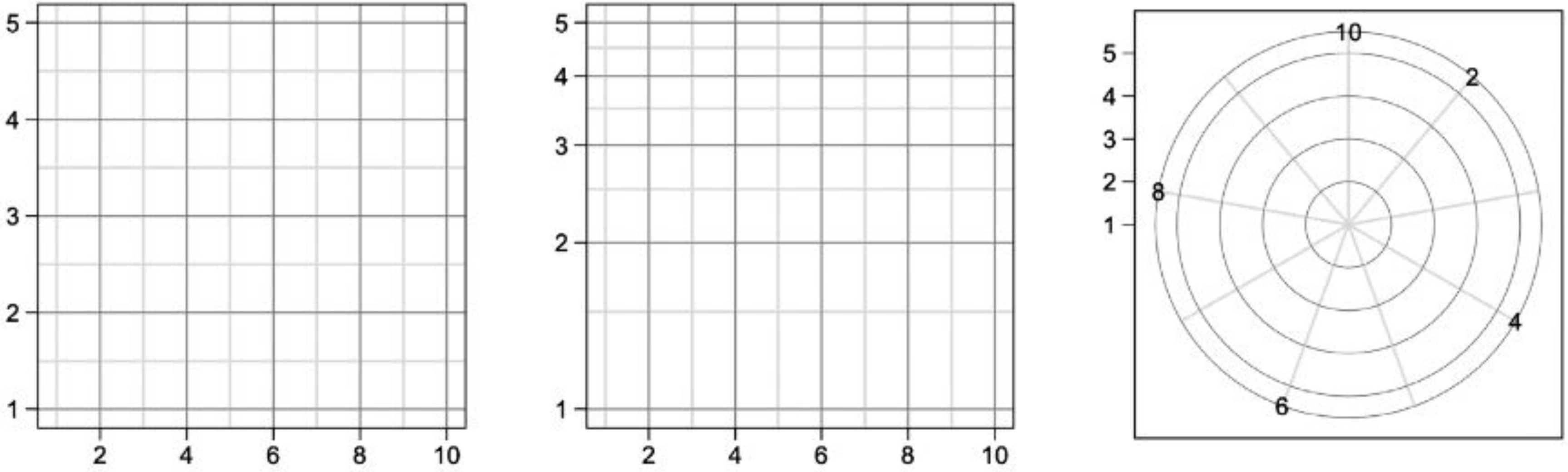


Figure 8. Examples of axes and grid lines for three coordinate systems: Cartesian, semi-log, and polar. The polar coordinate system illustrates the difficulties associated with non-Cartesian coordinates: it is hard to draw the axes well.

R

Podstawowe wizualizacje

Jednowymiarowe wizualizacje:

- histogram
- boxplot
- wykres skrzypcowy
- wykres słupkowy

Dwuwymiarowe wizualizacje:

- wykres punktowy / liniowy
- boxplot (wg kategorii)
- wykres mozaikowy
- *facets*

Warto:
https://journal.r-project.org/archive/2019/RJ-2019-033/RJ-2019-033.pdf