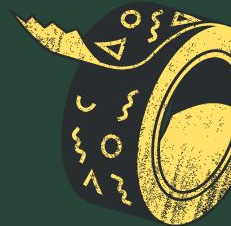




# Wstęp do Eksploracji Danych

Politechnika Warszawska

Anna Kozak



# Gramatyka języka wizualizacji

**Dlaczego projektujemy wykresy?**

# Dlaczego projektujemy wykresy?

“aby pokazać historie ukryte w danych”

# Trzy sposoby przedstawienia historii

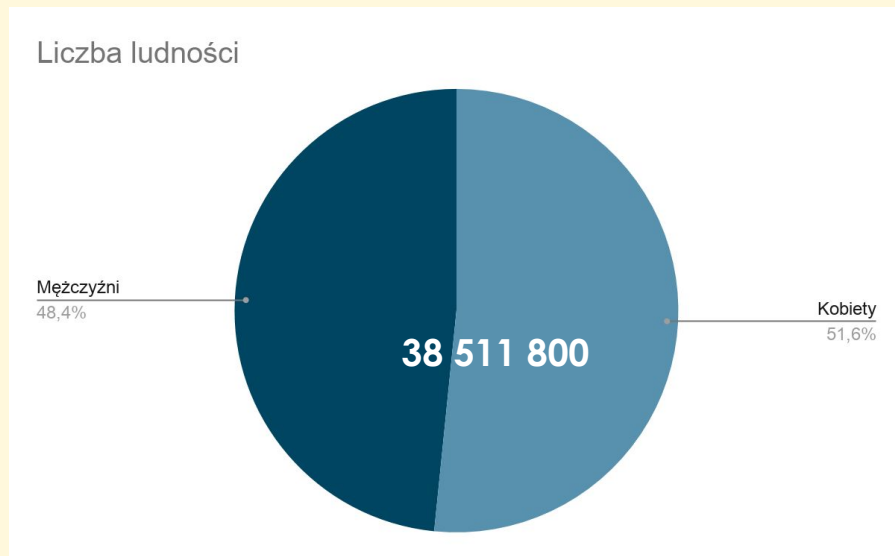
## 1) Opis słowny

*"W wyniku przeprowadzenia Narodowego Spisu Powszechnego w roku 2011 ustalono, że w Polsce mieszka 38 511 800 osób, z czego 48,4% to mężczyźni, a 51,6% to kobiety."*

## 2) Tabela

Liczba ludności Polski	W tym kobiet	W tym mężczyzn
38 511 800	51,6%	48,4%

## 3) Wykres



**Co w przypadku dużego zbioru danych?**

# Co w przypadku dużego zbioru danych?

Historia wyników polskich matur z lat 2010-2015.

## 1) Opis słowny

*“Wyniki matury z języka polskiego mają rozkład zbliżony do normalnego. W poszczególnych latach średnie tego rozkładu nieznacznie się różnią. Rozkład ten jest zaburzony w okolicy 21-22 punktów, czyli w pobliżu wartości stanowiących granicę zaliczenia (30% możliwych do uzyskania punktów). Praktycznie nie ma uczniów, którzy uzyskaliby jeden punkt poniżej progu zaliczenia, jest za to bardzo dużo osób, które zdały egzamin, otrzymując punkt więcej. Sugeruje to, że dosyć często osoby oceniające maturę, widząc, że do zaliczenia brakuje jednego–dwóch punktów, brakujące punkty “znajdowały”. W przypadku egzaminu z matematyki rozkłady są różne w różnych rocznikach i zdecydowanie nie przypominają rozkładu normalnego. W pobliżu progu zaliczenia również widzimy pewną nieregularność, największą w roku 2014. Jest ona jednak mniejsza niż w przypadku egzaminu z języka polskiego.”*

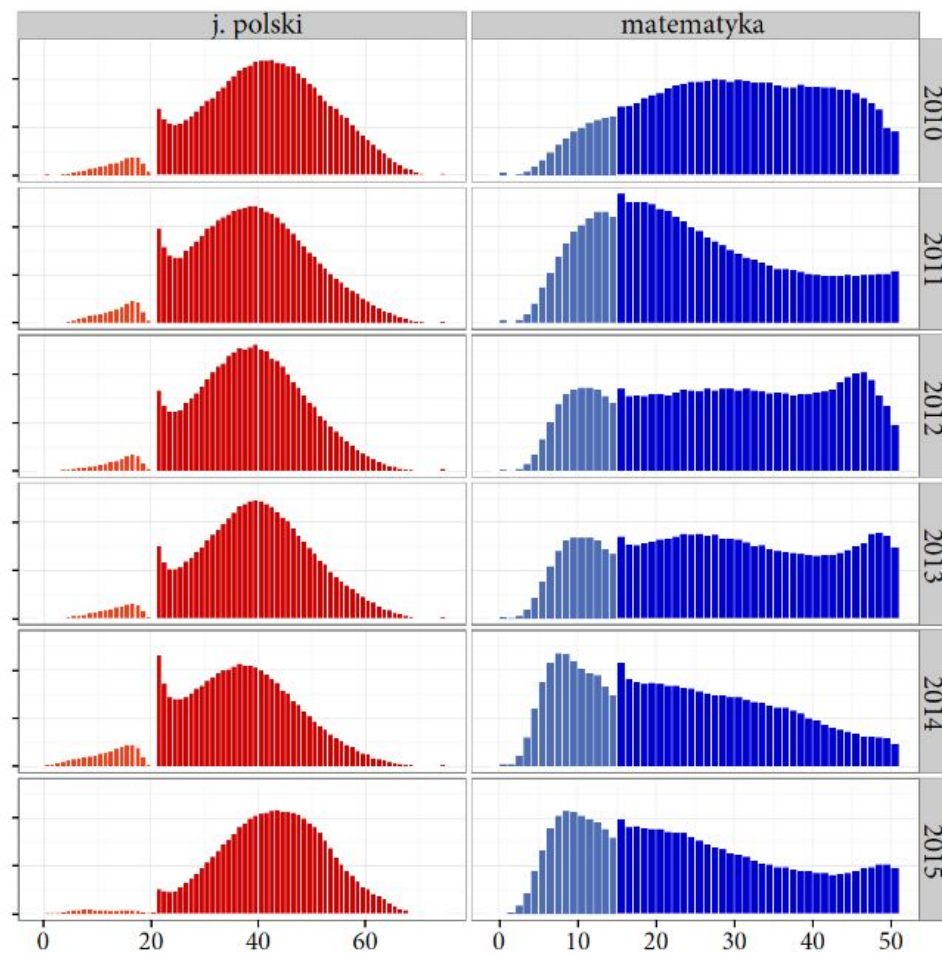
## 2) Tabela

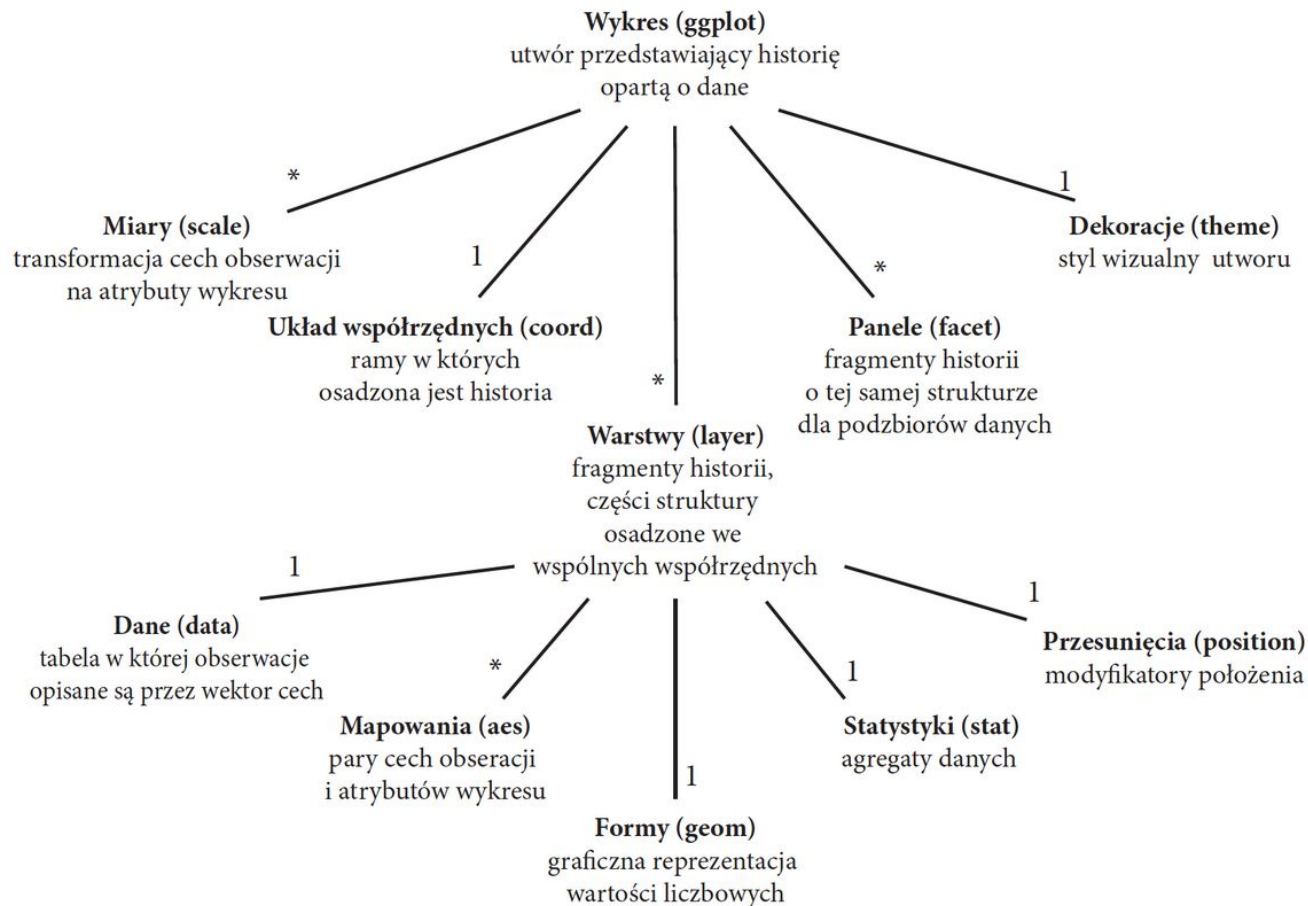
punkty	przedmiot	2010	2011	2012	2013	2014	2015
...	...	...	...	...	...	...	...
6	j. polski	0,09	0,09	0,09	0,09	0,25	0,16
7	j. polski	0,12	0,14	0,11	0,12	0,28	0,16
8	j. polski	0,16	0,18	0,12	0,14	0,34	0,19
9	j. polski	0,19	0,22	0,14	0,19	0,36	0,19
10	j. polski	0,23	0,27	0,18	0,21	0,40	0,17
11	j. polski	0,25	0,29	0,20	0,25	0,45	0,16
12	j. polski	0,28	0,31	0,23	0,28	0,47	0,15
13	j. polski	0,34	0,36	0,27	0,31	0,50	0,13
14	j. polski	0,37	0,41	0,32	0,37	0,61	0,13
15	j. polski	0,42	0,47	0,37	0,41	0,68	0,16
16	j. polski	0,49	0,57	0,45	0,45	0,73	0,16
17	j. polski	0,54	0,67	0,50	0,50	0,74	0,17
18	j. polski	0,54	0,62	0,46	0,44	0,63	0,14
19	j. polski	0,34	0,34	0,26	0,27	0,31	0,10
20	j. polski	0,13	0,09	0,09	0,09	0,07	0,06
21	j. polski	0,02	0,01	0,01	0,01	0,01	0,10
22	j. polski	1,90	2,72	2,43	2,28	3,76	0,90
23	j. polski	1,60	2,20	1,96	1,78	2,80	0,82
24	j. polski	1,46	1,95	1,80	1,56	2,36	0,81
25	j. polski	1,44	1,91	1,80	1,59	2,28	0,85
...	...	...	...	...	...	...	...



### 3) Wykres

Rozkład liczby punktów na maturze, poziom podstawowy





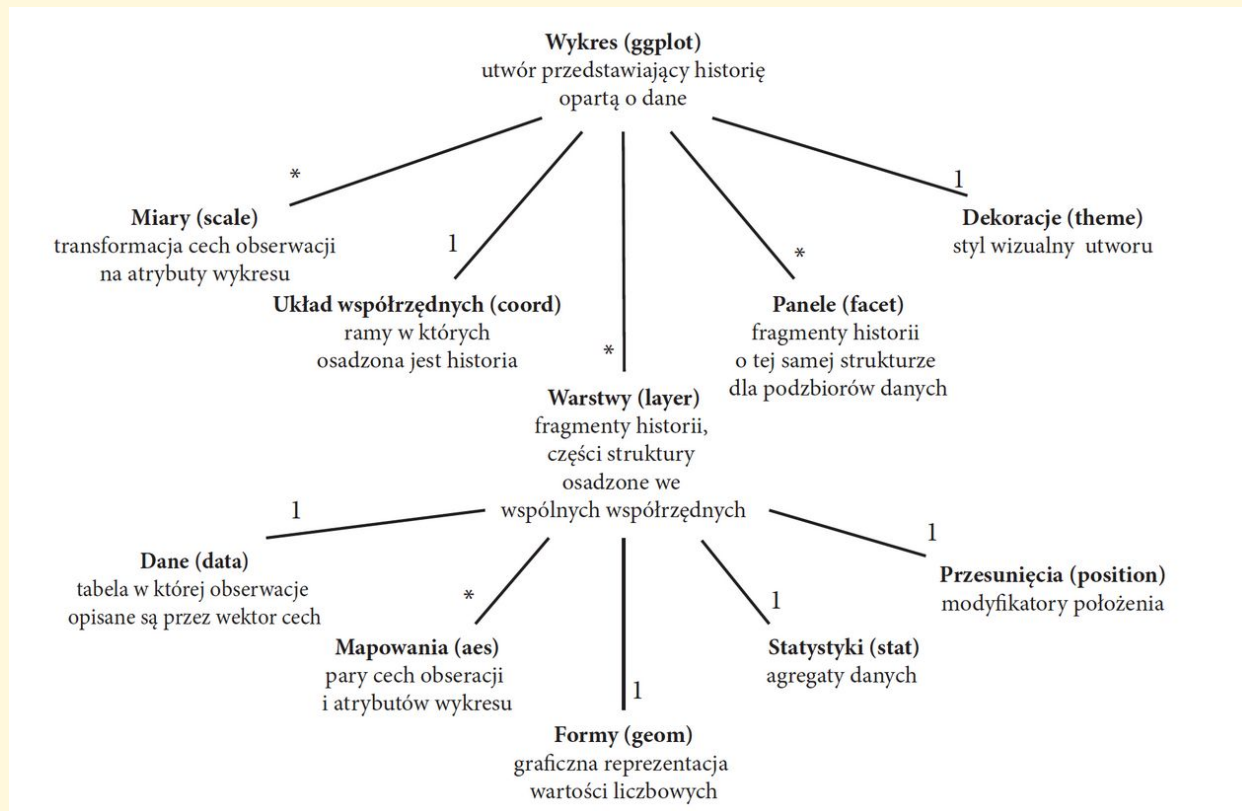
## Reprezentacja szeroka danych

przedmiot	rok_2010	rok_2011	rok_2012	rok_2013	rok_2014	rok_2015
j. polski	40.1	37.5	37.5	38.3	35.2	41.5
matematyka	29.2	24.1	27.9	27.3	22.3	

## Reprezentacja wąska danych

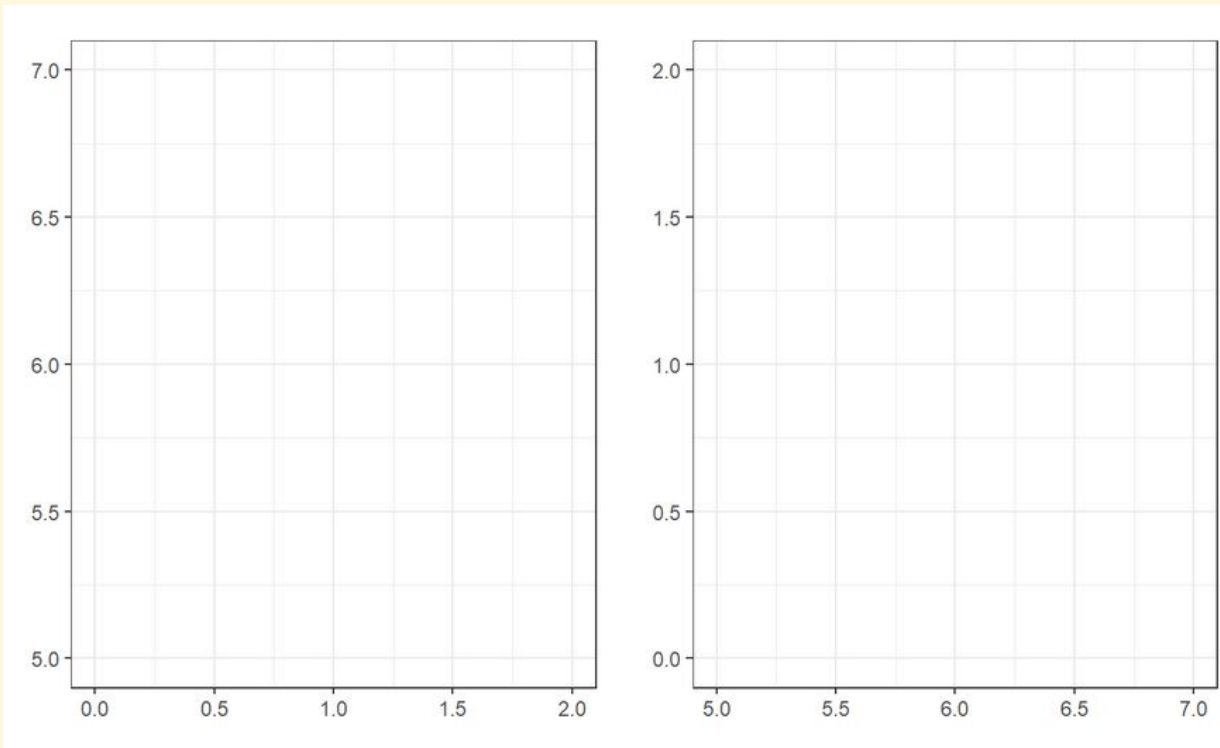
rok	przedmiot	srednia
rok_2010	j. polski	40.1
rok_2011	j. polski	37.5
rok_2012	j. polski	37.5
rok_2013	j. polski	38.3
rok_2014	j. polski	35.2
rok_2015	j. polski	41.5
rok_2010	matematyka	29.2
rok_2011	matematyka	24.1
rok_2012	matematyka	27.9
rok_2013	matematyka	27.3
rok_2014	matematyka	22.3

# Układ współrzędnych

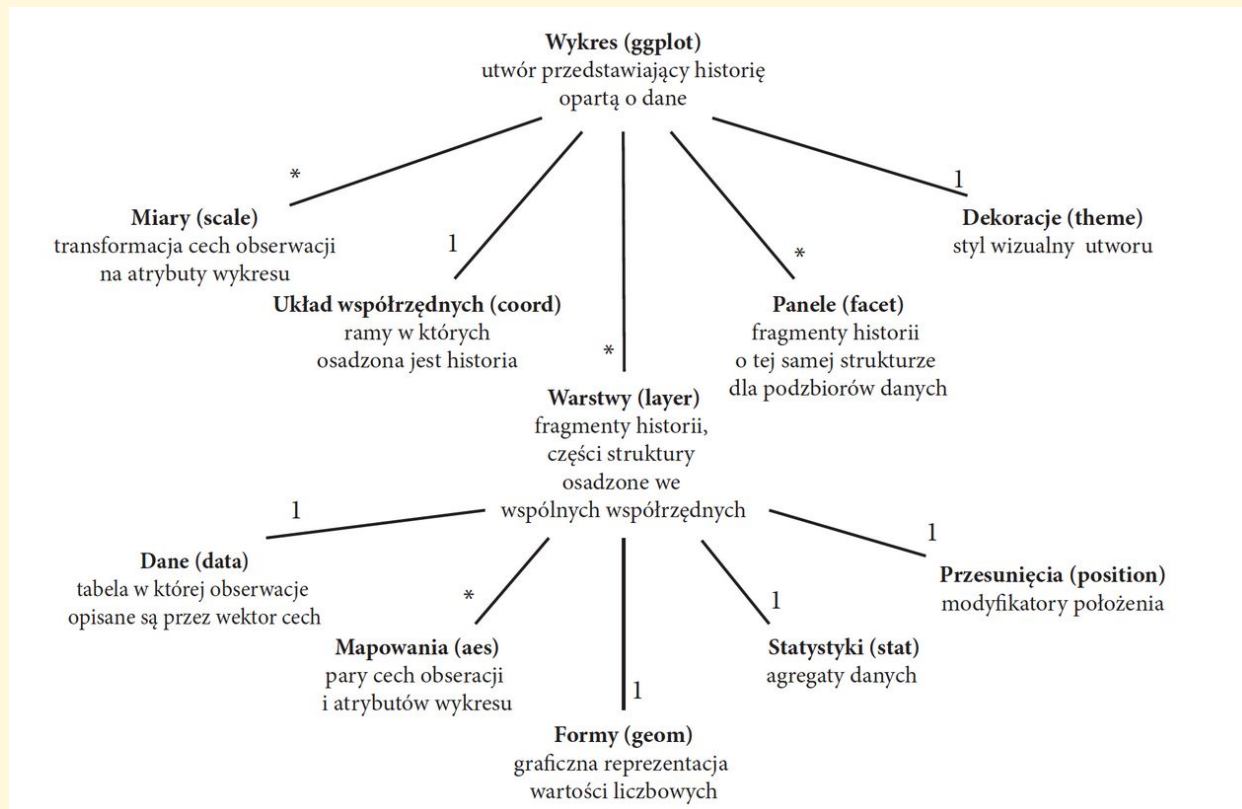


# Układ współrzędnych

Układ współrzędnych (coords): ramy, w których osadzona jest historia.



# Warstwy - Dane

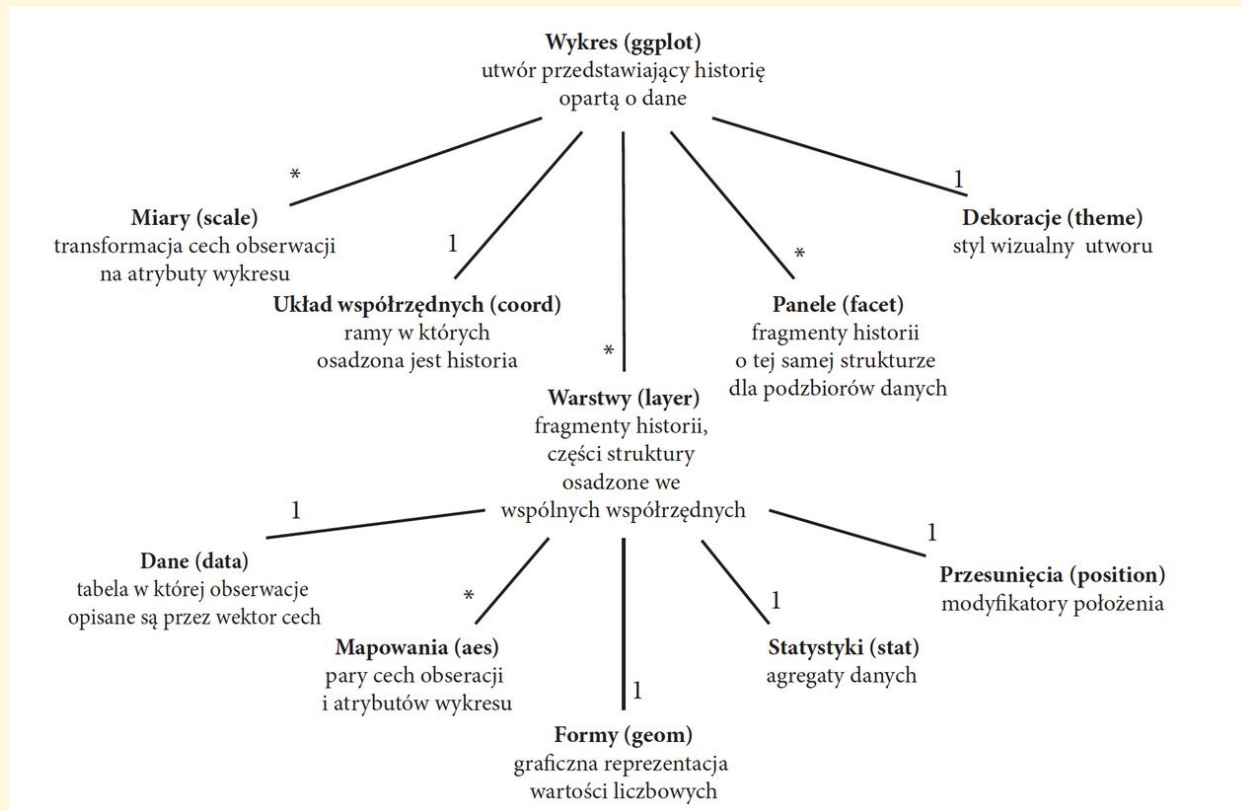


# Warstwy - Dane

Dane (data): tabela, w której obserwacje opisane są przez wektory cech.

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
## 1	5.1	3.5	1.4	0.2	setosa
## 2	4.9	3.0	1.4	0.2	setosa
## 3	4.7	3.2	1.3	0.2	setosa
## 4	4.6	3.1	1.5	0.2	setosa
## 5	5.0	3.6	1.4	0.2	setosa
## 6	5.4	3.9	1.7	0.4	setosa

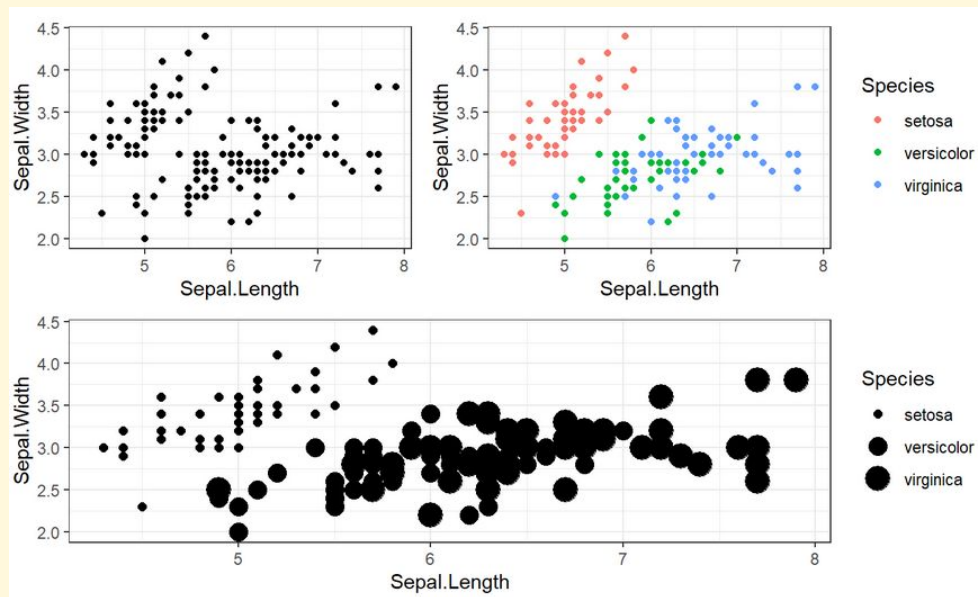
# Warstwy - Mapowania



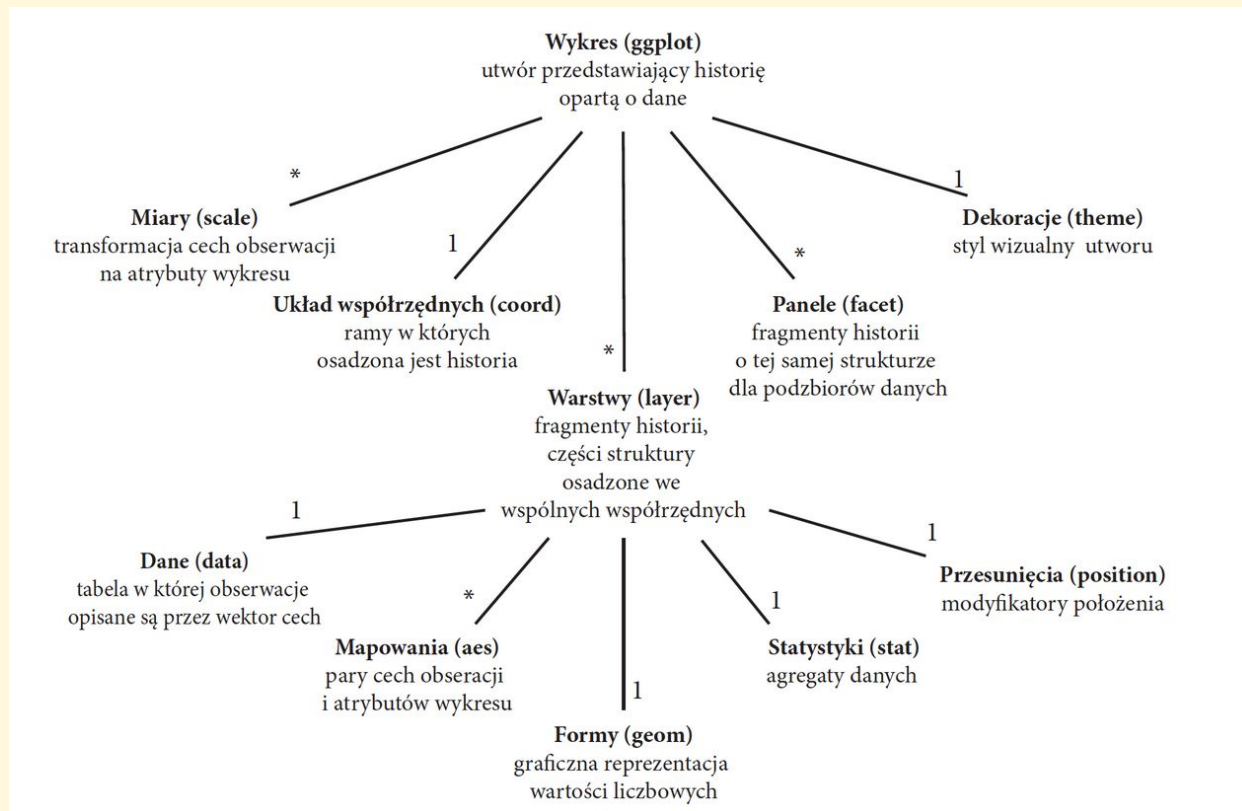


# Warstwy - Mapowania

Mapowania (aes): pary cech obserwacji i atrybutów wykresu.

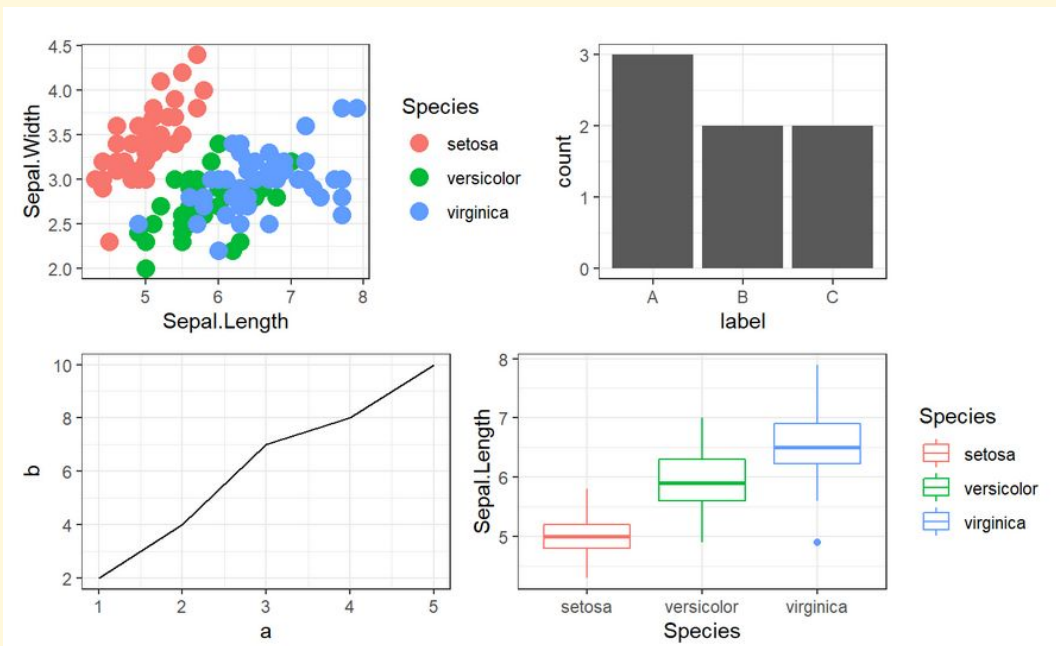


# Warstwy - Formy

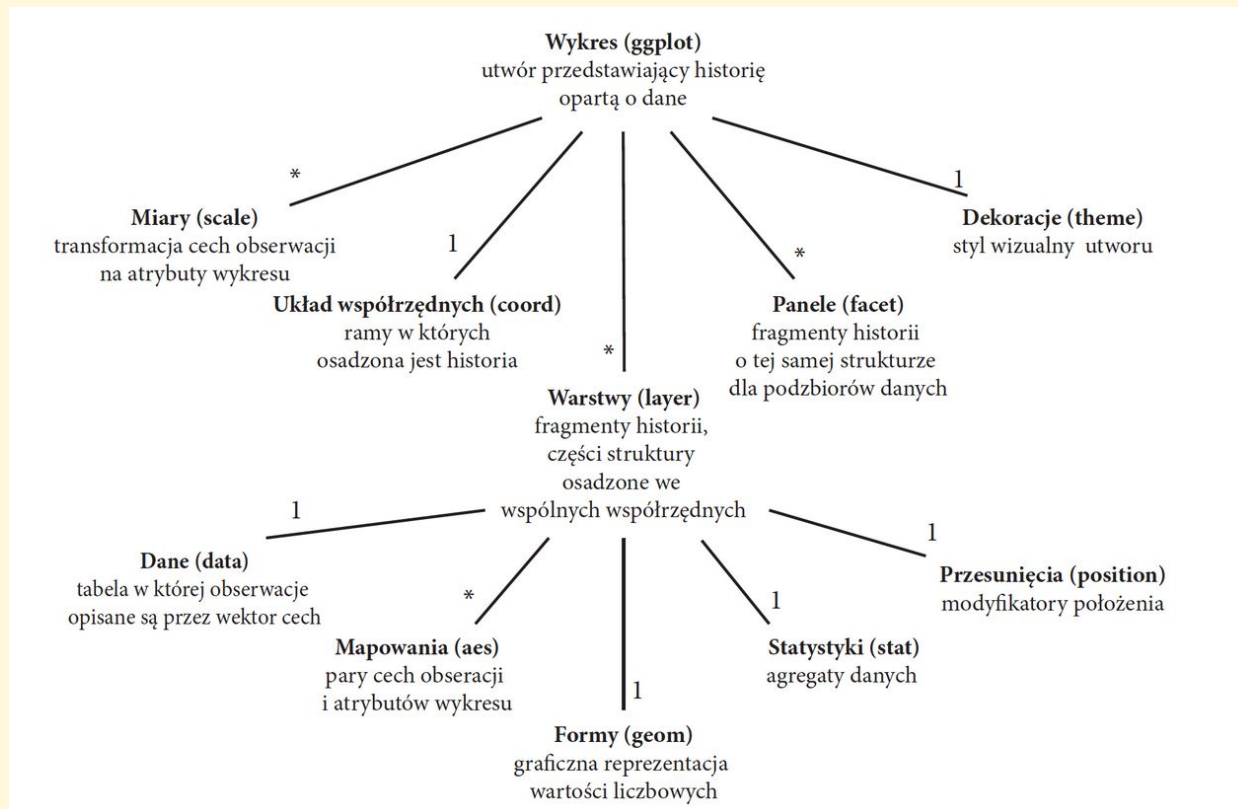


# Warstwy - Formy

Formy (geom): graficzna reprezentacja wartości liczbowych

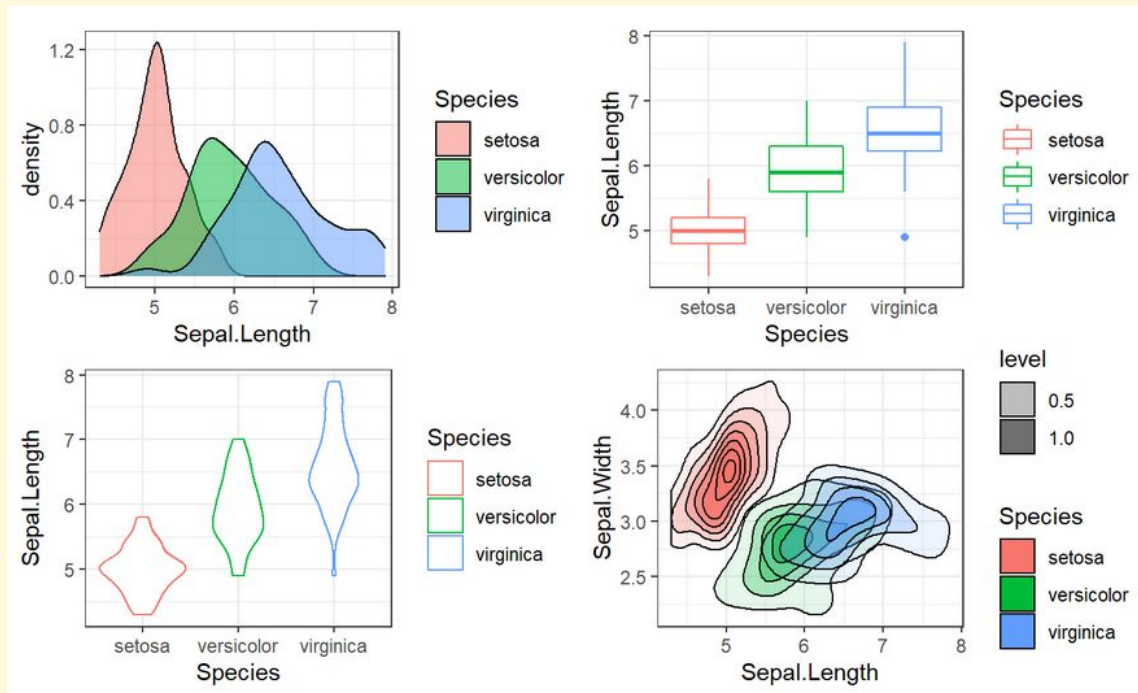


# Warstwy - Statystyki

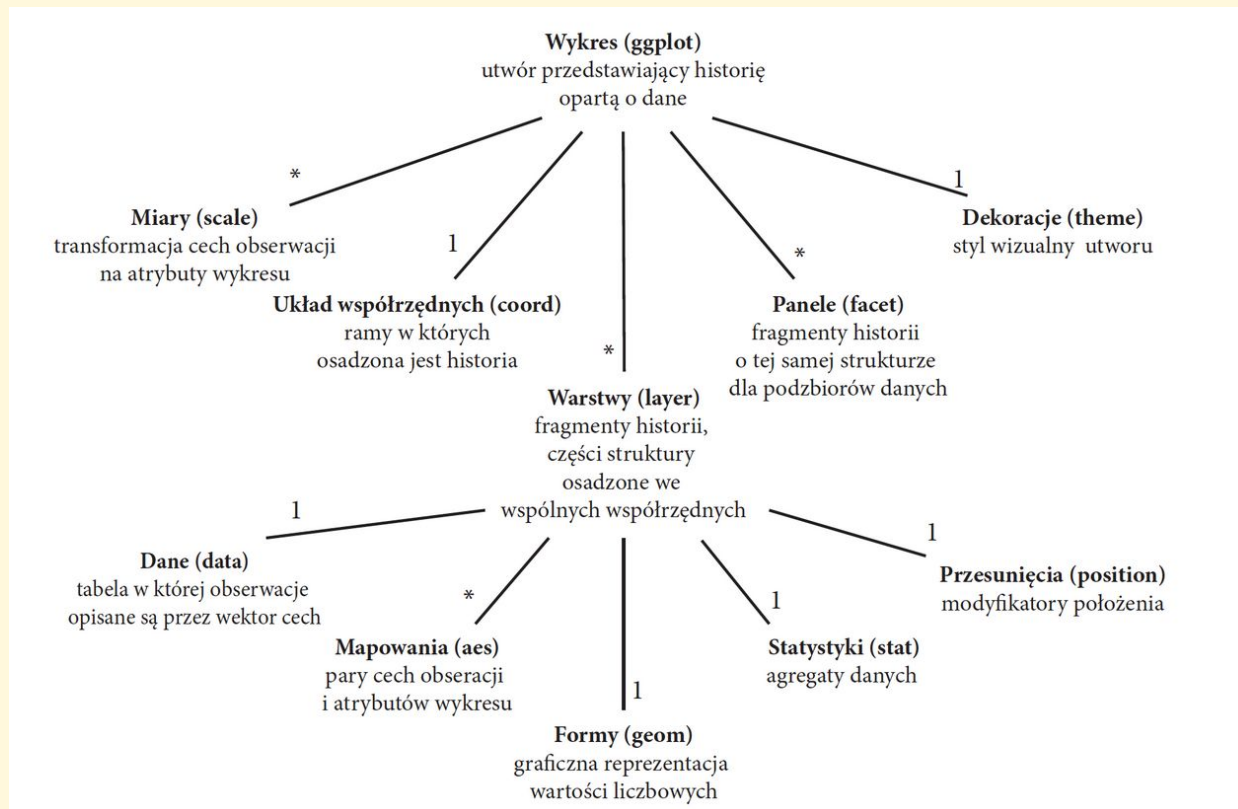


# Warstwy - Statystyki

Statystyki (stat): agregaty danych

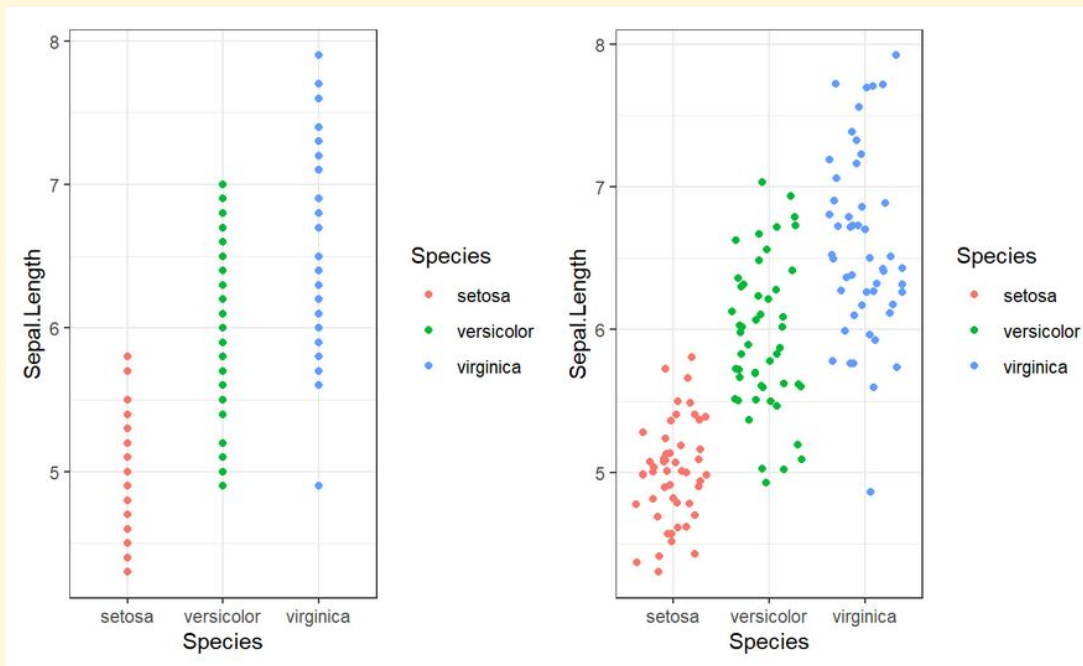


# Warstwy - Przesunięcia

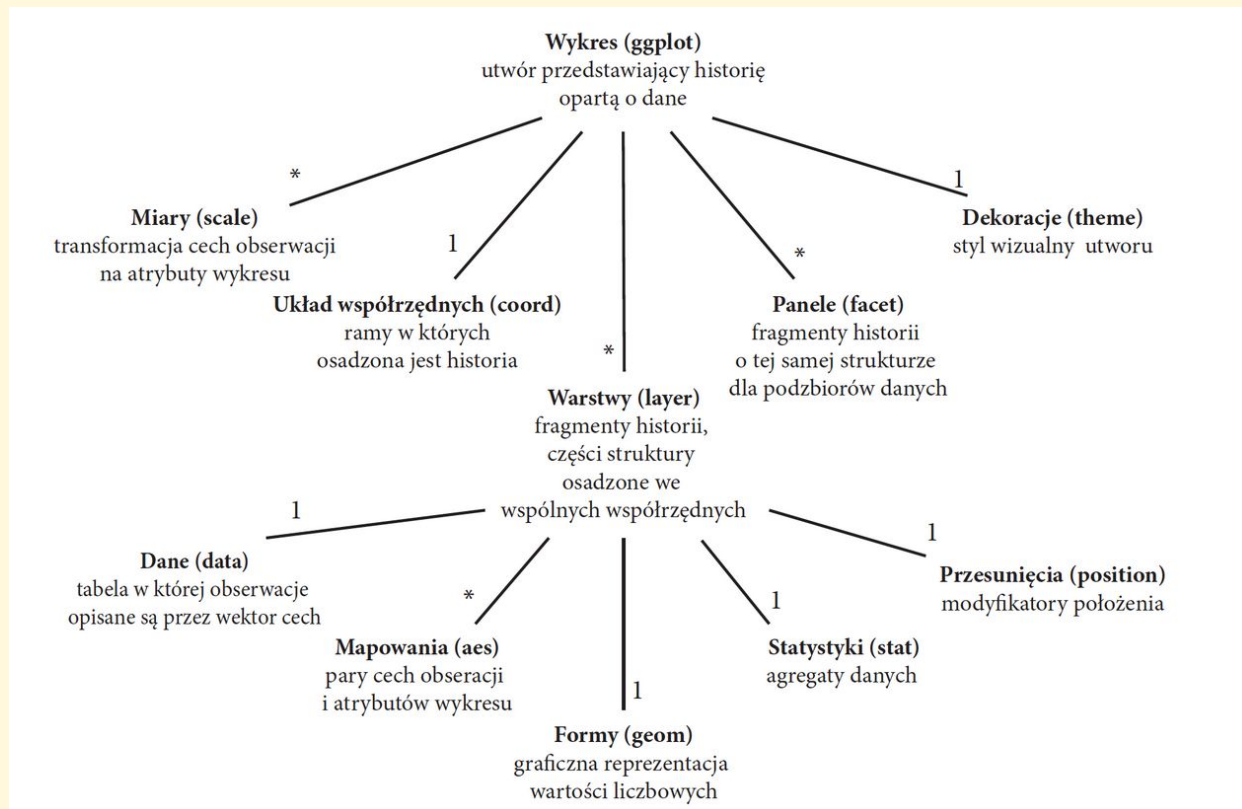


# Warstwy - Przesunięcia

Przesunięcia (position): modyfikatory położenia



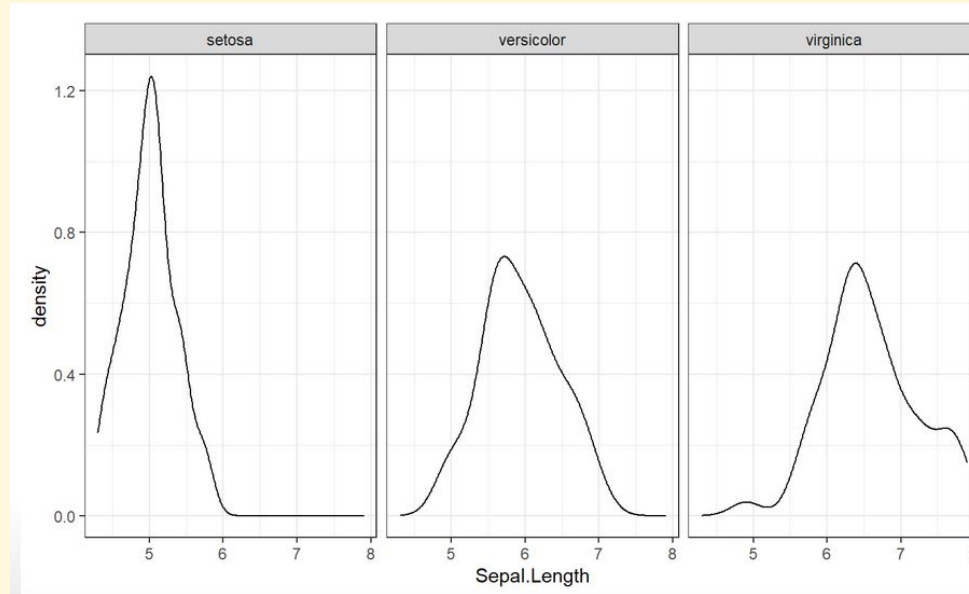
# Warstwy - Panele



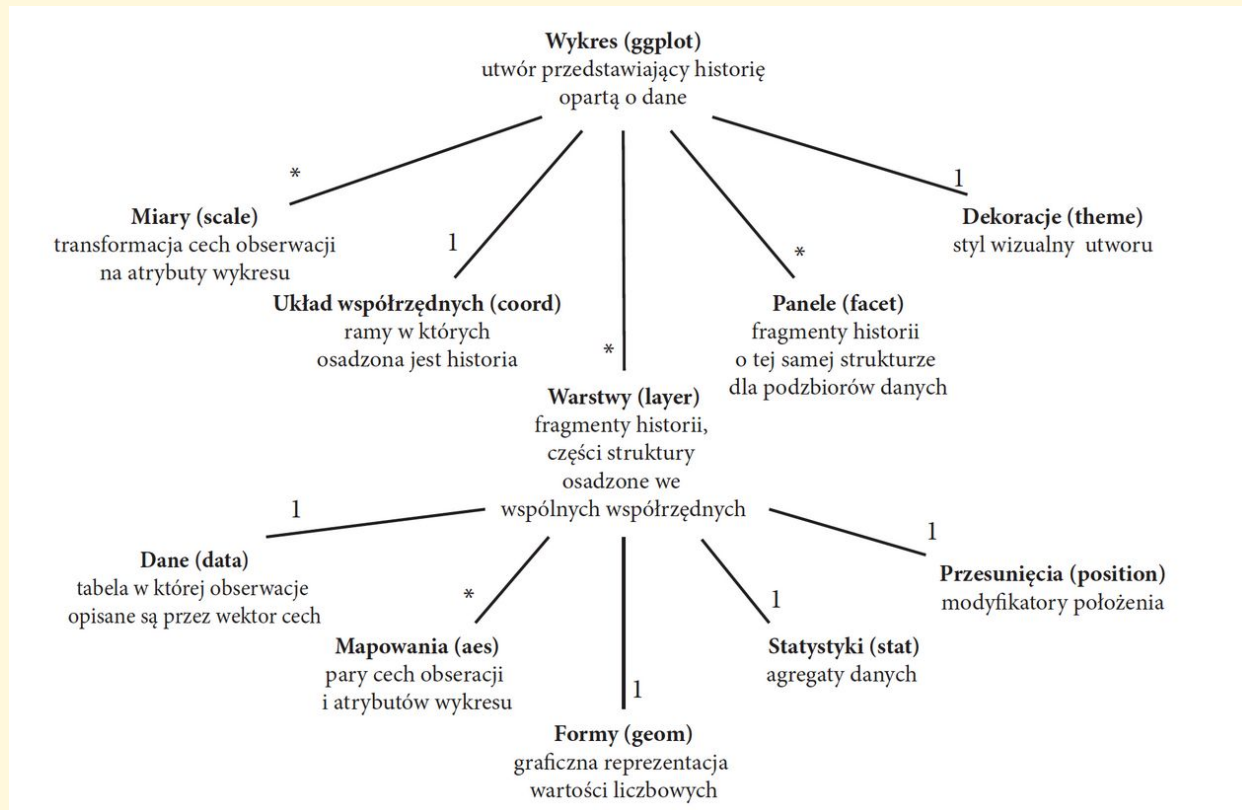


# Warstwy - Panele

Panele (facets): fragmenty historii o tej samej strukturze dla podzbiorów danych.

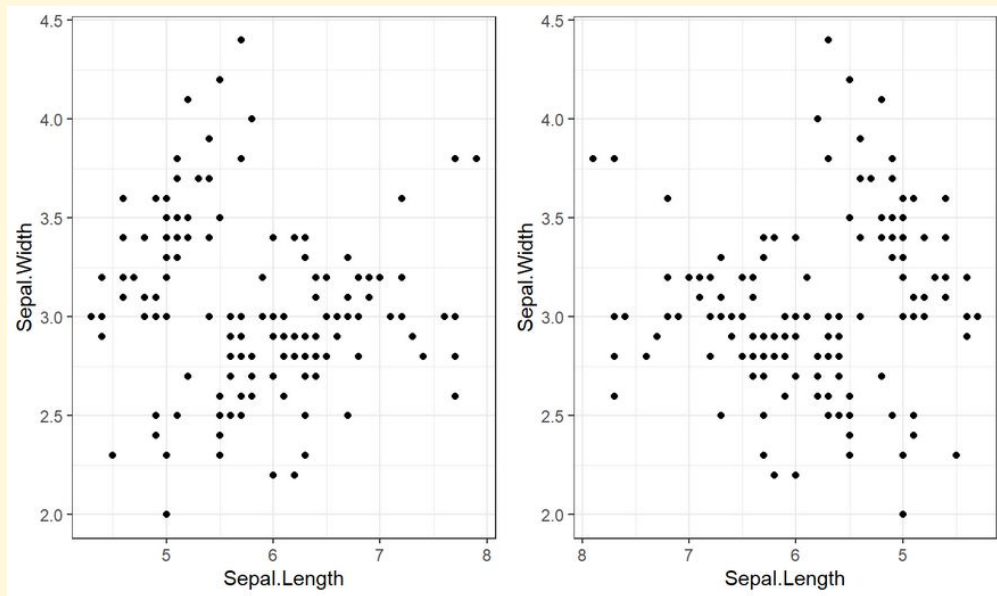


# Skale

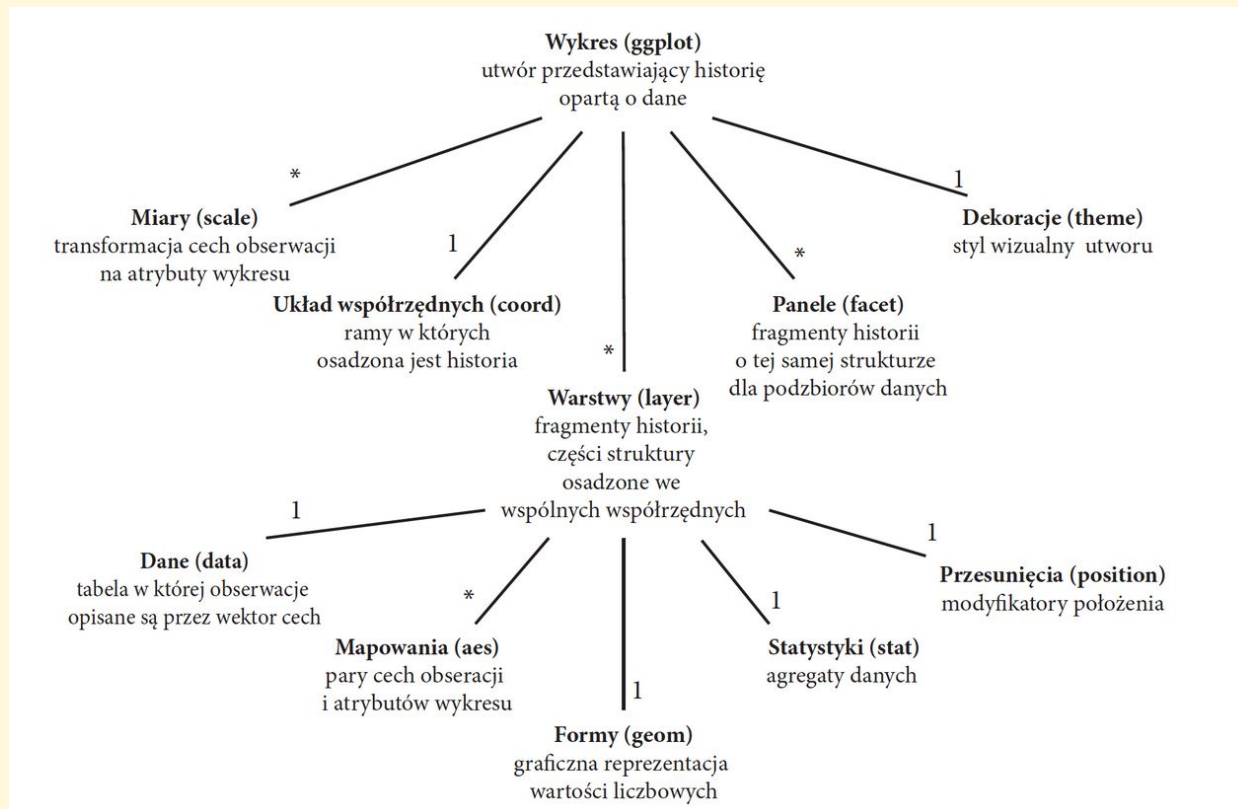


# Skale

Skale (scale): transformacja cech obserwacji na atrybuty wykresu.



# Dekoracje



# Dekoracje

Dekoracje (theme): styl wizualny.

