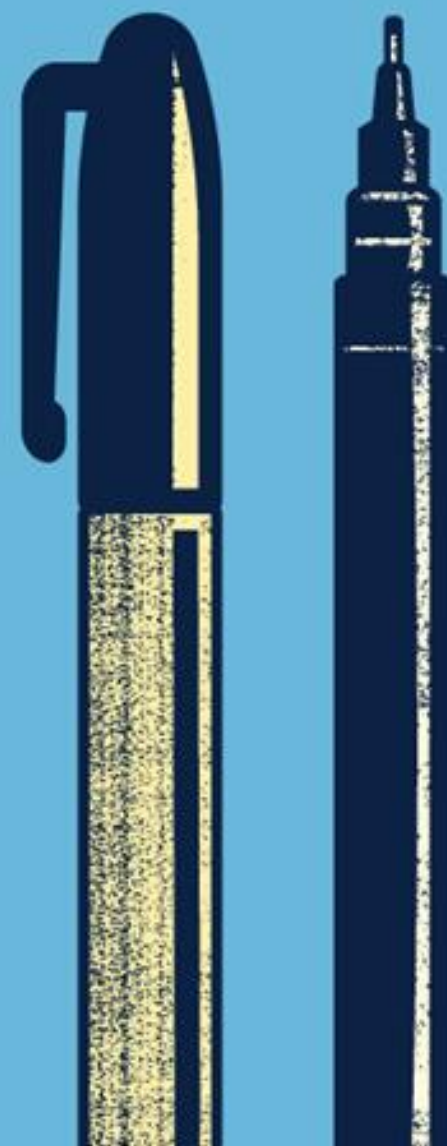




# Techniki Wizualizacji Danych

Politechnika Warszawska

Anna Kozak



# Anna Kozak



anna.kozak@pw.edu.pl

MS Teams

# Antoni Chudy

# Iza Danielewska

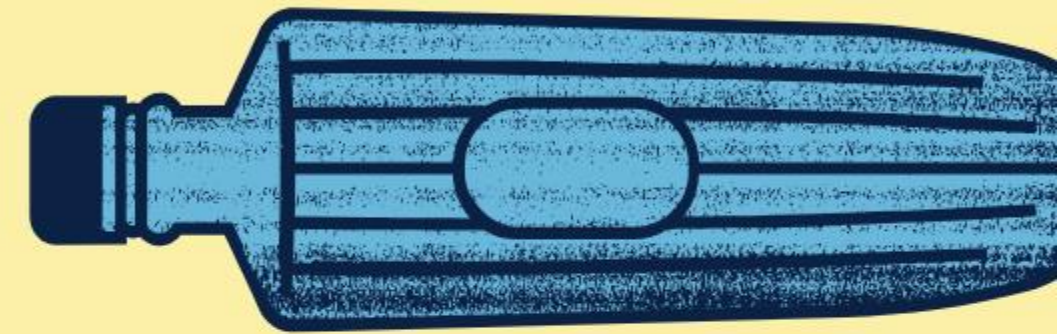
# Dawid Płodowski

# Katarzyna Woźnica



# Strona przedmiotu

<https://github.com/kozaka93/2025Z-DataVisualizationTechniques>



# Zajęcia

Wykład, s. 329 (środy)

Zajęcia projektowe, s. 329 (środy)

Laboratoria, s. 302, 303, 304 (środy, czwartki, piątki)

# Wykład

Na wykładzie będą przedstawione zarówno teoretyczne aspekty pracy z danymi, jak i praktyczne.

15 wykładów = 6 x 2h wykład + 3 x 1h wykład (w tym 2 x prezentacje projektów + 1 x test)

Test z wiedzy teoretycznej - nieobowiązkowy (10 pkt, Wykład 14).

# Zajęcia projektowe

- 2 projekty w ciągu semestru
- zespoły 3 osobowe, różne podczas 1 i 2 projektu
- projekt trwa 7-8 tygodni
- 25 pkt i 29 pkt za projekt (w tym 5 pkt za pracę na zajęciach projektowych)

# Laboratorium

- praca w R i Python
- powtórzenie operacji na danych (R: dplyr, tidyr; Python: pandas)
- wstęp do narzędzi pozwalających na estetyczne prezentowanie danych
- różne sposoby oceny zmiennych, danych, wizualizacji
- 6 x praca domowa (6 x 6 pkt)

Zajęcia laboratoryjne są obowiązkowe, w ciągu semestru dopuszczalne są co najwyżej dwie nieusprawiedliwione nieobecności.



# Ocena końcowa

Suma punktów z prac domowych, testu i projektów:

$$25 (P1) + 29 (P2) + 10 (T) + 6 \times 6 (PD) = 100$$

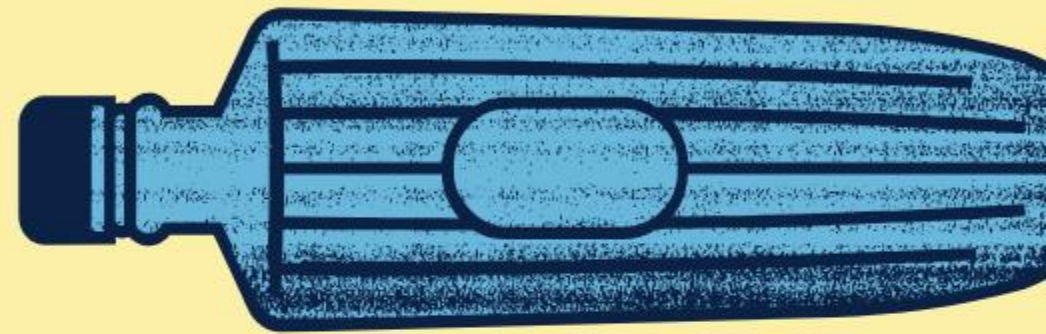
Aby zaliczyć kurs należy uzyskać ponad 50 punktów,  
w tym co najmniej 50% punktów z każdego z projektów.

Oceny będą wystawiane zgodnie z tabelą:

Ocena	3	3.5	4	4.5	5
Punkty	(50, 60]	(60, 70]	(70, 80]	(80, 90]	(90, $\infty$ )



# Pytania?



**Zanim wizualizacja to chwilę  
o eksploracji danych**

# Dane

Mogą być generowane przez:

- ?

# Dane

Mogą być generowane przez:

- banki,
- ubezpieczenia,
- portale społecznościowe,
- firmy telekomunikacyjne,
- szpitale,
- dane eksperymentalne,
- tekst,
- mapy,
- sklepy internetowe,
- ...



# Eksploracja danych - czym jest?

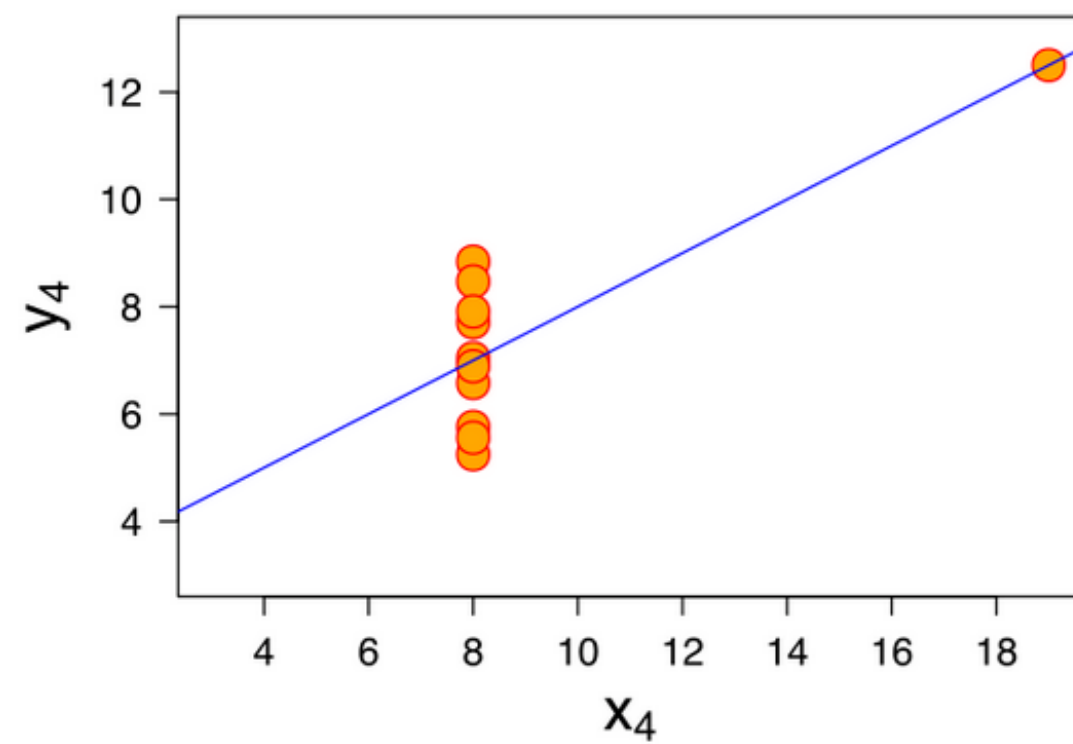
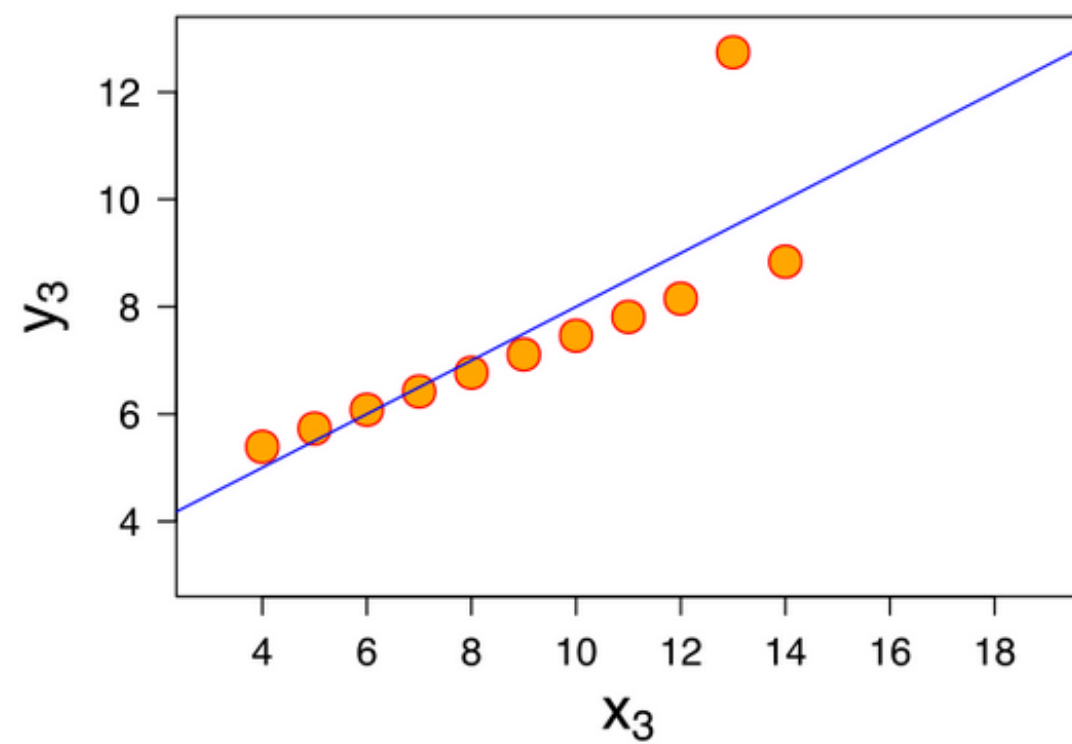
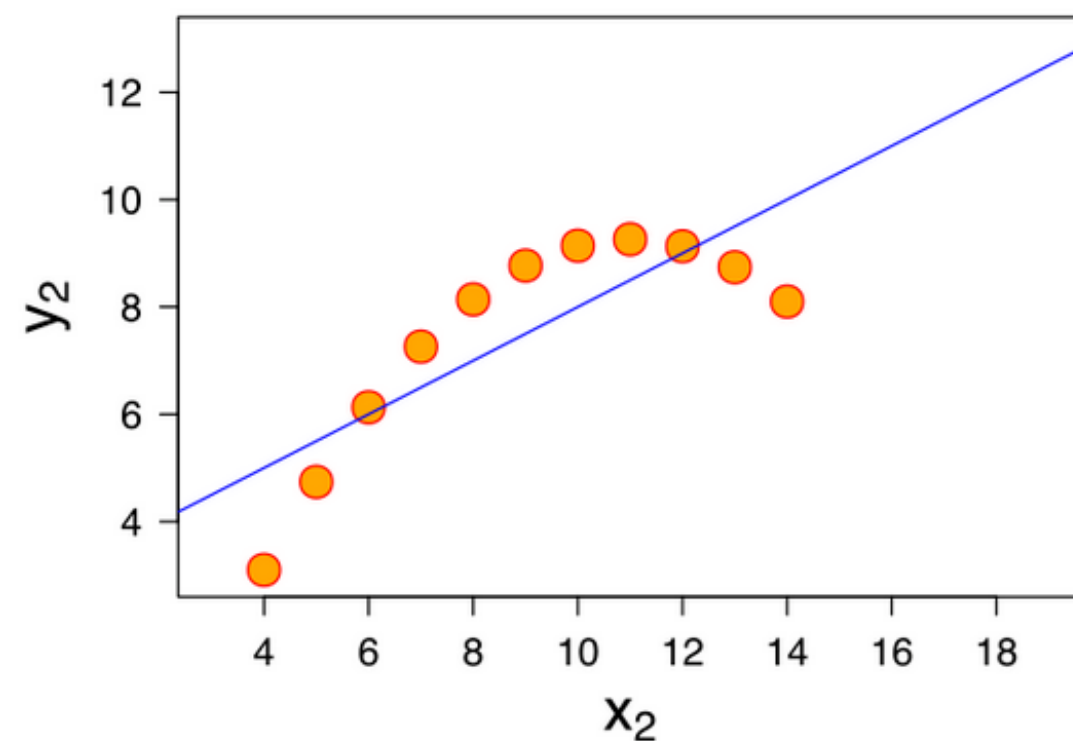
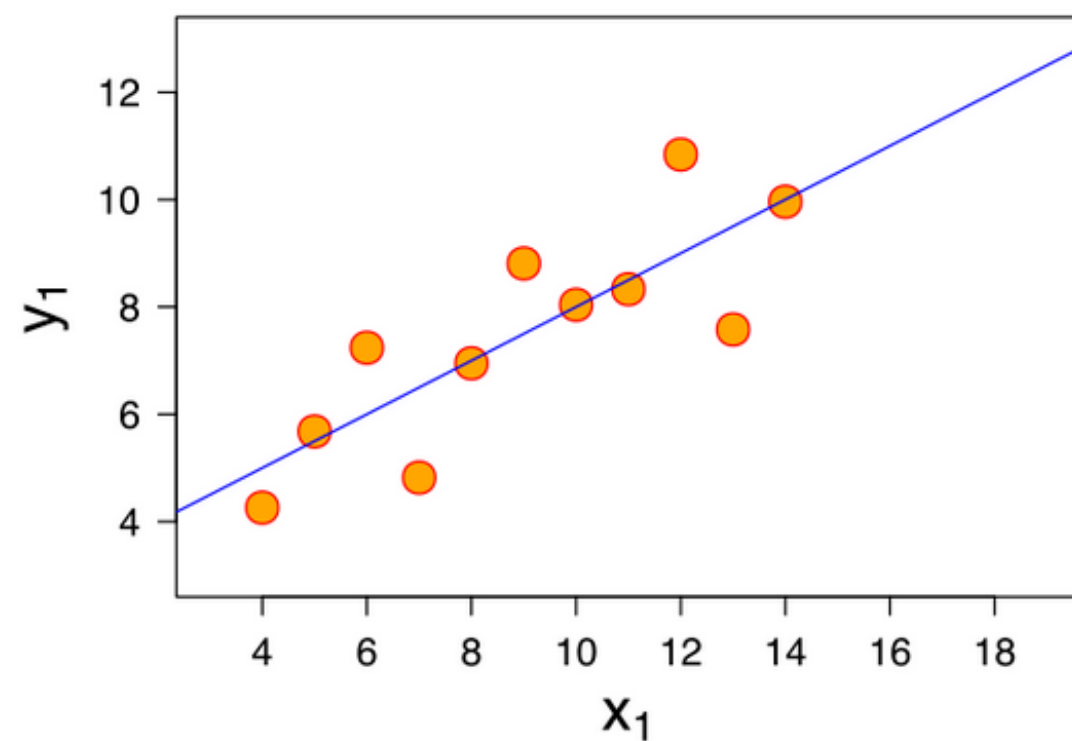
“proces odkrywania nietrywialnych, dotychczas nieznanych, potencjalnie użytecznych reguł, zależności, trendów”

**Cel:** analiza danych w celu lepszego ich zrozumienia

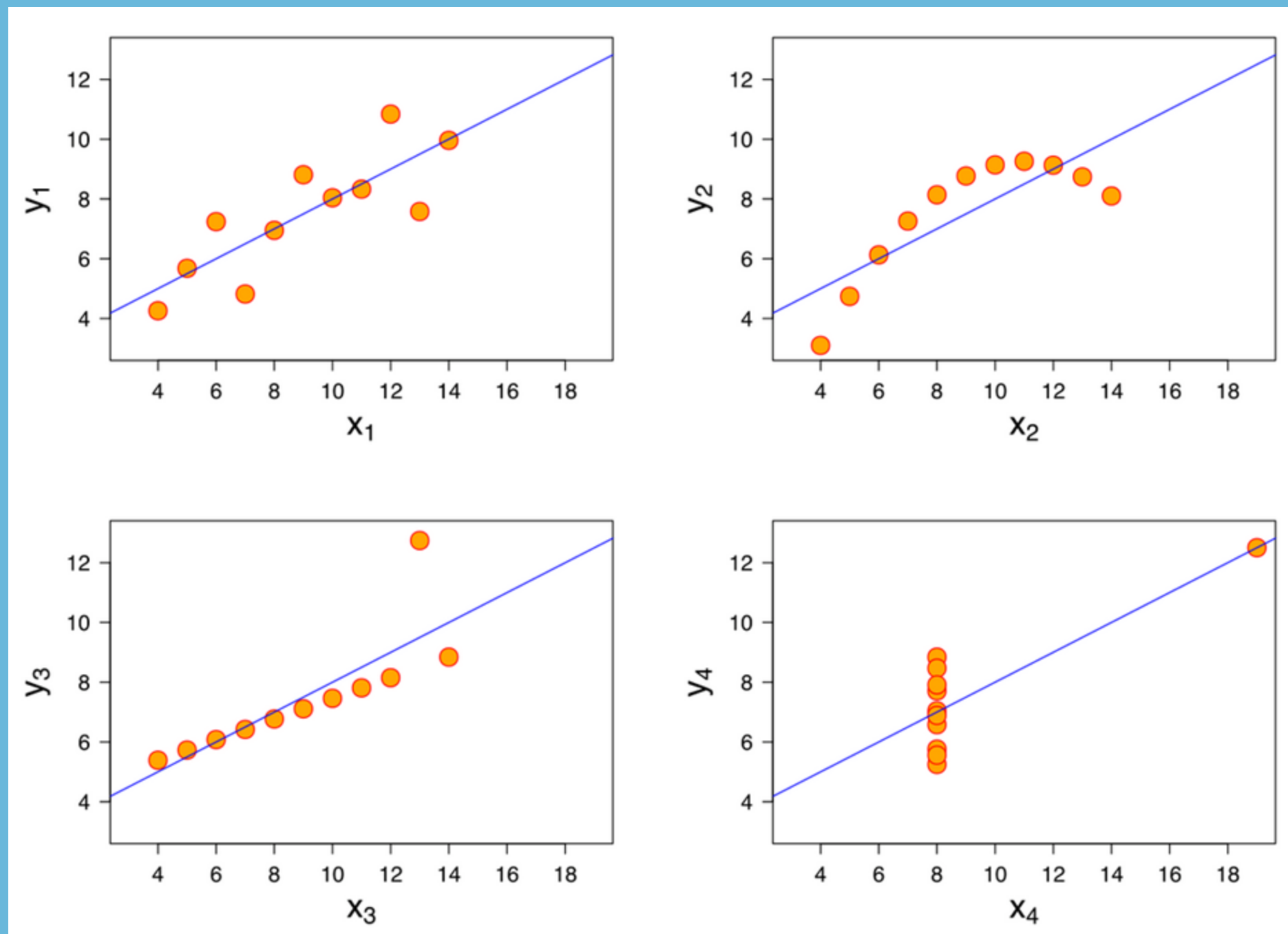
# Eksploracja danych - czym jest?

Na eksplorację danych składa się wiele dyscyplin, między innymi:

- bazy danych
- statystyka
- uczenie maszynowe
- techniki wizualizacji danych
- wyszukiwanie informacji



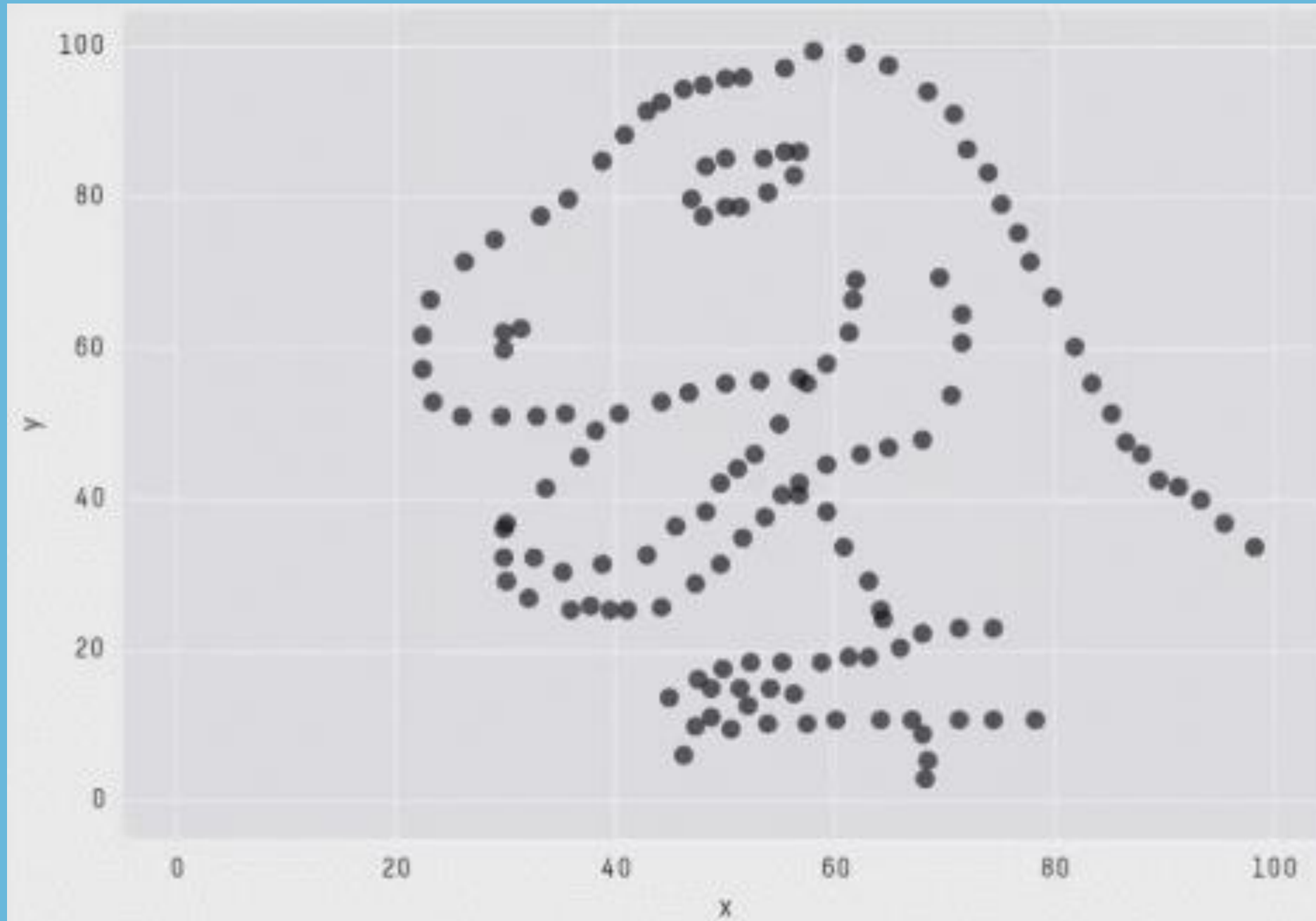
# Kwartet Anscombe' a



Cecha	Wartość
Średnia arytmetyczna zmiennej x	9
Wariancja zmiennej x	11
Średnia arytmetyczna zmiennej y	7.50 (identyczna do dwóch cyfr po przecinku)
Wariancja zmiennej y	4.122 lub 4.127 (identyczna do trzech cyfr po przecinku)
Współczynnik korelacji pomiędzy zmiennymi	0.816 (identyczny do trzech cyfr po przecinku)



# The Datasaurus Dozen



X Mean: 54.2659224  
Y Mean: 47.8313999  
X SD : 16.7649829  
Y SD : 26.9342120  
Corr. : -0.0642526

# Jak rozpoznać rodzaj zmiennej?

*“dane liczbowe to nie tylko liczby”*

# Typy danych

Zmienne jakościowe (nazywane również wyliczeniowymi, czynnikowymi lub kategorycznymi), to zmienne przyjmujące określoną liczbę wartości (najczęściej nie liczbowych). Zmienne te można dalej podzielić na:

- *binarne* (nazywane również dwumianowymi, dychotomicznymi) np. płeć (poziomy: kobieta/mężczyzna),
- *nominalne* (nazywane również zmiennymi jakościowymi nieuporządkowanymi) np. marka samochodu,
- *uporządkowane*, np. wykształcenie (poziomy: podstawowe/średnie/wyższe), ocena z przedmiotu.

# Typy danych

Zmienne ilościowe, z których można dodatkowo wyróżnić:

- *zliczenia* (liczba wystąpień pewnego zjawiska, opisywana liczbą całkowitą), np. liczba lat nauki, liczba wypadków,
- *ilorazowe*, czyli zmienne mierzone w skali, w której można dzielić wartości (ilorazy mają sens). Np. długość w metrach (coś jest 2 razy dłuższe, 10 razy krótsze itp.),
- *przedziałowe* (nazywane też interwałowymi), mierzone w skali, w której można odejmować wartości (wyznaczać długość przedziału).



# Struktura zbioru danych

ID	PŁEĆ	ZAWÓD	WZROST	DATA URODZENIA
ID_23	K	INFORMATYK	158	1978-03-12
ID_45	K	PRAWNIK	178	1989-05-29
ID_46	M	MATEMATYK	183	1991-01-19
ID_89	M	INFORMATYK	167	1982-02-20
ID_101	K	LEKARZ	163	1973-02-23

# Narzędzia do wizualizacji danych

- programistyczne (R, Python, JavaScript)
- programy graficzne (Inkscape, Gimp, Canva)
- programy dedykowane do wizualizacji danych (Tableau, Power BI)