

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I TELEKOMUNIKACJI

METODY ANALIZY I EKSPLORACJI DANYCH

Wykład 2 - Wizualizacje

DR INŻ. AGATA MIGALSKA



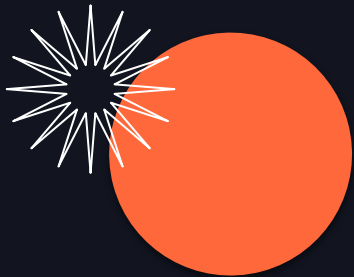
Wykład
2

01
CEL I
MOTYWACJA

02
TYPY WYKRESÓW

03
WSPANIAŁE
WIZUALIZACJE

04
BIBLIOTEKI



01

CEL I MOTYWACJA



**JEDEN OBRAZ
JEST WART
TYSIĄC SŁÓW**

80%
WZROK

jest dominującym zmysłem
u człowieka

13 ms

tyle czasu potrzebuje mózg, żeby przetworzyć jeden obraz

Źródło: <https://news.mit.edu/2014/in-the-blink-of-an-eye-0116>

WARTOŚĆ BIZNESOWA WIZUALIZACJI DANYCH

Analitycy danych, dostarczający spostrzeżenia / wnioski interesariuszom biznesowym:

32%

czasu spędzają na
przygotowywaniu raportów
i prezentacji

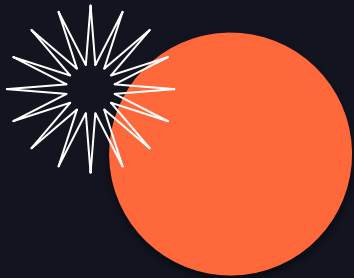
53% z nich

twierdzi, że decyzje
biznesowe opierają się na ich
spostrzeżeniach

64% z nich

twierdzi, że kierownictwo ma
problem ze zrozumieniem
znaczenia tego, co mówią

Źródło: State of Data Science 2021 - Anaconda <https://www.anaconda.com/state-of-data-science-2021>



2022

TYPY WYKRESÓW

* rozkład

- histogram

box and whiskers

violin plot

porównanie dwóch rozkładów

histogram stacked albo grouped

porównanie liczności różnych klas (kategorii)-

bar plot, kołowy wykres

heatmapa - np. korelacja

relacje pomiędzy dwoma zmiennymi -
scatterplot np z linią trendu

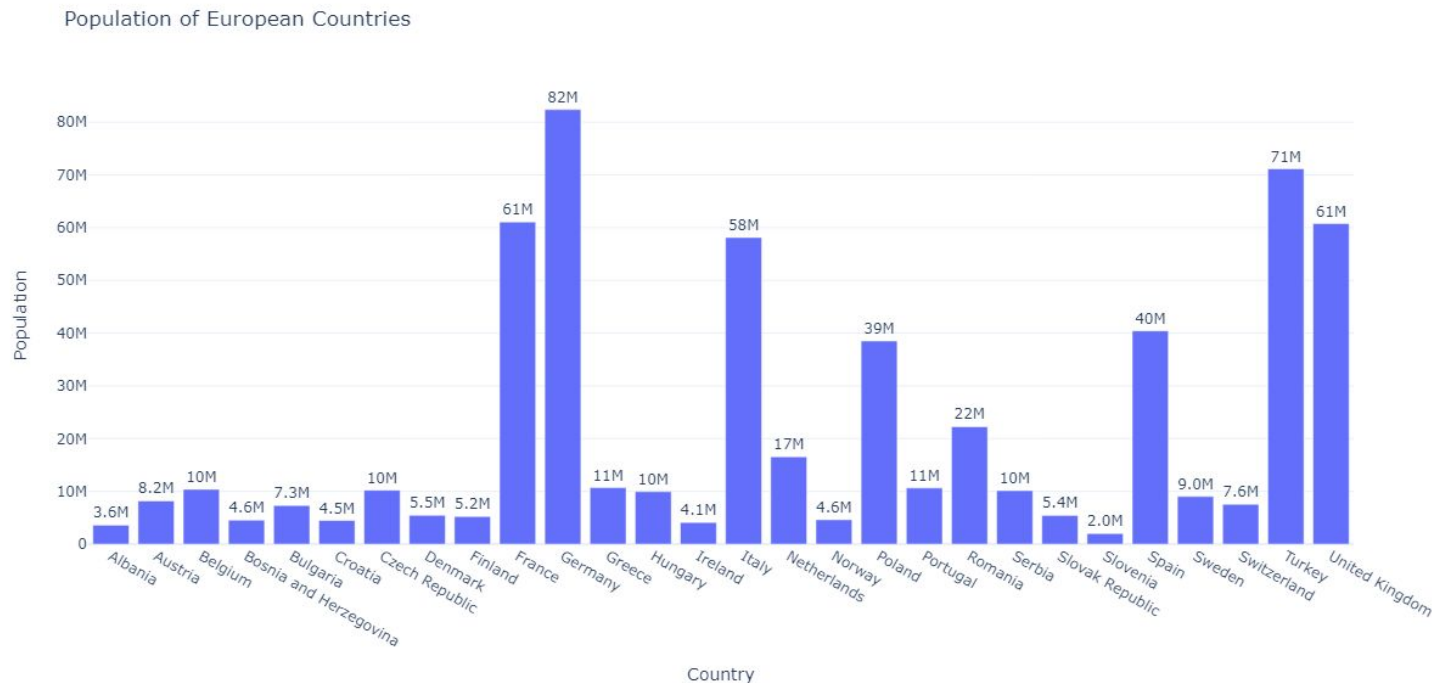
qqplot i pokazać skalę logarytmiczną

chcemy omówić skalę liniową i logarytmiczną

ROZKŁAD JEDNEJ ZMIENNEJ

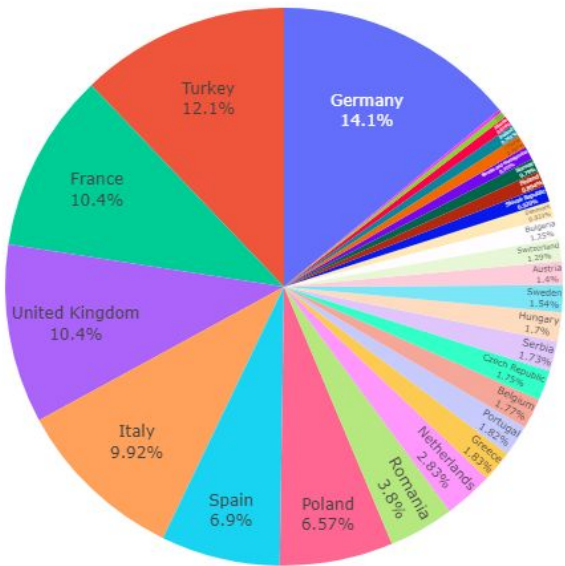


WYKRES SŁUPKOWY

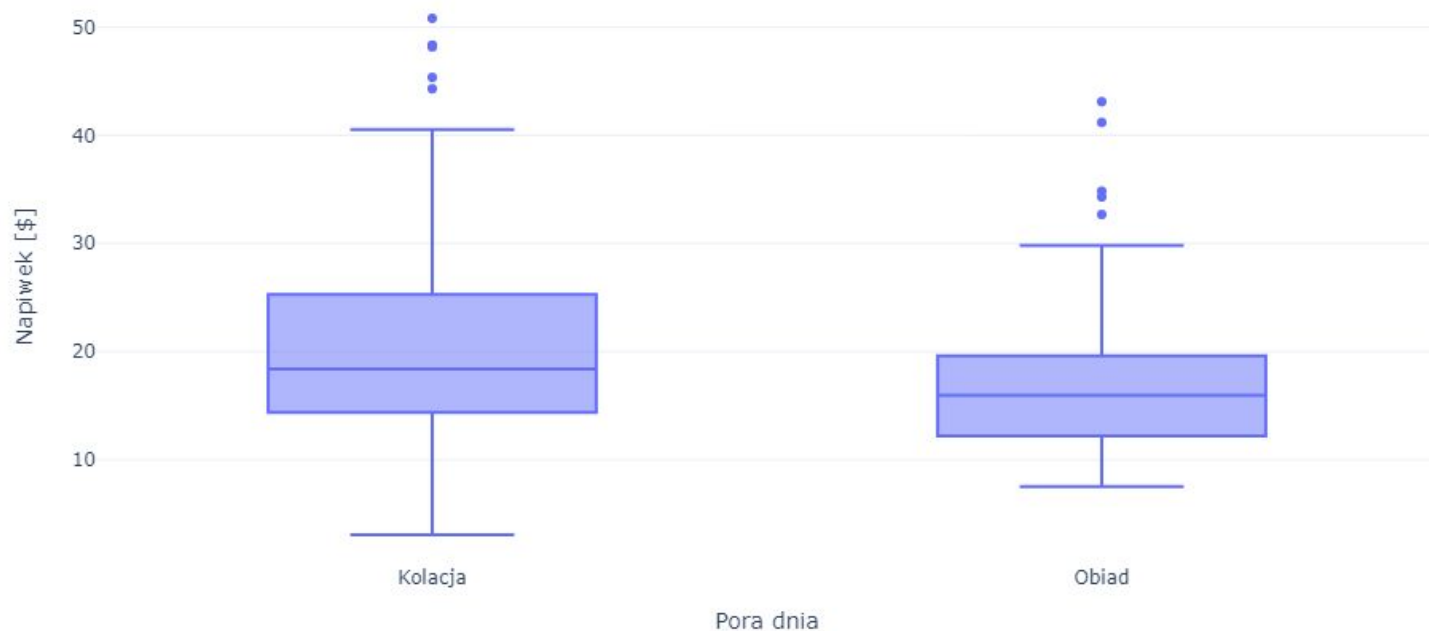


WYKRES KOŁOWY

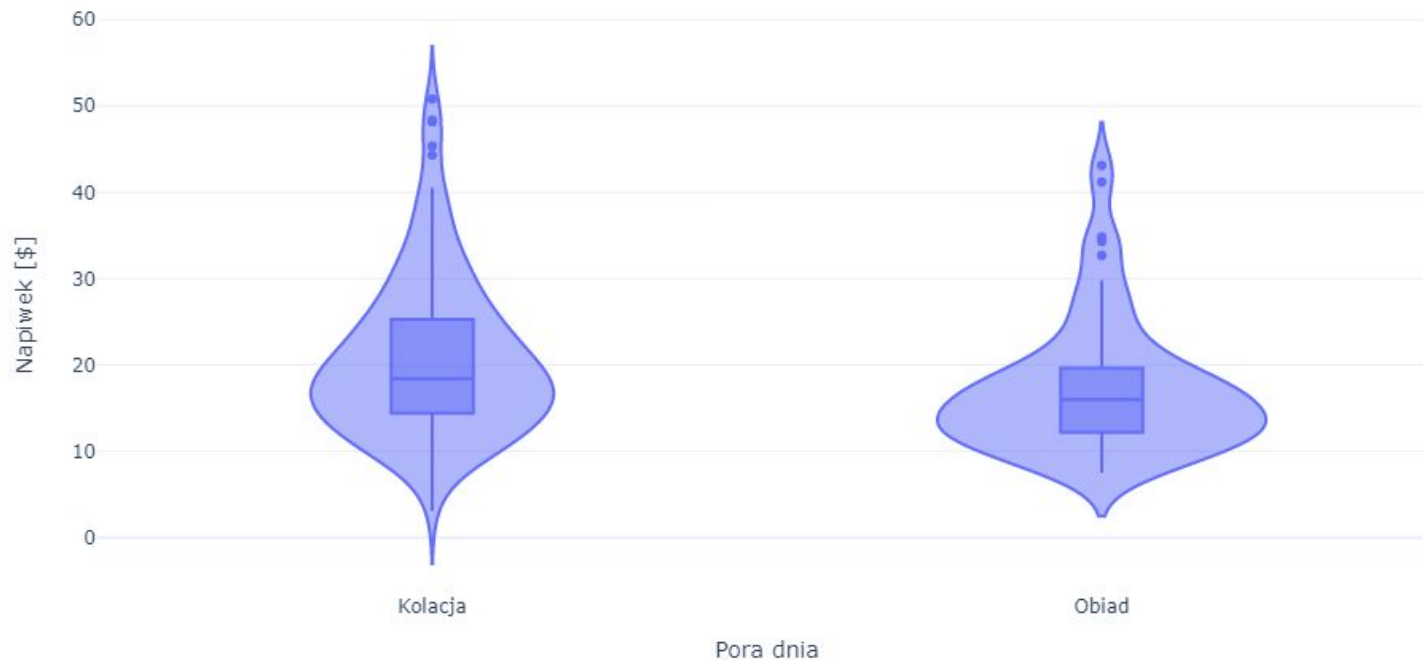
Population of European continent



WYKRES PUDEŁKOWY



WYKRES SKRZYPCOWY



RELACJE POMIĘDZY ZMIENNYMI



WYKRES PUNKTOWY

Oczekiwana długość życia a PKB per capita



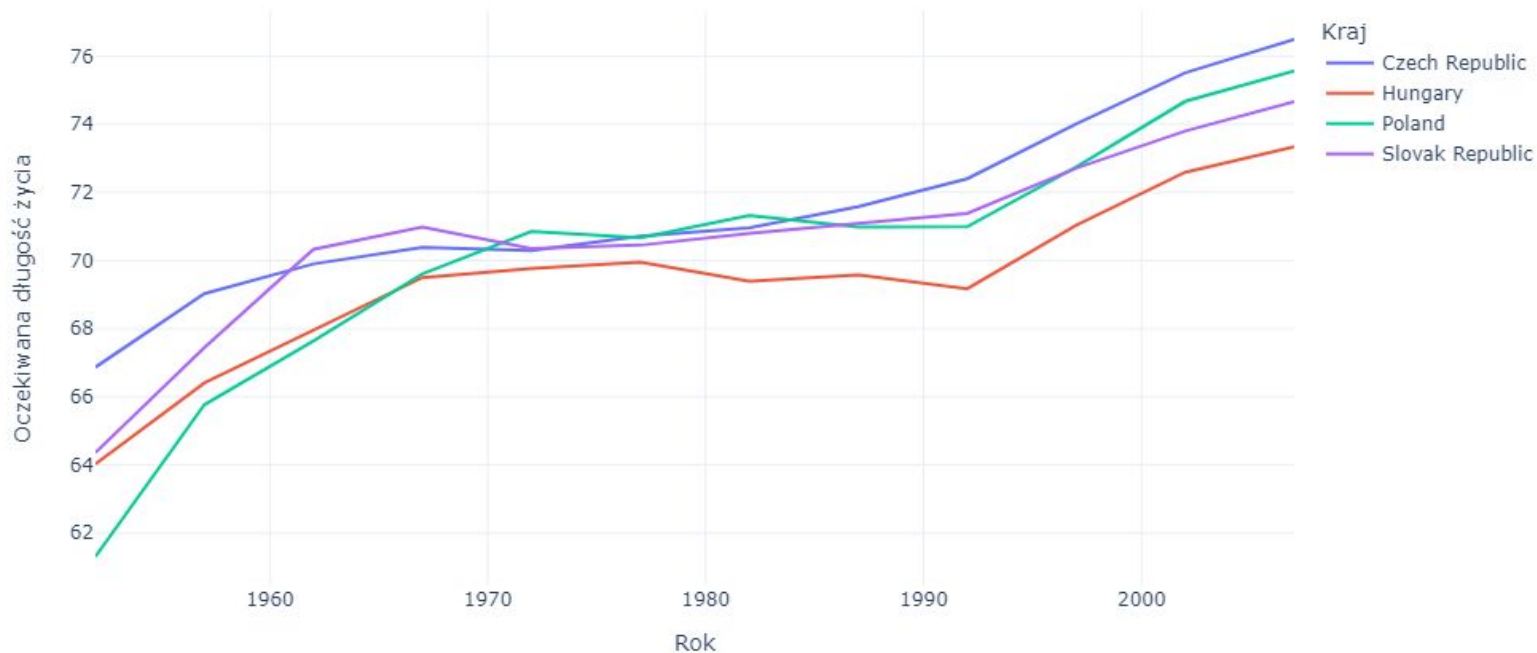
WYKRES PUNKTOWY

Oczekiwana długość życia a PKB per capita

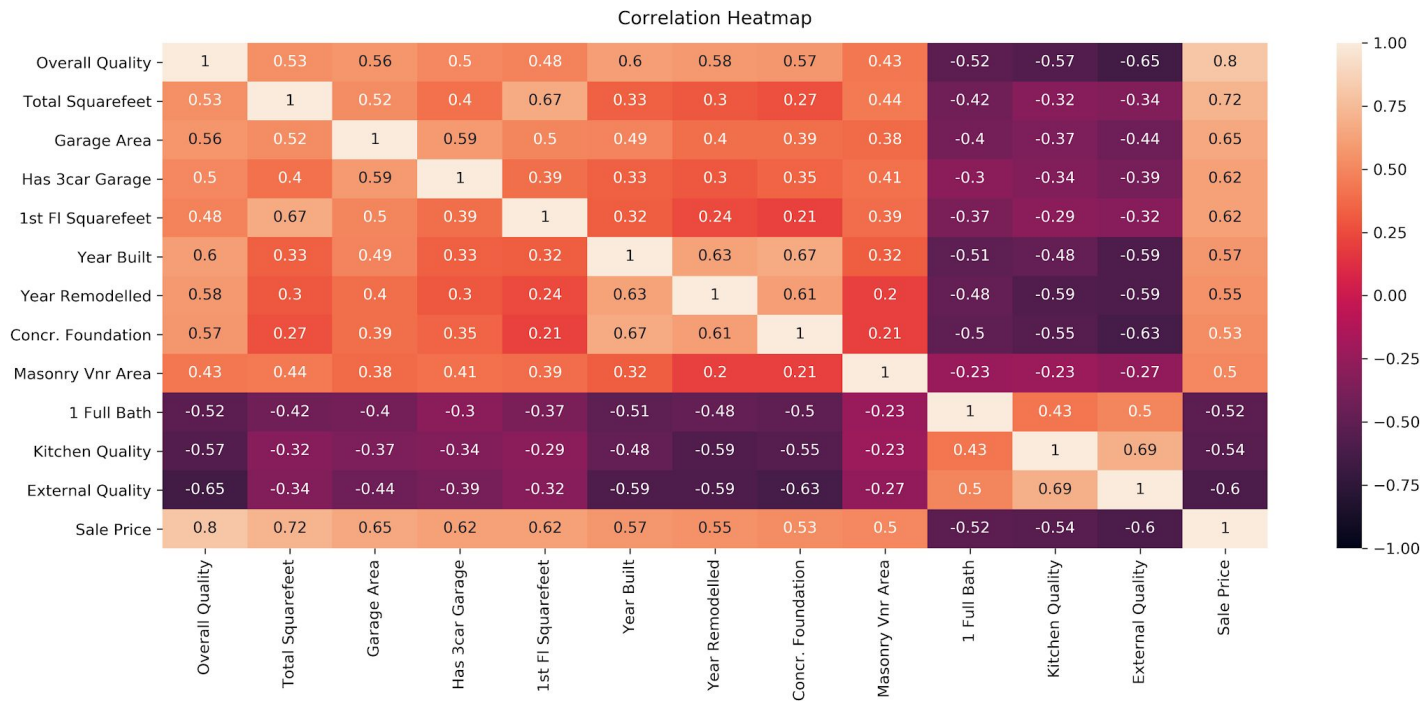


WYKRES LINIOWY

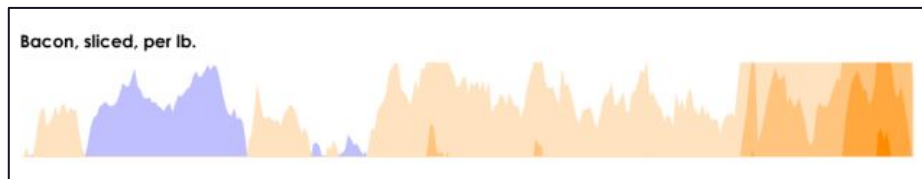
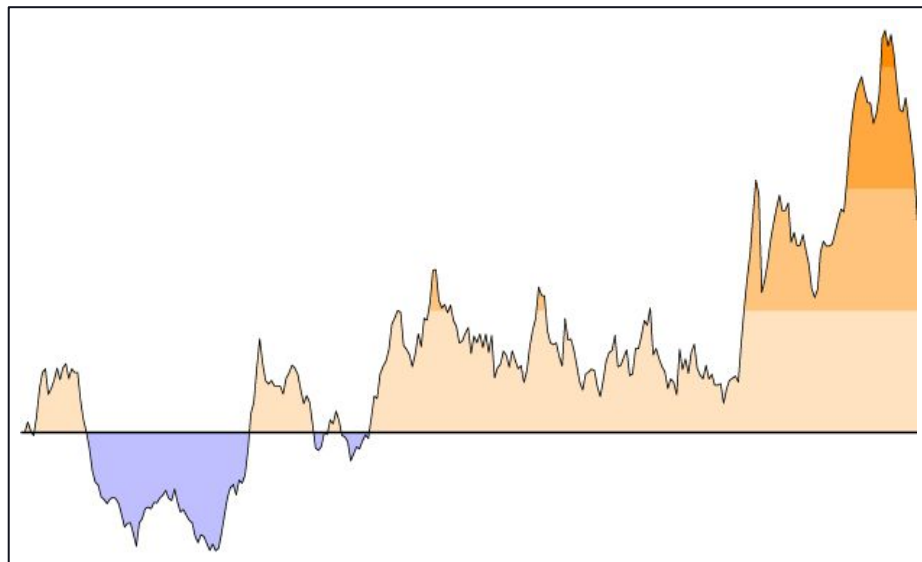
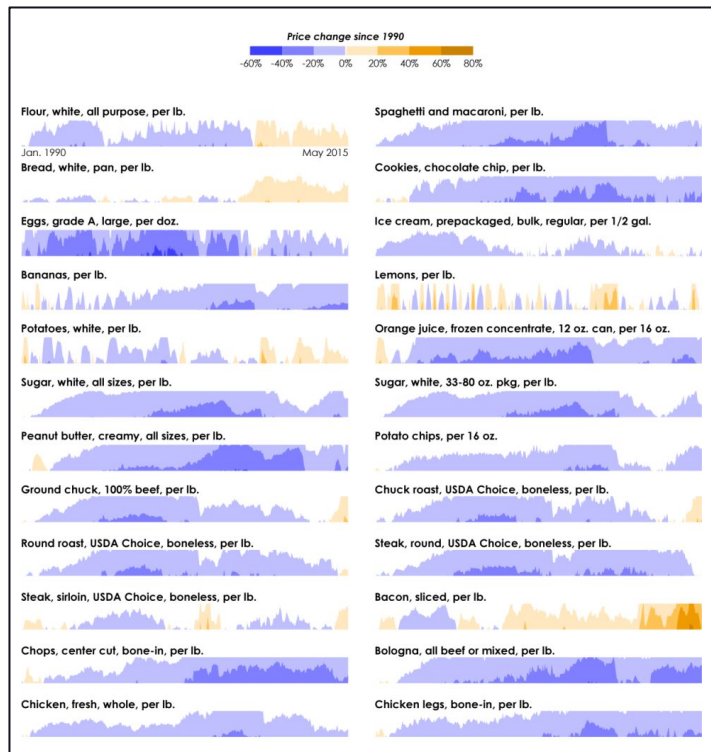
Oczekiwana długość życia w krajach Grupy Wyszehradzkiej



MAPA CIEPLNA



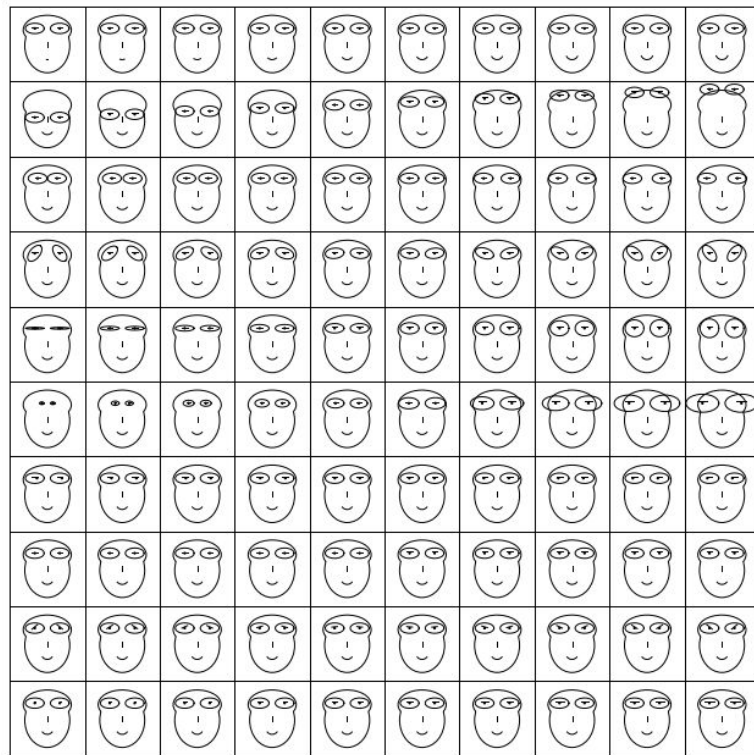
HORIZON GRAPH

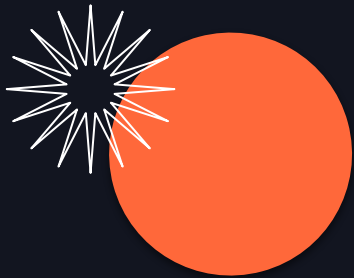


“SZALONE” PRZYKŁADY

Chernoff Faces

- Twarze Chernoffa wyświetlają wielowymiarowe dane w kształcie ludzkiej twarzy.
- Poszczególne części, takie jak oczy, uszy, usta i nos, reprezentują wartości zmiennych poprzez ich kształt, rozmiar, położenie i orientację.
- Uzasadnieniem dla używania twarzy jest to, że ludzie łatwo rozpoznają twarze i bez trudu zauważają drobne zmiany.





03

WSPANIAŁE WIZUALIZACJE

WSPANIAŁE WIZUALIZACJE

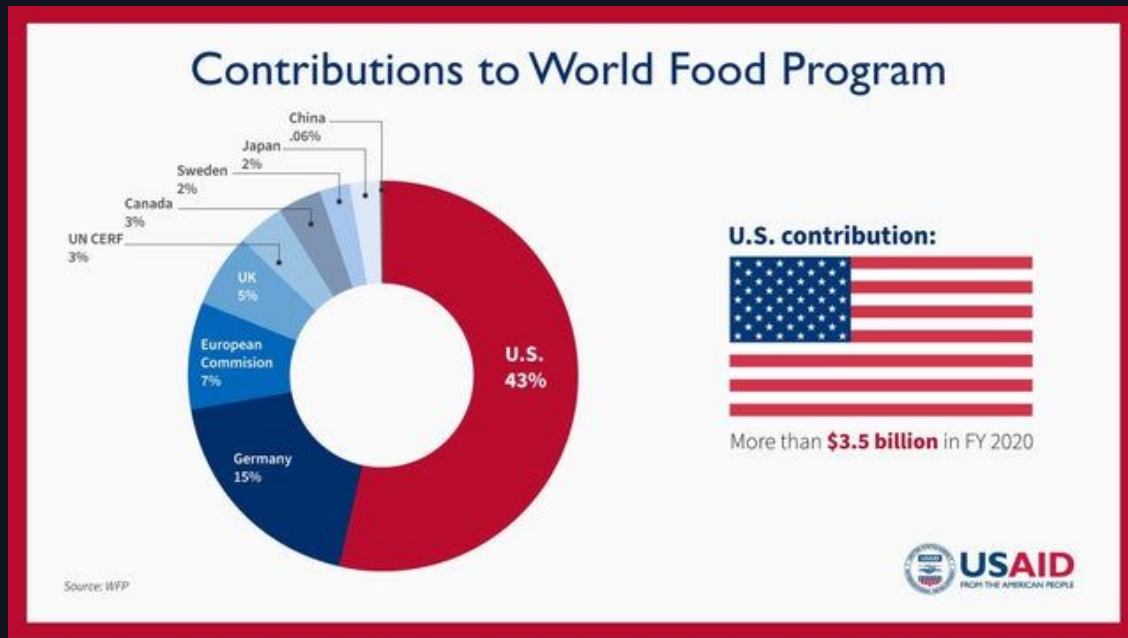
Pięć cech wspaniałych wizualizacji wg Alberto Cairo:

- Prawdomówność (Truthful)
 - Funkcjonalność (Functional)
 - Piękno (Beautiful)
 - Wnikliwość (Insightful)
 - Olśnienie (Enlightening)
-

PRAWDOMÓWNOŚĆ



PRAWDOMÓWNOŚĆ?



Wartości tekstowe
i zaznaczony obszar
powinny być spójne.

PRAWDOMÓWNOŚĆ?

PRESIDENTIAL ELECTIONS, 2013

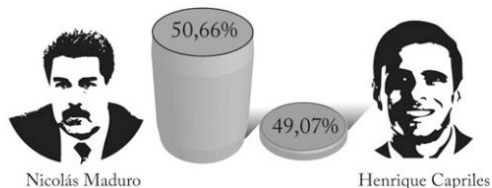


Fig. 5.3 Presidential election results in Venezuela, based on a graphic by Venezonala de Televisión. Notice the truncated Y-axis which greatly distorts the difference between the percentages of vote

PRESIDENTIAL ELECTIONS, 2013

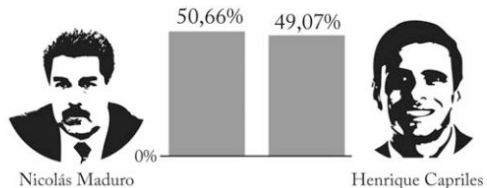


Fig. 5.4 An alternative version of the previous graphic in which a 0-baseline has been added, and the 3D effect has been removed

Oś Y na wykresach słupkowych powinna zawsze zaczynać się w zerze.

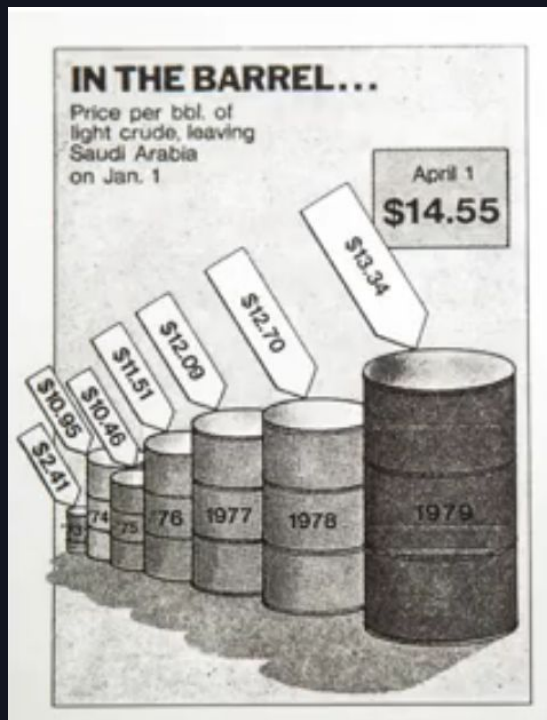
PRAWDOMÓWNOŚĆ?



Odległości pomiędzy kolejnymi punktami powinny zostać zachowane (patrz oś X).

(Dodatkowo: Oś Y na wykresach słupkowych powinna zaczynać się w zerze.)

PRAWDOMÓWNOŚĆ?



Czy cena baryłki ropy reprezentowana jest przez wysokość beczki? A może przez objętość beczki?

$$\text{Współczynnik kłamstwa} = \frac{\text{Efekt pokazany na grafice}}{\text{Efekt w danych}}$$

Współczynnik kłamstwa (zdefiniowany przez Edwarda Tufte) nie powinien być ani większy ani mniejszy od 1!

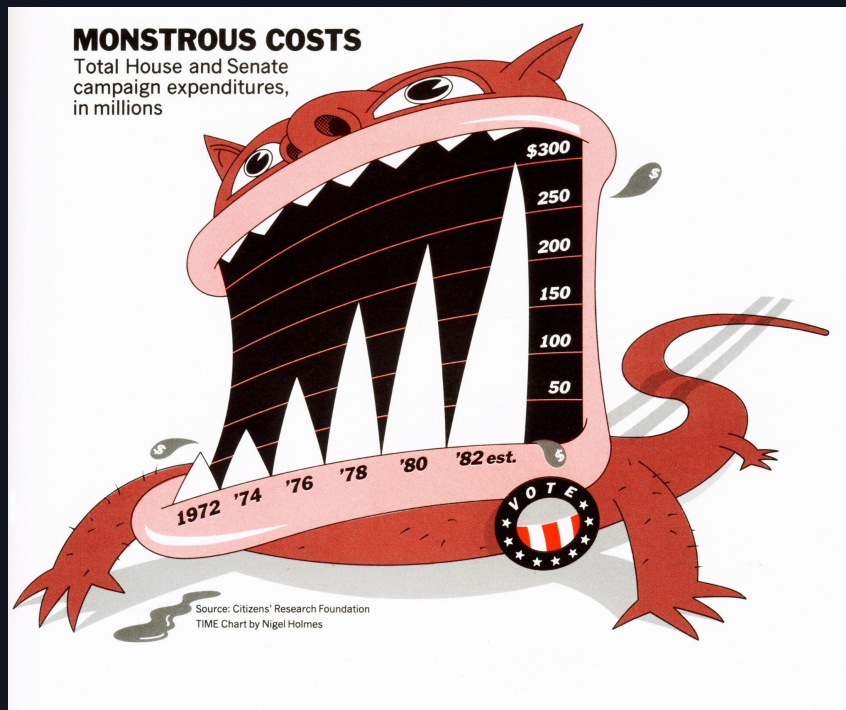
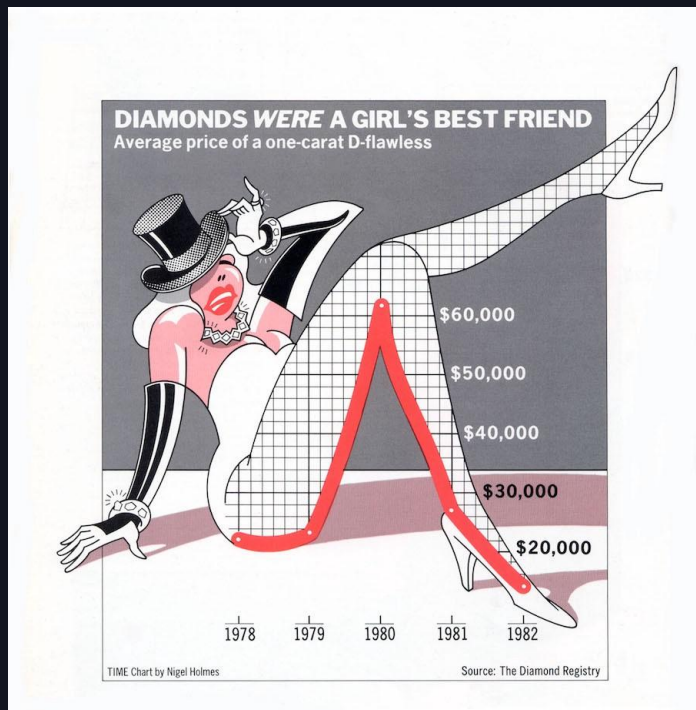


“With great power comes great responsibility”

—SPIDERMAN



CHARTJUNK



CHARTJUNK

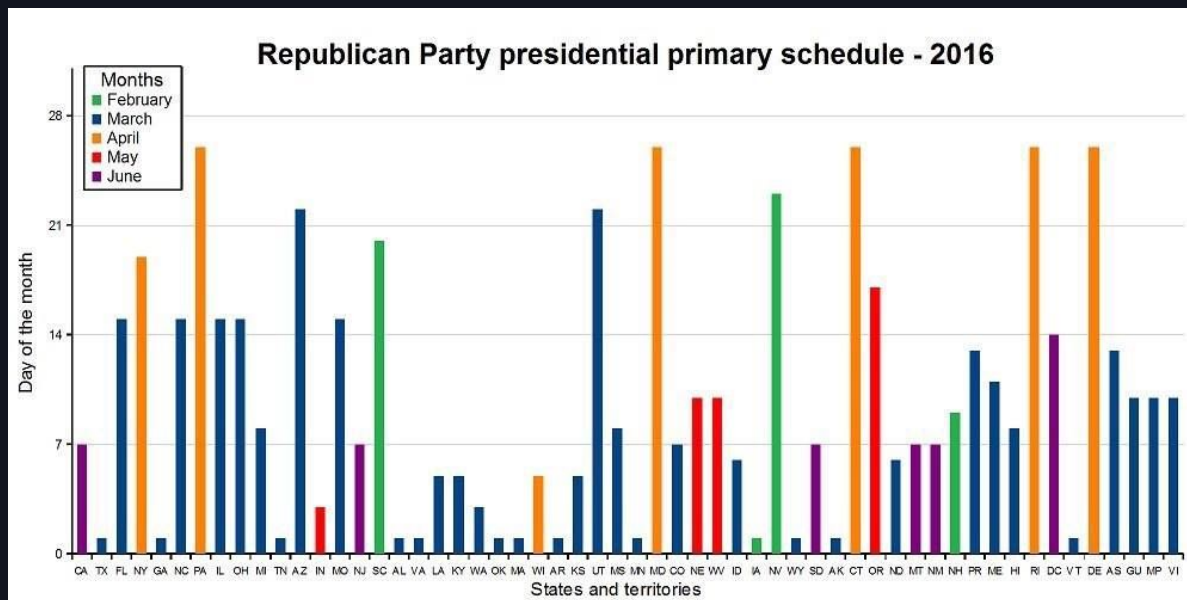
- Grafiki Nigela Holmesa (poprzedni slajd) przedstawiają dane w sposób zapadający w pamięć.
- Wg Edwarda Tufte, twórcy definicji współczynnika kłamstwa (lie factor), taki sposób prezentacji danych zawiera zbędne ozdoby.
- W badaniach porównujących “czyste” grafiki z grafikami Holmesa [Bateman et al. (2010)]:
 - badani lepiej pamiętali zawartość grafik Holmesa po 2-3 tygodniach,
 - badani wskazali subiektywnie, że grafiki Holmesa były przyjemniejsze, atrakcyjniejsze, łatwiejsze do zapamiętania i prostsze do opisanie.

[Bateman et. al (2010)] Bateman, S., Mandryk, R. L., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010, April). In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2573-2582).

FUNKCJONALNOŚĆ



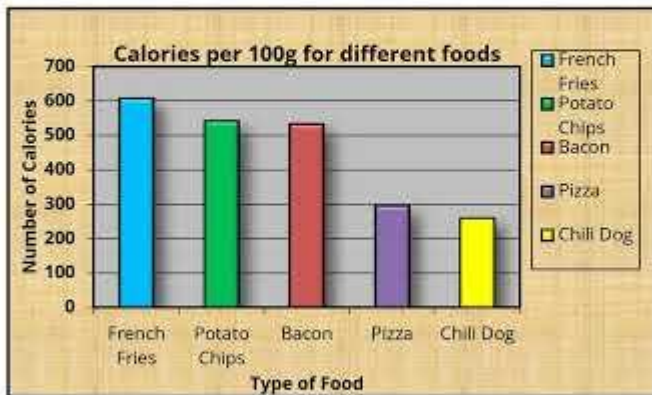
FUNKCJONALNOŚĆ?



Typ wykresu powinien
być dobrany do typu
prezentowanych danych

FUNKCJONALNOŚĆ

Remove backgrounds



Content by Darkzone Analytics

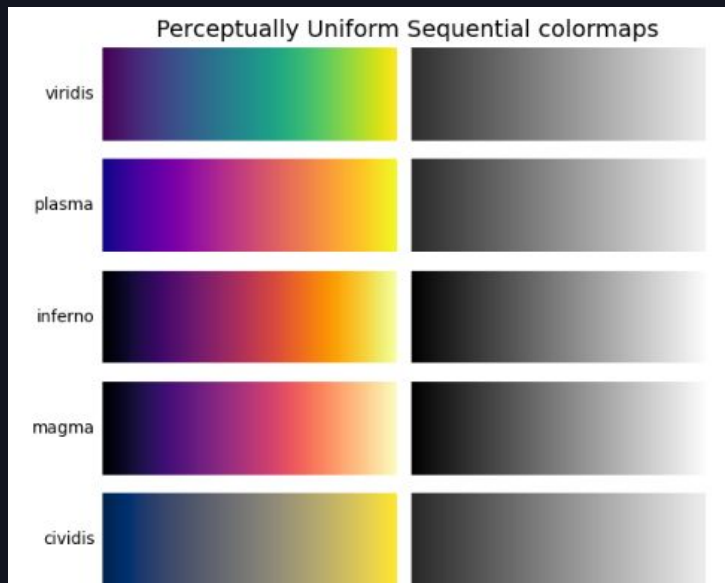
@ha.io/improve

Twitter: @darkzoneanalytics

Wykres zawiera tylko te elementy, które są niezbędne dla jego zrozumienia.

Wykres zawiera wszystkie elementy, które pozwalają go zrozumieć bez dodatkowego komentarza.

KOLOR



Percepcyjnie jednolita mapa kolorów

Równe kroki w danych są postrzegane jako równe kroki w przestrzeni kolorów.

Naukowcy odkryli, że ludzki mózg znacznie lepiej odbiera zmiany parametru jasności jako zmiany danych niż np. zmiany odcienia.

Zaburzenia widzenia barw

Najczęstszym zaburzeniem widzenia barw jest nierozróżnianie czerwieni i zieleni. Unikanie map kolorów z oboma tymi kolorami pozwoli uniknąć wielu problemów.

PIĘKNO / ESTETYKA



Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Mûlard, Inspecteur Général des Ponts et Chaussées en retraite. Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Thiers, de Leguér, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre.

Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davoust qui avaiem été détachés sur Minsk et Mohilow en ont rejoint vers Orscha et Witebsk, avaiem toujours marché avec l'armée.

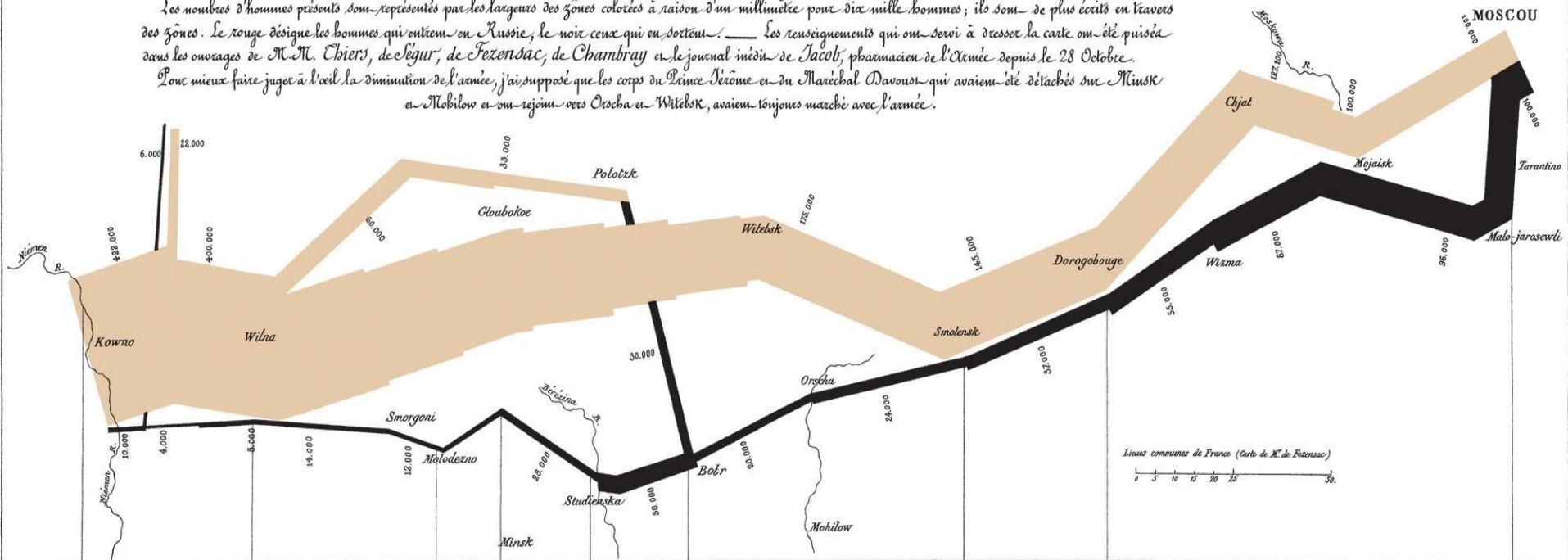
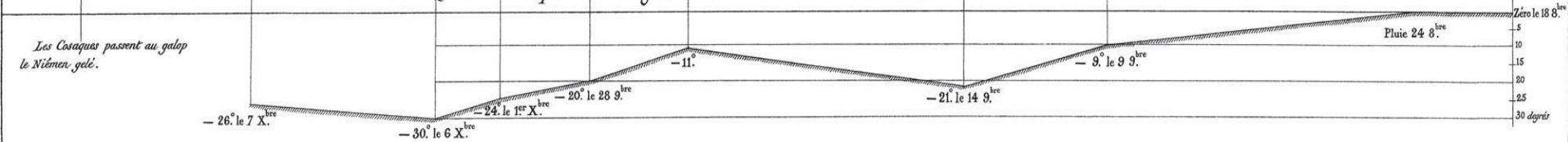


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.





“To może być najlepsza grafika statystyczna,
jaką kiedykolwiek narysowano.”

—**EDWARD TUFTE**,
emerytowany profesor
statystyki i informatyki
na Uniwersytecie Yale



DLACZEGO NAJLEPSZA?

- intuicyjna reprezentacja liczebności armii (1 mm = 6000 żołnierzy)
- położenie geograficzne na dwuwymiarowej powierzchni (długość i szerokość geograficzna)
- kolorystyczne oznaczenie kierunku marszu wojsk
 - "czerwony" w kierunku Rosji,
 - czarny w kierunku przeciwnym

- zarejestrowane temperatury
- zarejestrowane przeszkody topograficzne (rzeki)

*Les Cosaques passent au galop
le Niémen gelé.*

WNIKLIWOŚĆ

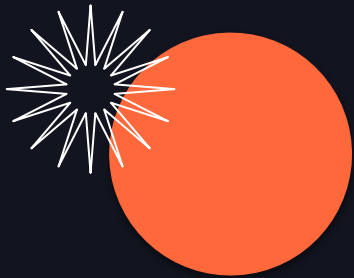
Doskonały wykres powinien towarzyszyć widzowi w wyciągnięciu wniosków z danych.



OLŚNIENIE

Doskonały wykres powinien dostarczyć widzowi olśnienia / momentu “aha”.





04

BIBLIOTEKI

BIBLIOTEKI



- Domyślna biblioteka
- Daje duże możliwości dopasowania wykresu do wymagań
- Domyślny silnik wykresów dla biblioteki pandas



- Nakładka na matplotlib
- Biblioteka dedykowana dla statystycznych wykresów
- Dostarcza estetycznych szablonów



- Interaktywna biblioteka do tworzenia wykresów z implementacją w kilku popularnych językach, w tym Pythonie
- Dostarcza estetycznych szablonów
- Daje pewne możliwości dopasowania wykresu do wymagań

THANKS!

**DZIĘKUJĘ
ZA UWAGĘ**

