

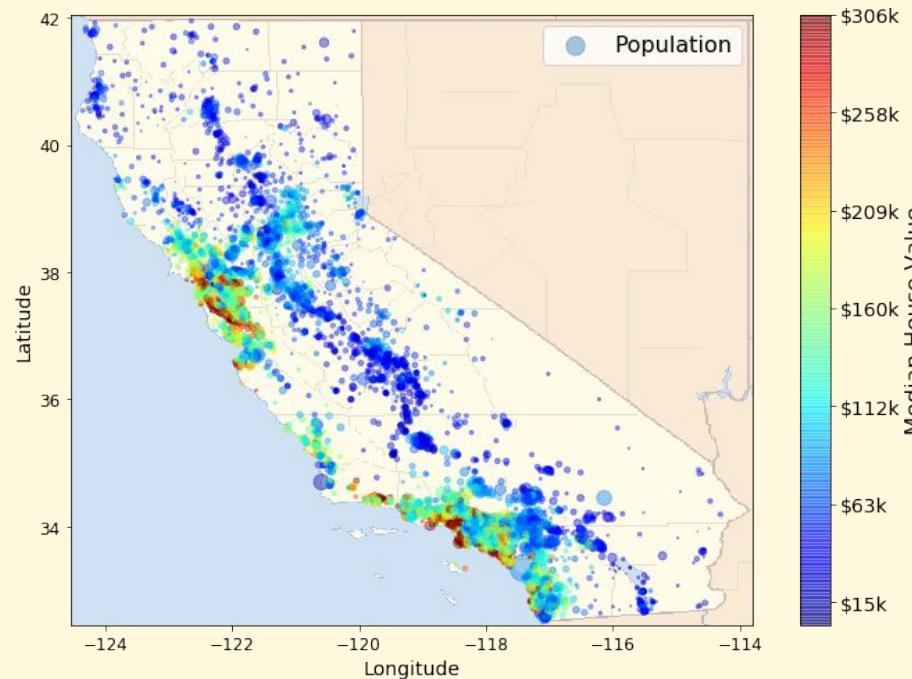
# Wstęp do Eksploracji Danych

Politechnika Warszawska  
Anna Kozak

Mapy - czy to aż tak skomplikowane?

# Motywacja

Mapy są świetnym sposobem do reprezentacji danych przestrzennych.



# Historia

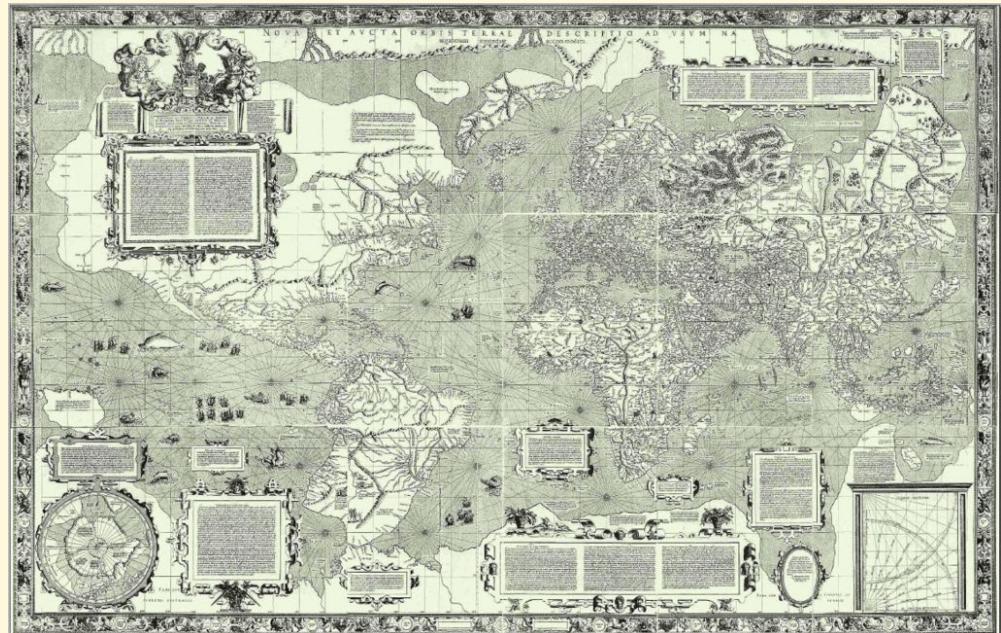
- Jedne z pierwszych tworzonych przez ludzi wizualizacji to wizualizacje kartograficzne - mapy.
- Tworzone je dla różnych potrzeb:
  - nawigacja
  - planowanie rozwoju miasta
  - planowanie sadzenia upraw



babilońska mapa świata  
(VI wiek p.n.e.)

# Jeszcze trochę o historii

- Do XVII wieku nie tworzono map tematycznych, tylko **ogólnogeograficzne** - przedstawiano państwa, miasta, drogi, rzeki itd.
- Kluczowymi zadaniami były:
  - odzwierciedlenie **trójwymiarowej** powierzchni Ziemi na **dwuwymiarowej** płaszczyźnie - szukanie odpowiedniej projekcji
  - poprawa dokładności i prawdziwości map



mapa świata Mercatora  
(1569)

# Teraźniejszość

- **Geowizualizacja** - dziedzina, która narodziła się wraz z przejściem kartografii w erę cyfrową; związana z narzędziami i technikami umożliwiającymi przedstawianie informacji geograficznych
- **GIS (geographic information system)** - system informacji geograficznej

*The relationship between GIS and cartography has been quite ambivalent in recent years.*

*In a special panel discussion on “Has GIS killed cartography?”, (...)  
cartographers answered this question by saying*

*“No, GIS has not killed cartography; cartography committed suicide instead”.*

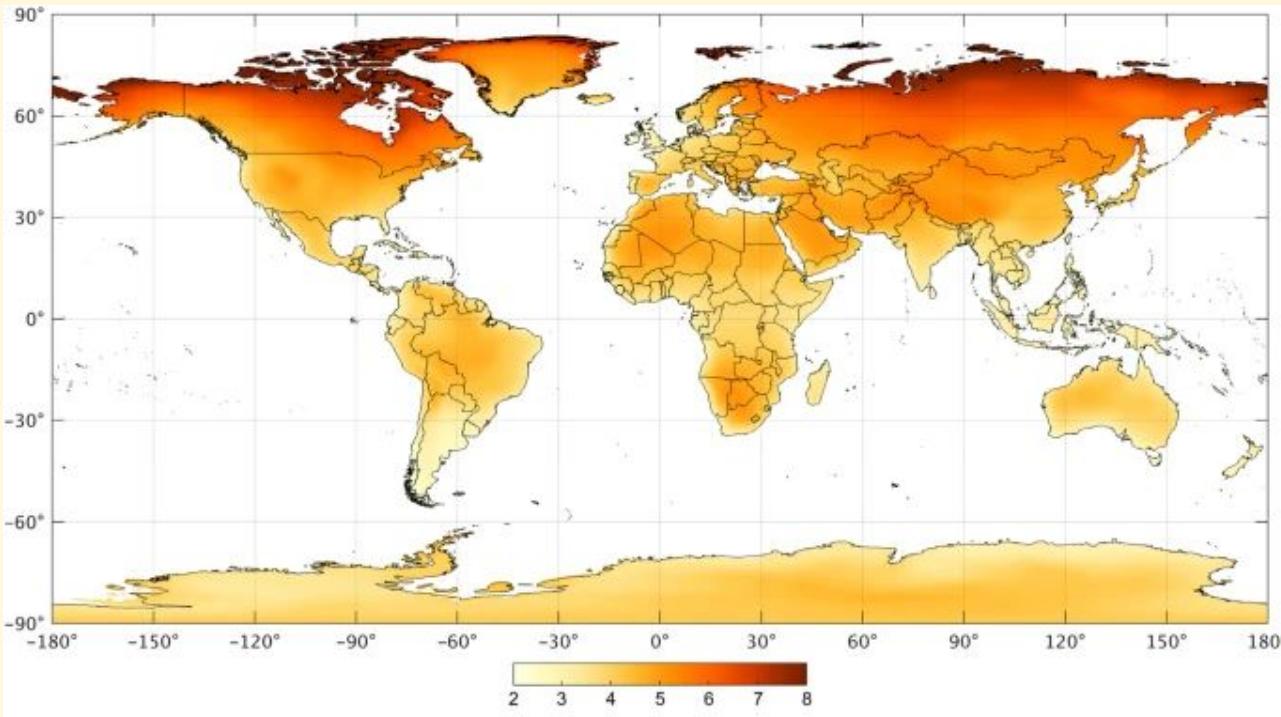
*The answer provided by GIS practitioners was*

*“No, GIS has not killed cartography; they got married”.*

# Teraźniejszość

## Zastosowania geowizualizacji:

- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia

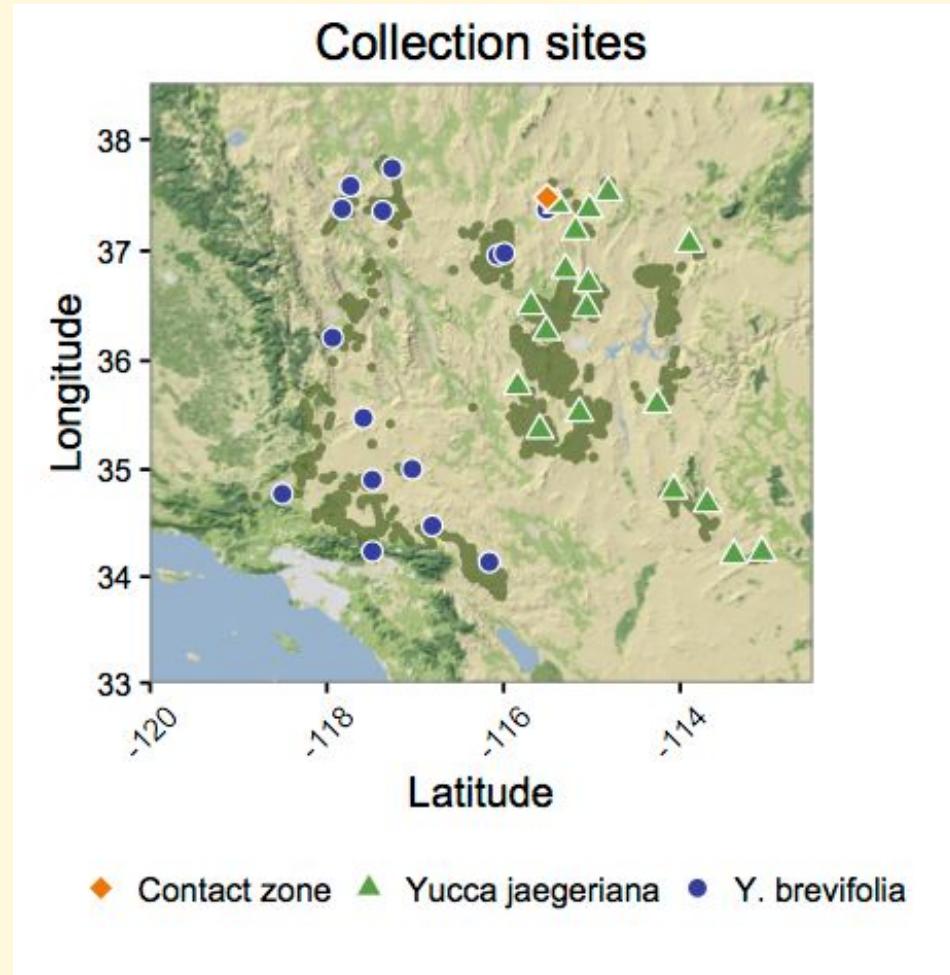


Beck, H., Zimmermann, N., McVicar, T. et al. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, 5, 180214 (2018).

# Teraźniejszość

## Zastosowania geowizualizacji:

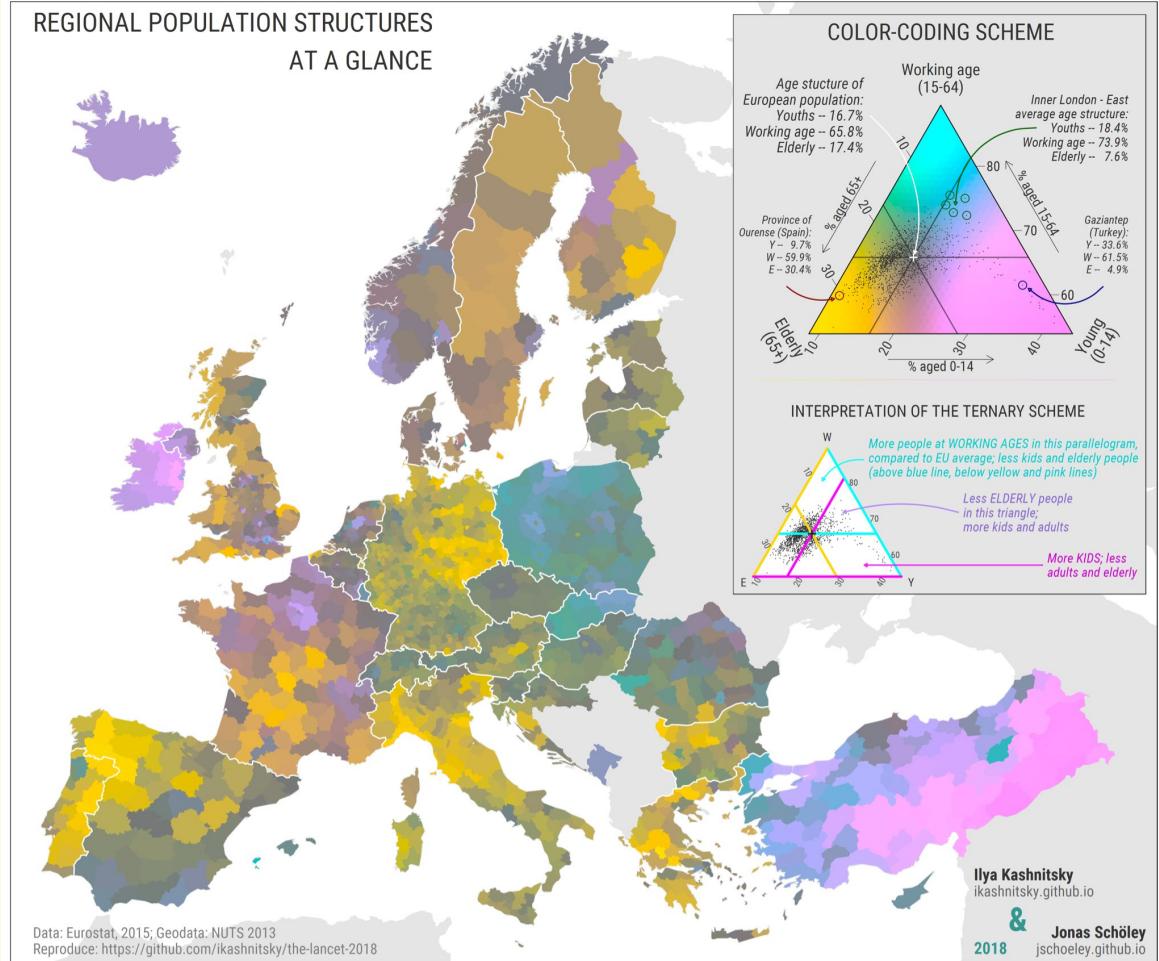
- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia



# Teraźniejszość

## Zastosowania geowizualizacji:

- klimatologia
- ekologia
- **demografia**
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia



Kashnitsky, I., Schöley, J. Regional population structures at a glance.  
*The Lancet*, 392(10143), 209–210 (2018).

# Teraźniejszość

## Zastosowania geowizualizacji:

- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- **transport**
- historia

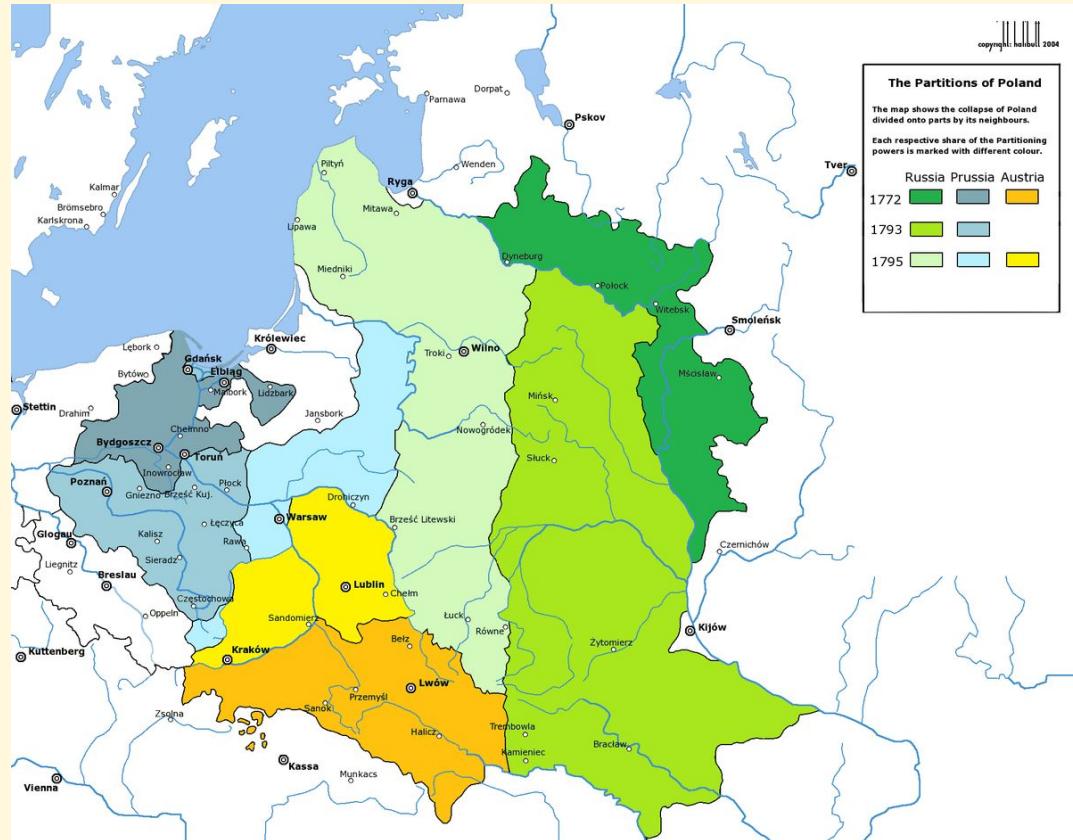


Opracowanie własne

# Teraźniejszość

## Zastosowania geowizualizacji:

- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia



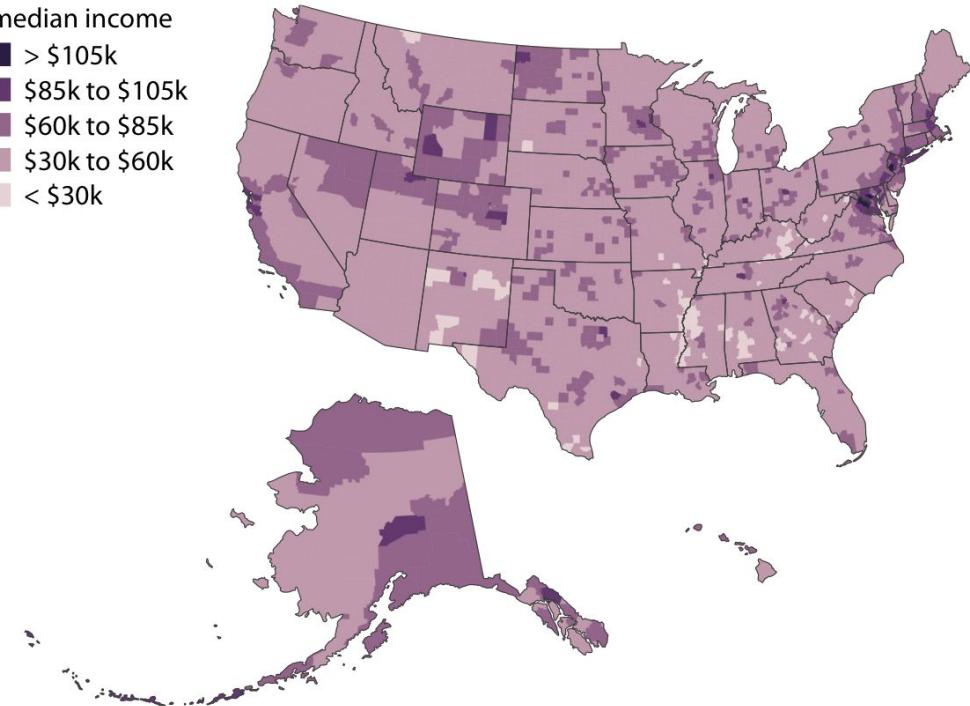
# Metody prezentacji danych na mapach

## KARTOGRAM (choropleth)

- wypełnienie kolorami powierzchni
- kolory kodują wartości (kategoryczne lub numeryczne)
- większe powierzchnie przykuwają większą uwagę, mniejsze powierzchnie mogą być niezauważalne
- często warto grupować zmienną ciągłą w dyskretne grupy (przedziały)
- kluczowa jest normalizacja - warto przedstawić gęstości, wskaźniki, względne wartości, należy unikać bezwzględnych wartości

median income

- > \$105k
- \$85k to \$105k
- \$60k to \$85k
- \$30k to \$60k
- < \$30k

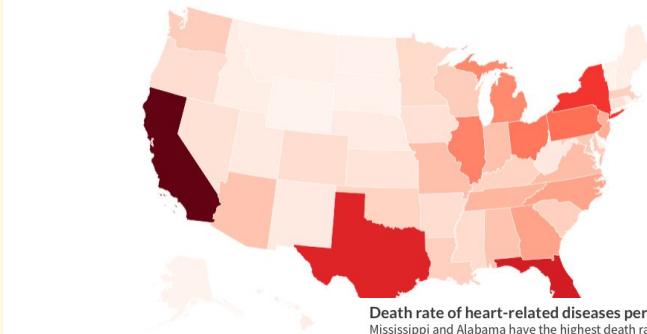


# Metody prezentacji danych na mapach

Total heart-disease related deaths per state

California, Texas, Florida, and New York registered the highest number of deaths

# of deaths 1068 — 93,037

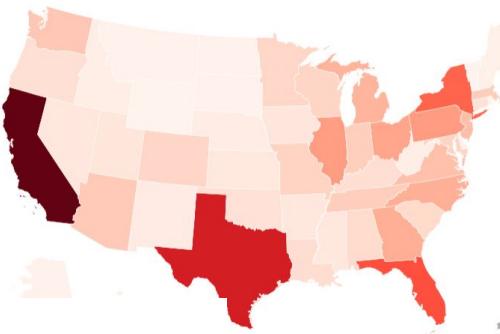


Source: U.S. Center for Disease Control and Prevention • A Flourish Projection map

Total population per state in 2020

California, Texas, Florida, and New York are the most populated states

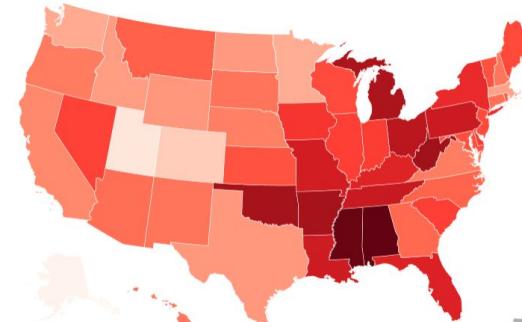
Population 582,328 — 39,368,078



Death rate of heart-related diseases per state

Mississippi and Alabama have the highest death rate in the country

Deaths per 10K people 14.61 — 38.87



Source: U.S. Center for Disease Control and Prevention • A Flourish Projection map

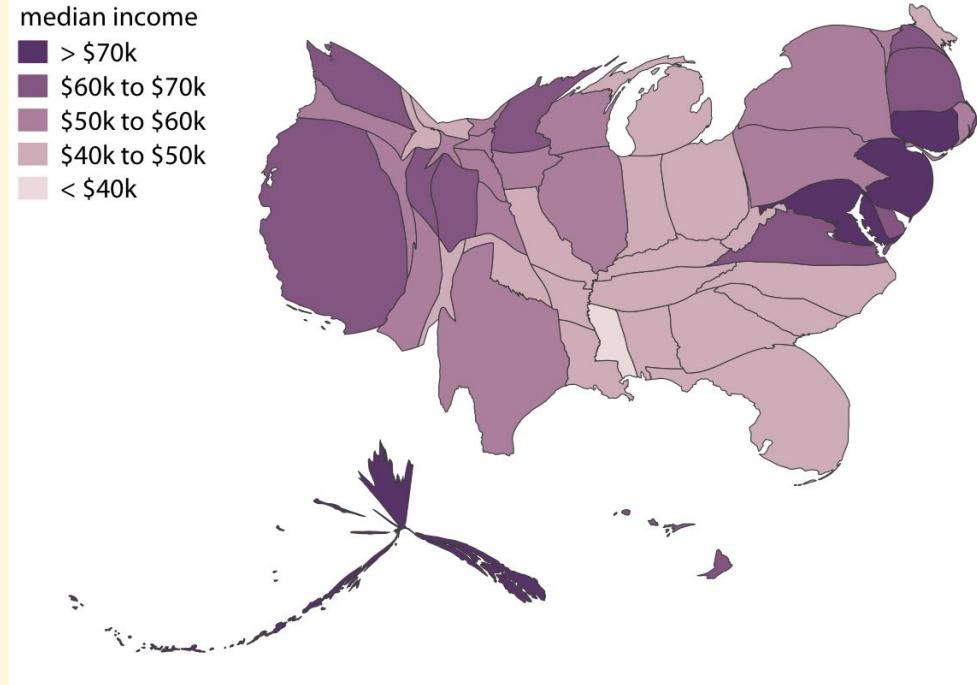
# Metody prezentacji danych na mapach

## KARTOGRAM ANAMORFICZNY (cartogram)

- zmiana geometrii regionów w celu reprezentacji danych, rozkładu zmiennej
- wypaczona przestrzeń geograficzna
- specjalny rodzaj – zgeometryzowane (wszystkie regiony o tym samym kształcie i rozmiarze)
- niwelują niektóre wady kartogramów typu choropleth

median income

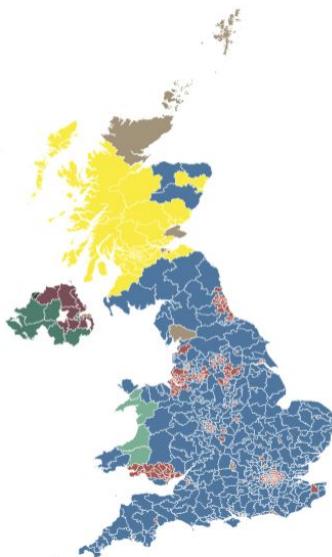
- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k



# Metody prezentacji danych na mapach

Results of the 2019 UK General Election

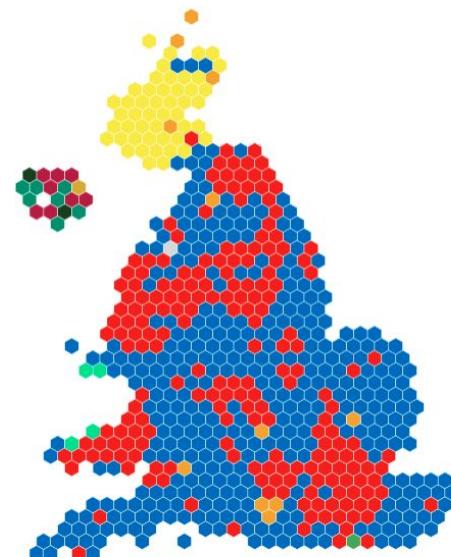
■ Con ■ Lab ■ LD ■ SNP ■ Green ■ PC ■ DUP ■ SF ■ SDLP ■ Alliance ■ Spk



Source: [ONS Geography Portal](#) and [OpenDataNI](#), House of Commons Library • A Flourish 3D map

Results of the 2019 UK General Election

■ Con ■ Lab ■ LD ■ SNP ■ Green ■ PC ■ DUP ■ SF ■ SDLP ■ Alliance ■ Spk

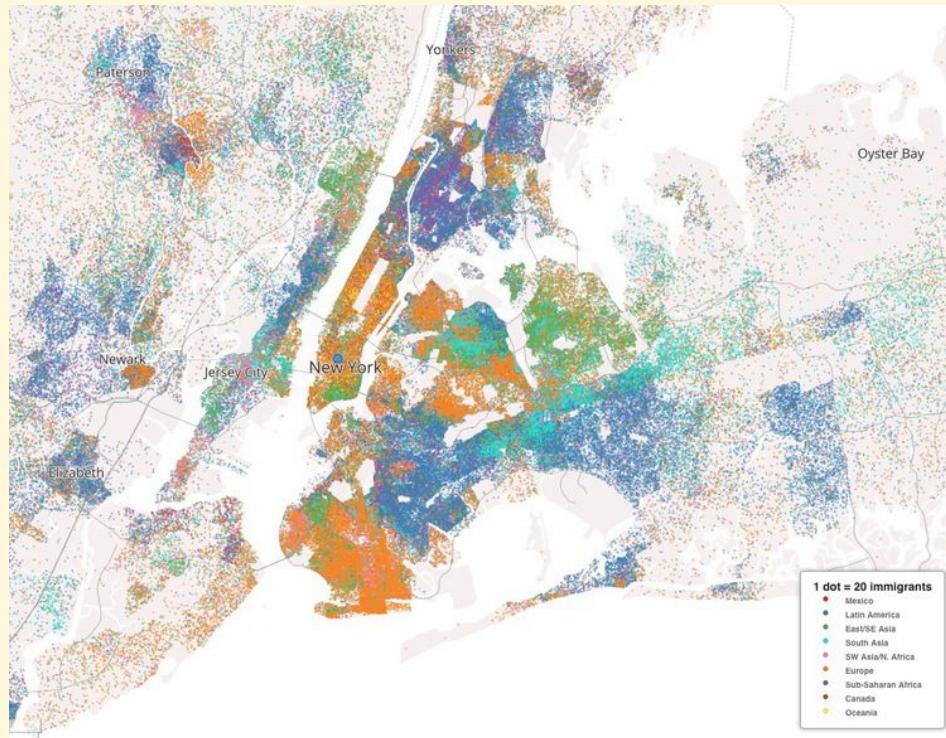


Source: [ODI Leeds](#), House of Commons Library • A Flourish 3D map

# Metody prezentacji danych na mapach

## MAPA KROPKOWA

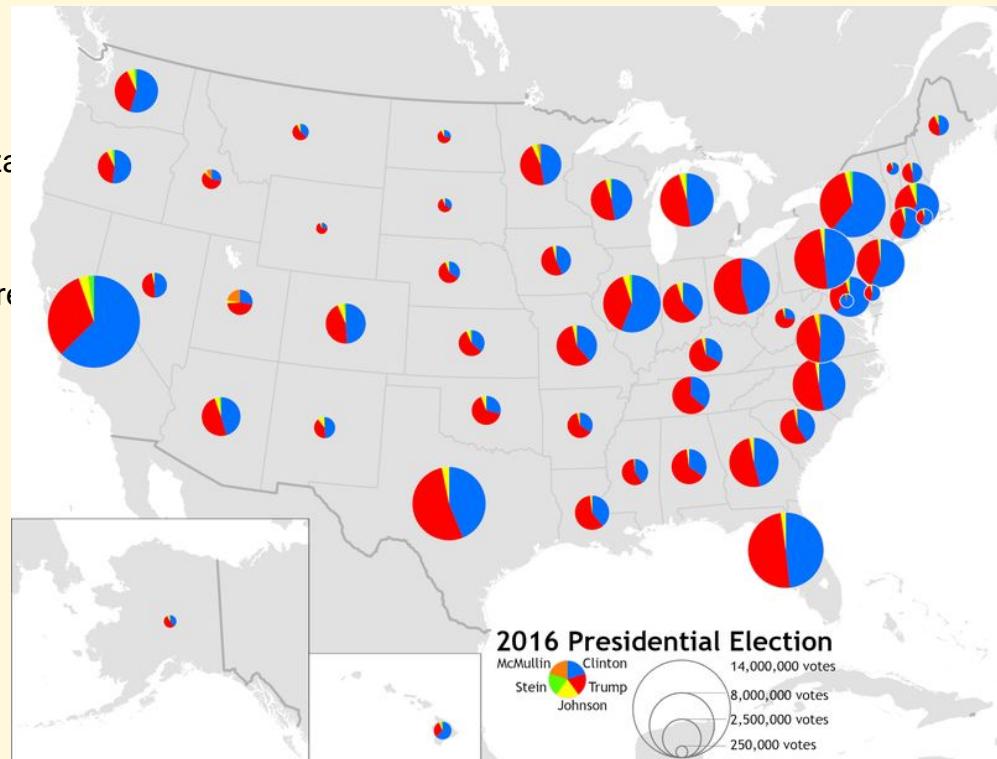
- informacja - umieszczenie znaków w odpowiednim miejscu na mapie
- jeden znak (kropka) - pewna liczba obiektów (niekoniecznie jeden)
- jedynie dane bezwzględne
- waga danego zjawiska, różna liczliwość obiektów → rozmiar znaku
- więcej niż jedno zjawisko, typ obiektu → kolor/typ znaków
- manipulując wielkością znaków, trzeba pamiętać, że ludzkie oko nie jest dobre w odróżnianiu powierzchni



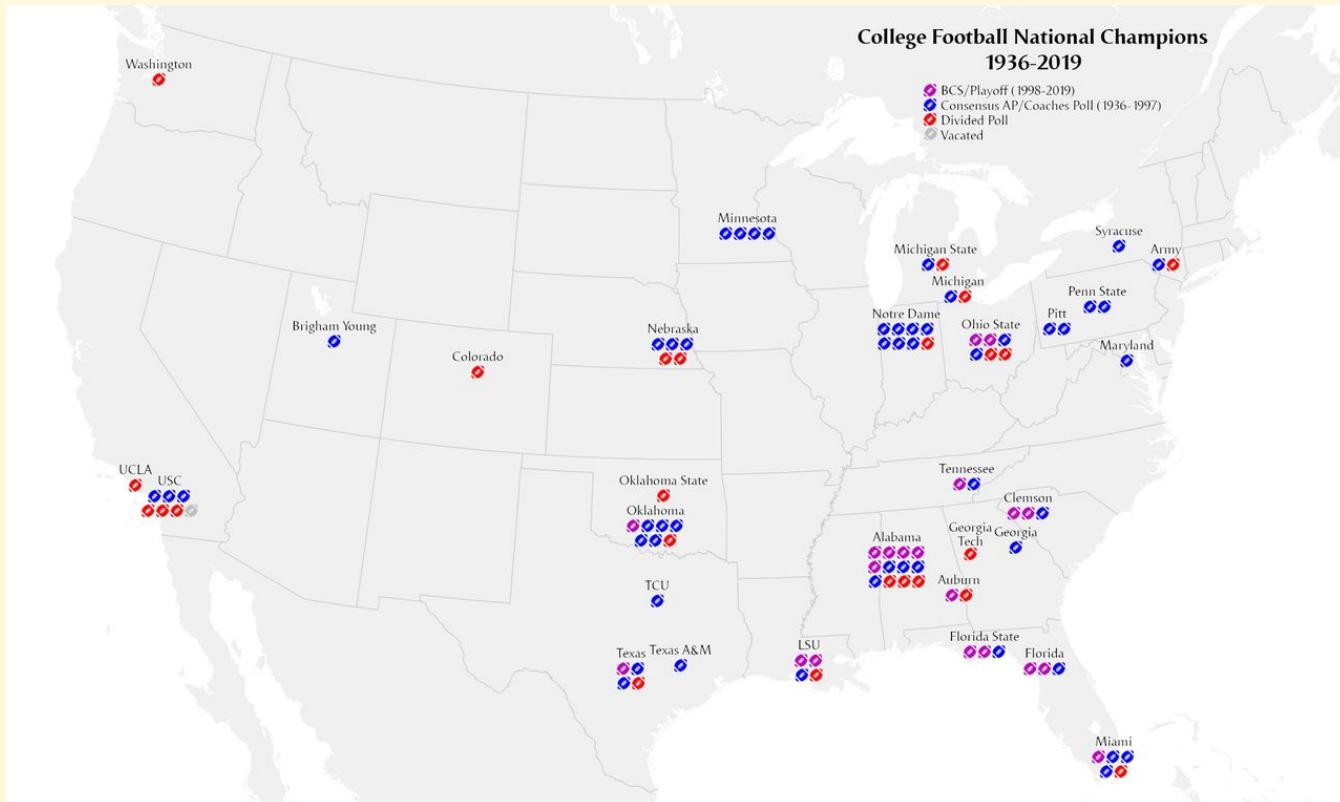
# Metody prezentacji danych na mapach

## KARTODIAGRAM

- umieszczenie diagramów lub wykresów w obrębie obiektów przestrzennych, których dotycza (chart maps)
- UWAGA!** często pojawiają się wykresy kołowe!
- inny typ: umieszczenie symboli/piktogramów w obrębie odpowiednich miejscach na mapie (isotype map)

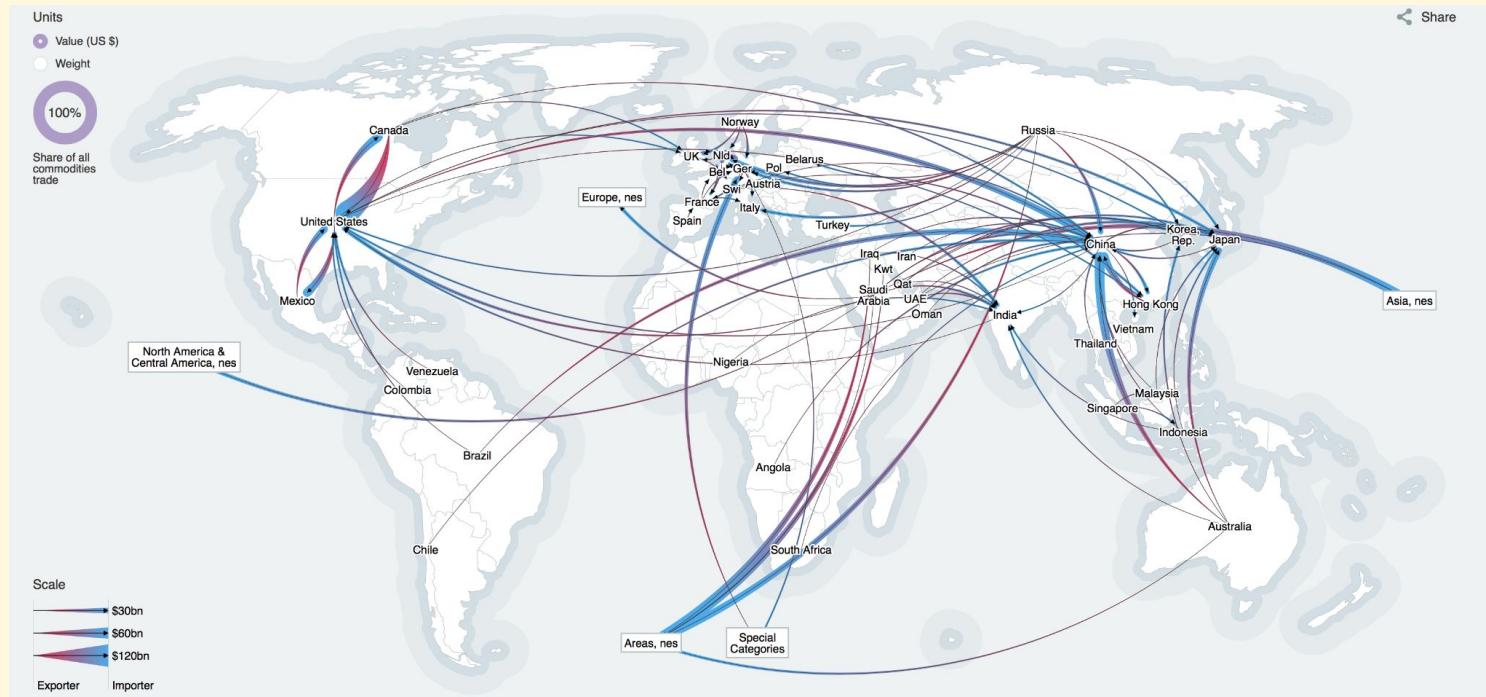


# Metody prezentacji danych na mapach



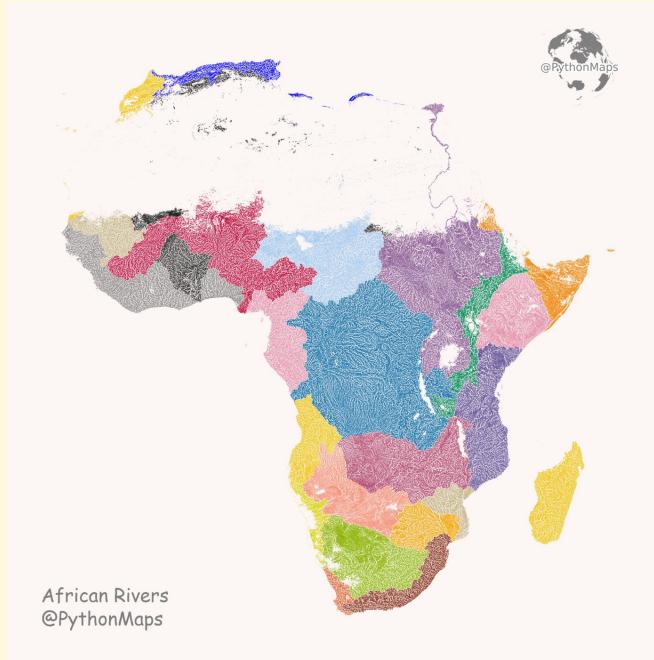
# Metody prezentacji danych na mapach

## MAPA PRZEPŁYWU

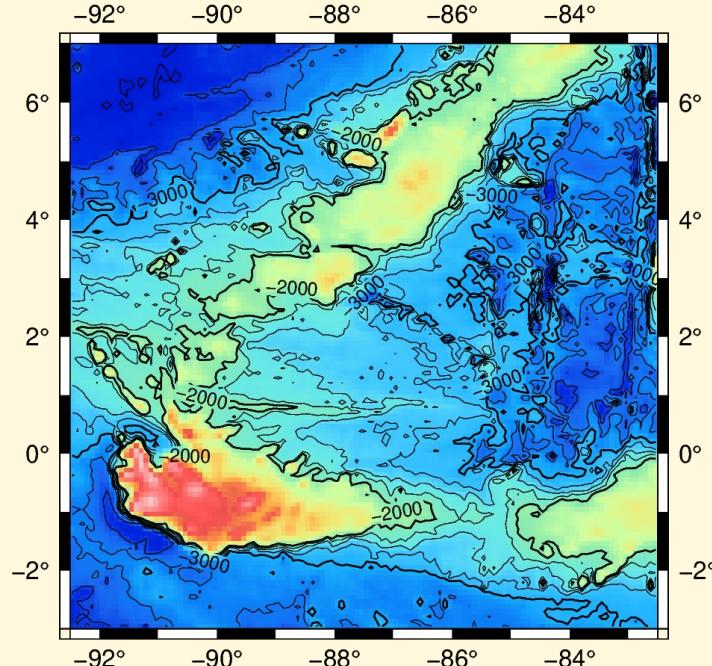


# Metody prezentacji danych na mapach

MAPA CHOROCHROMATYCZNA



MAPA IZOLINIOWA + MAPA CIEPŁA



<https://towardsdatascience.com/creating-beautiful-river-maps-with-python-37c9b5f5b74c>

<https://www.pygmt.org/v0.3.0/tutorials/contour-map.html>

# Koordynaty

- CRS = coordinate reference system
- **Dlaczego nie jeden układ?**
  - Ziemia – geoida
  - minimalizacja błędów i przekłamań
  - zależność od skali

## Ważne przykłady:

- WGS84 – układ jednolity dla całego świata wykorzystywany w GPS
- EPSG:3857 – układ stosowany w mapach internetowych

“Dziwne” błędy są zazwyczaj spowodowane brakiem ujednolicenia systemów odniesienia.

# Odwzorowania



# Odwzorowania

Właściwości do zachowania:

- kształty
- odległości
- kąty
- powierzchnie

Nie jest możliwe osiągnięcie **więcej niż jednej**  
z powyższych właściwości równocześnie!

Cel i typ mapy może przekładać się na wybór odpowiedniego odwzorowania.

# Odwzorowania

## ODWZOROWANIE MERKATORA

- wiernie zachowuje kąty
- używane w mapach nawigacyjnych
- używane w mapach Google
- zniekształcenia tym większe, im dalej od równika

## Kiedy stosować?

- dla dobrze rozpoznawalnej mapy świata
- kiedy rozmiar nie ma takiego znaczenia jak ogółem przedstawienie



# Odwzorowania

## ODWZOROWANIE ROBINSONA I MOLLWEIDEGO

- wiernie zachowują powierzchnię  
(problemy przy biegunach)

### Kiedy stosować?

- dla dobrze rozpoznawalnej mapy świata
- kiedy rozmiar obiektów jest istotny, ważniejsza jest powierzchnia niż kształt



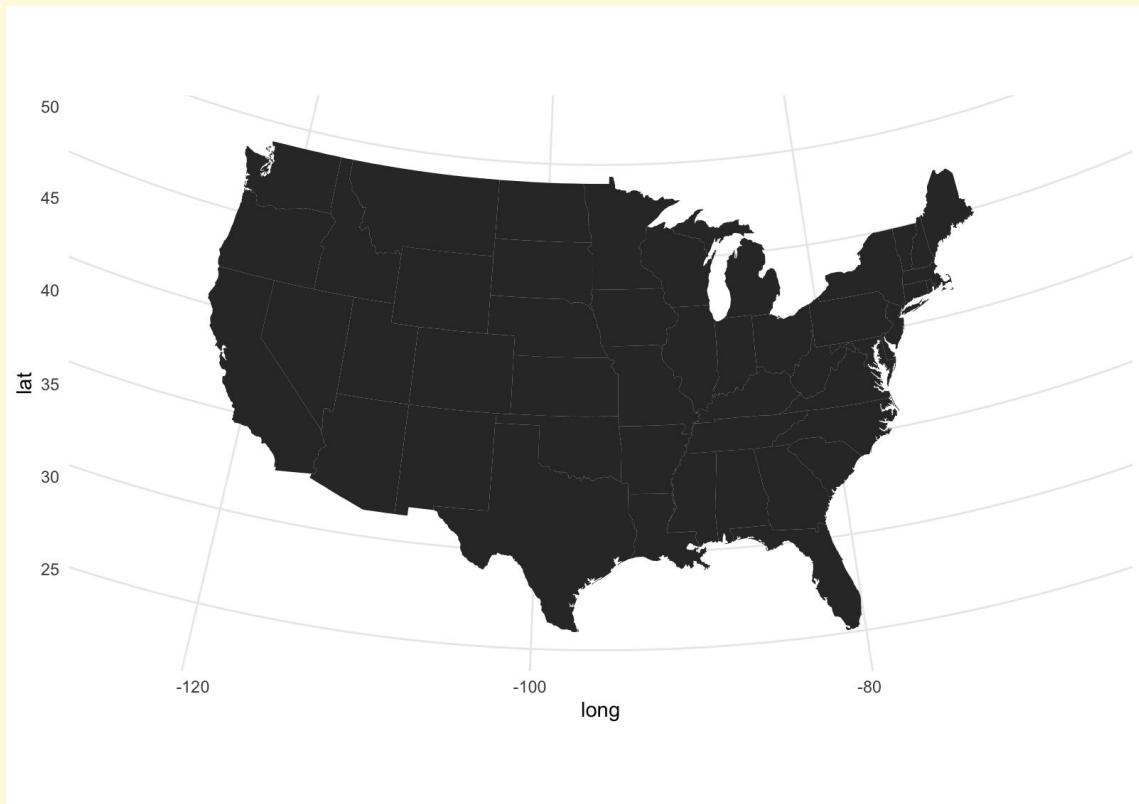
# Odwzorowania

## ODWZOROWANIA STOŽKOWE (NP. ALBERSA)

- wszystkie zniekształcenia są niewielkie pomiędzy dwoma wyróżnionymi równoleżnikami

### Kiedy stosować?

- dla przedstawienia obszarów o dużej rozciągłości równoleżnikowej (tzn. rozciągłości zachód-wschód)



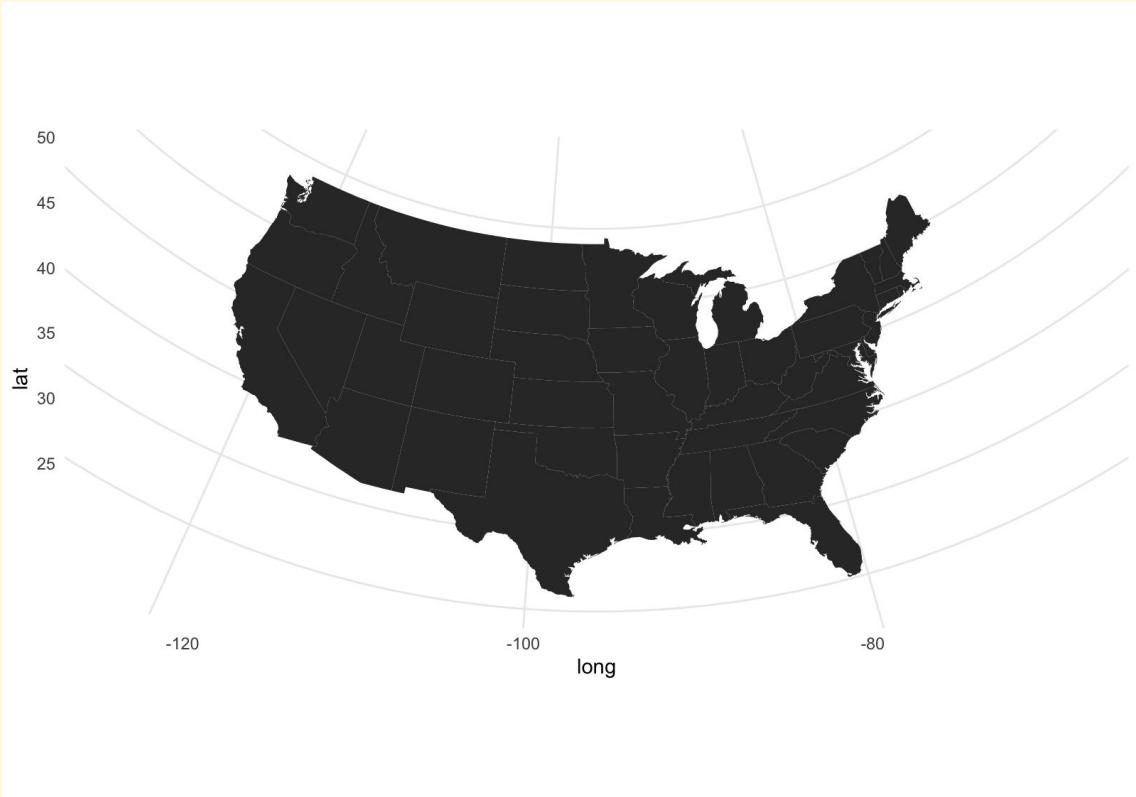
# Odwzorowania

## ODWZOROWANIA AZYMUTALNE

- zachowuje kierunki i odległości od centralnego punktu
- "symuluje" widok na Ziemię z kosmosu

## Kiedy stosować?

- na przykład do pokazania trasy samolotu



# Tworzenie map w R

## Sposób 1. (łatwy/średni)

użycie istniejących pakietów:

- statyczne wizualizacje: ggplot2 + maps (na laboratoriach), tmap, ggmap
- interaktywne wizualizacje: leaflet
- pobieranie danych przestrzennych: naturalearth, rnaturalearthdata
- klasy przestrzenne: sf, terra

## Sposób 2. (trudny)

wykorzystanie plików shape (*shapefiles*):

- pliki grafiki wektorowej dla danych geoprzestrzennych
- daje dużo większą swobodę i możliwości
- przydatne źródło: <https://github.com/Robinlovelace/Creating-maps-in-R>

# Więcej informacji

- <https://view.e.economist.com/?qs=f23f9794d30266884e9b9ef47429adf7cbe494594408510d45e193cdf3db52ac131bba428f03fa15eca8ebbe7f5e5d3955f1606c004bb72fc9f007f81a7ca1b6b5fb854d3cdca924755652b0fa24bffc>
- <https://help.flourish.studio/article/258-when-to-use-which-map-projection>
- <https://bookdown.org/mcwimberly/qdswr-book/mapping-with-ggplot2.html>
- <https://geocompr.robinlovelace.net/adv-map.html>

# Dashboards

“Dashboards are one of the most common use cases for data visualization, and their design and contexts of use are considerably different from exploratory visualization tools.”

# Czym jest dashboard?

*Definicja dashboardu ulega zmianom ...*

*"przeważnie wizualny ekran informacji, którego ludzie używają do szybkiego monitorowania bieżących warunków, które wymagają szybkiej reakcji, aby spełnić określona rolę"*

*"wizualne przedstawienie danych używanych do monitorowania warunków i/lub ułatwiania zrozumienia"*

wizualny



wizualna reprezentacja danych w postaci kafelkowego  
układu prostych wykresów i/lub dużych liczb

wizualny



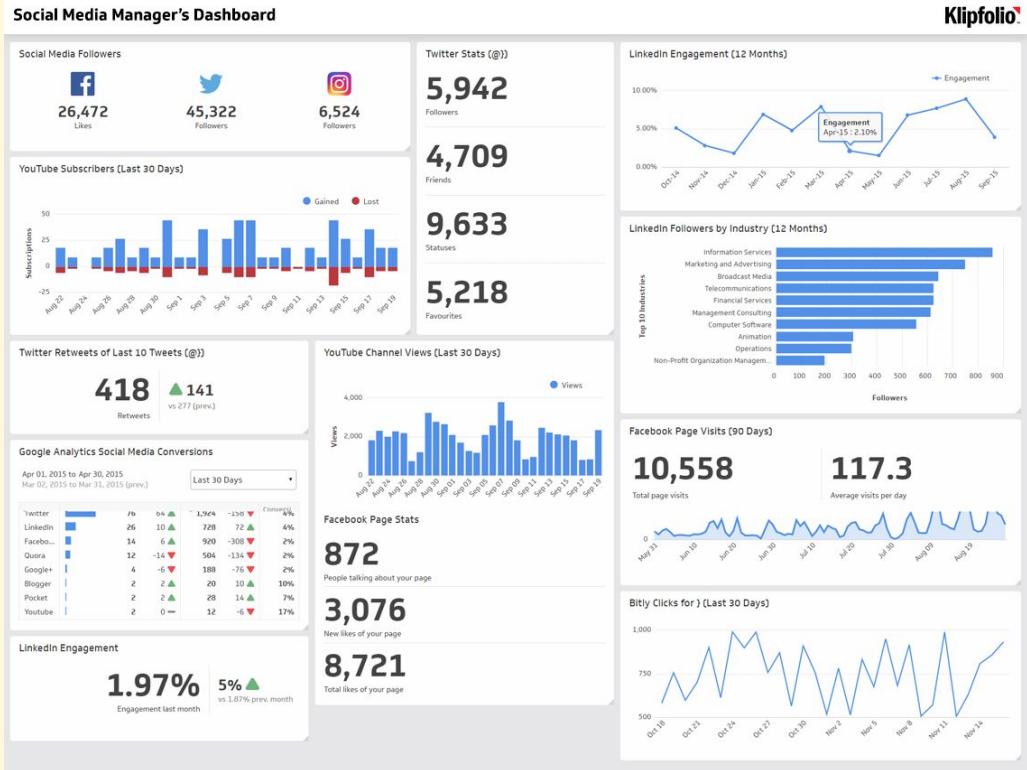
wizualna reprezentacja danych w postaci kafelkowego układu prostych wykresów i/lub dużych liczb

funkcjonalny

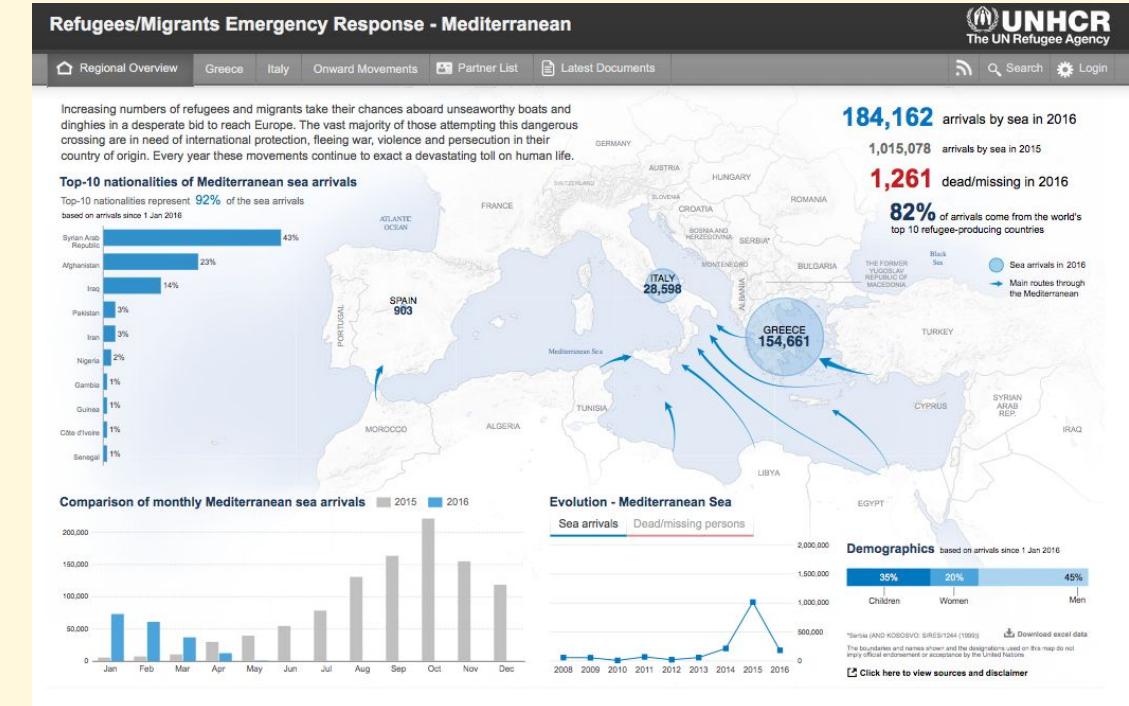


interaktywny ekran, który umożliwia monitorowanie w czasie rzeczywistym dynamicznie aktualizowanych danych

wizualny



funkcjonalny



# Zastosowanie

## → Wsparcie decyzyjne (strategiczne, taktyczne, operacyjne)

- ◆ pomoc organizacji w wyborze i ocenie strategii

"chcemy, aby użytkownicy z całego świata mogli kupować na naszej stronie"

- ◆ doskonalenie taktyki

"nasz CDN pomaga nam utrzymać globalną dostępność strony"

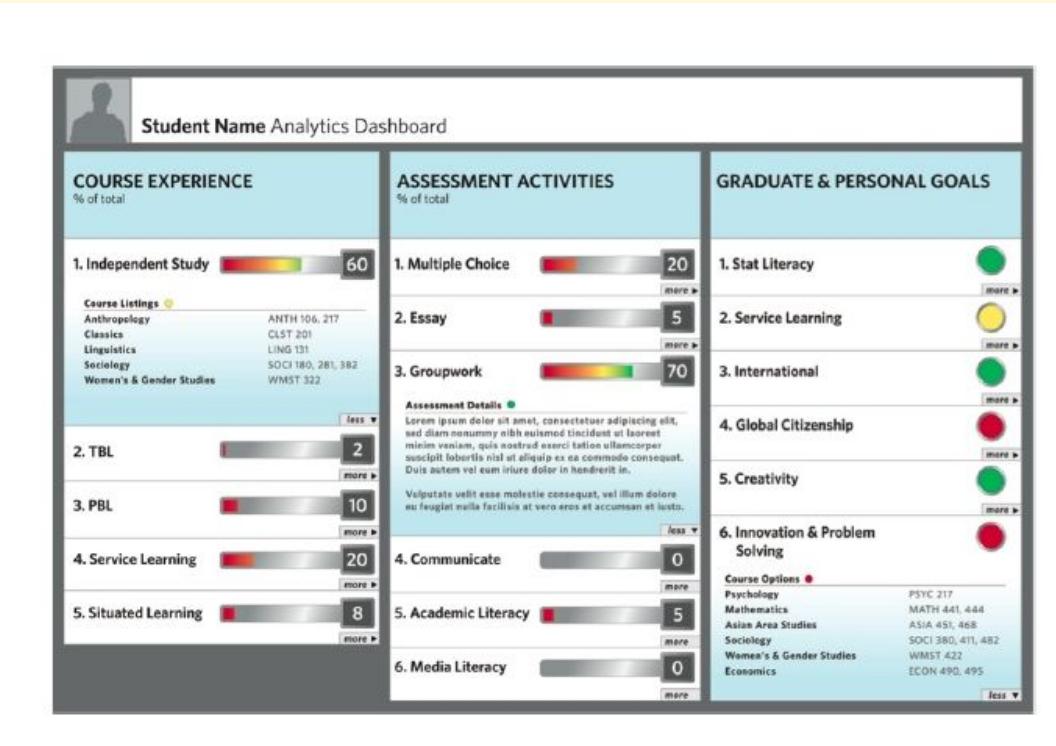
- ◆ ocena operacji

"użytkownicy w Seattle widzą powolną odpowiedź sieci"

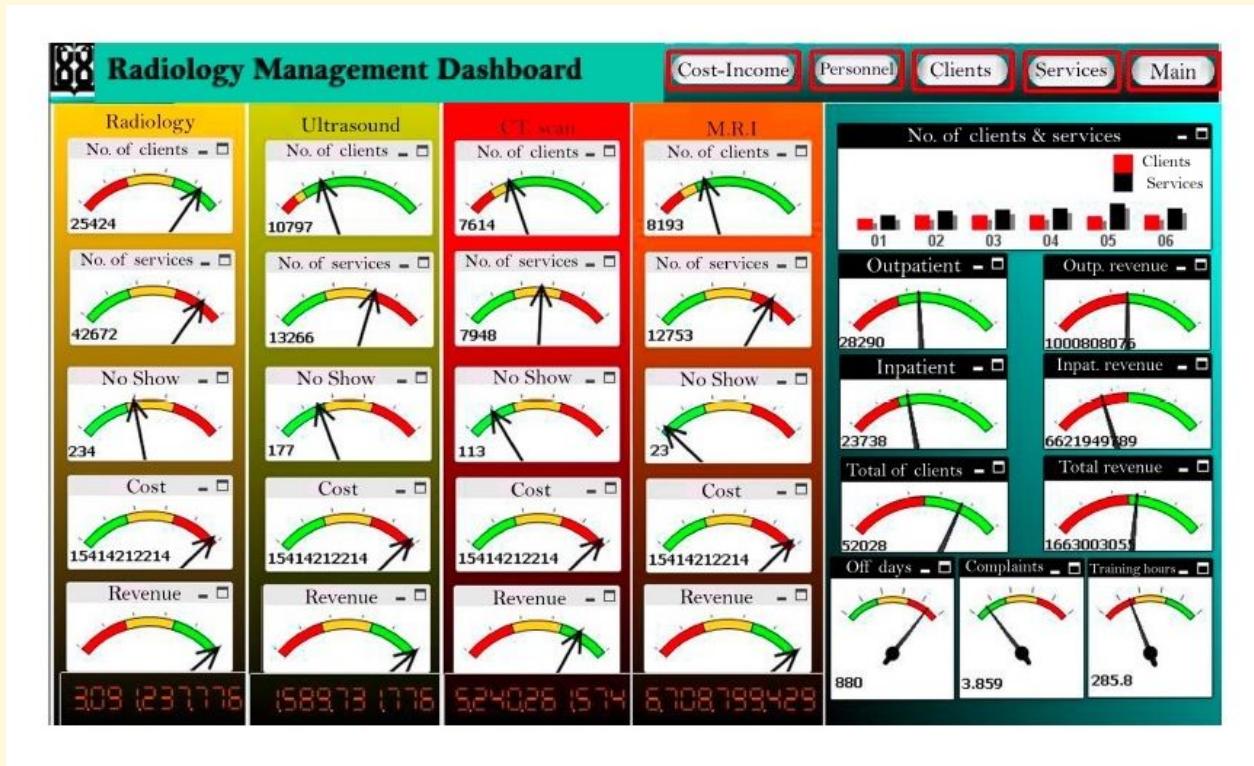
# Dashboard strategiczny



# Dashboard taktyczny



# Dashboard operacyjny



# Zastosowanie

- Komunikacja i uczenie
  - ◆ komunikacja lub edukowanie czytelnika, któremu może brakować kontekstu prezentowanych danych



# Odbiorca

- publiczny
  - ◆ ogólna konsumpcja, może opisywać dane istotne dla społeczeństwa

# Odbiorca

- publiczny
  - ◆ ogólna konsumpcja, może opisywać dane istotne dla społeczeństwa
- społeczny
  - ◆ szerokie zastosowanie dla wielu różnych osób w ramach danej struktury organizacyjnej, przy czym odbiorcy ci mają wspólny cel (np. wspieranie rentowności firmy)

# Odbiorca

- publiczny
  - ◆ ogólna konsumpcja, może opisywać dane istotne dla społeczeństwa
- społeczny
  - ◆ szerokie zastosowanie dla wielu różnych osób w ramach danej struktury organizacyjnej, przy czym odbiorcy ci mają wspólny cel (np. wspieranie rentowności firmy)
- organizacyjny
  - ◆ przypadki, w których dana osoba kontroluje dostęp do dashboardu dla wybranych przez siebie osób, identyfikując scenariusze wrażliwych danych lub analiz.

# Odbiorca

- publiczny
  - ◆ ogólna konsumpcja, może opisywać dane istotne dla społeczeństwa
- społeczny
  - ◆ szerokie zastosowanie dla wielu różnych osób w ramach danej struktury organizacyjnej, przy czym odbiorcy ci mają wspólny cel (np. wspieranie rentowności firmy)
- organizacyjny
  - ◆ przypadki, w których dana osoba kontroluje dostęp do dashboardu dla wybranych przez siebie osób, identyfikując scenariusze wrażliwych danych lub analiz.
- indywidualny
  - ◆ określają ilościowo daną osobę i zazwyczaj nie są udostępniane, z wyjątkiem zaufanych osób (np. lekarz lub planista finansowy)

# Wymagana umiejętności wizualizacji

Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

- niska
  - ◆ podstawowe typy wizualizacji, takie jak wykresy słupkowe i liniowe z panelami i agregacją

## A/B Test Dashboard

Klipfolio

### AB Test on Feature Y (Satisfaction)

#### Weighted Satisfaction Score

No Feature Y      Satisfaction Improvement  
**120.9**      **27%**      Feature Y  
153.8

#### Positive Satisfaction Rating

No Feature Y      Off vs On  
**78.4%**      **6.6%**      Feature Y  
85.0%  
See in Megapixel



#### AB Test Feature Y Funnels

Edit Widget → Save Widget

No Feature Y      Feature Y  
**61.4%**      **65.7%**  
+4.34% vs. Clicky Click

#### Edit Widget → > 6 month LTV

No Feature Y      Feature Y  
**44.0%**      **42.2%**  
-1.85% vs. No Feature Y

#### Edit Widget in Trial → Purchase

No Feature Y      Feature Y  
**3.54%**      **6.32%**  
2.78% vs. No Feature Y

### AB Test on Feature X (Satisfaction)

#### Weighted Satisfaction Score

Feature X Off      Satisfaction Improvement  
**147.3**      **5%**      Feature X On  
154.4

#### Positive Satisfaction Rating

Feature X Off      Off vs On  
**85.2%**      **2.3%**      Feature X On  
87.5%  
See in Megapixel



#### AB Test Feature X Funnels

Edit Widget → Save Widget

Feature X Off      Feature X On  
**62.3%**      **60.0%**  
-2.24% vs. No Feature X

#### Edit Widget → > 6 month LTV

Feature X Off      Feature X On  
**43.2%**      **40.3%**  
-2.91% vs. No Feature X

#### Edit Widget in trial → Purchase

Feature X Off      Feature X On  
**7.99%**      **5.05%**  
-2.94% vs. No Feature X

# Wymagana umiejętności wizualizacji

Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

- niska
  - ◆ podstawowe typy wizualizacji, takie jak wykresy słupkowe i liniowe z panelami i agregacją
- średnia
  - ◆ połączone podwójne osie, wykresy rozproszenia, miary skumulowane i mapy ciepła

# 50 Years Of Crime

Exploring five decades of crime in the United States  
(Source: Uniform Crime Reporting Website – Click to View)

Hover anywhere on charts to explore!

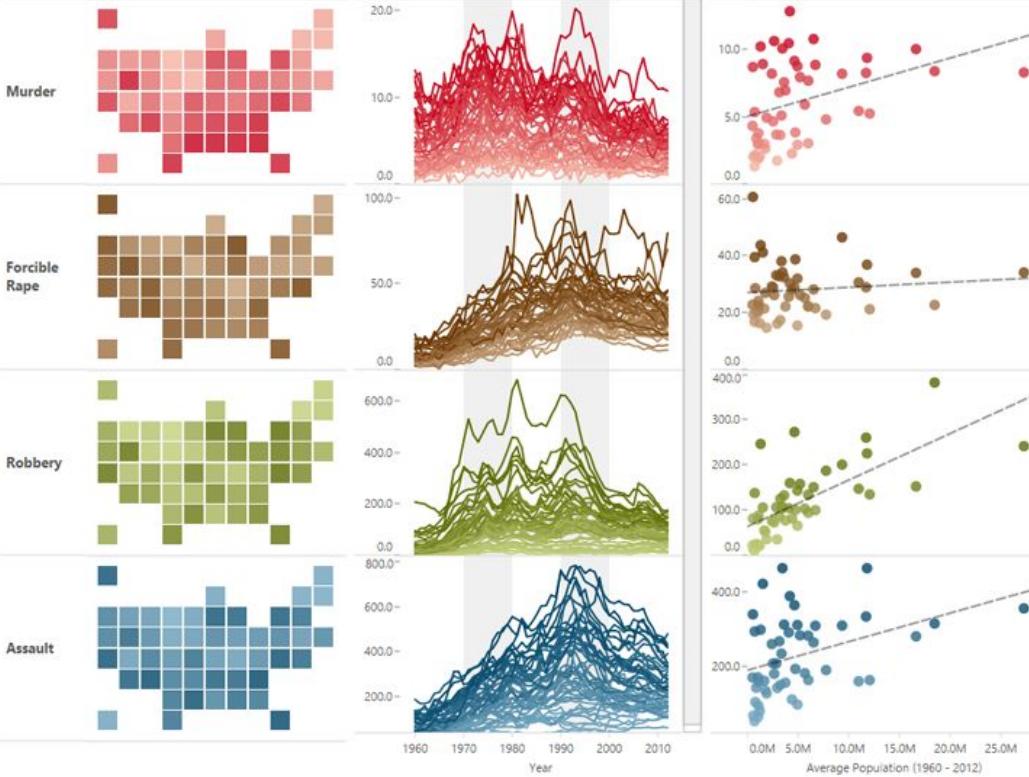
US Heat Map  
By Crime Rate

Crime Rate Trend  
By State (1960 – 2012)

Crime Rate vs  
State Population

Pick Crime Category  
 Property Crime  
 Violent Crime

All Crime Rates per 100,000 Population



# Wymagana umiejętności wizualizacji

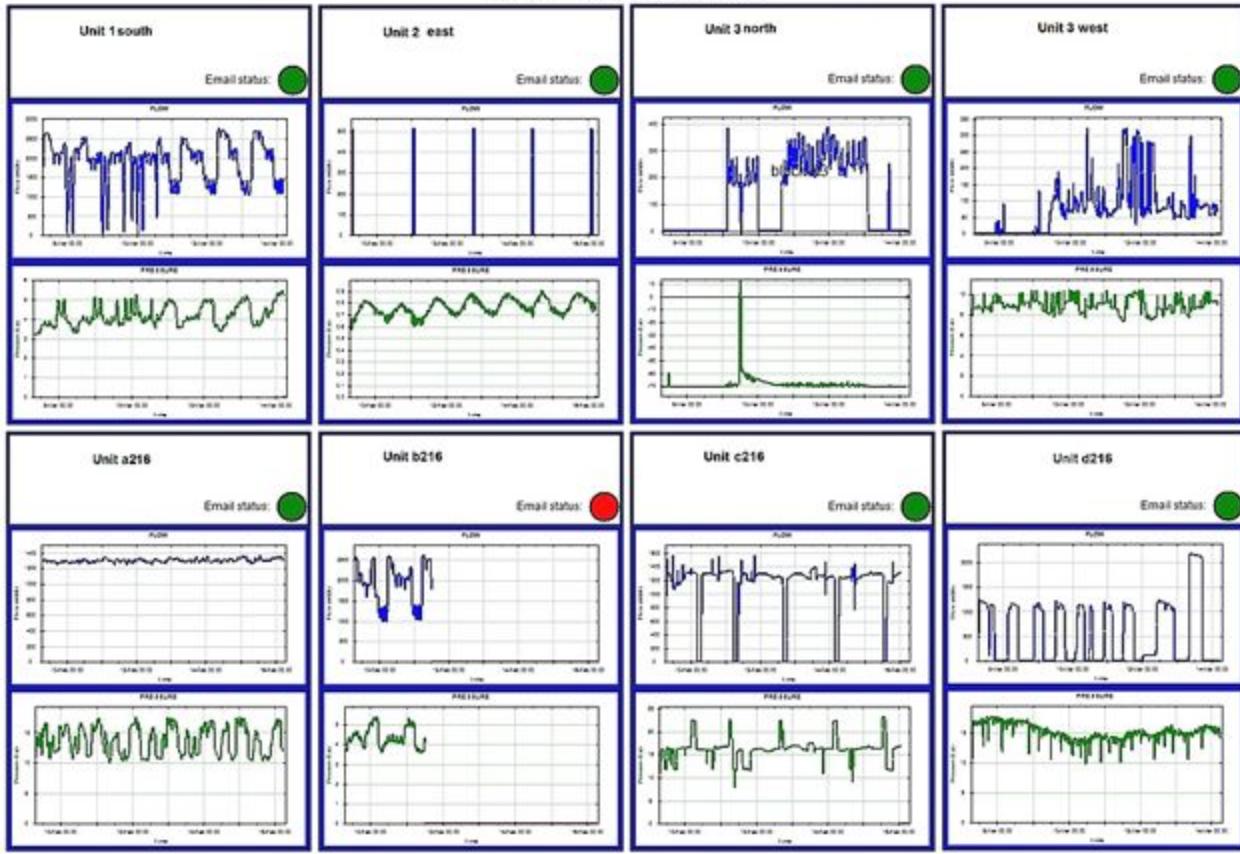
Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

- niska
  - ◆ podstawowe typy wizualizacji, takie jak wykresy słupkowe i liniowe z panelami i agregacją
- średnia
  - ◆ szerokie zastosowanie dla wielu różnych osób w ramach danej struktury organizacyjnej, przy czym odbiorcy ci mają wspólny cel (np. wspieranie rentowności firmy)
- wysoka
  - ◆ radar, treemapa i wizualizacje sieciowe, słupki błędów lub przedziały, połączone wykresy rozproszenia lub inne niestandardowe wizualizacje



A. Sarikaya, M. Correll, L. Bartram, M. Tory and D. Fisher, "What Do We Talk About When We Talk About Dashboards?", in *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, no. 1, pp. 682-692, Jan. 2019, doi: 10.1109/TVCG.2018.2864903.

WaMSS Dashboard Dashboard Graphic  
From 2016-03-07 07:00 to 2016-03-14 07:00



# Funkcje wizualne i interaktywność

- budowa i skład
  - ◆ modyfikacja konstrukcji i kompozycji widoków
  - ◆ elastyczność, która pozwala na dostosowanie rozmieszczenia widoków, reprezentacja wizualna i ich modyfikacja

# Funkcje wizualne i interaktywność

- budowa i skład
  - ◆ modyfikacja konstrukcji i kompozycji widoków
  - ◆ elastyczność, która pozwala na dostosowanie rozmieszczenia widoków, reprezentacja wizualna i ich modyfikacja
- kilka stron (multipage)
  - ◆ pozwalają na przełączanie się pomiędzy stronami, które mogą zawierać wizualizacje odnoszące się do różnych elementów procesu decyzyjnego lub pomagające dostarczyć niezbędny kontekst

# Funkcje wizualne i interaktywność

- budowa i skład
  - ◆ modyfikacja konstrukcji i kompozycji widoków
  - ◆ elastyczność, która pozwala na dostosowanie rozmieszczenia widoków, reprezentacja wizualna i ich modyfikacja
- kilka stron (multipage)
  - ◆ pozwalają na przełączanie się pomiędzy stronami, które mogą zawierać wizualizacje odnoszące się do różnych elementów procesu decyzyjnego lub pomagające dostarczyć niezbędny kontekst
- interaktywny interfejs
  - ◆ wybór podzbioru danych za pomocą slicerów i filtrów
  - ◆ obecności typowych komponentów interaktywnych

# Funkcje wizualne i interaktywność

- budowa i skład
  - ◆ modyfikacja konstrukcji i kompozycji widoków
  - ◆ elastyczność, która pozwala na dostosowanie rozmieszczenia widoków, reprezentacja wizualna i ich modyfikacja
- kilka stron (multipage)
  - ◆ pozwalają na przełączanie się pomiędzy stronami, które mogą zawierać wizualizacje odnoszące się do różnych elementów procesu decyzyjnego lub pomagające dostarczyć niezbędny kontekst
- interaktywny interfejs
  - ◆ wybór podzbioru danych za pomocą slicerów i filtrów
  - ◆ obecności typowych komponentów interaktywnych
- podświetlanie i adnotacje
  - ◆ opatrywanie dashboardów adnotacjami w celu ich późniejszego przeanalizowania

# Funkcje wizualne i interaktywność

- budowa i skład
  - ◆ modyfikacja konstrukcji i kompozycji widoków
  - ◆ elastyczność, która pozwala na dostosowanie rozmieszczenia widoków, reprezentacja wizualna i ich modyfikacja
- kilka stron (multipage)
  - ◆ pozwalają na przełączanie się pomiędzy stronami, które mogą zawierać wizualizacje odnoszące się do różnych elementów procesu decyzyjnego lub pomagające dostarczyć niezbędny kontekst
- interaktywny interfejs
  - ◆ wybór podzbioru danych za pomocą slicerów i filtrów
  - ◆ obecności typowych komponentów interaktywnych
- podświetlanie i adnotacje
  - ◆ opatrywanie dashboardów adnotacjami w celu ich późniejszego przeanalizowania
- modyfikacja danych
  - ◆ analizy "co by było gdyby", modelowanie i wprowadzanie danych mogą być przykładami zapisywania do źródła danych
  - ◆ dashboard inteligentnego domu, który umożliwia wyłączenie światła lub regulację termostatu



# Najważniejsze zasady tworzenia dashboardów

# **Najważniejsze zasady tworzenia dashboardów**

1. Przeanalizuj swoich odbiorców

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst
5. Nie próbuj umieszczać wszystkich informacji na jednej stronie

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst
5. Nie próbuj umieszczać wszystkich informacji na jednej stronie
6. Dobierz odpowiednie wykresy.

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst
5. Nie próbuj umieszczać wszystkich informacji na jednej stronie
6. Dobierz odpowiednie wykresy.
7. Starannie dobieraj układ graficzny

# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst
5. Nie próbuj umieszczać wszystkich informacji na jednej stronie
6. Dobierz odpowiednie wykresy.
7. Starannie dobieraj układ graficzny
8. Bądź ostrożny z kolorami - wybierz kilka i trzymaj się ich

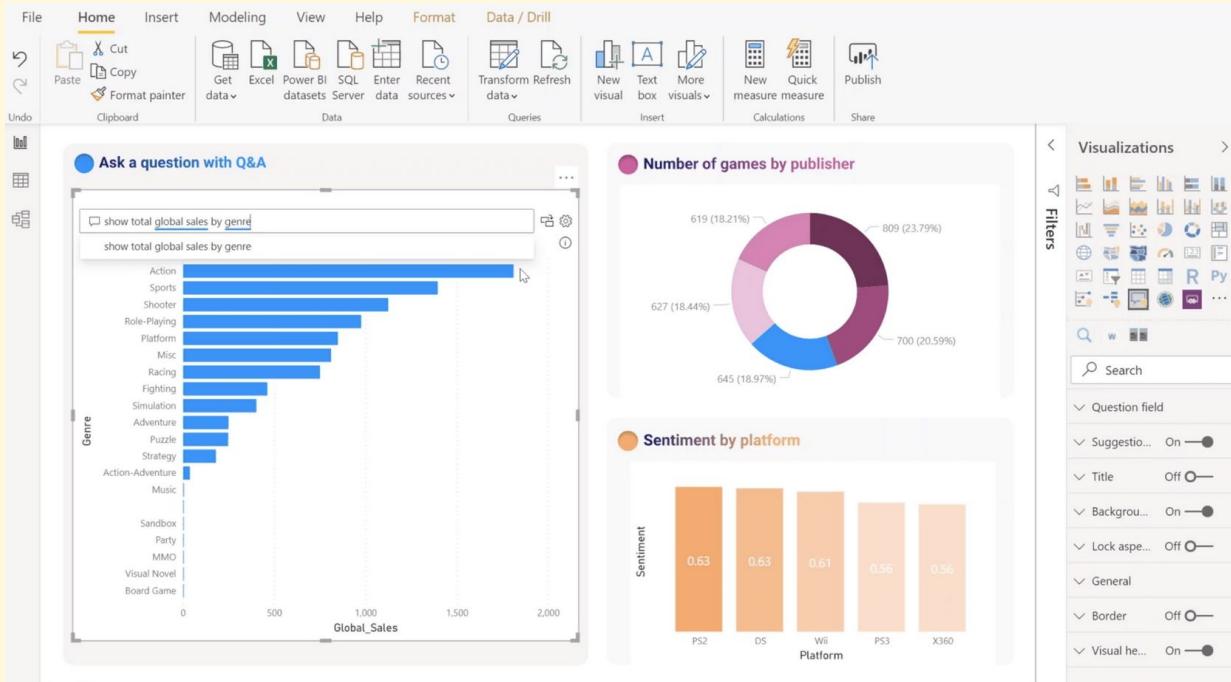
# Najważniejsze zasady tworzenia dashboardów

1. Przeanalizuj swoich odbiorców
2. Określ swoje cele
3. Opowiedz historię swoimi danymi
4. Zapewnij kontekst
5. Nie próbuj umieszczać wszystkich informacji na jednej stronie
6. Dobierz odpowiednie wykresy.
7. Starannie dobieraj układ graficzny
8. Bądź ostrożny z kolorami - wybierz kilka i trzymaj się ich
9. Animacje

# Narzędzia

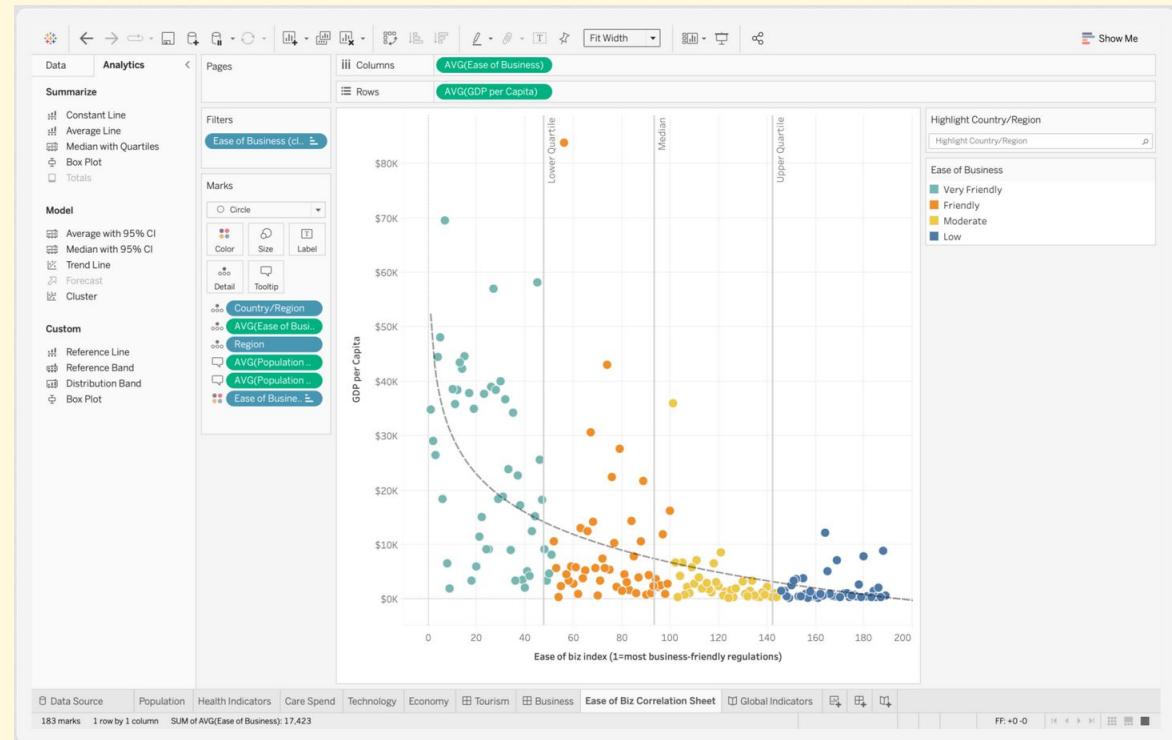
# Power BI

Łączy się z danymi i wizualizuje je za pomocą ujednoliconej, skalowej platformy do samoobsługowej i korporacyjnej analizy biznesowej (BI), która jest łatwa w użyciu i pozwala uzyskać dokładniejszy wgląd w dane.

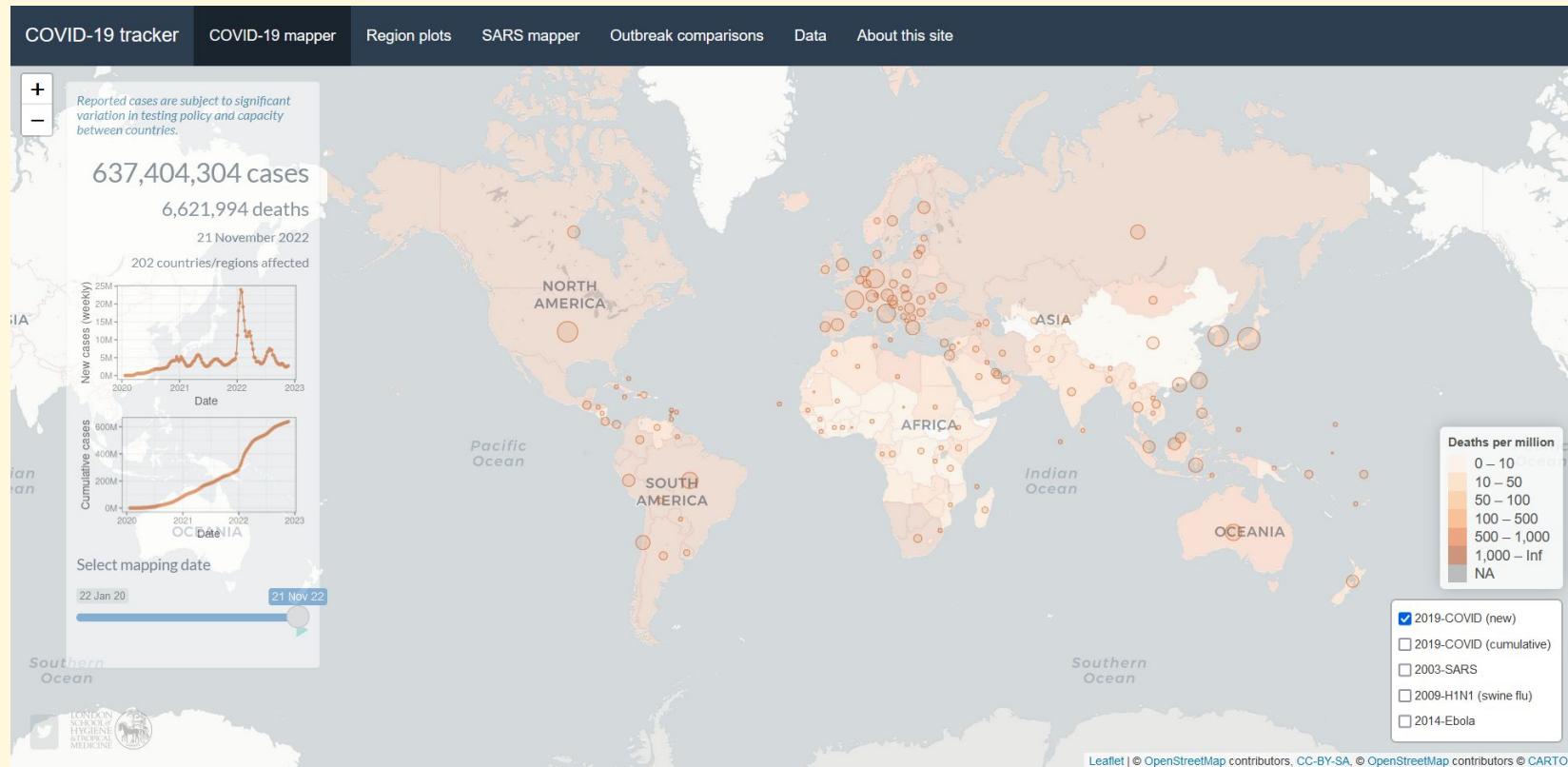


# Tableau

Narzędzie do wizualnej, opartej na podejściu Data Discovery analityki danych. Narzędzie BI błyskawicznie integruje się z dowolnymi źródłami danych i umożliwia ich analizę nie tylko specjalistom, ale każdemu, kto chce stworzyć przejrzystą wizualizację w oparciu o swoje dane.



# R Shiny



# Dash Python

