

Wstęp do Eksploracji Danych

Politechnika Warszawska
Anna Kozak

Kolory i skale

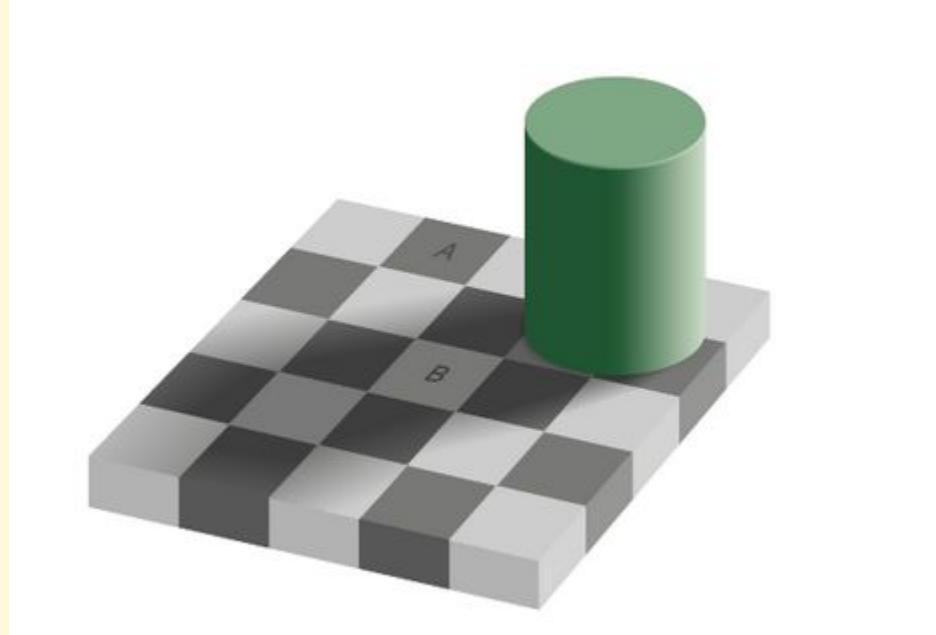
Kolory

Dlaczego dobór koloru jest ważny?

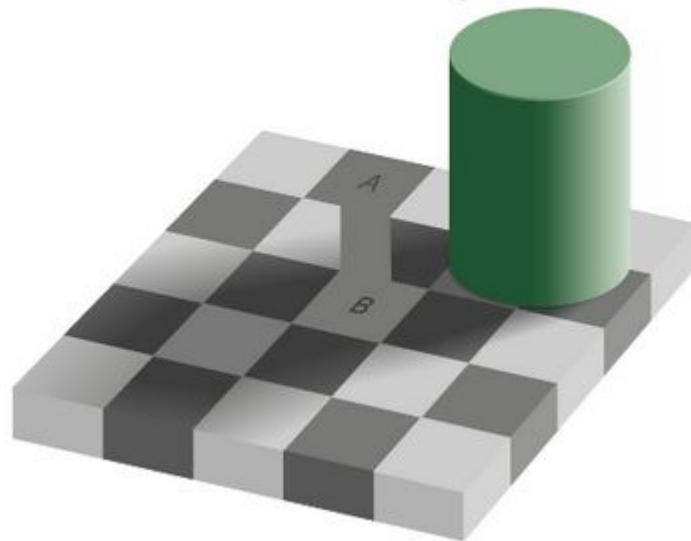
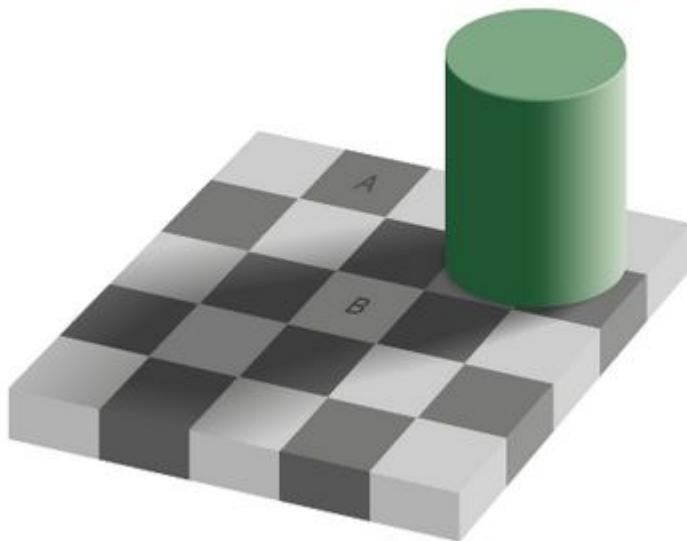
Podświadomie przypisujemy barwom znaczenia.

Postrzeganie barw różni się w zależności od oświetlenia jakości wydruku, ekranu lub projektora.

Wszystko jest względne



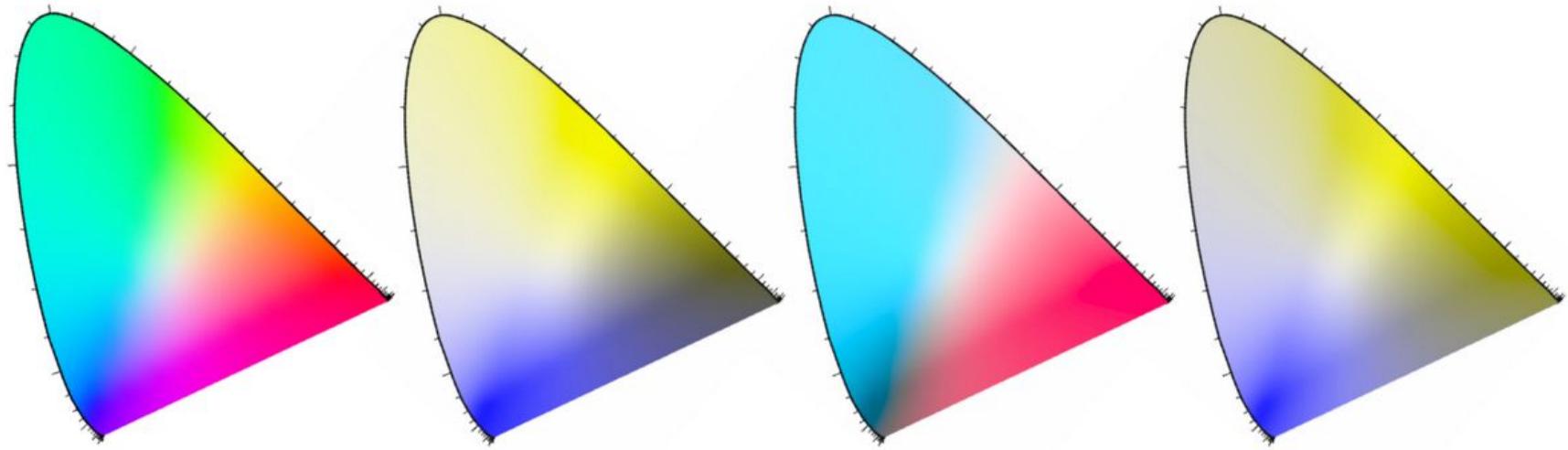
Wszystko jest względne



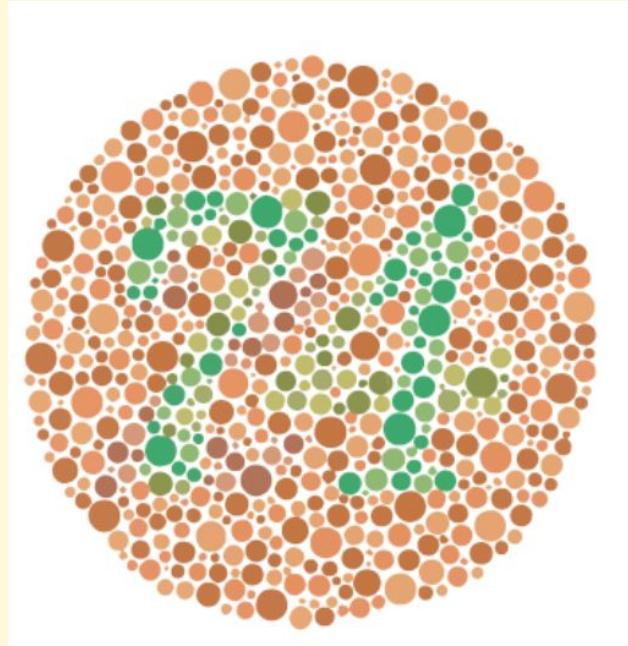
Wszystko jest względne



Zaburzenia w postrzeganiu barw



Zaburzenia w postrzeganiu barw



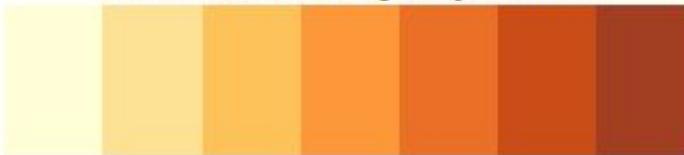
Zaburzenia w postrzeganiu barw

Jeżeli w naszej wizualizacji kolory pełnią kluczową funkcję to warto upewnić się, że przynajmniej osoby z typowymi dysfunkcjami widzenia kolorów będą w stanie odczytać informacje.

<https://projects.susielu.com/viz-palette>

Skale kolorystyczne

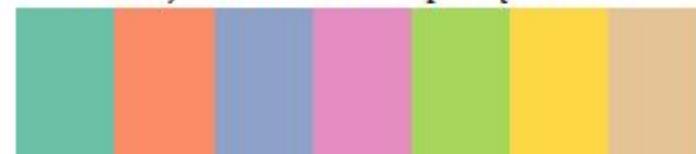
Skala ilościowa / uporządkowana



Skala uporządkowana rozbieżna



Skala jakościowa / nieuporządkowana

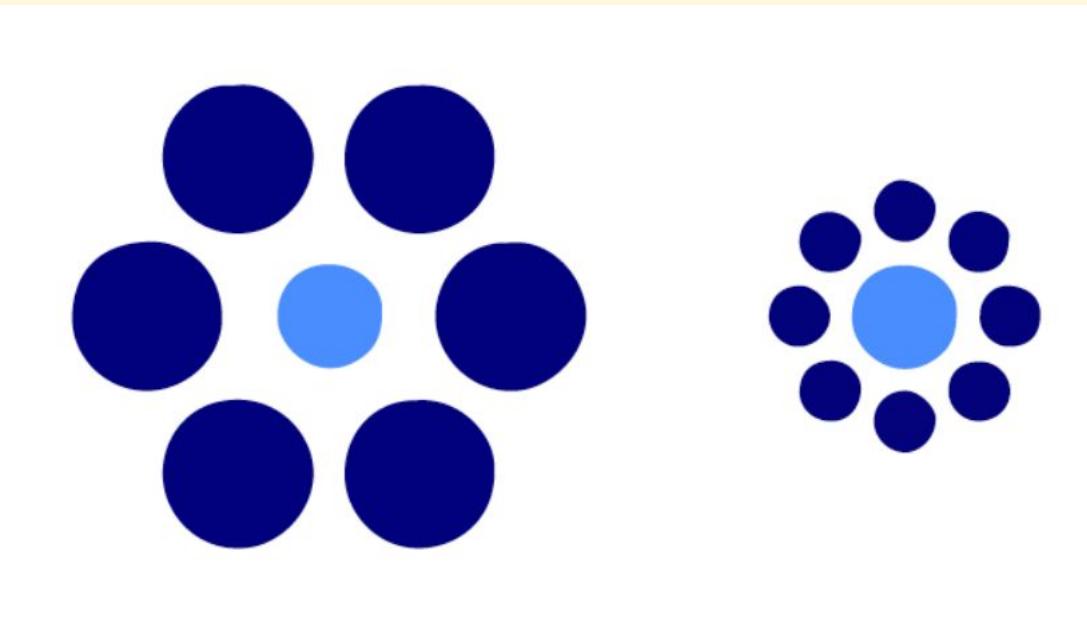


Skale

Pasma Macha



Iluzja Titchenera - Zniekształcenie postrzegania wielkości



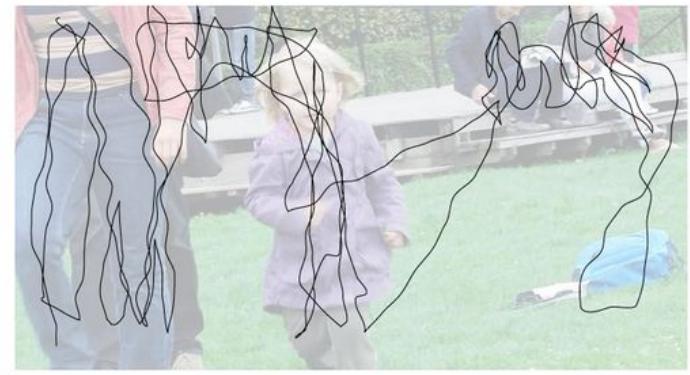
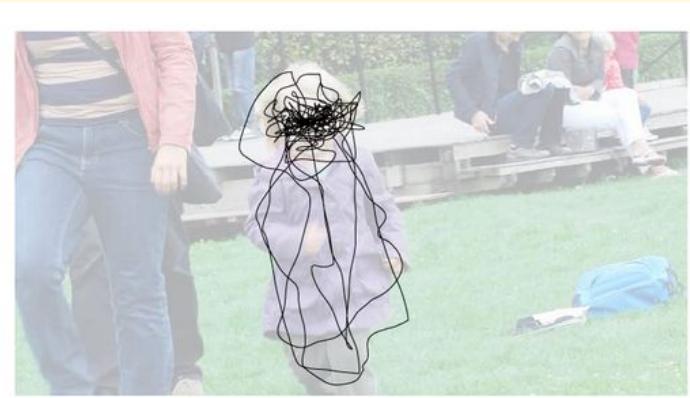
Obszar widzenia



Obszar widzenia



Fiksacje i sakady



Jak wiedza o strategiach przeglądania obrazu może pomóc w przygotowaniu lepszej prezentacji danych?

- Im więcej informacji, tym ważniejsze jest, by informacja była przedstawiana warstwowo.
- Jeżeli wykresowi towarzyszy słowna prezentacja, to warto powiedzieć, gdzie są interesujące elementy, ułatwi to ich lokalizację.
- Tytuł wykresu bardzo pomaga, ponieważ wstępnie informuje percepcję, czego wykres dotyczy i ułatwia wybór elementów wykresu do obserwacji.
- Wiele też zależy od tego, ile czasu odbiorca poświęci na analizę wykresu.

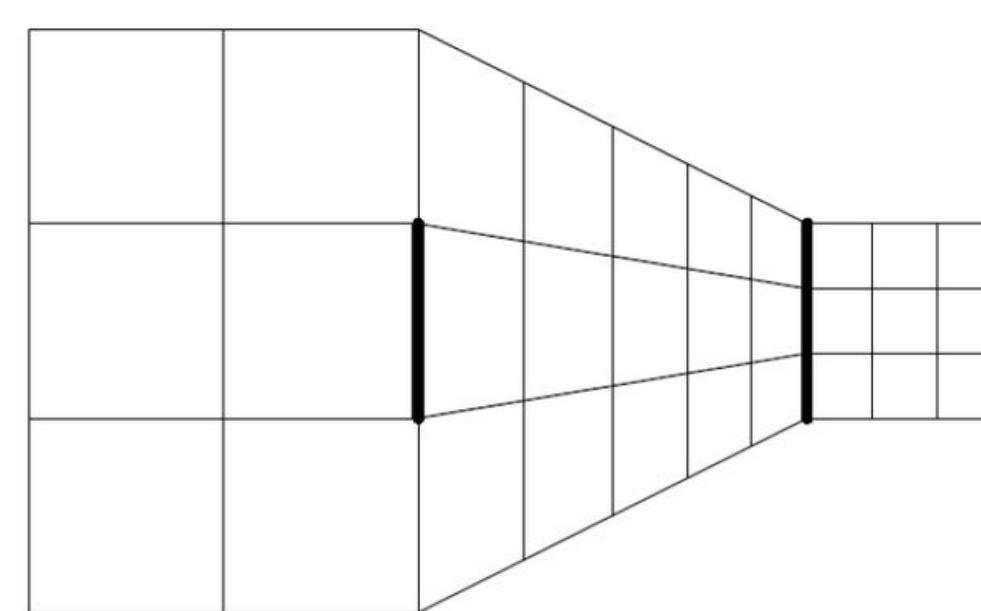
Iluzje

Pierwsza grupa iluzji dotyczy widzenia tego, czego nie ma. Im bardziej skomplikowany wykres, tym większa szansa, że coś przypadkowego zostanie uznane za ten “istotny” wzorzec.

Widzenie tego, czego nie ma



Pseudo perspektywa



Pseudo perspektywa

Jeżeli na wykresie znajdzie się cokolwiek, co może sugerować perspektywę, to zostanie dostrzeżona przez mózg. Automatycznie wpłynie to na zniekształconą ocenę wielkości.

Dlatego wszelkim trójwymiarowym wykresom, czy to kołowym, słupkowym czy piramidowym powinniśmy zdecydowanie powiedzieć: **NIE**.

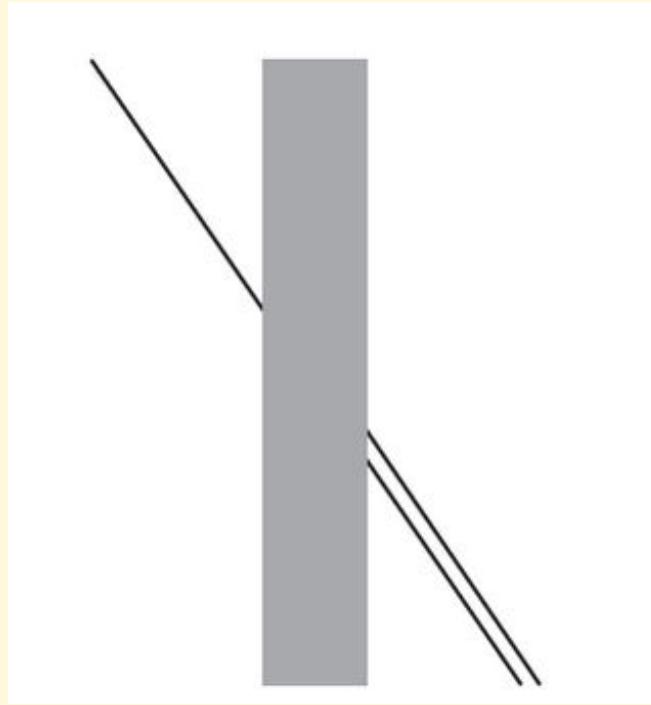
Problemy z kątami

Pewnych charakterystyk mózg nie jest w stanie dobrze ocenić. Dobrym przykładem są kąty.

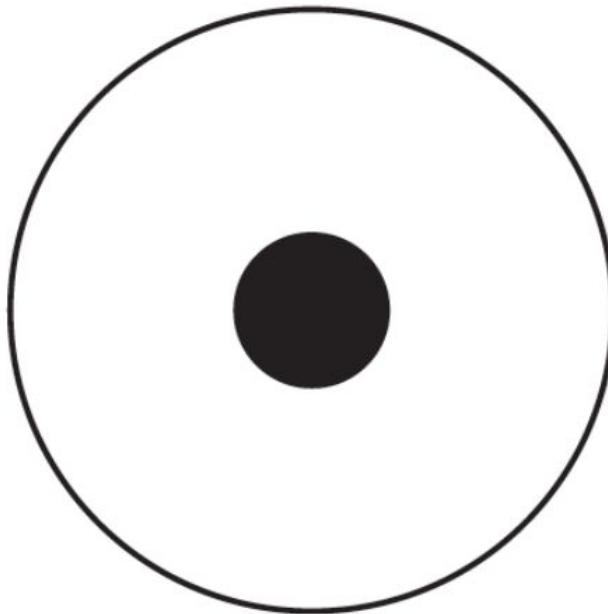
Ludzki mózg jest w stanie z dużą dokładnością ocenić, czy kąt jest bliski kątowi prostemu, ale ma duże problemy z oceną wielkości kątów ostrych i rozwartych.

Mózg ma skłonność do zawyżania wielkości kątów ostrych i zaniżania kątów rozwartych.

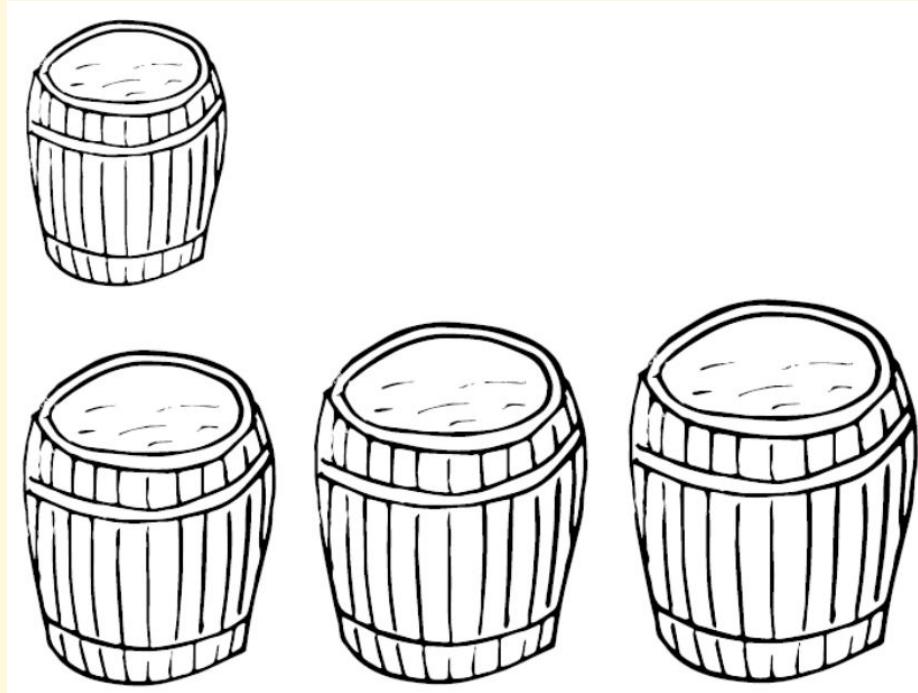
Problemy z kątami



Co łączy dentystę i dietetyka?



Ocena wielkości



Które województwo jest większe?



Które województwo jest większe?



warmińsko-mazurskie

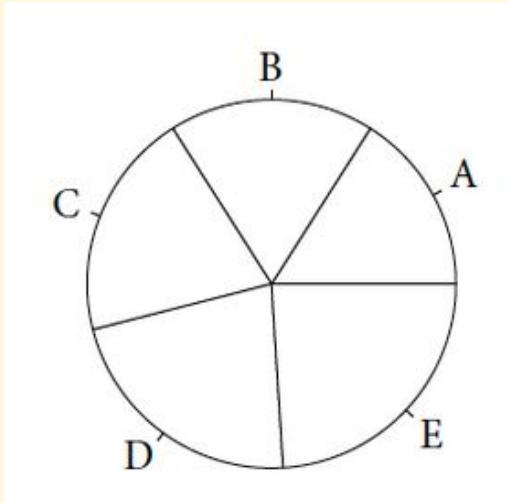
lubelskie

Hierarchia odczytywania charakterystyk

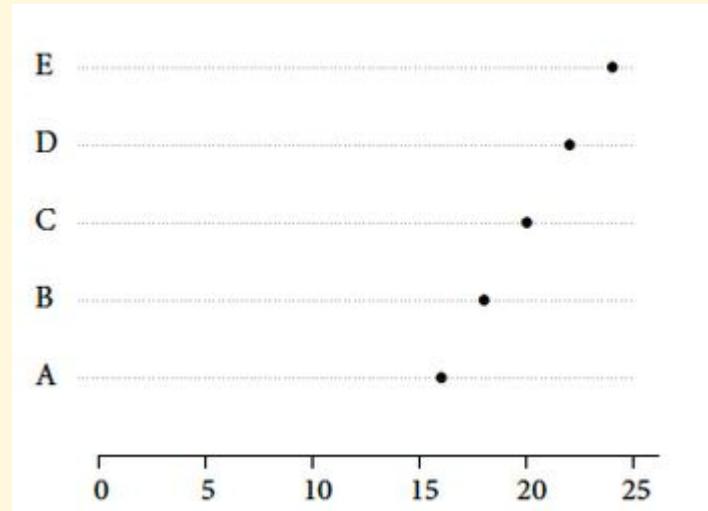
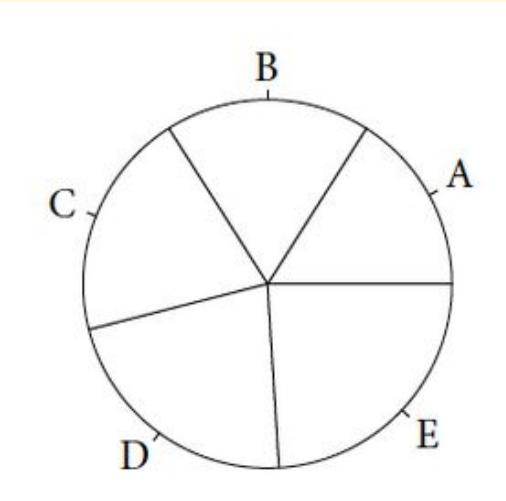
1. pozycje obiektów rozmieszczonych wzdłuż wspólnej skali (przykładowo wykres punktowy),
2. pozycje obiektów wzdłuż takiej samej, ale nie wspólnej skali (przykładowo sąsiadujące wykresy punktowe),
3. długości odcinków rozmieszczonych wzdłuż wspólnej skali,
4. długości odcinków wzdłuż takiej samej, ale nie wspólnej skali (o różnych punktach zaczepienia),
5. wielkości kątów i nachylenia (przy ocenie tempa wzrostu w wykresach liniowych),
6. powierzchnie,
7. objętości, gęstości, natężenia koloru,
8. sama barwa koloru.

Na bazie The Visual Decoding of Quantitative Information on Statistical Graphs [Journal of the Royal Statistical Society Series A, 150:192–229, 1987]

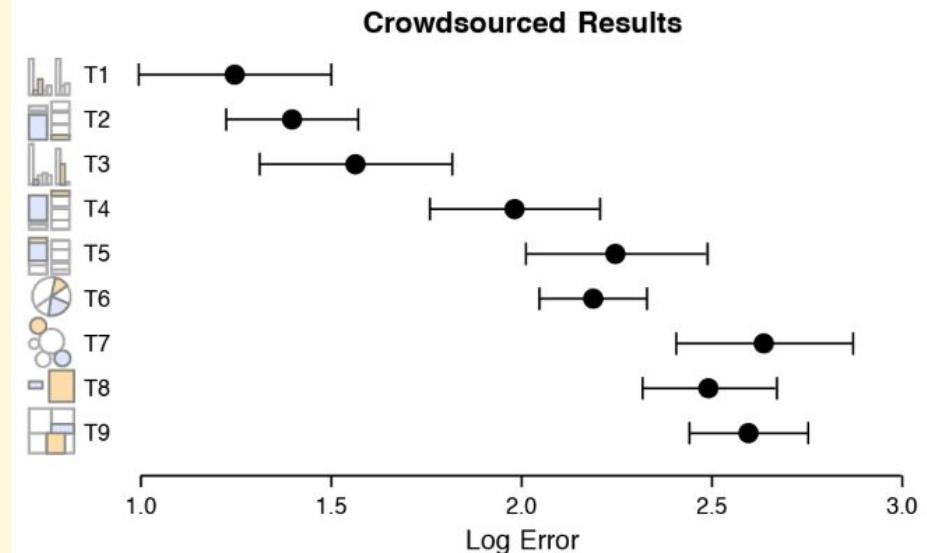
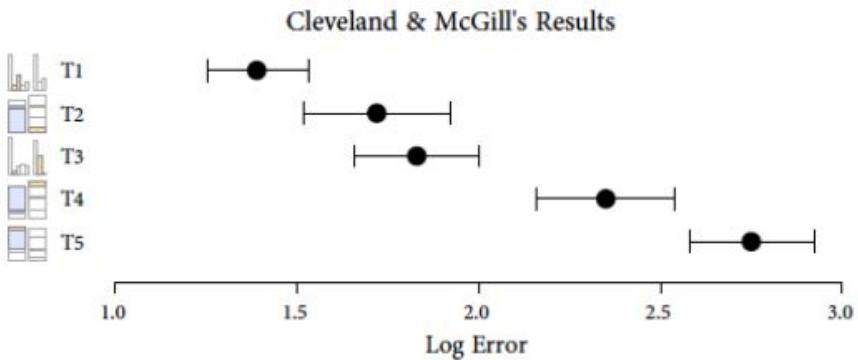
Porównanie



Badania nad percepcją



Badania nad percepcją



T1 – wysokości sąsiednich pasków,

T2 – pola prostokątów o wspólnej podstawie,

T3 – wysokości odległych pasków,

T4, T5 – pola prostokątów bez wspólnej podstawy,

T6 – pola wycinków koła,

T7 – pola kół,

T8, T9 – pola niewyrównanych prostokątów.

Do poczytania

Percepcja danych:

<http://www.biecek.pl/Eseje/indexDane.html>

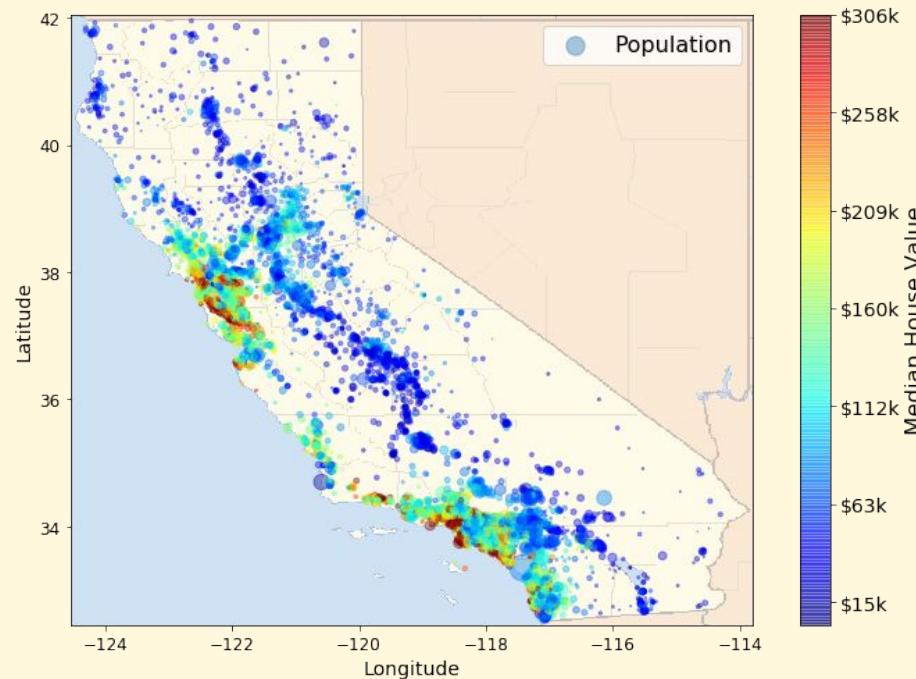
Percepcja kolorów:

<http://www.biecek.pl/Eseje/indexKolory.html>

Mapy - czy to aż tak skomplikowane?

Motywacja

Mapy są świetnym sposobem do reprezentacji danych przestrzennych.



Historia

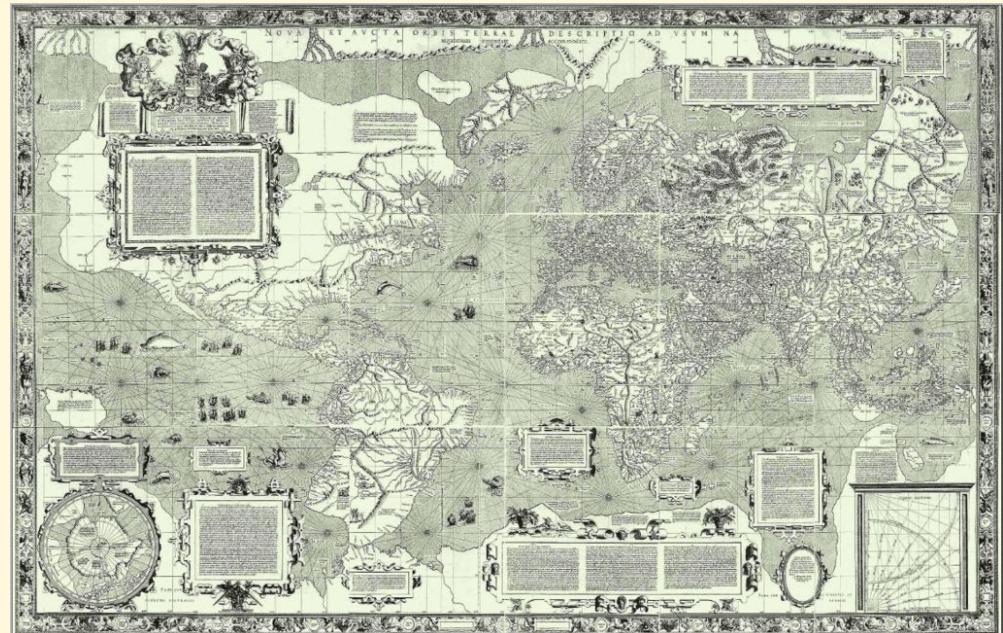
- Jedne z pierwszych tworzonych przez ludzi wizualizacji to wizualizacje kartograficzne - mapy.
- Tworzone je dla różnych potrzeb:
 - nawigacja
 - planowanie rozwoju miasta
 - planowanie sadzenia upraw



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

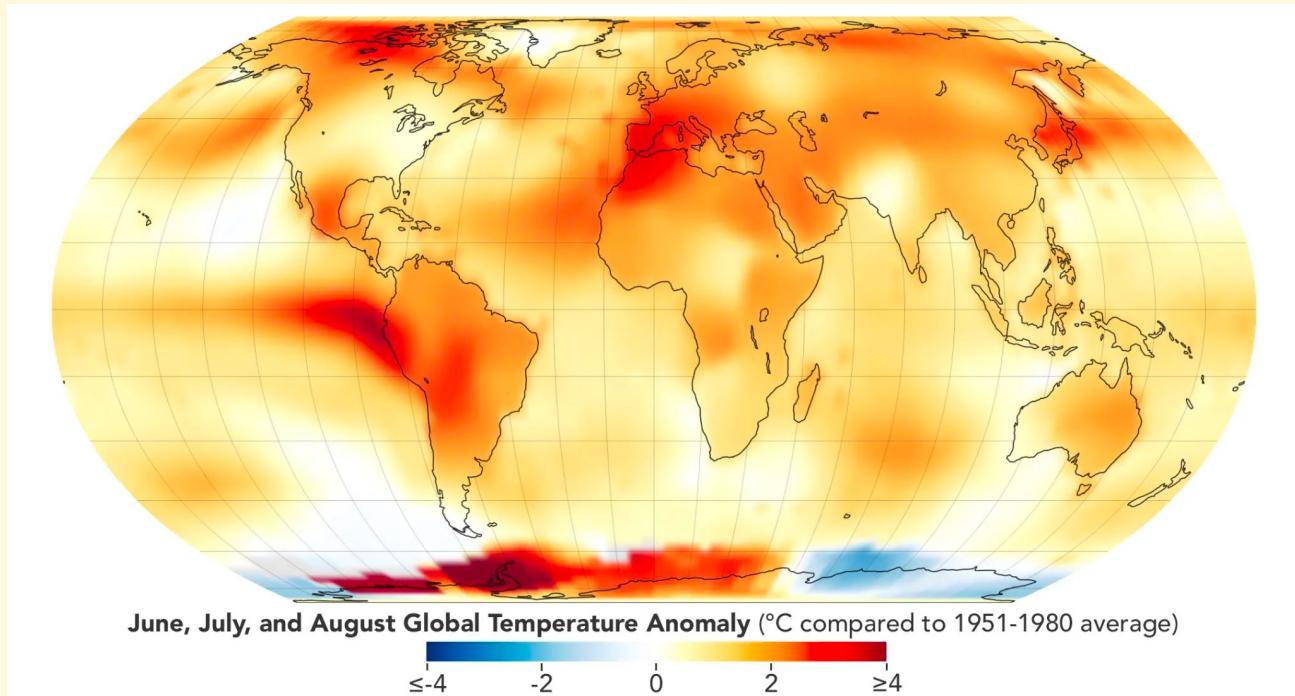


mapa świata Mercatora
(1569)

Teraźniejszość

Zastosowania geowizualizacji

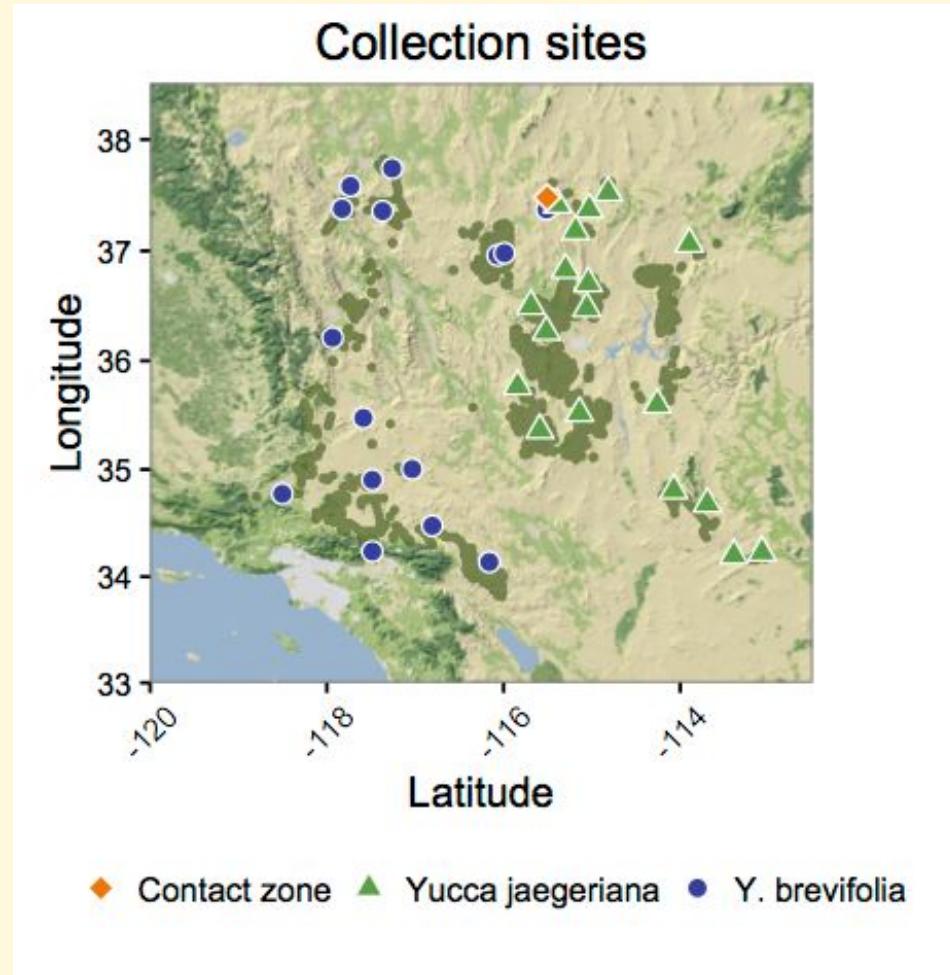
- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia



Teraźniejszość

Zastosowania geowizualizacji:

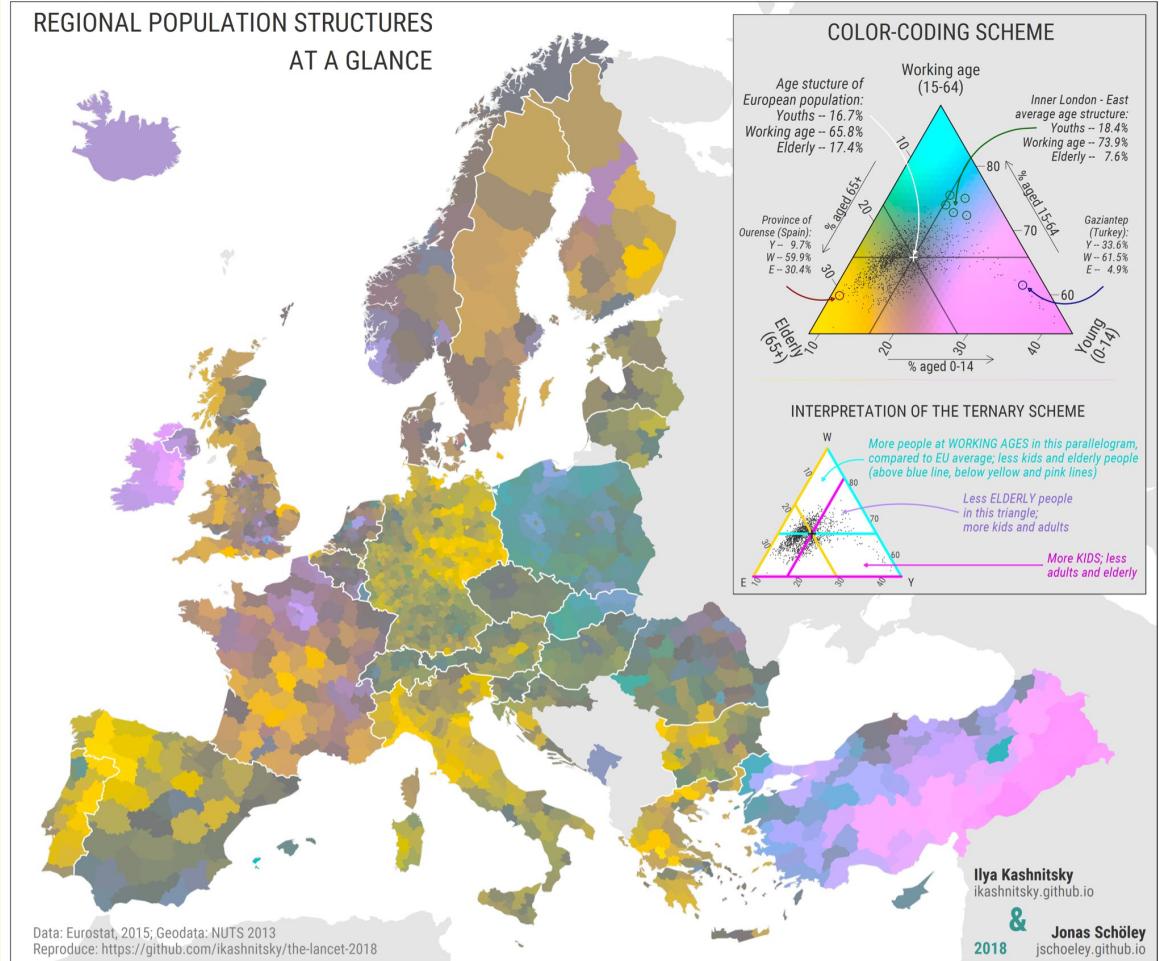
- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- transport
- historia



Teraźniejszość

Zastosowania geowizualizacji:

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

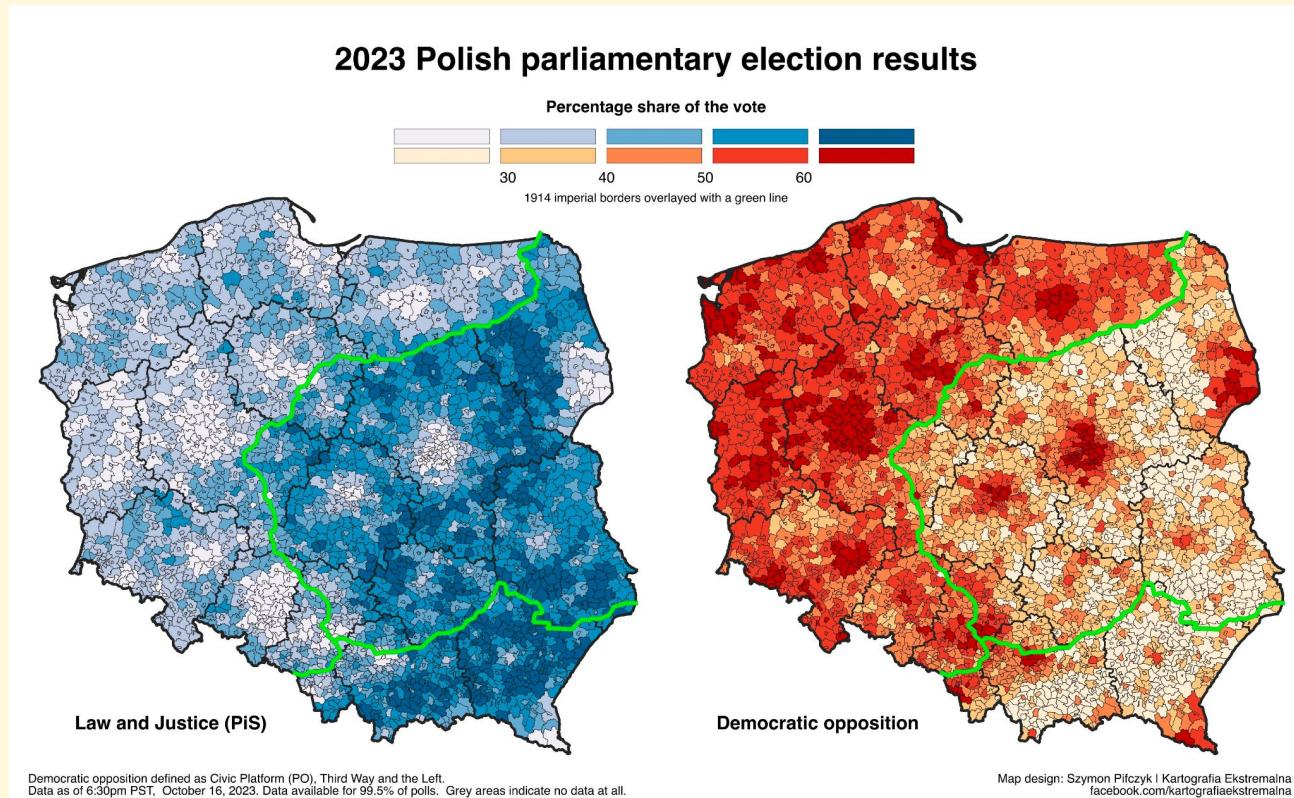


Kashnitsky, I., Schöley, J. Regional population structures at a glance.
The Lancet, 392(10143), 209–210 (2018).

Teraźniejszość

Zastosowania geowizualizacji:

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



Teraźniejszość

Zastosowania geowizualizacji:

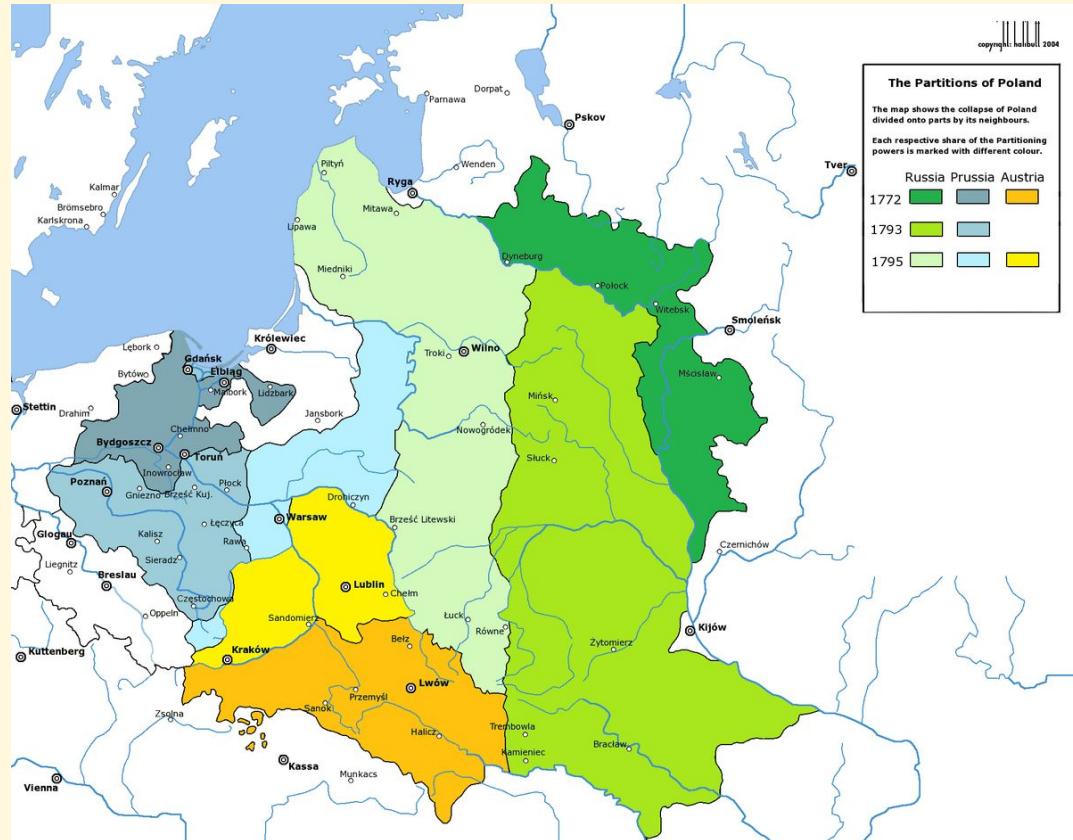
- klimatologia
- ekologia
- demografia
- ekonomia
- dziennikarstwo
- polityka
- **transport**
- historia



Teraźniejszość

Zastosowania geowizualizacji:

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



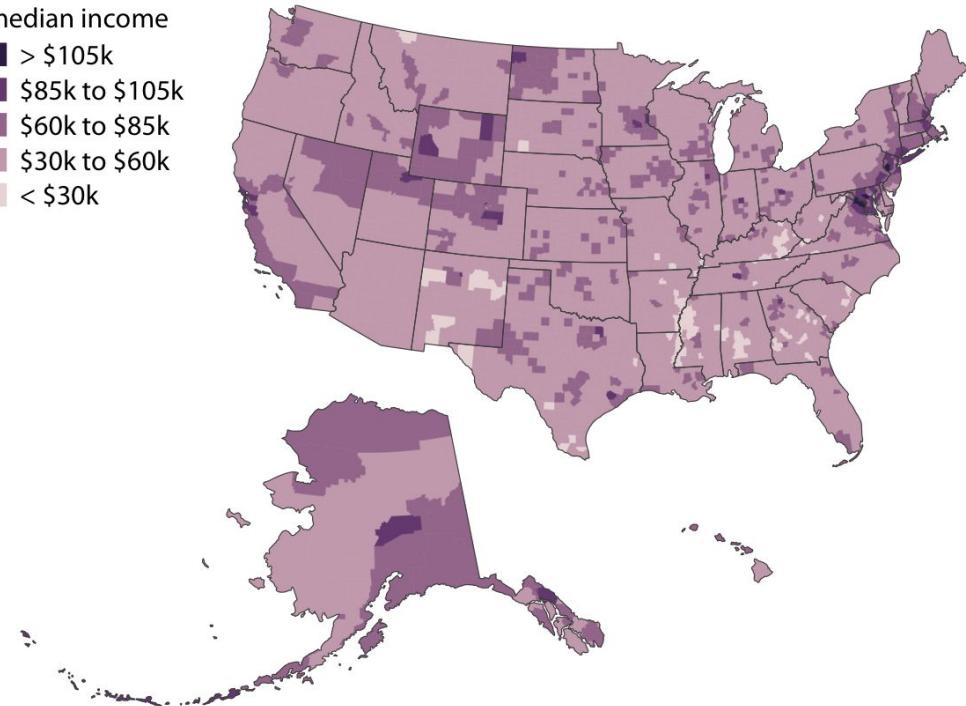
Metody prezentacji danych na mapach

KARTOGRAM (choropleth)

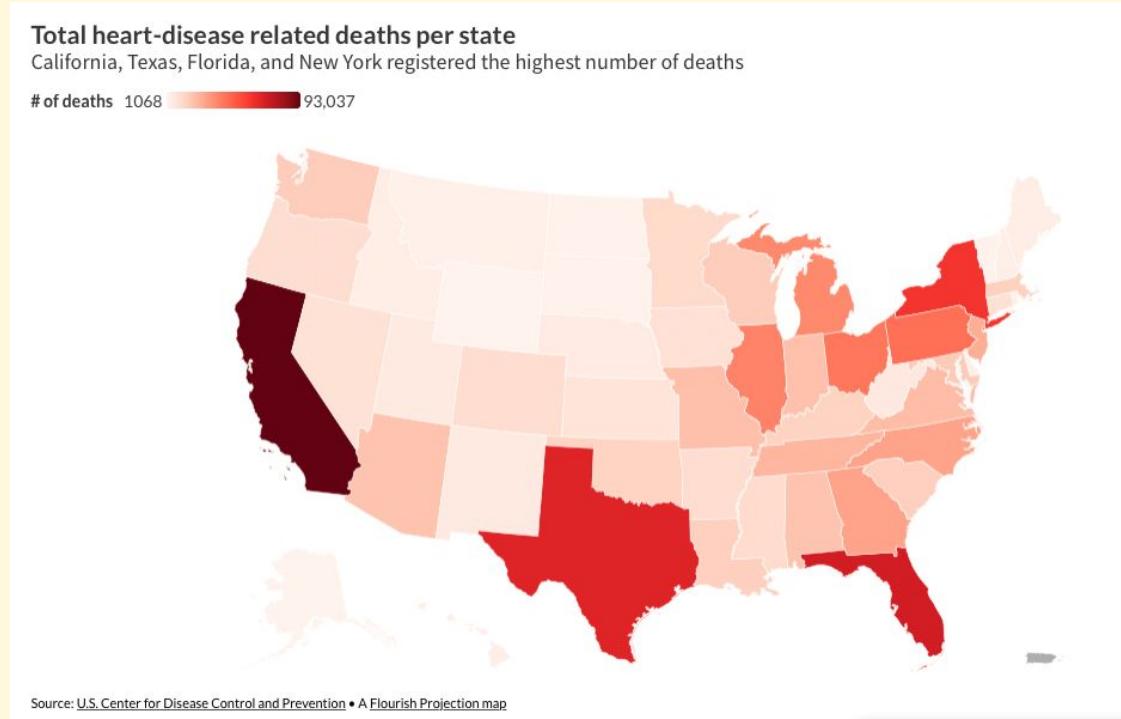
- wypełnienie kolorami powierzchni
- kolory kodują wartości (kategoryczne lub numeryczne)
- większe powierzchnie przykuwają większą uwagę, mniejsze powierzchnie mogą być niezauważalne
- często warto grupować zmienną ciągłą w dyskretne grupy (przedziały)
- kluczowa jest normalizacja - warto przedstawić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

median income

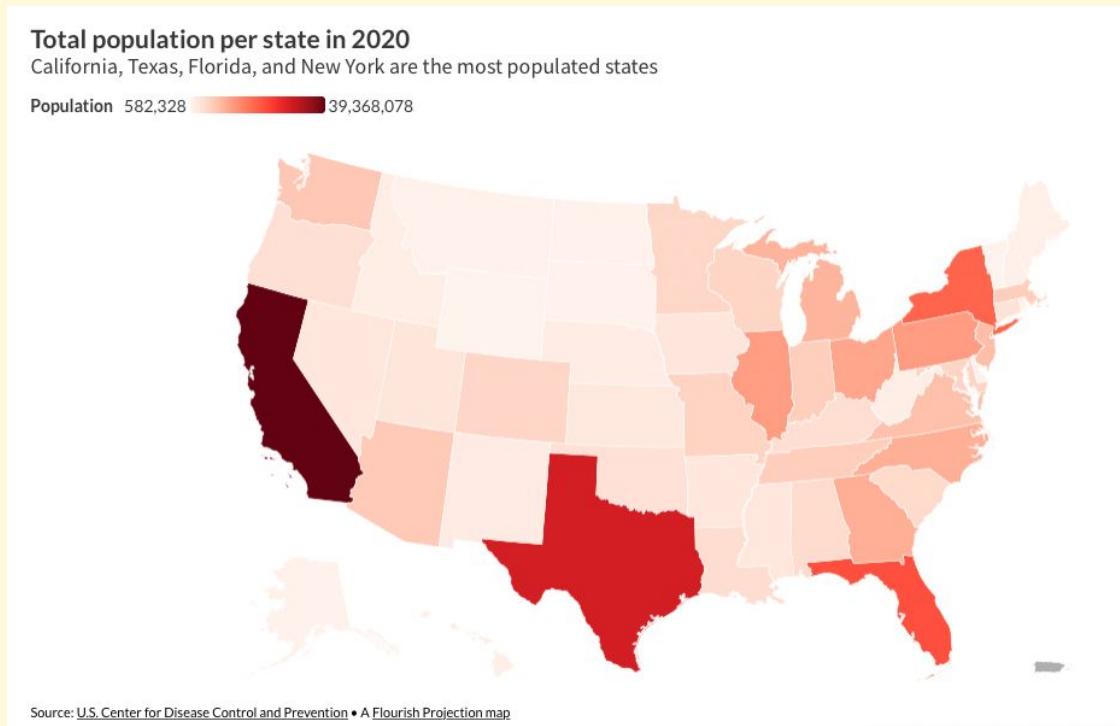
- > \$105k
- \$85k to \$105k
- \$60k to \$85k
- \$30k to \$60k
- < \$30k



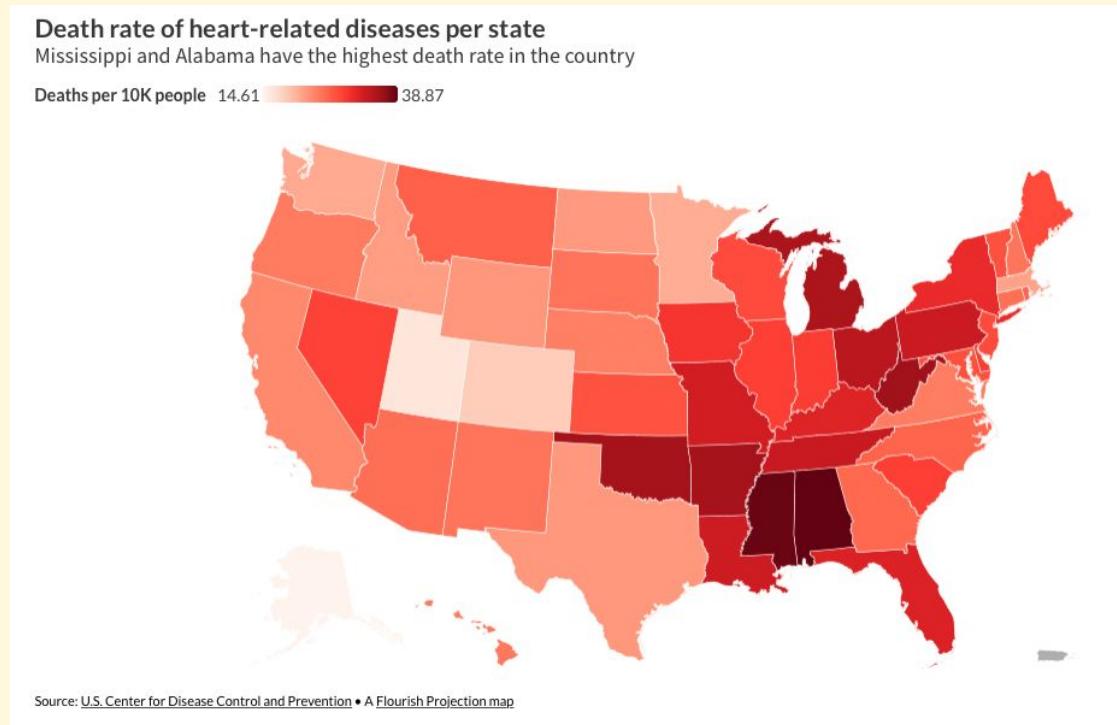
Metody prezentacji danych na mapach



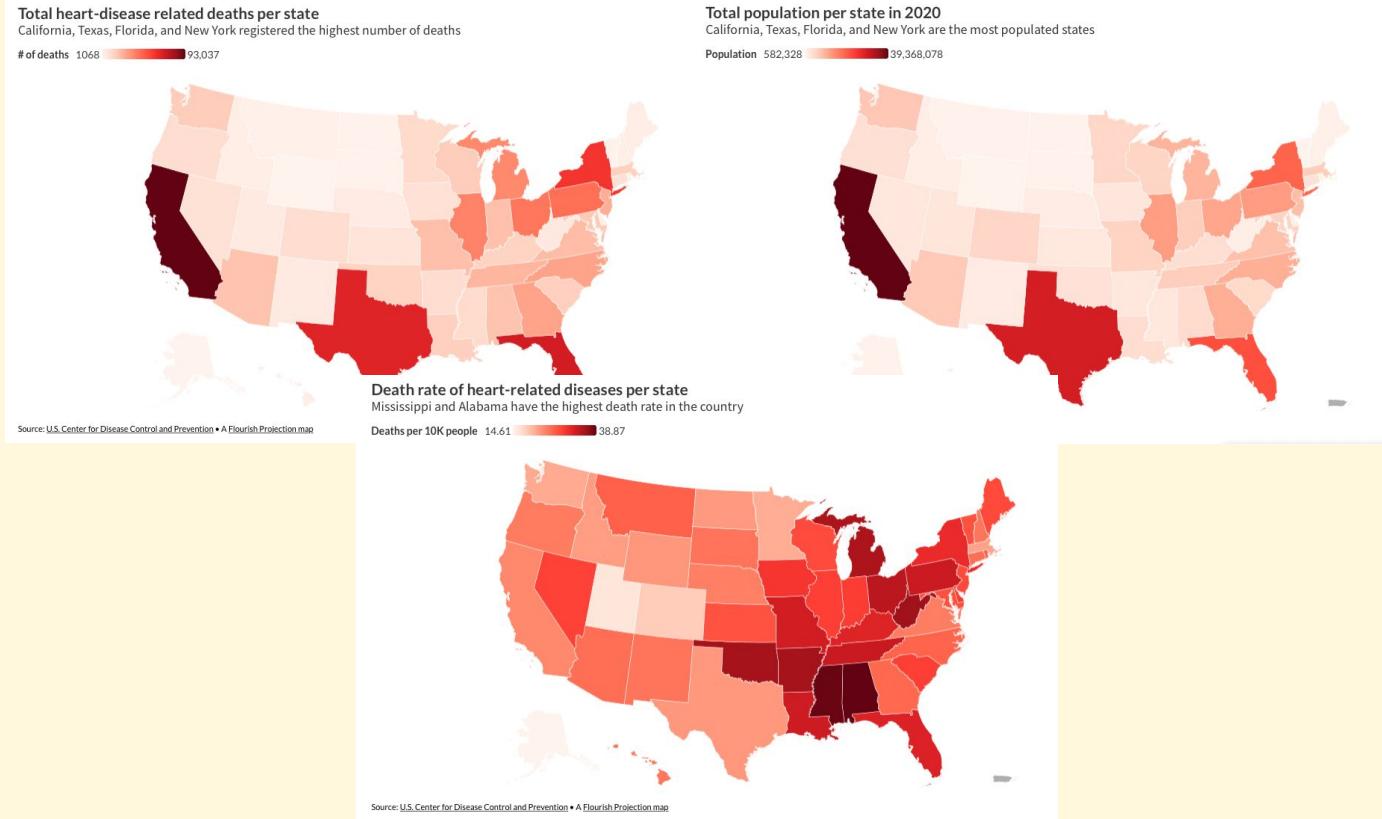
Metody prezentacji danych na mapach



Metody prezentacji danych na mapach



Metody prezentacji danych na mapach



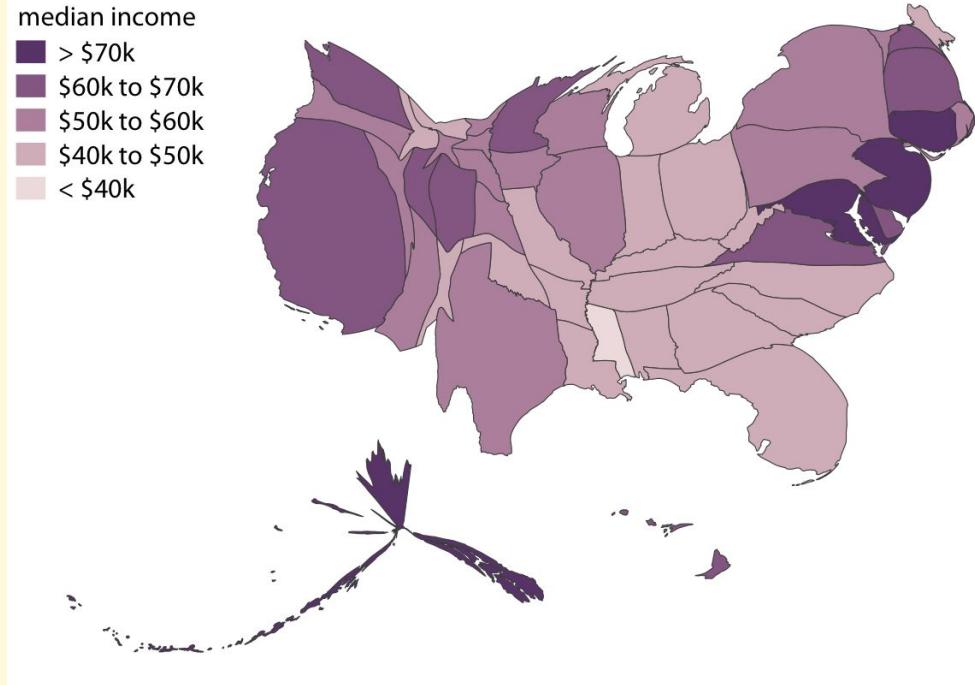
Metody prezentacji danych na mapach

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

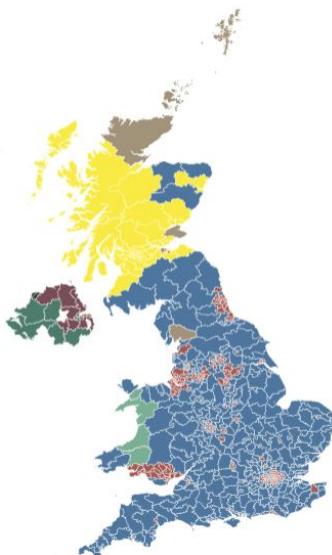
- > \$70k
- \$60k to \$70k
- \$50k to \$60k
- \$40k to \$50k
- < \$40k



Metody prezentacji danych na mapach

Results of the 2019 UK General Election

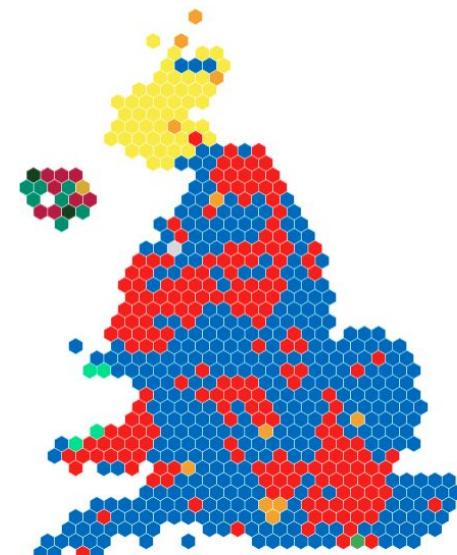
■ Con ■ Lab ■ LD ■ SNP ■ Green ■ PC ■ DUP ■ SF ■ SDLP ■ Alliance ■ Spk



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

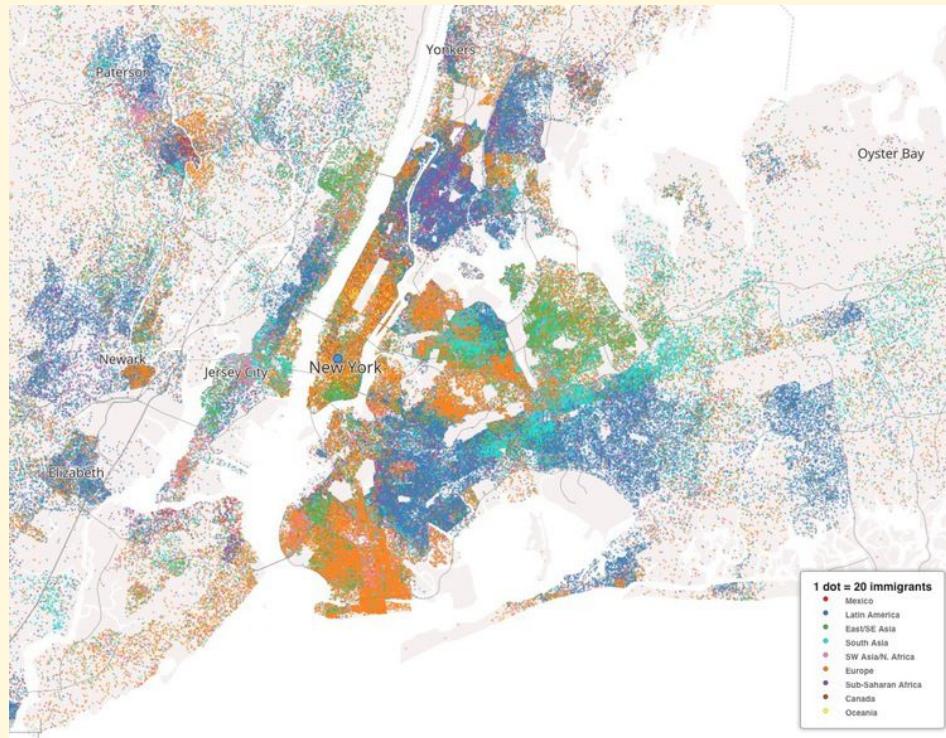


Source: [ODI Leeds](#), House of Commons Library • A Flourish 3D map

Metody prezentacji danych na mapach

MAPA KROPKOWA

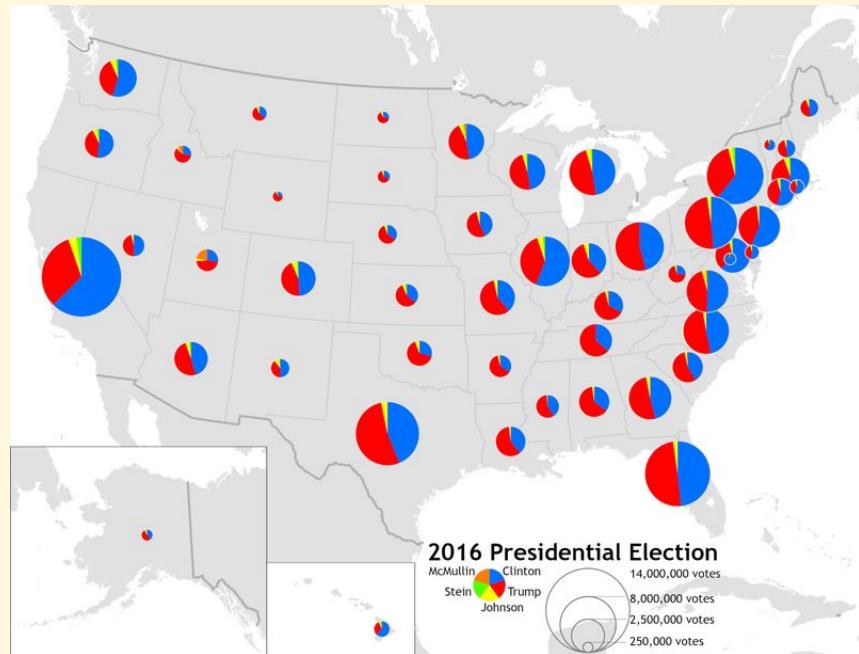
- informacja - umieszczenie znaków w odpowiednim miejscu na mapie
- jeden znak (kropka) - pewna liczba obiektów (niekoniecznie jeden)
- jedynie dane bezwzględne
- waga danego zjawiska, różna liczliwość obiektów → rozmiar znaku
- więcej niż jedno zjawisko, typ obiektu → kolor/typ znaków
- manipulując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



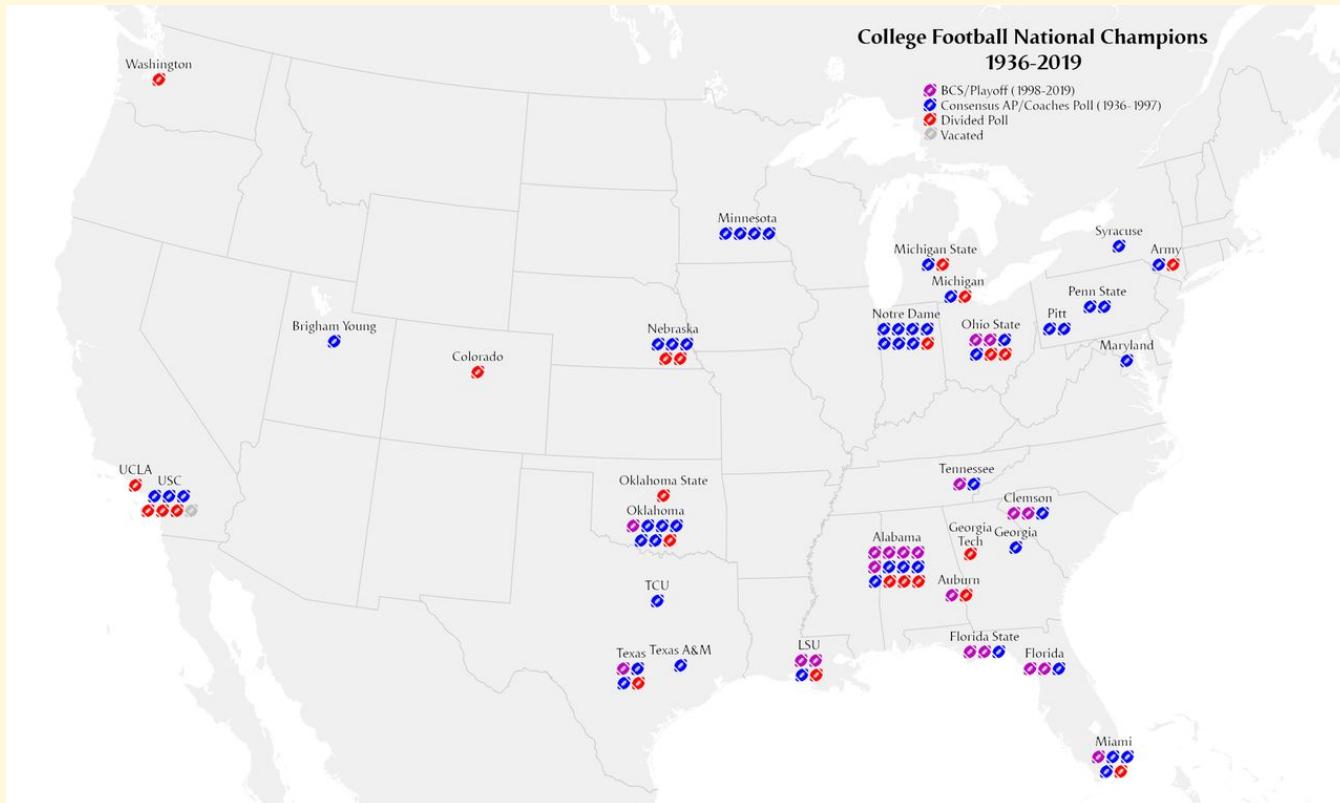
Metody prezentacji danych na mapach

KARTODIAGRAM

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

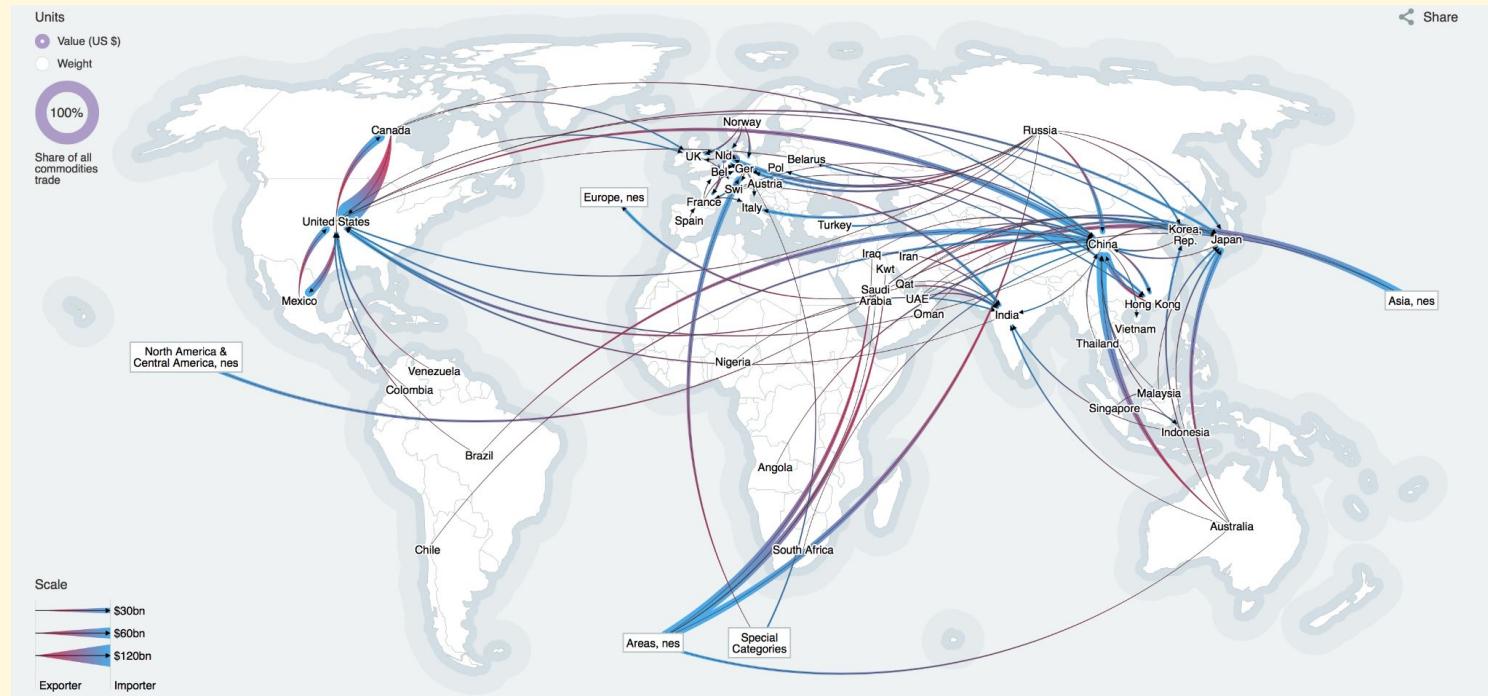


Metody prezentacji danych na mapach



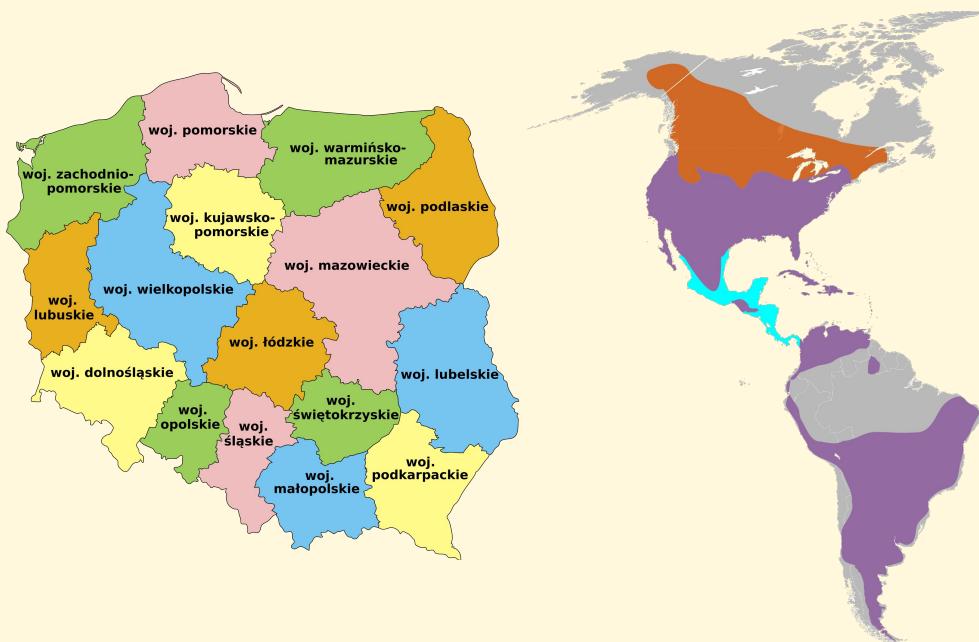
Metody prezentacji danych na mapach

MAPA PRZEPŁYWU

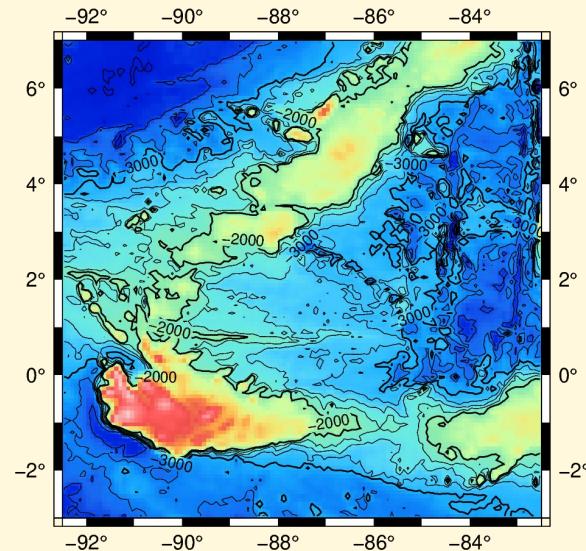


Metody prezentacji danych na mapach

MAPA CHOROCHROMATYCZNA/ZASIĘGÓW



MAPA IZOLINIOWA + MAPA CIEPŁA



Odwzorowania

Własności do zachowania:

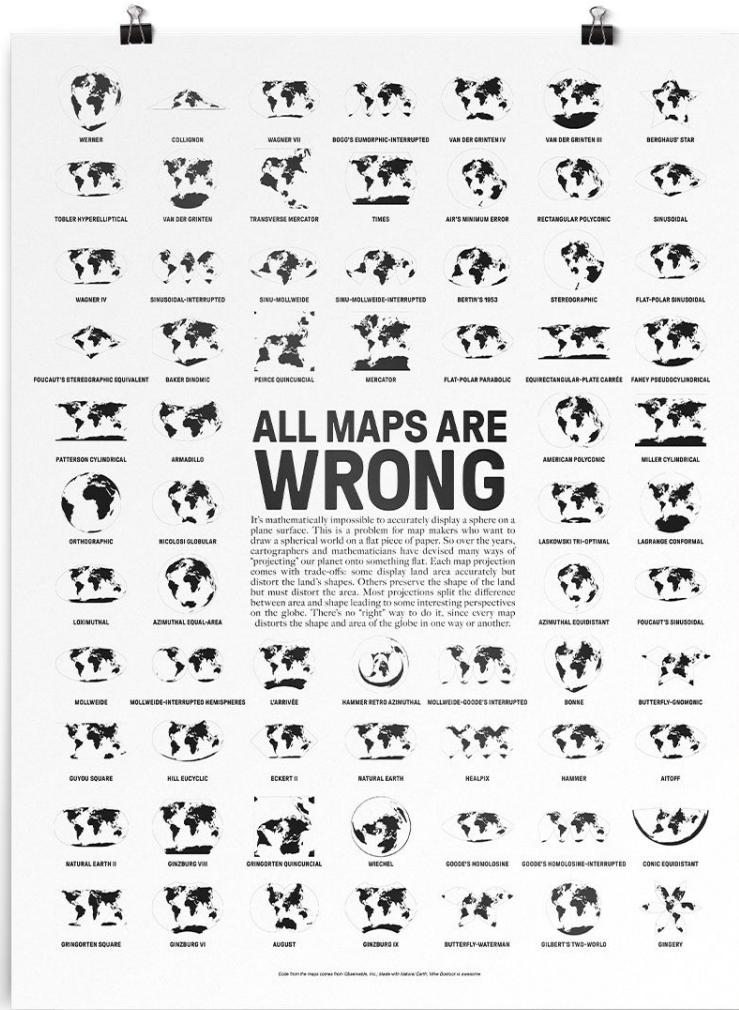
- kąty
- powierzchnie
- (odległości)

Problem: powierzchnia Ziemi nie jest izometryczna z płaszczyzną!

Nie jest możliwe zachowanie **więcej niż jednej własności** równocześnie.

Inna opcja: kompromis – pogodzenie się ze zniekształceniami różnego typu.

Johnny Harris
<https://store.dftba.com/products/all-maps-are-wrong-poster>



Odwzorowania



Koordynaty/systemy odniesienia

Cel: systematyzacja, monitorowanie różnych odwzorowań

Ważne przykłady:

- WGS84 (EPSG:4326) – układ jednolity dla całego świata wykorzystywany w GPS
- EPSG:3857 – układ stosowany w mapach internetowych

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

Odwzorowania

ODWZOROWANIE MERKATORA

- jedno z pierwszych (XVI wiek)
- wiernie zachowuje kąty
- używane w mapach nawigacyjnych
- zmodyfikowana wersja używana w mapach Google
- zniekształcenia tym większe, im dalej od równika

Kiedy stosować?

- gdy potrzeba zoomowania do lokalnych map
- mapy mniejszych państw
- kiedy rozmiar nie ma takiego znaczenia jak ogólne przedstawienie



Odwzorowania

ODWZOROWANIE MOLLWEIDEGO

- wiernie zachowuje powierzchnię
- bieguny są punktami

Kiedy stosować?

- dla dobrze rozpoznawalnej mapy świata (często w atlasach)
- mając dane na obu półkulach, dalej od równika
- kiedy najważniejsza jest powierzchnia, nie kształt



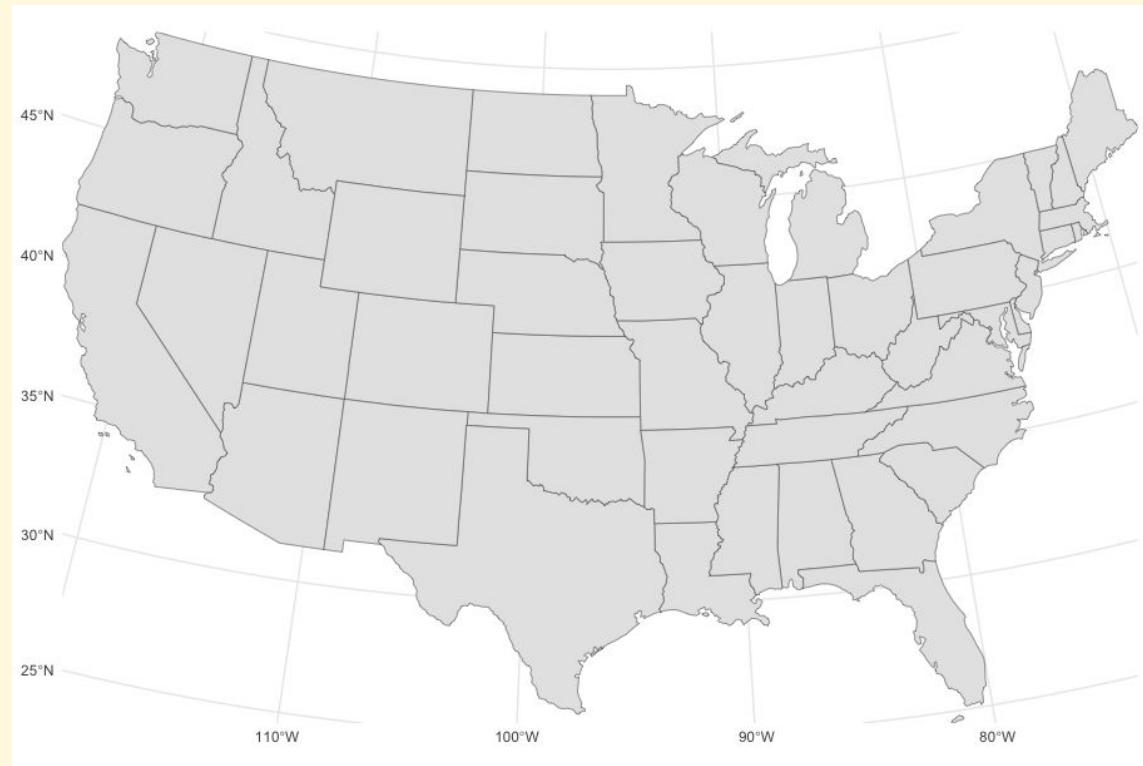
Odwzorowania

ODWZOROWANIA STOŽKOWE (NP. ALBERSA, LAMBERTA)

- wszystkie zniekształcenia są niewielkie pomiędzy dwoma wyróżnionymi równoleżnikami

Kiedy stosować?

- dla przedstawienia obszarów o dużej rozciągłości równoleżnikowej (wschód-zachód)
- USA, Kanada (oficjalne odwzorowania)



Odwzorowania

ODWZOROWANIA AZYMUTALNE (NP. ROWNOODLEGŁOŚCIOWE)

- zachowuje kierunki i odległości od wybranego centralnego punktu
- "symuluje" widok na Ziemię z kosmosu
- znajduje się na fladze ONZ (odwzorowanie scentrowane na biegunie północnym)

Kiedy stosować?

- do pokazania trasy samolotu
- do pokazania dystansu od epicentrum trzęsienia ziemi
- do pokazania zasięgu łączności radiowej





Jakie odwzorowanie wybrać?

- Nie istnieje jedno odwzorowanie, które pasowałoby do wszystkich celów.
- Każde odwzorowanie powoduje zniekształcenia, w szczególności w obszarach poza “centrum zainteresowania”.
- Wybór może zależeć od przeznaczenia, tematu mapy, np.:
 - kartogramy, w szczególności wykresy gęstości → zachowanie powierzchni
 - nawigacja → zachowanie kątów
 - zasięgi obiektów, zjawisk, loty → zachowanie odległości

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>

WHAT YOUR FAVORITE

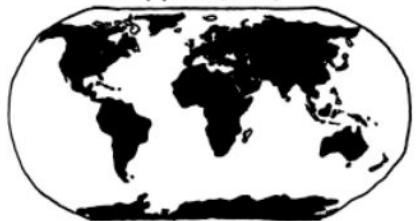
MAP PROJECTION SAYS ABOUT YOU

MERCATOR



YOU'RE NOT REALLY INTO MAPS.

ROBINSON



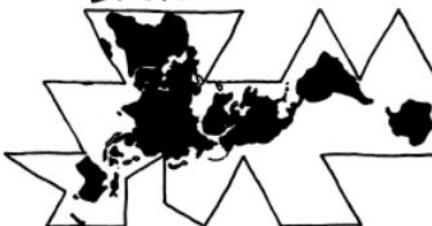
YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.

VAN DER GRIJNEN



YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!

Dymaxion



YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.

WINKEL-TRIPEL



NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A WT FAN SINCE LONG BEFORE "NAT GEO" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST-".

GOODE HOMOLOSONE



THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE'D JUST ELECT NORMAL PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE THAT ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY NEED TO.

HOBЫ-DYER



YOU WANT TO AVOID CULTURAL IMPERIALISM, BUT YOU'VE HEARD BAD THINGS ABOUT GALL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.

PLATE CARRÉE
(EQUIRECTANGULAR)



YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.

A GLOBE!



YES, YOU'RE VERY CLEVER.

PEIRCE QUINCUNCIAL



YOU THINK THAT WHEN WE LOOK AT A MAP, WHAT WE
REALLY SEE IS OURSELVES. AFTER YOU FIRST SAW
INCEPTION, YOU SAT SILENT IN THE THEATER FOR
SIX HOURS. IT FREAKS YOU OUT TO REALIZE THAT
EVERYONE AROUND YOU HAS A SKELETON INSIDE THEM.
YOU HAVE REALLY LOOKED AT YOUR HANDS.

WATERMAN BUTTERFLY



REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN
THE 1909 CAHILL MAP IT'S BASED— ... YOU HAVE A FRAMED
REPRODUCTION AT HOME?! WHOA. ... LISTEN, FORGET
THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?

GALL-PETERS



I HATE YOU.