



Wstęp do Eksploracji Danych

Politechnika Warszawska

Anna Kozak



Grafy

Grafy

- graph $G = (V, E)$
- vertices $V = \{v_1, v_2, \dots, v_n\}$
- edge $E = \{e_1, e_2, \dots, e_m\}$

Reprezentacje

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}\}$$

$$E = \{\{v_1, v_2\}, \{v_1, v_8\}, \{v_2, v_3\}, \{v_3, v_5\}, \{v_3, v_9\}, \\ \{v_3, v_{10}\}, \{v_4, v_5\}, \{v_4, v_6\}, \{v_4, v_9\}, \{v_5, v_8\}, \\ \{v_6, v_8\}, \{v_6, v_9\}, \{v_7, v_8\}, \{v_7, v_9\}, \{v_8, v_{10}\}, \\ \{v_9, v_{10}\}\}$$

Reprezentacje

$v_1:$ v_2, v_8

$v_2:$ v_1, v_3

$v_3:$ v_2, v_5, v_9, v_{10}

$v_4:$ v_5, v_6, v_9

$v_5:$ v_3, v_4, v_8

$v_6:$ v_4, v_8, v_9

$v_7:$ v_8, v_9

$v_8:$ $v_1, v_5, v_6, v_7, v_9, v_{10}$

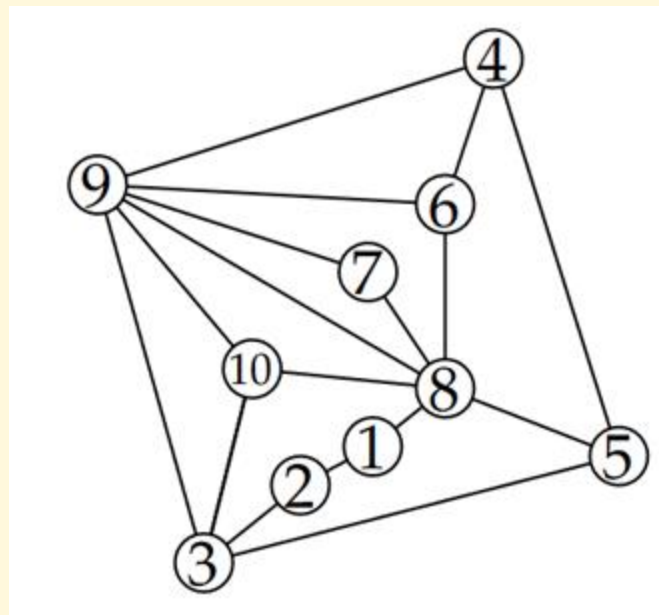
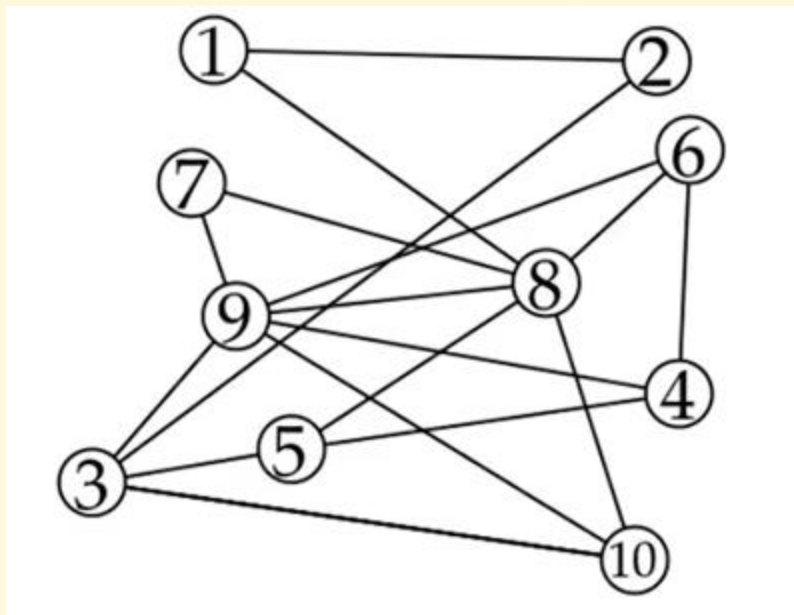
$v_9:$ $v_3, v_4, v_6, v_7, v_8, v_{10}$

$v_{10}:$ v_3, v_8, v_9

Reprezentacje

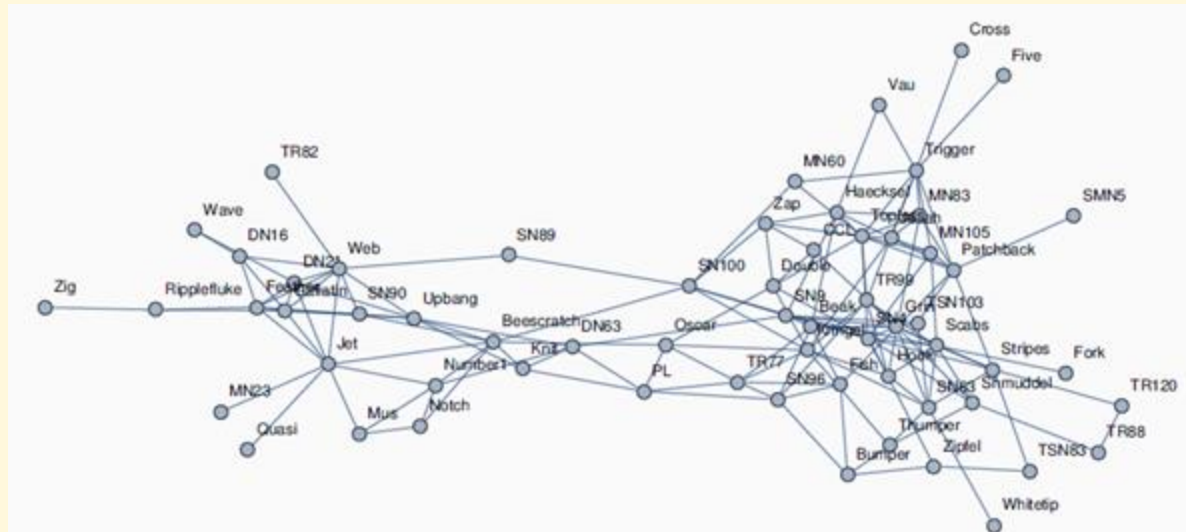
$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Reprezentacje



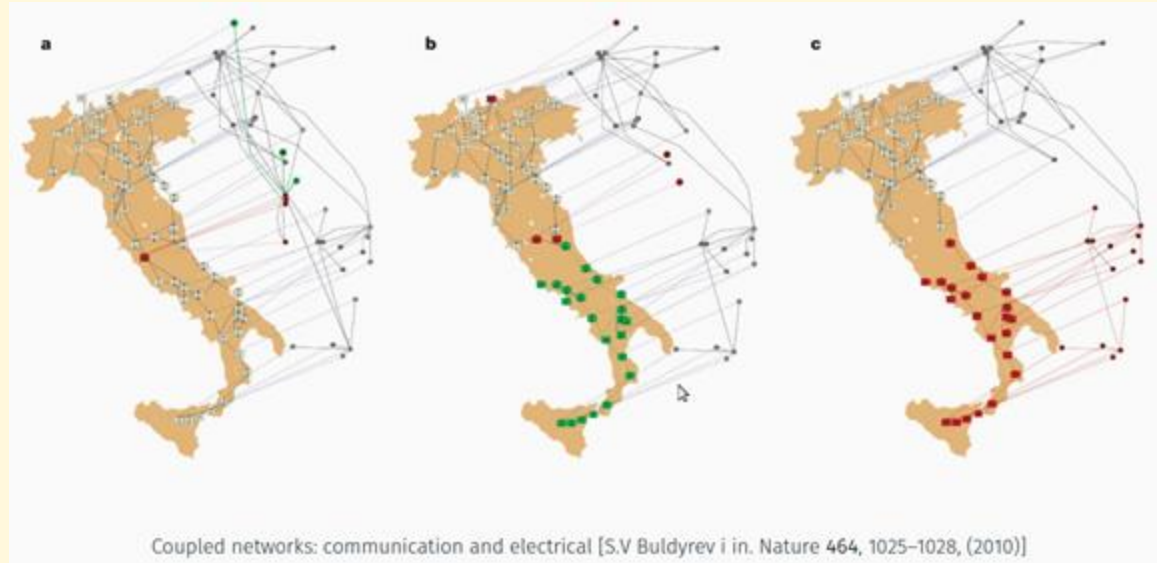
Grafy abstrakcyjne

- sieć społecznościowa
- sieć komunikacyjna
- sieć filogenetyczna



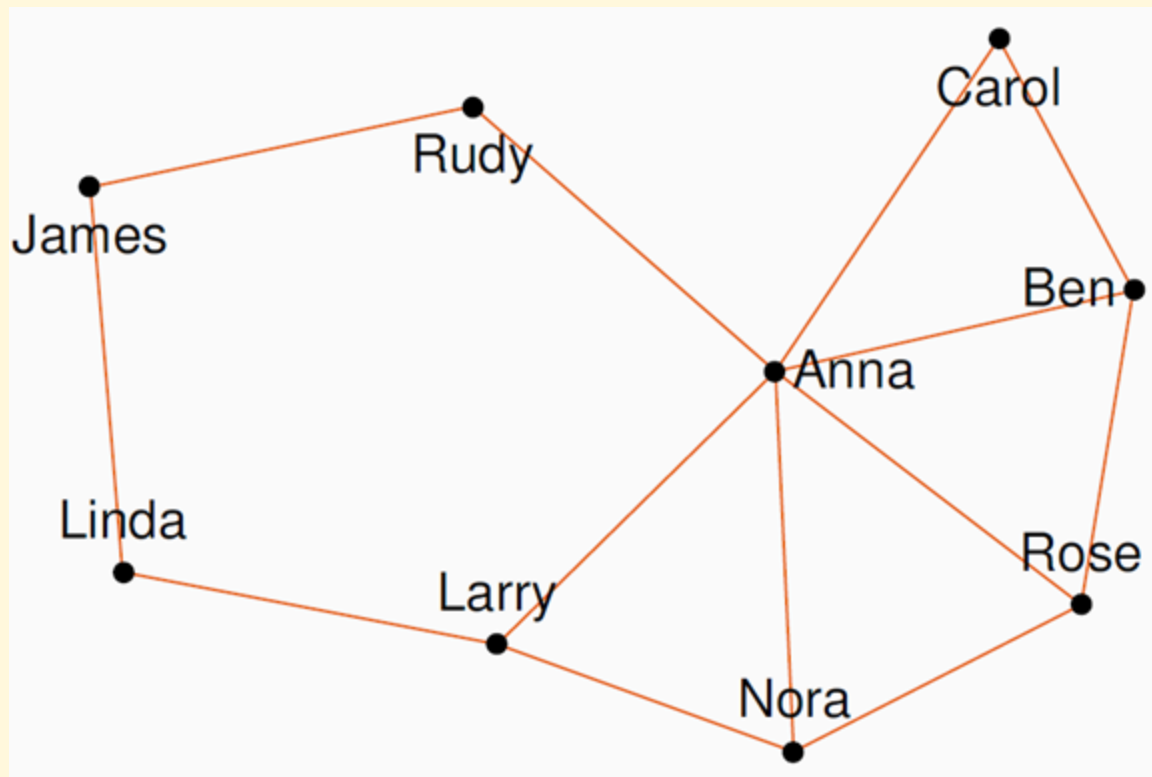
Grafy rzeczywiste

- sieć metra (transportu)
- sieć dróg
- sieć telekomunikacyjna



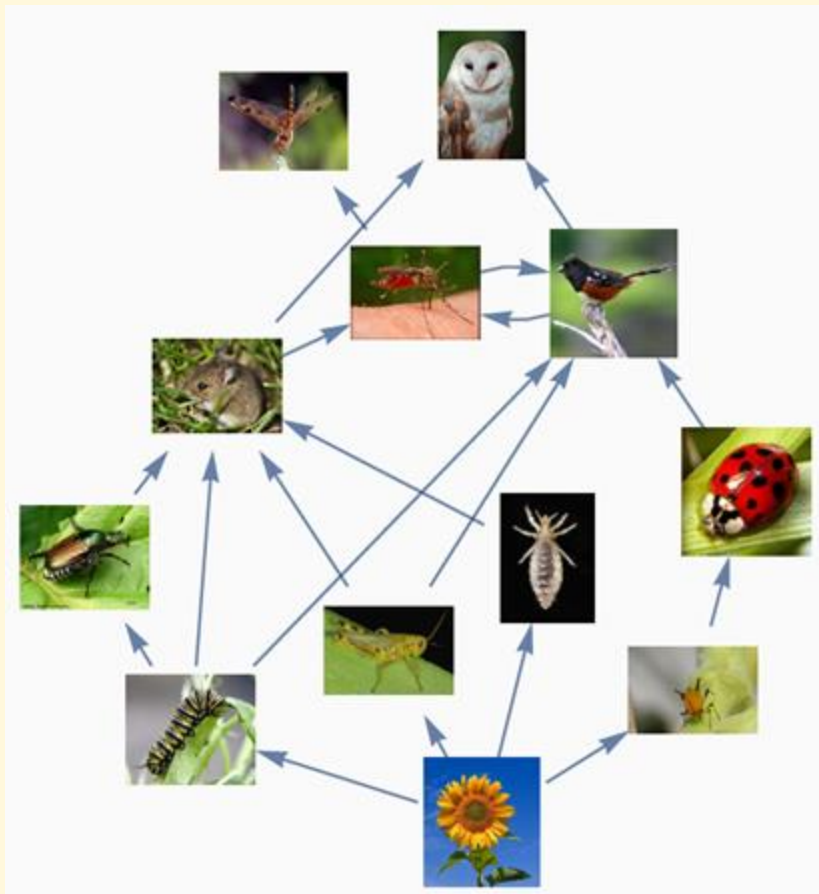
Rodzaje grafów

- grafy proste



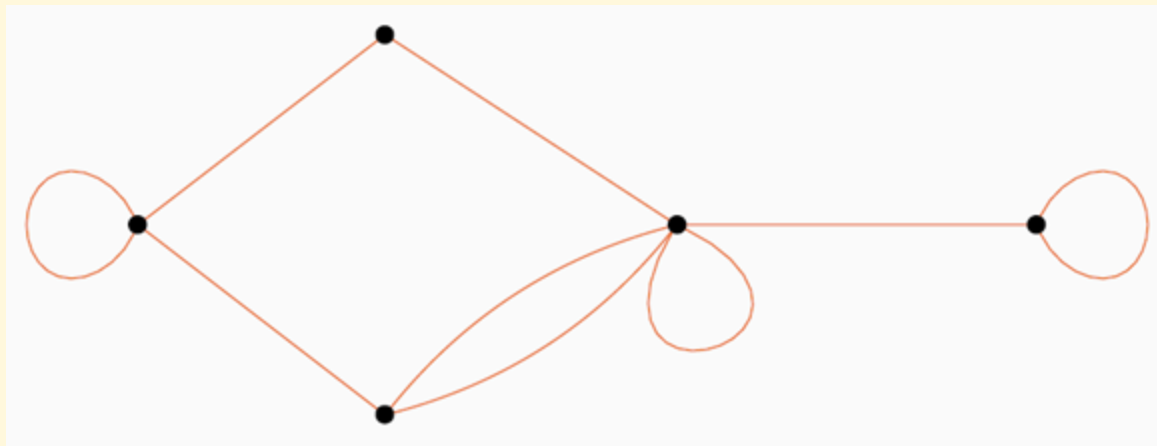
Rodzaje grafów

- grafy proste
- grafy skierowane



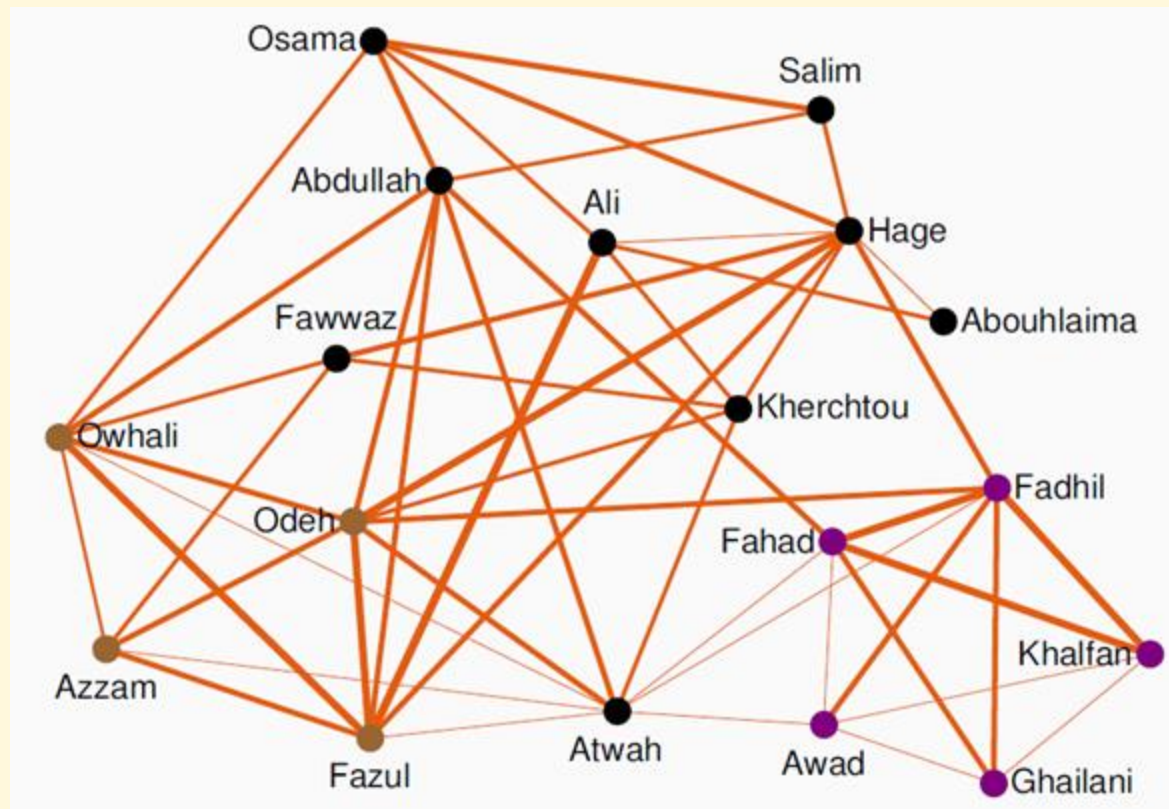
Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy



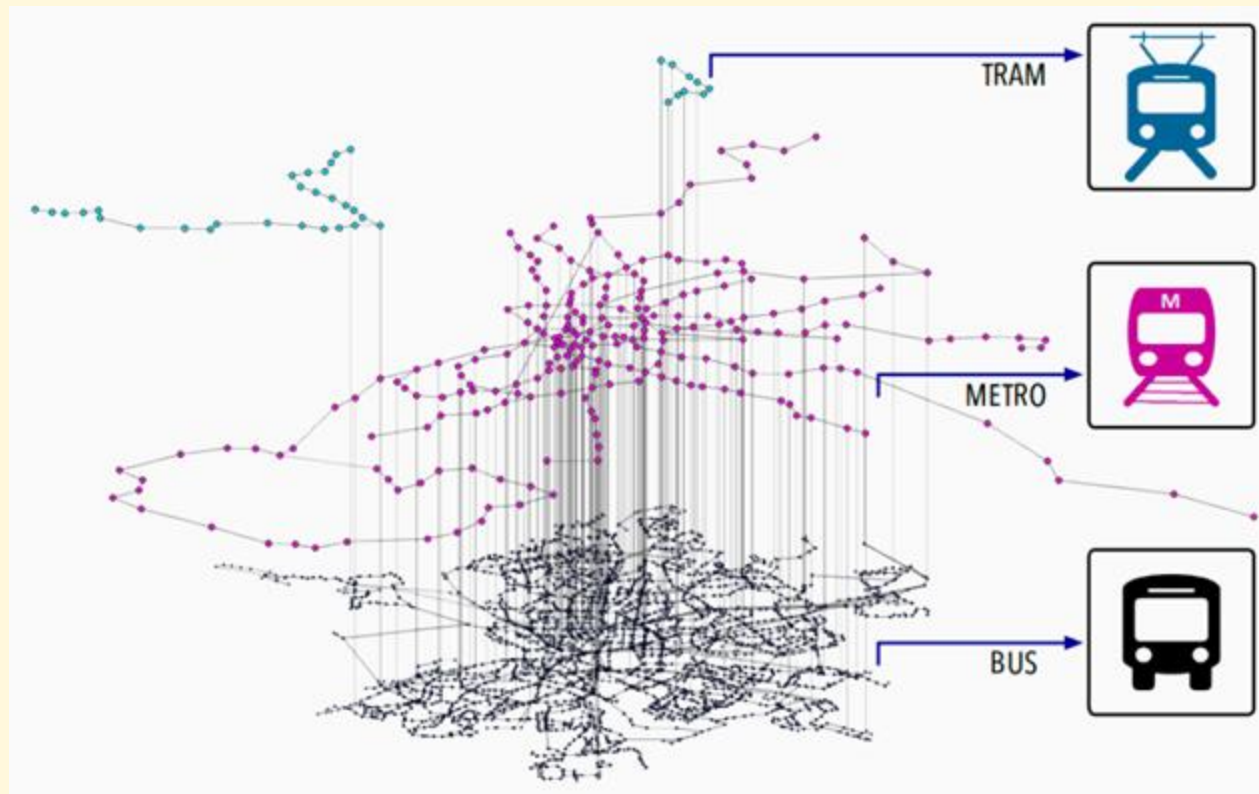
Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone



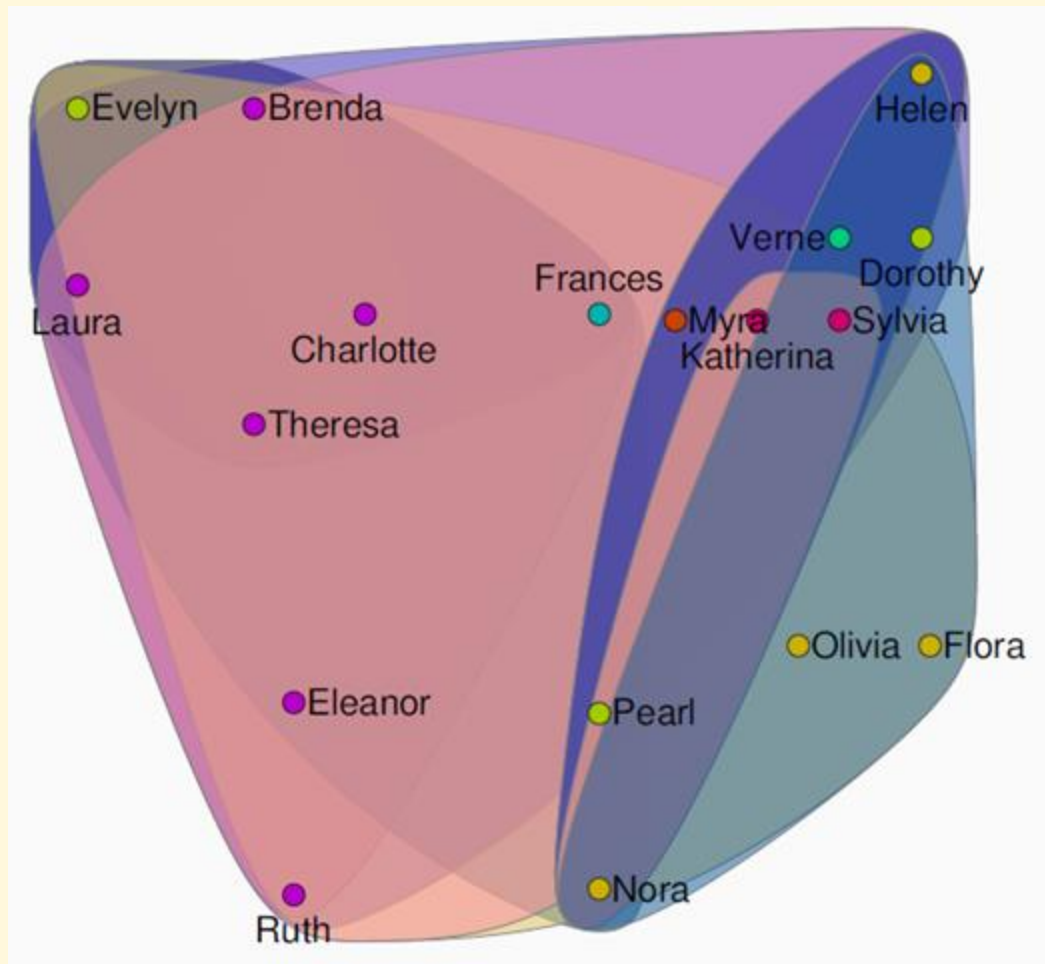
Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone
- grafy warstwowe



Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone
- grafy warstwowe
- hipergrafy



1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.
3. Niektóre grafy są zbyt duże, aby rysować je ręcznie.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.
3. Niektóre grafy są zbyt duże, aby rysować je ręcznie.

Potrzebujemy algorytmów, które automatycznie rysują grafy, aby były one bardziej dostępne dla ludzi.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.
4. Często spotykamy grafy dużych rozmiarów, które są trudne do wizualizacji, ze względu na: mnogość wierzchołków, nachodzenie labeli, nakładające się krawędzie, długi czas rysowania grafu.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.
4. Często spotykamy grafy dużych rozmiarów, które są trudne do wizualizacji, ze względu na: mnogość wierzchołków, nachodzenie labeli, nakładające się krawędzie, długi czas rysowania grafu.

Wizualizacja grafów jest skomplikowana i nie ma uniwersalnych reguł jak to robić. Niektórych informacji nie da się ukazać samym obrazkiem.

Jak mierzyć grafy?

1. Współczynnik klasteryzacji (clustering coefficient)
2. Długość najdłuższej ścieżki / średnica grafu
3. Gęstość sieci ($2E/N$)
4. Największy stopień wierzchołka (celebryta)
5. Indeks Hirsch'a
6. PageRank
7. Próg epidemii
8. Liczba Erdos'a
9. Liczba Bacon'a

Liczba Erdos'a

Minimalna ścieżka współautorstwa artykułów prowadząca od naukowca A do Paula Erdos'a. Wartość dla Paula Erdos'a jest równa 0.

MR Collaboration Distance = 3

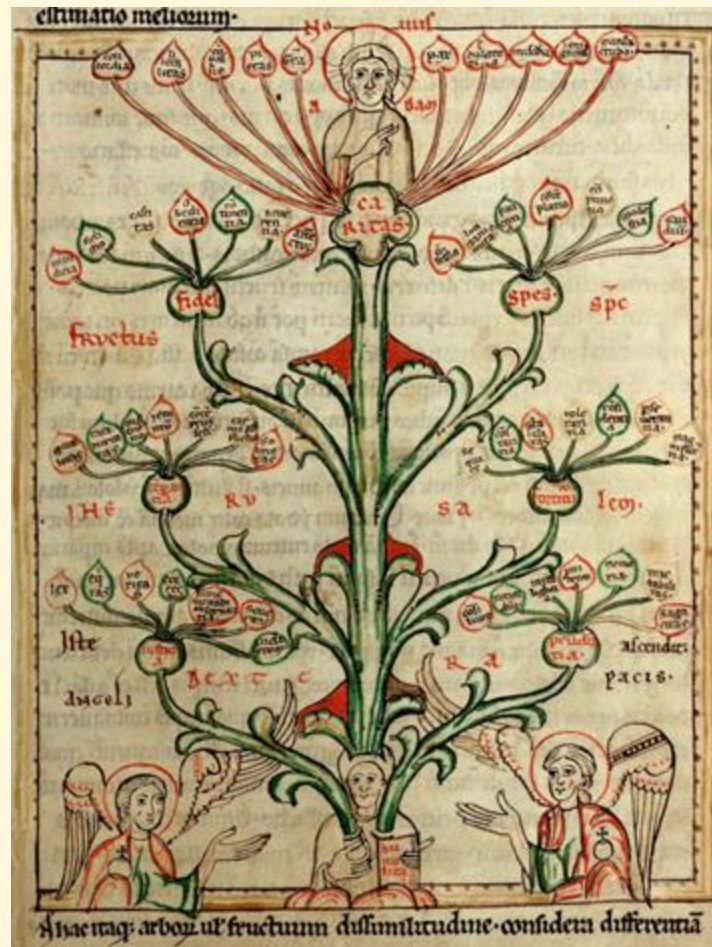
Grzegorzewski, Przemysław	coauthored with	Mesiar, Radko	MR3660830
Mesiar, Radko	coauthored with	Širáň, Jozef	MR4171596
Širáň, Jozef	coauthored with	Erdős, Paul ¹	MR1297187

MR Collaboration Distance = 4

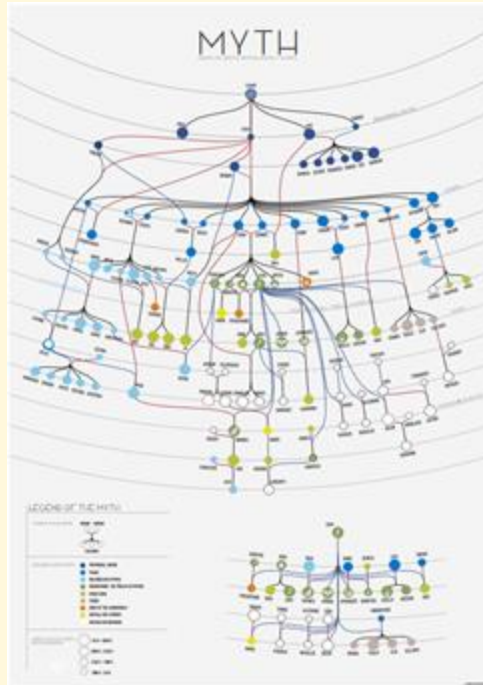
Cena, Anna	coauthored with	Gągolewski, Marek	MR4240238
Gągolewski, Marek	coauthored with	Mesiar, Radko	MR3158692
Mesiar, Radko	coauthored with	Širáň, Jozef	MR4171596
Širáň, Jozef	coauthored with	Erdős, Paul ¹	MR1297187

Drzewo cnót głównych

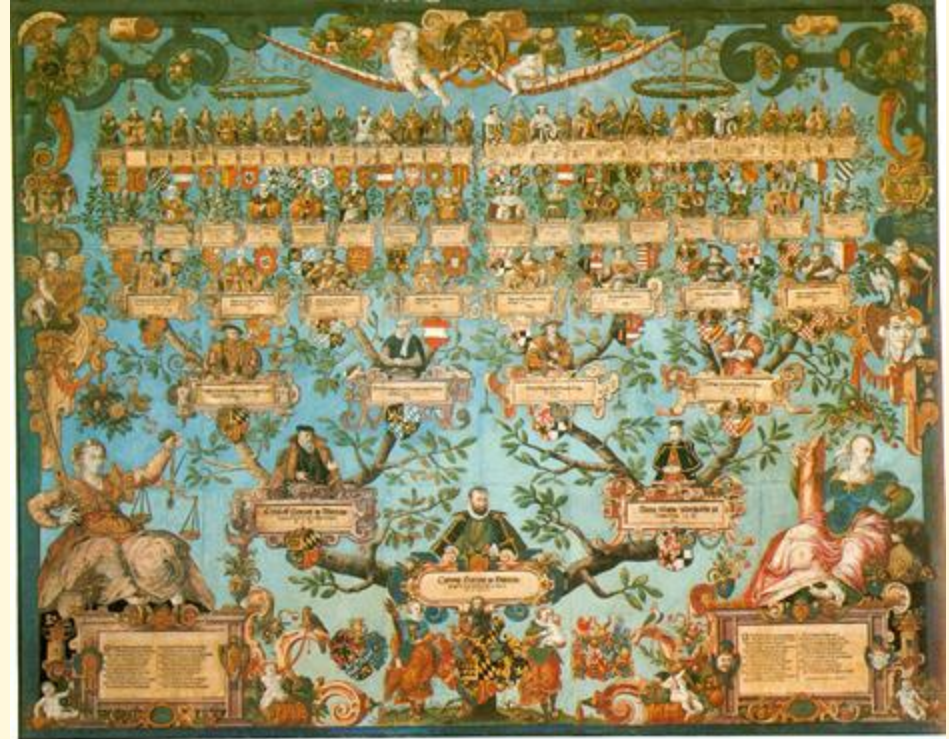
Cztery cnoty kardynalne umieszczone są niżej niż trzy cnoty teologiczne. Z każdej z cnót głównych wyrastają cnoty niższego rzędu. *Speculum Virginum*, XII w.



Drzewo genealogiczne (sieć społecznościowa)

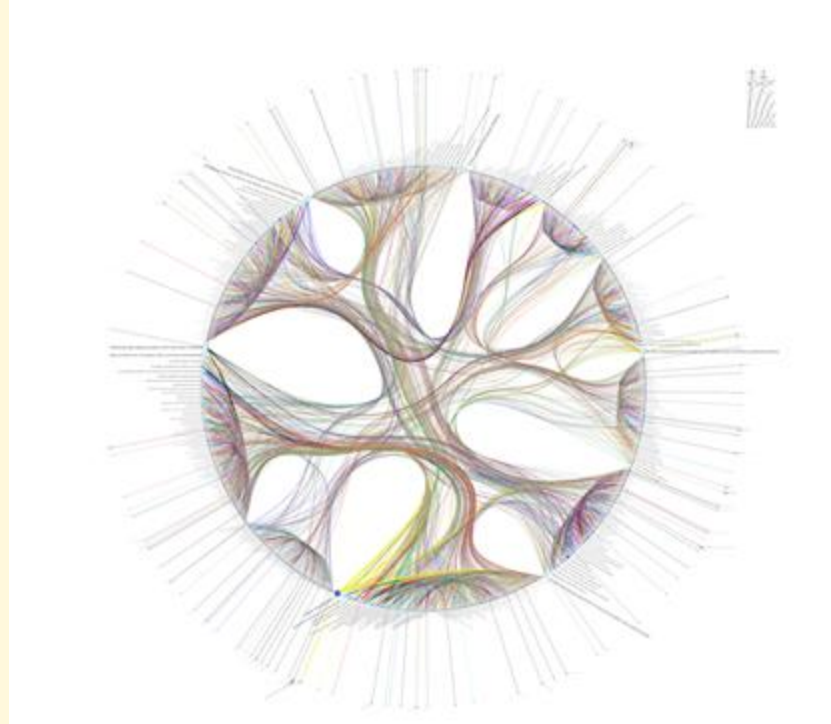


J. Klawitter, T. Mchedlidze, Link: go.uni-wue.de/myth-poster



Ahnentafel Herzog Ludwig von Württemberg, 1585

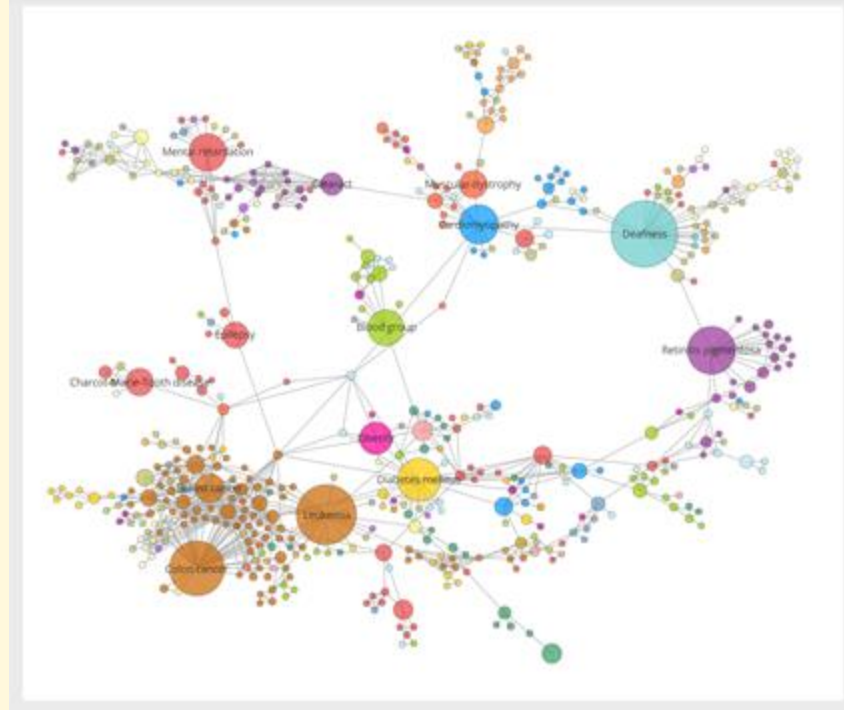
Graf cytowań (sieć społecznościowa)



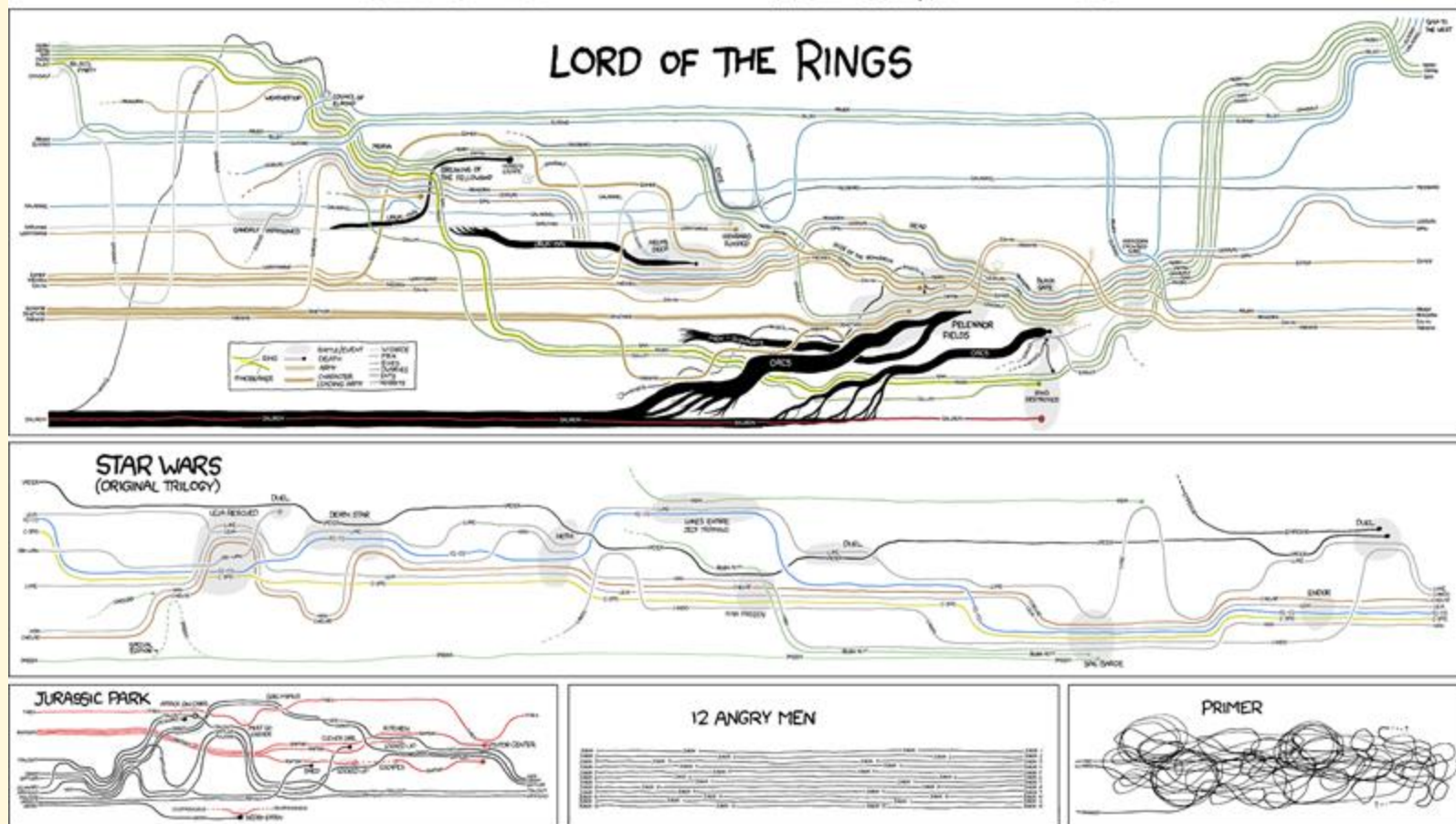
Kolej dużych prędkości w Europie (sieć komunikacyjna)



Sieć ludzkich chorób (bioinformatyka)



THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



- <https://seafire.rlp.net/f/f36d7e005a3c48a2bac2/>
- Przedmiot Social Networks & Recommendation Systems na studiach magisterskich Data Science - dr inż. Grzegorz Siudem
- The Anatomy of a Search Engine: Sergey Brin and Lawrence Page

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.”

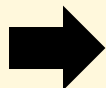
Czym jest dashboard?

Definicja dashboardu ulega zmianom ...

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

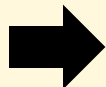
"wizualne przedstawienie danych używanych do monitorowania warunków i/lub ułatwiania zrozumienia"

wizualny



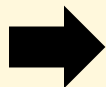
wizualna reprezentacja danych w postaci kafelkowego układu prostych wykresów i/lub dużych liczb

wizualny



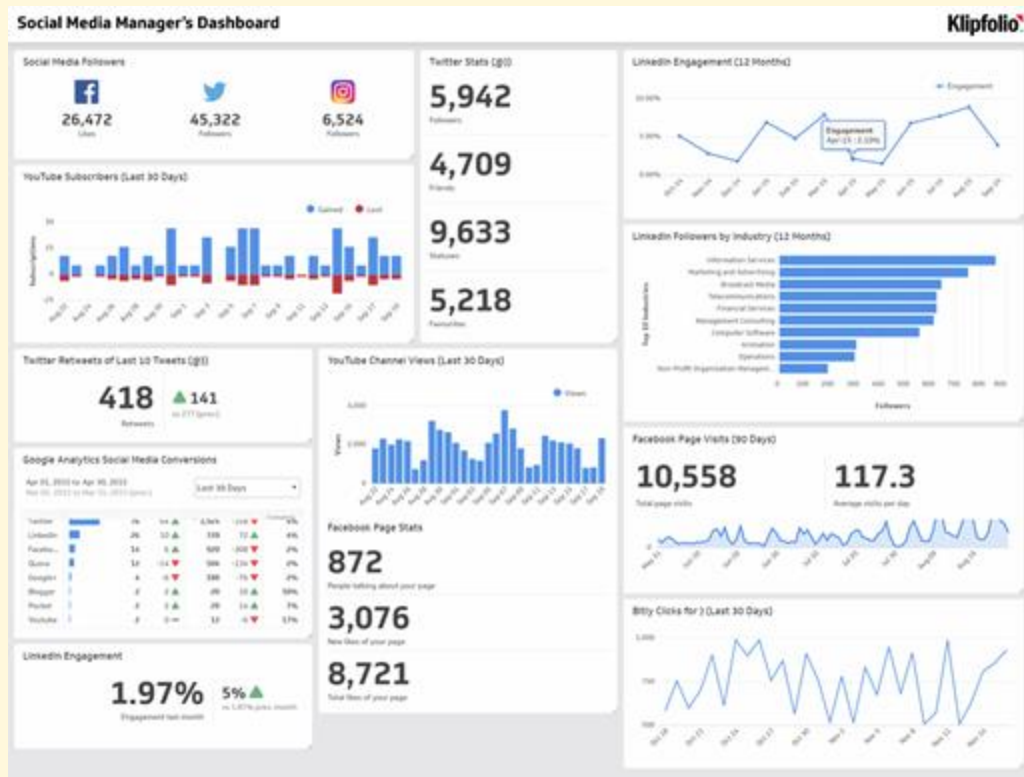
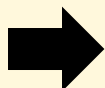
wizualna reprezentacja danych w postaci kafelkowego układu prostych wykresów i/lub dużych liczb

funkcjonalny

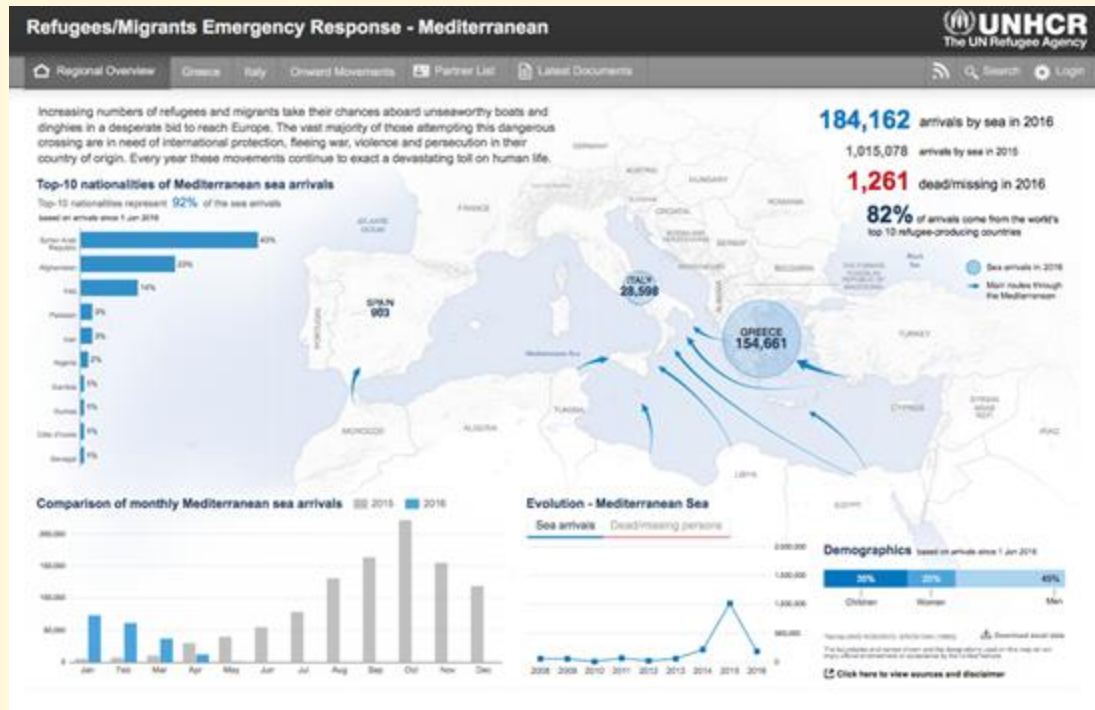
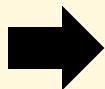


interaktywny ekran, który umożliwia monitorowanie w czasie rzeczywistym dynamicznie aktualizowanych danych

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"

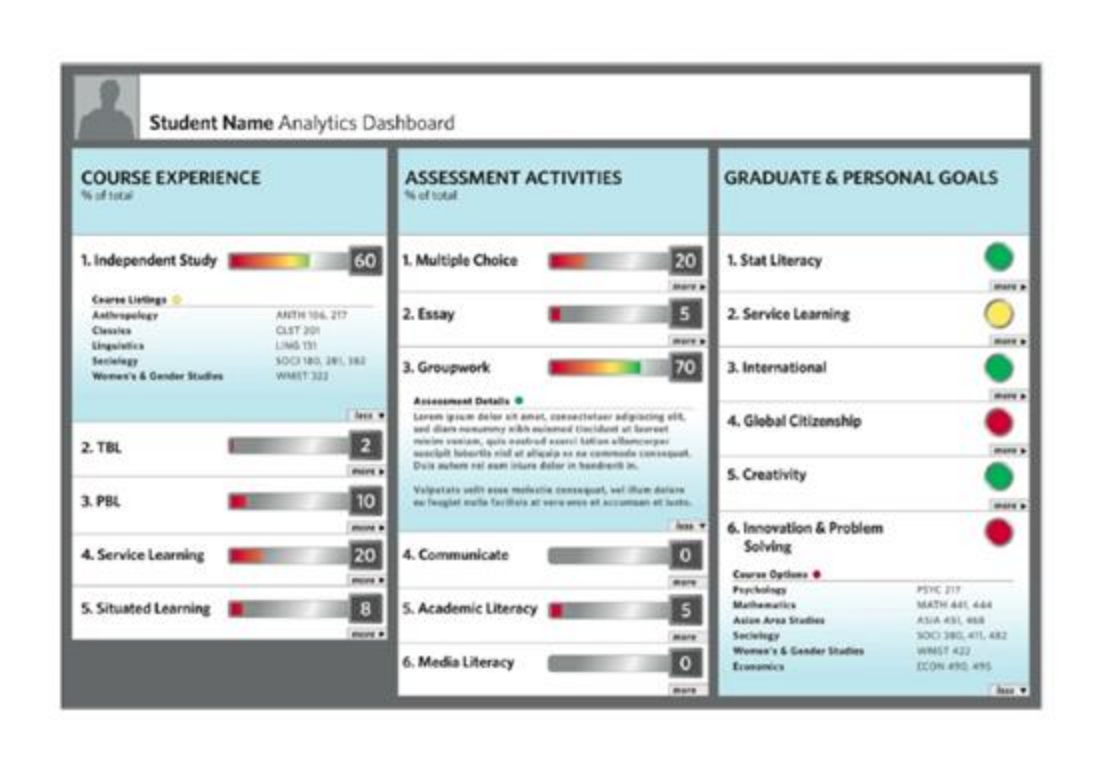
- ◆ **ocena operacji**

"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 umiejętność wizualizacji

Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

→ niska

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

AB Test on Feature Y (Satisfaction)

Weighted Satisfaction Score

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

Positive Satisfaction Rating

No Feature Y **78.4%** Diff vs On **6.6%** Feature Y **85.0%**
See in Widget



AB Test Feature Y Funnels

Edit Widget → Save Widget

No Feature Y **61.4%** Feature Y **65.7%**
4.38% vs. No Feature Y

Edit Widget → > 6 month LTV

No Feature Y **44.0%** Feature Y **42.2%**
-0.85% vs. No Feature Y

Edit Widget in Trial → Purchase

No Feature Y **3.54%** Feature Y **6.32%**
0.78% vs. No Feature Y

AB Test on Feature X (Satisfaction)

Weighted Satisfaction Score

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

Positive Satisfaction Rating

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



AB Test Feature X Funnels

Edit Widget → Save Widget

Feature X Off **62.3%** Feature X On **60.0%**
-2.34% vs. No Feature X

Edit Widget → > 6 month LTV

Feature X Off **43.2%** Feature X On **40.3%**
-2.92% vs. No Feature X

Edit Widget in trial → Purchase

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

Wymagana umiejętność wizualizacji

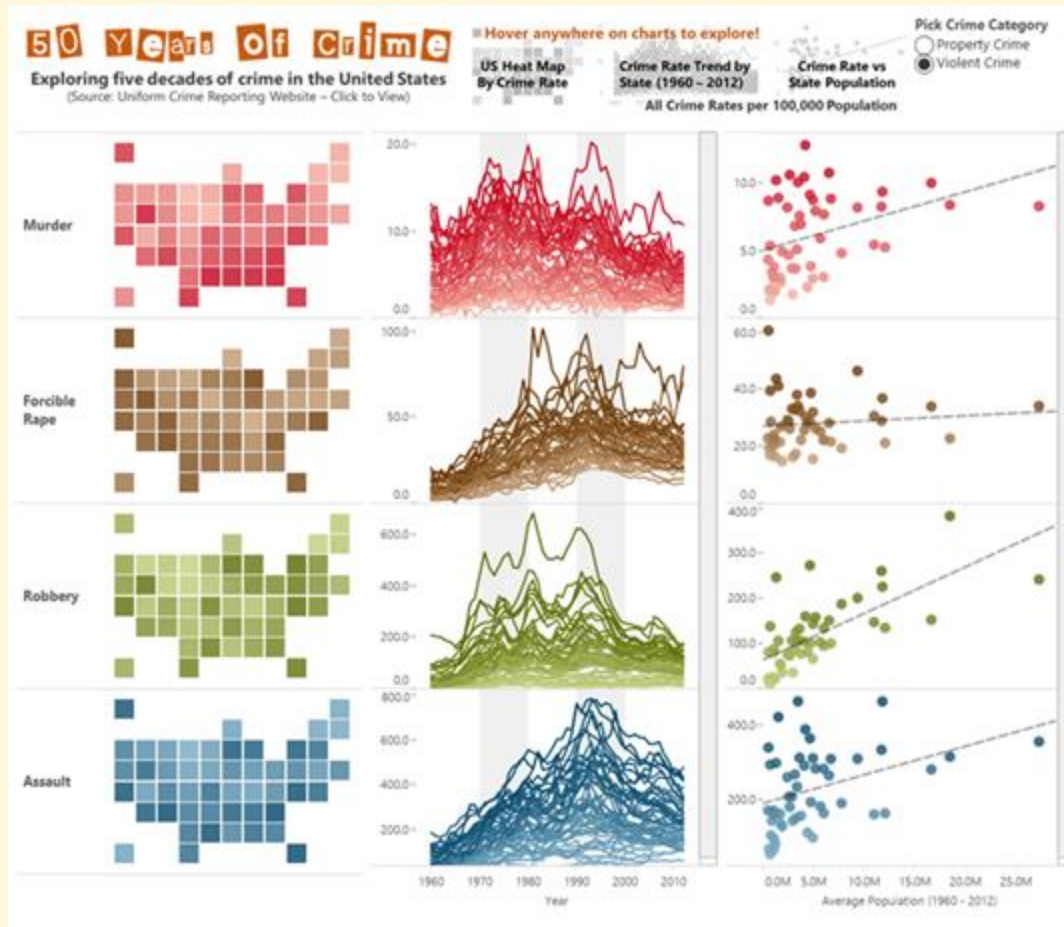
Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

→ 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



Wymagana umiejętność wizualizacji

Złożoność wizualizacji dostępnych w dashboardzie może ograniczać jego zrozumiałość.

→ niska

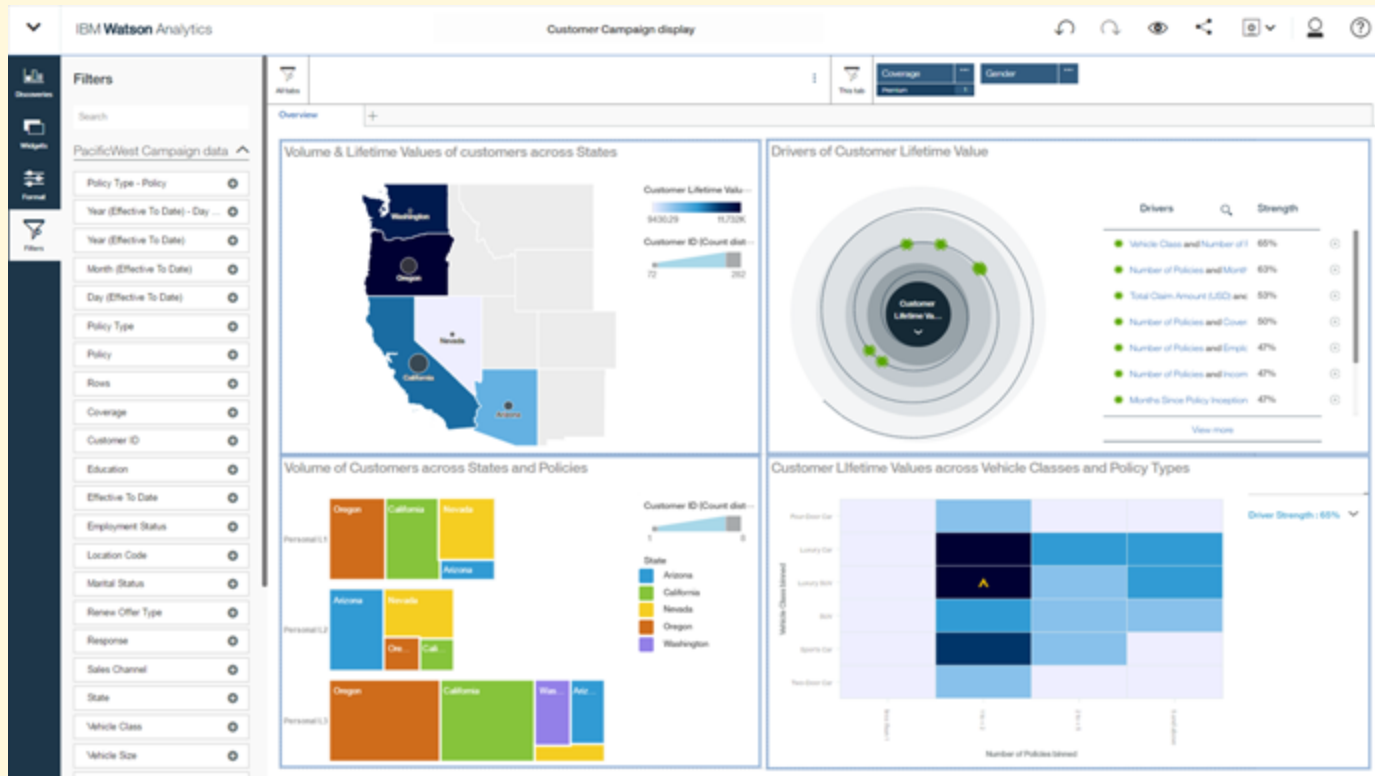
- ◆ podstawowe typy wizualizacji, takie jak wykresy słupkowe i liniowe z panelami i agregacją

→ średnia

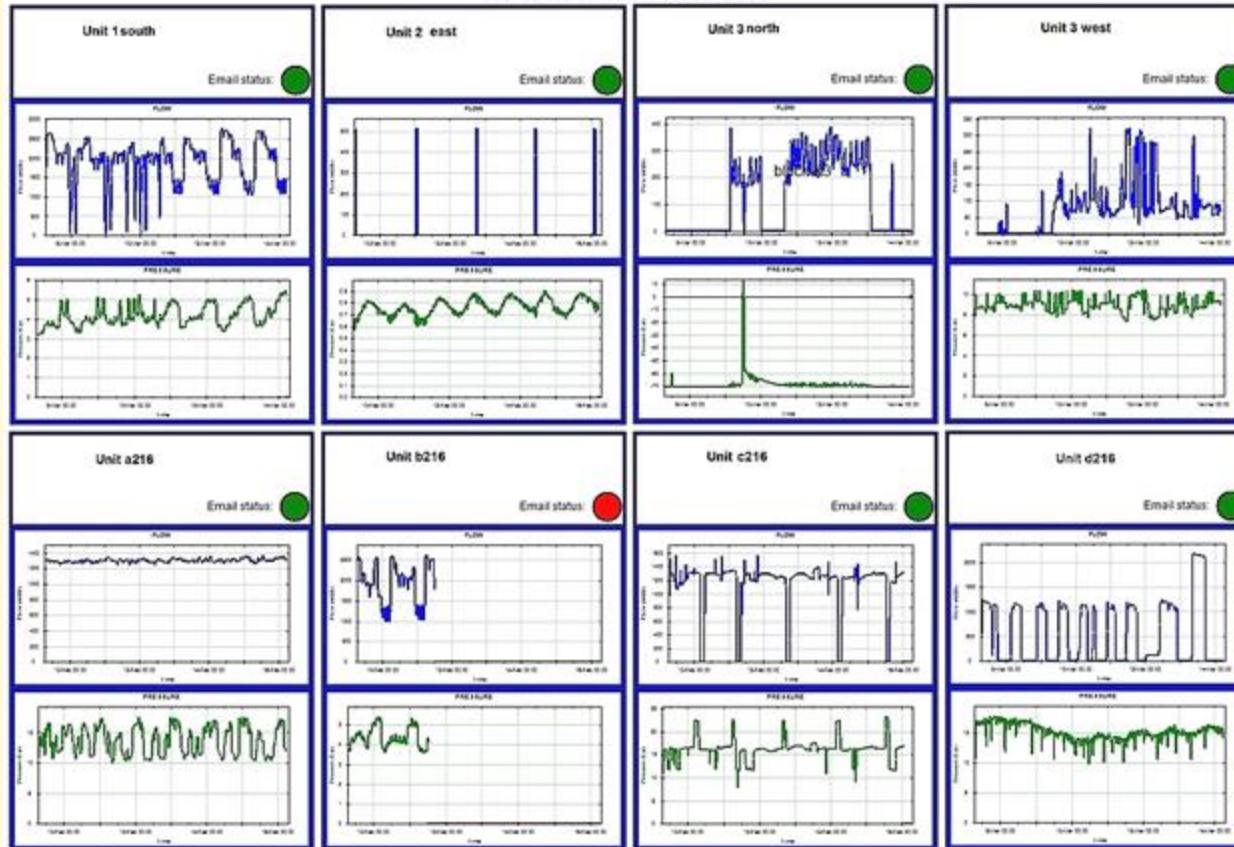
- ◆ szerokie zastosowanie dla wielu różnych osób w ramach danej struktury organizacyjnej, przy czym odbiorcy ci mają wspólny cel (np. wspieranie rentowności firmy)

→ wysoka

- ◆ radar, treemap i wizualizacje sieciowe, słupki błędów lub przedziały, połączone wykresy rozproszenia lub inne niestandardowe wizualizacje



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

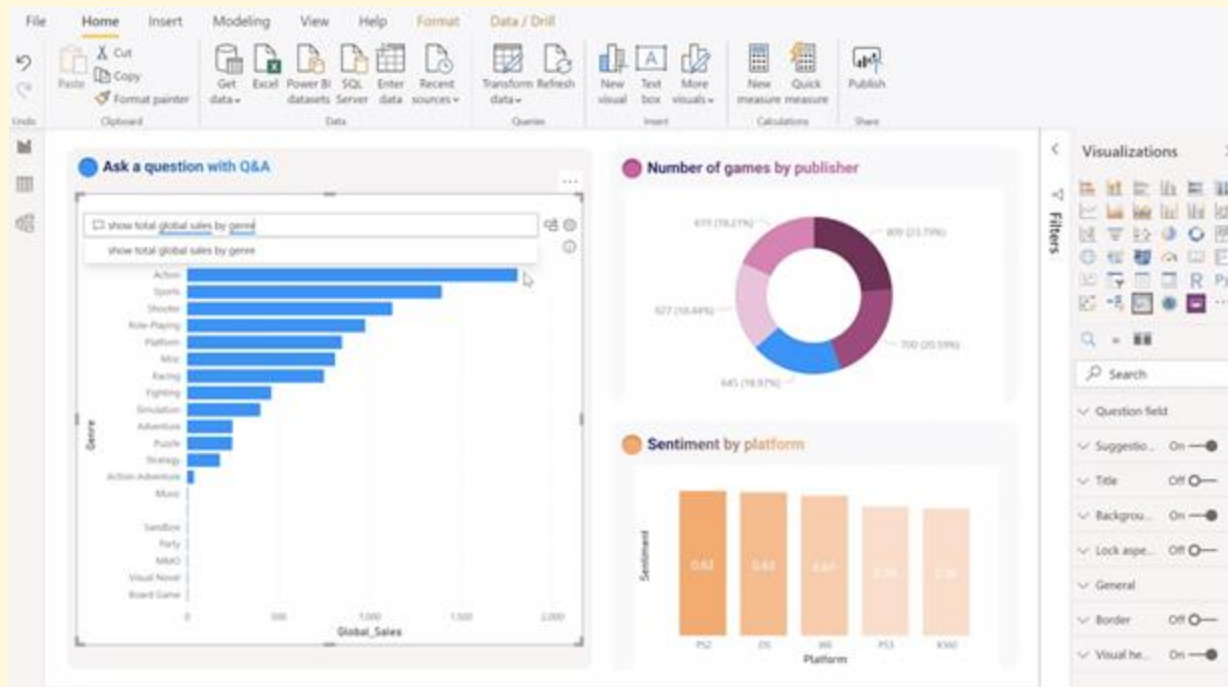
Najważniejsze zasady tworzenia dashboardów

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

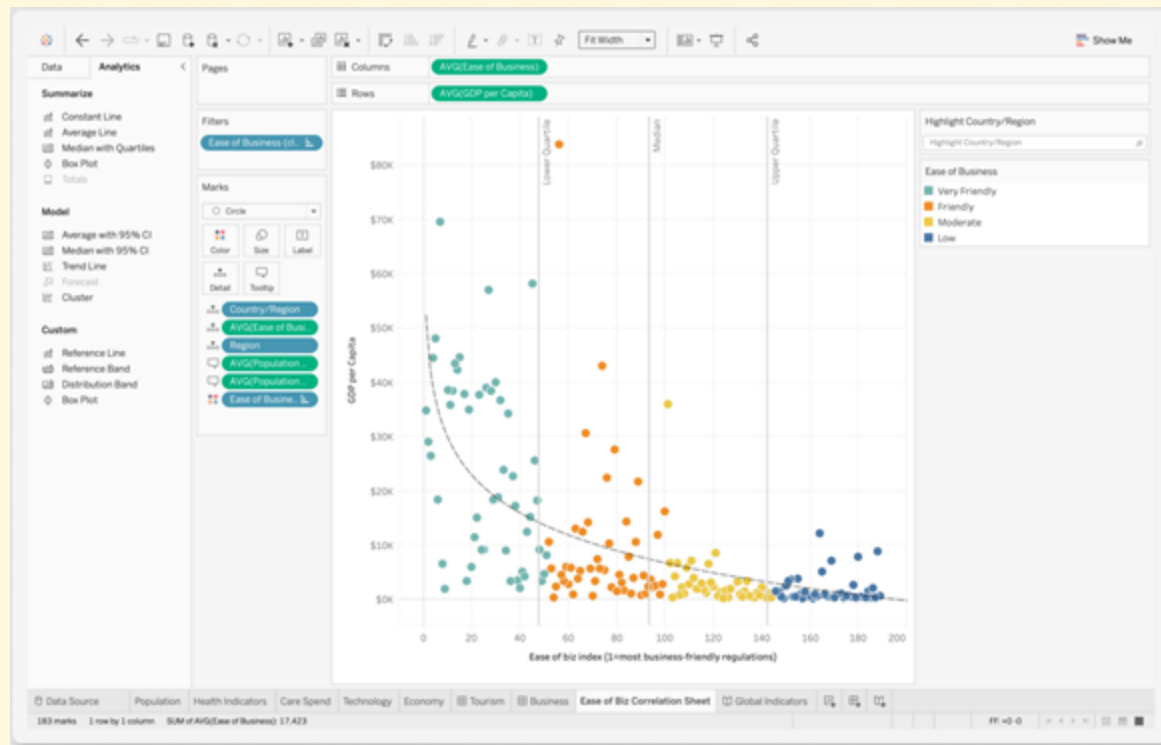
Power BI

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

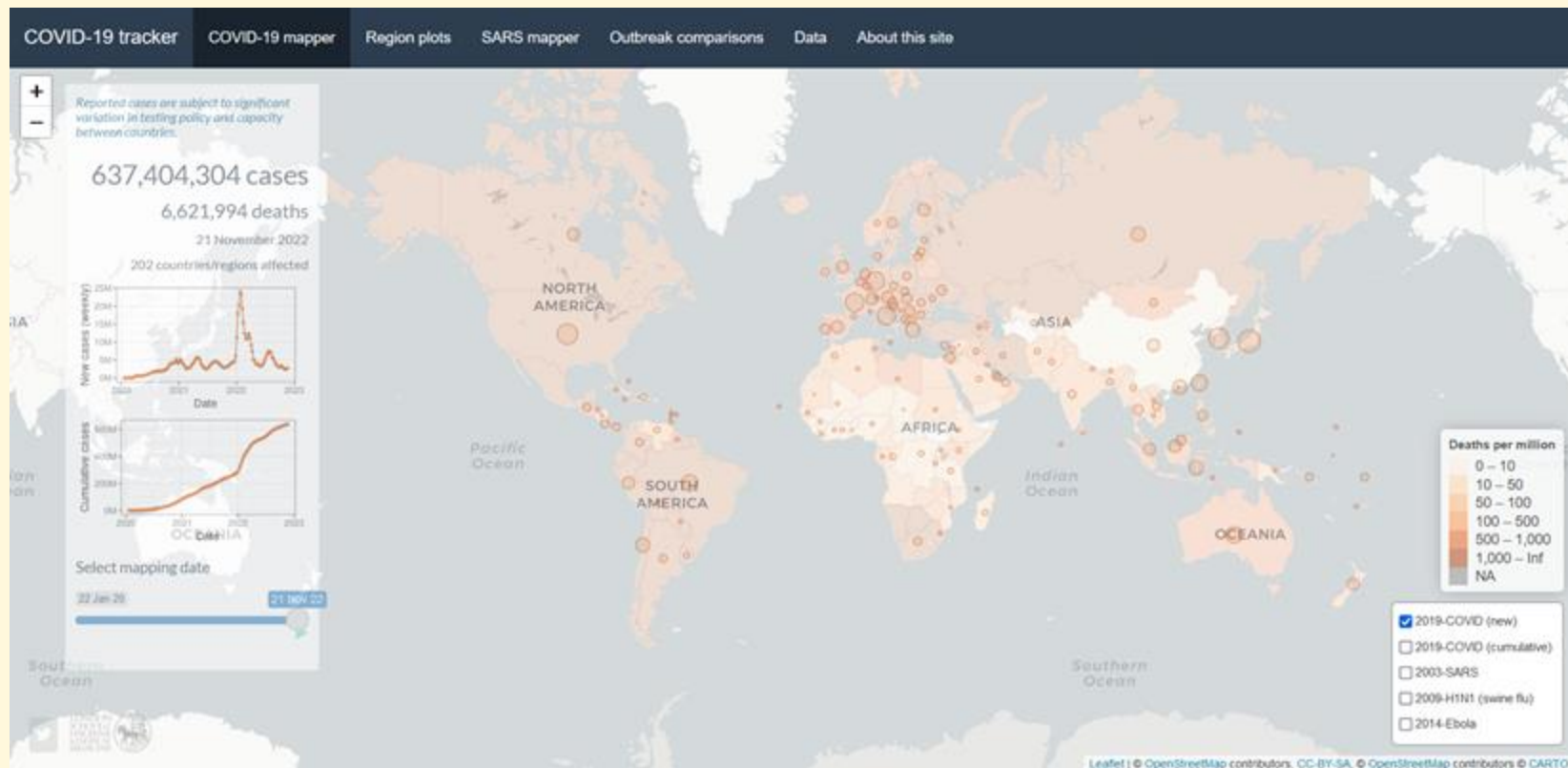


Tableau

Narzędzie do wizualnej, opartej na podejściu Data Discovery analizy 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

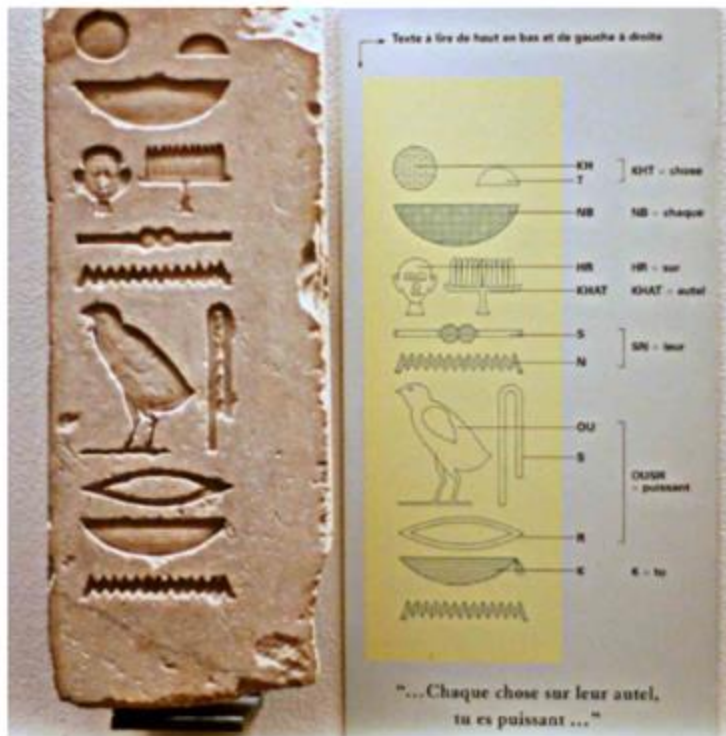


Historia grafiki statystycznej

Zaczniemy bardzo daleko...



Hieroglify



Tekst należy czytać
od lewej do prawej,
z góry na dół.

= rzecz

= wszystkie

= ołtarz

= ich

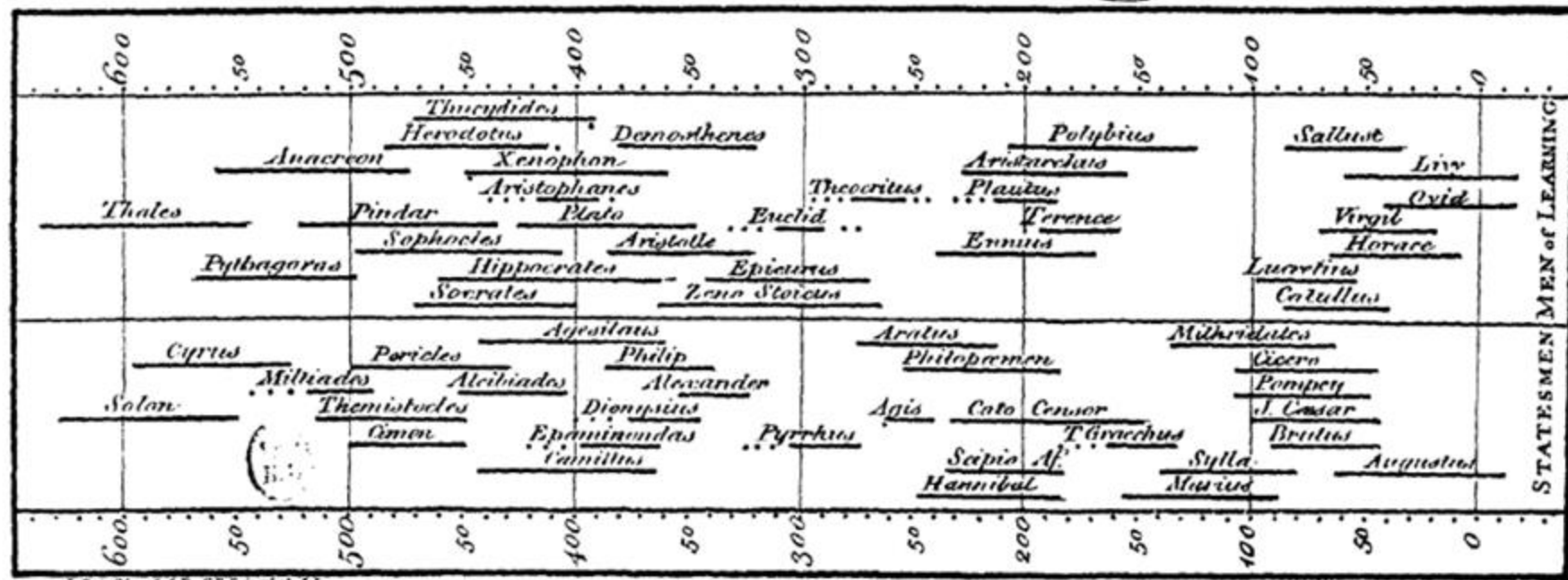
= potężny

= ty

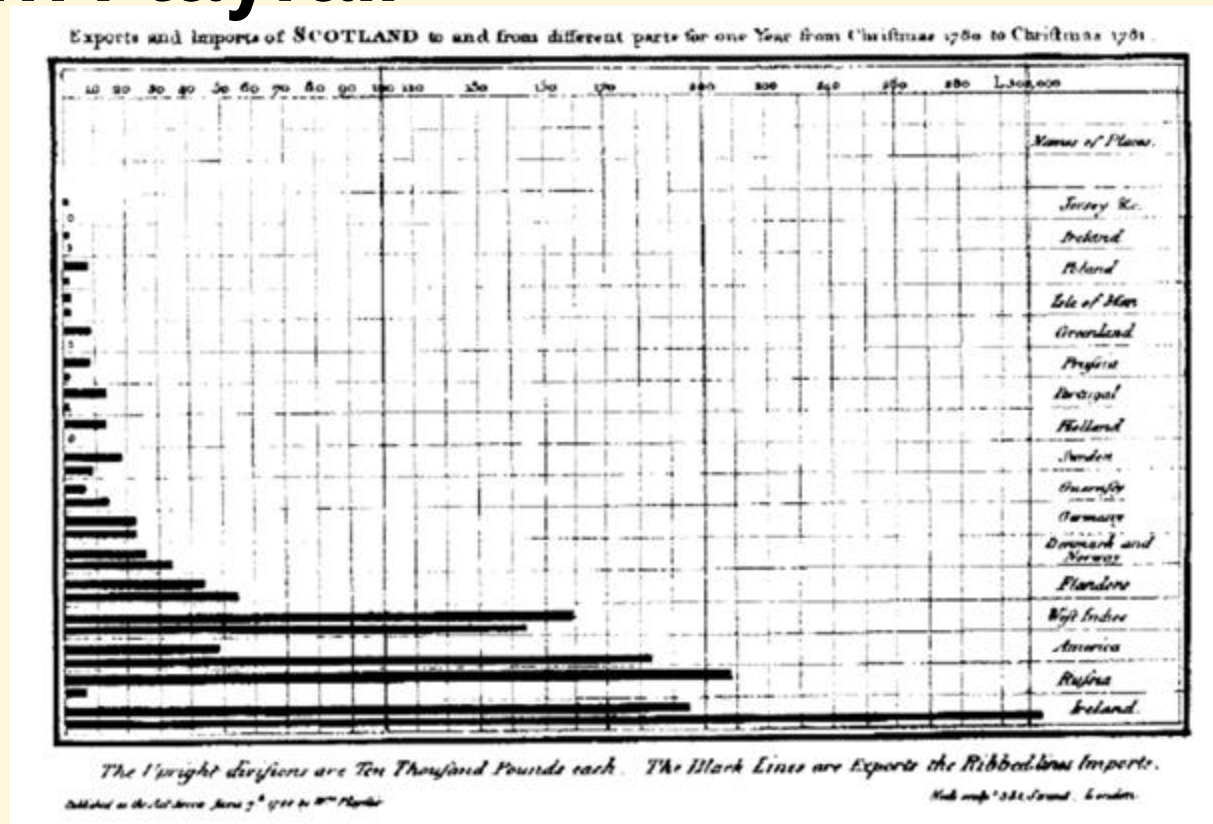
= Wszystko na ich ołtarz,
jesteś potężny

Joseph Priestley

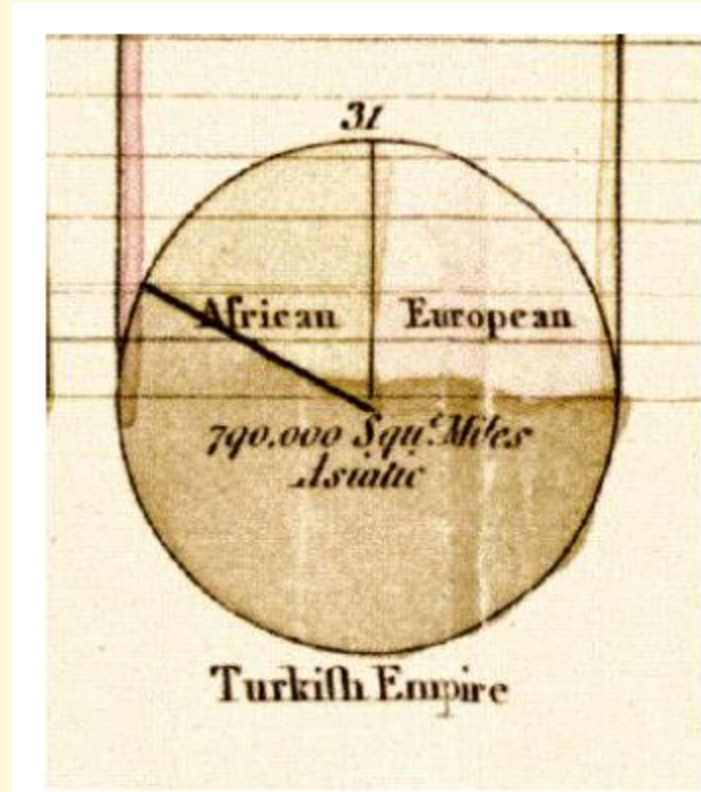
A Specimen of a Chart of Biography.



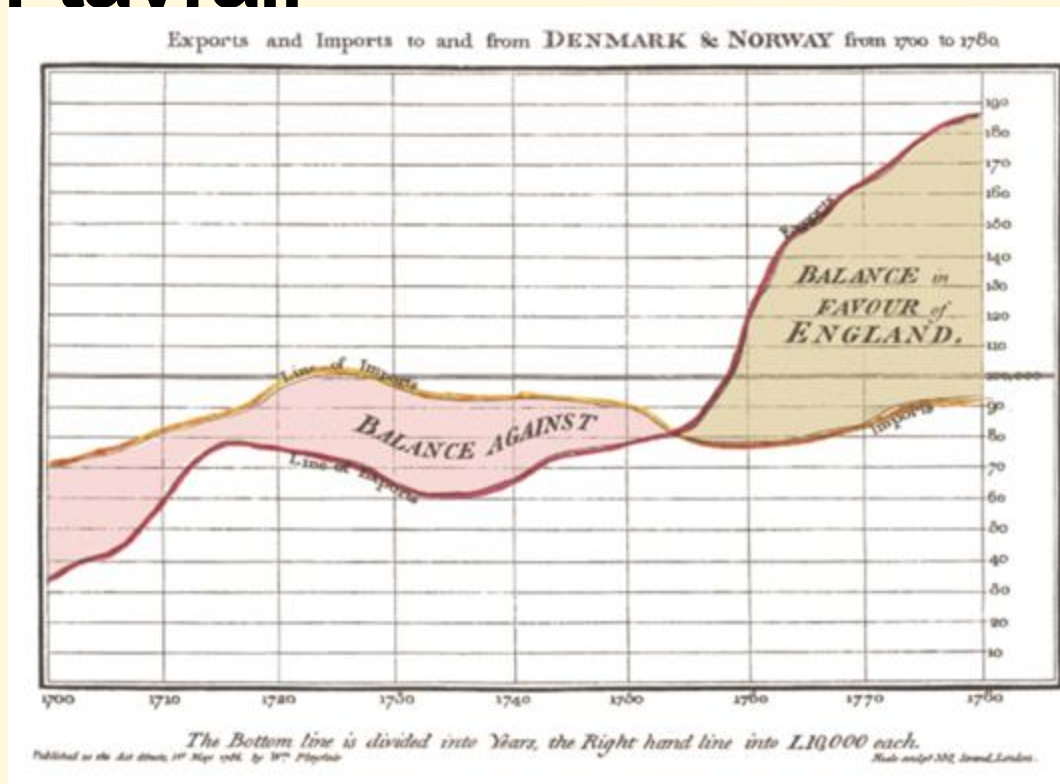
William Playfair



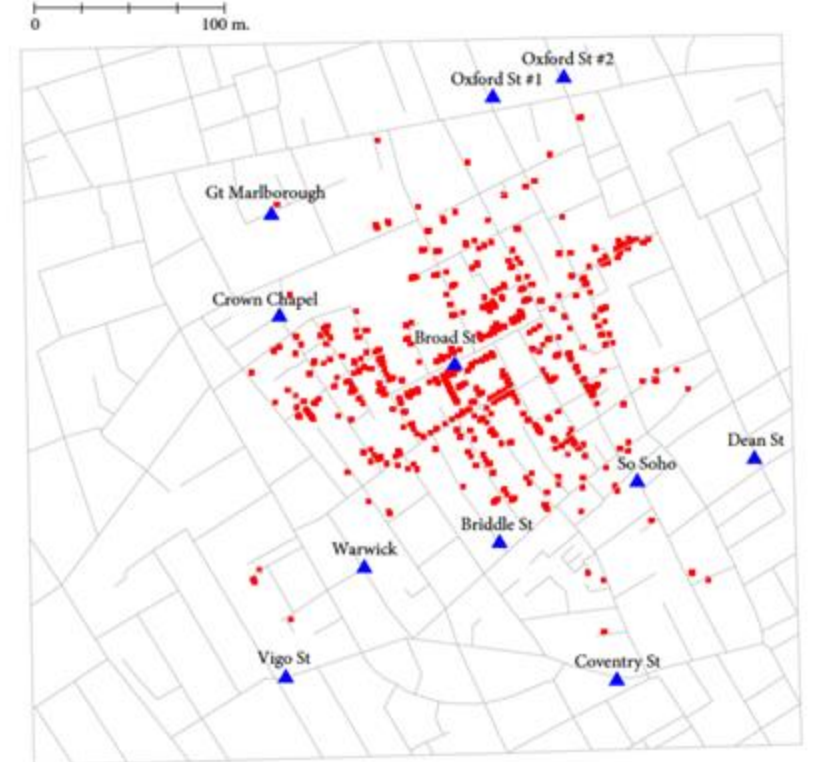
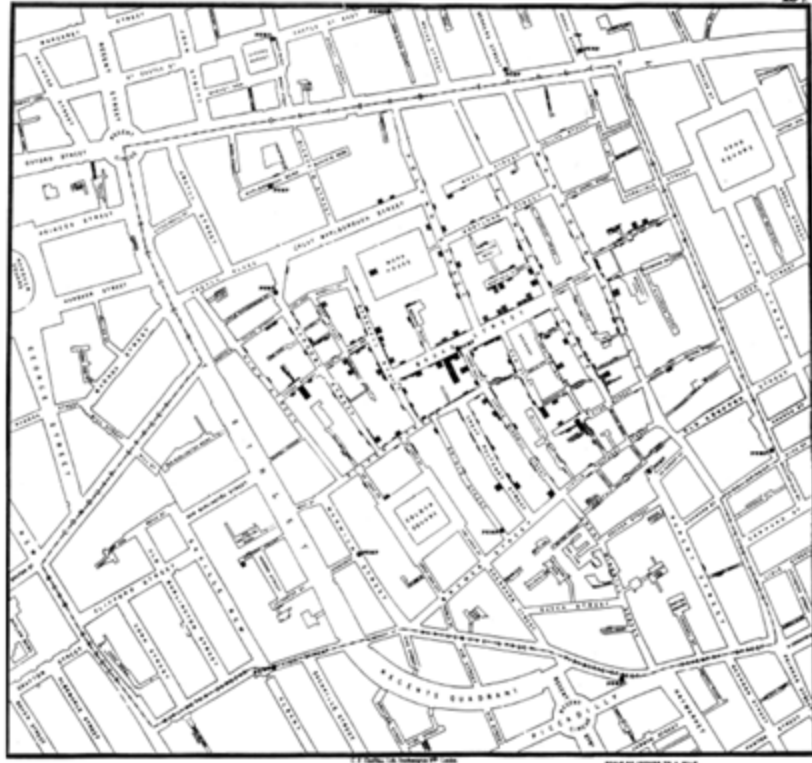
William Playfair



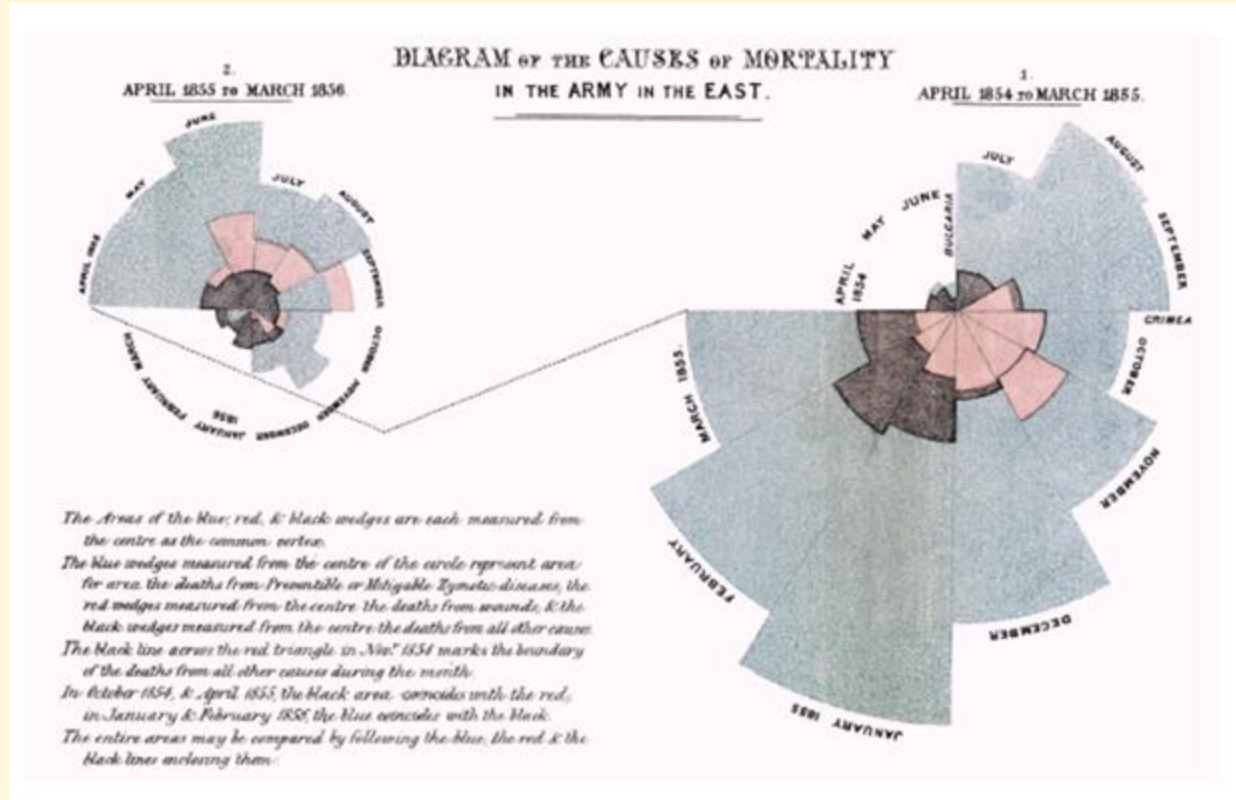
William Playfair



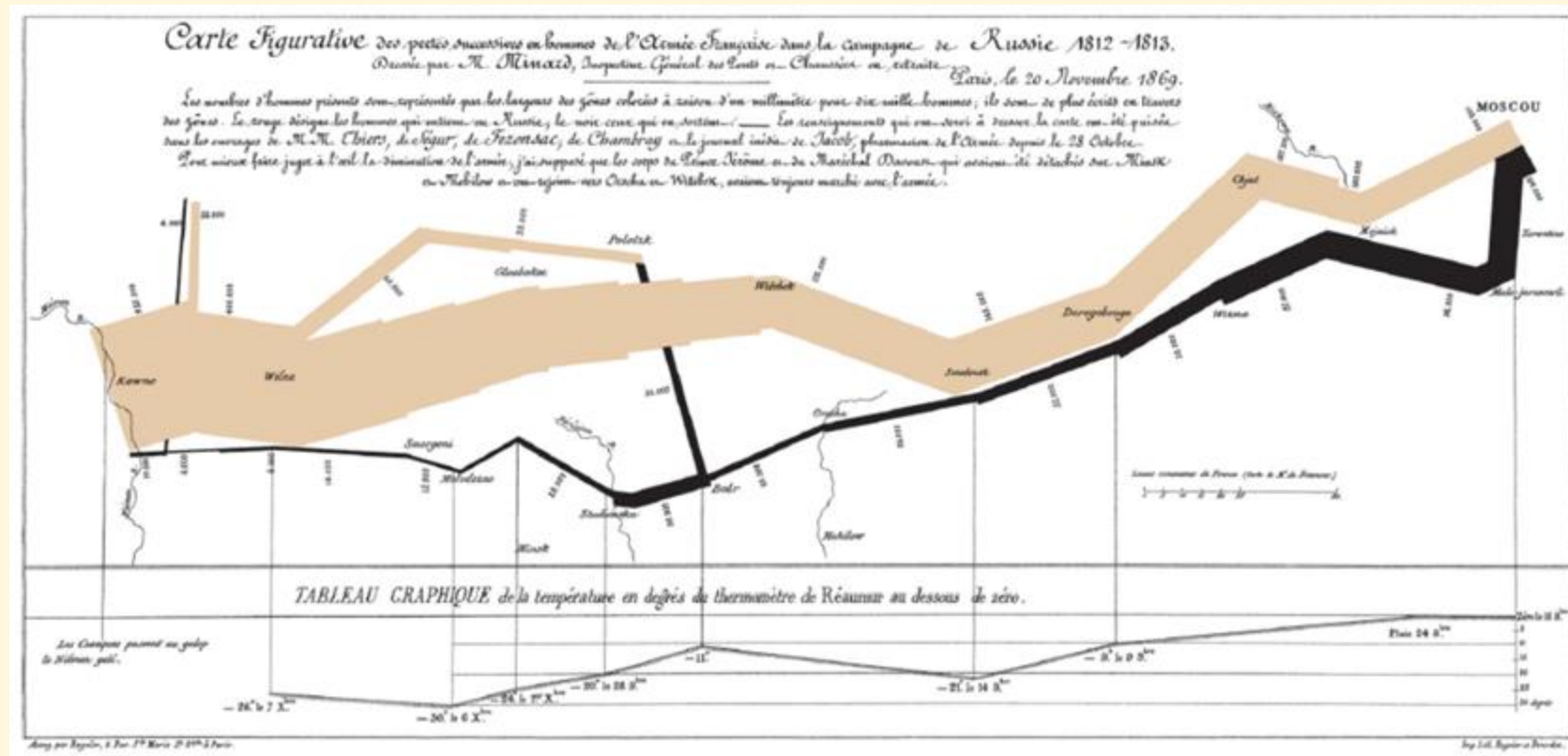
John Snow



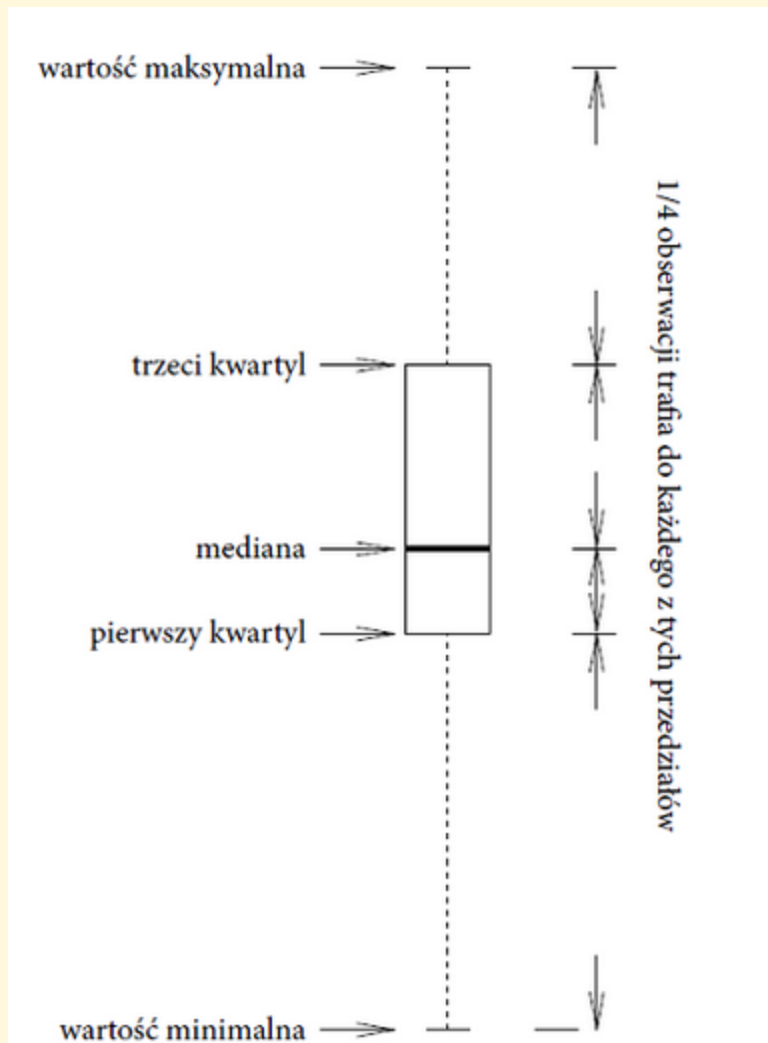
Florence Nightingale

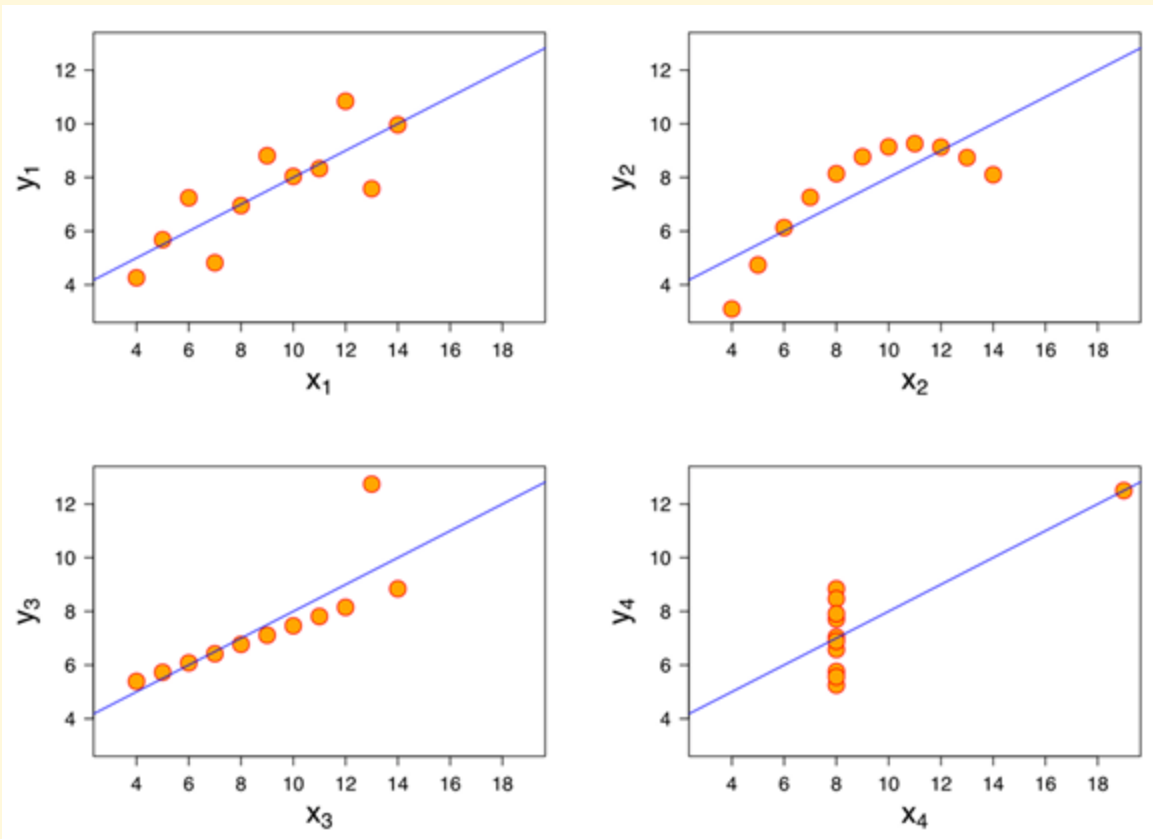


Charles Minard



John Tukey





Kwartet Anscombe'a

Edward Tufte

www.edwardtufte.com

