



Tomasz Jeruzalski

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych
Katedra Informatyki Gospodarczej
i Analiz Ekonomicznych
tjeruzalski@wne.uw.edu.pl

Karolina Chwalewska

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych
Katedra Informatyki Gospodarczej
i Analiz Ekonomicznych
kchwalewska@wne.uw.edu.pl

ANALIZA WIZUALNA W ADMINISTRACJI PUBLICZNEJ

Streszczenie: Artykuł przedstawia charakterystykę i zalety wykorzystania graficznej formy analizy danych potwierdzone empirycznym badaniem. Opisywane w nim trendy dotyczą wykorzystania wizualizacji w odniesieniu do najnowszych rozwiązań w dziedzinie analityki biznesowej. W artykule zostały opisane kryteria oceny wartości wizualizacji danych oraz metody wykorzystywane do uzyskiwania informacji z danych. Analiza empiryczna ma na celu wskazanie obszarów, których eksploracja może przyczynić się do stworzenia rozwiązań poprawiających sytuację np. na rynku pracy.

Słowa kluczowe: wizualizacja danych, przetwarzanie informacji, analiza danych.

Wprowadzenie

W związku z dynamicznie rosnącą ilością produkowanych w dzisiejszych czasach danych równoległe pojawia się coraz większy popyt na efektywniejsze sposoby pozyskiwania informacji z danych. Żyjemy w przełomowym momencie, ponieważ mamy większy niż kiedykolwiek dostęp do zasobów informacji. Jednak prawdziwym wyzwaniem jest uporządkowanie danych, tak aby stały się cennym źródłem wiedzy.

Bogactwo narodów nie zależy już jedynie od podstawowych, znanych od dawna czynników produkcji, takich jak ziemia, kapitał czy praca. Współcześnie źródłem uzyskiwania przewagi konkurencyjnej jest dostęp do wiedzy oraz możliwość jej szybkiego przekształcania. Pełne wykorzystanie zasobu informacji można osiągnąć dzięki rozpowszechnianiu danych wśród ludzi, którzy je rozumieją i tworzą.

Dyfuzja wiedzy jest normą, służy ogólnemu postępowi, ale tylko jeśli jest rozpowszechniana. Zwycięskimi okazują się instytucje, których członkowie są chłonni wiedzy, ponieważ to oni są źródłem innowacji.

Rozpowszechnianie wiedzy wymaga odpowiednich narzędzi. Jedną z najłatwiej przyswajalnych i prostych do interpretacji metod przedstawiania danych jest wizualizacja.

1. Kryteria oceny wartości wizualizacji danych

Wyznaczniki decydujące o wartości wizualnego przedstawienia danych można podzielić na trzy główne grupy. Pierwsze kryterium merytoryczne odnosi się do jakości przedstawianych danych. Kolejną część stanowi zbiór wyznaczników funkcjonalnych, które oceniają użyteczność, możliwość interpretacji oraz późniejszego wykorzystania zestawienia. Ostatnia grupa dotyczy aspektu analitycznego i interpretacyjnego.

1.1. Kryterium merytoryczne

Wartość informacji jest proporcjonalna do tego, jaki sens mogą im nadać odbiorcy i do jakich działań mobilizują. Najważniejszym celem analityków powinno być wydobycie i przedstawienie wartości zbieranych przez nich danych. Jest to jednak pojęcie bardzo subiektywne. Wizualizacja traci na wartości, jeżeli jest trudna do zrozumienia. Jednak określenia „trudny do zrozumienia” nie można mylić ze złożonością. Proste grafiki, które są łatwe do zrozumienia, lecz zawierają również niewielką ilość informacji, nie są wartościowe. Cenne wizualizacje pokazują złożone zjawiska, ale w taki sposób, aby odbiorca mógł łatwo je zrozumieć, poprzez np. skupienie uwagi na konkretnych aspektach lub podkreślenie wybranych perspektyw.

W przypadku wizualizacji wartość często identyfikowana jest z wrażeniem wzrokowym. Wiele z zasad przedstawionych w książce *Beautiful code* [Wilson, 2007, s. 9] można odnieść również do wizualizacji. Na przykład Kernighan utożsamia wartość kodu z jego zawartością, elegancją, efektywnością, użytecznością oraz zwięzłością (idealny kod powinien mieścić się na jednej stronie).

Tufte, profesor Uniwersytetu Yale oraz prekursor badań nad wizualizacjami, przedstawia koncepcję współczynnika *data/ink* [Tufte, 1983, s. 91-105], który zachęca projektantów grafik do zwiększania ilości tuszu na stronie, wykorzystywanego bezpośrednio do przedstawiania informacji. Im wyższy współczynnik, tym bardziej efektywna jest symbolika grafiki oraz głębsze znaczenie ma wizualizacja.

Wartość danych polega na ich głębi uzyskiwanej poprzez wyjawianie ukrytej poprzednio struktury i wzorów, które przyczyniają się do powstawania nowych pytań, inspiracji oraz zachęcania do odkrywania tematu.

Większość badaczy zajmujących się wizualizacjami skupia się na percepcyjnym i poznawczym aspekcie wykorzystania wizualizacji, zazwyczaj w kontekście interaktywnego systemu jednego użytkownika. W praktyce jednak wizualne analizy są często procesem społecznym. Uczestnicy dyskusji dochodzą ostatecznie do pewnych wniosków, uczą się od swoich partnerów. Co więcej, niektóre zestawy danych są na tyle duże, że interpretacja ich przez jedną osobę jest mało prawdopodobna. To sugeruje, że aby informacje były w pełni użyteczne przy podejmowaniu decyzji, powinny wspierać interakcje ludzkie.

Jako analitycy powinniśmy się starać, aby zachęcić użytkowników do pogłębiania wiedzy na dany temat. Adresaci powinni zrozumieć, że przedstawione dane dotyczą ich samych i tym samym wpływać na podejmowane decyzje. Podobny problem dotyczy całej sfery naukowej. Naukowcy opracowują niezwykle istotne zagadnienia prowadzące do postępu cywilizacyjnego oraz w większości przypadków ułatwiające nam życie. Jednak społeczeństwo nie docenia tych osiągnięć, tym samym nie jest chętne do wspierania finansowego działalności naukowej, ponieważ osiągnięcia przedstawiane są w sposób niezrozumiały i zbyt skomplikowany dla większości ludzi niezwiązanych bezpośrednio z tematem badań. Jednocześnie coraz częściej powstają próby łączenia dziedzin naukowych, biznesowych z obszarami artystycznymi dotyczącymi szeroko pojętej sztuki, aby przedstawiać zagadnienia z branż specjalistycznych w prosty i zrozumiały sposób, a tym samym rozpowszechniać działalność naukową.

1.2. Kryterium funkcjonalne

Kolejnym aspektem przy ocenie przekazywanych informacji jest ich użyteczność, możliwość interpretacji i późniejszego wykorzystania. W związku z tym sposób przedstawiania informacji powinien być nie tylko prosty i czytelny, ale również ułatwiający zapamiętanie przedstawionych informacji.

Nowak, profesor psychologii Uniwersytetu Warszawskiego, w swojej książce *Wyobrażeniowe mechanizmy przetwarzania informacji: myślenie przestrzenne* [Nowak, 1991, s. 15-30] porusza temat wyobrażeń jako formy przetwarzania informacji. Do opisu zjawiska wykorzystuje reprezentację w formie słownej oraz wyobrażeń. Pierwszy model przyjmuje sądy (najbliższy odpowiednik angielskiej nazwy zjawiska *propositional*) jako podstawową jednostkę reprezentacji poznawczej. W tym przypadku może to być np. stwierdzenie konkretnej właściwości obiektu, jak choćby „kanarek jest żółty”. Natomiast modele

wyobrażeń oprócz podstawowego kodu sądów uwzględniają również kod wyobrażeń, który często ograniczany jest tylko do wyobrażeń wzrokowych. Nowak w swojej książce porusza również aspekt przestrzenny jako część drugiego modelu. Swoje badania popiera innymi teoriami, które mówią m.in. o tym, że nie jesteśmy w stanie wyobrazić sobie koloru bez umiejscowienia go w konkretnym punkcie przestrzeni. Jednakże, aby podzielić się swoimi wyobrażeniami, również potrzebujemy specjalnego języka, który nam to umożliwi.

Jest wiele przykładów na to, że treści wyobrażone zapamiętywane są inaczej niż te przekazywane za pomocą słów. Podstawowe założenie dające fundament do interpretacji kolejnych wniosków przedstawił Paivio, mówiąc o zjawisku podwójnego kodowania [Paivio, 1971, s. 24].

Pierwszy eksperyment przedstawiający siłę wyobrażeń został przeprowadzony przez Kilpatricka w 1894 r. Badanie wykonano na dwóch grupach ludzi i polegało ono na zapamiętywaniu poszczególnych słów. Jednej z grup polecono, aby wyobrażała sobie słowa, które próbuje zapamiętać. Wyniki potwierdziły, że grupa, która otrzymała poradę, potrafiła lepiej odtworzyć słowa niż grupa, która nie otrzymała instrukcji. Dodatkowo potwierdzone zostało, że efekt instrukcji wyobrażenia był jeszcze silniejszy po upływie kilku dni. Efekt ten jest szczególnie silny, kiedy w wyobrażeniach elementy wchodzi z sobą w interakcję. Zjawisko to przedstawia w swoim badaniu Bower w 1970 r. [Nowak, 1991, s. 15-30].

Sposobem pobudzenia oraz właściwego ukierunkowania wyobrażeń jest wizualizacja. Ta forma prezentacji wyróżnia się na tle pozostałych dzięki możliwościom wpływania na sposób odbioru oraz przedstawiony kontekst. Symboliczne grafiki przedstawiające złożone informacje stymulują do analizy, tworzenia skojarzeń oraz doszukiwania się połączeń z dostępnymi nam źródłami informacji. Tworzenie kontekstu oraz odniesienie się do wiedzy z innych dziedzin poza matematyką i statystyką jest jednym z najważniejszych aspektów interpretacji danych. „Większość rzeczy, których chcemy dowiedzieć się o zjawiskach gospodarczych, można odkrywać i przedstawiać bez pomocy jakichkolwiek technicznych, nie mówiąc już o matematycznych, nakładek na zwyczajne sposoby myślenia i bez głębokiego wdawania się w dane statystyczne” [Schumpeter, 1933, s. 17].

Dane powinny służyć temu, aby ułatwić nam zrozumienie tematu, zamiast jeszcze bardziej komplikować. Informacje są niezwykle ważnym narzędziem. Jednak powinniśmy używać wizualizacji, aby uczyć oraz wzbudzać zainteresowanie. Tylko w ten sposób można zachęcić do pogłębiania wiedzy na dany temat. Naukowcy, statystycy, analitycy oraz graficy powinni przedstawiać historie ukryte w danych we właściwy sposób.

1.3. Kryterium analityczne i interpretacyjne

Sedláček w swojej książce pisze o tym, że matematyka stała się dzisiaj głównym przedmiotem zainteresowania ekonomistów [Sedláček, 2012, s. 297-321]. Jednak ekonomia zbudowana jest na podstawie nurtów filozoficznych, religijnych. Sedláček mówi o tym, że w centralnie planowanej gospodarce bloku sowieckiego wierzone, że zaawansowane możliwości technologiczne umożliwią „dyktowanie” optymalnych cen, które miały zastąpić mechanizmy rynkowe. Obecnie wszystkie najpotężniejsze systemy rynkowe kładą największy nacisk na modelowanie i ekonomiczne prognozowanie. W bardzo krótkim czasie nauka uważana za humanistyczną, wywodząca się z etyki, przekształciła się w matematyczną. Jednak matematyka powinna być jedynie językiem, a nie „narzędziem badawczym”. Takie podejście potwierdza również Marshall, jeden z twórców ekonomii matematycznej [Marshall, 1925, s. 30]. Jednocześnie w swoim podręczniku *Zasady ekonomiki* Marshall przeniósł swoje systemy formalne do dodatku na końcu książki. Lecz, jak tłumaczy jego uczeń Keynes, zrobił to, aby uniknąć wrażenia, że matematyka sama z siebie daje odpowiedzi na problemy z realnego świata.

W 1965 r. Fama postawił hipotezę o racjonalności rynku. Jednak Greenspan, wieloletni szef amerykańskiej Rezerwy Federalnej po kryzysie z 2008 r. ogłosił, że mylił się, będąc zwolennikiem wolnego rynku i minimalizacji ograniczeń.

Na ważny fakt zwraca uwagę Mini, przywołując postać Newtona, który wymyślił rachunek różniczkowy, aby uprościć sobie pracę przy rozwiązywaniu problemu fizycznego [Sedláček, 2012, s. 297-321]. Jednak obecnie postępujemy odwrotnie, tworząc modele człowieka (świata) i próbując dopasować je do matematyki.

W książce *The Secret Sins of Economics* McCloskey zwraca uwagę, że znaczna część współczesnej teorii ekonomicznej to jedynie intelektualna zabawa z przyjmowanymi założeniami. „Typowe sformułowanie w teorii ekonomicznej to »jeśli informacja jest symetryczna, to w grze istnieje równowaga« albo »jeśli ludzie przejawiają racjonalne oczekiwania w następujący sposób, to istnieje równowaga gospodarcza i polityka rządowa jest niepotrzebna«. (...) Podziwiam czystość matematyki i koncertów Mozarta. Jasne. Ale ekonomia ma badać świat, a nie czyste myślenie”. [McCloskey, 2002, s. 15].

W większości przypadków dzisiejsza ekonomia zawężana jest do ekonometrii. Jednak istnieje wiele dowodów na to, że przesada w stosowaniu matematyki do wyjaśniania zjawisk prowadzi jedynie do zaciemniania rzeczywistości. Leontief, laureat Nagrody Nobla z ekonomii stwierdza: „Niestety (...) bezkrytyczny entuzjazm wobec wzorów matematycznych często prowadzi do zasłaniania istotnych efemerycznych treści argumentu za robiącymi wrażenie znakami algebraicznymi (...). Żadna inna dziedzina badań empirycznych (niż ekonomia) nie

stosuje tak potężnego i skomplikowanego mechanizmu statystycznego z tak miernymi wynikami (...)" [Leontief, 1971, s. 1-7].

W tym samym duchu czesko-amerykański ekonomista Kmenta komentuje matematyczny aspekt analizy informacji: „ekonometria szeregów czasowych odrywa ekonometrię od ekonomii. Na przykład trudno sobie wyobrazić, że osoba z wykształceniem ekonomicznym powie o generowaniu PKB, że jest zależne od trendu czasowego i zakłóceń stochastycznych” [Kmenta, 2004, s. 47].

Sedláček jednoznacznie stwierdza, że żadne modele matematyczne nie pomogą analitykom przewidzieć krachu giełdowego, ponieważ żaden z nich nie jest w stanie zapisać za pomocą równań wszystkich zachowań ludzkich. Jednak jednocześnie Sedláček nie neguje matematyki ani ekonomii matematycznej. W swojej książce stara się raczej przypomnieć, że „ekonomia to coś więcej niż stosowanie wzorów matematycznych i jeśli chcemy rozmawiać o ludzkich zachowaniach, musimy znać różne podejścia ekonomiczne. Dlatego matematyka jest użyteczna, lecz niewystarczająca. Poszukiwanie kontekstu powinno być głównym zajęciem analityków. Wizualizacja danych pozwala skupić się jedynie na interpretacji przedstawianych informacji, pobudzając jednocześnie kreatywność poprzez estetyczną formę przekazu. Keynes stwierdził: „niedługo nadejdzie dzień, w którym zagadnienia ekonomiczne wrócą na należne im miejsce, a nasze serca i głowy zajmą się realnymi problemami – życia, relacji międzyludzkich, wszechświata, zachowań i religii” [Keynes, 1945-1946].

2. Wizualne metody wykorzystywane do uzyskiwania informacji z danych

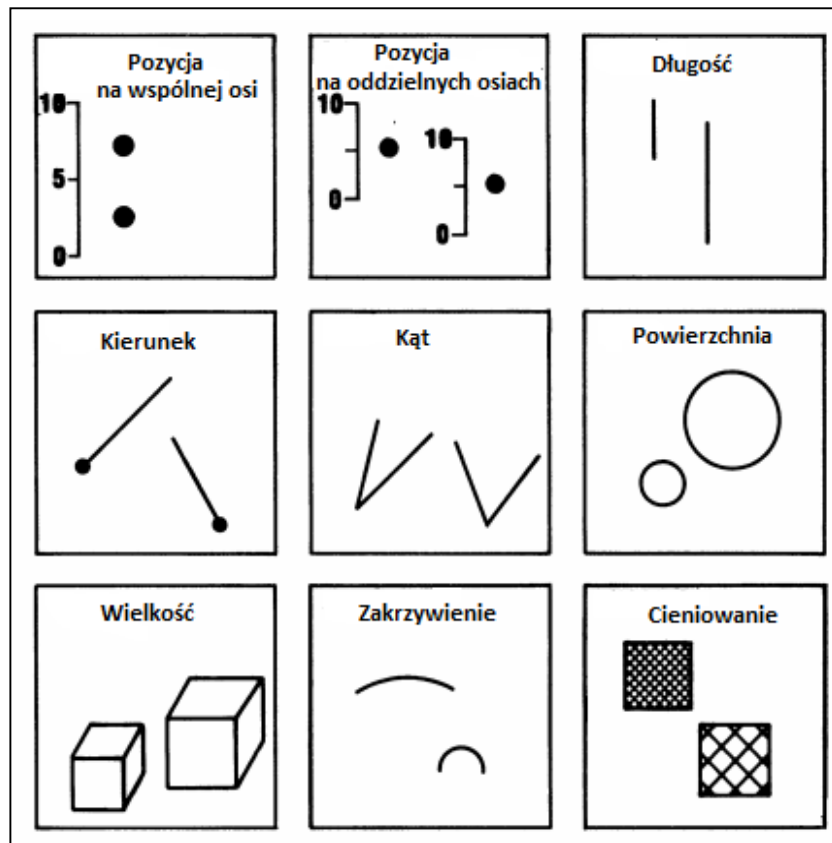
2.1. Wzorce rozwiązywania problemów

Jednym z wzorców upraszczających proces eksploracji danych, niezależnym od rodzaju oprogramowania, narzędzi, jest proces: *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Składa się on z sześciu etapów, przy czym wybór struktury następnego zależy od wyników [Larose, 2006, s. 6]:

- 1) Zrozumienie uwarunkowań biznesowych;
- 2) Zrozumienie danych;
- 3) Przygotowanie danych;
- 4) Modelowanie;
- 5) Ewolucja;
- 6) Wdrożenie.

Fakt, że ludzie są w stanie szybciej porównywać wizualne przedstawienie informacji niż surowe dane nie świadczy o tym, że wszystkie typy wizualizacji interpretowane są w ten sam sposób. Pewne typy graficznej prezentacji danych są łatwiej interpretowalne niż inne. Cleveland i McGill w swojej pracy zbadali typy prezentacji wizualnej, które są odbierane przez ludzi bardziej precyzyjnie lub mniej [Cleveland i McGill, 1984, s. 46].

Swoje wnioski przedstawili w uporządkowanej liście zilustrowanej na rys. 1. Na podstawie badania możemy wnioskować, że wiadomości prezentowane wzdłuż skali są łatwiejsze w odbiorze, niż przedstawiane za pomocą kierunku czy kąta. Najmniej zrozumiałymi wykresami są mapy termiczne. Jednak bardziej skomplikowane zestawienia, jak mapy termiczne, jednocześnie zawierają więcej informacji niż proste wykresy liniowe, dlatego nie zawsze można się kierować jedynie łatwością interpretacji.

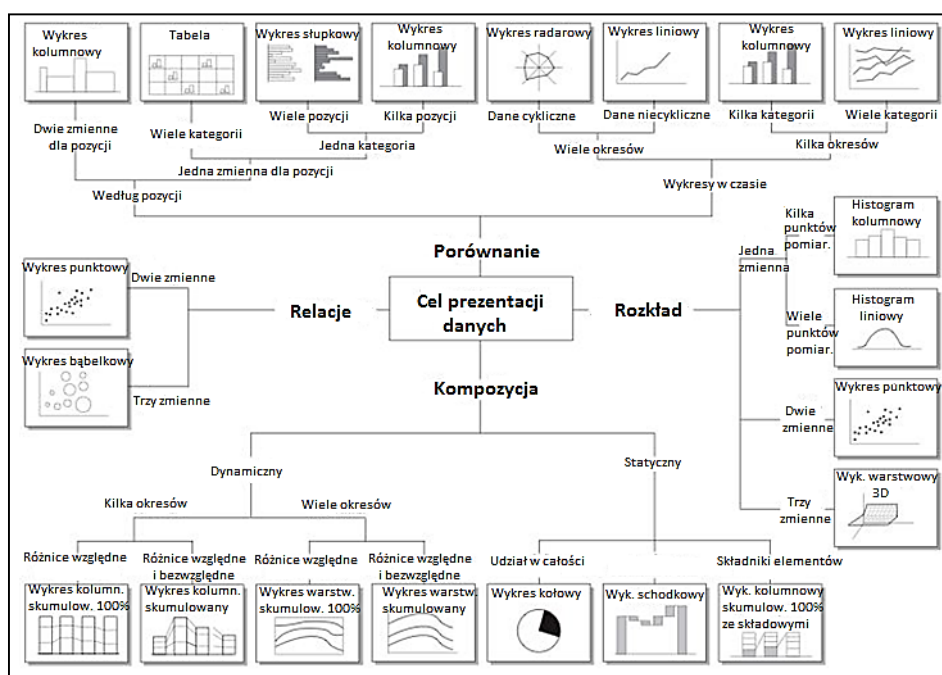


Rys. 1. Zależność między typem prezentacji wizualnej a efektywnością interpretacji

Źródło: Na podstawie Cleveland i McGill [1985, s. 13].

Powstało również zestawienie zawarte na rys. 2, na którym przedstawione zostały sugestie wykresów w zależności od rodzaju reprezentowanych danych. Według obrazu pierwszym kryterium przy wyborze odpowiedniego sposobu przedstawiania informacji jest główny cel prezentacji, tzn. porównanie, rozkład, relacje lub kompozycja. Kolejny krok polega na zdefiniowaniu bardziej szczegółowych danych, które mogą dotyczyć ilości przedstawianych zmiennych, dynamiki aktualizacji informacji, cykliczności, czy też ilości punktów pomiarowych. Dobór odpowiedniego wykresu ułatwia proces interpretacji danych.

Odpowiednie dobranie wykresu może być kluczem do prawidłowego przedstawienia danych.



Rys. 2. Sugestie wyboru wykresu w zależności od celu przedstawiania danych

Źródło: Na podstawie Simon [2014, s. 45].

W 2012 r. Lankow, Ritche i Crooks w prosty sposób udowodnili skuteczność formatowania danych [Lankow, Ritchie i Crooks, 2012, s. 67]. Przykład przedstawiony na rys. 3 został przeprowadzony jako eksperyment, którego celem było uświadomienie możliwości przyspieszenia analizy danych w zależności od potrzeb. Badanie przedstawia porównanie szybkości oszacowania występowania cyfry 7 w zbiorze. Wyróżnienie danego znaku innym kolorem zwiększa efektywność analizy kilkakrotnie.

2	1	4	3	9	5	6	7	8	2	3	6	5	9	4	0	1
6	7	9	3	4	9	0	5	6	2	5	8	4	0	5	2	6
9	8	2	6	3	5	9	3	2	9	3	7	2	6	3	4	8
8	1	6	2	3	8	7	9	5	0	2	3	9	2	8	4	3
0	9	1	8	5	4	2	9	4	7	4	6	8	4	0	2	9
3	9	2	7	3	6	6	5	2	9	4	0	4	9	4	8	6
5	2	4	3	6	4	8	1	0	3	9	4	8	4	7	3	2
8	6	2	3	0	8	7	3	6	2	5	4	4	8	3	5	0

Rys. 3. Efektywność formatowania danych

Źródło: Simon [2014, s. 35].

2.2. Eksploracyjna analiza danych

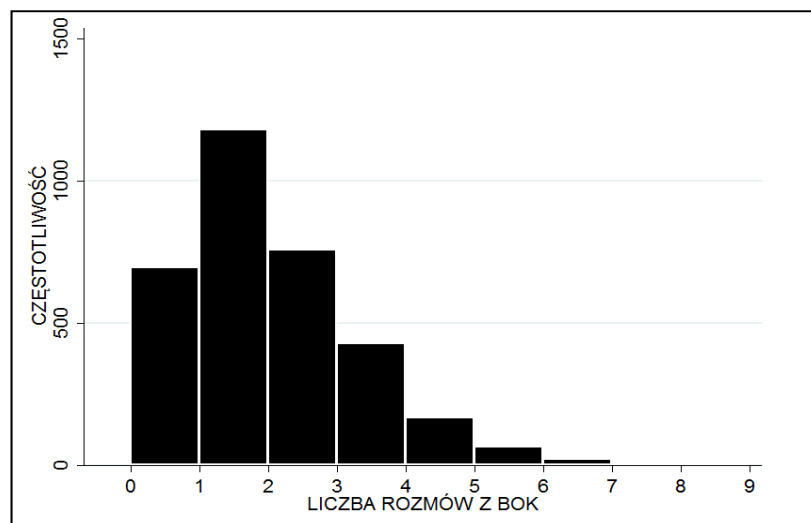
Podczas badania poszczególnych zagadnień analitycy często mają już określone *a priori* hipotezy. Konkretnie modele mają na celu sprawdzenie ich w relacji do poszczególnych zmiennych. W literaturze dostępna jest ogromna ilość statystycznych procedur testowania hipotez. Wśród nich wymienić możemy:

- test Z dla średniej populacji,
- test t dla średniej populacji,
- test χ^2 niezależności zmiennych losowych,
- analizę wariancji za pomocą testu F,
- test t dla współczynników opisujących równanie regresji,
- analizę szeregów czasowych,
- analizę testów nieparametrycznych.

Jednak często analitycy nie mają *a priori* hipotezy dotyczącej relacji między danymi. W szczególności, kiedy stoją oni przed zbadaniem dużej, nieznannej bazy danych, wolą zastosować eksploracyjną analizę danych (ang. *exploratory data analysis*, EDA) i wizualizację danych. Metody wizualne są szczególnie ważne w analizie danych ze względu na możliwości ludzkiego oka i mózgu. Pierwsza literatura dotycząca analizy danych zawiera wiele przykładów analizy metodami graficznymi. Jednak w ostatnich czasach nastąpiła pewnego rodzaju rewolucja. Dzięki zaawansowanym możliwościom technologicznym, które umożliwiają przeglądanie dużych baz danych w szybki i łatwy sposób, metody wizualne zaczęły być wykorzystywane znacznie szerzej.

Zazwyczaj proste wykresy, tabele czy rzuty, których przygotowanie zajmuje stosunkowo mniej czasu niż analiza modelu, odkrywają ważne relacje umożliwiające efektywniejsze prowadzenie dalszych badań.

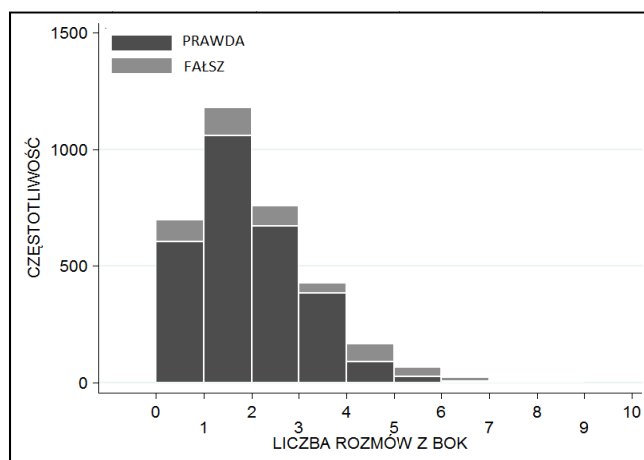
W książce *Odkrywanie wiedzy z danych* [Larose, 2006, s. 52] użyty został przykład danych z UCI Repository of Machine Learning Database na Uniwersytecie Kalifornijskim w Irwanie. Baza danych składa się z 20 zmiennych informujących o 3333 klientach, ze szczególnym uwzględnieniem informacji o tym, czy zrezygnowali z usługi, czy nie (ang. nazwa zmiennej *churn*). Analiza graficzna została wykorzystana do badania zmiennych ilościowych. Użyto trzech przykładów histogramów do uzyskania ogólnego rozkładu zmiennych. Histogram na rys. 4 obrazuje liczbę połączeń z biurem obsługi klienta. Wykres wskazuje na prawoskośność rozkładu, a moda równa się jednemu połączeniu. Jednak przedstawione na zestawieniu dane są zbyt ogólne, aby umożliwić uzyskanie wartościowych wniosków. Niezbędna jest rozbudowa wykresu poprzez dodanie zmiennych oraz normalizację. Histogramy często są również nieefektywne w przypadku zbyt małej ilości danych, ponieważ różnice pokazane na nim mogą prowadzić do zbyt pochopnych i fałszywych wniosków.



Rys. 4. Histogram liczby rozmów z BOK

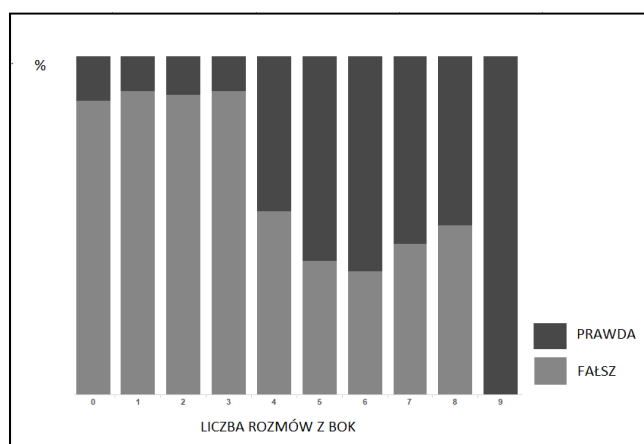
Źródło: Na podstawie danych z www.dataminingconsultant.com z wykorzystaniem programu Stata.

Na podstawie rys. 5 możemy zakładać, że procent rezygnacji może wzrastać wraz z ilością połączeń z BOK, jednak nie da się tego określić jednoznacznie. W tym celu dokonujemy normalizacji histogramu, tak aby każda kolumna (prostokąt) miała taką samą wysokość i szerokość, co zostało przedstawione na rys. 6.



Rys. 5. Histogram liczby rozmów z BOK z nałożoną informacją na temat rezygnacji z usługi
 Źródło: Na podstawie danych z www.dataminingconsultant.com z wykorzystaniem programu Stata.

Wnioski są teraz jednoznaczne. Klienci, którzy dzwoniли do BOK co najmniej cztery razy, wykazują wyraźnie wyższy odsetek rezygnacji niż klienci, którzy wykonywali mniej połączeń. Na podstawie tego badania można wyciągnąć wniosek, że powinno się dokładnie śledzić połączenia klientów z biurem obsługi, a przy trzecim połączeniu należy stosować szczególne metody zachęcające do pozostania lojalnym firmie.



Rys. 6. Znormalizowany histogram liczby rozmów z BOK z nałożoną informacją na temat rezygnacji z usługi

Źródło: Na podstawie danych z www.dataminingconsultant.com z wykorzystaniem programu Stata.

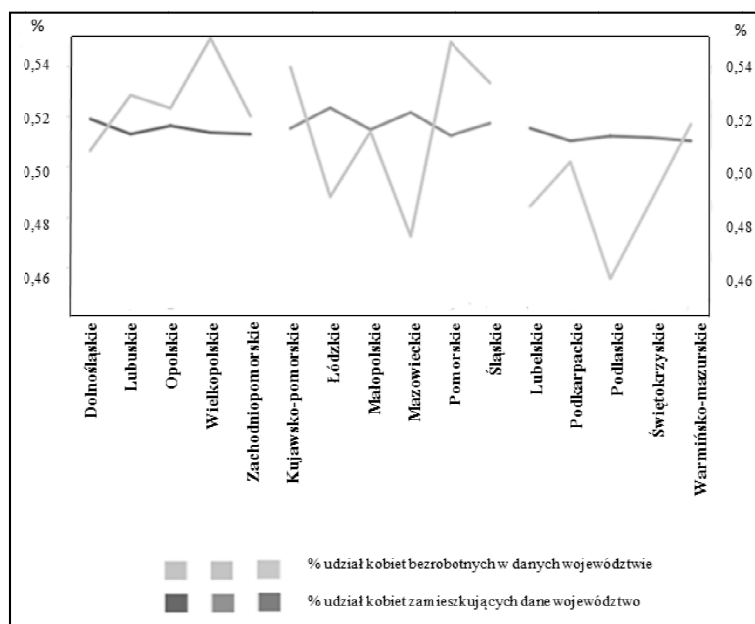
W powyższych badaniach nie zostały zastosowane żadne algorytmy eksploatacji danych. Pomimo to zebrane zostały cenne spostrzeżenia dotyczące relacji zmiennej opisującej rezygnację z usług firmy z innymi zmiennymi tylko przy zastosowaniu eksploracyjnej analizy danych.

3. Analiza rynku pracy na przykładzie poszczególnych województw Polski

Punkt ten zawiera badanie dotyczące zasobów siły roboczej w Polsce na poziomie województw. Przedstawione badanie jest analizą sytuacji poszczególnych województw Polski w 2013 r. Dane źródłowe pochodzą ze strony Głównego Urzędu Statystycznego, jednak przedstawione były do tej pory w postaci uszeregowanych kolumn i wierszy w pliku Excel.

3.1. Analiza bezrobocia na tle struktury społecznej

Pierwsze zestawienie na rys. 7 dotyczy struktury społeczeństwa oraz bezrobocia w poszczególnych województwach.



Rys. 7. Struktura społeczeństwa oraz bezrobocia w poszczególnych województwach w Polsce

Źródło: Na podstawie danych z Głównego Urzędu Statystycznego.

Dane zostały podzielone na trzy części według kryterium geograficznego. Linie, które przedstawione są ciemniejszym kolorem, przedstawiają procentowy udział kobiet w społeczeństwie danego województwa. Według zestawienia we wszystkich regionach współczynnik jest wyrównany i oscyluje wokół 52%. Natomiast linie o jaśniejszych barwach obrazują, jaki procent osób bezrobotnych stanowią kobiety.

Analiza wizualna pozwala nam na postawienie pierwszych hipotez:

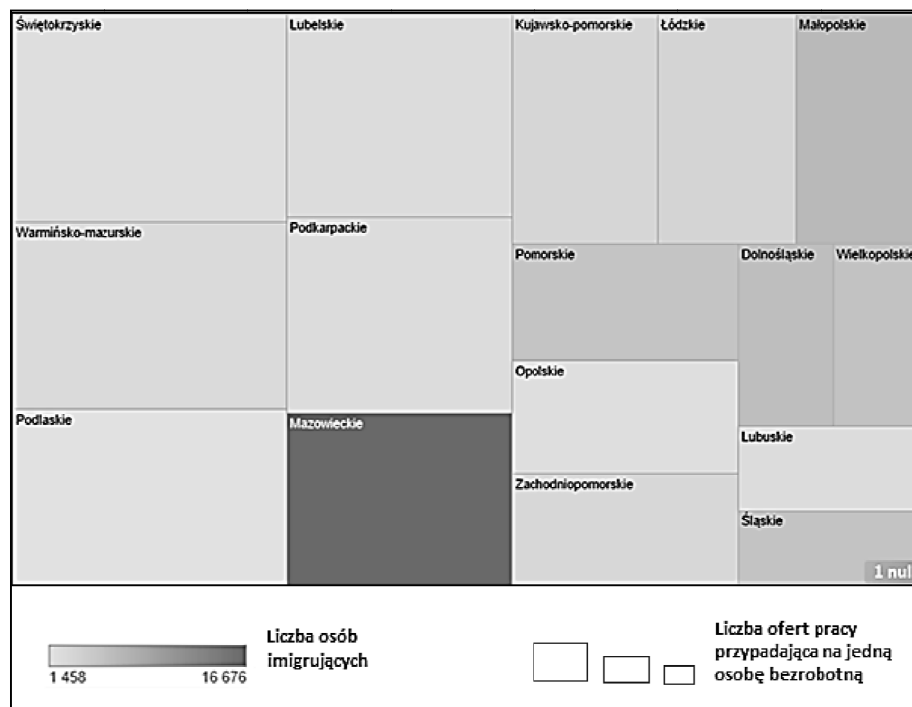
- w województwach zachodnich odsetek kobiet bezrobotnych jest znacznie wyższy niż w województwach wschodnich;
- mając jednocześnie dane dotyczące bezrobocia zestawione ze strukturą społeczeństwa danych regionów, wiemy, że różnice opisane przy pierwszej hipotezie nie wynikają ze zróżnicowania strukturalnego (tzn. większe bezrobocie kobiet na zachodzie nie jest spowodowane większą liczbą kobiet zamieszkujących ten region).

Postawione hipotezy są podstawą do dalszego badania cech wyróżniających osoby zamieszkujące dane regiony i poszukiwania prawdziwych przyczyn występujących różnic.

3.2. Analiza czynników determinujących migrację siły roboczej

Kolejna wizualizacja jest to zestawienie dotyczące liczby zarejestrowanych bezrobotnych przypadających na jedną ofertę pracy oraz imigracji osób w wieku produkcyjnym. Wielkość prostokąta obrazuje liczbę osób bezrobotnych przypadających na jedną ofertę pracy (im większy prostokąt, tym więcej bezrobotnych przypada na jedną ofertę pracy w danym regionie). Natomiast odcień figury odnosi się do liczby osób w wieku produkcyjnym, które imigrowały do danego województwa (ciemniejszy kolor oznacza większą liczbę osób imigrujących). Celem analizy tego zestawienia jest badanie korelacji liczby osób przypadających na jedną ofertę pracy i przybywających tam osób w wieku produkcyjnym.

Na podstawie rys. 8 można stwierdzić, że najgorsza sytuacja, czyli największa liczba osób przypadających na jedną ofertę pracy występuje w województwach: świętokrzyskim, warmińsko-mazurskim, podlaskim, lubelskim, podkarpackim oraz mazowieckim. Najmniejsza liczba osób na jedną ofertę pracy przypada w województwach śląskim, lubuskim, wielkopolskim, dolnośląskim i małopolskim. Najwięcej osób imigruje do województwa mazowieckiego, małopolskiego, dolnośląskiego, wielkopolskiego, pomorskiego, śląskiego. Najmniej osób przybywa do województwa opolskiego oraz podlaskiego.



Rys. 8. Liczba zarejestrowanych bezrobotnych przypadających na jedną ofertę pracy oraz imigracja osób w wieku produkcyjnym w Polsce na poziomie województw w 2013 r.

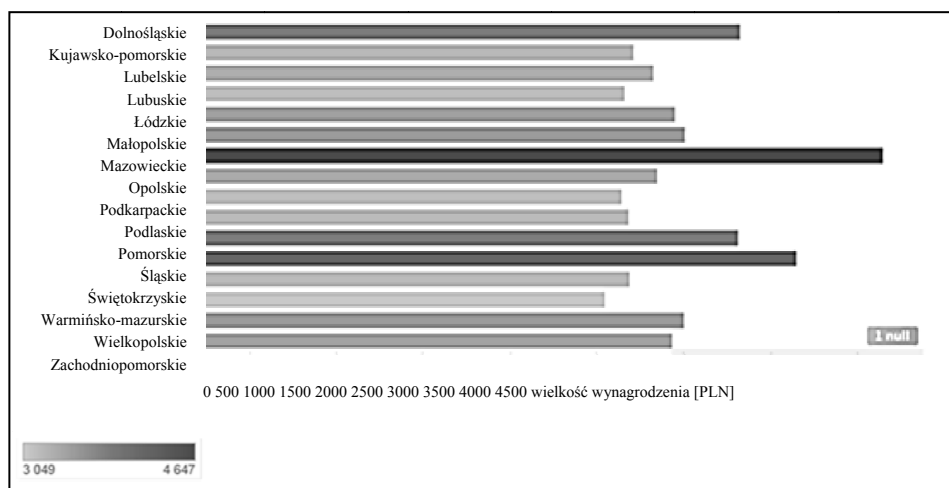
Źródło: Na podstawie danych z Głównego Urzędu Statystycznego.

Wysoka imigracja do województwa śląskiego, dolnośląskiego, wielkopolskiego oraz małopolskiego jest racjonalna pod względem stosunkowo małej liczby osób przypadających na jedną ofertę pracy. Jednocześnie niewykorzystaną szansą na znalezienie zatrudnienia jest województwo lubuskie, którego odpowiedni prostokąt na wykresie jest jednym z najmniejszych, co świadczy o małej liczbie osób bezrobotnych przypadających na jedną ofertę pracy, podczas gdy oznaczony jest również jasnym kolorem, co oznacza, że przybyło tam stosunkowo niewiele osób. Kolejną anomalią na wykresie jest województwo małopolskie, którego sytuację ilustruje duży prostokąt, co świadczy o dużej liczbie osób przypadających na jedną ofertę pracy, jednocześnie wyróżnia się najwyższym poziomem imigracji spośród wszystkich województw.

Jednak nie można oceniać postępowania migrantów jedynie na podstawie liczby osób bezrobotnych przypadających na jedną ofertę pracy. Na decyzję o przymieszczeniu wpływa również wiele innych czynników. Może to być również migracja w celu dalszego kształcenia. W związku z tym warto przedstawić te dane

również w kontekście innych miar. Kolejne zestawienie przedstawia przeciętne wynagrodzenia w sektorze przedsiębiorstw w poszczególnych województwach, im ciemniejszy kolor słupka, tym wyższy poziom wynagrodzenia brutto. Możemy sprawdzić, czy ta zmienna wyjaśnia anomalie występujące w poprzednim badaniu dotyczące województwa lubuskiego i mazowieckiego.

Na podstawie rys. 9 można stwierdzić, że najwyższe wynagrodzenia otrzymują mieszkańcy województwa mazowieckiego, natomiast najniższy poziom przeciętnej pensji występuje w województwie lubuskim, podkarpackim, świętokrzyskim oraz warmińsko-mazurskim. Zestawienie wyjaśnia anomalie dotyczące województwa mazowieckiego, które pomimo wysokiej liczby osób przypadających na jedną ofertę pracy wciąż przyciąga imigrantów wysokością zarobków. Natomiast niskie zarobki w województwie lubuskim wyjaśniają, dlaczego pomimo niewielkiej ilości osób przypadających na jedną ofertę pracy znikoma liczba osób w wieku produkcyjnym decyduje się na migrację do tego regionu.



Rys. 9. Średnie wynagrodzenie w poszczególnych województwach w Polsce

Źródło: Na podstawie danych z Głównego Urzędu Statystycznego.

3.3. Wnioski z przeprowadzonego badania

Na podstawie przeprowadzonego badania można stwierdzić, że w województwach zachodnich odsetek kobiet bezrobotnych jest znacznie wyższy niż w województwach wschodnich, przy czym nie wynika to ze struktury społecznej.

Druga część badania wskazuje na wysoką znaczącą zależność pomiędzy liczbą osób imigrujących do danego województwa, a liczbą osób przypadających na jedną ofertę pracy. Jednak nie jest to jedyne kryterium decydujące o kierunku mi-

gracji. Anomalia występująca w województwie mazowieckim i lubuskim wyjaśnia szczególny charakter wynagrodzenia.

Otrzymane wnioski są podstawą do prowadzenia dalszych, bardziej szczegółowych badań dotyczących zasobów siły roboczej. Uzyskane wyniki badania były niezauważalne przy wykorzystywaniu tradycyjnych metod analizy danych. Jest to potwierdzenie faktu, iż graficzne metody prezentacji informacji pozwalają na uzyskanie dodatkowej wiedzy oraz dają możliwość postawienia hipotez badanych bardziej zaawansowanymi, zmatematyzowanymi metodami.

Podsumowanie

W dzisiejszych czasach powstaje coraz więcej narzędzi umożliwiających analizę danych. Jednym z nich jest wizualizacja danych.

Główną zaletą wizualizacji jest możliwość odkrywczą analizy opartej na danych. W artykule przedstawiono studium praktyczne. Badanie dotyczące rynku pracy w Polsce ukazało strukturę bezrobocia oraz czynniki determinujące kierunki migracji. Takie informacje mogą przyczynić się do podjęcia kroków profilaktycznych w celu zapobiegania niepowołanym zjawiskom i poprawy sytuacji na rynku pracy.

Interpretacja dużych zbiorów danych zwykle nie jest prosta. Sposób analizy zależy od rodzaju oraz sposobu dalszego wykorzystania wyników. Wizualizacje nie są w stanie przedstawić wszystkich wymaganych informacji. Jednak wizualizacja danych pozwala na określenie głównych obszarów, którymi należy się dalej zajmować i umożliwia postawienie pierwszych hipotez.

Literatura

- Cleveland W., McGill R. (1985), *Graphical Perception and Graphical Methods for Analyzing Scientific Data*, American Association for the Advancement of Science.
- Keynes J.M. (1945-1946), *First Annual Report of the Arts Council*.
- Kmenta J. (2004), *Review of A Guide of Econometrics by Peter Kennedy*, „Business Economics”, Vol. 39/2.
- Lankow J., Ritchie J., Crooks R. (2012), *Infographics: The power of Visual Storytelling*, John Wiley & Sons, Hoboken.
- Larose D. (2006), *Odkrywanie wiedzy z danych*, WN PWN, Warszawa.
- Leontief W. (1971), *Theoretical Assumptions and Nonobserved Facts*, „American Economic Review”, Vol. 61.
- Marshall A. (1925), *Zasady ekonomiki*, M. Arct, Warszawa.

- McCloskey D. (2002), *The secret signs of economics*, Prickly Paradigm Press, Chicago.
- Nowak A. (1991), *Wyobrażeniowe mechanizmy przetwarzania informacji: myślenie przestrzenne*, Zakład Narodowy im. Ossolińskich, Wrocław.
- Paivio A. (1971), *Imagery and verbal processes*, Holt, Rinehart, and Winston, New York.
- Schumpeter J. (1933), *The Common Sense of Econometrics*, „Econometrica”, Vol. 1/1.
- Sedláček T. (2012), *Ekonomia dobra i zła*, Studio EMKA, Warszawa.
- Simon P. (2014), *The Visual Organization*, Wiley and SAS Business Series, John Wiley & Sons, Hoboken.
- Tufte E. (1983), *The Visual Display of Quantitative Information*, Graphics Press, Cheshire.
- Wilson G., Oram A. (2007), *Beautiful data*, O'Reilly Media, Sebastopol.

VISUAL ANALYTICS IN PUBLIC ADMINISTRATION

Summary: The article presents the characteristics and advantages of using graphical forms of data analysis confirmed by empirical study. It presents trends of visualization in the reference to the latest developments in the field of business intelligence. The article describes the criteria for assessing the value of data visualization methods used to obtain information from the data. Empirical analysis in the article aims to identify the areas where exploration can help to create solutions that improve the situation, for example in the labor market field.

Keywords: data visualization, data processing, data analysis.