

# Podstawy zasad wizualizacji danych na wykresach i publikowania wzorów statystycznych

Opracowanie **Dariusz K. Chojecki**

Redakcja językowa **Joanna Sygit**

## Wprowadzenie

Wizualizacja danych za pomocą wykresów jest często stosowana w badaniach historycznych opierających się na źródłach masowych lub przygotowanych już statystykach. Stanowi cenne uzupełnienie treści, pewnego rodzaju materiał dowodowy, na podstawie którego można przekonać odbiorcę do swoich racji, narzucić – w dobrym tego słowa znaczeniu – swój punkt widzenia, a przede wszystkim przyciągnąć jego uwagę, ilustrując dane zjawisko np. pod kątem struktury, rozwoju w czasie czy określonych współzależności. Innymi słowy, grafika skuteczniej oddziałuje na nasze zmysły niż odrębna tabela czy tekst. Mimo jej wagi *gros* tworzonych wykresów w fazie początkowej przypomina tytułowe brzydkie kaczątko z opowiadania Christiana Andersena.

Głównym celem tego minipodręcznika jest właśnie wskazanie na te elementy grafik, które mogą być przez nas tak zmodyfikowane, by wykres nabrał również cech małego dzieła sztuki, był estetyczny, przyciągał uwagę i dawał czytelnikowi możliwość jego lżejszego odbioru, a autorowi – przekazywania istotnego dla niego przesłania. Idąc śladem Cole Nussbauera i Knaflica, wszystkie zaprezentowane grafiki zostały przygotowane w popularnym programie Excel, który nadaje się do tworzenia *brzydkich kaczątek*, ale daje też możliwości pracy nad nimi tak, by stały się *pięknymi łabędziami* lub przynajmniej starały się je przypominać (o innych programach nie tu będzie mowy, zainteresowanego czytelnika wykorzystaniem zaawansowanych narzędzi zachęcić należy do sięgnięcia po **R** i **Tableau** i wiele innych). Przesłanie jest oczywiste: przed wysłaniem tekstu z wykresami do redakcji warto najpierw popracować nad grafikami, nie sugerując się ustawieniami automatycznymi, skoro program, siłą rzeczy, nie zna naszego zamysłu. W grę wchodzi również dobór odpowiedniego diagramu, najbardziej adekwatnego do naszych danych i tego, co chcemy przedstawić. Autor tego minipodręcznika ma za sobą również długą drogę

dochodzenia do zrozumienia, że forma nie powinna przerastać treści, a jedynie jej służyć.

Co do zasady, grafika nie funkcjonuje w pracach naukowych bez tabel i treści o charakterze narracyjnym. Żeby ją stworzyć, najpierw potrzebujemy danych, które są z reguły przedstawiane w formie tabeli. Ta zaś nie zawsze pojawia się we właściwym tekście czy w aneksie, stając się niekiedy cichym bohaterem naszej opowieści graficznej. Oczywiście słowo mówione lub tekst są niezbędne do tego, żeby przedstawić interesujący nas problem. Te elementy, szczególnie pierwszy, nie będą przedmiotem naszego zainteresowania. W przedstawianych zaleceniach nie zostanie bowiem podjęty temat, jak wykorzystywać wykresy podczas prezentowania wyników badań publiczności na seminariach, konferencjach, wykładach etc., co powinno cechować się pewną dynamiką i strukturą opowieści. Uwaga natomiast zostanie zwrócona na statyczne elementy grafik, czyli takie, które tworzą zamknięty, ukształtowany obraz zjawiska, osadzone w tekście naukowym. Prelegent może rozwijać swoją opowieść stopniowo podczas wykładu, dochodząc do finalnej grafiki, autor zaś jest niejako zobowiązany przedstawić obraz zjawiska bez stopniowania napięcia czy zwracania uwagi na poszczególne elementy na odrębnych slajdach. Wynika to nie tylko z faktu, że wykres z reguły powinien mieć eksplanacyjny (dowodowy, wyjaśniający) charakter, gdy poszukiwania określonych prawidłowości badacz ma już za sobą, lecz również z samych ograniczeń narzucanych przez redakcje bądź wydawnictwa co do ilości grafik i objętości dzieła. Obok aspektu finansowego ważną rolę odgrywa tu zachowanie określonych proporcji pomiędzy właściwym tekstem i innymi elementami w publikacji. Problem ten powoli traci na znaczeniu dzięki możliwości zamieszczania różnego rodzaju danych badawczych w repozytoriach przeznaczonych do ich gromadzenia. Tak czy inaczej omawiane tu formy wykresów będziemy traktować jako samodzielne części pracy naukowej, które powinny być również komunikatywne dla odbiorcy bez znajomości narracji tekstowej. Wynika z tego banalny wniosek (nie zawsze respektowany), że diagramy powinny być opatrzone tytułem, w którym znajdziemy odpowiedzi na następujące pytania: *Kto?* lub *Co?*, *Gdzie?*, *Kiedy?*, a ponadto: *Według czego?* Pierwsze trzy pytania są związane z cechami stałymi badanej zbiorowości. Na ich podstawie można np. w sposób jednoznaczny zidentyfikować populację, która jest przedmiotem naszego badania, za sprawą czego „staje się” ona unikatową, niepowtarzalną. Atrybuty te posiadają



wszystkie jednostki wchodzące w skład zbiorowości, natomiast różnią się one cechą lub cechami zmiennymi (w skrócie: zmienne), które wprowadzić należy do tytułu po słowie *według* bądź po wyrażeniu *w podziale na*. Oczywiście nieco inną strukturę będą miały już tytuły wykresów obrazujących rozwój zjawisk w czasie bądź współzależności. Niemniej jednak powinny także zawierać wymienione cechy stałe, by jednoznacznie identyfikować przedmiot badań. Temu ostatniemu, szczególnie w badaniach historycznych, służy również wskazanie źródła, na podstawie którego został przygotowany dany wykres. W tym wypadku jego zapis może być różny w zależności od tego, czy mamy do czynienia z archiwaliami, ankietami (rzadkość), statystykami urzędowymi czy opracowaniami etc. Nie zwalnia to nas jednak z obowiązku, żeby podać źródło danych w taki sposób, by odbiorca mógł je bez problemu zidentyfikować oraz w razie potrzeby dotrzeć do niego i jego treści (o ile dane nie są objęte jakąś specjalną klauzulą poufności).

### Wybór wykresu

Skoro już wiemy, jaką zbiorowość osób, faktów czy zdarzeń będziemy przedstawiać na naszych diagramach i na podstawie jakich danych to zrealizujemy, następnym naszym krokiem związanym z wizualizacją zjawiska – przy założeniu obrania jasnego celu badawczego – powinien być dobór właściwego wykresu. Ta umiejętność, wbrew pozorom, nie jest przedmiotem powszechnego nauczania. Zdarza się, że autorzy tworzą swoje grafiki bez głębszej refleksji nad nimi, nad ich zasadnością. Istnieje kilkadziesiąt typów wykresów i rzeczywiście może to przyprawić twórcę o pewien zawrót głowy. Warto zatem wskazać na te, które są najczęściej stosowane w wizualizacji zdecydowanej większości danych. Wśród nich należy wymienić takie, jak: kolumnowy i słupkowy (oraz ich odmiany dla wartości skumulowanych), wodospadowy, histogram, liniowy, stokowy (*slopegraph*), punktowy i jego odmianę „ważoną” – bąbelkowy, a ponadto skrzynkowy. Dodajmy do tego jeszcze wykres o nazwie *mapa drzewa*. Jest ich już dwanaście i ograniczmy się tylko do nich w tym miniprzewodniku. Nietrudno zauważyć, że wśród wymienionych brakuje najczęściej stosowanego diagramu kołowego, którego stosowanie, nie wchodząc w szczegóły, nie jest zalecane z uwagi na jego „atrakcyjną” formę znacząco obniżającą percepcję prezentowanych na nim danych. Na końcu pracy czytelnik znajdzie propozycje wykresów przygotowane na podstawie wybranych „zwykłych” danych z XIX i początku XX wieku

odnoszących się do miasta Trzcianka położonego w północnej części Wielkopolski. Niekiedy ukazane są również warianty tego samego typu wykresu. Przy wizualizacji danych użyto określonej kolorystyki. Należy jednak zaznaczyć, że w *Przeszości Demograficznej Polski – Poland's Demographic Past* zalecamy autorom przekazywanie do druku tzw. czarno-białych grafik i granie odcieniami szarości.

Najczęściej pracujemy na danych, które obrazują strukturę lub rozwój zjawiska w czasie. O tej pierwszej w uproszczeniu można mówić wówczas, gdy dysponujemy określonymi częściami tworzącymi całość zbiorowości. Mogą to być kategorie zmiennej jakościowej, np. płci, stanu cywilnego, typu gospodarstwa domowego itp. lub warianty zmiennej ilościowej, np. liczby dzieci, wieku, czasu realizacji zdarzenia itp. Wszystkie częstości bezwzględne czy względne (np. procentowe, promilowe) występowania poszczególnych kategorii czy wariantów powinny po zsumowaniu dać całość, ogólną liczebność, a w przypadku wartości względnych np. 100 albo 1000. Jeśli dysponujemy informacjami o strukturze, to do wizualizacji danych kategoriycznych możemy wykorzystać wykres kolumnowy (wykresy 1–3), słupkowy (wykresy 4–5) lub stokowy (wykresy 10–11), a także mapę drzewa (wykres 15). Kolejność nie jest przypadkowa. Pierwszy z nich jest najczęściej stosowany, gdyż ludzki umysł potrafi najlepiej rozróżniać elementy graficzne przy porównaniach ich wysokości. Wykres kolumnowy ma orientację pionową, natomiast wykres słupkowy – orientację poziomą, i tym się zasadniczo różnią. Wprawdzie percepcja tego drugiego jest nieco słabsza, ale posiada on jedną bardzo ważną zaletę – pozwala mianowicie przy poszczególnych kategoriach umiejscawiać dłuższe etykiety, łatwe do odczytania z uwagi na poziomą orientację tekstu, która jest naturalna dla naszego oka.

Utrzymanie poziomej orientacji etykiet w przypadku wykresów kolumnowych, szczególnie gdy jest wiele kategorii lub ich opisy są dłuższe, bywa często niemożliwe. W tej sytuacji program automatycznie stosuje ukośny tekst lub sam autor zmienia go na pionowy, co sprawia, że odbiór treści etykiet staje się trudniejszy. Pojawia się tu dylemat, który typ wykresu wybrać. Pragmatyka podpowiada, by stosować kolumnowy, o ile opis tekstowy kategorii pod osią  $x$  jest łatwy do przyswojenia przez odbiorcę i nie stanowi dla niego obciążenia poznawczego. To zalecenie również obowiązuje dla danych o charakterze ciągłym (np. czas, w tym wiek, odległość, powierzchnia) lub quasi-ciągłym (np.



liczba ludności, podatki, dochody), innymi słowy takich, które mają nieskończoną lub bardzo dużą liczbę możliwych wariantów i w związku z tym wymagają zapisu przedziałowego *od–do*. W starszych wersjach programu nie było histogramu. W związku z tym osoby posługujące się Excelem były zmuszone do stosowania wykresu kolumnowego, wykonania odpowiedniego grupowania i zapisu przedziałów jako zwykłych wartości tekstowych. W nowszych wersjach wprowadzono histogram, a w jego etykietach formalny zapis przedziałów. Użytkownik tym samym został niejako zwolniony z wykonywania grupowania, tj. określania częstości występowania poszczególnych wariantów; wystarczą indywidualne dane. Należy zaznaczyć, że wizualizacja danych ciągłych wymaga tego, by reprezentujące ich poziomy kolumny przylegały do siebie. O ile w histogramie jest to uwzględnione „z automatu”, o tyle w wykresie kolumnowym – już nie (milczące założenie, że *gros* danych ma charakter kategoriyczny). Żeby rozwiązać ten problem, należy ustawić szerokość przerwy na zero. Dlaczego o tym wspominam? Otóż histogram nie daje możliwości wyświetlania składowych dla poszczególnych wariantów *od–do*. W tej sytuacji należy zastosować specjalny rodzaj wykresu kolumnowego, o czym niżej.

Często zależy nam na porównaniu nie tylko samych kategorii pod względem wartości, lecz również ich składowych, nazywanych tu roboczo elementami. Przykład: dysponujemy danymi o płci oraz głównych przyczynach zgonów. W tym wypadku można zestawiać poszczególne wartości elementów obok siebie dla danej kategorii lub wykorzystać wykres skumulowany, by ukazać w „zwarty” sposób częstość bezwzględną albo względną (np. wyrażoną w procentach) występowania głównych przyczyn zgonów, które są tu elementami. Otrzymujemy dwie kolumny podzielone na odpowiednie fragmenty. Ich porównanie pozwala wyrobić sobie pierwszy pogląd na podobieństwa i różnice strukturalne, by odpowiedzieć na pytanie, czy płeć wpływa w jakiejś mierze na interesujące nas proporcje. Stąd już tylko krok do badania asocjacji.

Przy porównywaniu – najczęściej dwóch struktur – można z powodzeniem wykorzystać wykres stokowy, który jest swoistą odmianą wykresu liniowego dla danych kategoriycznych (wykresy 10–11). W tym wypadku składowe – elementy kategorii – są zastępowane znacznikami, które reprezentują konkretną wartość udziału. Strukturę jakiegoś zjawiska (np. nastawienia wobec stosowania antykoncepcji) możemy również rozpatrywać w kontekście jego kształtowania się przed lub po zaistnieniu określonego faktu, zdarzenia czy działania. Znaczniki



elementów o tym samym znaczeniu są łączone linią prostą, której kąt nachylenia pozwala automatycznie ocenić kierunek i siłę zmian (spadek, stagnacja bądź równowaga, wzrost). Stosowanie tego rodzaju wykresu ma pewne ograniczenie – traci on na przejrzystości, gdy znaczna część porównywanych elementów wykazuje znaczną odmienność pod względem kształtowania się poziomu wartości i dochodzi do licznych przecięć linii. Problem ten nabiera znaczenia, jeśli liczba porównywanych elementów jest znaczna.

Do prezentowania zagadnień strukturalnych możemy również wykorzystać *mapę drzewa*, zwłaszcza wtedy, gdy zamierzamy zwizualizować wiele kategorii zawierających ponadto elementy, których liczba może być również znaczna (wykres 15). Na myśl przychodzą tu dane o liczbie miejsc urodzenia ludności w podziale na nadrzędne i podrzędne jednostki administracyjne. Główne kategorie mogą stanowić tu parafie czy powiaty, a elementy – miejscowości. Nazwa tego typu wykresu jest nieco zwodnicza. W rzeczywistości chodzi o prostokąt podzielony na mniejsze prostokąty, które też mogą być podzielone na jeszcze mniejsze prostokąty. Ich obszar odpowiada wartości danej kategorii oraz elementów. Algorytm tworzenia wykresu pozwala zorientować się w hierarchii naszych danych. Od lewego górnego rogu do prawego dolnego patrząc, od najwyższej do najniższej liczebności, są ukazywane kategorie (w odniesieniu do całości) i ich elementy (w odniesieniu do danej kategorii). To ciekawe rozwiązanie umożliwia m.in. obserwowanie rozkładu proporcji wartości elementów w obrębie poszczególnych kategorii. Natomiast przy porównywaniu elementów o tym samym znaczeniu, ale funkcjonujących w różnych kategoriach, możemy natrafić na trudności w ocenie ich wkładu w całość. Wynika to z faktu, że prostokąty o tej samej lub zbliżonej powierzchni mają często różny kształt, co oczywiście jest związane z dążeniem do ich wkomponowania w całość. Przedstawienie tego rodzaju danych może być również wykorzystane do porównań w czasie, np. w dwóch momentach czy okresach (dwa odrębne wykresy zestawione jeden pod drugim).

Do porównań rozwoju zjawisk w czasie, dynamiki służy przede wszystkim wykres liniowy, na którym można zaprezentować także więcej niż jedną serię danych, np. kształtowanie się elementów ruchu naturalnego (wykresy 7–9). Stosując go, należy bezwzględnie pamiętać o tym, że jednostki czasu – miesiące, kwartały, półrocza, lata, dekady etc. – powinny odznaczać się jednorodnością, tzn. być takie same. Ich mieszanie ze sobą w szeregu czasowym wypacza bowiem

trend zjawiska. Tak też dzieje się w przypadku, gdy interwały albo momenty czasowe nie są jednakowe, np. 1686, 1687, 1688, 1701, 1705, a jednocześnie użytkownik w *typie osi* nie wybiera opcji *oś daty*, co pozwoliłoby uwzględnić wpływ czasu na rozwój analizowanego zjawiska. Często zdarza się, że po jej zastosowaniu znikają, jak za dotknięciem czarodziejskiej różdżki, gwałtowne wzrosty lub spadki. Mogą one również być wynikiem zmian terytorialnych jednostek administracyjnych (scaleń czy podziałów), dla których są agregowane dane. To także należy mieć na uwadze przy wizualizacji rozwoju zjawiska w czasie. Idealna sytuacja to taka, gdy mamy do dyspozycji kompletne dane dla jednorodnych jednostek czasu i niezmiennych swego obszaru jednostek administracyjnych. Jeśli jest inaczej, warto zastanowić się, czy przygotowany wykres liniowy będzie poprawnie prezentował zjawisko, czy wypaczenia jego nie będą zbyt wielkie, a tym samym obraz tendencji – pozorny. Jeśli jednak zdecydujemy się na przedstawienie grafiki, wówczas musi być ona opatrzona stosownym komentarzem. Przybliżenia wymaga jeszcze jeden problem związany z jednostkami czasu. Żeby program mógł je poprawnie interpretować, trzeba je zapisać w formacie liczbowym albo daty, dzięki czemu w *typie osi* można wybrać opcję *oś daty*. Niestety, gdy mamy do czynienia z zapisem tekstowym interwałów czasowych w postaci *od–do*, bardzo częstym w demografii historycznej, program nie jest w stanie rozpoznać ich znaczenia. Innymi słowy, przy niejednakowych rozpiętościach, np. 1831–1835 (5 lat), 1836–1840 (5 lat), 1841–1850 (10 lat), na wykresie Excela nie zostanie uwzględniony graficznie wpływ dłuższego przedziału czasu na rozwój analizowanego zjawiska. Pewnego rodzaju obejściem tego problemu jest zastosowanie wartości liczbowych wyznaczonych na podstawie środków przedziałów, tzn. 1833, 1838 i 1845,5. Uśrednione wartości można łączyć już liniami, zabieg ten wymaga jednak stosownego komentarza.

Warto również wspomnieć, że wykresy liniowe mogą być wykorzystywane do obrazowania sezonowości zjawisk. W przypadku wskaźników sezonowości, w obliczeniach których średni poziom analizowanych zdarzeń przyjmowany jest często za 100, wskazane jest, by w *opcjach osi* przecięcie z osią poziomą ustawić na tę właśnie wartość, dzięki czemu uzyskujemy, dosłownie mówiąc, linię odniesienia. Jest ona tworzona przez oś jednostek czasu. To, co znajduje się nad nią, przyjmuje wyższe odchylenia od średniej, to, co poniżej – niższe. Wykres ten można bez problemu przekształcić w warstwowy, pamiętając, żeby przy dwóch seriach danych, których obrysy wypełnień się przecinają, zastosować w

ich wypełnieniu barwą odpowiedni rodzaj przezroczystości. Pozwoli to dostrzec przebieg uprzednio zakrytej serii danych.

Często zachodzi potrzeba sprawdzenia, czy pomiędzy indywidualnymi wartościami dwóch zmiennych ilościowych zachodzi jakaś współzależność. Do tego celu możemy wykorzystać z powodzeniem wykres punktowy, który prezentuje znaczniki danych w układzie kartezjańskim (wykresy 12–13). Na podstawie układu wielu punktów (smugi) wnioskujemy wstępnie o istnieniu związku pomiędzy zmiennymi lub o jego braku (np. zestawienie udziału ludności polskojęzycznej i natężenia zgonów niemowląt w wybranych powiatach). Podejmujemy też decyzje o wykonaniu transformacji danych, a przede wszystkim o doborze odpowiedniej miary współzależności. Jeśli analizowanych jest wiele zmiennych, przedstawia się je na odrębnych miniwykresach w układzie panelowym. Przy takim typie wykresu trzeba zwrócić uwagę na to, że zdarza się występowanie danych o tych samych współrzędnych, np. punkt  $x = 10$  i  $y = 10$ . W tej sytuacji na diagramie jeden punkt zakryje drugi. Obraz naszego zjawiska zostanie zatem zredukowany. Aby temu przeciwdziałać, można zastosować dla każdego punktu wagi, które będą ukazywać częstość wystąpień nakładających się punktów. Przykładowo, może się tak zdarzyć, że większość punktów otrzyma wagę „1”, część „2”, a jeszcze inne, zapewne mniej liczne, wagę „3” etc. Uwzględnienie tego wymaga dodania do danych trzeciej kolumny, w której zostanie podana liczba wystąpień. Posłuży ona do skalowania wielkości znaczników punktowych wielkością kół. Co oczywiste, im większa waga, tym większa powierzchnia obiektów. Wykres bąbelkowy, stanowiący modyfikację wykresu punktowego, możemy zastosować także, gdy chcemy nadać wagę analizowanemu zjawisku, która nie wynika z liczby powtórzeń. Wspomniałem wcześniej o przykładowym zestawieniu udziału ludności polskojęzycznej i natężenia zgonów niemowląt w wybranych powiatach. Jak wiadomo, powiat powiatowi nierówny pod względem liczby ludności czy zdarzeń demograficznych. Aby to uwzględnić, można posłużyć się wagami w postaci liczby urodzeń żywych albo liczby zgonów niemowląt czy innymi, których użycie jest zasadne. Dzięki temu odbiorca otrzyma ważne, dodatkowe informacje; przykładowo bardziej istotna stanie się ocena zjawiska dla powiatu miejskiego Poznań, najbardziej zaludnionego i tym samym o największej bezwzględnej liczbie urodzeń i zgonów w grupie badanych jednostek administracyjnych, co sygnalizować będzie przypisany mu rozmiar bąbelka. W tym miejscu warto





zwrócić uwagę na dwie sprawy, a mianowicie na to, że dane dla wag powinny być posortowane od największych do najmniejszych, a ponadto obszary kół na diagramie powinny mieć widoczny obrys. Dzięki temu możemy uniknąć zakrywania jednego obiektu przez drugi bądź zlewania się ich.

Zarówno wykres punktowy, jak i histogram tworzone są na podstawie indywidualnych danych statystycznych, tzn. takich, które z reguły nie są wynikiem grupowania, ale mogą powstawać w drodze agregowania, np. liczba urodzeń, zgonów, zawartych małżeństw, stan ludności w poszczególnych parafiach. Tu jednostkami statystycznymi są jednostki administracyjne, nie zaś osoby, gospodarstwa domowe czy rodziny będące najczęściej podmiotem badań demografii historycznej. W grupie diagramów bazujących na indywidualnych danych szczególne znaczenie, obok wymienionych, ma wykres skrzynkowy, nazywany też pudełkowym (*boxplot*). Można go wykorzystać do wstępnej analizy danych (EDA), np. wieku osób, by sprawdzić lub porównać rozkład zmiennej ciągłej lub quasi-ciągłej, w tym wychwycić wartości odstające na tle innych (wykres 14). Zrozumienie jego znaczenia nie jest intuicyjne, wymaga bowiem pewnej wiedzy. Podstawą tej grafiki statystycznej, stosunkowo niedawno dodanej do Excela, są wartości przeciętnych pozycyjnych, tj. kwartyła pierwszego, drugiego (mediany) i trzeciego. Na diagramie tworzą one pudełko (skrzynkę), którego szerokość nie ma znaczenia. Dolną krawędzią pudełka jest wartość kwartyła pierwszego, górną – kwartyła trzeciego, zaś linią dzielącą – mediana. W środku pudełka znajduje się połowa obserwacji, poniżej i powyżej po jednej czwartej. Wąsy (zawiasy) skrzynki (pudełka) tworzą z reguły wartości minimalne i maksymalne. Jednak nie zawsze. Wysokość pudełka jest różnicą pomiędzy trzecim a pierwszym kwartyłem. Zgodnie z sugestią twórcy idei tej grafiki, jeśli jakaś wartość jest odległa o półtora wysokości pudełka od jego dolnej lub górnej krawędzi, wówczas jest uznawana za odstającą i oznaczana symbolem kółka. W tej sytuacji zawiasy są ustawiane na poziomie ostatniej wartości w zbiorze danych, która nie przekracza krytycznej wartości półtora wysokości pudełka. Na wykresie można również wyświetlić symbol średniej arytmetycznej, co pozwala dodatkowo ocenić wstępnie skośność rozkładu.

Ostatnim podstawowym diagramem jest wykres kaskadowy, który można również z powodzeniem zastosować do wizualizacji danych demograficznych, szczególnie do bilansu ludnościowego (wykres 6). Innymi słowy, pozwala on prześledzić stany ludności danej populacji i poziom zdarzeń demograficznych

mających wpływ na ich kształtowanie się, oczywiście poza początkowym. Dla przypomnienia, zdarzenia demograficzne, tj. urodzenia (+), zgony (-), napływ (+) i odpływ ludności (-), są składowymi bilansu ludnościowego. Ich liczbowy wymiar jest ukazywany na wykresie kaskadowym za pomocą prostokątów o odpowiedniej wysokości, przy czym stany ludności powinny przyjąć na nim kształt kolumn „usadowionych” na osi poziomej kategorii. Ażeby ich wysokość była przedstawiana od zera, należy zaznaczyć dany prostokąt symbolizujący stan ludności, a następnie w *opcjach serii danych* zaznaczyć pole *ustaw jako sumę*. Z równania bilansu ludnościowego wynika, że stan ludności w momencie  $t$  jest sumą wartości stanu ludności w momencie  $t-1$  i składowych pomiędzy obydwoma momentami czasu. Gdyby nasze dane były idealnie dokładne, wtedy ostatnia składowa, np. odpływ, byłaby połączona z sąsiadującym stanem ludności linią poziomą. W praktyce obserwujemy rozbieżność między wyznaczonym stanem populacji na podstawie bilansu ludności (z wykorzystaniem bieżącej rejestracji zdarzeń demograficznych) a stanem populacji uzyskanym na podstawie spisu powszechnego. Różnicę tę możemy również dodać do wykresu kaskadowego jako ostatnią składową; linia łącznika pomiędzy nią a sąsiednim wynikowym stanem ludności przyjmie wówczas orientację poziomą. W demografii historycznej rzadko ma się do dyspozycji dane z bieżącej rejestracji meldunkowej, w związku z czym z reguły szacuje się saldo migracji na podstawie ruchu naturalnego i stanów ludności. Wynik tych szacunków można również przenieść na wykres kaskadowy.

### Wygląd wykresu

Estetyka kojarzy nam się zazwyczaj z trudno definiowalnym pięknem, określoną strukturą i odpowiednimi proporcjami, a także doborem takich środków wyrazu, które pozwalają skutecznie przekazać idee i wywołać u odbiorcy pożądane przez twórcę odczucie, w naszym przypadku o charakterze pozytywnym, tj. niewywołujące dyskomfortu wskutek trudności, a nawet niekiedy niemożności przyswojenia treści o zbędnych elementach bądź mało właściwej formie. Jak widać, przy tworzeniu wykresów nauka spotyka się ze sztuką – autor wysłał do odbiorcy komunikat w głównej mierze oparty na percepcji kształtów, niejako predefiniowanej. Nasz umysł według zasad *gestaltu* rozróżnia obiekty na podstawie bliskości, podobieństwa, zamknięcia w przestrzeni, domknięcia, ciągłości i połączenia. Co ważne, wcale nie musimy sobie zdawać z tego sprawy, żeby właściwie interpretować grafiki, które zazwyczaj zawierają znaczki, linie,

prostokąty lub inne powierzchnie o określonym rozmiarze, kształcie czy wypełnieniu barwą, rzadziej szrafem. Na autorze jednak spoczywa obowiązek optymalnego wykorzystania tych elementów w przedstawieniu treści ich odbiorcom. Rzadko kiedy prezentujemy za pomocą wykresów jedną czy dwie wartości. Z reguły jest też więcej kategorii (wariantów) lub serii danych. Funkcjonują one w określonej przestrzeni i same dla siebie są odniesieniem. Zanim przejdziemy do ich syntetycznego opisu, a właściwie wskazania na pewne zasady, które warto respektować, by zapewnić skuteczniejszy odbiór treści, skupmy naszą uwagę na ich otoczeniu i spróbujmy sobie odpowiedzieć na kilka pytań, pamiętając o tym, że odpowiedzi są tylko pewną propozycją do rozważenia przy dopasowywaniu wykresów do naszych potrzeb.

*Jaki kształt powinien mieć właściwy obszar diagramu, czyli ten, na którym znajdują się elementy symbolizujące wartości?* Jeśli nie mamy do czynienia z wykresem punktowym, który wymaga z reguły kwadratu, zdecydowana większość grafik będzie wyglądać estetycznie przy zastosowaniu proporcji długości do wysokości o wartości 5:3 w orientacji poziomej (np. wykres kolumnowy). Relacje te są bliskie złotej liczbie znanej już starożytnym Grekom. Warto spojrzeć na ekran współczesnego telewizora lub komputera, żeby przekonać się, że tego rodzaju podejście nie jest obce projektantom. Dobrze jest, gdy tego rodzaju proporcję można również zastosować dla wykresu słupkowego. Bywa jednak, że powierzchnia wykresu przy dużej liczbie kategorii wymaga rozciągnięcia w pionie. W tej sytuacji wskazane jest, by wartość interesującego nas stosunku nie była większa w dzielniku od 5.

*Jakie wypełnienie powinien mieć obszar wykresu?* Białe. Każde inne w artykule naukowym jest zbędne, nic nie wnosi, a stanowi jedynie obciążenie poznawcze dla czytelnika. Za białym tłem przemawia fakt, że właściwy tekst publikacji naukowej jest osadzony właśnie na tym tle, a bardzo rzadko na szarym, co niekiedy dotyczy „wymków”, tj. komentarzy, wyjaśnień, dodatków, drobnych ekskursów w podręcznikach. U źródeł tego leży to, że na białym tle prawie wszystkie barwy i kształty są dobrze widoczne i rozpoznawalne. Istotne znaczenie posiada tu kontrast. Innymi słowy, szybciej przyciągną naszą uwagę czarne kropki na białym tle aniżeli na szarym.

*Które elementy wykresu mogą być zbędne?* Trudno nie zgodzić się z opinią, że wszystko to, co nie ma znaczenia dla odbioru treści lub nie jest niezbędne do

prawidłowego i szybkiego odczytania przesłania diagramu, powinno zostać usunięte z grafiki lub być umieszczone na niższym poziomie hierarchii wizualnej. W tym wypadku możemy wykorzystać zasady *gestaltu*, by redukować zbędne elementy, jeśli uznamy, że nie wpłynie to niekorzystnie na odbiór wykresu. Sytuacja ta dotyczy najczęściej różnego rodzaju obramowań, po których wyeliminowaniu i tak wiadomo, o co chodzi, a dzięki czemu odbiorca może lepiej skupić uwagę na właściwych treściach. Często zdarza się, że wykres jest przeładowany liniami siatki, także pomocniczymi. Zaleca się zatem, by korzystać z tych elementów oszczędnie lub w ogóle ich nie stosować, gdyż, szczególnie przy wykresach liniowych, stanowią one pewne obciążenie wizualne, zaburzają percepcję obserwowania tendencji. Jeśli zdecydujemy się na pozostawienie linii siatki, to nadajmy im lekki wygląd, stosując szarą barwę, niewielką grubość (włosowatość) czy przerywany kształt. Pozwoli to wysunąć na pierwszy plan rozmiary zjawiska, jego rozwój w czasie, strukturę, współzależności itp. w zależności od rodzaju wizualizowanych danych.

Żeby ukazać „precyzyjniej” rozmiary zjawiska, można stosować etykietowanie. Należy jednak pamiętać, że zbyt duża liczba wartości obniża percepcję, przytłacza odbiorcę szczegółami. Dlatego też warto rozważyć odpowiedni dobór etykiet, który, z punktu widzenia autora, będzie skutkował zwróceniem uwagi na to, co jest najważniejsze w przekazywanej treści. Oczywiście umiejscowienie etykiet z wartościami nie powinno istotnie zakłócać odbioru elementów graficznych ukazujących rozmiary zjawiska, dlatego też nie zawsze można liczyć na programowe opcje rozmieszczenia. W grę wchodzi wtedy samodzielne ich umiejscawianie, które pod względem redakcyjnym winno odznaczać się pewną spójnością (zmiana położenia, ale zachowanie podobnej bliskości względem znacznika). Rekomenduje się, by etykietowanie kolumn lub słupków znajdowało się wewnątrz tych obiektów graficznych, o ile pozwala na to ich układ. Wykres nie musi ukazywać wszystkich szczegółów liczbowych. Te powinny być zawarte w tabeli (źródło wykresu), która może być zamieszczona w strumieniu właściwego tekstu, aneksie bądź w repozytorium danych badawczych. Lekkość etykietom liczbowym umieszczonym w środku elementów graficznych można nadać, stosując dla ich wartości biały kolor, o ile pozwala na to wypełnienie stwarzające odpowiedni kontrast. Bardzo dobrze sprawdza się również pogrubienie czcionki.



Nie opowiadam się za tym, by likwidować osie, które obrazują skalę zjawiska lub stanowią linię, pod którą są umieszczone opisy kategorii, wariantów bądź jednostek czasu. Są one bowiem pewną „płaszczyzną” odniesienia umożliwiającą szybsze poznanie rozmiarów wizualizowanych danych, aczkolwiek ważne jest to, by pod względem formy funkcjonowały one na drugim planie. Jestem natomiast zwolennikiem tego, by na osiach wartości pomijać gęste znaczniki linii pomocniczych, a decydując się na znaczniki linii głównych, orientować je na zewnątrz, tj. w stronę wartości, do których się odnoszą. Te ostatnie w miarę możliwości powinny być wyrażone liczbami całkowitymi, bez zbędnych zer po przecinku, i separowane w przypadku liczb czterocyfrowych i większych. Wyświetlanie liczb na osi wartości należy dobrać tak, by pojawiło się ich około pięciu (np. 0, 25, 50, 75, 100). Skąd została wzięta liczba pięć? Otóż nasz umysł jest w stanie automatycznie wykonywać zliczanie elementów, których liczba nie przekracza pięciu, a zatem tyle, ile jest palców u jednej dłoni. Dowodzi tego eksperyment z jednakowymi kostkami do gry (przy wyrzuceniu z kubka sześciu kostek badany często nie jest w stanie podać błyskawicznie ich dokładnej liczby; padają szybkie odpowiedzi: pięć, sześć, siedem, przy czym udział nieprawidłowych odpowiedzi jest już znaczny).

Na podstawie powyższego można stwierdzić, że odciążenie funkcji poznawczych umysłu, a tym samym skuteczniejsze przyciągnięcie uwagi odbiorcy, jest możliwe wtedy, gdy liczba analizowanych elementów nie jest zbyt wysoka. Oczywiście pojawia się tu dylemat, czy warto redukować dane, by ułatwiać ich odbiór. Odpowiedź brzmi: należy poszukiwać złotego środka, iść na pewien kompromis i pamiętać, że nadmiarowe szczegóły można ukazać w innym miejscu. Pewnego rodzaju nadmiarowością jest również stosowanie dwóch osi  $x$ . I wcale nie chodzi tu o duplikowanie danych, bo tego zwyczajnie nie należy robić, ale o to, że na dwóch osiach mogą być przedstawiane dwie różne skale, dla dwóch różnych zjawisk, np. umieralności i śmiertelności. Tego rodzaju rozwiązanie jest kuszące, wygląda często efektownie, mimo to stanowi dla odbiorcy pewne obciążenie poznawcze, gdyż musi on włożyć więcej wysiłku w rozróżnienie wartości danych. W naszej kulturze czytamy od strony lewej do prawej. Wynika z tego, że skala wartości zjawiska powinna być umieszczana na grafice po lewej stronie. Jeśli istnieje zatem możliwość rezygnacji ze skali po prawej stronie, to rzecz jest zalecana, aczkolwiek nie jest to bezwzględny wymóg. Należy natomiast starać się przestrzegać zasady rozpoczynania osi od wartości zero w przypadku danych



na skali ilorazowej. Bywa, że odmienna praktyka jest związana z manipulacją danymi, gdyż dzięki temu można wywołać u odbiorcy pozorne wrażenie znaczących zmian. Oczywiście nie zawsze trzeba bezwzględnie stosować tę zasadę. Dotyczy to danych o niewielkiej zmienności, o wartościach „odległych” od zera. Mam tu na myśli np. poziom oczekiwanego dalszego przeciętnego trwania życia noworodka w regionach czy subregionach tego samego państwa w bliskich sobie przekrojach czasowych, ukazywany za pomocą wykresów pudełkowych. Tak czy inaczej to odstępstwo należy traktować jako potwierdzenie reguły wynikającej z faktu, że dane na skali ilorazowej – jak sama nazwa wskazuje – muszą mieć punkt odniesienia, by w ogóle mówić o proporcjach. Gdy punktu odniesienia nie ma, mamy do czynienia z danymi na skali przedziałowej, które pozwalają na określenie jedynie różnicy lub porządku, a zatem tego, co w przypadku liczb ma mniejszą moc poznawczą.

Nadawanie wykresom trzeciego wymiaru, tj. głębokości, i związanego z tym rzutowania, wydaje się początkującemu adeptowi wizualizacji danych (także kiedyś piszącemu te słowa) atrakcyjną formą przekazu informacji statystycznych. Oczywiście mamy tu do czynienia z imitacją trójwymiarowości, która nie wnosi żadnych wartości merytorycznych do odbioru danych, a ponadto utrudnia przybliżone odczytanie poziomu ich wartości. Mówiąc wprost, stosowanie tej formy jest zbędne, nadmiarowe i przeciąża odbiór zjawiska. Piękno tkwi w prostocie. Inaczej sprawy się mają w przypadku wizualizacji zjawiska, które rzeczywiście jest ukazywane w trzech wymiarach, np. liczba zgonów niemowląt w podziale na miesiące kalendarzowe, średnią temperaturę w stopniach Celsjusza i długość życia w dniach. W tej sytuacji uzasadnione jest zastosowanie wykresu trójwymiarowego, gdyż jego głębokość jest wynikiem konkretnych wartości zmiennej. Tak czy inaczej jest to przekształcenie przestrzeni trójwymiarowej na dwuwymiarową, które w sposób naturalny powoduje zniekształcenia, innymi słowy – niemożność w miarę dokładnej oceny wzajemnych relacji pomiędzy zmiennymi. Daje jednak możliwość wykrycia pewnych odmienności w kształtowaniu się zjawiska pod wpływem różnych zmiennych i bardzo przybliżonej oceny siły ich oddziaływania. Rzeczywiste wykresy 3D mają zatem bardziej charakter eksploracyjny aniżeli eksplanacyjny. Warto poszukiwać takich rozwiązań, dzięki którym trzy zmienne, nie mówiąc już o ich większej liczbie przekraczającej możliwości naszej percepcji, będą mogły zostać przedstawione na klasycznym diagramie typu 2D. Do tego celu można



wykorzystać np. wykres bąbelkowy. Pozwala on zwizualizować dane liczbowe w kontekście położenia znaczników w układzie kartezjańskim i ich wielkości. Natomiast na zwykłym wykresie punktowym trzecią zmienną może być kategoria, wyrażona barwą albo kształtem. Wynika z tego, że, odwołując się do położenia, wielkości, barwy, kształtu, a nawet stylu obramowania, istnieje możliwość wizualizowania więcej niż dwóch, a nawet trzech zmiennych na płaszczyźnie dwuwymiarowej. Niestety jakość tego przekazu nie jest (w pełni) zadawalająca z uwagi na niejednorodność sposobu przedstawiania danych. Należy zatem ostrożnie stosować tego rodzaju mieszane rozwiązania lub sięgać po specjalne wykresy, których zasady konstrukcji wymagają jednakże zaawansowanej wiedzy, tzn. nie są intuicyjne w odbiorze.

*Jaka powinna być barwa i kolejność elementów graficznych symbolizujących wartości?* To pytanie zależy oczywiście od typu wykresu. Najczęściej stosujemy kolumnowy i słupkowy oraz liniowy i na nich skupimy naszą uwagę. Naczelną zasadą jest to, by używać stonowanych barw, np. odcieni szarego czy niebieskiego. Wybór kolorów o intensywnym nasyceniu nie jest zalecany, podobnie jak stosowanie ich wielu różnych palet czy gradientów, których zmienność kolorów w obrębie danej kategorii jest pusta znaczeniowo. Zastosowanie tzw. ostrych kolorów, a tym samym pojawienie się na wykresie pstrokaczny, jest w odbiorze męczące, krzykliwe, niewyrafinowane (w pracy redakcyjnej miałem okazję spotkać się z diagramami, które w początkowej wersji miały kanarkowe tło i różowe kolumny, nie wspominając już o innych elementach). Ważna jest pewnego rodzaju subtelność i podkreślanie bardziej intensywnymi barwami tych elementów, na które chcemy rzeczywiście zwrócić uwagę odbiorcy. Wyróżnione elementy w przekazie wizualnym nie powinny jednak zajmować wiele miejsca, do jednej dziesiątej powierzchni, by skutecznie skupić na sobie wzrok czytelnika. Jednym słowem: umiar. Jest on ważnym elementem strategii graficznej, gdyż – powtórzmy – pozwala wyeksponować to, na czym istotnie nam zależy. Sojusznikiem jest tu również białe tło, dzięki któremu można wydobyć „na światło dzienne” elementy graficzne symbolizujące wartości. Jeśli te ostatnie zdominują tło swoją gęstością, rozmiarami czy etykietami, to przekaz naszej idei straci na swej wyrazistości, ponieważ na samym początku uwaga odbiorcy zostanie skierowana na to, co nie jest zasadniczą treścią wykresu.



Ciekawym rozwiązaniem jest stosowanie neutralnej bieli do oznaczania krawędzi kolumn histogramów czy *map drzewa*. Nadaje to większą lekkość grafice; niewprowadzenie bowiem kolejnego koloru – np. czarnego – pozwala zredukować obciążenie poznawcze. Wykorzystujemy to, co jest ogólnym tłem. To ostatnie stanowi pewnego rodzaju płaszczyznę odniesienia dla barw i kształtów; innymi słowy, scenę. Jeśli chcemy zatem, mówiąc w przenośni, wprowadzić na nią aktorów drugoplanowych, np. obok danych rzeczywistych dane prognozowane, to nie powinni oni zanadto rzucać się w oczy. Zmniejszenie intensywności barwy, grubości linii, zmiana konturu na delikatny kropkowy albo kreskowy wychodzą temu naprzeciw.

Problemy w odbiorze treści mogą pojawić się przy stosowaniu ciągłej skali barwnej z zakresu jednego koloru. Należy unikać tego rodzaju rozwiązania i sięgać po stopniowe tony, o odpowiednich składowych, które pozwalają naszemu zmysłowi wzroku dokonać odpowiedniego rozróżnienia. Nie wchodząc w szczegóły, jeśli korzystamy z barw określonego koloru, to trzeba brać pod uwagę fakt, że jesteśmy w stanie rozróżnić sąsiadujące ze sobą warianty barwne, gdy ich ogólna liczba z reguły nie przekracza pięciu (od w miarę słabego do w miarę silnego nasycenia). Warto dążyć do tego, by kolorów lub barw było stosunkowo niewiele, gdyż pozwala to zredukować obciążenie poznawcze i szybciej przyswajać treści grafiki – nie zniechęca, a zachęca do uważnego przyjrzenia się diagramowi. Z uwagi na osoby z niepełnosprawnością w rozróżnianiu kolorów warto zastanowić się nad rezygnacją z wypełnień czerwonych i zielonych na naszych wykresach.

Do ograniczenia liczby kolorów lub ich barw można wykorzystać kolejność elementów legendy, szczególnie wtedy, gdy wizualizowane są składowe poszczególnych kategorii lub serie czasowe, o ile ich liczba nie jest znaczna, nie przekracza umownie pięciu. Ważne jest, by legendę odpowiednio umiejscowić, stosując w zależności od potrzeby orientację poziomą albo pionową. W przypadku wykresów obrazujących rozwój zjawiska w czasie wykorzystanie legendy do redukcji kolorów czy kształtów jest trudniejsze, jeśli dochodzi do przecięć linii obrazujących dynamikę. W tej sytuacji lepiej posłużyć się etykietami serii danych dla wybranych znaczników, posiłkując się też liniami wiodącymi. Oczywiście nie chodzi tu o sygnalizowanie nazw serii dla każdego znacznika, przez który przebiega linia, a tylko dla jednego „reprezentanta” danej serii. Dla jakiego? Dla tego, w którego przypadku zastosowanie etykiety z linią





wiodącą będzie najczytelniejsze, dobrze wyróżni się na fragmencie białego tła, w możliwie niewielkim stopniu będzie zakłócało odbiór innych elementów graficznych. Jak widać, nie ma tu gotowej recepty, wszystko zależy bowiem od układu wizualizowanych danych i wolnej białej przestrzeni. Na tej ostatniej powinna być umieszczana zawartość ograniczonych do minimum opisów lub legendy, również w taki sposób, żeby nie wybijała się ona na pierwszy plan bądź nie przykrywała innych elementów graficznych. Niestety, nie zawsze jesteśmy w stanie spełnić ten warunek.

Kolejność elementów graficznych symbolizujących wartości na wykresach kolumnowych i słupkowych zależy zasadniczo od tego, czy dane przedstawiają zmienną jakościową (kategoryczną), czy zmienną ilościową (ciągłą, skokową) bądź quasi-ilościową (porządkową). W przypadku pierwszej możemy kierować się liczebnością kategorii lub ich składowych, porządkując je od największych do najmniejszych. Czynność ta pozwoli uzyskać bardziej uporządkowany obraz interesującego nas zjawiska, wyeliminuje lub złagodzi niekorzystny wpływ różnych poziomów danych – „poszarpanie” obrazu – na ich czytelność, a przede wszystkim da możliwość szybszego wychwycenia tego, co istotne i nieistotne. Jak zatem przeprowadzać sortowanie danych? W przypadku wykresów słupkowych dobrze zdaje egzamin umieszczenie dominującej liczebnie kategorii na samej górze, biorąc pod uwagę fakt, że czytanie rozpoczynamy od góry (sortowanie od najmniejszych do największych). Natomiast przy wykresach kolumnowych sprawdza się ulokowanie dominującej liczebnie kategorii w lewej części właściwego obszaru wykresu, skoro w naszym kręgu cywilizacyjnym piszemy i czytamy od lewej do prawej (sortowanie od największych do najmniejszych). Zaleceń tych nie można już stosować przy wizualizowaniu wartości zmiennych ilościowych lub quasi-ilościowych, gdyż tu porządek jest określany przez kolejność samych zmiennych wyrażonych punktowo czy przedziałowo, tj. od najmniejszych do największych ich wariantów (zazwyczaj) albo od największych do najmniejszych (rzadziej). Kolumny lub słupki powinny mieć jednolity kolor, np. szary; błędem w sztuce jest dobieranie różnych kolorów dla każdego z tych elementów graficznych. Zagrać kolorem należy tylko w sytuacji, gdy chcemy wskazać na jakąś wartość (wartości) odgrywającą w naszej narracji ważną rolę. Wyróżnianie – powtórzmy – sprawdza się jedynie wtedy, gdy jest oszczędnie stosowane. Jego efektywność jest również wysoka, jeśli do akcentowania stosujemy nie więcej niż dwa kolory, np. podstawowy o funkcji „tła” – szary

(zdecydowana większość elementów graficznych symbolizujących wartości), a wyróżniający – pomarańczowy lub niebieski (zdecydowana mniejszość). Ta strategia graficzna jest stosowana przez Cole Nussbauemer Knafliec, zaś oba zwracające na siebie uwagę kolory, co ciekawe, zagościły od jakiegoś czasu w domyślnym szablonie wykresów Excela.

### Słowo o wzorach

Zasady ogólne składania wzorów matematycznych są doskonale opisane w pracy Ewy Wolańskiej i in. zatytułowanej *Jak pisać i redagować. Poradnik redaktora. Wzory tekstów użytkowych*. Aby nie powielać treści precyzyjnie opisanych, ograniczymy się tu tylko do pewnego zasygnalizowania spraw związanych z formą, które przede wszystkim warto mieć na uwadze, prezentując wzory statystyczne (w szerszym znaczeniu: matematyczne). W czasopiśmie zaleca się ich wyodrębnienie z tekstu i przedstawienie w wolnej przestrzeni między akapitami z podaniem pod wzorem lub wzorami objaśnień do użytych oznaczeń, o ile ich znaczenie nie jest powszechnie znane. Wzory matematyczne powinny być przygotowane w tym samym kształcie, co zasadniczy tekst artykułu, z zachowaniem odpowiedniego kroju i rozmiaru oraz rozmieszczenia ich elementów, co umożliwia też wbudowany edytor równań w danym programie. Zdarza się, że mamy do czynienia z rozbudowanymi formułami. W przypadku przeniesień do następnego wiersza należy w nim ponowić znak operatora matematycznego (jeśli mamy do czynienia z mnożeniem, zalecane jest zastosowanie znaku „×”). Dla pewnego porządku wizualnego dobrze jest stosować w ciągach zawierających równości, a zapisanych w kilku liniach, ich wyśrodkowanie względem znaku równości. Rzecz nie jest obligatoryjna przy przekazywaniu propozycji artykułu do wstępnej oceny komitetu redakcyjnego i recenzentów, gdyż wymaga ona stosowania bardziej zaawansowanych metod formatowania, których elementem jest również stosowanie odpowiednich wcięć. Tak czy inaczej wstawione wzory muszą być edytowalne, innymi słowy – nie powinny być osadzonym obrazem.

Każdy wzór oznaczamy odpowiednią kolejną liczbą podaną w nawiasie półokrągłym bez kropki. Oznaczenie to powinno znajdować się w określonym miejscu i w określonej odległości od początku albo końca wzoru, by mogło być od niego jednoznacznie odróżnione. Często przyjmuje się też zasadę, że jest ono umieszczane przy prawym marginesie. W przypadku zapisywania we wzorach stałych wartości nie należy stosować separacji liczb. Tego rodzaju rozwiązanie



warto natomiast wprowadzać dla większych wartości ukazywanych na wykresach i tabelach, separując elementy pięciocyfrowych liczb – np. 37 251. Gdy w grupie danych pięcio- i więcej cyfrowych znajdzie się czterocyfrowa liczba, to również należy w odniesieniu do niej zastosować separację – np. 2 541. Wartości ujemne zapisujemy z użyciem półpauzy poprzedzającej liczbę bez spacji. Półpauzę również możemy wykorzystywać do oznaczenia okresów lub przedziałów: pomiędzy nią a liczbami nie dodaje się spacji. Zaleca się, by we wzorach liczby niecałkowite były wyrażane za pomocą ułamków dziesiętnych. Oczywiście ta uwaga nie dotyczy dzielenia sygnalizowanego kreską ułamkową, która musi być dłuższa od najdłuższego wyrażenia znajdującego się na niej lub pod nią. Wyrażenie ułamkowe ujęte w nawias powinno również być w całości nim objęte. Ta sama uwaga odnosi się do znaku pierwiastka. Symbole łacińskie i greckie zapisujemy kursywą. Przy ich iloczynach nie stosujemy spacji, chyba że oznaczenie składa się z dwóch lub więcej znaków, i z reguły pomijamy znak mnożenia, który również nie jest wymagany w sytuacji, gdy mamy iloczyn liczby i symbolu. Na koniec warto wspomnieć, że w wystawionym do oddzielnego wiersza wzorze, w którym jest znak sumowania, tj. sigma, oznaczenia początku i końca sumowania podajemy odpowiednio pod i nad sigma, natomiast po prawej stronie tego symbolu u dołu i góry wtedy, gdy wzór chcemy z jakichś powodów osadzić w strumieniu tekstu.

### Uwagi końcowe

Przedstawione zalecenia nie zamykają całego katalogu spraw do omówienia. Są tylko pewnym wyborem treści mających skłonić do refleksji nad znaczeniem formy w przedstawianiu różnego rodzaju danych liczbowych. Mają też zachęcić czytelnika posługującego się wbudowanymi szablonami do ich samodzielnej modyfikacji, by w możliwie najefektywniejszy sposób zaprezentować myśl przewodnią towarzyszącą wizualizacji własnych danych. Nie ma tu gotowych rozwiązań, są bardziej ogólne rady, z których czytelnik może skorzystać lub nie. Najważniejsza z nich, w ujęciu klasycznym, sprowadza się do tego, by przy tworzeniu wykresów kierować się tym, że piękno ukrywa się w prostocie, umiarze i odpowiednich proporcjach. Im więcej mało przydatnych elementów uda nam się wyeliminować z grafiki, im lepiej będą ulokowane w przestrzeni te, które symbolizują wartości zjawiska (i ich ewentualne opisy), im szybciej odbiorca skupi swoją uwagę na zasadniczym przesłaniu wykresu – tym bardziej powinniśmy być zadowoleni z efektów naszej pracy. Nim je zaprezentujemy

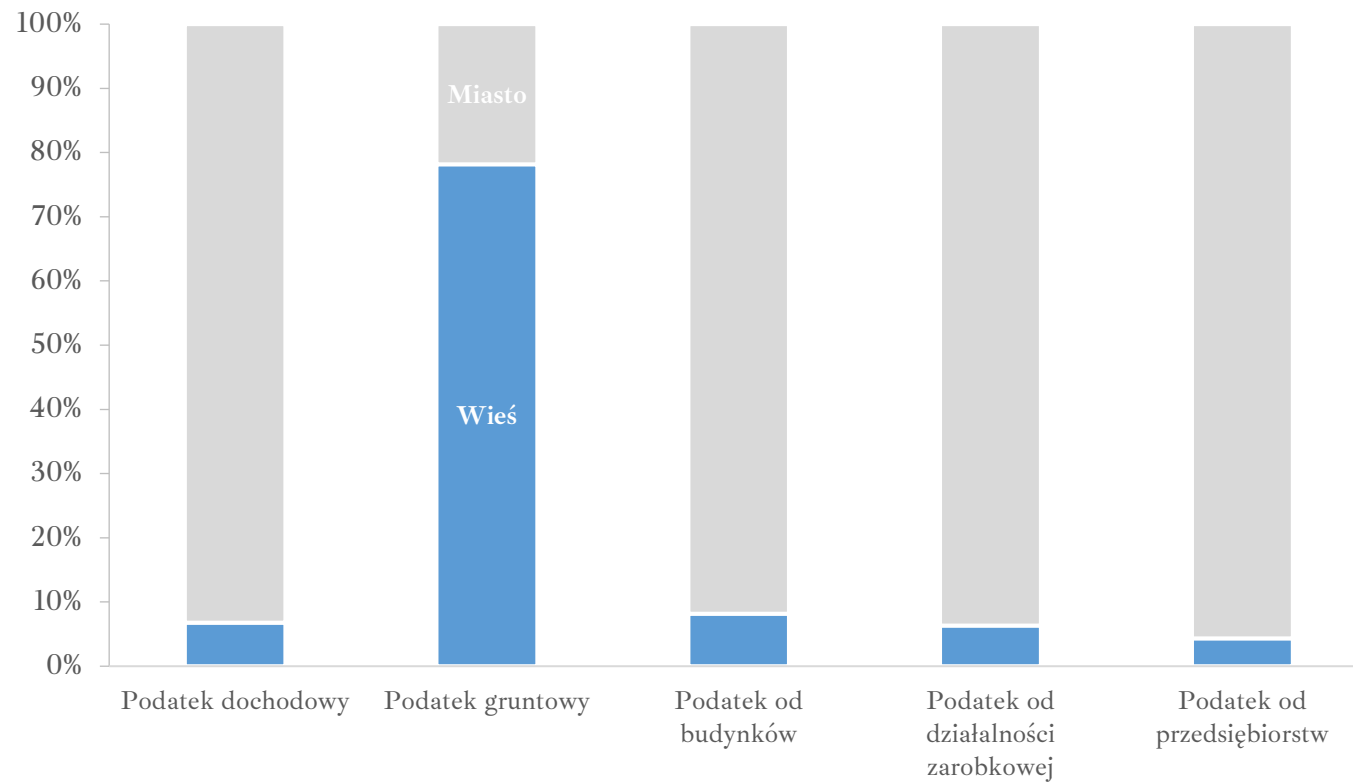
odbiorcy, warto najpierw przypatrzeć się różnym wariantom naszych grafik zarówno pod kątem doboru wykresu, jak i sposobu agregacji danych oraz samej formy ich przedstawiania. W przypadku znacznego stopnia komplikacji, szumu informacyjnego wynikającego z przebiegu i wielości danych obrazujących nasze zjawisko, śmiało decydujemy się na przekształcenie jednego wykresu w wiele ukazanych w jednym panelu. Po wybraniu najlepszego rozwiązania – jeśli jest taka możliwość, to również z pomocą pierwszych, testujących odbiorców – stosujemy konsekwentnie jego ogólne lub szczegółowe założenia do innych grafik w tekście, by były one spójne redakcyjnie, zaś same wykresy stały się w pełni samodzielnymi formami przekazu informacji.

## Podstawowa literatura

- Biecek, Przemysław. *Odkrywać: Ujawniać: Objasniać: Zbiór esejów o sztuce prezentowania danych*. Wydanie 2. Warszawa: Fundacja Naukowa Smarter Poland.pl, 2016.
- Knaflic, Cole Nussbaumer. *Storytelling danych: poradnik wizualizacji danych dla profesjonalistów*. Gliwice: Helion, 2019.
- Wilke, Claus. *Podstawy wizualizacji danych*. Gliwice: Helion, 2020.
- Wolańska, Ewa, Adam Wolański, Monika Zaśko-Zielińska, Anna Majewska-Tworek, Tomasz Piekot, red. *Jak pisać i redagować: poradnik redaktora, wzory tekstów użytkowych*. Wydanie 2. Warszawa: Wydawnictwo Naukowe PWN, 2022.

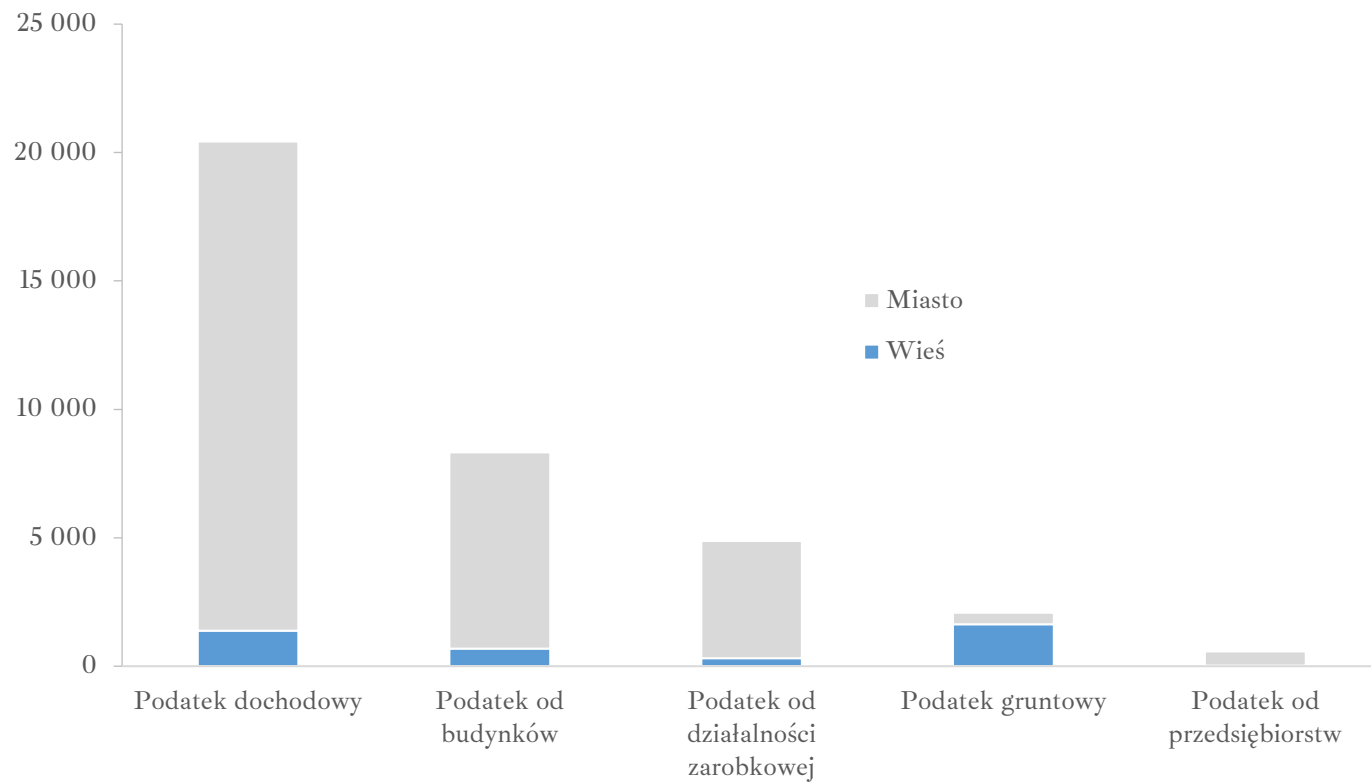
## Aneks (przykłady wykresów)

Wykres 1. Udział względny wysokości podatku w mieście i wsi Trzcianka w 1905 r. w podziale na poszczególne kategorie podatku



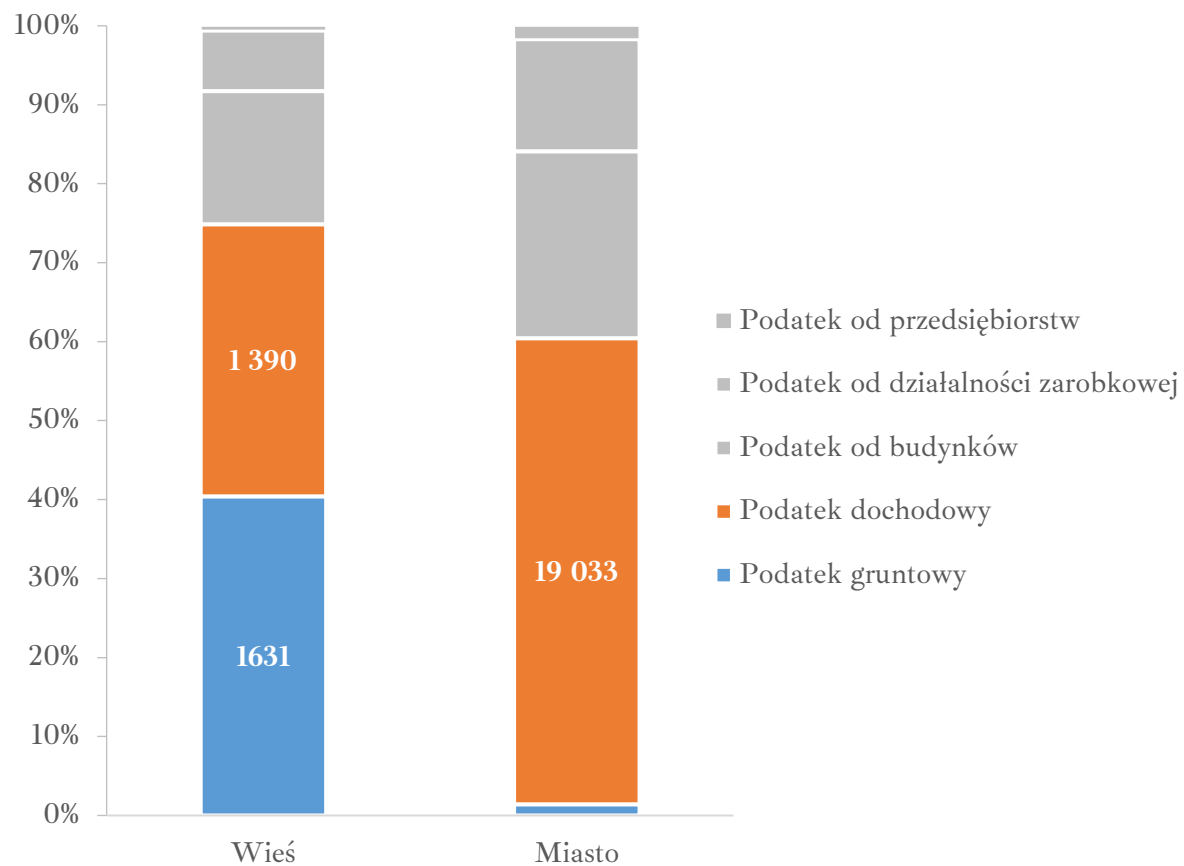
Źródło: opracowanie własne na podstawie Geheimes Staatsarchiv Preussischer Kulturbesitz, I HA. Rep. 77 Ministerium des Innern, Tit. 2686 Nr. 5.

Wykres 2. Udział bezwzględny wysokości podatku w mieście i wsi Trzcianka w 1905 r. w podziale na poszczególne kategorie podatku



Źródło: opracowanie własne na podstawie Geheimes Staatsarchiv Preussischer Kulturbesitz, I HA. Rep. 77 Ministerium des Innern, Tit. 2686 Nr. 5, nlb.

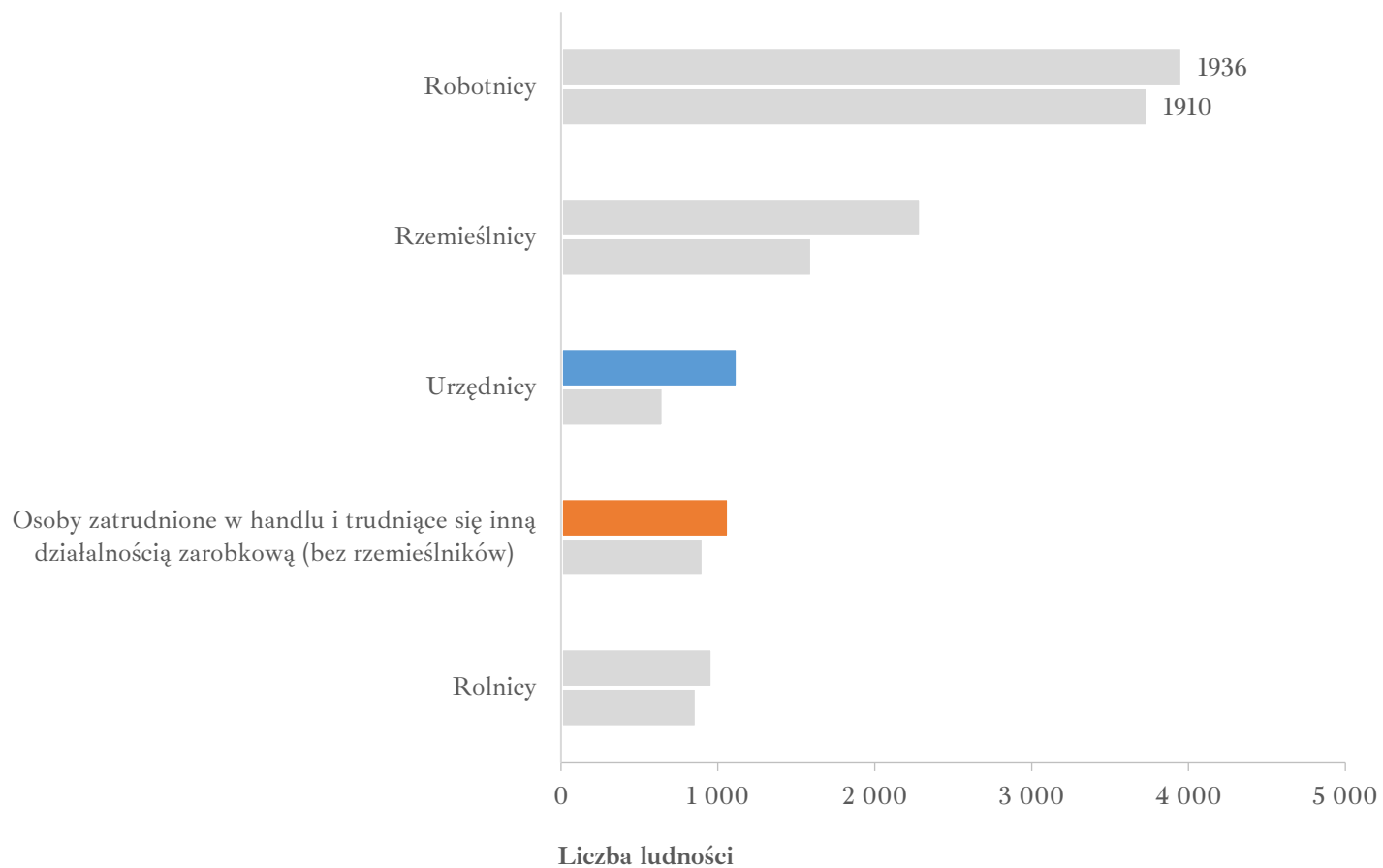
**Wykres 3. Udział względny wysokości podatku w markach w poszczególnych kategoriach w 1905 r. w podziale na miasto i wieś Trzcianka**



Źródło: opracowanie własne na podstawie Geheimes Staatsarchiv Preussischer Kulturbesitz, I HA. Rep. 77 Ministerium des Innern, Tit. 2686 Nr. 5, nlb.

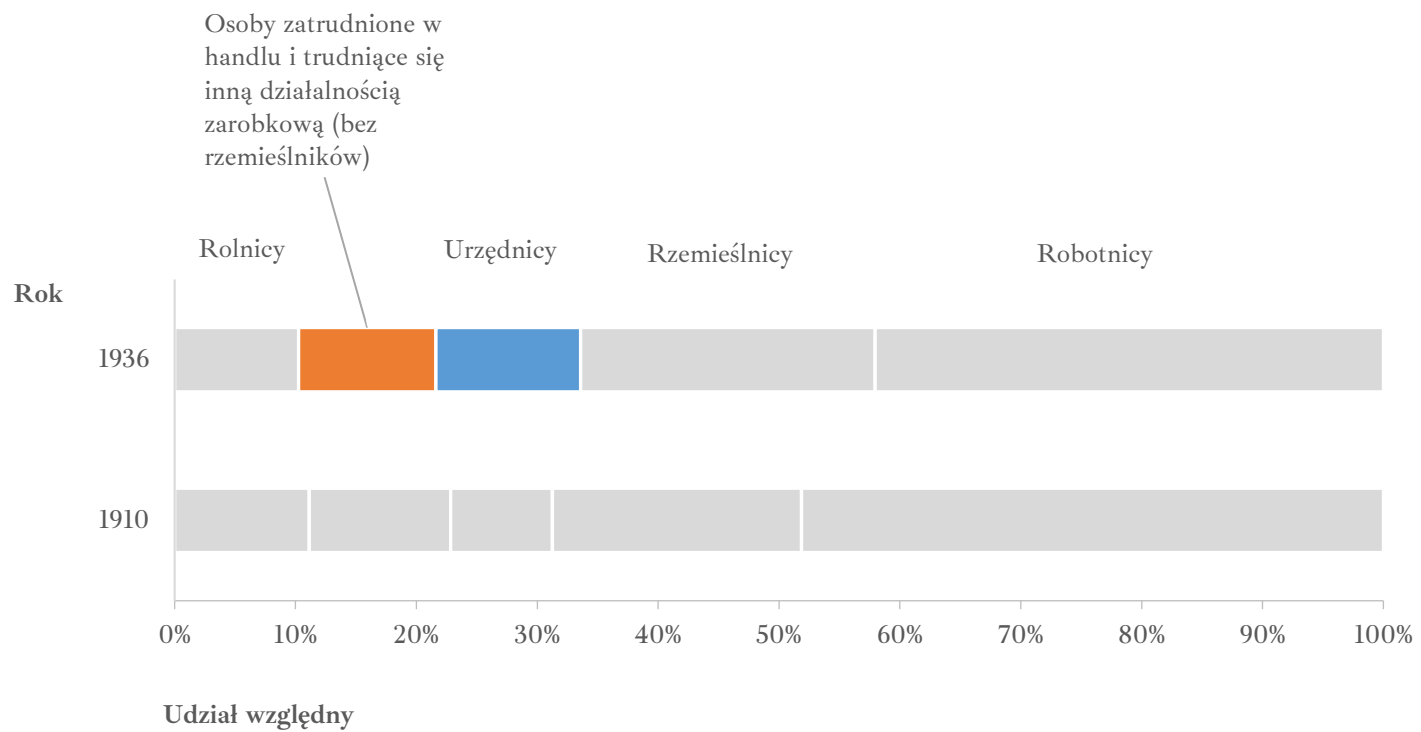


Wykres 4. Ludność miasta Trzcianka w 1910 i 1936 roku według źródeł utrzymania



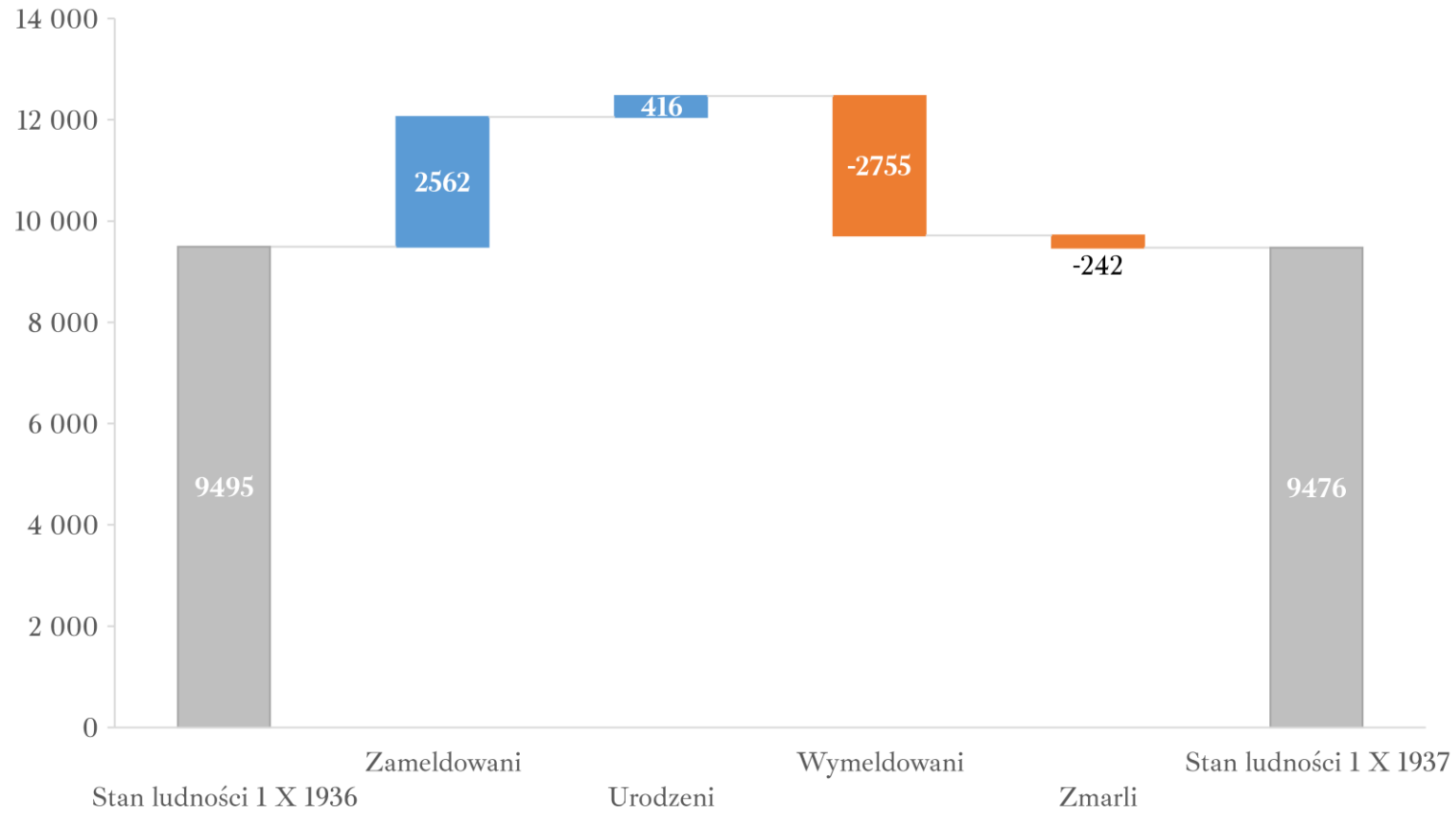
Źródło: opracowanie własne na podstawie Archiwum Państwowe w Poznaniu Oddział w Pile, Akta miasta Trzcianka, sygn. 55/15/0/9/300.

**Wykres 5. Ludność miasta Trzcianka w 1910 i 1936 roku według źródeł utrzymania**



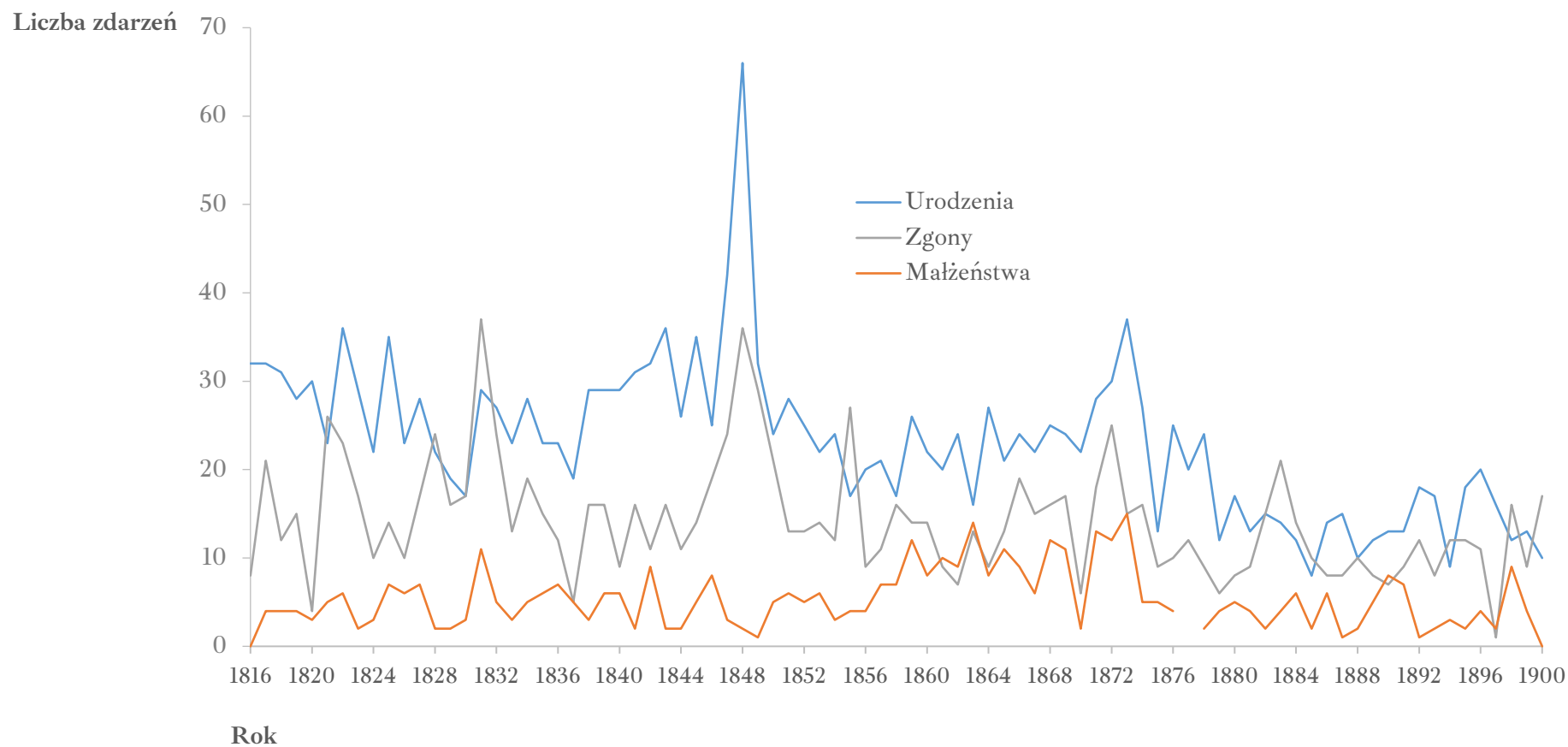
Źródło: opracowanie własne na podstawie Archiwum Państwowe w Poznaniu Oddział w Pile, Akta miasta Trzcianka, sygn. 55/15/0/9/300.

Wykres 6. Elementy bilansu ludnościowego miasta Trzcianka w okresie od 1 X 1936 do 1 X 1937 roku



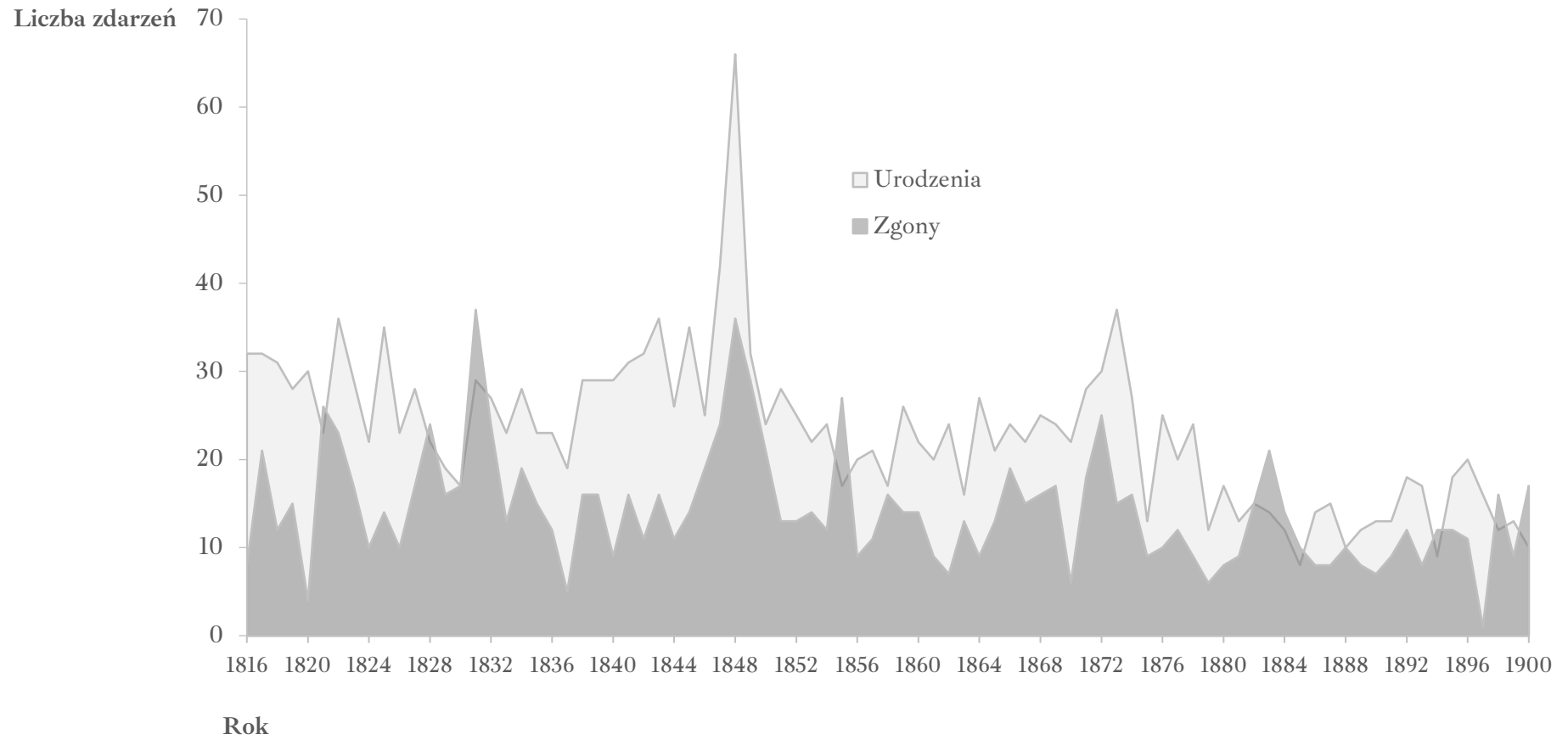
Źródło: opracowanie własne na podstawie Archiwum Państwowe w Poznaniu Oddział w Pile, Akta miasta Trzcianka, sygn. 55/15/0/9/300.

Wykres 7. Ruch naturalny w żydowskiej gminie wyznaniowej w Trzciance w latach 1816-1900



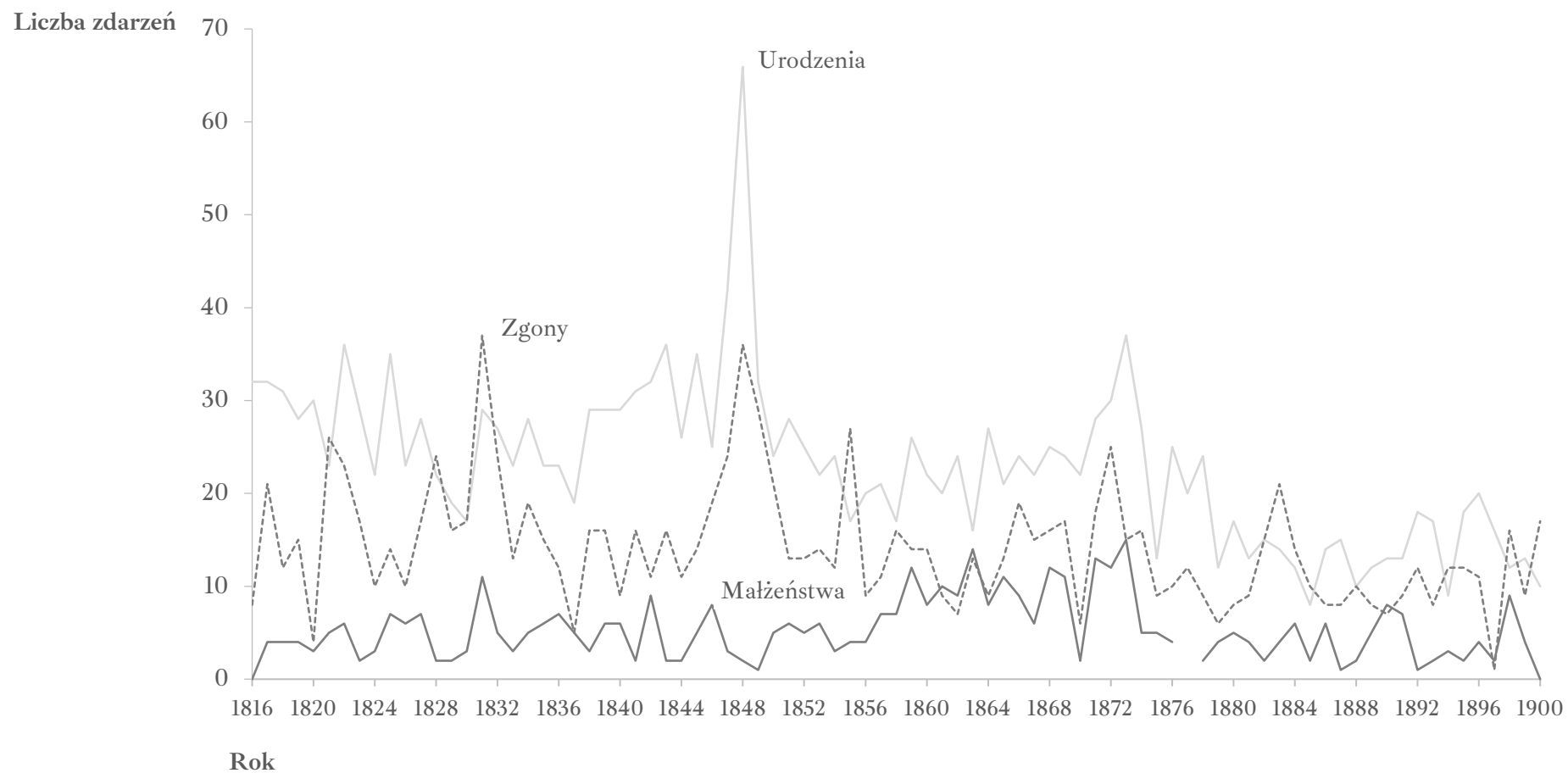
Źródło: opracowanie własne na podstawie Moses Löb Bamberger, *Geschichte der Juden in Schönlanke* (Berlin: Verlag von Louis Lamm, 1912), 34.

Wykres 8. Ruch naturalny w żydowskiej gminie wyznaniowej w Trzciance w latach 1816-1900



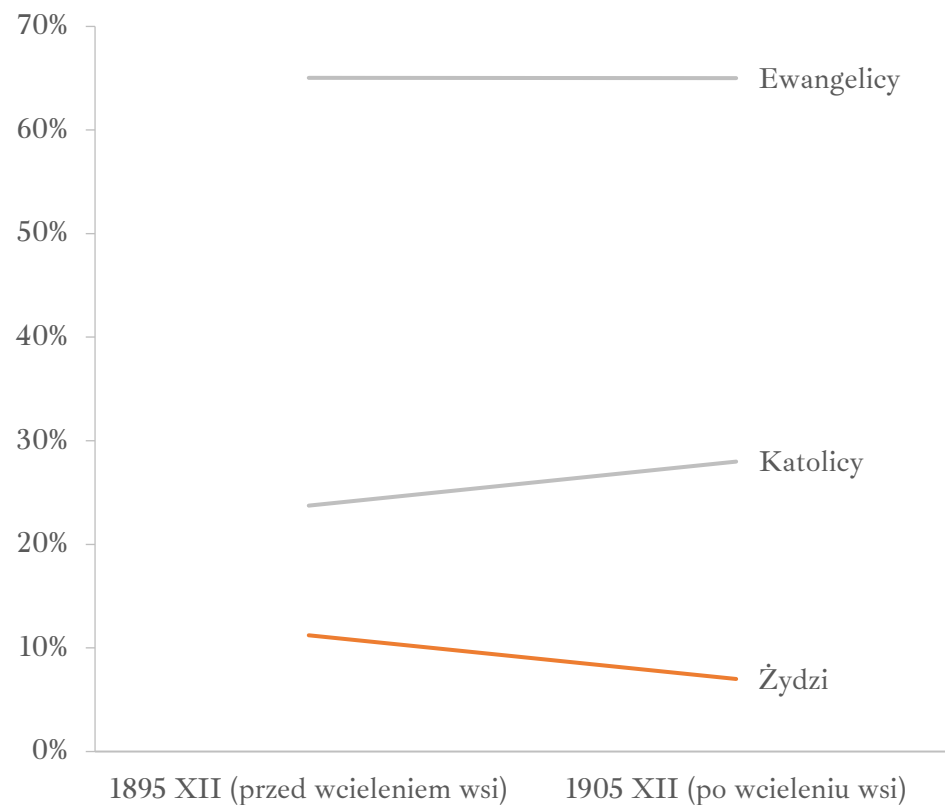
Źródło: opracowanie własne na podstawie Moses Löb Bamberger, *Geschichte der Juden in Schönlanke* (Berlin: Verlag von Louis Lamm, 1912), 34.

Wykres 9. Ruch naturalny w żydowskiej gminie wyznaniowej w Trzciance w latach 1816-1900



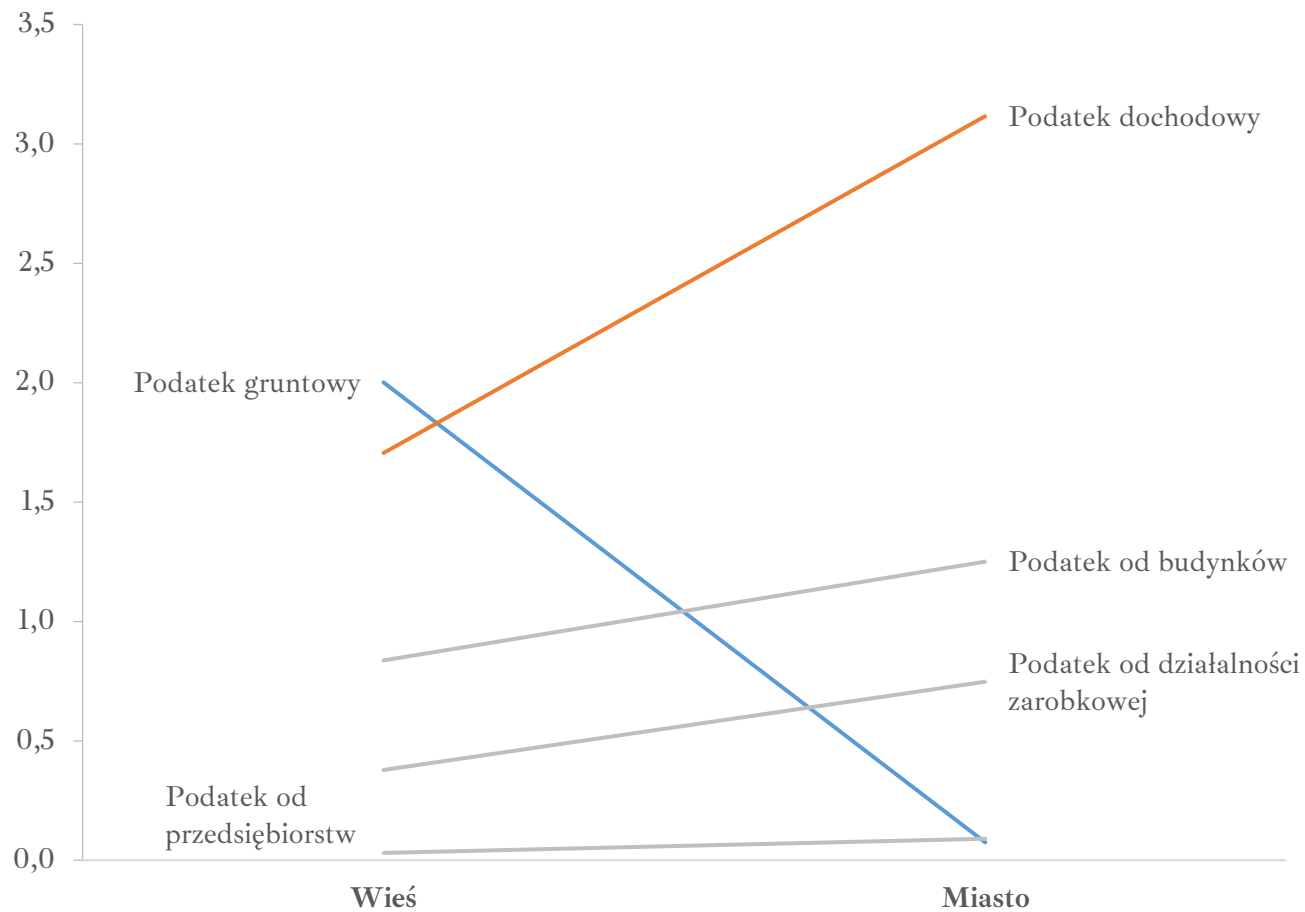
Źródło: opracowanie własne na podstawie Moses Löb Bamberger, *Geschichte der Juden in Schönlanke* (Berlin: Verlag von Louis Lamm, 1912), 34.

Wykres 10. Ludność miasta Trzcianka w 1895 i 1905 r.  
w podziale na grupy wyznaniowe



Źródło: opracowanie własne na podstawie *Gemeindelexikon für das Königreich Preussen. Auf Grund der Materialien der Volkszählung vom 2. Dezember 1895 und anderer amtlicher Quellen*, Band 5: *Provinz Posen* (Berlin: Verlag des Königlichen statistischen Bureau, 1898), 172–173; *Gemeindelexikon für das Königreich Preussen. Auf Grund der Materialien der Volkszählung vom 1. Dezember 1905 und anderer amtlicher Quellen*, Heft V: *Provinz Posen* (Berlin: Verlag des Königlich Preussischen Landesamte, 1908), 21–23.

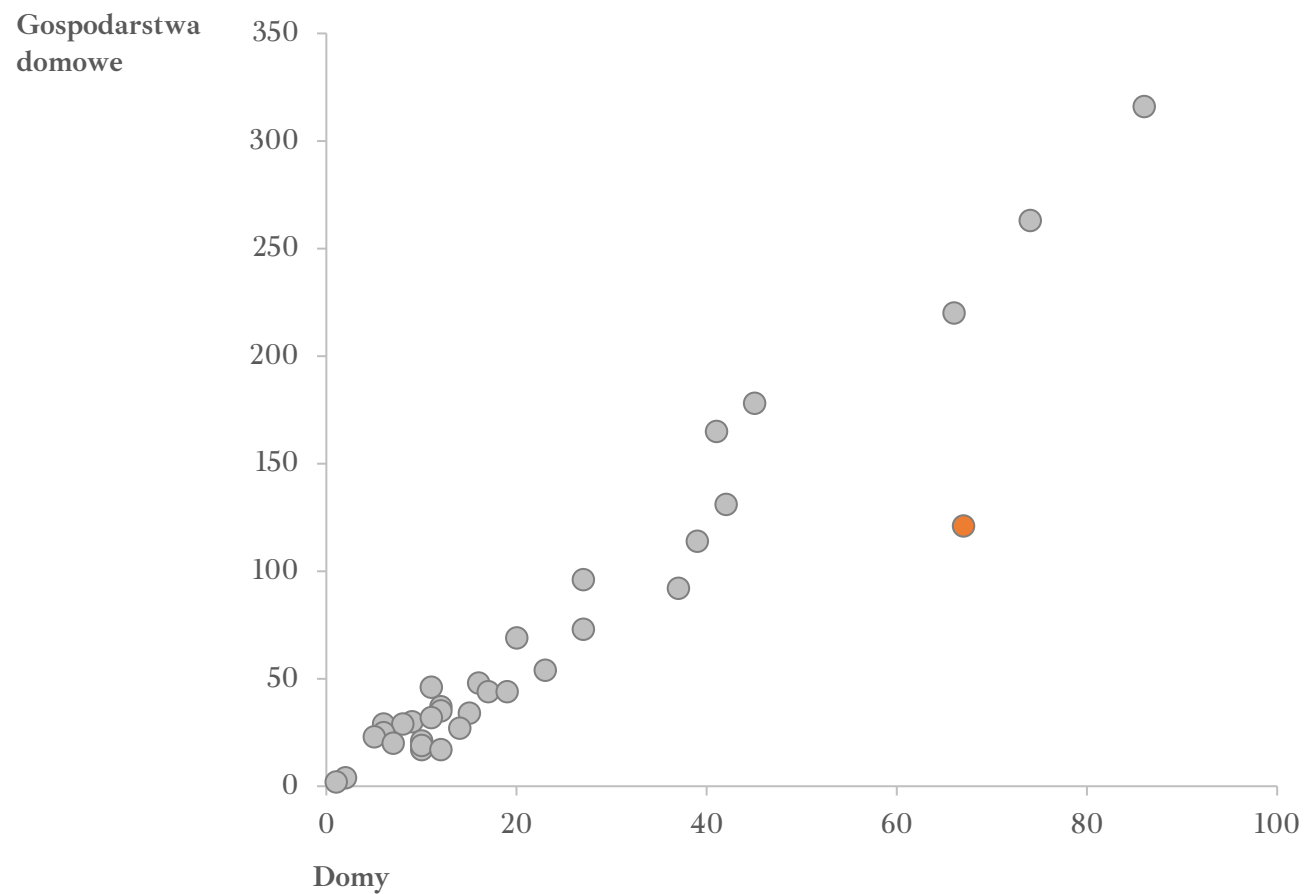
Wykres 11. Wysokość podatku w przeliczeniu na jednego mieszkańca w markach w 1905 roku w podziale na miasto i wieś Trzcianka



Źródło: opracowanie własne na podstawie Geheimes Staatsarchiv Preussischer Kulturbesitz, I HA. Rep. 77 Ministerium des Innern, Tit. 2686 Nr. 5.

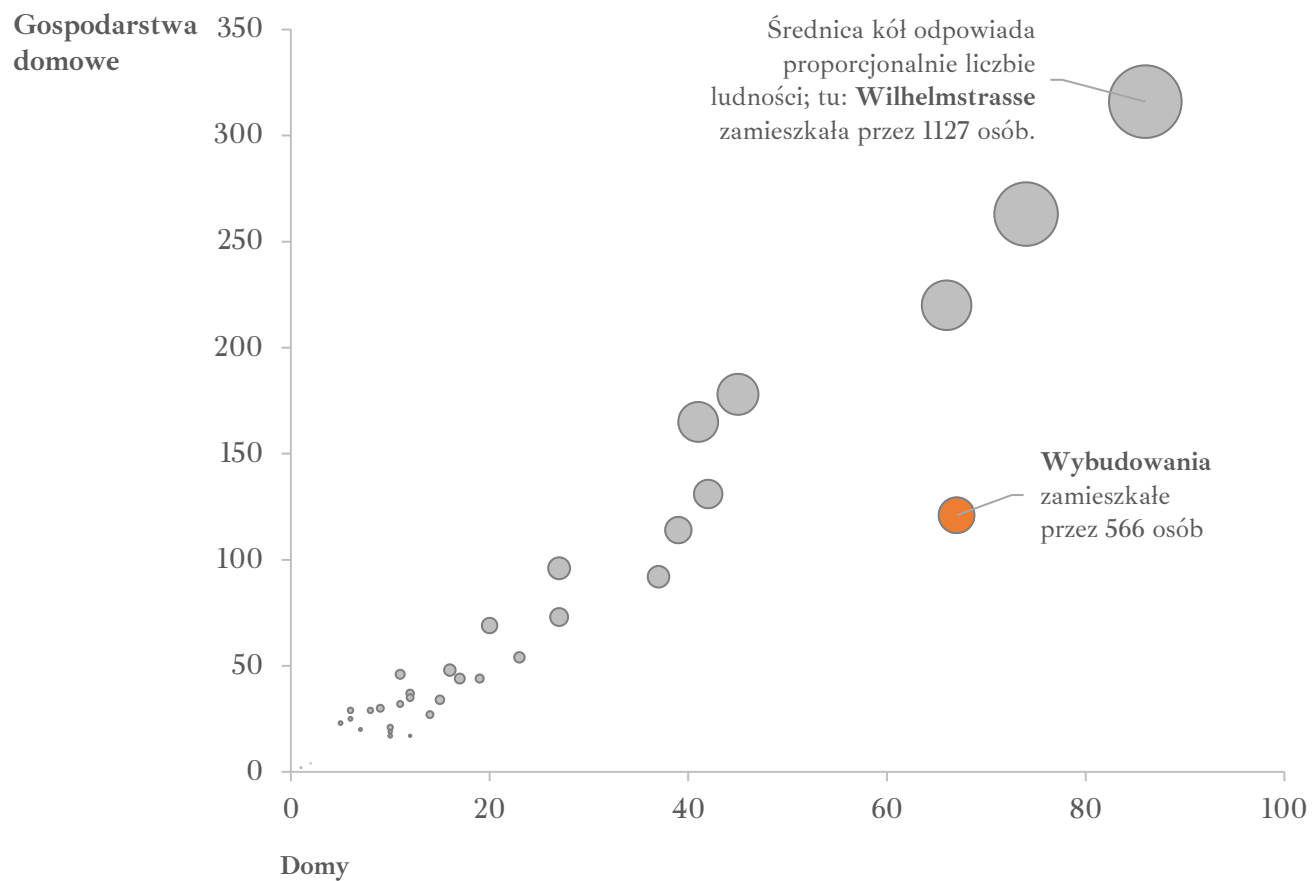


Wykres 12. Liczba domów i gospodarstw domowych na poszczególnych ulicach miasta Trzcianki w 1927 roku



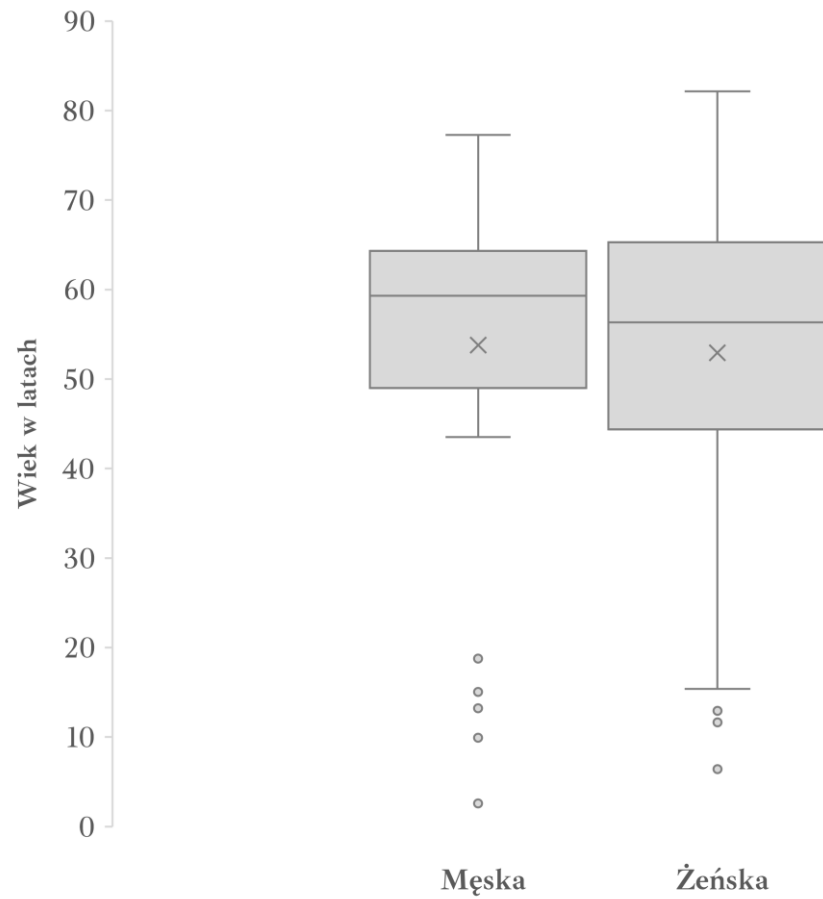
Źródło: opracowanie własne na podstawie Archiwum Państwowe w Poznaniu Oddział w Pile, Akta miasta Trzcianka, sygn. 55/15/0/3/518.

Wykres 13. Liczba domów i gospodarstw domowych ważona liczbą ludności na poszczególnych ulicach miasta Trzcianki w 1927 roku



Źródło: opracowanie własne na podstawie Archiwum Państwowe w Poznaniu Oddział w Pile, Akta miasta Trzcianka, sygn. 55/15/0/3/518.

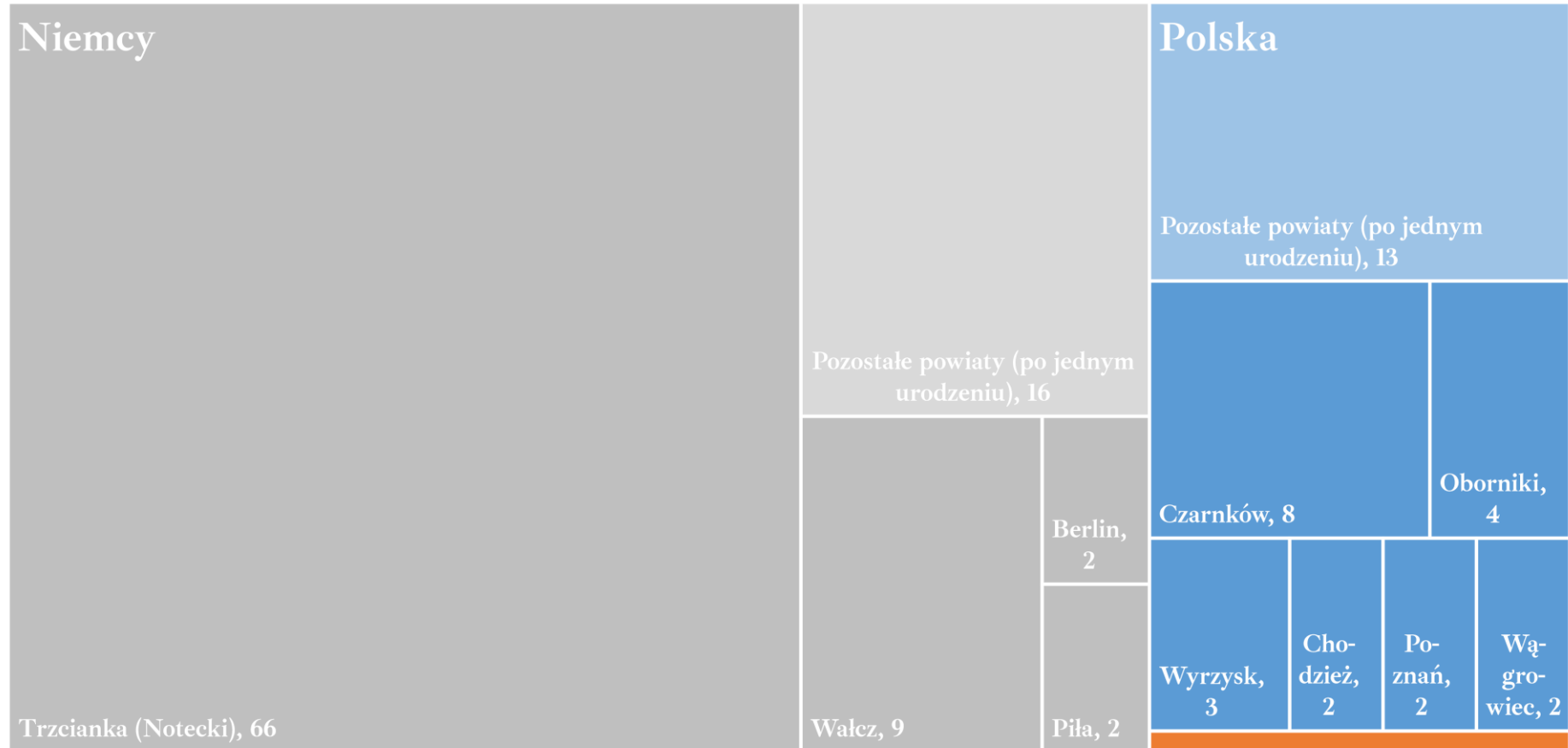
Wykres 14. Wiek ludności uznanej za żydowską w Trzciance  
17 maja 1939 roku w podziale na płeć



Źródło: opracowanie własne na podstawie *Mapping the Lives. A Central Memorial for the Persecuted in Europe 1933–1945*, dostęp 12.01.2023, <https://www.mappingthelives.org>.

Wykres 15. Liczba osób pochodzenia żydowskiego w Trzciance 17 maja 1939 roku według miejsca urodzenia (państwa, powiaty; granice z 1937 roku)

■ Niemcy ■ Polska ■ Węgry



Źródło: Opracowanie własne na podstawie *Mapping the Lives. A Central Memorial for the Persecuted in Europe 1933–1945*, dostęp 12.01.2023, <https://www.mappingthelives.org>.