

# Prognozowanie i symulacje

## Wykład 2:

Graficzna prezentacja  
danych czasowych  
Operacje na danych  
i danych czasowych

# Tematyka wykładu

- **prezentacja danych czasowych za pomocą różnych typów wykresów;**
- **elementy automatyzacji w tworzeniu wykresów;**
- **operacje na danych i zmiana struktury szeregu czasowego.**

## Definicja szeregu czasowego

**Szeregiem czasowym** będziemy nazywać ciąg obserwacji dowolnej cechy statystycznej, dokonanych w kolejnych momentach czasowych.

Wartości szeregu będziemy oznaczać symbolem:

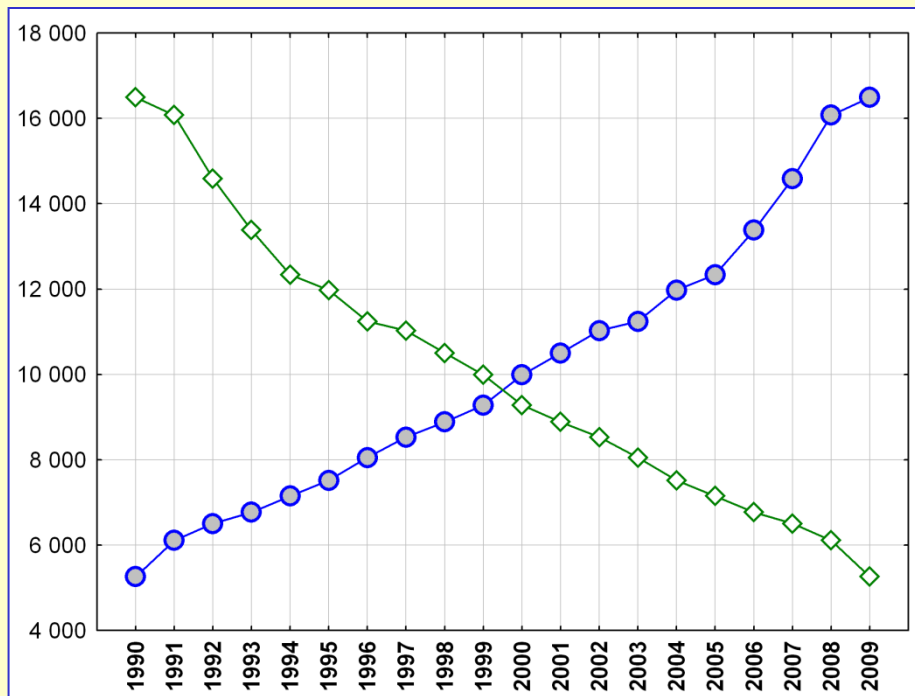
$$Y_t \quad (t = 1, \dots, T)$$

Tak zdefiniowany szereg czasowy ma charakter jednowymiarowy i obecnie zajmujemy się grupą metod statystycznych, które służą do prognozowania kolejnych jego wartości. Metody te mają swoją szczególną specyfikę, przede wszystkim ze względu na ściśle określony **cel analizy, czyli dokonanie prognozy**.

# Specyfika analizy danych czasowych

Sposób doboru metod analizy statystycznych jest, a przynajmniej winien być, bardzo mocno związany z charakterem zbioru danych oraz ich merytorycznym znaczeniem. Ten sam zbiór liczb musi być inaczej traktowany, gdy dotyczy danych czasowych, a inaczej, gdy dotyczy danych przekrojowych lub ankietowych.

Na zajęciach ze statystyki, jako najbardziej popularną metodę opisu danych liczbowych wskazano wyznaczanie statystyk opisowych, poniższy przykład pokazuje, że dla danych czasowych postępowanie takie nie zawsze ma sens.



Wykres przedstawia dynamikę dwóch zjawisk – nieco różniących się kierunkiem zmian!

Tymczasem statystyki opisowe dla obu szeregów czasowych są identyczne:

~~Średnia = 10085~~

~~Minimum = 5261~~

~~Maksimum = 16495~~

# Tworzenie prezentacji graficznej danych czasowych

W przypadku analizy danych czasowych, pierwszym i **niezbędnym** jej etapem, jest stworzenie odpowiedniego wykresu. Graficzny ogląd szeregu czasowego pozwala na wstępne zdefiniowanie czynników wpływających na zmienność badanego zjawiska w czasie, takich jak występowanie trendów, wahań sezonowych, cykli koniunkturalnych, etc.

Dane czasowe prezentuje się najczęściej w postaci wykresu liniowego (lub rzadziej słupkowego). Wykresy te są dostępne w programie *STATISTICA* w różnych wersjach, co zostało zobrazowane na kolejnych slajdach.

Omawiane narzędzia pozwalają na wizualizację przebiegu zmienności pojedynczego szeregu czasowego jak również porównywanie i analizę dynamiki dwóch lub większej liczby zmiennych.

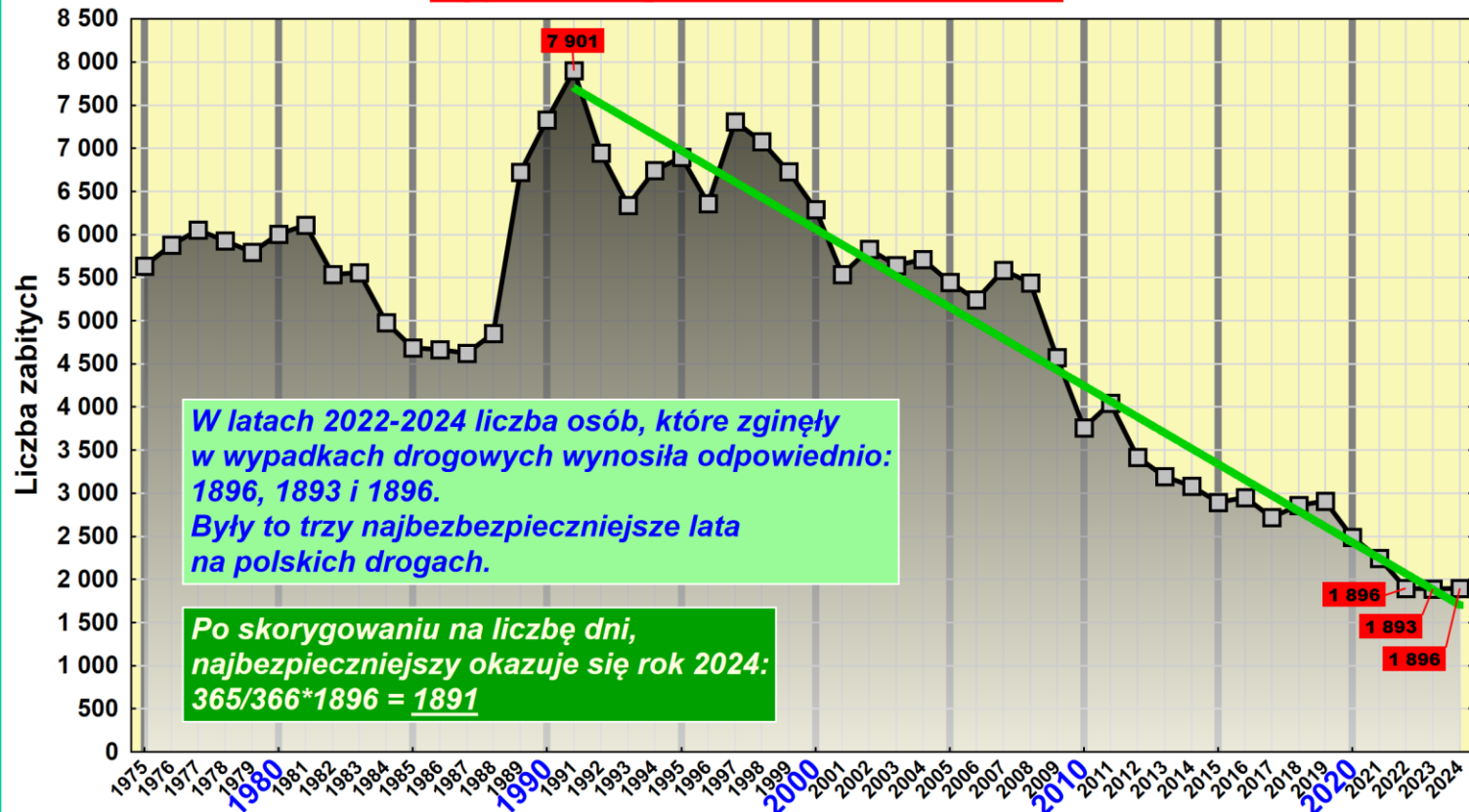
# Wykres liniowy

(Wykresy/Wykresy 2W/Wykresy liniowe (zmiennych))

Wykonujemy wykres w wersji podstawowej / Formatujemy wykres:

- na osi pionowej należy ustalić początek skali na wartości zerowej, jeśli tylko dla analizowanej wielkości zero jest absolutnym minimum;
- usuwamy robocze tytuły, a następnie osi pionowej oraz całemu wykresowi nadajemy odpowiednie tytuły;
- możemy włączyć wszystkie etykiety osi poziomej (lata) i ustalić sposób ich wyświetlania (równoległe bądź prostopadłe do osi – przy tej opcji możliwe jest też podanie dowolnego kąta).

**Wypadki drogowe w Polsce 1975-2024**

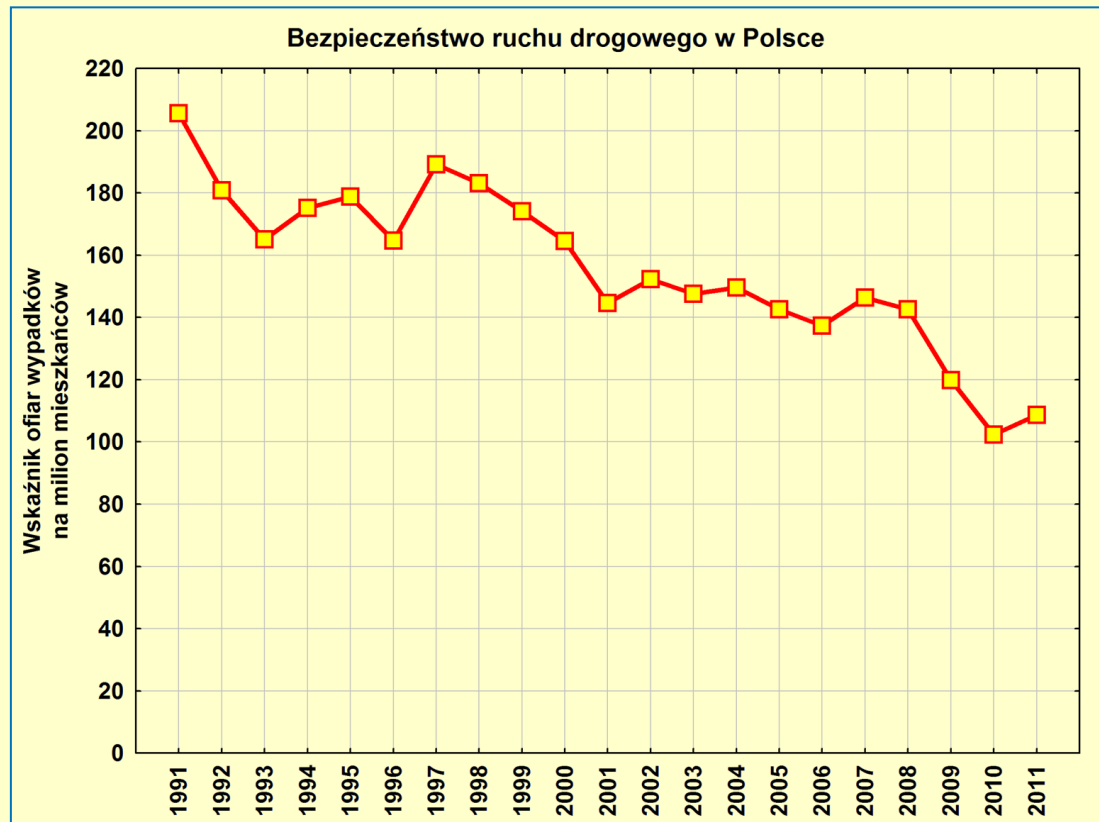
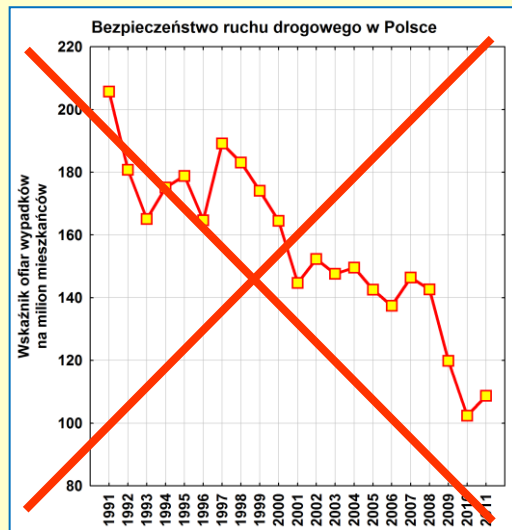
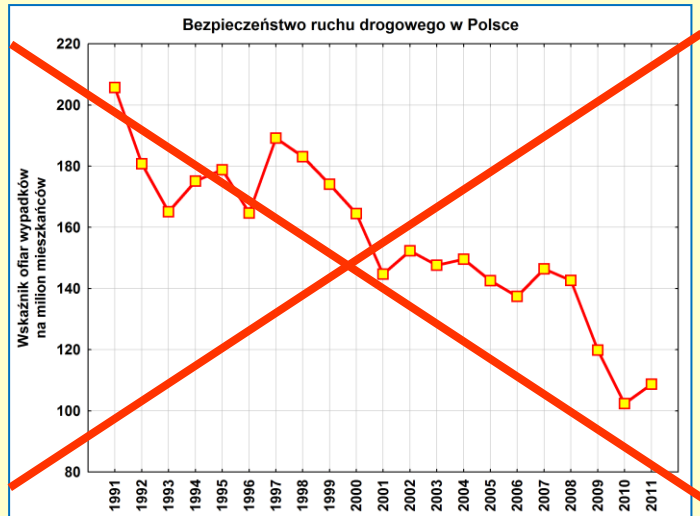


W latach 2022-2024 liczba osób, które zginęły w wypadkach drogowych wynosiła odpowiednio: 1896, 1893 i 1896. Były to trzy najbezpieczniejsze lata na polskich drogach.

Po skorygowaniu na liczbę dni, najbezpieczniejszy okazuje się rok 2024:  $365/366 \cdot 1896 = 1891$

# Jak (nie!) manipulować przekazem jaki niesie za sobą wykres?

To, jakie będą proporcje wykresu, czy skala nie będzie zbyt zawężona, może znacząco wpłynąć na przekaz „emocjonalny” jaki niesie ze sobą wykres.



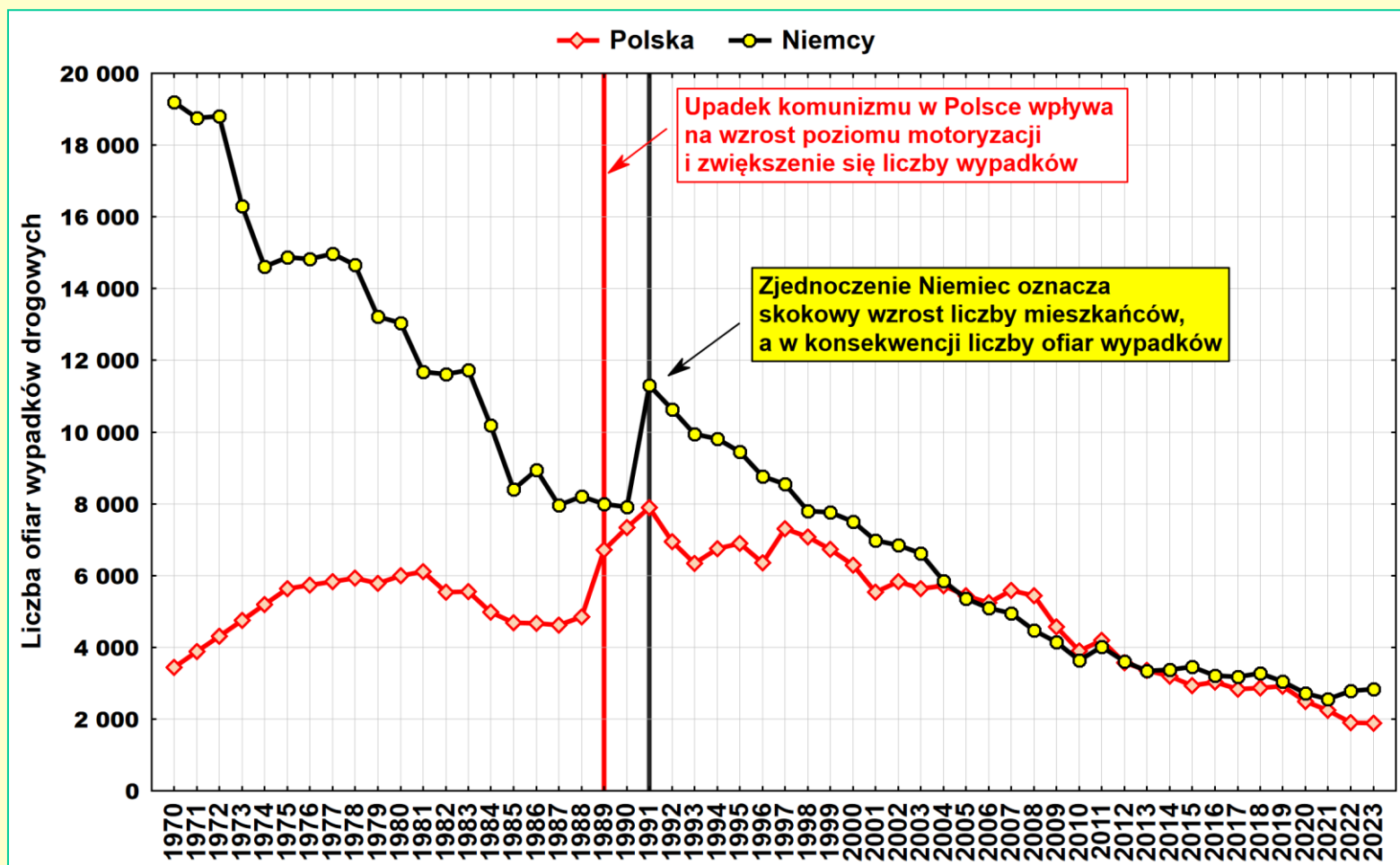
# Wykres liniowy w wersji wielokrotnej

Wykonujemy wykres *liniowy (zmiennych)* w wersji *wielokrotnej*.

Za pomocą tego wykresu można przedstawić zmienność w czasie kilku porównywalnych cech (wyrażonych w tych samych jednostkach i przyjmujących wartości z podobnego zakresu).

Legendę wykresu można dowolnie przesuwać.

W razie skasowania legendy można ją ponownie wstawić za pomocą poleceń *Wstaw / Legenda wykresu*.

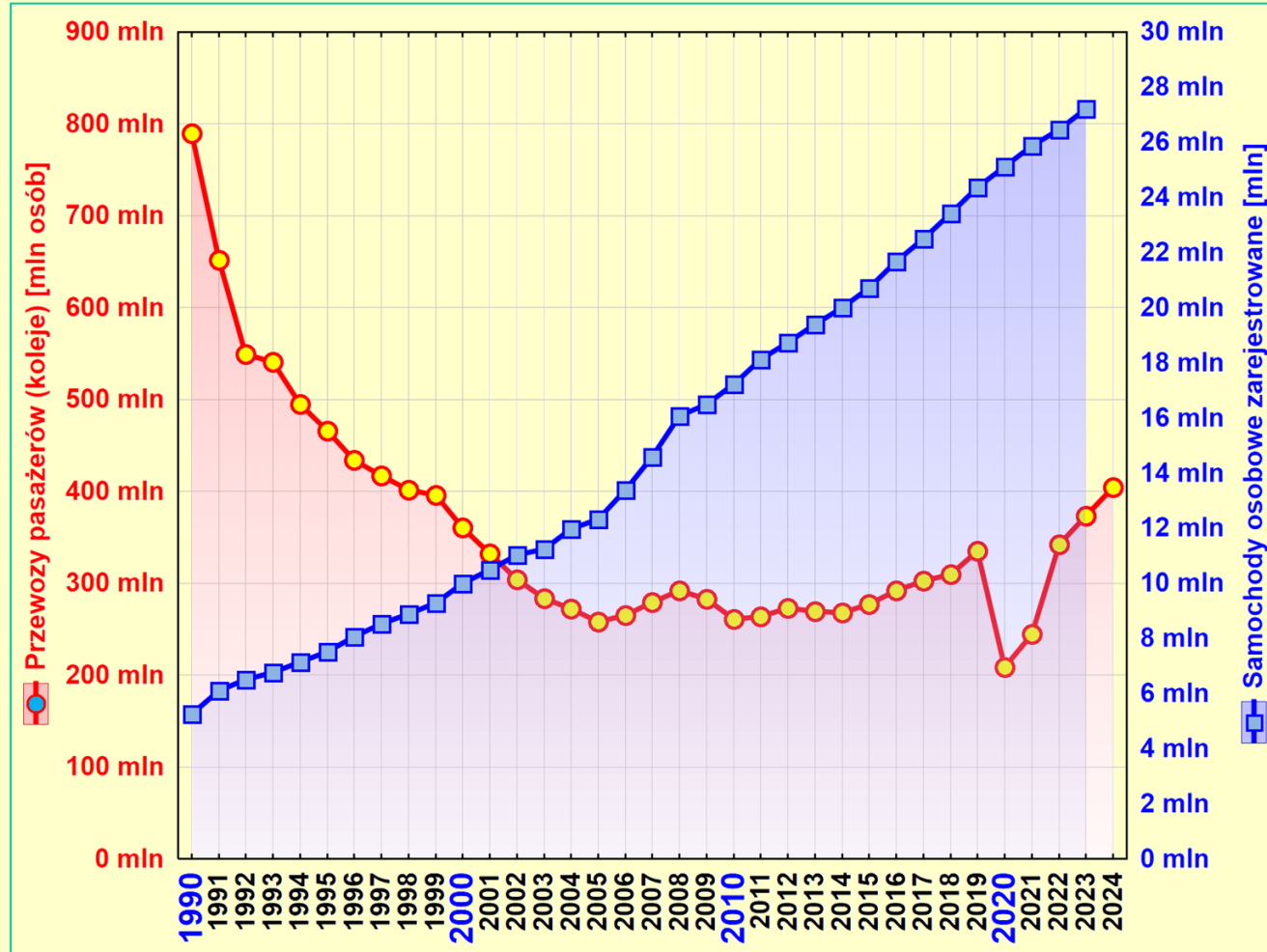


# Wykres liniowy w wersji podwójnej

Wykonujemy wykres *liniowy* (*zmiennych*) w wersji *podwójnej*.

Za pomocą tego wykresu można dokonywać porównania zmienności w czasie zjawisk, które są mierzone w zupełnie innych jednostkach (lub mających zupełnie inne zakresy wartości).

Przedstawiony obok wykres został odpowiednio sformatowany, tak by w maksymalny sposób ułatwić jego interpretację.



# Wykres liniowy w wersji potrójnej, poczwórnej...

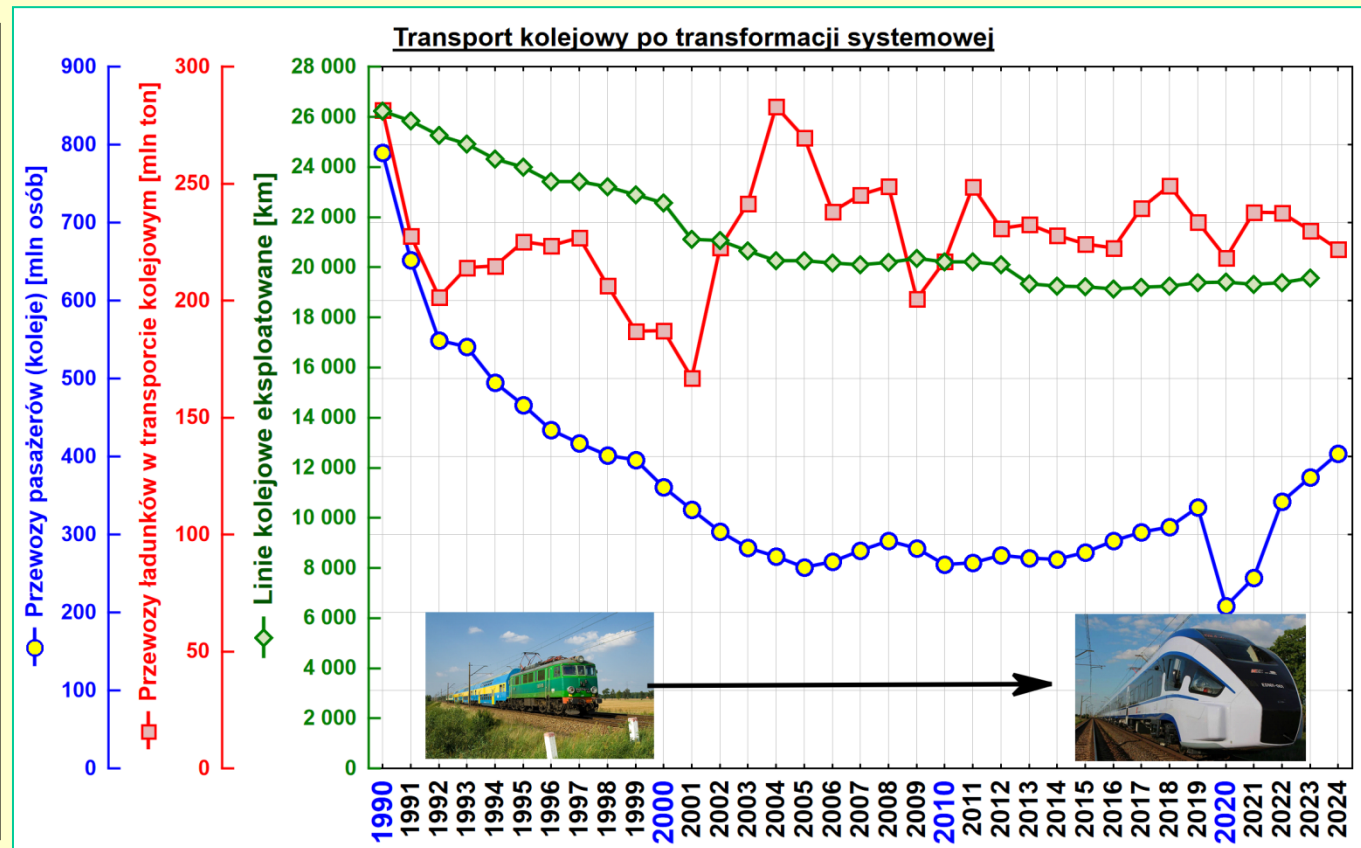
W programie *STATISTICA* istnieje możliwość wykorzystania większej liczby osi niż jest to „dopuszczalne” w standardowym wykresie podwójnym. Użytkownik może tworzyć i edytować swoje osie a następnie przypisywać do nich wykresy składowe. Poniżej zamieszczono przykład takiej „efektywnej” prezentacji.

Wykonujemy wykres *liniowy* (*zmiennych*) w wersji wielokrotnej. Następnie, w zakładce *Osie: ogólne* dodajemy dwukrotnie nową oś typu *Lewa Y*.

W zakładce *Wykres właściwy: ogólne* ustalamy przypisanie cech do osi jak na wykresie zamieszczonym obok.

Następnie nadajemy tytuły osiom (wraz ze znacznikami legendy).

Opcjonalnie możemy zmienić kolor czcionki i samej osi – tak jak na wykresie pod spodem.



# Wstawianie linii odniesienia

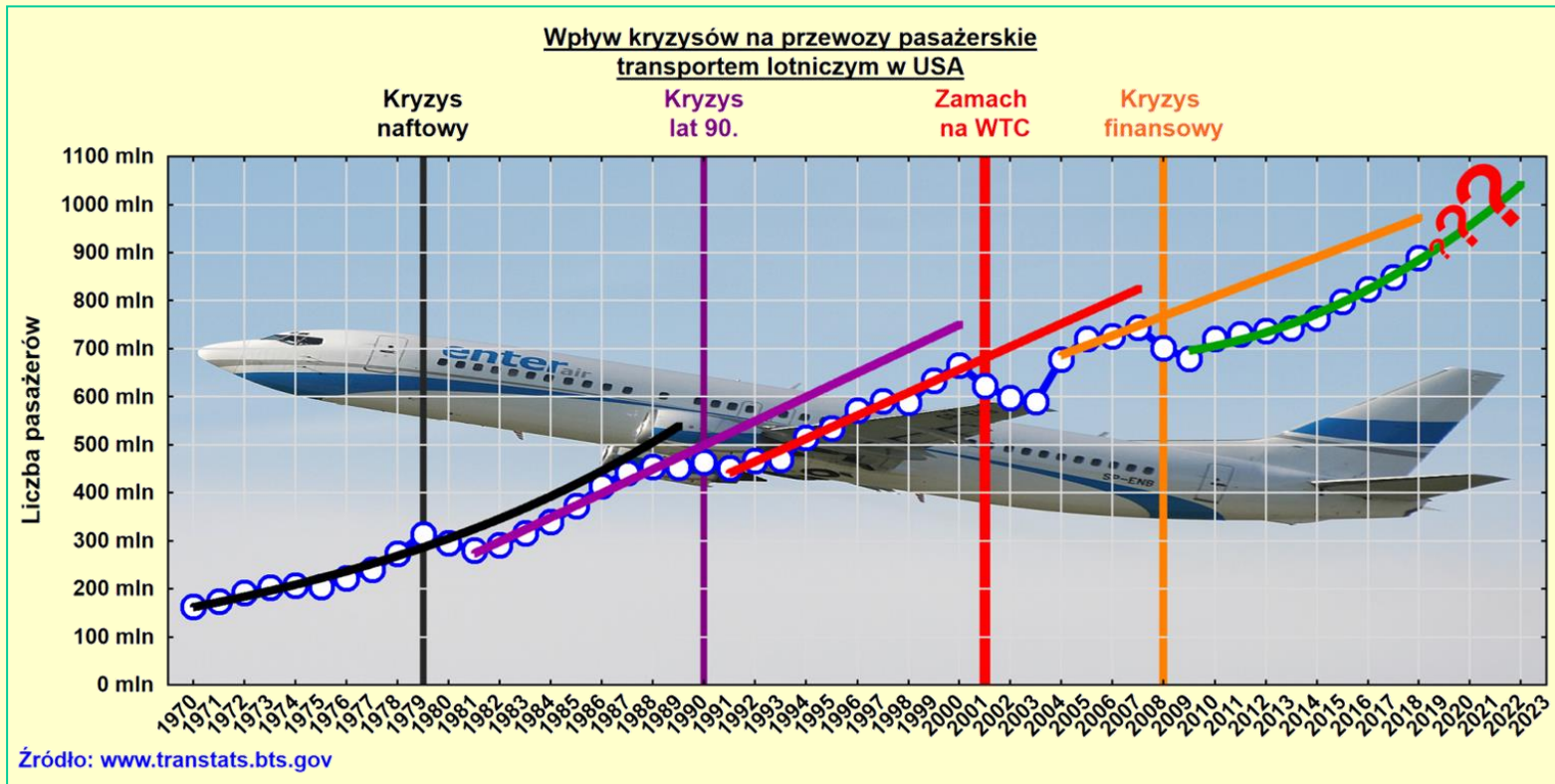
W programie *STATISTICA* istnieje możliwość wstawienia na wykresie tzw. linii odniesienia – w przypadku analizy danych czasowych można w ten sposób wyróżnić pewne okresy w przebiegu danego zjawiska lub pewne momenty zwrotne.

Aby dodać linie odniesienia znajdujemy odpowiednią zakładkę w *Opcjach wykresu* i dodajemy nową linię odniesienia (*Dodaj nową*).

Trzeba zwrócić uwagę, na której osi wstawiamy linie odniesienia!

Następnie ustalamy jej pozycję, wygląd linii oraz ewentualny opis.

**UWAGA:** Pozycje na wykresie liniowym określamy jako numer przypadku dla danej obserwacji w arkuszu danych!



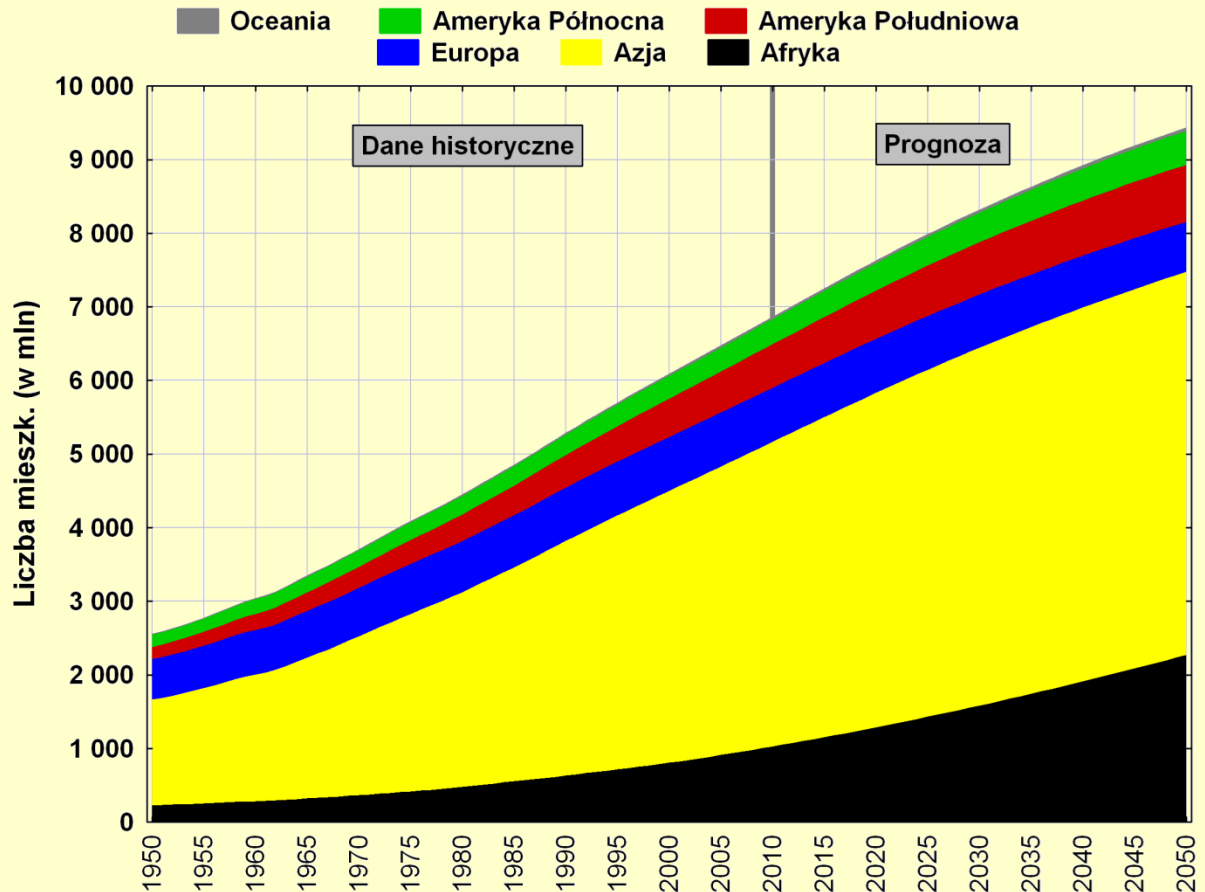
# Wykresy warstwowe

## Ludność zamieszkująca na poszczególnych kontynentach w latach 1950-2010 i 2011-2050

Wykres *sekwencyjny/nakładany* w wersji podstawowej.

Wykres ten można zastosować do prezentacji wartości wielu zmiennych, które można w logiczny sposób sumować.

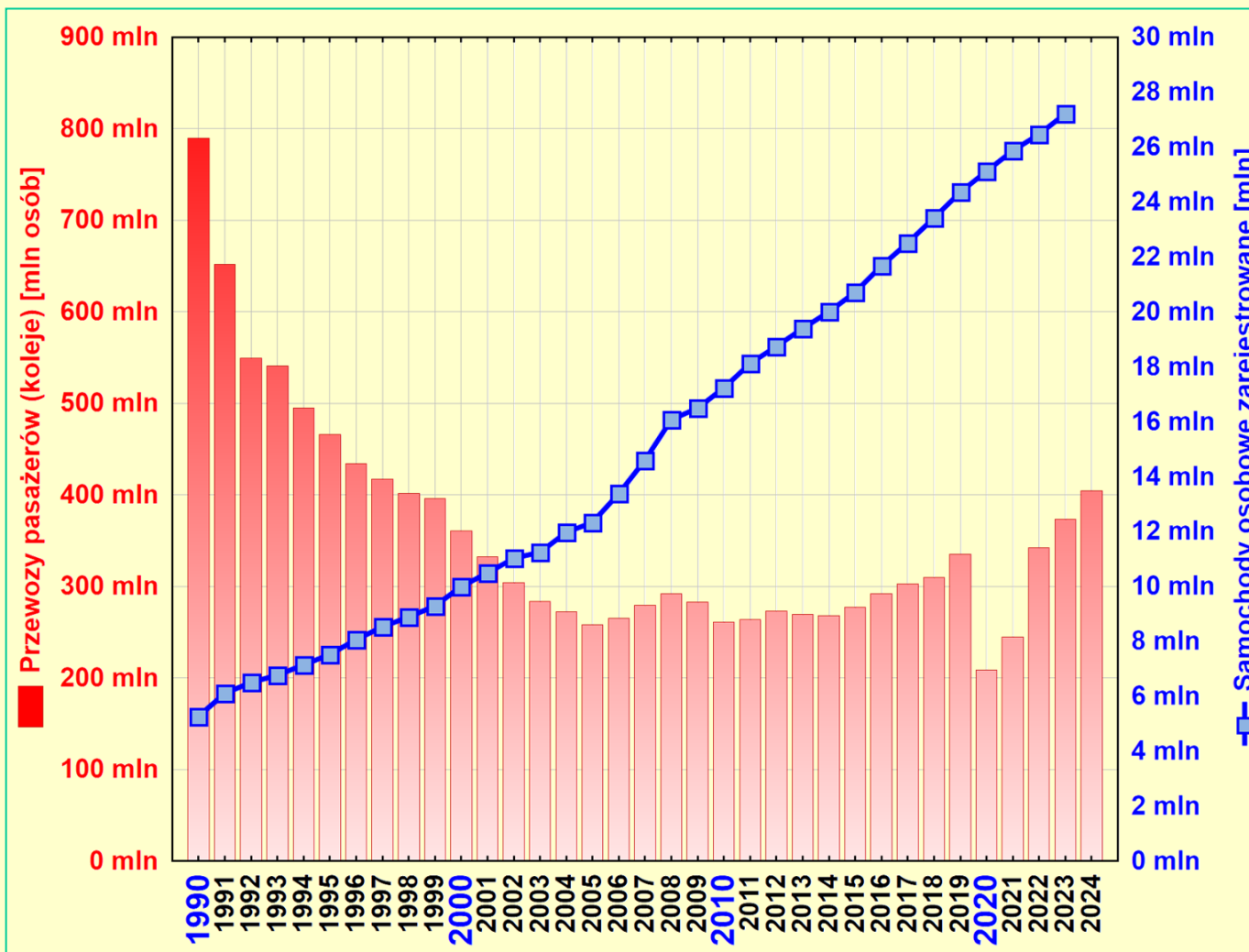
Co prawda porównanie poziomów poszczególnych zmiennych może być za jego pomocą nieco utrudnione, ale za to można ocenić trajektorię zmian ich sumy.



Źródło danych i prognoz: [www.census.gov](http://www.census.gov)

# Łączenie różnych typów wykresów

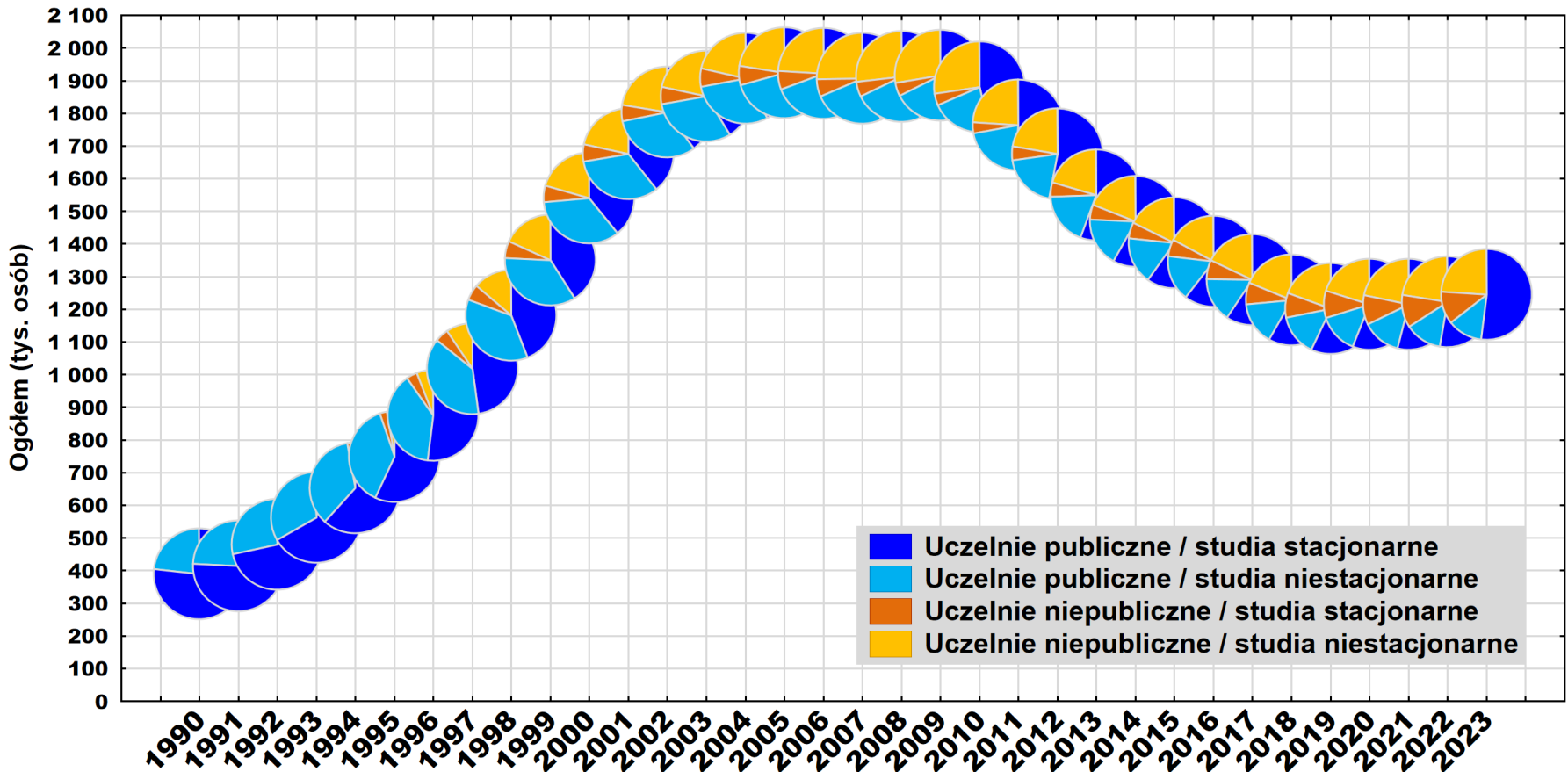
Poprzez zmianę wyświetlanych elementów można łączyć różne typy wykresu. Na przykład, z wykresu liniowego można zrobić wykresy „mieszany” – słupkowo-liniowy.



# Obrazkowy wykres rozrzutu

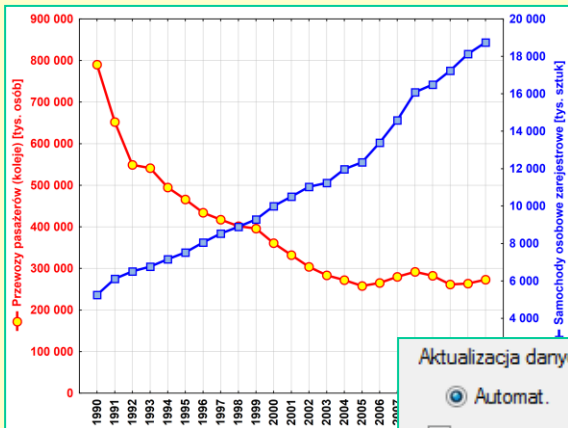
Obrazkowy wykres rozrzutu pozwala na przedstawienie dynamiki pewnego zjawiska czasowego wraz z informacją o jego strukturze. Przedstawiony poniżej przykład wykresu nie wyczerpuje wszystkich możliwości przekazywania informacji – możliwa jest jeszcze modyfikacja wielkości wykresów kołowych.

Liczba studentów w latach 1990-2023  
z podziałem względem rodzaju uczelni i typu studiów

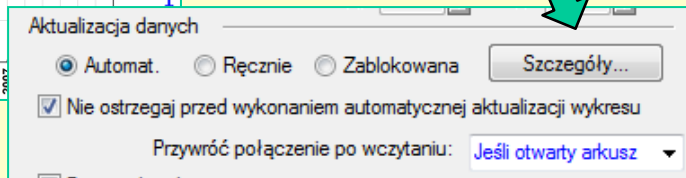


# Elementy automatyzacji w procesie tworzenia wykresów

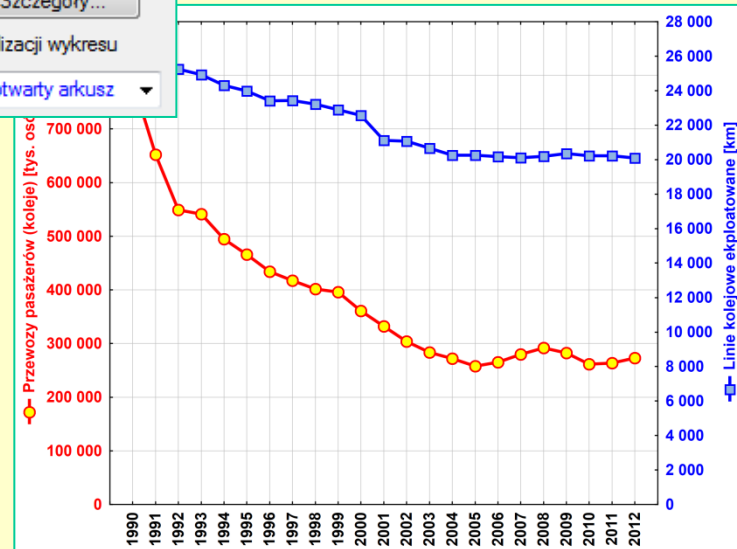
W programie *STATISTICA* można wykorzystać gotowy, sformatowany wykres jako „matrycę” do tworzenia kolejnych wykresów. W tym celu należy skopiować oryginalny wykres, a następnie wykorzystać możliwości aktualizacji danych (*Opcje wykresu / Układ*).



Za pomocą przycisku *Szczegóły* możemy dokonać zmiany cech prezentowanych na wykresie.



Warto nadmienić, iż „matrycę” wykresu można wykorzystać także do innego pliku danych, można także zmodyfikować ilość przypadków (lat) prezentowanych na wykresie.



# Przygotowanie arkusza danych czasowych (1)

## - operacje na datach

Za pomocą narzędzi *Operacje na datach* dostępnych za pomocą przycisku *Zmienne* można dokonywać skumulowania informacji o dacie zawartych w dwóch lub trzech kolumnach do jednej kolumny lub też rozdzielenia daty na dwie lub trzy kolumny.



	1 Rok	2 Miesiąc	3 Data
1	1998	1	Sty-1998
2	1998	2	Lut-1998
3	1998	3	Mar-1998
4	1998	4	Kwi-1998
5	1998	5	Maj-1998
6	1998	6	Cze-1998
7	1998	7	Lip-1998
8	1998	8	Sie-1998
9	1998	9	Wrz-1998
10	1998	10	Paź-1998
11	1998	11	Lis-1998
12	1998	12	Grz-1998



Operacje na datach

Daty liczbowo | **Daty tekstowo**

Metody operacji na datach

- Utwórz datę z 2 lub 3 zmiennych
- Podziel datę na 2 lub 3 zmienne

Zmienne źródłowe

Dzień:

Miesiąc:

Rok:

Zmienne docelowe

Nazwa:

Format:

Format zmiennej docelowej

Aby wybrać zmienną z listy, podwójnie kliknij pole lub naciśnij F2.

OK Anuluj

# **Przygotowanie arkusza danych czasowych (2)**

## **- kumulowanie danych czasowych**

Jeżeli szereg danych jest zbyt szczegółowy, można próbować agregować odpowiednio dane. W szczególności można dokonać zamiany danych dziennych na miesięczne lub roczne, danych miesięcznych na roczne, etc.

Oczywiście sposób przeprowadzania tej operacji zależy od tego czy analizujemy dane względne (wskaźniki) czy bezwzględne (wartości surowe).

Na przykład liczbę wypadków w poszczególnych miesiącach czy wartości transakcji z danego roku można zsumować, natomiast w przypadku stopy bezrobocia można je uśrednić albo wybrać jeden okres jako reprezentatywny (np. środek lub koniec roku).

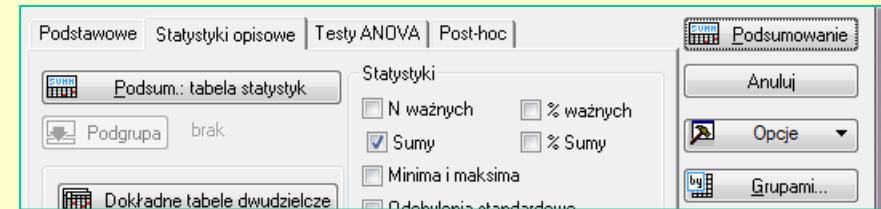
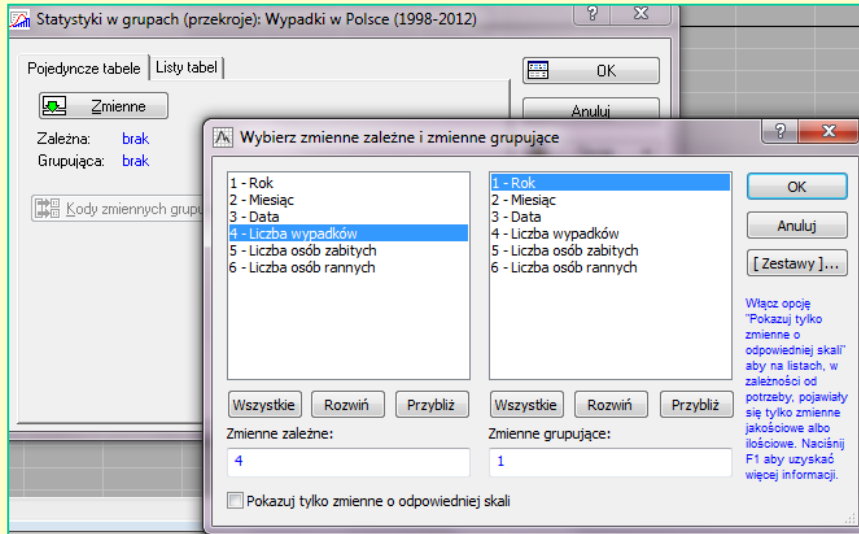
Do sumowania danych lub wyliczania średnich można wykorzystać analizę *Statystyki podstawie i tabele / Przekroje, prosta ANOVA*.

Do wycięcia danych należy zastosować warunki selekcji i polecenie *Dane / Podzbiór*.

# Przygotowanie arkusza danych czasowych (3)

## - agregacja danych czasowych

Poniżej opisano przykład zamiany danych miesięcznych o liczbie wypadków do danych rocznych.



Z arkusza wynikowego usuwamy kolumnę zawierającą wartości średnie oraz ostatni przypadek, zawierający podsumowanie wszystkich danych.

Rok	Liczba wypadków Średnie	Liczba wypadków Suma
1998	5154,583	61855,0
1999	4592,167	55106,0
2000	4777,833	57331,0
2001	4483,250	53799,0
2002	4632,500	53559,0
2003	4251,500	51078,0
2004	4145,750	51069,0
2005	4008,333	48100,0
2006	396,333	46876,0
2007	4123,000	49536,0
2008	4081,833	49054,0
2009	3683,000	44196,0
2010	3168,500	38025,0
2011	3254,833	39058,0
2012	3018,917	36227,0
2013	2616,571	18316,0
Ogół	4607,727	733185,0

Aby wynikowy arkusz był przedmiotem kolejnych analiz należy wykorzystać polecenie *Dane / Arkusz wejściowy*. Oczywiście arkusz wynikowy można zapisać w oddzielnej pliku i traktować go jako nową bazę danych.

# Przygotowanie arkusza danych czasowych (3)

## - nazwy przypadków i etykiety wykresów

Podczas tworzenia wykresów program *STATISTICA* nadaje opisy przypadków według ich nazw z arkusza danych. Jeśli nazw w arkuszu nie ma, na wykresie pojawiają się numery kolejnych przypadków. Często jednak istnieje potrzeba opisanie wykresów etykietami z pewnej zmiennej. Należy to uczynić podczas tworzenia wykresu.

**WAŻNE!!!**

W oknie wykresu należy w zakładce *Opcje* odnaleźć *Opcje etykietowania*, wybrać opcję *Ze zmiennej* i następnie wskazać zmienną, której wartości mają opisywać poszczególne przypadki.

Przykład takiego wykresu znajduje się obok.

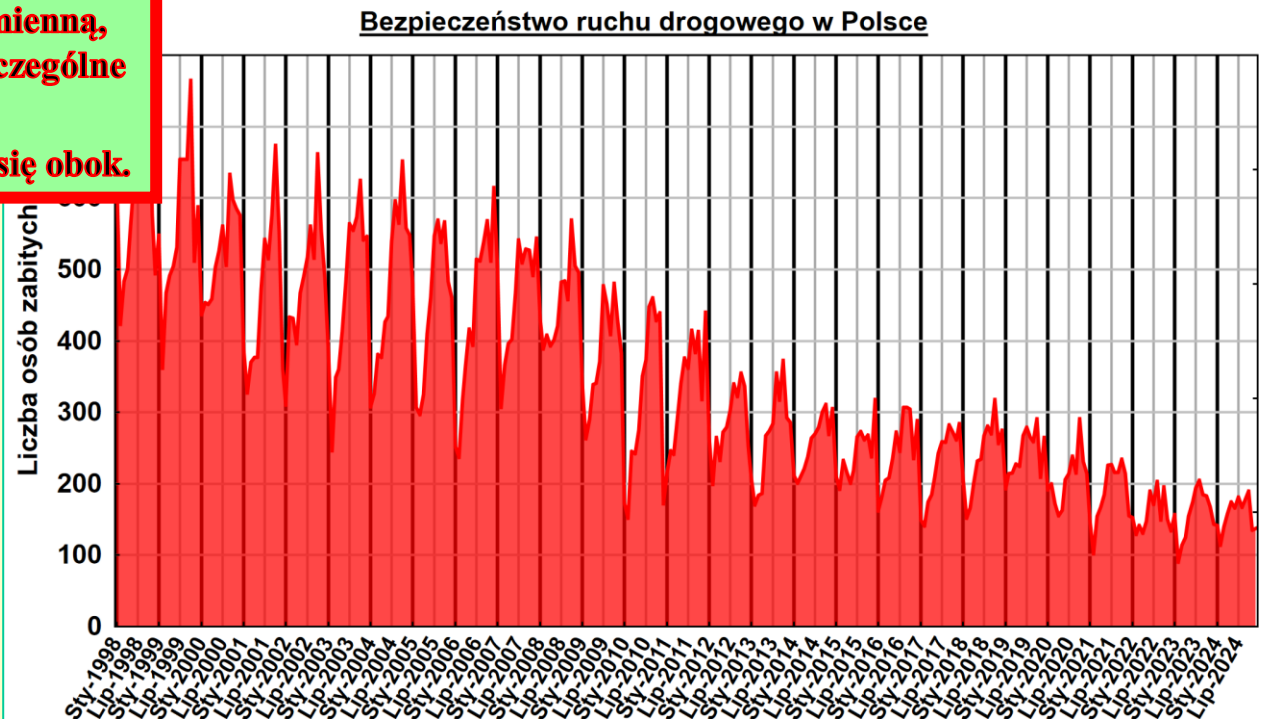
Podstawowe Więcej Wygląd Skategoryzowane Opcje 1 Opcje 2

Tytuł  
 Wyświetlaj domyślny tytuł

Tytuł użytkownika:

Stopka  
 Wyświetlaj stopkę

Opcje etykietowania  
Etykiety przypadków: Ze zmiennej



Źródło: [www.policja.pl](http://www.policja.pl)