

# **EKONOMETRIA PRZESTRZENNA**

## **PROJEKT ZESPOŁOWY**

### **1. Cel projektu**

Celem projektu jest zbadanie przestrzennej zależności i determinant zdawalności matury w powiatach Polski w latach 2020–2023. Projekt ma za zadanie sprawdzić, czy występują skupiska (hotspotty i coldspotty) w poziomie zdawalności matury w przestrzeni powiatów, oraz ocenić wpływ czynników społeczno-ekonomicznych (urbanizacja, liczba nauczycieli, wydatki na edukację, stopa bezrobocia, wynagrodzenia ogółem, liczba absolwentów) na zdawalność matury, uwzględniając ewentualne efekty przestrzenne przy pomocy modeli ekonometrii przestrzennej (SAR, SEM, ewentualnie Durbin).

### **2. Struktura badania**

#### **2. 1. Problem badawczy:**

Jakie są przestrzenne wzorce i determinanty zdawalności matury w powiatach Polski w latach 2020–2023 oraz w jakim stopniu czynniki społeczno-ekonomiczne (urbanizacja, zasoby kadrowe i finansowe w edukacji, warunki rynkowe pracy) tłumaczą obserwowane różnice i autokorelację przestrzenną w wynikach egzaminacyjnych?

Pytania badawcze:

1. Czy w latach 2020–2023 występuje istotna autokorelacja przestrzenna zdawalności matury w powiatach Polski (globalna i lokalna)?
2. Jakie czynniki społeczno-ekonomiczne (urbanizacja, liczba nauczycieli, wydatki na edukację, stopa bezrobocia, wynagrodzenia ogółem, liczba absolwentów) wpływają na zdawalność matury w przekroju powiatów w danym roku?
3. W których regionach występują skupiska wysokiej lub niskiej zdawalności (hotspotty/coldspotty)? Jakie są ich charakterystyki społeczno-ekonomiczne?

#### **2.2 Uzasadnienie wyboru problemu:**

Zdawalność matury jest istotnym wskaźnikiem poziomu edukacji w regionie, który wpływa na perspektywy rozwoju społeczno-gospodarczego. Zrozumienie przestrzennych wzorców wyników egzaminacyjnych może wspomóc decyzje o alokacji zasobów edukacyjnych (np.

nauczycieli, finansowania) i interwencje w regionach słabiej radzących sobie w edukacji. Wyniki analizy mogą pomóc decydentom (ministerstwo/dyrektorzy urzędów marszałkowskich/powiatowych) w efektywnym kierowaniu środków, by poprawić wyniki edukacyjne, szczególnie w powiatach o niskiej zdawalności i w ich sąsiedztwie.

Poziom przestrzenny analizy: NUTS 4 (powiaty Polski)

### 3. Przegląd literatury

W istniejącej literaturze dotyczącej kapitału ludzkiego i jakości edukacji w Polsce można zaobserwować coraz częstsze odwołania do zróżnicowania przestrzennego, choć nie wszystkie badania wykorzystują wprost narzędzia ekonometrii przestrzennej. Część prac, jak np. analiza Stalończyk (2018) dotycząca systemów wsparcia kapitału intelektualnego przez organizacje, czy raport OECD „Education at a Glance 2023”, koncentruje się głównie na aspektach instytucjonalnych i finansowych, a kwestie przestrzenne traktuje raczej ogólnikowo, jako tło analiz.

Inaczej przedstawia się sytuacja w badaniach takich jak Czyżewski i Polcyn (2016), którzy badają jakość edukacji na obszarach wiejskich Polski. W pracy tej wyraźnie uwzględniono różnice terytorialne i zastosowano podejście ilościowe, choć nie wykorzystano stricte modeli przestrzennych (jak SAR czy SEM), to jednak terytorialne ujęcie danych pozwala na identyfikację regionalnych dysproporcji. Z kolei Gałecka (2022), analizując produktywność szkolnictwa średniego, również zwraca uwagę na nierównomierność wyników edukacyjnych w zależności od lokalizacji, jednak nie stosuje formalnych narzędzi ekonometrii przestrzennej.

W literaturze poświęconej rynkowi pracy, jak prace Adamchika i Hyclaka (2016) czy analiza dotycząca opcji zewnętrznych (Earnings and Outside Options...), pojawia się problematyka regionalnych różnic płac, co wskazuje na istotność przestrzennego aspektu kapitału ludzkiego. Adamchik i Hyclak wykorzystują dane z podziałem regionalnym i analizują zróżnicowanie wynagrodzeń w Polsce, co może stanowić punkt wyjścia do dalszych analiz przestrzennych, choć sami autorzy nie posługują się bezpośrednio metodami ekonometrii przestrzennej.

Najwięcej potencjału do zastosowania metod przestrzennych mają prace takie jak „Przestrzenne zróżnicowanie kapitału ludzkiego w Polsce” (CEJSH 2001–2011), które już w tytule wskazują na przestrzenny charakter analizy. Choć publikacja ta nie zawiera formalnych

modeli przestrzennych, prezentuje rozkład wskaźników kapitału ludzkiego w układzie powiatowym i wojewódzkim, co pozwala na późniejsze wykorzystanie tych danych w zaawansowanej analizie przestrzennej.

Wreszcie, raporty lokalne, takie jak „Diagnoza społeczno-gospodarcza powiatu lidzbarskiego”, oraz dane CKE i OKE dotyczące zdawalności matur, dostarczają informacji w układzie przestrzennym (np. powiatowym), co umożliwia analizę zróżnicowania wyników edukacyjnych w ujęciu terytorialnym – jednakże w większości przypadków brakuje tam zastosowania narzędzi ekonometrii przestrzennej, mimo że dane są do tego dobrze przystosowane.

## 4. Opis danych

W niniejszym badaniu wykorzystano dane przekrojowe obejmujące wszystkie powiaty w Polsce (łącznie 380 jednostek administracyjnych, stan na rok 2022). Dane te mają charakter przestrzenny i zostały zebrane w formie uporządkowanego zbioru obserwacji z przypisanymi wartościami zmiennych ekonomiczno-społecznych oraz wskaźników edukacyjnych. Jednostką analizy w projekcie jest powiat – czyli drugi szczebel podziału administracyjnego Polski (NUTS-4 według nomenklatury unijnej).

### 4.1 Zmienna zależna

Zmienną objaśnianą (zależną) w modelu jest odsetek uczniów, którzy zdali egzamin maturalny w danym powiecie. Jest to wskaźnik wyrażony w procentach, który stanowi syntetyczną miarę efektywności systemu edukacji ponadgimnazjalnej (obecnie ponadpodstawowej) na poziomie lokalnym. Dane na temat zdawalności matury pochodzą z zasobów Centralnej Komisji Egzaminacyjnej oraz Głównego Urzędu Statystycznego.

### 4.2 Zmienne objaśniające

W modelu wykorzystano szereg zmiennych objaśniających, które potencjalnie mogą wpływać na wyniki egzaminów maturalnych. Są to między innymi:

Poziom urbanizacji (%) – udział ludności miejskiej w powiecie;

Liczba nauczycieli w powiecie - może pośrednio wpływać na wyniki w nauce.

Wydatki na edukację – jako pośrednia miara zaangażowania samorządów w finansowanie edukacji;

Stopa bezrobocia (%) – wskaźnik mogący pośrednio oddziaływać na motywację edukacyjną;

Średnie wynagrodzenie brutto - wskaźnik mogący mieć wpływ na lepszy dostęp do edukacji;

Ilość absolwentów uczelni wyższych ogółem – jako wskaźnik mogący mieć wpływ na motywację do nauki

Dane do zmiennych objaśniających pochodzą głównie z Banku Danych Lokalnych GUS oraz raportów Ministerstwa Edukacji i Nauki.

#### 4.3 Dane przestrzenne

W projekcie wykorzystano również dane przestrzenne w postaci pliku typu shapefile (format \*.shp), który zawiera granice administracyjne wszystkich powiatów w Polsce. Plik ten został pozyskany z otwartych zasobów GIS (Geographic Information System). Na jego podstawie możliwe było przypisanie współrzędnych geograficznych do każdej jednostki, a także wygenerowanie map tematycznych przedstawiających zmienność wskaźników w przestrzeni.

#### 4.4 Macierz wag przestrzennych

W celu uwzględnienia przestrzennego charakteru danych zastosowano macierze wag przestrzennych oparte na sąsiedztwie pierwszego rzędu, zdefiniowanym jako wspólna granica między powiatami. Macierze te zostały stworzone osobno dla każdego roku objętego analizą (2020–2023), co odpowiada strukturze danych panelowych zawierających obserwacje dla tych lat.

Dla każdego roku zastosowano funkcję poly2nb(), która generuje listę sąsiedztwa na podstawie geometrii granic powiatów, a następnie przekształcono ją do postaci listy wag przestrzennych za pomocą funkcji nb2listw(), przy zastosowaniu stylu "W", czyli normalizacji wierszowej. Styl ten powoduje, że suma wag w każdym wierszu macierzy wynosi 1, co jest standardowym podejściem przy estymacji modeli przestrzennych.

Dodatkowo, dla powiatów pozbawionych sąsiadów (np. miasta na prawach powiatu otoczone wodą) zastosowano parametr zero.policy = TRUE, aby uniknąć błędów przy generowaniu wag.

Macierze te posłużyły następnie do testowania obecności autokorelacji przestrzennej (testy Morana), a także jako istotny komponent modeli przestrzennych estymowanych w dalszej części pracy.

## 5. Analiza danych

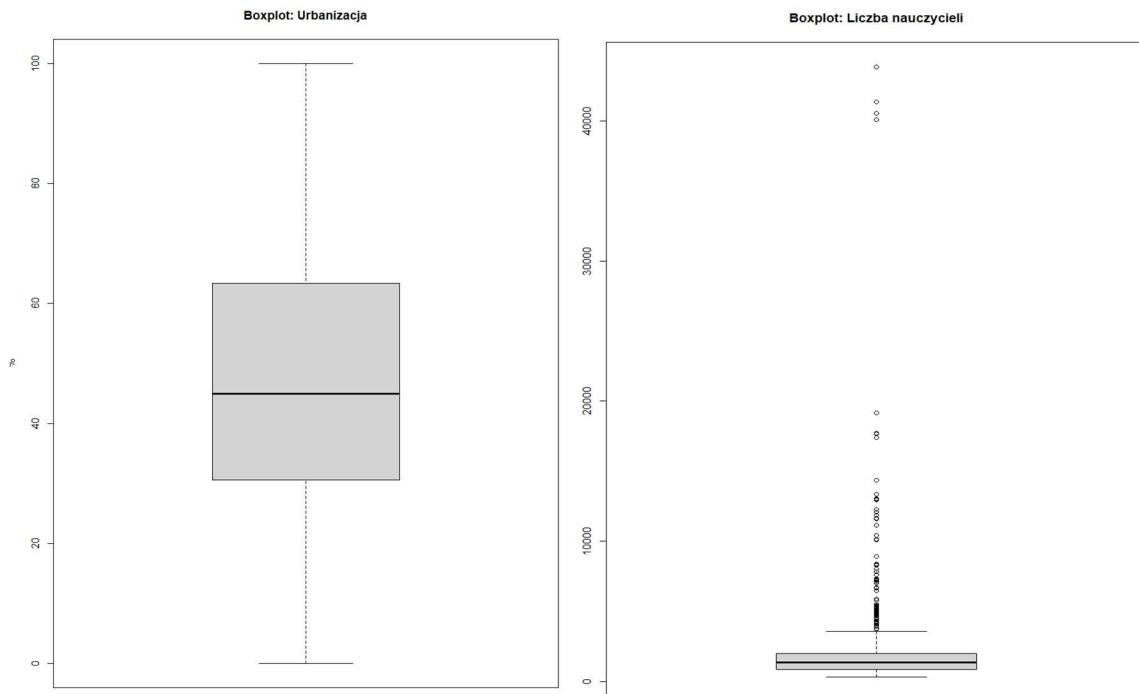
### 4.1 Wstępna analiza danych (analiza opisowa)

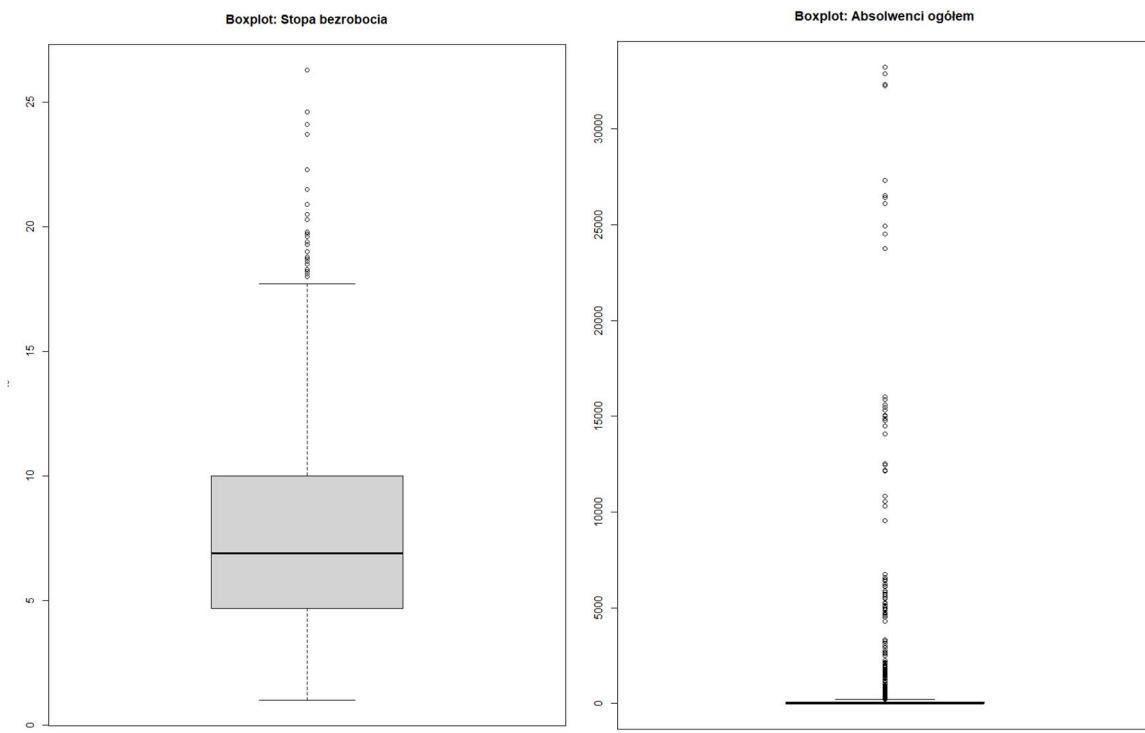
W pierwszym etapie przeprowadzono analizę braków danych. W tym celu posłużono się funkcją `colSums(is.na(dane))`, która pozwala sprawdzić liczbę wartości brakujących w każdej kolumnie zbioru danych. Wynik analizy nie wykazał żadnych brakujących obserwacji – wszystkie zmienne były完全に uzupełnione, co oznacza, że dane nie wymagały dodatkowego czyszczenia pod tym kątem.

```
Braki danych (NA):
> print(colSums(is.na(dane)))
      rok      powiat zdawalnosc_matury   urbanizacja_pct liczba_nauczycieli
      0          0            0                  0                      0
wydatki_edykacja stopa_bezrobocia wynagrodzenia_ogolem absolwenci_ogolem
      0          0            0                  0                      0
```

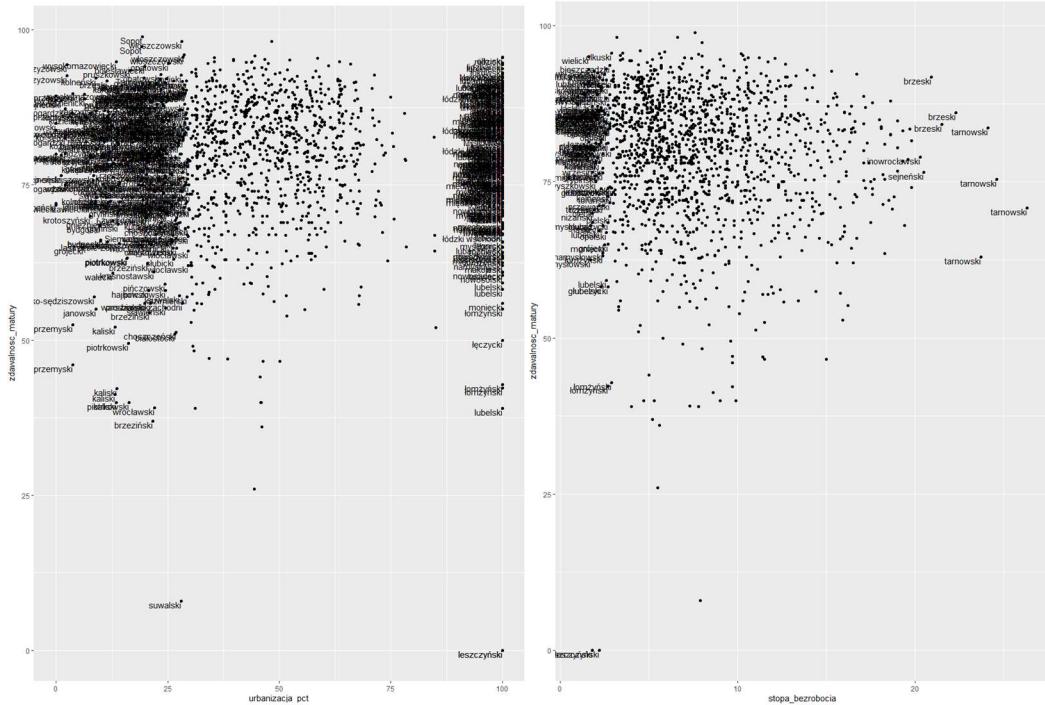
Kolejnym krokiem było wykrycie obserwacji odstających (ang. outliers), które mogłyby znacząco wpływać na wyniki modeli ekonometrycznych. Zastosowano kilka metod, aby uzyskać pełniejszy obraz ewentualnych anomalii w danych.

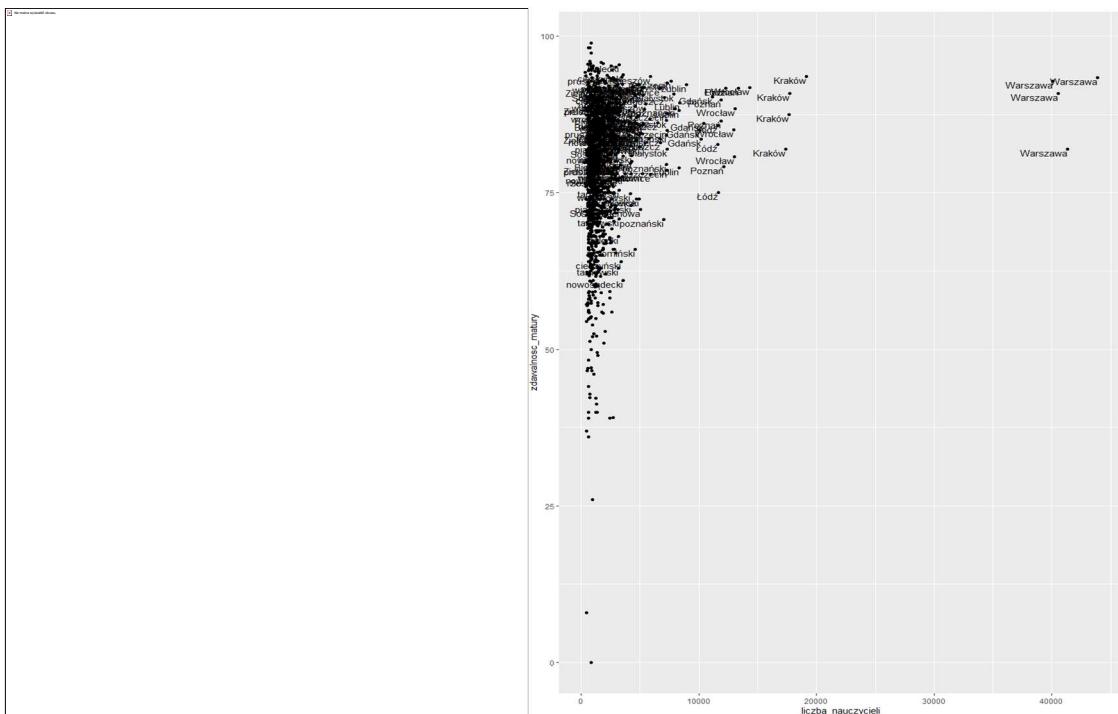
Po pierwsze, dla każdej ze zmiennych objaśniających wygenerowano wykresy pudełkowe (boxploty), które umożliwiają szybkie wizualne zidentyfikowanie obserwacji leżących poza typowym zakresem wartości. Przykładowe wykresy zostały zamieszczone poniżej:





Dodatkowo, wykorzystano wykresy rozrzutu, które pozwalają ocenić rozkład danych w przestrzeni oraz zauważać potencjalne przypadki odstające względem reszty obserwacji.





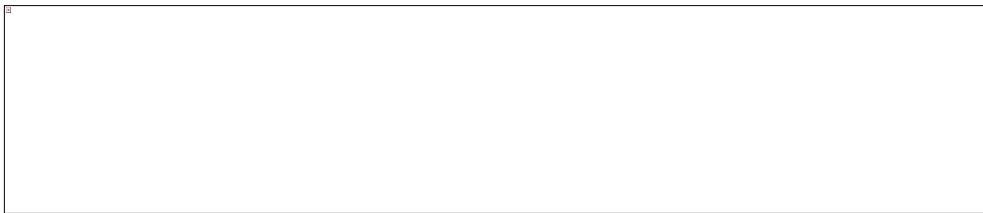
W kolejnym kroku zastosowano również podejście oparte na standaryzowanych wartościach z-score. Dla każdej zmiennej obliczono z-score, a następnie zidentyfikowano obserwacje, dla których wartość ta przekraczała ustalony próg (najczęściej  $\pm 3$ ), co może wskazywać na odstające obserwacje statystyczne.

```

> z_scores <- scale(dane$wynagrodzenia_ogolem)
> outliers <- which(abs(z_scores) > 3)
> cat("outliery - wynagrodzenia_ogolem (|z| > 3):\n")
outliery - wynagrodzenia_ogolem (|z| > 3):
> print(dane[outliers, c("powiat", "wynagrodzenia_ogolem")])
# A tibble: 14 × 2
  powiat      wynagrodzenia_ogolem
  <chr>          <dbl>
1 bełchatowski     9269.
2 Gdańsk            8916.
3 Jastrzębie-zdrój        0
4 Katowice         9106.
5 Kraków           9224.
6 Lubiński         12804.
7 Łęczyński        10420.
8 Płock              9152.
9 Warszawa          9626.
10 Jastrzębie-zdrój        0
11 Lubiński         11362.
12 Jastrzębie-zdrój     8988.
13 Lubiński         10077.
14 Lubiński          8920.

```

Uzupełnieniem analizy był test Grubbsa, który pozwala na formalną weryfikację obecności pojedynczych obserwacji odstających w rozkładzie danej zmiennej. Testy przeprowadzono osobno dla każdej zmiennej objaśniającej, a ich wyniki przedstawiono poniżej:



Dla wszystkich zmiennych uwzględnionych w analizie obliczono podstawowe miary statystyczne, takie jak: średnia, mediana, minimum, maksimum, odchylenie standardowe oraz współczynnik asymetrii.

Na podstawie uzyskanych wyników można zauważyc, że większość zmiennych cechuje się zróżnicowanym rozkładem — niektóre z nich wykazują silną asymetrię (np. wydatki\_na\_edykacjē, absolwenci\_ogółem), co sugeruje obecność obserwacji o bardzo wysokich wartościach w porównaniu do reszty. Inne zmienne, takie jak urbanizacja\_pct czy zdawalność\_matury, są bardziej symetryczne i rozproszone w sposób umiarkowany.

```
Zmienna: zdawalnosc_matury
Średnia: 79.15
Mediana: 80.95
Min: 0
Max: 98.92
Odch. stand.: 10.56
Asymetria: -1.69

Zmienna: stopa_bezrobocia
Średnia: 7.66
Mediana: 6.9
Min: 1
Max: 26.3
Odch. stand.: 4.08
Asymetria: 0.96

Zmienna: urbanizacja_pct
Średnia: 50.38
Mediana: 44.96
Min: 0
Max: 100
Odch. stand.: 27.4
Asymetria: 0.62

Zmienna: wynagrodzenia_ogolem
Średnia: 5599.77
Mediana: 5469.98
Min: 0
Max: 12804.03
Odch. stand.: 983.09
Asymetria: 0.93

Zmienna: liczba_nauczycieli
Średnia: 1883.43
Mediana: 1326.5
Min: 300
Max: 43868
Odch. stand.: 2703.22
Asymetria: 9.57

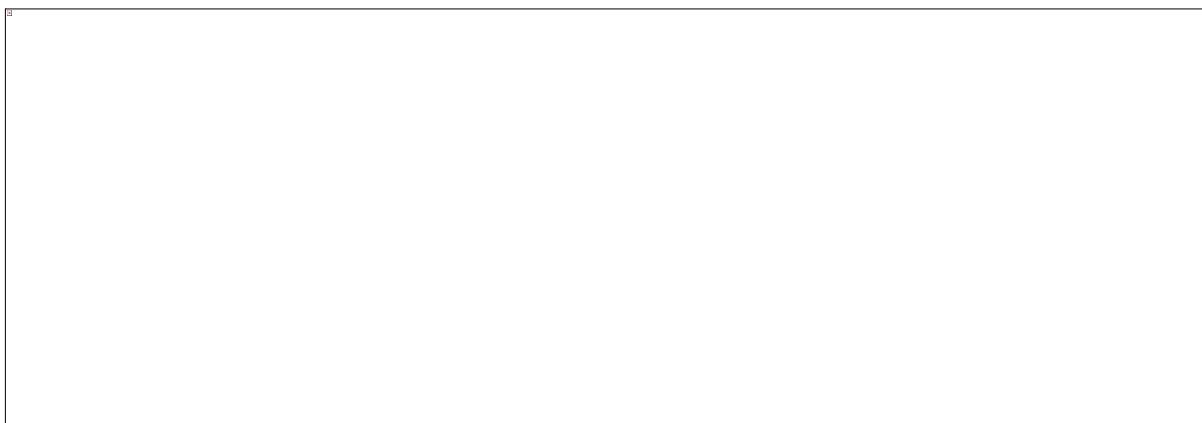
Zmienna: absolwenci_ogolem
Średnia: 645.07
Mediana: 0
Min: 0
Max: 33242
Odch. stand.: 3019.85
Asymetria: 7.31

Zmienna: wydatki_edykacja
Średnia: 118745593.4
Mediana: 36806833.08
Min: 7500
Max: 6660653006.59
Odch. stand.: 354291087.84
Asymetria: 10.46
```

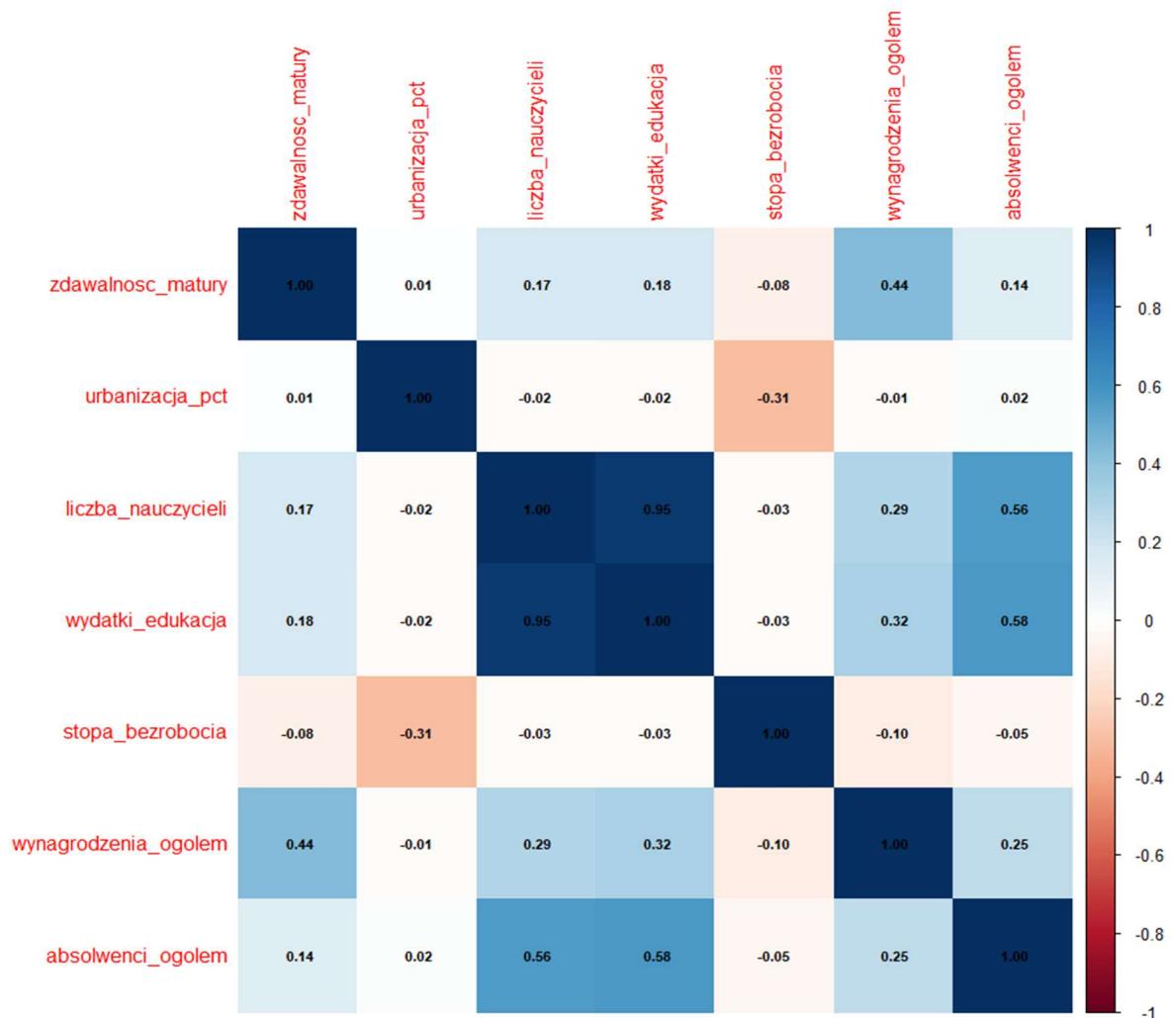
Przeprowadzono również analizę korelacji pomiędzy wybranymi zmiennymi ilościowymi, wykorzystując współczynnik korelacji Pearsona, którego wartości mieszczą się w przedziale od -1 do 1. Wyniki zaprezentowano w formie macierzy korelacji, co pozwoliło na identyfikację zależności pomiędzy badanymi zmiennymi.

Analiza wykazała silną dodatnią korelację (0,95) między liczbą nauczycieli a wydatkami na edukację, co sugeruje, że większe nakłady finansowe są powiązane z wyższą liczbą zatrudnionych nauczycieli. Zaobserwowano również znaczącą dodatnią korelację (0,58) między wydatkami na edukację a liczbą absolwentów, co może wskazywać na pozytywny wpływ poziomu finansowania na efektywność systemu edukacyjnego. Umiarkowana ujemna korelacja (-0,31) wystąpiła między stopą bezrobocia a poziomem urbanizacji, co może świadczyć o niższym poziomie bezrobocia w bardziej zurbanizowanych obszarach. Z kolei niska korelacja pomiędzy zdawalnością matury a pozostałymi zmiennymi sugeruje, że żaden pojedynczy czynnik nie odgrywa dominującej roli w kształtowaniu tego wskaźnika.

Przeprowadzona analiza korelacji pozwoliła na zidentyfikowanie istotnych powiązań pomiędzy zmiennymi i stanowi wartościowy punkt wyjścia do dalszych, bardziej zaawansowanych analiz statystycznych oraz budowy modeli predykcyjnych.



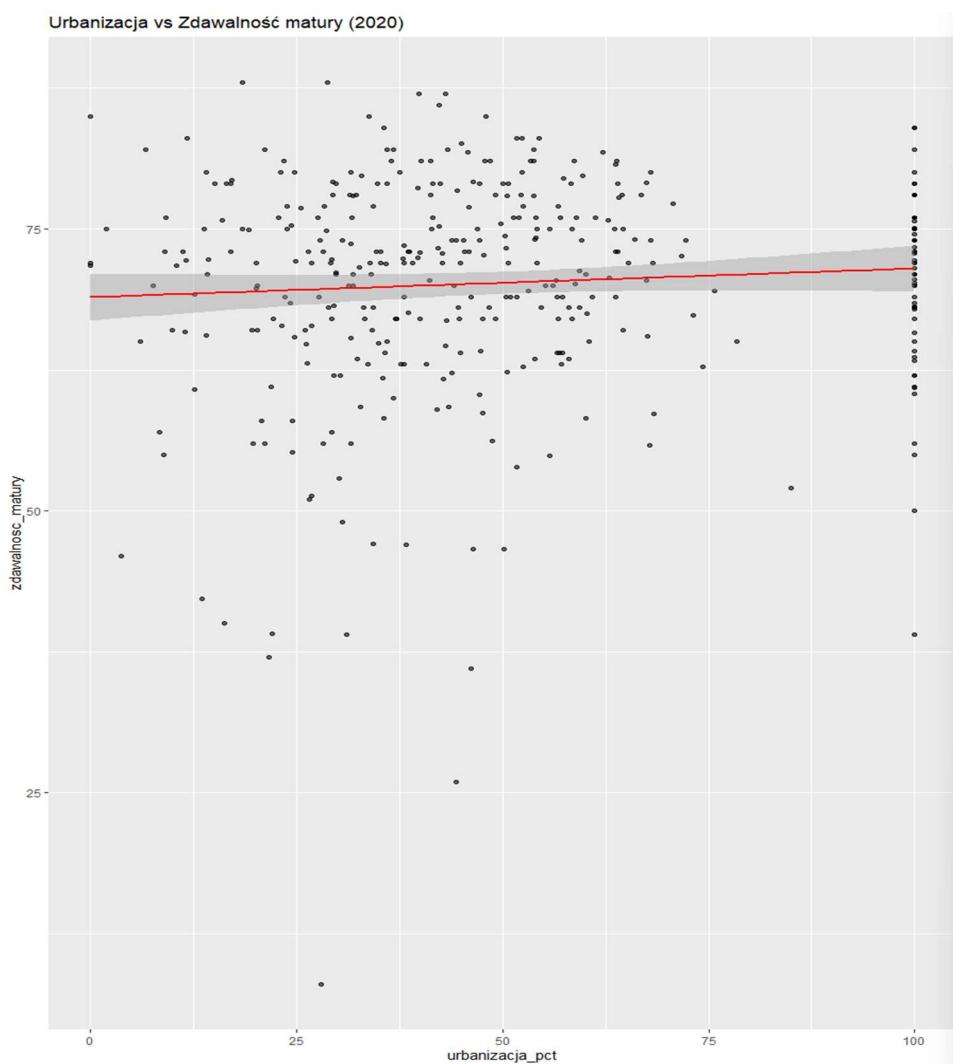
Korelacja została również przedstawiona na wykresie:



W ramach projektu wykonaliśmy również analizę wizualną danych w postaci wykresów rozrzutu i gęstości, aby lepiej zrozumieć zależności oraz rozkład badanych zmiennych.

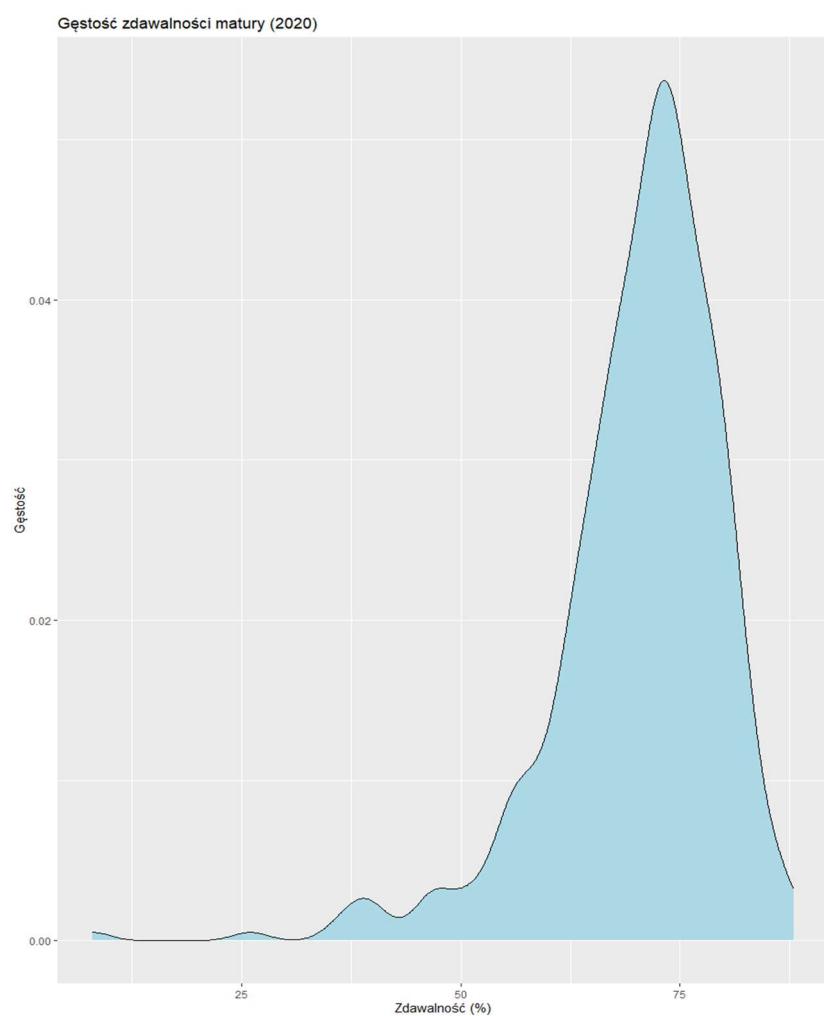
### **Wykres rozrzutu: urbanizacja vs zdawalność matury (2020)**

Na wykresie rozrzutu przedstawiono zależność między poziomem urbanizacji powiatów (na osi X) a średnią zdawalnością matury w 2020 roku (na osi Y). Każdy punkt reprezentuje jeden powiat. Dodatkowo, dodano linię regresji liniowej z przedziałem ufności (obszar zacieniowany na czerwono), która wskazuje na lekko rosnący trend — sugerując, że w bardziej zurbanizowanych obszarach średnia zdawalność matury może być nieco wyższa. Niemniej jednak, rozproszenie punktów pokazuje, że związek ten nie jest silny.



### **Wykres gęstości dla zdawalności matury:**

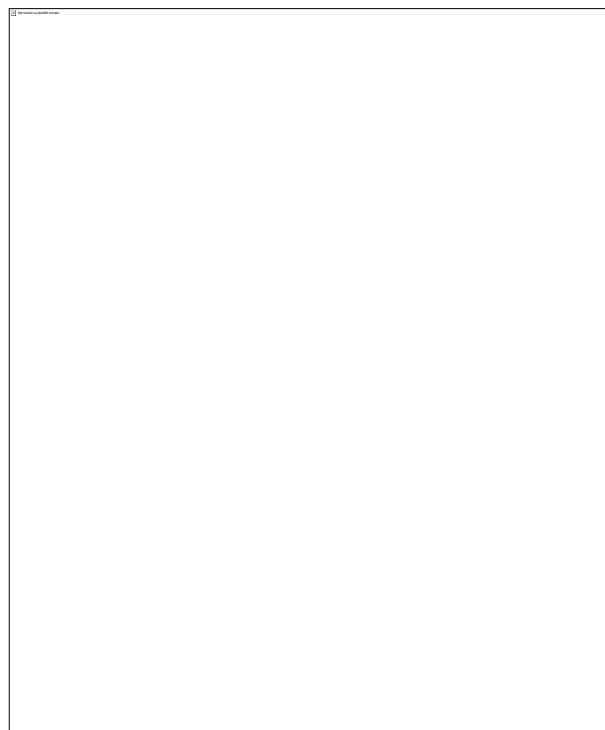
Na wykresie gęstości przedstawiono rozkład wartości zdawalności matury w 2020 roku w różnych powiatach. Oś X pokazuje procent zdawalności, a oś Y reprezentuje gęstość (czyli względnączęstość występowania danych wartości). Widzimy wyraźny szczyt w okolicach 70-75%, co oznacza, że większość powiatów osiągała wynik zdawalności właśnie w tym przedziale. Wykres ten umożliwia ocenę, jak bardzo zróżnicowane są wyniki zdawalności w Polsce.



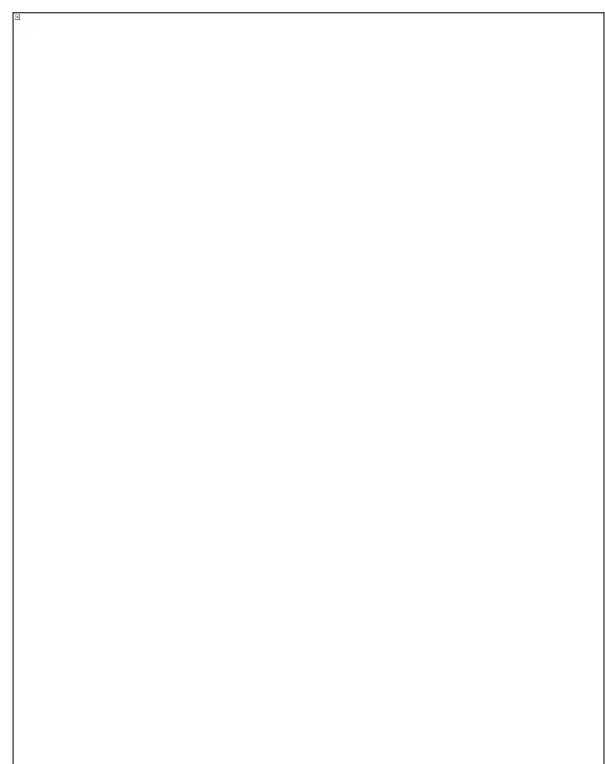
W celu wizualizacji zmiennej objaśnianej “zdawalność matury” sporządzono histogramy oraz kartogramy.

Histogramy:

Dla 2020:

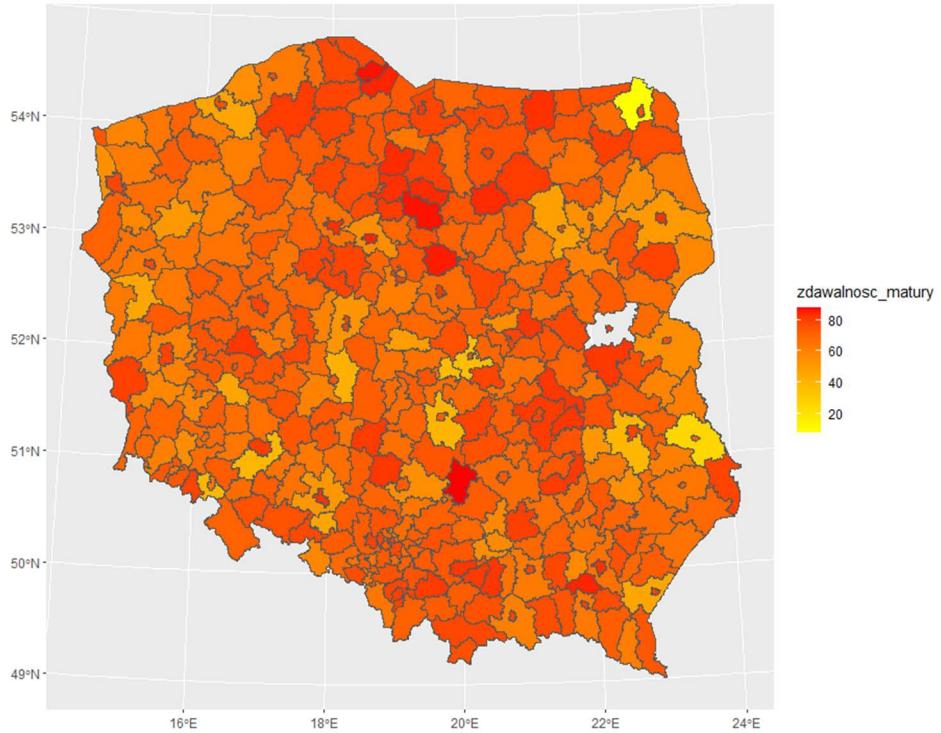


Dla 2023:

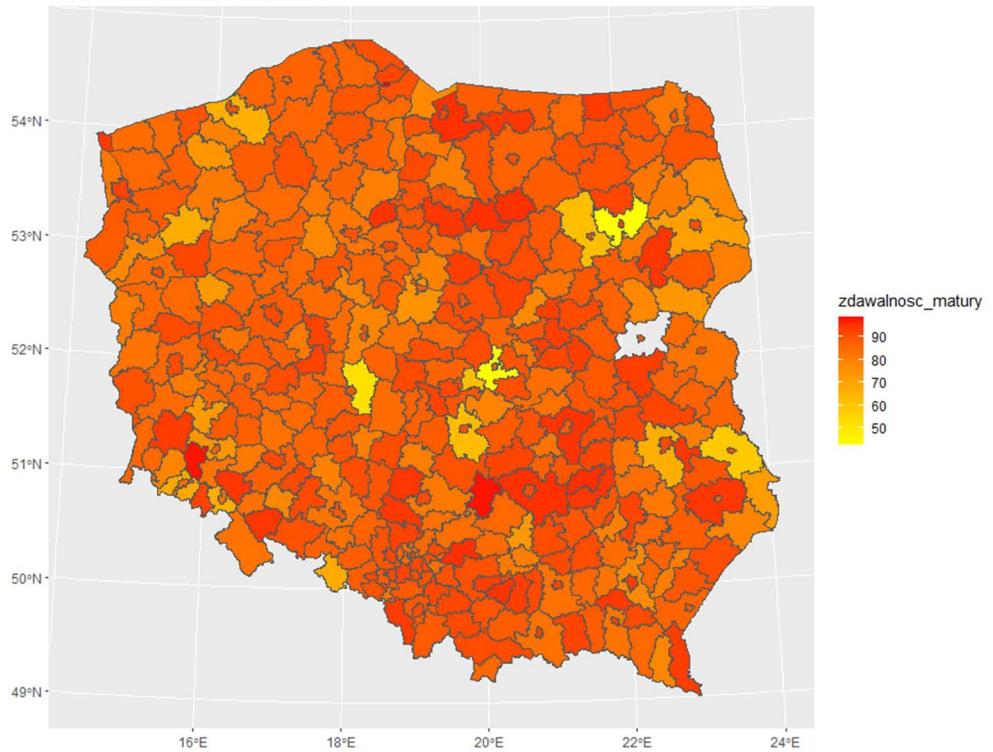


## Kartogramy:

Zdawalność matury w 2020 roku

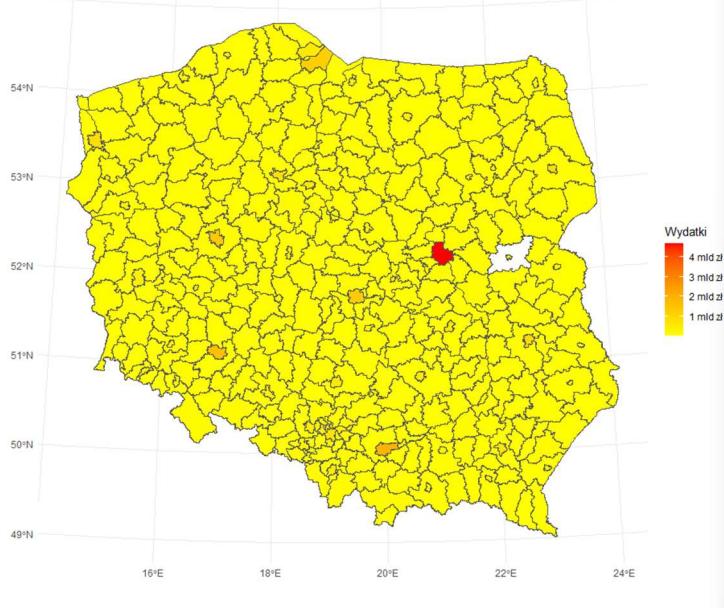


Zdawalność matury w 2023 roku

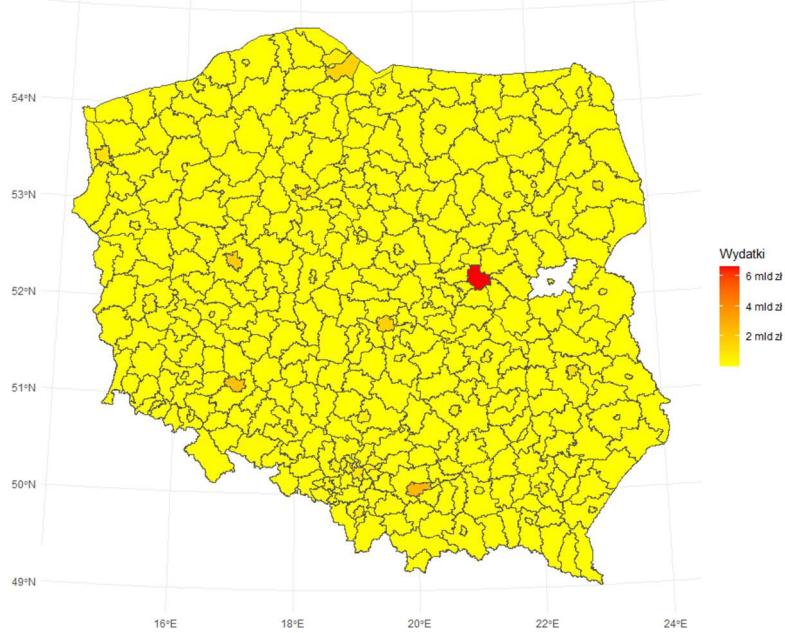


Analogiczne analizy przeprowadzono w odniesieniu do zmiennych objaśniających.  
Wydatki na edukację:

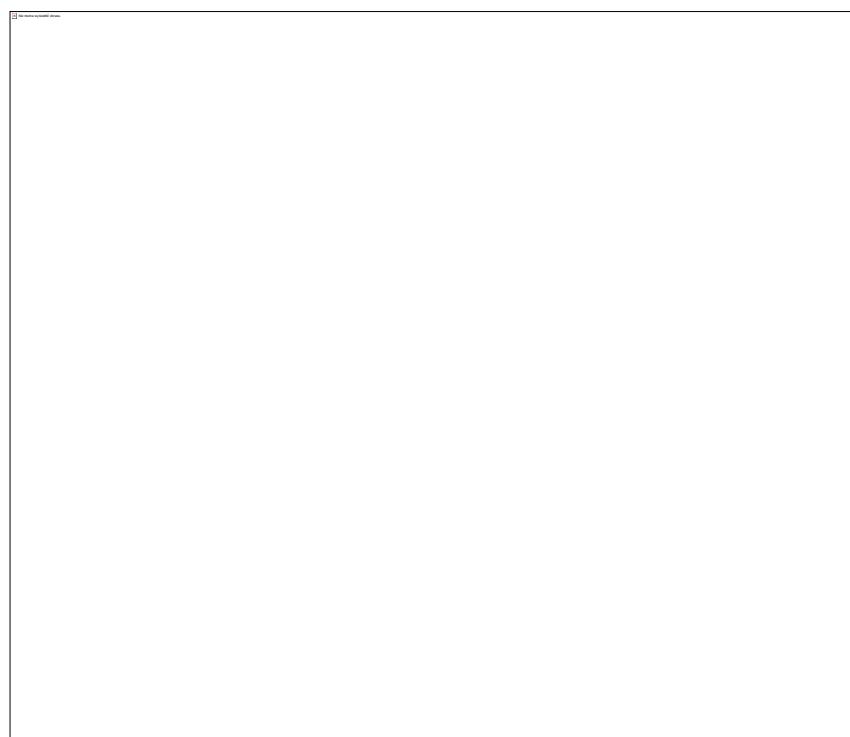
Wydatki na edukację w 2021 roku



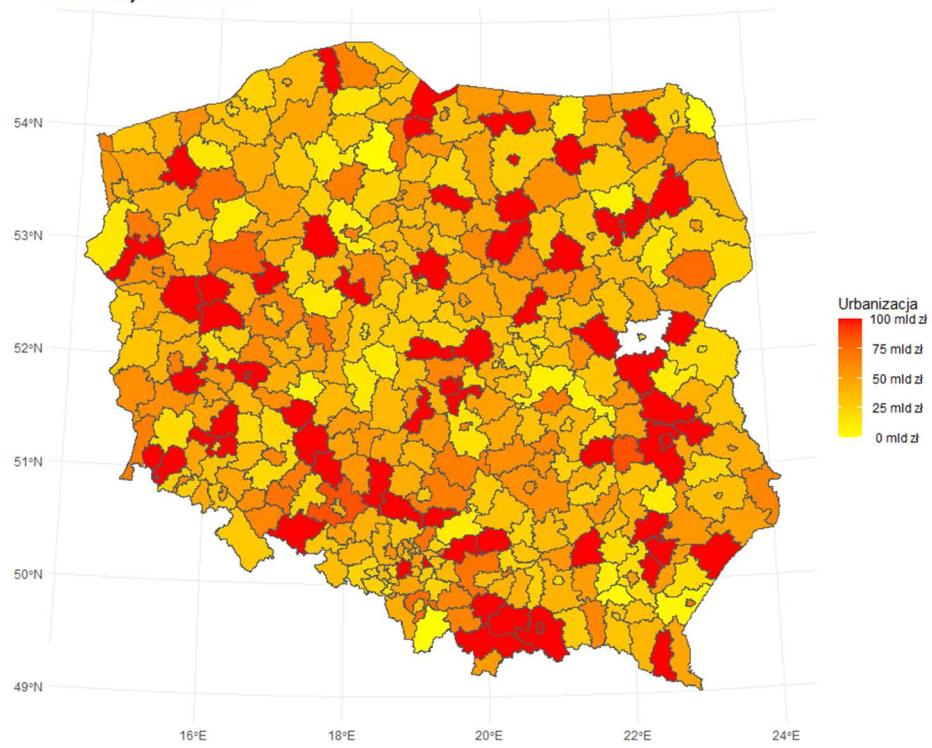
Wydatki na edukację w 2023 roku



Poziom urbanizacji:



Urbanizacja w 2023 roku

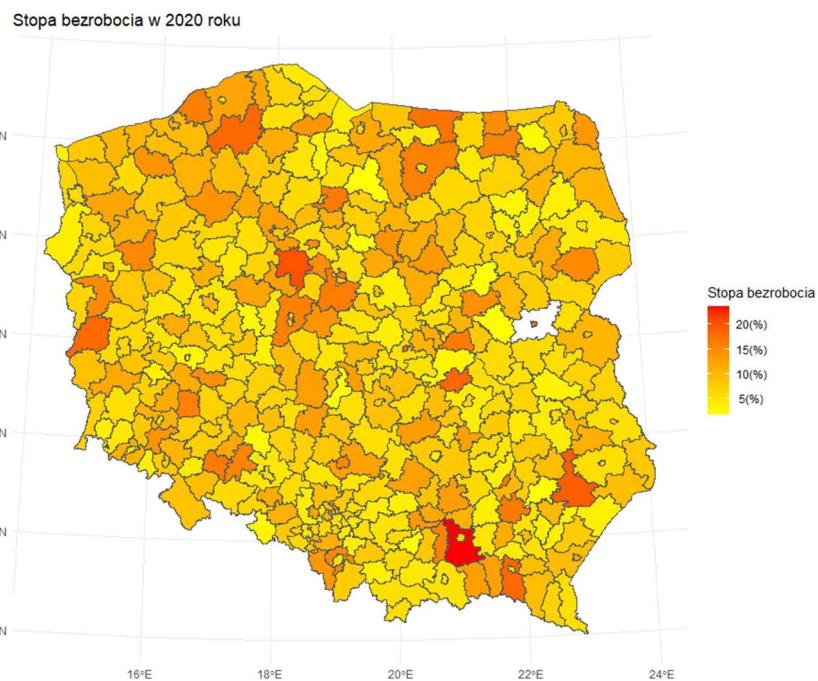


Liczba nauczycieli:

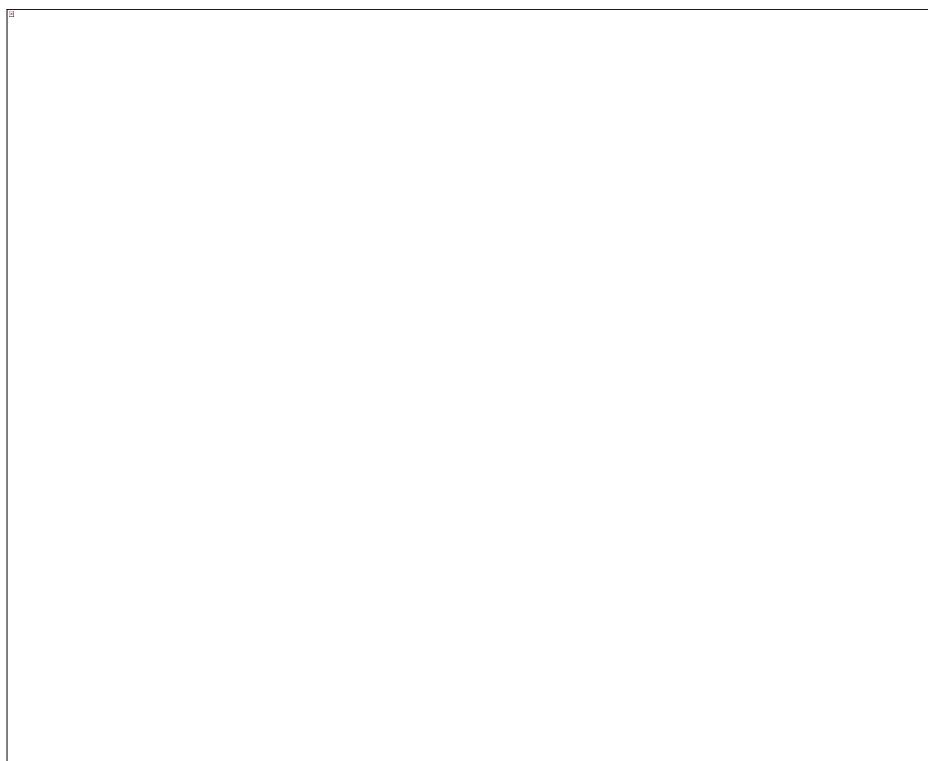
<input type="checkbox"/> 1
----------------------------

<input type="checkbox"/> 2
----------------------------

## Stopa bezrobocia:

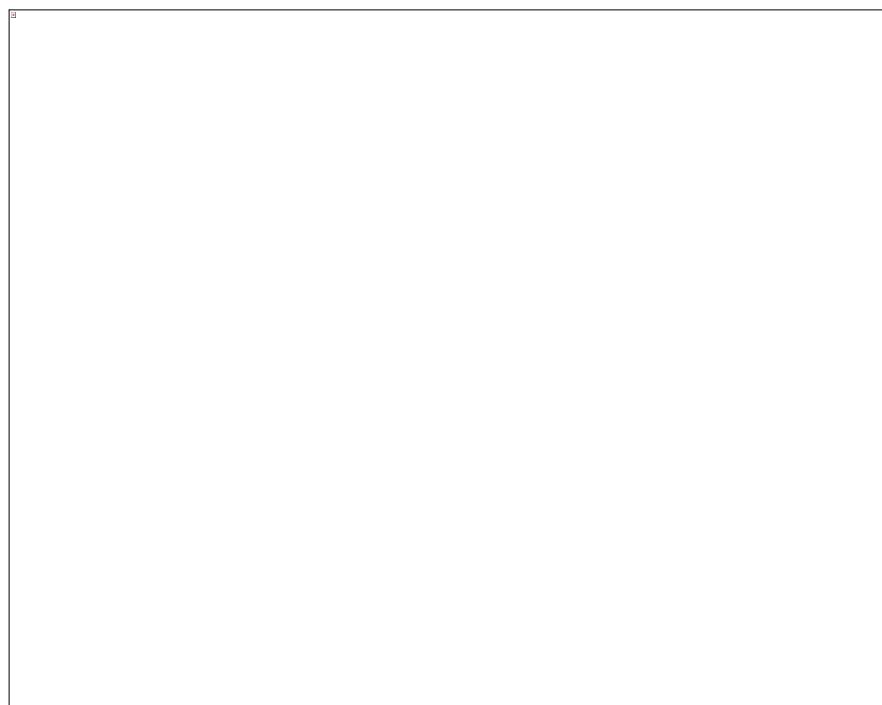
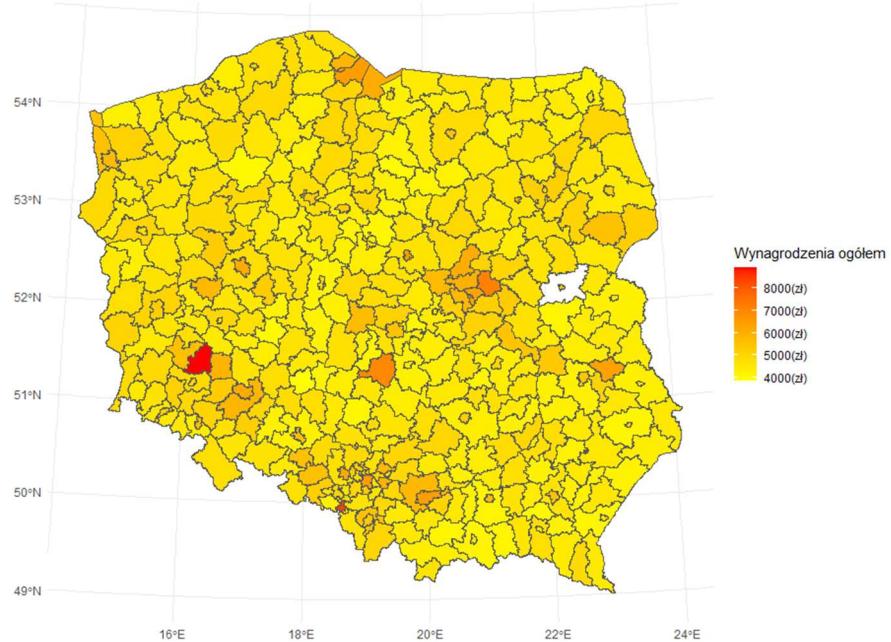


:

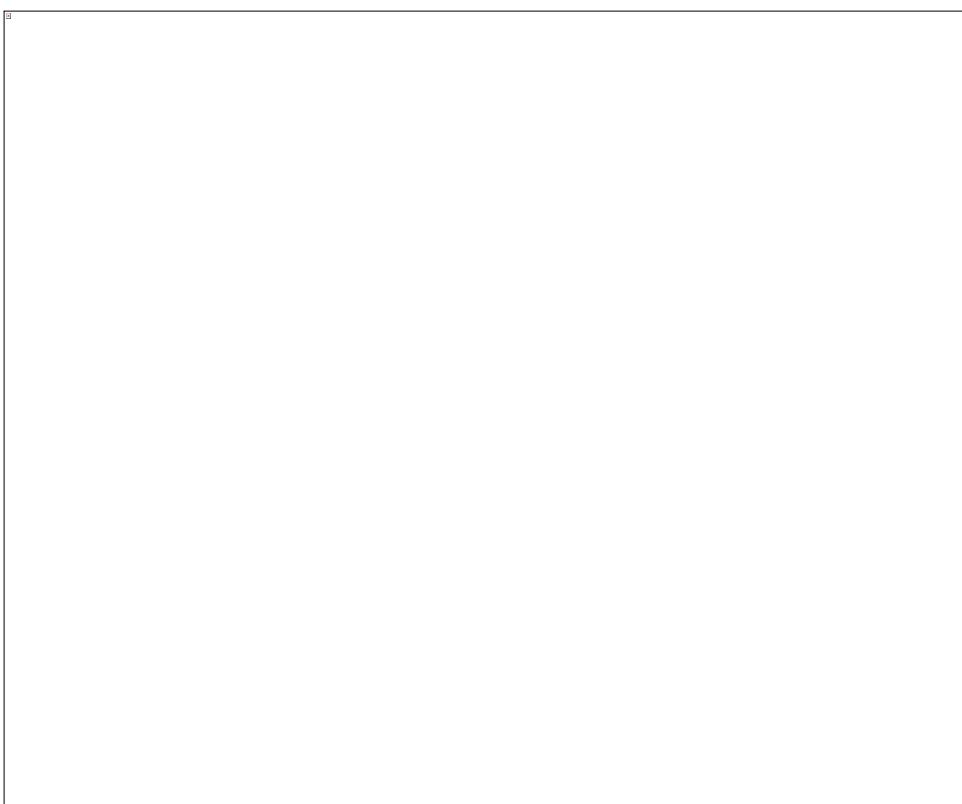
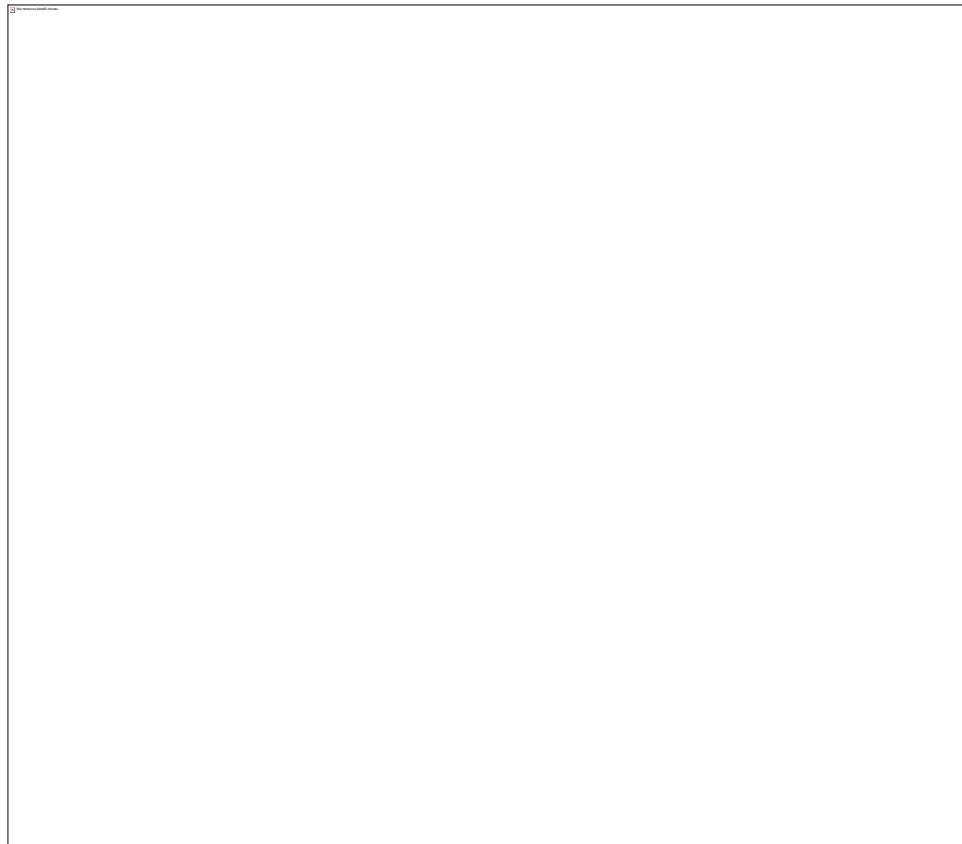


## Wynagrodzenia ogółem:

Wynagrodzenia ogółem w 2020 roku



Absolwenci ogółem:



## 5. Eksploracyjna analiza danych przestrzennych (ESDA)

### 5.1 Globalny wskaźnik autokorelacji przestrzennej Morana (Global Moran's I)

W ramach eksploracyjnej analizy danych przestrzennych (ESDA) przeprowadzono test globalnego wskaźnika Morana (Moran's I) dla siedmiu zmiennych społeczno-gospodarczych: zdawalność matury, wskaźnik urbanizacji, liczba nauczycieli, wydatki na edukację, stopa bezrobocia, wynagrodzenia ogółem oraz liczba absolwentów. Analizą objęto dane dla lat 2020–2023.

Globalny wskaźnik Morana umożliwia ocenę, czy jednostki przestrzenne (np. województwa) o zbliżonych wartościach danej cechy występują blisko siebie. Dodatni i istotny statystycznie wskaźnik świadczy o klasteryzacji przestrzennej, ujemny – o wzorze mozaikowym, natomiast wartość bliska零 wskazuje na brak autokorelacji.

Wyniki analizy według zmiennych:

#### 1. Zdawalność matury

2020:  $I = 0,0406$ ,  $p = 0,08995$

2021:  $I = 0,0122$ ,  $p = 0,3218$

2022:  $I = 0,0116$ ,  $p = 0,3278$

2023:  $I = 0,0167$ ,  $p = 0,2738$

We wszystkich latach obserwowane wartości wskaźnika były dodatnie, ale nieosiągające poziomu istotności statystycznej. Oznacza to brak wyraźnej autokorelacji przestrzennej – wyniki matur nie tworzą uporządkowanego wzoru przestrzennego.

```
=====      Globalny      Moran's      I:      zdawalnosc_matury      -      2020      =====
          Moran           I           test           under           randomisation
data:                                         mapa_roku[[zm]] 
weights:                                     listw
Moran      I      statistic      standard      deviate      =      1.3411,      p-value      =      0.08995
alternative                         hypothesis: greater
sample
Moran      I      statistic           Expectation
0.040603032   -0.002506266   0.001033295
estimates: Variance
```

```

=====      Globalny      Moran's      I:      zdawalnosc_maturity      -      2021      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]      listw
weights:

Moran      I      statistic      standard      deviate      =      0.46276,      p-value      =      0.3218
alternative      hypothesis:      greater
sample
Moran      I      statistic      Expectation      -0.002506266      estimates:      Variance
0.012174075                                         0.001006378

=====      Globalny      Moran's      I:      zdawalnosc_maturity      -      2022      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]      listw
weights:

Moran      I      statistic      standard      deviate      =      0.446,      p-value      =      0.3278
alternative      hypothesis:      greater
sample
Moran      I      statistic      Expectation      -0.0025062657      estimates:      Variance
0.0115704208      0.0009961822                                         0.001006378

=====      Globalny      Moran's      I:      zdawalnosc_maturity      -      2023      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]      listw
weights:

Moran      I      statistic      standard      deviate      =      0.60147,      p-value      =      0.2738
alternative      hypothesis:      greater
sample
Moran      I      statistic      Expectation      -0.002506266      estimates:      Variance
0.016742437      0.001024178                                         0.001006378

```

## 2. Wskaźnik urbanizacji

2020:  $I = 0,0422$ ,  $p = 0,08375$

2021:  $I = 0,0407$ ,  $p = 0,09124$

2022:  $I = 0,0399$ ,  $p = 0,09522$

2023:  $I = 0,0404$ ,  $p = 0,09279$

Wskaźnik urbanizacji utrzymywał się na stabilnym poziomie dodatnim, lecz nie osiągnął istotności statystycznej. Mimo to można zauważać delikatną tendencję do przestrzennego uporządkowania, choć niejednoznaczna.

=====	Globalny	Moran's	I:	urbanizacja_pct	-	2020	=====
Moran				test	under	randomisation	
data:							
weights:	listw	Moran	I statistic	standard	deviate =	1.3803,	p-value = 0.08375
alternative					hypothesis:		greater estimates:
sample							Variance
Moran	I statistic			Expectation			
0.042211454	-0.002506266			0.001049626			
=====	Globalny	Moran's	I:	urbanizacja_pct	-	2021	=====
Moran				test	under	randomisation	
data:							
weights:							mapa_roku[[zm]]
listw							
Moran	I statistic	standard	deviate =	1.3331,	p-value	= 0.09124	greater estimates:
alternative			hypothesis:				Variance
sample							
Moran	I statistic		Expectation				
0.040684741	-0.002506266		0.001049628				
=====	Globalny	Moran's	I:	urbanizacja_pct	-	2022	=====
Moran				test	under	randomisation	
data:							
weights:							mapa_roku[[zm]]
listw							
Moran	I statistic	standard	deviate =	1.3093,	p-value	= 0.09522	greater estimates:
alternative			hypothesis:				Variance
sample							
Moran	I statistic		Expectation				
0.039911500			-0.002506266				0.001049615
=====	Globalny	Moran's	I:	urbanizacja_pct	-	2023	=====
Moran				test	under	randomisation	
data:							
weights:							mapa_roku[[zm]]
listw							
Moran	I statistic	standard	deviate =	1.3237,	p-value	= 0.09279	greater estimates:
alternative			hypothesis:				Variance
sample							
Moran	I statistic		Expectation				
0.040379917			-0.002506266				0.001049601

### 3. Liczba nauczycieli

2020:  $I = 0,1162$ ,  $p = 3,119e-06$

2021:  $I = 0,1207$ ,  $p = 1,677e-06$

2022:  $I = 0,1247$ ,  $p = 9,33e-07$

2023:  $I = 0,1282$ ,  $p = 4,368e-07$

Wskaźnik Morana dla liczby nauczycieli wykazywał wyraźnie dodatnią i statystycznie istotną autokorelację przestrenną we wszystkich analizowanych latach. Wartości wskaźnika nieznacznie rosły w czasie, co wskazuje na narastającą przestrenną klasteryzację – województwa o podobnej liczbie nauczycieli sąsiadują ze sobą coraz wyraźniej.

```
===== Globalny Moran's I: liczba_nauczycieli - 2020 =====
Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw
Moran I statistic standard deviate = 4.5181, p-value = 3.119e-06
alternative hypothesis: greater
sample
Moran I statistic Expectation
 0.1162367458 -0.0025062657 0.0006907105
Variance

===== Globalny Moran's I: liczba_nauczycieli - 2021 =====
Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw
Moran I statistic standard deviate = 4.6479, p-value = 1.677e-06
alternative hypothesis: greater
sample
Moran I statistic Expectation
 0.1206941099 -0.0025062657 0.0007026141
Variance

===== Globalny Moran's I: liczba_nauczycieli - 2022 =====
Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw
Moran I statistic standard deviate = 4.7674, p-value = 9.33e-07
alternative hypothesis: greater
sample
Moran I statistic Expectation
 0.1247419356 -0.0025062657 0.0007124207
Variance
```

```

=====      Globalny      Moran's      I:      liczba_nauczycieli      -      2023      =====
          Moran           I           test        under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      4.9182,      p-value      =      4.368e-07
alternative                         hypothesis:
sample
Moran      I      statistic
                                         Expectation
0.1282418000   -0.0025062657   0.0007067472
                                         Variance
                                         Variance

```

#### 4. Wydatki na edukację

2020:  $I = -0,0008$ ,  $p = 0,4735$

2021:  $I = 0,0010$ ,  $p = 0,4467$

2022:  $I = 0,0014$ ,  $p = 0,4409$

2023:  $I = 0,0011$ ,  $p = 0,4447$

Wydatki na edukację nie wykazują przestrzennego uporządkowania. Wskaźniki Morana były bliskie zeru, a p-value nie wskazywało na istotność statystyczną – oznacza to przypadkowe rozmieszczenie tej zmiennej w przestrzeni.

```

=====      Globalny      Moran's      I:      wydatki_edukacja      -      2020      =====
          Moran           I           test        under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      0.066361,      p-value      =      0.4735
alternative                         hypothesis:
sample
Moran      I      statistic
                                         Expectation
-0.0008031233                           -0.0025062657
                                         Variance
                                         Variance
0.0006586734

```

```

=====      Globalny      Moran's      I:      wydatki_edukacja      -      2021      =====
          Moran           I           test        under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      0.13395,      p-value      =      0.4467
alternative                         hypothesis:
sample
Moran      I      statistic
                                         Expectation
0.0010345784   -0.0025062657   0.0006987427
                                         Variance
                                         Variance

```

```

=====          Globalny          Moran's        I:      wydatki_edukacja      -      2022      =====
          Moran           I           test         under      randomisation

data:                                 mapa_roku[[zm]]
weights:                                listw

Moran      I      statistic      standard      deviate      =      0.1488,      p-value      =      0.4409
alternative                         hypothesis:
sample
Moran      I      statistic      standard      deviate      =      0.1488,      p-value      =      0.4409
Expectation
Variance
0.0006926821

=====          Globalny          Moran's        I:      wydatki_edukacja      -      2023      =====
          Moran           I           test         under      randomisation

data:                                 mapa_roku[[zm]]
weights:                                listw

Moran      I      statistic      standard      deviate      =      0.13905,     p-value      =      0.4447
alternative                         hypothesis:
sample
Moran      I      statistic      standard      deviate      =      0.13905,     p-value      =      0.4447
Expectation
Variance
0.0006617168

```

## 5. Stopa bezrobocia

2020: I = -0,0130, p = 0,6267

2021: I = 0,0189, p = 0,2538

2022: I = 0,0123, p = 0,3237

2023: I = 0,0128, p = 0,3174

Stopa bezrobocia nie wykaz

Wartości wskaźnika są niskie i nieistotne statystycznie, co sugeruje brak trwałego wzorca przestrzennego.

```
Moran I test under randomisation
data: mapa_roku[[zm]]
weights: listw

Moran I statistic standard deviate = -0.32322, p-value = 0.6267
alternative hypothesis: greater
sample estimates:
Moran I statistic Expectation Variance
-0.012965030 -0.002506266 0.001047059
```

```

=====      Globalny      Moran's      I:      stopa_bezrobocia      -      2021      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]
weights:                                         listw

Moran      I      statistic      standard      deviate      =      0.6625,      p-value      =      0.2538
alternative
sample
Moran      I      statistic
0.018922672                                     expectation
                                         -0.002506266
estimates:
Variance
0.001046228

=====      Globalny      Moran's      I:      stopa_bezrobocia      -      2022      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]
weights:                                         listw

Moran      I      statistic      standard      deviate      =      0.45724,      p-value      =      0.3237
alternative
sample
Moran      I      statistic
0.012277163                                     expectation
                                         -0.002506266
estimates:
Variance
0.001045335

=====      Globalny      Moran's      I:      stopa_bezrobocia      -      2023      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]
weights:                                         listw

Moran      I      statistic      standard      deviate      =      0.47485,      p-value      =      0.3174
alternative
sample
Moran      I      statistic
0.012849690                                     expectation
                                         -0.002506266
estimates:
Variance
0.001045794

```

## 6. Wynagrodzenia ogółem

2020: I = 0,2427, p = 8,743e-15

2021: I = 0,2441, p = 6,094e-15

2022: I = 0,2493, p = 1,074e-15

2023: I = 0,2168, p = 1,004e-03

Wskaźnik Morana dla wynagrodzeń ogółem był najwyższy spośród wszystkich analizowanych zmiennych. W każdym roku autokorelacja była statystycznie istotna i dodatnia, przy czym wartości nieco spadły w 2023 roku. Wynika z tego wyraźna klasteryzacja przestrzenna poziomu wynagrodzeń w województwach.

```

===== Globalny Moran's I: wynagrodzenia_ogolem - 2020 =====

Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw

Moran I statistic standard deviate = 7.6679, p-value = 8.743e-15
alternative hypothesis: greater estimates: sample

Moran I statistic Expectation Variance

===== Globalny Moran's I: wynagrodzenia_ogolem - 2021 =====

Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw

Moran I statistic standard deviate = 7.7141, p-value = 6.094e-15
alternative hypothesis: greater estimates: sample

Moran I statistic Expectation Variance
0.244050579 -0.002506266 0.001021571

0.242733728 -0.002506266 0.001022897

===== Globalny Moran's I: wynagrodzenia_ogolem - 2022 =====

Moran I test under randomisation
data: mapa_roku[[zm]] weights: listw

Moran I statistic standard deviate = 7.9325, p-value = 1.074e-15
alternative hypothesis: greater estimates: sample

Moran I statistic Expectation Variance
0.249320629 -0.002506266 0.001007833

```

```

=====      Globalny      Moran's      I:      wynagrodzenia_ogolem      -      2023      =====
          Moran           I           test           under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      6.9222,      p-value      =      2.224e-12
alternative                         hypothesis:
sample
Moran      I      statistic                           Expectation
          Moran      I      statistic
          0.216829869   -0.002506266   0.001003998
estimates:                           Variance

```

## 7. Absolwenci ogółem

2020:  $I = -0,0114$ ,  $p = 0,6179$

2021:  $I = -0,0103$ ,  $p = 0,6041$

2022:  $I = -0,0104$ ,  $p = 0,6050$

2023:  $I = -0,0091$ ,  $p = 0,5875$

Wskaźniki Morana dla liczby absolwentów w każdym z lat pozostawały ujemne i nieistotne statystycznie, co sugeruje brak uporządkowania przestrzennego i możliwą mozaikowość – choć nie w sposób jednoznaczny.

```

=====      Globalny      Moran's      I:      absolwenci_ogolem      -      2020      =====
          Moran           I           test           under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      -0.29991,      p-value      =      0.6179
alternative                         hypothesis:
sample
Moran      I      statistic                           Expectation
          Moran      I      statistic
          -0.011401477   -0.002506266   0.000879695
estimates:                           Variance

```

```

=====      Globalny      Moran's      I:      absolwenci_ogolem      -      2021      =====
          Moran           I           test           under
data:                                         mapa_roku[[zm]]
weights:                                     listw
Moran      I      statistic      standard      deviate      =      -0.26392,      p-value      =      0.6041
alternative                         hypothesis:
sample
Moran      I      statistic                           Expectation
          Moran      I      statistic
          -0.0103291599   -0.0025062657   0.0008786036
estimates:                           Variance

```

```

=====      Globalny      Moran's      I:      absolwenci_ogolem      -      2022      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]
weights:                                         listw
Moran      I      statistic      standard      deviate      =      -0.26626,      p-value      =      0.605
alternative
sample
Moran      I      statistic
-0.0104087028
                           Expectation
                           -0.0025062657
                           Variance
                           0.0008808824
=====      Globalny      Moran's      I:      absolwenci_ogolem      -      2023      =====
          Moran           I           test        under      randomisation
data:                                           mapa_roku[[zm]]
weights:                                         listw
Moran      I      statistic      standard      deviate      =      -0.22101,      p-value      =      0.5875
alternative
sample
Moran      I      statistic
-0.0090660756   -0.0025062657   0.0008809725
                           Expectation
                           Variance

```

## 5.2 Obliczenie lokalnych wskaźników LISA i ich wizualizacja

W ramach analizy przestrzennej wyników egzaminu maturalnego za rok 2023 obliczono lokalne wskaźniki autokorelacji przestrzennej LISA. Celem tego etapu było zidentyfikowanie lokalnych klastrów wysokiej i niskiej zdawalności, a także obszarów, gdzie wyniki nie wykazywały istotnej przestrzennej zależności.

Na przygotowanej mapie zaznaczono trzy typy obszarów:

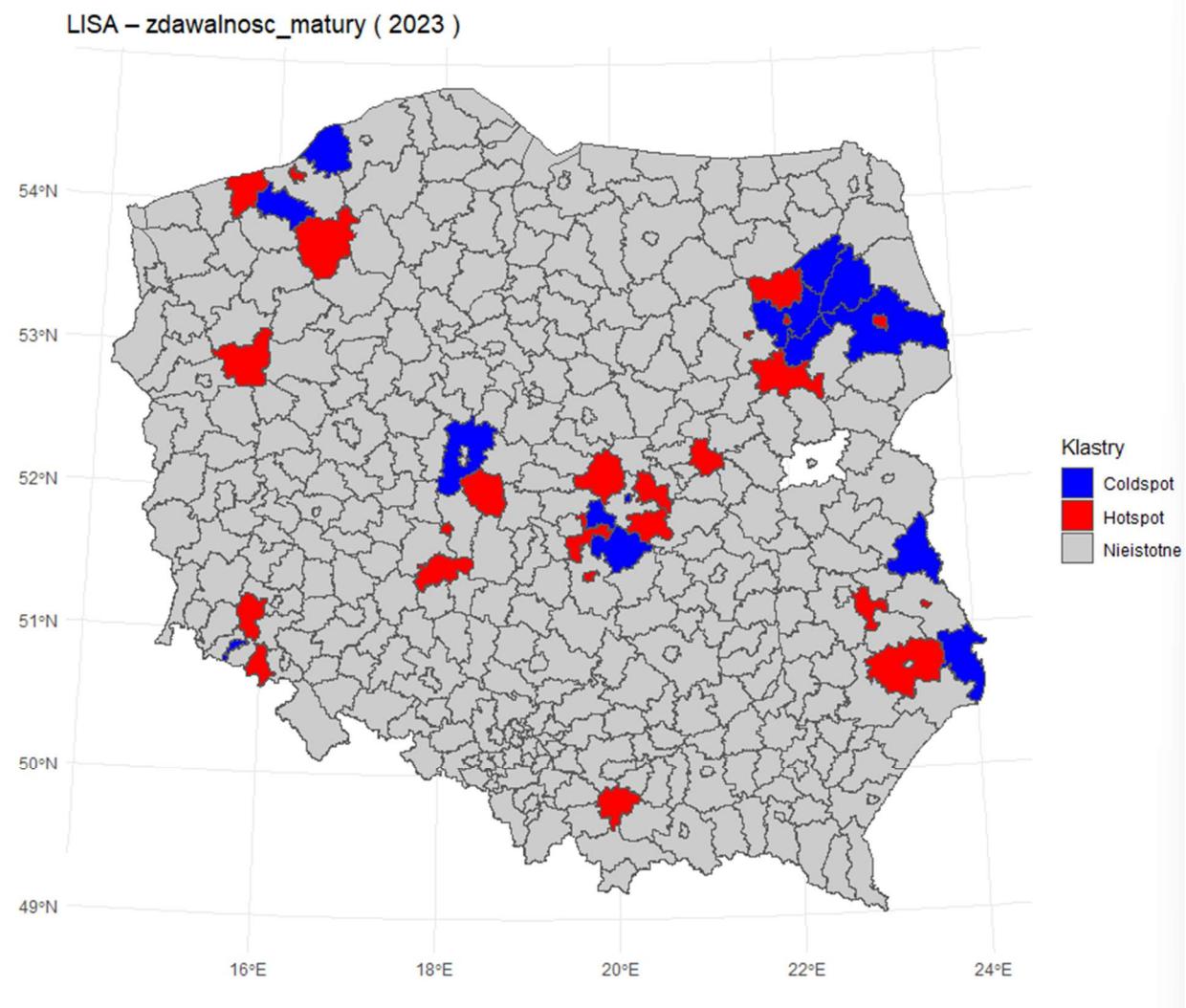
- Hotspotty (kolor czerwony) – powiaty o wysokiej zdawalności, otoczone przez inne jednostki o równie dobrych wynikach,
- Coldspotty (kolor niebieski) – powiaty o niskiej zdawalności, znajdujące się w sąsiedztwie innych obszarów o słabych wynikach,
- Nieistotne statystycznie (kolor szary) – powiaty, dla których nie stwierdzono istotnej autokorelacji przestrzennej.

Wyniki wskazują na wyraźne regionalne zróżnicowanie zdawalności. Do najbardziej widocznych coldspotów należą obszary w północno-wschodniej Polsce (w szczególności województwo warmińsko-mazurskie i część podlaskiego), południowo-wschodnia Polska (część woj. podkarpackiego i lubelskiego), a także fragmenty centralnej i południowo-zachodniej Polski. Sugeruje to istnienie regionów strukturalnie słabszych edukacyjnie, gdzie problemy mogą mieć charakter trwały i systemowy.

Hotspotty występują głównie w okolicach dużych ośrodków miejskich, takich jak Warszawa, Kraków, Lublin, Poznań, Wrocław czy Trójmiasto. Widoczne są również

klastry wysokiej zdawalności na Górnym Śląsku oraz w centralnej Polsce. Wskazuje to na koncentrację lepszych wyników w rejonach silniejszych gospodarczo i edukacyjnie, co może być związane z większą dostępnością do wysokiej jakości szkół średnich, lepszym zapleczem edukacyjnym oraz wyższym kapitałem społecznym i kulturowym uczniów.

W pozostałych powiatach nie stwierdzono istotnej przestrzennej zależności, co oznacza, że ich wyniki nie tworzą z sąsiadującymi obszarami spójnych klastrów.



W pozostałych powiatach nie stwierdzono istotnej przestrzennej zależności, co oznacza, że ich wyniki nie tworzą z sąsiadującymi obszarami spójnych klastrów. Duży

udział obszarów oznaczonych jako „nieistotne” jest zjawiskiem normalnym i może wynikać z relatywnej jednorodności przestrzennej danych

W poniższej tabeli przedstawione są istotne lokalne klastry LISA, czyli obserwacje, które charakteryzują się istotnym wskaźnikiem lokalnego Morana ( $p\text{-value} < 0,05$ ). Zostały one sklasyfikowane jako hotspots, jak na przykład Gdynia, lub coldspots, takie jak powiat krakowski, trzebnicki czy wrocławski. Tabela zawiera również kolumnę z geometrią, co potwierdza, że obiekty mają charakter przestrzenny i należą do klasy sf.

	Powiat	I <sub>i</sub>	P_value	Klastry	geometry
1	Gdynia	0.72884231	3.724964e-02	Hotspot MULTIPOLYGON (((471552.6 73...	
2	krakowski	-0.21827369	1.681864e-03	Coldspot MULTIPOLYGON (((574676.5 24...	
3	proszowicki	-0.24827710	7.663590e-04	Coldspot MULTIPOLYGON (((596694 2664...	
4	Siemianowice Śląskie	-0.21395661	4.137616e-02	Coldspot MULTIPOLYGON (((503603.3 26...	
5	Sopot	-0.07226971	1.957968e-05	Coldspot MULTIPOLYGON (((473466.3 72...	
6	średzki	-0.16672611	1.643585e-02	Coldspot MULTIPOLYGON (((325175.9 37...	
7	średzki	-0.16672611	1.643585e-02	Coldspot MULTIPOLYGON (((325175.9 37...	
8	trzebnicki	-0.16568052	1.712397e-02	Coldspot MULTIPOLYGON (((365680.1 36...	
9	wielicki	-0.54594150	1.926066e-07	Coldspot MULTIPOLYGON (((587870.6 24...	
10	wrocławski	-0.14563414	2.699640e-02	Coldspot MULTIPOLYGON (((372291.4 36...	

## 6. Modelowanie ekonometryczne

- Estymacja klasycznego modelu OLS (dla porównania)

Podczas przygotowania danych do estymacji modeli ekonometrycznych przeprowadzono analizę współzależności pomiędzy zmiennymi objaśniającymi. W tym celu obliczono **macierz korelacji Pearsona** dla wszystkich zmiennych ilościowych, co pozwoliło na zidentyfikowanie par zmiennych o bardzo wysokiej korelacji liniowej. Szczególnie wysoka wartość współczynnika korelacji została zaobserwowana między zmiennymi liczba\_nauczycieli a wydatki\_edukacja (**r = 0,95**) oraz między absolwenci\_ogółem a tymi samymi zmiennymi (rzędu 0,56–0,58).

Taka sytuacja wskazuje na występowanie **wielokolinearności**, czyli silnej liniowej współzależności pomiędzy zmiennymi objaśniającymi. Wielokolinearność jest jednym z głównych problemów w klasycznej analizie regresji i modelowaniu przestrzennym, ponieważ:

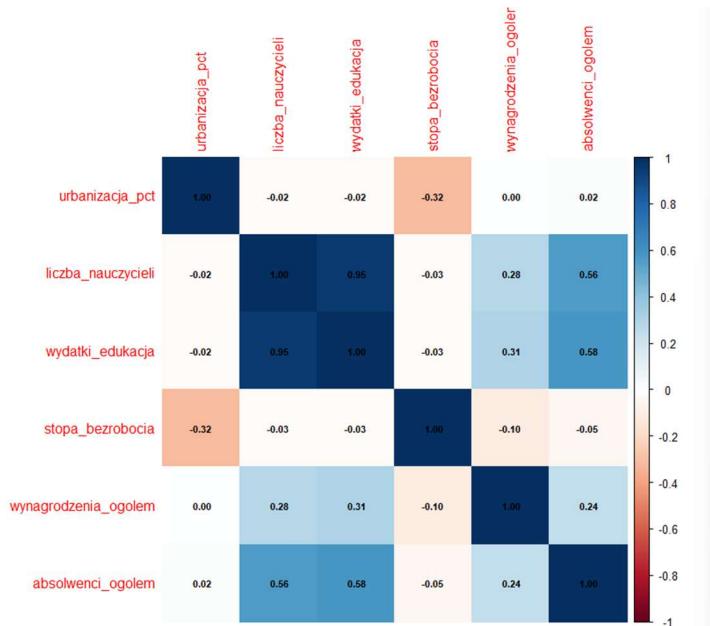
- **zaburza estymację współczynników regresji** – wartości estymatorów stają się niestabilne i bardzo wrażliwe na nawet niewielkie zmiany w danych;
- **prowadzi do dużych błędów standardowych** współczynników, co skutkuje niską istotnością statystyczną zmiennych;
- **utrudnia interpretację wyników** – nie wiadomo, która zmienna faktycznie wpływa na zmienną zależną, ponieważ ich informacje się nakładają;

- może prowadzić do błędów numerycznych, zwłaszcza w modelach przestrzennych (SAR, SEM), co w analizie ujawniło się błędem osobliwości macierzy (system is computationally singular).

W celu **zminimalizowania ryzyka wielokolinearności** oraz zapewnienia poprawnej estymacji modeli, podjęto decyzję o usunięciu najbardziej problematycznych zmiennych. Zachowano zmienne reprezentatywne o niższej korelacji, które wciąż miały potencjalne znaczenie teoretyczne i empiryczne w wyjaśnianiu zmiennej zależnej (zdawalnosc\_matury). Ostatecznie w estymowanych modelach uwzględniono:

- urbanizacja\_pct** – jako wskaźnik poziomu urbanizacji;
- wynagrodzenia\_ogolem** – jako wskaźnik dobrobytu i potencjału edukacyjnego regionu;
- stopa\_bezrobocia** – jako negatywny czynnik społeczno-ekonomiczny;
- absolwenci\_ogolem**

Dzięki tej selekcji uzyskano modele o większej stabilności, lepszym dopasowaniu oraz poprawnej interpretowalności ekonomicznej i statystycznej.



```

> # ===== MODEL OLS =====
> cat("\n==== MODEL OLS ====\n")

==== MODEL OLS ====
> model_ols <- lm(zdawalnosc_matury ~ urbanizacja_pct + wynagrodzenia_ogolem +
+                      stopa_bezrobocia + absolwenci_ogolem, data = mapa_dane_bal)
> summary(model_ols)

Call:
lm(formula = zdawalnosc_matury ~ urbanizacja_pct + wynagrodzenia_ogolem +
    stopa_bezrobocia + absolwenci_ogolem, data = mapa_dane_bal)

Residuals:
    Min      1Q  Median      3Q     Max 
-80.829 -4.219  1.495  6.135 29.825 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 5.455e+01 1.645e+00 33.154   <2e-16 ***
urbanizacja_pct -6.832e-03 9.173e-03 -0.745    0.457    
wynagrodzenia_ogolem 4.515e-03 2.482e-04 18.188   <2e-16 ***
stopa_bezrobocia -7.994e-02 6.093e-02 -1.312    0.190    
absolwenci_ogolem 1.205e-04 8.278e-05  1.456    0.146    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.458 on 1595 degrees of freedom
Multiple R-squared:  0.1917,    Adjusted R-squared:  0.1897 
F-statistic: 94.56 on 4 and 1595 DF,  p-value: < 2.2e-16

```

## Opis wyników klasycznego modelu OLS (2020–2023)

W ramach analizy statystycznej przeprowadzono estymację klasycznego modelu regresji liniowej (OLS), którego celem było określenie wpływu wybranych czynników społeczno-ekonomicznych na poziom **zdawalności matury** w polskich powiatach w latach **2020–2023**. Model został zbudowany na danych panelowych, obejmujących wyłącznie te jednostki, które były obecne w każdym z analizowanych lat, co pozwoliło zachować spójność próbki i zapewnić równowagę panelu.

Zmienną objaśnianą (zależną) była **zdawalność matury** wyrażona w procentach, natomiast zmiennymi objaśniającymi były:

- **urbanizacja\_pct** – udział ludności miejskiej w populacji powiatu,
- **wynagrodzenia\_ogolem** – przeciętne miesięczne wynagrodzenie brutto ogółem,
- **stopa\_bezrobocia** – poziom bezrobocia rejestrowanego,
- **absolwenci\_ogolem** – liczba absolwentów szkół ponadpodstawowych.

## **Wyniki modelu i ich interpretacja**

Model OLS okazał się statystycznie istotny na poziomie ogólnym, co potwierdza **wartość statystyki F (94,56)** oraz bardzo niska wartość p ( $< 2.2\text{e-}16$ ). Oznacza to, że przynajmniej jedna ze zmiennych objaśniających wywiera istotny wpływ na zdawalność matury.

Wskaźnik **R<sup>2</sup> = 0,1917** (oraz skorygowany R<sup>2</sup> = 0,1897) sugeruje, że około **19% zmienności zdawalności matury** zostało wyjaśnione przez model. Choć może się to wydawać niewielką wartością, należy pamiętać, że zdawalność matury jest zjawiskiem złożonym, zależnym również od wielu czynników niemierzalnych (np. motywacja uczniów, jakość nauczania), które nie zostały ujęte w modelu.

### **interpretacja współczynników regresji**

Z czterech analizowanych zmiennych, **jedynie przeciętne wynagrodzenie ogółem (wynagrodzenia\_ogolem)** okazało się istotnym statystycznie czynnikiem wpływającym na zdawalność matury ( $p < 0.001$ ). Wartość estymowanego współczynnika wyniosła **0.0045**, co oznacza, że wzrost przeciętnego wynagrodzenia o 1000 zł wiąże się ze **wzrostem zdawalności matury o około 4,5 punktu procentowego**. Wskazuje to, że **poziom dochodów w regionie może być silnym determinantem edukacyjnych sukcesów uczniów** – lepsze warunki materialne mogą oznaczać wyższe inwestycje w edukację (np. korepetycje, sprzęt, książki), stabilniejsze środowisko rodzinne oraz lepszy dostęp do zasobów edukacyjnych.

Pozostałe zmienne – czyli **urbanizacja, stopa bezrobocia** oraz **liczba absolwentów – nie osiągnęły istotności statystycznej**, co sugeruje, że ich wpływ na poziom zdawalności nie jest wyraźny lub jest tłumiony przez inne czynniki nieujęte w modelu.

### **Podsumowanie**

Model OLS dostarczył cennych informacji na temat czynników wpływających na zdawalność matury. Najważniejszym wnioskiem płynącym z tej części analizy jest **istotna rola wynagrodzeń** jako czynnika pozytywnie wpływającego na wyniki uczniów. Brak istotności pozostałych zmiennych może wynikać zarówno z ich słabszej korelacji ze zdawalnością, jak i z ewentualnych efektów przestrzennych, które zostaną uwzględnione i dokładniej zbadane w dalszej części analizy – w modelach przestrzennych typu **SAR** i **SEM**.

- Budowa modelu przestrzennego: SAR (model z opóźnieniem przestrzennym) lub SEM (model błędu przestrzennego)

```
> summary(model_sar)

Call:lagsarlm(formula = zdawalnosc_maturity ~ urbanizacja_pct + wynagrodzenia_ogolem +
  stopa_bezrobocia + absolwenci_ogolem, data = mapa_dane_bal,
  listw = listw, method = "eigen", zero.policy = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max 
-80.8201 -4.2741  1.3771  6.2199 29.0348 

Type: lag
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)    
(Intercept) 3.7178e+01 4.8747e+00 7.6267 2.398e-14  
urbanizacja_pct -6.8009e-03 9.1032e-03 -0.7471 0.4550  
wynagrodzenia_ogolem 4.5166e-03 2.5118e-04 17.9811 < 2.2e-16  
stopa_bezrobocia -7.0537e-02 6.0476e-02 -1.1664 0.2435  
absolwenci_ogolem 9.7700e-05 8.2396e-05 1.1857 0.2357  

Rho: 0.22162, LR test value: 17.91, p-value: 2.3164e-05
Asymptotic standard error: 0.06127
  z-value: 3.6172, p-value: 0.00029783
Wald statistic: 13.084, p-value: 0.00029783

Log likelihood: -5853.759 for lag model
ML residual variance (sigma squared): 87.976, (sigma: 9.3796)
Number of observations: 1600
Number of parameters estimated: 7
AIC: 11722, (AIC for lm: 11737)
LM test for residual autocorrelation
test value: 53.237, p-value: 2.9565e-13

> summary(model_sem)

Call:errorsarlm(formula = zdawalnosc_maturity ~ urbanizacja_pct + wynagrodzenia_ogolem +
  stopa_bezrobocia + absolwenci_ogolem, data = mapa_dane_bal,
  listw = listw, method = "eigen", zero.policy = TRUE)

Residuals:
    Min      1Q  Median      3Q     Max 
-81.2784 -4.1134  1.3754  5.8900 30.9339 

Type: error
Coefficients: (asymptotic standard errors)
              Estimate Std. Error z value Pr(>|z|)    
(Intercept) 5.2492e+01 1.6855e+00 31.1438 <2e-16  
urbanizacja_pct -9.5252e-03 9.3848e-03 -1.0150 0.3101  
wynagrodzenia_ogolem 4.9793e-03 2.5261e-04 19.7110 <2e-16  
stopa_bezrobocia -7.7403e-02 6.1683e-02 -1.2548 0.2095  
absolwenci_ogolem 1.1651e-04 8.7207e-05 1.3360 0.1816  

Lambda: 0.3696, LR test value: 51.453, p-value: 7.3341e-13
Asymptotic standard error: 0.063404
  z-value: 5.8293, p-value: 5.5648e-09
Wald statistic: 33.981, p-value: 5.5648e-09

Log likelihood: -5836.988 for error model
ML residual variance (sigma squared): 85.764, (sigma: 9.2609)
Number of observations: 1600
Number of parameters estimated: 7
AIC: 11688, (AIC for lm: 11737)
```

# Analiza przestrzenna – modele SAR i SEM (2020–2023)

Aby lepiej uwzględnić możliwe zależności przestrzenne pomiędzy powiatami, zastosowano dwa modele ekonometryczne uwzględniające strukturę przestrzenną: **model SAR** (z opóźnieniem przestrzennym zmiennej zależnej) oraz **model SEM** (z błędem przestrzennym). Modele te mają na celu uwzględnienie faktu, że wyniki edukacyjne w danym powiecie mogą być powiązane z wynikami w sąsiednich jednostkach administracyjnych.

## Model SAR (Spatial Autoregressive Model)

Model SAR zakłada, że **zdawalność matury w danym powiecie zależy nie tylko od jego własnych cech**, ale również od poziomu zdawalności w sąsiednich powiatach. Kluczowym elementem tego modelu jest **parametr rho ( $\rho$ )**, który informuje o sile autokorelacji przestrzennej w zmiennej zależnej.

### Najważniejsze wyniki:

- **Parametr rho ( $\rho$ ):** 0.2216 – istotny statystycznie ( $p < 0.001$ ), co oznacza, że zdawalność matury w danym powiecie **jest dodatnio powiązana ze zdawalnością w sąsiedztwie**.
- **Wynagrodzenia ogółem** – ponownie okazały się **jedyną istotną statystycznie zmienną** ( $p < 0.001$ ), z dodatnim wpływem (0.0045), co potwierdza wcześniejsze wyniki z modelu OLS.
- Pozostałe zmienne, takie jak **urbanizacja, stopa bezrobocia i liczba absolwentów**, były **nieistotne statystycznie**.
- **AIC (kryterium informacyjne Akaike'a): 11722**, co oznacza, że model SAR lepiej dopasowuje się do danych niż model OLS (dla którego  $AIC = 11737$ ).

### Wniosek:

Model SAR pokazuje, że istnieją **istotne efekty przestrzenne** – sukcesy edukacyjne w jednym powiecie są powiązane z sukcesami w sąsiednich. Włączenie tego efektu poprawiło dopasowanie modelu.

## Model SEM (Spatial Error Model)

Model SEM zakłada, że **zależność przestrzenna występuje nie w samej zmiennej zależnej, ale w błędach modelu** – czyli istnieją **niewidoczne czynniki przestrzenne**, które wpływają na wyniki, ale nie zostały uwzględnione jako zmienne objaśniające.

### Najważniejsze wyniki:

- **Parametr lambda ( $\lambda$ ):** 0.3696 – **silnie istotny statystycznie** ( $p < 0.001$ ), co potwierdza obecność autokorelacji przestrzennej w składniku resztowym modelu.
  - Ponownie, **wynagrodzenia ogółem** są **istotnie dodatnio skorelowane** ze zdawalnością ( $p < 0.001$ ).
  - Pozostałe zmienne pozostają **nieistotne statystycznie**.
  - **AIC modelu SEM wynosi 11688**, a więc **najniższy spośród wszystkich testowanych modeli** (OLS – 11737, SAR – 11722), co sugeruje, że **model SEM jest najlepiej dopasowany do danych**.
- 
- Opcjonalnie: model Durbin lub GWR (geograficznie ważona regresja)

## 7. Ocena modelu

- Testy autokorelacji reszt (Moran I dla reszt, testy LM)

```
library(spdep)
library(spatialreg)
library(lmtest)

# Moran's I test
moran.I(ols_resid, listw)
```

Moran I statistic standard deviate = 11.291, p-value < 2.2e-16  
alternative hypothesis: greater  
sample estimates:  
Moran I statistic      Expectation      Variance  
8.322084e-02      -6.253909e-04      5.514655e-05
- Porównanie dopasowania modeli (AIC,  $R^2$ )

```
library(spatialreg)
library(lmtest)
library(spatial
```

==== PORÓWNANIE MODELI: AIC ====  
> cat("AIC - OLS:", AIC(model\_ols), "\n")  
AIC - OLS: 11737.43  
> cat("AIC - SAR:", AIC(model\_sar), "\n")  
AIC - SAR: 11721.52  
> cat("AIC - SEM:", AIC(model\_sem), "\n")  
AIC - SEM: 11687.98

Model **SEM (Spatial Error Model)** osiągnął **najniższą wartość AIC**, co oznacza, że **najlepiej opisuje zmienność zdawalności matury** w ujęciu przestrzennym w latach 2020–2023. Ujęcie przestrzennej autokorelacji w składniku błędu pozwala na uchwycenie zależności, które są niewidoczne w klasycznym modelu OLS i tylko częściowo ujęte w SAR.

- Ocena statystyczna i ekonomiczna modelu, badanie istotności wyników

## 8. Wnioski i rekomendacje

- Kluczowe czynniki wpływające na badane zjawisko
- Znaczenie efektów przestrzennych
- Wnioski dla polityki publicznej / decyzji lokalnych
- Krytyczna ocena ograniczeń danych i metod

Bibliografia:

1. „System wsparcia kapitału intelektualnego przez organizacje...” (UWB, Stalończyk 2018)
2. „Education Quality and its Drivers in Rural Areas of Poland” (Czyżewski & Polcyn 2016)
3. „Earnings and Outside Options: Evidence from Poland” (lata 90.):
4. „Regional Wage Differentials in Poland” (Adamchik & Hyclak 2016)  
zróżnicowanie kapitału ludzkiego w Polsce” (Cejsh 2001–2011)
5. „Efektywność finansowania polskiego systemu oświaty – analiza porównawcza” (Sabina Waluś, 2018)
6. „Produktywność polskiego szkolnictwa średniego” (Małgorzata Gałecka, 2022)
7. Dane OECD „Education at a Glance 2023” – finansowanie edukacji w Polsce
8. Zdawalność matur to dane CKE i OKE
9. <https://powiatlidzbarski.pl/files/image/2023/Projekt-Za%C5%82%C4%85cznika-nr-1-%E2%80%9EDiagnoza-spo%C5%82eczno-gospodarcza-Powiatu-Lidzbarskiego%E2%80%9D.pdf>
10. „Przestrzenne zróżnicowanie kapitału ludzkiego w Polsce” (Cejsh 2001–2011)