

Aleksandra Femin

Struktura transportu i jej zmiany w Polsce
w latach 2010 - 2022

Spis treści

1.	Wstęp.....	6
2.	Opis użytych danych	7
3.	Wczytanie danych do środowiska R	9
3.1.	Instalowanie i ładowanie bibliotek.....	9
3.2.	Wczytanie danych z Excela.....	10
3.3.	Przygotowanie danych do dalszego przetwarzania	10
4.	Wyznaczanie podstawowych parametrów opisowych.....	11
4.1.	Wyznaczanie podstawowych parametrów dla przewozów ładunków	11
4.1.1.	Wartość maksymalna.....	11
4.1.2.	Wartość minimalna	11
4.1.3.	Odchylenie standardowe.....	12
4.1.4.	Średnia	13
4.1.5.	Mediana	13
4.1.6.	Kwartyle.....	14
4.1.7.	Rozstęp międzykwartylowy.....	15
4.1.8.	Wariancja	16
4.1.9.	Współczynnik korelacji	17
4.1.10.	Kowariancja	18
4.2.	Wyznaczanie podstawowych parametrów dla przewozów pasażerów	19
4.2.1.	Wartość maksymalna.....	19
4.2.2.	Wartość minimalna	19
4.2.3.	Odchylenie standardowe.....	20
4.2.4.	Średnia	20
4.2.5.	Mediana	21
4.2.6.	Kwartyle.....	21

4.2.7. Rozstęp międzykwartylowy.....	22
4.2.8. Wariancja	23
4.2.9. Współczynnik korelacji	23
4.2.10. Kowariancja	24
4.3. Podsumowanie podstawowych parametrów opisowych	25
5. Graficzna prezentacja danych	26
5.1. Wykresy kołowe	26
5.1.1. Wykres kołowy prezentujący udział rodzajów transportu w przewozach ładunków	26
5.1.2. Wykres kołowy prezentujący udział rodzajów transportu w przewozach pasażerów	27
5.2. Wykresy liniowe.....	28
5.2.1. Wykres liniowy prezentujący przewozy ładunków i pasażerów na przestrzeni lat	28
5.2.2. Wykres liniowy przedstawiający przewóz ładunków na przestrzeni lat z podziałem na rodzaj transportu	30
5.2.3. Wykres liniowy przedstawiający przewóz pasażerów na przestrzeni lat z podziałem na rodzaj transportu	31
5.3. Histogramy	33
5.3.1. Histogram przewozów ładunków	33
5.3.2. Histogram przewozów pasażerów	34
5.4. Wykresy słupkowe	35
5.4.1. Wykres słupkowy przewozów ładunków w poszczególnych latach z podziałem na rodzaj transportu	35
5.4.2. Wykres słupkowy przewozów pasażerów w poszczególnych latach z podziałem na rodzaj transportu	37
5.5. Wykresy punktowe	38
5.5.1. Wykres punktowy dla przewozów pasażerów (kolejowy vs samochodowy)	38

5.5.2. Wykres punktowy dla przewozów ładunków (kolejowy vs samochodowy)	39
5.6. Wykresy pudełkowe	40
5.6.1. Wykres pudełkowy przewozów ładunków dla transportu kolejowego, samochodowego i morskiego	40
5.6.2. Wykres pudełkowy przewozów ładunków dla transportu morskiego	42
5.7. Wykres rozkładu dla przewozów ładunków w transporcie kolejowym w dwóch okresach	44
5.8. Wykres skrzypcowy dla przewozów pasażerów transportem samochodowym w dwóch okresach	46
5.9. Mapy cieplne	48
5.9.1. Mapa cieplna korelacji między rodzajami transportu w przewozie ładunków	48
5.9.2. Mapa cieplna korelacji między rodzajami transportu w przewozie pasażerów	50
6. Weryfikacja hipotez statystycznych.....	52
6.1. Testowanie różnicy średnich przewozów pasażerów między transportem samochodowym a kolejowym	52
6.2. Testowanie różnicy średnich przewozów ładunków w transporcie kolejowym w dwóch okresach	53
6.3. Testy korelacji między przewozem ładunków transportem kolejowym a samochodowym.....	54
6.3.1. Test korelacji Pearsona:	55
6.3.2. Test korelacji Spearmana:.....	56
6.3.3. Test korelacji Kendalla:	56
6.4. Test korelacji Pearsona między przewozem pasażerów transportem kolejowym a samochodowym.....	57
6.5. Test normalności rozkładu danych za pomocą testu Shapiro-Wilka dla przewozu ładunków różnymi rodzajami transportu	58

6.6. Test zgodności rozkładu danych z rozkładem normalnym, przy użyciu testu Jarque-Bera dla różnych rodzajów transportu pasażerskiego	60
6.7. Test chi-kwadrat w analizie proporcji między transportem kolejowym a samochodowym w dwóch okresach	62
6.8. Test chi-kwadrat w analizie proporcji przewozów transportem samochodowym w dwóch okresach	63
6.9. Test regresji liniowej między ilością przewozów ładunków transportem kolejowym, a ilością przewozów ładunków transportem samochodowym	64
6.10. Test Wilcoxona-Manna-Whitneya dla danych dotyczących przewozów pasażerów transportem lotniczym w dwóch okresach	66
6.10.1. Test normalności rozkładu danych transportu lotniczego	67
6.10.2. Test Wilcoxona-Manna-Whitneya.....	68
Literatura	70
Spis rysunków	71
Spis tabel	74

1. Wstęp

Rozwój infrastruktury transportowej odgrywa fundamentalną rolę w dynamicznym funkcjonowaniu gospodarki, wpływając na mobilność społeczeństwa oraz efektywność przewozu towarów. Polska, będąca strategicznym punktem na mapie Europy, doświadczyła znaczących zmian w strukturze transportu w okresie od 2010 do 2022 roku. Analiza tych zmian jest kluczowa dla zrozumienia ewolucji sektorów transportowych, zarówno pod kątem przewozu ładunków, jak i transportu pasażerskiego.

Sektor transportowy w Polsce składa się z kilku kluczowych elementów, w tym transportu drogowego, kolejowego, morskiego i lotniczego, z których każdy ma swoje specyficzne cechy, ograniczenia i znaczenie dla krajowej i międzynarodowej gospodarki. Dynamika wzrostu gospodarczego, zmiany technologiczne, regulacje prawne oraz preferencje społeczne stanowiły istotne czynniki kształtujące strukturę i rozwój tych sektorów.

W kontekście przewozów ładunków, analiza zmian w transporcie drogowym, kolejowym i morskim odzwierciedlała rosnącą potrzebę efektywnego przemieszczania towarów w kraju i międzynarodowo. Zmiany w handlu międzynarodowym, rozwój centrów logistycznych oraz modernizacja infrastruktury transportowej miały istotny wpływ na zmiany w strukturze przewozów ładunków.

Z drugiej strony, w przypadku przewozów pasażerskich, transport drogowy, kolejowy i lotniczy odgrywały kluczową rolę w zapewnieniu mobilności jednostek oraz w rozwoju turystyki i biznesu. Zmiany w preferencjach podróżujących, rozwój technologii w sektorze lotniczym, a także modernizacja sieci kolejowej i dróg, miały wpływ na rozwój tych sektorów.

Celem projektu jest przeprowadzenie analitycznej oceny struktury transportu w Polsce w okresie lat 2010-2022, z uwzględnieniem dwóch głównych kategorii: przewozów ładunków i przewozów pasażerów. Projekt ten skupi się na analizie zmian w poszczególnych sektorach transportu, takich jak transport drogowy, kolejowy, morski w kontekście przewozów ładunków, oraz transport drogowy, kolejowy i lotniczy w odniesieniu do przewozów pasażerów. Projekt ten pozwoli zidentyfikować, czy przewozy ładunków i pasażerów w Polsce zmieniły się w badanym okresie, czy też pozostały na stałym poziomie, a może nawet się zmniejszyły.

2. Opis użytych danych

W projekcie zostały użyte dane o przewozach ładunków i pasażerów w Polsce. Dane pochodzą z biuletynu statystycznego GUS: <https://stat.gov.pl>.

**TABL. 52. PRZEWOZY ŁADUNKÓW I PASAŻERÓW
TRANSPORT OF GOODS AND PASSENGERS**

Okresy Periods	1	2	3	4	5	6	7	8
	Przewozy ładunków Transport of goods				Przewozy pasażerów Transport of passengers			
	Ogółem Grand total	transport kolejowy railway transport	transport samochodowy road transport	transport morski maritime transport	Ogółem Grand total	transport kolejowy railway transport	transport samochodowy road transport	transport lotniczy air transport
	tysiące ton	thousand tonnes			tysiące	thousand		
2010 M01	27925	14021	8722	702	72454	21789	50388	267
2010 M02	29278	15311	8781	657	71860	21111	50494	248
2010 M03	34031	18213	10426	715	80043	22315	57381	338
2010 M04	32918	17080	10359	542	74245	22074	51907	245
2010 M05	33803	17400	10555	586	70563	22319	47726	460
2010 M06	34833	17969	10857	598	67701	21218	45840	533

Rysunek 1 Fragment danych pliku Excel

Dane w tabeli podane są w tysiącach ton.

Przewozy ładunków odnoszą się do sumy masy ładunków przewiezionych poszczególnymi rodzajami transportu, tj.

- Przewozy ładunków transportem kolejowym: dane te obejmują sumę masy brutto wszystkich przesyłek handlowych i służbowych (tj. łącznie z wagą opakowań, a w przypadku przewozów w kontenerach również łącznie z wagą kontenerów) przewiezionych w wagonach towarowych w komunikacji wewnętrznej (krajowej) i międzynarodowej wszystkimi rodzajami pociągów. Dane nie obejmują przewozów manewrowych.
- Przewozy ładunków transportem samochodowym: dane o przewozach ładunków transportem samochodowym dotyczą całego transportu samochodowego, tj. zarobkowego i gospodarczego w sektorze publicznym i prywatnym. Przez transport samochodowy zarobkowy rozumie się świadczenie usług przewozowych za opłatą, a przez transport samochodowy gospodarczy - wykonywanie przewozów ładunków na własne potrzeby, tj. bez opłaty (wliczanych w koszt przedsiębiorstwa). Dane częściowo szacunkowe.

- Przewozy ładunków morską flotą transportową: przewozy ładunków wykonane w rejsach zakończonych statkami własnymi i statkami dzierżawionymi pływającymi pod banderą polską i obcą. Obejmują przewozy ładunków wykonane we wszystkich relacjach, tj. przywożonych do portów polskich z portów obcych, wywożonych z portów polskich do portów obcych, przewożonych pomiędzy portami obcymi, przewożonych pomiędzy portami polskimi. Nie ujmuje się przewozów ładunków wykonanych statkami przekazanymi w czarter obcemu (zagranicznemu) armatorowi.

Przewozy pasażerów definiuje się jako suma pasażerów przewiezionych poszczególnymi rodzajami transportu, tj.

- Przewozy pasażerów transportem kolejowym: liczba wszystkich pasażerów przewiezionych na sieci kolei normalnotorowych i wąskotorowych pociągami osobowymi, pośpiesznymi i ekspresowymi (łącznie z pasażerami przewiezionymi w wagonach sypialnych i w wagonach z miejscami do leżenia) z biletami płatnymi oraz pasażerów korzystających z bezpłatnych przejazdów na podstawie odpowiednich przepisów w ramach komunikacji wewnętrznej i międzynarodowej.
- Przewozy pasażerów transportem samochodowym: przewozy pasażerów taborem autobusowym to suma pasażerów przewiezionych w komunikacji krajowej i międzynarodowej autobusami w ramach komunikacji regularnej (tj. na podstawie rozkładu jazdy), regularnej specjalnej (np. przewozy pracownicze, szkolne) oraz w przewozach pozostałych (np. wycieczkowe). Dane o przewozach pasażerów transportem samochodowym nie obejmują przewozów osób taborem autobusowym przedsiębiorstw komunikacji miejskiej oraz taborem autobusowym będącym w transporcie samochodowym gospodarczym. Dane nie obejmują przewozów dokonywanych przez podmioty o liczbie pracujących do 9 osób.
- Przewozy pasażerów transportem lotniczym: dane o przewozach pasażerów transportem lotniczym obejmują pasażerów przewiezionych w lotach rozkładowych, dodatkowych i wynajętych (czarterowych) samolotami polskich przewoźników transportu lotniczego rozkładowego (transport pasażerów na stałych trasach i według stałych rozkładów lotów) oraz innych polskich przewoźników świadczących zarobkowo tego rodzaju usługi w komunikacji międzynarodowej i w komunikacji krajowej.

3. Wczytanie danych do środowiska R

3.1. Instalowanie i ładowanie bibliotek

```
1 #instalacja i wczytanie bibliotek
2 install.packages("openxlsx")
3 install.packages("dplyr")
4 install.packages("plotly")
5 install.packages("plotrix")
6 install.packages("ggplot")
7 install.packages("stringr")
8 install.packages("reshape2")
9 install.packages("scales")
10 library(openxlsx)
11 library(dplyr)
12 library(plotly)
13 library(plotrix)
14 library(ggplot2)
15 library(stringr)
16 library(reshape2)
17 library(scales)
```

Rysunek 2 Kod instalacji i ładowania bibliotek

Opis wykorzystanych bibliotek:

- 1) **openxlsx**: Biblioteka do obsługi plików Excel. Umożliwia odczyt, zapis i manipulacje danymi w formacie xlsx.
- 2) **dplyr**: Narzędzie do manipulacji danymi w języku R. Posiada funkcje do wykonywania popularnych operacji takich jak filtrowanie, sortowanie, grupowanie, łączenie i wiele innych, ułatwiając pracę z danymi.
- 3) **plotly**: Biblioteka do tworzenia interaktywnych wykresów i wizualizacji danych. Pozwala na generowanie interaktywnych wykresów, które można przeglądać, zoomować czy też eksportować.
- 4) **plotrix**: Biblioteka zawiera funkcje do tworzenia różnych typów wykresów w R, takich jak wykresy kołowe, histogramy, wykresy słupkowe itp. Wspiera różne opcje dostosowania wykresów.
- 5) **ggplot2**: Narzędzie do tworzenia wykresów, które opiera się na "grammar of graphics" (gramatyka grafiki). Pozwala na tworzenie estetycznych i elastycznych wykresów poprzez definiowanie warstw, estetyki, skali itp.
- 6) **stringr**: Pakiet do obsługi operacji na ciągach znaków w R. Zapewnia wiele przydatnych funkcji do manipulacji i analizy tekstu.
- 7) **reshape2**: Biblioteka do przekształcania i restrukturyzacji danych w R. Pomaga w przekształcaniu danych między różnymi formatami, co ułatwia analizę.
- 8) **scales**: Pakiet do obsługi skal, etykiet i przekształceń wizualnych w R. Zapewnia funkcje ułatwiające dostosowywanie skal i etykiet w wykresach.

3.2. Wczytanie danych z Excela

Za pomocą funkcji `read.xlsx` został wczytany plik z danymi, dodatkowo zostało określone z której zakładki oraz od którego wiersza chcemy wczytać dane.

```
15 #wczytuje dane z excelu od 6 wiersza; przpisuje do zmiennej  
16 dane <- read.xlsx("C:/Users/olafe/Desktop/Projekt_inz/dane.xlsx",  
17           sheet = 1, startRow = 6)
```

Rysunek 3 Kod wczytywania pliku danych

3.3. Przygotowanie danych do dalszego przetwarzania

Aby wczytane dane były czytelne, tworzymy wektor z nazwami, które później przypisujemy do odpowiednich kolumn.

```
20 #tworze wektor z nazwami, ktore pozniej przypisze do odpowiednich kolumn  
21 kolumny <- c("Okres", "ogolem - przewozy ladunkow", "transport kolejowy - ladunki",  
22           "transport samochodowy - ladunki", "transport morski - ladunki",  
23           "ogolem - przewozy pasazerow", "transport kolejowy - pasazerowie",  
24           "transport samochodowy - pasazerowie", "transport lotniczy - pasazerowie")  
25 colnames(dane) <- kolumny
```

Rysunek 4 Kod przypisujący nazwy do kolumn

	Okres	ogolem - przewozy ladunkow	transport kolejowy - ladunki	transport samochodowy - ladunki	transport morski - ladunki	ogolem - przewozy pasazerow	transport kolejowy - pasazerowie
1	2010 M01	27925	14021	8722	702	72454	21
2	2010 M02	29278	15311	8781	657	71860	21
3	2010 M03	34031	18213	10426	715	80043	22
4	2010 M04	32918	17080	10359	542	74245	22
5	2010 M05	33803	17400	10555	586	70563	22
6	2010 M06	34833	17969	10857	598	67701	21
7	2010 M07	36590	18850	11350	602	52201	20
8	2010 M08	36846	19096	11274	746	52380	21
9	2010 M09	37051	19949	11396	725	73698	21
10	2010 M10	38975	20938	11813	815	76944	23
11	2010 M11	37408	20191	11420	712	74312	22
12	2010 M12	32807	16989	10205	697	70250	20
13	2011 M01	34822	19095	10754	546	70512	21
14	2011 M02	35152	18593	11289	572	67703	20
15	2011 M03	39815	21555	12492	672	74645	22
16	2011 M04	38324	20098	12479	538	70037	21
17	2011 M05	39652	21443	12801	766	72097	23
18	2011 M06	39406	21398	13099	645	64725	21
19	2011 M07	40105	21160	13622	508	49437	20
20	2011 M08	40733	21453	13837	614	51833	21
21	2011 M09	40975	21215	14321	702	70984	21
22	2011 M10	42257	21196	14627	678	73916	24
23	2011 M11	40413	20809	13824	693	71959	23

Rysunek 5 Fragment wczytanych danych

4. Wyznaczanie podstawowych parametrów opisowych

4.1. Wyznaczanie podstawowych parametrów dla przewozów ładunków

4.1.1. Wartość maksymalna

- Funkcja max() służy do znalezienia największej wartości w danym zestawie danych.

```
31 #wartosci max  
32 max(dane$`transport kolejowy - ladunki`)
```

Rysunek 6 Funkcja wyznaczająca wartość maksymalną

```
> #wartosci max  
> max(dane$`transport kolejowy - ladunki`)  
[1] 22505  
> max(dane$`transport samochodowy - ladunki`)  
[1] 27256  
> max(dane$`transport morski - ladunki`)  
[1] 825
```

Rysunek 7 Wyniki kodu dla wartości maksymalnej

- Największa ilość przewiezionych ładunków w transporcie kolejowym wyniosła 22505.
- Największa ilość przewiezionych ładunków w transporcie samochodowym wyniosła 27256.
- Największa ilość przewiezionych ładunków w transporcie morskim wyniosła 825.

Wniosek: W transporcie samochodowym odnotowano największą ilość przewiezionych ładunków w porównaniu do innych środków transportu.

4.1.2. Wartość minimalna

- Funkcja min() służy do znalezienia najmniejszej wartości w danym zestawie danych.

```
36 #wartosci min  
37 min(dane$`transport kolejowy - ladunki`)
```

Rysunek 8 Funkcja wyznaczająca wartość minimalną

```

> #wartosci min
> min(dane$`transport kolejowy - ładunki`)
[1] 14021
> min(dane$`transport samochodowy - ładunki`)
[1] 8722
> min(dane$`transport morski - ładunki`)
[1] 438

```

Rysunek 9 Wyniki kodu dla wartości minimalnej

- Minimalna ilość przewiezionych ładunków w transporcie kolejowym wyniosła 14021.
- Minimalna ilość przewiezionych ładunków w transporcie samochodowym wyniosła 8722.
- Minimalna ilość przewiezionych ładunków w transporcie morskim wyniosła 438.

Wniosek: Transport morski odnotował najmniejszą ilość przewiezionych ładunków w porównaniu do pozostałych środków transportu.

4.1.3. Odchylenie standardowe

- Funkcja sd() oblicza odchylenie standardowe zbioru danych, co pomaga zrozumieć, jak bardzo wartości rozpraszały się wokół średniej. Jest miarą zmienności danych.

```

41 #odchylenie standardowe
42 sd(dane$`transport kolejowy - ładunki`)

```

Rysunek 10 Funkcja wyznaczająca odchylenie standardowe

```

> #odchylenie standardowe
> sd(dane$`transport kolejowy - ładunki`)
[1] 1589.582
> sd(dane$`transport samochodowy - ładunki`)
[1] 5008.331
> sd(dane$`transport morski - ładunki`)
[1] 88.59275

```

Rysunek 11 Wyniki kodu dla odchylenia standardowego

- Odchylenie standardowe w transporcie kolejowym wynosiło około 1589,582.
- Odchylenie standardowe w transporcie samochodowym wynosiło około 5008,331.
- Odchylenie standardowe w transporcie morskim wynosiło około 88,59275.

Wniosek: Transport samochodowy miał największą zmienność w przewożonych ładunkach, wyrażoną przez odchylenie standardowe.

4.1.4. Średnia

- Funkcja mean() oblicza średnią arytmetyczną wartości w zestawie danych. Jest używana do określenia przeciętnej wartości.

```
46 #średnia  
47 mean(dane$`transport kolejowy - ładunki`)
```

Rysunek 12 Funkcja wyznaczająca średnią

```
> #średnia  
> mean(dane$`transport kolejowy - ładunki`)  
[1] 19352.16  
> mean(dane$`transport samochodowy - ładunki`)  
[1] 18639.48  
> mean(dane$`transport morski - ładunki`)  
[1] 626.7308
```

Rysunek 13 Wyniki kodu dla średniej

- Średnia ilość przewiezionych ładunków w transporcie kolejowym wyniosła około 19352,16.
- Średnia ilość przewiezionych ładunków w transporcie samochodowym wyniosła około 18639,48.
- Średnia ilość przewiezionych ładunków w transporcie morskim wyniosła około 626,7308.

Wniosek: Transport kolejowy odnotował najwyższą średnią ilość przewiezionych ładunków w porównaniu do innych środków transportu.

4.1.5. Mediana

- Funkcja median() oblicza medianę, która jest środkową wartością w posortowanym zestawie danych. Jest używana do opisania centralnej tendencji danych.

```
51 #medianę  
52 median(dane$`transport kolejowy - ładunki`)
```

Rysunek 14 Funkcja wyznaczająca medianę

```

> #median
> median(dane$`transport kolejowy - ładunki`)
[1] 19476.5
> median(dane$`transport samochodowy - ładunki`)
[1] 18221
> median(dane$`transport morski - ładunki`)
[1] 620.5

```

Rysunek 15 Wyniki kodu dla mediany

- Mediana ilości przewiezionych ładunków w transporcie kolejowym wynosiła 19476,5.
- Mediana ilości przewiezionych ładunków w transporcie samochodowym wynosiła 18221.
- Mediana ilości przewiezionych ładunków w transporcie morskim wynosiła 620,5.

Wniosek: Mediana ilości przewiezionych ładunków w transporcie kolejowym była wyższa niż w innych środkach transportu.

4.1.6. Kwartyle

- Funkcja quantile() oblicza kwantyle, w tym przypadku 75% (trzeci kwartyl) i 25% (pierwszy kwartyl) danych. Pomaga w analizie rozkładu wartości w danych.

```

56 #kwartyl 3/4
57 quantile(dane$`transport kolejowy - ładunki`,3/4)
65 #kwartyl 1/4
66 quantile(dane$`transport kolejowy - ładunki`,1/4)

```

Rysunek 16 Funkcja obliczająca kwartyle

```

> #kwartyl 3/4
> quantile(dane$`transport kolejowy - ładunki`,3/4)
    75%
20483.25
> quantile(dane$`transport samochodowy - ładunki`,3/4)
    75%
23064.25
> quantile(dane$`transport morski - ładunki`,3/4)
    75%
697.75
> #kwartyl 1/4
> quantile(dane$`transport kolejowy - ładunki`,1/4)
    25%
18476.75
> quantile(dane$`transport samochodowy - ładunki`,1/4)
    25%
14313.25
> quantile(dane$`transport morski - ładunki`,1/4)
    25%
563.25

```

Rysunek 17 Wyniki kodu dla kwartyle

Kwartyle 3/4:

- Trzeci kwartyl ilości przewiezionych ładunków w transporcie kolejowym wynosił 20483,25.
- Trzeci kwartyl ilości przewiezionych ładunków w transporcie samochodowym wynosił 23064,25.
- Trzeci kwartyl ilości przewiezionych ładunków w transporcie morskim wynosił 697,75.

Wniosek: Trzeci kwartyl w transporcie samochodowym miał najwyższą wartość, co oznacza, że 75% danych były niższe od tego poziomu.

Kwartyle 1/4:

- Pierwszy kwartyl ilości przewiezionych ładunków w transporcie kolejowym wynosił 18476,75.
- Pierwszy kwartyl ilości przewiezionych ładunków w transporcie samochodowym wynosił 14313,25.
- Pierwszy kwartyl ilości przewiezionych ładunków w transporcie morskim wynosił 563,25.

Wniosek: Pierwszy kwartyl w transporcie morskim miał najniższą wartość, co oznacza, że 25% danych były niższe od tego poziomu.

4.1.7. Rozstęp międzykwartylowy

- Funkcja IQR() oblicza rozstęp międzykwartylowy, który jest różnicą między trzecim a pierwszym kwartylem danych. Pomaga w analizie rozproszenia wartości w obszarze centralnym danych.

```
66 #rozstep_miedzykwartylowy  
67 IQR(dane$`transport kolejowy - ładunki`)
```

Rysunek 18 Funkcja obliczająca rozstęp międzykwartylowy

```

> #rozstęp międzykwartylowy
> IQR(dane$`transport kolejowy - ładunki`)
[1] 2006.5
> IQR(dane$`transport samochodowy - ładunki`)
[1] 8751
> IQR(dane$`transport morski - ładunki`)
[1] 134.5

```

Rysunek 19 Wyniki kodu dla rozstępu międzykwartylowego

- Rozstęp międzykwartylowy ilości przewiezionych ładunków w transporcie kolejowym wynosił 2006.5.
- Rozstęp międzykwartylowy ilości przewiezionych ładunków w transporcie samochodowym wynosił 8751.
- Rozstęp międzykwartylowy ilości przewiezionych ładunków w transporcie morskim wynosił 134.5.

Wniosek: Transport samochodowy miał największy rozstęp międzykwartylowy, co oznacza duży zakres zmienności w ilości przewiezionych ładunków.

4.1.8. Wariancja

- Funkcja var() oblicza wariancję próbłową danych, co jest miarą rozproszenia danych. Im większa wariancja, tym większa zmienność.

```

71 #wariancja próbkowa
72 var(dane$`transport kolejowy - ładunki`)

```

Rysunek 20 Funkcja obliczająca wariancję

```

> #wariancja próbkowa
> var(dane$`transport kolejowy - ładunki`)
[1] 2526772
> var(dane$`transport samochodowy - ładunki`)
[1] 25083382
> var(dane$`transport morski - ładunki`)
[1] 7848.675

```

Rysunek 21 Wyniki kodu dla wariancji

- Wariancja próbkowa ilości przewiezionych ładunków w transporcie kolejowym wynosiła około 2526772.
- Wariancja próbkowa ilości przewiezionych ładunków w transporcie samochodowym wynosiła około 25083382.

- Wariancja próbкова ilości przewiezionych ładunków w transporcie morskim wynosiła około 7848,675.

Wniosek: Transport samochodowy miał największą wariancję próbłową, co oznacza, że dane były najbardziej rozproszone.

4.1.9. Współczynnik korelacji

- Funkcja cor() oblicza współczynnik korelacji między dwoma zestawami danych. Pomaga w zrozumieniu zależności między nimi.

```
76 #współczynnik korelacji
77 cor(dane$`transport kolejowy - ładunki`, dane$`transport samochodowy - ładunki`)
```

Rysunek 22 Funkcja obliczająca współczynnik korelacji

```
> #współczynnik korelacji
> cor(dane$`transport kolejowy - ładunki`, dane$`transport samochodowy - ładunki`)
[1] 0.3111442
> cor(dane$`transport kolejowy - ładunki`, dane$`transport morski - ładunki`)
[1] 0.4754244
> cor(dane$`transport samochodowy - ładunki`, dane$`transport morski - ładunki`)
[1] 0.4244659
```

Rysunek 23 Wyniki dla współczynnika korelacji

- Współczynnik korelacji między ilościami przewiezionych ładunków w transporcie kolejowym a samochodowym wyniósł około 0,3111442.
- Współczynnik korelacji między ilościami przewiezionych ładunków w transporcie kolejowym a morskim wyniósł około 0,4754244.
- Współczynnik korelacji między ilościami przewiezionych ładunków w transporcie samochodowym a morskim wyniósł około 0,4244659.

Wniosek: Istnieją pewne korelacje między ilościami przewiezionych ładunków w różnych środkach transportu, ale są one umiarkowane.

4.1.10. Kowariancja

- Funkcja cov() oblicza kowariancję między dwoma zestawami danych, co jest miarą współzmienności między nimi.

```
82 #kowariancja  
83 cov(dane$`transport kolejowy - ładunki`, dane$`transport samochodowy - ładunki`)
```

Rysunek 24 Funkcja obliczająca kowariancję

```
> #kowariancja  
> cov(dane$`transport kolejowy - ładunki`, dane$`transport samochodowy - ładunki`)  
[1] 2477067  
> cov(dane$`transport kolejowy - ładunki`, dane$`transport morski - ładunki`)  
[1] 66951.86  
> cov(dane$`transport samochodowy - ładunki`, dane$`transport morski - ładunki`)  
[1] 188336.3
```

Rysunek 25 Wyniki kodu dla kowariancji

- Kowariancja między transportem kolejowym a samochodowym w przewozie ładunków: kowariancja wynosząca około 2477067 sugeruje, że przewozy ładunków transportem kolejowym i samochodowym wykazywały pewien poziom współzmienności. Oznacza to, że wzrost liczby przewiezionych ładunków transportem kolejowym często wiązał się ze wzrostem liczby przewiezionych ładunków transportem samochodowym.
- Kowariancja między transportem kolejowym a morskim w przewozie ładunków: kowariancja wynosząca około 66951,86 sugeruje, że przewozy ładunków transportem kolejowym i morskim wykazywały pewien poziom współzmienności. Wzrost liczby przewiezionych ładunków transportem kolejowym często wiązał się ze wzrostem liczby przewiezionych ładunków transportem morskim.
- Kowariancja między transportem samochodowym a morskim w przewozie ładunków: kowariancja wynosząca około 188336,3 jest dodatnia i sugeruje, że przewozy ładunków transportem samochodowym i morskim wykazywały pewien poziom współzmienności. Wzrost liczby przewiezionych ładunków transportem samochodowym często wiązał się ze wzrostem liczby przewiezionych ładunków transportem morskim.

Wniosek: Kowariancja między ilościami przewiezionych ładunków sugeruje, że zmienne te są w pewien sposób współzmienne.

4.2. Wyznaczanie podstawowych parametrów dla przewozów pasażerów

4.2.1. Wartość maksymalna

```
> #wartosci max  
> max(dane$`transport kolejowy - pasazerowie`)  
[1] 32695  
> max(dane$`transport samochodowy - pasazerowie`)  
[1] 57381  
> max(dane$`transport lotniczy - pasazerowie`)  
[1] 3487
```

Rysunek 26 Wyniki kodu dla wartości maksymalnej

- Największa liczba pasażerów przewiezionych transportem kolejowym wyniosła 32695.
- Największa liczba pasażerów przewiezionych transportem samochodowym wyniosła 57381.
- Największa liczba pasażerów przewiezionych transportem lotniczym wyniosła 3487.

Wniosek: Transport samochodowy miał największą liczbę przewiezionych pasażerów w porównaniu do innych środków transportu.

4.2.2. Wartość minimalna

```
> #wartosci min  
> min(dane$`transport kolejowy - pasazerowie`)  
[1] 6097  
> min(dane$`transport samochodowy - pasazerowie`)  
[1] 4986  
> min(dane$`transport lotniczy - pasazerowie`)  
[1] 7
```

Rysunek 27 Wyniki kodu dla wartości minimalnej

- Minimalna liczba pasażerów przewiezionych transportem kolejowym wyniosła 6097.
- Minimalna liczba pasażerów przewiezionych transportem samochodowym wyniosła 4986.
- Minimalna liczba pasażerów przewiezionych transportem lotniczym wyniosła 7.

Wniosek: Transport lotniczy odnotował najmniejszą liczbę przewiezionych pasażerów w porównaniu do pozostałych środków transportu.

4.2.3. Odchylenie standardowe

```
> #odchylenie standardowe  
> sd(dane$`transport kolejowy - pasazerowie`)  
[1] 3896.608  
> sd(dane$`transport samochodowy - pasazerowie`)  
[1] 12925.29  
> sd(dane$`transport lotniczy - pasazerowie`)  
[1] 646.6259
```

Rysunek 28 Wyniki kodu dla odchylenia standardowego

- Odchylenie standardowe w transporcie kolejowym wynosiło około 3896,608.
- Odchylenie standardowe w transporcie samochodowym wynosiło około 12925,29.
- Odchylenie standardowe w transporcie lotniczym wynosiło około 646,6259.

Wniosek: Transport samochodowy miał największą zmienność w przewożonych pasażerach, wyrażoną przez odchylenie standardowe.

4.2.4. Średnia

```
> #średnia  
> mean(dane$`transport kolejowy - pasazerowie`)  
[1] 23392.34  
> mean(dane$`transport samochodowy - pasazerowie`)  
[1] 29180.1  
> mean(dane$`transport lotniczy - pasazerowie`)  
[1] 900.7115
```

Rysunek 29 Wyniki kodu dla średniej

- Średnia liczba pasażerów przewiezionych transportem kolejowym wyniosła około 23392,34.
- Średnia liczba pasażerów przewiezionych transportem samochodowym wyniosła około 29180,1.
- Średnia liczba pasażerów przewiezionych transportem lotniczym wyniosła około 900,7115.

Wniosek: Transport samochodowy odnotował najwyższą średnią liczbę przewiezionych pasażerów w porównaniu do innych środków transportu.

4.2.5. Mediana

```
> #mediana  
> median(dane$`transport kolejowy - pasazerowie`)  
[1] 23261.5  
> median(dane$`transport samochodowy - pasazerowie`)  
[1] 28999.5  
> median(dane$`transport lotniczy - pasazerowie`)  
[1] 707
```

Rysunek 30 Wyniki kodu dla mediany

- Mediana liczby pasażerów przewiezionych transportem kolejowym wynosiła 23261,5.
- Mediana liczby pasażerów przewiezionych transportem samochodowym wynosiła 28999,5.
- Mediana liczby pasażerów przewiezionych transportem lotniczym wynosiła 707.

Wniosek: Mediana liczby pasażerów przewiezionych transportem samochodowym była wyższa niż w innych środkach transportu.

4.2.6. Kwartyle

```
> #kwartył 3/4  
> quantile(dane$`transport kolejowy - pasazerowie`,3/4)  
75%  
25338.75  
> quantile(dane$`transport samochodowy - pasazerowie`,3/4)  
75%  
39969  
> quantile(dane$`transport lotniczy - pasazerowie`,3/4)  
75%  
1108.5  
> #kwartył 1/4  
> quantile(dane$`transport kolejowy - pasazerowie`,1/4)  
25%  
21783.5  
> quantile(dane$`transport samochodowy - pasazerowie`,1/4)  
25%  
17943.25  
> quantile(dane$`transport lotniczy - pasazerowie`,1/4)  
25%  
501.75
```

Rysunek 31 Wyniki kodu dla kwartylów

Kwartyle 3/4:

- Trzeci kwartyl liczby pasażerów przewiezionych transportem kolejowym wynosił 25338,75.
- Trzeci kwartyl liczby pasażerów przewiezionych transportem samochodowym wynosił 39969.
- Trzeci kwartyl liczby pasażerów przewiezionych transportem lotniczym wynosił 1108,5.

Wniosek: Trzeci kwartyl w transporcie samochodowym miał najwyższą wartość, co oznacza, że 75% danych były niższe od tego poziomu.

Kwartyle 1/4:

- Pierwszy kwartyl liczby pasażerów przewiezionych transportem kolejowym wynosił 21783,5.
- Pierwszy kwartyl liczby pasażerów przewiezionych transportem samochodowym wynosił 17943,25.
- Pierwszy kwartyl liczby pasażerów przewiezionych transportem lotniczym wynosił 501,75.

Wniosek: Pierwszy kwartyl w transporcie morskim miał najniższą wartość, co oznacza, że 25% danych były niższe od tego poziomu.

4.2.7. Rozstęp międzykwartylowy

```
> #rozstęp międzykwartylowy
> IQR(dane$`transport kolejowy - pasazerowie`)
[1] 3555.25
> IQR(dane$`transport samochodowy - pasazerowie`)
[1] 22025.75
> IQR(dane$`transport lotniczy - pasazerowie`)
[1] 606.75
```

Rysunek 32 Wyniki kodu dla rozstępu międzykwartylowego

- Rozstęp międzykwartylowy liczby pasażerów przewiezionych transportem kolejowym wynosił 3555,25.
- Rozstęp międzykwartylowy liczby pasażerów przewiezionych transportem samochodowym wynosił 22025,75.

- Rozstęp międzykwartylowy liczby pasażerów przewiezionych transportem lotniczym wynosił 606,75.

Wniosek: Transport samochodowy miał największy rozstęp międzykwartylowy, co oznacza duży zakres zmienności w liczbie przewiezionych pasażerów.

4.2.8. Wariancja

```
> #wariancja próbкова
> var(dane$`transport kolejowy - pasazerowie`)
[1] 15183554
> var(dane$`transport samochodowy - pasazerowie`)
[1] 167063170
> var(dane$`transport lotniczy - pasazerowie`)
[1] 418125
```

Rysunek 33 Wyniki kodu dla wariancji

- Wariancja próbкова liczby pasażerów przewiezionych transportem kolejowym wynosiła około 15183554.
- Wariancja próbкова liczby pasażerów przewiezionych transportem samochodowym wynosiła około 167063170.
- Wariancja próbкова liczby pasażerów przewiezionych transportem lotniczym wynosiła około 418125.

Wniosek: Transport samochodowy miał największą wariancję próbłową, co oznacza, że dane były najbardziej rozproszone.

4.2.9. Współczynnik korelacji

```
> #współczynnik korelacji
> cor(dane$`transport kolejowy - pasazerowie`, dane$`transport samochodowy - pasazerowie`)
[1] 0.03675242
> cor(dane$`transport kolejowy - pasazerowie`, dane$`transport lotniczy - pasazerowie`)
[1] 0.7259769
> cor(dane$`transport samochodowy - pasazerowie`, dane$`transport lotniczy - pasazerowie`)
[1] -0.3803805
```

Rysunek 34 Wyniki kodu dla współczynnika korelacji

- Współczynnik korelacji między liczbą pasażerów przewiezionych transportem kolejowym a samochodowym wyniósł około 0,03675242.
- Współczynnik korelacji między liczbą pasażerów przewiezionych transportem kolejowym a lotniczym wyniósł około 0,7259769.

- Współczynnik korelacji między liczbą pasażerów przewiezionych transportem samochodowym a lotniczym wyniósł około -0,3803805.

Wniosek: Istnieją pewne korelacje między liczbą przewiezionych pasażerów różnymi środkami transportu, przy czym transport kolejowy i lotniczy wykazują silniejszą korelację, podczas gdy transport samochodowy i lotniczy wykazują negatywną korelację.

4.2.10. Kowariancja

```
> #kowariancja
> cov(dane$`transport kolejowy - pasazerowie`,dane$`transport samochodowy - pasazerowie`)
[1] 1851028
> cov(dane$`transport kolejowy - pasazerowie`,dane$`transport lotniczy - pasazerowie`)
[1] 1829206
> cov(dane$`transport samochodowy - pasazerowie`,dane$`transport lotniczy - pasazerowie`)
[1] -3179155
```

Rysunek 35 Wyniki kodu dla kowariancji

- Kowariancja między transportem kolejowym a samochodowym: kowariancja wynosząca około 1851028 sugeruje, że przewozy pasażerów transportem kolejowym i samochodowym wykazywały pewien poziom współzmienności. Oznacza to, że wzrost liczby pasażerów przewiezionych transportem kolejowym często wiązał się ze wzrostem liczby pasażerów przewiezionych transportem samochodowym.
- Kowariancja między transportem kolejowym a lotniczym: kowariancja wynosząca około 1829206 sugeruje, że przewozy pasażerów transportem kolejowym i lotniczym również wykazywały pewien poziom współzmienności. Wzrost liczby pasażerów przewiezionych transportem kolejowym często wiązał się ze wzrostem liczby pasażerów przewiezionych transportem lotniczym.
- Kowariancja między transportem samochodowym a lotniczym: kowariancja wynosząca około -3179155 jest ujemna i sugeruje, że przewozy pasażerów transportem samochodowym i lotniczym miały tendencję do zmiany w przeciwnych kierunkach. Wzrost liczby pasażerów w jednym średku transportu często wiązał się ze spadkiem liczby pasażerów w drugim średku.

Wniosek: Kowariancja sugeruje, że zmienne te są w pewien sposób współzmienne.

4.3. Podsumowanie podstawowych parametrów opisowych

Parametr	Cecha	Przewozy ładunków			Przewozy pasażerów		
		transport kolejowy	transport samochodowy	transport morski	transport kolejowy	transport samochodowy	transport lotniczy
Wartość maksymalna		22505	27256	825	32695	57381	3487
Wartość minimalna		14021	8722	438	6097	4986	7
Odchylenie standardowe		1589,582	5008,331	88,59275	3896,608	12925,29	646,6259
Średnia		19352,16	18639,48	626,7308	23392,34	29180,1	900,7115
Mediana		19476,5	18221	620,5	23261,5	28999,5	707
Kwartyl 3/4		20483,25	23064,25	697,75	25338,75	39969	1108,5
Kwartyl 1/4		18476,75	14313,25	563,25	21783,5	17943,25	501,75
Rozstęp międzykwartylowy		2006,5	8751	134,5	3555,25	22025,75	606,75
Wariancja próbkowa		2526772	25083382	7848,675	15183554	167063170	418125

Tabela 1 Podsumowanie podstawowych parametrów opisowych cz.1

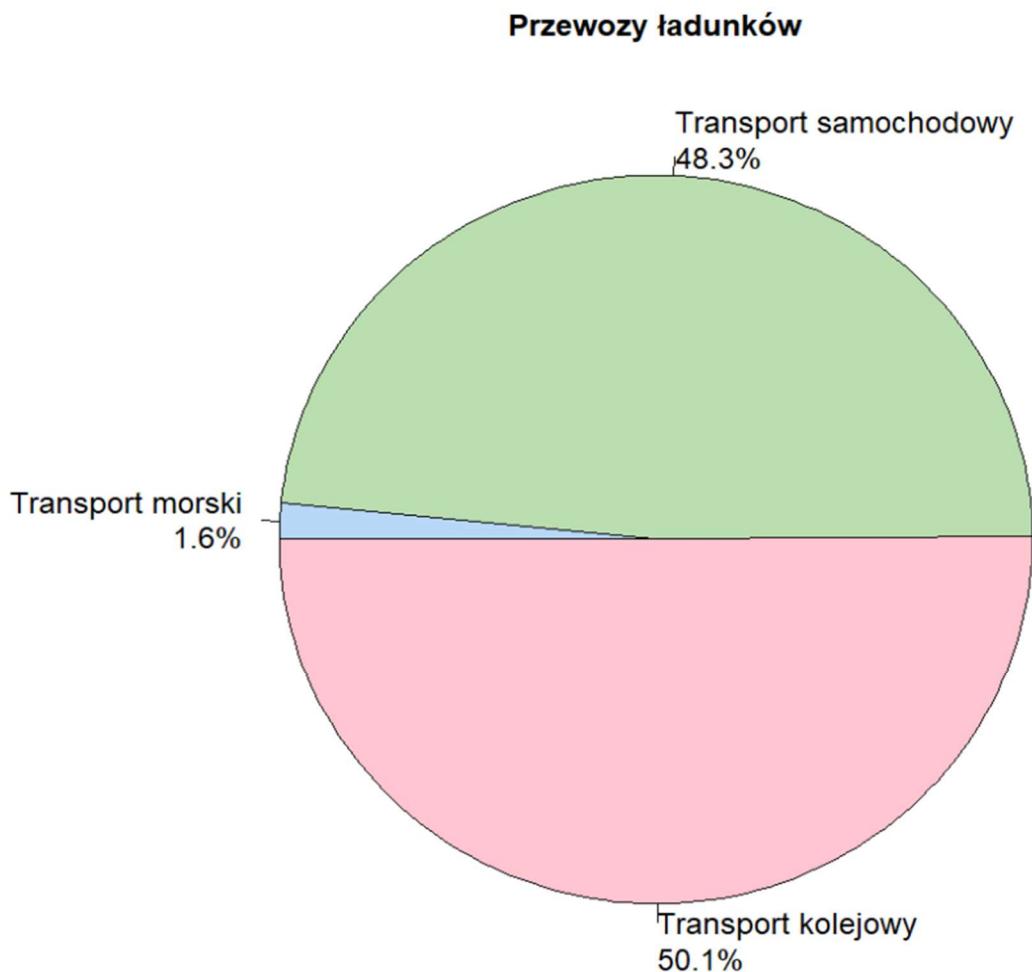
	Para cech	Współczynnik korelacji	Kowariancja
Przewozy ładunków	Transport kolejowy a samochodowy	0,3111442	2477067
	Transport kolejowy a morski	0,4754244	66951,86
	Transport samochodowy a morski	0,4244659	188336,3
Przewozy pasażerów	Transport kolejowy a samochodowy	0,03675242	1851028
	Transport kolejowy a lotniczy	0,7259769	1829206
	Transport samochodowy a lotniczy	-0,3803805	-3179155

Tabela 2 Podsumowanie podstawowych parametrów opisowych cz.2

5. Graficzna prezentacja danych

5.1. Wykresy kołowe

5.1.1. Wykres kołowy prezentujący udział rodzajów transportu w przewozach ładunków



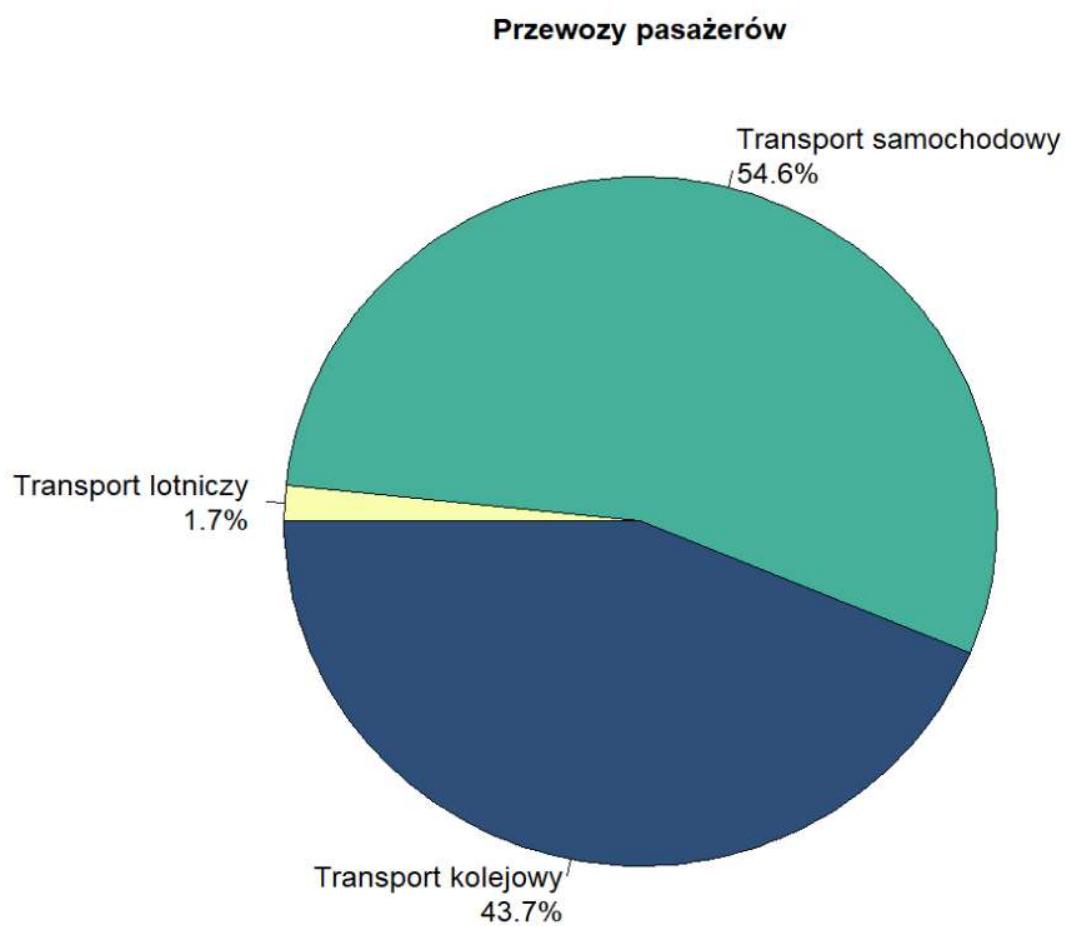
Rysunek 36 Wykres kołowy przewozów ładunków

Wykres pokazuje, że transport kolejowy reprezentuje główną siłę w przewozach ładunków z udziałem około 50,1%, następnie transport samochodowy z około 48,3%, podczas gdy udział transportu morskiego wynosi zaledwie 1,6%.

```
195 # Tworzenie wykresu kołowego
196 pie(x, labels = paste0(etykiety_ladunkow, "\n", round(100 * x / sum(x), 1), "%"),
197 col = hcl.colors(length(x), palette = "Pastel1"),
198 main = "Przewozy ładunków",
199 radius = -1,
200 cex = 1.2) # Zwiększa rozmiar etykiet
```

Rysunek 37 Kod generujący wykres kołowy

5.1.2. Wykres kołowy prezentujący udział rodzajów transportu w przewozach pasażerów

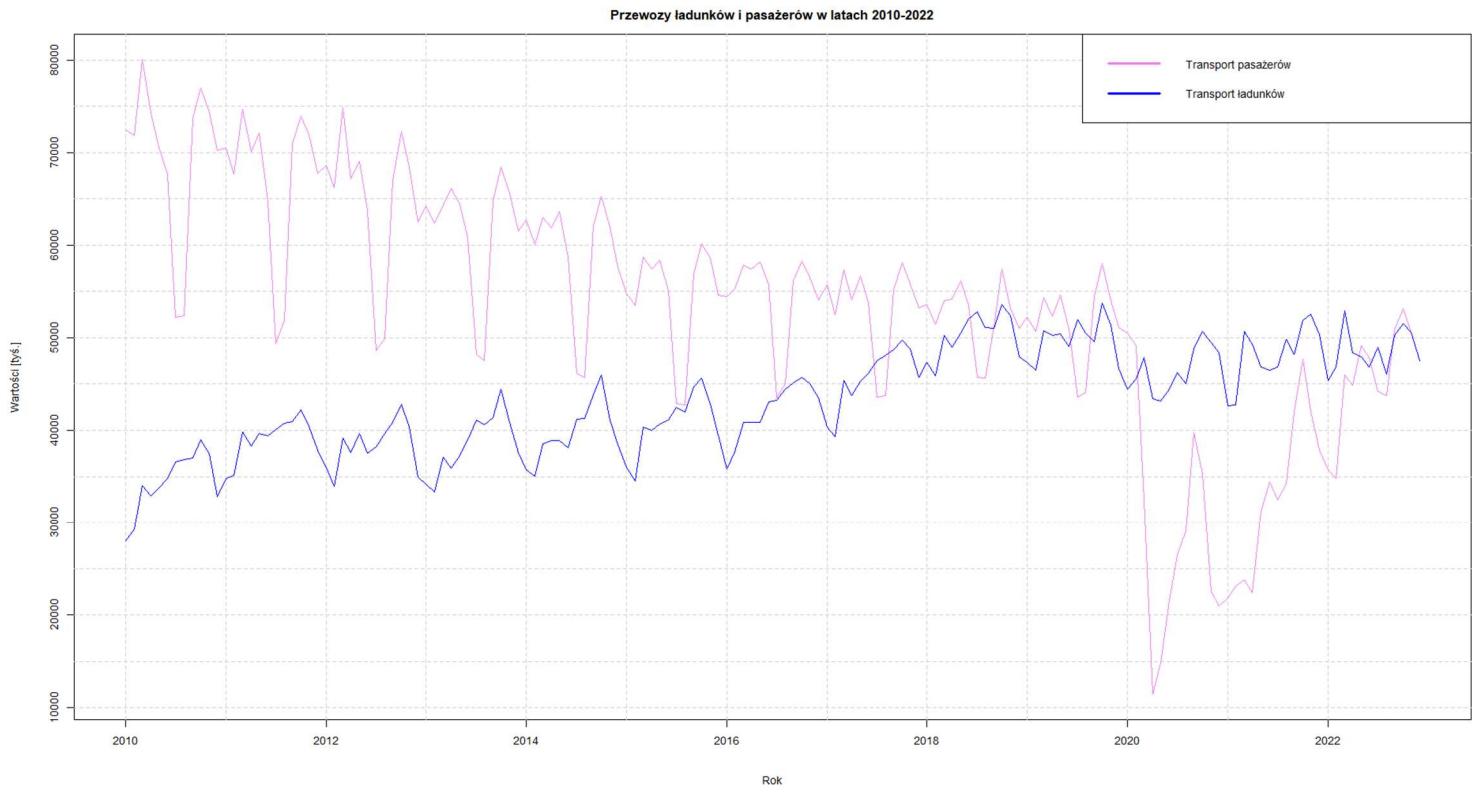


Rysunek 38 Wykres kołowy przewozów pasażerów

Wykres ukazuje, że transport samochodowy dominuje w przewozach pasażerów, reprezentując około 54,6% udziału. Transport kolejowy jest kolejnym istotnym uczestnikiem z udziałem około 43,7%, natomiast udział transportu lotniczego jest znacznie mniejszy, wynoszący około 1,7%.

5.2. Wykresy liniowe

5.2.1. Wykres liniowy prezentujący przewozy ładunków i pasażerów na przestrzeni lat



Rysunek 39 Wykres liniowy przewozów ładunków i pasażerów

Analiza danych dotyczących przewozu pasażerów wskazuje na tendencję spadkową w ilości przewożonych osób z roku na rok w pewnych okresach. Choć nie ma wyraźnego ciągłego trendu spadkowego przez cały okres analizy, to obserwuje się pewne lata, w których ilość przewożonych pasażerów zmniejszała się w porównaniu z poprzednimi latami. Zwłaszcza w ostatnich latach zanotowano spadki w przewozach, sugerujące tendencję malejącą w ilości przewożonych osób. Ta niestabilność i tendencja spadkowa mogą wynikać z różnych czynników, takich jak zmieniające się preferencje podróżujących, zmiany w warunkach ekonomicznych czy też wpływ zdarzeń zewnętrznych na sektor transportu, które mogły wpływać na zmniejszenie ruchu pasażerskiego.

Przewozy ładunków w badanym okresie wykazują ogólny trend wzrostowy. Zauważalne są regularne zmiany miesięczne, gdzie niektóre miesiące odznaczają się większymi lub mniejszymi ilościami przewożonych ładunków. Widoczne są okresowe skoki wartości, co może być rezultatem sezonowych trendów, świąt lub wyjątkowych zdarzeń gospodarczych. Pomimo tych wahań, istnieje długoterminowy wzrost w przewozach towarów, co może świadczyć o stabilności lub wzrastającej aktywności w sektorze transportu. Sezonowość również odgrywa istotną rolę, sugerując pewne miesięczne wzorce w

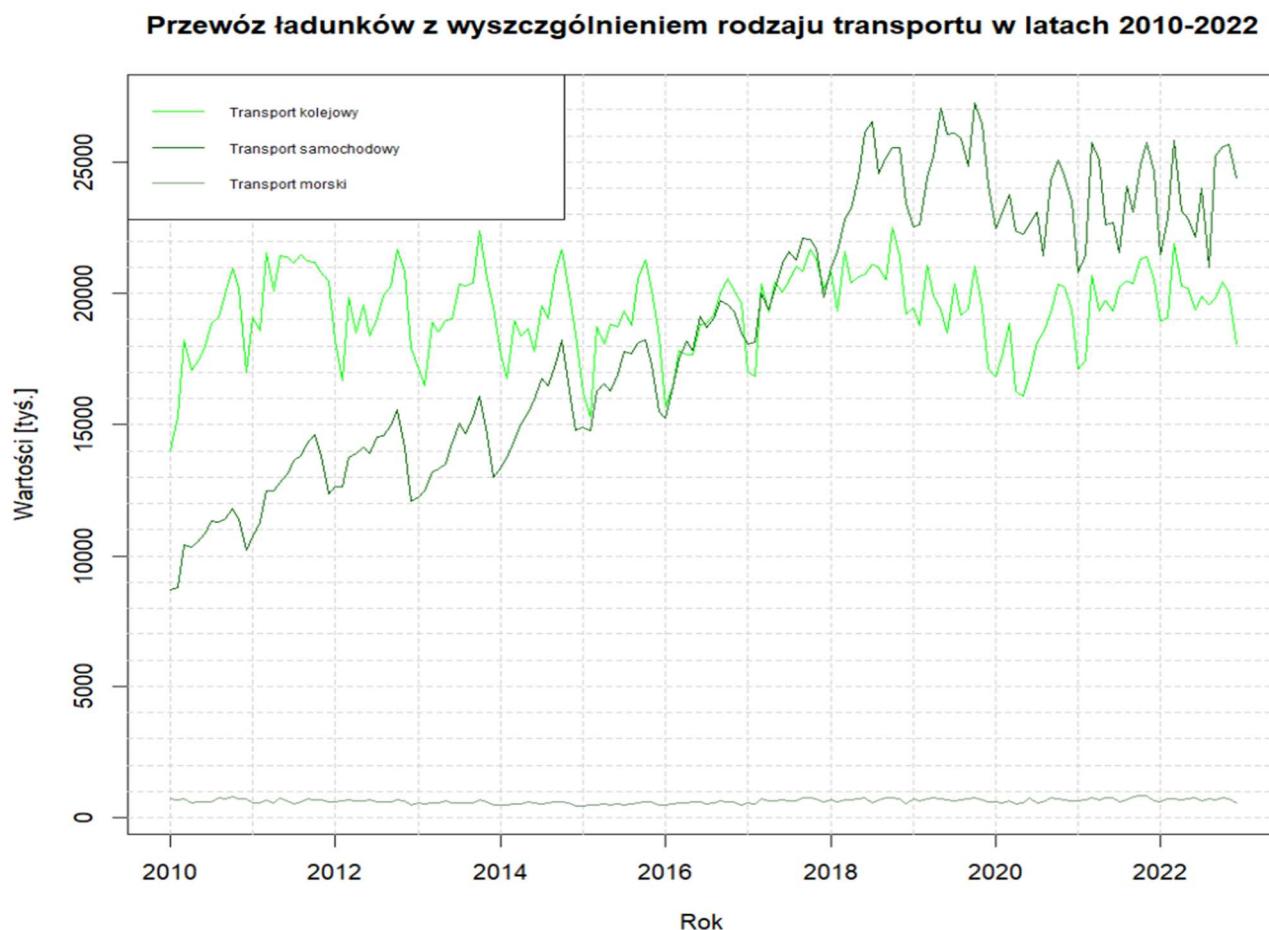
przewozach, które mogą być spowodowane cyklicznością w potrzebach przemysłowych lub konsumenckich. Jednak ogólnie rzecz biorąc, trend ten wykazuje tendencję do wzrostu w dłuższym okresie czasu.

Kod generujący wykresy liniowe dla określonych danych:

```
218 # Tworzenie wykresu liniowego
219 plot(tr_pasazerow,
220       type = "n",
221       main = "Przewozy ładunków i pasażerów w latach 2010-2022",
222       xlab = "Rok",
223       ylab = "Wartości [tyś.]",
224       col = "violet",
225       ylim = c(min(tr_ladunkow, tr_pasazerow), max(tr_ladunkow, tr_pasazerow)))
226
227 # Dodawanie linii poziomych w tle
228 abline(h = seq(0, max(tr_pasazerow), by = 5000), col = "lightgray", lty = 2)
229
230 # Dodawanie linii pionowych w tle
231 abline(v = seq(2010, 2022, by = 1), col = "lightgray", lty = 2)
232
233 # Rysowanie danych
234 lines(tr_pasazerow, col = "violet")
235 lines(tr_ladunkow, col = "blue")
236
237 legend(x = "topright", legend = c("Transport pasażerów", "Transport ładunków"),
238         col = c("violet", "blue"), lty = 1, cex = 1, lwd = 3)
239
```

Rysunek 40 Kod generujący wykres liniowy

5.2.2. Wykres liniowy przedstawiający przewóz ładunków na przestrzeni lat z podziałem na rodzaj transportu



Rysunek 41 Wykres liniowy przewozów ładunków z podziałem na rodzaj transportu

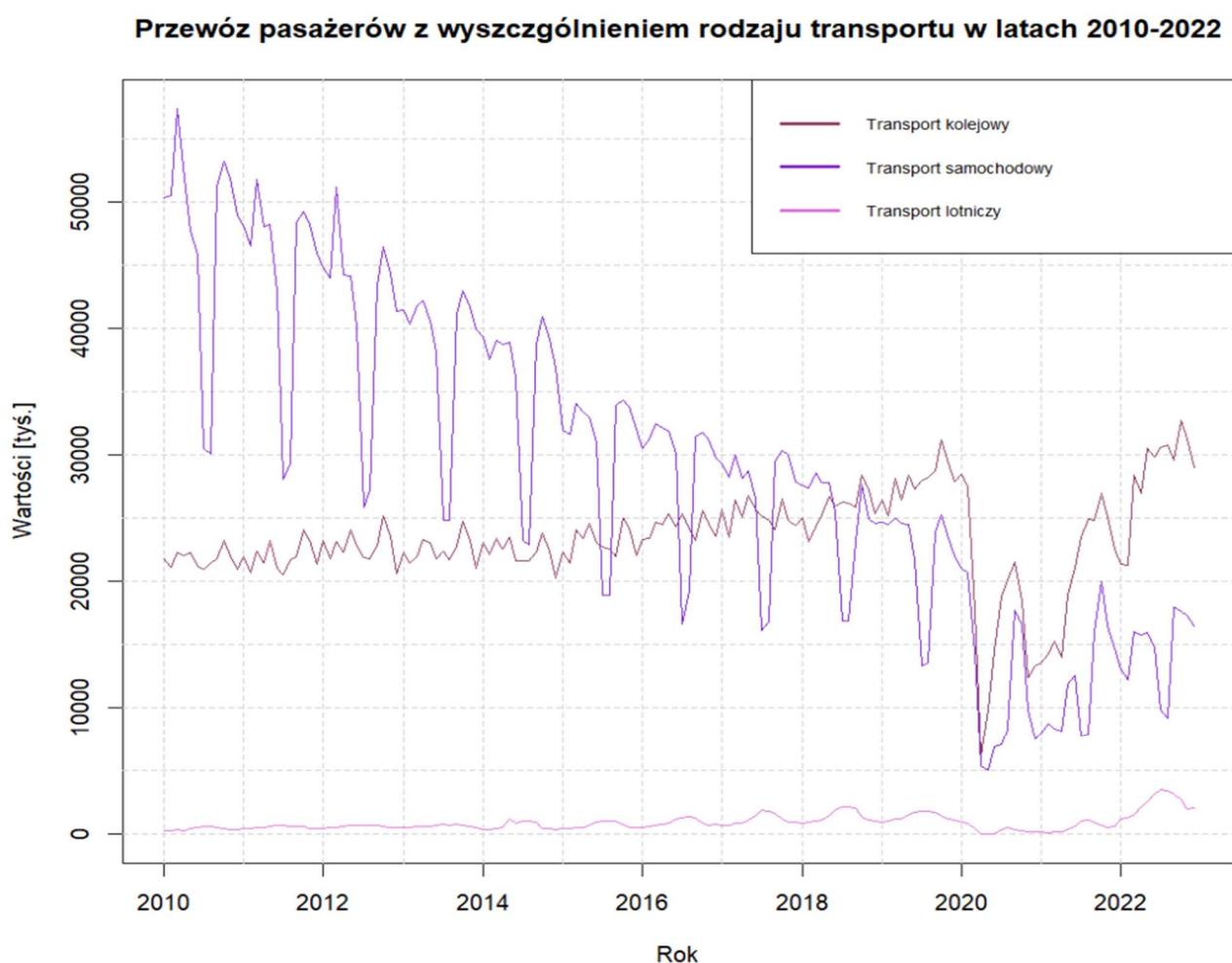
Analizując transport kolejowy zauważamy, że istnieje dość znacząca zmienność w przewozach. Obserwujemy pewne okresy wzrostu, poprzedzone często okresami spadków. Możemy zauważać, że w okolicach pewnych lat występowały wyraźne zmiany trendów, np. w okolicach lat 2010-2011, gdzie nastąpił wyraźny wzrost. Następnie w latach 2014-2015 było wyraźne osłabienie, a potem okres stabilności. Warto zwrócić uwagę na okresy wyższego wzrostu w latach 2017-2018 oraz późniejsze w 2021 roku, co może wskazywać na dynamiczne zmiany w sektorze transportu kolejowego.

Transport samochodowy ładunków w latach 2010-2022 wykazuje tendencję wzrostową z okresowymi zmianami w tempie przewozów. Pomimo krótkoterminowych fluktuacji i nieregularności w danych miesięcznych, widoczny jest ogólny trend wzrostowy. W niektórych okresach odnotowywano wyraźne skoki w przewozach, ale ogólnie linia trendu ukazuje stabilny, choć niejednolity rozwój. Możliwe jest również zauważenie okresów, w

których tempo wzrostu było niższe lub stabilizowało się, zanim ponownie nastąpił dynamiczniejszy wzrost przewozów.

Jeśli spojrzymy na skoki wartości w przewozach morskich, to te różnice nie są zazwyczaj dużymi liczbami. Widoczne wahania wskazują na pewne nieregularności, ale w większości przypadków różnice między wartościami są stosunkowo niewielkie. Można by uznać, że ogólny poziom przewozów w transporcie morskim utrzymuje się na względnie stabilnym poziomie, z pewnymi okresami wzrostów i spadków, ale bez znaczących skoków w dłuższej perspektywie czasowej.

5.2.3. Wykres liniowy przedstawiający przewóz pasażerów na przestrzeni lat z podziałem na rodzaj transportu



Rysunek 42 Wykres liniowy przewozów pasażerów z podziałem na rodzaj transportu

Transport kolejowy wykazywał tendencję do stabilności lub umiarkowanego wzrostu w okresie od 2010 do 2020 roku. Niemniej jednak, w 2020 roku zaobserwowano znaczący spadek liczby podróżnych (co jest zrozumiałe w kontekście globalnej pandemii COVID-19, która znaczco ograniczyła mobilność i funkcjonowanie sektora transportowego). Wzrost

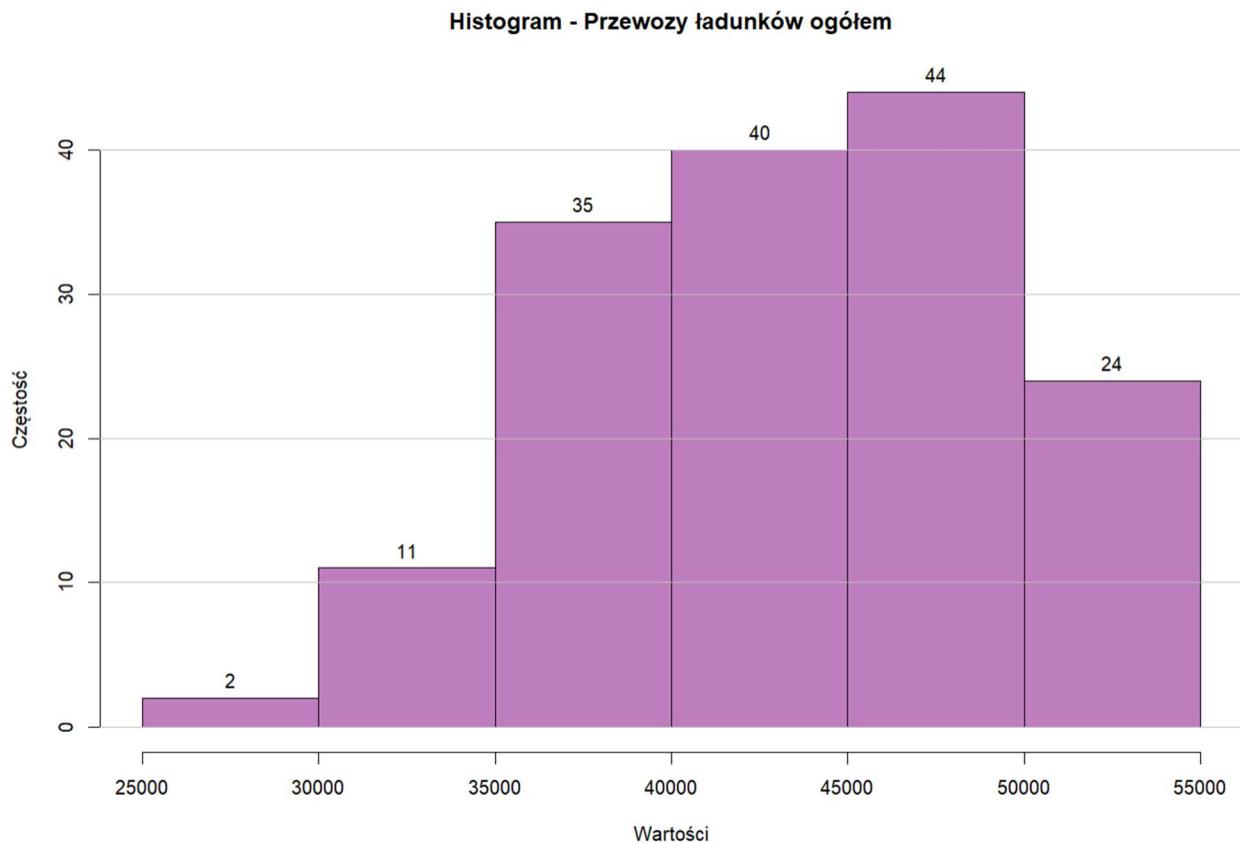
zanotowany w 2022 roku prawdopodobnie odzwierciedla stopniowe ożywienie sektora kolejowego.

Analiza danych z lat 2010-2022 wskazuje na spadek przewozów pasażerów transportem samochodowym w perspektywie długoterminowej. Obserwuje się tendencję malejącą, szczególnie wyraźną w 2020 r. Pandemia COVID-19 miała istotny wpływ na sektor transportu, co zaowocowało znacznym spadkiem liczby przewożonych osób tym środkiem transportu. Mimo pewnego wzrostu od 2020 do 2022 roku, dane wskazują na ogólny trend malejący w przewozach pasażerów transportem samochodowym.

Analiza danych dotyczących przewozów lotniczych w latach 2010-2022 wskazuje na zróżnicowane tendencje w transporcie lotniczym. Obserwuje się okresy zarówno wzrostu, jak i spadku liczby przewożonych pasażerów. Przed 2020 rokiem notowano tendencję wzrostową, co jest widoczne w coraz wyższych liczbach przewożonych osób rok po roku. Jednakże od roku 2020, pod wpływem pandemii COVID-19, transport lotniczy wykazywał znaczny spadek. Stopniowe odzyskiwanie ruchu lotniczego po roku 2020 jest widoczne w danych z lat kolejnych, gdzie liczba przewożonych pasażerów lotniczych stopniowo rośnie, co może być sygnałem stopniowego powrotu do normalności w sektorze transportu lotniczego.

5.3. Histogramy

5.3.1. Histogram przewozów ładunków



Rysunek 43 Histogram dla przewozów ładunków

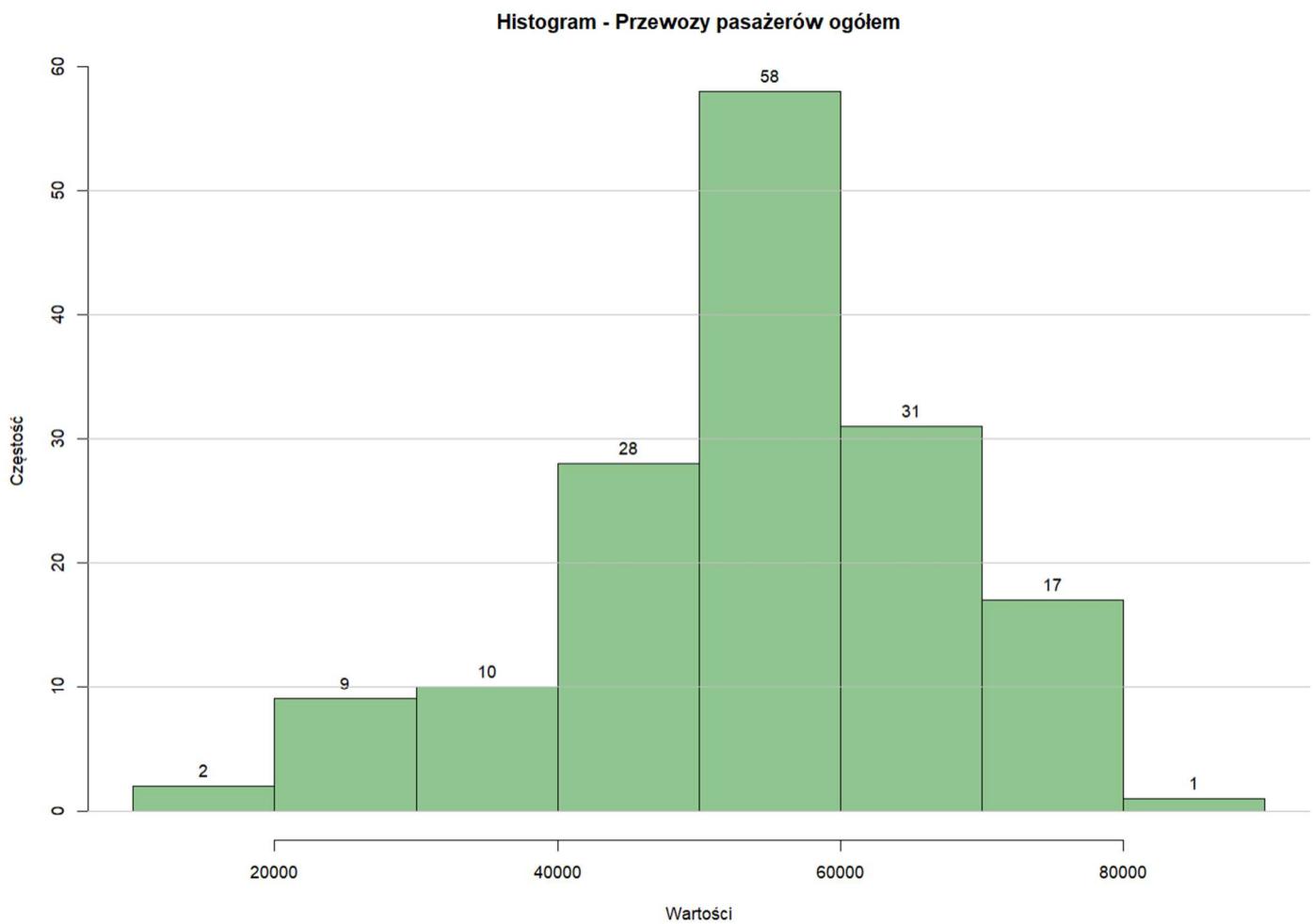
Analizując histogram, można stwierdzić, że większość danych przewozów ładunków ogółem zawiera się w przedziale między 40000 a 50000. Skrajne wartości (poniżej 35000 i powyżej 55000) występują znacznie rzadziej. To wskazuje na pewne zróżnicowanie w wartościach przewozów, ale też koncentrację w określonym przedziale wartości.

Poniżej przedstawiony został kod, który generuje histogramy:

```
318 ##### histogram
319
320 # Tworzenie histogramu dla przewozów ładunków ogółem
321 h <- hist(dane$ogółem - przewozy ładunków,
322           col = rgb(0.5, 0.0, 0.5, alpha = 0.5), # darkmagenta z półprzezroczystością
323           main = "Histogram - Przewozy ładunków ogółem",
324           xlab = "Wartości", ylab = "Częstość")
325
326 # Dodanie etykiet liczności na szczytach słupków
327 text(h$mid, h$counts, labels = h$counts, adj = c(0.5, -0.7))
328
329 # Dodanie ciemnych linii poziomych
330 abline(h = seq(0, max(h$counts), by = 10), col = "gray", lwd = 0.5)
```

Rysunek 44 Kod generujący histogram

5.3.2. Histogram przewozów pasażerów



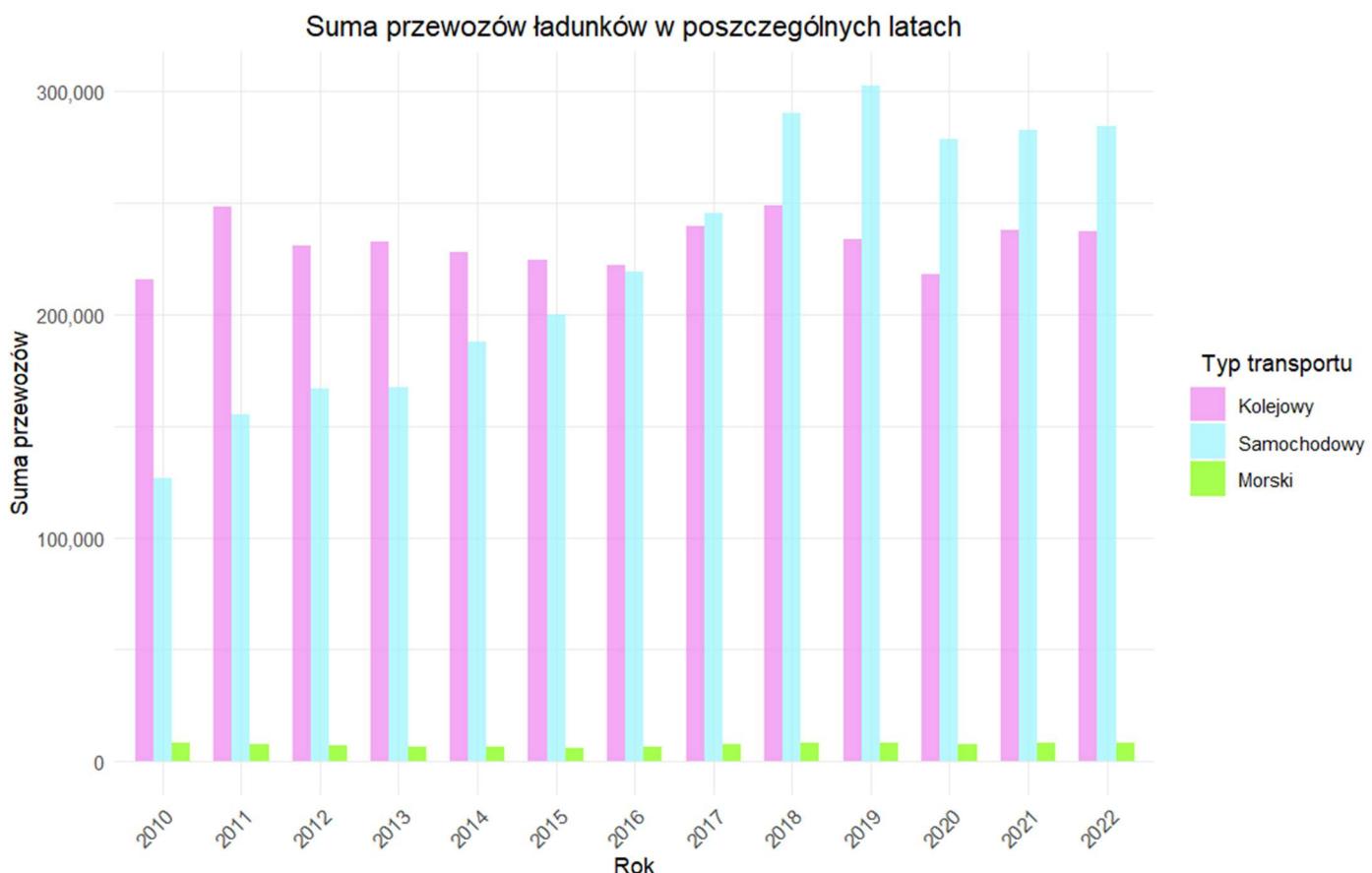
Rysunek 45 Histogram przewozów pasażerów

Analizując histogram przewozów pasażerów ogółem, można zauważyc, że większość wartości koncentruje się w przedziale między 40000 a 70000, co wskazuje na częstsze występowanie przewozów pasażerów o tych określonych wartościach. Wartości skrajne, zarówno te poniżej 40000, jak i powyżej 70000, są mniej liczne, co sugeruje, że są one mniej powszechne w analizowanych danych. Histogram ukazuje wyraźnie dominujący zakres wartości przewozów pasażerów, co wskazuje na ich charakterystyczny rozkład w badanej próbie danych.

5.4. Wykresy słupkowe

Wykres słupkowy pozwala porównać ilości przewozów w różnych rodzajach transportu w poszczególnych latach. Poprzez porównanie wysokości słupków dla różnych rodzajów transportu w tych samych latach, można ocenić, który rodzaj transportu był bardziej aktywny lub miał większy udział w przewozie ładunków lub pasażerów w danym okresie czasu.

5.4.1. Wykres słupkowy przewozów ładunków w poszczególnych latach z podziałem na rodzaj transportu



Rysunek 46 Wykres słupkowy przewozów ładunków

Można zauważyć, że transport kolejowy utrzymuje się cały czas na poziomie 200,000 – 250,000 i dominował do 2017 roku. Najniższe wartości zanotował w 2010 oraz w 2020 roku, natomiast najwyższą wartość w 2011 roku sięgającą 250,000. Jeśli chodzi o transport samochodowy zauważamy, że od 2010 roku zanotował duży wzrost, w 2017 roku suma przewozów ładunków transportem samochodowym, przewyższyła transport kolejowy. Najniższą wartość zanotowano w 2010 roku, natomiast najwyższą w 2019 roku. Transport morski jest znaczco niższy od innych typów transportu. Suma przewozów ładunków transportem morskim utrzymywała wartości na podobnym poziomie 5,000 – 10,000.

Kod dla wykresów słupkowych:

```
##### Wykres słupkowy przewozów ładunków w poszczególnych latach z podziałem na rodzaj transportu

# Grupowanie danych i sumowanie wartości dla każdego roku
dane_suma <- dane %>%
  mutate(Rok = substr(Okres, 1, 4)) %>% # Dodanie kolumny z rokiem
  group_by(Rok) %>%
  summarise(
    suma_transport_kolejowy = sum(`transport kolejowy - ładunki`),
    suma_transport_samochodowy = sum(`transport samochodowy - ładunki`),
    suma_transport_morski = sum(`transport morski - ładunki`)
  ) %>%
  ungroup()

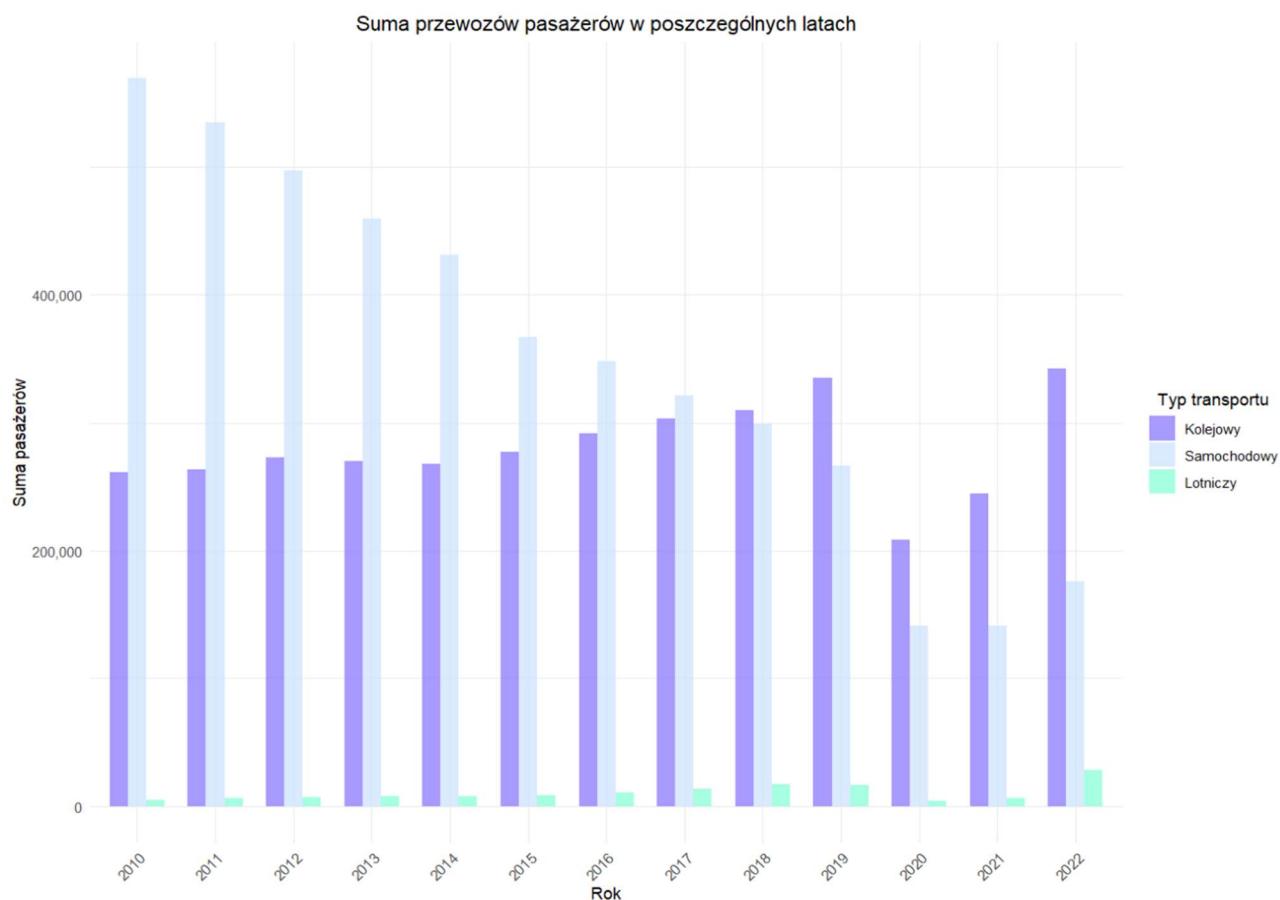
# Przekształcenie danych do formatu długiego
dane_suma_long <- melt(dane_suma, id.vars = "Rok", variable.name = "Typ_transportu", value.name = "Suma_ladunkow")

# Tworzenie wykresu słupkowego
wykres_suma <- ggplot(dane_suma_long, aes(x = Rok, y = Suma_ladunkow, fill = Typ_transportu)) +
  geom_bar(stat = "identity", position = "dodge", width = 0.7, alpha = 0.7) +
  labs(title = "Suma przewozów ładunków w poszczególnych latach",
       x = "Rok", y = "Suma przewozów") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_fill_manual(values = c("violet", "cadetblue1", "chartreuse1"),
                    name = "Typ transportu",
                    labels = c("Kolejowy", "Samochodowy", "Morski")) +
  scale_y_continuous(labels = scales::comma) +
  theme(legend.title = element_text(hjust = 0.5),
        plot.title = element_text(hjust = 0.5))

print(wykres_suma)
```

Rysunek 47 Kod generujący wykres słupkowy

5.4.2. Wykres słupkowy przewozów pasażerów w poszczególnych latach z podziałem na rodzaj transportu



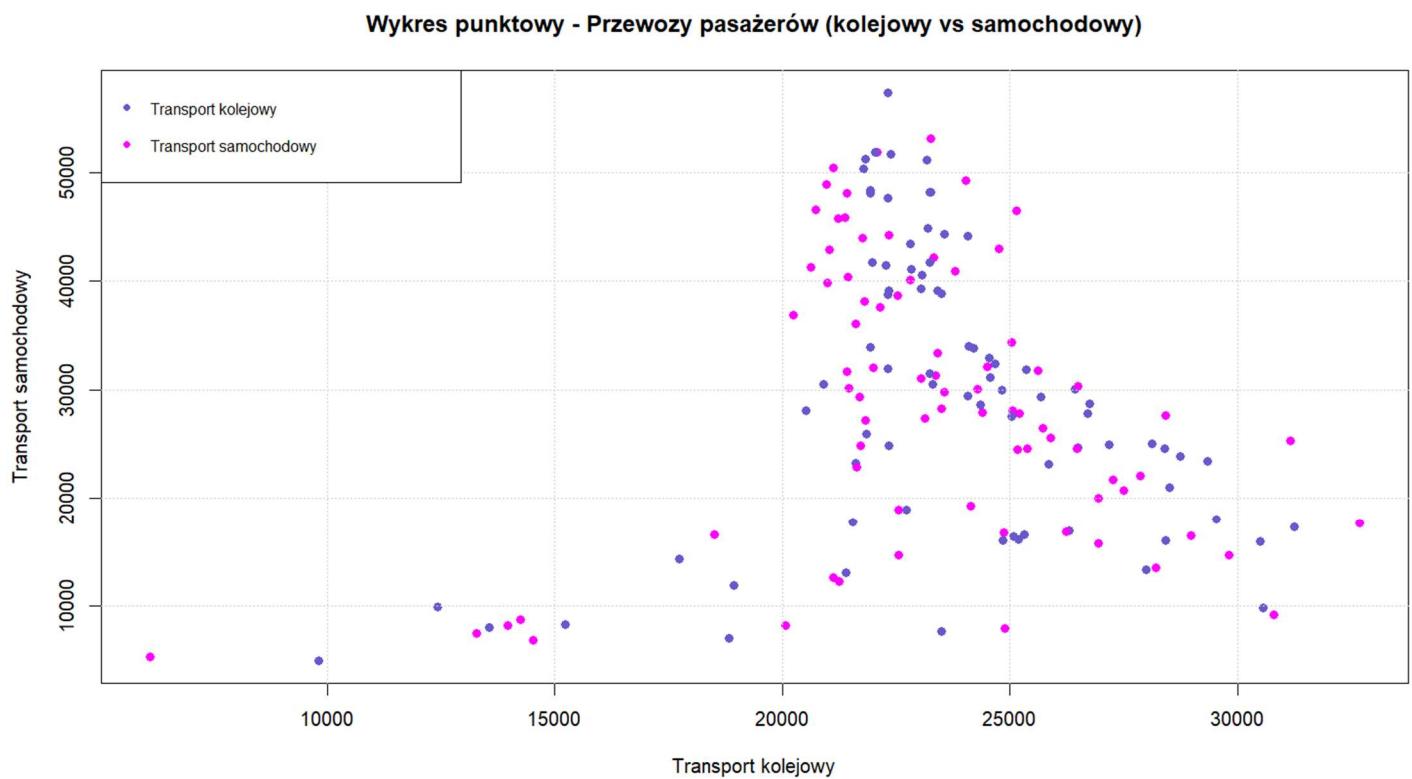
Rysunek 48 Wykres słupkowy przewozów pasażerów

Wykres przedstawia sumę przewozów pasażerów dla różnych typów transportu na przestrzeni lat. Zauważamy, że transport samochodowy w tych latach zanotował duży spadek. Do roku 2018 zdecydowanie dominował nad innymi typami transportu. Najwyższą wartość odnotował w 2010 roku, gdzie suma przewozów sięgała ponad 550,000. Najniższe wartości zanotował w 2020 oraz w 2021 roku, za to w 2022 roku suma przewozów lekko się zwiększyła. Jeśli chodzi o transport kolejowy zauważamy, że jego wartości na przestrzeni tych lat mieściły się w przedziale 250,000 – 350,000. Najniższą wartość zanotowano w 2020 roku, gdzie zauważamy znaczny spadek w porównaniu do lat poprzednich. natomiast najwyższą w 2019 oraz 2022 roku. Transport lotniczy jest znacznie niższy od innych typów transportu. Suma przewozów pasażerów transportem lotniczym utrzymywała wartości na podobnym poziomie, pomału wzrastając z roku na rok. W latach 2020 i 2021 zauważamy spadek wartości, w porównaniu do roku 2022, gdzie odnotowujemy najwyższą wartość dla tego typu transportu.

5.5. Wykresy punktowe

Wykres punktowy to rodzaj wykresu używanego do wizualizacji zależności między dwiema zmiennymi. Jest to graficzna reprezentacja danych, gdzie każdy punkt na wykresie reprezentuje jedną obserwację lub parę wartości z dwóch zmiennych.

5.5.1. Wykres punktowy dla przewozów pasażerów (kolejowy vs samochodowy)



Rysunek 49 Wykres punktowy przewozów pasażerów

Wykres punktowy, który został przedstawiony, obrazuje związek między dwoma zmiennymi: przewozem pasażerów transportem kolejowym (oś OX) i przewozem pasażerów transportem samochodowym (oś OY). Każdy punkt na wykresie reprezentuje parę wartości: ilość przewiezionych pasażerów danego rodzaju transportu dla danej obserwacji.

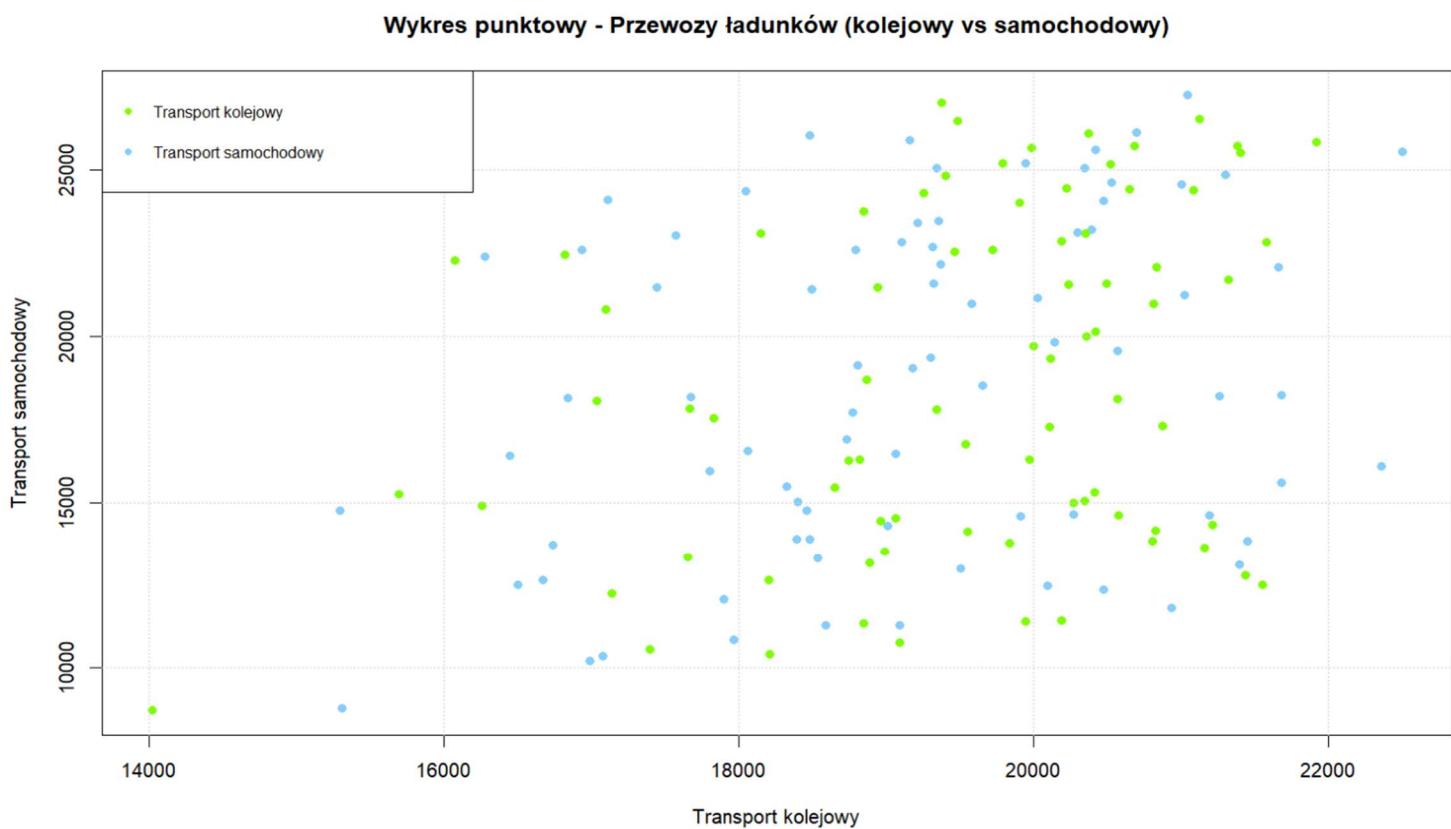
Analizując wykres, możemy zauważyć, że istnieje pewna zależność między przewozami pasażerów transportem kolejowym a samochodowym. Wartości tych przewozów niekoniecznie rosną lub maleją proporcjonalnie, ale możemy zauważyć, że im więcej pasażerów przewozi się transportem kolejowym, tym zazwyczaj większa jest także liczba pasażerów przewożonych transportem samochodowym.

Poniżej przedstawiony został kod, który generuje wykresy punktowe.

```
419 ##### Wykres punktowy - Przewozy pasażerów (kolejowy vs samochodowy)
420 plot(dane$`transport kolejowy - pasazerowie`,
421       dane$`transport samochodowy - pasazerowie`,
422       main = "Wykres punktowy - Przewozy pasażerów (kolejowy vs samochodowy)",
423       xlab = "Transport kolejowy",
424       ylab = "Transport samochodowy",
425       col = c("slateblue3", "magenta"), # Dwa różne kolory punktów
426       pch = 16, # Kształt punktów (tutaj: kółka)
427       cex = 1) # Rozmiar punktów
428
429 # Dodanie siatki
430 grid()
431
432 # Dodanie legendy
433 legend("topleft", legend = c("Transport kolejowy", "Transport samochodowy"),
434        col = c("slateblue3", "magenta"), pch = 16, cex = 0.8)
435
```

Rysunek 50 Kod generujący przewozy pasażerów

5.5.2. Wykres punktowy dla przewozów ładunków (kolejowy vs samochodowy)



Rysunek 51 Wykres punktowy przewozów ładunków

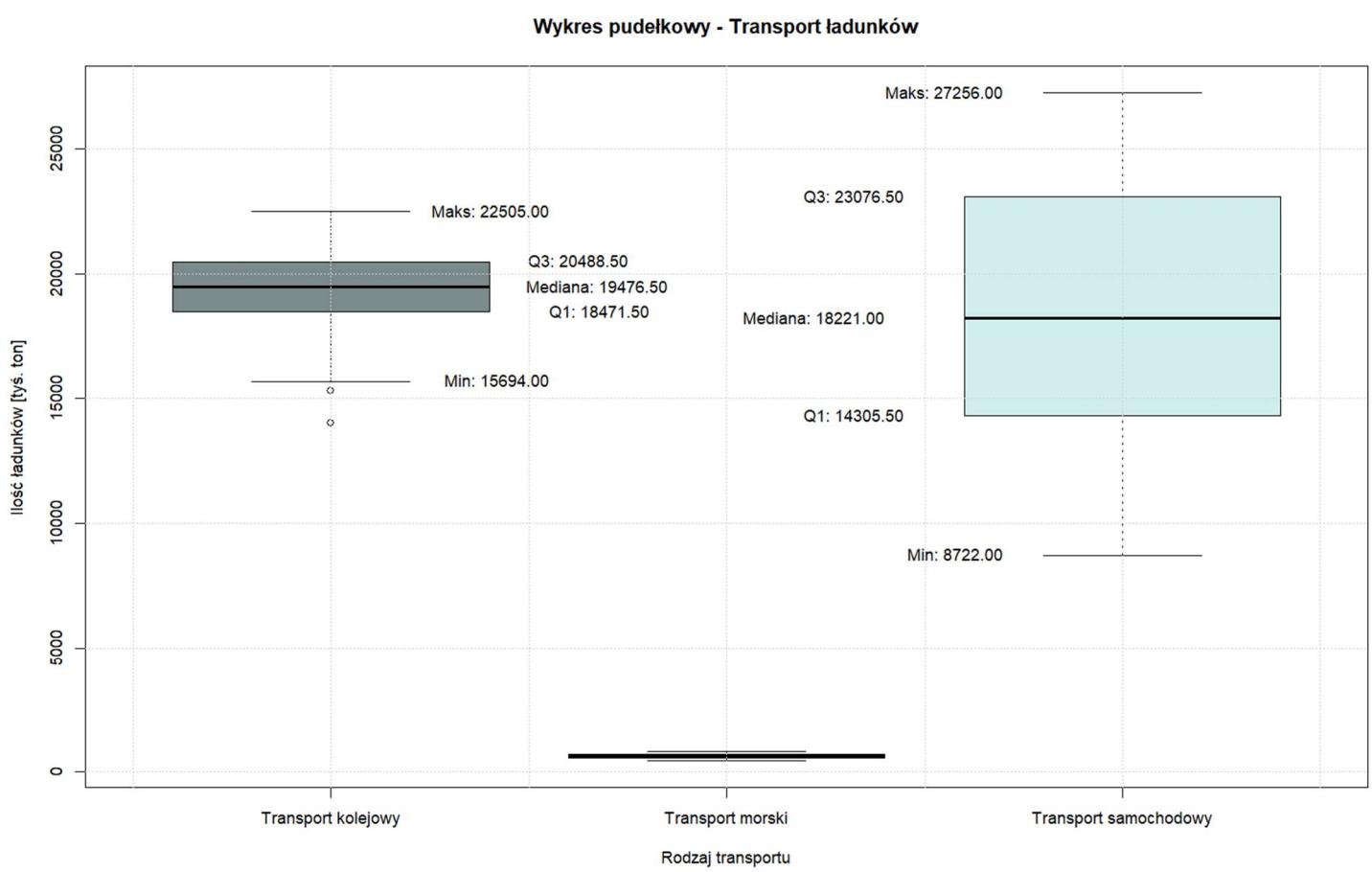
Analizując wykres, możemy stwierdzić, że istnieje tendencja do pewnego wzrostu ilości przewożonych ładunków zarówno transportem kolejowym, jak i samochodowym. Widoczne są pewne piki, co sugeruje okresy, w których następuje wzmożona aktywność przewozów. Wartości ilości przewożonych ładunków transportem kolejowym są często zbliżone do ilości przewożonych transportem samochodowym, ale nie zawsze rosną w sposób

proporcjonalny. Oznacza to, że w różnych przypadkach oba rodzaje transportu mogą przewozić podobne ilości ładunków.

5.6. Wykresy pudełkowe

Wykres pudełkowy umożliwia wizualizację kluczowych cech zbioru danych. Ukazuje medianę (środkową wartość), granice kwartyli: dolny kwartyl (Q1) stanowi 25% danych, a górny kwartyl (Q3) zawiera 75% danych, wąsy wychodzące z pudełka określają zakres wartości, a wartości odstające, znajdujące się poza wąsami, mogą stanowić wartości znacznie różniące się od reszty zbioru danych, będąc potencjalnymi punktami odstającymi.

5.6.1. Wykres pudełkowy przewozów ładunków dla transportu kolejowego, samochodowego i morskiego



Rysunek 52 Wykres pudełkowy przewozów ładunków

Transport Kolejowy:

- Mediana (Linia środkowa w pudełku): Wynosi około 19500 ton, co oznacza, że połowa danych mieści się w okolicach tej wartości.
- Pudełko (Kwartyle dolny i górny):
 - Dolny kwartyl (Q1) wynosi około 18500 ton, co oznacza, że 25% danych znajduje się poniżej tej wartości.
 - Górnny kwartyl (Q3) wynosi około 20000 ton, co oznacza, że 25% danych znajduje się powyżej tej wartości.
- Wąsy (Zakres): Wartości nieodstające znajdują się między około 16000 a 22500 ton.
- Wartości odstające (Outliers): Widoczne są pojedyncze wartości odstające poniżej około 16000 ton.

Transport Samochodowy:

- Mediana (Linia środkowa w pudełku): Wynosi około 18000 ton, co oznacza, że połowa danych mieści się w okolicach tej wartości.
- Pudełko (Kwartyle dolny i górny):
 - Dolny kwartyl (Q1) wynosi około 14000 ton, co oznacza, że 25% danych znajduje się poniżej tej wartości.
 - Górnny kwartyl (Q3) wynosi około 23000 ton, co oznacza, że 25% danych znajduje się powyżej tej wartości.
- Wąsy (Zakres): Wartości nieodstające znajdują się między około 9000 a 27000 ton.

5.6.2. Wykres pudełkowy przewozów ładunków dla transportu morskiego



Rysunek 53 Wykres pudełkowy przewozu ładunków transportem morskim

Transport Morski:

- Mediana (Linia środkowa w pudełku): Wynosi około 620 ton, co oznacza, że jest to środkowa wartość dla danych transportu morskiego.
- Pudełko (Kwartyle dolny i górny):
 - Dolny kwartyl (Q1) wynosi około 560 ton, co oznacza, że 25% danych jest poniżej tej wartości.
 - Górnny kwartyl (Q3) wynosi około 700 ton, co oznacza, że 25% danych jest powyżej tej wartości.
- Wąsy (Zakres): Wartości nieodstające znajdują się między około 430 a 830 ton.

Kod generujący wykres pudełkowy:

```
429 # Wykresy pudełkowe
430
431 # tworzenie wykresu pudełkowego dla transportu kolejowego, samochodowego i morskiego (przewoz ładunków)
432 # Tworzenie wykresu pudełkowego
433 wp1 <- boxplot(
434   dane$`transport kolejowy - ładunki`,
435   dane$`transport morski - ładunki`,
436   dane$`transport samochodowy - ładunki`,
437   main = "Wykres pudełkowy - Transport ładunków",
438   xlab = "Rodzaj transportu",
439   ylab = "Ilość ładunków [tys. ton]",
440   col = c("lightcyan1", "lightcyan2", "lightcyan3"),
441   names = c("Transport kolejowy", "Transport morski", "Transport samochodowy")
442 )
443
444 # Dodawanie parametrów opisowych na wykresie
445 text(1, wp1$stats[2, 1], labels = sprintf("Q1: %.2f", wp1$stats[2, 1]), pos = 4, offset = 11)
446 text(1, wp1$stats[3, 1], labels = sprintf("Mediana: %.2f", wp1$stats[3, 1]), pos = 2, offset = -18)
447 text(1, wp1$stats[4, 1], labels = sprintf("Q3: %.2f", wp1$stats[4, 1]), pos = 2, offset = -16)
448 text(1, wp1$stats[1, 1], labels = sprintf("Min: %.2f", wp1$stats[1, 1]), pos = 2, offset = -11)
449 text(1, wp1$stats[5, 1], labels = sprintf("Maks: %.2f", wp1$stats[5, 1]), pos = 2, offset = -12)
450
451 text(3, wp1$stats[2, 3], labels = sprintf("Q1: %.2f", wp1$stats[2, 3]), pos = 2, offset = 11)
452 text(3, wp1$stats[3, 3], labels = sprintf("Mediana: %.2f", wp1$stats[3, 3]), pos = 2, offset = 12)
453 text(3, wp1$stats[4, 3], labels = sprintf("Q3: %.2f", wp1$stats[4, 3]), pos = 2, offset = 11)
454 text(3, wp1$stats[1, 3], labels = sprintf("Min: %.2f", wp1$stats[1, 3]), pos = 2, offset = 6)
455 text(3, wp1$stats[5, 3], labels = sprintf("Maks: %.2f", wp1$stats[5, 3]), pos = 2, offset = 6)
456
457 # Dodanie siatki
458 grid()
```

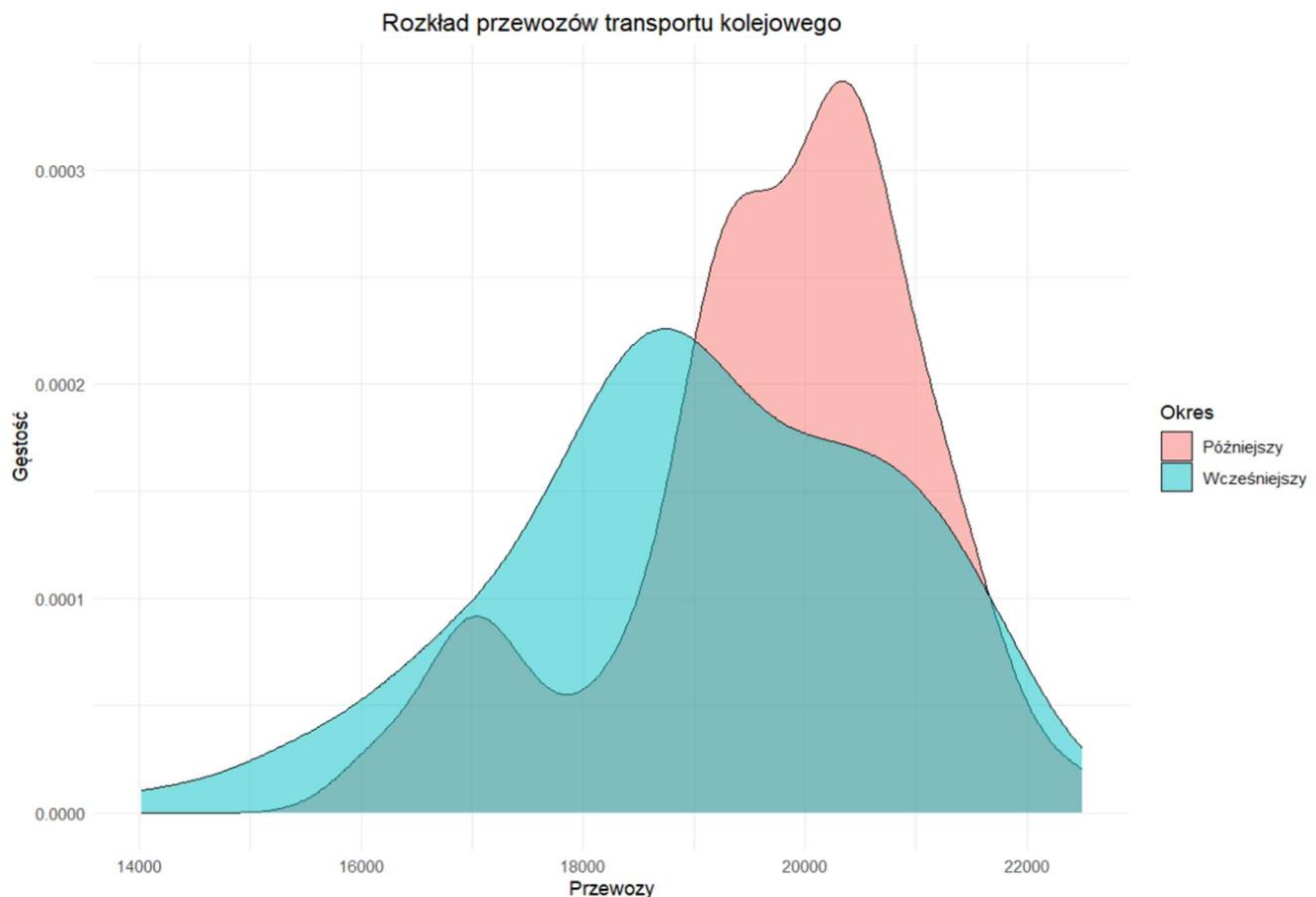
Rysunek 54 Kod generujący wykres pudełkowy

5.7. Wykres rozkładu dla przewozów ładunków w transporcie kolejowym w dwóch okresach

Wykres rozkładu jest używany do wizualizacji rozkładu danych. W przypadku wykresu gęstości dla transportu kolejowego w dwóch okresach (wcześniejszym i późniejszym), obrazuje różnice w dystrybucji ilości przewożonych ładunków między dwoma okresami czasowymi, pozwalając na wizualne porównanie rozkładu i ewentualne zmiany w dynamice transportu kolejowego w badanych latach.

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszysy): 2010-01 – 2016-06
- Okres drugi (późniejszy): 2016-07 – 2022-12



Rysunek 55 Wykres gęstości przewozów ładunków transportem kolejowym

Analizując wykres gęstości dla transportu kolejowego w dwóch okresach, można zauważyć:

- a) Różnice w rozkładzie: W pierwszym okresie przewozy wykazują szerszy zakres wartości, co sugeruje większą zmienność ilości przewożonych ładunków niż w drugim okresie. Rozkład w pierwszym okresie wydaje się być bardziej rozproszony.

- b) Przesunięcie średniej: Można zauważyc, że średnia wartość przewozów w drugim okresie jest wyższa niż w pierwszym okresie. Oznacza to potencjalny wzrost ilości przewożonych ładunków w transporcie kolejowym w drugim okresie w porównaniu do pierwszego.
- c) Skupienie wartości w drugim okresie: Wykres sugeruje, że wartości w drugim okresie są bardziej skupione wokół pewnego przedziału wartości niż w pierwszym okresie, co może oznaczać mniejszą zmienność ilości przewożonych ładunków.

Kod generujący wykres gęstości:

```

485 ##### wykres gęstości
486
487 # Tworzenie ramki danych dla każdego okresu
488 okres_wczesniejszy <- dane$`transport kolejowy - ładunki`[1:78] # Założmy, że wcześniejszy okres obejmuje
489 okres_pozniejszy <- dane$`transport kolejowy - ładunki`[79:156] # Założmy, że późniejszy okres obejmuje k
490
491 df_okresy <- data.frame(
492   Przewozy = c(okres_wczesniejszy, okres_pozniejszy),
493   Okres = rep(c("Wczesniejszy", "Późniejszy"), each = length(okres_wczesniejszy))
494 )
495
496 library(ggplot2)
497 library(scales) # Dodajemy bibliotekę do obsługi formatowania etykiet
498
499 ggplot(df_okresy, aes(x = Przewozy, fill = Okres)) +
500   geom_density(alpha = 0.5) +
501   labs(title = "Rozkład przewozów transportu kolejowego",
502       x = "Przewozy", y = "Gęstość") +
503   theme_minimal() +
504   theme(plot.title = element_text(hjust = 0.5)) +
505   scale_y_continuous(labels = comma) # Używamy funkcji comma do zmiany formatu etykiet na czytelniejszy
506

```

Rysunek 56 Kod generujący wykres gęstości

5.8. Wykres skrzypcowy dla przewozów pasażerów transportem samochodowym w dwóch okresach

Wykres skrzypcowy jest narzędziem wizualizacyjnym, które pozwala na zobrazowanie rozkładu i charakterystyki danych dla dwóch okresów przewozu pasażerów transportem samochodowym.

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszego): 2010-01 – 2016-06
- Okres drugi (późniejszego): 2016-07 – 2022-12



Rysunek 57 Wykres skrzypcowy przewozów pasażerów dla transportu samochodowego

Na wykresie można zaobserwować następujące cechy:

- Porównanie rozkładu: Wykres skrzypcowy dla obu okresów pozwala na porównanie gęstości rozkładu danych. Pierwszy okres wykazuje szerszy rozrzut wartości, obejmujący zakres od około 19 000 do około 58 000, z większym skupieniem wartości w zakresie wyższych liczb pasażerów. W przypadku drugiego okresu widoczny jest węższy zakres, skupiony głównie w przedziale od około 5 000 do około 32 000.
- Koncentracja danych: Wartości dla drugiego okresu są ogólnie niższe niż dla pierwszego. Wykres ten wskazuje na potencjalny spadek liczby przewożonych pasażerów.

- Zmiany w medianie: Mediana pierwszego okresu jest wyższa niż mediana drugiego. Sugeruje to, że wartości są bardziej skupione wokół wyższych liczb w pierwszym okresie niż w drugim.
- Rozkład kwartylowy: Wartości kwartylowe dla pierwszego okresu są wyższe niż dla drugiego okresu. Może to wskazywać, że pierwszy okres zawiera więcej wartości w wyższym przedziale liczb pasażerów, podczas gdy drugi okres charakteryzuje się mniejszym zakresem wartości w tym obszarze.

Widoczne różnice w rozkładzie wartości, medianie oraz kwartyli mogą wskazywać na zmniejszenie się liczby pasażerów w drugim okresie w porównaniu z pierwszym.

Kod generujący wykres skrzypcowy:

```

501 ##### WYKRES SKRZYPCOWY
502
503 # Tworzenie ramki danych dla dwóch okresów: wcześniejszego i późniejszego
504 okres_wczesniejszy <- dane$`transport samochodowy - pasazerowie`[1:78]
505 okres_pozniejszy <- dane$`transport samochodowy - pasazerowie`[79:length(dane$`transport samochodowy - pasazerowie`)]
506
507 # Tworzenie ramki danych
508 df_porownanie <- data.frame(
509   Przewozy = c(okres_wczesniejszy, okres_pozniejszy),
510   Okres = rep(c("Wcześniejszy", "Późniejszy"), each = length(okres_wczesniejszy)))
511 )
512
513 # Zmiana kolejności okresów w ramce danych i faktora
514 df_porownanie$Okres <- factor(df_porownanie$Okres, levels = c("Późniejszy", "Wcześniejszy"))
515
516 # Obliczenie wielkości próby
517 sample_size <- df_porownanie %>%
518   group_by(Okres) %>%
519   summarize(num = n())
520
521 # Rysowanie wykresu z dostosowanymi parametrami
522 df_porownanie %>%
523   left_join(sample_size) %>%
524   mutate(myaxis = paste0(Okres, "\n", "n=", num)) %>%
525   ggplot(aes(x = myaxis, y = Przewozy, fill = Okres)) +
526   geom_violin(width = 0.6, alpha = 0.8) +
527   geom_boxplot(width = 0.2, color = 'black', alpha = 0.4) +
528   scale_fill_manual(values = c("Późniejszy" = "lightblue", "Wcześniejszy" = "lightgreen")) +
529   theme_minimal() +
530   theme(
531     legend.position = "none",
532     plot.title = element_text(size = 11, hjust = 0.5), # wyśrodkowany tytuł
533   ) +
534   ggtitle("Wykres skrzypcowy dla przewozów pasażerów w dwóch okresach") +
535   xlab("Okres") # Opis osi x

```

Rysunek 58 Kod generujący wykres skrzypcowy

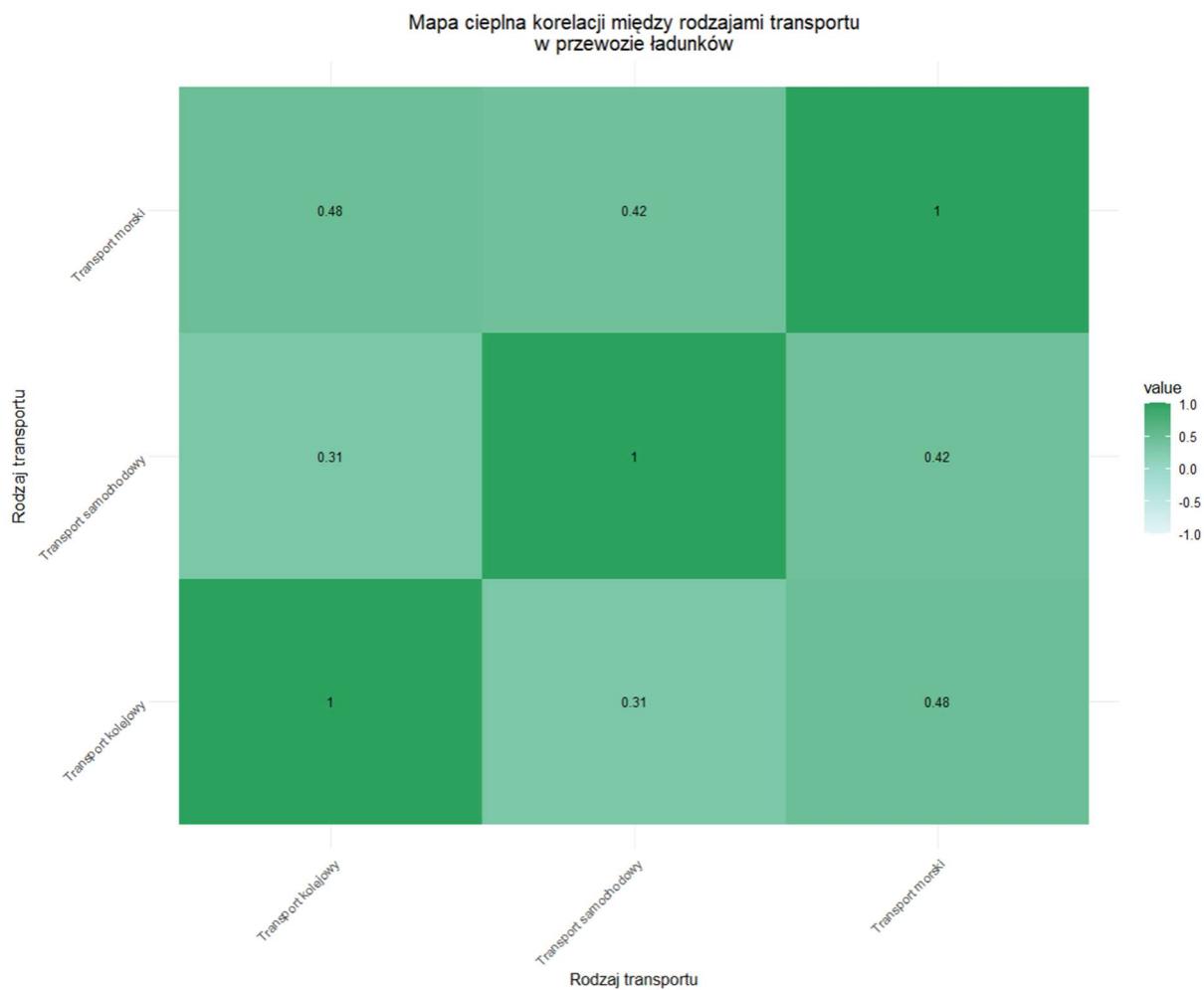
5.9. Mapy cieplne

Mapa ciepła (ang. heatmap) korelacji to wizualizacja, która przedstawia siłę związku (korelację) między różnymi zmiennymi w formie kolorowego obrazu. Każdy prostokąt na mapie reprezentuje parę zmiennych, a kolor wewnętrz prostokąta wskazuje na siłę korelacji między nimi.

Głównym celem mapy cieplnej korelacji jest zrozumienie wzajemnych zależności między zmiennymi. Zmienna korelacja może mieć wartość od -1 do 1:

- Korelacja bliska -1 oznacza silną negatywną zależność, czyli gdy jedna zmienna rośnie, druga maleje.
- Korelacja bliska 1 oznacza silną pozytywną zależność, czyli gdy jedna zmienna rośnie, druga też rośnie.
- Korelacja bliska 0 oznacza brak związku między zmiennymi.

5.9.1. Mapa cieplna korelacji między rodzajami transportu w przewozie ładunków

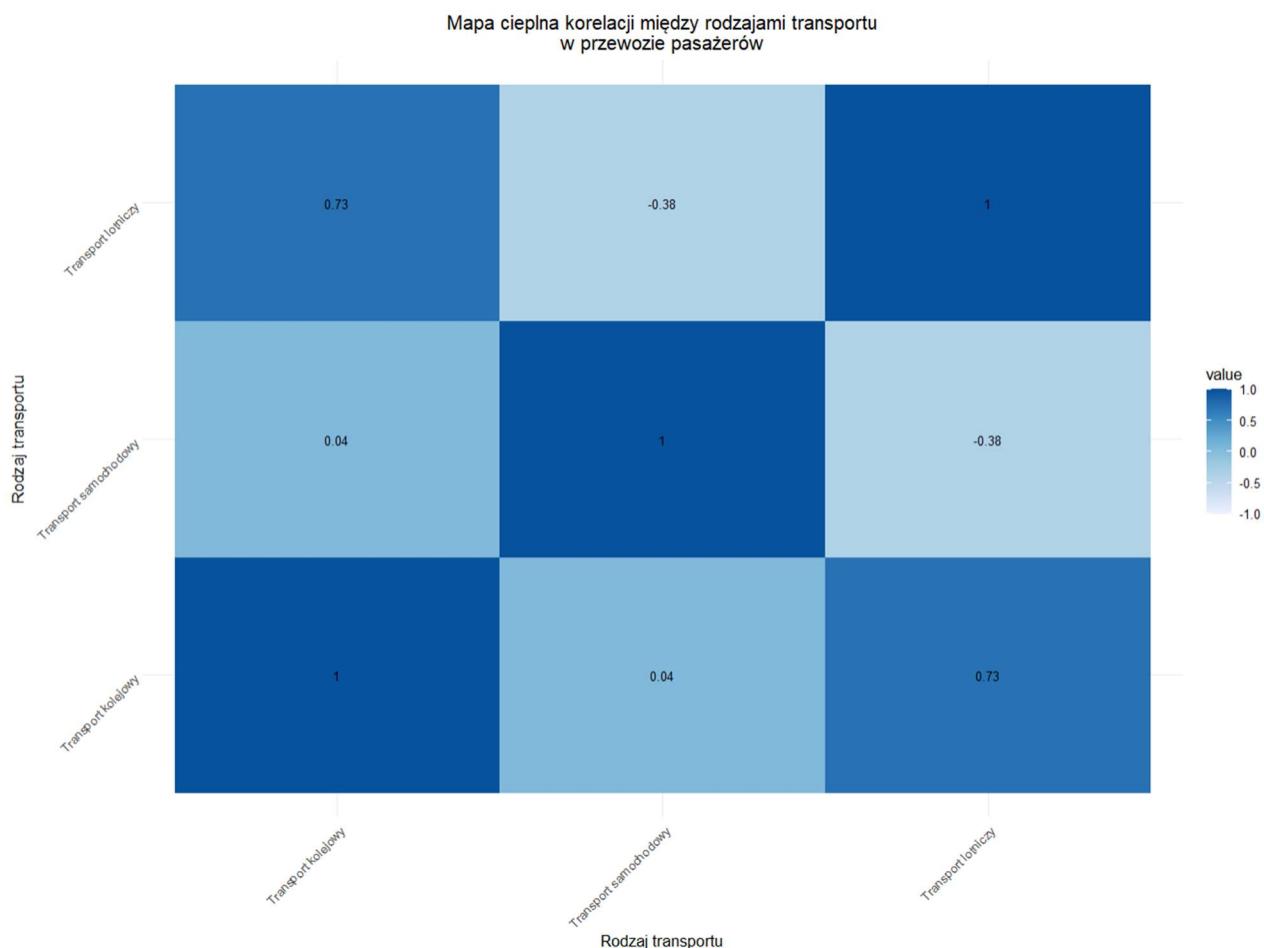


Rysunek 59 Mapa cieplna korelacji dla przewozu ładunków

Analiza wykresu korelacji między rodzajami transportu w przewozie ładunków wykazała umiarkowane związki między poszczególnymi metodami transportu. Wyniki wskazują na pewne współzależności między nimi, choć są one raczej umiarkowane niż silne.

- **Transport kolejowy i transport morski:** Istnieje umiarkowana dodatnia korelacja między transportem kolejowym a morskim (0,48). To sugeruje, że wzrost w transporcie kolejowym może częściowo korelować z zwiększeniem transportu morskiego, ale nie jest to bardzo silna zależność.
- **Transport samochodowy i transport morski:** Podobnie, istnieje umiarkowana dodatnia korelacja między transportem samochodowym a morskim (0,42). To sugeruje, że wzrost w transporcie samochodowym może być częściowo powiązany ze zwiększeniem transportu morskiego, choć ta zależność nie jest znacząca.
- **Transport kolejowy i transport samochodowy:** Wyniki wykazują słabą dodatnią korelację między transportem kolejowym a samochodowym (0,31). Wzrost w transporcie kolejowym może nieznacznie korelować ze zwiększeniem transportu samochodowego, jednakże ta zależność jest stosunkowo słaba w porównaniu do pozostałych.

5.9.2. Mapa cieplna korelacji między rodzajami transportu w przewozie pasażerów



Rysunek 60 Mapa cieplna korelacji dla przewozu pasażerów

Wartości korelacji pozwalają stwierdzić, że istnieją pewne związki między rodzajami transportu w przewozie pasażerów, ale większość tych związków jest albo bardzo słaba, albo nieistotna. Oznacza to, że zmiany w jednym rodzaju transportu nie są mocno powiązane z równoczesnymi zmianami w pozostałych rodzajach transportu w kontekście przewozu pasażerów.

- **Transport kolejowy i transport lotniczy:** Istnieje silna dodatnia korelacja między transportem kolejowym a lotniczym (0,73). To oznacza, że zmiany w jednym rodzaju transportu zwykle współwystępują z podobnymi zmianami w drugim rodzaju transportu.
- **Transport samochodowy i transport lotniczy:** Korelacja między tymi rodzajami transportu jest słabo ujemna (-0,38), co wskazuje na bardzo niewielką przeciwną zależność między nimi.

- **Transport kolejowy i transport samochodowy:** Korelacja między nimi jest bardzo słaba i prawie nieistotna (0,04), co sugeruje brak zauważalnej zależności między ilościami transportu kolejowego i samochodowego.

Kod generujący mapy ciepła:

```

540 # Wykresy map cieplnych
541
542 # Przygotowanie danych dla wykresu heatmap dla przewozu ładunków
543 dane_heatmap_ladunki <- dane[, c("transport kolejowy - ładunki",
544                               "transport samochodowy - ładunki", "transport morski - ładunki")]
545
546 correlation_matrix_ladunki <- cor(dane_heatmap_ladunki)
547
548 # Tworzenie wykresu heatmap dla przewozu ładunków
549 wykres_heatmap_ladunki <- ggplot(data = melt(correlation_matrix_ladunki),
550                                   aes(x = Var1, y = Var2, fill = value)) +
551   geom_tile() +
552   scale_fill_gradientn(colors = brewer.pal(3, "BuGn"),
553                         limits = c(-1, 1), na.value = "grey50") +
554   geom_text(aes(label = round(value, 2)), color = "black", size = 3) +
555   labs(title = "Mapa cieplna korelacji między rodzajami transportu\nw przewozie ładunków",
556         x = "Rodzaj transportu", y = "Rodzaj transportu") +
557   scale_x_discrete(labels = c("transport kolejowy - ładunki" = "Transport kolejowy",
558                             "transport samochodowy - ładunki" = "Transport samochodowy",
559                             "transport morski - ładunki" = "Transport morski")) +
560   scale_y_discrete(labels = c("transport kolejowy - ładunki" = "Transport kolejowy",
561                             "transport samochodowy - ładunki" = "Transport samochodowy",
562                             "transport morski - ładunki" = "Transport morski")) +
563   theme_minimal() +
564   theme(axis.text.x = element_text(angle = 45, hjust = 1),
565         axis.text.y = element_text(angle = 45, hjust = 1),
566         plot.title = element_text(hjust = 0.5)) # Wyśrodkowanie tytułu
567
568 print(wykres_heatmap_ladunki)

```

Rysunek 61 Kod generujący mapy cieplne

6. Weryfikacja hipotez statystycznych

6.1. Testowanie różnicy średnich przewozów pasażerów między transportem samochodowym a kolejowym

Do porównywania średnich wartości pomiędzy dwoma grupami możemy użyć **testu t-Studenta** dla dwóch niezależnych prób. W tym przypadku, chcemy porównać średnią ilość przewozów pasażerów transportem samochodowym z średnią ilością przewozów transportem kolejowym.

Hipotezy:

- **Hipoteza zerowa (H0):** Średnia ilość przewozów pasażerów transportem samochodowym jest równa średniej ilości przewozów transportem kolejowym.
- **Hipoteza alternatywna (H1):** Średnia ilość przewozów pasażerów transportem samochodowym różni się od średniej ilości przewozów transportem kolejowym.
- **Poziom istotności (alfa):** 0,05

```
407 # Testowanie średnich dla przewozów pasażerów transportem samochodowym i kolejowym
408 wynik_testu <- t.test(dane$`transport samochodowy - pasazerowie`,
409                         dane$`transport kolejowy - pasazerowie`)
410
411 # Wyświetlenie wyników testu
412 print(wynik_testu)
```

Rysunek 62 Kod wykonujący test t-Studenta

```
Welch Two Sample t-test

data: dane$`transport samochodowy - pasazerowie` and dane$`transport kolejowy - pasazerowie`
t = 5.3548, df = 182.94, p-value = 2.548e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3655.218 7920.308
sample estimates:
mean of x mean of y
29180.10 23392.34
```

Rysunek 63 Wynik kodu testu t-Studenta

Wyniki testu:

- Wartość t: Otrzymano wartość t równą 5,3548.
- Stopnie swobody (df): Stopnie swobody wynoszą 182,94.
- Wartość p: P-wartość wynosi 2,548e-07.
- Hipoteza alternatywna: Istnieje istotna różnica między średnimi ilościami przewozów pasażerów w obu rodzajach transportu.

- Przedział ufności: 95% przedział ufności dla różnicy średnich wynosi od 3655,218 do 7920,308.
- Estymatory próbkkowe: Średnia ilość przewozów pasażerów transportem samochodowym wynosi około 29180,10, a transportem kolejowym około 23392,34.

Wynik testu wskazuje na istotną statystyczną różnicę między średnimi ilościami przewozów pasażerów transportem samochodowym a transportem kolejowym. Otrzymana wartość p jest znaczco mniejsza od założonego poziomu istotności alfa = 0.05, co pozwala odrzucić hipotezę zerową na rzecz hipotezy alternatywnej, potwierdzając różnicę między średnimi ilościami przewozów pasażerów w obu rodzajach transportu.

6.2. Testowanie różnicy średnich przewozów ładunków w transporcie kolejowym w dwóch okresach

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszys): 2010-01 – 2016-06
- Okres drugi (późniejszy): 2016-07 – 2022-12

Hipotezy:

- **Hipoteza zerowa (H0):** Średnia ilość przewozów ładunków transportem kolejowym w dwóch okresach jest równa.
- **Hipoteza alternatywna (H1):** Średnia ilość przewozów ładunków transportem kolejowym w dwóch okresach nie jest równa.
- **Poziom istotności (alfa):** 0,05

```

417 # wybieram dane dotyczące przewozów ładunków transportem kolejowym
418 dane_kolejowe <- dane$`transport kolejowy - ladunki`
419
420 # Dzielę dane na dwie równe części - wcześniejszy i późniejszy okres
421 polowa_dlugosc <- length(dane_kolejowe) / 2
422 wczesniejszy_okres <- dane_kolejowe[1:polowa_dlugosc]
423 pozniejszy_okres <- dane_kolejowe[(polowa_dlugosc + 1):length(dane_kolejowe)]
424
425 # Przeprowadzam test t-Studenta dla dwóch niezależnych prób
426 wynik_testu <- t.test(wczesniejszy_okres, pozniejszy_okres)
427
428 # Wyświetl wyniki testu
429 print(wynik_testu)
```

Rysunek 64 Kod wykonujący test t-Studenta

```
Welch Two Sample t-test

data: wczesniejszy_okres and poznajeszy_okres
t = -2.5974, df = 147.02, p-value = 0.01035
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1143.1683 -155.2676
sample estimates:
mean of x mean of y
19027.55 19676.77
```

Rysunek 65 Wynik testu t-Studenta

Wyniki testu:

- Wartość t: Otrzymano wartość t równą -2,5974.
- Stopnie swobody (df): Stopnie swobody wynoszą 147,02.
- Wartość p: P-wartość wynosi 0,01035.
- Hipoteza alternatywna: Istnieje istotna różnica między średnimi ilościami przewozów w dwóch okresach.
- Przedział ufności: 95% przedział ufności dla różnicy średnich wynosi od -1143,1683 do -155,2676.
- Estymatory próbkkowe: Średnia ilość przewozów w pierwszym okresie wynosi około 19027,55, a w drugim okresie około 19676,77.

W oparciu o wynik testu, który wykazał istotną statystyczną różnicę między średnimi ilościami przewozów ładunków transportem kolejowym w dwóch okresach, **odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej**. W związku z tym możemy stwierdzić, że ilość przewozów ładunków transportem kolejowym nie jest taka sama w tych dwóch okresach czasowych.

6.3. Testy korelacji między przewozem ładunków transportem kolejowym a samochodowym

W niniejszym badaniu przeprowadzimy analizę korelacji między ilością przewożonych ładunków transportem kolejowym i samochodowym. Celem jest zbadanie ewentualnej zależności oraz stopnia powiązania pomiędzy tymi dwiema zmiennymi. Do przeprowadzenia analizy zastosowano trzy różne metody testów korelacji: Pearsona, Spearmana i Kendalla. Każda z tych metod pozwala na ocenę związku między zmiennymi, jednak różnią się w sposobie uwzględniania relacji liniowej oraz nieliniowej. Wyniki tych testów pomogą w zrozumieniu, czy istnieje silna, słaba czy brak korelacji między ilościami przewożonych ładunków za pomocą tych dwóch rodzajów transportu.

```

575 # Korelacja Pearsona
576 cor_pearson <- cor.test(dane$`transport samochodowy - ładunki`,
577                         dane$`transport kolejowy - ładunki`, method = "pearson")
578 print("Korelacja Pearsona:")
579 print(cor_pearson)
580
581 # Korelacja Spearmana
582 cor_spearman <- cor.test(dane$`transport samochodowy - ładunki`,
583                           dane$`transport kolejowy - ładunki`, method = "spearman")
584 print("Korelacja Spearmana:")
585 print(cor_spearman)
586
587 # Korelacja Kendalla
588 cor_kendall <- cor.test(dane$`transport samochodowy - ładunki`,
589                         dane$`transport kolejowy - ładunki`, method = "kendall")
590 print("Korelacja Kendalla:")
591 print(cor_kendall)

```

Rysunek 66 Kod wykonujący metody testów korelacji: Pearsona, Spearmana i Kendalla

6.3.1. Test korelacji Pearsona:

- **Hipoteza zerowa (H₀):** Brak istotnej korelacji między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Hipoteza alternatywna (H₁):** Istnieje istotna korelacja między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Poziom istotności (alfa):** 0,05

```

Pearson's product-moment correlation

data: dane$`transport samochodowy - ładunki` and dane$`transport kolejowy - ładunki`
t = 4.0629, df = 154, p-value = 7.701e-05
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.1619203 0.4464562
sample estimates:
        cor
0.3111442

```

Rysunek 67 Wynik testu korelacji Pearsona

Wartość współczynnika korelacji Pearsona wynosząca 0,311 wskazuje na istnienie umiarkowanej pozytywnej korelacji między tymi zmiennymi. Wynik testu jest istotny statystycznie ($p = 7,701\text{e-}05$, $\alpha = 0,05$), co oznacza **odrzucenie hipotezy zerowej na rzecz hipotezy alternatywnej**. Stwierdzono, że istnieje dodatni związek między ilością przewożonych ładunków transportem kolejowym i samochodowym.

6.3.2. Test korelacji Spearmana:

- **Hipoteza zerowa (H0):** Brak istotnej korelacji rangowej między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Hipoteza alternatywna (H1):** Istnieje istotna korelacja rangowa między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Poziom istotności (alfa):** 0,05

```
Spearman's rank correlation rho

data: dane$`transport samochodowy - ładunki` and dane$`transport kolejowy - ładunki`
S = 445622, p-value = 0.0001902
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.2956931
```

Rysunek 68 Wynik testu korelacji Spearmana

Zastosowanie testu korelacji Spearmana wykazało istnienie umiarkowanej dodatniej zależności ($\rho = 0,295$, $p = 0,0001902$) między przewozem ładunków transportem kolejowym i samochodowym. **Wartość p mniejsza od poziomu istotności $\alpha = 0,05$ skutkuje odrzuceniem hipotezy zerowej.** Wyniki sugerują istnienie statystycznie istotnego związku między ilościami przewożonych ładunków tymi dwoma rodzajami transportu.

6.3.3. Test korelacji Kendalla:

- **Hipoteza zerowa (H0):** Brak istotnej korelacji tau Kendalla między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Hipoteza alternatywna (H1):** Istnieje istotna korelacja tau Kendalla między przewozem ładunków transportem kolejowym a transportem samochodowym.
- **Poziom istotności (alfa):** 0,05

```
Kendall's rank correlation tau

data: dane$`transport samochodowy - ładunki` and dane$`transport kolejowy - ładunki`
z = 3.8893, p-value = 0.0001005
alternative hypothesis: true tau is not equal to 0
sample estimates:
tau
0.2099256
```

Rysunek 69 Wynik testu korelacji Kendalla

Test korelacji Kendalla wykazał, że wartość tau wynosząca 0,209 ($p = 0,0001005$) wskazuje na słabszy, lecz istotny statystycznie pozytywny związek między ilościami przewożonych ładunków transportem kolejowym i samochodowym. **Odrzucenie hipotezy zerowej na rzecz hipotezy alternatywnej** potwierdza istnienie korelacji, co sugeruje, że

wzrost przewozu ładunków w jednym rodzaju transportu wiąże się ze wzrostem w drugim rodzaju transportu.

6.4. Test korelacji Pearsona między przewozem pasażerów transportem kolejowym a samochodowym

Celem testu jest zbadanie stopnia związku pomiędzy liczbą pasażerów przewożonych transportem kolejowym a liczbą pasażerów przewożonych transportem samochodowym, aby zrozumieć, czy istnieje jakaś zależność między tymi dwoma rodzajami transportu.

Zostanie przeprowadzony test korelacji Pearsona pomiędzy liczbą pasażerów w transporcie kolejowym i liczbą pasażerów w transporcie samochodowym. Test ten umożliwi ocenę siły oraz kierunku ewentualnej zależności między tymi zmiennymi.

Wynik korelacji Pearsona może przyjąć wartości od -1 do 1.

- Wartość bliska 1 wskazuje na silną dodatnią korelację, czyli gdy jedna zmienna wzrasta, druga też rośnie.
- Wartość bliska -1 oznacza silną ujemną korelację, gdzie jedna zmienna wzrasta, a druga maleje.
- Wartość bliska零 oznacza brak korelacji między zmiennymi.

Hipotezy:

- **Hipoteza zerowa (H0):** Brak korelacji między liczbą pasażerów w transporcie samochodowym a transportem kolejowym.
- **Hipoteza alternatywna (H1):** Istnieje korelacja między liczbą pasażerów w transporcie samochodowym a transportem kolejowym.
- **Poziom istotności (alfa):** 0,05

```
432 #KORELACJA PEARSONA
433 # Przeprowadzenie testu korelacji Pearsona
434 wynik_korelacji <- cor.test(dane$"transport samochodowy - pasazerowie",
435                               dane$"transport kolejowy - pasazerowie", method = "pearson")
436
437 # Wyświetlenie wyników testu
438 print(wynik_korelacji)
```

Rysunek 70 Kod wykonujący test korelacji Pearsona

```

Pearson's product-moment correlation

data: dane$`transport samochodowy - pasazerowie` and dane$`transport kolejowy - pasazerowie`
t = 0.45639, df = 154, p-value = 0.6487
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.1210876 0.1927798
sample estimates:
cor
0.03675242

```

Rysunek 71 Wynik testu korelacji Pearsona

Wynik testu korelacji Pearsona sugeruje niewielką pozytywną korelację między liczbą pasażerów w transporcie samochodowym a transportem kolejowym. Wartość współczynnika korelacji wynosi około 0,0367, co wskazuje na bardzo słaby związek liniowy między tymi zmiennymi.

Dodatkowo, p-wartość, która wynosi 0,6487, jest znacznie większa od typowego poziomu istotności statystycznej (0,05). **Oznacza to, że nie ma wystarczających dowodów na odrzucenie hipotezy zerowej, czyli braku korelacji między liczbą pasażerów w transporcie samochodowym a transportem kolejowym.**

Wartość t-testu wynosi 0,45639, a stopnie swobody to 154. Przedział ufności dla korelacji mieści się w granicach od -0,1211 do 0,1928, co również potwierdza brak istotnej korelacji między tymi dwiema zmiennymi.

6.5. Test normalności rozkładu danych za pomocą testu Shapiro-Wilka dla przewozu ładunków różnymi rodzajami transportu

Test Shapiro-Wilka jest statystycznym testem używanym do oceny normalności rozkładu danych. Jest to test parametryczny, który pozwala stwierdzić, czy próbka pochodzi z populacji o rozkładzie normalnym.

```

643 ##### Test Shapiro-Wilka
644
645 #dla przewozu ładunków
646
647 #dla danych dotyczących transportu kolejowego
648 shapiro_test_kolejowy <- shapiro.test(dane$`transport kolejowy - ładunki`)
649 print("Test Shapiro-Wilka dla danych transportu kolejowego - pasażerowie:")
650 print(shapiro_test_kolejowy)
--
```

Rysunek 72 Kod wykonujący test Shapiro-Wilka

Hipotezy:

- **Hipoteza zerowa (H0):** Dane pochodzą z populacji o rozkładzie normalnym.
- **Hipoteza alternatywna (H1):** Dane nie pochodzą z populacji o rozkładzie normalnym.
- **Poziom istotności (alfa):** 0,05

Transport kolejowy:

```
Shapiro-Wilk normality test  
data: dane$`transport kolejowy - ładunki`  
W = 0.97036, p-value = 0.001931
```

Rysunek 73 Wynik testu Shapiro-Wilka dla transportu kolejowego

Test normalności Shapiro-Wilka dla danych dotyczących transportu kolejowego wykazał, że wartość p-value wynosi 0.001931, co oznacza istotną statystycznie różnicę od rozkładu normalnego na poziomie istotności 0.05. Wartość statystyki testowej (W) wynosi 0.97036. Zatem należy **odrzucić hipotezę o normalności rozkładu danych dotyczących transportu kolejowego**.

Transport samochodowy:

```
Shapiro-Wilk normality test  
data: dane$`transport samochodowy - ładunki`  
W = 0.94681, p-value = 1.237e-05
```

Rysunek 74 Wynik testu Shapiro-Wilka dla transportu samochodowego

Test normalności Shapiro-Wilka dla danych transportu samochodowego wykazał, że wartość p-value wynosi 1.237e-05, co jest znaczco mniejsze niż ustalony poziom istotności 0.05. Wartość statystyki testowej (W) wynosi 0.94681. **Odrzucenie hipotezy zerowej sugeruje, że dane dotyczące transportu samochodowego nie pochodzą z rozkładu normalnego.**

Transport morski:

```
Shapiro-Wilk normality test  
data: dane$`transport morski - ładunki`  
W = 0.98489, p-value = 0.08722
```

Rysunek 75 Wynik testu Shapiro-Wilka dla transportu morskiego

Test normalności Shapiro-Wilka dla danych dotyczących transportu morskiego wykazał, że wartość p-value wynosi 0,08722, co jest wyższe niż tradycyjny poziom istotności 0,05. Wartość statystyki testowej (W) wynosi 0,98489. Na podstawie tych wyników **nie ma podstaw do odrzucenia hipotezy zerowej**, co sugeruje, że dane transportu morskiego mogą pochodzić z rozkładu normalnego na poziomie istotności 0,05.

6.6. Test zgodności rozkładu danych z rozkładem normalnym, przy użyciu testu Jarque-Bera dla różnych rodzajów transportu pasażerskiego

Test Jarque-Bera to statystyczny test normalności, który bazuje na kołowej mierze spłaszczenia (kurtozie) i asymetrii danych. Jego celem jest sprawdzenie, czy dane liczbowe pochodzą z populacji o rozkładzie normalnym..

```
671 #dla danych dotyczących transportu kolejowego  
672 test_jarque_bera_kolejowy <- jarque.bera.test(dane$`transport kolejowy - pasazerowie`)  
673 print("Test Jarque-Bera dla danych transportu kolejowego:")  
674 print(test_jarque_bera_kolejowy)
```

Rysunek 76 Kod wykonujący test Jarque-Bera

Hipotezy:

- **Hipoteza zerowa (H0):** Badane dane pochodzą z rozkładu normalnego.
- **Hipoteza alternatywna (H1):** Badane dane nie pochodzą z rozkładu normalnego.
- **Poziom istotności (alfa):** 0,05

Transport kolejowy:

```
Jarque Bera Test  
data: dane$`transport kolejowy - pasazerowie`  
X-squared = 95.556, df = 2, p-value < 2.2e-16
```

Rysunek 77 Wyniki testu Jarque-Bera dla transportu kolejowego

- Wartość statystyki X-squared wynosi 95,556, z 2 stopniami swobody, i p-value mniejszym niż 2,2e-16.
- Wartość p-value jest znacząco mniejsza od przyjętego poziomu istotności ($\alpha = 0,05$), co oznacza **odrzucenie hipotezy zerowej**. Możemy stwierdzić, że dane dotyczące transportu kolejowego **nie pochodzą z rozkładu normalnego**.

Transport samochodowy:

```
Jarque Bera Test

data: dane$`transport samochodowy - pasazerowie`
X-squared = 5.5314, df = 2, p-value = 0.06293
```

Rysunek 78 Wyniki testu Jarque-Bera dla transportu samochodowego

- Wartość statystyki X-squared wynosi 5,5314, z 2 stopniami swobody, i p-value równym 0,06293.
- Wartość p-value (0,06293) jest większa od przyjętego poziomu istotności ($\alpha = 0,05$), co sugeruje **brak istotnych dowodów na odrzucenie hipotezy zerowej**. Istnieje słaba istotność statystyczna, sugerująca, że dane dotyczące transportu samochodowego **mogą pochodzić z rozkładu normalnego**.

Transport lotniczy:

```
Jarque Bera Test

data: dane$`transport lotniczy - pasazerowie`
X-squared = 162.33, df = 2, p-value < 2.2e-16
```

Rysunek 79 Wyniki testu Jarque-Bera dla transportu lotniczego

- Wartość statystyki X-squared wynosi 162,33, z 2 stopniami swobody, i p-value mniejszym niż 2,2e-16.
- Wartość p-value jest znacząco mniejsza od przyjętego poziomu istotności ($\alpha = 0,05$), co oznacza **odrzucenie hipotezy zerowej**. Dane dotyczące transportu lotniczego nie pochodzą z rozkładu normalnego.

Podsumowując, dla transportu kolejowego i lotniczego **odrzucamy hipotezę zerową o zgodności z rozkładem normalnym**, natomiast dla transportu samochodowego **nie mamy silnych dowodów na jej odrzucenie**, co sugeruje, że dane te mogą pochodzić z rozkładu normalnego.

6.7. Test chi-kwadrat w analizie proporcji między transportem kolejowym a samochodowym w dwóch okresach

Test chi-kwadrat ma na celu sprawdzenie, czy proporcje między przewozem ładunków transportem kolejowym a samochodowym różnią się w dwóch różnych okresach czasu (pierwsza połowa danych vs. druga połowa danych). Wykorzystano test chi-kwadrat, który porównuje dwie macierze kontyngencji (liczbę obserwacji w poszczególnych kategoriach). Test ten zwraca wartość p-value, która wskazuje na istotność statystyczną różnic w proporcjach między dwoma grupami czasowymi. Niska wartość p wskazuje na znaczącą różnicę.

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszy): 2010-01 – 2016-06
- Okres drugi (późniejszy): 2016-07 – 2022-12

```
688 #####Test chi-kwadrat
689
690 #dane
691 transport_kolejowy_ladunki <- dane$`transport kolejowy - ladunki`
692 transport_samochodowy_ladunki <- dane$`transport samochodowy - ladunki`
693
694 # Podział danych na dwie grupy czasowe (pierwsza połowa, druga połowa)
695 polowa_pierwsza <- length(transport_kolejowy_ladunki) / 2
696
697 transport_kolejowy_1 <- transport_kolejowy_ladunki[1:polowa_pierwsza]
698 transport_samochodowy_1 <- transport_samochodowy_ladunki[1:polowa_pierwsza]
699
700 transport_kolejowy_2 <- transport_kolejowy_ladunki[(polowa_pierwsza + 1):length(transport_kolejowy_ladunki)]
701 transport_samochodowy_2 <- transport_samochodowy_ladunki[(polowa_pierwsza + 1):length(transport_samochodowy_ladunki)]
702
703 # Przeprowadzenie testu chi-kwadrat dla dwóch grup czasowych
704 result <- chisq.test(rbind(
705   table(transport_kolejowy_1, transport_samochodowy_1),
706   table(transport_kolejowy_2, transport_samochodowy_2)
707 ))
```

Rysunek 80 Kod wykonujący test chi-kwadrat

Hipotezy:

- **Hipoteza zerowa (H0):** Proporcje przewozów ładunków transportem kolejowym i samochodowym w obu okresach czasowych są takie same.
- **Hipoteza alternatywna (H1):** Proporcje przewozów ładunków transportem kolejowym i samochodowym różnią się w dwóch różnych okresach czasowych.
- **Poziom istotności (alfa):** 0,05

Pearson's Chi-squared test

```
data: rbind(table(transport_kolejowy_1, transport_samochodowy_1), table(transport_kolejowy_2,
transport_samochodowy_2))
X-squared = 12012, df = 11935, p-value = 0.308
```

Rysunek 81 Wyniku testu chi-kwadrat

Na podstawie uzyskanej wartości p (0,308), która jest większa od ustalonego poziomu istotności (0,05), **nie mamy wystarczających dowodów na odrzucenie hipotezy zerowej**. Oznacza to, że nie ma istotnej statystycznie różnicy w proporcjach przewozów ładunków transportem kolejowym i samochodowym pomiędzy dwoma okresami czasowymi. Można przyjąć, że proporcje te pozostały zbliżone w obu okresach.

6.8. Test chi-kwadrat w analizie proporcji przewozów transportem samochodowym w dwóch okresach

Przeprowadzono test chi-kwadrat, aby ocenić ewolucję proporcji transportu pasażerów transportem samochodowym w dwóch różnych okresach. Dane dotyczące przewozów transportem samochodowym zostały podzielone na dwie grupy, reprezentujące dwa okresy czasowe. Test statystyczny porównuje proporcje tego rodzaju transportu między tymi okresami czasowymi, co pozwala na zrozumienie ewolucji i ewentualnych zmian w proporcjach tego rodzaju transportu w badanych okresach.

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszego): 2010-01 – 2016-06
- Okres drugi (późniejszego): 2016-07 – 2022-12

Hipotezy:

- **Hipoteza zerowa (H0):** Brak zmian w proporcjach transportu samochodowego między dwoma okresami czasowymi.
- **Hipoteza alternatywna (H1):** Istnieją istotne różnice w proporcjach transportu samochodowego między dwoma okresami czasowymi.
- **Poziom istotności (alfa):** 0,05

Pearson's Chi-squared test

```
data: table(transport_samochodowy_1, transport_samochodowy_2)
X-squared = 6006, df = 5929, p-value = 0.2388
```

Rysunek 82 Wynik testu chi-kwadrat

Wartość statystyki chi-kwadrat wynosi 6006 przy 5929 stopniach swobody, a p-wartość jest równa 0,2388. Wartość p-wartości wynosząca 0,2388 jest większa niż typowy poziom istotności statystycznej (np. 0,05). Oznacza to, że **brak jest wystarczających dowodów na odrzucenie hipotezy zerowej**. Nie ma wystarczających dowodów, aby

stwierdzić, że proporcje transportu samochodowego różnią się istotnie między dwoma okresami czasowymi.

6.9. Test regresji liniowej między ilością przewozów ładunków transportem kolejowym, a ilością przewozów ładunków transportem samochodowym

Test regresji, wykorzystujący analizę regresji liniowej, pozwala na zbadanie związku między dwiema zmiennymi, takimi jak ilość przewozów transportem kolejowym i ilość przewozów transportem samochodowym. Celem tego testu jest zrozumienie, czy wzrost lub spadek przewozów kolejowych wpływa na zmiany w przewozach samochodowych, oraz określenie, jak mocna jest ta zależność. Test ten pozwala na prognozowanie zmian w ilości przewozów samochodowych w oparciu o ilość przewozów kolejowych oraz na zrozumienie, czy istnieje statystycznie istotna zależność między tymi zmiennymi.

```
741 # Test regresji
742 # Przygotowanie danych
743 transport_kolejowy <- dane$`transport kolejowy - ładunki`
744 transport_samochodowy <- dane$`transport samochodowy - ładunki`
745
746 # Przeprowadzenie testu regresji
747 regresja <- lm(transport_samochodowy ~ transport_kolejowy)
748
749 # Wyświetlenie wyników testu
750 summary(regresja)
```

Rysunek 83 Kod wykonujący test regresji

Hipotezy:

- **Hipoteza zerowa (H0):** Brak istotnej statystycznie zależności między ilością przewozów ładunków transportem kolejowym a ilością przewozów ładunków transportem samochodowym.
- **Hipoteza alternatywna (H1):** Istnieje istotna statystycznie zależność między ilością przewozów ładunków transportem kolejowym a ilością przewozów ładunków transportem samochodowym.
- **Poziom istotności (alfa):** 0,05

```

> summary(regresja)

Call:
lm(formula = transport_samochodowy ~ transport_kolejowy)

Residuals:
    Min      1Q  Median      3Q     Max 
 -8381  -4527   396   4401  8387 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -331.9986  4685.1054 -0.071   0.944    
transport_kolejowy  0.9803    0.2413   4.063 7.7e-05 ***  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 4775 on 154 degrees of freedom
Multiple R-squared:  0.09681, Adjusted R-squared:  0.09095 
F-statistic: 16.51 on 1 and 154 DF,  p-value: 7.701e-05

```

Rysunek 84 Wynik testu regresji

- Residuals (Reszty):
 - Zbiór reszt (różnica między wartościami rzeczywistymi a przewidywanymi) pokazuje odchylenia wartości prognozowanych od rzeczywistych, przy czym minimalna wartość wynosi -8381, a maksymalna 8387.
- Coefficients (Współczynniki):
 - Wyraz wolny (-331,9986) to przecięcie się linii regresji z osią Y (transport samochodowy) w punkcie, gdy transport kolejowy wynosi 0, nie jest statystycznie istotny ($p > 0,05$), co oznacza, że dla wartości zerowych przewozów kolejowych nie możemy wnioskować o wartości przewozów samochodowych
 - Współczynnik dla transportu kolejowego (0,9803) oznacza, że każdy dodatkowy wzrost o jednostkę transportu kolejowego przewiduje wzrost transportu samochodowego o około 0,9803 jednostki.
- Residual standard error (Błąd standardowy reszt):
 - 4775 oznacza średni błąd prognoz, co sugeruje, że średnie odchylenie między danymi rzeczywistymi a przewidywanymi wynosi 4775.
- Multiple R-squared (R-kwadrat wielokrotny):
 - Wartość 0,09681 oznacza, że tylko około 9.68% zmienności transportu samochodowego można wyjaśnić przez transport kolejowy za pomocą modelu regresji liniowej.

- Adjusted R-squared (Skorygowany R-kwadrat):
 - Skorygowany R-kwadrat uwzględnia ilość zmiennych w modelu, w tym przypadku jest bardzo zbliżony do R-kwadrat, sugerując, że dodanie kolejnych zmiennych nie poprawiłoby modelu.
- F-statistic (Statystyka F):
 - Statystyka F wynosząca 16,51 jest wynikiem testu istotności regresji, co oznacza, że istnieje statystycznie istotny związek między przewozami kolejowymi a samochodowymi.
- p-value:
 - Bardzo niski wynik p-value (7,701e-05) oznacza, że mamy wystarczająco mocne dowody na odrzucenie hipotezy zerowej (brak związku między przewozami kolejowymi a samochodowymi) na rzecz alternatywnej hipotezy (istnieje związek między nimi).

W kontekście poziomu istotności alfa (0,05), **odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej**. Istnieje istotna statystycznie zależność między ilością przewozów ładunków transportem kolejowym a ilością przewozów ładunków transportem samochodowym.

Możemy stwierdzić, że ilość przewozów kolejowych ma istotny wpływ na ilość przewozów samochodowych w przypadku przewozu ładunków.

6.10. Test Wilcoxona-Manna-Whitneya dla danych dotyczących przewozów pasażerów transportem lotniczym w dwóch okresach

Test Wilcoxona-Manna-Whitneya, często zwany testem sumy rang, jest nieparametrycznym testem porównawczym, który sprawdza, czy dwie niezależne próbki pochodzą z tych samych populacji (czy pochodzą z tego samego rozkładu).

Wybór testu Wilcoxona-Manna-Whitneya dla porównania ilości przewozów lotniczych między dwoma okresami wynika z tego, że dane mogą nie być zgodne z założeniami testów parametrycznych. Przykładowo, w przypadku braku normalności rozkładu danych lub w obecności skrajnych wartości odstających, testy parametryczne mogą nie być adekwatne. Test Wilcoxona-Manna-Whitneya jest mniej wrażliwy na nienormalność danych i bardziej odporny na odstające wartości.

Cel wykonania testu Wilcoxona-Manna-Whitneya dla danych dotyczących przewozów lotniczych między dwoma okresami (wcześniejszym i późniejszym) polega na zbadaniu ewentualnych istotnych statystycznie różnic pomiędzy ilościami przewozów

lotniczych w tych dwóch okresach. Analiza ta ma na celu ustalenie, czy istnieje istotna różnica w przewozach lotniczych pomiędzy badanymi okresami czasowymi.

Dane zostały podzielone na dwa okresy:

- Okres pierwszy (wcześniejszy): 2010-01 – 2016-06
- Okres drugi (późniejszy): 2016-07 – 2022-12

6.10.1. Test normalności rozkładu danych transportu lotniczego

Dane zostaną poddane testowi normalności rozkładu z wykorzystaniem testu Shapiro-Wilka.

Wynik tego testu pozwoli określić, czy dane w obu okresach mają rozkład zbliżony do rozkładu normalnego.

Hipotezy:

- **Hipoteza zerowa (H0):** Dane dotyczące przewozów pasażerów transportem lotniczym w obu okresach pochodzą z populacji o rozkładzie normalnym.
- **Hipoteza alternatywna (H1):** Dane dotyczące przewozów pasażerów transportem lotniczym w obu okresach nie pochodzą z populacji o rozkładzie normalnym.
- **Poziom istotności (alfa):** 0,05

```
> # Test Shapiro-Wilka dla danych transportu lotniczego
> shapiro.test(dane$`transport lotniczy - pasazerowie`)

Shapiro-Wilk normality test

data: dane$`transport lotniczy - pasazerowie`
W = 0.84372, p-value = 1.317e-11
```

Rysunek 85 Wynik tesu Shapiro-Wilka

Wynik testu normalności Shapiro-Wilka dla danych dotyczących przewozów pasażerów transportem lotniczym wykazuje następujące wartości statystyczne:

- Wartość statystyki testowej (W) wynosi 0,84372.
- P-wartość (p-value) jest bardzo niska i wynosi 1,317e-11, co jest znaczająco mniejsze niż powszechnie przyjęty poziom istotności 0,05.

Wartość p-value (<0.001) jest znaczająco niższa od ustalonego poziomu istotności, co prowadzi do **odrzucenia hipotezy zerowej (H0) o normalności rozkładu danych**. Możemy stwierdzić, że dane dotyczące przewozów pasażerów transportem lotniczym prawdopodobnie

nie pochodzą z populacji o rozkładzie normalnym. Wartość p-value poniżej ustalonego poziomu istotności świadczy o silnej niezgodności danych z rozkładem normalnym. Oznacza to, że dalsze statystyczne analizy zakładające normalność rozkładu tych danych mogą być nieodpowiednie lub wymagałyby innego podejścia do analizy, dlatego dobrym podejściem jest wykonanie testu nieparametrycznego Wilcoxona-Manna-Whitneya.

6.10.2. Test Wilcoxona-Manna-Whitneya

Celem jest zbadanie ewentualnej różnicy w ilości przewozów pasażerów transportem lotniczym pomiędzy dwoma okresami czasowymi. Użycie testu Wilcoxona-Manna-Whitneya pozwala na porównanie median ilości przewozów między tymi okresami, nawet jeśli dane nie są z rozkładu normalnego.

Hipotezy:

- **Hipoteza zerowa (H0):** Brak istotnej statystycznie różnicy w ilości przewozów pasażerów transportem lotniczym między okresem wcześniejszym (2010-01 do 2016-06) a późniejszym (2016-07 do 2022-12).
- **Hipoteza alternatywna (H1):** Istnieje istotna statystycznie różnica w ilości przewozów pasażerów transportem lotniczym między okresem wcześniejszym a późniejszym.

```
761 # wydzielenie danych dla okresu wcześniejszego i późniejszego
762 dane_wczesniejszy <- dane$`transport lotniczy - pasazerowie`[1:78] # Dane dla okresu wcześniejszego
763 dane_pozniejszy <- dane$`transport lotniczy - pasazerowie`[79:length(dane$`transport lotniczy - pasazerowie`)]
764
765 # Przeprowadzenie testu Wilcoxona-Manna-Whitneya
766 wynik_testu <- wilcox.test(dane_wczesniejszy, dane_pozniejszy)
767
768 # Wyświetlenie wyniku testu
769 print(wynik_testu)
```

Rysunek 86 Kod wykonujący test Wilcoxona-Manna-Whitneya

```
> print(wynik_testu)

Wilcoxon rank sum test with continuity correction

data: dane_wczesniejszy and dane_pozniejszy
W = 1450.5, p-value = 1.708e-08
alternative hypothesis: true location shift is not equal to 0
```

Rysunek 87 Wynik testu Wilcoxona-Manna-Whitneya

Wynik testu:

- Wartość W = 1450,5
- Wartość p-value = 1,708e-08

Interpretacja:

Wartość $W = 1450,5$ odnosi się do statystyki testowej W . Wartość ta mierzy siłę zależności rang między dwiema próbками. Im wyższa wartość W , tym większa różnica w rangach między próbками.

Wartość p-value ($1,708\text{e-}08$) jest znaczco mniejsza od ustalonego poziomu istotności ($0,05$), co oznacza **odrzucenie hipotezy zerowej**. Odrzucenie hipotezy zerowej wskazuje na istnienie istotnej statystycznie różnicy w ilości przewozów pasażerów transportem lotniczym między okresem wcześniejszym (2010-01 do 2016-06) a późniejszym (2016-07 do 2022-12).

Literatura

1. Biecek P., *Wprowadzenie do analizy danych w R*, Wydawnictwo Naukowe PWN, Warszawa 2019
2. Murrell P., *R Graphics*, Chapman & Hall/CRC, 2006
3. *R for Data Science*, Wickham H., Grolemund G., O'Reilly Media, Inc., 2017
4. Verzani, J., *SimpleR - Using R for Introductory Statistics*, Chapman and Hall/CRC Press, 2004
5. Wickham H., *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York, 2016

Spis rysunków

Rysunek 1 Fragment danych pliku Excel.....	7
Rysunek 2 Kod instalacji i ładowania bibliotek	9
Rysunek 3 Kod wczytywania pliku danych	10
Rysunek 4 Kod przypisujący nazwy do kolumn	10
Rysunek 5 Fragment wczytanych danych	10
Rysunek 6 Funkcja wyznaczająca wartość maksymalną	11
Rysunek 7 Wyniki kodu dla wartości maksymalnej	11
Rysunek 8 Funkcja wyznaczająca wartość minimalną	11
Rysunek 9 Wyniki kodu dla wartości minimalnej	12
Rysunek 10 Funkcja wyznaczająca odchylenie standardowe	12
Rysunek 11 Wyniki kodu dla odchylenia standardowego	12
Rysunek 12 Funkcja wyznaczająca średnią	13
Rysunek 13 Wyniki kodu dla średniej.....	13
Rysunek 14 Funkcja wyznaczająca medianę	13
Rysunek 15 Wyniki kodu dla mediany	14
Rysunek 16 Funkcja obliczająca kwartyle	14
Rysunek 17 Wyniki kodu dla kwartyli.....	14
Rysunek 18 Funkcja obliczająca rozstęp międzykwartylowy	15
Rysunek 19 Wyniki kodu dla rozstępu międzykwartylowego	16
Rysunek 20 Funkcja obliczająca wariancję.....	16
Rysunek 21 Wyniki kodu dla wariancji	16
Rysunek 22 Funkcja obliczająca współczynnik korelacji	17
Rysunek 23 Wyniki dla współczynnika korelacji	17
Rysunek 24 Funkcja obliczająca kowariancję.....	18
Rysunek 25 Wyniki kodu dla kowariancji	18
Rysunek 26 Wyniki kodu dla wartości maksymalnej	19
Rysunek 27 Wyniki kodu dla wartości minimalnej	19
Rysunek 28 Wyniki kodu dla odchylenia standardowego	20
Rysunek 29 Wyniki kodu dla średniej.....	20
Rysunek 30 Wyniki kodu dla mediany	21
Rysunek 31 Wyniki kodu dla kwartyli.....	21
Rysunek 32 Wyniki kodu dla rozstępu międzykwartylowego	22
Rysunek 33 Wyniki kodu dla wariancji	23

Rysunek 34 Wyniki kodu dla współczynnika korelacji	23
Rysunek 35 Wyniki kodu dla kowariancji	24
Rysunek 36 Wykres kołowy przewozów ładunków	26
Rysunek 37 Kod generujący wykres kołowy	26
Rysunek 38 Wykres kołowy przewozów pasażerów	27
Rysunek 39 Wykres liniowy przewozów ładunków i pasażerów	28
Rysunek 40 Kod generujący wykres liniowy	29
Rysunek 41 Wykres liniowy przewozów ładunków z podziałem na rodzaj transportu	30
Rysunek 42 Wykres liniowy przewozów pasażerów z podziałem na rodzaj transportu	31

Rysunek 43 Histogram dla przewozów ładunków	33
Rysunek 44 Kod generujący histogram.....	33
Rysunek 45 Histogram przewozów pasażerów	34
Rysunek 46 Wykres słupkowy przewozów ładunków	35
Rysunek 47 Kod generujący wykres słupkowy.....	36
Rysunek 48 Wykres słupkowy przewozów pasażerów.....	37
Rysunek 49 Wykres punktowy przewozów pasażerów	38
Rysunek 50 Kod generujący przewozy pasażerów	39
Rysunek 51 Wykres punktowy przewozów ładunków	39
Rysunek 52 Wykres pudełkowy przewozów ładunków.....	40
Rysunek 53 Wykres pudełkowy przewozu ładunków transportem morskim	42
Rysunek 54 Kod generujący wykres pudełkowy	43
Rysunek 55 Wykres gęstości przewozów ładunków transportem kolejowym	44
Rysunek 56 Kod generujący wykres gęstości	45
Rysunek 57 Wykres skrzypcowy przewozów pasażerów dla transportu	
samochodowego	46

Rysunek 58 Kod generujący wykres skrzypcowy.....	47
Rysunek 59 Mapa cieplna korelacji dla przewozu ładunków	48
Rysunek 60 Mapa cieplna korelacji dla przewozu pasażerów	50
Rysunek 61 Kod generujący mapy cieplne	51
Rysunek 62 Kod wykonujący test t-Studenta.....	52
Rysunek 63 Wynik kodu testu t-Studenta	52
Rysunek 64 Kod wykonujący test t-Studenta.....	53
Rysunek 65 Wynik testu t-Studenta	54

Rysunek 66 Kod wykonujący metody testów korelacji: Pearsona, Spearmana i Kendalla	55
Rysunek 67 Wynik testu korelacji Pearsona	55
Rysunek 68 Wynik testu korelacji Spearmana.....	56
Rysunek 69 Wynik testu korelacji Kendalla	56
Rysunek 70 Kod wykonujący test korelacji Pearsona.....	57
Rysunek 71 Wynik testu korelacji Pearsona	58
Rysunek 72 Kod wykonujący test Shapiro-Wilka.....	58
Rysunek 73 Wynik testu Shapiro-Wilka dla transportu kolejowego	59
Rysunek 74 Wynik testu Shapiro-Wilka dla transportu samochodowego	59
Rysunek 75 Wynik testu Shapiro-Wilka dla transportu morskiego	60
Rysunek 76 Kod wykonujący test Jarque-Bera.....	60
Rysunek 77 Wyniki testu Jarque-Bera dla transportu kolejowego	60
Rysunek 78 Wyniki testu Jarque-Bera dla transportu samochodowego	61
Rysunek 79 Wyniki testu Jarque-Bera dla transportu lotniczego	61
Rysunek 80 Kod wykonujący test chi-kwadrat	62
Rysunek 81 Wyniku testu chi-kwadrat.....	62
Rysunek 82 Wynik testu chi-kwadrat.....	63
Rysunek 83 Kod wykonujący test regresji	64
Rysunek 84 Wynik testu regresji.....	65
Rysunek 85 Wynik tesu Shapiro-Wilka	67
Rysunek 86 Kod wykonujący test Wilcoxona-Manna-Whitneya	68
Rysunek 87 Wynik testu Wilcoxona-Manna-Whitneya.....	68

Spis tabel

Tabela 1 Podsumowanie podstawowych parametrów opisowych cz.1	25
Tabela 2 Podsumowanie podstawowych parametrów opisowych cz.2.....	25