Analiza zawartości CO2 i NO2 w powietrzu

Sofiya Rylova & Ewa Podlodowska

Etapy projektu

1 Zawartość dwutlenku węgla w powietrzu

1.1 Dane

Pierwszym tematem naszego projektu będzie opracowanie "Zestawu danych miesięcznych stężeń CO2".

Ten zestaw danych zawiera wybrane średnie miesięczne stężenia CO2 w Obserwatorium Mauna Loa w latach 1974-1987. Stężenia CO2 mierzono ciągłym analizatorem podczerwieni działu Geofizycznego Monitoringu Zmian Klimatu Laboratorium Zasobów Powietrza NOAA. Wybór miał na celu przybliżenie „warunków tła".

Ten zestaw danych otrzymano od Jima Elkinsa z NOAA w 1988 roku.

Każda linia zawiera stężenie CO2 (stosunek zmieszania w suchym powietrzu, wyrażony w skali ułamków molowych WMO X85, utrzymywanej przez "Scripps Institution of Oceanography"). Ponadto zawiera rok, miesiąc i wartość liczbową dla połączonego miesiąca i roku.

```r
library(rio)
dane <- read.table("C:/Users/nice2/Desktop/projekt_szeregi/dane.txt", quote="\"", comment.
dane <- dane[,-3]
```

```r
miesiac <- c("styczeń", "luty", "marzec", "kwiecień", "maj", "czerwiec", "lipiec", "sierpi

lata <- NULL

for(i in 1:14){
  lata <- c(lata, rep(1973+i, 12))
}
```

```r
colnames(dane) <- c("wartość", "lata", "miesiąc")
```

```r
knitr::kable(head(dane))
```

wartość	lata	miesiąc
333.13	1974.38	5
332.09	1974.46	6
331.10	1974.54	7
329.14	1974.63	8
327.36	1974.71	9
327.29	1974.79	10

1.2 Analiza danych za pomocą podstawowych statystyk:

```
knitr::kable(summary(dane[1]))
```

wartość
Min. :327.3
1st Qu.:334.1
Median :339.2
Mean :339.1
3rd Qu.:344.1
Max. :351.7

Z zestwienia widać, że najmniejsze stężenie CO2 wynosi 327.3, a największe 345.9.

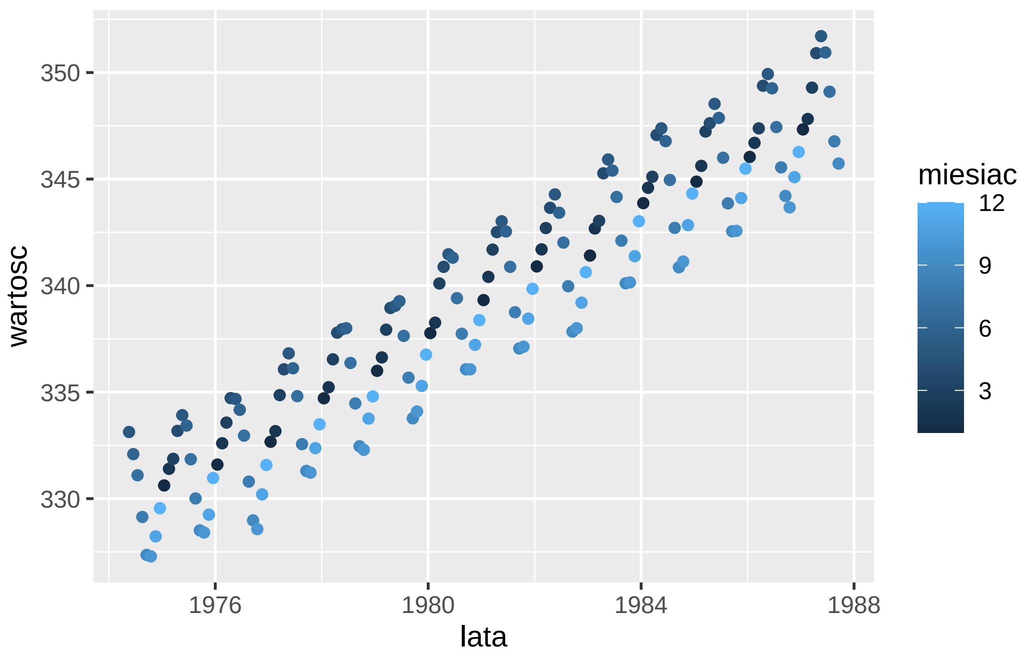
Warto dodać, że obecnie w naturalnym środowisku stężenie CO2 w powietrzu zwykle nie przekracza 400 ppm (0,04 %) objętości i takie stężenie jest najkorzystniejsze dla oddychającego człowieka.

Za próg bezpieczeństwa podczas 8-godzinnego dnia pracy przyjmuje się stężenie CO2 równe 5000 ppm. Jest to jednak próg bezpieczeństwa, a nie komfortu i wpływu na zdrowie. Narzekania na jakość powietrza z reguły pojawiają się w sytuacji w której stężenie CO2 przekracza 600-800 ppm, a nasilają powyżej 1000 ppm.

Dlatego tak ważna analiza CO2 w naturalnym średowisku i również w powieszczeniach.

1.3 Wizualizacja

```
library(tidyverse)
dane %>%
  ggplot(aes(x = lata, y = wartość, col = miesiąc))+
  geom_point()
```



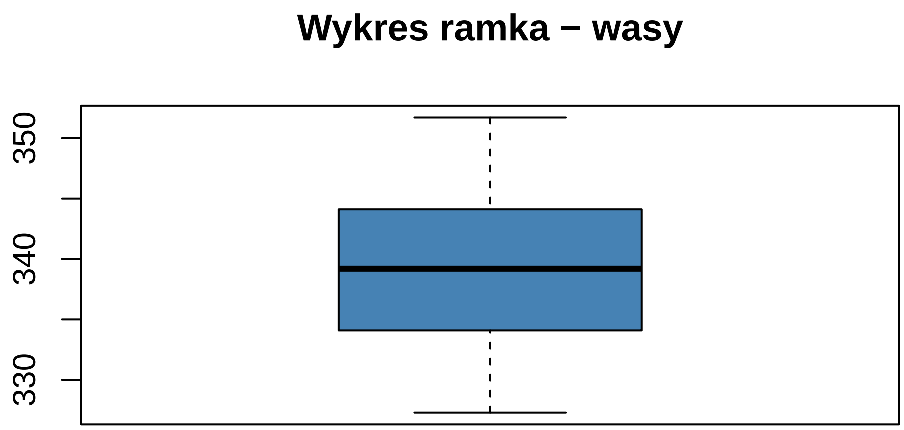
Widzimy zależnosc liniowową dodatnią między stężeniem CO2 i rokiem mierzenia stężenia.

Teraz przechodzimy do wykrycia sezonowości, w tym celu dokonamy przekształcenia ramki danych.

```
dane2 <- read.table("C:/Users/nice2/Desktop/projekt_szeregi/dane2.txt", quote="\"", commen
```

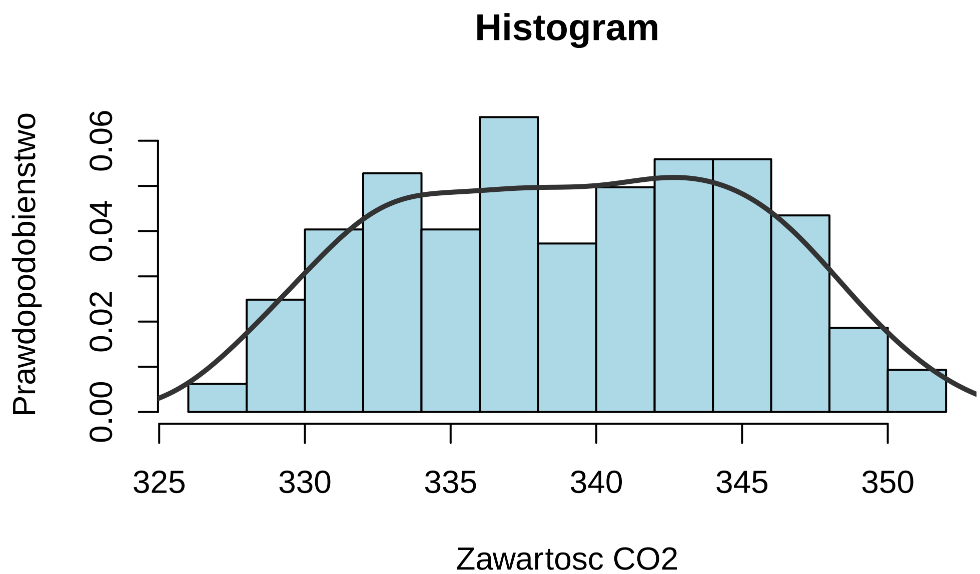
```
miesiac <- c("styczeń", "luty", "marzec", "kwiecień", "maj", "czerwiec", "lipiec", "sierpi

rownames(dane2) <- paste(rep(miesiac, 9), dane2[,3])

dane2 <- dane2[1]
colnames(dane2) <- "wartość"
```

```
boxplot(dane$wartość, col = "steelblue", main = "Wykres ramka - wąsy")
```



Rozkład cechuje się symetrią.

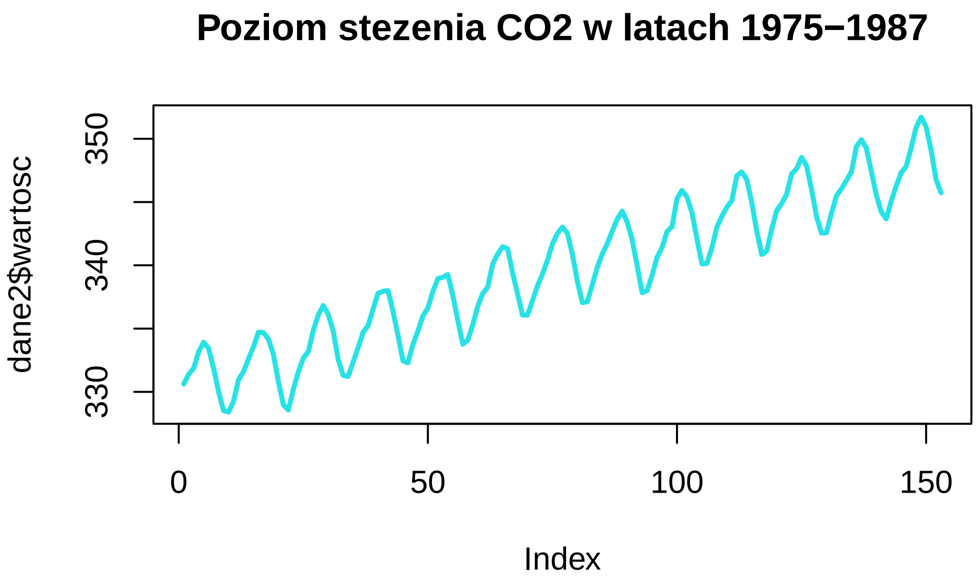
```
hist(dane$wartość, breaks = 9, col = "lightblue", main="Histogram",
xlab="Zawartość CO2", ylab="Prawdopodobieństwo", prob=T)
gestosc <- density(dane$wartość)
lines(x=gestosc$x, y=gestosc$y, col="grey20", lwd=2.5)
```



Rozkład nie jest symetryczny, występuje delikatna dualność.

Spróbujemy wykryć sezonowość za pomocą wykresów:

```
plot(dane2$wartość, col = 5, type = "l", main = "Poziom stężenia CO2 w latach 1975-1987",
```

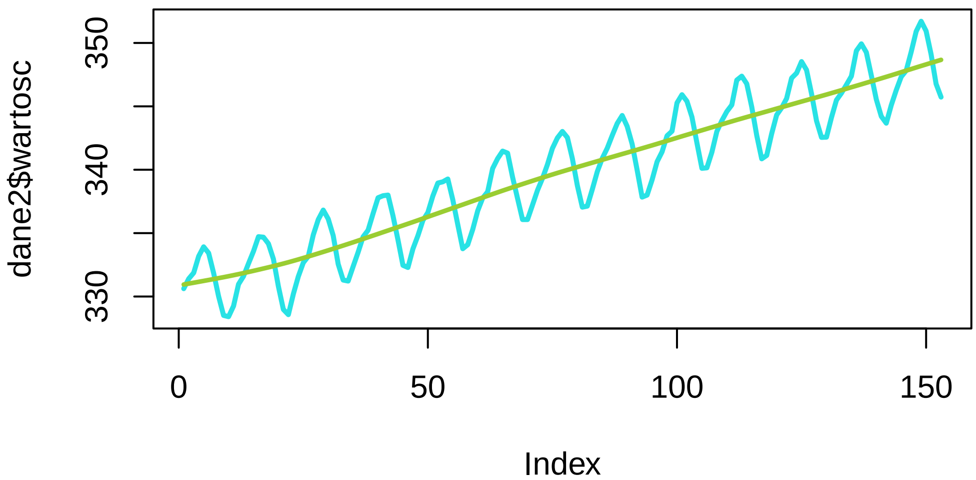


Możemy przepuszczać, że wahanie sezonowe mamy typu "addytewnego", nie będziemy tego wnioskować na podstawie jednego wykresu.

W środowisku R dostępne są także funkcje dotyczące filtrowania szeregów czasowych. Zajmiemy się tym w następnej kolejności.

Ponieważ mamy dane miesięczne, zaleca się stosowanie współczynnika `lambda` = 14400.

```
f <- FRAPO::trdhp(dane2$wartość, lambda=14400)
plot(dane2$wartość, col = 5, type = "l", main = "", lwd = 2.5)
lines(f,col="YellowGreen", lwd = 2.3)
```

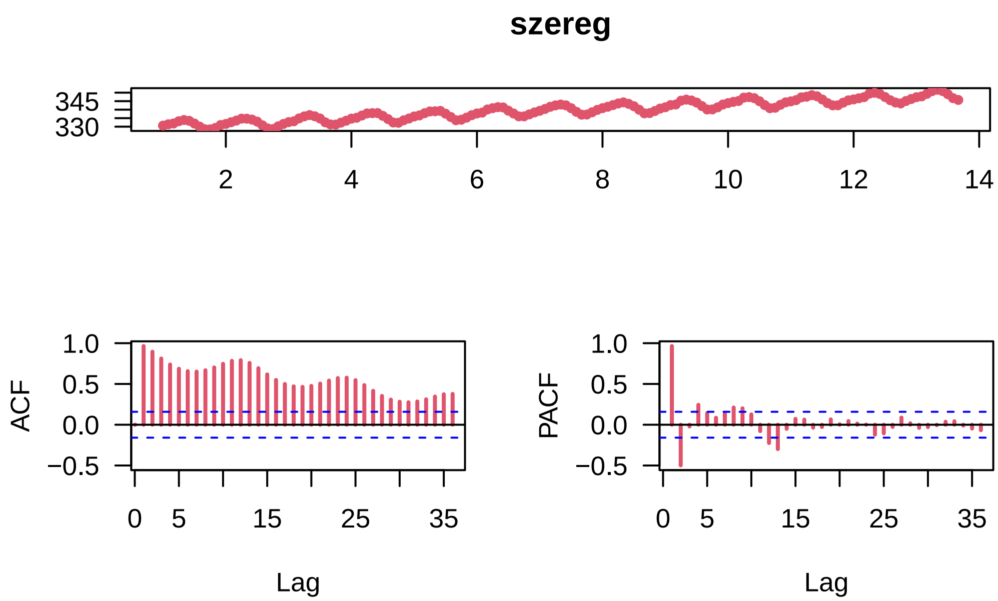


1.4 Identyfikacja trendu i sezonowości

1.4.1 Autokorelacja:

```
szereg <- ts(dane2$wartość, frequency = 12)
forecast::tsdisplay(szereg,col=2,lwd=2,las=1)
```

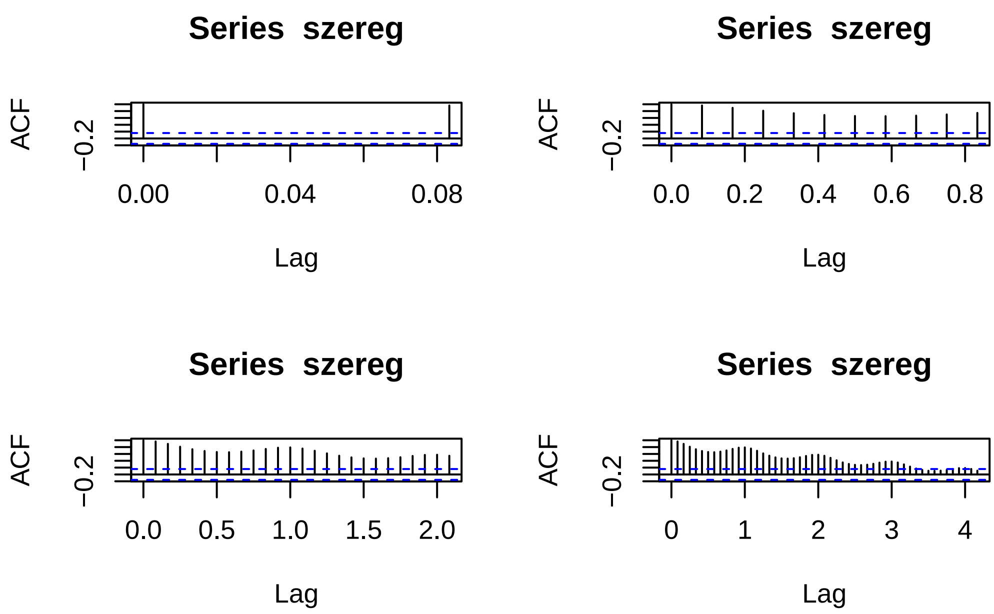
```
Registered S3 method overwritten by 'quantmod':
  method            from
  as.zoo.data.frame zoo
```



Na podstawie trzech wykresów: krzywej badanego zjawiska, funkcji autokorelacji ACF oraz funckcji autokorelacji PACF trudno nam powiedzieć o instnieniu trendu, iż funkcja ACF maleję wykładniczo wraz ze wzrostem parametru p.

Również możemy skorzystać z funkcji `forecast::Acf()`:

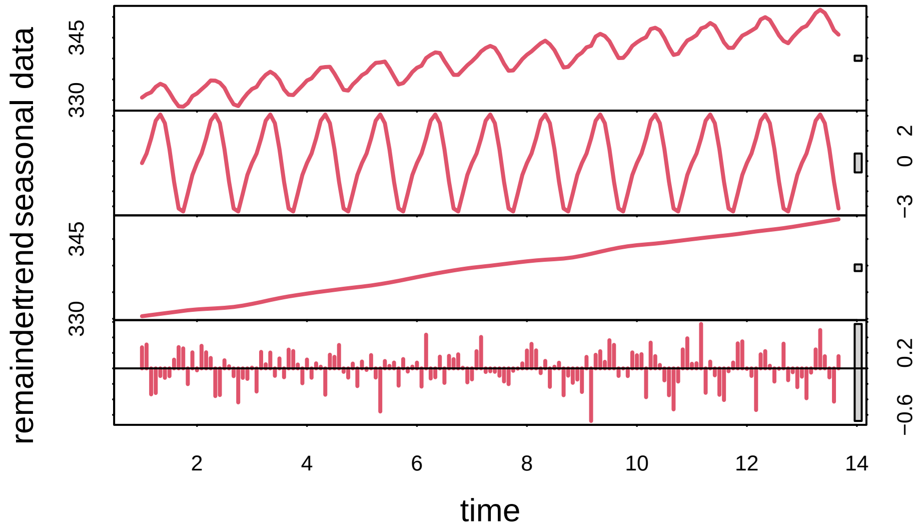
```
par(mfrow = c(2, 2))
acf(x = szereg, lag.max = 1, type = "correlation")
acf(x = szereg, lag.max = 10, type = "correlation")
acf(x = szereg, lag.max = 25, type = "correlation")
acf(x = szereg, lag.max = 50, type = "correlation")
```




```r
par(mfrow = c(1, 1))
```

Wykresy przedstawiają funkcję autokorelacji odpowiednie dla $\tau = \{1, 10, 25, 50\}$

```r
plot(stl(szereg,s.window="periodic"),col=2,lwd=2)
```



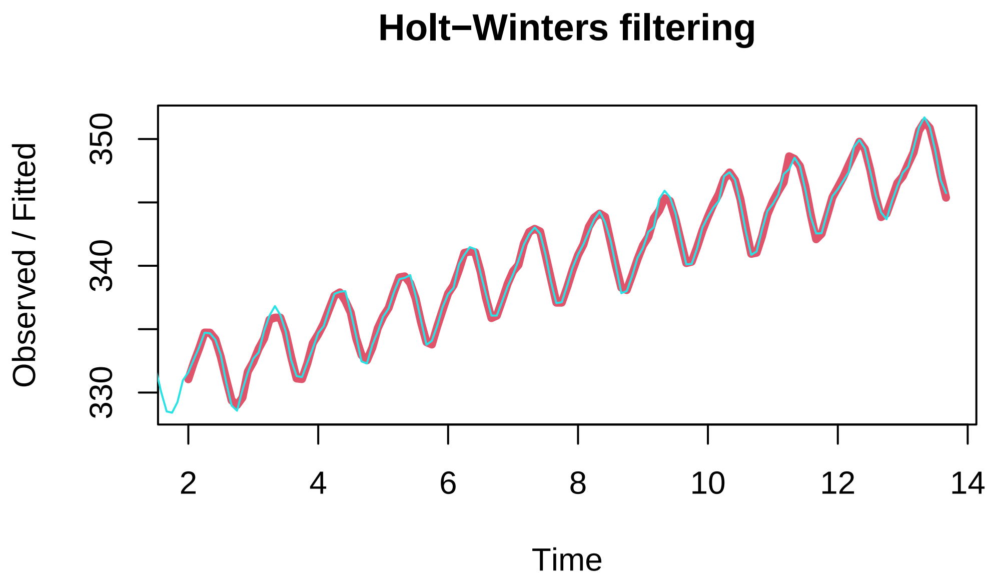
Na podstawie powyższych wykresów można

Wywnioskować, że nasze dane charakretyzują się sezonowością i pewnym trendem.

1.5 Predykcja

Aby przeprowadzić predykcję potrzebujemy zbudować model, decydujemy się na model Holt'a-Winters'a:

```
model <- HoltWinters(szereg)
plot(model, lwd = 4, col = 5)
```



Czerwona linia - dopasowane wartości, widzimy, że prawie idealnie się pokrywają.

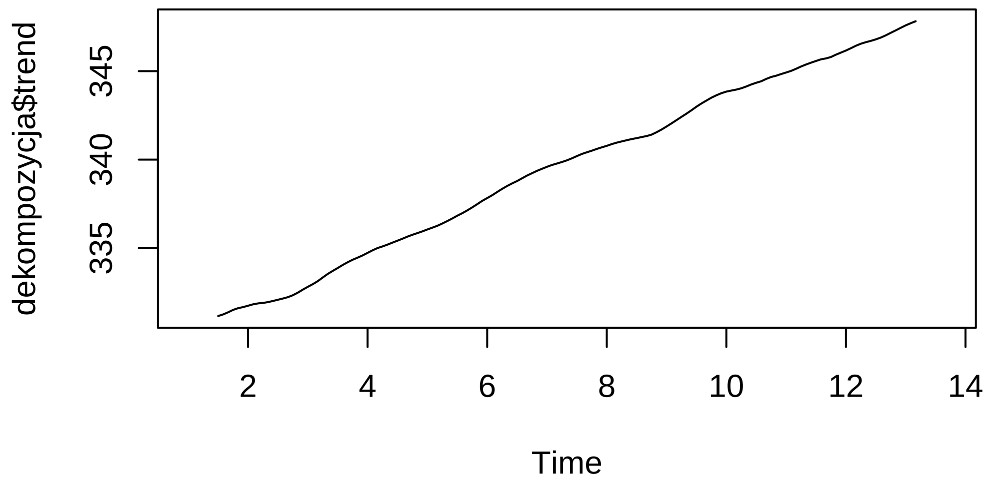
Dokonamy predykcję naszego modelu:

```
round(predict(model, n.ahead = 10),3)
```

```
       Jan     Feb     Mar     Apr     May     Jun     Jul Aug Sep     Oct
13                                                                 345.388
14 349.044 349.735 350.935 352.447 353.037 352.280 350.509
       Nov     Dec
13 346.879 348.213
14
```

Przedstawimy linię trendu:

```
dekompozycja <- decompose(szereg)
plot(dekompozycja$trend)
```



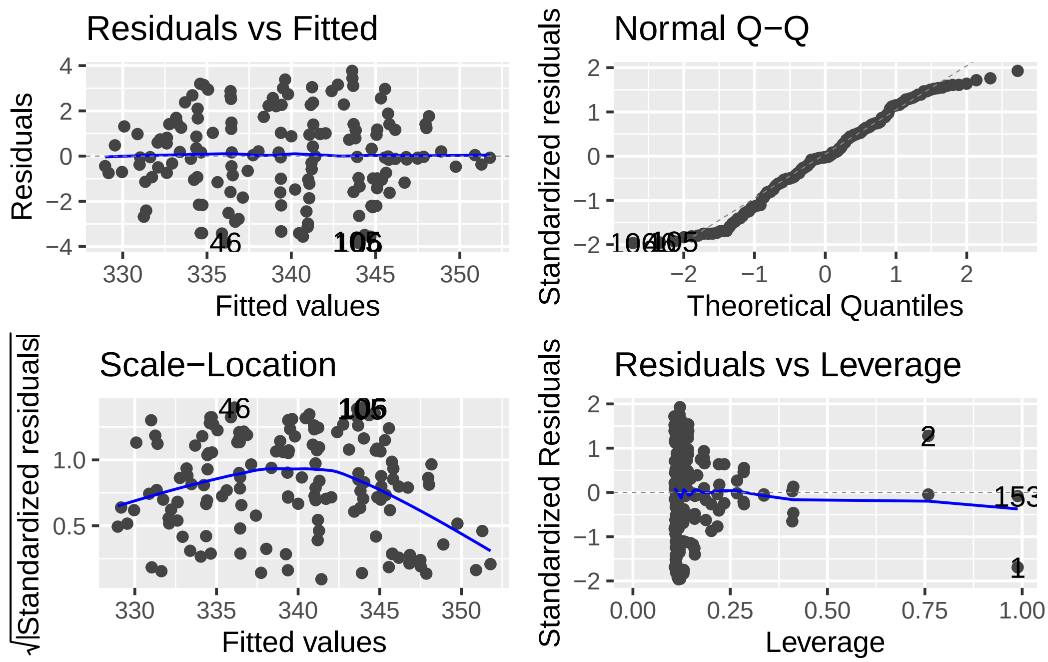
1.6 Dopasowanie trendu wielomianem

Spróbujemy dopasowanie wielomianem różnego stopnia.

Na podstawie analizy wyjaśniamy, że najlepsze AIC = -520.5642 dla wielomianu 15 stopnia.Najlepsze AIC = 401.4457 dla wielomianu 25 stopnia.

```
mod2 <- lm(szereg~poly(1:length(szereg), 25))
```

```
library(ggfortify)
autoplot(mod2)
```



Z wykresów diagnostycznych raczej można się spodziewać jednorodności wariancji, normalności rozkladu reszt, sprawdzimy to za pomocą testów:

1.6.1 Sprawdzenie założeń:

```
library(lmtest)
```

```
Ładowanie wymaganego pakietu: zoo
```

```
Dołączanie pakietu: 'zoo'
```

```
Następujące obiekty zostały zakryte z 'package:base':

    as.Date, as.Date.numeric
```

```
bptest(mod2)
```

```
	studentized Breusch-Pagan test

data:  mod2
BP = 37.103, df = 25, p-value = 0.05645
```

```r
gqtest(mod2)
```

```
	Goldfeld-Quandt test

data:  mod2
GQ = 1.4878, df1 = 51, df2 = 50, p-value = 0.08102
alternative hypothesis: variance increases from segment 1 to 2
```

```r
hmctest(mod2)
```

```
	Harrison-McCabe test

data:  mod2
HMC = 0.47801, p-value = 0.355
```

Wszystkie testy wykazały jednorodnośc wariancji.

```r
dwtest(mod2)
```

```
	Durbin-Watson test

data:  mod2
DW = 0.36734, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(mod2)
```

```
	Breusch-Godfrey test for serial correlation of order up to 1

data:  mod2
LM test = 101.93, df = 1, p-value < 2.2e-16
```

p-value jest bardzo niskie => błędy są zależne.

```r
tseries::kpss.test(mod2$residuals)
```

```
Warning in tseries::kpss.test(mod2$residuals): p-value greater than printed p-
value

    KPSS Test for Level Stationarity

data:  mod2$residuals
KPSS Level = 0.0091104, Truncation lag parameter = 4, p-value = 0.1
```

```r
tseries::adf.test(mod2$residuals)
```

```
Warning in tseries::adf.test(mod2$residuals): p-value smaller than printed p-
value

    Augmented Dickey-Fuller Test

data:  mod2$residuals
Dickey-Fuller = -12.962, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

1.7 Zbadamy staconarność:

```r
tseries::kpss.test(szereg)
```

```
    KPSS Test for Level Stationarity

data:  szereg
KPSS Level = 2.8957, Truncation lag parameter = 4, p-value = 0.01
```

Wniosek: Szereg nie jest stacjonarny.

Po zróżnicowaniu:

```r
tseries::adf.test(diff(szereg))
```

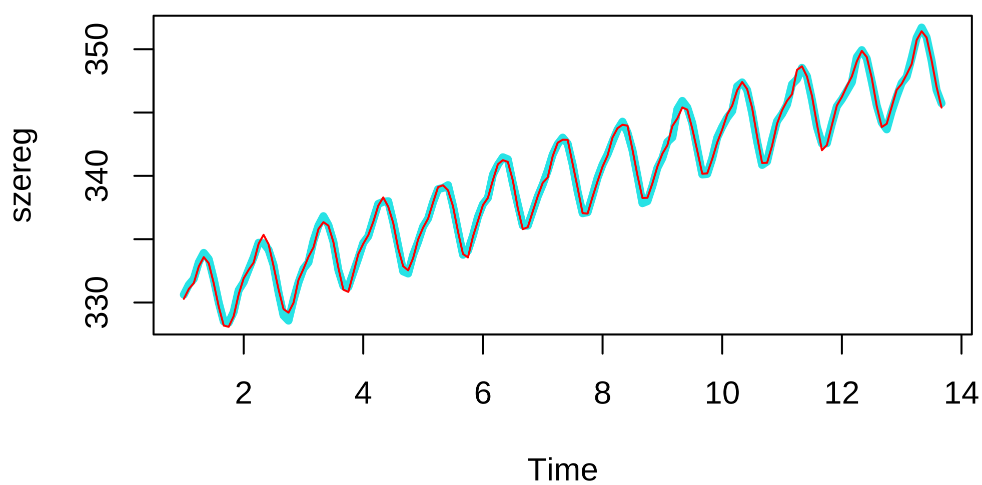
```
	Augmented Dickey-Fuller Test

data:  diff(szereg)
Dickey-Fuller = -9.9789, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Wniosek: Szereg jest stacjonarny.

Z powyższych wyników, decydujemy się ma model 'SARIMA'.

1.8 SARIMA

```r
sarima <- forecast::auto.arima(szereg, seasonal = TRUE)
plot(szereg, col = 5, lwd = 4)
lines(sarima$fitted, col = 'red', lwd = 1)
```



1.8.1 Wyznaczymy przedziały ufności na kolejne 12 miesięcy

```r
forecast::forecast(sarima, h = 12, level = 0.95)
```

```
       Point Forecast    Lo 95    Hi 95
Oct 13       345.3720 344.7302 346.0137
Nov 13       346.9551 346.2670 347.6433
Dec 13       348.2732 347.5275 349.0190
Jan 14       349.0337 348.2488 349.8186
Feb 14       349.6409 348.8287 350.4531
Mar 14       350.9995 350.1681 351.8310
Apr 14       352.2258 351.3806 353.0709
May 14       353.0117 352.1567 353.8666
Jun 14       352.2973 351.4353 353.1594
Jul 14       350.4455 349.5784 351.3126
Aug 14       348.2695 347.3987 349.1403
Sep 14       347.0660 346.1925 347.9394
```

```r
forecast::forecast(sarima, h = 12, level = 0.9)
```

```
       Point Forecast    Lo 90    Hi 90
Oct 13       345.3720 344.8334 345.9105
Nov 13       346.9551 346.3776 347.5326
Dec 13       348.2732 347.6474 348.8991
Jan 14       349.0337 348.3750 349.6924
Feb 14       349.6409 348.9593 350.3225
Mar 14       350.9995 350.3017 351.6973
Apr 14       352.2258 351.5165 352.9350
May 14       353.0117 352.2941 353.7292
Jun 14       352.2973 351.5739 353.0208
Jul 14       350.4455 349.7178 351.1732
Aug 14       348.2695 347.5387 349.0003
Sep 14       347.0660 346.3329 347.7990
```

1.8.2 Sprawdzenie założeń

1.8.2.1 Normalność reszt

```r
shapiro.test(sarima$residuals)
```

```

    Shapiro-Wilk normality test

data:  sarima$residuals
W = 0.98926, p-value = 0.2928
```

```r
nortest::ad.test(sarima$residuals)
```

```
	Anderson-Darling normality test

data:  sarima$residuals
A = 0.56483, p-value = 0.1414
```

```r
nortest::lillie.test(sarima$residuals)
```

```
	Lilliefors (Kolmogorov-Smirnov) normality test

data:  sarima$residuals
D = 0.0503, p-value = 0.45
```

Wniosek: reszty mają rozklad normalny.

1.8.2.2 Jednorodność wariancji

```r
t <- 1:length(szereg)
bptest(as.numeric(mod2$residuals)~t)
```

```
	studentized Breusch-Pagan test

data:  as.numeric(mod2$residuals) ~ t
BP = 0.00035519, df = 1, p-value = 0.985
```

Wariancja jest jednorodna.

```r
dwtest(sarima$residuals~t)
```

```
	Durbin-Watson test

data:  sarima$residuals ~ t
DW = 1.8728, p-value = 0.191
alternative hypothesis: true autocorrelation is greater than 0
```

```
bgtest(sarima$residuals~t, 3)
```

```
	Breusch-Godfrey test for serial correlation of order up to 3

data:  sarima$residuals ~ t
LM test = 0.86384, df = 3, p-value = 0.8341
```

Wniosek: Reszty są nie skorelowane.

Wniosek: Widzimy, że takie ważne i skomplikowane zjawisko jak stężenie CO2 można dobrze opisać naszym modelem, jest on dostatecznie dobry i skuteczny w celach predykcyjnych.

2 Zawartość dwutlenku azotu w powietrzu

2.1 Dane

Drugim tematem naszego projektu będzie opracowanie "Zestawu danych miesięcznych stężeń NO_2".

Dane pochodzą z urządzeń pomiarowych w Londynie. Pokazują średnie odczyty dla stężenia dwutlenku azotu. Jednostką jest mikrogram na metr sześcienny powietrza ($ug/m3$). Każda linia zawiera stężenie NO_2 oraz datę pomiaru. NO_2 to gaz, który cechuje się ostrym zapachem oraz specyficznym brunatnym zabarwieniem. To właśnie za jego sprawą smog przyjmuje nieestetyczne, brązowe zabarwienie. Gaz ten jest główną przyczyną powstawania smogu fotochemicznego w miastach o największym ruchu samochodowym. Tlenki azotu mają również związek z tworzeniem się efektu cieplarnianego oraz zjawiska kwaśnych deszczy zakwaszających gleby.

```
library(tidyverse)
dane <- read.csv("C:/Users/nice2/Desktop/projekt_szeregi/dane3.csv")
dane <- dane[,c(1,3)]
```

```
miesiac <- c("styczeń", "luty", "marzec", "kwiecień", "maj", "czerwiec", "lipiec", "sierpi
miesiac <- rep(miesiac, 11)
miesiac2 <- c(1:12)
miesiac2 <- rep(miesiac2, 11)
rok <- NULL
```

```
for(i in 1:11){
  rok <- c(rok, rep(2007+i, 12))
```

```
}
NO2 <- dane[,2]
dane2 <- cbind(miesiac, rok, NO2,miesiac2) %>% as.data.frame()
dane2[,3] <- as.numeric(dane2[,3])
dane2[,4] <- as.numeric(dane2[,4])
knitr::kable(head(dane2))
```

miesiac	rok	NO2	miesiac2
styczeń	2008	55.50269	1
luty	2008	75.92241	2
marzec	2008	55.61022	3
kwiecień	2008	61.75694	4
maj	2008	62.90323	5
czerwiec	2008	49.16111	6

2.2 Analiza danych za pomocą podstawowych statystyk:

Wyliczymy teraz średnie dla poszczególnych lat aby porównać je z roczną normą, która wynosi $40ug/m^3$ (zalecana norma przez WHO to $10ug/m^3$).

```
srednia <- NULL
for(i in 1:11){
  srednia[i] <- (sum(NO2[(i+12*(i-1)):(i*12)]))/12
}
lata <- NULL
for(i in 1:11){
  lata <- c(lata, 2007+i)
}
ramka <- cbind(lata,srednia)
knitr::kable(ramka)
```

lata	srednia
2008	57.012158
2009	52.608965
2010	47.826330
2011	39.926498
2012	38.251959
2013	32.360596
2014	28.194295

lata	srednia
2015	22.133377
2016	20.738135
2017	13.109213
2018	7.944984

Jak widzimy powyżej w latach 2008-2010 norma została przekroczona.

```
summary(dane2$NO2)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  38.95   49.12   55.71   55.76   61.77   75.92 
```

Z powyższych statystyk możemy odczytać minimalną miesięczną wartość stężenia dwutlenku azotu - 38.95 ug/m^3, a wartość maksymalna to 75.92 ug/m^3.

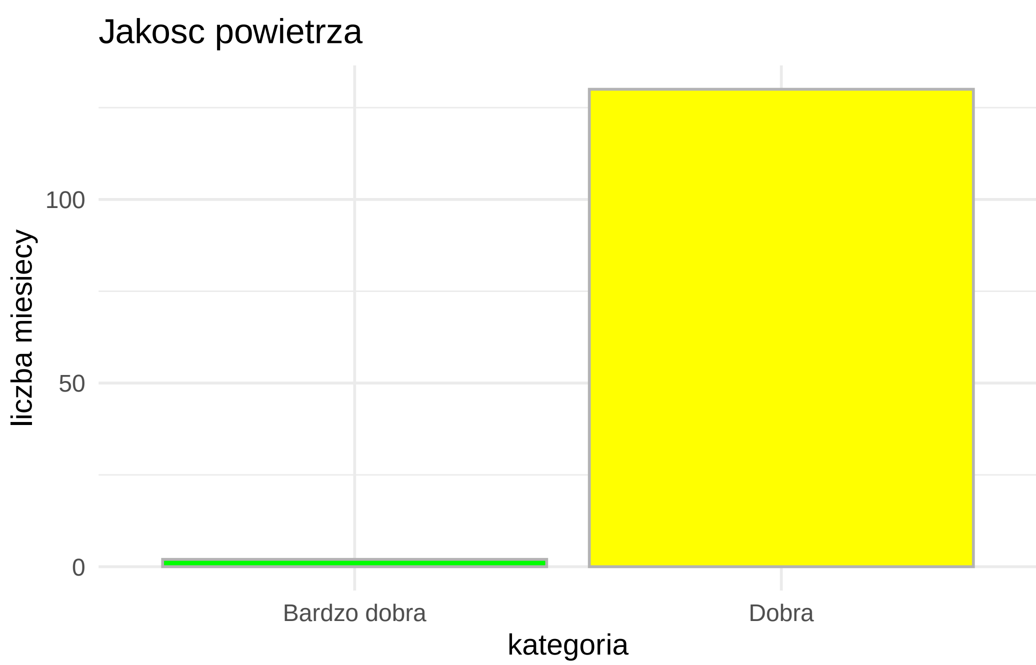
Przeanalizujmy teraz miesięczne normy

Indeks jakości powietrza	**PM10 [µg/m^3]**	**PM2,5 [µg/m^3]**	**O_3 [µg/m^3]**	**NO_2 [µg/m^3]**	**SO_2 [µg/m^3]**
Bardzo dobry	0 - 20	0 - 13	0 - 70	0 - 40	0 - 50
Dobry	20,1 - 50	13,1 - 35	70,1 - 120	40,1 - 100	50,1 - 100
Umiarkowany	50,1 - 80	35,1 - 55	120,1 - 150	100,1 - 150	100,1 - 200
Dostateczny	80,1 - 110	55,1 - 75	150,1 - 180	150,1 - 230	200,1 - 350
Zły	110,1 - 150	75,1 - 110	180,1 - 240	230,1 - 400	350,1 - 500
Bardzo zły	> 150	> 110	> 240	> 400	> 500
Brak indeksu	Indeks jakości powietrza nie jest wyznaczony z powodu braku pomiaru zanieczyszczenia dominującego w województwie.				

Ze statystyk opisowych widzimy, że nasze dane kwalifikują się jedynie do kategorii "Bardzo dobra" i "Dobra". Podzielimy więc dane i sprawdzimy liczebność poszczególnych grup.

```r
dane2["Kategoria"] <- as.factor(ifelse(dane2$NO2< 40, '1','2'))
```

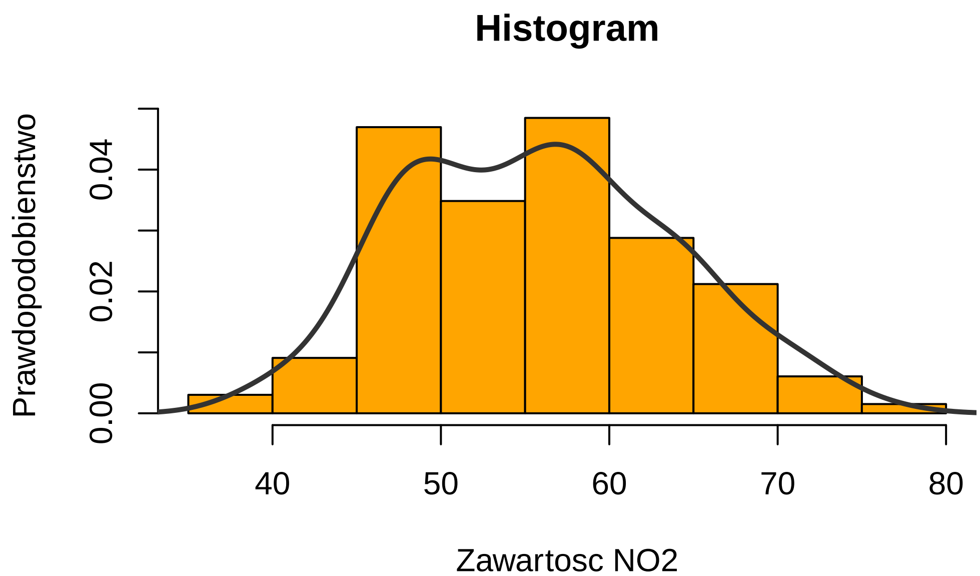
```r
library(ggplot2)

dane2 %>%
  ggplot(aes(x = Kategoria))+
  geom_bar(color = "gray70", fill = c("green", "yellow"))+
  scale_fill_manual("legend", values = c("Bardzo dobra" = "blue","Dobra" = "black"))+
  theme(axis.text.x = c("Bardzo dobra","Dobra"))+
  labs(title = "Jakość powietrza", x = "kategoria", y = "liczba miesięcy")+
  scale_x_discrete(labels = c("Bardzo dobra","Dobra"))+
  guides(fill=guide_legend(title="kategoria"))+
  scale_fill_discrete(breaks=c("1", "2"), labels=c("Bardzo dobra","Dobra"))+
  theme_minimal()
```



Jak widać najliczniejszą grupą jest "Dobra", a tylko dwa miesiace zakwalifikowały się do kategorii "Bardzo dobra".

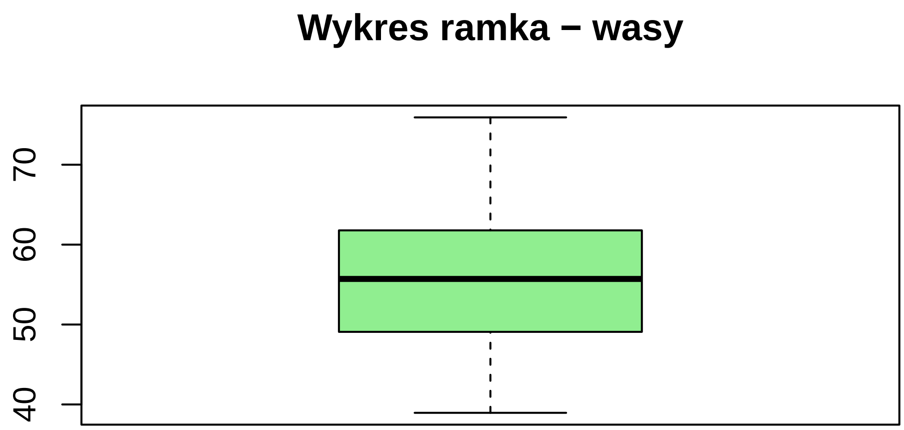
2.3 Wizualizacja

```
hist(NO2, breaks = 11, col = "orange", main="Histogram",
xlab="Zawartość NO2", ylab="Prawdopodobieństwo", prob=T)
gestosc <- density(NO2)
lines(x=gestosc$x, y=gestosc$y, col="grey20", lwd=2.5)
```



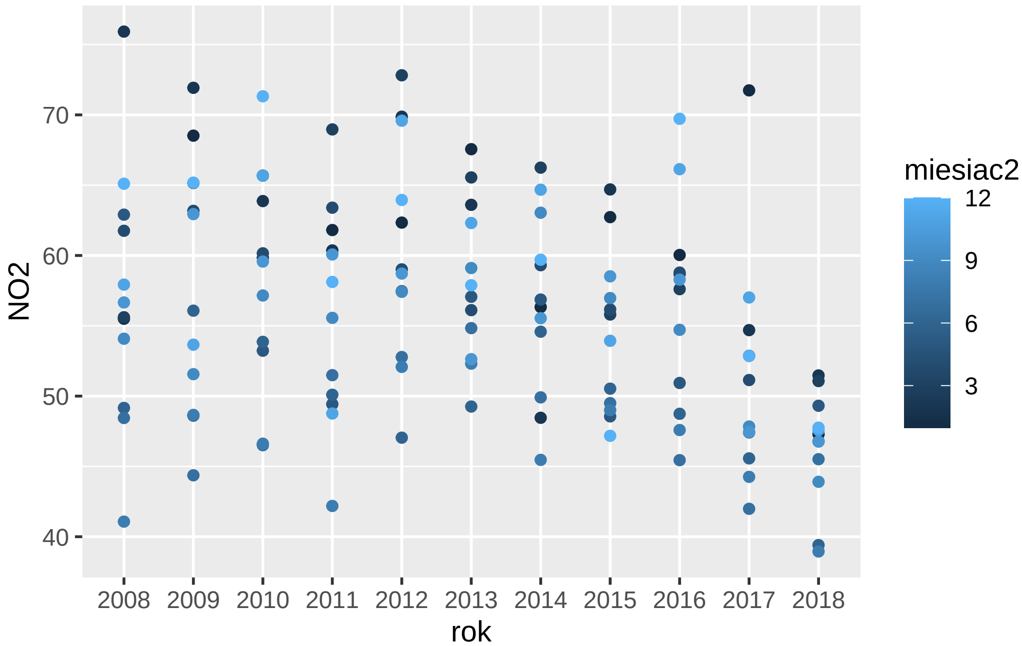
Możemy zaobserwować dualność rozkładu gęstości.

```
boxplot(NO2, col = "lightgreen", main = "Wykres ramka - wąsy")
```



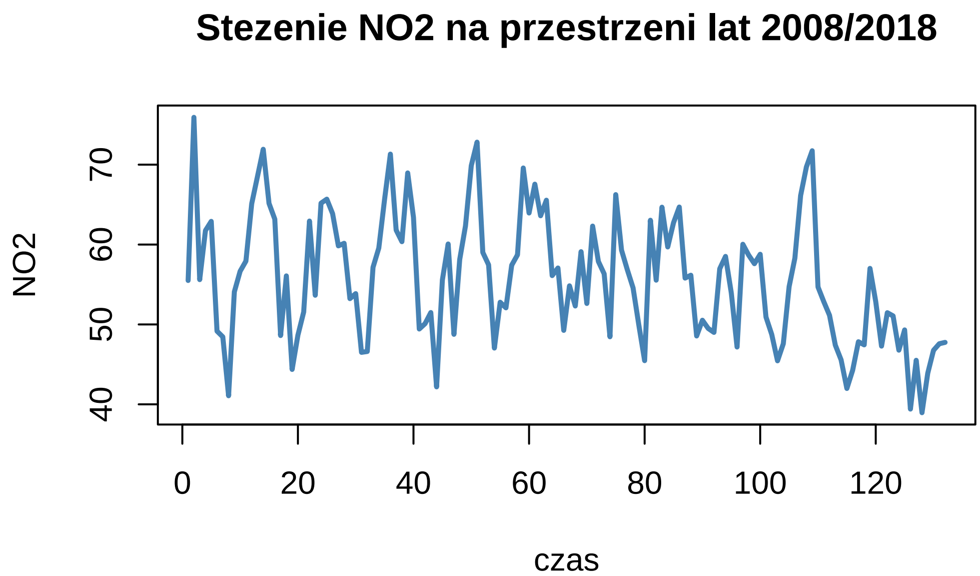
Rozkład cechuje się symetrią.

```
library(tidyverse)
dane2 %>%
  ggplot(aes(x = rok, y = NO2, col = miesiac2))+
  geom_point()
```



Ciężko jest nam odczytać postać zależności skłanialibyśmy się tu do postaci wielomianowej.

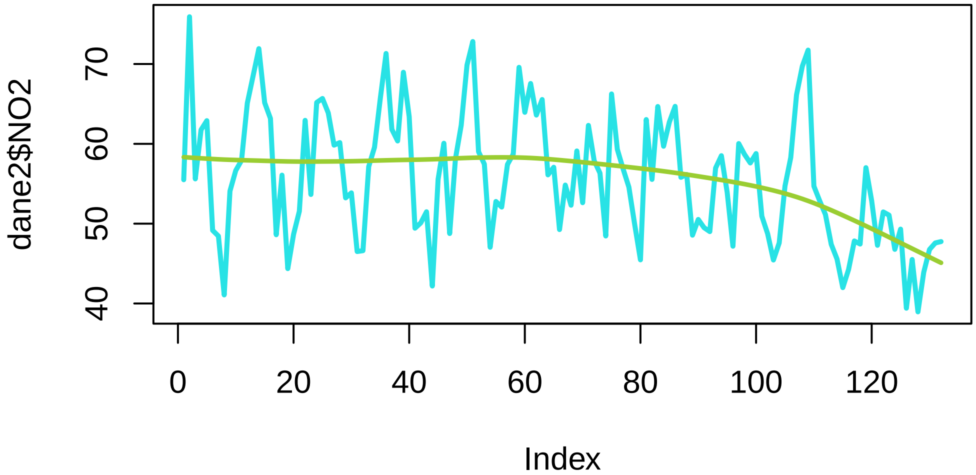
```
plot(NO2, col= "steelblue", main="Stężenie NO2 na przestrzeni lat 2008/2018", type="l", xl
```



Możemy tu dostrzec sezonowość i tendencję spadkową.

Ponieważ mamy dane miesięczne, zaleca się stosowanie współczynnika `lambda` = 14400.

```
f <- FRAPO::trdhp(dane2$NO2, lambda=14400)
plot(dane2$NO2, col = 5, type = "l", main = "", lwd = 2.5)
lines(f,col="YellowGreen", lwd = 2.3)
```

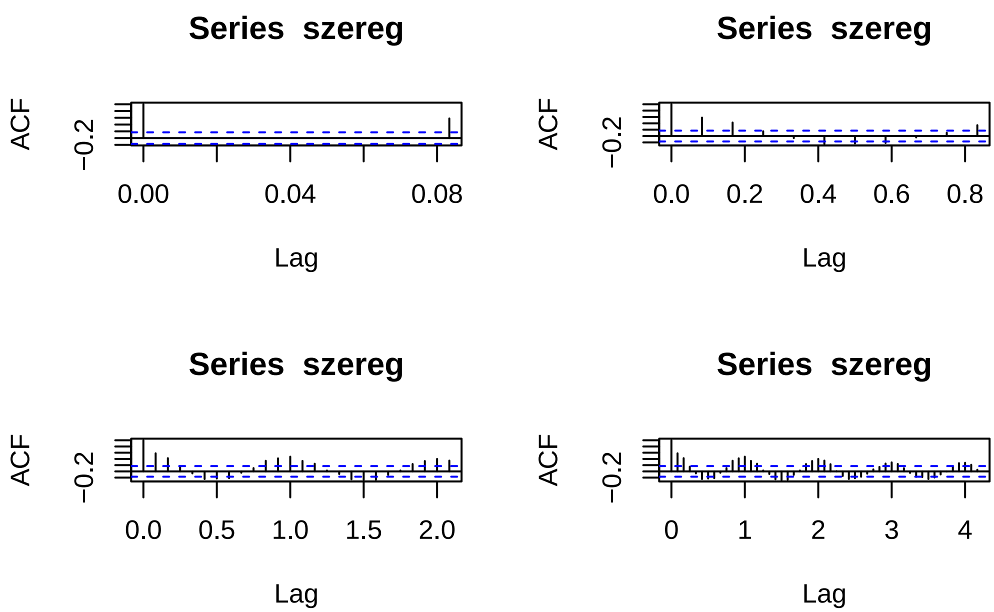


Na tym wykresie widzimy, że nasze wartości początkowo utrzymują stałą średnią, a potem spadają

2.4 Identyfikacja trendu i sezonowości

2.4.1 Autokorelacja:

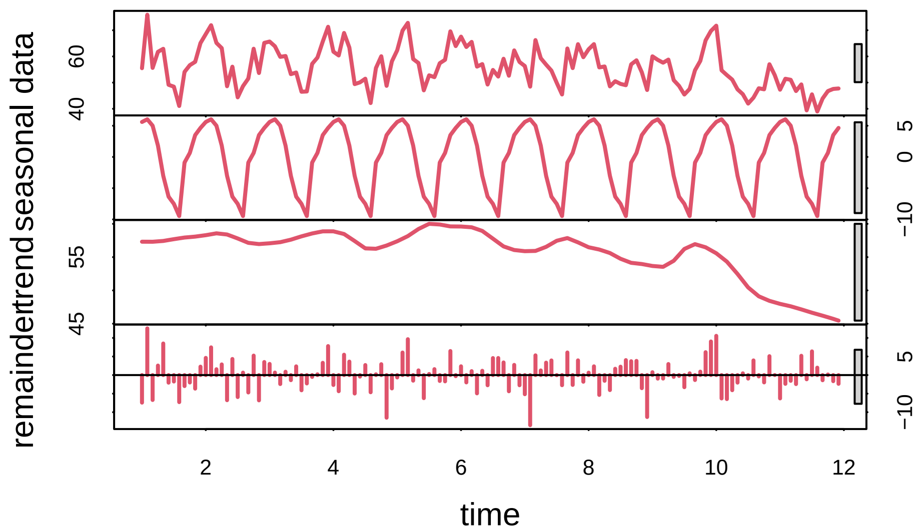
```
szereg <- ts(dane2$NO2, frequency = 12)
par(mfrow = c(2, 2))
acf(x = szereg, lag.max = 1, type = "correlation")
acf(x = szereg, lag.max = 10, type = "correlation")
acf(x = szereg, lag.max = 25, type = "correlation")
acf(x = szereg, lag.max = 50, type = "correlation")
```




```
par(mfrow = c(1, 1))
```

Wykresy przedstawiają funkcję autokorelacji odpowiednie dla $\tau = \{1, 10, 25, 50\}$

```
plot(stl(szereg,s.window="periodic"),col=2,lwd=2)
```

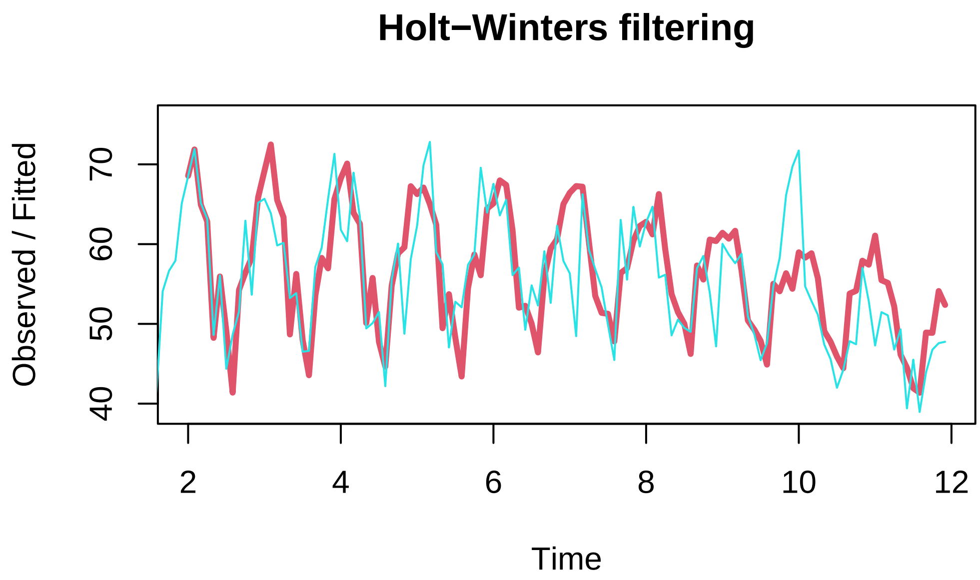


Na podstawie powyższych wykresów można wywnioskować, że nasze dane charakretyzują się sezonowością i pewnym trendem.

2.5 Predykcja

Aby przeprowadzić predykcję potrzebujemy zbudować model, decydujemy się na model Holt'a-Winters'a:

```
model <- HoltWinters(szereg)
plot(model, lwd = 3, col = 5)
```



Czerwona linia (dopasowane wartości) - widzimy, że jest dość dobrze dopasowana do naszych danych

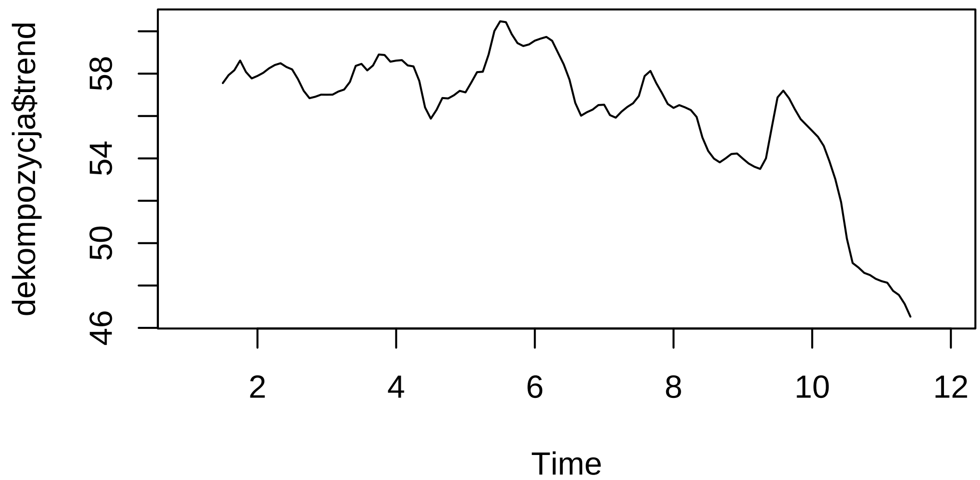
Dokonamy predykcji z użyciem naszego modelu:

```
round(predict(model, n.ahead = 10),3)
```

```
      Jan    Feb    Mar    Apr    May    Jun    Jul    Aug    Sep    Oct
12 53.269 50.333 49.883 46.479 42.775 38.716 38.522 36.183 42.946 43.684
```

Przedstawimy linię trendu:

```
dekompozycja <- decompose(szereg)
plot(dekompozycja$trend)
```



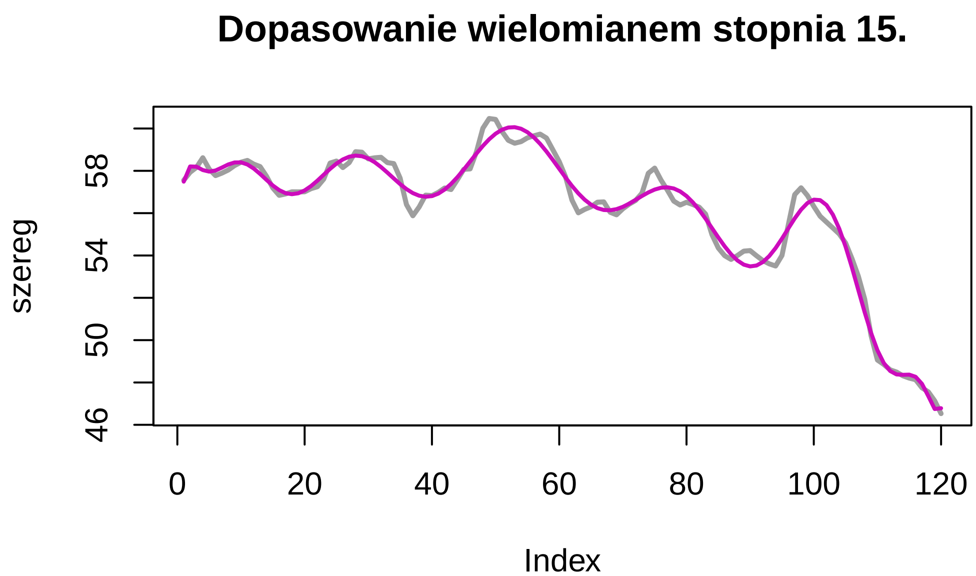
Spróbujemy dopasowanie wielomianem różnego stopnia.

2.6 Dopasowanie trendu wielomianem

Narysujemy trend

```
fit <- function(szereg, max.st){
  aic <- modele <- NULL
  t <- 1:length(szereg)
  for(i in 1:max.st){
    mod <- lm(szereg ~ poly(t, i))
    aic <- c(aic, AIC(mod))
    modele[[i]] <- mod
    }
  opt <- which(aic == min(aic))
  plot(x = szereg, type = "l", col = 8, lwd = 2.5)
  lines(modele[[opt]]$fitted.values, type = "l", col = 6, lwd = 2)
  title(sprintf("Dopasowanie wielomianem stopnia %i.", opt))
  cat("Najlepsze AIC = ", aic[opt], sprintf("dla wielomianu %i", opt), "stopnia.")
  return (modele[[opt]])
  }
```

```
trend <- as.numeric(dekompozycja$trend)
mod <- fit(na.omit(trend), max.st = 15)
```



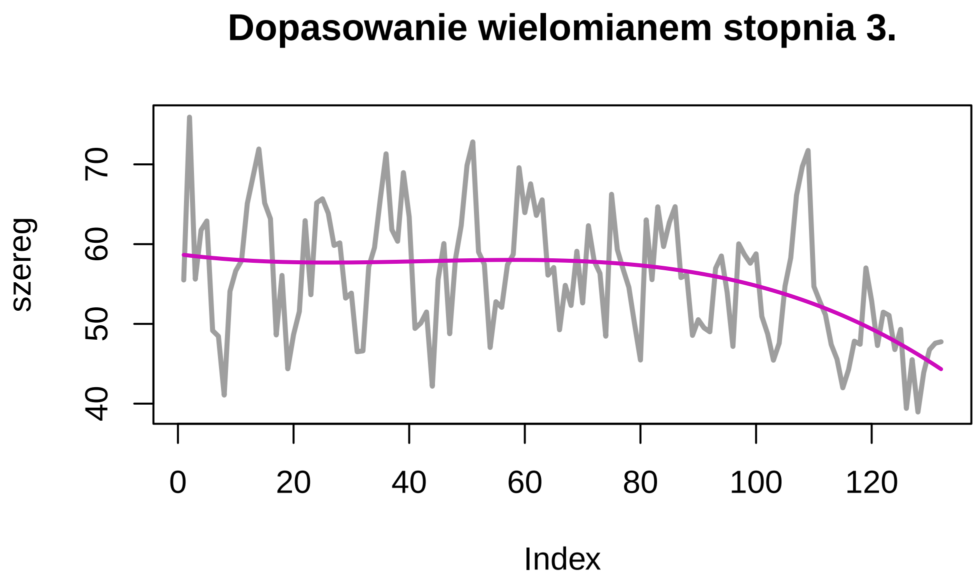

```
Najlepsze AIC =  177.43 dla wielomianu 15 stopnia.
```

```
AIC(mod)
```

```
[1] 177.43
```

AIC dla stopnia 15 to 177.43

```
mod2 <- fit(NO2, 15)
```



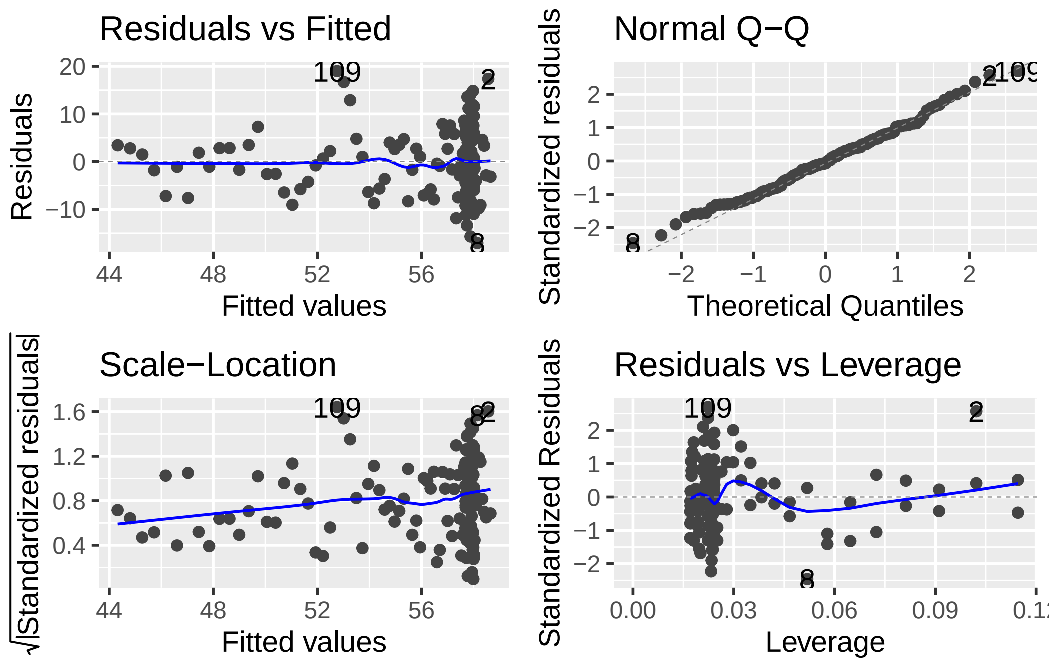

```
Najlepsze AIC =  898.9615 dla wielomianu 3 stopnia.
```

Indeks AIC

```
AIC(mod2)
```

```
[1] 898.9615
```

```
mod2 <- lm(szereg~poly(1:length(szereg), 3))
library(ggfortify)
autoplot(mod2)
```



2.6.1 Sprawdzenie założeń:

2.6.1.1 Jednorodność wariancji

```
library(lmtest)
bptest(mod2)
```

```
	studentized Breusch-Pagan test

data:  mod2
BP = 6.1829, df = 3, p-value = 0.103
```

```
gqtest(mod2)
```

```
	Goldfeld-Quandt test

data:  mod2
GQ = 0.58017, df1 = 62, df2 = 62, p-value = 0.9831
alternative hypothesis: variance increases from segment 1 to 2
```

```r
hmctest(mod2)
```

```
	Harrison-McCabe test

data:  mod2
HMC = 0.63082, p-value = 0.986
```

Wszystkie testy wykazały jednorodnośc wariancji.

2.6.1.2 Autokorelacja reszt

```r
dwtest(mod2)
```

```
	Durbin-Watson test

data:  mod2
DW = 1.0457, p-value = 3.302e-09
alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(mod2)
```

```
	Breusch-Godfrey test for serial correlation of order up to 1

data:  mod2
LM test = 29.901, df = 1, p-value = 4.547e-08
```

Występuje autokorelacja reszt

```r
library(tseries)
kpss.test(mod2$residuals)
```

```
	KPSS Test for Level Stationarity

data:  mod2$residuals
KPSS Level = 0.014925, Truncation lag parameter = 4, p-value = 0.1
```

```r
adf.test(mod2$residuals)
```

```
Warning in adf.test(mod2$residuals): p-value smaller than printed p-value

    Augmented Dickey-Fuller Test

data:  mod2$residuals
Dickey-Fuller = -7.5037, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

2.7 Zbadamy staconarność:

```r
tseries::kpss.test(szereg)
```

```
    KPSS Test for Level Stationarity

data:  szereg
KPSS Level = 0.72601, Truncation lag parameter = 4, p-value = 0.01118
```

Wniosek: Szereg nie jest stacjonarny

Po zróżnicowaniu:

```r
tseries::adf.test(diff(szereg))
```

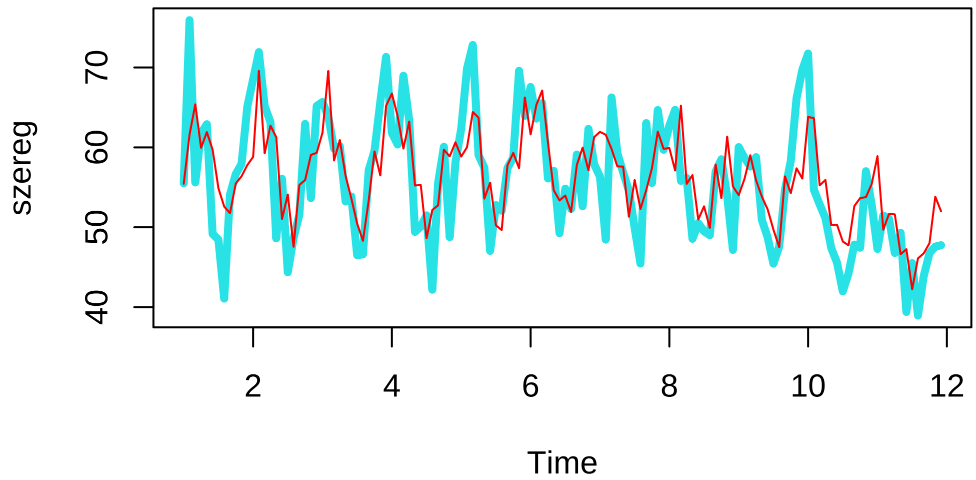
```
Warning in tseries::adf.test(diff(szereg)): p-value smaller than printed p-value

    Augmented Dickey-Fuller Test

data:  diff(szereg)
Dickey-Fuller = -7.3727, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Wniosek: Szereg jest stacjonarny po zróżnicowaniu.

Z powyższych wyników, decydujemy się ma model 'SARIMA'.

2.8 SARIMA

```
sarima <- forecast::auto.arima(szereg, seasonal = TRUE)
plot(szereg, col = 5, lwd = 4)
lines(sarima$fitted, col = 'red', lwd = 1)
```



2.8.1 Wyznaczymy przedziały ufności na kolejne 12 miesięcy

```
forecast::forecast(sarima, h = 12, level = 0.95)
```

```
       Point Forecast    Lo 95    Hi 95
Jan 12       50.74955 39.33052 62.16858
Feb 12       47.99318 35.93561 60.05075
Mar 12       47.63521 35.45100 59.81943
Apr 12       45.87413 33.63511 58.11315
May 12       45.54684 33.26787 57.82580
Jun 12       41.92715 29.61205 54.24226
Jul 12       42.72358 30.37345 55.07372
Aug 12       41.38823 29.00344 53.77302
Sep 12       44.01218 31.59291 56.43146
Oct 12       44.77698 32.32334 57.23062
Nov 12       47.94960 35.46170 60.43751
```

```
Dec 12       46.73324 34.21117 59.25531
```

```r
forecast::forecast(sarima, h = 12, level = 0.9)
```

```
       Point Forecast    Lo 90    Hi 90
Jan 12       50.74955 41.16640 60.33270
Feb 12       47.99318 37.87414 58.11221
Mar 12       47.63521 37.40990 57.86053
Apr 12       45.87413 35.60282 56.14544
May 12       45.54684 35.24200 55.85167
Jun 12       41.92715 31.59199 52.26232
Jul 12       42.72358 32.35902 53.08814
Aug 12       41.38823 30.99459 51.78188
Sep 12       44.01218 33.58960 54.43476
Oct 12       44.77698 34.32556 55.22840
Nov 12       47.94960 37.46943 58.42978
Dec 12       46.73324 36.22438 57.24209
```

2.8.2 Sprawdzenie założeń:

2.8.2.1 Normalność reszt

```r
shapiro.test(sarima$residuals)
```

```

	Shapiro-Wilk normality test

data:  sarima$residuals
W = 0.99362, p-value = 0.8206
```

```r
nortest::ad.test(sarima$residuals)
```

```

	Anderson-Darling normality test

data:  sarima$residuals
A = 0.1486, p-value = 0.9637
```

```r
nortest::lillie.test(sarima$residuals)
```

```
	Lilliefors (Kolmogorov-Smirnov) normality test

data:  sarima$residuals
D = 0.034426, p-value = 0.964
```

Wniosek: Reszty mają rozklad normalny.

2.8.2.2 Jednorodność wariancji

```r
t <- 1:length(szereg)
bptest(as.numeric(sarima$residuals)~t)
```

```
	studentized Breusch-Pagan test

data:  as.numeric(sarima$residuals) ~ t
BP = 3.0391, df = 1, p-value = 0.08128
```

Wniosek: Wariancja jest jednorodna.

2.8.2.3 Autokorelacja reszt

```r
dwtest(sarima$residuals~t)
```

```
	Durbin-Watson test

data:  sarima$residuals ~ t
DW = 2.1044, p-value = 0.6969
alternative hypothesis: true autocorrelation is greater than 0
```

```r
bgtest(sarima$residuals~t, 3)
```

```
	Breusch-Godfrey test for serial correlation of order up to 3

data:  sarima$residuals ~ t
LM test = 1.1126, df = 3, p-value = 0.774
```

Wniosek: Reszty są nie skorelowane.

Podsumowanie: Model spełnia założenia i jest dobrze dopasowany.