

# Projekt

Izabela Guła

2024-03-25

## Temat projektu

Tematem projektu jest interesujący mnie aspekt zmiany środowiska naturalnego, szczególnie zmian klimatu. Postanowiłam zbadać i przeanalizować, jakie czynniki wpływają na emisję dwutlenku węgla, podtlenku azotu oraz na wykorzystanie odnawialnych źródeł energii. Wybrałam kilka zmiennych, które moim zdaniem mogą mieć istotny wpływ na te aspekty. Moim celem jest zbadanie korelacji i powiązań między wybranymi zmiennymi a emisją gazów cieplarnianych oraz zużyciem energii odnawialnej. W większości będę się posługiwać wykresami i wyciągać z nich interesujące mnie kwestie i wnioski. Czy wybrane zmienne rzeczywiście mają wpływ na emisje gazów cieplarnianych i korzystanie z energii odnawialnej? Oto co zamierzam zbadać.

## Pakiety, których będę używać

```
library(dplyr)
library(countrycode)
library(mice)
library(tidyr)
library(gapminder)
library(ggplot2)
library(RColorBrewer)
library(ggrepel)
library(gridExtra)
library(cowplot)
```

## Zapoznanie się ze zmiennymi

### Populacja (pop)

Całkowita populacja oznaczająca liczbę mieszkańców na terytorium danego państwa.

### PKB per capita (PKB)

Produkt krajowy brutto na osobę. Dane skorygowane pod względem różnic w sile nabywczej. PKB wyrażone jest w międzynarodowych dolarach, skorygowane pod względem inflacji bazując na roku 2017 - parytet siły nabywczej bazowany jest na ICP z roku 2017.

### Procentowy udział rolnictwa w PKB (agriculture)

Jest to procentowy udział rolnictwa w wartości PKB. Obejmuje on leśnictwo, łowiectwo i rybołówstwo, a także uprawę roślin i produkcję zwierzęcą.

### Poziom urbanizacji (urban)

Jest to procentowy udział mieszkańców miast w ogólnej liczbie ludności.

### Emisja podtlenku azotu na osobę (nitrous\_per\_capita\_individua)

Jest to emisja gazu cieplarnego podtlenku azotu na osobę. Należy on do głównych gazów cieplarnianych. Jego emisja pochodzi głównie z rolnictwa. Będąc trzecim najważniejszym długotrwałym gazem cieplarnianym, podtlenek azotu w znacznym stopniu przyczynia się do globalnego ocieplenia i jest substancją, która znacznie zubaża ozon stratosferyczny. Zmienna została stworzona przeze mnie, dzieląc ogólną emisję podtlenku azotu przez populację. Ogólna emisje również zostawiłam w tabeli, ponieważ uznałam, że może przyda się w któryś obliczeniach.

**Emisja dwutlenku węgla na osobę (CO2)**

Emisja dwutlenku węgla pochodzącego ze spalania paliw kopalnianych w tonach na osobę (metryczne tony dwutlenku węgla na osobę).

**Powierzchnia terenów leśnych (woods)**

Procent całkowitej powierzchni terenów leśnych wyrażony w procentach.

**Produkcja energii odnawialnej (renewable)**

Procent użycia energii odnawialnej w stosunku do całkowitego użycia energii.

**Ślad materiałowy (material)**

Ślad materiałowy to całkowita ilość surowców, takich jak biomasa, paliwa kopalne, rudy metali i rudy niemetali, wydobywana co roku w celu zaspokojenia zapotrzebowania konsumpcyjnego krajów, wyrażona w tonach.

**Procentowy udział przemysłu w PKB (industry)**

Jest to procentowy udział przemysłu w PKB. Obejmuje wartość dodaną w górnictwie, przemyśle przetwórczym (również wykazywanym jako odrębna podgrupa), budownictwie, energii elektrycznej, wodzie i gazie.

**Kontynenty**

Postanowiłam podzielić świat na 4 części: Ameryki, Azję, Europe i Afrykę. Postanowiłam usunąć Australię i Oceanie z powodu brakujących i bardzo odstających wartości.

# Zapoznanie się ze strukturą pliku

Zestaw danych został stworzony z gapmintera, poprzez połączenie ze sobą wybranych zmiennych.

```
str(dane)
```

```
## 'data.frame':    1744 obs. of  14 variables:
##  $ kraj                : chr  "UAE" "Argentina" "Austria" "Burundi" ...
##  $ rok                  : chr  "1990" "1990" "1990" "1990" ...
##  $ pop                  : num  1.90e+06 3.26e+07 7.68e+06 5.48e+06 5.13e+06 9.1
3e+06 1.07e+08 8.77e+06 5.17e+05 2.71e+05 ...
##  $ PKB                  : num  75100 14300 37800 1280 2240 1170 1640 12600 2500
0 37900 ...
##  $ agriculture          : num  1.06 8.12 3.1 51.1 34.7 28 30.5 17.7 0.949 2.21
...
##  $ urban                : num  79.1 87 63 6.27 34.5 13.8 19.8 66.4 88.1 79.8
...
##  $ nitrous              : num  514 35100 4980 719 1550 4370 14700 8330 50.3 18.
9 ...
##  $ CO2                  : num  28.3 3.43 8.08 0.0384 0.137 0.0662 0.137 8.66 25
7.18 ...
##  $ woods                : num  3.45 12.9 45.8 10.8 42.9 28.2 14.8 30.1 0.32 50.
9 ...
##  $ renewable            : num  0 35.4 66.2 98.2 0 4.89 11.4 4.46 0 0 ...
##  $ material              : num  64.8 11.1 22.2 3.27 4.04 3.32 1.45 13.9 28.9 43.
4 ...
##  $ industry             : num  58.9 36 29.6 17.3 12.5 20.4 20.1 51.3 45.7 14.2
...
##  $ kontynent            : chr  "Asia" "Americas" "Europe" "Africa" ...
##  $ nitrous_per_capita_individual: num  0.000271 0.001077 0.000648 0.000131 0.000302 ...
##  - attr(*, "na.action")= 'omit' Named int [1:245] 1 2 3 6 8 10 17 18 25 26 ...
##  ..- attr(*, "names")= chr [1:245] "1" "2" "3" "6" ...
```

W moim pliku poza nazwami państw i kontynentów przeważają dane numeryczne. W sumie kolumn jest 14, a wartości 1553. Pozbyłam się brakujących danych NA1. Liczę, że nie przeszkodzi mi to w poprawnej analizie, gdyż danych jest nadal sporo.

## Poznanie zmiennych i ich rozkładu

W tym punkcie, za pomocą prostych statystyk i wykresów (w większości liniowych) poznam rozkład każdej zmiennej, co pozwoli wyciągnąć mi ciekawe wnioski i pomoże mi w stawianiu późniejszych hipotez.

### Populacja

```
population_in_thousands <- dane$pop/1000
summary(population_in_thousands)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      183    4520   10350   44227   30325 1390000
```

```

max_population <- max(dane$pop)
min_population <- min(dane$pop)

max_population_country <- filter(dane, pop == max_population) %>%
  mutate(population_in_thousands = pop/1000) %>%
  select(kraj, population_in_thousands)

min_population_country <- filter(dane, pop== min_population) %>%
  mutate(population_in_thousands = pop/1000) %>%
  select(kraj, population_in_thousands)

result_table <- bind_rows(max_population_country, min_population_country)
result_table

```

kraj	population_in_thousands
<chr>	<dbl>
China	1390000
Belize	183
2 rows	

Zmieniam dane, żeby populacja pokazywała mi się w tysiącach. W moim zestawie danych są państwa o bardzo dużej populacji, jak i tej małej. Największą populację posiadają Chiny, szczerze myślałam, że będą to Indie, jednak okazuje się, że nawet teraz Chiny wyprzedzają Indie z liczbą mieszkańców. Najmniejszą zauważoną populację posiada Brunei, państwo w Azji.

## PKB per capita

```
summary(dane$PKB)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      482     3430     9345    16523   23525   117000
```

```
sd(dane$PKB)
```

```
## [1] 18643.83
```

Jak można zauważyć, odchylenie standardowe PKB per capita jest duże, jednak jest to normalne, ponieważ w mojej bazie danych posiadam państwa bogate, jak i biedne.

Sprawdzę jednak, jak wygląda sytuacja, gdy pogrupujemy państwa na kontynenty.

```

pkb_dane <- dane %>%
  group_by(kontynent) %>%
  summarise(mean = mean(PKB), sd = sd(PKB))
pkb_dane

```

kontynent <chr>	mean <dbl>	sd <dbl>
Africa	4416.284	5348.93
Americas	14065.929	11387.03
Asia	17083.136	20930.42
Europe	32098.910	18960.26
4 rows		

Największe różnice w wartości PKB można zauważyć w Azji, najmniejsze w Afryce, jednak PKB w Afryce posiada najniższą średnia arytmetyczna.

## Udział rolnictwa w PKB

```
summary(dane$agriculture)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.0333	3.5350	9.7650	13.9086	21.6000	79.0000

W zestawie danych posiadam państwa, które mają naprawdę mały udział rolnictwa w PKB, bo jest to tylko 0,03%, oraz takie w których rolnictwo ma aż 56%, czyli ponad połowę udziału w PKB. Z mojej czystej ciekawości sprawdzę jakie to państwa.

```
max_rolnictwo <- max(dane$agriculture)
min_rolnictwo <- min(dane$agriculture)

max_rolnictwo_country <- filter(dane, agriculture == max_rolnictwo) %>%
  select(kraj, agriculture)

min_rolnictwo_country <- filter(dane, agriculture == min_rolnictwo) %>%
  select(kraj, agriculture)

wynik_rolnictwo <- bind_rows(max_rolnictwo_country, min_rolnictwo_country)
wynik_rolnictwo
```

kraj <chr>	agriculture <dbl>
Liberia	79.0000
Singapore	0.0333
2 rows	

Tak szczerze, to troszkę spodziewałam się takich wyników. Singapur jest to miasto-państwo w Azji Południowo-Wschodniej, gdzie sądziłam, że udziały rolnictwa są bardzo małe. Co do najwyższych udziałów rolnictwa to spodziewałam się, że będzie to jakieś państwo afrykańskie, gdzie głównie rozwija się rolnictwo. Gospodarka Konga jest być silnie uzależniona od rolnictwa z powodu tradycyjnej struktury gospodarczej opartej głównie na produkcji rolnej.

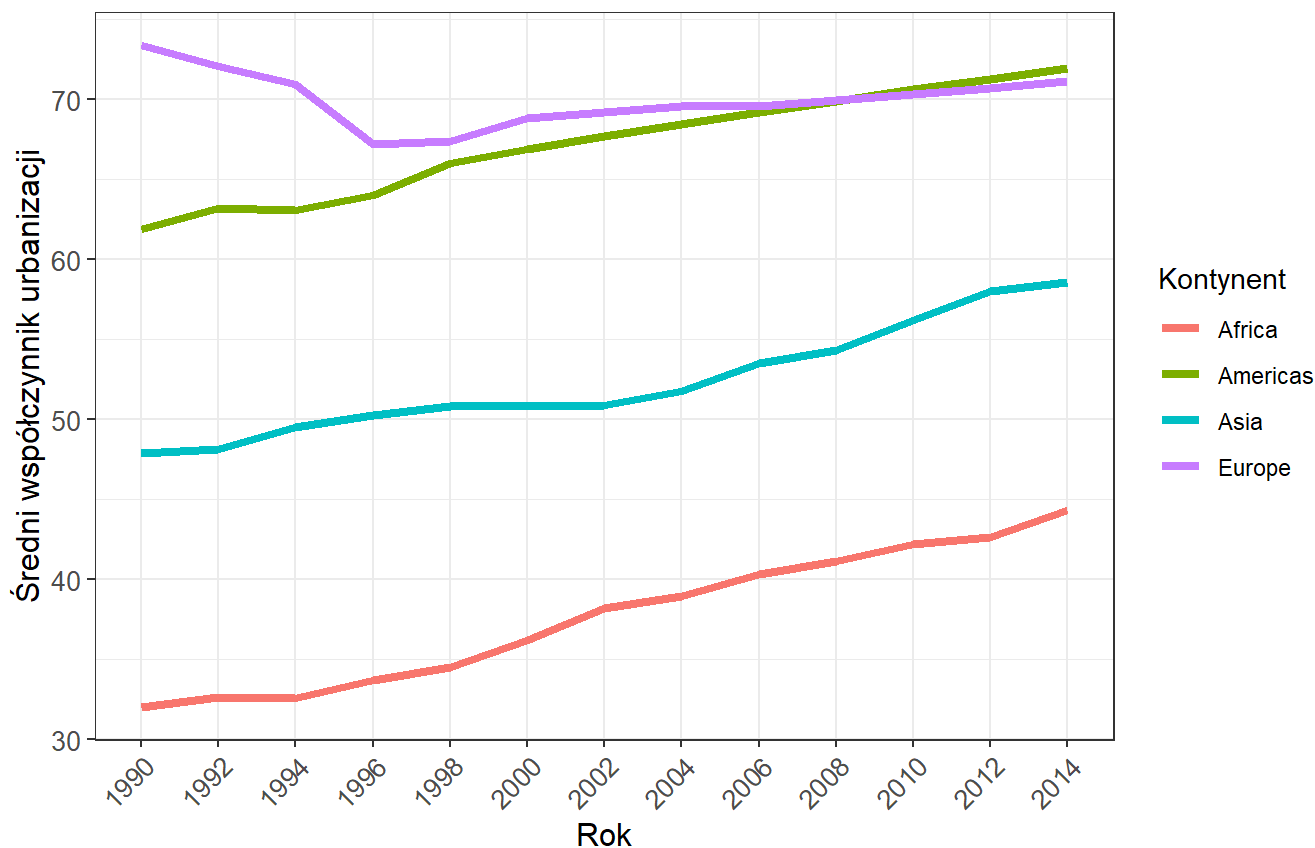
# Współczynnik urbanizacji

```
wykres_urban <- dane %>%
  group_by(kontynent, rok) %>%
  summarize(mean_urban = mean(urban, na.rm = TRUE))

ggplot(wykres_urban, aes(x = rok, y = mean_urban, color = kontynent, group = kontynent)) +
  geom_line(size = 1.5) +
  theme_bw() +
  labs(title = "Średni współczynnik urbanizacji w latach 1990 - 2014",
       subtitle = "na wybranych kontynentach",
       color = "Kontynent") +
  xlab("Rok") +
  ylab("Średni współczynnik urbanizacji") +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=14),
        axis.text.y = element_text(size=10),
        axis.text.x = element_text(angle = 45, hjust = 1),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        plot.title = element_text(size = 14, face = "bold"))
```

## Średni współczynnik urbanizacji w latach 1990 - 2014

na wybranych kontynentach



Uważam, że urbanizacja była do przewidzenia. Państwa na wszystkich kontynentach coraz bardziej się rozwijają i coraz więcej osób mieszka w miastach. Spodziewałam się najmniejszego wyniku w Afryce, natomiast najwyższego wyniku w Europie i Amerykach.

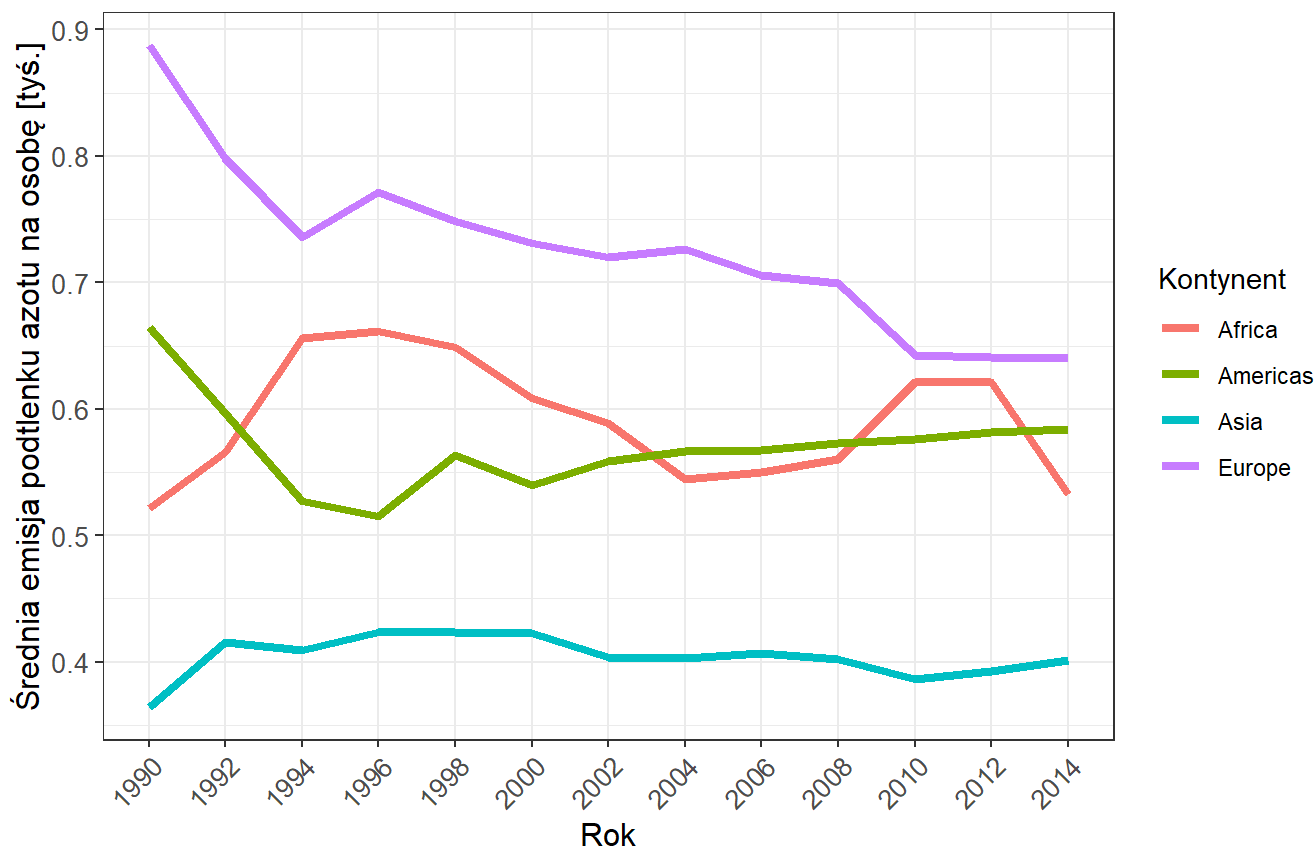
## Emisja podtlenku azotu na osobę

```
wykres_nitrous <- dane %>%
  group_by(kontynent, rok) %>%
  summarize(mean_nitrous = mean(nitrous_per_capita_individual, na.rm = TRUE))

ggplot(wykres_nitrous, aes(x = rok, y = mean_nitrous, color = kontynent, group = kontynent)) +
  geom_line(size = 1.5) +
  theme_bw() +
  labs(title = "Średnia emisja podtlenku azotu na osobę",
       subtitle = " w latach 1990 - 2014 na wybranych kontynentach",
       color = "Kontynent") +
  xlab("Rok") +
  ylab("Średnia emisja podtlenku azotu na osobę [tyś.]")+
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=14),
        axis.text.y = element_text(size=10),
        axis.text.x = element_text(angle = 45, hjust = 1),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        plot.title = element_text(size = 14, face = "bold")) +
  scale_y_continuous(labels = scales::number_format(scale = 1e3))
```

## Średnia emisja podtlenku azotu na osobę

w latach 1990 - 2014 na wybranych kontynentach

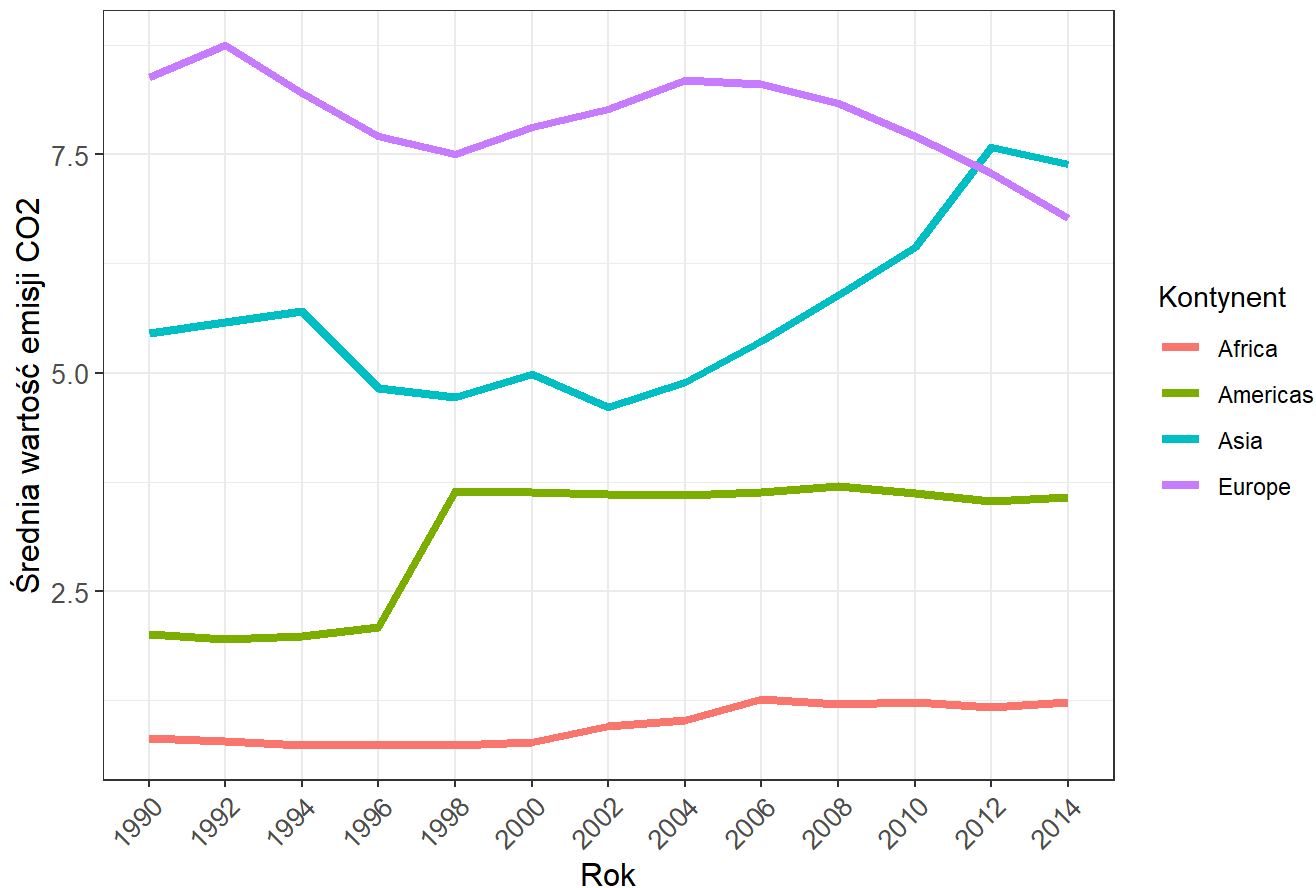


Aby ułatwić odczytywanie wykresu, zmieniłam formatowanie wartości dla emisji podtlenku azotu, są to małe liczby, dlatego dla ułatwienia są one wyświetlane w tysiącach. Tutaj powstałe wartości są bardzo ciekawe. Po pierwsze najwięcej podtlenku azotu produkuje Europa, Afryka i Ameryki. Natomiast o wiele mniejszą emisję niż pozostałe kontynenty możemy zaobserwować w Azji, nie sądziłam, że Azja będzie tutaj wartością odstającą ze względu na wiele rozwiniętych państw, które potencjalnie mogą być odpowiedzialne za wysoką emisję gazów cieplarnianych.

# Emisja dwutlenku węgla na osobę

```
wykres_CO2 <- dane %>%  
  group_by(kontynent, rok) %>%  
  summarize(mean_CO2 = mean(CO2, na.rm = TRUE))  
  
ggplot(wykres_CO2, aes(x = rok, y = mean_CO2, color = kontynent, group = kontynent)) +  
  geom_line(size = 1.5) +  
  theme_bw() +  
  labs(title = "Średnia wartość emisji CO2 na osobę",  
        color = "Kontynent") +  
  xlab("Rok") +  
  ylab("Średnia wartość emisji CO2") +  
  theme(axis.text=element_text(size=10),  
        axis.title=element_text(size=14),  
        axis.text.y = element_text(size=10),  
        axis.text.x = element_text(angle = 45, hjust = 1),  
        axis.title.x = element_text(size = 12),  
        axis.title.y = element_text(size = 12),  
        plot.title = element_text(size = 14, face = "bold"))
```

## Średnia wartość emisji CO2 na osobę



Z wykresu wynika, że najwyższą średnią emisję CO2 posiadają Azja i Europa. Cieszy mnie tendencja spadkowa w Europie w ciągu ostatnich lat. Najniższe wartości występują w Afryce, co według mnie było do przewidzenia.

Postanowiłam sprawdzić jakie państwa posiadają najwyższą średnią emisję CO2.



```
highest_emission <- dane %>%
  group_by(kraj) %>%
  summarize(srednia_CO2 = mean(CO2)) %>%
  top_n(n = 5, wt = srednia_CO2)
highest_emission
```

kraj	srednia_CO2
<chr>	<dbl>
Bahrain	23.43750
Kuwait	27.86667
Luxembourg	22.23750
Qatar	42.80000
UAE	26.45385
5 rows	

Pokazane wartości bardzo mnie zaintrygowały. Po sprawdzeniu okazało się, że są to państwa m.in. Bliskiego Wschodu, ale co łączy te wszystkie państwa? Okazuje się, że są one znane z wydobycia ropy naftowej czy gazu ziemnego. Wydobycie ropy naftowej ma związane ze sobą emisje dwutlenku węgla i inne gazów cieplarnianych. Proces wydobycia ropy naftowej oraz jej produkcji, transportu i rafinacji generuje emisję CO<sub>2</sub> do atmosfery. Stąd już wiemy, kto jest odpowiedzialny za największą emisję dwutlenku węgla do atmosfery.

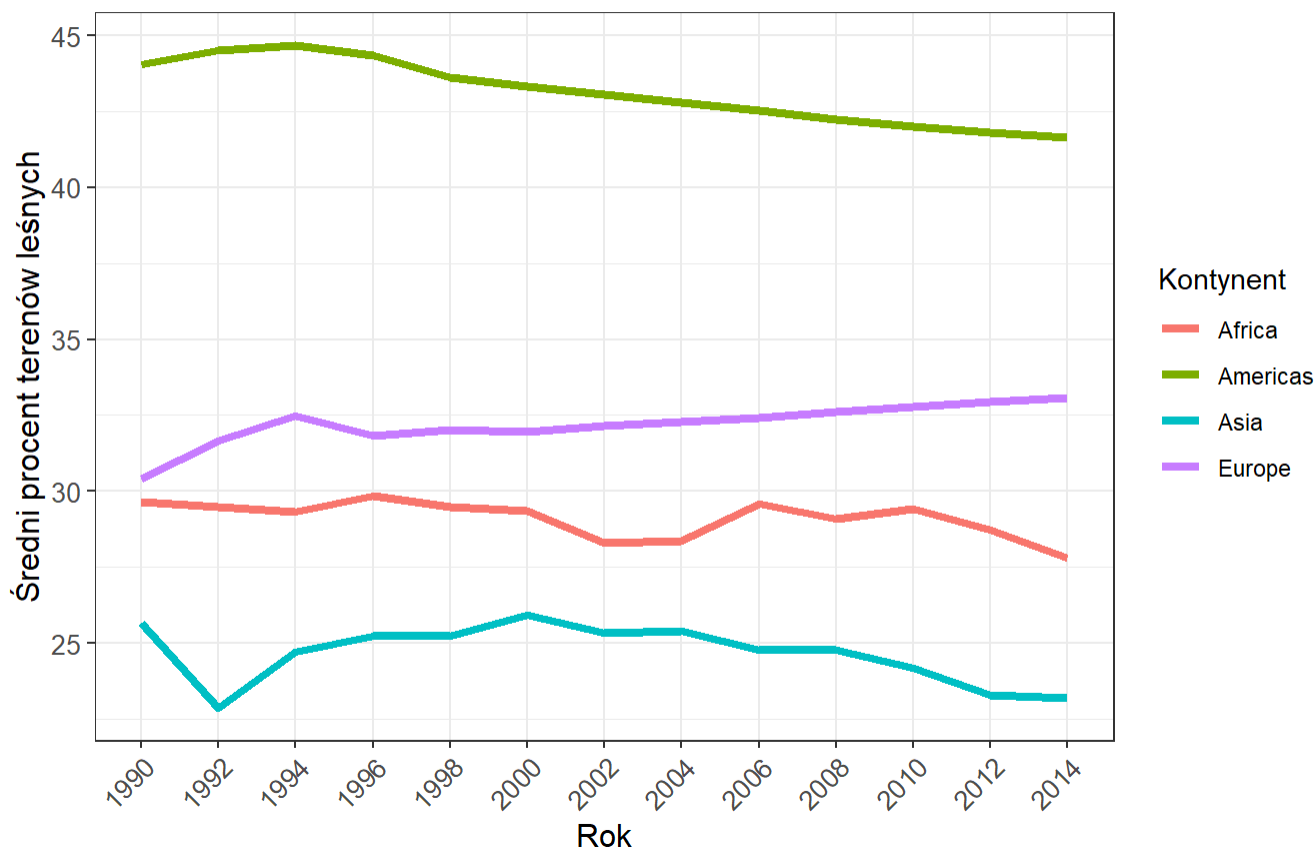
## Powierzchnia terenów leśnych

```
wykres_woods <- dane %>%
  group_by(kontynent, rok) %>%
  summarize(mean_woods = mean(woods, na.rm = TRUE))

ggplot(wykres_woods, aes(x = rok, y = mean_woods, color = kontynent, group = kontynent)) +
  geom_line(size = 1.5) +
  theme_bw() +
  labs(title = "Średni procent terenów leśnych w latach 1990 - 2014",
       subtitle = "na wybranych kontynentach",
       color = "Kontynent") +
  xlab("Rok") +
  ylab("Średni procent terenów leśnych") +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=14),
        axis.text.y = element_text(size=10),
        axis.text.x = element_text(angle = 45, hjust = 1),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        plot.title = element_text(size = 14, face = "bold"))
```

# Średni procent terenów leśnych w latach 1990 - 2014

na wybranych kontynentach



Najwięcej terenów leśnych jest na kontynentach Ameryki Północnej i Południowej. Rzeczą, która mnie zaskoczyła to, że w Azji występuje tak mało terenów leśnych. Jednak co nie jest już optymistyczne to spadek udziału terenów leśnych w Afryce. Jestem ciekawa, jakie są to państwa i czym jest spowodowany ten spadek.

```
africa_woods <- dane %>%
  filter(kontynent == "Africa") %>%
  group_by(kraj) %>%
  summarize(zmiana = max(woods) - min(woods)) %>%
  top_n(n = 5, wt = zmiana)
africa_woods
```

kraj	zmiana
<chr>	<dbl>
Benin	12.4
Cote d'Ivoire	13.6
Gambia	13.6
Malawi	10.7
Tanzania	10.1

5 rows

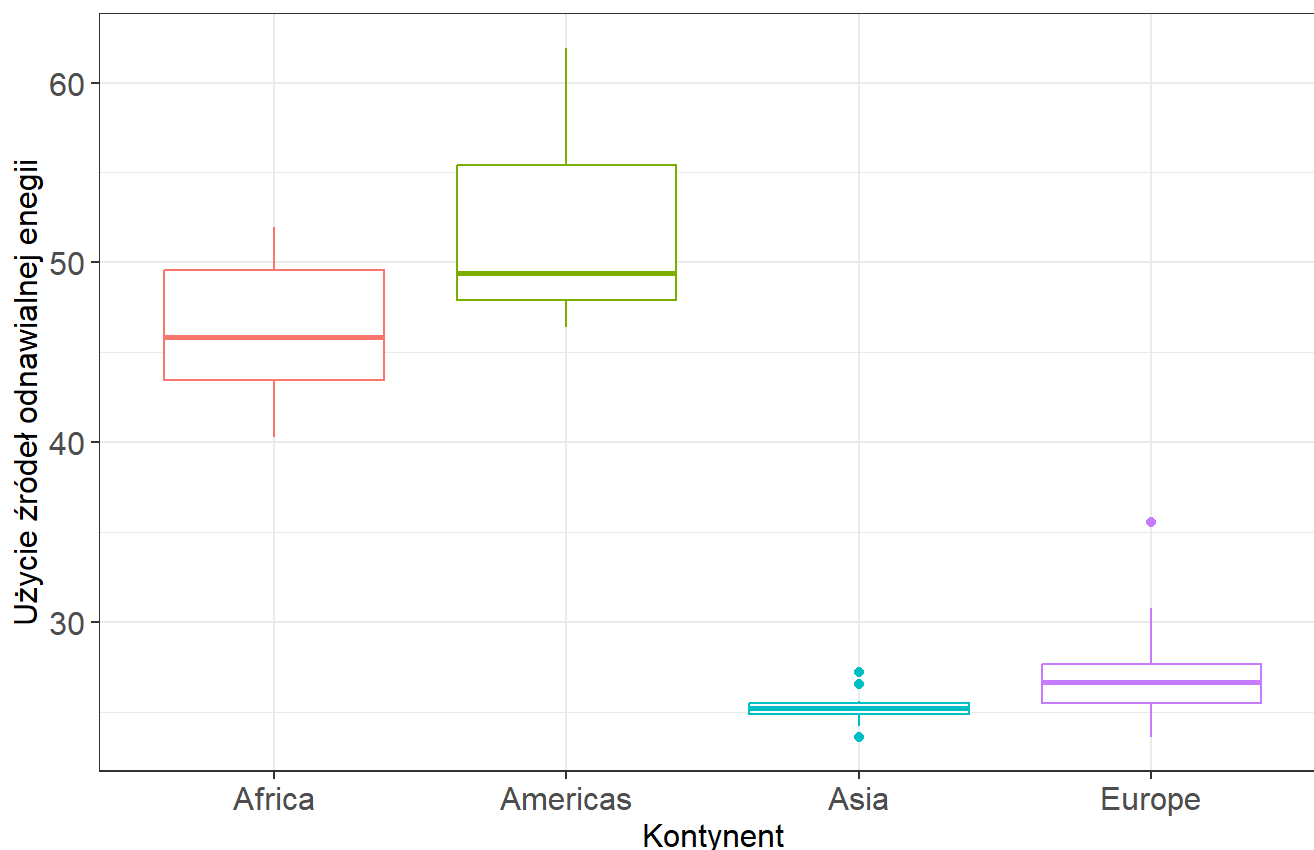
W niektórych państwach jest to spadek o ponad 10 punktów procentowych. Okazuje się, że te państwa leżą na terenach, gdzie występują lasy tropikalne. Duża część Afryki została poddana deforestacji w XX wieku. Była ona głównie efektem rozwoju rolnictwa i handlu, w którym przeważały produkty rolne, ale także przesiedleń w wyniku konfliktów. Szybki rozwój miast, szczególnie wzdłuż wybrzeża, także był przyczyną masowego wylesiania.

# Odnawialne źródła energii

```
wykres_renewable <- dane %>%
  group_by(kontynent, rok) %>%
  summarize(mean_ren = mean(renewable))
ggplot(wykres_renewable, aes(x = as.factor(kontynent), y = mean_ren, color = kontynent)) +
  geom_boxplot() +
  theme_bw() +
  theme(legend.position="none") +
  labs(title = "Średnia wartość użycia źródeł odnawialnej energii na kontynentach",
       subtitle = "średnia z lat 1990 - 2014",
       color = "Kontynent") +
  xlab("Kontynent") +
  ylab("Użycie źródeł odnawialnej energii") +
  theme(axis.text=element_text(size=12),
        plot.title = element_text(size = 14, face = "bold"),
        axis.title=element_text(size=12))
```

## Średnia wartość użycia źródeł odnawialnej energii na kontynentach

średnia z lat 1990 - 2014



Na podstawie wykresu pudełkowego mogę stwierdzić, że kontynentem z najniższym użyciem odnawialnej energii jest Azja. Wykres tego kontynentu jest bardzo „płaski”, oznacza to, że dane są słabo zróżnicowane. Dodatkowo klasyfikuje się on nisko, co pokazuje, że użycie energii odnawialnej w Azji jest małe. Najciekawszym przypadkiem jest Europa. Z wykresu możemy zauważyć, że posiada państwa z niskim użyciem energii odnawialnej, ale również istnieje wartość najbardziej odstająca, czyli państwo z naprawdę dużym użyciem energii odnawialnej.

```
srednia_renewable <- mean(dane$renewable[dane$rok >= 1990 & dane$rok <= 2014])
najwyzsza_wartosc <- dane$kraj[which.max(dane$renewable)]
najwyzsza_wartosc
```

```
## [1] "Ghana"
```

Najwyższym średnim użyciem energii odnawialnej może pochwalić się Dania. Jak się okazuje, w 2021 r. energia elektryczna ze źródeł odnawialnych stanowiła 71,9% duńskiego krajowego zaopatrzenia w energię elektryczną. Jest to naprawdę imponujący wynik i wiele państw powinno brać przykład.

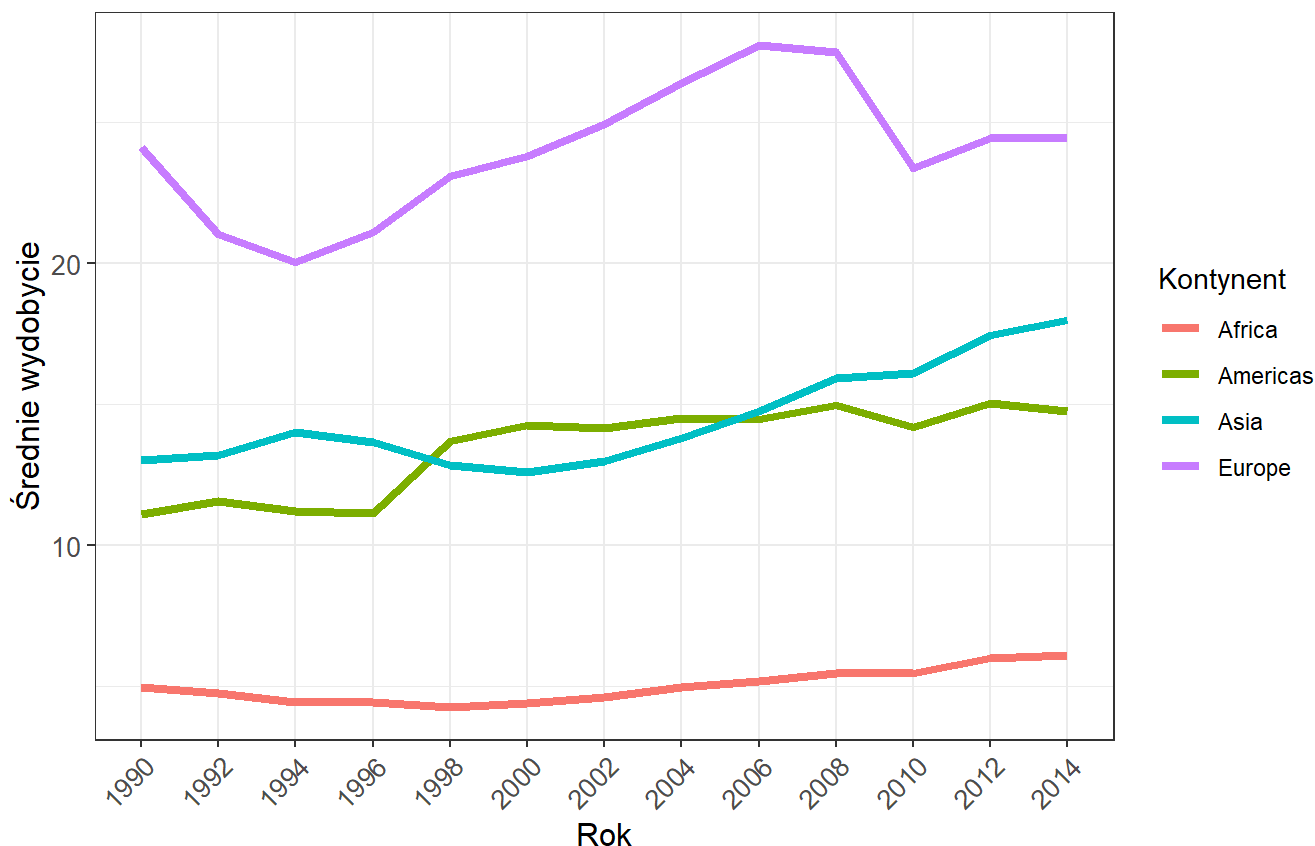
## Wydobycie surowców

```
wykres_materialy <- dane %>%
  group_by(kontynent, rok) %>%
  summarize(mean_materialy = mean(material, na.rm = TRUE))

ggplot(wykres_materialy, aes(x = rok, y = mean_materialy, color = kontynent, group = konty
nent)) +
  geom_line(size = 1.5) +
  theme_bw() +
  labs(title = "Średnie wydobycie surowców w latach 1990 - 2014",
       subtitle = "na wybranych kontynentach",
       color = "Kontynent") +
  xlab("Rok") +
  ylab("Średnie wydobycie") +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=14),
        axis.text.y = element_text(size=10),
        axis.text.x = element_text(angle = 45, hjust = 1),
        axis.title.x = element_text(size = 12),
        axis.title.y = element_text(size = 12),
        plot.title = element_text(size = 14, face = "bold"))
```

# Średnie wydobycie surowców w latach 1990 - 2014

na wybranych kontynentach



Powyższy wykres przedstawia średnie wydobycie surowców na kontynentach. Szczerze nie spodziewałam się, że Europa będzie tutaj tak wysoko, ponieważ myślałam, że na innych kontynentach wydobywa się więcej surowców. Nie spodziewałam się również, że Afryka znajdzie się najniżej. Afryka jest kontynentem bardzo bogatym w zasoby naturalne, których wielkość właściwie nie została oszacowana. Stosunkowo dobrze rozwiniętą gałęzią przemysłu jest wydobycie surowców, które w niektórych krajach wytwarza ponad 40% PKB. Jednakże gałąź ta jest bardzo wrażliwa na wahania cen. Dodatkowym problemem są niepokoje, powszechna korupcja i słabo rozwinięta infrastruktura.

## Hipotezy

### Hipoteza 1

Badanie mające na celu sprawdzenie zależności między wielkością PKB per capita a emisją dwutlenku węgla i emisją podtlenku azotu. Przeprowadzenie analizy, sprawdzającej, czy kraje z większym PKB per capita emitują więcej gazów cieplarnianych.

Osobiście uważam, że PKB może mieć wpływ na emisje dwutlenku węgla i podtlenku azotu. Myślę, że państwa posiadające wyższe PKB emitują więcej gazów cieplarnianych do atmosfery.

```

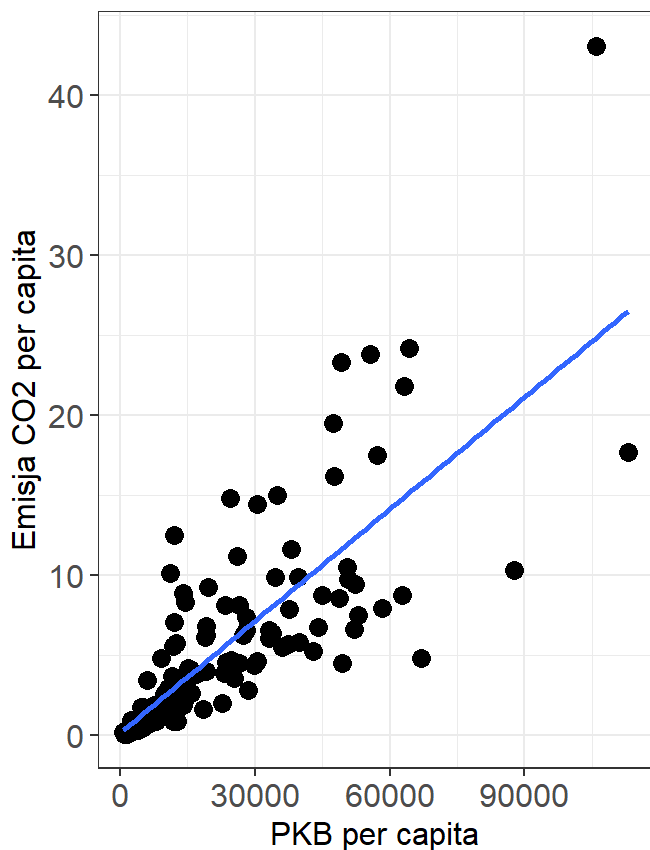
PKB_CO2 <- ggplot(dane %>% filter(rok == 2014), aes(x = PKB, y = CO2)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Zależność między emisją CO2
a PKB per capita w 2014 roku",
    x = "PKB per capita",
    y = "Emisja CO2 per capita",
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5)
  )

PKB_nitrous <- ggplot(dane %>% filter(rok == 2014), aes(x = PKB, y = nitrous_per_capita_in
dividual)) +
  geom_point(size = 3) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(
    title = "Zależność między emisją N2O
a PKB per capita w 2014 roku",
    x = "PKB per capita",
    y = "Emisja N2O per capita",
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 14, face = "bold",hjust = 0.5)
  )

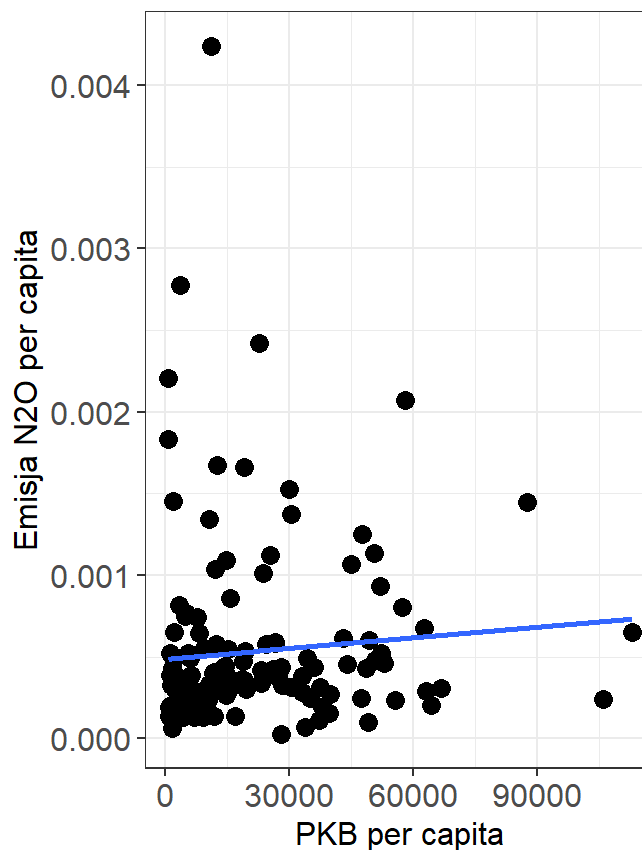
plot_grid(PKB_CO2, PKB_nitrous, nrow = 1)

```

**Zależność między emisją CO<sub>2</sub> a PKB per capita w 2014 roku**



**Zależność między emisją N<sub>2</sub>O a PKB per capita w 2014 roku**



Dodatkowo sprawdzę korelacje między zmiennymi, używając funkcji `cor()`.

```
cor(dane$PKB, dane$CO2)
```

```
## [1] 0.8358626
```

```
cor(dane$PKB, dane$nitrous_per_capita_individual)
```

```
## [1] 0.1097965
```

Wartość korelacji wynosząca 0,78 sugeruje umiarkowaną do silnej dodatnią zależność między zmiennymi PKB (produkt krajowy brutto) a emisją CO<sub>2</sub> na osobę. Oznacza to, że istnieje tendencja do wzrostu emisji dwutlenku węgla na osobę wraz ze wzrostem PKB. Na wykresie można zauważyć duże skupienie przy niskim poziomie PKB oraz niskiej emisji CO<sub>2</sub>. Linia trendu wskazuje wzrost PKB przy wzroście emisji dwutlenku węgla. Natomiast wartość korelacji między PKB a emisją podtlenku azotu wynosi 0,10, to oznacza to słabą zależność między tymi zmiennymi. Korelacja bliska zeru sugeruje, że zmiany w PKB nie są silnie związane ze zmianami w emisji podtlenku azotu. Linia trendu jest lekko pochylona i oznacza mały wzrost emisji podtlenku azotu względem wzrostu PKB. Podsumowując, okazuje się, że wartość PKB ma większy wpływ na emisje dwutlenku węgla niż emisję podtlenku azotu.

## Hipoteza 2

Badanie mające na celu sprawdzenie zależności między wielkością udziału rolnictwa w PKB per capita a emisją podtlenku azotu. Przeprowadzenie analizy, sprawdzającej, czy kraje z większym udziałem rolnictwa w PKB per capita emitują więcej podtlenku azotu.

Z informacji o podtlenku azotu dowiedziałam się, że duży wpływ na emisję podtlenku azotu na rolnictwo, dlatego spodziewam się, że wyższy udział rolnictwa będzie się wiązać z większą emisją N<sub>2</sub>O. Dodatkowo sprawdzę, jak zmieniał się udział rolnictwa w PKB na przestrzeni lat.



```

agriculture_data_1990 <- dane %>%
  filter(rok == 1990) %>%
  mutate(group = cut(agriculture, breaks = c(0, 10, 20, 30, 40, 50), labels = c("0-10%",
"11-20%", "21-30%", "31-40%", "41-50%"))) %>%
  na.omit()

agriculture_data_2014 <- dane %>%
  filter(rok == 2014) %>%
  mutate(group = cut(agriculture, breaks = c(0, 10, 20, 30, 40, 50), labels = c("0-10%",
"11-20%", "21-30%", "31-40%", "41-50%"))) %>%
  na.omit()

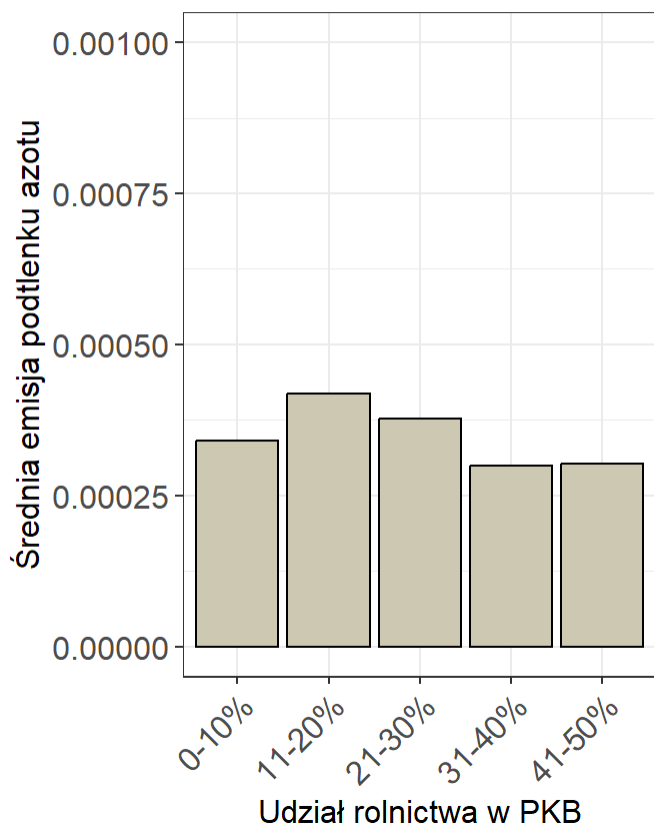
agriculture_nitr_1990 <- ggplot(agriculture_data_1990, aes(x = group, y = nitrous_per_capi
ta_individual)) +
  stat_summary(fun = mean, geom = "bar", fill = "cornsilk3", position = "dodge", colour =
"black") +
  labs(
    title = "Średnia emisja N2O w zależności
od udziału rolnictwa w PKB
w roku 1990",
    x = "Udział rolnictwa w PKB",
    y = "Średnia emisja podtlenku azotu"
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 12, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE)+
  ylim(0, 0.001)

agriculture_nitr_2014 <- ggplot(agriculture_data_2014, aes(x = group, y = nitrous_per_capi
ta_individual)) +
  stat_summary(fun = mean, geom = "bar", fill = "cornsilk3", position = "dodge", colour =
"black") +
  labs(
    title = "Średnia emisja N2O w zależności
od udziału rolnictwa w PKB
w roku 2014",
    x = "Udział rolnictwa w PKB",
    y = "Średnia emisja podtlenku azotu"
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 12, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE) +
  ylim(0, 0.001)

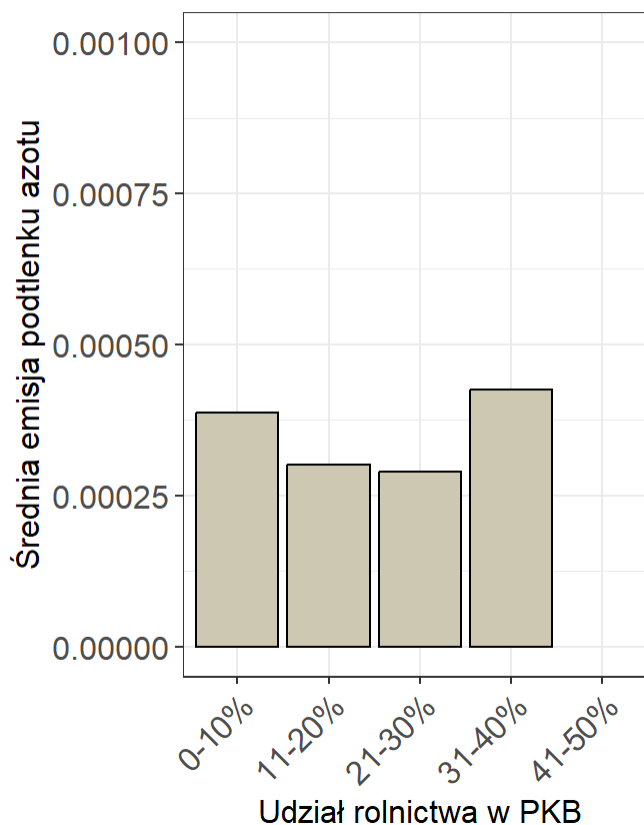
plot_grid( agriculture_nitr_1990, agriculture_nitr_2014, nrow = 1)

```

**Średnia emisja N<sub>2</sub>O w zależności od udziału rolnictwa w PKB w roku 1990**



**Średnia emisja N<sub>2</sub>O w zależności od udziału rolnictwa w PKB w roku 2014**



Co możemy wywnioskować z tego wykresu. Po pierwsze, w 2014 w zbiorze danych nie ma już państw, które mają udział rolnictwa w PKB powyżej 40%. W 1990 roku rozkład średniej emisji podtlenku azotu był bardzo podobny w zależności od udziału rolnictwa w PKB. Natomiast w 2014 roku możemy zauważyć, że państwa posiadające udział rolnictwa między 31% a 40% mają najwyższą emisję tego gazu cieplarnianego. Rolnictwo ma istotny wpływ na emisję podtlenku azotu, ponieważ jest on głównie związany z procesami związanej z nawożeniem azotowym i praktykami związanymi z gospodarką glebową. Okazuje się, że największy wpływ na jego emisję ma używanie nawozów azotanowych oraz hodowla zwierząt. Nawozy azotowe, takie jak mocznik, azotany i azotany amonowe, są powszechnie stosowane w rolnictwie, aby zwiększyć plony upraw poprzez dostarczenie roślinom azotu. Jednakże procesy związane z rozkładem tych nawozów w glebie mogą prowadzić do emisji N<sub>2</sub>O. Procesy nazywane denitryfikacją i nitryfikacją mogą prowadzić do emisji tego gazu do atmosfery. Natomiast jeśli chodzi o hodowlę zwierząt, w żołądkach zwierząt hodowlanych, takich jak krowy i owce, również przyczynia się do emisji N<sub>2</sub>O. Proces ten jest związany z metabolizmem azotowym zwierząt i może prowadzić do emisji N<sub>2</sub>O w postaci gazów wypuszczanych przez zwierzęta. Nie do końca zauważam tendencję wzrostową, jeżeli chodzi o większy udział rolnictwa i większą emisję N<sub>2</sub>O, dlatego sprawdzę jeszcze jedną rzecz.

```

dane_1990 <- dane %>%
  filter(rok == 1990)

dane_2014 <- dane %>%
  filter(rok == 2014)

mean_agriculture_1990 <- mean(dane_1990$agriculture)
mean_agriculture_2014 <- mean(dane_2014$agriculture)
mean_nitrous_1990 <- mean(dane_1990$nitrous_per_capita_individual)
mean_nitrous_2014 <- mean(dane_2014$nitrous_per_capita_individual)

wyniki_agri_nitr <- data.frame(
  Rok = c("1990", "2014"),
  Średnia_agriculture = c(mean_agriculture_1990, mean_agriculture_2014),
  Średnia_nitrous = c(mean_nitrous_1990, mean_nitrous_2014)
)

print(wyniki_agri_nitr)

```

```

##      Rok Średnia_agriculture Średnia_nitrous
## 1 1990          18.41313      0.0005564651
## 2 2014          10.74020      0.0005298706

```

Średni udział rolnictwa zmniejszył się, praktycznie o połowę. Mimo to, przez 24 lat rolnictwo przeszło wiele zmian zarówno na poziomie technologicznym, jak i społeczno-gospodarczym. Rozwój technologii w rolnictwie był niezwykle znaczący. Zastosowanie nowoczesnych maszyn rolniczych, technik uprawowych, nawożenia, ochrony roślin czy systemów nawadniających znacząco zwiększyło wydajność produkcji rolnej. Natomiast hodowla zwierząt bardzo zwiększyła się oraz stała się bardziej zintensyfikowana, co oznacza większe skupienie się na zwiększeniu produkcji mięsa, mleka, jajek i innych produktów zwierzęcych. Wykorzystuje się bardziej wydajne rasy zwierząt, a także wprowadza się nowoczesne technologie w żywieniu, hodowli i opiece nad zwierzętami. Prawdopodobnie takie czynniki, mimo zmniejszenia średniego udziału rolnictwa na świecie, wpływają na to, że średnia emisja tego gazu cieplarnianego się nie zmniejsza. Podsumowując, mogę stwierdzić, że na podstawie zmniejszającego się średniego udziału rolnictwa i mniej więcej podobnej średniej emisji podtlenku azotu, wielkość udziału rolnictwa w PKB może nie mieć wpływu na emisję tego gazu cieplarnianego. Wpływ na wielkość emisji mogą mieć inne czynniki takie jak np. zmiana sposobu uprawiania rolnictwa lub inne czynniki niezależne od rolnictwa.

## Hipoteza 3

Badanie mające na celu sprawdzenie zależności między wielkością udziałem przemysłu w PKB per capita a emisją CO<sub>2</sub>. Przeprowadzenie analizy, sprawdzającej, czy kraje z większym udziałem przemysłu w PKB per capita emitują więcej gazów cieplarnianych.

Jako iż bardzo dużo mówi się o tym, jak przemysł wpływa na emisję dwutlenku węgla, postanowiłam to sprawdzić. Osobiście uważam, że im państwo ma większy udział przemysłu w PKB, tym emituje więcej CO<sub>2</sub> do atmosfery.

```

industry_data_1990 <- dane %>%
  filter(rok == 1990) %>%
  mutate(group = cut(industry, breaks = c(0, 10, 20, 30, 40, 50, 60,70,80), labels = c("0-10%", "11-20%", "21-30%", "31-40%", "41-50%", "51-60%", "61-70%", "71-80%"))) %>%
  na.omit()

industry_data_2014 <- dane %>%
  filter(rok == 2014) %>%
  mutate(group = cut(industry, breaks = c(0, 10, 20, 30, 40, 50, 60,70,80), labels = c("0-10%", "11-20%", "21-30%", "31-40%", "41-50%", "51-60%", "61-70%", "71-80%"))) %>%
  na.omit()

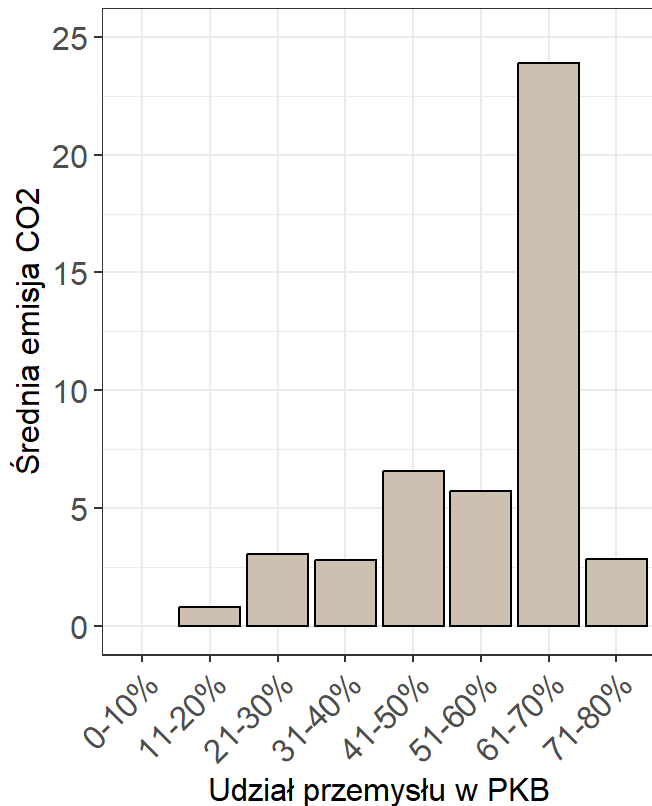
industry_CO2_1990 <- ggplot(industry_data_1990, aes(x = group, y = CO2)) +
  stat_summary(fun = mean, geom = "bar", fill = "antiquewhite3", position = "dodge", colour = "black") +
  labs(
    title = "Emisja CO2 w zależności
od udziału przemysłu w PKB
w roku 1990",
    x = "Udział przemysłu w PKB",
    y = "Średnia emisja CO2"
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE) +
  ylim(0, 25)

industry_CO2_2014 <- ggplot(industry_data_2014, aes(x = group, y = CO2)) +
  stat_summary(fun = mean, geom = "bar", fill = "antiquewhite3", position = "dodge", colour = "black") +
  labs(
    title = "Emisja CO2 w zależności
od udziału przemysłu w PKB
w roku 2014",
    x = "Udział przemysłu w PKB",
    y = "Średnia emisja CO2"
  ) +
  theme_bw() +
  theme(
    axis.text = element_text(size = 12),
    axis.title = element_text(size = 12),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_discrete(drop = FALSE) +
  ylim(0, 25)

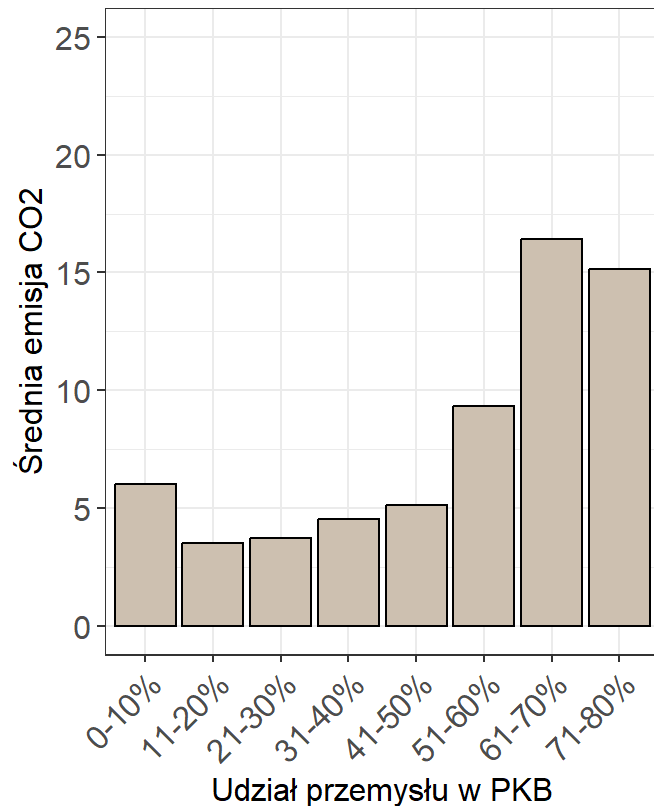
plot_grid(industry_CO2_1990, industry_CO2_2014, nrow = 1)

```

**Emisja CO<sub>2</sub> w zależności  
od udziału przemysłu w PKB  
w roku 1990**



**Emisja CO<sub>2</sub> w zależności  
od udziału przemysłu w PKB  
w roku 2014**



Początkowo, w 1990 roku, kraje charakteryzujące się wysokim udziałem przemysłu w PKB, oscylującym między 61 a 70%, były odpowiedzialne za znaczną część emisji dwutlenku węgla do atmosfery. Analizując dane z 2014 roku, można zauważyć, że wzrost udziału przemysłu w PKB wiązał się z istotnym zwiększeniem emisji dwutlenku węgla. To dobrze ilustruje, jak rozwój przemysłu miał wpływ na poziom emisji CO<sub>2</sub>. Pomimo postępu technologicznego, który bez wątpienia przyniósł wiele korzyści, wzrost produkcji w różnych sektorach przemysłu skutkował zwiększeniem emisji CO<sub>2</sub>. Szczególnie istotne jest, że rozwijające się kraje, dążące do zwiększenia produkcji, często polegają na energii pochodzącej z paliw kopalnych, co bezpośrednio przekłada się na większe emisje CO<sub>2</sub>. W sektorach przemysłu chemicznego i metalurgicznego procesy produkcyjne, wykorzystujące spalanie paliw kopalnych, wykazują tendencję do generowania znacznych ilości dwutlenku węgla. Podobnie, sektory ciężkiego przemysłu, takie jak produkcja stali, cementu czy chemikaliów, pozostają głównymi emitentami CO<sub>2</sub> z uwagi na intensywne procesy produkcyjne, które nadal opierają się głównie na spalaniu paliw kopalnych. Podsumowując, chociaż rozwój technologiczny przemysłu miał pozytywne skutki, towarzyszyły mu również negatywne konsekwencje dla środowiska, ze względu na zwiększenie emisji dwutlenku węgla do atmosfery. W dzisiejszych czasach z bardzo rozwiniętym przemysłem jego większy udział w PKB skutkuje większą emisją CO<sub>2</sub> do atmosfery.

## Hipoteza 4

Korelacja między współczynnikiem urbanizacji a emisją gazów cieplarnianych. Przeprowadzenie analizy, czy państwa, które posiadają wyższy współczynnik urbanizacji, emitują więcej gazów cieplarnianych.

Bardzo dużo mówi się o tym, ile miasta generują spalin i zanieczyszczeń, co może wiązać się też z emisją gazów cieplarnianych. Według mnie im wyższy stopień urbanizacji tym większa emisja gazów cieplarnianych, a na pewno CO<sub>2</sub>, co do podtlenku azotu jestem ciekawa wyników, gdyż nie znam wszystkich czynników wpływających na emisję tego gazu cieplarnianego.

```

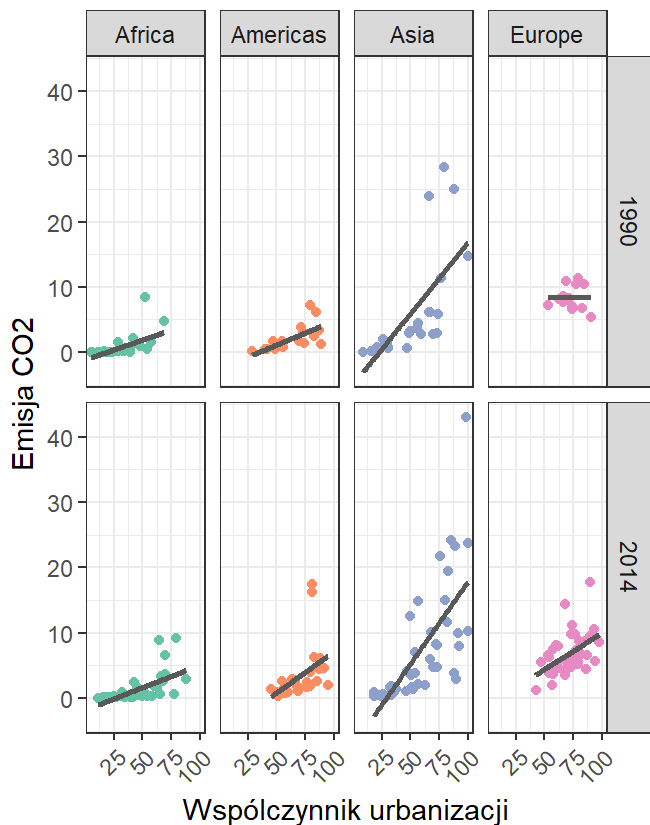
urban_CO2<-filter(dane, rok%in%c(1990, 2014)) %>%
  ggplot(aes(x = urban,
             y = CO2,
             col = kontynent)) +
  scale_color_brewer(palette = "Set2") +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, col = "gray35") +
  labs(x = "Współczynnik urbanizacji",
       y = "Emisja CO2",
       title="Zależność między współczynnikiem
urbanizacji a emisją CO2
w roku 1990 i roku 2014") +
  theme_bw()+
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  facet_grid(rok~kontynent)

urban_nitr<-filter(dane, rok%in%c(1990, 2014)) %>%
  ggplot(aes(x = urban,
             y = nitrous_per_capita_individual,
             col = kontynent)) +
  scale_color_brewer(palette = "Set2") +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, col = "gray35") +
  labs(x = "Współczynnik urbanizacji",
       y = "Emisja N2O",
       title="Zależność między współczynnikiem
urbanizacji a emisją N2O
w roku 1990 i roku 2014") +
  theme_bw()+
  theme(legend.position = "none",
        axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  facet_grid(rok~kontynent)

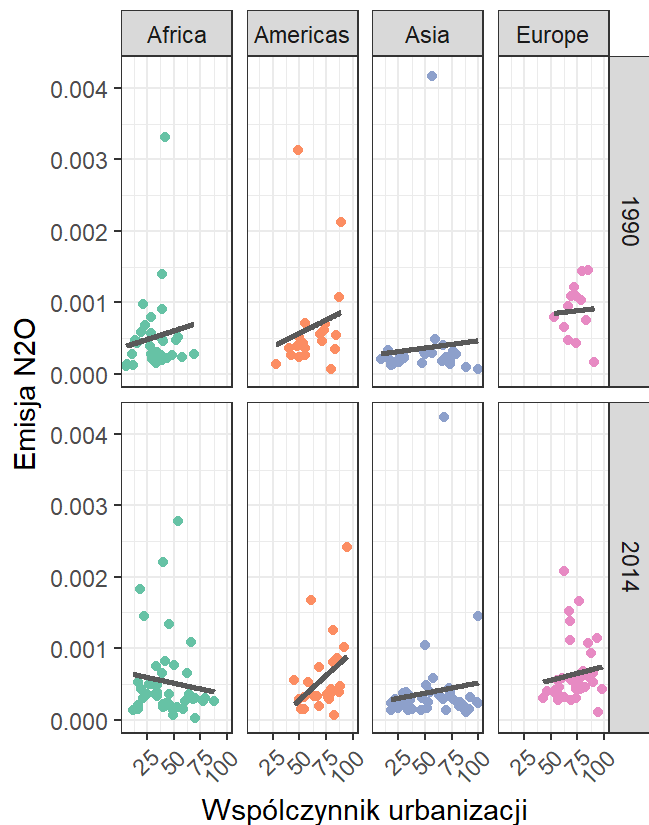
plot_grid(urban_CO2, urban_nitr, nrow = 1)

```

**Zależność między współczynnikiem urbanizacji a emisją CO2 w roku 1990 i roku 2014**



**Zależność między współczynnikiem urbanizacji a emisją N2O w roku 1990 i roku 2014**



Zacznę od emisji dwutlenku węgla. Na wykresie, dzięki linii trendu widzimy tendencję wzrostową, mówiącą o wpływie współczynnika urbanizacji na emisję CO2. Na przestrzeni lat możemy zaobserwować większy wpływ urbanizacji na emisję tego gazu cieplarnianego. Najbardziej widoczną zależność osiąga Azja, natomiast najmniej widoczne widzimy w Ameryce Północnej oraz Południowej. Jak można zauważyć, na obu kontynentach istnieją wartości odstające, zakładam, że są to państwa, w których nie tylko współczynnik urbanizacji wpływa na tak wysoką emisję dwutlenku węgla. Natomiast jeśli chodzi o emisję podtlenku potasu, tutaj korelacja jest mniejsza. Najbardziej widoczny jednak jest wzrost w Amerykach, co mnie zdziwiło, gdyż w przypadku emisji CO2 nie obserwujemy takiego wzrostu. Czyli prawdopodobnie istnieje jakiś inny czynnik, który wpływa na emisję podtlenku potasu, ale nie wpływa na emisję CO2. Sprawdzę teraz korelację między zmiennymi.

```
cor(dane$urban, dane$CO2)
```

```
## [1] 0.6350029
```

```
cor(dane$urban, dane$nitrous_per_capita_individual)
```

```
## [1] 0.1696858
```

Wysoka wartość współczynnika korelacji (0,56) pomiędzy wskaźnikiem urbanizacji a emisją CO2 wskazuje na umiarkowaną dodatnią zależność między tymi dwoma zmiennymi. Oznacza to, że istnieje pewna tendencja wzrostu emisji CO2 wraz ze wzrostem stopnia urbanizacji. W miastach często jest większe zużycie energii ze względu na transport, ogrzewanie, chłodzenie budynków itp. Wzrost urbanizacji może prowadzić do zwiększenia emisji CO2. Korelacja o wartości 0,16 między stopniem urbanizacji a emisją podtlenku azotu sugeruje umiarkowany, ale nieznaczący związek pomiędzy tymi dwoma zmiennymi. Istnieją oczywiście wartości odstające, które oczywiście mnie zaciekały.

```

najwyzsza_emisja_CO2 <- dane %>%
  filter(rok == 2014) %>%
  group_by(kontynent) %>%
  filter(CO2 == max(CO2)) %>%
  select(kontynent, kraj, CO2)

urban_CO2 <- dane %>%
  filter(rok == 2014) %>%
  group_by(kontynent, kraj) %>%
  filter(CO2 == max(CO2)) %>%
  select(kontynent, kraj, urban) %>%
  distinct()

max_emisja_CO2 <- najwyzsza_emisja_CO2 %>%
  left_join(urban_CO2, by = c("kontynent", "kraj"))

max_emisja_CO2

```

kontynent <chr>	kraj <chr>	CO2 <dbl>	urban <dbl>
Africa	Libya	9.23	79.0
Europe	Luxembourg	17.70	89.9
Asia	Qatar	43.10	98.9
Americas	USA	17.50	81.5
4 rows			

Możemy zaobserwować państwa z dużą emisją dwutlenku węgla i współczynnikiem urbanizacji wysokim i większym niż 50%. Bardzo zaintrygowało mnie pojawienie się Luksemburgu, jako państwa w Europie z największym współczynnikiem urbanizacji i największą emisją dwutlenku węgla. Okazuje się, że Luksemburg, choć jest niewielkim krajem pod względem powierzchni i populacji, jest jednym z najbardziej uprzemysłowionych krajów w Europie. Luksemburg jest jednym z najbogatszych krajów na świecie pod względem PKB per capita. Wysoki poziom życia i gospodarka, która opiera się na wielu sektorach, mogą wiązać się z większym zużyciem energii i zasobów, co może przyczynić się do wzrostu emisji CO2 na mieszkańca. Tak jak wcześniej mówiłam w Amerykach, nie obserwuje się raczej aż takiego wpływu współczynnika urbanizacji na emisję dwutlenku węgla, Trynidad i Tobago znalazł się na liście, ponieważ jest to państwo znane z bardzo dużego wydobycia ropy naftowej, która ma wpływ na emisję dwutlenku węgla, podobnie oczywiście sprawa wygląda z Katar, państwa znanego z dużego wydobycia ropy naftowej i gazu ziemnego, ale o wydobyciu więcej później.



```

najwyzsza_emisja_nitr <- dane %>%
  group_by(kontynent) %>%
  filter(nitrous_per_capita_individual == max(nitrous_per_capita_individual)) %>%
  select(kontynent, kraj, nitrous_per_capita_individual)

urban_nitr <- dane %>%
  group_by(kontynent, kraj) %>%
  filter(nitrous_per_capita_individual == max(nitrous_per_capita_individual)) %>%
  filter(rok == max(rok)) %>%
  select(kontynent, kraj, urban) %>%
  distinct()

max_emisja_nitr <- najwyzsza_emisja_nitr %>%
  left_join(urban_nitr, by = c("kontynent", "kraj"))

max_emisja_nitr

```

kontynent <chr>	kraj <chr>	nitrous_per_capita_individual <dbl>	urban <dbl>
Americas	Paraguay	0.003128079	48.7
Africa	Cameroon	0.004765625	42.0
Europe	Ireland	0.002989130	58.7
Asia	Mongolia	0.004244898	57.1
4 rows			

Wszystkie te państwa łączy najwyższa na kontynencie emisja podtlenku azotu oraz raczej średnia wartość współczynnika urbanizacji (średnia wynosi 58,66). Dlatego mogę stwierdzić, że współczynnik urbanizacji nie wpływa na emisję podtlenku potasu. Pozostaje mi szukać innych czynników, które mogą wpływać na emisję tego gazu cieplarnianego. Podsumowując, uważam, że współczynnik urbanizacji wpływa na emisję dwutlenku węgla poprzez to, że w obszarach miejskich zużywa się więcej energii, istnieje większe natężenie ruchu oraz obszary miejskie zazwyczaj koncentrują przemysł. Jednak nie znalazłam powiązań między współczynnikiem urbanizacji a emisją podtlenku azotu.

## Hipoteza 5

Korelacja między odsetkiem terenów leśnych a emisją CO<sub>2</sub> i podtlenku azotu. Przeprowadzenie analizy, aby sprawdzić, czy istnieje tendencja, że kraje z mniejszym udziałem powierzchni leśnych generują więcej CO<sub>2</sub>.

Jak większość wie, lasy są bardzo istotne w kwestii emisji dwutlenku węgla, ponieważ podczas procesu fotosyntezy pochłaniają one ten gaz cieplarniany. Dlatego w przypadku dwutlenku węgla uważam, że państwa posiadające więcej terenów leśnych mogą emitować mniej tego gazu. Jeśli chodzi o podtlenek azotu, to jestem ciekawa, czy udział terenów leśnych może mieć też wpływ na jego emisję.

```

woods_data <- dane_2014 %>%
  mutate(group = cut(woods, breaks = c(0, 20, 40, 60, 80, 100), labels = c("0-20%", "21-40%", "41-60%", "61-80%", "81-100%"))) %>%
  na.omit()

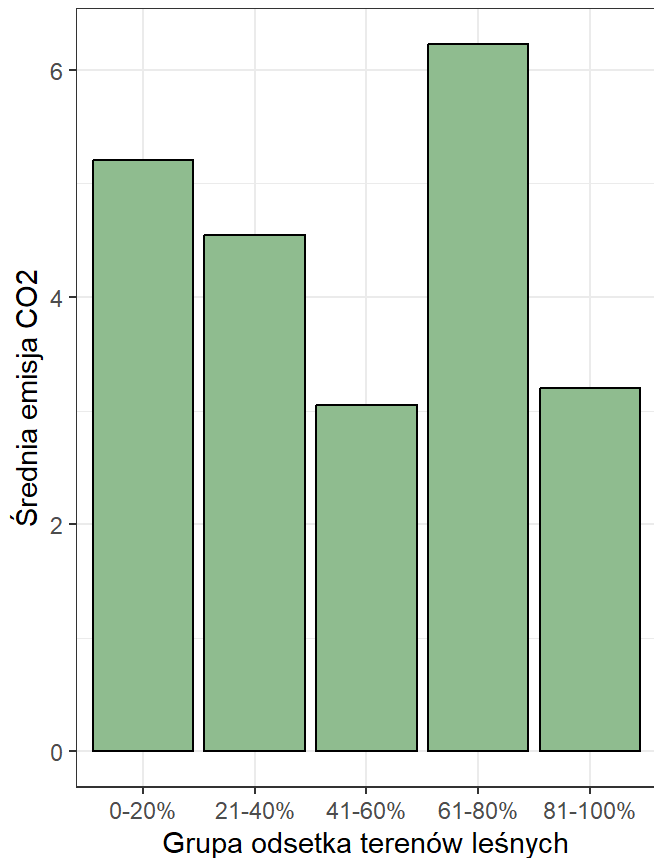
woods_CO2 <- ggplot(woods_data, aes(x = group, y = CO2)) +
  stat_summary(fun = mean, geom = "bar", fill = "darkseagreen", position = "dodge", colour = "black") +
  labs(
    title = "Średnia emisja CO2 w różnych grupach
odsetka terenów leśnych w 2014 roku",
    x = "Grupa odsetka terenów leśnych",
    y = "Średnia emisja CO2"
  ) +
  theme_bw() +
  theme(plot.title = element_text(size = 12, face = "bold", hjust = 0.5))

woods_nitr <- ggplot(woods_data, aes(x = group, y = nitrous_per_capita_individual)) +
  stat_summary(fun = mean, geom = "bar", fill = "darkseagreen", position = "dodge", colour = "black") +
  labs(
    title = "Średnia emisja N2O w różnych grupach
odsetka terenów leśnych w 2014 roku",
    x = "Grupa odsetka terenów leśnych",
    y = "Średnia emisja N2O"
  ) +
  theme_bw() +
  theme(plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  scale_y_continuous(labels = scales::number_format(scale = 1e3))

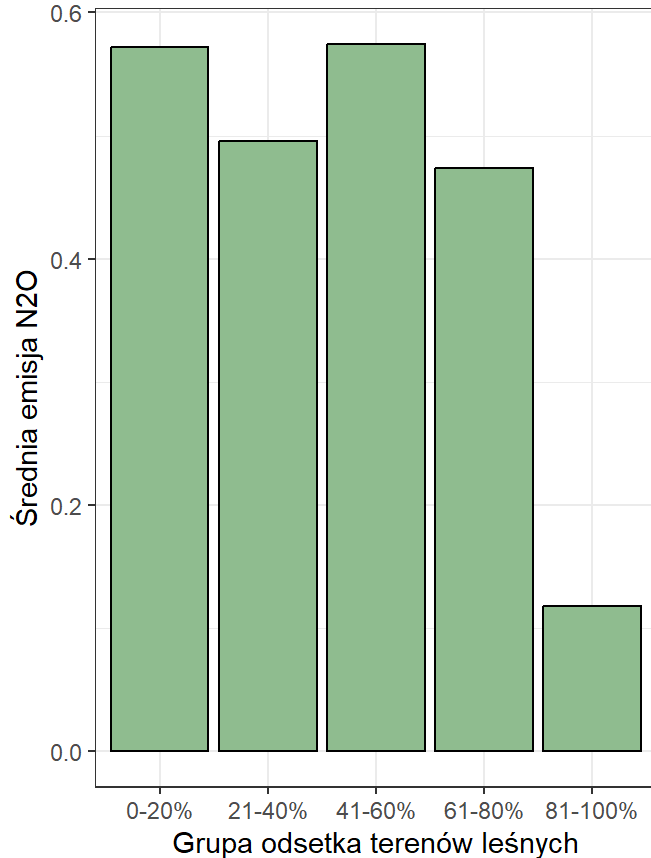
plot_grid(woods_CO2, woods_nitr, nrow = 1)

```

**Średnia emisja CO2 w różnych grupach odsetka terenów leśnych w 2014 roku**



**Średnia emisja N2O w różnych grupach odsetka terenów leśnych w 2014 roku**



Kolejne bardzo ciekawe wykresy, zacznę od tego z emisją dwutlenku węgla. Dane raczej pokazywałyby, że mniejsza wartość terenów leśnych wpływa na większą emisję dwutlenku węgla, jednak wszystko niszczy się przy odsetku 61-80% myślę, że mogą istnieć tam jakieś wartości odstające sprawiające, że istnieje tam taka wysoka emisja CO2. Postanowiłam posortować państwa z odsetkiem terenów leśnych w tym przedziale, jednak z największą emisją dwutlenku węgla i oto co otrzymałam.

```
countries_61_80_sorted <- woods_data %>%
  filter(group == "61-80%") %>%
  select(kraj, woods, CO2) %>%
  arrange(desc(CO2))

print(countries_61_80_sorted)
```

##	kraj	woods	CO2
## 1	Brunei	72.1	21.800
## 2	South Korea	65.1	11.600
## 3	Japan	68.5	9.880
## 4	Finland	73.6	8.730
## 5	Slovenia	62.0	6.550
## 6	Sweden	68.7	4.460
## 7	Brazil	60.5	2.590
## 8	Bhutan	71.2	1.330
## 9	Congo, Rep.	64.5	0.656
## 10	Lao	72.8	0.653
## 11	Zambia	61.8	0.277

Państwem prawdopodobnie odpowiedzialnym za tak drastyczny skok jest Brunei (jest to również państwo z najmniejszą populacją w moim zestawie danych), czym więc zajmuje się to państwo, że emituje tyle dwutlenku węgla do atmosfery. Brunei Darussalam jak się okazuje, jest znane z bardzo dużej emisji gazów cieplarnianych. Pomimo tego, że jest to jedno z najmniejszych państw w Azji, posiada jedno z najwyższych PKB per capita na świecie (w 2021 roku było to 23 miejsce na świecie). Co się tyczy emisji gazów cieplarnianych w Brunei, choć Brunei jest jednym z największych producentów ropy naftowej i gazu ziemnego, dlatego emisja CO2 w Brunei jest na wysokim poziomie.

Przejdę teraz do podtlenku azotu. Tutaj ciężko wypatrzyć jakąś zależność. Wykres utrzymuje wysokie wartości dotyczące emisji podtlenku azotu poza odsetkiem terenów leśnych wynoszących 81-100%. Czy może to oznaczać, że państwa z większą ilością terenów leśnych produkują mniejszą ilość gazów cieplarnianych do atmosfery?

```
countries_81_100_sorted <- woods_data %>%
  filter(group == "81-100%") %>%
  select(kraj, woods, nitrous_per_capita_individual) %>%
  arrange(desc(nitrous_per_capita_individual))

print(countries_81_100_sorted)
```

##	kraj	woods	nitrous_per_capita_individual
## 1	Gabon	91.6	2.639594e-04
## 2	Liberia	81.0	6.482301e-05
## 3	Equatorial Guinea	89.1	2.607692e-05

Postanowiłam przeprowadzić bardzo podobną analizę co wcześniej. Podane państwa posiadają naprawę dużą ilość terenów leśnych i małą emisją podtlenku azotu. Daje mi to pewną nadzieję, że może być to jeden z oczywiście wielu czynników wpływających na emisję podtlenku azotu. A co za tym idzie możliwy wniosek, że tereny leśne mogą być odpowiedzialne za mniejszą emisję tego gazu cieplarnianego. Z posiadanej przez większość wiedzy wiemy, że tereny zielone są ważnym elementem w cyklu życia dwutlenku węgla. Jednakże, kiedy lasy są niszczone poprzez wycinkę, spalanie lub degradację, uwolnione zostaje duże ilości zgromadzonego węgla, co przyczynia się do wzrostu emisji CO2. Czy może to też mieć wpływ na ten mniej popularny gaz cieplarniany, myślę, że tak. Podsumowując, mogę stwierdzić z posiadanej ogólnej wiedzy i wniosków wyciągniętych z analizy, że ilość terenów leśnych ma wpływ na emisję gazów cieplarnianych.

## Hipoteza 6

Związek między wydobywaniem surowców a emisją CO2 i N2O. Analiza mająca na celu zrozumienie, czy państwa o większym wydobywaniu surowców mają tendencję do większej emisji gazów cieplarnianych.

Tym razem przeprowadzę trochę inną analizę. Do obu wykresów wybiorę losowe państwa, które wydobywają dużo i mało surowców, następnie sprawdzę, czy istnieje zależność między wydobywaniem surowców a emisją poszczególnych gazów cieplarnianych. Według mnie większe wydobywanie surowców wiąże się z większą emisją gazów cieplarnianych.

Kraje wydobywające dużo surowców: Bahrajn, Arabia Saudyjska i Zjednoczone Emiraty Arabskie. Kraje wydobywające mało surowców: Nepal, Irak i Jemen. Aby ułatwić odczytywanie wykresu, państwa z dużym wydobywaniem będą oznaczone kolorem czerwonym a te z małym zielonym.

```

dane_material <- dane %>%
  filter(kraj %in% c("Bahrain", "Saudi Arabia", "UAE", "Nepal", "Iraq", "Yemen"))

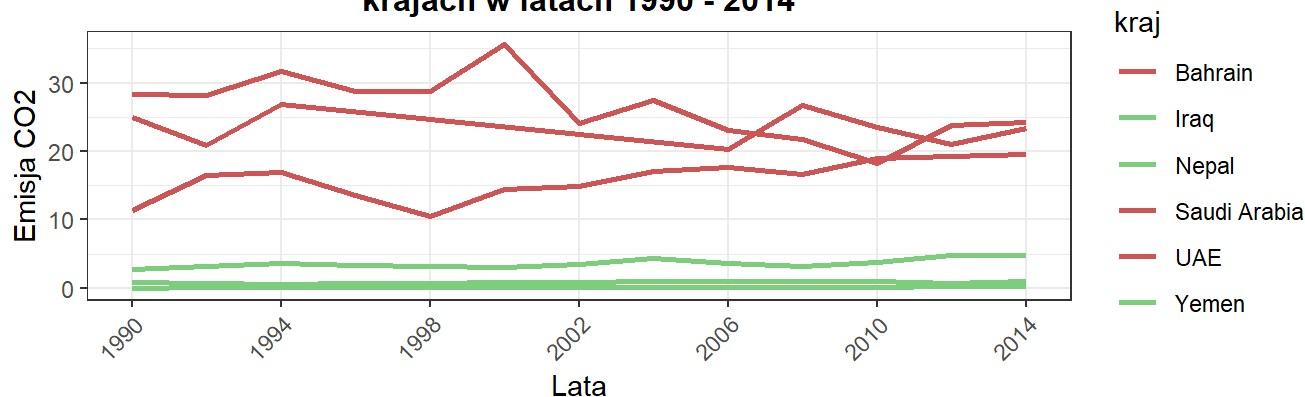
material_CO2 <- ggplot(data = dane_material, aes(x = rok, y = CO2, group = kraj, color = kraj)) +
  geom_line(linewidth = 1) +
  scale_x_discrete(breaks = c(1990, 1994, 1998, 2002, 2006, 2010, 2014)) +
  xlab("Lata") +
  ylab("Emisja CO2") +
  ggtitle("Emisja CO2 w wybranych\nkrajach w latach 1990 - 2014") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  scale_color_manual(values = c("indianred3", "palegreen3", "palegreen3", "indianred3", "indianred3", "palegreen3"))

material_nitr <- ggplot(data = dane_material, aes(x = rok, y = nitrous_per_capita_individual, group = kraj, color = kraj)) +
  geom_line(linewidth = 1) +
  scale_x_discrete(breaks = c(1990, 1994, 1998, 2002, 2006, 2010, 2014)) +
  xlab("Lata") +
  ylab("Emisja N2O") +
  ggtitle("Emisja N2O w wybranych\nkrajach w latach 1990 - 2014") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(size = 12, face = "bold", hjust = 0.5)) +
  scale_color_manual(values = c("indianred3", "palegreen3", "palegreen3", "indianred3", "indianred3", "palegreen3"))

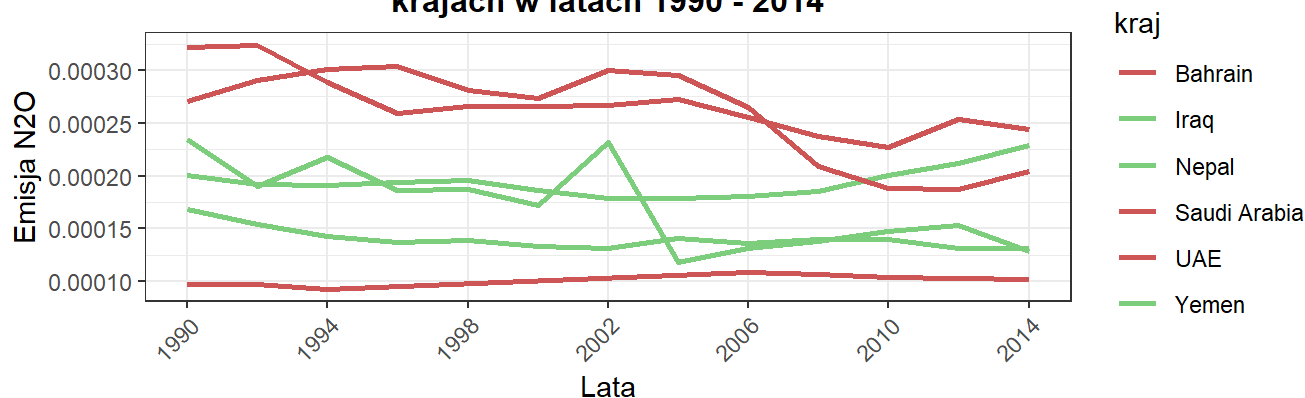
plot_grid(material_CO2, material_nitr, labels = NULL, nrow = 2)

```

**Emisja CO<sub>2</sub> w wybranych krajach w latach 1990 - 2014**



**Emisja N<sub>2</sub>O w wybranych krajach w latach 1990 - 2014**



Dzięki temu prostemu wykresowi możemy zobaczyć, jaką emisję dwutlenku węgla i podtlenku azotu posiadają państwa w zależności od ilości wydobytych surowców. Przy dwutlenku węgla sprawa jest naprawdę jasna, państwa z większą ilością wydobytych surowców emitują więcej dwutlenku węgla i na odwrót. Jednak przy emisji podtlenku potasu sprawa nie jest tak jasna. Bo na przykład Bahrain przy wysokim wydobyciu surowców emituje dużo dwutlenku węgla, jednak podtlenku potasu już nie, a patrząc na Zjednoczone Emiraty Arabskie widzimy już o wiele większą emisję podtlenku potasu. Jedynym wnioskiem, jaki przychodzi mi na myśl to, że emisja podtlenku potasu może zależeć od wydobywanych surowców i ich ilości. Po sprawdzeniu okazuje się, że Bahrain wydobywa przede wszystkim ropę naftową i gaz ziemny, tak samo, jak Zjednoczone Emiraty Arabskie. Prawdopodobnie istnieje jakiś inny czynnik wpływający na wartość emisji podtlenku potasu. Możliwe, że w tym przypadku Bahrain jest wartością odstającą lub jakiś inny czynnik wpływa na mniejszą emisję tego gazu cieplarnianego w tym państwie. Podsumowując, wielkość wydobycia surowców wpływa na emisję gazów cieplarnianych, na pewno większą korelację można zaobserwować w przypadku emisji dwutlenku węgla, jednak w przypadku podtlenku azotu, również widzę zależność, ale mniejszą. W przypadku emisji podtlenku azotu poza wartością odstającą, jaka jest Bahrain, widzę też większą emisję w państwach posiadających mniejsze wydobycie, czego w przypadku dwutlenku węgla nie obserwujemy.

## Hipoteza 7

Zależność między udziałem przemysłu i rolnictwa w PKB a wykorzystaniem odnawialnej energii. Który z tych czynników przyczynia się do większego wykorzystywania źródeł odnawialnej energii przez wybrane państwa.

Postanowiłam tę analizę przeprowadzić za pomocą podstawowych statystyk opisowych. Postanowiłam wybrać państwa z 2014 roku z najniższymi i najwyższymi wartościami dla przemysłu i rolnictwa. Następnie sprawdzę jakie użycie energii odnawialnej posiadają dane grupy, wyznaczając proste statystyki opisowe i następnie je analizując.

```

top_industry_countries <- dane %>%
  filter(rok == 2014) %>%
  top_n(40, industry) %>%
  arrange(desc(industry)) %>%
  select(kraj, renewable)

bottom_industry_countries <- dane %>%
  filter(rok == 2014) %>%
  arrange(industry) %>%
  slice_head(n = 40) %>%
  select(kraj, renewable)

top_agriculture_countries <- dane %>%
  filter(rok == 2014) %>%
  top_n(40, agriculture) %>%
  arrange(desc(agriculture)) %>%
  select(kraj, renewable)

bottom_agriculture_countries <- dane %>%
  filter(rok == 2014) %>%
  arrange(agriculture) %>%
  slice_head(n = 40) %>%
  select(kraj, renewable)

combined_summary_all <- bind_rows(
  bind_rows(
    dane %>%
      filter(rok == 2014, kraj %in% top_industry_countries$kraj) %>%
      summarize(
        srednia = mean(renewable, na.rm = TRUE),
        mediana = median(renewable, na.rm = TRUE),
        min = min(renewable, na.rm = TRUE),
        max = max(renewable, na.rm = TRUE),
        odchylenie_standardowe = sd(renewable, na.rm = TRUE)
      ) %>%
      mutate(Kategoria = "Najwyższy przemysł"),

    dane %>%
      filter(rok == 2014, kraj %in% bottom_industry_countries$kraj) %>%
      summarize(
        srednia = mean(renewable, na.rm = TRUE),
        mediana = median(renewable, na.rm = TRUE),
        min = min(renewable, na.rm = TRUE),
        max = max(renewable, na.rm = TRUE),
        odchylenie_standardowe = sd(renewable, na.rm = TRUE)
      ) %>%
      mutate(Kategoria = "Najniższy przemysł")
  ),

  bind_rows(
    dane %>%
      filter(rok == 2014, kraj %in% top_agriculture_countries$kraj) %>%

```

```

summarize(
  srednia= mean(renewable, na.rm = TRUE),
  mediana = median(renewable, na.rm = TRUE),
  min = min(renewable, na.rm = TRUE),
  max = max(renewable, na.rm = TRUE),
  odchylenie_standardowe = sd(renewable, na.rm = TRUE)
) %>%
mutate(Kategoria = "Najwyższe rolnictwo"),

dane %>%
  filter(rok == 2014, kraj %in% bottom_agriculture_countries$kraj) %>%
  summarize(
    srednia= mean(renewable, na.rm = TRUE),
    mediana = median(renewable, na.rm = TRUE),
    min = min(renewable, na.rm = TRUE),
    max = max(renewable, na.rm = TRUE),
    odchylenie_standardowe = sd(renewable, na.rm = TRUE)
  ) %>%
  mutate(Kategoria = "Najniższe rolnictwo")
)
)
combined_summary_all <- combined_summary_all %>%
  select(Kategoria, everything())

combined_summary_all

```

Kategoria <chr>	srednia <dbl>	mediana <dbl>	min <dbl>	max <dbl>	odchylenie_standardowe <dbl>
Najwyższy przemysł	27.17185	8.195	0	100.0	35.22470
Najniższy przemysł	42.19050	41.500	0	100.0	35.06083
Najwyższe rolnictwo	52.59683	60.450	0	100.0	37.46973
Najniższe rolnictwo	24.00015	15.300	0	97.7	26.36604
4 rows					

Najbardziej zastanawiające są te zera w wartości minimalnej, ale tak w moim zestawie danych istnieją państwa, które posiadają zerowe użycie energii odnawialnej. Niestety nie wiem, czy są to prawdziwe wartości, czy jednak takie są fakty, tak naprawdę minimalne wartości są dla nas najmniej interesujące. Zacznę od średniej, najwyższą średnią mogą pochwalić się państwa z najniższym udziałem przemysłu w PKB oraz państwa z najniższym udziałem rolnictwa w PKB, obie te wartości są dla mnie naprawdę ciekawe i niespodziewane. Najwyższą medianę wartości osiągają w tych samych kategoriach. Jeśli chodzi o największe wartości, to dołącza tutaj jeszcze najwyższe rolnictwo, jest to wartość odstająca i jest to Kenia. Kolejną statystyką jest odchylenie standardowe, w trzech grupach zauważamy wysokie i bardzo podobne odchylenie standardowe, oznacza to, że wartości w tej grupie są zróżnicowane. Jednak w grupie z najwyższym przemysłem odchylenie jest niskie, czyli wartości w tej grupie są bardzo podobne, ogólnie średnia w państwach z najwyższym przemysłem jest mała, w tych państwach użycie odnawialnych źródeł energii jest naprawdę niskie. Podsumowując państwa z najniższymi wartościami w przemyśle i rolnictwie mają najwyższe średnie użycie energii odnawialnej.

## Hipoteza 8

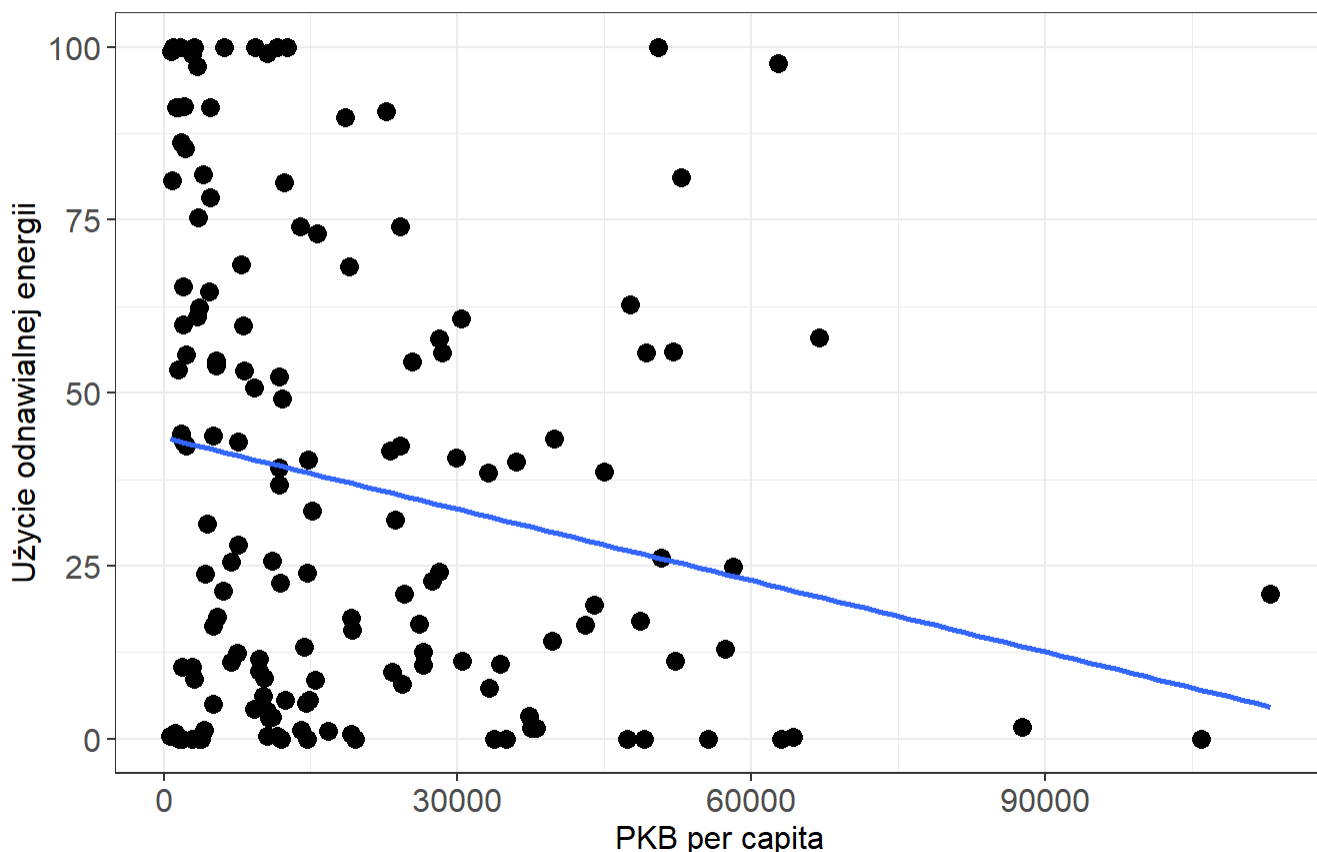


Zależność między PKB per capita a użytkowaniem odnawialnej energii. Czy państwa posiadające większą wartość produktu krajowego brutto starają się dbać o środowisko naturalne używając więcej odnawialnej energii.

Od początku projektu jestem bardzo optymistyczna i będę również teraz, uważam, że państwa z większym PKB wykorzystują więcej odnawialnej energii.

```
PKB_renewable_14 <- ggplot(dane %>% filter(rok == 2014), aes(x = PKB, y = renewable)) +  
  geom_point(size = 3) +  
  geom_smooth(method = "lm", se = FALSE) +  
  labs(  
    title = "Zależność między PKB per capita  
a użytkowaniem odnawialnej energii w 2014 roku",  
    x = "PKB per capita",  
    y = "Użycie odnawialnej energii",  
  ) +  
  theme_bw() +  
  theme(  
    axis.text = element_text(size = 12),  
    axis.title = element_text(size = 12),  
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5)  
  )  
PKB_renewable_14
```

**Zależność między PKB per capita  
a użytkowaniem odnawialnej energii w 2014 roku**



```
cor(dane$PKB, dane$renewable)
```

```
## [1] -0.257914
```

Wartość korelacji 0,16 wskazuje na stosunkowo niską dodatnią korelację między PKB per capita a wykorzystaniem odnawialnej energii. Korelacja o takiej wartości wskazuje na słaby związek między tymi dwiema zmiennymi. Można stwierdzić, że nie ma silnej zależności między wysokością PKB per capita a poziomem wykorzystania odnawialnych źródeł energii w badanych krajach. Mimo że wartość korelacji jest dodatnia, co sugeruje pewne powiązanie między tymi zmiennymi, to jednak jej niska wartość (0,16) świadczy o tym, że PKB per capita wpływa w małym stopniu na wykorzystanie odnawialnej energii. Z wykresu wynika, że państwa posiadające najwyższe użycie energii odnawialnej, wcale nie posiadają największego PKB per capita. Tak naprawdę to państwa z użyciem energii odnawialnej powyżej 20% posiadają PKB między 10000 a 60000, jest to wysokie PKB, jednak państwa o jeszcze wyższych wartościach nie osiągają takiego wyniku.

```
max_renewable <- dane_2014 %>%
  select(kraj, renewable, PKB) %>%
  arrange(desc(renewable)) %>%
  head(10)

print(max_renewable)
```

##	kraj	renewable	PKB
## 1	Albania	100.0	11600
## 2	Bhutan	100.0	9350
## 3	Ethiopia	100.0	1630
## 4	Iceland	100.0	50500
## 5	Lao	100.0	6150
## 6	Nepal	100.0	3150
## 7	Paraguay	100.0	12600
## 8	Congo, Dem. Rep.	99.9	999
## 9	Central African Republic	99.4	765
## 10	Namibia	99.1	10600

Dodatkowo z tej ramki danych możemy zauważyć, że wysokie użycie energii odnawialnej wcale nie musi zależeć od poziomu PKB. Średnie PKB w 2014 roku wyniosło 22342 i pare państw, które klasyfikują się do tej pierwszej dziesiątki osiągają PKB mniejsze (nawet znacznie mniejsze) od średniej. Podsumowując, niestety znowu byłam zbyt optymistyczna, ale zdałam sobie sprawę, że skoro państwa o wiele niższym PKB niż średnia potrafią zdać o środowisko, to dlaczego nie potrafią tego bardziej rozwinięte państwa. Wnioskiem z tego może być konieczność uwzględnienia różnych czynników, które mogą wpływać na stosunek między PKB a wykorzystaniem odnawialnych źródeł energii, nie ograniczając się jedynie do czynnika ekonomicznego.

## Zakończenie

Wybrałam temat projektu związany ze środowiskiem naturalnym i zmianami klimatu ze względu na moje głębokie zainteresowanie tym obszarem. W czasie analizy często kierowałam się optymistycznym podejściem (czasami nawet żalonym), co spowodowało, że kilka razy się zawiodłam. Głównym celem mojego badania było zbadanie, czy wybrane zmienne rzeczywiście mają związek z emisją gazów cieplarnianych oraz wykorzystaniem energii odnawialnej. W trakcie analizy potwierdziły się moje przypuszczenia, niektóre zmienne wykazywały istotny wpływ na emisję gazów cieplarnianych i użycie energii odnawialnej, podczas gdy inne nie miały takiego związku. Podsumowując, uważam, że podjęłam temat w interesujący sposób, co pozwoliło mi poszerzyć swoją wiedzę na temat złożoności relacji między czynnikami wpływającymi na środowisko.