

Akademia Górniczo-Hutnicza w Krakowie
Wydział Geologii, Geofizyki i Ochrony Środowiska



Sprawozdanie z Projektu 2
w ramach przedmiotu Przetwarzanie Danych Środowiskowych

autorzy: Maciej Bąk, Filip Giermek gr.1

Kraków/Olszyny, 04.05.2020

Przedmiot analizy

Przedmiotem analizy niniejszego projektu są tabele danych *dynow_t* oraz *dynow_p* pochodzące ze stacji meteorologicznej w Dynowie, składające się odpowiednio z wartości średniej temperatury oraz sumy opadów atmosferycznych dla poszczególnych miesięcy w latach 1951-2016.

1.1 Wczytanie danych i bibliotek

Wczytujemy odpowiednie biblioteki, które będą wykorzystywane w analizie. Do wczytania wykorzystujemy komendę *read.delim()*.

```
library(dplyr)

library(ggplot2)
library(tidyr)
dynow_temp<-read.delim("dynow_t.txt")
dynow_opady<-read.delim("dynow_p.txt")
```

1.2 Uzupełnienie brakujących wartości

W zaimportowanych pakietach danych znajdują się brakujące wartości. Zastępujemy je średnią arytmetyczną kolumn, w których dane braki występowały. Całość odbywa się w pętli.

```
#dla temperatury - "dynow_temp"
for(i in 2:13)
{
  tmp <- dynow_temp[,i] #dane z i-tej kolumny do tymczasowej zmiennej

  tmp[which(is.na(tmp))] <- mean(tmp, na.rm = T) #zmiana na średnią

  dynow_temp[,i] <- round(tmp, digits = 2) #powrót zmodyfikowanej kolumny na swoje miejsce
}

#dla opadów - "dynow_opady" - analogicznie jak wyżej
for(i in 2:13)
{
  tmp <- dynow_opady[,i]

  tmp[which(is.na(tmp))] <- mean(tmp, na.rm = T)

  dynow_opady[,i] <- round(tmp, digits = 2)
}
```

1.3 Połączenie baz danych o opadach i temperaturach

Aby móc w dalszej części wykonać analizę korelacji średniej temperatury i sumy opadów dla każdego miesiąca połączymy bazy danych. Najpierw zmienimy nazwy kolumn w bazie *dynow_temp* za pomocą funkcji *colnames()* dodając do każdego miesiąca '.temp', to samo dla bazy *dynow_opady* dodając '.opady' i łączymy bazy za pomocą funkcji *full.join()* zapisując powstałą bazę jako *dynow*.

```
colnames(dynow_temp)<-c("rok","I.temp","II.temp","III.temp","IV.temp","V.temp",
"VI.temp","VII.temp","VIII.temp","IX.temp","X.temp","XI.temp","XII.temp")

colnames(dynow_opady)<-c("rok","I.opady","II.opady","III.opady","IV.opady","V
.opady","VI.opady","VII.opady","VIII.opady","IX.opady","X.opady","XI.opady","
XII.opady")

dynow<-full_join(dynow_temp, dynow_opady, by="rok")
```

1.4 Utworzenie wykresów rozrzutu z trendem dla sumy opadów w miesiącu na przestrzeni lat

Do stworzenia wykresów rozrzutu wykorzystujemy pakiet *ggplot2* oraz zawartą w nim funkcję *ggplot()* oraz *geom_point()*. Aby dodać linię trendu zastosujemy funkcję *geom_smooth()*. W celu usprawnienia wykonywania wykresów zastosujemy pętlę *for*, której argument *i* będzie zmieniał się w przedziale od 2 do 13. Na początku do tymczasowej zmiennej *tmp* przypisujemy i-tą kolumnę ze zbioru danych *dynow_opady*. Kolejno tworzymy tymczasową ramkę danych *tmp_df* i do niej przypisujemy kolumnę z datą oraz wcześniej utworzoną zmienną *tmp* z danymi dla danego miesiąca. Następnie tworzymy wykres z tymczasowych zmiennych. Tytuł nadajemy z wykorzystaniem wcześniej utworzonego wektora *miesiace*. Na końcu zapisujemy wykres funkcją *ggsave()*. Zastosowanie funkcji *paste0* w ciele funkcji *ggsave()* pozwala na dynamiczną zmianę nazw poszczególnych plików. Wszystkie zastosowane pętle w późniejszych etapach przetwarzania danych działają na podobnej zasadzie.

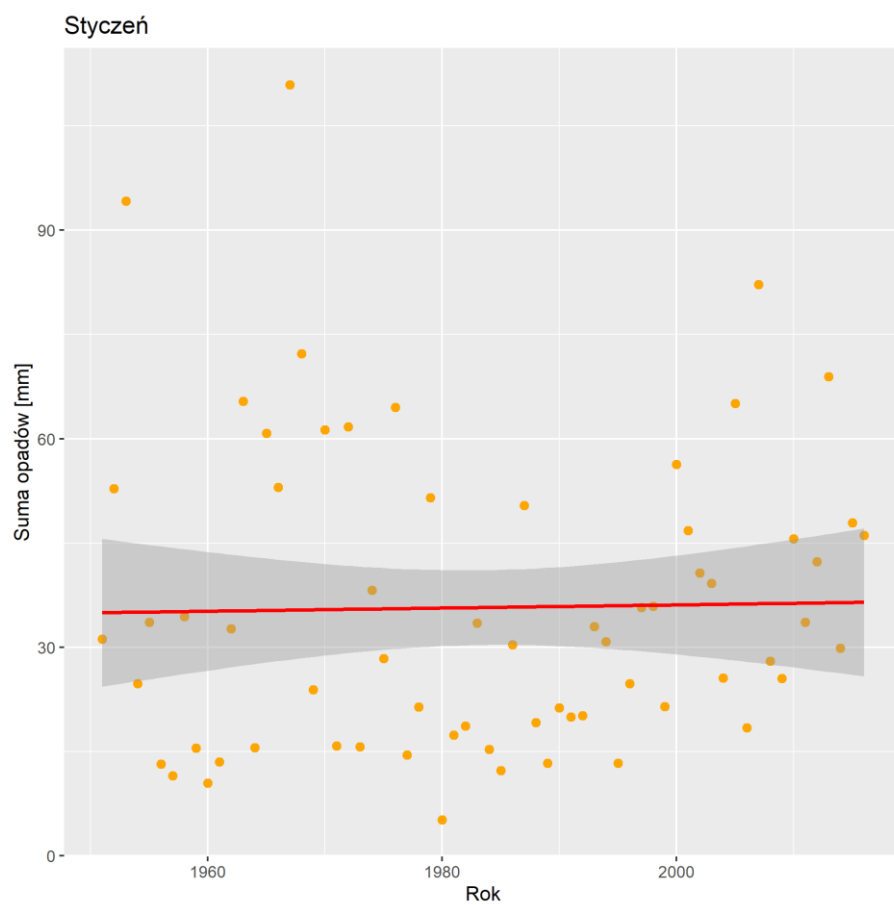
```
rok <- dynow_opady$rok
miesiace <- c("Styczeń", "Luty", "Marzec", "Kwiecień", "Maj", "Czerwiec",
"Lipiec", "Sierpień", "Wrzesień", "Październik", "Listopad", "Grudzień")
```

```
for(i in 2:13)
{
  tmp <- dynow_opady[,i] #dane z itej kolumny (kolejne miesiące)
```

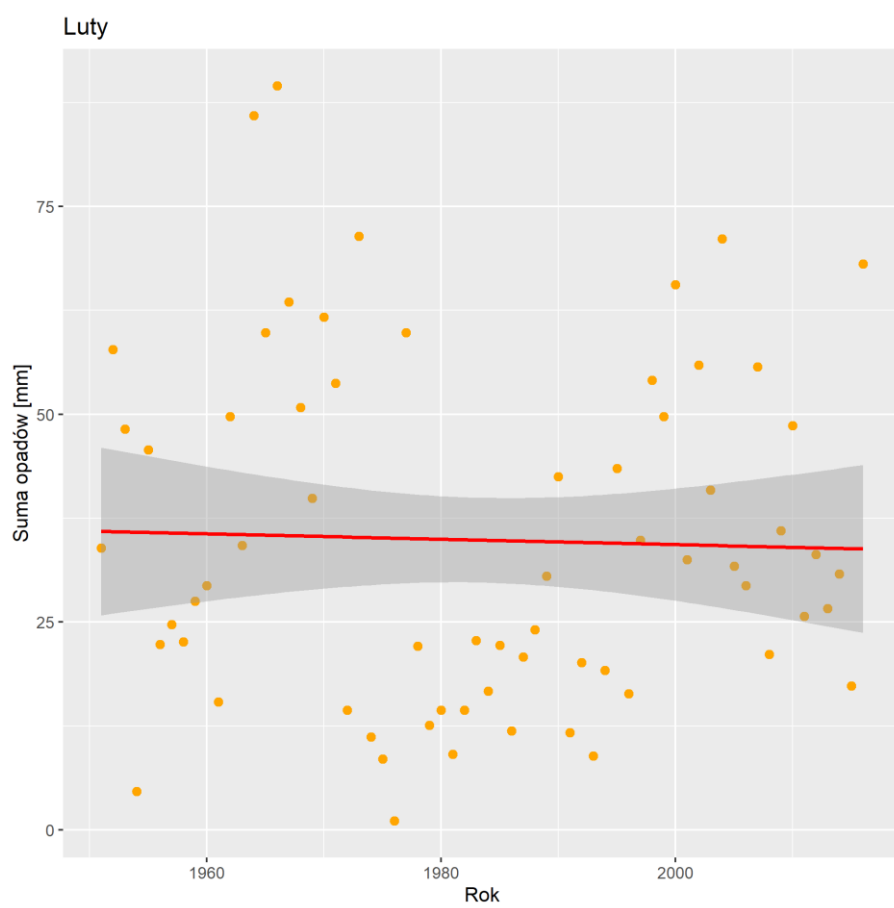
```
  tmp_df <- data.frame(rok, tmp) #tymczasowy df z rokiem i danymi dla
kolejnych miesięcy
```

```
  tmp_plot <-ggplot(tmp_df, aes(rok, tmp)) + #wykres
    geom_point(color="orange", size=2)+
    geom_smooth(method="lm", color="red")+
    xlab("Rok")+
    ylab("Suma opadów [mm])+
    ggtitle(miesiace[i-1])
```

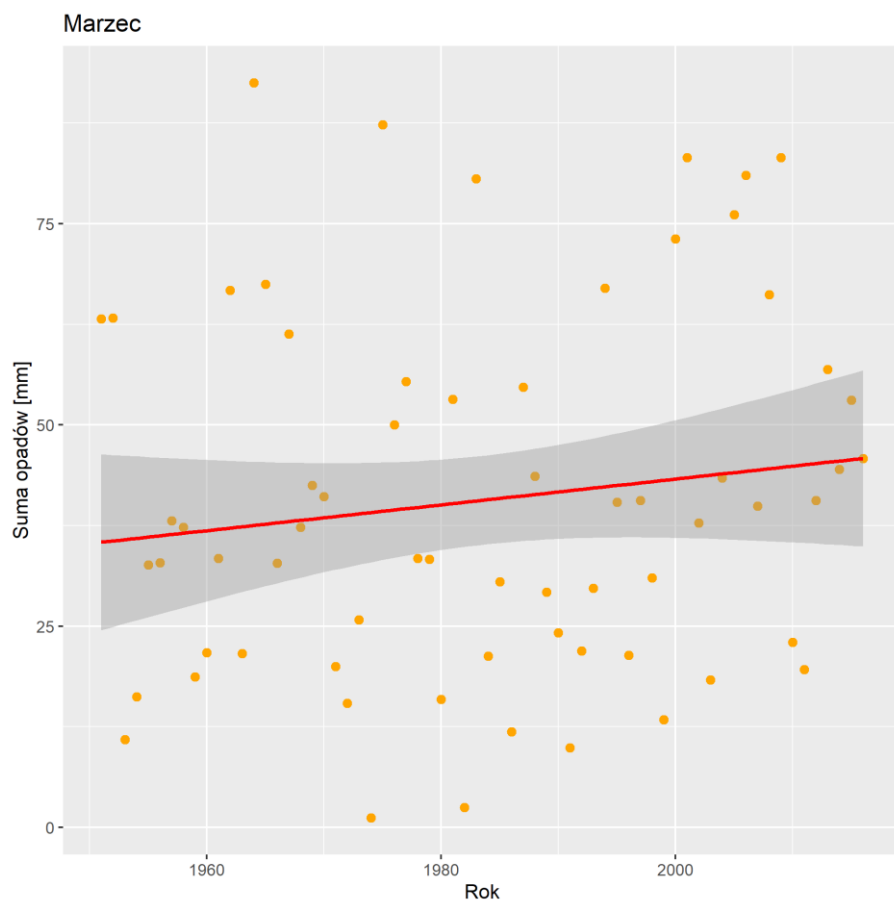
```
  ggsave(tmp_plot, filename = paste0("opady.trend_",miesiace[i-1],".png"))
}
```



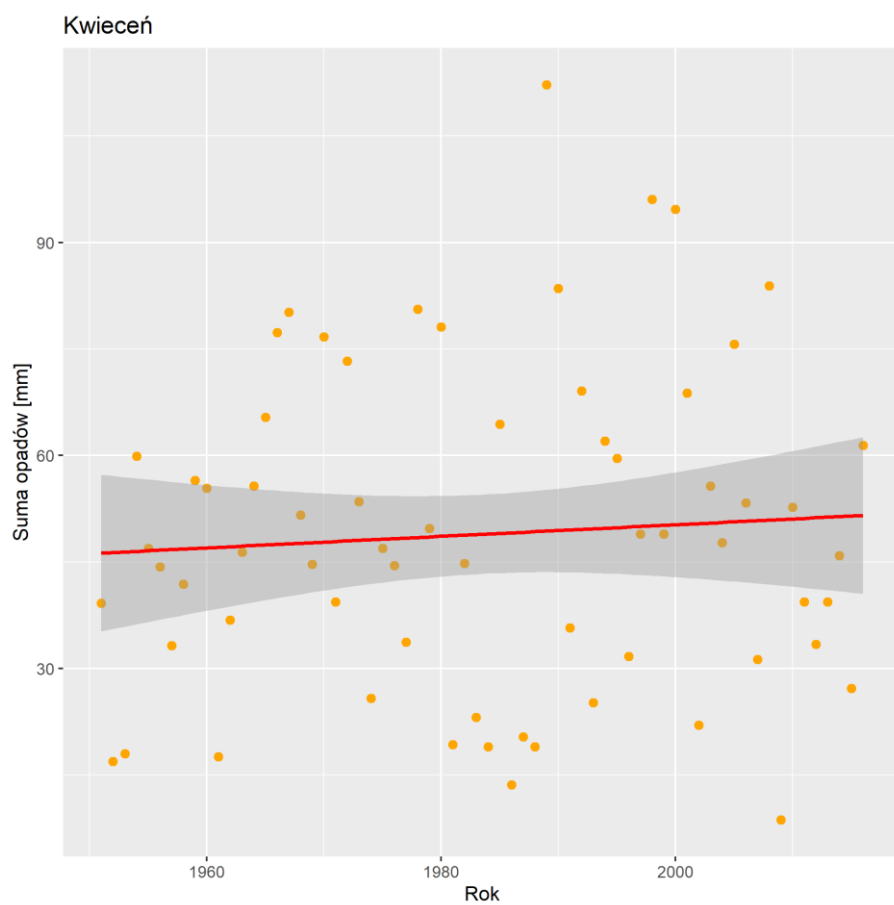
wykr. 1.4.1



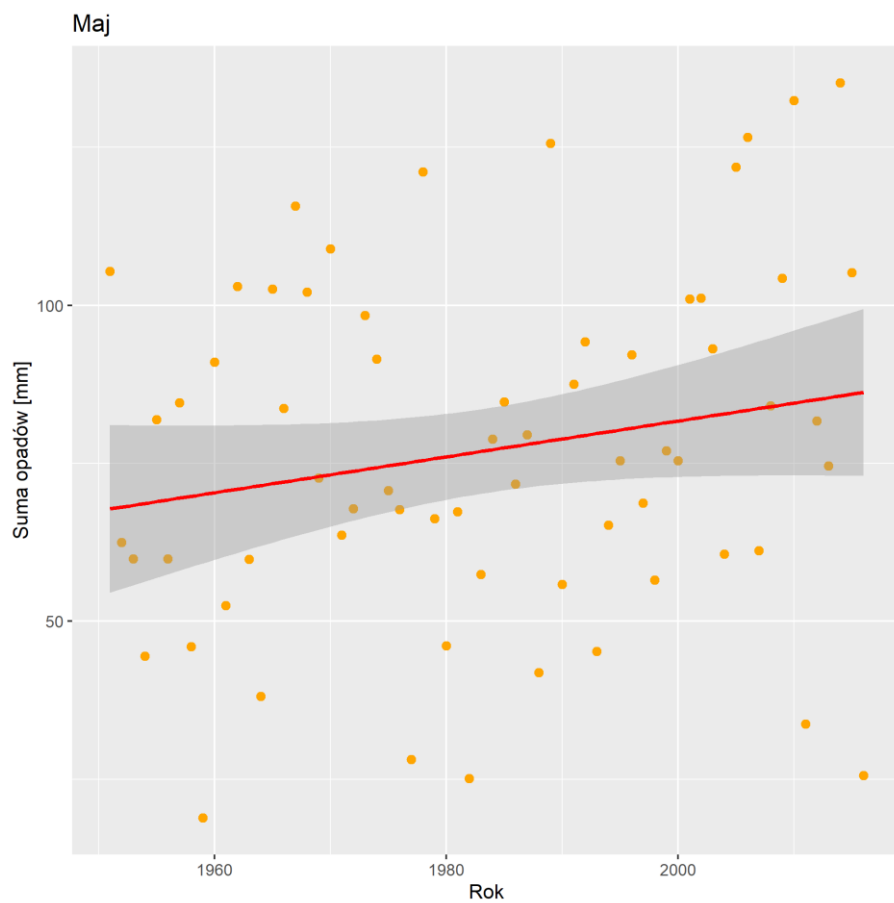
wykr. 1.4.2



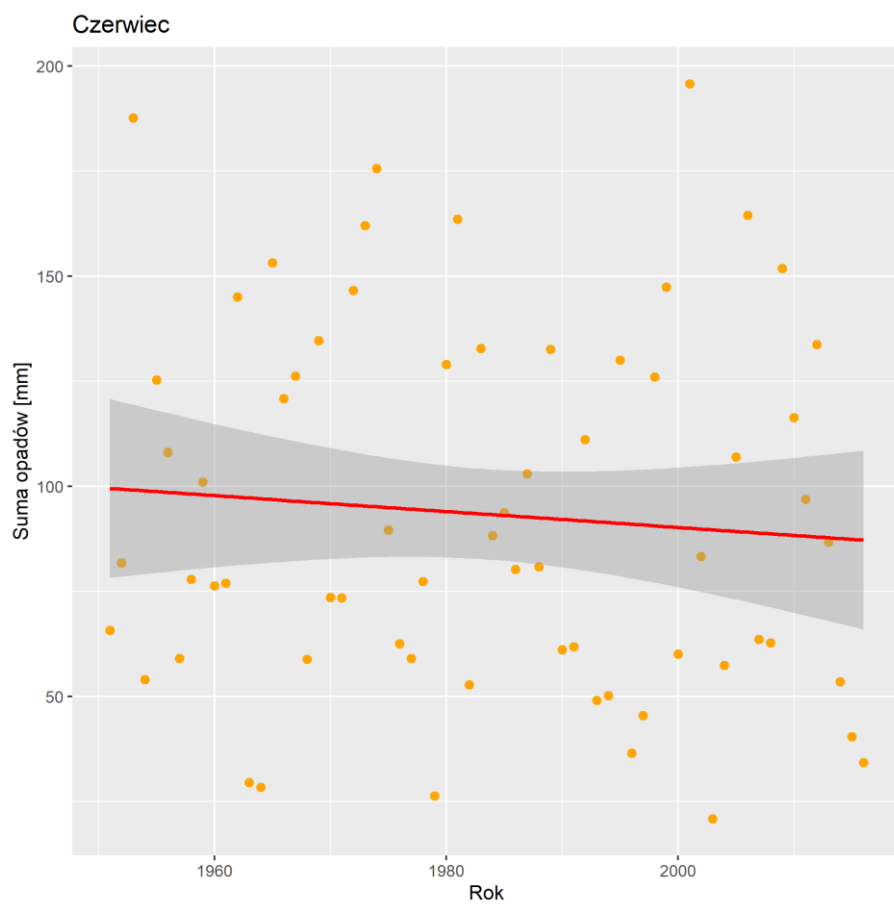
wykr. 1.4.3



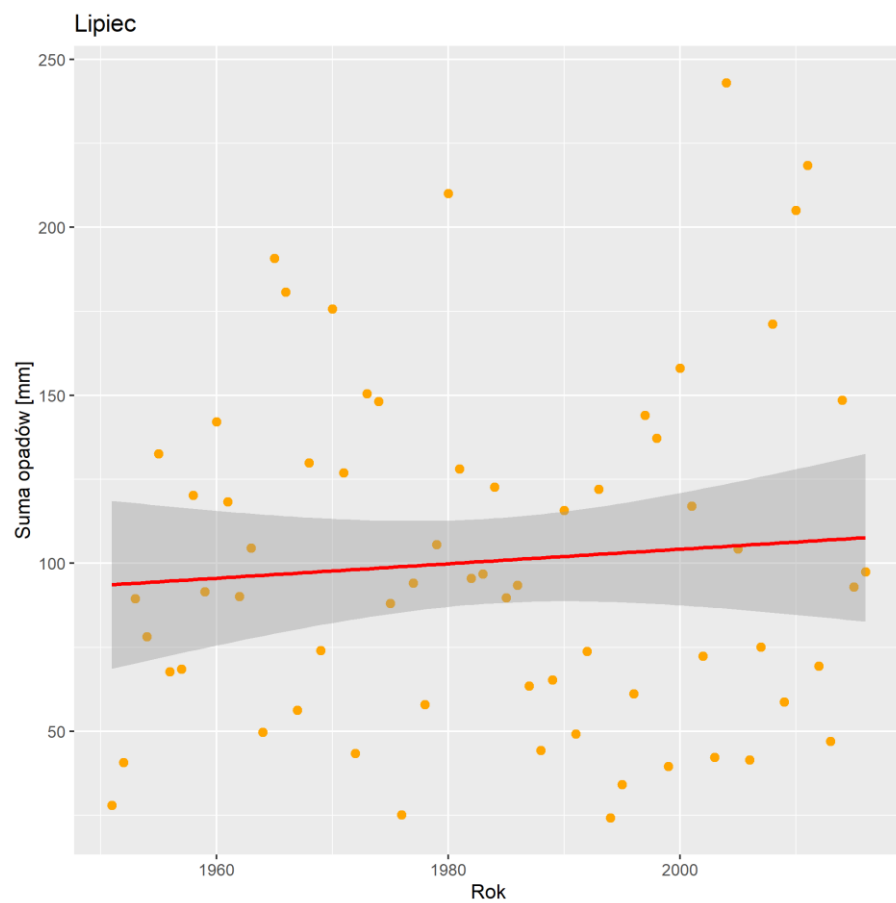
wykr. 1.4.4



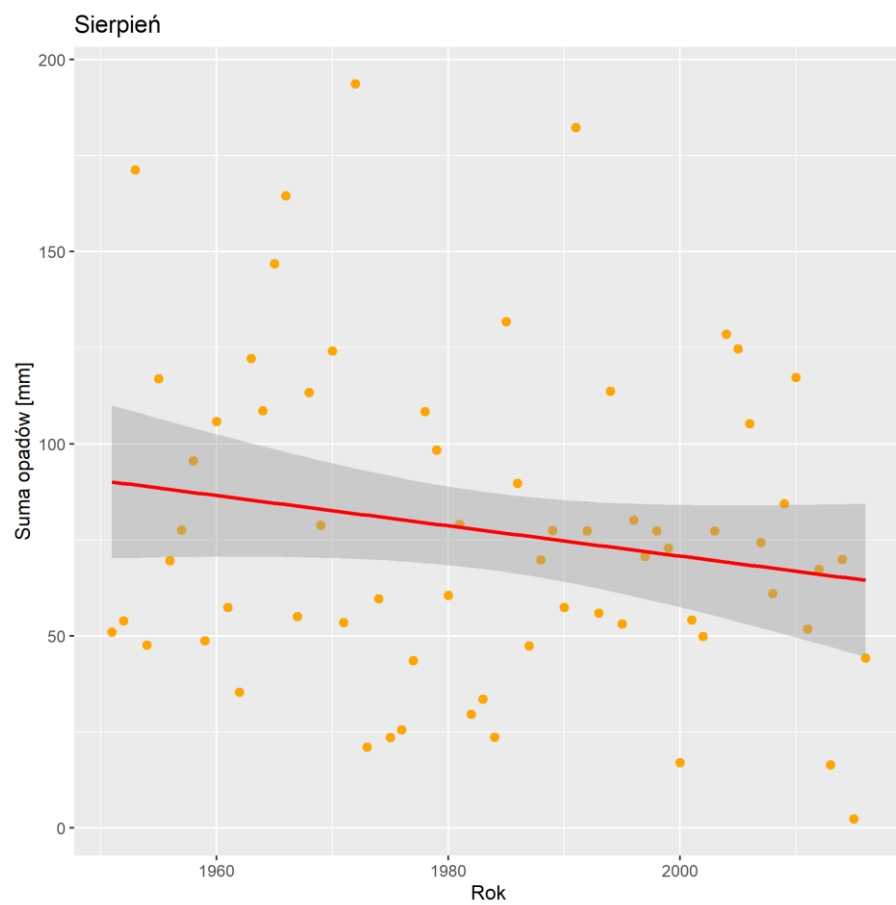
wykr. 1.4.5



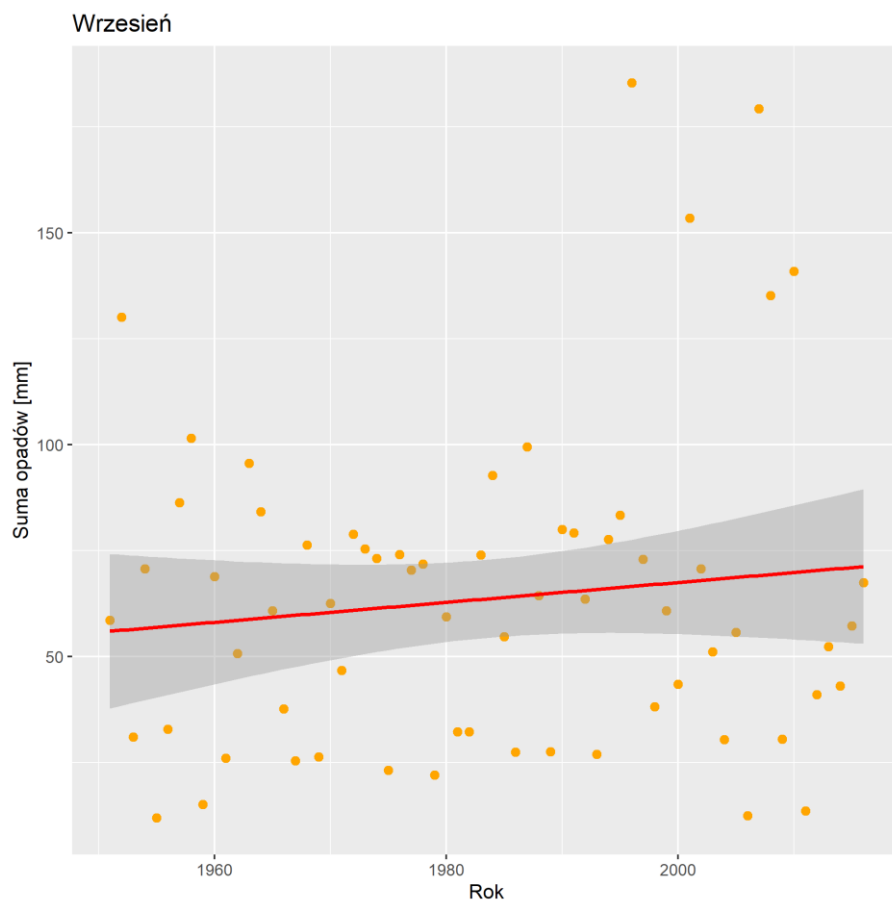
wykr. 1.4.6



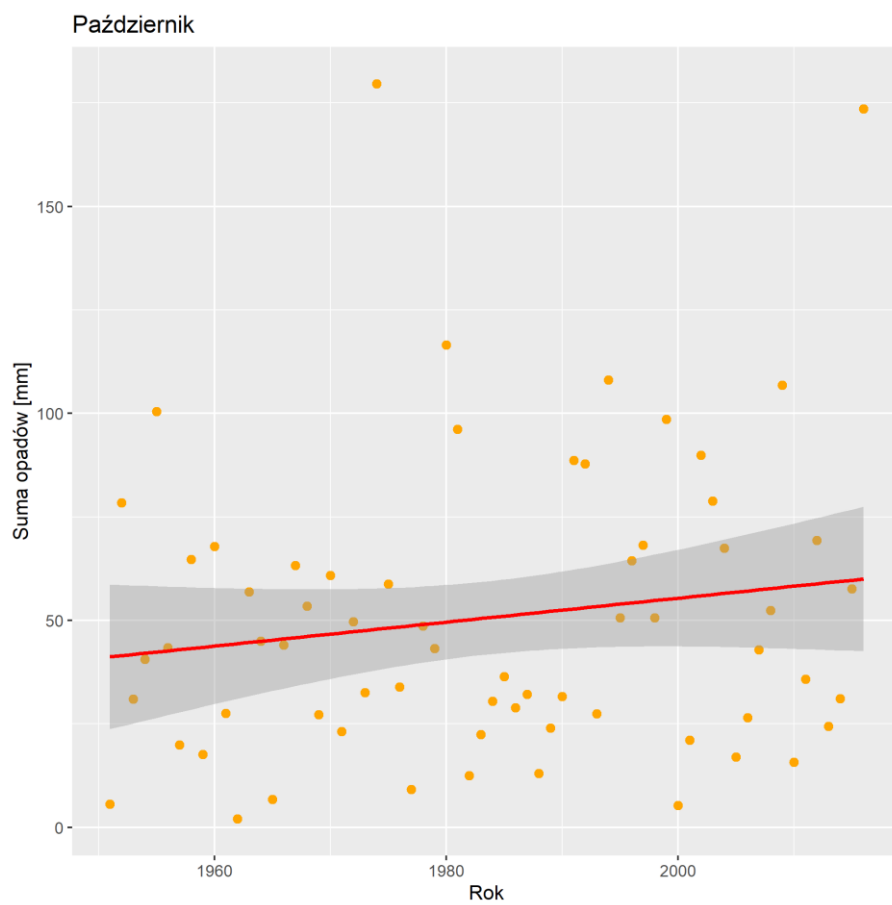
wykr. 1.4.17



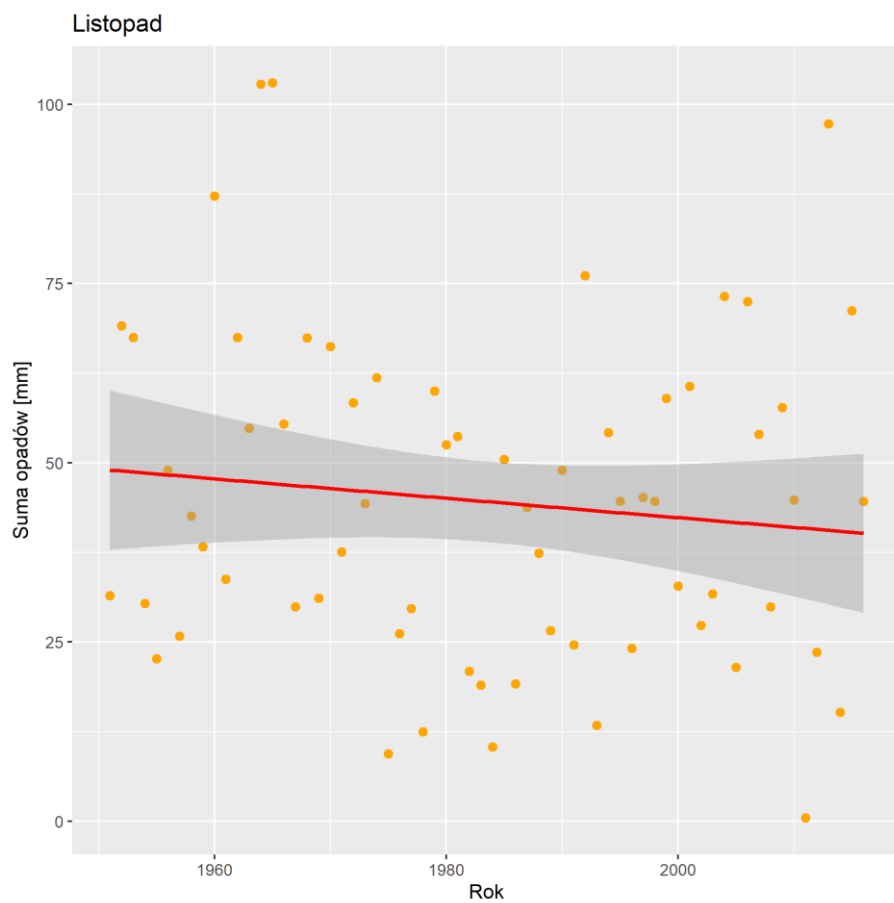
wykr. 1.4.8



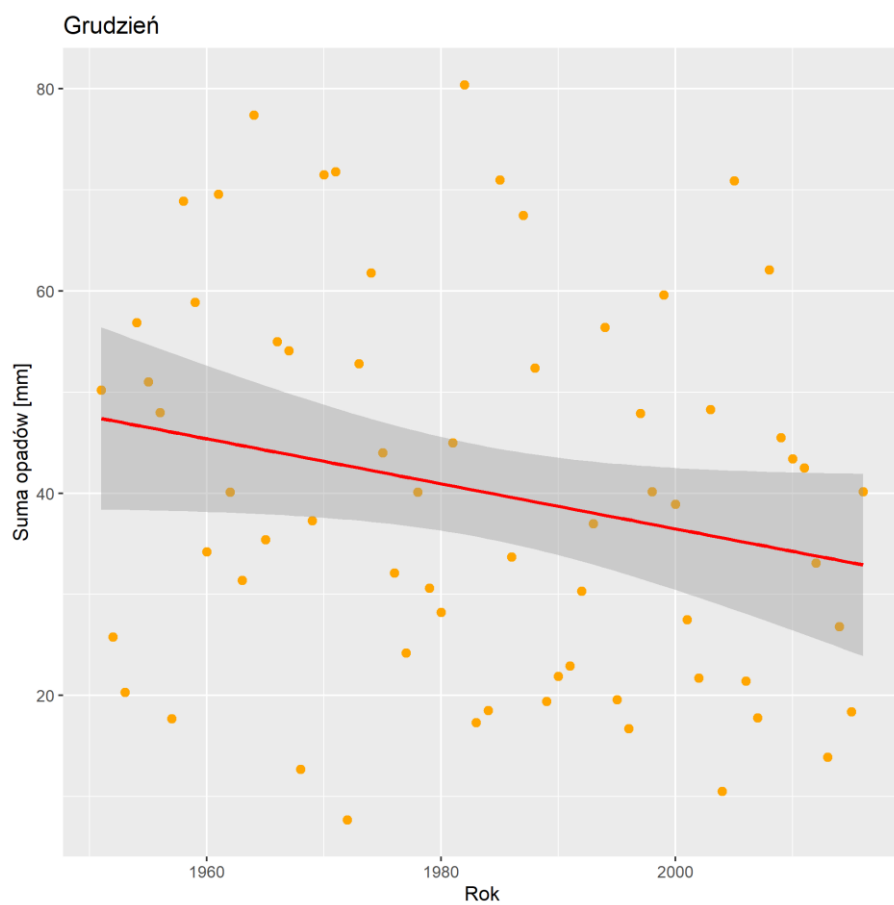
wykr. 1.4.9



wykr. 1.4.10



wykr. 1.4.11



wykr. 1.4.1 2

2.1 Utworzenie wykresów korelacji sumy opadów i średniej temperatury w miesiącu.

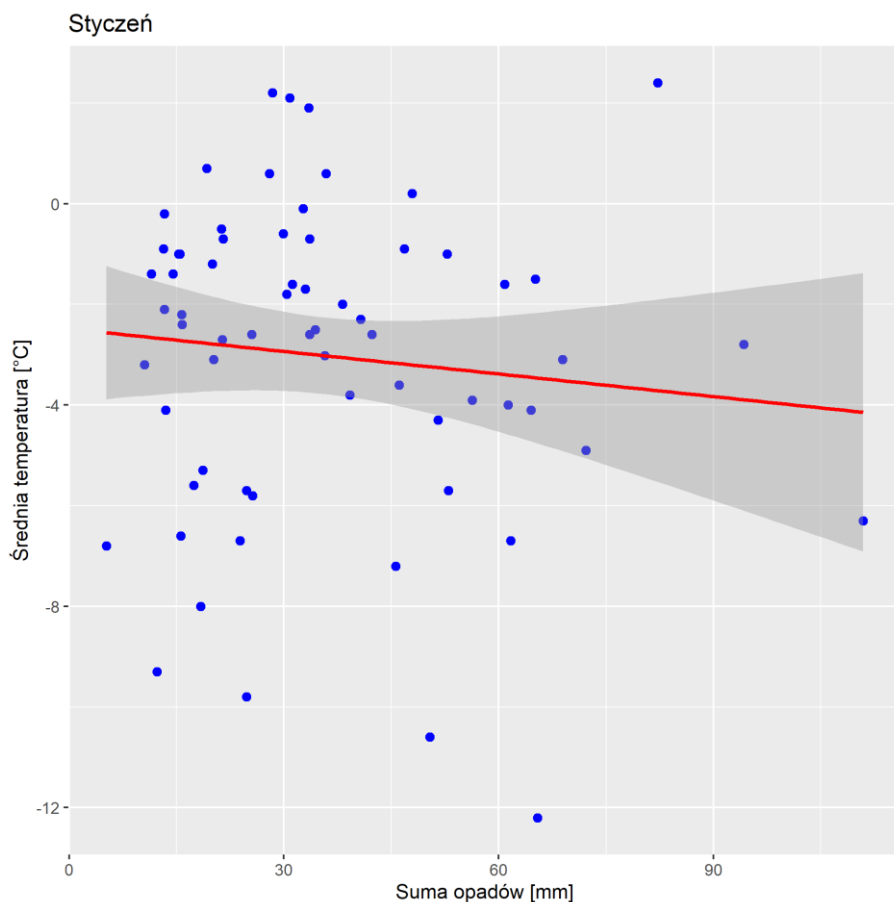
Postępujemy analogicznie jak w punkcie 1.4 wykorzystując funkcje: `ggplot()`, `geom_point()`, `geom_smooth()`. Całość powtarzamy w pętli.

```
for(i in 2:13)
{
  tmp1 <- dynow[,i] #dane z i-tej kolumny (kolejne miesiące) [temperatura]
  tmp2 <- dynow[,i+12] #dane z kolumn z opadami

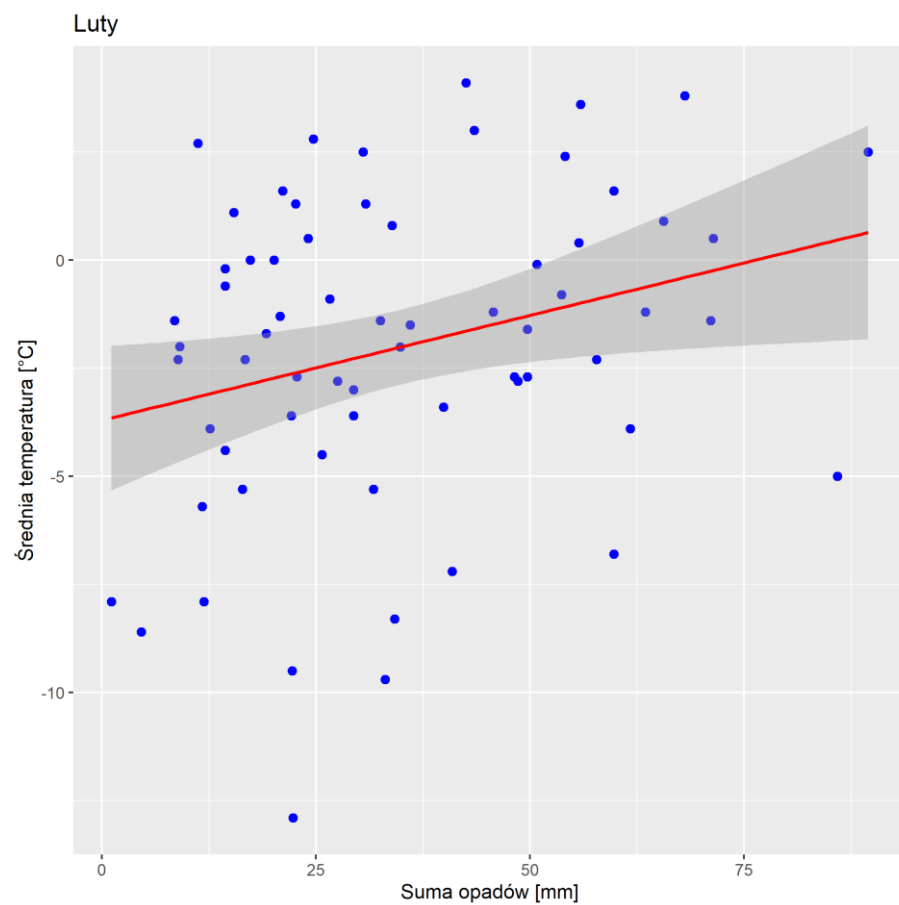
  tmp_df <- data.frame(tmp2, tmp1) #tymczasowy df

  tmp_plot <- ggplot(tmp_df, aes(tmp2, tmp1)) + #wykres
    geom_point(color="blue", size=2)+
    geom_smooth(method="lm", color="red")+
    xlab("Suma opadów [mm]") +
    ylab("Średnia temperatura [°C]") +
    ggtitle(miesiace[i-1])

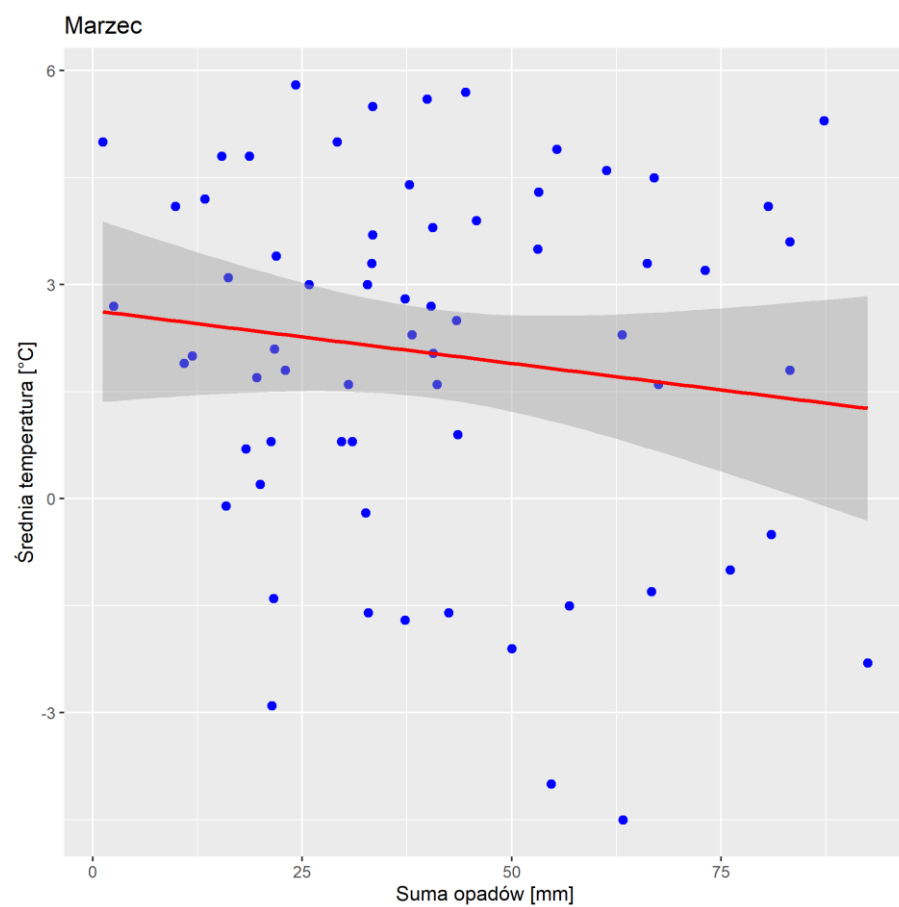
  ggsave(tmp_plot, filename = paste0("cor_temp-opady_", miesiace[i-1], ".png"))
  #zapisywanie
}
```



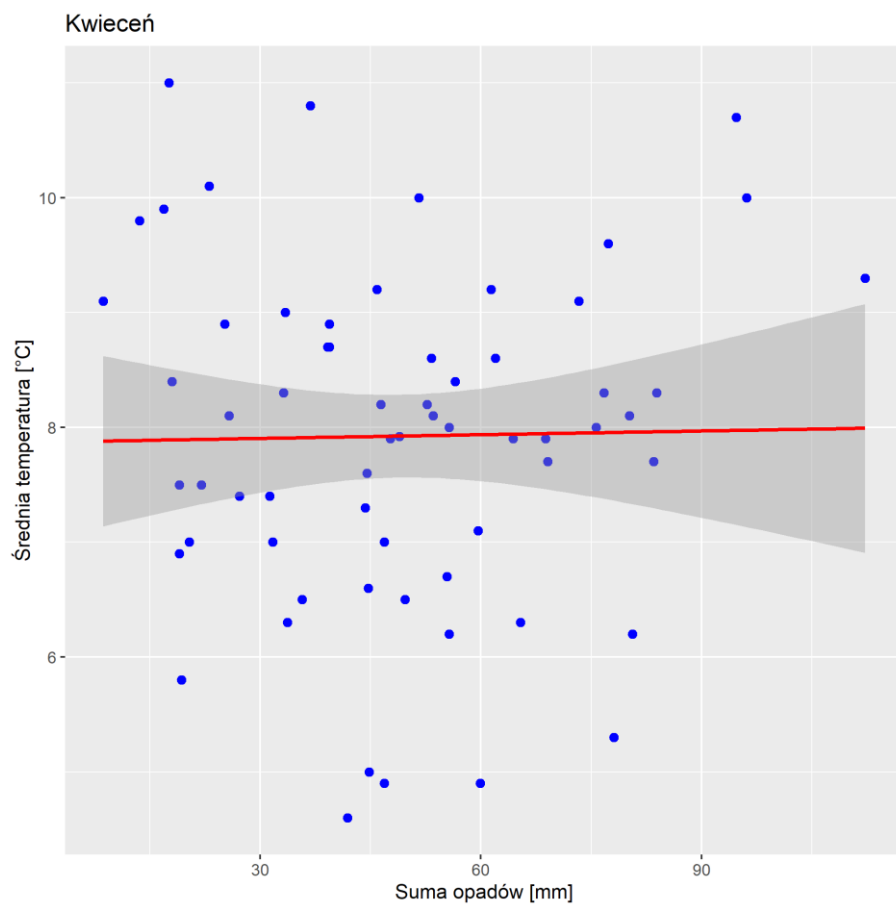
wykr. 2.1.1



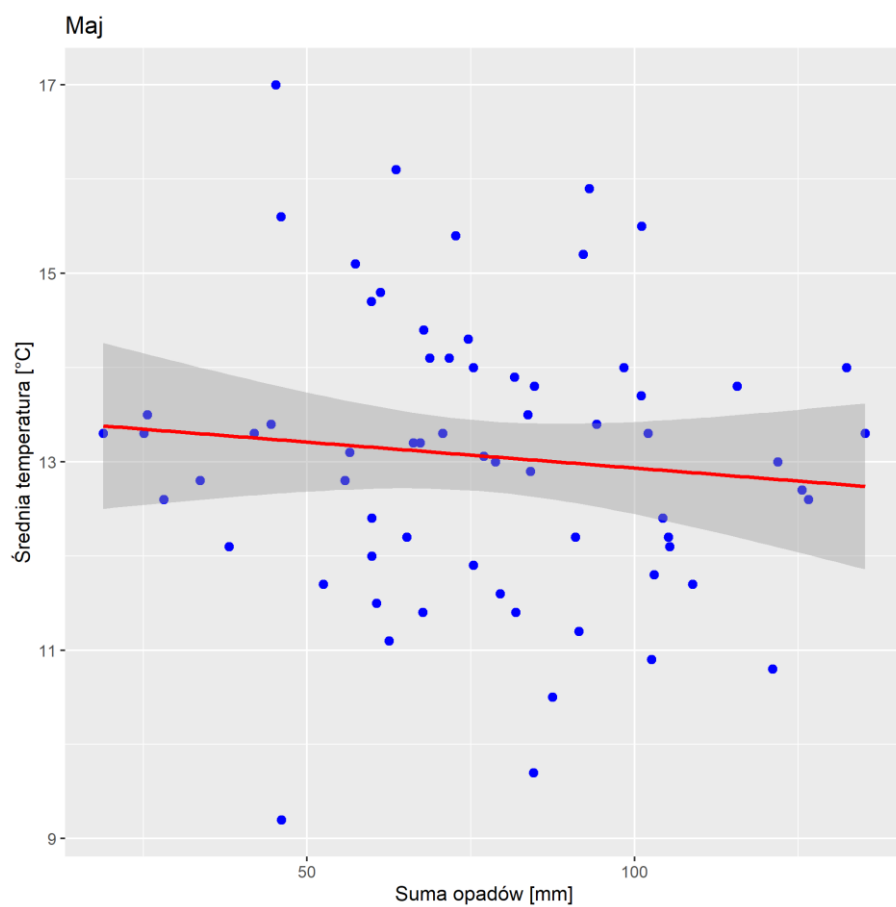
wykr. 2.1.2



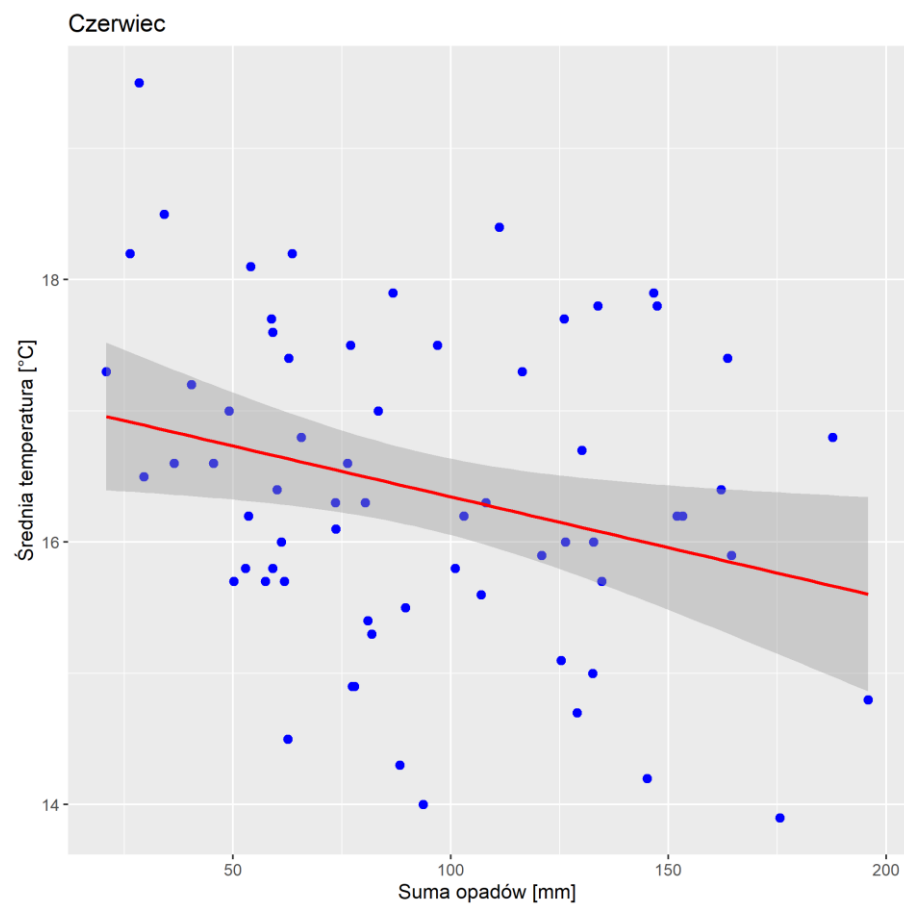
wykr. 2.1.3



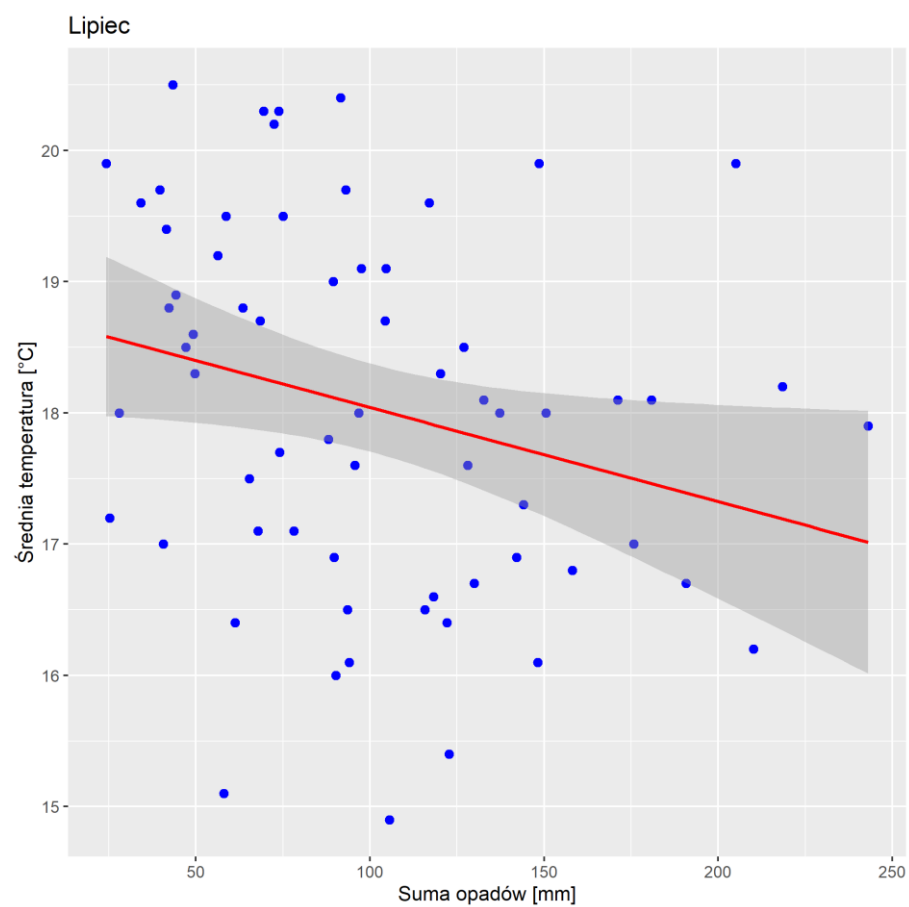
wykr. 2.1.4



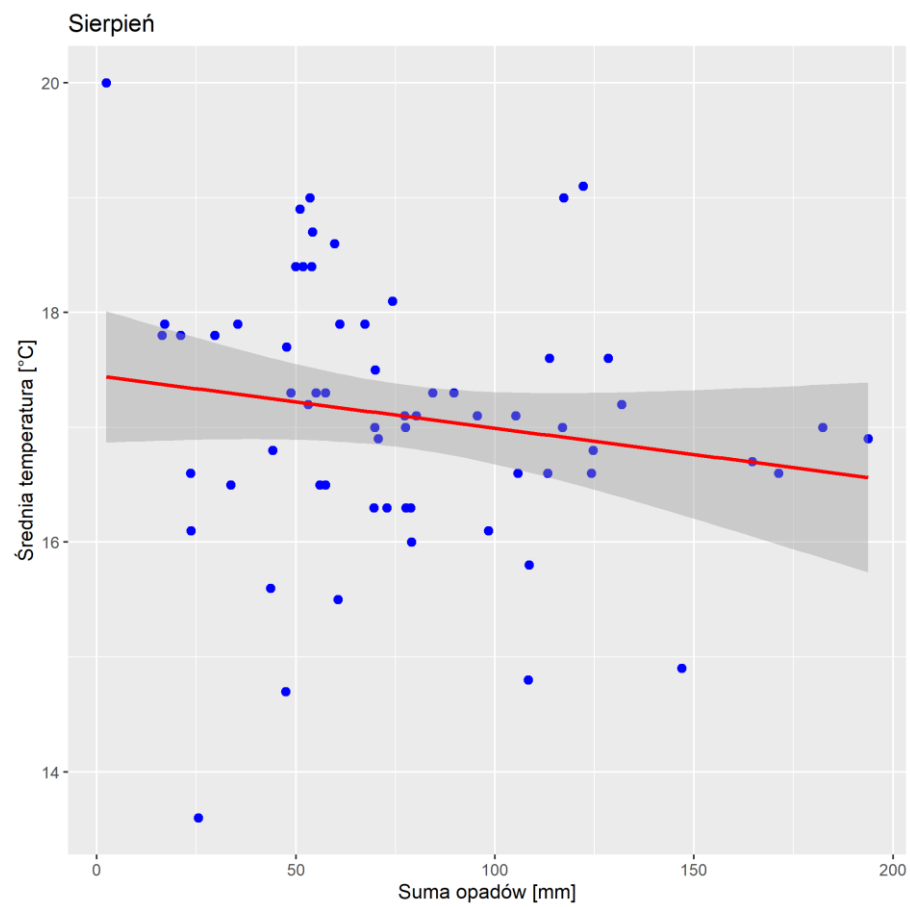
wykr. 2.1.5



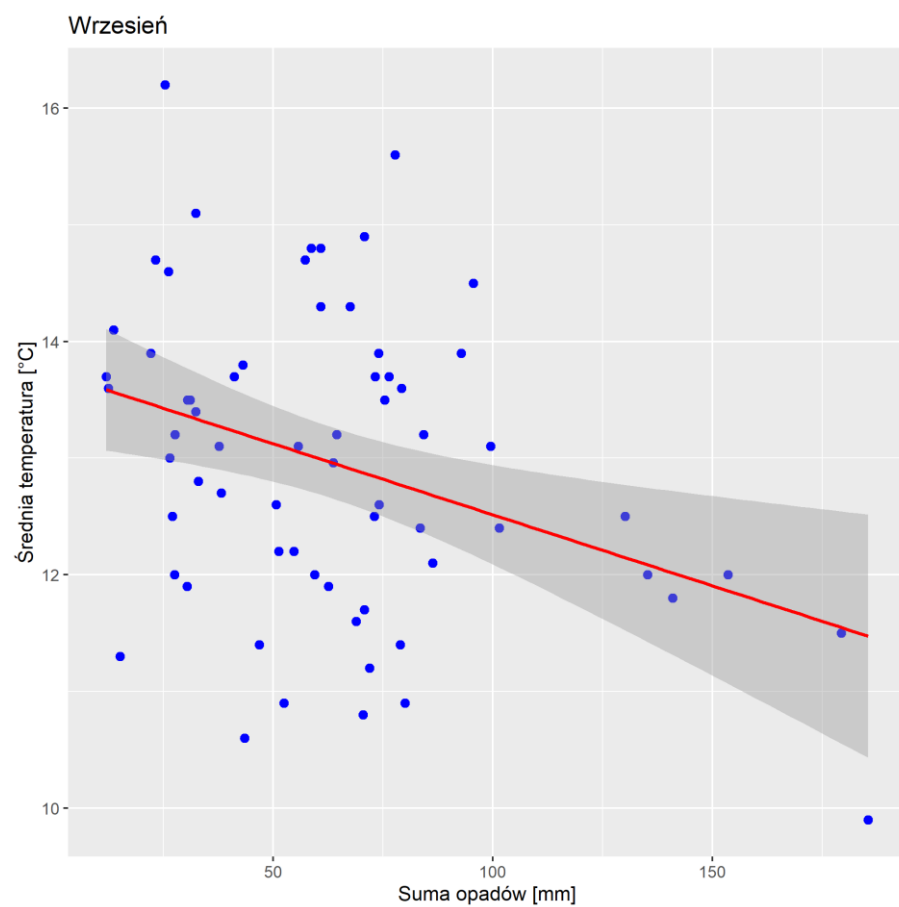
wykr. 2.1.6



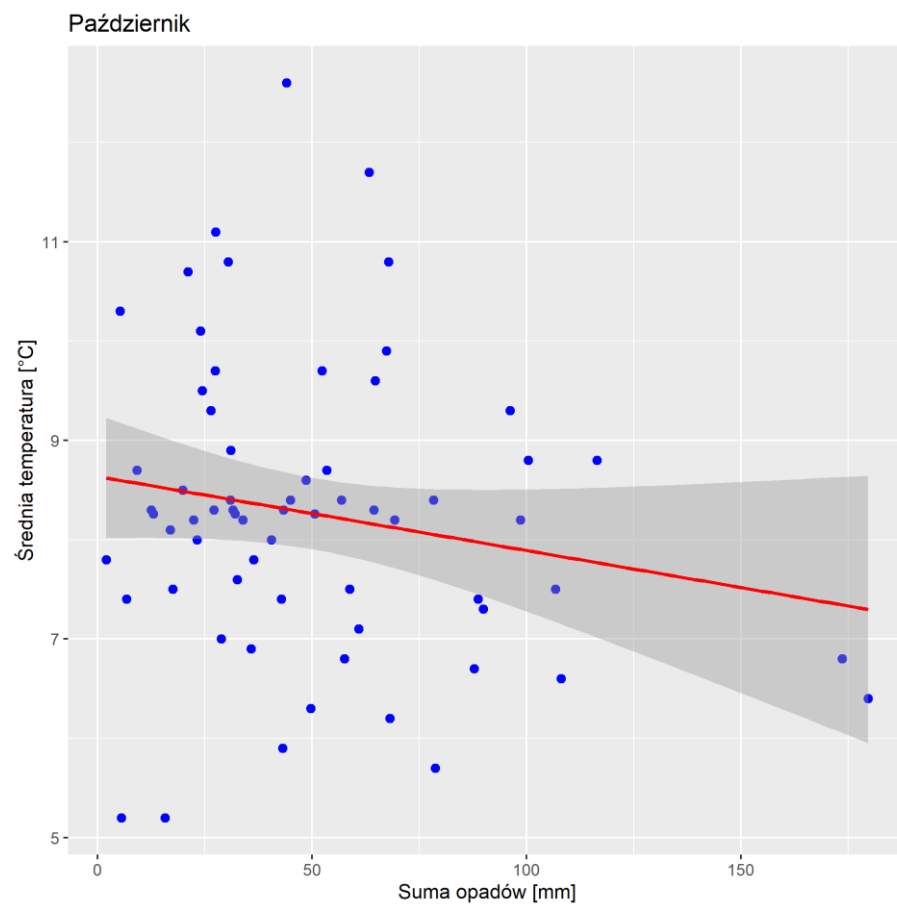
wykr. 2.1.77



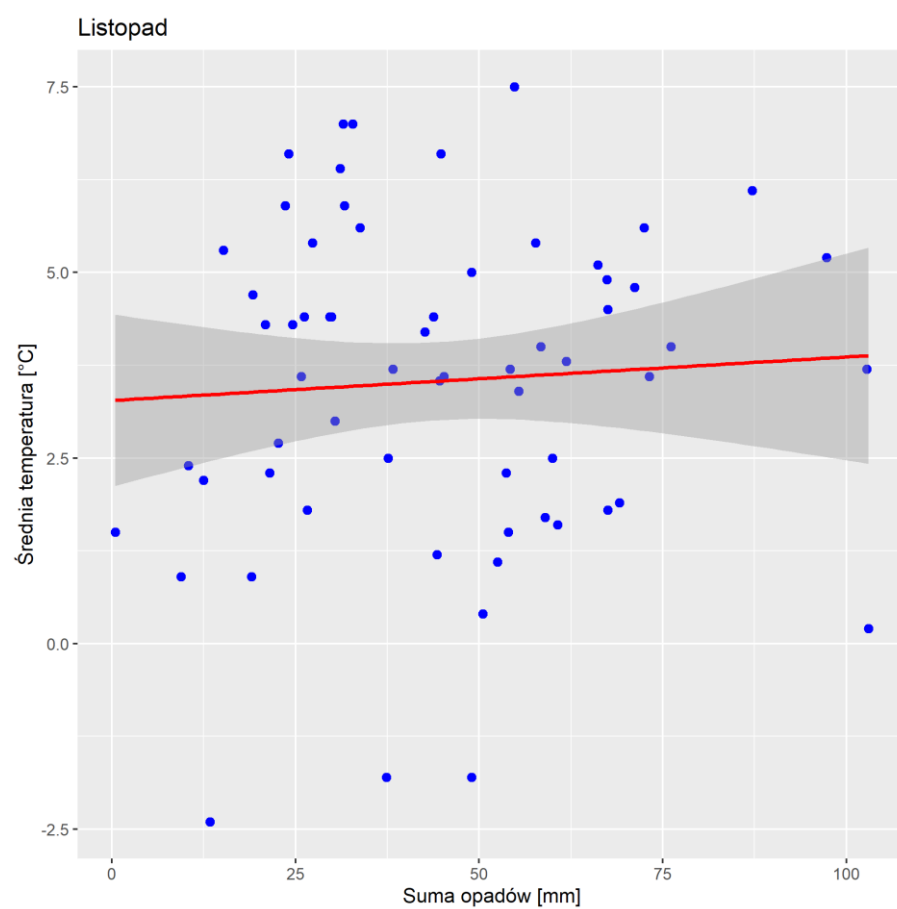
wykr. 2.1.8



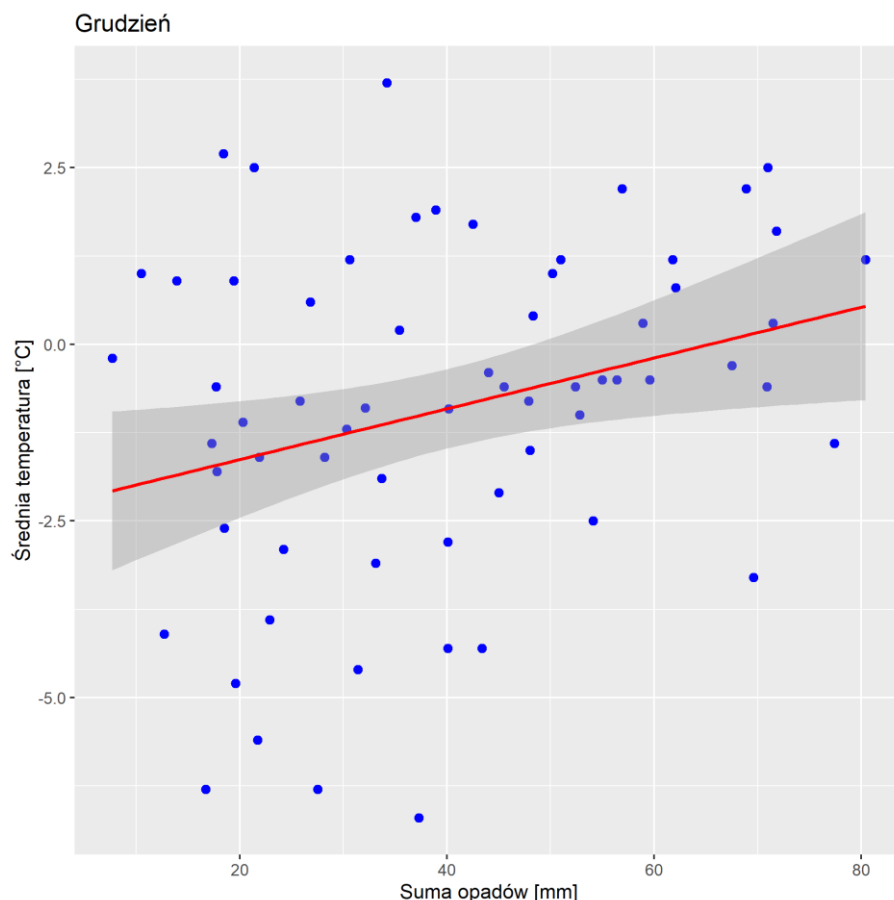
wykr. 2.1.9



wykr. 2.1.10



wykr. 2.1.11



wykr. 2.1.12

2.2 Utworzenie kolumn ze średnimi temperaturami oraz sumą opadów w porach roku i w okresie wegetacyjnym.

Kolumny dodajemy wykorzystując funkcję *mutate()*. Dokonałiśmy również zaokrąglenia danych do 2 miejsc po przecinku z pomocą funkcji *round()*.

```
#pory roku - temperatura
dynow <- dynow %>% as_tibble() %>%
  mutate(wiosna.temp.avg = (III.temp + IV.temp + V.temp)/3) #wiosna

dynow <- dynow %>% as_tibble() %>%
  mutate(lato.temp.avg = (VI.temp + VII.temp + VIII.temp)/3) #lato

dynow <- dynow %>% as_tibble() %>%
  mutate(jesien.temp.avg = (IX.temp + X.temp + XI.temp)/3) #jesień

dynow <- dynow %>% as_tibble() %>%
  mutate(zima.temp.avg = (XII.temp + I.temp + II.temp)/3) #zima

dynow <- dynow %>% as_tibble() %>%
  mutate(ow.temp.avg = (V.temp + VI.temp + VII.temp + VIII.temp + IX.temp)/5)

#pory roku - opady
dynow <- dynow %>% as_tibble() %>%
  mutate(wiosna.opady.sum = (III.opady + IV.opady + V.opady)) %>% #wiosna
```



```

round(digits = 2)

dynow <- dynow %>% as_tibble() %>%
  mutate(lato.opady.sum = (VI.opady + VII.opady + VIII.opady)) %>% #lato
  round(digits = 2)

dynow <- dynow %>% as_tibble() %>%
  mutate(jesien.opady.sum = (IX.opady + X.opady + XI.opady)) %>% #jesień
  round(digits = 2)

dynow <- dynow %>% as_tibble() %>%
  mutate(zima.opady.sum = (XII.opady + I.opady + II.opady)) %>% #zima
  round(digits = 2)

dynow <- dynow %>% as_tibble() %>%
  mutate(ow.opady.sum = (V.opady + VI.opady + VII.opady + VIII.opady + IX.opady)) %>% #ow
  round(digits = 2)

```

2.3 Utworzenie wykresów korelacji sumy opadów i średniej temperatury w porach roku i okresu wegetacyjnego.

Najpierw tworzymy ramkę danych *dynow_okresy*, w której oprócz kolumny rok, znajdują się średnie temperatury i sumy opadów dla pór roku i okresu wegetacyjnego.

```

dynow_okresy<-data.frame(dynow$rok, dynow$wiosna.temp.avg, dynow$lato.temp.avg,
  dynow$jesien.temp.avg, dynow$zima.temp.avg, dynow$ow.temp.avg, dynow$wiosna.opady.sum,
  dynow$lato.opady.sum, dynow$jesien.opady.sum, dynow$zima.opady.sum, dynow$ow.opady.sum)
View(dynow_okresy)

```

Następnie podobnie jak w punkcie 2.1 tworzymy wykresy korelacji za pomocą pętli. Wektor *okresy* pozwolił na odpowiednie nadanie tytułów i nazw plików.

```

okresy <- c("Wiosna", "Lato", "Jesień", "Zima", "Okres Wegetacyjny")

for(i in 2:6)
{
  tmp1 <- dynow_okresy[,i] #dane z i-tej kolumny (kolejne miesiące) [temp.]

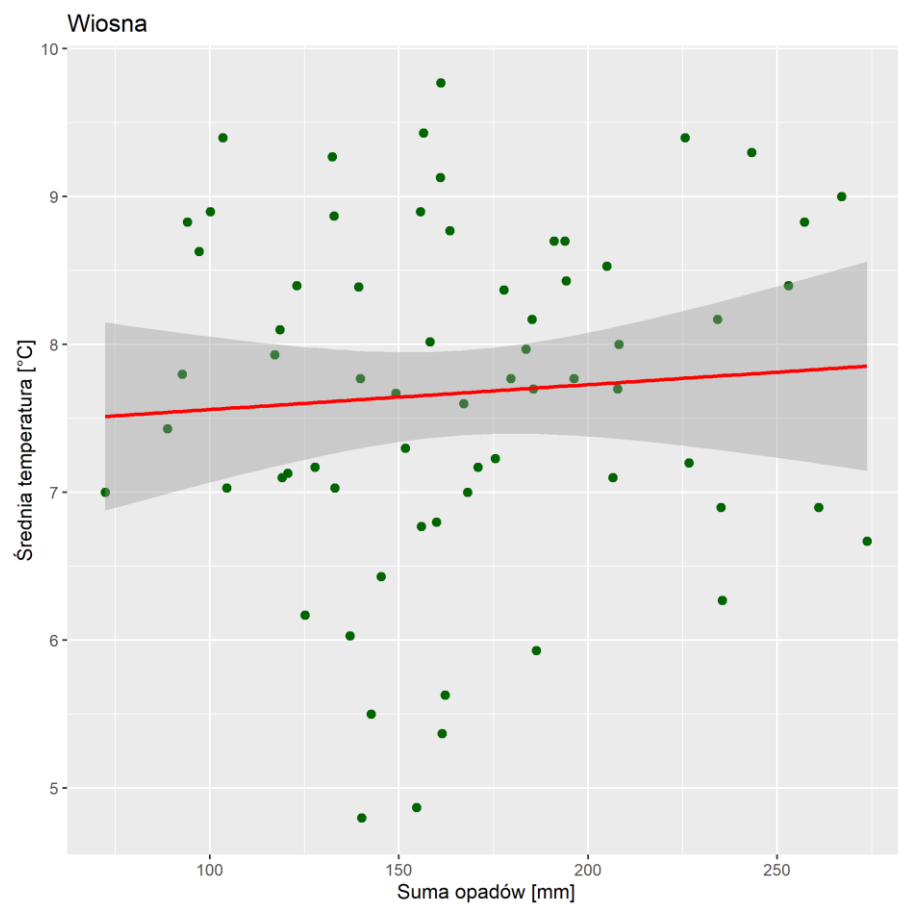
  tmp2 <- dynow_okresy[,i+5] #dane z kolumn z opadami

  tmp_df <- data.frame(tmp2, tmp1) #tymczasowy df

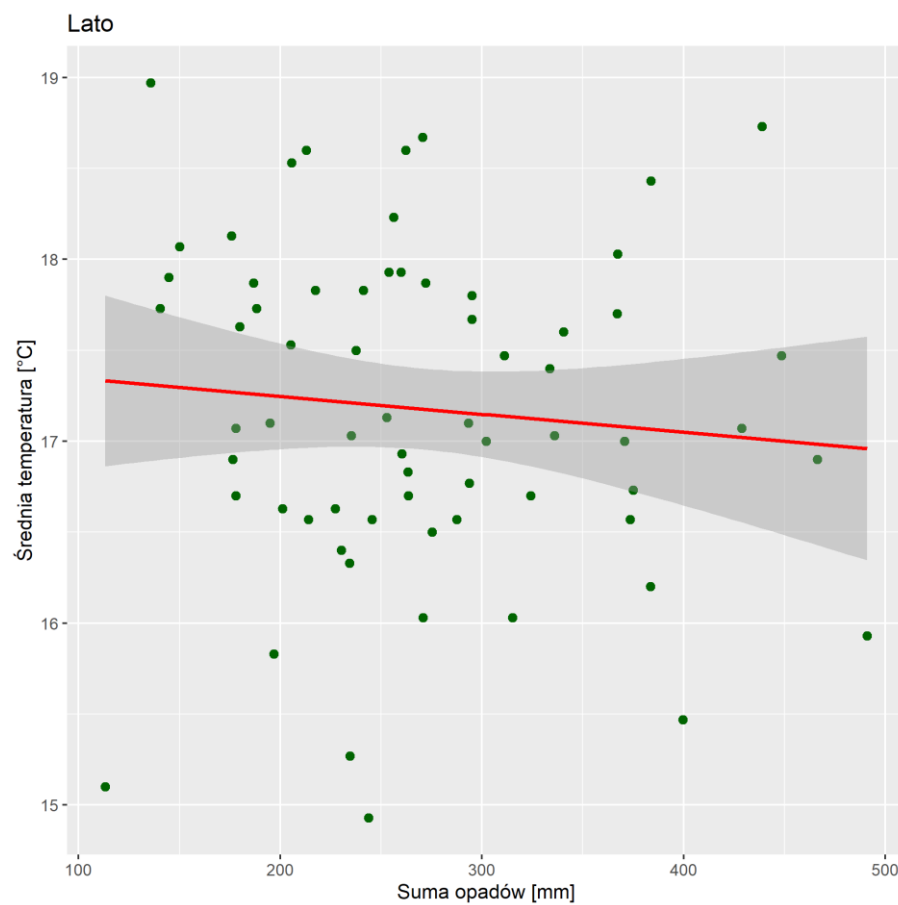
  tmp_plot <- ggplot(tmp_df, aes(tmp2, tmp1)) + #wykres
    geom_point(color="darkgreen", size=2)+
    geom_smooth(method="lm", color="red")+
    xlab("Suma opadów [mm])+
    ylab("Średnia temperatura [°C])+
    ggtitle(okresy[i-1]) +

  ggsave(tmp_plot, filename = paste0("cor_temp-opady_",okresy[i-1],".png"))
}

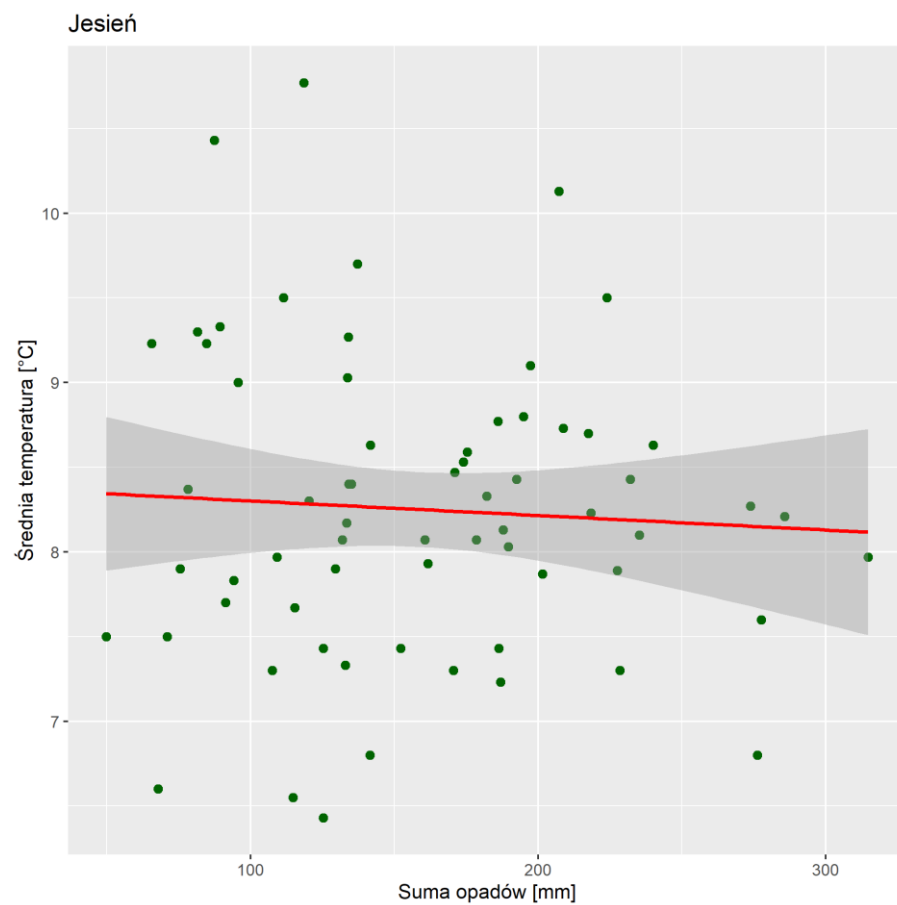
```



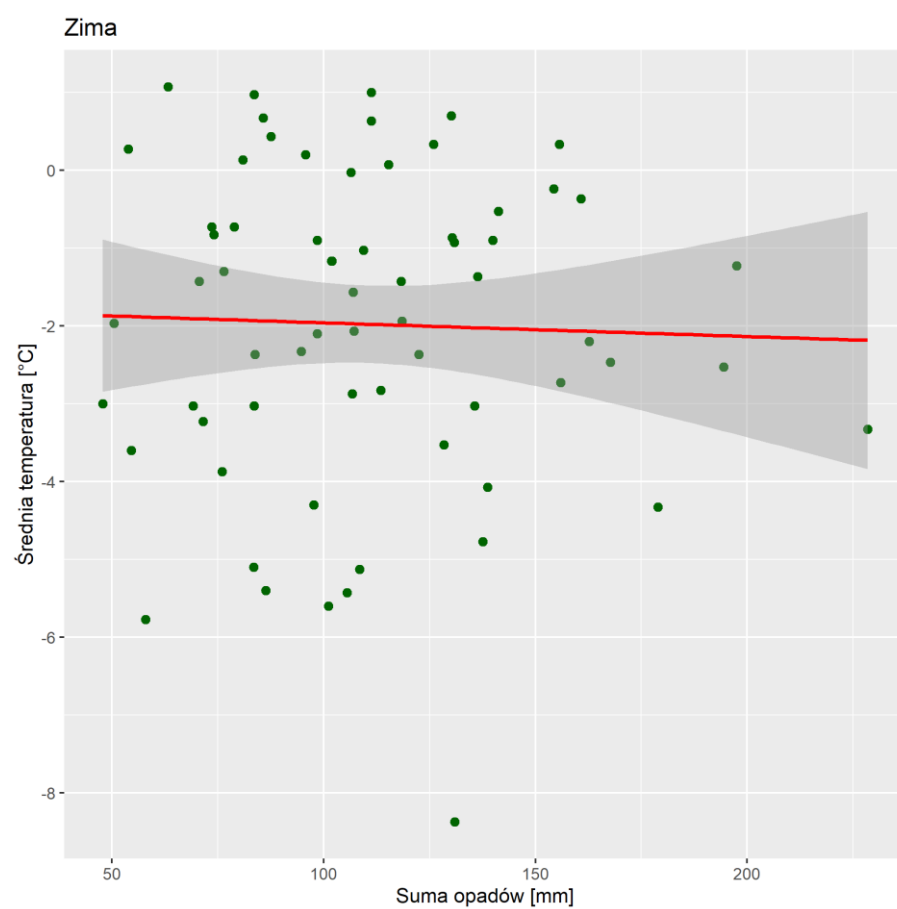
wykr. 2.3.1



wykr. 2.3.2



wykr. 2.3.3



wykr. 2.3.4



wykr. 2.3.5

2.4 Obliczanie współczynnika korelacji Pearsona i współczynnika rang Spearmana

Badamy korelację między średnią wartością temperatury powietrza a sumą opadów atmosferycznych. Współczynniki obliczymy za pomocą funkcji `cor()` a parametr `method` pozwoli na dobranie odpowiedniego typu korelacji. Wyniki wyświetlamy przy użyciu funkcji `print()`. Całość (podobnie jak w punkcie 1.4) odbywa się w pętli.

Dla miesięcy:

-wspł. Pearsona:

```
for(i in 2:13)
{
  tmp1 <- dynow[,i]

  tmp2 <- dynow[,i+12]

  wsp_p_mies <- cor(tmp1, tmp2, method="pearson")

  print(c(miesiace[i-1], wsp_p_mies))
}
```

```
## [1] "Styczeń"      "-0.105089117320161"
## [1] "Luty"         "0.2749129390447"
## [1] "Marzec"       "-0.131764856644808"
## [1] "Kwiecień"     "0.0166298470827561"
```

```
## [1] "Maj" "-0.0999160657534851"
## [1] "Czerwiec" "-0.278715641326083"
## [1] "Lipiec" "-0.261239427757128"
## [1] "Sierpień" "-0.168145919047691"
## [1] "Wrzesień" "-0.348145531588173"
## [1] "Październik" "-0.182029375130468"
## [1] "Listopad" "0.0627671075336259"
## [1] "Grudzień" "0.287109094971792"
```

-współ. Spearmana:

```
for(i in 2:13)
{
  tmp1 <- dynow[,i]

  tmp2 <- dynow[,i+12]

  wsp_p_mies <- cor(tmp1, tmp2, method="spearman")

  print(c(miesiace[i-1], wsp_p_mies))
}

## [1] "Styczeń" "-0.0891115992396651"
## [1] "Luty" "0.256128631679779"
## [1] "Marzec" "-0.0655287615308549"
## [1] "Kwiecień" "-0.0239542227727177"
## [1] "Maj" "-0.11198606458916"
## [1] "Czerwiec" "-0.254487765742238"
## [1] "Lipiec" "-0.319248259375156"
## [1] "Sierpień" "-0.215108543421992"
## [1] "Wrzesień" "-0.262456040052157"
## [1] "Październik" "-0.140210181420807"
## [1] "Listopad" "0.0595459434090823"
## [1] "Grudzień" "0.285685961849537"
```

Dla pór roku i okresu wegetacyjnego:

-współ. Pearsona:

```
j<-1
for(i in 2:6)
{
  tmp1 <- dynow_okresy[,i]

  tmp2 <- dynow_okresy[,i+5]

  wsp_p_okresy <- cor(tmp1, tmp2, method="pearson")

  print(c(okresy[j], wsp_p_okresy))

  j<-j+1
}

## [1] "Wiosna" "0.0697685576456831"
## [1] "Lato" "-0.0936774769374615"
## [1] "Jesień" "-0.0587811764763315"
```

```
## [1] "Zima" "-0.0328365778147387"
## [1] "Okres Wegetacyjny" "-0.139609510647012"
```

-wspł. rang Spearmana:

```
j<-1
for(i in 2:6)
{
  tmp1 <- dynow_okresy[,i]

  tmp2 <- dynow_okresy[,i+5]

  wsp_p_okresy <- cor(tmp1, tmp2, method="spearman")

  print(c(okresy[j], wsp_p_okresy))

  j<-j+1
}

## [1] "Wiosna" "0.0474768781284698"
## [1] "Lato" "-0.133782970413017"
## [1] "Jesień" "0.0110034662272375"
## [1] "Zima" "-0.0117018278546586"
## [1] "Okres Wegetacyjny" "-0.203231835322085"
```

2.5 Badanie istotności współczynników korelacji (w przypadku badania korelacji Pearsona).

Istotność korelacji pomiędzy średnią temperaturą a sumą opadów w kolejnych okresach roku sprawdzmy za pomocą funkcji *cor.test*. Przyjmujemy, że korelacja jest istotna gdy $p < 0.05$. Wynik działania funkcji prezentujemy poniżej. Kolejne wartości "p-value" prezentujemy w tabeli 2.1

```
cor.test(dynow$I.temp, dynow$I.opady)
```

```
##
## Pearson's product-moment correlation
##
## data: dynow$I.temp and dynow$I.opady
## t = -0.84539, df = 64, p-value = 0.401
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.3385118 0.1405177
## sample estimates:
## cor
## -0.1050891
```

Okres	p-value
Styczeń	0.401
Luty	0.02549
Marzec	0.2916
Kwiecień	0.8946
Maj	0.4247
Czerwiec	0.02344
Lipiec	0.03411
Sierpień	0.1772
Wrzesień	0.004177
Październik	0.1435
Listopad	0.6166
Grudzień	0.01942
Wiosna	0.5778
Lato	0.4544
Jesień	0.6392
Zima	0.7935
Okres wegetacyjny	0.2636

Tabela 2.1

3. Obserwacje

W ostatnim czasie kwestia ilości opadów na terenie Polski jest szeroko komentowana. Niedobór opadów stał się palącym problemem, w szczególności gdy stoimy na progu ekstremalnie suchego lata. O skali tego problemu mogliśmy przekonać się kiedy to zaledwie kilkanaście dni temu obserwowaliśmy pożar trawiący Biebrzański Park Narodowy.

Wykres rozrzutu pozwala obserwować przebieg zmian wartości sumy opadów na przestrzeni lat. Umieszczona na wykresach linia trendu umożliwia wnioskowanie o prawdopodobnej tendencji w przyszłości.

Tendencję wzrostową sumy opadów obserwujemy dla marca, kwietnia, maja, lipca, września i października. Minimalny wzrost wykazuje również styczeń. (wykresy 1.4.1 – 1.4.12)

Z wykresów korelacji wynika, że pośród miesięcy jedynie luty, listopad i grudzień wykazują tendencję dodatnią, jednocześnie żaden miesiąc nie wykazuje braku korelacji. (wykresy 2.1.1 – 2.1.12)

Wśród pór roku i okresu wegetacyjnego, jedynie wiosna wykazuje dodatnią tendencję. (wykresy 2.3.1 – 2.3.5).

Tabela 2.1 wykazuje, że współczynnik korelacji Pearsona w przypadku lutego, czerwca, lipca, września i grudnia jest istotny, czyli prawdopodobieństwo, że korelacja danych miesięcy jest prawdziwa jest wysokie. Wśród pór roku, żaden okres nie posiada istotnego współczynnika korelacji.

W kwestii różnic wartości współczynnika Pearsona i współczynnika rang Spearmana wyróżnia się marzec, sierpień, wrzesień oraz jesień. Interpretując wartość współczynnika korelacji rang Spearmana okazuje się, że w większości przypadków korelacja jest słaba (praktycznie brak związku). Największą istotność tego współczynnika wykazuje luty, lipiec oraz grudzień (punkt 2.4). Dla tych miesięcy wartość przekracza 0,2 co świadczy o występowaniu korelacji niskiej (zależność wyraźna).

4. Wnioski

Biorąc pod uwagę wyniki analizy wykonanej w ramach Projektu 1 interesującą tendencję wykazuje sierpień. Generalny wzrost temperatury w miesiącach wiosenno-letnich idzie w parze ze wzrostem opadów z wyjątkiem czerwca (wykr. 1.4.6) i sierpnia (wykr. 1.4.8), przy czym to w sierpniu są wyższe temperatury. Generuje to sprzyjające warunki dla powstawania suszy.

Wykres korelacji sumy opadów i średniej temperatury dla lutego (wykr.2.1.2) wykazuje tendencję wzrostową, a także znaczną istotność. Oznacza to że dla wyższych temperatur zanotowano większą sumę opadów.

Największą istotnością współczynnika korelacji Pearsona wyróżnia się grudzień (~ 0.0194).

Okres wegetacyjny nie wykazuje istotności współczynnika (punkt 2.4), jednocześnie tendencja korelacji jest ujemna (wykr. 2.3.5). Z tego powodu można założyć nieznaczną zależność występowania wyższej temperatury i mniejszej sumy opadów.

Różnice w wartości współczynnika korelacji Pearsona i rang Spearmana dla marca, sierpnia, września oraz jesieni prawdopodobnie wynika ze znacznej „odległości” między wartościami. Korelacja Pearsona opiera się na wartościach, a korelacja Spearmana nie korzysta z konkretnych wartości, ale z zależności między nimi.

5. Bibliografia

- prezentacje z przedmiotu PDŚ autorstwa Pani dr inż. Małgorzaty Danek
- <https://rstudio.com/resources/cheatsheets/>
- <https://bookdown.org/ndphillips/YaRrr/loops.html>