

Temperatura i Ciśnienie na Marsie

Filip Kosiorowski

2024-05-31

- 1.Dane
- 2. Parametry opisowe temperatury na Marsie.
- 3. Wykres temperatury
- 4. Histogram Temperatury
- 5. Wykres pudełkowy temperatury
- 6. Dystrybuanta temperatury
- 7. Parametry opisowe ciśnienia na Marsie
- 8. Wykres ciśnienia
- 9. Histogram ciśnienia
- 10. Wykres pudełkowy ciśnienia
- 11. Dystrybuanta ciśnienia
- 12. Hipoteza testu Wilcoxona
- 13. Hipoteza testu t-Studenta
- 14. Oto spis wszystkich funkcji użytych w analizie danych na temat temperatury i ciśnienia na Marsie, wraz z krótkim opisem każdej z nich:

1.Dane

```
#install.packages('readr')
#install.packages('ggplot2')
#install.packages('moments')
library(readr) # Załadowanie pakietu readr do wczytywania danych
library(ggplot2) # Załadowanie pakietu ggplot2 do tworzenia wykresów
library(moments) # Załadowanie pakietu moments do obliczania momentów statystycznych
library(rmarkdown)
#import danych

dane <- read.csv("C:\\Users\\kosio\\Downloads\\the-pudding-mars-weather\\the-pudding-mars-weather\\data\\mars_weather.csv", header = TRUE, sep = ",", na.strings = c("", "NA"))
```

Dane pochodzą ze strony <https://data.world/the-pudding/mars-weather> (<https://data.world/the-pudding/mars-weather>), oraz przedstawiają pomary temperatury oraz ciśnienia na powierzchni Marsa wykonane przez łazik Curiosity. Dane zostały podane w Lutym 2018 roku przez NASA i zmierzone przez Rover Environmental Monitoring Station (REMS). W danych mamy przedstawione odpowiednio datę pomiaru, SOL, tzn. dzień spędzony na Marsie, oraz odpowiednie pomiary

```
# Usuwanie brakujących wartości z kolumn 'terrestrial_date', 'max_temp' i 'min_temp'
dane_clean <- dane[complete.cases(dane$terrestrial_date, dane$max_temp, dane$min_temp), ]
# Tworzenie nowej ramki danych z kolumnami 'Data', 'temperatura_max' i 'temperatura_min'
Temperatura <- data.frame(Data = dane_clean$terrestrial_date,
                          temperatura_max = dane_clean$max_temp,
                          temperatura_min = dane_clean$min_temp)
# Obliczanie średniej temperatury dla każdej daty
Temperatura$temperatura_srednia <- rowMeans(Temperatura[c("temperatura_max", "temperatura_min")])
# Tworzenie nowej ramki danych z kolumnami 'Data' i 'cisnienie'
Pressure <- data.frame(Data = dane_clean$terrestrial_date, cisnienie = dane_clean$pressure)
```

2. Parametry opisowe temperatury na Marsie.

Te statystyki opisowe pomagają w zrozumieniu rozkładu temperatur w danych. Dominanta wskazuje najczęściej występującą wartość, mediana wskazuje wartość środkową, wariancja i odchylenie standardowe mierzą rozproszenie danych, a zakres i IQR pokazują rozpiętość wartości w zbiorze danych. Analiza tych parametrów może dostarczyć cennych informacji na temat charakterystyki temperatur w badanym zbiorze danych.

```

# Obliczanie dominanty dla minimalnej, maksymalnej i średniej temperatury
czestosc_min <- table(Temperatura$temperatura_min)
dominanta_min <- names(czestosc_min)[which.max(czestosc_min)]

czestosc_max <- table(Temperatura$temperatura_max)
dominanta_max <- names(czestosc_max)[which.max(czestosc_max)]

czestosci <- table(Temperatura$temperatura_srednia)
dominanta <- names(czestosci)[which.max(czestosci)]

# Obliczanie mediany dla maksymalnej i minimalnej temperatury
temperatura_max <- Temperatura$temperatura_max
temperatura_min <- Temperatura$temperatura_min
Mediana_max <- median(temperatura_max)
Mediana_min <- median(temperatura_min)

# Obliczanie wariancji dla maksymalnej i minimalnej temperatury
wariancja_temperatura_max <- var(Temperatura$temperatura_max)
wariancja_temperatura_min <- var(Temperatura$temperatura_min)

# Obliczanie odchylenia standardowego dla maksymalnej i minimalnej temperatury
odchylenie_standardowe_temperatura_max <- sd(Temperatura$temperatura_max)
odchylenie_standardowe_temperatura_min <- sd(Temperatura$temperatura_min)

# Obliczanie zakresu i rozstępu dla maksymalnej i minimalnej temperatury
zakres_temperatura_max <- range(Temperatura$temperatura_max)
rozstep_temperatura_max <- diff(zakres_temperatura_max)
zakres_temperatura_min <- range(Temperatura$temperatura_min)
rozstep_temperatura_min <- diff(zakres_temperatura_min)

# Obliczanie rozstępu międzykwartylowego dla maksymalnej i minimalnej temperatury
iqr_temperatura_max <- IQR(Temperatura$temperatura_max)
iqr_temperatura_min <- IQR(Temperatura$temperatura_min)
#Współczynnik zmienności temperatury dla maksymalnej i minimalnej temperatury
temperatura_sr <- mean(Temperatura$temperatura_srednia)
wspolczynnik_zmienności_tmax <- odchylenie_standardowe_temperatura_max / temperatura_sr
wspolczynnik_zmienności_tmin <- odchylenie_standardowe_temperatura_min / temperatura_sr
#Podanie wszystkich wymienionych wyżej wartości
cat("Dominanta: ", dominanta, "\n",
    "Dominanta_min: ", dominanta_min, "\n",
    "Dominanta_max: ", dominanta_max, "\n",
    "Mediana_max: ", Mediana_max, "\n",
    "Mediana_min: ", Mediana_min, "\n",
    "Wariancja temperatury maksymalnej: ", wariancja_temperatura_max, "\n",
    "Wariancja temperatury minimalnej: ", wariancja_temperatura_min, "\n",
    "Odchylenie standardowe temperatury maksymalnej: ", odchylenie_standardowe_temperatura_max, "\n",
    "Odchylenie standardowe temperatury minimalnej: ", odchylenie_standardowe_temperatura_min, "\n",
    "Zakres wartości temperatury maksymalnej wynosi od ", zakres_temperatura_max[1], " do ", zakres_temperatura_max[2], "\n",
    "Rozstęp wartości temperatury maksymalnej wynosi: ", rozstep_temperatura_max, "\n",
    "Zakres wartości temperatury minimalnej wynosi od ", zakres_temperatura_min[1], " do ", zakres_temperatura_min[2], "\n",

```

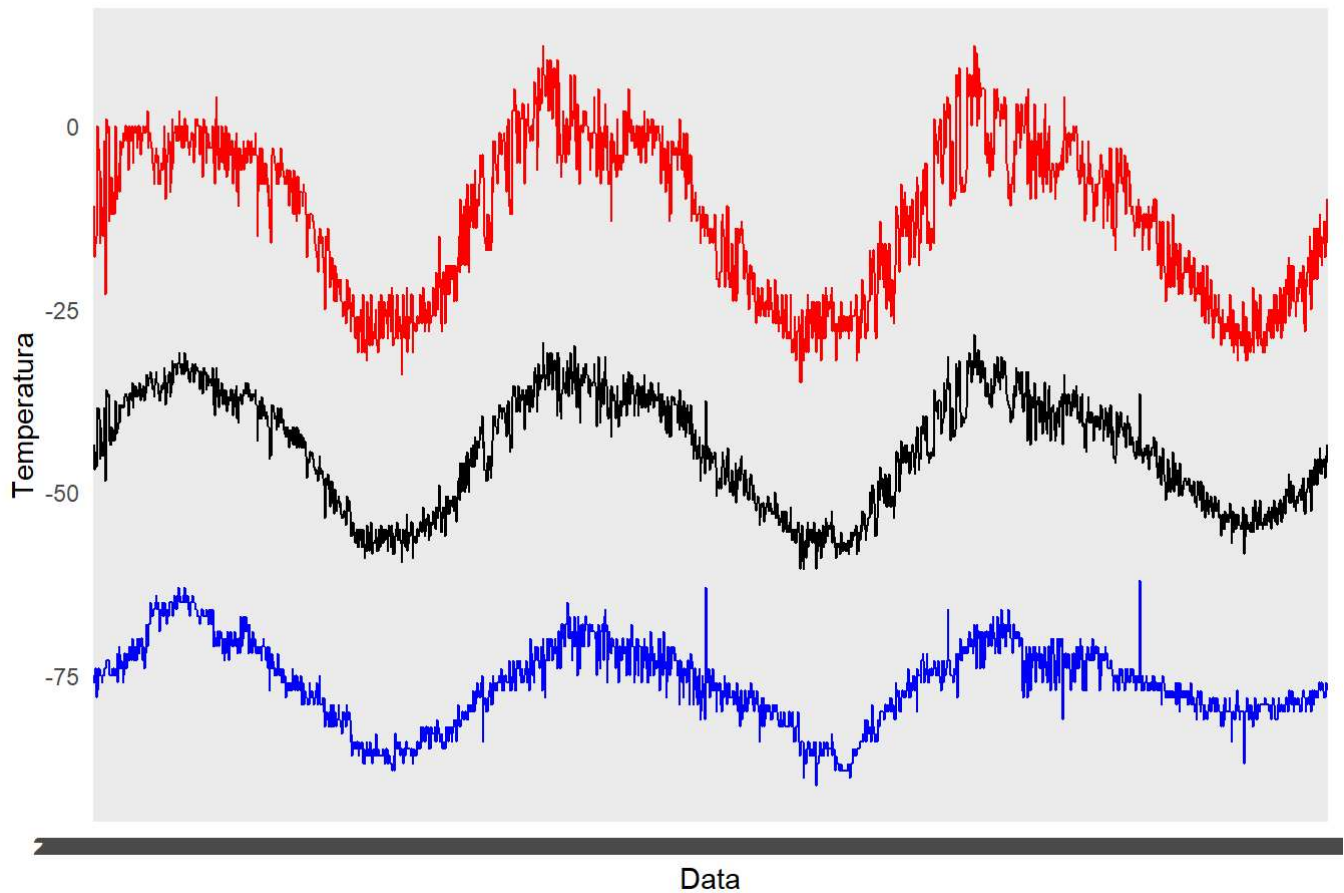
```
"Rozstęp wartości temperatury minimalnej wynosi: ", rozstep_temperatura_min, "\n",  
"Rozstęp międzykwartylowy temperatury maksymalnej wynosi: ", iqr_temperatura_max, "\n",  
"Rozstęp międzykwartylowy temperatury minimalnej wynosi: ", iqr_temperatura_min, "\n",  
"Współczynnik zmienności temperatury max wynosi: ", wspolczynnik_zmiennosci_tmax, "\n",  
"Współczynnik zmienności temperatury max wynosi: ", wspolczynnik_zmiennosci_tmax, "\n")
```

```
## Dominanta: -36.5  
## Dominanta_min: -77  
## Dominanta_max: -1  
## Mediana_max: -11  
## Mediana_min: -76  
## Wariancja temperatury maksymalnej: 114.4783  
## Wariancja temperatury minimalnej: 30.29509  
## Odchylenie standardowe temperatury maksymalnej: 10.69945  
## Odchylenie standardowe temperatury minimalnej: 5.504098  
## Zakres wartości temperatury maksymalnej wynosi od -35 do 11  
## Rozstęp wartości temperatury maksymalnej wynosi: 46  
## Zakres wartości temperatury minimalnej wynosi od -90 do -62  
## Rozstęp wartości temperatury minimalnej wynosi: 28  
## Rozstęp międzykwartylowy temperatury maksymalnej wynosi: 20  
## Rozstęp międzykwartylowy temperatury minimalnej wynosi: 8  
## Współczynnik zmienności temperatury max wynosi: -0.2414368  
## Współczynnik zmienności temperatury max wynosi: -0.2414368
```

3. Wykres temperatury

```
wykres_temp <- ggplot(Temperatura, aes(x = Data)) +  
  geom_line(aes(y = temperatura_max, group = 1), color = "red", linetype = "solid") +  
  geom_line(aes(y = temperatura_min, group = 1), color = "blue", linetype = "solid") +  
  geom_line(aes(y = temperatura_srednia, group = 1), color = "black", linetype = "solid") +  
  labs(x = "Data", y = "Temperatura", title = "Wykres temperatury") +  
  theme_minimal()  
wykres_temp
```

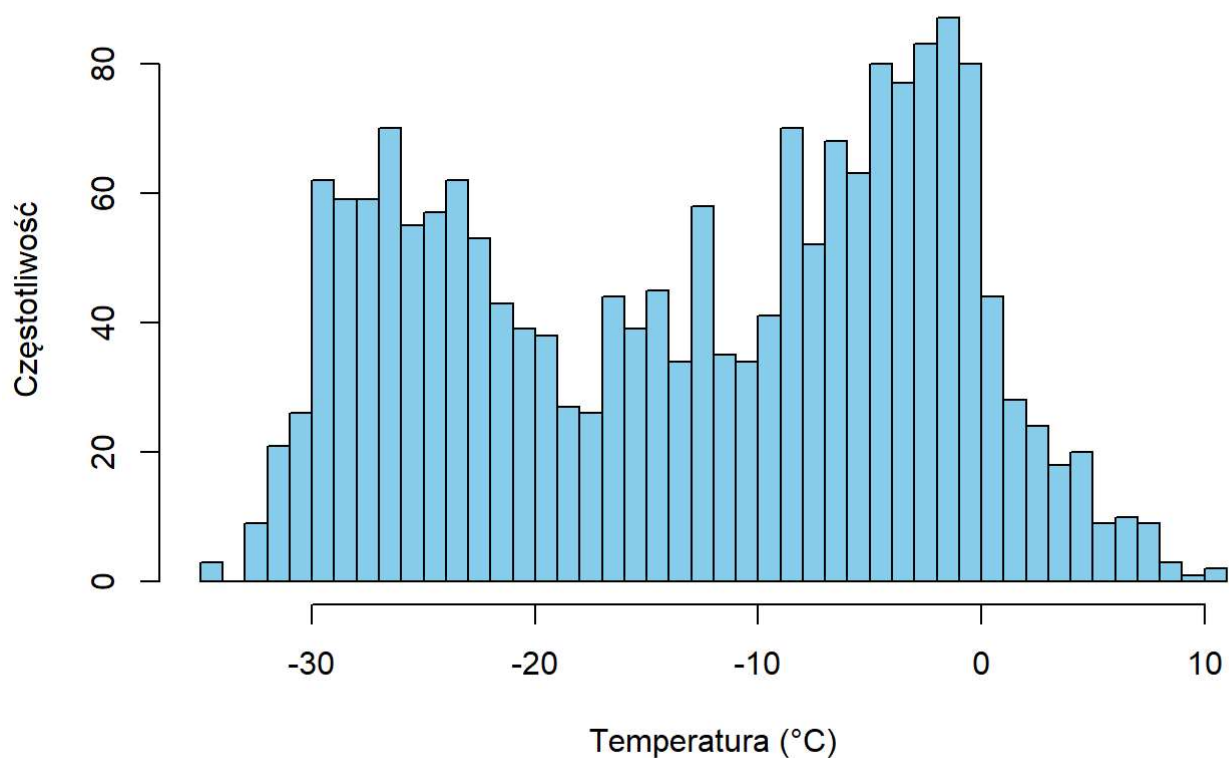
Wykres temperatury



4. Histogram Temperatury

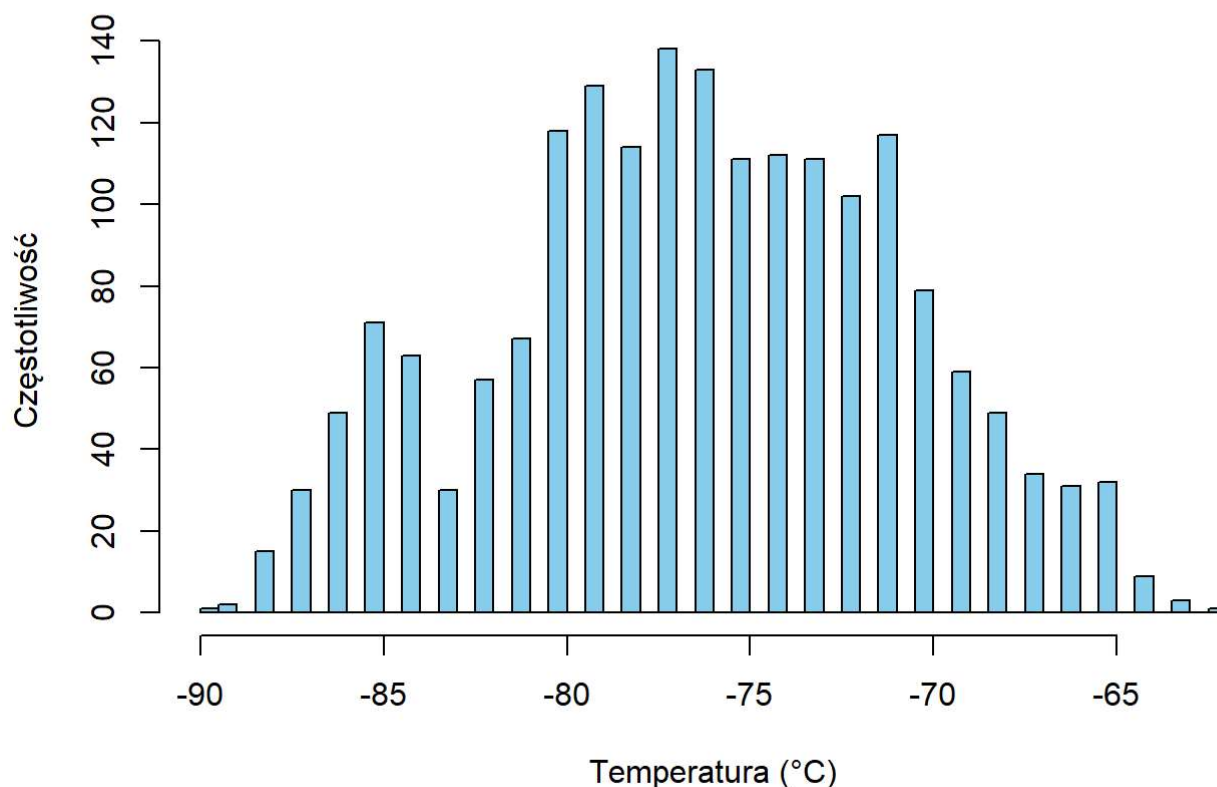
```
Histogram_max <- hist(temperatura_max,  
  main = "Histogram rozkładu maksymalnej temperatury",  
  xlab = "Temperatura (°C)",  
  ylab = "Częstotliwość",  
  col = "skyblue", # Kolor słupków histogramu  
  border = "black", # Kolor krawędzi słupków  
  breaks = 65) # Liczba przedziałów (słupków) na histogramie
```

Histogram rozkładu maksymalnej temperatury



```
Histogram_min <- hist(temperatura_min,  
  main = "Histogram rozkładu minimalnej temperatury",  
  xlab = "Temperatura (°C)",  
  ylab = "Częstotliwość",  
  col = "skyblue", # Kolor słupków histogramu  
  border = "black", # Kolor krawędzi słupków  
  breaks = 50) # Liczba przedziałów (słupków) na histogramie
```

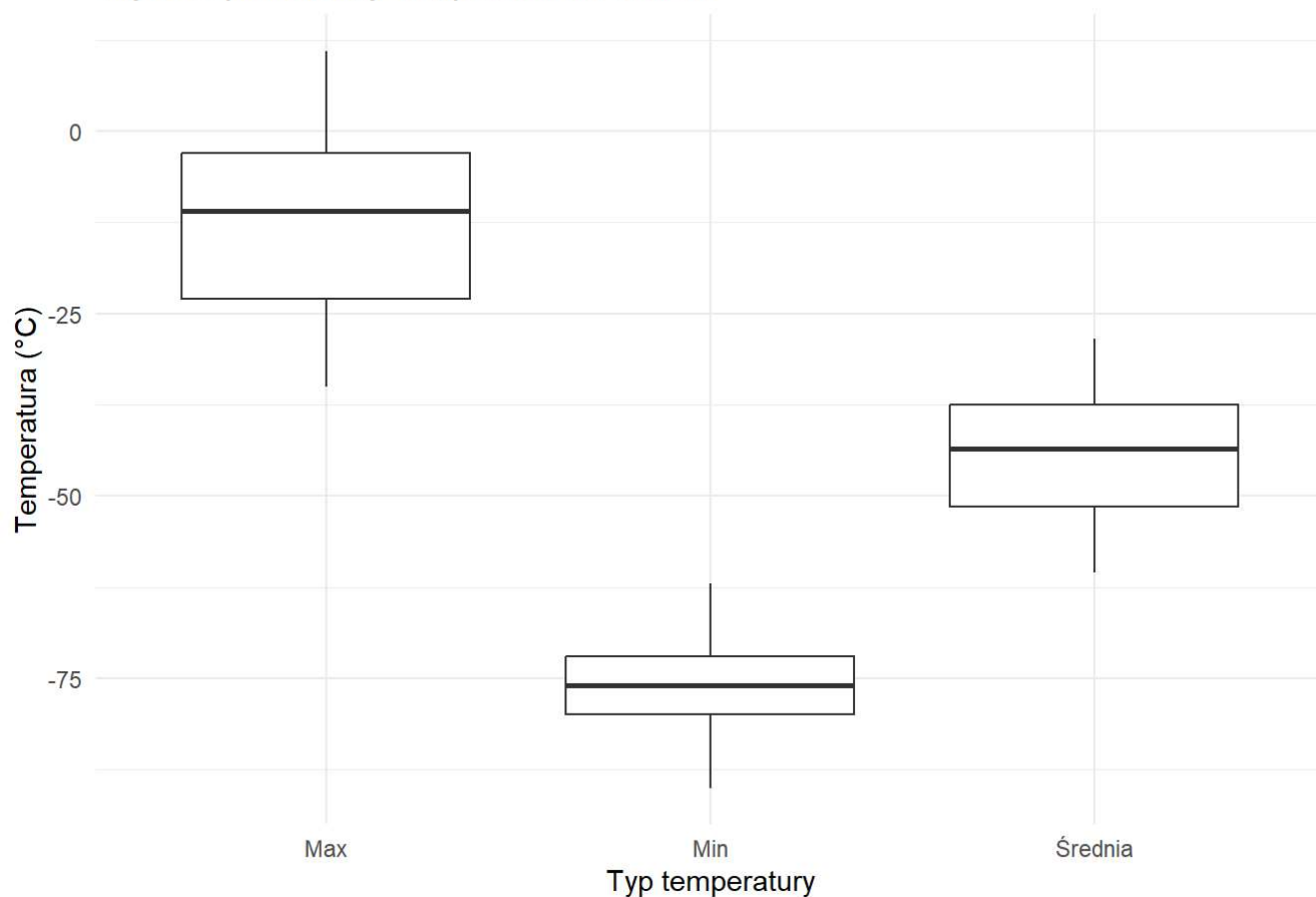
Histogram rozkładu minimalnej temperatury



5. Wykres pudełkowy temperatury

```
Temperatura_long <- data.frame(
  Data = rep(Temperatura$Data, 3),
  Typ_temperature = factor(rep(c("Max", "Min", "Średnia"), each = nrow(Temperatura))),
  Wartość = c(Temperatura$temperatura_max, Temperatura$temperatura_min, Temperatura$temperatura_srednia)
)
pudelkowy_wykres <- ggplot(Temperatura_long, aes(x = Typ_temperature, y = Wartość)) +
  geom_boxplot() +
  labs(title = "Wykres pudełkowy temperatur na Marsie",
       x = "Typ temperature",
       y = "Temperatura (°C)") +
  theme_minimal()
print(pudelkowy_wykres)
```

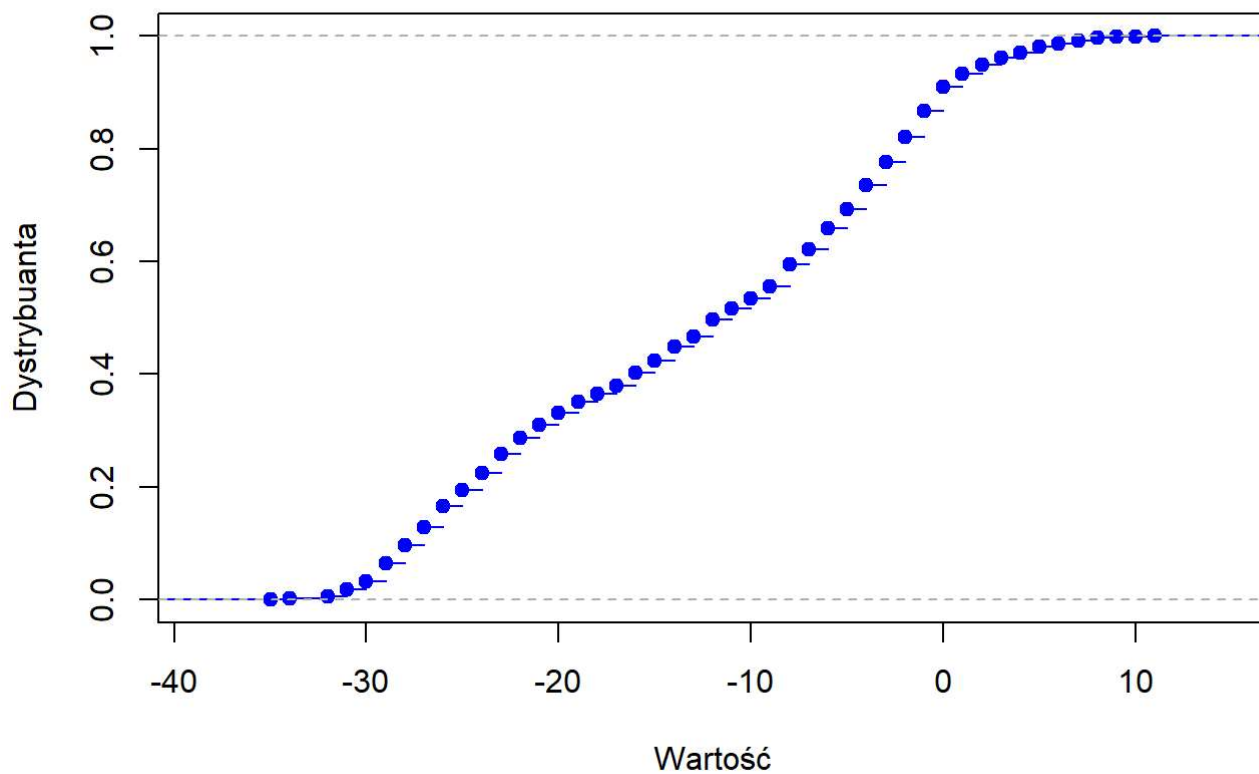
Wykres pudełkowy temperatur na Marsie



6. Dystrybuanta temperatury

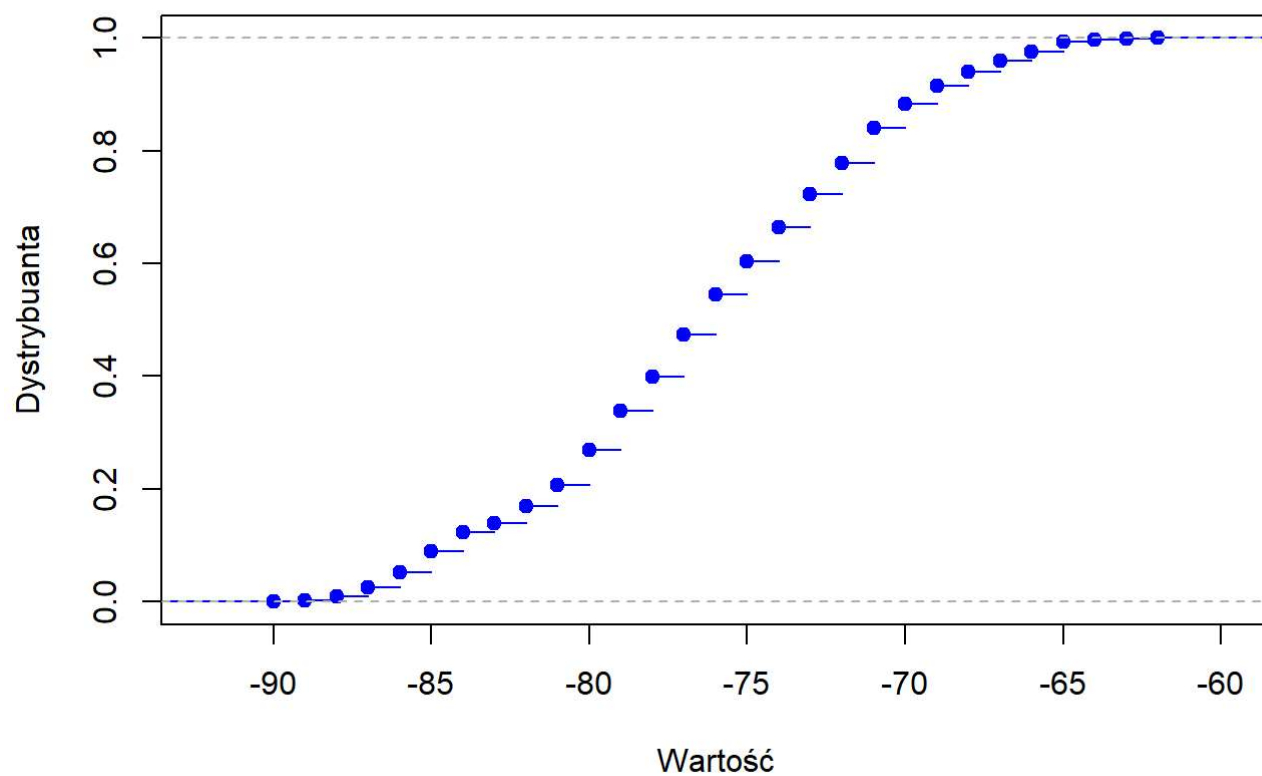
```
#dystrybuanta
dystrybuanta_max <- ecdf(temperatura_max)
plot(dystrybuanta_max,
     main = "Wykres dystrybuanty max",
     xlab = "Wartość",
     ylab = "Dystrybuanta",
     col = "blue")
```


Wykres dystrybuanty max



```
dystrybuanta_min <- ecdf(temperatura_min)
plot(dystrybuanta_min,
     main = "Wykres dystrybuanty min",
     xlab = "Wartość",
     ylab = "Dystrybuanta",
     col = "blue")
```

Wykres dystrybuanty min



7. Parametry opisowe ciśnienia na Marsie

```
#Parametry opisowe ciśnienia na Marsie
#srednia
Cisnienie_srednie <- mean(Pressure$cisnienie)
#Mediana
Mediana_pressure <- median(Pressure$cisnienie)

#dominanta
czestosci2 <- table(Pressure$cisnienie)
dominanta_cisnienie <- names(czestosci2)[which.max(czestosci2)]
#wariancja
wariancja_cisnienie <- var(Pressure$cisnienie)

#odchylenie standardowe
odchylenie_standardowe_cisnienie <- sd(Pressure$cisnienie)

#zakres
zakres_cisnienie <- range(Pressure$cisnienie)
rozstep_cisnienie <- diff(zakres_cisnienie)

#współczynnik zmienności
wspolczynnik_zmiennosci_c <- odchylenie_standardowe_cisnienie / Cisnienie_srednie

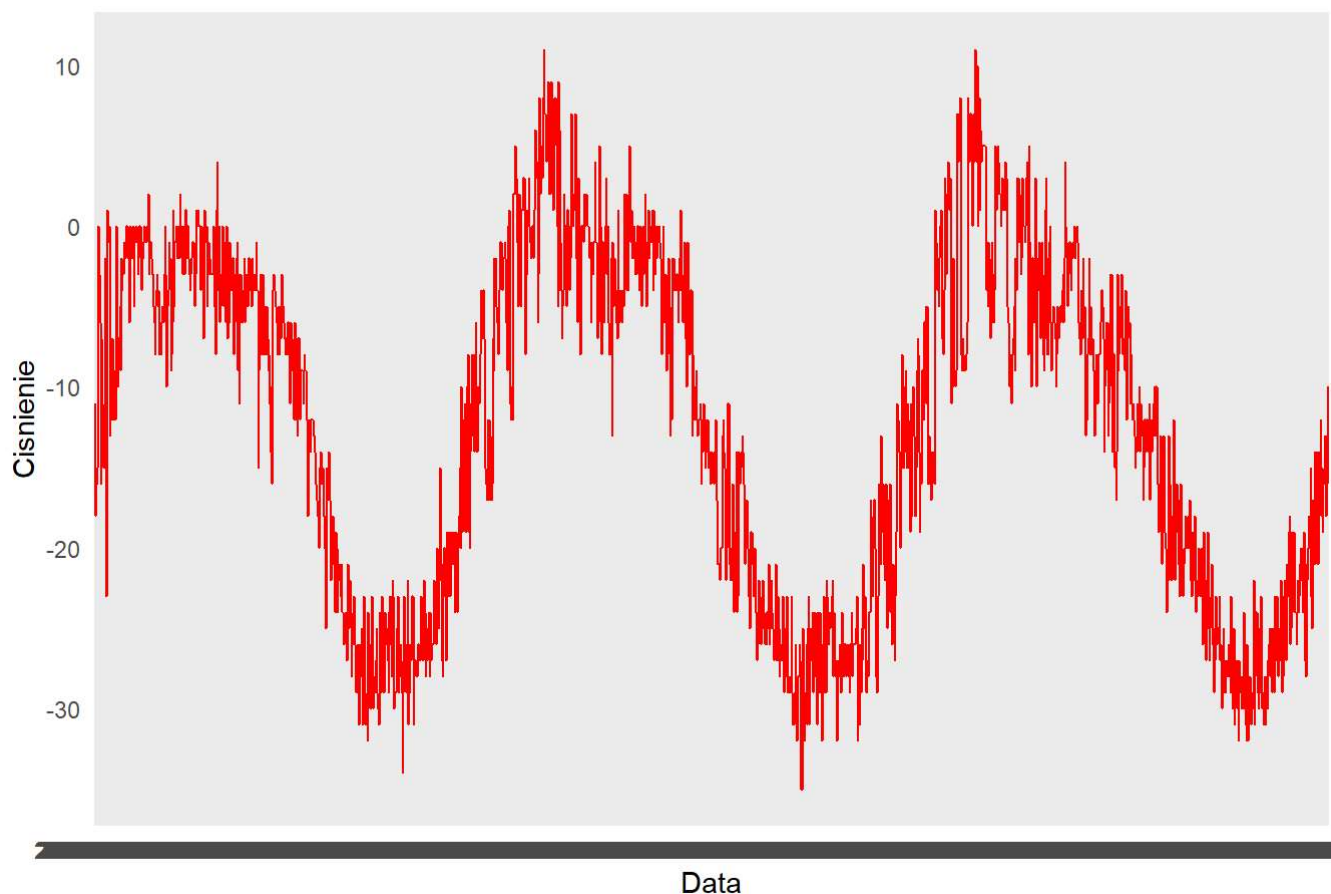
#Rozstęp międzykwartyłowy
iqr_cisnienie <- IQR(Pressure$cisnienie)
cat("Średnia ciśnienia: ", mean(Pressure$cisnienie), "\n",
    "Mediana ciśnienia: ", median(Pressure$cisnienie), "\n",
    "Dominanta ciśnienia: ", names(table(Pressure$cisnienie))[which.max(table(Pressure$cisnie
nie))], "\n",
    "Wariancja ciśnienia: ", var(Pressure$cisnienie), "\n",
    "Odchylenie standardowe ciśnienia: ", sd(Pressure$cisnienie), "\n",
    "Zakres wartosci ciśnienia: od ", range(Pressure$cisnienie)[1], " do ", range(Pressure$ci
snienie)[2], "\n",
    "Rozstęp wartosci ciśnienia: ", diff(range(Pressure$cisnienie)), "\n",
    "Rozstęp międzykwartyłowy ciśnienia: ", IQR(Pressure$cisnienie), "\n",
    "Współczynnik zmienności ciśnienia wynosi: ", wspolczynnik_zmiennosci_c, "\n")
```

```
## Średnia ciśnienia: 841.0664
## Mediana ciśnienia: 853
## Dominanta ciśnienia: 850
## Wariancja ciśnienia: 2943.413
## Odchylenie standardowe ciśnienia: 54.25323
## Zakres wartosci ciśnienia: od 727 do 925
## Rozstęp wartosci ciśnienia: 198
## Rozstęp międzykwartyłowy ciśnienia: 83
## Współczynnik zmienności ciśnienia wynosi: 0.06450528
```

8. Wykres ciśnienia

```
wykres_cisnienia <- ggplot(Pressure, aes(x = Data)) +  
  geom_line(aes(y = temperatura_max, group = 1), color = "red", linetype = "solid") +  
  labs(x = "Data", y = "Cisnienie", title = "Wykres ciśnienia") +  
  theme_minimal()  
wykres_cisnienia
```

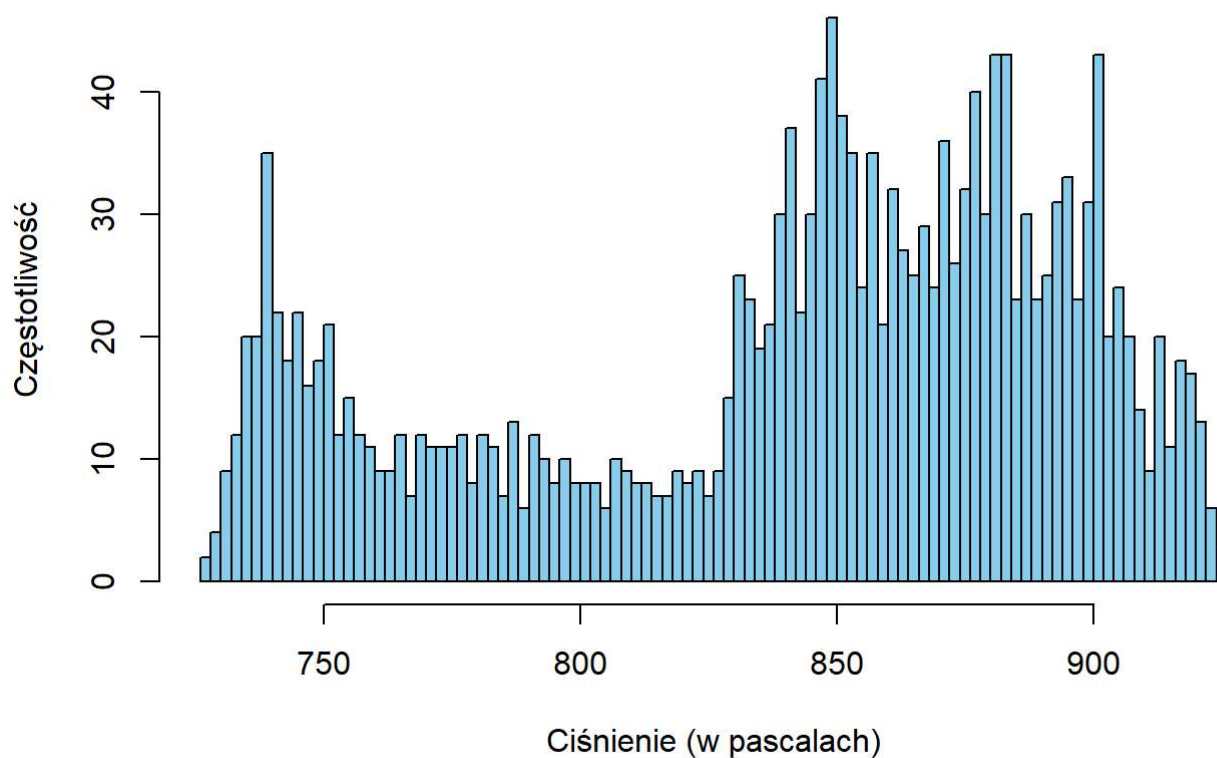
Wykres ciśnienia



9. Histogram ciśnienia

```
Histogram_cisnienie <- hist(Pressure$cisnienie,  
  main = "Histogram rozkładu cisnienia",  
  xlab = "Ciśnienie (w pascalach)",  
  ylab = "Częstotliwość",  
  col = "skyblue", # Kolor słupków histogramu  
  border = "black", # Kolor krawędzi słupków  
  breaks = 80) # Liczba przedziałów (słupków) na histogramie
```

Histogram rozkładu ciśnienia

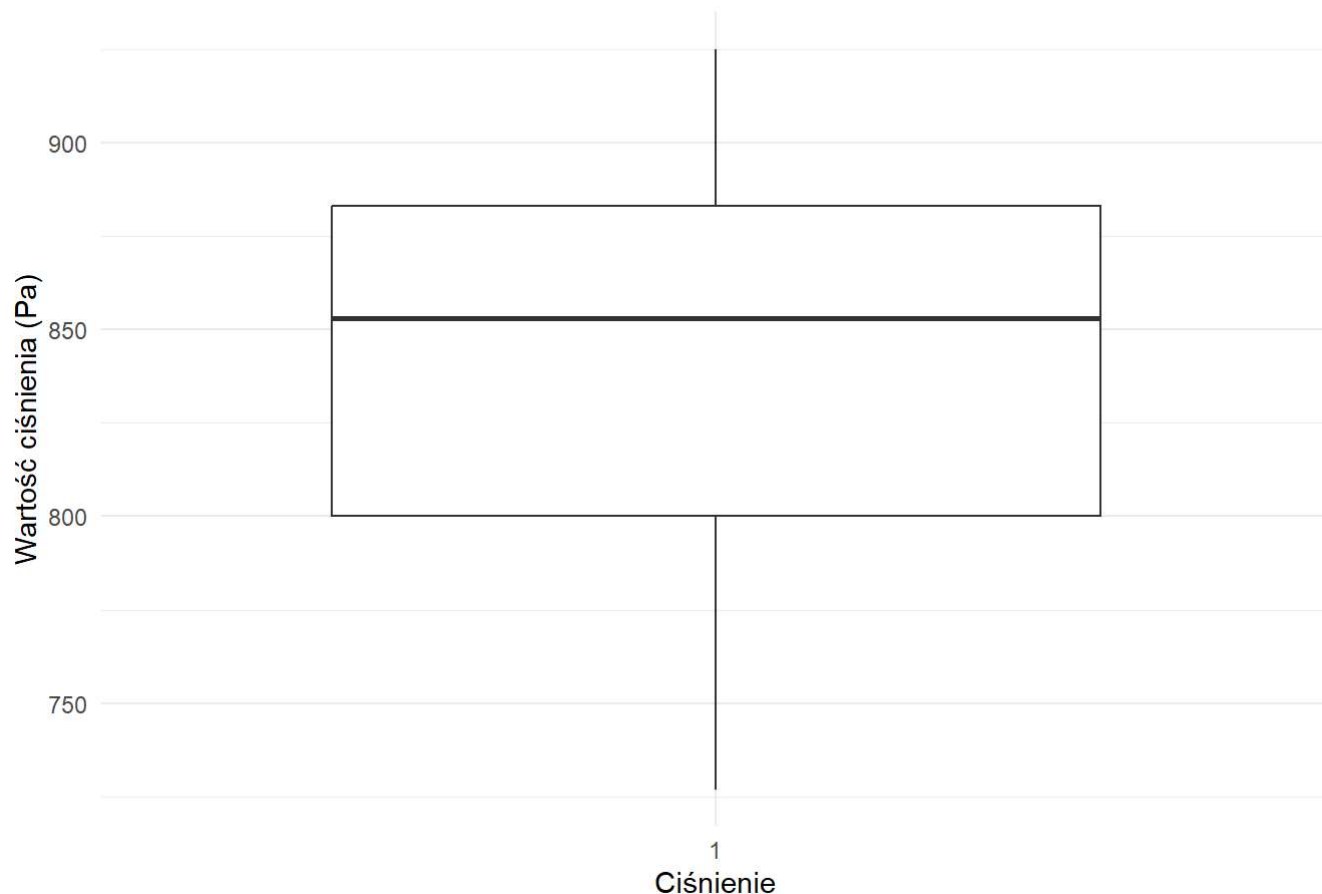


10. Wykres pudełkowy ciśnienia

```
Pressure_clean <- Pressure[complete.cases(Pressure$cisnienie), ]
pudelkowy_wykres_cisnienie <- ggplot(Pressure_clean, aes(x = factor(1), y = cisnienie)) +
  geom_boxplot() +
  labs(title = "Wykres pudełkowy ciśnienia na Marsie",
        x = "Ciśnienie",
        y = "Wartość ciśnienia (Pa)") +
  theme_minimal()

print(pudelkowy_wykres_cisnienie)
```

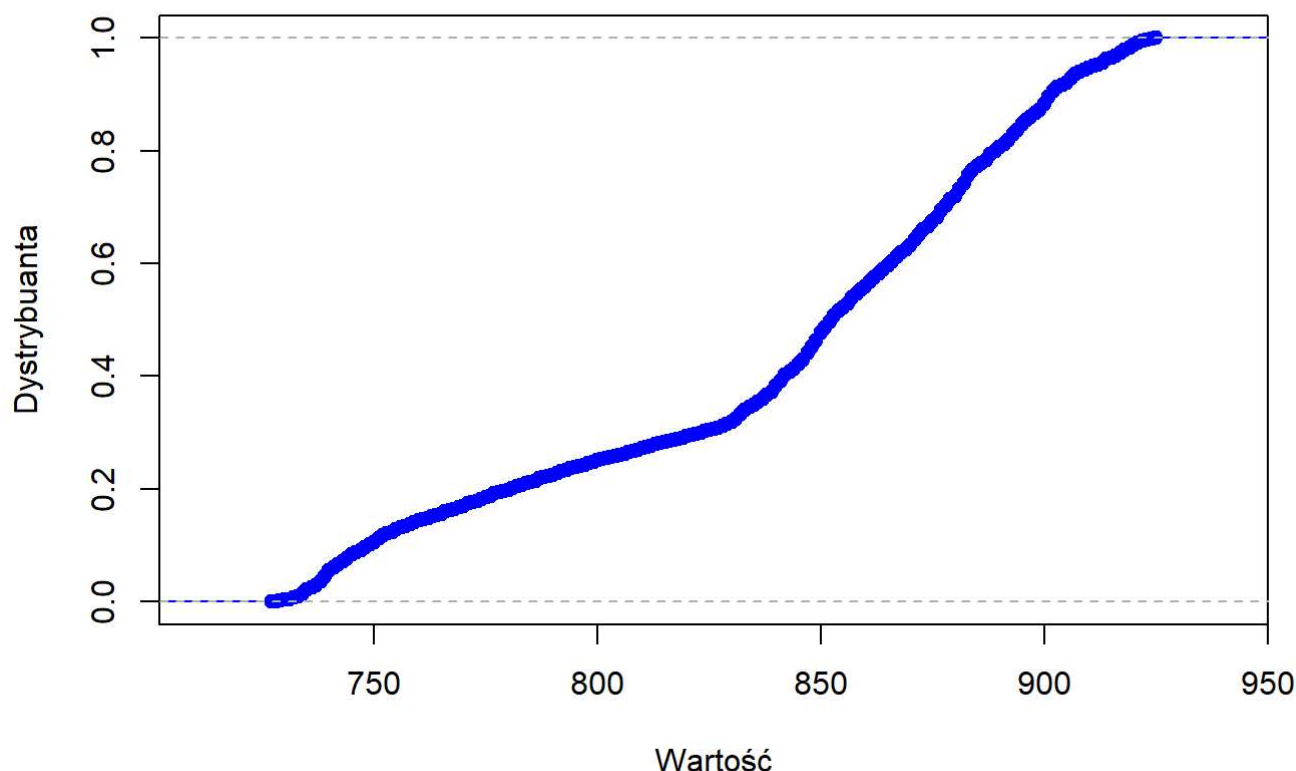
Wykres pudełkowy ciśnienia na Marsie



11. Dystrybuanta ciśnienia

```
dystybuanta_cisnienie <- ecdf(Pressure$cisnienie)
plot(dystybuanta_cisnienie,
     main = "Wykres dystrybuanty max",
     xlab = "Wartość",
     ylab = "Dystrybuanta",
     col = "blue")
```

Wykres dystrybuanty max



12. Hipoteza testu Wilcoxona

Na wykresie rozrzutu możemy zauważyć, że ciśnienia dla wybranych dat rozkładają się w różnych okresach. Na podstawie tego stwórzmy hipotezę.

Hipotezy: Hipoteza zerowa (H_0): Mediana wartości ciśnienia dla daty 2012-08-16 jest równa medianie wartości ciśnienia dla daty 2018-02-02. Hipoteza alternatywna (H_1): Mediana wartości ciśnienia dla daty 2012-08-16 jest różna od mediany wartości ciśnienia dla daty 2018-02-02. Chcemy ustalić, czy istnieje statystycznie istotna różnica między medianami wartości ciśnienia w dwóch wybranych datach. Jeśli wynik testu Wilcoxona wskazuje na istotną różnicę (wartość p jest mniejsza od ustalonego poziomu istotności, np. 0,05), możemy odrzucić hipotezę zerową i przyjąć, że mediany wartości ciśnienia są różne w tych dwóch datach. Jeśli wynik testu nie wskazuje na istotną różnicę (wartość p jest większa od poziomu istotności), nie mamy wystarczających dowodów, aby odrzucić hipotezę zerową.

Przeprowadzamy test Wilcoxona dla dwóch niezależnych prób, aby sprawdzić, czy mediany wartości ciśnienia dla wybranych dat są statystycznie różne. Użyjemy do tego funkcji `wilcox.test` z pakietu `stats`, który jest automatycznie załadowany do środowiska R.

```
date1 <- '2012-08-16'
date2 <- '2018-02-02'

pressure_date1 <- Pressure$cisnienie[Pressure$Data == date1]
pressure_date2 <- Pressure$cisnienie[Pressure$Data == date2]

wilcox_test_result <- wilcox.test(pressure_date1, pressure_date2)

print(wilcox_test_result)
```

```
##
## Wilcoxon rank sum exact test
##
## data: pressure_date1 and pressure_date2
## W = 0, p-value = 1
## alternative hypothesis: true location shift is not equal to 0
```

13. Hipoteza testu t-Studenta

Na wykresie gęstości możemy zauważyć, że najczęściej pojawiają się wyniki ciśnienia w okolicy 850 Pa. Na podstawie tego możemy stworzyć hipotezę.

Hipotezy: Hipoteza zerowa (H_0): Średnia wartość ciśnienia jest równa 400 Pa. Hipoteza alternatywna (H_1): Średnia wartość ciśnienia jest różna od 400 Pa. Test t-Studenta dla jednej próby (one-sample t-test) służy do sprawdzenia, czy średnia wartość w danym zbiorze danych (próbie) jest statystycznie różna od określonej wartości teoretycznej (hipotetycznej średniej). W naszym przypadku chcemy sprawdzić, czy średnie ciśnienie w zbiorze danych Pressure jest równe 400 Pa.

Jeśli wynik testu t-Studenta wskazuje, że różnica jest istotna (wartość p jest mniejsza od ustalonego poziomu istotności, np. 0,05), możemy odrzucić hipotezę zerową i przyjąć, że średnie ciśnienie różni się od 400 Pa. Jeśli wynik testu nie wskazuje na istotną różnicę (wartość p jest większa od poziomu istotności), nie mamy wystarczających dowodów, aby odrzucić hipotezę zerową.

Użyjemy do tego funkcji `t.test` z paczki `stats`, która jest automatycznie załadowana do środowiska R.

```
t_student <- t.test(Pressure$cisnienie, mu = 400)
print(t_student)
```

```
##
## One Sample t-test
##
## data: Pressure$cisnienie
## t = 351.28, df = 1866, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 400
## 95 percent confidence interval:
## 838.6039 843.5290
## sample estimates:
## mean of x
## 841.0664
```


14. Oto spis wszystkich funkcji użytych w analizie danych na temat temperatury i ciśnienia na Marsie, wraz z krótkim opisem każdej z nich:

1. `read.csv`: Funkcja używana do wczytania danych z pliku CSV do ramki danych w R.
2. `complete.cases`: Funkcja służąca do usuwania wierszy z brakującymi wartościami z ramki danych.
3. `rowMeans`: Funkcja obliczająca średnią wartość wierszy dla określonych kolumn w ramce danych.
4. `table`: Funkcja generująca tablicę częstości dla określonej kolumny w ramce danych.
5. `names`: Funkcja zwracająca lub ustawiająca nazwy dla obiektów R.
6. `which.max`: Funkcja zwracająca indeks największej wartości w wektorze.
7. `median`: Funkcja obliczająca medianę dla określonej wektora danych.
8. `var`: Funkcja obliczająca wariancję dla określonej wektora danych.
9. `sd`: Funkcja obliczająca odchylenie standardowe dla określonej wektora danych.
10. `range`: Funkcja zwracająca zakres wartości dla określonej kolumny w ramce danych.
11. `diff`: Funkcja obliczająca różnicę między dwiema wartościami w wektorze danych.
12. `IQR`: Funkcja obliczająca rozstęp międzykwartyłowy dla określonej kolumny w ramce danych.
13. `cat`: Funkcja służąca do wypisywania tekstu i wartości na konsoli R.
14. `ggplot`: Funkcja inicjująca tworzenie wykresu za pomocą pakietu `ggplot2`.
15. `geom_line`: Warstwa w `ggplot2` służąca do rysowania wykresu liniowego.
16. `labs`: Funkcja w `ggplot2` służąca do nadawania tytułów osiom i wykresom.
17. `theme_minimal`: Funkcja w `ggplot2` służąca do ustawienia minimalistycznego motywu wykresu.
18. `hist`: Funkcja służąca do tworzenia histogramu.
19. `factor`: Funkcja konwertująca wektory na faktory (kategorie).
20. `geom_boxplot`: Warstwa w `ggplot2` służąca do rysowania wykresu pudełkowego.
21. `print`: Funkcja wypisująca obiekty na konsoli R.
22. `ecdf`: Funkcja tworząca empiryczną dystrybucję dla określonego wektora danych.
23. `plot`: Funkcja rysująca różne typy wykresów w R.
24. `mean`: Funkcja obliczająca średnią wartość dla określonego wektora danych.
25. `wilcox.test`: Funkcja przeprowadzająca test Wilcoxa dla określonych grup danych.
26. `t.test`: Funkcja przeprowadzająca test t-studenta dla określonego wektora danych.
27. `library`: Funkcja załadowująca pakiety do środowiska R.
28. `header`: Argument funkcji `read.csv` określający, czy plik CSV zawiera nagłówki kolumn.
29. `sep`: Argument funkcji `read.csv` określający separator używany w pliku CSV.
30. `na.strings`: Argument funkcji `read.csv` określający wartości, które mają być interpretowane jako braki danych.