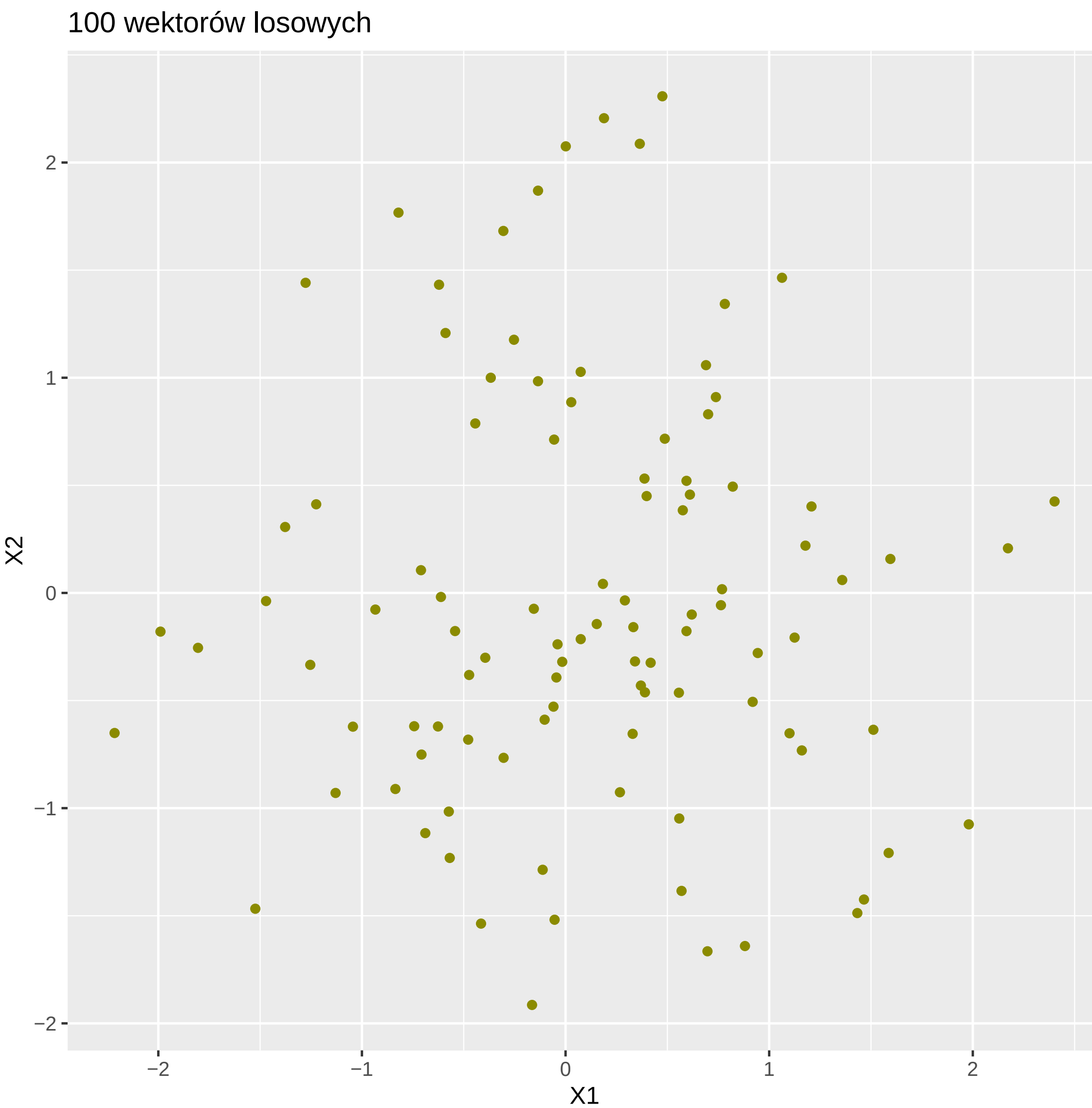
Zadanie 1. Rozważamy dwuwymiarowy rozkład normalny $N(0, \mathbb{1})$. Poniżej przedstawione jest 100 wektorów losowych pochodzących z tego rozkładu wygenerowanych przy użyciu funkcji *rnorm*.

```
sample <- matrix(rnorm(200), nrow = 100, ncol = 2)
```



Na powyższej chmurze możemy zaobserować duże rozproszenie punktów, co pociąga wniosek o braku korelacji między zmiennymi losowymi X_1, X_2. Możemy zatem przypuszczać niezależność zmiennych X_1, X_2, ponieważ wiemy, że dla rozkładu gaussowskiego nieskorelowanie zmiennych losowych pociąga ich niezależność.

Zadanie 2. Rozważamy otrzymaną powyżej chmurę, którą oznaczać będziemy przez X. W pierwszej części przekształcimy ją w chmurę Y z rozkładu $N(\mu, \Sigma)$, gdzie

$$\mu = (4, 2),\ \Sigma = \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}.$$

Musimy zatem znaleźć macierz A oraz wektor B, dla których spełniona będzie następująca zależność

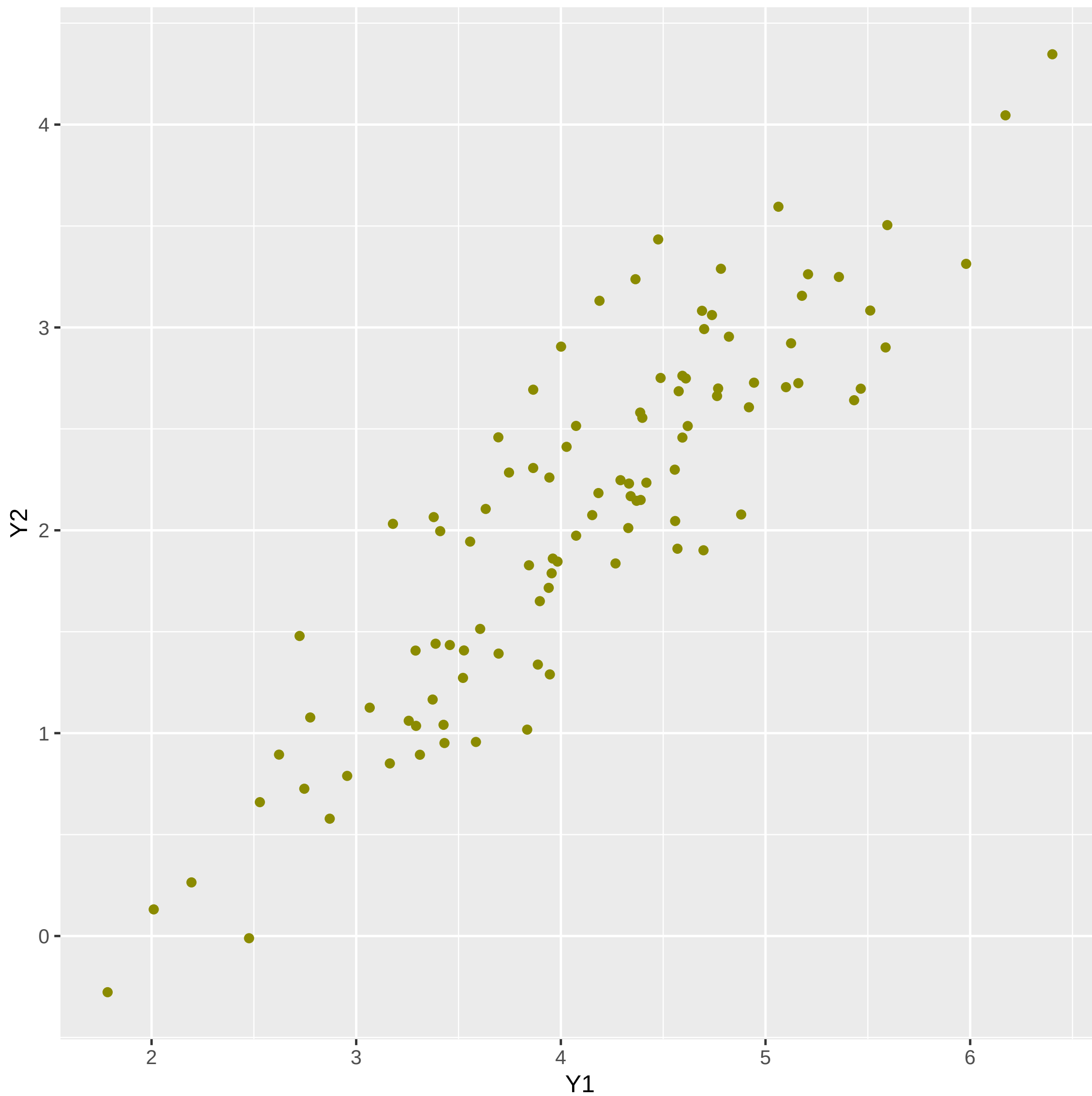
$$(4, 2)^T = A(0, 0)^T + B,\ \Sigma = AA^T,$$

ponieważ wówczas $Y = AX + B$. Oczywiście $B = (4, 2)^T$. Dalej zauważmy, że równanie $\Sigma = AA^T$ ma nieskończenie wiele rozwiązań. Przykładowym z nich jest

$$A = \begin{bmatrix} 1 & 0 \\ 0.9 & \sqrt{19} \cdot 10^{-1} \end{bmatrix}.$$

Po przekształceniu chmury X w opisany wcześniej sposób otrzymujemy chmurę Y przedstawioną na poniższym wykresie.

```
A <- matrix(c(1, 0, 0.9, sqrt(19)/10), byrow = TRUE, nrow = 2)
B <- c(4, 2)
transformed_cloud <- data.frame(t(sapply(c(1:100), function(i)
                                     A %*% sample[i, ] + B)))
```



Widzimy, że punkty układają się wzdłuż prostej, a zatem możemy wnioskować o skorelowaniu zmiennych $Y1$, $Y2$. Ponadto widzimy, że wraz ze wzrostem wartości $Y1$ rośnie wartość $Y2$. Zauważmy, że wynika to również z wartości macierzy Σ, tj. $\Sigma(2, 1) = 0.9$. Możemy więc przypuszczać zależność zmiennych $Y1, Y2$.

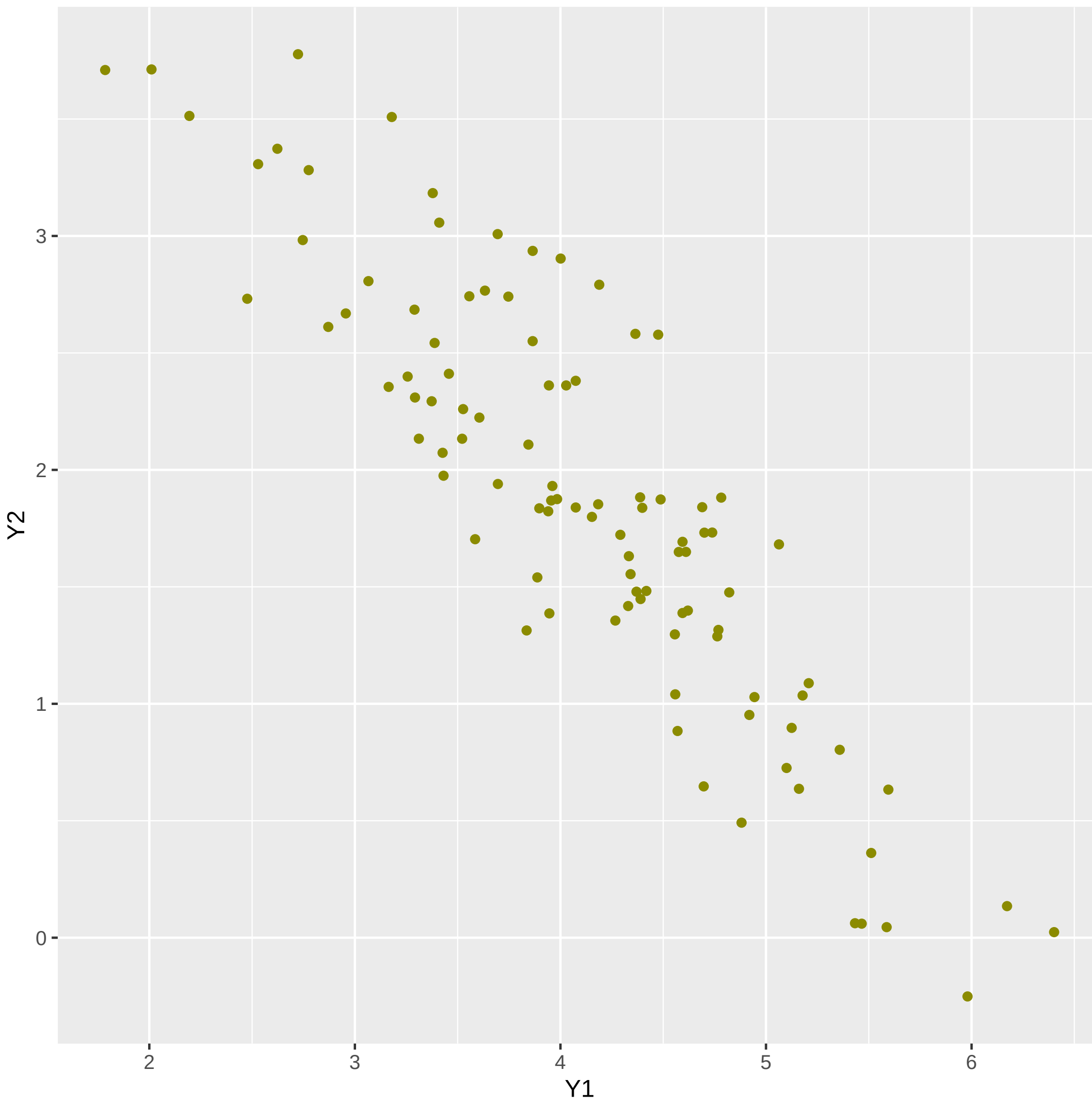
Powtarzając powyższe rozumowanie dla

$$\mu = (4, 2),\ \Sigma = \begin{bmatrix} 1 & -0.9 \\ -0.9 & 1 \end{bmatrix}$$

otrzymujemy

$$B = (4, 2),\ A = \begin{bmatrix} 1 & 0 \\ -0.9 & \sqrt{19} \cdot 10^{-1} \end{bmatrix}.$$

Otrzymana przez przekształcenie chmury X chmura Y ma następującą postać:



Podobnie jak we wcześniejszym przykładzie widzimy, że punkty układają się wzdłuż prostej, a więc zmienne $Y1$, $Y2$ są skorelowane. Ponadto widzimy, że wraz ze wzrostem wartości $Y1$ maleje wartość $Y2$. Zauważmy, że wynika to również z wartości macierzy Σ, tj. $\Sigma(2, 1) = -0.9$. Możemy więc przypuszczać zależność zmiennych $Y1, Y2$.

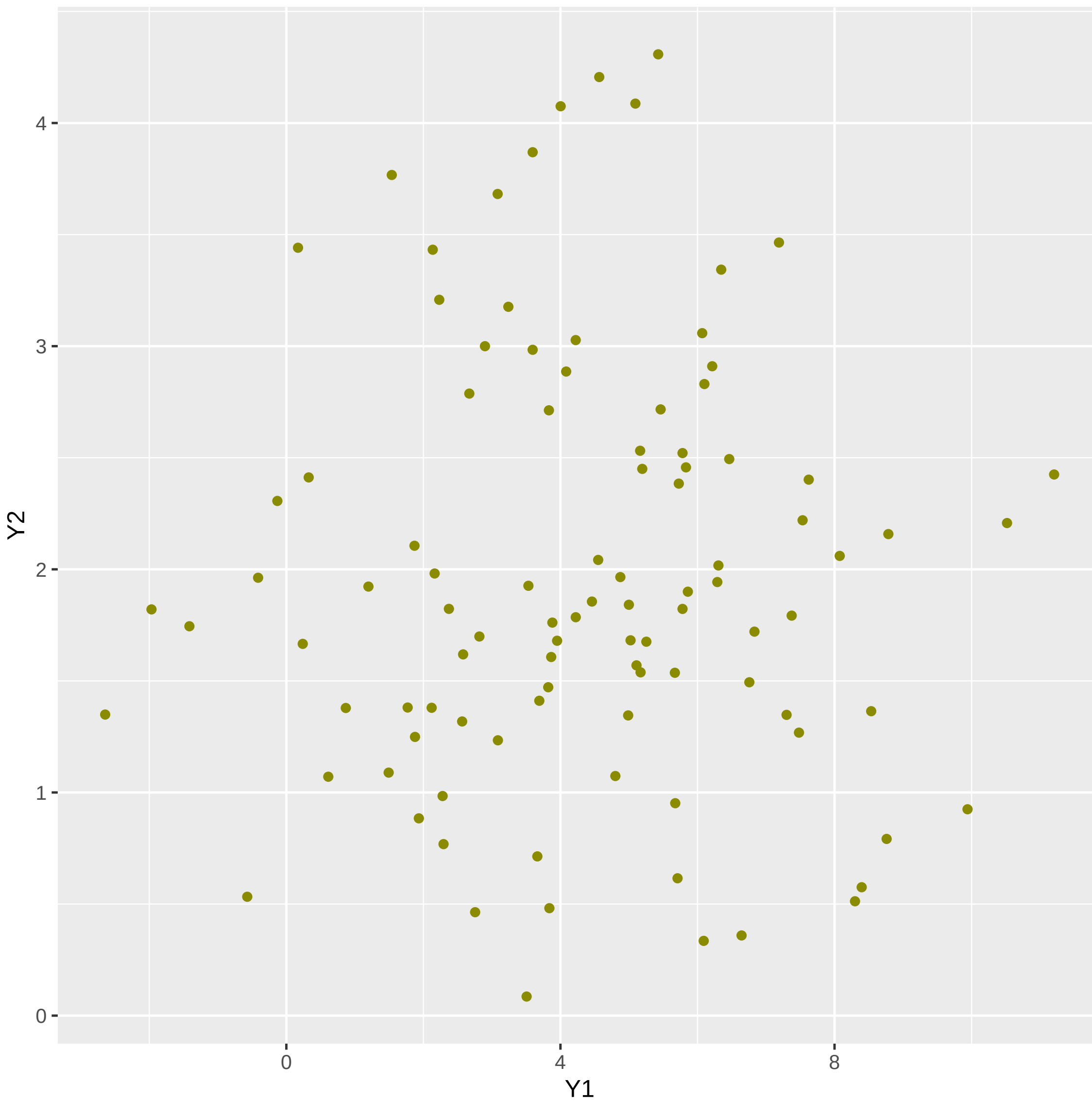
Na koniec powtarzając wcześniejsze rozumowania dla

$$\mu = (4, 2),\ \Sigma = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$

otrzymujemy

$$B = (4, 2), \; A = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}.$$

Otrzymana przez przekształcenie chmury X chmura Y ma następującą postać:



Na powyższej chmurze możemy zaobserować duże rozproszenie punktów, a zatem zmienne $Y1, Y2$ nie są skorelowane. Zauważmy, że wynika to również z wartości macierzy Σ, tj. $\Sigma(2,1) = 0$. Otrzymujemy więc wniosek o niezależności zmiennych $Y1, Y2$.

Zadanie 3. Zaczynamy od wygenerowania 200 wektorów losowych z rozkładu wielowymiarowego normalnego $N(0, \mathbb{1}_{100^2})$, które zapiszemy w macierzy $N_{200\times 100}$

```
X <- matrix(rnorm(20000, 0, 1), nrow = 200, ncol = 100)
```

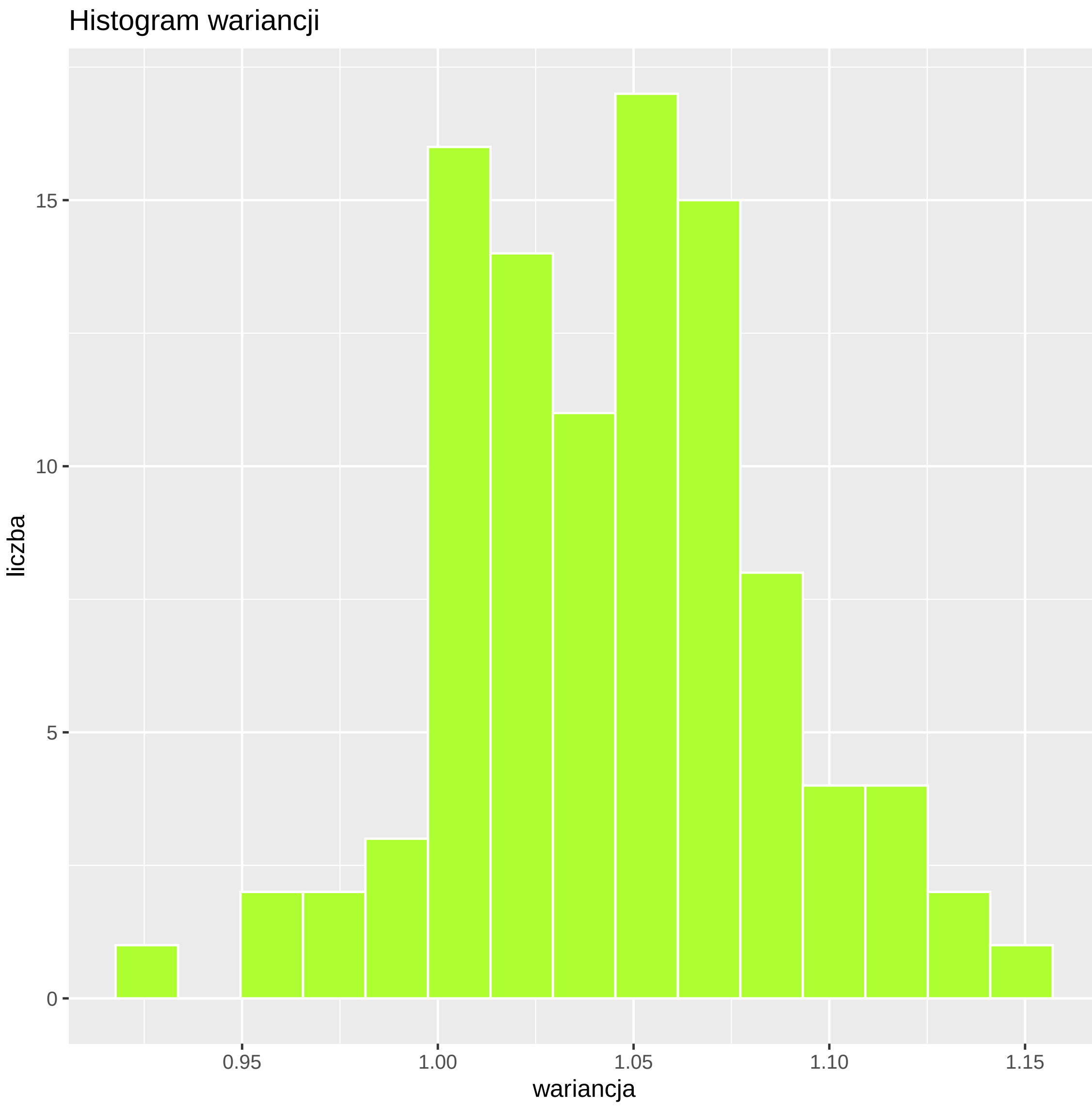
Następnie przy użyciu rozkładu Choleskiego wyznaczymy macierz A, dla której macierz $Y = XA$ zawierać będzie 200 wektorów z rozkładu wielowymiarowego normalnego $N(0, \Sigma_{100\times 100})$, gdzie $\Sigma(i,i) = 1$ i $\Sigma(i,j) = 0.9$ dla $i \neq j$.

```r
Y <- matrix(c(rep(c(1, rep(0.9, 100)), 99), 1), nrow = 100, ncol = 100)
A <- chol(Y)
M = X %*% A
```

Sprawdzanie wyniku rozpoczynamy od wyliczenia średniej oraz przedstawienia histogramu próbkowych wariancji współrzędnych M. Otrzymujemy, że średnia wynosi:

```
## [1] 1.045
```

Poniżej zaś predstawiamy histogram:



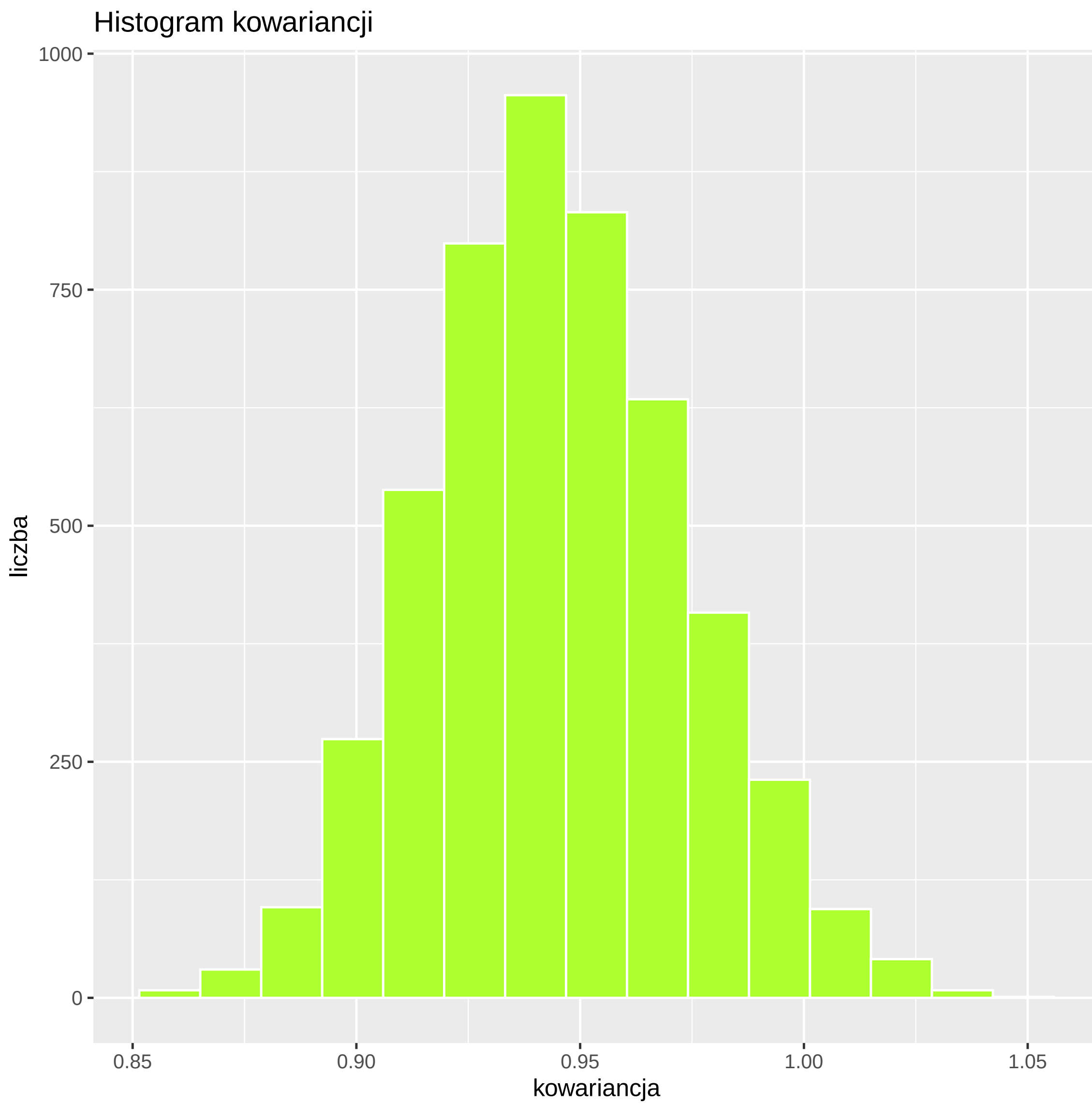
Na koniec przedstawimy średnią oraz histogram próbkowych kowariancji między różnymi współrzędnymi M.

```r
N = matrix(NA, 100, 100)
for (i in 1:100) {
  for (j in 1:100) N[i,j] = cov(M[,i],M[,j])
}
D=N[upper.tri(N)]
```

Średnia wynosi:

```
## [1] 0.9449308
```

Zaś histogram prezentuje się następująco:



Widzimy zatem, że uzyskana macierz A spełnia dane w zadaniu warunki.