МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНОМУ УНІВЕРСИТЕТІ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

Кафедра систем штучного інтелекту

**Лабораторна робота № 2**

з дисципліни

«Візуалізація даних»

Виконав:

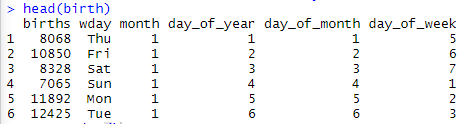
студент групи КН-208

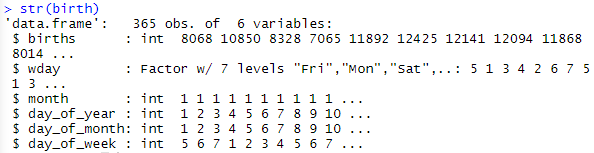
Келемен С. Й.

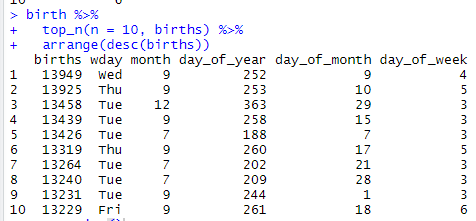
**Викладач:**

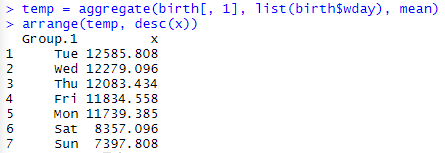
Бойко Н. І.

Львів – 2019 р.

Дослідимо структуру нашого датасету:  




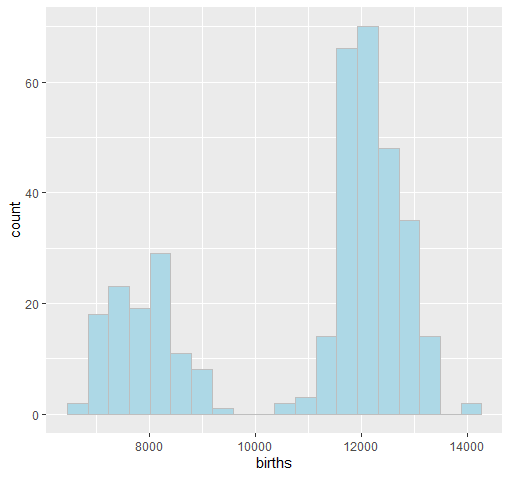
Знайдемо топ 10 днів в році, коли народилося найбільше дітей:  


Також знайдемо топ днів тижня по кількості народжуваних дітей:  


Побудуємо гістограму для кількості народжуваних:  
ggplot(birth, aes(x=births)) +

geom\_histogram(bins=20, color="grey", fill="lightblue") +

ylab("Days")



Знайдемо середнє значення:



Знайдемо середньоквадратичне відхилення:



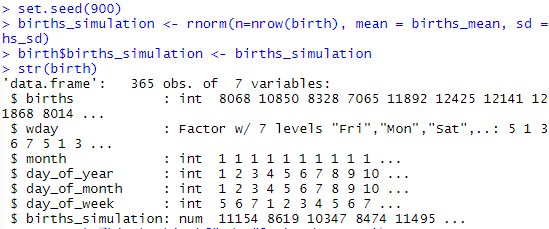
Згенеруємо нормальний розподіл, який має середнє  
значення births\_mean та середньоквадратичне відхилення  
births\_sd. Для цього використаємо функцію rnorm. Для того,  
щоб послідовність, яка генерується була сталою, при кожному  
виконанні нашого коду, встановимо параметр set.seed

set.seed(900)

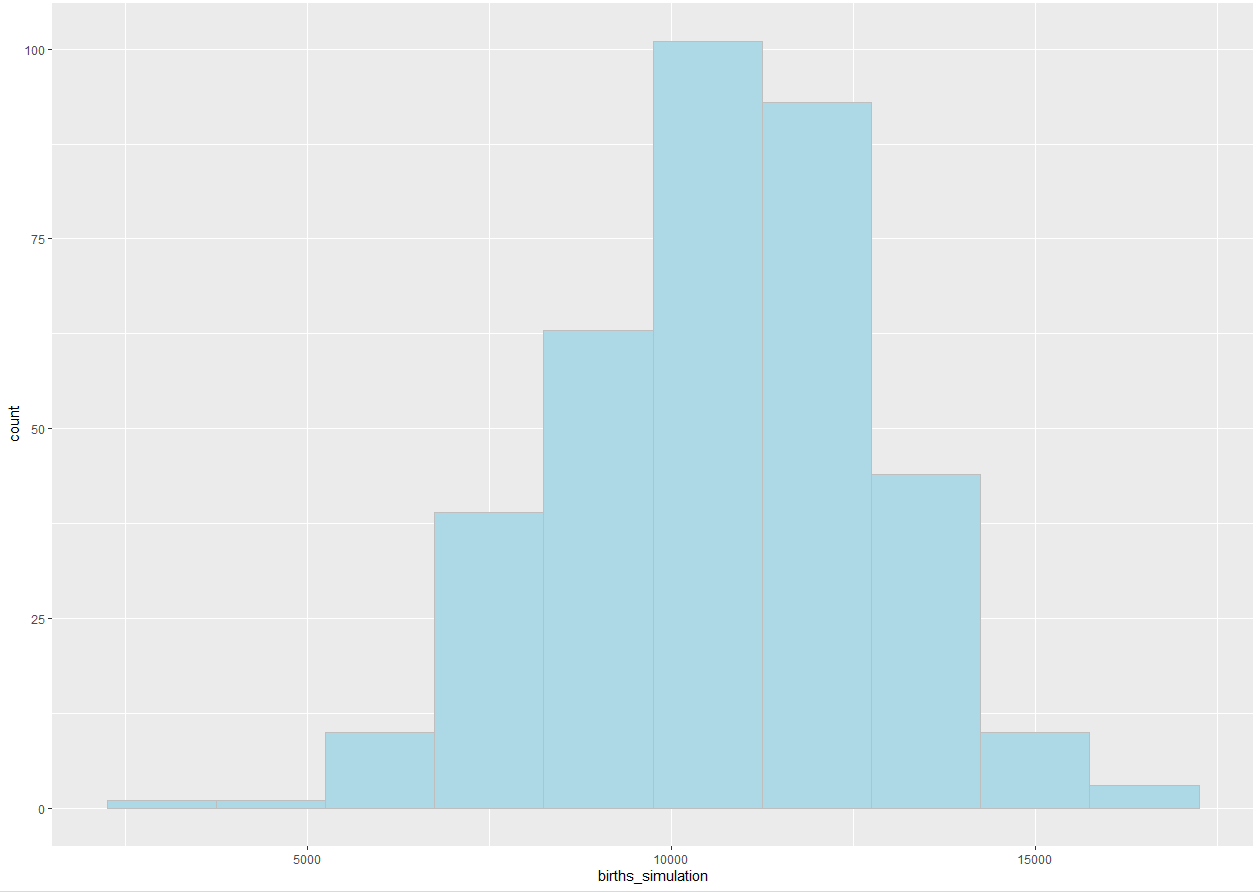
births\_simulation <- rnorm(n=nrow(birth), mean = births\_mean, sd = births\_sd)

birth$births\_simulation <- births\_simulation

str(birth)

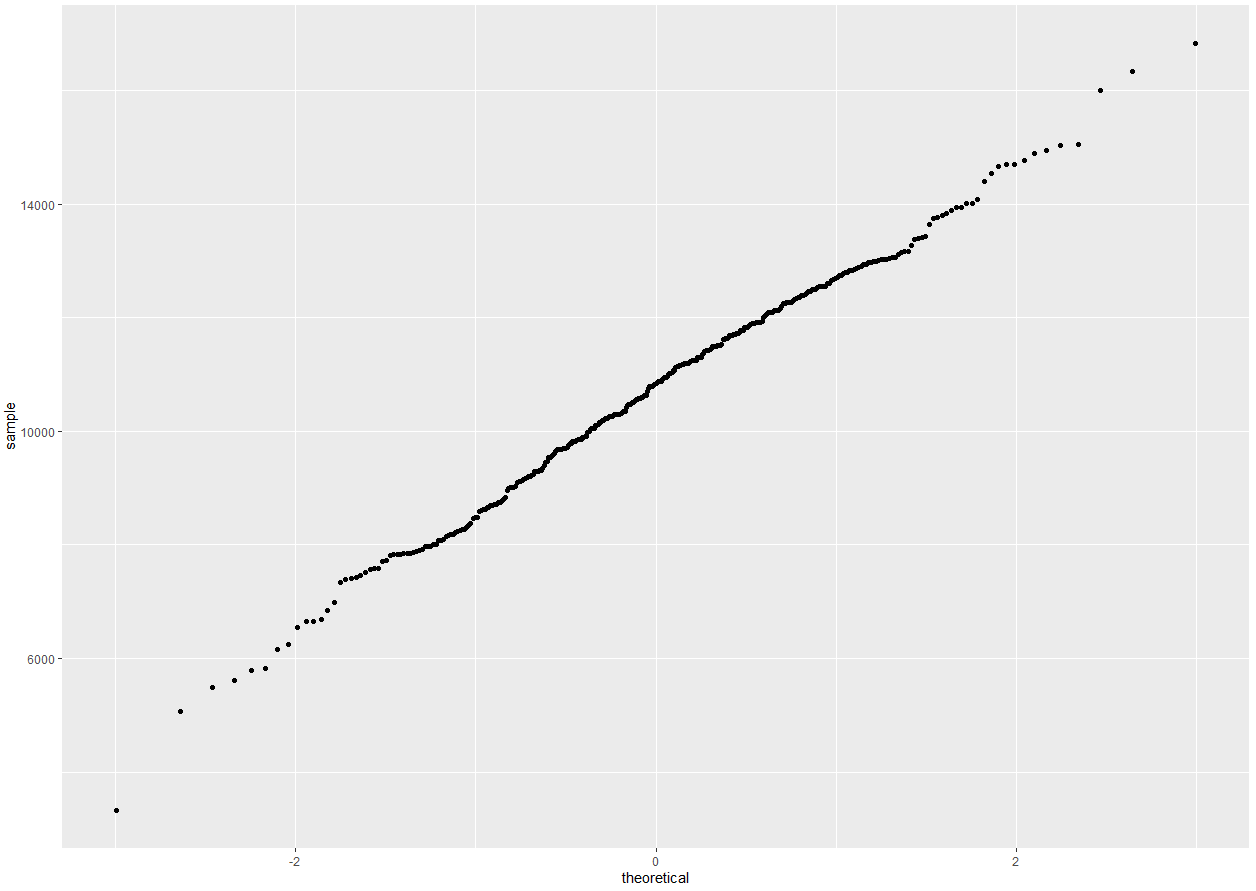


Побудуємо гістограму для цієї симуляції:  
ggplot(birth, aes(x=births\_simulation)) +  
geom\_histogram(bins=10, color="grey", fill="lightblue")

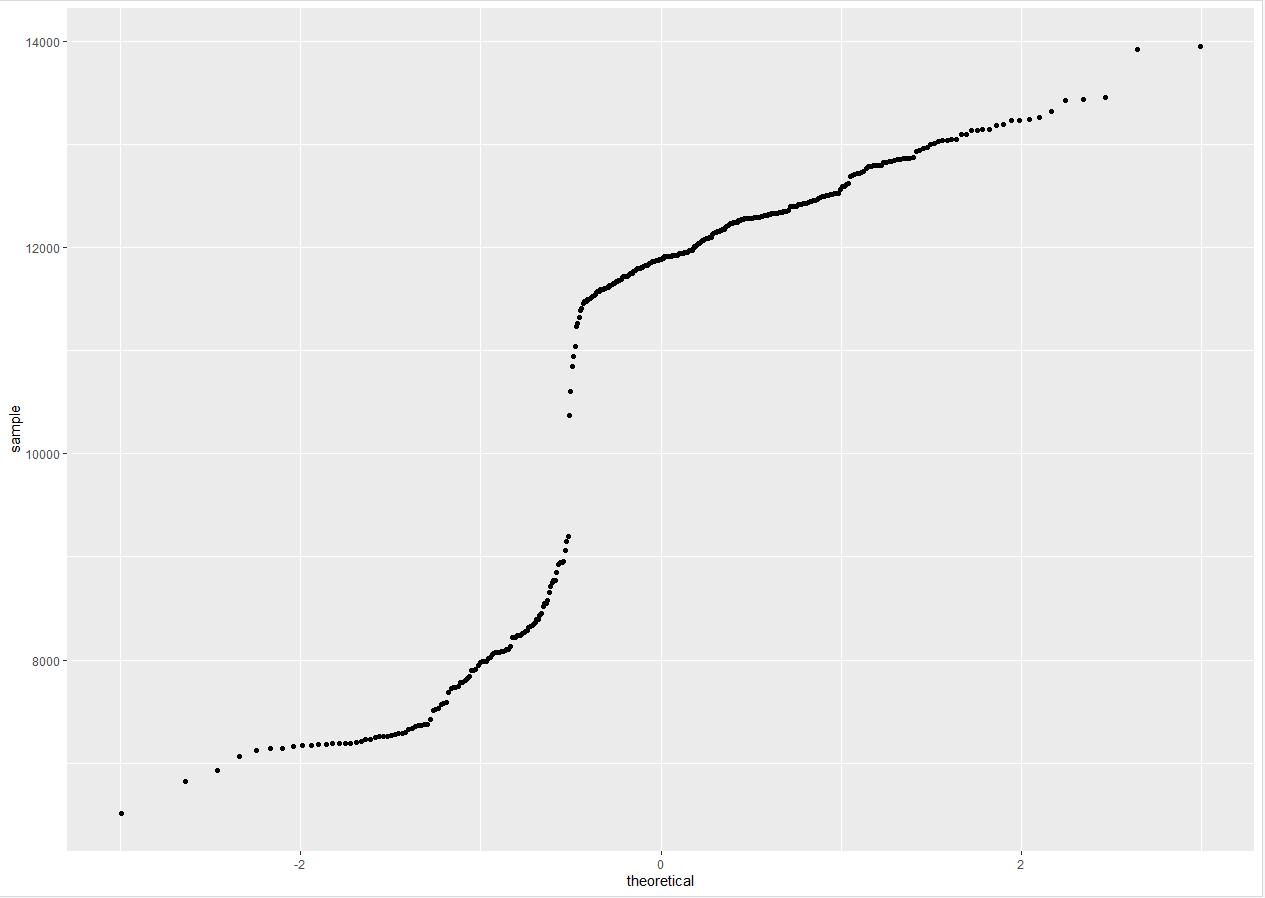


Для перевірки, чи є розподіл нормальним, використовується  
функція qqplot. Cкористаємося нею для перевірки чи є  
нормально розподілені дані народжуваності. Спочатку  
побудуємо qqplot для нашої симуляції births\_simulation:

ggplot(birth, aes(sample = births\_simulation)) + stat\_qq()



А тепер для справжніх даних:  
**ggplot**(movie\_body\_counts, **aes**(sample = IMDB\_Rating)) + **stat\_  
qq**()



**Висновок**

Знайшов середнє значення та середньоквадратичне відхилення для рейтингу даних фільмів, також побудував гістограму народжуваності та її симуляцію у вигляді нормального розподілу, в кінці перевірив чи нормально розподілена симуляція та оригінальні дані.