**Міністерство Освіти І НАУКИ України**

**Національний університет "Львівська політехніка"**

Інститут **ІКНІ**

Кафедра **СШІ**

**ЗВІТ**

До лабораторної роботи №4

**З дисципліни:** «Візуалізація даних»

**На тему:** «Аналіз даних та статистичне виведення на мові R»

**Виконав:**

ст. гр. КН-307

Шиманський П.С.

**Прийняв:**

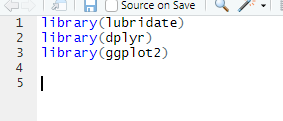
Шамуратов О.Ю.

Львів - 2019

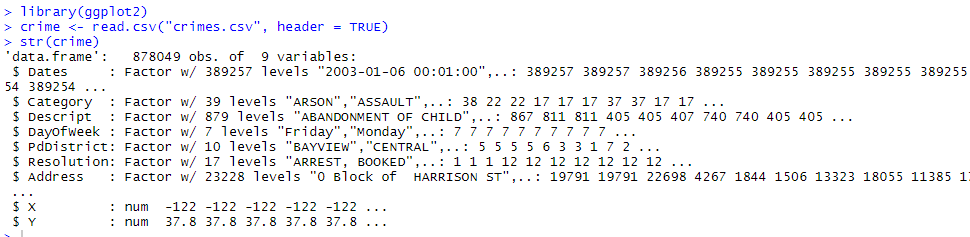
**Мета:** Аналіз даних та статистичне виведення на мові R.

**Хід роботи:**

1. Завантажив бібліотеки:



1. Завантажимо набір даних, який містить інформацію про злочини, які відбулися в Сан-Франциско (при потребі змініть шлях до файлу "crimes.csv"):



1. Маємо 878049 записи. Поле Dates містить інформацію про дату та час злочину. Нам достатньо лише дати. Для форматування дати будемо використовувати бібліотеку lubridate, зокрема функцію ymd\_hms(яка вказує, як перетворити стрічку в дату):



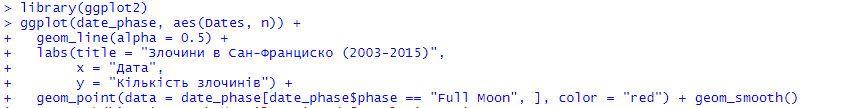
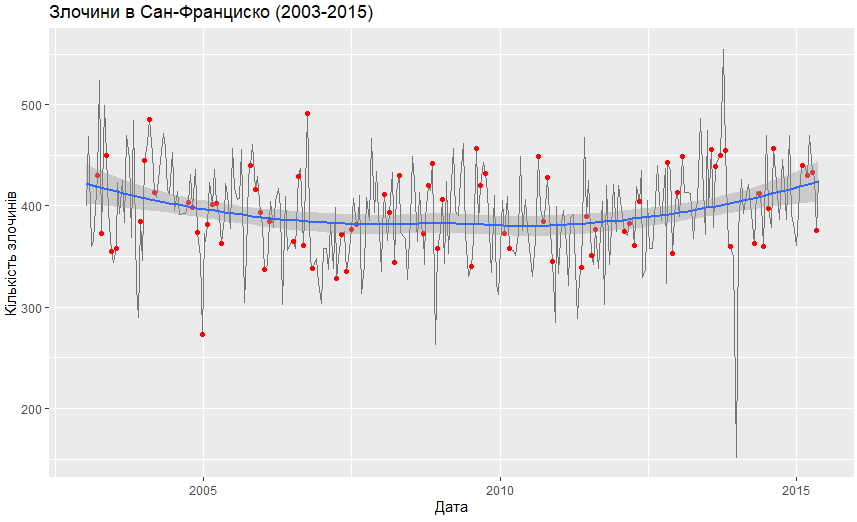
1. Завантажимо набір даних про фази Місяця (при потребі змініть шлях до файлу "moon.csv"):



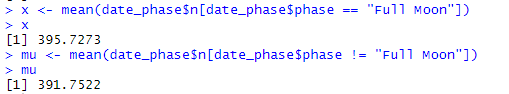
1. Перетворимо поле date в однаковій формат із полем Dates в crime:



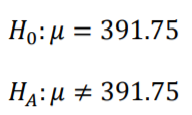
1. Об'єднаємо набори даних за датою. Тобто об'єднаємо інформацію про злочини та фазу місяця для кожної дати. Для цього використаємо функцію merge.
2. Побудуємо графік кількості злочинів по днях та позначимо значення для днів, коли спостерігався повний Місяць, червоними точками:

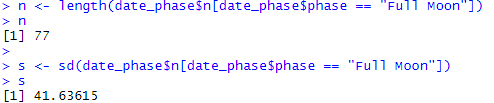
1. Давайте знайдемо середні значення для днів в котрі маємо фазу повного Місяця і решти фаз:



1. Отже, маємо, що середнє значення кількість злочинів у дні повного Місяця 395.73, в середня кількість у дні інших фаз - 391.75. Будемо вважати середню кількість злочинів у дні інших фаз оцінкою середнього значення генеральної сукупності.
2. Визначимо, чи середня кількість злочинів у дні повного Місяця статистично відрізняється від цього значення. Сформулюємо гіпотези нашого тесту:



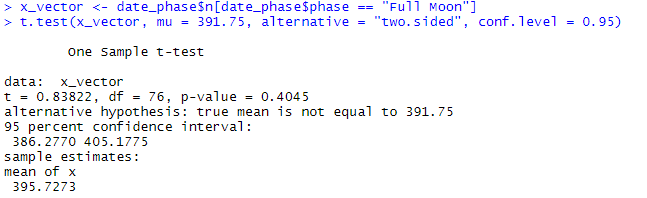
Знайдемо розмір та середньоквадратичне відхилення нашої вибірки:



1. Та знайдемо p-value:

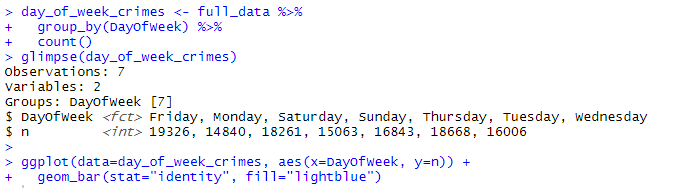
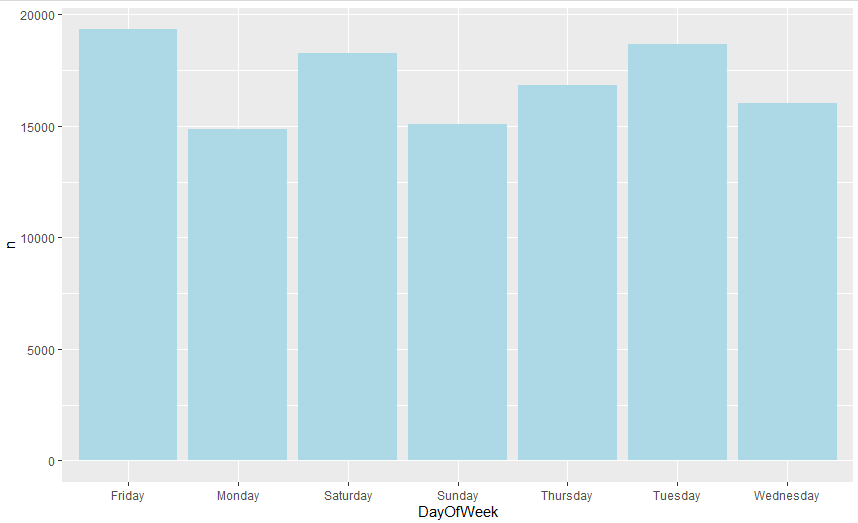


1. Також можемо використати функцію t-test, вказавши вектор значень та параметри mu = 391.75, alternative = "two.sided", conf.level = 0.95.



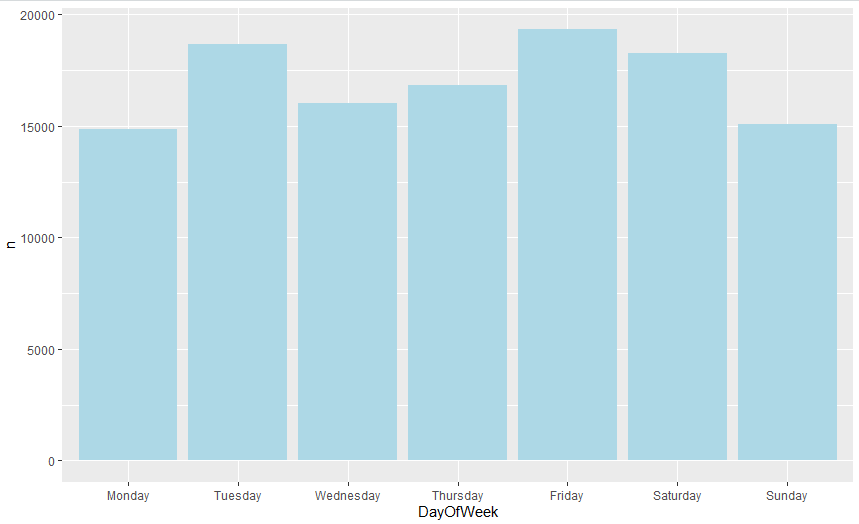
Отже, згідно нашого тесту, фаза Місяця не має впливає на середнє значення кількості злочинів. Можна також дослідити, чи впливає фаза Місяця на кількість злочинів, які можуть бути вчинені імпульсивно.

1. Давайте перевіримо, чи впливає день тижня на кількість злочинів. Обрахуємо кількість злочинів для кожного дня тижня та побудуємо графік:

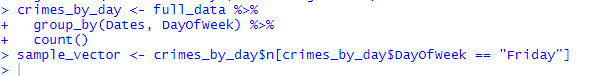
 

1. Змінимо порядок днів, на звичний для нас.





1. Перевіримо, чи середня кількість злочинів, вчинених по п'ятницях, відрізняється від середнього значення 391.75. Сформуємо нашу вибірку sample\_vector, яка міститиме значення кількості злочинів в Сан-Франциско для кожної п'ятниці:



Яка кількість ступенів вільності для цієї вибірки?

Відповідь: 45

Обрахуйте t-статистику.

Відповідь: 4.4

Обрахуйте p-value, результат заокругліть до четвертого знака після коми (X.XXXX)

Відповідь: 0.0001

Чи можемо ми вважати, що середня кількість злочинів вчинених по п'ятницях статистично відрізняється від значення 391.75?

Відповідь: Так

Побудуйте довірчий інтервал для рівня довіри 99%. Чи містить він значення нульової гіпотези ?

Відповідь: Ні

**Висновок:**

На лабораторній роботі я ознайомивя з аналізом даних та статистичним виведенням на мові R.