НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Кафедра автоматизації проектування енергетичних процесів і систем

Лабораторна робота № 5

З дисципліни: «Big Data та аналіз даних»

**Виконав:**

студент 5 курсу

групи ТР-02мп, ТЕФ

Круглий Д.В.

**Перевірив:**

Матичин І.І.

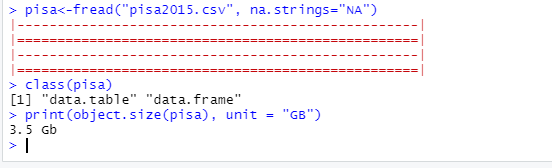
Київ – 2021

**Постановка задачі**

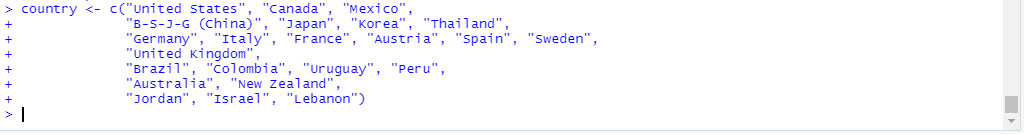
Використовуючи вибірку по 500 учнів з кожного регіону, на одному рисунку створити точкові діаграми (scatterplot) залежності балів з математики (змінна math) від соціально-економічного статусу (змінна ESCS) для різних регіонів і для різної статі учнів. Додати до графіків пряму лінійної регресії.

**Хід роботи**

1. Вчитання до змінної pisa файлу з даними pisa2015.csv:



1. Створення набору країн з яких буде робитися вибірка





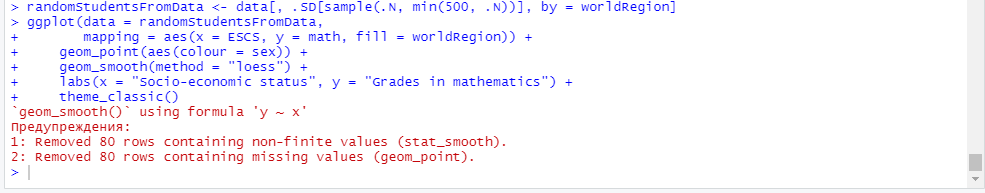
1. Занесення до змінної *data* записів з таблиці, де учні належать до країн зі змінної *country,* та мають дані по країнї, статі, балам з математики та соціально-економічного стану.



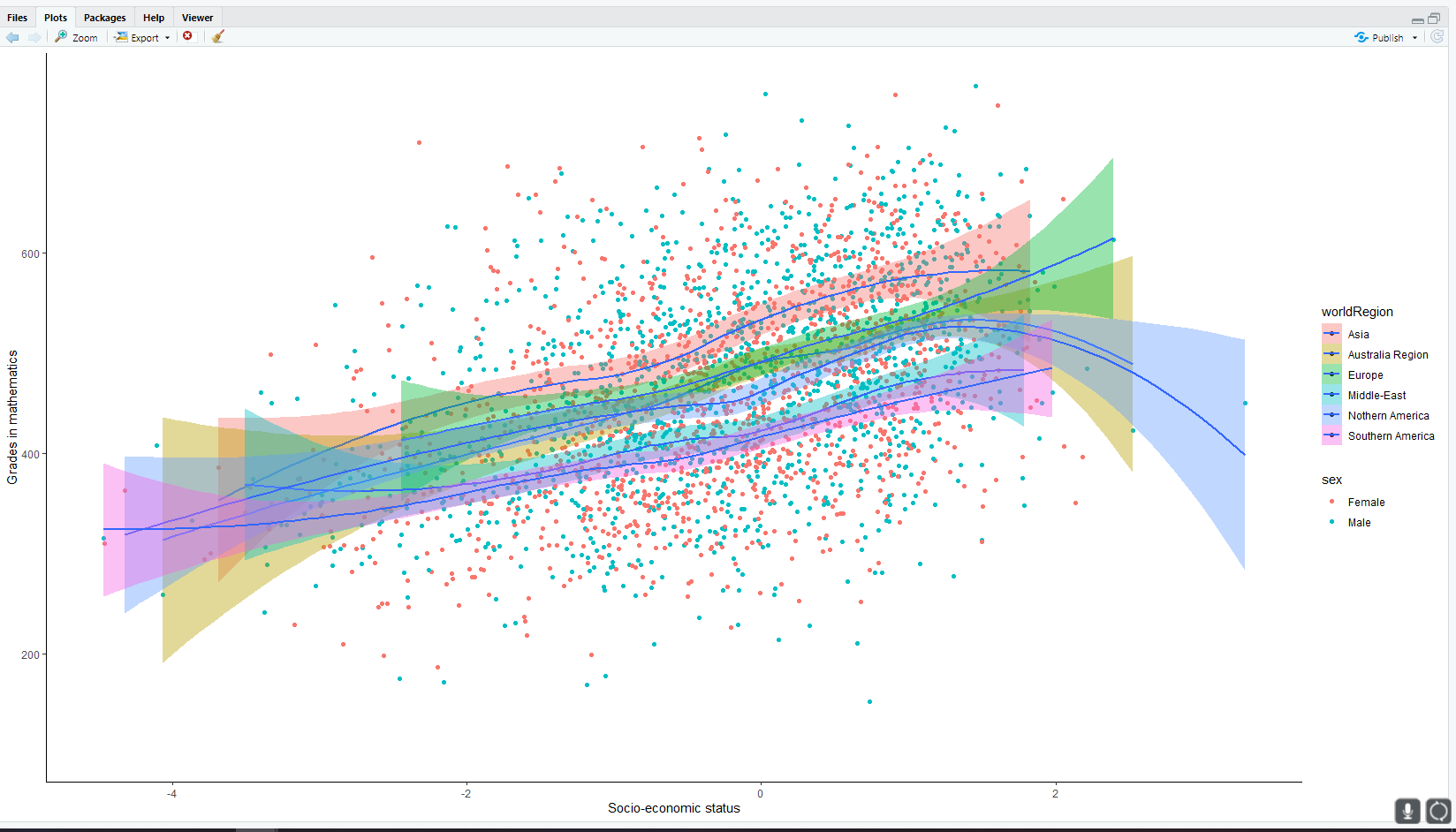
1. Об’єднання країн у регіони



1. Вибір по 500 випадкових студентів з кожного регіону. Створення точкової діаграми залежності балів з математики від соціально-економічного статусу для різних регіонів і для різної статі учнів.



1. Точкова діаграма з прямими регресії.



**Висновок:** виконавши дану лабораторну роботу, я отримав навички роботи з засобом створення графічного представлення даних ggplot2(), та було запрограмовано створення графіків залежності балів з математики від соціально-економічного статусу учнів різної статі, та побудовано прямі лінійної регресії. Зі створеної діаграми видно, що найкращі бали з математики, незалежно від соціально-економічного статусу, у середньому, мають учні з Азії, а найгірші – учні з Південної Америки та Ближнього Сходу. Також, можна зробити висновок, що бали з математики тим вище, чим кращий соціально-економічний статус учня.