**Міністерство Освіти І НАУКИ України**

**Національний університет "Львівська політехніка"**

Інститут **ІКНІ**

Кафедра **СШІ**

**ЗВІТ**

До лабораторної роботи №3

**З дисципліни:** «Візуалізація даних»

**На тему:** «Ознайомитися з аналізом даних та статичне виведення на мові R»

**Виконав:**

ст. гр. КН-307

Шиманський П.С.

**Прийняв:**

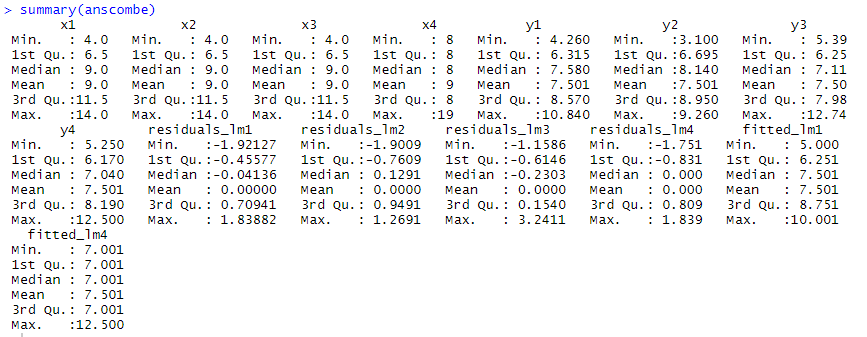
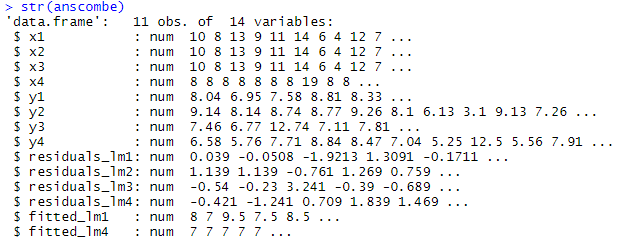
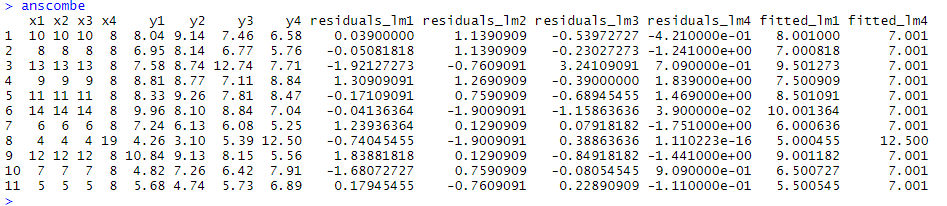
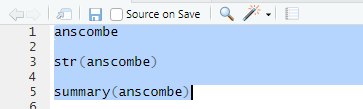
Шамуратов О.Ю.

Львів - 2019

**Мета:** Ознайомитися з аналізом даних та статичне виведення на мові R.

**Хід роботи:**

Досліджуємо набір даних anscombe.(Це вбудований у R датасет):



**Рис.1.** Дослідження набору даних anscombe

Коефіцієнт кореляції:



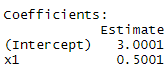
**Рис.2.** Коефіцієнт кореляції

Яке рівняння лінійної регресії для квартету Анскомбе?

ŷ=0.5x+3







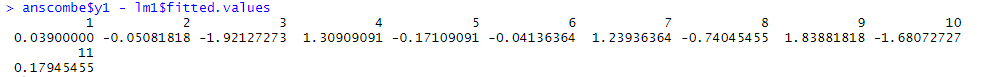
**Рис.2.** Відповідь

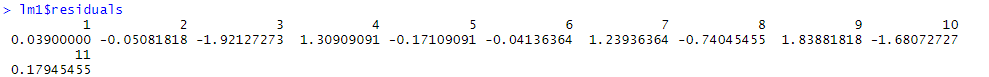
Знайдемо значення ^(fitted.values) згідно нашого рівняння лінійної регресії.



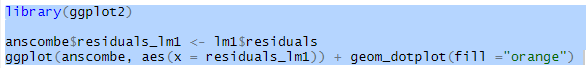


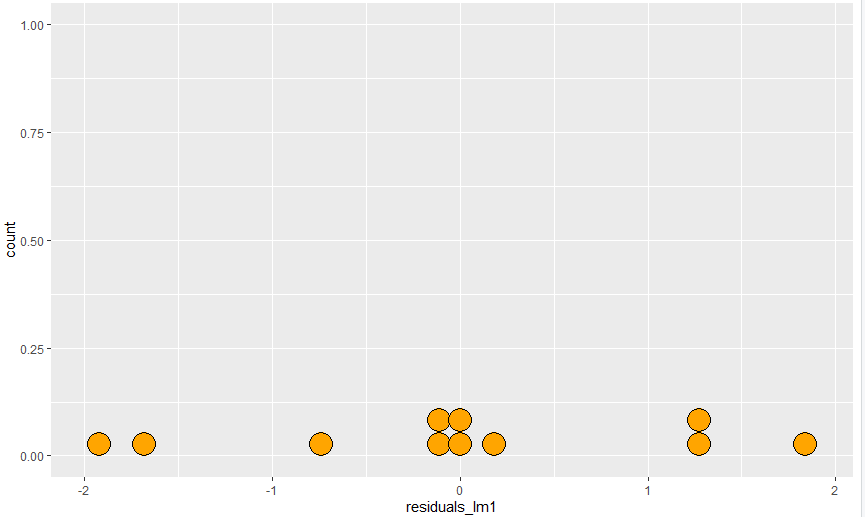
Знайдемо залишки. Це можна зробити віднявши fitted.values від реальних значень y1:



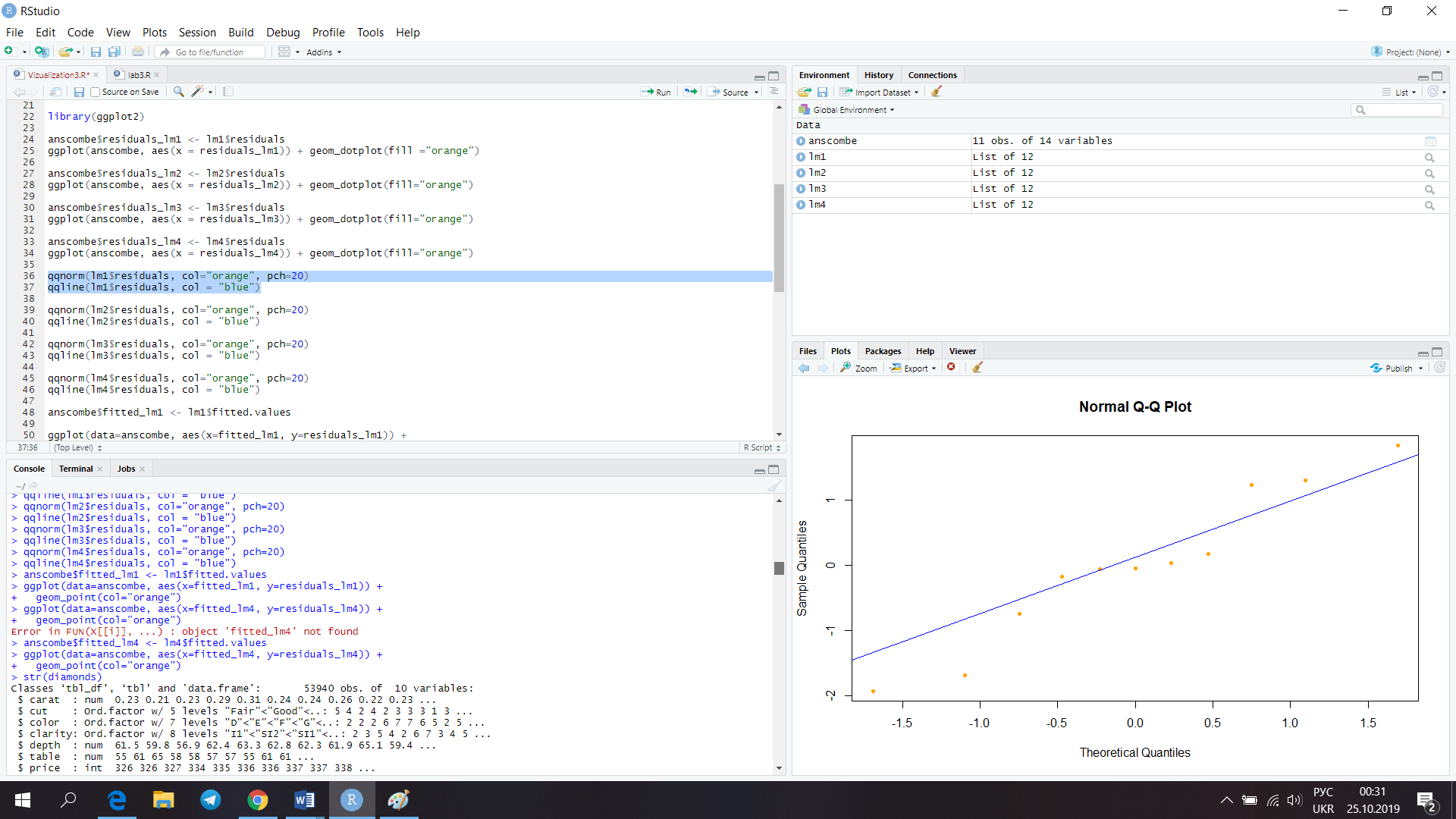


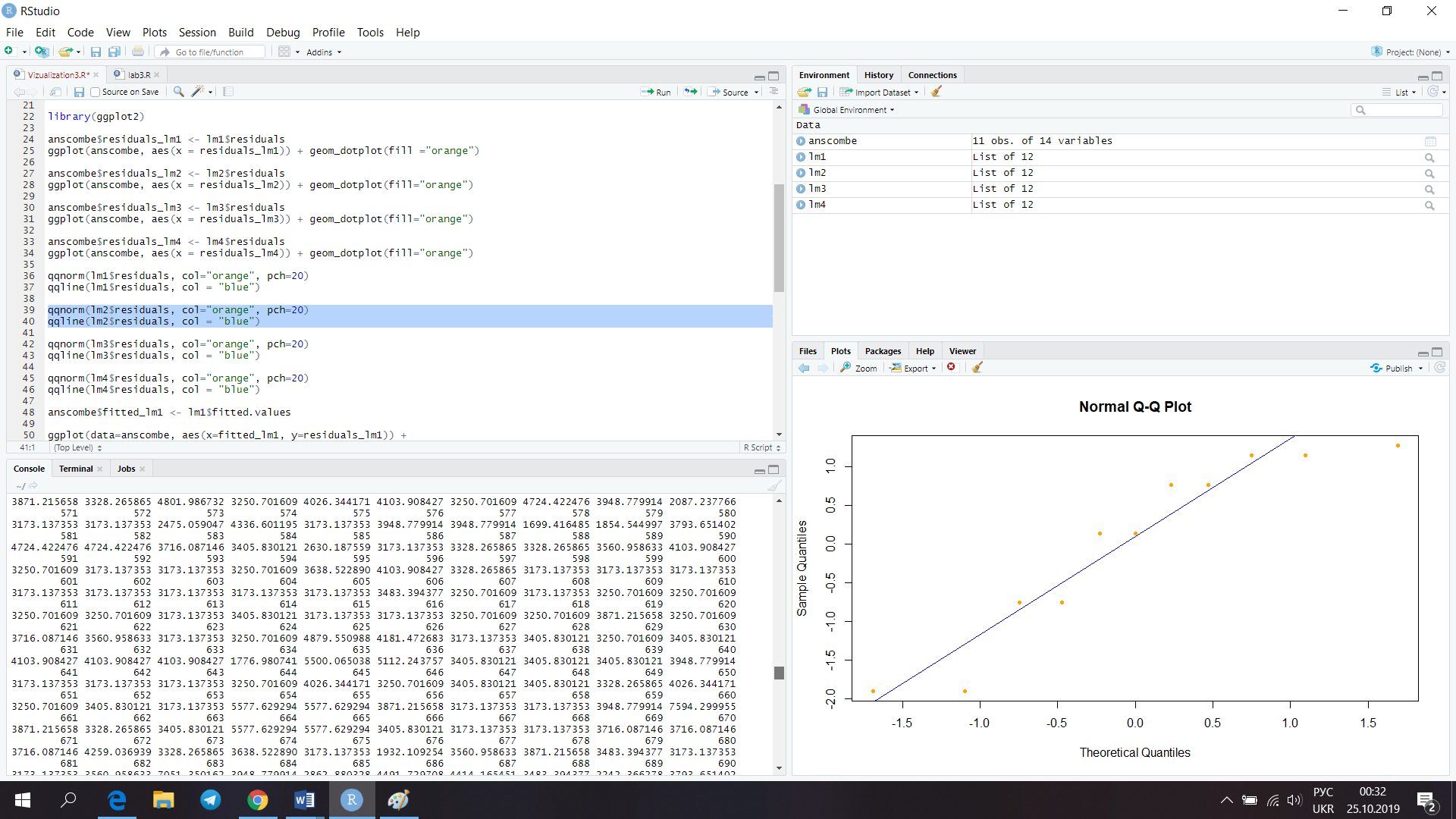
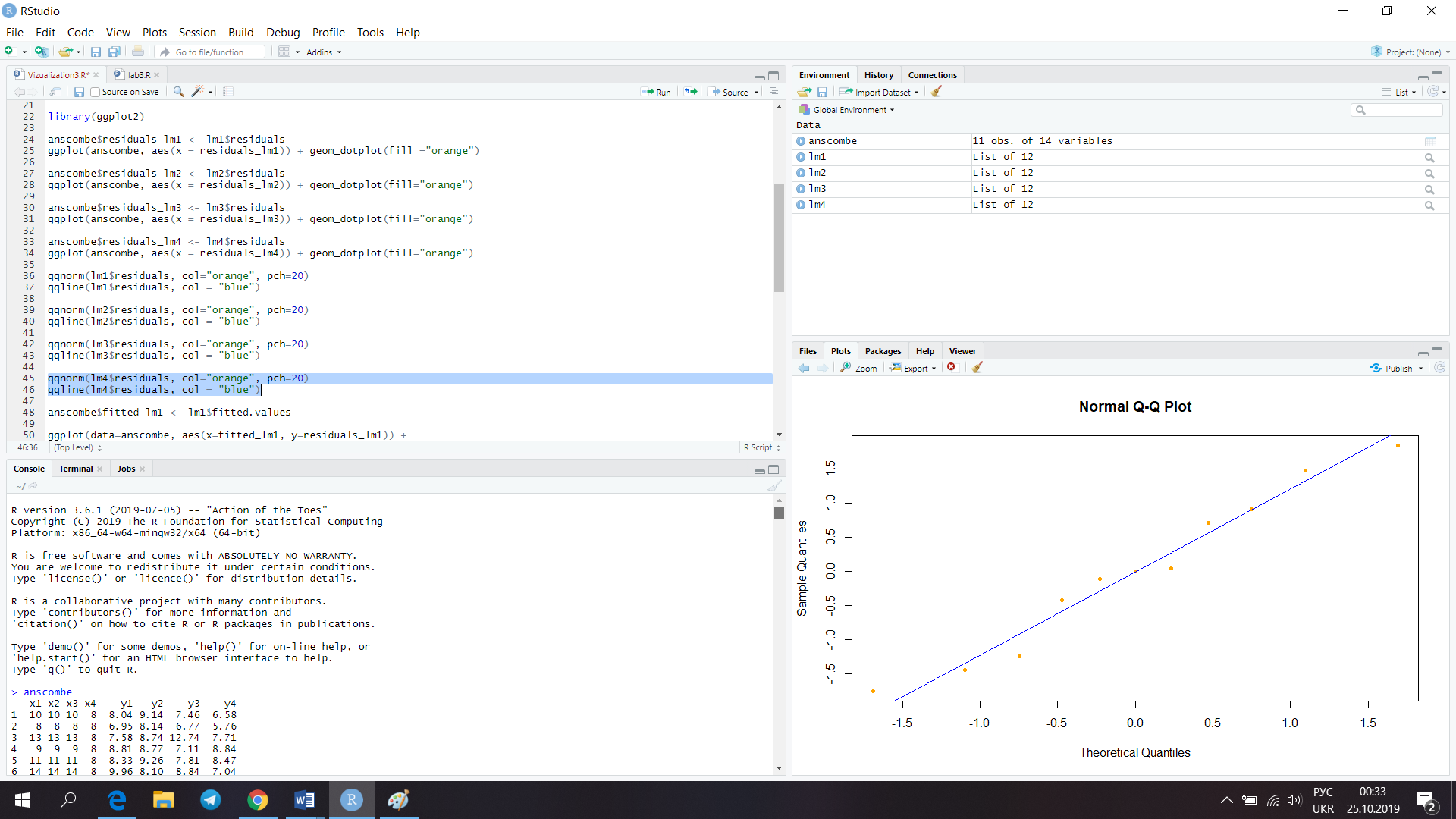
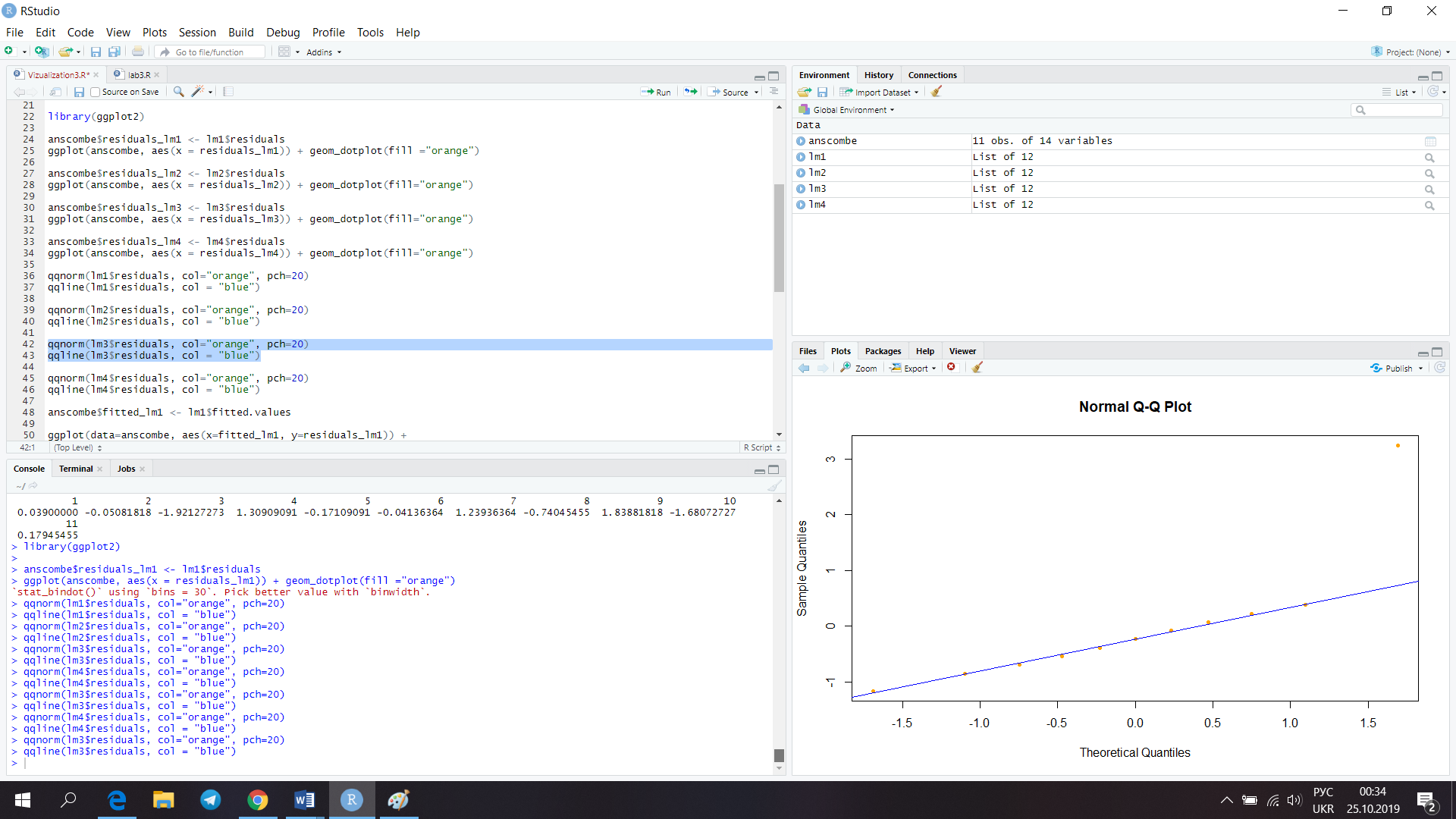
Найкраще оцінювати розподіл даних з допомогою гістограми, однак у нас всього одинадцять точок, тому можемо використати точковий графік.





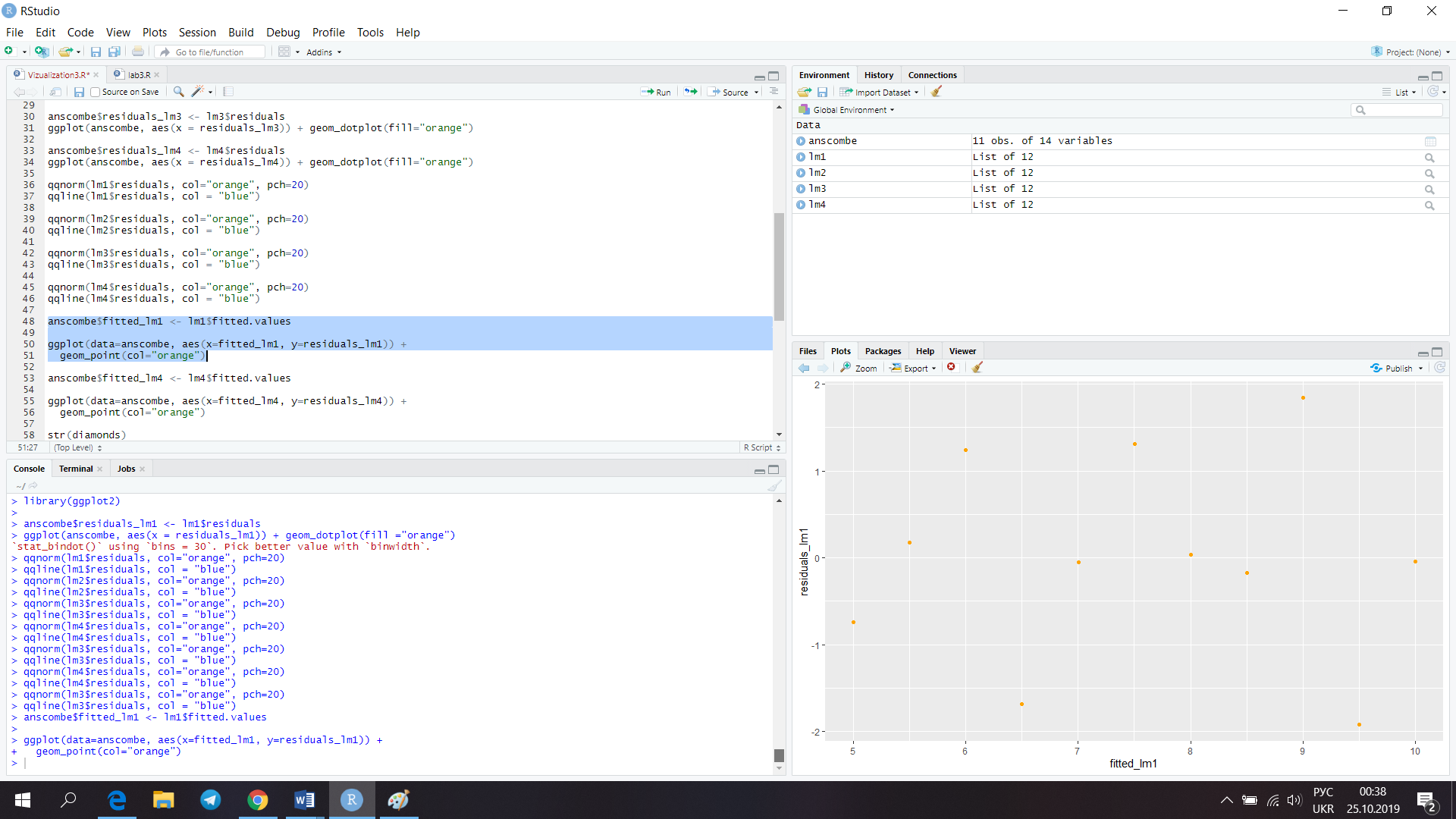
Оцінювати візуально розподіл даних для 11 точок досить тяжко, найбільше відповідають нормальному розподілу перший та третій набори даних.



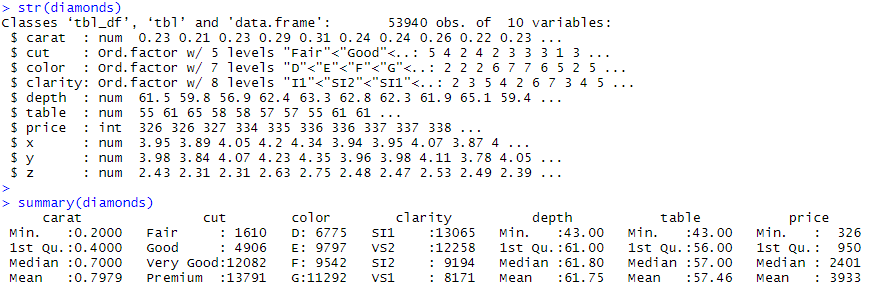
Оцінюємо варіативність залишків:

Для (x1, y1) умова сталості залишків виконується.



**Для датасету diamonds**

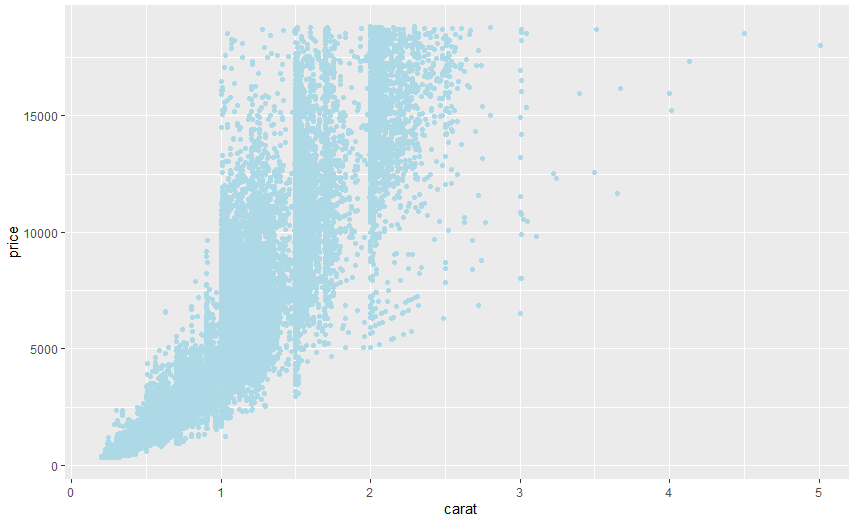
Подивимось на структуру набору даних:



В нас є 10 змінних та 53940 спостережень. Будемо досліджувати залежність між вагою (змінна carat) та ціною (змінна price).

**ggplot**(data=diamonds, **aes**(x=carat, y=price)) +

**geom\_point**(col="lightblue")



Знайдіть коефіцієнт кореляції між вагою (змінна carat) та ціною (змінна price) діамантів. Результат вкажіть в якості відповіді на питання 3.

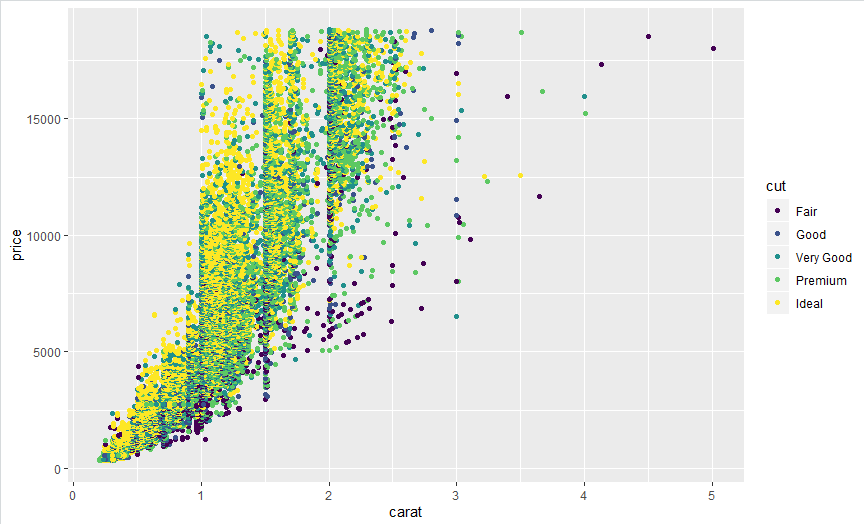
cor(diamonds$carat, diamonds$price)



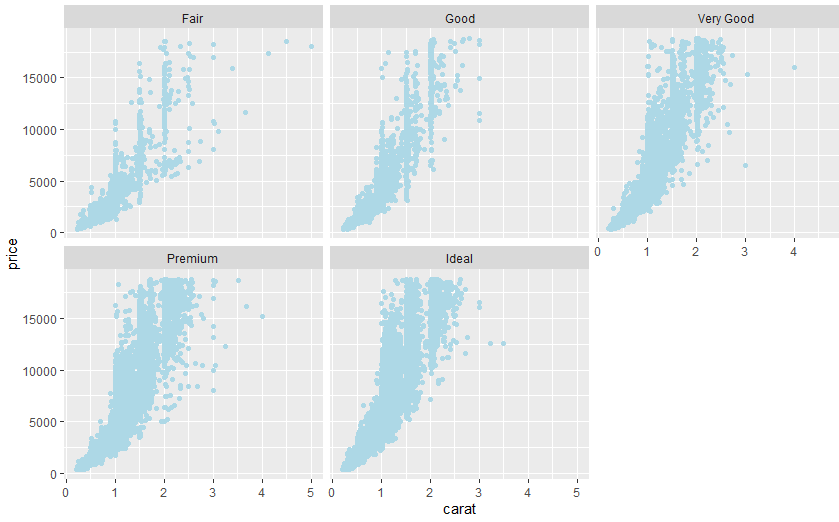
як розподілені вага та ціна в залежності від ступеня обробки діамантів:

**ggplot**(data=diamonds, **aes**(x=carat, y=price, col=cut)) +

**geom\_point**()

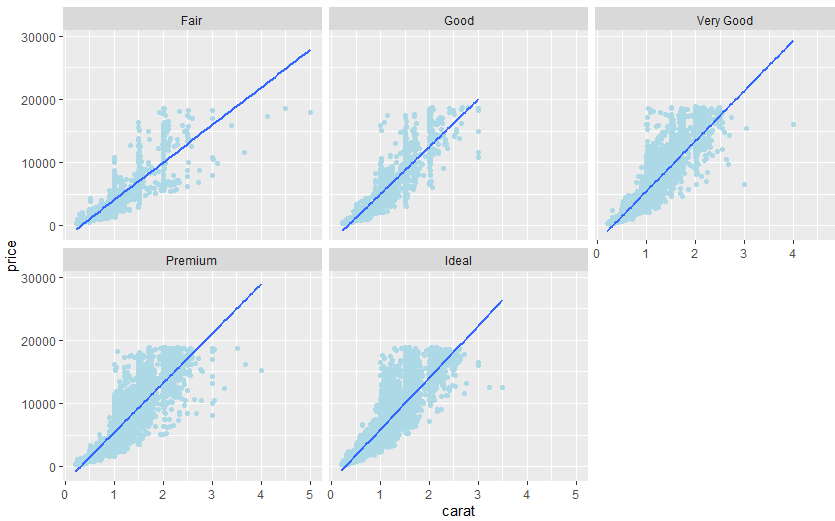


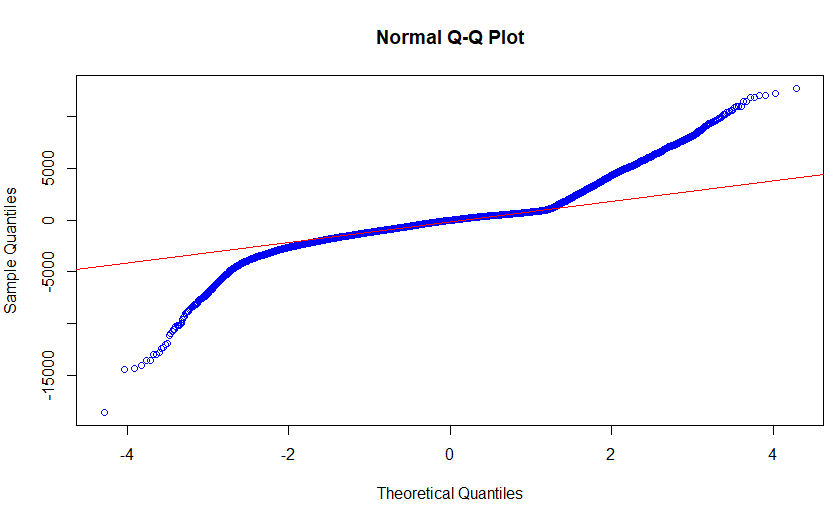
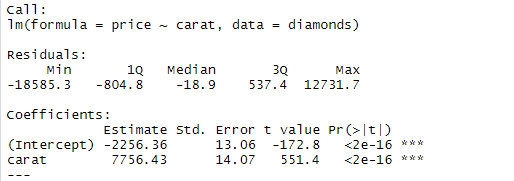
**ggplot**(data=diamonds, **aes**(x=carat, y=price)) + **geom\_point**(col="lightblue") + **facet\_wrap**(~cut)



Додамо до графіка ще лінію лінійної регресії:

**ggplot**(data=diamonds, **aes**(x=carat, y=price)) + **geom\_point**(col="lightblue") + **geom\_smooth**(method="lm", se=FALSE) + **facet\_wrap**(~cut)





Якою буде ціна ідеально (cut = Ideal) обробленого діаманта вагою 1 карат згідно моделі lin.diamond.ideal?

Відповідь: 5892

Якою буде ціна прийнятно (cut = Fair) обробленого діаманта вагою 1 карат згідно моделі lin.diamond.fair?

Відповідь: 4085

Висновок:

На лабораторній роботі я ознайомився з аналізом даних та статичним виведенням на мові R.