

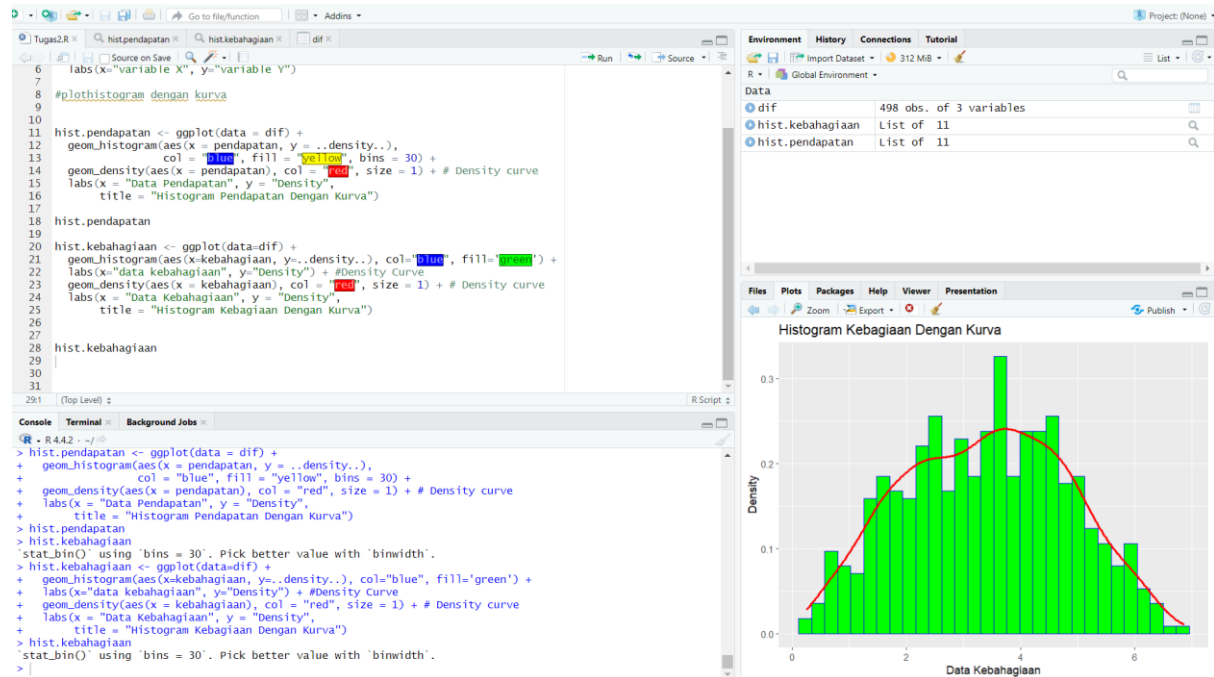
Nama : Dhafi Ade Al Faris

Tugas 2

ANALISIS DAN VISUALISASI DATA

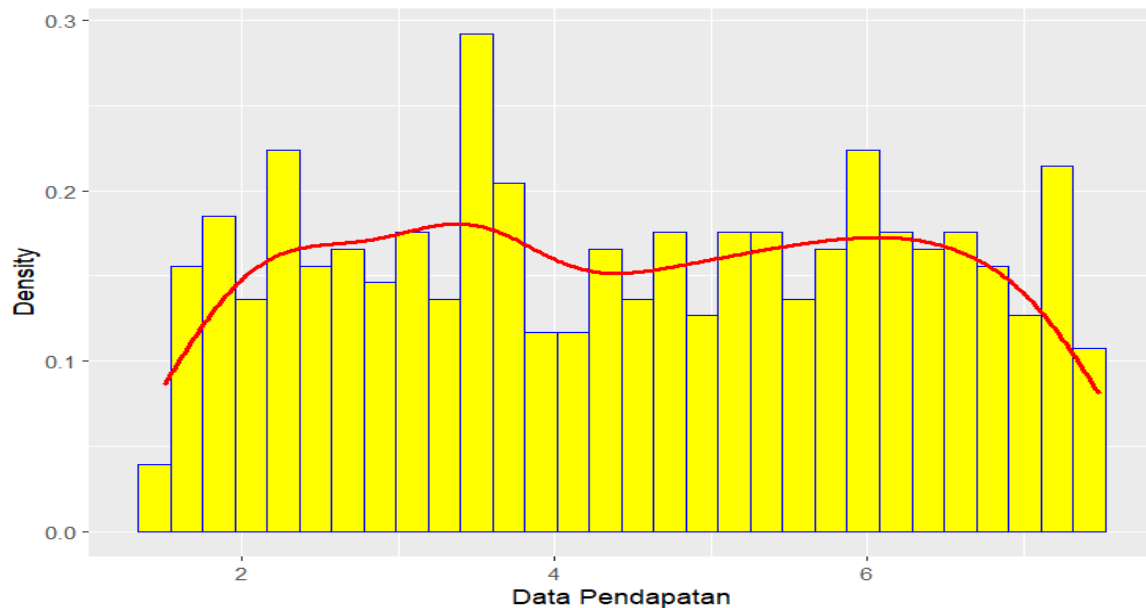
1.a. Gunakan (plot) histogram untuk cek apakah variabel terikat berdistribusi normal atau tidak. Jelaskan interpretasi histogram tersebut! (10 poin)

- >Kode Pengerjaan

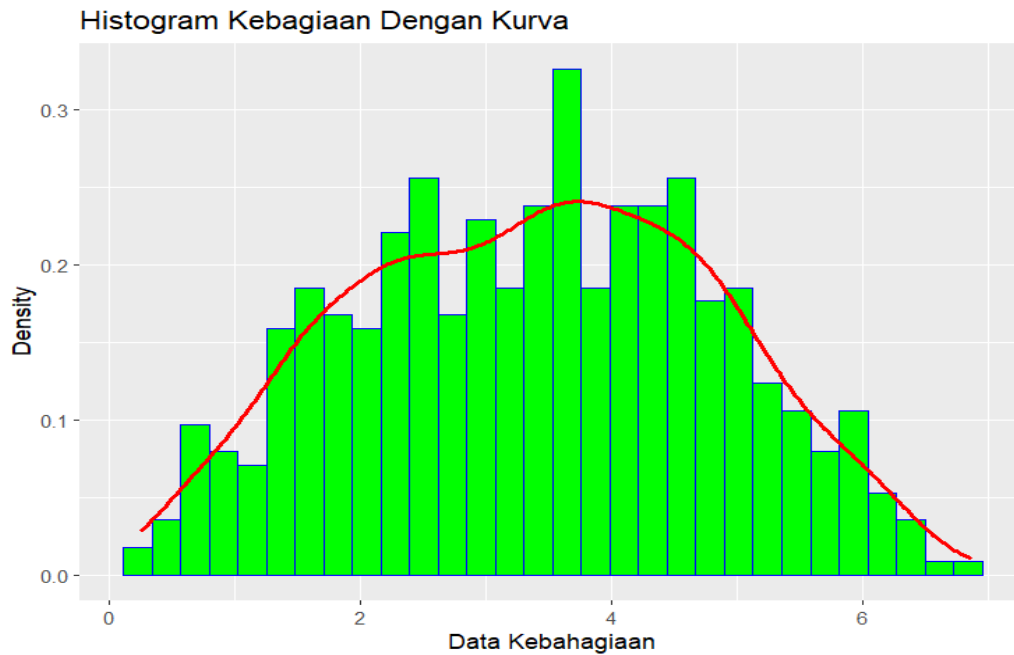


- Hasil Histogram Kolom Pendapatan (distribusi tidak normal)

Histogram Pendapatan Dengan Kurva



- Hasil Histogram Kolom Kebahagiaan (distribusi tidak normal)

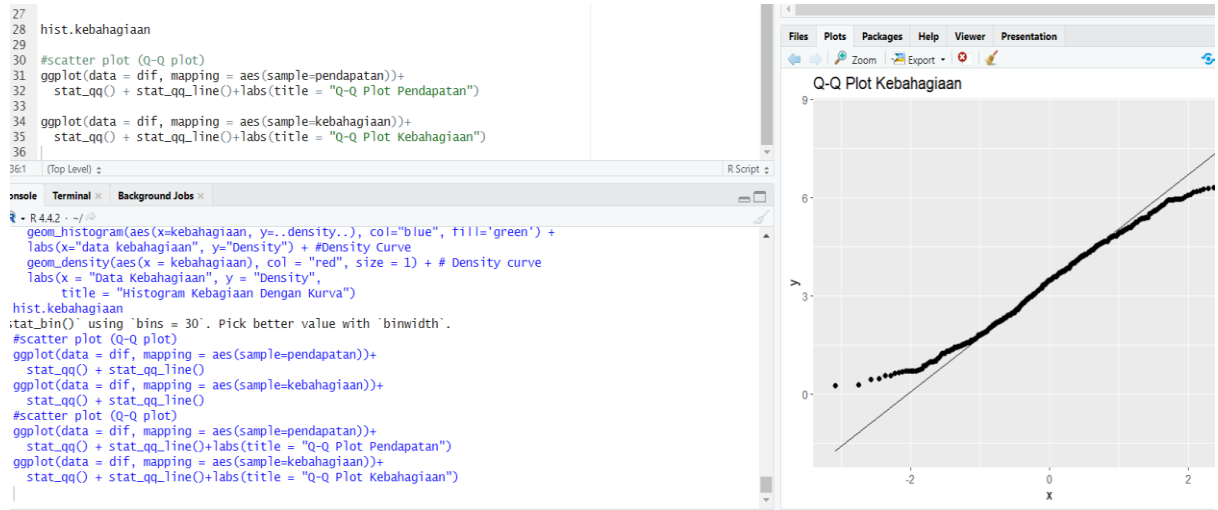


- Penjelasan Interpretasi ke dua histogram
 - Hasil histogram data pendapatan dan kebahagiaan di atas menunjukkan bahwa data berdistribusi tidak normal karena,
 - Bentuk Tidak Rata, bar tidak terlihat konsisten dan bentuk kurva tidak berbentuk lonceng sebagai mestinya data yang berdistribusi normal. Garis kurva pendapatan terlihat datar di bagian tengah dan garis kurva kebahagiaan terdapat beberapa lengkungan tidak normal.
 - Asimetris, sisi kiri dan kanan kurva kedua histogram terlihat sedikit tidak seimbang
 - Puncak tambahan, ada beberapa puncak selain puncak utama yang jelas
 - Ekor, Ujung distribusi tidak mengecil dengan halus seperti yang diharapkan

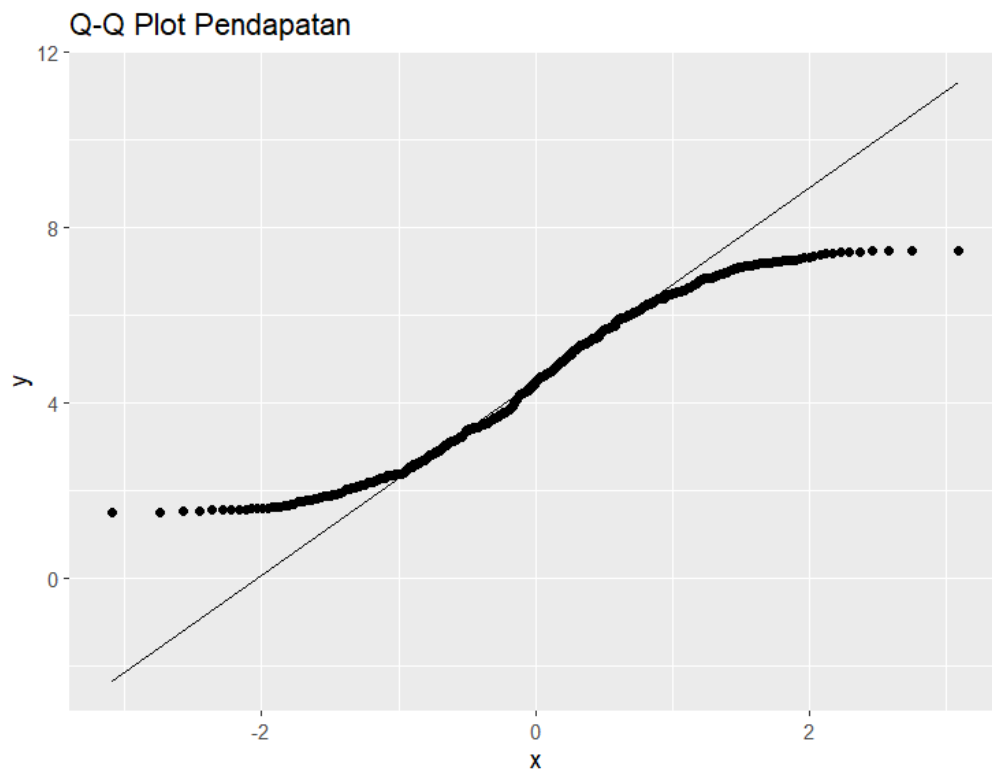
Perbedaan itu menunjukkan data sepenuhnya berdistribusi tidak normal.

b. Gunakan scatter plot untuk cek apakah variabel terikat dan bebas memiliki hubungan linear. Interpretasikan hasil grafik tersebut! (10 poin)

- kode pengerjaan menggunakan Q-Q Plot



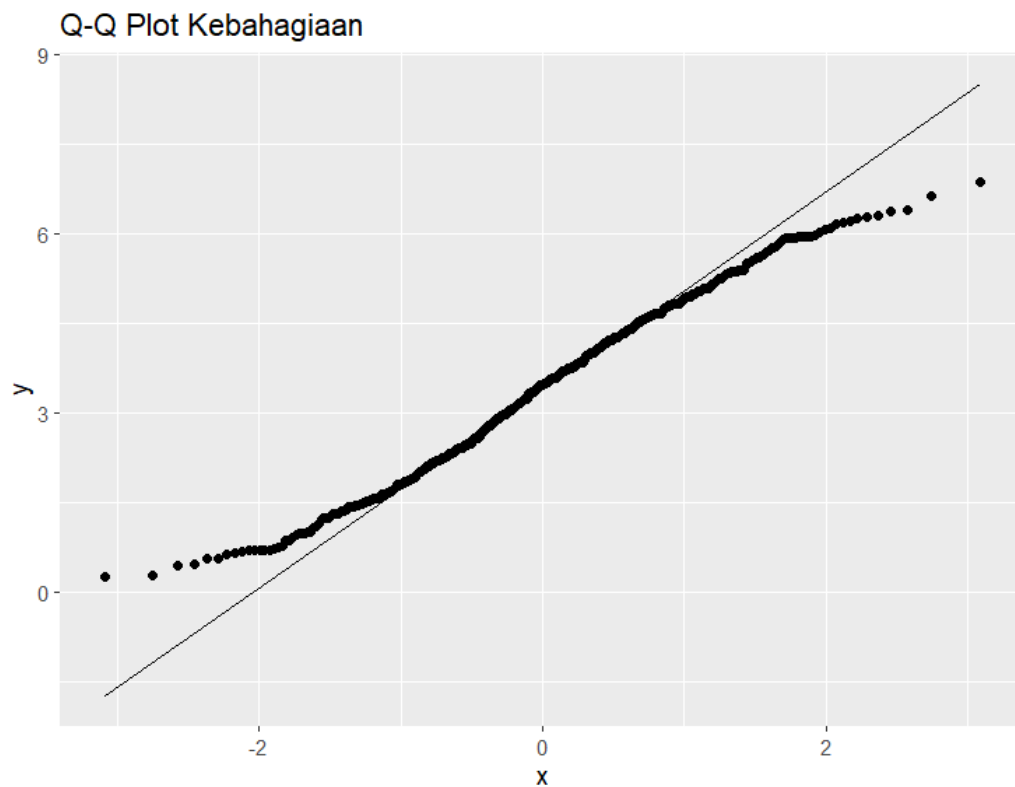
- Q-Q plot pendapatan



- Interpretasi

Dari hasil Q-Q Plot Pendapatan di atas menunjukkan kuantil-kuantil plot tersebut tidak sepenuhnya mengikuti garis linear terlihat pada kuantil sumbu x -1 dan seterusnya tidak lurus mengikuti garis linear dan sumbu x 1 dan seterusnya juga tidak lurus mengikuti garis linear, oleh sebab itu dapat dikatakan berdasarkan Q-Q Plot di atas distribusinya tidak normal

- Q-Q plot kebahagiaan



- Interpretasi

Dari hasil Q-Q Plot Kebahagiaan di atas juga sama yaitu menunjukkan kuantil-kuantil plot tersebut tidak sepenuhnya mengikuti garis linear terlihat pada kuantil disekitar sumbu x -1 dan seterusnya tidak lurus mengikuti garis linear dan disekitar sumbu x 1 dan seterusnya juga tidak lurus mengikuti garis linear, oleh sebab itu dapat dikatakan berdasarkan Q-Q Plot di atas distribusinya tidak normal.

c. Lakukan regresi linear sederhana dan jelaskan hasil regresi tersebut! (20 poin)

- Kode regrenasi sederhana

```
36
37
38 #Regresi Linear Sederhana
39 simpleregresi <- lm(kebahagiaan ~ pendapatan, dif)
40 summary(simpleregresi)
41 |
```

Variable y (dependant) = kebahagiaan

Variable x(independent) = pendapatan

- Output

```
> summary(simpleregresi)

Call:
lm(formula = kebahagiaan ~ pendapatan, data = dif)

Residuals:
    Min       1Q   Median       3Q      Max
-2.02292 -0.47796  0.03729  0.44564  2.38002

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.20906    0.08962   2.333  0.0201 *
pendapatan   0.71239    0.01870  38.089 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.724 on 496 degrees of freedom
Multiple R-squared:  0.7452,    Adjusted R-squared:  0.7447
F-statistic: 1451 on 1 and 496 DF,  p-value: < 2.2e-16
```

- Penjelasan

Dalam kasus ini regresi yang dibuat berdasarkan *pendapatan* (x) mempengaruhi *kebahagiaan* (y), di hasil output didapatkan nilai koefisien regresi pendapatan 0.71239 dan bernilai positif yang artinya jika pendapatan (var x) meningkat maka kebahagiaan (var y) juga ikut meningkat dan sebaliknya, jika pendapatan menurun maka kebahagiaan juga ikut menurun, selanjutnya juga terlihat p-value di didapatkan bernilai $<2e-16$ *** yang jauh dari nilai intercepny 5% berarti bisa disimpulkan pendapatan (var x) berpengaruh sangat signifikan terhadap kebahagiaan (var y) dan terdapat bintang 3 (***) juga bermaksud berpengaruh sangat signifikan. Kemudian didapatkan koefisien determinasi yaitu Multiple R-squared (R^2) bernilai 0.7452 atau 74.52% dimaksudkan sumbangan pengaruh pendapatan (var x) terhadap kebahagiaan (var y) sebesar 74.52%.

Kesimpulan

- Terdapat pengaruh yang sangat kuat antara pendapatan dan kebahagiaan
- Nilai p-value $<2e-16$ membuktikan pendapatan berpengaruh sangat signifikan terhadap kebahagiaan
- didapatkan koefisien determinasi (R^2) bernilai 74.52% berarti sumbangan pengaruh pendapatan terhadap kebahagiaan sebesar 74.52% dan sisanya 25,48% pengaruh berasal dari luar data ini.

d. Gunakan perintah plot(nama variabel regresi kalian) untuk cek homoskedastisitas. Interpretasikan hasil plot (20 poin).

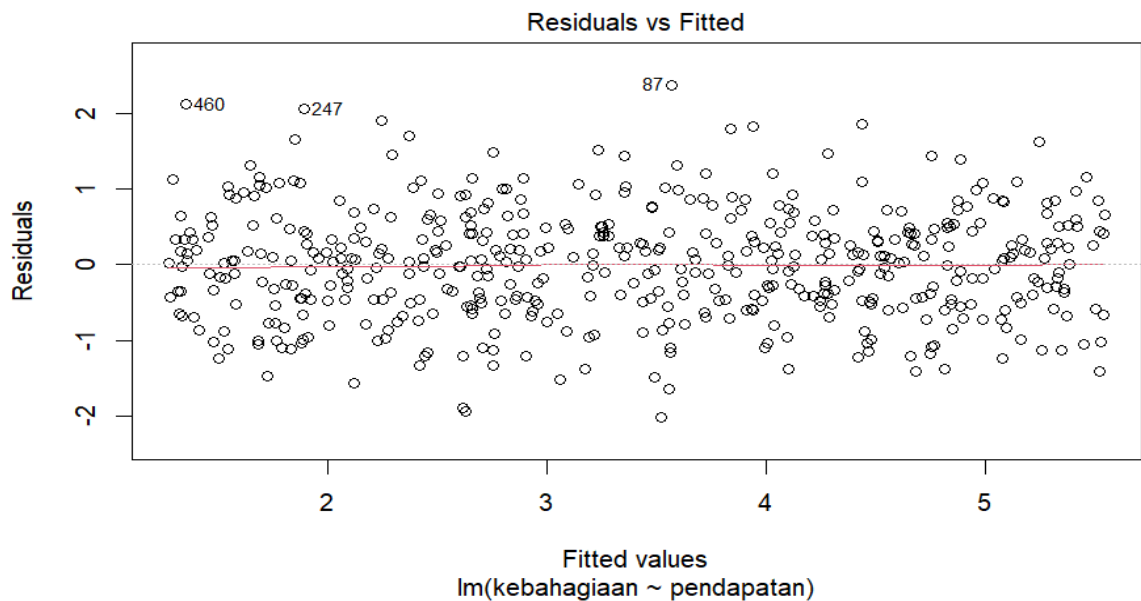
- Kode

#Regresi Linear Sederhana

```
simpleregresi <- lm(kebahagiaan ~ pendapatan, dif)
summary(simpleregresi)

plot(simpleregresi)
```

- Output plot() di R secara default ada 4 yang muncul, karena untuk mengecek homoskedastisitas menggunakan Residuals vs Fitted



- Interpretasi,

Berdasarkan plot di atas menunjukkan titik-titik residual tersebar acak di sekitar garis horizontal pada 0 tanpa pola yang jelas (menggumpal atau menyebar merata), maka model memenuhi asumsi homoskedastisitas (tidak ada masalah). Yang tidak kita inginkan adalah Heteroskedastisitas jika residual titik-titik berperntuk pola seperti corong atau fan.

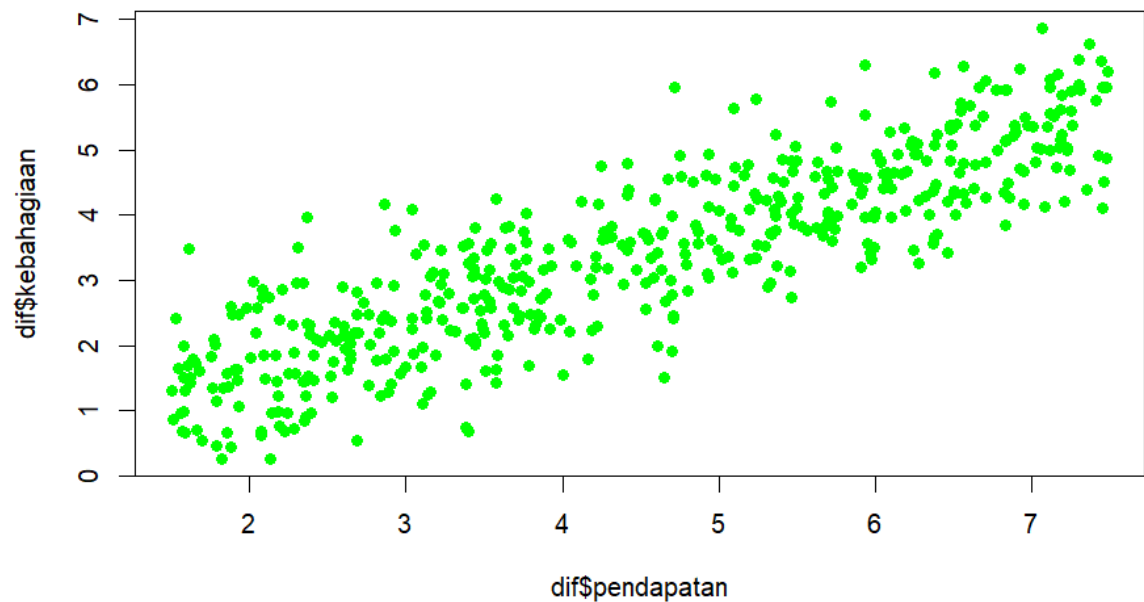
e. Visualisasikan hasil regresi menggunakan grafik dengan langkah-langkah berikut:

1. Plot data points pada grafik (10 poin)

- Kode

```
# 1. Plot data points pada grafik (scatter plot)
plot(dif$pendapatan, dif$kebahagiaan,
     pch = 19, col = "green")
```

- Output



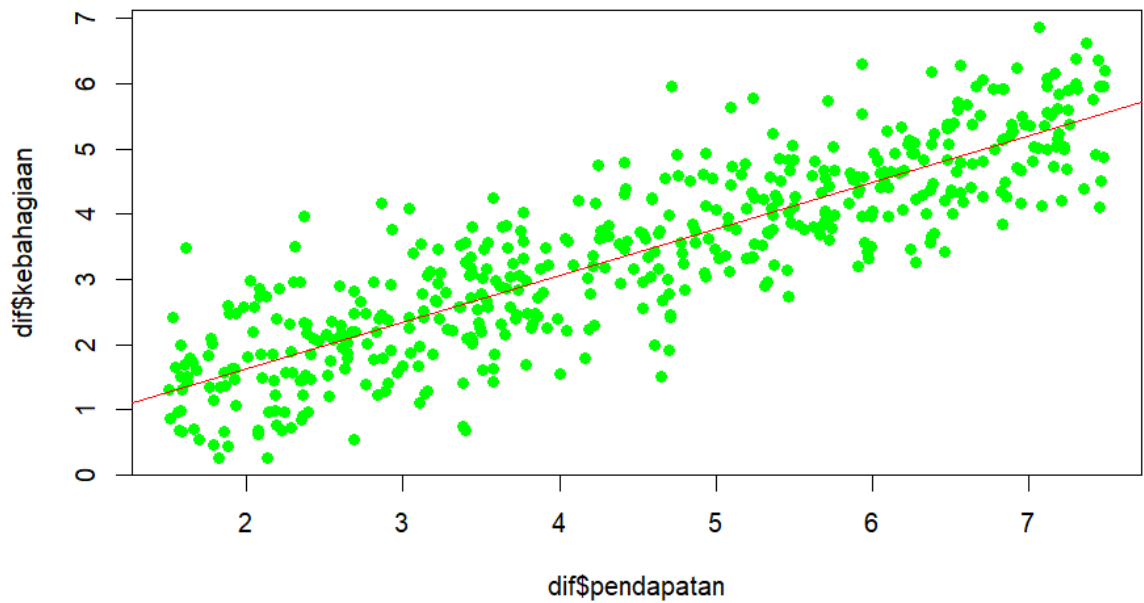
2. Tambahkan garis regresi linear pada data yang sudah diplot (10 poin)

- Kode

```
# 1. Plot data points pada grafik (scatter plot)
plot(dif$pendapatan, dif$kebahagiaan,
     pch = 19, col = "green")

# 2. Tambahkan garis regresi linear
abline(simpleregresi, col = "red")
|
```

- Output



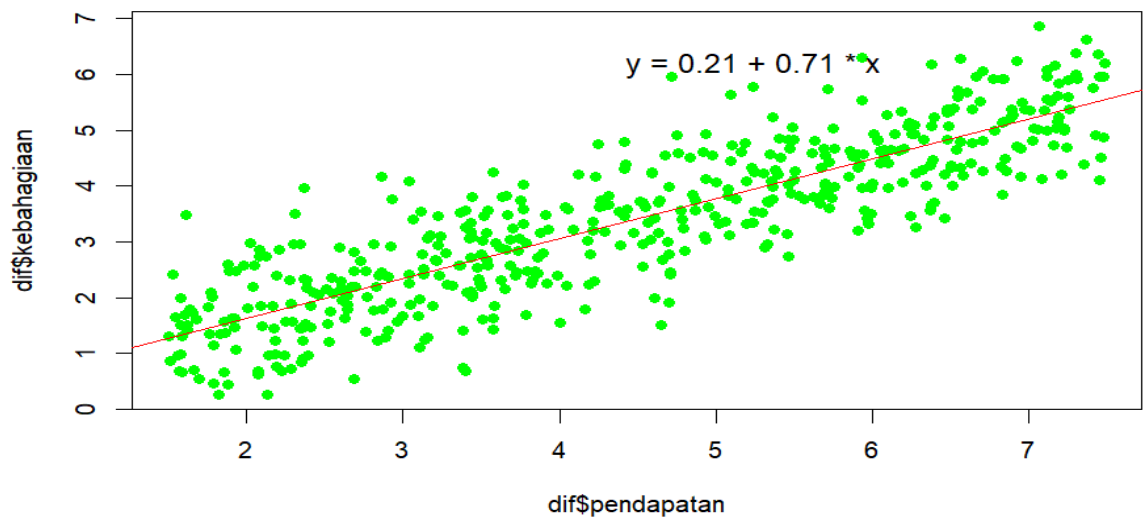
3. Tambah persamaan untuk garis linear (10 poin)

- Kode

```
# 3. Tambahkan persamaan untuk garis linear
equation <- paste("y = ", round(coef(simpleregresi)[1], 2),
                  " + ", round(coef(simpleregresi)[2], 2), " * x", sep = "")

# Dynamically add the equation to the plot at a reasonable position
text(x = max(dif$pendapatan) * 0.7, y = max(dif$kebahagiaan) * 0.9,
     labels = equation, col = "black", cex = 1.2)
```

- Output



4. Beri judul dan label yang sesuai untuk sumbu x dan y (10 poin)

- Kode

```
# 1. Plot data points pada grafik (scatter plot)
plot(dif$pendapatan, dif$kebahagiaan,
     pch = 19, col = "green",
     main = "Hubungan Pendapatan dengan Kebahagiaan", #5 judul dan sumbu y dan x
     xlab = "Pendapatan", ylab = "Kebahagiaan")
```

- Output

