* Fungsi untuk mengimport data

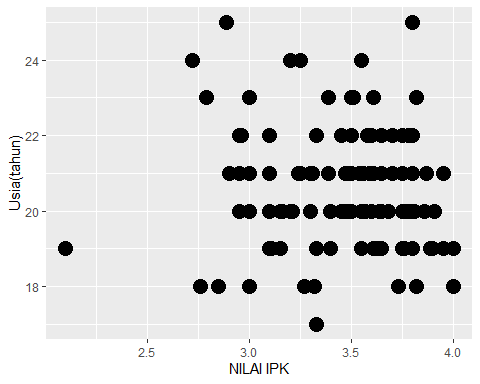
library(readxl)  
data <- read\_excel("C:/Users/Lenovo/Downloads/data.xlsx")  
data$IPK <- as.numeric(data$IPK)

* Fungsi untuk membuat scatterplot hubungan IPK dengan usia

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.1.2

ggplot(data, aes(x=IPK, y=Usia)) + geom\_point(size=5)+  
 labs(x="NILAI IPK", y="Usia(tahun)")



* Fungsi untuk mendeskripsikan variable baru antar data kategori

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.1.2

##   
## Attaching package: 'dplyr'

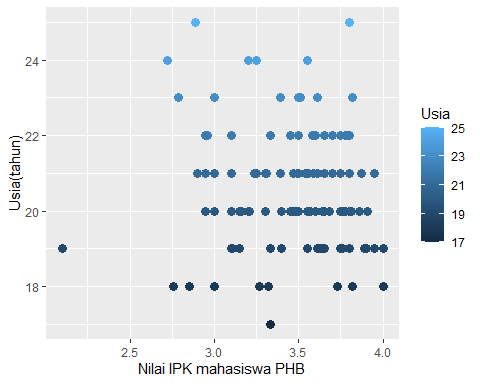
## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data = mutate(data, usia = case\_when(  
 data$`Usia` <= 20 ~ "usia muda",  
 data$`Usia` <= 25 ~ "usia pertengahan",  
 data$`Usia` >= 25 ~ "usia tua"  
),  
nilai = case\_when(  
 data$IPK <= 3 ~ "grade C ",  
 data$IPK <= 3.5 ~ "grade B",  
 data$IPK >= 3.5 ~ "grade A"  
)  
)

* Fungsi untuk membuat grafik plot(scatterplot)

ggplot(data, aes(x=IPK, y=`Usia`, color = Usia)) +  
 geom\_point(size=3)+  
 labs(x="Nilai IPK mahasiswa PHB", y="Usia(tahun)")



* Fungsi untuk membuat data yang baru dan menghapus data apabila data outlier

databaru = filter(data, `Usia` < 25)  
databaru

## # A tibble: 125 x 13  
## Nama `Jenis kelamin` Usia Prodi `Kesesuaian prodi` Pembiayaan `Kota asal`  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Fabil 0 20 2 1 0 0  
## 2 Amel 1 19 2 0 1 0  
## 3 Widia 1 21 3 1 0 1  
## 4 Sinta 1 20 7 0 0 1  
## 5 Ami 1 20 2 1 0 0  
## 6 Taufik 0 19 1 0 0 1  
## 7 Ziky 0 22 1 1 0 0  
## 8 Silvia 1 20 1 1 2 1  
## 9 Gilang 0 20 4 1 2 1  
## 10 Putri 0 22 3 0 0 0  
## # ... with 115 more rows, and 6 more variables: Kos <dbl>, Jam belajar <dbl>,  
## # Kendaraan <dbl>, IPK <dbl>, usia <chr>, nilai <chr>

* Fungsi untuk mengetahui nilai korelasi antar variable

cor(databaru$IPK, databaru$`Usia`)

## [1] -0.07827798

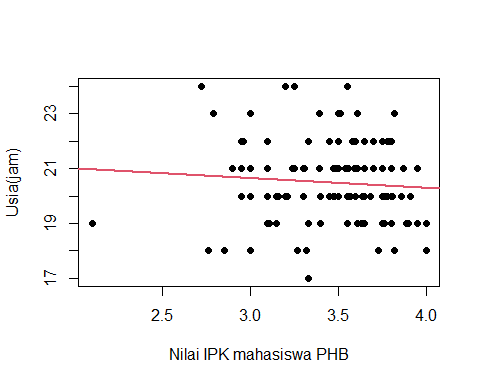
* Fungsi untuk mencari persamaan regresi

lm(`Usia`~IPK, databaru)

##   
## Call:  
## lm(formula = Usia ~ IPK, data = databaru)  
##   
## Coefficients:  
## (Intercept) IPK   
## 21.7257 -0.3529

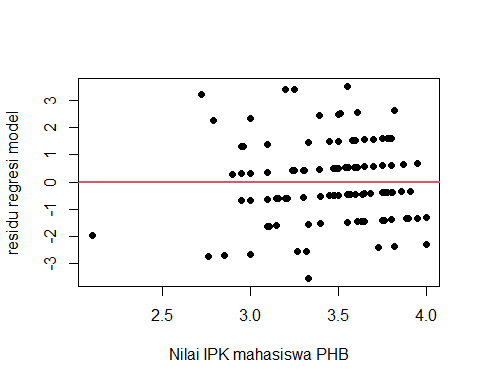
* Fungsi untuk membuat scatterplot dengan garis regresi dan korelasi

plot(databaru$IPK, databaru$`Usia`, pch=16,  
 xlab = "Nilai IPK mahasiswa PHB",  
 ylab = "Usia(jam)")  
abline(lm(databaru$`Usia` ~ databaru$IPK), col=2, lwd=2)  
text(3.4, 40, paste("y = 21.8135-0.3657x"))



* Fungsi untuk membuat plot residu(tidak ada outlier) dengan garis linear

model = lm(`Usia` ~ IPK, databaru)  
residu = resid(model)  
plot(databaru$IPK, residu, pch=16,  
 xlab = "Nilai IPK mahasiswa PHB",  
 ylab = "residu regresi model")  
abline(h=0, col=2, lwd=2)



* Fungsi untuk melihat nilai plot residu

summary(model)

##   
## Call:  
## lm(formula = Usia ~ IPK, data = databaru)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3.5506 -0.6847 -0.3848 0.6399 3.5270   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 21.7257 1.4090 15.419 <2e-16 \*\*\*  
## IPK -0.3529 0.4052 -0.871 0.386   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.458 on 123 degrees of freedom  
## Multiple R-squared: 0.006127, Adjusted R-squared: -0.001953   
## F-statistic: 0.7583 on 1 and 123 DF, p-value: 0.3856

* Fungsi untuk memprediksi data(co: hubungan usia dengan IPK)

nilaibaru = data.frame(IPK=3.85)  
predict(model, nilaibaru)

## 1   
## 20.36711