

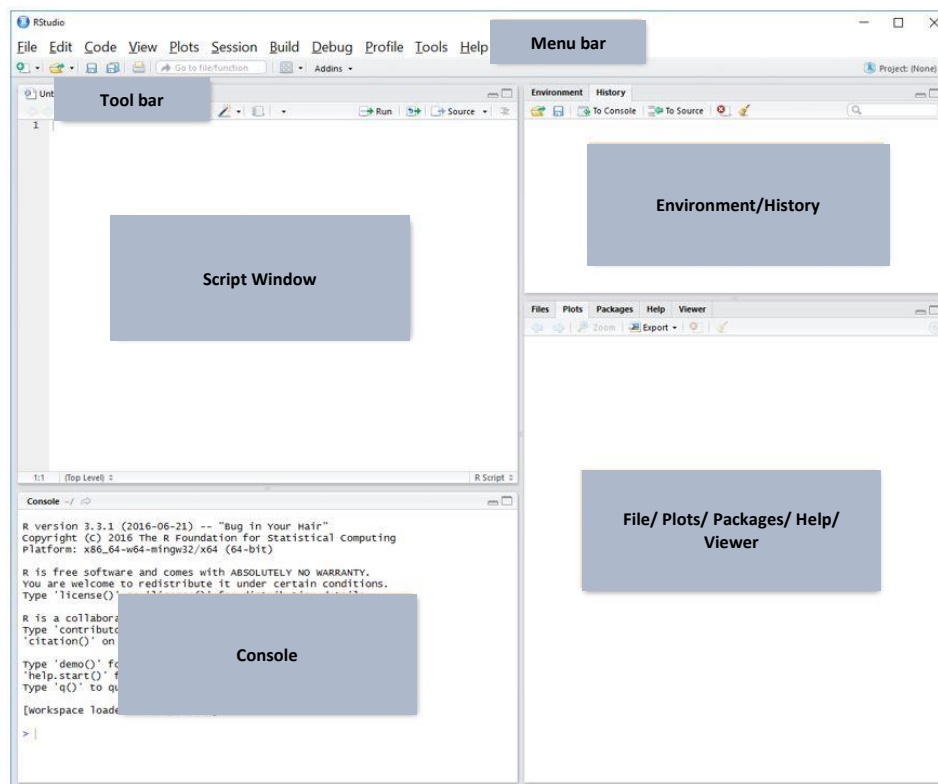
PENGENALAN RSTUDIO SEBAGAI APLIKASI PENGOLAH DATA

R dan RStudio Sebagai Aplikasi Pengolah Data

R adalah sebuah bahasa pemrograman berbasis statistika yang dikembangkan oleh Ross Ihaka dan Robert Gentleman pada tahun 1993. Saat ini, R dikenal sebagai salah satu aplikasi analisis data yang banyak digunakan oleh *data analyst* dalam melakukan persiapan data, pengolahan data, visualisasi data, dan pembuatan dashboard. R memiliki beberapa kemampuan, di antaranya:

- Gratis dan *Open Source*
- Tersedia banyak *package*
- Kemudahan dalam melakukan transformasi dan pemrosesan data
- Mampu menghasilkan grafik yang menarik

RStudio adalah *integrated development environment* (IDE) untuk R. Secara singkat, RStudio memudahkan kita untuk berkomunikasi dengan R sehingga kita bisa menjalankan fungsi-fungsi statistika dan *data science*.

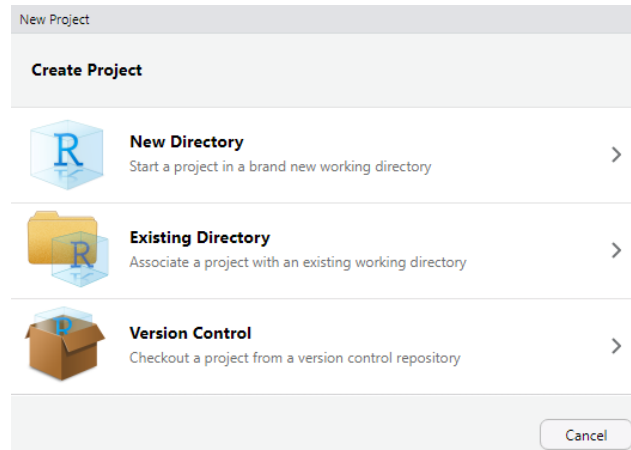


Membuat *Project* Baru di Dalam RStudio

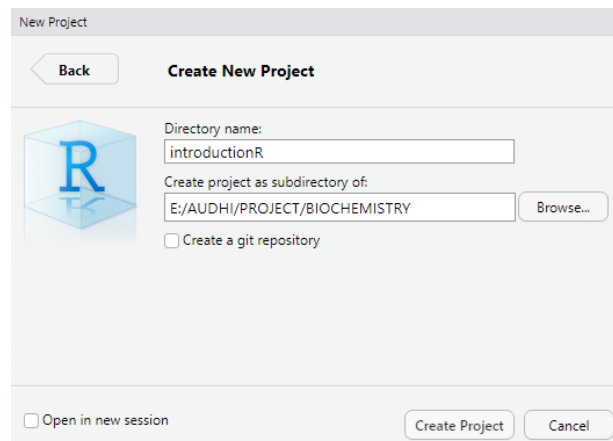
Sebelum melakukan analisis data, sebaiknya Anda mempersiapkan lingkungan pengembangan, berupa folder dimana Anda akan menyimpan data yang hendak dianalisis dan keluaran hasil analisis menggunakan RStudio. Langkah-langkah membuat *project* baru di RStudio, sebagai berikut:

- Pilih menu **File**, kemudian pilih **New Project...**
- Pilih jenis *project* yang akan dibuat

- a. *New Directory*: Apabila Anda belum mempunyai folder yang akan dijadikan sebagai *working directory* untuk *project* tersebut
- b. *Existing Directory*: Apabila folder Anda sudah ada sebelumnya
- c. *Version Control*: Apabila Anda akan membuat *project* berdasarkan repository yang sudah ada, misalnya di *GitHub* atau *SVN*.



3. Setelah Anda memilih *New Directory*, muncul pilihan *Project Type*. Pilih *New Project*
4. Tentukan nama *project* Anda. Kemudian, pilih menu *Browse* untuk menentukan dimana *project* Anda akan dibuat.



5. Klik **Create Project**. *Project* Anda akan muncul di RStudio

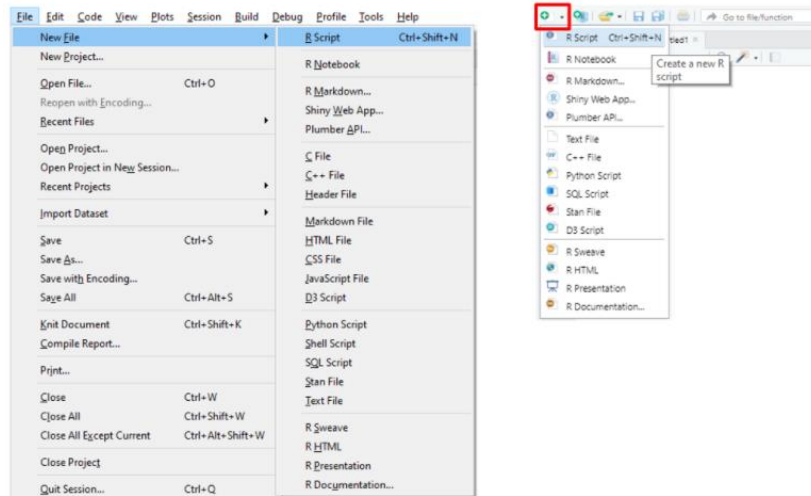
Mempersiapkan Data di Dalam Folder *Project*

Setelah Anda membuat *project* untuk analisis data, Anda perlu menempatkan data di dalam *working directory project* yang telah dibuat. Dalam kasus ini, *project* ditempatkan di dalam folder *E:/AUDHI/PROJECT/BIOCHEMISTRY/introductionR*. Hal yang perlu dilakukan cukup sederhana, yaitu menempatkan file Excel di dalam *folder project introductionR*.

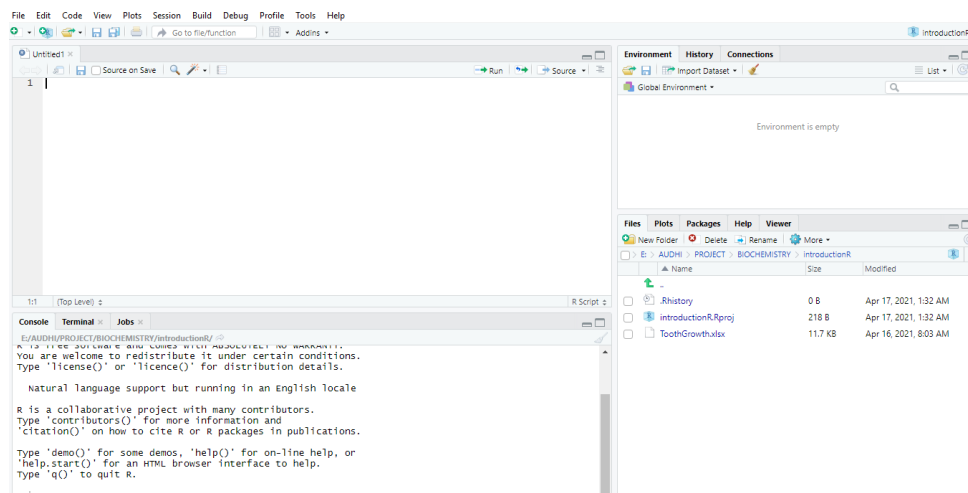
Membuat Script Baru

Setelah Anda telah membuat *project*, maka Anda dapat memulai membuat *script* yang digunakan untuk menuliskan perintah analisis data. Berikut langkah-langkah membuat script baru di RStudio

1. Pilih menu **File**, kemudian pilih **New File**
2. Kemudian, pilih **R Script** untuk membuat *script* baru



3. *Script window* akan muncul. *Script window* akan digunakan untuk menuliskan perintah atau script dalam analisis data



Membuka Data Menggunakan RStudio

Bahasa pemrograman R dapat membaca berbagai macam format data, seperti teks (txt), *Comma Separated Values* (CSV), Excel, SQL, SPSS, dan lain-lain. Untuk membuka data dengan format Excel, ikuti langkah-langkah berikut:

1. Buka file Excel Anda, kemudian **select** dan **copy** (CTRL + C)
2. Ketikkan perintah R di bawah untuk membaca atau *import* data yang telah tersalin ke dalam RStudio. Data akan tersimpan ke dalam variabel `my_data`

```
my_data = read.table(file = 'clipboard', sep = '\t', header = TRUE)
```

Untuk membuka data dengan format CSV, ikuti langkah-langkah berikut:

```
my_data = read.csv(file = 'anticancer.csv', sep = ',', header = TRUE)
```

Catatan:

- a. *Argument* `sep` diisi oleh pemisah atau separator dari data. Hal ini dapat dilihat jika Anda membuka file CSV dengan text editor, misalnya Notepad
 - b. *Argument* `header` diisi oleh `TRUE` mengindikasikan bahwa baris pertama akan dibaca sebagai nama kolom
3. Untuk melihat data yang telah dibaca oleh RStudio. Jalankan perintah `view(my_data)`. Apabila keluaran dari perintah yang Anda ketikkan seperti gambar berikut, maka langkah yang Anda lakukan sudah berhasil.

	i.Accession	r	Vero	MCF.7
1	KL	1	40.65	14.78
2	KL	2	32.66	14.93
3	KL	3	39.63	16.28
4	PK	1	35.98	15.36
5	PK	2	24.99	8.65
6	PK	3	25.95	9.56
7	BH	1	30.63	12.26
8	BH	2	21.32	16.24
9	BH	3	22.98	10.00
10	GK	1	46.99	36.64
11	GK	2	34.56	37.30

Showing 1 to 12 of 33 entries, 4 total columns

Eksplorasi Data Menggunakan Statistika Lima Serangkai

Eksplorasi data dilakukan untuk memahami karakteristik dari data yang hendak dilakukan analisis lebih lanjut.

1. Menampilkan dimensi dari data berupa jumlah baris dan kolom

```
dim(my_data)
```

```
[1] 33 4
```

Simpulan: Data terdiri atas 33 baris dan 4 kolom

2. Menampilkan nama kolom dari data

```
colnames(my_data)
```

```
[1] "Accession" "r" "Vero" "MCF.7"
```

Simpulan: Data terdiri atas 4 kolom, yaitu Accession, r, Vero, dan MCF.7.

3. Mengganti nama kolom pada data

```
colnames(my_data) = c('Accession', 'Repetition', 'Vero', 'MCF7')
```

```
colnames(my_data)
```

```
[1] "Accession" "Repetition" "Vero" "MCF7"
```

Simpulan: Setelah nama kolom diubah, data terdiri atas 4 kolom, yaitu Accession, Repetition, Vero, dan MCF7.

4. Menampilkan gambaran umum tiap kolom

```
str(my_data)
'data.frame':   33 obs. of  4 variables:
 $ Accession : chr  "KL" "KL" "KL" "PK" ...
 $ Repetition: int   1 2 3 1 2 3 1 2 3 1 ...
 $ Vero      : num  40.6 32.7 39.6 36 25 ...
 $ MCF7     : num  14.78 14.93 16.28 15.36 8.65 ...
```

Simpulan: Fungsi str() akan menampilkan gambaran umum dari masing-masing kolom di data. Dapat disimpulkan kolom Accession merupakan kategorik (chr berarti character) sedangkan Repetition, Vero, dan MCF7 merupakan numerik (int berarti integer dan num berarti numeric).

5. Mengubah tipe data di tiap kolom (kolom Accession dan Repetition)

```
my_data$Accession = factor(my_data$Accession)
my_data$Repetition = factor(my_data$Repetition)
str(my_data)
'data.frame':   33 obs. of  4 variables:
 $ Accession : Factor w/ 11 levels "BH","CB","DOXO",...: 5 5 5 10 10...
 $ Repetition: Factor w/ 3 levels "1","2","3": 1 2 3 1 2 3 1 2 3 1 ...
 $ Vero      : num  40.6 32.7 39.6 36 25 ...
 $ MCF7     : num  14.78 14.93 16.28 15.36 8.65 ...
```

Simpulan: Fungsi factor() akan mengubah tipe data pada kolom Accession dan Repetition menjadi karakter. Hal ini berbeda dengan chr dan lebih mudah untuk dilakukan analisis.

6. Menampilkan statistika lima serangkai tiap kolom

```
summary(my_data)
  Accession  Repetition      Vero      MCF7
BH      : 3      1:11      Min.   :11.70  Min.   : 2.37
CB      : 3      2:11      1st Qu.:18.38  1st Qu.: 9.56
DOXO    : 3      3:11      Median :30.63  Median :16.24
GK      : 3              Mean   :30.19  Mean   :27.68
KL      : 3              3rd Qu.:40.65  3rd Qu.:42.00
KP      : 3              Max.   :57.13  Max.   :66.09
(Other):15
```

Simpulan: Fungsi summary() akan memetakan kolom yang tergolong kategorik dan numerik. Apabila kolom memiliki tipe kategorik, maka akan muncul jumlah kategori sedangkan apabila kolom memiliki tipe numerik, maka akan muncul statistika lima serangkai.

- Kolom Vero memiliki nilai minimum 11.70 dan maksimum 57.13. Sebaran data simetris (Mean = Median)
- Kolom MCF7 memiliki nilai minimum 2.37 dan maksimum 66.09. Sebaran data menjulur ke kanan atau right skewed (Mean > Median)

7. Menampilkan nilai unik pada kolom dengan tipe data kategorik

```
unique(my_data$Accession)
[1] KL  PK  BH  GK  KP  PW  MD  LC  CB  MB  DOXO
Levels: BH CB DOXO GK KL KP LC MB MD PK PW
```

Simpulan: Kolom Accession memiliki 11 kategori, yaitu KL, PK, BH, GK, KP, PW, MD, LC, CB, MB, dan DOXO. Levels dapat diabaikan apabila di dalam 11 kategori ini tidak terdapat tingkatan kategori.

Eksplorasi Data Lanjutan

1. Memeriksa adanya duplikasi data

```
replications(Vero ~ Accession, data = my_data)
Accession
3
```

Simpulan: Untuk masing-masing kategori pada kolom Accession, terdapat duplikasi data sebanyak tiga observasi. Duplikasi data ini merupakan pengulangan di dalam data rancangan percobaan.

```
replications(Vero ~ Accession, data = my_data[1:30,])
$Accession
Accession
  BH  CB DOXO  GK  KL  KP  LC  MB  MD  PK  PW
  3   3   0   3   3   3   3   3   3   3   3
```

Simpulan: Perintah `my_data[1:30,]` bermakna Anda hanya membaca data hingga baris 30 sehingga terdapat 3 data dengan kategori DOXO yang tidak digunakan. Namun hal ini dimaksudkan untuk memperlihatkan persebaran dari masing-masing kategori Accession.

2. Rata-rata dan ragam Vero untuk kategori Accession

```
with(my_data, tapply(Vero, list(Accession), mean))
  BH      CB      DOXO      GK      KL      KP
24.97667 18.94000 45.93000 39.73667 37.64667 44.89333
  LC      MB      MD      PK      PW
15.55333 15.13667 15.13333 28.97333 45.17333
```

Simpulan: Kategori MB memiliki rata-rata Vero yang paling kecil sedangkan kategori DOXO memiliki rata-rata Vero yang paling besar.

```
with(my_data, tapply(Vero, list(Accession), var))
  BH      CB      DOXO      GK      KL      KP
24.659033 44.930100  7.050100 41.860633 18.910233 17.757233
  LC      MB      MD      PK      PW
10.072933  6.282033 20.470233 37.050433 118.012633
```

Simpulan: Kategori MB memiliki ragam Vero yang paling kecil sedangkan kategori PW memiliki ragam Vero yang paling besar.

3. Rata-rata dan ragam MCF7 untuk kategori Accession

```
with(my_data, tapply(MCF7, list(Accession), mean))
  BH      CB      DOXO      GK      KL      KP
12.833333  3.040000 65.046667 36.010000 15.330000 42.520000
  LC      MB      MD      PK      PW
 4.806667 11.133333 53.650000 11.190000 48.906667
```

Simpulan: Kategori CB memiliki rata-rata MCF7 yang paling kecil sedangkan kategori CB memiliki rata-rata MCF7 yang paling besar.

```
with(my_data, tapply(MCF7, list(Accession), var))
  BH      CB      DOXO      GK      KL      KP
 9.980933  0.477300  2.197033  2.873700  0.682500 110.203200
  LC      MB      MD      PK      PW
 1.112133  5.672233 103.682500 13.248700 80.405633
```

Simpulan: Kategori KL memiliki ragam MCF7 yang paling kecil sedangkan kategori KP memiliki ragam MCF7 yang paling besar.

Install *Package* untuk Melakukan Visualisasi Data

Package atau diterjemahkan menjadi pustaka merupakan koleksi dari berbagai macam fungsi dan data yang dikembangkan untuk keperluan analisis data yang spesifik. Beberapa contoh *package* yang umum dijumpai dalam melakukan pengolahan data, sebagai berikut:

- `ggplot2` digunakan untuk membuat visualisasi data yang interaktif dan variatif
- `dplyr` digunakan untuk melakukan manipulasi data (*data aggregation*, *data filtering*, dan sebagainya)

1. Instalasi *package* `ggplot2` untuk visualisasi data

```
install.packages('ggplot2')
library(ggplot2)
```

Catatan:

- a. Pastikan Anda terhubung dengan koneksi internet saat melakukan instalasi *package*
- b. Proses instalasi dikatakan berhasil apabila saat menjalankan perintah `library(ggplot2)` tidak ditemukan adanya *error* pada *console*.

2. Sebaran *vero* berdasarkan *Accession*

```
# Visualisasi sebaran vero dengan Box and whisker Plot
```

```
ggplot(data = my_data)+
  geom_boxplot(aes(x = reorder(Accession, vero),
                  y = vero))+
  labs(title = 'vero berdasarkan Accession',
       subtitle = ''Data Anti Kanker ',
       caption = 'Pengenalan RStudio')+
  xlab('Accession')+
  ylab('vero')+
  theme_minimal()
```

```
# Membuat data rata-rata
```

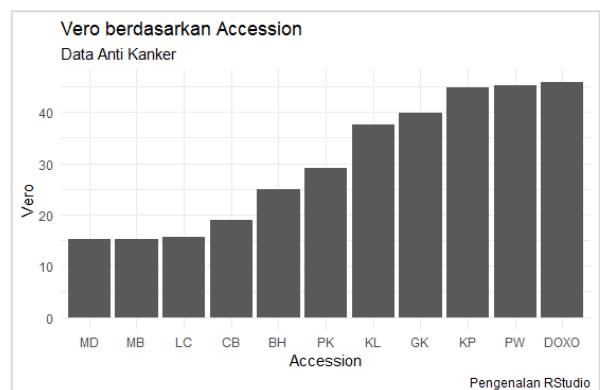
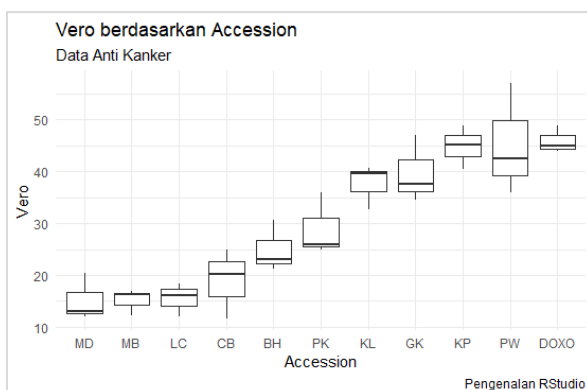
```
rata_vero = data.frame(as.table(
  with(my_data, tapply(vero, list(Accession), mean))))
```

```
# Mengubah nama kolom
```

```
colnames(rata_vero) = c('Accession', 'Rata')
```

```
# Visualisasi
```

```
ggplot(data = rata_vero)+
  geom_bar(aes(x = reorder(Accession, Rata),
                y = Rata),
          stat = 'identity')+
  labs(title = 'vero berdasarkan Accession',
       subtitle = 'Data Anti Kanker',
       caption = 'Pengenalan RStudio')+
  xlab('Accession')+
  ylab('vero')+
  theme_minimal()
```

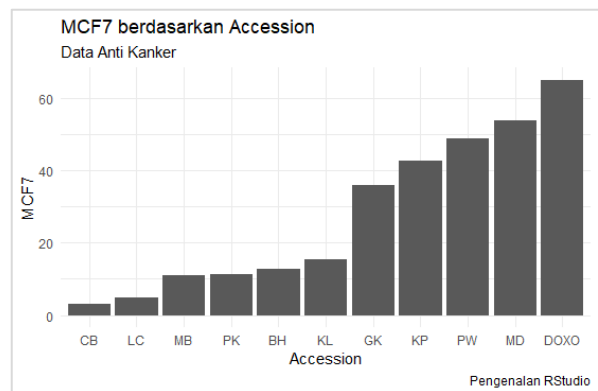
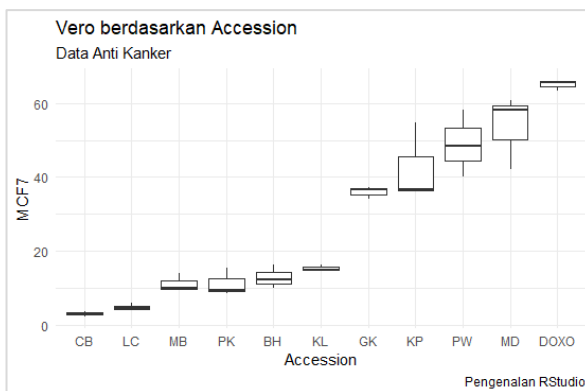


Simpulan: Sebaran kategori Accession sangat beragam, ditinjau dari nilai rata-rata hingga nilai ragam. Hal ini dapat menjadi indikasi bahwa faktor Accession berpengaruh terhadap nilai Vero.

3. Sebaran MCF7 berdasarkan Accession

```
# Visualisasi sebaran Vero dengan Box and whisker Plot
ggplot(data = my_data)+
  geom_boxplot(aes(x = reorder(Accession, MCF7),
                    y = MCF7))+
  labs(title = 'Vero berdasarkan Accession',
        subtitle = 'Data Anti Kanker ',
        caption = 'Pengenalan RStudio')+
  xlab('Accession')+
  ylab('MCF7')+
  theme_minimal()

# Membuat data rata-rata
rata_mcf7 = data.frame(as.table(
  with(my_data, tapply(MCF7, list(Accession), mean))))
# Mengubah nama kolom
colnames(rata_mcf7) = c('Accession', 'Rata')
# Visualisasi
ggplot(data = rata_mcf7)+
  geom_bar(aes(x = reorder(Accession, Rata),
                    y = Rata),
           stat = 'identity')+
  labs(title = 'MCF7 berdasarkan Accession',
        subtitle = 'Data Anti Kanker',
        caption = 'Pengenalan RStudio')+
  xlab('Accession')+
  ylab('MCF7')+
  theme_minimal()
```



Simpulan: Sebaran kategori Accession sangat beragam, ditinjau dari nilai rata-rata hingga nilai ragam. Hal ini dapat menjadi indikasi bahwa faktor Accession berpengaruh terhadap nilai MCF7.

Analysis of Variance (ANOVA) Menggunakan RStudio

Analysis of variance atau analisis sidik ragam adalah metode Statistika yang digunakan untuk menguji perbedaan rata-rata antara beberapa kelompok perlakuan. ANOVA diperkenalkan oleh Ronald Fisher. Dibandingkan dengan uji sample t (*t-test*) yang hanya dapat digunakan untuk menguji rata-rata antara dua kelompok perlakuan, ANOVA dapat menguji perbedaan lebih dari

dua kelompok perlakuan. ANOVA sering digunakan pada penelitian eksperimen dimana terdapat beberapa perlakuan. Hasil akhir dari ANOVA adalah *f-test* atau uji *f*.

Data yang digunakan pada penelitian Anda adalah data anti kanker dengan beberapa informasi yang dapat kita peroleh, sebagai berikut

- a. Faktor : Accession dengan 11 level perlakuan
- b. Output : vero dan MCF7
- c. Ulangan : Tiga kali ulangan per perlakuan

Sehingga rancangan percobaan yang digunakan adalah **Rancangan Percobaan Satu Faktor**, sedangkan rancangan lingkungan yang digunakan adalah **Rancangan Acak Lengkap dengan asumsi faktor lain yang memengaruhi output dianggap tetap**.

1. Membentuk hipotesis awal dan melakukan pengujian ANOVA

Hipotesis

H_0 : Tidak ada kategori Accession yang memengaruhi nilai Vero

H_1 : Minimal satu kategori Accession memengaruhi nilai Vero

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata (0.05)$

```
aov_out = aov(Vero ~ Accession, data = my_data)
```

```
aov_out
```

```
Call:
```

```
  aov(formula = Vero ~ Accession, data = my_data)
```

```
Terms:
```

```
                Accession Residuals
Sum of Squares  4973.809    694.111
Deg. of Freedom      10         22
```

```
Residual standard error: 5.616984
```

```
Estimated effects may be unbalanced
```

```
summary(aov_out)
```

```
          Df Sum Sq Mean Sq F value    Pr(>F)
Accession  10   4974    497.4   15.77 7.88e-08 ***
Residuals  22    694     31.6
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

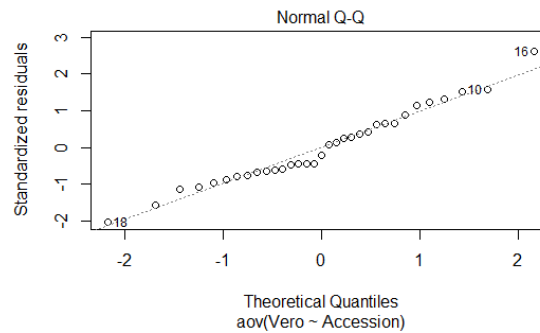
Keputusan: Tolak H_0 dikarenakan $pvalue (7.88e - 08) < taraf\ nyata (0.05)$

Simpulan: Cukup bukti untuk menyimpulkan bahwa minimal terdapat satu kategori Accession yang memberikan pengaruh terhadap nilai vero pada taraf nyata 5%.

2. Pengujian asumsi klasik ANOVA

Uji ANOVA akan valid apabila data terdistribusi Normal dan ragam antar kelompok perlakuan homogen. Sehingga setelah didapatkan keputusan bahwa minimal terdapat satu kategori Accession yang memberikan pengaruh terhadap nilai vero pada taraf nyata 5%, maka hasil tersebut akan diuji valid atau tidaknya.

- a. Uji Normalitas sisaan
`plot(aov_out, 2)`



Simpulan: Sisaan menyebar Normal dikarenakan seluruh titik berada di sekitar garis referensi 45°. Kesimpulan yang didapatkan secara eksploratif ini dapat diperkuat dengan melakukan uji formal menggunakan *Shapiro-Wilk test*.

Hipotesis

H_0 : Sisaan menyebar Normal

H_1 : Sisaan tidak menyebar Normal

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata\ (0.05)$

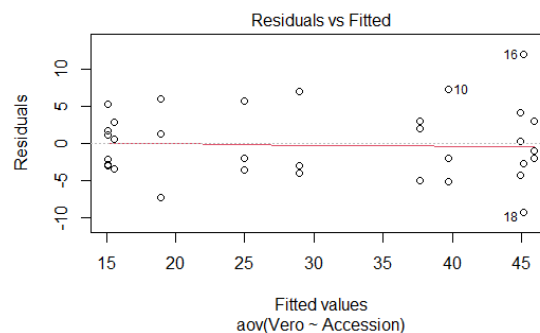
```
aov_residuals = residuals(object = aov_out)
shapiro.test(x = aov_residuals)
Shapiro-wilk normality test
```

```
data: aov_residuals
w = 0.97515, p-value = 0.634
```

Keputusan: Tidak tolak H_0 dikarenakan $pvalue\ (0.634) > taraf\ nyata\ (0.05)$

Simpulan: Tidak cukup bukti untuk menyimpulkan bahwa sisaan tidak menyebar Normal pada taraf nyata 5%.

- b. Uji asumsi kehomogenan ragam sisaan
`plot(aov_out, 1)`



Simpulan: Observasi pada baris 10, 16, dan 18 terdeteksi sebagai pencilan yang dapat memengaruhi hasil Normalitas sisaan dan kehomogenan ragam sisaan. Langkah yang

dapat dilakukan adalah menghapus observasi tersebut sehingga memenuhi asumsi kehomogenan ragam sisaan

Alternatif pengujian lain menggunakan *Levene's test* yang tidak sensitif terhadap penyimpangan dari sebaran Normalitas

Hipotesis

H_0 : Ragam sisaan homogen

H_1 : Ragam sisaan tidak homogen

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata (0.05)$

```
library(car)
leveneTest(Vero ~ Accession, data = my_data)
Levene's Test for Homogeneity of Variance (center = median)
  Df F value Pr(>F)
group 10  0.4244 0.9192
      22
```

Keputusan: Tidak tolak H_0 dikarenakan $pvalue (0.9192) > taraf\ nyata (0.05)$

Simpulan: Tidak cukup bukti untuk menyimpulkan bahwa ragam sisaan tidak homogen pada taraf nyata 5%.

Uji Lanjut atau Post-hoc Test Tuckey

Uji lanjut Tuckey atau dikenal dengan *Honestly Significance Diffirence (HSD)* merupakan salah satu metode uji lanjut yang diperkenalkan oleh Tuckey untuk membandingkan seluruh pasangan rata-rata kelompok perlakuan.

Hipotesis

H_0 : Tidak terdapat perbedaan antara masing – masing pasangan perlakuan

H_1 : Terdapat perbedaan antara masing – masing pasangan perlakuan

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata (0.05)$

```
TukeyHSD(aov_out)
Tukey multiple comparisons of means
 95% family-wise confidence level

Fit: aov(formula = Vero ~ Accession, data = my_data)

$Accession
      diff      lwr      upr      p adj
CB-BH -6.03666667 -22.4316452  10.3583119 0.9558219
DOXO-BH 20.953333333  4.5583548  37.3483119 0.0055501
GK-BH  14.760000000 -1.6349786  31.1549786 0.1028842
KL-BH  12.670000000 -3.7249786  29.0649786 0.2354465
KP-BH  19.916666667  3.5216881  36.3116452 0.0092839
LC-BH  -9.423333333 -25.8183119  6.9716452 0.6170715
MB-BH  -9.840000000 -26.2349786  6.5549786 0.5606118
MD-BH  -9.843333333 -26.2383119  6.5516452 0.5601612
PK-BH   3.996666667 -12.3983119  20.3916452 0.9977842
PW-BH  20.196666667  3.8016881  36.5916452 0.0080837
DOXO-CB 26.990000000 10.5950214  43.3849786 0.0002701
GK-CB  20.796666667  4.4016881  37.1916452 0.0060007
KL-CB  18.706666667  2.3116881  35.1016452 0.0167877
```

KP-CB	25.953333333	9.5583548	42.3483119	0.0004524
LC-CB	-3.386666667	-19.7816452	13.0083119	0.9994442
MB-CB	-3.803333333	-20.1983119	12.5916452	0.9985228
MD-CB	-3.806666667	-20.2016452	12.5883119	0.9985121
PK-CB	10.033333333	-6.3616452	26.4283119	0.5345645
PW-CB	26.233333333	9.8383548	42.6283119	0.0003934
GK-DOXO	-6.193333333	-22.5883119	10.2016452	0.9482356
KL-DOXO	-8.283333333	-24.6783119	8.1116452	0.7649395
KP-DOXO	-1.036666667	-17.4316452	15.3583119	1.0000000
LC-DOXO	-30.376666667	-46.7716452	-13.9816881	0.0000515
MB-DOXO	-30.793333333	-47.1883119	-14.3983548	0.0000421
MD-DOXO	-30.796666667	-47.1916452	-14.4016881	0.0000421
PK-DOXO	-16.956666667	-33.3516452	-0.5616881	0.0385972
PW-DOXO	-0.756666667	-17.1516452	15.6383119	1.0000000
KL-GK	-2.090000000	-18.4849786	14.3049786	0.9999930
KP-GK	5.156666667	-11.2383119	21.5516452	0.9845739
LC-GK	-24.183333333	-40.5783119	-7.7883548	0.0010980
MB-GK	-24.600000000	-40.9949786	-8.2050214	0.0008907
MD-GK	-24.603333333	-40.9983119	-8.2083548	0.0008892
PK-GK	-10.763333333	-27.1583119	5.6316452	0.4392661
PW-GK	5.436666667	-10.9583119	21.8316452	0.9777275
KP-KL	7.246666667	-9.1483119	23.6416452	0.8745188
LC-KL	-22.093333333	-38.4883119	-5.6983548	0.0031378
MB-KL	-22.510000000	-38.9049786	-6.1150214	0.0025456
MD-KL	-22.513333333	-38.9083119	-6.1183548	0.0025414
PK-KL	-8.673333333	-25.0683119	7.7216452	0.7164506
PW-KL	7.526666667	-8.8683119	23.9216452	0.8482275
LC-KP	-29.340000000	-45.7349786	-12.9450214	0.0000851
MB-KP	-29.756666667	-46.1516452	-13.3616881	0.0000695
MD-KP	-29.760000000	-46.1549786	-13.3650214	0.0000694
PK-KP	-15.920000000	-32.3149786	0.4749786	0.0619871
PW-KP	0.280000000	-16.1149786	16.6749786	1.0000000
MB-LC	-0.416666667	-16.8116452	15.9783119	1.0000000
MD-LC	-0.420000000	-16.8149786	15.9749786	1.0000000
PK-LC	13.420000000	-2.9749786	29.8149786	0.1775316
PW-LC	29.620000000	13.2250214	46.0149786	0.0000742
MD-MB	-0.003333333	-16.3983119	16.3916452	1.0000000
PK-MB	13.836666667	-2.5583119	30.2316452	0.1506184
PW-MB	30.036666667	13.6416881	46.4316452	0.0000607
PK-MD	13.840000000	-2.5549786	30.2349786	0.1504174
PW-MD	30.040000000	13.6450214	46.4349786	0.0000606
PW-PK	16.200000000	-0.1949786	32.5949786	0.0546369

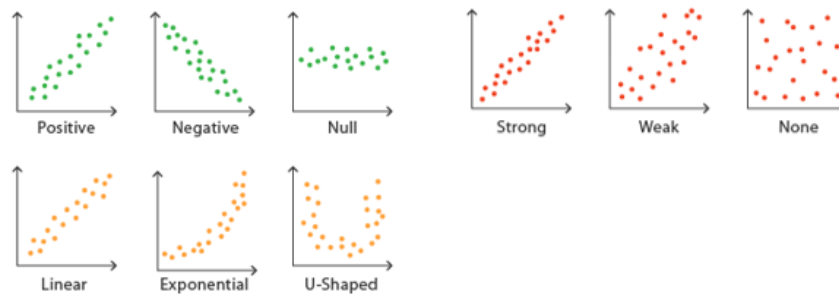
Keputusan: Tolak H_0 dikarenakan beberapa pasangan kelompok perlakuan memiliki $pvalue < taraf\ nyata\ (0.05)$

Simpulan: Cukup bukti untuk menyimpulkan bahwa terdapat perbedaan antara masing-masing pasangan perlakuan pada taraf nyata 5%. Beberapa pasangan kelompok perlakuan:

Analisis Korelasi Menggunakan RStudio

Secara sederhana, korelasi dapat diartikan sebagai hubungan. Dalam pengertian yang lebih luas, korelasi adalah ukuran yang digunakan untuk menggambarkan hubungan antara dua variabel kuantitatif. Hubungan dua variabel tersebut dapat terjadi karena adanya hubungan sebab akibat atau dapat pula terjadi karena kebetulan.

Dua variabel dikatakan berkorelasi apabila perubahan pada variabel yang satu akan diikuti perubahan pada variabel yang lain secara teratur dengan arah yang sama (korelasi positif) atau berlawanan (korelasi negatif).



1. Buka data dengan format CSV untuk dilakukan analisis korelasi

```
Data_korelasi = read.csv(file = 'data.csv', sep = ',', header = TRUE)
```

2. Menampilkan gambaran umum tiap kolom

```
str(Data_korelasi)
'data.frame': 15 obs. of 11 variables:
 $ geno      : chr  "A1" "A2" "A3" "A4" ...
 $ EY        : num  16.9 9 12.8 15.3 15.2 ...
 $ TPC       : num  15.27 5.76 5.28 5.09 16.82 ...
 $ RAC       : num  1.18 2.67 4.75 2.06 3.84 0.06 7.25 6.27 2.4 2.24 ...
 $ TFC       : num  2.41 1.24 1.87 1.99 1.72 0.1 0.84 1.72 0.89 1.13 ...
 $ SC        : num  1.21 1.82 0.53 0.73 2.17 0.61 1.45 0.36 0.55 0.87 ...
 $ DPPH      : num  15.43 7.87 12.37 14.28 13.75 ...
 $ FRAP      : num  1.6 0.42 1.28 0.82 0.91 0.07 1.6 1.36 0.83 1.19 ...
 $ HeLa      : num  52.6 57.5 52.4 55.7 53.9 ...
 $ MCF7      : num  36.7 57 50.9 41.5 58.6 ...
 $ GIA       : num  17.03 62.84 23.15 15.41 6.28 ...
```

Simpulan: Data terdiri atas 15 baris dan 11 kolom dimana dari 10 kolom memiliki tipe data numerik dan 1 kolom memiliki tipe data kategorik. Dapat disimpulkan bahwa kolom `geno` merupakan ID unik untuk 15 observasi.

3. Mencari koefisien korelasi terhadap seluruh kolom di data

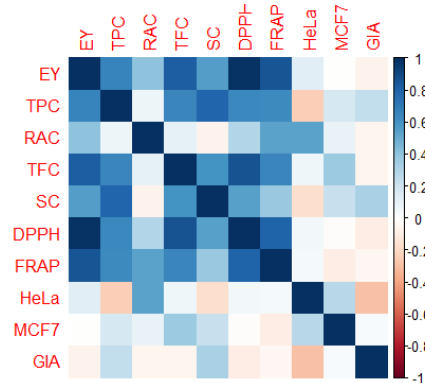
```
cor(data_korelasi[,-1], method = 'pearson')
```

	EY	TPC	RAC	TFC	SC
EY	1.00000000	0.66361092	0.40702032	0.82954879	0.5531678
TPC	0.66361092	1.00000000	0.07067672	0.65994919	0.7943288
RAC	0.40702032	0.07067672	1.00000000	0.10690578	-0.0668006
TFC	0.82954879	0.65994919	0.10690578	1.00000000	0.5924080
SC	0.55316783	0.79432876	-0.06680060	0.59240799	1.0000000
DPPH	0.98707752	0.64400793	0.29832948	0.86068854	0.5479680
FRAP	0.85230626	0.63642856	0.53822085	0.65503736	0.3756773
HeLa	0.12375984	-0.24461825	0.53702502	0.06222894	-0.1788720
MCF7	-0.00384829	0.17940778	0.09876573	0.36566412	0.2233705
GIA	-0.06513517	0.24624351	-0.05545468	-0.05012836	0.3287495

	DPPH	FRAP	HeLa	MCF7	GIA
EY	0.98707752	0.85230626	0.12375984	-0.00384829	-0.06513517
TPC	0.64400793	0.63642856	-0.24461825	0.17940778	0.24624351
RAC	0.29832948	0.53822085	0.53702502	0.09876573	-0.05545468
TFC	0.86068854	0.65503736	0.06222894	0.36566412	-0.05012836
SC	0.54796804	0.37567725	-0.17887197	0.22337049	0.32874947
DPPH	1.00000000	0.80704357	0.05350033	-0.01392927	-0.09000284
FRAP	0.80704357	1.00000000	0.04832341	-0.09157877	-0.04121886
HeLa	0.05350033	0.04832341	1.00000000	0.27808677	-0.29341500
MCF7	-0.01392927	-0.09157877	0.27808677	1.00000000	0.03295009
GIA	-0.09000284	-0.04121886	-0.29341500	0.03295009	1.00000000

4. Membuat visualisasi koefisien korelasi terhadap seluruh kolom di data

```
# Install package
install.packages('corrplot')
library(corrplot)
# Membuat visualisasi
matrix_corr = cor(data_korelasi[,-1], method = 'pearson')
corrplot(matrix_corr, method = 'color')
```



Simpulan: Variabel EY dan DPPH memiliki koefisien korelasi paling tinggi dan bernilai positif, artinya hubungan berbanding lurus. Selain itu, variabel HeLa, MCF7, dan GIA memiliki hubungan yang lemah, bahkan cenderung berbanding terbalik dengan variabel lainnya di dalam data.

5. Melakukan pengujian analisis korelasi pada kolom EY dan DPPH
 a. Uji asumsi Kenormalan data

Hipotesis

H_0 : Data menyebar Normal

H_1 : Data tidak menyebar Normal

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata (0.05)$

Kolom EY

```
shapiro.test(x = data_korelasi$EY)
Shapiro-wilk normality test
```

```
data: data_korelasi$EY
w = 0.92017, p-value = 0.1938
```

Kolom DPPH

```
shapiro.test(x = data_korelasi$DPPH)
Shapiro-wilk normality test
```

```
data: data_korelasi$DPPH
w = 0.93948, p-value = 0.3758
```

Keputusan: Tidak tolak H_0 dikarenakan $pvalue > taraf\ nyata (0.05)$

Simpulan: Tidak cukup bukti untuk menyimpulkan bahwa data tidak menyebar Normal pada taraf nyata 5%.

b. Uji analisis korelasi

Hipotesis

H_0 : Dua variabel tidak memiliki hubungan

H_1 : Dua variabel memiliki hubungan

Daerah kritis

Tolak H_0 apabila $pvalue < taraf\ nyata$ (0.05)

```
cor.test(data_korelasi$EY, data_korelasi$DPPH, method = 'pearson')  
Pearson's product-moment correlation
```

```
data: data_korelasi$EY and data_korelasi$DPPH  
t = 22.21, df = 13, p-value = 1.008e-11  
alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
 0.9604695 0.9958139  
sample estimates:  
      cor  
0.9870775
```

Keputusan: Tolak H_0 dikarenakan $pvalue$ ($1.008e - 11$) $< taraf\ nyata$ (0.05)

Simpulan: Cukup bukti untuk menyimpulkan bahwa variabel EY dan DPPH memiliki hubungan pada taraf nyata 5%.