

ANALISIS DATA COVID-19 DI JAWA TIMUR MENGGUNAKAN R

I. Latar Belakang

COVID-19 merupakan penyakit yang saat ini telah menjadi pandemi secara global. Kondisi menjadi semakin mengkhawatirkan karena hingga detik ini masih belum ditemukan vaksin yang efektif untuk virus penyebab COVID-19. Pemerintah di berbagai negara umumnya dengan sigap membentuk gugus tugas (*task force unit*) untuk menangani penyebaran COVID-19 di masyarakat, termasuk pemerintah di Indonesia.

Salah satu bentuk aksi yang dilakukan oleh pemerintah adalah dengan mengumpulkan dan menyediakan data pertumbuhan kasus COVID-19 kepada publik. Data pertumbuhan kasus tersebut tidak jarang juga dilengkapi dengan dasbor dan grafik visualisasi pendukung dengan harapan masyarakat dapat memahami informasi dengan mudah. Sebagai contoh adalah portal covid19.go.id besutan Gugus Tugas Penanganan COVID-19 Nasional serta portal data lainnya yang disediakan oleh masing-masing pemerintah daerah.

II. Tugas dan Langkah

Pada proyek ini diminta untuk melakukan eksplorasi dan analisis mengenai COVID-19 di Jawa Timur. Pengambilan data secara langsung menggunakan *Application Programming Interface* (API) yang tersedia di portal covid19.go.id dan diakses pada tanggal 25 Agustus 2020 (terakhir diperbaharui tanggal 24 Agustus 2020). Secara sederhana API dapat dipahami sebagai cara komputer untuk berkomunikasi dengan komputer lainnya sehingga dapat terjadi transaksi pertukaran data. Sehingga dalam proyek ini akan mempelajari teknik mengambil data dari API, teknik mempersiapkan data serta analisis dan visualisasi data.

Langkah yang akan digunakan sebagai berikut:

1. Mengimpor data *real-time* menggunakan API

Mengakses API

Rekapitulasi data COVID-19 di Jawa Timur tersedia dalam API publik yang beralamat https://data.covid19.go.id/public/api/prov_detail_JAWA_TIMUR.json dengan cara mengaksesnya menggunakan fungsi `GET()` dari paket `httr`.

```
> library(httr)
> resp <- GET("https://data.covid19.go.id/public/api/prov_detail_JAWA_TIMUR.json")
```

Mengevaluasi Respon

Saat menjalankan `GET()`, pada dasarnya hal yang terjadi adalah membuat sebuah permintaan kepada server penyedia API. Selanjutnya permintaan tersebut diproses dan dijawab oleh server sebagai sebuah respon. Ada tiga informasi utama dalam sebuah respon API, yaitu:

- i. Body berisikan konten atas permintaan yang telah dibuat
- ii. Status memiliki informasi apakah permintaan tersebut berhasil atau tidak dan dinyatakan dalam `status_code`. Ada beberapa jenis `status_code` yang umum dijumpai, antara lain:
 - 200 artinya permintaan sukses dipenuhi
 - 404 artinya berkas yang diminta tidak dapat ditemukan
 - 403 artinya akses permintaan ditolak
 - 500 artinya terjadi kesalahan pada server

```
> status_code(resp)
[1] 200
```

- iii. Headers umumnya mengandung informasi metadata dan dinyatakan dalam `headers()`.

```
> headers(resp)
$server
[1] "nginx"

$date
[1] "Tue, 25 Aug 2020 02:46:41 GMT"

$content-type
[1] "application/json"

$`last-modified`
[1] "Tue, 25 Aug 2020 02:23:31 GMT"

$`transfer-encoding`
[1] "chunked"

$connection
[1] "keep-alive"

$vary
[1] "Accept-Encoding"

$etag
[1] "w/\\"5f447623-9e90\\""

$content-encoding
[1] "gzip"

attr(,"class")
[1] "insensitive" "list"
```

Mengekstrak Isi Respon

Respon API dengan *status code* 200 menyatakan bahwa permintaan berhasil dipenuhi dan konten yang diminta tersedia untuk diekstrak. Selain itu telah diketahui lewat fungsi `headers()` bahwa konten yang diminta tersedia dalam bentuk *application/json*, yaitu berkas JSON. Selanjutnya mengekstrak konten tersebut menggunakan fungsi `content()`. Fungsi `content()` tersebut secara cerdas akan menyesuaikan output sesuai dengan jenis berkas yang tersedia serta fungsi `length()` dan `names()` digunakan untuk mengamati komponen serta nama komponennya.

```
> covid_raw <- content(resp, as = "parsed", simplifyvector = TRUE)
> length(covid_raw)
[1] 13
> names(covid_raw)
[1] "last_date"      "provinsi"      "kasus_total"
[4] "kasus_tanpa_tgl" "kasus_dengan_tgl" "meninggal_persen"
[7] "meninggal_tanpa_tgl" "meninggal_dengan_tgl" "sembuh_persen"
[10] "sembuh_tanpa_tgl" "sembuh_dengan_tgl" "list_perkembangan"
[13] "data"
```

2. Melakukan proses data *cleaning* dan transformasi data

Informasi Umum

Menggunakan variable `covid_raw` untuk menjawab pertanyaan-pertanyaan berikut:

- Berapa jumlah total kasus COVID-19 di Jawa Timur?
- Berapa persentase kematian akibat COVID-19 di Jawa Timur?
- Berapa persentase tingkat kesembuhan dari COVID-19 di Jawa Timur?

```
> covid_raw$kasus_total
[1] 30635
> covid_raw$meninggal_persen
[1] 7.165007
> covid_raw$sembuh_persen
[1] 78.18835
```

Perkembangan COVID-19

```
> covid <- covid_raw$list_perkembangan
> str(covid)
'data.frame': 160 obs. of 9 variables:
 $ tanggal      : num  1.58e+12 1.58e+12 1.58e+12 1.58e+12 1.58e+12 ...
 $ KASUS         : int   2 0 0 0 0 0 0 0 8 7 ...
 $ MENINGGAL     : int   0 0 0 0 0 0 1 2 0 0 ...
 $ SEMBUH        : int   0 1 0 0 0 0 0 2 2 0 ...
 $ DIRAWAT_OR_ISOLASI : int   2 -1 0 0 0 0 -1 -4 6 7 ...
 $ AKUMULASI_KASUS : int   2 2 2 2 2 2 2 10 17 ...
 $ AKUMULASI_SEMBUH : int   0 1 1 1 1 1 1 3 5 5 ...
 $ AKUMULASI_MENINGGAL : int   0 0 0 0 0 0 1 3 3 3 ...
 $ AKUMULASI_DIRAWAT_OR_ISOLASI : int   2 1 1 1 1 1 0 -4 2 9 ...
> head(covid)
  tanggal KASUS MENINGGAL SEMBUH DIRAWAT_OR_ISOLASI AKUMULASI_KASUS
1 1.584490e+12      2         0         0                2                2
2 1.584576e+12      0         0         1               -1                2
3 1.584662e+12      0         0         0                0                2
4 1.584749e+12      0         0         0                0                2
5 1.584835e+12      0         0         0                0                2
6 1.584922e+12      0         0         0                0                2
  AKUMULASI_SEMBUH AKUMULASI_MENINGGAL AKUMULASI_DIRAWAT_OR_ISOLASI
1                0                    0                    2
2                1                    0                    1
3                1                    0                    1
4                1                    0                    1
5                1                    0                    1
6                1                    0                    1
```

Setelah mengekstrak dan mengamati `covid`, terdapat kejanggalan data pada kolom tanggal dan format yang tidak konsisten sehingga akan dijinakkan terlebih dulu supaya lebih mudah diolah dan dianalisis. Ada beberapa tahapan untuk menjinakan data `covid`, yaitu:

- i. Mengaktifkan paket `dplyr`
- ii. Menggunakan operator pipe (`%>%`) untuk merangkai fungsi menjadi pipeline
- iii. Menghapus semua kolom yang berisi nilai kumulatif
- iv. Mengganti nama kolom “KASUS” menjadi “kasus_baru”
- v. Merubah format penulisan kolom MENINGGAL dan SEMBUH menjadi huruf kecil
- vi. Memperbaiki data pada kolom tanggal

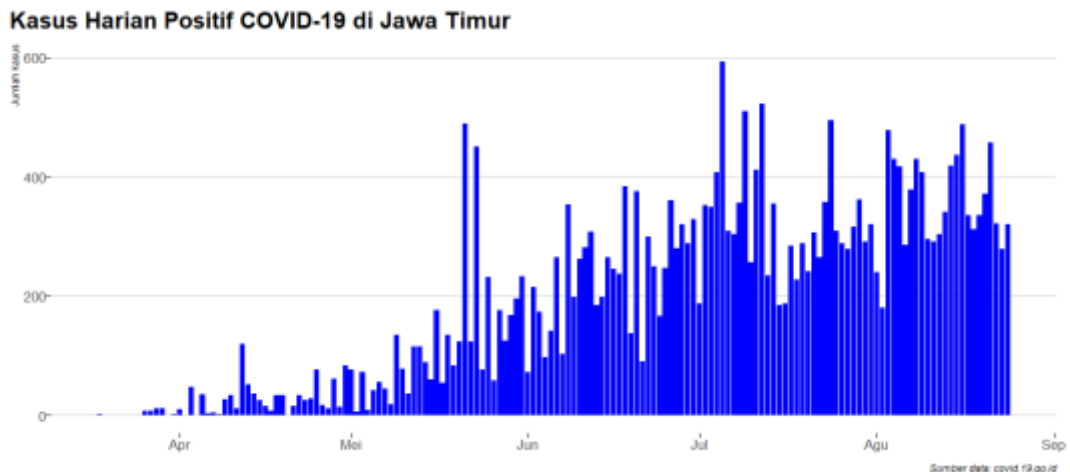
```
> library(dplyr)
> new_covid <-
+   covid %>%
+   select(-contains("DIRAWAT_OR_ISOLASI")) %>%
+   select(-starts_with("AKUMULASI")) %>%
+   rename(
+     kasus_baru = KASUS,
+     meninggal = MENINGGAL,
+     sembuh = SEMBUH
+   ) %>%
+   mutate(
+     tanggal = as.POSIXct(tanggal / 1000, origin = "1970-01-01"),
+     tanggal = as.Date(tanggal)
+   )
> str(new_covid)
'data.frame': 160 obs. of 4 variables:
 $ tanggal : Date, format: "2020-03-18" "2020-03-19" ...
 $ kasus_baru: int 2 0 0 0 0 0 0 0 8 7 ...
 $ meninggal : int 0 0 0 0 0 0 1 2 0 0 ...
 $ sembuh : int 0 1 0 0 0 0 0 2 2 0 ...
```

3. Membuat visualisasi data dengan grafik

Komponen untuk membuat visualisasi antara lain tabel data, kolom data, serta bentuk geometri untuk mempresentasikan data. Grafik merupakan salah satu cara untuk mempresentasikan data tersebut dalam bentuk lain yang harapannya lebih mudah dicerna. Untuk merancang visualisasi yang memiliki estetika akan digunakan paket `ggplot2` dan paket `hrbrthemes`.

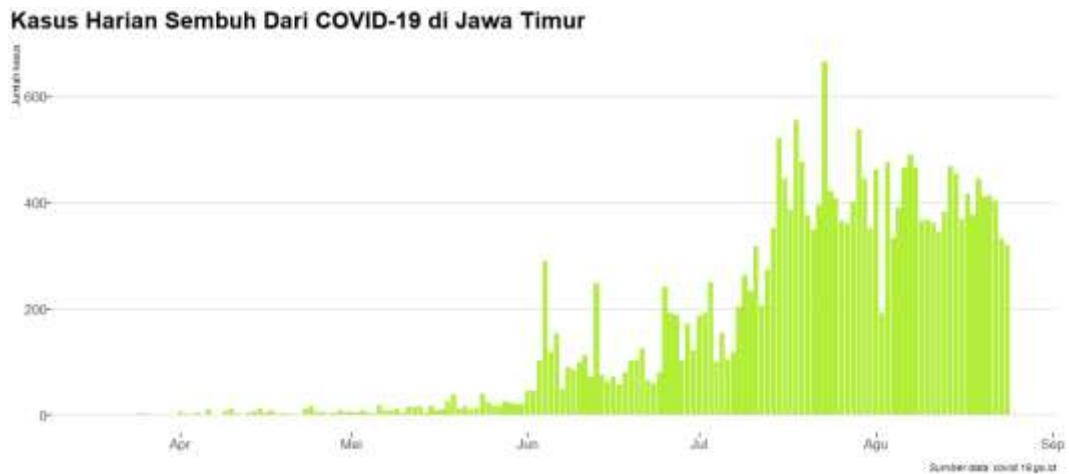
1) Grafik Kasus Harian positif COVID-19

```
> library(ggplot2)
> library(hrbrthemes)
> ggplot(new_covid, aes(tanggal, kasus_baru)) +
+   geom_col(fill = "blue") +
+   labs(
+     x = NULL,
+     y = "Jumlah kasus",
+     title = "Kasus Harian Positif COVID-19 di Jawa Timur",
+     caption = "Sumber data: covid.19.go.id"
+   ) +
+   theme_ipsum(
+     base_size = 13,
+     plot_title_size = 21,
+     grid = "Y",
+     ticks = TRUE
+   ) +
+   theme(plot.title.position = "plot")
```



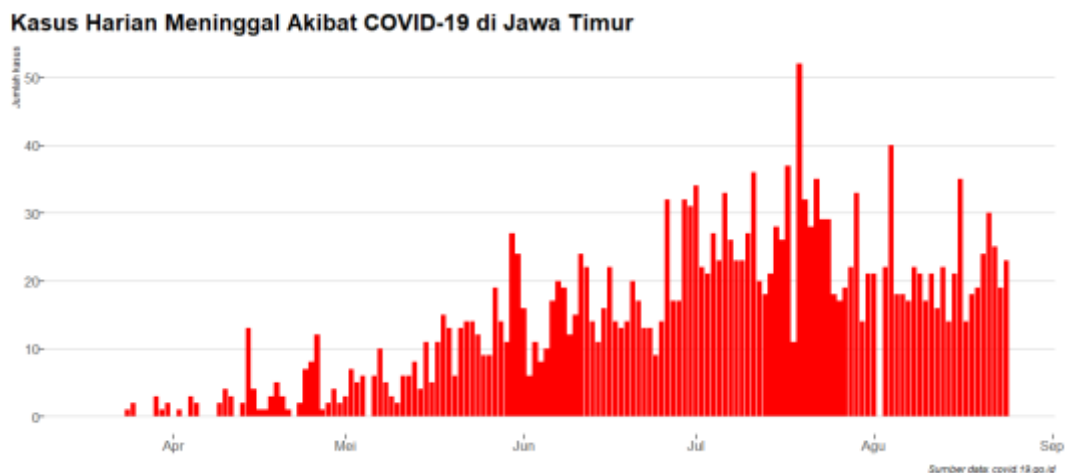
2) Grafik kasus harian sembuh dari COVID-19

```
> library(ggplot2)
> library(hrbrthemes)
> ggplot(new_covid, aes(tanggal, sembuh)) +
+   geom_col(fill = "olivedrab2") +
+   labs(
+     x = NULL,
+     y = "Jumlah kasus",
+     title = "Kasus Harian Sembuh Dari COVID-19 di Jawa Timur",
+     caption = "Sumber data: covid.19.go.id"
+   ) +
+   theme_ipsum(
+     base_size = 13,
+     plot_title_size = 21,
+     grid = "Y",
+     ticks = TRUE
+   ) +
+   theme(plot.title.position = "plot")
```



3) Grafik kasus harian meninggal akibat COVID-19

```
--
> library(ggplot2)
> library(hrbrthemes)
> ggplot(new_covid, aes(tanggal, meninggal)) +
+   geom_col(fill = "red") +
+   labs(
+     x = NULL,
+     y = "Jumlah kasus",
+     title = "Kasus Harian Meninggal Akibat COVID-19 di Jawa Timur",
+     caption = "Sumber data: covid.19.go.id"
+   ) +
+   theme_ipsum(
+     base_size = 13,
+     plot_title_size = 21,
+     grid = "Y",
+     ticks = TRUE
+   ) +
+   theme(plot.title.position = "plot")
```



4. Pola dan Dinamika COVID-19

Pada kasus ini, ada yang akhirnya sembuh namun tak sedikit yang meninggal akibat COVID-19. Sementara itu penambahan kasus baru terus terjadi. Hal ini memicu pertanyaan “Hingga saat ini ada berapa kasus yang masih aktif?”.

Jumlah kasus aktif dapat dihitung dengan cara mengurangi jumlah akumulasi positif dengan jumlah akumulasi sembuh dan jumlah akumulasi meninggal. Untuk menghitung nilai akumulasi dari suatu vektor numerik menggunakan fungsi `cumsum()`.

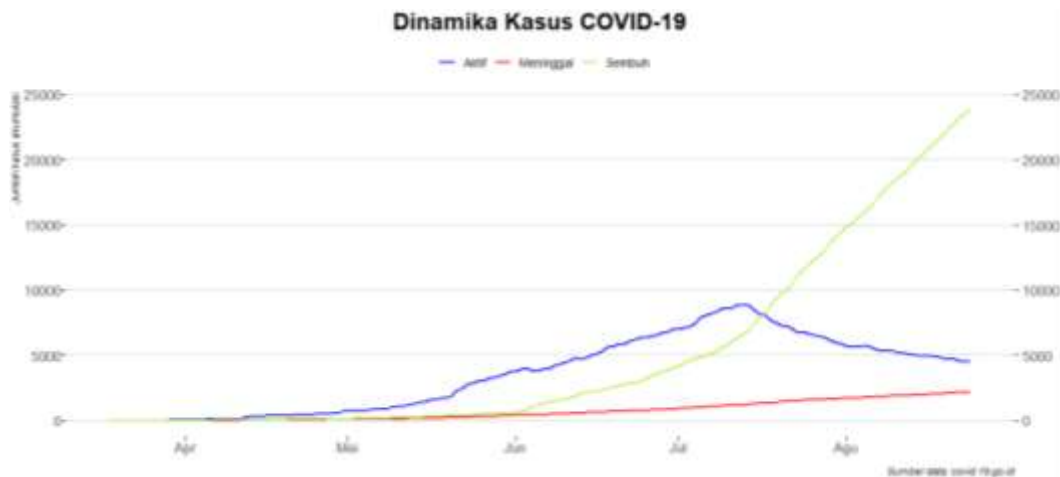
```
> library(dplyr)
> covid_akumulasi <-
+   new_covid %>%
+   transmute(
+     tanggal,
+     akumulasi_aktif = cumsum(kasus_baru) - cumsum(sembuh) - cumsum(meninggal),
+     akumulasi_sembuh = cumsum(sembuh),
+     akumulasi_meninggal = cumsum(meninggal)
+   )
```

Untuk mentransformasi data, gunakan fungsi `gather()` dari paket `tidyr`. Data tersebut akan dirubah dari yang semula berformat *wide* menjadi format *long*.

```
> library(dplyr)
> library(tidyr)
> dim(covid_akumulasi)
[1] 160 4
> covid_akumulasi_pivot <-
+   covid_akumulasi %>%
+   gather(
+     key = "kategori",
+     value = "jumlah",
+     -tanggal
+   ) %>%
+   mutate(
+     kategori = sub(pattern = "akumulasi_", replacement = "", kategori)
+   )
> dim(covid_akumulasi_pivot)
[1] 480 3
> glimpse(covid_akumulasi_pivot)
Rows: 480
Columns: 3
$ tanggal <date> 2020-03-18, 2020-03-19, 2020-03-20, 2020-03-21, 2020-03-22, 2020-03-23,
  2020-03-24, 2020-03-25, 2020-03-26, 2020-03-27, 2020-03-28, 2020-...
$ kategori <chr> "aktif", "aktif", "aktif", "aktif", "aktif", "aktif", "aktif", "aktif",
  "aktif", "aktif", "aktif", "aktif", "aktif", "aktif", "aktif", "ak...
$ jumlah <int> 2, 1, 1, 1, 1, 1, 0, -4, 2, 9, 20, 29, 29, 29, 34, 31, 79, 72, 104, 97,
  101, 102, 120, 138, 144, 263, 309, 325, 335, 346, 345, 373, 399, 3...
```


Grafik komparasi antara akumulasi kasus aktif, sembuh, dan meninggal

```
> library(ggplot2)
> library(hrbrthemes)
> ggplot(covid_akumulasi_pivot, aes(tanggal, jumlah, colour = (kategori))) +
+   geom_line(size = 0.9) +
+   scale_y_continuous(sec.axis = dup_axis(name = NULL)) +
+   scale_colour_manual(
+     values = c(
+       "aktif" = "blue",
+       "meninggal" = "red",
+       "sembuh" = "olivedrab2"
+     ),
+     labels = c("Aktif", "Meninggal", "Sembuh")
+   ) +
+   labs(
+     x = NULL,
+     y = "Jumlah kasus akumulasi",
+     colour = NULL,
+     title = "Dinamika Kasus COVID-19",
+     caption = "Sumber data: covid.19.go.id"
+   ) +
+   theme_ipsum(
+     base_size = 13,
+     plot_title_size = 21,
+     grid = "Y",
+     ticks = TRUE
+   ) +
+   theme(
+     plot.title = element_text(hjust = 0.5),
+     legend.position = "top"
+   )
+ )
```



5. Perkembangan COVID_19 dalam Rentang waktu pekanan

Menggunakan paket `lubridate` supaya lebih mudah bekerja dengan data tanggal dan menggunakan fungsi `week()` untuk mengekstrak informasi pekanan dalam satu tahun.

```

> library(dplyr)
> library(lubridate)
> covid_pekanan <- new_covid %>%
+   count(
+     tahun = year(tanggal),
+     pekan_ke = week(tanggal),
+     wt = kasus_baru,
+     name = "jumlah"
+   )
> glimpse(covid_pekanan)
Rows: 23
Columns: 3
$ tahun    <dbl> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020...
$ pekan_ke <dbl> 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, ...
$ jumlah   <int> 2, 41, 101, 280, 129, 253, 302, 484, 712, 1552, 1184, 1333, 1745...

```

“Apakah pekan ini lebih baik dari pekan kemarin?” untuk menjawab pertanyaan tersebut, perlu melakukan kalkulasi sederhana dengan tahapan berikut:

- i. Membuat kolom baru “jumlah_pekanlalu” yang berisi jumlah kasus baru dalam satu pekan sebelumnya
- ii. Menggunakan fungsi `lag()` dari paket `dplyr` untuk membuat kolom “jumlah_pekanlalu”
- iii. Mengganti nilai NA pada kolom “jumlah_pekanlalu” dengan nilai 0
- iv. Melakukan komparasi antara “jumlah” dengan kolom “jumlah_pekanlalu”. Hasilnya disimpan dalam kolom baru dengan nama “lebih_baik” yang berisi TRUE apabila jumlah kasus pekan ini lebih rendah dibandingkan jumlah kasus pekan lalu.

```

> library(dplyr)
> covid_pekanan <-
+   covid_pekanan %>%
+   mutate(
+     jumlah_pekanlalu = dplyr::lag(jumlah, 1),
+     jumlah_pekanlalu = ifelse(is.na(jumlah_pekanlalu), 0, jumlah_pekanlalu),
+     lebih_baik = jumlah < jumlah_pekanlalu
+   )
> glimpse(covid_pekanan)
Rows: 23
Columns: 5
$ tahun    <dbl> 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 2020, 20...
$ pekan_ke <dbl> 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, ...
$ jumlah    <int> 2, 41, 101, 280, 129, 253, 302, 484, 712, 1552, 1184, 13...
$ jumlah_pekanlalu <dbl> 0, 2, 41, 101, 280, 129, 253, 302, 484, 712, 1552, 1184,...
$ lebih_baik <lgl> FALSE, FALSE, FALSE, FALSE, TRUE, FALSE, FALSE, FALSE, F...

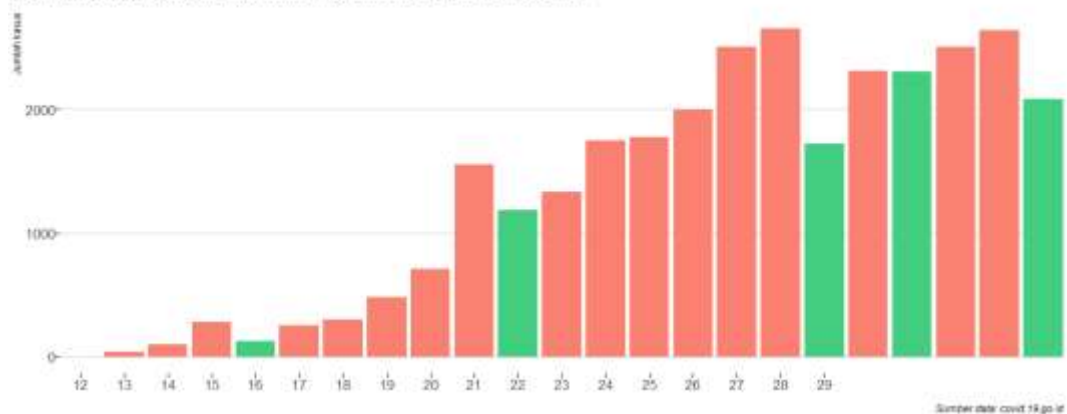
```

Grafik Kasus Pekan

```
> library(ggplot2)
> library(hrbrthemes)
> ggplot(covid_pekanan, aes(pekan_ke, jumlah, fill = lebih_baik)) +
+   geom_col(show.legend = FALSE) +
+   scale_x_continuous(breaks = 9:29, expand = c(0, 0)) +
+   scale_fill_manual(values = c("TRUE"="seagreen3", "FALSE"="salmon")) +
+   labs(
+     x = NULL,
+     y = "Jumlah kasus",
+     title = "Kasus Pekan Positif COVID-19 di Jawa Timur",
+     subtitle = "Kolom hijau menunjukkan penambahan kasus baru lebih sedikit",
+     caption = "Sumber data: covid.19.go.id"
+   ) +
+   theme_ipsum(
+     base_size = 13,
+     plot_title_size = 21,
+     grid = "Y",
+     ticks = TRUE
+   ) +
+   theme(plot.title.position = "plot")
```

Kasus Pekan Positif COVID-19 di Jawa Timur

Kolom hijau menunjukkan penambahan kasus baru lebih sedikit dibandingkan satu pekan sebelumnya



III. Kesimpulan

Provinsi Jawa Timur per tanggal 25 Agustus 2020 memiliki tren cukup baik COVID-19 dimana dari jumlah kasus total sebanyak 30.635 orang, persentase meninggal akibat COVID-19 sebanyak 7.16% dan persentase sembuh dari COVID-19 sebanyak 78.19%.

Pada grafik kasus harian positif COVID-19, sembuh dari COVID-19, dan meninggal akibat COVID-19 cenderung fluktuatif. Selain itu dilihat dari grafik komparasi, kasus aktif sudah mengalami penurunan, kasus sembuh meningkat setiap harinya, dan kasus meninggal perlu ditekan lagi agar tidak terjadi kenaikan lagi. Untuk kasus pekan positif COVID-19 untuk pekan ini lebih baik dibandingkan pekan sebelumnya.