

## Kuis\_DS-B

Torangto Situngkir\_123180122

##Intro ##Baca Petunjuk Terlebih Dahulu!

0. Cuci tangan dengan sabun hingga benar-benar bersih dengan durasi mencuci tangan kurang lebih 20 detik
1. Kerjakan soal-soal yang ada! Jangan lupa tulis NAMA dan NIM pada author!
2. Kuis terdiri dari 2 bagian yaitu bagian pertama dan bagian kedua
3. Jawablah dengan mengisi chunk dibawah soal!
4. Durasi pengerjaan sesuai selama 3 hari, dikumpulkan maksimal Rabu, 9 Desember 2020
5. No toleransi pengumpulan telat. Ingat, telat kemungkinan terburuk ga ada nilai kuis!
6. Misal soal rancu bisa menghubungi asisten terkait
7. Export hasil pekerjaan dalam format PDF/Word & sesuaikan nama file sesuai NIM masing-masing.

##BAGIAN PERTAMA

1. Load library apa saja yang kira-kira digunakan! Lalu gunakan data 'us\_contagious\_diseases'! **point 1**

```
library(dslabs)
library(tidyverse)

## -- Attaching packages ----- tidyverse
1.3.0 --

## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.4      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.0

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

data("us_contagious_diseases")
```

2. Tampilkan semua nama kolom pada data frame yang ada! **point 5**

```
names(us_contagious_diseases)

## [1] "disease"      "state"        "year"
## [2] "weeks_reporting"
## [5] "count"       "population"
```

3. Tampilkan tipe data pada kolom penyakit! **point 5**

```
class(us_contagious_diseases$disease)
```

```
## [1] "factor"
```

4. Tampilkan 10 data penyakit polio teratas diurutkan berdasarkan populasi dan terjadi pada antara tahun 1965 dan 1955! **point 11**

```
us_contagious_diseases %>%  
filter(year>1955 & year<1965,disease=="Polio") %>%  
arrange(desc(population)) %>%  
head(n=10)
```

```
##   disease      state year weeks_reporting count population  
## 1   Polio California 1964             25      2   17581336  
## 2   Polio   New York 1964             39     10   17572556  
## 3   Polio   New York 1963             48     12   17387034  
## 4   Polio   New York 1962             48     88   17190958  
## 5   Polio California 1963             52     19   17135662  
## 6   Polio   New York 1961             48    256   16988157  
## 7   Polio   New York 1960             52    280   16782304  
## 8   Polio California 1962             50     88   16676046  
## 9   Polio   New York 1959             39    353   16576443  
## 10  Polio   New York 1958             50    292   16371622
```

5. Klasifikasikan data tersebut berdasarkan jumlah kasusnya dengan kondisi : -jumlah kasus kurang dari 1000 dikategorikan sebagai "Biasa" -jumlah kasus lebih dari 2000 dikategorikan sebagai "Azab -jumlah kasus antara kedua kondisi diatas dikategorikan sebagai"Cobaan"

NB : jika dirasa jumlah data hasilnya terlalu banyak boleh menggunakan fungsi top\_n() atau head()**point 10**

```
klasifikasi = ""  
for(i in 1:nrow(us_contagious_diseases)){  
  if(us_contagious_diseases$count[i] < 1000) {  
    klasifikasi[i] <- "Biasa"  
  }  
  else if(us_contagious_diseases$count[i] > 2000) {  
    klasifikasi[i] <- "Azab"  
  }  
  else{  
    klasifikasi[i] <- "Cobaan"  
  }  
}  
head(klasifikasi,n=300)
```

```
## [1] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"  
"Biasa"  
## [9] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"  
"Biasa"
```

## [17]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [25]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [33]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [41]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [49]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [57]	"Cobaan"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [65]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [73]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [81]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [89]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [97]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [105]	"Cobaan"	"Cobaan"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Cobaan"						
## [113]	"Cobaan"	"Cobaan"	"Biasa"	"Azab"	"Cobaan"	"Biasa"	"Biasa"
	"Biasa"						
## [121]	"Biasa"	"Biasa"	"Cobaan"	"Cobaan"	"Cobaan"	"Biasa"	"Biasa"
	"Biasa"						
## [129]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [137]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [145]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [153]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [161]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [169]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [177]	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"	"Biasa"
	"Biasa"						
## [185]	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"
	"Azab"						
## [193]	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"
	"Azab"						
## [201]	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"
	"Azab"						
## [209]	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"	"Azab"
	"Azab"						

```
## [217] "Azab" "Azab" "Azab" "Cobaan" "Cobaan" "Cobaan" "Biasa"
"Biasa"
## [225] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [233] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [241] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [249] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [257] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [265] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [273] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [281] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [289] "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa" "Biasa"
"Biasa"
## [297] "Biasa" "Biasa" "Biasa" "Biasa"
```

6. Tambahkan variabel baru berisi data 'us\_contagious\_diseases' dengan tambahan kolom baru dengan nama "category" yang isinya merupakan implementasi nomor 5 dan kolom "rate" yang isinya merupakan hasil bagi jumlah kasus dengan populasi dikalikan 100000! **point 10**

```
newData <- us_contagious_diseases %>%
mutate(
  Category = klasifikasi,
  Rate = count/population*10^5
)
head(newData,n=10)
```

##	disease	state	year	weeks_reporting	count	population	Category	Rate
## 1	Hepatitis A	Alabama	1966	50	321	3345787	Biasa	9.594155
## 2	Hepatitis A	Alabama	1967	49	291	3364130	Biasa	8.650082
## 3	Hepatitis A	Alabama	1968	52	314	3386068	Biasa	9.273293
## 4	Hepatitis A	Alabama	1969	49	380	3412450	Biasa	11.135694
## 5	Hepatitis A	Alabama	1970	51	413	3444165	Biasa	11.991295
## 6	Hepatitis A	Alabama	1971	51	378	3481798	Biasa	10.856460
## 7	Hepatitis A	Alabama	1972	45	342	3524543	Biasa	9.703386
## 8	Hepatitis A	Alabama	1973	45	467	3571209	Biasa	

```

13.076804
## 9 Hepatitis A Alabama 1974          45    244    3620548    Biasa
6.739311
## 10 Hepatitis A Alabama 1975         46    286    3671246    Biasa
7.790271

```

7. Tampilkan kesimpulan pada data nomor 6 dikelompokkan berdasarkan negara bagian yang isinya nama negara bagian dan rata-rata rate per negara bagian! **point 8**

```

kesimpulan <- newData %>%
  group_by(state) %>%
  summarize(Rata_Rata=mean(Rate))

## `summarise()` ungrouping output (override with `.groups` argument)

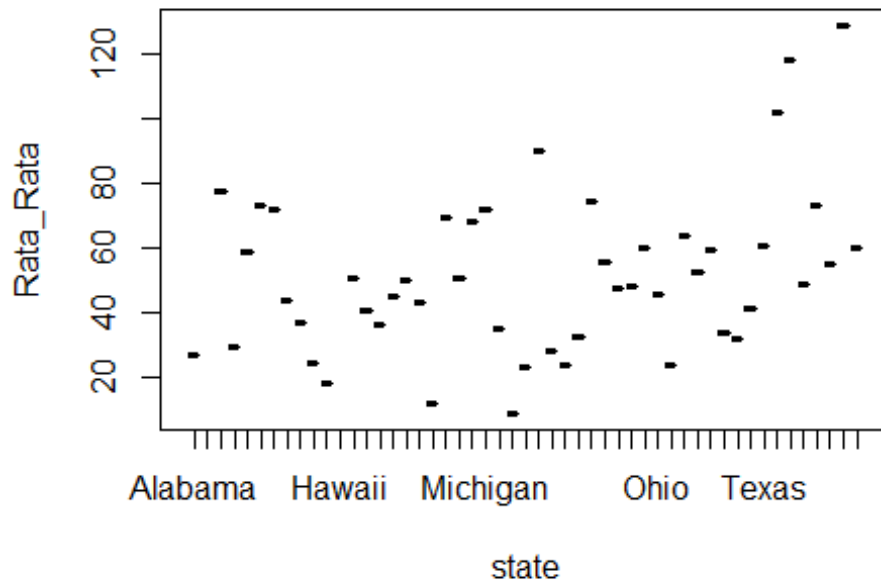
kesimpulan

## # A tibble: 51 x 2
##   state          Rata_Rata
##   <fct>          <dbl>
## 1 Alabama        27.3
## 2 Alaska          NA
## 3 Arizona        77.5
## 4 Arkansas        29.4
## 5 California      58.6
## 6 Colorado        73.2
## 7 Connecticut     72.1
## 8 Delaware        44.0
## 9 District Of Columbia 36.8
## 10 Florida        24.4
## # ... with 41 more rows

```

8. Lakukan visualisasi pada hasil nomor 7!(Bebas menggunakan plot, boxplot, hist, ggplot2 dsb) **point 5**

```
plot(kesimpulan)
```



## ##BAGIAN KEDUA

1. Load library tambahan untuk import file! **point 2**

```
library(readr)
library(ggplot2)
Covid19_Asean <- read_csv("Covid19_Asean.csv")

##
## -- Column specification -----
##
## cols(
##   Country = col_character(),
##   CountryCode = col_character(),
##   Confirmed = col_double(),
##   Recovered = col_double(),
##   Deaths = col_double()
## )
```

2. Tampilkan informasi rinci tentang struktur dataset yang digunakan! **point 5**

```
str(Covid19_Asean)

## tibble [10 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Country      : chr [1:10] "Indonesia" "Malaysia" "Myanmar" "Vietnam" ...
## $ CountryCode  : chr [1:10] "ID" "MY" "MM" "VN" ...
## $ Confirmed    : num [1:10] 2738 3963 22 249 2258 ...
## $ Recovered    : num [1:10] 204 1321 0 123 888 ...
## $ Deaths      : num [1:10] 221 63 1 0 27 0 0 6 177 0
```

```
## - attr(*, "spec")=
## .. cols(
## ..   Country = col_character(),
## ..   CountryCode = col_character(),
## ..   Confirmed = col_double(),
## ..   Recovered = col_double(),
## ..   Deaths = col_double()
## .. )
```

3. Tampilkan nama Negara dengan jumlah Penderita Covid-19 yang Terkonfirmasi dari paling banyak ke paling sedikit! **point 8**

```
Covid19_Asean %>%
  arrange(desc(Confirmed))

## # A tibble: 10 x 5
##   Country      CountryCode Confirmed Recovered Deaths
##   <chr>         <chr>         <dbl>     <dbl>   <dbl>
## 1 Malaysia     MY             3963      1321     63
## 2 Filipina     PH             3764        84    177
## 3 Indonesia    ID             2738       204    221
## 4 Thailand     TH             2258       888     27
## 5 Singapura    SG             1481       377      6
## 6 Vietnam      VN              249       123      0
## 7 Kamboja      KB              115        58      0
## 8 Myanmar      MM               22         0      1
## 9 Laos         LA               14         0      0
## 10 Timor Leste TL                1         0      0
```

4. Buat kolom baru bernama RateDeaths yang berisi rasio korban Covid-19 yang meninggal dengan yang terkonfirmasi! **point 11**

```
Covid19_Asean = mutate(Covid19_Asean, RateDeaths = Deaths / Confirmed)
```

5. Negara mana yang memiliki rasio kematian Covid-19 tertinggi dan terendah? Tampilkan nama negaranya. **point 11**

```
print("Tertinggi")

## [1] "Tertinggi"

filter(Covid19_Asean, RateDeaths == max(Covid19_Asean$RateDeaths))

## # A tibble: 1 x 6
##   Country      CountryCode Confirmed Recovered Deaths RateDeaths
##   <chr>         <chr>         <dbl>     <dbl>   <dbl>     <dbl>
## 1 Indonesia    ID             2738       204    221      0.0807

print("Terendah")

## [1] "Terendah"

filter(Covid19_Asean, RateDeaths == min(Covid19_Asean$RateDeaths))
```

```
## # A tibble: 4 x 6
##   Country      CountryCode Confirmed Recovered Deaths RateDeaths
##   <chr>        <chr>         <dbl>    <dbl>   <dbl>    <dbl>
## 1 Vietnam      VN             249      123     0         0
## 2 Laos         LA              14       0       0         0
## 3 Kamboja      KB             115      58       0         0
## 4 Timor Leste  TL              1       0       0         0
```

6. Tampilkan grafik plot antara penderita yang sembuh degan penderita yang terkonfirmasi Covid-19! **point 8**

```
Covid19_Asean %>%
  ggplot(aes(x = Recovered, y = Confirmed)) +
  geom_line()
```

