

Kuis_DS-B

Afifrendra Rifqi Nugraha_123190012

##Intro ##Baca Petunjuk Terlebih Dahulu!

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 12 jam, dikumpulkan maksimal Jum'at, 29 Oktober 2021 pukul 21.00 WIB
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 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr  0.3.4
## v tibble  3.1.4      v dplyr  1.0.7
## v tidyr   1.1.3      v stringr 1.4.0
## v readr   2.0.1      v forcats 0.5.1
```

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

```
library(dplyr)
data("us_contagious_diseases")
```

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

```
names(us_contagious_diseases)
```

```
## [1] "disease"      "state"        "year"         "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 rubella teratas diurutkan berdasarkan jumlah kasusnya dan terjadi pada antara tahun 2000 dan 2005! **point 11**

```
us_contagious_diseases %>% filter(disease=="Rubella",
                                year>=2000, year<=2005) %>%
  arrange(desc(count)) %>%
  top_n(10,count)
```

```
##   disease      state year weeks_reporting count population
## 1 Rubella North Carolina 2000           36      82    8049313
## 2 Rubella South Carolina 2000           40      14    4012012
## 3 Rubella California 2000           32      11   33871648
## 4 Rubella Massachusetts 2000           39       6   6349097
## 5 Rubella Texas 2000           34       5   20851820
## 6 Rubella Florida 2001           32       4   16272186
## 7 Rubella Alabama 2000           37       3   4447100
## 8 Rubella Illinois 2001           46       3   12501805
## 9 Rubella California 2002           29       2   34529758
## 10 Rubella Florida 2000           33       2   15982378
## 11 Rubella Georgia 2000            1       2    8186453
## 12 Rubella New Hampshire 2000          47       2   1235786
## 13 Rubella South Carolina 2001          25       2   4072049
```

5. Klasifikasikan data tersebut berdasarkan jumlah kasusnya dengan kondisi : -jumlah kasus kurang dari 500 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**

```
us_contagious_diseases %>%
  mutate(category = cut(count, c(0,500,2000,Inf),
                             labels = c("Biasa", "Cobaan", "Azab")))) %>%
  filter(category=="Biasa") %>% head()
```

```
##   disease      state year weeks_reporting count population category
## 1 Hepatitis A Alabama 1966           50     321    3345787 Biasa
## 2 Hepatitis A Alabama 1967           49     291    3364130 Biasa
## 3 Hepatitis A Alabama 1968           52     314    3386068 Biasa
## 4 Hepatitis A Alabama 1969           49     380    3412450 Biasa
## 5 Hepatitis A Alabama 1970           51     413    3444165 Biasa
## 6 Hepatitis A Alabama 1971           51     378    3481798 Biasa
```

```
us_contagious_diseases %>%
  mutate(category = cut(count, c(0,500,2000,Inf),
                             labels = c("Biasa", "Cobaan", "Azab")))) %>%
  filter(category=="Cobaan") %>% head()
```

```
##      disease  state year weeks_reporting count population category
## 1 Hepatitis A  Alaska 1976           13  1094      355341  Cobaan
## 2 Hepatitis A  Alaska 1977           39   551      365958  Cobaan
## 3 Hepatitis A  Arizona 1969           49   528      1712707  Cobaan
## 4 Hepatitis A  Arizona 1970           50   615      1770900  Cobaan
## 5 Hepatitis A  Arizona 1971           51   896      1837891  Cobaan
## 6 Hepatitis A  Arizona 1972           46   660      1913632  Cobaan
```

```
us_contagious_diseases %>%
  mutate(category = cut(count, c(0,500,2000,Inf),
                             labels = c("Biasa", "Cobaan", "Azab"))) %>%
  filter(category=="Azab") %>% head()
```

```
##      disease  state year weeks_reporting count population category
## 1 Hepatitis A  Arizona 1989           33  2009      3557380  Azab
## 2 Hepatitis A  California 1966          52  5933      18429575  Azab
## 3 Hepatitis A  California 1967          52  7480      18831882  Azab
## 4 Hepatitis A  California 1968          52 10821      19219725  Azab
## 5 Hepatitis A  California 1969          50  9051      19593348  Azab
## 6 Hepatitis A  California 1970          50  9422      19953134  Azab
```

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**

```
us_contagious_diseases<-us_contagious_diseases %>%
  mutate(category = cut(count, c(0,500,2000,Inf),
                           labels = c("Biasa", "Cobaan", "Azab")),
         rate = count/(population*100000))
us_contagious_diseases%>%head()
```

```
##      disease  state year weeks_reporting count population category
## 1 Hepatitis A  Alabama 1966           50   321      3345787  Biasa
## 2 Hepatitis A  Alabama 1967           49   291      3364130  Biasa
## 3 Hepatitis A  Alabama 1968           52   314      3386068  Biasa
## 4 Hepatitis A  Alabama 1969           49   380      3412450  Biasa
## 5 Hepatitis A  Alabama 1970           51   413      3444165  Biasa
## 6 Hepatitis A  Alabama 1971           51   378      3481798  Biasa
##      rate
## 1 9.594155e-10
## 2 8.650082e-10
## 3 9.273293e-10
## 4 1.113569e-09
## 5 1.199130e-09
## 6 1.085646e-09
```

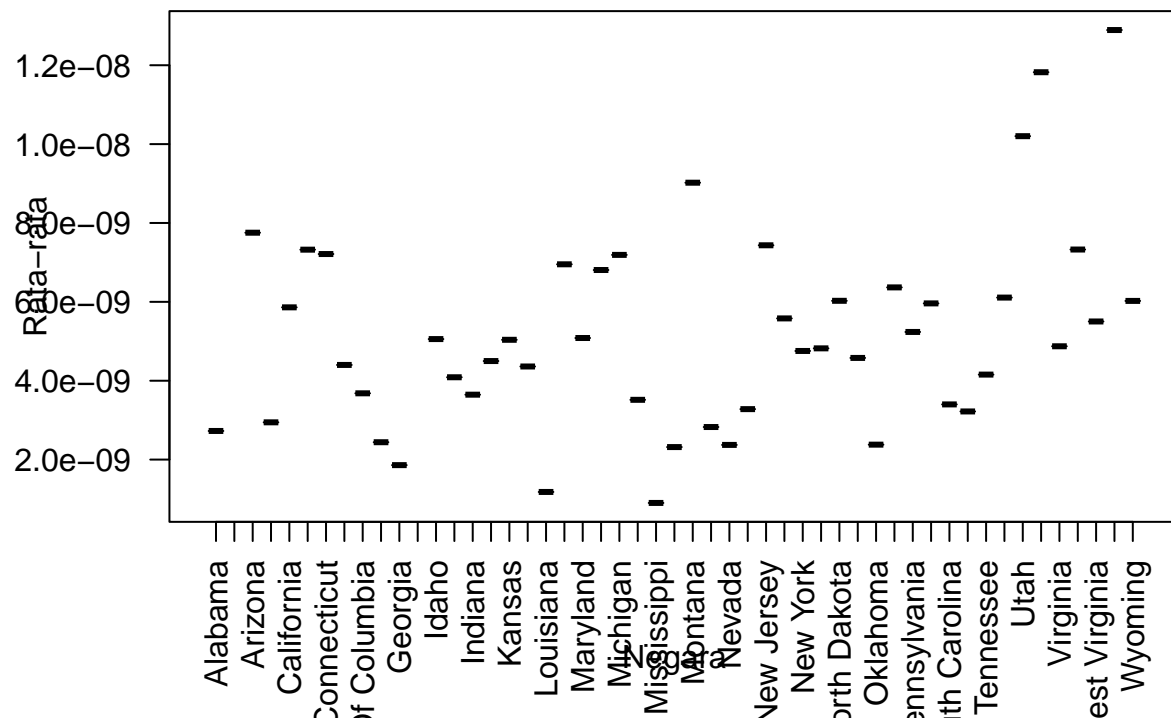
7. Tampilkan kesimpulan pada data nomor 6 dikelompokkan berdasarkan negara bagian yang isinya nama negara bagian dan rata-rata rate per negara bagian! **point 8**

```
us_contagious_diseases%>%group_by(state)%>%
  summarise(n=mean(rate))
```

```
## # A tibble: 51 x 2
##   state      n
##   <fct>    <dbl>
## 1 Alabama  2.73e-9
## 2 Alaska   NA
## 3 Arizona  7.75e-9
## 4 Arkansas 2.94e-9
## 5 California 5.86e-9
## 6 Colorado 7.32e-9
## 7 Connecticut 7.21e-9
## 8 Delaware 4.40e-9
## 9 District Of Columbia 3.68e-9
## 10 Florida 2.44e-9
## # ... with 41 more rows
```

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

```
plot(us_contagious_diseases%>%group_by(state)%>%
  summarise(n=mean(rate)),las=2,
  xlab = "Negara", ylab = "Rata-rata")
```



##BAGIAN KEDUA

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

```
library(readr)
worldwide_covid_data <- read_csv("worldwide_covid_data.csv")
```

```
## Rows: 196 Columns: 10
```

```
## -- Column specification -----
## Delimiter: ","
## chr (1): Country
## dbl (9): Total_Cases, Total_Deaths, Total_Recovered, Active_Cases, Total_Cas...

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

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

```
str(worldwide_covid_data)
```

```
## spec_tbl_df [196 x 10] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ Country          : chr [1:196] "Afghanistan" "Albania" "Algeria" "Andorra" ...
## $ Total_Cases      : num [1:196] 156071 182610 205990 15425 64033 ...
## $ Total_Deaths     : num [1:196] 7262 2888 5899 130 1702 ...
## $ Total_Recovered  : num [1:196] 128000 172464 141335 15205 52833 ...
## $ Active_Cases     : num [1:196] 20809 7258 58756 90 9498 ...
## $ Total_Cases_per1M_population: num [1:196] 3894 63546 4589 199217 1872 ...
## $ Deaths_per1M_population : num [1:196] 181 1005 131 1679 50 ...
## $ Total_Tests      : num [1:196] 771431 1289520 230861 193595 1092363 ...
## $ Tests_per1M_population : num [1:196] 19247 448738 5143 2500323 31933 ...
## $ Population       : num [1:196] 40080392 2873656 44892255 77428 34207984 ...
## - attr(*, "spec")=
## .. cols(
## .. Country = col_character(),
## .. Total_Cases = col_double(),
## .. Total_Deaths = col_double(),
## .. Total_Recovered = col_double(),
## .. Active_Cases = col_double(),
## .. Total_Cases_per1M_population = col_double(),
## .. Deaths_per1M_population = col_double(),
## .. Total_Tests = col_double(),
## .. Tests_per1M_population = col_double(),
## .. Population = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

3. Tampilkan 10 nama Negara dengan jumlah kasus Covid-19 yang terkonfirmasi dari paling banyak ke paling sedikit! **point 8**

```
worldwide_covid_data%>%
  select(Country, Total_Cases, Total_Deaths, Total_Recovered, Population)%>%
  top_n(10, Total_Cases) %>% arrange(desc(Total_Cases))
```

```
## # A tibble: 10 x 5
##   Country    Total_Cases Total_Deaths Total_Recovered Population
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 USA        46497719      759932      36375189  333558957
## 2 India      34215653      455684      33597339  1397864972
## 3 Brazil     21748984      606293      20944087  214549103
## 4 UK         8853227      139834      7198408   68355696
## 5 Russia     8316019      232775      7213584  146016918
## 6 Turkey     7909111       69559      7346279   85533610
## 7 France     7133766      117555      6921146   65463804
## 8 Iran       5877456      125519      5443243   85407642
## 9 Argentina  5283000      115866      5149181   45741769
## 10 Spain     5004143       87238      4859415   46778641
```

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

```
worldwide_covid_data <- worldwide_covid_data%>%mutate(
  Rate_Deaths = Total_Deaths/Total_Cases)
worldwide_covid_data %>%
  select(Country, Total_Cases, Total_Deaths, Total_Recovered, Rate_Deaths)
```

```
## # A tibble: 196 x 5
##   Country          Total_Cases Total_Deaths Total_Recovered Rate_Deaths
##   <chr>              <dbl>      <dbl>      <dbl>      <dbl>
## 1 Afghanistan      156071       7262      128000     0.0465
## 2 Albania           182610       2888      172464     0.0158
## 3 Algeria           205990       5899      141335     0.0286
## 4 Andorra           15425        130       15205     0.00843
## 5 Angola            64033       1702       52833     0.0266
## 6 Antigua and Barbuda 4031         99        3541     0.0246
## 7 Argentina         5283000     115866     5149181    0.0219
## 8 Armenia           298069       6112      263002     0.0205
## 9 Aruba             15848        171       15513     0.0108
## 10 Australia         163866       1669      130922     0.0102
## # ... with 186 more rows
```

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

```
worldwide_covid_data%>%
  select(Country, Rate_Deaths)%>%
  top_n(1, Rate_Deaths)
```

```
## # A tibble: 1 x 2
##   Country Rate_Deaths
##   <chr>      <dbl>
## 1 Yemen      0.191
```

```
worldwide_covid_data%>%
  select(Country, Rate_Deaths)%>%
  top_n(-1,Rate_Deaths)
```

```
## # A tibble: 1 x 2
##   Country Rate_Deaths
##   <chr>      <dbl>
## 1 Bhutan    0.00115
```

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

```
plot(log(worldwide_covid_data$Total_Recovered,10),
     log(worldwide_covid_data$Total_Cases,10), xlab="Sembuh(log10)",
     ylab="Kasus(log10)")
```

