Kuis\_DS-B

Rama Tri Agung\_123180053

## Intro Baca Petunjuk Terlebih Dahulu!

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

library(dslabs)  
library(ggplot2)  
data(us\_contagious\_diseases)

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

names(us\_contagious\_diseases)

## [1] "disease" "state" "year" "weeks\_reporting"  
## [5] "count" "population"

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

class(us\_contagious\_diseases$disease)

## [1] "factor"

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

us\_contagious\_diseases %>%  
 filter(disease == "Polio", year >= 1955, year <= 1965) %>%  
 arrange(desc(population)) %>%  
 top\_n(10)

## Selecting by population

## disease state year weeks\_reporting count population  
## 1 Polio California 1965 36 5 18012709  
## 2 Polio New York 1965 9 1 17743562  
## 3 Polio California 1964 25 2 17581336  
## 4 Polio New York 1964 39 10 17572556  
## 5 Polio New York 1963 48 12 17387034  
## 6 Polio New York 1962 48 88 17190958  
## 7 Polio California 1963 52 19 17135662  
## 8 Polio New York 1961 48 256 16988157  
## 9 Polio New York 1960 52 280 16782304  
## 10 Polio California 1962 50 88 16676046

1. 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**

Biasa <- filter(us\_contagious\_diseases, count < 1000)  
Biasa %>% head()

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

Azab <- filter(us\_contagious\_diseases, count > 2000)  
Azab %>% head()

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

Cobaan <- filter(us\_contagious\_diseases, count >= 1000 & count <= 2000)  
Cobaan %>% head()

## disease state year weeks\_reporting count population  
## 1 Hepatitis A Alaska 1976 13 1094 355341  
## 2 Hepatitis A Arizona 1978 48 1137 2504256  
## 3 Hepatitis A Arizona 1979 50 1788 2612561  
## 4 Hepatitis A Arizona 1985 42 1013 3185267  
## 5 Hepatitis A Arizona 1986 37 1015 3273618  
## 6 Hepatitis A Arizona 1987 47 1925 3363792

1. 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 = ifelse(count < 1000, "Biasa", ifelse(count > 2000, "Azab", "Cobaan")),  
 rate = (count / population) \* 10^5  
 )  
  
filter(us\_contagious\_diseases, category == "Cobaan") %>% head()

## disease state year weeks\_reporting count population category rate  
## 1 Hepatitis A Alaska 1976 13 1094 355341 Cobaan 307.87328  
## 2 Hepatitis A Arizona 1978 48 1137 2504256 Cobaan 45.40271  
## 3 Hepatitis A Arizona 1979 50 1788 2612561 Cobaan 68.43859  
## 4 Hepatitis A Arizona 1985 42 1013 3185267 Cobaan 31.80267  
## 5 Hepatitis A Arizona 1986 37 1015 3273618 Cobaan 31.00545  
## 6 Hepatitis A Arizona 1987 47 1925 3363792 Cobaan 57.22708

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

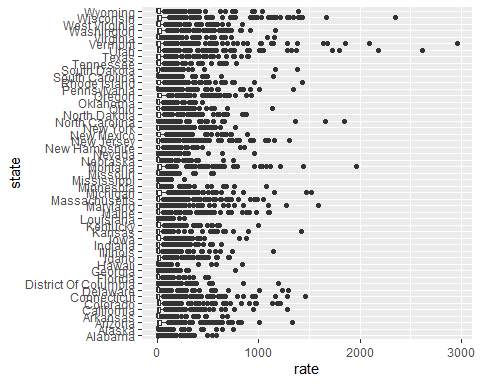
us\_contagious\_diseases %>%  
 select(state, rate) %>%  
 summary()

## state rate   
## Alabama : 315 Min. : 0.0000   
## Alaska : 315 1st Qu.: 0.3354   
## Arizona : 315 Median : 2.6018   
## Arkansas : 315 Mean : 50.9645   
## California: 315 3rd Qu.: 20.4358   
## Colorado : 315 Max. :2964.4269   
## (Other) :14175 NA's :214

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

ggplot(us\_contagious\_diseases, aes(x = rate, y = state)) + geom\_boxplot()

## Warning: Removed 214 rows containing non-finite values (stat\_boxplot).



## BAGIAN KEDUA

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

Covid19\_Asean <- read.csv(file = "Covid19\_Asean.csv")  
Covid19\_Asean

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

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

str(Covid19\_Asean)

## 'data.frame': 10 obs. of 5 variables:  
## $ Country : chr "Indonesia" "Malaysia" "Myanmar" "Vietnam" ...  
## $ CountryCode: chr "ID" "MY" "MM" "VN" ...  
## $ Confirmed : int 2738 3963 22 249 2258 14 115 1481 3764 1  
## $ Recovered : int 204 1321 0 123 888 0 58 377 84 0  
## $ Deaths : int 221 63 1 0 27 0 0 6 177 0

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

Covid19\_Asean %>%  
 arrange(desc(Confirmed)) %>%  
 select(Country)

## Country  
## 1 Malaysia  
## 2 Filipina  
## 3 Indonesia  
## 4 Thailand  
## 5 Singapura  
## 6 Vietnam  
## 7 Kamboja  
## 8 Myanmar  
## 9 Laos  
## 10 Timor Leste

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

Covid19\_Asean <- Covid19\_Asean %>% mutate(RateDeaths = Deaths / Confirmed \* 100)  
Covid19\_Asean

## Country CountryCode Confirmed Recovered Deaths RateDeaths  
## 1 Indonesia ID 2738 204 221 8.0715851  
## 2 Malaysia MY 3963 1321 63 1.5897048  
## 3 Myanmar MM 22 0 1 4.5454545  
## 4 Vietnam VN 249 123 0 0.0000000  
## 5 Thailand TH 2258 888 27 1.1957484  
## 6 Laos LA 14 0 0 0.0000000  
## 7 Kamboja KB 115 58 0 0.0000000  
## 8 Singapura SG 1481 377 6 0.4051317  
## 9 Filipina PH 3764 84 177 4.7024442  
## 10 Timor Leste TL 1 0 0 0.0000000

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

Covid19\_Asean %>%  
 filter(RateDeaths == max(RateDeaths)) %>%  
 select(Country)

## Country  
## 1 Indonesia

Covid19\_Asean %>%  
 filter(RateDeaths == min(RateDeaths)) %>%  
 select(Country)

## Country  
## 1 Vietnam  
## 2 Laos  
## 3 Kamboja  
## 4 Timor Leste

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

ggplot(Covid19\_Asean, aes(x = Recovered, y = Confirmed)) + geom\_step()

