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")

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

str(us\_contagious\_diseases)

## 'data.frame': 16065 obs. of 6 variables:  
## $ disease : Factor w/ 7 levels "Hepatitis A",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ state : Factor w/ 51 levels "Alabama","Alaska",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ year : num 1966 1967 1968 1969 1970 ...  
## $ weeks\_reporting: num 50 49 52 49 51 51 45 45 45 46 ...  
## $ count : num 321 291 314 380 413 378 342 467 244 286 ...  
## $ population : num 3345787 3364130 3386068 3412450 3444165 ...

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

class(us\_contagious\_diseases$disease)

## [1] "factor"

1. 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) %>% top\_n(10,count) %>% arrange(desc(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

1. 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

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 = 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

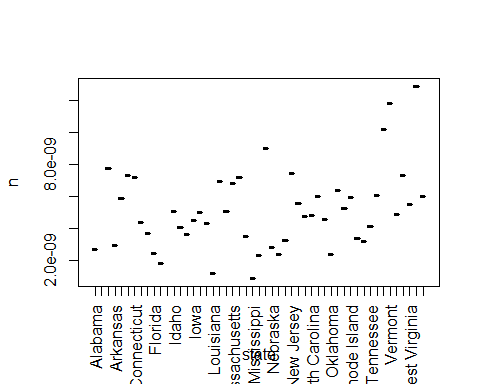
1. Tampilkan kesimpulan pada data nomor 6 dikelompokkan berdasarkan negara bagian yang isinya nama negara bagian dan rata-rata rate per negara bagian! **point 8**

us\_contagious\_diseases%>%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

1. 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=3)



##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.

1. 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>

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

worldwide\_covid\_data%>%top\_n(10, Total\_Cases) %>% arrange(desc(Total\_Cases))

## # A tibble: 10 x 10  
## Country Total\_Cases Total\_Deaths Total\_Recovered Active\_Cases  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 USA 46497719 759932 36375189 9362598  
## 2 India 34215653 455684 33597339 162630  
## 3 Brazil 21748984 606293 20944087 198604  
## 4 UK 8853227 139834 7198408 1514985  
## 5 Russia 8316019 232775 7213584 869660  
## 6 Turkey 7909111 69559 7346279 493273  
## 7 France 7133766 117555 6921146 95065  
## 8 Iran 5877456 125519 5443243 308694  
## 9 Argentina 5283000 115866 5149181 17953  
## 10 Spain 5004143 87238 4859415 57490  
## # ... with 5 more variables: Total\_Cases\_per1M\_population <dbl>,  
## # Deaths\_per1M\_population <dbl>, Total\_Tests <dbl>,  
## # Tests\_per1M\_population <dbl>, Population <dbl>

1. 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

## # A tibble: 196 x 11  
## Country Total\_Cases Total\_Deaths Total\_Recovered Active\_Cases  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Afghanistan 156071 7262 128000 20809  
## 2 Albania 182610 2888 172464 7258  
## 3 Algeria 205990 5899 141335 58756  
## 4 Andorra 15425 130 15205 90  
## 5 Angola 64033 1702 52833 9498  
## 6 Antigua and Barbuda 4031 99 3541 391  
## 7 Argentina 5283000 115866 5149181 17953  
## 8 Armenia 298069 6112 263002 28955  
## 9 Aruba 15848 171 15513 164  
## 10 Australia 163866 1669 130922 31275  
## # ... with 186 more rows, and 6 more variables:  
## # Total\_Cases\_per1M\_population <dbl>, Deaths\_per1M\_population <dbl>,  
## # Total\_Tests <dbl>, Tests\_per1M\_population <dbl>, Population <dbl>,  
## # Rate\_Deaths <dbl>

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

worldwide\_covid\_data%>%top\_n(1, Rate\_Deaths)

## # A tibble: 1 x 11  
## Country Total\_Cases Total\_Deaths Total\_Recovered Active\_Cases Total\_Cases\_per~  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Yemen 9711 1858 6309 1544 316  
## # ... with 5 more variables: Deaths\_per1M\_population <dbl>, Total\_Tests <dbl>,  
## # Tests\_per1M\_population <dbl>, Population <dbl>, Rate\_Deaths <dbl>

worldwide\_covid\_data%>%top\_n(-1,Rate\_Deaths)

## # A tibble: 1 x 11  
## Country Total\_Cases Total\_Deaths Total\_Recovered Active\_Cases Total\_Cases\_per~  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Bhutan 2617 3 2610 4 3343  
## # ... with 5 more variables: Deaths\_per1M\_population <dbl>, Total\_Tests <dbl>,  
## # Tests\_per1M\_population <dbl>, Population <dbl>, Rate\_Deaths <dbl>

1. 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), xlab="Sembuh(log10)",   
 ylab="Kasus(log10)")

