

Kuis_DS-F

DEMITRIES BASKHARA RIVALDO TOLLA_123180137

##Intro

0. Cuci tangan dengan sabun hingga benar-benar bersih dengan durasi mencuci tangan kurang lebih 20 dtk
1. Kerjakan Soal-soal yang ada! Jangan lupa author dinamai!
2. Kuis terdiri dari 2 bagian yaitu bagian pertama dan bagian kedua
3. Jawab dengan membuat chunk dibawah soal!
4. Durasi pengerjaan 3 hari sampai tanggal 13 Desember 2020
5. No toleransi pengumpulan telat, 11 jam loh bos telat ngisin-ngisini. Ingat, telat kemungkinan terburuk ga ada nilai kuis!
6. Misal soal rancu bisa menghubungi asisten terkait
7. Selamat memutus rantai gabut :) jangan lupa jaga kesehatan

##Bagian Pertama

1. Load library apa saja yang kira-kira digunakan! Lalu gunakan data 'us_contagious_diseases'! **point 1**

```
library(dslabs)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(data.table)
```

```
##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last
```

```
library(ggplot2)
data <- us_contagious_diseases
```

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

```
names(data)
```

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

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

```
class(data$disease)
```

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

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

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

```
##      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
```

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**

```
data %>%
  mutate(Klasifikasi = case_when(count < 1000 ~ "Biasa" , count > 2000 ~ "Azab" , TRUE ~ "Cobaan")) %>%
  head(n = 150)
```

##	disease	state	year	weeks_reporting	count	population	Klasifikasi
## 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
## 7	Hepatitis A	Alabama	1972	45	342	3524543	Biasa
## 8	Hepatitis A	Alabama	1973	45	467	3571209	Biasa
## 9	Hepatitis A	Alabama	1974	45	244	3620548	Biasa
## 10	Hepatitis A	Alabama	1975	46	286	3671246	Biasa
## 11	Hepatitis A	Alabama	1976	50	220	3721914	Biasa
## 12	Hepatitis A	Alabama	1977	43	206	3771085	Biasa
## 13	Hepatitis A	Alabama	1978	41	203	3817217	Biasa
## 14	Hepatitis A	Alabama	1979	47	257	3858703	Biasa
## 15	Hepatitis A	Alabama	1980	37	200	3893888	Biasa
## 16	Hepatitis A	Alabama	1981	40	138	3921581	Biasa
## 17	Hepatitis A	Alabama	1982	43	128	3942588	Biasa
## 18	Hepatitis A	Alabama	1983	43	178	3958271	Biasa
## 19	Hepatitis A	Alabama	1984	35	115	3970041	Biasa
## 20	Hepatitis A	Alabama	1985	22	37	3979341	Biasa
## 21	Hepatitis A	Alabama	1986	26	51	3987637	Biasa
## 22	Hepatitis A	Alabama	1987	36	81	3996407	Biasa
## 23	Hepatitis A	Alabama	1988	20	13	4007144	Biasa
## 24	Hepatitis A	Alabama	1989	32	56	4021358	Biasa
## 25	Hepatitis A	Alabama	1990	41	86	4040587	Biasa
## 26	Hepatitis A	Alabama	1991	38	39	4066003	Biasa
## 27	Hepatitis A	Alabama	1992	34	35	4097169	Biasa
## 28	Hepatitis A	Alabama	1993	27	40	4133242	Biasa
## 29	Hepatitis A	Alabama	1994	29	72	4173361	Biasa
## 30	Hepatitis A	Alabama	1995	38	75	4216645	Biasa
## 31	Hepatitis A	Alabama	1996	44	196	4262177	Biasa
## 32	Hepatitis A	Alabama	1997	26	40	4309007	Biasa
## 33	Hepatitis A	Alabama	1998	34	64	4356137	Biasa
## 34	Hepatitis A	Alabama	1999	35	52	4402529	Biasa
## 35	Hepatitis A	Alabama	2000	42	53	4447100	Biasa
## 36	Hepatitis A	Alabama	2001	41	66	4488959	Biasa
## 37	Hepatitis A	Alabama	2002	47	35	4528129	Biasa
## 38	Hepatitis A	Alabama	2003	47	14	4564878	Biasa
## 39	Hepatitis A	Alabama	2004	41	6	4599490	Biasa
## 40	Hepatitis A	Alabama	2005	47	34	4632262	Biasa
## 41	Hepatitis A	Alabama	2006	42	17	4663504	Biasa
## 42	Hepatitis A	Alabama	2007	48	20	4693536	Biasa
## 43	Hepatitis A	Alabama	2008	49	10	4722691	Biasa
## 44	Hepatitis A	Alabama	2009	49	10	4751307	Biasa
## 45	Hepatitis A	Alabama	2010	51	8	4779736	Biasa
## 46	Hepatitis A	Alabama	2011	27	7	4808274	Biasa
## 47	Hepatitis A	Alaska	1966	34	90	268476	Biasa
## 48	Hepatitis A	Alaska	1967	22	39	276180	Biasa
## 49	Hepatitis A	Alaska	1968	31	84	284068	Biasa
## 50	Hepatitis A	Alaska	1969	32	63	292136	Biasa
## 51	Hepatitis A	Alaska	1970	32	113	300382	Biasa
## 52	Hepatitis A	Alaska	1971	28	96	308810	Biasa
## 53	Hepatitis A	Alaska	1972	37	199	317458	Biasa

## 54	Hepatitis A	Alaska	1973	32	225	326376	Biasa
## 55	Hepatitis A	Alaska	1974	30	248	335621	Biasa
## 56	Hepatitis A	Alaska	1975	37	409	345253	Biasa
## 57	Hepatitis A	Alaska	1976	13	1094	355341	Cobaan
## 58	Hepatitis A	Alaska	1977	39	551	365958	Biasa
## 59	Hepatitis A	Alaska	1978	39	125	377187	Biasa
## 60	Hepatitis A	Alaska	1979	28	61	389118	Biasa
## 61	Hepatitis A	Alaska	1980	14	18	401851	Biasa
## 62	Hepatitis A	Alaska	1981	22	45	415458	Biasa
## 63	Hepatitis A	Alaska	1982	15	17	429850	Biasa
## 64	Hepatitis A	Alaska	1983	22	43	444879	Biasa
## 65	Hepatitis A	Alaska	1984	11	14	460375	Biasa
## 66	Hepatitis A	Alaska	1985	12	18	476138	Biasa
## 67	Hepatitis A	Alaska	1986	21	66	491942	Biasa
## 68	Hepatitis A	Alaska	1987	47	240	507531	Biasa
## 69	Hepatitis A	Alaska	1988	22	268	522622	Biasa
## 70	Hepatitis A	Alaska	1989	33	494	536903	Biasa
## 71	Hepatitis A	Alaska	1990	39	167	550043	Biasa
## 72	Hepatitis A	Alaska	1991	37	80	561769	Biasa
## 73	Hepatitis A	Alaska	1992	32	65	572122	Biasa
## 74	Hepatitis A	Alaska	1993	24	374	581250	Biasa
## 75	Hepatitis A	Alaska	1994	29	163	589317	Biasa
## 76	Hepatitis A	Alaska	1995	39	38	596513	Biasa
## 77	Hepatitis A	Alaska	1996	41	46	603039	Biasa
## 78	Hepatitis A	Alaska	1997	24	20	609112	Biasa
## 79	Hepatitis A	Alaska	1998	30	6	614959	Biasa
## 80	Hepatitis A	Alaska	1999	36	7	620817	Biasa
## 81	Hepatitis A	Alaska	2000	39	8	626932	Biasa
## 82	Hepatitis A	Alaska	2001	40	10	633510	Biasa
## 83	Hepatitis A	Alaska	2002	42	11	640570	Biasa
## 84	Hepatitis A	Alaska	2003	44	8	648078	Biasa
## 85	Hepatitis A	Alaska	2004	43	4	656003	Biasa
## 86	Hepatitis A	Alaska	2005	44	3	664309	Biasa
## 87	Hepatitis A	Alaska	2006	0	0	672961	Biasa
## 88	Hepatitis A	Alaska	2007	42	3	681920	Biasa
## 89	Hepatitis A	Alaska	2008	41	3	691147	Biasa
## 90	Hepatitis A	Alaska	2009	46	6	700599	Biasa
## 91	Hepatitis A	Alaska	2010	24	4	710231	Biasa
## 92	Hepatitis A	Alaska	2011	38	1	720004	Biasa
## 93	Hepatitis A	Arizona	1966	50	415	1572291	Biasa
## 94	Hepatitis A	Arizona	1967	48	390	1615310	Biasa
## 95	Hepatitis A	Arizona	1968	50	425	1661517	Biasa
## 96	Hepatitis A	Arizona	1969	49	528	1712707	Biasa
## 97	Hepatitis A	Arizona	1970	50	615	1770900	Biasa
## 98	Hepatitis A	Arizona	1971	51	896	1837891	Biasa
## 99	Hepatitis A	Arizona	1972	46	660	1913632	Biasa
## 100	Hepatitis A	Arizona	1973	22	160	1997501	Biasa
## 101	Hepatitis A	Arizona	1974	49	512	2088732	Biasa
## 102	Hepatitis A	Arizona	1975	50	437	2186365	Biasa
## 103	Hepatitis A	Arizona	1976	51	560	2289201	Biasa
## 104	Hepatitis A	Arizona	1977	51	929	2395761	Biasa
## 105	Hepatitis A	Arizona	1978	48	1137	2504256	Cobaan
## 106	Hepatitis A	Arizona	1979	50	1788	2612561	Cobaan
## 107	Hepatitis A	Arizona	1980	33	635	2718215	Biasa

## 108	Hepatitis A	Arizona	1981	43	434	2819039	Biasa
## 109	Hepatitis A	Arizona	1982	41	514	2915148	Biasa
## 110	Hepatitis A	Arizona	1983	44	845	3007444	Biasa
## 111	Hepatitis A	Arizona	1984	46	721	3097045	Biasa
## 112	Hepatitis A	Arizona	1985	42	1013	3185267	Cobaan
## 113	Hepatitis A	Arizona	1986	37	1015	3273618	Cobaan
## 114	Hepatitis A	Arizona	1987	47	1925	3363792	Cobaan
## 115	Hepatitis A	Arizona	1988	23	972	3457679	Biasa
## 116	Hepatitis A	Arizona	1989	33	2009	3557380	Azab
## 117	Hepatitis A	Arizona	1990	42	1645	3665228	Cobaan
## 118	Hepatitis A	Arizona	1991	37	837	3783229	Biasa
## 119	Hepatitis A	Arizona	1992	35	761	3911122	Biasa
## 120	Hepatitis A	Arizona	1993	26	662	4047937	Biasa
## 121	Hepatitis A	Arizona	1994	29	934	4192532	Biasa
## 122	Hepatitis A	Arizona	1995	40	990	4343559	Biasa
## 123	Hepatitis A	Arizona	1996	43	1397	4499438	Cobaan
## 124	Hepatitis A	Arizona	1997	27	1078	4658333	Cobaan
## 125	Hepatitis A	Arizona	1998	35	1224	4818136	Cobaan
## 126	Hepatitis A	Arizona	1999	36	601	4976457	Biasa
## 127	Hepatitis A	Arizona	2000	41	399	5130632	Biasa
## 128	Hepatitis A	Arizona	2001	40	341	5278300	Biasa
## 129	Hepatitis A	Arizona	2002	46	313	5419266	Biasa
## 130	Hepatitis A	Arizona	2003	44	237	5554021	Biasa
## 131	Hepatitis A	Arizona	2004	44	210	5683195	Biasa
## 132	Hepatitis A	Arizona	2005	47	240	5807553	Biasa
## 133	Hepatitis A	Arizona	2006	44	69	5927979	Biasa
## 134	Hepatitis A	Arizona	2007	51	163	6045475	Biasa
## 135	Hepatitis A	Arizona	2008	47	101	6161152	Biasa
## 136	Hepatitis A	Arizona	2009	48	72	6276227	Biasa
## 137	Hepatitis A	Arizona	2010	48	72	6392017	Biasa
## 138	Hepatitis A	Arizona	2011	48	19	6509700	Biasa
## 139	Hepatitis A	Arkansas	1966	52	299	1826028	Biasa
## 140	Hepatitis A	Arkansas	1967	47	295	1844745	Biasa
## 141	Hepatitis A	Arkansas	1968	38	166	1867112	Biasa
## 142	Hepatitis A	Arkansas	1969	34	189	1893248	Biasa
## 143	Hepatitis A	Arkansas	1970	42	212	1923295	Biasa
## 144	Hepatitis A	Arkansas	1971	46	236	1957252	Biasa
## 145	Hepatitis A	Arkansas	1972	46	278	1994452	Biasa
## 146	Hepatitis A	Arkansas	1973	44	141	2034009	Biasa
## 147	Hepatitis A	Arkansas	1974	49	444	2074967	Biasa
## 148	Hepatitis A	Arkansas	1975	49	409	2116289	Biasa
## 149	Hepatitis A	Arkansas	1976	50	449	2156856	Biasa
## 150	Hepatitis A	Arkansas	1977	47	406	2195463	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**

```
data_baru <- data
data_baru <- mutate(data_baru , Klasifikasi = case_when(count < 1000 ~ "Biasa" , count > 2000 ~ "Azab")
data_baru <- mutate(data_baru , rate = count / population * 10 ^ 5)
head(data_baru , n = 10)
```

```
##      disease  state year weeks_reporting count population Klasifikasi
```

```
## 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
## 7 Hepatitis A Alabama 1972      45   342   3524543   Biasa
## 8 Hepatitis A Alabama 1973      45   467   3571209   Biasa
## 9 Hepatitis A Alabama 1974      45   244   3620548   Biasa
## 10 Hepatitis A Alabama 1975      46   286   3671246   Biasa
##      rate
## 1    9.594155
## 2    8.650082
## 3    9.273293
## 4   11.135694
## 5   11.991295
## 6   10.856460
## 7    9.703386
## 8   13.076804
## 9    6.739311
## 10   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**

```
db.dt <- data.table(data_baru)
db.dt[,list(rata_rata=mean(rate , na.rm = TRUE)), by=state]
```

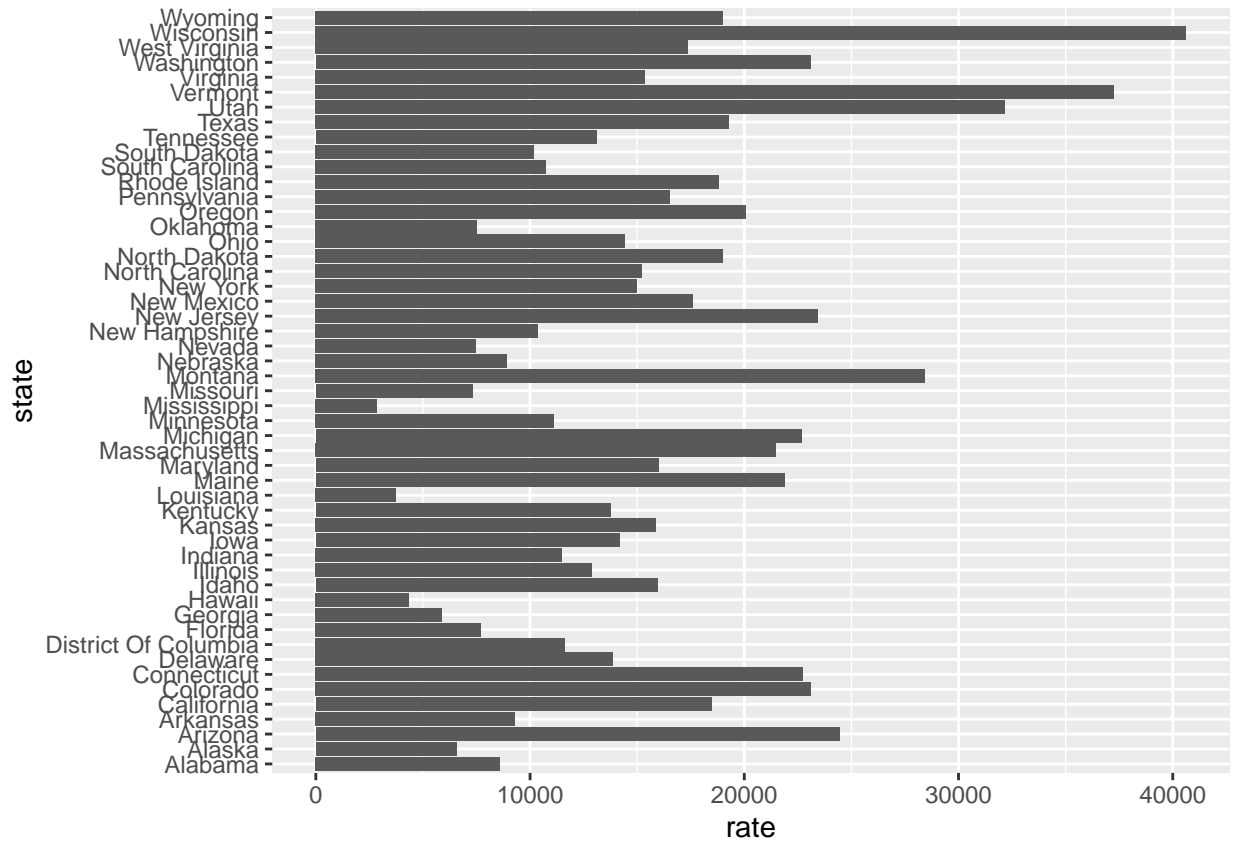
```
##      state rata_rata
## 1:      Alabama 27.254305
## 2:      Alaska 31.547713
## 3:      Arizona 77.532421
## 4:      Arkansas 29.449405
## 5:      California 58.608905
## 6:      Colorado 73.236255
## 7:      Connecticut 72.122288
## 8:      Delaware 43.998665
## 9: District Of Columbia 36.807179
## 10:      Florida 24.401025
## 11:      Georgia 18.593800
## 12:      Hawaii 20.889977
## 13:      Idaho 50.561081
## 14:      Illinois 40.876398
## 15:      Indiana 36.462964
## 16:      Iowa 44.990709
## 17:      Kansas 50.374806
## 18:      Kentucky 43.608362
## 19:      Louisiana 11.792634
## 20:      Maine 69.518367
## 21:      Maryland 50.827282
## 22:      Massachusetts 68.075267
## 23:      Michigan 71.909901
## 24:      Minnesota 35.156515
```

```
## 25:      Mississippi  9.007752
## 26:      Missouri   23.173617
## 27:      Montana    90.215212
## 28:      Nebraska   28.259611
## 29:      Nevada     23.717947
## 30:      New Hampshire 32.791969
## 31:      New Jersey  74.344977
## 32:      New Mexico  55.797871
## 33:      New York    47.540565
## 34:      North Carolina 48.203265
## 35:      North Dakota 60.266706
## 36:      Ohio        45.778918
## 37:      Oklahoma    23.786310
## 38:      Oregon      63.664772
## 39:      Pennsylvania 52.368000
## 40:      Rhode Island 59.607050
## 41:      South Carolina 33.989794
## 42:      South Dakota 32.206440
## 43:      Tennessee   41.553826
## 44:      Texas       61.079722
## 45:      Utah        102.012201
## 46:      Vermont     118.227946
## 47:      Virginia    48.730856
## 48:      Washington   73.265429
## 49:      West Virginia 55.024967
## 50:      Wisconsin   128.925467
## 51:      Wyoming     60.239481
##              state   rata_rata
```

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

```
ggplot(data_baru, aes(rate, state)) + geom_bar(stat = "identity")
```

```
## Warning: Removed 214 rows containing missing values (position_stack).
```



##Bagian Kedua

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

```
library(dslabs)
library(dplyr)
covid <- read.csv("Covid19_Asean.csv" , sep = "," , header = TRUE)
```

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

```
str(covid)
```

```
## '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
```

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


```
covid %>%
  mutate(rank = rank(-Confirmed)) %>%
  arrange(rank) %>%
  select(Country , rank)
```

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

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

```
covid %>%
  mutate(RateDeath = Deaths / Confirmed)
```

```
##      Country CountryCode Confirmed Recovered Deaths RateDeath
## 1    Indonesia         ID    2738      204     221 0.080715851
## 2    Malaysia         MY    3963     1321      63 0.015897048
## 3    Myanmar          MM      22         0       1 0.045454545
## 4    Vietnam          VN     249      123       0 0.000000000
## 5    Thailand         TH    2258     888      27 0.011957484
## 6     Laos           LA      14         0       0 0.000000000
## 7    Kamboja         KB     115         58       0 0.000000000
## 8    Singapura        SG    1481     377       6 0.004051317
## 9    Filipina         PH    3764       84     177 0.047024442
## 10 Timor Leste        TL        1         0       0 0.000000000
```

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

```
covid$Country[which.max(covid$Deaths / covid$Confirmed)]
```

```
## [1] "Indonesia"
```

```
covid$Country[which.min(covid$Deaths / covid$Confirmed)]
```

```
## [1] "Vietnam"
```

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

```
plot(covid$Recovered[order(covid$Recovered)] , covid$Confirmed[order(covid$Confirmed)] , type = "o")
```

