

Kuis_DS-B

Rivanka Desya fawwaz Assyiraj_123190018

##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(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(dslabs)
data("us_contagious_diseases")
```

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

```
colnames(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") %>% filter(year >= 2000 & year <= 2005) %>% arr
```

```
##      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
```

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**

```
head(iffelse(us_contagious_diseases$count<500, "Biasa", iffelse(us_contagious_diseases$count>2000, "Azab"
```

```
## [1] "Biasa" "Biasa" "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**

```
category = category = iffelse(us_contagious_diseases$count < 500 , "Biasa", iffelse(us_contagious_diseases
newdf = mutate(us_contagious_diseases,category , us_contagious_diseases, rate =((count/population)*1000
head(newdf, n=100)
```

```
##      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
## 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	Cobaan
## 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	Cobaan
## 97	Hepatitis A	Arizona	1970	50	615	1770900	Cobaan
## 98	Hepatitis A	Arizona	1971	51	896	1837891	Cobaan
## 99	Hepatitis A	Arizona	1972	46	660	1913632	Cobaan
## 100	Hepatitis A	Arizona	1973	22	160	1997501	Biasa
##	rate						
## 1	9.5941553						
## 2	8.6500819						
## 3	9.2732928						
## 4	11.1356943						
## 5	11.9912954						
## 6	10.8564598						
## 7	9.7033857						
## 8	13.0768040						
## 9	6.7393113						
## 10	7.7902707						
## 11	5.9109372						
## 12	5.4626188						
## 13	5.3180105						
## 14	6.6602690						
## 15	5.1362546						

## 16	3.5189889
## 17	3.2465984
## 18	4.4969129
## 19	2.8966955
## 20	0.9298022
## 21	1.2789529
## 22	2.0268206
## 23	0.3244206
## 24	1.3925644
## 25	2.1284036
## 26	0.9591729
## 27	0.8542484
## 28	0.9677633
## 29	1.7252282
## 30	1.7786653
## 31	4.5985889
## 32	0.9282881
## 33	1.4691916
## 34	1.1811393
## 35	1.1917879
## 36	1.4702741
## 37	0.7729462
## 38	0.3066895
## 39	0.1304492
## 40	0.7339827
## 41	0.3645328
## 42	0.4261180
## 43	0.2117437
## 44	0.2104684
## 45	0.1673733
## 46	0.1455824
## 47	33.5225495
## 48	14.1212253
## 49	29.5703846
## 50	21.5652984
## 51	37.6187654
## 52	31.0870762
## 53	62.6854576
## 54	68.9388926
## 55	73.8928732
## 56	118.4638511
## 57	307.8732823
## 58	150.5637259
## 59	33.1400605
## 60	15.6764786
## 61	4.4792722
## 62	10.8314198
## 63	3.9548680
## 64	9.6655495
## 65	3.0409992
## 66	3.7804166
## 67	13.4162157
## 68	47.2877519
## 69	51.2798925

```
## 70 92.0091711
## 71 30.3612627
## 72 14.2407288
## 73 11.3612132
## 74 64.3440860
## 75 27.6591376
## 76 6.3703557
## 77 7.6280307
## 78 3.2834684
## 79 0.9756748
## 80 1.1275464
## 81 1.2760555
## 82 1.5785070
## 83 1.7172206
## 84 1.2344193
## 85 0.6097533
## 86 0.4515971
## 87 0.0000000
## 88 0.4399343
## 89 0.4340611
## 90 0.8564100
## 91 0.5631970
## 92 0.1388881
## 93 26.3946051
## 94 24.1439724
## 95 25.5790341
## 96 30.8283904
## 97 34.7281044
## 98 48.7515310
## 99 34.4893898
## 100 8.0100085
```

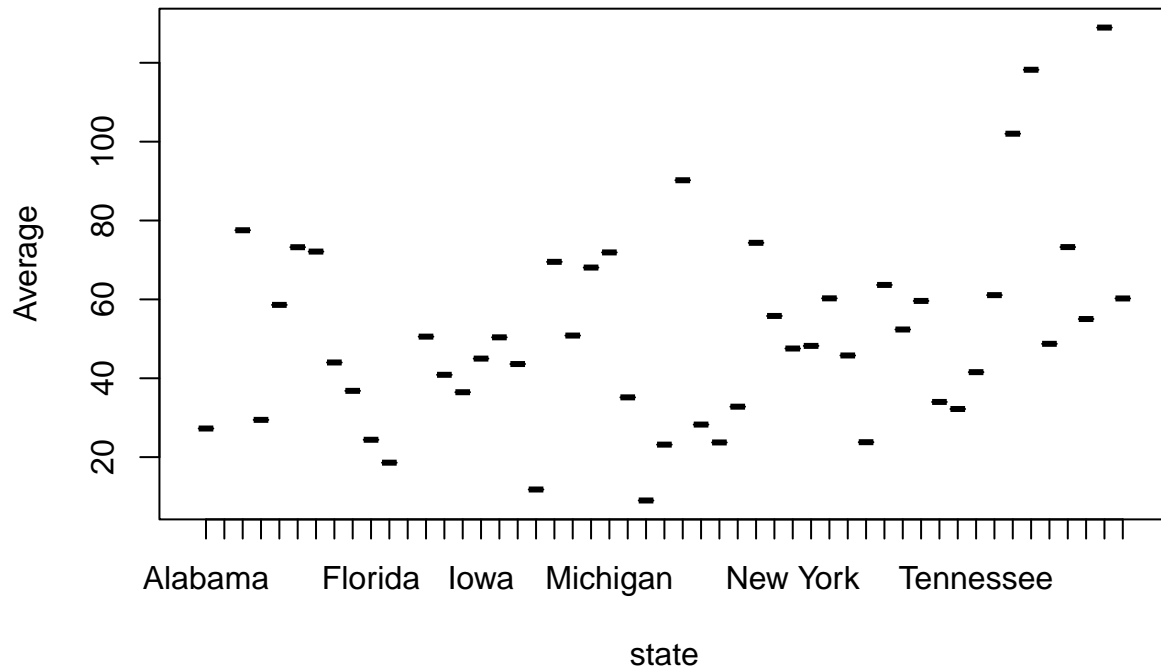
7. Tampilkan kesimpulan pada data nomor 6 dikelompokkan berdasarkan negara bagian yang isinya nama negara bagian dan rata-rata rate per negara bagian! **point 8**

```
x <- newdf %>%
  group_by(state) %>%
  summarize(Average = mean(rate))
print(x)
```

```
## # A tibble: 51 x 2
##   state      Average
##   <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(x)
```



##BAGIAN KEDUA

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

```
library(readr)
```

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

```
iris_csv <- read.csv("D:/worldwide_covid_data.csv")
str(iris_csv)
```

```
## 'data.frame':   196 obs. of  10 variables:
##  $ Country      : chr  "Afghanistan" "Albania" "Algeria" "Andorra" ...
##  $ Total_Cases   : int   156071 182610 205990 15425 64033 4031 5283000 298069 15848 163
##  $ Total_Deaths  : int    7262 2888 5899 130 1702 99 115866 6112 171 1669 ...
##  $ Total_Recovered : num  128000 172464 141335 15205 52833 ...
##  $ Active_Cases   : num   20809 7258 58756 90 9498 ...
##  $ Total_Cases_per1M_population: int   3894 63546 4589 199217 1872 40718 115496 100342 147611 6329 ..
##  $ Deaths_per1M_population    : int    181 1005 131 1679 50 1000 2533 2058 1593 64 ...
```

```
## $ Total_Tests          : num  771431 1289520 230861 193595 1092363 ...
## $ Tests_per1M_population : num  19247 448738 5143 2500323 31933 ...
## $ Population           : int  40080392 2873656 44892255 77428 34207984 98997 45741769 297054
```

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

```
iris_csv %>% arrange(desc(Total_Cases)) %>% top_n(10, Total_Cases) %>% select(Country)
```

```
##      Country
## 1      USA
## 2     India
## 3    Brazil
## 4      UK
## 5    Russia
## 6    Turkey
## 7    France
## 8     Iran
## 9  Argentina
## 10   Spain
```

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

```
iris_csv$Rate_Deaths <- iris_csv$Total_Deaths / iris_csv$Total_Cases
iris_csv[, "Rate_Deaths"]
```

```
##      [1] 0.046530105 0.015815125 0.028637312 0.008427877 0.026580045 0.024559663
##      [7] 0.021931857 0.020505319 0.010790005 0.010185151 0.014014227 0.013342486
##     [13] 0.028842266 0.005035516 0.017748366 0.008919104 0.007702134 0.019426559
##     [19] 0.018430713 0.006524029 0.017422222 0.001146351 0.037030580 0.045661863
##     [25] 0.012915020 0.027876842 0.006510520 0.040266473 0.014466302 0.001905239
##     [31] 0.009147860 0.023350916 0.015953893 0.016921587 0.008682063 0.034339846
##     [37] 0.022373751 0.047843631 0.025457077 0.034727144 0.014761679 0.012541180
##     [43] 0.020228545 0.008651437 0.010160334 0.004680865 0.017658935 0.007130459
##     [49] 0.013446252 0.006439150 0.010901334 0.018989435 0.063904635 0.056299578
##     [55] 0.031704608 0.012314899 0.006619594 0.008005030 0.026773011 0.017599934
##     [61] 0.012935343 0.007393264 0.016478673 0.006809931 0.015804669 0.006590636
##     [67] 0.034139974 0.014065505 0.021282727 0.009030214 0.021677655 0.033825549
##     [73] 0.013541858 0.024809823 0.012563224 0.022990380 0.025561558 0.028028282
##     [79] 0.027245099 0.017273538 0.035919382 0.002493200 0.013317998 0.033781410
##     [85] 0.021356008 0.011235385 0.012373762 0.006111290 0.006085756 0.027782289
##     [91] 0.011286977 0.024683544 0.010603318 0.012841061 0.012867111 0.020815618
##     [97] 0.005964933 0.014712631 0.001512778 0.014962763 0.013255953 0.030419483
##    [103] 0.049389090 0.014182818 0.017207334 0.014507231 0.010379173 0.022332043
##    [109] 0.037172554 0.011700810 0.002775675 0.035203728 0.012240553 0.015715157
##    [115] 0.021386909 0.009186955 0.008879348 0.075716300 0.022902125 0.010591350
##    [121] 0.004777881 0.014629778 0.015491060 0.012754309 0.037471415 0.027549492
##    [127] 0.014034343 0.008760005 0.024422195 0.004746567 0.012605042 0.033226837
##    [133] 0.013646903 0.035287308 0.004418575 0.013513913 0.022360472 0.010388324
##    [139] 0.015499572 0.011875643 0.035226718 0.091084781 0.015214024 0.025666107
##    [145] 0.016700114 0.002550871 0.006841297 0.028656448 0.027991158 0.013279852
```



```
## [151] 0.007850016 0.020056157 0.014285714 0.016699946 0.015114710 0.016014793
## [157] 0.025413752 0.008664004 0.005388029 0.018918074 0.001892850 0.016726137
## [163] 0.028102398 0.014567436 0.055479806 0.030473862 0.010819165 0.017433155
## [169] 0.025416930 0.013047910 0.077016750 0.022019445 0.012808672 0.012935930
## [175] 0.060034224 0.051709402 0.007257404 0.027761823 0.010090038 0.006117909
## [181] 0.009303756 0.029546712 0.035410870 0.008794794 0.002885847 0.025439629
## [187] 0.015794693 0.023225124 0.015471809 0.016343425 0.007108684 0.012017684
## [193] 0.024327865 0.191329420 0.017457834 0.035215937
```

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

```
tinggi = iris_csv$Country[which.max(iris_csv$Rate_Deaths)]
rendah = iris_csv$Country[which.min(iris_csv$Rate_Deaths)]
print(paste("Negara dengan rasio kematian tertinggi : ",tinggi) )
```

```
## [1] "Negara dengan rasio kematian tertinggi : Yemen"
```

```
print(paste("Negara dengan rasio kematian terendah : ",rendah) )
```

```
## [1] "Negara dengan rasio kematian terendah : Bhutan"
```

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

```
plot(iris_csv$Total_Recovered,iris_csv$Total_Cases)
```

