

Document Quarto Application Covid 19

Adama SANOGO & Ditchaba YEO

2024-05-21

Description des données

Chargement des données et affichage des premières lignes

```
library("tidyverse")
world_covid_19 <- read_delim("C:/Ditch/DATA SCIENCE/R/Projet_Shiny_App/App_Shiny/App_Shiny/d
                        delim = ",", na = c(""))
head(world_covid_19)
```

```
# A tibble: 6 x 10
  `Province/State` `Country/Region`   Lat   Long Date      Confirmed Deaths
  <chr>            <chr>            <dbl> <dbl> <date>      <dbl>   <dbl>
1 <NA>            Afghanistan      33.9  67.7 2020-01-22         0         0
2 <NA>            Albania          41.2  20.2 2020-01-22         0         0
3 <NA>            Algeria          28.0   1.66 2020-01-22         0         0
4 <NA>            Andorra          42.5   1.52 2020-01-22         0         0
5 <NA>            Angola          -11.2  17.9 2020-01-22         0         0
6 <NA>            Antigua and Barbuda 17.1 -61.8 2020-01-22         0         0
# i 3 more variables: Recovered <dbl>, Active <dbl>, `WHO Region` <chr>
```

Description :

- Province/State: Entités administratives ou Etats d'Etats Fédéraux(USA) ou possédant des territoires extérieurs(France)
- Country/Region : Les pays du monde
- Lat : La latitude(coordonnée géographique)
- Long : La longitude(coordonnée géographique)
- Date : Dates auxquelles les données sont enregistrées

- Confirmed : Le nombre de personnes ayant fait la maladie du Covid 19
- Deaths : Le nombre de personnes décédées de la maladie du Covid 19
- Recovered : Le nombre de personnes guéries de la maladie du Covid 19
- Active : Le nombre de personnes faisant la maladie du Covid 19
- WHO Region: Les différentes régions selon le découpage de l'OMS

Problématique : Notre objectif est de fournir un outils utile pour comprendre l'évolution de la propagation de la maladie du Covid 19 tant aux niveaux étatiques, continentaux, régionaux et mondiaux; suivre les tendances et les statistiques, et prendre des décisions informées pour la santé publique.

Résumé statistique :

```
summary(world_covid_19)
```

Province/State	Country/Region	Lat	Long
Length:49068	Length:49068	Min. : -51.796	Min. : -135.00
Class :character	Class :character	1st Qu.: 7.873	1st Qu.: -15.31
Mode :character	Mode :character	Median : 23.634	Median : 21.75
		Mean : 21.434	Mean : 23.53
		3rd Qu.: 41.204	3rd Qu.: 80.77
		Max. : 71.707	Max. : 178.06
Date	Confirmed	Deaths	Recovered
Min. : 2020-01-22	Min. : 0	Min. : 0.0	Min. : 0
1st Qu.: 2020-03-08	1st Qu.: 4	1st Qu.: 0.0	1st Qu.: 0
Median : 2020-04-24	Median : 168	Median : 2.0	Median : 29
Mean : 2020-04-24	Mean : 16885	Mean : 884.2	Mean : 7916
3rd Qu.: 2020-06-10	3rd Qu.: 1518	3rd Qu.: 30.0	3rd Qu.: 666
Max. : 2020-07-27	Max. : 4290259	Max. : 148011.0	Max. : 1846641
Active	WHO Region		
Min. : -14	Length:49068		
1st Qu.: 0	Class :character		
Median : 26	Mode :character		
Mean : 8085			
3rd Qu.: 606			
Max. : 2816444			

Le résumé statistique ci-dessus nous donne un aperçu des mesures centrales et de la dispersion des différentes variables du jeu de données.

Traitement des données :

modifions la position de la variable “Date” avec relocate()

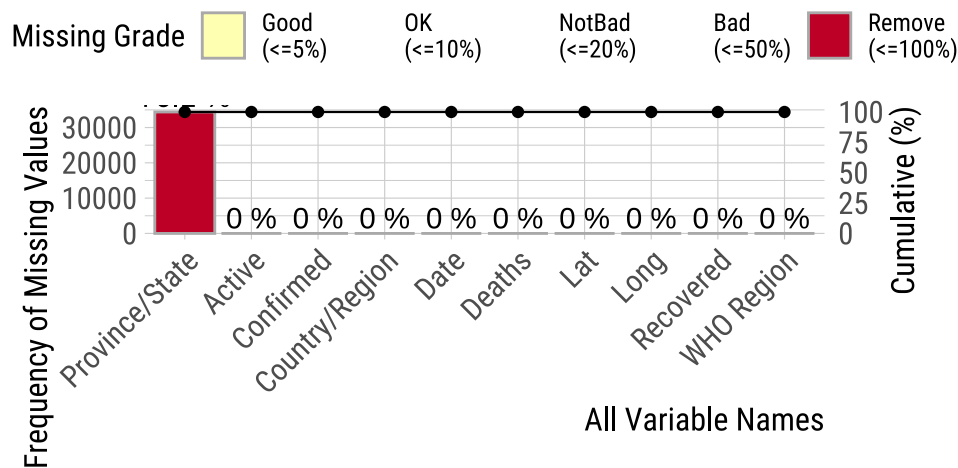
```
world_covid_19 <-  
  world_covid_19 |>  
  relocate(Date, .before = 'Province/State')  
#Affichage de world_world_19  
head(world_covid_19)
```

```
# A tibble: 6 x 10  
  Date       `Province/State` `Country/Region`   Lat   Long Confirmed Deaths  
  <date>     <chr>             <chr>             <dbl> <dbl>   <dbl> <dbl>  
1 2020-01-22 <NA>             Afghanistan       33.9  67.7     0      0  
2 2020-01-22 <NA>             Albania          41.2  20.2     0      0  
3 2020-01-22 <NA>             Algeria          28.0   1.66     0      0  
4 2020-01-22 <NA>             Andorra          42.5   1.52     0      0  
5 2020-01-22 <NA>             Angola          -11.2  17.9     0      0  
6 2020-01-22 <NA>             Antigua and Barbuda 17.1 -61.8     0      0  
# i 3 more variables: Recovered <dbl>, Active <dbl>, `WHO Region` <chr>
```

Repérer graphiquement les différentes variables pour repérer celles qui possèdent trop de valeurs manquantes en vue de les traiter.

```
library(dlookr)  
plot_na_pareto(world_covid_19)
```

Pareto chart with missing values



On voit que seule la variable Province/State a des données manquantes et en quantité importante, le nombre exact est 34404. Soit on supprime ou on laisse. La suppression ne va pas poser de problème car on a la longitude et la latitude des lieux. Dans la suite de notre travail, pour les Etats fédéraux nous utiliserons les *moyennes* des variables *Lat* et *Long* en lieu et place des valeurs de *Lat* et *Long*.

Suppression de la variable Province/State.

```
new_world_covid_19 <- world_covid_19[,-2]
```

Nous allons regrouper en fonction des continents le nombre total de: décès, de cas confirmés, de guéris et d'actifs

```
(total_continent <-
new_world_covid_19 |>
  group_by(`WHO Region`) |>
  summarise(Total_Confirmed = sum(Confirmed, na.rm = TRUE),
            Total_Deaths = sum(Deaths, na.rm = TRUE),
            Total_Recovered = sum(Recovered, na.rm = TRUE),
            Total_Active = sum(Active, na.rm = TRUE),
            ))
```

```
# A tibble: 6 x 5
```

	`WHO Region`	Total_Confirmed	Total_Deaths	Total_Recovered	Total_Active
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Africa	21791827	439978	11193730	10158119
2	Americas	402261194	19359292	157069444	225832458
3	Eastern Mediterrane~	74082892	1924029	48050703	24108160
4	Europe	248879793	19271040	123202075	106406678
5	South-East Asia	55118365	1458134	30030327	23629904
6	Western Pacific	26374411	932430	18861950	6580031

Regrouper en fonction des pays le nombre total de : décès, de cas confirmés, de guéris, non guéris et d'actifs

```
(total_country <-
  new_world_covid_19 |>
  group_by(`Country/Region`) |>
  summarise(Total_Confirmed = sum(Confirmed, na.rm = TRUE),
            Total_Deaths = sum(Deaths, na.rm = TRUE),
            Total_Recovered = sum(Recovered, na.rm = TRUE),
            Total_Active = sum(Active, na.rm = TRUE),
            ))
```

A tibble: 187 x 5

	`Country/Region`	Total_Confirmed	Total_Deaths	Total_Recovered	Total_Active
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1	Afghanistan	1936390	49098	798240	1089052
2	Albania	196702	5708	118877	72117
3	Algeria	1179755	77972	755897	345886
4	Andorra	94404	5423	69074	19907
5	Angola	22662	1078	6573	15011
6	Antigua and Barbuda	4487	326	2600	1561
7	Argentina	4450658	97749	1680024	2672885
8	Armenia	1587173	27089	857482	702602
9	Australia	960247	11387	711928	236932
10	Austria	2034986	71390	1638380	325216

i 177 more rows

Visualisation :

Maintenant nous pouvons faire la visualisation des données à partir de la construction de graphiques et de cartes

Exemple du code de la carte présentant les cas confirmés par pays dans le monde

```

library(leaflet)

data1 <- world_covid_19 |>
  group_by(`Country/Region`) |>
  summarise(
    latitude = mean(Lat, na.rm = TRUE),
    longitude = mean(Long, na.rm = TRUE),
    confirmed_total = sum(Confirmed, na.rm = TRUE))

ma_carte_0 <-
  leaflet() |>
  addTiles() |>
  setView(lng = 0, lat = 30, zoom = 2)

ma_carte <-
  ma_carte_0 |>
  addCircleMarkers(
    data = data1,
    lng = ~longitude,
    lat = ~latitude,
    color = "red",
    popup = ~paste(
      `Country/Region`, "<br>",
      "Cas confirmés:", confirmed_total),
    clusterOptions = markerClusterOptions(),
    radius = 5
  )

```

```

image_path <- "C:/Ditch/DATA SCIENCE/R/Projet_Shiny_App/graphe.jpg"
knitr::include_graphics(image_path)

```



Utilité et intérêt de l'application :

Notre application vise à fournir plusieurs fonctionnalités clés :

- Suivi des cas confirmés, des décès et des guérisons dans le monde entier.
- Visualisation des données sur les cartes et les graphiques pour mieux comprendre la propagation du virus
- Analyse des tendances et des modèles pour aider à prévoir les futures évolutions de la pandémie.

Présentation de l'application:

Notre application se présente comme une interface simple. Avec ses menus rangés horizontalement de la gauche vers la droite, et ses sous-sections accécibles grâce aux onglets, elle donne la possibilité de faire des filtres et des options pour affiner les données affichées selon les besoins de l'utilisateur.

Liens de l'application: https://yditchaba.shinyapps.io/App_Shiny/

Source des données: <https://www.kaggle.com/datasets/imdevskp/corona-virus-report>