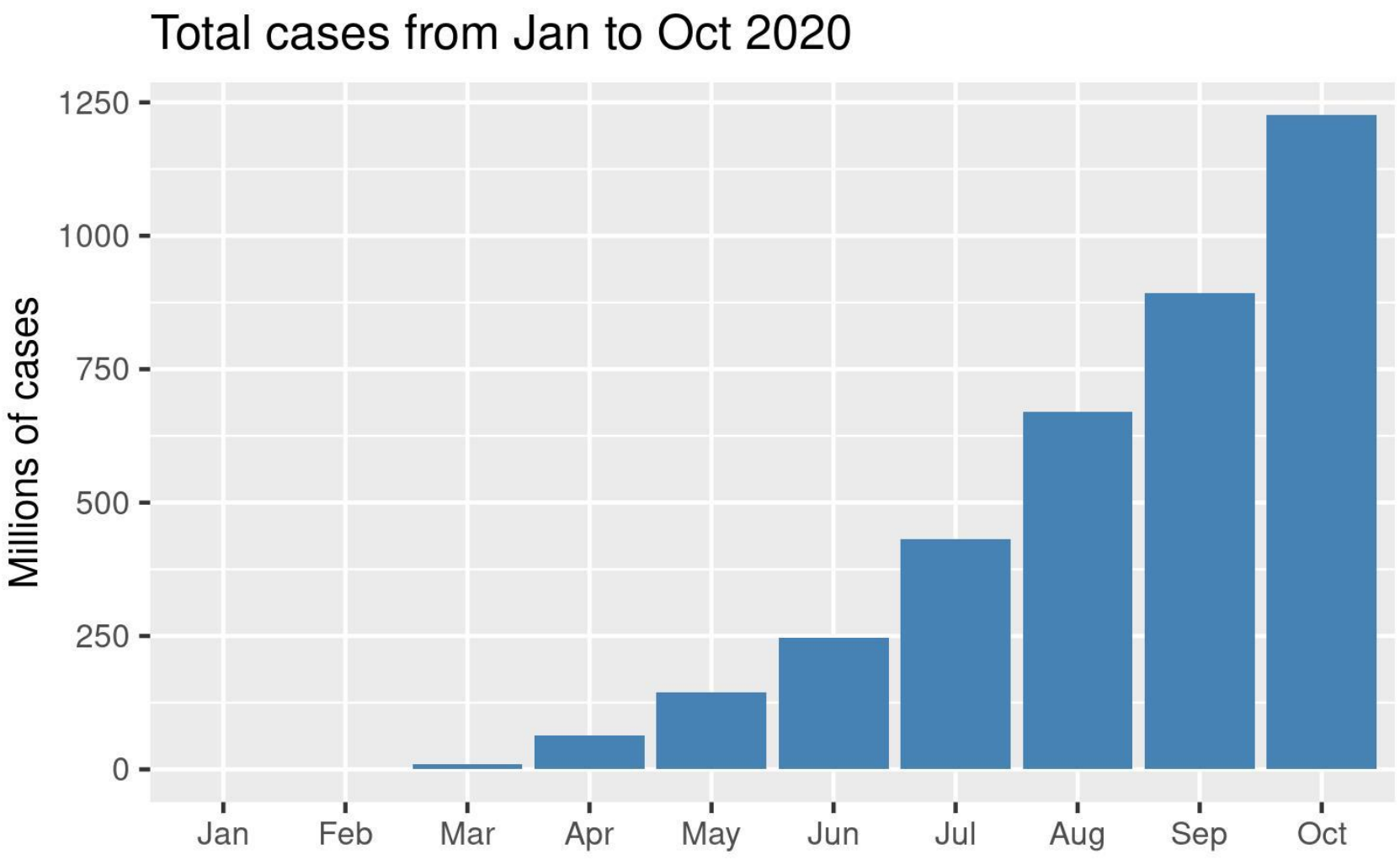
Jackie Pham

Covid19 - Data analysis using R

Data source:
https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

1. Three datasets include number of *confirmed cases*, *deaths* and *recovered cases* of each day during period of *22nd Jan to 9th Nov 2020*.
2. The analysis hereby is to:
    - Observe the number of cases increasing all over the world
    - Observe top 6 countries that have highest number of cases
    - Observe the cases in the US
    - Find the rate of change from 1st May to 31st Oct
    - Compare figures of northern countries (north of the equator) and southern countries


Total cases from Jan to Oct 2020
Millions of cases
1250
1000
750
500
250
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct

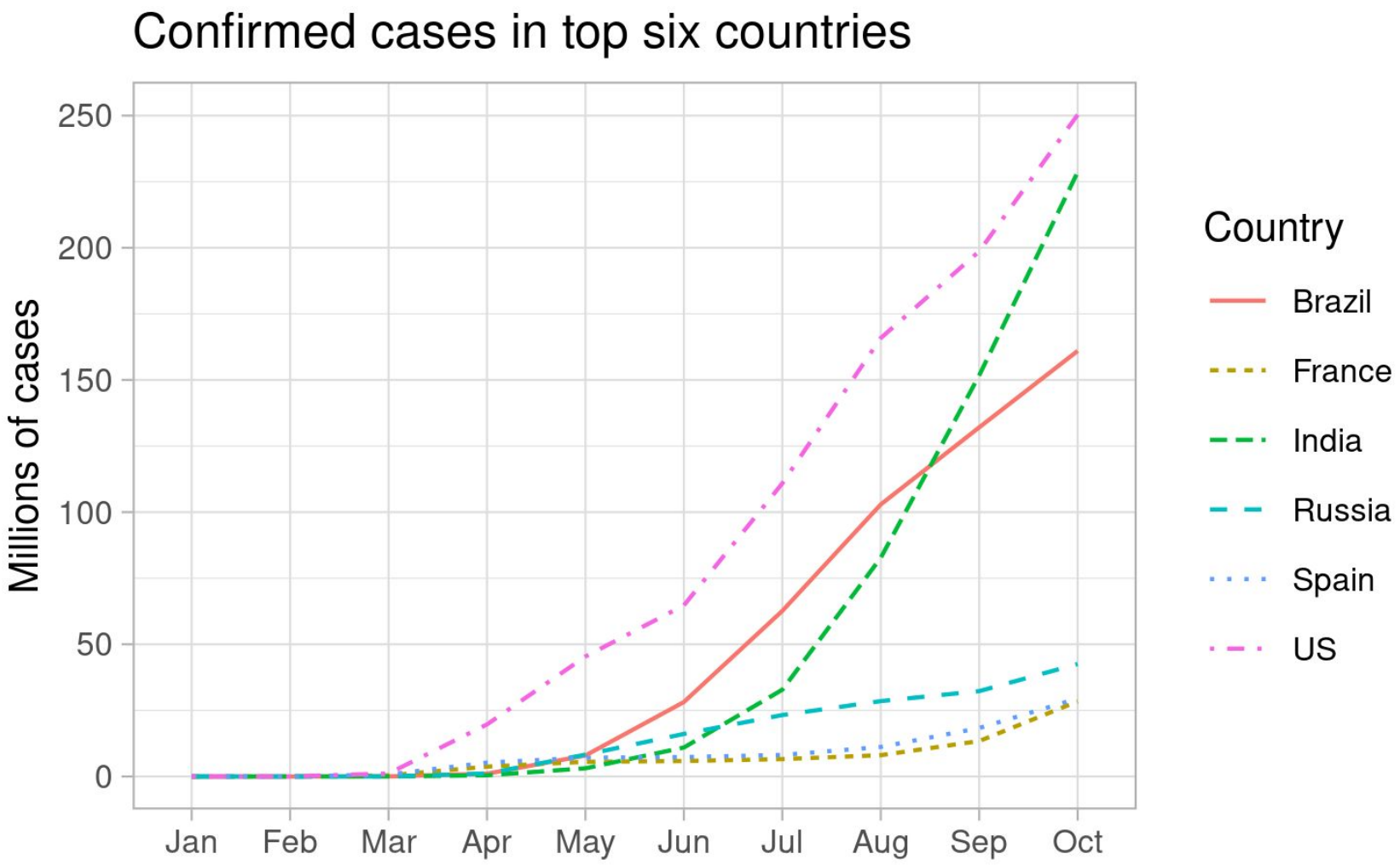
- On the last day of the observed period (9th Nov), the total confirmed cases around the world are **50,913,451** and distributed to northern countries and southern ones as follows:

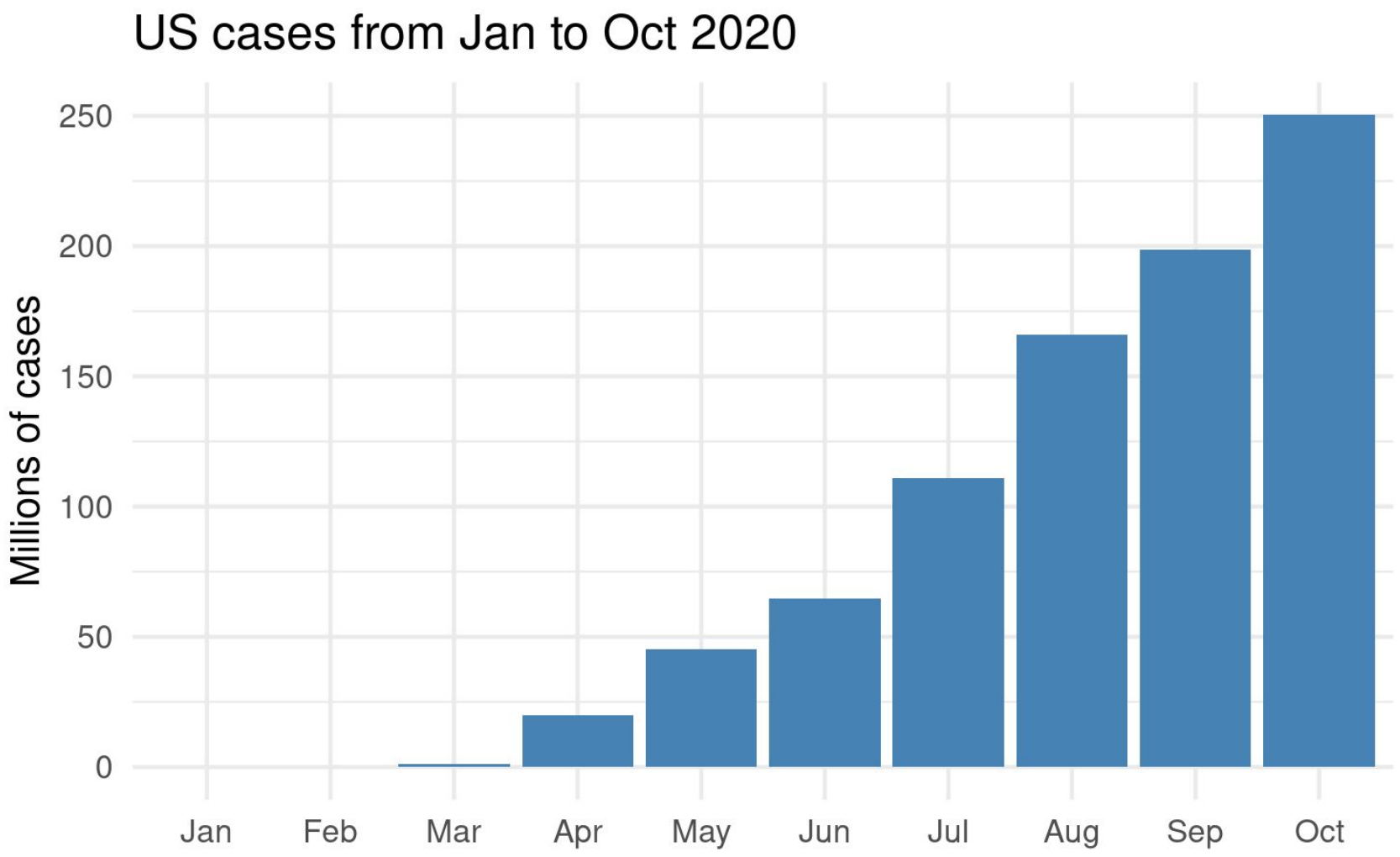
regions <fctr>	total_cases <dbl>	percentage <dbl>
Northern countries	40720033	79.97893
Southern countries	10192684	20.01963

Top 6 countries that had highest number of confirmed cases:

Country/Region <chr>	11/9/20 <dbl>
Spain	1381218
Russia	1781997
France	1856292
Brazil	5675032
India	8591730
US	10110552

The US had the highest number of confirmed cases among countries which was 10,110,552 on 9th Nov 2020


Confirmed cases in top six countries
Millions of cases
250
200
150
100
50
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Country
Brazil
France
India
Russia
Spain
US


US cases from Jan to Oct 2020
Millions of cases
250
200
150
100
50
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct

The rate of change May - Oct

- 114 out of 190 countries experienced an exponential strike in cases (have an increase by 1000%) within 6 months from 1st May to 31st Oct.
- When looking at the statistics we find that a large majority of those top countries are in Africa, some in the Middle Asia (Nepal, India, Pakistan, Bangladesh, ...) and some in South America (Brazil, Costa Rica, Mexico,...), followed by countries from Europe and South Asia.

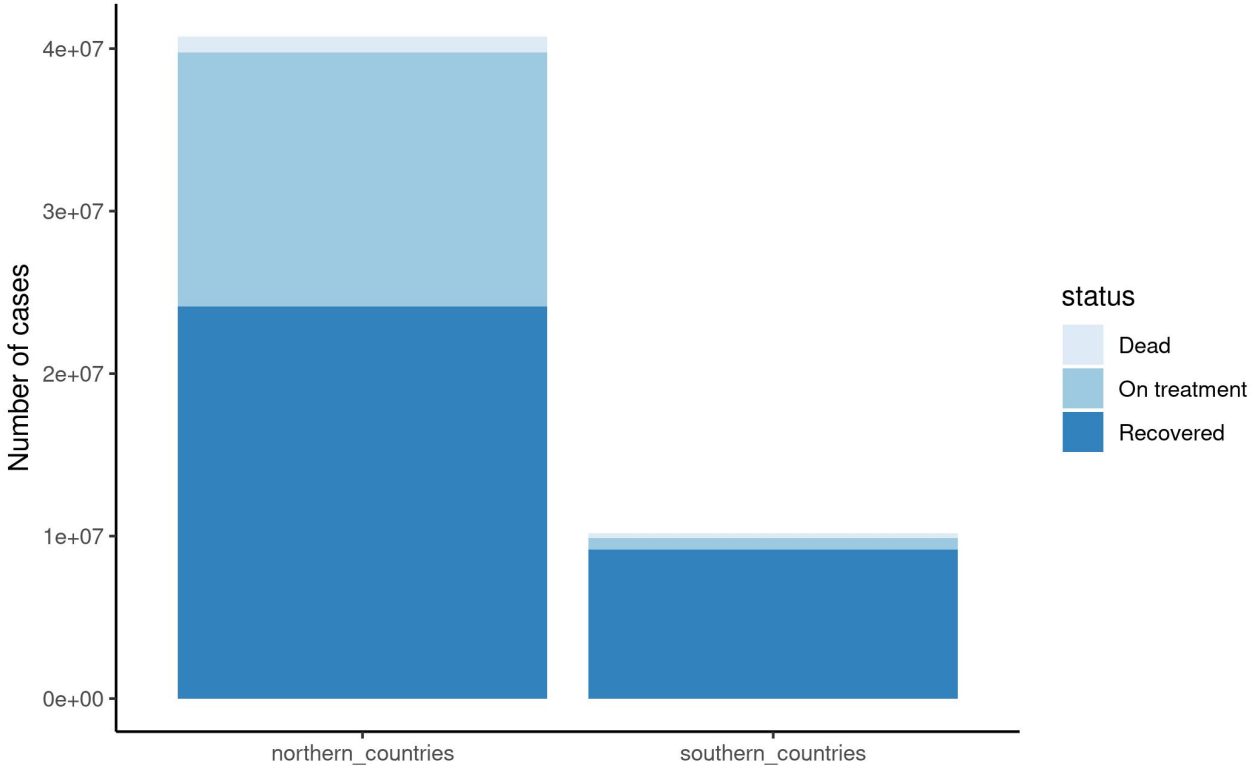
Africa have recorded the highest rate of change compared to other continents.

See variable may_oct in Rmd for more details

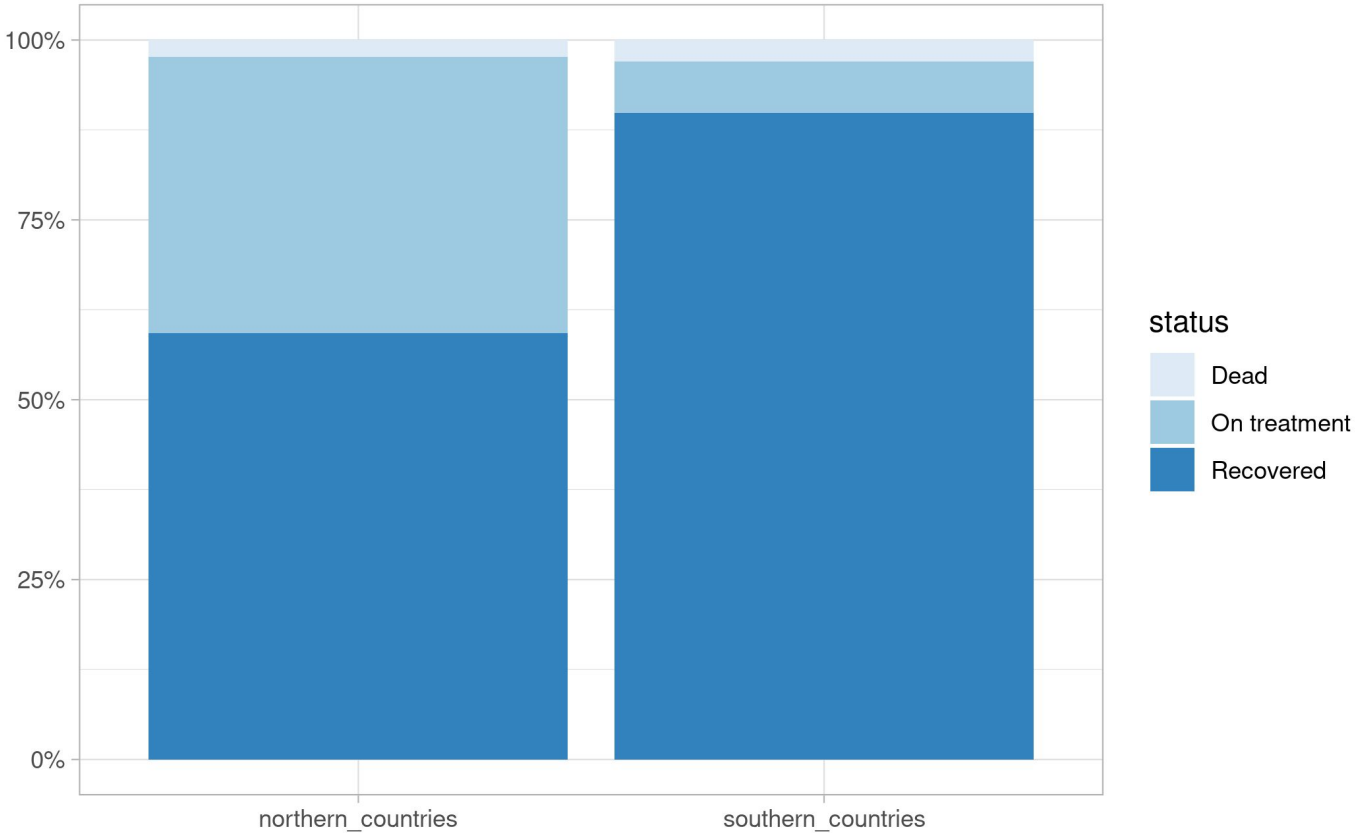
Now let’s take a look at all 3 dataframes to see the total numbers of *confirmed cases*, *deaths*, *under treatment* and *recovered cases*

By 9th Nov 2020 the numbers are as following:

regions <fctr>	confirmed <dbl>	deaths <dbl>	on_treatment <dbl>	recovered <dbl>
Northern countries	40720033	955048	15634207	24130778
Southern countries	10192684	308243	726474	9157967



- From this chart we can see the total confirmed cases in countries from the north of the equator are much higher than countries from the south.
- A majority of cases have been recovered in southern countries. We don’t see that in northern countries where there were still lots of cases on treatment.



- This chart is to illustrate the proportion of deaths, recovered and on-treatment cases in both areas.
- Again we can see a majority of cases have been recovered in southern countries.

By 9th Nov 2020:

- Total recovered cases are 33,289,404, which accounts for 65.38% of all confirmed cases all around the world.
- In northern countries, recovered cases are 24,130,778 (59.26% of confirmed cases) and sadly 955,048 people have died (2.35%)
- In southern countries, recovered cases are 9,157,967 (89.85% of confirmed cases) and 308,243 people have died (3.02%)