R Notebook

Code

Hide

```
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
```

Hide

```
library(readr)

# Load the data
confirmed <- read_csv('covid19_confirmed_global.csv')
```

```

── Column specification ────────────────────────────────────────────────────────
cols(
  .default = col_double(),
  `Province/State` = col_character(),
  `Country/Region` = col_character()
)
ℹ Use `spec()` for the full column specifications.
```

Hide

```
deaths <- read_csv('covid19_deaths_global.csv')
```

```

── Column specification ────────────────────────────────────────────────────────
cols(
  .default = col_double(),
  `Province/State` = col_character(),
  `Country/Region` = col_character()
)
ℹ Use `spec()` for the full column specifications.
```

Hide

```
recovered <- read_csv('covid19_recovered_global.csv')
```

```
── Column specification ────────────────────────────────────────────────────────
cols(
  .default = col_double(),
  `Province/State` = col_character(),
  `Country/Region` = col_character()
)
ℹ Use `spec()` for the full column specifications.
```

Hide

```
# Inspect the data
head(confirmed)
```

Province/State <chr>	Country/Region <chr>	Lat <dbl>	Long <dbl>	1/22/20 <dbl>	1/23/20 <dbl>	1/24/20 <dbl>
NA	Afghanistan	33.93911	67.70995	0	0	0
NA	Albania	41.15330	20.16830	0	0	0
NA	Algeria	28.03390	1.65960	0	0	0
NA	Andorra	42.50630	1.52180	0	0	0
NA	Angola	-11.20270	17.87390	0	0	0
NA	Antigua and Barbuda	17.06080	-61.79640	0	0	0

6 rows | 1-9 of 531 columns

Hide

```
head(deaths)
```

Province/State <chr>	Country/Region <chr>	Lat <dbl>	Long <dbl>	1/22/20 <dbl>	1/23/20 <dbl>	1/24/20 <dbl>
NA	Afghanistan	33.93911	67.70995	0	0	0
NA	Albania	41.15330	20.16830	0	0	0
NA	Algeria	28.03390	1.65960	0	0	0
NA	Andorra	42.50630	1.52180	0	0	0
NA	Angola	-11.20270	17.87390	0	0	0

NA	Antigua and Barbuda	17.06080	-61.79640	0	0	0

6 rows | 1-9 of 531 columns

Hide

```
head(recovered)
```

Province/State <chr>	Country/Region <chr>	Lat <dbl>	Long <dbl>	1/22/20 <dbl>	1/23/20 <dbl>	1/24/20 <dbl>
NA	Afghanistan	33.93911	67.70995	0	0	0
NA	Albania	41.15330	20.16830	0	0	0
NA	Algeria	28.03390	1.65960	0	0	0
NA	Andorra	42.50630	1.52180	0	0	0
NA	Angola	-11.20270	17.87390	0	0	0
NA	Antigua and Barbuda	17.06080	-61.79640	0	0	0

6 rows | 1-9 of 531 columns

Hide

```
# Sum the total number of cases on March 22nd
total_cases <- confirmed %>% select(`3/22/20`) %>% sum()
total_cases
```

```
[1] 344205
```

Hide

```
# Filter for countries in the northern hemisphere
northern_hemisphere_total_cases <- confirmed %>% filter(Lat > 0)%>%
  select('3/22/20')%>%
  sum()
northern_hemisphere_total_cases
```

```
[1] 336947
```

Hide

```
# Filter for Australia cases
australia_confirmed_cases <- confirmed %>% filter(`Country/Region` == 'Australia')
%>%
  select('3/22/20') %>%
  sum()
australia_confirmed_cases
```

```
[1] 1549
```

Hide

```
# Display percentages
northern_hemisphere_total_cases_percentage <- (northern_hemisphere_total_cases / t
otal_cases) * 100
northern_hemisphere_total_cases_percentage
```

```
[1] 97.89137
```

Hide

```
australia_confirmed_cases_percentage <- (australia_confirmed_cases / total_cases)
* 100
australia_confirmed_cases_percentage
```

```
[1] 0.4500225
```

Hide

```
# Group by countries
countries_grouped <- confirmed %>% group_by(`Country/Region`) %>% summarize_at(var
s(-Lat, -Long, -`Province/State`),sum)
countries_grouped
```

Country/Region <chr>	**1/22/20** <dbl>	**1/23/20** <dbl>	**1/24/20** <dbl>	**1/25/20** <dbl>	**1/26/20** <dbl>	**1/27/20** <dbl>	**1/28/20** <dbl>
Afghanistan	0	0	0	0	0	0	0
Albania	0	0	0	0	0	0	0
Algeria	0	0	0	0	0	0	0
Andorra	0	0	0	0	0	0	0
Angola	0	0	0	0	0	0	0
Antigua and Barbuda	0	0	0	0	0	0	0

Argentina	0	0	0	0	0	0	0
Armenia	0	0	0	0	0	0	0
Australia	0	0	0	0	4	5	5
Austria	0	0	0	0	0	0	0

1-10 of 195 rows | 1-10 of 528 columns Previous **1** 2 3 4 5 6 … 20 Next

Hide

```
# Filter the grouped data set for Australia
austrlia_grouped_confirmed <- countries_grouped %>% filter('Country/Region' == 'Au
stralia') %>%
  select('3/22/20')
australia_confirmed_cases
```

```
[1] 1549
```

Hide

```
# Group by countries
countries_grouped_recovered <- recovered %>% group_by(`Country/Region`) %>% summar
ize_at(vars(-Lat, -Long, -`Province/State`),sum)
countries_grouped_recovered
```

Country/Region <chr>	**1/22/20** <dbl>	**1/23/20** <dbl>	**1/24/20** <dbl>	**1/25/20** <dbl>	**1/26/20** <dbl>	**1/27/20** <dbl>	**1/28/20** <dbl>
Afghanistan	0	0	0	0	0	0	0
Albania	0	0	0	0	0	0	0
Algeria	0	0	0	0	0	0	0
Andorra	0	0	0	0	0	0	0
Angola	0	0	0	0	0	0	0
Antigua and Barbuda	0	0	0	0	0	0	0
Argentina	0	0	0	0	0	0	0
Armenia	0	0	0	0	0	0	0
Australia	0	0	0	0	0	0	0
Austria	0	0	0	0	0	0	0

1-10 of 195 rows | 1-10 of 528 columns Previous **1** 2 3 4 5 6 … 20 Next

Hide

```
us_grouped_confirmed <- countries_grouped %>% filter(`Country/Region` == 'US') %>% select('3/22/20')
us_grouped_confirmed
```

3/22/20
<dbl>
34898

1 row

Hide

```
us_grouped_recovered <- countries_grouped_recovered %>% filter(`Country/Region` == 'US') %>% select('3/22/20')
us_grouped_recovered
```

3/22/20
<dbl>
178

1 row

Hide

```
us_grouped_recovered_percentage <- us_grouped_recovered / us_grouped_confirmed * 100
us_grouped_recovered_percentage
```

3/22/20
<dbl>
0.5100579

1 row

Hide

```
uk_grouped_confirmed <- countries_grouped %>% filter(`Country/Region` == 'United Kingdom') %>% select('3/22/20')
uk_grouped_confirmed
```

3/22/20
<dbl>
10395

1 row

Hide

```
uk_grouped_recovered <- countries_grouped_recovered %>% filter(`Country/Region` == 'United Kingdom') %>% select('3/22/20')
uk_grouped_recovered
```

3/22/20
<dbl>
67

1 row

Hide

```
uk_grouped_recovered_percentage <- uk_grouped_recovered / uk_grouped_confirmed * 100
uk_grouped_recovered_percentage
```

3/22/20
<dbl>
0.6445406

1 row

Hide

```
# Filter to inspect the US row
countries_grouped_recovered %>% filter(`Country/Region` == 'United Kingdom')
```

Country/Region	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/3
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<
United Kingdom	0	0	0	0	0	0	0	0	

1 row | 1-10 of 528 columns

Hide

```
# Find the maximum number of confirmed and recovered cases
max_confirmed <- max(uk_grouped_confirmed)
max_confirmed
```

```
[1] 10395
```

Hide

Hide

```
max_recovered <- max(uk_grouped_recovered)
max_recovered
```

```
[1] 67
```

Hide

```
max_recovered/max_confirmed * 100
```

```
[1] 0.6445406
```

Hide

```
library(janitor)
```

```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

    chisq.test, fisher.test
```

Hide

```
transposed_confirmed <- transposed_confirmed %>% row_to_names(row_number = 1)
transposed_confirmed %>% head()
```

	Afghanistan <chr>	**Albania** <chr>	**Algeria** <chr>	**Ando…** <chr>	**Ang…** <chr>	**Antigua and Barbuda** <chr>	**Argentina** <chr>
1/22/20	0	0	0	0	0	0	0
1/23/20	0	0	0	0	0	0	0
1/24/20	0	0	0	0	0	0	0
1/25/20	0	0	0	0	0	0	0
1/26/20	0	0	0	0	0	0	0
1/27/20	0	0	0	0	0	0	0

6 rows | 1-9 of 171 columns

Hide

```
# Transform the columns to numeric values
transposed_confirmed <- apply(transposed_confirmed, 2, as.numeric) %>% as.data.fra
me()
head(transposed_confirmed)
```

	Afghanistan <dbl>	Albania <dbl>	Algeria <dbl>	Ando… <dbl>	Ang… <dbl>	Antigua and Barbuda <dbl>	Argentina <dbl>	Arme… <dbl>
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0

6 rows | 1-10 of 171 columns

Hide

```
library(ggplot2)
```

Hide

```
# Find the maximum number of confirmed cases in the US
transposed_confirmed %>% select(US) %>% max()
```

```
[1] 33272
```

Hide

```
# Find the maximum number of confirmed cases in the UK
transposed_confirmed %>% select('United Kingdom') %>% max()
```

```
[1] 5741
```

Hide

```
transposed_confirmed <- transposed_confirmed %>% mutate(date = c(1:nrow(transposed
_confirmed)))
head(select(transposed_confirmed, date))
```

date <int>

1	1
2	2
3	3
4	4
5	5
6	6

6 rows

Hide

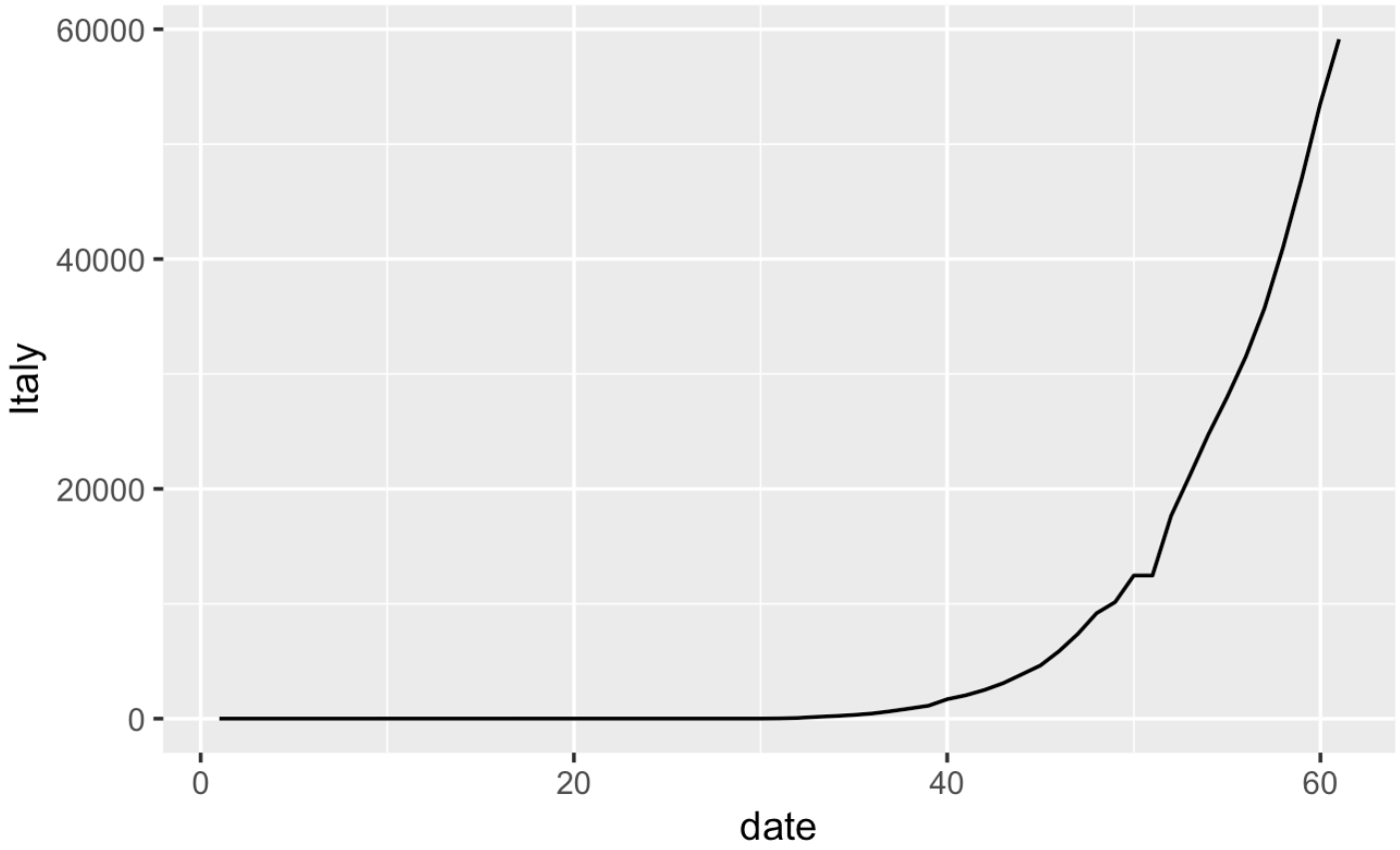
```
transposed_confirmed
```

Afghanistan <dbl>	Albania <dbl>	Algeria <dbl>	Ando… <dbl>	Ang… <dbl>	Antigua and Barbuda <dbl>	Argentina <dbl>	Arme… <dbl>
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

1-10 of 61 rows | 1-9 of 172 columns Previous **1** 2 3 4 5 6 7 Next

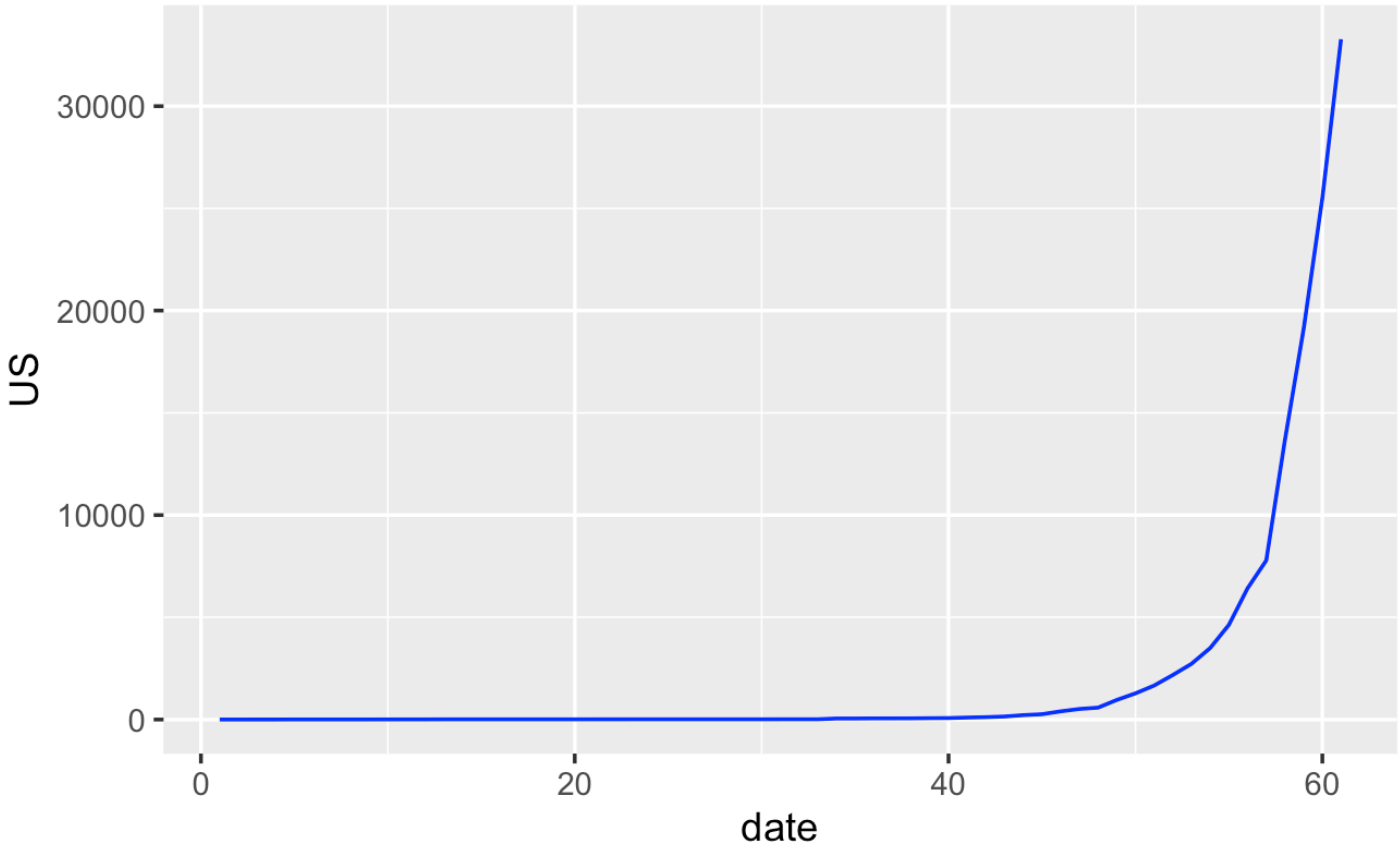
Hide

```
# Create a line graph with date on the X axis and number of cases in Italy on the
Y axis
transposed_confirmed %>% ggplot(aes(x = date, y = Italy)) + geom_line()
```



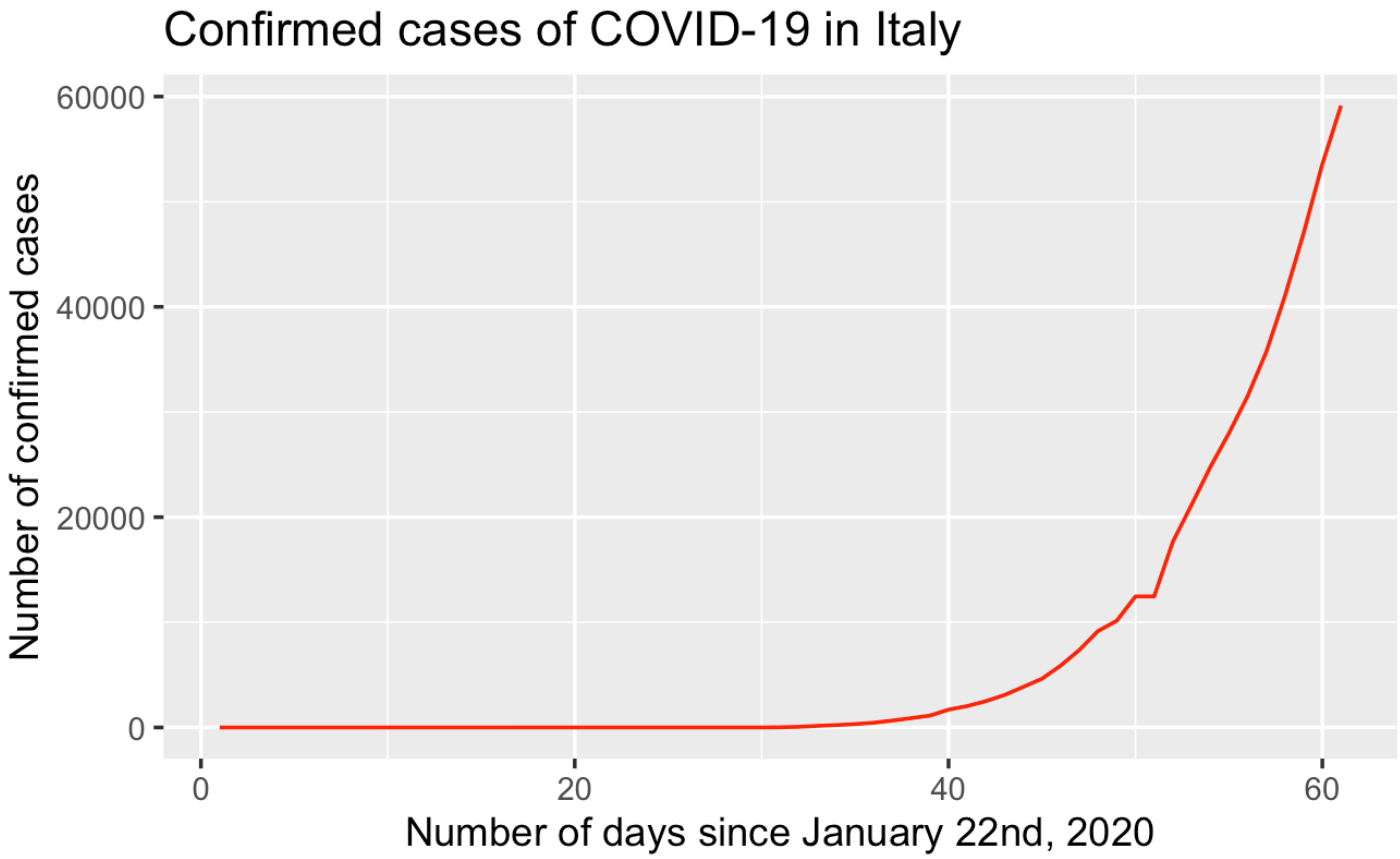
Hide

```
# Create a line graph with date on the X axis and number of cases in US on the Y a
xis
transposed_confirmed %>% ggplot(aes(x = date, y = US)) + geom_line(colour = "blue"
)
```



Hide

```
transposed_confirmed %>% ggplot(aes(x = date, y = Italy)) + geom_line(colour = "red") + labs(x = "Number of days since January 22nd, 2020", y = "Number of confirmed cases", title = "Confirmed cases of COVID-19 in Italy")
```


Confirmed cases of COVID-19 in Italy
Number of confirmed cases
60000
40000
20000
0
0
20
40
60
Number of days since January 22nd, 2020