

Branch: master ▾ seirR / data-raw / Examples / 02 Explore Data /

Create new file

Upload files

Find file

History

Jim Duggan and Jim Duggan Update help files		Latest commit 6537b26 7 hours ago
..		
📁 README_files/figure-gfm	Update help files	7 hours ago
📄 .DS_Store	Updates for first release	17 hours ago
📄 README.Rmd	Update	8 hours ago
📄 README.md	Update help files	7 hours ago

📄 README.md



## 🔗 Exploring available data

The package downloads up-to-date COVID-19 numbers from the [https://covid.ourworldindata.org/data/ecdc/full\\_data.csv](https://covid.ourworldindata.org/data/ecdc/full_data.csv).

This data is stored in the **data\_env** environment, and can be accessed. The following is the process for doing this

- First, load in the libraries, and include **ggplot2** for visualisation, and **dplyr** and **tidyr** for data manipulation.

```
library(seirR)
```

```
## Welcome to package seirR v0.0.0.1
```

```
## Checking https://covid.ourworldindata.org/data/ecdc/full_data.csv for data update...
```

```
## Loading https://covid.ourworldindata.org/data/ecdc/full_data.csv to global environment data_env
```

```
library(ggplot2)
```

```
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(tidyr)
```

- Explore the data contained in the R environment **data\_env**

```
data_env$covid_data
```

```
## # A tibble: 8,195 x 6
##   Date      Country ReportedNewCases ReportedNewDeat... ReportedTotalCa...
##   <date>    <chr>          <dbl>          <dbl>          <dbl>
## 1 2019-12-31 Afghan...          0            0            0
## 2 2020-01-01 Afghan...          0            0            0
## 3 2020-01-02 Afghan...          0            0            0
## 4 2020-01-03 Afghan...          0            0            0
## 5 2020-01-04 Afghan...          0            0            0
## 6 2020-01-05 Afghan...          0            0            0
## 7 2020-01-06 Afghan...          0            0            0
## 8 2020-01-07 Afghan...          0            0            0
## 9 2020-01-08 Afghan...          0            0            0
## 10 2020-01-09 Afghan...          0            0            0
## # ... with 8,185 more rows, and 1 more variable: ReportedTotalDeaths <dbl>
```

- Filter out a country of interest, and reshape to tidy data format

```
countries <- c("South Korea")

ds <- filter(data_env$covid_data,
              Country %in% countries)

ds_piv <- ds %>% pivot_longer(~c(Date, Country), names_to = "Measure", values_to = "Number") %>%
  filter(Measure %in% c("ReportedNewCases", "ReportedNewDeaths"))
```

- Explore the re-shaped dataset

```
ds_piv

## # A tibble: 186 x 4
##   Date      Country Measure      Number
##   <date>    <chr>    <chr>    <dbl>
## 1 2019-12-31 South Korea ReportedNewCases      0
## 2 2019-12-31 South Korea ReportedNewDeaths      0
## 3 2020-01-01 South Korea ReportedNewCases      0
## 4 2020-01-01 South Korea ReportedNewDeaths      0
## 5 2020-01-02 South Korea ReportedNewCases      0
## 6 2020-01-02 South Korea ReportedNewDeaths      0
## 7 2020-01-03 South Korea ReportedNewCases      0
## 8 2020-01-03 South Korea ReportedNewDeaths      0
## 9 2020-01-04 South Korea ReportedNewCases      0
## 10 2020-01-04 South Korea ReportedNewDeaths      0
## # ... with 176 more rows
```

- Visualise some of the results

```
ggplot(ds_piv, aes(x = Date, y = Number, colour = Measure)) +
  geom_point() + geom_line()
```

