Assessment 1 info track

1. Installing covid19 package



WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/markm/AppData/Local/R/win-library/4.3’

(as ‘lib’ is unspecified)

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/ggplot2\_3.5.1.zip'

Content type 'application/zip' length 4952503 bytes (4.7 MB)

downloaded 4.7 MB

package ‘ggplot2’ successfully unpacked and MD5 sums checked

Error in install.packages : ERROR: failed to lock directory ‘C:\Users\markm\AppData\Local\R\win-library\4.3’ for modifying

Try removing ‘C:\Users\markm\AppData\Local\R\win-library\4.3/00LOCK’







WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/markm/AppData/Local/R/win-library/4.3’

(as ‘lib’ is unspecified)

also installing the dependencies ‘R.oo’, ‘R.methodsS3’, ‘R.utils’, ‘data.table’

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/R.oo\_1.26.0.zip'

Content type 'application/zip' length 972504 bytes (949 KB)

downloaded 949 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/R.methodsS3\_1.8.2.zip'

Content type 'application/zip' length 83288 bytes (81 KB)

downloaded 81 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/R.utils\_2.12.3.zip'

Content type 'application/zip' length 1411368 bytes (1.3 MB)

downloaded 1.3 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/data.table\_1.16.0.zip'

Content type 'application/zip' length 2447609 bytes (2.3 MB)

downloaded 2.3 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/COVID19\_3.0.3.zip'

Content type 'application/zip' length 35811 bytes (34 KB)

downloaded 34 KB

package ‘R.oo’ successfully unpacked and MD5 sums checked

Error in install.packages : ERROR: failed to lock directory ‘C:\Users\markm\AppData\Local\R\win-library\4.3’ for modifying

Try removing ‘C:\Users\markm\AppData\Local\R\win-library\4.3/00LOCK’

unlink("C:/Users/markm/AppData/Local/R/win-library/4.3/00LOCK", recursive = TRUE)

having trouble finding ID so just use the function Covid19

analyse the data and then evaluate this in relation to the questions we are answering

what are the broad types of data:

- case data (new, total)

- mortality data (daily, total)

- hospitalization data (currently admitted, ICU admisssions)

- Testing data (number of test, outcomes)

- vaccination data (coverage, efficacy)

- demographics (age, gender, pre-existing conditions) - how different characteristics of the population were differently impacted

main types of resources:

- data sources (OWID, John hopkinds, WHO and CDC)

- r libraries

What types of problems can the resources solve?

- monotor the spread of the virus

- time-series forecasting (ARIME, exponential smoothing) can predict future cases

- can be compared with economic activity, employment and GDP

general data science resources that might be applicable?

- time series analysis - ARIMA / Exponential smoothingv/vrolbing averages

- machine learning methods / regression models used to predict cases / deaths

- data visualization: ggplot2

- predicted modelling: SEIR models

- dashboards - R shiny/Tableau

Comparisons: [COVID-19 (Coronavirus) Data Hub | Tableau](https://www.tableau.com/covid-19-coronavirus-data-resources#:~:text=The%20Global%20COVID-19%20tracker%20provided%20key%20metrics%20on%20where)

- case counts vs mortality data

- geographic comparison / Population Density vs. COVID-19 Case Growth:

- correlation between covid cases and mortality rate

- demographic factors and mortality rate (age, gender, population density) - this can include comparison between countries with younger vs older populations / Correlation between COVID-19 Cases and Death Rates by Age Group

- seasonal trends

potential graphs:

- line plots with rolling averages for cases / deaths

- heatmap of case growth by country - intensity of case growth - darker => higher case counts

- Bar Plot of Mortality Rate by Age Group

- population density vs case rate with bubble plot

Example project notes:

* I have analysed in depth 2 sets of data from different sources, with a brief look at another data set. These data sets are listed in the resources section at the bottom of the document.

Necessary packages

[r code]

Anlysis of dataset 1

[import]

Initial analysis:

Data types, columns, missing data counts , data quality report: type. Count, unique values with columns down side

Head command

Make any changes necessary to the DS e.g. timestamps i.e. make It nicer / renaming

Specific analysis on [file size]

Treeplot using squarify

Specific analysis of \_\_

Anlysis of Dataset 2….

Pie chart – death tolls for example but maybe normalise this figure / isolate larger frequencies (<https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/#15.-Ordered-Bar-Chart>)

[The R Graph Gallery – Help and inspiration for R charts (r-graph-gallery.com)](https://r-graph-gallery.com/#:~:text=The%20R%20Graph%20Gallery%20boasts%20the%20most%20extensive%20compilation%20of) [The R Graph Gallery – Help and inspiration for R charts (r-graph-gallery.com)](https://r-graph-gallery.com/#:~:text=The%20R%20Graph%20Gallery%20boasts%20the%20most%20extensive%20compilation%20of)

[Top 50 ggplot2 Visualizations - The Master List (With Full R Code) (r-statistics.co)](https://r-statistics.co/Top50-Ggplot2-Visualizations-MasterList-R-Code.html?utm_content=cmp-true)