

# Statistical Computing with R

## Masters in Data Science 503 (S6)

### First Batch, SMS, TU, 2021

Shital Bhandary

Associate Professor

Statistics/Bio-statistics, Demography and Public Health Informatics

Patan Academy of Health Sciences, Lalitpur, Nepal

Faculty, Data Analysis and Decision Modeling, MBA, Pokhara University, Nepal

Faculty, FAIMER Fellowship in Health Professions Education, India/USA.

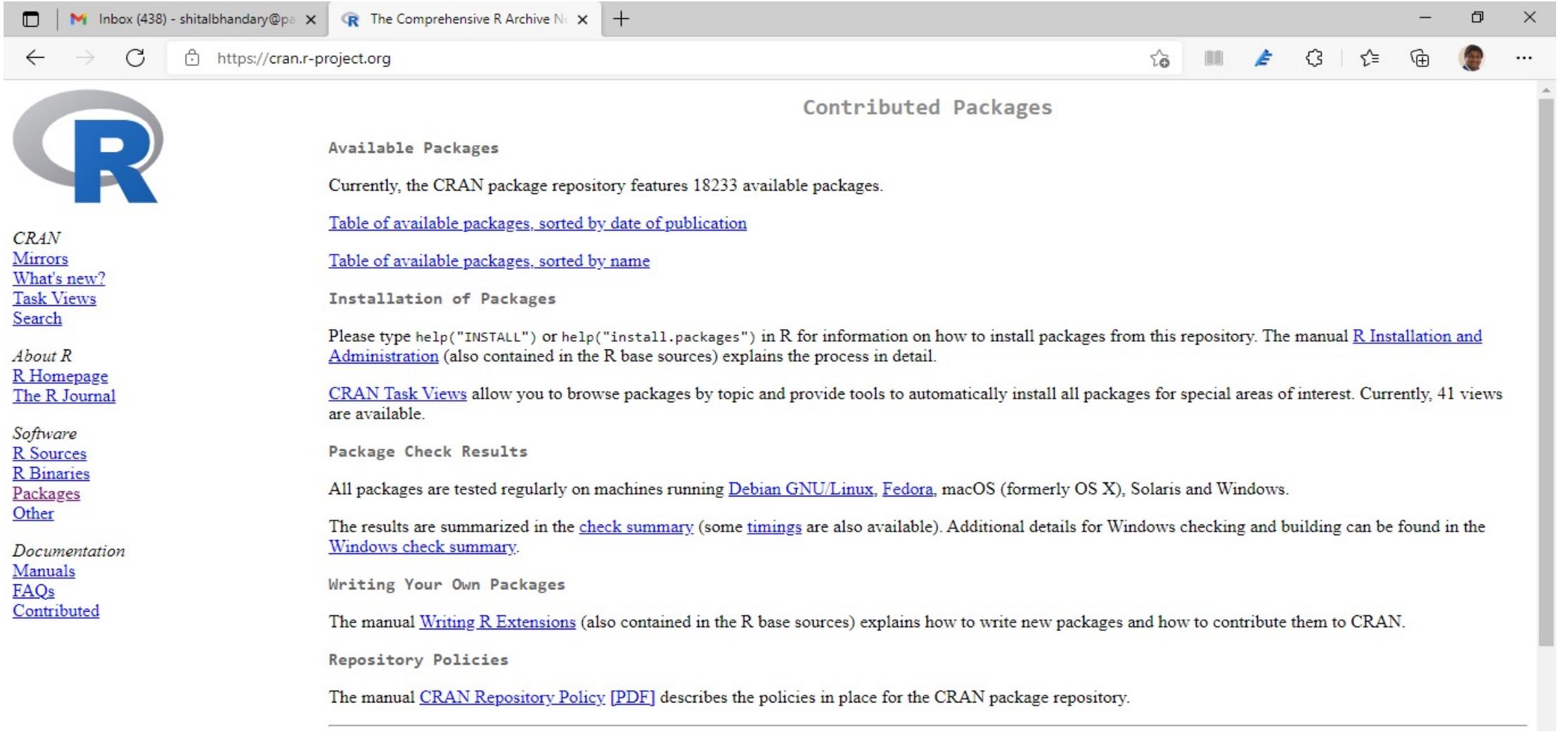
# Review Preview (Unit 2, Part 1)

- Package in R
- Using package in R
- Reading data in R
- Reviewing data in R

# What is a “package” in R?

- In R, the fundamental unit of shareable code is the package.
- A package bundles together code, data, documentation, and tests, and is easy to share with others.
- As of June 2019, there were over 14,000 packages available on the **Comprehensive R Archive Network**, or CRAN, the public clearing house for R packages.
- This huge variety of packages is one of the reasons that R is so successful: the chances are that someone has already solved a problem that you’re working on, and you can benefit from their work by downloading their package.

# Available “packages” from CRAN:



The screenshot shows a web browser window with the URL <https://cran.r-project.org>. The page title is "Contributed Packages". On the left, there is a sidebar with links to various CRAN resources like CRAN Mirrors, Task Views, and Documentation. The main content area contains sections for Available Packages, Installation of Packages, Package Check Results, Writing Your Own Packages, and Repository Policies.

**Contributed Packages**

**Available Packages**

Currently, the CRAN package repository features 18233 available packages.

[Table of available packages, sorted by date of publication](#)

[Table of available packages, sorted by name](#)

**Installation of Packages**

Please type `help("INSTALL")` or `help("install.packages")` in R for information on how to install packages from this repository. The manual [R Installation and Administration](#) (also contained in the R base sources) explains the process in detail.

[CRAN Task Views](#) allow you to browse packages by topic and provide tools to automatically install all packages for special areas of interest. Currently, 41 views are available.

**Package Check Results**

All packages are tested regularly on machines running [Debian GNU/Linux](#), [Fedora](#), macOS (formerly OS X), Solaris and Windows.

The results are summarized in the [check summary](#) (some [timings](#) are also available). Additional details for Windows checking and building can be found in the [Windows check summary](#).

**Writing Your Own Packages**

The manual [Writing R Extensions](#) (also contained in the R base sources) explains how to write new packages and how to contribute them to CRAN.

**Repository Policies**

The manual [CRAN Repository Policy \[PDF\]](#) describes the policies in place for the CRAN package repository.

# How to install, use and get help about any package from CRAN?

- We can install packages of CRAN in R using:

```
install.packages("packagename")
```

e.g. `install.packages("dplyr")`

- We can then use the installed packages in R using:

```
library(packagename)
```

i.e. `library(dplyr)`

- We can get help on the installed packages in R using:

```
?packagement or help(package = "packagement")
```

e.g. `?dplyr` OR `help(package = "dplyr")`

# After typing library(dplyr) and ?dplyr in R:

The screenshot shows a web browser window with multiple tabs open. The active tab displays the documentation for the `dplyr` package. The URL in the address bar is `127.0.0.1:19490/library/dplyr/html/dplyr-package.html`. The page title is `dplyr-package {dplyr}` and the subtitle is `dplyr: A Grammar of Data Manipulation`. The content includes sections for `Description`, `Author(s)`, `Maintainer`, `Authors`, `Other contributors`, and `See Also`. It also lists useful links and provides a link to the package index.

dplyr-package {dplyr}

dplyr: A Grammar of Data Manipulation

**Description**

To learn more about `dplyr`, start with the vignettes: `browseVignettes(package = "dplyr")`

**Author(s)**

**Maintainer:** Hadley Wickham [hadley@rstudio.com](mailto:hadley@rstudio.com) ([ORCID](#))

Authors:

- Romain François ([ORCID](#))
- Lionel Henry
- Kirill Müller ([ORCID](#))

Other contributors:

- RStudio [copyright holder, funder]

**See Also**

Useful links:

- <https://dplyr.tidyverse.org>
- <https://github.com/tidyverse/dplyr>
- Report bugs at <https://github.com/tidyverse/dplyr/issues>

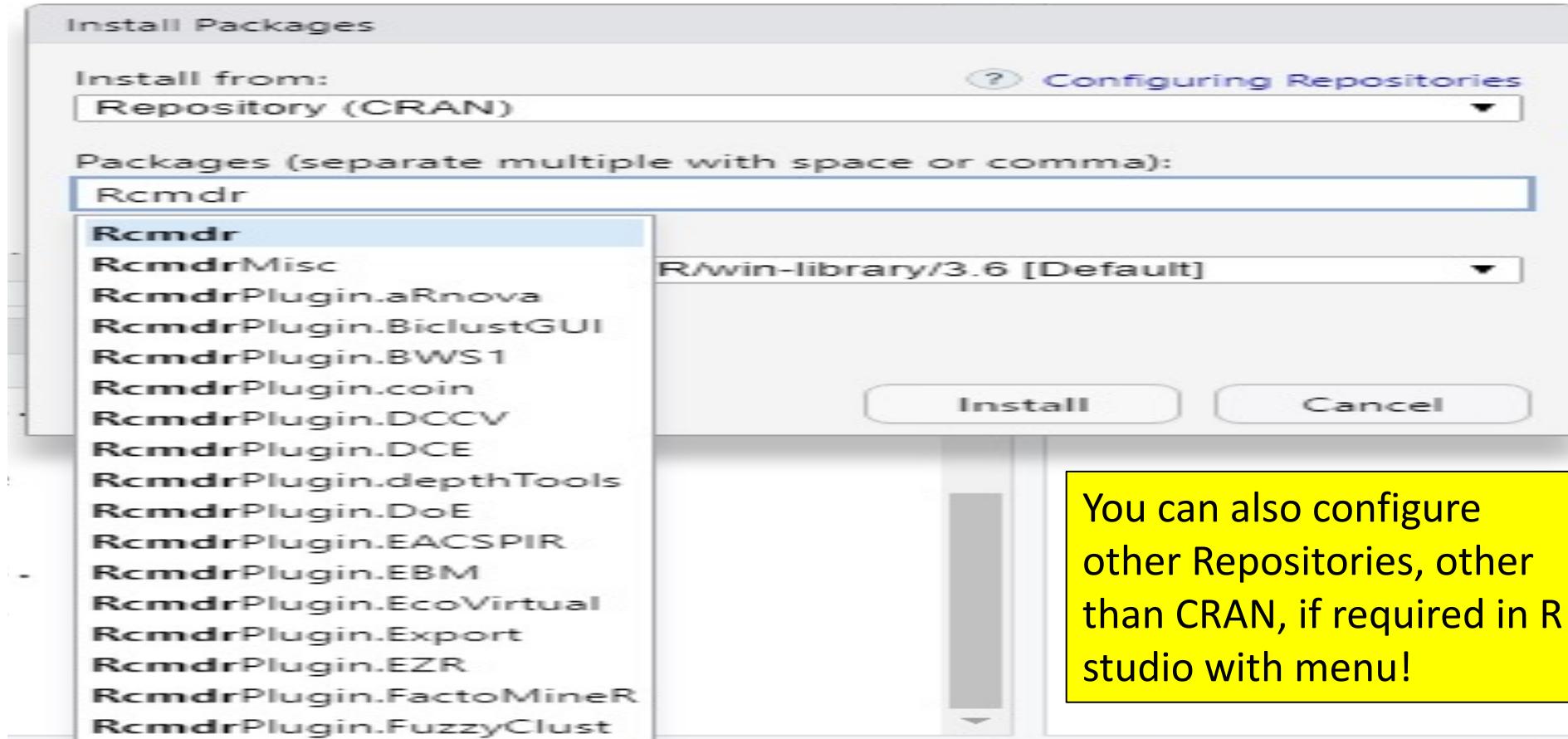
---

[Package `dplyr` version 1.0.2 [Index](#)]

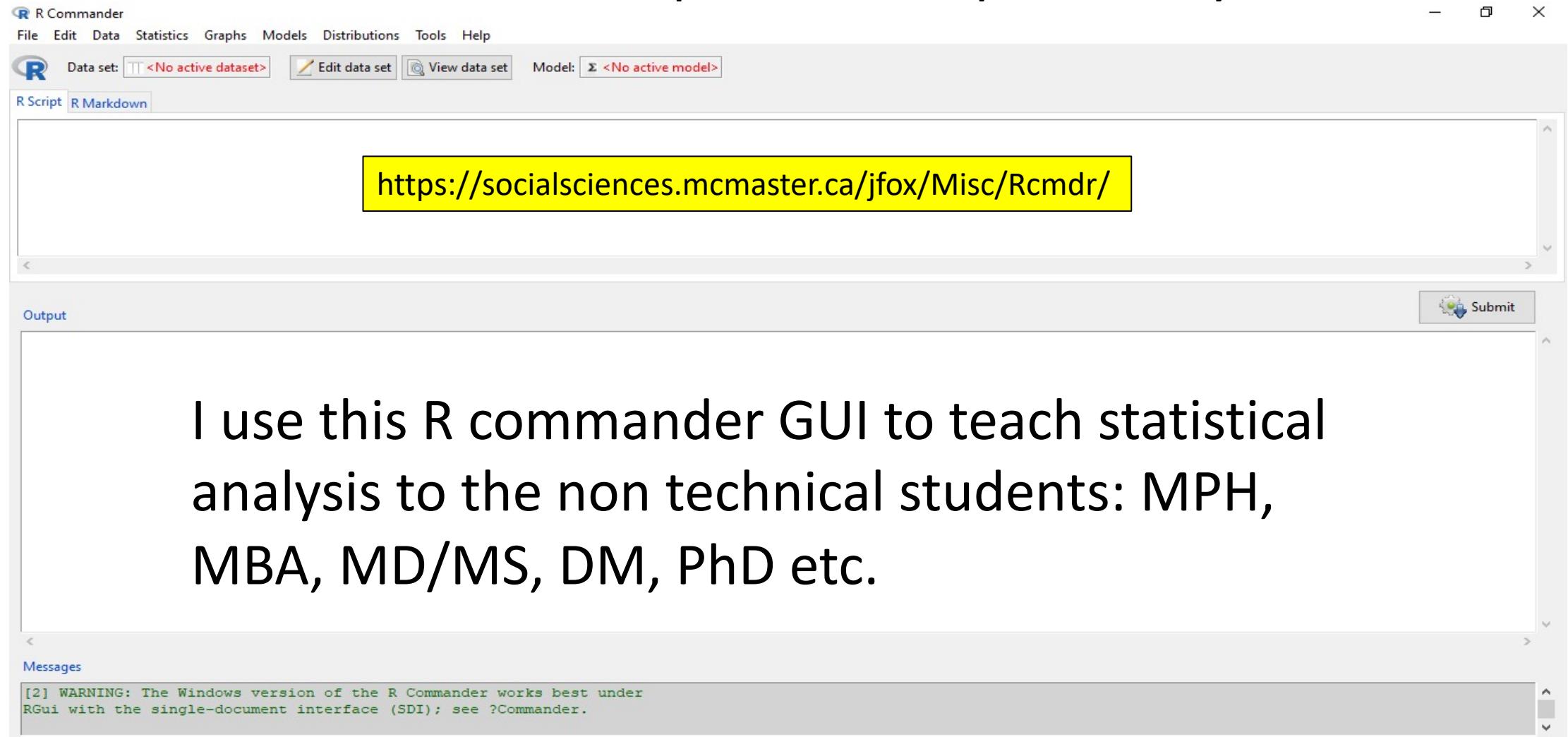
Let us try to install a very popular R package to do statistical analysis among researchers:

- `install.packages("Rcmdr")` in R Console or R Studio
- This will ask you to select a mirror, always select the first mirror
- No need to select mirror near to us as internet speed is “good” now
- Load this package using `library(Rcmdr)` command after successful installation

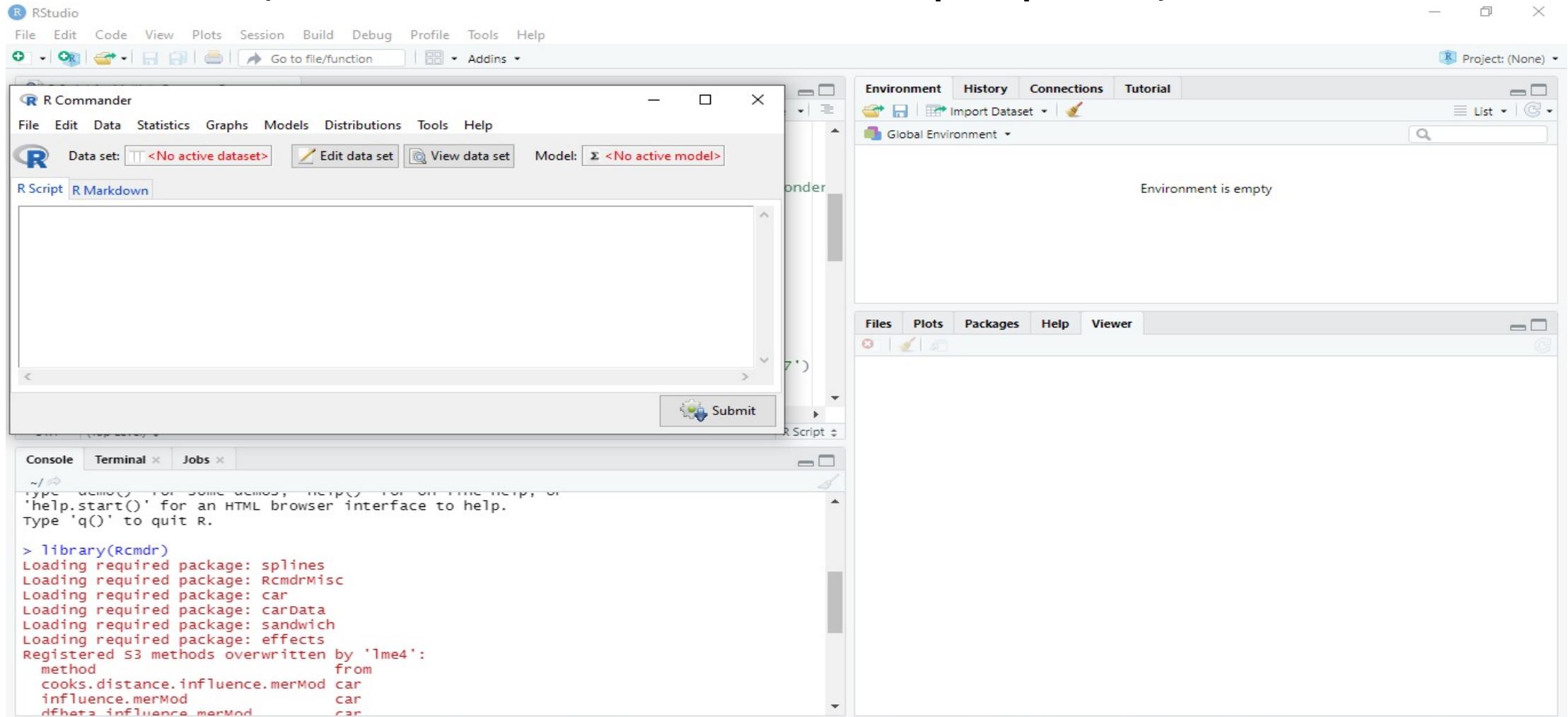
In R Studio, you can use menu to install any package from CRAN: Tools -> Install packages



When library(Rcmdr) is used in R console:  
A new GUI will be opened separately from R!



# I got this after typing library(Rcmdr) in R Studio (Windows 10, Dell Optiplex):



# “Packages” details at CRAN:

---

## **Related Directories**

### [Archive](#)

Previous versions of the packages listed above, and other packages formerly available.

### [Orphaned](#)

Packages with no active maintainer, see the corresponding [README](#).

### [bin/windows/contrib](#)

Windows binaries of contributed packages

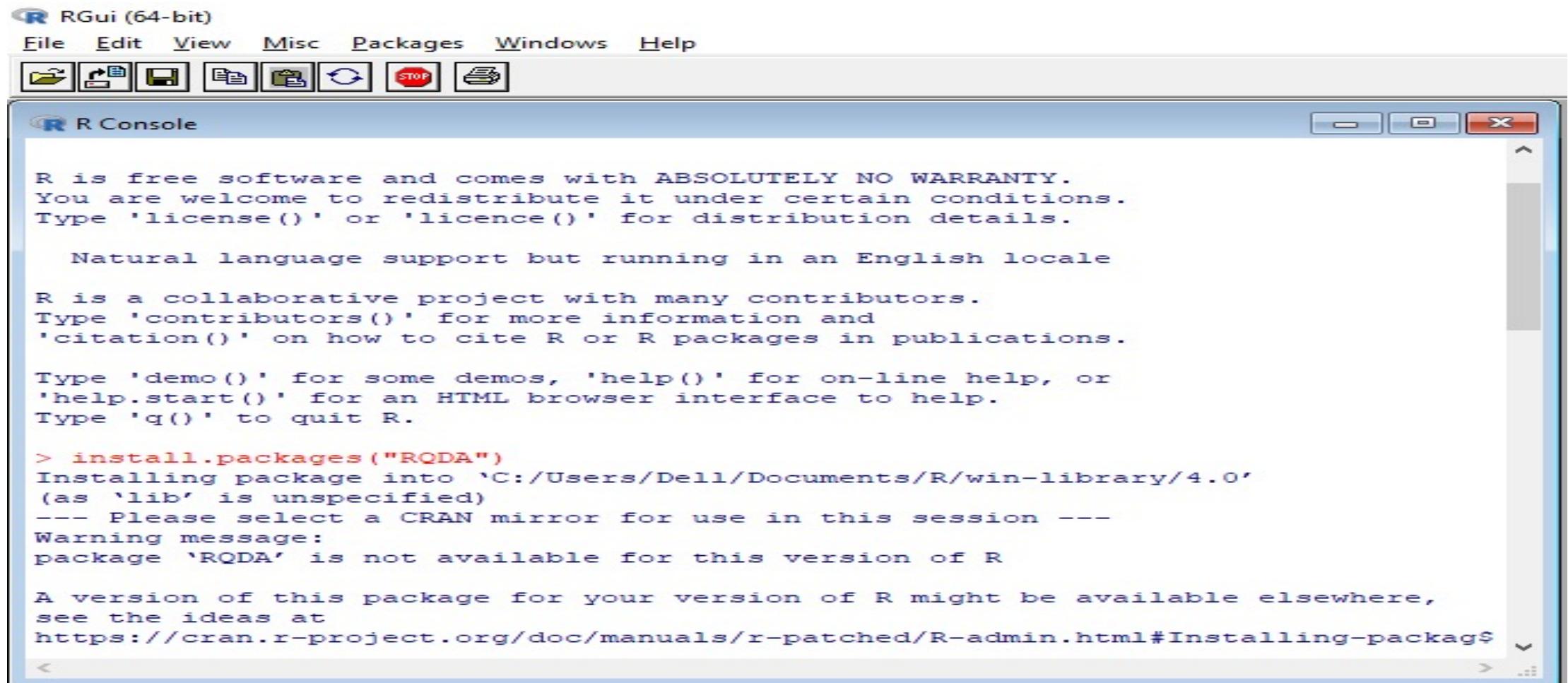
### [bin/macosx/contrib](#)

macOS High Sierra binaries of contributed packages

### [bin/macosx/el-capitan/contrib](#)

OS X El Capitan binaries of contributed packages

Let us try to install another very popular package for doing Qualitative Data Analysis in R (RQDA):



The screenshot shows the RGui (64-bit) application window. The title bar reads "RGui (64-bit)". Below it is a menu bar with "File", "Edit", "View", "Misc", "Packages", "Windows", and "Help". A toolbar with various icons follows. The main area is titled "R Console". It displays the standard R startup message about being free software with no warranty, followed by information about natural language support, contributors, and help resources. Then, a command is entered: > `install.packages("RQDA")`. The console outputs the path to the library, prompts for a CRAN mirror, and displays a warning message stating that the package 'RQDA' is not available for the current version of R. It also provides a link for more information.

```
R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.  
  
Natural language support but running in an English locale  
  
R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.  
  
Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.  
  
> install.packages("RQDA")  
Installing package into 'C:/Users/Dell/Documents/R/win-library/4.0'  
(as 'lib' is unspecified)  
--- Please select a CRAN mirror for use in this session ---  
Warning message:  
package 'RQDA' is not available for this version of R  
  
A version of this package for your version of R might be available elsewhere,  
see the ideas at  
https://cran.r-project.org/doc/manuals/r-patched/R-admin.html#Installing-packages
```

Kindly visit this website and try to fix it:

- <https://rqda.r-forge.r-project.org/> #RQDA general package info
- If it did not help then you can visit here for most recent version:
  - <https://rqda.github.io/RQDA/> # github site of RQDA developer
  - `install.packages("devtools")` #Install it only if you haven't done it yet!
  - `devtools::install_github("RQDA/RQDA", INSTALL_opts = "--no-multiarch")`

# My MPH/PhD students used this approach to install RQDA package in PC (Windows OS):



Shital Bhandary

Jun 10, 2020 (Edited Jun 11, 2020)

⋮

I could install RQDA in R 3.6.1 as follows and hope you can do the same too in your PC:

1. `install.packages("RGtk2") #R 3.4.0 and higher required`
2. `install.packages("https://cran.r-project.org/bin/windows/contrib/3.3/RQDA_0.3-1.zip"); repos=NULL` #RQDA was not available for R3.6.1 from CRAN so using older version
3. `install.packages("RSQLite") #As this package was build for higher version # library(RQDA) command informed that RSQLite could not be loaded so installed it again for R3.6.1`
4. `install.packages("https://cran.r-project.org/bin/windows/contrib/3.3/gWidgetsRGtk2_0.0-86.zip"); repos=NULL` #This package is also removed from the CRAN and old version is used
5. `install.packages("https://cran.r-project.org/bin/windows/contrib/3.3/gWidgets_0.0-54.zip"); repos=NULL` #This package is also removed from the CRAN and old version is used
6. `install.packages("cairoDevice") #This package was required to run gWidgets`
7. `library(RQDA) # Running now`



1 class comment



Shital Bhandary Jun 10, 2020

Portable version of RQDA is available from the developer's dropbox here:

[https://www.dropbox.com/s/5zebadz41dep09k/RQDA\\_0.3\\_1.rar?dl=0](https://www.dropbox.com/s/5zebadz41dep09k/RQDA_0.3_1.rar?dl=0)

# Use of RQDA in MPH Thesis/Publication:

 **JNHR**  
Journal of Nepal Health Research Council

[HOME](#)   [ANNOUNCEMENTS](#)   [ABOUT JNHR](#)   [EDITORIAL TEAM](#)   [ISSUES](#) ▾   [INSTRUCTION & GUIDELINES](#)   [CONTACT](#)   [!\[\]\(5e53eb9f193269f9de9b9d9917d9677c\_img.jpg\) Search](#)

Home / Archives / Vol 19 No 2 (2021): Vol. 19 No. 1 Issue 51 Apr-Jun 2021 / Original Article

**Factors Affecting Psychoactive Substances Use Among Bachelor Level Students of Nepalgunj SubMetropolitan City**

**Puspa Basnet**  
School of Public Health, Patan Academy of Health Sciences, Lagenkhel, Lalitpur, Nepal

**Reshu Agrawal Sagtani**  
School of Public Health, Patan Academy of Health Sciences, Lagenkhel, Lalitpur, Nepal

**Shital Bhandary**  
School of Public Health, Patan Academy of Health Sciences, Lagenkhel, Lalitpur, Nepal

**DOI:** <https://doi.org/10.33314/jnhr.v19i2.3431>

 [PDF](#)

Published  
2021-09-06

How to Cite  
Basnet, P., Agrawal Sagtani, R., & Bhandary, S. (2021). Factors Affecting Psychoactive Substances Use Among Bachelor Level Students of Nepalgunj SubMetropolitan City. *Journal of Nepal Health Research Council*, 19(2), 288-294.  
<https://doi.org/10.33314/jnhr.v19i2.3431>

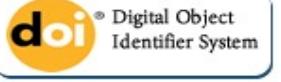
[Make a Submission](#)

**JNHR is indexed in and associated with:**









# Publication with RQDA: Factors Affecting Psychoactive Substances Use Among Bachelor Level Students of Nepalganj Sub Metropolitan City

The screenshot shows a web browser window with the following details:

- Address bar: Not secure | jnhrc.com.np/index.php/jnhrc/article/view/3431/1113
- Title bar: Factors Affecting Psychoactive Substances Use Among Bachelor Level Students of Nepalganj SubMetropolitan City
- Toolbar: Includes back, forward, search, and download icons.
- Content area:
  - Header: Factors Affecting Psychoactive Substances Use Among Bachelor Level Students
  - Text (left column): Describes the study design, including snowball sampling, purposive selection of parents/guardians and teachers, and audio-recording of interviews.
  - Text (right column): Details the selection of students from different streams, the number of participants (13 students, 6 parents/guardians, 6 teachers), and the time period (August-September 2020). It also mentions transcription, translation, and deductive thematic analysis using RQDA software.
  - Section header: RESULTS

# “Packages” from Bioconductor repository: To work with Bioinformatics!

The screenshot shows the official Bioconductor website at [www.bioconductor.org](http://www.bioconductor.org). The page features a teal header with the Bioconductor logo and navigation links for Home, Install, Help, Developers, and About. Below the header, there are four main content boxes: 'About Bioconductor' (describing tools for genomic analysis), 'Install' (listing packages and installation guides), 'Learn' (Master Bioconductor tools), 'News' (a list of recent announcements), 'Use' (bioinformatic solutions), and 'Develop' (contribute to Bioconductor).

**About Bioconductor**

Bioconductor provides tools for the analysis and comprehension of high-throughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, and an active user community. Bioconductor is also available as an [AMI](#) (Amazon Machine Image) and [Docker](#) images.

**News**

- Bioconductor [3.14](#) release schedule announced. Please view for important deadlines.
- Bioconductor [Bioc 3.13](#) Released.
- Bioconductor [browsable code base](#) now available.
- See our [google calendar](#) for events, conferences, meetings, forums, etc. Add your event with email to events at bioconductor.org.
- Bioconductor [F1000 Research Channel](#) is available.
- Orchestrating single-cell analysis with Bioconductor ([abstract](#); [website](#)) and other [recent literature](#).

**Install >**

- Discover [2042 software packages](#) available in Bioconductor release 3.13.

Get started with Bioconductor

- [Install Bioconductor](#)
- [Get support](#)
- [Latest newsletter](#)
- [Follow us on twitter](#)
- [Install R](#)

**Learn >**

Master Bioconductor tools

- [Courses](#)
- [Support site](#)
- [Package vignettes](#)
- [Literature citations](#)
- [Common work flows](#)
- [FAQ](#)
- [Community resources](#)
- [Videos](#)

**Use >**

Create bioinformatic solutions with Bioconductor

- [Software](#), [Annotation](#), and [Experiment](#) packages
- [Docker](#) and [Amazon](#) machine images
- Latest [release announcement](#)
- Use Bioconductor in the [AnVIL](#). See our [project updates](#).
- [Community Slack](#) sign-up
- [Support site](#)
- [Events calendar](#); [email events at](#)

**Develop >**

Contribute to Bioconductor

- [Developer resources](#)
- [Use Bioc 'devel'](#)
- 'Devel' packages
- [Package guidelines](#)
- [New package submission](#)
- [Git source control](#)
- [Build reports](#)
- [Browsable code base](#)

# NCBI Repository: Free to use and analyze!

The screenshot shows a web browser window with the URL <https://www.ncbi.nlm.nih.gov/sars-cov-2/>. The page title is "SARS-CoV-2 Sequence Resources". Below it, a section titled "Genome Reference Sequence (NC\_045512)" is displayed. This section contains four items, each with a blue button to its right:

- NCBI RefSeq** SARS-CoV-2 genome annotation **Download Annotation**
- NCBI RefSeq** SARS-CoV-2 genome sequence record **View Record**
- NCBI RefSeq** SARS-CoV-2 genome graphical display **View Display**
- NCBI Gene** SARS-CoV-2 curated gene records **View Records**

---

**Explore the Data**

Search a **BLAST** database of Betacoronavirus nucleotide sequences **Run BLAST**

**Feedback**

↑  
Back To Top

# GISAID: Needs registration and “strict” code of conduct to be followed!

The screenshot shows the GISAID website homepage with the following sections:

- In Focus:** First human case of influenza A/H10N3 virus infection. On May 28, 2021, a resident of the City of Zhenjiang in Jiangsu Province, China, was diagnosed with an influenza A/H10N3 virus infection. This is the first human case of infection with a low pathogenic avian influenza virus of this subtype. How the person was infected is unclear. There is no evidence that the virus spreads easily between humans. A/H10 influenza viruses are abundant in wild birds globally, and occasionally transmit to poultry.  
[> read more](#)
- hCoV-19 Submission Tracking:** Two world maps showing submission distribution.
- hCoV-19 Tracking of Variants:** A world map showing the tracking of COVID-19 variants.
- Genomic epidemiology of hCoV-19:** Phylogeny, Diversity, and Transmission maps.
- hCoV-19 data sharing via GISAID:** 3,649,841 submissions.
- Enabled by data shared via GISAID:** CoVerage, Helmholtz Centre for Infection Research (Germany), Epidemiological dynamics of SARS-CoV-2 lineages.

# Preliminary analysis of a COVID virus in R:



Shital Bhandary <shitalbhandary@pahs.edu.np>

to Shreekrishna, Ira ▾

Jan 10, 2021, 10:42 PM



Dear Shreekrishna sir and Ira,

Please look at this article from Bangladesh based on 3 whole genome sequence data. Let us discuss if we can do it too.

I have downloaded the most recent hcov-19 complete genome file from [GISAIID](#) website in FASTA format and imported/read in the R software 4.0.2 using the [SeqinR](#) package. I got following results:

```
attr("name")
[1] "hCoV-19/Nepal/70N/2020|EPI_ISL_754073|2020-08-03"
Interpretation:
Sample Identifier: hCoV-19/Nepal/70N/2020|EPI_ISL_754073
Sample collection date: 2020-08-03

1. > length(covnep1seq)
[1] 29509

2. > table(covnep1seq)
covnep1seq
   a    c    g    k    n    r    t    w    y
 8527 5271 5635  2 854  3 9213  1  3

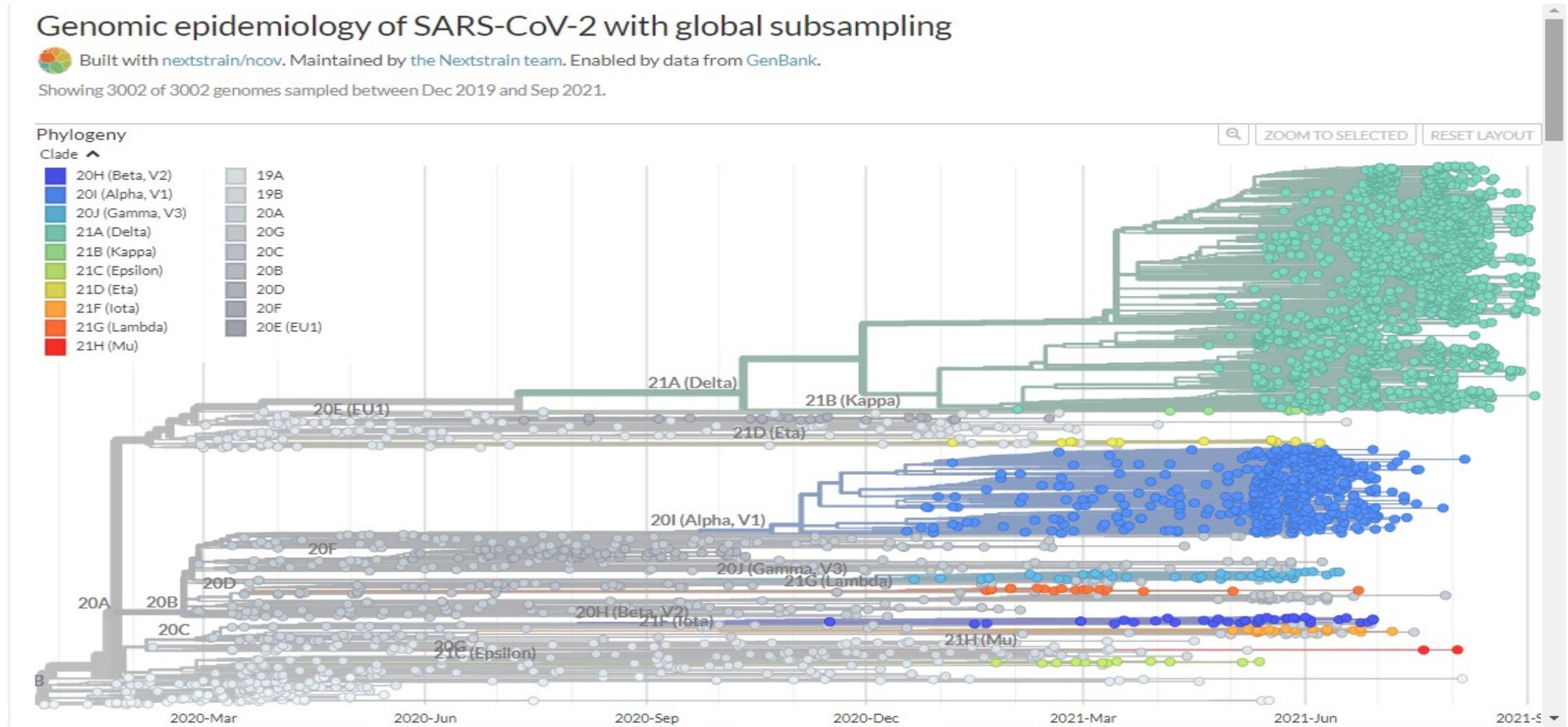
3. > GC(covnep1seq)
[1] 0.3807163
```

If you are interested in Bioinformatics then visit this site:

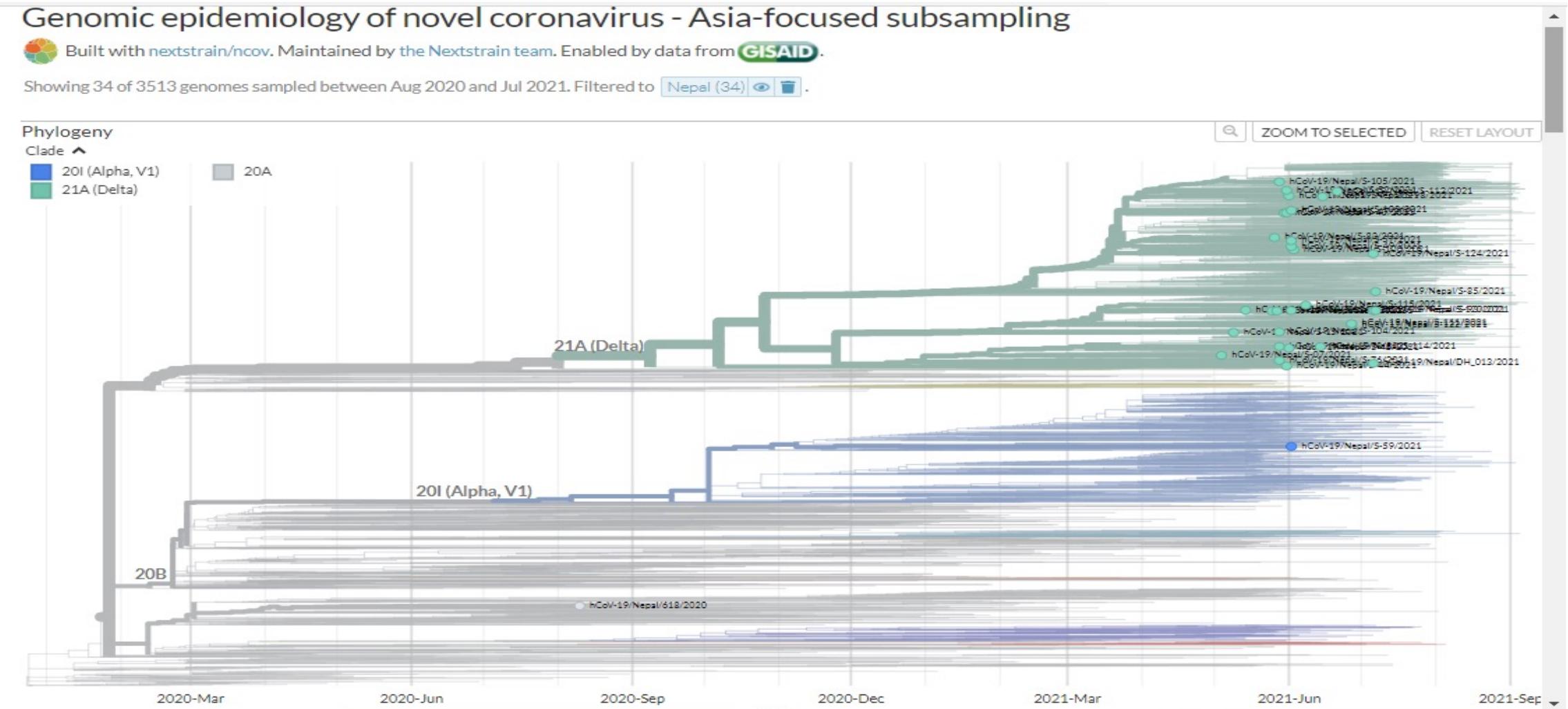
<https://a-little-book-of-r-for-bioinformatics.readthedocs.io/en/latest/src/chapter1.html>

[https://raunakms.github.io/#publications\\_selected](https://raunakms.github.io/#publications_selected)

# Or Use Nextstrain: Open Global sub-sampling!



Or Use NextStrain: SARS-COV-2 in Nepal



# Reading data in R/R Studio:

- Text files: R base, readr etc. #Already covered in Unit 1
- Excel files: readXL, openxls etc. #Already covered in Unit 1
- SPSS, Stata, SAS files: foreign, haven etc. #Already covered in Unit 1
- JSON files: rjson, jsonlite, RJSONIO etc.

Where, JSON = JavaScriptObjectNotation, used a lot in websites!

# Creating a JSON file using a Text Editor (Notepad):

[https://www.tutorialspoint.com/r/r\\_json\\_files.htm](https://www.tutorialspoint.com/r/r_json_files.htm)

- `Install.packages("rjson")`
- Create a JSON file in text editor e.g. notepad

```
{  
  "ID": ["1", "2", "3", "4", "5", "6", "7", "8"],  
  "Name": ["Rick", "Dan", "Michelle", "Ryan", "Gary", "Nina", "Simon", "Guru"],  
  "Salary": ["623.3", "515.2", "611", "729", "843.25", "578", "632.8", "722.5"],  
  "StartDate": ["1/1/2012", "9/23/2013", "11/15/2014", "5/11/2014", "3/27/2015",  
    "5/21/2013", "7/30/2013", "6/17/2014"],  
  "Dept": ["IT", "Operations", "IT", "HR", "Finance", "IT", "Operations", "Finance"]  
}
```

- save with .json extension e.g. `jason_data.json`

Read the created JSON file in R and Convert it as data.frame for further manipulation in R:

- library("rjson")
- data <- fromJSON(file = "jason\_data.json")  
    # jason\_data.json must be in the working directory of R!
- print(data)
  
- Covert to data frame:
- jason\_data\_frame <- as.data.frame(data)
- print(jason\_data\_frame)                          #Get summary, histogram of salary,  
  # Average salary by department  
  #Frequency distribution of all variables

# Reading JSON file from URL: Web API

<https://www.geeksforgeeks.org/working-with-json-files-in-r-programming/>

- `install.packages("jsonlite")` #Package “RJSONIO” also works!
- `library(jsonlite)`
- `Raw <- fromJSON("https://data.ny.gov/api/views/9a8c-vfzj/rows.json?accessType=DOWNLOAD")` #Large list!
- `food_market <- Raw[['data']]` #Large Matrix, 28472 rows and 24 columns!
- `Names <- food_market[,14]` #Large characters, Col 14 only!
- `heads(Names)` #Few names from Column 14!

# What more can you do with the food\_market data?

- Try: `table(Names)`
- Try: `table(V19)`
- Try: `table(farm_market$V19)`
- What is “atomic vector”?
- Convert the `farm_market` data to `data.frame` and get summary, create plots of all the “useful” variables and compute appropriate averages too!

# Web scrapping in R: A Simple (barebones) Example

<https://www.dataquest.io/blog/web-scraping-in-r-rvest>

- The recommended package for web scrapping in R is “rvest”
- `install.packages("rvest")`
- `library(rvest)`
- `simple <- read_html("https://dataquestio.github.io/web-scraping-pages/simple.html")`
- `simple %>%`  
 `html_nodes("p") %>%`  
 `html_text()`

# How to do complex web scrapping? Inspect the HTML elements!

The screenshot shows a web browser window with two tabs: 'Inbox (58) - shitalbhandary@gmail.com' and 'COVID-19 pandemic in Nepal - en.wikipedia.org'. The main content is the Wikipedia article 'COVID-19 pandemic in Nepal'. A context menu is open over the first paragraph, listing options like Back, Forward, Reload, Save as..., Print..., Cast..., Send to your devices, Create QR Code for this page, Translate to English, View page source, and Inspect. To the right of the article, there is a sidebar with a summary of COVID-19 statistics in Nepal.

**COVID-19 pandemic in Nepal**

From Wikipedia, the free encyclopedia

This article needs to be **updated**. The reason given for this is: This article is outdated.. Please help update this to reflect recent events or newly available information. You can edit this article, or you can discuss it on its talk page.

The **COVID-19 pandemic in Nepal** is part of the worldwide pandemic of coronavirus disease 2019 (COVID-19). The first case in Nepal was confirmed on 23 January 2020, a 22-year-old Chinese student, who had returned to Kathmandu from Wuhan on 9 January, tested positive for COVID-19 in South Asia.<sup>[3]</sup> Nepal's first case of local transmission was confirmed on 14 May. A country-wide lockdown came into effect on 24 March 2020, and ended on 15 May. The Ministry of Health and Population (MoHP) has confirmed a total of 783,910 cases and 11,028 deaths across 35 countries.<sup>[1]</sup> In the meantime, 4,085,796 real-time RT-PCR (RT-qPCR) tests have been conducted in the country.<sup>[1]</sup> The viral disease has been detected in all provinces and districts of the country, with Bagmati Province being the worst hit province and district respectively. As for Nepalese abroad, as of 26 July 2020, the Non-Resident Nepali Association has reported a total of 12,667 confirmed cases, 16,190 recoveries, and 161 deaths across 35 countries.<sup>[5]</sup>

Between January and March, Nepal took steps to prevent a widespread outbreak of the disease while preparing for it by procuring essential supplies, equipment and medicine, upgrading health infrastructure, training medical personnel, and spreading public awareness. Starting in mid-January, Nepal established health-desks at Tribhuvan International Airport as well as on border checkpoints with India. Land borders with India as well as China were later completely sealed off, and all international flights were suspended. All academic examinations were cancelled, and schools and colleges were closed. Quarantine centres and temporary hospitals are being set up across the country. Laboratory facilities are being upgraded and expanded. Hospitals have been setting up ICU units and

| COVID-19 pandemic in Nepal |   |
|----------------------------|---|
| Disease                    | COVID-19  |
| Virus strain               | SARS-CoV-2  |
| Location                   | Nepal   |
| First outbreak             | Wuhan, Hubei, China                                     |
| Index case                 | Kathmandu, Bagmati Province                             |
| Arrival date               | 9 January 2020<br>(1 year, 7 months, 4 weeks and 1 day) |
| Date                       | 23 January 2020   |
| Confirmed cases            | 783,910 (18 September) <sup>[1]</sup>                   |
| Active cases               | ▲ 25,082 (18 September) <sup>[1]</sup>                  |
| Recovered                  | 747,800 (18 September) <sup>[1]</sup>                   |
| Deaths                     | 11,028 (18 September) <sup>[1]</sup>                    |

# What to do now?

Inbox (58) - shitalbhandary@gmail.com x COVID-19 pandemic in Nepal - V +

en.wikipedia.org/wiki/COVID-19\_pandemic\_in\_Nepal

Not logged in Talk Contributions Create account Log in

1 The tidy text format | Text Mining with R https://www.tidytextmining.com/tidytext.html

Sources Network > 1 | Settings : X

Apps YouTube Maps Gmail How To Remove A... cov-glue.cvr.gla.ac... News Translate

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA The Free Encyclopedia

## COVID-19 pandemic in Nepal

From Wikipedia, the free encyclopedia

This article needs to be **updated**. The reason given is: Almost everything except the stat in the infobox is outdated.. Please help update this to reflect recent events or newly available information. (September 2021)

The **COVID-19 pandemic in Nepal** is part of the worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case in Nepal was confirmed on 23 January 2020 when a 31-year-old student, who had returned to Kathmandu from Wuhan on 9 January, tested positive for the disease.<sup>[2]</sup> It was also the first recorded case of COVID-19 in South Asia.<sup>[3]</sup> Nepal's first case of local transmission was confirmed on 4 April in Kailali District. The first death occurred on 14 May. A country-wide lockdown came into effect on 24 March 2020, and ended on 21 July 2020.<sup>[4]</sup> As of 18 September 2021, the Ministry of Health and Population

| COVID-19 pandemic in Nepal |   |
|----------------------------|---|
| Disease                    | COVID-19  |
| Virus strain               | SARS-CoV-2  |
| Location                   | Nepal   |
| First outbreak             | Wuhan, Hubei, China                                     |
| Index case                 | Kathmandu, Bagmati Province                             |
| Arrival date               | 9 January 2020<br>(1 year, 7 months, 4 weeks and 1 day) |
| Date                       | 23 January 2020   |
| Confirmed cases            | 783,910 (18 September) <sup>[1]</sup>                   |
| Active cases               | ▲ 25,082 (18 September) <sup>[1]</sup>                  |
| Recovered                  | 747,800 (18 September) <sup>[1]</sup>                   |

1 The tidy text format | Text Mining with R https://www.tidytextmining.com/tidytext.html

html.client-js.ve-available lang="en" dir="ltr">  
body class="mediawiki ltr sitedir-ltr mw-hide-empty-elt ns-0 ns-subject mw-editable page-COVID-19\_pandemic\_in\_Nepal rootpage-COVID-19\_pandemic\_in\_Nepal skin-vector action-view skin-vector-legacy">

<a id="top"></a>

This article needs to be **updated**. The reason given is: Almost everything except the stat in the infobox is outdated.. Please help update this to reflect recent events or newly available information. (September 2021)

# COVID-19 pandemic in Nepal

The **COVID-19 pandemic in Nepal** is part of the worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first case in Nepal was confirmed on 23 January 2020 when a 31-year-old student, who had returned to Kathmandu from Wuhan on 9 January, tested positive for the disease.<sup>[2]</sup> It was also the first recorded case of COVID-19 in South Asia.<sup>[3]</sup> Nepal's first case of local transmission was confirmed on 4 April in Kailali District. The first death occurred on 14 May. A country-wide lockdown came into effect on 24 March 2020, and ended on 21 July 2020.<sup>[4]</sup> As of 18 September 2021, the Ministry of Health and Population

| COVID-19 pandemic in Nepal |   |
|----------------------------|---|
| Disease                    | COVID-19  |
| Virus strain               | SARS-CoV-2  |
| Location                   | Nepal   |
| First outbreak             | Wuhan, Hubei, China                                     |
| Index case                 | Kathmandu, Bagmati Province                             |
| Arrival date               | 9 January 2020<br>(1 year, 7 months, 4 weeks and 1 day) |
| Date                       | 23 January 2020   |
| Confirmed cases            | 783,910 (18 September) <sup>[1]</sup>                   |
| Active cases               | ▲ 25,082 (18 September) <sup>[1]</sup>                  |
| Recovered                  | 747,800 (18 September) <sup>[1]</sup>                   |

# We/you need to scrap this data (table) in R: And create plots, get summaries etc.

Inbox (58) - shitalbhandary@gmail.com × COVID-19 pandemic in Nepal - V × +

en.wikipedia.org/wiki/COVID-19\_pandemic\_in\_Nepal

Apps YouTube Maps Gmail How To Remove A... cov-glue.cvr.gla.ac... News Translate 1 The tidy text form... Welcome to Text Mi... Reading list

Data [edit]

This article needs to be **updated**. Please help update this article to reflect recent events or newly available information. (July 2021)

The table below documents the daily growth and change of laboratory-confirmed COVID-19 cases, deaths, and recoveries and real-time RT-qPCR tests in Nepal, since the first confirmed case on 23 January 2020:

| Date   | Confirmed cases |     |        | Recoveries |     | Deaths |     | RT-PCR tests |     | TPR    | RR   | CFR           | Ref. |
|--------|-----------------|-----|--------|------------|-----|--------|-----|--------------|-----|--------|------|---------------|------|
|        | Total           | New | Active | Total      | New | Total  | New | Total        | New |        |      |               |      |
| 23 Jan | 1               | +1  | 1      | 0          | 0   | 0      | 0   |              |     | 0%     | 0%   | [170]         |      |
| 24 Jan | 1               | 0   | 1      | 0          | 0   | 0      | 0   |              |     | 0%     | 0%   |               |      |
| 25 Jan | 1               | 0   | 1      | 0          | 0   | 0      | 0   |              |     | 0%     | 0%   |               |      |
| 26 Jan | 1               | 0   | 1      | 0          | 0   | 0      | 0   |              |     | 0%     | 0%   |               |      |
| 27 Jan | 1               | 0   | 1      | 0          | 0   | 0      | 0   |              |     | 0%     | 0%   |               |      |
| 28 Jan | 1               | 0   | 1      | 0          | 0   | 0      | 0   | 3            |     | 33.33% | 0%   | [171]         |      |
| 29 Jan | 1               | 0   | 0      | 1          | +1  | 0      | 0   | 4            | +1  | 25%    | 100% | 0% [172][173] |      |
| 30 Jan | 1               | 0   | 0      | 1          | 0   | 0      | 0   | 5            | +1  | 20%    | 100% | 0% [174]      |      |
| 31 Jan | 1               | 0   | 0      | 1          | 0   | 0      | 0   | 5            | 0   | 20%    | 100% | 0% [175]      |      |
| 1 Feb  | 1               | 0   | 0      | 1          | 0   | 0      | 0   |              |     | 100%   | 0%   |               |      |
| 2 Feb  | 1               | 0   | 0      | 1          | 0   | 0      | 0   | 5            |     | 20%    | 100% | 0% [176]      |      |
| 3 Feb  | 1               | 0   | 0      | 1          | 0   | 0      | 0   |              |     | 100%   | 0%   |               |      |
| 4 Feb  | 1               | 0   | 0      | 1          | 0   | 0      | 0   | 14           |     | 7.14%  | 100% | 0% [177]      |      |

# Hint: I have already given the code!

- Also see these posts:
- <https://kyleake.medium.com/wikipedia-data-scraping-with-r-rvest-in-action-3c419db9af2d>
- <https://www.engineeringbigdata.com/web-scraping-wikipedia-world-population-rvest-r/>
- <https://stackoverflow.com/questions/33360634/how-to-scrape-data-from-wikipedia-using-r>

# What are the “ethical issues” with the use of web scrapping/ web APIs?

- Self-learning!
- <https://research.library.gsu.edu/c.php?g=1050939&p=7628916>
- <https://towardsdatascience.com/ethics-in-web-scraping-b96b18136f01>
- <https://blogs.mulesoft.com/api-integration/strategy/ethics-of-apis/>

# Question/Queries?

# Thank you!

@shitalbhandary