Progress:

Campy-COVID Project
Metrics generation

Steps

simply click to run!

www.BANDICAM.com

vjayaman/Metrics-CGM-ECC

github.com/vjayaman/Metrics-CGM-ECC

Search or jump to... / Pull requests Issues Marketplace Explore

vjayaman / Metrics-CGM-ECC Public

Pin | Unwatch 1 | Fork 0 | Star 0

Code | Issues | Pull requests | Actions | Projects 1 | Wiki | Security | Insights | Settings

main | 24 branches | 16 tags | Go to file | Add file | Code

vjayaman Renamed 7a and 7b 379c26b 1 hour ago 343 commits

ModificationNotes	Removed cumbersome files from ModificationNotes - saved copies else...	4 months ago
inputs	Adjusted some of the report code	1 hour ago
report_specific	Updated heatmap script - uses real names and saves heatmap clusterin...	3 months ago
scripts	Renamed 7a and 7b	1 hour ago
tests	Updated interval framework to include a set0 - works for months	4 months ago
Epiquant_Report.Rmd	Files as parameters, etc.	1 hour ago
README.md	Updated readme	2 months ago
collect_metrics.rscript	Renamed 7a and 7b	1 hour ago
directory_structure.exe	Directory re-structuring done but for heatmaps	10 hours ago
environment_setup.rscript	Directory re-structuring done but for heatmaps	10 hours ago
form.html	Directory re-structuring done but for heatmaps	10 hours ago

README.md

Imported to GitLab - cscscience setup. Will update here as well with each major change.

About

No description, website, or topics provided.

Readme
0 stars
1 watching
0 forks

Releases 11

Version4.2.8 Latest
1 hour ago

+ 10 releases

Packages

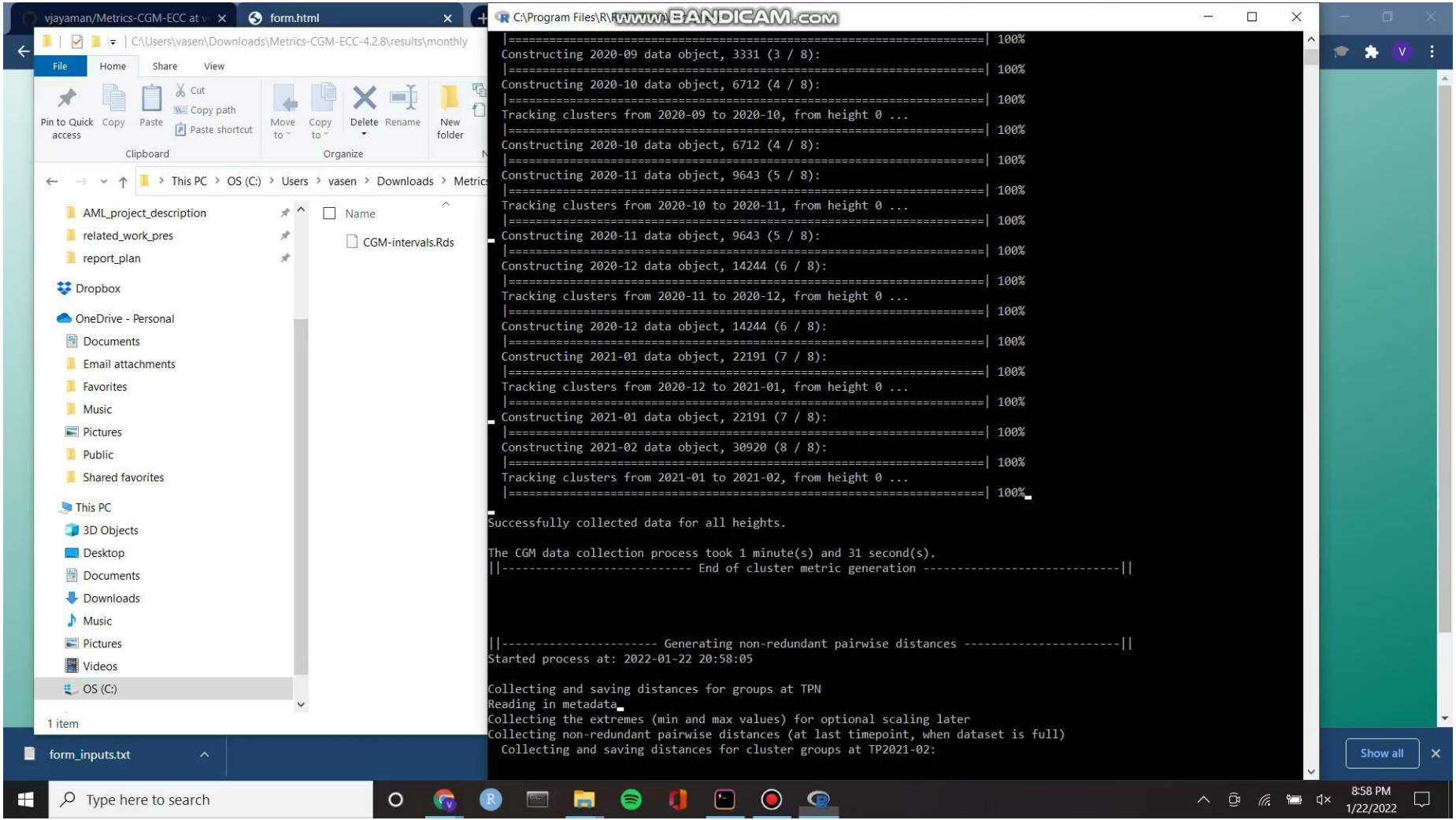
No packages published
Publish your first package

Languages

R 93.0% HTML 3.4% C++ 1.4% Other 2.2%

Type here to search

8:53 PM 1/22/2022


vjayaman/Metrics-CGM-ECC at v
form.html
C:\Program Files\R\R
www.BANDICAM.com
C:\Users\vasen\Downloads\Metrics-CGM-ECC-4.2.8\results\monthly
File
Home
Share
View
Pin to Quick access
Copy
Paste
Cut
Copy path
Paste shortcut
Clipboard
Move to
Copy to
Delete
Rename
New folder
Organize
This PC
OS (C:)
Users
vasen
Downloads
Metrics
AML_project_description
related_work_pres
report_plan
Dropbox
OneDrive - Personal
Documents
Email attachments
Favorites
Music
Pictures
Public
Shared favorites
This PC
3D Objects
Desktop
Documents
Downloads
Music
Pictures
Videos
OS (C:)
1 item
Name
CGM-intervals.Rds
|==========| 100%
Constructing 2020-09 data object, 3331 (3 / 8):
Constructing 2020-10 data object, 6712 (4 / 8):
Tracking clusters from 2020-09 to 2020-10, from height 0 ...
Constructing 2020-10 data object, 6712 (4 / 8):
Constructing 2020-11 data object, 9643 (5 / 8):
Tracking clusters from 2020-10 to 2020-11, from height 0 ...
Constructing 2020-11 data object, 9643 (5 / 8):
Constructing 2020-12 data object, 14244 (6 / 8):
Tracking clusters from 2020-11 to 2020-12, from height 0 ...
Constructing 2020-12 data object, 14244 (6 / 8):
Constructing 2021-01 data object, 22191 (7 / 8):
Tracking clusters from 2020-12 to 2021-01, from height 0 ...
Constructing 2021-01 data object, 22191 (7 / 8):
Constructing 2021-02 data object, 30920 (8 / 8):
Tracking clusters from 2021-01 to 2021-02, from height 0 ...
Successfully collected data for all heights.
The CGM data collection process took 1 minute(s) and 31 second(s).
||---------- End of cluster metric generation ----------||
||---------- Generating non-redundant pairwise distances ----------||
Started process at: 2022-01-22 20:58:05
Collecting and saving distances for groups at TPN
Reading in metadata
Collecting the extremes (min and max values) for optional scaling later
Collecting non-redundant pairwise distances (at last timepoint, when dataset is full)
Collecting and saving distances for cluster groups at TP2021-02:
form_inputs.txt
Show all
Type here to search
8:58 PM
1/22/2022

www.BANDICAM.com

Releases · vjayaman/Metrics-CGM | form.html

C:\Program Files\R\R-4.1.2\bin\Rscript.exe

```
||------------------------------ Starting heatmap generation ------------------------------||
Saving plots for cluster TP2_h000_c144 (158, size 12) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c092 (768, size 24) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c088 (134, size 26) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c073 (214, size 102) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c008 (1100, size 2096) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c003 (215, size 7020) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c004 (1159, size 9960) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...

Heatmaps saved to results/monthly/heatmaps

Epitables saved to results/monthly/epitables

Clustering done in the heatmaps saved to results/monthly/heatmaps/clustering

||------------------------------ Completed heatmap generation ------------------------------||


Press [enter] to close the window:
```

C:\Users\vasen\Downloads\Metrics-CGM-ECC-4.2.8\results\monthly

New item · Easy access · Properties · Open · Edit · History · Select all · Select none · Invert selection

Metrics-CGM-ECC-4.2.8 › results › monthly

Search monthly

	Date modified	Type	Size
	1/22/2022 8:48 PM	File folder	
	1/22/2022 8:28 PM	File folder	
	1/22/2022 8:48 PM	File folder	
	1/22/2022 8:44 PM	File folder	
sults	1/22/2022 8:23 PM	File folder	
ds	1/22/2022 8:21 PM	RDS File	44 KB
s	1/22/2022 8:18 PM	RDS File	37 KB
	1/22/2022 8:23 PM	RDS File	25 KB
ster_results.tsv	1/22/2022 8:23 PM	TSV File	4,085 KB
ster_results.tsv	1/22/2022 8:23 PM	TSV File	632 KB

10 items

form_inputs.txt

Show all

Type here to search

8:50 PM
1/22/2022

www.BANDICAM.com

Releases · vjayaman/Metrics-CGM | form.html

C:\Program Files\R\R-4.1.2\bin\Rscript.exe

```
||------------------------------ Starting heatmap generation ------------------------------||
Saving plots for cluster TP2_h000_c144 (158, size 12) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c092 (768, size 24) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c088 (134, size 26) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c073 (214, size 102) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c008 (1100, size 2096) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c003 (215, size 7020) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...
Saving plots for cluster TP2_h000_c004 (1159, size 9960) - after last TP ...
   Generating heatmap, density plot, and frequency histogram for temporal data ...
   Generating heatmap, density plot, and frequency histogram for geographical data ...
   Generating heatmap, density plot, and frequency histogram for (temp+geo)/2 data ...

Heatmaps saved to results/monthly/heatmaps

Epitables saved to results/monthly/epitables

Clustering done in the heatmaps saved to results/monthly/heatmaps/clustering

||------------------------------ Completed heatmap generation ------------------------------||



Press [enter] to close the window:
```

C:\Users\vasen\Downloads\Metrics-CGM-ECC-4.2.8\results\monthly

New item | Easy access | Properties | Open | Edit | History | Select all | Select none | Invert selection

Metrics-CGM-ECC-4.2.8 › results › monthly

Search monthly

	Date modified	Type	Size
	1/22/2022 8:48 PM	File folder	
	1/22/2022 8:28 PM	File folder	
	1/22/2022 8:48 PM	File folder	
	1/22/2022 8:44 PM	File folder	
sults	1/22/2022 8:23 PM	File folder	
ds	1/22/2022 8:21 PM	RDS File	44 KB
s	1/22/2022 8:18 PM	RDS File	37 KB
	1/22/2022 8:23 PM	RDS File	25 KB
ster_results.tsv	1/22/2022 8:23 PM	TSV File	4,085 KB
uster_results.tsv	1/22/2022 8:23 PM	TSV File	632 KB

* impli 10 items

form_inputs.txt

Show all

Type here to search

8:50 PM
1/22/2022

Epiquant_Report.Rmd

Knit on Save | Knit | Run

```
---
title: "SARS_CoV2 EpiQuant Analysis Report"
output:
  bookdown::html_document2:
    number_sections: FALSE
params:
  strain_results: "Merged_strain_results"
  lineage_file: "inputs/GISAID Lineages (770000 isolates).csv"
  long_file_dir: "multiset"
  vector_file: "inputs/Vector_Order.csv"
  country_code_file: "inputs/Country_Code_1.csv"
  epitables_dir: "results/monthly/epitables"
---

```{r setup, include=FALSE}

knitr::opts_chunk$set(echo = FALSE,message = FALSE,warning = FALSE,eval=TRUE)

```


```{r load-packages,warning = FALSE}
libs <- c("openxlsx", "reshape2", "plyr", "dplyr", "tidyr", "grid", "gridExtra",
          "ggplot2", "RColorBrewer", "gplots", "readr", "scatterplot3d",
          "heatmap3", "devtools", "knitr", "ggpubr", "ggrepel", "stringr",
          "magrittr", "tibble", "purrr", "summarytools", "lubridate", "tinytex",
          "png", "patchwork", "tidyverse", "ggridges", "bookdown", "foreign",
          "data.table", "forcats", "plotly")
y <- lapply(libs, require, character.only = TRUE); rm(libs); rm(y)

## for tables
pacman::p_load(
  rio,          # File import
  here,         # File locator
  skimr,        # get overview of data
  tidyverse,    # data management + ggplot2 graphics
  gtsummary,    # summary statistics and tests
  rstatix,      # summary statistics and statistical tests
  janitor,      # adding totals and percents to tables
  scales        # easily convert proportions to percents
```
```

Epiquant_Report.Rmd

Knit on Save | Knit | Run

```
4. ECC direction classes vs compass rose (see [Figure 4.4](#Fig.4.4-link))

```{r}
# IF RUNNING CHUNK BY CHUNK INTERACTIVELY, RUN THIS PART FIRST:
# params <- data.frame(strain_results = "Merged_strain_results",
#                      lineage_file = "inputs/GISAID Lineages (770000 isolates).csv",
#                      long_file_dir = "multiset",
#                      vector_file = "inputs/Vector_Order.csv",
#                      country_code_file = "inputs/Country_Code_1.csv",
#                      epitables_dir = "results/monthly/epitables")

# LOADING DATA
Lineages <- read_csv(params$lineage_file) %>%
  filter (T0 != "#N/A") %>% select("Strain", "T0.original","Pango_lineage")


ECC_long_base <- file.path("results", params$long_file_dir, params$strain_results) %>%
  list.files(., full.names = TRUE, pattern = "tsv") %>%
  map(read_tsv) %>% reduce(rbind)


ECC_Month_base <- file.path("results", "monthly", params$strain_results) %>%
  list.files(., full.names = TRUE, pattern = "tsv") %>%
  map(read_tsv) %>% reduce(rbind)
first_monthday <- grep("set0", ECC_Month_base$interval, value = TRUE) %>% unique() %>%
  as.character() %>% gsub("set0-", "", .)
ECC_Month_base$interval <- gsub('set0', first_monthday, ECC_Month_base$interval)
ECC_Month_base <- ECC_Month_base %>% separate(interval, into = c('Month_1', 'Month_2'), sep = 8)
ECC_Month_base <- merge(x = ECC_Month_base, y = Lineages, by = "Strain", all.x = TRUE)


ECC_Week <- file.path("results", "weekly", params$strain_results) %>%
  list.files(., full.names = TRUE, pattern = "tsv") %>%
  map(read_tsv) %>% reduce(rbind)

```
```

Epiquant_Report.Rmd

Knit on Save | Knit | Run

```
ECC_Week <- file.path("results", "weekly", params$strain_results) %>%
  list.files(., full.names = TRUE, pattern = "tsv") %>%
  map(read_tsv) %>% reduce(rbind)

ECC_Week <- ECC_Week %>%
  mutate(interval = gsub('set0', "00", interval),
         Week1  = isoweek(Date),
         Year_Week = strftime(ECC_Week$Date, format = "%G-%V")) %>%
  merge(x = ., y = Lineages, by = "Strain", all.x = TRUE) %>%
  mutate_if(is.numeric, ~ round(., digit = 2))

# create new variables Timepoints, Multistrains, Week and Running day
ECC_Week <- ECC_Week %>%
  mutate(Multistrain_Cluster = ifelse(ECC_Week$`TP2 cluster size (1)` > 2,"Multistrain","Singleton" ),
         Cluster_Size = ECC_Week$`TP2 cluster size (1)`,
         Geo_ECC = ECC_Week$TP2_ECC.0.0.1,Temp_ECC=ECC_Week$TP2_ECC.0.1.0,
         Delta_Geo_ECC = ECC_Week$delta_ECC_0.0.1,
         Delta_Temp_ECC=ECC_Week$delta_ECC_0.1.0,
         Cluster_Size1 = ECC_Week$`TP1 cluster size + 1 (2)`,
         Geo_ECC1 = ECC_Week$TP1_ECC.0.0.1,
         Temp_ECC1 = ECC_Week$TP1_ECC.0.1.0,
         "Cluster growth by size"=ECC_Week$`Actual cluster growth (TP2 size - TP1 size)`,
         "Novel_growth by rate"=ECC_Week$`Novel growth = (TP2 size) / (TP2 size - number of novels)`)


WX <- list.files(params$epitables_dir, full.names = TRUE)
Whisker_Data <- lapply(WX[3:5], function(x_i) {
  cluster_name <- strsplit(x_i, split = "/") %>% unlist() %>% extract2(4) %>% gsub(".Rds", "", .)
  readRDS(x_i) %>% add_column(Cluster = cluster_name)
}) %>% bind_rows() %>% select(Cluster, Temp.Dist, Geog.Dist) %>%
  separate(Cluster, sep = 13, into = c("Cluster", "Cluster1"))


country_code <- read_csv(params$country_code_file)
Vector_Order <- read_csv(params$vector_file)

# ------------------------------------------------------------------------------------------
# HOW TO READ IN THE ORIGINAL CLUSTER NAMES and collect cluster sizes: ---------------------
# ------------------------------------------------------------------------------------------
# tpn <- readRDS("inputs/processed/allTP2.Rds")
# lineages <- as.data.table(tpn$original[,c("Strain","T0")]) %>%
#   merge.data.table(tpn$new_cols[,c("Strain","0")], ., by = "Strain") %>%
#   set_colnames(c("Strain", "New_colname", "Original_colname"))
```

181:79 | Chunk 3 | R Markdown

Progress:

Campy-COVID Project
Metrics generation