# Take-home Exercise 1: Analysing and Visualising Spatio-temporal Patterns of COVID-19 in DKI Jakarta, Indonesia

This handout provides the context, the task, the expectation and the grading criteria of Take-home Exercise 1. Students must review and understand them before getting started with the take-home exercise.

## Contents

## Setting the Scene

Since late December 2019, an outbreak of a novel coronavirus disease (COVID-19; previously known as 2019-nCoV) was reported in Wuhan, China, which had subsequently affected 210 countries worldwide. In general, COVID-19 is an acute resolved disease but it can also be deadly, with a 2% case fatality rate.

The COVID-19 pandemic in Indonesia is part of the ongoing worldwide pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The virus was confirmed to have reached Indonesia on March 2, 2020. It started with two cases in March. As of July 31 2021, there had been 3,409,658 cumulative confirmed cases of COVID-19 in Indonesia and 94,119 reported cumulative deaths. All cases were spread in 34 provinces in Indonesia. Among all the provinces, DKI Jakarta (Indonesian: Daerah Khusus Ibukota Jakarta and in English: Special Capital Region of Jakarta) contributed close to 24% of the cumulative confirmed cases.

Despite its compactness, the cumulative confirmed cases were not evenly distributed within DKI Jakarta as Figure below reveals at the district and sub-district levels. The question was where were the sub-districts with relatively higher number of confirmed cases and how they changed over time.

# Objectives

The take-home exercise aims to embark you the skill of perform geospatial analysis to "Think Globally, Act Locally". You will also acquire the skills of documenting the analysis procedures conform to reproducible best practice using RMarkdown. Last but not least, you will also learn the skill of communicating your work via web post.

The specific objective of the exercise is to reveal the spatio-temporal patterns of monthly cumulative confirmed COVID-19 rate and death rate at *sub-district* or *kelurahan*.

# The Task

In this take-home exercise, you are tasked to reveal the spatia-temporal patterns of COVID-19 case at DKI Jakarta. The temporal interval must be at *month* (last day of the month) and the geographic should be at *sub-district* or *kelurahan* in Indonesia language. The specific tasks are as follow:

## Data Extraction, Wrangling and Integration

### APSTIAL DATA

- Download the daily COVID-19 data of DKI Jakarta,

- Extract ID_KEL, Nama_provinsi, nama_kota, nama_kecamatan nama_kelurahan, POSITIF (cumulative confirmed cases) and meninggal (cumulative death cases) from `data` worksheet of the daily COVID-19 data, and

- Integrate the daily data into a data frame by month.

- Download `DKI Jarkata` shapefile,

- Import the shapefile into R and save it in a simple feature data frame format, (Note: The national Projected Coordinates Systems of Indonesia is DGN95 / Indonesia TM-3 zone 54.1)

- Exclude all the outer islands from the DKI Jakarta sf data frame, and

- Retain the first nine fields in the DKI Jakarta sf data frame. The ninth field *JUMLAH_PEN = Total Population*.

- Combine the geospatial and aspatial data frame into simple feature data frame.

- Calculate the cumulative confirmed cases rate (i.e. cases per 10000 population) and the cumulative death rate by month.

## Thematic Mapping

- Plot maps to show the spatio-temporal distribution of cumulative confirmed cases rate and death rate at sub-direct level by using appropriate thematic mapping technique provided by tmap package.

## Analytical Mapping

- Plot analytical maps to show the spatio-temporal distribution of cumulative confirmed cases rate and death rate at sub-direct level by using appropriate thematic mapping technique provided by tmap package.

# The Data

For the purpose of this assignment, the following data should be used:

- Open Data Covid-19 Provinsi DKI Jakarta. This portal provides daily update of COVID-19 measures at both sub-district and district level. For the purpose of this exercise, data at the sub-district level will be used. This data is available in `data` worksheet.

- Indonesia Geospatial. This portal provides a comprehensive collection of geospatial data mainly in ESRI shapefile format at different geographical levels. For the purpose of this study, the **Shapefile (SHP) Batas Desa Provinsi DKI Jakarta** provided at PODES 2019 geospatial layer should be used.

# Grading Criteria

This exercise will be graded by using the following criteria:

- **Geospatial Data Wrangling (20 marks):** This is an important aspect of geospatial analytics. You will be assessed on your ability to employ appropriate R functions from various R packages specifically designed for modern data science such as readxl, tidyverse (tidyr, dplyr, ggplot2), sf just to mention a few of them, to perform the entire geospatial data wrangling processes, including. This is not limited to data import, data extraction, data cleaning and data transformation. Besides assessing your ability to use the R functions, this criterion also includes your ability to clean and derive appropriate variables to meet the analysis need. (Warning: All data are like vast grassland full of land mines. Your job is to clear those mines and not to step on them).

- **Geospatial Analysis (30 marks):** In this exercise, you are expected to use the appropriate thematic and analytics mapping techniques and R functions introduced in class to analysis the geospatial data prepared. You will be assessed on your ability to derive analytical maps by using appropriate rate mapping techniques.

- **Geovisualisation (20 marks):** In this section, you will be assessed on your ability to communicate the complex spatial statistics results in business friendly visual representations. This course is geospatial centric, hence, it is important for you to demonstrate your competency in using appropriate geovisualisation techniques to reveal and communicate the findings of your analysis.

- **Reproducibility (20 marks):** This is an important learning outcome of this exercise. You will be assessed on your ability to provide a comprehensive documentation of the analysis procedures in the form of code chunks of RMarkdown. It is important to note that it is not enough by merely providing the code chunk without any explanation on the purpose and R function(s) used.

- **Bonus (10 marks):** Demonstrate your ability to employ methods beyond what you had learned in class to gain insights from the data. The methods used must be geospatial in nature.

# Submission Instructions

- The write-up of the take-home exercise must be in **distill** or **blogdown** format. You are required to publish the write-up on **Netlify**.

- The R project of the take-home exercise must be pushed onto your Github repository.

- You are required to provide the links to Netlify service of the take-home exercise write-up and github repository on eLearn.

# Due Date

10th September 2021 (Friday), 11.59pm (midnight).

## Reference

Health Ministry of Indonesia. Published 2020. Accessed March 30, 2020. https://covid19.go.id/

JAKARTA's COVID-19 RESPONSE TEAM. Accessed August 12, 2021.

## Learning from senior

You are advised to review these sample submissions prepared by your seniors.

- Take-home Exercise 1: Geographic Analysis of the Supply and Demand of Childcare Services in Singapore by Xiao Rong Wong

- Take-home Exercise 1: Understanding and Analysing Changes in the Supply of Childcare Services by Rajiv Abraham Xavier

## Q & A

### Q1:

I noted that the data link you provided will lead us to a webpage of daily updates (https://riwayat-file-covid-19-dki-jakarta-jakartagis.hub.arcgis.com/).

When you mention to "Download the daily COVID-19 data of DKI Jakarta" Do you mean to do this manually or are we to use a google drive connector since all of these links are linked to a google drive sheet.

> *Prof. Kam: You are expected to download the file one by one. The data are cumulative. Hence, to download the data by month, you only need to download every last day of the month. In this case, there are only 19 files to download (assuming that 31 January 2020 - 31 July 2021). I have downloaded the data manual to try out myself and it won't take more than 15minutes to download all of them. In fact, for your information, I have downloaded all the daily data until yesterday. This is the job of a data analyst* 😔

### Q2:

In addition, when you mention "Integrate the daily data into a data frame by month" Do you mean we have to download multiple sheets and combine the data? If so does this mean we need to download for the last 30 days? or multiple months?

> *Prof Kam: Refer to earlier reply above.*

(I am unsure if this is the case or whether you refer to a specific feature. Unfortunately i do not understand the data at all given that most of the features are in a foreign language)

> *Prof. Kam: Yes, you have to download multiple workbooks and extract the worksheet and columns you need and then combine them into an integrated data file. This is what we called data wrangling in Data Science. You will learn how to do it using readxl, tydyr and dplyr package. With regard the foreign language, less worry because translation will be provided. In will be included in the revised version of the Take-home Exercise 1 handout latest by tomorrow morning 9.00am.*

## Q3:

Lastly, when you mention "Download the DKI Jarkata shapefile" Is this manual downloading or do we need to dynamically download by code and extract and import from there?

> *Prof. Kam: Personally, I think download manually is easier because I just need to click on the link provided. After downloading, you are required to use sf package of r to parse the data into R and save it as a simple feature data frame. You will learn more on how to work with R package in next week lesson.*

## Q4:

In this case, there are only 19 files to download (assuming that 31 January 2020 - 31 July 2021). On the webpage of daily updates (https://riwayat-file-covid-19-dki-jakarta-jakartagis.hub.arcgis.com/), I am only able to see the data file starting from 25 March 2020 onwards. In addition, the link to access the data file for 31 January 2021 is broken.

> *Prof Kam: Since the data only starts on 25 March 2020, then there is nothing wrong the first month data is on 31st March 2020. Likewise if 31st January 2021 is not available, there is nothing wrong 30th January 2021 data set is used. More important, you are required to document these two issues on your report.*

## Q5:

Some of the excel files had duplicate columns with the same name. For the `Meninggal` column, can I confirm that we look at the column without N/A values?

> *Prof Kam: We are interested in analysing the geographical distribution of **Cumulative Death**. In this*

> *Prof Kam: We are interested in analysing the geographical distribution of* **Cumulative Death**. *In this case, the* `Meninggal` *without NA values should be used.*

# Peer Learning

---

[AVULA NIHARIKA](#)

[CHIN SHAN PEI SARAH](#)

[DARRYL KWOK BING HENG](#)

[GOH JIA XIN GENICE](#)

[JEROME GOH TING CHUAN](#)

[KWEK YI CHEN](#)

[MEGAN SIM TZE YEN](#)

- **What is so special**: The data important, extracting, wrangling and processing process were appropriately performed and comprehensively documented. The choropleth maps were well designed. All process are highly reproducible.

[NGAH XIN YAN](#)

[NOR AISYAH BINTE AJIT](#)

- **What is so special**: Appropriate used of functions to reduce copy-paste of codes. Ability to apply analytical mapping techniques appropriately.

[TEO JUN PENG](#)

[TOH JUN LONG](#)

[WONG WEI LING](#)

[XU PENGTAI](#)

[YU YILING](#)