# Collaborative Data Science and COVID-19

Assoc Prof Peter Julian Cayton, PhD

School of Statistics, University of the Philippines Diliman

Member, UP COVID-19 Pandemic Response Team

Member, LEADS for Health Security and Resilience

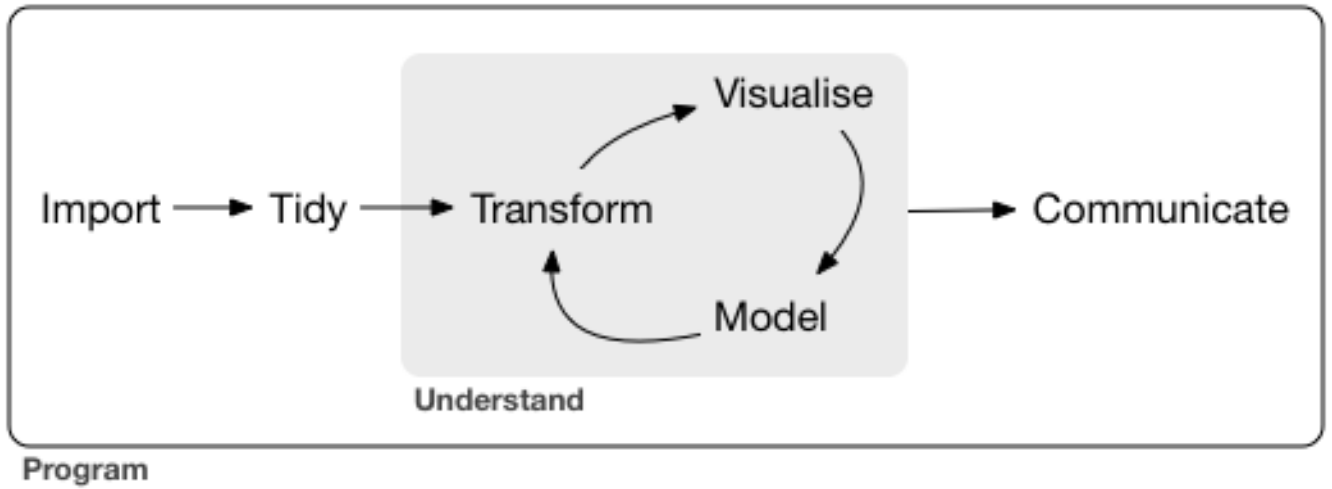



# Flow of the Presentation

- A Typical Data Science/Analytics Workflow
- Membership to Data-Driven Teams
- Example Projects
- Demonstrating the Workflow with a Team using Example Projects
- Closing Remarks

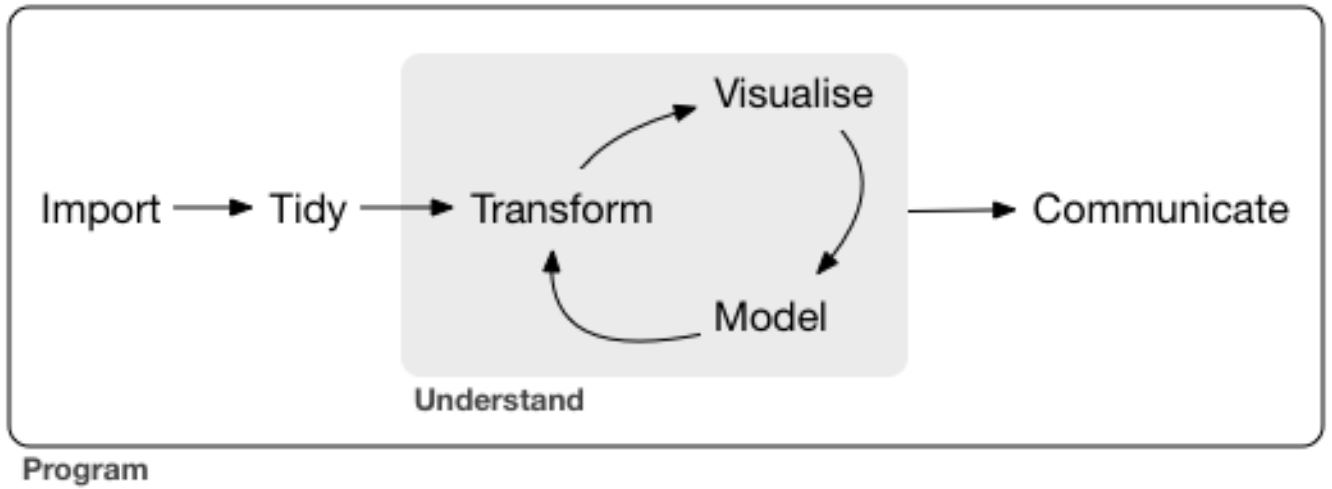
# A Typical Data Science/Analytics Workflow

# A Typical Data Science/Analytics Workflow



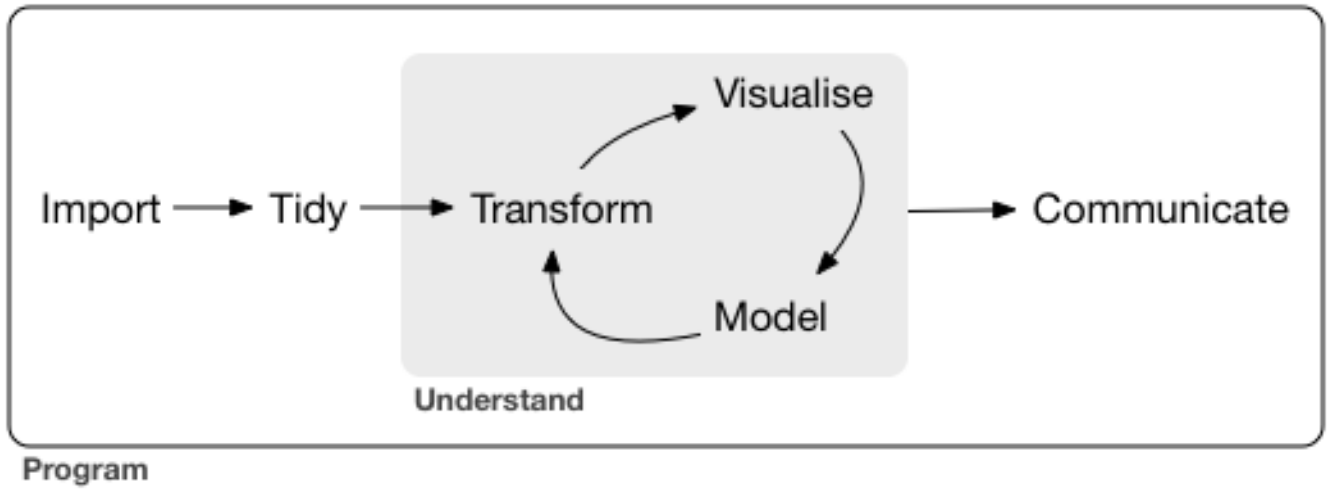

- Source: Wickham, H and Grolemund, G (2017). R for Data Science. O'Reilly. https://r4ds.had.co.nz/introduction.html

# A Typical Data Science/Analytics Workflow



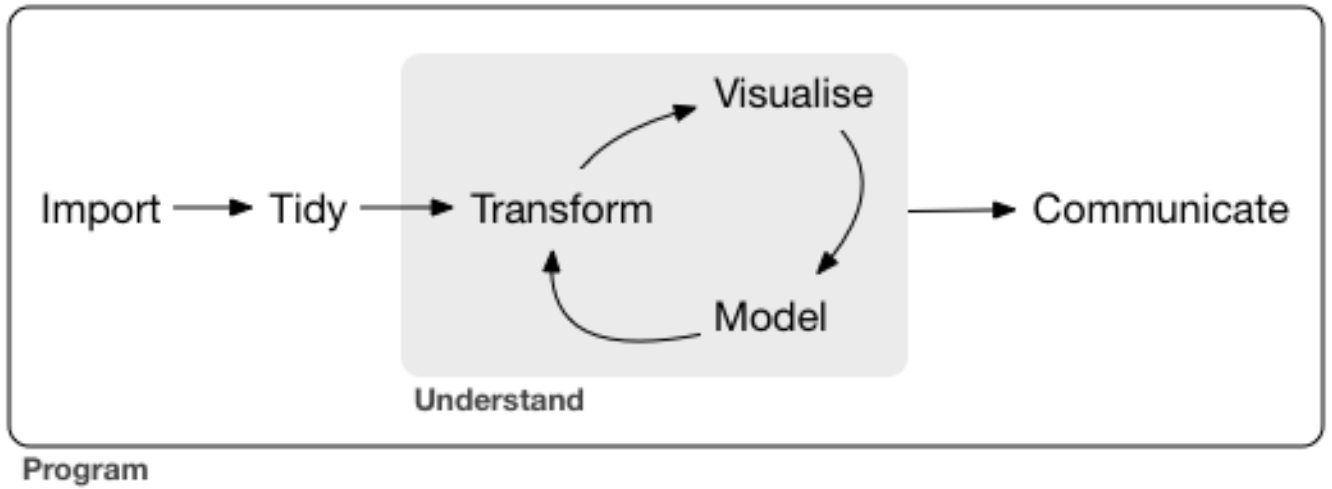

- Import
  - Extracting the data from an internal database, a file, an online website, or thru a web application programming interface (API), to be loaded in R/RStudio

# A Typical Data Science/Analytics Workflow



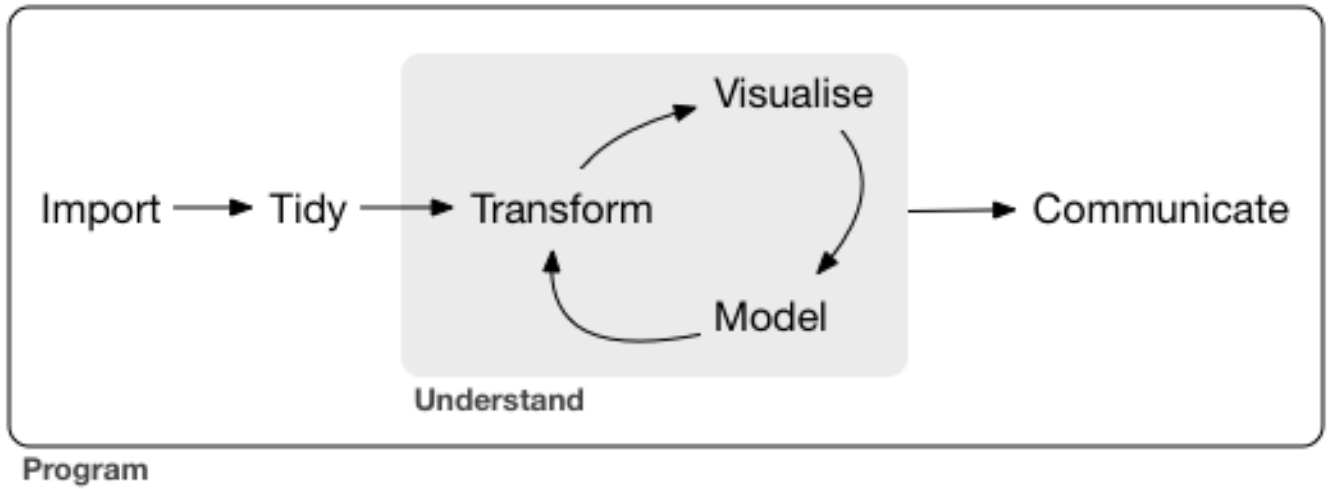

- Tidy
  - Arranging the data into a neat data structure, with variable as columns and data points as rows.
  - Included in this step would be data cleaning, data augmentation, missing data imputation, and others

# A Typical Data Science/Analytics Workflow



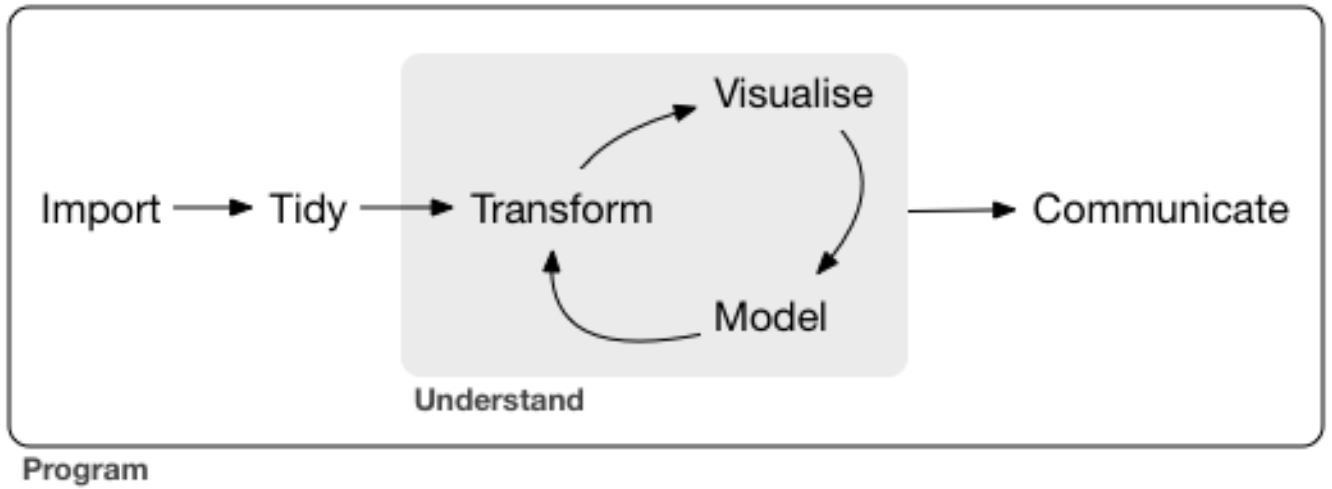

- Understand
  - Generally, the steps to extract insights from data after tidying up.

# A Typical Data Science/Analytics Workflow



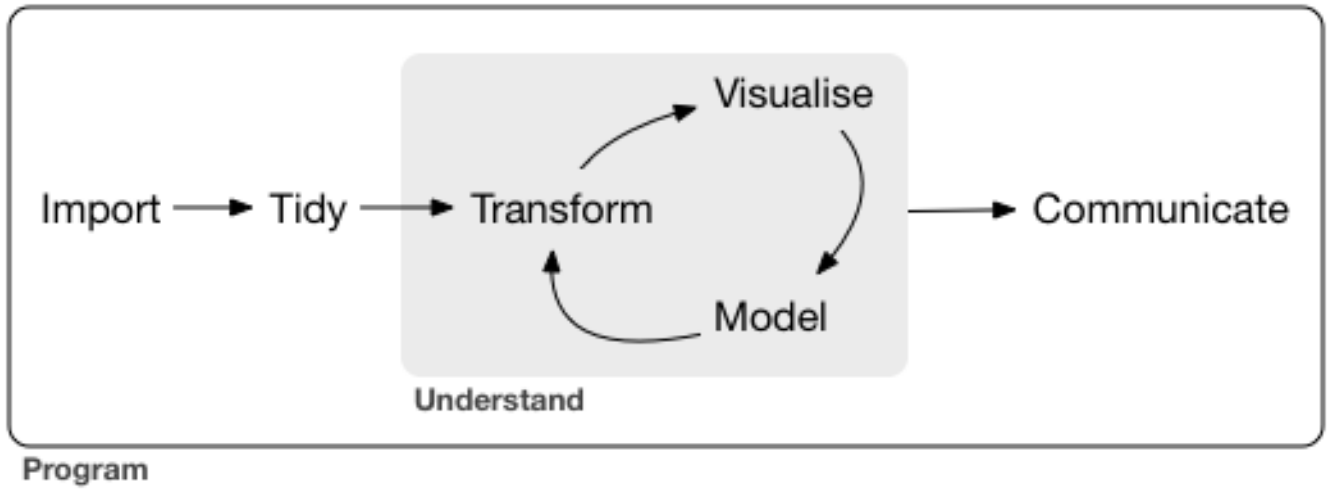

- Transform
  - Processing the data in preparation for further steps. Examples are:
    - Narrowing the data, e.g., by region or by age,
    - Computing new variables, e.g., length of days until recovery, or delays in reporting cases
    - Aggregating data, e.g., counting cases or solving rates/means

# A Typical Data Science/Analytics Workflow



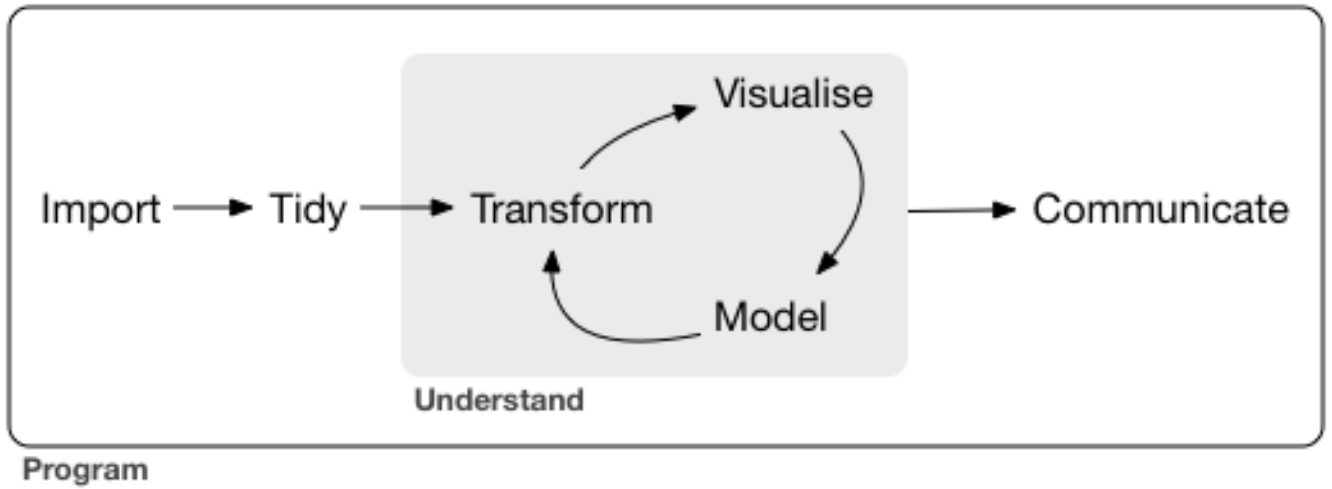

- Visualize
  - Plot data into graphs so that patterns and features may be explored and insights be extracted from what is seen.

# A Typical Data Science/Analytics Workflow



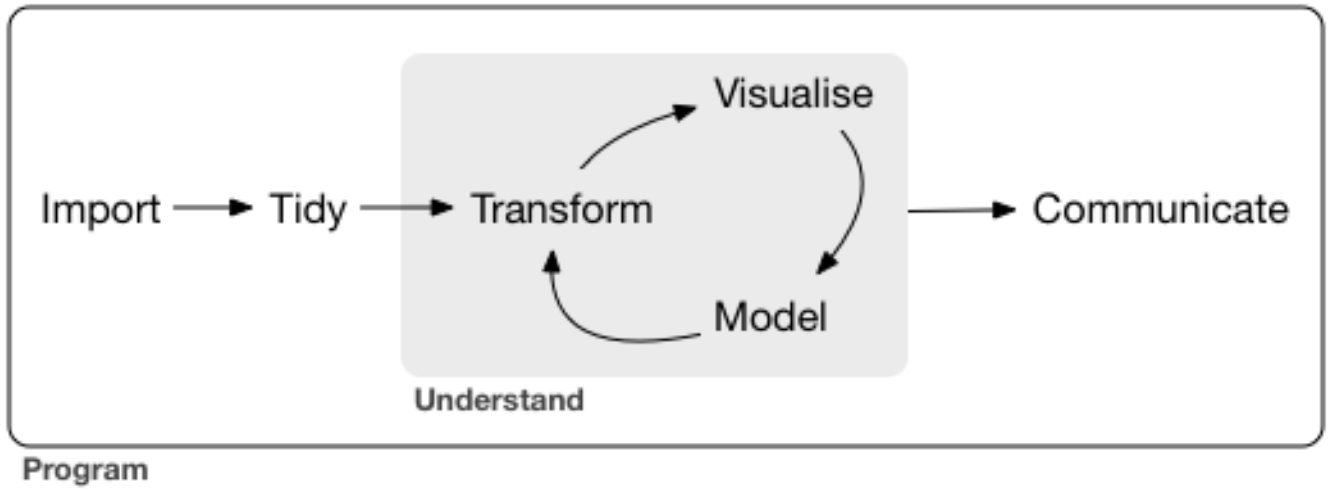

- Model
  - When necessary, models help in summarizing the complex relationships and the patterns found from visualizations
  - Designing models for prediction or forecasting

# A Typical Data Science/Analytics Workflow




- Communicate
  - These include:
    - Writing reports, creating dashboards, making presentations, compilations, etc.

# A Typical Data Science/Analytics Workflow




- Program
  - All these processes to be encapsulated in a data science project plan
  - Possible to be encapsulated in one software, but it's not impossible to use more than one depending on team members' capabilities to process and analyze data.

# Membership to Data-Driven Teams



# About the UP Pandemic Response Team

# About UP COVID-19 Pandemic Response Team

- Leaders:
  - Dr Teodoro Herbosa, UP System
  - Dr Alfredo Mahar Lagmay, Executive Director, UP Resilience Institute
- Who we are
  - 200+ experts and volunteers from the whole UP System, from Baguio to Davao
  - spanning multiple fields: political scientists, statisticians, mathematicians, geographers, geologists, medical doctors, linguists, economists, etc.
- Next slides from Dr Lagmay: Our Portfolio

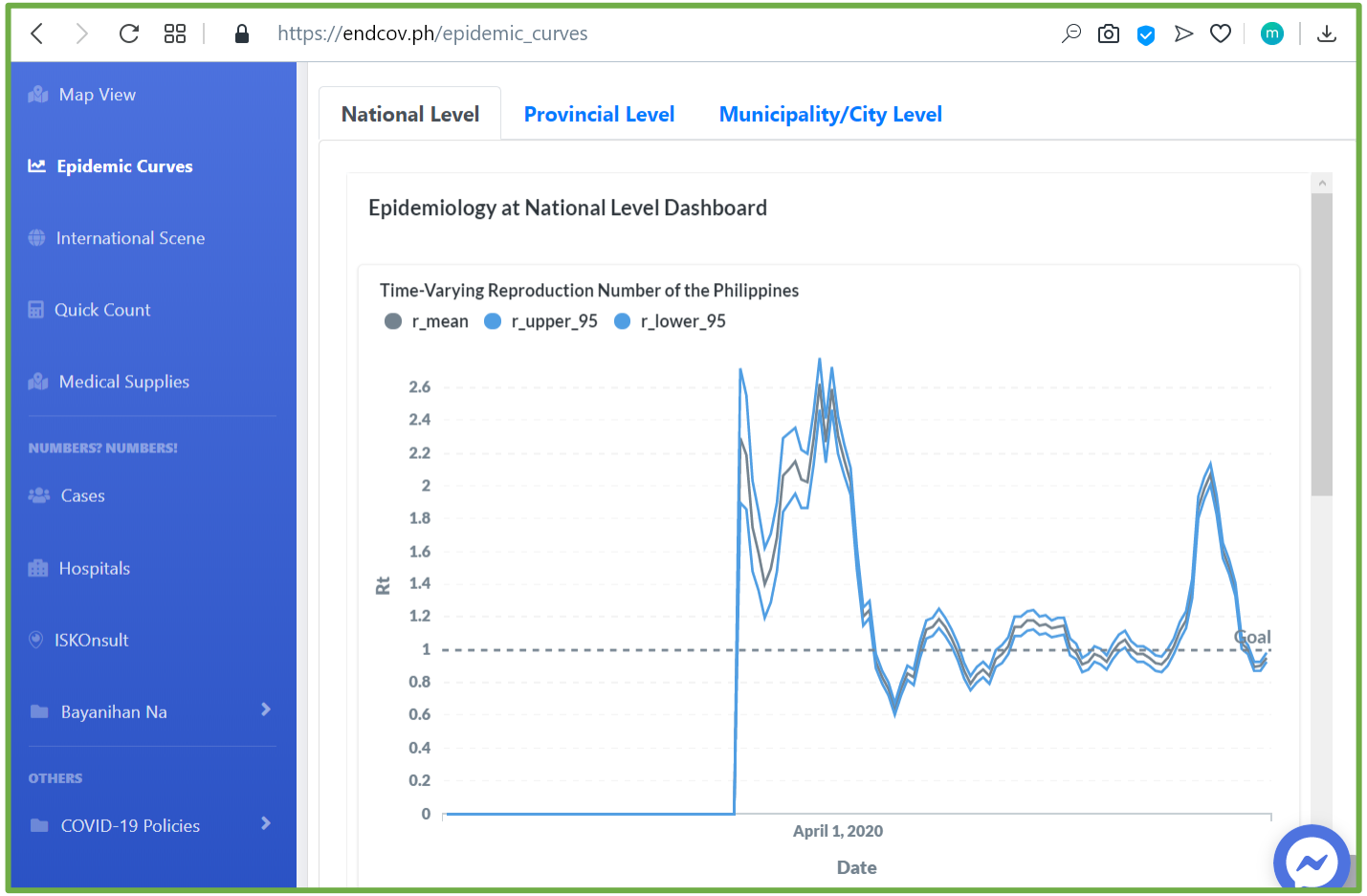
UP PANDEMIC RESPONSE TEAM
Prompt. Reliable. Transparent.

# endcov.ph

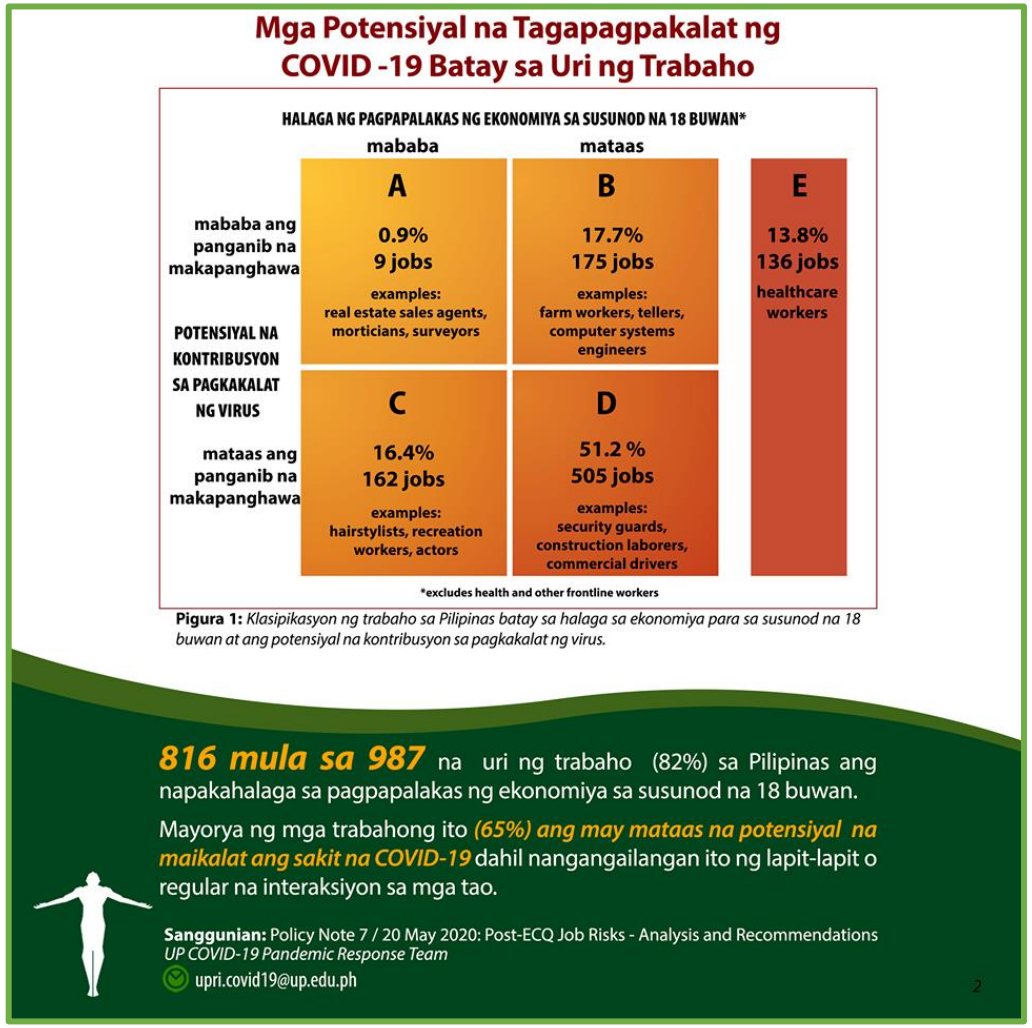
**University of the Philippines**
April 2 ·

The UP COVID-19 Pandemic Response Team has created a web portal (http://endcov.ph/) mapping out COVID-19 cases in the country in clear, useful detail.

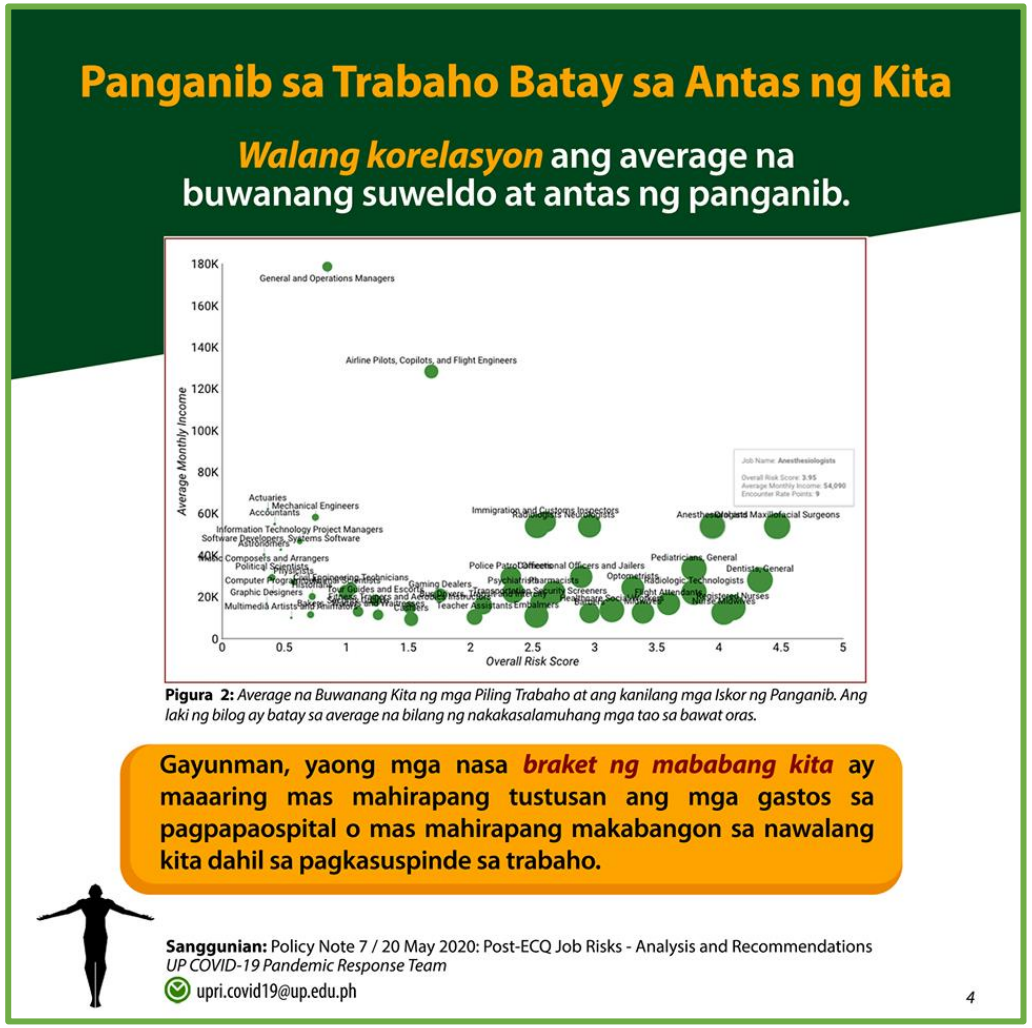
Read more here: https://www.up.edu.ph/up-launches-web-portal-map-for-covid.../

# Yani, the Endcov bot

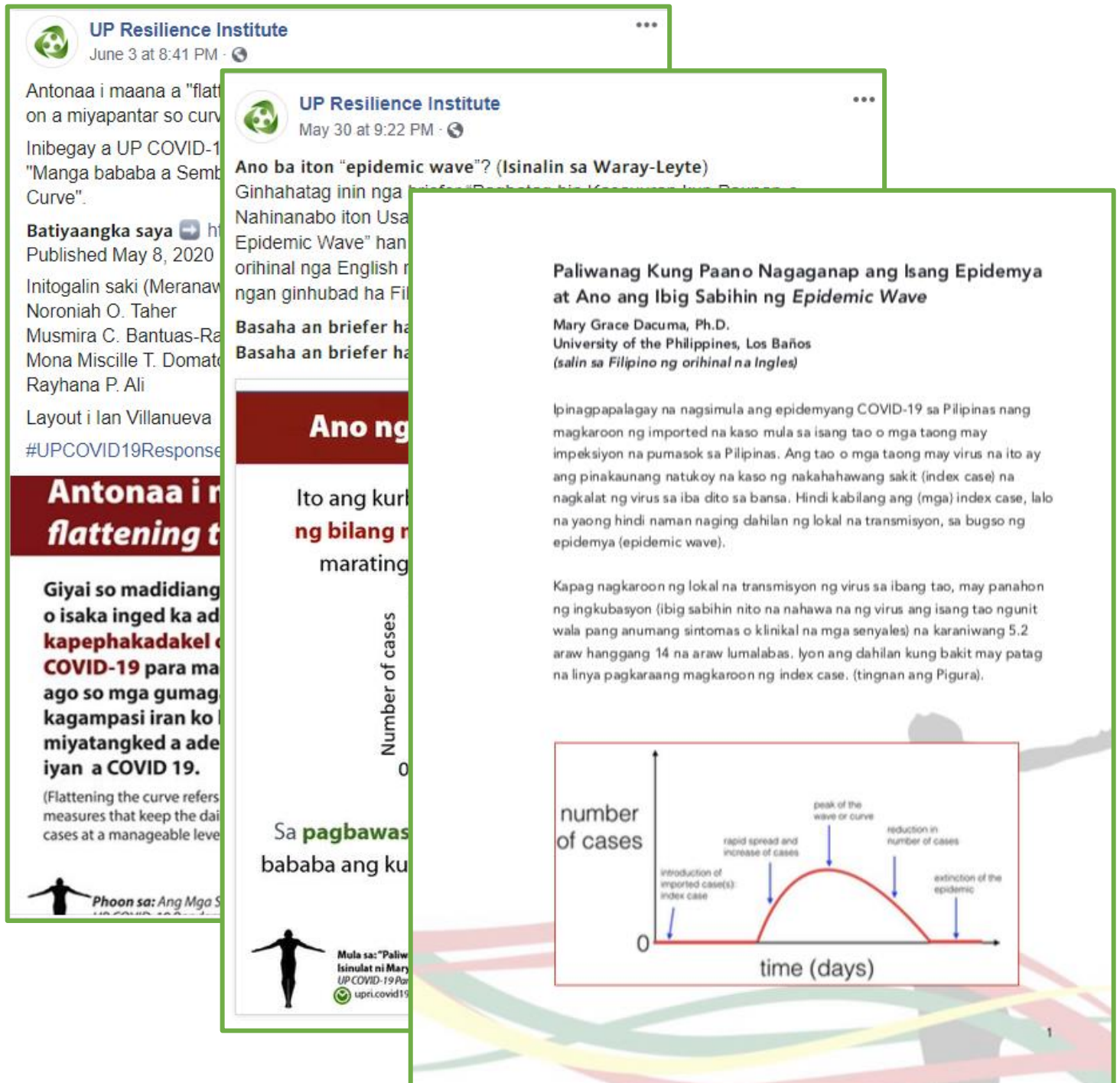



**Pigura 1:** *Klasipikasyon ng trabaho sa Pilipinas batay sa halaga sa ekonomiya para sa susunod na 18 buwan at ang potensiyal na kontribusyon sa pagkakalat ng virus.*



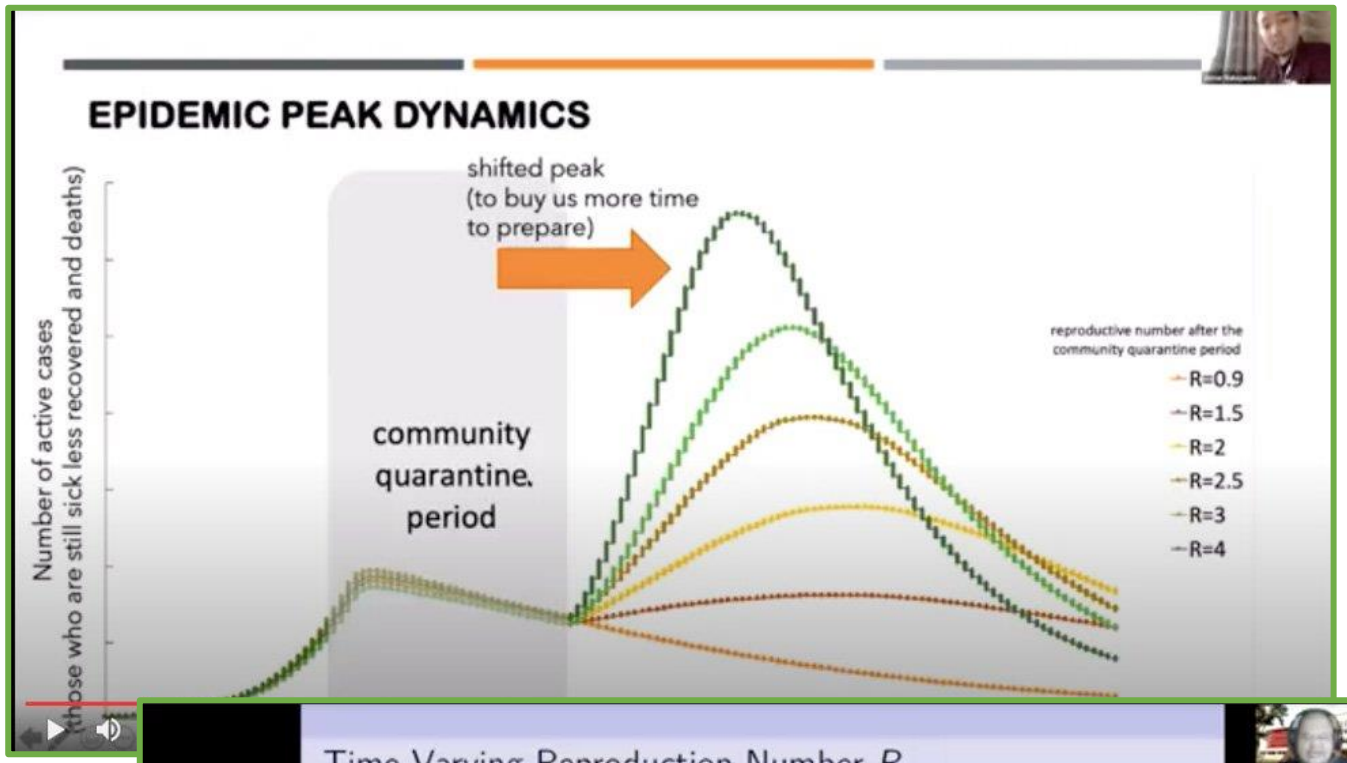

**Pigura 2:** *Average na Buwanang Kita ng mga Piling Trabaho at ang kanilang mga Iskor ng Panganib. Ang laki ng bilog ay batay sa average na bilang ng nakakasalamuhang mga tao sa bawat oras.*

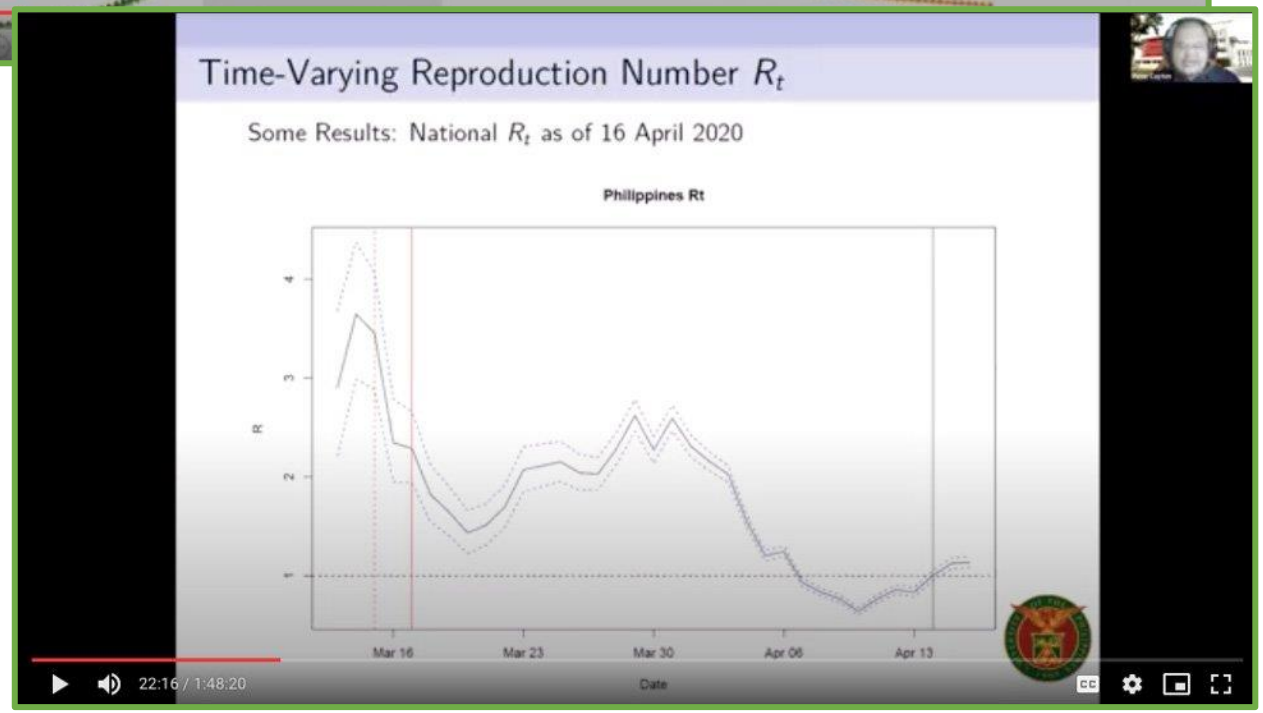
To better inform the general public, the team has also redesigned their studies into **animation, video formats and flashcards** to be more accessible.




**Policy notes** are now available in English as well as Tagalog, Ilokano, Bikol Sentral, Waray, Cebuano, Hiligaynon, Aklanon, Kapampangan, Itawis, Chavacano de Zamboanga, Meranaw, and Bahasa Sug.

# COVID-19 Threshold Model



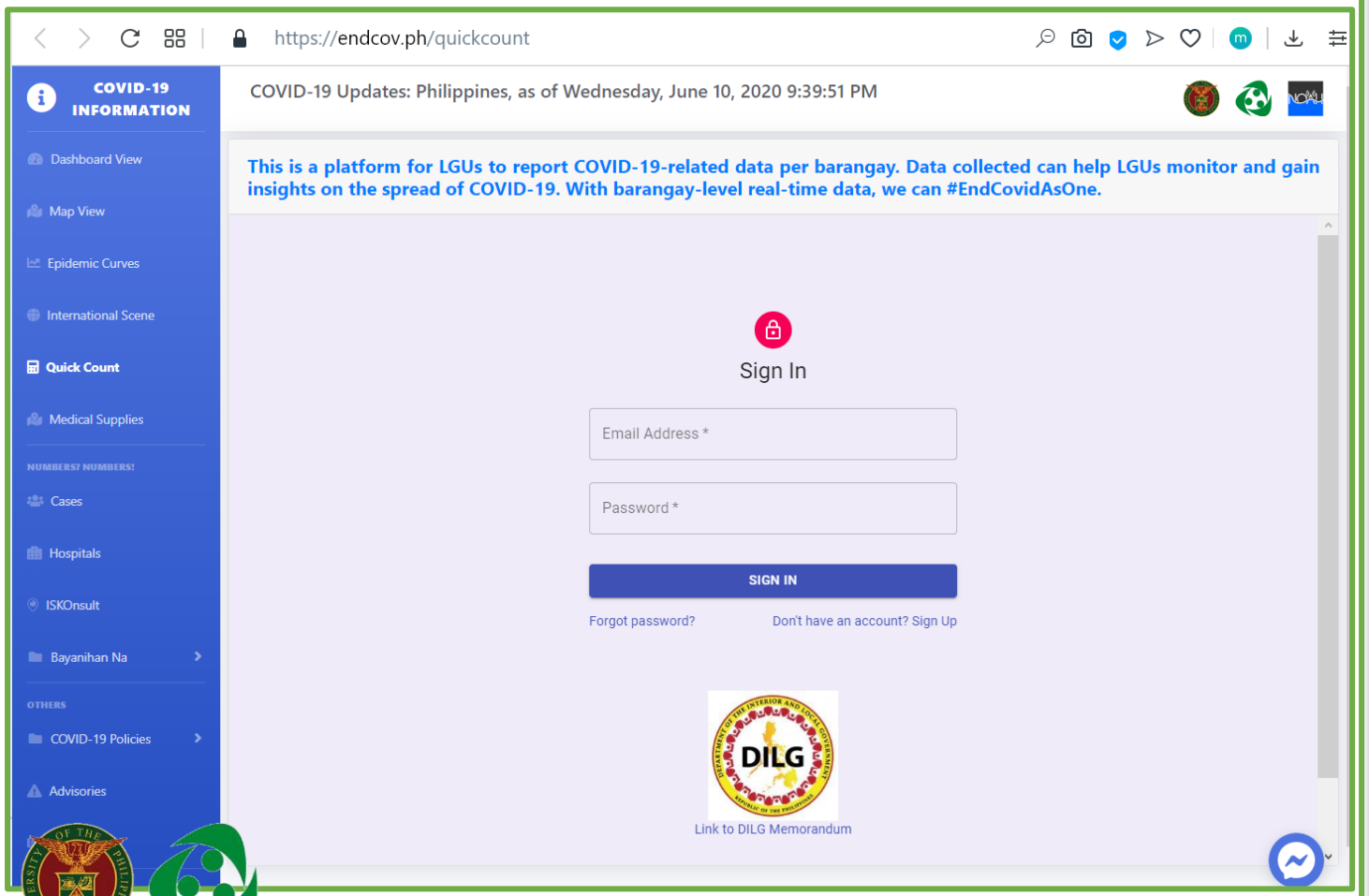



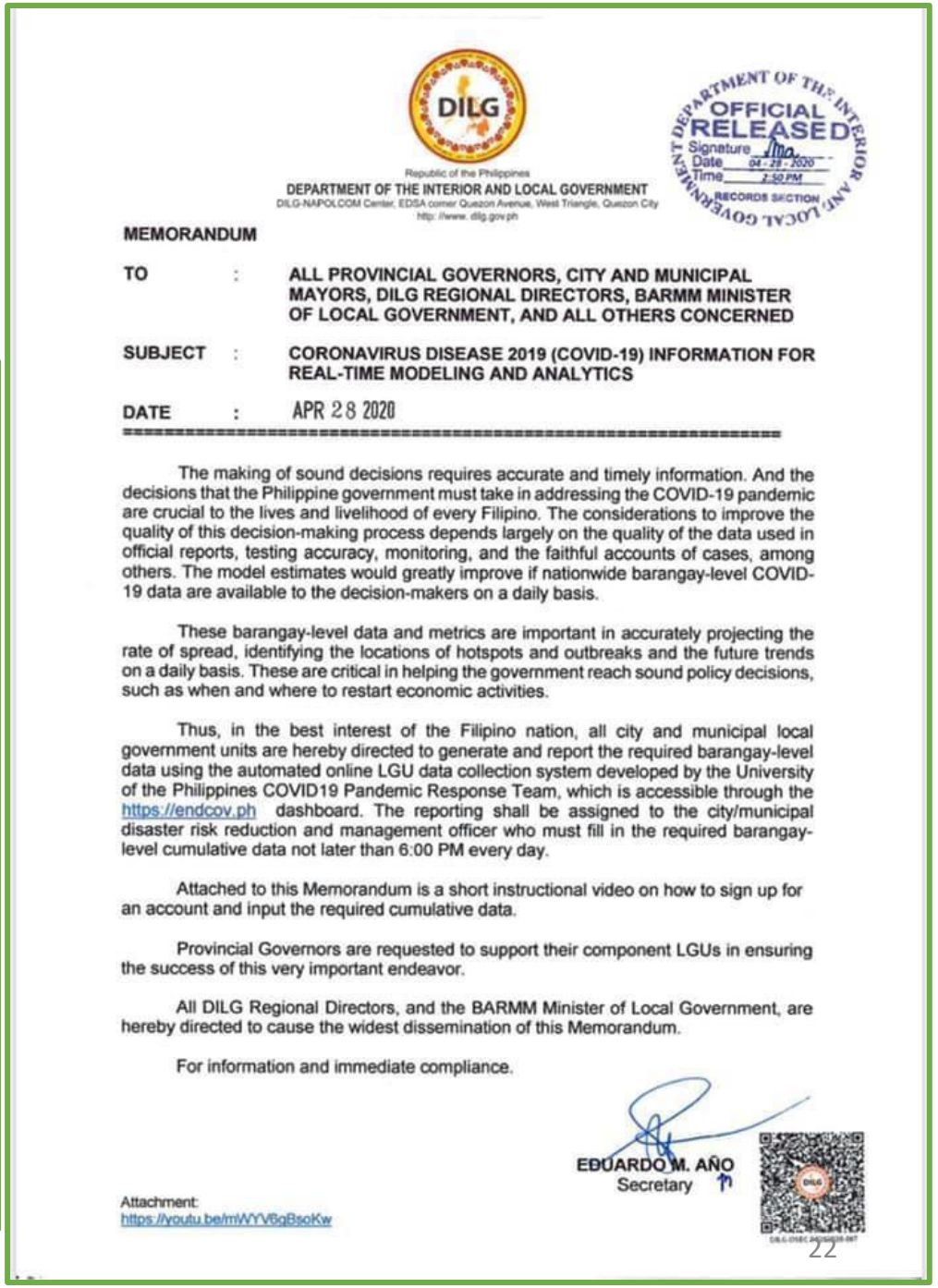

The response team has drafted a number of **policy notes** based on the results of their studies, including recommendations for a graduated activation of the ECQ that depends on the level of risk per area.

# Working with the national government

Team members have also **given presentations to the President and the Inter-Agency Task Force on Emerging Infectious Disease (IATF)**. Given their highly specialized competencies, members have been called to join the IATF Technical Working Group on Anticipatory and Forward Planning.

# Coordination with the LGUs



Republic of the Philippines
DEPARTMENT OF THE INTERIOR AND LOCAL GOVERNMENT
DILG-NAPOLCOM Center, EDSA corner Quezon Avenue, West Triangle, Quezon City
http://www.dilg.gov.ph

OFFICIAL RELEASED
Signature
Date 04-28-2020
Time 2:50 PM
RECORDS SECTION
DEPARTMENT OF THE INTERIOR AND LOCAL GOVERNMENT

**MEMORANDUM**

**TO** : **ALL PROVINCIAL GOVERNORS, CITY AND MUNICIPAL MAYORS, DILG REGIONAL DIRECTORS, BARMM MINISTER OF LOCAL GOVERNMENT, AND ALL OTHERS CONCERNED**

**SUBJECT** : **CORONAVIRUS DISEASE 2019 (COVID-19) INFORMATION FOR REAL-TIME MODELING AND ANALYTICS**

**DATE** : APR 28 2020

The making of sound decisions requires accurate and timely information. And the decisions that the Philippine government must take in addressing the COVID-19 pandemic are crucial to the lives and livelihood of every Filipino. The considerations to improve the quality of this decision-making process depends largely on the quality of the data used in official reports, testing accuracy, monitoring, and the faithful accounts of cases, among others. The model estimates would greatly improve if nationwide barangay-level COVID-19 data are available to the decision-makers on a daily basis.

These barangay-level data and metrics are important in accurately projecting the rate of spread, identifying the locations of hotspots and outbreaks and the future trends on a daily basis. These are critical in helping the government reach sound policy decisions, such as when and where to restart economic activities.

Thus, in the best interest of the Filipino nation, all city and municipal local government units are hereby directed to generate and report the required barangay-level data using the automated online LGU data collection system developed by the University of the Philippines COVID19 Pandemic Response Team, which is accessible through the https://endcov.ph dashboard. The reporting shall be assigned to the city/municipal disaster risk reduction and management officer who must fill in the required barangay-level cumulative data not later than 6:00 PM every day.

Attached to this Memorandum is a short instructional video on how to sign up for an account and input the required cumulative data.

Provincial Governors are requested to support their component LGUs in ensuring the success of this very important endeavor.

All DILG Regional Directors, and the BARMM Minister of Local Government, are hereby directed to cause the widest dissemination of this Memorandum.

For information and immediate compliance.

**EDUARDO M. AÑO**
Secretary

Attachment:
https://youtu.be/mWYV6gBsoKw

Aside from working with national and local government units, the UP response team has been collaborating with academics from the National University of Singapore, University of California Davis and University College London as well as local academic institutions.

On the international front, the team has joined the Forecast-based Warning, Analysis and Response Network (FOREWARN), an organization of academics, scientists and humanitarian workers.

# About UP COVID-19 Pandemic Response Team

- Projects I am Involved In:

1. Epidemic Curves of the UP COVID-19 Pandemic Response Team
2. ENDCOV Map on current active cases and probability of outbreak
3. LGU Data Analytics Group
4. Compendium of COVID-19 Statistics for Island Groups, Regions, Provinces, Cities, and Municipalities
5. PSPHP Graphs, Time Varying-R Dashboard with the LEADS 4 Health Security and Resilience

UP PANDEMIC RESPONSE TEAM
Prompt. Reliable. Transparent.

# About UP COVID-19 Pandemic Response Team

- Platform:
  - https://endcov.up.edu.ph
- Contact Us:
  - upri.covid19@up.edu.ph

UP PANDEMIC RESPONSE TEAM
Prompt. Reliable. Transparent.


# About the L4H Consortium

# About the L4H Consortium

- Leading Evidence-based Actions through Data Science for Health Security and Resilience
- A consortium of data scientists, physicians, mathematicians, and epidemiologists, convened by the Philippine Society of Public Health Physicians (PSPHP)
- website: https://leads4health.org/l4h/

# About the L4H Consortium

# About the L4H Consortium

## Our Members

- **Jason V. Alacapa, MD, MBA, MPH, MHM** - Consortium Co-convenor, Philippine Society of Public Health Physicians; CEO, metaHealth Insights and Innovation, Inc.
- **Geminn Louis C. Apostol, MD, MBA** - Assistant Professor and Environmental Health Specialist, Ateneo School of Medicine and Public Health
- **DJ Darwin R. Bandoy, DVD, PhD(c)** - Assistant Professor, University of the Philippines
- **Peter Julian A. Cayton, PhD** - Associate Professor II, University of the Philippines Diliman
- **Lester Sam Araneta Geroy, MD, MPH, MSc** - President, Philippine Society of Public Health Physicians
- **Dominic Ligot** - Chief Technology Officer, CirroLytix Research Services; Board Member, AAP; Board Member, PCIJ; Lecturer, University of Asia and the Pacific
- **Jason D. Ligot, MD** - Director, Organic Intelligence
- **Robert Neil F. Leong, MSc, PhD(c)** - Chief Data Officer, metaHealth Insights and Innovation, Inc.; PhD Researcher, The University of New South Wales - Sydney; Assistant Professor 2, Mathematics and Statistics Dept. - De La Salle University
- **Lionel A. Peters, MD, MPM** - Member, Philippine Society of Public Health Physicians
- **Michael Angelo B. Promentilla, PhD** - Professor and Research Lead, Waste and Resource Management Unit, Center for Engineering and Sustainable Development Research, De La Salle University
- **Jomar F. Rabajante, PhD** - Professor, University of the Philippines Los Baños and University of the Philippines Open University
- **Miguel Antonio S. Salazar, MD, MSc, Dr.sc.hum.(c)** - Consortium Co-Convenor, Philippine Society of Public Health Physicians
- **Xerxes T. Seposo, MENRM, MPH, PhD** - Assistant Professor, Nagasaki University
- **Jan Gil G. Sarmiento, MSc (ongoing)** - Instructor, University of the Philippines Diliman
- **Theresa Rosario Tan, MABA (ongoing)** - Associate Consultant, CirroLytix Research Services
- **April Anne S. Tigue, MSc, PhD(c)** - Research staff, Waste and Resource Management Unit, Center for Engineering and Sustainable Development Research, De La Salle University

# About the L4H Consortium

Our Institutions

PSPHP
PHILIPPINE SOCIETY OF PUBLIC HEALTH PHYSICIANS

DE LA SALLE UNIVERSITY MANILA

THE UPLB BIOMATHEMATICS INITIATIVE

UNIVERSITY OF THE PHILIPPINES 1908

ASM+PH
ATENEO SCHOOL OF MEDICINE & PUBLIC HEALTH

META HEALTH
CONFIDENCE IN DECISION MAKING

AAP ANALYTICS
Association of the Philippines

CIRROLYTIX

# Example Projects

# Example Projects

1. Epidemiological Statistics for COVID-19 (w/UP PRT)

> Compiling epidemiological statistics for website and PDF statistical reports outputs

2. Short-term Forecasting of COVID-19 Cases (w/ L4H)

> Forecasting for national and regional reported case counts of COVID-19

3. Mapping COVID-19 Cases (w/ UP PRT)

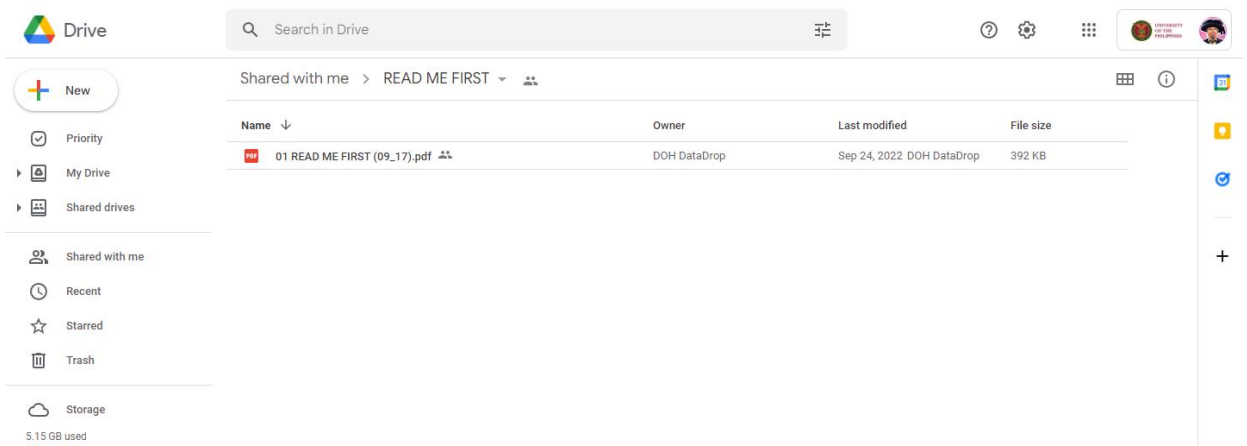
> Mapping the distribution of active cases and case per 100,000 population by barangay in the Philippines

# Demonstrating the Workflow with a Team using Example Projects

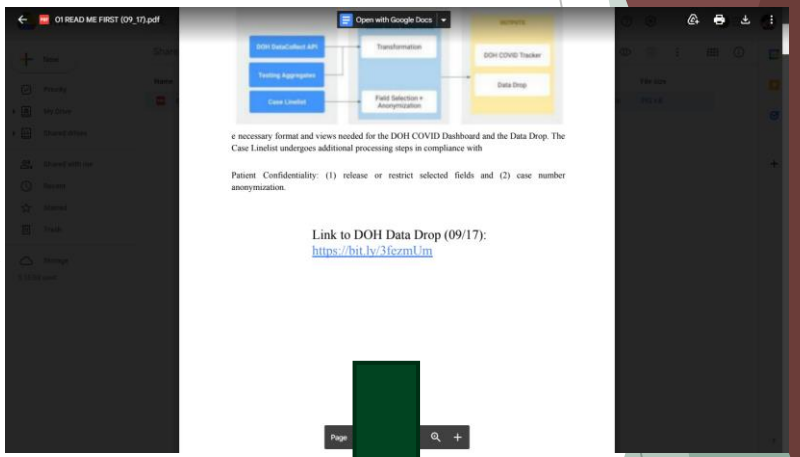
# Demonstrating the Workflow with a Team using Example Projects

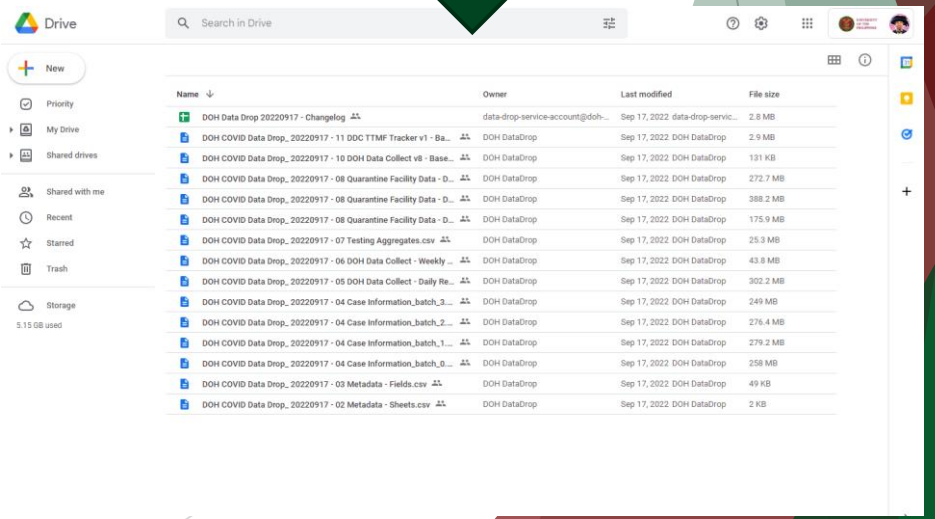
| Project | Import |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Primary Data Source:<br><br>DOH Data Drop for the Case Information Dataset [https://bit.ly/DataDropPH]<br><br>Secondary Data Source:<br><br>PSGC Data with 2020 Population Counts [https://psa.gov.ph/classification/psgc/downloads/2_PSGC%202Q-2022-Publication-Datafile.xlsx]<br><br>Researchers:<br>Trixie Delmendo, UPRI;<br>Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | |
| Mapping COVID-19 Cases (w/ UP PRT) | |

# Demonstrating the Workflow with a Team using Example Projects





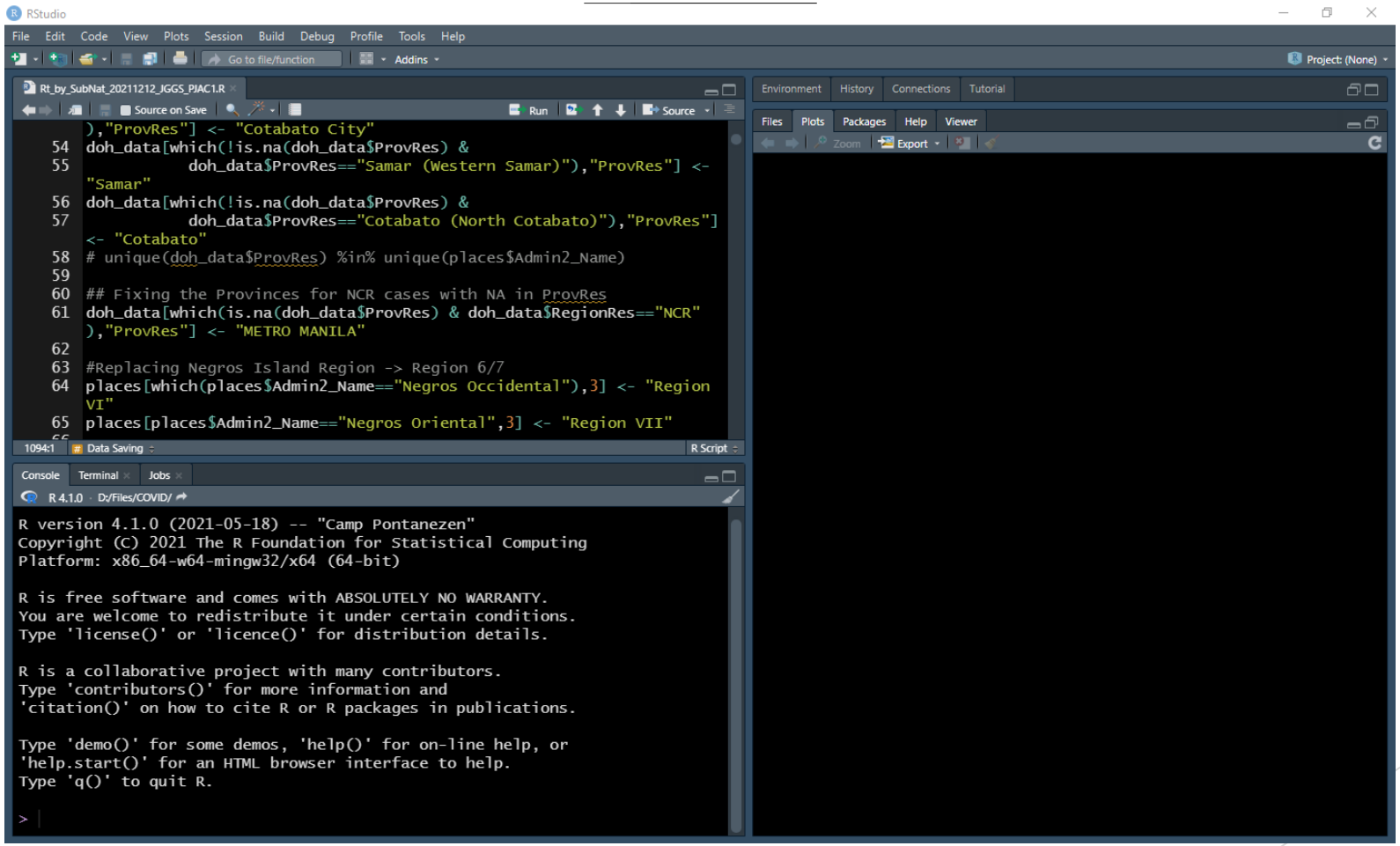


# Demonstrating the Workflow with a Team using Example Projects

| Project | Tidy |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Procedures:<br>➢ Reconcile and correct residence variable matched with PSGC data, assuming Barangay PSGC as supreme location variable<br>➢ Reconcile between dataset versions to add missing recovery dates<br><br>Researchers:<br>Trixie Delmendo, UPRI;<br>Jan Gil Sarmiento, UPD SS<br>Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | |
| Mapping COVID-19 Cases (w/ UP PRT) | |

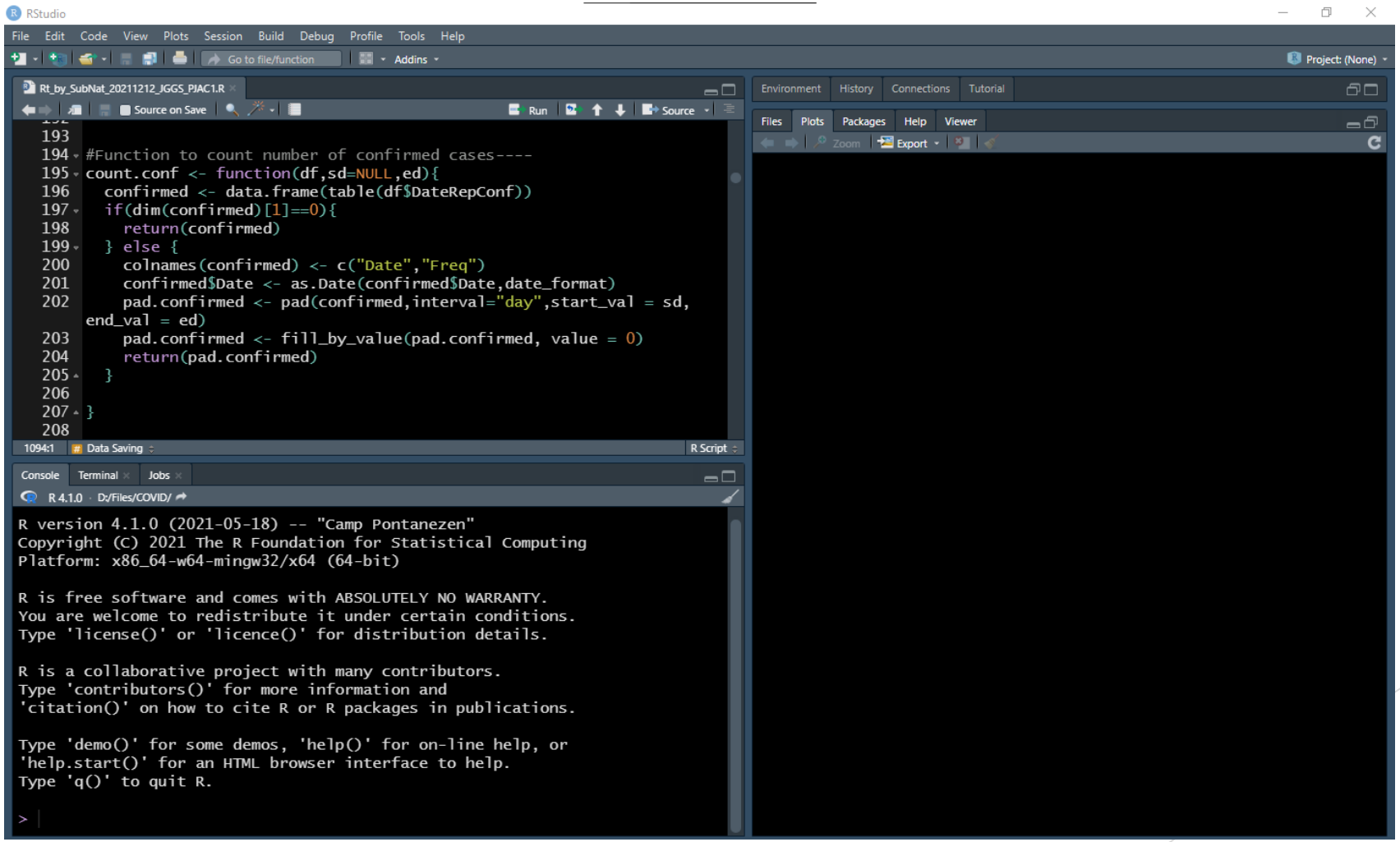
# Demonstrating the Workflow with a Team using Example Projects



# Demonstrating the Workflow with a Team using Example Projects

| Project | Transform |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Procedures:<br>➢ Group data by geographic hierarchy: 1) National, 2) Island Group, 3) Regional, 4) Provincial, 5) City/Municipality<br>➢ Compute for cumulative and new counts for cases, recoveries, deaths, and active cases per date per hierarchy<br>➢ Compute for case rates such as recovery rates, fatality rates, and doubling times<br><br>Researchers: Jan Gil Sarmiento, UPD SS; Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | Procedures:<br>➢ Group data by geographic hierarchy: 1) National and 2) Regional<br>➢ Compute for new counts for cases per date per hierarchy<br><br>Researchers: Jan Gil Sarmiento, Simon Bismonte, Maryliz Zubiri, and Nicole Uy |
| Mapping COVID-19 Cases (w/ UP PRT) | Procedures:<br>➢ Group data by geographic location: Barangay via PSGC<br>➢ Compute for total counts for cases, recoveries, deaths, and active cases per barangay<br>➢ Compute for case per 100,000 population using PSGC population data and case counts by barangay.<br><br>Researchers: Trixie Delmendo, UP RI; Peter Julian Cayton, UP PRT |

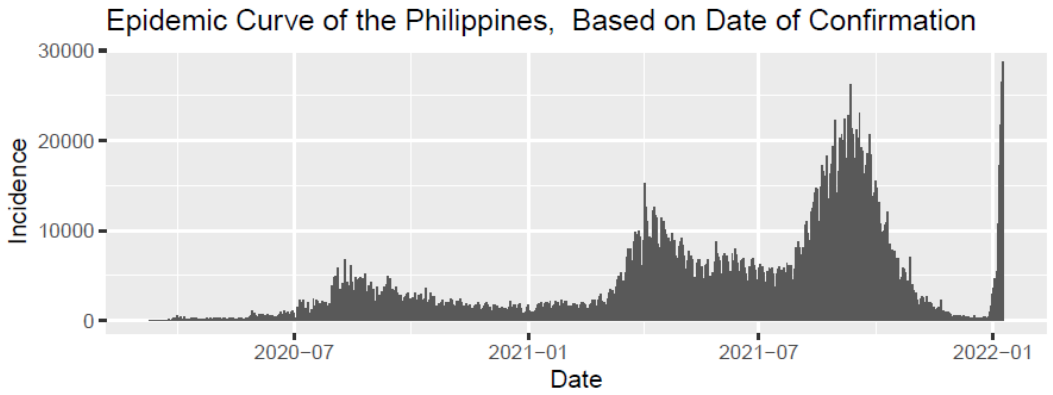
# Demonstrating the Workflow with a Team using Example Projects

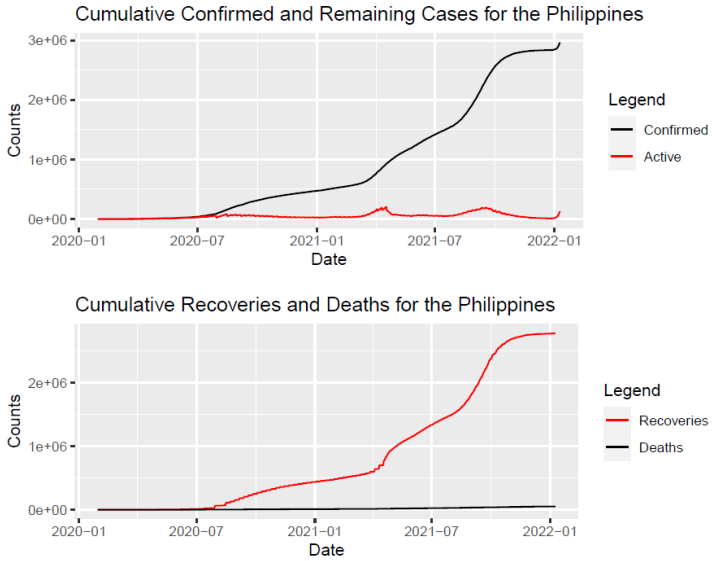


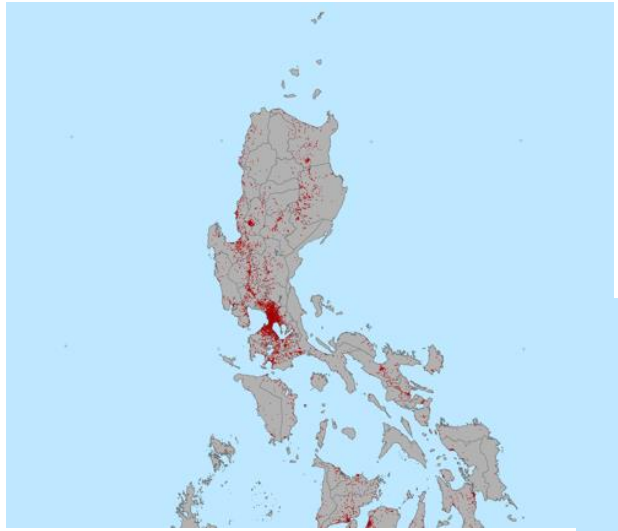
# Demonstrating the Workflow with a Team using Example Projects

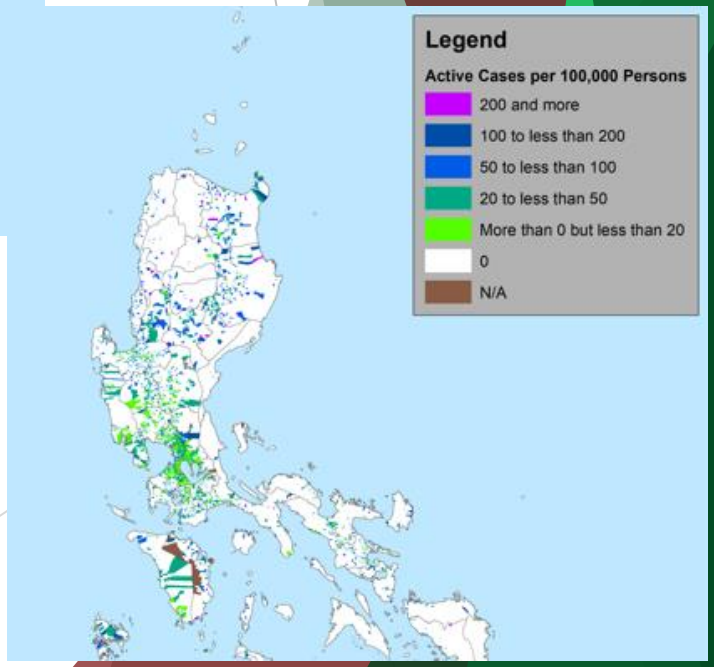
| Project | Visualize |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Procedures:<br>➢ Generate plots of cases, recoveries, deaths, and active cases per hierarchy<br><br>Researchers: Jan Gil Sarmiento, UPD SS; Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | Procedures:<br>➢ Generate plots of cases per hierarchy<br><br>Researchers: Jan Gil Sarmiento, Simon Bismonte, Maryliz Zubiri, and Nicole Uy |
| Mapping COVID-19 Cases (w/ UP PRT) | Procedures:<br>➢ Produce a map of active case distribution in which case dots are randomly distributed within their barangay of residence<br>➢ Produce a color map of case per 100,000 by barangay of residence with the barangay colors done by belonging into an interval<br><br>Researchers:<br>Trixie Delmendo, UP RI; Steffanie Chua, UP RI; Jake Mendoza, UP RI; Feye Andal, UP RI; UP Resilience Youth Mappers; Peter Julian Cayton, UP PRT |

# Demonstrating the Workflow with a Team using Example Projects



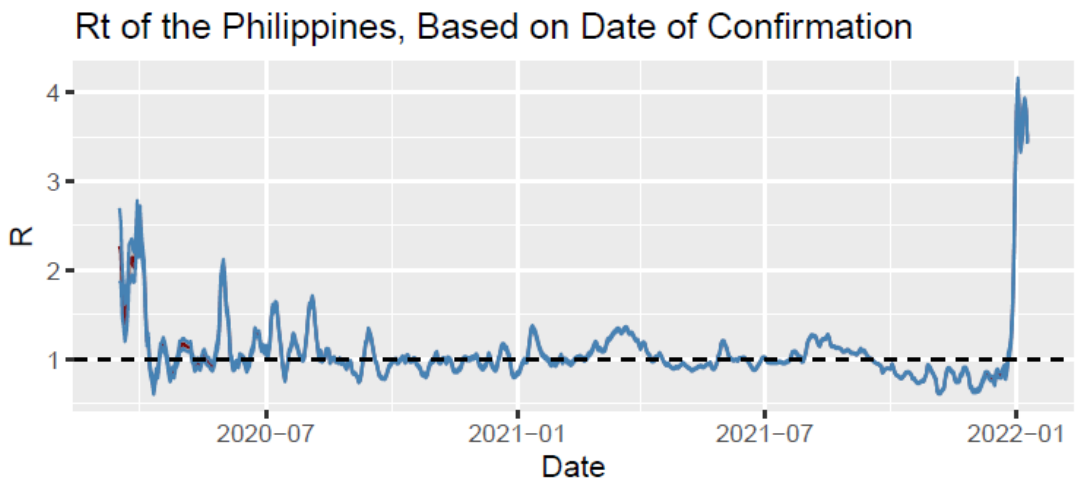



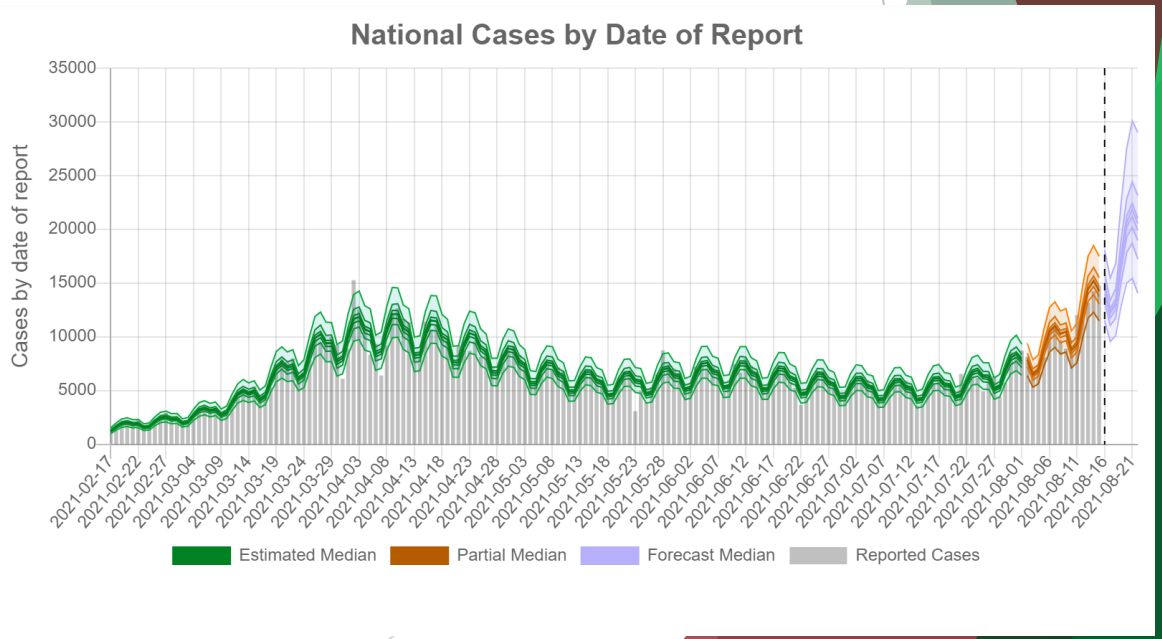


# Demonstrating the Workflow with a Team using Example Projects

| Project | Model |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Procedures:<br>➢ Generate the time-varying reproduction number based on an existing Bayesian estimation technique<br><br>Researchers: Jan Gil Sarmiento, UPD SS; Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | Procedures:<br>➢ Generate the short-term forecasts assuming a model based on the Epiforecasts methodology<br><br>Researchers: Jan Gil Sarmiento, Simon Bismonte, Maryliz Zubiri, Peter Julian Cayton, Robert Neil Leong |
| Mapping COVID-19 Cases (w/ UP PRT) | (none) |

# Demonstrating the Workflow with a Team using Example Projects

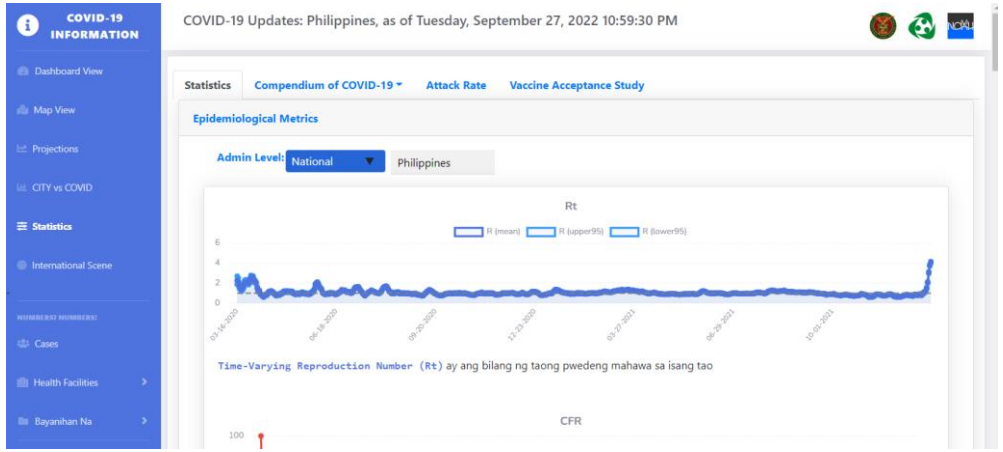


# Demonstrating the Workflow with a Team using Example Projects

| Project | Communicate |
|---|---|
| Epidemiological Statistics for COVID-19 (w/UP PRT) | Procedures:<br>➢ Publish a website page for uploading the statistics<br>➢ Generate a PDF report file<br><br>Researchers: Feye Andal and developers from the UP RI; Peter Julian Cayton, UP PRT |
| Short-term Forecasting of COVID-19 Cases (w/ L4H) | Procedure:<br>➢ Publish the results through the L4H website<br><br>Researchers: Dominic Ligot, Mark Toledo, and Angelica Mhay Salazar |
| Mapping COVID-19 Cases (w/ UP PRT) | Procedure:<br>➢ Produce maps based on the results<br><br>Researchers: Trixie Delmendo, UP RI; Steffanie Chua, UP RI; Jake Mendoza, UP RI; mappers and research of the UP RI and UP RI Youthmappers, |

# Demonstrating the Workflow with a Team using Example Projects




Compendium of Philippine COVID-19 Statistics

as of Jan 9, 2022

Volume I: Island Groups, Regions, and Provinces

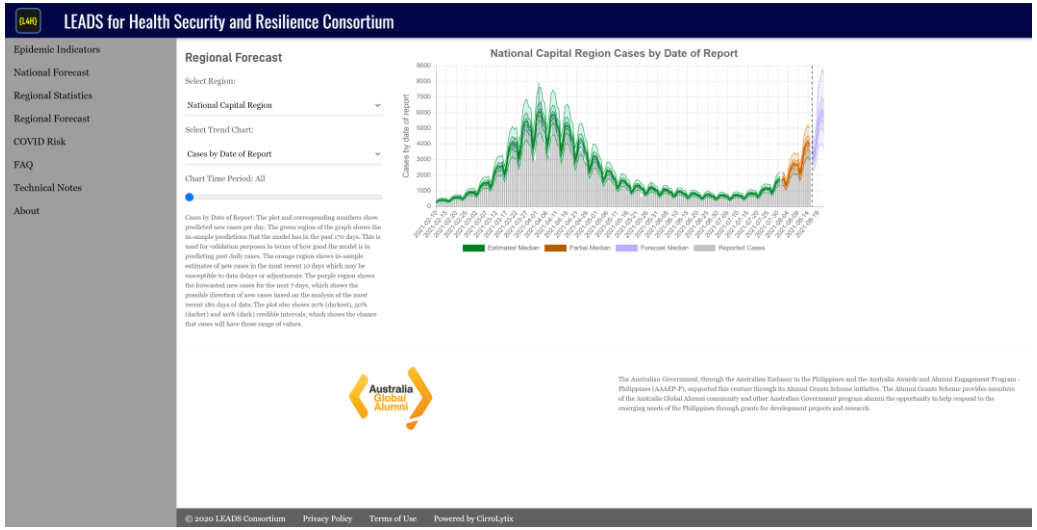
Peter Julian Cayton, Jan Gil Sarmiento, Trixie Delmendo, UP COVID-19 PRT, and L4H

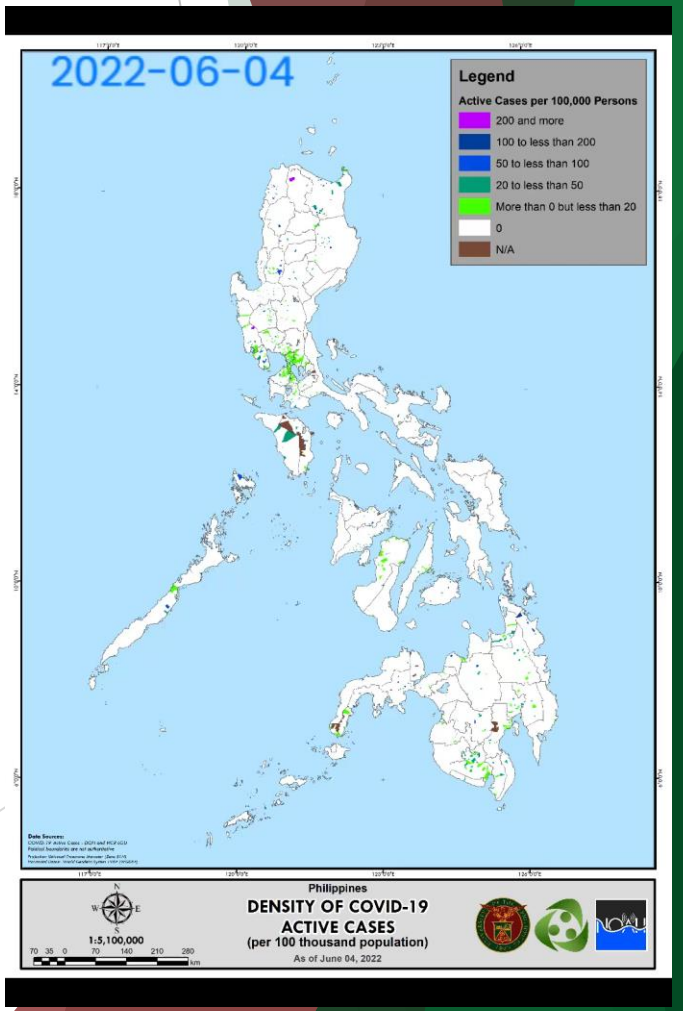
2022-01-10

### About the Authors

Peter Julian Cayton is an associate professor at the School of Statistics, University of the Philippines Diliman. His Google Scholar page is at https://scholar.google.com.au/citations?user=KtH_mGEAAAAJ. His email is pacayton(at)up.edu.ph.

Jan Gil Sarmiento is a Master of Science (Statistics) student and former instructor at the School of Statistics, University of the Philippines Diliman.



# Closing Remarks

# Closing Remarks

- Data science involves not only technical skills such as programming, statistics, and subject matter knowledge, but also communication and team collaboration
- By making data science projects in teams, the burden per person is reduced and high-quality data products are produced.
- Utilize each member's comparative advantage skills to maximize quality of output
- Build cohesion and empathy as a team of analysts and scientists producing data products

# Thanks and stay safe always!