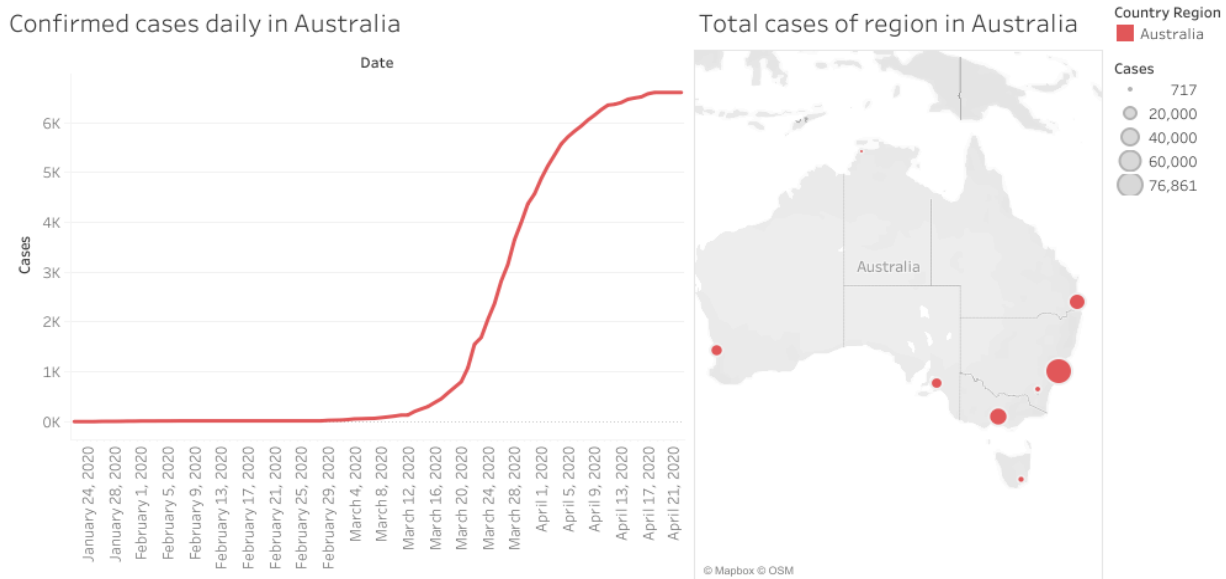


BTTH Buổi 1 (DS103.K21.2) – 16/04/2020

Họ và tên: Võ Hoàng Thông – MSSV: 18521462

BTTH1: Các bạn vẽ đồ thị và nhận xét sơ bộ số ca dương tính (confirmed) cho 1 quốc gia bất kỳ mà các bạn chọn (code mẫu theo file demo.R đã gửi trên TEAMS)

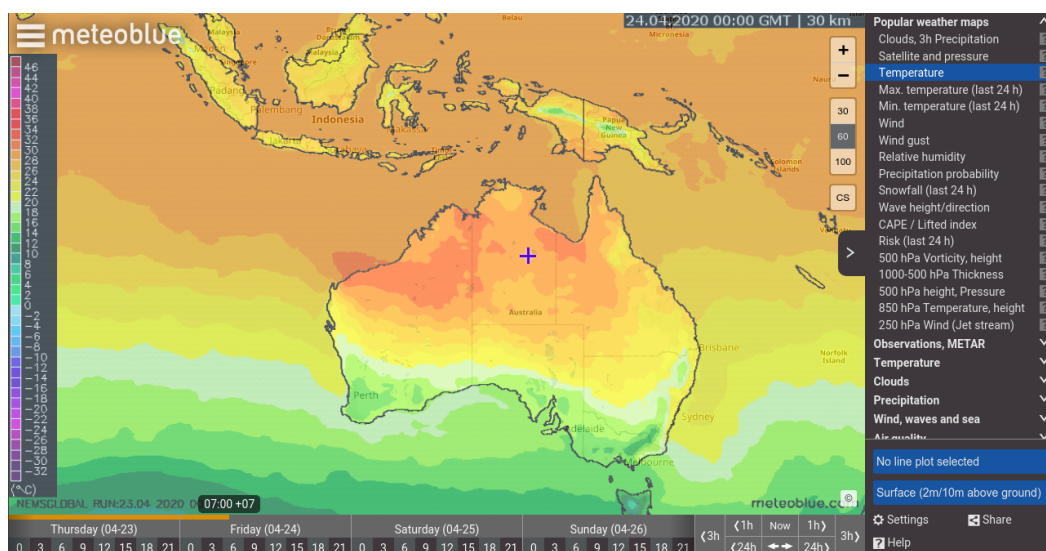
Biểu đồ dưới đây là số ca dương tính và tổng số ca nhiễm theo vùng của Australia.



Nhận xét:

- + Số ca nhiễm có dấu hiệu tăng theo thời gian, đỉnh điểm là khoảng tháng 4.
- + Ở giai đoạn tháng 1 và tháng 2, số ca nhiễm có mức độ tăng chưa đáng kể.
- + Các vùng dịch có xu hướng phân bố về phía Đông Bắc và ven biển của Australia. Ở đây có thể lý giải, Australia đang trong giai đoạn nắng nóng và nhiệt độ, cụ thể từ nửa trái của quốc gia này cho đến phần cực phía Tây Bắc có biểu đồ nhiệt độ nắng nóng từ từ 20 độ C đến cao nhất là 38 độ C.

Từ dữ kiện trên, ta có thể suy luận rằng, COVID19 bị suy yếu đáng kể trong khu vực nắng nóng có nhiệt độ trung bình cao.



Hình ảnh biểu đồ nhiệt độ của Australia.

Nguồn dữ liệu: <https://data.world/covid-19-data-resource-hub/>

Nguồn ảnh biểu đồ nhiệt: <https://www.meteoblue.com/en/weather/map/maxmintemp/world>

BTTH2: Các bạn xây dựng 1 tidy data cho 1 bộ dataset bất kỳ tự chọn (Tidy data gồm: code book, instruciton list, tidy data, raw data). Trình bày như mục số 2 của file Bai thuc hanh 2.pdf. (Chỉ chọn 1 bộ thôi nhe).

Bộ dữ liệu: **Car Evaluation Database**

Nguồn: <https://archive.ics.uci.edu/ml/machine-learning-databases/car/>

Codebook:

Codebook được mô tả trong file **car.names**.

Bảng CODE BOOK mô tả dữ liệu

Thông tin	Nội dung
Tên bộ dữ liệu	Car Evaluation Database
Nguồn thu thập và cách thức thu thập	None
Số thuộc tính	6
Thông tin tên các thuộc tính	Buying (buying price): giá mua; định dạng: v-high, high, med, low. Maint (price of the maintenance): giá bảo trì; định dạng: v-high, high, med, low. Doors(number of doors): số lượng cửa xe; định dạng: 2, 3, 4, 5-N. Persons (capacity in terms of persons to carry): Số lượng người tối đa có thể chứa của xe; định dạng số nguyên: 2,4, ... N Lug_boot (the size of luggage boot): Kích cỡ khoang phía sau xe; định dạng: small, med, big. Safety (estimated safety of the car): Ước lượng độ an toàn của xe; định dạng: low, med, high.
Thông tin tác giả	Marko Bohanec (marko.bohanec@ijs.si)

Raw data:

Raw data gồm file: **car.data**.

Tidy data:

Tidy data sẽ được lưu lại thành file: **car.csv**.

Instruction list: File Preprocessing-data.python

Code Python

```
import pandas as pd
file = open("Datasets/car.data","r")
dataset = pd.DataFrame(columns=['buying', 'maint', 'doors', 'persons', 'lug_boot', 'safety'])
```

```
for line in file.readlines():
    data = line;
    data = data.split(',')
    dataset = dataset.append({'buying': data[0], 'maint':data[1], 'doors':data[2], 'persons': data[3],
                              'lug_boot': data[4], 'safety': data[5]}, ignore_index=True)
dataset.to_csv('car.csv', index=False)
```