UPJŠ ASSETS CHALLENGE

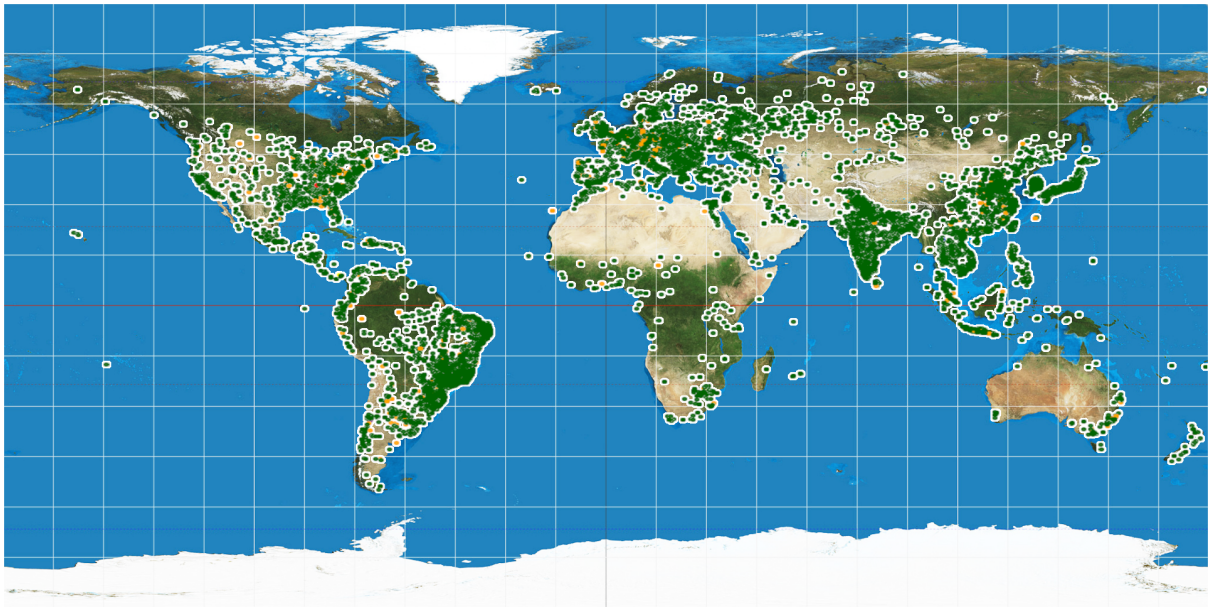
Kamil Iwanowski, Mateusz Mazurkiewicz

Hack Kosice

Introduction

INTRODUCTION

Cybersecurity plays a vital role in today's world. In order for companies to have an overview of cyber threats and attacks targeting their computer network, it is necessary for them to effectively evaluate the threat data. The team analyzed one-week dataset from real computer networks in the Czech Republic containing millions of security alerts and described their insights in this document.

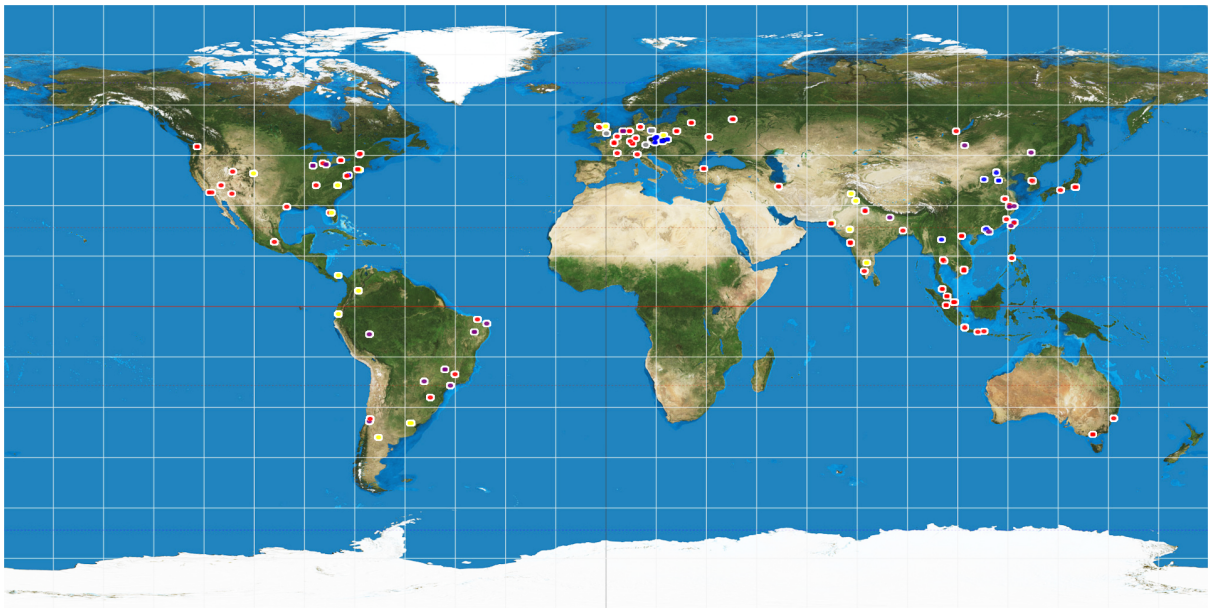


Geolocation data

The UPJŠ dataset includes considerable amount of information about locations of sources of the attacks, including their countries, cities, and even coordinates. Out of 11.75M registered attacks, 3.74M have been tracked. Locations of all of those alerts were plotted against the world map on the image above. Most of the attacks are classified as 'Recon' or 'Attempt' in the IDEA system (Intrusion Detection Extensible Alert), used by the dataset (more on that later). Vast majority of those are attempts at guessing the passwords or testing the systems for weak points, not very harmful comparing to, for example, 'Malware' category. The sources of more dangerous attacks are plotted on the map below. The difference is immediately visible.

COUNTRIES OF ORIGIN

The country with highest count of attack attempts is China. Local contribution to cybersecurity threat is relatively low, most of it being foreign. Out of 3.74M tracked attacks, 23.0k of them were from within the country (0.6%). The detailed breakdown is contained in the tables on the next page.



Attack attempts per country:

Total	**3.74M (100%)**
China	519.8k (13.9%)
United States	502.3k (13.4%)
Vietnam	318.7k (8.5%)
Japan	292.8k (7.8%)
India	255.5k (6.8%)
Brazil	195.4k (5.2%)
Canada	161.6k (4.3%)
Indonesia	159.2k (4.3%)
France	123.0k (3.3%)
Hong Kong	112.4k (3.0%)
Russia	109.3k (2.9%)
Other	991.1k (26.5%)

Attack attempts per continent:

Asia	2.21M (59.1%)
Northern America	674.4k (18.0%)
Europe	504.9k (13.5%)
Southern America	263.5k (7.0%)
Africa	67.9k (1.8%)
Oceania	22.4k (0.6%)

The distribution of (known) alert sources over the world is heavily unequal.

DANGEROUS ATTACKS LOCATIONS

3'036'003 out of 3'741'089 (81.2%) tracked attacks were assigned 'Recon.Scanning' event type classification. It is defined as 'attacks that send requests to a system to discover weak points' in the IDEA standard. It encompasses port scanning, host sweeping and other methods of collecting information about hosts. 83.2% of alerts from China are of this type.

Interestingly, this is true for only 59.7% of attack attempts from USA. As much as 26.1% of USA-based attacks are of 'Attempt.Exploit' nature, that is 'attempts to compromise a system or to disrupt any service by exploiting vulnerabilities with a standardised identifier'. In fact, most of all tracked exploit attacks in Czech Republic were pinpointed to USA (131238, 51.6%), the other Northern American country, Canada, lagging just behind as source of 32.7% of all tracked exploit attack.

Those are significant observations. Northern America leads the way with malware as well. Only 10 out of 3591 located malware alerts were from China. 588 were from USA, and 202 from Canada. 633 are attributed to Indonesia and 475 to Vietnam.

On the other hand, the most harmless attacks were from Japan, over 99.2% of the alerts being caused by scanning, 112 by exploits, 24 by malware, and 0 by ransomware. For context to those numbers, consult tables above.

Attacks data

Intrusion Detection Extensible Alert is a human-readable format for automated security incident reporting. The provided JSON serialized dataset takes 5.55GB of memory uncompressed, boasting more than 11.7M incidents. An example of such incident, taken from `cesnet.cz` website, is:


```
{
    "Format": "IDEA0",
    "ID": "4390fc3f-c753-4a3e-bc83-1b44f24baf75",
    "CreateTime": "2012-11-03T10:00:02Z",
    "DetectTime": "2012-11-03T10:00:07Z",
    "WinStartTime": "2012-11-03T05:00:00Z",
    "WinEndTime": "2012-11-03T10:00:00Z",
    "EventTime": "2012-11-03T07:36:00Z",
    "CeaseTime": "2012-11-03T09:55:22Z",
    "Category": ["Fraud.Phishing"],
    "Ref": ["cve:CVE-1234-5678"],
    "Confidence": 1,
    "Note": "Synthetic example",
    "ConnCount": 20,
    "Source": [
        {
            "Type": ["Phishing"],
            "IP4": ["192.168.0.2-192.168.0.5", "192.168.0.10/25"],
            "IP6": ["2001:0db8:0000:0000:0000:ff00:0042::/112"],
            "Hostname": ["example.com"],
            "URL": ["http://example.com/cgi-bin/killemall"],
            "Proto": ["tcp", "http"],
            "AttachHand": ["att1"],
            "Netname": ["ripe:IANA-CBLK-RESERVED1"]
        }
    ],
    "Target": [
        {
            "Type": ["Backscatter", "OriginSpam"],
            "Email": ["innocent@example.com"],
            "Spoofed": true
        },
        {
            "IP4": ["10.2.2.0/24"],
            "Anonymised": true
        }
    ],
    "Attach": [
        {
            "Handle": "att1",
            "FileName": ["killemall"],
            "Type": ["Malware"],
            "ContentType": "application/octet-stream",
            "Hash": ["sha1:0c4a38c3569f0cc632e74f4c"],
            "Size": 46,
            "Ref": ["Trojan-Spy:W32/FinSpy.A"],
            "ContentEncoding": "base64",
            "Content": "TVpqdXN0a2lkZGluZwo="
        }
    ],
```

```
    "Node": [
        {
            "Name": "cz.cesnet.kippo-honey",
            "Type": ["Protocol", "Honeypot"],
            "SW": ["Kippo"],
            "AggrWin": "00:05:00"
        }
    ]
}
```

ATTACK TYPES

The following table completes information contained in the last subsection of the geolocation data section.

Attack attempts per IDEA classification:

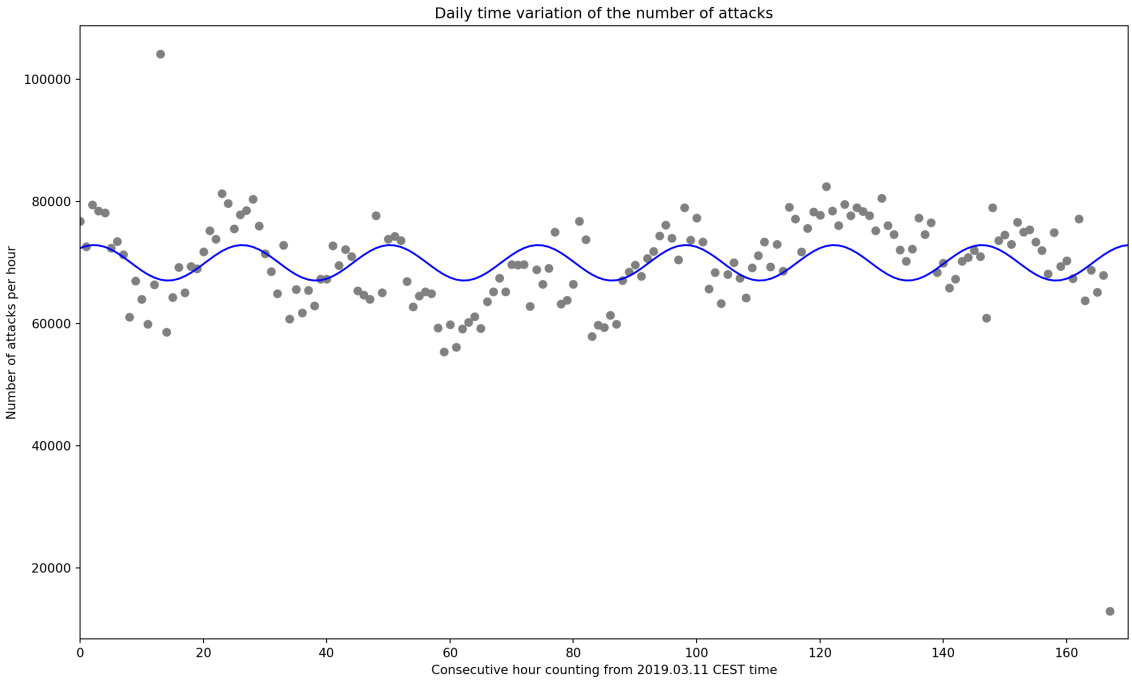
Classification	Total (% of all)	Tracked (% tracked)
Recon.Scanning	9'030.2k (76.9%)	3'036.0k (33.6%)
Attempt.Login	1'540.9k (13.1%)	435.7k (28.3%)
Attempt.Exploit	846.5k (7.2%)	254.2k (30.0%)
Malware	125.0k (1.1%)	3.6k (2.9%)
Availability.DoS	41.2k (0.4%)	5.3k (13.0%)
Anomaly.Traffic	16.6k (0.1%)	1.1k (6.6%)
Malware.Ransomware	9.3k (0.1%)	2.0k (21.0%)
Availability.DDoS	9.2k (0.1%)	1.4k (15.4%)
Intrusion.Botnet	5.3k (under 0.1%)	1.0k (18.4%)
Vulnerable.Config	0.7k (under 0.1%)	0.5k (77.8%)

What is remarkable at the first glance is the extremely low proportion of tracked malware attacks. Only 3.6k out of over 125k malware alerts have assigned location. As discussed in previous section, few malware attacks are known to be performed from China, especially compared to its Recon and Attempt numbers. It is reasonable to assume that a large chunk of malware attacks without location come from this country.

Definitions of classifications can be found at `https://idea.cesnet.cz/en/classifications`.

TIMING

The next thing the team analyzed was the time frame of the attacks. They are not evenly space throughout the day; instead the count of alerts oscillated throughout the day everyday.



The model in graph above was obtained by Ordinary Least Squares. The curve is described by

$$y = 69940 + 1622.6 \cdot sin(2\pi t) + 2408.4 \cdot cos(2\pi t) \quad (1)$$

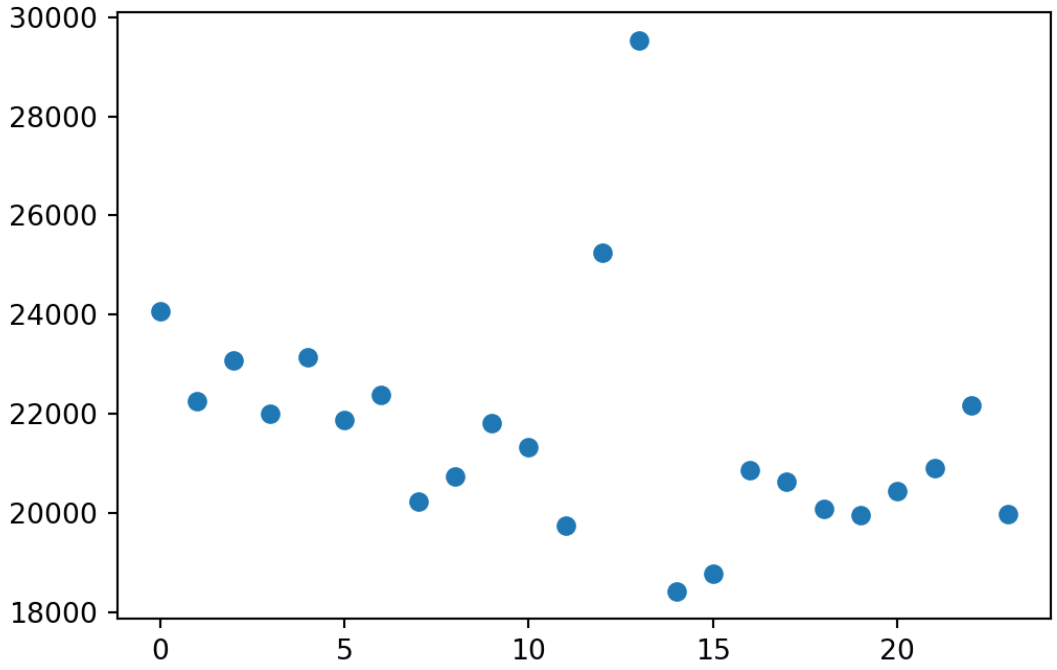
Where y is predicted number of alerts and t is time of the day (current hour divided by 24, that is: 6:00 is $t = 0.25$, 12:30 is $t = 0.521$). As can be seen, the model does not explain much of dependent variable variation. It would be much better, if not for the severe time constraint and experienced technical difficulties. However, the p-values are promising (0.000, 0.054, 0.005, respecitvely for three constants included in Equation 1).

Note

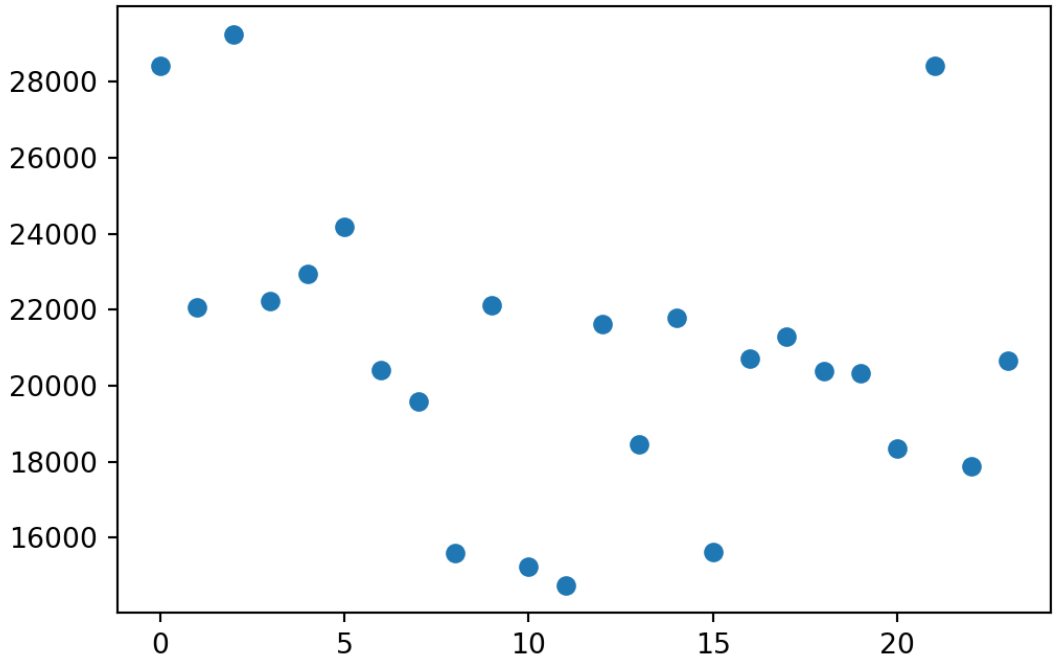
All clock time is in CEST, local to Czech Republic.

Very interesting outcome is obtained when this data is divided by the country of origin of the attack. Undeniably, activity of the threat actors depend on their geographic location. However, most of the alerts are raised at night (in the target's location).

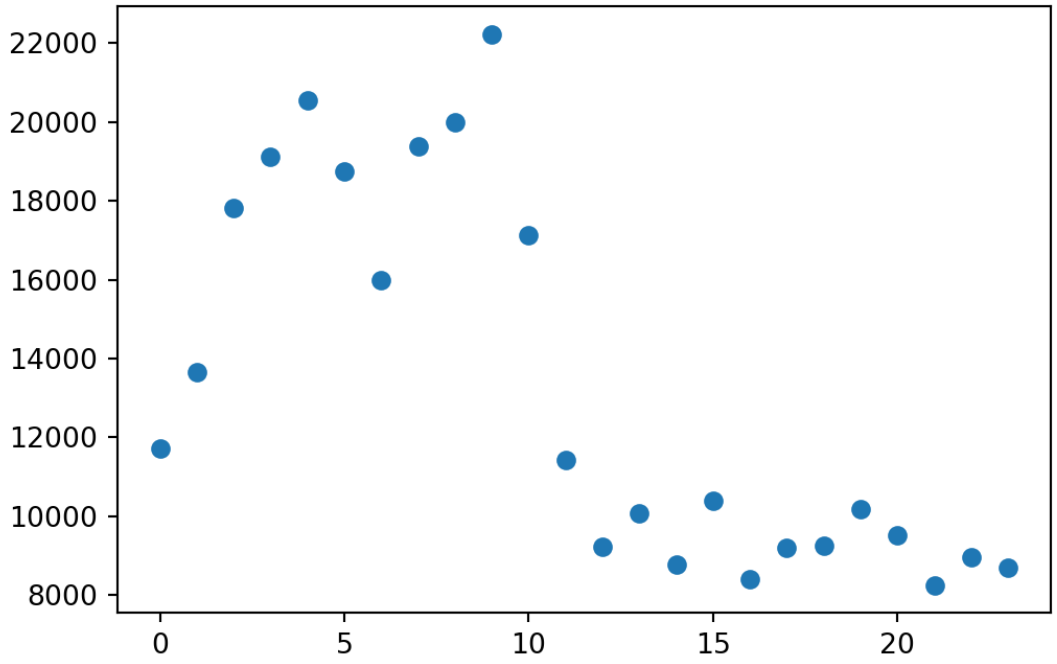
China


30000
28000
26000
24000
22000
20000
18000
0
5
10
15
20

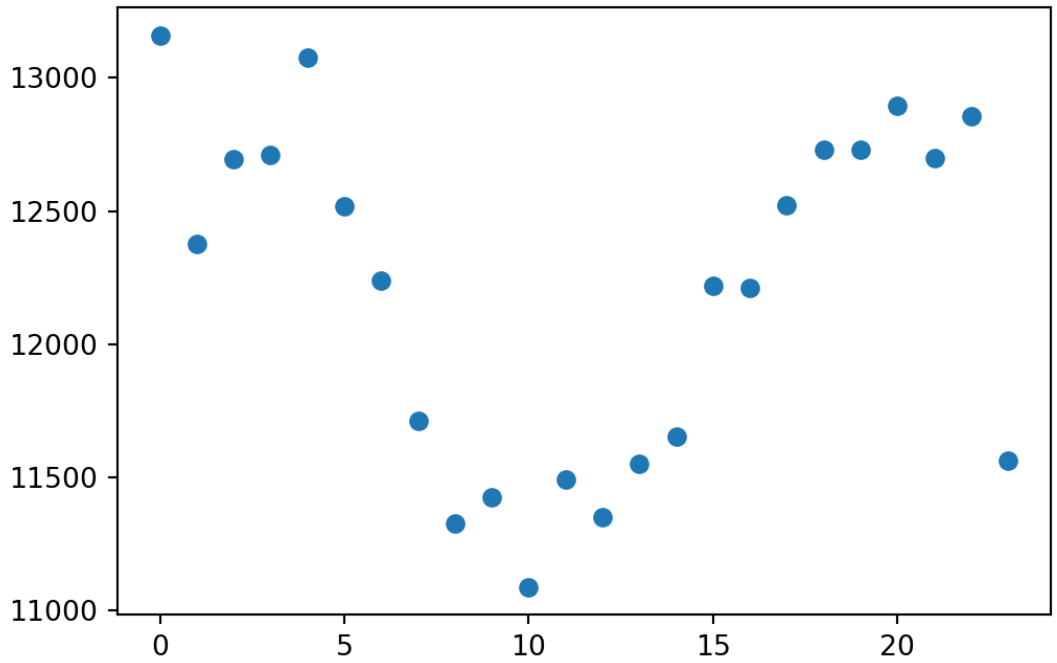
USA


28000
26000
24000
22000
20000
18000
16000
0
5
10
15
20

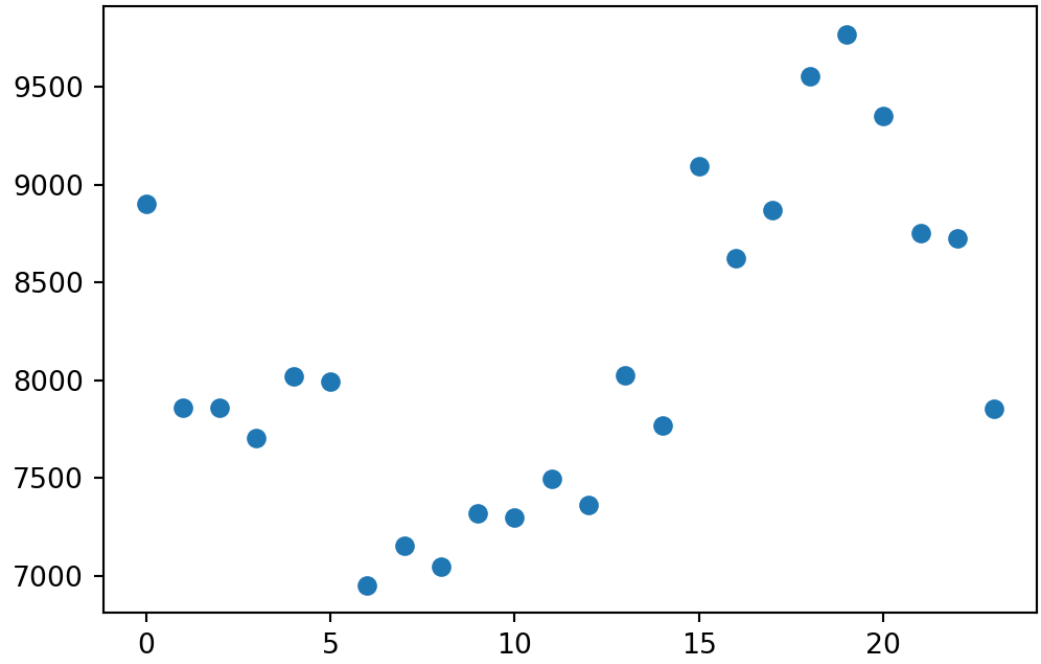
Vietnam



Japan



Brazil



Canada