detection_of_IoT_botnet_attacks_N_BaIoT Data Set exploration

Yulia Zamyatina

April, 2021

Dataset description

According to UCI MAchine Learning Repository [1], this data set is the collection of real traffic data, gathered from 9 commercial IoT-devices authentically infected by Mirai and BASHLITE (gafgyt).

The data set has 115 attributes:

1. It has 5 time-frames: L5, L3, L1, L0.1 and L0.01.
2. The statistics extracted from each stream for each time-frame:
   - *weight*: the weight of the stream (can be viewed as the number of items observed in recent history)
   - *mean*
   - *std (variance)*
   - *radius*: the root squared sum of the two streams' variances
   - *magnitude*: the root squared sum of the two streams' means
   - *covariance*: an approximated covariance between two streams
   - *pcc*: an approximated correlation coefficient between two streams
3. It has following stream aggregations:
   - *MI*: ("Source MAC-IP" in N-BaIoT paper) Stats summarizing the recent traffic from this packet's host (IP + MAC)
   - *H*: ("Source IP" in N-BaIoT paper) Stats summarizing the recent traffic from this packet's host (IP)
   - *HH*: ("Channel" in N-BaIoT paper) Stats summarizing the recent traffic going from this packet's host (IP) to the packet's destination host.
   - *HH_jit*: ("Channel jitter" in N-BaIoT paper) Stats summarizing the jitter of the traffic going from this packet's host (IP) to the packet's destination host.
   - *HpHp*: ("Socket" in N-BaIoT paper) Stats summarizing the recent traffic going from this packet's host+port (IP) to the packet's destination host+port. Example 192.168.4.2:1242 -> 192.168.4.12:80

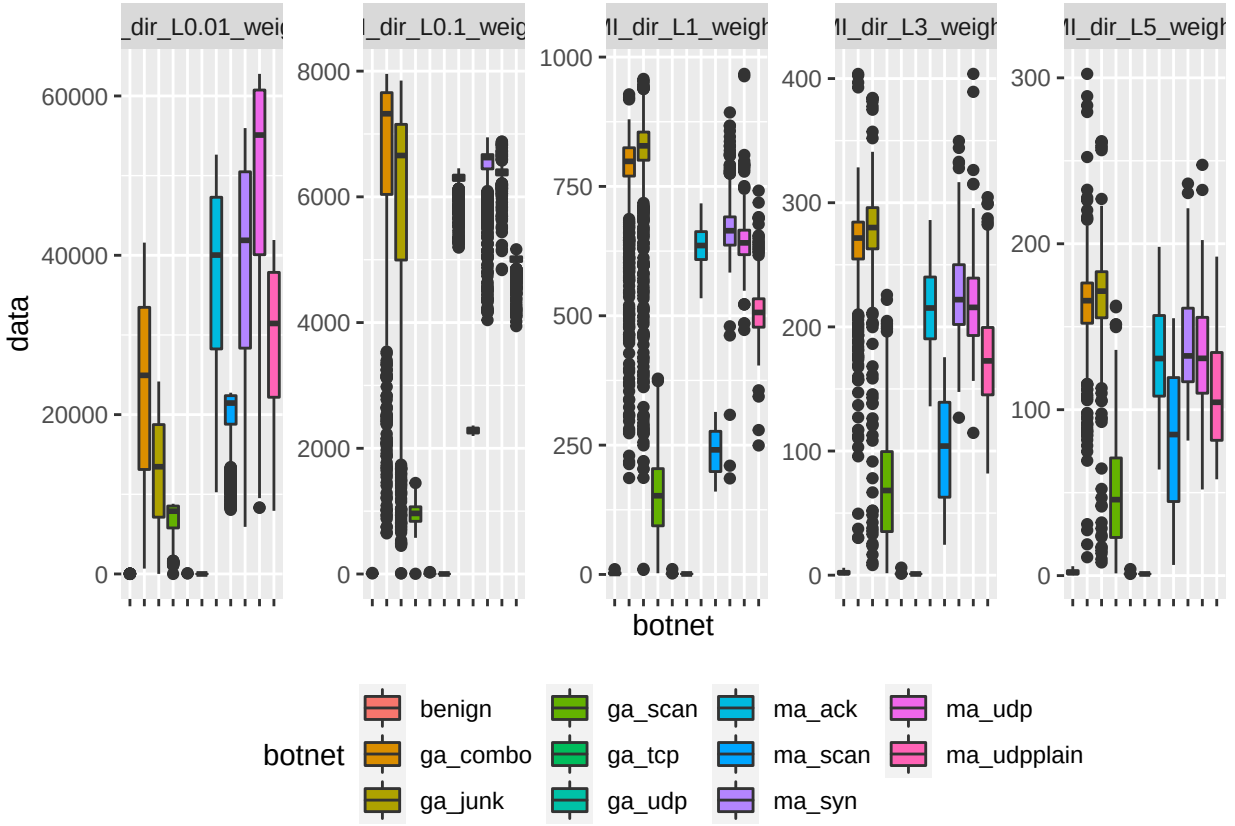
Thus, the column *'MI_dir_L5_weight'* in the data set shows the weight of the recent traffic from the packet's host for L5 time-frame.

The data set consists of *.csv files, each representing a benign traffic or an attack. When I gathered *.csv files together in one data set, I added *'botnet'* column, where I keep information about the attacks from the different botnets. The dataset contains *combo*, *junk*, *scan*, *tcp* and *udp* gafgyt attacks, and *ack*, *scan*, *syn*, *udp* and *udpplain* mirai attacks. I used *'ga'* prefix for gafgyt attacks and *'ma'* for mirai attacks in the *'botnet'* column.

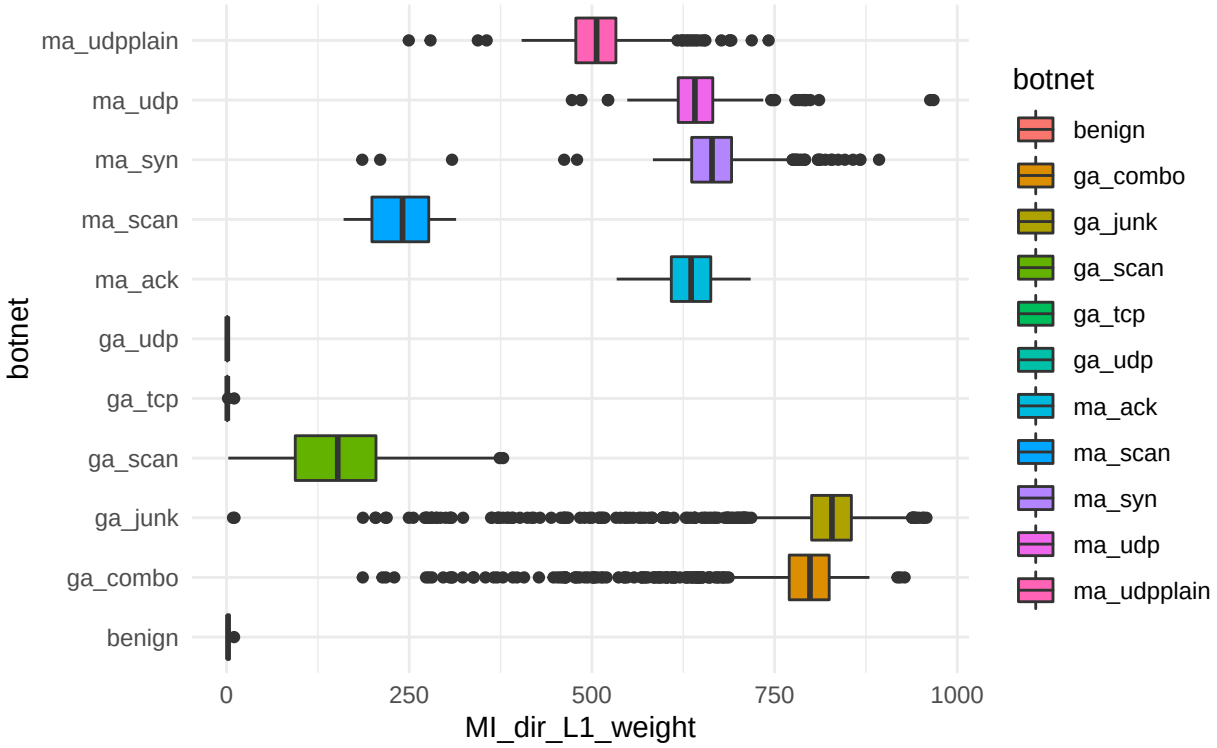
[1] detection_of_IoT_botnet_attacks_N_BaIoT Data Set

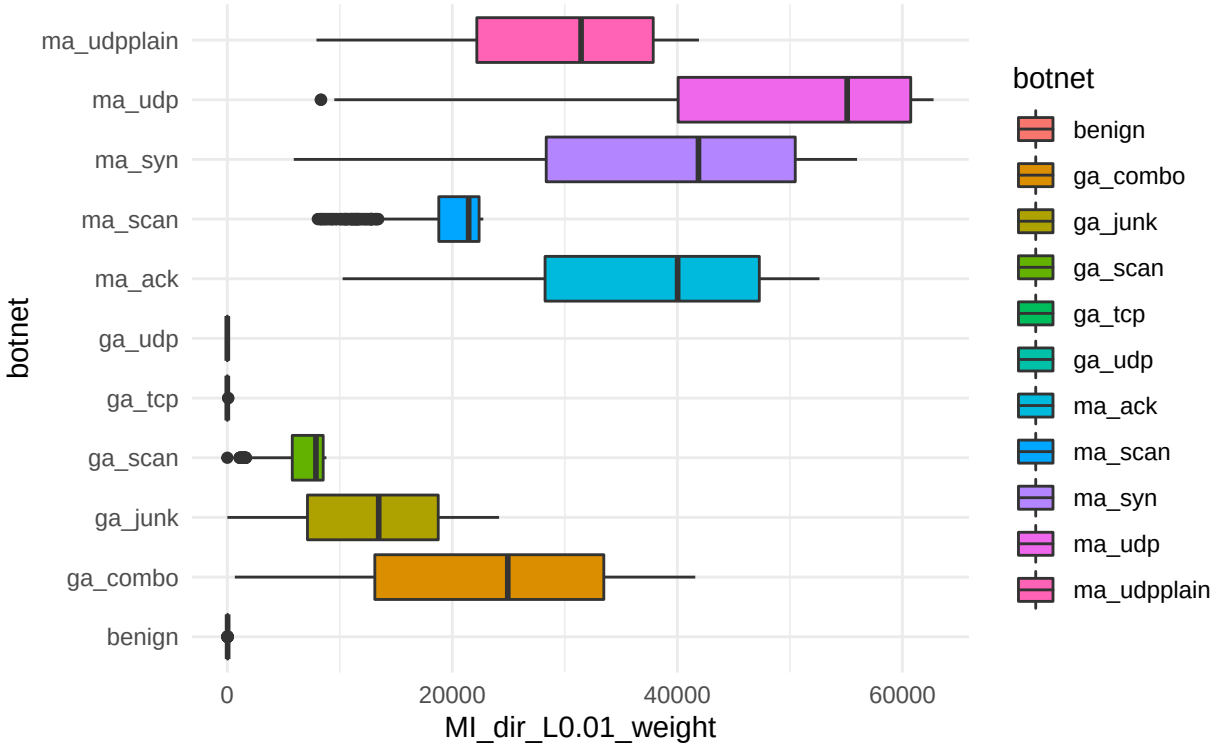
Below I explore the data only from Danmini doorbell device. Because the whole dataset is huge, I made a sample about 1000 rows for each botnet just for an illustrative purpose. Otherwise, all plots will be heavy.

The few first columns contain the data for *MI* stream, and I start my research from *weight* data for L5 - L0.01 time-frames (MI_dir_L5_weight - MI_dir_L0.01_weight columns):



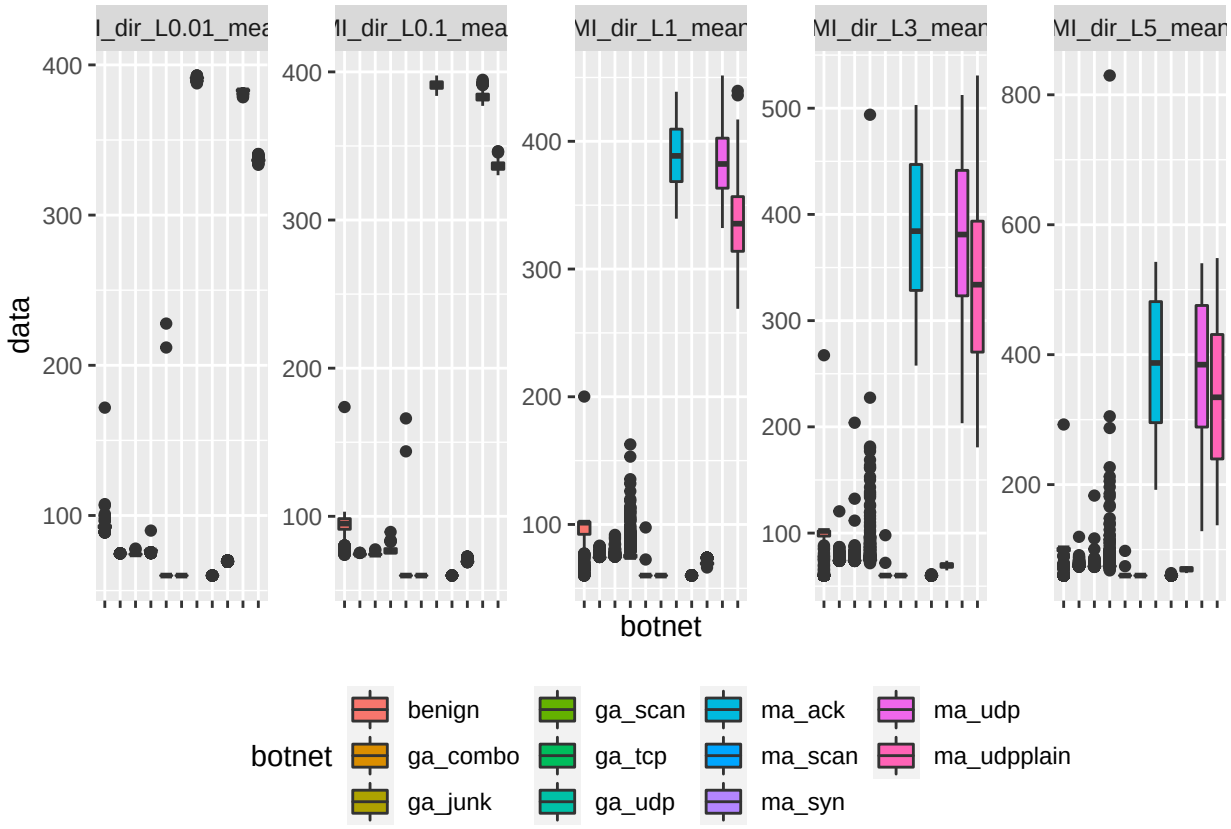
The plot shows that using only the *weight* attribute, I can easily separate **benign traffic**, **ga_tcp** and **ga_udp** attacks from the other attacks. Their boxplots look like points on the plot, that is, their medians are close to 0, they do not have a large IQR, there are no outliers. This is more clearly seen at the small time-frames, let's view close up the *weight* attribute for L1 and L0.01 time-frames:



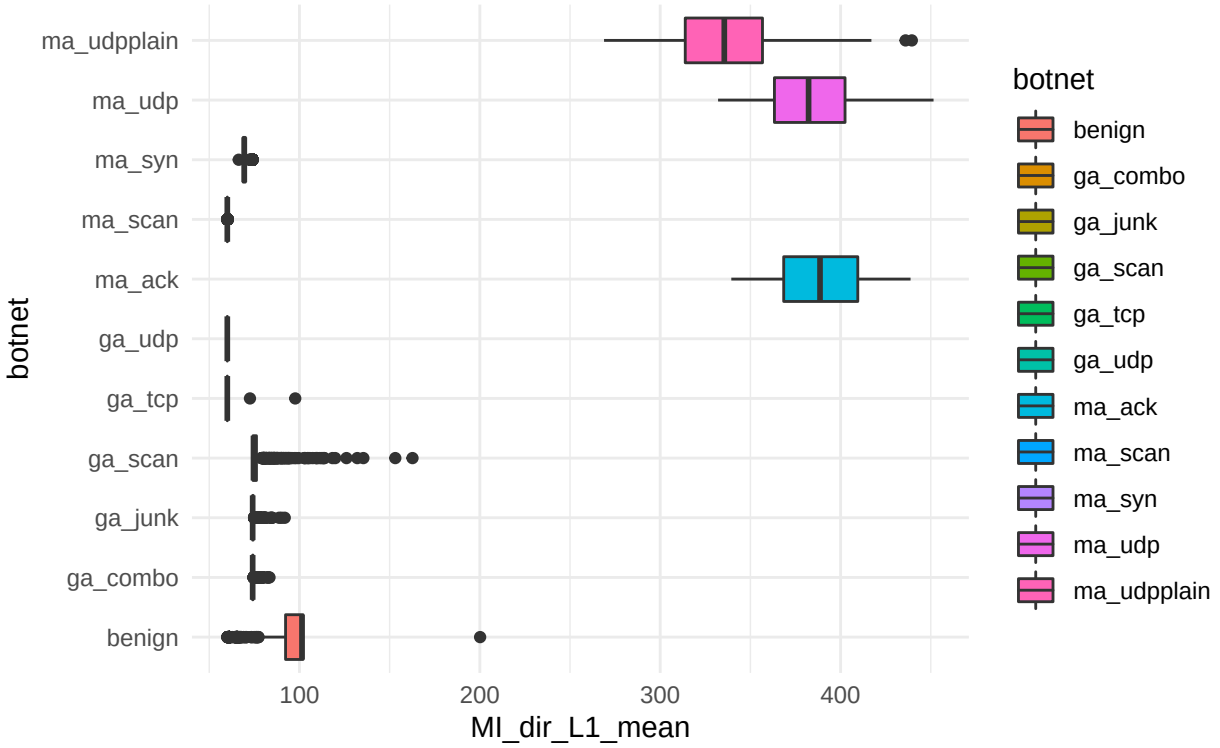


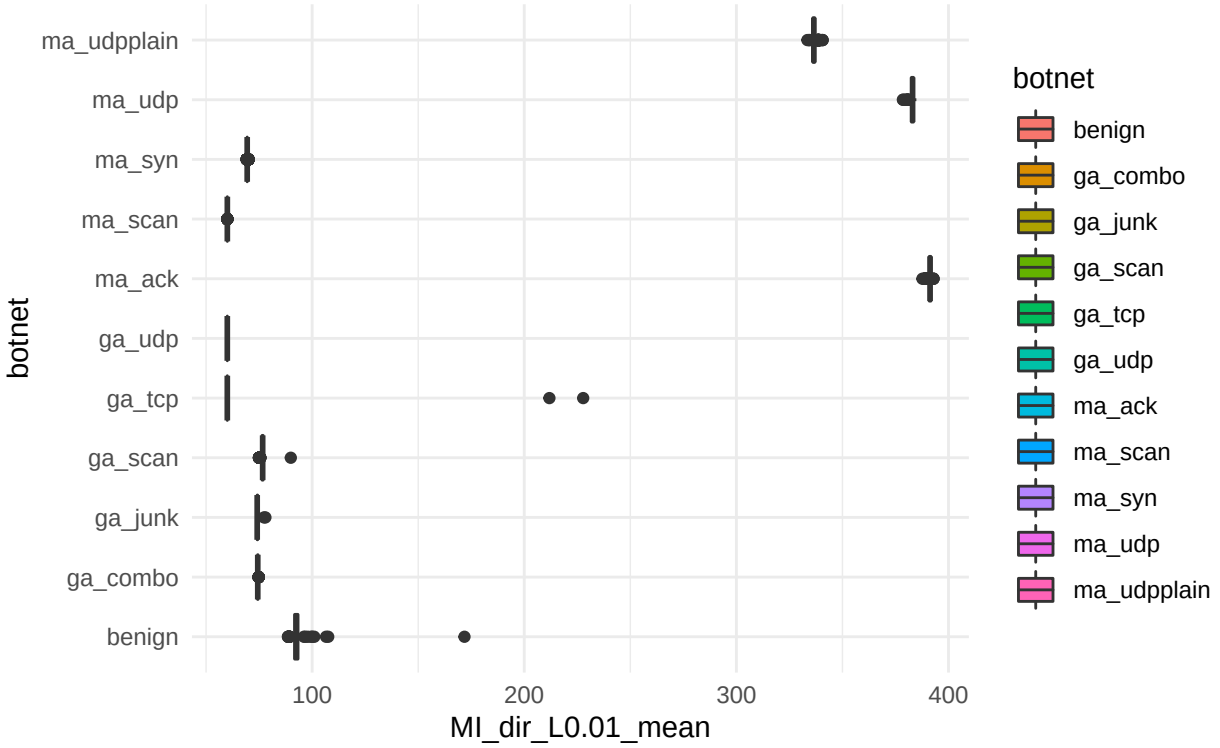
The **ga_tcp** an **ga_udp** disguise themselves well as **benign traffic**. So, I have to find at least one another attribute, that can help me separate **benign traffic** from **ga_tcp** and **ga_udp** attacks.

Next, I would like to explore *mean* attribute for the same stream (MI_dir_L5_mean - MI_dir_L0.01_mean):



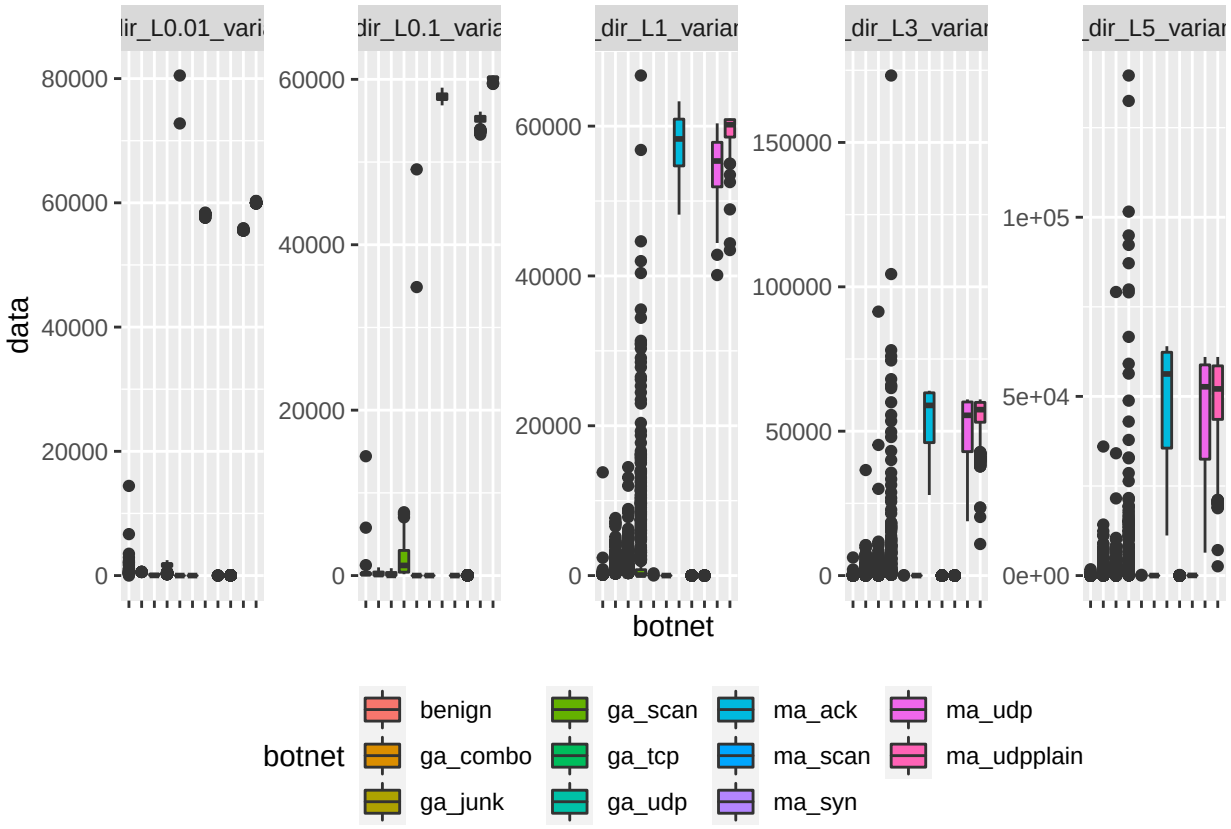
This plot shows that *mean* attribute can help me separate **benign traffic** from **ga_tcp** and **ga_udp**, as median for **benign traffic** is higher that for these attacks. Let's check this on the small time-frames (L1 and L0.01):





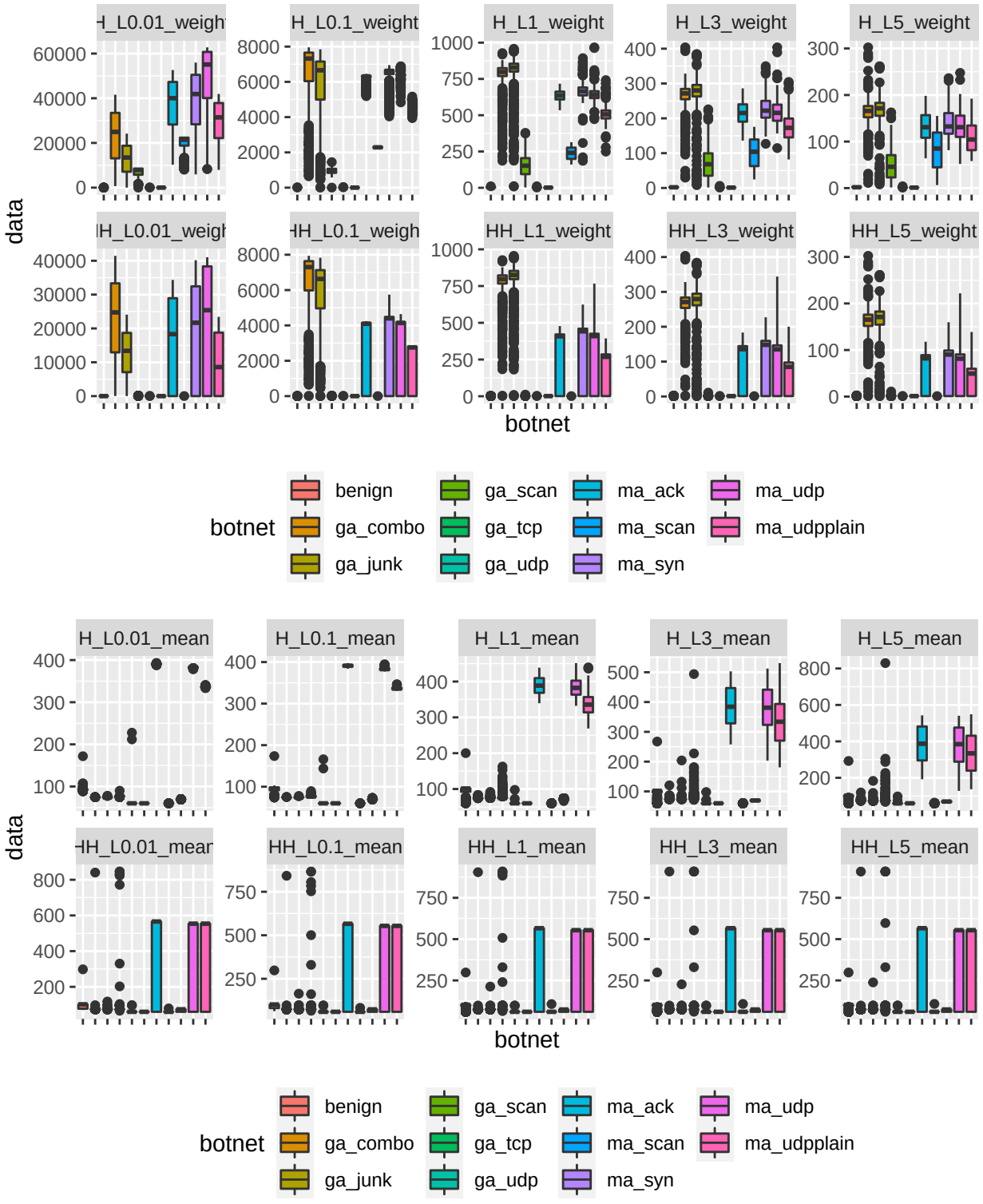
Therefore, I still need one more attribute as all of them have outliers, and using only the *weight* and *mean* attributes will not help clearly separate **benign traffic** from the attacks.

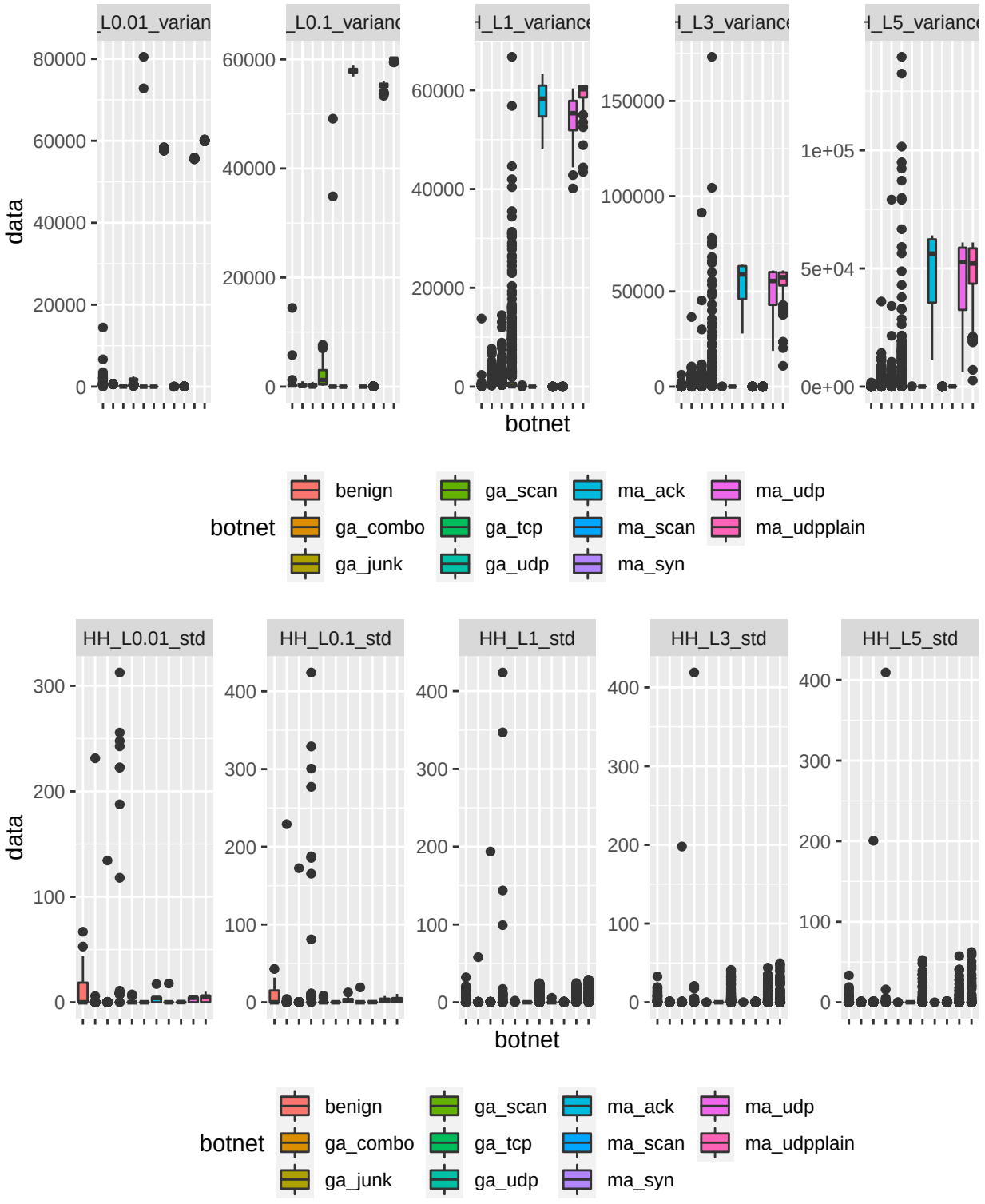
After that let's explore *variance* attribute (MI_dir_L5_variance - MI_dir_L0.01_variance):



This plot shows that *variance* attribute doesn't give a new information how to separate **benign traffic** from attacks, so I can easily remove this attribute if I need to reduce the data frame (or martix) dimension.

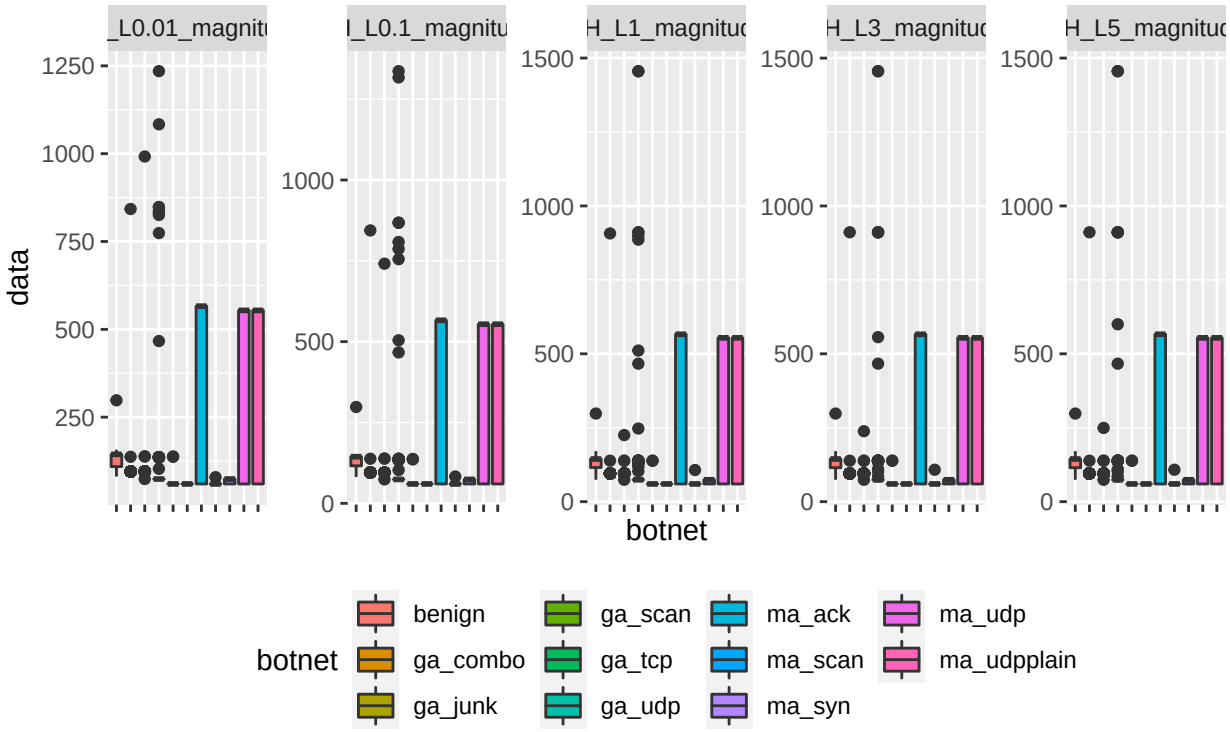
I have finished with MI stream and can explore statistics for H and HH streams. In the same way as for the previous stream, I will consider *weight*, *mean* and *variance* (or *std*) attributes:


H_L0.01_weigh
H_L0.1_weight
H_L1_weight
H_L3_weight
H_L5_weight
HH_L0.01_weigh
HH_L0.1_weigh
HH_L1_weight
HH_L3_weight
HH_L5_weight
data
botnet
botnet
benign
ga_combo
ga_junk
ga_scan
ga_tcp
ga_udp
ma_ack
ma_scan
ma_syn
ma_udp
ma_udpplain
H_L0.01_mean
H_L0.1_mean
H_L1_mean
H_L3_mean
H_L5_mean
HH_L0.01_mean
HH_L0.1_mean
HH_L1_mean
HH_L3_mean
HH_L5_mean
data
botnet
botnet
benign
ga_combo
ga_junk
ga_scan
ga_tcp
ga_udp
ma_ack
ma_scan
ma_syn
ma_udp
ma_udpplain

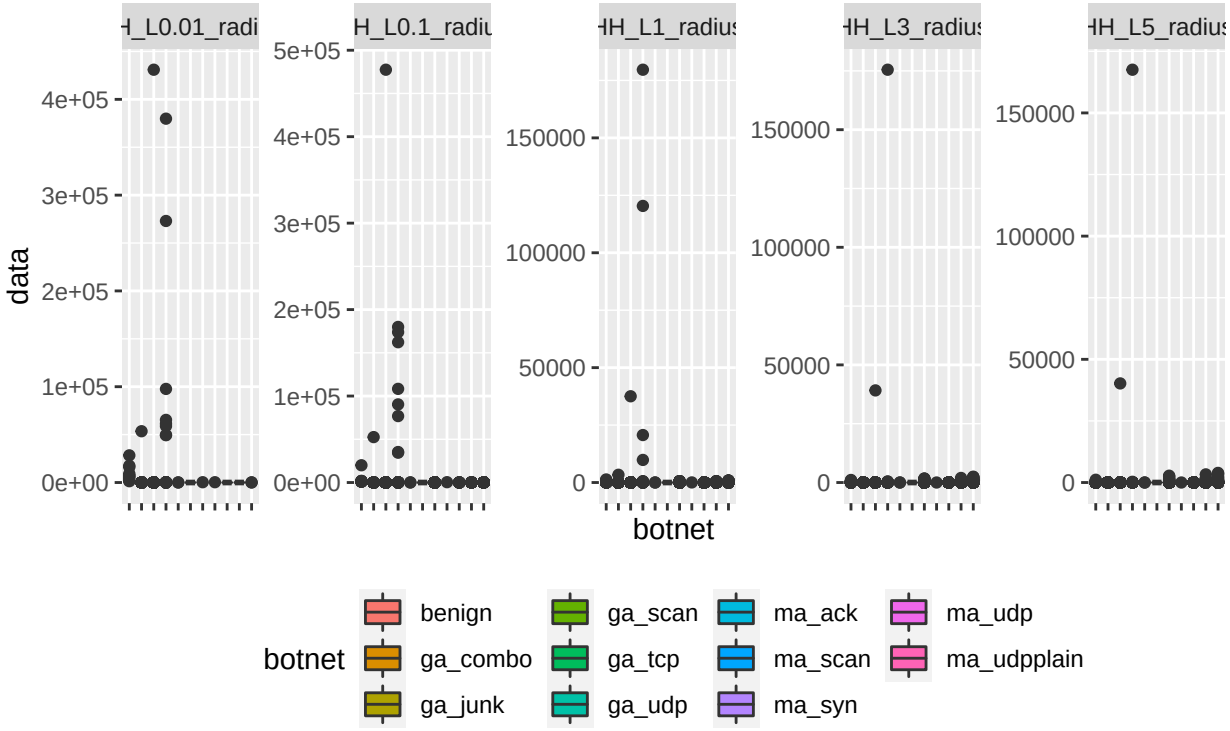


All plots show that I can use *weight* and *mean* attributes to separate **benign traffic** from the attacks, and remove *variance* or *std* attributes if I need to reduce the data frame (or matrix) dimension.

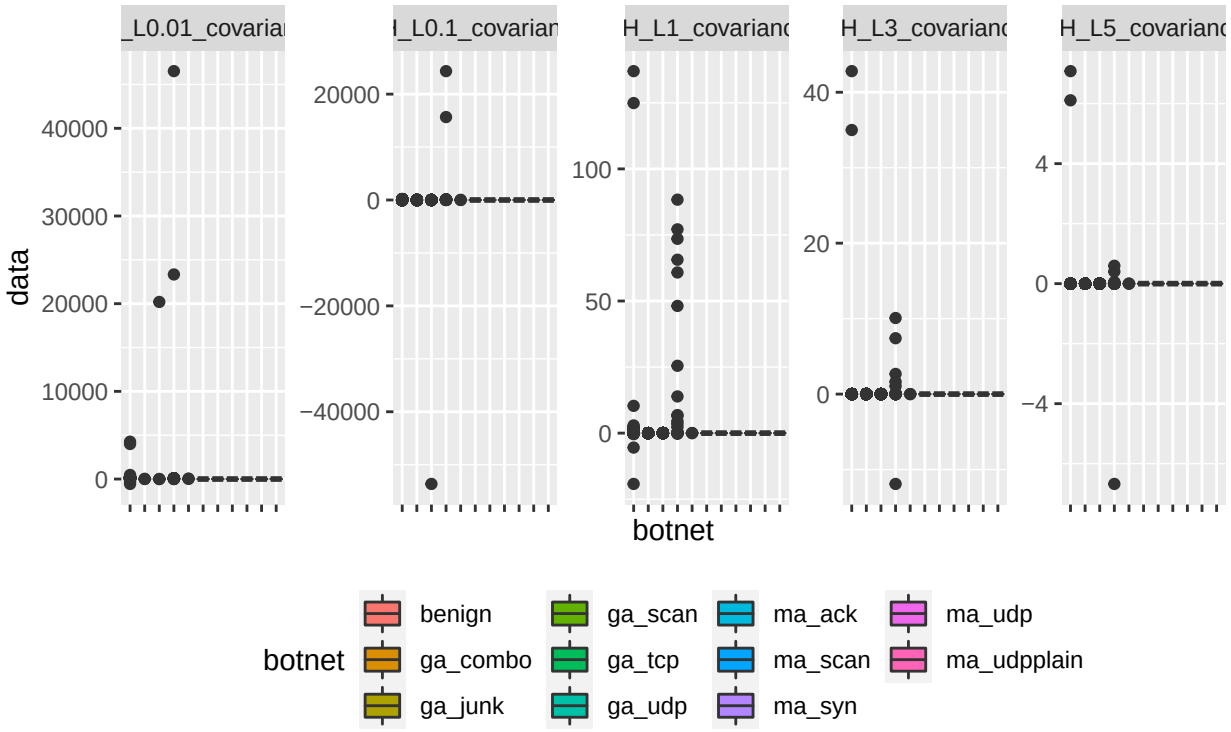
HH-stream also has *magnitude*, *radius*, *covariace* and *pcc* attributes, let's explore them:



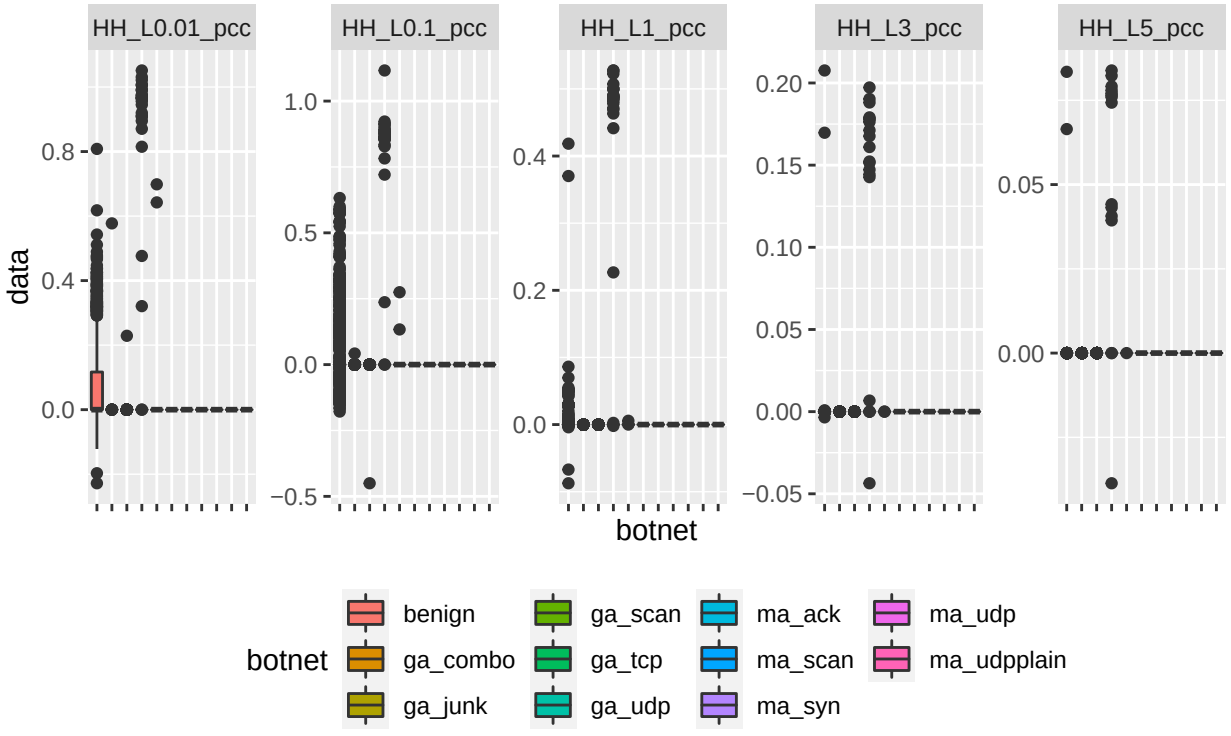
The plot shows that *magnitude* attribute can be used for separation **benign traffic** from the attacks.



The plot shows that *radius* attribute doesn't give any information how to separate **benign traffic** from the attacks and can be removed if needed.



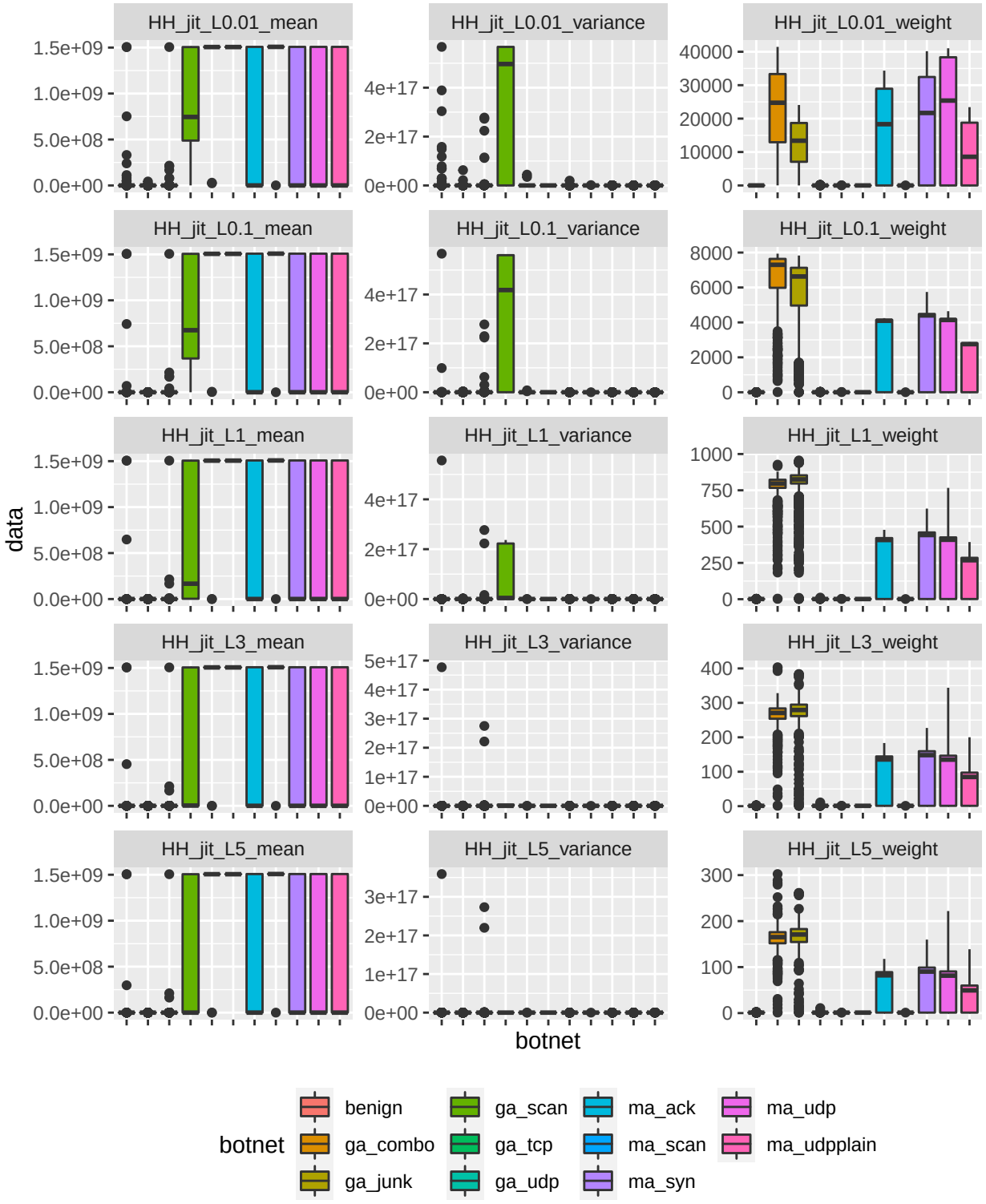
Covariance attribute can be used for separation attacks from **benign traffic**, and, probably, it's the one I've looked for.



The plot shows that, probably, *pcc* attrubute can also be used for separation.

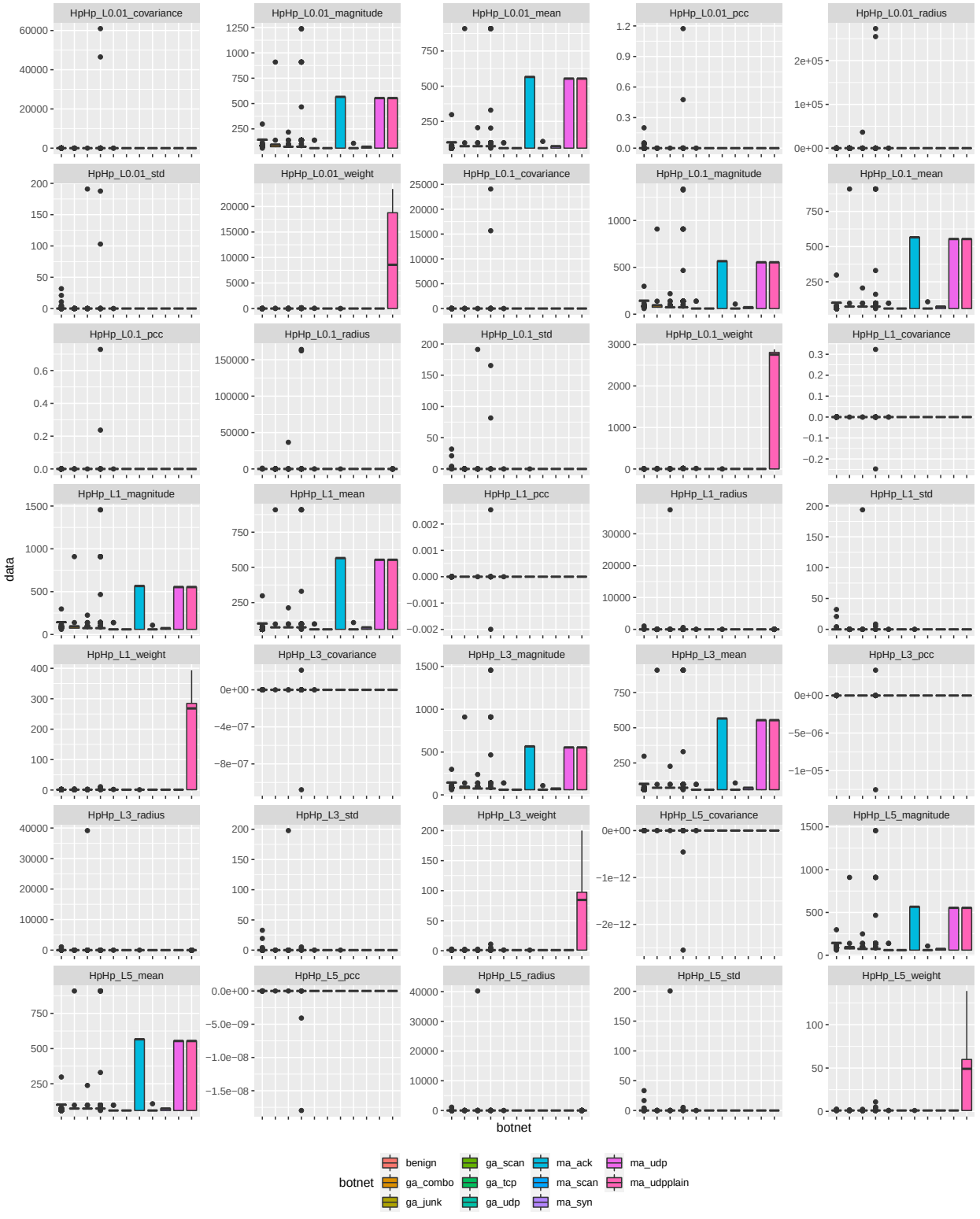
Next stream for exploration is HH_jit. And because I followed the same pattern when I explored other

streams, now I can make a quick glimpse.



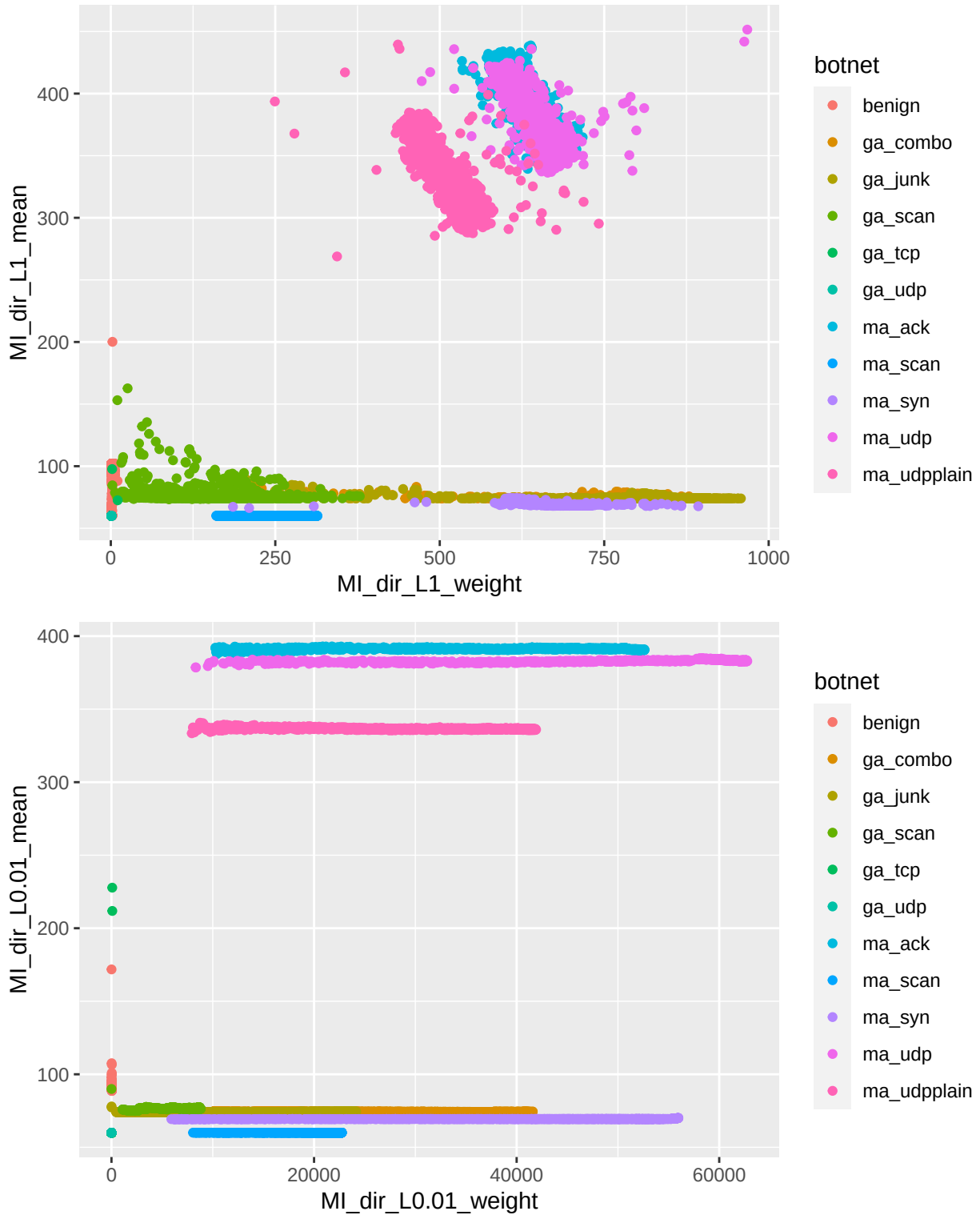
These plots have no new information how to separate **benign traffic** from attacks.

Now, I have left only HpHp strean for exploration, and again, I can make a quick glimpse:

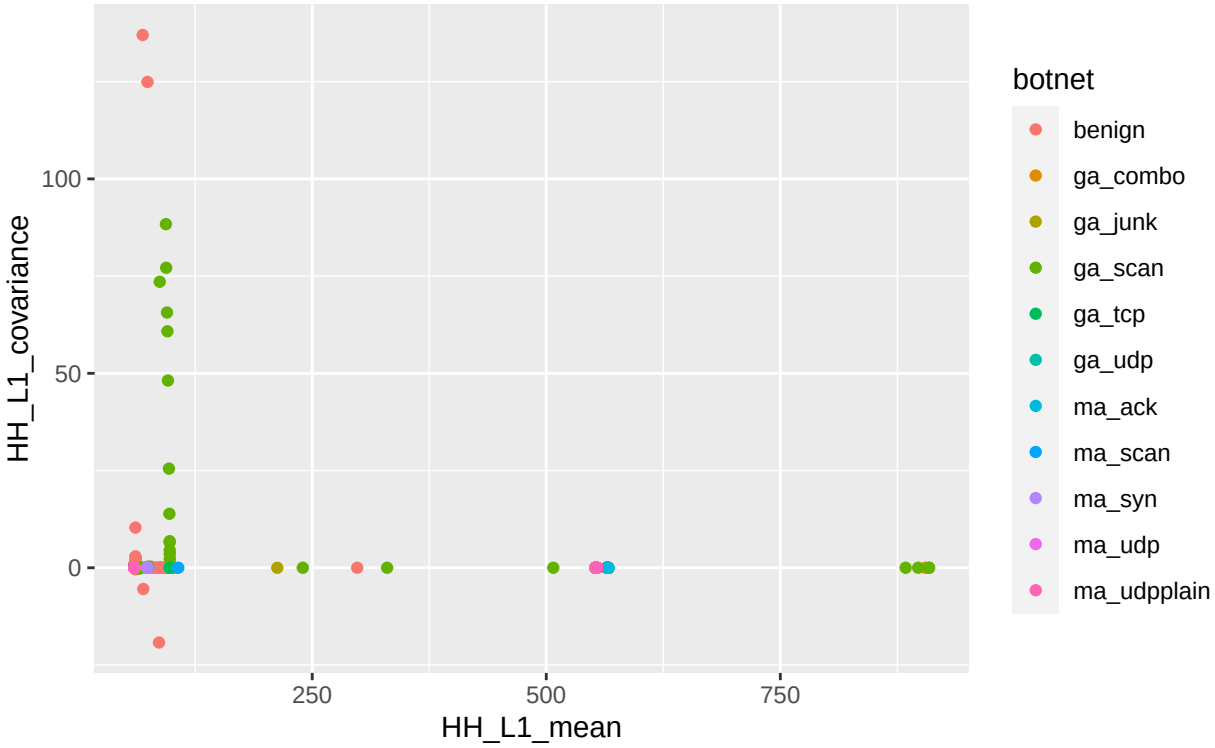


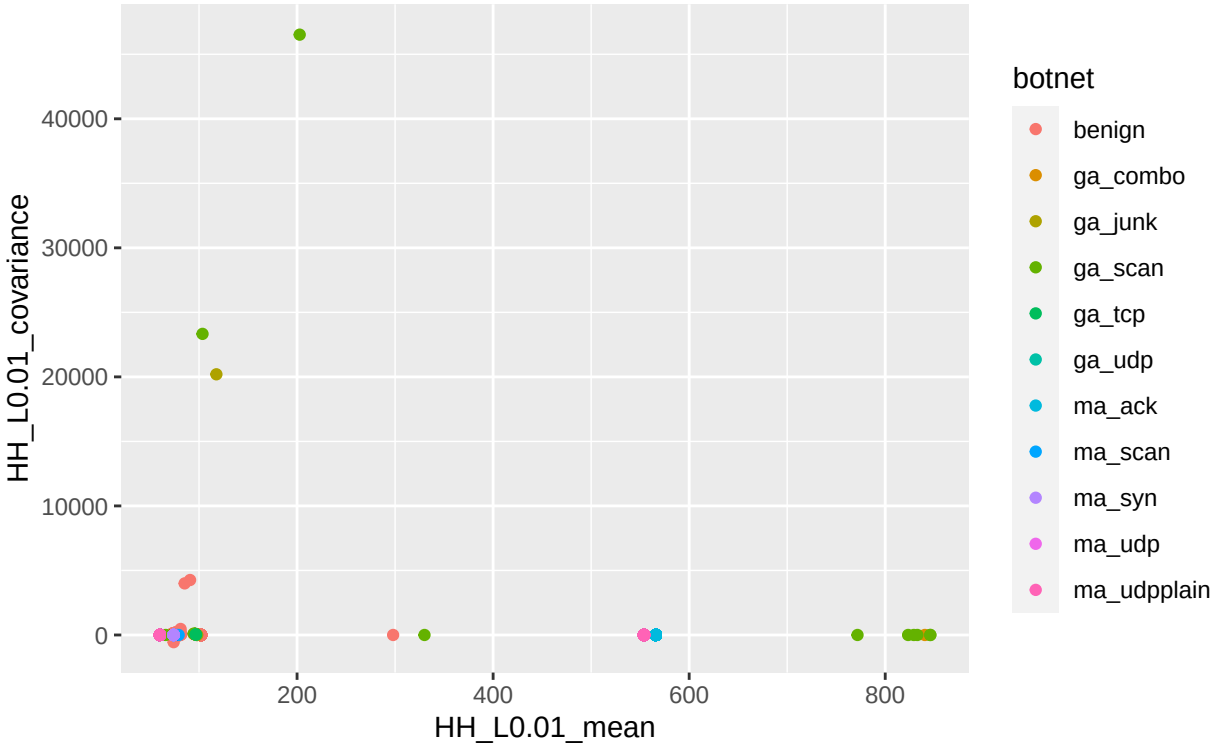
I didn't expect, that this stream gives me new information how to separate **benign traffic** from the attacks.

I've noticed that all the plots have more pronounced data at the small time-frames. I would like to explore what data will give me some attribute combinations, like *weight* vs. *mean* exactly on these small time-frames.



Pair *weight-mean* for L0.01 time-frame shows how easily some attacks can be separated from **benign traffic**, compared with L1 time-frame.





Pair *mean-covariance* better shows how **ga_tcp** and **ga_udp** can be separated from **benign traffic**.

In a result, all explorations show that I can use *weight*, *mean* and *covariance* attributes to make a decision how to separate **benign traffic** from attacks, and remove other statistic attributes if I need to reduce the data frame (or matrix) dimension.