

Nejdřív jsem napsal Python skript, který stáhne potřebné json soubory, změní formátování jejich obsahu a převede je do bulk formátu pro Elasticsearch. Tedy například

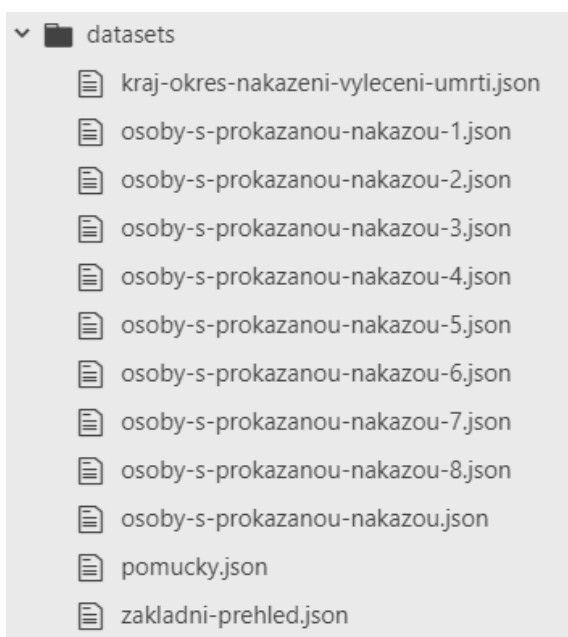
```
{"pomucka": "Brýle", "kraj_nuts_kod": "CZ010", "mnozstvi": 17770}

| ->

{"index": {"_index": "pomucky", "_type": "pomucky"}}

{"pomucka": "Br\u00fdle", "kraj_nuts_kod": "CZ-10", "mnozstvi": 17770}
```

Tento script upraví kódy krajů a okresu, případně přidá dekodovaný název kraje. Pokud je soubor dost velký, rozdělí výstup do několika souborů (podle dokumentace: "Bulk requests will yield much better performance than single-document index requests. " ... "Beware that too large bulk requests might put the cluster under memory pressure when many of them are sent concurrently"). Výstupem skriptu je tedy sada json souborů v bulk formátu:



Následně jsem vytvořil základní docker-compose obsahující standardní image Elasticsearch a Kibana. Po spuštění jsem poslal všechny připravené soubory:

```
$ for FILE in *; do echo curl -XPUT 'http://localhost:9200/_bulk' -H 'Content-Type: application/x-ndjson' --data-binary @$FILE; done
```

Stačilo v Kibana vytvořit indexy a pustit se do vizualizace.

Poznámka: předchozí kroky by šlo udělat i pomocí Logstash, ale přišlo mi to zbytečné, jelikož veškerá data už jsou v json formátu a v rámci úkolu je potřebujeme poslat pouze jednou staticky.

zakladni-prehled

Data

Options

Metrics

▶ Metric

Max datum

🔍

✖

▶ Metric

Max aktivni_pripady

🔍

✖

▶ Metric

Max aktualne_hospitalizovani

🔍

✖

▶ Metric

Max potvrzene_pripady_celkem

🔍

✖

▶ Met...

Max potvrzene_pripady_dnesni_den

🔍

✖

▶ Metric

Max provedene_testy_celkem

🔍

✖

▶ Met...

Max provedene_testy_vcerejsi_den

🔍

✖

▶ Met...

Max potvrzene_pripady_vcerejsi_den

🔍

✖

▶ Metric

Max umrti

🔍

✖

▶ Metric

Max vyleceni

🔍

✖

Add metrics

Buckets

Select buckets type

Split Group

Cancel

October 27th 2020, 01:00:00.000 170,528 6,191

datum

aktivni pripady

aktualne hospitalizovani

277,215 8,845 2,118,316 34,946 10,273

potvrzene pripady celkem

potvrzene pripady dnesni den

provedene testy celkem

provedene testy vcerejsi den

potvrzene pripady vcerejsi den

2,517

umrti

104,170

vyleceni

zakladni-prehled

Data

Options

Metrics

▶ Metric

Max datum

🔍

✖

▶ Metric

Max aktivni_pripady

🔍

✖

▶ Metric

Max aktualne_hospitalizovani

🔍

✖

▶ Metric

Max potvrzene_pripady_celkem

🔍

✖

▶ Met...

Max potvrzene_pripady_dnesni_den

🔍

✖

▶ Metric

Max provedene_testy_celkem

🔍

✖

▶ Met...

Max provedene_testy_vcerejsi_den

🔍

✖

▶ Met...

Max potvrzene_pripady_vcerejsi_den

🔍

✖

▶ Metric

Max umrti

🔍

✖

▶ Metric

Max vyleceni

🔍

✖

Add metrics

Buckets

Select buckets type

Split Group

Cancel

October 27th 2020, 01:00:00.000 170,528 6,191

datum

aktivni pripady

aktualne hospitalizovani

277,215 8,845 2,118,316 34,946 10,273

potvrzene pripady celkem

potvrzene pripady dnesni den

provedene testy celkem

provedene testy vcerejsi den

potvrzene pripady vcerejsi den

2,517

umrti

104,170

vyleceni

osoby-s-prokazanou-nakazou

Data

Options

Value

Aggregation

Count

Custom Label

počet nakažených

Advanced

Buckets

shape field

Aggregation

Terms

Field

kraj_nuts_kod.keyword

Order By

metric: počet nakažených

Order

Descent

Size

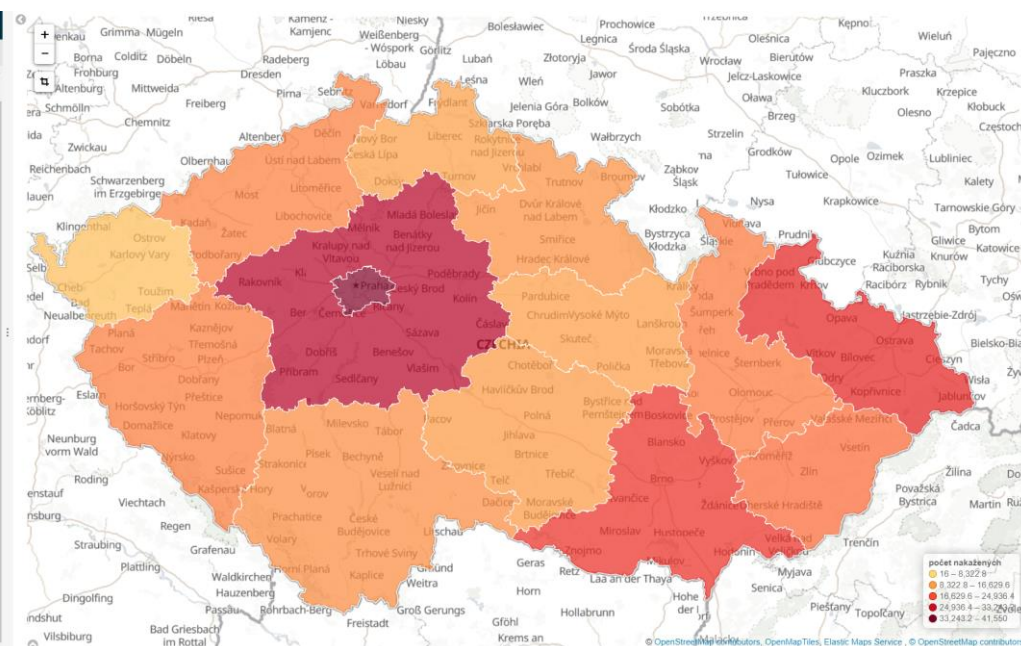
100

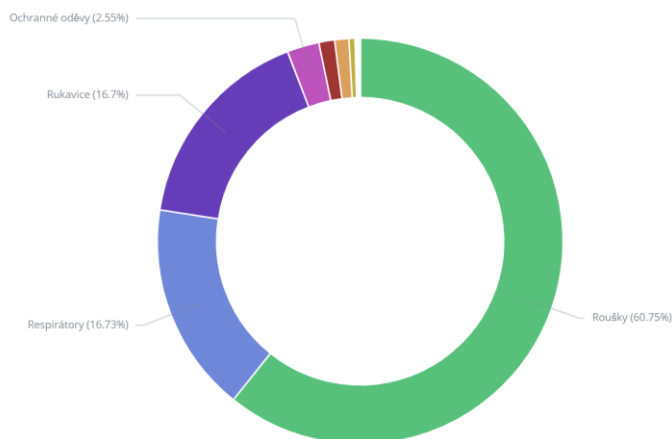
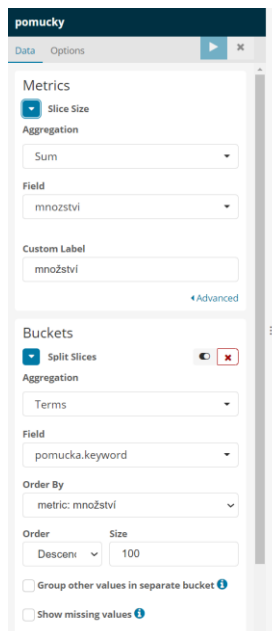
Group other values in separate bucket

Show missing values

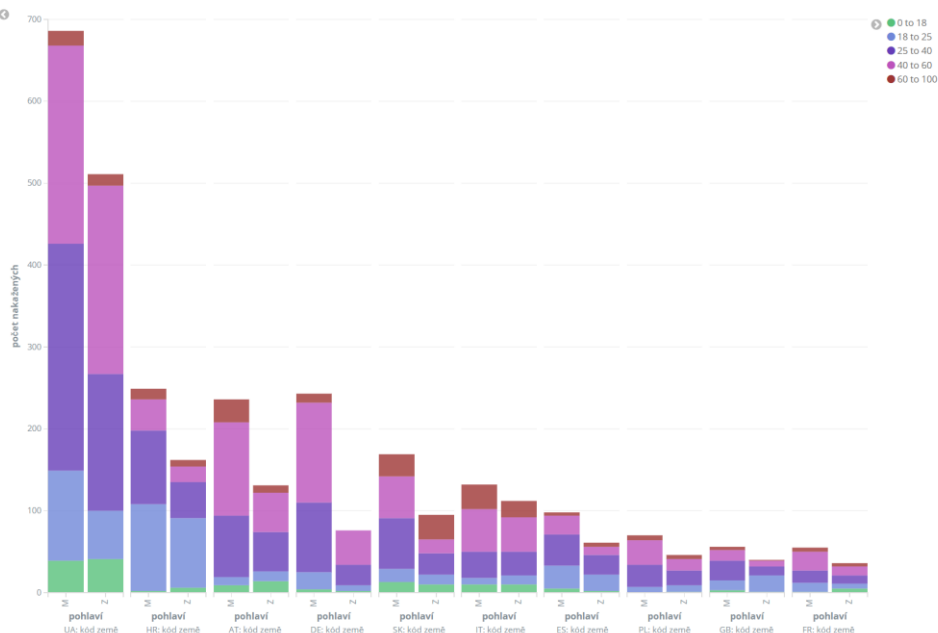
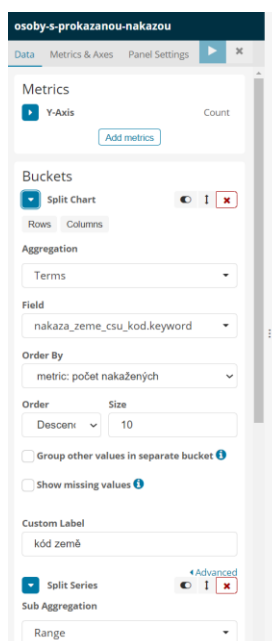
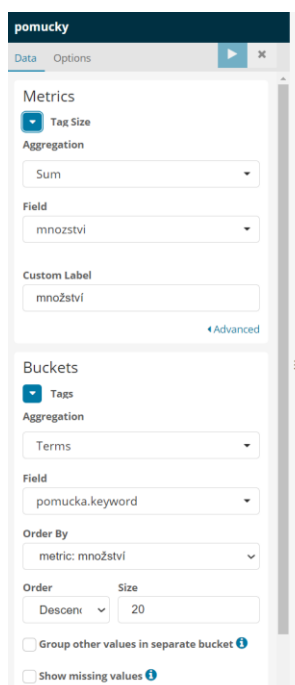
Custom Label

Advanced





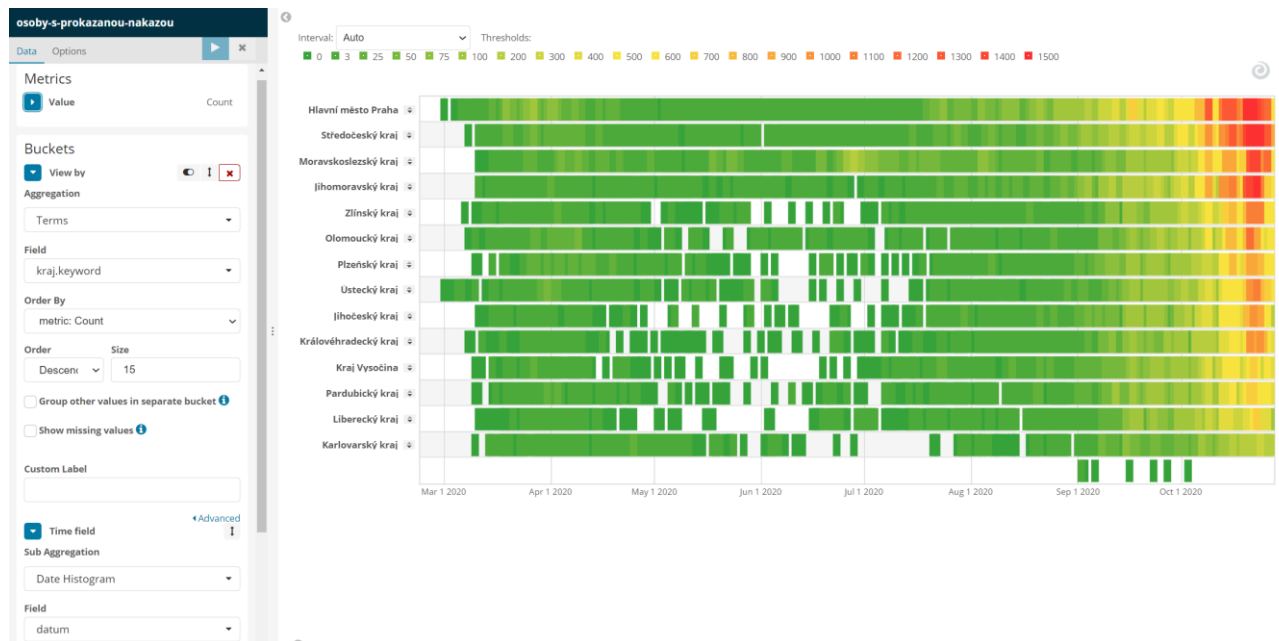
- Roušky
- Respirátory
- Rukavice
- Ochranné oděvy
- Odběrové sady
- Rychlotesty
- Návrky na obuv
- Brýle
- Šitý
- Dezinfekce (I)
- Filtry k polomaskám
- Polomasky
- Ostatní
- Teploměry
- Termokamery
- Plicní ventilátory
- Příslušenství k plicním...



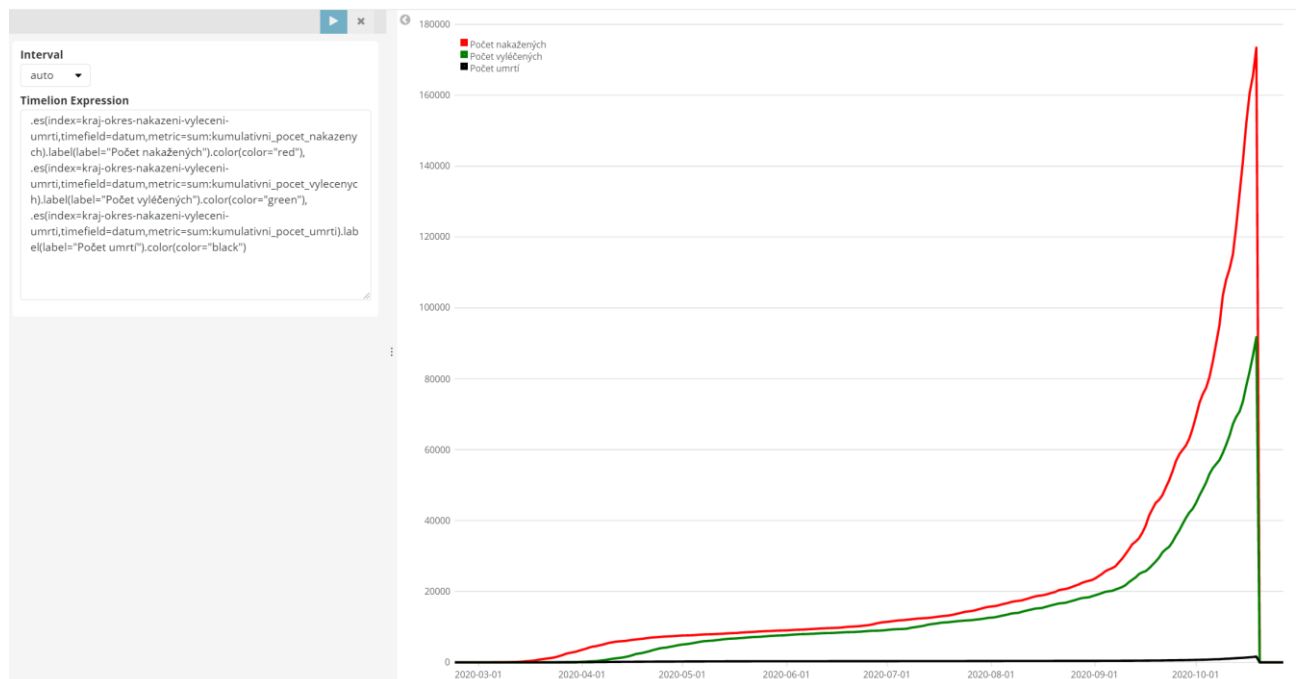
Na vytvoření následující vizualizace jsem použil plugin Swimlane:

```
$ docker exec -it kibana bash
```

```
$ bin/kibana-plugin install https://github.com/prelert/kibana-swimlane-  
vis/releases/download/v6.2.4/prelert_swimlane_vis-6.2.4.zip
```



Nakonec jsem zkusil vytvořit graf pomocí Timelion. Tento graf nevypadá na konci moc šikovně, kvůli tomu, že dataset "COVID-19: Přehled epidemiologické situace dle hlášení krajských hygienických stanic podle okresu" není tak aktuální jako ostatní použité.



[illegible]