

Project:

Suvinisu fenoloogiline areng maakonniti 2018. a suvel

Valentina Sagris

Oktoober, 2019

1. Sissejuhatus

Projekti eesmärgiks on R-tarkvara kursuses omandatud oskuste kinnitamine ja nende rakendamine erialase analüüsi läbiviimisel ja analüüsi tulemuste vormistamisel.

Oma projektis uurisin põllukultuuride fenoloogilist arengut 2018 a. kevadel ja suvel. Püstitatud hüpoteesiks on hästi tuntud fenomeni tõestamine ja illustreerimine: Kesk- ja Kagu-Eesti põllukultuurid arenevad üks kuni kaks nädalat varem, kui mujal Eestis.

Selleks kasutasin satelliitseire andmeid, mis pärinevad Copernicuse programmi Sentinel-2 sensorilt. Fenoloogilistes ja muudes taimkatte uuringutes on laialdaselt kasutatav nn. normeeritud vahe vegetatsiooniindeks (normalized difference vegetation index, NDVI), mida arvutatakse punase ja infrapunase spektraalkanaleid kombineerides. Indeks on väga tundlik taimestiku klorofüllisalduse suhtes - mida kõrgem klorofüllisaldus, seda kõrgem indeks. Projektis kasutasin eelnevalt arvutatud põldude keskmiseid NDVI väärtusi csv formaadis ja andmete kaardiga sidumiseks põldude tsentroidide kihti *shape* formaadis. Andmete generaliseerimiseks maakonna tasandile kasutasin maakonna andmekihti *shape* formaadis. Ruumiliste andmetega töötamiseks kasutasin R paketi nimega *sf* ehk '*simple feature*'.

2. Andmed

NDVI andmeid lugesin sisse vi dataframe-i, ja kuna need on „pikas“ formaadis, viisin need üle „laia“ formaati:

```

>str(ts2)
data.frame': 8219 obs. of 173 variables:
 $ id : int 172323 172430 172463 172493 172604 172814 172850 17285
3 172854 172905 ...
 $ 2018-04-07: num NA NA NA 0.248 NA ...
 $ 2018-04-09: num NA NA NA NA NA NA NA NA NA NA ...
 $ 2018-04-10: num NA NA NA NA NA NA NA NA NA NA ...
 $ 2018-04-11: num NA NA NA NA NA NA NA NA NA NA ...
 $ 2018-04-12: num NA 0.251 NA 0.23 NA ...
 $ 2018-04-13: num NA NA 0.232 0.228 0.225 ...
 $ 2018-04-14: num NA NA NA NA NA NA NA NA NA NA ...
 $ 2018-04-15: num NA NA NA NA NA NA NA NA NA NA ...
 $ 2018-04-18: num NA NA 0.207 NA NA ...
 $ 2018-04-20: num 0.159 0.256 0.186 NA NA ...

```

Kuna andmed on üsna hõredad (väärtused puuduvad peamiselt pilvede tõttu), agregeerisin neid dekaadide ehk 10 päeva kaupa, leides igale dekaadile keskmise. Mõnedel põldudel juuni teises ja kolmandas dekaadis ei olnud ühtegi väärtust ja seetõttu on tulemuseks NaN.

```

>dec
   id      mai1      mai2      mai3      juuni1      juuni2      juuni3
1 172323 0.1355857 0.1347820 0.2873837 0.3959755 0.6108550      NaN
2 172430 0.3616643 0.2478800 0.3101625 0.4508540 0.7123870 0.7619350
3 172463 0.2085790 0.3324083 0.5384820 0.6047255 0.6493920      NaN
4 172493 0.1627837 0.1614400 0.3120915 0.5082950      NaN 0.6593507
5 172604 0.2360223 0.2357230 0.3191395 0.4812083 0.6138540 0.7119425
6 172814 0.2671485 0.2950243 0.6229827 0.7166000 0.7258020 0.8237080
7 172850 0.2105045 0.1942197 0.1942010 0.2591875 0.2938420 0.5370970
8 172853 0.2718677 0.2482412 0.2014490 0.2137280 0.4475500      NaN
9 172854 0.2951350 0.2317260 0.1939120 0.2035510 0.4530835      NaN
10 172905 0.3152480 0.3306260 0.5180300 0.6838975 0.6475690 0.7160690

```

Põldude tsentroidide andmekiht on järgmise struktuuriga:

```

> str(pld)
Classes 'sf' and 'data.frame': 8221 obs. of 4 variables:
 $ f_id : num 13657459 13657455 13659107 13659108 13659110 ...
 $ id : num 208360 208377 208408 208405 208411 ...
 $ geometry_v: Factor w/ 1 level "True": 1 1 1 1 1 1 1 1 1 1 ...
 $ geometry :sfc_POINT of length 8221;

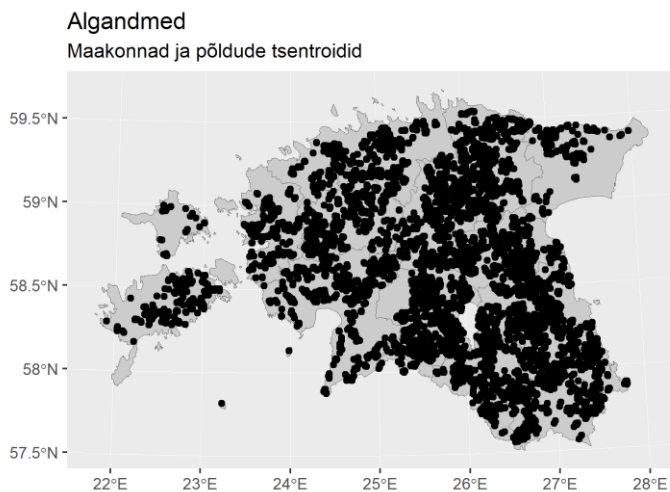
```

Maakondade andmekiht on järgmise struktuuriga:

```

> str(mk)
Classes 'sf' and 'data.frame': 15 obs. of 5 variables:
 $ MNIMI : Factor w/ 15 levels "Harju maakond",...: 3 1 4 12 14 8 13 ..
 $ MKOOD : Factor w/ 15 levels "0037","0039",...: 3 1 4 12 14 85 5 ...
 $ Shape_Leng: num 347726 722983 314129 394609 423453 ...
 $ Shape_Area: num 3.34e+09 4.34e+09 2.60e+09 2.99e+09 3.42e+09 ...
 $ geometry :sfc_MULTIPOLYGON of length 15;

```

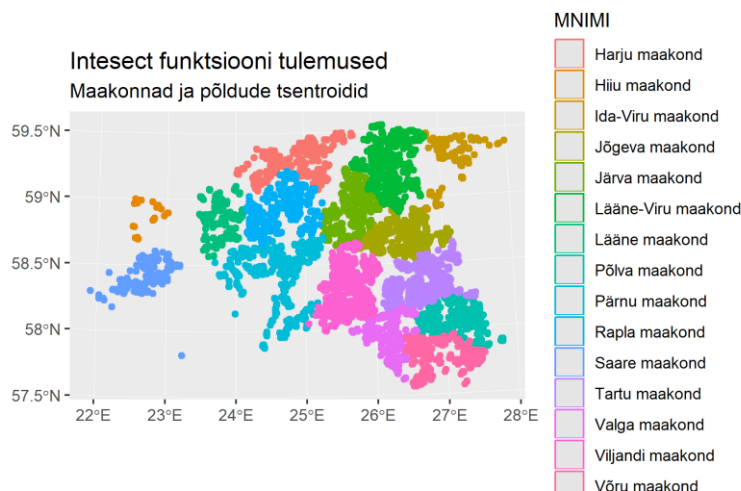


Joonis 1. Ruumilised algandmed – maakondade polügoonid ja põldude tsentroidid.

3. Metoodika

Projekti analüüsimetoodika koosneb järgmistest sammudest:

1. Samm. Leidsin, mis maakonda jääb iga põllu tsentroid, kasutasin selleks sf paketi funktsiooni *st_intersection*. Tulemuseks on dataframe nimega *out*



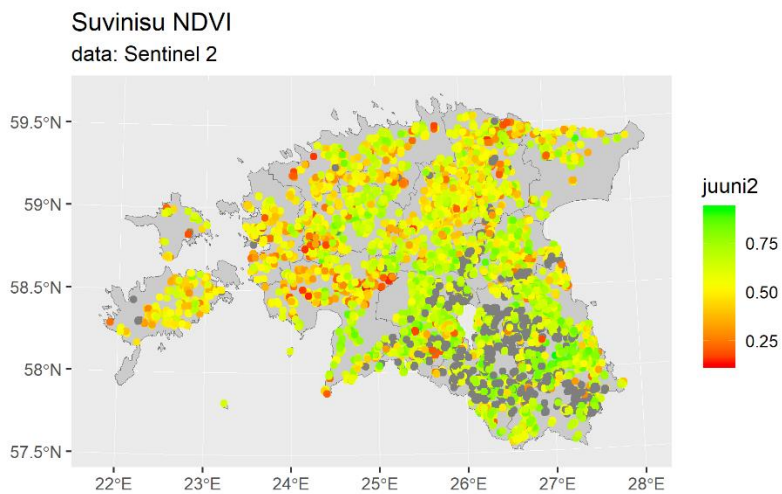
Joonis 2. Põldude jagunemine maakondadesse.

2. Samm. Liitsin tsentroididele NDVI dekaadide andmed, kasutades funktsiooni *merge*
3. Samm. Arvutasin maakonna keskmise NDVI, kasutades funktsiooni *aggregate* (eelnevalt kustutasin NaN)

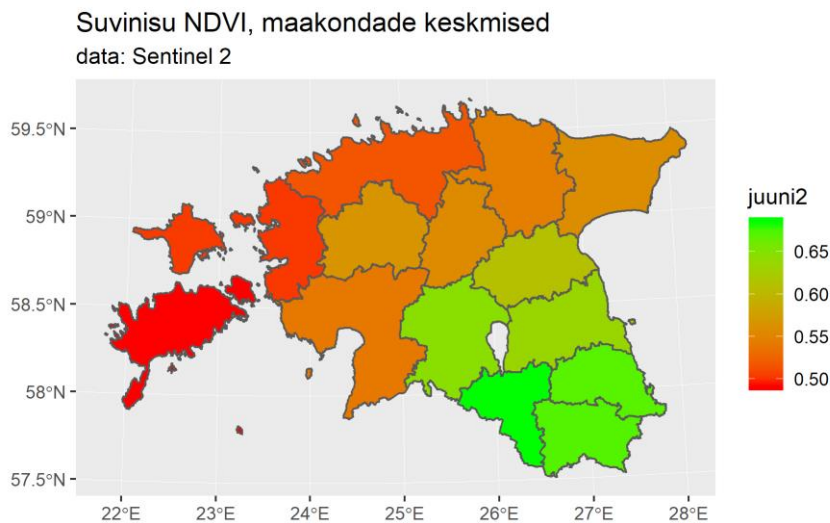
4. Tulemused

Joonisel 3 on toodud põldude kaupa NDVI väärtused juunikuu teises dekaadis. Osa põlde, peamiselt Kesk- ja Kagu Eestis, olid kogu dekaadi ulatuses pilvedega kaetud (halli värviga). Joonisel tuleb hästi välja, et Kesk- ja Kagu-Eestis on suvinisu rohkem arenenud, järgmisena kasvab vili hästi Virumaal, Harju ja saared on väga heterogeensed ning viimasena selles reas

tulevad Pärnu ja Läänemaa. Samas joonisel on näha, et kõrvuti võivad asuda põllud väga erinevas arenguastmes viljaga, ja tegelikult pilt ei ole nii ühtlane piirkonniti nagu seda inimsilm näeb.



Joonis 3. Suvinisu NDVI põldude kaupa.



Joonis 4. Suvinisu NDVI, maakondade keskmised.

Joonisel 4 on sama perioodi andmed, kuid juba agregeerituna maakonna tasemel. Pilt on kergemini loetav ja R tuleb selle joonistamisega palju kiiremini toime. Selgesti on näha loode-kagu suunalist trendi.

Kokkuvõte

R tarkava on tõhus töörisk andmeanalüüsiks, see on paindlik tänu oma arvukatele lisapakettidele. Selle graafiline väljund võimaldab tulemusi hõlpsasti visualiseerida. Projekti käigus saime näitlikustada teadmisi Eesti põllukultuuride fenoloogiast.