

# 1 Teht.1

## 1.1 10-means, järjestetty data

Within cluster sum of squared errors: 4.4954

Clustered Instances  
0 1460 ( 13%)  
1 334 ( 3%)  
2 3798 ( 33%)  
3 3 ( 0%)  
4 54 ( 0%)  
5 3302 ( 29%)  
6 37 ( 0%)  
7 783 ( 7%)  
8 22 ( 0%)  
9 1672 ( 15%)

Tutkitaan sisältöä:

Klusteri 0: aika suuri. Hero < 0 eli mutanteilla geeni passiivisempi kuin terveillä. Muuttujat korreloimattomia. Klusteri 1: melko pieni. Hero ≥ 1 eli mutanteilla geeni aktiivisempi. Muuttujat korreloimattomia. Klusteri 2: Iso. Hero < 0 eli eli mutanteilla geeni passiivisempi. Muuttujat korreloimattomia. Klust. 3 vain 3 alkiota! (Ucp1, Alb, Pck1) rero2 aika iso (> 15), toiset pieniä. Eli rottavanhuksilla geeni paljon aktiivisempi. Klusteri 4: pieni. rero2 > 1. Rottavanhuksilla geeni aktiivisempi. r(Hero, Rero1) = 0.44, toiset korreloimattomia. Klusteri 5: Iso. Kaikki lähellä 0:aa. Jollain yhden (rero1:n) arvo hieman suurempi. Muuttujat korreloimattomia. Klusteri 6: Suuri. hero > 2 eli mutanteilla geeni aktiivisempi. Muuttujat korreloimattomia. (suurin r(hero, rero1) = 0.20) Klusteri 7: Keskikokoinen. Hero hieman koholla. Muuttujat korreloimattomia. Klusteri 8: Pieni (22 alkiota). Erikoinen! roero1:n arvot valtavan suuria! Jopa 179 (Rrm2). Toiset kaksi muuttujaa hyvin pieniä. Eli vauvarotilla geeni tosi aktiivinen ja hiljenee aikuisena. Klusteri 9: Iso. Hero koholla eli mutanteilla geeni aktiivisempi. Muuttujat korreloimattomia.

Sama mutta density-based (saadaan jakaumat): (tuli aavistuksen eri klusterointi)

Fitted estimators (with ML estimates of variance):

Cluster: 0 Prior probability: 0.1273

Attribute: hiero Normal Distribution. Mean = -0.4044 StdDev = 0.1142

Attribute: roero1 Normal Distribution. Mean = 0.2108 StdDev = 1.0221

Attribute: roero2 Normal Distribution. Mean = -0.02 StdDev = 0.1644

Taulukko 1: 10-means kaikki 3 muuttujaa

Attribute	Full Data (11465)	0 (1460)	1 (334)	2 (3798)	3 (3)	4 (54)	5 (3302)	6 (37)	7 (783)	8 (22)	9 (1672)
hiero	0.0863	-0.4044	1.3563	-0.1265	0.07	0.0624	0.0809	2.7281	0.7785	0.0718	0.3735
roero1	0.341	0.2108	0.1865	0.3209	-0.4867	0.7131	0.165	-0.0151	0.1588	56.0305	0.2289
roero2	0.0194	-0.02	-0.0061	0.0051	16.2567	2.6706	0.0125	-0.0027	0.0039	0.3382	-0.006

Cluster: 1 Prior probability: 0.0292  
Attribute: hiero Normal Distribution. Mean = 1.3563 StdDev = 0.2199  
Attribute: roero1 Normal Distribution. Mean = 0.1865 StdDev = 0.964  
Attribute: roero2 Normal Distribution. Mean = -0.0061 StdDev = 0.2132  
Cluster: 2 Prior probability: 0.3311  
Attribute: hiero Normal Distribution. Mean = -0.1265 StdDev = 0.0665  
Attribute: roero1 Normal Distribution. Mean = 0.3209 StdDev = 1.8748  
Attribute: roero2 Normal Distribution. Mean = 0.0051 StdDev = 0.1655  
Cluster: 3 Prior probability: 0.0003  
Attribute: hiero Normal Distribution. Mean = 0.07 StdDev = 0.207  
Attribute: roero1 Normal Distribution. Mean = -0.4867 StdDev = 0.1801  
Attribute: roero2 Normal Distribution. Mean = 16.2567 StdDev = 0.98  
Cluster: 4 Prior probability: 0.0048  
Attribute: hiero Normal Distribution. Mean = 0.0624 StdDev = 0.2702  
Attribute: roero1 Normal Distribution. Mean = 0.7131 StdDev = 4.2777  
Attribute: roero2 Normal Distribution. Mean = 2.6706 StdDev = 1.3797  
Cluster: 5 Prior probability: 0.2878  
Attribute: hiero Normal Distribution. Mean = 0.0809 StdDev = 0.0714  
Attribute: roero1 Normal Distribution. Mean = 0.165 StdDev = 1.0267  
Attribute: roero2 Normal Distribution. Mean = 0.0125 StdDev = 0.1757  
Cluster: 6 Prior probability: 0.0033  
Attribute: hiero Normal Distribution. Mean = 2.7281 StdDev = 1.1784  
Attribute: roero1 Normal Distribution. Mean = -0.0151 StdDev = 0.3518  
Attribute: roero2 Normal Distribution. Mean = -0.0027 StdDev = 0.1714  
Cluster: 7 Prior probability: 0.0683  
Attribute: hiero Normal Distribution. Mean = 0.7785 StdDev = 0.1349  
Attribute: roero1 Normal Distribution. Mean = 0.1588 StdDev = 0.6243  
Attribute: roero2 Normal Distribution. Mean = 0.0039 StdDev = 0.2096  
Cluster: 8 Prior probability: 0.002  
Attribute: hiero Normal Distribution. Mean = 0.0718 StdDev = 0.2966  
Attribute: roero1 Normal Distribution. Mean = 56.0305 StdDev = 33.1648  
Attribute: roero2 Normal Distribution. Mean = 0.3382 StdDev = 1.1313  
Cluster: 9 Prior probability: 0.1458  
Attribute: hiero Normal Distribution. Mean = 0.3735 StdDev = 0.0976  
Attribute: roero1 Normal Distribution. Mean = 0.2289 StdDev = 0.998  
Attribute: roero2 Normal Distribution. Mean = -0.006 StdDev = 0.1744

## 1.2 Vain rottamuuttajat

Tulee ihan erilaisia!

Taulukko 2: 10-means (vain rottien attr)

Attribute	Full Data (11466)	0 (8776)	1 (326)	2 (376)	3 (264)	4 (340)	5 (281)	6 (683)	7 (110)	8 (75)	9 (235)
roero1	0.0	0.0	0.4	-0.38	0.98	-0.25	-0.0	-0.09	0.04	-0.88	-0.41
roero2	0.0	-0.05	-0.09	0.04	-0.11	0.06	0.07	0.0	0.04	0.18	-0.13

0 8885 ( 77%)  
1 290 ( 3%)  
2 361 ( 3%)  
3 263 ( 2%)  
4 329 ( 3%)  
5 278 ( 2%)  
6 669 ( 6%)  
7 106 ( 1%)  
8 53 ( 0%)  
9 232 ( 2%)

## 2 Teht.2

### 2.1 Average-link, $K = 1000$

Esimerkkejä:

Outliereita:  $2*\sin^4 + \pi$  yksin huomasi, c++osasi, osasi rautaa, dataa, raapustaa shell, lähellä

usein sama taivutusmuoto (jos ei muuten): tehnyt, tekemiseen/tekemisessä  
kauniita klustereita: koulutyö, ohjelmointikieli/-kurssi, internet, ohjelma, perus-alkuiset

Toimi varsin hyvin! (Eikä juuri eroa vaikka sotki datan välillä.)

Kun  $K=500$ , ei yhtä hyviä.

### 2.2 Ward ja $K=1000$

Omituisia. Miksi esim. kokeilu, aiheisiin, iltana ja freehand samassa?

Näyttää kuin olisi eri etäisyysmitta! (mutta ei ollut)

### 2.3 Complete ja $K=1000$

Nättejä tuloksia, samantap. kuin average, mutta ehkä jakanut joitain average (samankantaisia) klustereita osiin?

## 3 Teht.3

Mansikilla outo outlier alussa klo 14–15. (Samaan aikaan koko navetalla paljon vierailuja.)

Mustikki syö harevmmmin mutta enemmän.

Koko navetta 24h keskiarvot: iso piikki aamulla n. 8 aikaan, seuraavat huiput eivät yhtä selviä, klo 15 ja 21 aikaan. Klo 5 aikaan hiljaisinta.

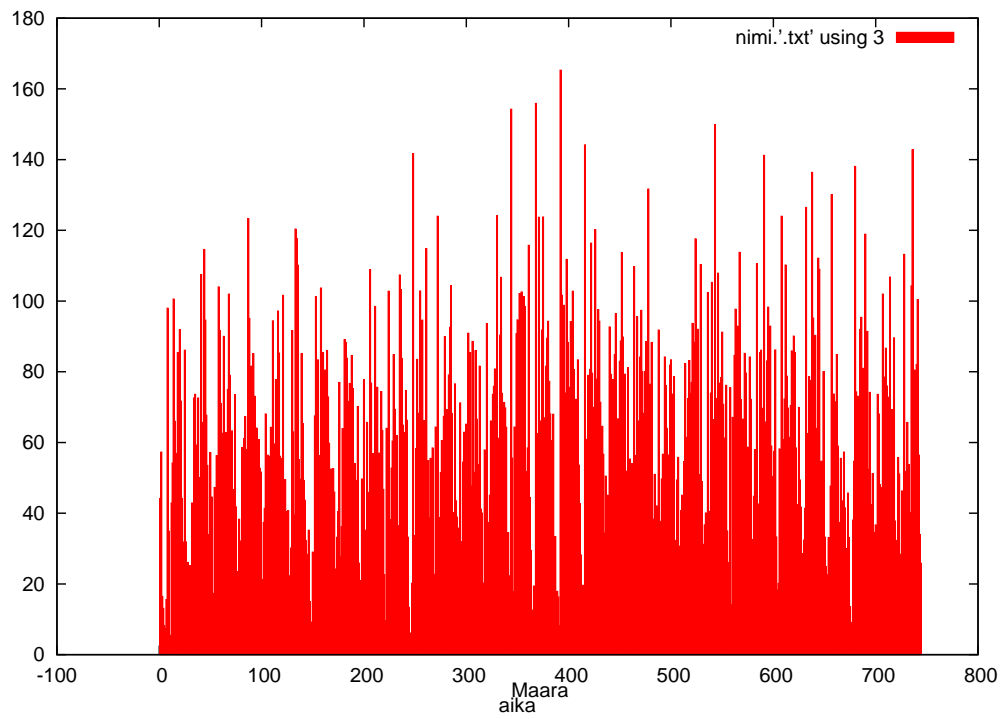
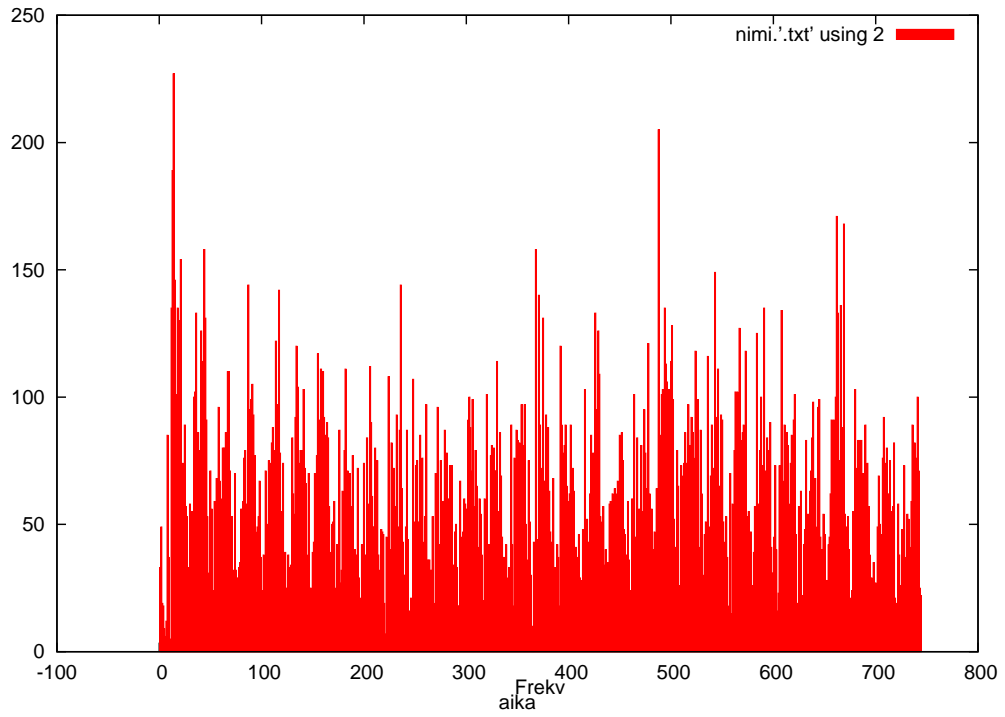
Mansikki ja Mustikki 24h keskiarvot: Mansikilla paljon enemmän vierailuja ja isot erot vrk-aikojen välillä: suurin huippu n. 19 aikaan, 0–6 hiljaista. Mustikilla vierailuja aika tasaisesti, jopa klo 1 eräs huippu. Mansikin määrissä sama kuvio, lisääntyvät klo 19 kohden ja sitten taas laskua. Mustikillakin suurin piikki illalla, mutta vasta klo 20? Muuallakin korkeita piikkejä.

Määrien korrelogrammi: Mansikki näyttää säännöllisemmältä: selvempi piikki 24h kohdalla. Mustikin piikki 23h kohdalla eli vuorokausi pikemminkin 23h??? (Mutta ei selvä.)

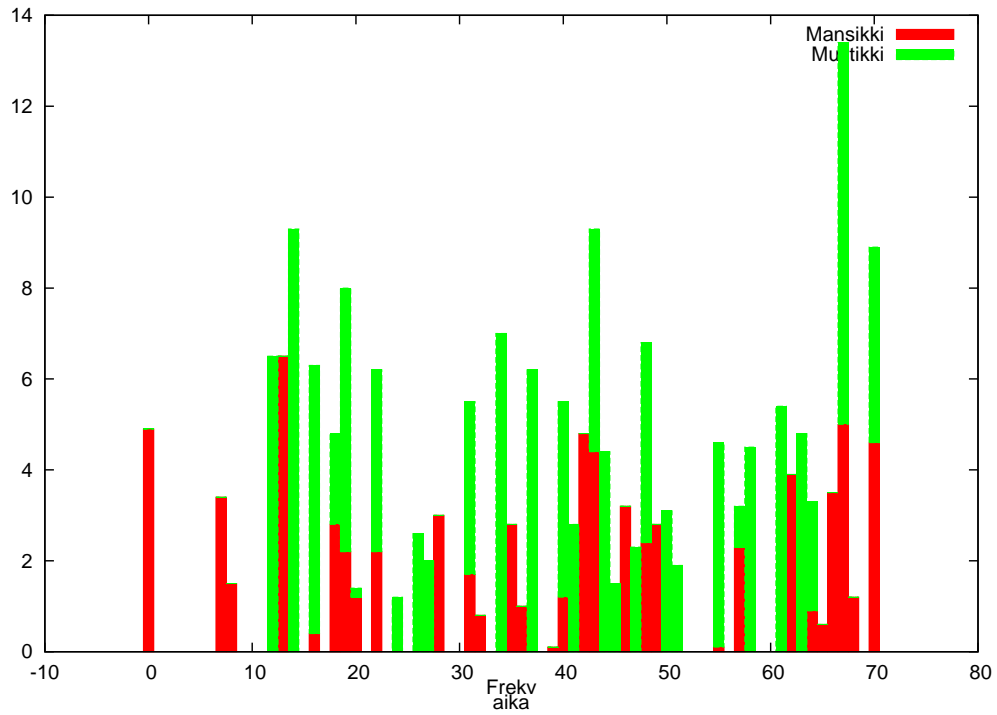
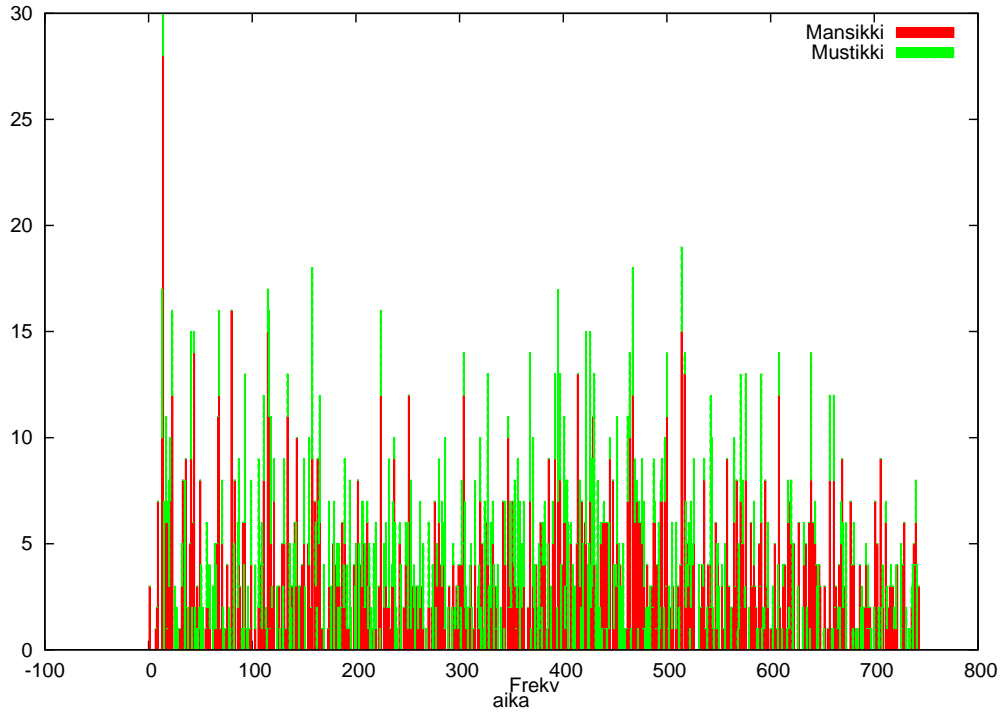
Huom! 0-kohdan neg. piikillä ei merkitystä.

Frekvenssien korrelogrammi: Kummallakin korkea piikki 24h kohdalla, mutta Mustikin korkein piikki 3h kohdalla..?

Kuva 1: Koko navetta

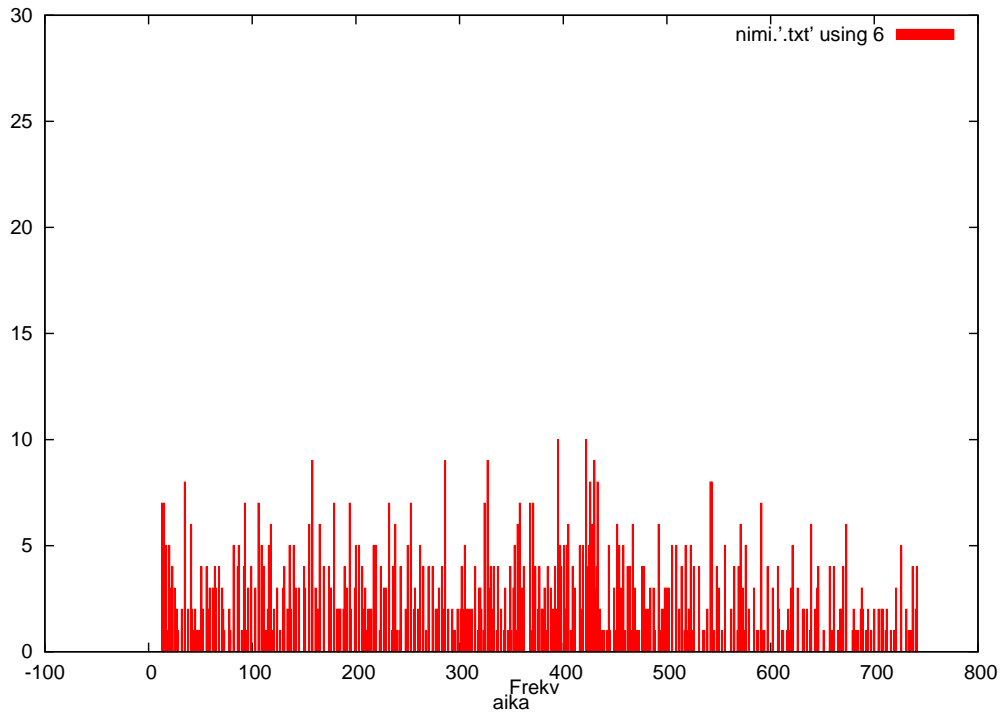
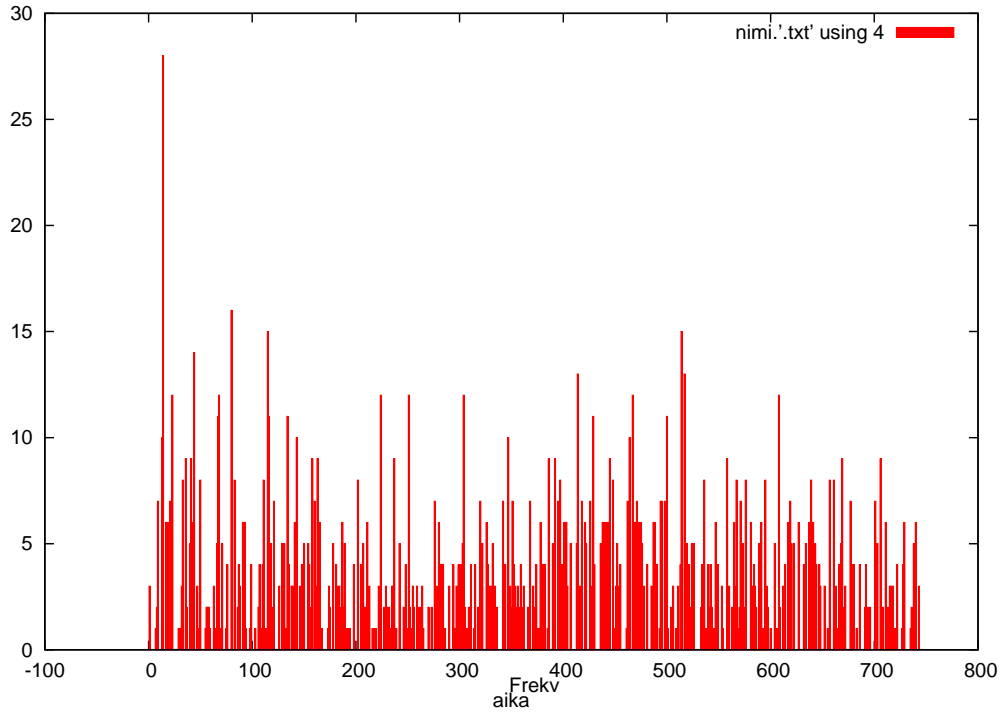


Kuva 2: Frekvenssit: Mansikki ja Mustikki

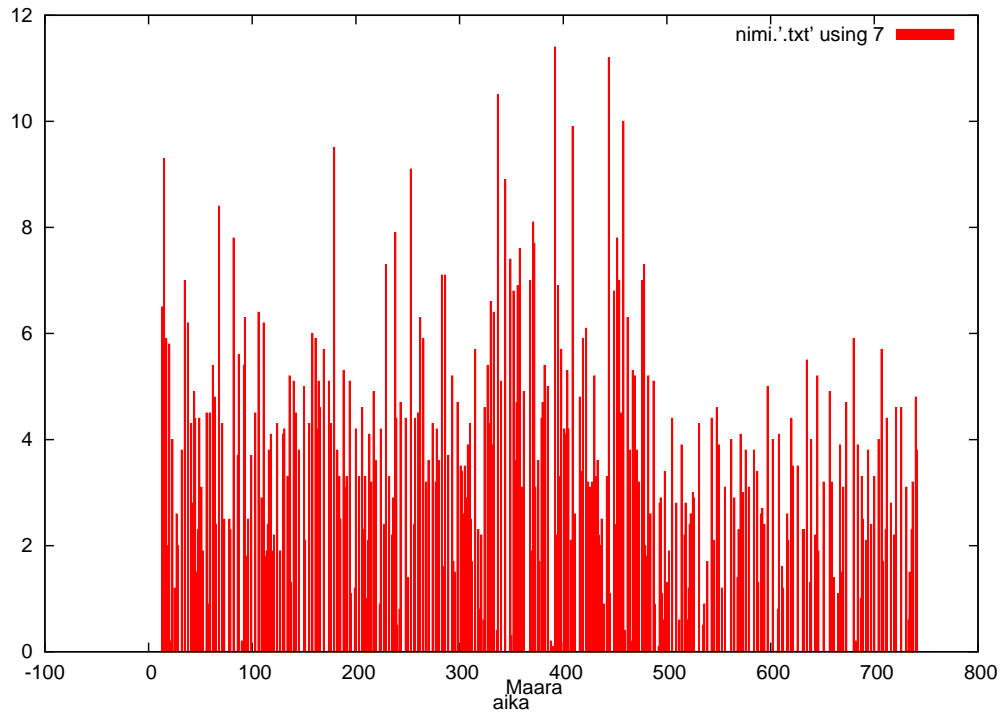
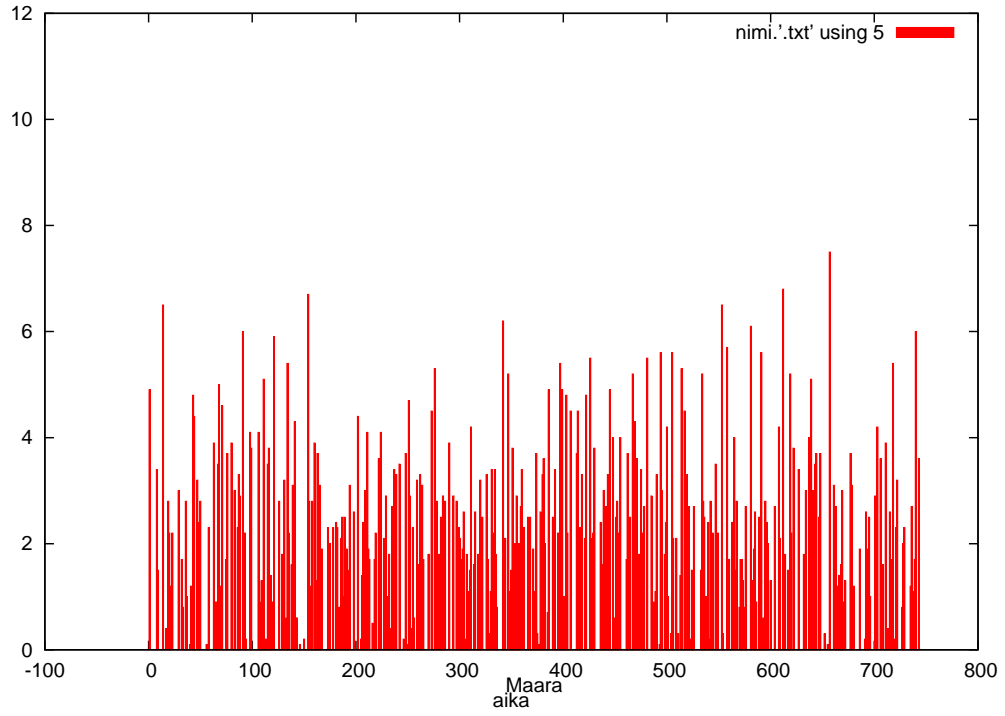




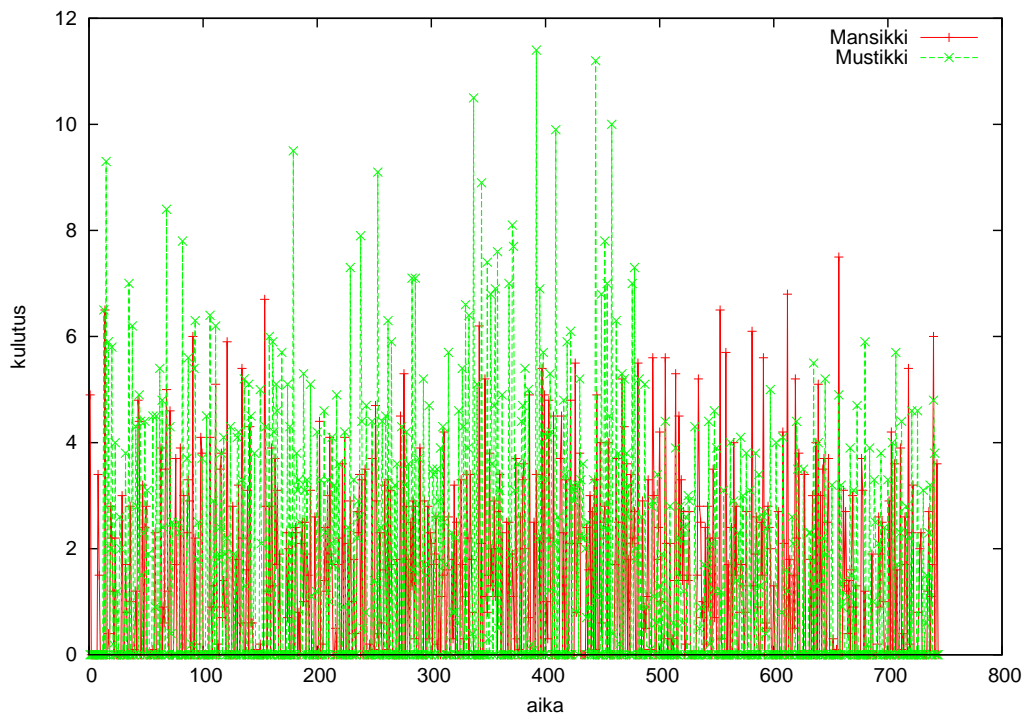
Kuva 3: Frekvenssit: Mansikki ja Mustikki



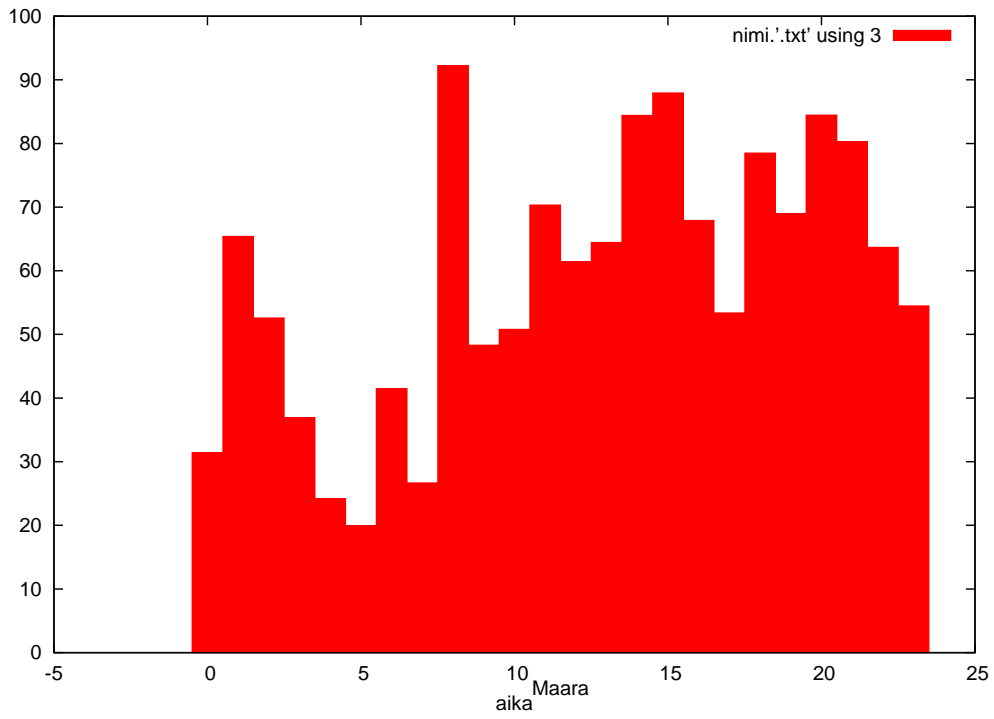
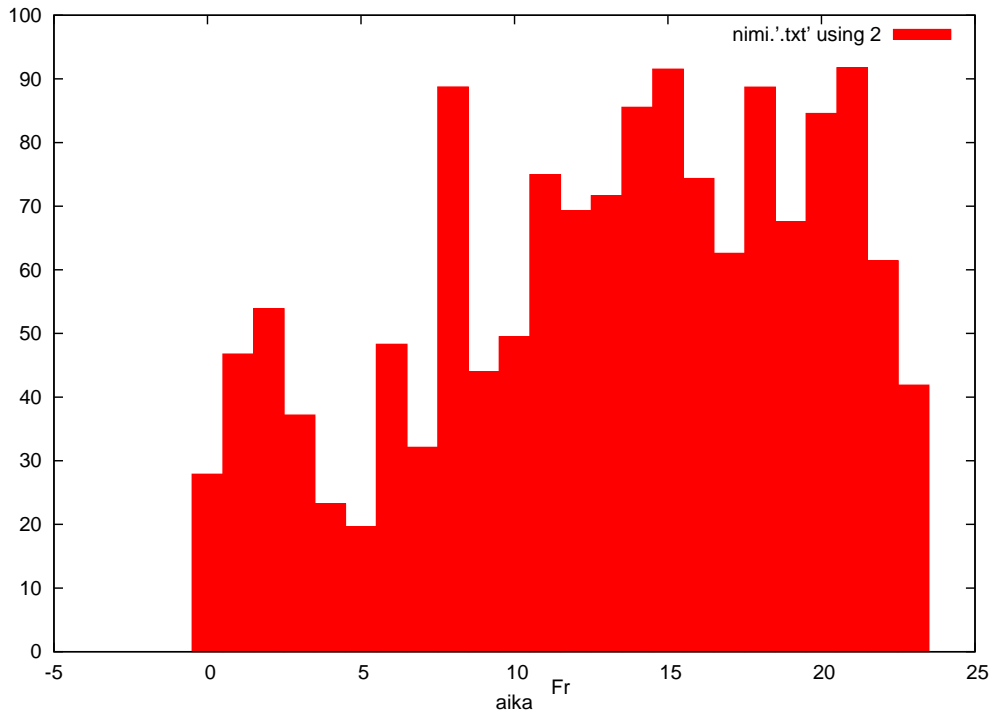
Kuva 4: Määrät: Mansikki ja Mustikki



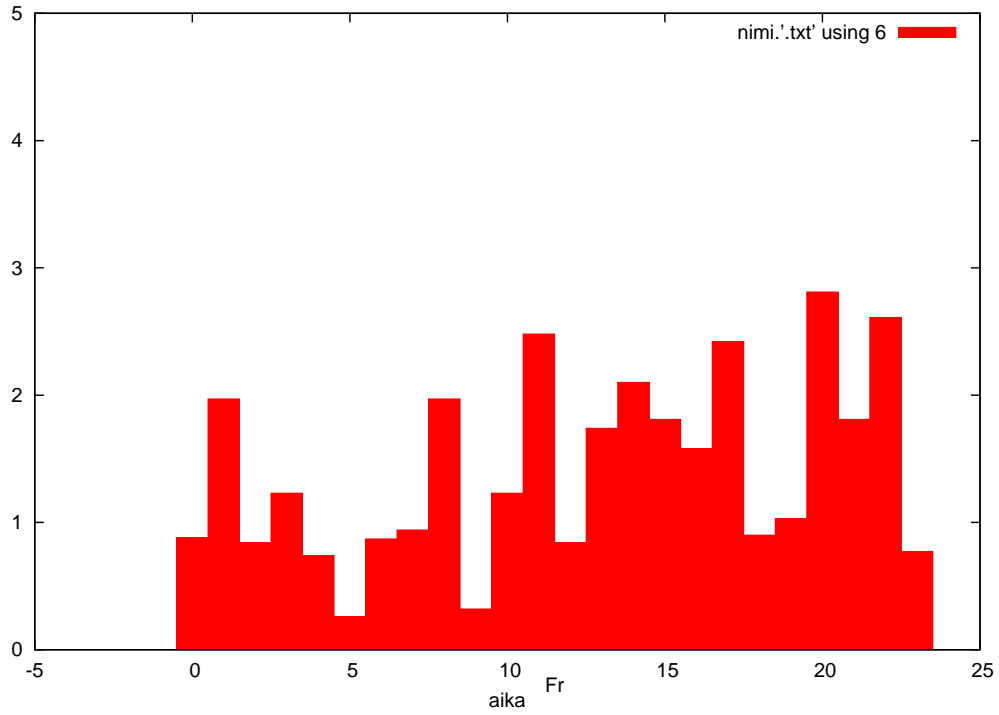
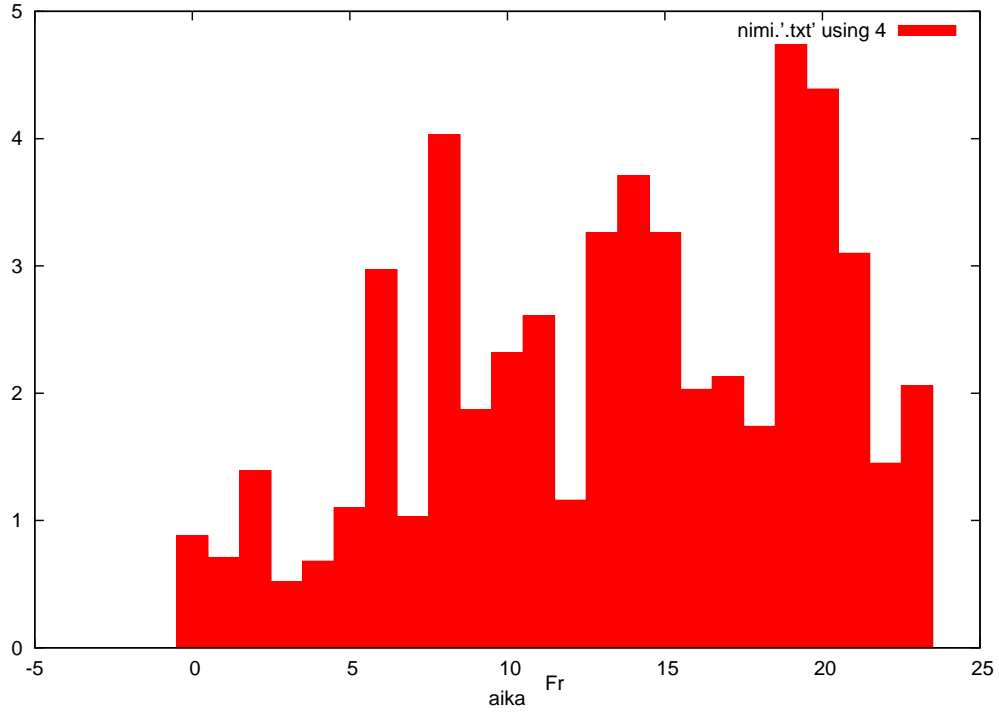
Kuva 5: Määrät: Mansikki ja Mustikki



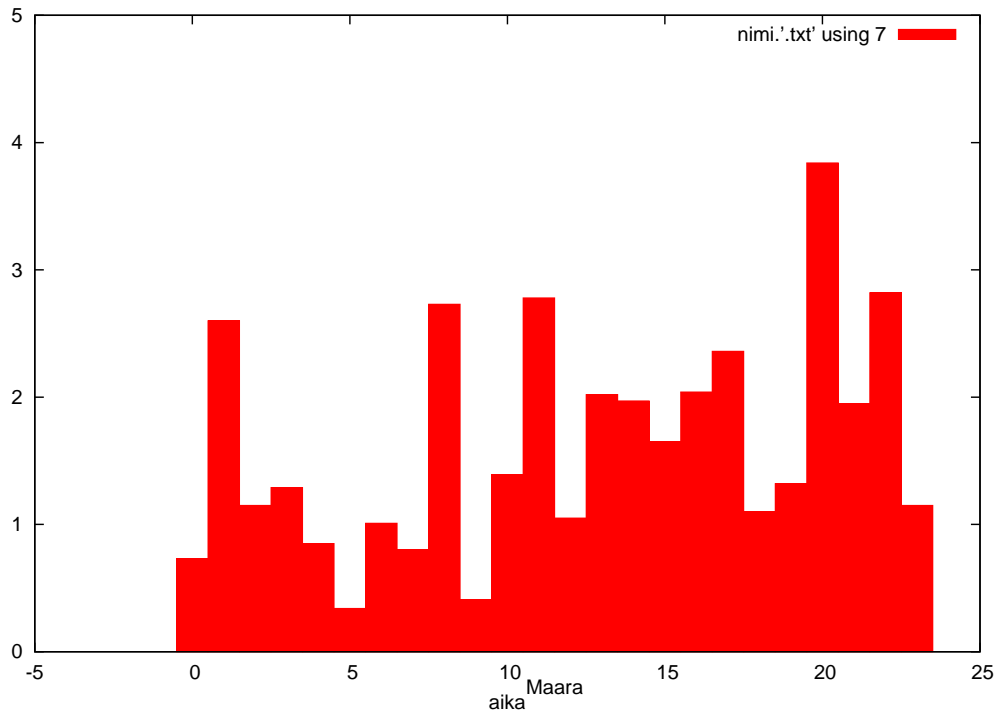
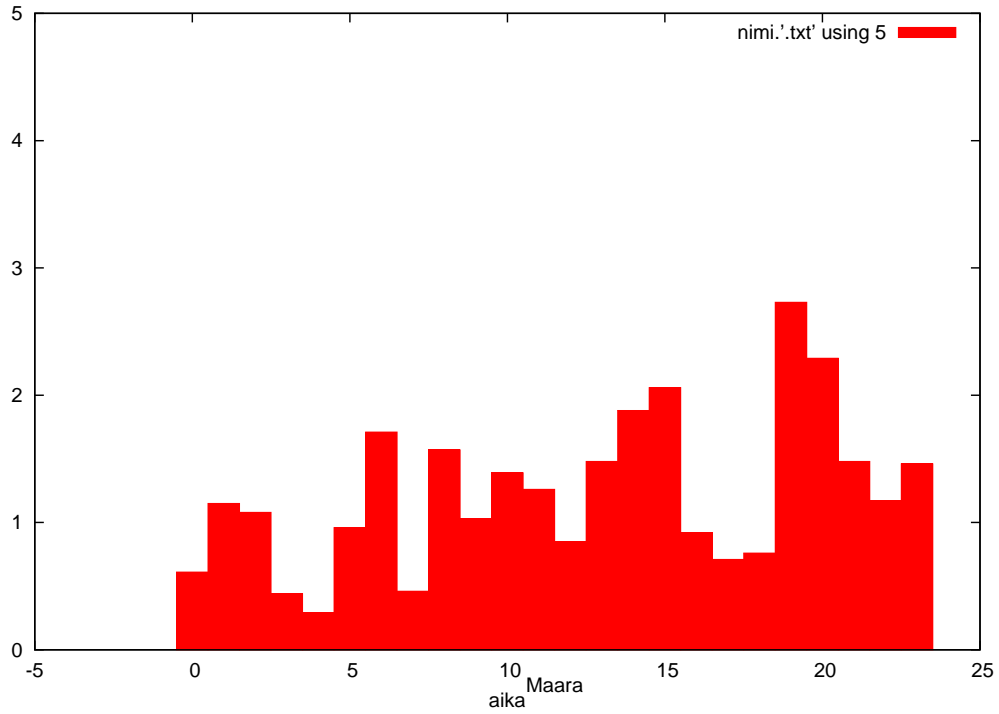
Kuva 6: Koko navetta



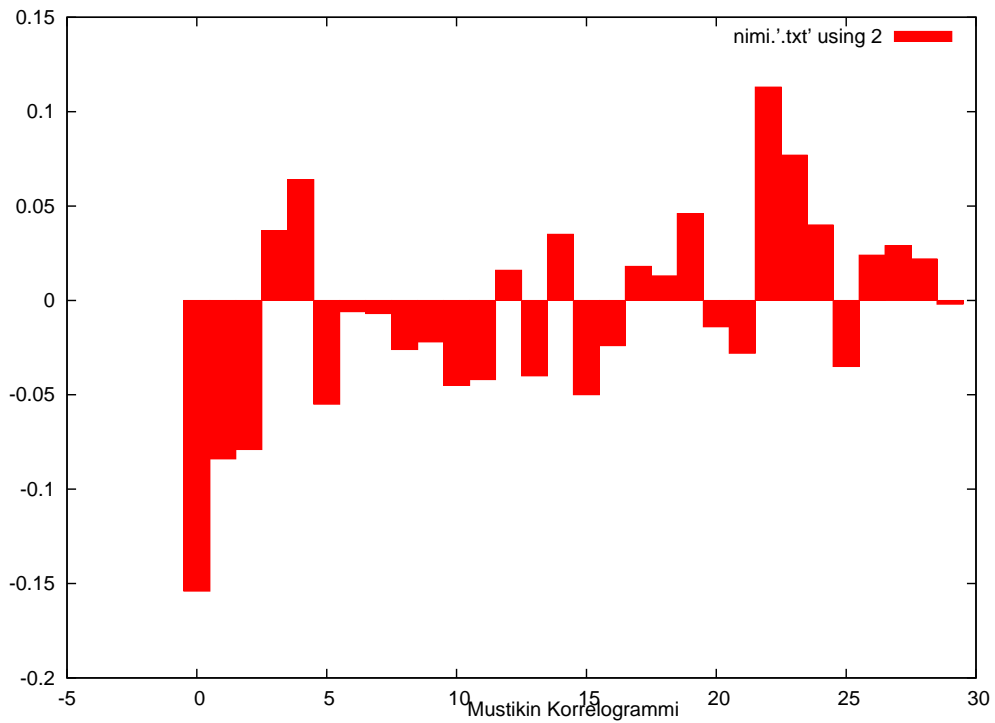
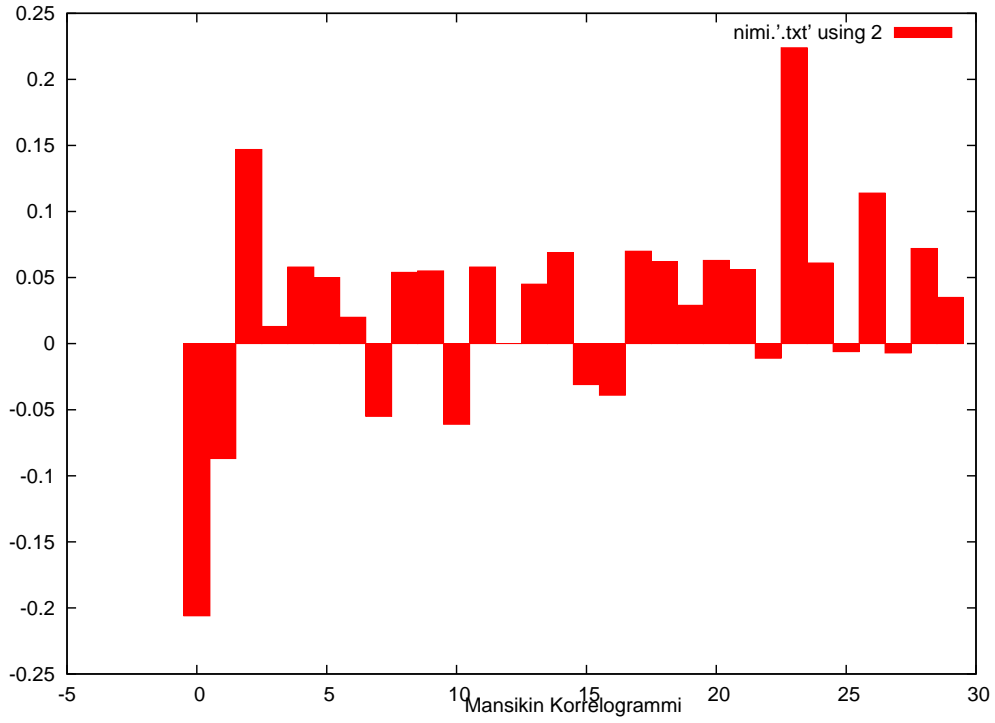
Kuva 7: Frekvenssit: Mansikki ja Mustikki



Kuva 8: Määrät: Mansikki ja Mustikki



Kuva 9: Määrien korrelogrammit: Mansikki ja Mustikki



Kuva 10: Frekvenssien korrelogrammit: Mansikki ja Mustikki

