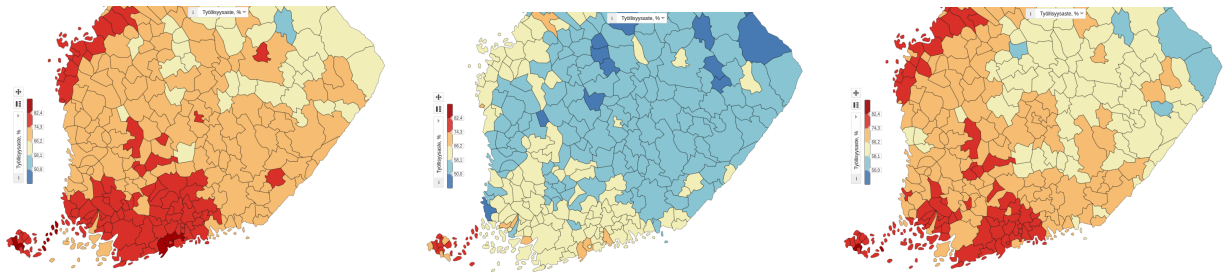
Project

Teemu Saravirta - 716051

4/11/2021

Introduction

Finland had not fully recovered on a employment level basis from the 1990s recession by the time 2008 economic crisis hit. Like other countries, Finland experienced a sharp downturn but not a mass unemployment like seen in 1990-1994. Contrasting with many countries though, Finland only experienced about a year of increasing employment, before actually falling again in the years up to 2014. In 2018, Finnish employment had finally recovered to almost equivalent level to 1989, but the key difference, the jobs were increasingly concentrated. In 1989 employment in Southern Finland blanketed the whole region, but in 2018 it was highly concentrated the municipalities *surrounding* the main cities of Finland. As seen in Figure 1, all the municipalities surrounding Helsinki have a high employment level, but Helsinki itself is lagging behind. The same happens also in Turku and Tampere, as well as in other large cities such as Jyväskylä, Lappeenranta and Vaasa.



This divergence is the most apparent in recent years, so I will focus on the available data from 2015 onwards - the period starting from the end of the last downturn. Data does not cover covid-19 period yet, so the whole period is characterized by some of the strongest growth Finland has seen in a long time. The question becomes: **What are the characteristics of these surrounding municipalities?** I have done extensive exploratory analysis on a tool available at the source of the data, where I explored changes for the whole Finland across the period of data available (1987-2018). From there I assumed to have found that there were certain variables that showed clear trends:

- percentage of people under 15 y.o.
- expenditure in social and health activities
- municipal debt per capita
- employment level

The last of which shows heuristically the clearest distinction between the major cities and their surrounding regions.

As all Finnish cities are heterogeneous to each other, I decided so scope this to just the largest cities, and I will pick the surrounding municipalities based on Figure 1, including all the municipalities in the 74.3-82.4% employment- bin. The large cities are Turku, Tampere, as well as Helsinki, Espoo and Vantaa grouped as *Capital cities*.

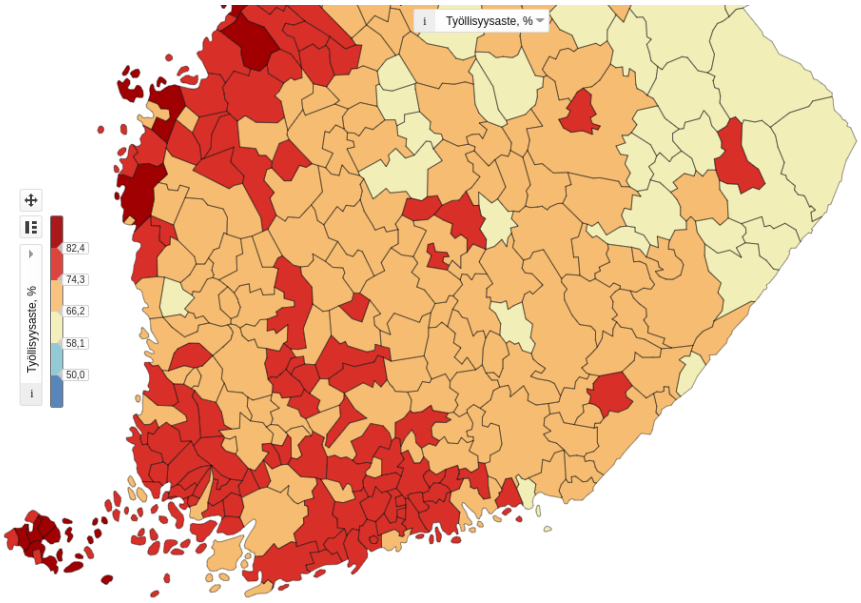


Figure 1: Employment in 2018

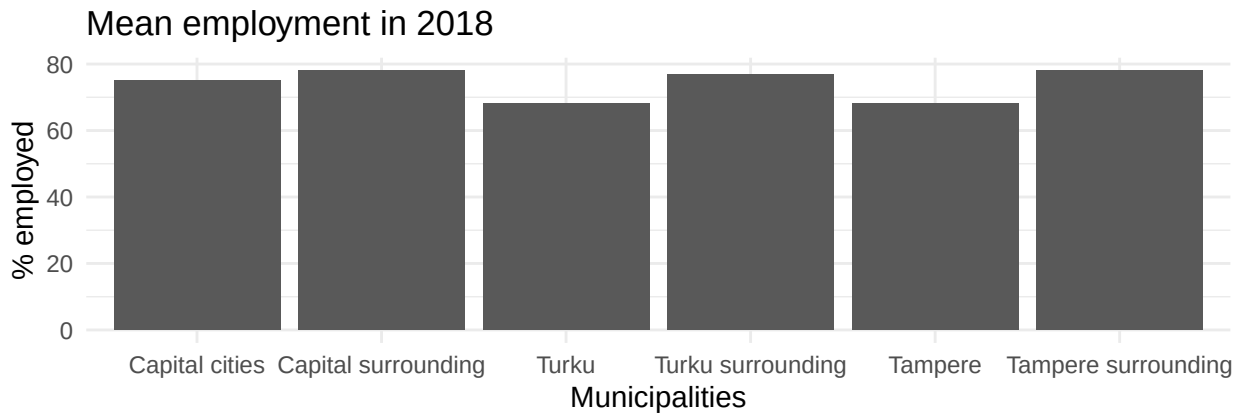
Data source

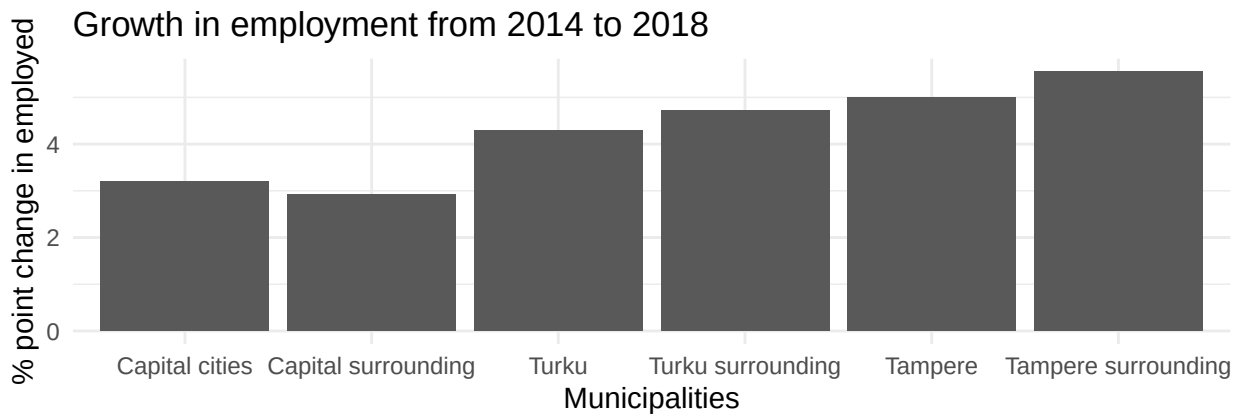
Source of data is Tilastokeskus (source). The exact link to select variables etc is *here*.

KPIs of Finnish municipalities over time from 1987 to 2019, as well as the aggregated KPIs from the whole of Finland. Municipalities in the data are based on the 2021 municipality division. Data is included from 397 municipalities + aggregated to cover Finland, and it has 421 variables across time covering 20 years from 1987-2018. I selected a subset of this data as mentioned above.

Univariate analysis

I start by seeing what are the differences between the chosen regions in their employment levels. I use % employed, which measures the % of 15-64 y.o. that are employed.





As we can see from the above, the surrounding municipalities have on average considerably higher employment levels. In Capital cities it was 3%, in Turku 9%, and in Tampere 10%.

Interestingly the change in employment around the capital cities is smaller than that of the capital. Difference between the cities and surrounding municipalities being:

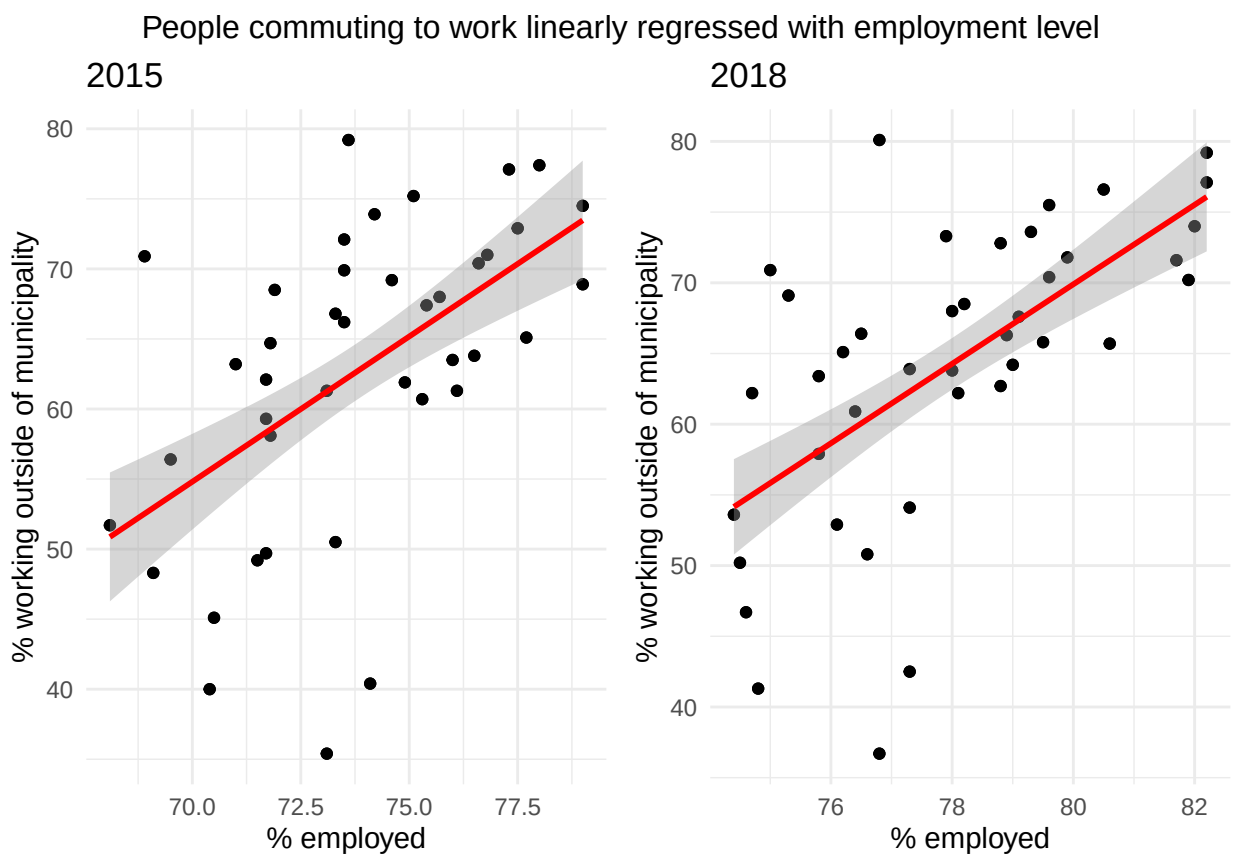
- Capital -0.28
- Turku 0.43
- Tampere 0.55

Bivariate analysis

Here I chose to see if I can find a linear dependency between *the % of people of the surrounding municipalities that are commuting outside of their home municipality to work* and *

```
##
## Call:
## lm(formula = risk2015 ~ out2015, data = lm_data.final)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.0536 -1.6209 -0.2983  1.8325  4.3252
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 65.06998    1.55528  41.838  < 2e-16 ***
## out2015      0.13940    0.02444   5.703 1.91e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.38 on 80 degrees of freedom
## Multiple R-squared:  0.289,  Adjusted R-squared:  0.2802
## F-statistic: 32.52 on 1 and 80 DF,  p-value: 1.906e-07
```



For both 2015 and 2018, with highly different economic situations, the results are statistically significant. However, looking at the graphs, there is significant amount of outliers, especially in the municipalities with a lower employment level. The municipalities with a high employment are tightly linked with the regression line, suggesting that the best doing municipalities are the ones that can make the most out of commuting to their neighboring city. However this is only correlation between the two variables, and such such no contains no causal implications.

Multivariate analysis

I selected to use corrgrams and PCA to find possible correlations, as the data was quantitative, and of high dimension. The initial selection by the author was done based on which variables could:

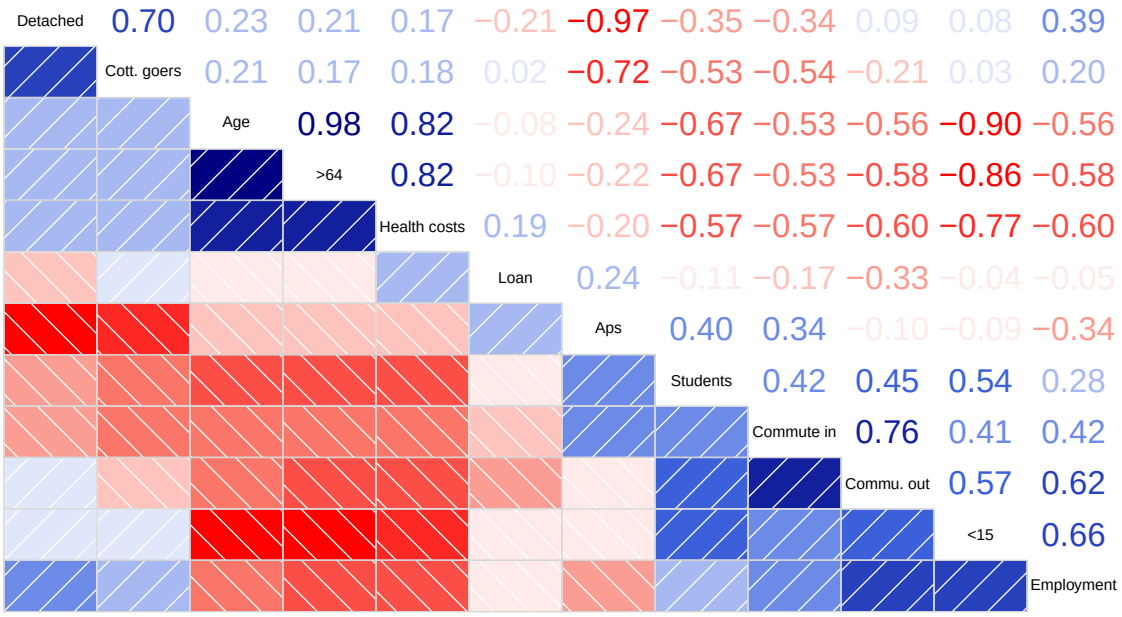
1. Show variations between the surrounding municipalities, or between them and cities
2. Display key structural trends

Explanations for the variables:

- Loan: per capita dept held by the municipality and companies owned by it
- Employment: % of 15-64 y.o. employed
- Health costs: Expenditure in social and health activities
- Detatched: % of Household-Dwelling Units (HDU, consist of occupants of a dwelling) living in single family houses, two family houses, and two story houses with two HDUs living in different floors.
- Aps: % of HDUs living in apartment buildings
- Cott. goers: percentage of cottages owned by out-of-municipality persons
- <15: % of residents being under 15
- >64: % of residents being 65 or higher
- Age: average age of residents
- Students: % of residents 15 or above being students
- Commu. out: % of workers living in the municipality commuting out
- Commute in: % of workers working in the municipality being from other municipalities

---

Exploring just the surrounding municipalities



Let's start by looking at employment levels. Interestingly it, and other variables, correlate similarly with both commuters coming in and commuters coming out of these surrounding municipalities. Furthermore,

these correlate strongly with each other, so with better connectivity both the employees coming in and going out increase, in conjunction with employment levels.

The largest positive correlation with employment levels is the % of people under 15 in the municipality. Likewise positive correlations are found with the % of people living in detached homes, and a smaller one with the % of population being students.

One possible explanation is that people with relatively stable jobs have high confidence in future, and such the other changes stem from family increases. Important distinction is that neither the % of people living in detached houses nor apartment buildings correlate with % people under 15, but they have opposing mid-sized correlations with employment levels. So the well-to-do municipalities seem to be the ones where the people have the confidence in future, and can afford to live in a detached home.

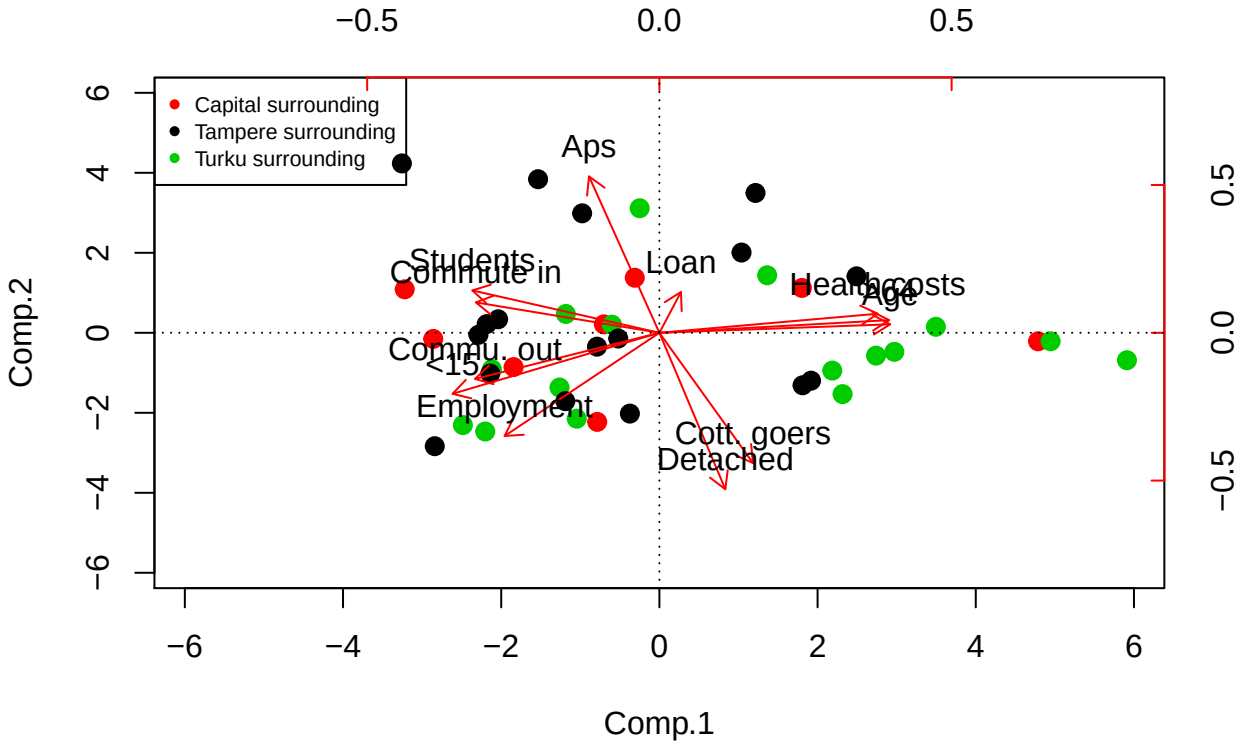
Interestingly smaller but flipped effects for the municipalities are seen the value of debt per capita the municipality and all the companies it has ownership of (Loans variable). The average amount seems to go up with detached homes and down with apartment buildings. So while the detached homes are associated with better-off residents, they also correlate with higher debt levels in the municipality. This intuitively makes sense as the infrastructure costs imposed on municipalities grow as people live further apart.

Lastly to mention, out-of-municipality cottage goers is associated with higher employment level and higher debt per capita. And it seems like these cottage goers do not like students, apartment buildings, or people commuting to work in the municipality while they are going there to the cottage.

PCA with surrounding municipalities

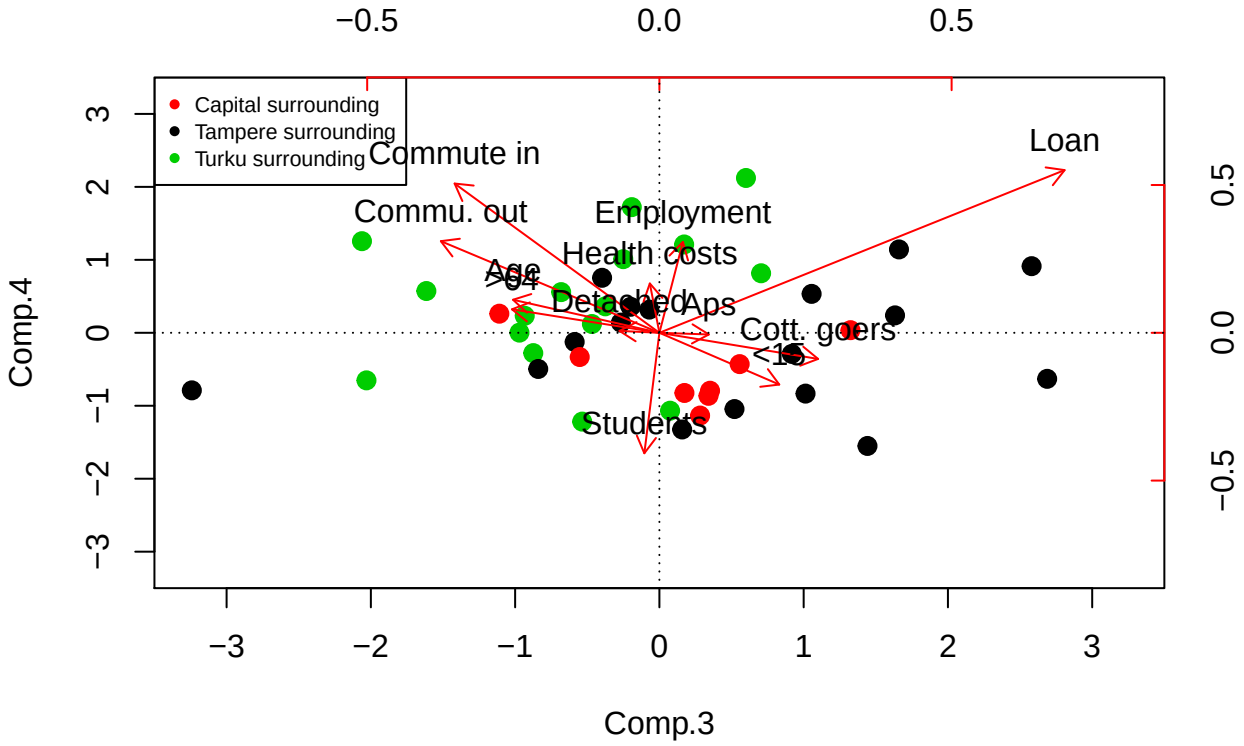
For the surrounding municipalities, PCA 1 and 2 explain together 71.7% of variance. With 4 we achieve 89.2%. From below you can see which components best explain the variation in the selected variables (larger absolute values = explain well).

```
##              Comp.1 Comp.2 Comp.3 Comp.4
## Loan           0.04   0.14   0.69   0.55
## Employment    -0.26  -0.35   0.04   0.31
## Health costs   0.37   0.06  -0.02   0.17
## Detached       0.11  -0.53  -0.07   0.01
## Aps           -0.12   0.53   0.08  -0.01
## Cott. goers    0.16  -0.44   0.27  -0.09
## <15           -0.35  -0.21   0.21  -0.18
## >64            0.39   0.04  -0.25   0.08
## Age            0.39   0.03  -0.25   0.11
## Students      -0.32   0.14  -0.03  -0.41
## Commu. out    -0.32  -0.16  -0.37   0.31
## Commute in    -0.31   0.10  -0.35   0.50
```



1st and 2nd principal component

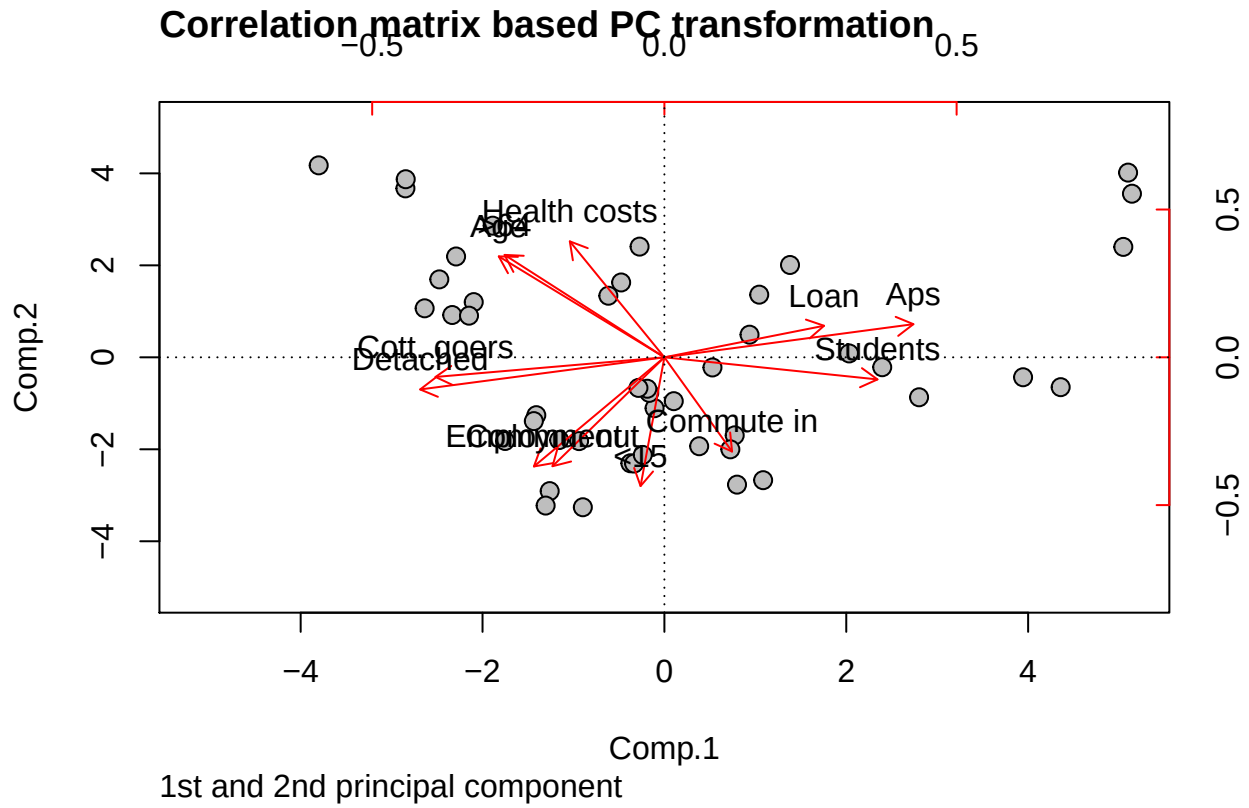
Based on how the first component captures variables related to age, I interpret Comp 1 to be something akin to "relative structural healthiness" of the municipality, focusing on dependency ratio. Component 2 then captures more the physical make-up of the municipality, with high emphasis on housing type. Most of the correlations were already visible from corrgram, but we can additionally see from here that Turku region seems to be correlating with higher health costs. Verifying it, the avg over the surrounding areas for Turku, Tampere, and Capital are in order: 3.2k, 3.0k, 2.9k.



3rd and 4th principal component

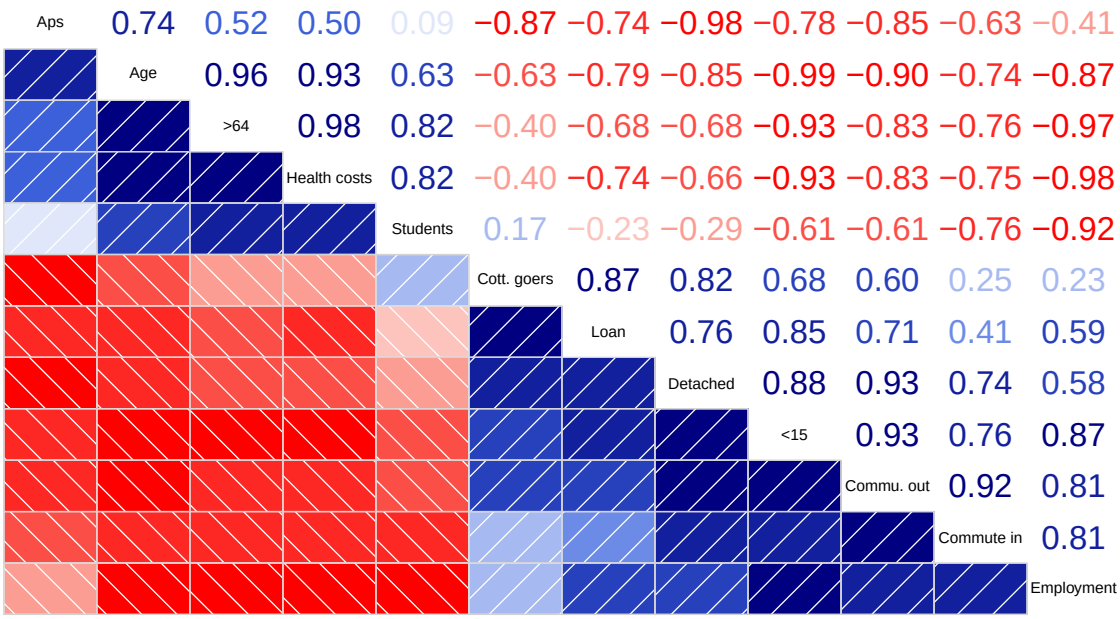
Interestingly, as municipal debt was not in the corrgram a strongly correlating variable, the PCA 1 and 2 did not capture it very well, but it is the main variance captured in both 3 and 4. I wouldn't say there are any suprises here, maybe that figuring out the debt level on a municipality is not as straight-forward as one would assume.

PCA with cities and surrounding municipalities



Now with cities factored in, the correlation structure looks very different. Loans now strongly correlate with students and % of residents living in apartments, and likewise the opposite with arriving cottage goers and detached homes. I now assume the Comp 1 to be a general measure for cities, separating them from the surrounding municipalities. Now I take the Comp 2 to be relative structural health of a municipality. As component 1 is almost impartial to % of residents under 15 y.o., I take that there are both cities and surrounding municipalities with youthful populations. Then the Comp 2 distinguishes these municipalities based on the health expenditure, youthfulness, and connectivity. But this is again to be taken with a grain of salt, as now commute in and health costs are some of the strongest negatively correlating variables, something that was not seen with just he surrounding municipalities. This suggest that Comp 2 also has a good deal of *'cityness'* as a factor. But on the other hand, with more people commuting in, the municipality is 'more structurally healthy' as it brings money flows to the municipality. Health costs are quite the opposite, as they might be more linked to the place of residence than that of workplace.

Correlation for just the cities



As we only have three data points, PCA isn't applicable to measuring just the cities, and correlation structure is much more pronounced that with more data points. Corrgram should also be taken with a grain of salt, as n=3.

But to highlight one key finding, debt level is much more strongly correlated (in relation to the correlation between other variables) with health costs than in surrounding municipalities where it was almost neutral.

But we can also see from the corrgram that amount of students is correlating strongly with the health costs of the city, suggesting that we are now capturing more the relative size of cities than that of differences between the variables themselves.

Discussion

The suggestions reached are biased on that they are based on the author's assumptions about what makes a 'structurally healthy municipality', as well as other interpretations, such as how stable jobs lead to confidence in future leading to family increases. These assumptions and interpretations are based on the limited knowledge of the author, about a highly complex system of society. If there was wording that can be interpreted as suggesting that the correlation studied is causation, that was not intended.

PCA does not lend itself well to few data points over a multitude of dimensions, so comparing the cities themselves is hard. This was circumvented by doing comparative assessment of just the surrounding municipalities, surrounding municipalities and cities, and just the cities. Resulting is large change in the PCA results. Much of the analysis of the municipalities PCA is valid to the combined one, and the resulting interpretation of relative structural health and connectivity seem to be satisfactory. Cities however tend to be surrounded by more well-off/ less well-off municipalities, which would warrant more study into what exactly is making certain municipalities less well-off.