PCA on Finnish Municipalities

Dimensionality reduction with PCA on Finnish municipality demographics

Niko Miller

07.04.2022

Contents

1 Introduction

1.1 Motivation

Tilastokeskus provides data on 32 demographic variables of Finnish municipalities [Statistics Finland, 2022]. Those variables are in many cases correlated and some are perhaps better in explaining differences between municipalities. Hence, an interesting question is whether we could reduce dimensionality of the data and describe municipalities using only a few variables.

1.2 Objective and research question

1.3 Scope

2 Data and Exploratory Analysis

I used PXWEB API [Magnusson et al., 2019] to retrieve a data set on key ratios for all of Finland's municipalities during 1987-2019 according to the 2020 municipality classification. The data set is provided by Tilastokeskus [Statistics Finland, 2022].

2.1 Cleaning and Manipulation

2.2 Univariate Analysis

2.3 Bivariate Analysis

3 Principal Component Analysis (PCA)

3.1 Principal Components 1 and 2

4 Discussion and Conclusions

References

Mans Magnusson, Markus Kainu, Janne Huovari, and Leo Lahti. pxweb: R tools for px-web api, 2019.

Statistics Finland. Kuntien avainluvut muuttujina alue 2020, tiedot ja vuosi, 2022. URL https://pxwebapi2.stat.fi/PXWeb/api/v1/fi/Kuntien_avainluvut/2020/kuntien_avainluvut_2020_aikasarja.px. [Data accessed 2022-04-06 16:59:07 using pxweb R package 0.13.1].