

# PCA on Finnish Municipalities

Dimensionality reduction with PCA on Finnish municipality demographics

Niko Miller

08.04.2022

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Objective and research question . . . . .	1
1.3	Scope . . . . .	1
<b>2</b>	<b>Data and Exploratory Analysis</b>	<b>1</b>
2.1	Cleaning and Manipulation . . . . .	1
2.2	Variables and descriptive statistics . . . . .	1
2.3	Univariate Analysis . . . . .	2
2.4	Bivariate Analysis . . . . .	2
<b>3</b>	<b>Principal Component Analysis (PCA)</b>	<b>2</b>
3.1	Principal Components 1 and 2 . . . . .	2
<b>4</b>	<b>Discussion and Conclusions</b>	<b>2</b>

# 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 Variables and descriptive statistics

```
vars <- c('Prop.Urban.Areas', 'Pop', 'Pop.Change', 'Prop.Below15', 'Prop.15to64', 'Prop.Over64', 'Prop.Swedish', 'Prop.Foreign', 'Pop.Growth', 'Migr.Gain', 'Families', 'Households', 'Prop.Households.rowSmall', 'Prop.Households.Rent', 'Prop.Educ.Degree2', 'Prop.Educ.High', 'Employed', 'Empl.Rate', 'Prop.Households.Empl', 'Prop.Unempl', 'Prop.Pension', 'Support.Ratio', 'Jobs')
translations <- rep("NA", length(vars))
var.df <- data.frame(vars, translations)

knitr::kable(var.df, col.names = c("Variable", "Explanation"))
```

Variable	Explanation
Prop.Urban.Areas	NA
Pop	NA
Pop.Change	NA
Prop.Below15	NA
Prop.15to64	NA
Prop.Over64	NA
Prop.Swedish	NA
Prop.Foreign	NA
Pop.Growth	NA
Migr.Gain	NA
Families	NA
Households	NA
Prop.Households.rowSmall	NA
Prop.Households.Rent	NA
Prop.Educ.Degree2	NA
Prop.Educ.High	NA
Employed	NA
Empl.Rate	NA
Prop.Households.Empl	NA
Prop.Unempl	NA
Prop.Pension	NA
Support.Ratio	NA
Jobs	NA

Variable	Explanation
Prop.Primary.Sector	NA
Prop.Secondary.Sector	NA
Prop.Services.Sector	NA
Jobs.Self.Suff	NA
Margin.Citizen	NA
Loan.Citizen	NA
Concern.Loan.Citizen	NA
Educ.Cult.Citizen	NA
Soc.Health.Citizen	NA

## 2.3 Univariate Analysis

## 2.4 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](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].