

# **Relation Between Economic Conditions and Voting Results in Finland**

Ali Onur Ozkan

25/09/2024

## **Contents:**

- 1) Introduction
- 2) Description of Data
- 3) Data Handling
- 4) Results
  - 4.1) Correlation Analysis
  - 4.2) Income vs. Voting
  - 4.3) Linear Regression Analysis:
  - 4.4) Residual Analysis
  - 4.5) Hypothetical Prediction
- 5) Conclusion

## **1) Introduction**

Knowing the link between economic indicators and voting results is important to understand the citizens' concerns and party preferences in Finland. Based on this observation, this report aims to identify specific income indicators that determine voter behavior, anchoring the research on the 2023 elections, based on income statistics from 2017. Therefore, the project tries to identify any specific income indicators that might impact voter behavior and, thus, shed light on the economic profiles of supporters for various parties. Looking at data sets from the recent 2023 elections with income statistics from 2017 will be beneficial to understand these dynamics.

## **2) Description of Data**

Two data sets are used in this project:

**Election Results:** This dataset includes the share of votes in each municipality received by each of the 23 parties.

**Income Statistics:** For every municipality, this dataset includes average income, median income, earned income, capital income, total taxes, state tax, municipal tax, and income after tax.

### 3) Data Handling

First of all, it was necessary to be very careful while handling the data and make sure the data read and processed correctly, with the special characters in the Finnish language a proper encoding of files had to be used: Windows-1252. Translation of column headings from Finnish to English was done. The datasets were read; then both were combined into one by matching the municipalities for ease of use. Municipalities without election data were excluded.

Part of the Code Used:

```
ek2023 <- read.csv("ek2023.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE, fileEncoding = "Windows-1252")
```

```
tulot2017 <- read.csv("tulot2017.csv", header = TRUE, sep = ",", stringsAsFactors = FALSE, fileEncoding = "Windows-1252")
```

```
colnames(ek2023) <- trimws(tolower(colnames(ek2023)))
```

```
colnames(tulot2017) <- trimws(tolower(colnames(tulot2017)))
```

```
ek2023$value <- trimws(tolower(ek2023$value))
```

```
tulot2017$value <- trimws(tolower(tulot2017$value))
```

```
merged_data <- merge(ek2023, tulot2017, by = "alue")
```

### 4) Results

Given the above facts, two major parties were observed in our data to be quite consistently correlated with many of the economic indicators: KESK, or Centre Party; and VIHR, or Green League. Analysis of those economic factors allows to develop a deeper understanding of the socioeconomic profiles of the voters for those parties and how the economic standing of a voter influences their voting behavior for these parties.

**4.1) Correlation Analysis:** The patterns among different economic indicators and party support are studied.

KESK: The more the income, the less the support for KESK; it would, therefore, imply a negative relationship between income and support for KESK.

VIHR: The increase in average income is followed by an increase in the support to VIHR. That means an increase in probability when income increases-which indicates more likelihood of being a VIHR voter.

Part of the Code Used:

```
numeric_columns <- c("tulot", "mediaanitulos", "ansiotulos", "sdp", "ps", "kok", "kesk", "vihr", "vas")
```

```
merged_data[numeric_columns] <- lapply(merged_data[numeric_columns], as.numeric)
```

```
correlation_matrix <- cor(merged_data[, numeric_columns], use = "complete.obs")
```

```
print(correlation_matrix)
```

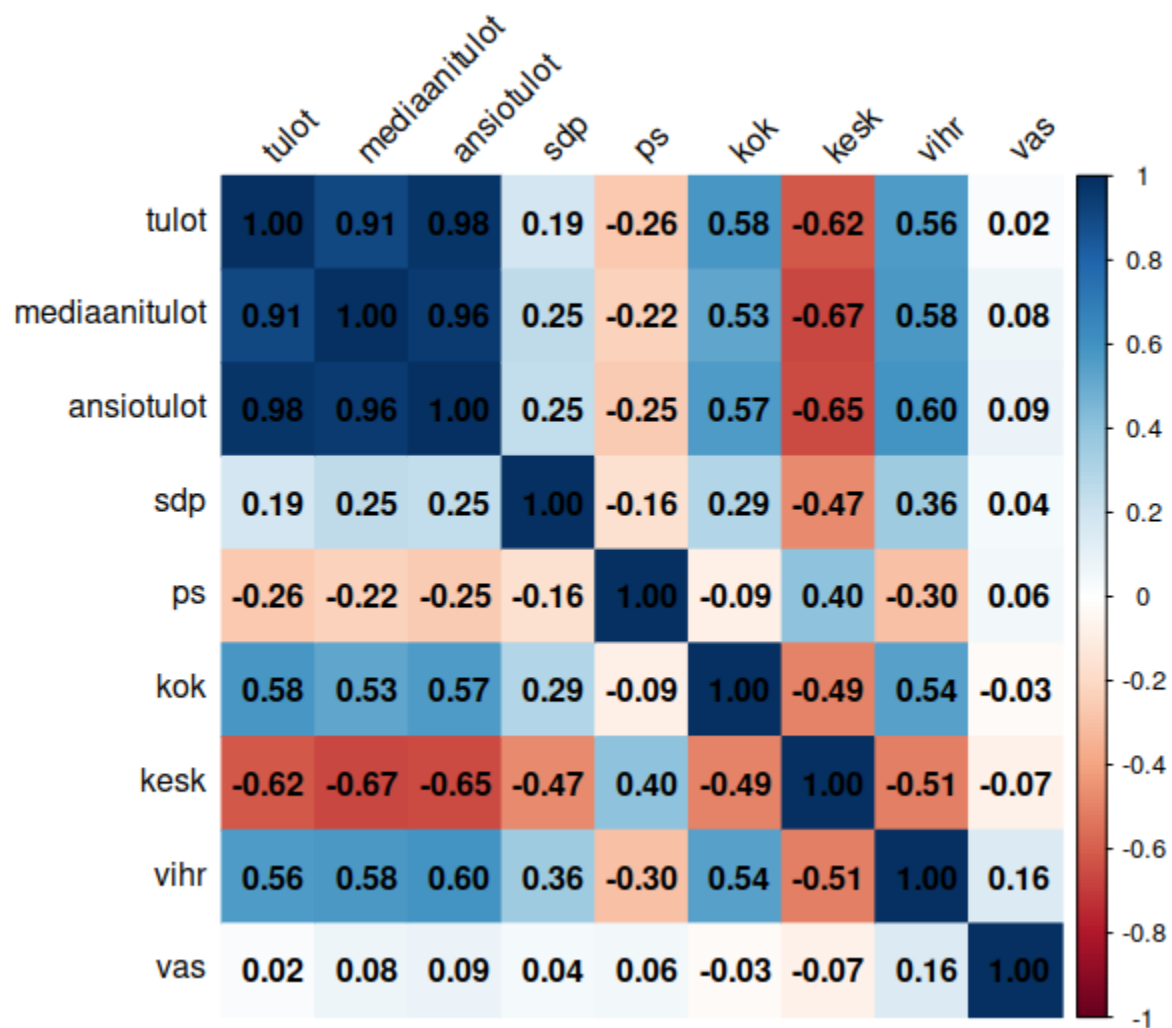
```
corrplot::corrplot(correlation_matrix, method = "color", addCoef.col = "black", tl.col = "black", tl.srt = 45)
```

The correlation plot:

Average Income (‘tulot’)

Median Income (‘mediaanitulos’)

Earned Income (‘ansiotulos’)



**4.2) Income vs. Voting:** The scatter plots are plotted to see the relationship between average income and support for different parties, focusing on KESK and VIHR.

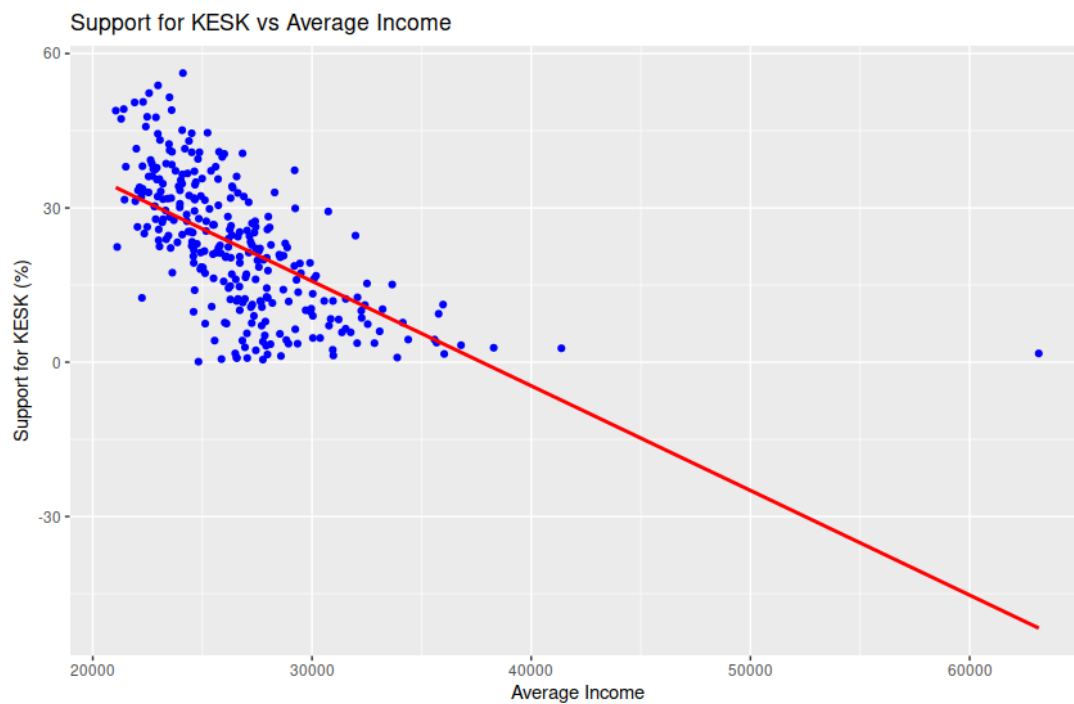
The output on the plots supports the claim made in the correlation analysis. While the support for KESK is in a negative relation with the average income, the support for VIHR is in a positive relation with the average income. Additionally, in both plots, the municipality 'Kaunainen' (the dot on the rightmost side) is a clear outlier due to its extremely high-income value.

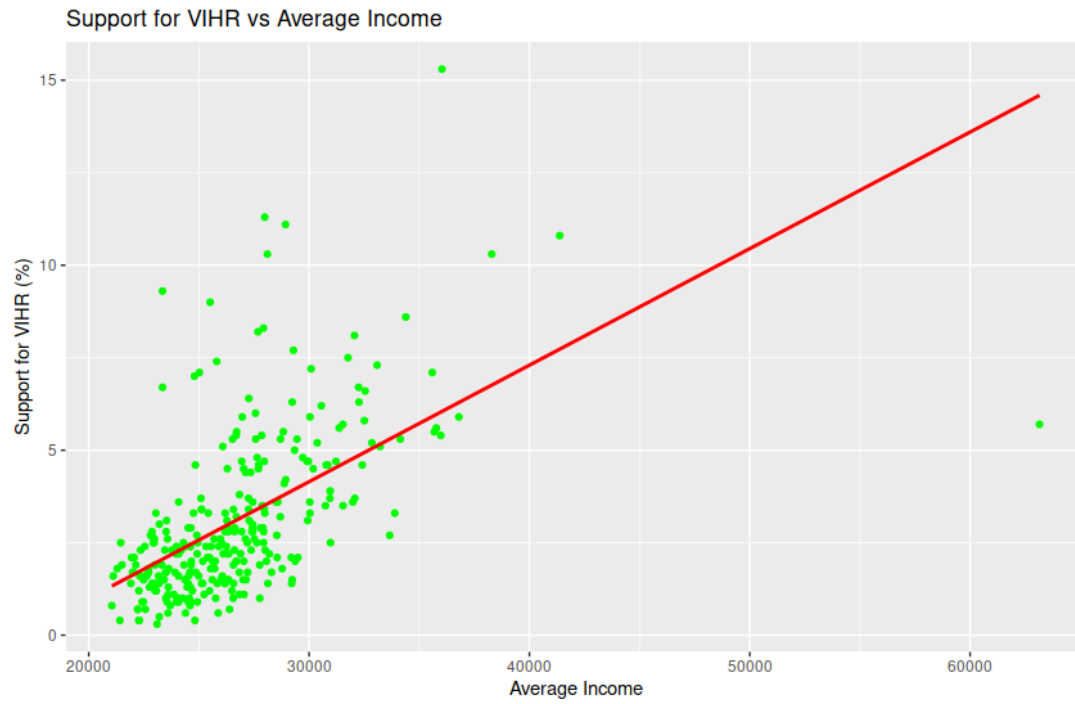
Part of the Code Used:

```
ggplot(merged_data, aes(x = tulot, y = kesk)) +  
  geom_point(color = "blue") +  
  geom_smooth(method = "lm", se = FALSE, color = "red") +  
  labs(title = "Support for KESK vs Average Income", x = "Average Income", y = "Support for KESK (%)")
```

```
ggplot(merged_data, aes(x = tulot, y = vihr)) +  
  geom_point(colour = "green") +  
  geom_smooth(method = "lm", se = FALSE, colour = "red") +  
  labs(title = "Support for VIHR vs Average Income", x = "Average Income", y = "Support for VIHR (%)")
```

Scatter plots for KESK and VIHR:





**4.3) Linear Regression Analysis:** Linear regression analysis was performed to measure the relationship between economic indicators and party support for KESK and VIHR.

Results have been suggesting that, out of all the political parties, only KESK and VIHR showed a strong correlation with economic indicators. Economic factors explain the profiles of the voters. However, all by itself, it is not able to explain the behavior of voting. Social policy also helps in shaping the decisions of the voters.

Part of the Code Used:

```
model_kesk <- lm(kesk ~ tulot + mediaanitulut + ansiotulot, data = merged_data)
summary(model_kesk)
```

```
model_vihr <- lm(vihr ~ tulot + mediaanitulut + ansiotulot, data = merged_data)
summary(model_vihr)
```

Summary Statistics of the Regression Models for KESK and VIHR:

**KESK Regression Results:**

Intercept: The baseline of support of KESK is 85.99%.

Average Income (`^tulot`): A very minimal positive but insignificant effect on KESK support ( $p = 0.41349$ ). So, to predict the support for KESK, the average income isn't a strong data.

Median Income (`^mediaanitulut`): Significant negative effect on KESK support where  $p = 0.00712$ , which suggests that municipalities with a higher median income tend to support KESK less.

Earned Income (`^ansiotulot`): is not significant. ( $p = 0.23166$ ).

R-square: It can be inferred that 44.8% of the variance in KESK support can be explained by these economic indicators, because the multiple R-squared is 0.4484.

#### VIHR Regression Results:

Intercept: The base level of support for VIHR is -5.91%, showing a baseline disadvantage in municipalities with the lowest incomes.

Average Income ( $\hat{\text{tulot}}$ ): Significant negative effect on VIHR support ( $p = 0.030592$ ), suggesting that higher income reduces support for VIHR.

Earned Income ( $\hat{\text{ansiotulot}}$ ): This had a significant positive effect ( $p = 0.000759$ ), meaning people having higher earned income are more likely to support VIHR.

R-square: It can be inferred that 36.8% of the variance in VIHR support can be explained by these economic indicators, because the multiple R-squared is 0.3681.

#### Interpretation of Residuals:

Residual distribution gives some insight into the model fit. Residual statistics were presented for both KESK and VIHR such as Min, 1Q, Median, etc., which show that although the model explains a good part of variance, there is still some deviation from the model forecast.

### 4.4) Residual Analysis

To check the fit of the models, residuals were examined:

#### KESK Residuals:

Min: -28.577

1Q: -6.171

Median: -0.331

3Q: 5.958

Max: 28.633

The great range in residuals indicates that the model captures some of the variation but also misses much of the nuance in voter behavior.

#### VIHR Residuals:

Min: -5.771

1Q: -1.085

Median: -0.324

3Q: 0.594

Max: 9.440

The residuals for VIHR are closer, providing a better fit to the model, but there are still outliers, including residuals close to 9.44.

**4.5) Hypothetical Prediction:** The support for KESK and VIHR is predicted using hypothetical income data to highlight how income influences party support.

Part of the Code Used:

```
hypothetical_data <- data.frame(tulot = 30000, mediaanitulot = 25000, ansiotulot = 28000)
predicted_kesk <- predict(model_kesk, newdata = hypothetical_data)
cat("Predicted support for KESK: ", predicted_kesk, "%
")
predicted_vihr <- predict(model_vihr, newdata = hypothetical_data)
cat("Predicted support for VIHR: ", predicted_vihr, "%
")
```

The predicted support percentages for KESK and VIHR based on hypothetical data, showing that KESK having stronger support than VIHR under the circumstances below:

If (Average Income) = 30000, (Median Income) = 25000, (Average Earned Income) = 28000

Predicted support for KESK: 14.50401 %

Predicted support for VIHR: 4.467677 %

## 5) Conclusion

It is inferred from the findings is that only a few parties crossed the filters placed on them. Of 23 parties, only two showed a steady and strong correlation between the economic indicators and their voting percentage. The rest of the parties did not have a strong likelihood of depending on economic policies to change the tide of voter preference in the election.

While the economic indicators give some insight into the economic profiles of the party supporters, they are not sufficient for the explanation of the voting behavior on their own; in addition, social policy is also an important factor and may shape the decisions of the voters.