

# R-Bootcamp Report

Leonid Gavriluk\*

Bernardo Freire Barboza da Cruz<sup>†</sup>

21 February, 2022

---

\*Hochschule Luzern, leonid.gavrilyuk@stud.hslu.ch

<sup>†</sup>Hochschule Luzern, bernardo.dacruz@stud.hslu.ch

# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction.</b>  | <b>3</b>  |
| 1.1      | Project goal . . . . .  | 3         |
| 1.2      | Data . . . . .  | 3         |
| <b>2</b> | <b>Data Preparation</b>   | <b>4</b>  |
| 2.1      | Data set 1: Turnout at the city and municipal council elections since 2006, by city district. . | 4         |
| 2.2      | Data set 2: Municipal elections vote share, by party and electoral district since 1913. . . . . | 5         |
| 2.3      | Data set 3: Wealth distribution of the population in Zürich, by district . . . . .              | 6         |
| 2.4      | Data set 4: Income distribution of the population in Zürich, by district . . . . .              | 8         |
| 2.5      | Combine wealth and income data . . . . .  | 10        |
| 2.6      | Education data . . . . .  | 11        |
| <b>3</b> | <b>Merging Data</b>   | <b>14</b> |
| <b>4</b> | <b>Data Visualization</b>   | <b>15</b> |
| <b>5</b> | <b>Fit Model</b>  | <b>16</b> |
| <b>6</b> | <b>Chapter of Choice TBD</b>  | <b>17</b> |
| <b>7</b> | <b>References</b>   | <b>18</b> |

# 1 Introduction.

## 1.1 Project goal

The goal is to understand the dynamics in the performance of each part in each electoral district of the city. The time span in interest is **2006-2018** ??? The questions to be answered: the portrait of the voters, identification of relationship (if any) between such characteristics as voters' education or age and the election results of the party.

## 1.2 Data

To perform analysis, six data sets covering the time period 2006-2022 were retrieved from the Open Data platform of the City of Zürich:

- Turnout at the city and municipal council elections, by city district, since 2006. (*Beteiligung am Urnengang der Stadt- und Gemeinderatswahlen nach Stadtquartier*)
- Municipal elections vote share, by party and electoral district, since 1913. (*Gemeinderatswahlen Stimmenanteil nach Partei und Wahlkreis*)
- Turnout at the city and municipal council elections, by age and gender, since 2006. ?? ??

The data sets did not have the same content and were not organised in the same way. Each data set contained some irrelevant information - for example, historical election data since 1913 whereas the time span of interest is 2006-2018. This required data preparation activities. Each data set was prepared, subsequently all of them were merged into one final table.

## 2 Data Preparation

### 2.1 Data set 1: Turnout at the city and municipal council elections since 2006, by city district.

The data set of dimensions 136x7 reflects how many people from each city district participated in five last elections (2006-2022).

Table 1: Original data set: Turnout in the city and municipal council elections since 2006, by city district

| Jahr | QNr | Qname                 | Sberechtigigte | Nteilnehmende | teilnehmende | Beteiligung |
|------|-----|-----------------------|----------------|---------------|--------------|-------------|
| 2006 | 11  | Rathaus (Kreis 1)     | 1974           | 1186          | 788          | 39.9        |
| 2006 | 12  | Hochschulen (Kreis 1) | 377            | 232           | 145          | 38.5        |
| 2006 | 13  | Lindenhof (Kreis 1)   | 1335           | 962           | 373          | 27.9        |
| 2006 | 14  | City (Kreis 1)        | 597            | 440           | 157          | 26.3        |
| 2006 | 21  | Wollishofen (Kreis 2) | 10115          | 6168          | 3947         | 39.0        |
| 2006 | 23  | Leimbach (Kreis 2)    | 3123           | 1997          | 1126         | 36.1        |

Examination of the dataset revealed an important issue: The territorial entity in this dataset is a city district - “*Stadtquartier*”. However, in Zürich, the elections are held across 12 electoral districts - “*Wahlkreise*”. Each electoral district consists of several “*Stadtquartiers*”. For example, electoral district “Kreis 1+2” unites six city districts (they are shown in the “Qname” column in the Table 1 above).

To address the issue, the following manipulations were performed:

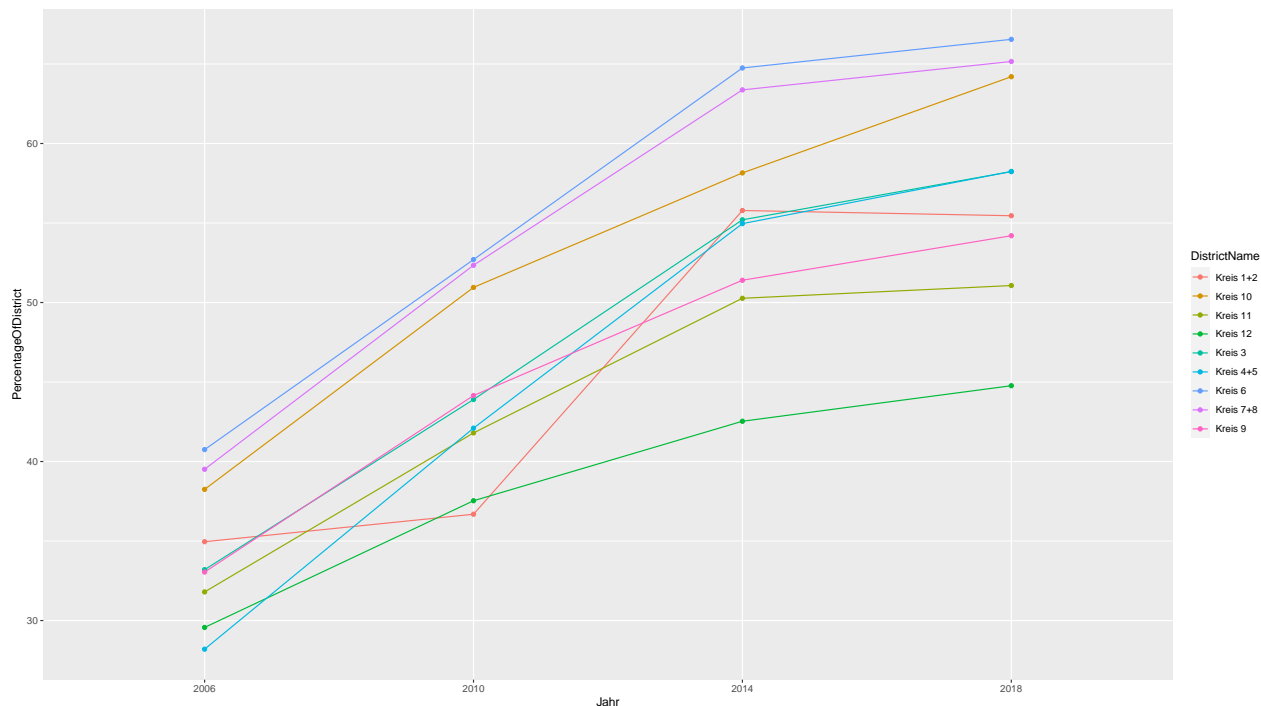
- In the column “Qname”, the electoral district is specified in the parentheses: e.g. “Rathaus (Kreis 1)”. The content of the parentheses was extracted with the **stringr** package and stored in the created column “DistrictName”.
- Rename the electoral districts to reflect the fact that some city districts are merged for the elections - for example “Kreis 1” and “Kreis 2” became “Kreis 1+2”. This was done with the **dplyr mutate()** function.
- Calculate the mean values for each group and year using **dplyr group\_by()** and **summarise()** functions (saved as a separate data frame). The results is saved as a separate data frame, with the new column “PercentageOfDistrict”.
- Calculate percentage of each district’s voters relative to the total number of the city voters using **dplyr group\_by()** %>% **summarise()** %>% **transmute()** functions. The results is saved as a separate data frame, with the new column “PercentageOfCity”.
- Merge the dataframes with the **dplyr inner\_join()**.
- Change “Jahr” and “DistrictName” to factors with the built-in **as.factor()**

Table 2: Final dataset: Turnout by electoral district

| Jahr | DistrictName | PercentageOfDistrict | PercentageOfCity |
|------|--------------|----------------------|------------------|
| 2006 | Kreis 1+2    | 34.95714             | 11.436022        |
| 2006 | Kreis 10     | 38.25000             | 12.176240        |
| 2006 | Kreis 11     | 31.80000             | 13.336140        |
| 2006 | Kreis 12     | 29.56667             | 5.654176         |
| 2006 | Kreis 3      | 33.20000             | 11.112771        |
| 2006 | Kreis 4+5    | 28.20000             | 7.482310         |

The analysis shows that the amount of voters in each district has grown since 2006 - Zürich residents are becoming more active in exercising their right to vote. Kreis 6 has the highest percentage of active citizens,

followed by Kreis 7+8 and Kreis 10. Residents of Kreis 12 are the least interested in the elections people in Zürich. However, when it comes to the percentage of all city voters, Kreis 12 contributes the third largest portion of votes. This means, politicians should mobilize residents of this district to gain votes on the city and national level elections. Other important districts are Kreis 7+8 (gives most votes) and Kreis 11 (second place).



## 2.2 Data set 2: Municipal elections vote share, by party and electoral district since 1913.

The data set of dimensions 5100x6 shows the results of each party at the elections since 1913. Naturally, it contains a lot of irrelevant information because of the changes that happened since 1913. For example, for each year there are separate rows for “Kreis 1 (before 2002)”, “Kreis 2 (before 2002) and”Kreis 1+2 (after 2006)” - the districts were united in 2002. Some political parties do not exist anymore; some parties changed their names. Additionally, the table is in the long format whereas the final data set requires it to be wide.

Table 3: Original data set: Municipal elections vote share, by party and electoral district since 1913.

| Jahr | Partei        | ParteiNr | Wahlkreis    | WahlkreisSort | Stimmenanteil |
|------|---------------|----------|--------------|---------------|---------------|
| 1913 | SP            | 1        | Stadt Zürich | 0             | 39.1          |
| 1913 | BGB/SVP       | 2        | Stadt Zürich | 0             | NA            |
| 1913 | FDP           | 3        | Stadt Zürich | 0             | 38.6          |
| 1913 | GPS           | 4        | Stadt Zürich | 0             | NA            |
| 1913 | GLP           | 5        | Stadt Zürich | 0             | NA            |
| 1913 | CVP/Die Mitte | 6        | Stadt Zürich | 0             | 7.9           |

The following manipulations were performed:

- Choose only relevant time period (years 2006-2018) and eight still existing parties using the **%in%** operator.

- Change the names of the electoral districts. For example, “Kreis 11 (ab 1974)” became simply “Kreis 11” - the content in the parentheses was removed with the **stringr str\_replace()** function.
- Rename the columns with the **dplyr %>% rename()** function to keep them in line with the other tables (e.g. “Wahlkreis” to “DistrictName”).
- Remove unnecessary columns with the **dplyr %>% select()** function.
- Transform the table into wide format with the **tidyverse pivot\_wider()** function.
- Merge the set with the first table “Turnout by electoral district” (see part 2.1.)

Table 4: Final dataset: Municipal elections vote share, by party and electoral district, 2006-2018.

| Year | DistrictName | SP   | BGB/SVP | FDP  | GPS  | GLP  | CVP/Die Mitte | AL   | EVP |
|------|--------------|------|---------|------|------|------|---------------|------|-----|
| 2006 | Kreis 1 + 2  | 30.1 | 16.2    | 23.1 | 13.1 | 2.4  | 7.7           | 2.5  | 3.0 |
| 2010 | Kreis 1 + 2  | 28.0 | 16.9    | 19.3 | 13.7 | 8.6  | 5.8           | 2.6  | 1.9 |
| 2014 | Kreis 1 + 2  | 26.6 | 16.2    | 21.0 | 11.4 | 10.5 | 5.1           | 4.9  | 1.9 |
| 2018 | Kreis 1 + 2  | 30.8 | 13.1    | 19.8 | 12.4 | 10.4 | 4.3           | 6.3  | 1.5 |
| 2006 | Kreis 3      | 37.5 | 18.2    | 8.6  | 14.3 | 2.3  | 7.1           | 6.1  | 2.3 |
| 2010 | Kreis 3      | 34.2 | 16.9    | 8.1  | 13.5 | 10.4 | 5.1           | 6.9  | 1.7 |
| 2014 | Kreis 3      | 32.1 | 15.0    | 10.5 | 12.8 | 10.4 | 4.0           | 9.8  | 1.4 |
| 2018 | Kreis 3      | 35.8 | 10.3    | 10.9 | 14.1 | 10.4 | 3.2           | 12.1 | 1.6 |

The analysis shows that

### 2.3 Data set 3: Wealth distribution of the population in Zürich, by district

The data set of dimensions 756x8 variables reflects how the wealth distribution in absolute terms changed over time per district and per tax class. The following table shows the accumulated wealth distribution across all districts of the city of Zürich between the years 1999 and 2019.

Table 5: Original data set: Distribution wealth tax per category, district and year

| SteuerJahr | KreisSort | KreisLang | SteuerTarifSort |
|------------|-----------|-----------|-----------------|
| 1999       | 1         | Kreis 1   | 0               |
| 1999       | 1         | Kreis 1   | 1               |
| 1999       | 1         | Kreis 1   | 2               |
| 1999       | 2         | Kreis 2   | 0               |
| 1999       | 2         | Kreis 2   | 1               |
| 1999       | 2         | Kreis 2   | 2               |

| SteuerTarifSort | SteuerTarifLang        | SteuerVermoegeen_p50 | SteuerVermoegeen_p25 | SteuerVermoegeen_p75 |
|-----------------|------------------------|----------------------|----------------------|----------------------|
| 0               | Grundtarif             | 23.0                 | 0                    | 174                  |
| 1               | Verheiratetentarif     | 182.0                | 22                   | 711                  |
| 2               | Einelternfamilientarif | 27.5                 | 0                    | 283                  |
| 0               | Grundtarif             | 37.0                 | 3                    | 186                  |
| 1               | Verheiratetentarif     | 148.0                | 33                   | 458                  |
| 2               | Einelternfamilientarif | 7.0                  | 0                    | 61                   |

Examining the previous showed data revealed that the data set contains a lot of irrelevant information. For example, the columns “KreisSort” and “KreisLang” are redundant, since the first is simply the encoding of

the second in number. The same applies for the the columns “*SteuerTarifSort*” and “*SteuerTarifLang*”, since the first here again, is the encoding of the second in number. However, the columns “*KreisLang*” and “*SteuerTarifLang*” are redundant and therefore, dropped. For practical reasons the columns “*SteuerVermoegeen\_p25*” and “*SteuerVermoegeen\_p75*” were dropped as well.

Moreover, the columns “*SteuerTarifSort*” and “*KreisSort*” are converted to factors, since those columns are automatically defined as integer number. Additionally, the names of the columns do not correspond with the previous data set and therefore, we have to change the following:

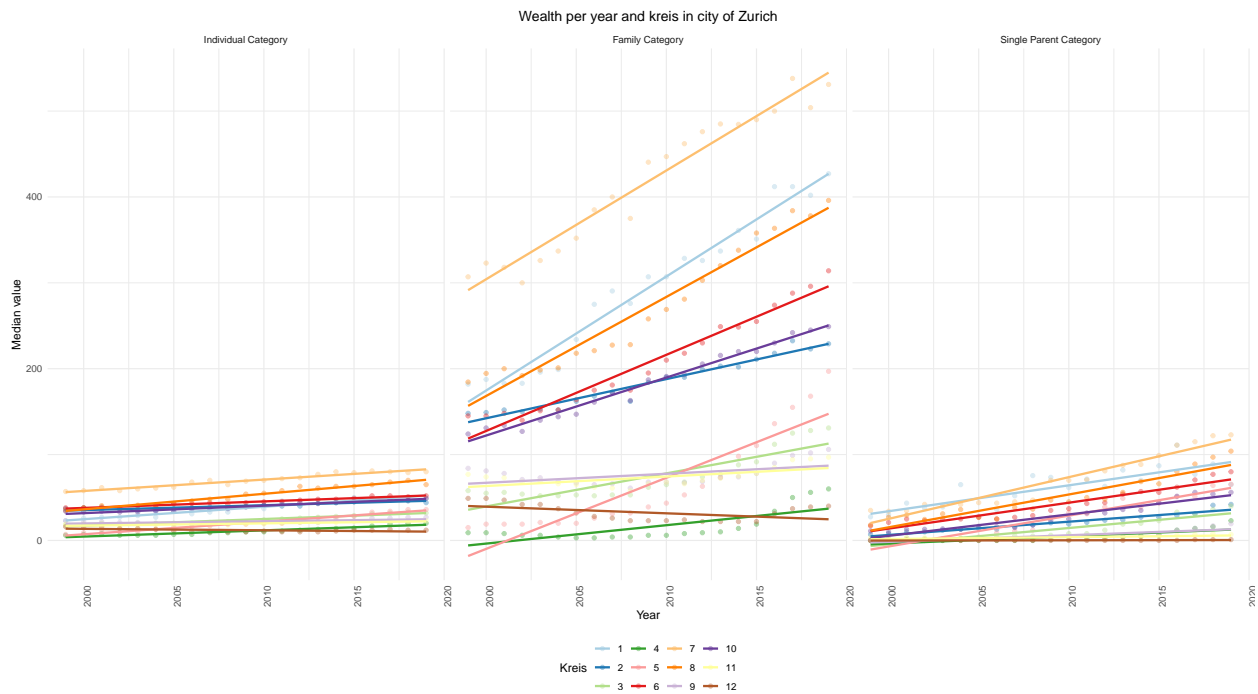
| Original Column Name | New Column Name |
|----------------------|-----------------|
| KreisSort            | DistrictNumber  |
| SteuerJahr           | Year            |
| SteuerVermoegeen_p50 | Wealth          |
| SteuerTarifSort      | Category        |

Finally, after all modifications, the data set looks as follows:

Table 8: Final data set: Distribution wealth tax per category, district and year

| Year | DistrictNumber | TaxCategory | Wealth |
|------|----------------|-------------|--------|
| 1999 | 1              | 0           | 23.0   |
| 1999 | 1              | 1           | 182.0  |
| 1999 | 1              | 2           | 27.5   |
| 1999 | 2              | 0           | 37.0   |
| 1999 | 2              | 1           | 148.0  |
| 1999 | 2              | 2           | 7.0    |

The next graph shows the distribution of wealth per year, tax category and district in Zurich.



The previous graph shows the distribution of wealth by district, year and tax category. Since the ordinate is scaled for all tax categories with the same values, it is clearly visible the difference of accumulated wealth between the districts across all and those differences seems to have a clear trend in increasing. The biggest differences between districts can be seen in the *family category* subgraph. The highest values have been observed for district 7 and the lowest values for district 12. District 7, however, has the highest values among all tax categories. District 12 shows as well across all categories the lowest values of accumulated as well.

## 2.4 Data set 4: Income distribution of the population in Zürich, by district

The following data set of dimensions 756x8 variables reflects how the income distribution in absolute terms changed over time per district and per tax class. The following table shows the accumulated income distribution across all districts of the city of Zürich between the years 1999 and 2019.

Table 9: Original data set: Distribution income tax per category, district and year

| SteuerJahr | KreisSort | KreisLang | SteuerTarifSort |
|------------|-----------|-----------|-----------------|
| 1999       | 1         | Kreis 1   | 0               |
| 1999       | 1         | Kreis 1   | 1               |
| 1999       | 1         | Kreis 1   | 2               |
| 1999       | 2         | Kreis 2   | 0               |
| 1999       | 2         | Kreis 2   | 1               |
| 1999       | 2         | Kreis 2   | 2               |

| SteuerTarifLang        | SteuerEinkommen_p50 | SteuerEinkommen_p25 | SteuerEinkommen_p75 |
|------------------------|---------------------|---------------------|---------------------|
| Grundtarif             | 37.8                | 17.40               | 64.80               |
| Verheiratetentarif     | 83.4                | 52.00               | 130.20              |
| Einelternfamilientarif | 46.7                | 26.05               | 87.05               |
| Grundtarif             | 37.9                | 19.90               | 58.20               |
| Verheiratetentarif     | 69.7                | 49.10               | 101.40              |
| Einelternfamilientarif | 39.2                | 21.90               | 58.90               |

Examining the previous showed data revealed that the data set contains a lot of irrelevant information. For example, the columns “*KreisSort*” and “*KreisLang*” are redundant, since the first is simply the encoding of the second in number. The same applies for the columns “*SteuerTarifSort*” and “*SteuerTarifLang*”, since the first column, here again, is the encoding of the second in number. However, the columns “*KreisLang*” and “*SteuerTarifLang*” are redundant and therefore, dropped. For practical reasons the columns “*SteuerEinkommen\_p25*” and “*SteuerEinkommen\_p75*” were dropped as well.

Moreover, the columns “*SteuerTarifSort*” and “*KreisSort*” are converted to factors, since those columns are automatically defined as integer number. Additionally, the names of the columns do not correspond with the previous data set and therefore, we have to change the following:

| Original Column Name | New Column Name |
|----------------------|-----------------|
| KreisSort            | DistrictNumber  |
| SteuerJahr           | Year            |
| SteuerEinkommen_p50  | Income          |
| SteuerTarifSort      | Category        |

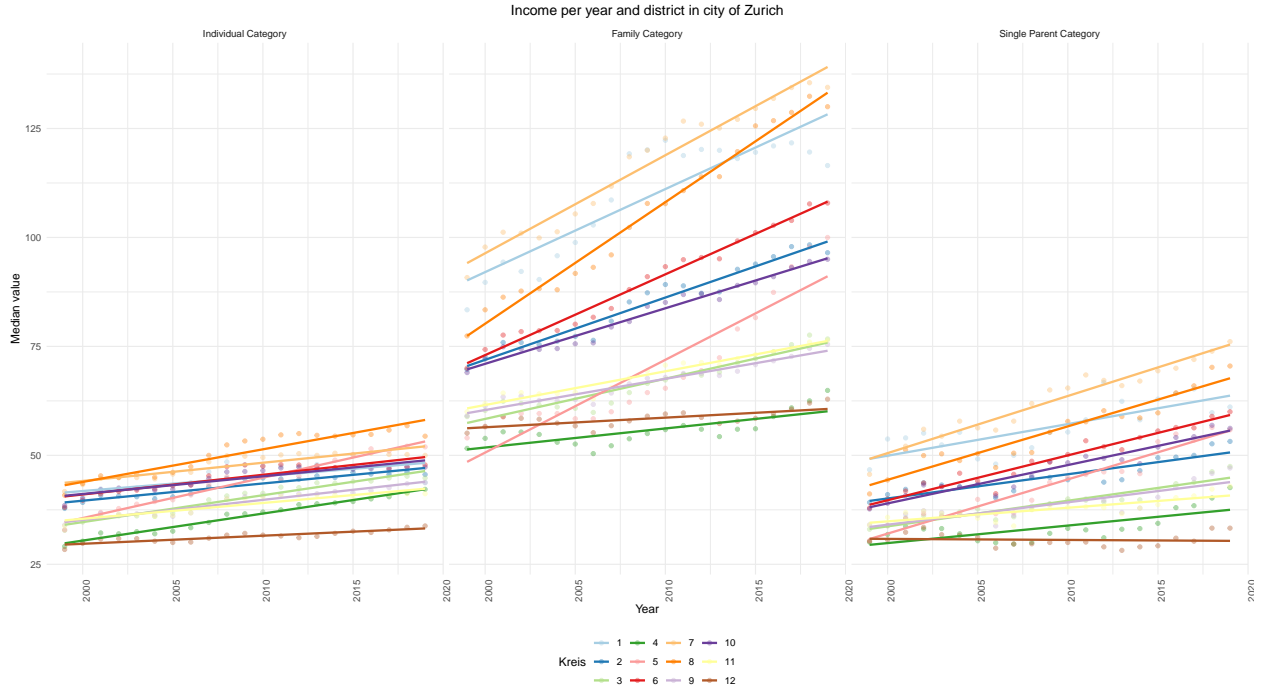
Finally, after all modifications, the data set looks as follows:



Table 12: Final data set: Distribution income tax per category, district and year

| Year | DistrictNumber | TaxCategory | Income |
|------|----------------|-------------|--------|
| 1999 | 1              | 0           | 37.8   |
| 1999 | 1              | 1           | 83.4   |
| 1999 | 1              | 2           | 46.7   |
| 1999 | 2              | 0           | 37.9   |
| 1999 | 2              | 1           | 69.7   |
| 1999 | 2              | 2           | 39.2   |

The next graph shows the distribution of income per year, tax category and district in Zurich.



The previous graph shows the distribution of income by district, year and tax category. Since the ordinate is scaled for all tax categories with the same values, it is clearly visible the difference of accumulated income between the districts across all and those differences seems to have a clear trend in increasing. The biggest differences between districts can be see in the *family category* subgraph. The highest values have been observed for district 7 and the lowest values for district 12. District 7, however, has the highest values among all tax categories. District 12 shows as well across all categories the lowest values of accumulated as well.

## 2.5 Combine wealth and income data

The aim of this part is to combine the income and wealth data into one data frame. As a first step, both data frames are combined using the *merge()* function. The merged data sets of income and wealth data, contains the income data and wealth data per district and tax category combined.

Table 13: Merged income and wealth data

| Year | DistrictNumber | TaxCategory | Income | Wealth |
|------|----------------|-------------|--------|--------|
| 1999 | 1              | 0           | 37.8   | 23.0   |
| 1999 | 1              | 1           | 83.4   | 182.0  |
| 1999 | 1              | 2           | 46.7   | 27.5   |
| 1999 | 10             | 0           | 38.5   | 34.0   |
| 1999 | 10             | 1           | 69.0   | 124.0  |
| 1999 | 10             | 2           | 37.7   | 11.0   |

Thereafter, the district name has to be modified in order to match if the data presented in sections concerning *Data set 1* and *Data set 2*: The name of the districts corresponds to the prefix “*Kreis*” followed by “+” and the corresponding number. In the case of the *wealth* and *income* data frames, those values have to be transformed and combined as the following correspondence tables shows:

Table 14: Correspondence table between district names and district numbers

| Number | Name        |
|--------|-------------|
| 1      | Kreis 1 + 2 |
| 2      | Kreis 1 + 2 |
| 3      | Kreis 3     |
| 4      | Kreis 4 + 5 |
| 5      | Kreis 4 + 5 |
| 6      | Kreis 6     |
| 7      | Kreis 7 + 8 |
| 8      | Kreis 7 + 8 |
| 9      | Kreis 9     |
| 10     | Kreis 10    |
| 11     | Kreis 11    |
| 12     | Kreis 12    |

Thereafter, several transformations have to be done as follows in order to transform the data into a mergable format:

- Pivot to a longer format the tables and combine the tax values
- Group by year, district name and tax type
- Sum the values amount all categories within one district
- Pivot to wide format again to averaged values

After the step-by-step implementation of those transformation steps one gets the following data frame:

Table 15: Final wealth and income data frame per district and year

| Year | DistrictName | SumTaxIncome | SumTaxWealth |
|------|--------------|--------------|--------------|
| 1999 | Kreis 1 + 2  | 314.70       | 424.5        |
| 1999 | Kreis 3      | 124.75       | 76.0         |

| Year | DistrictName | SumTaxIncome | SumTaxWealth |
|------|--------------|--------------|--------------|
| 1999 | Kreis 4 + 5  | 228.10       | 37.0         |
| 1999 | Kreis 6      | 146.00       | 200.0        |
| 1999 | Kreis 7 + 8  | 337.55       | 636.5        |
| 1999 | Kreis 9      | 127.15       | 108.0        |

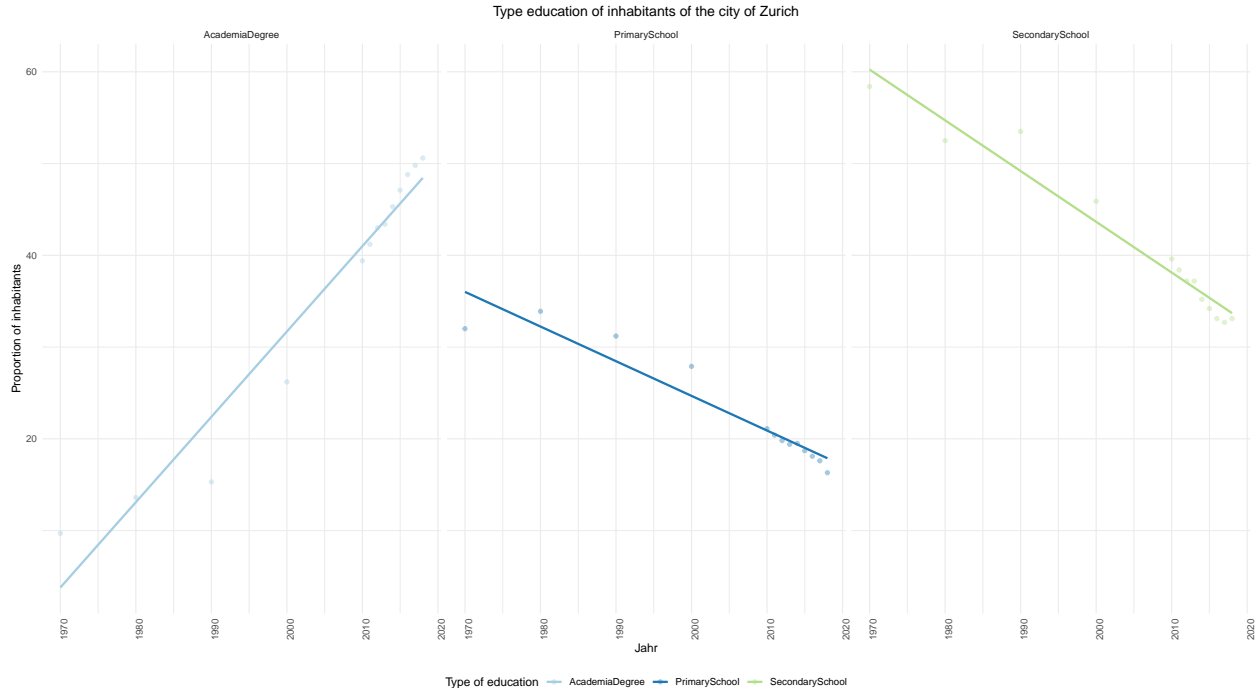
## 2.6 Education data

The following data set contains 39x5 variables on type of education and year of the complete city of Zürich. This data set allows us to try to infer the change in education per type in the city of Zürich and interpolate the change of education type on a district respectively on a neighborhood level.

Table 16: Original data set: Education distribution per category and year

| Jahr | Bildungsstand         | AntBev | untAntBevKI | obAntBevKI |
|------|-----------------------|--------|-------------|------------|
| 1970 | Obligatorische Schule | 32.0   | NA          | NA         |
| 1970 | Sekundarstufe II      | 58.4   | NA          | NA         |
| 1970 | Tertiärstufe          | 9.7    | NA          | NA         |
| 1980 | Obligatorische Schule | 33.9   | NA          | NA         |
| 1980 | Sekundarstufe II      | 52.5   | NA          | NA         |
| 1980 | Tertiärstufe          | 13.6   | NA          | NA         |

The following graphs shows the change of education in the city of Zürich between the year 1970 and 2018.



The previous graph shows, that the number of inhabitants without an academical degree is decreasing in the course of time since 1970 in the City of Zürich, while the number of persons having a university degree is increasing. First, the values of the previous data frame have to be re-arranged using the `pivot_wide()` and the names of the education type are renamed to match the correspondence table

Table 17: Pivoted data set: Education distribution per category and year

| Year | PrimarySchool | SecondarySchool | AcademiaDegree | DeltaYear |
|------|---------------|-----------------|----------------|-----------|
| 1970 | 32.0          | 58.4            | 9.7            | 0         |
| 1980 | 33.9          | 52.5            | 13.6           | 10        |
| 1990 | 31.2          | 53.5            | 15.3           | 20        |
| 2000 | 27.9          | 45.9            | 26.2           | 30        |
| 2010 | 21.1          | 39.6            | 39.4           | 40        |
| 2011 | 20.4          | 38.4            | 41.2           | 41        |

In order to be able to reproduce this phenomena in the previous data set, the parameters of theses changes is going to be calculated using the  $lm()$  function. The used formula is the following:

$$N_{AntBev} = c + \frac{\Delta d_i}{\Delta t} \cdot t$$

Applying the previous showed formula to the other data set one gets the following values

Table 18: Coefficients of the linear model depending on the education type

|             | PrimarySchool | SecondarySchool | AcademiaDegree |
|-------------|---------------|-----------------|----------------|
| (Intercept) | 36.0127064    | 60.240920       | 3.7915859      |
| DeltaYear   | -0.3777745    | -0.552921       | 0.9300644      |

The previous table shows the coefficients of the linear model of each education type. Before further transformation, we load a further data frame. The data set of dimensions 138x6 variables reflects how the education of the inhabitants changed over time per district and per education class. The following table shows the proportion of education per districts of the city of Zürich for the year 2021. The data set is updated on a yearly basis.

Table 19: Original data set: Education distribution per category and district for the year 2021

| RaumSort | RaumLang | Bildungsstand | AntBev | untAntBevKI | obAntBevKI |
|----------|----------|---------------|--------|-------------|------------|
| 10       | Kreis 1  | 1             | 9.8    | 6.7         | 12.9       |
| 10       | Kreis 1  | 2             | 26.8   | 22.2        | 31.4       |
| 10       | Kreis 1  | 3             | 63.4   | 58.4        | 68.4       |
| 11       | Rathaus  | 1             | 9.9    | 5.9         | 13.9       |
| 11       | Rathaus  | 2             | 27.7   | 21.8        | 33.6       |
| 11       | Rathaus  | 3             | 62.3   | 55.9        | 68.7       |

The column “*RaumSort*” encodes the district number and corresponding neighborhood within the district. The last digit describes the neighborhood and the first one (or two) digit the district number. The column “*RaumLang*” describes the district and neighborhood name and the column “*Bildungsstand*” describes the type of education as follows:

Table 20: Education ecoding: Education number key vs. meaning in words

| EducationNumber | Meaning          |
|-----------------|------------------|
| 1               | Primary school   |
| 2               | Secondary school |
| 3               | Academia degree  |

Where type *1* describes the nine complete year of the mandatory school program, *2* describes either a professional degree or high-school diploma and *3* an academical degree (Bachelor's degree and higher). The columns "*untAntBevKI*" and "*obAntBevKI*" describes the lower and upper portion of the confidence interval of the values per education. Summarizing, the data frame is going to be updated as follows:

- "*Bildungsstand*" is update with the meaning as described in the previous table
- Columns "*untAntBevKI*" and "*obAntBevKI*" are dropped
- Column "*Year*" is added with value *2021*
- Column "*RaumSort*" is transformed to column "*DistrictNumer*"
- District designation adapted
- District summarized according to voting data frames

Those transformations steps leads to the following data frame:

Table 21: Transformed data frame for education in year 2021

| Year | DistrictName | PrimarySchool | SecondarySchool | AcademiaDegree |
|------|--------------|---------------|-----------------|----------------|
| 2021 | Kreis 1 + 2  | 12.5          | 29.7            | 57.8           |
| 2021 | Kreis 10     | 12.4          | 33.7            | 54.0           |
| 2021 | Kreis 11     | 20.3          | 35.5            | 44.2           |
| 2021 | Kreis 12     | 29.2          | 39.6            | 31.1           |
| 2021 | Kreis 3      | 17.2          | 31.6            | 51.2           |
| 2021 | Kreis 4 + 5  | 13.9          | 26.9            | 59.3           |
| 2021 | Kreis 6      | 10.0          | 25.4            | 64.6           |
| 2021 | Kreis 7 + 8  | 8.4           | 26.4            | 65.2           |
| 2021 | Kreis 9      | 20.0          | 38.6            | 41.5           |

The previous table shows the transformed data for the year 2021. Now, taking into account the previously calculated *lm coefficients* one can extrapolate values for the missing years of the data frame.

### 3 Merging Data

## 4 Data Visualization

## 5 Fit Model



## 6 Chapter of Choice TBD

## 7 References