Project proposal

Team number 4

1. Introduction

Public housing is government-owned housing units provided to low-income families, the elderly, and individuals with disabilities at affordable rent. The primary goal of public housing is to offer safe, decent, and sanitary living conditions for those who cannot afford market-rate housing. Nowadays, multiple public housing companies (most of them are funded by the Israeli government) own a bit more than 50,000 public housing apartments around Israel. Since the Ministry of Construction permits eligible individuals to obtain a housing unit only in the city where they are registered, there are currently approximately 4,300 eligible individuals on the waiting list for an apartment. Additionally, there are 25,000 more eligible individuals (including 6,000 Holocaust survivors) with an average waiting time of 3 to 4 years to receive an apartment, and even longer in Jerusalem and the central region. Those stats raises several questions about the public housing procurement policy lead by the government, our project will focus on this main question:

Is there a relation between general elections results per city, to the public housing procurement policy in the same city? Our project aims to find a relation (if exist) between the elected government policy regarding public housing, to the top voted parties in each city. Does the government buy more apartments in cities who voted for her? Does it buy less apartments in cities which didn't vote for her? In which cities does the government sell more housing units? In which regions there are more vacant apartments? Did those regions vote for the elected government? Since 2015, Israel has had 5 national elections, 4 governments and 7 different ministers leading the Ministry of Construction. Research up to this point, focused on the shortage of apartments and the long waiting time, but we aim to find whether there is a hidden government policy behind the scenes, based on the elected government vote rate in each city. Our approach to answer the question is to combine multiple data sets regarding public housing (property purchases, vacant apartments, etc) and data sets about general elections per city (most elected parties per city). Then, we wish to find patterns which can indicate certain government's policies to help us find an answer.

2. Data

Elections results: information about the last elections results in every ballot box in the country. A brief explanation about the variables:

- ballot id – id of each ballot box. an integer number.
- Committee symbol – the ballot's geographical areas. Each area gets a number in range 1 to 20, ballot boxes inside IDF bases get 99.
- settlement name – the settlement of the ballot box.

- settlement ID – settlement's id of each ballot box.
- Ballot num – a float number that identifies the ballot box in its settlement. It can repeat itself in different settlements.
- BZB – number of eligible voters registered to the ballot box.
- Voters – number of votes in the ballot box.
- Rejected – number of uncounted votes in the ballot box.
- Accepted – number of counted votes in the ballot box.
- The rest of the variables are all the parties and how many votes each party got.

Properties by rooms: Each record represents a settlement, a floor number, and the number of public housing properties in the country which are from that settlement and on that floor number. A brief explanation about the variables:

- CityLMSName – string of the settlement's name.
- Floor – floor number. Integer.
- OneRooms – number of one room apartments that exist in the settlement and on the floor number.
- TwoRooms – number of two rooms apartments that exists in the settlement and on the floor number. And it continues this way until TenRooms, and MoreRooms (more than 10) variables.

Property purchases: Information about public housing property purchasing per settlement & year.
A brief explanation about the variables:

- Lamas code – identification number of a settlement.
- Year – 2017 – 2024.
- Lamas name – the settlement's name.
- 3 rooms apartments – number of purchased 3 rooms apartments.
- 4+ rooms apartments – number of purchased 4 or more rooms apartments.
- average price (NIS) 3 rooms apartments – the average purchase price of 3 rooms apartment.
- average price (NIS) 4+ rooms apartments – the average purchase price of 4 or more rooms apartment.

Vacant apartments: List of public housing apartments that have been available for more than 3 months.
A brief explanation about the variables:

- CityLmsName – settlement's name.
- CityLmsCode – settlement id, number.
- NumOfRooms – number of rooms in the apartment (a float number because there're apartments with half-rooms).
- Floor – settlement 's floor number.
- TotalArea – area of the apartment in square meter.
- CompanyName – the company that owns the apartment.
- StatusChangeDate – date the apartment has been available from.

3. Preliminary results

Source code

Property purchases: The **mean cost** of a 3 rooms apartment is 1,405,986 NIS, and the mean cost of a 4 rooms apartments is 1,671,485 NIS. There are **5 companies** that own available apartments, Amidar with 233 apartments, Amigur with 34, Co. Ltd with 8 and flint and shakmona with 1 each. Most of the apartment inventory consists of **2, 3 or 4 rooms** apartments, with a distribution of 29.3%, 48.6% and 15% accordingly.
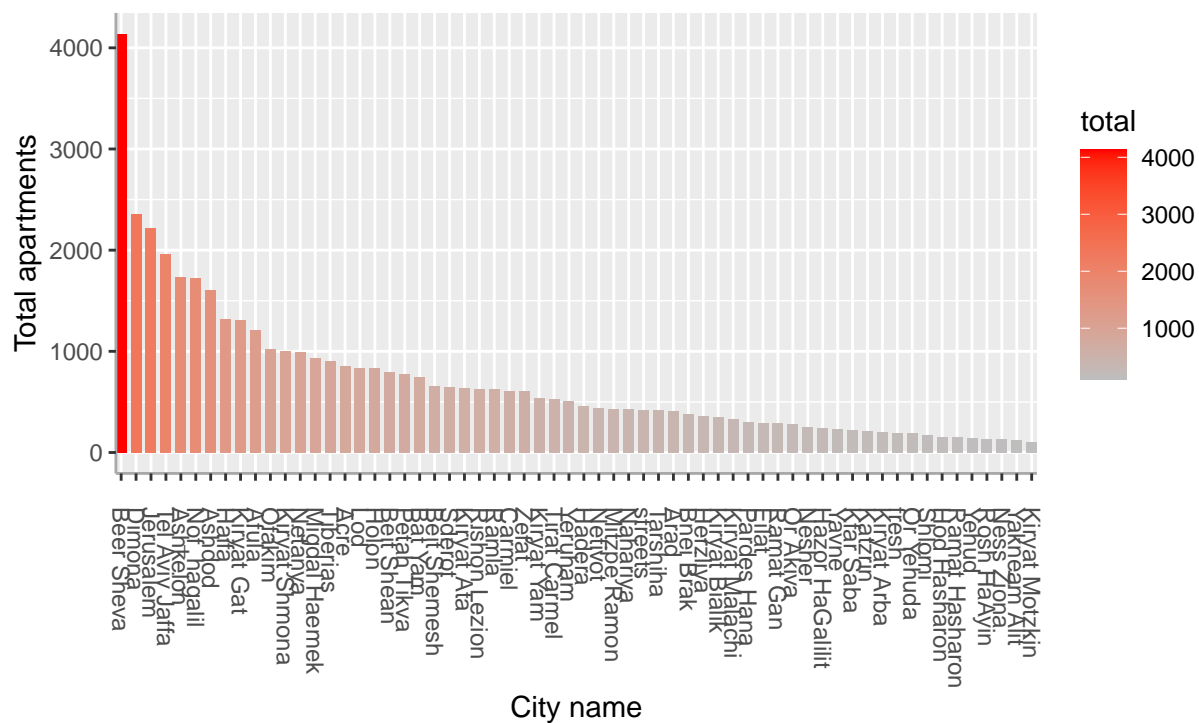

Number of available apartments by city
collected in May 24
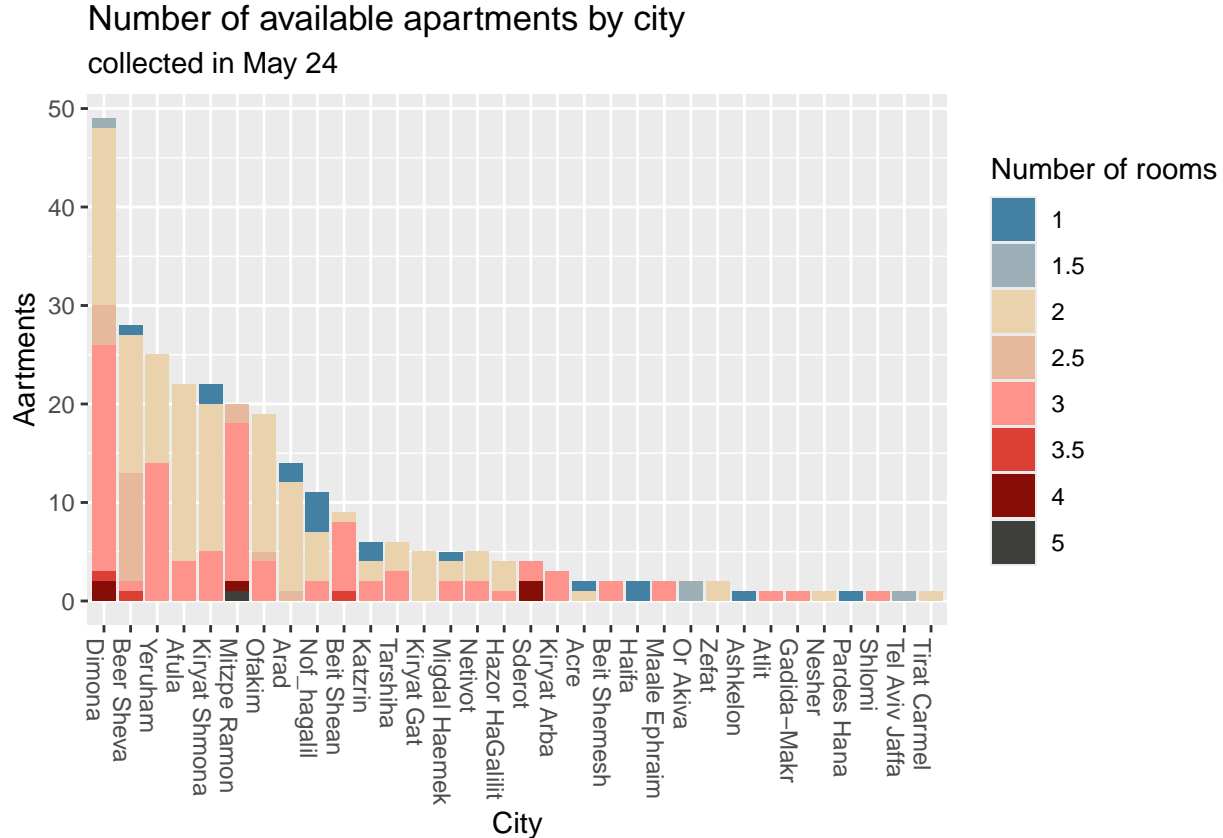
Apartments bought each year by City
From 2017 to 2024
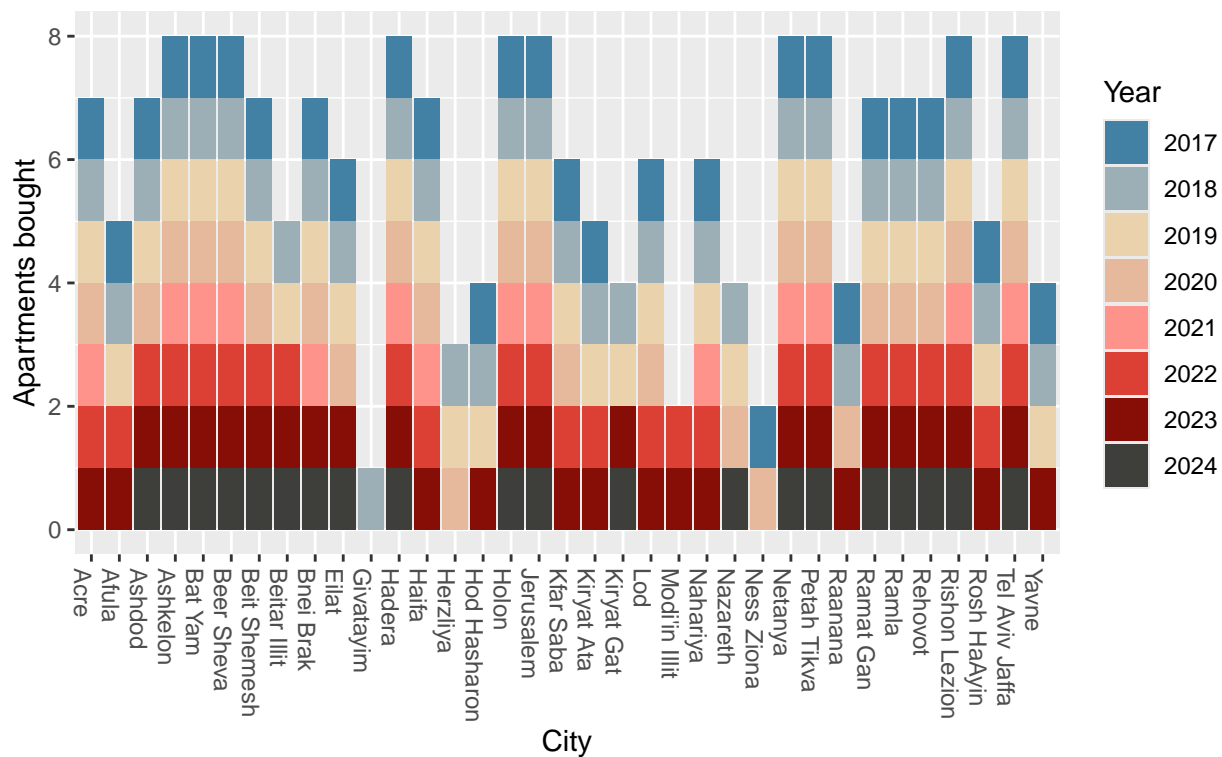
Total apartments by City
Where there are more than 100 apartments in the city

Looking at the data visualization and some summary statistics we can see that there are some

concentrations for the current public housing system, where a major portion of the apartments are 2 and 3 rooms which may not help those who need theme. Furthermore there is 1 company owning most of the apartments and a lot of the apartments are located in southern Israel in cities such as Be'er Sheva or Dimona.

4. Data analysis plan

1. Outcome (Response, Y) and Predictor (Explanatory, X) Variables

a. First Relation to examine : Outcome (Y): The number of houses bought per city per year. Explanatory (X): Segmentation of voting percentages in that city.
b. Second Relation to examine : Outcome (Y): The number of houses bought in a certain city in relation to the percentage of apartments available out of the total number of existing apartments in the city. Explanatory (X): Segmentation of voting percentages in that city.

2. Comparison Groups

- Temporal Comparison: Different years where election results were different.
- Spatial Comparison: Different cities in the same year with varying voting patterns.

3. Methods to Answer the Question

- **Data Preprocessing:** Ensuring all data is in a uniform language, Predicting or imputing null values to complete the dataset. Removing duplicates, correcting errors, and ensuring consistency. Creating new variables or modifying existing ones to better fit the analysis
- **Descriptive Analysis:** Describe the current public housing procurement policy, its components, and influencing factors. make variable overview: Detail the apartment inventory, political state of the city, and other relevant variables.
- **Statistical Analysis:**
  - Logistic Regression: To model the probability of a certain number of houses being bought based on voting percentages.
  - Linear Regression: To measure the linear relationship between the number of houses bought and the voting percentages.
  - Non-linear Regression: For more complex relationships that cannot be adequately modeled by linear regression.
  - Comparative Analysis: Using t-tests or ANOVA to compare housing procurement policies across different cities and years based on their voting patterns.
  - Correlation Analysis: To check the strength and direction of the relationship between voting patterns and housing purchases.

4. Necessary Results to Support the Hypothesis*

a. Comparison Analysis: Identify and quantify differences in public housing procurement policies by comparing cities with different political leanings.
b. Inventory Analysis: Measure the percentage of vacant apartments and correlate this with housing purchases.

5. Teamwork and Division of Work

- Dori: Data Collection and Preprocessing - Gather and clean data, handle translations, manage missing values.
- Daniel: Descriptive and Exploratory Analysis - Conduct descriptive analysis, visualize data, and provide initial insights.
- Ofek: Statistical Analysis - Perform logistic, linear, and non-linear regression analyses, comparative and correlation analyses.
- Shahaf: Reporting and Presentation - Consolidate findings, prepare the final report, and create visual presentations for clarity and communication of results.

Differences from a Team of Three: A team of four allows for more specialisation and a more in-depth analysis, reducing individual workloads and increasing the quality and detail of the project. In contrast, a team of three may need to limit the scope of analysis and take on multiple roles, potentially leading to slower progress and less detailed insights. Collaboration and Communication: Tedious work will happen offline where each person will work at home. We will share a Git repository to keep the R code updated. Bi-weekly Zoom meetings will be held to go over the tasks needed to be done, ensuring everyone is on track and any issues can be addressed promptly. This regular communication is essential, especially considering Daniel's Miloim (reserve duty) commitments. During these meetings, we will review progress, plan next steps, and ensure alignment among all team members. By maintaining clear and consistent communication, we aim to work efficiently and effectively, despite being in different locations.

Appendix

Data README

```
# SISE2601 Project data description
================
Team 4

Elect results:
Rows: 12,545
Columns: 51

Properties by rooms:
Rows: 843
Columns: 14

Property purchases:
Rows: 208
Columns: 7

Vacant apartments:
Rows: 277
Columns: 14
```