

Data Analysis on the prices of houses in King County

Customised for Erin Robinson

Clemens Bretscher, 13.10.2021

Who is Erin Robinson?

- Erin Robinson wants to **buy and sell** houses in **poor neighborhoods**.
- He wants to his **costs back** and only wants a **small profit**, he wants to act **socially responsible**

My mission

What are the key investigations?

- Give recommendations on buying/selling houses
- Targeted district: poor neighborhood
- The price of the house should be low
- The house should be in a condition that it can be easily renovated

Available Data

- **One table** of data with **20 factors** for each listed house
- About **21600 houses** listed

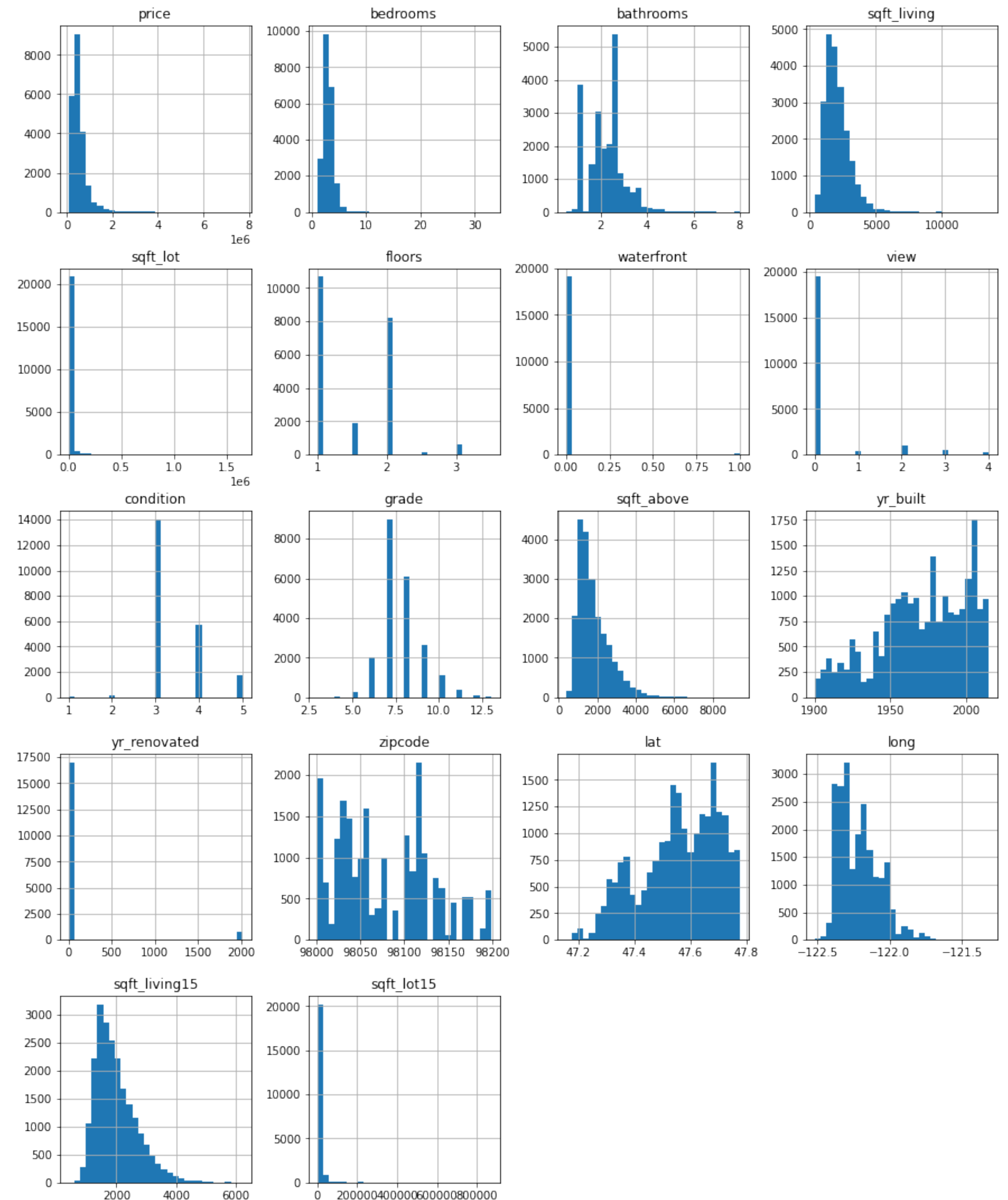
Data summary

Factors:

- Id of the house
- The date at which the house was last sold
- Price of the house
- Number of bedrooms
- Number of bathrooms
- Square feet of living space
- Square feet of lot space
- Number of floors
- If there is a body of water in front of the house
- If the house has been viewed already
- Overall condition of the house
- Overall grade given to the housing unit, based on King County grading system
- Square footage of house apart from the basement
- Square feet of the basement
- Built year
- Year when the house was last renovated
- Zipcode
- Latitude coordinate
- Longitude coordinate
- The square footage of interior housing living space for the nearest 15 neighbors
- The square footage of the land lots of the nearest 15 neighbors

Quality of data

- There is numerical and categorical data
- Some factors follow a similar distribution like the price.



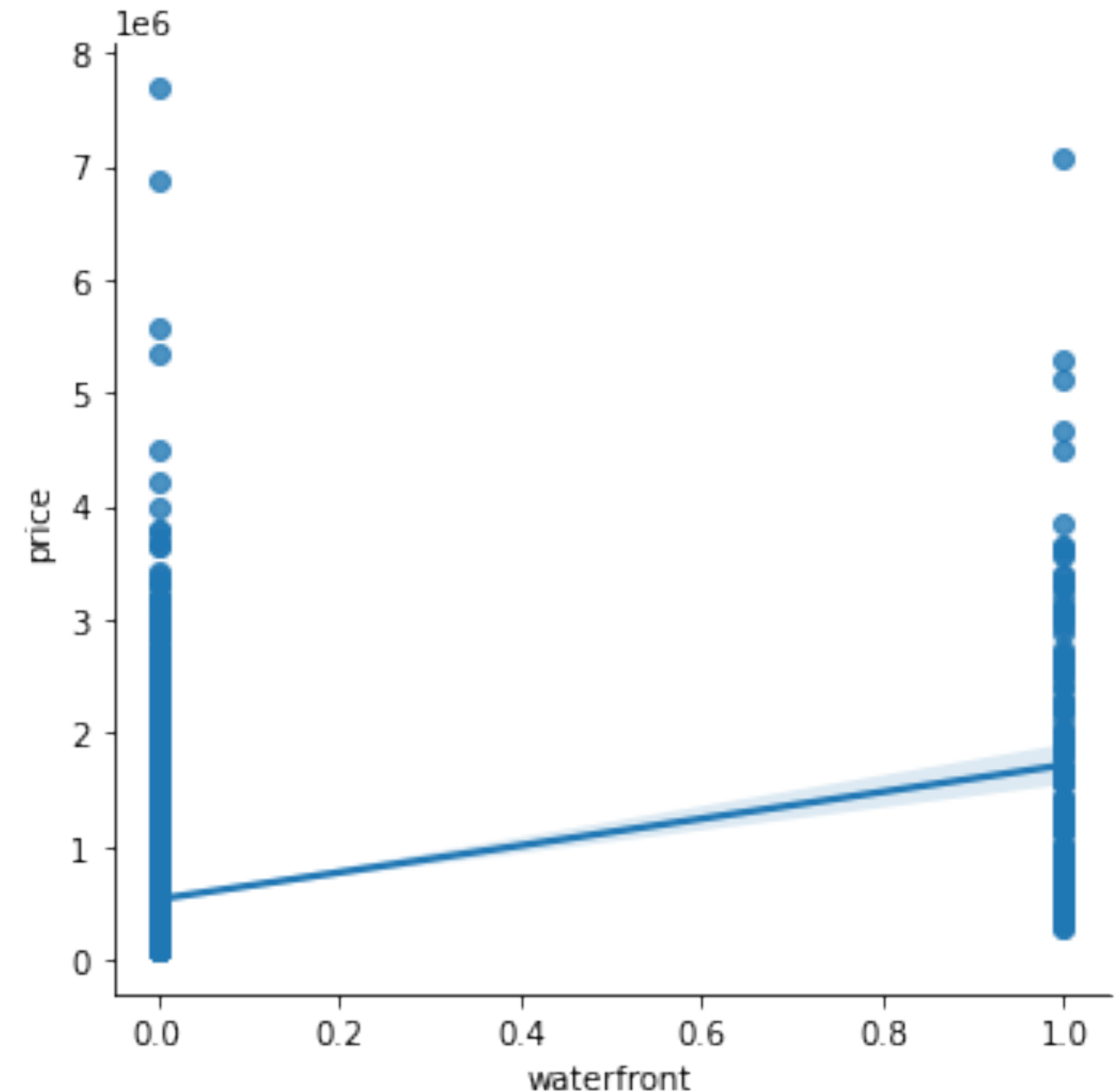
Hypotheses

- H_01 : When the house is next to a **body of water** the price will be higher
- H_02 : The worse the **condition or the grade** the lower the price will be
- H_03 : The price depends on the **zip code**

H₀1

When the house is next to a **body of water** the price will be higher

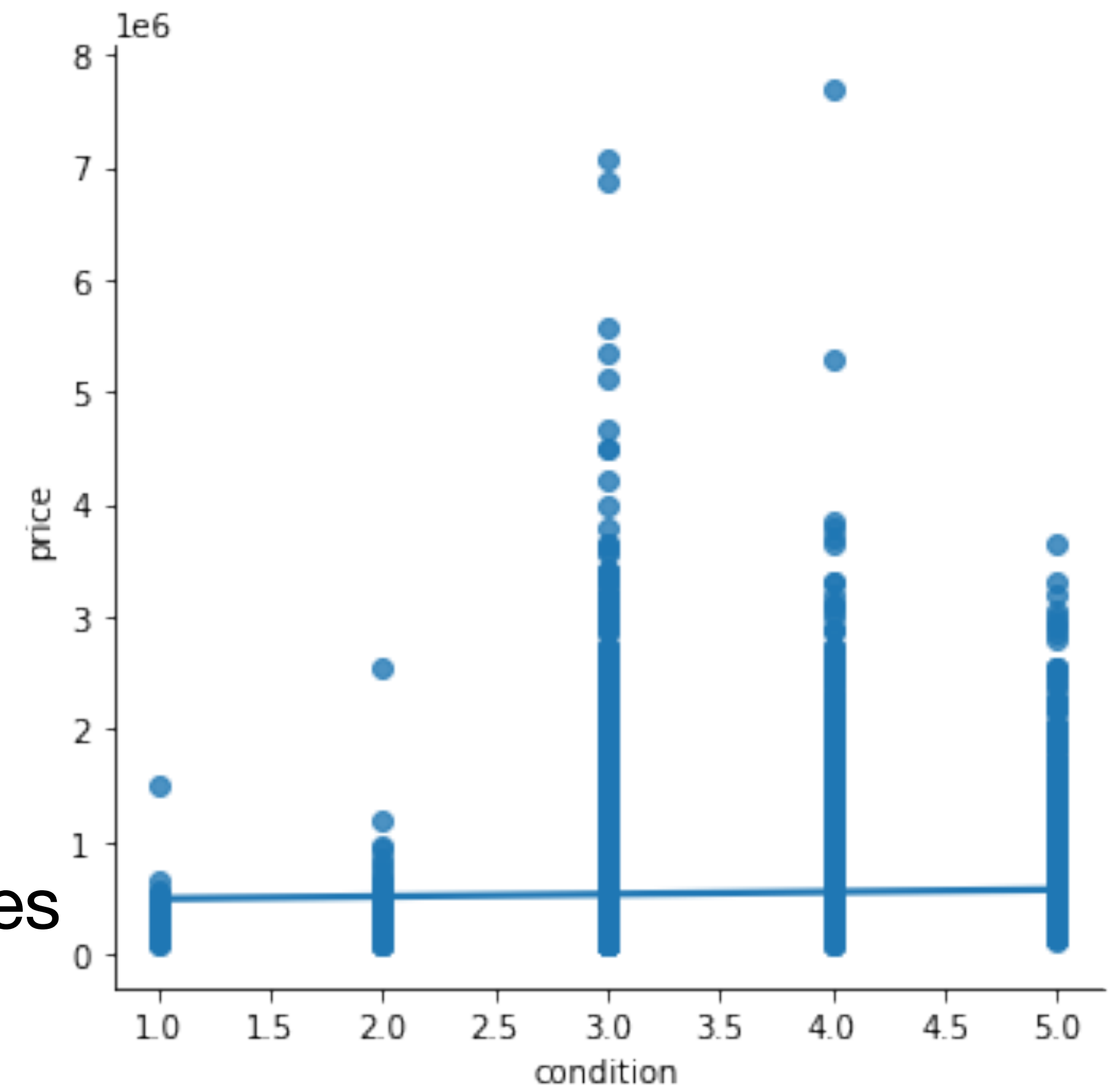
- OLS model:
 - t-value : 39.855
 - $P > |t|$: 0.000
 - R^2 : 0.076
- Model: $\text{price} = 5.326 \times 10^5 + 1.185 \times 10^6 \times \text{waterfront}$
- Interpretation: A waterfront is an indicator for more expensive houses. However, of course many expensive houses do not have a waterfront.



H₀2

The worse the **condition** or the grade the lower the price will be

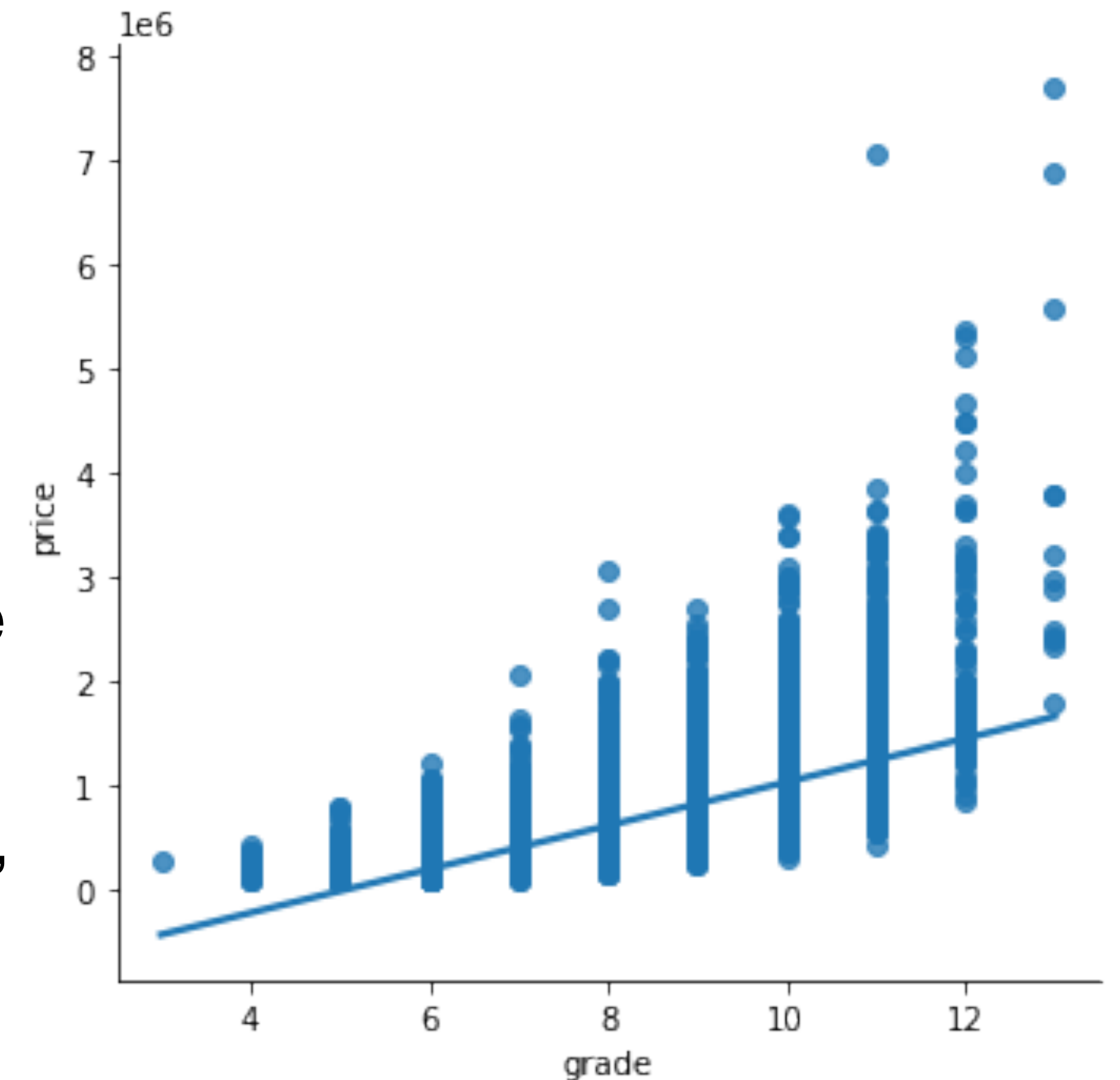
- OLS model:
 - t-value : 5.302
 - $P > |t|$: 0.000
 - R^2 : 0.001
- Model: $\text{price} = 4.709 \times 10^5 + 2.036 \times 10^4 \times \text{condition}$
- Interpretation: The condition of the house does not seem to reflect the price of the house.



H₀2

The worse the condition or **the grade** the lower the price will be

- OLS model:
 - t-value : 131.895
 - $P > |t|$: 0.000
 - R^2 : 0.446
- Model: $\text{price} = -1.061 \times 10^6 + 2.092 \times 10^5 \times \text{grade}$
- Interpretation: The grade seems to be a good predictor of the price. In contrast to the condition, the grade of the house is based on King County grading system, which seems to be reliable



H₀3

The price depends on the **zip code**

- Mean prices of individual zip codes were calculated and plotted
- OLS model:
 - t-value : 12.434
 - $P > |t|$: 0.000
 - R^2 : 0.695
- Model: $\text{price} = 1.338 \times 10^5 + 1.238 \times 10^4 \times \text{idx_zipcode}$
- Interpretation: The zipcodes can be categorized depending on their mean price.

