

Visualize

Isaac Verminck

2018-04-30

```
read_tidy_listings()
```

Rental prices by municipality

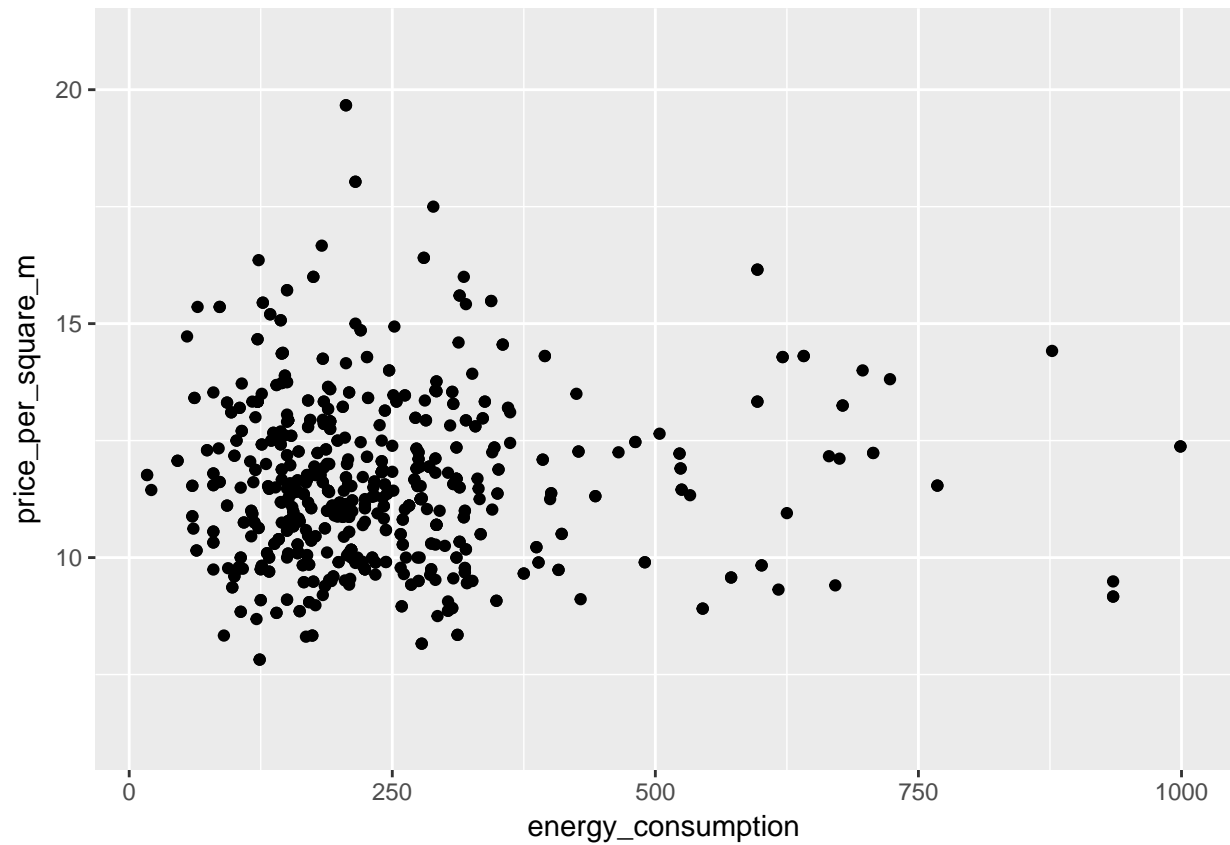
We want to know if some municipalities are more expensive than others. Of course we want to avoid comparing apples and oranges so we first have to find out what variables influence the price per square meter. For example the current state could make a large difference (people will pay more for a recently renovated appartement than they would for a appartement to be renovated).

The variables in our data having a possible influence are the current state, the energy consumption level and the construction year. Current state is categorical, the other two are continuous.

We start off with a naive approach where we don't take any confounding variables into account and we gradually peel back confounding variables.

Energy consumption

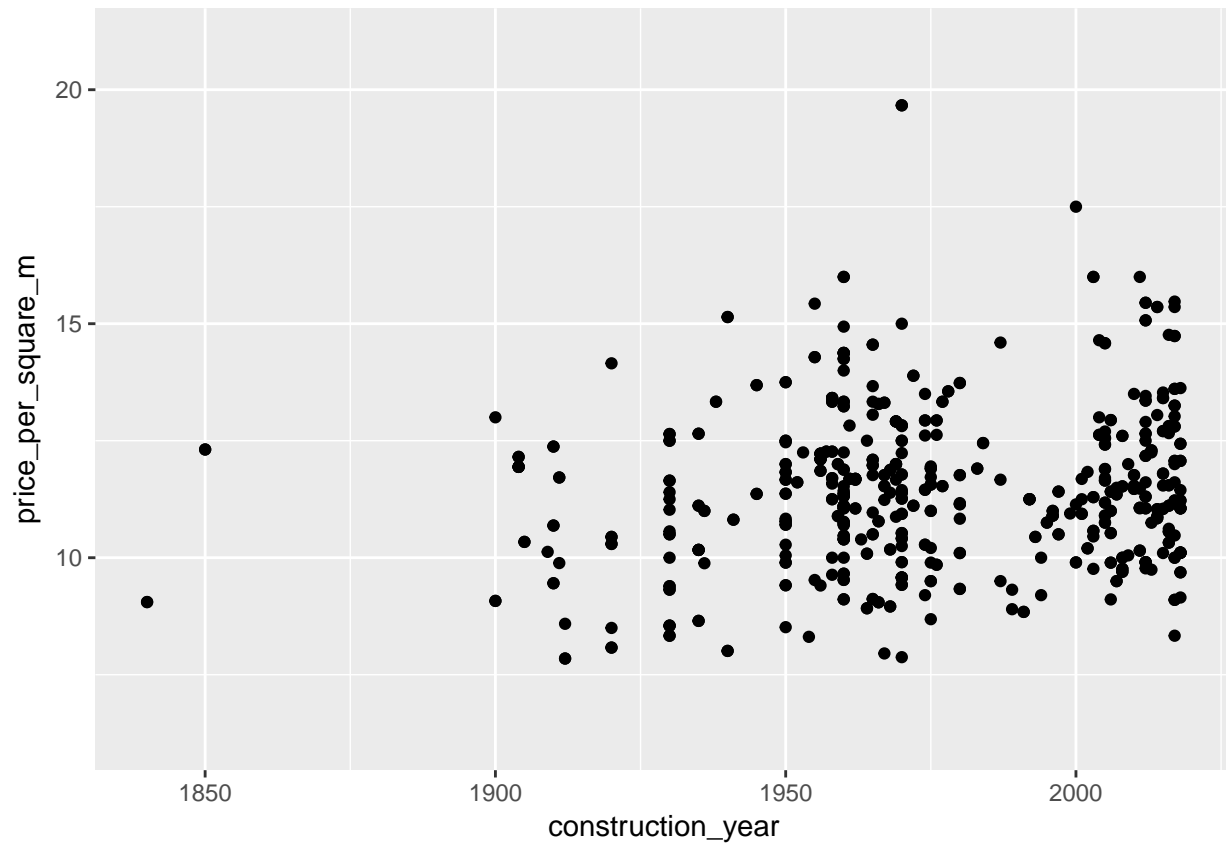
```
ggplot(data = tidy_listings,  
       aes(x = energy_consumption, y = price_per_square_m)) +  
  geom_point()
```



Energy consumption doesn't seem to have a large influence on the final price.

Construction year

```
ggplot(data = tidy_listings,  
       aes(x = construction_year, y = price_per_square_m)) +  
  geom_point()
```

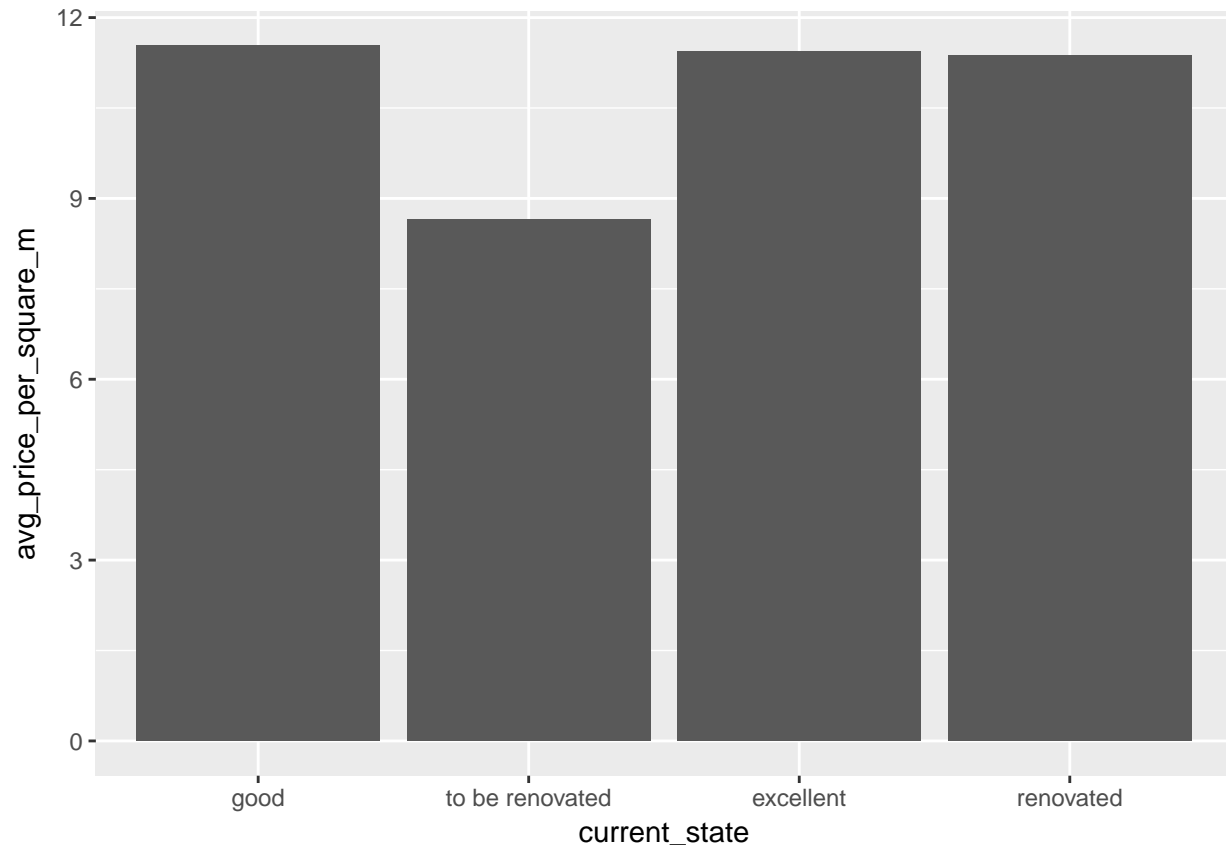


Construction year doesn't seem to have too much of an influence either.

Current state

```
df_influence_current_state <- tidy_listings %>%
  filter(!is.na(current_state)) %>%
  group_by(current_state) %>%
  summarise(avg_price_per_square_m = mean(price_per_square_m, na.rm = TRUE))

ggplot(data = df_influence_current_state,
       aes(x = current_state, y = avg_price_per_square_m)) +
  geom_col()
```



People don't seem to appreciate a great difference between apartments that are in a good state or in an excellent state. The only influence we notice is for apartments to be renovated, but there aren't enough of those to make any valid conclusions.

Municipality

```
df_influence_municipality <- tidy_listings %>%
  filter(!is.na(municipality)) %>%
  group_by(municipality) %>%
  summarise(avg_price_per_square_m = mean(price_per_square_m, na.rm = TRUE)) %>%
  arrange(avg_price_per_square_m) %>%
  mutate(municipality = factor(x = municipality, levels = municipality)) # so municipality is displayed in

ggplot(data = df_influence_municipality,
  aes(x = municipality, y = avg_price_per_square_m)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

