Projekt Zaliczeniowy 1

Michał Hiszpański, Nikola Kubiczek, Adrian Mirecki

2026-04-29

Data Fields

1. `city` - the name of the city where the property is located
2. `type` - type of the building
3. `squareMeters` - the size of the apartment in square meters
4. `rooms` - number of rooms in the apartment
5. `floor` / floorCount - the floor where the apartment is located and the total number of floors in the building
6. `buildYear` - the year when the building was built
7. `latitude`, `longitude` - geo coordinate of the property
8. `centreDistance` - distance from the city centre in km
9. `poiCount` - number of points of interest in 500m range from the apartment (schools, clinics, post offices, kindergartens, restaurants, colleges, pharmacies)
10. `[poiName]Distance` - distance to the nearest point of interest (schools, clinics, post offices, kindergartens, restaurants, colleges, pharmacies)
11. `ownership` - the type of property ownership
12. `condition` - the condition of the apartment
13. `has[features]` - whether the property has key features such as assigned parking space, balcony, elevator, security, storage room
14. `price` - offer price in Polish Zloty apartments_pl_YYYY_MM.csv: sale price apartments_rent_pl_YYYY_MM.csv: monthly rent

```
files <- list.files("rent", pattern = "\\.csv$", full.names = TRUE)
files <- sort(files)

df_list <- lapply(seq_along(files), function(i) {
  d <- read.csv(files[i])
  d$month_id <- i
  d
})

rent <- do.call(rbind, df_list)

has_cols <- grep("^has", names(rent), value = TRUE)

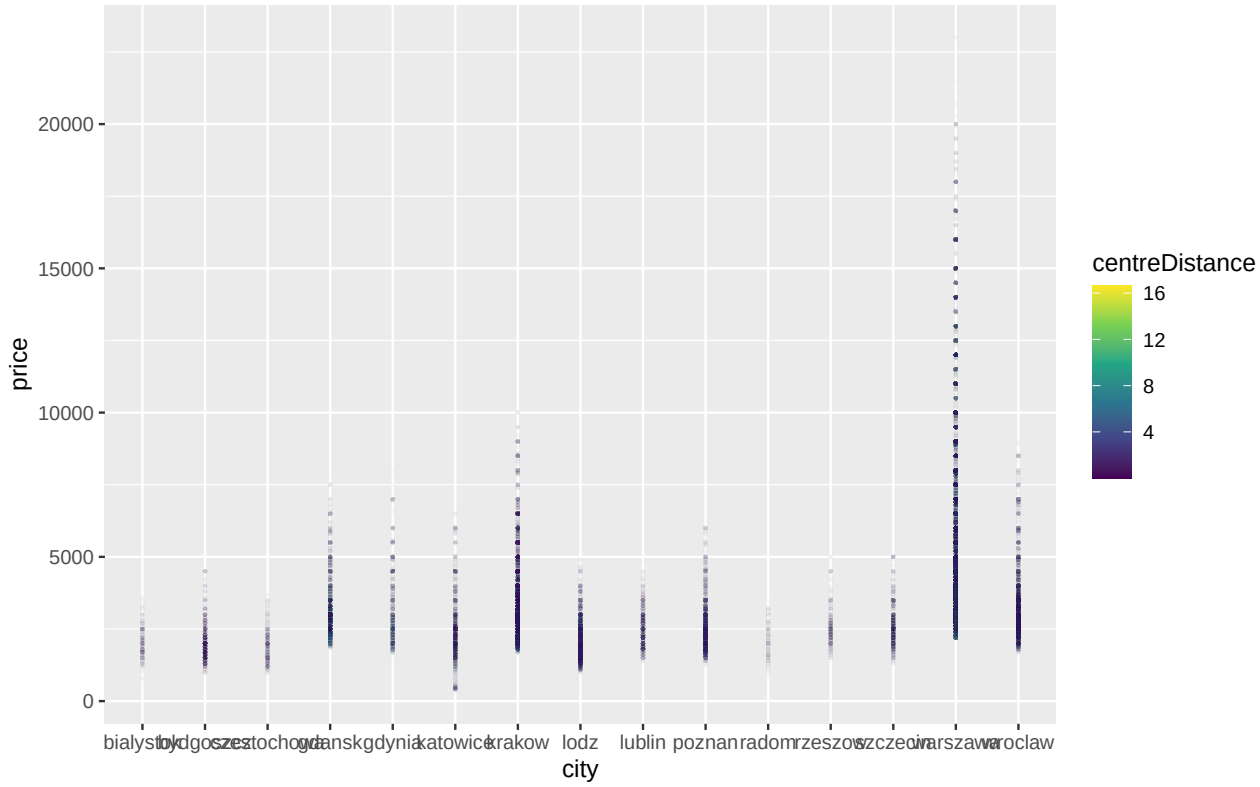
rent[has_cols] <- lapply(rent[has_cols], function(x) {
  tolower(x) == "yes"
})
summary(rent)
```

```
##       id               city               type            squareMeters
##  Length:70847       Length:70847       Length:70847       Min.   : 25.00
##  Class :character   Class :character   Class :character   1st Qu.: 40.00
##  Mode  :character   Mode  :character   Mode  :character   Median : 50.00
```

```
##                                                        Mean   : 55.25
##                                                        3rd Qu.: 64.00
##                                                        Max.   :150.00
##
##      rooms           floor         floorCount       buildYear
##  Min.   :1.000   Min.   : 1.00   Min.   : 1.000   Min.   :1850
##  1st Qu.:2.000   1st Qu.: 2.00   1st Qu.: 4.000   1st Qu.:1984
##  Median :2.000   Median : 3.00   Median : 5.000   Median :2010
##  Mean   :2.385   Mean   : 3.42   Mean   : 5.687   Mean   :1998
##  3rd Qu.:3.000   3rd Qu.: 4.00   3rd Qu.: 7.000   3rd Qu.:2020
##  Max.   :6.000   Max.   :30.00   Max.   :30.000   Max.   :2024
##                  NA's   :8613    NA's   :1379     NA's   :19682
##     latitude       longitude     centreDistance      poiCount
##  Min.   :49.98   Min.   :14.45   Min.   : 0.020   Min.   :  0.00
##  1st Qu.:51.08   1st Qu.:18.65   1st Qu.: 1.770   1st Qu.:  9.00
##  Median :52.19   Median :19.96   Median : 3.380   Median : 17.00
##  Mean   :51.79   Mean   :19.69   Mean   : 3.856   Mean   : 25.21
##  3rd Qu.:52.25   3rd Qu.:21.01   3rd Qu.: 5.420   3rd Qu.: 30.00
##  Max.   :54.58   Max.   :23.20   Max.   :16.620   Max.   :210.00
##
##  schoolDistance   clinicDistance   postOfficeDistance kindergartenDistance
##  Min.   :0.0020   Min.   :0.0010   Min.   :0.0010     Min.   :0.0010
##  1st Qu.:0.1740   1st Qu.:0.3020   1st Qu.:0.2350     1st Qu.:0.1530
##  Median :0.2850   Median :0.5500   Median :0.3880     Median :0.2590
##  Mean   :0.3688   Mean   :0.7763   Mean   :0.4702     Mean   :0.3272
##  3rd Qu.:0.4490   3rd Qu.:0.9700   3rd Qu.:0.5850     3rd Qu.:0.4010
##  Max.   :4.8560   Max.   :4.9900   Max.   :4.9390     Max.   :4.7510
##  NA's   :15       NA's   :50       NA's   :26         NA's   :43
##  restaurantDistance collegeDistance pharmacyDistance  ownership
##  Min.   :0.001      Min.   :0.004   Min.   :0.0010   Length:70847
##  1st Qu.:0.087      1st Qu.:0.520   1st Qu.:0.1360   Class :character
##  Median :0.180      Median :0.971   Median :0.2310   Mode  :character
##  Mean   :0.261      Mean   :1.287   Mean   :0.3089
##  3rd Qu.:0.329      3rd Qu.:1.809   3rd Qu.:0.3820
##  Max.   :4.961      Max.   :5.000   Max.   :4.9860
##  NA's   :154        NA's   :874     NA's   :80
##  buildingMaterial    condition         hasParkingSpace hasBalcony
##  Length:70847       Length:70847       Mode :logical   Mode :logical
##  Class :character   Class :character   FALSE:43075     FALSE:27341
##  Mode  :character   Mode  :character   TRUE :27772     TRUE :43506
##
##
##
##
##  hasElevator     hasSecurity     hasStorageRoom      price
##  Mode :logical   Mode :logical   Mode :logical   Min.   :  346
##  FALSE:26784     FALSE:60614     FALSE:59833     1st Qu.: 2500
##  TRUE :44063     TRUE :10233     TRUE :11014     Median : 3100
##                                                  Mean   : 3853
##                                                  3rd Qu.: 4490
##                                                  Max.   :23000
##
##     month_id
##  Min.   :1.000
```

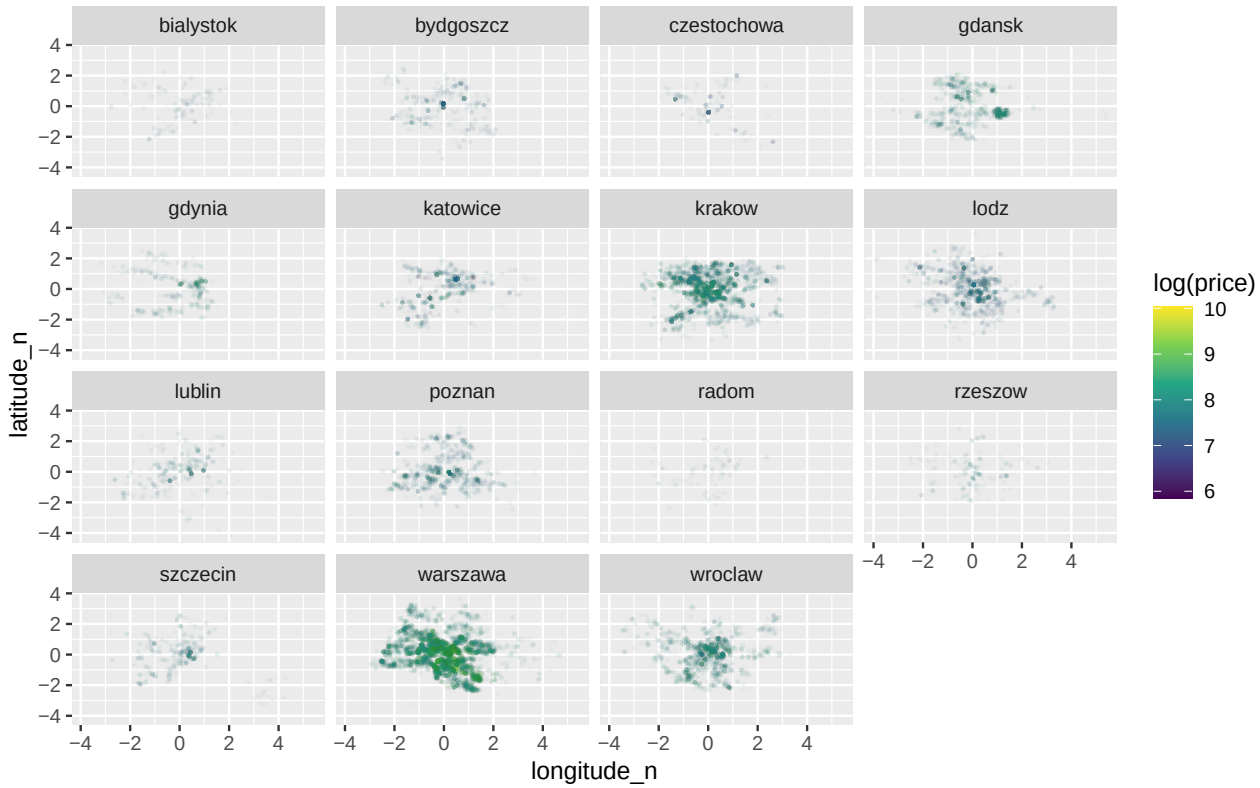
```
##  1st Qu.:3.000
##  Median :5.000
##  Mean   :4.556
##  3rd Qu.:7.000
##  Max.   :8.000
##
```

```
ggplot(rent) + geom_point(aes(x=city, y=price, col=centreDistance), alpha=0.01, size=0.2) + scale_color_
```

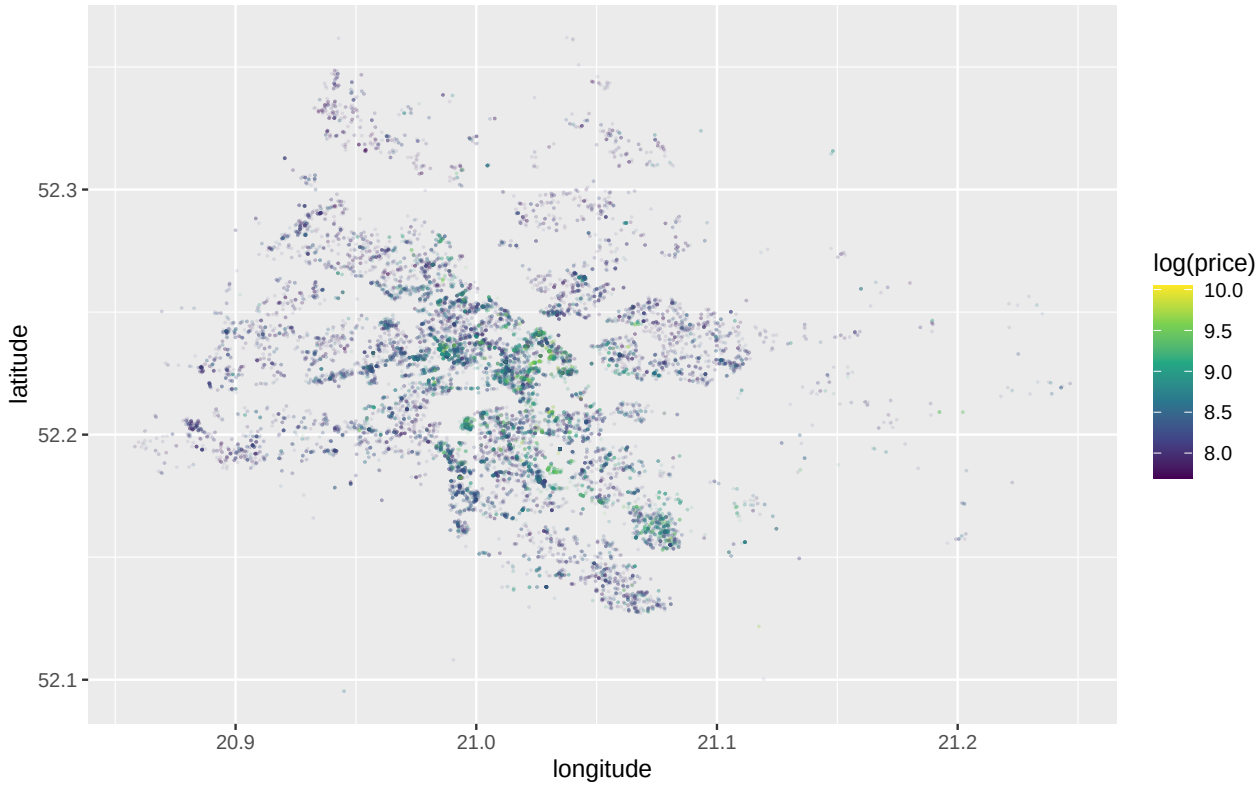



```
rent_norm <- rent %>%
  group_by(city) %>%
  mutate(
    latitude_n  = as.numeric(scale(latitude)),
    longitude_n = as.numeric(scale(longitude)),
    price_n = (price - min(price)) / (max(price) - min(price))
  ) %>%
  mutate() %>%
  ungroup()

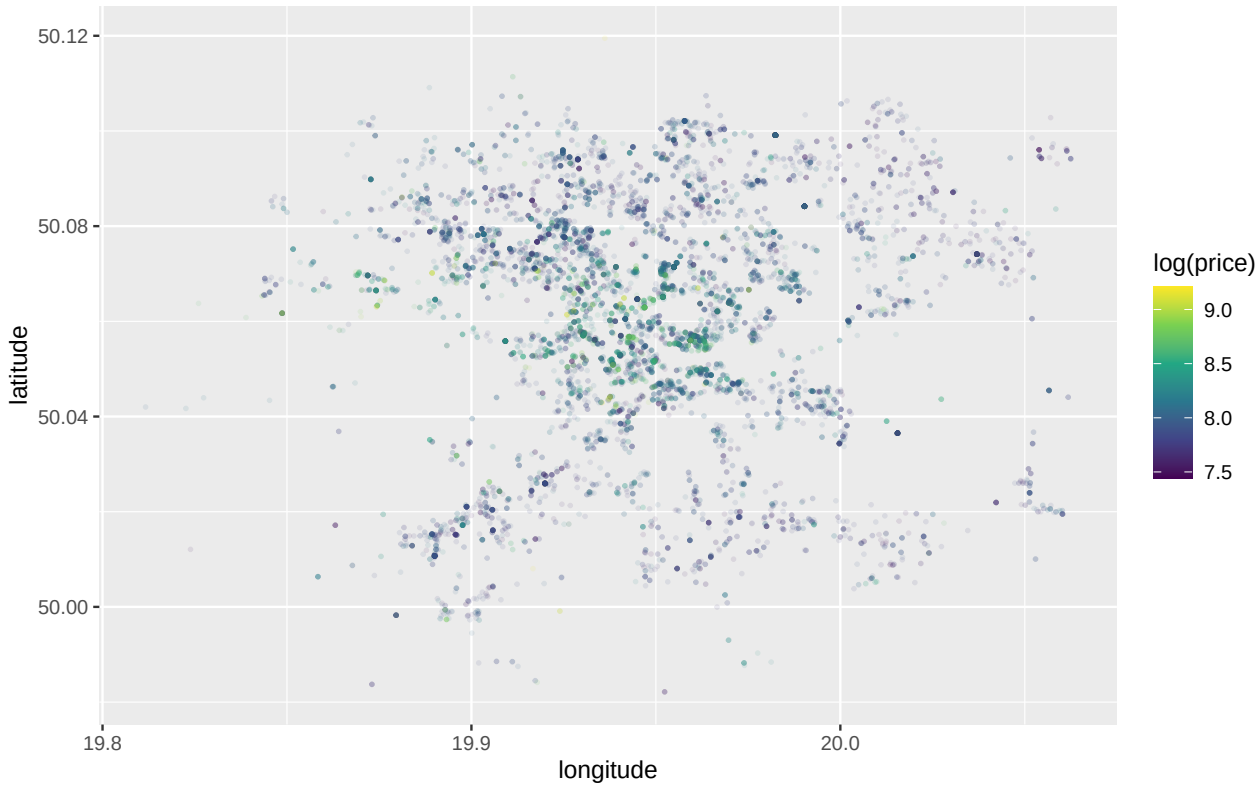
ggplot(rent_norm) + geom_point(aes(x=longitude_n, y=latitude_n, col=log(price)), alpha=0.01, size=0.2) +
```



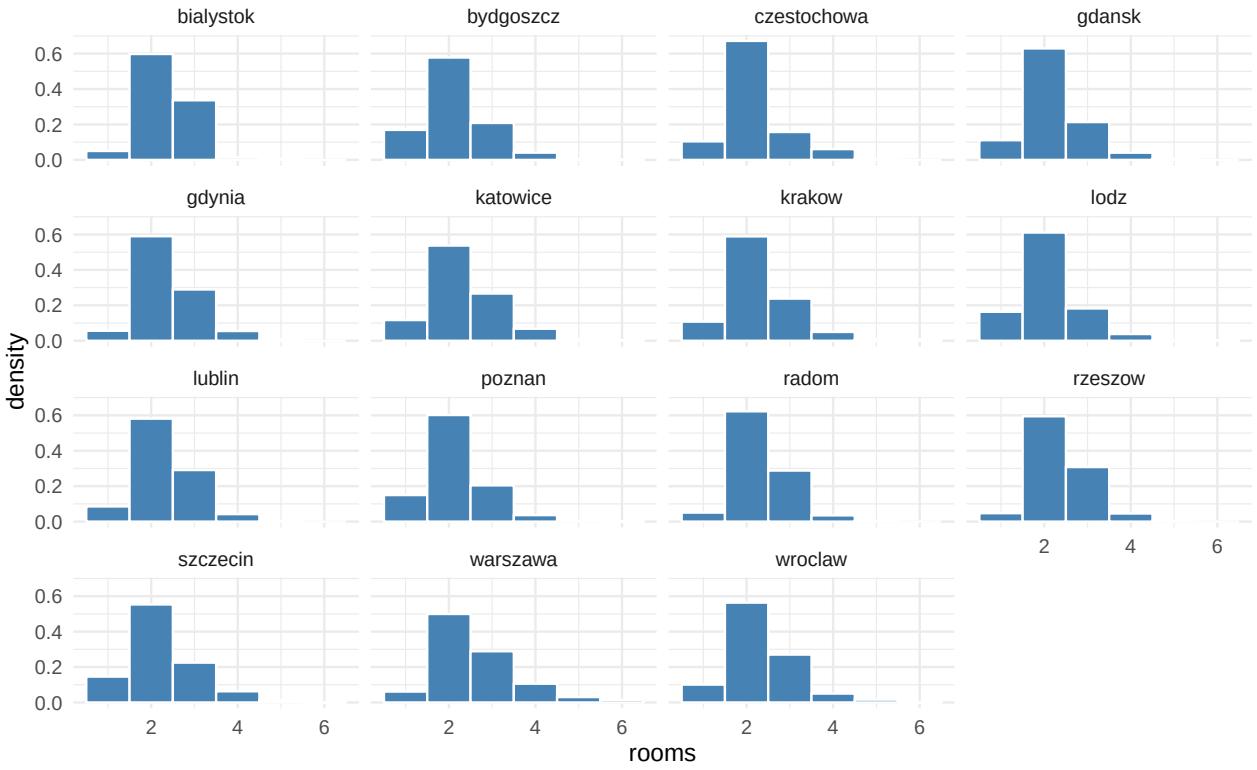

```
ggplot(rent[rent$city=="warszawa", ]) + geom_point(aes(x=longitude, y=latitude, col=log(price)), alpha=0
```



```
ggplot(rent[rent$city=="krakow", ]) + geom_point(aes(x=longitude, y=latitude, col=log(price)), alpha=0.
```




```
ggplot(rent, aes(x = rooms)) +
  geom_histogram(aes(y = after_stat(density)), binwidth = 1, fill = "steelblue", color = "white") +
  facet_wrap(~ city) +
  theme_minimal()
```


bialystok
bydgoszcz
czestochowa
gdansk
gdynia
katowice
krakow
lodz
lublin
poznan
radom
rzeszow
szczecin
warszawa
wroclaw
density
rooms
0.0
0.2
0.4
0.6
2
4
6