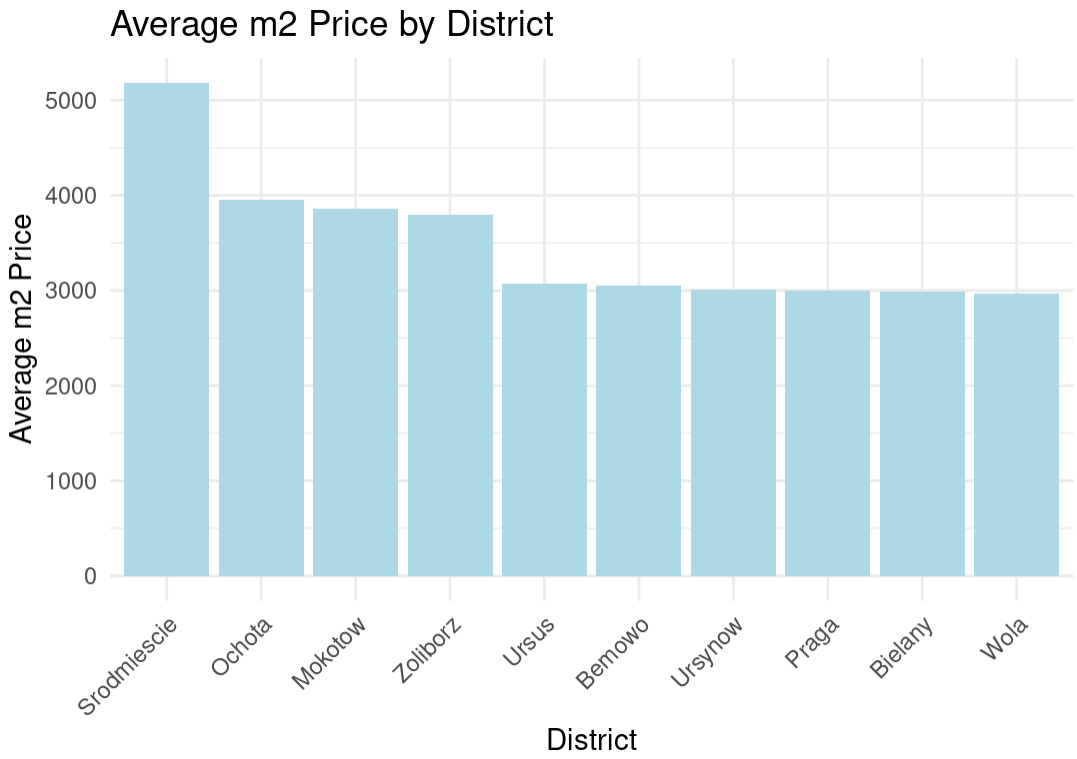
Data Visualization Assignment 2

Ezgi Altıntop
41623432038

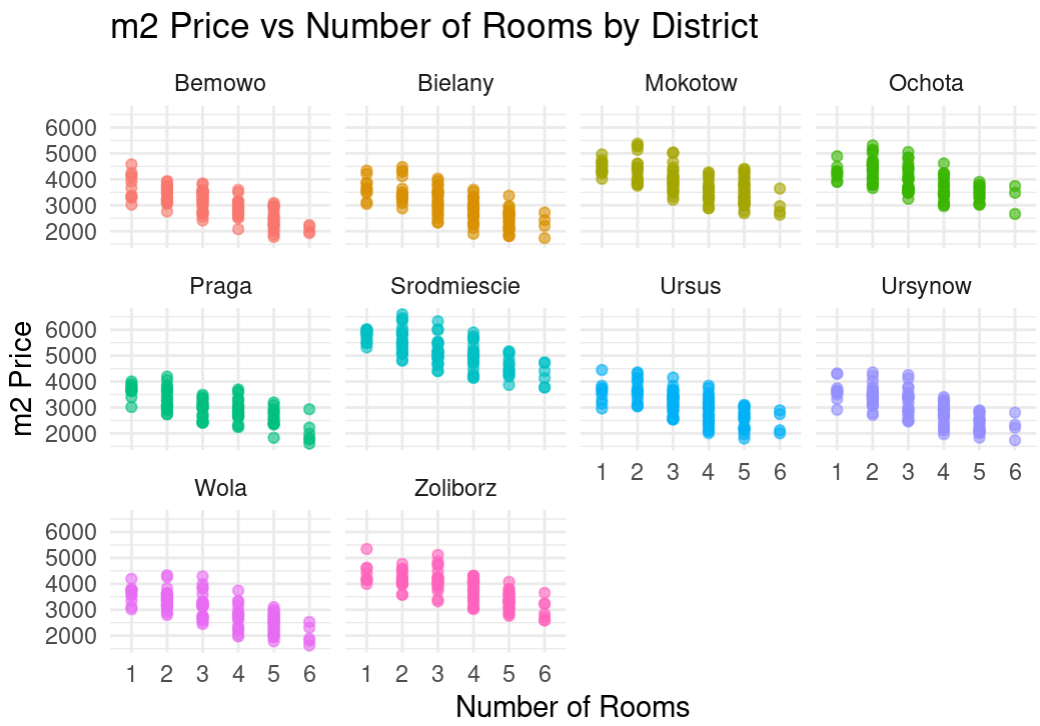
Case1: Using the apartments dataset in the DALEX package, perform visualizations that can answer the questions below:

-Visualize and interpret the average square meter (m2.price) prices of houses by districts. (10 + 10 points)



I compared the average m2 prices by districts in the graph. As a result, we can easily see that the district with the most expensive average m2 price is Srodmiescie by sorting from largest to smallest. When we consider all the districts we have, we can say that the average price is 3000.

-Investigate and visualize how the average square meter (m2.price) prices of houses vary according to the number of rooms (no.rooms) in the house by districts, and provide your interpretation. (15 + 15 points)

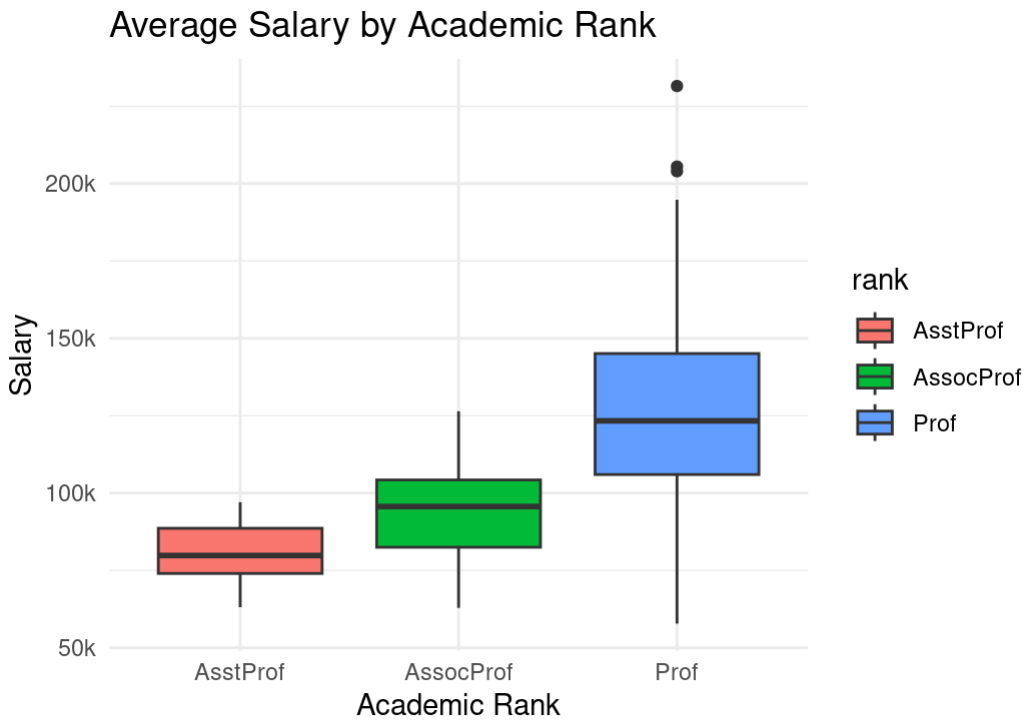


Each plot shows the variation of m2 price with the number of rooms for each district.

Here we can see how much the m2 price of each district is affected by the number of rooms. Generally, we can say that the distributions in all districts are similar. The prices of 1-2-3 room apartments are distributed higher than those of 4-5-6 rooms. We can see that the number of 6 room apartments is especially low and the m2 prices are cheaper. If you are looking for a 3 and 4 room apartment in the most expensive district, Srodmiescie, you can find a cheap or very expensive house since the distribution range is wide.

Case 2) Using the Salaries dataset in the carData package, perform visualizations that can answer the questions below:

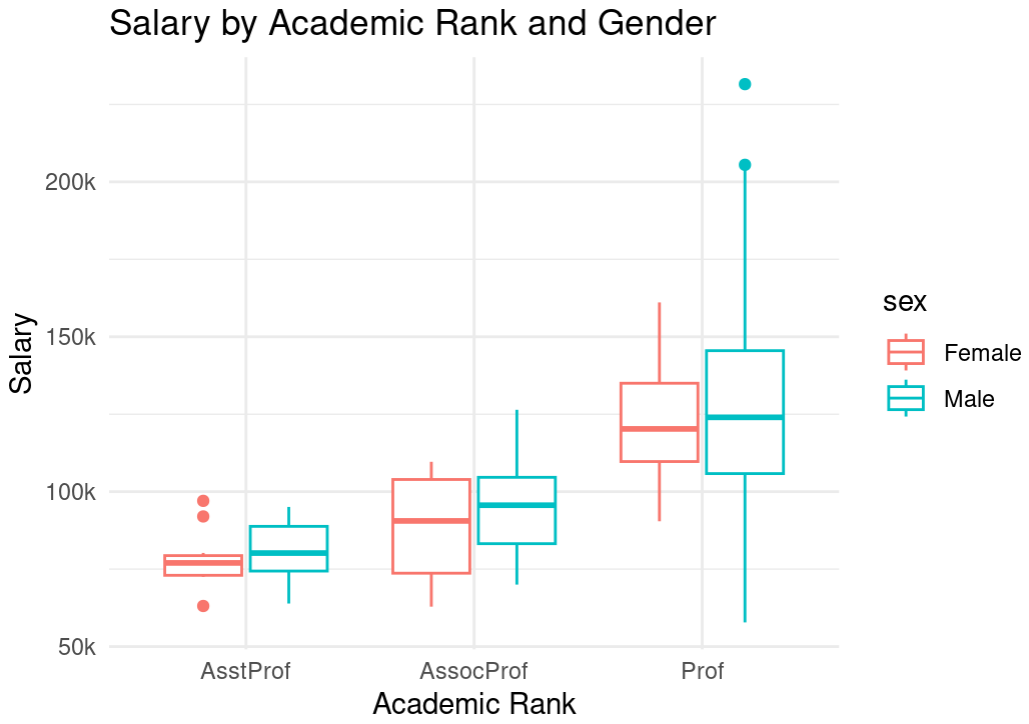
-Visualize and interpret the average salary according to academic rank (rank). (10 + 10 points)



Boxplot showing the distribution of salary by academic rank

We can see very clearly that the highest paid rank is Prof. Here I used a box plot so we can see the average salaries and how their ranges are. We can say that the ranges of AsstProfs are closer, that is, they receive similar salaries. When we look at the box plot of AssocProfs, we can understand that the salaries outside the quartiles and median are higher and lower from the length of the line. We said that the widest and longest box plot is Prof., we see that there are especially profs with outlines. It can be interpreted that these people may be world-famous profs.

-Investigate and visualize how the average salary varies by academic rank (rank) according to gender (sex), and provide your interpretation. (15 + 15 points)

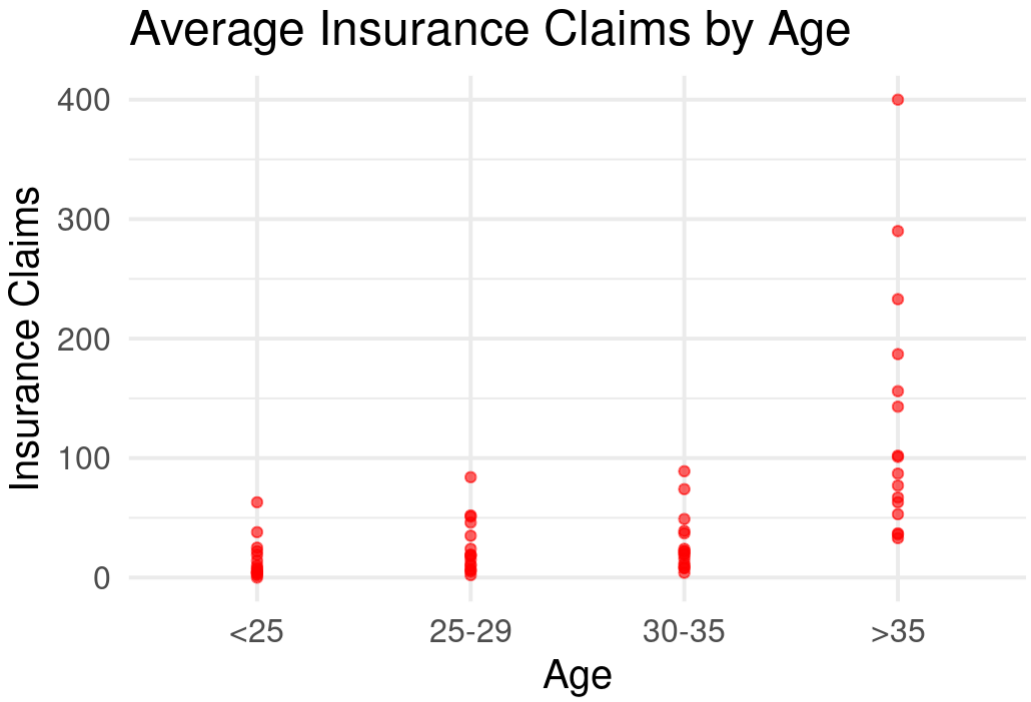


Boxplot showing the salary distribution by academic rank and gender

In these graphs, male gender is indicated in blue and female gender is indicated in red. As we see in the graphs, we can say that men receive more salaries than women in each rank distinction. We can say that this difference is gradually widening and that there is a greater difference in Profs.

Case 3) Using the Insurance dataset in the MASS package, perform visualizations that can answer the questions below:

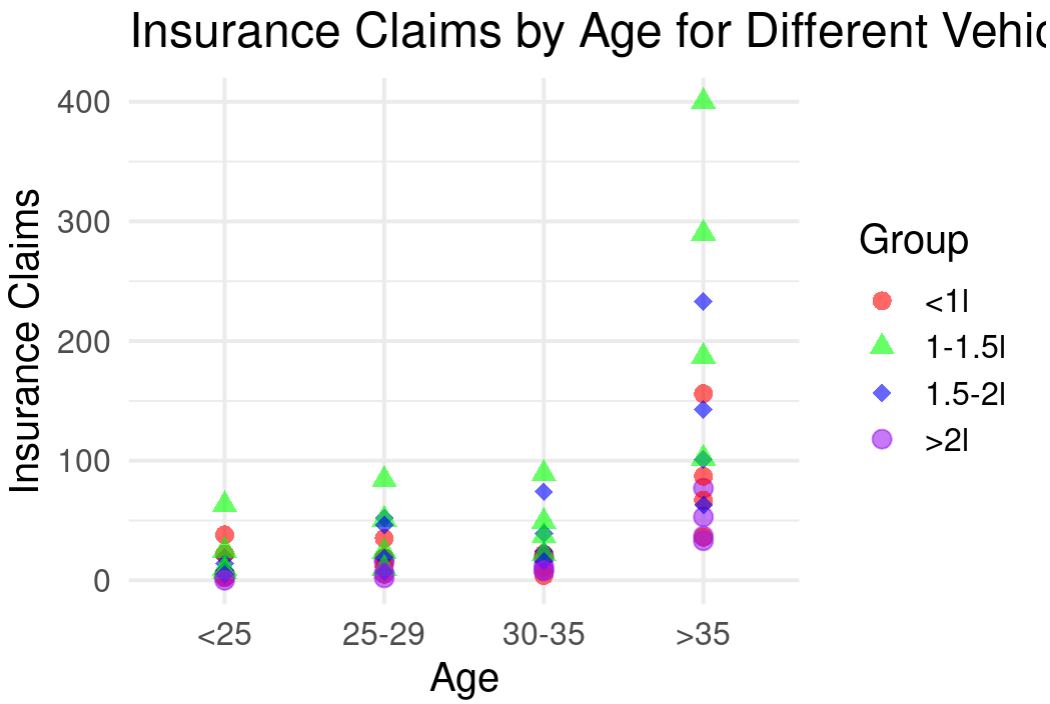
-Visualize and interpret the average number of policies (Claims) according to age (Age). (10 + 10 points)



relationship between age and insurance claims with a smooth trend line.

In this graph, I compared the number of policies according to age groups and we can say that those over 35 have the most claims. For example, we can say that those under 25 have similar numbers of claims and that their average is less than 60.

-Investigate and visualize how the average number of policies (Claims) varies by age (Age) according to the vehicle engine size (Group), and provide your interpretation. (15 + 15 points)



on, and colors/shapes represent different vehicle groups.

Here we can see the preferences of age groups according to engine size. As can be seen, younger ages preferred the group over 2 and we can say that those over 35 preferred the group 1-1.5 more.