Exam Template

Student

7/12/2020

3-Visual story telling part 1: green buildings

The case Over the past decade, both investors and the general public have paid increasingly close attention to the benefits of environmentally conscious buildings. There are both ethical and economic forces at work here. In commercial real estate, issues of eco-friendliness are intimately tied up with ordinary decisions about how to allocate capital. In this context, the decision to invest in eco-friendly buildings could pay off in at least four ways.

Every building has the obvious list of recurring costs: water, climate control, lighting, waste disposal, and so forth. Almost by definition, these costs are lower in green buildings. Green buildings are often associated with better indoor environments—the kind that are full of sunlight, natural materials, and various other humane touches. Such environments, in turn, might result in higher employee productivity and lower absenteeism, and might therefore be more coveted by potential tenants. The financial impact of this factor, however, is rather hard to quantify ex ante; you cannot simply ask an engineer in the same way that you could ask a question such as, "How much are these solar panels likely to save on the power bill?" Green buildings make for good PR. They send a signal about social responsibility and ecological awareness, and might therefore command a premium from potential tenants who want their customers to associate them with these values. It is widely believed that a good corporate image may enable a firm to charge premium prices, to hire better talent, and to attract socially conscious investors. Finally, sustainable buildings might have longer economically valuable lives. For one thing, they are expected to last longer, in a direct physical sense. (One of the core concepts of the green-building movement is "life-cycle analysis," which accounts for the high front-end environmental impact of ac- quiring materials and constructing a new building in the first place.) Moreover, green buildings may also be less susceptible to market risk—in particular, the risk that energy prices will spike, driving away tenants into the arms of bolder, greener investors. Of course, much of this is mere conjecture. At the end of the day, tenants may or may not be willing to pay a premium for rental space in green buildings. We can only find out by carefully examining data on the commercial real-estate market.

The file greenbuildings.csv contains data on 7,894 commercial rental properties from across the United States. Of these, 685 properties have been awarded either LEED or EnergyStar certification as a green building. You can easily find out more about these rating systems on the web, e.g. at www.usgbc.org. The basic idea is that a commercial property can receive a green certification if its energy efficiency, carbon footprint, site selection, and building materials meet certain environmental benchmarks, as certified by outside engineers.

A group of real estate economists constructed the data in the following way. Of the 1,360 green-certified buildings listed as of December 2007 on the LEED or EnergyStar websites, current information about building characteristics and monthly rents were available for 685 of them. In order to provide a control population, each of these 685 buildings was matched to a cluster of nearby commercial buildings in the CoStar database. Each small cluster contains one green-certified building, and all non-rated buildings within a quarter-mile radius of the certified building. On average, each of the 685 clusters contains roughly 12 buildings, for a total of 7,894 data points.

The columns of the data set are coded as follows:

CS.PropertyID: the building's unique identifier in the CoStar database. cluster: an identifier for the building cluster, with each cluster containing one green-certified building and at least one other non-green-certified building within a quarter-mile radius of the cluster center. size: the total square footage of available rental space in the building. empl.gr: the year-on-year growth rate in employment in the building's geographic region. Rent: the rent charged to tenants in the building, in dollars per square foot per calendar year. leasing.rate: a measure of occupancy; the fraction of the building's available space currently under lease. stories: the height of the building in stories. age: the age of the building in years. renovated: whether the building has undergone substantial renovations during its lifetime. class.a, class.b: indicators for two classes of building quality (the third is Class C). These are relative classifications within a specific market. Class A buildings are generally the highest-quality properties in a given market. Class B buildings are a notch down, but still of reasonable quality. Class C buildings are the least desirable properties in a given market. green.rating: an indicator for whether the building is either LEED- or EnergyStar-certified. LEED, Energystar: indicators for the two specific kinds of green certifications. net: an indicator as to whether the rent is quoted on a "net contract" basis. Tenants with net-rental contracts pay their own utility costs, which are otherwise included in the quoted rental price. amenities: an indicator of whether at least one of the following amenities is available on-site: bank, convenience store, dry cleaner, restaurant, retail shops, fitness center. cd.total.07: number of cooling degree days in the building's region in 2007. A degree day is a measure of demand for energy; higher values mean greater demand. Cooling degree days are measured relative to a baseline outdoor temperature, below which a building needs no cooling. hd.total07: number of heating degree days in the building's region in 2007. Heating degree days are also measured relative to a baseline outdoor temperature, above which a building needs no heating. total.dd.07: the total number of degree days (either heating or cooling) in the building's region in 2007. Precipitation: annual precipitation in inches in the building's geographic region. Gas.Costs: a measure of how much natural gas costs in the building's geographic region. Electricity.Costs: a measure of how much electricity costs in the building's geographic region. cluster.rent: a measure of average rent per square-foot per calendar year in the building's local market. The goal An Austin real-estate developer is interested in the possible economic impact of "going green" in her latest project: a new 15-story mixed-use building on East Cesar Chavez, just across I-35 from downtown. Will investing in a green building be worth it, from an economic perspective? The baseline construction costs are $100 million, with a 5% expected premium for green certification.

The developer has had someone on her staff, who's been described to her as a "total Excel guru from his undergrad statistics course," run some numbers on this data set and make a preliminary recommendation. Here's how this person described his process.

I began by cleaning the data a little bit. In particular, I noticed that a handful of the buildings in the data set had very low occupancy rates (less than 10% of available space occupied). I decided to remove these buildings from consideration, on the theory that these buildings might have something weird going on with them, and could potentially distort the analysis. Once I scrubbed these low-occupancy buildings from the data set, I looked at the green buildings and non-green buildings separately. The median market rent in the non-green buildings was 25 per square foot per year, while the median market rent in the green buildings was 27.60 per square foot per year: about 2.60 more per square foot. (I used the median rather than the mean, because there were still some outliers in the data, and the median is a lot more robust to outliers.) Because our building would be 250,000 square feet, this would translate into an additional 250000 x 2.6 = 650000 of extra revenue per year if we build the green building.

Our expected baseline construction costs are 100 million, with a 5% expected premium for green certification. Thus we should expect to spend an extra 5 million on the green building. Based on the extra revenue we would make, we would recuperate these costs in 5000000/650000 = 7.7 years. Even if our occupancy rate were only 90%, we would still recuperate the costs in a little over 8 years. Thus from year 9 onwards, we would be making an extra $650,000 per year in profit. Since the building will be earning rents for 30 years or more, it seems like a good financial move to build the green building.

The developer listened to this recommendation, understood the analysis, and still felt unconvinced. She has therefore asked you to revisit the report, so that she can get a second opinion.

Do you agree with the conclusions of her on-staff stats guru? If so, point to evidence supporting his case. If not, explain specifically where and why the analysis goes wrong, and how it can be improved. Do you see the possibility of confounding variables for the relationship between rent and green status? If so, provide evidence for confounding, and see if you can also make a picture that visually shows how we might "adjust" for such a confounder. Tell your story in pictures, with appropriate introductory and supporting text.

Note: this is intended as an exercise in visual and numerical story-telling. Your approach should rely on pictures and/or tables, not a regression model. Tell a story understandable to a non-technical audience. Keep it concise.

Firstly, to fact check the Excel guru, I will calculate median rent for green buildings:

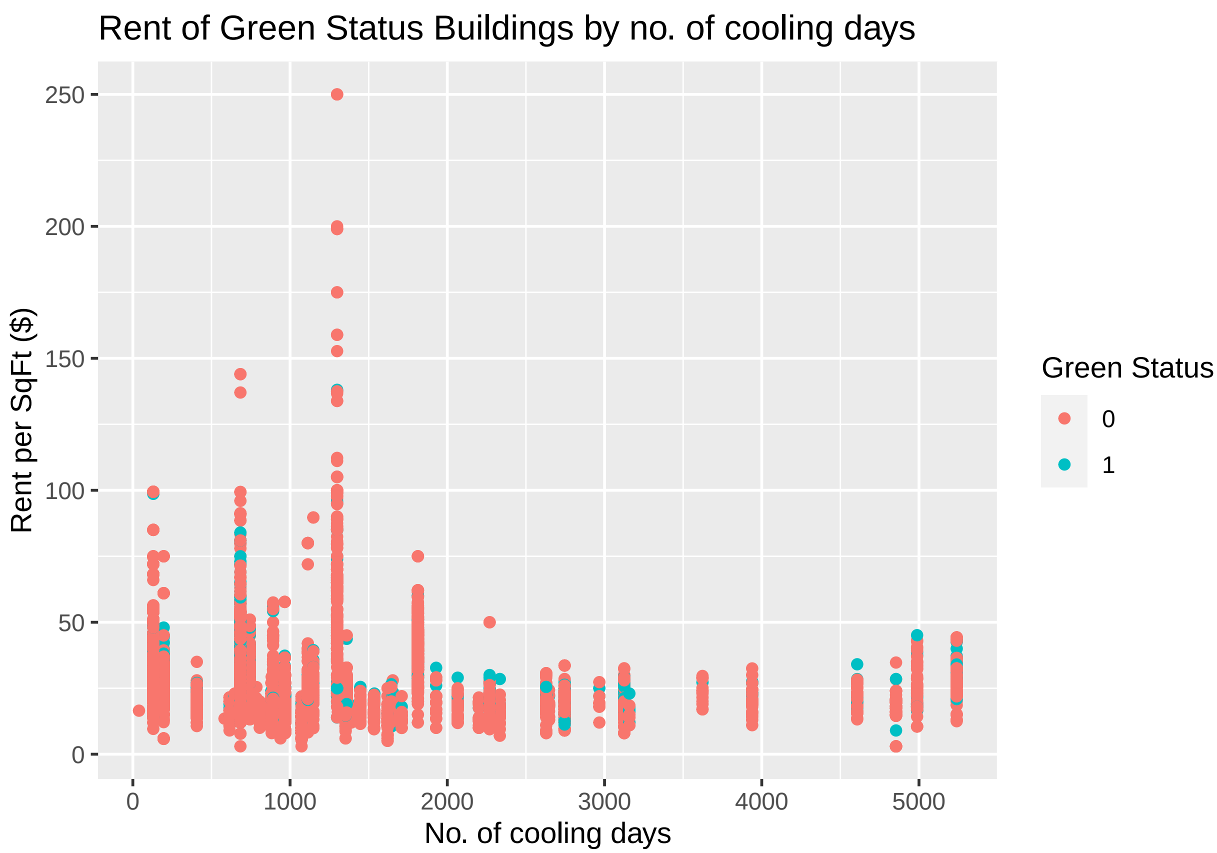
```
## [1] 27.6
```

```
## [1] 25
```

Turns out the guru calculated his median correctly

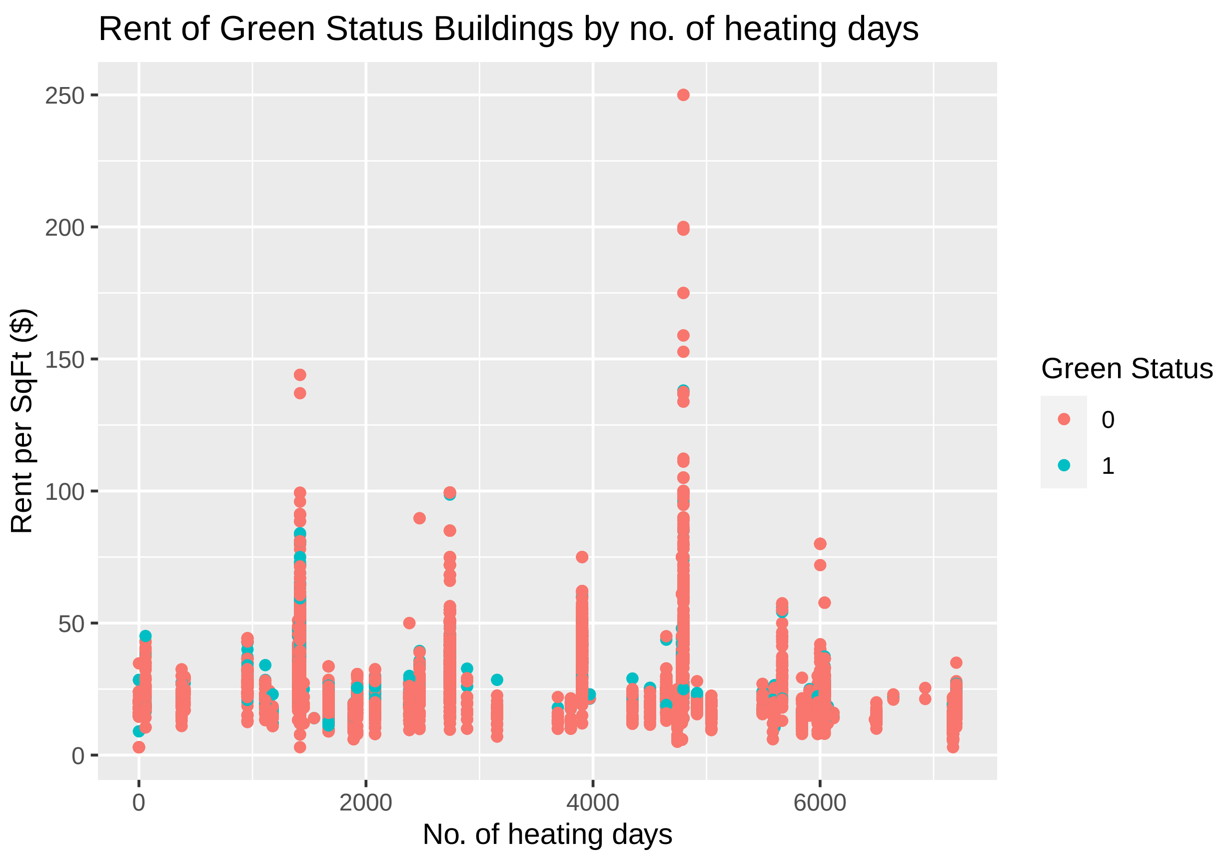
But in his analysis, I feel that he should've considered other variables before giving a conclusion. I shall firstly find out variables which are confounding with Rent by checking whether they are correlated to Rent or not using scatter plots

#First approach - check for correlation between cooling days, heating days, degree days

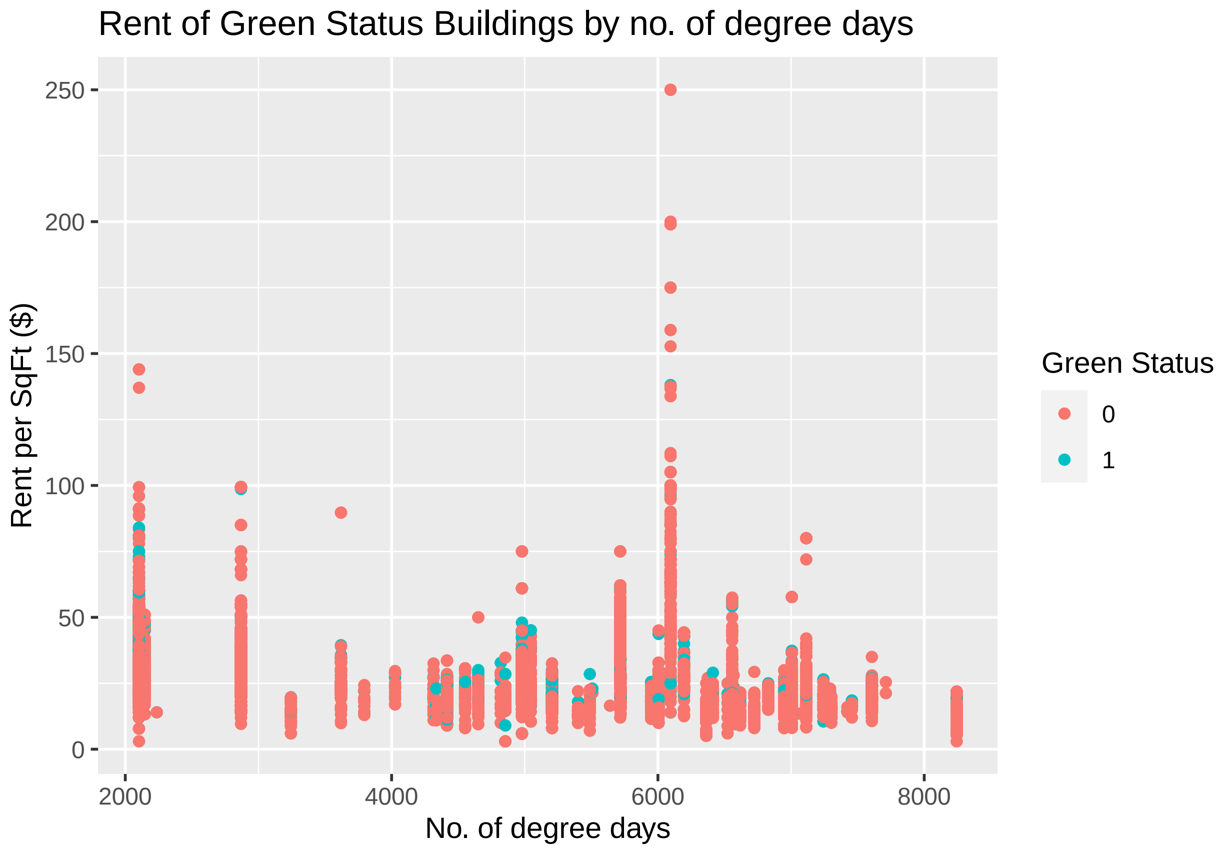


* green buildings have a higher rent across all kinds of cooling days (barring 1000-2000 days),so we can conclude that no of cooling days does not contribute to the rent of the building

- No correlation observed between number of cooling days and rent



From the graph we see that green_rated buildings charge a higher rent when: * The no of heating degree days are high, implying that there is a need for heating on most days. This implies that the savings in energy costs are higher for a green building, thereby having a higher rent * No Correlation observed between number of heating days and Rent

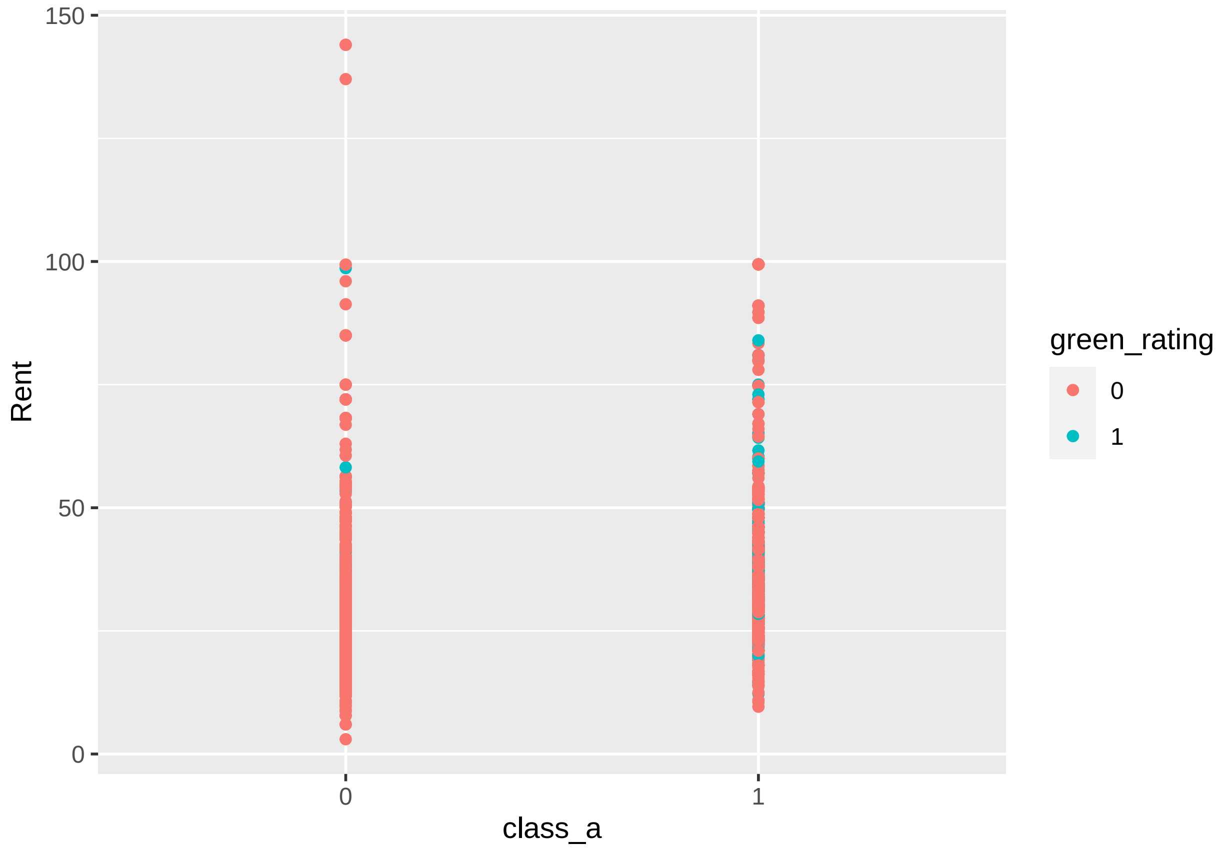


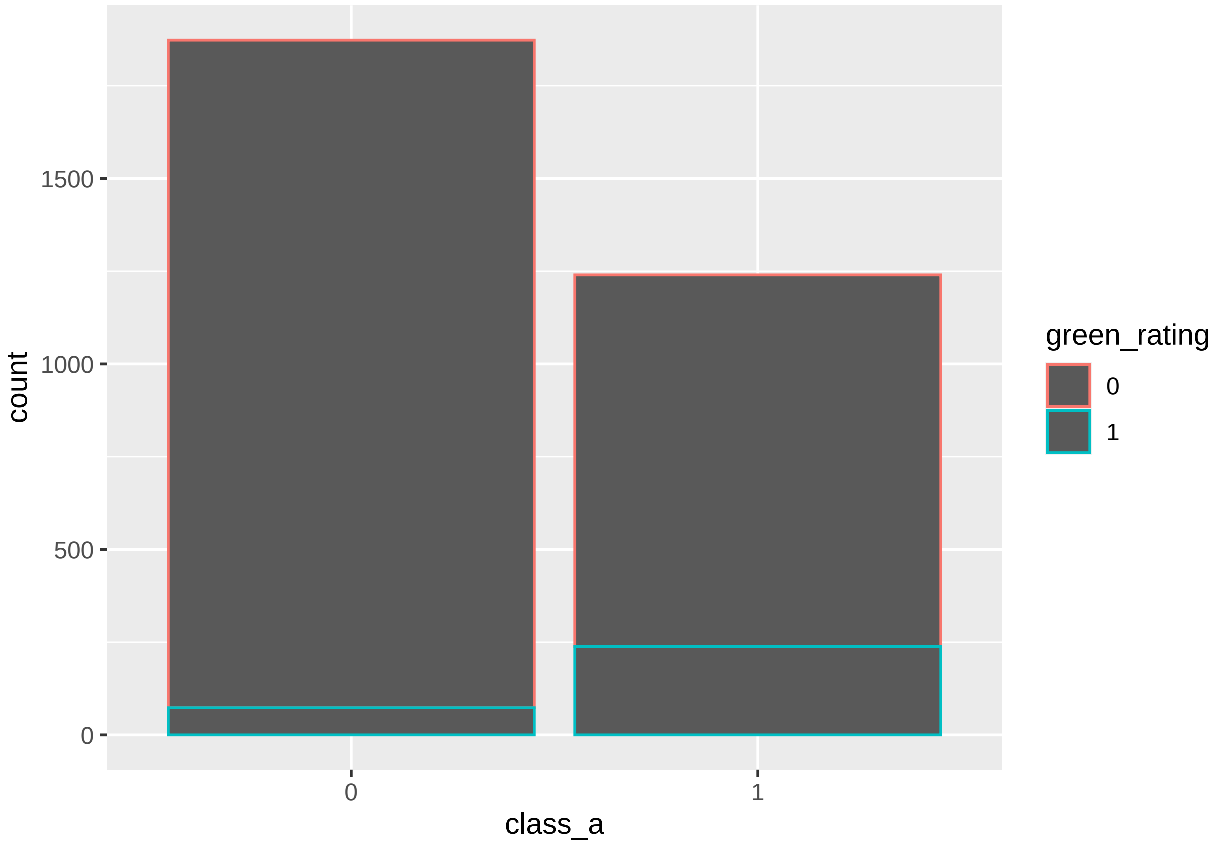
- We can now confidently say that places that have more than 4000q degree days (extreme temperatures), green buildings charge a higher rent. One possible reason for higher rent could be higher savings in energy costs.

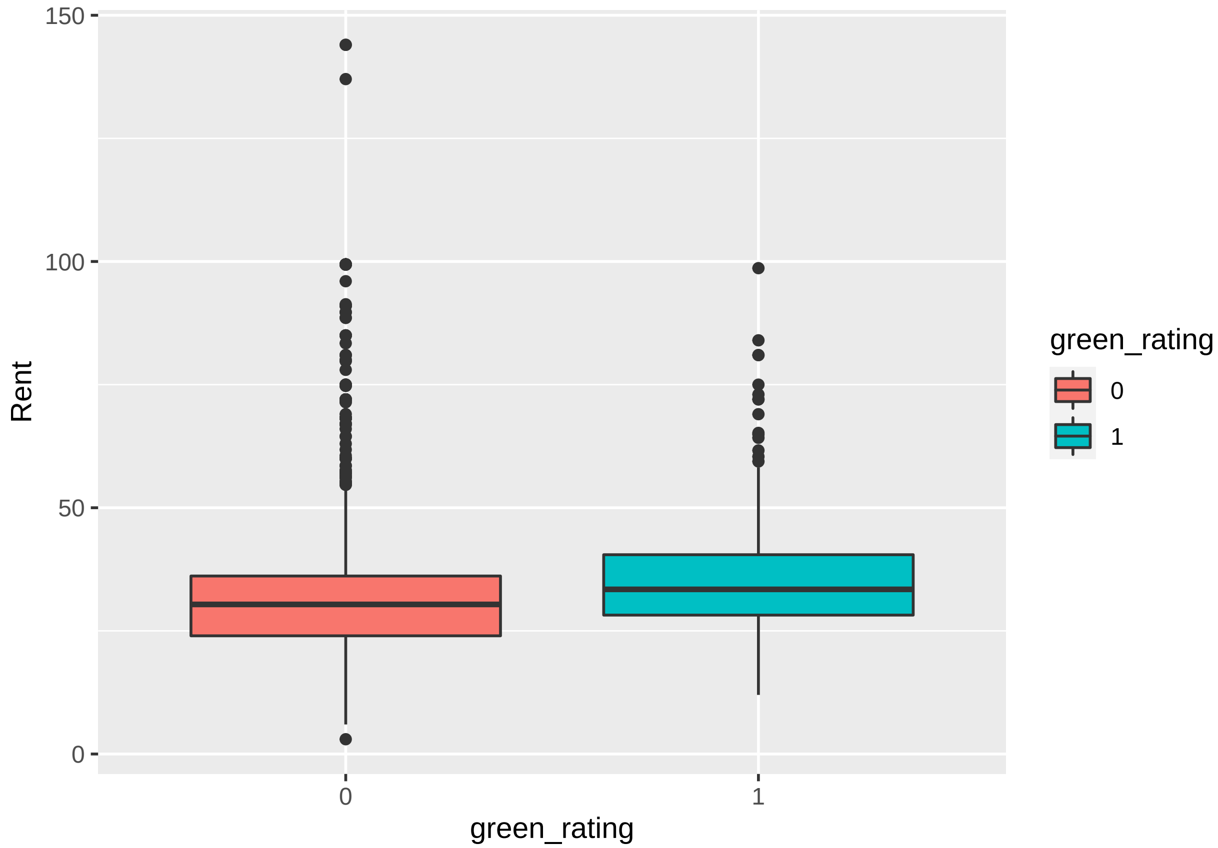
We will now hold the degree days constant and check if it is a confounding variable for Rent for different intervals of degree days.

It is possible that because buildings with degree days > 2000 are better built, and hence charge a premium rent

holding degree days < 4000 constant

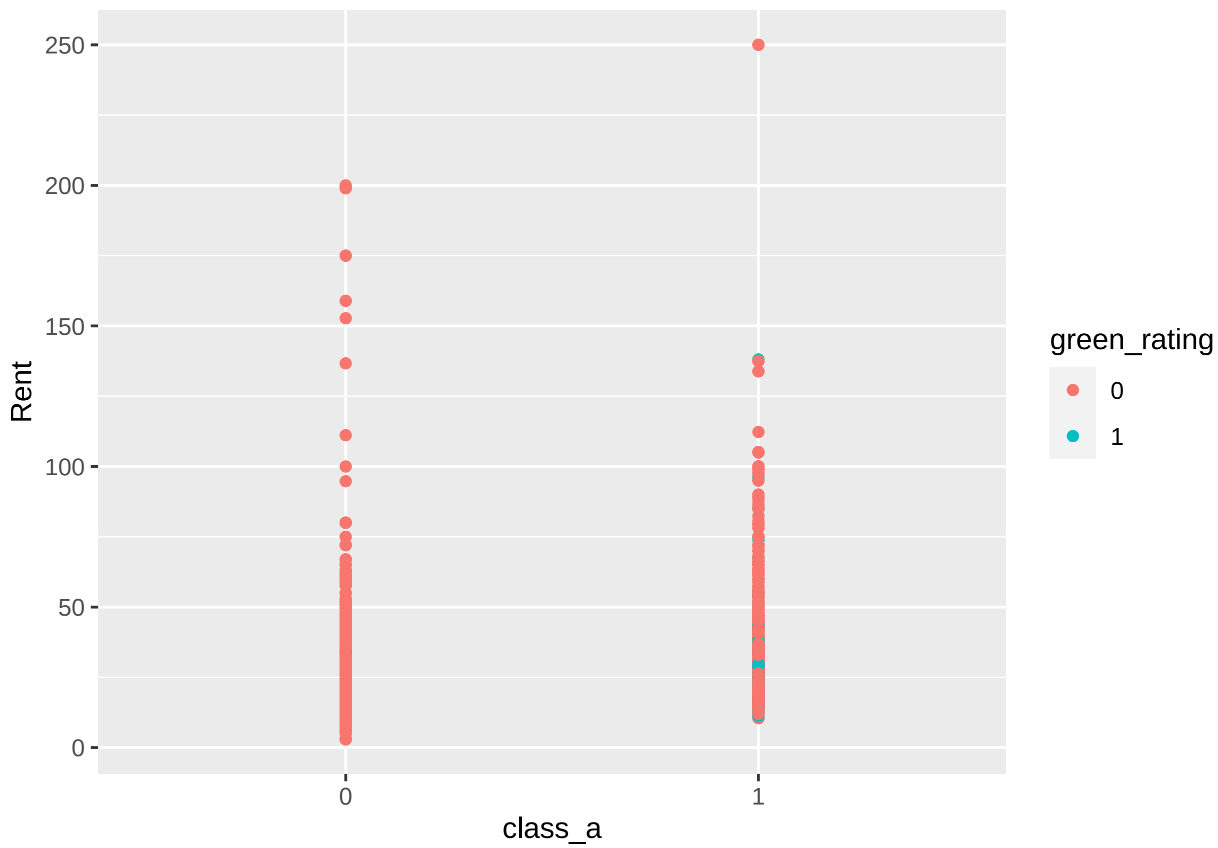

150
100
50
0
Rent
0
1
class_a
green_rating
0
1


1500
1000
500
0
count
0
1
class_a
green_rating
0
1

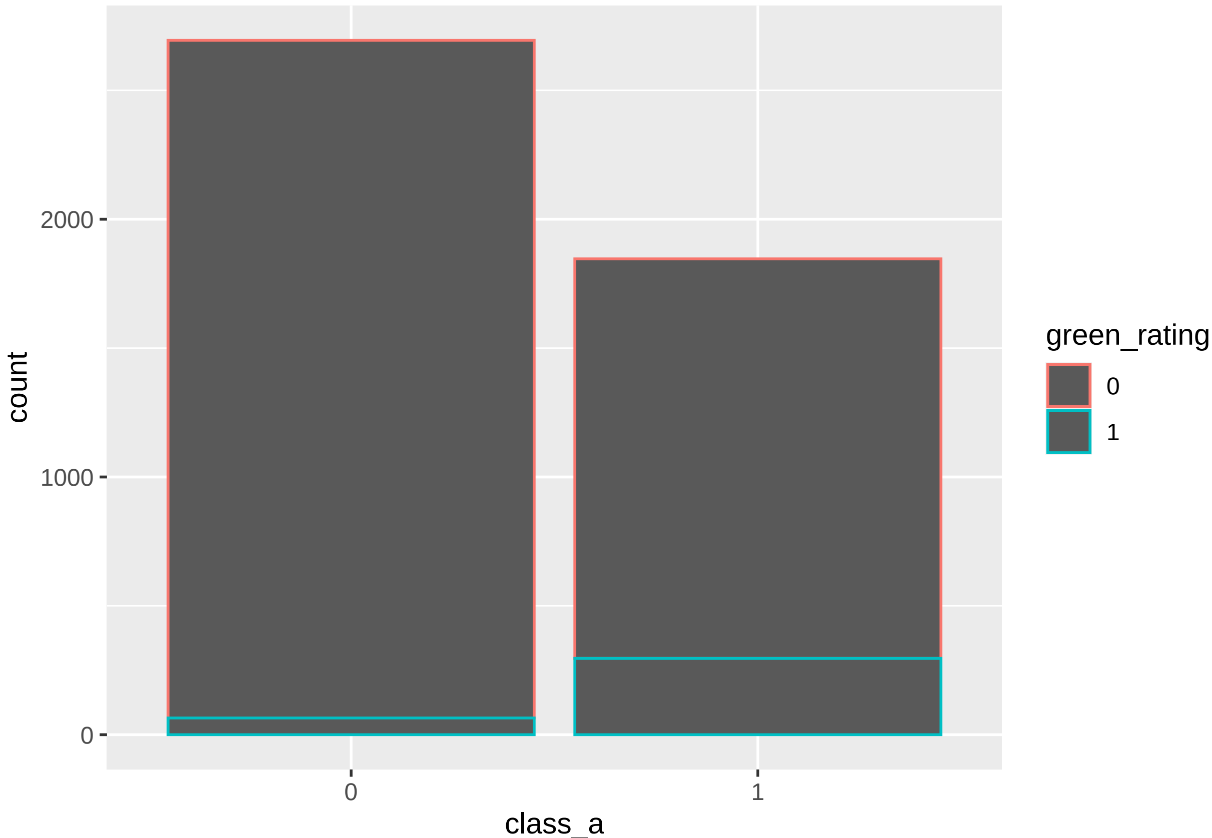


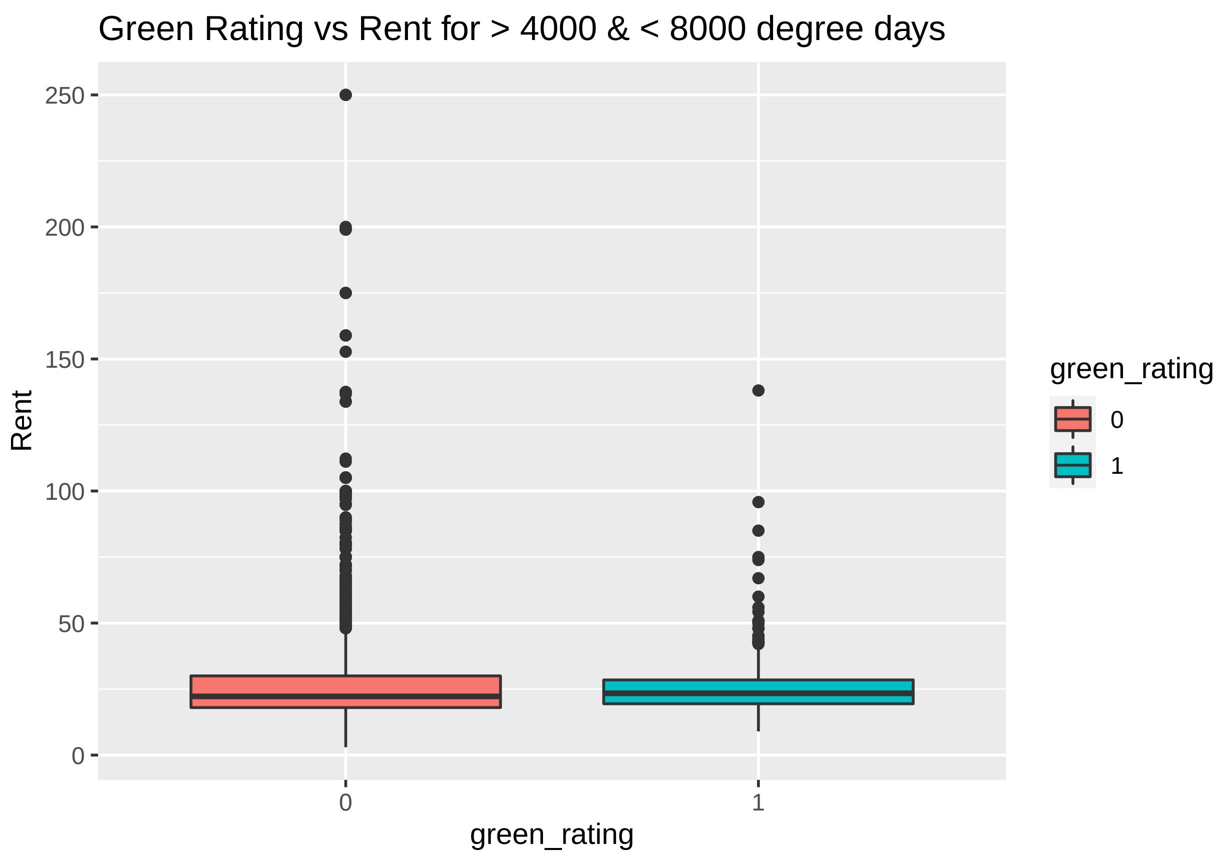
This is telling us that there is almost equal rent associated with a green building and non green building, if it is class_a and in an area with degree days < 4000

now we will try for >4000 and <8000

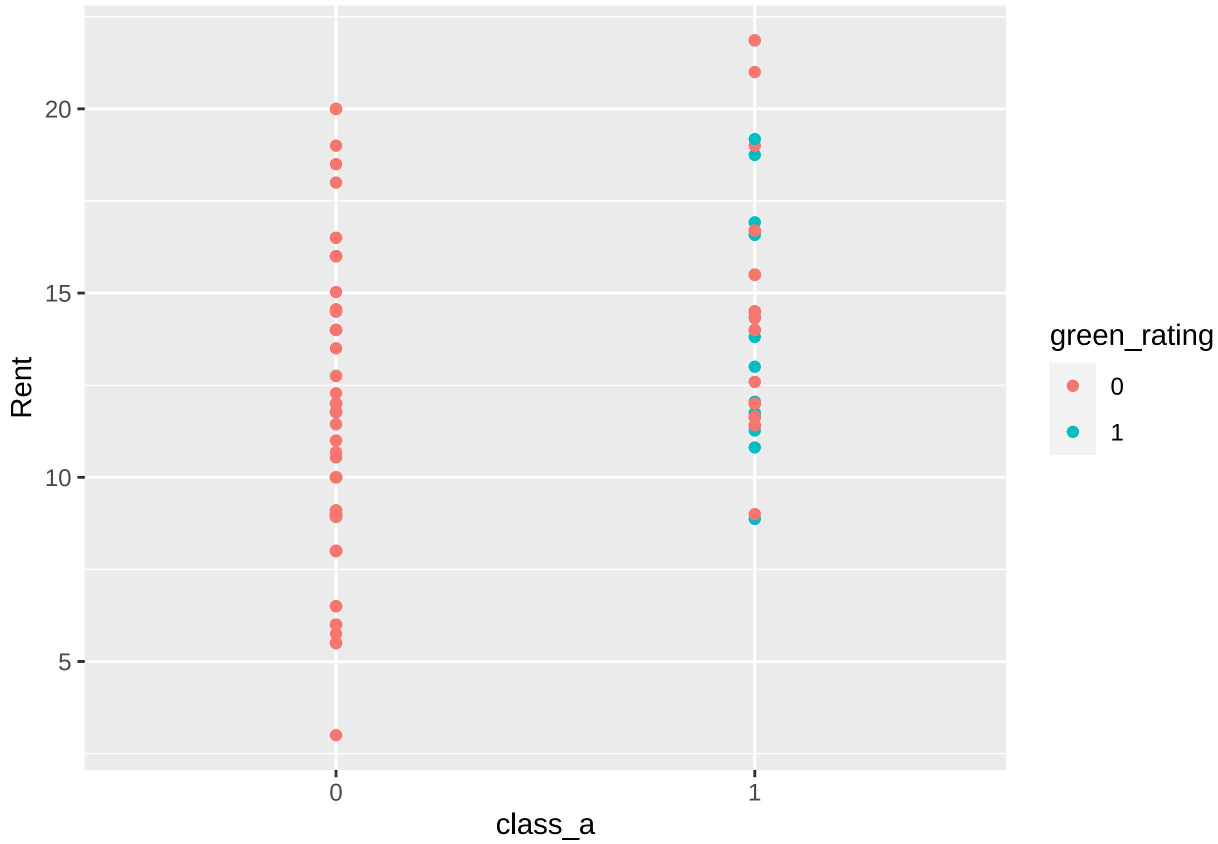


250
200
150
100
50
0
Rent
0
1
class_a
green_rating
0
1

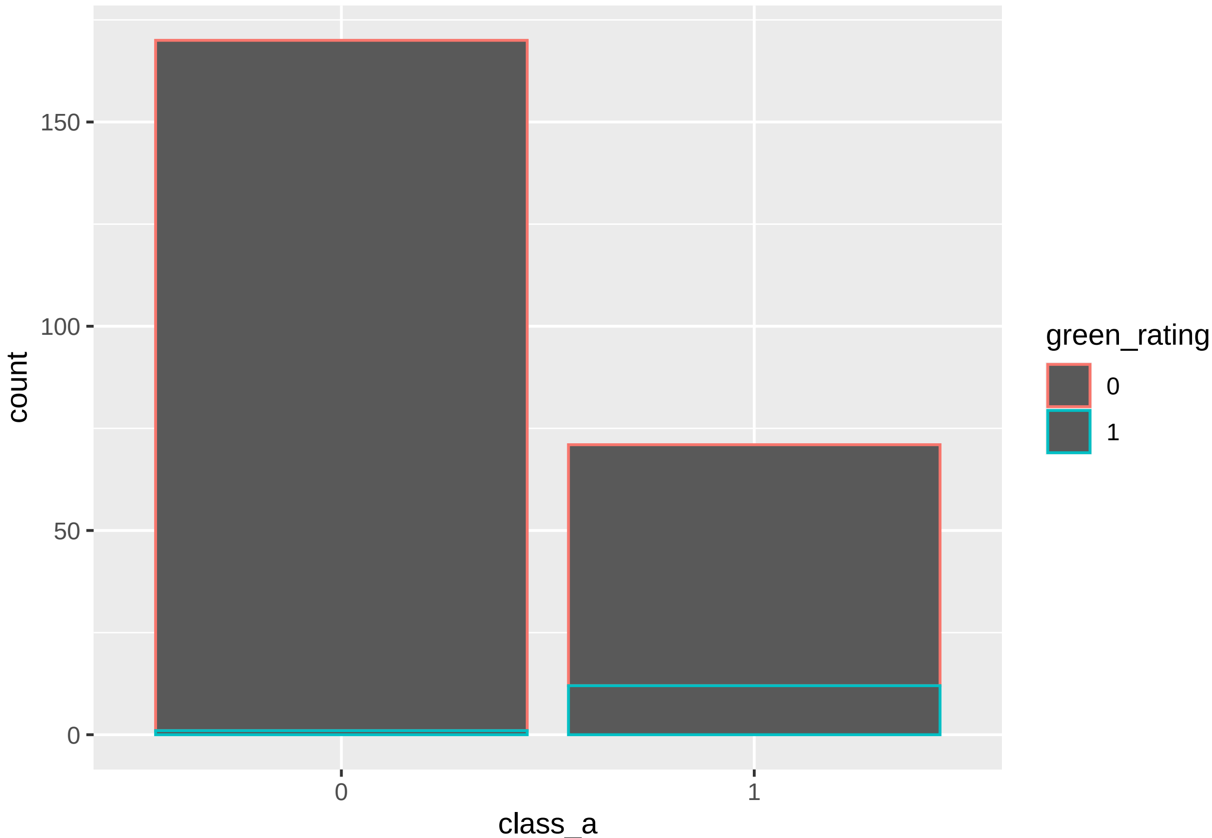

count
2000
1000
0
0
1
class_a
green_rating
0
1

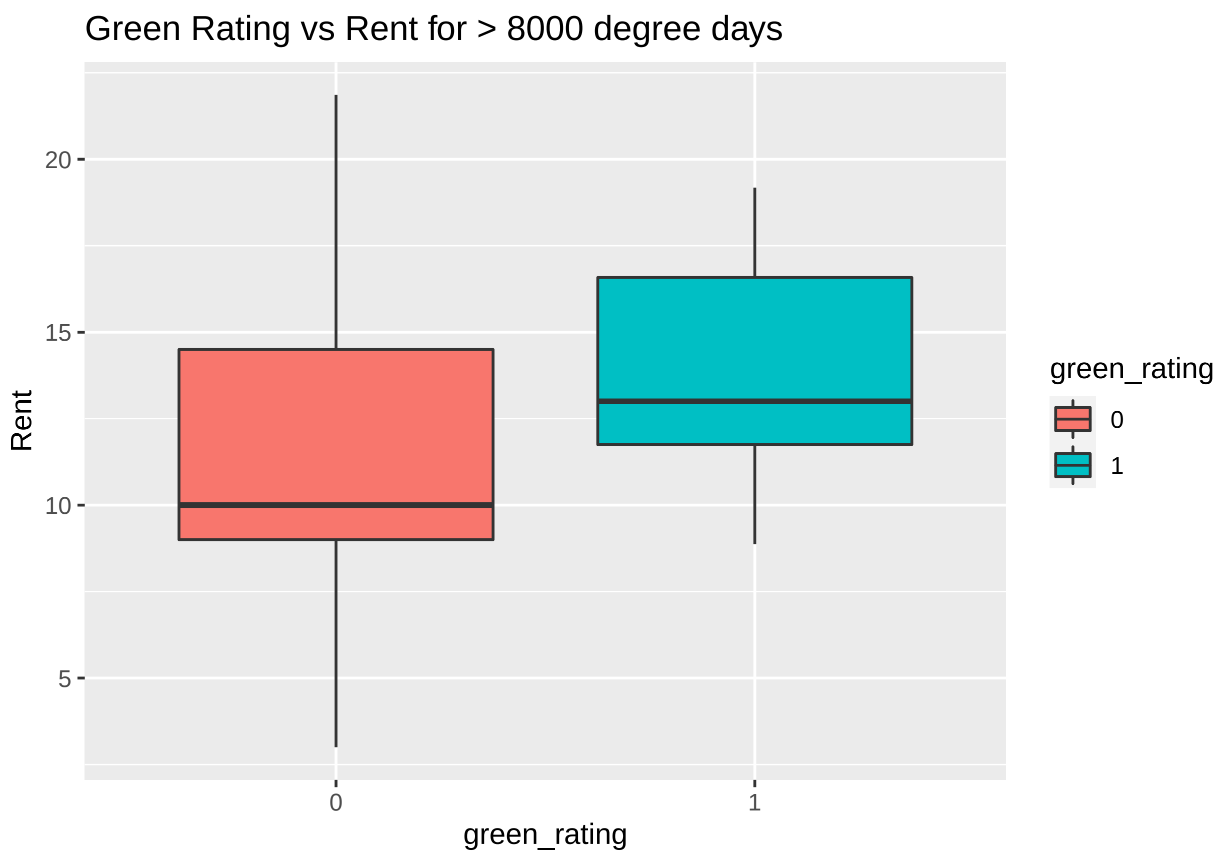


the Rent is comparable for moderate degree days conditions, now we will check for extreme conditions where there are more than 8000 degree days.



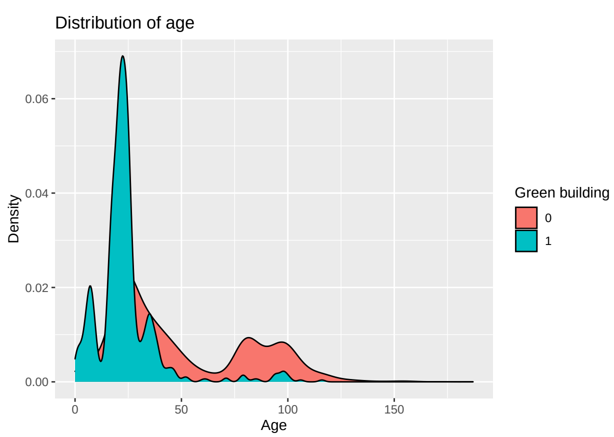
green_rating
0
1
Rent
20
15
10
5
0
1
class_a

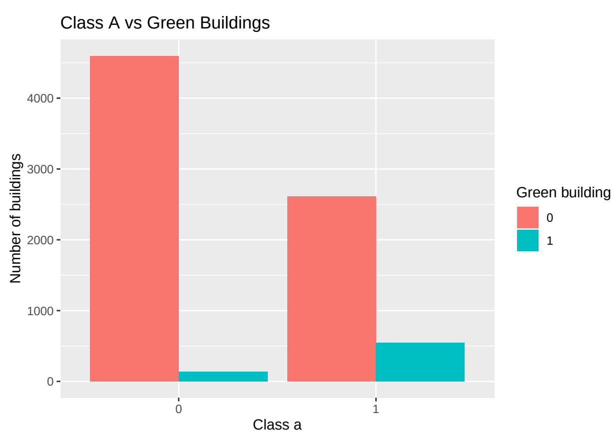

150
100
50
0
count
0
1
class_a
green_rating
0
1

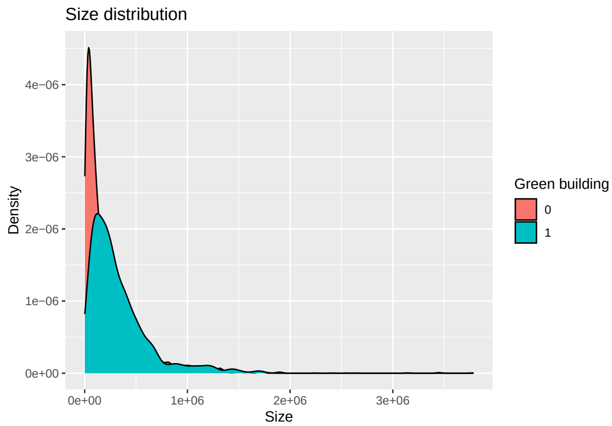


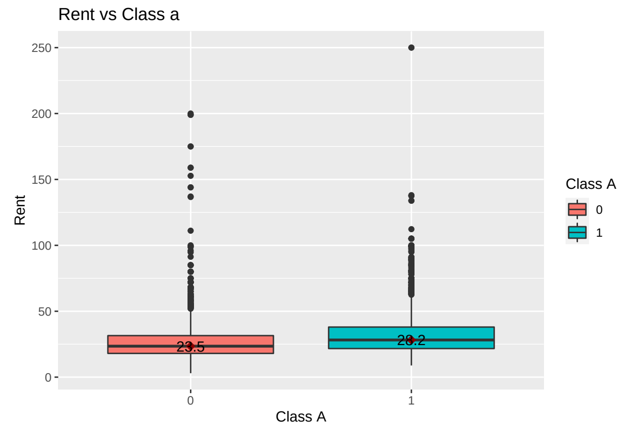
Here we can again confirm that we see rent for green buildings higher for class a In all areas with greater than > 8000 degree days, we see that across the class of the buildings, the rent is higher for green_rated buildings.

Comparing boxplots we can conclude that degree days is a confounding variable which should've been a part of the guru's analysis and we should invest in a green building if they are going to be built in areas with a high number of degree days (>8000), ie areas with extremes of temperature.









Observations: Most of the green buildings are newly built as compared to non-green buildings. There are more class a buildings in green buildings than non-class a buildings. More number of green buildings have higher size than non-green buildings. There is a difference in the rent of class a and non-class a buildings,class a buildings charging more rent.

###Conclusion:

We can conclude that the guru should've taken other variables such as number of degree days, class a b into his consideration.

It would've been financially more feasible if the building is a class a building built in areas with higher number of degree days / more extreme weather conditions, thereby convincing renters to pay higher rent in expectations of having higher savings.

#4 CAPMETRO

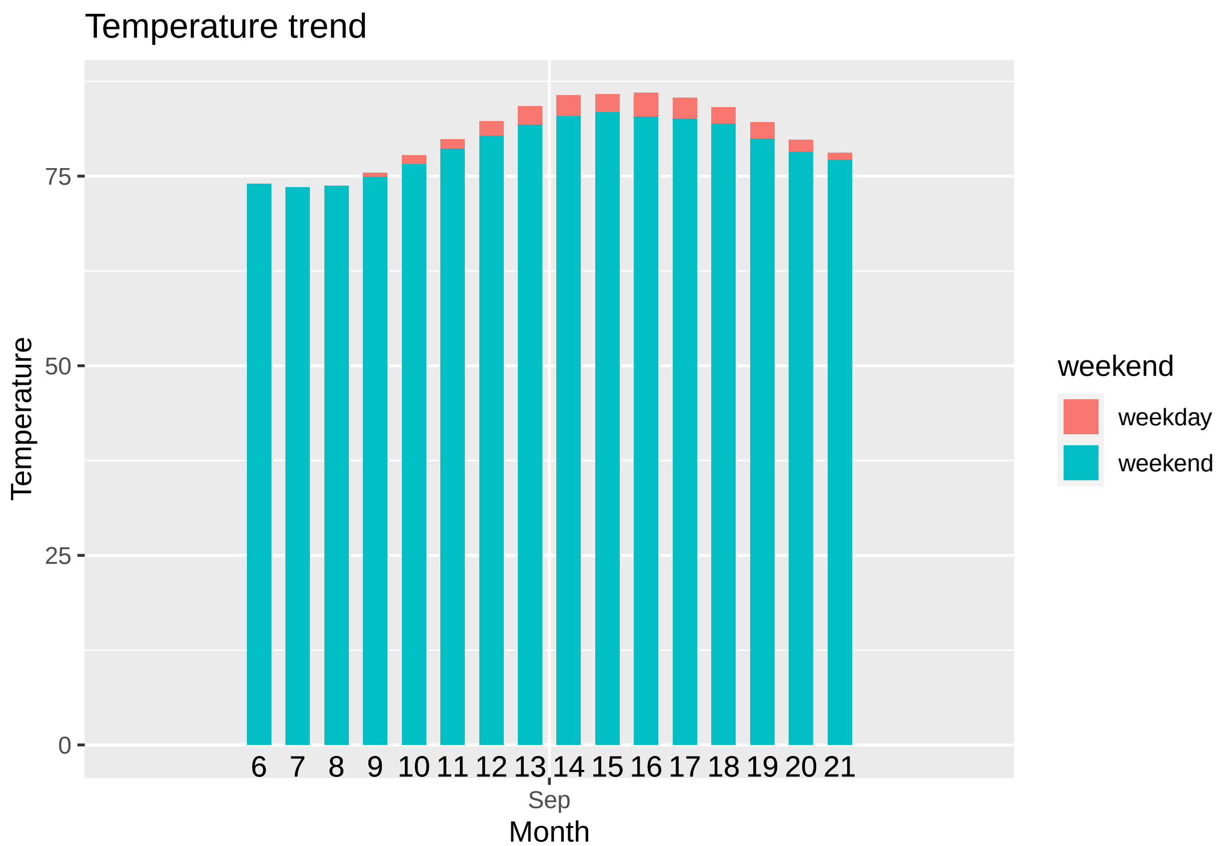
```
## Rows: 5824 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (3): day_of_week, month, weekend
## dbl  (4): boarding, alighting, temperature, hour_of_day
## dttm (1): timestamp
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
##    timestamp                      boarding         alighting
##  Min.   :2018-09-01 06:00:00   Min.   :  0.00   Min.   :  0.00
##  1st Qu.:2018-09-23 17:56:15   1st Qu.: 13.00   1st Qu.: 13.00
##  Median :2018-10-16 13:52:30   Median : 33.00   Median : 28.00
##  Mean   :2018-10-16 13:52:30   Mean   : 51.51   Mean   : 47.65
##  3rd Qu.:2018-11-08 09:48:45   3rd Qu.: 79.25   3rd Qu.: 64.00
##  Max.   :2018-11-30 21:45:00   Max.   :288.00   Max.   :304.00
##  day_of_week         temperature     hour_of_day       month
##  Length:5824        Min.   :29.18   Min.   : 6.00   Length:5824
##  Class :character   1st Qu.:59.20   1st Qu.: 9.75   Class :character
##  Mode  :character   Median :72.75   Median :13.50   Mode  :character
##                     Mean   :69.28   Mean   :13.50
##                     3rd Qu.:79.29   3rd Qu.:17.25
##                     Max.   :97.64   Max.   :21.00
##    weekend
##  Length:5824
```

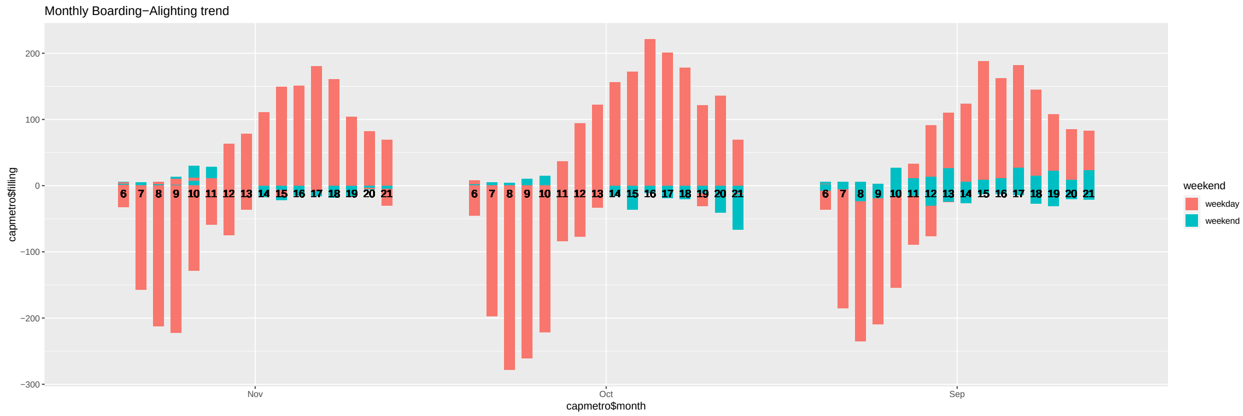
```
##   Class :character
##   Mode  :character
##
##
##
```

For all 3 months, we can see that weekdays have more boarding than weekends. More boarding is observed in the middle of the day between 12pm to 6pm, this might be the time when the buses are in most use.

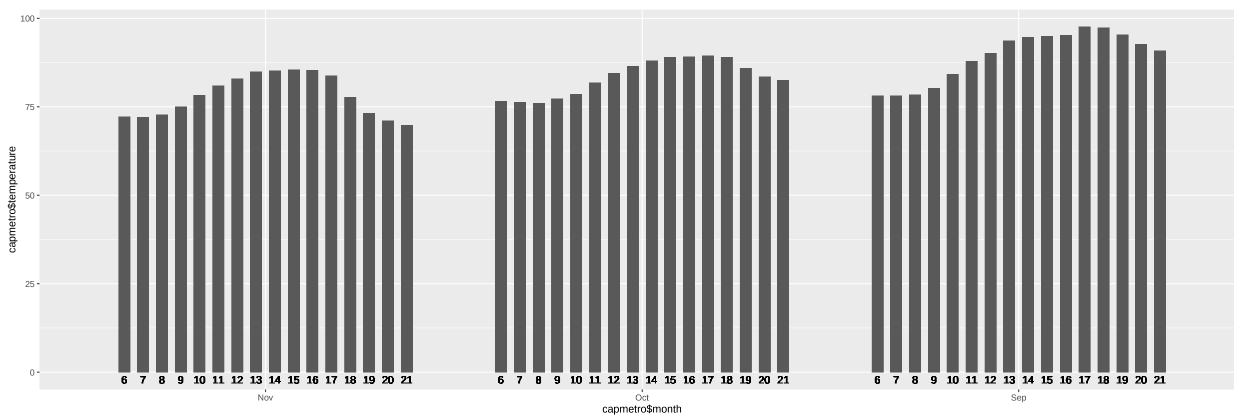
An observation I see while traveling on the buses is their advertisement : come in, cool off. Maybe because temperatures are more during the afternoon, people tend to use buses We will check to see if the temperature graph for a month aligns with this

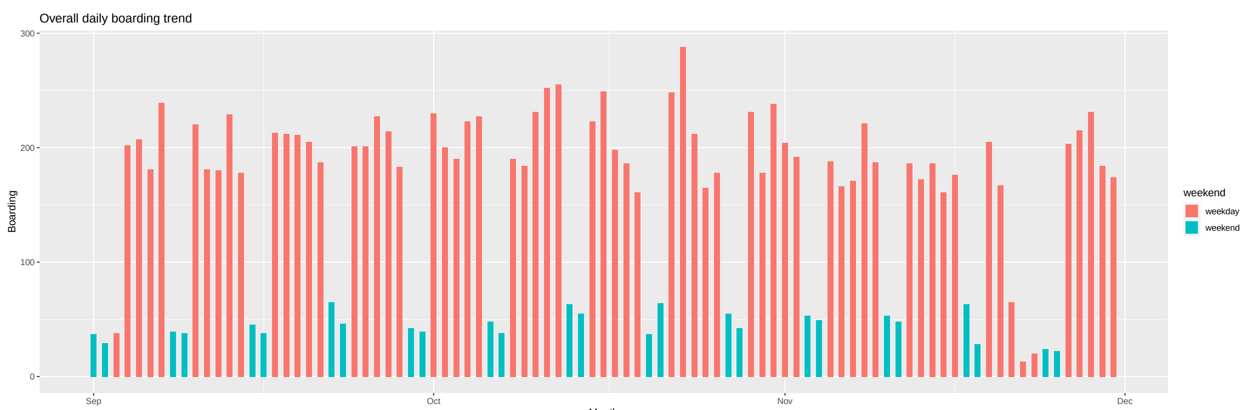


As we can see, the temperatures are more during the same time period boarding is at its peak. This concludes that temperatures along with the advertisement compels people to board more during the afternoon.

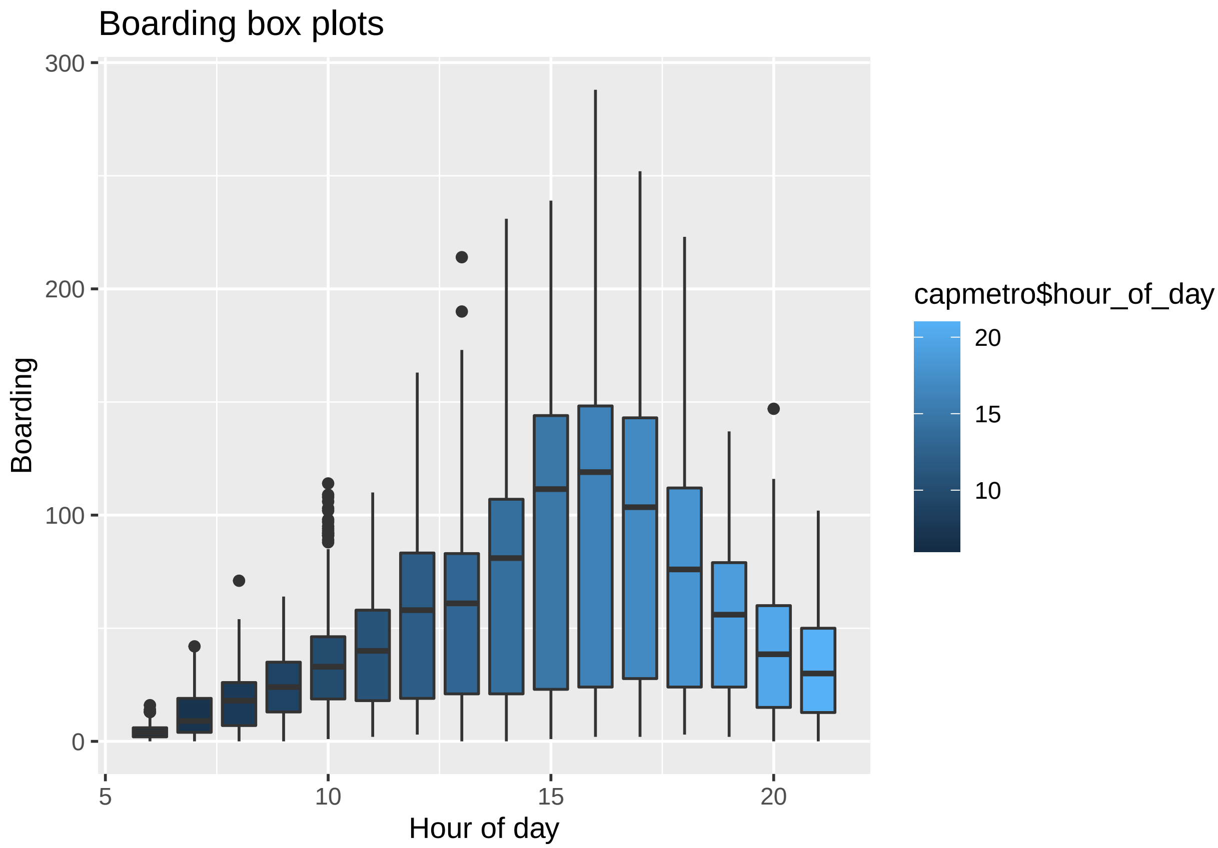


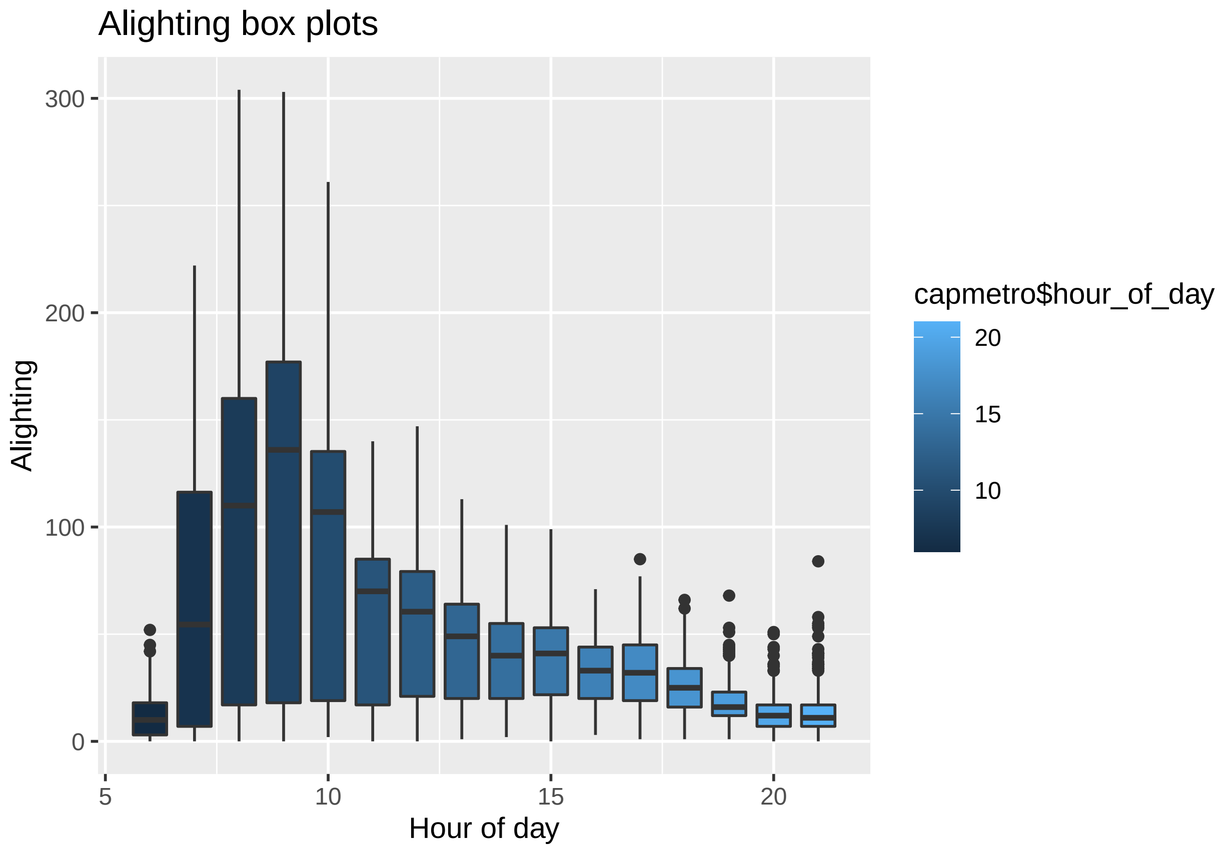
In morning, boarding-alighting is negative. Hence we can infer that the buses would be empty between time



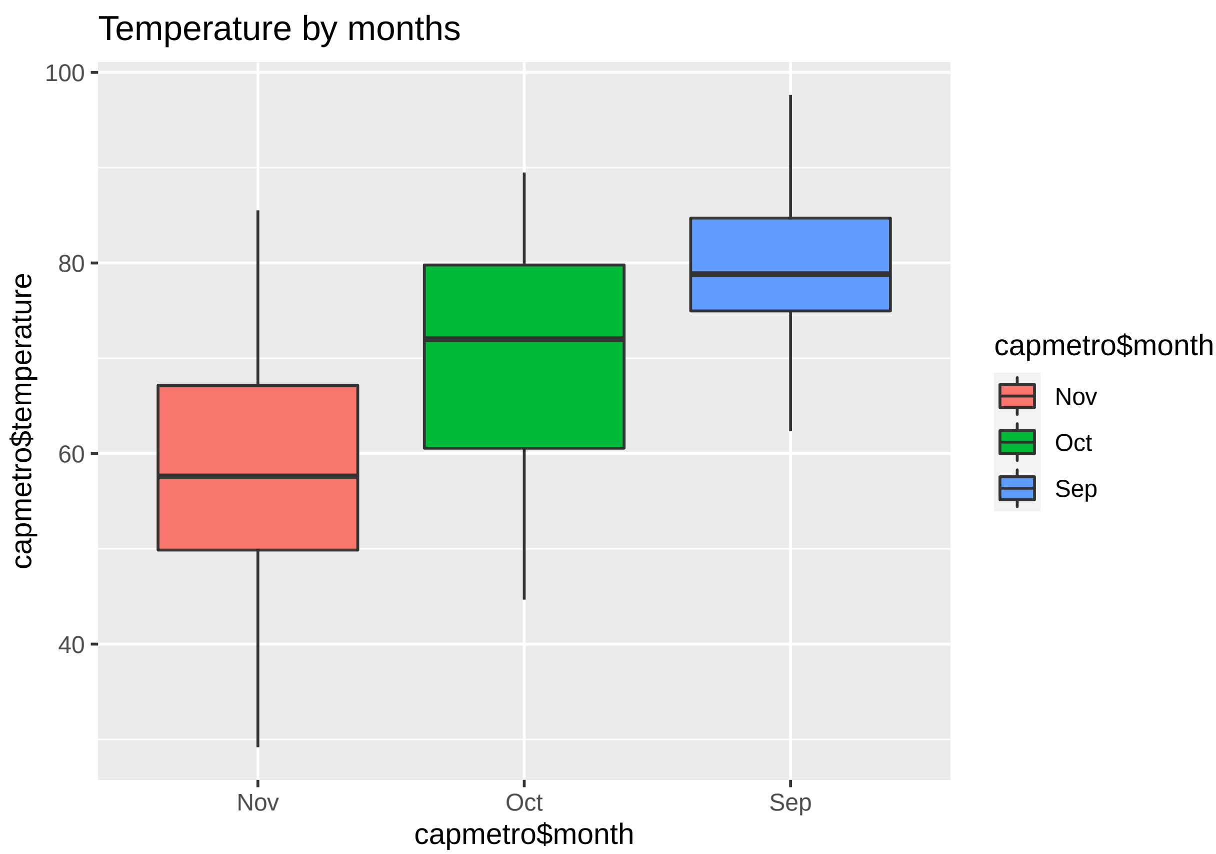


Here is the total line trend of boarding. We see a continuous pattern in the boarding trend, one exception being a prolonged drop in boarding in late November.

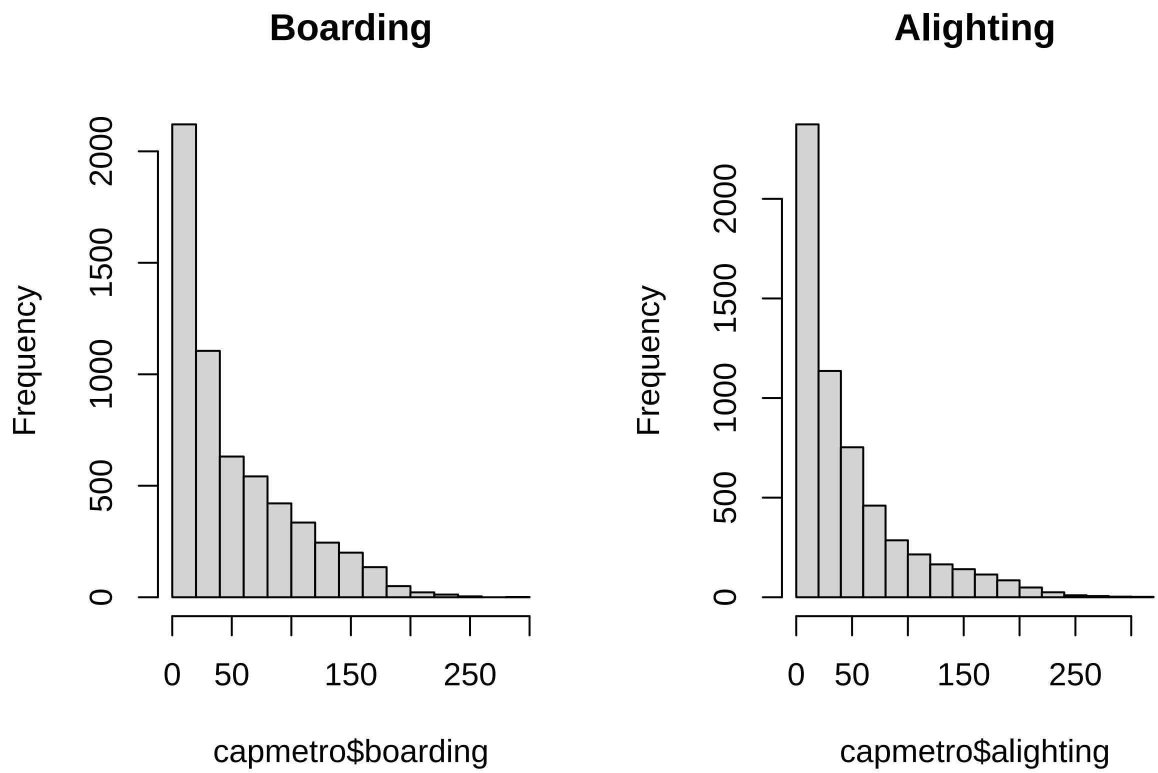

Boarding box plots
Boarding
Hour of day
300
200
100
0
5
10
15
20
capmetro$hour_of_day
20
15
10



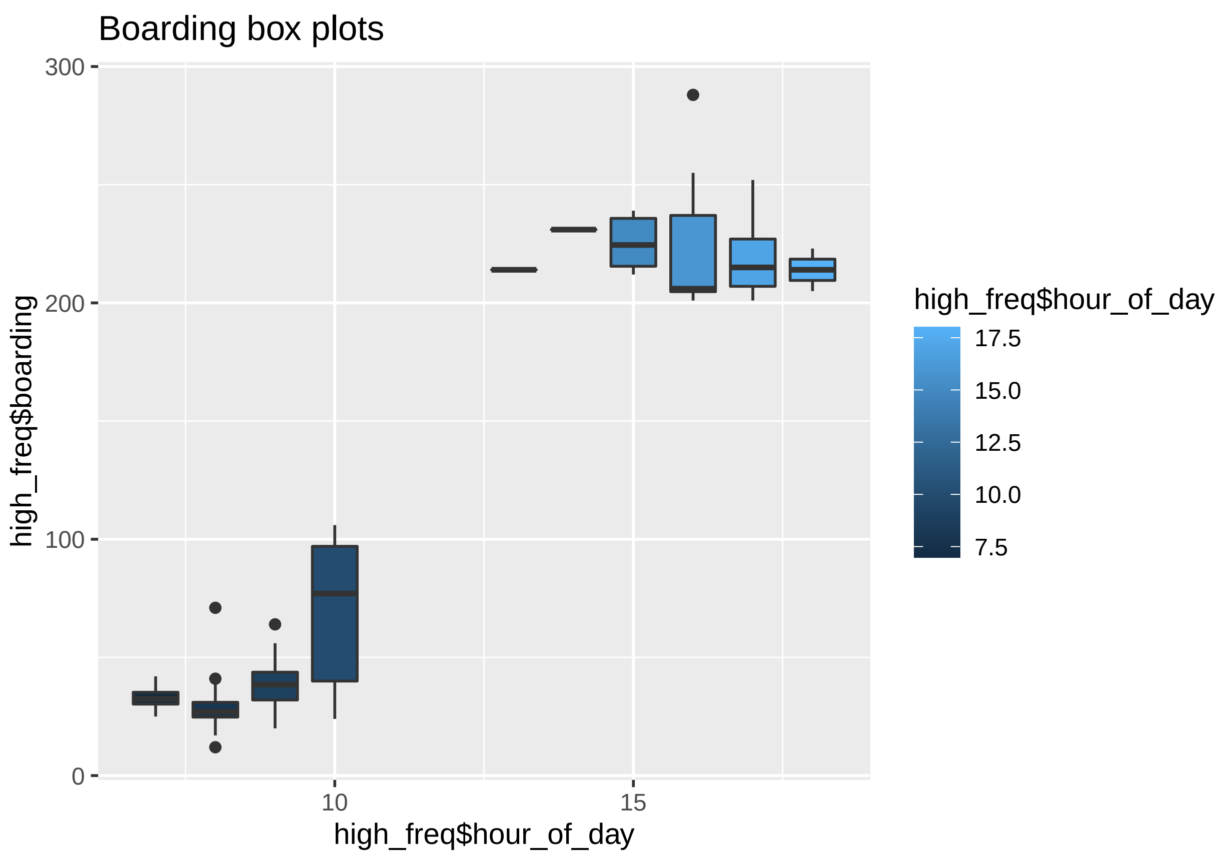
Boarding is at its peak in the middle of the day, alighting is at its peak at the start of the day.

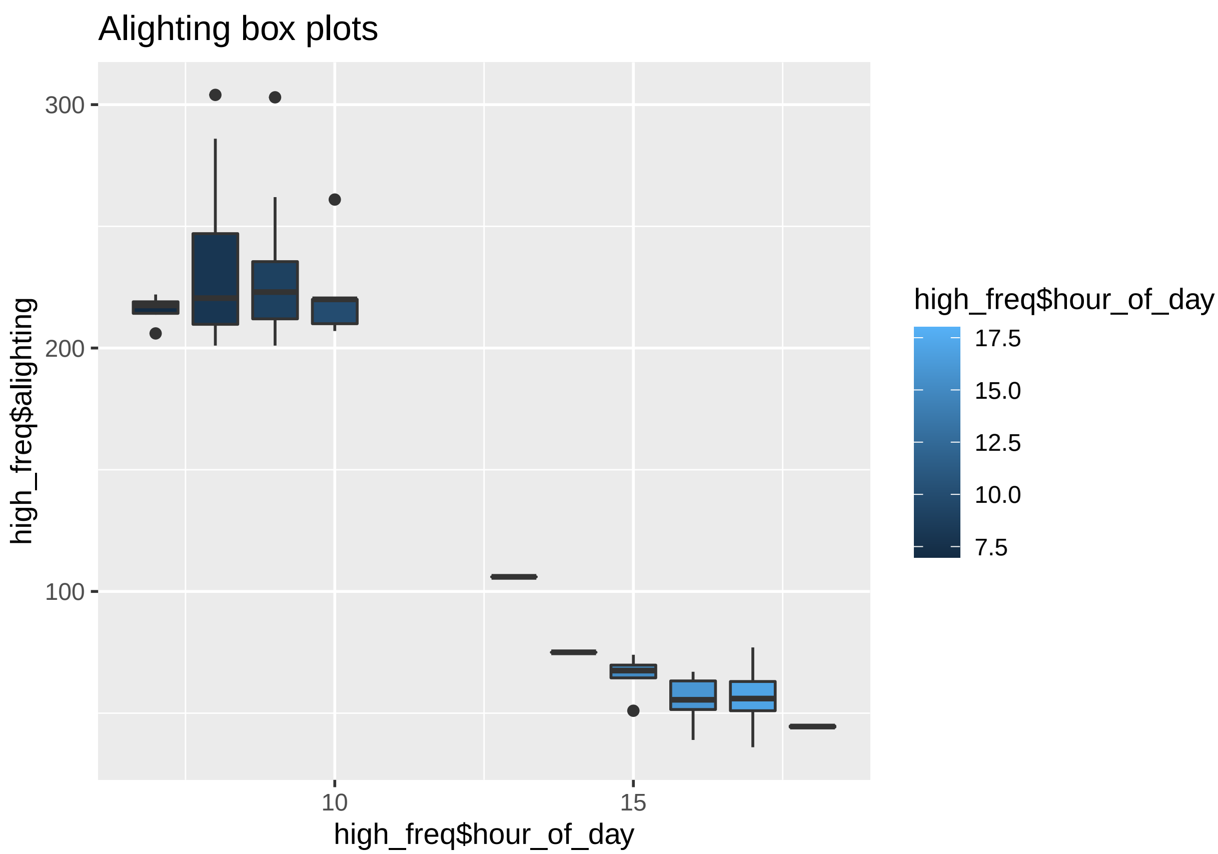


Average temperature is highest in Sep, followed by Oct and Nov



There are also some instances where more than 200 people are either boarding or alighting


Boarding box plots
high_freq$boarding
300
200
100
0
10
15
high_freq$hour_of_day
high_freq$hour_of_day
17.5
15.0
12.5
10.0
7.5



High boarding(>200) happens usually around 3pm and high alighting happens usually around 8:30am