Visual story telling part 1: green buildings

Kristen Lowe and Siboney Cardoso

2024-08-07

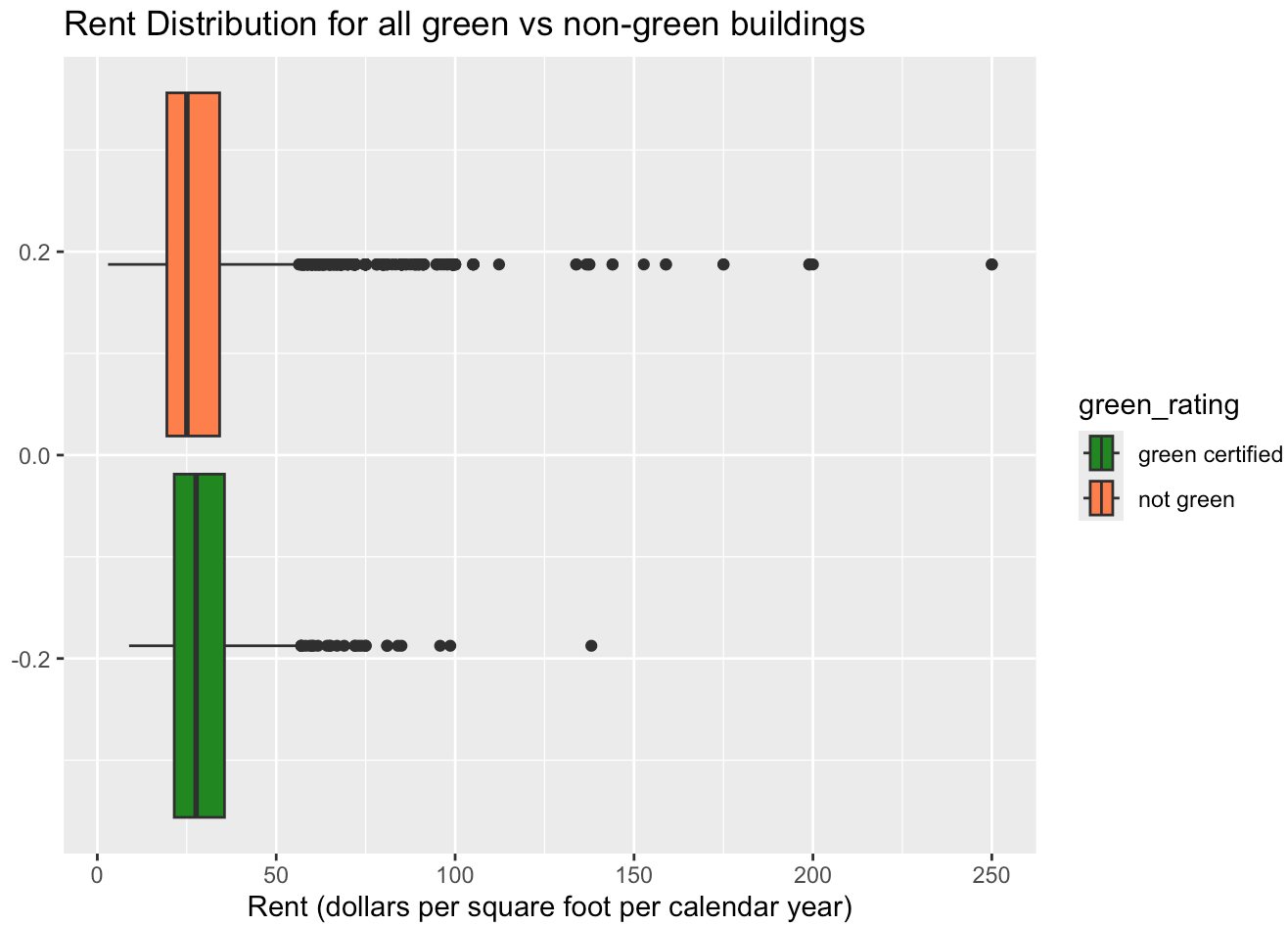
```
library(tidyverse)
```

```
greenbuildings <- read_csv('data/greenbuildings.csv', show_col_types = FALSE)

# clean the data
greenbuildings <- greenbuildings %>%
  filter(leasing_rate >= 10)

# look at green vs non-green buildings
greenbuildings %>%
  mutate(green_rating = ifelse(green_rating == 1, "green certified", "not green")) %>%
  ggplot(aes(x = Rent, fill = green_rating)) +
  geom_boxplot() +
  labs(x = "Rent (dollars per square foot per calendar year)",
       title = "Rent Distribution for all green vs non-green buildings") +
  scale_fill_discrete(name = "Green Status") +
  scale_fill_manual(values = c("green certified" = "forestgreen", "not green" = "cora
l"))
```

```
## Scale for fill is already present.
## Adding another scale for fill, which will replace the existing scale.
```



While the median market rent in non-green buildings is slightly lower than in the green buildings, it does not appear to be a significant difference. Moreover, if you look at the entire range of rent values, there are a number of non-green buildings with extremely high rent, while the green buildings don't reach these values.

```
# look at green vs non-green buildings, grouped by building cluster
non_green_clustered_rent <- greenbuildings %>%
  filter(green_rating == 0) %>%
  group_by(cluster) %>%
  summarise(nonGreenRent = mean(Rent))

greenbuildings %>%
  filter(green_rating == 1) %>%
  dplyr::select(cluster, Rent) %>%
  merge(non_green_clustered_rent, by = 'cluster') %>%
  pivot_longer(cols = c(Rent, nonGreenRent),
               names_to = "green_rating",
               values_to = "rent") %>%
  mutate(green_rating = ifelse(green_rating == "Rent", "green certified", "not green (cl
ustered)")) %>%
  ggplot(aes(x = rent, fill = green_rating)) +
  geom_boxplot() +
  labs(x = "Rent (dollars per square foot per calendar year)",
       title = "Rent Distribution for green vs clustered non-green buildings") +
  scale_fill_discrete(name = "Green Status") +
  scale_fill_manual(values = c("green certified" = "forestgreen", "not green (clustere
d)" = "coral"))
```

```
## Scale for fill is already present.
## Adding another scale for fill, which will replace the existing scale.
```

When we group the non-green buildings by their cluster before plotting the distribution, the market rent disparities between the two groups becomes more apparent: green buildings do indeed tend to have higher rent prices.

```
# calculations

green_building_cost = 105000000
regular_building_cost = 100000000

green_rent = median((greenbuildings %>% filter(green_rating == 1))$Rent)
regular_rent = median(non_green_clustered_rent$nonGreenRent)

area = 250000

green_rev_per_yr = area * green_rent
reg_rev_per_yr = area * regular_rent

print(paste("Number of years to pay off green building:", green_building_cost / green_re
v_per_yr))
```

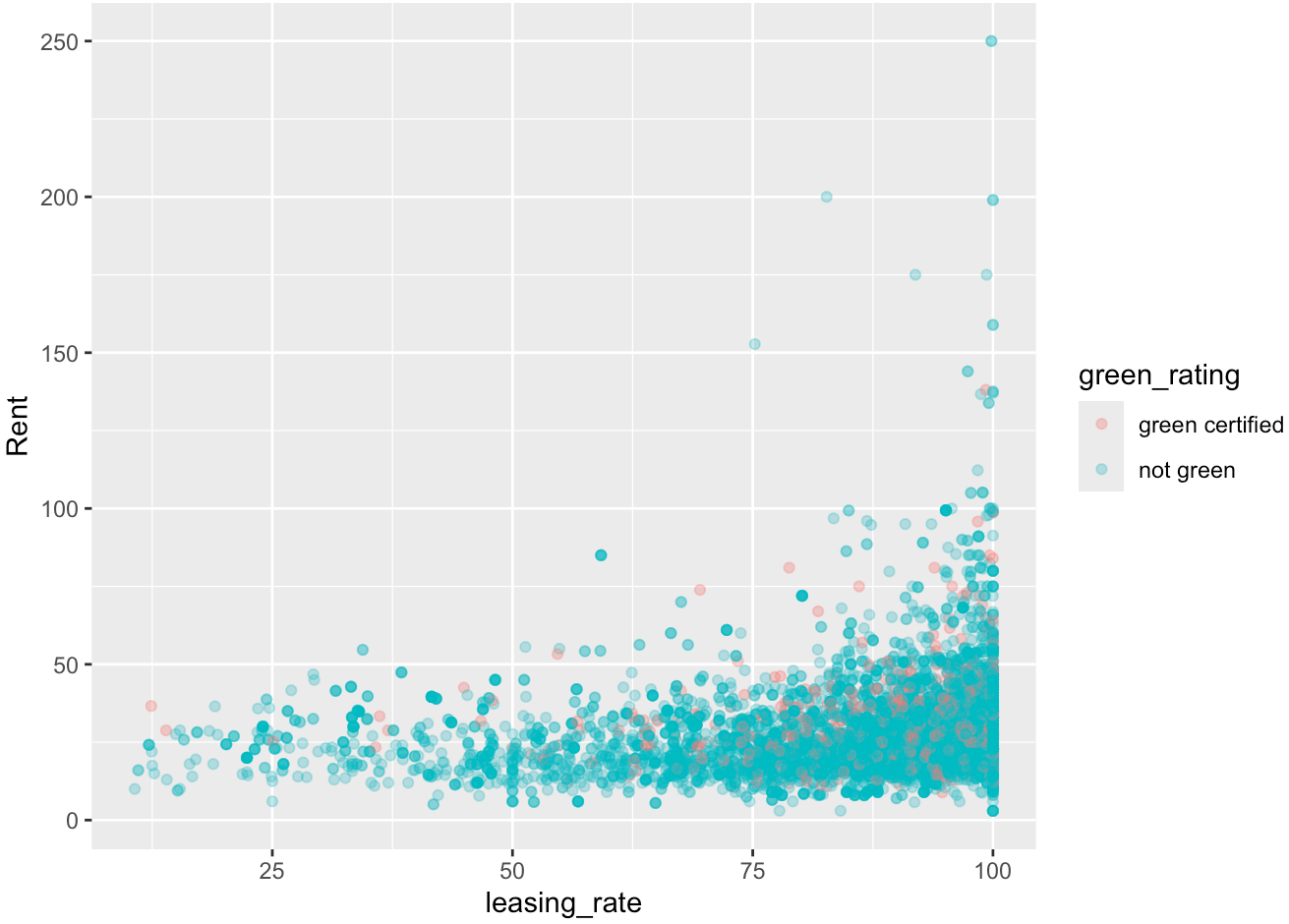
```
## [1] "Number of years to pay off green building: 15.2173913043478"
```

```
print(paste("Number of years to pay off non-green building:", regular_building_cost / re
g_rev_per_yr))
```

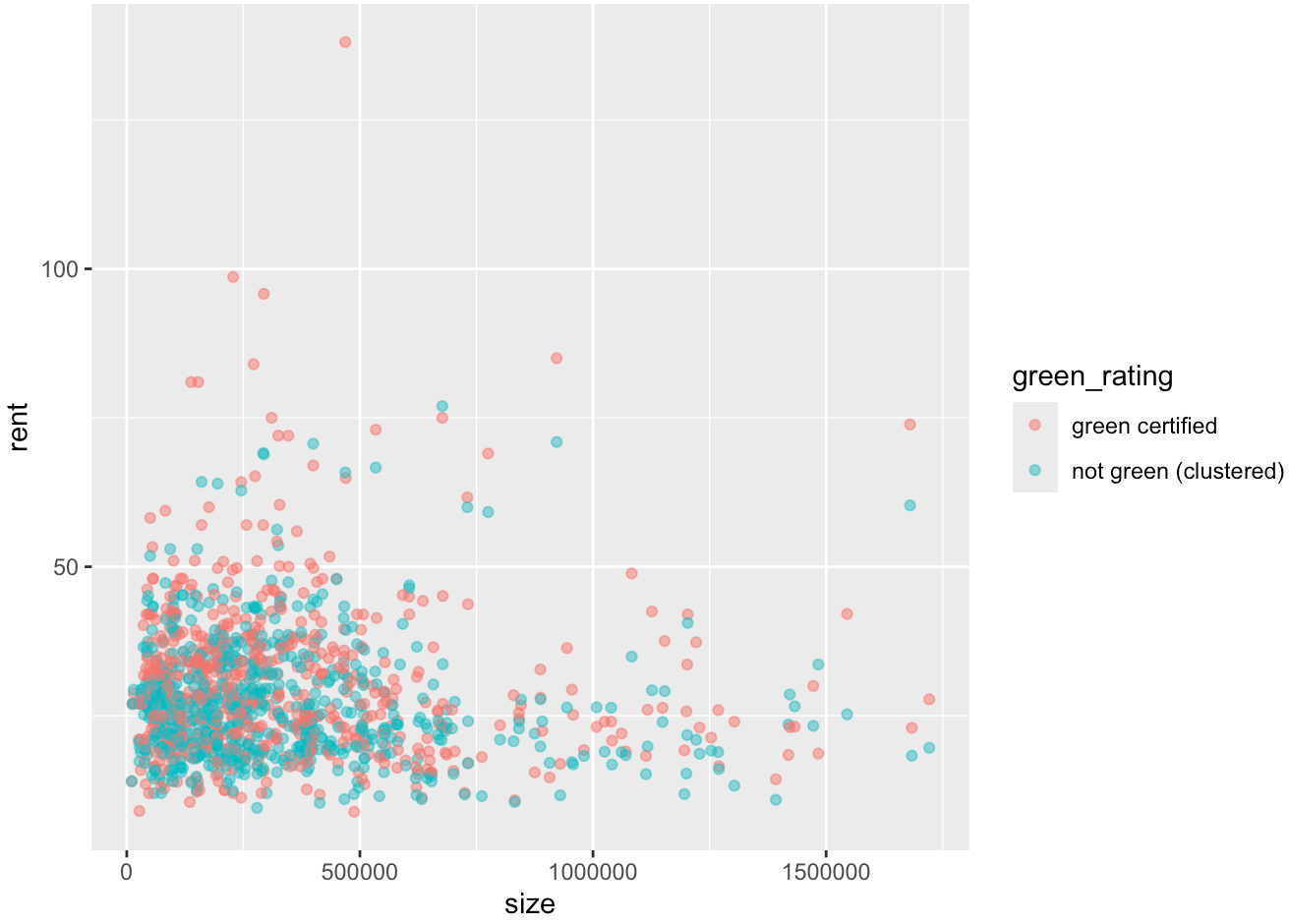
```
## [1] "Number of years to pay off non-green building: 15.8856235107228"
```

Based on my calculations, I agree with the on-staff stats guru. The green building will take less time to pay off, and will make more money for the developer in the long run. However, the stats guru did not directly compare the green building's cost to the regular building's cost, so I decided to include that in my analysis.

```
greenbuildings %>%
  mutate(green_rating = ifelse(green_rating == 1, "green certified", "not green")) %>%
  ggplot(aes(x = leasing_rate,
             y = Rent,
             color = green_rating)) +
  geom_point(alpha = 0.3)
```



```
greenbuildings %>%
  filter(green_rating == 1) %>%
  merge(non_green_clustered_rent, by = 'cluster') %>%
  dplyr::select(-green_rating) %>%
  pivot_longer(cols = c(Rent, nonGreenRent),
               names_to = "green_rating",
               values_to = "rent") %>%
  mutate(green_rating = ifelse(green_rating == "Rent", "green certified", "not green (cl
ustered)")) %>%
  ggplot(aes(x = size,
             y = rent,
             color = green_rating)) +
  geom_point(alpha = 0.5)
```



Upon examining the data for confounding variables, I was not able to find any variables with an obvious visually confounding relationship (example plots above).