

Exploratory_Analysis

2023-01-31

```
library(tidyverse)
library(dplyr)
library(scales)
library(readxl)
Austin_Housing_Data <- read_excel("/Users/naimasagar/Green Housing Project/Austin_Housing_Data.xlsx")
data <- Austin_Housing_Data
```

Square Feet vs Closing Price - Green Building Verification

```
data %>% select ('Square Feet', 'Close Price', 'Green Building Verification') %>%
  rename(sqft = 'Square Feet', close_price = 'Close Price', green_verification = 'Green Building Verification') %>%
  mutate(green_verification = as.character(green_verification)) %>%
  ggplot(aes(x = sqft, y = close_price, color = green_verification)) + geom_point(color = 'black', alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(trans = 'log', labels = label_number(accuracy = 1)) +
  scale_y_continuous(trans = 'log', labels = label_number(accuracy = 1)) +
  labs(title = "Square Feet vs Closing Price", x = "Square Feet", y = "Close Price ($)", color = "Green Building Verification") +
  theme(plot.title = element_text(hjust = 0.5))
```



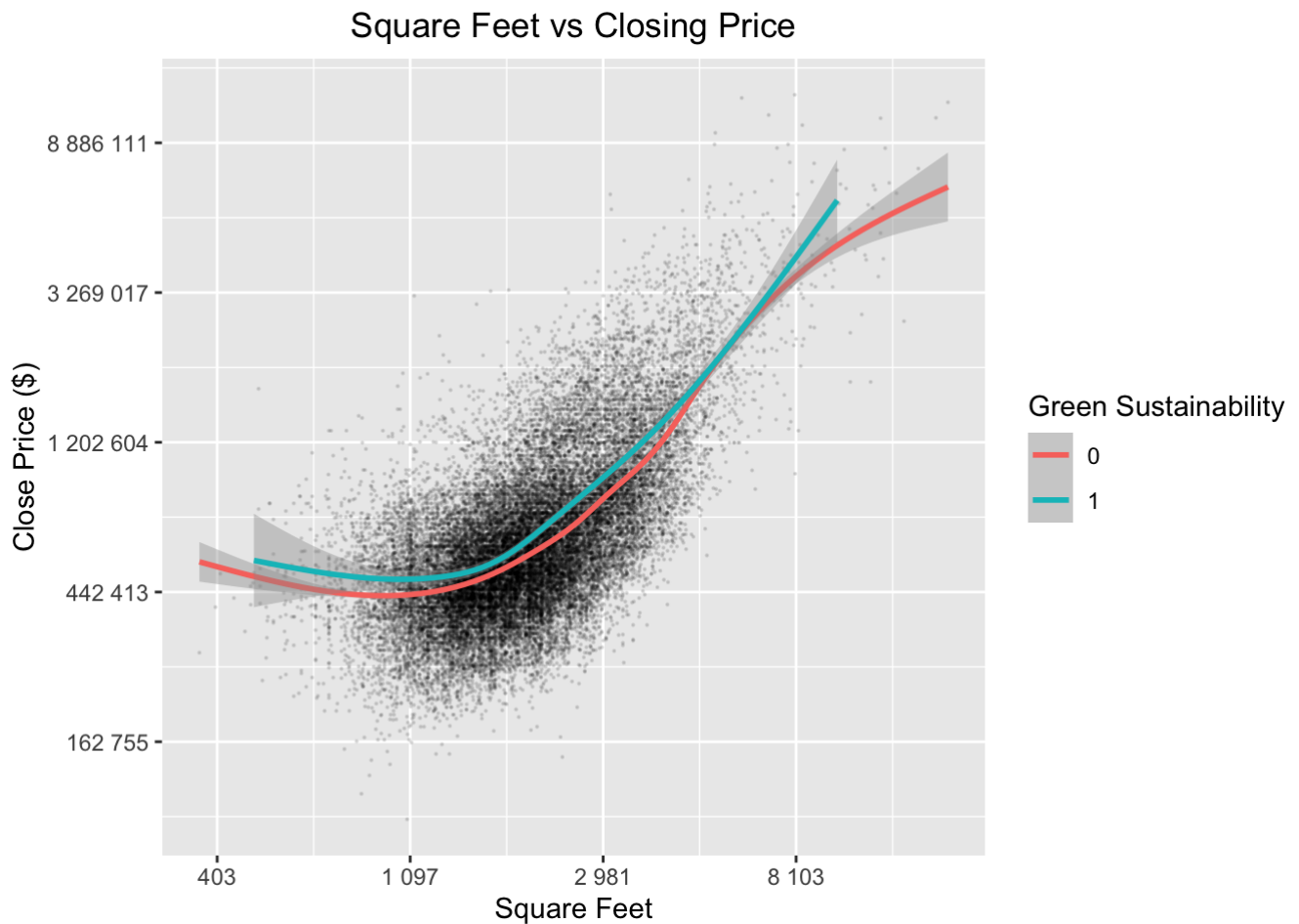
Square Feet vs Closing Price - Green Energy Efficient

```
data %>% select ('Square Feet', 'Close Price', 'Green Energy Efficient') %>%
  rename(sqft = 'Square Feet', close_price = 'Close Price', green_energy = 'Green Energy
Efficient') %>%
  mutate(green_energy = as.character(green_energy)) %>%
  ggplot(aes(x = sqft, y = close_price, color = green_energy)) + geom_point(color =
'black', alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(trans='log', labels = label_number(accur
acy = 1)) +
  scale_y_continuous(trans='log', labels = label_number(accuracy = 1)) +
  labs(title = "Square Feet vs Closing Price", x = "Square Feet", y = "Close Price
($)", color = "Green Energy Efficient") +
  theme(plot.title = element_text(hjust = 0.5))
```



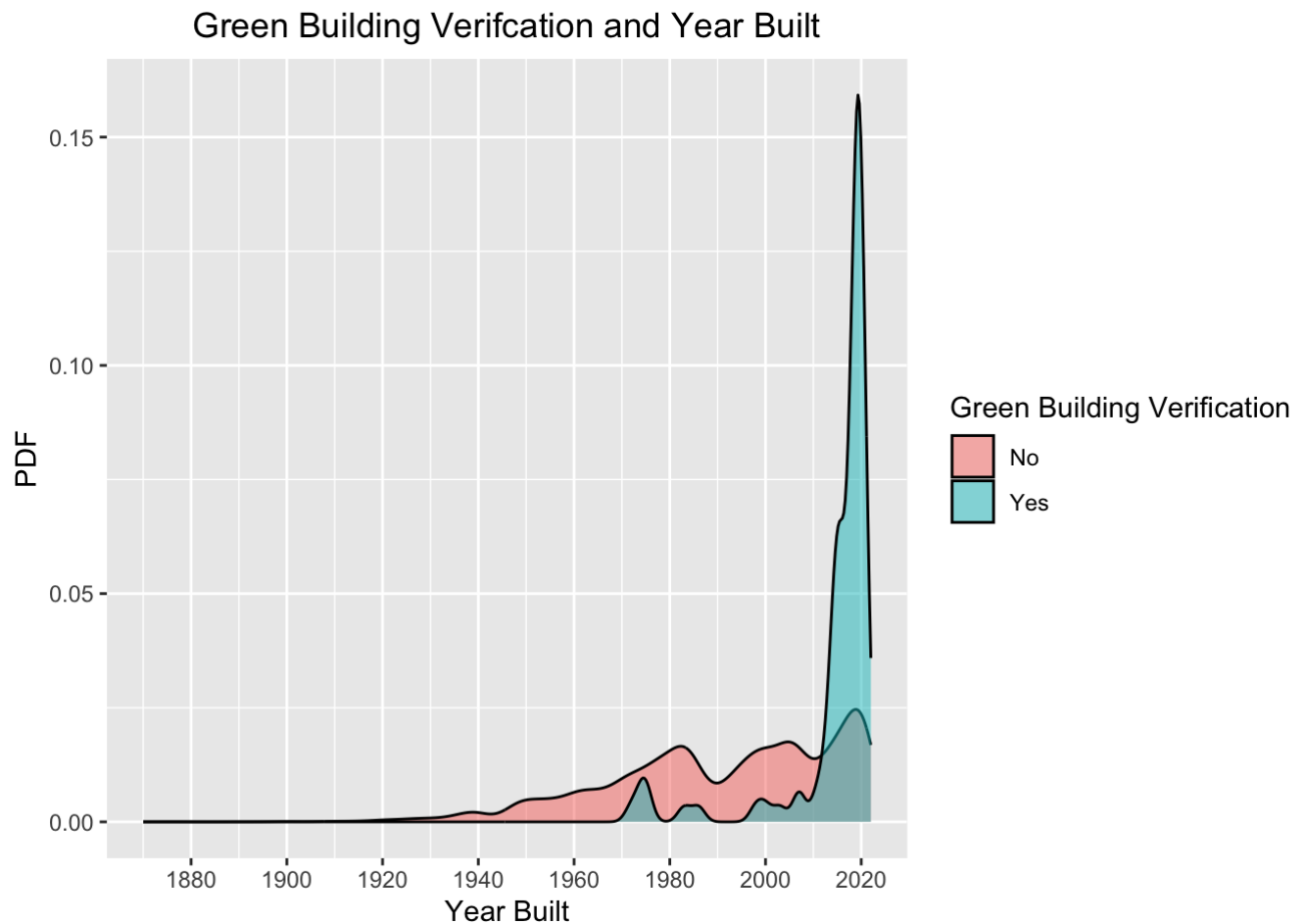
Square Feet vs Closing Price - Green Sustainability

```
data %>% select ('Square Feet', 'Close Price', 'Green Sustainability') %>%
  rename(sqft = 'Square Feet', close_price = 'Close Price', green_sustainability = 'Green Sustainability') %>%
  mutate(green_sustainability = as.character(green_sustainability)) %>%
  ggplot(aes(x = sqft, y = close_price, color = green_sustainability)) + geom_point(color='black', alpha = 0.1, size = 0.1) +
  geom_smooth(size = 1) + scale_x_continuous(trans='log', labels = label_number(accuracy = 1)) +
  scale_y_continuous(trans='log', labels = label_number(accuracy = 1)) +
  labs(title = "Square Feet vs Closing Price", x = "Square Feet", y = "Close Price ($)", color = "Green Sustainability") +
  theme(plot.title = element_text(hjust = 0.5))
```



Year Built and Green Building Verification PDF

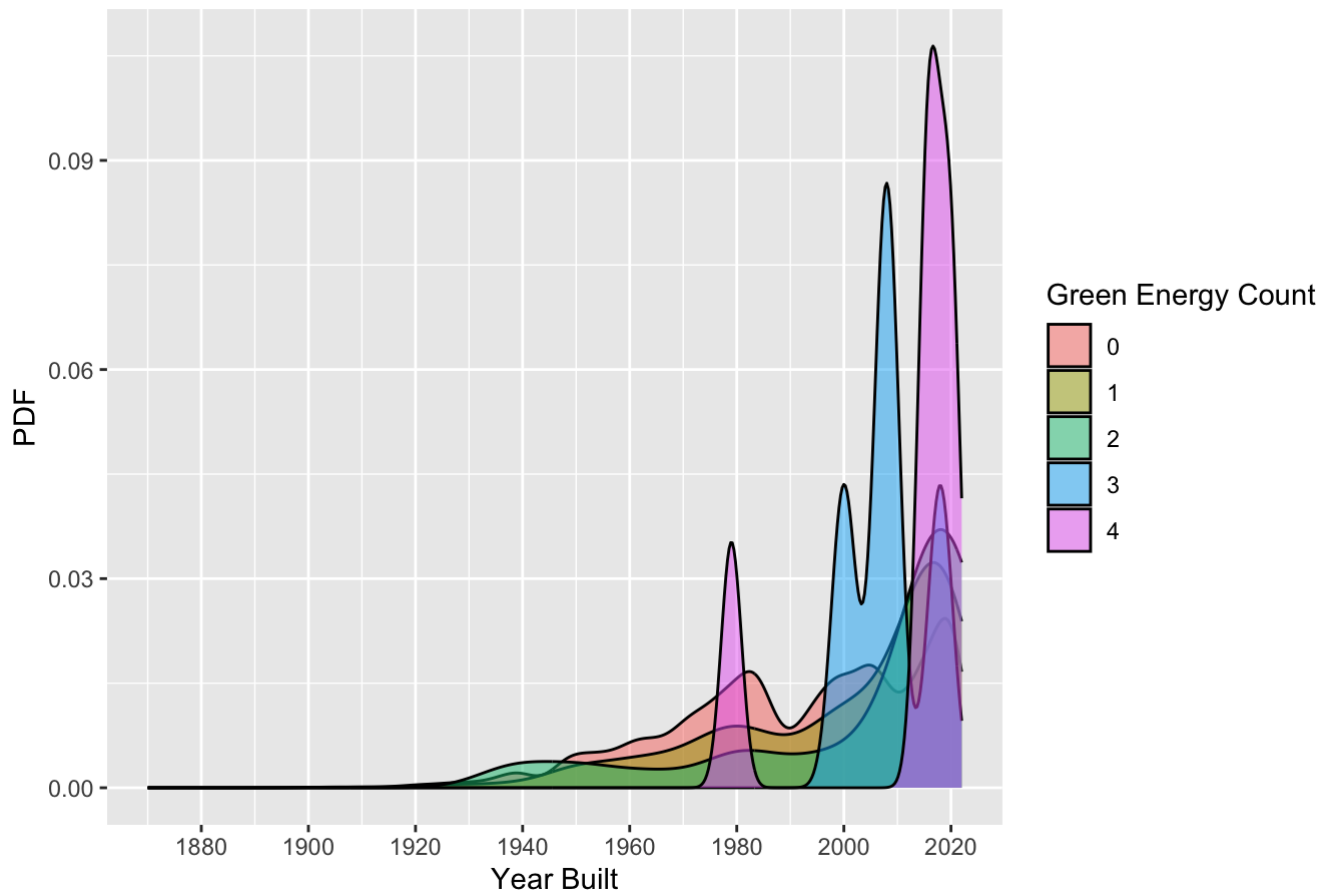
```
data %>% select ('Green Building Verification', 'Year Built') %>%
  rename(green_verification = 'Green Building Verification', year_built = 'Year Built') %>%
  mutate(green_verification = as.character(green_verification)) %>%
  ggplot(aes(x = year_built, fill = green_verification)) + geom_density(alpha = 0.5)
+
  labs(title = "Green Building Verification and Year Built", x = "Year Built", y = "PDF") +
  scale_fill_discrete(name = "Green Building Verification", labels = c("No", "Yes"))
+
  scale_x_continuous(n.breaks = 10) + theme(plot.title = element_text(hjust = 0.5))
```



Year Built and Green Energy Count PDF

```
data %>% select ('Green Energy Count', 'Year Built') %>%  
  rename(green_count = 'Green Energy Count', year_built = 'Year Built') %>%  
  mutate(green_count = as.character(green_count)) %>%  
  ggplot(aes(x = year_built, fill = green_count)) + geom_density(alpha = 0.5) +  
  labs(title = "Green Energy Count and Year Built", x = "Year Built", y = "PDF") +  
  scale_fill_discrete(name = "Green Energy Count") +  
  scale_x_continuous(n.breaks = 10) +  
  theme(plot.title = element_text(hjust = 0.5))
```

Green Energy Count and Year Built



Year Built and Green Energy Efficient PDF

```
data %>% select ('Green Energy Efficient', 'Year Built') %>%
  rename(green_energy = 'Green Energy Efficient', year_built = 'Year Built') %>%
  mutate(green_energy = as.character(green_energy)) %>%
  ggplot(aes(x = year_built, fill = green_energy)) + geom_density(alpha = 0.5) +
  labs(title = "Green Energy Efficiency and Year Built", x = "Year Built", y = "PDF")
+
  scale_fill_discrete(name = "Green Energy Efficient", labels = c("No", "Yes")) +
  scale_x_continuous(n.breaks = 10) +
  theme(plot.title = element_text(hjust = 0.5))
```

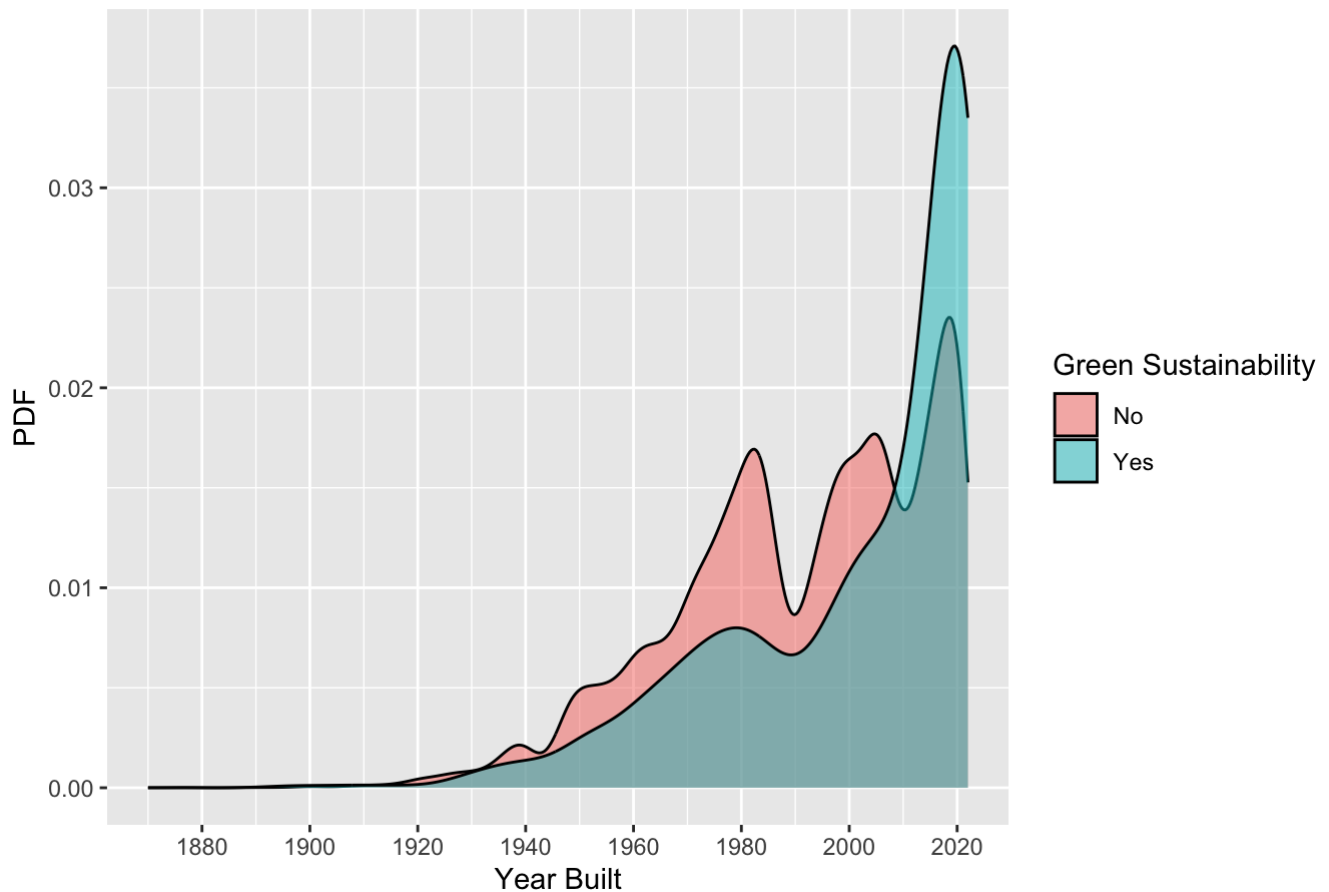
Green Energy Efficiency and Year Built



Year Built and Green Sustainability PDF

```
data %>% select ('Green Sustainability', 'Year Built') %>%  
  rename(green_sustainability = 'Green Sustainability', year_built = 'Year Built') %  
>%  
  mutate(green_sustainability = as.character(green_sustainability)) %>%  
  ggplot(aes(x = year_built, fill = green_sustainability)) + geom_density(alpha = 0.  
5) +  
  labs(title = "Green Sustainability and Year Built", x = "Year Built", y = "PDF") +  
  scale_fill_discrete(name = "Green Sustainability", labels = c("No", "Yes")) +  
  scale_x_continuous(n.breaks = 10) +  
  theme(plot.title = element_text(hjust = 0.5))
```

Green Sustainability and Year Built



Latitude, Longitude & Green Energy Efficient

```
data %>% select ('Longitude', 'Latitude', 'Green Energy Efficient') %>%  
  rename(longitude = 'Longitude', latitude = 'Latitude', green_energy = 'Green Energy  
Efficient') %>%  
  mutate(green_energy = as.character(green_energy)) %>%  
  ggplot(aes(x = longitude, y = latitude, color = green_energy)) +  
  geom_point(alpha = 0.5, size = 0.5) + xlim(-98.1, -97.5) + ylim(30.1, 30.55) +  
  labs(title = "Green Energy Efficient and Location", x = "Longitude", y = "Latitud  
e", color = "Green Energy Efficient") +  
  theme(plot.title = element_text(hjust = 0.5))
```

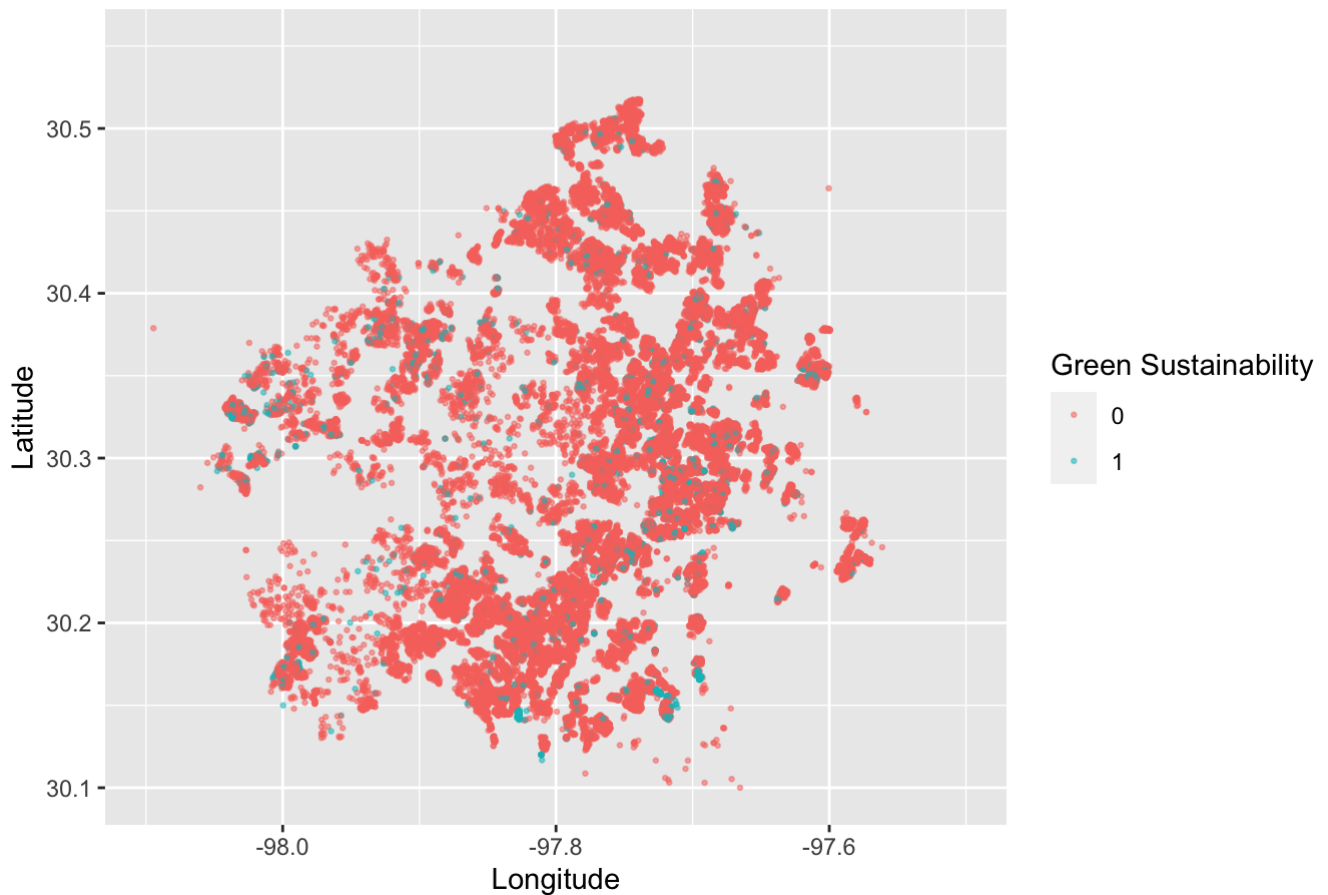

Green Energy Efficient and Location



Latitude, Longitude & Green Sustainability

```
data %>% select ('Longitude', 'Latitude', 'Green Sustainability') %>%  
  rename(longitude = 'Longitude', latitude = 'Latitude', green_sustainability = 'Green Sustainability') %>%  
  mutate(green_sustainability = as.character(green_sustainability)) %>%  
  ggplot(aes(x = longitude, y = latitude, color = green_sustainability)) +  
  geom_point(alpha = 0.5, size = 0.5) + xlim(-98.1, -97.5) + ylim(30.1, 30.55) +  
  labs(title = "Green Sustainability and Location", x = "Longitude", y = "Latitude",  
color = "Green Sustainability") +  
  theme(plot.title = element_text(hjust = 0.5))
```

Green Sustainability and Location



Binary Variable Correlation Matrix

```
data %>% select("Green Building Verification", "Green Energy Efficient", "Green Sustainability", "Accessibility Features", "Homeowner Association", "Occupant Type", "Patio & Porch Features", "Security Features") %>%  
  na.omit() %>% cor() %>% as.data.frame() %>% rownames_to_column() %>% pivot_longer(-1) %>%  
  ggplot(aes(rowname, name, fill=value)) + geom_tile() +  
  xlab("") + ylab("") + coord_fixed() +  
  scale_fill_gradient2(low="red", mid="white", high="blue") +  
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```

