

Final Project Milestone 2: Analyses

Daniel Pinckney

11/17/2020

Data Processing

```
colnames(MoCo_MD_LEED_Cert)
```

Renaming the columns to make them workable

```
## [1] "Building Name"           "Path"  
## [3] "Certification date"      "City"  
## [5] "State"                  "Country"  
## [7] "Rating system"          "Version"  
## [9] "Certification level (name)" "Certification level (numeric)"
```

```
MoCo_MD_LEED_Cert <- MoCo_MD_LEED_Cert %>%  
  rename(  
    name = "Building Name",  
    path = "Path",  
    cert_date = "Certification date",  
    city = "City",  
    state = "State",  
    country = "Country",  
    build_type = "Rating system",  
    version = "Version",  
    cert_name = "Certification level (name)",  
    cert_num = "Certification level (numeric)"  
  ) %>%  
  select(name, cert_date, build_type, cert_name, cert_num)
```

```
MoCo_MD_LEED_Cert <- MoCo_MD_LEED_Cert %>%  
  mutate(new_build = case_when(  
    build_type == "Commercial Interiors" ~ 0,  
    build_type == "New Construction" ~ 1,  
    build_type == "Core and Shell" ~ 1,  
    build_type == "Schools - New Construction" ~ 1,  
    build_type == "Existing Buildings" ~ 0,  
    build_type == "Retail - New Construction" ~ 1,
```

```

    build_type == "Retail - Commercial Interiors" ~ 0
  )) %>%
  drop_na(new_build)

summarize(MoCo_MD_LEED_Cert,
           mean_new_build = mean(new_build))

```

Making the independent variable

```

## # A tibble: 1 x 1
##   mean_new_build
##           <dbl>
## 1           0.612

```

```

MoCo_MD_LEED_Cert %>%
  group_by(new_build) %>%
  summarize(
    mean_score = mean(cert_num)
  )

```

Working with the Dependent Variable

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```

## # A tibble: 2 x 2
##   new_build mean_score
##       <dbl>     <dbl>
## 1         0         2.46
## 2         1         2.52

```

```

summarize(MoCo_MD_LEED_Cert,
           mean_score = mean(cert_num))

```

```

## # A tibble: 1 x 1
##   mean_score
##       <dbl>
## 1         2.5

```

Analysis

```

MoCo_MD_LEED_Cert %>%
  group_by(new_build) %>%
  summarize(
    level_mean = mean(cert_num),
    level_se = sd(cert_num) / sqrt(n())
  ) %>%

```

```

pivot_wider(names_from = new_build, values_from = c(level_mean, level_se)) %>%
mutate(
  ate_level_diff = level_mean_1 - level_mean_0,
  ate_level_se = sqrt(level_se_1 ^ 2 + level_se_0 ^ 2)
) %>%
select(ate_level_diff, ate_level_se)

```

ATE Calculations

```
## 'summarise()' ungrouping output (override with '.groups' argument)
```

```

## # A tibble: 1 x 2
##   ate_level_diff ate_level_se
##         <dbl>         <dbl>
## 1         0.0592         0.129

```

```

ggplot(MoCo_MD_LEED_Cert, aes(cert_num)) +
  geom_histogram(fill = "seagreen", binwidth = 1, color = "black") +
  labs(y = "Number of Certified Buildings", x = "Certification Level",
       title = "Distribution of Building Certification Levels")

```



Plot

```
MoCo_MD_LEED_Cert_new <- MoCo_MD_LEED_Cert %>%
  filter(new_build == 1)

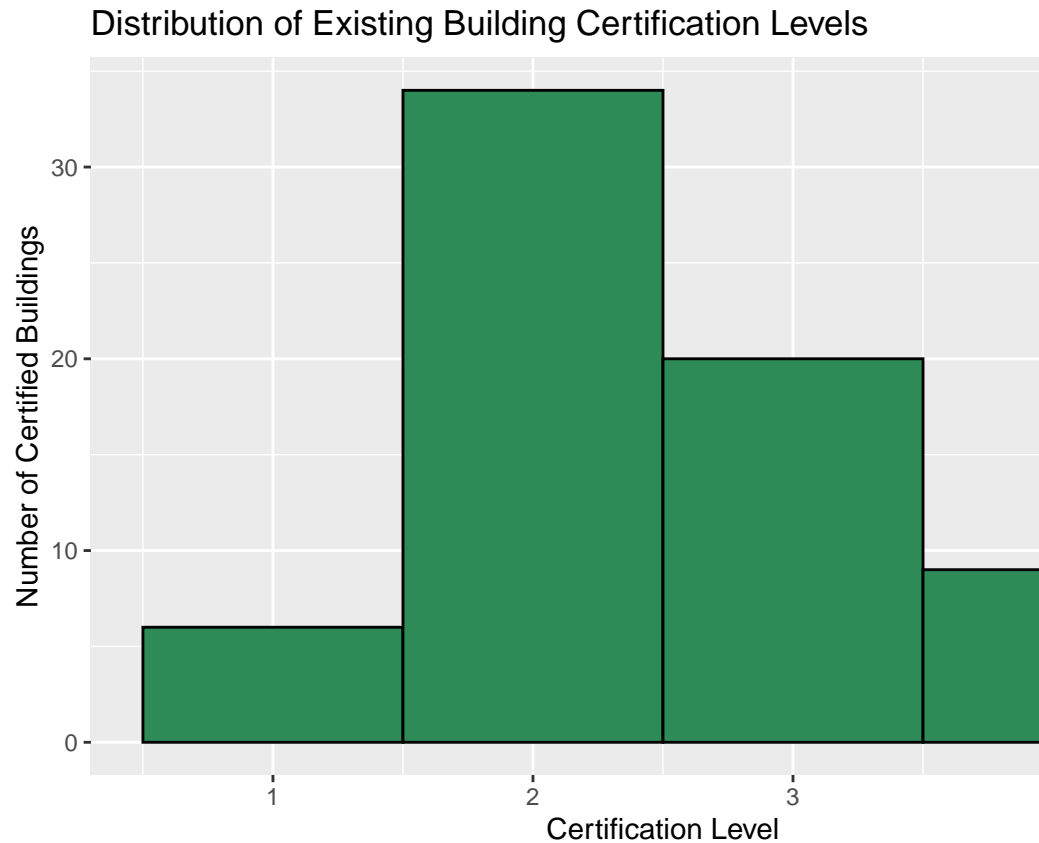
ggplot(MoCo_MD_LEED_Cert_new, aes(cert_num)) +
  geom_histogram(fill = "seagreen", binwidth = 1, color = "black") +
  labs(y = "Number of Certified Buildings", x = "Certification Level",
       title = "Distribution of New Building Certification Levels")
```



New Building Histogram

```
MoCo_MD_LEED_Cert_exist <- MoCo_MD_LEED_Cert %>%
  filter(new_build == 0)

ggplot(MoCo_MD_LEED_Cert_exist, aes(cert_num)) +
  geom_histogram(fill = "seagreen", binwidth = 1, color = "black") +
  labs(y = "Number of Certified Buildings", x = "Certification Level",
       title = "Distribution of Existing Building Certification Levels")
```



Existing Building Histogram

```
lin_fit <- lm(cert_num ~ new_build, data = MoCo_MD_LEED_Cert)
summary(lin_fit)
```

Regression

```
##
## Call:
## lm(formula = cert_num ~ new_build, data = MoCo_MD_LEED_Cert)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5229 -0.5229 -0.4638  0.5362  1.5362
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.46377    0.10119  24.347  <2e-16 ***
## new_build    0.05917    0.12931   0.458    0.648
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8406 on 176 degrees of freedom
## Multiple R-squared:  0.001188,    Adjusted R-squared:  -0.004487
## F-statistic: 0.2094 on 1 and 176 DF,  p-value: 0.6478
```

```
# I am not sure what to do with the regression for this data, but here it is  
  
# also considering doing a regression like lm(cert_num ~ new_build + cert_date),  
# but the date data is strange
```