

# Assignment\_One

Evan Trowbridge

4/24/2021

```
library(readr)
library(janitor)
library(tidyverse)
library(skimr) # For reviewing data
library(plm) # For panel regressions
library(stargazer) # For nice regression tables
library(sf) # For spatial data
```

Part A

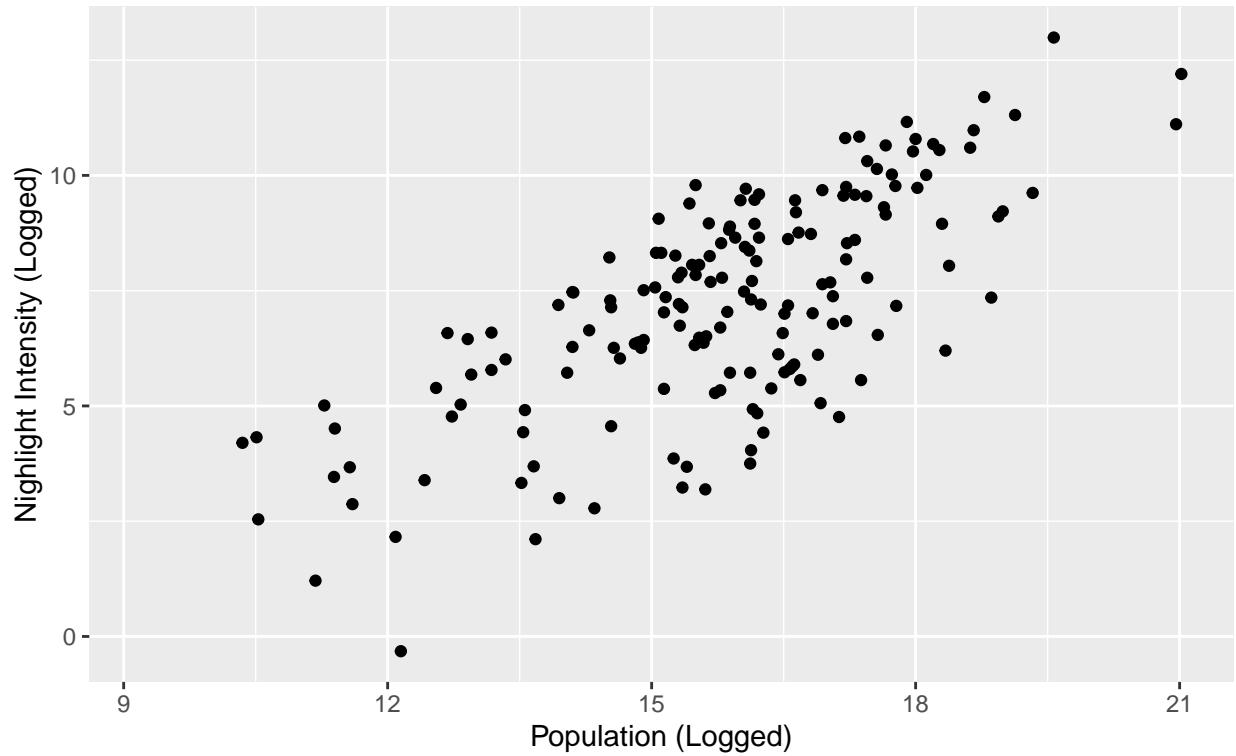
```
df_2a <- df_light %>% filter(date==2012)

# Figure 1: Population
df_2a %>%
  ggplot(aes(x = log_population, y = log_dmsp)) +
  geom_point() +
  labs(
    x = "Population (Logged)",
    y = "Nightlight Intensity (Logged)",
    title = "Figure 1: Nightlight Intensity by Population",
    subtitle = "All values are logged"
  )

## Warning: Removed 11 rows containing missing values (geom_point).
```

Figure 1: Nightlight Intensity by Population

All values are logged

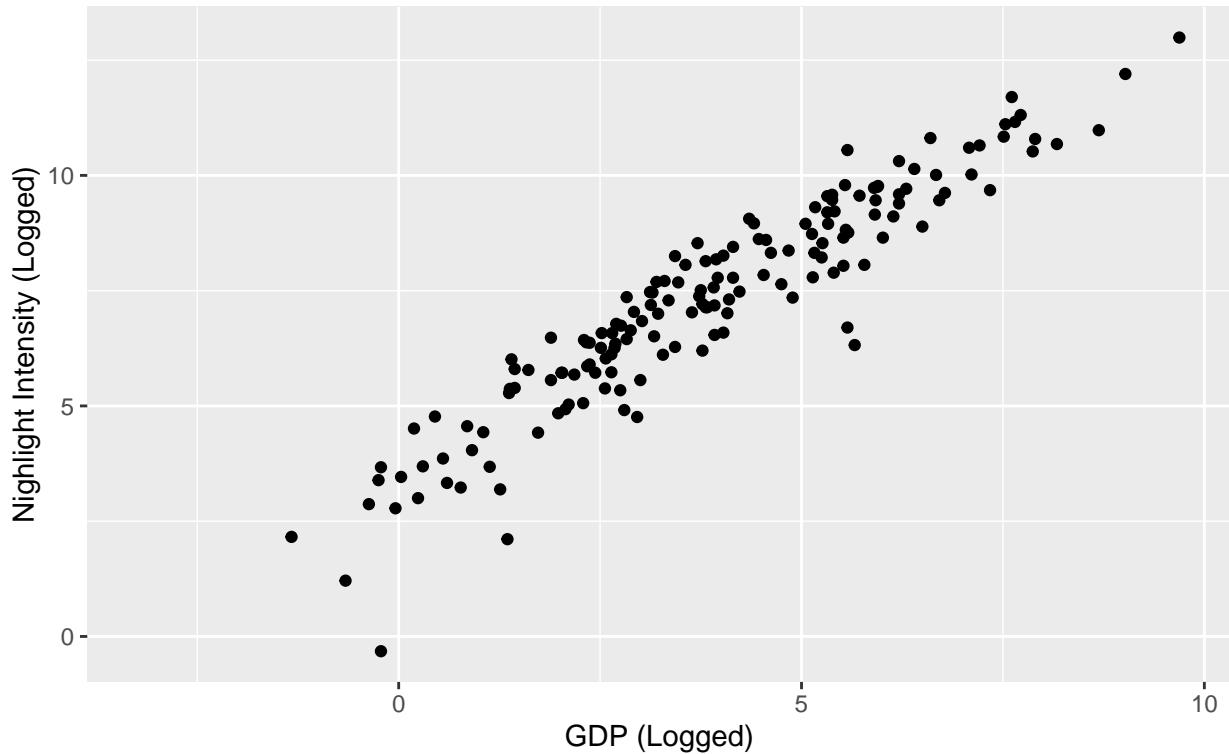


```
# Figure 2: GDP
df_2a %>%
  ggplot(aes(x = log_gdp, y = log_dmsp)) +
  geom_point() +
  labs(
    x = "GDP (Logged)",
    y = "Nightlight Intensity (Logged)",
    title = "Figure 2: Nightlight Intensity by GDP",
    subtitle = "All values are logged"
  )
```

```
## Warning: Removed 18 rows containing missing values (geom_point).
```

Figure 2: Nightlight Intensity by GDP

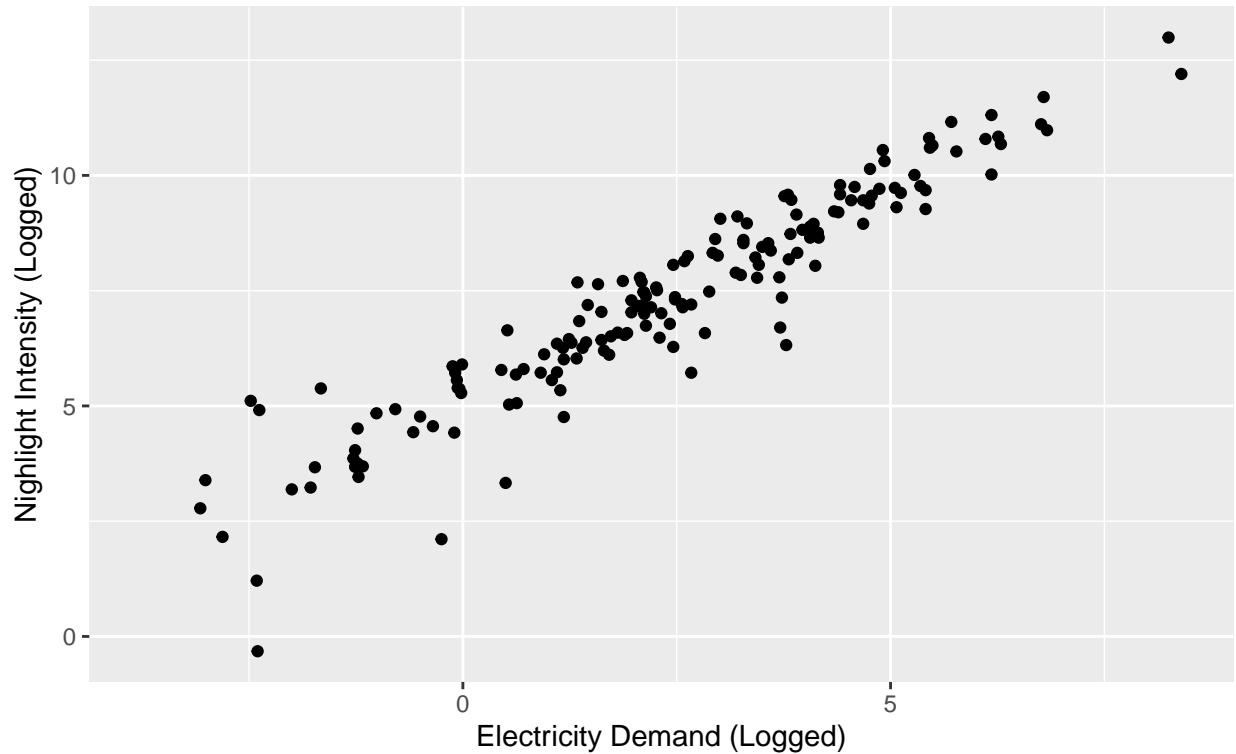
All values are logged



```
# Figure 2: Electricity Demand
df_2a %>%
  ggplot(aes(x = log_elec, y = log_dmsp)) +
  geom_point() +
  labs(
    x = "Electricity Demand (Logged)",
    y = "Nightlight Intensity (Logged)",
    title = "Figure 3: Nightlight Intensity by Electricity Demand",
    subtitle = "All values are logged"
  )
```

## Warning: Removed 15 rows containing missing values (geom\_point).

Figure 3: Nightlight Intensity by Electricity Demand  
All values are logged



Part B

```
df_2b <- df_light

# Using set.seed to make results reproducible
set.seed(123)

# Selecting the 150 countries
df_150_countries <- df_2b %>%
  select(country) %>%
  distinct() %>%
  sample_n(150) %>%
  mutate(training_group = TRUE)

df_2b <- left_join(df_2b, df_150_countries, by = "country") %>%
  distinct() %>%
  mutate(
    training_group =
      if_else(training_group == TRUE, TRUE, FALSE, missing = FALSE),
    date = as_factor(date)) %>%
  filter(
    !is.na(log_gdp) &
    !is.na(log_population) &
    !is.na(log_dmsp))

# Converting to plm panel indexed dataframe
```

```

p_2b <- pdata.frame(df_2b, index = c("date")) %>%
  filter(
    !is.na(log_gdp) &
    !is.na(log_population) &
    !is.na(log_dmfp)) %>%
  select(date, log_gdp, log_population, log_dmfp, training_group)

# Dataframe for countries in training group
p_2b_150 <- p_2b %>% filter(training_group==TRUE)

# Dataframe for countries in prediction group
p_2b_30 <- p_2b %>% filter(training_group==FALSE)

# Regression model
# I use 0 as the intercept since this uses fixed effects
# Reference: https://stackoverflow.com/questions/65702581/predict-out-of-sample-on-fixed-effects-mode
gdp_mod <- plm(log_gdp ~ 0 + log_population + log_dmfp,
                 data = p_2b_150,
                 model = "within")

p_2b_30 <- predict(gdp_mod, newdata = p_2b_30) %>%
  tibble() %>%
  rename(pred_value = ".") %>%
  bind_cols(p_2b_30)

cor(p_2b_30$log_gdp, p_2b_30$pred_value)

```

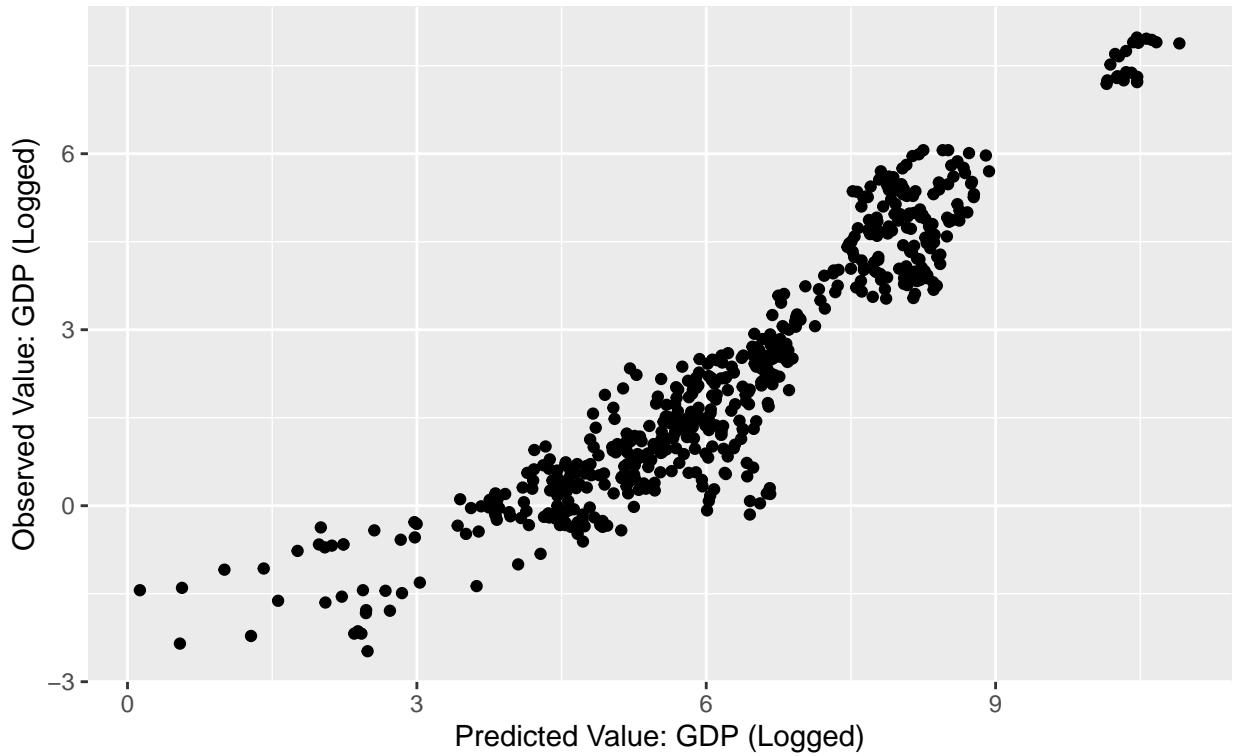
## [1] 0.9363949

```

p_2b_30 %>%
  ggplot(aes(x=pred_value, y =log_gdp)) +
  geom_point() +
  labs(
    x = "Predicted Value: GDP (Logged)",
    y = "Observed Value: GDP (Logged)",
    title = "Figure 4: Predicted Values and Observed Values",
    subtitle = "For 30 Countries Not in Training Data"
  )

```

**Figure 4: Predicted Values and Observed Values  
For 30 Countries Not in Training Data**



```
stargazer(gdp_mod, title = "Results of 150-Country Sample", type = "latex")
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
% Date and time: Wed, Apr 28, 2021 - 10:40:54 PM

Table 1: Results of 150-Country Sample

<i>Dependent variable:</i>	
	log_gdp
log_population	0.114*** (0.011)
log_dmisp	0.799*** (0.009)
Observations	2,906
R <sup>2</sup>	0.857
Adjusted R <sup>2</sup>	0.856
F Statistic	8,617.072*** (df = 2; 2885)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

```

df_2c <- df_light

df_2c <- df_2c %>%
  filter(
    !is.na(log_gdp) &
    !is.na(dictator) &
    !is.na(log_dmsp)) %>%
  select(date, country, log_gdp, log_dmsp, dictator)

# Converting to plm panel indexed dataframe
p_2c <- pdata.frame(df_2c) %>%
  filter(
    !is.na(log_gdp) &
    !is.na(dictator) &
    !is.na(log_dmsp)) %>%
  select(country, date, log_gdp, log_dmsp, dictator)

gdp_mod_2c <- plm(log_gdp ~ log_dmsp * dictator,
                     index = c("country", "date"),
                     data = p_2c,
                     model = "within")

summary(gdp_mod_2c)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log_gdp ~ log_dmsp * dictator, data = p_2c, model = "within",
##       index = c("country", "date"))
##
## Unbalanced Panel: n = 22, T = 150-165, N = 3489
##
## Residuals:
##      Min. 1st Qu. Median 3rd Qu. Max.
## -2.60382 -0.54824 -0.06181  0.47770  3.72289
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## log_dmsp                  0.9083131  0.0064162 141.5648 < 2.2e-16 ***
## dictator                   0.5935240  0.0988103   6.0067 2.089e-09 ***
## log_dmsp:dictator        -0.1582315  0.0139789 -11.3193 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    16849
## Residual Sum of Squares: 2134.6
## R-Squared:               0.87331
## Adj. R-Squared:          0.87243
## F-statistic: 7959.62 on 3 and 3464 DF, p-value: < 2.22e-16

df_2c %>% mutate(
  dictator_text = if_else(
    dictator == 1,

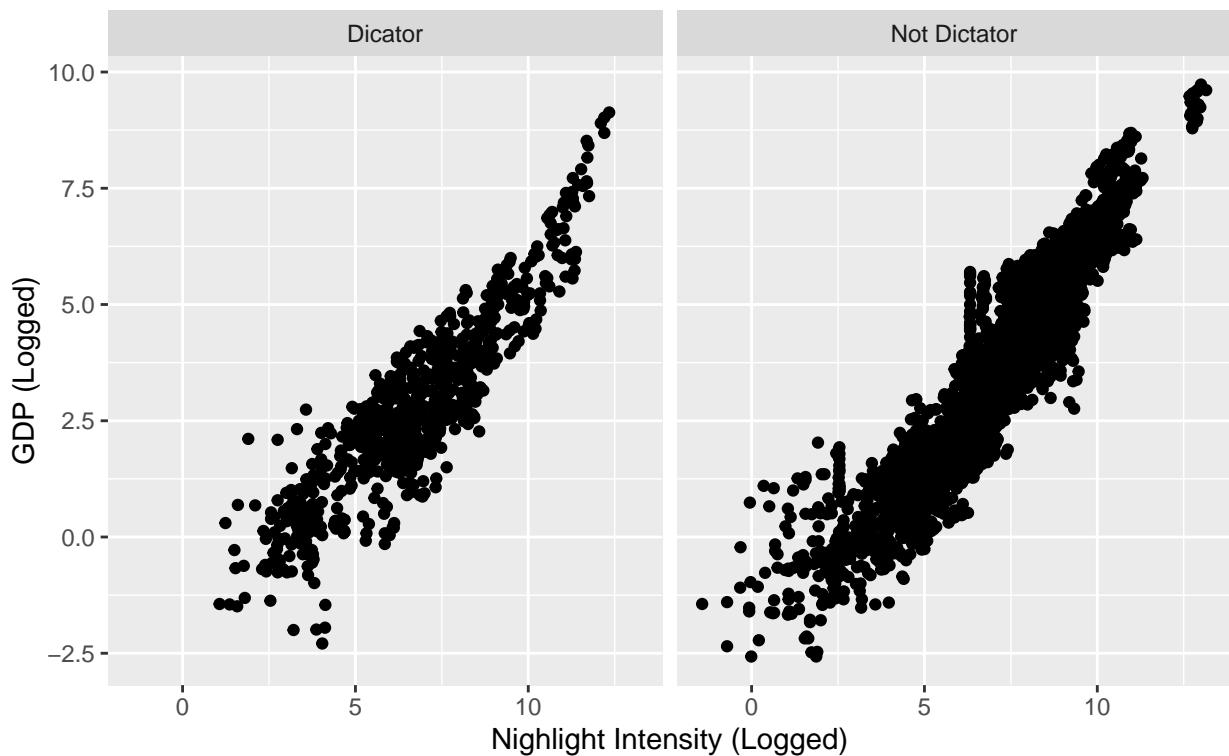
```

```

  "Dicator",
  "Not Dicatator")) %>%
ggplot(aes(x=log_dmsp, y =log_gdp)) +
geom_point() +
labs(
  x = "Nightlight Intensity (Logged)",
  y = "GDP (Logged)",
  title = "Figure 5: Relationship Between Nightlight Intensity and GDP",
  subtitle = "Comparing Between Governments With and Without Dictators"
) +
facet_grid(cols=vars(dictator_text))

```

**Figure 5: Relationship Between Nightlight Intensity and GDP**  
**Comparing Between Governments With and Without Dictators**



```
stargazer(gdp_mod_2c, title = "GDP Correlation with Nightlights and Dictatorial Government", type = "la
```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
% Date and time: Wed, Apr 28, 2021 - 10:40:55 PM

```
india_shape <- st_read("india_nl.shp")
```

```

## Reading layer 'india_nl' from data source 'C:\Users\edtro\OneDrive\Documents\GitHub\energy_dev_world'
## Simple feature collection with 641 features and 72 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:            xmin: 68.11009 ymin: 6.755698 xmax: 97.4091 ymax: 37.0503
## geographic CRS: GCS_Unknown

```

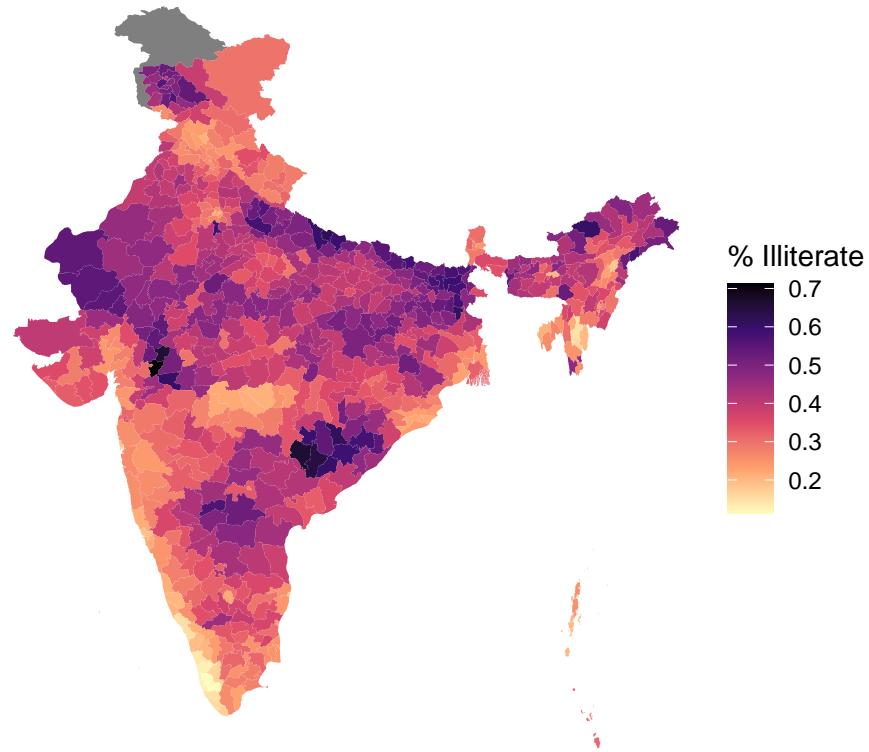
Table 2: GDP Correlation with Nightlights and Dictatorial Government

<i>Dependent variable:</i>	
	log_gdp
log_dmfp	0.908*** (0.006)
dictator	0.594*** (0.099)
log_dmfp:dictator	-0.158*** (0.014)
Observations	3,489
R <sup>2</sup>	0.873
Adjusted R <sup>2</sup>	0.872
F Statistic	7,959.620*** (df = 3; 3464)

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

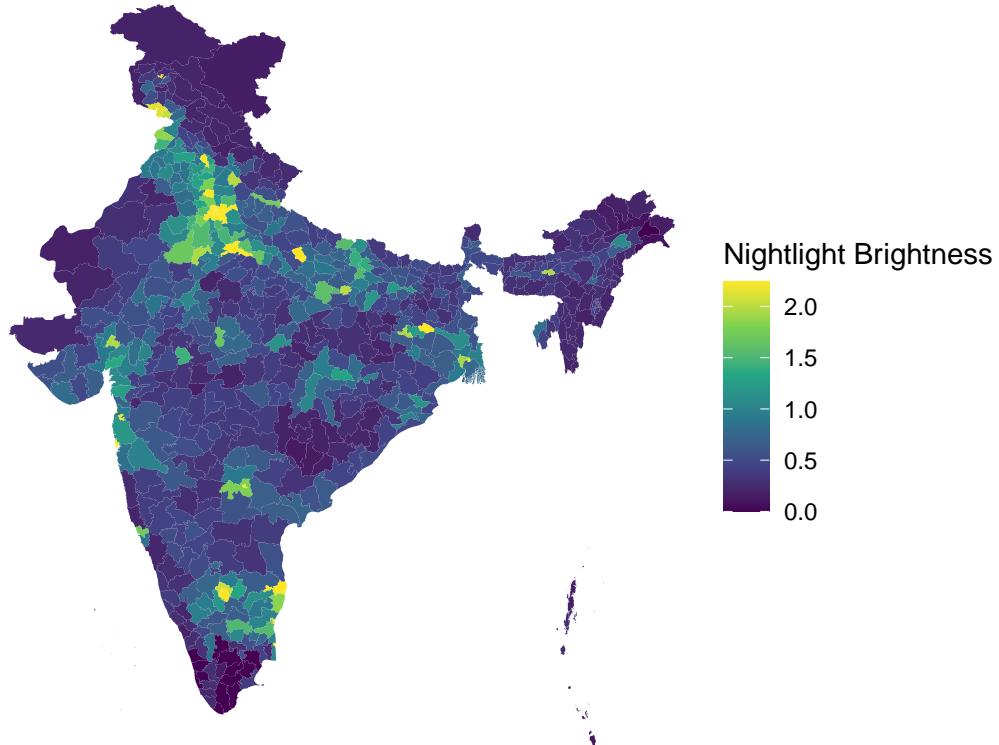
```
india_shape %>%
  ggplot(aes(fill = PERC_ILLT)) +
  geom_sf(color = NA) +
  scale_fill_viridis_c(
    option = "magma",
    direction = -1,
    name = "% Illiterate") +
  theme_void() +
  labs(title = "Figure 6: Illiteracy in India by District")
```

Figure 6: Illiteracy in India by District



```
india_shape %>%
  ggplot(aes(fill = NL)) +
  geom_sf(color = NA) +
  scale_fill_viridis_c(
    name = "Nightlight Brightness") +
  theme_void() +
  labs(title = "Figure 7: Nightlight in India by District")
```

Figure 7: Nightlight in India by District



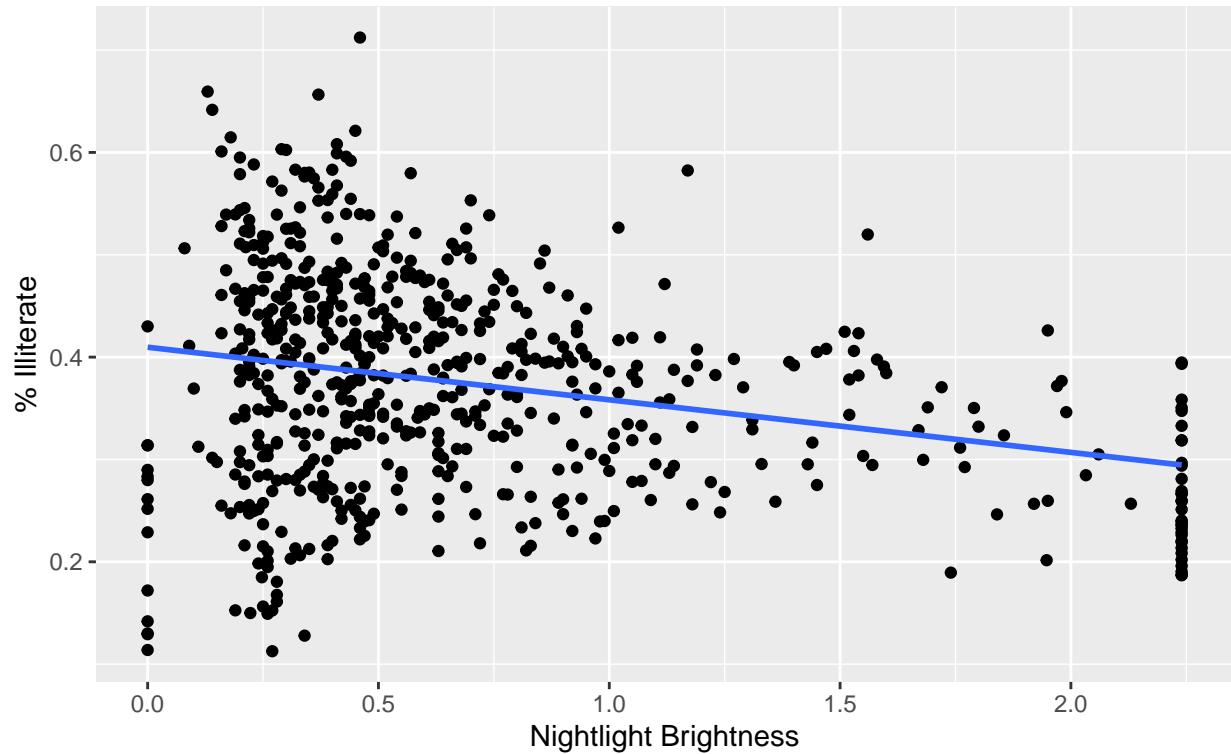
```
india_shape %>%
  ggplot(aes(x=NL, y =PERC_ILLT)) +
  geom_point() +
  stat_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Nightlight Brightness",
    y = "% Illiterate",
    title = "Figure 8: Comparing Nightlight Brightness with Illiteracy Rates",
    subtitle = "By District in India"
  )
```

```
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 1 rows containing non-finite values (stat_smooth).

## Warning: Removed 1 rows containing missing values (geom_point).
```

Figure 8: Comparing Nightlight Brightness with Illiteracy Rates  
By District in India



```
state_gdp <- read_csv("state_gdp.csv")
```

```
##  
## -- Column specification -----  
## cols(  
##   StateName = col_character(),  
##   GDP = col_double()  
## )  
  
df_avg_nl <- india_shape %>%  
  tibble() %>%  
  select(STATE_UT, NL) %>%  
  group_by(STATE_UT) %>%  
  mutate(avg_state_nl = mean(NL)) %>%  
  ungroup() %>%  
  rename(StateName = STATE_UT) %>%  
  select(StateName, avg_state_nl) %>%  
  distinct() %>%  
  mutate(StateName =  
    if_else(StateName == "Tamilnadu",  
           "Tamil Nadu",  
           StateName)) %>%  
  full_join(state_gdp) %>%  
  filter(StateName != "India")  
  
## Joining, by = "StateName"
```

```

model_2f <- lm(GDP ~ avg_state_nl, data = df_avg_nl)

summary(model_2f)

## 
## Call:
## lm(formula = GDP ~ avg_state_nl, data = df_avg_nl)
## 
## Residuals:
##    Min     1Q Median     3Q    Max 
## -271871 -224664 -95654 100367 1027159 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 230435     89302    2.580   0.0152 *  
## avg_state_nl 26877      99878    0.269   0.7898    
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
## 
## Residual standard error: 297600 on 29 degrees of freedom
##   (6 observations deleted due to missingness)
## Multiple R-squared:  0.002491,  Adjusted R-squared:  -0.03191 
## F-statistic: 0.07241 on 1 and 29 DF,  p-value: 0.7898

```

```

df_avg_nl %>%
  ggplot(aes(x=avg_state_nl, y =GDP)) +
  geom_point() +
  stat_smooth(method = "lm", se = FALSE) +
  labs(
    x = "Nightlight Brightness",
    y = "GDP",
    title = "Figure 9: Comparing Nightlight Brightness with GDP",
    subtitle = "By State in India"
  )

```

```

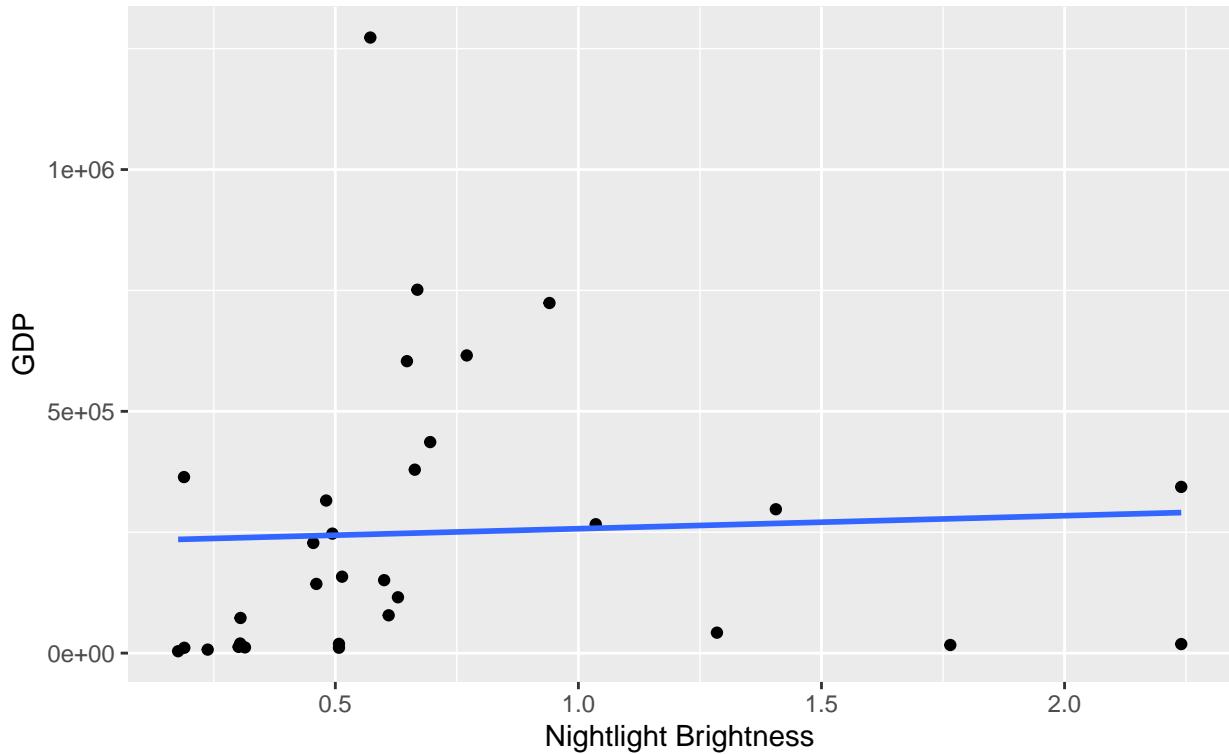
## `geom_smooth()` using formula 'y ~ x'

## Warning: Removed 6 rows containing non-finite values (stat_smooth).

## Warning: Removed 6 rows containing missing values (geom_point).

```

Figure 9: Comparing Nightlight Brightness with GDP  
By State in India



```
df_avg_nl_darker <- df_avg_nl %>% filter(avg_state_nl < 1)

model_2f_darker <- lm(GDP ~ avg_state_nl, data = df_avg_nl_darker)

summary(model_2f_darker)
```

```
## 
## Call:
## lm(formula = GDP ~ avg_state_nl, data = df_avg_nl_darker)
## 
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -305013 -100423  -26967   30259  925166 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -189266     141493  -1.338  0.19409    
## avg_state_nl  938929     268730   3.494  0.00196 ** 
## --- 
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 261200 on 23 degrees of freedom
##   (2 observations deleted due to missingness)
## Multiple R-squared:  0.3467, Adjusted R-squared:  0.3183 
## F-statistic: 12.21 on 1 and 23 DF,  p-value: 0.001957
```

```

stargazer(model_2f, model_2f_darker,
           title = "Indian State GDP Correlation with Nightlight Brightness",
           covariate.labels = "Avg. State Nightlight Brightness",
           dep.var.labels = "State GDP",
           column.labels = c("All States", "States With Brightness < 1"),
           type = "latex")

```

% Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu  
% Date and time: Wed, Apr 28, 2021 - 10:41:05 PM

Table 3: Indian State GDP Correlation with Nightlight Brightness

	<i>Dependent variable:</i>	
	State GDP	
	All States	States With Brightness < 1
	(1)	(2)
Avg. State Nightlight Brightness	26,877.060 (99,877.550)	938,928.900*** (268,729.700)
Constant	230,434.500** (89,301.700)	-189,265.900 (141,492.900)
Observations	31	25
R <sup>2</sup>	0.002	0.347
Adjusted R <sup>2</sup>	-0.032	0.318
Residual Std. Error	297,581.300 (df = 29)	261,198.900 (df = 23)
F Statistic	0.072 (df = 1; 29)	12.208*** (df = 1; 23)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Questions: can my entity fixed effect for 2c just be a country and not the average gdp. The results seem the same

iS “NL” nightlight?