Assignment_One

Evan Trowbridge

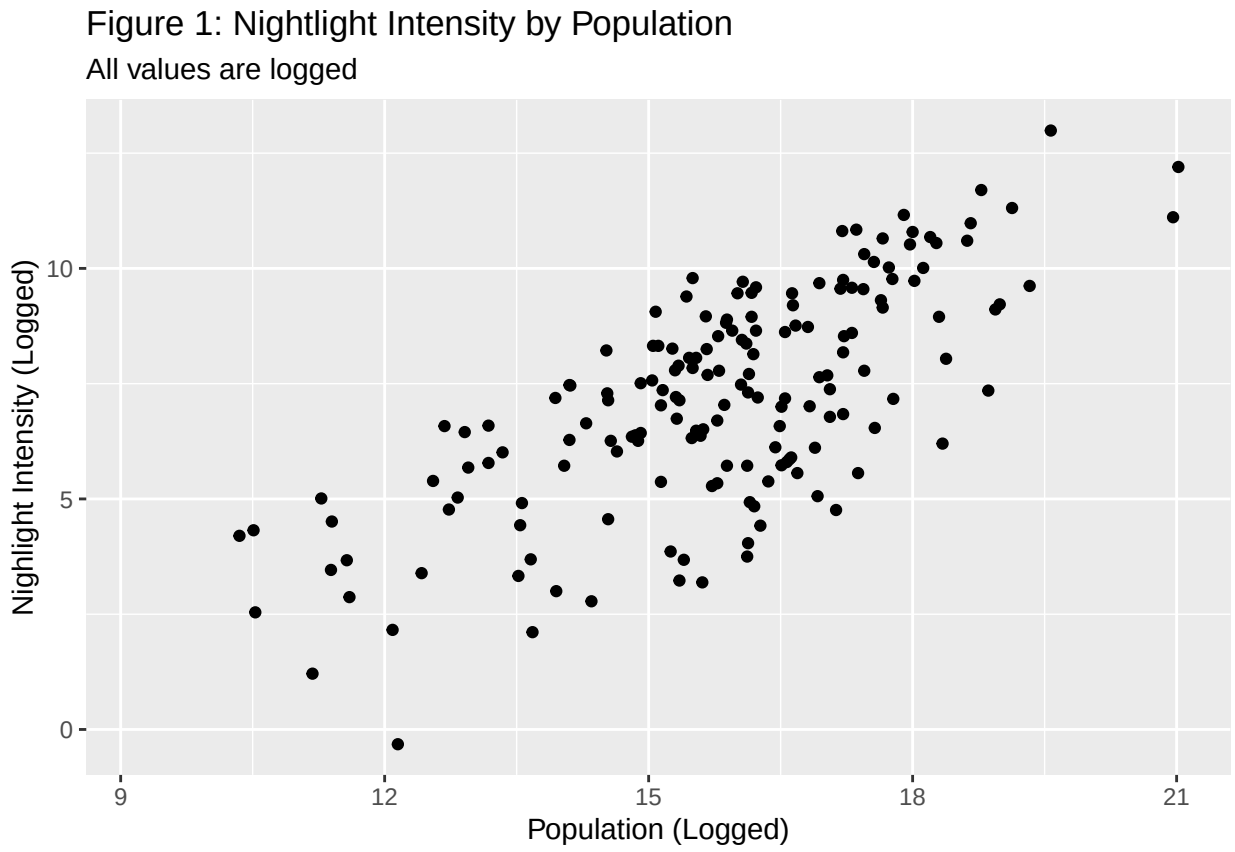
4/24/2021

```
library(readr)
library(janitor)
library(tidyverse)
library(skimr) # For reviewing data
library(plm)
library(stargazer)
```

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 = "Nighlight Intensity (Logged)",
    title = "Figure 1: Nightlight Intensity by Population",
    subtitle = "All values are logged"
  )
```

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

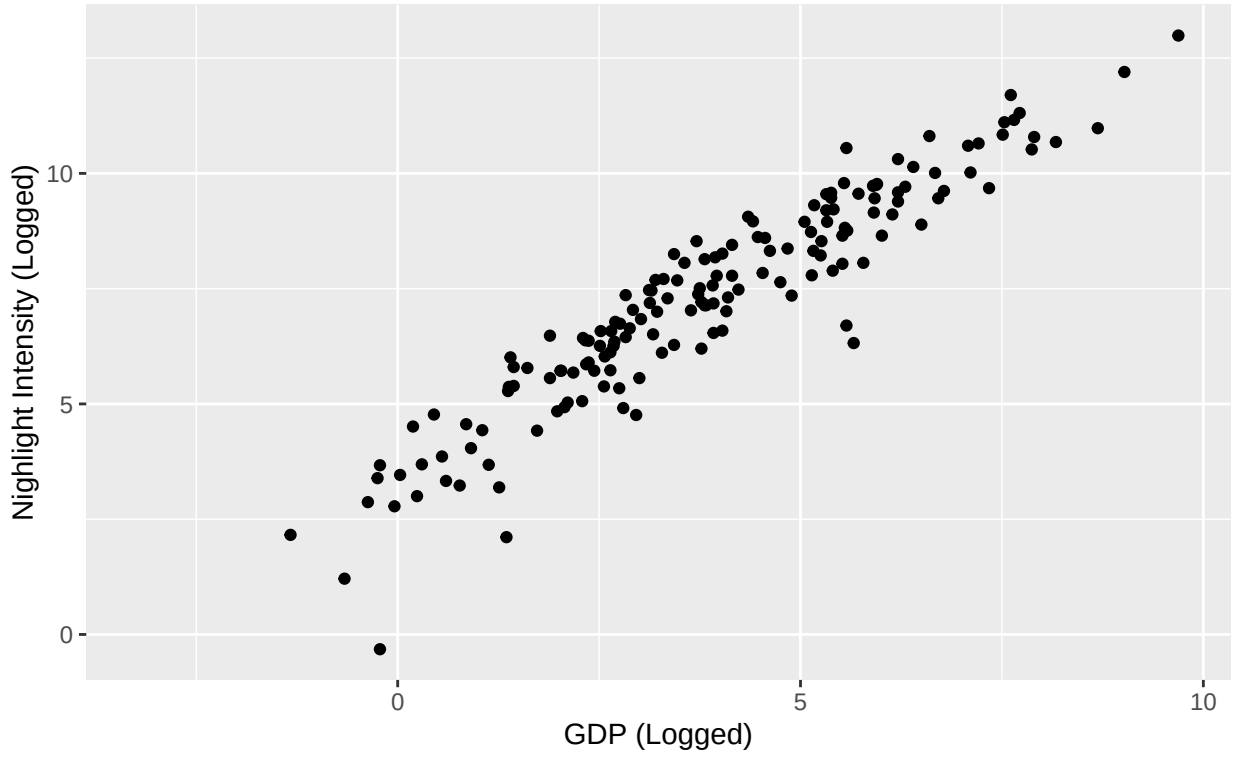



```
# Figure 2: GDP
df_2a %>%
  ggplot(aes(x = log_gdp, y = log_dmsp)) +
  geom_point() +
  labs(
    x = "GDP (Logged)",
    y = "Nighlight Intensity (Logged)",
    title = "Figure 2: Nightlight Intensity by GDP",
    subtitle = "All values are logged"
  )
```

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

Figure 2: Nightlight Intensity by GDP

All values are logged

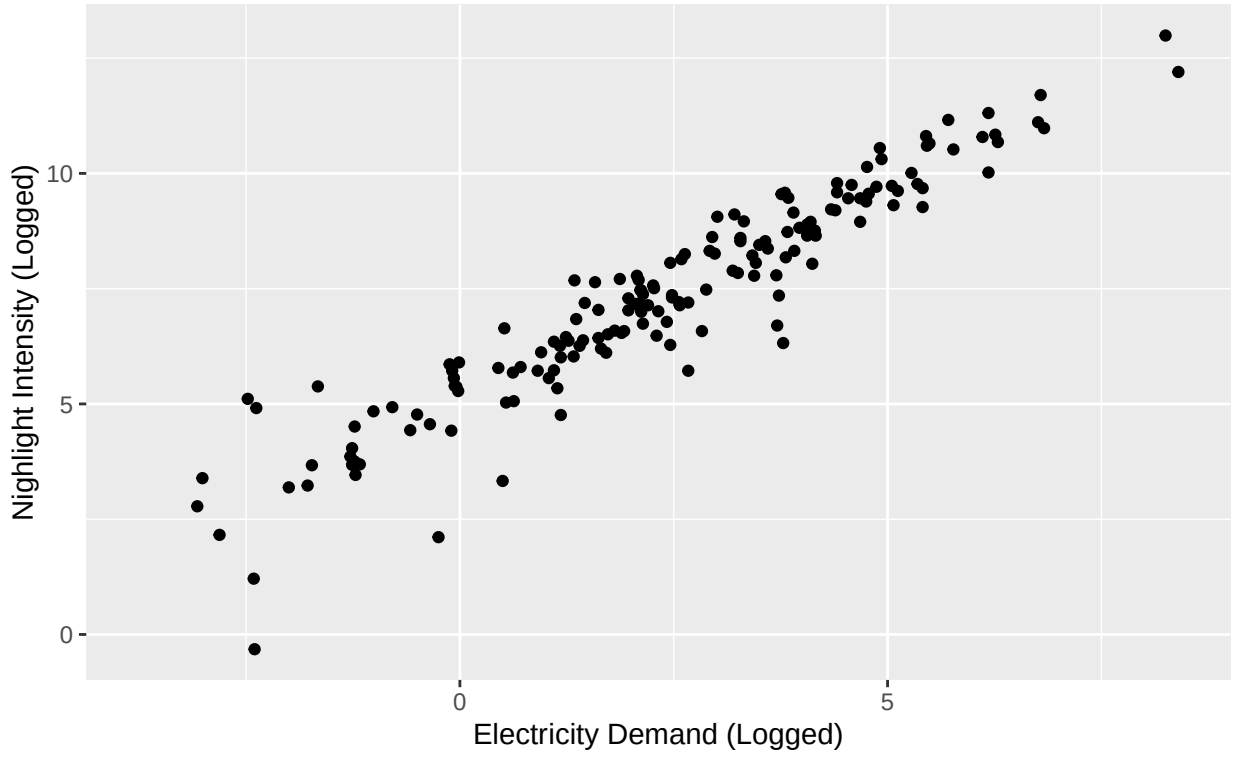



```
# Figure 2: Electricity Demand
df_2a %>%
  ggplot(aes(x = log_elec, y = log_dmsp)) +
  geom_point() +
  labs(
    x = "Electricity Demand (Logged)",
    y = "Nighlight Intensity (Logged)",
    title = "Figure 3: Nightlight Intensity by Electricity Demand",
    subtitle = "All values are logged"
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

```
## 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 may 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 indexed dataframe
```

```r
p_2b <- pdata.frame(df_2b, index  = c("date")) %>%
  filter(
    !is.na(log_gdp) &
    !is.na(log_population) &
    !is.na(log_dmsp)) %>%
  select(date, log_gdp, log_population, log_dmsp, 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_dmsp,
               data = p_2b_150,
               model = "within")

summary(gdp_mod)
```

```
## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = log_gdp ~ 0 + log_population + log_dmsp, data = p_2b_150,
##     model = "within")
##
## Unbalanced Panel: n = 19, T = 69-163, N = 2906
##
## Residuals:
##     Min.  1st Qu.   Median  3rd Qu.     Max.
## -2.60275 -0.55330 -0.06189  0.44946  3.44331
##
## Coefficients:
##                 Estimate Std. Error t-value  Pr(>|t|)
## log_population 0.1137698  0.0108646  10.472 < 2.2e-16 ***
## log_dmsp       0.7988370  0.0094053  84.935 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    13599
## Residual Sum of Squares: 1950.1
## R-Squared:      0.8566
## Adj. R-Squared: 0.85561
## F-statistic: 8617.07 on 2 and 2885 DF, p-value: < 2.22e-16
```

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

```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harva
## % Date and time: Wed, Apr 28, 2021 - 1:24:51 PM
```

```
## \begin{table}[!htbp] \centering
##   \caption{Results of 150-Country Sample}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & log\_gdp \\
## \hline \\[-1.8ex]
##  log\_population & 0.114$^{***}$ \\
##   & (0.011) \\
##   & \\
##  log\_dmsp & 0.799$^{***}$ \\
##   & (0.009) \\
##   & \\
## \hline \\[-1.8ex]
## Observations & 2,906 \\
## R$^{2}$ & 0.857 \\
## Adjusted R$^{2}$ & 0.856 \\
## F Statistic & 8,617.072$^{***}$ (df = 2; 2885) \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

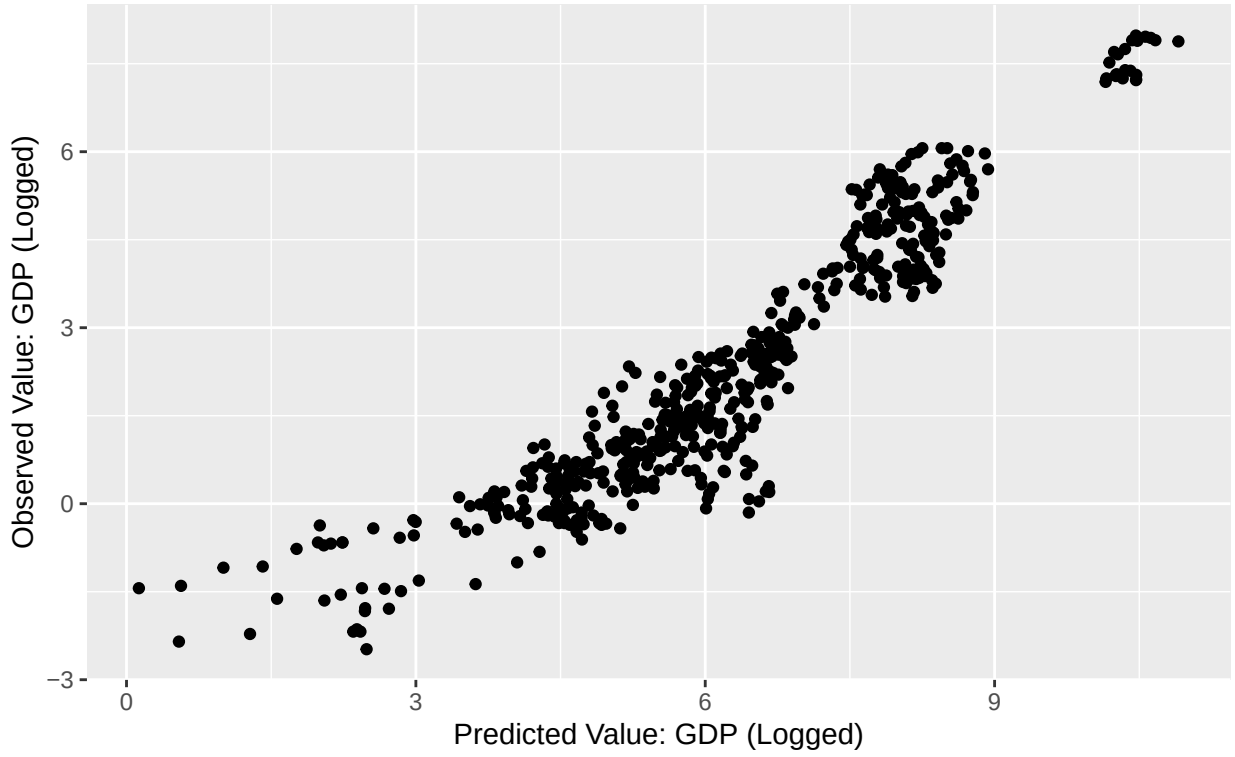
```
p_2b_30 <- predict(gdp_mod, newdata = p_2b_30) %>%
  tibble() %>%
  rename(pred_value = ".") %>%
  bind_cols(p_2b_30)

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 - 1:24:51 PM

Table 1: Results of 150-Country Sample

	Dependent variable:
	log_gdp
log_population	0.114***
	(0.011)
log_dmsp	0.799***
	(0.009)
Observations	2,906
R^2	0.857
Adjusted R^2	0.856
F Statistic	8,617.072*** (df = 2; 2885)
Note:	*p<0.1; **p<0.05; ***p<0.01