Lab 7 /Assignment 3

Task 1

Seperating pre and post 2012 data.

```
df_pre_2012 = pv_df %>%  filter(year < 2012)
df_post_2012 = pv_df %>% filter(year > 2012)
```

Creating the variable cum_cap, which calculates the cumulative sum of nameplate.

```
df_pre_2012 = df_pre_2012 %>% arrange(date) %>% mutate(
  cum_cap = cumsum(nameplate)
)

df_post_2012 = df_post_2012 %>% arrange(date) %>% mutate(
  cum_cap = cumsum(nameplate)
)
```

Removing zero values

```
df_pre_2012 = df_pre_2012 %>% filter(cost_per_kw != 0)
df_post_2012 = df_post_2012 %>% filter(cost_per_kw != 0)
```

Creating the variables *log2_cum_cap* and *log2_cost_per_kw*

```
df_pre_2012["log2_cum_cap"] = log2(df_pre_2012$cum_cap)
df_pre_2012["log2_cost_per_kw"] = log2(df_pre_2012$cost_per_kw)

df_post_2012["log2_cum_cap"] = log2(df_post_2012$cum_cap)
df_post_2012["log2_cost_per_kw"] = log2(df_post_2012$cost_per_kw)
```

Linear models pre and post 2012.

```
learning_mod_pre = lm(log2_cost_per_kw~log2_cum_cap, data = df_pre_2012)
summary(learning_mod_pre)
```

```
##
## Call:
## lm(formula = log2_cost_per_kw ~ log2_cum_cap, data = df_pre_2012)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2249 -0.2209 -0.0405  0.2050  3.7030
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  13.855415   0.018759  738.61   <2e-16 ***
## log2_cum_cap -0.062405   0.001121  -55.65   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3449 on 45087 degrees of freedom
## Multiple R-squared:  0.06426,    Adjusted R-squared:  0.06424
## F-statistic:  3096 on 1 and 45087 DF,  p-value: < 2.2e-16
```

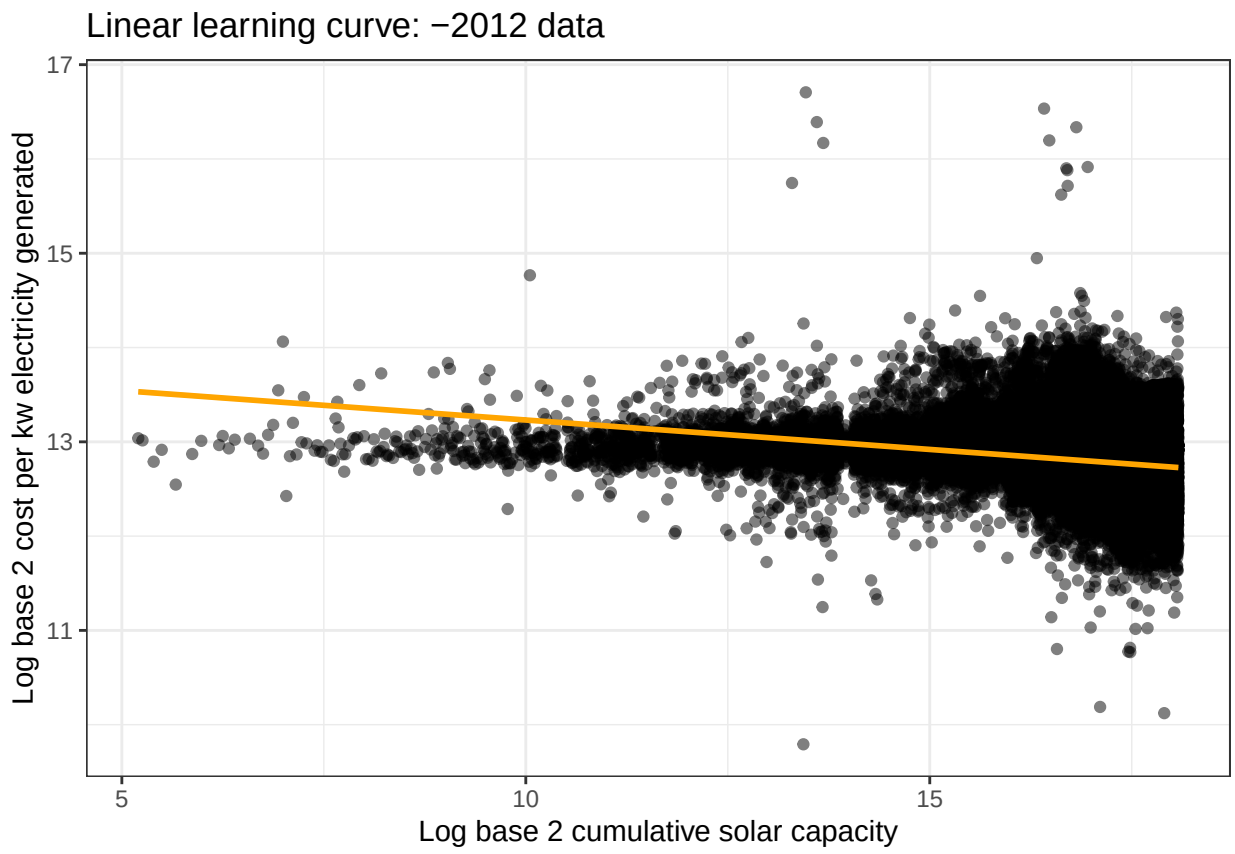
```
learning_mod_post = lm(log2_cost_per_kw~log2_cum_cap, data = df_post_2012)
summary(learning_mod_post)
```

```
##
## Call:
## lm(formula = log2_cost_per_kw ~ log2_cum_cap, data = df_post_2012)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2028 -0.1447  0.0288  0.1166  2.4923
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.6596888  0.0160133  790.58   <2e-16 ***
## log2_cum_cap -0.0251671  0.0009516  -26.45   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2848 on 39885 degrees of freedom
## Multiple R-squared:  0.01723,    Adjusted R-squared:  0.01721
## F-statistic: 699.4 on 1 and 39885 DF,  p-value: < 2.2e-16
```

Plotting linear learning curves pre and post 2012.

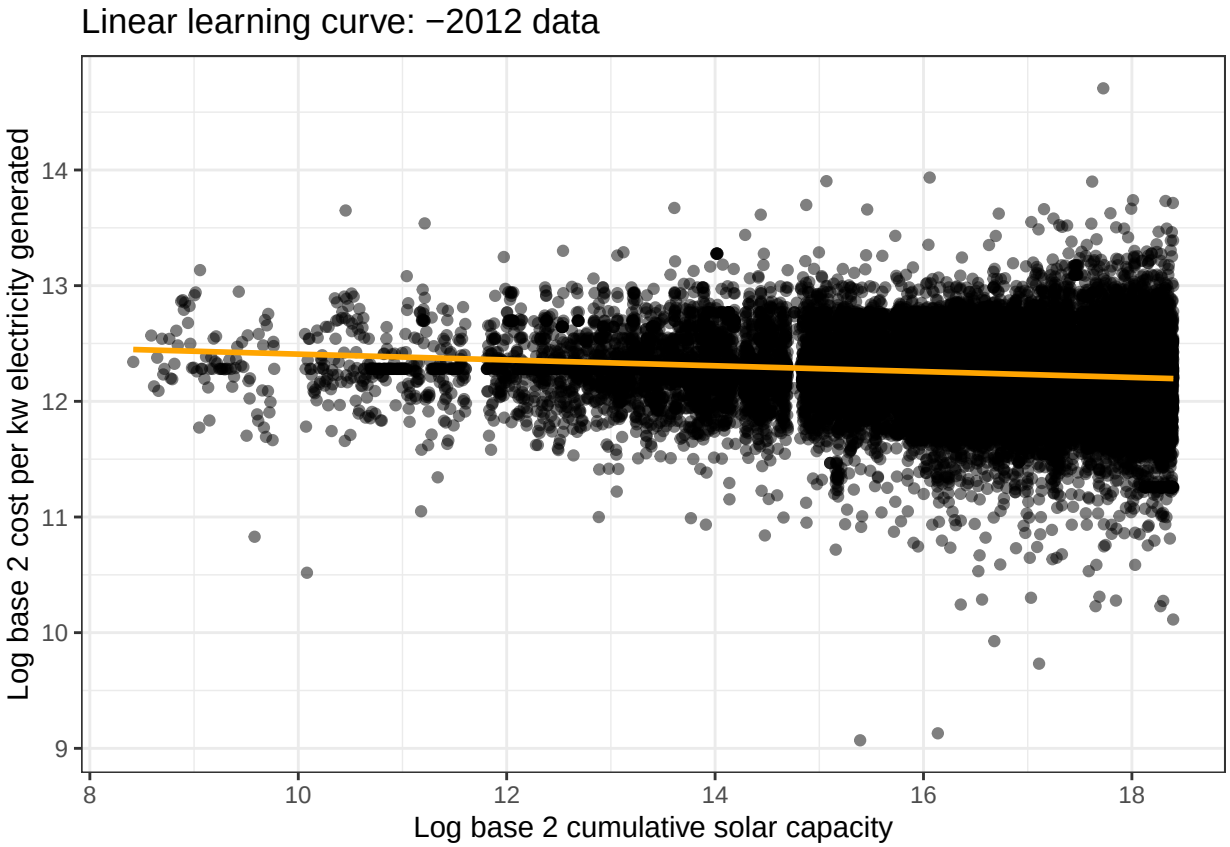
```
df_pre_2012  %>%  ggplot(aes(x = log2_cum_cap, y = log2_cost_per_kw)) +
  geom_point(alpha=.5, color = "black") +
  geom_smooth(method = "lm", color = "orange") +
  labs(title = "Linear learning curve: -2012 data",
       x     = "Log base 2 cumulative solar capacity",
       y     = "Log base 2 cost per kw electricity generated") +
  theme_bw()
```

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




```
df_post_2012  %>%  ggplot(aes(x = log2_cum_cap, y = log2_cost_per_kw)) +
  geom_point(alpha=.5, color = "black") +
  geom_smooth(method = "lm", color = "orange") +
  labs(title = "Linear learning curve: -2012 data",
       x     = "Log base 2 cumulative solar capacity",
       y     = "Log base 2 cost per kw electricity generated") +
  theme_bw()
```

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


Linear learning curve: −2012 data
Log base 2 cost per kw electricity generated
14
13
12
11
10
9
8
10
12
14
16
18
Log base 2 cumulative solar capacity