EC 502 HW3 R Notebook

Laya Dang (pd03@bu.edu)

Data Set-up

Import tidyverse

```
suppressPackageStartupMessages({
  library(tidyverse)
  library(ggplot2)
})
```

Load the dataset:

```
country_data = read.csv("cross_country_data.csv")
head(country_data, 5)
```

```
##   Country Year SavingsRate   RealGDP LaborForce HumanCapital PhysicalCapital
## 1 Albania 1980   0.3934749  9981.908  0.9249071     2.395581        42834.52
## 2 Albania 1981   0.4208891 10555.427  0.9514943     2.445612        45231.09
## 3 Albania 1982   0.4482948 10866.670  0.9771618     2.496687        47943.00
## 4 Albania 1983   0.4195175 10986.740  1.0006725     2.548830        50831.43
## 5 Albania 1984   0.3768510 10849.229  1.0246358     2.588237        53751.48
```

Calculate (real) GDP per worker in 1981 and 2010:

```
country_data <- country_data %>%
  transform(output_per_worker = RealGDP / LaborForce)

country_data_gdp_per_worker_1981_2010 <- country_data %>%
  filter(Year %in% c(1981, 2010)) %>%
  select(Country, Year, output_per_worker) %>%
  pivot_wider(names_from = Year,
              values_from = output_per_worker,
              names_prefix = "output_per_worker_")

head(country_data_gdp_per_worker_1981_2010, 5)
```

```
## # A tibble: 5 x 3
##   Country   output_per_worker_1981 output_per_worker_2010
```

```
##   <chr>                        <dbl>                 <dbl>
## 1 Albania                     11094.                24309.
## 2 Argentina                   24359.                37175.
## 3 Australia                   52401.                72687.
## 4 Austria                     49189.                71226.
## 5 Bahrain                     74160.                43394.
```

Average output per worker and average savings rate:

```
summary_country_data <- country_data %>%
  group_by(Country) %>%
  summarise_at(vars(SavingsRate), list(savings_mean = mean))

head(summary_country_data, 5)
```

```
## # A tibble: 5 x 2
##   Country   savings_mean
##   <chr>            <dbl>
## 1 Albania          0.279
## 2 Argentina        0.185
## 3 Australia        0.287
## 4 Austria          0.280
## 5 Bahrain          0.259
```

Average growth rate of the labor force and average growth rate of GDP per worker (from 1981 to 2010):

t = 1981

T = 2010 - 1981 = 29

formula: 1/29 * ln(val at 2010 / val at 1981)

(log in R has base e)

```
# GET SAVINGS MEAN
summary_country_data <- country_data %>%
  group_by(Country) %>%
  summarise_at(vars(SavingsRate), list(savings_mean = mean))

# GET AVG GROWTH RATE OF LABOR (n) and AVG GROWTH RATE OF GDP PER WORKER (g)
summary_country_data <- country_data %>%
  filter(Year %in% c(1981, 2010)) %>%
  select(Country, Year, LaborForce, output_per_worker) %>%
  pivot_wider(names_from = Year,
              values_from = c(output_per_worker, LaborForce),
              names_sep = "_",
              names_prefix = "year_") %>%
  transform(avg_growth_labor = (1/29) * log(LaborForce_year_2010 / LaborForce_year_1981)) %>%
  transform(avg_growth_gdp_per_worker = (1/29) * log(output_per_worker_year_2010 / output_per_worker_yea
  left_join(summary_country_data, by = "Country")

head(summary_country_data %>% select(Country, avg_growth_labor, avg_growth_gdp_per_worker, savings_mean
```

```
##     Country avg_growth_labor avg_growth_gdp_per_worker savings_mean
## 1   Albania     -0.001273320                0.02705051    0.2787880
## 2 Argentina      0.012385833                0.01457744    0.1849724
## 3 Australia      0.020416269                0.01128414    0.2874728
## 4   Austria      0.009319133                0.01276498    0.2803114
## 5   Bahrain      0.055696186               -0.01847930    0.2586320
```

Question (1) Descriptive Statistics

a) number of countries

```r
country_data %>%  summarise(num_countries = n_distinct(Country))
```

```
##   num_countries
## 1           118
```

b) mean (across countries) of output per worker 1981, output per worker 2010, savings rate, growth of labor force, growth rate of GDP per worker

```r
summary_country_data %>%
  summarise_at(vars(output_per_worker_year_1981,
                    output_per_worker_year_2010,
                    savings_mean,
                    avg_growth_labor,
                    avg_growth_gdp_per_worker),
               list(mean = mean)) %>%
  mutate(across(everything(), ~ round(., digits = 4))) %>%
  pivot_longer(cols=everything(), names_to = "variable", values_to = "value")
```

```
## # A tibble: 5 x 2
##   variable                            value
##   <chr>                               <dbl>
## 1 output_per_worker_year_1981_mean 27043.
## 2 output_per_worker_year_2010_mean 34551.
## 3 savings_mean_mean                   0.200
## 4 avg_growth_labor_mean               0.0225
## 5 avg_growth_gdp_per_worker_mean      0.0113
```

c) standard deviation of vals above

```r
summary_country_data %>%
  summarise_at(vars(output_per_worker_year_1981,
                    output_per_worker_year_2010,
                    savings_mean,
                    avg_growth_labor,
                    avg_growth_gdp_per_worker), list(sd = sd)) %>%
  mutate(across(everything(), ~ round(., digits = 4))) %>%
  pivot_longer(cols=everything(), names_to = "variable", values_to = "value")
```

```
## # A tibble: 5 x 2
##   variable                           value
##   <chr>                              <dbl>
## 1 output_per_worker_year_1981_sd 37953.
## 2 output_per_worker_year_2010_sd 33102.
## 3 savings_mean_sd                  0.0688
## 4 avg_growth_labor_sd              0.0131
## 5 avg_growth_gdp_per_worker_sd     0.0181
```

d) min and max of vals above

Min:

```
summary_country_data %>%
  summarise_at(vars(output_per_worker_year_1981,
                    output_per_worker_year_2010,
                    savings_mean,
                    avg_growth_labor,
                    avg_growth_gdp_per_worker),
               list(min = min)) %>%
  mutate(across(everything(), ~ round(., digits = 4))) %>%
  pivot_longer(cols=everything(), names_to = "variable", values_to = "value")
```

```
## # A tibble: 5 x 2
##   variable                          value
##   <chr>                             <dbl>
## 1 output_per_worker_year_1981_min 887.
## 2 output_per_worker_year_2010_min 968.
## 3 savings_mean_min                  0.0479
## 4 avg_growth_labor_min             -0.0141
## 5 avg_growth_gdp_per_worker_min    -0.0347
```

Max:

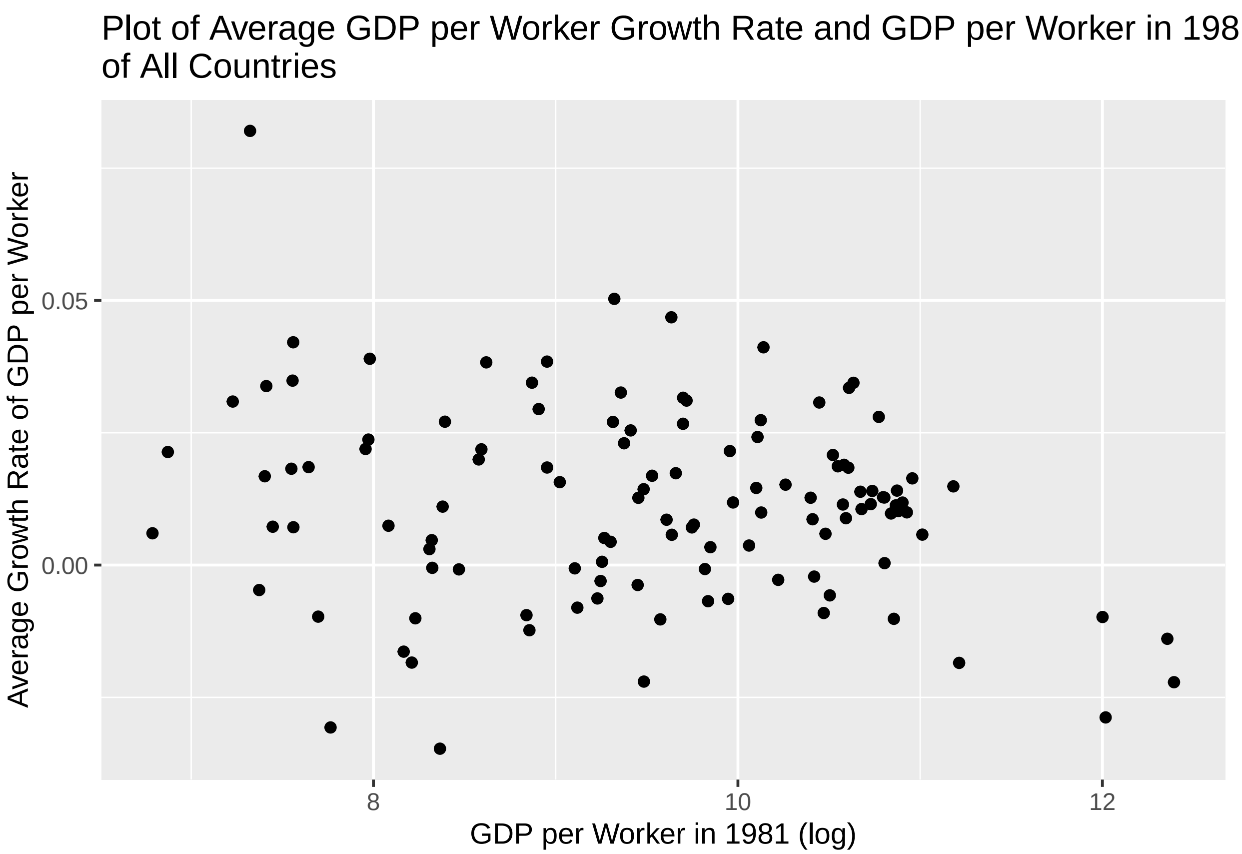
```
summary_country_data %>%
  summarise_at(vars(output_per_worker_year_1981,
                    output_per_worker_year_2010,
                    savings_mean,
                    avg_growth_labor,
                    avg_growth_gdp_per_worker),
               list(max = max)) %>%
  mutate(across(everything(), ~ round(., digits = 4))) %>%
  pivot_longer(cols=everything(), names_to = "variable", values_to = "value")
```

```
## # A tibble: 5 x 2
##   variable                             value
##   <chr>                                <dbl>
## 1 output_per_worker_year_1981_max 241086.
## 2 output_per_worker_year_2010_max 155184.
## 3 savings_mean_max                     0.446
## 4 avg_growth_labor_max                 0.083
## 5 avg_growth_gdp_per_worker_max        0.082
```

Question (2) Unconditional Convergence

Plot:

```
summary_country_data %>%
  ggplot(aes(x=log(output_per_worker_year_1981), y=avg_growth_gdp_per_worker)) +
  geom_point() +
  labs(title="Plot of Average GDP per Worker Growth Rate and GDP per Worker in 1981\nof All Countries "
        x ="GDP per Worker in 1981 (log)", y = "Average Growth Rate of GDP per Worker")
```



OLS Regression:

```
ols_data <- data.frame(
  # y value:
  avg_growth_gdp_per_worker = summary_country_data[['avg_growth_gdp_per_worker']],
  # x value:
  log_output_per_worker_year_1981 = log(summary_country_data[['output_per_worker_year_1981']])
)

head(ols_data, 5)
```

```
##   avg_growth_gdp_per_worker log_output_per_worker_year_1981
## 1                0.02705051                        9.314117
## 2                0.01457744                       10.100645
## 3                0.01128414                       10.866678
## 4                0.01276498                       10.803426
## 5               -0.01847930                       11.213982
```

```r
ols_model <- lm(avg_growth_gdp_per_worker~log_output_per_worker_year_1981, data=ols_data)
summary(ols_model)
```

```
##
## Call:
## lm(formula = avg_growth_gdp_per_worker ~ log_output_per_worker_year_1981,
##     data = ols_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04915 -0.01204  0.00226  0.01044  0.06475
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      0.037238   0.012788   2.912  0.00431 **
## log_output_per_worker_year_1981 -0.002724   0.001332  -2.045  0.04312 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0179 on 116 degrees of freedom
## Multiple R-squared:  0.0348, Adjusted R-squared:  0.02648
## F-statistic: 4.182 on 1 and 116 DF,  p-value: 0.04312
```

From the above OLS model, we get the formula where g = 0.037238 + -0.002724 * log(Y/L) + 0.0179

The estimate for b1(hat) is -0.002724 and its standard error is 0.001332

This implies that the living standards in nations that are initially poorer (lower GDP per worker in 1981) fall behind compared to wealthier nations, indicated by the negative b1 coefficient. However, the low R-squared value (3.48%) suggests that while GDP per worker in 1981 does have some predictive power, there are likely other important factors not included in the model.

Question (3) MRW Revisited

```r
ols_data_mrw <- data.frame(
  # y value:
  avg_growth_gdp_per_worker = summary_country_data[['avg_growth_gdp_per_worker']],
  # x1 value:
  log_output_per_worker_year_1981 = log(summary_country_data[['output_per_worker_year_1981']]),
  # x2 value:
  log_avg_growth_labor_plus_0.05 = log(summary_country_data[['avg_growth_labor']] + 0.05 )
)

head(ols_data_mrw, 5)
```

```
##   avg_growth_gdp_per_worker log_output_per_worker_year_1981
## 1                0.02705051                        9.314117
## 2                0.01457744                       10.100645
## 3                0.01128414                       10.866678
## 4                0.01276498                       10.803426
## 5               -0.01847930                       11.213982
```

```
##   log_avg_growth_labor_plus_0.05
## 1                      -3.021529
## 2                      -2.774417
## 3                      -2.653331
## 4                      -2.824823
## 5                      -2.247186
```

```
ols_model_mrw <- lm(avg_growth_gdp_per_worker~log_output_per_worker_year_1981+log_avg_growth_labor_plus_
summary(ols_model_mrw)
```

```
##
## Call:
## lm(formula = avg_growth_gdp_per_worker ~ log_output_per_worker_year_1981 +
##     log_avg_growth_labor_plus_0.05, data = ols_data_mrw)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.047635 -0.009087 -0.001348  0.008121  0.059462
##
## Coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -0.062898   0.022800  -2.759  0.00675 **
## log_output_per_worker_year_1981 -0.003519   0.001218  -2.889  0.00462 **
## log_avg_growth_labor_plus_0.05  -0.040776   0.007993  -5.101 1.34e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01623 on 115 degrees of freedom
## Multiple R-squared:  0.2129, Adjusted R-squared:  0.1992
## F-statistic: 15.55 on 2 and 115 DF,  p-value: 1.051e-06
```

The estimate for b1(hat) is -0.003519 and its standard error is 0.001218, which implies a higher impact of savings on living standards than before, as this value is more negative.

The value of alpha is implied to be 0.003519, which means labor share is around (1 - 0.003519) 0.996481, or the majority of output. This is not consistent with the standard assumption of 2/3.

The R-squared is 0.2129, which indicates that the model does not do a good job of indicating the variance in the model, and there is not much predictive power the coefficients have. The model only explains 21.29% of the variance.

Question (4) HJ Levels Accounting

A = (Y/L) * (K/Y)^(1-a/a) * (L/H)

where a = 1/3:

A = (Y/L) * (K/Y)^(2/9) * (L/H)

Calculate parts of the equation:

```
hj_data <- country_data %>%
  filter(Year == 2010) %>%
  transform(capital_per_output = PhysicalCapital / RealGDP) %>%  # capital per output in 2010
```

```r
  transform(recip_human_cap = 1 / HumanCapital) %>%  # reciprocal of human capital per person
  select(Country, Year, capital_per_output, recip_human_cap, HumanCapital)

head(hj_data %>% select(-HumanCapital), 5)
```

```
##     Country Year capital_per_output recip_human_cap
## 1   Albania 2010           4.034723       0.3328644
## 2 Argentina 2010           3.398882       0.3547817
## 3 Australia 2010           3.613600       0.2951154
## 4   Austria 2010           3.385438       0.3521654
## 5   Bahrain 2010           3.982973       0.3500738
```

Estimate A:

```r
hj_data <- hj_data %>%
  left_join(summary_country_data %>% select(Country, output_per_worker_year_2010), by="Country") %>%
  transform(A_estimate = output_per_worker_year_2010 * capital_per_output^(2/9) * recip_human_cap) %>%
  transform(log_A_estimate = log(A_estimate))

head(hj_data %>%  select(A_estimate, log_A_estimate), 5)
```

```
##   A_estimate log_A_estimate
## 1   11031.92       9.308548
## 2   17309.58       9.759015
## 3   28538.78      10.259019
## 4   32890.93      10.400952
## 5   20652.43       9.935588
```

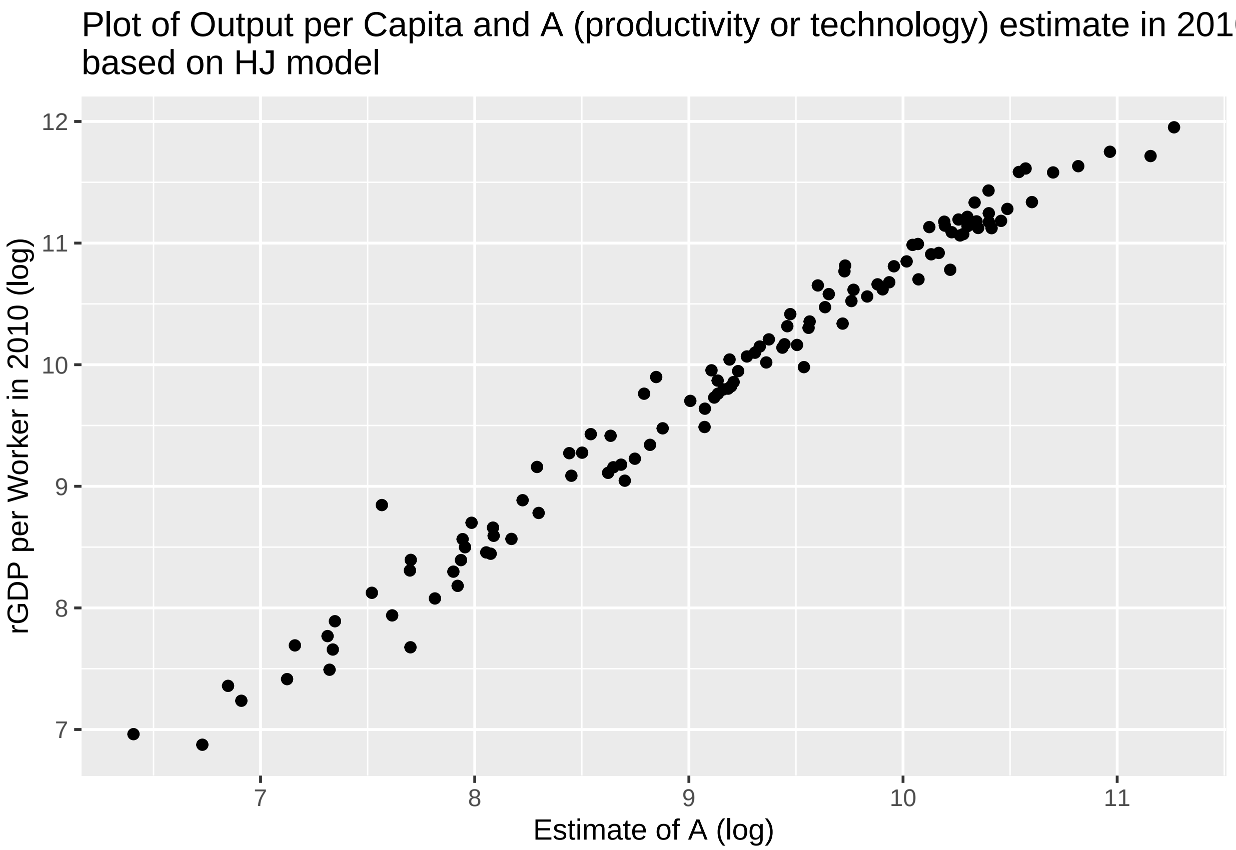
Summary statistics:

```r
hj_data %>%
  summarise_at(vars(log_A_estimate), list(log_A_mean = mean, log_A_sd = sd))
```

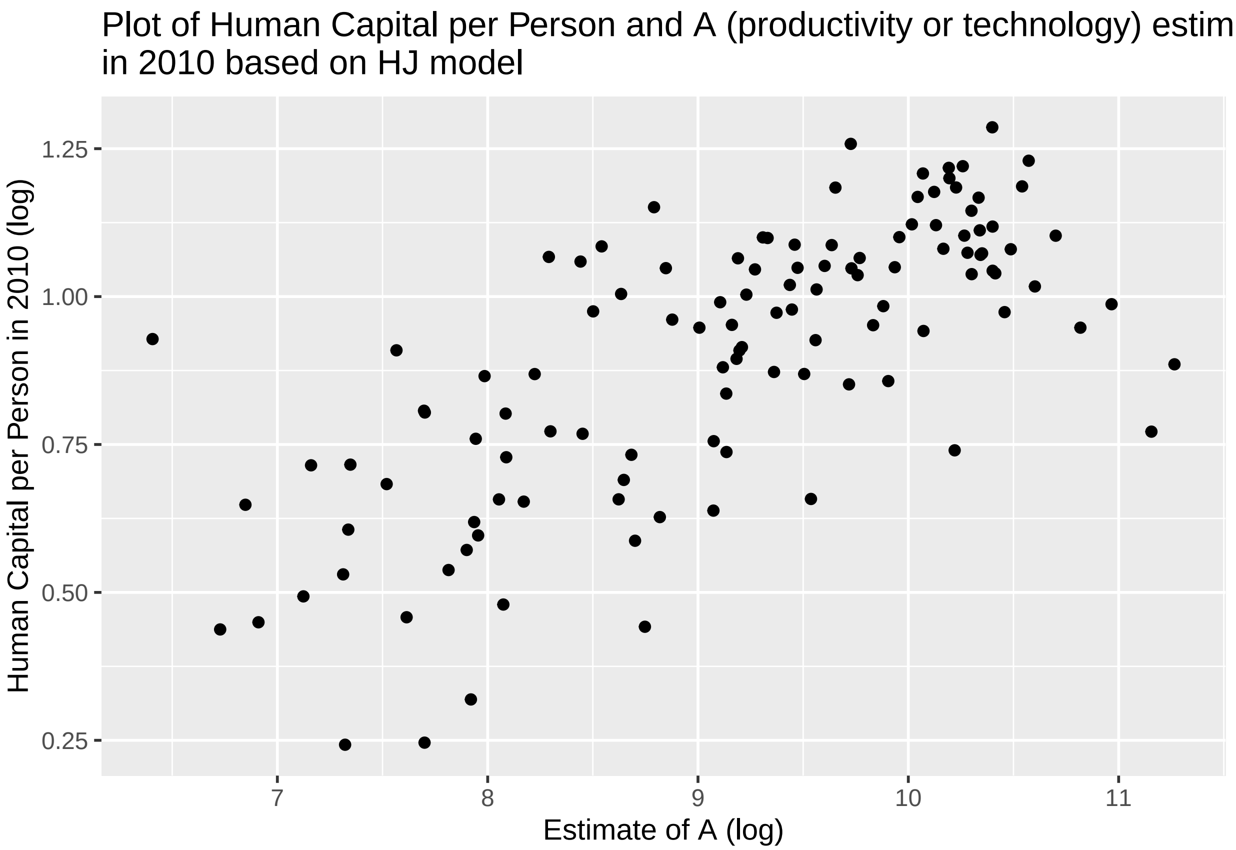
```
##   log_A_mean log_A_sd
## 1   9.155138 1.105463
```

Plots:

```r
hj_data %>%
  ggplot(aes(x=log_A_estimate, y=log(output_per_worker_year_2010))) +
  geom_point() +
  labs(title="Plot of Output per Capita and A (productivity or technology) estimate in 2010 \nbased on H
        x ="Estimate of A (log)", y = "rGDP per Worker in 2010 (log)")
```




```
hj_data %>%
  ggplot(aes(x=log_A_estimate, y=log(HumanCapital))) +
  geom_point() +
  labs(title="Plot of Human Capital per Person and A (productivity or technology) estimate \nin 2010 ba
        x ="Estimate of A (log)", y = "Human Capital per Person in 2010 (log)")
```




```r
hj_ols_model <- lm(log(output_per_worker_year_2010)~log_A_estimate, data=hj_data)
summary(hj_ols_model)
```

```
##
## Call:
## lm(formula = log(output_per_worker_year_2010) ~ log_A_estimate,
##     data = hj_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53717 -0.11691 -0.00643  0.11088  0.78249
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -0.44376    0.13704  -3.238  0.00157 **
## log_A_estimate  1.12432    0.01486  75.654  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1777 on 116 degrees of freedom
## Multiple R-squared:  0.9801, Adjusted R-squared:   0.98
## F-statistic:  5724 on 1 and 116 DF,  p-value: < 2.2e-16
```

The estimated b1(hat), or the coefficient for A, is 1.12432, and since this is a positive coefficient, the model suggests that living standards (Y/L) is higher in nations with higher productivity (A). The R-squared is 0.9801, which means the model is a good fit and explains 98.01% of the variance in the data.