Is Energy in Synergy with Life

All_Of_Us

9/14/2021

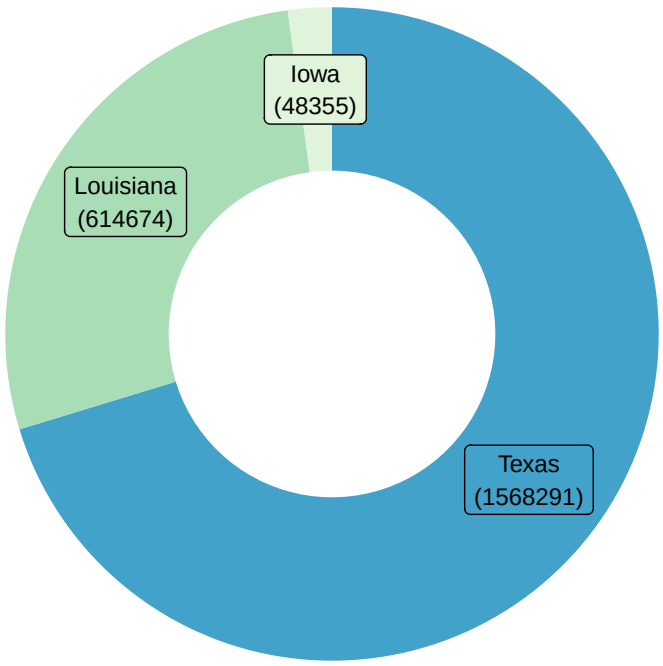
```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.2     v dplyr   1.0.7
## v tidyr   1.1.3     v forcats 0.5.1
## v readr   2.0.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# Reading the energy dataset
energy_data <- read_csv("https://think.cs.vt.edu/corgis/datasets/csv/energy/energy.csv",
                        show_col_types = FALSE)
energy_df <- as.tibble(energy_data)
```

```
## Warning: `as.tibble()` was deprecated in tibble 2.0.0.
## Please use `as_tibble()` instead.
## The signature and semantics have changed, see `?as_tibble`.
```

```
# Creating the arrays for renewable and non renewable resources
consumption_energy_renewable_industrial <- c('Consumption.Industrial.Solar','Consumption.Industrial.Wind
                                             'Consumption.Industrial.Hydropower',
                                             'Consumption.Industrial.Liquefied Petroleum Gases')
consumption_energy_non_renewable_industrial <- c('Consumption.Industrial.Coal','Consumption.Industrial.

# Data Manipulation for Visualization
energy_mutate_df <- energy_df %>% mutate(Total_Consumption_renewable_industrial = rowSums(energy_df[cons
Total_Consumption_non_renewable_industrial = rowSums(energy_df[consumption_energy_non_renewable_industri


# Manipulating the data for Consumption of Renewable Resources in Industrial sector(Donut Chart)
top3_Consumption_renewable_industrial <- energy_mutate_df %>% filter(Year == 2014) %>% slice_max(Total_
  select('State','Year','Consumption.Industrial.Solar','Consumption.Industrial.Wind','Consumption.Indust

# Visualizing the data in the form of Donut chart
data <- top3_Consumption_renewable_industrial
data_per_thou <- data$Total_Consumption_renewable_industrial
data$fraction <- data_per_thou / sum(data_per_thou)
# Compute the cumulative percentages (top of each rectangle)
```

```
data$ymax <- cumsum(data$fraction)
# Compute the bottom of each rectangle
data$ymin <- c(0, head(data$ymax, n=-1))
# Compute label position
data$labelPosition <- (data$ymax + data$ymin) / 2
# Compute a good label
data$label <- paste0(data$State,"\n(", data_per_thou,")")
# Make the plot
 top3_states_rene_chart <- ggplot(data, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=State)) +
  geom_rect() +
  geom_label( x=3.5, aes(y=labelPosition, label=label), size=3) +
  scale_fill_brewer(palette=4) +
  coord_polar(theta="y") +
  xlim(c(2, 4)) + ggtitle("Top 3 States Consumption of Renewable resources(in BTU)")+ theme_void() +
  theme(legend.position = "none")
top3_states_rene_chart
```

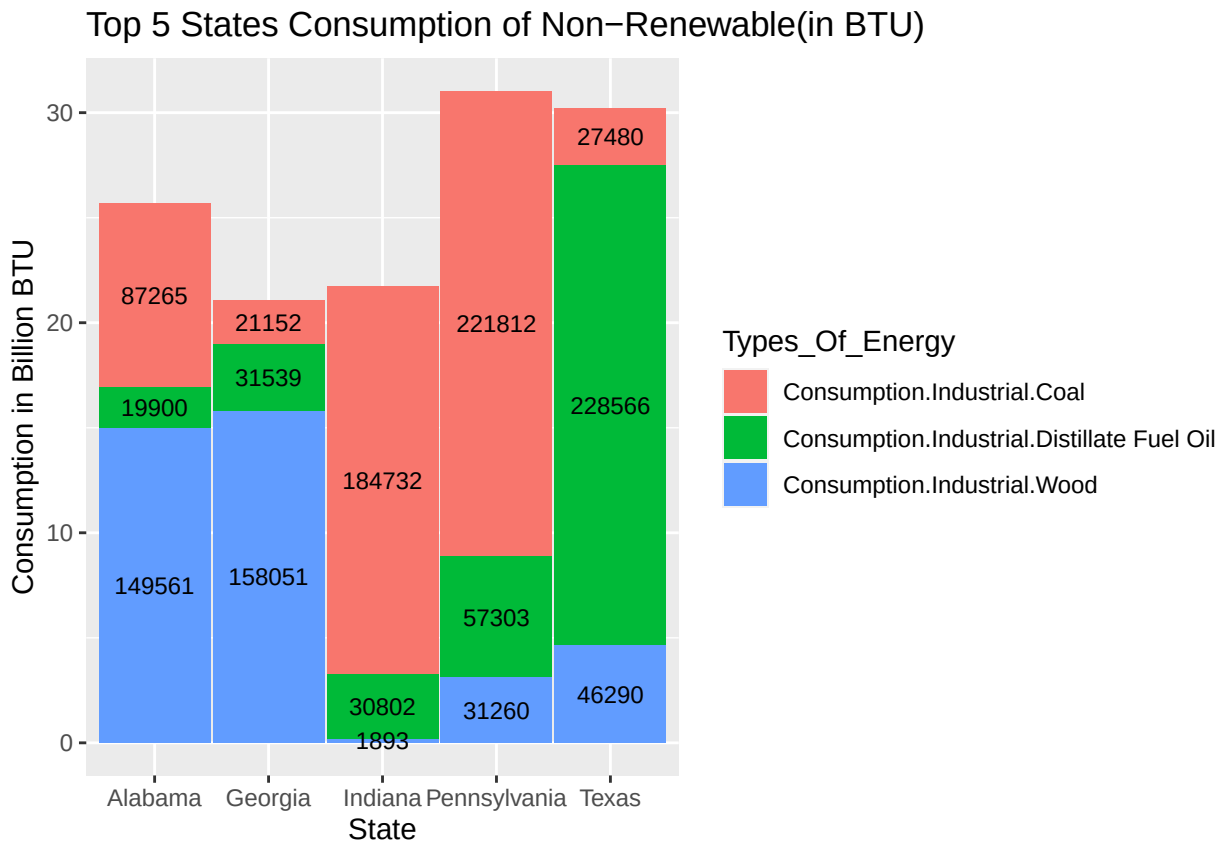



```
# Manipulating the data for Consumption of Non-Renewable Resources in Industrial sector
top5_Consumption_non_renewable_industrial <- energy_mutate_df %>% filter(Year == 2014) %>%
  slice_max(Total_Consumption_non_renewable_industrial, n = 5) %>%
  select('State','Year','Consumption.Industrial.Coal','Consumption.Industrial.Distillate Fuel Oil','Cons
```

```
## Rows: 5
## Columns: 5
## $ State                                          <chr> "Pennsylvania", "Texas", ~
```

```
## $ Year                                          <dbl> 2014, 2014, 2014, 2014, 2~
## $ Consumption.Industrial.Coal                   <dbl> 221812, 27480, 87265, 184~
## $ `Consumption.Industrial.Distillate Fuel Oil` <dbl> 57303, 228566, 19900, 308~
## $ Consumption.Industrial.Wood                   <dbl> 31260, 46290, 149561, 189~
```

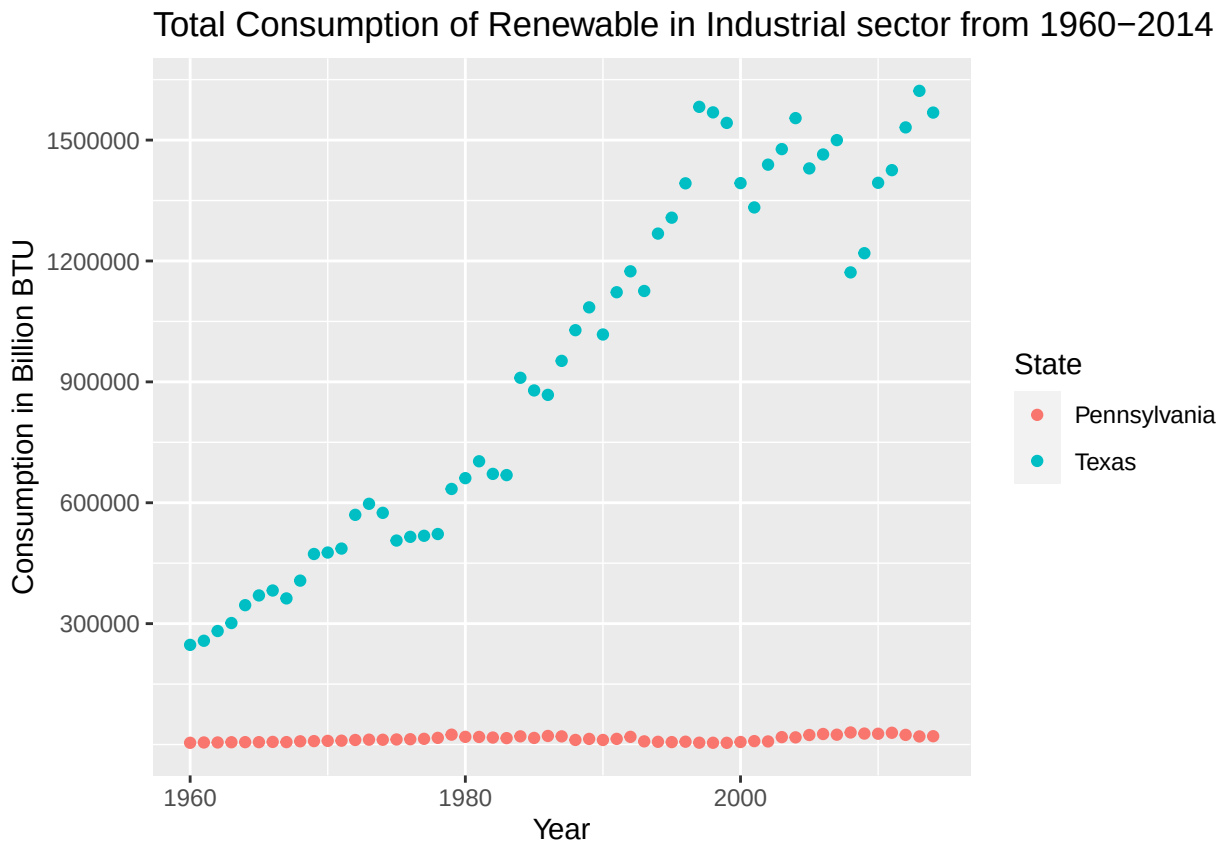
```
# Visualizing the data in the form of Stack Bar chart
pivot_data <- top5_Consumption_non_renewable_industrial %>%
    pivot_longer(cols = -c(State, Year),
                 names_to = "Types_Of_Energy",
                values_to = "Consumption")
top5_states_non_rene_chart <- ggplot(pivot_data, aes(x=State, y = Consumption/10000, fill = Types_Of_Ene
  geom_text(size = 3, position = position_stack(vjust = 0.5))+ xlab("State")  + ylab("Consumption in Bil
top5_states_non_rene_chart
```



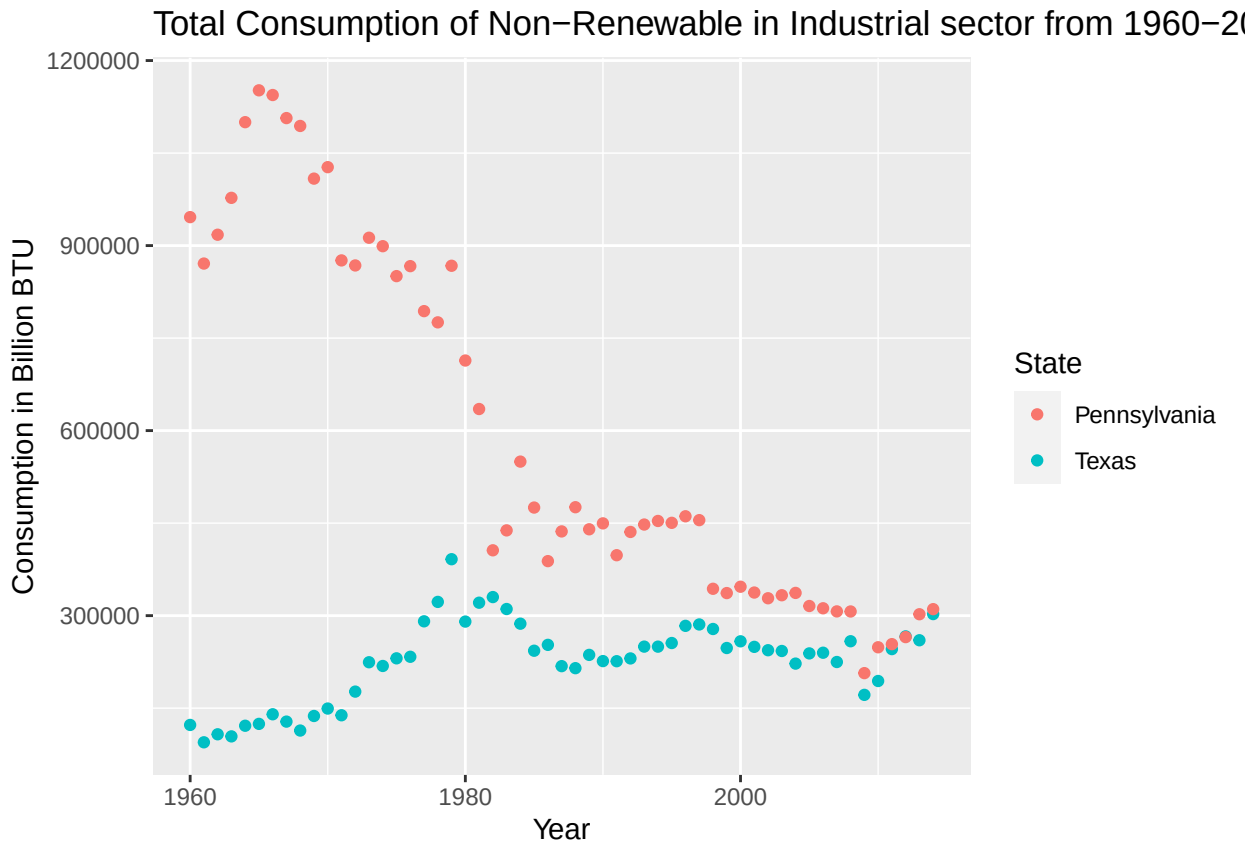

```
## Analysis of Energy Dataset
#  Who are the best and Worst States in consumption of energy resources?
states <- c("Texas", "Pennsylvania")
rene_non_rene_state <-filter(energy_mutate_df, State  %in% states) %>% select('State','Year','Total_Cons
```

```
## Rows: 110
## Columns: 4
## $ State                                        <chr> "Texas", "Pennsylvania", "T~
## $ Year                                         <dbl> 1960, 1960, 1961, 1961, 196~
## $ Total_Consumption_renewable_industrial       <dbl> 247302, 4303, 257481, 5061,~
## $ Total_Consumption_non_renewable_industrial <dbl> 122759, 946164, 94620, 8708~
```

```
# Gathering the data for both renewable and non-renewable resources usage states
# Comparing the two states based upon their total consumption of all renewable resources
c_rene_indust_graph <- ggplot(rene_non_rene_state) +
  geom_point(aes(x=Year,y=Total_Consumption_renewable_industrial, fill=State,color=State),stat="identity
   scale_y_continuous(breaks=c(300000,600000,900000,1200000,1500000,1800000)) +
  ggtitle("Total Consumption of Renewable in Industrial sector from 1960-2014") + xlab("Year") +  ylab(
c_rene_indust_graph
```




```
# Comparing the two states based upon their total consumption of all non_renewable resources
c_non_rene_indust_graph <- ggplot(rene_non_rene_state) +
  geom_point(aes(x=Year,y=Total_Consumption_non_renewable_industrial, fill=State,color=State),stat="ider
  scale_y_continuous(breaks=c(300000,600000,900000,1200000,1500000,1800000)) +
  ggtitle("Total Consumption of Non-Renewable in Industrial sector from 1960-2014") + xlab("Year") +  yl
c_non_rene_indust_graph
```



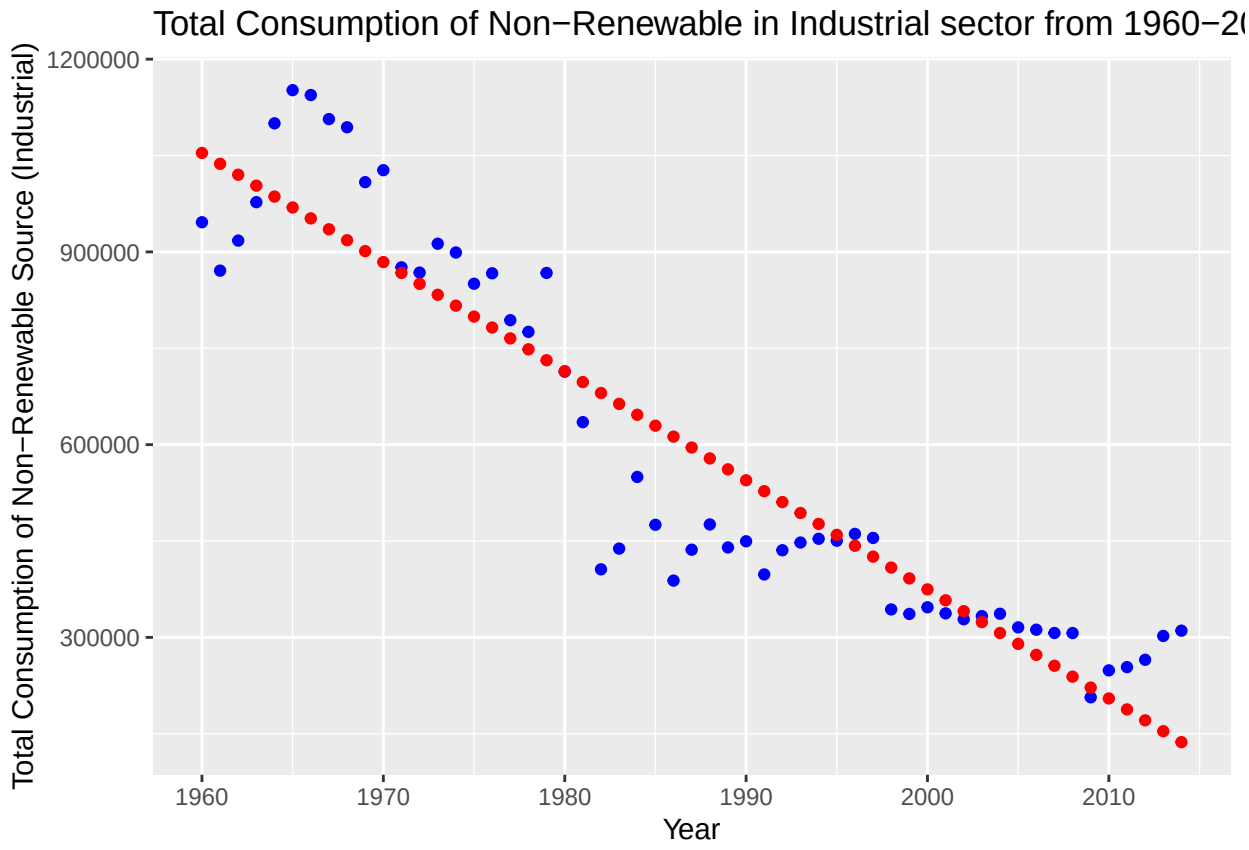

```
### Linear model for consumption of Non-Renewable source of energy for Pennsylvania over the years.

rene_non_rene_state_penn <- rene_non_rene_state %>% filter(State == 'Pennsylvania')
model <- lm(Total_Consumption_non_renewable_industrial ~ Year,data = rene_non_rene_state_penn)
summary(model)
```

```
##
## Call:
## lm(formula = Total_Consumption_non_renewable_industrial ~ Year,
##     data = rene_non_rene_state_penn)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -274487  -70052    9284   73653  191986
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34339265.6  1876308.4   18.30   <2e-16 ***
## Year          -16982.3      944.3  -17.98   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 111200 on 53 degrees of freedom
## Multiple R-squared:  0.8592, Adjusted R-squared:  0.8566
## F-statistic: 323.4 on 1 and 53 DF,  p-value: < 2.2e-16
```

```
x_years <- seq(1960, 2025)
new_df <- tibble(years = x_years)
rene_non_rene_state_pred <- rene_non_rene_state_penn %>% mutate(pred = predict(model))
rene_non_rene_plot <- rene_non_rene_state_pred %>% ggplot() + geom_point(aes(x = Year, y = Total_Consump
xlab("Year") +  ylab("Total Consumption of Non-Renewable Source (Industrial)") + ggtitle("Total Consumpt
```

```
## Scale for 'x' is already present. Adding another scale for 'x', which will
## replace the existing scale.
```

```
rene_non_rene_plot
```




```
# Bias
# Missing data for other renewable resources
# Not using all columns in the dataset
```