# Evaluating Human Development Against Renewable Energy Usage

Lars Arienti, Selin Bayraktaroglu, Valentina Herrero,
Heather Tzou, Nisha Nanda

# Introduction

**Research Plan:**

This project evaluated the Human Development Index (HDI) against renewable energy share of total energy consumption (RE%) with the goal of understanding whether countries with higher HDI use more or less renewable energy than countries with a lower HDI.

We hypothesized that countries that were highly developed would be more advanced in their technological efforts and would have greater financial resources. Renewable energy is still a developing technology and has high costs so we theorized that countries with high HDI would be able to afford the renewable energy.

Given these same advanced technological and fiscal resources, we hypothesized countries with higher HDI will have faster development in the last decade, while underdeveloped countries will see a decline in renewable energy share as their power consumption requirements increase.

## Our Informed Hypothesis

| Countries with higher HDI | Countries with lower HDI |
|---|---|
| >0.75 | <0.55 |
| ↑ % RE | ↓ % RE |

# Research Questions

**Do countries with a higher HDI consume more renewable energy (relative to other countries)?**

**Alternate Hypothesis:** There is a positive correlation between HDI and Renewable Energy Consumption

**Null Hypothesis:** There is no correlation between HDI and Renewable Energy Consumption

**Do different HDI levels affect renewable energy consumption?**

**Alternate Hypothesis:** There is a difference in renewable energy consumption by HDI levels

**Null Hypothesis:** There is no difference in renewable energy consumption by HDI levels

**Do different HDI levels affect which countries are most effective at increasing renewable energy share; developed or developing countries?**

**Alternate Hypothesis:** Developed countries will have the greatest increases in renewable energy share in the past decade. Underdeveloped countries will see a decline in renewable energy share as their power consumption requirements increase

**Null Hypothesis:** There is a correlation between the development of a country and renewable energy share in the last decade

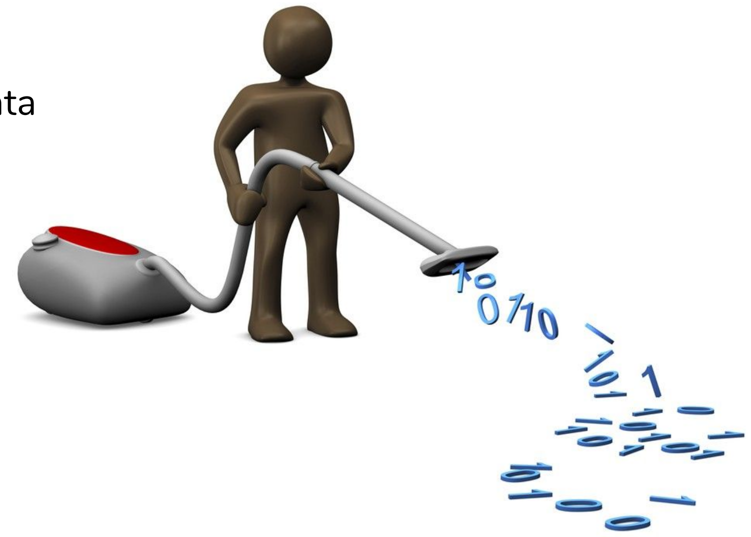
**Sources:**

- Human Development Index
- Environmental Sustainability, Dimension: Renewable Energy
  - Our World in Data
  - United Nations

*Datasets were sourced from the Human Development Index to serve as an indicator of a country's development. The HDI score was aligned with each country's renewable energy use percentage (of total energy usage).*

# Data Cleanup & Exploration

1. The exploration and cleanup process
2. Insights while exploring data
3. Problems that arose after exploring the data



# 1. The Exploration & Clean-Up Process

1. Reading the first dataset “Human Development Index”

```python
#Reading and cleaning "Human Development Index" Data
humanDevIndex=pd.read_csv("Resources/HumanDevelopmentIndex.csv")
humanDevIndex.head()
```

| | HDI Rank (2018) | Country | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | .. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 170 | Afghanistan | 0.298 | 0.304 | 0.312 | 0.308 | 0.303 | 0.327 | 0.331 | 0.335 | .. |
| 1 | 69 | Albania | 0.644 | 0.625 | 0.608 | 0.611 | 0.617 | 0.629 | 0.639 | 0.639 | .. |
| 2 | 82 | Algeria | 0.578 | 0.582 | 0.589 | 0.593 | 0.597 | 0.602 | 0.610 | 0.619 | .. |
| 3 | 36 | Andorra | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | .. |
| 4 | 149 | Angola | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | .. |

```python
# Replace nan's with zero:
humanDevIndex=humanDevIndex.fillna(0)

# Rename column "Country" to match with other dataset, which will help with 'merge'
humanDevIndex=humanDevIndex.rename(columns={"Country":"Country Name"})

# Changing the number format to 'float'
humanDevIndex=humanDevIndex.iloc[:,0:28]
humanDevIndex.iloc[:,2:28]=humanDevIndex.iloc[:,2:28].astype(float)
for i in range(2,28):
    humanDevIndex.iloc[:,i]=humanDevIndex.iloc[:,i].map("{:.2f}".format)

humanDevIndex.head()
```

| | HDI Rank (2018) | Country Name | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | ... | 2006 | 2007 | 2008 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 170 | Afghanistan | 0.30 | 0.30 | 0.31 | 0.31 | 0.30 | 0.33 | 0.33 | 0.34 | ... | 0.42 | 0.43 | 0.44 |
| 1 | 69 | Albania | 0.64 | 0.62 | 0.61 | 0.61 | 0.62 | 0.63 | 0.64 | 0.64 | ... | 0.71 | 0.72 | 0.72 |
| 2 | 82 | Algeria | 0.58 | 0.58 | 0.59 | 0.59 | 0.60 | 0.60 | 0.61 | 0.62 | ... | 0.70 | 0.71 | 0.71 |
| 3 | 36 | Andorra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 0.83 | 0.83 | 0.83 |
| 4 | 149 | Angola | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 0.47 | 0.48 | 0.49 |

# 1. The Exploration & Clean-Up Process

## 2. Reading the second dataset "Renewable Energy Consumption"

```python
# Reading "Energy Consumption" Data
energyCons=pd.read_csv("Resources/TheWorldBank/EnergyConsumption.csv")
energyCons.head()
```

| | Series Name | Series Code | Country Name | Country Code | 1990 [YR1990] | 1991 [YR1991] | 1992 [YR1992] | 1993 [YR1993] |
|---|---|---|---|---|---|---|---|---|
| 0 | Renewable energy consumption (% of total final... | EG.FEC.RNEW.ZS | Afghanistan | AFG | 15.924532 | 17.036444 | 26.521629 | 30.585667 |
| 1 | Renewable energy consumption (% of total final... | EG.FEC.RNEW.ZS | Albania | ALB | 25.518088 | 32.998174 | 46.813329 | 51.152042 |
| 2 | Renewable energy consumption (% of total final... | EG.FEC.RNEW.ZS | Algeria | DZA | 0.177228 | 0.323692 | 0.289771 | 0.476342 |
| 3 | Renewable energy consumption (% of total final... | EG.FEC.RNEW.ZS | American Samoa | ASM | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 4 | Renewable energy consumption (% of total final... | EG.FEC.RNEW.ZS | Andorra | AND | 14.273550 | 14.273550 | 14.308754 | 13.918955 |

```python
# Replace nan's with zero:
energyCons=energyCons.fillna(0)

# Rename year columns for a cleaner look
energyCons=energyCons.iloc[:,2:30]
rename_map = {}
for columns in energyCons.columns:
    if '[' in columns:
        rename_map[columns] = columns.split()[0]
energyCons=energyCons.rename(columns=rename_map)

# Change the number format to float
#energyCons=energyCons.dropna(how='any')
energyCons.dtypes
energyCons.iloc[:,2:28]=energyCons.iloc[:,2:28].astype(float)
for i in range(2,28):
    energyCons.iloc[:,i]=energyCons.iloc[:,i].map("{:.2f}".format)

energyCons.head()
```

| | Country Name | Country Code | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | ... | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Afghanistan | AFG | 15.92 | 17.04 | 26.52 | 30.59 | 32.80 | 35.08 | 37.95 | 41.43 | ... | 37.14 | 33.86 | 21.34 | 17.81 |
| 1 | Albania | ALB | 25.52 | 33.00 | 46.81 | 51.15 | 51.46 | 50.61 | 51.64 | 55.95 | ... | 31.71 | 32.10 | 35.91 | 37.22 |
| 2 | Algeria | DZA | 0.18 | 0.32 | 0.29 | 0.48 | 0.40 | 0.42 | 0.39 | 0.47 | ... | 0.41 | 0.41 | 0.30 | 0.31 |
| 3 | American Samoa | ASM | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 0.00 | 0.00 | 0.00 | 0.00 |
| 4 | Andorra | AND | 14.27 | 14.27 | 14.31 | 13.92 | 14.56 | 14.48 | 14.10 | 14.19 | ... | 17.49 | 16.94 | 17.42 | 17.52 |

# 1. The Exploration & Clean-Up Process

3. Merging two datasets

```
# Merging two datasets
# Columns with HDI: Human Development Index
# Columns with RE: Renewable Energy Consumption Rate
merged_df=pd.merge(humanDevIndex,energyCons,on="Country Name",how='outer',suffixes=("_HDI","_RE"))
merged_df=merged_df.dropna(how='any')
merged_df.head()
```

| | HDI Rank (2018) | Country Name | 1990_HDI | 1991_HDI | 1992_HDI | 1993_HDI | 1994_HDI | 1995_HDI | 1996_HDI | 1997_HDI | ... | 2006_RE | 2007_RE | 2008_RE | 2009_RE | 2010_ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 170.0 | Afghanistan | 0.30 | 0.30 | 0.31 | 0.31 | 0.30 | 0.33 | 0.33 | 0.34 | ... | 37.14 | 33.86 | 21.34 | 17.81 | 14 |
| 1 | 69.0 | Albania | 0.64 | 0.62 | 0.61 | 0.61 | 0.62 | 0.63 | 0.64 | 0.64 | ... | 31.71 | 32.10 | 35.91 | 37.22 | 37 |
| 2 | 82.0 | Algeria | 0.58 | 0.58 | 0.59 | 0.59 | 0.60 | 0.60 | 0.61 | 0.62 | ... | 0.41 | 0.41 | 0.30 | 0.31 | 0 |
| 3 | 36.0 | Andorra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 17.49 | 16.94 | 17.42 | 17.52 | 19 |
| 4 | 149.0 | Angola | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 65.02 | 61.60 | 58.11 | 55.75 | 54 |

# 1. The Exploration & Clean-Up Process

4. Cleaning merged dataset

```
# Drop NAs and remove column "Country Code"
merged_df=merged_df.dropna(how='any')
del merged_df["Country Code"]
```

```
# Remove the row with an error
merged_df=merged_df.drop([33],axis=0)
```

```
# Make sure all of the numbers are in float format
merged_df.iloc[:,2:52]=merged_df.iloc[:,2:52].astype(float)
```

```
# Show the columns
merged_df.columns
```

```
Index(['HDI Rank (2018)', 'Country Name', '1990_HDI', '1991_HDI', '1992_HDI',
       '1993_HDI', '1994_HDI', '1995_HDI', '1996_HDI', '1997_HDI', '1998_HDI',
       '1999_HDI', '2000_HDI', '2001_HDI', '2002_HDI', '2003_HDI', '2004_HDI',
       '2005_HDI', '2006_HDI', '2007_HDI', '2008_HDI', '2009_HDI', '2010_HDI',
       '2011_HDI', '2012_HDI', '2013_HDI', '2014_HDI', '2015_HDI', '1990_RE',
       '1991_RE', '1992_RE', '1993_RE', '1994_RE', '1995_RE', '1996_RE',
       '1997_RE', '1998_RE', '1999_RE', '2000_RE', '2001_RE', '2002_RE',
       '2003_RE', '2004_RE', '2005_RE', '2006_RE', '2007_RE', '2008_RE',
       '2009_RE', '2010_RE', '2011_RE', '2012_RE', '2013_RE', '2014_RE',
       '2015_RE'],
      dtype='object')
```

# 1. The Exploration & Clean-Up Process

## 5. Adding extra columns: Avg HDI, Avg RE Consumption, Total RE Consumption

```
# Create two new columns, which calculates avg HDI index and energy consumption for each countries
merged_df["Avg HDI"]=merged_df.iloc[:,2:27].mean(axis=1)
merged_df["Avg Energy Consumption"]=merged_df.iloc[:,28:52].mean(axis=1)
merged_df["Total Energy Consumption"]=merged_df.iloc[:,28:52].sum(axis=1)
```

```
merged_df.head()
```

| | HDI Rank (2018) | Country Name | 1990_HDI | 1991_HDI | 1992_HDI | 1993_HDI | 1994_HDI | 1995_HDI | 1996_HDI | 1997_HDI | ... | 2015_RE | Avg HDI | Avg Energy Consumption | Total Energy Consumption |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 170.0 | Afghanistan | 0.30 | 0.30 | 0.31 | 0.31 | 0.30 | 0.33 | 0.33 | 0.34 | ... | 18.42 | 0.3836 | 32.726250 | 785.43 |
| 1 | 69.0 | Albania | 0.64 | 0.62 | 0.61 | 0.61 | 0.62 | 0.63 | 0.64 | 0.64 | ... | 38.62 | 0.6856 | 40.551667 | 973.24 |
| 2 | 82.0 | Algeria | 0.58 | 0.58 | 0.59 | 0.59 | 0.60 | 0.60 | 0.61 | 0.62 | ... | 0.06 | 0.6652 | 0.377500 | 9.06 |
| 3 | 36.0 | Andorra | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 19.75 | 0.4924 | 16.119167 | 386.86 |
| 4 | 149.0 | Angola | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | ... | 49.57 | 0.3020 | 66.917083 | 1606.01 |

# 2. Insights While Exploring Data

**Segmenting data to understand RE consumption trend better**

| Development | Avg Energy Consumption | Median Energy Consumption | Var Energy Consumption | Std Energy Consumption |
|---|---|---|---|---|
| Low Human Development | 1226 | 1283.89 | 626874.991004 | 791.754375 |
| Medium Human Development | 757 | 750.95 | 264173.052221 | 513.977677 |
| High Human Development | 331 | 243.71 | 99370.795217 | 315.231336 |
| Very High Human Development | 331 | 177.11 | 157914.497661 | 397.384571 |

| | Avg Energy Consumption | Median Energy Consumption | Var Energy Consumption | Std Energy Consumption |
|---|---|---|---|---|
| **Total:** | 795.84 | 570.67 | 513872.300163 | 716.848868 |

```
# Grouping HDIs into 4 groups: Low, Medium, High and Very High Human Development:
    # HDI of less than 0.550 for low human development,
    # 0.550-0.699 for medium human development,
    # 0.700-0.799 for high human development
    # and 0.800 or greater for very high human development.

# Creating a column "Development" to segment countries by their development level:
# Creating bins and labels for each segments
bins=[0,0.549,0.699,0.799,1]
labels=["Low Human Development","Medium Human Development", "High Human Development","Very High Human Development"]
merged_df["Development"]=pd.cut(merged_df["Avg HDI"],bins,labels=labels)
```

# 3. Problems after exploring the data

**We were almost losing 27 countries (including USA) because of using dropna before merging datasets**

**BEFORE:**

- Used dropna before merging

```
# Number of countries in the merged dataset
merged_df["Country Name"].count()
```

```
127
```

We were losing 27 countries as a result of excessive dropping

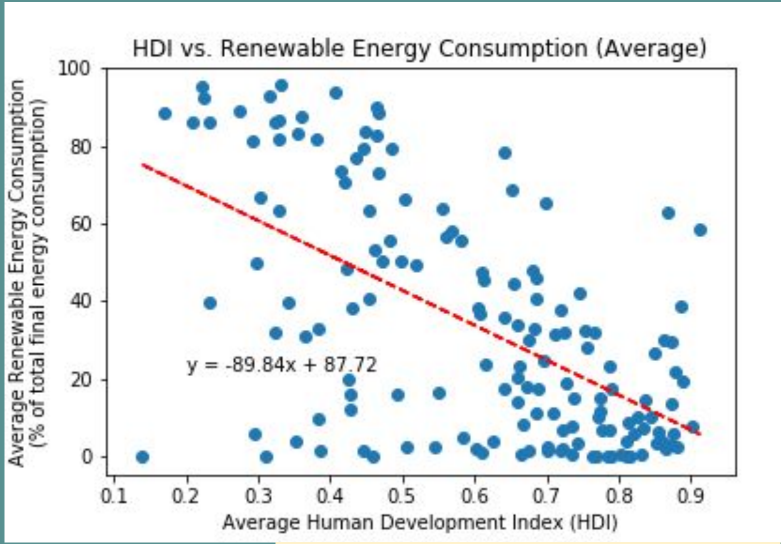
**AFTER:**

- Filled na's with zeros before merging
- Used dropna after merging

```
# Number of countries in the merged dataset
merged_df["Country Name"].count()
```

```
164
```

# Data Analysis: Looking at the Relationship between HDI and Renewable Energy Consumption




Correlation coefficient: ≈ -0.589
p-value ≈ 1.428e-15

```python
# Creating a scatter plot to see the relationship betweeen HDI and Renewable Energy Consumption:

# Creating labels and titles
plt.title("HDI vs. Renewable Energy Consumption (Average)")
plt.xlabel("Average Human Development Index (HDI)")
plt.ylabel(r"Average Renewable Energy Consumption" + "\n" + "(% of total final energy consumptio
n)")

# Creating x and y axis
x_values = merged_df["Avg HDI"]
y_values = merged_df["Avg Energy Consumption"]

# Regression analysis
(slope, intercept, rvalue, pvalue, stderr) = linregress(x_values, y_values)
regress_values = x_values * slope + intercept
line_eq = "y = " + str(round(slope, 2)) + "x + " + str(round(intercept, 2))
plt.scatter(x_values, y_values)
plt.plot(x_values, regress_values, "r--")
plt.annotate(line_eq, (0.20, 22), fontsize = 10, color = "black")

plt.savefig("Final_Images/firstgraph.png")

print(f"Correlation coefficient: {rvalue} \n")
print(f"P-value: {pvalue} \n")
```

# Data Analysis: Looking at the Relationship between HDI and Renewable Energy Consumption by HDI levels

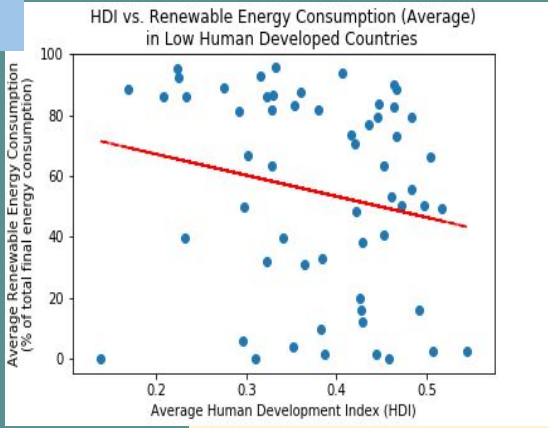
1) **Low Human Development**:
HDI < 0.550

2) **Medium Human Development**:
0.550 ≤ HDI ≤ 0.699

3) **High Human Development**:
0.700 ≤ HDI ≤ 0.799

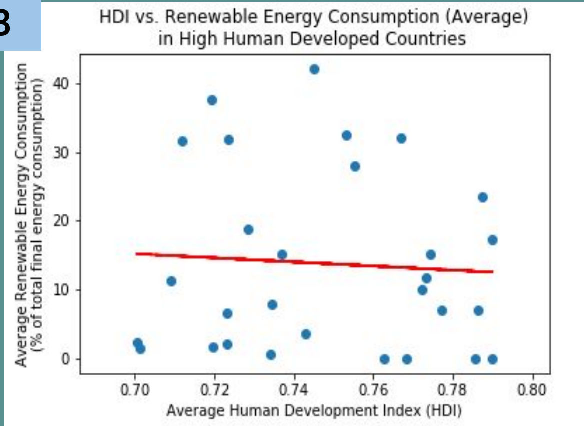
4) **Very High Human Development**:
HDI ≤ 0.800

1



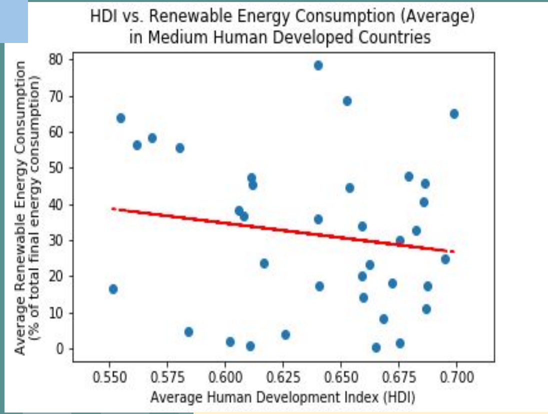

Correlation coefficient: ≈ -0.200

3



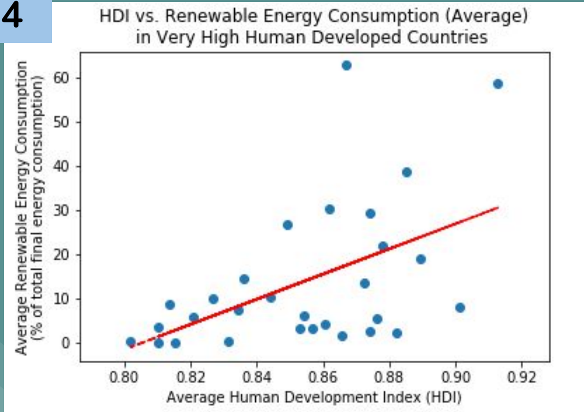

Correlation coefficient: ≈ -0.065

2



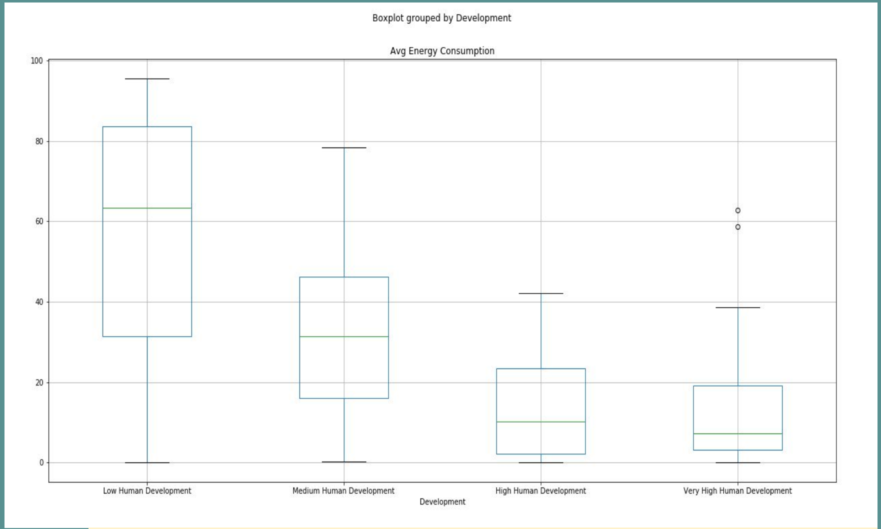

Correlation coefficient: ≈ -0.160

4




Correlation coefficient: ≈ 0.505

# ANOVA Test on the Different HDI Groups (Low, Medium, High, & Very High Human Development)




F_onewayresult(statistic ≈ 26.654, **p-value ≈ 7.666e-14**

```
# Creating a boxplot for each segment:
merged_df.boxplot("Avg Energy Consumption", by="Development", figsize = (20,10))
group1 = merged_df.loc[merged_df["Development"]=="Low Human Development"]["Avg Energy Consumptio
n"]
group2 = merged_df.loc[merged_df["Development"]=="Medium Human Development"]["Avg Energy Consumpt
ion"]
group3 = merged_df.loc[merged_df["Development"]=="High Human Development"]["Avg Energy Consumptio
n"]
group4 = merged_df.loc[merged_df["Development"]=="Very High Human Development"]["Avg Energy Consu
mption"]

# Perform ANOVA to  test the validity of the null hypothesis which states a commonly accepted cla
im about a population.
stats.f_oneway(group1,group2,group3,group4)
```

Well that's a bummer...

It seems like high HDI countries use LESS renewable energy than lower HDI countries. Let's see if this is a growing trend or a fading trend.

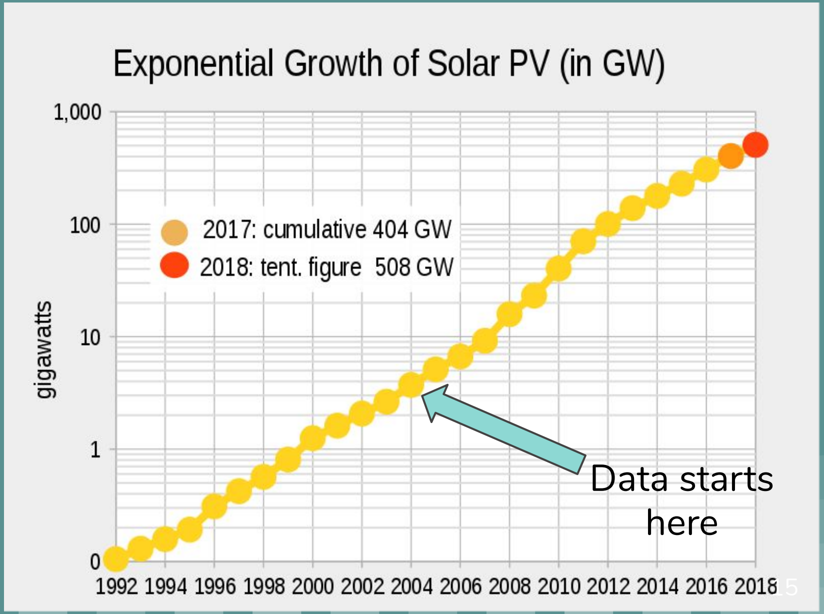
How will we answer this question?

Compare the HDI of a country to RE% over time.

To ensure that we look at time periods where renewable energy sources, such as solar panels, are readily available, we will drop some of our older data to examine this trend.

Specifically, compare the avg RE% for 2004-2009 to avg RE% for 2010-2015 and the avg HDI from 2004-2015

https://en.wikipedia.org/wiki/Growth_of_photovoltaics:



## Before we show you the results, ask yourself:

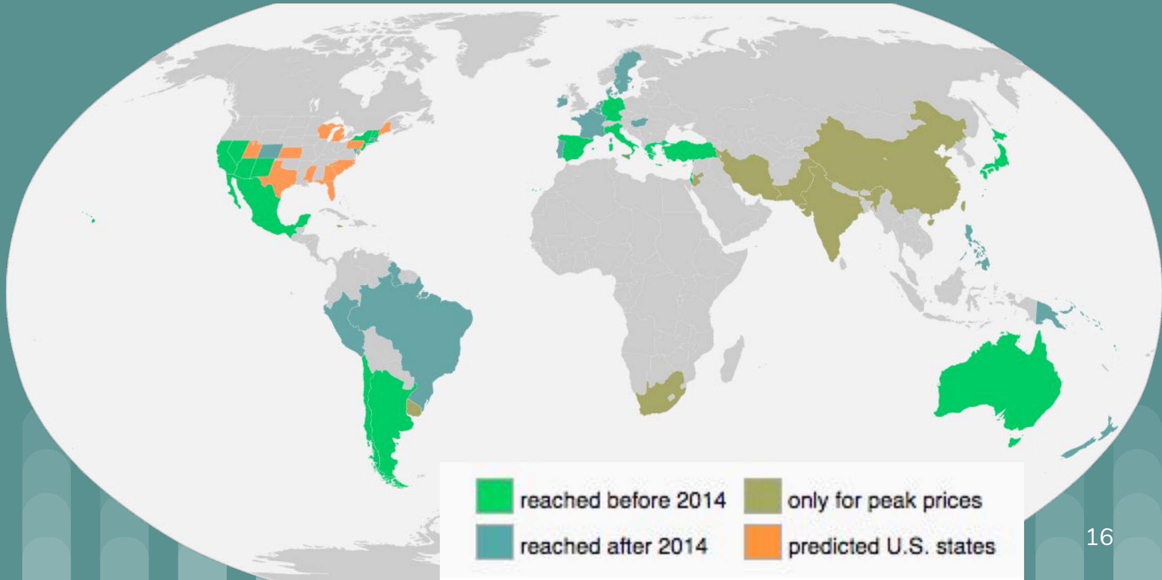
Will developed countries tend to use more or less renewable energy over time?

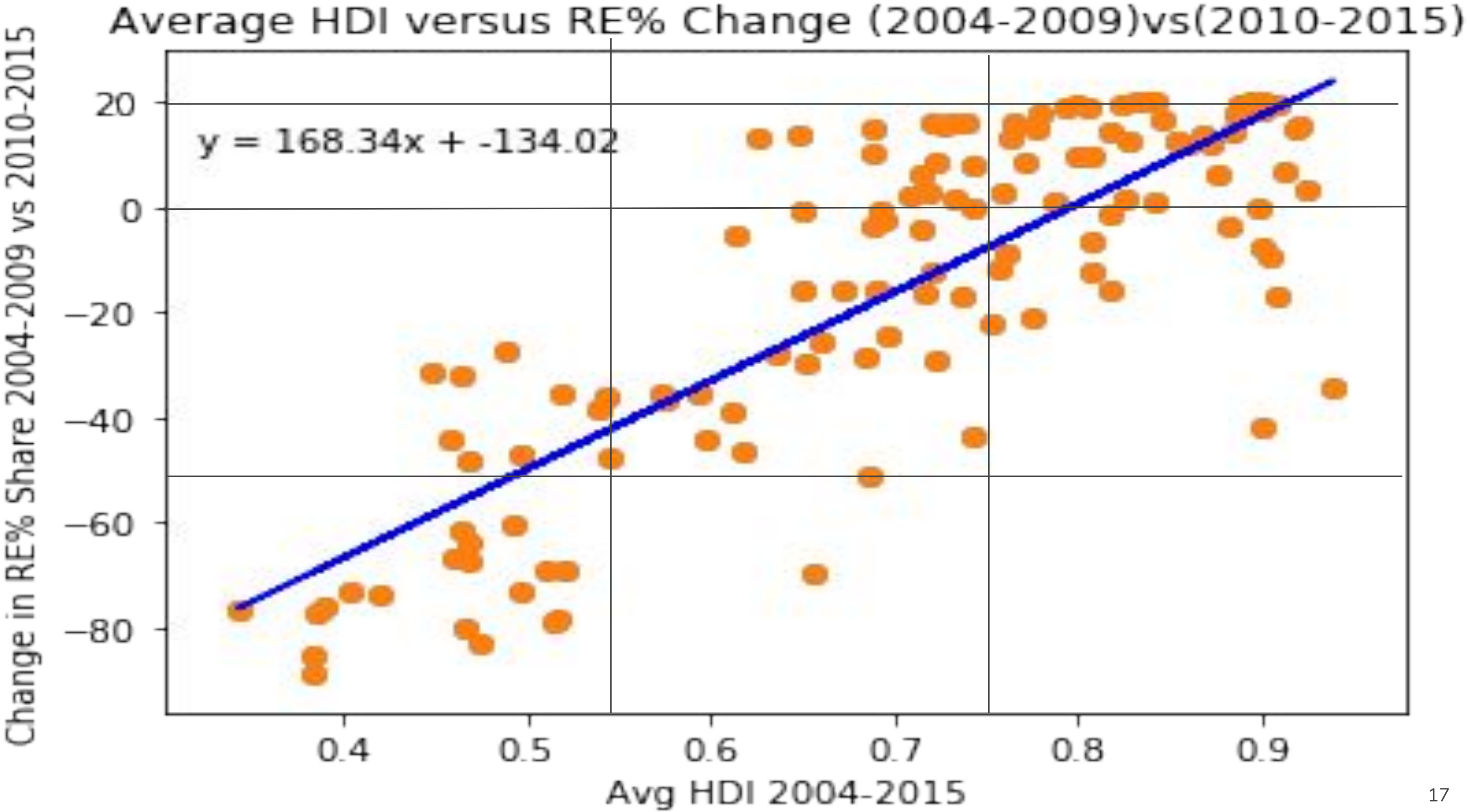
...What about less developed and developing countries?

Who can afford to purchase for renewable energy production?

**Grid parity** (or socket **parity**) occurs when an alternative energy source can generate power at a levelized cost of electricity (LCOE) that is less than or equal to the price of power from the electricity **grid**.

Ultimately, this is an economic problem, as higher electricity costs incentivize the development of private alternative energy sources.





Average HDI versus RE% Change (2004-2009)vs(2010-2015)
y = 168.34x + -134.02
Change in RE% Share 2004-2009 vs 2010-2015
Avg HDI 2004-2015
20
0
−20
−40
−60
−80
0.4
0.5
0.6
0.7
0.8
0.9

# Discussion

## Insights from Data

- High HDI and Very High HDI associated with lower renewable energy consumption
- Increase in HDI is correlated with higher growth over time of RE% share

## Inferences Drawn

- Underdeveloped countries rely on alternative energy sources as their power consumption needs increase
- More developed countries see an increase in renewable energy share as they’ve acquired the technology and resources to grow

## Future Trends

- Renewable energy shares are going to be prominent within highly developing countries

# Post Mortem

## Complications Along the Way

- Normalizing and merging datasets
    - Organizing data into groups to reduce workload and draw clearer conclusions
    - Narrowing the scope of our study

## Future Questions

➔ What type of renewable energy sources are used most commonly compared against HDI levels?

➔ How do HDI levels compare to that of fossil fuel consumption?

➔ Do higher HDI countries have a higher diversity in energy sources?

# Resources

HDI dataset: http://hdr.undp.org/en/data

RE Consumption dataset: https://data.worldbank.org/indicator/eg.fec.rnew.zs

Grouping HDI: http://hdr.undp.org/en/content/human-development-report-2019-readers-guide

# Questions?