

INM402 Coursework Submission

Name: *Ling-Yun, Huang*

Email: *Ling-Yun.Huang@city.ac.uk*

1. Research Questions

In 2016, the Taiwanese government announced a series of energy goals aimed to achieve by 2025. This includes reaching 20% of total power generation from renewable energy sources.

1. How has Taiwan's renewable energy generation changed over the past 20 years, especially before and after 2016? What factors might explain these changes?
2. With only two years left, how far is Taiwan to its energy goals? Based on current progress, is it possible to achieve the goals by 2025?

2. The Visualization

The dataset used in this report originates from Taiwan's [open data platform](#) and included records of power generation from various sources from 2005 to 2023. Used Python to analyse the prediction data and import into this report.

Figure 1-1. Renewable Energy Generation Over Time in Taiwan

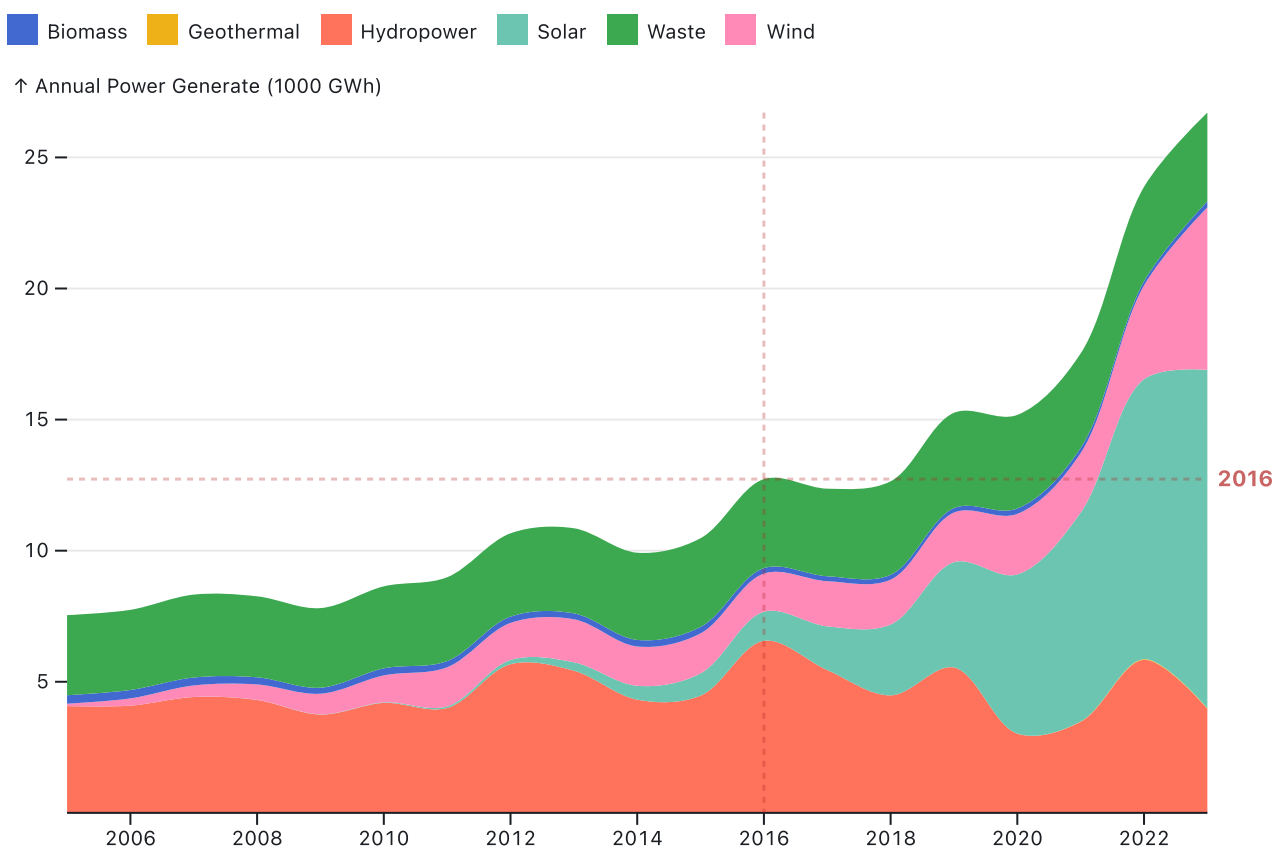


Figure 1-2. The Power Generation in Solar and Wind Sources

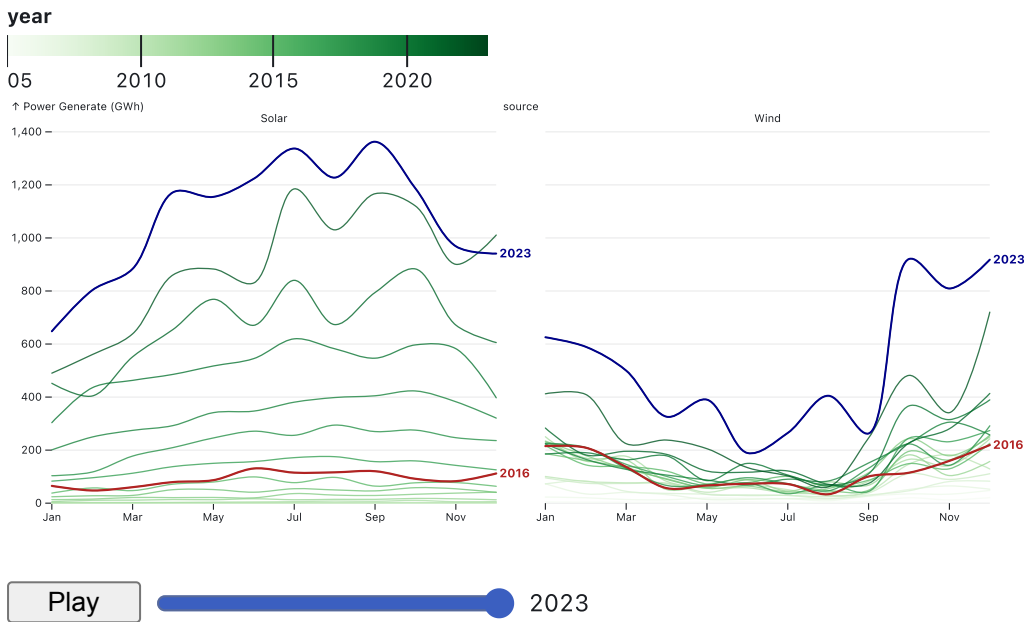
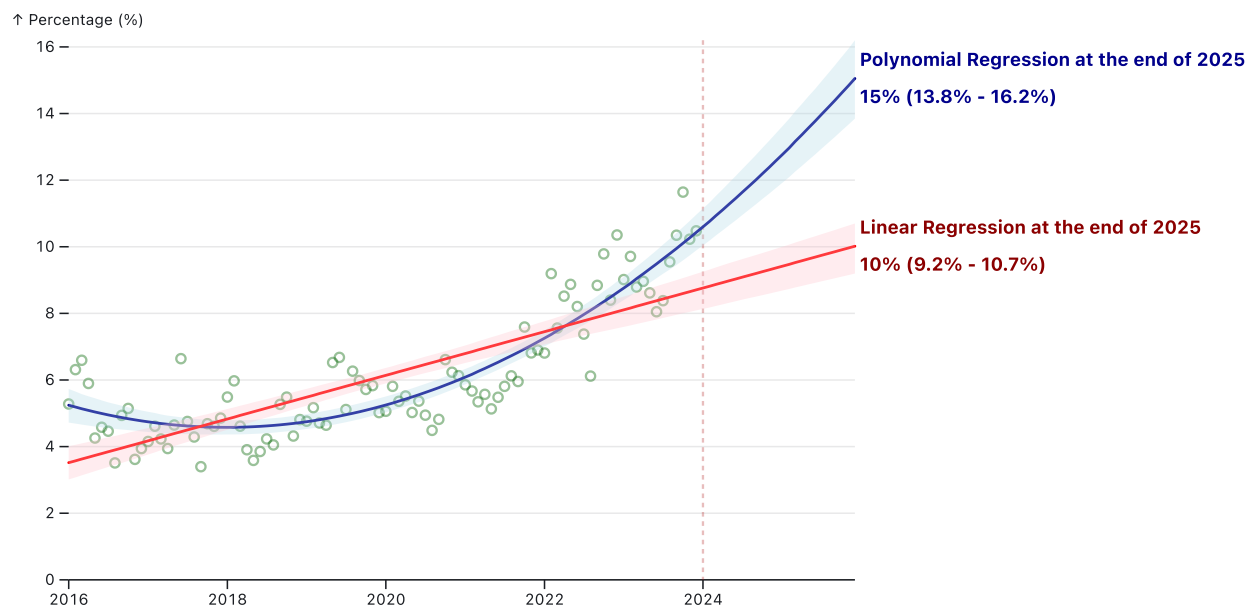


Figure 2. The Prediction of Renewable Power Generation



3. Insights

1. In Figure 1-1, the trend of renewable power generation in Taiwan shows a consistent annual increase, especially with a noticeable accelerate after 2016. A closer examination of individual energy sources reveals distinct patterns, with significant contributions from Solar and Wind sources. Solar energy, in particular, emerges as a major contributor to this upward trend.
2. To evaluate the contributions of Solar and Wind energy, Figure 1-2 offers deeper insights in both sources.
 - In Solar energy, the generation initially starts at a minimal level, showing slight improvement before 2016. During this period, no clear seasonal

pattern is observed. However, after 2016, there's a significant increase in Solar energy generation. Additionally, a seasonal pattern emerges, with higher generation in summer and lower in winter, becoming increasingly evident each year.

- Conversely, Wind energy also begins at a low level but doesn't show as much improvement after 2016 compared to Solar energy. The most significant enhancement is observed in the past two years (2022-2023). Similar to Solar energy, Wind energy shows a seasonal pattern, with higher generation in winter and lower in summer, becoming more evident over time.
- To encourage more people to join the renewable power generation increase, the Taiwanese government has announced several policies to support this initiative. These policies include encouraging individuals or companies to install solar panels on house or factory roofs and providing financial support for wind infrastructure planting. These policies may have contributed to the insights mentioned above. Additionally, since it is easier to install solar panels than wind infrastructure, solar energy has shown immediate improvement compared to wind energy.

3. Figure 2 shows two types of regression predictions: linear and polynomial, these can help me to answer the second question. By the end of 2025, the linear regression prediction is expected to reach 15% with a 95% confidence interval from 9.2% to 10.7% on renewable power generation. On the other hand, the polynomial regression prediction is expected to reach 15% with a 95% confidence interval from 13.8% to 16.2%. With closer examination of the raw data, the polynomial line demonstrates a closer fit than the linear line. In addition, the Taiwan government's financial support for related projects continues. While the 20% target may not be reached by 2025, it is possible by 2027.

4. Design Justification

The visualisations in this coursework follow Jonathan Corum's "[See, Think, Design, Produce](#)" process, which includes understanding and explaining the data through visualisation. For the first question, two things I wanted to show, an overview of the renewable power generation and the deeper insights for each sources. For the second question, I wanted to not only show the prediction but the reason why I choose this statistical approach to analysis.

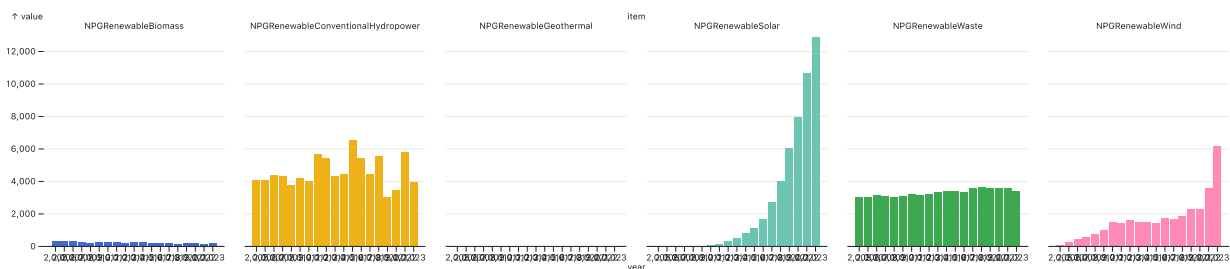
1. It's been a challenge to find a single chart that effectively addresses the first two goals. Each design option comes with trade-offs. For example, I tried multi-faced bar chart (see reject 1 below), although in this figure I could easily see how each sources generated across two decades. However, I couldn't see the overall value with this choice. The final decision is a combination of a stacked area chart with line charts. The stacked area chart allow me to show the overall generation and also provided the detail of individual sources ([Munzner, 2014](#)). But it also has its drawbacks, such as it being difficult to see individual changes because the values of each category are additive and they can be greatly affected by other categories. To

overcome this problem, I added another figure with line charts to further extract deeper insights.

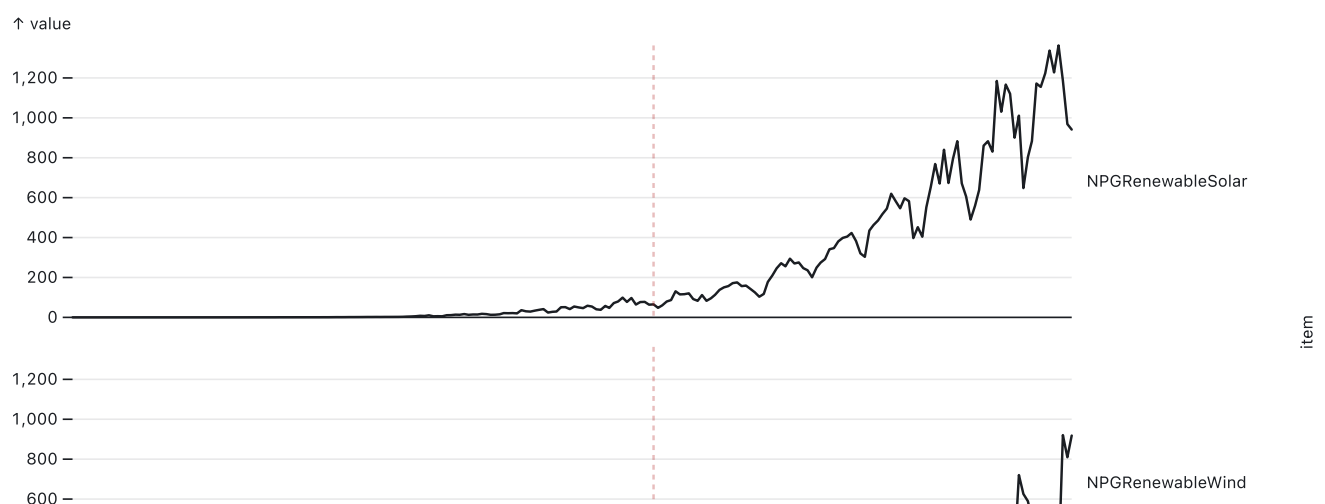
2. To deal with stacked area charts, I need the data in each source to start at absolute zero. First, I tried plotting all the data on a line chart (see reject 2 below), through this chart we can see how both sources change over time, it also shows the seasonal patterns of both sources model. However, we can already see the trend in the first graph, and it takes effort to observe that the seasonal patterns are different for the two sources. So, how do I design a plot that includes these insights but also achieve "preattentive processing" (Ware, 2021), to delivered the message more easily. Inspired by Lecture 6 material, I chose to split the data by year and put each year's data on a line chart. By overlaying the data from each year, we can see not only the seasonal patterns, but also how the value change for each year. I also highlight the year of 2016, to make the comparison more easily for viewers.
3. For the prediction data, I use Python to do the analysis part, linear and polynomial regression, to analyse and predict the data, and generate visualisations based on the results. This process follows the Wilkinson's Grammar of Graphics (Wilkinson, 2010, 2012). I know the polynomial regression would more suitable for this prediction task compare to linear regression. However, I still get both results displayed on the plot. By showing the different possibilities, I provided an invitation for the audiences to join the prediction process. Like what Corum says: "The goal is to communicate that understanding".

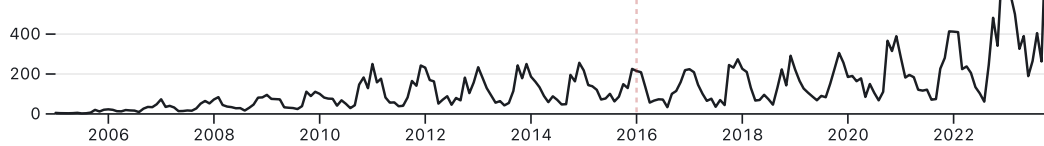
Below are the reject designs I mention above.

reject 1



reject 2





5. Validation

1. In Munzner's book "Visualisation Analysis and Design" (Munzner, 2014), she discusses the concept of the "decision-making loop", which combines both the "computer in the loop" and the "human in the loop". My designs follow this framework, incorporating previous insights into later figures to reach a higher level in Bloom's domain taxonomies (Bloom et al., 1956). For example, we observed that the growth of renewable energy is mainly driven by solar and wind energy, so subsequent charts were carefully designed to reveal deeper insights. In addition, for prediction data, we use our understanding of the different trends before and after 2016 to make predictions based on data after 2016 to ensure accuracy.
2. On the other hand, by choosing to display only solar and wind data, I acknowledge that I may be ignoring the potential future impact of other energy sources. The limitations of the stacked area chart, combined with the other visualisations I presented, may bias the viewer to only consider the positive aspects of the data and ignore the potential negative effects. For example, the instability of hydropower energy in recent years (see reject 1) remains unexplored, and questions of its causes and potential future changes remain unanswered. Given that hydropower accounts for a large portion of renewable energy generation, its instability may be worthy of discussion and analysis.
3. I used different y-axis scales in Figures 1 and 2, representing absolute values and percentages. This difference could cause a problem, particularly if total power generation varies significantly each year. The insights derived from Figure 1 and applied in Figure 2 may be affected if the observed trends do not persist with the change in scale. However, I would argue that despite the yearly increase in total power generation, the significant change in renewable power generation after 2016 is much higher. And the pattern still holds in percentage scale in this case. Nevertheless, it's important to be cautious about such differences to ensure the accuracy of the analysis (Alberto Cairo "How Charts Lie").

6. References

Bloom, B., Engelhart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. David McKay.

Munzner, T. (2014). *Visualization Analysis and Design*. CRC press.
<https://go.exlibris.link/9jMy6fQG>

Ware, C. (2021). *Information visualization: perception for design, 4th Edition*. Morgan Kaufmann. <https://library.city.ac.uk/443/record-b2679546>

Wilkinson, L. (2012). *The Grammar of Graphics*. Springer. <https://go.exlibris.link/Y2gQqNZG>

```
powerGenerateM = ▶ Array(228) [Object, Object, Object, Object, Object, Object,
predictData = ▶ Array(120) [Object, Object, Object, Object, Object, Object, Obj
powerGenerateY = ▶ Array(19) [Object, Object, Object, Object, Object, Object, (
powerTidyM = ▶ Array(10488) [Object, Object, Object, Object, Object, Object, Ol
powerTidyY = ▶ Array(874) [Object, Object, Object, Object, Object, Object, Objc
```

What are the different types of renewable energy sources available, and how do they work? How does renewable energy compare to traditional fossil fuels in terms of cost and environmental impact? What policies and incentives are in place to promote the adoption of renewable energy? What are the challenges and limitations of integrating renewable energy into existing energy grids? How can we improve energy storage technologies to address the intermittency of renewable sources like solar and wind? What are the socioeconomic impacts of transitioning to renewable energy, including job creation and energy access? What role can innovations in renewable energy play in mitigating climate change? How can communities and individuals participate in and benefit from renewable energy projects? What are the environmental considerations associated with large-scale deployment of renewable energy infrastructure? How can we ensure equitable access to renewable energy resources globally?

- **Prompt:** I have monthly data and I want to use linear regression and polynomial regression to analysis and predict following 24 months data, how do I do it in python and also how do I output csv file with analysis and prediction data and 95% CI.

- **Response:**

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
import statsmodels.api as sm

# Load data
data = pd.read_csv('your_data.csv')

# Extract features and target variable
X = data[['month']].values
y = data['value'].values

# Linear Regression
model_linear = LinearRegression()
model_linear.fit(X, y)

# Polynomial Regression
degree = 2 # Adjust the degree as needed
poly_features = PolynomialFeatures(degree=degree)
X_poly = poly_features.fit_transform(X)
model_poly = LinearRegression()
model_poly.fit(X_poly, y)

# Predict future values
future_months = np.arange(1, 25).reshape(-1, 1)
linear_predictions = model_linear.predict(future_months)
poly_predictions = model_poly.predict(poly_features.transform(future_months))

# Output to CSV
predictions_df = pd.DataFrame({
    'Month': future_months.flatten(),
    'Linear_Predictions': linear_predictions,
    'Poly_Predictions': poly_predictions
})
predictions_df.to_csv('predictions.csv', index=False)

# Confidence Intervals
X_with_intercept = sm.add_constant(X_poly)
model = sm.OLS(y, X_with_intercept)
result = model.fit()
prediction_interval = result.get_prediction(poly_features.transform(future_mon
predictions_df['Lower_CI'] = prediction_interval.summary_frame(alpha=0.05)['ob
predictions_df['Upper_CI'] = prediction_interval.summary_frame(alpha=0.05)['ob
```

- I have also used Grammarly to check the accuracy of my english grammar.

Do not edit below this line

```
import {task} from "@jwolondon/schema-teaching"

import {icons} from "6ccb7a5a129abf60"

import {readData, T, table} from "@jwolondon/teaching-utilities"

import {Scrubber} from "@mbostock/scrubber"

import {References} from "@jwolondon/references-class"

bib = ►References {refObject: Array(5), citeFac: f(...)}

refs = `

@book{munzner_visualization_2014,
  title={Visualization Analysis and Design},
  author={Munzner, Tamara},
  year={2014},
  publisher={CRC press},
  url={https://go.exlibris.link/9jMy6fQG}
}

@book{bloom_taxonomy_1956,
  title={Taxonomy of educational objectives, handbook I: The cognitive domain}
  author={Bloom, Benjamin and Engelhart, Max and Furst, Edward and Hill, Walke}
  year={1956},
  publisher={David McKay}
}

@book{ware_information_2021,
  title={Information visualization: perception for design, 4th Edition},
  author={Ware, Colin}, Show 24 truncated lines

<style>

<link>
```