Analysis of the Relationship Between Wind Speed and Humidity

December 2024

- **List**

About Dataset

Content

The Dataset is fully dedicated for the developers who want to train the model on Weather Forecasting for Indian climate. This dataset provides data from **1st January 2013** to **24th April 2017** in the city of Delhi, India. The 4 parameters here are

meantemp, humidity, windspeed, meanpressure.

Work process

Alright, so far we have prepared the data. We want to find the relationship between wind speed and humidity using software or the R programming language. First, we will display a portion of the data and obtain some statistics such as mean, variance, etc., for the two target variables. We will also calculate a correlation coefficient between the two target variables. After these steps, we will generate a series of plots such as scatter plots, bar graphs, and finally, we will create a network graph. In the end, we will provide a general conclusion**.**

introduction

I have an idea to see if there is a change in humidity as the wind speed increases. That's why I have gathered this dataset and I want to work on this hypothesis.

We will answer this hypothesis with several steps. First, we will calculate the correlation coefficient. If it is close to one, it means that as one variable increases, the other also increases, and vice versa. If it is close to negative one, it means that as one variable increases, the other decreases. Then, we will display bar charts for the two variables. Next, we will check the normality of the distributions. If the distributions are normal, it is much better. Finally, we will examine the line charts and the network graph.

Data

Our data consists of 5 variables, and we are analyzing the two variables: wind speed and humidity. Each variable has 115 observations. Below, we will display a portion of the data.

date	meantemp	humidity	wind_speed	meanpressure
2017-01-01	15.91304	85.86957	2.743478	59.000
2017-01-02	18.50000	77.22222	2.894444	1018.278
2017-01-03	17.11111	81.88889	4.016667	1018.333
2017-01-04	18.70000	70.05000	4.545000	1015.700
2017-01-05	18.38889	74.94444	3.300000	1014.333
2017-01-06	19.31818	79.31818	8.681818	1011.773

Summary

humidity	Min	1st Qu	Median	Mean	3rd Qu	Max.
	17.75	39.62	57.75	56.26	71.90	95.83

wind_speed	Min	1st Qu	Median	Mean	3rd Qu	Max.
	1.387	5.564	8.069	8.144	10.069	19.314

Based on the information above, the humidity variable has a minimum of 17.75 and a maximum of 95.83 And the wind speed variable has a minimum of 1.397 and a maximum of 19.31And the mean and standard deviation for the two variables are shown .

Normality

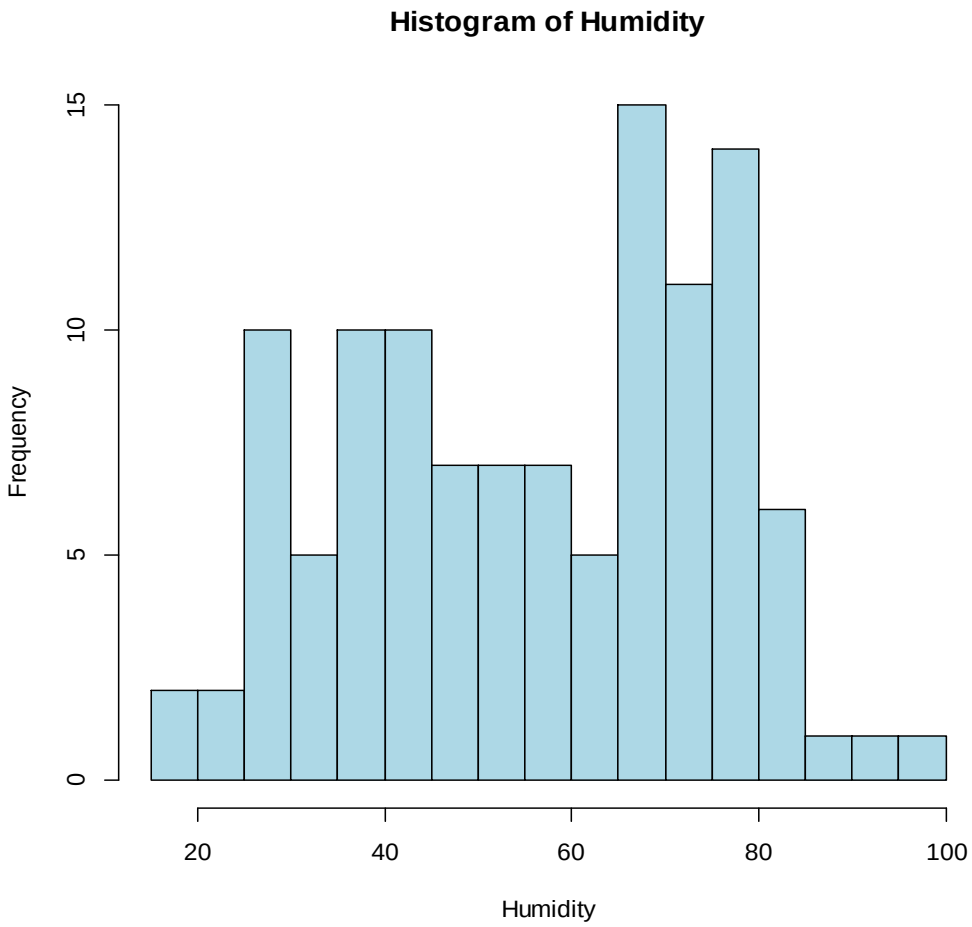
Here, we will check the normality of the data using the Shapiro-Wilk test and by examining histograms

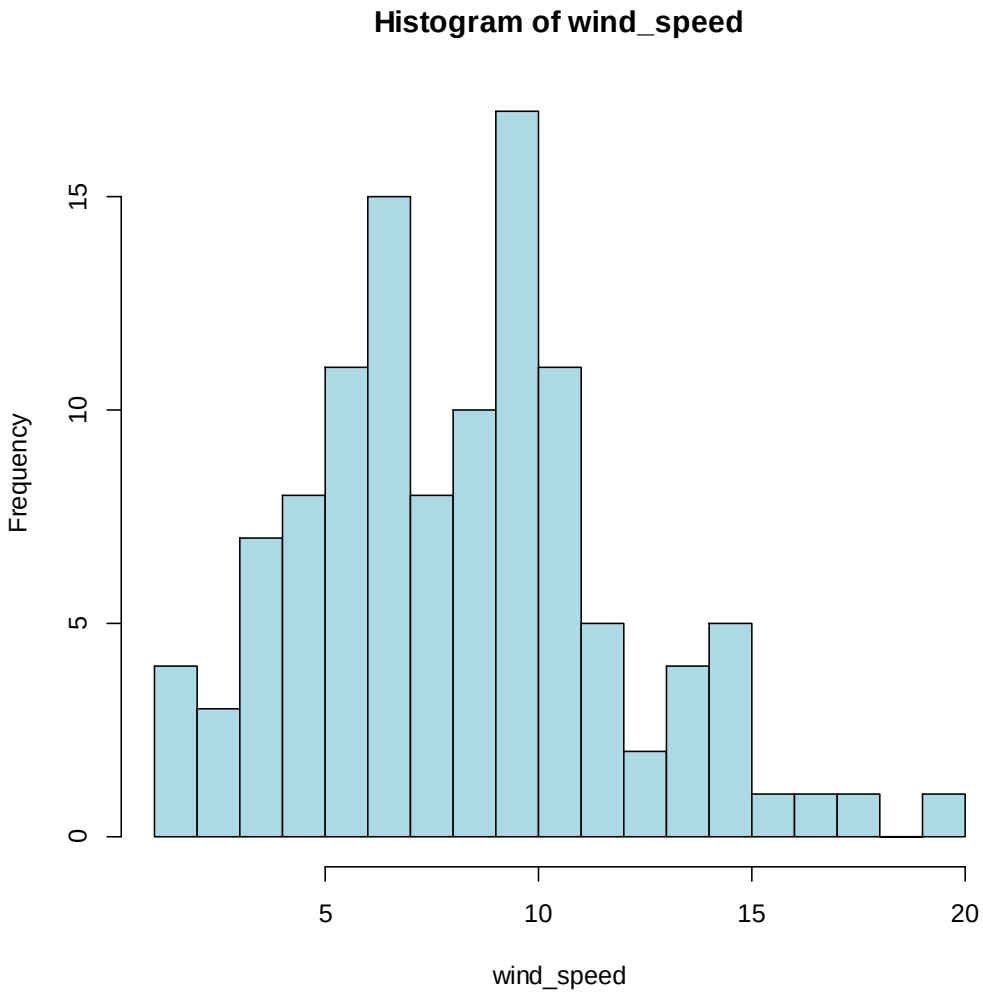
data: data$humidity

W = 0.95813, **p-value** = 0.07278

data: data$wind_speed

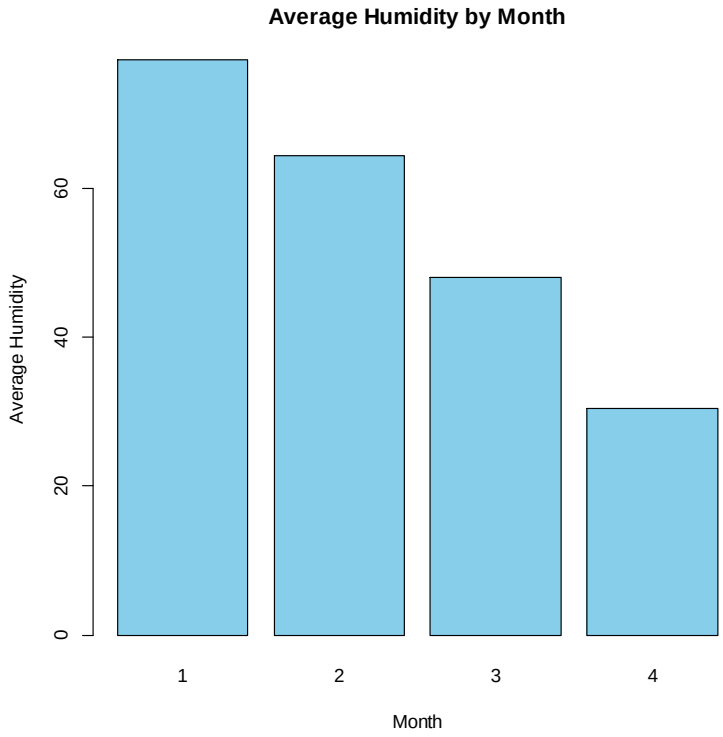
W = 0.97865, **p-value** = 0.06527

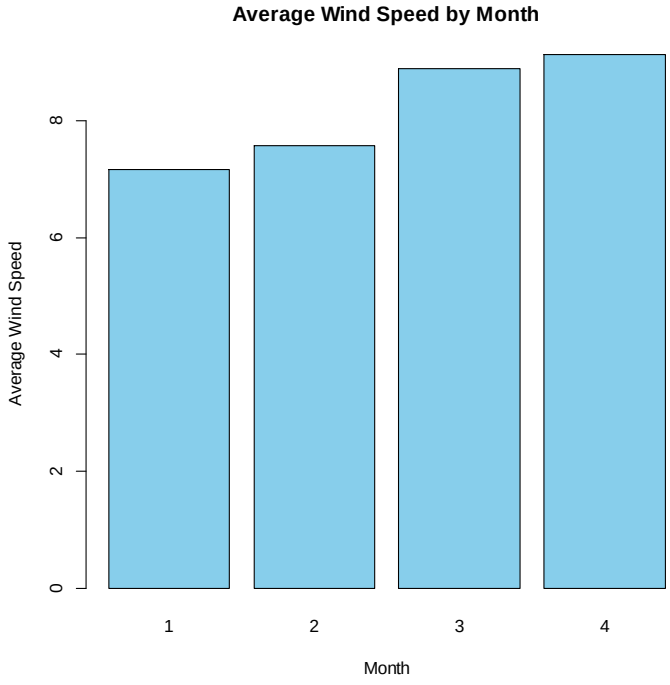

Histogram of Humidity
Frequency
0
5
10
15
20
40
60
80
100
Humidity



Based on the obtained p-values, since the p-values for both variables are higher than 0.05, we conclude that their distributions are normal. Additionally, the histograms for both variables also indicate normality, as they approximately have a bell-shaped curve.

Bar


Average Humidity by Month
Average Humidity
0
20
40
60
1
2
3
4
Month


Average Wind Speed by Month
Average Wind Speed
0
2
4
6
8
1
2
3
4
Month

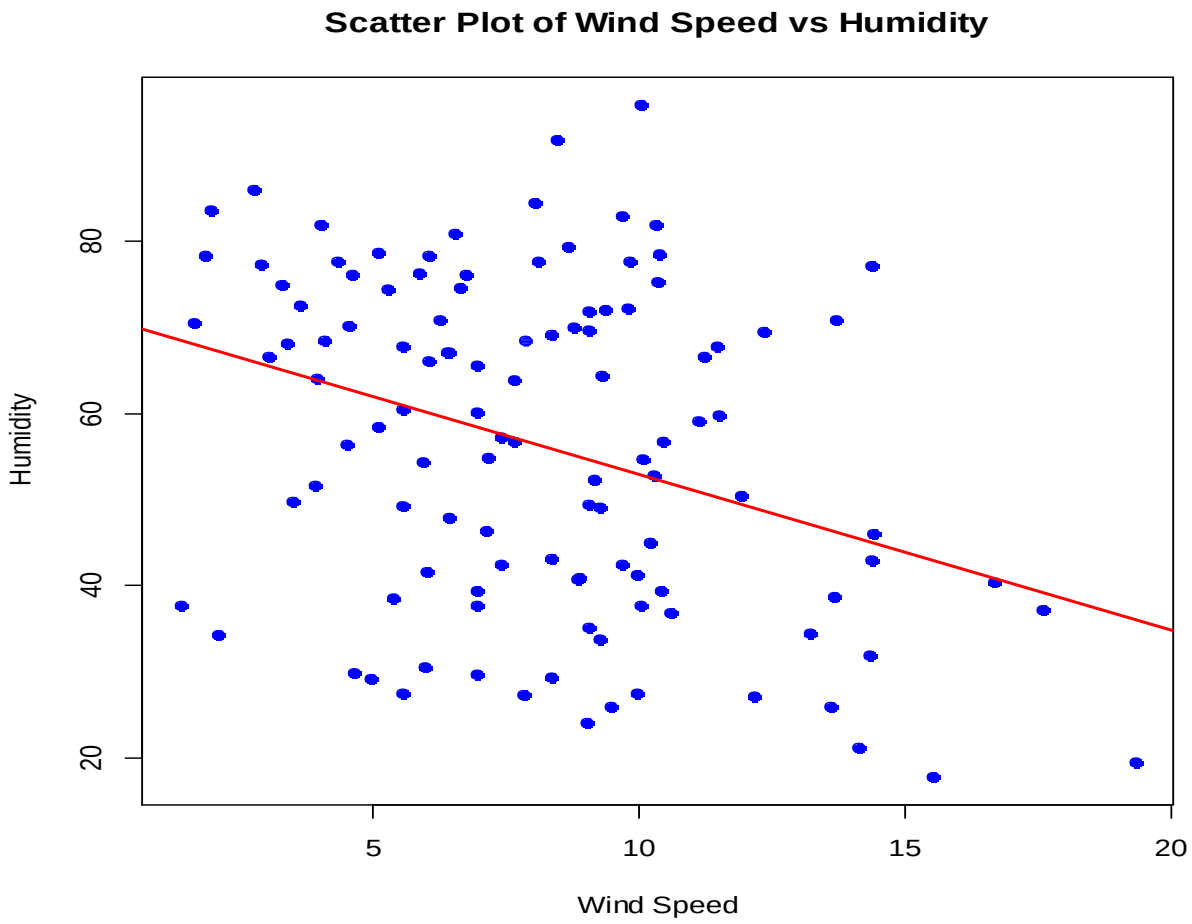
The first bar chart shows the average humidity by month. It indicates that humidity is highest in the first month and gradually decreases over the next three months. The second bar chart illustrates the average wind speed by month. It shows that wind speed is relatively stable in the first two months and increases slightly in the third and fourth months.

Correlation

" Correlation : -0.340506569064783 "

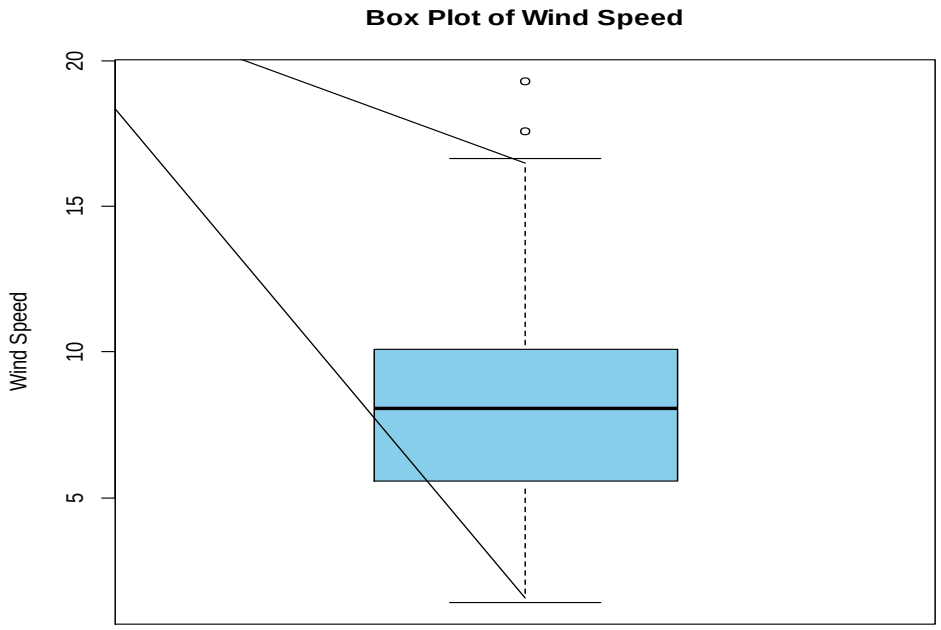
correlation coefficient of -0.34 means that there is a weak negative relationship between the two variables, humidity and wind speed. This indicates that, generally, as wind speed increases, the humidity tends to decrease, but this relationship is not very strong.

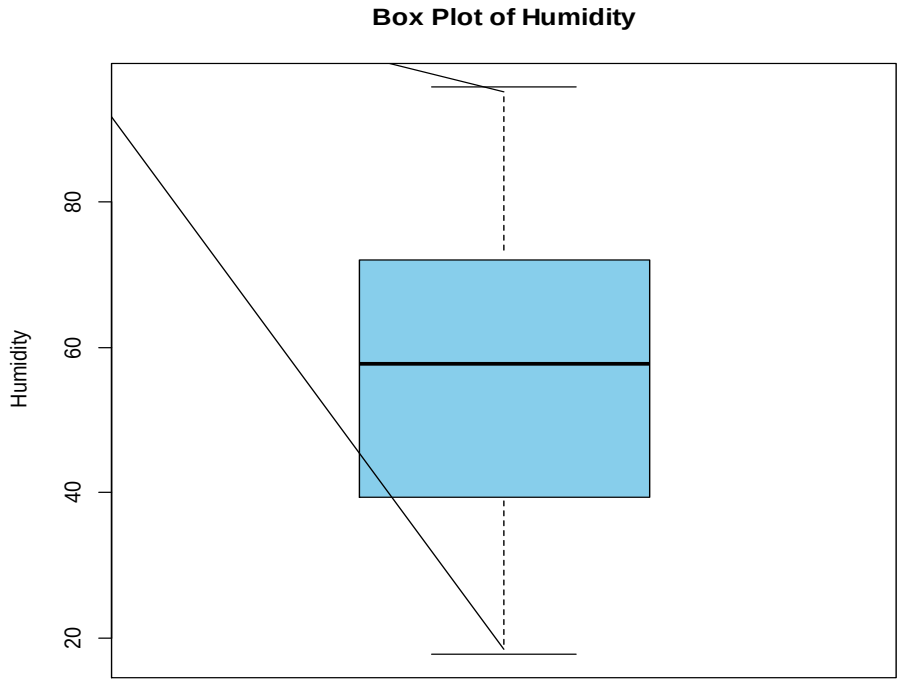
Scatter plot



This scatter plot shows the relationship between wind speed and humidity. Each point represents an observation, with wind speed on the x-axis and humidity on the y-axis. The red line is the line of best fit, indicating a negative correlation. As wind speed increases, humidity tends to decrease, which is consistent with the correlation coefficient of -0.34.

Box Plot

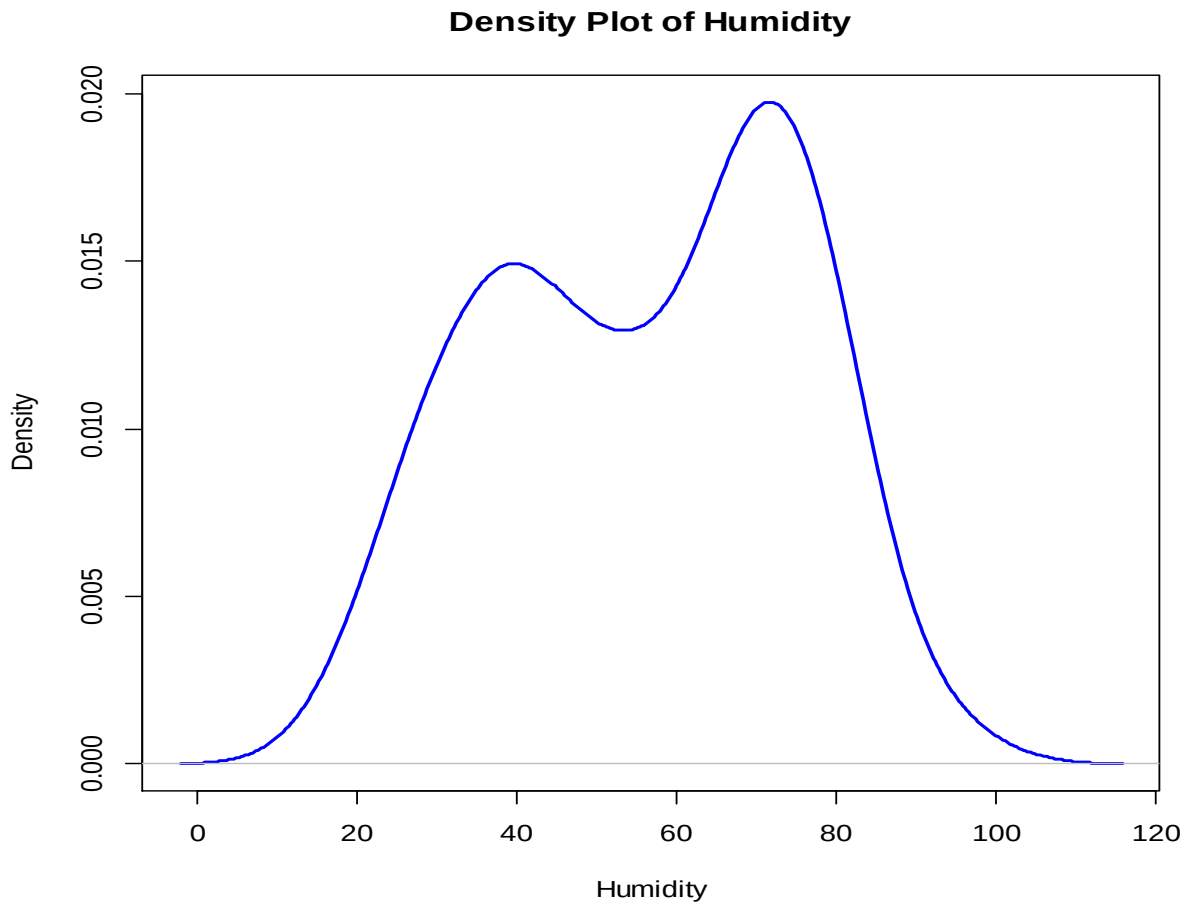

Box Plot of Wind Speed
Wind Speed
20
15
10
5

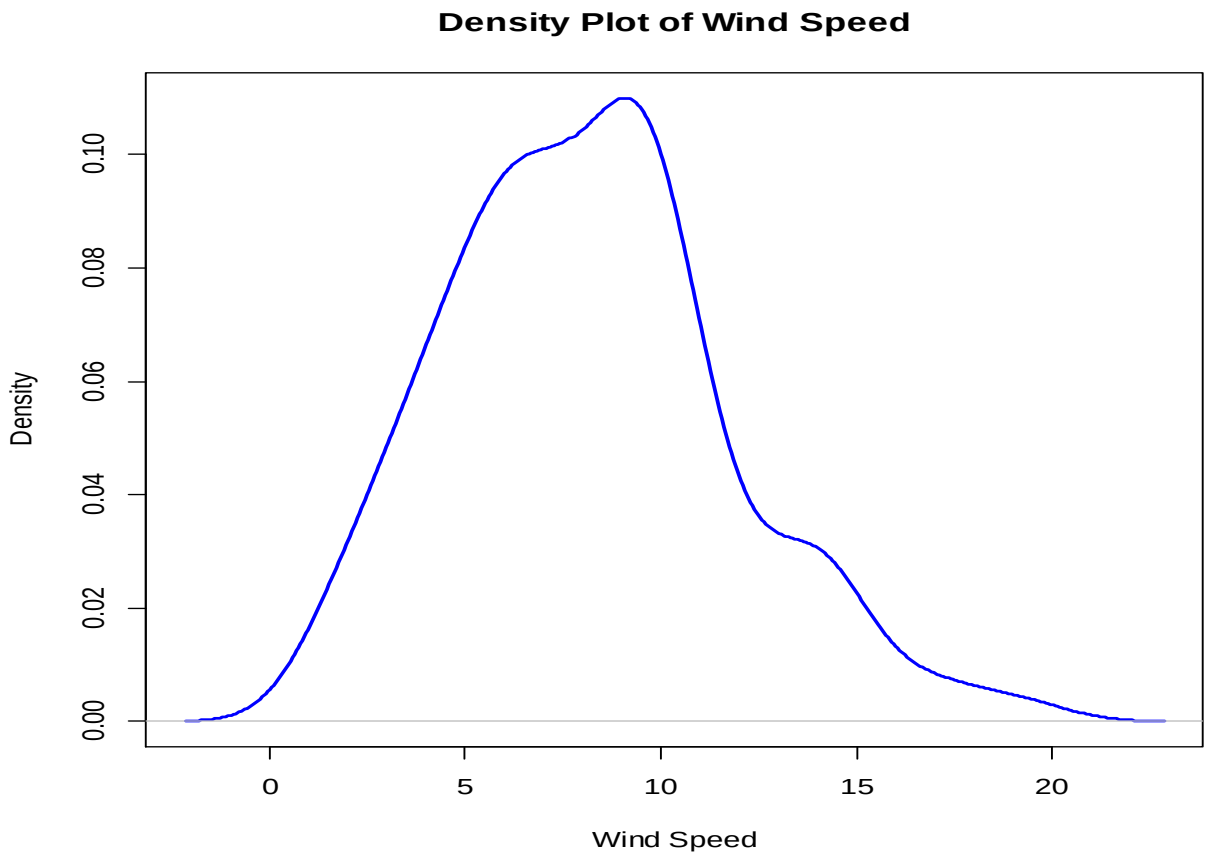

Box Plot of Humidity
Humidity
80
60
40
20

The first box plot displays the distribution of wind speed. The median wind speed is around 10, with an interquartile range (IQR) between approximately 7 and 13. There are a few outliers above 15, indicating some instances of higher wind speeds. The overall spread of the data shows moderate variability.

The second box plot shows the distribution of humidity. The median humidity is around 55, with an IQR between approximately 40 and 70. The data is more spread out compared to wind speed, indicating higher variability in humidity levels. There are no significant outliers, and the distribution appears fairly symmetrical.

Density plot

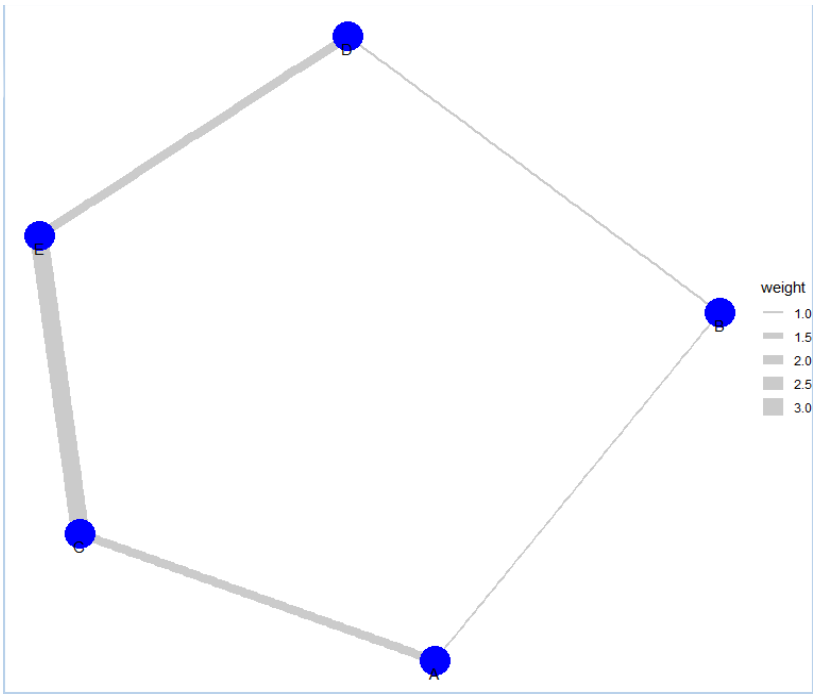


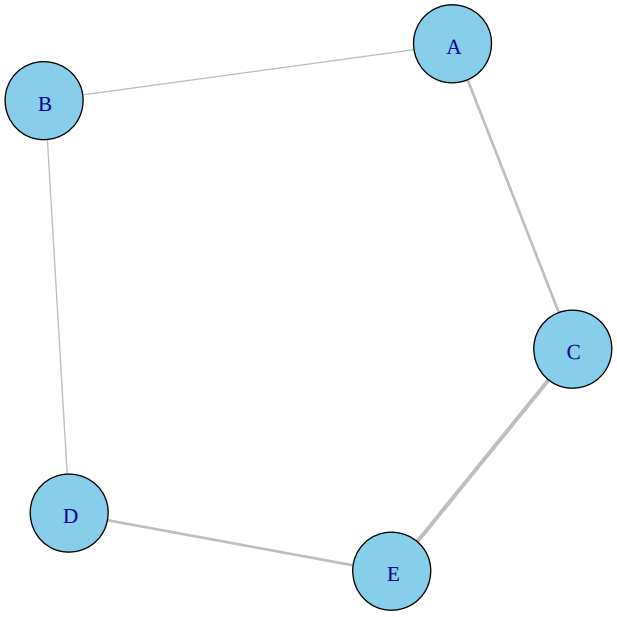


The first density plot illustrates the distribution of humidity. The plot shows a bimodal distribution with two peaks, one around 35 and another around 60, indicating that there are two common levels of humidity in the data. The distribution is fairly spread out, suggesting variability in humidity levels.

The second density plot represents the distribution of wind speed. This plot shows a unimodal distribution with a peak around 8, indicating that this is the most common wind speed in the data. The distribution has a long tail towards higher wind speeds, suggesting that while higher wind speeds are less common, they do occur. The overall shape indicates moderate variability in wind speed.

Network graph





The first graph is a weighted network diagram. The nodes represent different entities, and the edges represent the relationships between them. The thickness of the edges indicates the weight or strength of the relationship, with thicker edges representing stronger connections. For example, the connection between two nodes with a weight of 3.0 is stronger compared to the connection with a weight of 1.0. The blue circles represent the nodes, and the varying shades of gray indicate the different weights of the edges.

The second graph is a simple network diagram. It shows the relationships between five nodes labeled A, B, C, D, and E. Each node is connected to at least one other node, forming a pentagon shape. The edges are uniform in thickness, indicating that all connections are of equal strength or weight. This diagram provides a clear visualization of how the nodes are interconnected without emphasizing the strength of the connections.

Conclusion

Assuming normal data distribution and appropriate coefficient of variation, the analysis indicates a weak negative correlation ($r=-0.34r = -0.34r=-0.34$) between wind speed and humidity. This suggests that as wind speed increases, humidity tends to decrease, but the relationship is not strong. The suitable distribution and graphical analyses confirm a weak yet meaningful correlation. Further investigation into other influencing factors is recommended.