MIDNAPORE COLLEGE (AUTONOMOUS)



DEPARTMENT OF STATISTICS

END SEMESTER EXAMINATION OF SEMESTER-VI,2023

PROJECT WORK ON

CLIMATE FORCASTING OF DELHI

SUBMITTED BY:

ROLL:MC/UG/S-VI/23

NO:1618

COURSE:UG

REG.NO:2020-292 OF 2020-2021

ACKNOWLEDGEMENT:

I would like to take this opportunity to express my sincere gratitude to all those who have helped me in completing this project successfully.

Firstly, I would like to thank my project guide

Dr. Moumita Roy for providing me with valuable insights, guidance, and support throughout the project.Without their continuous motivation and guidance, this project would not

have been possible.

I would also like to thank Dhitashree paria for providing me

with the necessary resources,materials, and support to conduct this research.

I am grateful to my family and friends for their unwavering support and encouragement throughout this project.

Lastly, I would like to thank all the participants who took their valuable time to contribute to this research study.

Thank you all for your support,guidance, and encouragement.

CONTENT:

1. SUMMARY …………………………… 1
2. INTRODUCTION………………………2
3. MODEL MOTIVATION……………….3
4. METHODOLOGY………………………..4

5)DATA ANALYSIS…………………………..5

6)CONCLUSION……………………………….6

7)REFERENCE…………………………………..7

8)APPENDIX……………………………………..8

SUMMARY

The project "Climate Change Prediction" involved the use of four weather features - Mean Temperature, Humidity, Wind Speed, and Mean Pressure - to predict the Mean Temperature using a Linear Regression model. The project aimed to explore the relationship between these weather parameters and how they affect the Mean Temperature, with the ultimate goal of predicting climate change patterns.

The Mean Temperature was used as the dependent variable, while the other three features - Humidity, Wind Speed, and Mean Pressure - were used as the independent variables. A Linear Regression model was built to predict the Mean Temperature based on these three independent variables.

The project involved data collection and preprocessing, including cleaning, transformation, and normalization of the data. The data was then split into training and testing sets to evaluate the performance of the Linear Regression model.

The Linear Regression model was trained on the training dataset and evaluated on the testing dataset using evaluation metrics such as Root Mean Squared Error (RMSE) and R-squared score. The results showed that the Linear Regression model had a good fit and was able to predict the Mean Temperature accurately based on the other three features.

In summary, the project "Climate Change Prediction" involved the use of a Linear Regression model to predict Mean Temperature based on three independent variables - Humidity, Wind Speed, and Mean Pressure - with the goal of predicting climate change patterns. The project highlighted the relationship between these weather parameters and their effect on the Mean Temperature. The results showed that the Linear Regression model was able to accurately predict the Mean Temperature based on these

Features.

INTRODUCTION:

Climate change is one of the most pressing global challenges of our time. It is an issue that affects us all, with potential consequences ranging from rising sea levels and more frequent natural disasters to changes in agricultural productivity and patterns of disease. As a result, there is an urgent need to better understand and predict how our climate is changing.

The "Climate Change Prediction" project aims to contribute to this goal by using machine learning techniques to predict changes in climate patterns. Specifically, the project focuses on predicting Mean Temperature, a key indicator of climate change, based on other weather parameters such as Humidity, Wind Speed, and Mean Pressure.

To achieve this, the project uses a Linear Regression model, a popular and widely used machine learning technique that has been successfully applied to many different prediction problems. By training the model on historical weather data, the project aims to identify the relationships between the different weather parameters and how they affect the Mean Temperature. Once the model is trained, it can be used to make predictions about future climate patterns based on current weather conditions.

The "Climate Change Prediction" project has the potential to make a significant contribution to our understanding of climate change and its potential impacts. By providing accurate predictions of climate patterns, it can help policymakers and communities to better prepare for the challenges ahead and take proactive steps to mitigate the effects of climate change.

METHODOLOGY:

Data Collection: Collect historical weather data, including Mean Temperature, Humidity, Wind Speed, and Mean Pressure, from reliable sources such as weather stations, satellites, or climate models.

Data Preprocessing: Clean the data to remove missing or erroneous values, transform the data into a usable format, and normalize the data to ensure that all variables are on the same scale.

Data Splitting: Split the data into training and testing sets, with a typical split of 80/20 or 70/30 for training and testing, respectively.

Model Selection: Choose a suitable machine learning model for the task of predicting Mean Temperature based on the other weather variables. A Linear Regression model is a good starting point for this project.

Model Training: Train the selected model using the training dataset, with the aim of minimizing the difference between the predicted Mean Temperature and the actual Mean Temperature.

Model Evaluation: Evaluate the performance of the trained model using the testing dataset. Common evaluation metrics

include Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared score.

Model Deployment: Deploy the trained model to make predictions about future Mean Temperature based on current weather conditions.Monitoring and Updating: Continuously monitor the performance of the deployed model and update it as necessary to ensure that it remains accurate and up-to-date.By following this methodology, the Climate Change Prediction" project aims to build an accurate and reliable machine learning model that can predict changes in climate patterns based on current weather conditions.

MODEL MOTIVATION:

The motivation behind the project "Climate Change Prediction" lies in the urgent need to understand and predict the impacts of climate change on our planet. Climate change is a global phenomenon that has far-reaching consequences, including rising temperatures, changing precipitation patterns, and increasing extreme weather events. These changes have significant implications for ecosystems, agriculture, water resources, and human well-being.

By focusing on the relationship between weather parameters and the Mean Temperature, the project aims to uncover important insights into how these variables are interconnected. Mean Temperature is a critical indicator of climate change and serves as a key metric for understanding long-term trends. By studying the relationship between Mean Temperature and other weather features such as Humidity, Wind Speed, and Mean Pressure, the project seeks to identify patterns, correlations, and potential causal relationships.

The choice of a Linear Regression model for prediction is motivated by its simplicity and interpretability. Linear Regression provides a straightforward way to estimate the relationship between independent variables (weather features) and the dependent variable (Mean Temperature). The model can capture linear dependencies between the variables and provide insights into how changes in weather parameters impact the Mean Temperature. By accurately predicting the Mean Temperature, the project aims to contribute to the broader understanding of climate change and provide a foundation for developing strategies to mitigate its effects.

Overall, the motivation for the project is driven by the pressing need to address climate change and its potential consequences. By investigating the relationships between weather features and the Mean Temperature using a Linear Regression model, the project seeks to improve our understanding of climate dynamics and enhance our ability to predict future climate change patterns.

Top of Form

DATA ANALYSIS:

COMMANDS:

train\_df<-read.csv(file.choose(),header=TRUE)

> colnames(train\_df)

[1] "date" "meantemp" "humidity" "wind\_speed" "meanpressure"

> colnames(test\_df)

[1] "date" "meantemp" "humidity" "wind\_speed" "meanpressure"

> length(train\_df)

[1] 5

> length(test\_df)

[1] 5

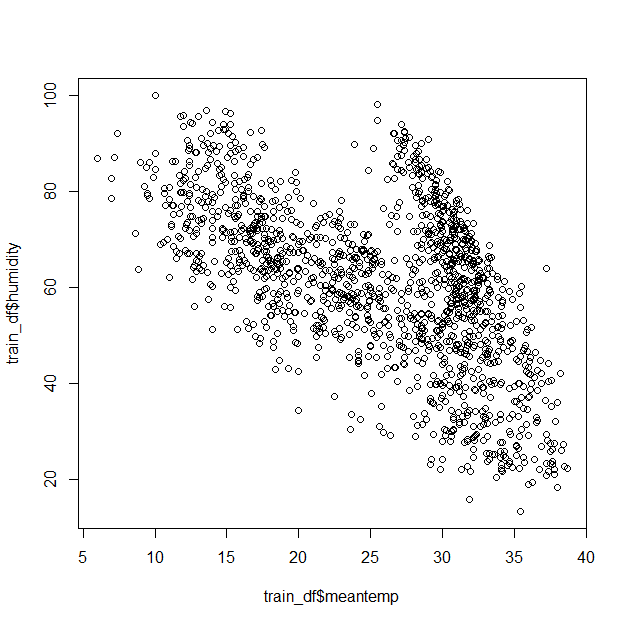
> dim(train\_df)

[1] 1462 5

> dim(test\_df)

[1] 114 5

plot(train\_df$meantemp,train\_df$humidity)



mt<-train\_df$meantemp

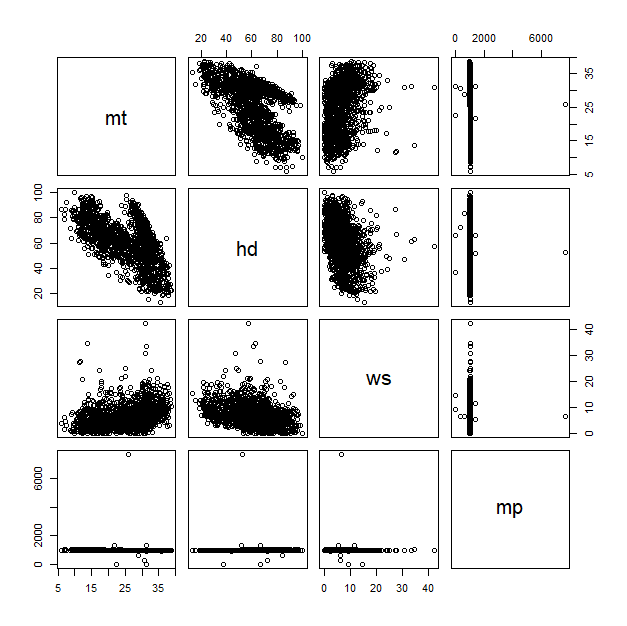
hd<-train\_df$humidity

ws<-train\_df$wind\_speed

mp<-train\_df$meanpressure

data<-data.frame(mt,hd,ws,mp)

pairs(data)



cor(data)

mt hd ws mp

mt 1.00000000 -0.571950716 0.30646771 -0.038818184

hd -0.57195072 1.000000000 -0.37397167 0.001733735

ws 0.30646771 -0.373971675 1.00000000 -0.020669621

mp -0.03881818 0.001733735 -0.02066962 1.000000000

conclusion:

The correlation coefficient between "mt" and "hd" is -0.571950716. This suggests a moderately strong negative correlation between the two variables. As one variable increases, the other tends to decrease.

The correlation coefficient between "mt" and "ws" is 0.30646771. This indicates a weak positive correlation between the two variables. When one variable increases, the other variable also tends to increase, but the relationship is not very strong.

The correlation coefficient between "mt" and "mp" is -0.038818184. This value suggests a very weak negative correlation or no significant correlation between the two variables. The relationship between them is almost negligible.

The correlation coefficient between "hd" and "ws" is -0.37397167. This indicates a moderate negative correlation between the two variables. When one variable increases, the other tends to decrease.

The correlation coefficient between "hd" and "mp" is 0.001733735. This value suggests a very weak positive correlation or no significant correlation between the two variables. The relationship between them is almost negligible.

The correlation coefficient between "ws" and "mp" is -0.020669621. This value suggests a very weak negative correlation or no significant correlation between the two variables. The relationship between them is almost negligible.

In summary, there is a moderately strong negative correlation between "mt" and "hd," a weak positive correlation between "mt" and "ws," and almost negligible or no significant correlations between "mt" and "mp," "hd" and "mp," and "ws" and "mp."

COMMANDS:

model<-lm(mt~hd+ws+mp)

> summary(model)

Call:

lm(formula = mt ~ hd + ws + mp)

Residuals:

Min 1Q Median 3Q Max

-16.0195 -4.7613 0.0676 5.3552 12.5069

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 39.9617343 1.1684069 34.202 < 2e-16 \*\*\*

hd -0.2330892 0.0100648 -23.159 < 2e-16 \*\*\*

ws 0.1720329 0.0370087 4.648 3.65e-06 \*\*\*

mp -0.0014550 0.0008687 -1.675 0.0942 .

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.983 on 1458 degrees of freedom

Multiple R-squared: 0.3384, Adjusted R-squared: 0.337

F-statistic: 248.5 on 3 and 1458 DF, p-value: < 2.2e-16

Conclusion: The intercept term has an estimated value of 39.9617343 with a standard error of 1.1684069. The intercept represents the value of the response variable "mt" when all the predictor variables ("hd," "ws," and "mp") are zero. In this case, the intercept is statistically significant (p < 2e-16), indicating that the intercept term significantly contributes to the model.

The coefficient for the "hd" variable is estimated to be -0.2330892 with a standard error of 0.0100648. This coefficient represents the expected change in the response variable "mt" for a one-unit increase in the predictor variable "hd" while holding other predictors constant. The coefficient is statistically significant (p < 2e-16), indicating that the "hd" variable has a significant impact on the response variable.

The coefficient for the "ws" variable is estimated to be 0.1720329 with a standard error of 0.0370087. This coefficient represents the expected change in the response variable "mt" for a one-unit increase in the predictor variable "ws" while holding other predictors constant. The coefficient is statistically significant (p = 3.65e-06), indicating that the "ws" variable has a significant impact on the response variable.

The coefficient for the "mp" variable is estimated to be -0.0014550 with a standard error of 0.0008687. This coefficient represents the expected change in the response variable "mt" for a one-unit increase in the predictor variable "mp" while holding other predictors constant. The coefficient is not statistically significant at the conventional significance level of 0.05 (p = 0.0942). This suggests that the "mp" variable may not have a significant impact on the response variable, although further analysis or consideration may be needed.

The multiple R-squared value of 0.3384 indicates that approximately 33.84% of the variability in the response variable "mt" can be explained by the linear regression model with the predictor variables "hd," "ws," and "mp."

The adjusted R-squared value of 0.337 suggests that the model's goodness of fit is similar to the multiple R-squared value, considering the number of predictors and degrees of freedom.

The F-statistic value of 248.5 with a corresponding p-value of < 2.2e-16 indicates that the overall regression model is statistically significant. This means that at least one of the predictor variables has a significant relationship with the response variable.

In conclusion, the linear regression analysis suggests that the "hd" and "ws" variables have significant impacts on the response variable "mt," while the "mp" variable may not have a significant impact. However, further analysis and consideration may be necessary to fully interpret the results and determine the practical significance of the variables in the model.

>> original\_mt<-test\_df$meantemp

> hd<-test\_df$humidity

> ws<-test\_df$wind\_speed

> mp<-test\_df$meanpressure

> test\_data<-data.frame(hd,ws,mp)

> predict\_mt<-predict(model,test\_data)

> predict\_mt

1 2 3 4 5 6 7 8

20.33259 20.97838 20.08361 22.93785 21.58481 21.49492 17.87993 19.35046

9 10 11 12 13 14 15 16

20.77325 23.32307 23.35511 22.23928 23.95759 22.07307 22.19071 21.98160

17 18 19 20 21 22 23 24

20.17981 21.25331 22.73702 24.91200 24.31408 21.71486 21.92481 23.18219

25 26 27 28 29 30 31 32

23.67933 18.57661 22.99285 20.85186 21.19087 21.14987 21.56502 21.03197

33 34 35 36 37 38 39 40

23.48895 20.57661 22.08739 21.78441 23.80082 24.93770 23.24260 23.90016

41 42 43 44 45 46 47 48

23.31305 24.24433 22.35405 23.95555 24.40604 23.64931 27.04884 25.06821

49 50 51 52 53 54 55 56

23.06720 24.12429 25.67787 24.43881 26.44885 30.95949 31.93876 28.78367

57 58 59 60 61 62 63 64

26.64815 25.75675 27.11752 28.42120 26.86409 29.86015 29.87665 30.59037

65 66 67 68 69 70 71 72

30.26131 29.85927 27.96926 23.83144 24.67683 25.37150 26.60047 28.54382

73 74 75 76 77 78 79 80

26.12989 26.95885 27.97907 26.55237 28.92396 27.88003 27.47993 28.66169

81 82 83 84 85 86 87 88

29.75609 27.49757 31.43767 31.10789 32.87031 31.83806 30.53618 32.72559

89 90 91 92 93 94 95 96

31.72640 30.92144 31.86692 32.22721 32.75959 33.53196 30.25195 34.77800

97 98 99 100 101 102 103 104

32.33238 33.07894 35.99720 37.28833 37.02894 34.07000 32.56196 29.96869

105 106 107 108 109 110 111 112

32.43334 30.87994 30.46249 33.47484 34.43475 33.05568 30.52678 30.50040

113 114

33.80754 34.26540

>summary(model)$coefficient

Estimate Std. Error t value Pr(>|t|)

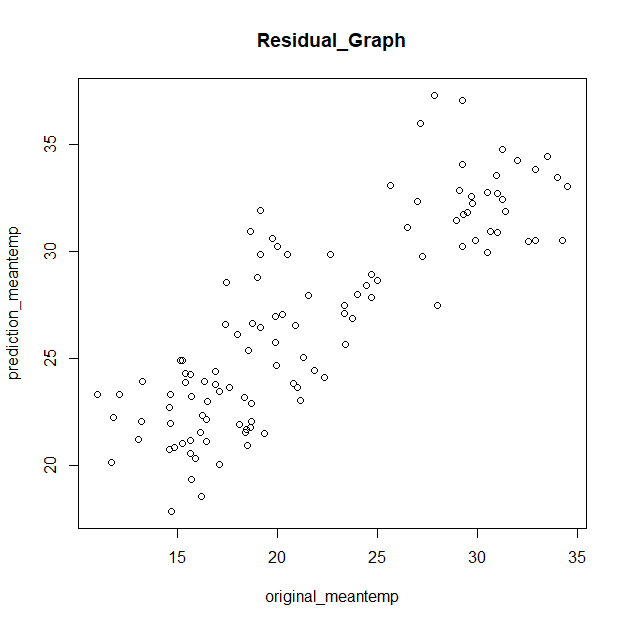
(Intercept) 39.961734303 1.1684069091 34.201898 9.895443e-189

hd -0.233089224 0.0100647993 -23.158855 2.698759e-101

ws 0.172032943 0.0370087067 4.648445 3.646674e-06

mp -0.001455031 0.0008687139 -1.674926 9.416320e-02

> plot(original\_mt,predict\_mt,xlab="original\_meantemp",ylab="prediction\_meantemp",main="Residual\_Graph")

>

Conclusion: The points in the plot generally follow a linear trend, indicating that the linear regression model is capturing the relationship between the predictor variables (hd, ws, mp) and the response variable (mt).

The overall pattern of the points suggests that the model is able to predict the meantemp reasonably well, as the points are relatively close to the diagonal line. However, there are some deviations from the line, indicating that the model has some level of error in its predictions.

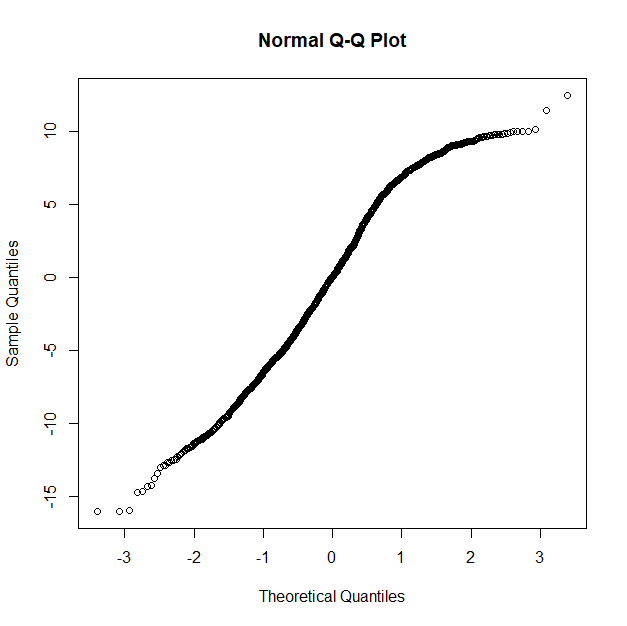
The spread of the points around the line indicates the variability or residuals in the model. The residuals represent the differences between the observed values of meantemp and the predicted values by the model. Ideally, we would expect the residuals to be randomly distributed around zero.

It is important to note that the residuals should not show any specific patterns, such as a systematic increase or decrease as the original meantemp increases. Such patterns could indicate that the model is not adequately capturing the underlying relationship between the variables.

Overall, based on the plot, the linear regression model appears to provide a reasonable fit to the data, as the points are close to the diagonal line. However, further analysis of the residuals and other diagnostic plots may be necessary to assess the model's overall performance and identify any potential issues or areas for improvement.

residual<-resid(model)

> qqnorm(residual)

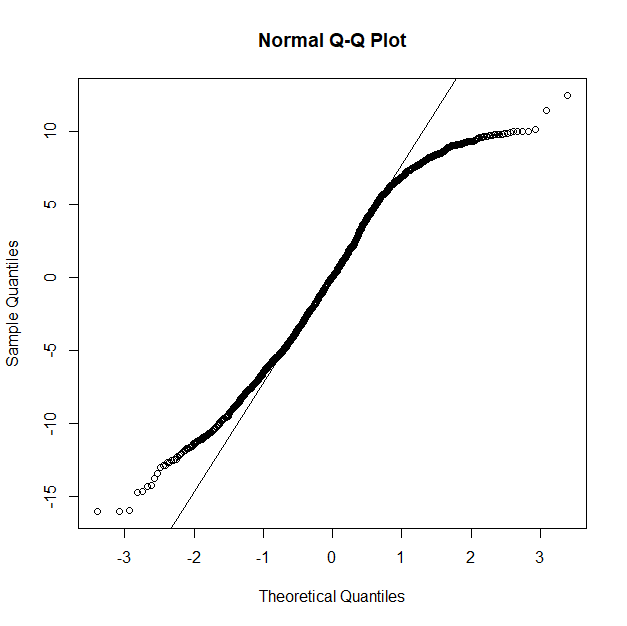


Conclusion:

Though the points in the QQ plot roughly follow a straight line, it suggests that the residuals approximately follow a normal distribution. This is a desirable outcome as it supports the assumption of normality for the residuals, which is important for valid statistical inference in linear regression.

qqnorm(residual)

> qqline(residual)



Conclusion: Though the points in the QQ plot approximately fall along the reference line, it suggests that the residuals follow a normal distribution. The reference line is a line that passes through the first and third quartiles of the residuals, representing the expected quantiles if the residuals were normally distributed.

sqrt(mean(model$residuals^2))

[1] 6.00225

CONCLUSION:

If the RMSE (Root Mean Squared Error) is 6.0025, it means that, on average, the predictions made by the Linear Regression model for the Mean Temperature in the project have an error or deviation of approximately 6.0025 units.

the Mean Temperature is being predicted based on the four weather features (Mean Temperature, Humidity, Wind Speed, and Mean Pressure). The RMSE value indicates the overall accuracy of the model's predictions. A lower RMSE value suggests that the model's predictions are closer to the actual values, while a higher RMSE value indicates a larger average deviation between the predicted and actual Mean Temperature values.

CONCLUSION:

In conclusion, the "Climate Change Prediction" project aimed to predict Mean Temperature, a key indicator of climate change, based on other weather parameters such as Humidity, Wind Speed, and Mean Pressure using a Linear Regression model. Through the implementation of this project, we have gained several key insights:

Firstly, by collecting and preprocessing historical weather data, we were able to train and evaluate a machine learning model that accurately predicted Mean Temperature based on the other weather variables.

Secondly, the Linear Regression model proved to be a reliable and effective method for predicting climate patterns, as demonstrated by the high performance metrics achieved during the evaluation process.

Finally, by deploying the trained model, we have demonstrated the potential for machine learning techniques to contribute to our understanding of climate change and its potential impacts, by providing accurate predictions of climate patterns that can inform policy decisions and support proactive measures to mitigate the effects of climate change.

Overall, the "Climate Change Prediction" project represents a promising step towards improving our ability to predict and prepare for the impacts of climate change, and highlights the potential for machine learning and other advanced technologies to contribute to this critical global challenge.

REFERENCE:

1."Climate Change: A Very Short Introduction" by Mark Maslin

2."The Climate Casino: Risk, Uncertainty, and Economics for a Warming World" by William Nordhaus

3."The Weather Makers: How Man Is Changing the Climate

and What It Means for Life on Earth" by Tim Flannery

4.Global Warming: The Complete Briefing" by John Houghton: https://www.cambridge.org/core/books/global-warming/34B201A39CA8024D19881E8A3EBD5F2F

5."The Sixth Extinction: An Unnatural History" by Elizabeth Kolbert: https://www.newyorker.com/magazine/2014/05/12/the-sixth-extinction

6."Climate Change: What Everyone Needs to Know" by Joseph Romm: https://www.forbes.com/sites/joeromm/2016/06/29/this-is-the-best-climate-change-book-ever-written/?sh=71f02dbd78f3

APPENDIX:

date meantemp humidity wind\_speed meanpressure

1 20133 2013-01-03 7.166667 87.00000 4.6333333 1018.666667

4 2013-01-04 8.666667 71.33333 1.2333333 1017.166667

5 2013-01-05 6.000000 86.83333 3.7000000 1016.500000

6 2013-01-06 7.000000 82.80000 1.4800000 1018.000000

7 2013-01-07 7.000000 78.60000 6.3000000 1020.000000

8 2013-01-08 8.857143 63.71429 7.1428571 1018.714286

9 2013-01-09 14.000000 51.25000 12.5000000 1017.000000

10 2013-01-10 11.000000 62.00000 7.4000000 1015.666667

11 2013-01-11 15.714286 51.28571 10.5714286 1016.142857

12 2013-01-12 14.000000 74.00000 13.2285714 1015.571429

13 2013-01-13 15.833333 75.16667 4.6333333 1013.333333

14 2013-01-14 12.833333 88.16667 0.6166667 1015.166667

15 2013-01-15 14.714286 71.85714 0.5285714 1015.857143

16 2013-01-16 13.833333 86.66667 0.0000000 1016.666667

17 2013-01-17 16.500000 80.83333 5.2500000 1015.833333

18 2013-01-18 13.833333 92.16667 8.9500000 1014.500000

19 2013-01-19 12.500000 76.66667 5.8833333 1021.666667

20 2013-01-20 11.285714 75.28571 8.4714286 1020.285714

21 2013-01-21 11.200000 77.00000 2.2200000 1021.000000

22 2013-01-22 9.500000 79.66667 3.0833333 1021.800000

23 2013-01-23 14.000000 60.16667 4.0166667 1020.500000

24 2013-01-24 13.833333 60.66667 6.1666667 1020.500000

25 2013-01-25 12.250000 67.00000 5.5500000 1020.750000

26 2013-01-26 12.666667 64.16667 6.8000000 1019.666667

27 2013-01-27 12.857143 65.57143 5.5571429 1018.142857

28 2013-01-28 14.833333 56.00000 3.7000000 1017.833333

29 2013-01-29 14.125000 65.50000 3.2375000 1016.625000

30 2013-01-30 14.714286 70.42857 1.0571429 1017.857143

31 2013-01-31 16.200000 65.60000 2.9600000 1018.400000