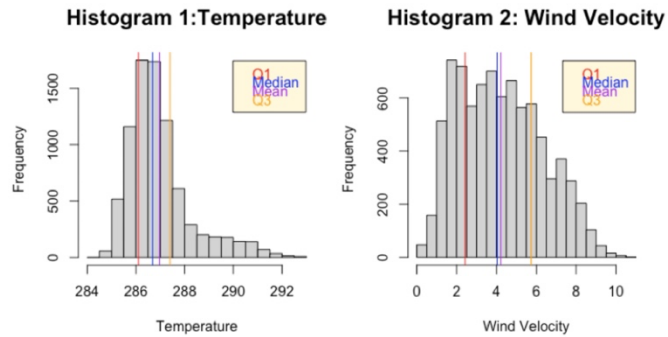


STAT0004, 2021: cover page

Please describe how each member of your group has contributed in this project:

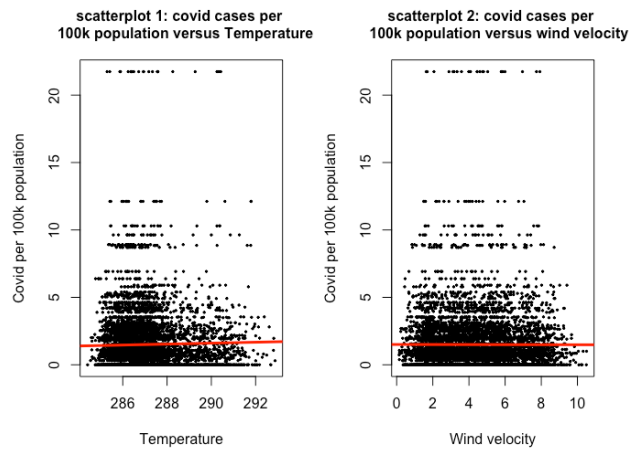
Everyone contributed equally.

Task 1



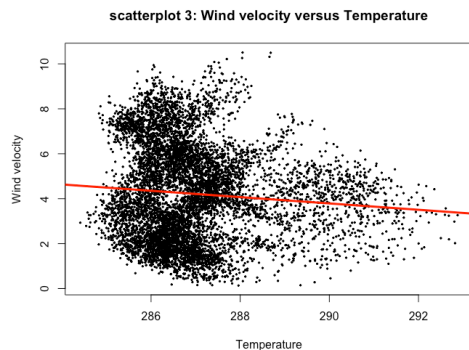
The skewness of temperature is calculated as 1.430036, indicating that the data is positively skewed. From the histogram, we can conclude that temperatures in the UK fluctuate between 285 and 288, in which the lower quartile and upper quartile lie. Moreover, the frequency of measurements decreases as the temperature observed becomes extreme. The mean and median are 287 and 286.7, respectively.

The skewness of wind velocity is 0.3306, showing that the data is slightly positively skewed. From histogram 2, we can see that the values of wind velocity fluctuate between 2.4240 and 5.7403, in which the lower quartile and upper quartile lie. Furthermore, the mean and median are 4.2190 and 4.0294.



From scatterplot 1, although there seems to have a positive association between covid cases per 100k population and temperature, the correlation is so weak that the association is negligible. The correlation coefficient between temperature and covid cases per 100k population is 0.021, which suggests nearly no relationship between these two variables. The second scatterplot shows that the correlation coefficient is very close to zero, which means there is no association between wind velocity and covid cases.

Since the absolute value of coefficient (0.021) between temperature and covid cases is slightly higher, it might suggest that the temperature affects more than the wind velocity on covid infections. (The coefficient in scatterplot 2 is -0.0021)



The negative relationship between wind velocity and temperature indicates that higher temperature is associated with lower wind velocity in scatterplot 3. What's more, the strongest association is between wind velocity and temperature among the three scatterplots, whose correlation coefficient is -0.088.

Task 2

(a) The measurements of average temperature or velocity in England and non-England areas are shown below separately.

STAT0004, 2021: written report

	2020-06-15	2020-06-16	2020-06-17	2020-06-18	2020-06-19	2020-06-20	2020-06-21	2020-06-22	2020-06-23	2020-06-24	2020-06-25
avg temp England	286.916100	286.649183	286.510510	286.417415	286.431462	287.066523	287.089931	286.896803	288.114201	289.151026	289.852715
avg temp other	286.928231	286.673008	286.618708	286.595130	286.422736	286.999805	286.990916	286.802871	287.926617	288.913682	289.790026
avg v.wind England	1.449500	1.556171	2.153087	2.290100	2.327094	4.820298	5.645670	6.656025	4.782199	2.717471	3.362257
avg v.wind other	1.329956	1.394371	2.228533	2.194408	2.109448	4.906284	5.651331	6.918127	4.866216	2.291540	2.994460

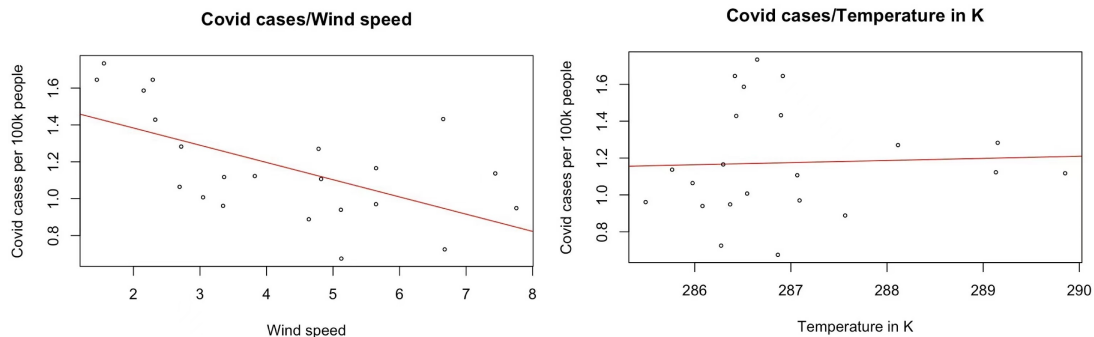
	2020-06-26	2020-06-27	2020-06-28	2020-06-29	2020-06-30	2020-07-01	2020-07-02	2020-07-03	2020-07-04	2020-07-05	2020-07-06
avg temp England	289.132281	287.564655	286.274127	286.296110	286.544559	285.977481	285.489180	286.080368	286.864779	286.368117	285.764524
avg temp other	288.884561	287.431710	286.184330	286.252586	286.488800	285.904899	285.507229	286.045597	286.774970	286.232793	285.717087
avg v.wind England	3.823623	4.635758	6.678058	5.646376	3.045786	2.693634	3.346757	5.117661	5.123902	7.755114	7.438746
avg v.wind other	3.637942	4.397708	6.532891	5.731817	2.778447	2.413845	3.398501	5.336075	5.352004	7.821391	7.348998

(b) By carrying out the student-t test, we test the presenter's claims about the average wind velocity and temperature across England and non-England areas separately. We set the null hypothesis as the rest of the UK is warmer than England and the rest of the UK is windier than England, respectively, for each test. For the temperature, the result shows an associated p-value of 0.003871, which is significant (<0.05). Therefore, the null hypothesis is rejected. For the velocity, it has an associated p-value of 0.06575, which is larger than the significance level (0.05). Hence there is insufficient evidence to reject the null hypothesis.

(c) In conclusion, there is enough evidence to support the presenter's claims that England is warmer in the period 2020-06-15 to 2020-07-06 compared to the rest of the UK and the rest of the UK was windier than England.

Task 3

(a) We build two simple linear models testing the correlation between covid cases, wind speed, and temperature separately. To test the epidemiologist's claims, we use two hypothesis tests, with the null hypothesis H_0 claiming that there is no monotonic linear relationship between the daily number of covid infections per 100k population in England and wind velocity.



Based on the plot conducted above, we could conclude that the covid infections and wind velocity have a negative linear relationship. The plot for covid cases and temperature, in contrast, composes a horizontal line, indicating no correlation between the two variables.

The p-value for linearity between covid cases and wind speed is 0.003424, which is significant (<0.01), and hence there is enough evidence to reject H_0 and accept H_1 .

The p-value for linearity between covid cases and temperature is 0.8442, greater than the significance level (0.01). Therefore, there is insufficient evidence to reject H_0 .

(b) In conclusion, there is enough evidence to support the epidemiologist's claim about the monotonic linear relationship between the daily number of covid infections per 100k population in England and wind velocity. However, there is no sufficient evidence to deduct the monotonic linear relationship between the daily number of covid cases and temperature in England.