DATA SCIENCE REPORT ON CLIMATE DATA

REPORT

Arinze Ezirim*
School of Collective Intelligence
University Mohammed VI Polytechnic
Rabat, Morocco
arinze.ezirim@um6p.ma

1 Study Description

The dataset for analysis was retrieved from `https://userpage.fu-berlin.de/soga/data/raw-data/Climfrance.csv`. It contained data on latitude, longitude, altitude, station, etc., from France with about 36 observations and 12 unique features, which were used to estimate the mean annual temperature in different stations in France.

2 Data Description and Exploration

The dataset has twelve variables:

- **Station name**
- **Altitude**
- **Latitude**
- **Longitude**
- **Mean Temperature**
- **Maximum Temperature**
- **Minimum Temperature**
- **Relative Humidity**
- **Mean Precipitation**
- **Maximum Precipitation within 24 hours**
- **Number of Rainy days**
- **Number of Sunshine hours per year**

Table 1: Summary Statistics of Climate Data

	Altitude	Latitude	Longitude	T-Mean
Min	3.00	42.55	-4.41	-1.20
1st Quartile	46.25	44.00	0.608	10.47
Median	96.00	45.80	3.100	11.15
Mean	249.94	46.19	2.763	11.10
3rd Quartile	214.00	48.16	5.228	12.35
Max	2860.0	50.73	9.48	16.40

Below is the map showing the stations in France for the data collection.

*

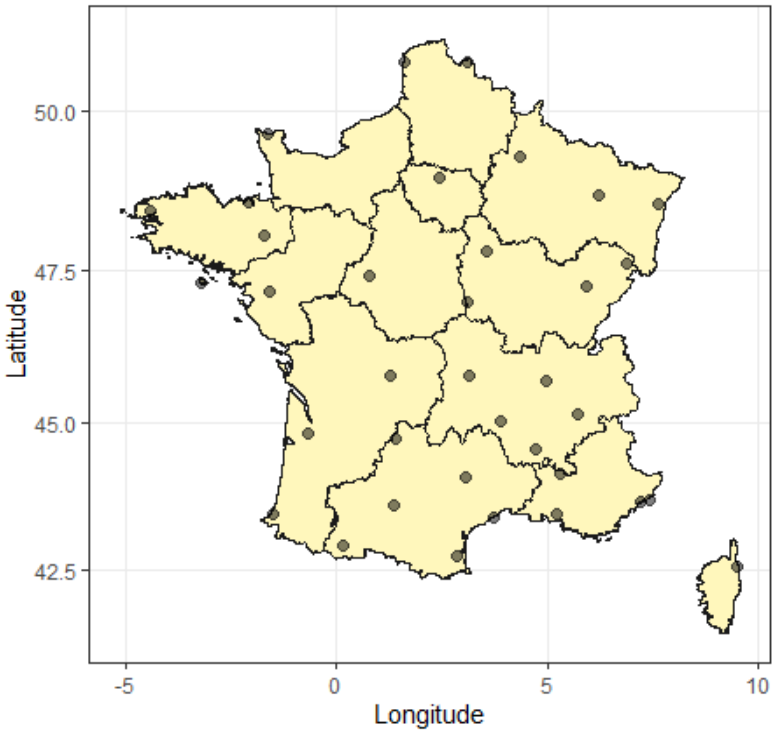

Figure 1: Map showing the stations in France

3 Model 1

Model$_1 = lm(t - mean\ altitude + latitude + longitude, data = climfrar)$

3.1 Model Coefficients

Table 2: Regression Coefficients and Statistical Significance

	Estimate	Std Error	T value	Pr(¿—t—)
Intercept	37.2650	2.6220	14.212	7.29e-15
Altitude	-0.00641	0.00087	-7.383	3.17e-08
Latitude	-0.5340	0.05575	-9.577	1.24e-10
Longitude	0.0321	0.03957	0.811	0.424

Mean-Annual-Temperature = 37.265 - 0.006414*altitude - 0.534*latitude + 0.0321*longitude

3.2 Model Interpretation

- In Model 1:
  - The intercept shows the average mean annual temperature of 37.265 for altitude, latitude and longitude with zero values (a= 37.265, SE = 2.622, p¡0.001)
  - Altitude is negatively associated with mean annual temperature; for every increase in altitude by one unit, the average mean annual temperature decreases by 0.0064 units(b= (b = 0.0064, SE = 0.00087, p¡0.001)
  - Latitude is negatively associated with mean annual temperature; for every increase in latitude by one unit, the average mean average temperature decreases by 0.534 units(c= (c = 0.534, SE = 0.0557, p¡0.001)

- Longitude is positively associated with mean annual temperature; for every increase in longitude by one unit, the average mean average temperature increases by 0.0321 units(d= (d = 0.0321, SE = 0.03957, p¿0.001)

4 Model 2

$\text{Model}_2 = lm(t - mean\ altitude + latitude, data = climfrar)$

4.1 Model Coefficients

Table 3: Regression Coefficients and Statistical Significance

	Estimate	Std Error	T value	Pr(¿—t—)
Intercept	37.9148	2.4829	15.27	5.68e-16
Altitude	-0.00626	0.00084	-7.42	2.34e-08
Latitude	-0.54653	0.05326	-10.26	1.72e-11

Mean-Annual-Temperature = 37.9147 - 0.00626*altitude - 0.547*latitude

4.2 Prediction

Table 4: Predicted mean annual temperature

Location	Estimate	Lower bound	Upper bound
Mont-Ventoux	6.1657	3.792	8.539
Pic-du-midi	-3.4473	-8.348	1.453

4.3 Prediction Interpretation

- In Model 2:
  - The predicted mean annual temperature of Mont-Ventoux is 6.17 with a 95% confidence that the predicted mean annual temperature falls within the range of 3.79 to 8.54.
  - The predicted mean annual temperature of Pic-du-midi is -3.45 with a 95% confidence that the predicted mean annual temperature falls within the range of -8.35 to 1.45.

5 3DScatterplot for Model 2

5.1 Model Summary Output

Table 5: Regression Coefficients and Statistical Significance

	Estimate	Std Error	T value	Pr(¿—t—)
Intercept	37.9148	2.4829	15.27	5.68e-16
Altitude	-0.00626	0.00084	-7.42	2.34e-08
Latitude	-0.54653	0.05326	-10.26	1.72e-11

5.2 Model 2 Result Interpretation

- In Model 2:
  - The intercept shows the mean annual temperature of 37.915 for a altitude, latitude and longitude with zero values (a= 37.915, SE = 2.483, p¡0.001)
  - Altitude is negatively associated with mean annual temperature, for every increase in altitude by one unit the average mean annual temperature decreases by 0.00636 units(b= 0.00364,SE = 0.000844, p¡0.001)

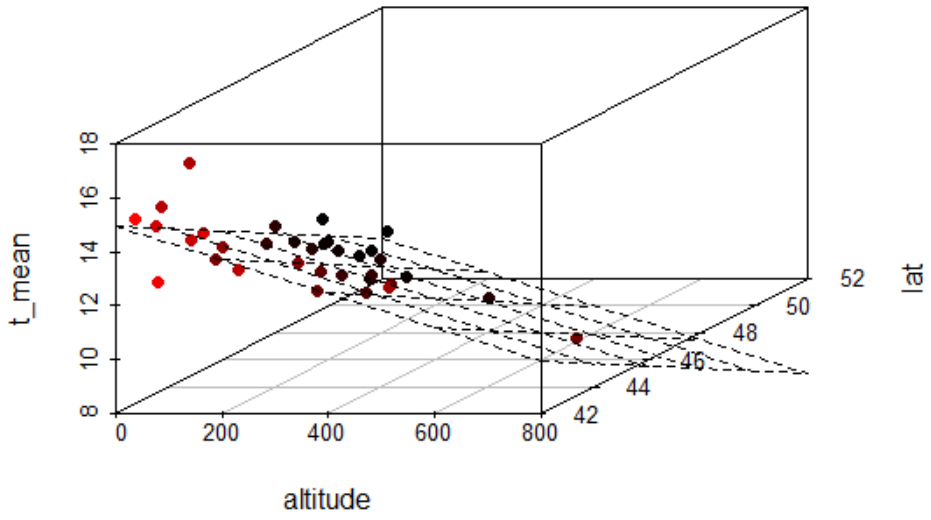

Figure 2: 3D Scatterplot for Model 2

Table 6: Model Fit Statistics

F Statistics	P value	R squared	Adjusted R squared
75.26	1.268e-12	0.8292	0.8182

- Latitude is negatively associated with mean annual temperature, for every increase in latitude by one unit the average mean annual temperature decreases by 0.547 units(c= 0.547,SE = 0.0533, p¡0.001)
- The R squared value of 83% shows the proportion of variance in the mean annual temperature that is explained by altitude and latitude.
- The associated p value of 1.268e-12 suggests that the independent variables altitude and latitude are significantly related to mean annual temperature.

R-Code

Link to the R code: `https://github.com/Youngprof3/climate_data_analysis`