Investigating the Minimum Temperature of Minas Gerais's Mesoregions (2018-2022)

Alexandre de Carvalho Assunção

November 2023

Introduction

This project aims to analyze the minimum temperature in the mesoregions of the state of Minas Gerais during the years 2018-2022. There are over 60 working meteorologic stations within Minas Gerais providing the minimum temperature they registered from 2018 to 2022. They collected the data daily, and so for better visualization, I used the yearly mean temperature for each station. By the end of this project, I hope to see if there are any noticeable patterns of temperature change.

Data Processing

The data for the meteorologic stations was obtained from the INMET tables for the meteorologic stations with the limits of Minas Gerais. These tables were divided into two sections: a) a header containing information about the station, such as the station's name, code, and, more importantly, its latitude and longitude, and b) a data frame of over twenty meteorologic variables, from which I extracted the minimum temperature and the date, which are the relevant variables for this project.

The data provided by the tables was far from perfect. Every single year, every single station failed to measure the temperature in at least some hour of the day, registering instead a dummy variable on the order of -9999, which I needed to find and treat as missing data instead of a temperature below absolute zero. I chose to do this with the Z-Score metric:

The Z-score (Z) for the variable `Min_Temp` is calculated using the formula:

$$Z = \frac{X - \mu}{\sigma}$$

where:

- Z is the Z-score,
- X is the individual data point of `Min_Temp`,

- μ is the mean of `Min_Temp`,
- σ is the standard deviation of `Min_Temp`.

If the Z-score (Z) for the variable `Min_Temp` does not satisfy $-3 < Z$ or $Z > 3$, I treat it as an outlier since it's over three standard deviations from the mean. This proved sufficient to find when the stations failed to register the minimum temperature.

However, the Z-score doesn't work when the station fails to register the minimum temperature throughout the *entire* year. Since all data is around -999, every single row falls within three σ from μ. These tables I removed manually.

After removing outliers, it was a simple matter to use `geopandas` to create the following data frame, which I used for the rest of this project:

Year	Station	Latitude	Longitude	Yearly μ_{Temp}	Yearly σ_{Temp}
2018	BARBACENA	-21.22888888	-43.76694443	17.848	3.669
2018	ARAXÁ	-19.60555555	-46.94944444	21.163	3.659
2018	MONTES CLAROS	-16.71666666	-43.86666666	23.151	4.424
2018	...	...	...	...	

Table 1: Average Minimum Yearly Temperature by Station

The Geographic Database

I want to begin this section by justifying the use of Minas Gerais's mesoregions for this project, which seems like an arbitrary choice, given IBGE provides shapefiles for microregions and municipalities. If we're investigating temperature variations, wouldn't it make more sense to use smaller regions? Well...

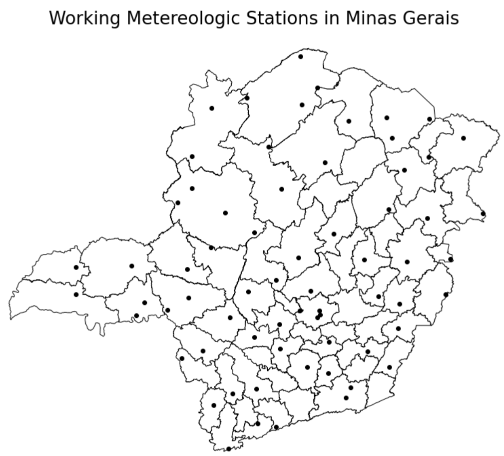

Figure 1: Stations over MG's Microregions.

As we can see, several microregions don't have any stations in them and the case for municipalities is even worse. But, when we use the mesoregions:

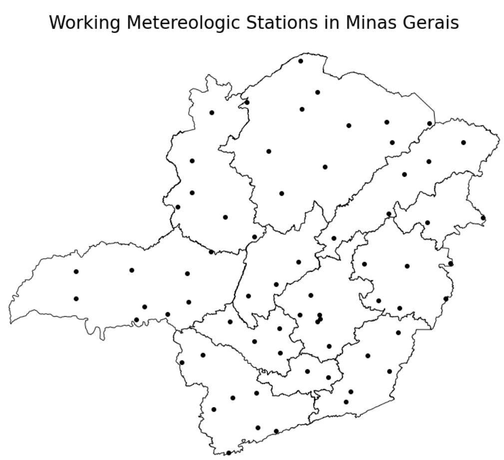

Figure 2: Stations over MG's Mesoregions.

Now every mesoregion has at least one station. We can assign them meaningful values. So let's discuss how to do that. We have a) a data frame with latitude, longitude, and average temperature, and a shapefile of the state of Minas Gerais. Luckily, all the functions we need to join them into one geographic data frame can be found in Python's `geopandas` package.

I did this in three steps:

- Turn the latitude and longitude tables of our non-geographic data frame, currently two `floats`, into the geometric point type of `geopandas`:

```
geometry = gpd.points_from_xy(df['Longitude'], df['Latitude'])
stations = gpd.GeoDataFrame(df, geometry=geometry)
#
```

Figure 3: Turning floats into xy points with `geopandas`.

- Reproject the points into the appropriate projection, which I found in the Minas Gerais's shapefile *.prj* file.

```
stations.crs = 'EPSG:4674'
stations = stations.to_crs(MG.crs)
```

Figure 4: Reprojects the points.

- Finally we merge the data frames by their coordinates and take the average minimum temperature by region with the following query:

```
SELECT MEAN(Stations.Min_Temp())
FROM Merged Map
GROUP BY Cod_Region
```

In `geopandas` we can use the following code:

```
map_with_stations = gpd.sjoin(stations, MG, how='left', op='within')

avg_temp_by_region = map_with_stations.groupby('CD_MESO')['Mean_Temp'].mean().reset_index()

map_with_avg_temp = MG.merge(avg_temp_by_region, on='CD_MESO', how='left')
```

Figure 5: Merges data frames and calculates mean temperature by region.

1 Results

After creating the merged map, producing the results was only a matter of plotting the data frames using Python's `matplotlib` package. To look at the code you can check the MapMaker.py file included with this report. Here are the maps:

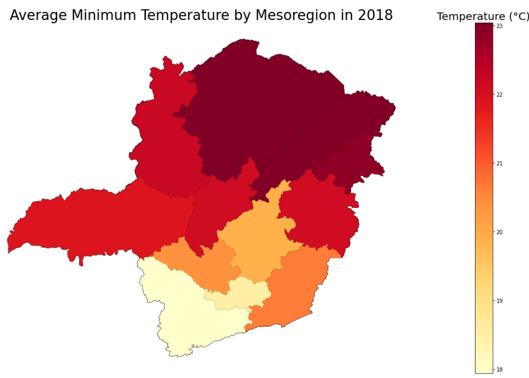

Figure 6: Minimum Temperature 2018.

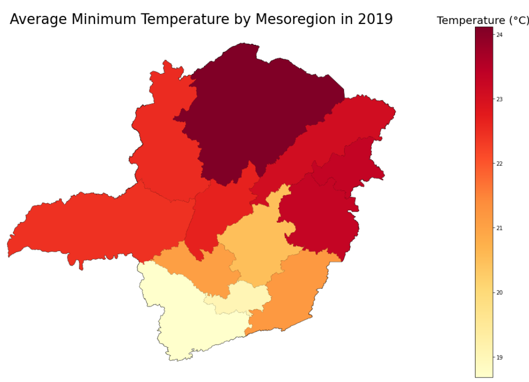

Figure 7: Minimum Temperature 2019

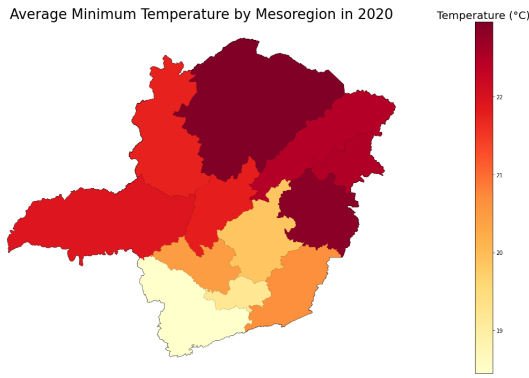

Figure 8: Minimum Temperature 2020

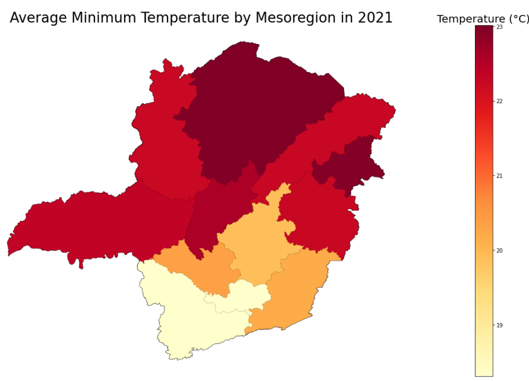

Figure 9: Minimum Temperature 2021

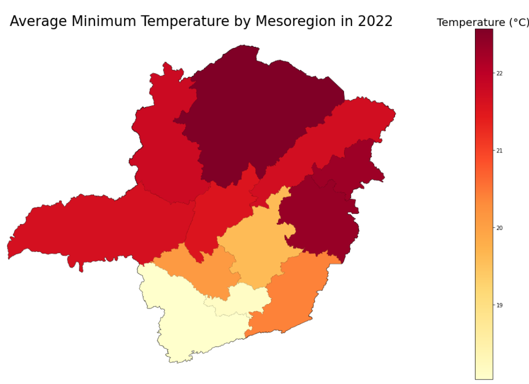

Figure 10: Minimum Temperature 2022

2 Conclusion

We can observe some obvious patterns, such as the northern parts of the state having higher temperatures than the southern ones, but unfortunately, as the table below shows, there is no universal trend for the entire state, at least, not within the period studied.

Year	Yearly μ_{Temp}	Yearly σ_{Temp}
2018	21.21	1.73
2019	21.82	1.73
2020	21.24	1.49
2021	21.22	1.73
2022	20.84	1.58

Table 2: State-wide Yearly Average Minimum Temperature.

The shapefiles produced in this project and the processed tables can be found in the
Resultados_Shapefiles folder, uploaded on Google Drive since it was too big to submit on Moodle. You can access it by clicking on the folder name in this report.
The Python code, `tp2_DataProcessing.py`, and `tp2_MapMaker.py` are included with this submission and can be executed after adjusting the paths for the inputs.

Thank you for reading.