Second Task- Visualization on weather data

Maria Cicone, Giorgia Leonardi, Miro Confalone

### Data manipulation

We retrieved the three .csv files from the Github project of our colleagues regarding the weather and decided to merge them into a single one using the left\_join() function. We found some NAs and filled them with the appropriate values. Given that we found a dataset with a large number of different cities (51), we decided to group their respective countries in regions by creating a small dataframe and attaching it to the “df” dataset through the aforementioned function. In this way, we added a new column to the dataset.

library(tidyverse)  
library(readr)  
library(lubridate)  
library(hms)  
library(RColorBrewer)  
library(harrypotter)  
library(viridis)

daily\_weather <- read\_csv("https://raw.githubusercontent.com/MarcoAmadori1/API\_project/main/6\_daily.csv")  
countries<-read\_csv("https://raw.githubusercontent.com/MarcoAmadori1/API\_project/main/results.csv")  
hourly\_data<-read\_csv("https://raw.githubusercontent.com/MarcoAmadori1/API\_project/main/hourly.csv")  
  
countries<-replace\_na(data=countries,replace=list(country="Vatican City",country\_code="va"))  
  
  
countries\_reviewed<- countries%>%  
 select(country, country\_code, capital\_cities, lat, long)  
joined\_data <- hourly\_data %>%  
 left\_join(countries\_reviewed, by="capital\_cities")   
  
joined\_data<-select(joined\_data, select=-c("...1"))  
  
daily\_reviewed<-select(daily\_weather, select=-c("...1", "weathercode"))  
  
df<- joined\_data%>%  
 left\_join(daily\_reviewed, by=c("time","capital\_cities"))  
  
df<-df %>%  
 fill(everything())  
  
country<-c("Albania","Andorra","Armenia","Austria","Azerbaijan","Belarus","Belgium","Bosnia and Herzegovina","Bulgaria","Croatia","Cyprus","Czechia","Denmark","Estonia","Finland","France","Georgia","Germany","Greece","Hungary","Iceland","Ireland","Italy","Kazakhstan","Kosovo","Latvia","Liechtenstein","Lithuania","Luxembourg","Malta","Moldova","Monaco","Montenegro","Netherlands","North Macedonia","Norway","Poland","Portugal","Romania","Russia","San Marino","Serbia","Slovakia","Slovenia","Spain","Sweden","Switzerland","Turkey","Ukraine","United Kingdom", "Vatican City")  
region<-c("Southern Europe","Southern Europe","Caucasian Region","Central Europe","Caucasian Region","Eastern Europe","Western Europe","Southern Europe","Eastern Europe","Southern Europe","Southern Europe","Central Europe","Northern Europe","Northern Europe","Northern Europe","Western Europe","Caucasian Region","Central Europe","Southern Europe","Central Europe","Northern Europe","Northern Europe","Southern Europe","Eastern Europe","Southern Europe","Northern Europe","Western Europe","Northern Europe","Western Europe","Southern Europe","Eastern Europe","Western Europe","Southern Europe","Western Europe","Southern Europe","Northern Europe","Central Europe","Southern Europe","Eastern Europe","Eastern Europe","Southern Europe","Southern Europe","Central Europe","Central Europe","Southern Europe","Northern Europe","Central Europe","Caucasian Region","Eastern Europe","Northern Europe", "Southern Europe")  
geographic\_position<-data.frame(country, region)  
  
df<-df%>%  
 left\_join(geographic\_position, by="country")

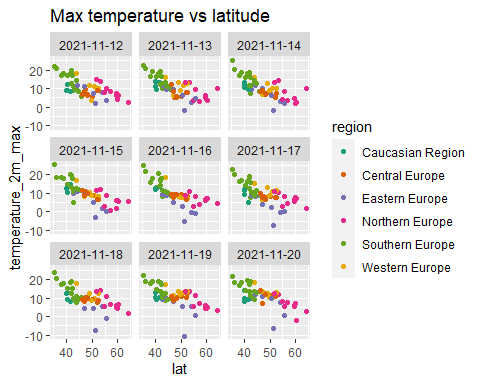
The following table resumes the 24 numeric variables of the final dataframe.

|  |  |  |
| --- | --- | --- |
| Selected variables | Description | Unit |
| pressure\_msl | Atmospheric air pressure reduced to sea level | hPa |
| temperature\_2m | Air temperature at 2 meters above ground | °C (°F) |
| snow\_height | Snow height on the ground | meters |
| direct\_radiation | Direct solar radiation as average of the preceding hour on the horizontal plane and the normal plane (perpendicular to the sun) | W/m² |
| weathercode | The most severe weather condition on a given day | WMO code |
| winddirection\_10m | Wind direction at 10, 80, 120 or 180 meters above ground | ° |
| precipitation | Total precipitation (rain, showers, snow) sum of the preceding hour | mm (inch) |
| relativehumidity\_2m | Relative humidity at 2 meters above ground | % |
| windspeed\_10m | Wind speed on 10 meters is the standard level. | km/h (mph, m/s, knots) |
| apparent\_temperature | Apparent temperature is the perceived feels-like tempertature combinding wind chill factor, realtive humidity and solar radition | °C (°F) |
| cloudcover | Total cloud cover as an area fraction | % |
| temperature\_2m\_max, temperature\_2m\_min | Maximum and minimum daily air temperature at 2 meters above ground | °C (°F) |
| apparent\_temperature\_max, apparent\_temperature\_min | Maximum and minimum dailt apparent temperature | °C (°F) |
| sunset, sunset | Sun rise and set times | iso8601 |
| precipitation\_sum | Sum of daily precipitation | mm |
| precipitation\_hours | The number of hours with rain |  |
| windspeed\_10m\_max, windgusts\_10m\_max | Maximum wind speed and gusts on a day | hours |
| winddirection\_10m\_dominant | Dominant wind direction | ° |
| shortwave\_radiation\_sum | The sum of solar radiation on a given day in Mega Joules | MJ/m² |
| lat | latitude |  |
| long | longitude |  |

### Data visualization

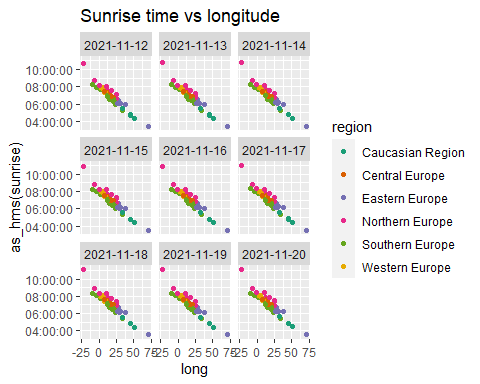
We firstly chose to compare the most relevant variables in order to visualize the relations among them.

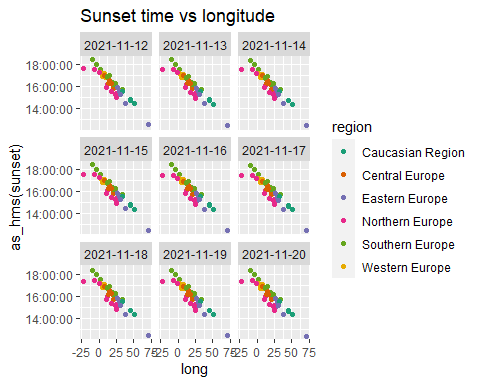
The first correlation we tried to explore was the one between the max temperature and the latitude of the city.



As we can see on the plots, there is inverse correlation between the two variables.

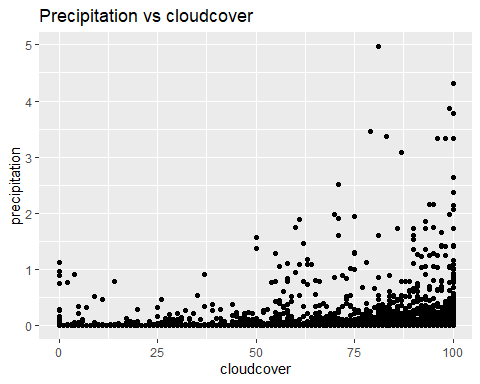
We proceeded with the same kind of analysis on sunset and sunrise time versus the cities’ longitude.





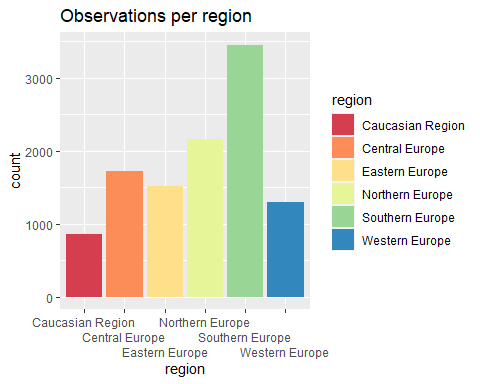
It is evident that there is an inverse correlation between the sunset/sunrise time and the longitude too.

Then we took precipitation and cloudcover as variables and plotted them.



We see there could be direct correlation between the two. The higher the percentages of cloudcover, the more precipitations there will be.

Lastly, we decided to visualize the number of observations per region.



We can immediately see that the majority of observations is related to Southern Europe. Conversely, the observations on the remaining regions were more or less equally distributed, with the Caucasian region registering the lowest number.

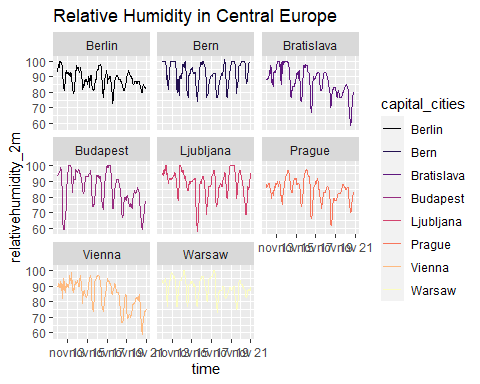
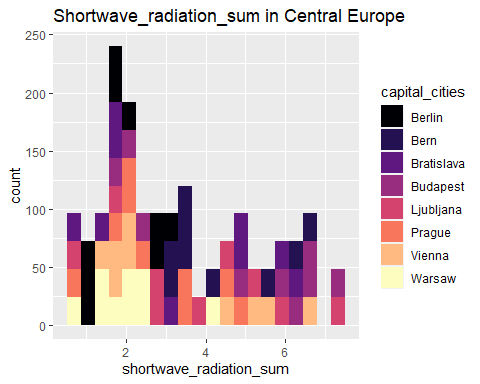
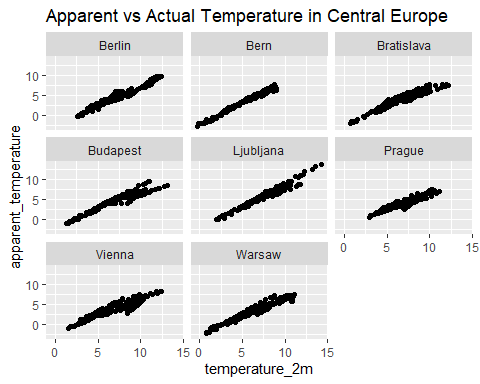
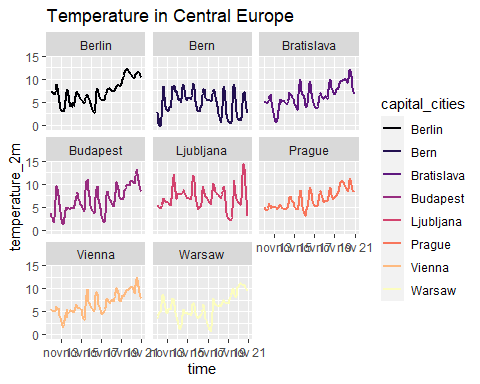
Considering that the number of observations is more manageable when dividing it by regions, we opted for further visualization on these smaller datasets.

### Visualization by regions

We created new dataframes through the filter() function. The four graphs we built represented:1) a time series of temperature; 2)A scatterplot of actual vs apparent temperature; 3) A histogram of the radiation sum; 4) a times series of relative humidity.

All these graphs visualize the difference among the region’s capital cities through the usage of color. For the second dataframe, given that the number of cities was equal to 16, we created a customized palette in order to color it.

### Central Europe



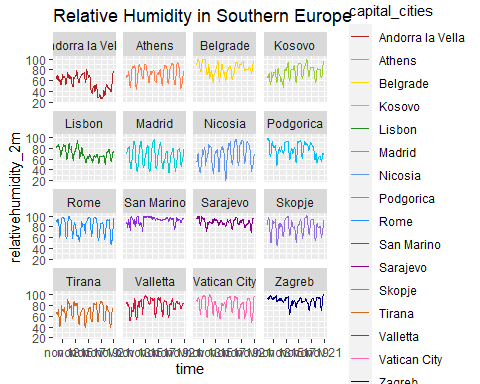
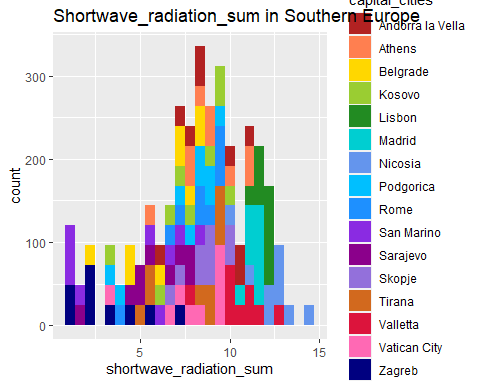
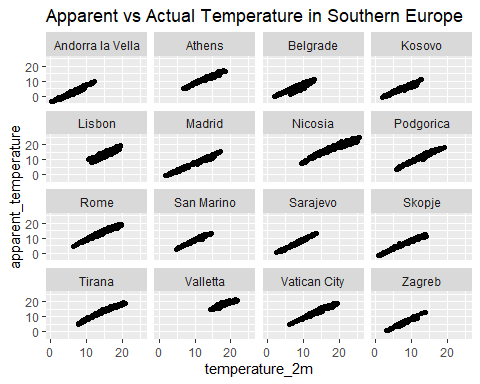
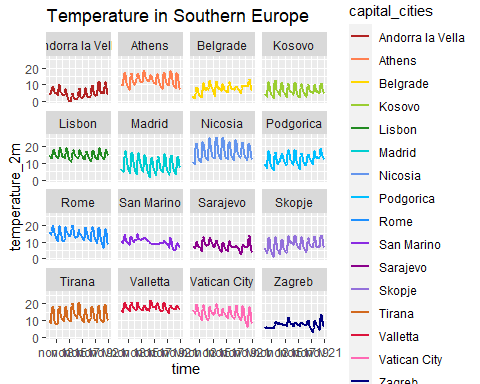
The first graph shows that Ljubljana was the city where the highest temperatures were recorded. The others registered similar trends.

The second graph showed that the apparent temperature was quite similar to the actual one.

The third one shows that the city with the largest amount of daily radiations was Budapest. On the opposite side, Warsaw registered the lowest number of radiations.

The last graph depicts Bern as the city with the highest and most consistent level of humidity.

### Southern Europe



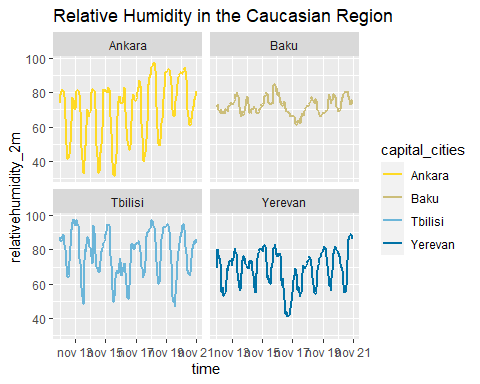
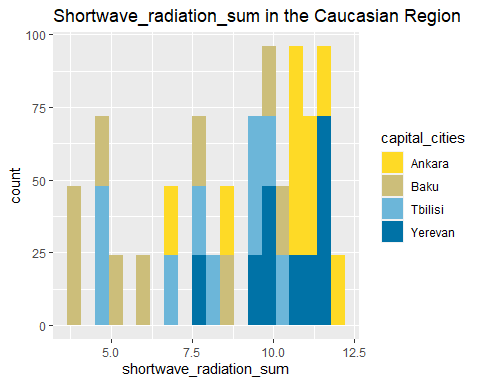
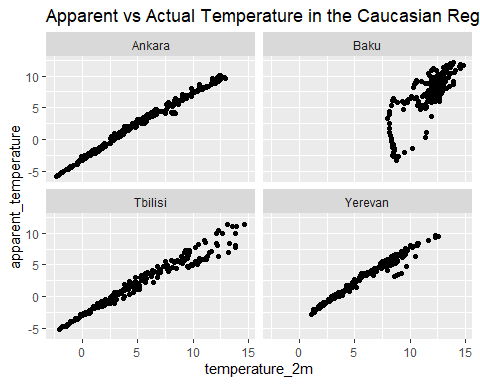
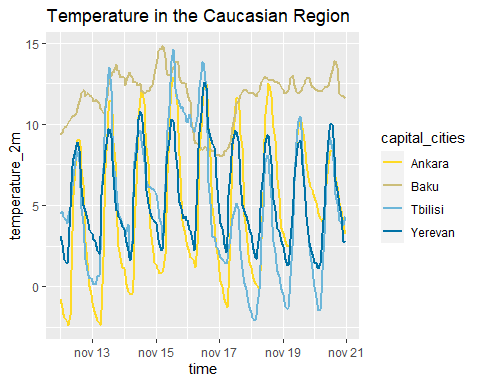
The first graph shows that Nicosia and La Valletta were the cities where the highest temperatures were recorded. Moreover, Zagreb and San Marino were the ones with the lowest temperature variation between day and night.

The second graph shows that the apparent temperature was quite similar to the actual one.

The third one shows that the cities with the largest amount of daily radiations were Lisbon, Nicosia and La Valletta. On the opposite side, San Marino, Zagreb and Sarajevo registered the lowest number of radiations.

The last graph depicts San Marino as the city with the highest and most consistent level of humidity, together with Sarajevo.

### The Caucasian Region



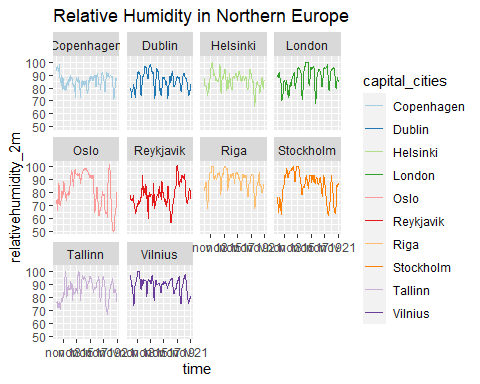
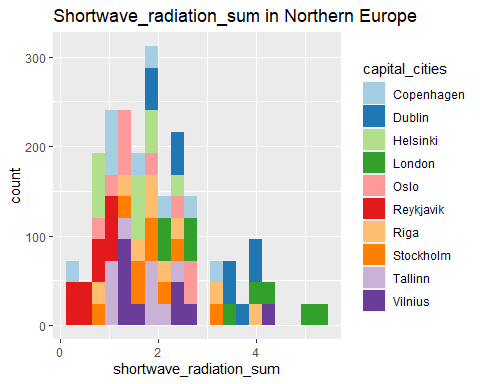
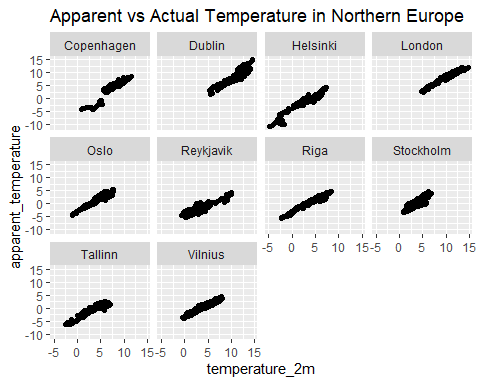
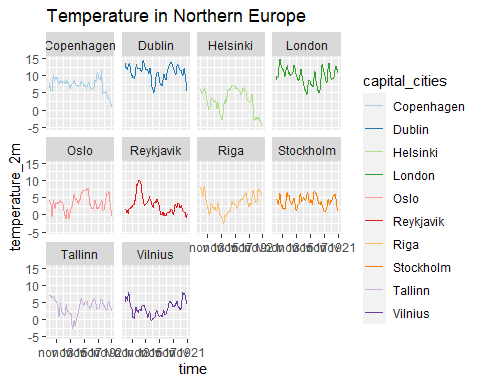
The first graph shows that Baku and Tbilisi were the cities here the highest temperatures were recorded, but the temperature in the first one was more consistent, registering the lowest temperature variation between day and night.

The second graph showed that the apparent temperature was quite similar to the actual one everywhere but in Baku, where the apparent temperature was fount to be considerably lower than the actual one several times.

The third one shows that the cities with the largest amount of daily radiations were Ankara and Yerevan.On the opposite side, Baku registered the lowest number of radiations.

The last graph depicts Tbilisi as the city with the highest level of humidity, while Baku shows the most consistent values.

### Northern Europe



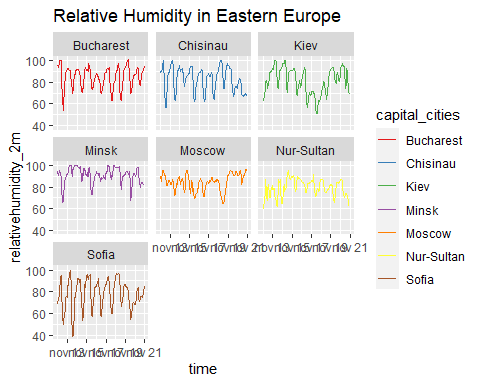
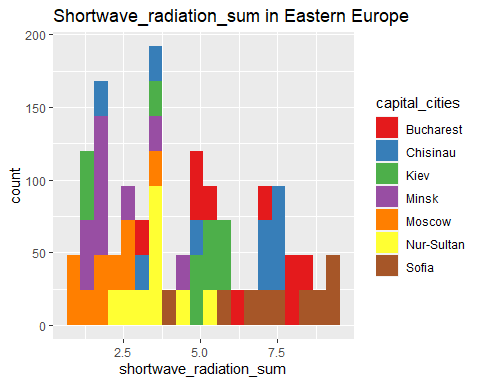
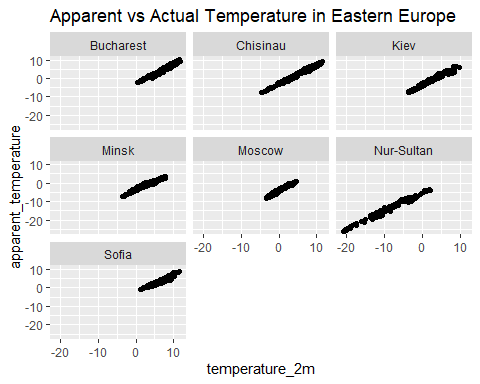
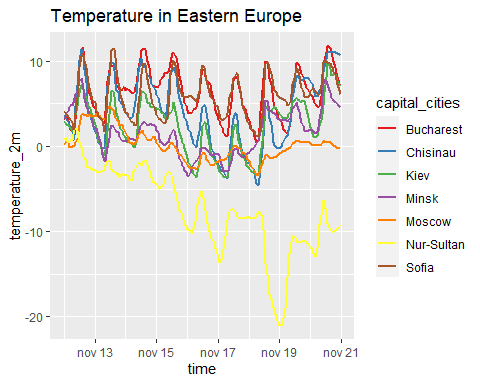
The first graph shows that Dublin and London were the cities where the highest temperatures were recorded.Copenhagen registered the lowest temperature variation between day and night. Helsinki and Tallin recorded the lowest temperature.

The second graph shows that the apparent temperature was quite similar to the actual one, but in Copenhagen.Generally speaking, though, the lines appear to be thicker than the ones of the other regions, which means that the apparent temperature swings more in relation to a single value of actual temperature.

The third one shows that the city with the largest amount of daily radiations was London.On the opposite side Reykjavik registered the lowest number of radiations.

The last graph depicts Copenhagen and Reykjavik as the cities with the lowest level of humidity.

### Eastern Europe



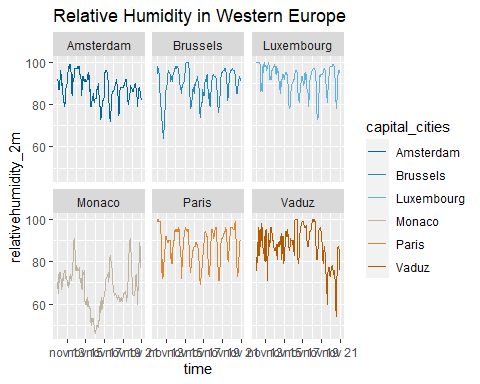
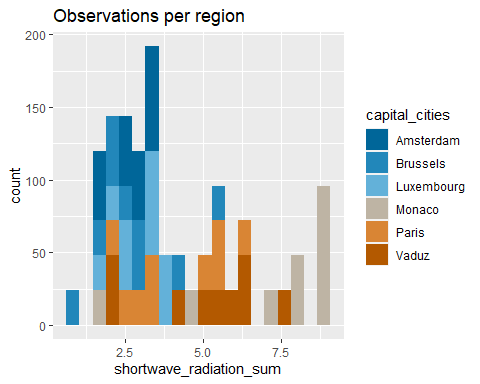
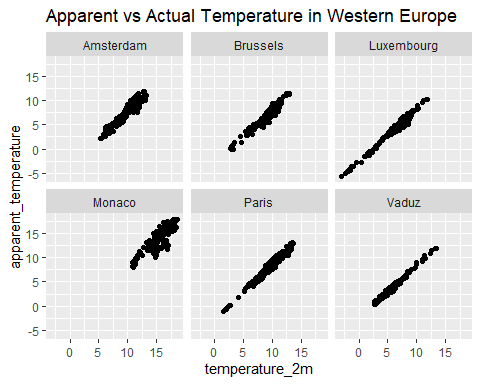
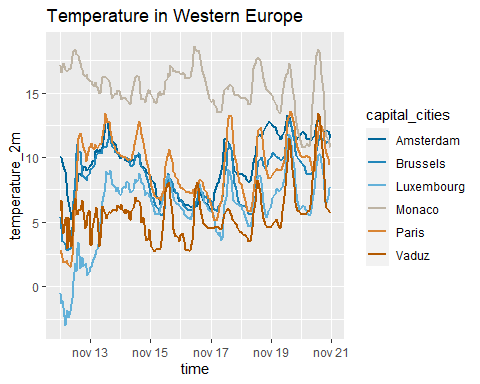
The first graph shows that Bucharest and Sofia were the cities where the highest temperatures were recorded. Moreover, Moscow was the one with the lowest temperature variation between day and night. Nur-Sultan registered the lowest temperature.

The second graph showed that the apparent temperature was quite similar to the actual one.

The third one shows that the cities with the largest amount of daily radiations were Sofia and Bucharest, especially the first of the two.On the opposite side Moscow and Minsk registered the lowest number of radiations.

The last graph depicts Minsk as the city with the highest level of humidity.

### Western Europe



The first graph shows that Monaco was the city where the highest temperatures were recorded and with the lowest temperature variation between day and night.

The second graph showed that the apparent temperature was quite similar to the actual one. Generally speaking, though, the lines appear to be thicker than the ones of other regions, which means that the apparent temperature swings more in relation to a single value of actual temperature, especially in Monaco.

The third one shows that the city with the largest amount of daily radiations was Monaco.On the opposite side, Brussels registered the lowest number of radiations.

The last graph depicts Luxembourg as the city with the highest and most consistent level of humidity.The lowest level of humidity was registered by Monaco.