QIANQIAN MENG 23212722

February 16, 2024

0.1 Exploratory data analysis of the Irish weather

For this assignment you will perform an exploratory data analysis (EDA) of historic weather data from Met Eireann, Ireland's main meteorological service.

[1]:
```
# Load in necessary packages
import numpy as np
import pandas as pd
from pandas import DataFrame, Series


import matplotlib.pyplot as plt
import seaborn as sns
```

1. Load in the weather.csv dataset into Python as a pandas DataFrame. Describe the data. How many years of recordings are included? What is the temporal resolution of the data? Which weather measurements are reported? (8 marks)

Load in the weather.csv dataset into Python as a pandas DataFrame:

[2]:
```
path = '/Users/qianqianmeng/Desktop/Python Study/MIDTERM/' #Define my file path␣
 ↪where the weather.csv dataset is located.
skip_rows = list(range(0, 11))#list contains first 11 rows with not data, I␣
 ↪want to skip
weather = pd.read_csv(path + 'weather_1819.csv', skiprows=skip_rows)#Read the␣
 ↪the weather.csv file into a pandas DataFrame, skip the first 11 rows
```

Describe the data:

The weather.csv dataset is stored in a pandas DataFrame with a RangeIndex that extends from 0 to 2920, and it consists of a total of 10 columns. The variables included are 'day', 'month', 'year', 'station', 'maxtp', 'mintp', 'rain', 'wdsp', 'hg', and 'sun'. The data types of these columns are as follows: six columns are of the float64 data type, two are int64, and two are objects. The entire DataFrame occupies approximately 228.3 KB of memory. If the 'Non-Null Count' of any column is less than 2920, it indicates that there are some missing values in that column."

[3]:
```
weather.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2920 entries, 0 to 2919
Data columns (total 10 columns):
```

```
 #   Column   Non-Null Count  Dtype
---  ------   --------------  -----
 0   day      2920 non-null   int64
 1   month    2920 non-null   object
 2   year     2920 non-null   int64
 3   station  2920 non-null   object
 4   maxtp    2902 non-null   float64
 5   mintp    2902 non-null   float64
 6   rain     2897 non-null   float64
 7   wdsp     2915 non-null   float64
 8   hg       2904 non-null   float64
 9   sun      2913 non-null   float64
dtypes: float64(6), int64(2), object(2)
memory usage: 228.3+ KB
```

How many years of recordings are included?

two years of recordings are included, 2018 and 2019

[4]:
```
weather.year.unique()
```

[4]:
```
array([2018, 2019])
```

What is the temporal resolution of the data?

The '0 days' value indicates that there are multiple records within each day; however, since no finer time units than 'year', 'month', and 'day' are recorded, the temporal resolution of the data is daily.

[5]:
```
weather['date'] = weather['day'].astype(str) + '-' + weather['month'] + '-' +␣
 ↪weather['year'].astype(str)#from last section, day and year is stored as␣
 ↪integer, convert them to string
weather['date'] = pd.to_datetime(weather['date'], format='%d-%b-%Y')  #convert␣
 ↪the 'date' string into a pandas datetime object
weather.date.value_counts() #it shows there are four records in each day
weather = weather.sort_values('date') #sort the dataset in ascending order by␣
 ↪the 'date' column
print(weather.date.value_counts()) #there are four records in each day
weather['date_diff'] = weather['date'].diff() #get the difference between␣
 ↪consecutive dates
print(weather['date_diff'].value_counts()) #to see unique differences
```

```
date
2018-01-01    4
2019-04-25    4
2019-04-27    4
2019-04-28    4
2019-04-29    4
             ..
2018-09-03    4
2018-09-04    4
```

```
2018-09-05    4
2018-09-06    4
2019-12-31    4
Name: count, Length: 730, dtype: int64
date_diff
0 days    2190
1 days     729
Name: count, dtype: int64
```

Which weather measurements are reported?

'maxtp', 'mintp', 'rain', 'wdsp','hg', 'sun'

[6]:
```
weather.columns #get the column names
```

[6]:
```
Index(['day', 'month', 'year', 'station', 'maxtp', 'mintp', 'rain', 'wdsp',
       'hg', 'sun', 'date', 'date_diff'],
      dtype='object')
```

2. Determine how many missing values there are in each column of the dataset. Can you think of a reason why these values are missing? Discuss different strategies for filling the missing values, highlighting the advantages and disadvantages of each strategy, in the context of this dataset. (8 marks)

   Note: You do not need to implement any of your suggested strategies.

There are missing values in the columns 'maxtp', 'mintp', 'rain', 'wdsp', 'hg', and 'sun', with the corresponding numbers of missing entries being 18, 18, 23, 5, 16, and 7, respectively. These values may be missing due to the removal of outliers or perceived errors during the data manipulation process.

Discuss different strategies for filling the missing values, highlighting the advantages and disadvantages of each strategy, in the context of this dataset: 1. Replacing with an arbitrary value. Pros: This allows for an educated guess regarding the missing values. By using a distinct arbitrary value, it is easy to track where data has been imputed. Cons: The arbitrary value may not accurately represent the actual weather conditions and could distort the data analysis. 2. Fill with mean value. Pros: Easy to implement in code and can be a quick way to address missing data. Cons: Since there are outliers in the measurements, then using the mean to fill missing values might not be appropriate. 3. Forward fill and backward fill. Pros: Since the data is recorded daily over two years, utilizing forward fill or backward fill can preserve the seasonal trends in weather patterns. Cons: Given the variability of the Irish weather, these methods might not capture sudden, short-term fluctuations accurately.

[7]:
```
weather.isnull().sum()#check null first, and use .sum() to count
```

[7]:
```
day           0
month         0
year          0
station       0
maxtp        18
mintp        18
```

```
rain         23
wdsp          5
hg           16
sun           7
date          0
date_diff     1
dtype: int64
```

3. Write code to answer the following questions: (15 marks)
    1. At what station and on what date was the highest wind speed recorded?
    2. At what station and on what date was the highest maximum air temperature recorded?
    3. At what station and on what date was the largest amount of rain recorded?

A: The highest wind speed of 28.5 was recorded at Dublin Airport on 2018-03-02.

[8]:
```
weather.sort_values(by='wdsp',ascending=False) # Sort the DataFrame by the␣
↪'wdsp' column in descending order.
```

[8]:
```
      day month  year          station  maxtp  mintp  rain  wdsp    hg  sun  \
60      2   mar  2018   Dublin Airport   -0.2   -1.2   5.6  28.5  50.0  0.0
59      1   mar  2018   Dublin Airport   -0.5   -5.1  12.2  26.0  47.0  0.0
1436    8   dec  2019  Shannon Airport   10.1    5.3   7.8  25.2  58.0  0.6
2593    8   feb  2019    Knock Airport    9.0    4.1  13.2  24.7  60.0  1.5
481    27   apr  2019   Dublin Airport   11.7    5.0   8.0  24.0  46.0  6.9
…   …   …   …              …    …    …   …   …   …  …
134    15   may  2018   Dublin Airport   17.4    4.5   NaN   NaN  25.0  3.4
2387   17   jul  2018    Knock Airport   17.5    9.5   4.9   NaN  19.0  1.8
1755   23   oct  2018     Cork Airport   15.3    6.5   0.0   NaN  29.0  4.0
300    28   oct  2018   Dublin Airport    9.8   -2.9   0.0   NaN  22.0  5.6
1811   18   dec  2018     Cork Airport   11.0    3.7  11.0   NaN  52.0  3.0

           date date_diff
60   2018-03-02    0 days
59   2018-03-01    0 days
1436 2019-12-08    0 days
2593 2019-02-08    0 days
481  2019-04-27    1 days
…         …       …
134  2018-05-15    0 days
2387 2018-07-17    0 days
1755 2018-10-23    0 days
300  2018-10-28    0 days
1811 2018-12-18    0 days

[2920 rows x 12 columns]
```

B: The highest maximum air temperature of 32.0 was recorded at Shannon Airport on 2018-06-28.

[9]:
```
weather.sort_values(by='maxtp',ascending=False) # Sort the DataFrame by the␣
↪'maxtp' column
```

[9]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	\
908	28	jun	2018	Shannon Airport	32.0	12.4	0.0	4.9	20.0	15.6	
909	29	jun	2018	Shannon Airport	31.1	15.5	0.0	5.1	15.0	15.5	
907	27	jun	2018	Shannon Airport	30.8	15.4	0.0	5.2	16.0	14.8	
906	26	jun	2018	Shannon Airport	29.3	16.0	0.0	6.4	18.0	13.4	
2368	28	jun	2018	Knock Airport	28.9	16.6	0.0	6.2	15.0	15.2	
…	…	…	…	…	…	…	…	…	…	…	…
2009	4	jul	2019	Cork Airport	NaN	NaN	0.0	6.0	17.0	14.7	
1286	11	jul	2019	Shannon Airport	NaN	NaN	1.0	8.2	21.0	4.7	
1300	25	jul	2019	Shannon Airport	NaN	NaN	4.2	15.6	31.0	3.1	
2031	26	jul	2019	Cork Airport	NaN	NaN	2.1	9.4	22.0	10.1	
2077	10	sep	2019	Cork Airport	NaN	NaN	0.1	8.9	26.0	2.7	

	date	date_diff
908	2018-06-28	1 days
909	2018-06-29	0 days
907	2018-06-27	0 days
906	2018-06-26	1 days
2368	2018-06-28	0 days
…	…	…
2009	2019-07-04	1 days
1286	2019-07-11	0 days
1300	2019-07-25	0 days
2031	2019-07-26	0 days
2077	2019-09-10	0 days

[2920 rows x 12 columns]

C: The largest amount of rain of 54.6 was recorded at Cork Airport on 2019-04-15.

[10]:
```
weather.sort_values(by='rain',ascending=False) # Sort the DataFrame by the␣
↪'rain' column
```

[10]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	\
1929	15	apr	2019	Cork Airport	8.3	5.2	54.6	19.6	45.0	0.0	
2111	14	oct	2019	Cork Airport	13.1	9.8	41.2	9.3	26.0	0.0	
1998	23	jun	2019	Cork Airport	13.9	12.6	39.1	12.5	27.0	0.0	
2148	20	nov	2019	Cork Airport	9.1	6.9	33.7	11.1	25.0	0.0	
1381	14	oct	2019	Shannon Airport	13.3	7.6	33.4	5.7	17.0	0.0	
…	…	…	…	…	…	…	…	…	…	…	…
2846	19	oct	2019	Knock Airport	11.8	6.5	NaN	7.6	25.0	3.1	
670	2	nov	2019	Dublin Airport	10.6	2.8	NaN	6.1	20.0	1.1	
2172	14	dec	2019	Cork Airport	6.1	0.5	NaN	11.0	37.0	1.8	
2184	26	dec	2019	Cork Airport	11.6	7.6	NaN	13.0	37.0	0.6	
2918	30	dec	2019	Knock Airport	9.6	3.4	NaN	8.4	21.0	0.0	

```
           date date_diff
1929 2019-04-15    0 days
2111 2019-10-14    1 days
1998 2019-06-23    0 days
2148 2019-11-20    1 days
1381 2019-10-14    0 days
…          …       …
2846 2019-10-19    0 days
670  2019-11-02    0 days
2172 2019-12-14    0 days
2184 2019-12-26    0 days
2918 2019-12-30    0 days

[2920 rows x 12 columns]
```

4. Create a numerical summary (mean, standard deviation, minimum, maximum, etc.) for each of the weather measurements. Discuss and interpret your results. (8 marks)

maxtp: mean: The average maximum temperature across all the days is approximately 13.28°C. std: The standard deviation is 5.15, which indicates that the daily maximum temperatures typically vary by ±5.15°C from the mean. min and max: The temperatures ranged from -1.8°C to 32°C. IQR: Half of the days had a maximum temperature between 9.5°C and 17.1°C.

mintp: mean: The average minimum temperature across all the days is approximately 6.43°C. std: The standard deviation is 4.37, indicating that the daily minimum temperatures typically vary by ±4.37°C from the mean. min and max: The temperatures ranged from -7°C to 18.9°C. IQR: Half of the days had a minimum temperature between 3.1°C and 9.6°C.

rain: mean: On average, there was about 3.06 mm of rainfall. min and max: Rainfall measurements ranged from 0 mm (no rain) to a high of 54.6 mm. IQR: Half of the days had rainfall amounts between 0 mm and 4 mm.

wdsp: mean: The average wind speed was around 9.48 knots. min and max: Wind speeds ranged from 2.3 knots to 28.5 knots. IQR: Half of the days experienced wind speeds between 6.5 knots and 11.8 knots.

hg: mean: The average of the highest gusts was around 25.44 knots. min and max: Highest gusts ranged from 7 knots to a very high of 84 knots. IQR: Half of the days had highest gusts between 19 and 30 knots.

sun: mean: On average, there were about 3.78 hours of sunshine per day. min and max: The duration of sunshine ranged from 0 hours (no sunshine) to 15.9 hours. IQR: Half of the days recorded sunshine durations between 0.3 hours and 6.3 hours.

[11]:
```
drop_columns = ['day', 'year', 'date_diff'] #since .describe() provides a␣
 ↪summary of every numerical column, drop the irrelevant ones for this␣
 ↪question.
weather_measurements = weather.drop(columns=drop_columns) # Drop the specified␣
 ↪columns from the 'weather' DataFrame and create a new DataFrame.
```

```
weather_measurements.describe()
```

[11]:

	maxtp	mintp	rain	wdsp	hg	\
count	2902.000000	2902.000000	2897.000000	2915.000000	2904.000000	
mean	13.283150	6.432977	3.063583	9.481475	25.443871	
min	-1.800000	-7.000000	0.000000	2.300000	7.000000	
25%	9.500000	3.100000	0.000000	6.500000	19.000000	
50%	12.800000	6.400000	0.700000	8.900000	24.000000	
75%	17.100000	9.600000	4.000000	11.800000	30.000000	
max	32.000000	18.900000	54.600000	28.500000	84.000000	
std	5.146289	4.368755	5.053881	3.820605	9.278313	

	sun	date
count	2913.000000	2920
mean	3.783797	2018-12-31 12:00:00
min	0.000000	2018-01-01 00:00:00
25%	0.300000	2018-07-02 00:00:00
50%	2.600000	2018-12-31 12:00:00
75%	6.300000	2019-07-02 00:00:00
max	15.900000	2019-12-31 00:00:00
std	3.850012	NaN

5. Create a graphical summary for each of the weather measurements. Discuss your plots in relation to the summary statistics found in question 4. (10 marks)

For 'maxtp' and 'mintp', their histograms show roughly normal distributions with mean-centered, confirmed by the summary statistics as 13.28°C and 6.43°C, respectively. Approximately 95% of the data falls within two standard deviations. 'maxtp' temperatures ranged from -1.8°C to 32°C, while 'mintp' ranged from -7°C to 18.9°C.

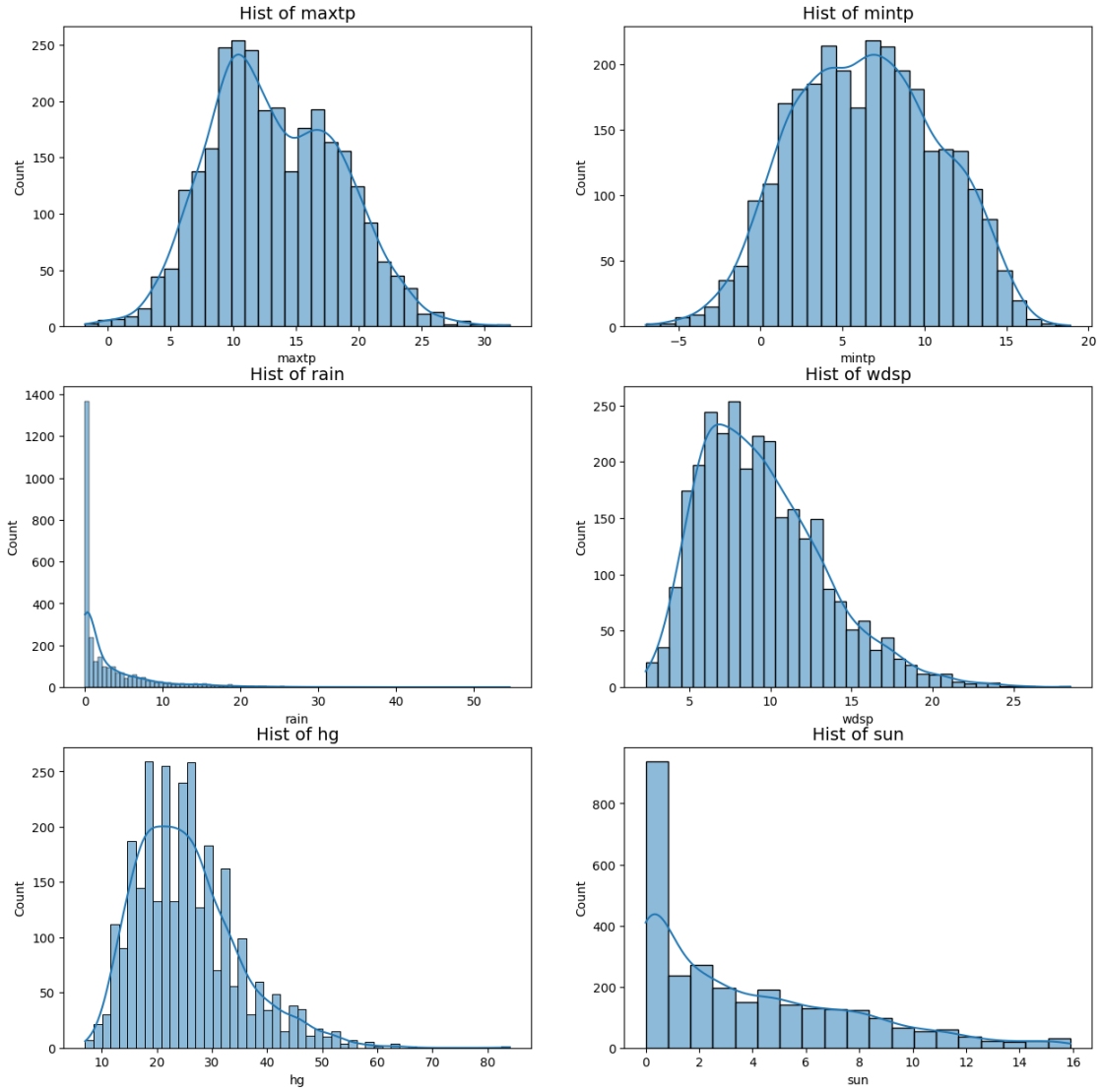
The remaining histograms are right-skewed. Among them, 'hg' has the widest data spread, as it has the largest range, confirmed by its standard deviation of 9.28 in the summary statistics. The 'rain' histogram is heavily right-skewed, with the mean at approximately 3.06 mm, which may not be very representative due to the skew. Its histogram suggests that the most common rainfall is between 0-5 mm, aligning with the IQR showing that half of the days had rainfall between 0 mm and 4 mm. The maximum recorded rainfall is 54.6 mm, and the minimum is no rain, at 0 mm. The 'sun' histogram is less skewed compared to 'rain'; on most days, there isn't much sunshine, as the average sunshine duration per day is about 3.78 hours, ranging from 0 to a maximum of 15.9 hours.

The histograms for 'wdsp' and 'hg' are not as heavily skewed to the right. Their corresponding averages are 9.48 knots for wind speed and 25.44 knots, respectively. Wind speeds ranged from 2.3 to 28.5 knots, with the highest gusts recorded between 7 knots and a very strong 84 knots. Half of the days experienced wind speeds between 6.5 and 11.8 knots and highest gusts between 19 and 30 knots.

[12]:
```
plt.figure(figsize=(15,15))

plt.subplot(3,2,1) # Create a subplot in a 3x2 grid at position 1
```

```
sns.histplot(weather_measurements['maxtp'], kde=True) #add kde line
plt.xlabel('maxtp')
plt.title('Hist of maxtp',fontsize=14)


plt.subplot(3,2,2) #position 2
sns.histplot(weather_measurements['mintp'], kde=True)
plt.xlabel('mintp')
plt.title('Hist of mintp',fontsize=14)

plt.subplot(3,2,3) #position 3
sns.histplot(weather_measurements['rain'], kde=True)
plt.xlabel('rain')
plt.title('Hist of rain',fontsize=14)

plt.subplot(3,2,4) #position 4
sns.histplot(weather_measurements['wdsp'], kde=True)
plt.xlabel('wdsp')
plt.title('Hist of wdsp',fontsize=14)

plt.subplot(3,2,5) #position 5
sns.histplot(weather_measurements['hg'], kde=True)
plt.xlabel('hg')
plt.title('Hist of hg',fontsize=14)

plt.subplot(3,2,6) #position 6
sns.histplot(weather_measurements['sun'], kde=True)
plt.xlabel('sun')
plt.title('Hist of sun',fontsize=14)
```

[12]: Text(0.5, 1.0, 'Hist of sun')

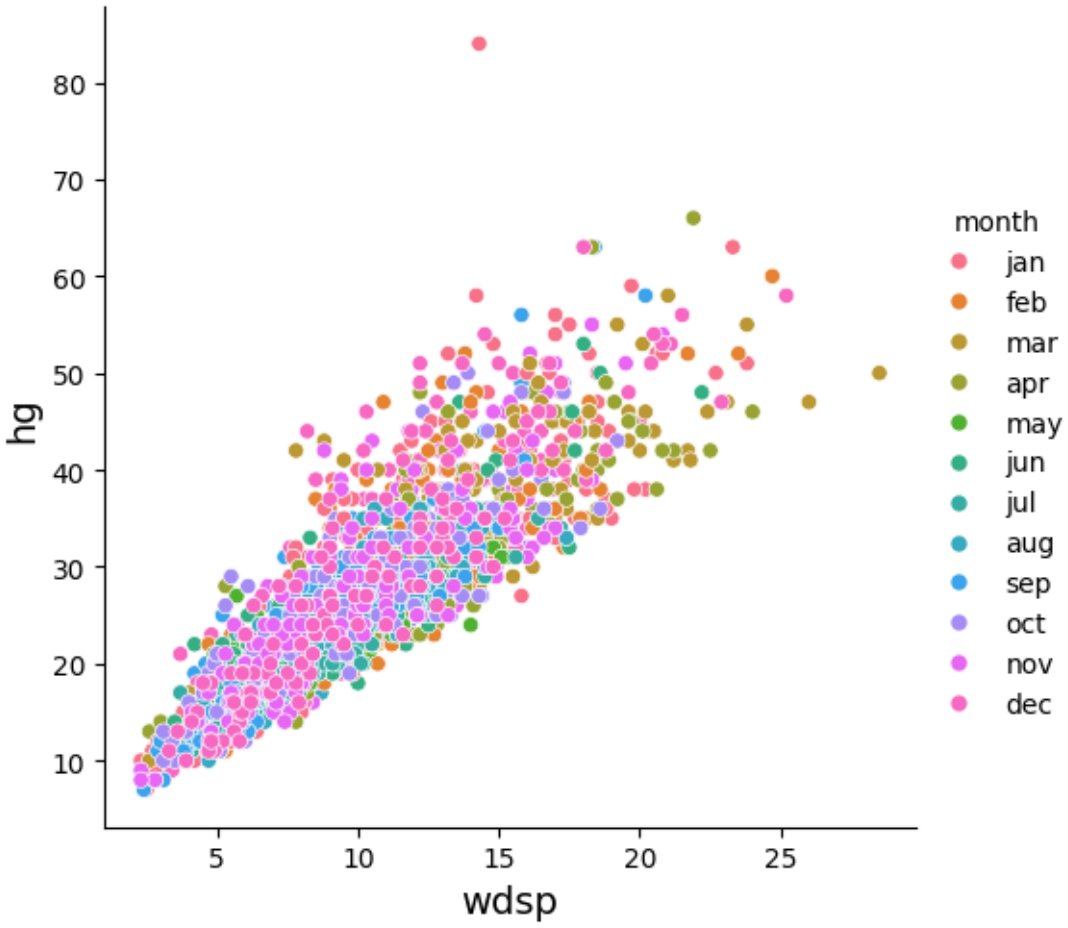


6. Produce a scatter plot of the mean wind speed versus the highest gust and colour your points based on month. Interpret your plot. (8 marks)

Based on my analysis of the plot, there appears to be a positive correlation between the mean wind speed and the highest gusts. Additionally, the data suggests that, throughout the year, a higher mean wind speed generally coincides with more intense gusts. However, the plot does not reveal any clear relationship between the highest gusts or mean wind speed and the months of the year.

[13]:
```
sns.relplot(data=weather_measurements, x='wdsp', y="hg", hue='month') #hue␣
  ↪argument here to colour points based on month
plt.xlabel('wdsp',fontsize=14)
plt.ylabel('hg',fontsize=14)
```

[13]:
```
Text(38.95087152777777, 0.5, 'hg')
```



7. Compute the daily temperature range, and add this as an additional variable to your DataFrame. Print out the last 10 rows of your DataFrame to show that the column has been added correctly. (5 marks)

[14]:
```
weather['daily_temperature_range'] = weather['maxtp'] - weather['mintp'] #add␣
↪new column daily_temperature_range to weather data frame
weather.tail()
```

[14]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	\
728	30	dec	2019	Dublin Airport	11.0	2.9	0.2	6.9	20.0	0.5	
1459	31	dec	2019	Shannon Airport	8.6	2.7	0.1	5.6	16.0	0.0	
729	31	dec	2019	Dublin Airport	9.0	4.1	0.0	6.2	16.0	0.2	
2189	31	dec	2019	Cork Airport	8.7	5.5	0.4	7.5	19.0	0.0	
2919	31	dec	2019	Knock Airport	6.0	1.3	0.0	7.1	18.0	2.1	

	date	date_diff	daily_temperature_range
728	2019-12-30	0 days	8.1
1459	2019-12-31	1 days	5.9

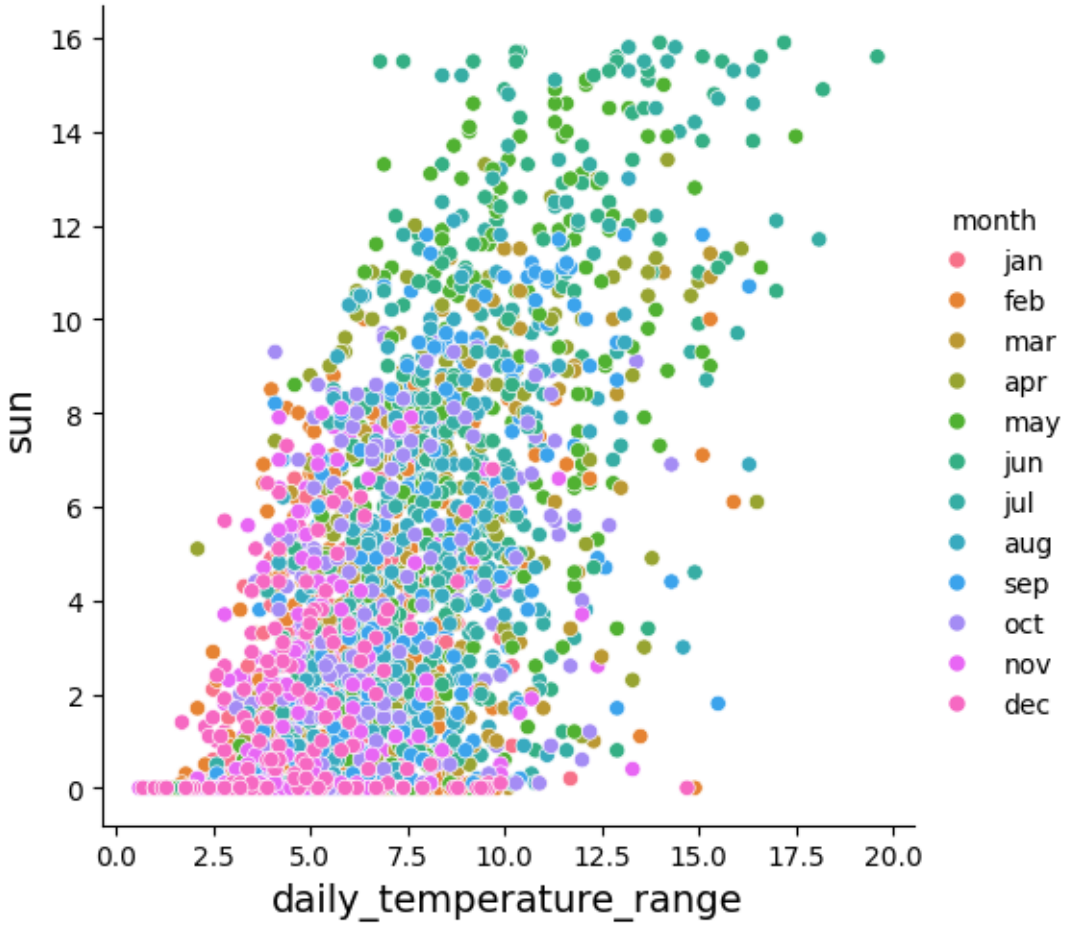
```
729  2019-12-31    0 days                    4.9
2189 2019-12-31    0 days                    3.2
2919 2019-12-31    0 days                    4.7
```

8. Plot the daily temperature range versus the hours of sunlight per day, colouring the points based on month. Interpret your plot. (8 marks)

In summer (May, June, July, August), when there is more sunlight, the daily temperature range tends to be wider. However, according to my plot, there appears to be no linear relationship between the daily temperature range and the number of sunlight hours per day. This suggests that a wider daily temperature range does not necessarily correlate with longer periods of sunlight.

[15]:
```
sns.relplot(data=weather, x='daily_temperature_range', y='sun', hue='month')
plt.xlabel('daily_temperature_range',fontsize=14)
plt.ylabel('sun',fontsize=14)
```

[15]:
```
Text(38.97096527777777, 0.5, 'sun')
```



9. Perform a comparative analysis of the weather at Dublin Airport, Shannon Airport and Cork

Airport. (20 marks)

For full marks on this question you should create numerical and graphical summaries of the weather measurements at each weather station and discuss how the weather differs (or is similar) across these locations.

Shannon Airport has higher maximum air temperatures(14.16) and higher minimum air temperatures(7.26) than the other two airports. It has larger medians for both maximum and minimum temperatures, which suggests a generally warmer climate.

Cork Airport experiences higher mean rainfall at 3.47 mm and also records the highest maximum rainfall at 54.6 mm. Its boxplot shows a larger median, suggesting it has more rain than the other two airports. On the other hand, Dublin Airport exhibits the least variability in rainfall, with the smallest mean of 2.15 mm and the smallest median, indicating it is the driest among the three.

The average wind speed and highest gust measurements are quite similar across all three airports. However, Cork Airport has the highest mean wind speed and the highest recorded gust, as well as a slightly larger median, making it the windiest airport.

Cork Airport also enjoys a slightly longer mean sunshine duration at 4.23 hours, with a higher median and greater variability, indicating more average sunshine than another two airports.

[16]:
```
weather_measurements = weather.drop('daily_temperature_range', axis=1) #drop␣
 ↪the column daily_temperature_range
# data for each weather station
dublin = weather_measurements[weather_measurements['station'] == 'Dublin␣
 ↪Airport']
shannon = weather_measurements[weather_measurements['station'] == 'Shannon␣
 ↪Airport']
cork = weather_measurements[weather_measurements['station'] == 'Cork Airport']

# numerical summaries for each weather station
dublin_summary = dublin.describe()
shannon_summary = shannon.describe()
cork_summary = cork.describe()

print(f"Numerical summary for Dublin Airport: ")
print(dublin_summary)

print(f"Numerical summary for Shannon Airport: ")
print(shannon_summary)

print(f"Numerical summary for Cork Airport: ")
print(cork_summary)
```

```
Numerical summary for Dublin Airport:
              day         year       maxtp       mintp        rain  \
count  730.000000   730.000000  730.000000  730.000000  726.000000
mean    15.720548  2018.500000   13.592603    5.724247    2.147383
min      1.000000  2018.000000   -0.500000   -5.800000    0.000000
```

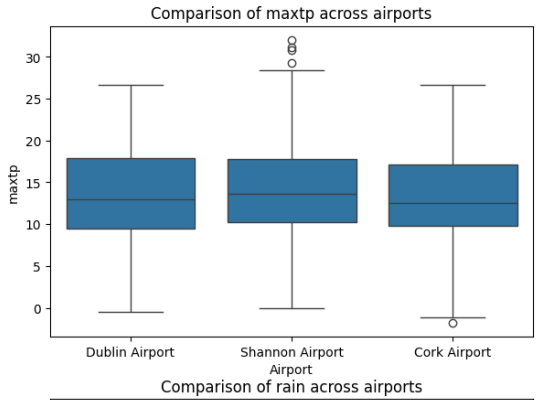
```
25%      8.000000  2018.000000    9.500000    2.400000    0.000000
50%     16.000000  2018.500000   13.000000    5.600000    0.200000
75%     23.000000  2019.000000   17.875000    8.900000    2.300000
max     31.000000  2019.000000   26.700000   17.800000   24.200000
std      8.802278     0.500343    5.265505    4.488748    4.120571

             wdsp          hg         sun                 date  \
count  728.000000  728.000000  729.000000                  730
mean     9.542720   24.943681    4.046091  2018-12-31 12:00:00
min      3.000000    9.000000    0.000000  2018-01-01 00:00:00
25%      6.700000   19.000000    0.600000  2018-07-02 06:00:00
50%      8.800000   24.000000    2.900000  2018-12-31 12:00:00
75%     11.500000   30.000000    6.600000  2019-07-01 18:00:00
max     28.500000   56.000000   15.900000  2019-12-31 00:00:00
std      3.829729    8.438033    3.952654                  NaN

                      date_diff
count                       729
mean   0 days 05:51:36.296296296
min              0 days 00:00:00
25%              0 days 00:00:00
50%              0 days 00:00:00
75%              0 days 00:00:00
max              1 days 00:00:00
std    0 days 10:19:02.410817267
Numerical summary for Shannon Airport:
              day         year       maxtp       mintp        rain  \
count  730.000000   730.000000  723.000000  723.000000  721.000000
mean    15.720548  2018.500000   14.160028    7.267082    2.844383
min      1.000000  2018.000000    0.000000   -5.000000    0.000000
25%      8.000000  2018.000000   10.250000    3.850000    0.000000
50%     16.000000  2018.500000   13.600000    7.200000    0.800000
75%     23.000000  2019.000000   17.750000   10.850000    3.900000
max     31.000000  2019.000000   32.000000   18.900000   33.400000
std      8.802278     0.500343    5.106169    4.442844    4.402685

             wdsp          hg         sun                 date  \
count  730.000000  724.000000  728.000000                  730
mean     9.264658   24.968232    3.900275  2018-12-31 12:00:00
min      2.300000    7.000000    0.000000  2018-01-01 00:00:00
25%      6.100000   18.000000    0.500000  2018-07-02 06:00:00
50%      8.750000   24.000000    2.700000  2018-12-31 12:00:00
75%     11.700000   30.000000    6.500000  2019-07-01 18:00:00
max     25.200000   66.000000   15.600000  2019-12-31 00:00:00
std      4.013313    9.576674    3.883124                  NaN

                      date_diff
count                       730
```

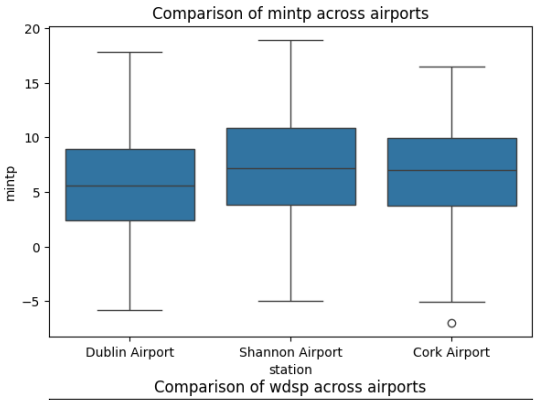
```
mean    0 days 06:50:18.082191780
min               0 days 00:00:00
25%               0 days 00:00:00
50%               0 days 00:00:00
75%               1 days 00:00:00
max               1 days 00:00:00
std     0 days 10:50:26.128233934
Numerical summary for Cork Airport:
              day         year       maxtp       mintp        rain  \
count  730.000000   730.000000  723.000000  723.000000  723.000000
mean    15.720548  2018.500000   13.262517    6.955325    3.470124
min      1.000000  2018.000000   -1.800000   -7.000000    0.000000
25%      8.000000  2018.000000    9.750000    3.750000    0.000000
50%     16.000000  2018.500000   12.500000    7.000000    0.700000
75%     23.000000  2019.000000   17.100000    9.900000    4.250000
max     31.000000  2019.000000   26.700000   16.500000   54.600000
std      8.802278     0.500343    4.894583    4.203288    6.137343

             wdsp          hg         sun                 date  \
count  728.000000  726.000000  727.000000                  730
mean     9.758516   25.976584    4.225860  2018-12-31 12:00:00
min      2.900000   10.000000    0.000000  2018-01-01 00:00:00
25%      6.700000   19.000000    0.300000  2018-07-02 06:00:00
50%      9.100000   24.500000    3.100000  2018-12-31 12:00:00
75%     12.200000   32.000000    7.250000  2019-07-01 18:00:00
max     23.800000   63.000000   15.700000  2019-12-31 00:00:00
std      3.749587    9.336514    4.055068                  NaN

                     date_diff
count                      730
mean    0 days 04:59:50.136986301
min               0 days 00:00:00
25%               0 days 00:00:00
50%               0 days 00:00:00
75%               0 days 00:00:00
max               1 days 00:00:00
std     0 days 09:45:05.427125996
```

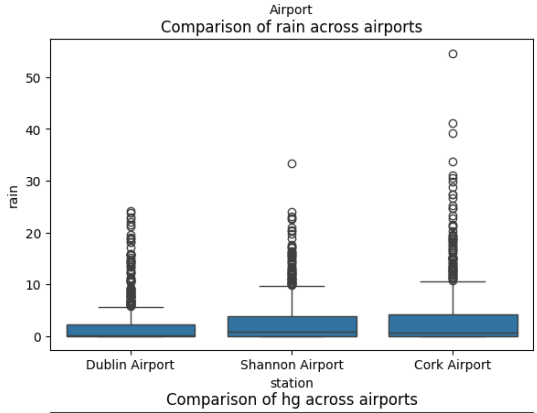
[17]:
```
plt.figure(figsize=(15,15))
# Create a boxplot for maxtp at each airport
plt.subplot(3,2,1)
sns.boxplot(x='station', y='maxtp', data=weather[weather['station'].
  ↪isin(['Dublin Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of maxtp across airports')
plt.xlabel('Airport')
plt.ylabel('maxtp')
```

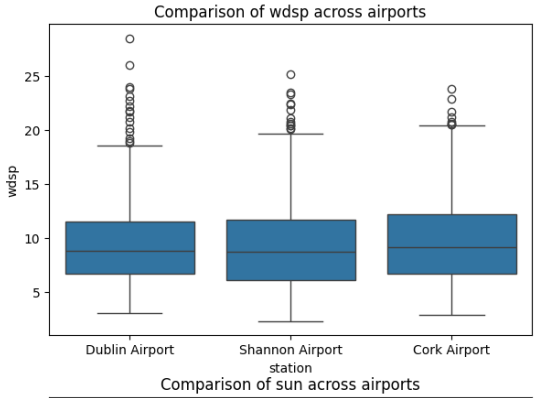
```
# Create a boxplot for mintp at each airport
plt.subplot(3,2,2)
sns.boxplot(x='station', y='mintp', data=weather[weather['station'].
 ↪isin(['Dublin Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of mintp across airports')
plt.ylabel('mintp')

# Create a boxplot for rain at each airport
plt.subplot(3,2,3)
sns.boxplot(x='station', y='rain', data=weather[weather['station'].
 ↪isin(['Dublin Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of rain across airports')
plt.ylabel('rain')

# Create a boxplot for wind speed at each airport
plt.subplot(3,2,4)
sns.boxplot(x='station', y='wdsp', data=weather[weather['station'].
 ↪isin(['Dublin Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of wdsp across airports')
plt.ylabel('wdsp')

# Create a boxplot for hg at each airport
plt.subplot(3,2,5)
sns.boxplot(x='station', y='hg', data=weather[weather['station'].isin(['Dublin␣
 ↪Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of hg across airports')
plt.ylabel('hg')

# Create a boxplot for sun at each airport
plt.subplot(3,2,6)
sns.boxplot(x='station', y='sun', data=weather[weather['station'].isin(['Dublin␣
 ↪Airport', 'Shannon Airport', 'Cork Airport'])])
plt.title('Comparison of sun across airports')
plt.ylabel('sun')
```

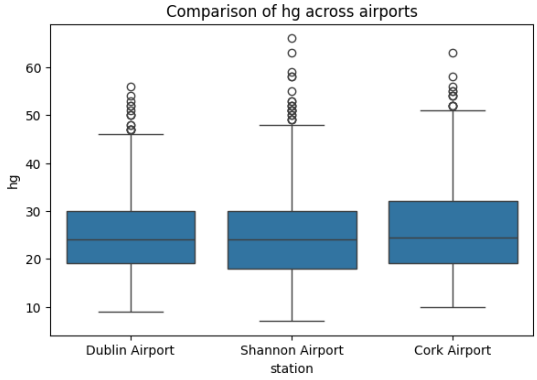
[17]: `Text(0, 0.5, 'sun')`

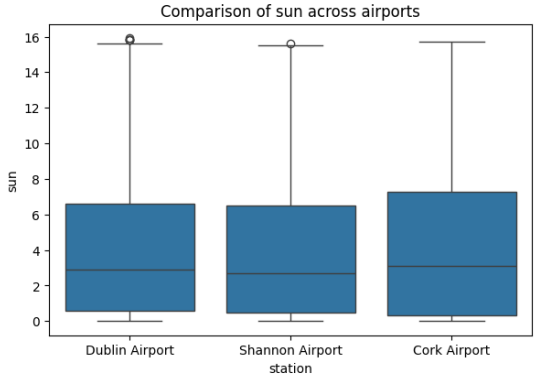

Comparison of maxtp across airports
maxtp
Dublin Airport
Shannon Airport
Cork Airport
Airport


Comparison of mintp across airports
mintp
Dublin Airport
Shannon Airport
Cork Airport
station


Comparison of rain across airports
rain
Dublin Airport
Shannon Airport
Cork Airport
station


Comparison of wdsp across airports
wdsp
Dublin Airport
Shannon Airport
Cork Airport
station


Comparison of hg across airports
hg
Dublin Airport
Shannon Airport
Cork Airport
station


Comparison of sun across airports
sun
Dublin Airport
Shannon Airport
Cork Airport
station

[ ]:

[ ]:

[ ]:

[ ]: