Irish Weather Data Analysis - EDA in Python

June 25, 2025

[ ]:
```
#Exploratory data analysis of the Irish weather
#Load in necessary packages
import numpy as np
import pandas as pd
from pandas import DataFrame, Series
import seaborn as sns
import matplotlib.pyplot as plt
```

1. Load in the weather_1819.csv dataset into Python as a pandas DataFrame.

[3]:
```
#Import data as dataframe
path = ''
weather_df = pd.read_csv(path+'weather_1819.csv', skiprows=11)
weather_df.head()
```

[3]:

	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun
0	1	jan	2018	Dublin Airport	7.5	3.2	0.6	18.5	41.0	2.7
1	2	jan	2018	Dublin Airport	11.1	3.4	8.4	17.0	54.0	0.8
2	3	jan	2018	Dublin Airport	8.1	4.6	1.3	23.8	51.0	0.9
3	4	jan	2018	Dublin Airport	9.3	3.4	10.7	14.6	48.0	1.3
4	5	jan	2018	Dublin Airport	6.7	-1.0	0.0	6.6	16.0	3.3

[4]:
```
#Find unique years in the data set. Use len() to find the number of years.
unique_years=weather_df.year.unique()
print('The data represents Irish weather for {} years i.e.: {}'.
 ↪format(len(unique_years), unique_years))
```

```
The data represents Irish weather for 2 years i.e.: [2018 2019]
```

[5]:
```
#Create a date column - in a copy dataframe based on current data
month_map = {'jan': 1, 'feb': 2, 'mar': 3, 'apr': 4,'may': 5, 'jun': 6, 'jul':␣
 ↪7, 'aug': 8,'sep': 9, 'oct': 10, 'nov': 11, 'dec': 12}
weather_df_copy = weather_df.copy()
weather_df_copy['month'] = weather_df['month'].map(month_map)

#Sort based on the new date column
weather_df_copy['date'] = pd.to_datetime(weather_df_copy[['year', 'month',␣
 ↪'day']]).sort_values()
```

```
#Finding unique values to see difference in days between adjacent dates of␣
 ↪collecting data
unique_dates_df = pd.DataFrame(weather_df_copy['date'].unique(),␣
 ↪columns=['date'])

#Print head of new dataframe to analyse temporal resolution
unique_dates_df.head()
```

[5]:
```
        date
0 2018-01-01
1 2018-01-02
2 2018-01-03
3 2018-01-04
4 2018-01-05
```

From the above data we can see that the temporal resolution is 1 day (daily).

[6]:
```
#The full form for the weather measurements are available in the 5-10 rows in␣
 ↪csv file.
weather_measurements=pd.read_csv('weather_1819.
 ↪csv',skiprows=3,nrows=6,usecols=[0])

print('Weather measurements reported are: \n',weather_measurements,'\n')

#We can also get the abbreviated weather measurements from the dataframe we␣
 ↪created earlier.
print('These are abbreviated as following columns in the weather dataframe: {}'.
 ↪format((weather_df.columns[4:]).tolist()))
```

```
Weather measurements reported are:
     station: location of measurement
0  maxtp: Maximum Air Temperature (C)\t
1     mintp: Minimum  Air Temperature (C)
2         rain: Precipitation Amount (mm)
3            wdsp: Mean Wind Speed (knot)
4                 hg: Highest Gust (knot)
5          sun: Sunshine duration (hours)

These are abbreviated as following columns in the weather dataframe: ['maxtp',
'mintp', 'rain', 'wdsp', 'hg', 'sun']
```

2. Determine how many missing values there are in each column of the dataset.

[7]:
```
#Get count of missing values
weather_df.isnull().sum()
```

[7]:
```
day          0
month        0
```

```
year        0
station     0
maxtp      18
mintp      18
rain       23
wdsp        5
hg         16
sun         7
dtype: int64
```

The above list provides the number of missing values for each columns. date parameters and station columns does not have any missing values. However, weather measurements do have missing values in them. This could be due to:

- Missing Values in the CSV i.e, 'Blanks' in the data (truly missing data)
- From a data standpoint, these missing values maybe caused by measurement limitations, reporting issues/human errors, instrument errors etc.

To fill the missing values we can do the following:

- Using fillna with fixed values: We can specify different fill values for each weather measurement column (e.g., weather_df.fillna({'wdsp': 10, 'sun': 0})).
  - Advantage: This method is easy to implement and improves data appearance.
  - Disadvanatge: In real scenario like weather with several variations among data, it does not make sense to put a constant fixed data. It may introduce bias and misrepresent the actual variations in the data.
- Using fillna with column means: Replacing missing values with the mean of the respective columns (ex: weather_df.fillna({'rain': weather_df['rain'].mean()})).
  - Advantage: Provides a more representative solution.
  - Disadvantage: This approach might not accurately represent the actual data and could still mask notable variations in weather patterns.
- Using Backfill (bfill) and Forward Fill (ffill): (ex: weather_df['rain'].fillna(method='bfill')).
  - Advantage: These methods can better reflect seasonal changes by using nearby values (say 2 consecutive days in Spring/Summer).
  - Disadvantage: However, they may not represent actual weather conditions if drastic changes occur between readings in consecutive days which can often occur in case of weather.
- Using dropna: If we choose to remove rows with missing values (ex: weather_df.dropna(subset=['rain'])):
  - Advantage: We can achieve a cleaner dataset.
  - Disadvantage: This leads to the loss of potentially valuable data, which can affect overall analysis.

3. 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?

[8]:
```
#create tuple for use in loop
```

```
max_tuple=[('wdsp', 'Highest windspeed', 'knots'),('maxtp', 'Highest maximum␣
 ↪air temperature', 'Celsius'),('rain', 'Largest amount of rain', 'mm')]

#Use loop to calculate the mentioned values
for column, fullform, unit in max_tuple:
    max_row = weather_df_copy.loc[weather_df_copy[column].idxmax()]
    print(f"{fullform} ({max_row[column]} {unit}), was recorded in␣
 ↪{max_row['station']} station on {max_row['date'].date()}.")
```

```
Highest windspeed (28.5 knots), was recorded in Dublin Airport station on
2018-03-02.
Highest maximum air temperature (32.0 Celsius), was recorded in Shannon Airport
station on 2018-06-28.
Largest amount of rain (54.6 mm), was recorded in Cork Airport station on
2019-04-15.
```

4. Create a numerical summary (mean, standard deviation, minimum, maximum, etc.) for each of the weather measurements.

[9]:
```
weather_df.drop(columns=['year', 'day']).describe()
```

[9]:

	maxtp	mintp	rain	wdsp	hg	sun
count	2902.000000	2902.000000	2897.000000	2915.000000	2904.000000	2913.000000
mean	13.283150	6.432977	3.063583	9.481475	25.443871	3.783797
std	5.146289	4.368755	5.053881	3.820605	9.278313	3.850012
min	-1.800000	-7.000000	0.000000	2.300000	7.000000	0.000000
25%	9.500000	3.100000	0.000000	6.500000	19.000000	0.300000
50%	12.800000	6.400000	0.700000	8.900000	24.000000	2.600000
75%	17.100000	9.600000	4.000000	11.800000	30.000000	6.300000
max	32.000000	18.900000	54.600000	28.500000	84.000000	15.900000

Numerical summary interpretation (Fig 1.):

- **count**: The count of observations of each weather measurements are given, and these are not consistent indicating missing values in the dataframe (as we had seen earlier).

Analysis on individual weather measurements:

Maximum temperature (maxtp):

- **mean**: The average maximum temperature is 13.28°C.

- **Standard deviation (sd)**: A value of ~5°C, just like minimum temperature, could be considered a high deviation in this context. This makes sense as temperatures can change significantly on a daily basis.
- **Minimum (min)**: The lowest maximum air temperature of -1.8°C. This shows that every day the temperature went below 0 at some point.
- **25% (25th percentile)**: 25% of the days have maximum temperature 9.5°C or below.
- **50% (50th percentile/Median)**: 50% of the data has maximum temperatures below 12.8°C.
- **75% (75th percentile)**: 75% of the days have a maximum temperature lower than ~17°C.
- **Maximum (max)**: The highest maximum temperature is 32°C. This is very high above the mean (13.28°C) and the 75th percentile (~17°C). These could be rare case/s (outliers).

Minimum temperature (min):

- **mean**: The average temperature is 6.43°C.
- **sd**: A value of 4.3°C could be considered a high deviation in this context, which makes sense as temperatures can change significantly on a daily basis.
- **min**: The lowest minimum temperature is -7°C and this represents extreme cold day/s (probably night times). This is very less compared to mean of 6.43°C, showing these extreme cold days are occasional.
- **25%**: Shows that a 25% of the days are at ~3°C or lower minimum temperature.
- **50%**: The mean is 6.43°C which is very close to the median of ~6.4°C. This could mean that minimum temperature distribution is almost symmetrical.
- **75%**: Shows that 75% of the days have minimum temperature below 9.6°C, i.e., most days are moderately cold.
- **max**: The highest minimum temperature is ~19°C and this much higher than the mean (~6.4°C). This could mean that warm days are also occasional.

Precipitation amount (rain):

- **mean**: Average precipitation amount across the data in 2018-2019 is ~3mm.
- **sd**: The sd for rain is high at ~5mm compared to the mean of ~3mm, showing that rainfall measurements in this dataset change significantly from the mean.
- **min**: We can see that for the minimum value for rain (precipitation amount) is 0 mm and this makes sense as there could be no rain days.
- **25%**: Also known as 1st quartile. For rain, this value is 0mm. This indicates that no rain days are frequent and at least the lowest 25% consists of days with no rainfall.
- **50%**: As per data, half of the days have less than 0.7mm amount of rainfall.
- **75%**: Rainfall(~4mm) is closer to mean (~3mm) than median (0.7mm). This would mean that these high rainfalls are occasional.
- **max**: The day/s with maximum rainfall is having a value of 54.6mm. This is considerably higher than both the mean (~3mm) and the 75th percentile (4mm).

Mean wind speed (wdsp):

- **mean**: Average mean wind speed is 9.48 knots.
- **sd**: 3.82 knots.
- **min**: The minimum mean wind speed is 2.3 knots which represents day/s with less wind.
- **25%**: 25% of the days have mean wind speeds among the data below 6.5 knots.
- **50%**: Half of the days have mean wind speeds below 8.9 knot.
- **75%**: 75% of data have mean wind speed below 11.8 knots.

- **max**: The maximum mean wind speed is recorded at 28.5 knots. This is nearly 3 times the mean (9.48 knots). Therefore, these could be outliers or rare cases.

Highest gust (hg):

- **mean**: Average is ~25.44 knots.
- **sd**: We can observe that this value is maximum in case of highest gust (hg) over the weather entire dataset from 2018-2019 (9.27 knots), showing rapid changes from mild to extremely gusty.
- **min**: A minimum gust of 7 knots is seen. This shows there is always some wind. This value is very lower than the mean of 25.44 knots, meaning on average, winds are much stronger.
- **25%**: 25% of the data shows value below 19 knots.
- **50%**: The value is 24 knots, showing half of the days have wind strengths below this.
- **75%**: 75% of the days have gust below ~30 knots.
- **max**: The maximum highest gust level is 84 knots. This is very much higher than the 75th percentile (30 knots) and mean (25.44 knots). This shows occasional high gusty day/s.

Sunshine duration (sun):

- **mean**: On average, we got about 3.78 hours of sunshine each day in 2018-2019 in Ireland.
- **sd**: Variability in daily sunshine duration is 3.85 hours.
- **min**: We can see minimum value of sunshine duration is 0 hours and this makes sense as there could be days with limited or no sunshine. This suggests cloudy or rainy days.
- **25%**: The 1st quartile value is 0.3 hours, which shows that 25% of the days have very little sunshine.
- **50%**: Half of the days have 2.6 hours of sunshine or less.
- **75%**: For a significant portion of the data sunshine duration is limited at 6.3 hours.
- **max**: Sunshine is maximum for ~16 hours, which is very higher than the mean (3.78 hours) and the 75th percentile (6.3 hours). This could be occasional or seasonal sunny day/s.

5. Create a graphical summary for each of the weather measurements.

[10]:
```
sns.set_style("darkgrid")
plt.figure(figsize=(12, 18))

colors = sns.color_palette("Set2", 6)


#Create tuple for using later in loop
plot_titles =(('maxtp','Maximum temperature','Celsius'),('mintp','Minimum␣
 ↪temperature','Celsius'),('rain','Precipitation amount','mm'),
              ('wdsp','Mean wind speed','knots'),('hg', 'Highest gust',␣
 ↪'knots'),('sun','Sunshine duration','hours'))

sns.set_style("darkgrid")
plt.figure(figsize=(12, 18))
colors = sns.color_palette("Set2", 6)

#Loop for generating plot of each weather measurements
for i in range(len(plot_titles)):
```

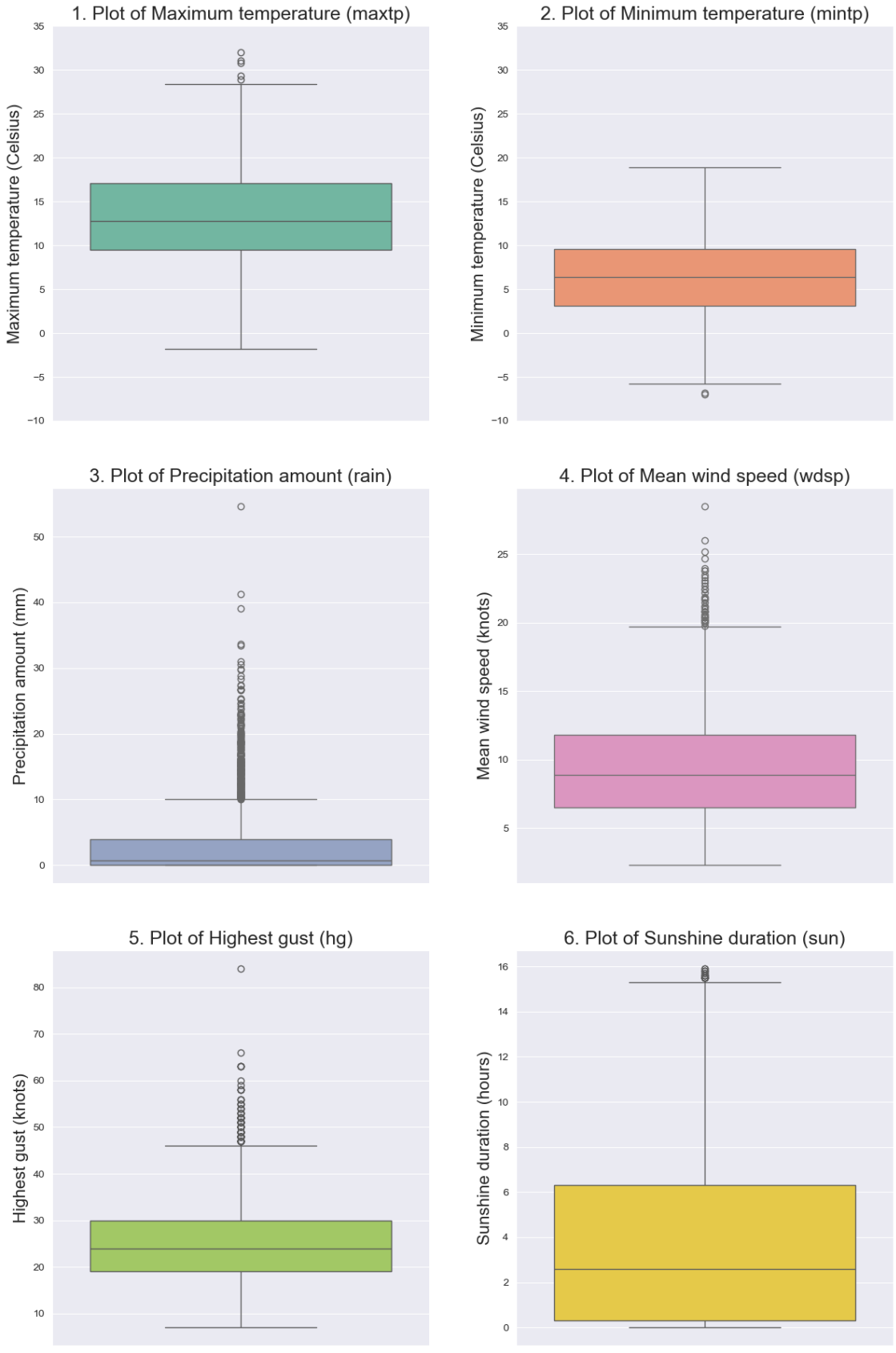
```
    measure, label, unit = plot_titles[i]

    plt.subplot(3,2,1+i)
    if measure=='maxtp' or measure=='mintp':
        plt.ylim(-10,35) #To align temperature measurements for comparison
    sns.boxplot(data=weather_df, y=measure, color=colors[i])
    plt.ylabel(f'{label} ({unit})', size=16)
    plt.title(f'{i+1}. Plot of {label} ({measure})', size=18)

#Provide appropriate spacing and main title
plt.tight_layout(h_pad=4, w_pad=4)
plt.suptitle('Fig 1. Boxplots of Irish weather measurements from 2018-2019',␣
 ↪y=1.04, size=20)
plt.show()
```

```
<Figure size 1200x1800 with 0 Axes>
```

Fig 1. Boxplots of Irish weather measurements from 2018-2019



Graphical summary observations (Fig 2.):

Maximum temperature (maxtp):

- maxtp (maximum temperature) has a considerably higher median compared to mintp. This means there is a wide difference range between daily maximums and minimums.
- The spread of maxtp is is also wider than that of mintp. This shows more variations in maximum temperatures.
- maxtp has outliers above upper whisker. These could be days with extreme high temperature (this is backed up by numerical summary we had seen earlier around 32°C).
- This boxplot suggests that distribution of maximum temperature is right skewed (median is a little placed towards bottom). We had seen in numerical summary mean is 13.28°C and median is 12.28°C, which suggests the same – median is below mean. This means there are a large number of records on the lower readings than higher readings of maximum temperature.

Minimum temperature (mintp):

- Like mentioned in the numerical summary, minimum temperature has an almost symmetrical distribution (mean and median are very close ~6.4°C).
- It has outliers below lower whisker populated at around minimum value of minimum temperatures we had observed in the numerical summary, i.e, -7°C.
- The whole box (IQR) is placed below 10°C (3rd quartile) showing most days are moderately cold as mentioned in numerical summary.

Precipitation amounts (rain):

- The median of rainfall is quite low. The placement of median towards the bottom in the box (IQR or inter quartile region) shows a skewed distribution where many days have little or no rain. This information was also observed in numerical summary where we saw mean and 1st quartile was at 0mm, and 3rd quartile was much closer to mean than median.
- However, there are a lot of outliers observed in the plot. These could be days with heavy rains.
- Rain shows extreme values, suggesting occasional heavy rain events that could appear as outliers in a boxplot. We can see the exact value of maximum amount for rain was at 54.6mm from numerical summary.

Mean wind speed (wdsp):

- Mean wind speed have slightly lower median. This indicates a slight skewness.
- There are a lot of outliers. One that is particularly far in the maximum we observed in numerical summary which is ~85 knots.

Highest gust (hg):

- hg has a considerably high median compared to mean wind speeds.
- We see that the whisker start above 0. If we look at numerical summary or boxplot we can see that the minimum gust 7 knots. This means (as mentioned earlier), there is always some wind.
- There are many outliers and one particular outlier is very far from the rest of the data. This is observed at around 84 knots, as provided in numerical summary as well.

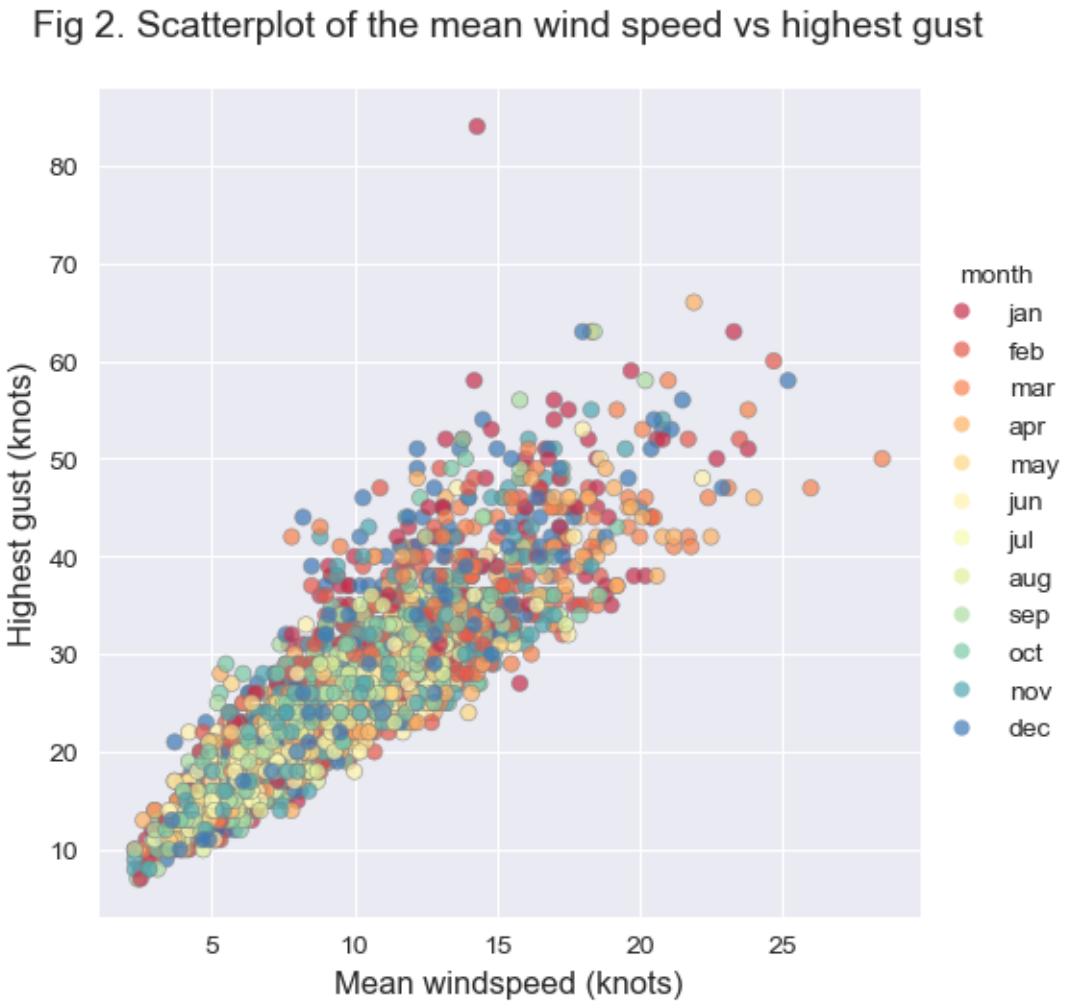
- The distribution for high gust is also almost symmetrical like minimum temperature. This is backup by the mean and median values we observed in numerical summary, which are fairly close (24-25 knots).

Sunshine duration (sun):

- We can see lower whisker of the box start at 0 hours. This is as mentioned in numerical summary, where these values represent days with no or almost no sunshine.
- It particularly have a narrower range with fewer outliers, reflecting less extreme variation.

6. Produce a scatter plot of the mean wind speed versus the highest gust.

[11]:
```
sns.relplot(data=weather_df, y='hg', x='wdsp', hue='month', palette='Spectral',␣
 ↪alpha=0.7, edgecolor='grey')
plt.ylabel('Highest gust (knots)', fontsize=12)
plt.xlabel('Mean windspeed (knots)', fontsize=12)
plt.title('Fig 2. Scatterplot of the mean wind speed vs highest gust',␣
 ↪fontsize=14, pad=20)
plt.show()
```



- From Fig 2., we can a positive linear relationship between mean wind speed (wdsp) and highest gust (hg). We can observe very closely packed points towards a direction with a positive slope.
- However, the spread of the points also indicates little variations in gust for a given wind speed.
- One particular point is located far away from this group or cluster of other points. This is an outlier. This points show a relatively very high highest gust for a particular wind speed and does not follow the general trend. From the color of the point we can see that this outlier falls in data collected from January. This value could affect our analysis if not properly considered.
- The cluster of points in the low wind readings of the plot are mostly green and yellow which represents data from months June - September. This shows a seasonal pattern of comparitiley lesser wind months.
- On the higher value side, we can see more orange, red, and blue colors. This indicates that during the November to April period, the weather in Ireland is very windy.

7. Compute the daily temperature range, and add this as an additional variable to the DataFrame.

[12]:
```
weather_df['temp_range']=weather_df['maxtp']-weather_df['mintp']

#Print last 10 rows of dataframe. We can also use below code within print()
weather_df.tail(10)
```

[12]:

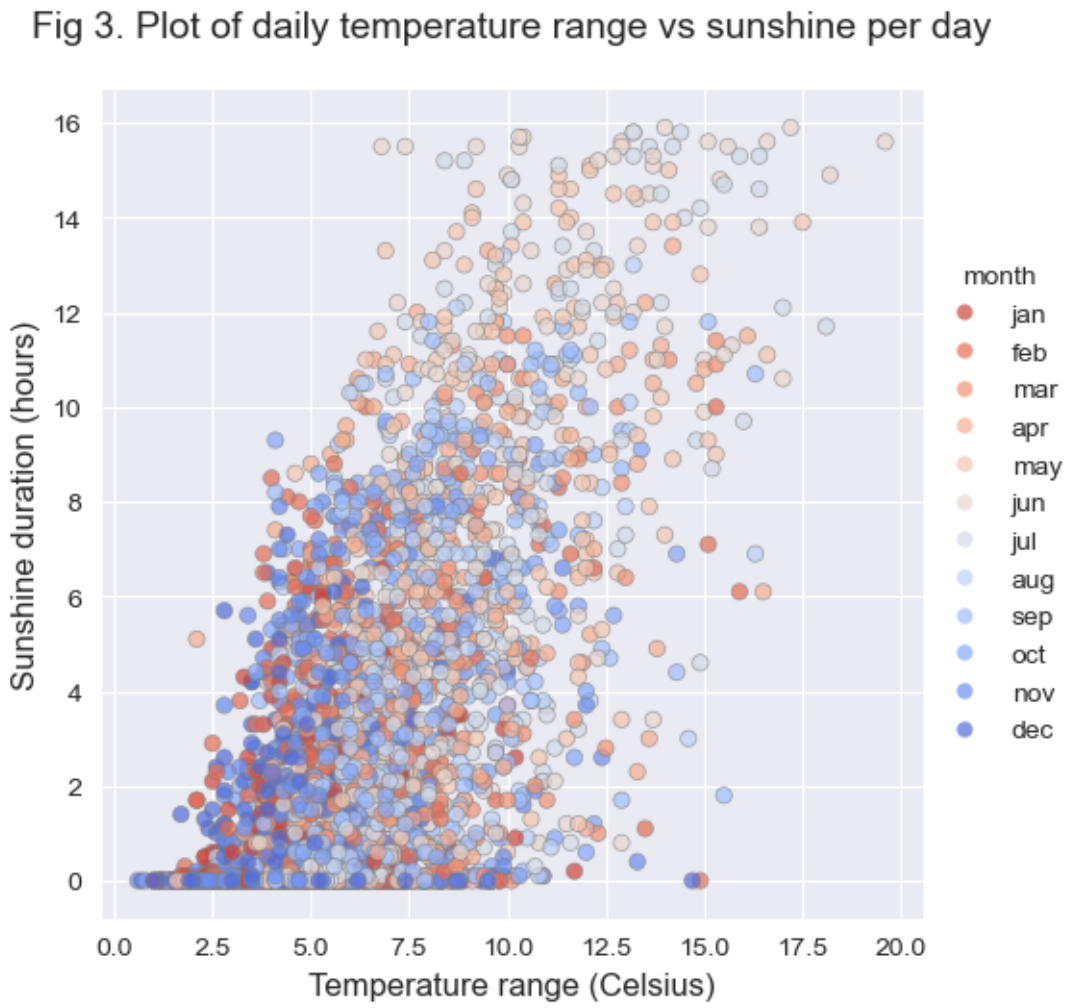
	day	month	year	station	maxtp	mintp	rain	wdsp	hg	sun	\
2910	22	dec	2019	Knock Airport	7.0	2.7	4.8	8.0	20.0	3.1	
2911	23	dec	2019	Knock Airport	6.9	3.0	5.0	10.5	29.0	2.7	
2912	24	dec	2019	Knock Airport	6.6	3.1	0.7	8.4	24.0	2.9	
2913	25	dec	2019	Knock Airport	4.9	1.4	0.0	9.0	32.0	0.0	
2914	26	dec	2019	Knock Airport	10.0	4.8	6.2	12.2	34.0	0.0	
2915	27	dec	2019	Knock Airport	10.8	9.8	3.7	12.8	29.0	0.0	
2916	28	dec	2019	Knock Airport	9.8	7.9	0.0	14.8	30.0	0.0	
2917	29	dec	2019	Knock Airport	10.5	7.0	0.0	12.8	32.0	0.1	
2918	30	dec	2019	Knock Airport	9.6	3.4	NaN	8.4	21.0	0.0	
2919	31	dec	2019	Knock Airport	6.0	1.3	0.0	7.1	18.0	2.1	

	temp_range
2910	4.3
2911	3.9
2912	3.5
2913	3.5
2914	5.2
2915	1.0
2916	1.9

```
2917         3.5
2918         6.2
2919         4.7
```

8. Plot the daily temperature range versus the hours of sunlight per day, colouring the points based on month.

[16]:
```
sns.relplot(data=weather_df,  x='temp_range', y='sun', hue='month',␣
 ↪palette='coolwarm_r', edgecolor='grey', alpha=0.7)
plt.xlabel('Temperature range (Celsius)', fontsize=12)
plt.ylabel('Sunshine duration (hours)', fontsize=12)
plt.title('Fig 3. Plot of daily temperature range vs sunshine per␣
 ↪day',fontsize= 14,pad=20)
plt.show()
```



- From Fig 3., we can see a positive relation between temperature range and sunshine duration.

This relationship is not as strong as one we saw in Fig 2.

- The wide spread suggestions high variations in sunshine for a given temperature range.
- We can see a lot of outliers in this case, which means that sunshine duration varied significantly for a given temperature range compared to general trend.
- Based on coloring of months, we can see towards the start and end of the year, very little sunshine as well as low temperature ranges are observed. This represents winter season changes in months October - February. The little to no sunshine suggests that it will be mostly dark throughout the day in this period.
- The months in the middle such as May - July show higher temperatures indicating summer in Ireland.
- Overall, we can see that temperature range fall below 20. and sunshine is not more than ~16 hours.

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

[14]:
```
#Filter based on station and create a new dataframe
weather_df_new = weather_df[weather_df.station.isin(['Dublin Airport', 'Shannon␣
 ↪Airport', 'Cork Airport'])][weather_df.columns[3:10]]

#use describe() to get summary statistics and groupby for grouping by station.␣
 ↪Transpose() will give a cleaner look and reset index() updates individual␣
 ↪records with weather measure it represents.
weather_df_new_describe=weather_df_new.groupby('station').describe().
 ↪transpose().reset_index()

#Rename column level_0 and level_1 to measure and statistic. use sort_values()␣
 ↪for sorting. Rest_index() again for new index after sorting.
weather_df_new_describe.columns = ['measure', 'statistic'] +␣
 ↪list(weather_df_new_describe.columns[2:5])
weather_df_new_describe.sort_values(by=['measure', 'statistic']).
 ↪reset_index(drop=True, inplace=True)

#see values of the new dataframe (this gives numerical summary of each weather␣
 ↪measurement by station)
weather_df_new_describe
```

[14]:

	measure	statistic	Cork Airport	Dublin Airport	Shannon Airport
0	maxtp	count	723.000000	730.000000	723.000000
1	maxtp	mean	13.262517	13.592603	14.160028
2	maxtp	std	4.894583	5.265505	5.106169
3	maxtp	min	-1.800000	-0.500000	0.000000
4	maxtp	25%	9.750000	9.500000	10.250000
5	maxtp	50%	12.500000	13.000000	13.600000
6	maxtp	75%	17.100000	17.875000	17.750000

7	maxtp	max	26.700000	26.700000	32.000000
8	mintp	count	723.000000	730.000000	723.000000
9	mintp	mean	6.955325	5.724247	7.267082
10	mintp	std	4.203288	4.488748	4.442844
11	mintp	min	-7.000000	-5.800000	-5.000000
12	mintp	25%	3.750000	2.400000	3.850000
13	mintp	50%	7.000000	5.600000	7.200000
14	mintp	75%	9.900000	8.900000	10.850000
15	mintp	max	16.500000	17.800000	18.900000
16	rain	count	723.000000	726.000000	721.000000
17	rain	mean	3.470124	2.147383	2.844383
18	rain	std	6.137343	4.120571	4.402685
19	rain	min	0.000000	0.000000	0.000000
20	rain	25%	0.000000	0.000000	0.000000
21	rain	50%	0.700000	0.200000	0.800000
22	rain	75%	4.250000	2.300000	3.900000
23	rain	max	54.600000	24.200000	33.400000
24	wdsp	count	728.000000	728.000000	730.000000
25	wdsp	mean	9.758516	9.542720	9.264658
26	wdsp	std	3.749587	3.829729	4.013313
27	wdsp	min	2.900000	3.000000	2.300000
28	wdsp	25%	6.700000	6.700000	6.100000
29	wdsp	50%	9.100000	8.800000	8.750000
30	wdsp	75%	12.200000	11.500000	11.700000
31	wdsp	max	23.800000	28.500000	25.200000
32	hg	count	726.000000	728.000000	724.000000
33	hg	mean	25.976584	24.943681	24.968232
34	hg	std	9.336514	8.438033	9.576674
35	hg	min	10.000000	9.000000	7.000000
36	hg	25%	19.000000	19.000000	18.000000
37	hg	50%	24.500000	24.000000	24.000000
38	hg	75%	32.000000	30.000000	30.000000
39	hg	max	63.000000	56.000000	66.000000
40	sun	count	727.000000	729.000000	728.000000
41	sun	mean	4.225860	4.046091	3.900275
42	sun	std	4.055068	3.952654	3.883124
43	sun	min	0.000000	0.000000	0.000000
44	sun	25%	0.300000	0.600000	0.500000
45	sun	50%	3.100000	2.900000	2.700000
46	sun	75%	7.250000	6.600000	6.500000
47	sun	max	15.700000	15.900000	15.600000

[15]:
```
#Create tuple for using later in loop
plot_titles =(('maxtp','Maximum temperature','Celsius'),('mintp','Minimum␣
 ↪temperature','Celsius'),('rain','Precipitation amount','mm'),
              ('wdsp','Mean wind speed','knots'),('hg', 'Highest gust',␣
 ↪'knots'),('sun','Sunshine duration','hours'))
```

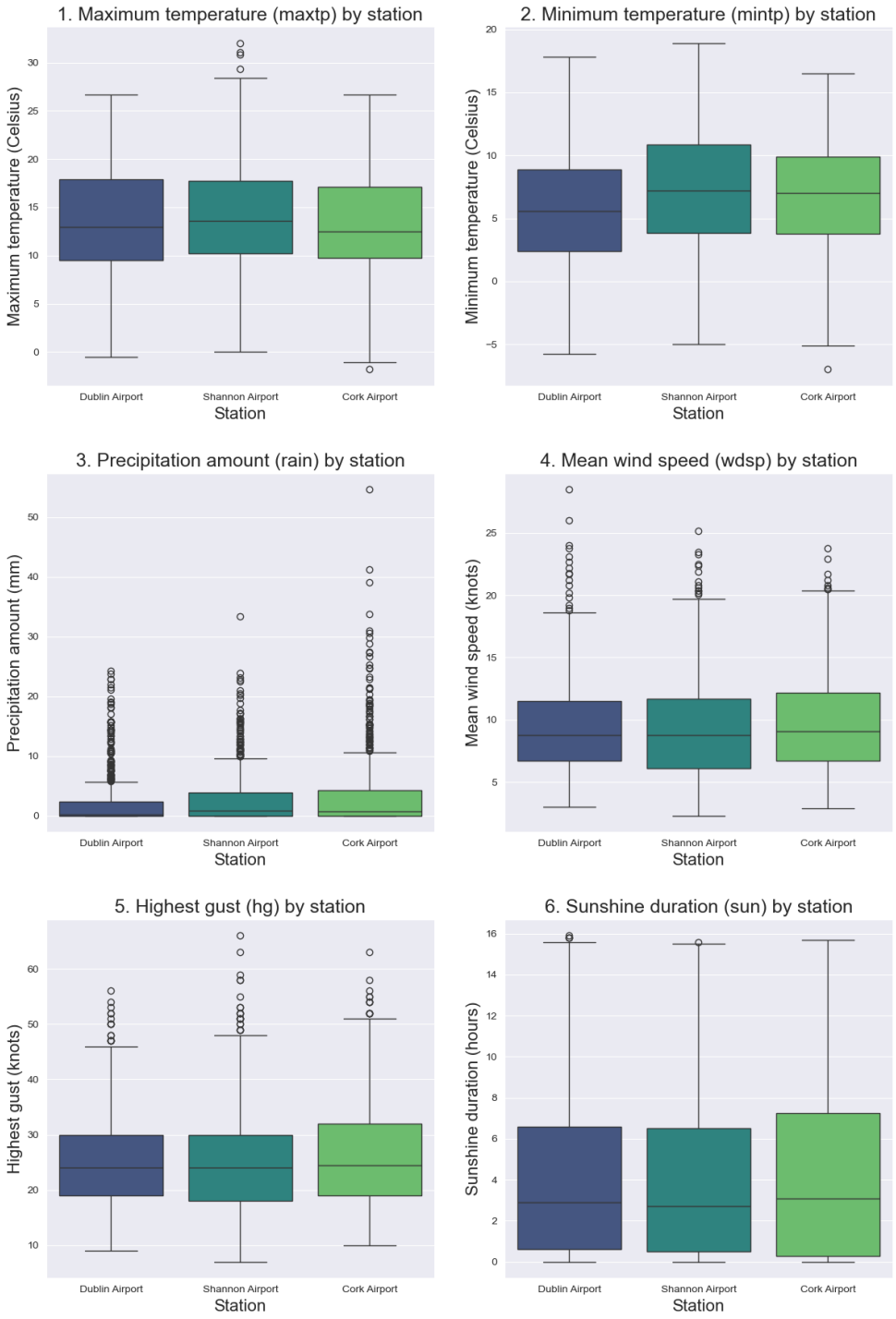
```
plt.figure(figsize=(12, 18))

#Loop for generating boxplot of each weather measurement by station
for i in range(len(plot_titles)):
    measure, label, unit = plot_titles[i]
    plt.subplot(3,2,1+i)
    sns.boxplot(data=weather_df_new, x='station', y=measure, hue='station',␣
 ↪palette='viridis')
    plt.ylabel(f'{label} ({unit})', size=16)
    plt.xlabel('Station', size=16)
    plt.title(f'{i+1}. {label} ({measure}) by station', size=18)

#Provide appropriate spacing and main title
plt.suptitle('Fig 4. Grouped boxplots of Irish weather measurements from␣
 ↪2018-2019 in Dublin, Shannon and Cork airports', y=1.02, size=20)
plt.tight_layout(h_pad=3,w_pad=3)

plt.show()
```

Fig 4. Grouped boxplots of Irish weather measurements from 2018-2019 in Dublin, Shannon and Cork airports



Maximum temperature (maxtp):

- Fig 4. illustrates that Shannon airport has the highest median temperature at 13.6°C.

- The box (IQR region) is widest for Dublin Airport, implying more variations in temperature within the middle compared to Cork and Shannon airports.
- From whiskers we can see that Cork experiences the lowest extremes. However, the upper whisker for Shannon airport is at a higher temperature (32.0°C as we can see from numerical summary).
- Also we can immediately notice is that Shannon airport has got outlier above the upper whisker. For Dublin and Cork airport stations, this outlier presence is not seen.
- This means that significantly high temperatures occurred in Shannon airport compared to Cork and Dublin according to the 2018-2019 weather report.

Minimum temperature (mintp):

- Shannon airport has seen the highest median minimum temperature at 7.2°C.
- Its IQR is also wider than Dublin and Cork airports. This implies that there are great variations in minimum temperature that occurred within the middle of the data compared to Dublin and Cork.
- Noticeably, Cork has outliers below the lower whisker. From numerical summary, we can confirm the same and see that this occurred at temperature of -7.0°C.
- Compared to other measures, minimum temperature seems to have very less outliers.

Precipitation amount (rain):

- All 3 stations seemed to have right skewed distribution as median is placed much towards the bottom.
- From Fig 4, we can confirm that Cork airport has the highest average rainfall. The height of whisker and presence of outliers is very high compared to Dublin and Shannon airports. This indicates that at Cork airport had experience extreme rainfall.
- From numerical summary, we can also see that mean of rain is also largest in Cork (~3.45mm).
- The IQR for Dublin is very narrow compared to Cork and Shannon. This means there are less variations in precipitation amounts in the middle of data from 25%-75%. Dublin also has less outliers indicating lesser extremes of rainfall.

Mean wind speed (wdsp):

- All 3 stations seems to have somewhat similar distributions.
- Cork airport has the highest median (~9.7knots). It also has the highest mean (~9.8knots).
- The IQR is narrow at Shannon airport data than Cork and Dublin, indicating lesser variations.
- Noticeably, the outliers are highest in Dublin airport than Cork and Shannon, which means Dublin experienced higher extremes of mean wind speed.

Highest gust (hg):

- The distribution looks somewhat similar between the 3 and they have similar medians and IQRs.
- However, Cork airport has wider IQR region, which means it has more variations in the middle 50% of data.
- Shannon airport has the highest outliers, and therefore we can say that it experienced highest extremes of gusts. From the table, we can see that maximum value of highest gust that occurred here is as much as 66knots.

Sunshine duration (sun):

- The distribution looks somewhat similar between the 3 airports and they have similar medians and IQRs. We can also see from the table that the means of sunshine durations are very close to each other in the 3 stations (~4hours).
- However, Cork airport shows more variations compared to the other two, as we can the box (IQR) is larger for Cork.
- The higher ends for all 3 stations appear to occur around similar readings (15.5-16 hours). Dublin and Shannon airports seems to have outliers, but Cork has a higher whisker.