Graph representing cases of COVID-19 per region in France

This document uses data taken from this data base: https://www.kaggle.com/lperez/coronavirus-france-dataset

It represents the number of cases per region at the start of the pandemic in France. It has been made to identify which region was the most affected by the virus, but also to see which region had the least cases. It has been made using python.

Author: Arthur Gatignol (arthur.gatignol@gmail.com)

Version: 0.1

In [5]:

```python
import pandas as pd
```

In [7]:

```python
df = pd.read_csv('patient.csv')
```

In [53]:

```python
# Importing the libraries to plot the graph
import matplotlib.pyplot as plt
import numpy as np
```

Variables

In the cell below, I am showing defining variables that will be used for plotting the graph, and which take in count the number of cases per region

In [59]:

```python
# Variables that will be used for the graph
Bretagne = 0
Occitanie = 0
IDF = 0
GrandEst = 0
Reunion = 0
Martinique = 0
Guadeloupe = 0
Guyane = 0
NouvelleAquitaine = 0
Corse = 0
Auvergne = 0
Bourgogne = 0
CentreVal = 0
HautFrance = 0
Normandie = 0
PACA = 0
PaysDeLaLoire = 0
StMartin = 0
StBart = 0
```

Algorithm

Below is the main algorithm that goes through the 'regions' column and will increment the above variables by 1 if the name of a certain region is mentionned within the table.

In [64]:

```python
# For loop that goes through the column of the regions
for i in df['region']:

    # For 'Nouvelle Aquitaine'
    if i == 'Nouvelle-Aquitaine':
```

```python
        NouvelleAquitaine = NouvelleAquitaine + 1

    # For 'Occitanie'
    elif i == 'Occitanie':
        Occitanie = Occitanie + 1

    # For 'Hauts de France'
    elif i == 'Hauts-de-France':
        HautFrance = HautFrance + 1

    # For 'Normandie'
    elif i == 'Normandie':
        Normandie = Normandie + 1

    # For 'Pays-de-la-Loire'
    elif i == 'Pays-de-la-Loire':
        PaysDeLaLoire = PaysDeLaLoire + 1

    # For 'Bourgogne-Franche-Comté'
    elif i == 'Bourgogne-Franche-Comté':
        Bourgogne = Bourgogne + 1

    # For 'Bretagne'
    elif i == 'Bretagne':
        Bretagne = Bretagne + 1

    # For 'PACA'
    elif i == 'PACA':
        PACA = PACA + 1

    # For 'Centre-Val de Loire'
    elif i == 'Centre-Val de Loire':
        CentreVal = CentreVal + 1

    # For 'Auvergne-Rhône-Alpes'
    elif i == 'Auvergne-Rhône-Alpes':
        Auvergne = Auvergne + 1

    # For 'Saint-Martin'
    elif i == 'Saint-Martin':
        StMartin = StMartin + 1

    # For 'Saint-Barthélémy'
    elif i == 'Saint-Barthélémy':
        StBart = StBart + 1

    # For 'Guyane'
    elif i == 'Guyane':
        Guyane = Guyane + 1

    # For 'Martinique'
    elif i == 'Martinique':
        Martinique = Martinique + 1

    # For 'Grand-Est'
    elif i == 'Grand-Est':
        GrandEst = GrandEst + 1

    # For 'Ile-de-France'
    elif i == 'Ile-de-France':
        IDF = IDF + 1

    # For 'Corse'
    elif i == 'Corse':
        Corse = Corse + 1

    # For 'La Reunion'
    elif i == 'La Reunion':
        Reunion = Reunion + 1

    # For 'Guadeloupe'
    elif i == 'Guadeloupe':
        Guadeloupe = Guadeloupe + 1

    # For 'Guyane'
    elif i == 'Guyane':
        Guyane = Guyane + 1
```

Plotting the graph

Below is the actual plotting of the graph and more globally how I have analysed the data to determine which region of France as the most cases of COVID-19.

In [62]:

```python
# Here I have learned from different documentations how
# I could change plot a bar graph using python and
# matplotlib.

regions = ['Nouvelle-Aquitaine', 'Occitanie', 'Hauts-de-France', 'Normandie', 'Pays-de-la-Loire',
           'Bourgogne-Franche-Comté', 'Bretagne', 'PACA', 'Centre-Val de Loire',
           'Auvergne-Rhône-Alpes', 'Saint-Martin', 'Saint-Barthélémy', 'Guyane',
           'Martinique', 'Grand-Est', 'Ile-de-France', 'Corse', 'La Reunion',
           'Guadeloupe',]

cases = [NouvelleAquitaine, Occitanie, HautFrance, Normandie, PaysDeLaLoire, Bourgogne,
         Bretagne, PACA, CentreVal, Auvergne, StMartin, StBart, Guyane, Martinique,
         GrandEst, IDF, Corse, Reunion, Guadeloupe]

y_pos = np.arange(len(regions))


plt.barh(y_pos, cases, align='center', alpha=0.5)
plt.yticks(y_pos, regions)
plt.xlabel('Number of COVID-19 cases')
plt.title('Number of cases of COVID-19 per region')

plt.show()
```
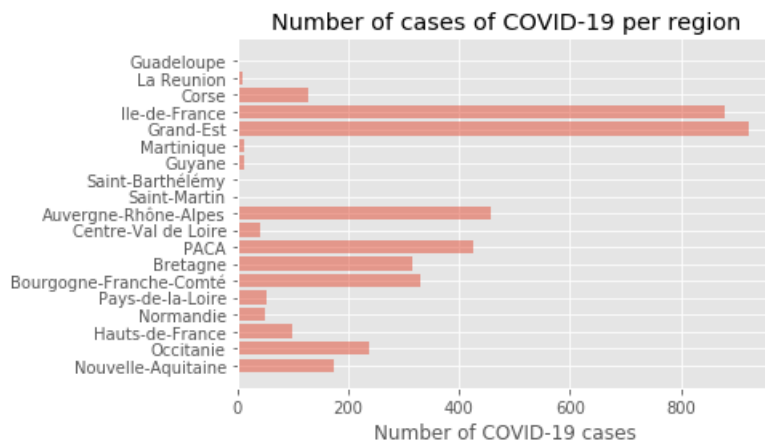

Conclusion

In conclusion, we can see that the region that had the most cases at the start of the pandemic was 'Grand-Est', which is due to a religious gathering in Mulhouse. It is closely followed by 'Ile de France', as it is the capital. Indeed, all the people that went on holidays in Italy, and the ones that from Mulhouse, came back directly to Paris, contaminating other people living in the capital.

Finally, we can say that most cases came in france from Italy and it is mostly due to the February holidays, where most Parisians travel down to Italy to do some skiing. Some of them caught the virus there and brought it back to the capital.

Below are the number of cases per region:

In [63]:

```python
for i in range(len(regions)):
    print(f"{regions[i]}: {cases[i]} cases of COVID-19.")
```

```
Nouvelle-Aquitaine: 174 cases of COVID-19.
Occitanie: 236 cases of COVID-19.
Hauts-de-France: 98 cases of COVID-19.
Normandie: 50 cases of COVID-19.
```

```
Pays-de-la-Loire: 52 cases of COVID-19.
Bourgogne-Franche-Comté: 330 cases of COVID-19.
Bretagne: 314 cases of COVID-19.
PACA: 424 cases of COVID-19.
Centre-Val de Loire: 40 cases of COVID-19.
Auvergne-Rhône-Alpes: 456 cases of COVID-19.
Saint-Martin: 4 cases of COVID-19.
Saint-Barthélémy: 2 cases of COVID-19.
Guyane: 12 cases of COVID-19.
Martinique: 12 cases of COVID-19.
Grand-Est: 922 cases of COVID-19.
Ile-de-France: 880 cases of COVID-19.
Corse: 126 cases of COVID-19.
La Reunion: 10 cases of COVID-19.
Guadeloupe: 2 cases of COVID-19.
```