# Programming for Data Analytics
# Lab 07.01: knock weather

Lecturer: Andrew Beatty

In this lab I want to see if there is any relation between the windspeed and the month in knock

I am going to use the dataset

"https://cli.fusio.net/cli/climate_data/webdata/mly4935.csv"

**WARNING: this is not the dataset I asked you to use for the assignment**

## Get the date

1. Look at the csv file and realise that the first 19 lines do not contain data, so we need to skip them

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt


df = pd.read_csv("https://cli.fusio.net/cli/climate_data/webdata/mly4935.csv",
skiprows=19)
print(df.head(3))
```

2. Is there any correlation between the mean temperature and the month. The answer is no,

```
corrtemp = df["month"].corr(df["meant"])
print(corrtemp)
```

but that is for a number of reasons (why do you think)

### Tidy the wind data

3. If you look at the data you will see that there are some windspeeds missing, we need to drop those rows, lets make a new dataset which only has the month and wind speed.

```
cleandf = df[["month","wdsp"]]
```

4. Great, now lets drop the NAs

```
cleandf.dropna(inplace=True)
```

Arghhhhh that did not work at all, I went down quite a rabbit hole trying to fix this

# pandas does not recognise the '' as an NA so I am converting them to na before droping them
#cleandf['wdsp'].replace('', np.nan, inplace=True)
# that did not work, I spent a while bashing my head to try and get this to work
# for some reason I could not get regex to work
# I searched and found
# https://stackoverflow.com/questions/29314033/drop-rows-containing-empty-cells-from-a-pandas-dataframe
# cleandf[cleandf['wdsp'].str.strip().astype(bool)]

```
cleandf['wdsp']= cleandf.loc[:,('wdsp')].replace(' ', np.nan)
cleandf.dropna(inplace=True)
```

5. Now we just need to convert the wind speed to floats

```
cleandf['wdsp'] = cleandf['wdsp'].astype(float)
```

## Now we can analyse

6. Is there a Correlation

```
corrwind = cleandf["month"].corr(cleandf["wdsp"])
print (f"wind correlation {corrwind}")
```

7. No there does not seem to be, that is strange let's do some regression

```
sns.set_style('whitegrid')
#sns.scatterplot(x='total_bill',y='tip',data=dataset)
sns.lmplot(x='month', y='wdsp', order=3, data=cleandf)
plt.show()
```

8. That's better.