# COMP1204: Data Management
# Coursework One: Hurricane Monitoring

Zhengbin Lu
zl15g22

March 17, 2023

## 1 Introduction

This coursework covers three parts, Unix scripting for file processing, report writing using Latex, and use of Git for version control.

In this work, we're required to process data from the hurricane as a KML file by using Unix script, extract the important information, reformat the data and convert it into a CSV file, resolve the Git conflict by using version control and writing the report in latex format.

This report details how the Unix script work and the plots of hurricane after processing the data. Meanwhile, it contains the usage of resolving conflict.

## 2 Create CSV Script

Here is the detail of the Unix script. It cleans the data from KML file, reformat it and finally outputs into a CSV file.

```
#!/bin/bash

#fetching the arguement of the input and output file
input=$1
output=$2

# Displays the file which being processed, and the output file name.
echo "Converting $input -> $output"

#Output the header line which contains the basic information
# ">" - to initialize the file
echo "Timestamp,Latitude,Longitude,MinSeaLevelPressure,MaxIntensity" > $output

#cat - displays the content of the file
#grep - fetches the line which contains the keyword "UTC,N,mb,knots"
#sed - deletes the lines which contain the useless information by the keywords
#sed - deletes the irrevelent words and notations
#paste - merges every four lines into one line, each line is separated by a space
#awk - prints the column which contains the relevant information
#sed - replaces ";" by a space
#sed - adjusts the format and replaces the unnecessary space between the words
#sed - deletes the space which at the end of the line
#">>" - to append the csv file
cat $input | grep -E 'UTC|N|mb|knots' | sed -E '/<(\/)?dtg|AL|TS|name|Name|Num|<TR
    >|TD/d' | sed -E 's/(]]>)|<(\/)?(B|td|tr|table)>//g' | paste -d ' ' - - - - |
```

```
    awk '{print $1, $2, $3, $4, ",", $5, $6, $7, $8, ",",$9, $10, "," $14, $15}' |
    sed 's/;/ /g' | sed -E 's/\s*,\s*/,/g' | sed 's/[[:space:]]*$//'>> $output

#Output the finish information
echo "Done!"
```

**Listing 1:** CSV Script

# 3 Storm Plots

The plots of three hurricanes, al012020, al102020, and al112020 are as follow:

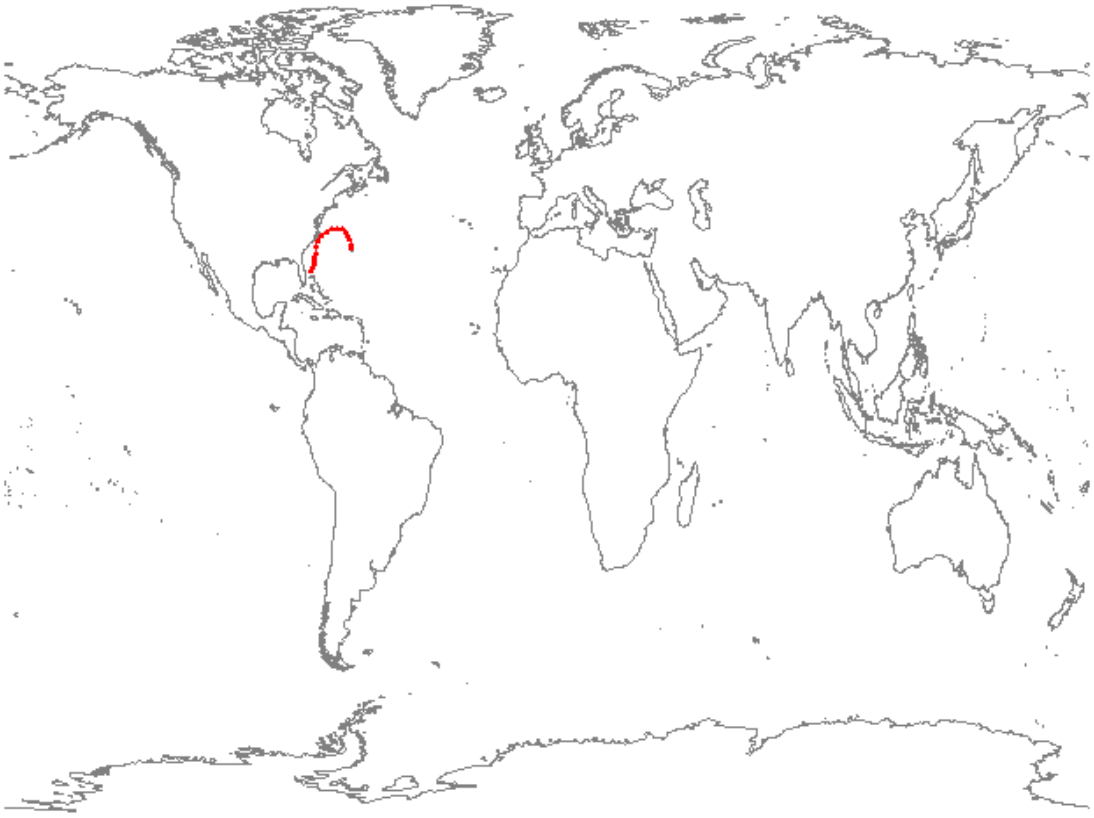

**Figure 1:** Storm plot for al012020.kml

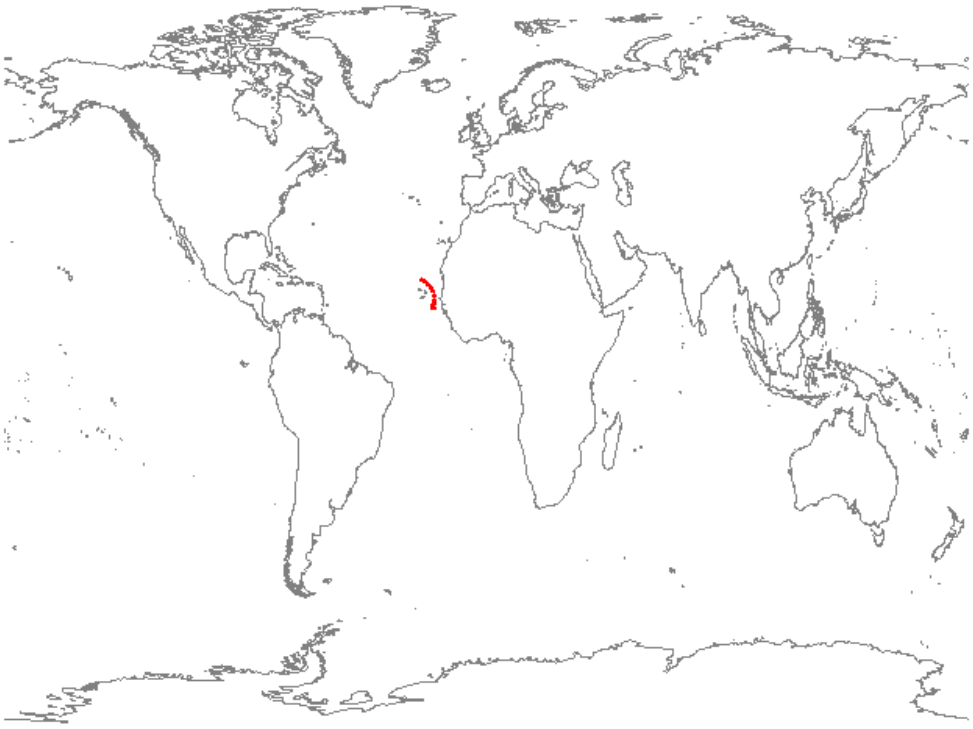

**Figure 2:** Storm plot for al102020.kml

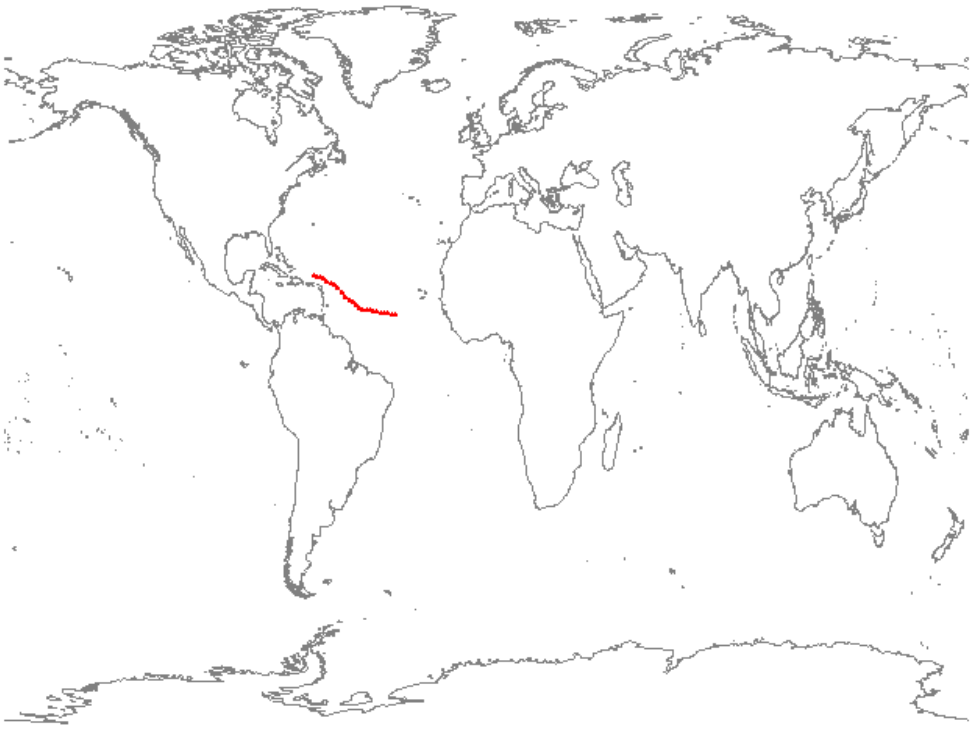

**Figure 3:** Storm plot for al112020.kml

# 4 Git usage

```
import pandas as pd
import matplotlib.pyplot as plt
import os
import glob
import math
user_key = 997

def plot_all_csv_pressure():
    path = os.getcwd()
    csv_files = glob.glob(os.path.join(path, '*.csv'))

    for f in csv_files:
        storm = pd.read_csv(f)
        storm['Pressure'].plot()
        plt.show()
```

**Listing 2:** Branch master

```
import pandas as pd
import os
import glob
import matplotlib.pyplot as plt
user_key= 274

def plot_all_csv_pressure():
    path = os.getcwd()
    csv_files = glob.glob(os.path.join(path, '*.csv'))

    fr f in csv_files:
        storm = pd.read_csv(f)
        storm['Pressure'].plot()
        plt.show()

def plot_all_csv_intensity():
    path = os.getcwd()
    csv_files = glob.glob(os.path.join(path, '*.csv'))

    for f in csv_files:
        storm = pd.read_csv(f)
        storm['Intensity'].plot()
        plt.show()
```

**Listing 3:** Branch python-addon

After executing the command and creating the conflict, I tried to merge the branch by using the 'git merge' command, and there was an error of version conflict. I modified the file by using vim, removed the "user key = 997" from the master branch, and the wrong function which was used to compute the pressure in the branch python-addon. Then I added the "user key = 274" and the "intensity function" from the branch python-addon, reordered the rest of the codes, removed the redundant part of the code, and resolved the conflict. After that, I added the comment by command 'git commit -am ' and pushed it into the remote repository by the command 'git push'.

The code after conflict resolving as follow:

```
import pandas as pd
import matplotlib.pyplot as plt
import os
```

```
import glob
import math
user_key = 997
user_key= 274

def plot_all_csv_pressure():
    path = os.getcwd()
    csv_files = glob.glob(os.path.join(path, '*.csv'))

    for f in csv_files:
        storm = pd.read_csv(f)
        storm['Pressure'].plot()
        plt.show()

def plot_all_csv_intensity():
    path = os.getcwd()
    csv_files = glob.glob(os.path.join(path, '*.csv'))

    for f in csv_files:
        storm = pd.read_csv(f)
        storm['Intensity'].plot()
        plt.show()
```

**Listing 4:** Script after conflict resolving