

COMP1204: Data Management

Coursework One: Hurricane Monitoring

Daniel Braghis
32161204

March 8, 2022

1 Introduction

This coursework covers three key topics: Unix, Latex, and Git. The coursework is divided into three parts: Unix scripting for file processing, report writing using Latex and use of Git for version control. We have been provided data about storms in KML format and were tasked with extracting the data of interest and converting it to CSV format for later processing.

This document details how the bash script that accomplishes the task works and shows the resulting plots of the storm data refined with the script. Also, in the process, this report aims to demonstrate the use of LaTeX for report writing.

2 Create CSV Script

Below is the full script used to retrieve the data from KML files with comments explaining each command's purpose. The script is split in 3 main parts: cleaning the data, organizing it in the correct format and exporting the result to the CSV file with the correct column headers.

```
1 #!/bin/bash
2
3 # Getting arguments for input and output file
4 # Variable for regex patterns for later use delimited by '|'
5 INPUT="$1"
6 OUTPUT="$2"
7 REGEX=".*UTC.*\|.*<N>.*\|.*<S>.*\|.*<W>.*\|.*<E>.*\|.*knots.*\|.*mb"
8
9 # Conversion information and starting progress bar
10 echo "Converting ${INPUT} -> ${OUTPUT}..."
11 echo -ne '[          ] (0%)\r'
12 sleep 0.3
13
14 # Cleaning up the input
15 # cat - prints contents of file
16 # sed - removes the html style tags
17 # awk - assignment in awk removes leading and trailing spaces/tabs
18 # grep - prints the lines that match the regex expressions
19 # sed - removes the leading ', ' for West/East elements
20 echo -ne '[#####          ] (33%)\r'
21 CLEAN=$(sed -e 's/<[^>]*>//g' ${INPUT} | awk '{ $1=$1 }; {print}' | grep -o $
    {REGEX} | sed -e 's/, //g')
22 sleep 0.3
```

```

23
24 # Formatting the output
25 # echo - get the contents of the cleaned input
26 # sed - remove the 2 duplicate dates from every 7 line block
27 # tr - replaces new lines with commas
28 # sed - puts the new line back to create rows of 5 values of interest
29 echo -ne '[#####
                ] (66%)\r'
30 RESULT=$( echo "${CLEAN}" | sed -n '3~7p;4~7p;5~7p;6~7p;7~7p' | tr "\n" " ,
        " | sed 's/,/\n/5;P;D')
31 sleep 0.3
32
33 # Append labels and output the result to csv
34 echo -ne '[#####
                ] (100%)\r'
35 APPEND=$'Timestamp,Latitude,Longitude,MinSeaLevelPressure,MaxIntensity\n'
36 echo "${APPEND}${RESULT}" > "${OUTPUT}"
37
38 # Stop the progress bar
39 echo -ne '\n'

```

Listing 1: CSV Script

3 Storm Plots

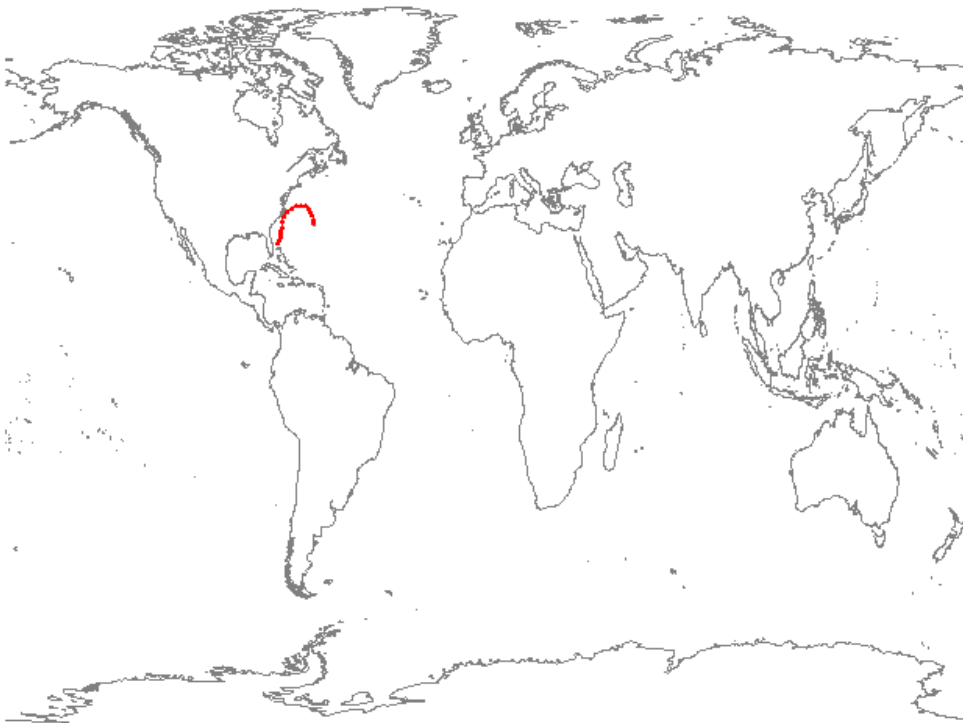


Figure 1: Storm plot for al012020.kml



Figure 2: Storm plot for al102020.kml

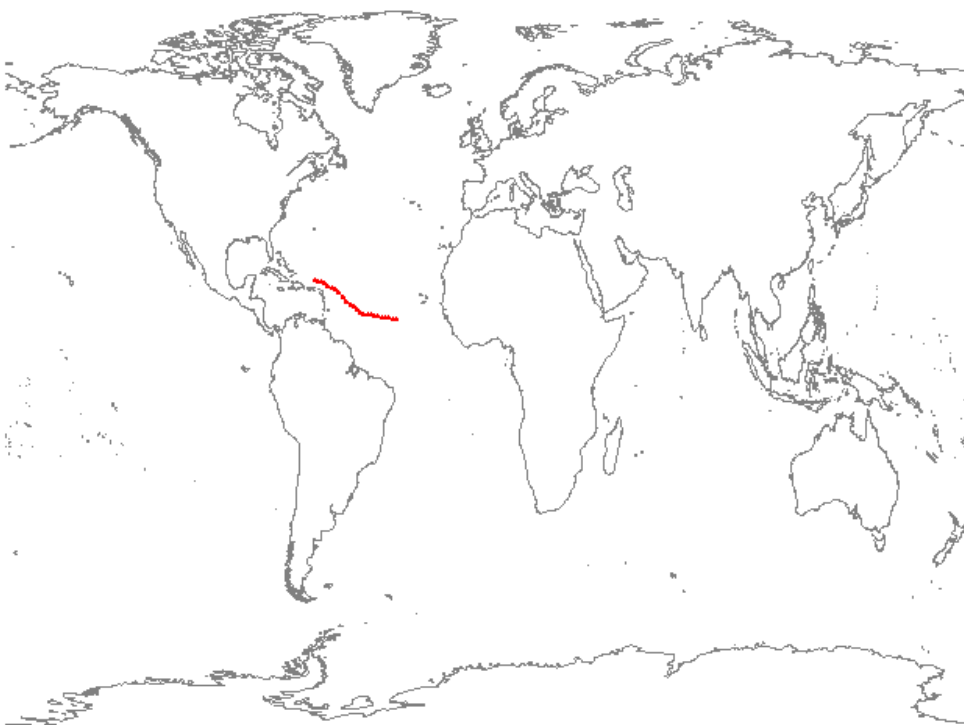


Figure 3: Storm plot for al112020.kml



Figure 4: Storm plot for al122020.kml

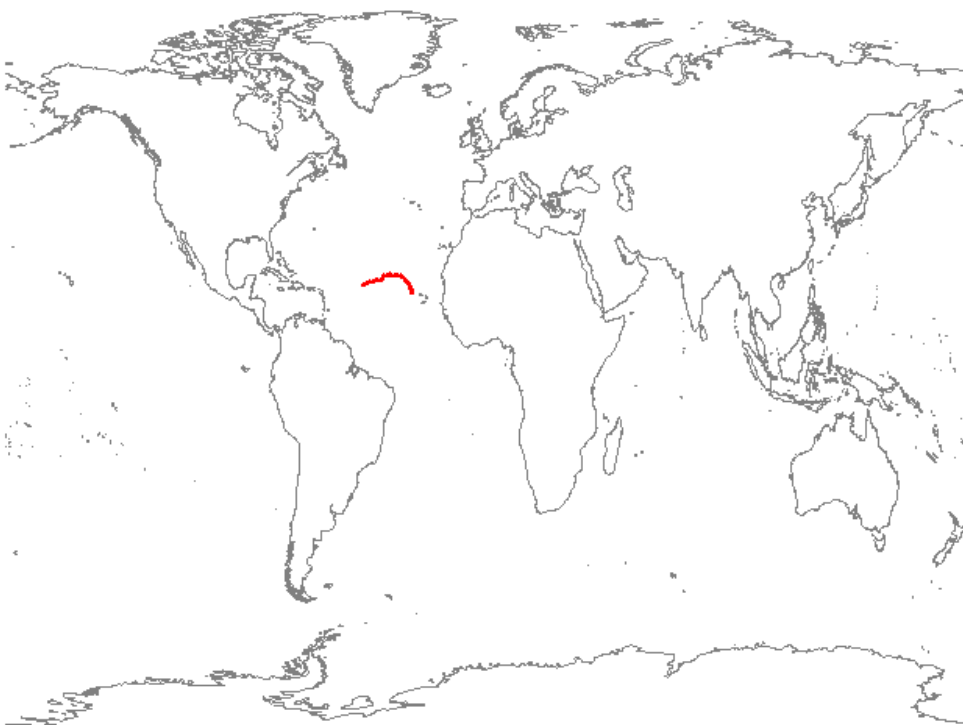


Figure 5: Storm plot for al212020.kml