COMP1204: Data Management
Coursework One: Hurricane Monitoring

Akhil Maram

32565887

February 28, 2022

1   Introduction

The aim of this coursework is to create a bash script capable of extracting useful data from hurricane reports. Hurricane reports are given as kml files and the bash script must extract the timestamps, latitude, longitude, minimum sea level pressure, and maximum intensity. This data is then presented in a csv file. The csv file is then processed through a different bash script to display the trajectory of the hurricane over the given timeframe.

2   Create CSV Script

Bash Script 1: Pseudocode for create_csv.sh

```bash
1  #!/bin/bash
2
3  #Extract all timestamps and add them into timestamp.txt
4  cat $1 | grep 'UTC' | sed -e '/<dtg>/d' -e '/tr/d' -e 's/<name>//g' -e 's/<\/name>//g' |
   ↪  awk '{print $1 " " $2 " " $3 " " $4}' > timestamp.txt
5  echo 'Timestamps collected'
6
7  #Extract all latitudes and add them with units into lat.txt
8  cat $1 | grep '<lat>' | sed -e '/tr/d' -e 's/<lat>//g' -e 's/<\/lat>//g' | awk '{print $1
   ↪  " N"}' > lat.txt
9  echo 'Latitudes collected'
10
11 #Extract all longitudes and add them with units into long.txt
12 cat $1 | grep '<lon>' | sed -e '/tr/d' -e 's/<lon>//g' -e 's/<\/lon>//g' | awk '{print $1
   ↪  " W"}' > long.txt
13 echo 'Longitudes collected'
14
15 #Extract minSeaLevelPressure with units and add them into minSeaLevelPres.txt
16 cat $1 | grep 'minSeaLevelPres' | sed -e 's/<minSeaLevelPres>//g' -e
   ↪  's/<\/minSeaLevelPres>//g' | awk '{print $1 " mb"}' > minSeaLevelPres.txt
17 echo 'MinSeaLevelPressures collected'
18
19 #Extract maxIntensity with units and add them into maxIntensity.txt
20 cat $1 | grep 'knots' | sed 's/<tr><td>//g' | awk '{print $1 " "  $2}' | sed 's/;//g' >
   ↪  maxIntensity.txt
21 echo 'MaxIntensities collected'
22
23 echo 'Converting' $1 '-> ' $2
24
25 #Merge timestamp.txt, lat.txt, long.txt, minSeaLevelPres.txt, and maxIntensity.txt
   ↪  with',' and result stored in csv file
26 paste -d , timestamp.txt lat.txt long.txt minSeaLevelPres.txt maxIntensity.txt > $2
27
28 #Add the appropriate heading to newly created csv file
29 sed -i '1 i Timestamp,Latitude,Longitude,MinSeaLevelPressure,MaxIntensity' $2
30
31 #Remove unnecessary files
32 rm timestamp.txt lat.txt long.txt minSeaLevelPres.txt maxIntensity.txt
33
34 echo 'Done!'
```

The pseudocode above takes 2 arguments: the first argument is the name of the kml file to process and the second argument is the name of the csv file storing the extracted information.

Line 4 uses grep to extract lines containing 'UTC' and removes all lines with the tag '<dtg >'. The name tags are then removed with sed and awk prints out each line again to remove indentations. The final set of data includes all the timestamps, which are then added into a txt file called 'timestamp.txt'.

Line 8 uses grep to search for all lines with the '<lat >' tag. Sed removes line with the '/tr' tag and removes the latitude tags. The direction is then appended onto each line (North) using awk and the final set of data is added into a txt file called 'lat.txt'.

Line 12 uses grep to search for all lines with the '<lon >' tag. Sed removes line with the '/tr' tag and removes the longitude tags. The direction is then appended onto each line (West) and the final set of data is added into a txt file called 'long.txt'.

Line 16 uses grep to search for all lines with the phrase 'minSeaLevelPres' and sed then removes minSeaLevelPres tags. This expresses the minSeaLevelPressure in mb and Hg. To report the pressure in mb, awk is used to extract the first number on each line and append the units 'mb'. The final set of pressures is then added to a txt file called 'minSeaLevelPres.txt'.

Line 20 uses grep to search for all lines with the units 'knots' as the max velocity of the winds has to be reported in knots. Sed is then used to remove all the '<tr ><tg >' tags. Awk then takes the first 2 values on each line to present all the maxIntensities with the relevant units. This set of data is added into a txt file called 'maxIntensity.txt'.

At this stage, there are 5 txt files each with different information. The paste command on line 26 takes all 5 files and merges them with the comma. The resulting file is then added into a csv file specified as the second argument on the command line.

The sed command on line 29 adds the appropriate column headings in the csv file.

As the csv file now has all the required information, the command on line 32 deletes the 5 temporary txt files.

All echo statements in the pseudocode print messages to the command line for debugging purposes.

3 Storm Plots

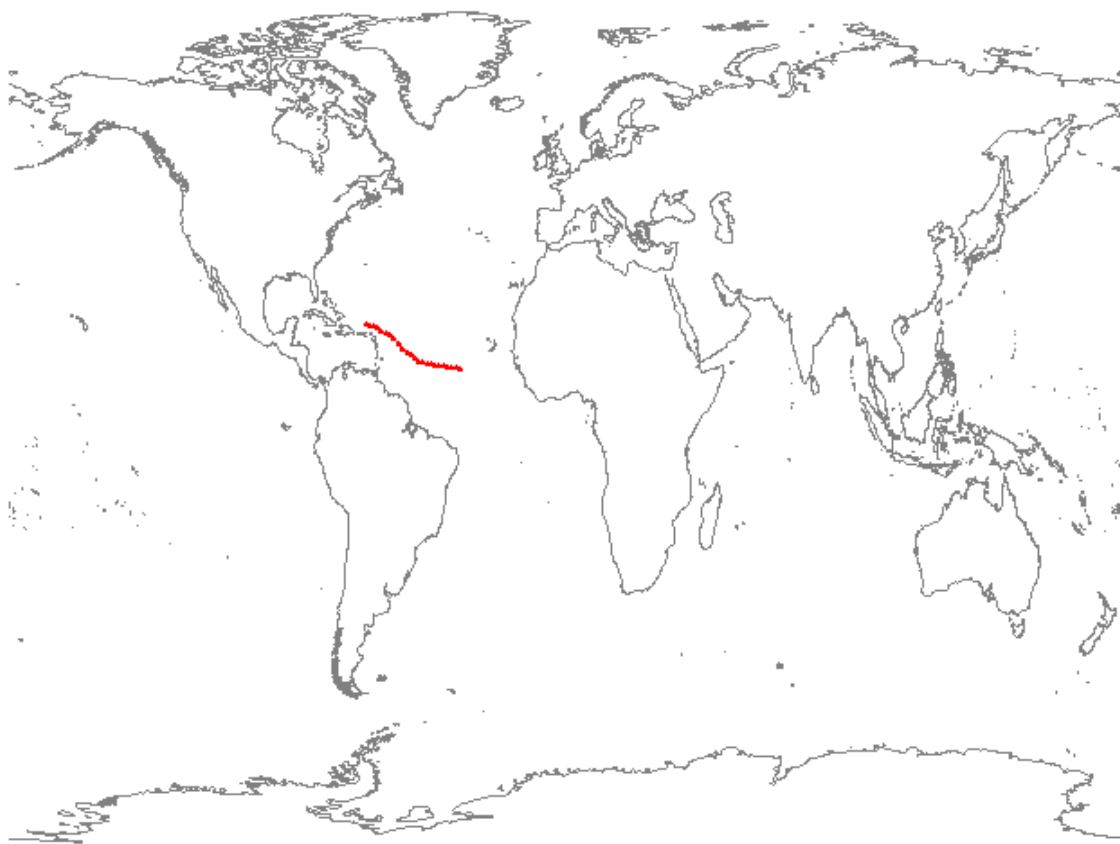Figure 1: Plot of hurricane trajectory from al102020.kml
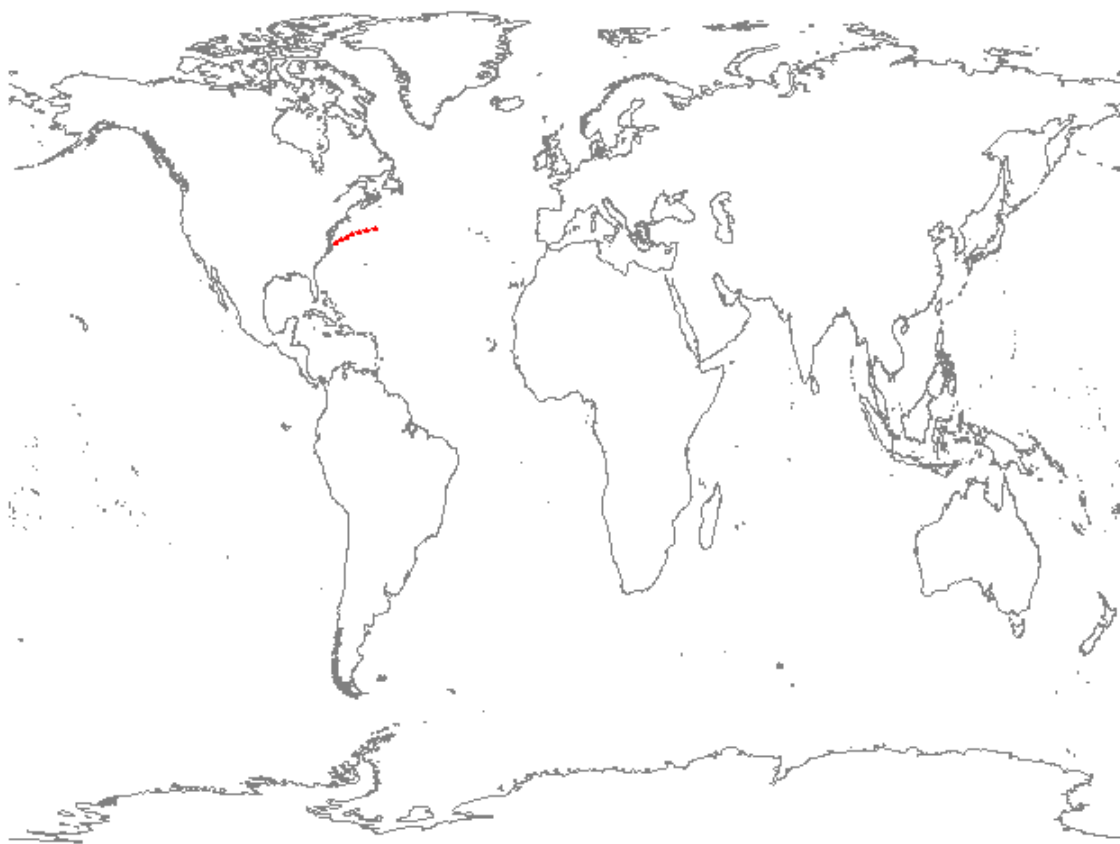
Figure 2: Plot of hurricane trajectory from al112020.kml

Figure 3: Plot of hurricane trajectory from al122020.kml