Exploratory analysis of Atlantic hurricanes

Background

Hurdat2 is a database of hurricanes and similar storms produced by the US National Oceanic and Atmospheric Administration (NOAA). It contains data on

- the track each storm followed (latitude-longitude coordinates)
- the wind speed in knots (1 knot = 0.52 m/s) and the classification of the storm at each point along its track.
- the point at which each storm crossed onto land if it did cross onto land ("landfall")

and a number of other variables, all organised by a unique ID code for each storm and sometimes a name. Databases like this are crucial for understanding how climate change has affected patterns of extreme weather and predicting how the choices we make about carbon emissions shape extreme weather in the future.

The original database is stored in a somewhat awkward format, but I have tidied it and you will find the full dataset for the Atlantic Ocean region in the dataframe hurrs inside the file Hurdata_tidy.RData.

The documentation from NOAA is at https://www.nhc.noaa.gov/data/hurdat/hurdat2-format-nov2019.pdf
Background information on the Hurdat2 project and hurricane science in general are at https://www.noaa.gov/education/resource-collections/weather-atmosphere/hurricanes https://www.aoml.noaa.gov/hrd/hurdat/Data_Storm.html

I have removed some variables and simplified the classification system. The data set can be found in "/processed_data" we will use the word "storm" to mean "anything with a unique ID in the database" and the column is Hurricane is TRUE/FALSE depending on whether the storm system had reached hurricane status (based on wind speed) at a given point in time.

Question 1: What's in this dataset?

(a) Load the dataframe. Pick one famous hurricane that you know the name and make a dataframe similar to hurrs but containing only the data for your chosen hurricane. (Note that the people who name hurricanes have reused names sometimes.)

(b) Write a function called Hurdat2_summary that takes the hurrs dataframe as an input and returns a **named list** containing

- the total number of storms
- the range of years covered
- a vector containing the number of track points for each storm

Test your function by running it for three cases:

- the full hurrs dataframe
- a version that has been filtered to contain only storms from a single year
- the one-hurricane dataframe you made in (a)

(c) Make a histogram of the number of track points, using the output of your function for the hurrs case.

(d) What year did they begin naming hurricanes and other storms? (Answer this with R code and the hurrs dataframe–you can check if you're right using Google).

Question 2: Mapping

Here is some code that plots all the storms in the database atop a simple world map.

```
# you might need to run # install.packages("maps") # and
restart R to make this work. hurricaneBasemap = function()
{ library(maps) world = map_data("world") basemap =
ggplot() + coord_fixed(xlim=c(-130,30),ylim=c(0,90)) +
    geom_polygon(data=world, aes(x=long,y=lat,group=group), color="gray50", fill="white")
  return(basemap)
} hurricaneBasemap() + geom_path(data=hurrs,
aes(x=longitude,y=latitude,group=id,color=windspeed)) +
scale_color_distiller(type="seq",direction=1,palette="YlOrRd")
```

a. Figure out how it works and add a helpful one-line comment to each of the 9 lines of code.
b. Now write a function that takes a year as input and makes a similar plot for only the hurricanes during that year. The function doesn't need to return a value, but it does need to generate a plot. Test your function by making the plot for a year of your choice.

Question 3: The local perspective (

world.cities in the maps package gives information on lots of world cities including their latitude-longitude (lat-long) coordinates. You can access this data with the lines

```
library(maps) cities =
world.cities
```

Now cities is a dataframe in your Environment.

a. The formula for calculating distance *D* in km from latitude-longitude coordinates is

$$\mathrm{D} = 111.325\sqrt{dx^2 + dy^2}$$

where

$$\mathrm{dx} = (x_2 - x_1)\cos\left(\frac{\pi}{180}\frac{y_1+y_2}{2}\right)$$

$$\mathrm{dy} = y_2 - y_1$$

and *x* is longitude and *y* latitude, in degrees. (The $\pi/180$ is there to convert from degrees to radians.) In other words, one degree of latitude is always 111.325 km, but the number of km in one degree of longitude varies as the cosine of the mean latitude.

Write a function that calculates distance between (lon1,lat1) and (lon2,lat2). Test it.

b. Adapt your function so that instead of (lon2,lat2) it takes the name of a city as input, finds its latitude– longitude coordinates, and returns the distance to (lat1,lon1). Test it. Give this function a new name and have it call the function from part (a).

c. A common mistake in working with latitudes and longitudes is to swap them around. It isn't always possible to catch user errors of this sort, but one test is that longitude can take values from -180 to 180 while latitude is always -90 to 90. Modify your function to check whether it's likely that the user has swapped lat1 and lon1 and if so, provide a useful warning message.
d. Choose an eastern North American or Caribbean city, and use your function to determine which storms passed within 100 km of it. Plot those storms on a map like those in Q1, with a dot on top showing where the city is.

Question 4: Climate change

a. Globally, the number of hurricane-strength storms has been clearly increasing because of climate change. Is this true in the Atlantic specifically? Plot the number of storms in each year.
b. The total number isn't the only aspect of hurricane statistics that might change with climate change. What about the geographical area affected? Make a plot that lets you visually evaluate whether the latitude range where storms cross the coastline (make landfall) is expanding in recent decades. Does your plot suggest any concerns about data quality and data coverage in this database?
c. Is peak wind speed increasing over time? Answer this using whatever kind of plot or table you think is most helpful.
d. What about the seasonal timing of hurricanes? First, make a histogram of storms by month over the entire dataset. This should show that there is a distinct hurricane season—they tend to happen at a certain time of year. Use a bin width of 1 month.
e. Next, add a "decade" column to the hurrs dataframe. (You can calculate decade from year in one line of code: Google it if you have never seen this rounding trick.)
f. Finally, make a plot that shows whether "hurricane season" is expanding into earlier and later months of the year over time.