

DELIVERABLE 1

Due Wednesday, November 9th at 2:59 a.m.

This is an individual deliverable (see syllabus)

As tropical cyclones make their way over the Atlantic and Pacific oceans during hurricane season, having an accurate forecast ahead of time is crucial in providing decision-makers the opportunity to undertake life-saving actions. Throughout your Masters program, you will learn various predictive tools that are useful in a wide variety of situations. In this deliverable, we focus on descriptive analytics pertaining to past tropical cyclones.

A tropical cyclone is a storm characterized by strong winds and heavy rain. Tropical cyclones originate from the oceans in tropical areas and grow bigger as they move. Some tropical cyclones make landfall (i.e., they touch land) and they are very closely monitored as they can be life-threatening and destructive.

A tropical cyclone is classified into one of 7 categories based on its wind speed (see table below):

- A category 0 tropical cyclone is also called a tropical depression
- A category 1 tropical cyclone is also called a tropical storm
- A tropical cyclone in categories 2-6 is called a hurricane

<i>Category</i>	<i>Name</i>	<i>Wind speed (in knots)</i>
0	Tropical Depression	0-34
1	Tropical Storm	35-63
2	Hurricane	64-82
3	Hurricane	83-95
4	Hurricane	96-113
5	Hurricane	114-135
6	Hurricane	>= 136

Tropical cyclones originate from different 'ocean basins' around the world. In this midterm, we focus on those originating from the Eastern Pacific and North Atlantic basins, since these are of particular interest to the United States.

The Data

The file <**tropical_cyclones.csv**> consists of 269 records corresponding to almost all tropical cyclones that made landfall in North America between 1980 and September 2020. Of these, 85 records are cyclones that originated in the Eastern Pacific basin and the remaining 184 records are cyclones that originated in the North Atlantic basin. Tropical cyclones that did not make landfall are not included in this dataset.

The dataset has the following 14 columns:

Column name	Description
Year	year the tropical cyclone originated
Month	month the tropical cyclone originated
Name	name assigned to the tropical storm; if none were assigned the entry is "NOT_NAMED"
Basin	North Atlantic basin or Eastern Pacific basin
distance_48	distance in kilometers (km) from landfall location 48 hours prior to landfall
distance_24	distance in kilometers (km) from landfall location 24 hours prior to landfall
center_speed_48	speed of the center of the tropical cyclone in knots 48 hours prior to landfall (note: think of it as how fast the cyclone is moving)
center_speed_24	speed of the center of the tropical cyclone in knots 24 hours prior to landfall (note: think of it as how fast the cyclone is moving)
wind_speed_48	maximum sustained wind speed of the tropical cyclone in knots 48 hours prior to landfall
wind_speed_24	maximum sustained wind speed in the tropical cyclone in knots 24 hours prior to landfall
wind_speed_landfall	maximum sustained wind speed in the tropical cyclone in knots at landfall
category_landfall	category of the tropical cyclone at landfall (0, 1, 2, 3, 4, 5, or 6) (see table on page 2)
hurricane_landfall	0 if the tropical cyclone was category 0 or 1 at landfall; 1 if the tropical cyclone was category 2, 3, 4, 5, or 6 at landfall
damage	estimate of the financial damage caused by the tropical cyclone in billions of dollars; note that these estimates are available for only <i>some</i> tropical cyclones. When data is missing, it indicates that the damage was smaller than \$1billion or that the tropical cyclone was too recent to provide an estimate of the damages.

Use R to answer the following questions:

1. Which month of the year has the highest number of tropical cyclones making landfall?
2. Compare the tropical cyclone activity across the North Atlantic and the Eastern Pacific basins:
 - i. Which of the two basins has, on average, the more severe tropical cyclones as measured by the maximum sustained wind speed at landfall?
 - ii. Which of the two basins has the higher proportion of hurricanes among all tropical cyclones originating in that basin (a hurricane is indicated by 1 in the hurricane_landfall column)?
3. Assume that the maximum sustained wind speed at landfall of a tropical cyclone follows a normal distribution:
 - i. Estimate and report the mean and standard deviation of that distribution across all tropical cyclones in the dataset.
 - ii. Using the mean and standard deviation you computed in (i), calculate the probability that the maximum sustained wind speed at landfall is less than or equal to 34 knots (that is, the cyclone is a tropical depression)? How does this compare with the actual proportion of tropical depressions in the dataset?
4. Your dataset contains a damage column that represents the total damage (in \$billions) caused by a tropical cyclone when it made landfall. We provide below an R command and output that calculates the average of that column by basin.

```
> summarize(group_by(df, basin), av_damage = mean(damage, na.rm = TRUE))
# A tibble: 2 x 2
  basin          av_damage
  <chr>          <dbl>
1 Eastern Pacific    2.45
2 North Atlantic   21.0
```

One interpretation of this output is that cyclones from the North Atlantic basin are approximately ten times stronger than cyclones from the Eastern Pacific basin. What are some other reasons that possibly explain the difference in average damage observed in this output?

5. To understand the potential impact of climate change on tropical cyclone activity it may be useful to look at the evolution of a few metrics over the past 40 years. Answer the following questions based only on the data we have provided.
 - i. Has the annual number of tropical cyclones making landfall been changing over the past 40 years? There are 3 possibilities: the annual number is increasing, there is no significant change, or it is decreasing. Identify which possibility is best supported by the data and explain why.
 - ii. Has the average maximum sustained wind speed at landfall (that is, annual average of wind_speed_landfall) of tropical cyclones been changing over the past 40 years? There are 3 possibilities: the average maximum sustained wind speed at landfall is increasing, there is no significant change, or it is decreasing. Identify which of possibility is best supported by the data and explain why.