[R] EDA

## Basic Library

Just to get started with the [R]EDA assignment, We are going to include the Basic Libraries we required for it. Which is “tidyverse” in our case.

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

## Dataset Storms

Now, We have used the dataset named “Storms” which comes under the package of “dplyr”. Now, only the data sets are in the package of “dataset” are not converted into the tibble form, but other than that all the datasets are already in the tibble form, just like my “storms”. So, lets add the dataset in one variable, even though we can use it without assigning it to same variable, we should have good practice for it, in case we want to change the variable name.

And the dataset has 10010 observations in the 13 different Variables.

storms <- storms

## Questions.

Here are some of the questions, that we are going to get answers form the dataset storms, and visualize them. 1. Are storms getting more frequent by the year? 2. Are storms getting stronger by the year? 3. What months see the most number of storms?

## Info of Missing values.

Now, here I performing the checks to determine the quality of my “storms” data.

storms %>% summarise(Total\_Count = n())

## # A tibble: 1 x 1  
## Total\_Count  
## <int>  
## 1 10010

filter(storms, is.na(ts\_diameter) | is.na(hu\_diameter)) %>%  
 summarise(Missing\_Count = n())

## # A tibble: 1 x 1  
## Missing\_Count  
## <int>  
## 1 6528

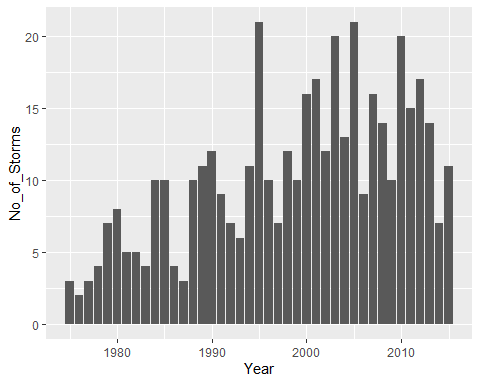
As you can see only 3482 rows have the ts\_diameter and hu\_diameter records, other 6528 rows have that column value missing, thus those variables might not be of much use when we have to find something or have to make conclusion depending on columns data.

## Answer 1

(storm\_freq <- group\_by(storms, Year = year) %>%  
 distinct(name) %>%  
 summarise(No\_of\_Storms = n()))

## # A tibble: 41 x 2  
## Year No\_of\_Storms  
## <dbl> <int>  
## 1 1975 3  
## 2 1976 2  
## 3 1977 3  
## 4 1978 4  
## 5 1979 7  
## 6 1980 8  
## 7 1981 5  
## 8 1982 5  
## 9 1983 4  
## 10 1984 10  
## # ... with 31 more rows

ggplot(data = storm\_freq) +  
 geom\_col(mapping = aes(x = Year, y = No\_of\_Storms))



**Conclusion 1**: Yes, the Storms are getting more frequent by the year.

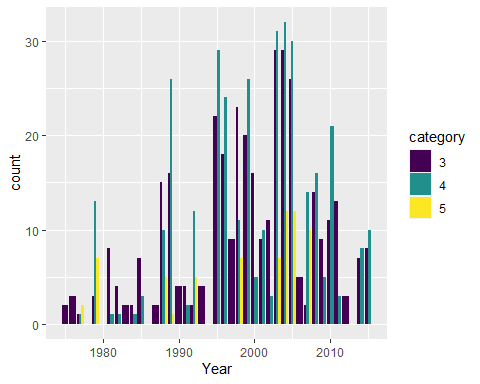
The above trend does show increase in the number of storms per year. As normal observer we can see that this data is distributed from 1975 to 2015, as shown in the graph by the X axis, and on the y axis we have No\_of\_Storms, which seems rising up as we move along X axis bar.

## Answer 2

(storm\_cat <- filter(storms, category > 2) %>% group\_by(Year = year))

## # A tibble: 779 x 14  
## # Groups: Year [36]  
## name year month day hour lat long status category wind pressure  
## <chr> <dbl> <dbl> <int> <dbl> <dbl> <dbl> <chr> <ord> <int> <int>  
## 1 Caroline 1975 8 31 0 24 -97 hurricane 3 100 973  
## 2 Caroline 1975 8 31 6 24.1 -97.5 hurricane 3 100 963  
## 3 Belle 1976 8 8 18 29.5 -75.3 hurricane 3 100 958  
## 4 Belle 1976 8 9 0 30.9 -75.3 hurricane 3 105 957  
## 5 Belle 1976 8 9 6 32.5 -75.2 hurricane 3 105 959  
## 6 Anita 1977 9 1 18 25.2 -95.5 hurricane 3 110 945  
## 7 Anita 1977 9 2 0 24.6 -96.2 hurricane 5 140 931  
## 8 Anita 1977 9 2 6 24.2 -97.1 hurricane 5 150 926  
## 9 Anita 1977 9 2 12 23.7 -98 hurricane 4 120 940  
## 10 David 1979 8 28 0 12.2 -52.9 hurricane 4 115 947  
## # ... with 769 more rows, and 3 more variables: ts\_diameter <dbl>,  
## # hu\_diameter <dbl>, Year <dbl>

ggplot(data = storm\_cat) +  
 geom\_bar(mapping = aes(x = Year, fill = category), position = "dodge")

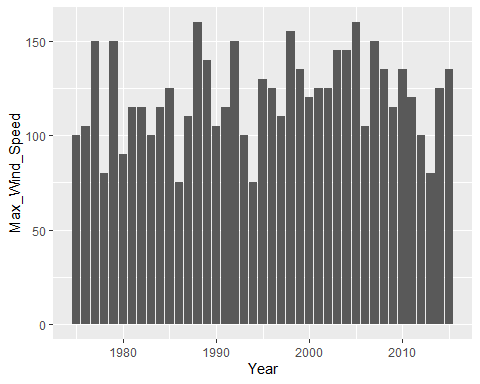


Above graph does show higher number of category 3 and above storms per year. And, yes storms are getting stronger by the year, from given data set you can see that by watching high frequency of higher category number increasing as we move to recent years.

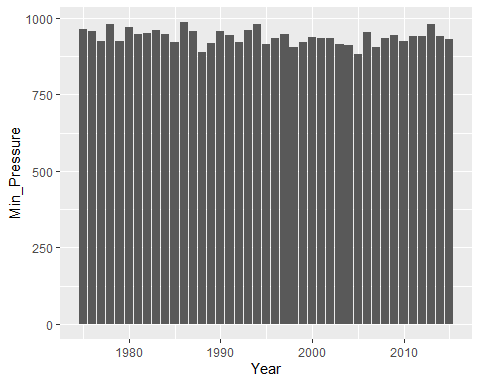
(storm\_wind <- storms %>%  
 group\_by(Year = year) %>%  
 summarise(  
 Max\_Wind\_Speed = max(wind),  
 Min\_Pressure = min(pressure)  
 )  
)

## # A tibble: 41 x 3  
## Year Max\_Wind\_Speed Min\_Pressure  
## <dbl> <int> <int>  
## 1 1975 100 963  
## 2 1976 105 957  
## 3 1977 150 926  
## 4 1978 80 980  
## 5 1979 150 924  
## 6 1980 90 970  
## 7 1981 115 946  
## 8 1982 115 950  
## 9 1983 100 962  
## 10 1984 115 949  
## # ... with 31 more rows

ggplot(data = storm\_wind) +   
 geom\_col(mapping = aes(x = Year, y = Max\_Wind\_Speed))



ggplot(data = storm\_wind) +   
 geom\_col(mapping = aes(x = Year, y = Min\_Pressure))



**Conclusion 2**: From the above 3 graphs, the numbers of stronger storms are increasing, but the storms themselves cannot be conclusively said to be getting stronger.

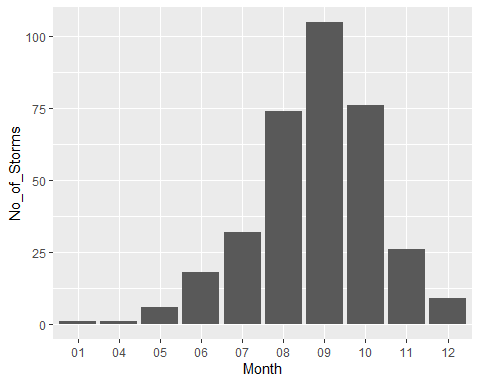
This dataset is from NHC (national Hurricane Center) which is the part of the National Oceanic Atmospheric Administration (NOAA), that says that Category on the dataset which is Saffir-Simpson storm category which is estimated from wind speed, in our dataset it is from (-1 to 5), so more the number stronger the stronger the wind speed.

## Answer 3

(storm\_month <- storms %>%   
 mutate(  
 mon = if\_else(  
 month < 10,   
 str\_pad(month, width = 2, side = "left", pad = "0"),   
 paste(month)  
 )  
 ) %>%   
 group\_by(Month = mon) %>%   
 distinct(name) %>%   
 summarise(No\_of\_Storms = n())  
)

## # A tibble: 10 x 2  
## Month No\_of\_Storms  
## <chr> <int>  
## 1 01 1  
## 2 04 1  
## 3 05 6  
## 4 06 18  
## 5 07 32  
## 6 08 74  
## 7 09 105  
## 8 10 76  
## 9 11 26  
## 10 12 9

ggplot(data = storm\_month) +  
 geom\_col(mapping = aes(x = Month, y = No\_of\_Storms))



**Conclusion 3**: From the above graph, the Atlantic monsoon months, i.e., from July to November, see high number of tropical storms while winter months see almost none.