Hurricane Analysis

Yaniv Bronshtein and Venkata Raju Datla

9/22/2021

Import the necessary libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(gridExtra) #For arranging ggplots
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

Read in the hurricane data manually including both Category 4 and Category 5 Hurricane Data

```
path = "C:/Users/rajuk/OneDrive/Desktop/Class work/02 Fall 2021/03 Prob and stats/01 Home work/01 Hurri

hurricane_df <- read_csv("hurricane_data.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Hurricane = col_character(),
##   Wind = col_double(),
##   Year = col_double(),
##   Category = col_double()
## )
```

Create a calculated column for the decade to get per-decade data

```
hurricane_df['decade']=as.numeric(substr(hurricane_df$Year,1,3))*10
```

Split the hurricane_df into category 4 and category 5 data. In addition, generate the by decade median windspeed dataframes

```
cat4_df <- hurricane_df %>%
  filter(Category == 4)

cat5_df <- hurricane_df %>%
  filter(Category == 5)

cat_4df_by_decade <- cat4_df %>%
  group_by(decade) %>%
  summarise(Median_Wind = median(Wind))

cat_5df_by_decade <- cat5_df %>%
  group_by(decade) %>%
  summarise(Median_Wind = median(Wind))
```

Create a function to generate scatter plots by year or by decade

```
get_windspeed_scatter <- function(data, title, by_year) {
  if(by_year) {
    return(
      ggplot(data=data, mapping = aes(x = Year, y = Wind)) +
        geom_point() +
        ggtitle(title) +
        xlab("Year") +
        ylab("WindSpeed (km/h)") #+
    )

  }
  else {
    return(ggplot(data=data, mapping = aes(x = decade, y = Median_Wind)) +
             geom_point() +
             ggtitle(title) +
             xlab("Decade") +
             ylab("Median Wind Speed (km/h)") #+
    )

  }
}
```

Let's see some summary statistics first for the data

```
summary(cat4_df)
```

```
##   Hurricane              Wind            Year         Category     decade
##  Length:127         Min.   :215.0   Min.   :1853   Min.   :4   Min.   :1850
##  Class :character   1st Qu.:220.0   1st Qu.:1933   1st Qu.:4   1st Qu.:1930
##  Mode  :character   Median :220.0   Median :1963   Median :4   Median :1960
##                     Mean   :227.3   Mean   :1963   Mean   :4   Mean   :1959
##                     3rd Qu.:240.0   3rd Qu.:2000   3rd Qu.:4   3rd Qu.:2000
##                     Max.   :250.0   Max.   :2021   Max.   :4   Max.   :2020
```

```
summary(cat5_df)
```

```
##   Hurricane              Wind            Year         Category     decade
##  Length:36          Min.   :260.0   Min.   :1924   Min.   :5   Min.   :1920
##  Class :character   1st Qu.:260.0   1st Qu.:1954   1st Qu.:5   1st Qu.:1950
##  Mode  :character   Median :275.0   Median :1984   Median :5   Median :1980
```

```
##                     Mean   :274.2   Mean   :1979   Mean   :5   Mean   :1974
##                     3rd Qu.:280.0   3rd Qu.:2005   3rd Qu.:5   3rd Qu.:2000
##                     Max.   :305.0   Max.   :2019   Max.   :5   Max.   :2010
```

```
summary(cat_4df_by_decade)
```

```
##      decade      Median_Wind
##  Min.   :1850   Min.   :215.0
##  1st Qu.:1892   1st Qu.:220.0
##  Median :1935   Median :223.8
##  Mean   :1935   Mean   :226.0
##  3rd Qu.:1978   3rd Qu.:230.0
##  Max.   :2020   Max.   :240.0
```

```
summary(cat_5df_by_decade)
```

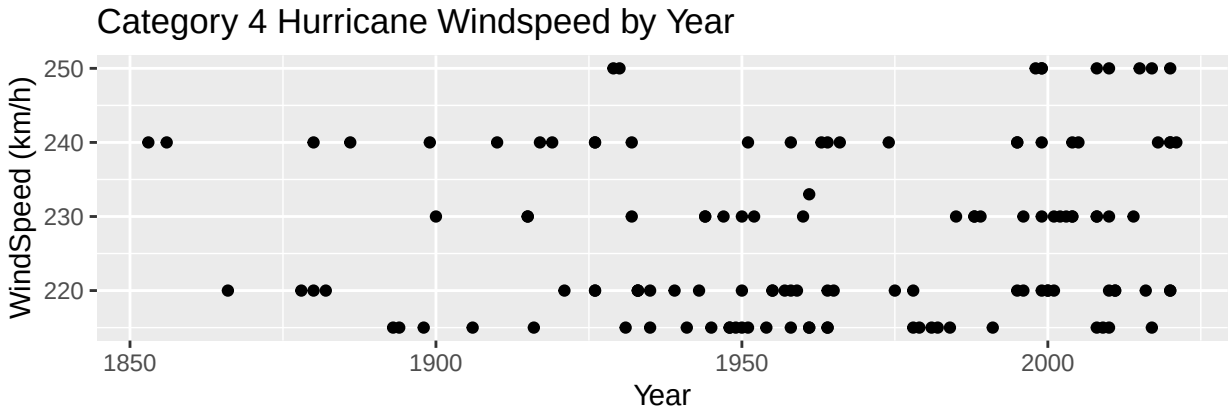
```
##      decade      Median_Wind
##  Min.   :1920   Min.   :260.0
##  1st Qu.:1950   1st Qu.:265.0
##  Median :1970   Median :275.0
##  Mean   :1968   Mean   :274.7
##  3rd Qu.:1990   3rd Qu.:280.0
##  Max.   :2010   Max.   :295.0
```

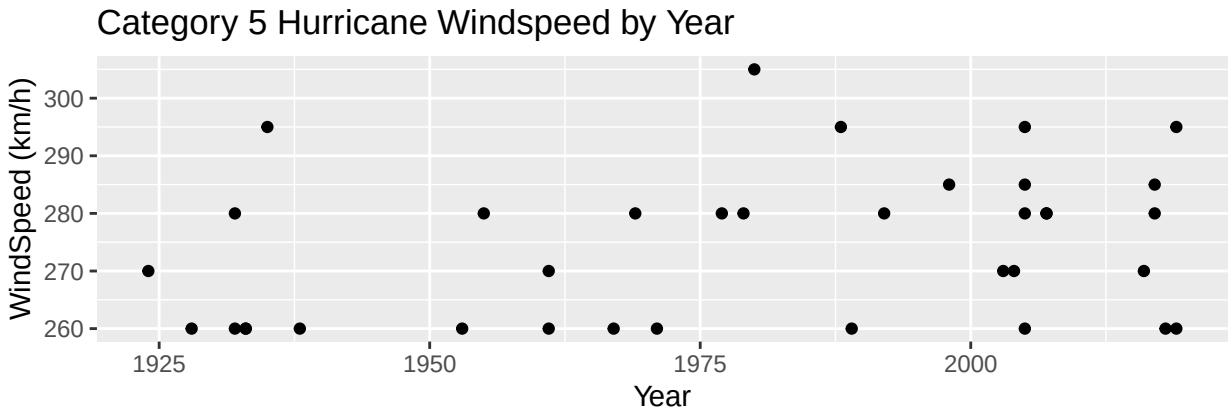
Generate the scatter plots for by year data

```
cat_4scatter <- get_windspeed_scatter(cat4_df, "Category 4 Hurricane Windspeed by Year", TRUE)

cat_5scatter <- get_windspeed_scatter(cat5_df, "Category 5 Hurricane Windspeed by Year", TRUE)

grid.arrange(cat_4scatter,cat_5scatter)
```



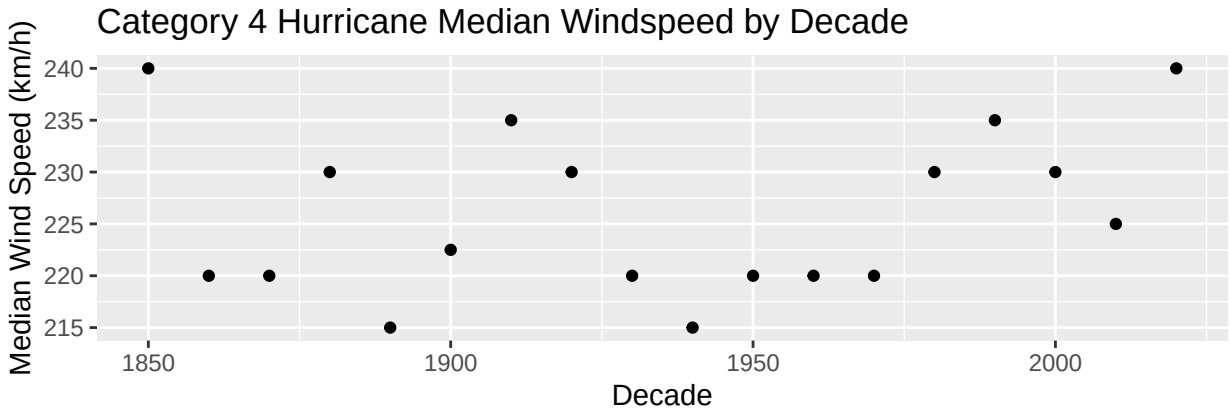


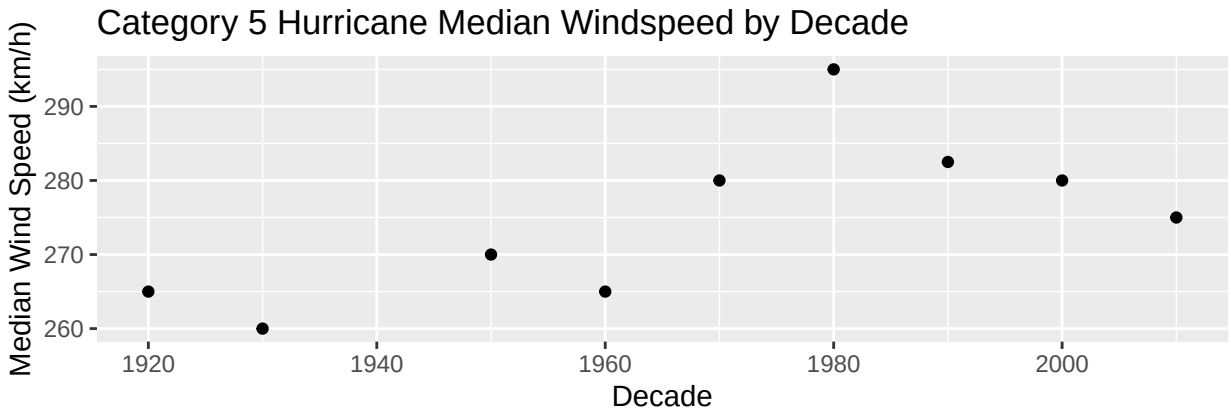
Generate the scatter plots for by decade data

```
cat_4scatter_by_decade <- get_windspeed_scatter(cat_4df_by_decade, "Category 4 Hurricane Median Windspe

cat_5scatter_by_decade <- get_windspeed_scatter(cat_5df_by_decade, "Category 5 Hurricane Median Windspe

grid.arrange(cat_4scatter_by_decade,cat_5scatter_by_decade)
```





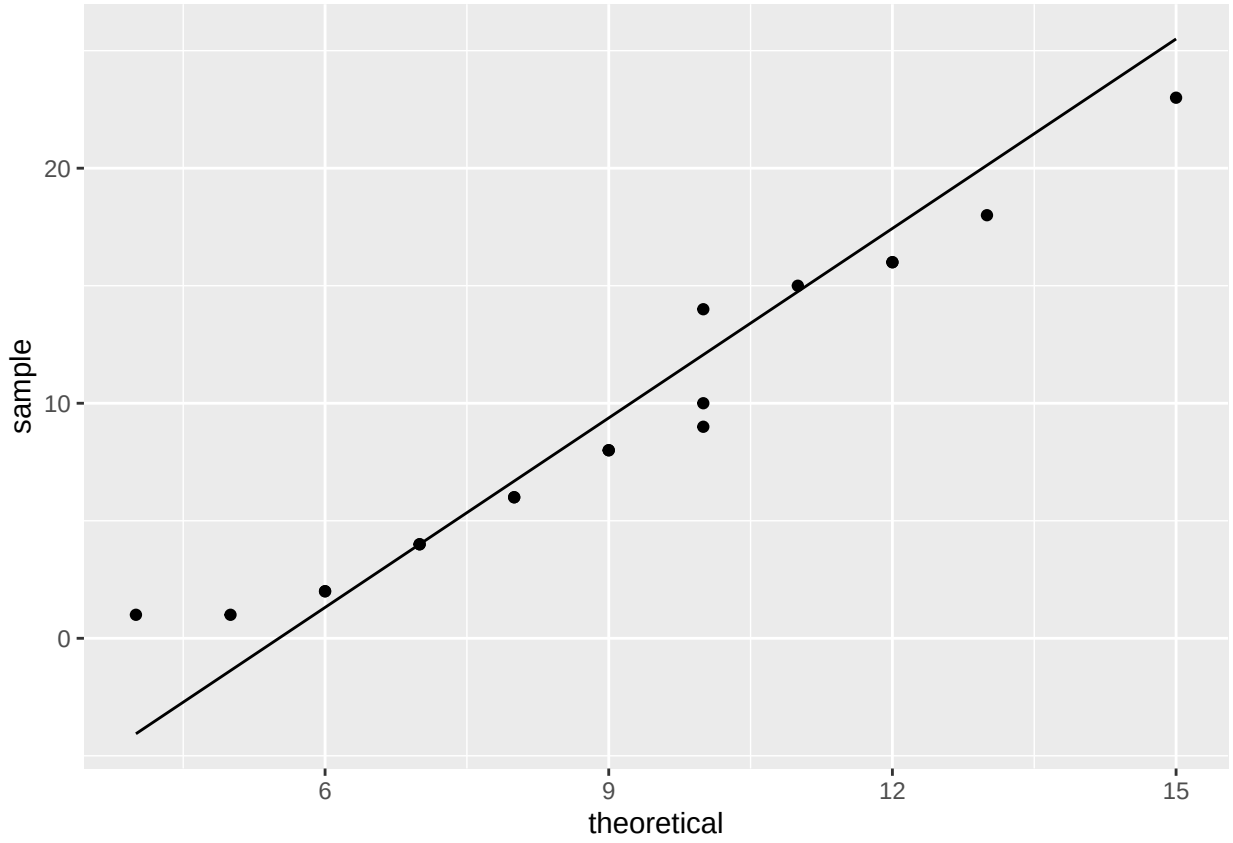
Create a dataframe to get hurricane frequency data

```
dec_freq = hurricane_df %>% count(decade) %>% rename(freq="n")
dec_freq
```

```
## # A tibble: 18 x 2
##    decade  freq
##  *  <dbl> <int>
##  1   1850     2
##  2   1860     1
##  3   1870     1
##  4   1880     4
##  5   1890     4
##  6   1900     2
##  7   1910     6
##  8   1920     8
##  9   1930    16
## 10   1940     9
## 11   1950    16
## 12   1960    15
## 13   1970     8
## 14   1980    10
## 15   1990    14
## 16   2000    23
## 17   2010    18
## 18   2020     6
```

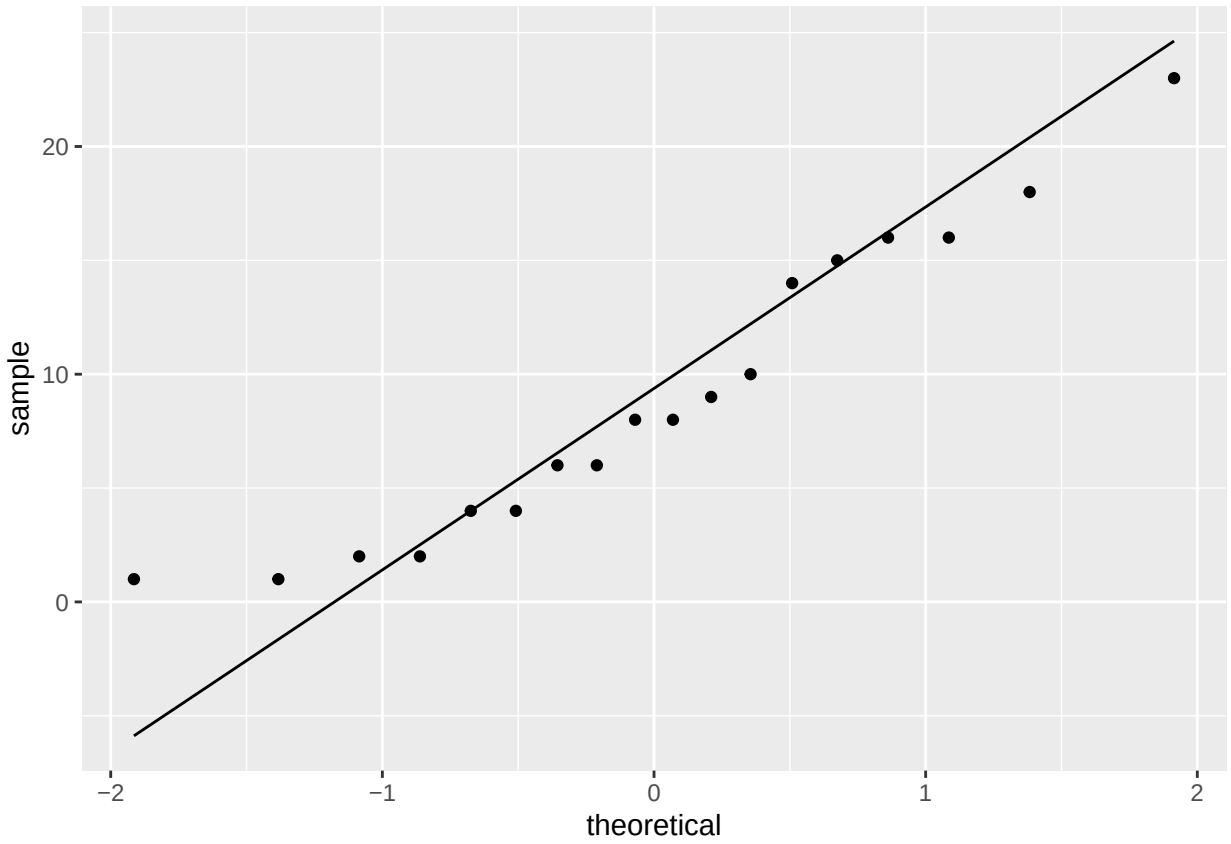
Generate the qqplot for the Poisson Distribution

```
ggplot(data = dec_freq,
       mapping = aes(sample = freq)) +
  geom_qq(distribution = stats::qpois,
          dparams = list(lambda = mean(dec_freq$freq))) +
  geom_qq_line(distribution = stats::qpois,
               dparams = list(lambda = mean(dec_freq$freq)))
```



Generate the qqplot for the Normal Distribution

```
ggplot(data = dec_freq,
       mapping = aes(sample = freq)) +
  geom_qq() +
  geom_qq_line()
```



Based on the plots Poisson is better because the datapoints are close to the 45 degree line