i everyone,

**1. Recap:**

**Same as last week's dataset, you can skip to the discussion:**

**File: [weather.cs](https://my.uclaextension.edu/users/189080/files/6366205?wrap=1&verifier=ehhwoYJ474il2SIGKMjo6uou5O7n58CFvabYCPN2" \t "_blank)**[**Download weather.cs**](https://my.uclaextension.edu/users/189080/files/6366205/download?verifier=ehhwoYJ474il2SIGKMjo6uou5O7n58CFvabYCPN2&download_frd=1)

weather <- read.csv('weather.csv')  
weather\_2016 <- weather[weather$Date.Year == '2016',]

I take out the weather dataset from ‘https://corgis-edu.github.io/corgis/csv/weather/’, But I only focus on the average temperature (°F) in LA in 2016. That’s to say, I have to get rid of 2017 data at first. I created a variable to store the dataset that only includes data from 2016.

**2. Discussion:**

before starting my discussion, I will arrange my data at first:

LA <- weather\_2016[weather\_2016$Station.City == 'Los Angeles',]  
CA <- weather\_2016[weather\_2016$Station.State == 'California',]  
Long\_beach <- weather\_2016[weather\_2016$Station.City == 'Long Beach',]  
LA\_LB = rbind(LA, Long\_beach)

library(multcomp)

**A.stem-and-leaf plot:**

stem(LA$Data.Temperature.Avg.Temp, scale = 1)

Table

Description automatically generated

From the stem data, we can see the data range is within 55-74, and T=63(F) was the mode of the overall LA's average dataset.

**B. Histogram with overlaid density plot:**

hist(LA$Data.Temperature.Avg.Temp, breaks = seq(50,80,2),    
     xlab = 'Temperature',col=rgb(1,0,1,0.25), ylim = c(0,12),   
     main = "LA and Long Beach's weather")  
hist(Long\_beach$Data.Temperature.Avg.Temp, breaks = seq(50,80,2),   
     col=rgb(0,1,1,0.5), add = T)  
legend ('topright', c('Los Angeles', 'Long Beach'),   
        col = c(rgb(1,0,1,0.25), rgb(0,1,1,0.5)), lwd = 10)

Chart, histogram

Description automatically generated

From the above density plot, the most common average temperature in Los Angeles and Long Beach is around 64-66 (F), ~10 counts

par(mfrow=c(1,3))  
hist( LA$Data.Temperature.Avg.Temp, freq = F, breaks = seq(50,80,2),    
     xlab = 'Temperature', col=rgb(1,0,1,0.25), ylim = c(0,0.13),  
     main = "Los Angeles's weather")  
dens1 <- density (LA$Data.Temperature.Avg.Temp, bw = 2)  
lines(dens1, lwd = 2)  
legend ('topright', c('Density', 'Lines'),   
        col = c(rgb(1,0,1,0.25), 'Black'), lwd = c(10,2))  
hist(Long\_beach$Data.Temperature.Avg.Temp, freq = F, ylim = c(0,0.13),  
     breaks = seq(50,80,2), xlab = 'Temperature', col=rgb(0,1,1,0.5),    
     main = "Long Beach's weather")  
dens2 <- density (Long\_beach$Data.Temperature.Avg.Temp, bw = 2)  
lines(dens2, lwd = 2)  
legend ('topright', c('Density', 'Lines'),   
        col = c(rgb(0,1,1,0.25), 'Black'), lwd = c(10,2))  
hist( LA\_LB$Data.Temperature.Avg.Temp, freq = F, breaks = seq(45,80,2),    
      xlab = 'Temperature', col=rgb(1,1,0,0.25), ylim = c(0,0.13),  
      main = "LA and Long beach's weather")  
dens3 <- density (LA\_LB$Data.Temperature.Avg.Temp, bw = 2)  
lines(dens3, lwd = 2)  
legend ('topright', c('Density', 'Lines'),   
        col = c(rgb(1,1,0,0.25), 'Black'), lwd = c(10,2))

Chart, histogram

Description automatically generated

Text

Description automatically generated

from the density dataset,  64-66 is occupied around 10% of the dataset

Besides that, even the Both Los Angeles and Long Beach's density plot do not look like a normal distribution, when they combine into one dataset, it becomes a normal distribution.

**C. Bar plot:**

LA\_LB$season[LA\_LB$Date.Month <=3] <- 'Winter'  
LA\_LB$season[LA\_LB$Date.Month >3 & LA$Date.Month <=6] <- 'Spring'  
LA\_LB$season[LA\_LB$Date.Month >6 & LA$Date.Month <=9] <- 'Summer'  
LA\_LB$season[is.na(LA\_LB$season)] <- 'Fall'  
LA\_LB$season <- as.factor(LA\_LB$season)  
LA\_LB$season <- factor(LA\_LB$season, levels = c('Winter', 'Spring', 'Summer', 'Fall'))

avgtemp2 <-aggregate(Data.Precipitation~Station.City\*season, data = LA\_LB, FUN = sum)  
avgtemp2

barplot (avgtemp2$Data.Precipitation, ylab = 'Average precipitation(inches)',  
         names.arg = c('Winter','', 'Spring','', 'Summer','', 'Fall',''), col = c('violetred1', 'deepskyblue1'),   
         , space = c(1,0,1,0,1,0,1,0),   
         main = "2016 LA and Long Beach's precipitation by season")  
legend ('center', c('Long Beach', 'Los Angeles'),   
        col = c('violetred1', 'deepskyblue1'), lwd = 10)

Chart, bar chart

Description automatically generated

Based on precipitation from Los Angeles and Long Beach, The 'Fall' and the 'Winter' season has the most precipitation