

R Notebook

```
#Arjun Bhan  
library(ggplot2)  
library(tidyverse)
```

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

```
## v tibble 3.0.6      v dplyr 1.0.4  
## v tidyr 1.1.2      v stringr 1.4.0  
## v readr 1.4.0      v forcats 0.5.1  
## v purrr 0.3.4
```

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

```
library(RColorBrewer)  
library(Stat2Data)
```

```
WeatherAUS<- read.csv(file.choose(), stringsAsFactors = FALSE)
```

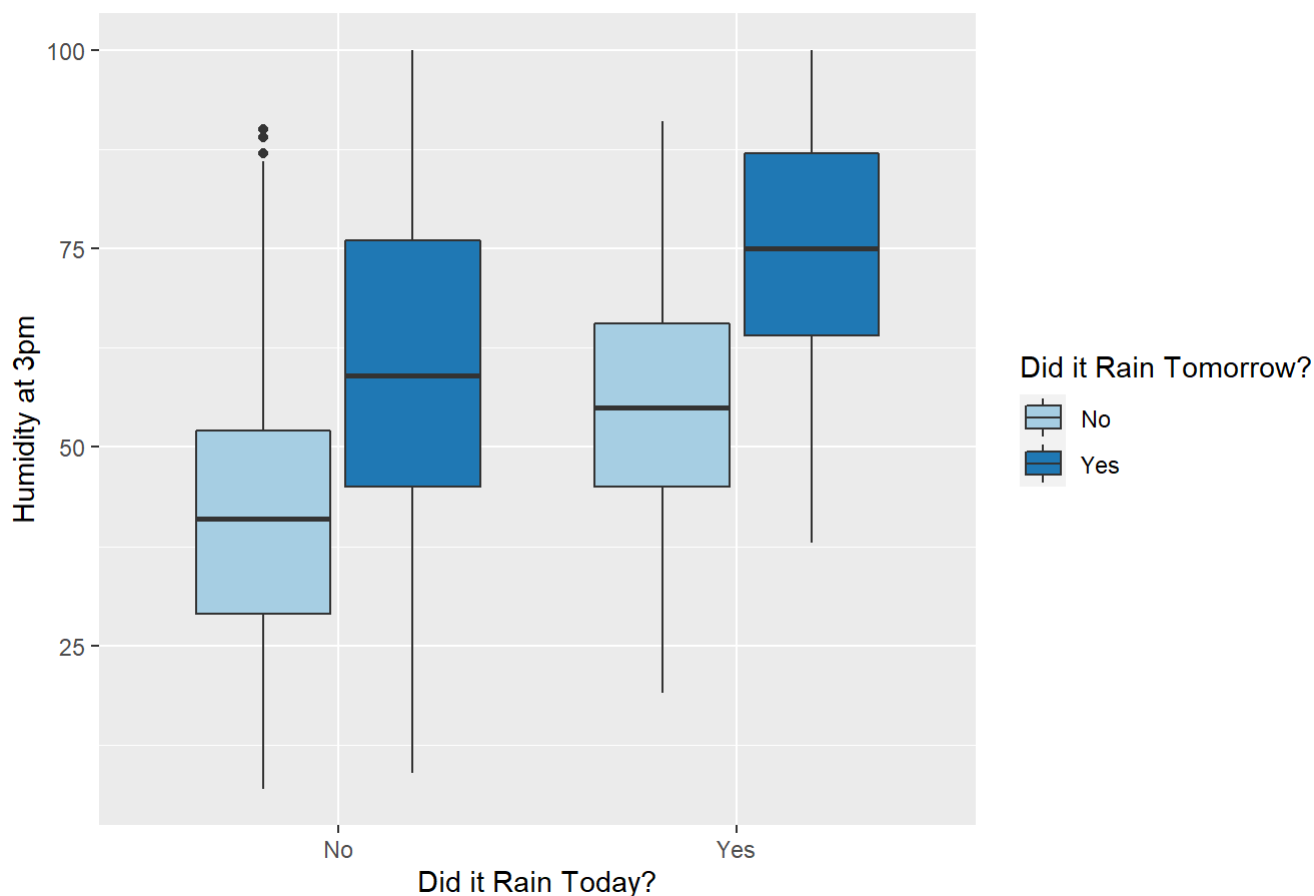
```
WeatherAUS<-WeatherAUS %>% separate(Date, c("years","months","days"),sep = "-")
```

```
WeatherAlbur=filter(WeatherAUS,Location == "Albury")  
WeatherLaun=filter(WeatherAUS,Location == "Launceston")
```

```
WeatherAlbur<-filter(WeatherAlbur,!is.na(WeatherAlbur$RainTomorrow))  
WeatherAlburGra1<-filter(WeatherAlbur,!is.na(WeatherAlbur$RainToday))  
WeatherAlburGra1<-filter(WeatherAlburGra1,!is.na(WeatherAlburGra1$Humidity3pm))
```

```
ggplot(WeatherAlburGra1,aes(x=RainToday,y=Humidity3pm, fill=RainTomorrow))+geom_boxplot()+labs(t  
itle="The effect of humidity on likelihood to rain today and tomorrow",x="Did it Rain Today?",y=  
"Humidity at 3pm",fill="Did it Rain Tomorrow?")+scale_fill_brewer(palette="Paired")
```

The effect of humidity on likelihood to rain today and tomorrow



#From this data, we can see that humidity is higher on days when it rained and on the following day. Since humidity is related to if it rains today and tomorrow, it is likely that these two variables are also positively correlated.

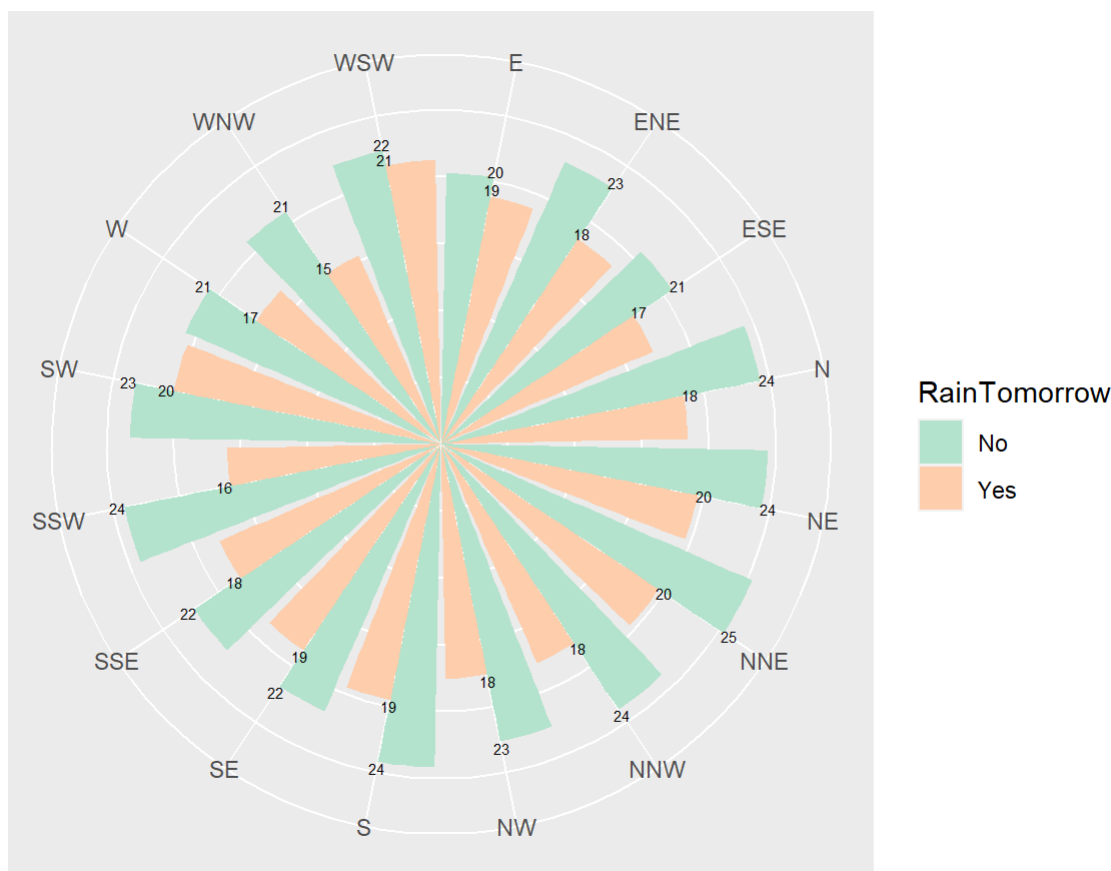
```
WeatherAUS_Rain_Wind<-WeatherAlbur%>% group_by(WindGustDir,RainTomorrow) %>% summarize(Temp3pm=mean(Temp3pm))
```

`summarise()` has grouped output by 'WindGustDir'. You can override using the `.groups` argument.

```
WeatherAUS_Rain_Wind<-filter(WeatherAUS_Rain_Wind,!is.na(WindGustDir))
WeatherAUS_Rain_Wind<-filter(WeatherAUS_Rain_Wind,!is.na(RainTomorrow))
WeatherAUS_Rain_Wind<-filter(WeatherAUS_Rain_Wind,!is.na(Temp3pm))
```

```
ggplot(WeatherAUS_Rain_Wind,aes(x=WindGustDir, y=Temp3pm,fill=RainTomorrow))+geom_col(position = "dodge")+geom_text(aes(label=round(Temp3pm,digits=0)),position = position_nudge(y=.5),size =2,)+coord_polar()+scale_fill_brewer(palette="Pastel2")+theme(axis.ticks = element_blank(),axis.text.y = element_blank(),axis.title.x=element_blank(),axis.title.y=element_blank())+labs(title="The effect of temp at 3 pm on whether it will rain for all wind gust directions")
```

The effect of temp at 3 pm on whether it will rain for all wind gust directions



#From this graph we can tell that when temperature at 3 pm is lower there is greater chance of rain. This is true for whichever direction the wind gust is blowing. As temperature changes for the direction of the wind gust, my hypothesis is that there is whether it rains tomorrow varies with temperature at 3 pm and the wind gust direction.

```
WeatherAlbur$months<-as.integer(WeatherAlbur$months)
```

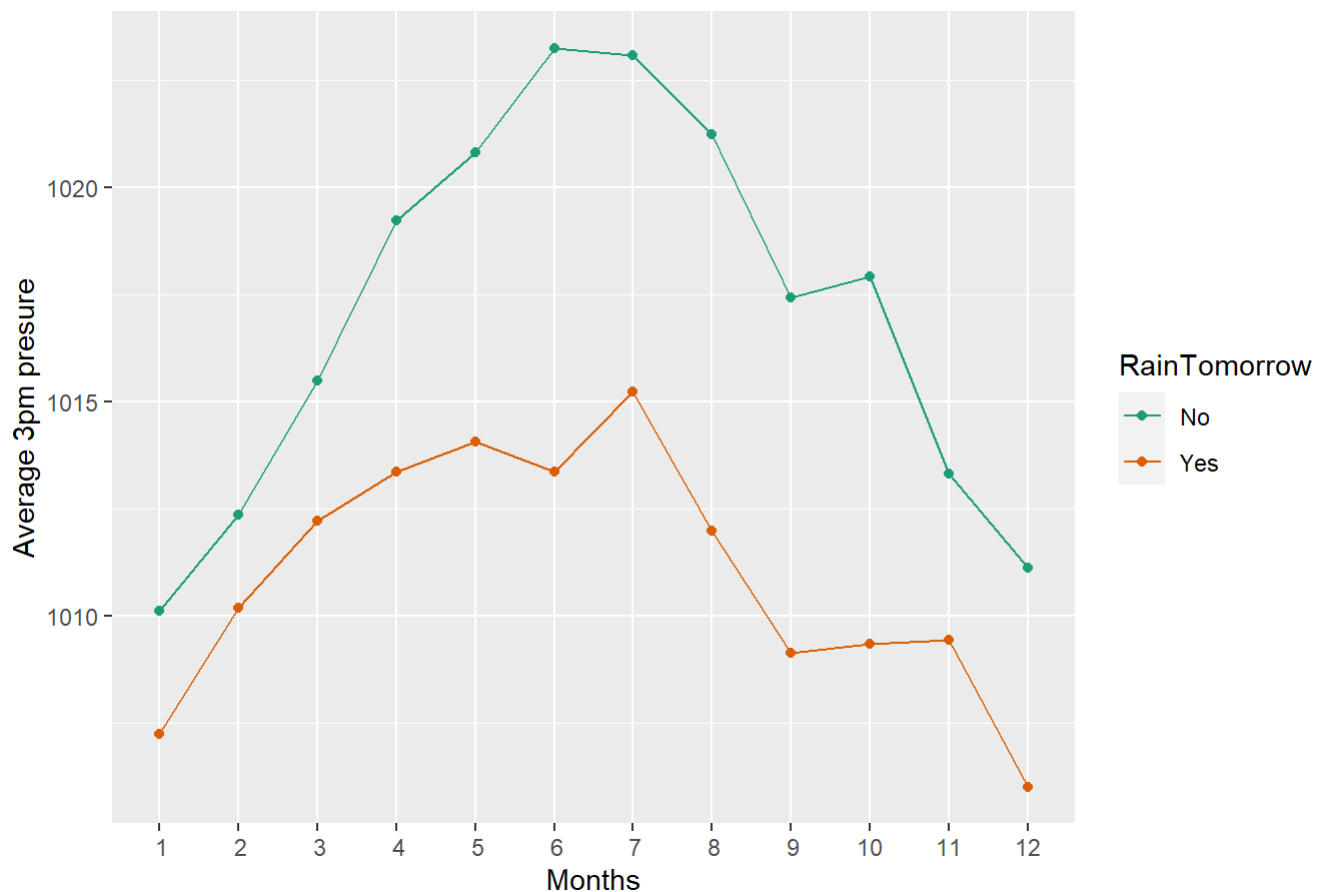
```
WeatherAUSMonthGro<-filter(WeatherAlbur,!is.na(WeatherAlbur$months))
WeatherAUSMonthGro<-filter(WeatherAUSMonthGro,!is.na(WeatherAlbur$Pressure3pm))
WeatherAUSMonthGro<-WeatherAUSMonthGro%>% group_by(months,RainTomorrow) %>% summarize(MeanPressure3pm=mean(Pressure3pm))
```

`summarise()` has grouped output by 'months'. You can override using the `.groups` argument.

```
ggplot(WeatherAUSMonthGro,aes(x=months,y=MeanPressure3pm,color=RainTomorrow))+geom_point()+geom_line()+labs(title="The effect of pressure on likelihood to rain tomorrow for each month",x="Months",y="Average 3pm pressure",fill="Rain Tomorrow")+scale_x_discrete(limits=c(1:12))+scale_color_brewer(palette="Dark2")
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```

The effect of pressure on likelihood to rain tomorrow for each month



As we can see from the graph, for every month the average 3pm pressure is higher when it does not rain the next day than when it does. This means that pressure and chance of rain are negatively correlated. As we can see from the graph, during June and July have the highest pressure. During these months it's much more likely not to rain the next day than it is for it to rain.

```
WeatherLaun$months<-as.integer(WeatherLaun$months)
WeatherLaunMonthGro<-filter(WeatherLaun,!is.na(WeatherLaun$months))
WeatherLaunMonthGro<-filter(WeatherLaunMonthGro,!is.na(WeatherLaunMonthGro$Pressure3pm))
WeatherLaunMonthGro<-WeatherLaunMonthGro%>% group_by(months,RainTomorrow) %>% summarize(MeanPressure3pm=mean(Pressure3pm))
```

`summarise()` has grouped output by 'months'. You can override using the `.groups` argument.

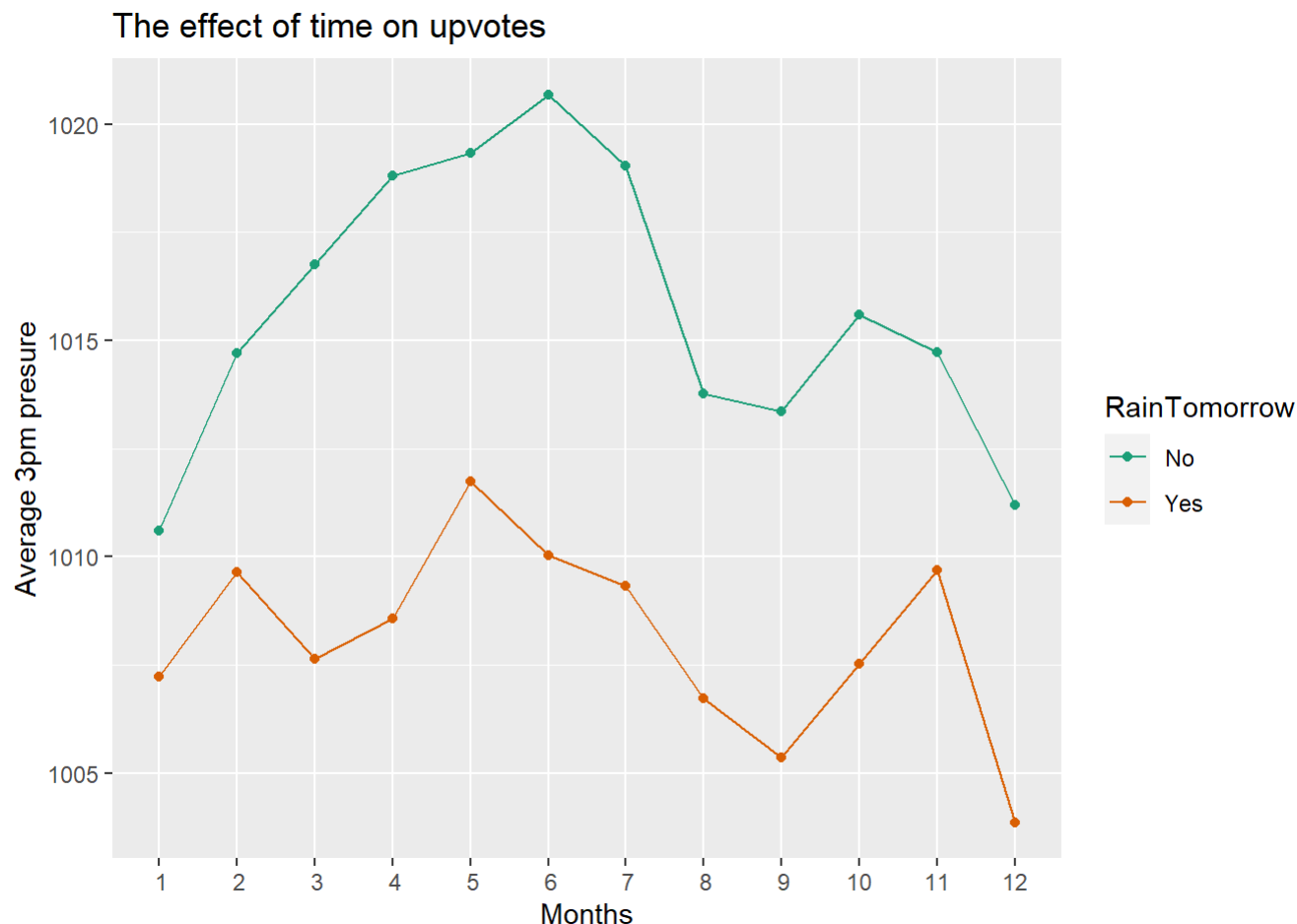
```
WeatherLaun$months<-as.integer(WeatherLaun$months)

WeatherLaunMonthGro<-filter(WeatherLaun,!is.na(WeatherLaun$months))
WeatherLaunMonthGro<-filter(WeatherLaunMonthGro,!is.na(WeatherLaunMonthGro$Pressure3pm))
WeatherLaunMonthGro<-filter(WeatherLaunMonthGro,!is.na(WeatherLaunMonthGro$RainTomorrow))
WeatherLaunMonthGro<-WeatherLaunMonthGro%>% group_by(months,RainTomorrow) %>% summarize(MeanPressure3pm=mean(Pressure3pm))
```

`summarise()` has grouped output by 'months'. You can override using the `.groups` argument.

```
ggplot(WeatherLaunMonthGro,aes(x=months,y=MeanPressure3pm,color=RainTomorrow))+geom_point()+geom_line()+labs(title="The effect of time on upvotes",x="Months",y="Average 3pm pressure",fill="Rain Tomorrow")+scale_x_discrete(limits=c(1:12))+scale_color_brewer(palette="Dark2")
```

```
## Warning: Continuous limits supplied to discrete scale.  
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```



Like the previous graph, for every month the average 3pm pressure is higher when it doesn't rain than when it does. As before, I hypothesize that the mean pressure and chance of rain are negatively correlated. As we can see that June and May have the highest pressure for this region so it is likely that the chance of rain is less likely during these months. Although the data point changed between the two locations, the pressure being lower on days without rain tomorrow is true. The only change in my conclusion is that I believe that June and May will have the lowest chance of rain when in the previous graph I thought July and June would.