Auckland Cycle analysis

Lawrence May

Total number of cyclists counted for each day, with rainfall.

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.0
## v tidyr   1.1.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

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

```
#Read in bike data
bike_files<- list.files("/Users/lawrence/Google Drive/UNI/current/Stats 369/a1/",pattern=".csv",full=TRU
bike_data<-map(bike_files,read_csv)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Date = col_character()
## )

## See spec(...) for full column specifications.

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Date = col_character()
## )

## See spec(...) for full column specifications.

## Parsed with column specification:
## cols(
##   .default = col_double(),
##   Date = col_character()
## )

## See spec(...) for full column specifications.
```

```r
#Join into one dataframe
bike_data <- dplyr::bind_rows(bike_data[[1]],bike_data[[2]],bike_data[[3]])

bike <- subset(bike_data, select = -c(Date))
bike[is.na(bike)] <- 0

#Take sum of all cyclists on each day
bike<-rowSums(bike)
bike_data <- data.frame(bike_data$Date,bike)
bike_data %>% rename(Date = bike_data.Date, Count = bike) -> bikes

#Read in rain data
rain_files<- list.files("/Users/lawrence/Google Drive/UNI/current/Stats 369/a1/",pattern=".txt",full=TRU
rain_data<-map(rain_files,read_csv,skip=9)
```

```
## Parsed with column specification:
## cols(
##   Station = col_double(),
##   `Date(NZST)` = col_double(),
##   `Time(NZST)` = col_character(),
##   `Amount(mm)` = col_double(),
##   `Period(Hrs)` = col_double(),
##   Freq = col_character()
## )
```

```
## Warning: 12 parsing failures.
##   row     col  expected                                                actual
## 35091 Station a double  UserName is = tslumley                               '/Users/lawrence/Google Di
## 35091 NA      6 columns 1 columns                                            '/Users/lawrence/Google Di
## 35092 Station a double  Total number of rows output = 35090                  '/Users/lawrence/Google Di
## 35092 NA      6 columns 1 columns                                            '/Users/lawrence/Google Di
## 35093 Station a double  Number of rows remaining in subscription = 1793302 '/Users/lawrence/Google Di
## ..... ....... ......... ..................................................... ....................
## See problems(...) for more details.
```

```
## Parsed with column specification:
## cols(
##   Station = col_double(),
##   `Date(NZST)` = col_double(),
##   `Time(NZST)` = col_character(),
##   `Amount(mm)` = col_double(),
##   `Period(Hrs)` = col_double(),
##   Freq = col_character()
## )
```

```
## Warning: 12 parsing failures.
##   row     col  expected                                                actual
## 17305 Station a double  UserName is = tslumley                               '/Users/lawrence/Google Di
## 17305 NA      6 columns 1 columns                                            '/Users/lawrence/Google Di
## 17306 Station a double  Total number of rows output = 17304                  '/Users/lawrence/Google Di
## 17306 NA      6 columns 1 columns                                            '/Users/lawrence/Google Di
## 17307 Station a double  Number of rows remaining in subscription = 1775998 '/Users/lawrence/Google Di
```

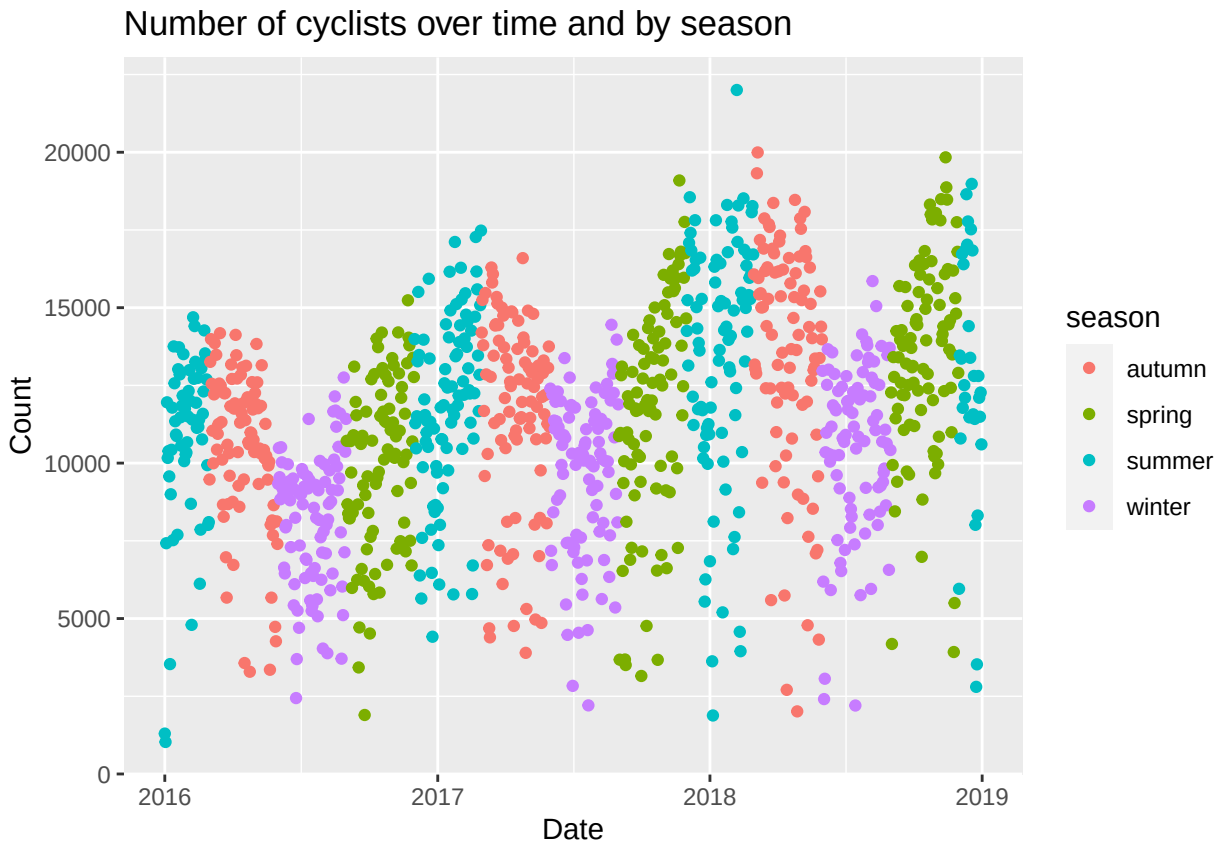
```
## ..... ....... ......... ..................................................... ......................
## See problems(...) for more details.
```

```
#Join into one dataframe
rain <- dplyr::bind_rows(rain_data[[1]],rain_data[[2]])
rain %>% filter(Station == 22719) -> rain1
rain %>% filter(Station == 37852) -> rain2

#Aggregate daily rainfall for each locationn
aggregate(rain1$`Amount(mm)` ~ rain1$`Date(NZST)`, data=rain1, FUN=sum)  -> rain1
aggregate(rain2$`Amount(mm)` ~ rain2$`Date(NZST)`, data=rain2, FUN=sum)  -> rain2

colnames(rain1) <- c("Date", "Rain")
colnames(rain2) <- c("Date", "Rain")
rain <- rbind(rain1,rain2)

#Take average of daily rain in Mangere and Albany to get approximation for average rainfall across Auck
aggregate(Rain ~ Date, data=rain, FUN=mean)  -> rain

bikes %>% separate(col=Date, into=c("dow","day","month","year")) %>% unite(Date,c(day,month,year), sep =

#Change date to same format and join on date
bikes$Date <- as.Date(bikes$Date, format='%d %b %Y')
rain$Date <- as.Date(as.character(rain$Date), format='%Y%m%d')
combined_df <- inner_join(bikes,rain,by = 'Date')
head(combined_df)
```

```
##   dow       Date month year Count  Rain
## 1 Fri 2016-01-01   Jan 2016  1299 20.25
## 2 Sat 2016-01-02   Jan 2016  1030 19.15
## 3 Sun 2016-01-03   Jan 2016  7423  6.80
## 4 Mon 2016-01-04   Jan 2016 11956  0.05
## 5 Tue 2016-01-05   Jan 2016 10167  0.00
## 6 Wed 2016-01-06   Jan 2016 10387  0.00
```

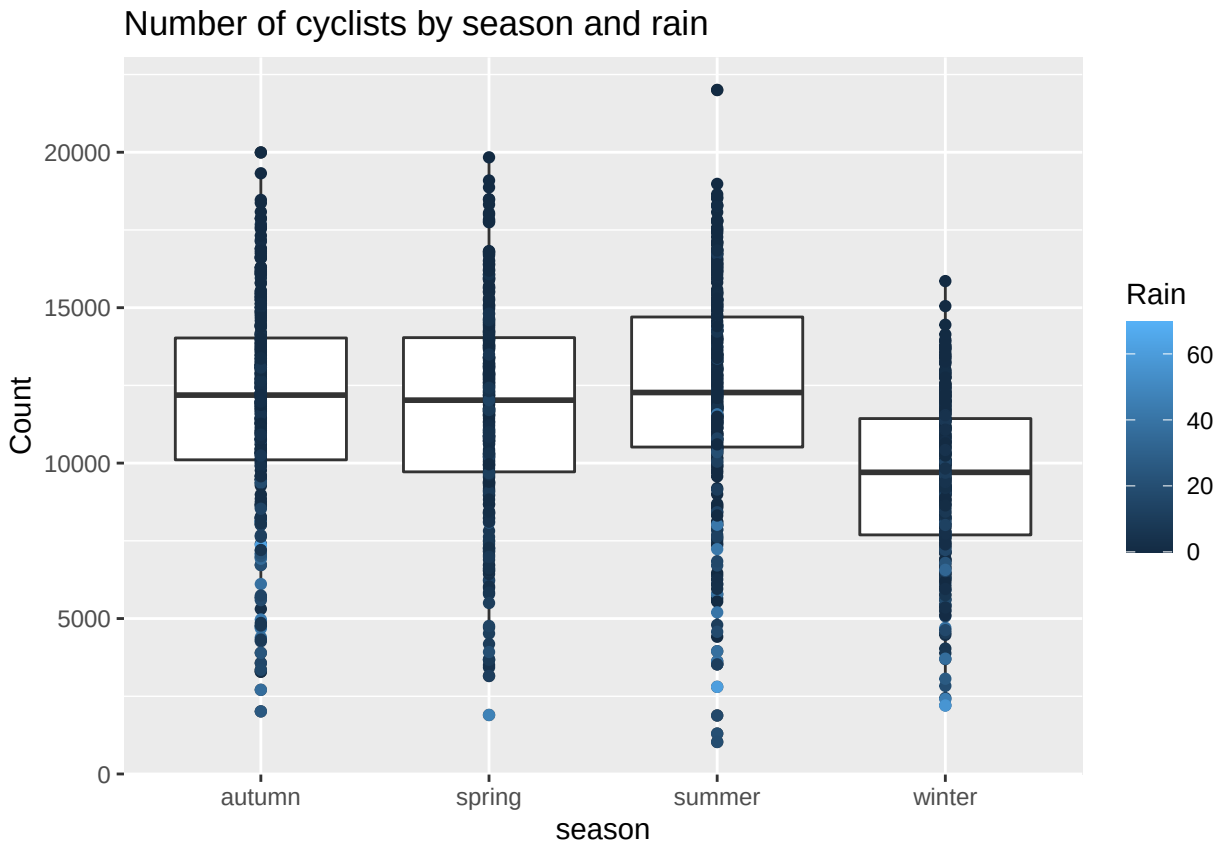
Some graphical analysis

```
#Create season factor variable
spring <- c("Sep", "Oct", "Nov")
summer <- c("Dec", "Jan", "Feb")
autumn <- c("Mar", "Apr", "May")
winter <- c("Jun", "Jul", "Aug")

#Insert into dataframe
combined_df <- mutate(combined_df, season = factor(case_when(month %in% spring ~ "spring",
                                                month %in% summer ~ "summer", month %in% autumn ~ "autumn"
                                                month %in% winter ~ "winter",
                                                TRUE                ~ NA_character_)))

ggplot(combined_df, aes(Date, Count, color = season)) + geom_point()+ggtitle("Number of cyclists over t
```



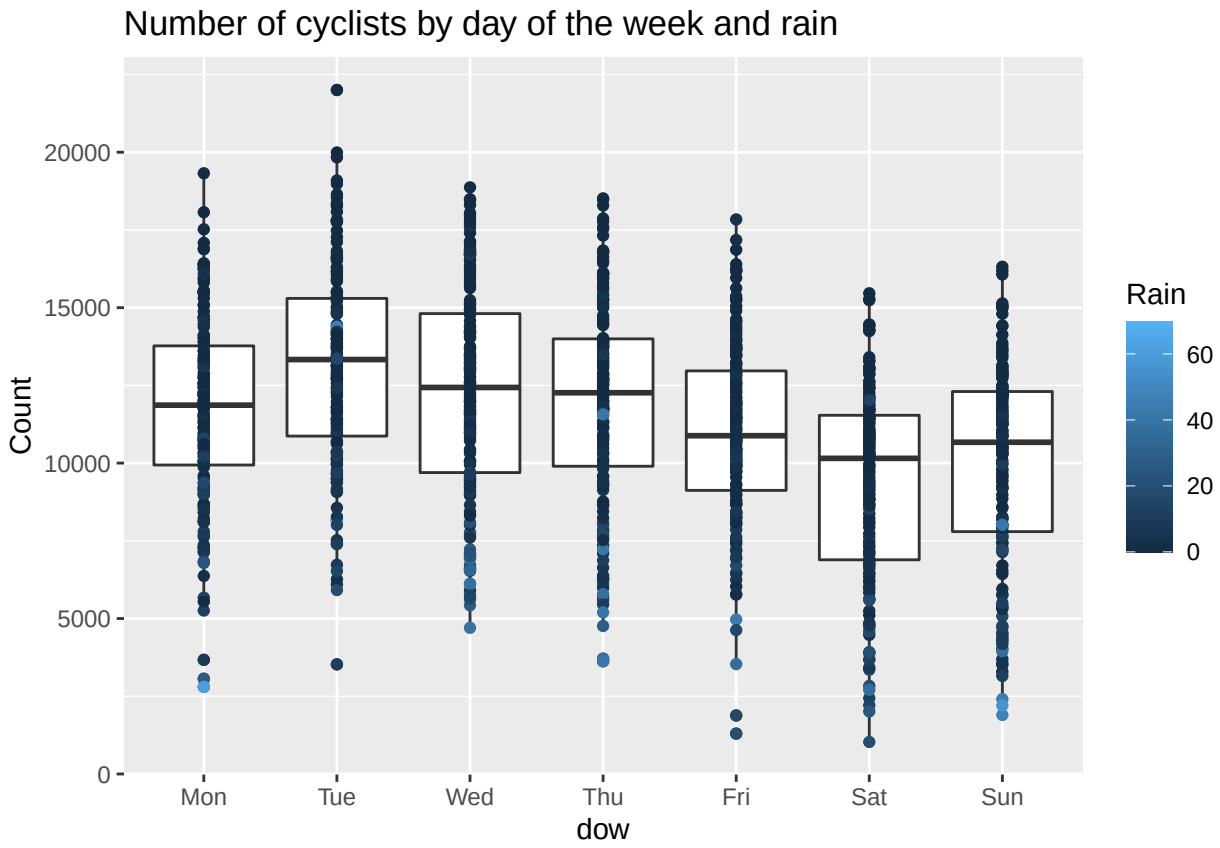
There appears to be an increasing trend of the number of cyclists over time, with a strong seasonal component (higher number of cyclist in spring and summer).

```
ggplot(combined_df, aes(season, Count, color = Rain)) + geom_boxplot()+ggtitle("Number of cyclists by s
```



There is a strong seasonal pattern with most cyclists in the summer, spring and autumn months and far less in the winter months. Rain does appear to have a negative effect as well, and seems to be able to explain many of the negative outliers (light blue = heavy rain).

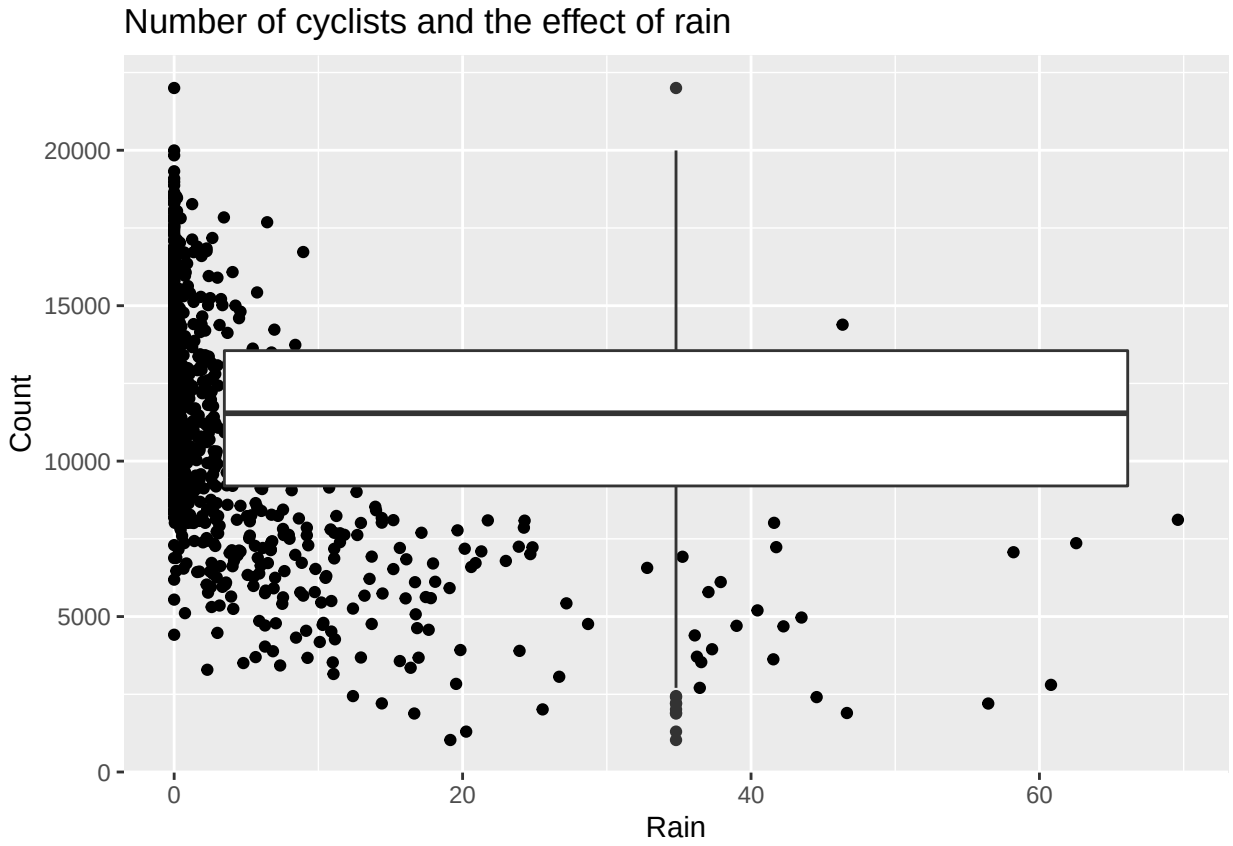
```
combined_df$dow<-factor(combined_df$dow , levels=c('Mon','Tue','Wed','Thu','Fri','Sat','Sun'))
ggplot(combined_df, aes(dow, Count, color = Rain)) + geom_boxplot()+ggtitle("Number of cyclists by day
```



There also appears to be a Day of the week effect, with most cyclists riding on Tuesdays and the least amount of Saturdays.

```
ggplot(combined_df, aes(Rain, Count)) + ggtitle("Number of cyclists and the effect of rain")+ geom_poin
```

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```



More rain leads to less than average number of cyclists.

Simple regression model to estimate number of cyclists and effect of rain on cycle activity

```
combined_df$year <- as.numeric(combined_df$year)
bike.lm <- lm(Count ~ year + season + dow + Rain, data = combined_df)
summary(bike.lm)
```

```
##
## Call:
## lm(formula = Count ~ year + season + dow + Rain, data = combined_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10259.7  -1232.2    194.5   1358.7   9367.8
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.121e+06  1.623e+05 -19.228  < 2e-16 ***
## year          1.554e+03  8.048e+01  19.308  < 2e-16 ***
## seasonspring -3.924e+02  1.859e+02  -2.111 0.034968 *
## seasonsummer  3.098e+02  1.859e+02   1.667 0.095891 .
## seasonwinter -2.361e+03  1.850e+02 -12.765  < 2e-16 ***
## dowTue        1.403e+03  2.457e+02   5.710 1.46e-08 ***
## dowWed        8.775e+02  2.459e+02   3.569 0.000374 ***
## dowThu        4.781e+02  2.458e+02   1.945 0.051996 .
## dowFri       -7.154e+02  2.453e+02  -2.917 0.003608 **
## dowSat       -2.246e+03  2.453e+02  -9.157  < 2e-16 ***
```

```
## dowSun       -1.574e+03  2.453e+02  -6.417 2.07e-10 ***
## Rain         -2.249e+02  8.229e+00 -27.328  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2173 on 1084 degrees of freedom
## Multiple R-squared:  0.6097, Adjusted R-squared:  0.6057
## F-statistic: 153.9 on 11 and 1084 DF,  p-value: < 2.2e-16
```

There is a highly significant effect of year on cyclists, the number of cyclists increases by 1500 every year, on average. There is also a strong seasonal effect, most notably of winter which decreases the number of cyclists by 2400 compared to the baseline autumn. As expected by the graph, there is also a strong effect of day of the week, most notably Tuesday which, on average, has 1400 cyclists more than a typical Monday.

Based on the model, rain does have a large, and highly significant, effect on people cycling in Auckland: each additional millimeter of rain precipitation per day resulted in 225 people less cycling per day. This can be seen in the graphs as well: On days with higher than usual precipitation (i.e more than 40mm) the count of people cycling is significantly less than what is usually the case in that month.