Assignment 4

Caolan McDonagh

28/03/2022

Data Visualisation - Assignment 4

This is an R Markdown document.

Question 1

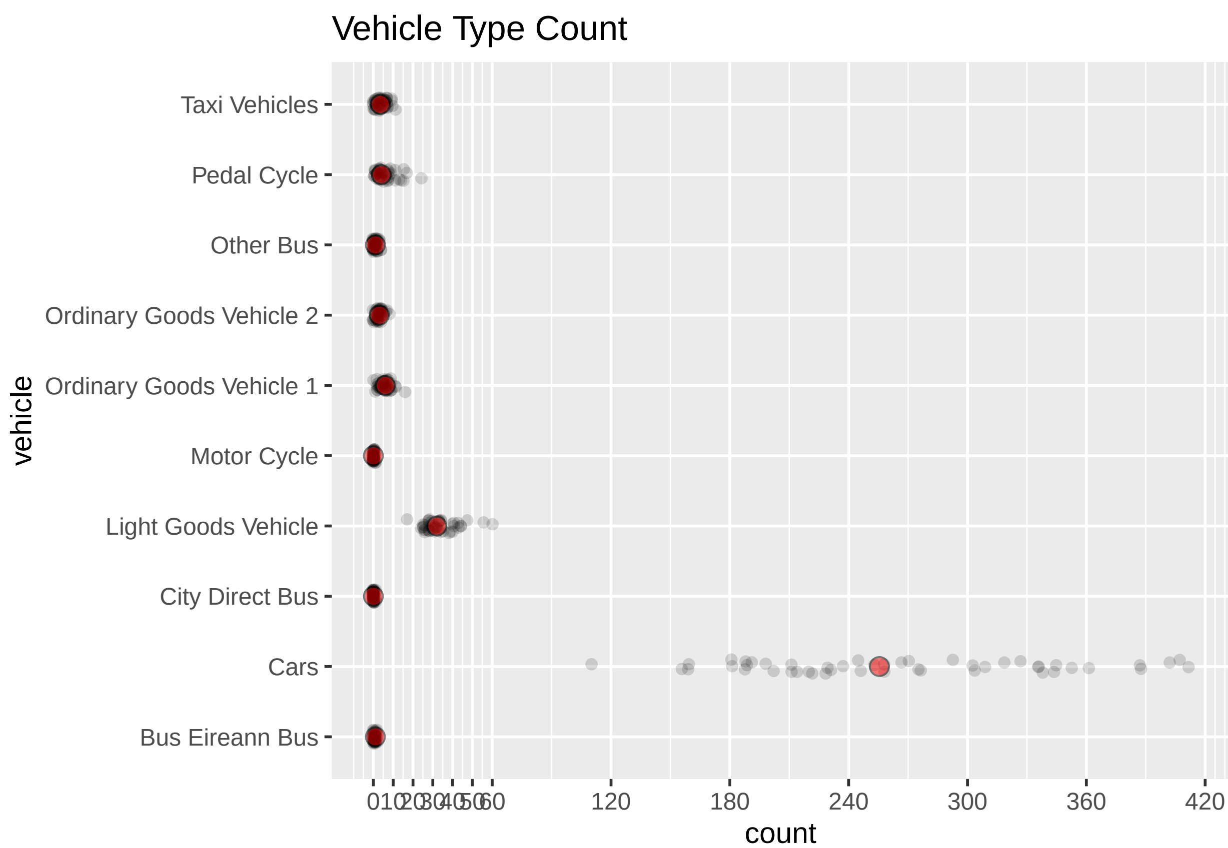
Present a visualisation of the distributions of vehicles per 15 minute interval per vehicle type contributing to traffic at this junction. Vehicles types must be labelled with their full name (not abbreviations such as 'PCL')

For Q1 I experimented with a few visualizations, but settled on the below four. The first two are both strip plots, the third is a 3D Scatter plot (intractable) and the fourth is a stacked density chart. The benefit of using a strip plot for the distribution of vehicles, is we can compare with reasonable accuracy, what the vehicle distribution looks like from a quick overview. This plot lacks fine details in my opinion, specifically because of the much larger counts seen by cars compared to any other vehicle passing through this junction. There is a single red dot per strip, showing the median count of the given vehicle type. This is good for giving a general comparison between the distribution of vehicles. The second strip plot is a mirror, except for the orientation of the data, lack of a break separating cars from the rest of the data and the replacement of the median red dot with a box plot which allows us to see where core spread of these vehicles are within the count. The 3D Scatter plot is in my opinion the best way of visualizing distribution of vehicles as there is an interactive element to it, allowing the user to zoom and inspect each vehicle or time accordingly. This third access representing time is also a bonus over the other plots as you can contrast the distribution at a given time. Each vehicle category is clearing colored on their axes, allowing for clear comparisons in distribution. I also included the stacked density chart as it is fairly good at visualizing the distribution of vehicles. Cars is clearly the majority of the traffic seen coming through this junction, followed by light goods vehicles and bicycles, and so on. It does not deal with the scale difference in cars and smaller count vehicles such as buses as they essentially squashed into invisibility. with proper scaling work done to cars, this could be much more useful, as shown the extra plot after, where it is scaled down to suit more of the data (no cars).

Overall I believe it hard to visualize the distribution of vehicles proportionally as the cars severely outnumber the rest of the vehicles seen coming through this junction. Proper break/scaling techniques are required to make it legible on a graph.

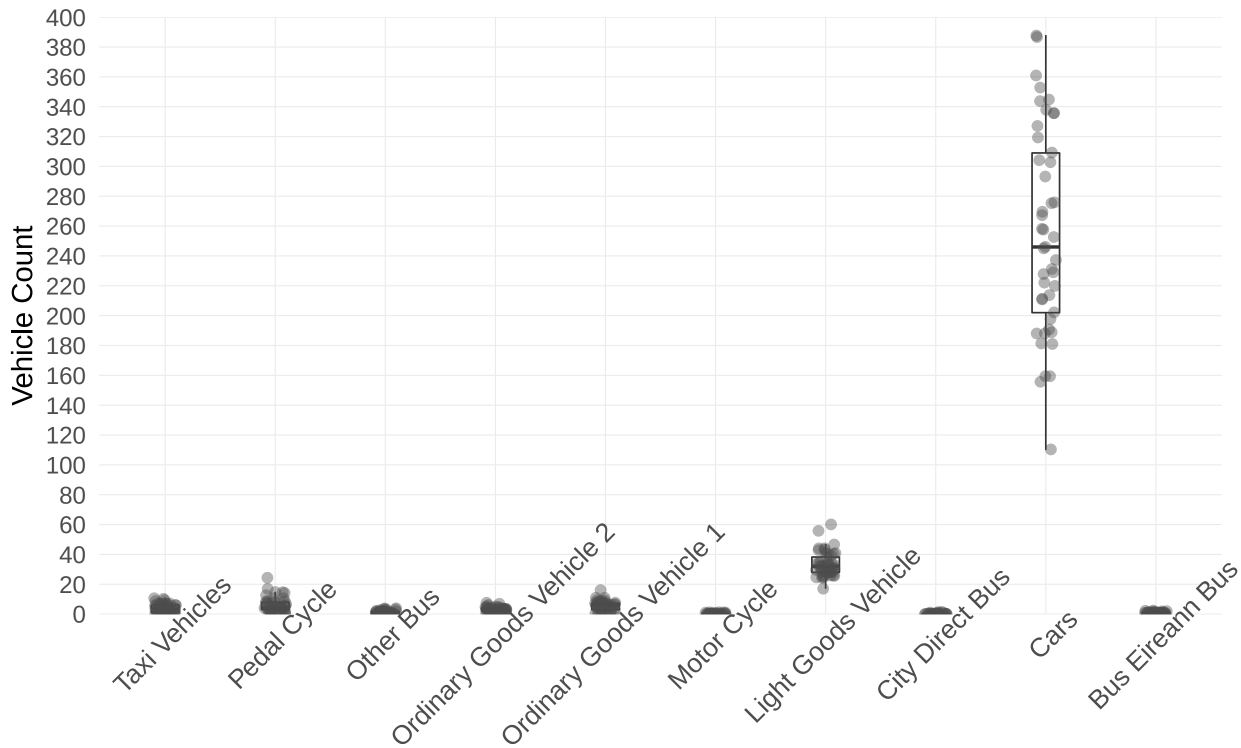
```
library(tinytex)
library(ggplot2)
library(ggridges)
library(ggdist)
library(plotly)
library(readr)
library(plyr)
library(dplyr)
```

```r
library(knitr)
library(kableExtra)
library(lattice)
library(colorspace)
library(ggrepel)
library(ggbreak)
library(lubridate)
library(viridis)

Dataset = read_csv('Junction Turning Counts 2016 Outside DSI_LONG_FORMAT.csv')

Dataset = Dataset  %>%
    mutate(vehicle = recode(vehicle, PCL = 'Pedal Cycle', MCL = 'Motor Cycle'
                           , CAR =  'Cars', TAXI =  'Taxi Vehicles'
                           , LGV =  'Light Goods Vehicle'
                           , OGV1 =  'Ordinary Goods Vehicle 1'
                           , OGV2 =  'Ordinary Goods Vehicle 2'
                           , CDB =  'City Direct Bus'
                           , BEB =  'Bus Eireann Bus', OB =  'Other Bus'))


aggdataSum <-aggregate(count ~ TIME + vehicle, data = Dataset, FUN = sum,
                       na.rm = TRUE)


ggplot(aggdataSum, aes(y=count, x=vehicle)) +

  geom_point(position = position_jitter(width = 0.1), alpha = 0.15)+
  stat_summary(fun.y = median, geom="point", fill = "red", shape = 21, size= 3
             , alpha = .5) +

  scale_y_continuous(breaks = c(0,10,20,30,40,50,60,120,180,240,300,360,420)) +
  coord_flip() +

  ggtitle("Vehicle Type Count")
```



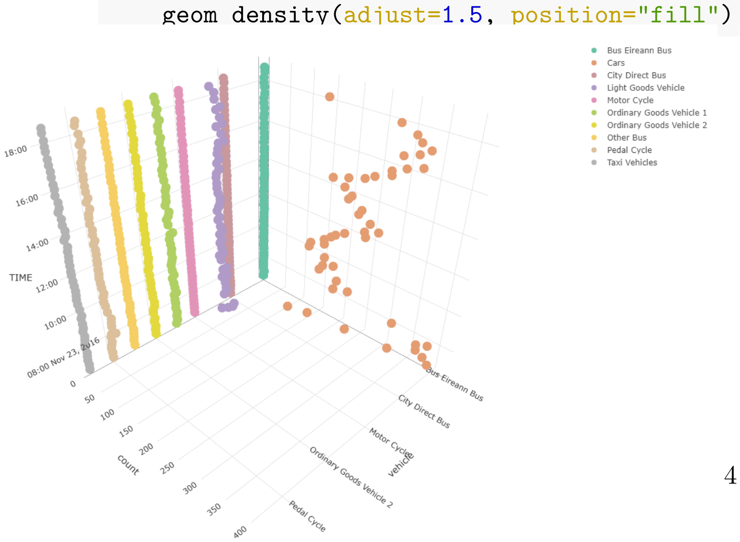

```
#################

ggplot(aggdataSum,
       aes(x = vehicle, y = count)) +
  geom_boxplot(
    width = .25,
    size=0.3,
    outlier.shape = NA
  ) +
   geom_point(col= "grey30",
              size = 1.4,
              alpha = 0.42,
              position = position_jitter(seed = 1, width = .1)) +

  scale_y_continuous(expand=c(0,0),breaks = seq(0,400,by=20),
                   limits= c(0,400),
                   name = "Vehicle Count") +
   scale_x_discrete(limits = rev)+

   theme_minimal() +
  theme(panel.grid.minor.y = element_blank(),
        panel.grid.major.x = element_line(size = 0.2),
        panel.grid.minor.x = element_blank(),
        panel.grid.major.y = element_line(size = 0.2),
        axis.text.x = element_text(size = 10, angle = 45),
        axis.title.x = element_blank(),
```

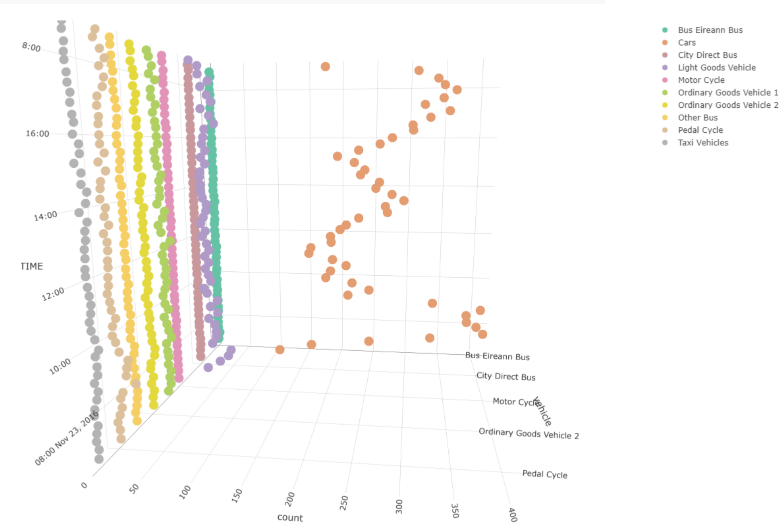
```
        panel.spacing = unit(0.5, "lines"),
        plot.margin = margin(t = 4, r = 10, b = 4, l = 4, "pt"),
        plot.title = element_text(size=10))
```

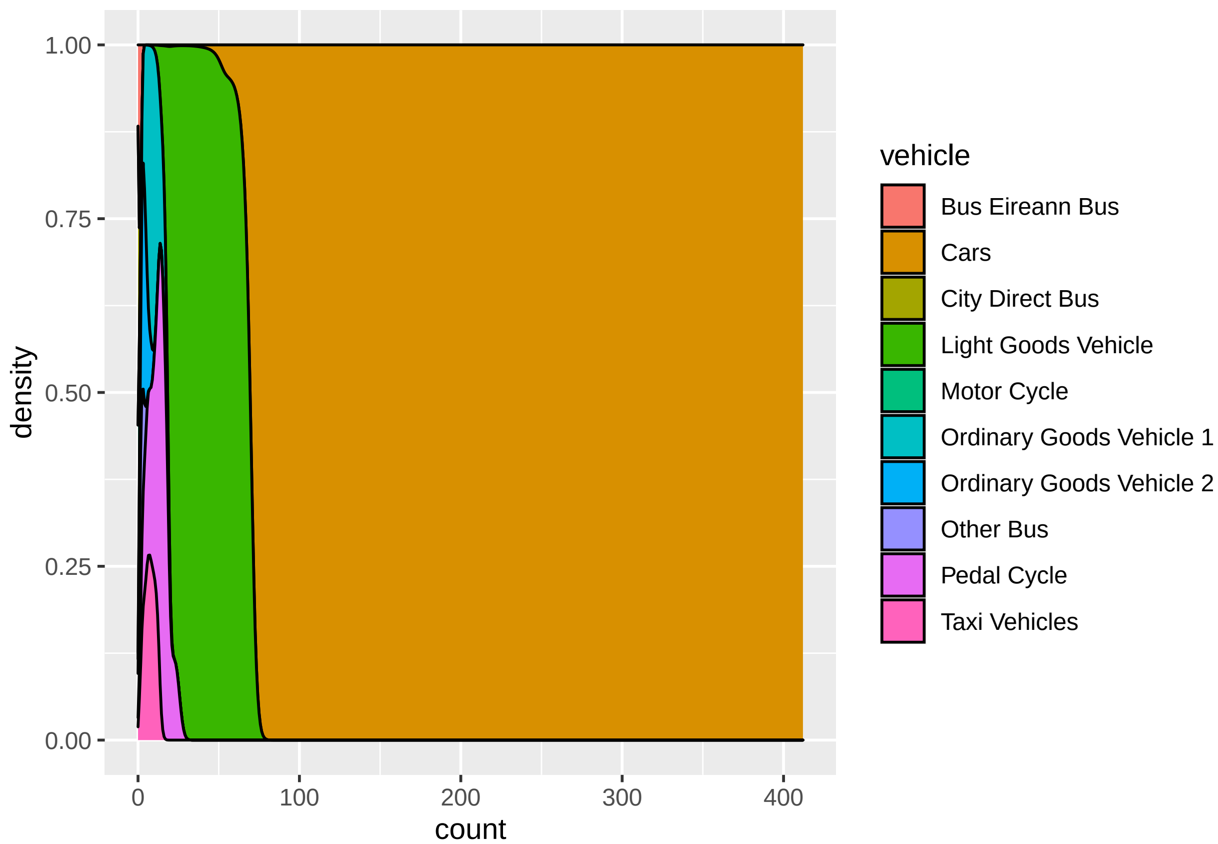



```
##############
 plot_ly(aggdataSum, y=~count, x=~vehicle, z=~TIME, type="scatter3d",
         mode="markers", color=~vehicle) %>% layout(xaxis = list(tickfont =
                                                     list(size = 1)),
        yaxis = list(tickfont = list(size = 1)))
```

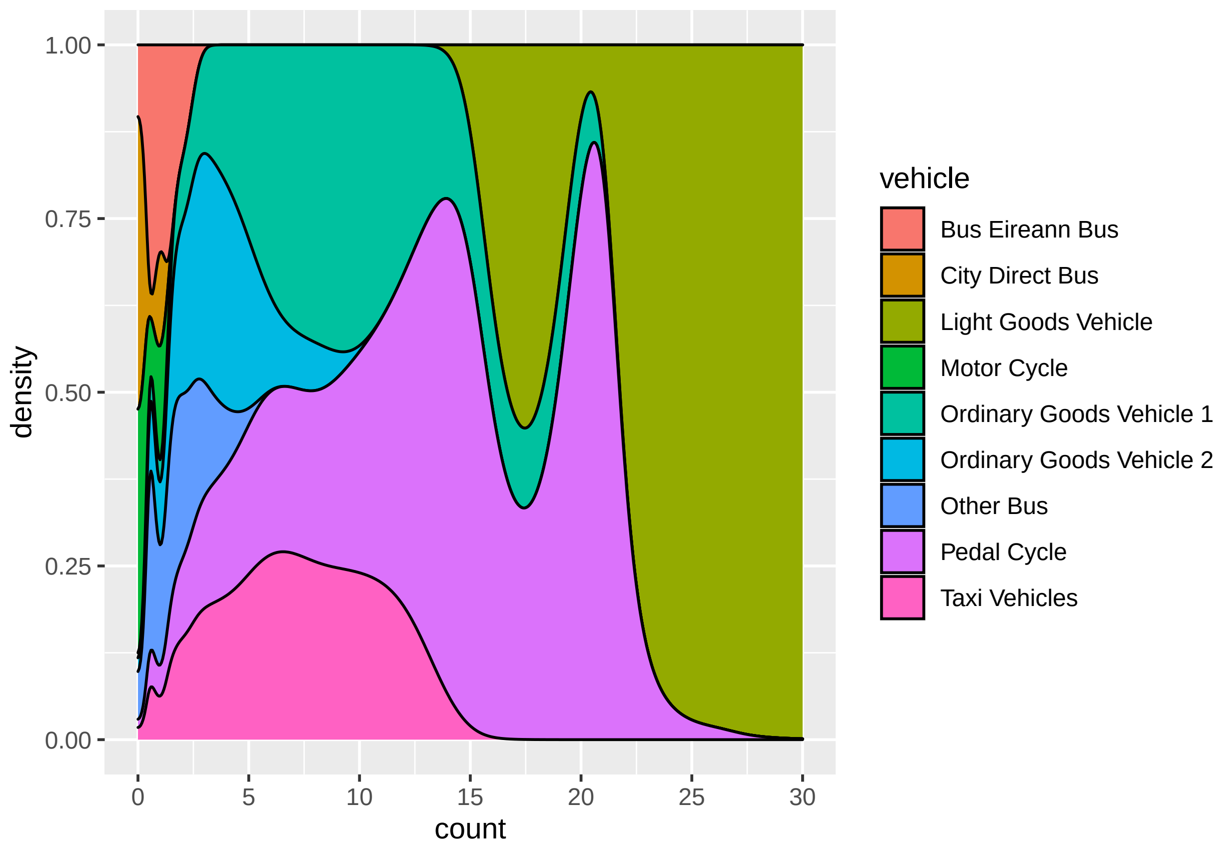
```
##############
ggplot(data=aggdataSum, aes(x=count, group=vehicle, fill=vehicle)) +
    geom_density(adjust=1.5, position="fill")
```








```
ggplot(data=aggdataSum, aes(x=count, group=vehicle, fill=vehicle)) +
    geom_density(adjust=1.5, position="fill") +
    #scale_x_continuous(breaks = c(0,10,20,30,40,50,60,120,240,360,420))
    scale_x_continuous(breaks = seq(0,30,5), limits = c(0,30))
```



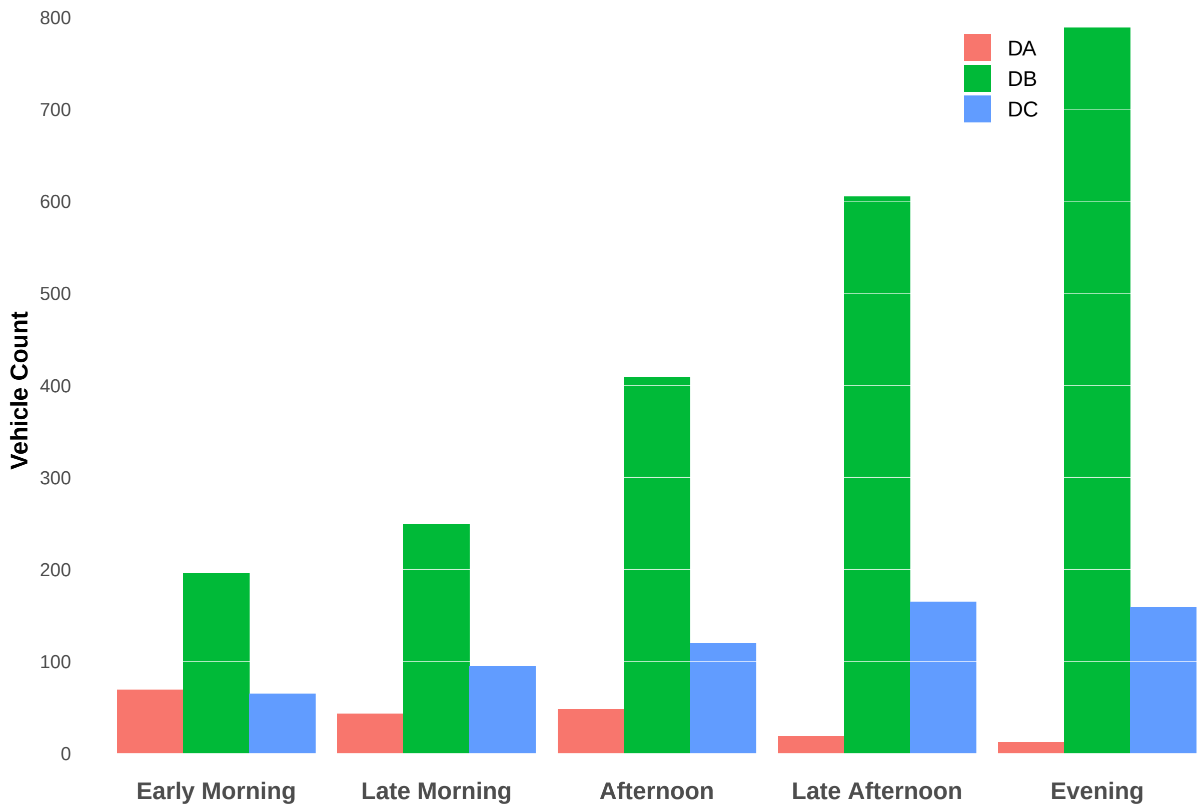
Question 2

Present a visualization that shows how the proportions of the traffic coming from D divides into the roads indicated by A, B and C at different times of the day. The times of the day are early morning (7 to 9.30 am), late morning (9.30 to 12 noon), afternoon (12 to 14.30), late afternoon (14.30 to 17.00) and evening (17.00 to 19.00)

I picked a simple grouped bar chart. Each group is the given time of day grouping. These groups then house the three divides of DA,DB and DC. We can then clearly see the proportion of traffic that takes a given route (DA,DB,DC) and also compare that directly to other times of day. I think the group bar is the best plot for this application as we can obviously pull patterns from the plot. DB is always the most popular according to the plot. If we check a map, we can see DB is leading onto the N59 (out from the city), probably the best route home for any people living North/West of Galway city. You can also see this route gets more popular as the day progresses. This is again probably due to the fact people are finishing work and heading home/out of the city. The reverse is seen with DA, which heads from the city into the industrial park. It is most popular in the morning, and again this is probably due to people heading into work. These visual clues are easy to read in the group bar chart and this is why I believe this to be the best for Q2.

```
#Show proportion of traffic from D to A/B/C against one another. Broken into 5 time slots.

TimeOfDayData <- Dataset %>%
    mutate(ymd_hms = ymd_hms(TIME),
         TimeOfDay = case_when(TIME < ymd_hms('2016-11-23 09:30:00')
                                ~ 'Early Morning',
                              TIME >=  ymd_hms('2016-11-23 09:30:00')
```

```
                             & TIME <  ymd_hms('2016-11-23 12:00:00')
                             ~ 'Late Morning',
                             TIME >=  ymd_hms('2016-11-23 12:00:00')
                             & TIME <  ymd_hms('2016-11-23 14:30:00')
                             ~ 'Afternoon',
                             TIME >=  ymd_hms('2016-11-23 14:30:00')
                             & TIME <  ymd_hms('2016-11-23 17:00:00')
                             ~ 'Late Afternoon',
                             TIME >=  ymd_hms('2016-11-23 17:00:00')
                             ~ 'Evening'))


TimeOfDayData <- TimeOfDayData %>%
  filter(grepl("DA|DB|DC", turn))

#TimeOfDayData <- aggregate(TimeOfDayData$count, by=list(turn=TimeOfDayData$turn), FUN=sum)

TimeOfDayData <- aggregate(count ~ TimeOfDay + turn, TimeOfDayData, sum)



ggplot(TimeOfDayData, aes(x = TimeOfDay, y=count, fill=turn)) +
  geom_bar(position="dodge", stat="identity")  +

  scale_y_continuous(name = "Vehicle Count", breaks = seq(0,800, by=100)) +

  scale_x_discrete(limits = c('Early Morning','Late Morning','Afternoon',
                              'Late Afternoon','Evening')) +

  theme_classic() +

  theme(
        axis.line.x = element_blank(),
        axis.ticks.x = element_blank(),
        axis.ticks.y = element_blank(),
        axis.line.y= element_blank(),
        axis.title.x=element_blank(),
        axis.text.x = element_text(  vjust = 5, size=9, face="bold"),
        axis.text.y = element_text( size=7),
        axis.title.y = element_text(size=9, face="bold"),
        legend.text = element_text(size=8),
        legend.title = element_blank(),
        legend.position = c(0.8,0.9),
        legend.key.size = unit(0.8,"line"),
        plot.title=element_text( hjust=0.00, face='bold', size=11),
        panel.background = element_blank(),
    panel.grid.major.y = element_line(size = 0.1, linetype = 'solid', colour =
                                        "white"),
     panel.ontop = TRUE
        )
```



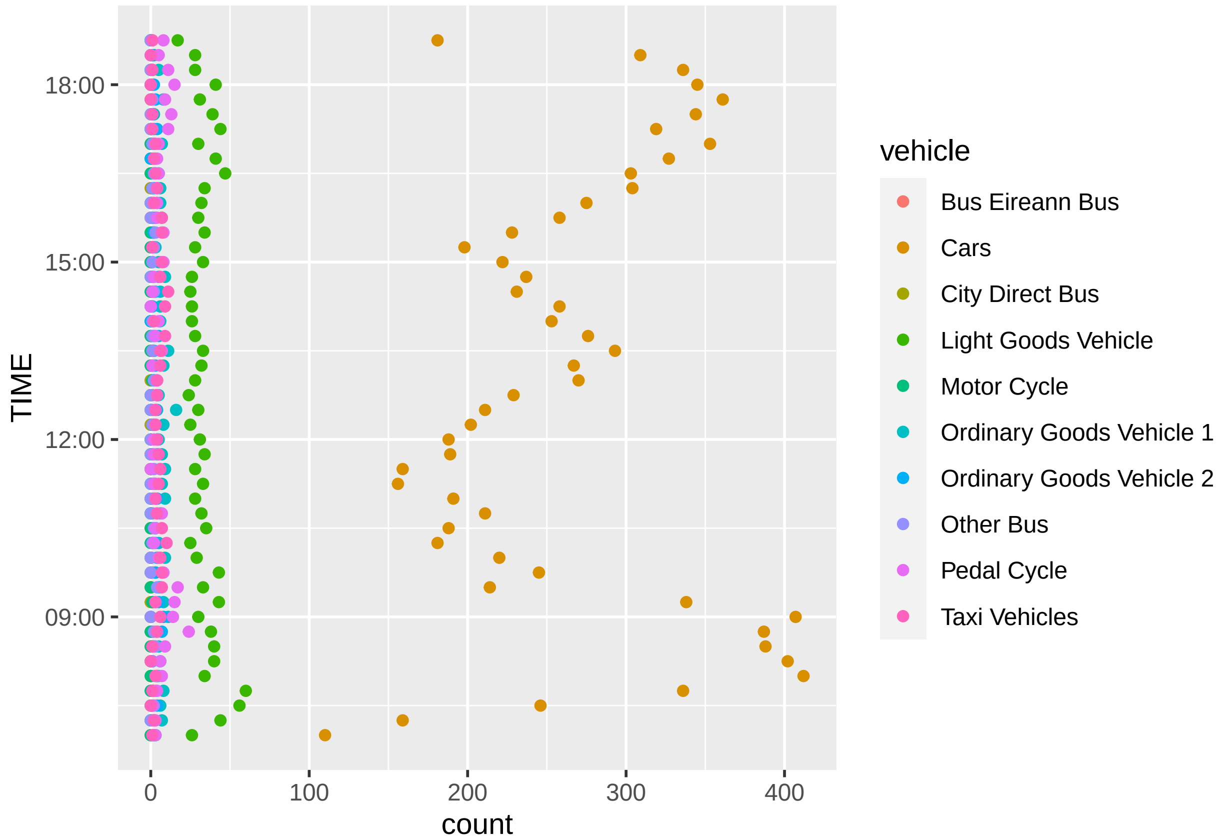
Question 3

Present a visualization of the volume of vehicles at this junction per vehicle type at each time stamp in the data set. This should be a single plot. A line graph, area or bar plots are not acceptable solutions. The reader should be able to perceive any patterns in the data.

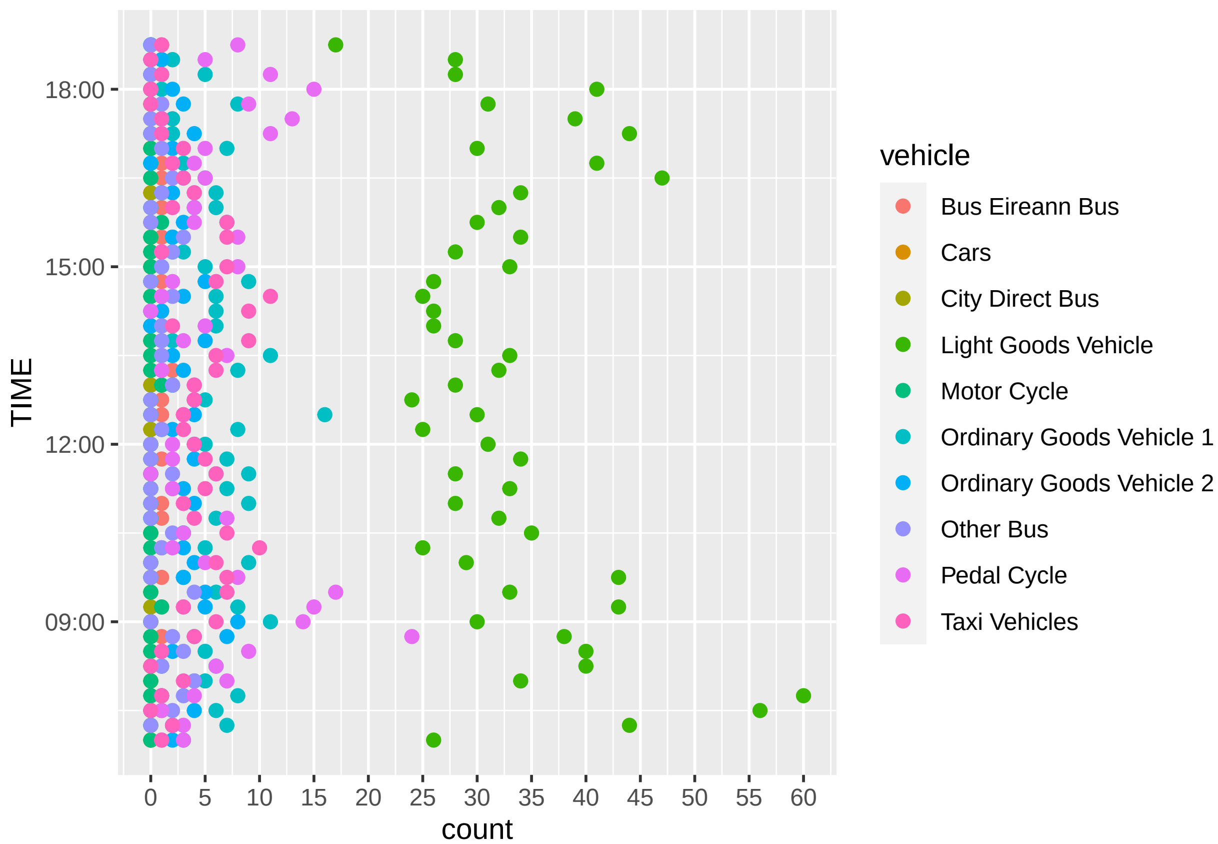
For Q3 I chose to use a multiple strip plot. Where each strip is a given time interval (every 15mins). This allows me to plot the given vehicle count per 15min, allowing us to extract patterns in the data through time. Each vehicle is given colored to allow us to distinguish. The most obvious pattern visible to a reader is the count of cars throughout the day. The count has obvious peaks and valleys, the key peaks being early morning, where we can see the volume of cars is at its highest. This is due to people making their way to work, dropping kids to school etc… This is seen again around lunch hour and then at the end of the work/school day. This pattern is replicated on a smaller level in light goods vehicles, again following the hours of a workday. This is more obvious if we remove cars from the plot. This closer look shows some more patterns that couldn't be resolved from the full plot. Things like buses are very linear in their volume as they are usually at a constant count, passing through the junction as part of a circuit on a regular basis.

The multiple strip plot and use of dots lets the reader summarize the rough count of a given vehicle at a fairly accurate time stamp during the day.

```
#Volume of vehicles at each time

ggplot(aggdataSum, aes(y=TIME, x=count, label=vehicle)) +
  geom_point(aes(col=vehicle), size=1.5)
```




```
ggplot(aggdataSum, aes(y=TIME, x=count, label=vehicle)) +
  geom_point(aes(col=vehicle), size=2) +
  scale_x_continuous(breaks = seq(0,60,5), limits = c(0,60))
```



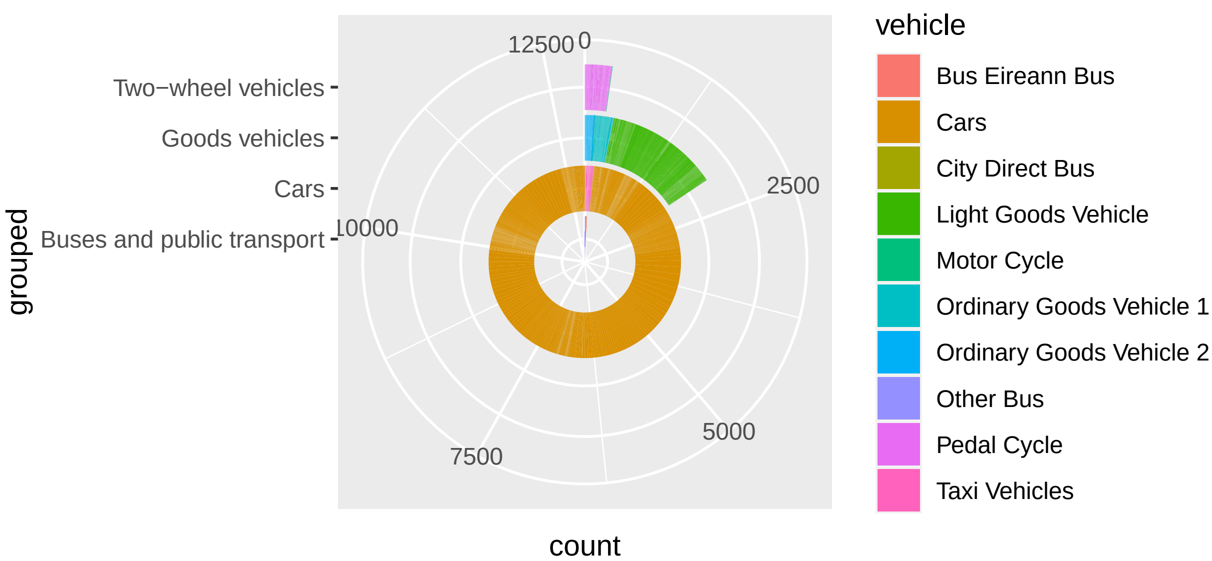
Question 4

The vehicle types can be grouped into the following categories Two-wheel vehicles: PCL, MCL Cars : Car, Taxi Goods vehicles: LGV, OGV1, OGV2 Buses and public transport : CDB, BEB, OB Present a visualization that shows the proportion of these categories of vehicles and their subcategories at this junction over the full 12 hour period.

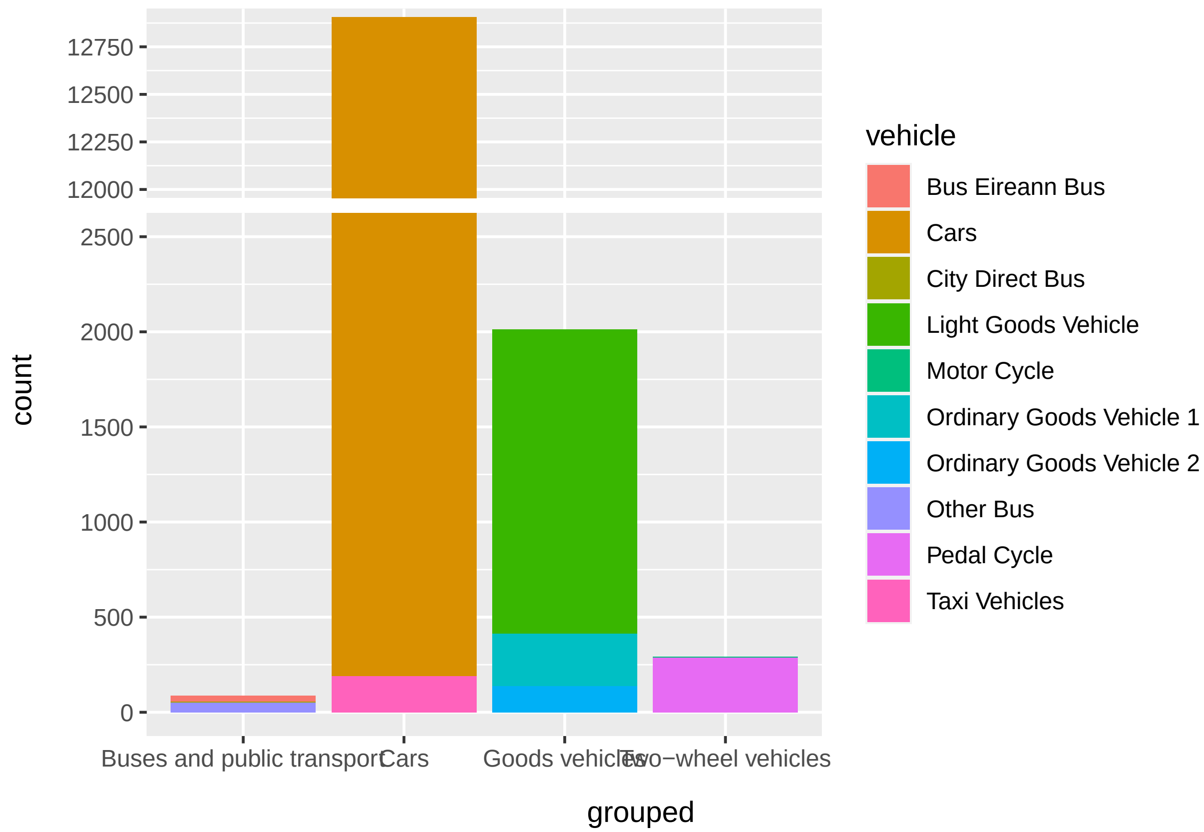
For the visualization of proportion of vehicle categories I chose two plots, the first of which is a multi-layered pie chart. This was more so an experiment. I believe this plot could work very well for Q4, if not for the scaling issue between cars and the other vehicle groups. As seen in the plot, we have 4 distinct radial layers, each representing a given vehicle group. We could accurately contrast the proportion of these categories and subcategories if the scale wasn't blown off by the count of cars (Above 12,500. Next highest is below 2500). Due to this issue I implemented a grouped and stacked bar chart, with appropriate use of a break on the count. Each vehicle group has its own bar, then to represent the sub category, I can simple stack the given vehicles within their group. With the break in the cars category, the rest of the data is saved by the scaling up to the high count of cars. We can clearly make out the distribution of both grouped vehicles, and their sub categories. E.g bicycles make up the majority of Two-Wheel vehicles, the same can be said with Light Goods Vehicles and Cars in each of their categories.

```
#Group vehicles into the 4 groups
#Proportion per these groups, and sub categories at the junction over all time.

vehiclesGrouped<- Dataset %>%
  mutate(NumerisedVehicle = case_when(
    vehicle == 'Pedal Cycle' ~ 1,
```

```r
    vehicle == 'Motor Cycle' ~ 2,
    vehicle == 'Cars' ~ 3,
    vehicle == 'Taxi Vehicles' ~ 4,
    vehicle == 'Light Goods Vehicle' ~ 5,
    vehicle == 'Ordinary Goods Vehicle 1' ~ 6,
    vehicle == 'Ordinary Goods Vehicle 2' ~ 7,
    vehicle == 'City Direct Bus' ~ 8,
    vehicle == 'Bus Eireann Bus' ~ 9,
    vehicle == 'Other Bus' ~ 10,
  ))


vehiclesGrouped<- vehiclesGrouped %>%
         mutate(grouped = case_when(NumerisedVehicle == 1| NumerisedVehicle ==
                                     2 ~ 'Two-wheel vehicles',
                             NumerisedVehicle == 3| NumerisedVehicle ==4  ~
                               'Cars',
                             NumerisedVehicle == 5| NumerisedVehicle ==6|
                               NumerisedVehicle ==7 ~ 'Goods vehicles',
                             NumerisedVehicle == 8| NumerisedVehicle ==9|
                               NumerisedVehicle ==10 ~
                               'Buses and public transport'))




ggplot(vehiclesGrouped, aes(x = grouped, y = count, fill = vehicle)) +
          geom_col() +
          coord_polar("y")
```




```r
ggplot(vehiclesGrouped, aes(x = grouped, y = count, fill = vehicle)) +
          geom_col() +
          scale_y_break(c(2500, 12000)) +
          coord_polar("y")
```



12750
12500
12250
12000
2500
2000
1500
1000
500
0
count
Buses and public transport
Cars
Goods vehicles
Two−wheel vehicles
grouped
vehicle
Bus Eireann Bus
Cars
City Direct Bus
Light Goods Vehicle
Motor Cycle
Ordinary Goods Vehicle 1
Ordinary Goods Vehicle 2
Other Bus
Pedal Cycle
Taxi Vehicles