Activity_1

5645242_Daniel_Guo

2025-02-05

1

https://www.kaggle.com/datasets/abdelrahmanemad594/premier-league-season-2024

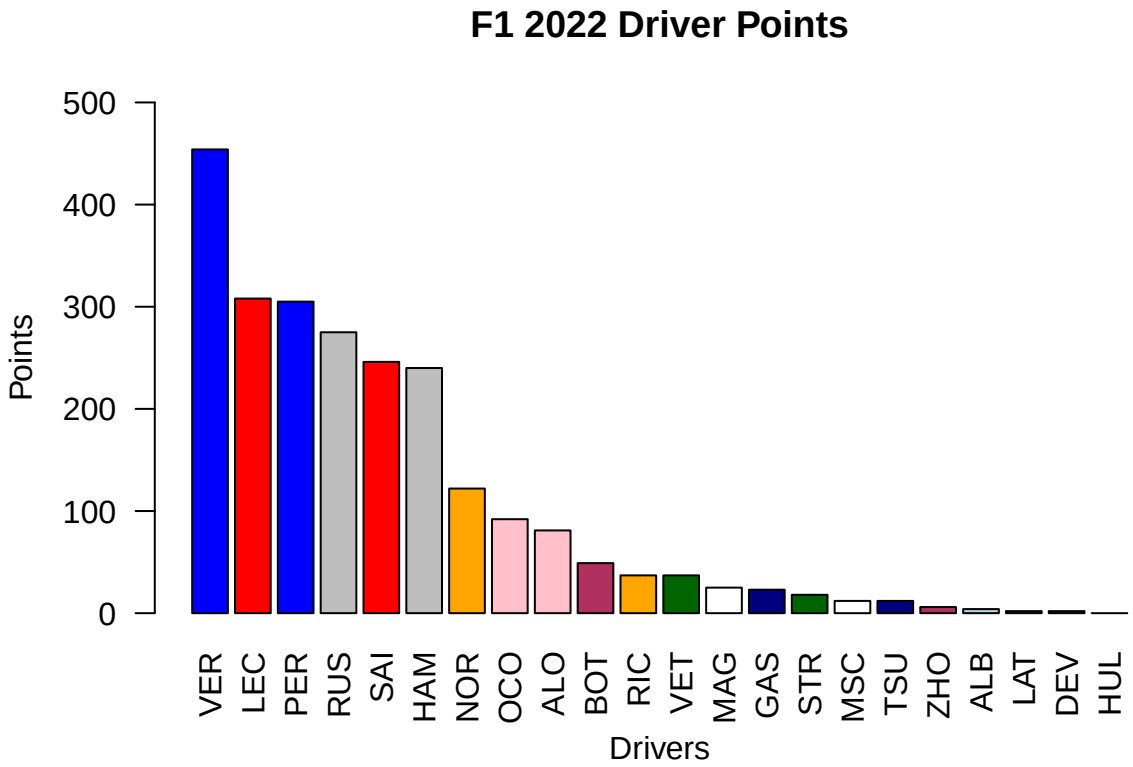
```
F1_2022 <- read.csv(
  "/Users/danielguo/Desktop/University/Year 1/ST117/Activity 1/F1_2022_Data.csv",
  sep = ",")
F1_2022
```

```
##    POS Driver.Code      Driver.Name                  Constructor Points
## 1    1         VER   Max Verstappen         Red Bull Racing RBPT    454
## 2    2         LEC  Charles Leclerc                     Ferrari    308
## 3    3         PER     Sergio Perez         Red Bull Racing RBPT    305
## 4    4         RUS   George Russell                    Mercedes    275
## 5    5         SAI     Carlos Sainz                     Ferrari    246
## 6    6         HAM   Lewis Hamilton                    Mercedes    240
## 7    7         NOR     Lando Norris              McLaren Mercedes    122
## 8    8         OCO     Esteban Ocon               Alpine Renault     92
## 9    9         ALO  Fernando Alonso               Alpine Renault     81
## 10  10         BOT  Valtteri Bottas           Alfa Romeo Ferrari     49
## 11  11         RIC Daniel Ricciardo              McLaren Mercedes     37
## 12  12         VET Sebastian Vettel Aston Martin Aramco Mercedes     37
## 13  13         MAG  Kevin Magnussen                  Haas Ferrari     25
## 14  14         GAS     Pierre Gasly              AlphaTauri RBPT     23
## 15  15         STR     Lance Stroll Aston Martin Aramco Mercedes     18
## 16  16         MSC  Mick Schumacher                  Haas Ferrari     12
## 17  17         TSU     Yuki Tsunoda              AlphaTauri RBPT     12
## 18  18         ZHO      Zhou Guanyu           Alfa Romeo Ferrari      6
## 19  19         ALB  Alexander Albon            Williams Mercedes      4
## 20  20         LAT  Nicholas Latifi            Williams Mercedes      2
## 21  21         DEV    Nyck De Vries            Williams Mercedes      2
## 22  22         HUL  Nico Hulkenberg Aston Martin Aramco Mercedes      0
##    Pole.Positions No.of.Fastest.Laps Wins Podiums DNFs
## 1               7                  5   15      17    2
## 2               9                  3    3      11    3
## 3               1                  3    2      11    3
## 4               2                  4    1       8    1
## 5               3                  2    1       9    6
## 6               0                  2    0       9    2
## 7               0                  2    0       1    2
## 8               0                  0    0       0    2
## 9               0                  0    0       0    6
## 10              0                  0    0       0    6
```

```
## 11              0                   0    0       0    3
## 12              0                   0    0       0    3
## 13              0                   0    0       0    5
## 14              0                   0    0       0    3
## 15              0                   0    0       0    3
## 16              0                   0    0       0    3
## 17              0                   0    0       0    6
## 18              0                   1    0       0    5
## 19              0                   0    0       0    5
## 20              0                   0    0       0    5
## 21              0                   0    0       0    0
## 22              0                   0    0       0    0
```

```
Constructor_colors <- c(
  "Red Bull Racing RBPT" = "blue",
  "Mercedes" = "grey",
  "Ferrari" = "red",
  "McLaren Mercedes" = "orange",
  "Alpine Renault" = "pink",
  "Aston Martin Aramco Mercedes" = "darkgreen",
  "AlphaTauri RBPT" = "navy",
  "Alfa Romeo Ferrari" = "maroon",
  "Haas Ferrari" = "white",
  "Williams Mercedes" = "lightblue"
)

barplot(F1_2022$Points,
        names.arg = F1_2022$Driver.Code,
        main = "F1 2022 Driver Points",
        xlab = "Drivers",
        ylab = "Points",
        col = Constructor_colors[F1_2022$Constructor],
        las = 2,
        border = "black",
        ylim = c(0,500))
```



I wanted to show the number of points each driver won across the 2022 season, while having the bar chart in the colours of the constructors also reflects the number of points each constructors won across this season.

Unlike a histogram, a bar chart is perfect for categorical data such as each driver.It clearly shows who won the 2022 season with the ranking of every drivers.

Clearly, Max Verstappen and its constructor won the 2022 season with a distant margin, followed by Leclerc and his constructor Ferrari. The top three constructors lead the rest of the pack by a very large margin, difference between Hamilton and Norris.

2.

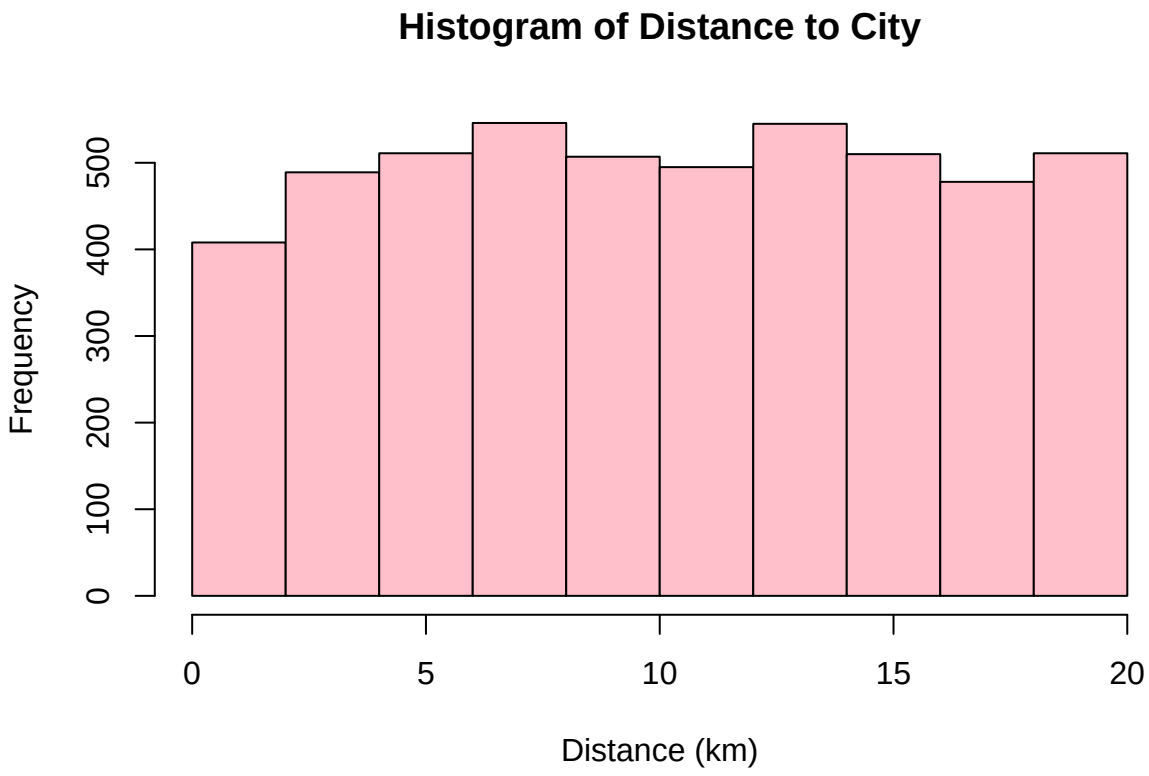
https://www.kaggle.com/datasets/vivekattri/global-ev-charging-stations-dataset

```
EV_Charging_Station <- read.csv(
  "/Users/danielguo/Desktop/University/Year 1/ST117/Activity 1/detailed_ev_charging_stations.csv",
  sep = ",")
head(EV_Charging_Station)
```

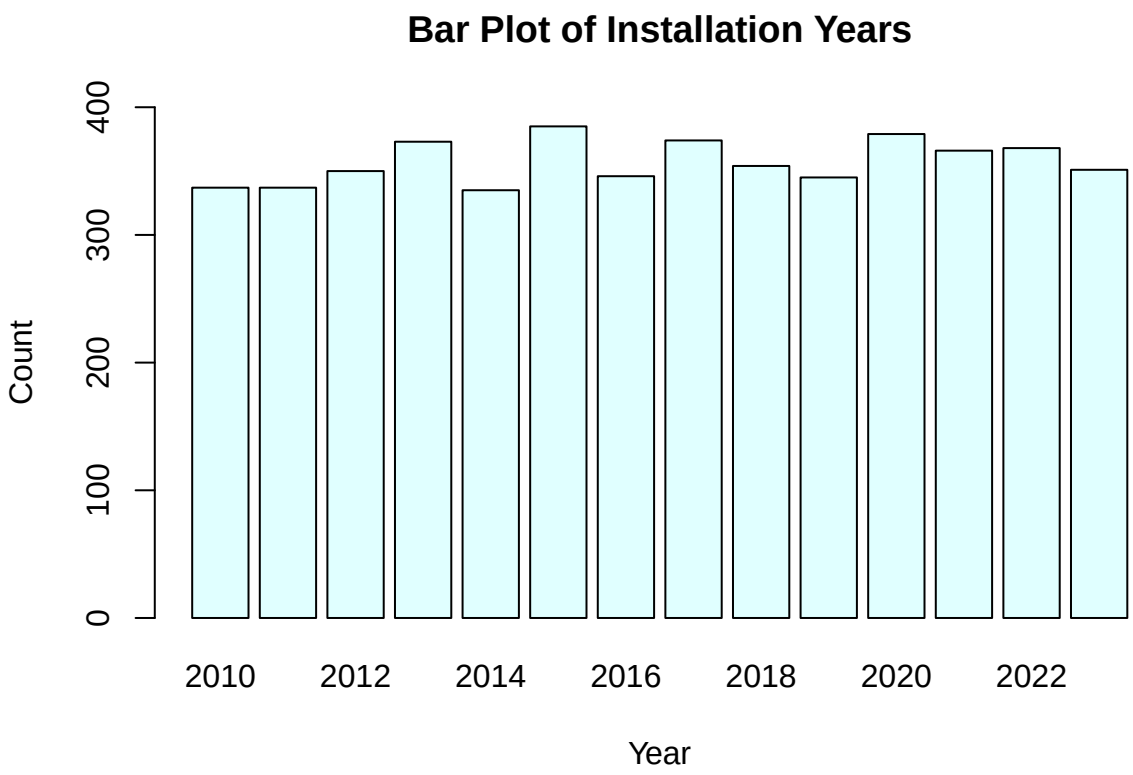
```
##   Station.ID  Latitude  Longitude                              Address
## 1   EVS00001 -33.40100   77.97497      4826 Random Rd, City 98, Country
## 2   EVS00002  37.86186 -122.49030 8970 San Francisco Ave, San Francisco
## 3   EVS00003  13.77609  100.41278             5974 Bangkok Ave, Bangkok
## 4   EVS00004  43.62825  -79.46894             6995 Toronto Ave, Toronto
## 5   EVS00005  19.11987   72.91337               5704 Mumbai Ave, Mumbai
## 6   EVS00006 -23.69501  -46.54819         1545 São Paulo Ave, São Paulo
##      Charger.Type Cost..USD.kWh. Availability Distance.to.City..km.
## 1      AC Level 2           0.27   9:00-18:00                 4.95
## 2 DC Fast Charger           0.19         24/7                 4.96
## 3      AC Level 2           0.48   6:00-22:00                 8.54
## 4      AC Level 1           0.41   9:00-18:00                13.28
```

```
## 5      AC Level 2           0.11   9:00-18:00                  9.76
## 6      AC Level 2           0.22   6:00-22:00                  2.02
##   Usage.Stats..avg.users.day. Station.Operator Charging.Capacity..kW.
## 1                          35            EVgo                    350
## 2                          83            EVgo                    350
## 3                          24      ChargePoint                     50
## 4                          70        Greenlots                    350
## 5                          19            EVgo                    350
## 6                          39          Ionity                    350
##   Connector.Types Installation.Year Renewable.Energy.Source Reviews..Rating.
## 1    CCS, CHAdeMO              2013                     Yes              4.0
## 2   Tesla, Type 2              2010                     Yes              3.9
## 3     Type 2, CCS              2019                      No              3.6
## 4          Type 2              2010                     Yes              4.2
## 5             CCS              2015                     Yes              3.7
## 6           Tesla              2016                     Yes              3.2
##   Parking.Spots Maintenance.Frequency
## 1             7              Annually
## 2             2               Monthly
## 3             9              Annually
## 4             7               Monthly
## 5             6              Annually
## 6             9             Quarterly
```

```r
hist(EV_Charging_Station$Distance.to.City..km.,
     main = "Histogram of Distance to City",
     xlab = "Distance (km)",
     col = "pink",
     breaks = 10)
```

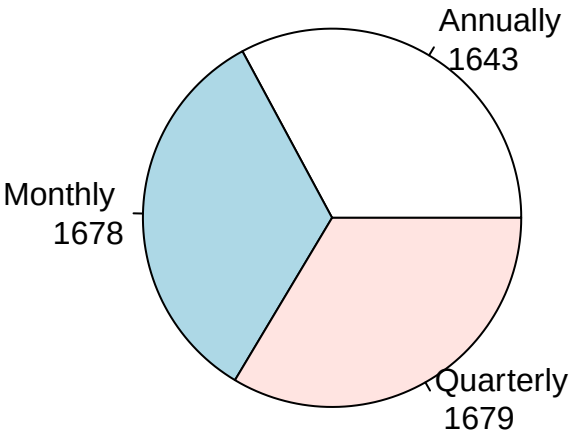


```r
barplot(table(EV_Charging_Station$Installation.Year),
        main = "Bar Plot of Installation Years",
        xlab = "Year",
        ylab = "Count",
        col = "lightcyan",
        border = "black",
        ylim = c(0,400))
```

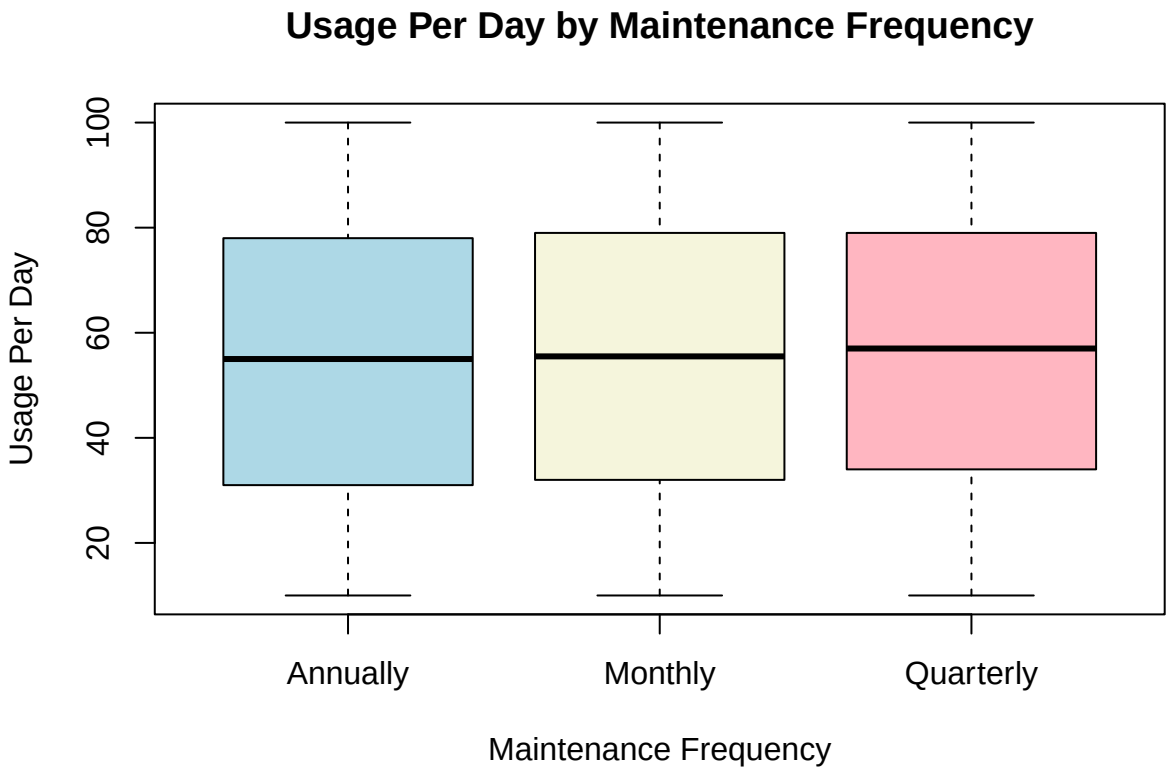



```r
maintenance_table <- table(EV_Charging_Station$Maintenance.Frequency)
pie(maintenance_table,
    main = "Pie Chart of Maintenance Frequency",
    labels = paste(names(maintenance_table), "\n", maintenance_table))
```

Pie Chart of Maintenance Frequency




```
boxplot(EV_Charging_Station$Usage.Stats..avg.users.day. ~ EV_Charging_Station$Maintenance.Frequency,
        main = "Usage Per Day by Maintenance Frequency",
        xlab = "Maintenance Frequency",
        ylab = "Usage Per Day",
        col = c("lightblue", "beige", "lightpink"),
        border = "black")
```



3.

The written answer for the small data set is in section 1, under the first plot.

i.

1. Distance to cities show the availability of the EV stations to citizens in suburban area, travelling between cities. A wide range of distance can be essential for the use of EV and can encourage more EV consumptions.
2. Installation years show the year which the EV charging stations are installed, which could indicate the how recency and the quality of the charging stations.
3. Maintenance frequency shows the operational reliability of EV stations.
4. How usage per day may differ with different maintenance frequency. Further, box plot shows if most stations are used (if the median is high, many stations have high usage.) Further, a long lower whisker can show that some are under-utilised.

ii.

1. Histogram works best for numerical ranges, like distances in km). It shows urban rural comparison by showing where charging stations are mostly located. Useful for policy decisions or expanding infrastructure in areas with less stations. It can detects outliers: if there are stations located extremely far from cities.
2. Unlike a histogram, a bar chart is perfect for categorical data like each year. It can show the trend over the years.
3. Pie chart shows the proportion of each maintenance frequency, can show if most stations need frequent or rare maintenance.
4. A series of box plot allows you to compare how the distribution of usage per day differs, including the median and max and min values, based on maintenance frequency. Median represents the middle value of charging station usage, a higher of which shows a higher use of the stations with the corresponding maintenance frequency.

iii.

1. Relative similar number of stations at each distance except marginally less closser to the city, 0-2km. However, no data outside of 20km radius of the city, therefore more infrastructures may need to be planned in areas 20km outside of the cities.
2. There is no trend of increase or decrease across the years, although several peaks in 2013, 2015, 2017, 2020, which is surprising because the increasing risks of climate change and increase encouragement of EV usage.
3. Even spread of each Maintenance Frequency, which indicates that there is a relatively large proportion that are less maintenance, so that some of them may require more maintenance.
4. Even whisker and no outlier, so that no stations are extremely under-utilised. Annually and Monthly have relatively similar box plot, while quarterly has a higher median, so that the stations has more use, which is unusual.