R Notebook

Import libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.2      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Load Data

```
df <- read.csv("./assignment (8).csv")

head(df)
```

```
##   playground   type  SES     avgPb    avgAs     avgZn   avgSr   avgCd
## 1        h15  mulch high   8.24875  5.10750   129.980  23.530 5.00625
## 2        h15 rubber high  45.39000  7.76625  9722.780  57.545 6.45750
## 3        h15   soil high 336.60000 34.64875   627.575 114.080 8.71500
## 4        h20 rubber high  13.40750  5.74500 16312.205  90.430 7.24500
## 5        h20   sand high  13.64000 10.26625    20.830  95.400 9.32250
## 6        h20   soil high  40.61500 44.62625    46.135 107.505 8.80500
##   peeling_paint rubber_condition num_homes num_homes_paint distance EJ_criteria
## 1             0                2         1               0 73.14203           0
## 2             0                2         1               0 81.73490           0
## 3             0                2         1               0 87.05013           0
## 4             0                0         2               0 99.04874           0
## 5             0                0         2               0 99.04874           0
## 6             0                0         2               0 99.04874           0
##                            notes
## 1
## 2 Some damaged sections of rubber
## 3
## 4
## 5                     outside box
## 6
```

```
str(df)
```

```
## 'data.frame':    85 obs. of  15 variables:
##  $ playground      : chr  "h15" "h15" "h15" "h20" ...
##  $ type            : chr  "mulch" "rubber" "soil" "rubber" ...
##  $ SES             : chr  "high" "high" "high" "high" ...
##  $ avgPb           : num  8.25 45.39 336.6 13.41 13.64 ...
##  $ avgAs           : num  5.11 7.77 34.65 5.75 10.27 ...
##  $ avgZn           : num  130 9722.8 627.6 16312.2 20.8 ...
##  $ avgSr           : num  23.5 57.5 114.1 90.4 95.4 ...
##  $ avgCd           : num  5.01 6.46 8.71 7.25 9.32 ...
##  $ peeling_paint   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ rubber_condition: int  2 2 2 0 0 0 0 0 0 0 ...
##  $ num_homes       : int  1 1 1 2 2 2 3 3 3 0 ...
##  $ num_homes_paint : int  0 0 0 0 0 0 1 1 1 0 ...
##  $ distance        : num  73.1 81.7 87.1 99 99 ...
##  $ EJ_criteria     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ notes           : chr  " " "Some damaged sections of rubber" " " " " ...
```

As per the metadata, playground, type, SES, peeling_paint, rubber_condition, EJ_criteria are categorical variable so changing it to factors would be appropriate.

```
df$playground <- as.factor(df$playground)
df$type <- as.factor(df$type)
df$SES <- as.factor(df$SES)
df$peeling_paint <- as.factor(df$peeling_paint)
df$rubber_condition <- as.factor(df$rubber_condition)
df$EJ_criteria <- as.factor(df$EJ_criteria)

str(df)
```

```
## 'data.frame':    85 obs. of  15 variables:
##  $ playground      : Factor w/ 28 levels "h15","h20","h3",..: 1 1 1 2 2 2 3 3 3 4 ...
##  $ type            : Factor w/ 4 levels "mulch","rubber",..: 1 2 4 2 3 4 1 2 4 1 ...
##  $ SES             : Factor w/ 3 levels "high","low","medium": 1 1 1 1 1 1 1 1 1 1 ...
##  $ avgPb           : num  8.25 45.39 336.6 13.41 13.64 ...
##  $ avgAs           : num  5.11 7.77 34.65 5.75 10.27 ...
##  $ avgZn           : num  130 9722.8 627.6 16312.2 20.8 ...
##  $ avgSr           : num  23.5 57.5 114.1 90.4 95.4 ...
##  $ avgCd           : num  5.01 6.46 8.71 7.25 9.32 ...
##  $ peeling_paint   : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
##  $ rubber_condition: Factor w/ 3 levels "0","1","2": 3 3 3 1 1 1 1 1 1 1 ...
##  $ num_homes       : int  1 1 1 2 2 2 3 3 3 0 ...
##  $ num_homes_paint : int  0 0 0 0 0 0 1 1 1 0 ...
##  $ distance        : num  73.1 81.7 87.1 99 99 ...
##  $ EJ_criteria     : Factor w/ 4 levels "0","1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
##  $ notes           : chr  " " "Some damaged sections of rubber" " " " " ...
```

Summary

```
df.grouping <- df %>% group_by(SES, type) %>%
                   summarise(avg_pb = mean(avgPb),
                             .groups = 'drop')

df.grouping
```

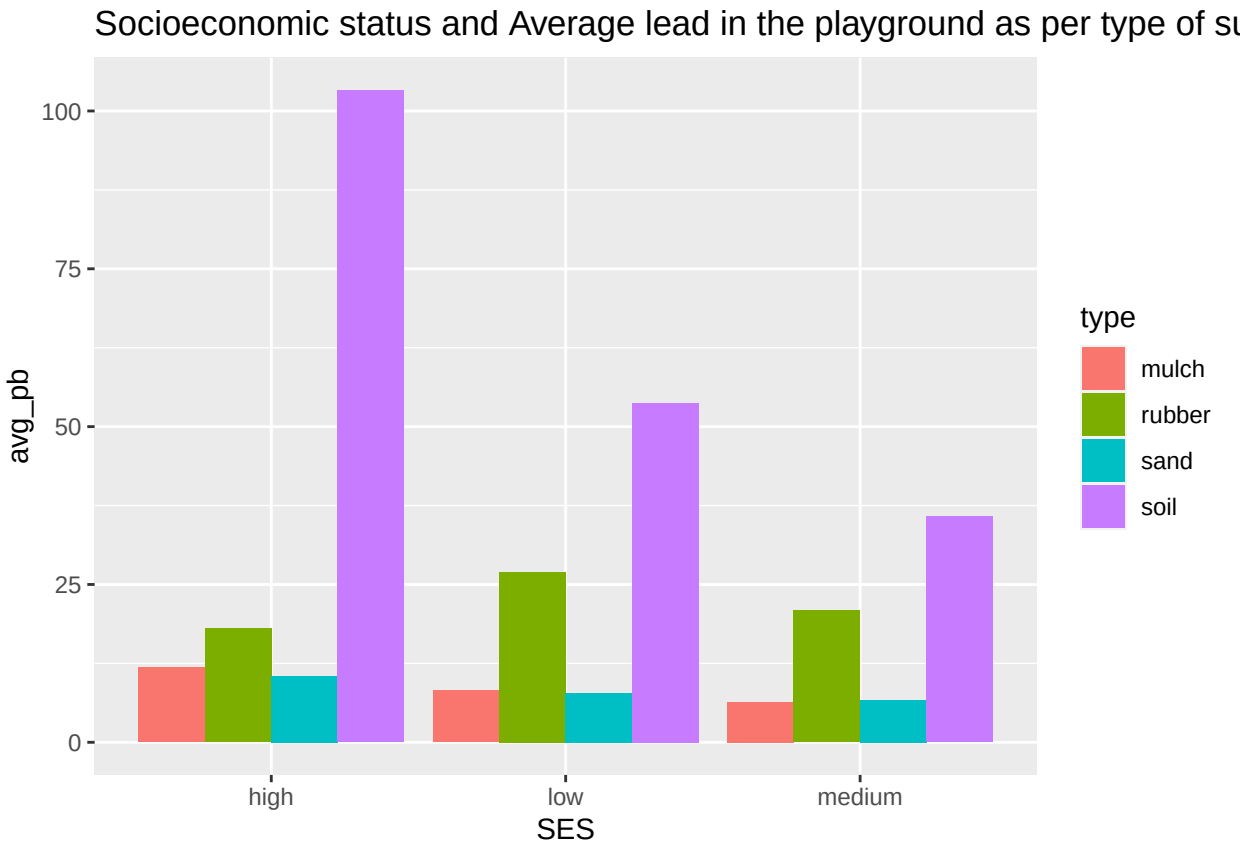
```
## # A tibble: 12 x 3
##    SES    type   avg_pb
##    <fct>  <fct>   <dbl>
##  1 high   mulch   11.8
##  2 high   rubber  18.1
##  3 high   sand    10.5
##  4 high   soil   103.
##  5 low    mulch    8.23
##  6 low    rubber  27.0
##  7 low    sand     7.83
##  8 low    soil    53.7
##  9 medium mulch    6.35
## 10 medium rubber  20.8
## 11 medium sand     6.74
## 12 medium soil    35.8
```

Question

Is the amount of lead present in playground surface influenced by the socioeconomic status of the neighborhood and the type of surface of the playground?

Part A

```
ggplot(data=df.grouping, aes(x=SES, y=avg_pb, fill=type)) +
geom_bar(stat="identity", position=position_dodge()) + ggtitle("Socioeconomic status and Average lead i
```



PART B

We have created a bar graph with SES on the x-axis and avg_pb (Average lead) on the y-axis. We have used the type of surface as the filler to identify average lead in different surface type as per their SES classification i.e., high, medium, and low. From the above graph we can see the maximum lead is in surface is for type soil in every category of SES. Though it is highest i.e. 103 ug/g in SES category high. We can also see that the second highest position goes to SES category low for surface type soil with an average lead of 53.72 ug/g. Medium SES has the lowest lead in all the surface type categories, for comparison surface type soil has 35.7 ug/g. In conclustion, SES condition of high and low which are the extremes have high impact on the avgerage lead in the surface specially in surface type soil.