

1) Write a R script that visualizes different perspectives of the data we covered during the lecture, using Yelp dataset of 'business,' and 'users.'

```
library(ggplot2)

print(users_yelp_dataset)

ggplot(users_yelp_dataset) +
  geom_bar(aes(x=yelping_since), fill="gray")
```

This function is used to find the correlation between the number of reviews and number of fans

```
print(cor(users_yelp_dataset$review_count,
users_yelp_dataset$fans))
```

Training a linear regression model and viewing the co-efficients

```
my.lm = lm(fans ~ review_count + fans,
data=users_yelp_dataset)

coeffs = coefficients(my.lm)

print(coeffs)
```

```
ggplot(users_yelp_dataset) +  
geom_bar(aes(x=average_stars),  
fill="gray")
```

Function for doing clustering on the user dataset

```
userClusters =  
kmeans(users_yelp_dataset[,c(3,11)], 4)
```

Visualising the clusters

```
ggplot(users_yelp_dataset,  
aes(average_stars, fans,  
color=userClusters$cluster)) +  
geom_point()
```

2) Submit a screenshot of the generated visualization, as in a 'jpg' or 'png' file format.



