INFSCI 2809: Spatial Data Analytics Project 1 Due: January 31, 2019

Spatial Data Manipulation Obtain the file "pgh_streets.shp", which is one of the files in "pgh_streets.zip", and:

a. Find the total number of road segments. [5 points]

Output:

```
length(streets)
[1] 22222
```

b. Calculate minimum, maximum, and mean segment lengths. [5 points]

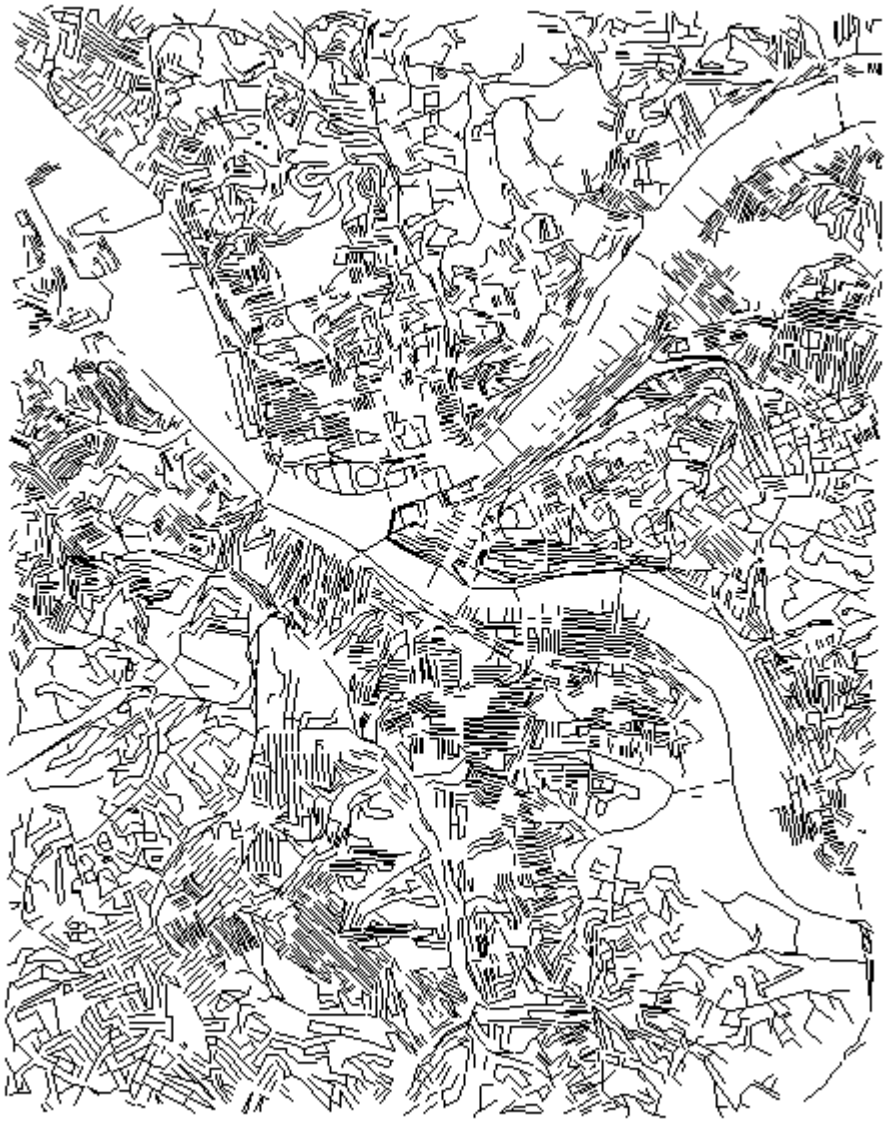
```
# Min, max, and mean segment lengths (in miles)
min(streets$LENGTH)
max(streets$LENGTH)
mean(streets$LENGTH)
```

Output:

```
> min(streets$LENGTH)
[1] 3e-04
> max(streets$LENGTH)
[1] 1.46654
> mean(streets$LENGTH)
[1] 0.05979852
```

c. Filter out the segments that are below the mean length that you calculated in (b) and then create a map showing the remaining segments. [5 points]

```
# Filter out segments below mean length and create map showing
greater_mean=streets[streets$LENGTH>=mean(streets$LENGTH),]
plot(greater_mean)
```

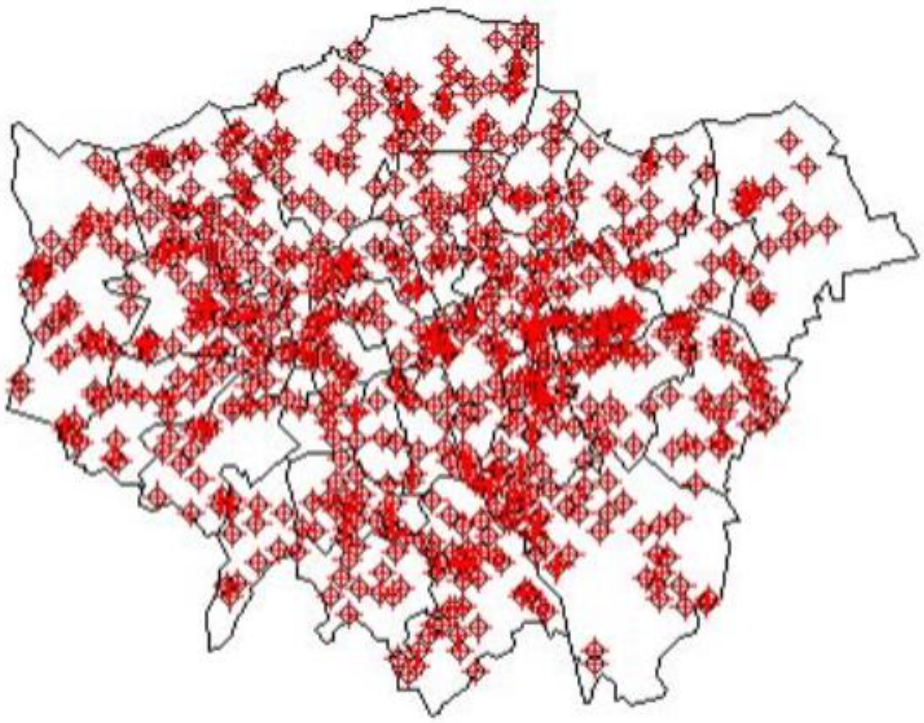


Points: 25 [10 points for R script]

Spatial Data Aggregation Use the data in “lnd.RData” and “stations.RData” and aggregate the values contained in points (stations) to correspond with polygons (lnd) and:

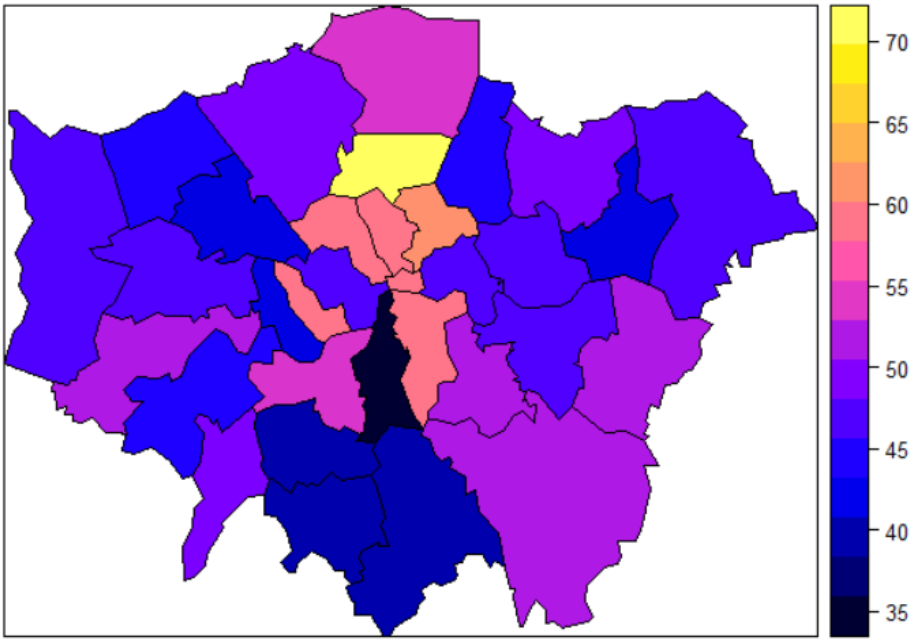
a. Aggregate the data based on the mean of the point values. Create a map and prepare a report on the result. [10 points]

```
load("C:/Users/rvais/Wdirectory/SPA/lnd.RData")
load("C:/Users/rvais/Wdirectory/SPA/stations.RData")

plot(lnd)
points(stations)
plot(stations,col=”red”,add=TRUE)
```

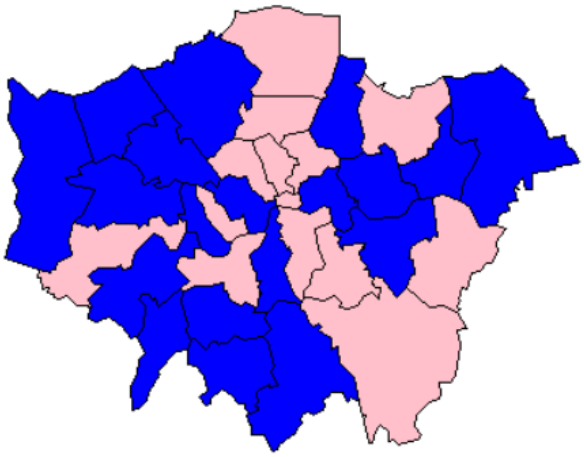


```
# Mean in each polygon
st1<-stations[c('Value')]
st_X=stations[c('coords.x1')]
st_Y=stations[c('coords.x2')]
Xaggr=aggregate(st_X, lnd, mean)
```

```
Yaggr=aggregate(st_Y, lnd, mean)
aggr=aggregate(st1,lnd, mean)
data=data.frame(aggr, Xaggr, Yaggr)

# Plot aggregated values
spplot(aggr)
```

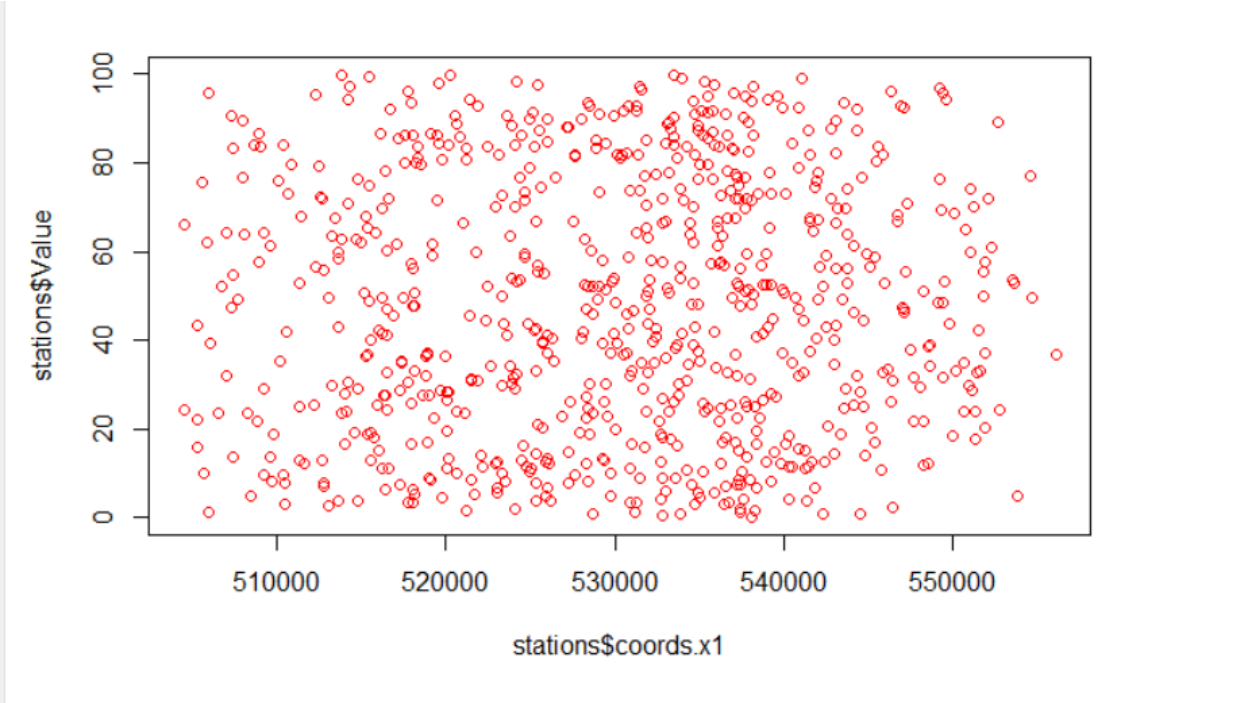


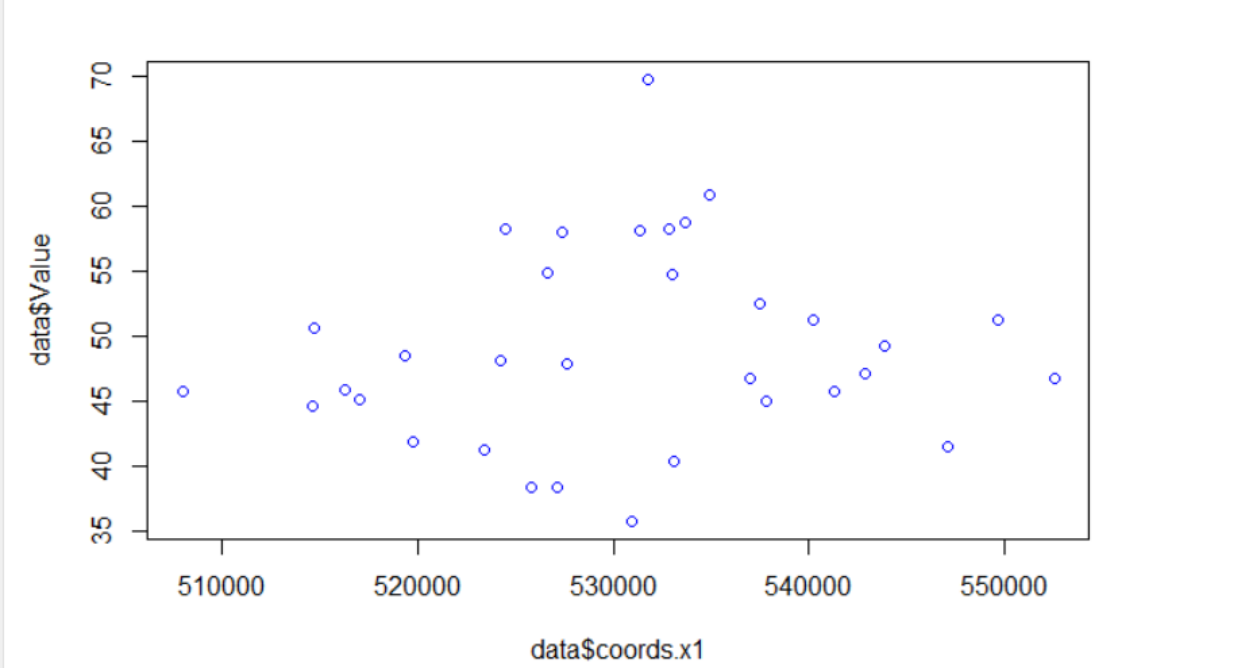

```
#OR
### Regions with below average values are blue; regions with above values are pink.
plot(aggr, col=ifelse(aggr$Value<mean(aggr$Value),'blue','pink'))
```



b. Run regression on the point values before and after aggregation. Prepare a report on the result [20 points] Points: 40 [10 points for R script]

```
 # Plot before regression to display effect of aggregation
#Plotting co-ordinate x1 against value before and after aggregation
plot(x=stations$coords.x1, y=stations$Value,col='red')
plot(x=data$coords.x1, y=data$Value,col='blue')
```

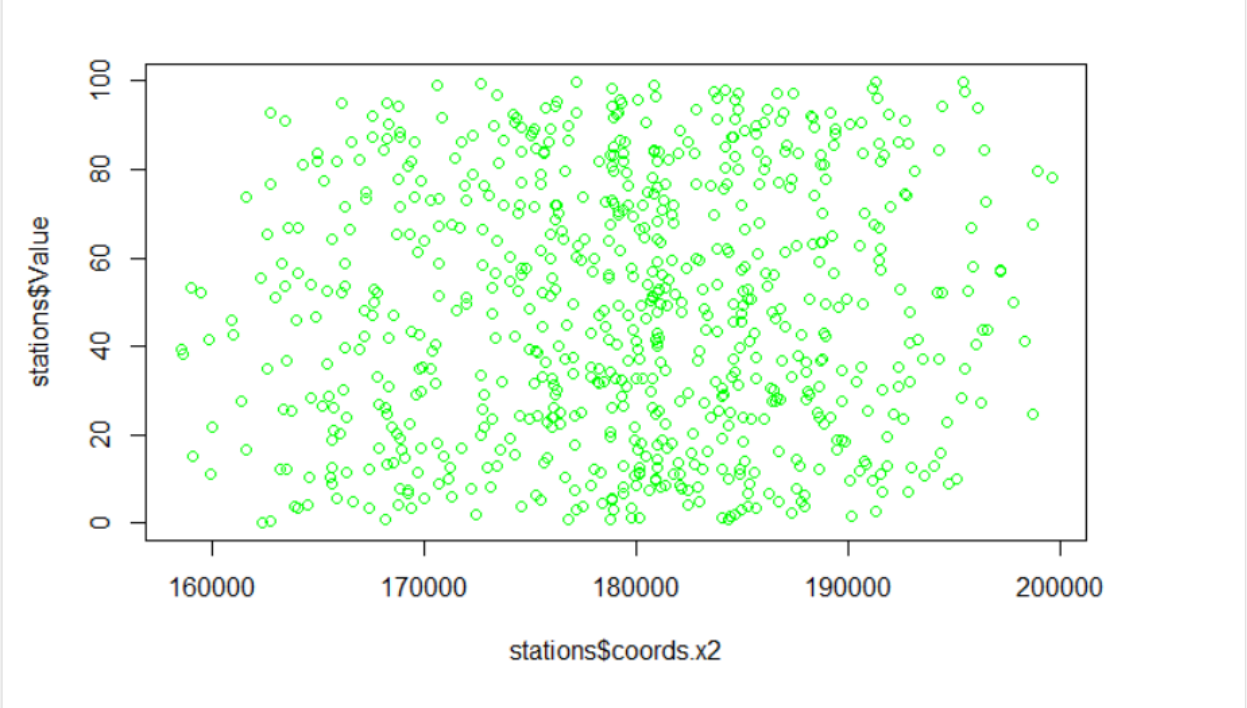


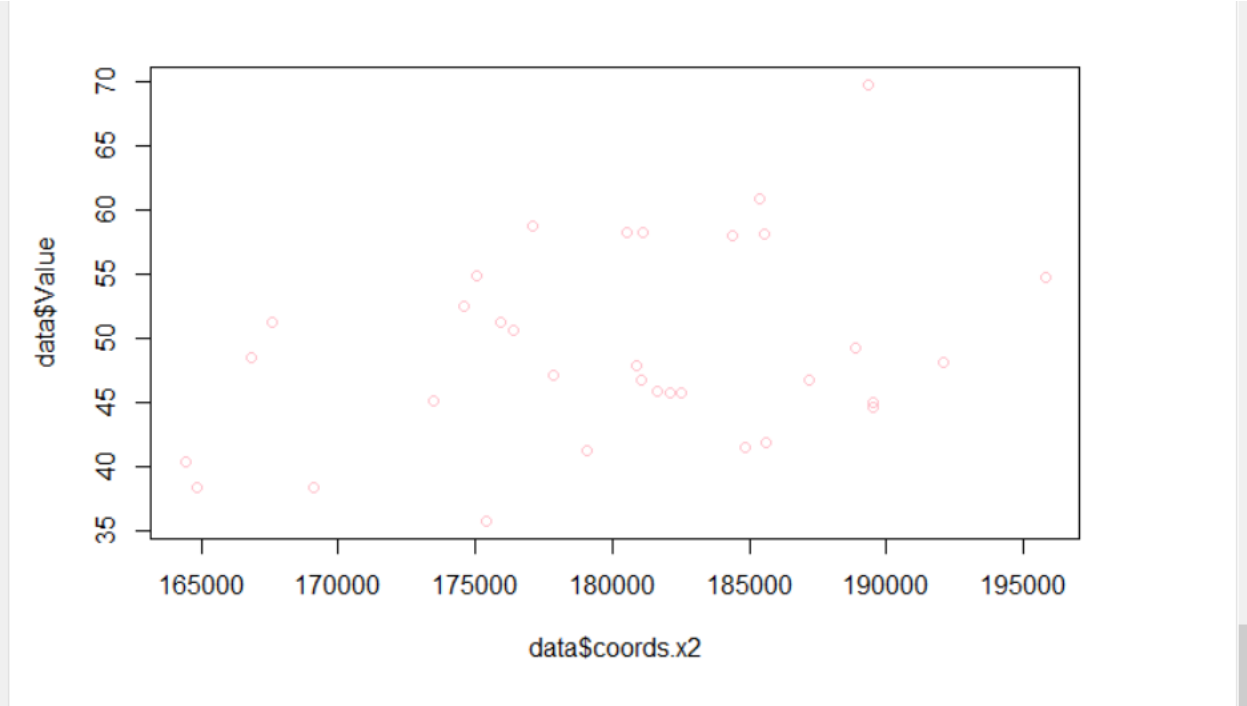


#Plotting co-ordinate x2 against value before and after aggregation

plot(x=stations$coords.x2, y=stations$Value,col='green')

plot(x=data$coords.x2, y=data$Value,col='pink')





Regressions before and after aggregation;After applying lm the,effect is stronger after aggregation.

#Regression before applying aggregation functions for co-ordinate x1

summary(lm(data=stations, Value~coords.x1))

#Regression after applying aggregation functions for co-ordinate x1

summary(lm(data=data, Value~coords.x1))

#Regression before applying aggregation functions for co-ordinate x2

summary(lm(data=stations, Value~coords.x2))

#Regression after applying aggregation functions for co-ordinate x2

summary(lm(data=data, Value~coords.x2))

```
Call:
lm(formula = Value ~ coords.x1, data = stations)

Residuals:
    Min      1Q  Median      3Q     Max
-48.289 -24.065  -1.048  25.178  53.529

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.029e-01  4.763e+01   0.002    0.998
coords.x1   8.997e-05  8.981e-05   1.002    0.317

Residual standard error: 28.63 on 729 degrees of freedom
Multiple R-squared:  0.001375,  Adjusted R-squared:  4.914e-06
F-statistic: 1.004 on 1 and 729 DF,  p-value: 0.3168


Call:
lm(formula = Value ~ coords.x1, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-13.368  -4.244  -1.050   5.499  20.576
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 8.705e+00  6.686e+01   0.130    0.897
coords.x1   7.617e-05  1.260e-04   0.605    0.550

Residual standard error: 7.649 on 31 degrees of freedom
Multiple R-squared:  0.01165,   Adjusted R-squared:  -0.02023
F-statistic: 0.3654 on 1 and 31 DF,  p-value: 0.5499


Call:
lm(formula = Value ~ coords.x2, data = stations)

Residuals:
    Min      1Q  Median      3Q     Max
-48.221 -24.557  -0.718  25.237  53.152

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.275e+01  2.183e+01   0.584    0.560
coords.x2   1.954e-04  1.215e-04   1.608    0.108

Residual standard error: 28.6 on 729 degrees of freedom
Multiple R-squared:  0.003532,  Adjusted R-squared:  0.002166
F-statistic: 2.584 on 1 and 729 DF,  p-value: 0.1084


Call:
lm(formula = Value ~ coords.x2, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-11.816  -4.792  -1.887   6.138  17.731

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.5009990 29.0595494  -0.293   0.7718
coords.x2    0.0003198  0.0001611   1.985   0.0561 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7.248 on 31 degrees of freedom
Multiple R-squared:  0.1127,    Adjusted R-squared:  0.0841
F-statistic: 3.938 on 1 and 31 DF,  p-value: 0.05611
```

```

By applying regression on the data, we conclude that after applying regression the, effect is stronger after aggregation. The R squared stats also helps us to conclude that effect is increasing in the aggregated model.
```