Analysis

Chathurangi

5/14/2020

Table of Contents

Feature Selection for highest and lowest growth groups

Here we consider the population growth model and identity high positive growth and low positive growth groups. We apply feature detection techiniques to both of these groups

Data preprocessing and Spliting Train and test

```
groups<-read.table("zipfit.csv",header= T, stringsAsFactors = F, sep=",",fill
=T, quote = "")

head(groups)
```

```
##   X0.10112235885300533 X10040
## 1            0.1045679  11211
## 2            0.1057603  11219
## 3            0.1068443  11367
## 4            0.1099202  11205
## 5            0.1111075  11230
## 6            0.1118197  11204
```

```
colnames(groups)<-c("growth", "ZipCode")
groups<-groups[order(groups$growth),]
#groups$ZipCode<-factor(groups$ZipCode)

highID<-head(groups,30)
lowID<- tail(groups,30)


feature<-read.table("nycjackpotbyzip.csv",header= T, stringsAsFactors = F, se
p=",",fill=T, quote = "")

DATAFull<-feature[((feature$geo_id %in% highID$ZipCode)|(feature$geo_id %in%
lowID$ZipCode)), ]


#high growth rate =Infect2 and Infect1 otherwise
DATAFull$Infection<-ifelse(DATAFull$geo_id %in% highID$ZipCode, "Infect2","In
fect1" )


#read density data
density<-read.table("popbyzipdensity.csv",header= T, stringsAsFactors = F, se
p=",",fill=T, quote = "")
DATAFull$ZipCode<-DATAFull$geo_id

DATA<-merge(DATAFull,density[,2:5], by="ZipCode" )
```

```
#missing values frequencies
#freq.na(DATA)

#                                                          missing   %
#pop_5_years_over                                              60 100
#pop_15_and_over                                               60 100
#pop_never_married                                             60 100
#pop_now_married                                               60 100
#pop_separated                                                 60 100
#pop_widowed                                                   60 100
#pop_divorced                                                  60 100
#speak_only_english_at_home                                    60 100
#speak_spanish_at_home                                         60 100
#speak_spanish_at_home_low_english                             60 100

#we can remove following factors

DATA<-DATA[,!(colnames(DATA) %in% c("pop_5_years_over","pop_15_and_over","pop
_never_married","pop_now_married","pop_separated","pop_widowed","pop_divorced
","speak_only_english_at_home","speak_spanish_at_home","speak_spanish_at_home
_low_english"))]




# To simulate a train and test set we are going to split randomly this data s
et into 60% train and 40% test.

set.seed(125)
train_index <- sample(1:nrow(DATA), 0.6 * nrow(DATA))
test_index <- setdiff(1:nrow(DATA), train_index)



# Build X_train, y_train, X_test, y_test
x_train <- DATA[train_index, -c(1,2,3,232)]
y_train <- DATA[train_index, "Infection"]

X_test <- DATA[test_index, -c(1,2,3,232)]
y_test <- DATA[test_index, "Infection"]
```

Using support vector machanism

Train the SVM model

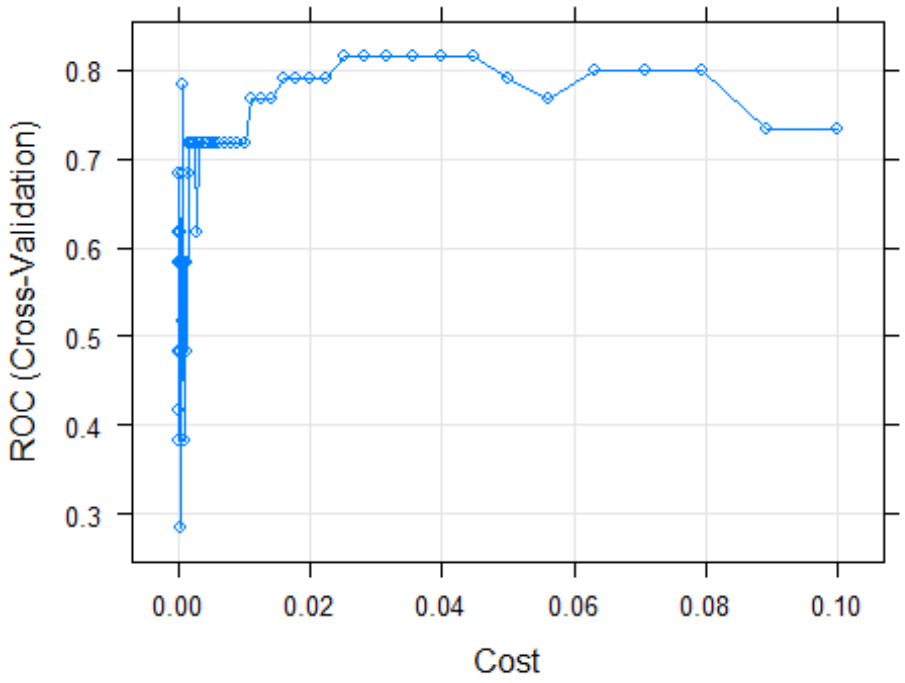
```
#we use support vector machine


fitControl <- trainControl(method="cv",
```

```
                           number=10,
                           classProbs=T,
                           summaryFunction=twoClassSummary)

set.seed(1234) # for reproducible results

## evaluate on train set based on area under the ROC (AUC)


SVM <- train(x=x_train,
             y=y_train ,
             method="svmLinear2",
             trControl=fitControl,
             tuneGrid=expand.grid(cost=10^(seq(-4.5, -1, by = 0.05))),
             metric='ROC')
## summary of performance across each value of tuning parameters
SVM

## Support Vector Machines with Linear Kernel
##
##  36 samples
## 231 predictors
##   2 classes: 'Infect1', 'Infect2'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 33, 33, 31, 32, 32, 33, ...
## Resampling results across tuning parameters:
##
##   cost          ROC        Sens       Spec
##   3.162278e-05  0.6833333  1.0000000  0.00
##   3.548134e-05  0.6166667  1.0000000  0.00
##   3.981072e-05  0.6833333  1.0000000  0.00
##   4.466836e-05  0.6833333  1.0000000  0.00
##   5.011872e-05  0.6833333  1.0000000  0.00
##   5.623413e-05  0.6166667  1.0000000  0.00
##   6.309573e-05  0.5833333  1.0000000  0.00
##   7.079458e-05  0.4833333  1.0000000  0.00
##   7.943282e-05  0.5833333  1.0000000  0.00
##   8.912509e-05  0.4833333  1.0000000  0.00
##   1.000000e-04  0.5833333  1.0000000  0.00
##   1.122018e-04  0.6833333  1.0000000  0.00
##   1.258925e-04  0.4166667  1.0000000  0.00
##   1.412538e-04  0.6166667  1.0000000  0.00
##   1.584893e-04  0.3833333  1.0000000  0.00
##   1.778279e-04  0.5833333  1.0000000  0.00
##   1.995262e-04  0.6166667  1.0000000  0.00
##   2.238721e-04  0.2833333  1.0000000  0.00
##   2.511886e-04  0.4833333  1.0000000  0.00
##   2.818383e-04  0.4833333  1.0000000  0.00
```

```
##   3.162278e-04  0.5833333  1.0000000  0.00
##   3.548134e-04  0.5833333  1.0000000  0.00
##   3.981072e-04  0.5833333  1.0000000  0.00
##   4.466836e-04  0.5833333  1.0000000  0.00
##   5.011872e-04  0.6833333  1.0000000  0.00
##   5.623413e-04  0.7833333  1.0000000  0.00
##   6.309573e-04  0.6833333  1.0000000  0.00
##   7.079458e-04  0.5833333  1.0000000  0.00
##   7.943282e-04  0.5166667  0.9666667  0.00
##   8.912509e-04  0.3833333  0.9666667  0.00
##   1.000000e-03  0.5833333  0.9666667  0.05
##   1.122018e-03  0.5833333  0.9666667  0.05
##   1.258925e-03  0.4833333  0.9666667  0.05
##   1.412538e-03  0.6833333  0.9666667  0.10
##   1.584893e-03  0.7166667  0.9666667  0.15
##   1.778279e-03  0.7166667  0.9666667  0.30
##   1.995262e-03  0.7166667  0.9666667  0.30
##   2.238721e-03  0.7166667  0.9333333  0.30
##   2.511886e-03  0.7166667  0.9333333  0.30
##   2.818383e-03  0.6166667  0.9333333  0.30
##   3.162278e-03  0.7166667  0.9333333  0.30
##   3.548134e-03  0.7166667  0.9333333  0.30
##   3.981072e-03  0.7166667  0.9333333  0.30
##   4.466836e-03  0.7166667  0.8833333  0.30
##   5.011872e-03  0.7166667  0.8333333  0.30
##   5.623413e-03  0.7166667  0.8333333  0.30
##   6.309573e-03  0.7166667  0.8333333  0.30
##   7.079458e-03  0.7166667  0.8333333  0.30
##   7.943282e-03  0.7166667  0.8333333  0.50
##   8.912509e-03  0.7166667  0.7833333  0.60
##   1.000000e-02  0.7166667  0.7833333  0.60
##   1.122018e-02  0.7666667  0.7833333  0.70
##   1.258925e-02  0.7666667  0.7833333  0.70
##   1.412538e-02  0.7666667  0.7833333  0.75
##   1.584893e-02  0.7916667  0.7833333  0.75
##   1.778279e-02  0.7916667  0.7833333  0.75
##   1.995262e-02  0.7916667  0.7833333  0.75
##   2.238721e-02  0.7916667  0.7833333  0.75
##   2.511886e-02  0.8166667  0.7833333  0.75
##   2.818383e-02  0.8166667  0.7833333  0.75
##   3.162278e-02  0.8166667  0.7833333  0.75
##   3.548134e-02  0.8166667  0.8333333  0.75
##   3.981072e-02  0.8166667  0.8333333  0.75
##   4.466836e-02  0.8166667  0.8333333  0.75
##   5.011872e-02  0.7916667  0.8333333  0.75
##   5.623413e-02  0.7666667  0.8333333  0.75
##   6.309573e-02  0.8000000  0.8333333  0.75
##   7.079458e-02  0.8000000  0.8333333  0.75
##   7.943282e-02  0.8000000  0.7833333  0.75
##   8.912509e-02  0.7333333  0.7500000  0.75
```

```
##   1.000000e-01  0.7333333  0.7500000  0.65
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was cost = 0.02511886.
```

```
plot(SVM, metric = "ROC")
```




```
SVM$bestTune
```

```
##          cost
## 59 0.02511886
```

```
SVM$finalModel
```

```
##
## Call:
## svm.default(x = as.matrix(x), y = y, kernel = "linear", cost = param$cost,
##     probability = classProbs)
##
##
## Parameters:
##    SVM-Type:  C-classification
##  SVM-Kernel:  linear
##        cost:  0.02511886
##
## Number of Support Vectors:  26
```

```
#ROC was used to select the optimal model using the largest value.The final v
alue used for the model was cost = 0.02238721
```

Validation of the SVM model

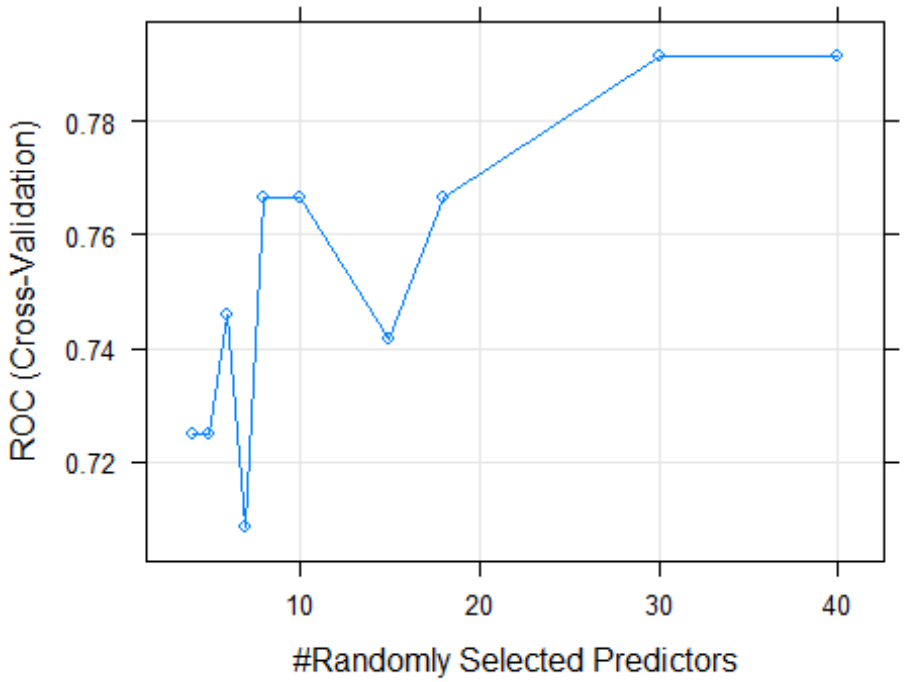
```
## predicting the validation data:
predSVM <- predict(SVM,X_test)

## or predicting using the probabilities (nice because you can get ROC)
probsSVM <- extractProb(list(model=SVM),
                        testX=X_test,
                        testY=y_test)

probsSVM$obs <- probsSVM$obs
probsSVM$pred <- probsSVM$pred

## Calculating Accuracy
mean(probsSVM$obs==probsSVM$pred)

## [1] 0.8333333

## see classification prob for each sample in validation set
## pred column shows model predicted label if cutoff for calling label = 0.5
table(probsSVM$obs, probsSVM$pred) # This is the confusion matrix

##
##           Infect1 Infect2
##   Infect1      30       0
##   Infect2      10      20

## summary of performance result on validation set
twoClassSummary(probsSVM, lev = levels(probsSVM$obs))

##       ROC      Sens      Spec
## 0.9555556 1.0000000 0.6666667

#      ROC      Sens      Spec
#0.9555556 1.0000000 0.7333333
```

Using Random Forest Model

Train the Random Forests Model

```
## RandomForest with 10x cross validation
fitControl <- trainControl(method="cv",
                           number=10,
                           classProbs=T,
                           summaryFunction=twoClassSummary)


set.seed(1234) # for reproducible results
```

```r
## evaluate on train set based on area under the ROC (AUC)
RF <- train(x=x_train,
            y=y_train,
            method="rf",
            trControl=fitControl,
            tuneGrid=expand.grid(mtry=c(4,5,6,7,8, 10, 15, 18,30,40)),
            metric='ROC')
## summary of performance across each value of tuning parameters
RF

## Random Forest
##
##  36 samples
## 231 predictors
##   2 classes: 'Infect1', 'Infect2'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 33, 33, 31, 32, 32, 33, ...
## Resampling results across tuning parameters:
##
##   mtry  ROC        Sens       Spec
##    4    0.7250000  0.8833333  0.55
##    5    0.7250000  0.8833333  0.55
##    6    0.7458333  0.8333333  0.55
##    7    0.7083333  0.8666667  0.50
##    8    0.7666667  0.8833333  0.40
##   10    0.7666667  0.8333333  0.55
##   15    0.7416667  0.8333333  0.50
##   18    0.7666667  0.8333333  0.50
##   30    0.7916667  0.8000000  0.50
##   40    0.7916667  0.8000000  0.50
##
## ROC was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 30.

plot(RF, metric = "ROC")
```




```
RF$bestTune

##   mtry
## 9   30

RF$finalModel

##
## Call:
##  randomForest(x = x, y = y, mtry = param$mtry)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 30
##
##         OOB estimate of  error rate: 33.33%
## Confusion matrix:
##         Infect1 Infect2 class.error
## Infect1      18       4   0.1818182
## Infect2       8       6   0.5714286
```

Validation of the Random Forests model

```
## predicting the validation data:
predRF <- predict(RF,X_test)

## or predicting using the probabilities (nice because you can get ROC)
probsRF <- extractProb(list(model=RF),
                       testX=X_test,
```

```
                    testY=y_test)

## removing trainings data
#probsRF <- probsRF[probsRF$dataType!='Training',]

## Make sure the levels are appropriate for twoClassSummary(), ie case group
is first level
levs <- c("Infect2", "Infect1")
probsRF$obs <- factor(probsRF$obs, levels = levs)
probsRF$pred <- factor(probsRF$pred, levels = levs)

## Calculating Accuracy
mean(probsRF$obs==probsRF$pred)

## [1] 0.8833333

## see classification prob for each sample in validation set
## pred column shows model predicted label if cutoff for calling label = 0.5
table(probsRF$obs, probsRF$pred)

##
##           Infect2 Infect1
##   Infect2      25       5
##   Infect1       2      28

## summary of performance result on validation set
twoClassSummary(probsRF, lev = levels(probsRF$obs))

##       ROC      Sens      Spec
## 0.9700000 0.8333333 0.9333333
```

Summary of models and varaibles importance

Summary of two models

RandomForest

SVM1

AUC

0.97

0.95

Sensitivity

0.83

1

Specitivity

0.93

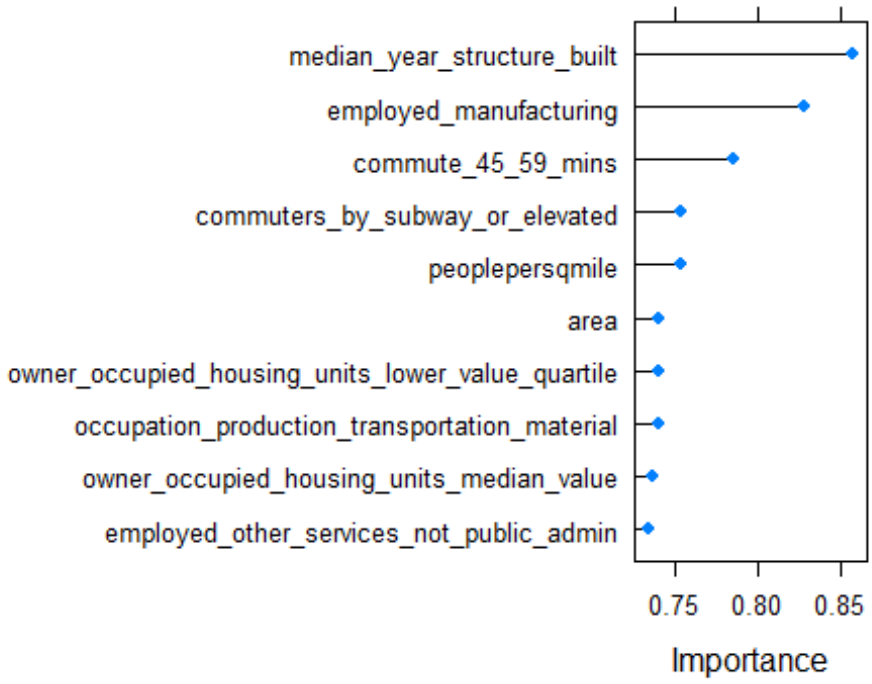
0.73

Variables importance using SVM and Random Forests

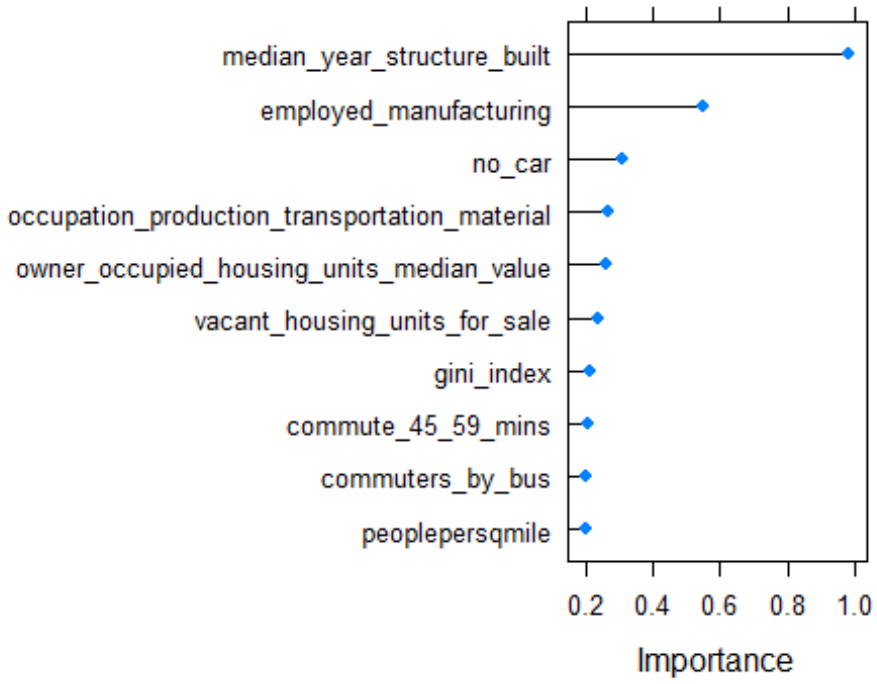
Based on the importance values of SVM and RF, we think following 14 features are most significant for higher infectious rates.

- median_year_structure_built
- employed_manufacturing
- commute_45_59_mins
- commuters_by_subway_or_elevated
- peoplepersqmile
- area
- owner_occupied_housing_units_lower_value_quartile
- occupation_production_transportation_material
- owner_occupied_housing_units_median_value
- employed_other_services_not_public_admin
- no car
- vacant_housing_units
- commuters_by_bus
- gini index

```
#Variable Importance Top 10 variables in SVM

SVMImp <- varImp(SVM, scale = FALSE)
#head(SVMImp)
#plot(SVMImp, top = 10)
plot(SVMImp, top = 10)
```




```r
#Variable Importance Top 10 variables in RF
RFImp <- varImp(RF, scale = FALSE)
#head(RFImp)
#plot(RFImp, top = 20)
plot(RFImp, top = 10)
```




```
#par(mfrow=c(1,2))
#par(mar=c(1, 1, 1, 1))
```

Fitting Logistic models

univarible and multivaraiable models using all 14 different selected features from Random forest models and SVM models

- If we go through the OR of following univaraible and multivaraible models then we can conlclude that only area (OR= 0.41 (CI= 0.19-0.72, p=0.010)) and gini index are significantly affect the high postive rates

```
DATA$InfectionBinary<-ifelse(DATA$Infection=="Infect2",1,0)

DATa<-as.data.frame(DATA)
# Explanatory or confounding variables
explanatory = c("median_year_structure_built","employed_manufacturing","commu
te_45_59_mins","commuters_by_subway_or_elevated","peoplepersqmile","area","ow
ner_occupied_housing_units_lower_value_quartile","owner_occupied_housing_unit
s_median_value","occupation_production_transportation_material","employed_oth
er_services_not_public_admin","no_car","vacant_housing_units","commuters_by_b
us","gini_index")
```

```r
# Explanatory variable of interest
explanatory_multi= c("area","gini_index")

#Dependent Variables
dependent = "InfectionBinary" #


finalfit.glm(DATA,dependent,explanatory, explanatory_multi,metrics = TRUE)
```

```
## [[1]]
##                          Dependent: InfectionBinary
##                        median_year_structure_built        [0.0,1990.0]
##                             employed_manufacturing      [127.0,3160.0]
##                                commute_45_59_mins     [778.0,14863.0]
##                    commuters_by_subway_or_elevated     [312.0,29314.0]
##                                      peoplepersqmile   [4884.2,125797.9]
##                                                 area          [0.6,13.7]
##  owner_occupied_housing_units_lower_value_quartile  [18400.0,705000.0]
##          owner_occupied_housing_units_median_value [27600.0,1073700.0]
##      occupation_production_transportation_material      [878.0,8780.0]
##           employed_other_services_not_public_admin      [323.0,5293.0]
##                                             no_car     [326.0,32435.0]
##                               vacant_housing_units      [321.0,6449.0]
##                                   commuters_by_bus      [776.0,9915.0]
##                                         gini_index           [0.4,0.6]
##       unit     value             OR (univariable)        OR (multivariable
)
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.002)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.01, p=0.001)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.001)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.001)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.016)
-
##  Mean (sd) 0.5 (0.5)    0.54 (0.29-0.84, p=0.024) 0.41 (0.19-0.72, p=0.010
)
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.002)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.001)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.024)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.011)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.003)
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.091)
```

```
-
##  Mean (sd) 0.5 (0.5)    1.00 (1.00-1.00, p=0.007)
-
##  Mean (sd) 0.5 (0.5) 0.09 (0.00-2060.55, p=0.637) 0.00 (0.00-0.47, p=0.047
)
##
## [[2]]
##
##  Number in dataframe = 60, Number in model = 60, Missing = 0, AIC = 75.1,
C-statistic = 0.756, H&L = Chi-sq(8) 3.55 (p=0.896)
```

GLM models using Backward elemination (step AIC)

- Here we used backward eleminination using all around all 232 factors
- The final model according to the step wise AIC criteria includes following 18 different factors
- (male_pop + median_age + male_under_5 + male_5_to_9 + male_22_to_24 + male_25_to_29 + male_30_to_34 + male_40_to_44 + male_45_to_49 + male_50_to_54 + male_55_to_59 + male_75_to_79 + female_55_to_59 + female_62_to_64 + white_pop + black_pop + asian_pop + hispanic_pop)

GLM Model Comparisons

+model1:glm model with only significant parameters obtained from RF and SVM seeems the best model

- area and gini_index are significant parameters

```
#First we normalize data




# glm model with only  significant parameters obtained from RF and SVM
model1<-glm(InfectionBinary~area+gini_index ,data = DATAFF,family = binomial(
link = "logit"),control=glm.control(maxit=50))

# Final GLM models using Backward elemination (step AIC)
model2<-glm(InfectionBinary~ male_pop + median_age + male_under_5 + male_5_to
_9 +
    male_22_to_24 + male_25_to_29 + male_30_to_34 + male_40_to_44 +
    male_45_to_49 + male_50_to_54 + male_55_to_59 + male_75_to_79 +
    female_55_to_59 + female_62_to_64 + white_pop + black_pop +
    asian_pop + hispanic_pop,data = DATAFF,family = binomial(link = "logit"),
control=glm.control(maxit=50))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

#Here first remove outliars
cooksd <- cooks.distance(model2)
```

```r
#In this plot, what seems to be a dark thick black line is actually all our d
ata points. In the right-top corner we see also what seems to be 3 outlier, o
r a bunch of them grouped
outliers <- rownames(DATAFF[cooksd > 4*mean(cooksd, na.rm=T), ])
print(outliers)#only three outliars
```

```
## [1] "1"  "2"  "14"
```

```r
DATAFF<-DATAFF[!(rownames(DATAFF) %in% outliers),]

# Final GLM models using Backward elemination (step AIC)
model2<-glm(InfectionBinary~ male_pop + median_age + male_under_5 + male_5_to
_9 +
    male_22_to_24 + male_25_to_29 + male_30_to_34 + male_40_to_44 +
    male_45_to_49 + male_50_to_54 + male_55_to_59 + male_75_to_79 +
    female_55_to_59 + female_62_to_64 + white_pop + black_pop +
    asian_pop+ hispanic_pop ,data = DATAFF,family = binomial(link = "logit"))
```

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```r
export_summs(model1, model2, exp=TRUE, error_format = "[{conf.low}, {conf.hig
h}], p={p.value}")
```

	Model 1	Model 2
(Intercept)	3218.91 *	Inf
	[3.14, 3304794.26], p=0.02	[0.00, Inf], p=1.00
area	0.41 *	
	[0.21, 0.81], p=0.01	
gini_index	0.00 *	
	[0.00, 0.83], p=0.05	
male_pop		0.69

	[0.00, 1256342 3866148 4998912 0024468 8426604 6800248 4204882 4280062 2442486 2048006 04226.00], p=1.00
median_age	0.00
	[0.00, Inf], p=1.00
male_under_5	2.19
	[0.00, 5172847 0272206 4003124 2660688 4440404 0846600 4064024 0046604 4244028 8062202 0640268 8220642 4026282 4648802 4824464 6828426 6222088 8288066 4608424 6062280 2600066 0406806 2462448 0268680 8402646 0808.00], p=1.00

male_5_to_9	1.64	[0.00, 4357102 2002192 1999200 4426808 0862000 8226622 6246686 6404660 8442440 4888460 2284866 2646204 0844888 6260648 6426006 4488062 0806620 4442468 6064688 2442028 2644884 2622424 2646866 8688022 4406662 0.00], p=1.00
male_22_to_24	1.63	

	[0.00, 1932807 9791930 0009328 2600040 8262248 6842680 6024228 2284640 8246408 6280244 8486048 4004464 2624664 6886224 6208088 6222480 02248.00], p=1.00
male_25 _to_29	2.49

	[0.00, 3427464 4367330 8009706 0846664 4862084 0864244 6468282 0002606 0602682 4408804 2688800 6604044 6068282 0822846 8808888 2640662 6684044 0866686 8428642 8220260 4640266 8802000 0484002 0646048 8888442 4664688 8288288 4266402 0220484. 00], p=1.00
male_30 _to_34	1.55

	[0.00, 1375125 1801273 4007112 8444200 0668428 0862682 2486424 4088266 4428022 2408640 6684486 4008488 6846806 4484.00], p=1.00
male_40 _to_44	2.48

	[0.00, 1251257 2019048 4995282 8024686 6000868 8466424 6844248 4044800 6664042 2622444 6686262 0622480 6642648 2402026 8262020 2882608 4426202 2886462 2442464 6200002 0088024 0662884 2086826 8204044 6884826 6484808 6088602 0666808 2424024 8660440 8062260 4000020 4.00], p=1.00
male_45_to_49	2.58

	[0.00, 8577633 2732124 2964662 6820468 4626028 4246064 6628844 0682660 8024820 0008044 8066202 0426002 6000684 2682848 0466820 4842006 0622020 6422002 2844860 8828246 8840888 4602060 8800046 6004064 8488440 4422646 8608628 6288882 88.00], p=1.00
male_50 _to_54	1.27

	[0.00, 5203081 4281383 9977608 2262646 8688600 8206482 4626220 4820020 2444682 8008620 6820206 6686426 6040684 8420888 8804884 8404260 4880624. 00], p=1.00
male_55 _to_59	1.88
	[0.00, 1407266 4875453 1005194 8002862 0086622 2606426 0002220 6042644 0448688 8284442 6402466 4260840 0628488 2822240 4026688 4488804 0062442 2880648 6642262 0062488 088004.0 0], p=1.00
male_75 _to_79	6.00

	[0.00, Inf], p=1.00
female_55_to_59	1.36
	[0.00, 3185366766488679964064844200068602848668460266066844828824288620086626066880228406424404628280488088282046220228824244466440288266468826268804048 46.00], p=1.00
female_62_to_64	2.07

		[0.00, 1567839 8411378 0008714 0404866 0022820 0684442 8844004 2666208 8004264 2484604 8240642 8468086 4644880 4204600 0840826 8662622 2068068 6600206 0204684 4624840 4442060 8682066 2406860 8004846 0086068 6.00], p=1.00
white_pop	0.95	
		[0.00, 6374820 4685313 5953364 824820.00], p=1.00
black_pop	0.95	
		[0.00, 1293949 9134444 8011634 4080200.00], p=1.00

asian_pop		0.95 [0.00, 3012182101973440080208860606.00], p=1.00
hispanic_pop		0.93 [0.00, 8075327457750959318406246666.00], p=1.00
N	60	57
AIC	75.13	38.00
BIC	81.42	76.82
Pseudo R2	0.28	1.00

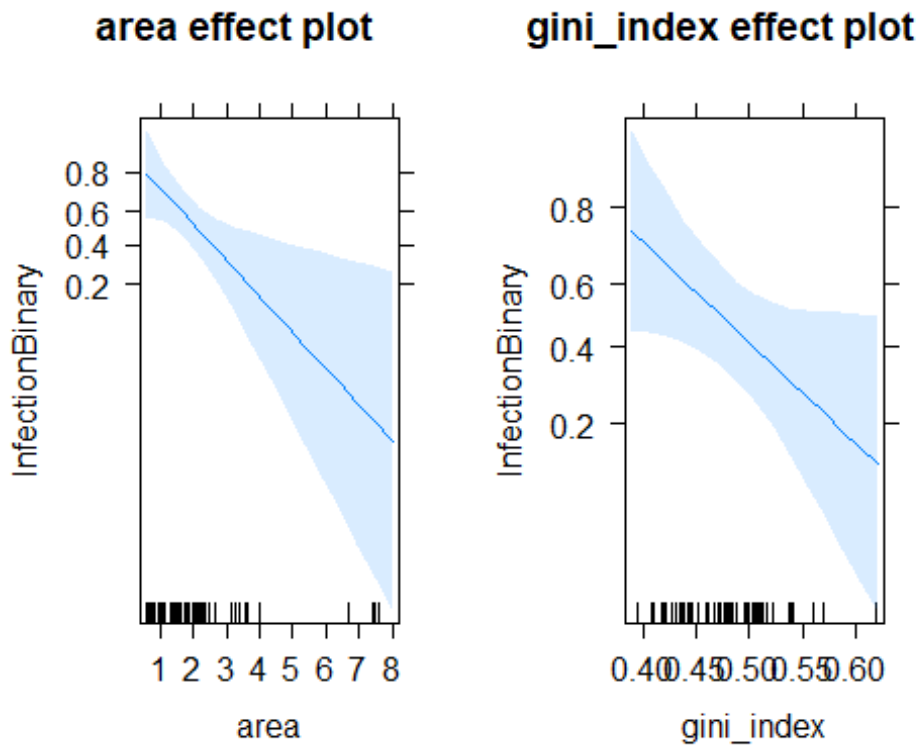
*** p < 0.001; ** p < 0.01; * p < 0.05.

```
summ(model1,exp = T,digits = getOption("jtools-digits", default = 3),confint = getOption("summ-confint", TRUE), conf.method = getOption("summ-conf.method", c("Wald")))
```

Observations
60
Dependent variable
InfectionBinary
Type
Generalized linear model
Family
binomial
Link
logit
$\chi^2(2)$
14.044
Pseudo-R^2 (Cragg-Uhler)

0.278
Pseudo-R^2 (McFadden)
0.169
AIC
75.133
BIC
81.416
exp(Est.)
2.5%
97.5%
z val.
p
(Intercept)
3218.911
3.135
3304794.257
2.283
0.022
area
0.412
0.209
0.812
-2.563
0.010
gini_index
0.000
0.000
0.834
-1.988
0.047
Standard errors: MLE

```
par(mfrow=c(1,1))
par(mar=c(1, 1, 1, 1))
plot(allEffects(model1))
```



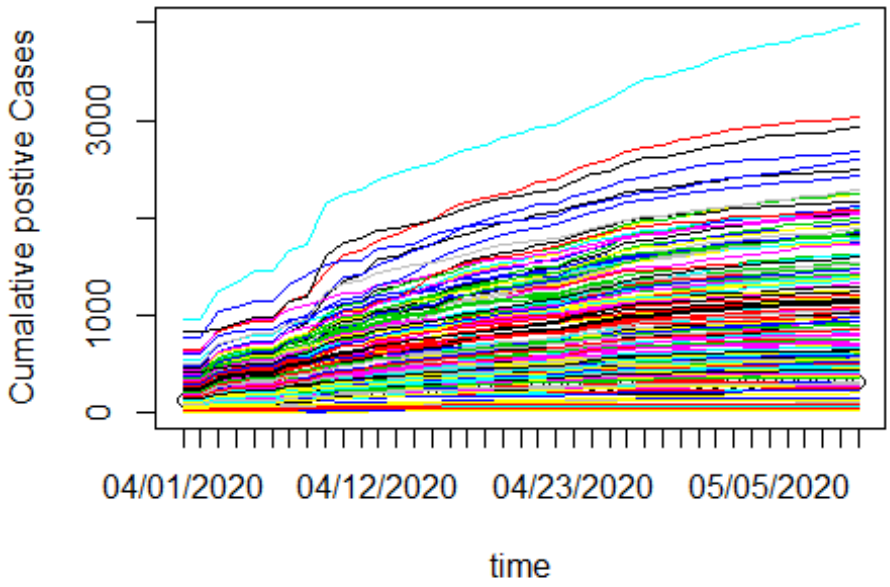
Identify Clusters of Longitudinal Trajectories

Here aim is to dentify zipcodes with significantly different rates of change of postive cases using traj Package.

Data preprocessing

```
ff<-read.table("time-test-by-zcta_borough.csv",header= T, stringsAsFactors =
F, sep=",",fill=T, quote = "")
ff<-na.omit(ff)
#ffLast<-ff[ff$Date=="05/10/2020",]
#head(ffLast)
ff<-data.frame(ff)

ff1<-ff[,c(2,3,5)]


ff1<-data.frame(ff1)

dat<-ff1 %>%
  group_by_at(vars(-Positive)) %>%  # group by everything other than the valu
e column.
  mutate(row_id=1:n()) %>% ungroup() %>%  # build group index
  spread(key=Date, value=Positive) %>%    # spread
  dplyr::select(-row_id)
```

```r
y.data<-dat[,2:40]

#y.data<-scale(y.data)

#This will plot the trajectories
plot(unlist(y.data[1,]), ylim=c(0,max(y.data, na.rm=T)), ylab="Cumalative pos
tive Cases", xaxt='n', xlab="time" )

for(i in 1:dim(y.data)[1]){ lines(unlist(y.data[i,]), col=i+1)}
axis(1, at=1:39, labels=colnames(y.data))
```




```r
dat<-na.omit(dat)
colnames(dat)<-c("ZipCodes", paste0("X", 1:39, "X"))
dat$ZipCodes<-factor(dat$ZipCodes)
#head(dat)


dattime = matrix(rep(1:39,dim(dat)[1] ), ncol=39, byrow=dim(dat)[1])
dattime<-data.frame(dattime)
dattime$ZipCodes<-dat$ZipCodes
dattime<-dattime[,c("ZipCodes",colnames(dattime)[1:39])]
dat = as.data.frame(dat)
```

Trajectory analysis

- The first step in the analysis consists of the computing 24 measures of each trajectory.

- In the second step of the analysis, a factor analysis is performed to select a subset of measures that describes the main features of the trajectories. The function step2factors is used to perform the factor analysis.
- The step2factors has identified measures m2 (Mean-over-time), m4 (Coefficient of variation (CV)) and m16 (Ratio of the maximum absolute difference to the mean-over-time) as the main factors of this set of trajectories. Measures 5,6, 14,15 and 13 were not considered because they were too correlated with other measures (measures with a correlation higher than 0.95 are omitted from the factor analysis).

```
#https://cran.r-project.org/web/packages/traj/vignettes/trajVignette.pdf
s1 = step1measures(dat, dattime, ID = TRUE)

## [1] "Correlation of m1 and m5 : 1"
## [1] "Correlation of m1 and m6 : 1"
## [1] "Correlation of m1 and m14 : 1"
## [1] "Correlation of m3 and m9 : 1"
## [1] "Correlation of m5 and m6 : 1"
## [1] "Correlation of m5 and m14 : 1"
## [1] "Correlation of m6 and m14 : 1"
## [1] "Correlation of m11 and m15 : 1"
## [1] "Correlation of m12 and m13 : 1"

s2 = step2factors(s1)

## [1] "m5 is removed because it is perfectly correlated with m1"
## [2] "m6 is removed because it is perfectly correlated with m1"
## [3] "m14 is removed because it is perfectly correlated with m1"
## [4] "m9 is removed because it is perfectly correlated with m3"
## [5] "m6 is removed because it is perfectly correlated with m5"
## [6] "m14 is removed because it is perfectly correlated with m5"
## [7] "m14 is removed because it is perfectly correlated with m6"
## [8] "m15 is removed because it is perfectly correlated with m11"
## [9] "m13 is removed because it is perfectly correlated with m12"
## [1] "Computing reduced correlation e-values..."

head(s2$factors)
```

ID	m2	m4	m16
10001	247	25.8	0.0931
10002	665	31.7	0.101
10003	324	24	0.102
10004	24.1	13.9	0.125
10005	43.2	22	0.0926
10006	14.5	40.1	0.344

```
#Once this step is done, the third step of the procedure consists in clusteri
ng the trajectories based on the measures identified in the factor analysis. H
ere number of clusters 3

s3 = step3clusters(s2, nclusters = 3)
head(s3$clusters)
```

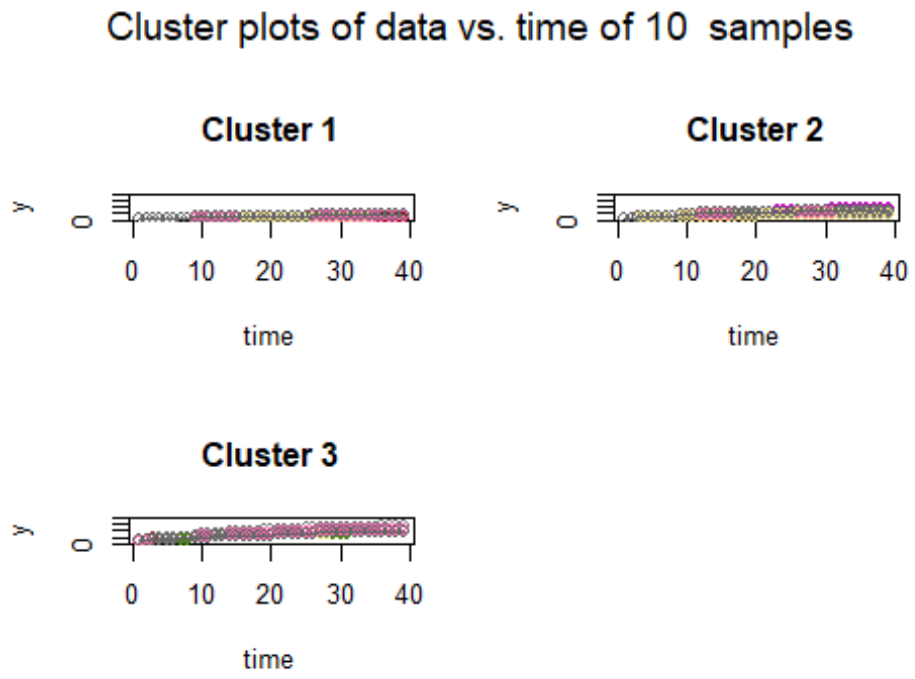
ID	cluster
10001	1
10002	2
10003	1
10004	1
10005	1
10006	1

```
#Cluster Memberships
s3$clust.distr

##
##  1  2  3
## 91 55 31
```

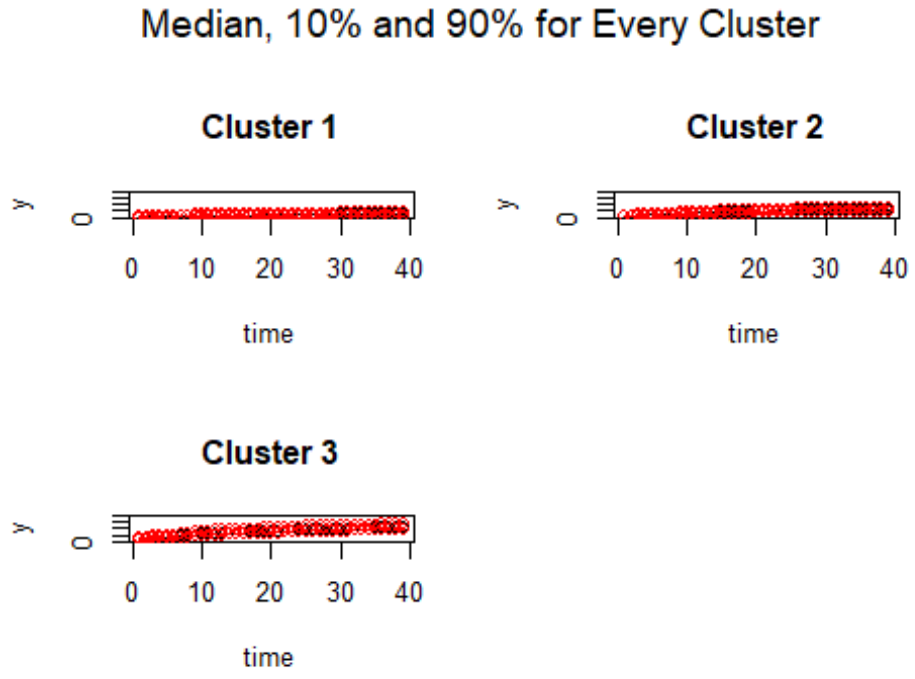
Plotting the traj object

- Here we are plotting 10 random trajectories from each cluster and plots them using randomly selected colours.
- Also we can plot,the mean trajectory of every cluster, and the median trajectory of every cluster with 10th and 90th percentiles

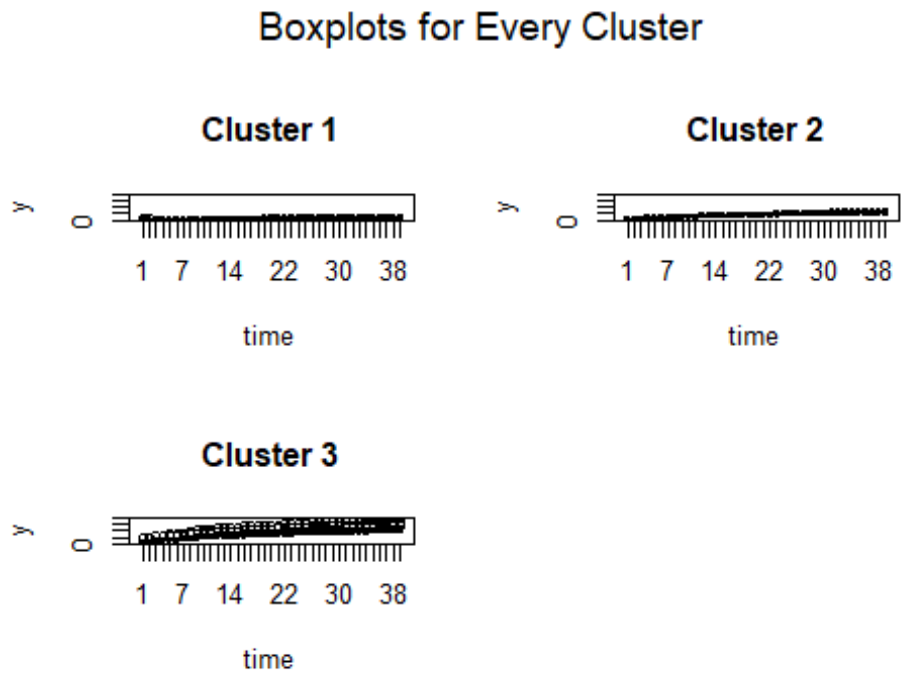
```
plot(s3)
```




```
plotMedTraj(s3)
```




```
plotBoxplotTraj(s3)
```




```
#https://towardsdatascience.com/covid-19-infection-in-italy-mathematical-models-and-predictions-7784b4d7dd8d
```

Now we analze the basic charectristics of each groups

Comparison of groups using p values

- Here most of the factors are significant in these three groups (most of the cases p <0.001)
- Significant features: Total population, households, gender, gender by older age groups, Race, inclcome level, poverty status, vacant housing units, higher rent, commute, do not having car, commuter by public, less than college, male by education level, employment type, area
- Furture investigation is needed identitify why these covarites related to rate of positive cases.

```
feature<-read.table("nycjackpotbyzip.csv",header= T, stringsAsFactors = F, sep=",",fill=T, quote = "")
feature$ZipCode<-feature$geo_id
density<-read.table("popbyzipdensity.csv",header= T, stringsAsFactors = F, sep=",",fill=T, quote = "")
DATA11<-merge(feature,density,by="ZipCode",all.x=T)

GR<-s3$clusters
GR$ZipCode<-GR$ID
```

```r
GROUPFULL<-merge(GR,DATA11, by="ZipCode",all.x=T )

GROUPFULL<-GROUPFULL[,!(colnames(GROUPFULL) %in% c("pop_5_years_over","pop_15
_and_over","pop_never_married","pop_now_married","pop_separated","pop_widowed
","pop_divorced","speak_only_english_at_home","speak_spanish_at_home","speak_
spanish_at_home_low_english","speak_spanish_at_home","speak_spanish_at_home_l
ow_english","aggregate_travel_time_to_work","renter_occupied_housing_units_pa
ying_cash_median_gross_rent","owner_occupied_housing_units_lower_value_quarti
le","owner_occupied_housing_units_median_value","owner_occupied_housing_units
_upper_value_quartile","median_rent","percent_income_spent_on_rent"))]


#missing values frequencies
#freq.na(GROUPFULL)

GROUPFULL<-data.frame(GROUPFULL)


GROUPFULL$cluster<-factor(GROUPFULL$cluster)

pred<-GROUPFULL[, -c(1,2,4,5)]
pred<-pred[,-c(1)]
myVars <- colnames(pred)


#Finding significant factors using chi-square tests for all 230 predictors
tab <- CreateTableOne(vars = myVars, strata = "cluster" , data = GROUPFULL,fa
ctorVars="cluster")
tab

##
Stratified by cluster
##
1
##   n
91
##   total_pop (mean (SD))
29220.54 (15867.86)
##   households (mean (SD))
12502.38 (7971.21)
##   male_pop (mean (SD))
13976.04 (7465.93)
##   female_pop (mean (SD))
15244.49 (8486.89)
##   median_age (mean (SD))
38.49 (4.78)
##   male_under_5 (mean (SD))
863.60 (542.41)
##   male_5_to_9 (mean (SD))
```

```
649.90 (392.90)
##   male_10_to_14 (mean (SD))
646.29 (394.51)
##   male_15_to_17 (mean (SD))
369.64 (240.71)
##   male_18_to_19 (mean (SD))
255.07 (312.59)
##   male_20 (mean (SD))
139.99 (119.56)
##   male_21 (mean (SD))
141.26 (115.96)
##   male_22_to_24 (mean (SD))
544.13 (402.92)
##   male_25_to_29 (mean (SD))
1453.19 (1032.36)
##   male_30_to_34 (mean (SD))
1435.70 (1058.11)
##   male_35_to_39 (mean (SD))
1186.74 (771.96)
##   male_40_to_44 (mean (SD))
982.92 (586.45)
##   male_45_to_49 (mean (SD))
976.54 (563.65)
##   male_50_to_54 (mean (SD))
914.23 (474.67)
##   male_55_to_59 (mean (SD))
850.46 (478.57)
##   male_65_to_66 (mean (SD))
265.07 (181.43)
##   male_67_to_69 (mean (SD))
346.75 (218.50)
##   male_70_to_74 (mean (SD))
442.77 (305.74)
##   male_75_to_79 (mean (SD))
314.67 (199.37)
##   male_80_to_84 (mean (SD))
218.11 (175.82)
##   male_85_and_over (mean (SD))
217.48 (188.50)
##   female_under_5 (mean (SD))
810.70 (498.05)
##   female_5_to_9 (mean (SD))
656.77 (424.84)
##   female_10_to_14 (mean (SD))
615.95 (391.69)
##   female_15_to_17 (mean (SD))
362.19 (242.99)
##   female_18_to_19 (mean (SD))
292.07 (463.61)
##   female_20 (mean (SD))
```

```
158.09 (143.47)
##   female_21 (mean (SD))
165.10 (149.21)
##   female_22_to_24 (mean (SD))
640.40 (455.49)
##   female_25_to_29 (mean (SD))
1634.91 (1264.97)
##   female_30_to_34 (mean (SD))
1544.71 (1137.77)
##   female_35_to_39 (mean (SD))
1180.91 (748.26)
##   female_40_to_44 (mean (SD))
998.04 (608.40)
##   female_45_to_49 (mean (SD))
954.05 (545.01)
##   female_50_to_54 (mean (SD))
938.70 (497.19)
##   female_55_to_59 (mean (SD))
905.21 (516.94)
##   female_60_to_61 (mean (SD))
381.73 (229.09)
##   female_62_to_64 (mean (SD))
476.27 (281.69)
##   female_65_to_66 (mean (SD))
336.22 (222.21)
##   female_67_to_69 (mean (SD))
431.29 (298.84)
##   female_70_to_74 (mean (SD))
595.25 (391.52)
##   female_75_to_79 (mean (SD))
432.15 (306.34)
##   female_80_to_84 (mean (SD))
322.78 (225.04)
##   female_85_and_over (mean (SD))
411.00 (286.99)
##   white_pop (mean (SD))
14171.81 (11970.58)
##   population_1_year_and_over (mean (SD))
28877.35 (15671.35)
##   population_3_years_over (mean (SD))
28198.49 (15310.70)
##   pop_16_over (mean (SD))
24733.66 (13792.99)
##   pop_25_years_over (mean (SD))
21909.41 (12410.65)
##   pop_25_64 (mean (SD))
17575.87 (10121.69)
##   not_us_citizen_pop (mean (SD))
3856.75 (2322.61)
##   black_pop (mean (SD))
```

```
3840.99 (5943.90)
##   asian_pop (mean (SD))
4472.10 (4068.27)
##   hispanic_pop (mean (SD))
5750.47 (5204.10)
##   amerindian_pop (mean (SD))
41.40 (56.52)
##   other_race_pop (mean (SD))
221.67 (394.03)
##   two_or_more_races_pop (mean (SD))
710.51 (586.42)
##   hispanic_any_race (mean (SD))
5750.47 (5204.10)
##   not_hispanic_pop (mean (SD))
23470.07 (13977.58)
##   asian_male_45_54 (mean (SD))
270.76 (303.88)
##   asian_male_55_64 (mean (SD))
242.74 (295.48)
##   black_male_45_54 (mean (SD))
289.29 (413.79)
##   black_male_55_64 (mean (SD))
247.91 (388.21)
##   hispanic_male_45_54 (mean (SD))
369.92 (334.36)
##   hispanic_male_55_64 (mean (SD))
263.87 (222.97)
##   white_male_45_54 (mean (SD))
938.71 (776.42)
##   white_male_55_64 (mean (SD))
846.95 (740.26)
##   median_income (mean (SD))
91822.67 (40059.64)
##   income_per_capita (mean (SD))
60628.64 (36923.83)
##   income_less_10000 (mean (SD))
830.14 (658.17)
##   income_10000_14999 (mean (SD))
550.42 (474.07)
##   income_15000_19999 (mean (SD))
439.82 (302.98)
##   income_20000_24999 (mean (SD))
434.00 (297.51)
##   income_25000_29999 (mean (SD))
379.20 (238.16)
##   income_30000_34999 (mean (SD))
390.43 (240.80)
##   income_35000_39999 (mean (SD))
339.92 (205.60)
##   income_40000_44999 (mean (SD))
```

```
364.82 (214.83)
##   income_45000_49999 (mean (SD))
298.55 (196.11)
##   income_50000_59999 (mean (SD))
671.54 (378.41)
##   income_60000_74999 (mean (SD))
942.19 (566.30)
##   income_75000_99999 (mean (SD))
1358.96 (823.64)
##   income_100000_124999 (mean (SD))
1136.48 (744.60)
##   income_125000_149999 (mean (SD))
882.19 (601.33)
##   income_150000_199999 (mean (SD))
1178.82 (918.91)
##   income_200000_or_more (mean (SD))
2304.90 (2716.46)
##   pop_determined_poverty_status (mean (SD))
28669.96 (15587.68)
##   poverty (mean (SD))
3814.84 (3092.01)
##   gini_index (mean (SD))
0.49 (0.06)
##   housing_units (mean (SD))
14220.59 (9327.50)
##   occupied_housing_units (mean (SD))
12502.38 (7971.21)
##   housing_units_renter_occupied (mean (SD))
8051.55 (6200.04)
##   vacant_housing_units (mean (SD))
1718.21 (1710.17)
##   vacant_housing_units_for_rent (mean (SD))
374.16 (373.15)
##   vacant_housing_units_for_sale (mean (SD))
90.84 (94.37)
##   dwellings_1_units_detached (mean (SD))
1235.86 (1498.69)
##   dwellings_1_units_attached (mean (SD))
682.10 (882.67)
##   dwellings_2_units (mean (SD))
1420.95 (1554.41)
##   dwellings_3_to_4_units (mean (SD))
1086.53 (1474.69)
##   dwellings_5_to_9_units (mean (SD))
1181.97 (1792.54)
##   dwellings_10_to_19_units (mean (SD))
1147.56 (1376.80)
##   dwellings_20_to_49_units (mean (SD))
2136.11 (2384.15)
##   dwellings_50_or_more_units (mean (SD))
```

```
5306.87 (6426.65)
##   mobile_homes (mean (SD))
16.41 (26.61)
##   housing_built_2005_or_later (mean (SD))
145.12 (290.17)
##   housing_built_2000_to_2004 (mean (SD))
264.53 (418.63)
##   housing_built_1939_or_earlier (mean (SD))
1326.04 (1192.82)
##   median_year_structure_built (mean (SD))
1421.88 (880.06)
##   married_households (mean (SD))
4496.01 (2755.38)
##   nonfamily_households (mean (SD))
6209.88 (5441.46)
##   family_households (mean (SD))
6292.51 (3363.22)
##   households_public_asst_or_food_stamps (mean (SD))
1430.05 (1250.93)
##   male_male_households (mean (SD))
80.85 (123.53)
##   female_female_households (mean (SD))
26.44 (37.27)
##   children (mean (SD))
4975.03 (2902.49)
##   children_in_single_female_hh (mean (SD))
1230.44 (1162.22)
##   rent_burden_not_computed (mean (SD))
377.91 (272.36)
##   rent_over_50_percent (mean (SD))
1754.20 (1265.06)
##   rent_40_to_50_percent (mean (SD))
606.02 (474.28)
##   rent_35_to_40_percent (mean (SD))
458.25 (353.71)
##   rent_30_to_35_percent (mean (SD))
684.09 (574.53)
##   rent_25_to_30_percent (mean (SD))
845.24 (697.91)
##   rent_20_to_25_percent (mean (SD))
968.15 (813.09)
##   rent_15_to_20_percent (mean (SD))
993.88 (838.29)
##   rent_10_to_15_percent (mean (SD))
760.54 (670.91)
##   rent_under_10_percent (mean (SD))
603.26 (679.70)
##   owner_occupied_housing_units (mean (SD))
4450.84 (3249.49)
##   million_dollar_housing_units (mean (SD))
```

```
489.22 (563.04)
##   mortgaged_housing_units (mean (SD))
2483.68 (1797.34)
##   different_house_year_ago_different_city (mean (SD))
995.82 (1170.21)
##   different_house_year_ago_same_city (mean (SD))
2447.77 (1780.47)
##   families_with_young_children (mean (SD))
1874.32 (1155.15)
##   two_parent_families_with_young_children (mean (SD))
1388.23 (990.28)
##   two_parents_in_labor_force_families_with_young_children (mean (SD))
875.36 (694.91)
##   two_parents_father_in_labor_force_families_with_young_children (mean (SD
))   457.36 (341.09)
##   two_parents_mother_in_labor_force_families_with_young_children (mean (SD
))    38.27 (43.71)
##   two_parents_not_in_labor_force_families_with_young_children (mean (SD))
17.23 (32.85)
##   one_parent_families_with_young_children (mean (SD))
486.09 (440.00)
##   father_one_parent_families_with_young_children (mean (SD))
87.21 (87.99)
##   father_in_labor_force_one_parent_families_with_young_children (mean (SD)
)     75.58 (75.63)
##   commute_less_10_mins (mean (SD))
657.60 (538.20)
##   commute_10_14_mins (mean (SD))
843.63 (635.38)
##   commute_15_19_mins (mean (SD))
1211.59 (976.68)
##   commute_20_24_mins (mean (SD))
1780.70 (1533.41)
##   commute_25_29_mins (mean (SD))
887.71 (845.80)
##   commute_30_34_mins (mean (SD))
2682.54 (2022.39)
##   commute_35_44_mins (mean (SD))
1732.81 (1505.63)
##   commute_60_more_mins (mean (SD))
2694.76 (1790.09)
##   commute_45_59_mins (mean (SD))
2103.91 (1894.77)
##   commuters_16_over (mean (SD))
14595.26 (8693.23)
##   walked_to_work (mean (SD))
2104.87 (2666.76)
##   worked_at_home (mean (SD))
855.48 (862.92)
##   no_car (mean (SD))
```

```
7905.29 (7713.86)
##   no_cars (mean (SD))
7322.85 (6862.09)
##   one_car (mean (SD))
3654.58 (2226.68)
##   two_cars (mean (SD))
1174.84 (1146.16)
##   three_cars (mean (SD))
265.23 (306.27)
##   four_more_cars (mean (SD))
84.89 (111.82)
##   commuters_by_public_transportation (mean (SD))
8412.29 (6520.78)
##   commuters_by_bus (mean (SD))
1118.56 (862.05)
##   commuters_by_car_truck_van (mean (SD))
3412.20 (2851.81)
##   commuters_by_carpool (mean (SD))
541.37 (461.11)
##   commuters_by_subway_or_elevated (mean (SD))
6937.10 (6307.11)
##   commuters_drove_alone (mean (SD))
2870.82 (2435.26)
##   group_quarters (mean (SD))
723.23 (1332.87)
##   associates_degree (mean (SD))
1189.53 (702.35)
##   bachelors_degree (mean (SD))
6522.99 (5000.33)
##   high_school_diploma (mean (SD))
3177.88 (2010.85)
##   less_one_year_college (mean (SD))
627.31 (348.06)
##   masters_degree (mean (SD))
3620.63 (3333.91)
##   one_year_more_college (mean (SD))
1865.31 (990.10)
##   less_than_high_school_graduate (mean (SD))
2497.80 (1974.63)
##   high_school_including_ged (mean (SD))
3752.81 (2344.14)
##   bachelors_degree_2 (mean (SD))
6522.99 (5000.33)
##   bachelors_degree_or_higher_25_64 (mean (SD))
10303.22 (8669.98)
##   graduate_professional_degree (mean (SD))
5453.66 (5483.39)
##   some_college_and_associates_degree (mean (SD))
3682.14 (1930.91)
##   male_45_64_associates_degree (mean (SD))
```

```
210.88 (147.04)
##   male_45_64_bachelors_degree (mean (SD))
833.76 (619.34)
##   male_45_64_graduate_degree (mean (SD))
779.99 (811.62)
##   male_45_64_less_than_9_grade (mean (SD))
204.24 (184.31)
##   male_45_64_grade_9_12 (mean (SD))
245.13 (201.07)
##   male_45_64_high_school (mean (SD))
761.14 (499.12)
##   male_45_64_some_college (mean (SD))
467.63 (266.73)
##   male_45_to_64 (mean (SD))
3502.77 (1853.65)
##   employed_pop (mean (SD))
15777.46 (9617.55)
##   unemployed_pop (mean (SD))
918.36 (610.14)
##   pop_in_labor_force (mean (SD))
16703.95 (10069.91)
##   not_in_labor_force (mean (SD))
8029.71 (4182.63)
##   workers_16_and_over (mean (SD))
15450.75 (9451.35)
##   armed_forces (mean (SD))
8.12 (20.20)
##   civilian_labor_force (mean (SD))
16695.82 (10066.71)
##   employed_agriculture_forestry_fishing_hunting_mining (mean (SD))
13.49 (20.00)
##   employed_arts_entertainment_recreation_accommodation_food (mean (SD))
1521.54 (1091.81)
##   employed_construction (mean (SD))
598.30 (495.56)
##   employed_education_health_social (mean (SD))
3553.07 (2057.73)
##   employed_finance_insurance_real_estate (mean (SD))
2166.44 (2018.06)
##   employed_information (mean (SD))
923.32 (968.71)
##   employed_manufacturing (mean (SD))
499.87 (330.95)
##   employed_other_services_not_public_admin (mean (SD))
729.20 (489.86)
##   employed_public_administration (mean (SD))
554.54 (321.68)
##   employed_retail_trade (mean (SD))
1257.62 (717.95)
##   employed_science_management_admin_waste (mean (SD))
```

```
2947.85 (2586.75)
##   employed_transportation_warehousing_utilities (mean (SD))
663.60 (468.70)
##   employed_wholesale_trade (mean (SD))
348.64 (264.07)
##   occupation_management_arts (mean (SD))
8823.70 (7359.64)
##   occupation_natural_resources_construction_maintenance (mean (SD))
646.14 (536.96)
##   occupation_production_transportation_material (mean (SD))
920.16 (681.75)
##   occupation_sales_office (mean (SD))
3130.89 (1709.66)
##   occupation_services (mean (SD))
2256.56 (1498.12)
##   management_business_sci_arts_employed (mean (SD))
8823.70 (7359.64)
##   sales_office_employed (mean (SD))
3130.89 (1709.66)
##   in_grades_1_to_4 (mean (SD))
1052.32 (641.62)
##   in_grades_5_to_8 (mean (SD))
1002.70 (619.41)
##   in_grades_9_to_12 (mean (SD))
1020.55 (649.69)
##   in_school (mean (SD))
5978.88 (3345.76)
##   in_undergrad_college (mean (SD))
1489.19 (1298.41)
##   Row (mean (SD))
80.26 (56.37)
##   population (mean (SD))
28069.26 (15447.01)
##   area (mean (SD))
1.10 (0.80)
##   peoplepersqmile (mean (SD))
41501.24 (33128.73)
##
Stratified by cluster
##
2
##   n
55
##   total_pop (mean (SD))
57564.22 (18467.86)
##   households (mean (SD))
20537.58 (7985.91)
##   male_pop (mean (SD))
27472.24 (8807.11)
##   female_pop (mean (SD))
```

```
30091.98 (9843.75)
##   median_age (mean (SD))
37.52 (4.21)
##   male_under_5 (mean (SD))
1900.33 (865.95)
##   male_5_to_9 (mean (SD))
1734.13 (730.85)
##   male_10_to_14 (mean (SD))
1674.40 (666.97)
##   male_15_to_17 (mean (SD))
973.27 (355.20)
##   male_18_to_19 (mean (SD))
666.22 (263.63)
##   male_20 (mean (SD))
386.36 (169.39)
##   male_21 (mean (SD))
384.33 (136.95)
##   male_22_to_24 (mean (SD))
1277.67 (567.84)
##   male_25_to_29 (mean (SD))
2548.65 (1228.14)
##   male_30_to_34 (mean (SD))
2215.75 (943.97)
##   male_35_to_39 (mean (SD))
1839.87 (733.86)
##   male_40_to_44 (mean (SD))
1736.16 (637.53)
##   male_45_to_49 (mean (SD))
1724.55 (587.74)
##   male_50_to_54 (mean (SD))
1766.02 (575.06)
##   male_55_to_59 (mean (SD))
1691.42 (553.75)
##   male_65_to_66 (mean (SD))
531.16 (246.55)
##   male_67_to_69 (mean (SD))
666.64 (306.24)
##   male_70_to_74 (mean (SD))
844.35 (344.14)
##   male_75_to_79 (mean (SD))
628.91 (253.76)
##   male_80_to_84 (mean (SD))
408.56 (242.75)
##   male_85_and_over (mean (SD))
352.69 (225.47)
##   female_under_5 (mean (SD))
1788.75 (771.37)
##   female_5_to_9 (mean (SD))
1598.02 (640.34)
##   female_10_to_14 (mean (SD))
```

```
1536.55 (601.87)
##   female_15_to_17 (mean (SD))
964.07 (349.49)
##   female_18_to_19 (mean (SD))
645.44 (358.07)
##   female_20 (mean (SD))
374.18 (203.04)
##   female_21 (mean (SD))
370.05 (181.26)
##   female_22_to_24 (mean (SD))
1299.45 (609.51)
##   female_25_to_29 (mean (SD))
2610.36 (1186.72)
##   female_30_to_34 (mean (SD))
2205.60 (918.11)
##   female_35_to_39 (mean (SD))
2050.53 (748.42)
##   female_40_to_44 (mean (SD))
1869.65 (632.44)
##   female_45_to_49 (mean (SD))
1965.69 (644.70)
##   female_50_to_54 (mean (SD))
2039.36 (663.49)
##   female_55_to_59 (mean (SD))
1954.24 (622.27)
##   female_60_to_61 (mean (SD))
788.84 (281.21)
##   female_62_to_64 (mean (SD))
1033.45 (401.76)
##   female_65_to_66 (mean (SD))
636.15 (264.60)
##   female_67_to_69 (mean (SD))
855.40 (375.23)
##   female_70_to_74 (mean (SD))
1128.96 (492.25)
##   female_75_to_79 (mean (SD))
911.65 (370.98)
##   female_80_to_84 (mean (SD))
663.05 (323.17)
##   female_85_and_over (mean (SD))
802.53 (453.72)
##   white_pop (mean (SD))
15556.85 (14830.75)
##   population_1_year_and_over (mean (SD))
56860.89 (18215.35)
##   population_3_years_over (mean (SD))
55368.60 (17671.53)
##   pop_16_over (mean (SD))
46705.60 (15042.19)
##   pop_25_years_over (mean (SD))
```

```
39991.00 (13288.60)
##   pop_25_64 (mean (SD))
31560.95 (10411.64)
##   not_us_citizen_pop (mean (SD))
9577.87 (6071.78)
##   black_pop (mean (SD))
13442.27 (14226.71)
##   asian_pop (mean (SD))
9013.40 (11863.60)
##   hispanic_pop (mean (SD))
17568.38 (12216.65)
##   amerindian_pop (mean (SD))
106.76 (104.41)
##   other_race_pop (mean (SD))
705.47 (1746.19)
##   two_or_more_races_pop (mean (SD))
1148.16 (715.76)
##   hispanic_any_race (mean (SD))
17568.38 (12216.65)
##   not_hispanic_pop (mean (SD))
39995.84 (18343.64)
##   asian_male_45_54 (mean (SD))
610.15 (897.76)
##   asian_male_55_64 (mean (SD))
605.82 (773.57)
##   black_male_45_54 (mean (SD))
889.02 (869.30)
##   black_male_55_64 (mean (SD))
815.22 (774.45)
##   hispanic_male_45_54 (mean (SD))
999.11 (706.16)
##   hispanic_male_55_64 (mean (SD))
791.42 (555.95)
##   white_male_45_54 (mean (SD))
947.29 (967.59)
##   white_male_55_64 (mean (SD))
988.76 (1039.62)
##   median_income (mean (SD))
56874.42 (19551.45)
##   income_per_capita (mean (SD))
28636.95 (9383.79)
##   income_less_10000 (mean (SD))
2083.73 (1519.80)
##   income_10000_14999 (mean (SD))
1467.16 (966.04)
##   income_15000_19999 (mean (SD))
1131.85 (627.17)
##   income_20000_24999 (mean (SD))
1051.11 (552.18)
##   income_25000_29999 (mean (SD))
```

```
946.00 (462.43)
##   income_30000_34999 (mean (SD))
894.07 (389.05)
##   income_35000_39999 (mean (SD))
849.42 (383.86)
##   income_40000_44999 (mean (SD))
811.22 (323.44)
##   income_45000_49999 (mean (SD))
679.25 (294.79)
##   income_50000_59999 (mean (SD))
1333.04 (528.53)
##   income_60000_74999 (mean (SD))
1737.51 (699.31)
##   income_75000_99999 (mean (SD))
2258.13 (836.19)
##   income_100000_124999 (mean (SD))
1606.91 (681.97)
##   income_125000_149999 (mean (SD))
1098.02 (539.13)
##   income_150000_199999 (mean (SD))
1268.09 (716.62)
##   income_200000_or_more (mean (SD))
1322.07 (1339.88)
##   pop_determined_poverty_status (mean (SD))
56502.53 (18446.89)
##   poverty (mean (SD))
11753.20 (7400.18)
##   gini_index (mean (SD))
0.48 (0.06)
##   housing_units (mean (SD))
22245.58 (8766.66)
##   occupied_housing_units (mean (SD))
20537.58 (7985.91)
##   housing_units_renter_occupied (mean (SD))
13761.44 (7966.96)
##   vacant_housing_units (mean (SD))
1708.00 (1070.05)
##   vacant_housing_units_for_rent (mean (SD))
431.11 (384.05)
##   vacant_housing_units_for_sale (mean (SD))
125.58 (91.82)
##   dwellings_1_units_detached (mean (SD))
2424.69 (2362.03)
##   dwellings_1_units_attached (mean (SD))
1724.27 (1599.00)
##   dwellings_2_units (mean (SD))
3047.76 (2120.88)
##   dwellings_3_to_4_units (mean (SD))
2034.80 (2246.48)
##   dwellings_5_to_9_units (mean (SD))
```

```
1192.80 (1436.46)
##   dwellings_10_to_19_units (mean (SD))
1275.05 (1270.46)
##   dwellings_20_to_49_units (mean (SD))
3649.58 (3070.80)
##   dwellings_50_or_more_units (mean (SD))
6856.49 (6340.38)
##   mobile_homes (mean (SD))
29.05 (24.28)
##   housing_built_2005_or_later (mean (SD))
98.22 (118.86)
##   housing_built_2000_to_2004 (mean (SD))
292.02 (267.11)
##   housing_built_1939_or_earlier (mean (SD))
2315.67 (1395.90)
##   median_year_structure_built (mean (SD))
1351.07 (912.04)
##   married_households (mean (SD))
7678.40 (3655.06)
##   nonfamily_households (mean (SD))
7498.64 (4525.92)
##   family_households (mean (SD))
13038.95 (4247.22)
##   households_public_asst_or_food_stamps (mean (SD))
5036.02 (3354.52)
##   male_male_households (mean (SD))
47.60 (68.64)
##   female_female_households (mean (SD))
30.65 (33.22)
##   children (mean (SD))
12169.51 (4666.24)
##   children_in_single_female_hh (mean (SD))
4157.96 (2608.13)
##   rent_burden_not_computed (mean (SD))
764.80 (480.90)
##   rent_over_50_percent (mean (SD))
4031.67 (2322.85)
##   rent_40_to_50_percent (mean (SD))
1164.73 (704.91)
##   rent_35_to_40_percent (mean (SD))
839.49 (518.56)
##   rent_30_to_35_percent (mean (SD))
1281.93 (859.93)
##   rent_25_to_30_percent (mean (SD))
1347.36 (839.35)
##   rent_20_to_25_percent (mean (SD))
1385.85 (863.78)
##   rent_15_to_20_percent (mean (SD))
1341.40 (798.20)
##   rent_10_to_15_percent (mean (SD))
```

```
1022.22 (673.58)
##   rent_under_10_percent (mean (SD))
581.98 (545.45)
##   owner_occupied_housing_units (mean (SD))
6776.15 (3860.48)
##   million_dollar_housing_units (mean (SD))
479.04 (582.47)
##   mortgaged_housing_units (mean (SD))
4018.00 (2455.07)
##   different_house_year_ago_different_city (mean (SD))
838.22 (837.50)
##   different_house_year_ago_same_city (mean (SD))
3713.85 (1815.79)
##   families_with_young_children (mean (SD))
4133.24 (1804.03)
##   two_parent_families_with_young_children (mean (SD))
2424.73 (1509.21)
##   two_parents_in_labor_force_families_with_young_children (mean (SD))
1331.35 (831.50)
##   two_parents_father_in_labor_force_families_with_young_children (mean (SD
))   947.22 (723.25)
##   two_parents_mother_in_labor_force_families_with_young_children (mean (SD
))    90.02 (83.09)
##   two_parents_not_in_labor_force_families_with_young_children (mean (SD))
56.15 (62.92)
##   one_parent_families_with_young_children (mean (SD))
1708.51 (956.76)
##   father_one_parent_families_with_young_children (mean (SD))
324.69 (198.67)
##   father_in_labor_force_one_parent_families_with_young_children (mean (SD)
)    273.47 (171.55)
##   commute_less_10_mins (mean (SD))
931.56 (513.60)
##   commute_10_14_mins (mean (SD))
1230.33 (580.82)
##   commute_15_19_mins (mean (SD))
1573.27 (631.33)
##   commute_20_24_mins (mean (SD))
1959.13 (853.02)
##   commute_25_29_mins (mean (SD))
811.16 (528.18)
##   commute_30_34_mins (mean (SD))
3751.16 (1752.97)
##   commute_35_44_mins (mean (SD))
2491.09 (1294.20)
##   commute_60_more_mins (mean (SD))
7645.49 (2986.46)
##   commute_45_59_mins (mean (SD))
4245.16 (2070.78)
##   commuters_16_over (mean (SD))
```

```
24638.36 (8202.12)
##   walked_to_work (mean (SD))
2049.80 (1723.55)
##   worked_at_home (mean (SD))
963.00 (722.84)
##   no_car (mean (SD))
10563.38 (7897.74)
##   no_cars (mean (SD))
10718.45 (7392.30)
##   one_car (mean (SD))
6649.73 (2926.10)
##   two_cars (mean (SD))
2444.31 (1645.43)
##   three_cars (mean (SD))
540.33 (549.07)
##   four_more_cars (mean (SD))
184.76 (215.58)
##   commuters_by_public_transportation (mean (SD))
14484.87 (6942.74)
##   commuters_by_bus (mean (SD))
2938.53 (1054.24)
##   commuters_by_car_truck_van (mean (SD))
7529.16 (3845.10)
##   commuters_by_carpool (mean (SD))
1313.56 (926.61)
##   commuters_by_subway_or_elevated (mean (SD))
11065.98 (7189.69)
##   commuters_drove_alone (mean (SD))
6215.60 (3362.47)
##   group_quarters (mean (SD))
1315.84 (1729.94)
##   associates_degree (mean (SD))
2698.69 (911.53)
##   bachelors_degree (mean (SD))
7582.18 (3916.57)
##   high_school_diploma (mean (SD))
8502.22 (3008.02)
##   less_one_year_college (mean (SD))
1413.16 (527.15)
##   masters_degree (mean (SD))
3651.18 (2682.73)
##   one_year_more_college (mean (SD))
4432.42 (1320.14)
##   less_than_high_school_graduate (mean (SD))
8632.02 (5259.90)
##   high_school_including_ged (mean (SD))
10189.11 (3338.83)
##   bachelors_degree_2 (mean (SD))
7582.18 (3916.57)
##   bachelors_degree_or_higher_25_64 (mean (SD))
```

```
10746.27 (6534.91)
##   graduate_professional_degree (mean (SD))
5043.42 (4322.06)
##   some_college_and_associates_degree (mean (SD))
8544.27 (2340.69)
##   male_45_64_associates_degree (mean (SD))
418.80 (179.20)
##   male_45_64_bachelors_degree (mean (SD))
1058.65 (567.42)
##   male_45_64_graduate_degree (mean (SD))
741.47 (652.72)
##   male_45_64_less_than_9_grade (mean (SD))
728.05 (531.31)
##   male_45_64_grade_9_12 (mean (SD))
800.56 (518.98)
##   male_45_64_high_school (mean (SD))
1947.75 (671.65)
##   male_45_64_some_college (mean (SD))
1007.49 (276.22)
##   male_45_to_64 (mean (SD))
6702.78 (2121.13)
##   employed_pop (mean (SD))
26313.73 (8914.50)
##   unemployed_pop (mean (SD))
2173.89 (1015.20)
##   pop_in_labor_force (mean (SD))
28506.00 (9565.54)
##   not_in_labor_force (mean (SD))
18199.60 (6249.60)
##   workers_16_and_over (mean (SD))
25601.36 (8730.54)
##   armed_forces (mean (SD))
18.38 (36.31)
##   civilian_labor_force (mean (SD))
28487.62 (9564.96)
##   employed_agriculture_forestry_fishing_hunting_mining (mean (SD))
22.40 (27.93)
##   employed_arts_entertainment_recreation_accommodation_food (mean (SD))
3044.11 (1854.71)
##   employed_construction (mean (SD))
1420.04 (761.30)
##   employed_education_health_social (mean (SD))
7496.24 (2538.38)
##   employed_finance_insurance_real_estate (mean (SD))
2007.91 (1036.36)
##   employed_information (mean (SD))
795.75 (662.84)
##   employed_manufacturing (mean (SD))
856.24 (506.78)
##   employed_other_services_not_public_admin (mean (SD))
```

```
1543.24 (743.34)
##   employed_public_administration (mean (SD))
1066.09 (452.37)
##   employed_retail_trade (mean (SD))
2582.95 (854.63)
##   employed_science_management_admin_waste (mean (SD))
3071.25 (1639.78)
##   employed_transportation_warehousing_utilities (mean (SD))
1850.07 (674.96)
##   employed_wholesale_trade (mean (SD))
557.45 (332.90)
##   occupation_management_arts (mean (SD))
9546.36 (5459.47)
##   occupation_natural_resources_construction_maintenance (mean (SD))
1756.04 (848.60)
##   occupation_production_transportation_material (mean (SD))
2666.29 (1072.28)
##   occupation_sales_office (mean (SD))
5649.56 (1724.22)
##   occupation_services (mean (SD))
6695.47 (2652.26)
##   management_business_sci_arts_employed (mean (SD))
9546.36 (5459.47)
##   sales_office_employed (mean (SD))
5649.56 (1724.22)
##   in_grades_1_to_4 (mean (SD))
2683.36 (1073.66)
##   in_grades_5_to_8 (mean (SD))
2628.89 (1026.67)
##   in_grades_9_to_12 (mean (SD))
2711.71 (991.87)
##   in_school (mean (SD))
14136.91 (4906.44)
##   in_undergrad_college (mean (SD))
3423.89 (1326.24)
##   Row (mean (SD))
97.00 (47.96)
##   population (mean (SD))
56074.73 (18389.03)
##   area (mean (SD))
1.90 (1.52)
##   peoplepersqmile (mean (SD))
42963.95 (27300.43)
##
Stratified by cluster
##
3
##   n
31
##   total_pop (mean (SD))
```

```
84131.45 (15059.57)
##   households (mean (SD))
28469.16 (5241.76)
##   male_pop (mean (SD))
39925.45 (7968.88)
##   female_pop (mean (SD))
44206.00 (7573.60)
##   median_age (mean (SD))
35.24 (4.21)
##   male_under_5 (mean (SD))
3187.68 (1079.99)
##   male_5_to_9 (mean (SD))
2834.87 (776.41)
##   male_10_to_14 (mean (SD))
2789.55 (570.30)
##   male_15_to_17 (mean (SD))
1656.90 (424.01)
##   male_18_to_19 (mean (SD))
981.97 (358.86)
##   male_20 (mean (SD))
601.90 (232.35)
##   male_21 (mean (SD))
561.58 (175.08)
##   male_22_to_24 (mean (SD))
1726.19 (478.21)
##   male_25_to_29 (mean (SD))
3482.39 (1109.52)
##   male_30_to_34 (mean (SD))
3188.68 (1192.32)
##   male_35_to_39 (mean (SD))
2771.16 (1007.23)
##   male_40_to_44 (mean (SD))
2589.26 (851.63)
##   male_45_to_49 (mean (SD))
2482.03 (644.29)
##   male_50_to_54 (mean (SD))
2534.16 (485.00)
##   male_55_to_59 (mean (SD))
2316.94 (498.12)
##   male_65_to_66 (mean (SD))
695.81 (239.30)
##   male_67_to_69 (mean (SD))
854.26 (286.09)
##   male_70_to_74 (mean (SD))
1033.42 (315.60)
##   male_75_to_79 (mean (SD))
750.29 (261.99)
##   male_80_to_84 (mean (SD))
492.74 (196.62)
##   male_85_and_over (mean (SD))
```

```
428.71 (212.69)
##   female_under_5 (mean (SD))
3134.52 (1022.32)
##   female_5_to_9 (mean (SD))
2778.90 (789.36)
##   female_10_to_14 (mean (SD))
2785.48 (767.86)
##   female_15_to_17 (mean (SD))
1571.81 (378.85)
##   female_18_to_19 (mean (SD))
961.68 (332.94)
##   female_20 (mean (SD))
576.52 (173.93)
##   female_21 (mean (SD))
600.16 (207.15)
##   female_22_to_24 (mean (SD))
1839.55 (481.36)
##   female_25_to_29 (mean (SD))
3634.65 (1099.63)
##   female_30_to_34 (mean (SD))
3373.52 (974.37)
##   female_35_to_39 (mean (SD))
3053.58 (756.56)
##   female_40_to_44 (mean (SD))
2791.23 (572.68)
##   female_45_to_49 (mean (SD))
2781.68 (601.58)
##   female_50_to_54 (mean (SD))
2813.71 (659.29)
##   female_55_to_59 (mean (SD))
2671.52 (627.68)
##   female_60_to_61 (mean (SD))
1080.61 (288.28)
##   female_62_to_64 (mean (SD))
1433.35 (392.18)
##   female_65_to_66 (mean (SD))
903.39 (212.45)
##   female_67_to_69 (mean (SD))
1110.71 (380.66)
##   female_70_to_74 (mean (SD))
1431.32 (364.66)
##   female_75_to_79 (mean (SD))
1114.61 (365.34)
##   female_80_to_84 (mean (SD))
809.32 (279.18)
##   female_85_and_over (mean (SD))
954.19 (410.43)
##   white_pop (mean (SD))
18162.13 (21616.50)
##   population_1_year_and_over (mean (SD))
```

```
82969.23 (14815.14)
##   population_3_years_over (mean (SD))
80428.45 (14228.47)
##   pop_16_over (mean (SD))
65538.29 (11876.80)
##   pop_25_years_over (mean (SD))
55542.19 (10628.59)
##   pop_25_64 (mean (SD))
44963.42 (9365.12)
##   not_us_citizen_pop (mean (SD))
16087.23 (8262.08)
##   black_pop (mean (SD))
24639.74 (23472.96)
##   asian_pop (mean (SD))
8423.32 (10247.88)
##   hispanic_pop (mean (SD))
31187.39 (20859.49)
##   amerindian_pop (mean (SD))
172.45 (163.54)
##   other_race_pop (mean (SD))
411.03 (278.15)
##   two_or_more_races_pop (mean (SD))
1119.94 (466.31)
##   hispanic_any_race (mean (SD))
31187.39 (20859.49)
##   not_hispanic_pop (mean (SD))
52944.06 (22380.87)
##   asian_male_45_54 (mean (SD))
628.58 (785.11)
##   asian_male_55_64 (mean (SD))
529.10 (681.25)
##   black_male_45_54 (mean (SD))
1586.94 (1363.32)
##   black_male_55_64 (mean (SD))
1398.00 (1280.03)
##   hispanic_male_45_54 (mean (SD))
1883.87 (1252.50)
##   hispanic_male_55_64 (mean (SD))
1335.94 (823.54)
##   white_male_45_54 (mean (SD))
970.45 (1135.07)
##   white_male_55_64 (mean (SD))
1044.52 (1195.99)
##   median_income (mean (SD))
50204.71 (15568.20)
##   income_per_capita (mean (SD))
24080.68 (6894.54)
##   income_less_10000 (mean (SD))
2813.39 (1421.44)
##   income_10000_14999 (mean (SD))
```

```
2125.87 (975.83)
##   income_15000_19999 (mean (SD))
1652.97 (570.30)
##   income_20000_24999 (mean (SD))
1505.39 (411.42)
##   income_25000_29999 (mean (SD))
1417.16 (346.47)
##   income_30000_34999 (mean (SD))
1389.29 (350.18)
##   income_35000_39999 (mean (SD))
1266.87 (317.12)
##   income_40000_44999 (mean (SD))
1296.13 (317.12)
##   income_45000_49999 (mean (SD))
1048.29 (243.38)
##   income_50000_59999 (mean (SD))
2130.45 (468.94)
##   income_60000_74999 (mean (SD))
2582.26 (652.48)
##   income_75000_99999 (mean (SD))
3133.68 (983.59)
##   income_100000_124999 (mean (SD))
2112.65 (860.94)
##   income_125000_149999 (mean (SD))
1336.03 (724.84)
##   income_150000_199999 (mean (SD))
1439.10 (903.90)
##   income_200000_or_more (mean (SD))
1219.65 (1187.11)
##   pop_determined_poverty_status (mean (SD))
83152.52 (15191.50)
##   poverty (mean (SD))
18590.84 (8589.38)
##   gini_index (mean (SD))
0.47 (0.04)
##   housing_units (mean (SD))
30640.39 (5877.82)
##   occupied_housing_units (mean (SD))
28469.16 (5241.76)
##   housing_units_renter_occupied (mean (SD))
20429.19 (6673.28)
##   vacant_housing_units (mean (SD))
2171.23 (969.51)
##   vacant_housing_units_for_rent (mean (SD))
581.90 (362.75)
##   vacant_housing_units_for_sale (mean (SD))
144.19 (120.10)
##   dwellings_1_units_detached (mean (SD))
2344.74 (2260.00)
##   dwellings_1_units_attached (mean (SD))
```

```
2639.19 (2631.74)
##   dwellings_2_units (mean (SD))
5201.23 (3907.48)
##   dwellings_3_to_4_units (mean (SD))
3954.23 (2882.25)
##   dwellings_5_to_9_units (mean (SD))
1781.81 (1920.59)
##   dwellings_10_to_19_units (mean (SD))
1442.84 (910.44)
##   dwellings_20_to_49_units (mean (SD))
5365.13 (4095.70)
##   dwellings_50_or_more_units (mean (SD))
7843.10 (4802.98)
##   mobile_homes (mean (SD))
44.77 (33.30)
##   housing_built_2005_or_later (mean (SD))
144.35 (235.66)
##   housing_built_2000_to_2004 (mean (SD))
431.65 (546.10)
##   housing_built_1939_or_earlier (mean (SD))
3421.55 (1276.47)
##   median_year_structure_built (mean (SD))
1448.74 (868.57)
##   married_households (mean (SD))
10529.23 (3873.98)
##   nonfamily_households (mean (SD))
9254.84 (2924.24)
##   family_households (mean (SD))
19214.32 (3413.47)
##   households_public_asst_or_food_stamps (mean (SD))
7926.26 (3817.50)
##   male_male_households (mean (SD))
40.55 (57.80)
##   female_female_households (mean (SD))
42.90 (35.97)
##   children (mean (SD))
20739.71 (5354.74)
##   children_in_single_female_hh (mean (SD))
7557.29 (4209.56)
##   rent_burden_not_computed (mean (SD))
1061.48 (450.62)
##   rent_over_50_percent (mean (SD))
6551.94 (2548.23)
##   rent_40_to_50_percent (mean (SD))
1841.13 (658.56)
##   rent_35_to_40_percent (mean (SD))
1316.81 (516.28)
##   rent_30_to_35_percent (mean (SD))
1784.03 (656.85)
##   rent_25_to_30_percent (mean (SD))
```

```
1954.23 (655.58)
##   rent_20_to_25_percent (mean (SD))
2070.97 (772.54)
##   rent_15_to_20_percent (mean (SD))
1887.90 (679.87)
##   rent_10_to_15_percent (mean (SD))
1310.13 (516.50)
##   rent_under_10_percent (mean (SD))
650.58 (340.85)
##   owner_occupied_housing_units (mean (SD))
8039.97 (5395.48)
##   million_dollar_housing_units (mean (SD))
404.48 (583.81)
##   mortgaged_housing_units (mean (SD))
5130.52 (3806.35)
##   different_house_year_ago_different_city (mean (SD))
798.65 (443.64)
##   different_house_year_ago_same_city (mean (SD))
4985.71 (1628.61)
##   families_with_young_children (mean (SD))
7141.10 (2356.05)
##   two_parent_families_with_young_children (mean (SD))
3977.84 (2415.19)
##   two_parents_in_labor_force_families_with_young_children (mean (SD))
2062.42 (1062.58)
##   two_parents_father_in_labor_force_families_with_young_children (mean (SD
))  1618.45 (1094.62)
##   two_parents_mother_in_labor_force_families_with_young_children (mean (SD
))   236.90 (388.39)
##   two_parents_not_in_labor_force_families_with_young_children (mean (SD))
60.06 (54.85)
##   one_parent_families_with_young_children (mean (SD))
3163.26 (1702.74)
##   father_one_parent_families_with_young_children (mean (SD))
509.61 (250.26)
##   father_in_labor_force_one_parent_families_with_young_children (mean (SD)
)    453.74 (220.15)
##   commute_less_10_mins (mean (SD))
1254.74 (541.22)
##   commute_10_14_mins (mean (SD))
1733.90 (677.10)
##   commute_15_19_mins (mean (SD))
2230.84 (805.66)
##   commute_20_24_mins (mean (SD))
2594.55 (732.79)
##   commute_25_29_mins (mean (SD))
1126.16 (654.25)
##   commute_30_34_mins (mean (SD))
5729.45 (2758.67)
##   commute_35_44_mins (mean (SD))
```

```
3597.03 (2149.54)
##   commute_60_more_mins (mean (SD))
11101.84 (3569.90)
##   commute_45_59_mins (mean (SD))
6173.84 (2691.55)
##   commuters_16_over (mean (SD))
35542.35 (8332.50)
##   walked_to_work (mean (SD))
2838.32 (1781.03)
##   worked_at_home (mean (SD))
1077.26 (566.67)
##   no_car (mean (SD))
15408.42 (7410.60)
##   no_cars (mean (SD))
14991.87 (5772.17)
##   one_car (mean (SD))
9645.48 (2631.76)
##   two_cars (mean (SD))
3071.61 (2306.19)
##   three_cars (mean (SD))
578.03 (552.17)
##   four_more_cars (mean (SD))
182.16 (234.09)
##   commuters_by_public_transportation (mean (SD))
21289.16 (7332.42)
##   commuters_by_bus (mean (SD))
4549.68 (2000.80)
##   commuters_by_car_truck_van (mean (SD))
10750.23 (5016.28)
##   commuters_by_carpool (mean (SD))
1791.13 (738.25)
##   commuters_by_subway_or_elevated (mean (SD))
16130.23 (7559.03)
##   commuters_drove_alone (mean (SD))
8959.10 (4452.02)
##   group_quarters (mean (SD))
1247.55 (935.83)
##   associates_degree (mean (SD))
3944.61 (979.32)
##   bachelors_degree (mean (SD))
9035.94 (4235.02)
##   high_school_diploma (mean (SD))
14232.71 (4395.62)
##   less_one_year_college (mean (SD))
2123.35 (978.49)
##   masters_degree (mean (SD))
3766.48 (2252.04)
##   one_year_more_college (mean (SD))
6470.71 (1712.51)
##   less_than_high_school_graduate (mean (SD))
```

```
12457.65 (4598.02)
##   high_school_including_ged (mean (SD))
16707.84 (4827.78)
##   bachelors_degree_2 (mean (SD))
9035.94 (4235.02)
##   bachelors_degree_or_higher_25_64 (mean (SD))
11949.87 (6142.71)
##   graduate_professional_degree (mean (SD))
4802.10 (2954.98)
##   some_college_and_associates_degree (mean (SD))
12538.68 (2902.33)
##   male_45_64_associates_degree (mean (SD))
598.10 (245.53)
##   male_45_64_bachelors_degree (mean (SD))
1188.87 (589.42)
##   male_45_64_graduate_degree (mean (SD))
671.71 (427.91)
##   male_45_64_less_than_9_grade (mean (SD))
1124.84 (633.15)
##   male_45_64_grade_9_12 (mean (SD))
1087.61 (354.46)
##   male_45_64_high_school (mean (SD))
3161.13 (990.13)
##   male_45_64_some_college (mean (SD))
1465.84 (366.97)
##   male_45_to_64 (mean (SD))
9298.10 (1925.90)
##   employed_pop (mean (SD))
37604.26 (8743.21)
##   unemployed_pop (mean (SD))
3159.26 (1120.82)
##   pop_in_labor_force (mean (SD))
40781.32 (8846.92)
##   not_in_labor_force (mean (SD))
24756.97 (4150.01)
##   workers_16_and_over (mean (SD))
36619.61 (8575.28)
##   armed_forces (mean (SD))
17.81 (19.27)
##   civilian_labor_force (mean (SD))
40763.52 (8848.82)
##   employed_agriculture_forestry_fishing_hunting_mining (mean (SD))
45.48 (43.09)
##   employed_arts_entertainment_recreation_accommodation_food (mean (SD))
4348.45 (2271.47)
##   employed_construction (mean (SD))
2366.13 (1500.45)
##   employed_education_health_social (mean (SD))
11085.90 (3125.18)
##   employed_finance_insurance_real_estate (mean (SD))
```

```
2440.32 (868.15)
##   employed_information (mean (SD))
869.61 (819.90)
##   employed_manufacturing (mean (SD))
1321.00 (563.86)
##   employed_other_services_not_public_admin (mean (SD))
2160.55 (934.38)
##   employed_public_administration (mean (SD))
1380.58 (607.47)
##   employed_retail_trade (mean (SD))
3911.29 (981.16)
##   employed_science_management_admin_waste (mean (SD))
3808.13 (1920.86)
##   employed_transportation_warehousing_utilities (mean (SD))
3131.74 (1027.15)
##   employed_wholesale_trade (mean (SD))
735.06 (381.89)
##   occupation_management_arts (mean (SD))
11063.16 (5262.61)
##   occupation_natural_resources_construction_maintenance (mean (SD))
2914.58 (1602.99)
##   occupation_production_transportation_material (mean (SD))
4602.16 (1523.38)
##   occupation_sales_office (mean (SD))
7916.03 (1750.43)
##   occupation_services (mean (SD))
11108.32 (3443.78)
##   management_business_sci_arts_employed (mean (SD))
11063.16 (5262.61)
##   sales_office_employed (mean (SD))
7916.03 (1750.43)
##   in_grades_1_to_4 (mean (SD))
4575.87 (1223.62)
##   in_grades_5_to_8 (mean (SD))
4491.26 (1028.85)
##   in_grades_9_to_12 (mean (SD))
4550.81 (1137.03)
##   in_school (mean (SD))
22410.03 (4771.93)
##   in_undergrad_college (mean (SD))
4963.03 (1219.23)
##   Row (mean (SD))
100.45 (35.28)
##   population (mean (SD))
81495.39 (14496.31)
##   area (mean (SD))
2.57 (2.42)
##   peoplepersqmile (mean (SD))
45911.35 (23887.80)
##
```

```
Stratified by cluster
##
p
##   n
##   total_pop (mean (SD))
<0.001
##   households (mean (SD))
<0.001
##   male_pop (mean (SD))
<0.001
##   female_pop (mean (SD))
<0.001
##   median_age (mean (SD))
0.003
##   male_under_5 (mean (SD))
<0.001
##   male_5_to_9 (mean (SD))
<0.001
##   male_10_to_14 (mean (SD))
<0.001
##   male_15_to_17 (mean (SD))
<0.001
##   male_18_to_19 (mean (SD))
<0.001
##   male_20 (mean (SD))
<0.001
##   male_21 (mean (SD))
<0.001
##   male_22_to_24 (mean (SD))
<0.001
##   male_25_to_29 (mean (SD))
<0.001
##   male_30_to_34 (mean (SD))
<0.001
##   male_35_to_39 (mean (SD))
<0.001
##   male_40_to_44 (mean (SD))
<0.001
##   male_45_to_49 (mean (SD))
<0.001
##   male_50_to_54 (mean (SD))
<0.001
##   male_55_to_59 (mean (SD))
<0.001
##   male_65_to_66 (mean (SD))
<0.001
##   male_67_to_69 (mean (SD))
<0.001
##   male_70_to_74 (mean (SD))
<0.001
```

```
##   male_75_to_79 (mean (SD))
<0.001
##   male_80_to_84 (mean (SD))
<0.001
##   male_85_and_over (mean (SD))
<0.001
##   female_under_5 (mean (SD))
<0.001
##   female_5_to_9 (mean (SD))
<0.001
##   female_10_to_14 (mean (SD))
<0.001
##   female_15_to_17 (mean (SD))
<0.001
##   female_18_to_19 (mean (SD))
<0.001
##   female_20 (mean (SD))
<0.001
##   female_21 (mean (SD))
<0.001
##   female_22_to_24 (mean (SD))
<0.001
##   female_25_to_29 (mean (SD))
<0.001
##   female_30_to_34 (mean (SD))
<0.001
##   female_35_to_39 (mean (SD))
<0.001
##   female_40_to_44 (mean (SD))
<0.001
##   female_45_to_49 (mean (SD))
<0.001
##   female_50_to_54 (mean (SD))
<0.001
##   female_55_to_59 (mean (SD))
<0.001
##   female_60_to_61 (mean (SD))
<0.001
##   female_62_to_64 (mean (SD))
<0.001
##   female_65_to_66 (mean (SD))
<0.001
##   female_67_to_69 (mean (SD))
<0.001
##   female_70_to_74 (mean (SD))
<0.001
##   female_75_to_79 (mean (SD))
<0.001
##   female_80_to_84 (mean (SD))
<0.001
```

```
##   female_85_and_over (mean (SD))
<0.001
##   white_pop (mean (SD))
0.435
##   population_1_year_and_over (mean (SD))
<0.001
##   population_3_years_over (mean (SD))
<0.001
##   pop_16_over (mean (SD))
<0.001
##   pop_25_years_over (mean (SD))
<0.001
##   pop_25_64 (mean (SD))
<0.001
##   not_us_citizen_pop (mean (SD))
<0.001
##   black_pop (mean (SD))
<0.001
##   asian_pop (mean (SD))
0.003
##   hispanic_pop (mean (SD))
<0.001
##   amerindian_pop (mean (SD))
<0.001
##   other_race_pop (mean (SD))
0.023
##   two_or_more_races_pop (mean (SD))
<0.001
##   hispanic_any_race (mean (SD))
<0.001
##   not_hispanic_pop (mean (SD))
<0.001
##   asian_male_45_54 (mean (SD))
0.002
##   asian_male_55_64 (mean (SD))
<0.001
##   black_male_45_54 (mean (SD))
<0.001
##   black_male_55_64 (mean (SD))
<0.001
##   hispanic_male_45_54 (mean (SD))
<0.001
##   hispanic_male_55_64 (mean (SD))
<0.001
##   white_male_45_54 (mean (SD))
0.986
##   white_male_55_64 (mean (SD))
0.495
##   median_income (mean (SD))
<0.001
```

```
##   income_per_capita (mean (SD))
<0.001
##   income_less_10000 (mean (SD))
<0.001
##   income_10000_14999 (mean (SD))
<0.001
##   income_15000_19999 (mean (SD))
<0.001
##   income_20000_24999 (mean (SD))
<0.001
##   income_25000_29999 (mean (SD))
<0.001
##   income_30000_34999 (mean (SD))
<0.001
##   income_35000_39999 (mean (SD))
<0.001
##   income_40000_44999 (mean (SD))
<0.001
##   income_45000_49999 (mean (SD))
<0.001
##   income_50000_59999 (mean (SD))
<0.001
##   income_60000_74999 (mean (SD))
<0.001
##   income_75000_99999 (mean (SD))
<0.001
##   income_100000_124999 (mean (SD))
<0.001
##   income_125000_149999 (mean (SD))
0.001
##   income_150000_199999 (mean (SD))
0.343
##   income_200000_or_more (mean (SD))
0.008
##   pop_determined_poverty_status (mean (SD))
<0.001
##   poverty (mean (SD))
<0.001
##   gini_index (mean (SD))
0.432
##   housing_units (mean (SD))
<0.001
##   occupied_housing_units (mean (SD))
<0.001
##   housing_units_renter_occupied (mean (SD))
<0.001
##   vacant_housing_units (mean (SD))
0.271
##   vacant_housing_units_for_rent (mean (SD))
0.031
```

```
##   vacant_housing_units_for_sale (mean (SD))
0.016
##   dwellings_1_units_detached (mean (SD))
0.001
##   dwellings_1_units_attached (mean (SD))
<0.001
##   dwellings_2_units (mean (SD))
<0.001
##   dwellings_3_to_4_units (mean (SD))
<0.001
##   dwellings_5_to_9_units (mean (SD))
0.216
##   dwellings_10_to_19_units (mean (SD))
0.522
##   dwellings_20_to_49_units (mean (SD))
<0.001
##   dwellings_50_or_more_units (mean (SD))
0.096
##   mobile_homes (mean (SD))
<0.001
##   housing_built_2005_or_later (mean (SD))
0.489
##   housing_built_2000_to_2004 (mean (SD))
0.139
##   housing_built_1939_or_earlier (mean (SD))
<0.001
##   median_year_structure_built (mean (SD))
0.856
##   married_households (mean (SD))
<0.001
##   nonfamily_households (mean (SD))
0.009
##   family_households (mean (SD))
<0.001
##   households_public_asst_or_food_stamps (mean (SD))
<0.001
##   male_male_households (mean (SD))
0.058
##   female_female_households (mean (SD))
0.090
##   children (mean (SD))
<0.001
##   children_in_single_female_hh (mean (SD))
<0.001
##   rent_burden_not_computed (mean (SD))
<0.001
##   rent_over_50_percent (mean (SD))
<0.001
##   rent_40_to_50_percent (mean (SD))
<0.001
```

```
##   rent_35_to_40_percent (mean (SD))
<0.001
##   rent_30_to_35_percent (mean (SD))
<0.001
##   rent_25_to_30_percent (mean (SD))
<0.001
##   rent_20_to_25_percent (mean (SD))
<0.001
##   rent_15_to_20_percent (mean (SD))
<0.001
##   rent_10_to_15_percent (mean (SD))
<0.001
##   rent_under_10_percent (mean (SD))
0.875
##   owner_occupied_housing_units (mean (SD))
<0.001
##   million_dollar_housing_units (mean (SD))
0.771
##   mortgaged_housing_units (mean (SD))
<0.001
##   different_house_year_ago_different_city (mean (SD))
0.500
##   different_house_year_ago_same_city (mean (SD))
<0.001
##   families_with_young_children (mean (SD))
<0.001
##   two_parent_families_with_young_children (mean (SD))
<0.001
##   two_parents_in_labor_force_families_with_young_children (mean (SD))
<0.001
##   two_parents_father_in_labor_force_families_with_young_children (mean (SD
)) <0.001
##   two_parents_mother_in_labor_force_families_with_young_children (mean (SD
)) <0.001
##   two_parents_not_in_labor_force_families_with_young_children (mean (SD))
<0.001
##   one_parent_families_with_young_children (mean (SD))
<0.001
##   father_one_parent_families_with_young_children (mean (SD))
<0.001
##   father_in_labor_force_one_parent_families_with_young_children (mean (SD)
)  <0.001
##   commute_less_10_mins (mean (SD))
<0.001
##   commute_10_14_mins (mean (SD))
<0.001
##   commute_15_19_mins (mean (SD))
<0.001
##   commute_20_24_mins (mean (SD))
0.008
```

```
##   commute_25_29_mins (mean (SD))
0.151
##   commute_30_34_mins (mean (SD))
<0.001
##   commute_35_44_mins (mean (SD))
<0.001
##   commute_60_more_mins (mean (SD))
<0.001
##   commute_45_59_mins (mean (SD))
<0.001
##   commuters_16_over (mean (SD))
<0.001
##   walked_to_work (mean (SD))
0.244
##   worked_at_home (mean (SD))
0.361
##   no_car (mean (SD))
<0.001
##   no_cars (mean (SD))
<0.001
##   one_car (mean (SD))
<0.001
##   two_cars (mean (SD))
<0.001
##   three_cars (mean (SD))
<0.001
##   four_more_cars (mean (SD))
0.001
##   commuters_by_public_transportation (mean (SD))
<0.001
##   commuters_by_bus (mean (SD))
<0.001
##   commuters_by_car_truck_van (mean (SD))
<0.001
##   commuters_by_carpool (mean (SD))
<0.001
##   commuters_by_subway_or_elevated (mean (SD))
<0.001
##   commuters_drove_alone (mean (SD))
<0.001
##   group_quarters (mean (SD))
0.030
##   associates_degree (mean (SD))
<0.001
##   bachelors_degree (mean (SD))
0.027
##   high_school_diploma (mean (SD))
<0.001
##   less_one_year_college (mean (SD))
<0.001
```

```
##   masters_degree (mean (SD))
0.973
##   one_year_more_college (mean (SD))
<0.001
##   less_than_high_school_graduate (mean (SD))
<0.001
##   high_school_including_ged (mean (SD))
<0.001
##   bachelors_degree_2 (mean (SD))
0.027
##   bachelors_degree_or_higher_25_64 (mean (SD))
0.587
##   graduate_professional_degree (mean (SD))
0.768
##   some_college_and_associates_degree (mean (SD))
<0.001
##   male_45_64_associates_degree (mean (SD))
<0.001
##   male_45_64_bachelors_degree (mean (SD))
0.008
##   male_45_64_graduate_degree (mean (SD))
0.761
##   male_45_64_less_than_9_grade (mean (SD))
<0.001
##   male_45_64_grade_9_12 (mean (SD))
<0.001
##   male_45_64_high_school (mean (SD))
<0.001
##   male_45_64_some_college (mean (SD))
<0.001
##   male_45_to_64 (mean (SD))
<0.001
##   employed_pop (mean (SD))
<0.001
##   unemployed_pop (mean (SD))
<0.001
##   pop_in_labor_force (mean (SD))
<0.001
##   not_in_labor_force (mean (SD))
<0.001
##   workers_16_and_over (mean (SD))
<0.001
##   armed_forces (mean (SD))
0.040
##   civilian_labor_force (mean (SD))
<0.001
##   employed_agriculture_forestry_fishing_hunting_mining (mean (SD))
<0.001
##   employed_arts_entertainment_recreation_accommodation_food (mean (SD))
<0.001
```

```
##   employed_construction (mean (SD))
<0.001
##   employed_education_health_social (mean (SD))
<0.001
##   employed_finance_insurance_real_estate (mean (SD))
0.488
##   employed_information (mean (SD))
0.685
##   employed_manufacturing (mean (SD))
<0.001
##   employed_other_services_not_public_admin (mean (SD))
<0.001
##   employed_public_administration (mean (SD))
<0.001
##   employed_retail_trade (mean (SD))
<0.001
##   employed_science_management_admin_waste (mean (SD))
0.173
##   employed_transportation_warehousing_utilities (mean (SD))
<0.001
##   employed_wholesale_trade (mean (SD))
<0.001
##   occupation_management_arts (mean (SD))
0.252
##   occupation_natural_resources_construction_maintenance (mean (SD))
<0.001
##   occupation_production_transportation_material (mean (SD))
<0.001
##   occupation_sales_office (mean (SD))
<0.001
##   occupation_services (mean (SD))
<0.001
##   management_business_sci_arts_employed (mean (SD))
0.252
##   sales_office_employed (mean (SD))
<0.001
##   in_grades_1_to_4 (mean (SD))
<0.001
##   in_grades_5_to_8 (mean (SD))
<0.001
##   in_grades_9_to_12 (mean (SD))
<0.001
##   in_school (mean (SD))
<0.001
##   in_undergrad_college (mean (SD))
<0.001
##   Row (mean (SD))
0.062
##   population (mean (SD))
<0.001
```

```
##   area (mean (SD))
<0.001
##   peoplepersqmile (mean (SD))
0.777
##
Stratified by cluster
##
test
##   n
##   total_pop (mean (SD))
##   households (mean (SD))
##   male_pop (mean (SD))
##   female_pop (mean (SD))
##   median_age (mean (SD))
##   male_under_5 (mean (SD))
##   male_5_to_9 (mean (SD))
##   male_10_to_14 (mean (SD))
##   male_15_to_17 (mean (SD))
##   male_18_to_19 (mean (SD))
##   male_20 (mean (SD))
##   male_21 (mean (SD))
##   male_22_to_24 (mean (SD))
##   male_25_to_29 (mean (SD))
##   male_30_to_34 (mean (SD))
##   male_35_to_39 (mean (SD))
##   male_40_to_44 (mean (SD))
##   male_45_to_49 (mean (SD))
##   male_50_to_54 (mean (SD))
##   male_55_to_59 (mean (SD))
##   male_65_to_66 (mean (SD))
##   male_67_to_69 (mean (SD))
##   male_70_to_74 (mean (SD))
##   male_75_to_79 (mean (SD))
##   male_80_to_84 (mean (SD))
##   male_85_and_over (mean (SD))
##   female_under_5 (mean (SD))
##   female_5_to_9 (mean (SD))
##   female_10_to_14 (mean (SD))
##   female_15_to_17 (mean (SD))
##   female_18_to_19 (mean (SD))
##   female_20 (mean (SD))
##   female_21 (mean (SD))
##   female_22_to_24 (mean (SD))
##   female_25_to_29 (mean (SD))
##   female_30_to_34 (mean (SD))
##   female_35_to_39 (mean (SD))
##   female_40_to_44 (mean (SD))
##   female_45_to_49 (mean (SD))
##   female_50_to_54 (mean (SD))
##   female_55_to_59 (mean (SD))
```

```
##   female_60_to_61 (mean (SD))
##   female_62_to_64 (mean (SD))
##   female_65_to_66 (mean (SD))
##   female_67_to_69 (mean (SD))
##   female_70_to_74 (mean (SD))
##   female_75_to_79 (mean (SD))
##   female_80_to_84 (mean (SD))
##   female_85_and_over (mean (SD))
##   white_pop (mean (SD))
##   population_1_year_and_over (mean (SD))
##   population_3_years_over (mean (SD))
##   pop_16_over (mean (SD))
##   pop_25_years_over (mean (SD))
##   pop_25_64 (mean (SD))
##   not_us_citizen_pop (mean (SD))
##   black_pop (mean (SD))
##   asian_pop (mean (SD))
##   hispanic_pop (mean (SD))
##   amerindian_pop (mean (SD))
##   other_race_pop (mean (SD))
##   two_or_more_races_pop (mean (SD))
##   hispanic_any_race (mean (SD))
##   not_hispanic_pop (mean (SD))
##   asian_male_45_54 (mean (SD))
##   asian_male_55_64 (mean (SD))
##   black_male_45_54 (mean (SD))
##   black_male_55_64 (mean (SD))
##   hispanic_male_45_54 (mean (SD))
##   hispanic_male_55_64 (mean (SD))
##   white_male_45_54 (mean (SD))
##   white_male_55_64 (mean (SD))
##   median_income (mean (SD))
##   income_per_capita (mean (SD))
##   income_less_10000 (mean (SD))
##   income_10000_14999 (mean (SD))
##   income_15000_19999 (mean (SD))
##   income_20000_24999 (mean (SD))
##   income_25000_29999 (mean (SD))
##   income_30000_34999 (mean (SD))
##   income_35000_39999 (mean (SD))
##   income_40000_44999 (mean (SD))
##   income_45000_49999 (mean (SD))
##   income_50000_59999 (mean (SD))
##   income_60000_74999 (mean (SD))
##   income_75000_99999 (mean (SD))
##   income_100000_124999 (mean (SD))
##   income_125000_149999 (mean (SD))
##   income_150000_199999 (mean (SD))
##   income_200000_or_more (mean (SD))
##   pop_determined_poverty_status (mean (SD))
```

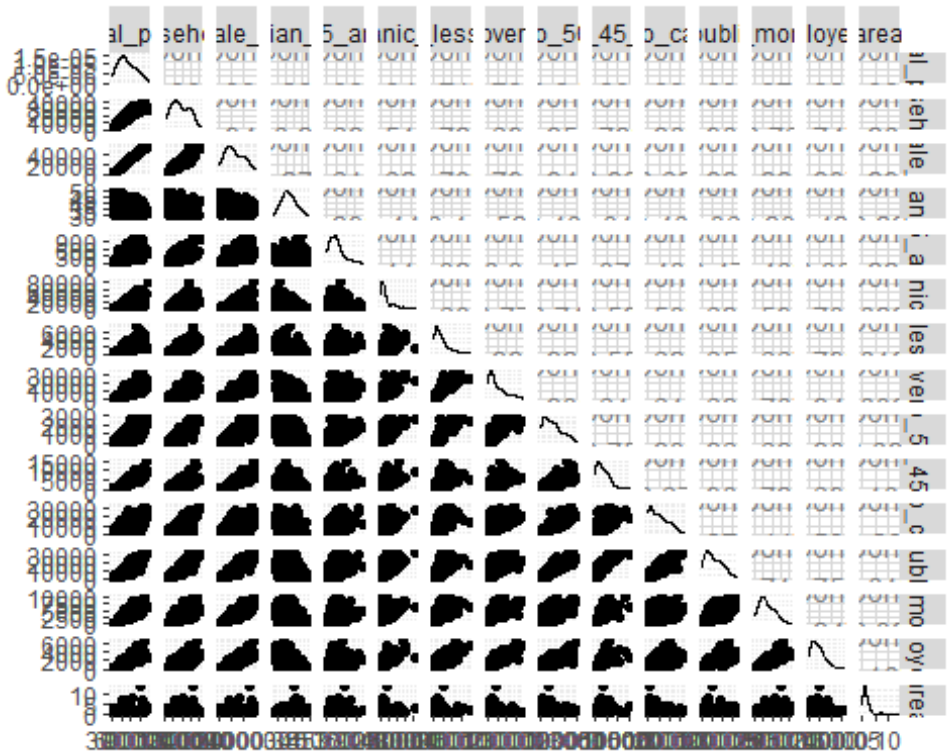
```
##   poverty (mean (SD))
##   gini_index (mean (SD))
##   housing_units (mean (SD))
##   occupied_housing_units (mean (SD))
##   housing_units_renter_occupied (mean (SD))
##   vacant_housing_units (mean (SD))
##   vacant_housing_units_for_rent (mean (SD))
##   vacant_housing_units_for_sale (mean (SD))
##   dwellings_1_units_detached (mean (SD))
##   dwellings_1_units_attached (mean (SD))
##   dwellings_2_units (mean (SD))
##   dwellings_3_to_4_units (mean (SD))
##   dwellings_5_to_9_units (mean (SD))
##   dwellings_10_to_19_units (mean (SD))
##   dwellings_20_to_49_units (mean (SD))
##   dwellings_50_or_more_units (mean (SD))
##   mobile_homes (mean (SD))
##   housing_built_2005_or_later (mean (SD))
##   housing_built_2000_to_2004 (mean (SD))
##   housing_built_1939_or_earlier (mean (SD))
##   median_year_structure_built (mean (SD))
##   married_households (mean (SD))
##   nonfamily_households (mean (SD))
##   family_households (mean (SD))
##   households_public_asst_or_food_stamps (mean (SD))
##   male_male_households (mean (SD))
##   female_female_households (mean (SD))
##   children (mean (SD))
##   children_in_single_female_hh (mean (SD))
##   rent_burden_not_computed (mean (SD))
##   rent_over_50_percent (mean (SD))
##   rent_40_to_50_percent (mean (SD))
##   rent_35_to_40_percent (mean (SD))
##   rent_30_to_35_percent (mean (SD))
##   rent_25_to_30_percent (mean (SD))
##   rent_20_to_25_percent (mean (SD))
##   rent_15_to_20_percent (mean (SD))
##   rent_10_to_15_percent (mean (SD))
##   rent_under_10_percent (mean (SD))
##   owner_occupied_housing_units (mean (SD))
##   million_dollar_housing_units (mean (SD))
##   mortgaged_housing_units (mean (SD))
##   different_house_year_ago_different_city (mean (SD))
##   different_house_year_ago_same_city (mean (SD))
##   families_with_young_children (mean (SD))
##   two_parent_families_with_young_children (mean (SD))
##   two_parents_in_labor_force_families_with_young_children (mean (SD))
##   two_parents_father_in_labor_force_families_with_young_children (mean (SD
))
##   two_parents_mother_in_labor_force_families_with_young_children (mean (SD
```

```
))
##   two_parents_not_in_labor_force_families_with_young_children (mean (SD))
##   one_parent_families_with_young_children (mean (SD))
##   father_one_parent_families_with_young_children (mean (SD))
##   father_in_labor_force_one_parent_families_with_young_children (mean (SD)
)
##   commute_less_10_mins (mean (SD))
##   commute_10_14_mins (mean (SD))
##   commute_15_19_mins (mean (SD))
##   commute_20_24_mins (mean (SD))
##   commute_25_29_mins (mean (SD))
##   commute_30_34_mins (mean (SD))
##   commute_35_44_mins (mean (SD))
##   commute_60_more_mins (mean (SD))
##   commute_45_59_mins (mean (SD))
##   commuters_16_over (mean (SD))
##   walked_to_work (mean (SD))
##   worked_at_home (mean (SD))
##   no_car (mean (SD))
##   no_cars (mean (SD))
##   one_car (mean (SD))
##   two_cars (mean (SD))
##   three_cars (mean (SD))
##   four_more_cars (mean (SD))
##   commuters_by_public_transportation (mean (SD))
##   commuters_by_bus (mean (SD))
##   commuters_by_car_truck_van (mean (SD))
##   commuters_by_carpool (mean (SD))
##   commuters_by_subway_or_elevated (mean (SD))
##   commuters_drove_alone (mean (SD))
##   group_quarters (mean (SD))
##   associates_degree (mean (SD))
##   bachelors_degree (mean (SD))
##   high_school_diploma (mean (SD))
##   less_one_year_college (mean (SD))
##   masters_degree (mean (SD))
##   one_year_more_college (mean (SD))
##   less_than_high_school_graduate (mean (SD))
##   high_school_including_ged (mean (SD))
##   bachelors_degree_2 (mean (SD))
##   bachelors_degree_or_higher_25_64 (mean (SD))
##   graduate_professional_degree (mean (SD))
##   some_college_and_associates_degree (mean (SD))
##   male_45_64_associates_degree (mean (SD))
##   male_45_64_bachelors_degree (mean (SD))
##   male_45_64_graduate_degree (mean (SD))
##   male_45_64_less_than_9_grade (mean (SD))
##   male_45_64_grade_9_12 (mean (SD))
##   male_45_64_high_school (mean (SD))
##   male_45_64_some_college (mean (SD))
```

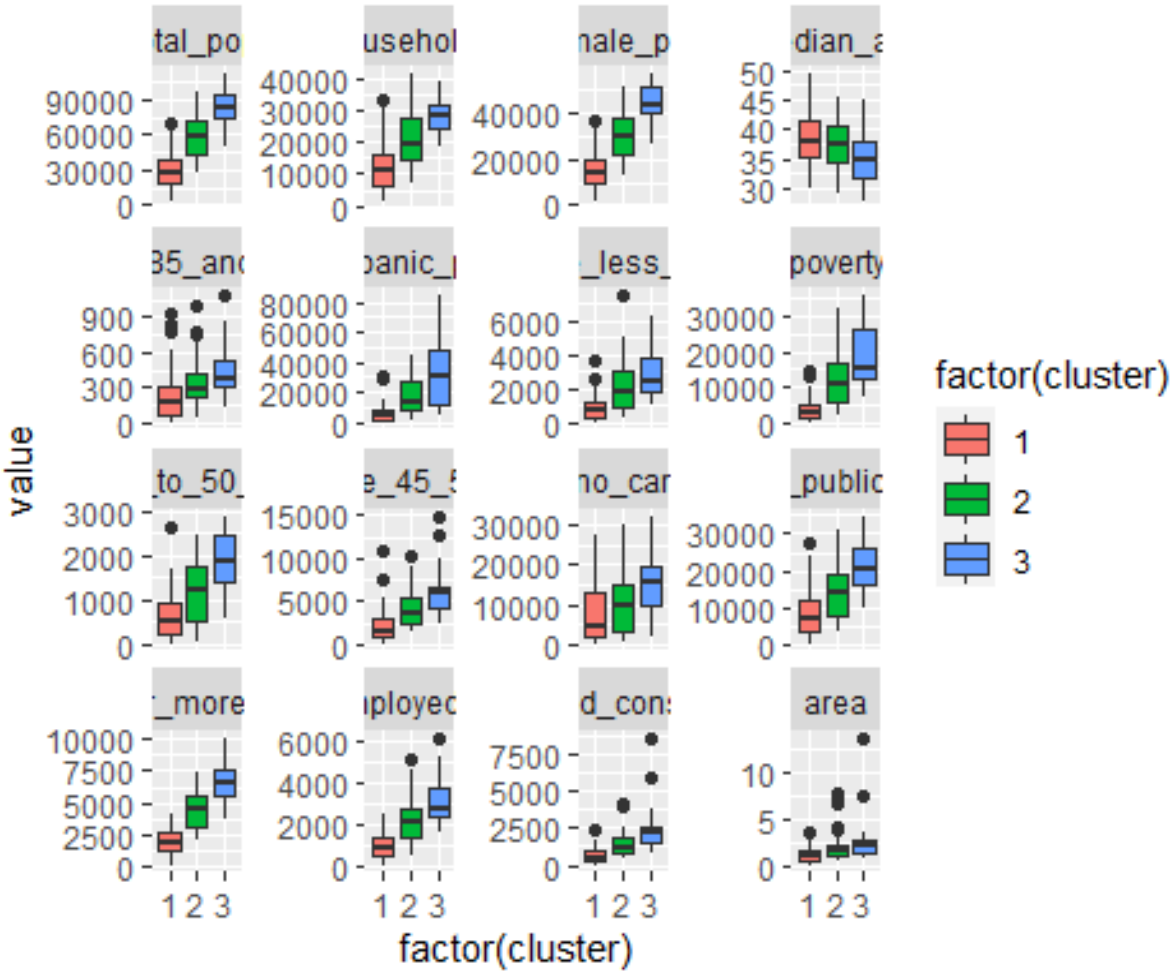
```
##   male_45_to_64 (mean (SD))
##   employed_pop (mean (SD))
##   unemployed_pop (mean (SD))
##   pop_in_labor_force (mean (SD))
##   not_in_labor_force (mean (SD))
##   workers_16_and_over (mean (SD))
##   armed_forces (mean (SD))
##   civilian_labor_force (mean (SD))
##   employed_agriculture_forestry_fishing_hunting_mining (mean (SD))
##   employed_arts_entertainment_recreation_accommodation_food (mean (SD))
##   employed_construction (mean (SD))
##   employed_education_health_social (mean (SD))
##   employed_finance_insurance_real_estate (mean (SD))
##   employed_information (mean (SD))
##   employed_manufacturing (mean (SD))
##   employed_other_services_not_public_admin (mean (SD))
##   employed_public_administration (mean (SD))
##   employed_retail_trade (mean (SD))
##   employed_science_management_admin_waste (mean (SD))
##   employed_transportation_warehousing_utilities (mean (SD))
##   employed_wholesale_trade (mean (SD))
##   occupation_management_arts (mean (SD))
##   occupation_natural_resources_construction_maintenance (mean (SD))
##   occupation_production_transportation_material (mean (SD))
##   occupation_sales_office (mean (SD))
##   occupation_services (mean (SD))
##   management_business_sci_arts_employed (mean (SD))
##   sales_office_employed (mean (SD))
##   in_grades_1_to_4 (mean (SD))
##   in_grades_5_to_8 (mean (SD))
##   in_grades_9_to_12 (mean (SD))
##   in_school (mean (SD))
##   in_undergrad_college (mean (SD))
##   Row (mean (SD))
##   population (mean (SD))
##   area (mean (SD))
##   peoplepersqmile (mean (SD))
```

Comparison of groups(boxplots)

```
ggpairs(GROUPFULL[, c("total_pop","households","female_pop","median_age", "male_85_and_over", "hispanic_pop", "income_less_10000","poverty","rent_40_to_50_percent","commute_45_59_mins","no_car","commuters_by_public_transportation", "one_year_more_college","unemployed_pop","area")])
```



```
#There do not seem to be any strong linear relations among our continuous pre
dictors
#If most your predictors appear independent of each other, that is fine. It s
hapes your expectations of the model. For example, if they are independent, t
he estimate for one predictor should not change much when you enter another p
redictor (although the standard error and significance tests may). We can get
all of this information and intuition about what and how to model are data by
simply viewing it.



tmp <- melt(GROUPFULL[, c("cluster","total_pop","households","female_pop","me
dian_age", "male_85_and_over", "hispanic_pop", "income_less_10000","poverty",
"rent_40_to_50_percent","commute_45_59_mins","no_car","commuters_by_public_tr
ansportation","one_year_more_college","unemployed_pop","employed_construction
","area")],id.vars="cluster")
ggplot(tmp, aes(factor(cluster), y = value, fill=factor(cluster))) +
  geom_boxplot() +
  facet_wrap(~variable, scales="free_y")
```


tal_po
usehol
nale_p
dian_a
35_and
panic_
less
poverty
_to_50_
e_45_5
no_car
_public
_more
ployed
d_con
area
value
factor(cluster)
1
2
3
1 2 3