Example

Exploring the Data

Data came from Stat 495 final project. (use info from project...). Needed a sample of 1000...

Importing the data:

```
#using data from final stat 495 project
library(readr)
data_subset <- read_csv("CopyOfdata_subset.csv")

set.seed(1)
#getting a sample of 1000 observations
mysample <- data_subset[sample(1:nrow(data_subset), 1000,
   replace=FALSE),]
```

Picking variables to focus on–> expanding conclusions from Stat 495 project

```
#only keeping the variables I want to look at
myvars <- c("Latitude_tri", "Longitude_tri", "poor_or_fair_health", "poor_physical_health_days", "physi
smallsample <- mysample[myvars]
```

Applying CLARA

Step 1: finding k

```
#finding k with project data, using Elbow Method
pkgs <- c("factoextra",  "NbClust")
install.packages(pkgs)

library(factoextra)
library(NbClust)
library(ggplot2)

# Elbow method
fviz_nbclust(new, kmeans, method = "wss") +
    geom_vline(xintercept = 4, linetype = 2)+
  labs(subtitle = "Elbow method")
```

Step 2: Run CLARA function

```
new<- na.omit(smallsample)

## run CLARA
clarasamp <- clara(new[1:6], 4)
```

```
## print components of clarax
print(clarasamp)
summary(clarasamp)
```

```
## plot clusters
plot(new, col = clarasamp$cluster)
## plot centers
points(clarasamp$centers, col = 1:2, pch = 8)
```

```r
#plotting clara
factoextra::fviz_cluster(clarasamp)
```

Evaluation of CLARA

Model to Predict Cluster

First, had to include a cluster variable in the original data set, using the data provided by the CLARA function.

```r
#adding each data point's cluster #
cluster<- clarasamp$clustering
cluster_data<- cbind(new, cluster)
```