

# Components of Healthy Longevity across Chinese Elders

Jessica Yu

## Introduction

We live in a capitalistic society that is built for the younger population, but aging is an inevitable process that everyone will go through; although there is not much of an emphasis on elder well being, it is imperative that we understand how to take care of the aging population and support all members of our community. Living long is not the only aspect of aging, but rather the kind of life lived is important too. Elders are not a monolithic group, and current analysis should not treat them as such. There are few datasets that exist that capture the details and day-to-day habits of elders. In this analysis, using the Chinese Longitudinal Healthy Longevity Survey data, we hope to group different types of elders based on a large number of characteristics, and identify what main differences exist between them. We also want to pinpoint underlying variables that contribute to overall quality of life. We will be employing factor analysis and clustering as multivariate analysis methods for our data. These two questions are asked under a broader picture of wanting to understand the elder population as a whole.

Ultimately with this information, we can better understand an often forgotten sub-population of individuals.

## Preliminary Analysis

The Chinese Longitudinal Healthy Longevity Survey is a large scale assessment of health status and quality of life of a sample of Chinese elders (age 65+) from 1998 to 2014 (Zeng 2014). Elders were randomly sampled across 22 provinces in China, and are representative of the elder Chinese population. Data and information were collected through detailed interviews of the elders. There are over 877 variables in the original dataset, and a majority are factor type variables. For our analysis, we needed to use only numeric variables, so the dataset was narrowed down to 16 numeric variables of interest, and the details of each variable are listed below. The variables take on a large range of values, and include measurements of health, education, and access to healthcare.

`num_people_living_with` - number of people the elder is living with `age_began_drinking` - age when elder started drinking `years_of_schooling` - years of total schooling `heart_rate` - heart rate (beats/min) `attempts_repeat_correctly` - attempts to repeat the names of three objects correctly `num_natural_teeth` - number of natural teeth `num_children` - number of children `age_started_exercise` - age when started to exercise `age_started_physlabor` - age when started physical labor `waist_circumference` - waist circumference in centimeters `hours_sleep` - how many hours slept normally `household_income` - total household income

`hours_grandchildren_help_elder` - how many hours last week did grandchildren spend helping elder `distance_from_hospitalkm` - distance from home to hospital in kilometers `inpatient_costs` - money spent on inpatient costs in the past year

The original dataset also has 7192 observations (elders). As this was a survey administered to elders at their own consent and discretion, many of the variables of interest contained NA values; once we omitted observations with NA values, we narrowed our dataset down to 268 observations. While this isn't the most ideal size for analysis and generalizability, we can still use our multivariate analysis methods and draw some interesting insights.

```
data <- read_csv("YuJData.csv")

e <- na.omit(data)

data_cleaned <- e %>%
  filter(
    num_people_living_with < 22,
    age_began_drinking < 102, years_of_schooling < 21, heart_rate < 99,
    num_natural_teeth < 45,
    num_children < 17, age_started_exercise < 102, age_started_physlabor < 90,
    waist_circumference < 150,
    hours_sleep < 25, household_income < 88880, hours_grandchildren_help_elder < 888,
    inpatient_costs < 88888, age_began_drinking > 0,
    age_started_exercise > 0
  ) %>%
  select(-X1)
```

In the dataset, NA inputs were oftentimes coded as different values ranging from '888', '-1', and '999'. We filtered the dataset to remove these NA values, which reduced our dataset as aforementioned.

## Data and Variance Summary

```
options(scipen = 999)
summary(data_cleaned)
```

```
## num_people_living_with age_began_drinking years_of_schooling heart_rate
## Min. :0.0 Min. : 6.0 Min. : 0.0 Min. :17.0
## 1st Qu.:1.0 1st Qu.:19.0 1st Qu.: 0.0 1st Qu.:68.0
## Median :2.0 Median :21.5 Median : 3.0 Median :72.0
## Mean :2.3 Mean :27.5 Mean : 3.3 Mean :73.4
## 3rd Qu.:3.0 3rd Qu.:30.0 3rd Qu.: 6.0 3rd Qu.:80.0
## Max. :9.0 Max. :90.0 Max. :17.0 Max. :98.0
## attempts_repeat_correctly num_natural_teeth num_children
## Min. :0.00 Min. : 0.00 Min. : 0.00
## 1st Qu.:0.00 1st Qu.: 0.00 1st Qu.: 3.00
## Median :0.00 Median : 6.00 Median : 4.00
## Mean :1.08 Mean : 9.73 Mean : 4.29
## 3rd Qu.:1.00 3rd Qu.:18.00 3rd Qu.: 6.00
## Max. :9.00 Max. :34.00 Max. :12.00
## age_started_excercise age_started_physlabor waist_circumfrence hours_sleep
## Min. : 9.0 Min. : 8.0 Min. : 27.0 Min. : 1.0
## 1st Qu.:22.8 1st Qu.:15.0 1st Qu.: 74.0 1st Qu.: 6.0
## Median :60.0 Median :17.0 Median : 81.0 Median : 8.0
## Mean :49.8 Mean :18.9 Mean : 81.1 Mean : 7.8
## 3rd Qu.:70.0 3rd Qu.:20.0 3rd Qu.: 88.0 3rd Qu.:10.0
## Max. :99.0 Max. :71.0 Max. :112.0 Max. :16.0
## household_income hours_grandchildren_help_elder TRUEAGE
## Min. : 0 Min. : 0.0 Min. : 67.0
## 1st Qu.:13080 1st Qu.: 0.0 1st Qu.: 77.0
## Median :30000 Median : 1.0 Median : 84.0
## Mean :31865 Mean : 19.4 Mean : 84.7
## 3rd Qu.:50000 3rd Qu.: 15.0 3rd Qu.: 91.0
## Max. :85000 Max. :168.0 Max. :114.0
## distance_from_hospitalkm inpatient_costs
## Min. : 0.00 Min. : 0
## 1st Qu.: 1.00 1st Qu.: 0
## Median : 1.00 Median : 0
## Mean : 1.96 Mean : 3843
## 3rd Qu.: 2.00 3rd Qu.: 2025
## Max. :23.00 Max. :80000
```

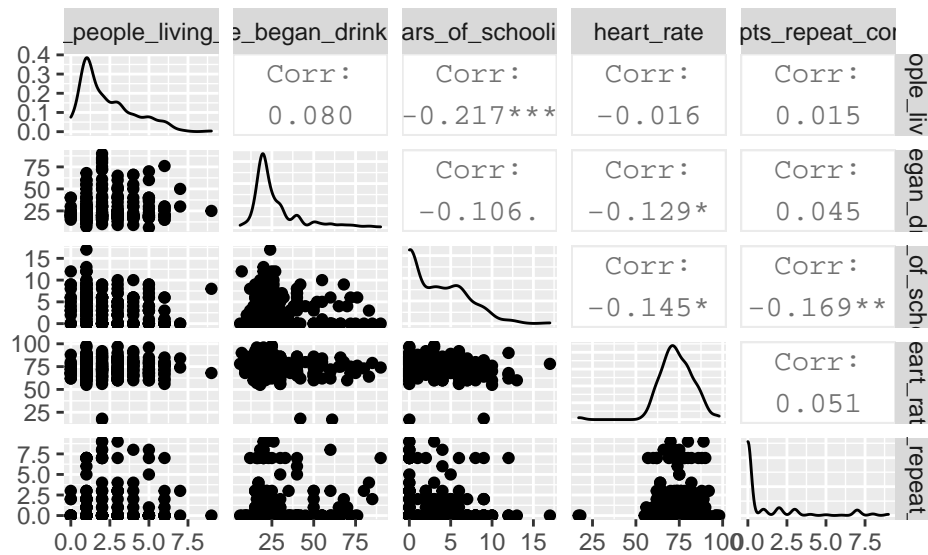
```
apply(data_cleaned, MARGIN = 2, FUN = var)
```

```
## num_people_living_with age_began_drinking
## 2.89183 223.47078
## years_of_schooling heart_rate
## 11.48509 97.39296
## attempts_repeat_correctly num_natural_teeth
## 4.67174 108.39369
## num_children age_started_excercise
## 4.19804 540.30984
## age_started_physlabor waist_circumfrence
## 70.03995 128.73062
## hours_sleep household_income
## 5.86936 454235939.32987
## hours_grandchildren_help_elder TRUEAGE
## 1700.43887 90.04163
## distance_from_hospitalkm inpatient_costs
## 4.13570 111882967.86808
```

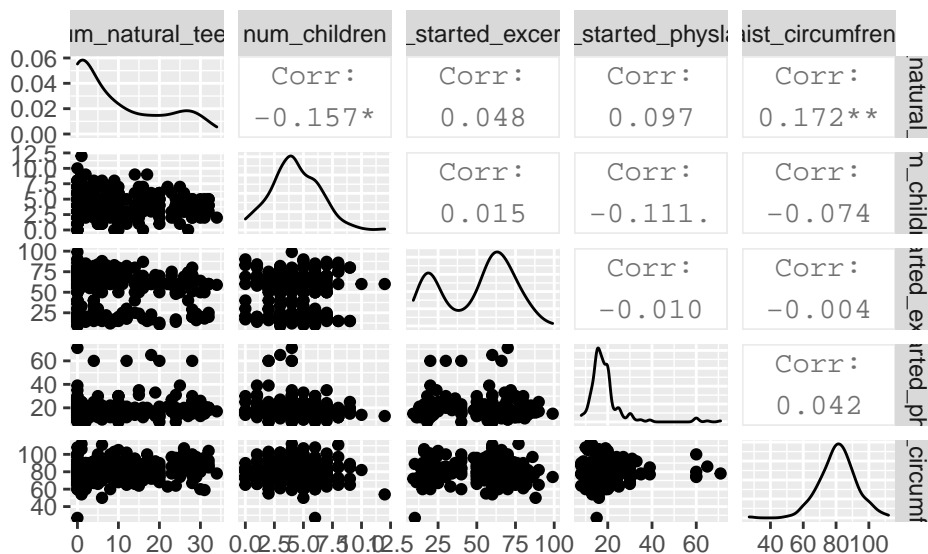
Looking at overview of the variables in the dataset, it seems that for the most part these values for the variables makes sense intuitively (values for number of teeth, heart rate, etc. make sense given the diverse sample we have but are also realistically valid). There are no missing values, which will be crucial in clustering and factor analysis. Many of the variables have high variance, which can be beneficial when identifying distinct clusters in cluster analysis.

## Bivariate Scatterplots and Correlations

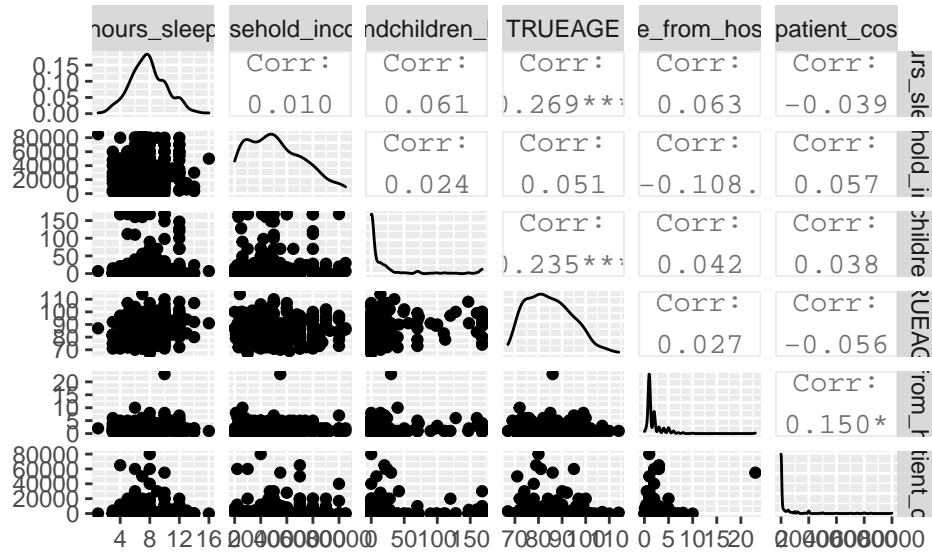
```
GGally::ggpairs(data_cleaned[, 1:5])
```



```
GGally::ggpairs(data_cleaned[, 6:10])
```



```
GGally::ggpairs(data_cleaned[, 11:16])
```



From the scatterplot matrix, it seems that the variables have a wide range of distribution types. It really depends on the context of the variables. For example, waist circumference and hours slept are more normally distributed because it is hard to have abnormal values for those variables. Household income has right skew which is a pattern we normally see with income. Number of children are counts that tend to be on the lower side, except for some more rare cases. All in all, the univariate analysis doesn't raise any major concerns for our future analysis.

However, in terms of bivariate analysis, there is more concern. There seems to be weak correlation between most of the variables, which will impact the effectiveness of factor analysis. The strongest relationships are the negative relationship between age and the number of teeth, as well as the positive correlation between number of people the elder is living with and their household income.

We will proceed with caution when doing factor analysis and anticipate that the factors produced may not be interpretable.

## Methods

### Clustering

Our first method that we will be using is clustering. The purpose of clustering is to identify distinct groups with the data; we use it in order find natural distinct groups of elders so we can zero in on different types of elders that may need more attention or assistance compared to others, and ultimately help us get a more detailed understanding of the sample. There are a variety of different ways to cluster, but we will employ two different types of clustering methods on our dataset. The first one is agglomerative hierarchical clustering, which starts with each elder in their own cluster, and fuses with others into larger clusters until there is only one large cluster left. Hierarchical clustering produces a dendrogram, a tree-like diagram, which allows the user to identify how many clusters to use for analysis. The “fusing” process can be based on a variety of different algorithms. In this analysis we will proceed with Ward’s method, which joins clusters of elders together based on smallest within cluster sum of squares values, with an aim to create the most homogeneous groups.

The second clustering method is k-means clustering. It begins with partitioning observations into a pre-specified number of clusters (unlike in hierarchical clustering), and observations are moved to different clusters until the WGSS (within-group sum of squares) is minimized. While it can oftentimes be a subjective decision to choose how many clusters in the final output, we will use the `clValid` package that determines the best number of clusters for each method based on internal measures such as decreasing connectivity and increasing the silhouette value. Based on that output, we will determine the number of clusters to use for k-means and where to “cut” the dendrogram for hierarchical clustering. For both of these methods, we will proceed with standardizing the variables to prevent any one variable from dominating any of our analysis.

To assess the strength and validity of the two methods we will interpret the silhouette coefficient along with comparing the identified clusters with a handful of variables to better understand how the algorithms classified the elders.

### Factor Analysis

Our second method is factor analysis, which aims to analyze the relationship between the variables in the dataset with a “latent” variable that does not exist in the dataset, but is related to them. In our analysis, we want to investigate the latent variable of the quality of life. We have variables that are related to the quality of life such as health and education measures, so we can use factor analysis to tie the two together. Factor analysis is very helpful in allowing us to study a concept that is not directly expressed in the data, but that is still important. These insights, along with clustering analysis, help us dive deeper into understanding the older population and specific trends and characteristics that may appear among them.

Because we don’t have any specific predictions or assumptions about our factors, we will conduct exploratory factor analysis. As mentioned previously, factor analysis requires variables to be correlated with one another; because the variables of the dataset seem to be weakly correlated we may see some issues with factors. Regardless, we will proceed with our analysis. Because our data does not follow multivariate normality due to the skewness in some of the univariate analysis, we will use the `fa` function within the `psych` package to find factors, and will be using the correlation matrix for ease of interpretation and standardization of variables. We will use hypothesis testing embedded in the `fa` function to determine the number of factors, while also considering keeping the number of factors low for easier interpretation. The type of rotation for factor analysis can be a subjective choice as well, but we will use varimax rotation, which is the most popular form that emphasizes a simple structure model for the factors.

## Results

### Clustering

```
compare <- clValid(scale(data_cleaned), 2:9,
  clMethods =
    c("hierarchical", "kmeans"), metric = "euclidean", method = "ward",
  validation = c("internal")
)
summary(compare)
```

```
##
## Clustering Methods:
## hierarchical kmeans
##
## Cluster sizes:
## 2 3 4 5 6 7 8 9
##
## Validation Measures:
```

		2	3	4	5	6	7	8	9
## hierarchical	Connectivity	131.753	172.065	190.850	210.354	212.971	213.780	240.429	261.570
##	Dunn	0.137	0.123	0.124	0.130	0.130	0.149	0.149	0.149
##	Silhouette	0.079	0.023	0.038	0.050	0.057	0.070	0.074	0.055
## kmeans	Connectivity	121.996	197.582	190.992	203.673	217.038	215.357	243.249	254.596
##	Dunn	0.085	0.133	0.133	0.133	0.133	0.146	0.146	0.149
##	Silhouette	0.105	0.067	0.082	0.090	0.095	0.096	0.098	0.091
##									
##	Optimal Scores:								
##									
##	Score	Method	Clusters						
##	Connectivity	121.996 kmeans	2						
##	Dunn	0.149 hierarchical	7						
##	Silhouette	0.105 kmeans	2						

First, we use the `clValid` package to determine the best number of clusters to use for k-means clustering and where the cut-off number of clusters is for hierarchical clustering. The output shows that 7 clusters is the most optimal for hierarchical clustering, and for k-means, 2 clusters is the optimal amount.

```
elder.dist <- dist(scale(data_cleaned))

hward <- hclust(elder.dist, method = "ward.D")

plot(hward)
```

## Cluster Dendrogram



elder.dist  
hclust (\*, "ward.D")

```
wardSol <- (cutree(hcward, k = 7))
summary(as.factor(wardSol))
```

```
## 1 2 3 4 5 6 7
## 84 76 54 17 26 6 5
```

```
wardsil <- silhouette(wardSol, elder.dist)
summary(wardsil)
```

```
## Silhouette of 268 units in 7 clusters from silhouette.default(x = wardSol, dist = elder.dist) :
## Cluster sizes and average silhouette widths:
##      84      76      54      17      26      6      5
## -0.0118218 0.0909485 0.0937653 0.1497067 0.0832441 0.3980656 0.1444004
## Individual silhouette widths:
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.14983 -0.00366  0.07167  0.07016  0.14460  0.48047
```

```
summary(wardsil)$avg.width
```

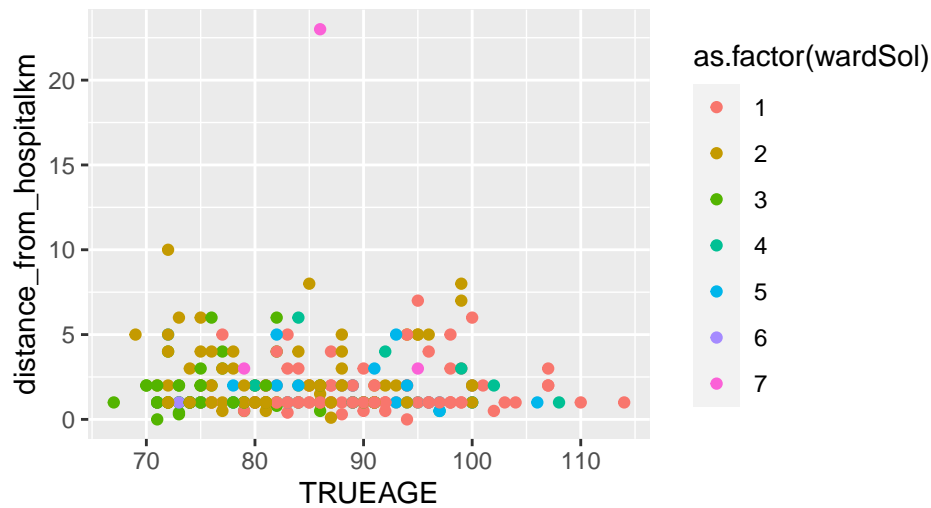
```
## [1] 0.0701572
```

After cutting the dendrogram output into 7 clusters, we can look at the average silhouette widths and plot some bivariate relationships between variables to identify what may have been some of the criteria for distinguishing groups of elders. The clusters vary wildly in size, with two of the clusters having less than 10 observations, and two of the clusters having over 70. The silhouette widths are also extremely low aside from one cluster, indicating that there is no significant structure in clustering. Cluster 1 has a negative silhouette width, which means some observations in the cluster would be better suited in another cluster.

```
ggplot(aes(x = TRUEAGE, y = distance_from_hospitalkm, colour = as.factor(wardSol)),
      data = data_cleaned) +
  geom_point() +
  ggtitle("Age vs Distance From Hospital ")
```

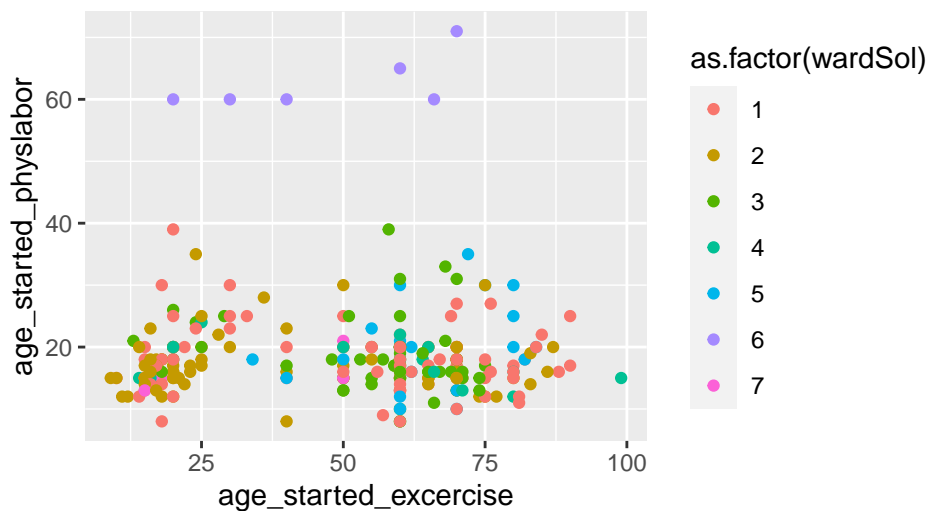


### Age vs Distance From Hospital



```
ggplot(aes(x = age_started_exercise, y = age_started_physlabor, colour = as.factor(wardSol)),
  data = data_cleaned) +
  geom_point() +
  ggtitle("Age Started Exercise vs Age Started Physical Labor")
```

### Age Started Exercise vs Age Started Physical Labor



In addition, plotting some of the variables in the dataset while also coloring by the clusters as indicated by the dendrogram produced no visible pattern or classification rules we can see. All of this suggests to us that the elders in our dataset were not able to be successfully classified using hierarchical clustering. We move on to k-means.

```
set.seed(2020)
Ksol <- kmeans(scale(data_cleaned), centers = 2)
Ksol$centers
```

```
## num_people_living_with age_began_drinking years_of_schooling heart_rate
```

```
## 1      0.513900      0.159679      -0.389346  0.0291394
## 2      -0.392187      -0.121860      0.297132 -0.0222380
##  attempts_repeat_correctly num_natural_teeth num_children
## 1      0.434382      -0.521157      0.301555
## 2      -0.331502      0.397725      -0.230134
##  age_started_exercise age_started_physlabor waist_circumfrence hours_sleep
## 1      0.0499732      -0.162936      -0.445335  0.268203
## 2      -0.0381375      0.124346      0.339861  -0.204681
##  household_income hours_grandchildren_help_elder TRUEAGE
## 1      0.238608      0.451320  0.774576
## 2      -0.182096      -0.344428 -0.591124
##  distance_from_hospitalkm inpatient_costs
## 1      -0.00488439      -0.117653
## 2      0.00372756      0.089788
```

```
kmeanssil <- silhouette(Ksol$cluster, elder.dist)
summary(kmeanssil)
```

```
## Silhouette of 268 units in 2 clusters from silhouette.default(x = Ksol$cluster, dist = elder.dist) :
## Cluster sizes and average silhouette widths:
##      116      152
## 0.0495023 0.1527343
## Individual silhouette widths:
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -0.0585  0.0479  0.1051  0.1081  0.1645  0.3002
```

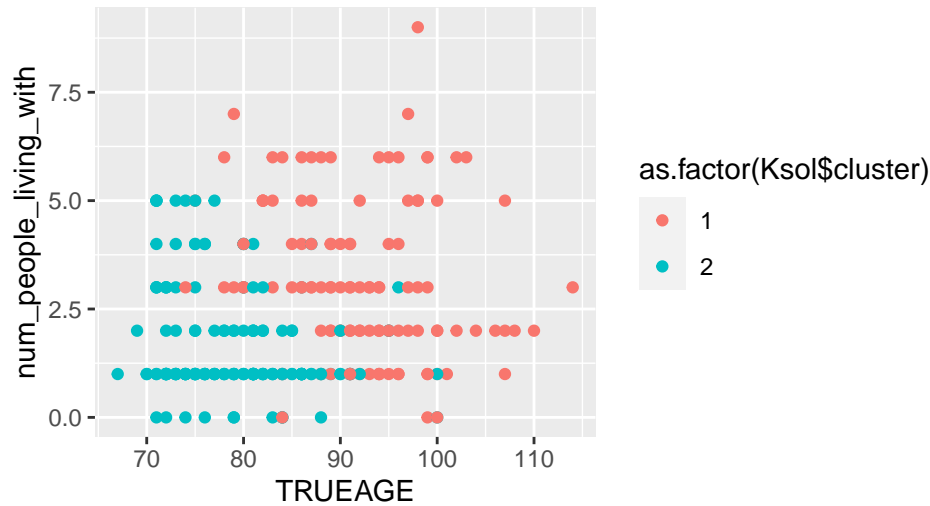
```
summary(kmeanssil)$avg.width
```

```
## [1] 0.108052
```

K-means clustering produced two clusters, each of relatively equal size. The silhouette values for both clusters are extremely low just like in hierarchical clustering, which suggests the invalidity of k-means. However, just to visualize how k-means clustered elders, we can plot some variables colored by their k-means cluster assignment.

```
ggplot(aes(x = TRUEAGE, y = num_people_living_with, colour = as.factor(Ksol$cluster)),
      data = data_cleaned) +
  geom_point() +
  ggtitle("Age VS Number of People Living with Elder")
```

### Age VS Number of People Living with Elder



```
ggplot(aes(x = waist_circumfrence, y = age_began_drinking, colour = as.factor(Ksol$cluster)),
  data = data_cleaned) +
  geom_point() +
  ggtitle("Age Began Drinking vs Waist Circumfrence")
```

### Age Began Drinking vs Waist Circumfrence



We can see a pretty distinct separation between younger elders versus older elders in the first plot, with the first cluster representing younger elders. This distinction is much more plausible compared to comparing health status of elders like in the second graph. Although we can hypothesize what the clustering rule was, we must recognize that, based on the low overall silhouette value of .11, k-means clustering was ultimately not an effective method either.

The lack of success in clustering may suggest several possibilities. We may not have a sufficient sample size, or enough variables, or perhaps elders in this dataset happened to share a lot of similarities that the clustering was not able to separate. More complicated and careful analysis may be needed to understand this complex group of individuals.

## Factor Analysis

```
myfa <- fa(cov(data_cleaned), nfactors = 5, rotate = "varimax")
myfa
```

```
## Factor Analysis using method = minres
## Call: fa(r = cov(data_cleaned), nfactors = 5, rotate = "varimax")
## Standardized loadings (pattern matrix) based upon correlation matrix
##
```

	MR1	MR3	MR2	MR4	MR5	h2	u2	com
## num_people_living_with	0.20	-0.01	0.70	0.00	0.03	0.537	0.463	1.2
## age_began_drinking	0.09	0.07	0.12	-0.15	-0.02	0.053	0.947	3.2
## years_of_schooling	-0.36	-0.08	-0.20	-0.19	0.01	0.215	0.785	2.2
## heart_rate	0.09	-0.06	-0.06	0.67	0.01	0.471	0.529	1.1
## attempts_repeat_correctly	0.17	0.97	-0.01	0.14	-0.07	0.997	0.003	1.1
## num_natural_teeth	-0.55	-0.11	0.02	0.10	0.10	0.334	0.666	1.2
## num_children	0.29	0.00	0.08	-0.01	0.06	0.096	0.904	1.2
## age_started_exercise	0.04	0.06	0.01	0.16	-0.04	0.033	0.967	1.5
## age_started_physlabor	-0.23	0.01	0.05	-0.12	-0.17	0.099	0.901	2.6
## waist_circumference	-0.30	-0.17	-0.09	-0.10	0.15	0.158	0.842	2.6
## hours_sleep	0.29	0.07	-0.05	0.09	0.03	0.104	0.896	1.4
## household_income	-0.05	0.02	0.50	-0.07	-0.01	0.260	0.740	1.1
## hours_grandchildren_help_elder	0.20	0.36	0.04	-0.05	0.07	0.183	0.817	1.7
## TRUEAGE	0.74	0.26	0.07	0.00	-0.05	0.620	0.380	1.3
## distance_from_hospitalkm	0.05	0.04	-0.15	-0.11	0.49	0.284	0.716	1.3
## inpatient_costs	-0.04	-0.02	0.11	0.01	0.35	0.135	0.865	1.3

```
##
##
```

	MR1	MR3	MR2	MR4	MR5
## SS loadings	1.43	1.21	0.87	0.62	0.45
## Proportion Var	0.09	0.08	0.05	0.04	0.03
## Cumulative Var	0.09	0.16	0.22	0.26	0.29
## Proportion Explained	0.31	0.26	0.19	0.14	0.10
## Cumulative Proportion	0.31	0.58	0.77	0.90	1.00

```
##
## Mean item complexity = 1.6
## Test of the hypothesis that 5 factors are sufficient.
##
## The degrees of freedom for the null model are 120 and the objective function was 1.48
## The degrees of freedom for the model are 50 and the objective function was 0.15
##
## The root mean square of the residuals (RMSR) is 0.03
## The df corrected root mean square of the residuals is 0.04
##
## Fit based upon off diagonal values = 0.95
## Measures of factor score adequacy
##
```

	MR1	MR3	MR2	MR4	MR5
## Correlation of (regression) scores with factors	0.83	0.99	0.77	0.72	0.60
## Multiple R square of scores with factors	0.69	0.97	0.60	0.52	0.36
## Minimum correlation of possible factor scores	0.38	0.95	0.20	0.04	-0.27

The output for factor analysis is shown above. While the maximum number of factors we can use where degrees of freedom is still greater than 0 is 10 factors, this makes interpreting extremely hard. Because of this, we use 5 factors in our analysis, even though we may have been able to use more. We can see that the

factors seem to follow a weak simple structure for the most part. A handful of variables are loaded the most for only 1 factor, and the factors seem different enough.

Even with 5 factors, interpreting them is quite difficult. We could say that the first factor simply represents longevity, as both age and number of natural teeth (which is a negative loading) are large components of this factor. The second factor is mainly a measure of mental agility, as `attempts_repeat_correctly` represents how many times it took an elder to repeat a set of words correctly. The third factor represents family support, as both household income and number of people the elder is living with have the greatest loadings. The fourth factor is a health measure, as heart rate is the largest loading for this factor. The fifth factor is related to health care costs, with both distance from nearest hospital and inpatient costs being the largest components of this factor.

```
factor.stats(f = myfa)
```

```
## Call: fa.stats(r = r, f = f, phi = phi, n.obs = n.obs, np.obs = np.obs,
##      alpha = alpha, fm = fm)
##
## Test of the hypothesis that 5 factors are sufficient.
##
## The degrees of freedom for the model is 50 and the fit was 0.15
##
## Measures of factor score adequacy
## Correlation of scores with factors      0.83 0.99 0.77 0.72 0.6
## Multiple R square of scores with factors 0.69 0.97 0.6 0.52 0.36
## Minimum correlation of factor score estimates 0.38 0.95 0.2 0.04 -0.27
```

Regardless of how these factors are interpreted, it is important to recognize that the fit of the model is not ideal, so realistically it is not a valid analysis. Goodness of fit measures indicate that the fit was .15, which is very poor. In addition, the variation that each factor explains is also incredibly small, with their combined variation equaling .29. Additional research is needed to investigate other goodness of fit measures in the `psych` package that can help us better understand our factor analysis results.

Even though it is not statistically insightful, the interpretation of the factors does suggest that our original claim that elders are multi-faceted is relatively supported; an assessment of the quality of life cannot just be their age alone, but rather, a multitude of factors must be considered. With only 16 variables in our dataset, we were still able to define 5 distinct components that represent issues and characteristics that define elders.

## Conclusion

Our original goal was to determine whether or not we could better understand the aging population as a whole, and we utilized two methods, clustering and factor analysis, in hopes of differentiating elders and finding underlying components of quality of living. Our clustering attempt was not very successful, as the clusters produced did not seem to strongly separate elders into any distinct categories. Since there were only 268 elders and 16 variables, it may be challenging to identify extremely distinct groups, especially since variables such as hours of sleep and heart rate can't realistically vary too much. There are several hundred other factor variables in this dataset that could have a lot of potential if used, and can be considered for future analysis. On the other hand, factor analysis produced 5 components to measuring the quality of living that gives us a more in-depth view of an elder's life. Healthcare access, longevity, health status, family support, and mental agility are all important pillars in aging that can be used to characterize elders. Although we weren't able to categorize elders, we were able to reduce the dimension of our dataset down into a few groups of variables that paint an insightful picture about their lives.

Our analysis was exploratory in nature and cannot tell us definitively which elders are living "better" lives vs those that are not. As a whole it mainly shows us that in policy, law, and media, there is needs to be more nuance when describing the population. Because this study was conducted in China, where culturally there is more respect for the older population compared to Western countries as well as different socioeconomic conditions, we can expect the data to look very different. Thus, we cannot generalize this analysis to elders across the world. In future analysis it could be valuable to look at and aggregate elders across different countries to understand the aging process universally. Despite the limitations, this analysis is a first step in understanding a growing population and working towards bridging gaps in our community.

## Citations

Zeng, Yi, Vaupel, James, Xiao, Zhenyu, Liu, Yuzhi, and Zhang, Chunyuan. Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998-2014. Inter-university Consortium for Political and Social Research [distributor], 2017-04-11. <https://doi.org/10.3886/ICPSR36692.v1>