2011 US World Value Survey Data Analysis Report

Author of the overall report
Yuanyuan Lin (Lynn)
Oct,2019

A. Abstract

World Value Survey is a common questionnaire survey which aims to study the political attitudes and values of individuals

B. Introduction

This report gives an overview of the quantitative data analysis of 244 survey questions that were conducted by a network of social scientists at Michigan University in 2011. The purpose of this report is to know how United States whole society affects and changes participants' attitude, values and belief on economy, education, work, family, politics, culture, national identity, environment, gender equality, religion, security, impact of globalization, wiliness to join different organizations, tendency to immigrate, characteristics of democracy, morality, happiness toward life, social network, science/technology development.

1. Data Source

World Values Survey website contains questionnaire, codebook, statistical data files on the Data&Documentation section in almost 100 countries. I access the 2011 United States raw data on Wave 6 and obtain the data in Csv file. The survey was conducted from 06.09.2011-07.05.2011. There are in total 70.86% cooperation rate with 2232 cases completed out of 3150 cases.

C. Methodology

Small protest of the World Values Survey has verified the data accuracy and time of the survey, dataset itself has no mistakes but some redundant questions with similar information. To answer this question, I analyzed what participants have answered on World Value Survey questions, used exploratory data analysis and exploratory factor analysis to explores people's values and beliefs in the United States.

1. EDA

Several exploratory data analysis (EDA) plots for people's point of view on education, work, religion, gender value is in discussion section.

2. EFA

Exploratory Factor Analysis (EFA) is used to. measure, monitors and analyzes World Value Survey Dataset in the following major components: Tolerance of ethnic minorities and

foreigners, attitudes toward economy, education, work, family, politics, culture, national identity, environment, gender equality, religion, security, impact of globalization, wiliness to join different organizations, tendency to immigrate, characteristics of democracy, morality, happiness toward life, social network, science/technology development.

Discussion

After doing data cleaning including removing NA values and deleting columns that have questions have not been asked, the whole dataset contains 244 variables. Then, Exploratory Factor Analysis (EFA) has been used so that 244 variables were condensed into only twenty major factors that are valuable to analyze.
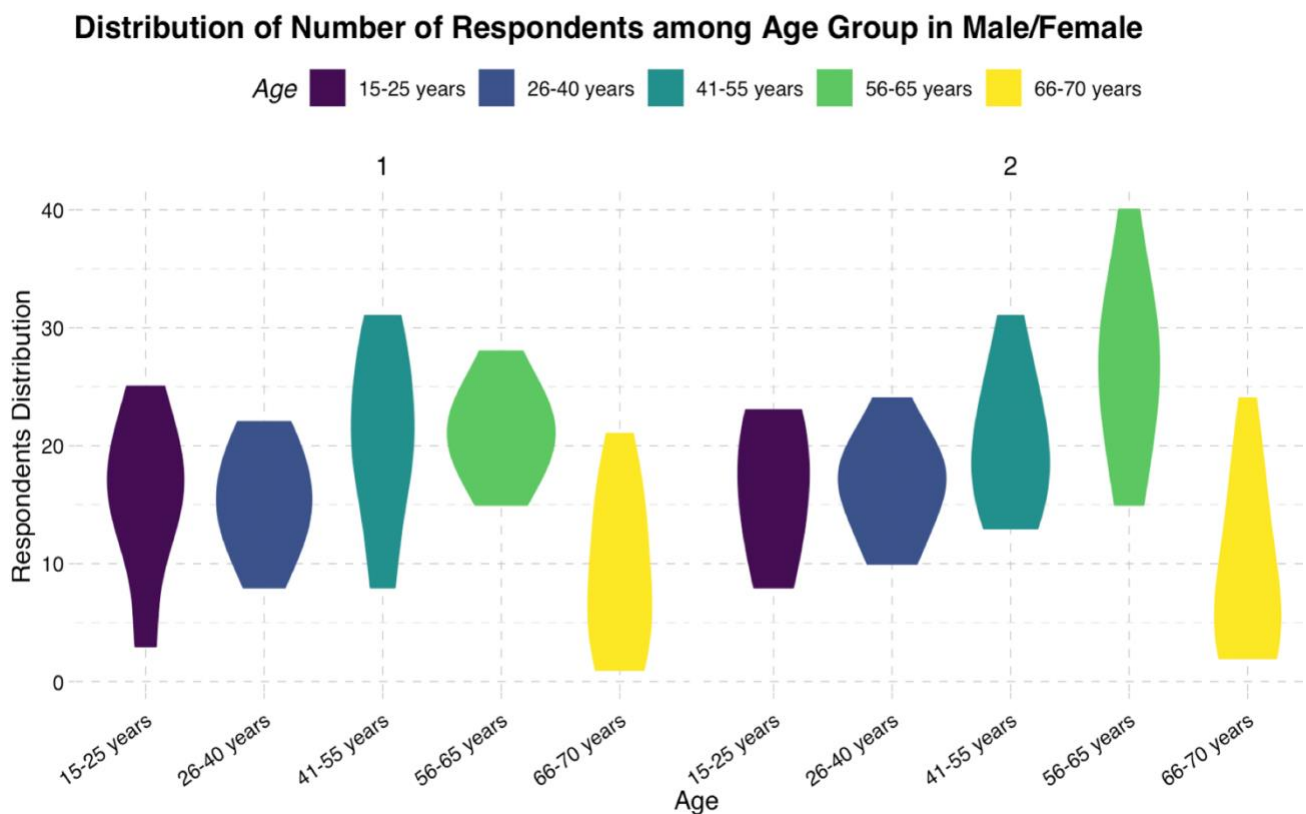
1. Gender values

Table 1.1

The violin plot shows the number of respondents in five age groups. Number one stands for male and the other is female. On Table 1.1, the number of male participants aged between 56-65 years is relatively large than female in the same age group. Both male/female participants who are 66-70 years are in a small amount with about twenty at most. Most of respondents are age in 26-65 years. Perhaps not surprisingly, Viewpoint from middle-age people mostly
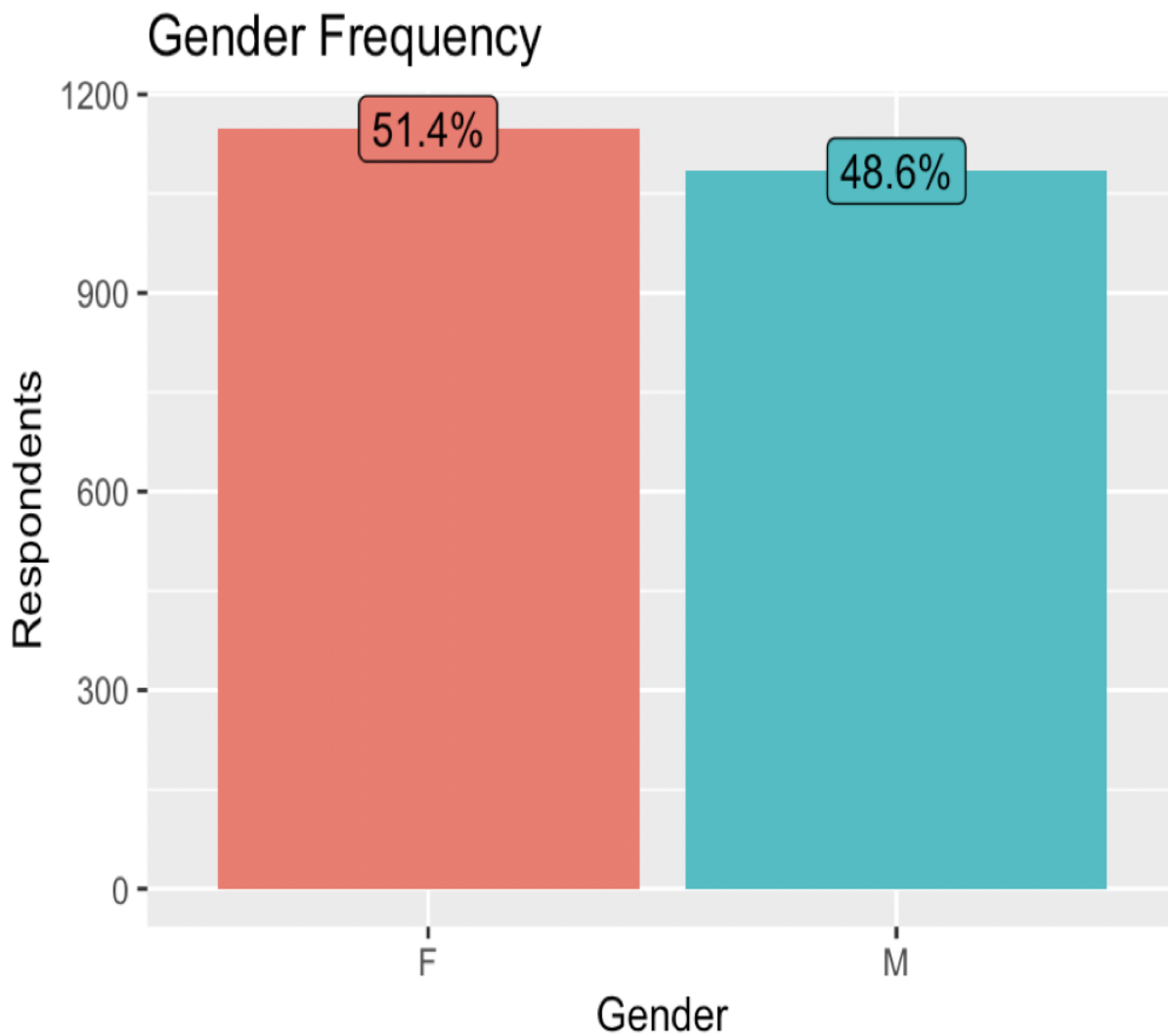
could be representative.

Table 1.2

Gender frequency of the number of respondents is shown in Table 1.2. About 51.4% of the total participants are female and the rest of 48.6% are male.

3. Religion

Whether You are a Religious Person or Not

Table 2.1

This violin plot shows the distribution of 2232 US citizens religion belief regarding on whether they think they are religious person or not. On the Y-axis in this chart, the unit of 1,2 and 3 stand for religious, not religious and atheist. Note that on Table 2.1, the age range with the largest number of people who are religious people are between 66 to 70 years whereas 15-25 years has the smallest amount of religious people. Apart from age group in 15-25 years and 66-70 years, the largest amount of people who do not belief in any religion are in 41-55 years. The result of non-religious belief is almost the same in 26-40 years and 56-65 years.

4. Education

On Table 3.2 explains the x-axis in the Table 3.1. Number one stands for no formal education and number ten stands for university-level education, with degree. As the unit number increases, highest education level increases. Together, two gender plots both show the general pattern of large amount of complete secondary: university-preparatory type and some university-level education, without degree.
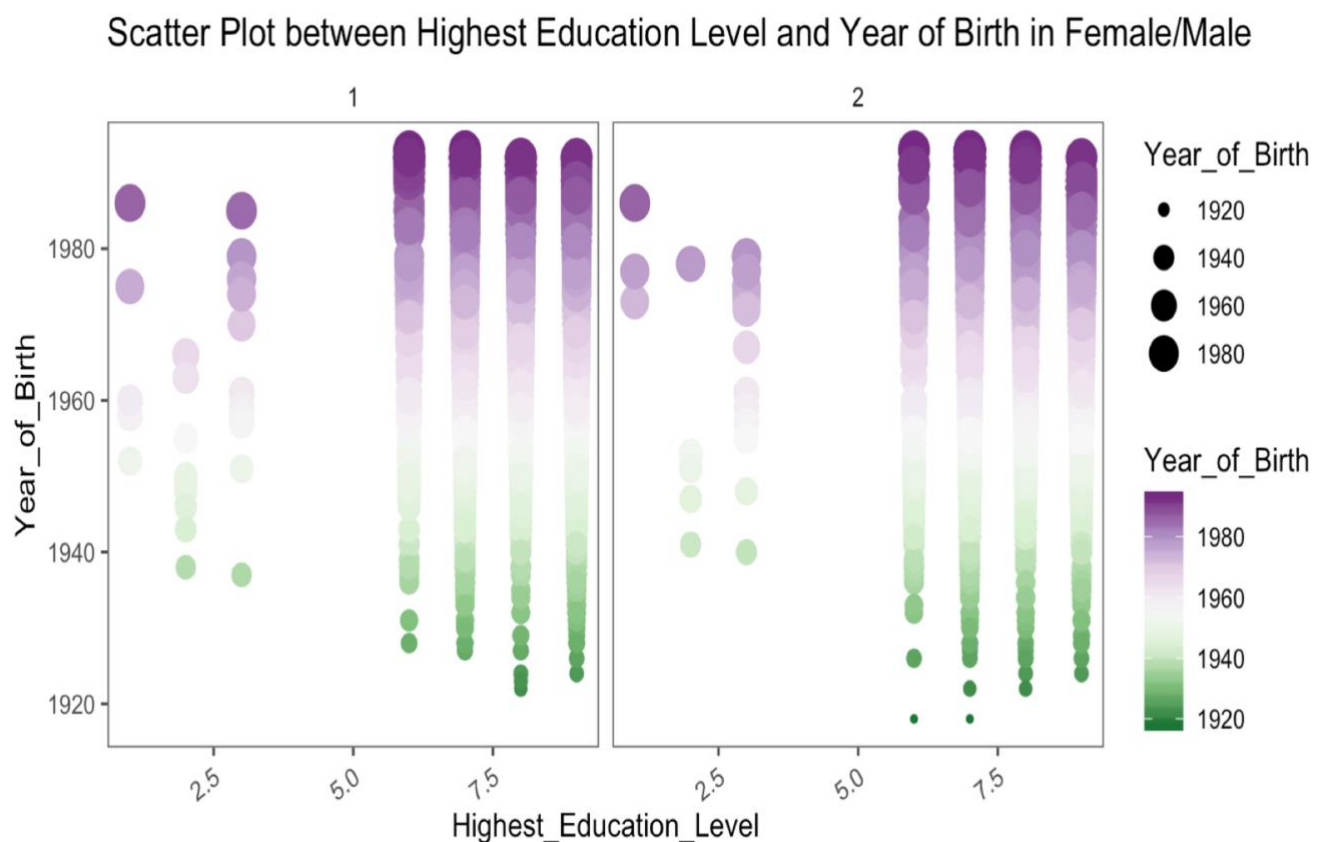

Table 3.1

1 No formal education
2 Incomplete primary school
3 Complete primary school
4 Incomplete secondary school: technical/vocational type
5 Complete secondary school: technical/vocational type
6 Incomplete secondary: university-preparatory type
7 Complete secondary: university-preparatory type
8 Some university-level education, without degree
9 University-level education, with degree
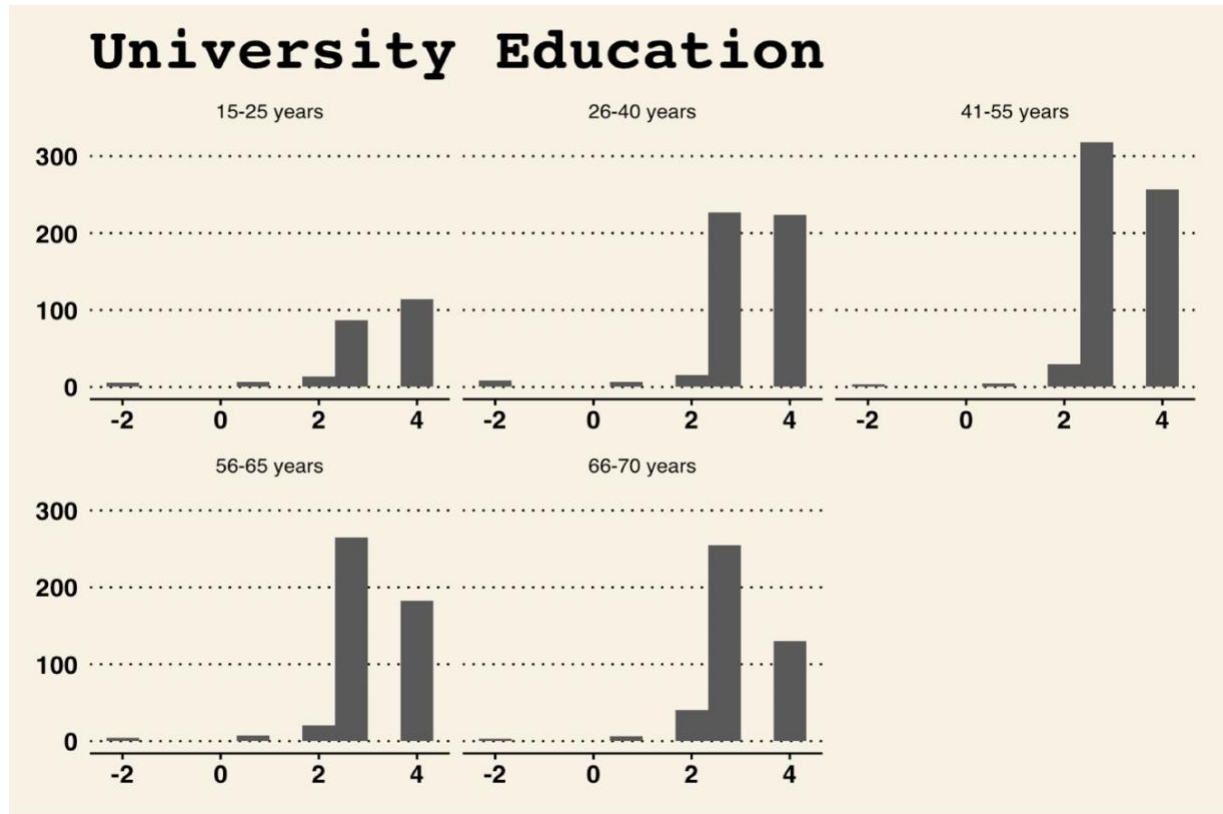
Table 3.2

University Education

Table 3.3

This particular grouped bar graph shows what participants think whether university education is more important for a boy than a girl. There are five categories regarding on the agreement of boys are more important to take university education. Negative two stands for no response from the respondents. The correct answer for agreeing and strongly agree that university education is more important for boys than girls are 1 and 2. Obviously, there are only small percent of participants think that it is not as important for girls to take university education as boys. It appears that the largest number of voting are among 41-55 years with vote for number three and four. This makes sense, because the middle-aged people tend to focus more on importance of gender equality than other groups.

5. Happiness and life satisfaction
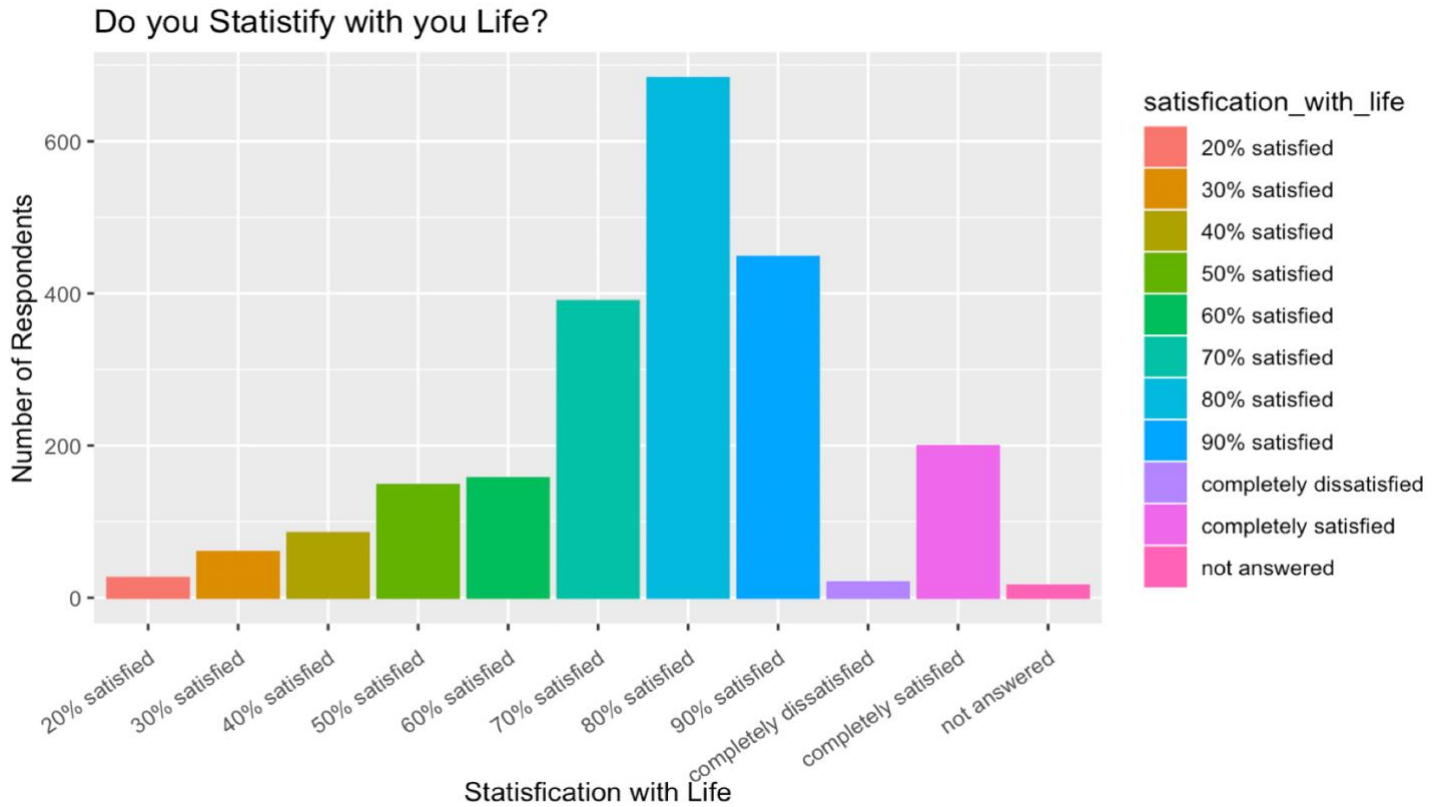
Do you Statistify with you Life?

Table 4.1

The chart gives information about the satisfaction with life among 2232 respondents. The tallest vertical bar above (around 680) stands for the number of respondents who are 80% satisfied with their life. Overall, it is perhaps unsurprising that most U.S citizens are around 70-90 percent satisfied with their life, but there are still around two hundred participants completely satisfied with life which may reflect the a good well-being of living condition.

6.Work

Age and Employment Status by Gender

Table 5.1

The pattern of two genders (number one is male, number two is female) are noticeable different. Overall, the scatter plots appear to be especially informative, it tends to confirm that most female are doing manual tasks no matter what age they are. Only a small amount of female is doing the type of work in between mostly manual work and mostly intellectual work. Unfortunately, there are no female respondents who characterize their major work as the most intellectual.

6. Exploratory Factor Analysis (EFA)

Exploratory Factor Analysis (EFA) is used in 2011 US World Value Survey data analysis report. Due to large number of variables with multicollinearity relationship, it is essential to reduce data to a smaller set of summary variables. EFA helps explore the underlying structure and similarity among 244 variables. From Parallel Analysis Screen Plot in Table 6.1, optimized factor number is around twenty which is the intersection between FA Actual Data and FA Resampled Data. After arrived at a probable number of factors, it is time to perform factor analysis at number twenty. Output showing factors and loadings are in Table 6.2. Note that there are MR1-MR20 numbers for each column variable. By looking at the largest result in each row, 244 variables are grouped into twenty categories. Then, validation of factor analysis is used in current stage. The root mean square of residuals (RMSR) is 0.02 which is pretty closer to 0. RMSEA (root mean square error of approximation) index is 0.042. As it is below 0.05, it is a good model fit based on RMSEA. Only Tucker-Lewis Index (TLI) is slightly a little smaller than the normal number but the other two conditions are satisfied. New twenty factors are formed depending on the variable loading. On Table 6.4 and Table 6.5 are named- factors after establishing the adequacy of the factors.
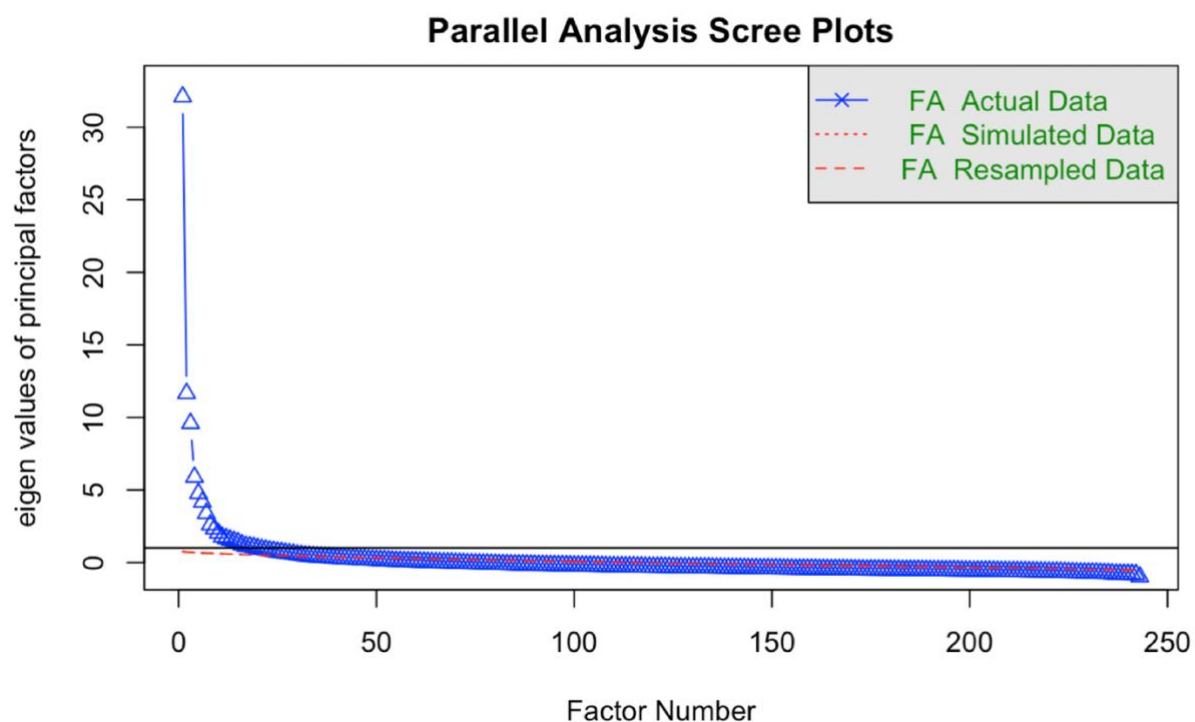

Table 6.1

	MR1	MR3	MR2	MR15	MR14	MR5
	:S3: AsIs>	<S3: AsIs>	<S3: AsIs>	<S3: AsIs>	<S3: AsIs>	<S3: AsIs>
V4..Important.in.life..Family	-0.13	0.19	-0.01	-0.03	0.04	0.10
V5..Important.in.life..Friends	-0.13	0.09	0.08	0.00	0.07	0.03
V6..Important.in.life..Leisure.time	-0.07	-0.02	0.03	-0.03	0.01	0.08
V7..Important.in.life..Politics	0.00	0.08	-0.02	-0.02	-0.18	0.26
V8..Important.in.life..Work	-0.02	0.13	-0.08	-0.04	0.05	0.08
V9..Important.in.life..Religion	-0.03	0.82	-0.03	0.01	0.02	0.03
V10..Feeling.of.happiness	-0.09	0.09	-0.02	-0.01	0.01	0.03
V11..State.of.health..subjective.	-0.02	0.06	0.02	-0.04	-0.07	0.02
V12..Important.child.qualities..independence	-0.01	-0.22	-0.07	-0.06	0.08	0.00
V13..Important.child.qualities..Hard.work	-0.04	-0.05	-0.03	-0.13	0.07	0.02
V14..Important.child.qualities..Feeling.of.responsibility	0.00	-0.12	-0.04	-0.02	-0.06	-0.06
V15..Important.child.qualities..Imagination	-0.03	-0.12	-0.01	0.00	-0.08	0.02
V16..Important.child.qualities..Tolerance.and.respect.for.other.people	0.10	-0.05	0.11	0.02	0.06	-0.03
V17..Important.child.qualities..Thrift.saving.money.and.things	0.00	-0.10	0.02	-0.01	-0.01	0.04
V18..Important.child.qualities..Determination..perseverance	-0.01	-0.09	0.05	-0.07	-0.12	0.03
V19..Important.child.qualities..Religious.faith	-0.05	0.69	0.02	0.00	-0.04	0.02
V20..Important.child.qualities..Unselfishness	-0.01	0.06	0.13	0.02	-0.04	0.00
V21..Important.child.qualities..Obedience	-0.04	0.17	0.07	0.06	0.06	-0.04
V22..Important.child.qualities..Self.expression	-0.01	-0.15	0.02	0.08	0.01	-0.03

Table 6.2

```
The degrees of freedom for the null model are  29403  and the objective function was  150.45 with
Chi Square of  323349.2
The degrees of freedom for the model are 24733  and the objective function was  54.14

The root mean square of the residuals (RMSR) is  0.02
The df corrected root mean square of the residuals is  0.02

The harmonic number of observations is  2232 with the empirical chi square  57859.76  with prob <
0
The total number of observations was  2232  with Likelihood Chi Square =  115639.7  with prob <
0

Tucker Lewis Index of factoring reliability =  0.63
RMSEA index =  0.042  and the 90 % confidence intervals are  0.04 NA
BIC =  -75067.86
Fit based upon off diagonal values = 0.98
Measures of factor score adequacy
                                                 MR1  MR3  MR2 MR15 MR14  MR5  MR4  MR9  MR7
MR18  MR6 MR19 MR12 MR17 MR16  MR8 MR20 MR10 MR13 MR11
Correlation of (regression) scores with factors  0.97 0.97 0.97 0.96 0.94 0.94 0.93 0.97 0.92
0.93 0.92 0.92 0.92 0.92 0.92 0.91 0.92 0.90 0.92 0.89
Multiple R square of scores with factors         0.94 0.94 0.94 0.92 0.89 0.89 0.87 0.94 0.86
0.86 0.85 0.85 0.84 0.85 0.85 0.83 0.84 0.81 0.85 0.80
Minimum correlation of possible factor scores     0.88 0.88 0.88 0.84 0.78 0.77 0.75 0.89 0.71
0.73 0.71 0.71 0.69 0.70 0.71 0.67 0.68 0.62 0.69 0.59
```

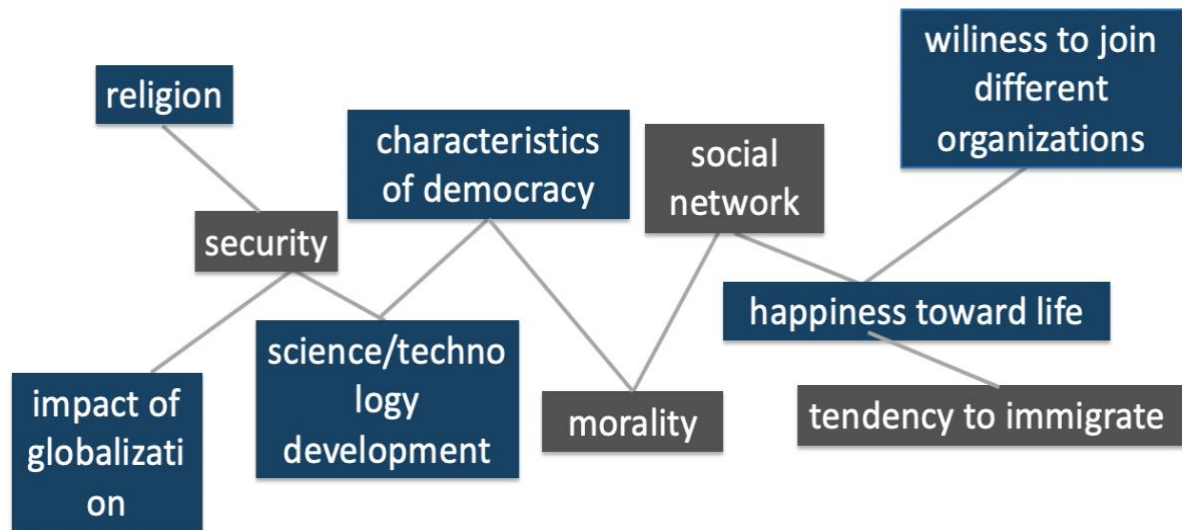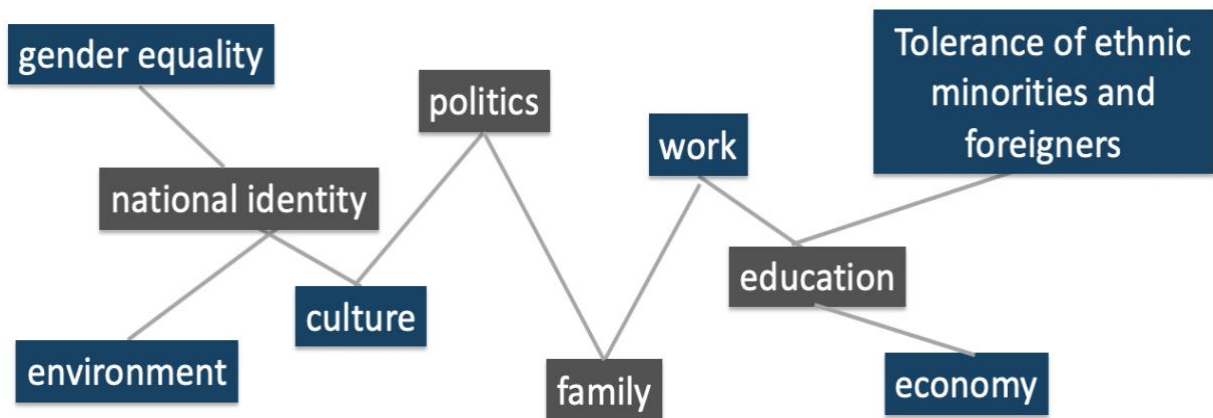## Table 6.3

Table 6.4

Table 6.5

Results

Individual Values certainly differ among gender, education background, income and ethnicity. One of reasons that U.S citizens answered the questionnaires in certain ways is because of the existence of emancipative values. Emancipative values in one of the major cultural components of human empowerment in the United States. Emancipative values help increase people's capabilities, aspirations and entitlements to exercise freedom. Generally, there are four categories of defining personal value and belief, including traditional values, secular-rational values, survival values and self-expression values. Firstly, traditional values have preferences on religion, parent-child ties, deference to traditional family values and rejection to suicide, abortion and divorce. US society places emphasis on traditional values which have a high level of national pride. The opposite preferences to traditional values are called secular-rational values. As existential security in US increases, the society cultivates more on traditional value. Additionally, there are two opposite values which belong to level of trust and tolerance. American has strong emphasis on self-expression values. The self-expression value emphasis on high tolerance of gender equality, foreigners, homosexuality, high priority on decision-making participation in political and economic life and environment protection. On the contrary, survival value focus on low acceptance of outsiders with weak sense of happiness. Since US society policy makers always try to build civil society with rising individual agency, the society moves focus from survival value to self-expression value. Also, there are a few limitations with respect to the data analysis of World Value Survey. On the basis of this data analysis report, the analysis does not extend to other waves of survey. Moreover, the EDA and EFA sections do not seek to trace back to the older time for further US survey investigations. Discussion in this report would be more interesting if do further research and data analysis cross the countries.

Conclusion

The exploratory data analysis and exploratory factorial analysis on 2011 United States World Value Survey help demonstrate people's points of view in economy, education, work, family, politics, culture, national identity, environment, gender equality, religion, security etc. It is essential to understand and appreciate their distinctive point of views. Culture zone pattern in the world based on value difference. Value difference include different senses of existential security and individual agency. For example, Protestant societies of Northern Europe has strong emphasis on the secular-rational values while Islamic societies of the Middle East focus on traditional values. Further data analysis on other countries surveys responses would help to test whether nowadays technological and economic changes would change people's value and belief and help better understand the motivations for several particular upheaval all over the world, such as French civil unrest, Yugoslav wars., etc.
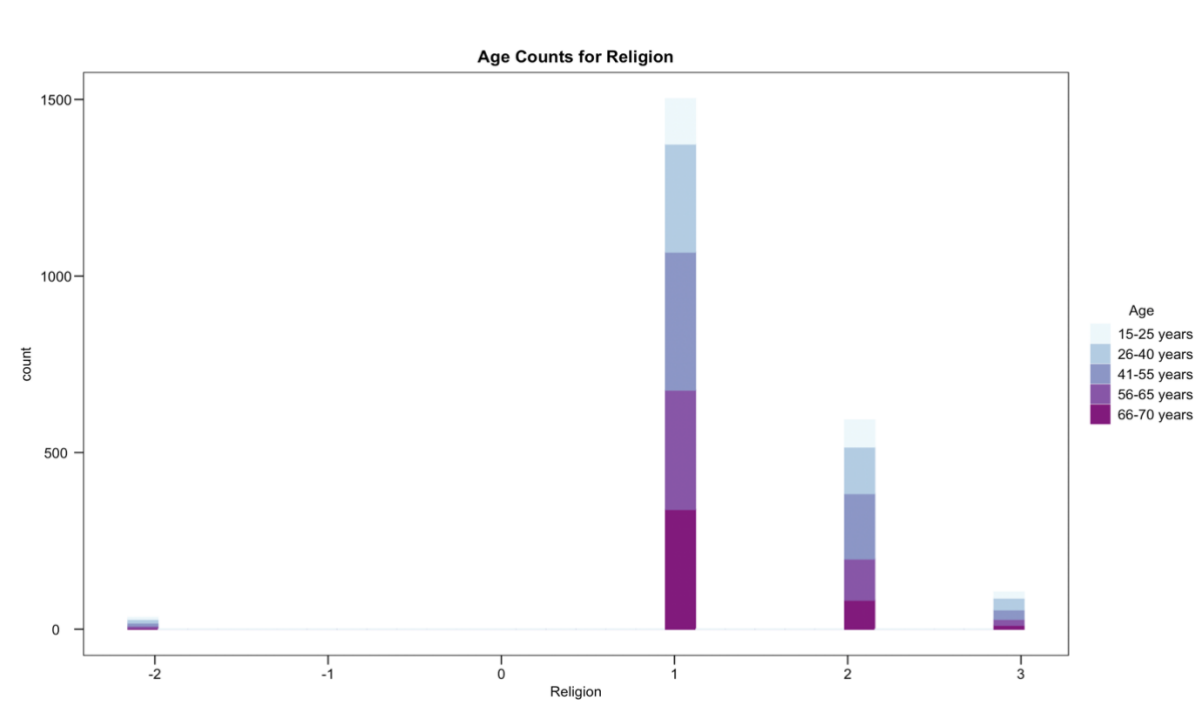
Reference

Alesina, A; Giuliano, P; Nunn, N (2010), The Origins of Gender roles — Women and the Plough, World Values Survey.

Alexander, A; Welzel, C (2010), "Empowering Women: The Role of Emancipative Values", European Sociological Review, 27 (3).

Appendix

More EDA

Code

```
Survey<-read.csv("Survey.csv")
Survey<-na.omit(Survey)
#head(Survey)
Survey1<-Survey[,(1:243)]
show<-Survey %>%
   filter(!is.na(Survey$V4..Important.in.life..Family)) %>%
   count(Survey$V4..Important.in.life..Family)

show[1,1]='not answered'
show[2,1]='very important'
show[3,1]='rather important'
```

```r
show[4,1]='not very important'
show[5,1]='not at all important'
show<-as.data.frame(show)
show
show1<-Survey %>%
    filter(!is.na(Survey$V23..Satisfaction.with.your.life)) %>%
    count(Survey$V23..Satisfaction.with.your.life)

show1[1,1]='not answered'
show1[2,1]='completely dissatisfied'
show1[3,1]='20% satisfied'
show1[4,1]='30% satisfied'
show1[5,1]='40% satisfied'
show1[6,1]='50% satisfied'
show1[7,1]='60% satisfied'
show1[8,1]='70% satisfied'
show1[9,1]='80% satisfied'
show1[10,1]='90% satisfied'
show1[11,1]='completely satisfied'
show1<-as.data.frame(show1)
show1<- rename(show1, satisfication_with_life = `Survey$V23..Satisfaction.with.your.life`)
show1
ggplot(show1) +
  aes(x = satisfication_with_life, fill = satisfication_with_life,
        colour = satisfication_with_life, weight = n) +
  geom_bar() +
  scale_fill_hue() +
  scale_color_hue() +
  theme_gray()+
  theme( axis.text.x    = element_text(angle=35, hjust=1, vjust=0.9))+
  labs(x = "Statisfication with Life", y = "Number of Respondents") +
    ggtitle("Do you Statistify with you Life?")
ggplot(show1) +
  aes(x = satisfication_with_life, y = n, fill = satisfication_with_life, colour =
satisfication_with_life) +
  geom_boxplot() +
  scale_fill_hue() +
  scale_color_hue() +
  theme_gray()+
  theme( axis.text.x    = element_text(angle=35, hjust=1, vjust=0.9))+
  labs(x = "Statisfication with Life", y = "Number of Respondents") +
    ggtitle("Box?")
show2<-Survey %>%
    filter(!is.na(Survey$V240..Sex) & !is.na(Survey$V242..Age)) %>%
```

```r
  group_by(Survey$V240..Sex, Survey$V242..Age) %>%
  count()
show2$`Survey$V242..Age` <- cut(show2$`Survey$V242..Age`,
                     breaks = c(-Inf, 25, 40, 55, 65,Inf),
                     labels = c("15-25 years", "26-40 years", "41-55 years", "56-65
years", "66-70 years"),
                     right = FALSE)
show2<-as.data.frame(show2)

#show2[show2 =='1']<-"Male"
#show2[show2 =='2']<-"Female"
#show2
show2<- rename(show2, Age =`Survey$V242..Age`)
#show2
ggplot(show2) +
  aes(x = Age, y = n, fill = Age, colour = Age) +
  geom_violin(adjust = 1.6, scale = "area") +
  scale_fill_viridis_d(option = "viridis") +
  scale_color_viridis_d(option = "viridis") +
  theme_pander() +
  theme(legend.position = "top") +
  facet_wrap(vars(`Survey$V240..Sex`))+
  theme( axis.text.x    = element_text(angle=35, hjust=1, vjust=0.9))+
  labs(x = "Age", y = " Respondents Distribution") +
   ggtitle("Distribution of Number of Respondents among Age Group in Male/Female")
Survey<- rename(Survey, Highest_Education_Level =
V248..Highest.educational.level.attained)
Survey<- rename(Survey, Year_of_Birth = V241..Year.of.birth)
knitr::opts_chunk$set(echo = FALSE, fig.height=1.5, fig.width=8)
ggplot(Survey) +
  aes(x = Highest_Education_Level, y = Year_of_Birth, colour = Year_of_Birth, size =
Year_of_Birth) +
  geom_point() +
  scale_color_distiller(palette = "PRGn") +
  theme_few() +
  facet_wrap(vars(V240..Sex))+
theme( axis.text.x    = element_text(angle=35, hjust=1, vjust=0.9))+
   ggtitle("Scatter Plot between Highest Education Level and Year of Birth in Female/Male")
show4<-Survey[,(9:19)]
show4<- rename(show4, Independence = V12..Important.child.qualities..independence)
show4<- rename(show4, Hard_Work = V13..Important.child.qualities..Hard.work)
show4<- rename(show4, Responsibility =
V14..Important.child.qualities..Feeling.of.responsibility)
show4<- rename(show4, Imagination = V15..Important.child.qualities..Imagination)
```

```r
show4<- rename(show4, Tolerance =
V16..Important.child.qualities..Tolerance.and.respect.for.other.people)
show4<- rename(show4, Thrift =
V17..Important.child.qualities..Thrift.saving.money.and.things)
show4<- rename(show4, Determination =
V18..Important.child.qualities..Determination..perseverance)
show4<- rename(show4, Religious_faith = V19..Important.child.qualities..Religious.faith)
show4<- rename(show4,Unselfishness = V20..Important.child.qualities..Unselfishness)
show4<- rename(show4, Obedience = V21..Important.child.qualities..Obedience)
show4<- rename(show4, Self_expression = V22..Important.child.qualities..Self.expression)
#show4
condition <- c("Independence" , "Hard_Work" ,
"Responsibility","Imagination","Tolerance","Thrift","Determination","Religious_faith","Uns
elfishness","Obedience","Self_expression")
y1<-count(show4,Independence)
y2<-count(show4,Hard_Work)
y2
y<-data.frame(cbind(y1,y2))
y
Education<-Survey$V52..A.university.education.is.more.important.for.a.boy.than.for.a.girl
Age<-Survey$V242..Age

show5_new<-data.frame(cbind(Education,Age))

show5_new$Age <- cut(show5_new$Age,
                      breaks = c(-Inf, 25, 40, 55, 65,Inf),
                      labels = c("15-25 years", "26-40 years", "41-55 years", "56-65
years", "66-70 years"),
                      right = FALSE)
#show5_new
ggplot(show5_new) +
  aes(x = Education, fill = Education, colour = Education) +
  geom_histogram(bins = 10L) +
  scale_fill_distiller(palette = "Set3") +
  scale_color_distiller(palette = "Set3") +
  theme_wsj() +
  facet_wrap(vars(Age), scales = "free_x")+
  labs(x = "Age in Range", y = " Number of Respondents ") +
  ggtitle("University Education")
multi = Survey
Survey$is_ds = as.character(ifelse(Survey$V240..Sex==1,"M","F"))
foo <- Survey %>%
  filter(!is.na(is_ds)) %>%
  group_by(is_ds) %>%
```

```r
  count()

global_ds <- foo %>%
  filter(is_ds == "M") %>%
  .$n / sum(foo$n) * 100


p1 <- foo %>%
  mutate(percentage = str_c(as.character(round(n/sum(foo$n)*100,1)), "%")) %>%
  ggplot(aes(is_ds, n, fill = is_ds)) +
  geom_col() +
  geom_label(aes(label = percentage), position = position_dodge(width = 1))+
  labs(x = "Gender", y = "Respondents") +
  theme(legend.position = "none") +
  ggtitle("Gender Frequency")
p1
exp_code = Survey$V242..Age
expml = as.character(Survey$V229..Employment.status)
multi$is_ds=as.character(Survey$V240..Sex)
multi$is_ds = ifelse(multi$is_ds==1,"M","F")
multi$exp_ml=expml
multi$exp_code = exp_code
M=sum(as.numeric(ifelse(Survey$V240..Se=="1",1,0)))
Fe = 2232-1084
get_binCI <- function(x,n){
  as_data_frame(setNames(as.list(binom.test(x,n)$conf.int),c("lwr","upr")))}
multi %>%
  filter(!is.na(exp_code) & !is.na(exp_ml) & !is.na(is_ds)) %>%
  group_by(is_ds, exp_code, exp_ml) %>%
  count() %>%
  spread(is_ds, n, fill = 0) %>%
  mutate(frac = M/(M+Fe)*100,
      lwr = get_binCI(M,(M+Fe))[[1]]*100,
          upr = get_binCI(M,(M+Fe))[[2]]*100,
      sum = M+Fe,
          err = mean(c(frac-lwr, upr-frac))
          ) %>%
  ggplot(aes(exp_code, exp_ml, size = sum, color = frac)) +
  geom_point() +
  scale_color_gradient(low = "#ffdf3f", high = "#DB7093") +
  scale_size(trans = "log10") +
  theme(axis.text.x   = element_text(angle=45, hjust=1, vjust=0.9)) +
  labs(color = "Gender", size = "# replies",
        x = "Age", y = "Employment Status") +
  ggtitle("Age and Employment Status by Gender")
```

```
Religion<-Survey$V147..Religious.person
Age<-Survey$V242..Age

show6_new<-data.frame(cbind(Religion,Age))

show6_new$Age <- cut(show6_new$Age,
                         breaks = c(-Inf, 25, 40, 55, 65,Inf),
                         labels = c("15-25 years", "26-40 years", "41-55 years", "56-65
years", "66-70 years"),
                         right = FALSE)
#show6_new
#esquisser(p1)
ggplot(show6_new) +
  aes(x = Age, y = Religion, fill = Age, colour = Age) +
  geom_violin(adjust = 1L, scale = "area") +
  scale_fill_viridis_d(option = "plasma") +
  scale_color_viridis_d(option = "plasma") +
  theme_par()+
theme( axis.text.x    = element_text(angle=25, hjust=1, vjust=0.9))+
   ggtitle("Whether You are a Religious Person or Not")
paralle<-fa.parallel(Survey1,fm='minres',fa='fa')
factor <- fa(Survey1,nfactors = 20,rotate = "oblimin",fm="minres")
print(factor)
```