Group 6: Sean Atkinson, Sameer Khan, Parker Davis

Assignment #1 Political Opinion Research

Abstract:

The data provided as a Pew research survey is analyzed to ascertain the sentiment of the voting population at that time. Politics in the United States has become increasingly polarized over the past few decades. Political leaning can often be determined by a person's response to questions pertaining to certain "wedge issues" regarding the state of the country. Therefore, it is possible to get a sense of which political party might be favored in an upcoming election based on the answers to a few key questions asked to survey respondents. This analysis provides a method to understand the principal factors contributing to this overall sentiment and how such issues could be addressed in future campaigns.

Keywords: Principal Component Analysis (PCA), Exploratory Factor Analysis (EFA), Election Sentiment

Introduction:

Team 6 political consultants have been tasked with analyzing past survey results to understand the sentiment of surveyed personnel and their perspective on both the government and President Donald Trump. The analysis will review the largest responses in survey results to understand the highly fragile analysis of presidential approval, perspective of the country's invasiveness, and overall approval of how the country is being run. BBC News has requested an assessment to understand the political climate better and build a portfolio profile of similar surveys to gain an advantage in understanding the political leadings of elections in the future.

The business problem is to understand if specific sentiment in a range of questions can be analyzed to predict the next favored political party.

Literature Review:

Similar approaches to sentiment have been gathered via social media, specifically Twitter (Bose, Dey, Roy, & Sarddar, 2019). The research team utilized political channels and specifically identified tweets to build sentiment analysis using the NRC Emotion Lexicon. The team identified eight emotions combined with a deep learning tool to analyze sentiment into positive, negative, and neutral categories.

Methods:

To build the components of the surveyed results, the results have been cataloged according to a completeness assessment. In the phone survey, many questions were not answered to provide a complete set of data points to complete the analysis. Only 30 questions were answered entirely and supplied as a dataset of the surveyed results. Team 6 started our principal component analysis using the dataset as mentioned above.

In assessing the resulting data, a few noticeable issues were identified. Combinations of multiple questions were answered with the "Don't Know/Refused" selection chosen. These survey question results were removed. The data also included multiple questions with low volumes of answers, so a targeted set of answers for questions included in this analysis was 600. Only including variables with responses over 600. These questions can be seen in Table 1.

Question #	Question text	Response	Count of responses

1	All in all, are you satisfied or dissatisfied with the way things are going in this country today?	Dissatisfied	930
2	Do you approve or disapprove of the way Donald Trump is handling his job as President?	Disapprove	792
20	Some people say they are basically content with the federal government, others say they are frustrated, and others say they are angry. Which of these best describes how you feel?	Frustrated	875
25	How much of the time do you think you can trust the government in Washington to do what is right? Just about always, most of the time, or only some of the time?	Only some of the time	1046
47	From what you have seen or heard about events in the new Congress, in general, do you think the Democrats in Congress are keeping the promises they made during the campaign, or not?	No, not keeping promises	692
50a	Most corporations make a fair and reasonable amount of profit	Statement 2	627
50c	We should pay less attention to problems overseas and concentrate on problems here at home	Statement 2	664
60	And thinking about politics and elections, would you say that personally insulting political opponents is	Never fair game	952
64	How fair do you think our present federal tax system is? Overall, would you say that our tax system is	Moderately fair	624
65	Please tell me how much, if at all, each of the following bothers you about the federal tax system? The feeling that some poor people don't pay their fair share	Not at all	617
66	At the present time, do you think religion as a whole is increasing its	Losing influence	1081

	influence on American life or losing its influence?		
68	As I name some groups, please tell me whether you feel each one is generally FRIENDLY toward religion, NEUTRAL toward religion, or UNFRIENDLY toward religion.	The Supreme Court Neutral towards religion	904
69	In your opinion, should churches and other houses of worship	Should express their view on day-to-day social and political questions	610
70	Do you approve or disapprove of the tax law passed by Donald Trump and Congress in 2017?	Disapprove	703

Table 1: Survey questions with over 600 responses

The resulting 14 questions were selected given the cut off for a great number of resulting answers as a larger proportion of answers would point to a stronger reaction to the question and provide an answer in the positive or negative and not neutral. Principal component analysis was utilized to identify variances between the questions. The selection of 14 questions was justified after a PCA was performed on the original 30 questions which resulted in low loadings for each component, thus making it very difficult to identify similarities between the question variables. Principal components (PC) are measures of variance across the variables where the first PC represents the largest variance in the selected dataset. Due to the first PC demonstrating such a large variance subsequent PC's demonstrate lower and lower variance between the variables. It is important to find a delicate balance in how many PC's to use so that the model can be easy to understand and process while still maintaining a high level of relational information. For this research 3 PC's were used to achieve the goal.

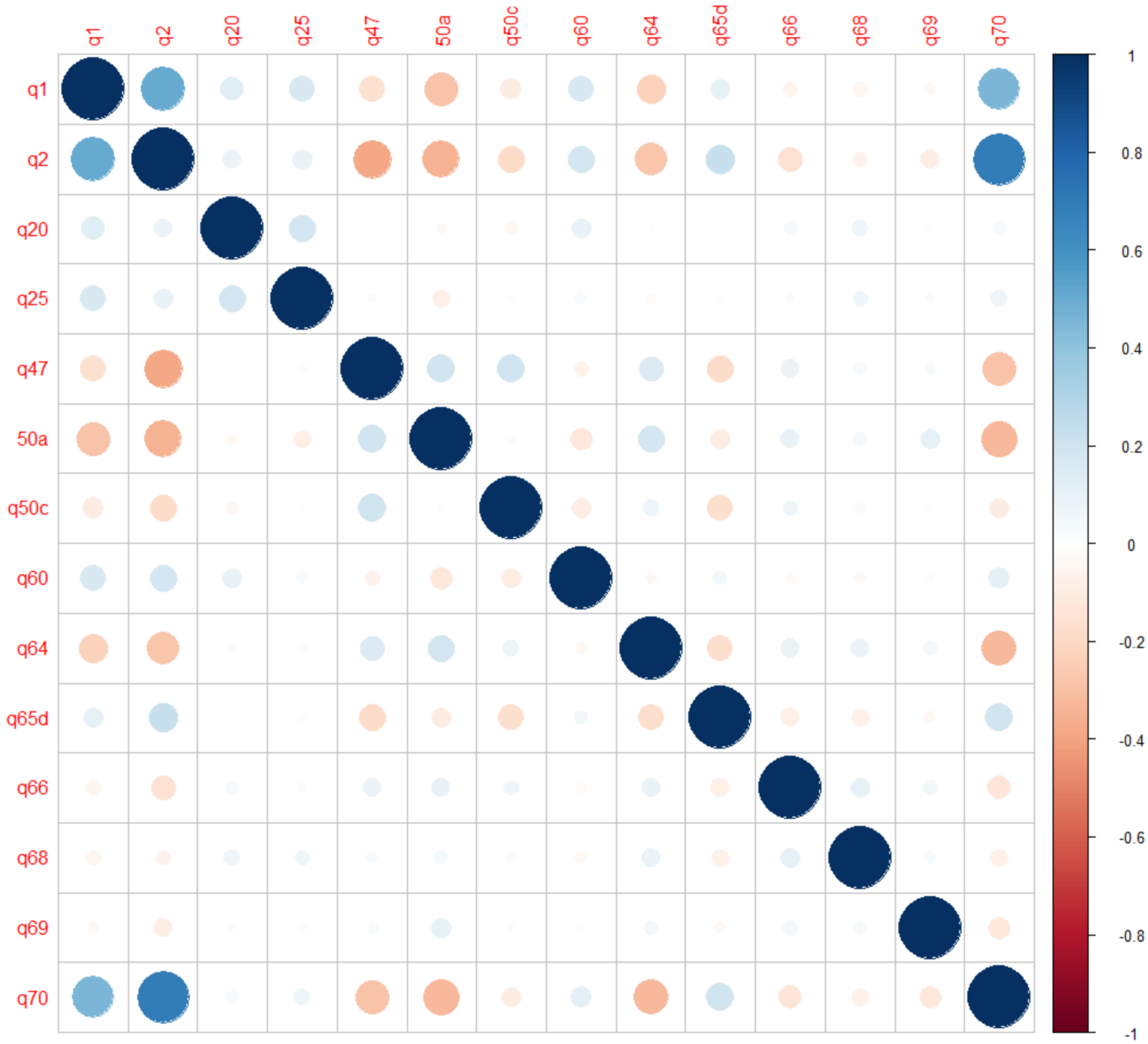

Figure 1 Correlation Matrix

Figure 1 provides a visualization of the variables from the reduced dataset. The highest correlations of 2 variables are indicated with both color (positive and negative correlation) and circle size (strength of correlation) . As part of our preliminary analysis we can identify higher correlations between Q1, Q2 and Q70. This will be used as inference as we explore other techniques in PCA and EFA.

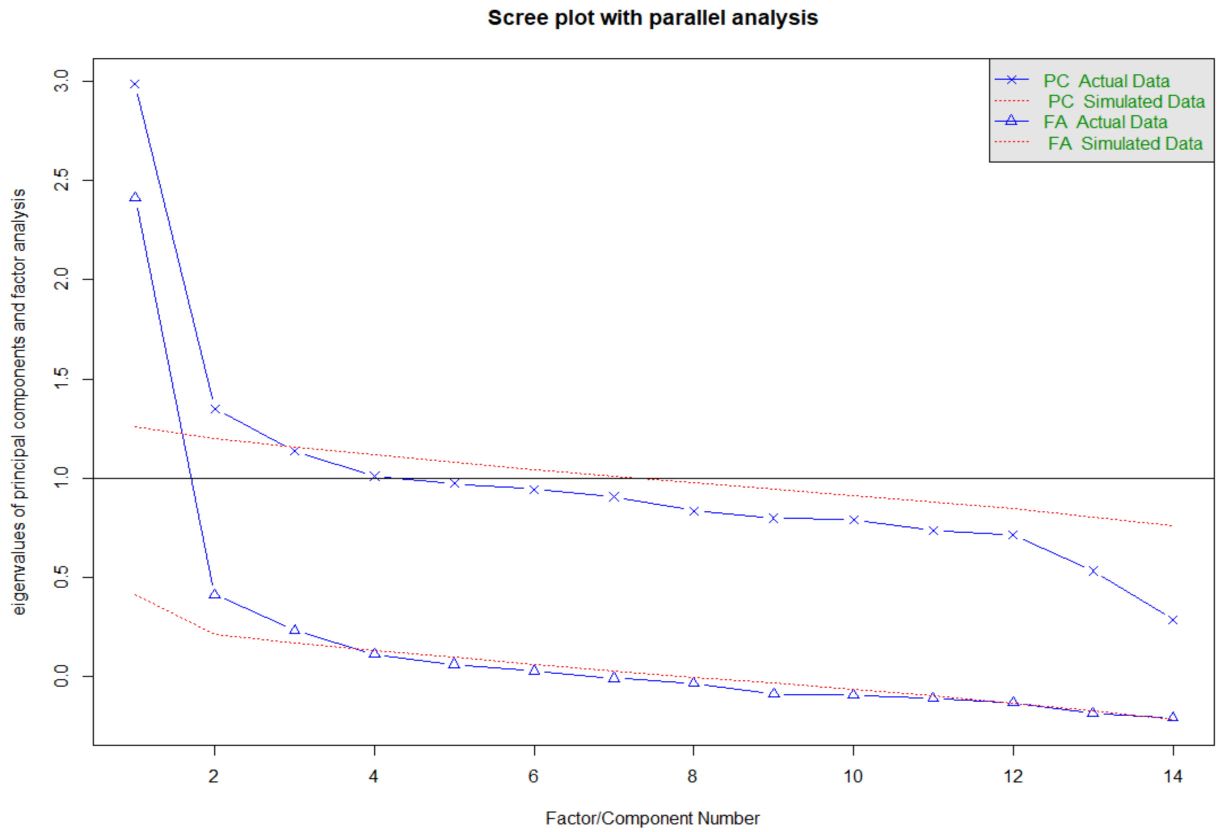

Figure 2 Parallel PCA and EFA graph

A method to understand how to manage the components of the data set is shown in Figure 2. Using a plotting method to graphically identify both the number of principal components for analysis and the number of factors recommended. The graph shows a scree plot indicating the number of components for PCA to be utilized. In this analysis 4 components are identified. In the same graph we also have in parallel produced the scree plot for EFA. The number of factors to retain shows only a single factor. As part of the research method Team 6 will review the results and confirm the output of Figure 2.

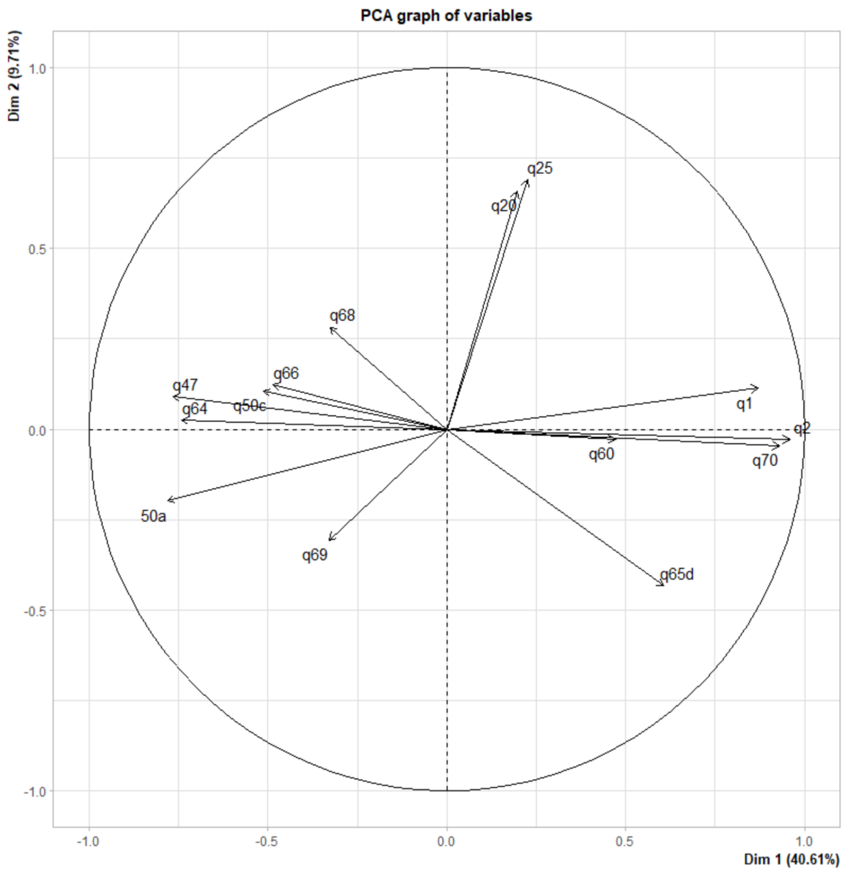

Figure 3 Correlation loading graph

In the loading graph we can identify those variables that are influenced by the principal component. This infers that closely related variables will have a very small angle of difference and those variables on the right side if the graph is positively correlated. This representation is used as a conformation of the correlation matrices and shows the positive correlation of Q1, Q2, and Q70.

Exploratory Factor Analysis

With the review of principal components Team 6 will work through the statistical methods of Factor Analysis. These methods applied to the underlying research will define a number of factors to be categorized as a factor used for analysis. The Scree plot from Figure 2 indicates a single factor to be identified. This output will be assessed in the following analysis.

Using combinations of analysis variables the following tests were performed:

Test	**Number of factors**	**Rotation**	**Factoring method**
fa1	4	None	principal analysis
fa2	4	Varimax	principal analysis
fa3	2	Varimax	principal analysis
fa4	2	Varimax	maximum likelihood
fa5	1	Varimax	maximum likelihood

Table 2: Exploratory Factor Analysis variables

Test output

In the following analysis we are looking at specific indicators:

1. Looking for Chi square close to the number of observations 1500
2. TLI > 0.9
3. RMSEA - < 0.05

Test	**Empirical Chi Square**	**Tucker Lewis**	**RMSEA**
fa1	71.31	0.978	0.021
fa2	71.31	0.978	0.021
fa3	242.49	0.923	0.038
fa4	255.48	0.928	0.037
fa5	599.74	0.879	0.048

Table 3: Exploratory Factor Analysis results

To complement this analysis the utility of the factor loadings represented in Figure 4-13 provide indicators of how many components should be used; the columnar data has a cut off of 0.5 for the components to be considered loaded to our analysis and confirming alignment. Observation of the PA components for 4 factors shows no value in PA2, 3, 4.

Test #1: fa1 analysis

```
Factor Analysis using method =  pa
Call: fa(r = reduced.pew, nfactors = 4, rotate = "none", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
                                                                   PA1   PA2   PA3   PA4    h2   u2 com
q1Dissatisfied                                                    0.59  0.25 -0.01 -0.11 0.418 0.58 1.4
q2Disapprove                                                      0.87  0.03  0.00  0.18 0.783 0.22 1.1
q20Frustrated                                                     0.10  0.36  0.21 -0.05 0.186 0.81 1.9
q25Only some of the time                                          0.12  0.37  0.07 -0.07 0.158 0.84 1.4
q47No, not keeping promises                                      -0.42  0.20 -0.16 -0.12 0.260 0.74 2.0
q50aStatement #2                                                 -0.44 -0.05  0.11  0.08 0.212 0.79 1.2
q50cStatement #2                                                 -0.23  0.18 -0.42  0.03 0.265 0.74 2.0
q60Never fair game                                                0.21  0.09  0.12 -0.09 0.073 0.93 2.4
q64Moderately fair                                               -0.38  0.11  0.09  0.14 0.184 0.82 1.6
q65dNot at all                                                    0.29 -0.22  0.15 -0.10 0.168 0.83 2.7
q66Losing influence                                              -0.18  0.17  0.05  0.09 0.075 0.93 2.6
q68dNeutral toward religion                                      -0.11  0.19  0.09  0.22 0.105 0.90 2.8
q69Should express their views on day-to-day social and political questions -0.12  0.04  0.11  0.04 0.030 0.97 2.4
q70Disapprove                                                     0.77  0.00 -0.19  0.08 0.635 0.36 1.1
```

Figure 4: Factor Analysis for 4 factors using principal analysis

PA2, PA3, and PA4 for unrotated analysis is not providing us with valuable loadings using this number of factors. We can adjust the rotation of the analysis using the varimax function. PA1 has Q1, Q2, Q70 has highly loaded variables in the analysis so is providing some inference to align to previous analysis showing a single factor recommended by exploratory factor analysis as seen in Figure 5

```
                       PA1  PA2  PA3  PA4
SS loadings           2.47 0.53 0.37 0.18
Proportion Var        0.18 0.04 0.03 0.01
Cumulative Var        0.18 0.21 0.24 0.25
Proportion Explained  0.70 0.15 0.10 0.05
Cumulative Proportion 0.70 0.85 0.95 1.00
```

Figure 5: Eigenvalue results of fa1 analysis

Test #2: fa2 analysis

Using rotation – varimax with 4 factors

```
Factor Analysis using method =  pa
Call: fa(r = reduced.pew, nfactors = 4, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
                                                                       PA1   PA3   PA2   PA4    h2   u2 com
q1Dissatisfied                                                        0.48 -0.11  0.40 -0.12 0.418 0.58 2.2
q2Disapprove                                                          0.80 -0.34  0.18 -0.01 0.783 0.22 1.5
q20Frustrated                                                         0.00 -0.05  0.41  0.12 0.186 0.81 1.2
q25Only some of the time                                              0.07  0.07  0.38  0.08 0.158 0.84 1.2
q47No, not keeping promises                                          -0.32  0.40  0.04  0.02 0.260 0.74 1.9
q50aStatement #2                                                     -0.40  0.05 -0.15  0.17 0.212 0.79 1.7
q50cStatement #2                                                      0.00  0.51 -0.08  0.04 0.265 0.74 1.1
q60Never fair game                                                    0.11 -0.13  0.20 -0.05 0.073 0.93 2.5
q64Moderately fair                                                   -0.32  0.12 -0.04  0.26 0.184 0.82 2.3
q65dNot at all                                                        0.14 -0.33 -0.01 -0.20 0.168 0.83 2.1
q66Losing influence                                                  -0.14  0.11  0.07  0.20 0.075 0.93 2.7
q68dNeutral toward religion                                          -0.05  0.05  0.07  0.31 0.105 0.90 1.2
q69Should express their views on day-to-day social and political questions -0.14 -0.03  0.02  0.10 0.030 0.97 2.0
q70Disapprove                                                         0.76 -0.16  0.11 -0.14 0.635 0.36 1.2
```

Figure 6: Factor Analysis for 4 factors with Varimax rotation

PA3 now introduces Q50a as loaded showing some indications that other variables may have some inference to consider in the analysis of building multiple factors for the assessment criteria. PA1 now has Q2, Q70 as highly loaded, with the utility of a rotation of negatively correlated variables we start to identify stronger loadings.

```
                       PA1  PA3  PA2  PA4
SS loadings           1.88 0.73 0.60 0.34
Proportion Var        0.13 0.05 0.04 0.02
Cumulative Var        0.13 0.19 0.23 0.25
Proportion Explained  0.53 0.21 0.17 0.10
Cumulative Proportion 0.53 0.74 0.90 1.00
```

Figure 7: Eigenvalue results of fa2 analysis

The output provides an eigenvalue of 1.88, the result is high but not aligned to that of unrotated factor analysis as we are addressing the need for SS loadings to identify the highest value with alignment to our other measure of building a number of factors with the greatest number of statistical indicators. These values point to a single principle component as the cut off is 1

Test #3: fa3 analysis

```
Factor Analysis using method =  pa
Call: fa(r = reduced.pew, nfactors = 2, rotate = "varimax", fm = "pa")
Standardized loadings (pattern matrix) based upon correlation matrix
                                                                   PA1   PA2    h2   u2 com
q1Dissatisfied                                                    0.44  0.47 0.408 0.59 2.0
q2Disapprove                                                      0.77  0.37 0.734 0.27 1.4
q20Frustrated                                                    -0.05  0.36 0.130 0.87 1.0
q25Only some of the time                                         -0.05  0.40 0.161 0.84 1.0
q47No, not keeping promises                                      -0.45 -0.01 0.203 0.80 1.0
q50aStatement #2                                                 -0.38 -0.21 0.192 0.81 1.6
q50cStatement #2                                                 -0.24  0.01 0.056 0.94 1.0
q60Never fair game                                                0.15  0.17 0.053 0.95 2.0
q64Moderately fair                                               -0.39 -0.05 0.158 0.84 1.0
q65dNot at all                                                    0.34 -0.06 0.122 0.88 1.1
q66Losing influence                                              -0.24  0.09 0.066 0.93 1.2
q68dNeutral toward religion                                      -0.17  0.12 0.044 0.96 1.8
q69Should express their views on day-to-day social and political questions -0.13  0.00 0.017 0.98 1.0
q70Disapprove                                                     0.70  0.29 0.579 0.42 1.3
```

Figure 8: Factor Analysis for 2 factors with Varimax rotation

```
                      ML2  ML1
SS loadings           1.21 1.05
Proportion Var        0.40 0.35
Cumulative Var        0.40 0.75
Proportion Explained  0.53 0.47
Cumulative Proportion 0.53 1.00
```

Figure 9: Eigenvalue results of fa3 analysis

Using 2 factors with the varimax rotation still being utilized shows that no significant loads are within the second PA2 factor. This along with the SS loading value of 1.21 shows the methods applied are not identifying a stronger statistical significance of eigenvalue to that of previous analysis.

Test #4: fa4 analysis

```
Factor Analysis using method =  ml
Call: fa(r = reduced.pew, nfactors = 2, rotate = "varimax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
                                                                   ML1   ML2    h2   u2 com
q1Dissatisfied                                                    0.44  0.46 0.397 0.60 2.0
q2Disapprove                                                      0.80  0.35 0.759 0.24 1.4
q20Frustrated                                                    -0.06  0.36 0.132 0.87 1.0
q25Only some of the time                                         -0.05  0.40 0.165 0.84 1.0
q47No, not keeping promises                                      -0.45  0.00 0.207 0.79 1.0
q50aStatement #2                                                 -0.37 -0.21 0.179 0.82 1.6
q50cStatement #2                                                 -0.22  0.01 0.049 0.95 1.0
q60Never fair game                                                0.14  0.18 0.053 0.95 1.9
q64Moderately fair                                               -0.37 -0.05 0.141 0.86 1.0
q65dNot at all                                                    0.32 -0.07 0.107 0.89 1.1
q66Losing influence                                              -0.24  0.09 0.063 0.94 1.3
q68dNeutral toward religion                                      -0.15  0.11 0.036 0.96 1.9
q69Should express their views on day-to-day social and political questions -0.13  0.00 0.018 0.98 1.0
q70Disapprove                                                     0.73  0.29 0.612 0.39 1.3
```

Figure 10: Factor Analysis for 2 factors with Varimax rotation and maximum likelihood

Another variable to utilize in factor analysis is the use of the “maximum likelihood” factoring method. This method applies linear combinations to form factors. In Test #4, the test is utilized with 2 factors and provides a goodness of fit for the variables. Maximum likelihood provides a technique to understand the Chi Squared output to show how well the factors align.

```
                       ML1  ML2
SS loadings           2.11 0.81
Proportion Var        0.15 0.06
Cumulative Var        0.15 0.21
Proportion Explained  0.72 0.28
Cumulative Proportion 0.72 1.00
```

Figure 11: Eigenvalue results of fa4 analysis

Test #4 shows no loadings for the ML2 category of factors but has enhanced the Eigenvalue to 2.11. The use of Maximum likelihood will be used for Test 5 as a factoring method and will address a single factors utility for the dataset and underlying business requirement as part of this report.

Test #5: fa5 analysis

Moving to one factor with rotation and factoring method of maximum likelihood. The following test will address the original scree plots output and determine if a single factor will provide an optimized approach for this dataset.

```
Factor Analysis using method =  ml
Call: fa(r = reduced.pew, nfactors = 1, rotate = "varimax", fm = "ml")
Standardized loadings (pattern matrix) based upon correlation matrix
                                                                      ML1     h2   u2 com
q1Dissatisfied                                                       0.57 0.3289 0.67   1
q2Disapprove                                                         0.88 0.7665 0.23   1
q20Frustrated                                                        0.09 0.0086 0.99   1
q25Only some of the time                                             0.11 0.0128 0.99   1
q47No, not keeping promises                                         -0.41 0.1708 0.83   1
q50aStatement #2                                                    -0.42 0.1764 0.82   1
q50cStatement #2                                                    -0.20 0.0391 0.96   1
q60Never fair game                                                   0.20 0.0413 0.96   1
q64Moderately fair                                                  -0.36 0.1289 0.87   1
q65dNot at all                                                       0.26 0.0694 0.93   1
q66Losing influence                                                 -0.18 0.0321 0.97   1
q68dNeutral toward religion                                         -0.09 0.0081 0.99   1
q69Should express their views on day-to-day social and political questions -0.12 0.0144 0.99   1
q70Disapprove                                                        0.78 0.6105 0.39   1
```

Figure 12: Factor Analysis for 1 factor with Varimax rotation and maximum likelihood

```
                ML1
SS loadings    2.41
Proportion Var 0.17
```

Figure 13: Eigenvalue results of fa5 analysis

Given these values, EFA's utility will utilize a single factor and group variables Q1, Q2, and Q70 into a sentiment factor. The responses to the questions of both presidential performance, attitude towards the country, and specific legislation will provide insight into the underlying sentiment of the survey takers. The single factor provides a stronger goodness of fit test via the Chi Squared measure and a comparative Eigenvalue to that of the original no rotation parallel analysis with four factors (see Table 3).

Results:

EFA shows one factor as a viable solution, given the loadings from Test #5, provides the following combination of questions to measure sentiment via the survey analysis (See Figure 14) . The single factor approach provides 3 variables to be utilized to build a single factor for analysis. The utility of PCA identified a specific pattern of components to be utilized and in combination with the statistical techniques of EFA the team was able to identify both a method to understand sentiment based on a reduced number of survey responses and where a relationship between specific questions can be applied as a sentiment factor.

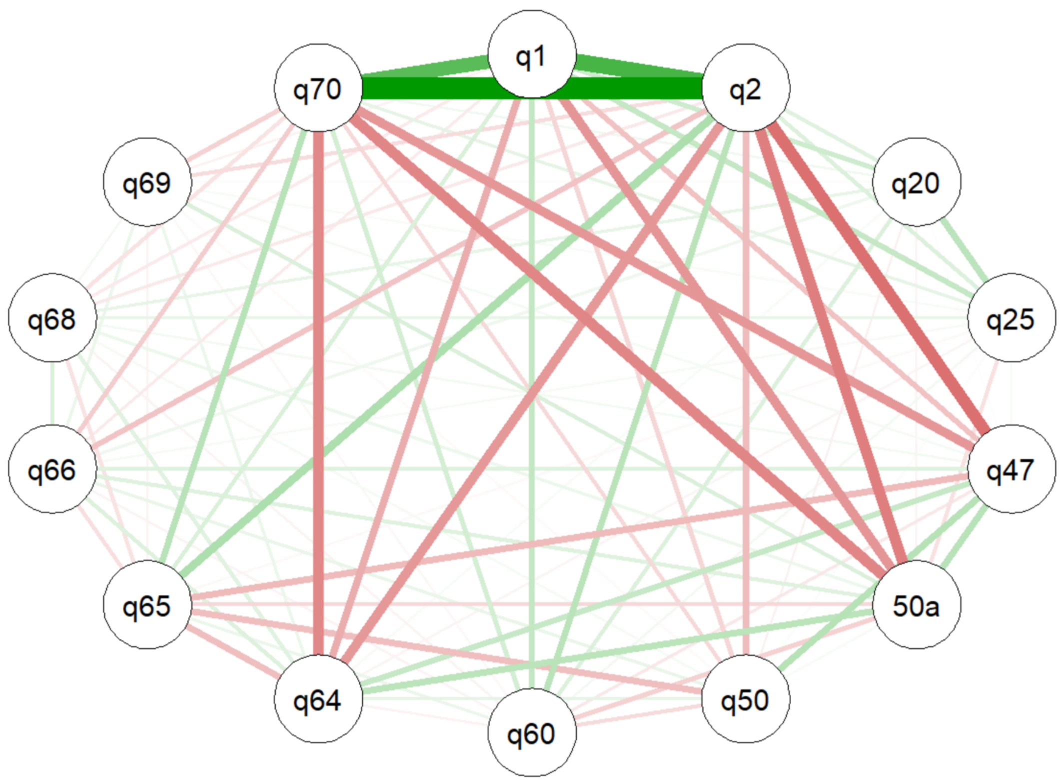

Figure 14 Qgraph to identify relationships between variables

Conclusion:

Understanding the survey's political sentiment has provided specific insight into relevant components to specifically address the cumulative effect of a few vital variables and the questions about presidential and government function. The analysis has provided several key considerations and question types to encourage participation and completion of survey data. Answers to these questions also indicate specific sentiment toward incumbent presidents and contribute to presidential elections' predictive decisions. For non-US-based media outlets, understanding the American Political consciousness is important concerning the sentiment of US citizens' perspective to both the current government structure and presidential effectiveness. The analysis draw a conclusion for future inference with respect to US political climate the BBC would structure specific survey questions with the following questions:

1: All in all, are you satisfied or dissatisfied with the way things are going in this country today?

2: Do you approve or disapprove of the way **[Current US President]** is handling his job as President?

3: Do you approve or disapprove of the **[Law, Bill, Act]** passed by **[Current US president]** and Congress in **[Year]**?

The following as a sentiment factor will identify current political opinions and views of the survey takers and can provide the BBC with indicators for utilization in reviewing the likelihood of the current president and their political party maintaining their position in the White House. With respect to Q3 research will need to be undertaken to understand specific legislation that could be polarizing for survey takers as a means to build specific indicators of sentiment.

Team 6 would like to thank the BBC for the opportunity to provide this report, your business and continued support is appreciated.

References:

Bose, R., Dey, R. K., Roy, S., & Sarddar, D. (2019). Analyzing political sentiment using Twitter data. In *Information and communication technology for intelligent systems* (pp. 427-436). Springer, Singapore.

Complete Guide to Factor Analysis (updated 2022). Qualtrics. (2022, January 18). Retrieved January 25, 2022, from https://www.qualtrics.com/experience-management/research/factor-analysis/