RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
Recruiting the Cyber Operations Officer (17A) Survey

This survey is being conducted to obtain relevant information from current Army Cyber Operations Officers (17As) on how they would assess the recruitment and application process for becoming a 17A. This survey will address background and experience history as well as motivation for becoming 17As. No PII will be recorded with data and all analysis will be reported anonymously in aggregate. All data will be collected and analyzed in a virtual computing environment maintained in accordance with applicable data at rest regulations and standards.

Welcome, we are requesting your participation in a brief survey created by Sean Gibbs and Wallace Nicholson, students at the Naval Postgraduate School (NPS). This survey is a part of our research on the Army’s recruitment of the Cyber Operations Officer. Your input will assist us in assessing the current Army’s human resource model (HRM) for recruiting the Cyber Operations Officer (17A). The survey will help us better understand the current population of Army Cyber Operations Officers and more accurately assess the Army’s HRM for recruiting these officers. The survey results will also assist us in conducting comparative analysis, determining how the Army HRM for recruiting Cyber Operations Officers aligns with best practices. In addition, this study has the potential to provide the groundwork for the improvement of the current Army HRM for recruiting Cyber Operations Officers that could positively impact long term effectiveness.

There are 40 questions in this survey.

Consent

1 [] Introduction. You are invited to participate in a research study entitled Recruiting the Army Cyber Leader: an analysis and evaluation of the effectiveness of the Army’s human resource model (HRM) used for recruiting the Cyber Operations Officer (17A). The purpose of the research is to evaluate the Army’s use of its current human resource model to recruit Cyber Operations Officers.

Procedures. This study will be conducted using three research methods: collection and analysis of data from the Cyber Branch; comparative analysis between Army recruitment of 17As and other government / commercial sector organizations recruitment of personnel with comparable skill sets; and a survey of current 17A Cyber Operations Officers.

- Subjects who choose to participate in this study will only be a part of the survey portion and will only be responsible for completing and submitting the survey. The survey will ask respondents about the following: basic demographic information, academic/professional background and experience, assessment of the 17A recruitment process, job satisfaction, current duty position / assignment and motivations for becoming a 17A.
- The expected duration of each subject’s involvement in this study is purely based on the amount of time it takes for them to complete the survey, which should take no longer than 15 minutes. The survey opens on December 12, 2016 and closes on February 1, 2017.
- The survey will be sent out to all current cyber operations officers for voluntary participation in this study.

Location. The survey will take place online and all other methods of research will take place from the Naval Postgraduate School in Monterey, CA.

Voluntary Nature of the Study. Your participation in this study is strictly voluntary. If you choose to
participate you can change your mind at any time and withdraw from the study. You will not be penalized in any way or lose any benefits to which you would otherwise be entitled if you choose not to participate in this study or to withdraw.

Potential Risks and Discomforts. The potential risks of participating in this study are: There is minimal risk of breach of confidentiality. No PII will be recorded with data, all analysis will be reported anonymously in aggregate. All data will be collected and analyzed in a virtual computing environment maintained in accordance with applicable data at rest regulations and standards.

Anticipated Benefits. The primary benefit of this study will be fully understanding the Army HRM for recruiting Cyber Operations Officer and determining whether or not it aligns with best practices. In addition, this study has the potential to provide the groundwork for the improvement of the current Army HRM for recruiting Cyber Operations Officers (or other specialty skill sets) that could positively impact long term effectiveness. You will not directly benefit from your participation in this research.

Confidentiality & Privacy Act. Any information that is obtained during this study will be kept confidential to the full extent permitted by law. All efforts, within reason, will be made to keep your personal information in your research record confidential but total confidentiality cannot be guaranteed. [All data will be collected and analyzed in a virtual computing environment maintained on the NPS secure server.]

Points of Contact. If you have any questions or comments about the research, or you experience an injury or have questions about any discomforts that you experience while taking part in this study please contact the Principal Investigator, Mr. Steve Mullins, (831) 656-7939, sjmullin@nps.edu. Questions about your rights as a research subject or any other concerns may be addressed to the Navy Postgraduate School IRB Chair, Dr. Larry Shattuck, 831-656-2473, lgshattu@nps.edu.

Statement of Consent. I have read the information provided above. I have been given the opportunity to ask questions and all the questions have been answered to my satisfaction. I have been provided a copy of this form for my records and I agree to participate in this study. I understand that by agreeing to participate in this research and signing this form, I do not waive any of my legal rights. *

Please choose only one of the following:

☐ Yes
☐ No
Demographics

Basic information on demographic of respondents

2 [] Please select your age group. *

Please choose only one of the following:

- 18-24
- 25-31
- 32-38
- 39-45
- 46+

3 [] Please select your gender. *

Please choose only one of the following:

- Female
- Male

4 [] How long have you been in the Army? *

Please choose only one of the following:

- Less than 1 year
- 1 - 5 years
- 6 - 10 years
- 11 - 15 years
- 16 - 20 years
- More than 20 years
5. What is your current rank? * Please choose only one of the following:
   - 2LT
   - 1LT
   - CPT
   - MAJ
   - LTC
   - COL
   - GEN

6. How long have you held your current rank? * Please choose only one of the following:
   - Less than 1 year
   - 1 - 2 years
   - 3 - 4 years
   - 5 - 6 years
   - More than 6 years
Background / Experience
This group of questions will focus on respondents' academic, technical and military background and experience.

7 [ ] Do you hold a Bachelor of Science (B.S.) or Bachelor of Arts (B.A.) undergraduate degree? *

Please choose all that apply:

- B.S.
- B.A.
- No undergraduate degree
- Other: ____________________________

8 [ ] Is your undergraduate degree from a Science, Technology, Engineering or Mathematics (STEM) Field? *

Please choose only one of the following:

- Yes
- No

9 [ ] What was your undergraduate major? *

Please write your answer here: ____________________________

10 [ ] Do you hold a Master of Science (M.S.), Master of Arts (M.A.) or Master of Business Administration (M.B.A.) degree? *

Please choose all that apply:

- M.S.
- M.A.
- M.B.A.
- No graduate degree
- Other: ____________________________
11 [ ] Is your graduate degree from a STEM program?
Please choose only one of the following:

- Yes
- No

12 [ ] What was your graduate major?
Please write your answer here:

13 [ ] Do you hold a PhD?
Please choose only one of the following:

- Yes
- No

14 [ ] What type of PhD do you hold?
Please write your answer here:
15 [] What, if any, IT Certifications do you currently hold? Select all that apply: *

Please choose all that apply:

- NET+
- CCNA
- CCNP
- SEC+
- CASP
- GSLC
- CISSP
- CISM
- CEH
- PMP
- GSEC
- NONE
- Other: ________________________________________________

16 [] What was your previous MOS, if applicable?

Please write your answer here:

__________________________________________________________________

17 [] Do you have any experience in the following IT / Cyber related fields? Select all that apply: *

Please choose all that apply:

- Computer Network Defense (CND)
- Network Management
- Information Assurance
- Coding / Programming
- Ethical "White Hat" Hacking
- Database Creation / Management
- Network Engineering
- None
- Other: ________________________________________________________
What [Officer] leadership or key developmental positions have you held in the Army prior to becoming a 17A? Check all that apply: *

Please choose all that apply:

☐ Platoon / Section Leader
☐ Company Commander
☐ Executive / Operations Officer (XO/S3)
☐ Staff Primary
☐ Battalion Level Commander / Director
☐ Brigade Level Commander / Director
☐ Deputy Commander / Director
☐ None
☐ Other: ________________________________
19 [] How were you made aware of the Army’s recruitment of Officers for the Cyber Branch (17A)? *

Please choose all that apply:

☐ Branch Manager / Human Resources Command
☐ Chain of Command
☐ Friend
☐ Social Media / Internet
☐ Other: ____________________________

20 [] To what extent do you agree or disagree with the following statements: *

Please choose the appropriate response for each item:

The 17A application packet requirements allowed me to effectively represent my relevant skills.

Strongly disagree Disagree Neutral Agree Strongly Agree Not Applicable

My CoC was supportive of my participation in the application / recruitment process.

The Cyber Branch proponent was easily accessible during the application / recruitment process.
21 To what extent do you agree or disagree with the following statements: *

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>My current duty position and job responsibilities are in line with my expectations of those of a Cyber Operations Officer (17A) based on the application / recruitment process.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I have the technical skills required to perform my assigned duties as expected in my current duty position.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I have the technical skills and experience to advance as a 17A.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Current Duty Position / Assignment

22 [] Are you currently assigned to a position designated for a 17A on your unit's MTOE? *

Please choose only one of the following:

- Yes
- No
- Unknown

23 [] If yes, what is your current duty title? *

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '22 [CDPA1]' (Are you currently assigned to a position designated for a 17A on your unit's MTOE?)

Please choose only one of the following:

- Automation Management Officer
- Automation Operations Officer
- Battalion Commander
- Battle Captain / Watch Officer
- BDE Staff Primary
- BN / BDE S3
- BN / BDE XO
- BN Staff Primary
- Branch Chief
- Brigade Commander
- Company Commander
- Chief of Staff
- CORPS Staff Primary
- Cyber Defense Manager
- Cyber Exercise Planner
- Cyber Network Defense Manager
- Cyber Planner
- Cyber Protection Team Leader
- Cyber Protection Team Member
- Cyber Research Scientist
24 [] Did you attend MOS specific training prior to arrival at your current duty station? *

Please choose only one of the following:

- Yes
- No

25 [] If yes, which training / military education did you attend? *

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '24 [CDPA3]' (Did you attend MOS specific training prior to arrival at your current duty station?)

Please choose only one of the following:

- Cyber Basic Officer Leaders Course (CBOLC)
- 17A Qualification Course
- Other
Job Satisfaction / Performance

26 [ ] Are you currently rated by a 17A? *

Please choose only one of the following:

- Yes
- No

27 [ ] If yes, to what extent do you agree or disagree with the following statements: *

Only answer this question if the following conditions are met:
Answer was ‘Yes’ at question ’26 [JOBPER2]’ (Are you currently rated by a 17A?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>My 17A rater possesses the technical skill required for their duty position.</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>My 17A rater provides me with sufficient and sound technical counsel to perform my duties.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

28 [ ] If no, to what extent do you agree or disagree with the following statements: *

Only answer this question if the following conditions are met:
Answer was ’No’ at question ’26 [JOBPER2]’ (Are you currently rated by a 17A?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>My rater possesses the technical skill required to supervise me as a 17A.</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>My rater provides me sufficient and sound technical counsel to progress as a 17A.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
**29 [ ]Are you currently senior rated by a 17A?** *

Please choose **only one** of the following:

- Yes
- No

**30 [ ]If yes, to what extent do you agree or disagree with the following statements:** *

**Only answer this question if the following conditions are met:**
Answer was 'Yes' at question '29 [JOBPER3]' (Are you currently senior rated by a 17A?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- My 17A senior rater possesses the technical skill required for their duty position.
- My 17A senior rater provides me sufficient and sound technical counsel to progress as a 17A.

**31 [ ]If no, to what extent do you agree or disagree with the following statements:** *

**Only answer this question if the following conditions are met:**
Answer was 'No' at question '29 [JOBPER3]' (Are you currently senior rated by a 17A?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- My senior rater possesses the technical skill required to supervise me as a 17A.
- My senior rater provides me sufficient and sound technical counsel to progress as a 17A.
32 [] Do you currently rate any 17As? *

Please choose only one of the following:

- Yes
- No

33 [] If yes, to what extent do you agree or disagree with the following statements: *

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '32 [JOBPER4]' (Do you currently rate any 17As?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>17A(s) I rate possess the technical skills required for them to perform their assigned duties.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17A(s) I rate perform as expected when given technical tasks related to their assigned duties.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

34 [] Do you currently senior rate any 17As? *

Please choose only one of the following:

- Yes
- No

35 [] If yes, to what extent do you agree or disagree with the following statement: *

Only answer this question if the following conditions are met:
Answer was 'Yes' at question '34 [JOBPER5]' (Do you currently senior rate any 17As?)

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>17A(s) I senior rate have the technical skill and desire required to advance in this field.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Motivation

36 []To what extent do you agree or disagree with the following statements: *

Please choose the appropriate response for each item:

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>I became a 17A because I am very passionate about the mission of the Cyber Branch and the role of the Cyber Operations Officer.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I became a 17A because I have the technical experience and expertise to excel in the Cyber Branch.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I became a 17A because I wanted to gain technical skills and experience that would make me marketable after I leave the military.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I became a 17A because I was no longer satisfied with my previous branch.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I became a 17A because of the potential for greater opportunities for advancement in the Cyber Branch.</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

37 []In order of importance, rank the following attributes required for a 17A: *

All your answers must be different.

Please number each box in order of preference from 1 to 5

- Tactical Experience
- Leadership Experience
- Technical Experience
- STEM Degree
- IT Certifications
38 [] What was the most difficult / frustrating part of the 17A application / recruitment process?
Please write your answer here:

39 [] What could be done to improve the 17A application / recruitment process?
Please write your answer here:
Please utilize the text box below to provide any additional topics or questions you think should be added to this survey to more accurately assess the application/recruitment process for Cyber Operations Officers.

Please write your answer here:
Thank you for your participation. The Principal Investigator for this research is Mr. Steve Mullins, (831) 656-7939. This research has been approved by the Naval Postgraduate School Institutional Review Board (IRB). The IRB Chair is Dr. Larry Shattuck (831) 656-2473.


Submit your survey.
Thank you for completing this survey.
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
I wrote this function to find out how many factors we need, depending on the data set. It is called a ‘scree’ plot and can graphically tell you the optimal number of factors using the eigenvalues of the data’s correlation matrix.

```r
myscree<-function(indata){
ev <- eigen(cor(indata)) # get eigenvalues
ap <- parallel(subject=nrow(indata), var=ncol(indata), rep=100, cent=.05)
nS <- nScree(x=ev$values, aparallel=ap$eigen$qevpea)
plotnScree(nS)
}
```

Now, we need to pull in the data. Make sure you put the data somewhere you can reach. For R, the data has to be in .csv format.

```r
data1<-read.csv("C:/Users/jsshingl1/Documents/Projects/Survey 17A/Raw Survey Data.csv")
```

First, we are going to look at all of the data without subsetting. Some exploratory data analysis is a good start.

```r
numRanks<-data1%>%
group_by(Rank)%>%
tally()
numRanks$PercRanks<-numRanks$n/sum(numRanks$n)
numRanks
```

Thus, most of your observastions came from CPTS.

We are going to look at this for all of the different catagories. What is the mean and std deviation for each of the questions by rank? Sex?

```r
data1 %>%
group_by(Rank) %>%
summarise (mean_20A = mean(X20A), var_20A=var(X20A))
```

```r
# A tibble: 6 x 3
#  Rank mean_20A var_20A
#1 1 1.647059 4.2426471
#2 2 3.500000 2.1666667
#3 3 3.614286 1.4287785
```
data1 %>%
group_by(Rank) %>%
summarise(mean_20B = mean(X20B), var_20B = var(X20B))

# A tibble: 6 x 3
#  Rank mean_20B  var_20B
#  <int>  <dbl>   <dbl>
# 1     1 2.058824 4.9338235
# 2     2 3.954545 2.6168831
# 3     3 4.057143 2.2865424
# 4     4 4.214286 2.1236934
# 5     5 4.343750 2.0393145
# 6     6 4.888889 0.1111111

data1 %>%
group_by(Rank) %>%
summarise(mean_20C = mean(X20C), var_20B = var(X20C))

# A tibble: 6 x 3
#  Rank mean_20C  var_20C
#  <int>  <dbl>   <dbl>
# 1     1 1.882353 3.860294
# 2     2 3.363636 1.766234
# 3     3 3.357143 2.319876
# 4     4 3.452381 1.765970
# 5     5 3.343750 4.168347
# 6     6 4.000000 1.750000

data1 %>%
group_by(Rank) %>%
summarise(mean_21A = mean(X21A), var_21A = var(X21A))

# A tibble: 6 x 3
#  Rank mean_21A  var_21A
#  <int>  <dbl>   <dbl>
# 1     1 2.764706 3.816176
# 2     2 2.818182 3.489177
# 3     3 2.728571 3.128157
# 4     4 3.666667 2.471545
# 5     5 3.531250 2.321573
# 6     6 4.111111 1.111111

data1 %>%
group_by(Rank) %>%
summarise(mean_21B = mean(X21B), var_21B = var(X21B))

# A tibble: 6 x 3
#  Rank mean_21B  var_21B
#  <int>  <dbl>   <dbl>
# 1     1 3.882353 1.6102941
# 2     2 3.545455 3.4025974
# 3     3 3.728571 3.3600414
# 4     4 4.095238 1.050987
data1 %>%
  group_by(Rank) %>%
  summarise(mean_21C = mean(X21C), var_21C = var(X21C))

# # A tibble: 6 x 3
# #   Rank mean_21C var_21C
# #  <int>   <dbl>    <dbl>
# 1      1   3.941  0.809
# 2      2   4.000  2.190
# 3      3   4.114  1.610
# 4      4   4.167  0.776
# 5      5   3.938  1.351
# 6      6   4.333  0.750

data1 %>%
  group_by(Rank) %>%
  summarise(mean_36A = mean(X36A), var_36A = var(X36A))

# # A tibble: 6 x 3
# #   Rank mean_36A var_36A
# #  <int>   <dbl>    <dbl>
# 1      1   3.882  2.485
# 2      2   4.091  2.182
# 3      3   4.643  0.581
# 4      4   4.881  0.156
# 5      5   4.781  0.176
# 6      6   5.000  0.000

data1 %>%
  group_by(Rank) %>%
  summarise(mean_36B = mean(X36B), var_36B = var(X36B))

# # A tibble: 6 x 3
# #   Rank mean_36B var_36B
# #  <int>   <dbl>    <dbl>
# 1      1   3.824  2.654
# 2      2   4.045  2.045
# 3      3   4.314  1.001
# 4      4   4.119  1.181
# 5      5   4.156  0.846
# 6      6   4.333  0.750

data1 %>%
  group_by(Rank) %>%
  summarise(mean_36C = mean(X36C), var_36C = var(X36C))

# # A tibble: 6 x 3
# #   Rank mean_36C var_36C
# #  <int>   <dbl>    <dbl>
# 1      1   3.294  2.846
# 2      2   4.046  2.141
# 3      3   3.287  2.380
# 4      4   3.547  1.619
# 5      5   3.000  2.323
```r
# A tibble: 6 x 3
#  Rank mean_36D var_36D
#  <int>   <dbl>    <dbl>
#1     1  0.4706  0.8897
#2     2  2.6364  4.2424
#3     3  3.1714  1.9412
#4     4  2.5476  1.8148
#5     5  2.5625  1.2863
#6     6  2.3333  2.0000

data2<-data1[,c(1:4,10:15, 26:32, 36)]
cor(data2[,c(1:4,10:15, 26:32, 36)])
```

Now for some correlation plots. I found a nice package that make some really nice ones. I am also going to include Rank and

```
data2<-data1[,c(1:4,10:15, 26:32, 36)]
cor(data2[,c(1:4,10:15, 26:32, 36)])
```
<table>
<thead>
<tr>
<th></th>
<th>X36A</th>
<th>X36B</th>
<th>X36C</th>
<th>X36D</th>
<th>X36E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certs</td>
<td>0.145472540</td>
<td>0.228449367</td>
<td>0.131835247</td>
<td>0.031804254</td>
<td>0.112442152</td>
</tr>
<tr>
<td>Experience</td>
<td>0.192259535</td>
<td>-0.016870503</td>
<td>-0.03647009</td>
<td>0.10000000</td>
<td>0.06596468</td>
</tr>
<tr>
<td>OPExperience</td>
<td>0.192259535</td>
<td>-0.016870503</td>
<td>-0.03647009</td>
<td>0.10000000</td>
<td>0.06596468</td>
</tr>
<tr>
<td>Age</td>
<td>0.29691740</td>
<td>0.006779977</td>
<td>-0.08168687</td>
<td>0.08103889</td>
<td>0.09927932</td>
</tr>
<tr>
<td>M.F</td>
<td>-0.03105079</td>
<td>-0.098788510</td>
<td>-0.04599898</td>
<td>-0.16528379</td>
<td>0.04931653</td>
</tr>
<tr>
<td>TIS</td>
<td>0.234235633</td>
<td>0.072881166</td>
<td>-0.09243088</td>
<td>0.11796254</td>
<td>-0.01960941</td>
</tr>
<tr>
<td>Rank</td>
<td>0.313961121</td>
<td>0.057869480</td>
<td>-0.08465124</td>
<td>0.15817929</td>
<td>-0.01514267</td>
</tr>
<tr>
<td>X20A</td>
<td>0.281095361</td>
<td>0.213862231</td>
<td>0.09670597</td>
<td>0.27128354</td>
<td>0.14516250</td>
</tr>
<tr>
<td>X20B</td>
<td>0.269436999</td>
<td>0.180065951</td>
<td>0.05278756</td>
<td>0.16249718</td>
<td>0.04350882</td>
</tr>
<tr>
<td>X20C</td>
<td>0.164836581</td>
<td>0.176396625</td>
<td>0.14906870</td>
<td>0.05813199</td>
<td>0.06571135</td>
</tr>
<tr>
<td>X21A</td>
<td>0.126501761</td>
<td>0.089547456</td>
<td>0.01462158</td>
<td>-0.02628114</td>
<td>-0.02060765</td>
</tr>
<tr>
<td>X21B</td>
<td>0.115997771</td>
<td>0.286394819</td>
<td>-0.13490036</td>
<td>-0.04506151</td>
<td>-0.08723452</td>
</tr>
<tr>
<td>X21C</td>
<td>0.115021751</td>
<td>0.364245996</td>
<td>-0.09119579</td>
<td>-0.03467009</td>
<td>-0.11236593</td>
</tr>
<tr>
<td>X36A</td>
<td>1.000000000</td>
<td>0.540426800</td>
<td>0.21381082</td>
<td>0.24683256</td>
<td>0.18566258</td>
</tr>
<tr>
<td>X36B</td>
<td>0.540426800</td>
<td>1.000000000</td>
<td>0.07627077</td>
<td>0.11471948</td>
<td>0.18660407</td>
</tr>
<tr>
<td>X36C</td>
<td>0.213810821</td>
<td>0.076270773</td>
<td>1.00000000</td>
<td>0.17753609</td>
<td>0.33183250</td>
</tr>
<tr>
<td>X36D</td>
<td>0.246832561</td>
<td>0.114719478</td>
<td>0.17753609</td>
<td>1.00000000</td>
<td>0.17476232</td>
</tr>
<tr>
<td>X36E</td>
<td>0.185662581</td>
<td>0.186064071</td>
<td>0.33183250</td>
<td>0.17476232</td>
<td>1.00000000</td>
</tr>
<tr>
<td>Certs</td>
<td>0.145472540</td>
<td>0.228449367</td>
<td>0.13183525</td>
<td>0.03180425</td>
<td>0.112442152</td>
</tr>
<tr>
<td>Experience</td>
<td>0.094841385</td>
<td>0.359717623</td>
<td>-0.07676999</td>
<td>-0.03872512</td>
<td>0.06596468</td>
</tr>
<tr>
<td>OPExperience</td>
<td>0.192259535</td>
<td>-0.016870503</td>
<td>-0.03647009</td>
<td>0.10000000</td>
<td>0.06596468</td>
</tr>
<tr>
<td>Age</td>
<td>0.047423987</td>
<td>0.105577727</td>
<td>0.72281821</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M.F</td>
<td>0.044207250</td>
<td>-0.226694211</td>
<td>-0.15009959</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIS</td>
<td>-0.009327496</td>
<td>0.094456886</td>
<td>0.67486239</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>0.029053795</td>
<td>0.111523877</td>
<td>0.73890319</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X20A</td>
<td>0.047830457</td>
<td>0.04244444</td>
<td>0.21369219</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X20B</td>
<td>0.013254899</td>
<td>0.046515811</td>
<td>0.21124985</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X20C</td>
<td>0.088070445</td>
<td>0.02183183</td>
<td>0.15288753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X21A</td>
<td>-0.048946624</td>
<td>0.125955171</td>
<td>0.20045665</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X21B</td>
<td>0.101989816</td>
<td>0.21445700</td>
<td>0.16331759</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X21C</td>
<td>0.055418697</td>
<td>0.223947244</td>
<td>-0.03243205</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X36A</td>
<td>0.145472540</td>
<td>0.098441383</td>
<td>0.19225953</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X36B</td>
<td>0.228449367</td>
<td>0.359717621</td>
<td>-0.01687050</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X36C</td>
<td>0.131835247</td>
<td>-0.07676999</td>
<td>-0.1123175</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X36D</td>
<td>0.031804254</td>
<td>-0.03872512</td>
<td>0.05030855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X36E</td>
<td>0.112442152</td>
<td>0.06596468</td>
<td>0.11843457</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Certs</td>
<td>1.000000000</td>
<td>0.163601311</td>
<td>-0.03646918</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>0.1636013091</td>
<td>1.00000000</td>
<td>0.02664683</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Finally, a little bit of factor analysis

```r
myScree(data1[,c(10:15, 26:30)])
```
Non Graphical Solutions to Scree Test

This graph says that we should use a three factor model for these 12 questions. Let's give it a try!

```r
fa1 <- factanal(data1[, c(10:15, 26:30)], factors = 3, rotation = "varimax")
fa1
```

```
## Call:
## factanal(x = data1[, c(10:15, 26:30)], factors = 3, rotation = "varimax")
##
## Uniquenesses:
##   X20A X20B X20C X21A X21B X21C X36A X36B X36C X36D X36E
## 0.446 0.631 0.665 0.717 0.369 0.417 0.487 0.375 0.873 0.869 0.848
##
## Loadings:
##   Factor1 Factor2 Factor3
## X20A    0.709   0.224
## X20B    0.109   0.576   0.158
## X20C    0.559   0.152
## X21A    0.401   0.338
## X21B    0.791
## X21C    0.763
## X36A   -0.179   0.198   0.664
## X36B    0.431   0.659
## X36C   -0.162   0.107   0.299
## X36D    0.217   0.280
## X36E   -0.130   0.360
##
```
## Factor1 Factor2 Factor3
## SS loadings 1.648 1.371 1.284
## Proportion Var 0.150 0.125 0.117
## Cumulative Var 0.150 0.274 0.391
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 41.31 on 25 degrees of freedom.
## The p-value is 0.0213

Subsetting now: 17A as Rater. Notice that this allows me to add question 27A and 27B into the factor analysis.

```r
subdata1 <- data1 %>%
  filter(!is.na(X27A))
myscree(subdata1[, c(10:17, 26:30)])
```

**Non Graphical Solutions to Scree Test**

- Eigenvalues (>mean = 4)
- Parallel Analysis (n = 3)
- Optimal Coordinates (n = 1)
- Acceleration Factor (n = 1)

##This says 1, but we are going to still use 3.

```r
fa1 <- factanal(subdata1[, c(10:17, 26:30)], factors = 3, rotation = "varimax")
fa1
```

## Call:
## `factanal(x = subdata1[, c(10:17, 26:30)], factors = 3, rotation = "varimax")`
##
## # Uniquenesses:
## X20A X20B X20C X21A X21B X21C X27A X27B X36A X36B X36C X36D
## 0.752 0.767 0.872 0.826 0.078 0.661 0.105 0.447 0.507 0.443 0.943 0.857
## X36E
## Loadings:
## Factor1 Factor2 Factor3
## X20A 0.192 0.456
## X20B 0.102 0.464
## X20C 0.157 0.314
## X21A 0.330 0.249
## X21B 0.152 0.946
## X21C 0.118 0.195 0.536
## X27A 0.927 0.142 0.123
## X27B 0.734 0.118
## X36A 0.694
## X36B 0.176 0.646 0.330
## X36C 0.193 -0.136
## X36D -0.163 0.336
## X36E 0.271 0.246 -0.226
##
## SS loadings 1.737 1.729 1.462
## Proportion Var 0.134 0.133 0.112
## Cumulative Var 0.134 0.267 0.379
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 30.39 on 42 degrees of freedom.
## The p-value is 0.909

Subsetting now: 17A not Rater. We get to add 28A and 28B.

```r
subdata1 <- data1 %>%
  filter(is.na(X27A))
myscree(subdata1[,c(10:15,18:19,26:30)])
```
Non Graphical Solutions to Scree Test

Eigenvalues
Eigenvalues (>mean = 4)
Parallel Analysis (n = 4)
Optimal Coordinates (n = 4)
Acceleration Factor (n = 4)

fa1 <- factanal(subdata1[, c(10:15, 18:19, 26:30)], factors = 4, rotation = "varimax")
fa1

## This one says 4, so 4 it is!

fa1 <- factanal(subdata1[, c(10:15, 18:19, 26:30)], factors = 4, rotation = "varimax")
fa1

## Call:
## factanal(x = subdata1[, c(10:15, 18:19, 26:30)], factors = 4, rotation = "varimax")
##
## Uniquenesses:
## X20A X20B X20C X21A X21B X21C X28A X28B X36A X36B X36C X36D
## 0.387 0.534 0.516 0.647 0.429 0.237 0.005 0.119 0.442 0.414 0.731 0.817
## X36E
## 0.785
##
## Loadings:
## Factor1 Factor2 Factor3 Factor4
## X20A 0.166 0.732 0.223
## X20B 0.107 0.667
## X20C 0.677 0.132
## X21A 0.361 0.370 0.293
## X21B 0.273 0.686 -0.156
## X21C 0.868
## X28A 0.986 0.121
## X28B 0.912 0.136 0.177
## X36A 0.180 0.256 0.178 0.654
## X36B 0.501 0.576
## X36C -0.146 0.484
Subsetting now: 17A Senior Rater. We get to add 30A and 30B.

going with 2. Best p-value

fra1 <- factanal(subdata1[,c(10:15,20:21,26:30)], factors=2, rotation="varimax")

fra1

## Call:
## factanal(x = subdata1[, c(10:15, 20:21, 26:30)], factors = 2, rotation = "varimax")
##
## ## Uniquenesses:
## X20A X20B X20C X21A X21B X21C X30A X30B X36A X36B X36C X36D
## 0.753 0.503 0.874 0.983 0.931 0.965 0.005 0.349 0.465 0.486 0.967 0.874
## X36E
## 0.868
##
## Loadings:
## Factor1 Factor2
## X20A 0.489
## X20B 0.699
## X20C 0.339 0.106
## X21A 0.101
## X21B 0.221 0.144
## X21C 0.165
## X30A 0.996
## X30B 0.807
## X36A 0.705 0.194
## X36B 0.708 0.112
## X36C 0.147 0.104
## X36D 0.352
## X36E 0.276 0.236
##
## Loadings:
## Factor1 Factor2
## SS loadings 2.152 1.825
## Proportion Var 0.166 0.140
## Cumulative Var 0.166 0.306
##
## Test of the hypothesis that 2 factors are sufficient.
## The chi square statistic is 95.58 on 53 degrees of freedom.
## The p-value is 0.000305

Subsetting now: 17A not Senior Rater. We get to add 31A and 31B.

```r
subdata1 <- data1 %>%
  filter(is.na(X30A))
myscree(subdata1[, c(10:15, 22:23, 26:30)])
```
## Non Graphical Solutions to Scree Test

### Components
- Eigenvalues
- Eigenvalues (>mean = 4 )
- Parallel Analysis (n = 4 )
- Optimal Coordinates (n = 4 )
- Acceleration Factor (n = 1 )

---

```r
##Try 4. Go with 3.
fal<-factanal(subdata1[,c(10:15,22:23,26:30)], factors=3, rotation="varimax")
fal
```

---

```
## Call:
## factanal(x = subdata1[, c(10:15, 22:23, 26:30)], factors = 3, rotation = "varimax")
```

---

```
## Uniquenesses:
## X20A X20B X20C X21A X21B X21C X31A X31B X36A X36B X36C X36D
## 0.329 0.701 0.579 0.662 0.511 0.215 0.058 0.073 0.777 0.617 0.947 0.881
## X36E
## 0.947
```

---

```
## Loadings:
##     Factor1 Factor2 Factor3
## X20A  0.213   0.120   0.782
## X20B  0.206   0.106   0.495
## X20C    0.640
## X21A  0.400   0.395   0.150
## X21B  0.288   0.613  -0.175
## X21C   0.869  -0.160
## X31A  0.965   0.102
## X31B   0.957
## X36A  0.124   0.342   0.300
## X36B   0.599   0.150
## X36C   0.220
```

13
## X36D 0.345
## X36E 0.212
##
## Factor1 Factor2 Factor3
## SS loadings 2.21 1.82 1.674
## Proportion Var 0.17 0.14 0.129
## Cumulative Var 0.17 0.31 0.439
##
## Test of the hypothesis that 3 factors are sufficient.
## The chi square statistic is 91.62 on 42 degrees of freedom.
## The p-value is 1.5e-05

After relooking at the data, this is my recommendation:

1) Create a score for each of the factors, using all of the data from the dataset (all 190ish observations)

\[
\begin{align*}
\text{w1sum} &< -0.401 + 0.791 + 0.763 \\
\text{w1.1} &< -0.401 / \text{w1sum} \\
\text{w2.1} &< -0.791 / \text{w1sum} \\
\text{w3.1} &< -0.763 / \text{w1sum} \\
\text{w2sum} &< -0.709 + 0.576 + 0.559 \\
\text{w1.2} &< -0.709 / \text{w2sum} \\
\text{w2.2} &< -0.576 / \text{w2sum} \\
\text{w3.2} &< -0.559 / \text{w2sum} \\
\text{w3sum} &< -0.664 + 0.659 \\
\text{w1.3} &< -0.664 / \text{w3sum} \\
\text{w2.3} &< -0.659 / \text{w3sum}
\end{align*}
\]

\[
\begin{align*}
\text{data1$FactorScore1} &< -(\text{w1.1} \times \text{data1$X21A} + \text{w2.1} \times \text{data1$X21B} + \text{w3.1} \times \text{data1$X21C}) \\
\text{data1$FactorScore2} &< -(\text{w1.2} \times \text{data1$X20A} + \text{w2.2} \times \text{data1$X20B} + \text{w3.2} \times \text{data1$X20C}) \\
\text{data1$FactorScore3} &< -(\text{w1.3} \times \text{data1$X36A} + \text{w2.3} \times \text{data1$X36B})
\end{align*}
\]

Now, based on the demographic data, predict the factor score. We can do this by splitting the data into 2 sets, the training set and the test set.

\[
\begin{align*}
\text{traindata1} &< \text{sample_n(data1, 0.7*dim(data1)[1])} \\
\text{testdata1} &< \text{data1[-as.numeric(rownames(traindata1))]},]
\end{align*}
\]

Now we can do a little regression and prediction with these data sets and see how we did:

\[
\begin{align*}
\text{lmFact1} &< \text{lm(traindata1$FactorScore1~traindata1$Rank+traindata1$BS+traindata1$Certs+traindata1$Experience)} \\
\text{summary(lmFact1)}
\end{align*}
\]

Get Rid of Rank, No.UG.Degree, and Rank.1, whatever that is.

\[
\begin{align*}
\text{lmFact1} &< \text{lm(traindata1$FactorScore1~traindata1$Age+traindata1$M.F+traindata1$TIS+traindata1$TIG)} \\
\text{summary(lmFact1)}
\end{align*}
\]

## Call:
\[
\text{lm(formula = traindata1$FactorScore1 - traindata1$Age + traindata1$M.F + traindata1$TIS + traindata1$TIG)}
\]

## Residuals:
\[
\begin{align*}
\text{Min} & \quad \text{1Q} & \quad \text{Median} & \quad \text{3Q} & \quad \text{Max} \\
-2.5316 & -0.4323 & 0.1966 & 0.6588 & 1.8267
\end{align*}
\]

## Coefficients:
## Call:
```
lm(formula = traindata1$FactorScore1 ~ traindata1$Age + traindata1$M.F +
    traindata1$TIS)
```

## Residuals:
```
Min 1Q Median 3Q Max
-2.5779 -0.3973 0.1612 0.6675 1.7721
```

## Coefficients:
```
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           4.48501   0.38719  11.583   < 2e-16 ***
traindata1$Age       -0.26460   0.12397  -2.134   0.03470 *
traindata1$M.F       -0.53175   0.25854  -2.057   0.04172 *
traindata1$TIS       0.27841   0.10280   2.708   0.00768 **
traindata1$TIG        0.05303   0.07229   0.733   0.46460
```

## Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

## Residual standard error: 0.9332 on 129 degrees of freedom  
## Multiple R-squared: 0.09246, Adjusted R-squared: 0.06432  
## F-statistic: 3.286 on 4 and 129 DF, p-value: 0.01334

---

Get rid of TIG. Probably taken care of by TIS
```
lmFact1<-.lm(traindata1$FactorScore1 ~ traindata1$Age + traindata1$M.F + traindata1$TIS)
```

```
summary(lmFact1)
```

## Call:
```
lm(formula = traindata1$FactorScore1 ~ traindata1$Age + traindata1$M.F +
    traindata1$TIS)
```

## Residuals:
```
Min 1Q Median 3Q Max
-2.5779 -0.3973 0.1612 0.6675 1.7721
```

## Coefficients:
```
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           4.5137  0.3845    11.739   < 2e-16 ***
traindata1$Age       -0.2604  0.1236   -2.107   0.03707 *
traindata1$M.F       -0.5377  0.2579   -2.085   0.03906 *
traindata1$TIS       0.2935  0.1005    2.920   0.00413 **
```

## Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  

## Residual standard error: 0.9315 on 130 degrees of freedom  
## Multiple R-squared: 0.08868, Adjusted R-squared: 0.06765  
## F-statistic: 4.217 on 3 and 130 DF, p-value: 0.006992

---

No TIS
```
lmFact1<-.lm(traindata1$FactorScore1 ~ traindata1$Age + traindata1$M.F)
```

```
summary(lmFact1)  ##This sucks.
```

## Call:
```
lm(formula = traindata1$FactorScore1 ~ traindata1$Age + traindata1$M.F)
```

## Residuals:
```
Min 1Q Median 3Q Max
-2.7662 -0.3885 0.1726 0.6782 1.5217
```

## Coefficients:
```
                      Estimate Std. Error t value Pr(>|t|)    
(Intercept)           4.36374  0.39187   11.136   <2e-16 ***
```

15
# Train data statistics

```r
## traindata1$Age  0.03210 0.07447 0.431 0.667
## traindata1$M.F  -0.49086 0.26474 -1.854 0.066 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9579 on 131 degrees of freedom
## Multiple R-squared: 0.02891, Adjusted R-squared: 0.01408
## F-statistic: 1.95 on 2 and 131 DF, p-value: 0.1464
```

I am going to try a different method.

```r
attach(traindata)

treeFact1 <- rpart(FactorScore1 ~ Rank + Age + TIS + BS + Certs + Experience + MS + OPEXperience, cp = 0.014)

rpart.plot(treeFact1, main = "TREE Factor 1")
```

### TREE Factor 1

![Tree Factor 1 Diagram](image)

```r
# text(treeFact1)

treeFact2 <- rpart(FactorScore2 ~ Rank + Age + TIS + BS + Certs + Experience + MS + OPEXperience, cp = .014)

printcp(treeFact2)
```

```r
##
## Regression tree:
## rpart(formula = FactorScore2 ~ Rank + Age + TIS + BS + Certs + Experience + MS + OPEXperience, cp = 0.014)
##
## Variables actually used in tree construction:
## [1] BS Rank
```
## Root node error: 222.67/134 = 1.6617
##
## n= 134
##
## | CP | nsplit | rel error | xerror  | xstd  |
##|-----|---------|------------|---------|-------|
##| 1   | 0.2622  | 0          | 1.0000  | 1.0132 | 0.1652 |
##| 2   | 0.0168  | 1          | 0.7378  | 0.7768 | 0.1194 |
##| 3   | 0.0140  | 2          | 0.7209  | 0.8434 | 0.1183 |

rpart.plot(treeFact2, main="TREE Factor 2")

TREE Factor 2

<table>
<thead>
<tr>
<th>3.6</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7</td>
<td>11%</td>
</tr>
<tr>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>Rank &lt; 1.5</td>
<td>no</td>
</tr>
<tr>
<td>3.8</td>
<td>89%</td>
</tr>
<tr>
<td>BS &gt;= 0.5</td>
<td></td>
</tr>
<tr>
<td>3.8</td>
<td>78%</td>
</tr>
<tr>
<td>4.3</td>
<td>11%</td>
</tr>
</tbody>
</table>

# text(treeFact2)

treeFact3<-rpart(FactorScore3~Rank+Age+TIS+BS+Certs+Experience+MS+OPExperience, cp=0.006)
printcp(treeFact3)

## Regression tree:
## rpart(formula = FactorScore3 ~ Rank + Age + TIS + BS + Certs + Experience + MS + OPExperience, cp = 0.006)
##
## Variables actually used in tree construction:
## [1] Certs Experience Rank TIS
##
## Root node error: 77.213/134 = 0.57622
##
## n= 134
# CP nsplit rel error xerror xstd
# 1 0.1319868 0 1.00000 1.01957 0.36495
# 2 0.1015174 1 0.86801 1.04588 0.34135
# 3 0.0313155 2 0.76650 1.06879 0.37665
# 4 0.0277206 3 0.73518 1.05419 0.37626
# 5 0.0146558 4 0.70746 0.97559 0.33947
# 6 0.0133507 5 0.69280 0.98246 0.33928
# 7 0.0095105 6 0.67945 0.96975 0.33940
# 8 0.0060000 7 0.66994 0.95753 0.33977

rpart.plot(treeFact3, main='TREE Factor 3')
## 

### [1] 1.399393

#### (Mean of the difference between predicted and actual)

```
mean(resultsdfFact1$diff)
```

### [1] -0.2620682

#### (Median of the difference between predicted and actual)

```
median(resultsdfFact1$diff)
```

### [1] 0.1728327

Second Model (Factor 2):

```
resultsdfFact2 <- data.frame(
  Actual = testdata1$FactorScore2,
  Prediction = predict(treeFact2, newdata = testdata1)
)
resultsdfFact2$diff <- resultsdfFact2$Actual - resultsdfFact2$Prediction
```

### [1] 1.08705

#### (Standard Deviation of the difference between predicted and actual)

```
sd(resultsdfFact2$diff)
```

### [1] -0.1821727

#### (Mean of the difference between predicted and actual)

```
mean(resultsdfFact2$diff)
```

### [1] 0.1613028

Third Model (Factor 3):

```
resultsdfFact3 <- data.frame(
  Actual = testdata1$FactorScore3,
  Prediction = predict(treeFact3, newdata = testdata1)
)
resultsdfFact3$diff <- resultsdfFact3$Actual - resultsdfFact3$Prediction
```

### [1] 1.171089

#### (Standard Deviation of the difference between predicted and actual)

```
sd(resultsdfFact3$diff)
```

### [1] -0.1375787

#### (Mean of the difference between predicted and actual)

```
mean(resultsdfFact3$diff)
```

### [1] 0.1100923

Final thoughts: the models don’t seem too bad, but I wouldn’t promote/demote anyone based on the results. They could be a good start, but the trees are a little overfit (not something that you have to mention, but
you could say that more data is needed for a better model).
SUPPLEMENTAL TO THESIS
(4 OF 10: TRAC-MONTEREY 17A SURVEY ANALYSIS (SENTIMENT_TEXT ANALYSIS))

RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
Overview

The purpose of this write-up is to document the text analysis support provided by the author to the 17A Survey Thesis team. We include all of the code needed to support future reproduction of the analysis and encourage future research to move the ball further forward. For brevity this write-up only covers the analysis of Question 38, the results for Question 39 can easily be obtained by substituting variable names.

Setting up the R Environment

Here we load all the required packages. Note that in this instance the rJava package requires us to point to a specific jre file - the individual system requirements to implement this step will vary. We also set the random seed to support reproduction of several of the network graph we will generate later.

```r
library(tibble)
library(tm)
library(wordcloud)
library(dplyr)
library(RColorBrewer)
library(widyr)
library(tidytext)

library(ggplot2)
library(igraph)
library(ggforce)
library(ggraph)

options(java.home="C:\Program Files\Java\jdk1.8.0_91\jre")
library(rJava)
library(ngram)
library(qdap)

set.seed(1234)
```

Loading and subsetting data

Here we load the data from a pre-formatted csv. The 17A Survey Thesis team previously conduct manual coding of each response to the questions of interest (Q38 and Q39) to identify the response as dealing with Selection, Recruitment, or None. We then subset the full data set to those response corresponding to each of the coded labels for Question 38.

```r
data <- read.csv("C:\Users\nparker\Desktop\19A\19A_data.csv",
stringsAsFactors = FALSE)

###Subset the data
Pre-processing the text responses.

Here we define a function that will execute our pre-processing of the dataframe object into the Corpus object that the `wordcloud` package expects as input. As part of the pre-processing we first transform all characters to lower case, remove punctuation, remove extra whitespace, and remove the stopwords. In addition to the standard SMART and english stopword sets we also remove the following words: technical, process, and branch. Early exploratory analysis indicated that these words were generically structural within the dataset and their processes only introduced unnecessary noise in the dataset.

The function takes a concatenated vector of the responses of interest, the column names of the vectors, and any additional stopwords. The output is a term-document matrix (tdm).

```r
extraStopwords <- c("technical", "process", "branch")
corpFunction <- function(input_data, c_list, extraStopwords){
  c <- VCorpus(VectorSource(input_data))
  c <- tm_map(c, content_transformer(tolower))
  c <- tm_map(c, removePunctuation)
  c <- tm_map(c, content_transformer(stripWhitespace))
  c <- tm_map(c, removeWords, tm::stopwords("SMART"))
  c <- tm_map(c, removeWords, tm::stopwords("english"))
  c <- tm_map(c, removeWords, extraStopwords)
  c <- tm_map(c, PlainTextDocument)
  cloud.tdm = TermDocumentMatrix(c)
  cloud.tdm.m = as.matrix(cloud.tdm)
  colnames(cloud.tdm.m) <- c_list
  return(cloud.tdm.m)
}
```

Generating Comparison and Commonality Wordclouds.

Here we use the `wordcloud` package to generate the comparison and commonality wordclouds.

```r
cloud.38 <- c(sel.38.data, recruit.38.data)
c_list.38 <- c("Selection", "Recruiting")
```
tdm.38.m <- corpFunction(cloud.38, c_list.38, extraStopwords)

comp.38 <- comparison.cloud(tdm.38.m, min.words = 4, max.words = 50, random.order= FALSE, title.size = 2, rot.per = 0)

common.38 <- commonality.cloud(tdm.38.m, min.words = 4, max.words = 15, random.order= FALSE)
These visualization give a qualitative sense of the data but lack some of the quantitative data that some may find helpful. To support this requirement we also generate bar graphs of the most common words.

```r
sel.38.d <- data.frame(word = rownames(tdm.38.m), freq=tdm.38.m[,1])

sel.38.d <- sel.38.d[order(-sel.38.d$freq),]

rec.38.d <- data.frame(word = rownames(tdm.38.m), freq=tdm.38.m[,2])

rec.38.d <- rec.38.d[order(-rec.38.d$freq),]

```
Most frequent words – Q38 – Selection

rec.38.bar <- barplot(rec.38.d[1:15,]$freq, las = 2, names.arg = rec.38.d[1:15,]$word, col = "lightblue", main = "Most frequent words- Q38 - Recruiting", ylab = "Word frequencies")
Correlation and co-occurrence

We now leverage so work conducted by LTC Melanie Vinton to visualize the co-occurrence and correlation of words within responses. His full documentation is published at http://dcoe.army.mil/files/9415/0170/2779/text_correlation.html.

We first develop two functions (one for co-occurrence and one for correlation) to process that data into the format our visualization calls expect. Here we use the tidytext package to replicate the same pre-processing steps that we used the tm package to complete in the wordcloud functions above.

Both functions take the following as inputs: a single vector of freetext responses, the base set of stopwords, and any extra stopwords. The correlation function also includes a cutoff value as an input. Note that now we are using all responses to Question 38 that had either a Selection or Recruiting tag.

cococurrenceFunction <- function(input_data, stops, extraStopwords){
  d <- input_data
  d$id <- seq(1,nrow(d))
  d.tidy <- unnest_tokens(tbl = d, output = word, input = Response)
  d.tidy <- filter(d.tidy, word != "")
  d.tidy <- dplyr::anti_join(d.tidy, stops)
```r
d.tidy <- dplyr::anti_join(d.tidy, data.frame(word = extraStopwords))

d.tidy.pairs <- pairwise_count(d.tidy, word, id, sort = TRUE, upper = FALSE)

return(d.tidy.pairs)
}

corFunction <- function(input_data, stops, extraStopwords, cutoff){
  d <- input_data
  d$id <- seq(1,nrow(d))
  d.tidy <- unnest_tokens(tbl = d, output = word, input = Response)
  d.tidy <- filter(d.tidy, word != "")
  d.tidy <- dplyr::anti_join(d.tidy, stops)
  d.tidy <- dplyr::anti_join(d.tidy, data.frame(word = extraStopwords))
  text_cor <- filter(group_by(d.tidy, word), n() > cutoff)
  text_cor2 <- pairwise_cor(text_cor, word, id, sort = TRUE, upper = FALSE)
  return(text_cor2)
}

### Co-occurance Q38
Q38 <- rbind(sel.38.data, recruit.38.data)

Q38.pairs <- cooccuranceFunction(Q38, stop_words, extraStopwords)

## Joining, by = "word"
## Joining, by = "word"
Q38.pairs.2 <- filter(Q38.pairs, n > 2)

Q38.g <- graph_from_data_frame(Q38.pairs.2)

Q38.g2 <- ggraph(Q38.g, layout = "fr") +
  geom_edge_link(aes(edge_alpha = n, edge_width = n), edge_colour = "cyan4") +
  geom_node_point(size = 5) +
  geom_node_text(aes(label = name), repel = TRUE,
                 point.padding = unit(0.2, "lines")) +
  ggtitle("Word pairings in Q38 - Selection and Recruiting Tagged Responses")+
  theme_void()

Q38.g2
```
Word pairings in Q38 – Selection and Recruiting Tagged Responses

###Q38 correlation

```r
Q38.cor <- corFunction(Q38, stop_words, extraStopwords, 7)

## Joining, by = "word"
## Joining, by = "word"
Q38.cor2 <- filter(Q38.cor, correlation > 0.1)

Q38.c <- graph_from_data_frame(Q38.cor2)

Q38.c2 <- ggraph(Q38.c, layout = "fr") +
  geom_edge_link(aes(edge_width = correlation, edge_alpha = correlation),
                edge_colour = "royalblue") +
  geom_node_point(size = 5) +
  geom_node_text(aes(label = name), repel = TRUE,
                 point.padding = unit(0.2, "lines")) +
  ggtitle("Word Correlations for Q38 - Recruiting and Selection Tags") +
  theme_void()

Q38.c2
```
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Age</th>
<th>n</th>
<th>Age Group</th>
<th>n</th>
<th>Average Age by Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1LT</td>
<td>23</td>
<td>1</td>
<td>21</td>
<td>2</td>
<td>24.97</td>
</tr>
<tr>
<td>1LT</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>11</td>
<td>25.46</td>
</tr>
<tr>
<td>1LT</td>
<td>25</td>
<td>6</td>
<td>24</td>
<td>14</td>
<td>23.7</td>
</tr>
<tr>
<td>1LT</td>
<td>27</td>
<td>5</td>
<td>25</td>
<td>13</td>
<td>24.7</td>
</tr>
<tr>
<td>1LT</td>
<td>28</td>
<td>2</td>
<td>26</td>
<td>13</td>
<td>25.3</td>
</tr>
<tr>
<td>1LT</td>
<td>30</td>
<td>1</td>
<td>27</td>
<td>11</td>
<td>25.4</td>
</tr>
<tr>
<td>1LT</td>
<td>31</td>
<td>1</td>
<td>28</td>
<td>18</td>
<td>24.5</td>
</tr>
<tr>
<td>2LT</td>
<td>21</td>
<td>2</td>
<td>29</td>
<td>15</td>
<td>24.2</td>
</tr>
<tr>
<td>2LT</td>
<td>22</td>
<td>11</td>
<td>30</td>
<td>15</td>
<td>25.4</td>
</tr>
<tr>
<td>2LT</td>
<td>23</td>
<td>6</td>
<td>31</td>
<td>20</td>
<td>24.7</td>
</tr>
<tr>
<td>2LT</td>
<td>24</td>
<td>1</td>
<td>32</td>
<td>10</td>
<td>25.4</td>
</tr>
<tr>
<td>2LT</td>
<td>25</td>
<td>3</td>
<td>33</td>
<td>25</td>
<td>24.9</td>
</tr>
<tr>
<td>2LT</td>
<td>26</td>
<td>2</td>
<td>34</td>
<td>8</td>
<td>24.6</td>
</tr>
<tr>
<td>2LT</td>
<td>27</td>
<td>1</td>
<td>35</td>
<td>26</td>
<td>25.1</td>
</tr>
<tr>
<td>2LT</td>
<td>28</td>
<td>2</td>
<td>36</td>
<td>13</td>
<td>25.3</td>
</tr>
<tr>
<td>2LT</td>
<td>29</td>
<td>3</td>
<td>37</td>
<td>18</td>
<td>25.5</td>
</tr>
<tr>
<td>2LT</td>
<td>30</td>
<td>2</td>
<td>38</td>
<td>24</td>
<td>25.7</td>
</tr>
<tr>
<td>2LT</td>
<td>31</td>
<td>1</td>
<td>39</td>
<td>15</td>
<td>25.5</td>
</tr>
<tr>
<td>2LT</td>
<td>32</td>
<td>1</td>
<td>40</td>
<td>19</td>
<td>25.5</td>
</tr>
<tr>
<td>2LT</td>
<td>33</td>
<td>1</td>
<td>41</td>
<td>11</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>26</td>
<td>5</td>
<td>42</td>
<td>10</td>
<td>25.4</td>
</tr>
<tr>
<td>CPT</td>
<td>27</td>
<td>5</td>
<td>43</td>
<td>9</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>28</td>
<td>14</td>
<td>44</td>
<td>9</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>29</td>
<td>12</td>
<td>45</td>
<td>10</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>30</td>
<td>12</td>
<td>46</td>
<td>3</td>
<td>25.3</td>
</tr>
<tr>
<td>CPT</td>
<td>31</td>
<td>18</td>
<td>47</td>
<td>6</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>32</td>
<td>13</td>
<td>48</td>
<td>3</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>33</td>
<td>9</td>
<td>50</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>34</td>
<td>4</td>
<td>51</td>
<td>2</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>35</td>
<td>9</td>
<td>52</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>36</td>
<td>3</td>
<td>53</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>37</td>
<td>4</td>
<td>54</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>38</td>
<td>5</td>
<td>55</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>39</td>
<td>4</td>
<td>56</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>40</td>
<td>5</td>
<td>57</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>41</td>
<td>1</td>
<td>58</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>CPT</td>
<td>44</td>
<td>4</td>
<td>59</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>32</td>
<td>6</td>
<td>60</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>33</td>
<td>15</td>
<td>61</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>34</td>
<td>4</td>
<td>62</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>35</td>
<td>17</td>
<td>63</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>36</td>
<td>10</td>
<td>64</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>37</td>
<td>14</td>
<td>65</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>38</td>
<td>19</td>
<td>66</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>39</td>
<td>8</td>
<td>67</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>40</td>
<td>9</td>
<td>68</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>41</td>
<td>4</td>
<td>69</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>42</td>
<td>6</td>
<td>70</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>43</td>
<td>3</td>
<td>71</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>44</td>
<td>2</td>
<td>72</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>MAJ</td>
<td>45</td>
<td>1</td>
<td>73</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>39</td>
<td>3</td>
<td>74</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>40</td>
<td>5</td>
<td>75</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>41</td>
<td>6</td>
<td>76</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>42</td>
<td>4</td>
<td>77</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>43</td>
<td>6</td>
<td>78</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>44</td>
<td>3</td>
<td>79</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>45</td>
<td>8</td>
<td>80</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>46</td>
<td>3</td>
<td>81</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>47</td>
<td>6</td>
<td>82</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>48</td>
<td>2</td>
<td>83</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>50</td>
<td>1</td>
<td>84</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>51</td>
<td>2</td>
<td>85</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>52</td>
<td>1</td>
<td>86</td>
<td>1</td>
<td>25.5</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>53</td>
<td>1</td>
<td>87</td>
<td>1</td>
<td>25.5</td>
</tr>
</tbody>
</table>

**Average Age Total**: 34.08

**Age Ranges**

<table>
<thead>
<tr>
<th>Age Ranges</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-24</td>
<td>34</td>
</tr>
<tr>
<td>25-31</td>
<td>105</td>
</tr>
<tr>
<td>32-38</td>
<td>134</td>
</tr>
<tr>
<td>39-45</td>
<td>83</td>
</tr>
<tr>
<td>45+</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>373</td>
</tr>
</tbody>
</table>

**Average Age by Rank**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Average Age by Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1LT</td>
<td>24.97</td>
</tr>
<tr>
<td>2LT</td>
<td>25.46</td>
</tr>
<tr>
<td>CPT</td>
<td>23.7</td>
</tr>
<tr>
<td>MAJ</td>
<td>37.1</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>44.20</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Average Age by Gender**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>2LT</td>
<td>25.06</td>
<td>25.13</td>
</tr>
<tr>
<td>1LT</td>
<td>25.34</td>
<td>25.60</td>
</tr>
<tr>
<td>CPT</td>
<td>32.48</td>
<td>32.42</td>
</tr>
<tr>
<td>MAJ</td>
<td>37.11</td>
<td>37.19</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>44.20</td>
<td>47.00</td>
</tr>
<tr>
<td></td>
<td>34.08</td>
<td>34.42</td>
</tr>
</tbody>
</table>
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
<table>
<thead>
<tr>
<th>RANK_PDE</th>
<th>PN_SEX_CD</th>
<th>Number</th>
<th>Percent Total</th>
<th>Percent Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1LT</td>
<td>F</td>
<td>6</td>
<td>1.609%</td>
<td>15.38%</td>
</tr>
<tr>
<td>1LT</td>
<td>M</td>
<td>33</td>
<td>8.847%</td>
<td>84.62%</td>
</tr>
<tr>
<td>2LT</td>
<td>F</td>
<td>4</td>
<td>1.072%</td>
<td>11.11%</td>
</tr>
<tr>
<td>2LT</td>
<td>M</td>
<td>32</td>
<td>8.579%</td>
<td>88.89%</td>
</tr>
<tr>
<td>CPT</td>
<td>F</td>
<td>11</td>
<td>2.949%</td>
<td>8.59%</td>
</tr>
<tr>
<td>CPT</td>
<td>M</td>
<td>117</td>
<td>31.367%</td>
<td>91.41%</td>
</tr>
<tr>
<td>MAJ</td>
<td>F</td>
<td>16</td>
<td>4.290%</td>
<td>13.45%</td>
</tr>
<tr>
<td>MAJ</td>
<td>M</td>
<td>103</td>
<td>27.614%</td>
<td>86.55%</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>F</td>
<td>1</td>
<td>0.268%</td>
<td>1.96%</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>M</td>
<td>50</td>
<td>13.405%</td>
<td>98.04%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>373</td>
<td>100.000%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ASG_UNT_MJR_CMD_CD</th>
<th>PN_SEX_CD</th>
<th>n</th>
<th>CMD Percentage</th>
<th>Total Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCYBER</td>
<td>F</td>
<td>2</td>
<td>8.70%</td>
<td>0.54%</td>
</tr>
<tr>
<td>ARCYBER</td>
<td>M</td>
<td>21</td>
<td>91.30%</td>
<td>5.63%</td>
</tr>
<tr>
<td>USARCENT</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>USARAF</td>
<td>M</td>
<td>2</td>
<td>100.00%</td>
<td>0.54%</td>
</tr>
<tr>
<td>USINSCOM</td>
<td>F</td>
<td>12</td>
<td>19.35%</td>
<td>3.22%</td>
</tr>
<tr>
<td>USINSCOM</td>
<td>M</td>
<td>50</td>
<td>80.65%</td>
<td>13.40%</td>
</tr>
<tr>
<td>USATEC</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>USAIMCOM</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>ODD AGENCIES</td>
<td>M</td>
<td>3</td>
<td>100.00%</td>
<td>0.80%</td>
</tr>
<tr>
<td>SOC COM</td>
<td>M</td>
<td>2</td>
<td>100.00%</td>
<td>0.54%</td>
</tr>
<tr>
<td>DIRECTED MIL OVERSTRENGTH</td>
<td>M</td>
<td>4</td>
<td>100.00%</td>
<td>1.07%</td>
</tr>
<tr>
<td>USAREUR</td>
<td>M</td>
<td>6</td>
<td>100.00%</td>
<td>1.61%</td>
</tr>
<tr>
<td>21ST TSC</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>FORSCOM</td>
<td>F</td>
<td>2</td>
<td>6.45%</td>
<td>0.54%</td>
</tr>
<tr>
<td>FORSCOM</td>
<td>M</td>
<td>29</td>
<td>93.55%</td>
<td>7.77%</td>
</tr>
<tr>
<td>20TH SPT</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>NETCOM</td>
<td>F</td>
<td>11</td>
<td>10.68%</td>
<td>2.95%</td>
</tr>
<tr>
<td>NETCOM</td>
<td>M</td>
<td>92</td>
<td>89.32%</td>
<td>24.66%</td>
</tr>
<tr>
<td>JOINT ACTIVITIES (LESS NATO)</td>
<td>F</td>
<td>1</td>
<td>8.33%</td>
<td>0.27%</td>
</tr>
<tr>
<td>JOINT ACTIVITIES (LESS NATO)</td>
<td>M</td>
<td>11</td>
<td>91.67%</td>
<td>2.95%</td>
</tr>
<tr>
<td>USMA</td>
<td>F</td>
<td>2</td>
<td>7.69%</td>
<td>0.54%</td>
</tr>
<tr>
<td>USMA</td>
<td>M</td>
<td>24</td>
<td>92.31%</td>
<td>6.43%</td>
</tr>
<tr>
<td>USARPAC</td>
<td>F</td>
<td>1</td>
<td>8.33%</td>
<td>0.27%</td>
</tr>
<tr>
<td>USARPAC</td>
<td>M</td>
<td>11</td>
<td>91.67%</td>
<td>2.95%</td>
</tr>
<tr>
<td>CSA</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>FEDERAL AGENCIES</td>
<td>M</td>
<td>10</td>
<td>100.00%</td>
<td>2.68%</td>
</tr>
<tr>
<td>USASOC</td>
<td>M</td>
<td>7</td>
<td>100.00%</td>
<td>1.88%</td>
</tr>
<tr>
<td>TRADOC</td>
<td>F</td>
<td>7</td>
<td>11.67%</td>
<td>1.88%</td>
</tr>
<tr>
<td>TRADOC</td>
<td>M</td>
<td>53</td>
<td>88.33%</td>
<td>14.21%</td>
</tr>
<tr>
<td>US ARMY WAR COLLEGE</td>
<td>M</td>
<td>2</td>
<td>100.00%</td>
<td>0.54%</td>
</tr>
<tr>
<td>US ARMY SECURITY ASSISTANCE CENTER</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>ZZZ</td>
<td>M</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>373</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Females</td>
<td>Total Males</td>
<td>RACE_CD</td>
<td>PN_SEX_CD</td>
<td>n</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td>----------------------------------------------</td>
<td>-----------</td>
<td>---</td>
</tr>
<tr>
<td>38</td>
<td>335</td>
<td>American Indian/Alaskan Native</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>American Indian/Alaskan Native</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asian</td>
<td>F</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Black or African American</td>
<td>F</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Native Hawaiian or other Pacific Islander</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>White</td>
<td>F</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UNKNOWN</td>
<td>F</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DOD Officer Percentage</th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Army Officer Percentage</td>
<td>F</td>
<td>M</td>
</tr>
</tbody>
</table>

| 373 | 373 | 17.0% | 83.0% |
|     |     | 16.8% | 83.2% |
NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

SUPPLEMENTAL TO THESIS
(7 OF 10: PDE DATA_GENDERACE_APP)

RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by
Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
## RACE and SEX By Race Percentage

<table>
<thead>
<tr>
<th>RACE_CD</th>
<th>PN_SEX_CD</th>
<th>n</th>
<th>By Race Percentage</th>
<th>Total Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Indian/Alaskan Native</td>
<td>F</td>
<td>1</td>
<td>50.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
<td>M</td>
<td>1</td>
<td>50.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>Asian</td>
<td>F</td>
<td>3</td>
<td>9.09%</td>
<td>0.80%</td>
</tr>
<tr>
<td>Asian</td>
<td>M</td>
<td>30</td>
<td>90.91%</td>
<td>8.04%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>F</td>
<td>9</td>
<td>23.08%</td>
<td>2.41%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>M</td>
<td>30</td>
<td>76.92%</td>
<td>8.04%</td>
</tr>
<tr>
<td>Native Hawaiian or other Pacific Islander</td>
<td>F</td>
<td>1</td>
<td>100.00%</td>
<td>0.27%</td>
</tr>
<tr>
<td>White</td>
<td>F</td>
<td>20</td>
<td>7.27%</td>
<td>5.36%</td>
</tr>
<tr>
<td>White</td>
<td>M</td>
<td>255</td>
<td>92.73%</td>
<td>68.36%</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>F</td>
<td>4</td>
<td>17.39%</td>
<td>1.07%</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>M</td>
<td>19</td>
<td>82.61%</td>
<td>5.09%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>373</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### DOD Officer Percentage

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOD Officer</td>
<td>17.0%</td>
<td>83.0%</td>
</tr>
</tbody>
</table>

### Army Officer Percentage

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>Army Officer</td>
<td>16.8%</td>
<td>83.2%</td>
</tr>
</tbody>
</table>
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
<table>
<thead>
<tr>
<th>AGENT_MJR_CMD_CD</th>
<th>PN SEX CD</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCYBER</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>ARCYBER</td>
<td>M</td>
<td>21</td>
</tr>
<tr>
<td>USARCENT</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>USARAF</td>
<td>M</td>
<td>2</td>
</tr>
<tr>
<td>USINSCOM</td>
<td>F</td>
<td>12</td>
</tr>
<tr>
<td>USINSCOM</td>
<td>M</td>
<td>50</td>
</tr>
<tr>
<td>USATEC</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>USAIMCOM</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>DOD AGENCIES</td>
<td>M</td>
<td>3</td>
</tr>
<tr>
<td>SOCCOM</td>
<td>M</td>
<td>2</td>
</tr>
<tr>
<td>DIRECTED MIL OVERSTRENGTH</td>
<td>M</td>
<td>4</td>
</tr>
<tr>
<td>USAREUR</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>21ST TSC</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>FORSCOM</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>FORSCOM</td>
<td>M</td>
<td>29</td>
</tr>
<tr>
<td>20TH SPT</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>NETCOM</td>
<td>F</td>
<td>11</td>
</tr>
<tr>
<td>NETCOM</td>
<td>M</td>
<td>92</td>
</tr>
<tr>
<td>JOINT ACTIVITIES (LESS NATO)</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>JOINT ACTIVITIES (LESS NATO)</td>
<td>M</td>
<td>11</td>
</tr>
<tr>
<td>USMA</td>
<td>F</td>
<td>2</td>
</tr>
<tr>
<td>USMA</td>
<td>M</td>
<td>24</td>
</tr>
<tr>
<td>USARPAC</td>
<td>F</td>
<td>1</td>
</tr>
<tr>
<td>USARPAC</td>
<td>M</td>
<td>11</td>
</tr>
<tr>
<td>OSA</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>FEDERAL AGENCIES</td>
<td>M</td>
<td>10</td>
</tr>
<tr>
<td>USASOC</td>
<td>M</td>
<td>7</td>
</tr>
<tr>
<td>TRADOC</td>
<td>F</td>
<td>7</td>
</tr>
<tr>
<td>TRADOC</td>
<td>M</td>
<td>53</td>
</tr>
<tr>
<td>US ARMY WAR COLLEGE</td>
<td>M</td>
<td>2</td>
</tr>
<tr>
<td>US ARMY SECURITY ASSISTANCE CENTER</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>ZZZ</td>
<td>M</td>
<td>1</td>
</tr>
</tbody>
</table>
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins
Alejandro Hernandez

Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
<table>
<thead>
<tr>
<th>ASG_UNIT_MJR_CMD_CD</th>
<th>n</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARCYBER</td>
<td>223</td>
<td>5.17%</td>
</tr>
<tr>
<td>USARENT</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>USARAF</td>
<td>2</td>
<td>0.54%</td>
</tr>
<tr>
<td>USINSCOM</td>
<td>62</td>
<td>15.64%</td>
</tr>
<tr>
<td>USATEC</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>USARCOM</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>DOD AGENCIES</td>
<td>3</td>
<td>0.80%</td>
</tr>
<tr>
<td>USOM</td>
<td>2</td>
<td>0.54%</td>
</tr>
<tr>
<td>DIRECTED MIL OVERSTRENGTH</td>
<td>4</td>
<td>1.07%</td>
</tr>
<tr>
<td>USARFOR</td>
<td>4</td>
<td>1.07%</td>
</tr>
<tr>
<td>21ST TSC</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>FORSCOM</td>
<td>31</td>
<td>8.31%</td>
</tr>
<tr>
<td>20TH SPT</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>NETCOM</td>
<td>105</td>
<td>27.64%</td>
</tr>
<tr>
<td>JOINT ACTIVITIES (LESS NATO)</td>
<td>12</td>
<td>3.22%</td>
</tr>
<tr>
<td>USAR</td>
<td>26</td>
<td>6.97%</td>
</tr>
<tr>
<td>USARPAF</td>
<td>12</td>
<td>3.22%</td>
</tr>
<tr>
<td>USA</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>DoD AGENCIES</td>
<td>10</td>
<td>2.68%</td>
</tr>
<tr>
<td>USASOC</td>
<td>17</td>
<td>4.50%</td>
</tr>
<tr>
<td>TRADOC</td>
<td>60</td>
<td>16.09%</td>
</tr>
<tr>
<td>US ARMY WAR COLLEGE</td>
<td>2</td>
<td>0.54%</td>
</tr>
<tr>
<td>US ARMY SECURITY ASSISTANCE CENTER</td>
<td>1</td>
<td>0.27%</td>
</tr>
<tr>
<td>UNKNOWN</td>
<td>1</td>
<td>0.27%</td>
</tr>
</tbody>
</table>

17A Assignment proportion by Major Command
RECRUITING THE CYBER LEADER: AN EVALUATION OF THE HUMAN RESOURCE MODEL USED FOR RECRUITING THE ARMY’S “CYBER OPERATIONS OFFICER”

by

Wallace C. Nicholson
Sean A. Gibbs

September 2017

Thesis Co-Advisors: Steve Mullins Alejandro Hernandez
Second Reader: William Hatch

Approved for public release. Distribution is unlimited.
<table>
<thead>
<tr>
<th>RANK, POS</th>
<th>EDU_LVL, CD</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1LT</td>
<td>Baccalaureate degree</td>
<td>34</td>
<td>9.115%</td>
</tr>
<tr>
<td>1LT</td>
<td>Masters degree</td>
<td>4</td>
<td>1.072%</td>
</tr>
<tr>
<td>1LT</td>
<td>Doctorate degree</td>
<td>1</td>
<td>0.268%</td>
</tr>
<tr>
<td>1LT</td>
<td>Completed one semester of college, no high school diploma</td>
<td>3</td>
<td>0.804%</td>
</tr>
<tr>
<td>2LT</td>
<td>Baccalaureate degree</td>
<td>33</td>
<td>8.311%</td>
</tr>
<tr>
<td>2LT</td>
<td>Masters degree</td>
<td>2</td>
<td>0.536%</td>
</tr>
<tr>
<td>CPT</td>
<td>Baccalaureate degree</td>
<td>89</td>
<td>23.861%</td>
</tr>
<tr>
<td>CPT</td>
<td>Masters degree</td>
<td>38</td>
<td>10.188%</td>
</tr>
<tr>
<td>CPT</td>
<td>Doctorate degree</td>
<td>1</td>
<td>0.268%</td>
</tr>
<tr>
<td>MA</td>
<td>Baccalaureate degree</td>
<td>42</td>
<td>11.260%</td>
</tr>
<tr>
<td>MA</td>
<td>Masters degree</td>
<td>73</td>
<td>19.571%</td>
</tr>
<tr>
<td>MA</td>
<td>Doctorate degree</td>
<td>4</td>
<td>1.072%</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>Baccalaureate degree</td>
<td>7</td>
<td>1.877%</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>Masters degree</td>
<td>33</td>
<td>8.847%</td>
</tr>
<tr>
<td>COL/LTC</td>
<td>Doctorate degree</td>
<td>13</td>
<td>3.490%</td>
</tr>
</tbody>
</table>