Report

KK

2019 7 13

0. Data Loading

```
JOB = read.csv("D:\\Dropbox\\DATA SET\\new-york-city-current-job-postings\\nyc-jobs.csv"
,stringsAsFactors = FALSE)

JOB$Posting.Date = as.Date(JOB$Posting.Date, format = "%Y-%m-%d %H:%M")
summary(JOB$Posting.Date)
```

```
##         Min.      1st Qu.       Median         Mean      3rd Qu. 
## "2011-06-24" "2017-03-24" "2018-05-31" "2017-12-16" "2018-10-18" 
##         Max.         NA's 
## "2019-07-15"         "17" 
```

```
JOB$Post..Until = as.Date(JOB$Post..Until, format = "%Y-%m-%d")
```

1. Library

```
library(ggplot2)
library(dplyr)
library(reshape)
library(tseries)
library(tm)
library(stringr)
library(qgraph)
```

2. Exploratory Data Analysis

```
format(JOB$Posting.Date[1],"%y")
```

```
## [1] "11"
```

```
format(JOB$Posting.Date[1],"%m")
```

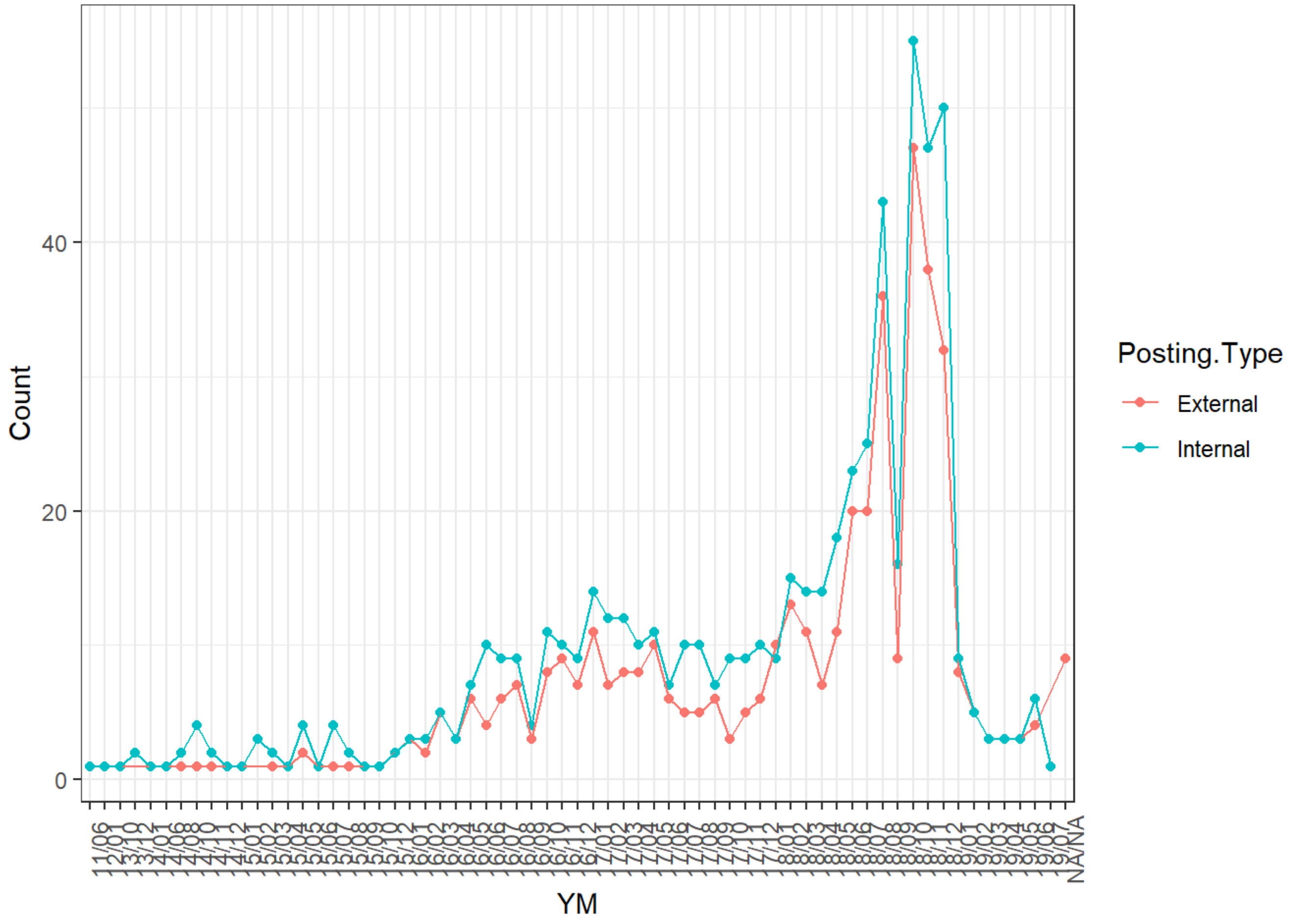
```
## [1] "06"
```

```
JOB$YM = paste(format(JOB$Posting.Date,"%y"),
               format(JOB$Posting.Date,"%m"), sep="/")
```

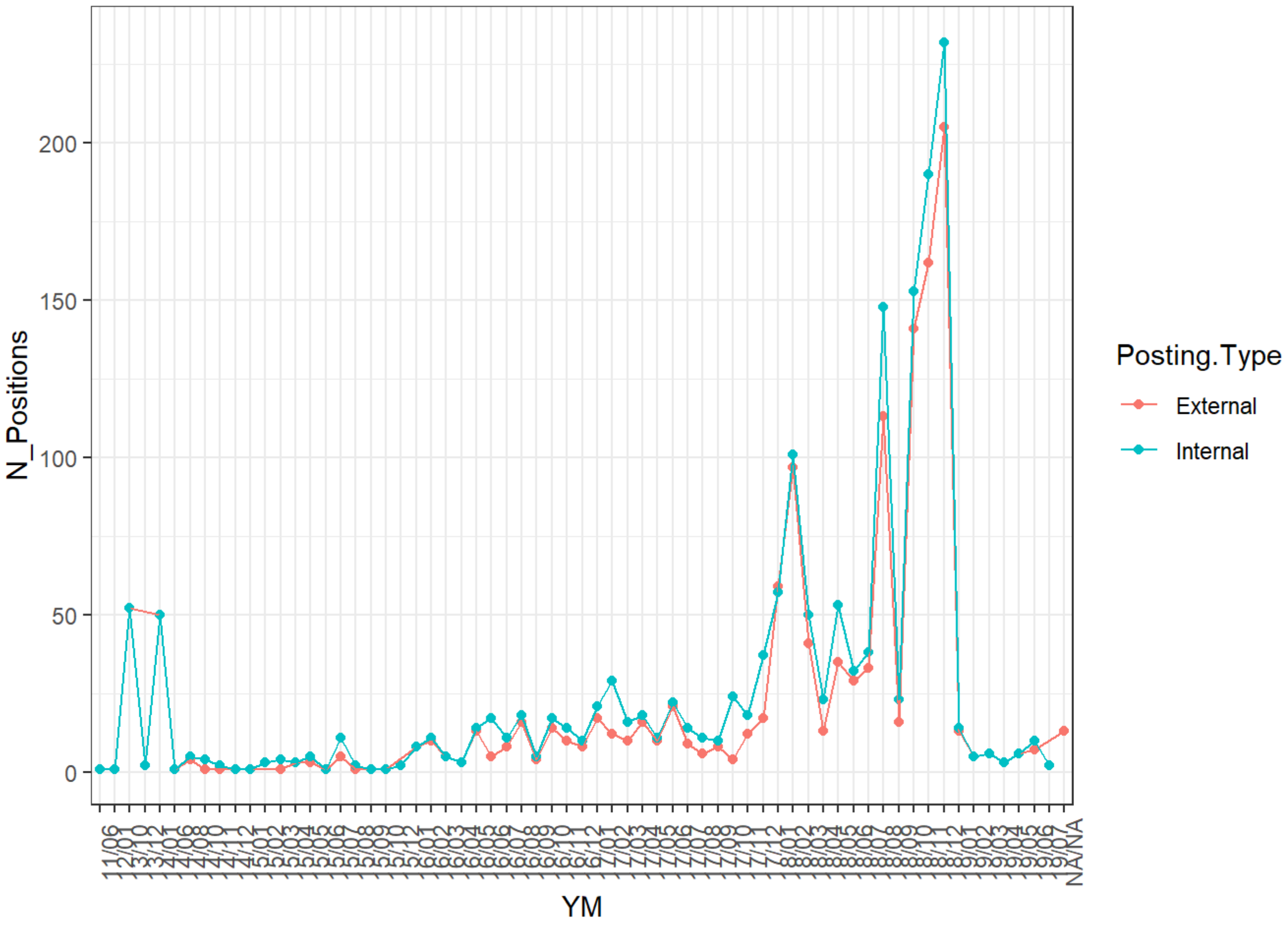
```
summary(as.factor(JOB$Posting.Type))
```

```
## External Internal
##      447      609
```

```
# summary(as.factor(JOB$Business.Title))

JOB_YM  = JOB %>%
  group_by(YM,Posting.Type) %>%
  summarise(Count = length(Posting.Date),
            N_Positions = sum(X..Of.Positions, na.rm = TRUE)) %>%
  arrange(YM)



ggplot(JOB_YM[-nrow(JOB_YM),]) +
  geom_point(aes(x = YM, y= Count,col = Posting.Type)) +
  geom_line(aes(x = YM, y = Count,col = Posting.Type, group = Posting.Type)) +
  # geom_point(aes(x = YM, y= N_Positions,col = Posting.Type)) +
   # geom_line(aes(x = YM, y = N_Positions,col = Posting.Type, group = Posting.Type)) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90))
```



```
ggplot(JOB_YM[-nrow(JOB_YM),]) +
  # geom_point(aes(x = YM, y= Count,col = Posting.Type)) +
  # geom_line(aes(x = YM, y = Count,col = Posting.Type, group = Posting.Type)) +
  geom_point(aes(x = YM, y= N_Positions,col = Posting.Type)) +
   geom_line(aes(x = YM, y = N_Positions,col = Posting.Type, group = Posting.Type)) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 90))
```



Category

```
length(levels(as.factor(JOB$Business.Title)))
```

```
## [1] 486
```

```
CORPUS = Corpus(VectorSource(JOB$Business.Title))
CORPUS_TM = tm_map(CORPUS, tolower)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS, tolower): transformation drops
## documents
```

```
CORPUS_TM = tm_map(CORPUS_TM, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_TM, removeNumbers): transformation
## drops documents
```

```
CORPUS_TM = tm_map(CORPUS_TM, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_TM, removePunctuation):
## transformation drops documents
```

```
CORPUS_TM = tm_map(CORPUS_TM, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_TM, stripWhitespace): transformation
## drops documents
```

```
CORPUS_TM = tm_map(CORPUS_TM, removeWords,
                    c(stopwords("english"),"my","custom","words"))
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_TM, removeWords,
## c(stopwords("english"), : transformation drops documents
```

```
TDM = TermDocumentMatrix(CORPUS_TM)

inspect(TDM)
```

```
## <<TermDocumentMatrix (terms: 498, documents: 1056)>>
## Non-/sparse entries: 3353/522535
## Sparsity           : 99%
## Maximal term length: 23
## Weighting          : term frequency (tf)
## Sample             :
##             Docs
## Terms        1042 1043 480 510 511 657 658 680 851 852
##   analyst       0    0   0   0   0   0   0   0   0   0
##   assistant     0    0   0   0   0   0   0   0   0   0
##   associate     0    0   0   0   0   1   1   0   0   0
##   civil         0    0   0   0   0   0   0   0   0   0
##   director      0    0   0   0   0   0   0   0   0   0
##   engineer      0    0   0   0   0   0   0   0   0   0
##   manager       0    0   0   0   0   0   0   0   0   0
##   project       0    0   0   0   0   0   0   0   0   0
##   senior        0    0   0   0   0   0   0   0   0   0
##   specialist    0    0   0   0   0   0   0   0   0   0
```

```
TDM = as.matrix(TDM)

D = rowSums(TDM)

sort(D[1:20],decreasing = TRUE)
```

```
##         analyst         manager       assistant        director 
##             130             129              80              69 
##       associate            unit       executive        business 
##              56              34              21              16 
##        contract          worker     maintenance     development 
##              14              11              10               8 
##       technical servicesheating         painter       temporary 
##               4               2               2               2 
##         account         chemist            cost      estimating 
##               1               1               1               1
```

- 직업 타이틀 이름 종류가 486개이므로 묶어주는 작업이 필요함
    1. Analysy (anal)
    2. Manage (manag)
    3. Assistance (assist)
    4. Director (dir)
    5. Execut (execut)
    6. Business (busi)
    7. Contract (contra)
    8. Maintenance (Maint)

```
str_detect(str = "ABCD" ,"A")
```

```
## [1] TRUE
```

```
Job_Class = function(x){

  if( str_detect(str = x , "anal")){

    y = "Analyst"
  }else if(str_detect(str = x , "manag")){

    y = "Management"
  }else if(str_detect(str = x , "assist")){

    y = "Assistance"
  }else if(str_detect(str = x , "dir")){

    y = "Director"

  }else if(str_detect(str = x , "exec")){

    y = "Execute"

  }else if(str_detect(str = x , "busi")){

    y = "Business"
  }else if(str_detect(str = x , "contra")){

    y = "Contract"
  }else if(str_detect(str = x , "Maint")){

    y = "Maintanenence"

  }else{

    y = "ETC"
  }
  return(y)
}


JOBS = rownames(TDM)

Job_Class(JOBS[1])
```

```
## [1] "ETC"
```

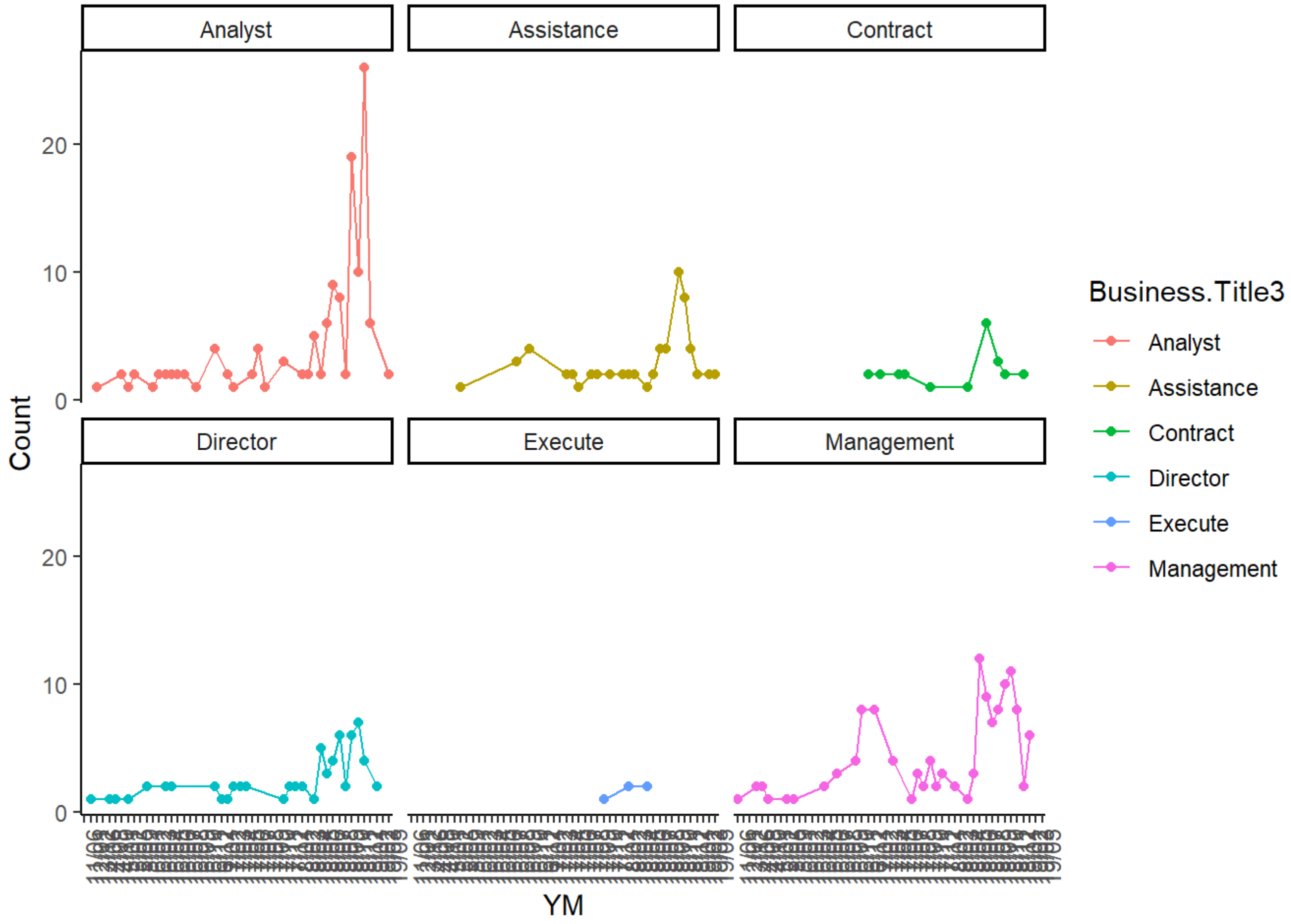
```
JOB$Business.Title2 = tolower(JOB$Business.Title)



Job_Class(JOB$Business.Title2)
```

```
## Warning in if (str_detect(str = x, "anal")) {: length > 1 이라는 조건이 있
## 고, 첫번째 요소만이 사용될 것입니다
## Warning in if (str_detect(str = x, "manag")) {: length > 1 이라는 조건이 있
## 고, 첫번째 요소만이 사용될 것입니다
```

```
## [1] "Management"
```

```
BT = c()
for( k in 1:nrow(JOB)){

  BT = c(BT, Job_Class(JOB$Business.Title2[k]))

}

JOB$Business.Title3 = BT
```

```
JOB %>%
  group_by(YM,Business.Title3) %>%
  summarise(Count = length(YM)) %>%
  filter(Business.Title3 != "ETC" ) %>%
  filter(!str_detect(YM, "NA")) %>%
  ggplot() +
  geom_point(aes(x = YM, y = Count, col = Business.Title3)) +
  geom_line(aes(x = YM, y = Count, col = Business.Title3, group = Business.Title3)) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 90)) +
  facet_wrap(~ Business.Title3)
```



- 요즘 인기가 많은 Analyst에 대해 분석을 진행

Exploratory Data Analysis 2

```
CORPUS_PS = Corpus(VectorSource(JOB$Preferred.Skills))

CORPUS_PS_TM = tm_map(CORPUS_PS, tolower)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_PS, tolower): transformation drops
## documents
```

```
CORPUS_PS_TM = tm_map(CORPUS_PS_TM, removeNumbers)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_PS_TM, removeNumbers): transformation
## drops documents
```

```
CORPUS_PS_TM = tm_map(CORPUS_PS_TM, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_PS_TM, removePunctuation):
## transformation drops documents
```

```
CORPUS_PS_TM = tm_map(CORPUS_PS_TM, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_PS_TM, stripWhitespace):
## transformation drops documents
```

```
CORPUS_PS_TM = tm_map(CORPUS_PS_TM, removeWords,
                    c(stopwords("english"),"my","custom","words"))
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_PS_TM, removeWords,
## c(stopwords("english"), : transformation drops documents
```

```
TDM = TermDocumentMatrix(CORPUS_PS_TM)
inspect(TDM)
```

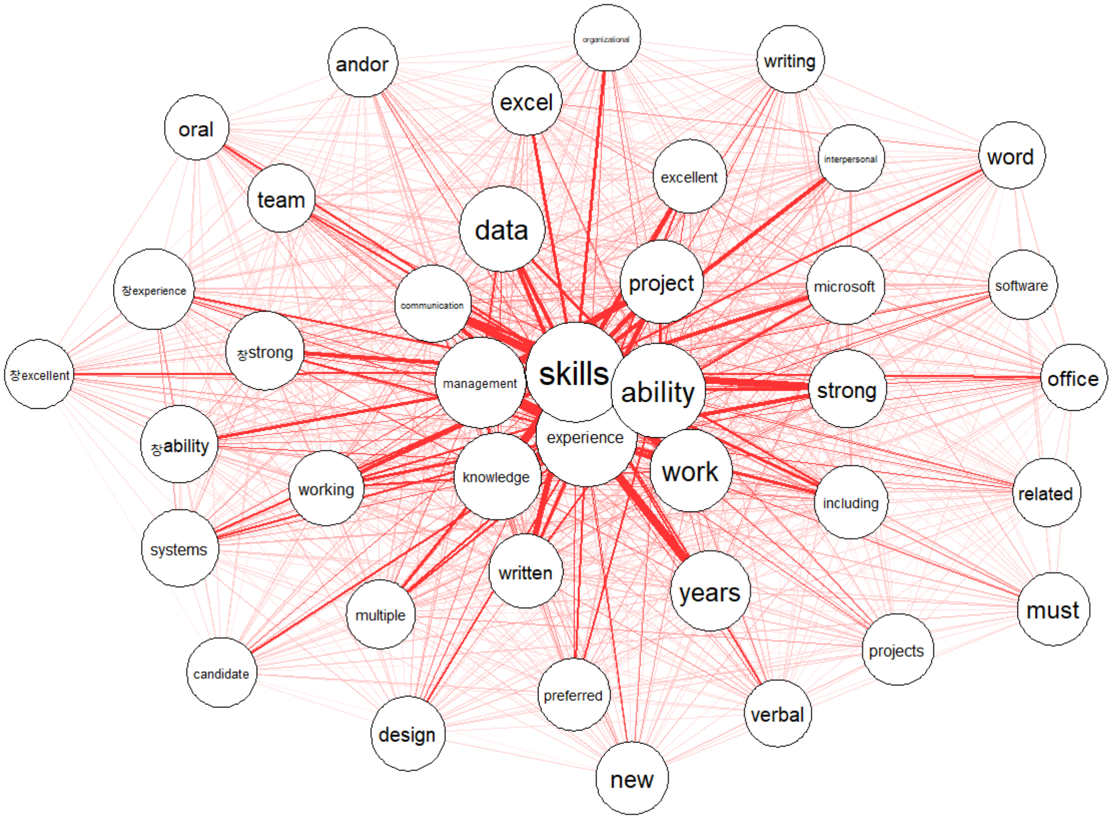
```
## <<TermDocumentMatrix (terms: 3607, documents: 1056)>>
## Non-/sparse entries: 39701/3769291
## Sparsity           : 99%
## Maximal term length: 27
## Weighting          : term frequency (tf)
## Sample             :
##                Docs
## Terms            112 113 31 32 486 545 546 698 879 905
##   ability          4   4  3  3   2   5   5   3   1   1
##   communication    1   1  0  0   1   2   2   0   0   0
##   experience       7   7  6  6   4   3   3   2   4   8
##   knowledge        1   1  4  4   0   3   3   5   1   1
##   management      11  11  0  0  12   0   0   0   4   1
##   microsoft        0   0  0  0   0   0   0   0   0   0
##   skills           5   5  3  3   2   3   3   4   3   1
##   work             2   2  1  1   2   4   4   4   1   0
##   written          0   0  0  0   1   2   2   1   0   0
##   years            0   0  0  0   0   1   1   0   0   0
```

```
TDM = as.matrix(TDM)

# TDM = as.data.frame(TDM)

# TDM$Words = rownames(TDM)

# TDM_G = TDM %>%
#   group_by(Words) %>%
#   summarise_each(funs(sum))
```

- One Hot Encoding

```
Convert_One = function(x){

  y = ifelse(x > 0 , 1, 0)
  return(y)
}

TDM_OHE = apply(TDM,2,Convert_One)
```

```
Word_Count = rowSums(TDM)
Word_Count <- order(Word_Count, decreasing=T)
word.freq1_J <- TDM[Word_Count[1:40],]
co.matrix1_J <- word.freq1_J %*% t(word.freq1_J)

qgraph(co.matrix1_J, labels=rownames(co.matrix1_J), diag=F,
       layout='spring', edge.color="#FF3333", vsize=log(diag(co.matrix1_J)),alpha = 0.5)
```



```
print(JOB$Minimum.Qual.Requirements[1])
```

```
## [1] "1.\tA baccalaureate degree from an accredited college and two years of experienc
e in community work or community centered activities in an area related to the duties de
scribed above; or  2.\tHigh school graduation or equivalent and six years of experience
in community work or community centered activities in an area related to the duties as d
escribed above; or  3.\tEducation and/or experience which is equivalent to \"1\" or \"2
\" above. However, all candidates must have at least one year of experience as described
in \"1\" above."
```

```
print(JOB$Minimum.Qual.Requirements[3])
```

```
## [1] "1. Three years of full-time satisfactory experience as a mechanic, journey perso
n or helper in the electrical trades, the mechanical trades, or the construction or main
tenance of buildings; or  2. A satisfactory combination of education and experience that
is equivalent to \"1\" above. Education may be substituted for experience on the basis t
hat each one year of full-time training in the electrical, mechanical, or construction t
rades in a trade school or vocational high school approved by a State창?p꿺s Department o
f Education or a recognized accrediting organization, may be substituted for six months
of the experience described in \"1\" above. However, all candidates must have a minimum
of two years of experience as described in \"1\" above."
```

```
print(JOB$Minimum.Qual.Requirements[160])
```

```
## [1] "1. A four-year high school diploma or its educational equivalent, and   (a) six
months of full-time mainframe, mini-computer or LAN/WAN computer operations experience o
r service desk/desk top support experience acquired in the past one year and three month
s; or  (b) graduation from an approved technical school (approximately 675 hours) with a
specialization in mainframe/mini-computer operations and/or technical support acquired i
n the past five years and three months; or (c) currently valid A+ or Network+ certificat
ion; or  2. An associate degree or 60 semester credits from an accredited college includ
ing or supplemented by 12 semester credits in computer technology; or  3. Education and/
or experience that is equivalent to 창???창???or 창???창???above. Undergraduate credit may
be substituted for experience on the basis of 30 semester credits, from an accredited co
llege, including or supplemented by six credits in computer technology for three months
of experience. However, all candidates must have a high school diploma or its educationa
l equivalent.     Experience which primarily involves performing computer data entry an
d/or using word processing, spread sheet and/or database applications as an end user is
not acceptable towards meeting the education and experience requirements for this examin
ation."
```

```
JOB$Minimum.Qual.Requirements2 = JOB$Minimum.Qual.Requirements



JOB$Minimum.Qual.Requirements2 = gsub('\"1\"',"",JOB$Minimum.Qual.Requirements2)

for(i in 1:5){
  JOB$Minimum.Qual.Requirements2 = gsub(paste('\"',i,'\"',sep = ""),"",JOB$Minimum.Qual.
Requirements2)
}

JOB$Minimum.Qual.Requirements2[1]
```

```
## [1] "1.\tA baccalaureate degree from an accredited college and two years of experienc
e in community work or community centered activities in an area related to the duties de
scribed above; or  2.\tHigh school graduation or equivalent and six years of experience 
in community work or community centered activities in an area related to the duties as d
escribed above; or  3.\tEducation and/or experience which is equivalent to  or  above. H
owever, all candidates must have at least one year of experience as described in  abov
e."
```

```
summary(as.factor(JOB$Business.Title3))
```

```
##    Analyst Assistance   Contract   Director        ETC    Execute 
##        135         66         23         67        625          5 
## Management 
##        135 
```

```
Analyst = JOB %>%
  filter(Business.Title3 == "Analyst")

print(Analyst$Minimum.Qual.Requirements[1])
```

```
## [1] "1. A baccalaureate degree from an accredited college and six months of satisfact
ory full-time professional experience in procurement of goods, services, construction or
construction-related services, or professional, technical or administrative experience i
n contract negotiation/management; or  2. An associate degree or completion of 60 semest
er credits from an accredited college, and 18 months of satisfactory, full-time professi
onal experience as described in 창???창???above; or  3. A four-year high school diploma o
r its educational equivalent and two and one-half years of satisfactory full time profes
sional experience as described in 창???창???above; or  4. A combination of education and/
or experience equivalent to 창???창??? 창???창??? or 창???창???above. College education may
be substituted for professional experience under 창???창???or 창???창???above at the rate
of 30 semester credits from an accredited college for 6 months of experience. However, a
ll candidates must have at least a four year high school diploma or its educational equi
valent and 6 months of the experience described in 창???창???above.  SPECIAL NOTES:  To b
e eligible for placement in Assignment Level II, individuals must have, after meeting th
e minimum requirements, either one year served at Assignment Level I or one additional y
ear of the experience described in \"1\" above.  To be eligible for placement in Assignm
ent Level III, individuals must have, after meeting the minimum requirements, either one
year served at Assignment Level II or two additional years of the experience described i
n \"1\" above, at least one year of which must have been supervisory, or spent performin
g professional procurement duties equivalent to those performed at Assignment Level II
I."
```

```
print(Analyst$Minimum.Qual.Requirements[3])
```

```
## [1] "1. A baccalaureate degree from an accredited college and three years of satisfac
tory full-time progressively responsible  clerical/administrative  experience, one year
of which must have been in an administrative capacity or supervising staff performing  c
lerical/administrative  work of more than moderate difficulty;  or  2. An associate degr
ee or 60 semester credits from an accredited college and four years of satisfactory full
-time progressively responsible  clerical/administrative  experience including one year
of the administrative supervisory experience described in  \"1\"  above;  or  3. A four-
year high school diploma or its educational equivalent approved by a State's department
of education or a recognized accrediting organization and five years of satisfactory ful
l-time progressively responsible  clerical/administrative  experience including one year
of the administrative supervisory experience as described in  \"1\"  above;  4. Educatio
n and/or experience equivalent to  \"1\",  \"2\",  or  \"3\"  above.  However, all candi
dates must possess the one year of administrative or supervisory experience as described
in  \"1\"  above.  Education above the high school level may be substituted for the gene
ral clerical/administrative experience (but not for the one year of administrative or su
pervisory experience described in  \"1\"  above) at a rate of 30 semester credits from a
n accredited college for 6 months of experience up to a maximum of 3했쩍 years."
```

```
print(Analyst$Minimum.Qual.Requirements[160])
```

```
## [1] NA
```

```
length(unlist(strsplit(Analyst$Minimum.Qual.Requirements2[1],";")))
```

```
## [1] 4
```

```
strsplit(Analyst$Minimum.Qual.Requirements2[10],";")
```

```
## [[1]]
## [1] "1. A master창?p꽒s degree from an accredited college or university, accredited by
regional, national, professional or specialized agencies recognized as accrediting bodie
s by the U.S. Secretary of Education and the Council for Higher Education Accreditation
(CHEA) in economics, finance, accounting, business or public administration, human resou
rces management, management science, operations research, organizational behavior, indus
trial psychology, statistics, personnel administration, labor relations, psychology, soc
iology, human resources development, political science, urban studies or a closely relat
ed field, and one year of satisfactory full-time professional experience in one or a com
bination of the following: working with the budget of a large public or private concern
in budget administration, accounting, economic or financial administration, or fiscal or
economic research"
## [2] " or in management or methods analysis, operations research, organizational resea
rch or program evaluation"
## [3] " or in personnel or public administration, recruitment, position classification,
personnel relations, labor relations, employee   benefits, staff development, employment
program planning/administration, labor market research, economic planning, social servic
es program planning/evaluation, or fiscal management"
## [4] " or in a related area"
## [5] " or  2. A baccalaureate degree from an accredited college or university, accredi
ted by regional, national, professional or specialized agencies recognized as accreditin
g bodies by the U.S. Secretary of Education and the Council for Higher Education Accredi
tation (CHEA) and three years of satisfactory full-time professional experience in the a
reas described in 창???\" above."
```

```
Length = c()
MR = list()

for(k in 1:nrow(Analyst)){

  A  = length(unlist(strsplit(Analyst$Minimum.Qual.Requirements2[k],";")))
  Length = c(Length,A)
  MR[[k]] = unlist(strsplit(Analyst$Minimum.Qual.Requirements2[k],";"))

}
```

```
A_MR1 = c()
A_MR2 = c()
A_MR3 = c()
A_MR4 = c()
A_MR5 = c()
A_MR6 = c()
A_MR7 = c()
A_MR8 = c()

for(k in 1:nrow(Analyst)){

  A_MR1 = c(A_MR1,MR[[k]][1])
  A_MR2 = c(A_MR2,MR[[k]][2])
  A_MR3 = c(A_MR3,MR[[k]][3])
  A_MR4 = c(A_MR4,MR[[k]][4])
  A_MR5 = c(A_MR5,MR[[k]][5])
  A_MR6 = c(A_MR6,MR[[k]][6])
  A_MR7 = c(A_MR7,MR[[k]][7])
  A_MR8 = c(A_MR8,MR[[k]][8])

}

Analyst = Analyst %>%
  mutate(  A_MR1 = A_MR1,
            A_MR2 = A_MR2,
            A_MR3 = A_MR3,
            A_MR4 = A_MR4,
            A_MR5 = A_MR5,
            A_MR6 = A_MR6,
            A_MR7 = A_MR7,
            A_MR8 = A_MR8
  )
```

A_MR1

```
Corpus_AMR1 = Corpus(VectorSource(Analyst$A_MR1))

CORPUS_AMR1_TM= tm_map(Corpus_AMR1, tolower)
```

```
## Warning in tm_map.SimpleCorpus(Corpus_AMR1, tolower): transformation drops
## documents
```

```
CORPUS_AMR1_TM = tm_map(CORPUS_AMR1_TM, removePunctuation)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_AMR1_TM, removePunctuation):
## transformation drops documents
```

```
CORPUS_AMR1_TM = tm_map(CORPUS_AMR1_TM, stripWhitespace)
```

```
## Warning in tm_map.SimpleCorpus(CORPUS_AMR1_TM, stripWhitespace):
## transformation drops documents
```

```
CORPUS_AMR1_TM = tm_map(CORPUS_AMR1_TM, removeWords,
                    c(stopwords("english"),"my","custom","words"))
```

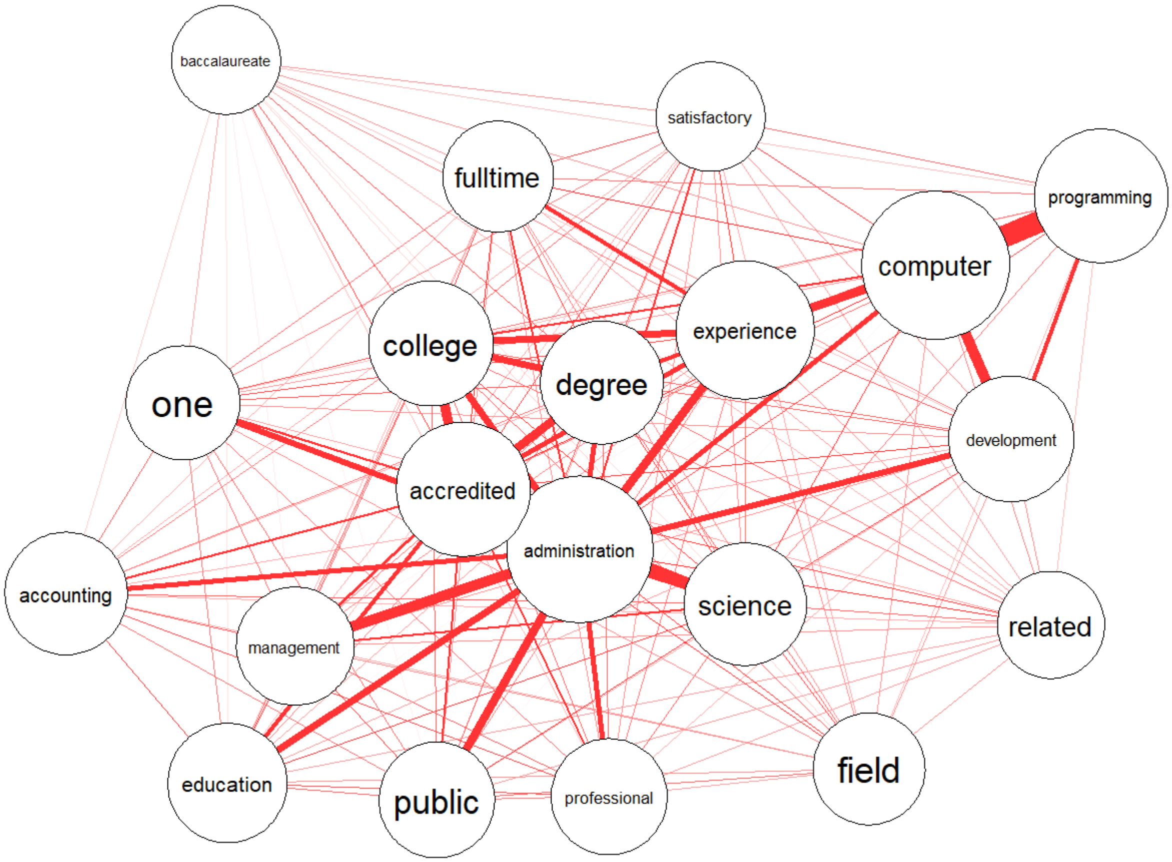
```
## Warning in tm_map.SimpleCorpus(CORPUS_AMR1_TM, removeWords,
## c(stopwords("english"), : transformation drops documents
```

```
TDM = TermDocumentMatrix(CORPUS_AMR1_TM)
inspect(TDM)
```

```
## <<TermDocumentMatrix (terms: 251, documents: 135)>>
## Non-/sparse entries: 3903/29982
## Sparsity           : 88%
## Maximal term length: 22
## Weighting          : term frequency (tf)
## Sample             :
##                  Docs
## Terms            10 11 25 28 5 6 60 61 8 9
##   accredited      2  2  2  2 2 2  2  2 2 2
##   administration  4  4  4  4 4 4  4  4 4 4
##   baccalaureate   0  0  0  0 0 0  0  0 0 0
##   college         1  1  1  1 1 1  1  1 1 1
##   computer        0  0  0  0 0 0  0  0 0 0
##   degree          1  1  1  1 1 1  1  1 1 1
##   experience      1  1  1  1 1 1  1  1 1 1
##   fulltime        1  1  1  1 1 1  1  1 1 1
##   satisfactory    1  1  1  1 1 1  1  1 1 1
##   science         2  2  2  2 2 2  2  2 2 2
```

```
TDM = as.matrix(TDM)
D = rowSums(TDM)
D = sort(D,decreasing = TRUE)
```

```
Word_Count = rowSums(TDM)
Word_Count <- order(Word_Count, decreasing=T)
word.freq1_J <- TDM[Word_Count[1:20],]
co.matrix1_J <- word.freq1_J %*% t(word.freq1_J)

qgraph(co.matrix1_J, labels=rownames(co.matrix1_J), diag=F,
       layout='spring', edge.color="#FF3333", vsize=log(diag(co.matrix1_J)) * 2,alpha =
0.5)
```



```
Analyst$Experience = ifelse(str_detect(tolower(Analyst$A_MR1),"experi"),1,0)
```

Statistical Test

Hypothesis Test

가설1. 분석가의 연봉이 타 직업군에 높을까?
귀무가설 : 같다
대립가설 : 다르다

```
summary(JOB$Salary.Range.From)
```

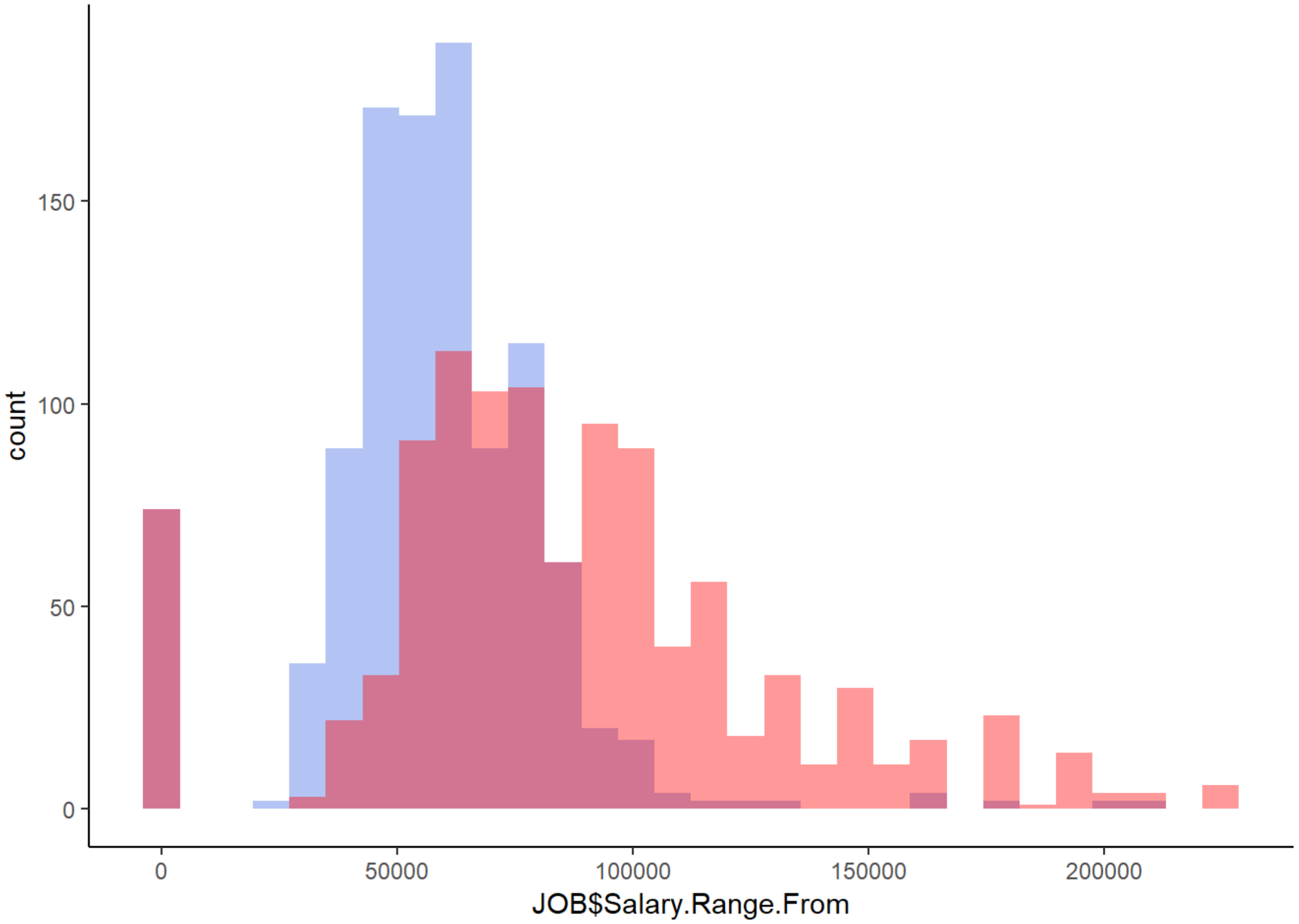
```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##      8.75  47860.00  57065.00  57219.93  69522.75 209585.00
```

```
summary(JOB$Salary.Range.To)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max. 
##     10.36  61104.00  80000.00  85906.75 105000.00 224749.00
```

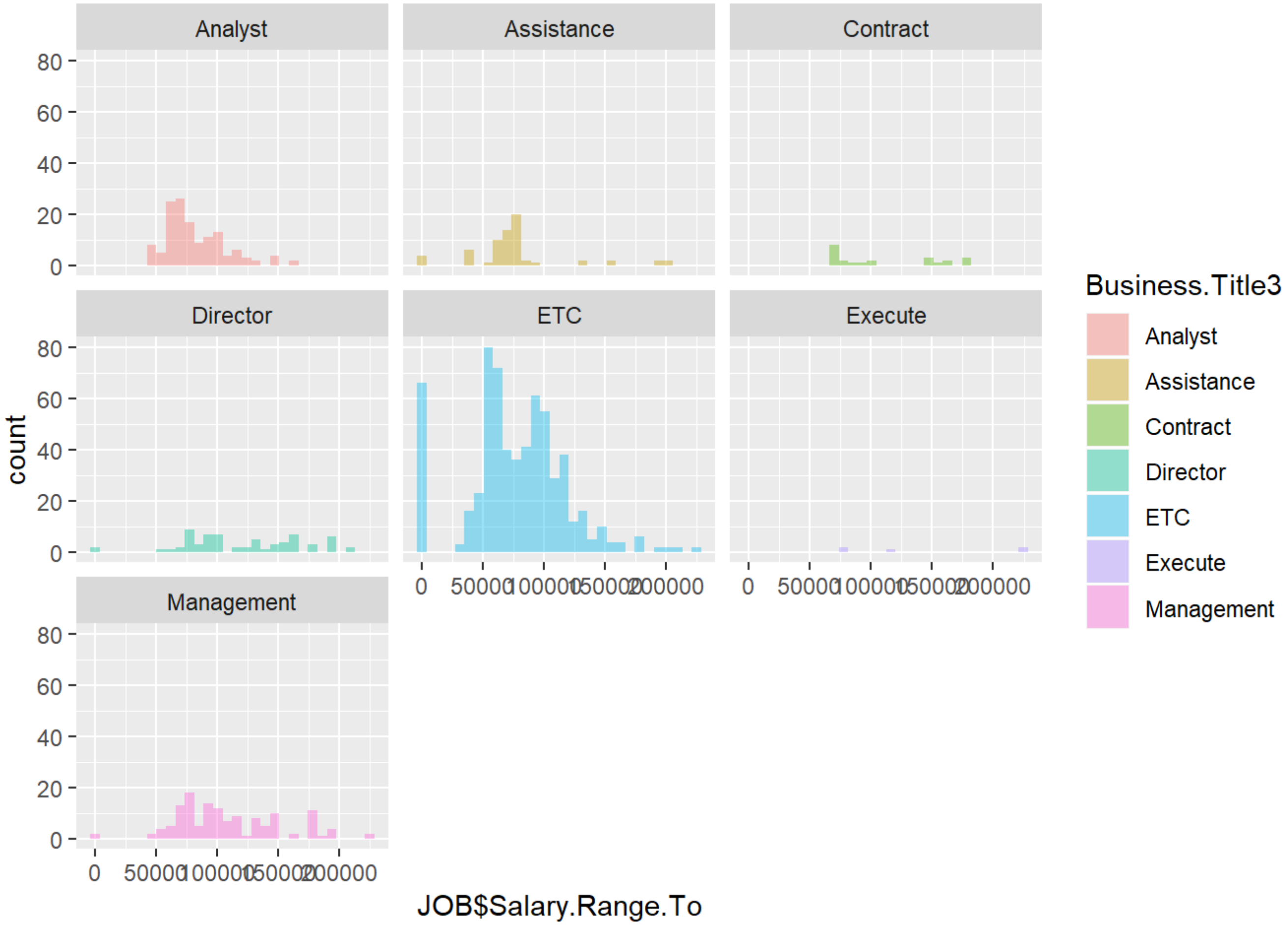
```
ggplot(JOB) +
  geom_histogram(aes(x = JOB$Salary.Range.From), fill = 'royalblue', alpha = 0.4) +
  geom_histogram(aes(x = JOB$Salary.Range.To), fill = 'red', alpha = 0.4) +
  theme_classic()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

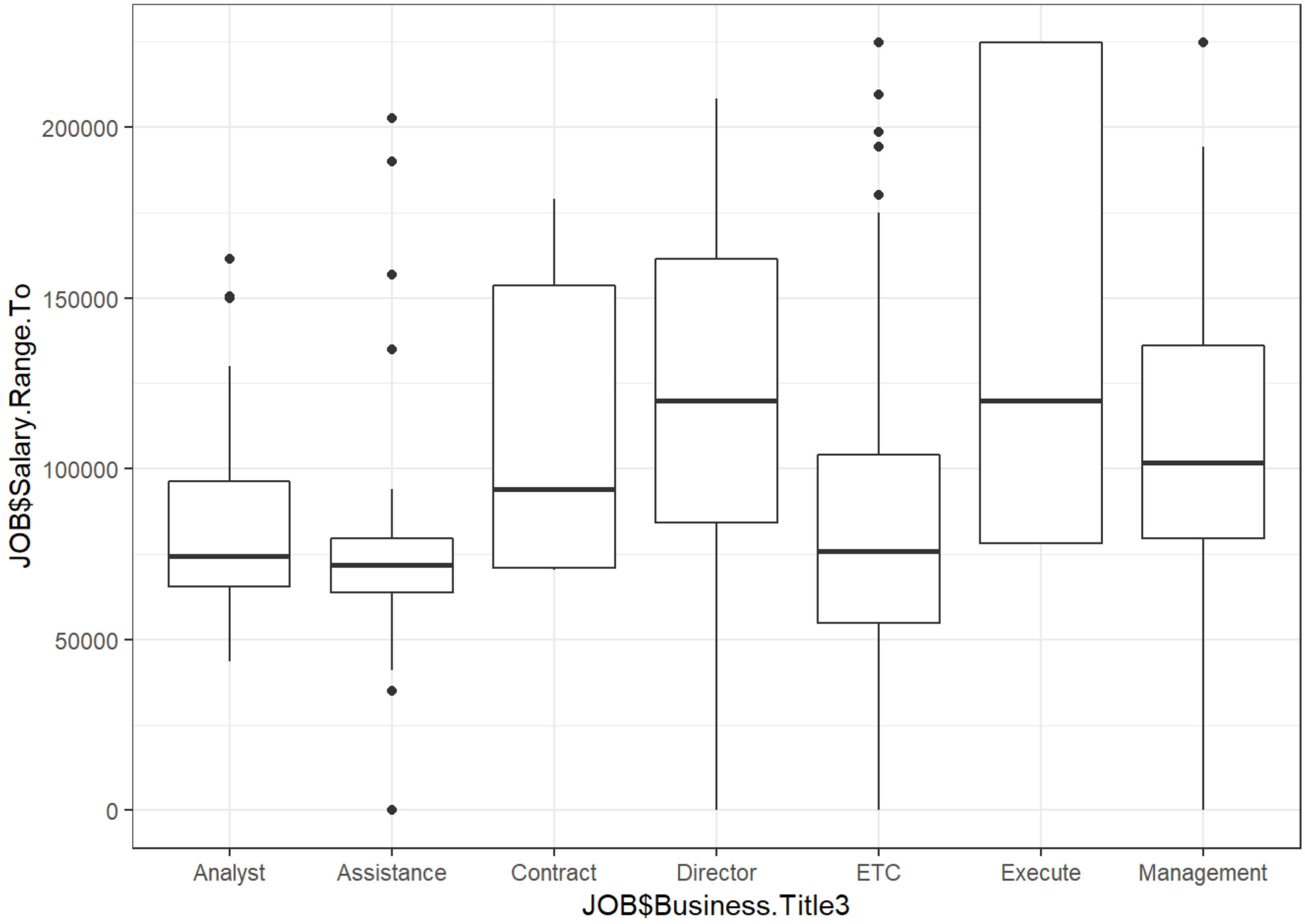


```
ggplot(JOB) +
  geom_histogram(aes(x = JOB$Salary.Range.To, fill = Business.Title3), alpha = 0.4) +
  facet_wrap( ~ Business.Title3)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
ggplot(JOB) +
  geom_boxplot(aes(x = JOB$Business.Title3, y = JOB$Salary.Range.To)) +
  theme_bw()
```



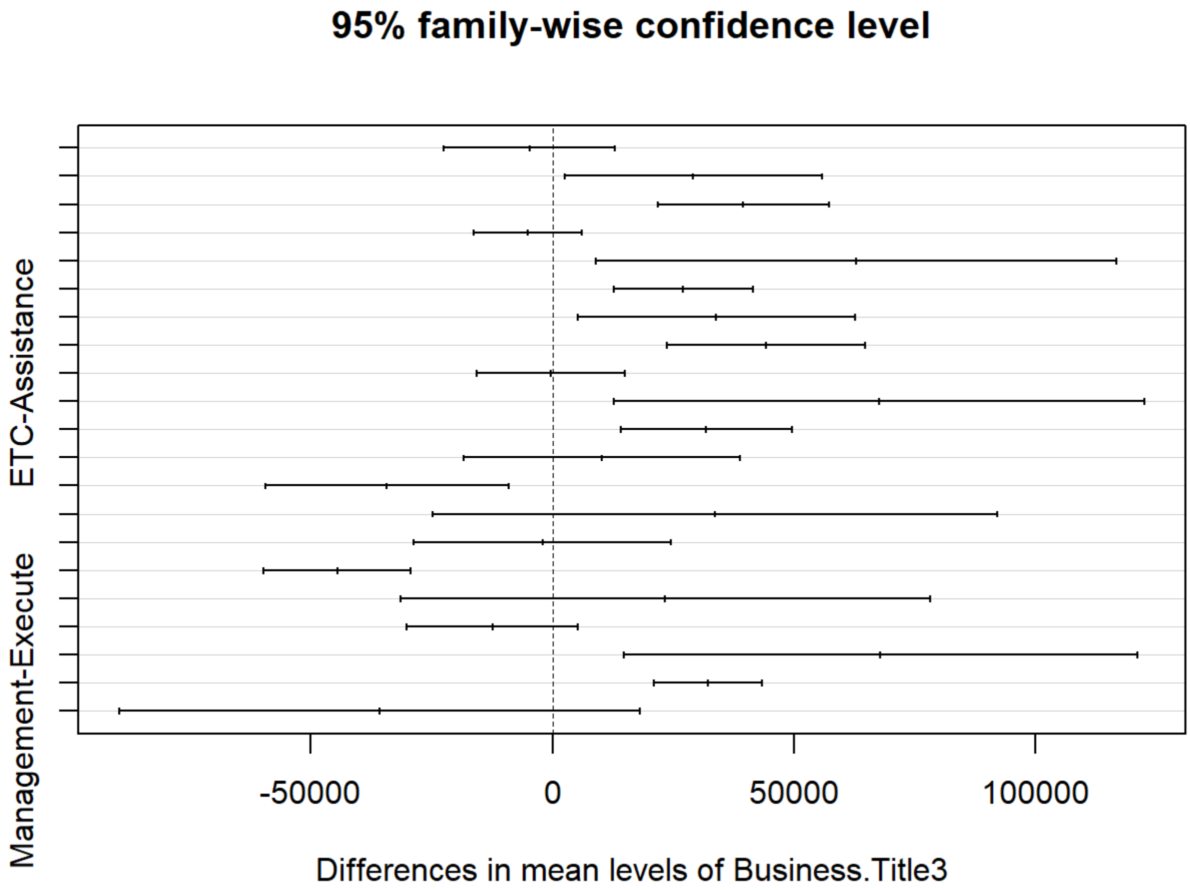
```
ANOVA = aov(Salary.Range.To ~ Business.Title3, data = JOB)
summary(ANOVA)
```

```
##                   Df    Sum Sq   Mean Sq F value Pr(>F)    
## Business.Title3    6 2.468e+11 4.114e+10   25.63 <2e-16 ***
## Residuals       1049 1.684e+12 1.605e+09                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TUKEY = TukeyHSD(ANOVA)

TUKEY
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Salary.Range.To ~ Business.Title3, data = JOB)
##
## $Business.Title3
##                              diff        lwr        upr     p adj
## Assistance-Analyst     -4760.9706 -22536.901  13014.960 0.9858326
## Contract-Analyst       29200.7997   2503.280  55898.319 0.0215983
## Director-Analyst       39542.3548  21855.746  57228.963 0.0000000
## ETC-Analyst            -5026.9884 -16259.402   6205.425 0.8416629
## Execute-Analyst        62901.8519   9002.282 116801.422 0.0105174
## Management-Analyst     27130.1566  12724.891  41535.423 0.0000007
## Contract-Assistance    33961.7703   5304.667  62618.873 0.0087323
## Director-Assistance    44303.3253  23778.020  64828.630 0.0000000
## ETC-Assistance          -266.0178 -15583.953  15051.917 1.0000000
## Execute-Assistance     67662.8224  12766.232 122559.413 0.0052800
## Management-Assistance  31891.1272  14115.197  49667.058 0.0000029
## Director-Contract      10341.5551 -18260.228  38943.338 0.9372756
## ETC-Contract          -34227.7881 -59355.720  -9099.856 0.0011979
## Execute-Contract       33701.0522 -24697.659  92099.763 0.6130128
## Management-Contract    -2070.6430 -28768.162  24626.876 0.9999880
## ETC-Director          -44569.3431 -59783.532 -29355.154 0.0000000
## Execute-Director       23359.4971 -31508.235  78227.230 0.8709277
## Management-Director   -12412.1981 -30098.807   5274.411 0.3695218
## Execute-ETC            67928.8402  14789.222 121068.459 0.0031839
## Management-ETC         32157.1450  20924.732  43389.558 0.0000000
## Management-Execute    -35771.6952 -89671.265  18127.875 0.4406941
```

```
plot(TUKEY)
```



```
Reg = lm(Salary.Range.To ~ Experience , data = Analyst)
summary(Reg)
```

```
##
## Call:
## lm(formula = Salary.Range.To ~ Experience, data = Analyst)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -44754 -11378    209   6896  70393
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    66012       3299  20.012  < 2e-16 ***
## Experience     25093       4070   6.165 8.01e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 22370 on 132 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.2235, Adjusted R-squared:  0.2177
## F-statistic:    38 on 1 and 132 DF,  p-value: 8.009e-09
```

최근 몇년 간 구직자리 변화 비교

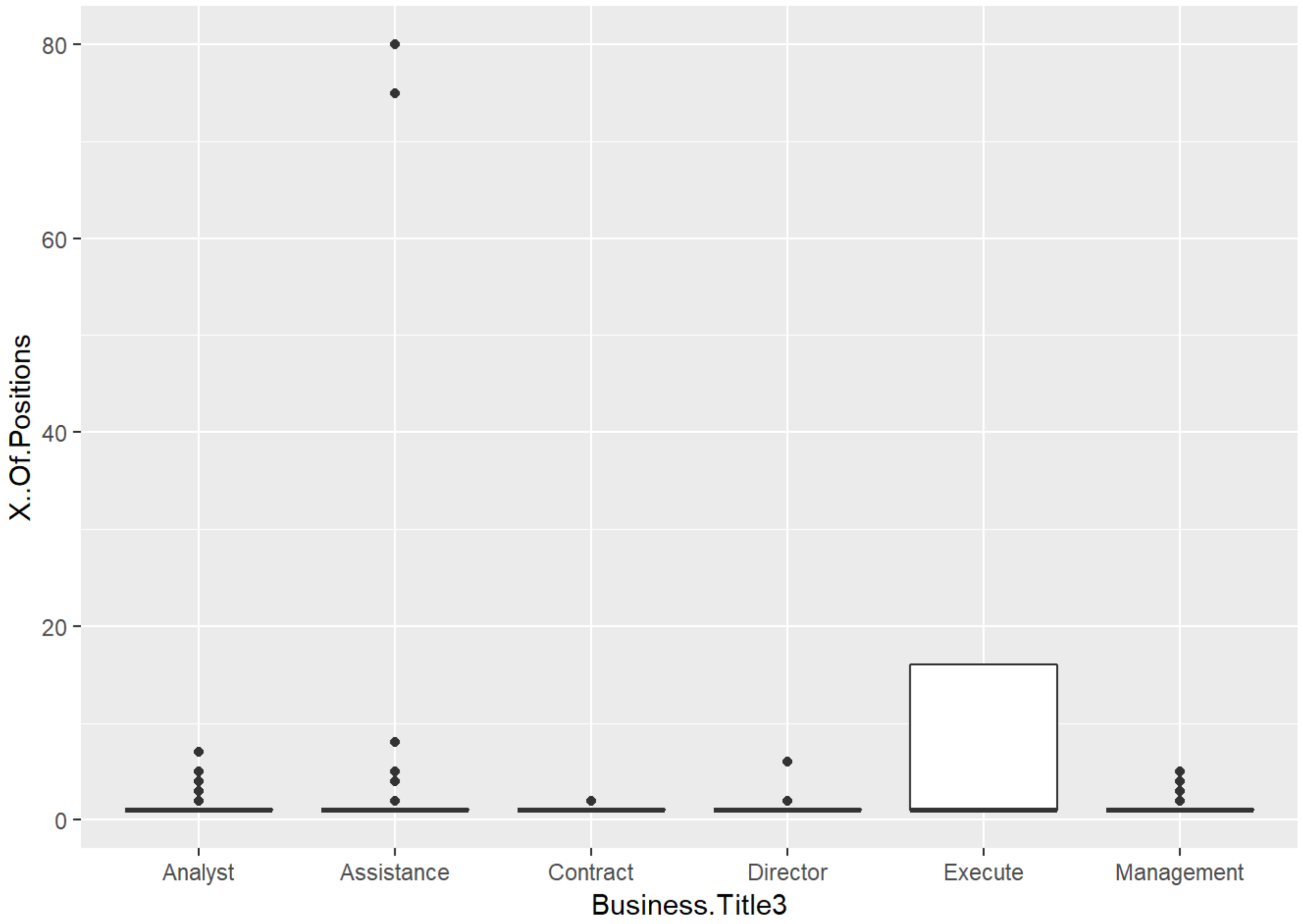
귀무가설 : 각 직업군 별로 채용인원수가 같다.

대립가설 : 다르다.

```
JOB_2017 = JOB %>%
  filter(Posting.Date > '2017-01-01')

Anova = aov(JOB_2017$X..Of.Positions ~ JOB_2017$Business.Title3)

summary(Anova)
```

```
##                           Df Sum Sq Mean Sq F value  Pr(>F)
## JOB_2017$Business.Title3   6   2009   334.8    3.63 0.00146 **
## Residuals                833  76842    92.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
JOB %>%
  filter(Business.Title3 != 'ETC') %>%
ggplot() +
  geom_boxplot(aes(x = Business.Title3, y = X..Of.Positions))
```


X..Of.Positions
Analyst
Assistance
Contract
Director
Execute
Management
Business.Title3