

# correlation

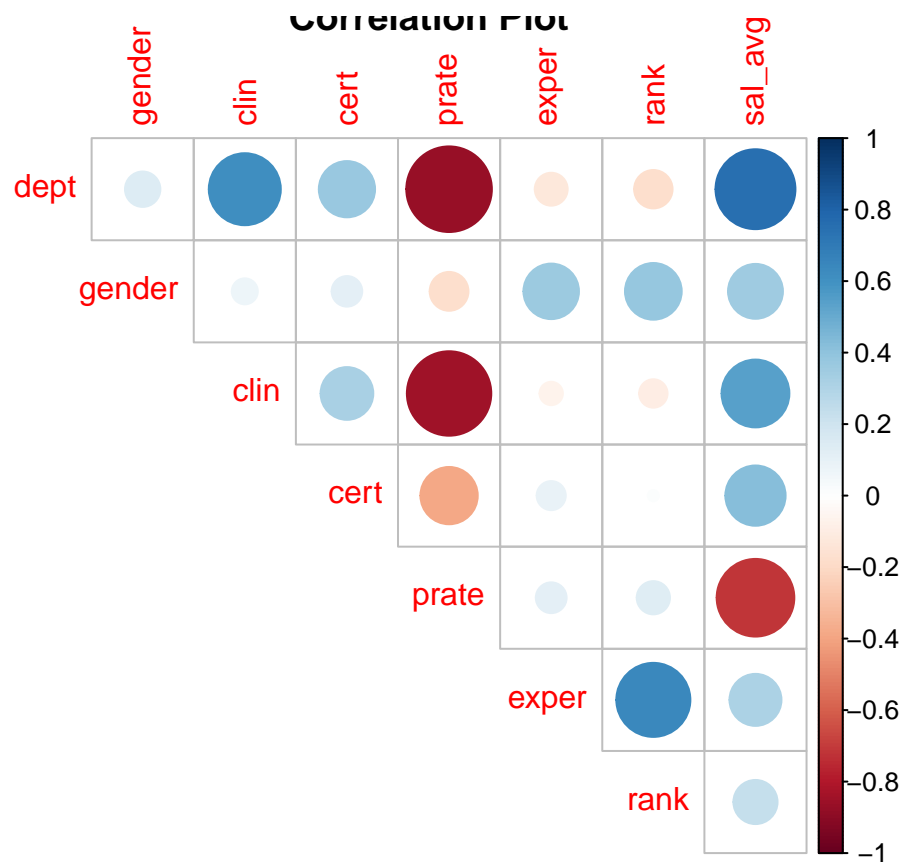
Yingxi Ji

12/5/2019

```
lawsuit = read_csv("./Lawsuit.csv") %>%  
  janitor::clean_names() %>%  
  mutate(sal_avg = (sal94 + sal95)/2)  
  
sapply(lawsuit, function(na) sum(is.na(na)))
```

```
##      id    dept gender  clin  cert  prate  exper  rank  sal94  
##      0      0      0      0      0      0      0      0      0  
##  sal95 sal_avg  
##      0      0
```

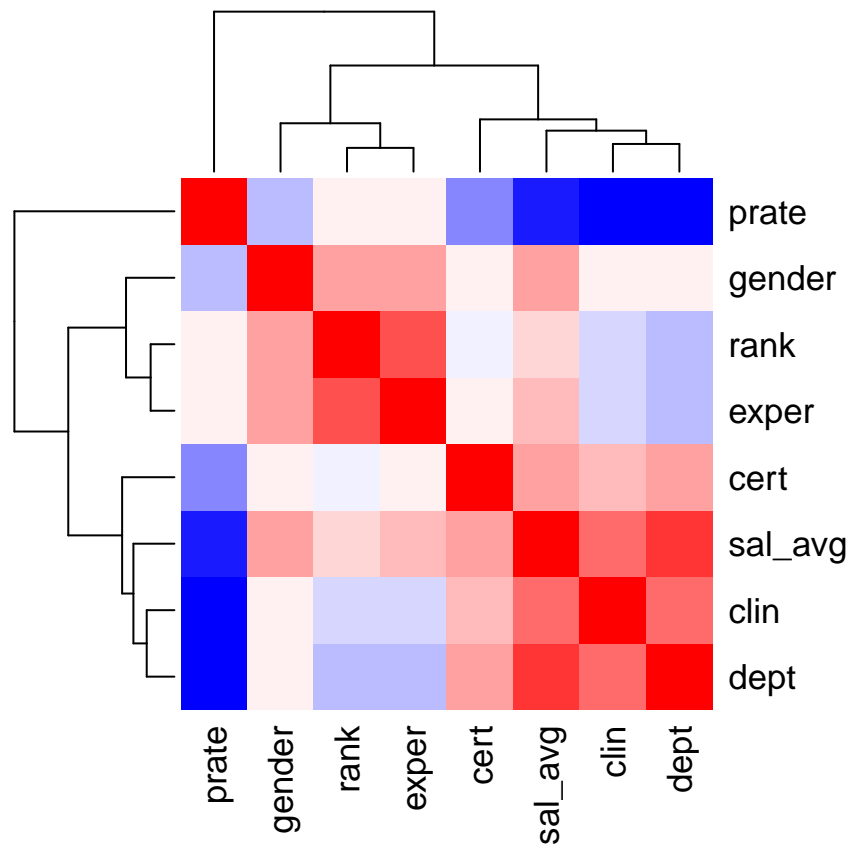
```
lawsuit_cor = lawsuit %>%  
  select(-id, -sal94, -sal95)  
cor(lawsuit_cor)%>%  
  corrplot(method = "circle", type = "upper", diag=FALSE, title = "Correlation Plot")
```



```
## Matrix
res <- cor(lawsuit_cor)
round(res, 2)
```

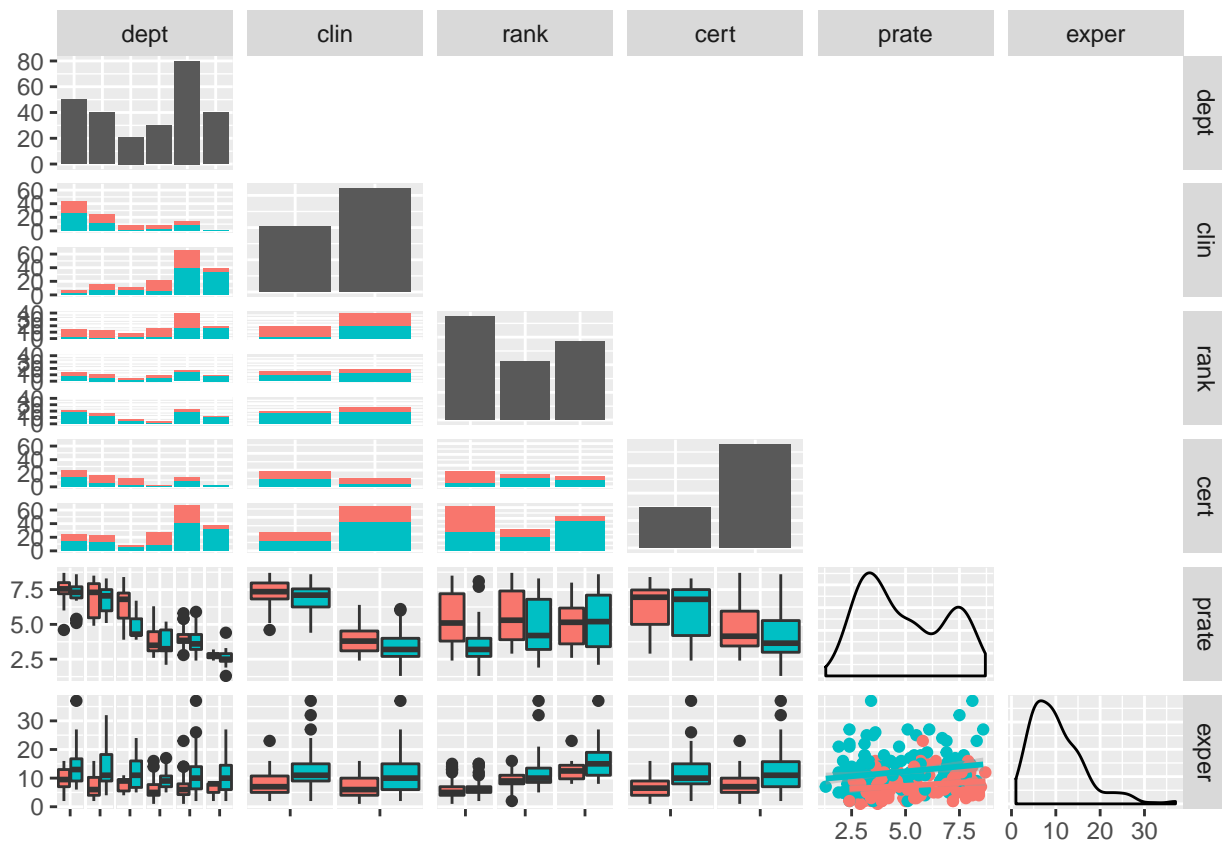
```
##      dept gender  clin  cert prate exper  rank sal_avg
## dept      1.00   0.15  0.61  0.37 -0.87 -0.13 -0.17   0.76
## gender    0.15   1.00  0.08  0.11 -0.18  0.36  0.38   0.36
## clin      0.61   0.08  1.00  0.33 -0.84 -0.06 -0.10   0.54
## cert      0.37   0.11  0.33  1.00 -0.39  0.10  0.01   0.43
## prate     -0.87  -0.18 -0.84 -0.39  1.00  0.11  0.13  -0.71
## exper     -0.13  0.36 -0.06  0.10  0.11  1.00  0.65   0.32
## rank      -0.17  0.38 -0.10  0.01  0.13  0.65  1.00   0.23
## sal_avg    0.76  0.36  0.54  0.43 -0.71  0.32  0.23   1.00
```

```
## Heatmap
col<- colorRampPalette(c("blue", "white", "red"))(20)
heatmap(x = res, col = col, symm = TRUE)
```



```
lawsuit_cor$dept = as.factor(lawsuit_cor$dept)
lawsuit_cor$gender = as.factor(lawsuit_cor$gender)
lawsuit_cor$clin = as.factor(lawsuit_cor$clin)
lawsuit_cor$cert = as.factor(lawsuit_cor$cert)
lawsuit_cor$rank = as.factor(lawsuit_cor$rank)
lawsuit_cor$sal_avg = as.factor(lawsuit_cor$sal_avg)
```

```
lawsuit_cor%>%
  GGally::ggpairs(
    columns = c("dept", "clin", "rank", "cert", "prate", "exper"),
    lower = list(
      continuous = "smooth",
      combo = "box",
      mapping = aes(color = gender)
    ),
    upper = "blank"
  )
```



There may be a collinearity issue where we can see from the plot above.

prate is highly correlated with dept and clin.

clin is highly correlated with dept.

rank is highly correlated with exper.