prosperity_1125

Jonathan Kang (jk36)

2022-11-25

```
library(dplyr)
library(tidyverse)
library(readxl)
library(sqldf)
```

```
## Warning in doTryCatch(return(expr), name, parentenv, handler): unable to load shared object '/Library
##   dlopen(/Library/Frameworks/R.framework/Resources/modules//R_X11.so, 0x0006): Library not loaded: /
##   Referenced from: /Library/Frameworks/R.framework/Versions/4.1/Resources/modules/R_X11.so
##   Reason: tried: '/opt/X11/lib/libSM.6.dylib' (no such file), '/Library/Frameworks/R.framework/Resou
```

```
library(ggplot2)
library(GGally)
library(reshape2)
library(MASS)
library(mclust)
```

Import Data

```
free_prosp_full_dataset <- read_excel("Freedom-and-Prosperity-Indexes-Full-Data-Set.xlsx",
    na = "no data")
free_prosp_data <- sqldf("SELECT Region as region,
              Country as country,
              `Freedom rank 2021` as freedom_rank,
              `Freedom category 2021` as freedom_category,
              `Freedom score 2021` as freedom_score,
              `Economic Freedom score 2021` as economic_freedom_score,
              `Political Freedom score 2021` as political_freedom_score,
              `Legal Freedom score 2021` as legal_freedom_score,
              `Prosperity rank 2021` as prosperity_rank,
              `Prosperity category 2021` as prosperity_cateogry,
              `Prosperity score 2021` as prosperity_score,
              `Income score 2021` as income_score,
              `Environment score 2021` as environment_score,
              `Minority Rights  score 2021` as minority_rights_score,
              `Health score 2021` as health_score,
              `Happiness score 2021` as happiness_score
       FROM free_prosp_full_dataset;
      ")
glimpse(free_prosp_data)
```

```
## Rows: 174
## Columns: 16
## $ region                   <chr> "AP", "ECA", "MENA", "SSA", "AME", "ECA", "AP"~
## $ country                  <chr> "Afghanistan", "Albania", "Algeria", "Angola",~
## $ freedom_rank             <dbl> 164, 66, 138, 131, 60, 83, 15, 12, 153, 50, 14~
## $ freedom_category         <chr> "Unfree", "Mostly Free", "Mostly Unfree", "Mos~
## $ freedom_score            <dbl> 24.10239, 63.58576, 40.27716, 41.59470, 65.830~
## $ economic_freedom_score   <dbl> 17.08333, 77.29435, 39.78885, 54.49857, 62.943~
## $ political_freedom_score <dbl> 33.32030, 61.86086, 35.71293, 34.56298, 79.885~
## $ legal_freedom_score      <dbl> 21.903547, 51.602055, 45.329692, 35.722562, 54~
## $ prosperity_rank          <dbl> 171, 60, 107, 167, 46, 68, 12, 14, 101, 89, 76~
## $ prosperity_cateogry      <chr> "Unprosperous", "Mostly Prosperous", "Mostly U~
## $ prosperity_score         <dbl> 22.15571, 56.04416, 42.37210, 26.27789, 61.607~
## $ income_score             <dbl> 0.3277097, 6.0444229, 4.0538900, 2.3182425, 10~
## $ environment_score        <dbl> 27.70019, 50.19225, 48.98121, 11.85785, 59.340~
## $ minority_rights_score    <dbl> 45.58054, 93.72385, 38.33682, 67.93933, 98.614~
## $ health_score             <dbl> 37.17013, 81.38809, 75.93969, 25.30787, 75.254~
## $ happiness_score          <dbl> 0.00000, 48.87218, 44.54887, 23.96617, 64.0977~
```

No ties in ranking

Exploratory Visualizations

Overall Dataset

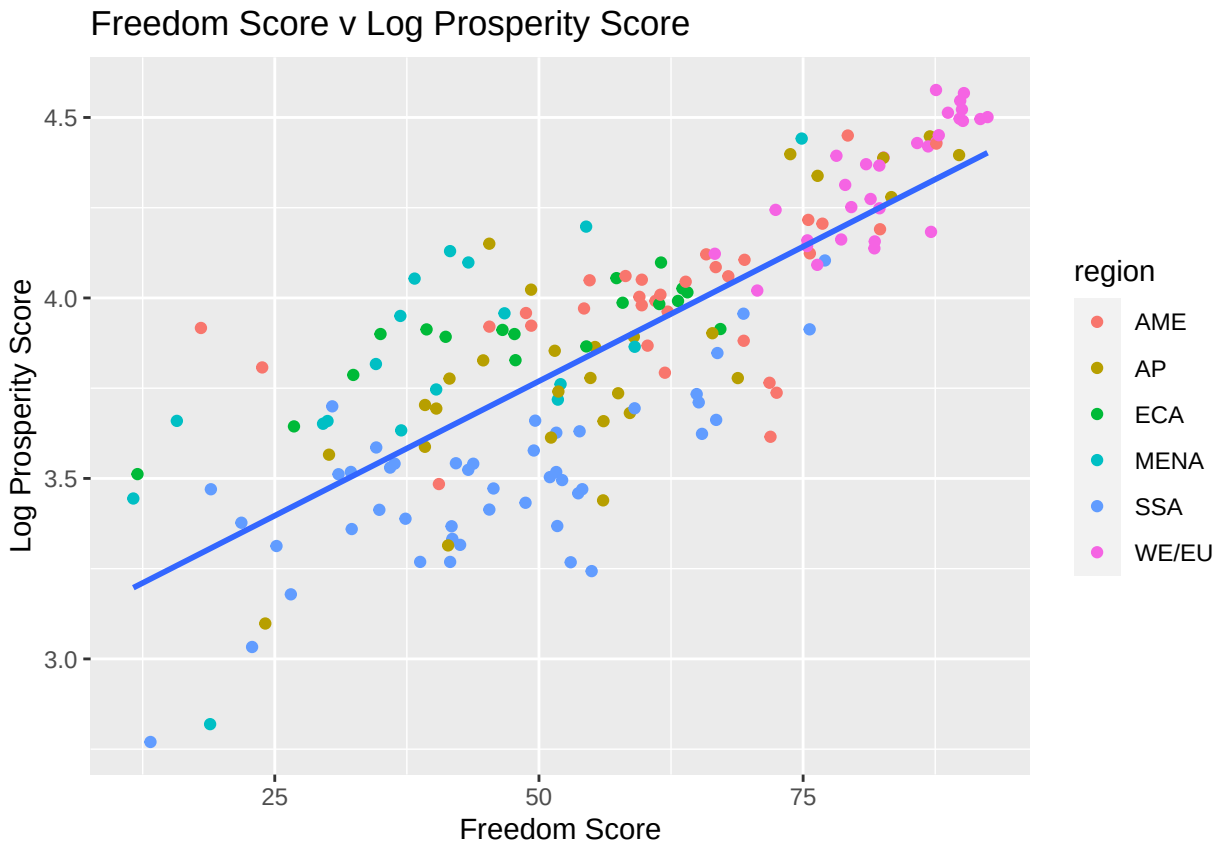
```
## RANK
# lin_model = lm(freedom_rank ~ prosperity_rank, free_prosp_data)
# lin_model$coefficients
#
# plot(free_prosp_data$freedom_rank, free_prosp_data$prosperity_rank)
# # Check normalities, outliers, heteroscedasticity
# plot(lin_model)
#
# ggplot(data = free_prosp_data) +
#   geom_point(aes(x = freedom_rank, y = prosperity_rank)) +
#   geom_abline()
# cor(free_prosp_data$freedom_rank, free_prosp_data$prosperity_rank)

## SCORES
### Apply log transformation to satisfy normality assumptions
model_data <- free_prosp_data %>%
  dplyr::select(region, freedom_score, prosperity_score) %>%
  mutate(log_prosperity = log(prosperity_score))
lin_model = lm(freedom_score ~ log_prosperity, model_data)

## Check normalities, outliers, heteroscedasticity
# plot(model_data$freedom_score, model_data$log_prosperity)
# plot(lin_model)
summary(lin_model)
```

```
##
```

```
## Call:
## lm(formula = freedom_score ~ log_prosperity, data = model_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -40.707  -6.733   1.665   8.090  26.036
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -108.209      9.571  -11.31   <2e-16 ***
## log_prosperity   42.618      2.462   17.31   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.08 on 172 degrees of freedom
## Multiple R-squared:  0.6353, Adjusted R-squared:  0.6332
## F-statistic: 299.6 on 1 and 172 DF,  p-value: < 2.2e-16
```

```
ggplot(data = model_data, aes(x = (freedom_score), y = log_prosperity)) +
  geom_point(aes(color = region)) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Freedom Score v Log Prosperity Score") +
  xlab("Freedom Score") +
  ylab("Log Prosperity Score")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

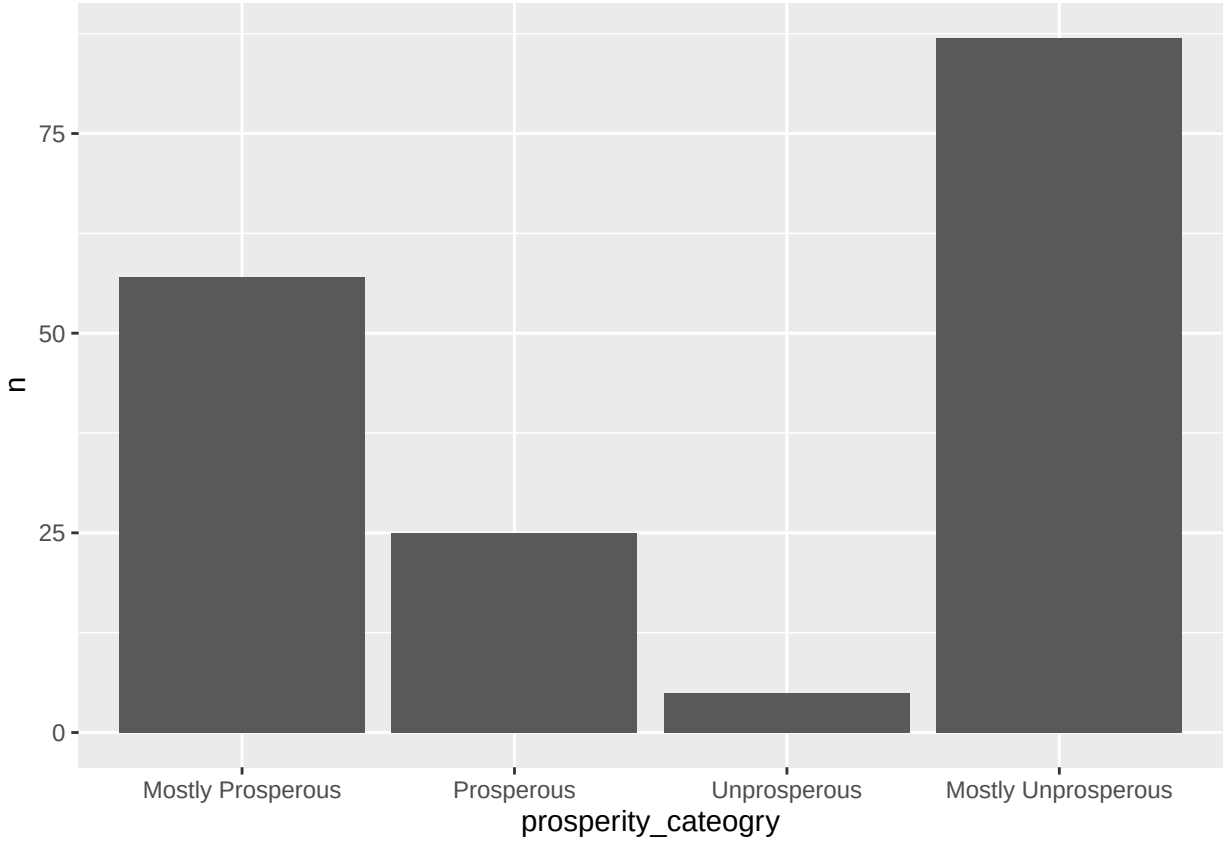


Multiple R-squared: 0.6353

```
table(free_prosp_data$prosperity_cateogry, free_prosp_data$freedom_category)
```

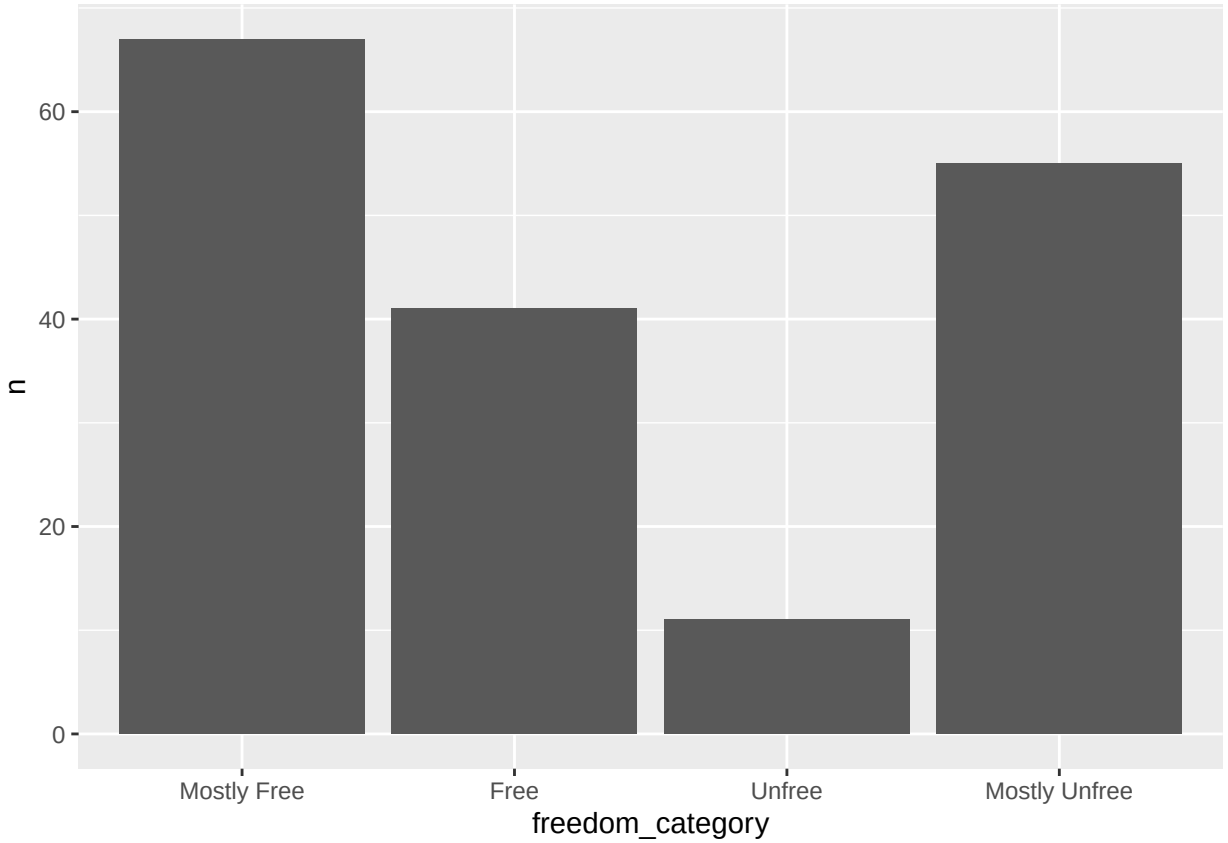
```
##
##                        Free Mostly Free Mostly Unfree Unfree
##   Mostly Prosperous      18          27            11      1
##   Mostly Unprosperous     0          38            43      6
##   Prosperous             23           2             0      0
##   Unprosperous            0           0             1      4
```

```
prosp_counts = free_prosp_data %>% dplyr::select(prosperity_cateogry) %>% group_by(prosperity_cateogry)
order = c("Mostly Prosperous", "Prosperous", "Unprosperous", "Mostly Unprosperous")
prosp_counts <- prosp_counts %>%
  mutate(prosperity_cateogry = factor(prosperity_cateogry, levels = order)) %>%
  arrange(prosperity_cateogry)

ggplot(prosp_counts) +
  geom_col(aes(x = prosperity_cateogry, y = n))
```




```
freedom_counts = free_prosp_data %>% dplyr::select(freedom_category) %>% group_by(freedom_category) %>%
order = c("Mostly Free", "Free", "Unfree", "Mostly Unfree")
freedom_counts <- freedom_counts %>%
  mutate(freedom_category = factor(freedom_category, levels = order)) %>%
  arrange(freedom_category)
```

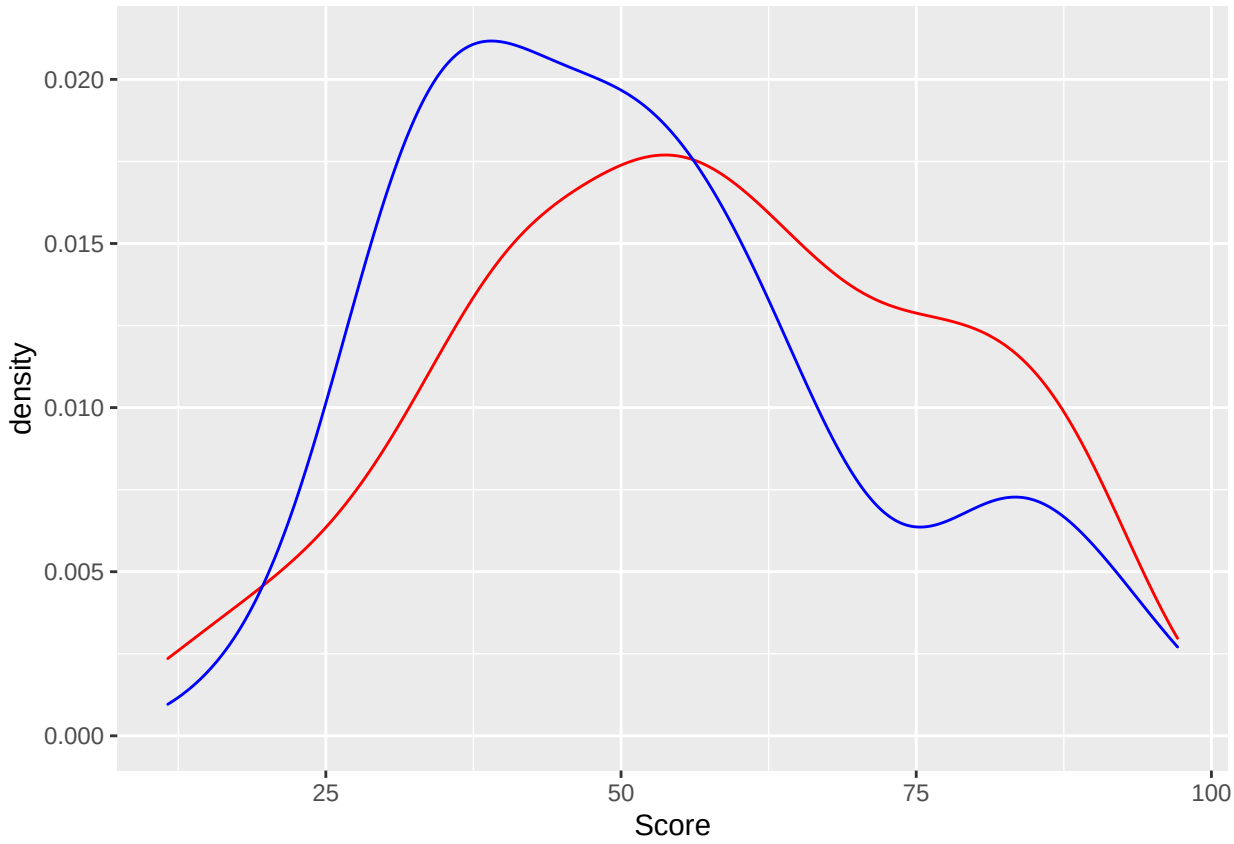
```
ggplot(freedom_counts) +
  geom_col(aes(x = freedom_category, y = n))
```



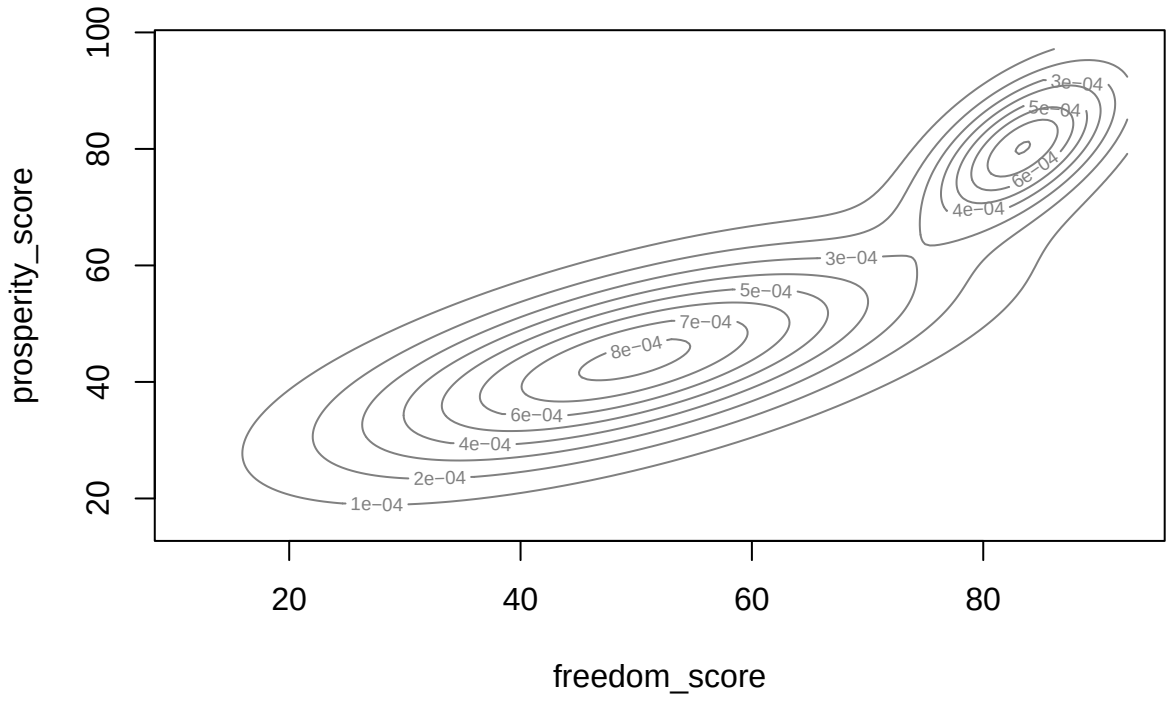
Results show that there is a possibility we can cluster into 2 groups by comparing the scores.
We will proceed with an unsupervised learning method:

Cluster the groups by using Mixture of Multivariate Gaussian Model

```
ggplot(free_prosp_data) +
  geom_density(aes(x = freedom_score), color="Red") +
  geom_density(aes(x = prosperity_score), color="Blue") +
  xlab("Score")
```



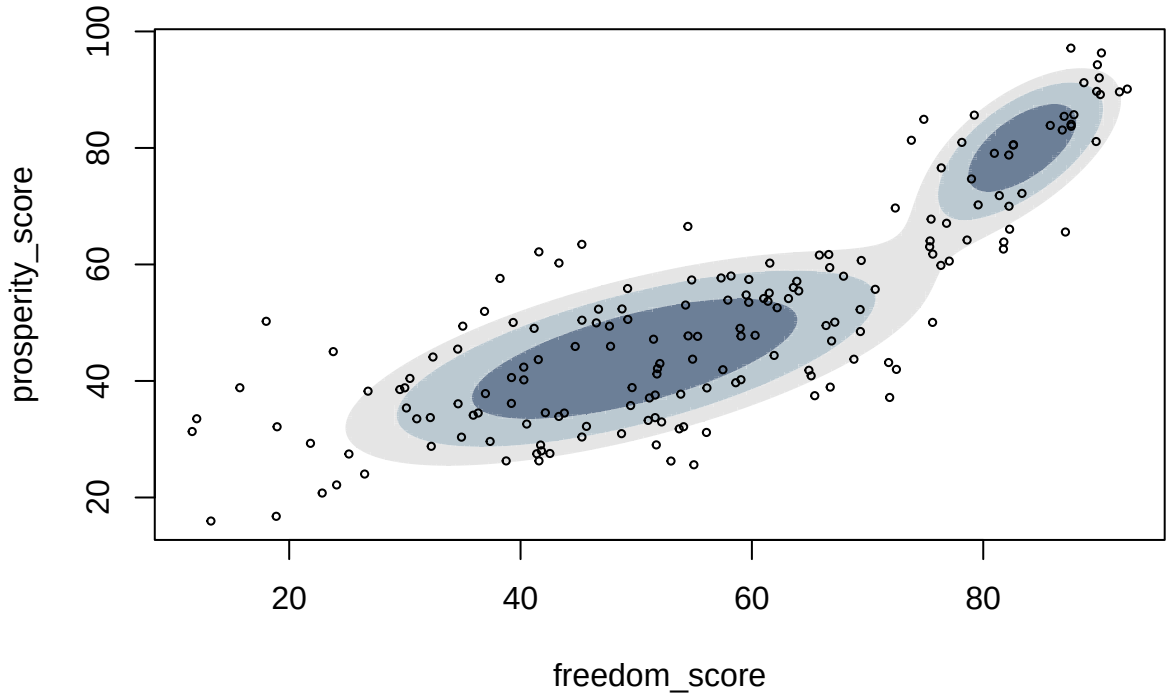

```
# gauss_data <- free_prosp_data %>% select(freedom_score, prosperity_score)
# plot(mclustBIC(model_data[c(2,3)]))
gaussian_mix_model = densityMclust(data=model_data[c(2,3)], G=2)
```



```
summary(gaussian_mix_model)
```

```
## ----------------------------------------------------
## Density estimation via Gaussian finite mixture modeling
## ----------------------------------------------------
##
## Mclust VEV (ellipsoidal, equal shape) model with 2 components:
##
##  log-likelihood   n df       BIC       ICL
##       -1396.483 174 10 -2844.556 -2858.369
```

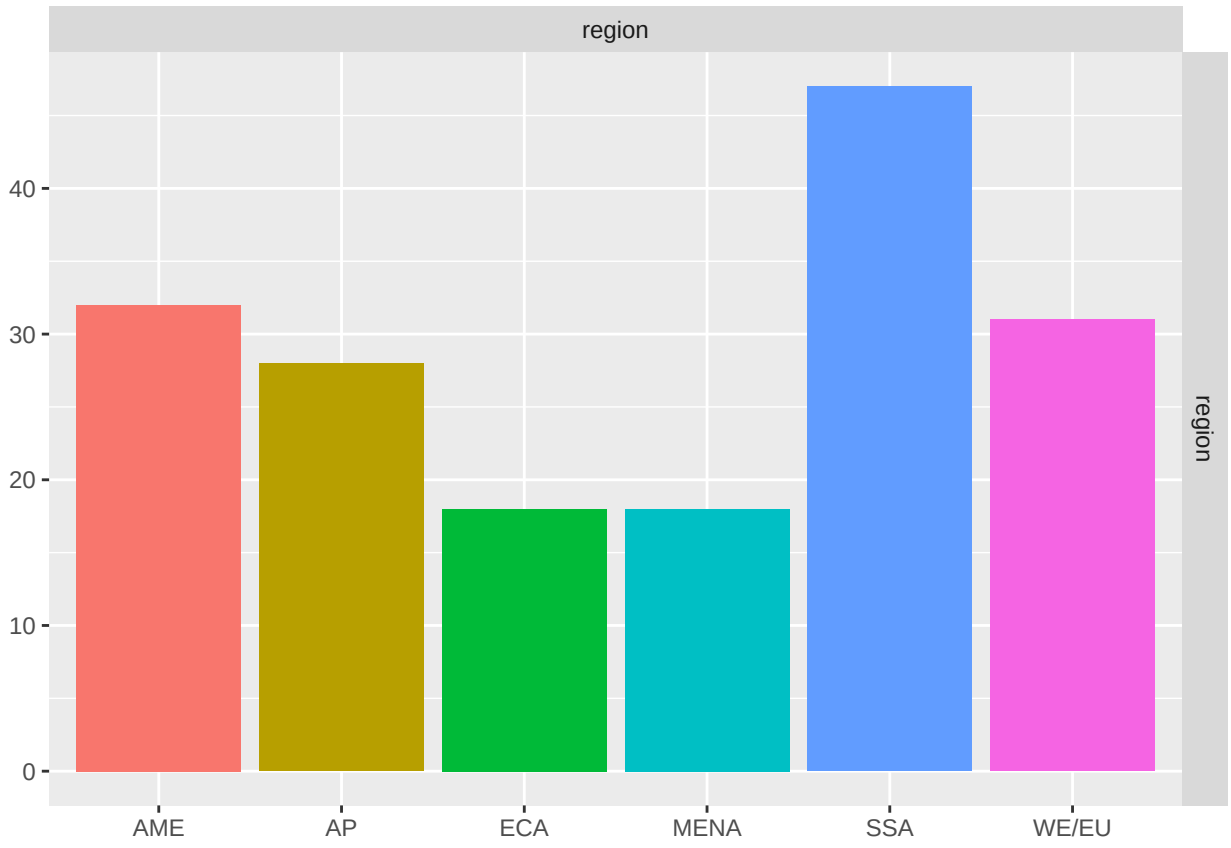
```
plot(gaussian_mix_model, what = "density", type = "hdr", data = model_data[c(2,3)], points.cex = 0.5)
```



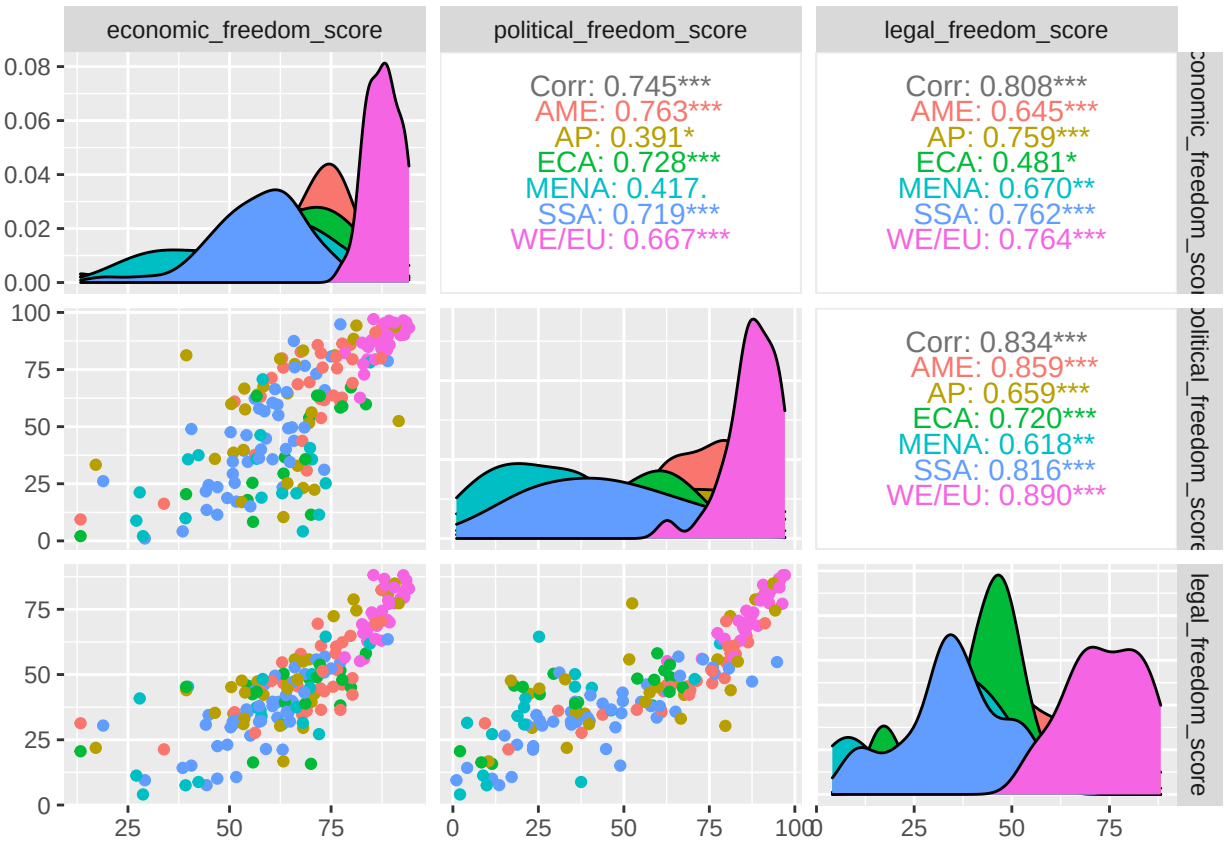
Economic

Overall

```
econ_subs_dataset <- free_prosp_data %>%
  dplyr::select(region, economic_freedom_score, political_freedom_score, legal_freedom_score)
ggpairs(data=econ_subs_dataset,1, mapping=aes(color=region))
```

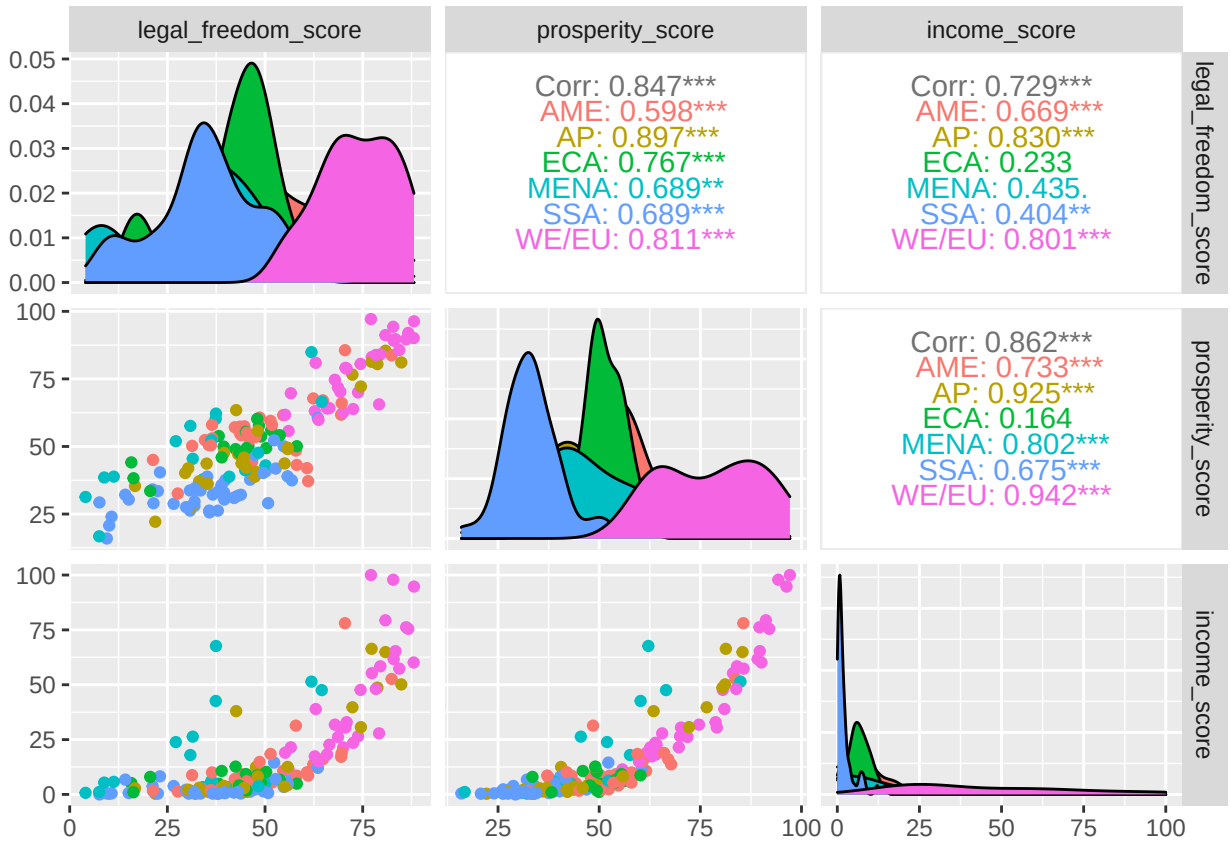



```
ggpairs(data=econ_subs_dataset,2:4, mapping=aes(color=region))
```



Political

```
polit_subs_dataset <- free_prosp_data %>%
  dplyr::select(region, legal_freedom_score, prosperity_score, income_score, environment_score, minority
# ggpairs(data=polit_subs_dataset, mapping=aes(color=region))
ggpairs(data=polit_subs_dataset,2:4, mapping=aes(color=region))
```




```
ggpairs(data=polit_subs_dataset,5:8, mapping=aes(color=region))
```


environment_score
minority_rights_score
health_score
happiness_score
nvironment_scor
norityrights_sco
health_score
happiness_score
Corr: 0.188*
AME: 0.131
AP: 0.308
ECA: 0.423
MENA: 0.531*
SSA: 0.111
WE/EU: 0.235
Corr: 0.886***
AME: 0.734***
AP: 0.895***
ECA: 0.389
MENA: 0.733***
SSA: 0.684***
WE/EU: 0.918***
Corr: 0.793***
AME: 0.667***
AP: 0.680***
ECA: 0.210
MENA: 0.575*
SSA: 0.381*
WE/EU: 0.684***
Corr: 0.111
AME: 0.042
AP: 0.367
ECA: 0.540*
MENA: 0.593**
SSA: −0.108
WE/EU: 0.210
Corr: 0.331***
AME: 0.200
AP: 0.559**
ECA: 0.128
MENA: 0.261
SSA: 0.247
WE/EU: 0.437*
Corr: 0.766***
AME: 0.808***
AP: 0.747***
ECA: −0.065
MENA: 0.655**
SSA: 0.197
WE/EU: 0.655***
0.075
0.050
0.025
0.000
100
75
50
25
0
0
25
50
75
100