MCCA Screening Assessment

Erika Xu

Feb 17, 2025

```
library(tidyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(colorblindr)
```

```
## Loading required package: colorspace
```

The unit of observation for the combined dataset is the firm. Some initial observations I find interesting within the dataset is that there are only five variables that provide detail about each firm. Binary_has_dei_program indicates weather or no a firm has a DEI program. These is count_attorneys, represents the total number of attorneys at the given firm, count_female, the total number of females, and count_minority, the total number of minorities at the given firm. And mcca_score seems to indicate how a firm scored against MCCA's Scorecard.

Also, there are no given dates, and there are duplicates for each firm entry. Therefore, I assume that this dataset represent a snapshot of a firm's DEI positioning at a single point in time, rather than reflecting a firm's DEI progress over time.

With this dataset, I would assume each attribute/column about each firm will give some information about their MCCA score. I am curious about why we are looking at these specific attributes in this dataset or if other variables can be more effective in illustrating charts/ graphs about a firm's MCCA score. Additionally, I am curious about what is categorized under "county_minority" since no prior details are provided about these variables.

```
attributes <- read.csv("data/dat_attributes.csv")
score <- read.csv("data/dat_score.csv")

nrow (attributes)
```

```
## [1] 820
```

```
colnames (attributes)
```

```
## [1] "count_attorneys"        "binary_has_dei_program" "firm_name"
## [4] "count_female"           "count_minority"
```

```
nrow (score)
```

```
## [1] 820
```

```
colnames (score)
```

```
## [1] "firm_name"  "mcca_score"
```

```
combined <- full_join(attributes, score, by = "firm_name") |>
  select(firm_name, everything())


combined <- combined |>
  relocate(binary_has_dei_program, .after = firm_name)
```

From the combined dataset, I would like to know whether or not having a DEI program (binary_has_dei_program) impacts a firm's score on the MCCA scorecard (mcca_score). In this case, having a score greater or equal to 0.60 will be passing.

Around 169 (~20%) of the total firms have a DEI program and 651 (~80%) do not have a DEI program. For firms with a DEI program, 68 (40%) have a passing MCCA score. For firms that do not have DEI programs, 254 (39%) have a passing MCCA score. This suggests that having a DEI program does not necessarily correlate strongly to a passing MCCA score since the passing rates between firms with DEI programs and those without are very small.

Next, I would like to further explore how variables, like firm size (count_attorneys) and diversity metrics (count_female and count_minority) impact the MCCA score. Maybe there will be some underlying patterns there.

```
#DEI programs and mcca_score
combined |>
  count(binary_has_dei_program)
```

```
##   binary_has_dei_program   n
## 1                  FALSE 651
## 2                   TRUE 169
```

```
sum(combined$binary_has_dei_program == TRUE & combined$mcca_score >= 0.60)
```

```
## [1] 68
```

```
68/169
```

```
## [1] 0.4023669
```

```
sum(combined$binary_has_dei_program == FALSE & combined$mcca_score >= 0.60)
```

```
## [1] 254
```

```
254/651
```

```
## [1] 0.390169
```

```
sum(combined$mcca_score > 0.600)
```

```
## [1] 322
```

There seems to be no correlations between MCCA scores (mcca_score) and firm size. Whether or not a firm has a DEI program also does not appear to correlate with the firm size.

```
#group firms by size
summary(combined$count_attorneys)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0   264.8   502.5   502.8   747.0  1000.0
```

```
combined <- combined |>
  mutate(firm_size = cut(count_attorneys,
                         breaks = c(1, 264, 502, 747, 1000),
                         labels = c("Very Small", "Small", "Medium", "Large")))
summary(combined$firm_size)
```

```
## Very Small      Small     Medium      Large       NA's
##        204        205        206        204          1
```

```
#MCCA Score vs Firm Size
ggplot(combined, aes(x = count_attorneys, y = mcca_score, color = binary_has_dei_program)) +
  geom_point() +
  facet_wrap(~ firm_size, scales = "free") +
  labs(x = "Number of Attorneys (Firm Size)",
       y = "MCCA Score",
       title = "MCCA Score vs Firm Size") +
  theme_minimal() +
  scale_fill_OkabeIto()
```

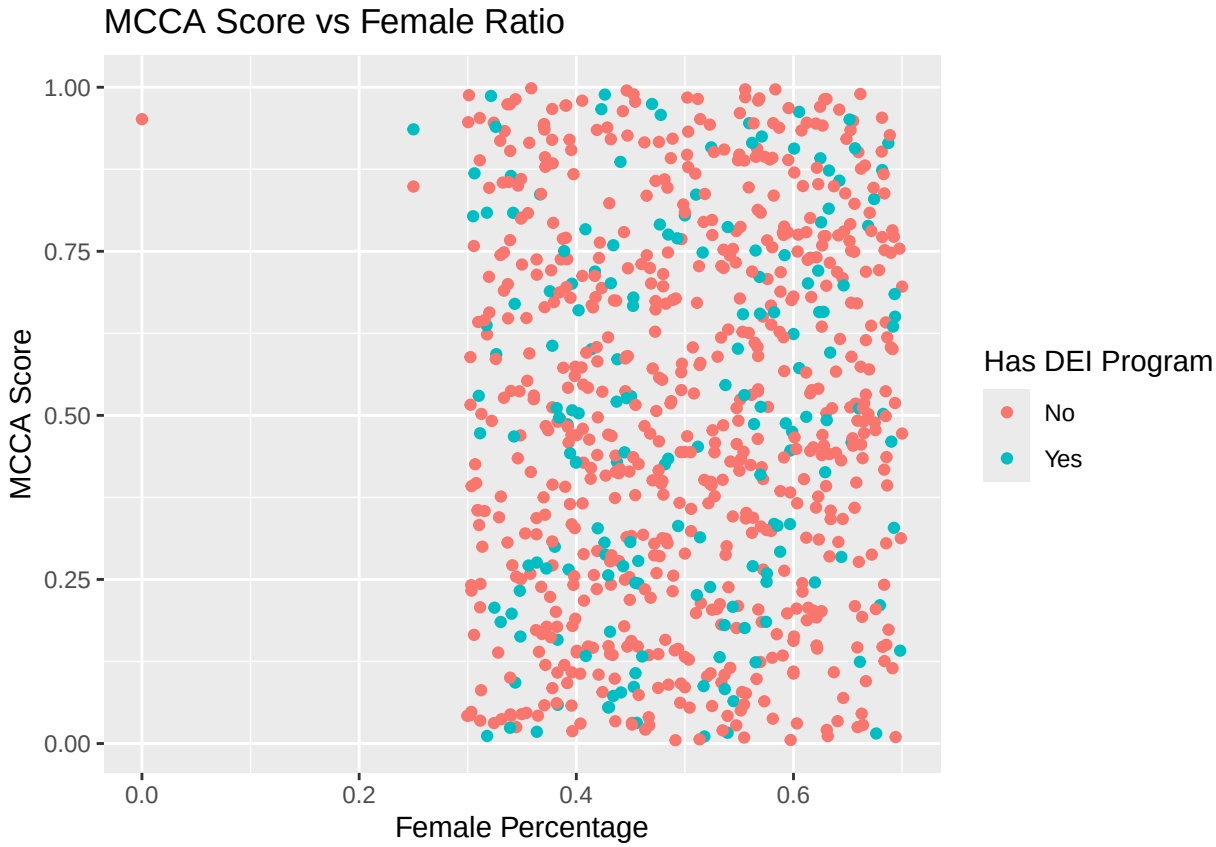
After reviewing the scatter plots for potential correlations between MCCA scores (mcca_score) and the diversity metrics in the dataset (count_female and count_minority), there are no correlations. Whether or not a firm has a DEI program also does not appear to correlate with the percentage of female or minority attorneys. Although, it is worth noting that most firms have 30% or more female attorneys.

```
#female_percent
combined <- combined |>
```

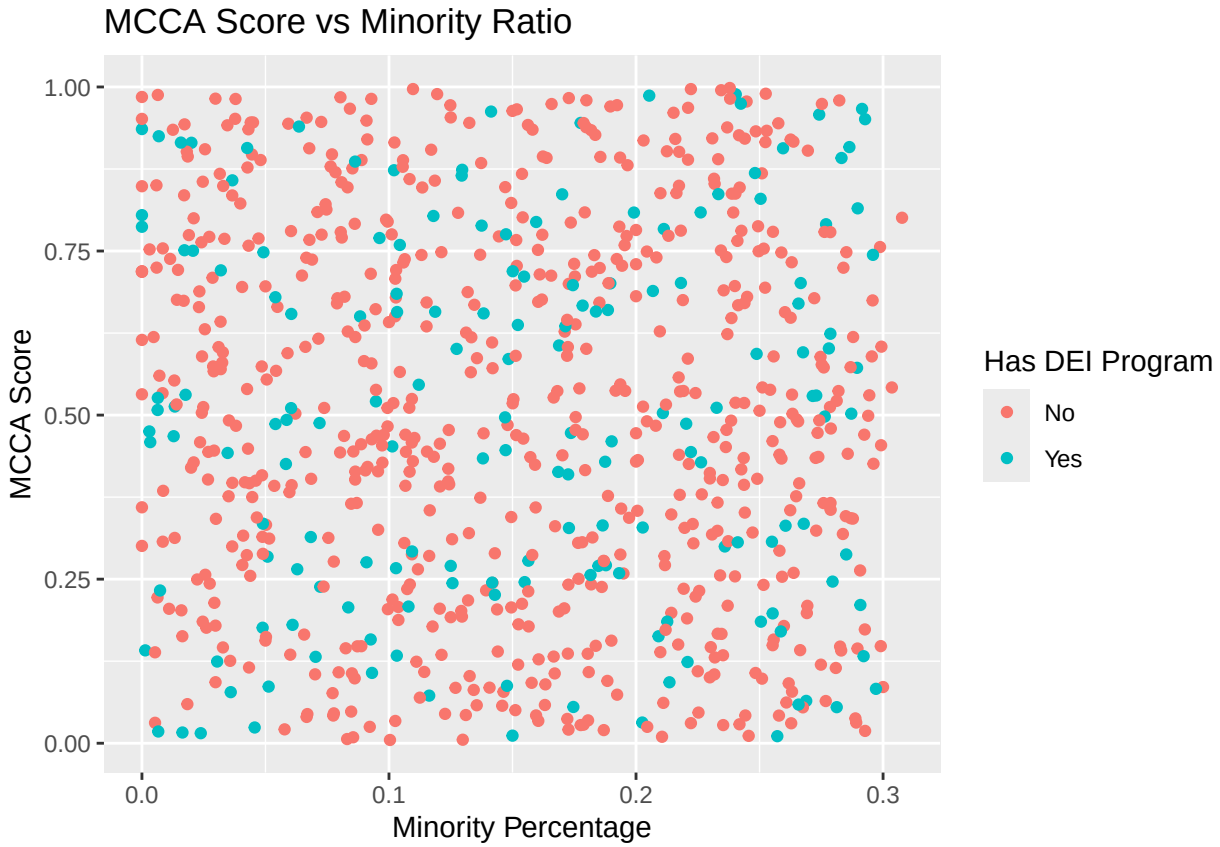
```
  mutate(female_percent = count_female / count_attorneys)

#minority_percent
combined <- combined |>
  mutate(minority_percent = count_minority / count_attorneys)

#MCCA Score vs Female Ratio
ggplot(combined, aes(x = female_percent, y = mcca_score)) +
  geom_point(aes(color = factor(binary_has_dei_program))) +
  labs(x = "Female Percentage",
       y = "MCCA Score",
       title = "MCCA Score vs Female Ratio") +
  scale_color_discrete(name = "Has DEI Program", labels = c("No", "Yes")) +
  scale_fill_OkabeIto()
```




```
#MCCA Score vs Minority Ratio
ggplot(combined, aes(x = minority_percent, y = mcca_score)) +
  geom_point(aes(color = factor(binary_has_dei_program))) +
  labs(x = "Minority Percentage",
       y = "MCCA Score",
       title = "MCCA Score vs Minority Ratio") +
  scale_color_discrete(name = "Has DEI Program", labels = c("No", "Yes")) +
  scale_fill_OkabeIto()
```



So, from what I've gathered, a firm's MCCA score has no correlation to whether or not it has a DEI program, firm size, female ratio, or minority ratio. Now, I have to ask: What does the MCCA score represent and how is it scored? Are there specific criteria or metrics that influence a firm's score that isn't included in this dataset?

```
#side quests
female <- combined |>
  filter(female_percent == 0)

high_scoring_firms <- combined |>
  filter(mcca_score > 0.80)

low_scoring_firms <- combined |>
  filter(mcca_score < 0.20)
```