How Might Race Influence Employee Experience?

An Analysis by Angie Gupta Wilen for LLO-8200 Fall 2020

INTRODUCTION

The year 2020 will be memorable for many reasons, to include economic volatility, the global covid pandemic, and the increased unrest associated with racial and socioeconomic inequity across the United States.

This paper takes aim at a specific element of inequity which is inequity within the workplace. Using a sample data set reflecting common Human Resources roster data, this study explores the impact of race as it influences a number of elements of performance and success within an organization. These “success” elements include performance rating, compensation levels, and employee satisfaction.

This exploration reconciles well with existing literature on this subject. One large meta-analysis found that performance ratings differed between Black and White employees after controlling for a number of variables (Roth, Huffcutt, and Bobko, 2003). Further, BLS data frequently shows both a gender and race disparity in income levels. For example, when the Obama administration asked the EEOC to begin gathering pay data that includes race and gender, they considered 2013 BLS data which showed that Black men, Hispanic men, Black women, and Hispanic women earn 75.1%, 67.2%, 64.0%, and 54% of pay as compared to a White male colleague (Slone, 2016). And finally, another study found that Black and Latino engagement survey respondents reported higher levels of perceived discrimination experiences, an element that was negatively correlated with engagement (Jones, Ni, & Wilson, 2009).

While these more macro level studies support a rationale for exploring how race might correlate with employee success metrics, this analysis seeks to take these broad themes and test them in a more micro setting. That is, this paper explores whether we find evidence of these themes within a single employer, based on a point-in-time HR data set.

Before beginning this analysis, relevant libraries are loaded here.

```r
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------
-------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.1     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------------
-------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(rvest)
```

```
## Loading required package: xml2
```

```
##
## Attaching package: 'rvest'
```

```
## The following object is masked from 'package:purrr':
##
##     pluck
```

```
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```r
library(tigris)
```

```
## To enable
## caching of data, set `options(tigris_use_cache = TRUE)` in your R script or .Rprofile.
```

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

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

```r
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

```r
library(readxl)

library(ggplot2)
library(readr)
library('tidyr')
library('dplyr')
library(dplyr)

library(ModelMetrics)
```

```
##
## Attaching package: 'ModelMetrics'
```

```
## The following object is masked from 'package:base':
##
##     kappa
```

```r
library(modelr)
```

```
##
## Attaching package: 'modelr'
```

```
## The following objects are masked from 'package:ModelMetrics':
##
##     mae, mse, rmse
```

```r
library(knitr)
```

DATA

Traditional Human Resources (HR) data can entail confidential and sensitive data points. Date of birth, age, compensation, and performance rating, for example, will rarely if ever be publicly retrievable data fields from any HR information system.

Happily, the website Kaggle hosts a wealth of datasets on a variety of topics, including HR. From here, I acquired the "HRDataset_v13" at this address: https://www.kaggle.com/rhuebner/human-resources-data-set?select=HRDataset_v13.csv (https://www.kaggle.com/rhuebner/human-resources-data-set?select=HRDataset_v13.csv). The dataset revolves around a fictitious company, Dental Magic, but reflects similar data fields and observations one might find in many HR data sets. This synthetic data set serves as an excellent opportunity to avoid violating privacy and data sensitivity issues associated with real HR data sets while also affording an opportunity to analyze the research questions at hand.

I begin by loading the data to explore its format and the variables available to work with.

```r
DentalMagicHR<-read_csv("C:/Users/angie/Dropbox/Vanderbilt/LLO8200-master/Research Project/Denta
lMagicHR.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   EmpID = col_double(),
##   MarriedID = col_double(),
##   MaritalStatusID = col_double(),
##   GenderID = col_double(),
##   EmpStatusID = col_double(),
##   DeptID = col_double(),
##   PerfScoreID = col_double(),
##   FromDiversityJobFairID = col_double(),
##   PayRate = col_double(),
##   Termd = col_double(),
##   PositionID = col_double(),
##   Zip = col_double(),
##   ManagerID = col_double(),
##   EngagementSurvey = col_double(),
##   EmpSatisfaction = col_double(),
##   SpecialProjectsCount = col_double(),
##   DaysLateLast30 = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
save(DentalMagicHR,file="DentalMagicHR.Rdata")
```

In reviewing the data (now called DentalMagicHR), we see that the data is already tidy. Specifically, each row represents a unique observation, or in this case, the name and characteristics associated with a single employee. Additionally, each column has a title that is intuitive in what variable it represents.

The data includes 35 unique variables and 310 observations. Of these variables, this analysis will focus on the following.

1. Race (titled “RaceDesc” in this dataset). This discrete independent variable has five values: American Indian, Asian, Black, Hispanic, Two or more, and White.

2. Compensation (titled “PayRate” in this dataset). THis dependent variable is a continuous variable ranging from 14 (low pay) to 80 (highest pay).

3. Performance Rating (titled “PerformanceScoreID” in this dataset). This dependent variable has four possible values including “1” which equals “PIP”, “2” which equals “Needs Improvement”, “3” which equals “Fully Meets”, and “4” which equals “Exceeds”).

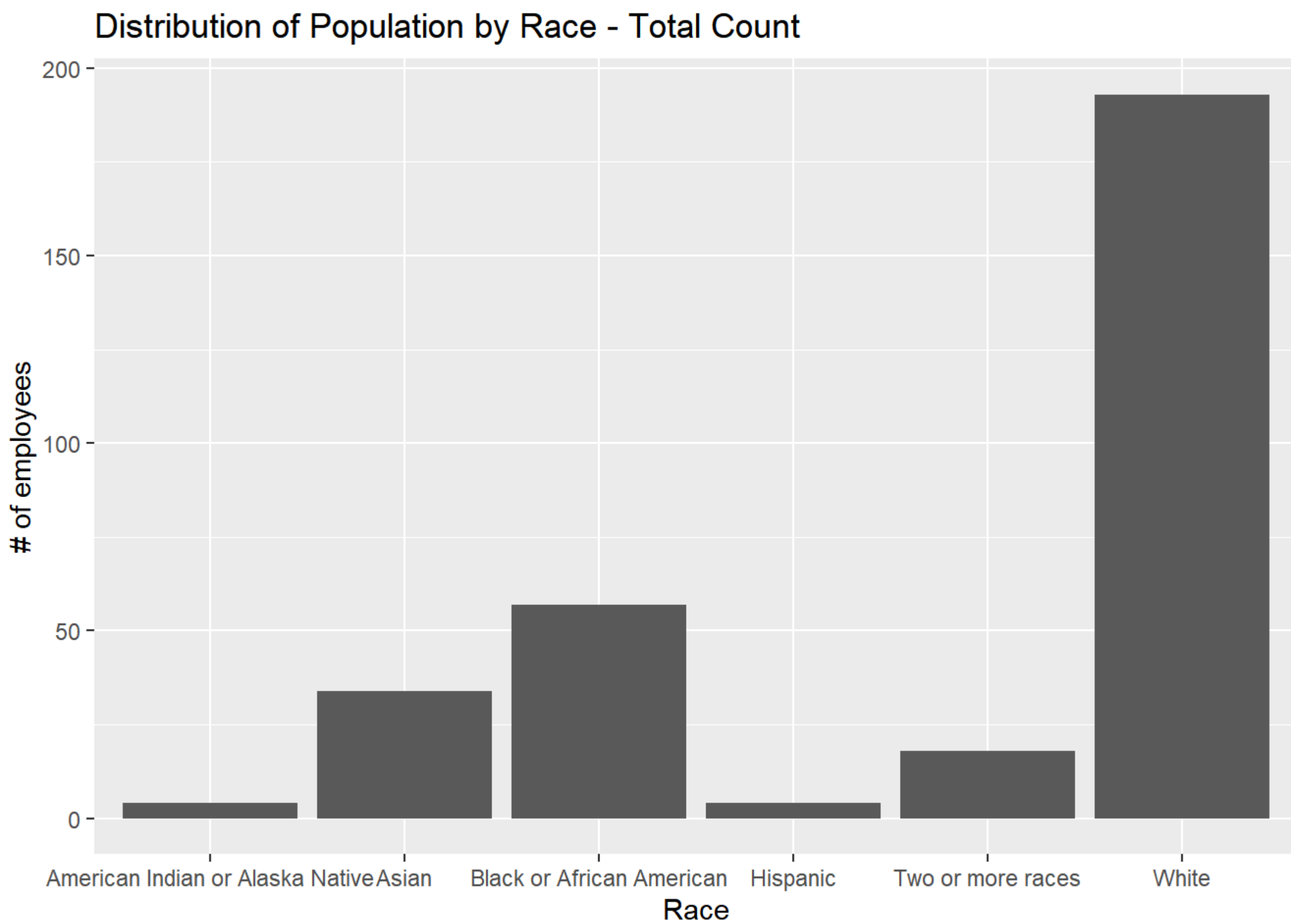
4. Employee Satisfaction (titled “EmpSatisfaction” in this dataset). This dependent variable is a discrete variable ranging from 1 (unsatisfied) to 5 (very satisfied).

EXPLORATORY DATA ANALYSIS

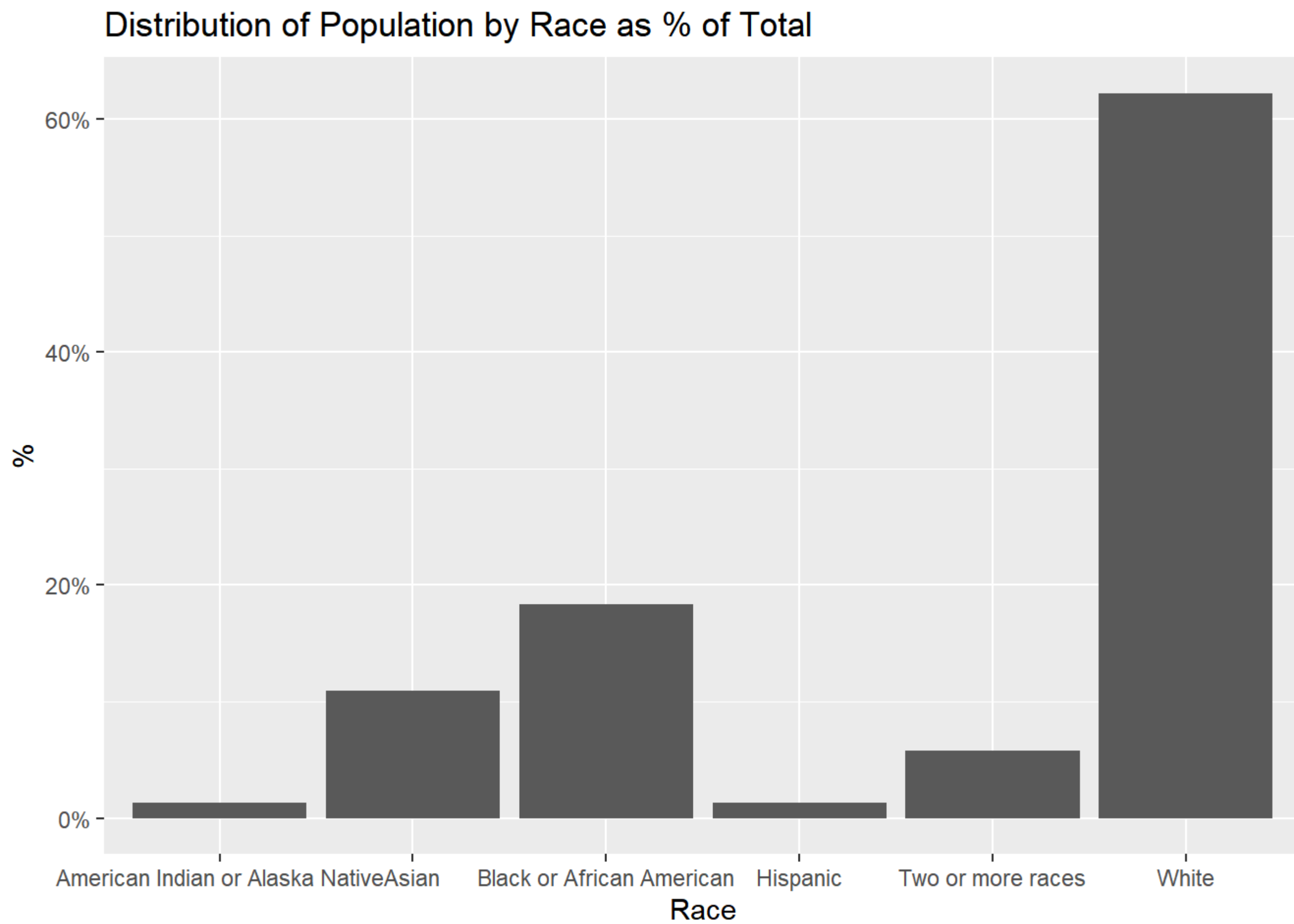
Prior to building models to better understand whether (and if so how) race influences performance ratings, compensation, and employee satisfaction, I conducted an exploratory data analysis to better understand the predictive potential of these variables.

Prior to exploring the relationship between race and other variables, let's first understand the distribution of race in this population.

```
DentalMagicHR$RaceDesc<-as.factor(DentalMagicHR$RaceDesc)


## A visual to show the total DentalMagic employee population headcount by race.
gg <-ggplot(DentalMagicHR,aes(x=RaceDesc)) +
      geom_bar() +
      ggtitle("Distribution of Population by Race - Total Count") +
      ylab("# of employees") +
      xlab("Race")

gg
```




```
##A second visual to shoDistribution of Population by Race as % of Totalw the total DentalMagic
 racial representation as a percentage (%) of total population.
gg2<-ggplot(DentalMagicHR,aes(RaceDesc))+
  geom_bar(aes(y=(..count..)/sum(..count..)))+
  scale_y_continuous(labels=scales::percent)+
  ggtitle("Distribution of Population by Race as % of Total")+
  ylab("%")+
  xlab("Race")

gg2
```



I can see that while six races are represented at this organization, the organization is predominantly White (with nearly 200 White employees), followed in population size by Black/African American employees (over 50), and then Asian employees (over 30). The relatively low representation of employees who identify as American Indian or Alaskan Native (AIAN), Hispanic, and Two or more races (2+) should be kept in mind as their small sample size may skew data analysis to follow.

While one could argue that combining these small data sets, (AIAN, Hispanic, 2+) for our analysis, I am choosing not to. While this limits insights derived in our analysis about these populations, I am consciously choosing to explore their “success” experience as disparate groups. In sum, I choose to limit the power of my analysis for these races in return for an opportunity to see the depth and variability in the data.

Now, let’s conduct some exploratory analysis to understand the relationship between race and the three variables - performance, compensation, and satisfaction.

First, I explored the distribution of performance ratings as a function of race.

```
## First, we'll look at a summary table of the distribution of performance ratings by race.
prop.table(table(DentalMagicHR$RaceDesc,DentalMagicHR$PerfScoreID),margin=1)
```

```
##
##                                            1          2          3          4
##   American Indian or Alaska Native 0.00000000 0.00000000 0.50000000 0.50000000
##   Asian                            0.02941176 0.05882353 0.79411765 0.11764706
##   Black or African American        0.01754386 0.14035088 0.75438596 0.08771930
##   Hispanic                         0.00000000 0.25000000 0.75000000 0.00000000
##   Two or more races                0.11111111 0.00000000 0.72222222 0.16666667
##   White                            0.04145078 0.03626943 0.80310881 0.11917098
```

```
## We can see with a summary view that there is a difference in average performance scores when
 analyzed by race.
DentalMagicHR%>%group_by(RaceDesc)%>%summarise(mean_performance_by_race=mean(PerfScoreID))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

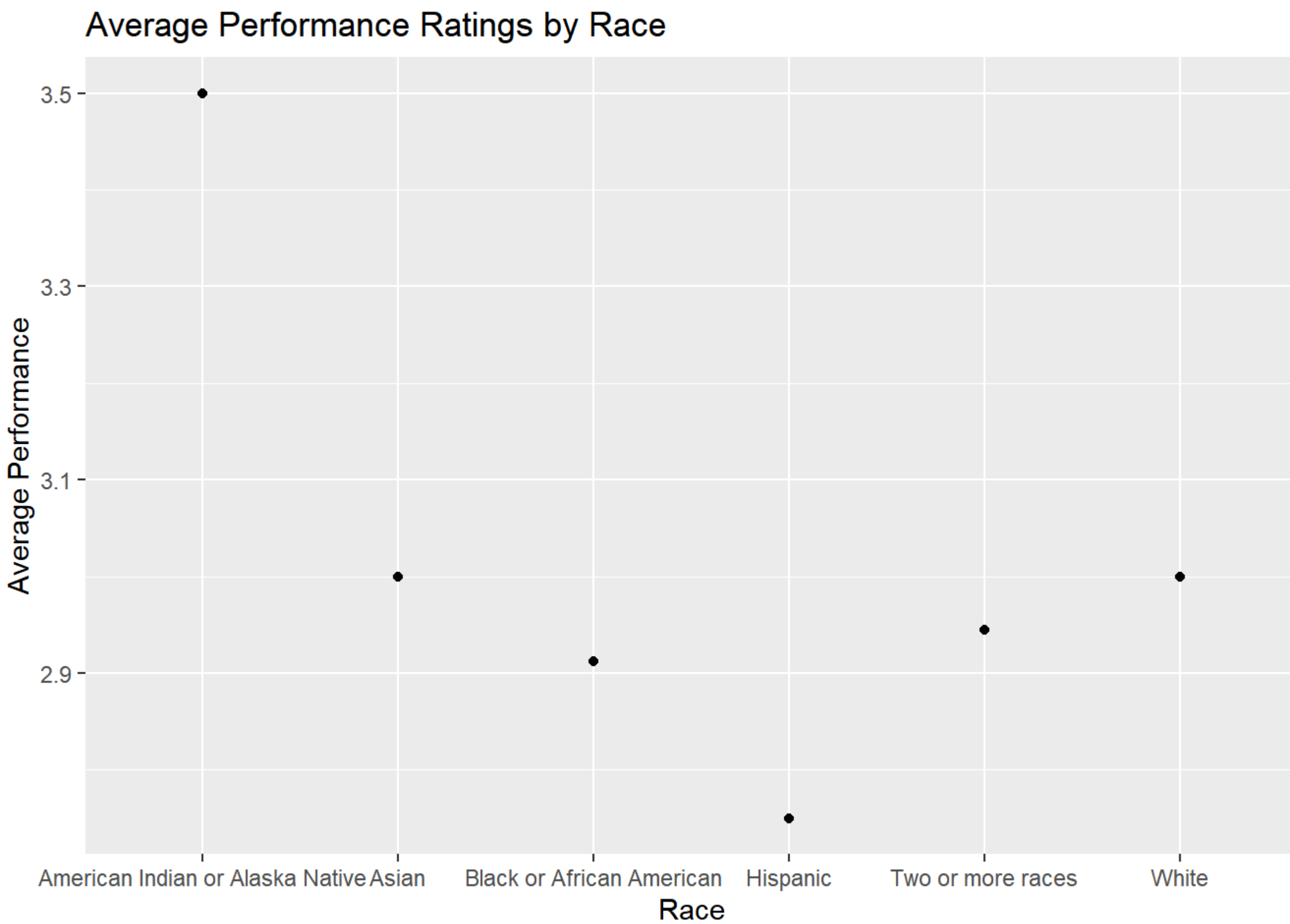
```
## # A tibble: 6 x 2
##   RaceDesc                         mean_performance_by_race
##   <fct>                                               <dbl>
## 1 American Indian or Alaska Native                     3.5
## 2 Asian                                                3
## 3 Black or African American                            2.91
## 4 Hispanic                                             2.75
## 5 Two or more races                                    2.94
## 6 White                                                3
```

Looking at the distribution of performance by race, we can see that there are differences in distribution. Races with the highest proportion of high performers (rated 4) are AIAN, Two or more races, White, and Asian. Races with the highest proportion of PIPs include the Two or more race population, and White population. Another view will show the average performance rating by race to consider this analysis in a different way.

```
## We can see with a summary view that there is a difference in average performance scores when
 analyzed by race.
DentalMagicHR%>%group_by(RaceDesc)%>%mutate(mean_performance_by_race=mean(PerfScoreID))
```

```
## # A tibble: 310 x 36
## # Groups:   RaceDesc [6]
##    Employee_Name  EmpID MarriedID MaritalStatusID GenderID EmpStatusID DeptID
##    <chr>          <dbl>     <dbl>           <dbl>    <dbl>       <dbl>  <dbl>
##  1 Brown, Mia    1.10e9         1               1        0           1      1
##  2 LaRotonda, W~ 1.11e9         0               2        1           1      1
##  3 Steans, Tyro~ 1.30e9         0               0        1           1      1
##  4 Howard, Este~ 1.21e9         1               1        0           1      1
##  5 Singh, Nan    1.31e9         0               0        0           1      1
##  6 Smith, Leigh~ 7.11e8         1               1        0           5      1
##  7 Bunbury, Jes~ 1.50e9         1               1        0           5      6
##  8 Carter, Mich~ 1.40e9         0               0        0           1      6
##  9 Dietrich, Je~ 1.41e9         0               0        0           1      6
## 10 Digitale, Al~ 1.31e9         1               1        1           1      6
## # ... with 300 more rows, and 29 more variables: PerfScoreID <dbl>,
## #   FromDiversityJobFairID <dbl>, PayRate <dbl>, Termd <dbl>, PositionID <dbl>,
## #   Position <chr>, State <chr>, Zip <dbl>, DOB <chr>, Sex <chr>,
## #   MaritalDesc <chr>, CitizenDesc <chr>, HispanicLatino <chr>, RaceDesc <fct>,
## #   DateofHire <chr>, DateofTermination <chr>, TermReason <chr>,
## #   EmploymentStatus <chr>, Department <chr>, ManagerName <chr>,
## #   ManagerID <dbl>, RecruitmentSource <chr>, PerformanceScore <chr>,
## #   EngagementSurvey <dbl>, EmpSatisfaction <dbl>, SpecialProjectsCount <dbl>,
## #   LastPerformanceReview_Date <chr>, DaysLateLast30 <dbl>,
## #   mean_performance_by_race <dbl>
```

```
## Let's also look at this as a visual.
DentalMagicHR<-DentalMagicHR%>%group_by(RaceDesc)%>%
  mutate(mean_performance_by_race=mean(PerfScoreID))%>%
  ungroup()

ggperf <-ggplot(DentalMagicHR,aes(x=RaceDesc,y=mean_performance_by_race)) +
       geom_point() +
       ggtitle("Average Performance Ratings by Race") +
       ylab("Average Performance") +
      xlab("Race")

ggperf
```



The table and graph clearly shows a difference in average performance rating by race. Of note… - Here, we see that the Hispanic population has the lowest ratings on average, while AIAN show the highest average ratings. While the visual highlights how wide a gap exists with respect to the Hispanic and American Indian population, we must remember that these two populations are also the smallest in terms of representation as percentage of overall employee population. This would mean that fewer employees are driving the average performance rating number for theses two populations, as compared to others. One cannot put too much weight to this insight.

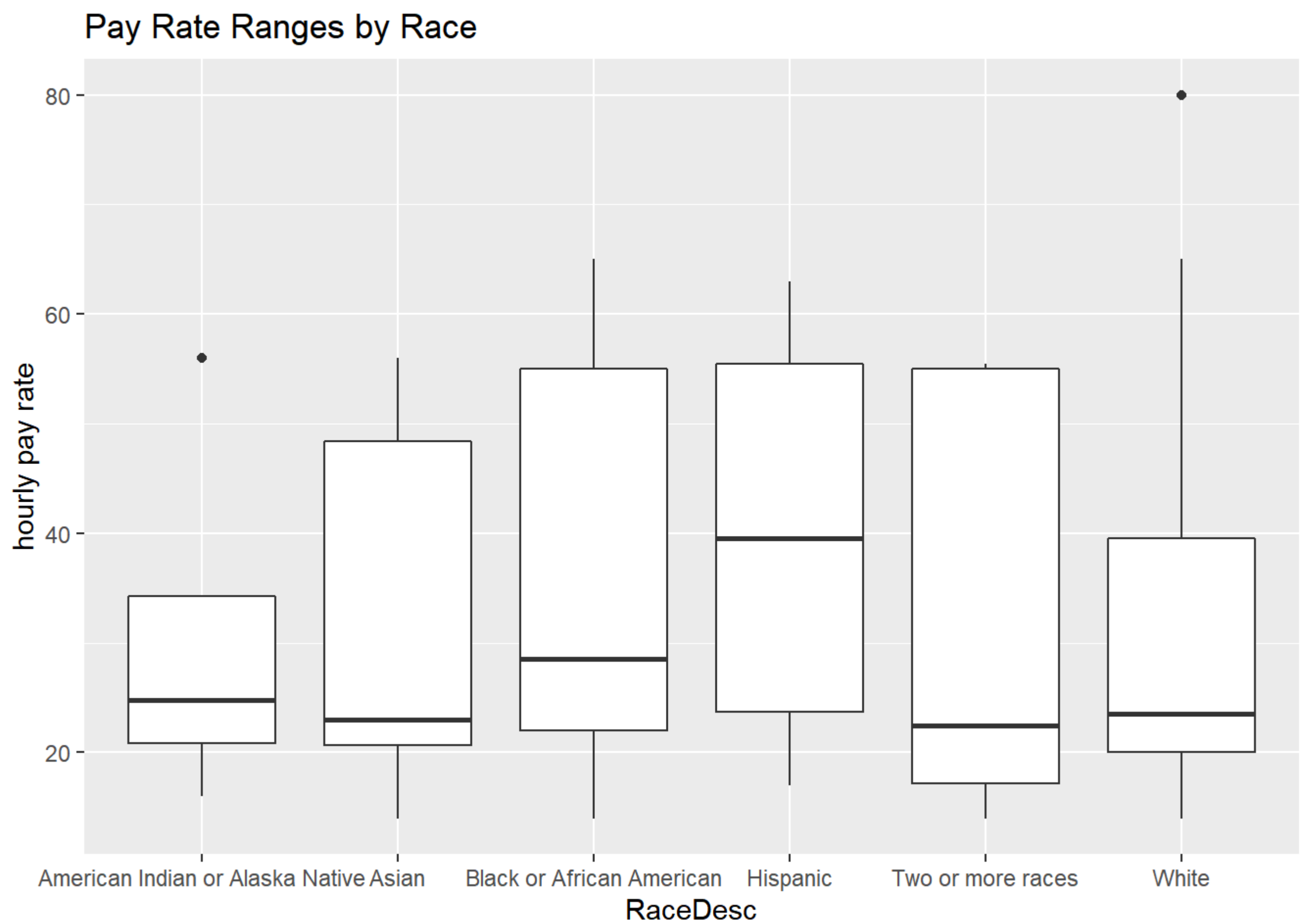
- Additionally, because 3.0 represents “Meets Expectations”, we see that Black | African American as well as Two or more race population, on average, have a below average performance rating.

Second, I explored the distribution of compensation levels as a function of race.

```
ggpay <-ggplot(DentalMagicHR,aes(x=RaceDesc,y=PayRate)) +
       geom_boxplot() +
       ggtitle("Pay Rate Ranges by Race") +
       ylab("hourly pay rate")
       xlab("race")
```

```
## $x
## [1] "race"
##
## attr(,"class")
## [1] "labels"
```

```
ggpay
```



Here, we note a few things.
- The median pay for the Hispanic population is much higher than all other races, but again, due to small sample size I cannot put weight into this observation.

- Median pay for the “Two or more races” population is the lowest.
- We have a few outliers in the data that are highly compensated (one AIAN observation and one White observation). The AIAN observation, additionally, is from a small sample size of AIAN to begin with adding to its outlier status.
- Black and Two or more race populations share relatively equal pay distributions at the 25th, 50th, and 75th percentile (as does Hispanic, but again, low sample size here).
- Asian, and White populations have a lower median pay, as does the Two or more races population (despite this population’s 75th percentile rivaling Hispanic and African American populations).

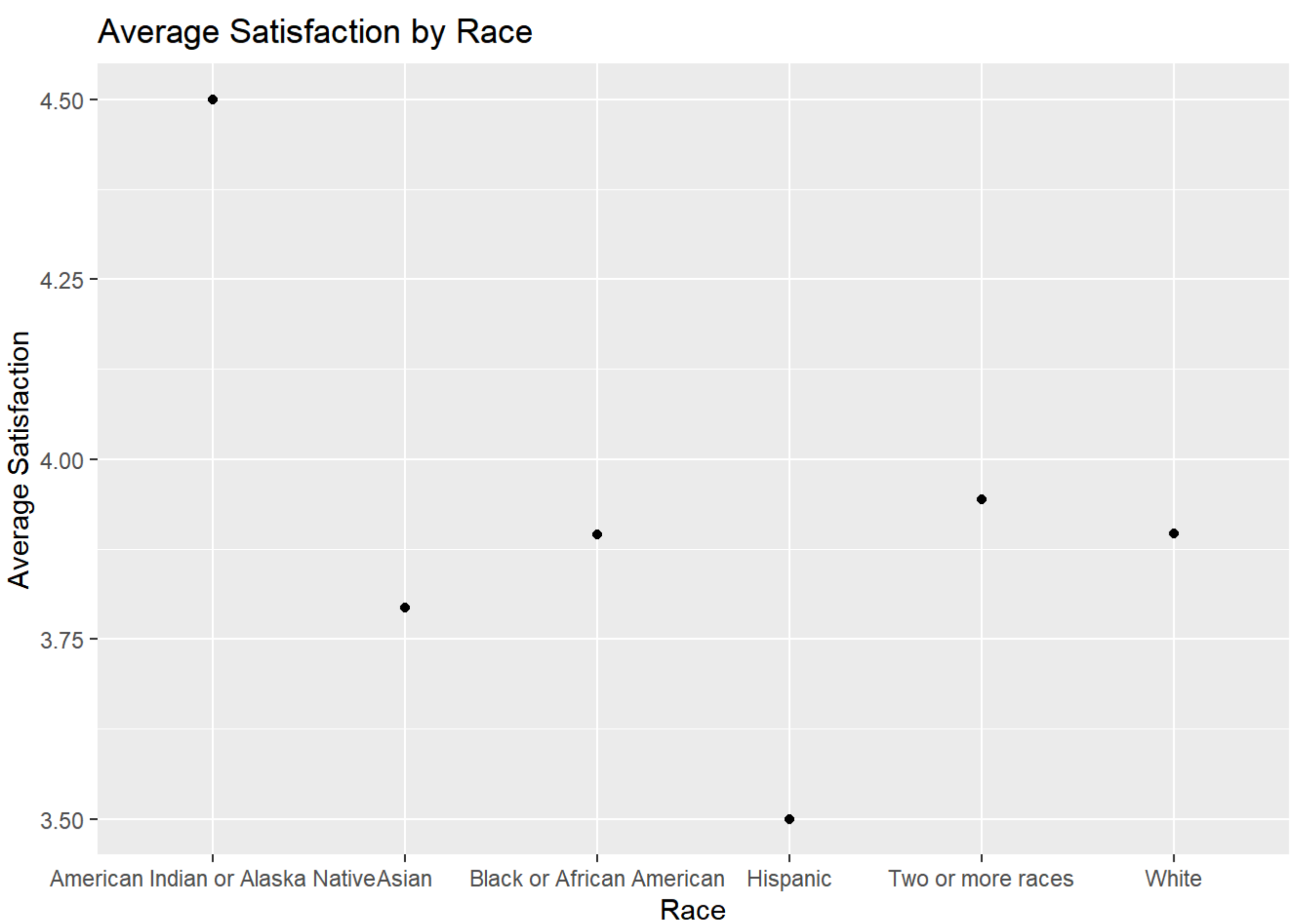
Finally, I explored the distribution of employee satisfaction as a function of race.

```
## I first want to understand the distribution of satisfaction  by race.
## We can see with a summary view that there is a difference in average satisfaction scores when analyzed by race.


DentalMagicHR%>%group_by(RaceDesc)%>%summarise(mean_satisfaction_by_race=mean(EmpSatisfaction))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 2
##   RaceDesc                         mean_satisfaction_by_race
##   <fct>                                                <dbl>
## 1 American Indian or Alaska Native                      4.5
## 2 Asian                                                 3.79
## 3 Black or African American                             3.89
## 4 Hispanic                                              3.5
## 5 Two or more races                                     3.94
## 6 White                                                 3.90
```

```
## Let's also look at this as a visual.
DentalMagicHR<-DentalMagicHR%>%group_by(RaceDesc)%>%
  mutate(mean_satisfaction_by_race=mean(EmpSatisfaction))%>%
  ungroup()

ggsat <-ggplot(DentalMagicHR,aes(x=RaceDesc,y=mean_satisfaction_by_race)) +
       geom_point() +
       ggtitle("Average Satisfaction by Race") +
       ylab("Average Satisfaction") +
       xlab("Race")
ggsat
```



Here, I quickly note that the average satisfaction reported by employees differs by race. Specifically,

- The small sample size Hispanic population on average reports a far lower satisfaction rating (close to 3.5) as compared to others, whereas the American Indian | Alaskan population reports the highest satisfaction ratings, at around 4.5. Again, while the visual highlights how wide a gap these two populations show, we must remember that these two populations are also the smallest in terms of representation as percentage of overall employee population. This would mean that fewer employees are driving the satisfaction ratings for theses two populations, as compared to others.

- Generally, for most of the population (represented by Asian, Black, White, and Two or more race identities) show little variance. Therefore, while I will explore whether race predicts satisfaction, I am likely not to find significance, given that the satisfaction variable has little range.

ANALYTIC PLAN - APPROACH

I have built models to test predictions based on key insights and assumptions from the exploratory data analysis. These key insights and assumptions are as follows.

1. Assumption: success is defined as high performance ratings, high satisfaction, and high compensation.

2. Assumption: The racial distribution at DentalMagicHR is skewed. The organization is predominantly White (>60%), followed by Black (>15%), Asian (>10%), and two or more races (>5%).

3. Assumption: Hispanic and AIAN represent small percentages, conclusions about them must consider this. Must limit any conclusions drawn as simply evidence for further research / dialogue.

Ex: Biggest variability in avg performance rating by race was in Hispanic and AIAN populations Ex: Biggest variability in employee satisfaction was within Hispanic and AIAN populations

4. Insight: Asian, White, and 2+ race population have the lowest median pay rates. They represent ~80% of the population, suggesting 20% of the population (made up of AIAN, Black, and Hispanic populations) have higher average pay.

5. Insight: Little variability in the satisfaction averages across race. This may limit any model's ability to prove that race influences employee satisfaction

ANALYTIC PLAN - EXPLORATION

Based on the work thus far, I will develop models to explore the following.

- Exploration 1: How well does race predict compensation levels?
- Exploration 2: How well does race predict employee satisfaction levels?
- Exploration 3: How well does race predict performance ratings?

To do this, I will create variables of success as defined as follows.

- Employee satisfaction levels, using the EmpSatisfaction variable, can interpret these ratings as 1=strongly dissatisfied; 2=dissatisfied; 3=neutral; 4=satisfied; 5=highly satisfied.

- Employee Performance ratings uses the "PerfScoreID" variable and can be interpreted as 1=Performance Improvement Plan (PIP); 2= Needs Improvement; 3 = Fully Meets; 4=Exceeds

- Pay rate, using the PayRate variable, is a continuous variable defining the hourly pay of each individual in the roster.

Let's start with Exploration 1 - How well does race predict compensation levels?

```
##Creating a linear model to test the relationship between compensation (dependent) and race (in
dependent).
##Because gender inequity in pay is a well documented phenomenon, I am controlling for this vari
able.

DentalMagicHR$race.f <-factor(DentalMagicHR$RaceDesc)
is.factor(DentalMagicHR$race.f)
```

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

```
lin_mod_comp<-lm(PayRate~relevel(race.f, ref="White")+as.factor(Sex), data=DentalMagicHR)

summary(lin_mod_comp)
```

```
##
## Call:
## lm(formula = PayRate ~ relevel(race.f, ref = "White") + as.factor(Sex),
##     data = DentalMagicHR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -23.642 -11.249  -6.159  12.905  51.841
##
## Coefficients:
##                                                                Estimate
## (Intercept)                                                     28.1589
## relevel(race.f, ref = "White")American Indian or Alaska Native   0.1711
## relevel(race.f, ref = "White")Asian                              1.6349
## relevel(race.f, ref = "White")Black or African American          5.3934
## relevel(race.f, ref = "White")Hispanic                           8.5236
## relevel(race.f, ref = "White")Two or more races                  1.5150
## as.factor(Sex)M                                                  4.0901
##                                                                Std. Error
## (Intercept)                                                        1.3306
## relevel(race.f, ref = "White")American Indian or Alaska Native     7.6941
## relevel(race.f, ref = "White")Asian                                2.8337
## relevel(race.f, ref = "White")Black or African American            2.2958
## relevel(race.f, ref = "White")Hispanic                             7.7135
## relevel(race.f, ref = "White")Two or more races                    3.7539
## as.factor(Sex)M                                                    1.7539
##                                                                t value Pr(>|t|)
## (Intercept)                                                     21.162   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   0.022   0.9823
## relevel(race.f, ref = "White")Asian                              0.577   0.5644
## relevel(race.f, ref = "White")Black or African American          2.349   0.0195
## relevel(race.f, ref = "White")Hispanic                           1.105   0.2700
## relevel(race.f, ref = "White")Two or more races                  0.404   0.6868
## as.factor(Sex)M                                                  2.332   0.0204
##
## (Intercept)                                                    ***
## relevel(race.f, ref = "White")American Indian or Alaska Native
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American        *
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## as.factor(Sex)M                                                *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.23 on 303 degrees of freedom
## Multiple R-squared:  0.03901,    Adjusted R-squared:  0.01998
## F-statistic:  2.05 on 6 and 303 DF,  p-value: 0.05904
```

```
DentalMagicHR<-DentalMagicHR%>%
  add_predictions(lin_mod_comp)%>% ## Add in predictions from the model
  rename(predcomp_lm=pred)
```

In Exploration 1, after controlling for Gender, I do find some significance. Specifically, holding gender constant, race identity as Black is significantly positively correlated to PayRate as compared to White employees. Holding gender constant, a Black employee at DentalMagicHR makes on average 5.39 more on their hourly pay rate as compared to white employees. And based on an R-square value of .03901, this accounts for approximately 4% of the variability in Pay Rate. I've added the predicted compensation level from this model to my data, calling this data "predcomp_lm".

Let's continue with exploration 2: How well does race predict employee satisfaction levels?

```
##Creating a linear model to test the relationship between satisfaction and race (independent).
While one could argue both that satisfaction is continuous, one could also argue that it is not
 as employees can only enter sentiment from 1-5.  I'll treat it as continuous at this stage as i
t makes more intuitive sense.

## I initially test using the linear model using solely the race variable.  I find no significan
ce.
lin_mod_sats<-lm(EmpSatisfaction~relevel(race.f, ref="White"),data=DentalMagicHR)

summary(lin_mod_sats)
```

```
##
## Call:
## lm(formula = EmpSatisfaction ~ relevel(race.f, ref = "White"),
##     data = DentalMagicHR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.8964 -0.8964  0.1036  1.1036  1.2059
##
## Coefficients:
##                                                                   Estimate
## (Intercept)                                                       3.896373
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.603627
## relevel(race.f, ref = "White")Asian                             -0.102255
## relevel(race.f, ref = "White")Black or African American         -0.001636
## relevel(race.f, ref = "White")Hispanic                          -0.396373
## relevel(race.f, ref = "White")Two or more races                  0.048071
##                                                                 Std. Error
## (Intercept)                                                       0.065770
## relevel(race.f, ref = "White")American Indian or Alaska Native   0.461565
## relevel(race.f, ref = "White")Asian                               0.169943
## relevel(race.f, ref = "White")Black or African American           0.137741
## relevel(race.f, ref = "White")Hispanic                            0.461565
## relevel(race.f, ref = "White")Two or more races                   0.225182
##                                                                 t value Pr(>|t|)
## (Intercept)                                                       59.242   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.308    0.192
## relevel(race.f, ref = "White")Asian                              -0.602    0.548
## relevel(race.f, ref = "White")Black or African American          -0.012    0.991
## relevel(race.f, ref = "White")Hispanic                           -0.859    0.391
## relevel(race.f, ref = "White")Two or more races                   0.213    0.831
##
## (Intercept)                                                     ***
## relevel(race.f, ref = "White")American Indian or Alaska Native
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9137 on 304 degrees of freedom
## Multiple R-squared:  0.009645,   Adjusted R-squared:  -0.006643
## F-statistic: 0.5921 on 5 and 304 DF,  p-value: 0.706
```

```
## I now test after controlling variables that intuitively make sense as potential influencers t
o engagement, including employee rating(PerformanceScore) and ManagerID (given literature sugges
ting managers have substantial influence on individual engagement)
lin_mod_sats<-lm(EmpSatisfaction~relevel(race.f,ref="White")+
                   as.factor(PerformanceScore)+
                  as.factor(ManagerID),data=DentalMagicHR)

summary(lin_mod_sats)
```

```
##
## Call:
## lm(formula = EmpSatisfaction ~ relevel(race.f, ref = "White") +
##     as.factor(PerformanceScore) + as.factor(ManagerID), data = DentalMagicHR)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.5239 -0.8184  0.0000  0.8337  2.6333
##
## Coefficients:
##                                                                Estimate
## (Intercept)                                                    3.686300
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.459859
## relevel(race.f, ref = "White")Asian                           -0.175409
## relevel(race.f, ref = "White")Black or African American       -0.008819
## relevel(race.f, ref = "White")Hispanic                        -0.369076
## relevel(race.f, ref = "White")Two or more races                0.078394
## as.factor(PerformanceScore)Fully Meets                        -0.154125
## as.factor(PerformanceScore)Needs Improvement                  -0.529251
## as.factor(PerformanceScore)PIP                                -1.746172
## as.factor(ManagerID)2                                          0.146190
## as.factor(ManagerID)3                                          0.643235
## as.factor(ManagerID)4                                          0.296453
## as.factor(ManagerID)5                                          0.805879
## as.factor(ManagerID)6                                         -0.088100
## as.factor(ManagerID)7                                          0.535173
## as.factor(ManagerID)9                                         -0.532175
## as.factor(ManagerID)10                                         0.739398
## as.factor(ManagerID)11                                         0.426545
## as.factor(ManagerID)12                                         0.127553
## as.factor(ManagerID)13                                         0.536911
## as.factor(ManagerID)14                                         0.465431
## as.factor(ManagerID)15                                         0.929140
## as.factor(ManagerID)16                                         0.460012
## as.factor(ManagerID)17                                         0.583636
## as.factor(ManagerID)18                                         0.305232
## as.factor(ManagerID)19                                         0.550230
## as.factor(ManagerID)20                                         0.462198
## as.factor(ManagerID)21                                         0.634155
## as.factor(ManagerID)22                                         0.583808
## as.factor(ManagerID)30                                         1.467825
## as.factor(ManagerID)39                                         0.594072
##                                                               Std. Error
## (Intercept)                                                     0.390842
## relevel(race.f, ref = "White")American Indian or Alaska Native   0.450115
## relevel(race.f, ref = "White")Asian                             0.177256
## relevel(race.f, ref = "White")Black or African American         0.138738
## relevel(race.f, ref = "White")Hispanic                          0.450416
## relevel(race.f, ref = "White")Two or more races                 0.226617
## as.factor(PerformanceScore)Fully Meets                          0.162918
## as.factor(PerformanceScore)Needs Improvement                    0.261388
## as.factor(PerformanceScore)PIP                                  0.297584
## as.factor(ManagerID)2                                           0.406323
```

```
## as.factor(ManagerID)3                                              0.939991
## as.factor(ManagerID)4                                              0.408897
## as.factor(ManagerID)5                                              0.488621
## as.factor(ManagerID)6                                              0.559653
## as.factor(ManagerID)7                                              0.422564
## as.factor(ManagerID)9                                              0.704628
## as.factor(ManagerID)10                                             0.455572
## as.factor(ManagerID)11                                             0.402234
## as.factor(ManagerID)12                                             0.406944
## as.factor(ManagerID)13                                             0.465975
## as.factor(ManagerID)14                                             0.401237
## as.factor(ManagerID)15                                             0.613394
## as.factor(ManagerID)16                                             0.398020
## as.factor(ManagerID)17                                             0.427523
## as.factor(ManagerID)18                                             0.398342
## as.factor(ManagerID)19                                             0.400626
## as.factor(ManagerID)20                                             0.398812
## as.factor(ManagerID)21                                             0.427687
## as.factor(ManagerID)22                                             0.402789
## as.factor(ManagerID)30                                             0.930685
## as.factor(ManagerID)39                                             0.428004
##                                                                  t value Pr(>|t|)
## (Intercept)                                                        9.432  < 2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.022   0.3079
## relevel(race.f, ref = "White")Asian                               -0.990   0.3233
## relevel(race.f, ref = "White")Black or African American          -0.064   0.9494
## relevel(race.f, ref = "White")Hispanic                            -0.819   0.4133
## relevel(race.f, ref = "White")Two or more races                    0.346   0.7297
## as.factor(PerformanceScore)Fully Meets                            -0.946   0.3450
## as.factor(PerformanceScore)Needs Improvement                      -2.025   0.0439
## as.factor(PerformanceScore)PIP                                    -5.868 1.28e-08
## as.factor(ManagerID)2                                              0.360   0.7193
## as.factor(ManagerID)3                                              0.684   0.4944
## as.factor(ManagerID)4                                              0.725   0.4691
## as.factor(ManagerID)5                                              1.649   0.1002
## as.factor(ManagerID)6                                             -0.157   0.8750
## as.factor(ManagerID)7                                              1.266   0.2064
## as.factor(ManagerID)9                                             -0.755   0.4508
## as.factor(ManagerID)10                                             1.623   0.1057
## as.factor(ManagerID)11                                             1.060   0.2899
## as.factor(ManagerID)12                                             0.313   0.7542
## as.factor(ManagerID)13                                             1.152   0.2502
## as.factor(ManagerID)14                                             1.160   0.2471
## as.factor(ManagerID)15                                             1.515   0.1310
## as.factor(ManagerID)16                                             1.156   0.2488
## as.factor(ManagerID)17                                             1.365   0.1733
## as.factor(ManagerID)18                                             0.766   0.4442
## as.factor(ManagerID)19                                             1.373   0.1708
## as.factor(ManagerID)20                                             1.159   0.2475
## as.factor(ManagerID)21                                             1.483   0.1393
## as.factor(ManagerID)22                                             1.449   0.1484
## as.factor(ManagerID)30                                             1.577   0.1159
## as.factor(ManagerID)39                                             1.388   0.1663
##
```

```
## (Intercept)                                                         ***
## relevel(race.f, ref = "White")American Indian or Alaska Native
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## as.factor(PerformanceScore)Fully Meets
## as.factor(PerformanceScore)Needs Improvement                        *
## as.factor(PerformanceScore)PIP                                      ***
## as.factor(ManagerID)2
## as.factor(ManagerID)3
## as.factor(ManagerID)4
## as.factor(ManagerID)5
## as.factor(ManagerID)6
## as.factor(ManagerID)7
## as.factor(ManagerID)9
## as.factor(ManagerID)10
## as.factor(ManagerID)11
## as.factor(ManagerID)12
## as.factor(ManagerID)13
## as.factor(ManagerID)14
## as.factor(ManagerID)15
## as.factor(ManagerID)16
## as.factor(ManagerID)17
## as.factor(ManagerID)18
## as.factor(ManagerID)19
## as.factor(ManagerID)20
## as.factor(ManagerID)21
## as.factor(ManagerID)22
## as.factor(ManagerID)30
## as.factor(ManagerID)39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8599 on 271 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.1991, Adjusted R-squared:  0.1104
## F-statistic: 2.246 on 30 and 271 DF,  p-value: 0.0003722
```

```
## While I find no statistical significance on the basis of race, I find some related to certain
performance ratings after controlling for other variables (which is discussed below).  I will ad
d the predictions of this linear model to the file.
DentalMagicHR<-DentalMagicHR%>%
  add_predictions(lin_mod_sats)%>%
  rename(predsats_lin=pred)

##Mutating prediction to zero decimals, rounding down
DentalMagicHR<-DentalMagicHR%>%mutate(RoundedSatsPrediction=ifelse(predsats_lin<2,1,ifelse(preds
ats_lin<3,2,ifelse(predsats_lin<4,3,ifelse(predsats_lin<5,4,5)))))
```

```
## While I argue that satisfaction can be viewed through a linear model, I ensure adequate due d
iligence by running the logit model as well.
##install.packages("nnet")
#library(nnet)


log_mod_sats<-glm(EmpSatisfaction~relevel(race.f,ref="White")+
                  as.factor(PerformanceScore)+
                  as.factor(ManagerID),data=DentalMagicHR)

summary(log_mod_sats)
```

```
##
## Call:
## glm(formula = EmpSatisfaction ~ relevel(race.f, ref = "White") +
##     as.factor(PerformanceScore) + as.factor(ManagerID), data = DentalMagicHR)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5239  -0.8184   0.0000   0.8337   2.6333
##
## Coefficients:
##                                                                Estimate
## (Intercept)                                                    3.686300
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.459859
## relevel(race.f, ref = "White")Asian                           -0.175409
## relevel(race.f, ref = "White")Black or African American       -0.008819
## relevel(race.f, ref = "White")Hispanic                        -0.369076
## relevel(race.f, ref = "White")Two or more races                0.078394
## as.factor(PerformanceScore)Fully Meets                        -0.154125
## as.factor(PerformanceScore)Needs Improvement                  -0.529251
## as.factor(PerformanceScore)PIP                                -1.746172
## as.factor(ManagerID)2                                          0.146190
## as.factor(ManagerID)3                                          0.643235
## as.factor(ManagerID)4                                          0.296453
## as.factor(ManagerID)5                                          0.805879
## as.factor(ManagerID)6                                         -0.088100
## as.factor(ManagerID)7                                          0.535173
## as.factor(ManagerID)9                                         -0.532175
## as.factor(ManagerID)10                                         0.739398
## as.factor(ManagerID)11                                         0.426545
## as.factor(ManagerID)12                                         0.127553
## as.factor(ManagerID)13                                         0.536911
## as.factor(ManagerID)14                                         0.465431
## as.factor(ManagerID)15                                         0.929140
## as.factor(ManagerID)16                                         0.460012
## as.factor(ManagerID)17                                         0.583636
## as.factor(ManagerID)18                                         0.305232
## as.factor(ManagerID)19                                         0.550230
## as.factor(ManagerID)20                                         0.462198
## as.factor(ManagerID)21                                         0.634155
## as.factor(ManagerID)22                                         0.583808
## as.factor(ManagerID)30                                         1.467825
## as.factor(ManagerID)39                                         0.594072
##                                                               Std. Error
## (Intercept)                                                     0.390842
## relevel(race.f, ref = "White")American Indian or Alaska Native   0.450115
## relevel(race.f, ref = "White")Asian                             0.177256
## relevel(race.f, ref = "White")Black or African American         0.138738
## relevel(race.f, ref = "White")Hispanic                          0.450416
## relevel(race.f, ref = "White")Two or more races                 0.226617
## as.factor(PerformanceScore)Fully Meets                          0.162918
## as.factor(PerformanceScore)Needs Improvement                    0.261388
## as.factor(PerformanceScore)PIP                                  0.297584
## as.factor(ManagerID)2                                           0.406323
```

```
## as.factor(ManagerID)3                                              0.939991
## as.factor(ManagerID)4                                              0.408897
## as.factor(ManagerID)5                                              0.488621
## as.factor(ManagerID)6                                              0.559653
## as.factor(ManagerID)7                                              0.422564
## as.factor(ManagerID)9                                              0.704628
## as.factor(ManagerID)10                                             0.455572
## as.factor(ManagerID)11                                             0.402234
## as.factor(ManagerID)12                                             0.406944
## as.factor(ManagerID)13                                             0.465975
## as.factor(ManagerID)14                                             0.401237
## as.factor(ManagerID)15                                             0.613394
## as.factor(ManagerID)16                                             0.398020
## as.factor(ManagerID)17                                             0.427523
## as.factor(ManagerID)18                                             0.398342
## as.factor(ManagerID)19                                             0.400626
## as.factor(ManagerID)20                                             0.398812
## as.factor(ManagerID)21                                             0.427687
## as.factor(ManagerID)22                                             0.402789
## as.factor(ManagerID)30                                             0.930685
## as.factor(ManagerID)39                                             0.428004
##                                                                  t value Pr(>|t|)
## (Intercept)                                                        9.432  < 2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.022   0.3079
## relevel(race.f, ref = "White")Asian                               -0.990   0.3233
## relevel(race.f, ref = "White")Black or African American           -0.064   0.9494
## relevel(race.f, ref = "White")Hispanic                            -0.819   0.4133
## relevel(race.f, ref = "White")Two or more races                    0.346   0.7297
## as.factor(PerformanceScore)Fully Meets                            -0.946   0.3450
## as.factor(PerformanceScore)Needs Improvement                      -2.025   0.0439
## as.factor(PerformanceScore)PIP                                    -5.868 1.28e-08
## as.factor(ManagerID)2                                              0.360   0.7193
## as.factor(ManagerID)3                                              0.684   0.4944
## as.factor(ManagerID)4                                              0.725   0.4691
## as.factor(ManagerID)5                                              1.649   0.1002
## as.factor(ManagerID)6                                             -0.157   0.8750
## as.factor(ManagerID)7                                              1.266   0.2064
## as.factor(ManagerID)9                                             -0.755   0.4508
## as.factor(ManagerID)10                                             1.623   0.1057
## as.factor(ManagerID)11                                             1.060   0.2899
## as.factor(ManagerID)12                                             0.313   0.7542
## as.factor(ManagerID)13                                             1.152   0.2502
## as.factor(ManagerID)14                                             1.160   0.2471
## as.factor(ManagerID)15                                             1.515   0.1310
## as.factor(ManagerID)16                                             1.156   0.2488
## as.factor(ManagerID)17                                             1.365   0.1733
## as.factor(ManagerID)18                                             0.766   0.4442
## as.factor(ManagerID)19                                             1.373   0.1708
## as.factor(ManagerID)20                                             1.159   0.2475
## as.factor(ManagerID)21                                             1.483   0.1393
## as.factor(ManagerID)22                                             1.449   0.1484
## as.factor(ManagerID)30                                             1.577   0.1159
## as.factor(ManagerID)39                                             1.388   0.1663
##
```

```
## (Intercept)                                                          ***
## relevel(race.f, ref = "White")American Indian or Alaska Native
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## as.factor(PerformanceScore)Fully Meets
## as.factor(PerformanceScore)Needs Improvement                         *
## as.factor(PerformanceScore)PIP                                       ***
## as.factor(ManagerID)2
## as.factor(ManagerID)3
## as.factor(ManagerID)4
## as.factor(ManagerID)5
## as.factor(ManagerID)6
## as.factor(ManagerID)7
## as.factor(ManagerID)9
## as.factor(ManagerID)10
## as.factor(ManagerID)11
## as.factor(ManagerID)12
## as.factor(ManagerID)13
## as.factor(ManagerID)14
## as.factor(ManagerID)15
## as.factor(ManagerID)16
## as.factor(ManagerID)17
## as.factor(ManagerID)18
## as.factor(ManagerID)19
## as.factor(ManagerID)20
## as.factor(ManagerID)21
## as.factor(ManagerID)22
## as.factor(ManagerID)30
## as.factor(ManagerID)39
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.7393489)
##
##     Null deviance: 250.17  on 301  degrees of freedom
## Residual deviance: 200.36  on 271  degrees of freedom
##   (8 observations deleted due to missingness)
## AIC: 797.13
##
## Number of Fisher Scoring iterations: 2
```

```
## The summary shows similar results (both models are limited in terms of race as a predictor of satisfaction but do find significance with respect to PIP and Needs Improvement ratings).
```

In Exploration 2, I see no sigificant relationship between race and employee satisfaction in both models. I opt to use the linear model in my interpretation and predictions.

I controlled for common influencers of satisfaction, including the Manager as well as performance rating. Perhaps not surprisingly, those with low performance ratings (Need Improvement, and PIP) had statistically significant negative relationship to satisfaction, after holding race and ManagerID constant. Better stated, holding race and manager ID constant, an employee with a Needs Improvement rating has, on average, a half point lower engagement score (on a 5 point scale) as compared to those with an Exceeds rating and an employee with a PIP rating has, on average, a 1.74 lower engagement score as compared to those with an Exceeds rating.

The significance in correlation for PIP and Needs Improvement rating has its argument boosted by a high Multiple R-Square of 0.1991, suggesting nearly 20% of the variance in engagement can be accounted for by these low ratings. Intuitively, this makes sense.

I entered the predicted satisfaction level (based on this model) to my data file, calling it predsats_lin.

Finally, let's start exploration 3: How well does race predict performance ratings?

```
## Performance ratings could carry a similar logic to satisfaction in calling it a continuous va
riable (with each unit higher in rating signifying better performance).  While imperfect, I will
consider it.  However, I will test it both through logit and linear given this variable's  chara
cteristics.

## Linear
lin_mod_rat<-lm(PerfScoreID~relevel(race.f, ref="White"),data=DentalMagicHR)
summary(lin_mod_rat)
```

```
##
## Call:
## lm(formula = PerfScoreID ~ relevel(race.f, ref = "White"), data = DentalMagicHR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00000  0.00000  0.00000  0.08772  1.08772
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                 3.000e+00
## relevel(race.f, ref = "White")American Indian or Alaska Native  5.000e-01
## relevel(race.f, ref = "White")Asian                         2.676e-16
## relevel(race.f, ref = "White")Black or African American    -8.772e-02
## relevel(race.f, ref = "White")Hispanic                     -2.500e-01
## relevel(race.f, ref = "White")Two or more races            -5.556e-02
##                                                            Std. Error
## (Intercept)                                                 4.154e-02
## relevel(race.f, ref = "White")American Indian or Alaska Native  2.915e-01
## relevel(race.f, ref = "White")Asian                         1.073e-01
## relevel(race.f, ref = "White")Black or African American     8.700e-02
## relevel(race.f, ref = "White")Hispanic                      2.915e-01
## relevel(race.f, ref = "White")Two or more races             1.422e-01
##                                                            t value Pr(>|t|)
## (Intercept)                                                 72.215   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.715   0.0874
## relevel(race.f, ref = "White")Asian                          0.000   1.0000
## relevel(race.f, ref = "White")Black or African American     -1.008   0.3141
## relevel(race.f, ref = "White")Hispanic                      -0.858   0.3918
## relevel(race.f, ref = "White")Two or more races             -0.391   0.6964
##
## (Intercept)                                                ***
## relevel(race.f, ref = "White")American Indian or Alaska Native .
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5771 on 304 degrees of freedom
## Multiple R-squared:  0.01616,    Adjusted R-squared:  -1.832e-05
## F-statistic: 0.9989 on 5 and 304 DF,  p-value: 0.4186
```

```
##Logit
log_mod_rat<-glm(PerfScoreID~relevel(race.f, ref="White"),data=DentalMagicHR)
summary(log_mod_rat)
```

```
##
## Call:
## glm(formula = PerfScoreID ~ relevel(race.f, ref = "White"), data = DentalMagicHR)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.00000   0.00000   0.00000   0.08772   1.08772
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                 3.000e+00
## relevel(race.f, ref = "White")American Indian or Alaska Native  5.000e-01
## relevel(race.f, ref = "White")Asian                         2.676e-16
## relevel(race.f, ref = "White")Black or African American    -8.772e-02
## relevel(race.f, ref = "White")Hispanic                     -2.500e-01
## relevel(race.f, ref = "White")Two or more races            -5.556e-02
##                                                            Std. Error
## (Intercept)                                                 4.154e-02
## relevel(race.f, ref = "White")American Indian or Alaska Native  2.915e-01
## relevel(race.f, ref = "White")Asian                         1.073e-01
## relevel(race.f, ref = "White")Black or African American     8.700e-02
## relevel(race.f, ref = "White")Hispanic                      2.915e-01
## relevel(race.f, ref = "White")Two or more races             1.422e-01
##                                                            t value Pr(>|t|)
## (Intercept)                                                 72.215   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.715   0.0874
## relevel(race.f, ref = "White")Asian                          0.000   1.0000
## relevel(race.f, ref = "White")Black or African American     -1.008   0.3141
## relevel(race.f, ref = "White")Hispanic                      -0.858   0.3918
## relevel(race.f, ref = "White")Two or more races             -0.391   0.6964
##
## (Intercept)                                                ***
## relevel(race.f, ref = "White")American Indian or Alaska Native .
## relevel(race.f, ref = "White")Asian
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3330784)
##
##     Null deviance: 102.92  on 309  degrees of freedom
## Residual deviance: 101.26  on 304  degrees of freedom
## AIC: 546.88
##
## Number of Fisher Scoring iterations: 2
```

```
## I find no significance in either model.
```

```
## I'm going to now test this model with additional variables that might improve model accuracy.
Specifically, ManagerID (to better understand whether a specific manager bias influences rating
s), Gender, and PayRate (to explore whether more seniority in pay is more strongly associated wi
th higher ratings)

## Linear
lin_mod_rat<-lm(PerfScoreID~relevel(race.f, ref="White")+as.factor(GenderID)+as.factor(ManagerI
D)+PayRate,data=DentalMagicHR)
summary(lin_mod_rat)
```

```
##
## Call:
## lm(formula = PerfScoreID ~ relevel(race.f, ref = "White") + as.factor(GenderID) +
##     as.factor(ManagerID) + PayRate, data = DentalMagicHR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.00651 -0.08595  0.02083  0.14177  1.18229
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                 2.9350293
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.6029079
## relevel(race.f, ref = "White")Asian                        -0.0400968
## relevel(race.f, ref = "White")Black or African American    -0.1120384
## relevel(race.f, ref = "White")Hispanic                     -0.3926557
## relevel(race.f, ref = "White")Two or more races             0.0098344
## as.factor(GenderID)1                                       -0.0657731
## as.factor(ManagerID)2                                      -0.0411639
## as.factor(ManagerID)3                                      -0.0923636
## as.factor(ManagerID)4                                       0.0009575
## as.factor(ManagerID)5                                       0.0493390
## as.factor(ManagerID)6                                       0.1961392
## as.factor(ManagerID)7                                      -0.2885346
## as.factor(ManagerID)9                                      -0.2597030
## as.factor(ManagerID)10                                     -0.0442421
## as.factor(ManagerID)11                                     -0.1138100
## as.factor(ManagerID)12                                     -0.1333420
## as.factor(ManagerID)13                                     -0.1119040
## as.factor(ManagerID)14                                      0.0023057
## as.factor(ManagerID)15                                     -0.5296704
## as.factor(ManagerID)16                                     -0.0211193
## as.factor(ManagerID)17                                     -0.5056582
## as.factor(ManagerID)18                                      0.1191602
## as.factor(ManagerID)19                                      0.0046227
## as.factor(ManagerID)20                                     -0.0721733
## as.factor(ManagerID)21                                     -0.1790834
## as.factor(ManagerID)22                                     -0.1475745
## as.factor(ManagerID)30                                     -0.0932001
## as.factor(ManagerID)39                                     -0.0685605
## PayRate                                                     0.0056490
##                                                            Std. Error
## (Intercept)                                                 0.2803978
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.3061944
## relevel(race.f, ref = "White")Asian                         0.1216149
## relevel(race.f, ref = "White")Black or African American     0.0937832
## relevel(race.f, ref = "White")Hispanic                      0.3086909
## relevel(race.f, ref = "White")Two or more races             0.1571448
## as.factor(GenderID)1                                        0.0725576
## as.factor(ManagerID)2                                       0.3261446
## as.factor(ManagerID)3                                       0.6503557
## as.factor(ManagerID)4                                       0.3075856
## as.factor(ManagerID)5                                       0.3785836
```

```
## as.factor(ManagerID)6                                        0.3857690
## as.factor(ManagerID)7                                        0.3187043
## as.factor(ManagerID)9                                        0.5263967
## as.factor(ManagerID)10                                       0.3523575
## as.factor(ManagerID)11                                       0.2754352
## as.factor(ManagerID)12                                       0.2747244
## as.factor(ManagerID)13                                       0.3543537
## as.factor(ManagerID)14                                       0.2762080
## as.factor(ManagerID)15                                       0.4632138
## as.factor(ManagerID)16                                       0.2735536
## as.factor(ManagerID)17                                       0.3523765
## as.factor(ManagerID)18                                       0.2740239
## as.factor(ManagerID)19                                       0.2752936
## as.factor(ManagerID)20                                       0.2740113
## as.factor(ManagerID)21                                       0.3516943
## as.factor(ManagerID)22                                       0.2757554
## as.factor(ManagerID)30                                       0.6411026
## as.factor(ManagerID)39                                       0.2934445
## PayRate                                                      0.0058973
##                                                             t value Pr(>|t|)
## (Intercept)                                                  10.467   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.969    0.050
## relevel(race.f, ref = "White")Asian                          -0.330    0.742
## relevel(race.f, ref = "White")Black or African American      -1.195    0.233
## relevel(race.f, ref = "White")Hispanic                       -1.272    0.204
## relevel(race.f, ref = "White")Two or more races               0.063    0.950
## as.factor(GenderID)1                                         -0.906    0.365
## as.factor(ManagerID)2                                        -0.126    0.900
## as.factor(ManagerID)3                                        -0.142    0.887
## as.factor(ManagerID)4                                         0.003    0.998
## as.factor(ManagerID)5                                         0.130    0.896
## as.factor(ManagerID)6                                         0.508    0.612
## as.factor(ManagerID)7                                        -0.905    0.366
## as.factor(ManagerID)9                                        -0.493    0.622
## as.factor(ManagerID)10                                       -0.126    0.900
## as.factor(ManagerID)11                                       -0.413    0.680
## as.factor(ManagerID)12                                       -0.485    0.628
## as.factor(ManagerID)13                                       -0.316    0.752
## as.factor(ManagerID)14                                        0.008    0.993
## as.factor(ManagerID)15                                       -1.143    0.254
## as.factor(ManagerID)16                                       -0.077    0.939
## as.factor(ManagerID)17                                       -1.435    0.152
## as.factor(ManagerID)18                                        0.435    0.664
## as.factor(ManagerID)19                                        0.017    0.987
## as.factor(ManagerID)20                                       -0.263    0.792
## as.factor(ManagerID)21                                       -0.509    0.611
## as.factor(ManagerID)22                                       -0.535    0.593
## as.factor(ManagerID)30                                       -0.145    0.885
## as.factor(ManagerID)39                                       -0.234    0.815
## PayRate                                                       0.958    0.339
##
## (Intercept)                                                 ***
## relevel(race.f, ref = "White")American Indian or Alaska Native *
## relevel(race.f, ref = "White")Asian
```

```
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## as.factor(GenderID)1
## as.factor(ManagerID)2
## as.factor(ManagerID)3
## as.factor(ManagerID)4
## as.factor(ManagerID)5
## as.factor(ManagerID)6
## as.factor(ManagerID)7
## as.factor(ManagerID)9
## as.factor(ManagerID)10
## as.factor(ManagerID)11
## as.factor(ManagerID)12
## as.factor(ManagerID)13
## as.factor(ManagerID)14
## as.factor(ManagerID)15
## as.factor(ManagerID)16
## as.factor(ManagerID)17
## as.factor(ManagerID)18
## as.factor(ManagerID)19
## as.factor(ManagerID)20
## as.factor(ManagerID)21
## as.factor(ManagerID)22
## as.factor(ManagerID)30
## as.factor(ManagerID)39
## PayRate
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5912 on 272 degrees of freedom
##   (8 observations deleted due to missingness)
## Multiple R-squared:  0.06678,    Adjusted R-squared:  -0.03272
## F-statistic: 0.6712 on 29 and 272 DF,  p-value: 0.9015
```

```
## I find some significance in this model after exploring likely factors I sought to control fo
r.  Specifically, after controlling for gender, manager, and payrate, there is a statistically s
ignificant relationship between rating and status as  American Indian / Alaska Native (AI/AN).
 The R-Square value suggests that identity as AI/AN influences up to 7% of variability in perfor
mance rating.

##I will add the prediction to the model
DentalMagicHR<-DentalMagicHR%>%
  add_predictions(lin_mod_rat)%>%
  rename(predrat_lin=pred)

##Mutating prediction to zero decimals, rounding down
DentalMagicHR<-DentalMagicHR%>%mutate(RoundedRatingPrediction=ifelse(predrat_lin<2,1,ifelse(pred
rat_lin<3,2,ifelse(predrat_lin<4,3,4))))

## I also checked this using the logit model, but found (as anticipated) similar significance as
well.
log_mod_rat<-glm(PerfScoreID~relevel(race.f, ref="White")+as.factor(GenderID)+as.factor(ManagerI
D)+PayRate,data=DentalMagicHR)
summary(log_mod_rat)
```

```
##
## Call:
## glm(formula = PerfScoreID ~ relevel(race.f, ref = "White") +
##     as.factor(GenderID) + as.factor(ManagerID) + PayRate, data = DentalMagicHR)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.00651  -0.08595   0.02083   0.14177   1.18229
##
## Coefficients:
##                                                              Estimate
## (Intercept)                                                 2.9350293
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.6029079
## relevel(race.f, ref = "White")Asian                        -0.0400968
## relevel(race.f, ref = "White")Black or African American    -0.1120384
## relevel(race.f, ref = "White")Hispanic                     -0.3926557
## relevel(race.f, ref = "White")Two or more races             0.0098344
## as.factor(GenderID)1                                       -0.0657731
## as.factor(ManagerID)2                                      -0.0411639
## as.factor(ManagerID)3                                      -0.0923636
## as.factor(ManagerID)4                                       0.0009575
## as.factor(ManagerID)5                                       0.0493390
## as.factor(ManagerID)6                                       0.1961392
## as.factor(ManagerID)7                                      -0.2885346
## as.factor(ManagerID)9                                      -0.2597030
## as.factor(ManagerID)10                                     -0.0442421
## as.factor(ManagerID)11                                     -0.1138100
## as.factor(ManagerID)12                                     -0.1333420
## as.factor(ManagerID)13                                     -0.1119040
## as.factor(ManagerID)14                                      0.0023057
## as.factor(ManagerID)15                                     -0.5296704
## as.factor(ManagerID)16                                     -0.0211193
## as.factor(ManagerID)17                                     -0.5056582
## as.factor(ManagerID)18                                      0.1191602
## as.factor(ManagerID)19                                      0.0046227
## as.factor(ManagerID)20                                     -0.0721733
## as.factor(ManagerID)21                                     -0.1790834
## as.factor(ManagerID)22                                     -0.1475745
## as.factor(ManagerID)30                                     -0.0932001
## as.factor(ManagerID)39                                     -0.0685605
## PayRate                                                     0.0056490
##                                                            Std. Error
## (Intercept)                                                 0.2803978
## relevel(race.f, ref = "White")American Indian or Alaska Native  0.3061944
## relevel(race.f, ref = "White")Asian                         0.1216149
## relevel(race.f, ref = "White")Black or African American     0.0937832
## relevel(race.f, ref = "White")Hispanic                      0.3086909
## relevel(race.f, ref = "White")Two or more races             0.1571448
## as.factor(GenderID)1                                        0.0725576
## as.factor(ManagerID)2                                       0.3261446
## as.factor(ManagerID)3                                       0.6503557
## as.factor(ManagerID)4                                       0.3075856
## as.factor(ManagerID)5                                       0.3785836
```

```
## as.factor(ManagerID)6                                          0.3857690
## as.factor(ManagerID)7                                          0.3187043
## as.factor(ManagerID)9                                          0.5263967
## as.factor(ManagerID)10                                         0.3523575
## as.factor(ManagerID)11                                         0.2754352
## as.factor(ManagerID)12                                         0.2747244
## as.factor(ManagerID)13                                         0.3543537
## as.factor(ManagerID)14                                         0.2762080
## as.factor(ManagerID)15                                         0.4632138
## as.factor(ManagerID)16                                         0.2735536
## as.factor(ManagerID)17                                         0.3523765
## as.factor(ManagerID)18                                         0.2740239
## as.factor(ManagerID)19                                         0.2752936
## as.factor(ManagerID)20                                         0.2740113
## as.factor(ManagerID)21                                         0.3516943
## as.factor(ManagerID)22                                         0.2757554
## as.factor(ManagerID)30                                         0.6411026
## as.factor(ManagerID)39                                         0.2934445
## PayRate                                                        0.0058973
##                                                               t value Pr(>|t|)
## (Intercept)                                                    10.467   <2e-16
## relevel(race.f, ref = "White")American Indian or Alaska Native   1.969    0.050
## relevel(race.f, ref = "White")Asian                            -0.330    0.742
## relevel(race.f, ref = "White")Black or African American        -1.195    0.233
## relevel(race.f, ref = "White")Hispanic                         -1.272    0.204
## relevel(race.f, ref = "White")Two or more races                 0.063    0.950
## as.factor(GenderID)1                                           -0.906    0.365
## as.factor(ManagerID)2                                          -0.126    0.900
## as.factor(ManagerID)3                                          -0.142    0.887
## as.factor(ManagerID)4                                           0.003    0.998
## as.factor(ManagerID)5                                           0.130    0.896
## as.factor(ManagerID)6                                           0.508    0.612
## as.factor(ManagerID)7                                          -0.905    0.366
## as.factor(ManagerID)9                                          -0.493    0.622
## as.factor(ManagerID)10                                         -0.126    0.900
## as.factor(ManagerID)11                                         -0.413    0.680
## as.factor(ManagerID)12                                         -0.485    0.628
## as.factor(ManagerID)13                                         -0.316    0.752
## as.factor(ManagerID)14                                          0.008    0.993
## as.factor(ManagerID)15                                         -1.143    0.254
## as.factor(ManagerID)16                                         -0.077    0.939
## as.factor(ManagerID)17                                         -1.435    0.152
## as.factor(ManagerID)18                                          0.435    0.664
## as.factor(ManagerID)19                                          0.017    0.987
## as.factor(ManagerID)20                                         -0.263    0.792
## as.factor(ManagerID)21                                         -0.509    0.611
## as.factor(ManagerID)22                                         -0.535    0.593
## as.factor(ManagerID)30                                         -0.145    0.885
## as.factor(ManagerID)39                                         -0.234    0.815
## PayRate                                                         0.958    0.339
##
## (Intercept)                                                   ***
## relevel(race.f, ref = "White")American Indian or Alaska Native *
## relevel(race.f, ref = "White")Asian
```

```
## relevel(race.f, ref = "White")Black or African American
## relevel(race.f, ref = "White")Hispanic
## relevel(race.f, ref = "White")Two or more races
## as.factor(GenderID)1
## as.factor(ManagerID)2
## as.factor(ManagerID)3
## as.factor(ManagerID)4
## as.factor(ManagerID)5
## as.factor(ManagerID)6
## as.factor(ManagerID)7
## as.factor(ManagerID)9
## as.factor(ManagerID)10
## as.factor(ManagerID)11
## as.factor(ManagerID)12
## as.factor(ManagerID)13
## as.factor(ManagerID)14
## as.factor(ManagerID)15
## as.factor(ManagerID)16
## as.factor(ManagerID)17
## as.factor(ManagerID)18
## as.factor(ManagerID)19
## as.factor(ManagerID)20
## as.factor(ManagerID)21
## as.factor(ManagerID)22
## as.factor(ManagerID)30
## as.factor(ManagerID)39
## PayRate
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3495483)
##
##     Null deviance: 101.881  on 301  degrees of freedom
## Residual deviance:  95.077  on 272  degrees of freedom
##   (8 observations deleted due to missingness)
## AIC: 570.01
##
## Number of Fisher Scoring iterations: 2
```

In Exploration 3, I find a significant positive relationship between the AIAN subgroup and performance rating when compared to the White population, and after holding gender, manager, and pay rate constant. Holding gender and manager ID constant, employees who identify as AIAN on average have a .6 higher rating as compared to white employees on a 4 point rating scale. The R-Square value suggests that identity as AI/AN influences up to 7% of variability in performance rating.
This is an interesting finding, given the small representation of the AIAN population. I have decided to keep this insight in this analysis to engage further discussion and analysis in the future, but recognize that a very small sample size has contributed to this significant finding. I have created a predicted rating variable called predrat_lin.

CONCLUDING REMARKS

Let’s summarize the findings from the models.

Exploration 1: How well does race predict compensation levels?

- Holding gender constant, a Black employee at DentalMagicHR makes on average $5.39 more on their hourly pay rate as compared to white employees. And based on an R-square value of .03901, this accounts for approximately 4% of the variability in Pay Rate.
- Predicted compensation level for this model has been added as “predcomp_lm”.

Exploration 2: How well does race predict employee satisfaction levels?

- I see no significant relationship between race and employee satisfaction in both linear and logit models.
- Holding race and manager ID constant, an employee with a Needs Improvement rating has, on average, a half point lower engagement score (on a 5 point scale) as compared to those with an Exceeds rating
- Holding race and manager ID constant, an employee with a PIP rating has, on average, a 1.74 lower engagement score (on a 5 point scale) as compared to those with an Exceeds rating.
- The significance in correlation for PIP and Needs Improvement rating has its argument boosted by a high Multiple R-Square of 0.1991, suggesting nearly 20% of the variance in engagement can be accounted for by these low ratings.
- Predicted satisfaction level (based on this model) to my data file, calling it predsats_lin.

Exploration 3: How well does race predict performance ratings?

- Holding gender and manager ID constant, employees who identify as AIAN on average have a .6 higher rating as compared to white employees on a 4 point rating scale. The R-Square value suggests that identity as AI/AN influences up to 7% of variability in performance rating.
- This is an interesting finding, given the small representation of the AIAN population. I decided to keep this insight in this analysis to engage further discussion and analysis in the future, but recognize that a very small sample size has contributed to this significant finding.
- I have created a predicted rating variable called predrat_lin.

LESSONS LEARNED AND FINAL THOUGHTS

A few thoughts here. First, limitations…

- I used fake data for this exercise, limiting interpretation or ability to conclude as consistent with real life scenarios.
- Small population size in general (~300) makes the depth and breadth of this analysis limited. For example, the significance of the AIAN population with respect to performance is quite limited in terms of utility.
- Single location vs. nationwide or international: One could argue that because this dataset is from an employer in a single location, it may not adequately translate to phenomena in other parts of the country or the world.

Also, the analysis offered me a few “ahas”.

- The process of analyzing might be useful in real life analysis. The analytical discpline and the skill of regression in R is useful.
- Finding nothing is still a finding!

Finally, I made some compromises in this analysis, but they came with cost.

- I debated whether to keep or toss small populations (AIAN & Hispanic). I decided to keep it and made a finding about AIAN, but this also was at a cost because of the small population size.

- Linear v Logit on ordinal variables (performance & satisfaction): I made the decision to treat these as linear, but they are not continuous variables.
- Limited variability & range on success metrics (what is “success” at an organization, exactly?): Here, I tried to build models around success as defined by pay, satisfaction, and performance. But the latter two variables had little range and variability to begin with, limiting the likelihood of any findings.

REFERENCES

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. Journal of Applied Psychology, 88(4), 694–706. https://doi.org/10.1037/0021-9010.88.4.694 (https://doi.org/10.1037/0021-9010.88.4.694)

Slone, K. P. (2016, Feb 12). Pay inequality issue of race, not just sex. The Indianapolis Recorder.

Jones, J. R., Ni, J., & Wilson, D. C. (2009). Comparative effects of race/ethnicity and employee engagement on withdrawal behavior. Journal of Managerial Issues, 21(2), 195+. https://link.gale.com/apps/doc/A202072966/AONE?u=tel_a_vanderbilt&sid=AONE&xid=b5fd67f8 (https://link.gale.com/apps/doc/A202072966/AONE?u=tel_a_vanderbilt&sid=AONE&xid=b5fd67f8)