PENNSTATE

Tel: (814) 863-0281
Fax: (814) 863-7114

Statistical Consulting Center
Department of Statistics
Eberly College of Science

The Pennsylvania State University
323 Thomas Building
University Park, Pa 16802

April 21th, 2018

From: Enying Gao, Yuhe Tian, Kaiyi Wu, Ruiqi Zhang, Undergraduate Student Consultants

To: Professor Kirsten Eilertson

Re: STAT 470 final project

PROJECT TITLE: Burnout Level among Faculty at Penn State Eberly College of Science

1.0 - PROJECT DESCRIPTION

For the final project, our group propose to explore Penn State Eberly College of Science instructors' burnout level, as well as factors attribute to instructors' burnout. According to James (2001), higher education has been described as the fastest growing service in both public and private sector for many years, and along with this trend, teaching in college is recognized as a stressful occupation. In the United States, approximately 20% of faculty members experiences high level of stress because of strong teaching demands (Padilla & Thompson, 2015). As denoted by Tang, Au, Schwarzer and Schmitz (2001), constant stress may lead to burnout, which has a direct effect on both negative physical and mental health. Burnout in educational staff has been subject to a large number of researches, while our group aim to provide insight for the Penn State Eberly College of Science community exclusively. We hope our project could help improve Penn State Eberly College of Science instructors' health condition and therefore develop a better education system.

This study will first examine the overall burnout level for Penn State Eberly College of Science instructors. After obtained the result, our group will use a correlational design to examine the relationship between instructors' burnout level and the explanatory variables of interest, which includes instructors' years of teaching, weekly working hours, job position, teaching discipline, main culture influence, age and gender. Correlational design helps us to examine the relationship between different variables in a nature setting. Since the variables of interests for this study depend on the individual instructor and could only be measured instead of being manipulated, a correlational design will be the best choice.

Participants for this study are volunteered faculties from Penn State Eberly College of Science. According to Penn State Budget Office (2017), our potential participants are 490 faculties. Based on a survey response rate of 10%, we expect a total of 50 participants. In order to promote faculty's involvement, there is a 5% chance for each participant to win a $10 Starbucks gift card.

All Penn State Eberly College of Science faculties were invited to participate in this study. After permission been granted from each department, participating faculty were recruited via email. A detailed briefing of the study was sent to every faculty via their institutional email addresses. The briefing indicated the motivation and purpose of this study. It also acknowledged participants that the survey is only designed for currently employed faculty and will take approximately 5 minutes to complete. Participants' anonymity will be highlighted. No information would be given to the college or department in order to assure confidentiality. The survey was open for ten days.

1.1 - RESEARCH QUESTIONS

Question 1: What is faculty's burnout Penn State Eberly College of Science?
Question 2: What are the potential factors that influence falculty's burnout in Eberly College of Science?

1.2 – VARIABLES

We received 39 responses from the survey. 21 variables were collected in raw data (see Appendix A.2 for Original Data.csv). We dropped the second observation because the subject left many questions blank and dropped the sixteenth observation because the subject's answers are not consistent with each other (age twenty with twenty years of teaching experience).

The burnout level was measured with the emotional exhaustion subscale of the Maslach Burnout Inventory-Educators Survey (MBI-ES) (Maslach, Jackson, & Leiter, 1996). The questionnaire consisted of nine items (see Appendix A). Participants were asked to rate each of the nine questions on a seven-point Likert-type scale anchored from 0 = "Never", 1= "A few times a year or less", 2 = "Once a month or less", 3= "A few times a month", 4 = "Once a week", 5 = "A few times a week" to 6 = "Every day". The total score of the nine items was used to represent each participant's burnout level.

Teaching years was measured by asking participants to report the number of years they have been teaching, while PSU teaching years records number of years they have been teaching in Penn State. Working hour per week was measured by asking participants to report the number of hours they work per week. There is one missing value in teaching years and one in working hours. Additional teaching related information such as job position and academic department of employment was collected. Missing department information was replaced by unknown.

Culture influence was measured by asking participant's ethnicity. Ethnicity is defined as the term for the culture of people in a given geographic region, including their language, heritage, religion and customs. Our group believe that ethnicity is the best way to measure instructors' main culture of influence. Participants also reported other demographics information such as age and gender in this section. In order to replace the missing value in age, we took the mean difference between age and teaching year, and add this mean difference and teaching to get an approximation for age.

The transformed dataset is recorded in an excel file (Model Data.csv) in Appendix A.2, and the R code is offered in Appendix B.1.

Table of Variables

No.	Variable Name	Type	Description
1	subject	Numeric	A subject identifier, assigned by the researchers
2	gender	Categorical	Female, Male, and Other
3	age	Numeric	Recorded in years
4	position	Categorical	Instructors' job positions
5	status	Categorical	Instructors' working status, full-time or part-time
6	department	Categorical	Instructors' department of employment
7	teachyr	Numeric	Self-reported number of teaching years in college
8	PSUteachyr	Numeric	Self-reported number of teaching years at PSU
9	tenure	Categorical	Instructors' tenure status, tenured, tenure track, non-tenure track
10	workhr	Numeric	Self-reported average working hour per week
11	responsibility	Categorical	Self-reported main working responsibility on 1 – 5 scale. 1 means focusi more on teaching, 5 means focusing more on research.
12	ethnicity	Categorical	Main culture influence of each subject
13	burnout	Numeric	Total score by adding up the scores each subject gets in the burnout survey
14	BurnoutDegree	Categorical	Degree of faculty's burnout. Burnout degree is low if burnout $\leq$ 16, moderate if 17 $\leq$burnout $\leq$ 27, high if burnout > 28

Table 1: *This table includes the name, type and description of all variables maintained in the dataset and all variables we used in our analysis.*

2.0 - EXPLORATORY DATA ANALYSIS (EDA)

The purpose of EDA is to gain an introductory understanding of the survey results and to identify variables that influence Penn State Eberly College of Science instructors' burnout level.

We received 39 responses from Eberly College of Science. As mentioned before in Section 1.2 - Variables, there're 10 responses in the dataset that contain missing values and we dropped 2. Such circumstances are understandable as we expected some human errors and granted respondents freedom to leave items blank if they feel given information is sensitive.

After applying methods we discussed before to missing values in our initial dataset (See Appendix A.2 for link to 'Original Data.csv'), we obtained 37 responses (See Appendix A.2 for link to 'Model Data.csv'), in which 16 are females and 21 are males. Since our obtained sample size is small, the majority of our respondents are White/Caucasian with full-time status. With this being said, there're only 2 responses from Asian/Pacific Islander and 1 response from part-time employee.

Figure 1 illustrates the distribution of responses with respect to variable burnout degree. Among all the responses we received, approximately 20% of instructors scored to experience high burnout level every day. However, majority of instructors (approximately 60%) belongs to category 'Low', which suggests that most employees are satisfied with their daily jobs. Further analysis on factors relating to instructors' burnout level will be discussed in detail in Section 3.0 – Statistical Analysis. More visualization of responses in term of other variables can be found in Appendix A.3.

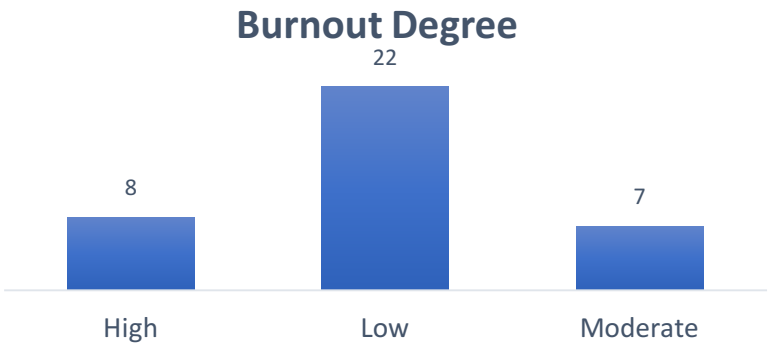


Figure 1: *The graph represents number of responses in terms of Burnout Degree.*

The following table (Table 2) displays the descriptive statistics for quantitative variables of interest. The average burnout for faculty in Eberly College of Science is 17.38, which falls on the edge of low and moderate burnout category. We can also see that the statistics of variable teachyr and PSUteachyr are really close, indicating possible existence of collinearity.

Descriptive Statistics for quantitative variables

Variable	N	Mean	Standard Deviation	Minimum	Maximum
age	37	44.37	10.72	30	68
teachyr	37	13.41	11.07	1	37
PSUteachyr	36	12.19	10.31	1	37

workhr	36	51.69	8.67	40	70
burnout	37	17.38	10.82	0	44

Table 2: *The table displays the count, mean, standard deviation, and range of the quantitative variables*

As a result, we examined correlation between those two variables. It turned out that variable PSUteachyr has a strong positive relationship with variable teachyr (See Figure 2). Thus, we decided to include only one of those two variables into our model to avoid collinearity problem. The variable we chose in the end is PSUteachyr because the purpose of this study is to find factors that are related to instructors' burnout level in Eberly College of Science at Penn State University Park.

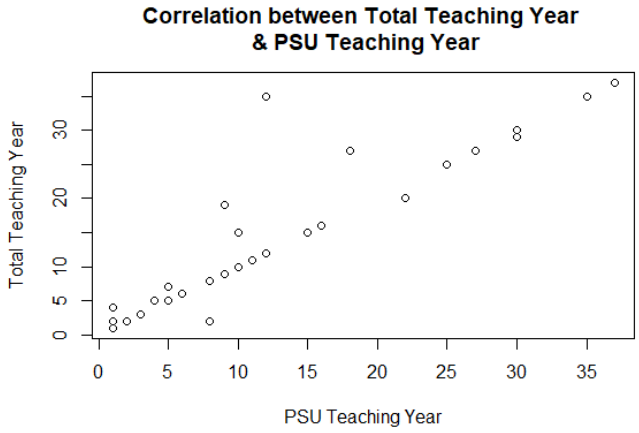


Figure 2: *The graph illustrates the correlation between Total Teaching Year and PSU Teaching Year.*

Next, we examined relationship among working hour, teaching year and burnout. Figure 3 shows that as working hour increases, professors are more likely to experience stress. Figure 4 illustrates that longer teaching year at Penn State contributes to lower burnout level.

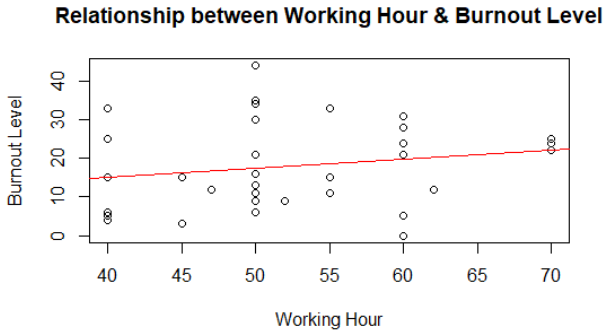


Figure 3: *The graph illustrates the relationship between Working Hour & Burnout Level.*

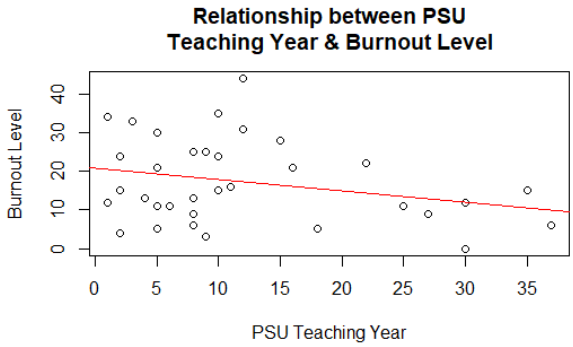


Figure 4: *The graph illustrates the relationship between Teaching Year & Burnout Level.*

Then we plotted variable burnout against all other variables and obtained following informative figures. Figure 5 demonstrates that burnout level is similar among male and female professors. Figure 6 implies that professors who are in tenure track experience higher stress than those who're not in tenure track and those who're already tenured. Figure 7 suggests that full professors score the lowest on burnout level comparing to professors in other teaching positions. Figure 8 specifies that professors in department Biochemistry, Biology and Chemistry tend to be more stressful than those in other departments. Figure 9 expresses that professors who involve in both teaching and researching appear to experience high level of stress in Pennsylvania State University Park.

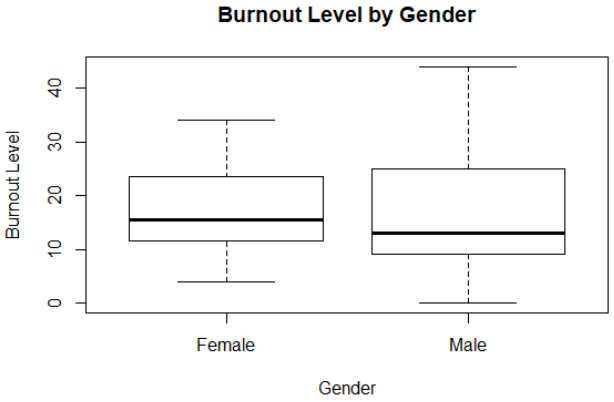


Figure 5: *The boxplot compares professors' Burnout Level by Gender.*

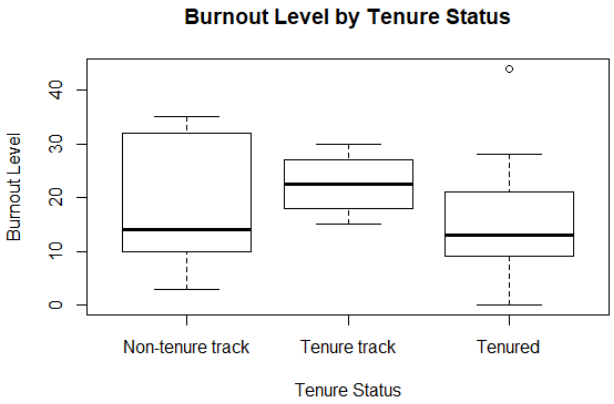


Figure 6: *The boxplot compares professors' Burnout Level by Tenure Status.*

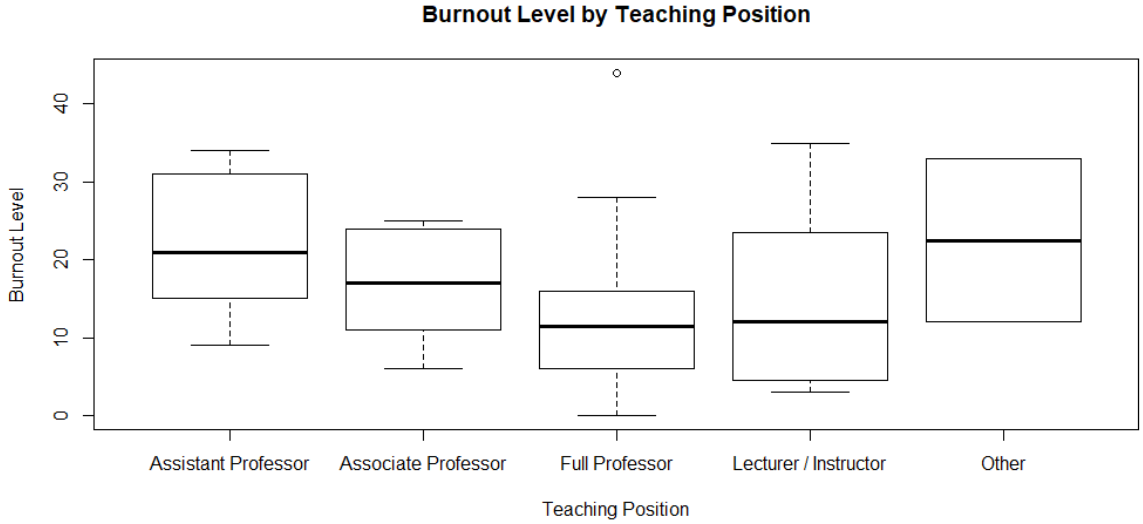


Figure 7: *The boxplot compares professors' Burnout Level by Teaching Positions.*

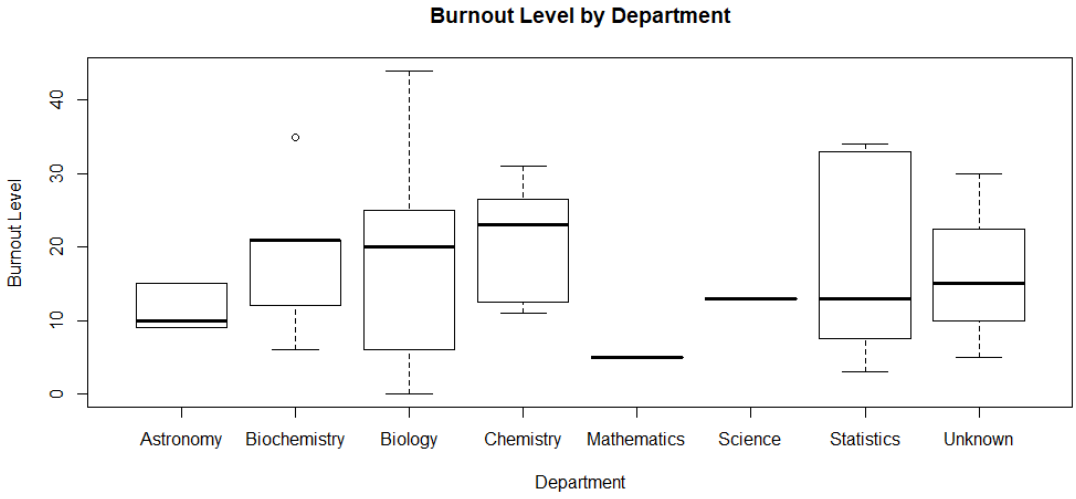


Figure 8: *The boxplot compares professors' Burnout Level by Department in Eberly College of Science.*

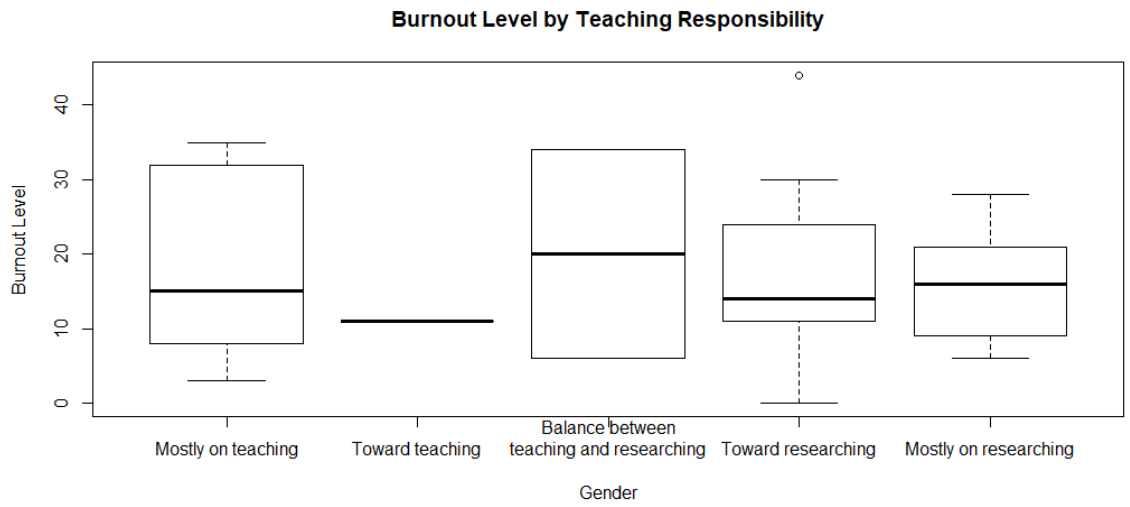


Figure 9: *The boxplot compares professors' Burnout Level by Teaching Responsibility.*

3.0 –STATISTICAL ANALYSIS

In order to address our second research question – what are the potential factors that influence faculty's burnout level in Eberly College of Science, a Multiple Linear Regression Model was used. The model determined how our variables of interest contributes to the burnout level of Eberly College of Science's faculty (See Appendix C for full model).

By conducting exploratory data analysis, we found that faculty's burnout level showed evident variation among tenure status, job position, working responsibility, and department of employment. Moreover, faculty's years of teaching at Penn State seemed to be negatively related to their burnout level, and working hour per week seemed to be positively related to burnout level.

Previous research conducted by Padilla and Thompson (2015) indicated that time allocation variables and perceived pressure contribute to faculty burnout. Time allocation variables include total teaching year at Penn State and average working hour per week. That is to say, perceived pressure is mainly influenced by job demands. Since job demand might be different among job position, tenure status and across department, we decide that using the variable responsibility, which directly measure working, is a better operationalization for examining faculty's burnout.

Based on information provided by EDA and previous researches, we fit our final model based on total teaching year at Penn State, working hour per week, and work responsibility. The assumptions required to build a Multiple Linear Regression Model were checked (See appendix C for details), and the results are summarized below in table3.

Results of Factors Attributes to Faculty's Burnout Level in Eberly College of Science

Factor	**Coefficient**	**P-value**	**Significance**
PSUteachyr	-0.372	0.056	Yes
workhr	0.597	0.016	Yes
responsibility2	-13.583	0.229	No
responsibility3	1.196	0.883	No
responsibility4	-4.420	0.347	No
responsibility5	-4.116	0.493	No

Table 3: *The table above shows the results from the regression model for faculty's burnout level. (For working responsibility, we set responsibility1- mostly teaching as default, responsibility2 indicate the faculty's work responsibility is toward teaching, responsibility3 as both teaching and research, responsibility4 as toward research, and responsibility5 as mostly research)*

In our test, we choose p-value of 0.1 as the cutoff line. If we had a p-value smaller or equal to 0.1, we can determine the corresponding variable has significant contribution to faculty's burnout level. Moreover, the sign of coefficients for a significant explanatory variable can also help in drawing conclusion, where a positive coefficient implies an increase in faculty's burnout level, where a negative coefficient implies a decrease in faculty's burnout level.

According to our model build for faculty's burnout level, the number of years teaching at Penn State had a coefficient of -0.372 and a p-value of 0.056. The above information indicated that there is a negative relationship between faculty's burnout level in Eberly College of Science and the total years faculty works at Penn State. Thus, the longer the faculty teaches at Penn State, the lower the burnout level the faculty has.

Our model also showed that working hour per week had a coefficient of 0.597 and a p-value of 0.016. The above information indicated there is a positive relationship between faculty's burnout level in Eberly College of Science and faculty's average working hour per week. Thus, the longer the faculty works per week, the more burnout the faculty experience.

4.0 – RECOMMENDATIONS

Question 1: What is the burnout level among faculty at Eberly College of Science?
As we concluded in EDA, the average burnout for faculty in Eberly College of Science is 17.38, which falls on the edge of low and moderate burnout category. The EDA also showed that more than half of faculties have a low burnout level, indicating a health mental status and a relative flexible working situation. Around 20% of faculties may have high burnout level and face the challenge of overload pressure, while others show a moderate burnout level.

Question 2: What variables should we use to predict faculty's burnout level at Eberly College of Science?
From our analysis, the teaching year at penn state and the average working hour per week have significant influence on faculty's burnout level. The burnout level will increase with a increasing working hour and the burnout will decrease as the teaching year growing. We want to suggest faculties to find a balance between work and life, especially at the first few years entering the position.

5.0 - RESOURCES

In order to obtain a clear and accurate conclusion for this designed experiment, Statistical Software "RStudio" and Microsoft Excel were used to process and analyze the data.

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.
This program was used mainly in Exploratory Data Analysis (EDA) and Statistical Analysis sections in this report. Our group used RStudio to perform the linear models and to generate various box-plots and tables.

Microsoft Excel (2016)
This computer software was used to organize Dr. Carlson's collected data. The "Study 1, Trial-Level Data" spreadsheet provided 2784 observations and 15 variables of interest.

The following additional resources were also used to complete the study:

James, A. (2011). Managing Stress in Education: A Comprehensive Guide for Staff and Students.
Maslach, C., Jackson, S. E., & Leiter, M. P. (1996). Maslach Burnout Inventory: Manual (3rd Edn). Palo Alto, CA: Consulting Psychologists Press.
Padilla, M. A., & Thompson, J. N. (2015). Burning Out Faculty at Doctoral Research Universities. Stress and Health, 32(5), 551-558.
Tang, C, S., Au, W., Schwarzer, R., & Schmitz, G. (2001). Mental health outcomes of job stress among Chinese teachers: role of stress resource factors and burnout. Journal of Organizational Behavior, 22, 887-901.

6.0 – CONSIDERATIONS

In order to fully understand the scale and effectiveness of this study, there are several limitations we need to consider.

First, the quality of sample data and its representativeness has to be noted. In this survey, we collected 37 effective observations in total, given a target population with size of 490. The response rate is lower than 10% and the sample size is relative small. We also failed to delivery and collect response from Department of Physics and Mathematics due to communication problems, losing one third potential respondents (87 from Physics and 67 from Mathematics). Thus, the sample we got may not be representative enough for the population of faculties in Eberly College of Science. The name and questions of our questionnaire involve several negative description words, which may lead to bias of the respondents. Some potential respondents who are suffering from burnout problem or feel bad about the questionnaire design would reject to fulfill the survey. We also get some feedback indicating a lost in respondent due to information privacy consideration. All those concerns may hurt the sample's representativeness.

Further, it would be helpful to denote that people's responses may varies based on their own explanation of the questions, which may differ from what we expected. For instance, some respondent would consider their teaching assistant or lecturing experience in post-graduate periods as effective teaching year while some would not. Some respondents would calculate their working hours per week accurately and some would give approximations. The data collected from the survey is not necessary to be a fully reflection of the real situation.

Despite the findings of this research, it should be noted that the burnout level may have been impacted by factors not considered in the design of this survey. There exist multiple stress source in people's life, from family burden to peer pressure. Those factors out of college would definitely have influence on facilities's burnout level, which have not been covered in this study. We may want to expand our study to other factors in further research.

We would like to express our great appreciation to Dr. Kirsten Eilertson for her patient instruction and useful critiques in this research work. We would also like to thank Penn State Statistical Consulting Center for their supporting in the statistical assistance.

Appendix A

Section A.1: Original Dataset and Scales

The following table (Table A1) describes the original data collected from the survey.

Table A1: Original Variables

No.	Variable Name	Type	Description
1	subject	Numeric	A subject identifier, assigned by the researchers
2	gender	Categorical	Female, Male, and Other
3	age	Numeric	Recorded in years
4	position	Categorical	Instructors' job positions
5	status	Categorical	Instructors' working status, full-time or part-time
6	department	Categorical	Instructors' department of employment
7	teachyr	Numeric	Self-reported number of teaching years in college
8	PSUteachyr	Numeric	Self-reported number of teaching years at PSU
9	tenure	Categorical	Instructors' tenure status, tenured, tenure track, non-tenure track
10	workhr	Numeric	Self-reported average working hour per week
11	responsibility	Categorical	Self-reported main working responsibility on 1 – 5 scale. 1 means focusing more on teaching, 5 means focusing more on research.
12	ethnicity	Categorical	Main culture influence of each subject
13	Question 1	Ordinal	Answers to nine survey questions regarding burnout level. 0 – Never 1 – A few times a year of less 2 – Once a month or less 3 – A few times a month 4 – Once a week 5 – A few times a week 6 – Everyday
14	Question 2		
15	Question 3		
16	Question 4		
17	Question 5		
18	Question 6		
19	Question 7		
20	Question 8		
21	Question 9		

Table 1: *This table includes the name, type and description of all variables maintained in the dataset and all variables we used in our analysis.*

Section A.2: Original and Recoded dataset

The dataset embedded below is the dataset used in our analysis and is created from the original dataset. R-code used to generate new dataset is included in Appendix B.1.

The datasets can be found from following link:
https://drive.google.com/open?id=1ve-K5AMmlDaH8XlDn0bUjICbiWRwvhRY

Section A.3: Addition information for EDA

The following figures (Figure 1-7) generated from Excel illustrate the distribution of responses with respect to variable gender, employee status, ethnicity, department, position, main working responsibility, and tenure status.

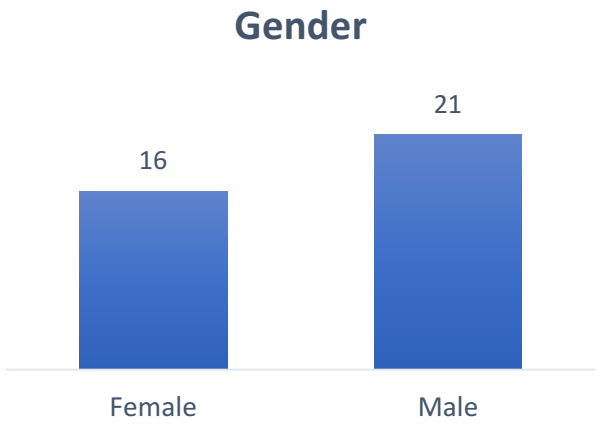


Figure 1: *The graph represents number of responses in terms of Gender.*

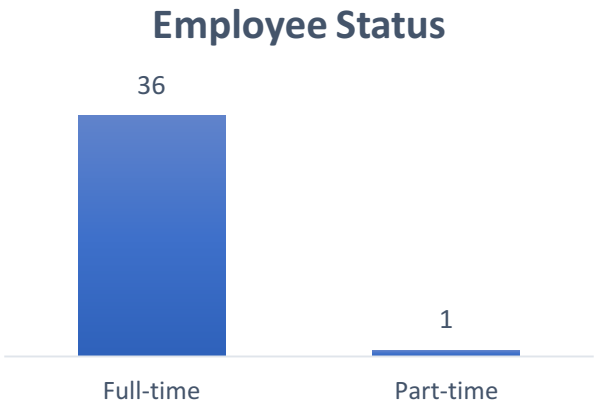


Figure 2: *The graph represents number of responses in terms of Employee Status*

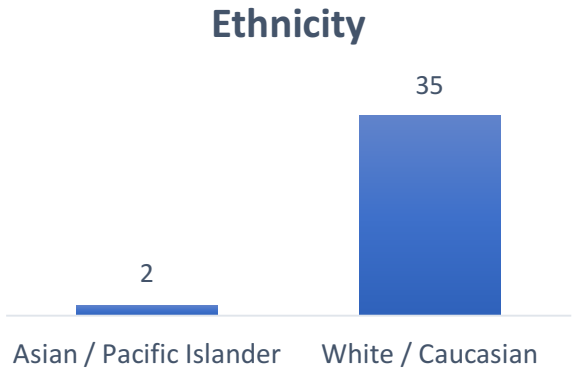


Figure 3: *The graph represents number of responses in terms of Ethnicity.*

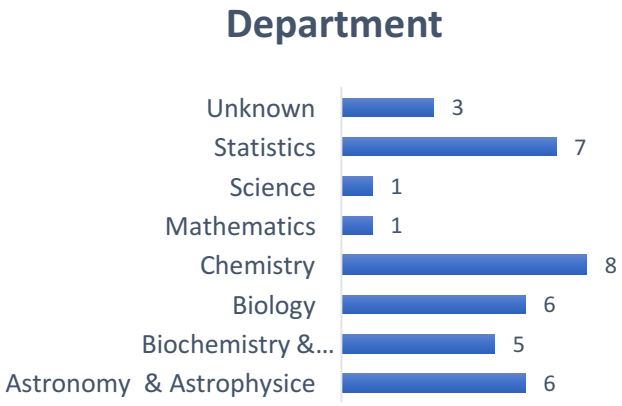


Figure 4: *The graph represents number of responses in terms of Department.*

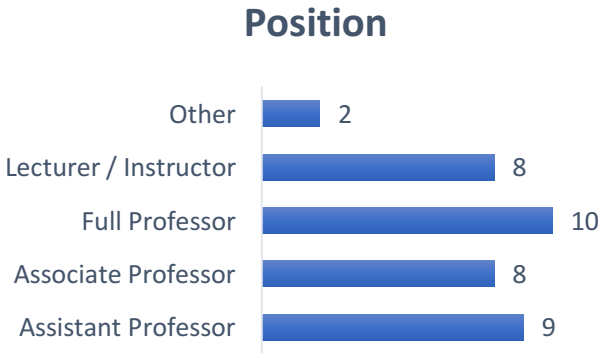


Figure 5: *The graph represents number of responses in terms of Position.*

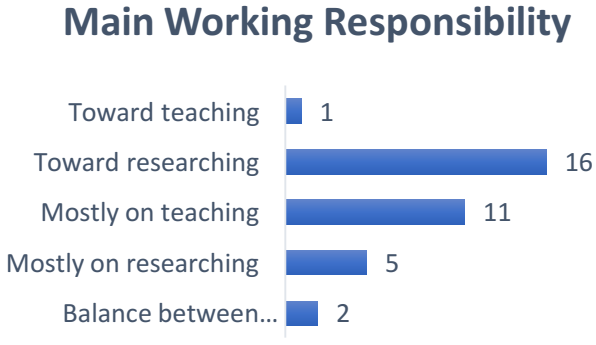


Figure 6: *The graph represents number of responses in terms of Main Working Responsibility.*

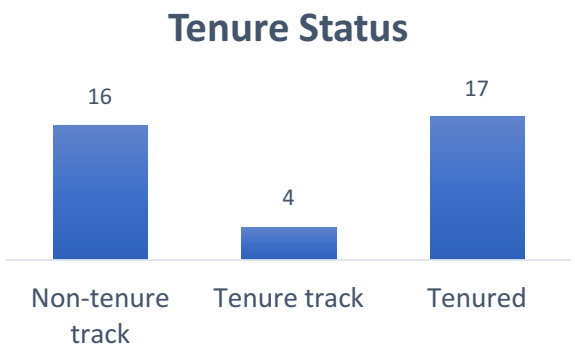


Figure 7: *The graph represents number of responses in terms of Tenure Status.*

Figure 8 generated from Rstudio compares professors' burnout level with respect to ethnicity.

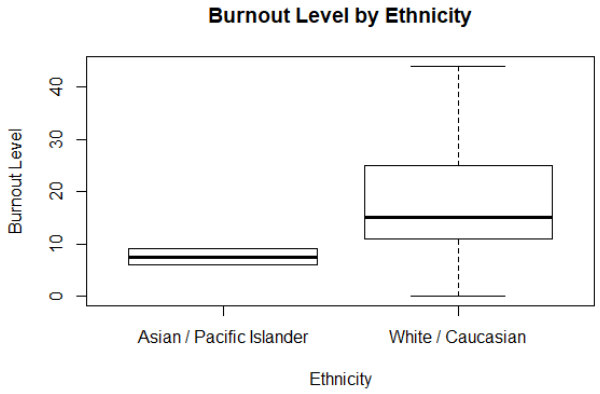


Figure 8: *The boxplot compares professors' Burnout Level by Ethnicity*

The following graphs generated from Rstudio represent the employee structure of Pennsylvania State University Park by looking at the relationship among professors' teaching position, teaching responsibility and tenure status. From Figure 9 we can notice that the major working responsibility for Lecture/Instructor and Assistant Professor is teaching, whereas Associate Professor and Full Professor focus more on researching. From Figure 10, we can notice that all Associate Professor and Full Professor have already gained 'Tenured' status, whereas other professors are either in tenure track or non-tenure track.

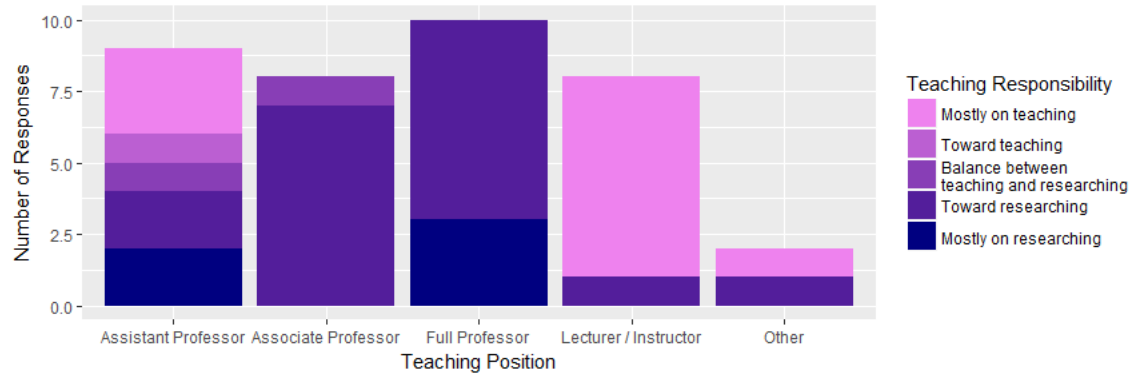


Figure 9: *The bar graph illustrates professors' Teaching Responsibility among different Teaching Positions.*

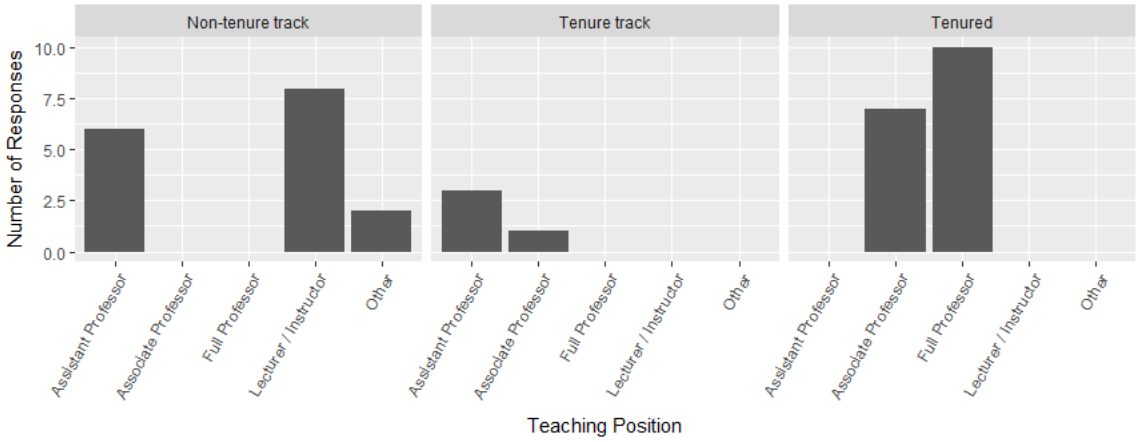


Figure 10: *The bar graph illustrates professors' Tenure Status among different Teaching Positions.*

Appendix B

Section B.1: R code used to transform original dataset into model dataset

```
library(data.table)
data<-fread("Original Data.csv")
  ### deal with NA and define burnout scores and type
datasub <- data[,2:12]

datasub$burnout <- data$Question1 + data$Question2 + data$Question3 +
  data$Question4 + data$Question5 + data$Question6 + data$Question7 +
  data$Question8 + data$Question9       #the score for burnout is sum of all question

datasub$BurnoutDegree[datasub$burnout < 17] <- "Low"
datasub$BurnoutDegree[(datasub$burnout >= 17)&(datasub$burnout<=27)] <- "Moderate"
datasub$BurnoutDegree[datasub$burnout > 27] <- "High"

#sub missing value
x <- datasub$age
y <- datasub$PSUteachyr
z <- x - y
ageM <- mean(z, na.rm = TRUE)

for (i in 1:length(x) ) {
  if (is.na(x[i]))
  { x[i] <-  ageM + y[i] }
}
datasub$age <- x
####################################
x <-data$department
for (i in 1:length(x)) {
  if (is.na(x[i]))
  {
    x[i] <- "Unknown"
  }
  else if (nchar(x[i])<3)
  {x[i]<-"Unknown"}
}
datasub$department <- x

#####Final data
datamodel <- datasub[-c(2,16),]
write.csv(datamodel, "Model Data.csv", row.names = FALSE)
```

Section B.2: R code used to graph plots in EDA

```
# read final data
data<-read.csv('Model Data.csv')
attach(data)

# Correlation
plot(PSUteachyr,teachyr,main = 'Correlation between Total Teaching Year \n & PSU Teaching Year',
```

```
    xlab='PSU Teaching Year', ylab='Total Teaching Year')
plot(burnout~workhr, main = 'Relationship between Working Hour & Burnout Level',
    xlab='Working Hour', ylab='Burnout Level')
abline(lm(burnout~workhr), col='red')
plot(burnout~PSUteachyr, main = 'Relationship between PSU\nTeaching Year & Burnout Level',
    xlab='PSU Teaching Year', ylab='Burnout Level')
abline(lm(burnout~PSUteachyr), col='red')

# Boxplots
boxplot(burnout~gender, main = 'Burnout Level by Gender', xlab='Gender', ylab='Burnout Level')
boxplot(burnout~ethnicity, main = 'Burnout Level by Ethnicity', xlab='Ethnicity', ylab='Burnout Level')
boxplot(burnout~position, main = 'Burnout Level by Teaching Position', xlab='Teaching Position', ylab='Burnout
Level')
boxplot(burnout~tenure, main = 'Burnout Level by Tenure Status', xlab='Tenure Status', ylab='Burnout Level')
boxplot(burnout~department, main = 'Burnout Level by Department', xlab='Department', ylab='Burnout Level',
names=c('Astronomy', Biochemistry','Biology','Chemistry','Mathematics','Science','Statistics','Unknown'))
boxplot(burnout~responsibility, names=c('Mostly on teaching', 'Toward teaching',
                                'Balance between teaching and researching',
                                'Toward researching', 'Mostly on researching'),
                              main = 'Burnout Level by Teaching Responsibility', xlab='Gender', ylab='Burnout Level')

# Bar charts
ggplot(data=data,aes(x=position,fill=as.factor(responsibility))) +
  labs(x='Teaching Position', y='Number of Responses') +
  geom_bar() +   scale_fill_manual(name='Teaching Responsibility',
              values=scales::seq_gradient_pal("violet", "navy", "Lab")(seq(0,1,length.out=5)),
              breaks=c("1", "2", "3","4","5"),
              labels=c("Mostly on teaching", "Toward teaching", "Balance between\nteaching and researching",
"Toward researching", "Mostly on researching"))
ggplot(data=data,aes(x=position))+
  geom_bar() + labs(x='Teaching Position', y='Number of Responses') +
  theme(axis.text.x=element_text(angle=60,hjust=1))+ facet_wrap(~tenure,ncol=3)
```

Section B.3: R code used to build model and check assumptions required to build a Multiple Linear Regression Model

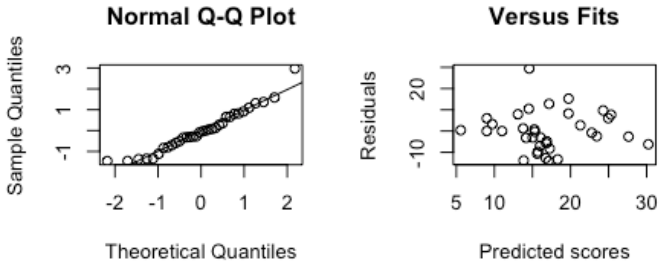
```
#read data
data<-read.csv('Model Data.csv')
#delete missing variables
data<-na.omit(data)
#check variable type
str(data)
#treat variable responsibility as a factor
data$responsibility<-as.factor(data$responsibility)

attach(data)
library(car)

#full model
fullmodel<-
lm(burnout~gender+age+ethnicity+PSUteachyr+status+workhr+position+responsibility+tenure+department)
```

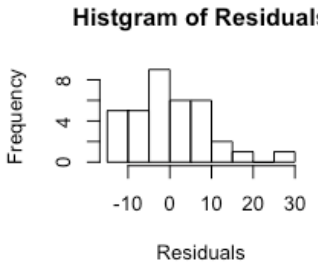
```
summary(fullmodel)
vif(fullmodel)
#check assumption for full model
#residual plots
par(mfrow=c(2,2))
fullmodel.std<-rstandard(fullmodel)
qqnorm(fullmodel.std)
qqline(fullmodel.std)
plot(fitted(fullmodel),residuals(fullmodel),xlab="Predicted scores",ylab="Residuals",main="Versus Fits")
hist(fullmodel$residuals,xlab="Residuals",ylab="Frequency",main="Histgram of Residuals")
plot(fullmodel$residuals,xlab="Observation Number",ylab="Residuals",main="Versus Order")

#final model
model<-lm(burnout~PSUteachyr+workhr+responsibility)
summary(model)
vif(model)
#check assumption for final model
#residual plots
par(mfrow=c(2,2))
model.std<-rstandard(model)
qqnorm(model.std)
qqline(model.std)
plot(fitted(model),residuals(model),xlab="Predicted scores",ylab="Residuals",main="Versus Fits")
hist(model$residuals,xlab="Residuals",ylab="Frequency",main="Histgram of Residuals")
plot(model$residuals,xlab="Observation Number",ylab="Residuals",main="Versus Order")
```

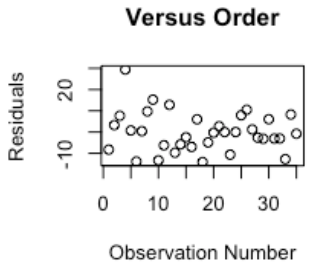
Appendix C

Section C.1: Output and residual plot for final model

```
> model<-lm(burnout~PSUteachyr+workhr+responsibility)
> summary(model)

Call:
lm(formula = burnout ~ PSUteachyr + workhr + responsibility)

Residuals:
     Min       1Q   Median       3Q      Max
-14.1299  -6.0168  -0.3135   6.0182  29.4240

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6.3638    11.1715  -0.570   0.5735
PSUteachyr       -0.3720     0.1869  -1.990   0.0564 .
workhr            0.5965     0.2323   2.567   0.0159 *
responsibility2 -13.5826    11.0384  -1.230   0.2288
responsibility3   1.1962     8.0673   0.148   0.8832
responsibility4  -4.4203     4.6251  -0.956   0.3474
responsibility5  -4.1161     5.9235  -0.695   0.4929
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.25 on 28 degrees of freedom
Multiple R-squared:  0.2539,    Adjusted R-squared:  0.09405
F-statistic: 1.588 on 6 and 28 DF,  p-value: 0.1875

> vif(model)
                   GVIF Df GVIF^(1/(2*Df))
PSUteachyr     1.142094  1        1.068688
workhr         1.315680  1        1.147031
responsibility 1.333319  4        1.036613
```







The following assumptions were checked to use Multiple Linear Regression Model for Burnout

Assumption1: There is a linear relationship between Explanatory variables and the response variable. We checked the qq-plot for assumption1, most dots fit to a straight line expect the two on both ends.

Assumption2: The errors are independent. As we scanned the Versus Fit plot from left to right, predicted scores between 15 to 20 had residuals below 0, and the variation of the residuals appeared to be inconstant. However, given the small sample size, there was little to suggest violation of the normality assumption.

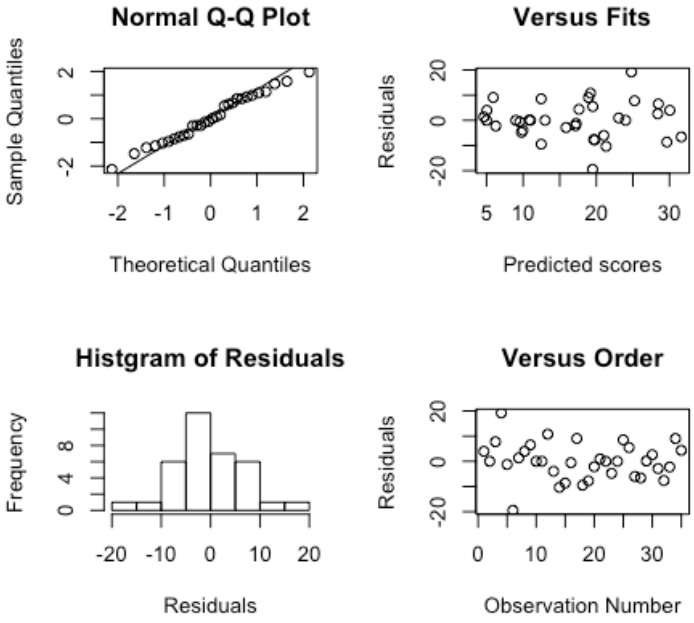
Assumption3: The errors at each set of values of the predictors are normally distributed. The histogram was a little skewed to the right. Still, given the small sample size, there was little to suggest violation of the normality assumption.

Assumption4: The errors at each set of values of the predictors have equal variances. As we scanned the Versus Order plot from left to right, the residuals remained approximately 0, the variation of the residuals appeared to be roughly constant, and there was no excessively outlying point.

Section C.2: Output and residual plot for full model

```
> fullmodel<-lm(burnout~gender+age+ethnicity+PSUteachyr+status+workhr+position+responsibility+tenure+department)
> summary(fullmodel)

Call:
lm(formula = burnout ~ gender + age + ethnicity + PSUteachyr +
    status + workhr + position + responsibility + tenure + department)

Residuals:
    Min      1Q  Median      3Q     Max
-19.494  -4.368   0.000   4.163  19.207

Coefficients:
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                                11.1768    51.0911   0.219    0.831
genderMale                                  1.8387     6.3878   0.288    0.779
age                                        -0.9935     1.2118  -0.820    0.430
ethnicityWhite / Caucasian                  7.7846    17.7030   0.440    0.669
PSUteachyr                                  0.2124     1.0771   0.197    0.847
statusPart-time                             8.6684    18.0628   0.480    0.641
workhr                                      0.9754     0.5458   1.787    0.101
positionAssociate Professor               -17.2213    20.5582  -0.838    0.420
positionFull Professor                     -7.9577    26.0663  -0.305    0.766
positionLecturer / Instructor             -10.1790    11.0891  -0.918    0.378
positionOther                             -13.0155    18.5304  -0.702    0.497
responsibility2                           -25.6749    20.9666  -1.225    0.246
responsibility3                            -4.8801    16.7312  -0.292    0.776
responsibility4                           -15.0781    12.6620  -1.191    0.259
responsibility5                           -21.5790    13.6265  -1.584    0.142
tenureTenure track                         -0.2473    12.6519  -0.020    0.985
tenureTenured                              10.6523    24.8899   0.428    0.677
departmentBiochemistry & Molecular Biology 13.6528    13.6948   0.997    0.340
departmentBiology                           6.7838    10.5222   0.645    0.532
departmentChemistry                         3.9932    12.7811   0.312    0.761
departmentMathematics                     -12.1153    17.1312  -0.707    0.494
departmentScience                          12.6590    18.2552   0.693    0.502
departmentStatistics                       -3.2165    14.3747  -0.224    0.827
departmentUnknown                           0.6568    11.2776   0.058    0.955

Residual standard error: 12.88 on 11 degrees of freedom
Multiple R-squared:  0.5376,    Adjusted R-squared:  -0.4292
F-statistic: 0.5561 on 23 and 11 DF,  p-value: 0.8865
```

```
> vif(fullmodel)
                     GVIF Df GVIF^(1/(2*Df))
gender           2.108615  1        1.452107
age             32.602273  1        5.709840
ethnicity        3.562923  1        1.887571
PSUteachyr      24.038779  1        4.902936
status           1.910830  1        1.382328
workhr           4.603718  1        2.145628
position       513.025047  4        2.181561
responsibility  48.437959  4        1.624233
tenure          72.728670  2        2.920293
department     108.750046  7        1.397846
```



The following assumptions were checked to use Multiple Linear Regression Model for Burnout

Assumption1: There is a linear relationship between Explanatory variables and the response variable. We checked the qq-plot for assumption1, most dots fit to a straight line, with a few points off on both ends.

Assumption2: The errors are independent. As we scanned the Versus Fit plot from left to right, the residuals remained approximately 0, the variation of the residuals appeared to be roughly constant, and there was no excessively outlying point.

Assumption3: The errors at each set of values of the predictors are normally distributed. The histogram showed that the data appeared to be normal distributed.

Assumption4: The errors at each set of values of the predictors have equal variances. As we scanned the Versus Order plot from left to right, the residuals remained approximately 0, the variation of the residuals appeared to be roughly constant, and there was no excessively outlying point.