Final Project

Evelyn Delph, Thomas James, & Nick Mobley

4/21/2020

Introduction

Abstract

Are undergraduate students receiving the upmost quality of education from there professors at Indiana University? Since instate students pay $10,948 for tuition and out of state students pay $36,512, it is important to determine if such expenditures are worth the cost. For the purposes of this report, we are examining if an instructors' grade distrubition for undergraduate courses is indicative instructors' salaries. For this study, the quality of education that an instructor provides is determined by their average GPA. We are interested in studying grade distrubtion and salary across other variables including: the department, the course level, the instructors' level of teaching, and the year. The goals of this report aim to answer the following questions:

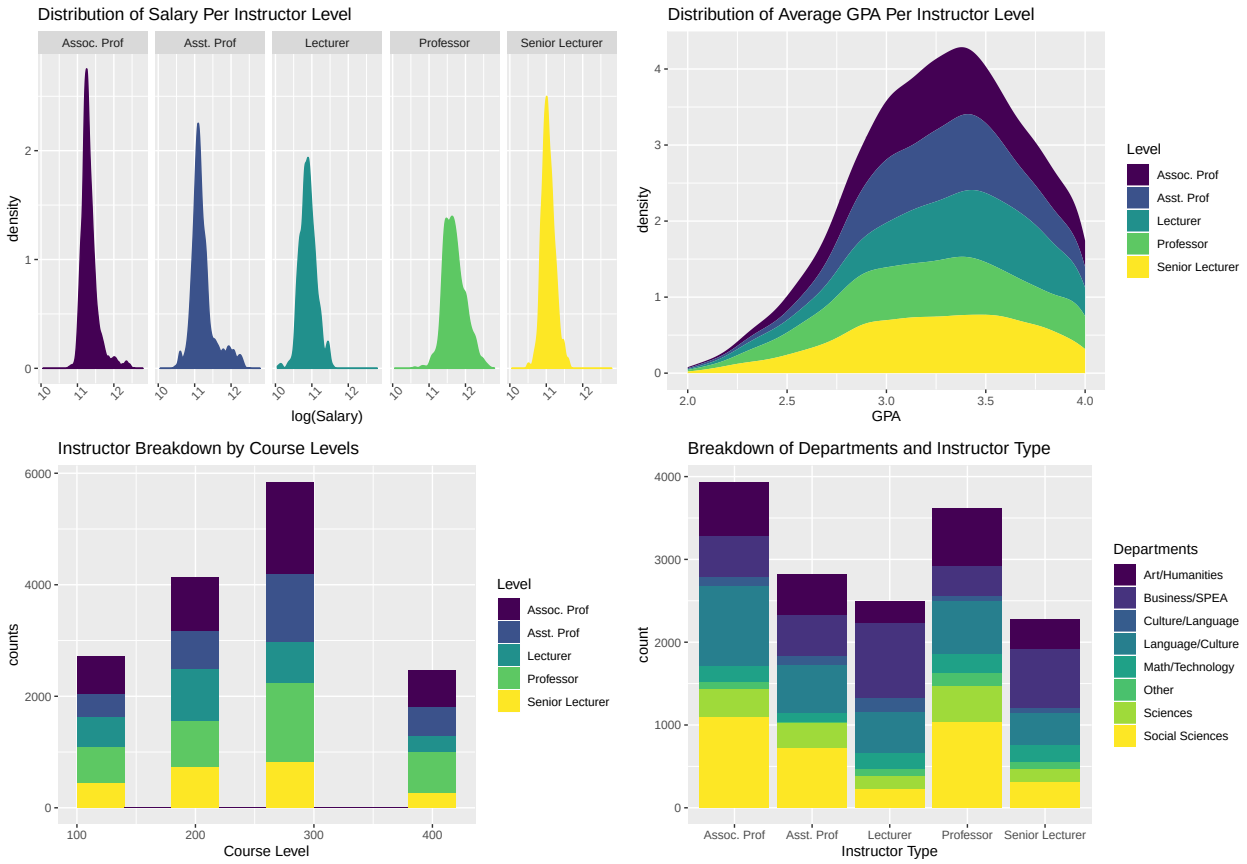
- Does GPA impact an instructor's salary?
- Do GPA and salary relationships vary across course levels?
- Do GPA and salary relationships vary across years?
- Do GPA and salary relationships vary across different departments?

We will answer these questions with numerous EDA techniques to study the relationships among these variables. We aim to achieve a better understanding if the quality of education impacts instructors' salaries.

Description of Data

Our master data file consists of 9 grade distribution and 9 salary files from the IU Registrar between the years 2010-2018. The datasets were joined by the instructor. The variables of interest pulled from these datasets include: Name of Instructor, Department, Instructor Level, Salary, Year, Average GPA, Number of Students, and Course Level.

```
DESCRIPTION
Total Rows: 15,163
Total Columns: 7
1. Name of Instructor: 1,942 unique instructors (string)
2. Departments: 100 departments at IU categorized into 8 groups (string)
3. Instructor Level: The title of an instructor (string)
4. Salary: The annual salary of a professor (integer)
5. Average GPA: Average GPA grouped by course and instructor, between 2.0-4.0 (double)
6. Number of Students: Number of students grouped by course and instructor (integer)
7. Course Level: Denotes 100, 200, 300, and 400 level classes (integer)
```



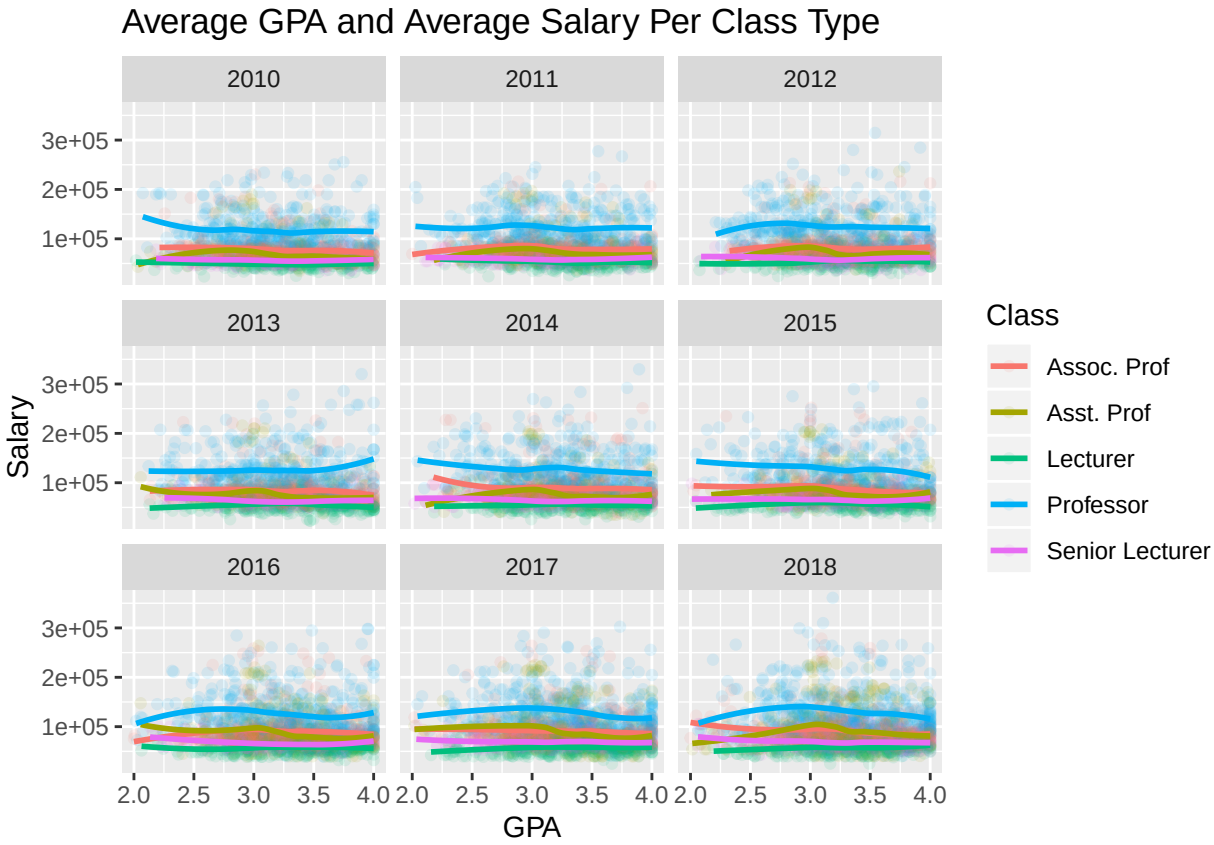
Modeling

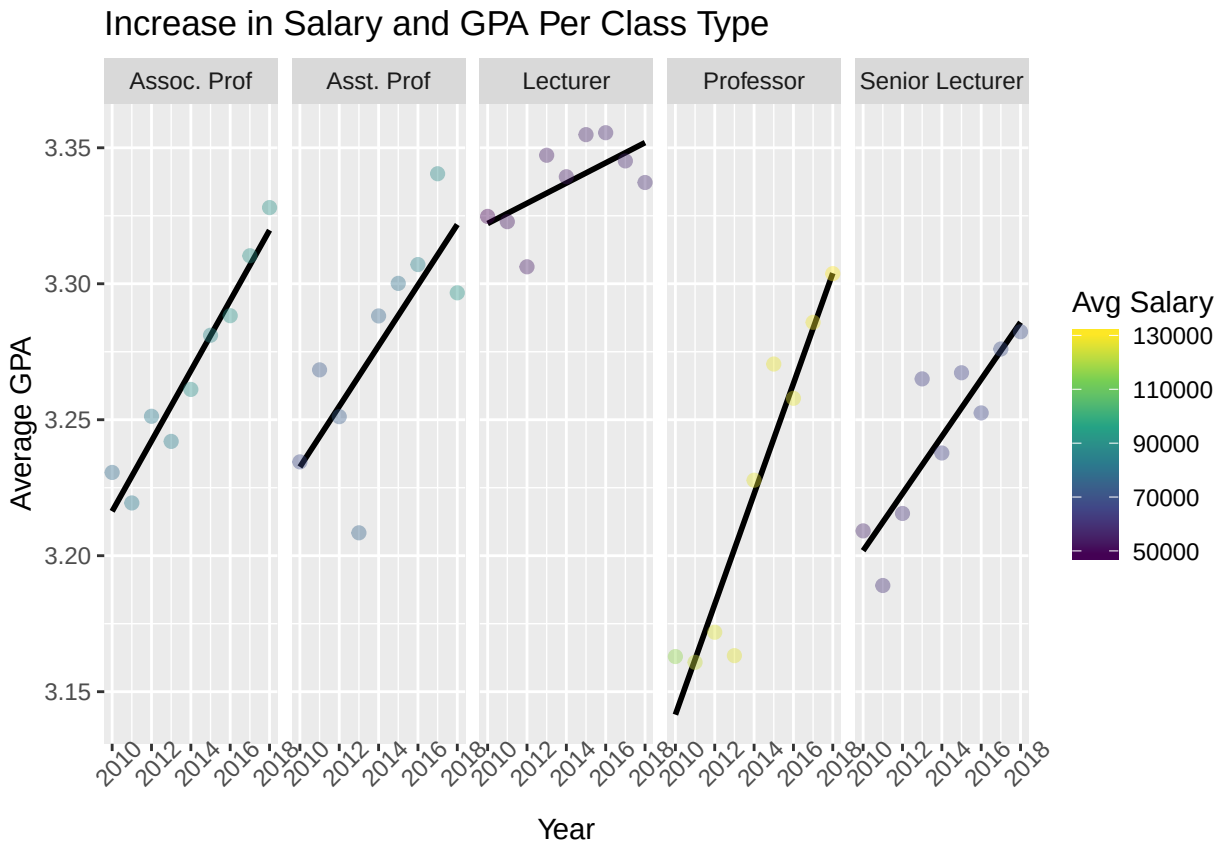
. . . Have to work on this. . .

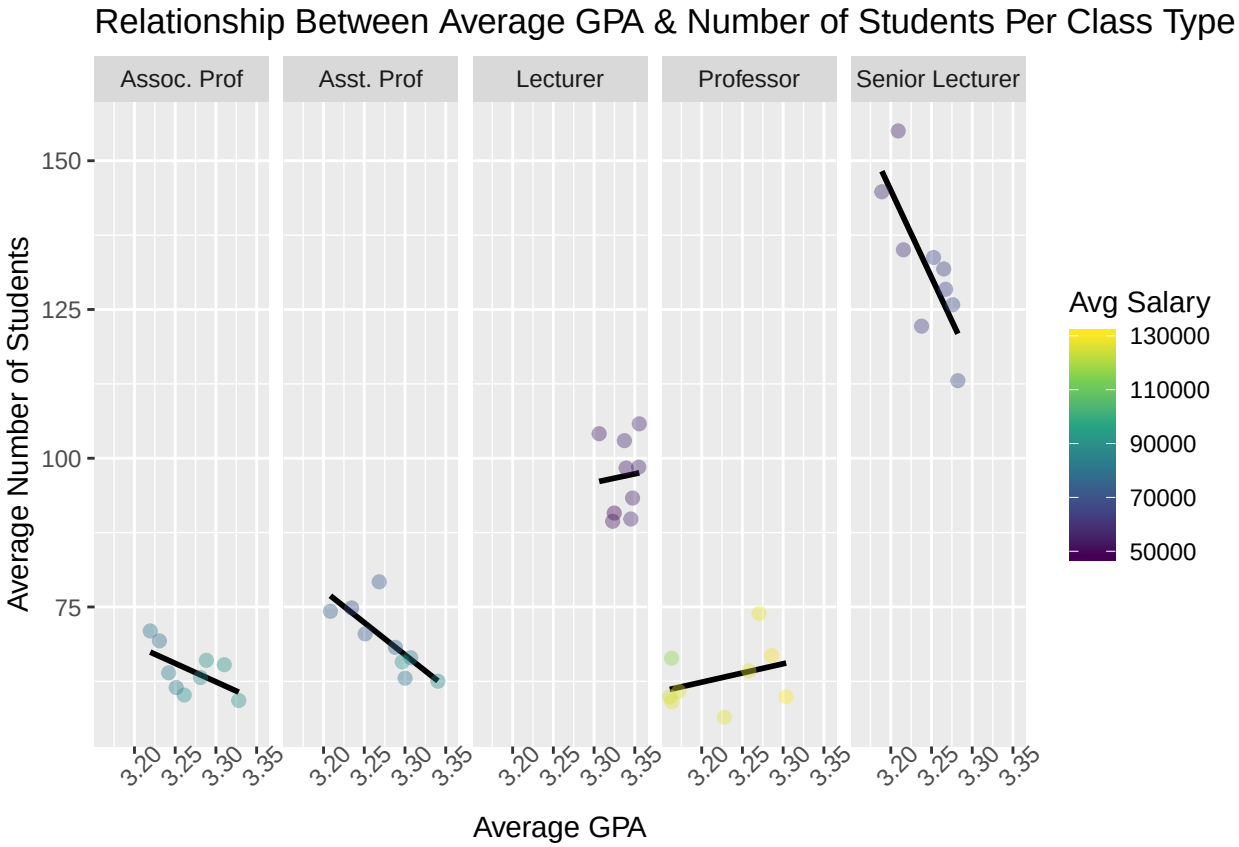
- First graph: Can we see a visible trend between gpa and salary over the 8 year span?
- Second graph: Can we see a visible trend between gpa, salary, and number of students per class type?

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

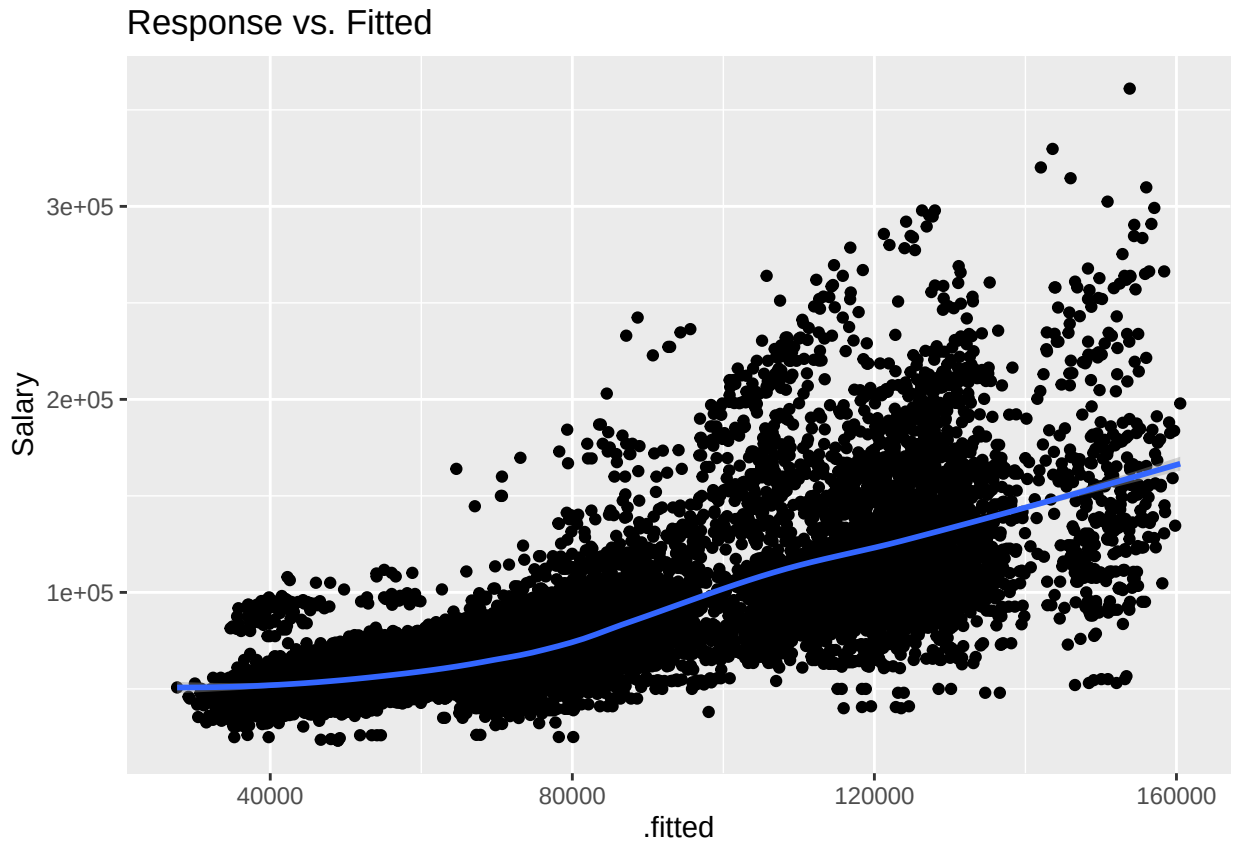

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```


Average GPA and Average Salary Per Class Type
2010
2011
2012
2013
2014
2015
2016
2017
2018
3e+05
2e+05
1e+05
Salary
2.0
2.5
3.0
3.5
4.0
GPA
Class
Assoc. Prof
Asst. Prof
Lecturer
Professor
Senior Lecturer

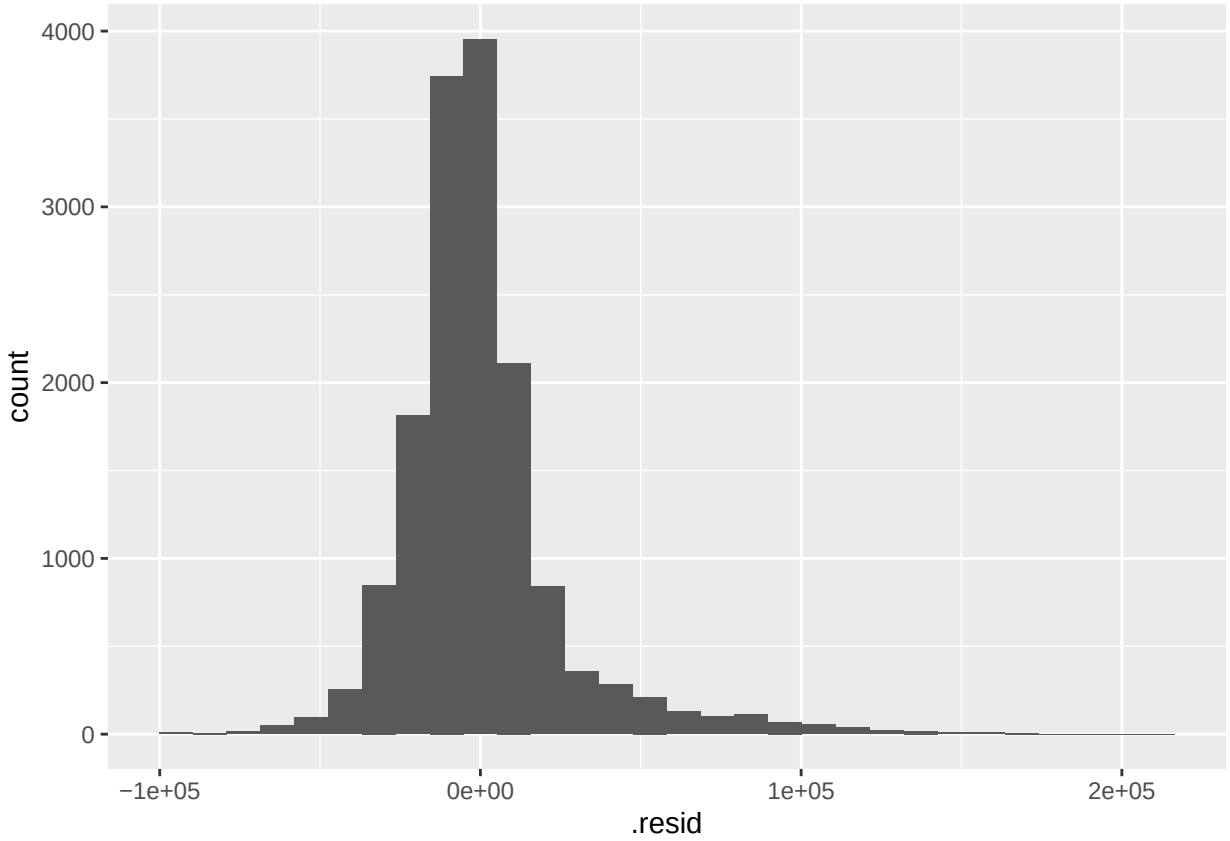

Increase in Salary and GPA Per Class Type
Assoc. Prof
Asst. Prof
Lecturer
Professor
Senior Lecturer
Average GPA
3.35
3.30
3.25
3.20
3.15
2010
2012
2014
2016
2018
Year
Avg Salary
130000
110000
90000
70000
50000


Relationship Between Average GPA & Number of Students Per Class Type
Assoc. Prof
Asst. Prof
Lecturer
Professor
Senior Lecturer
Average Number of Students
150
125
100
75
3.20
3.25
3.30
3.35
Average GPA
Avg Salary
130000
110000
90000
70000
50000

Part 2: Models/Predictions

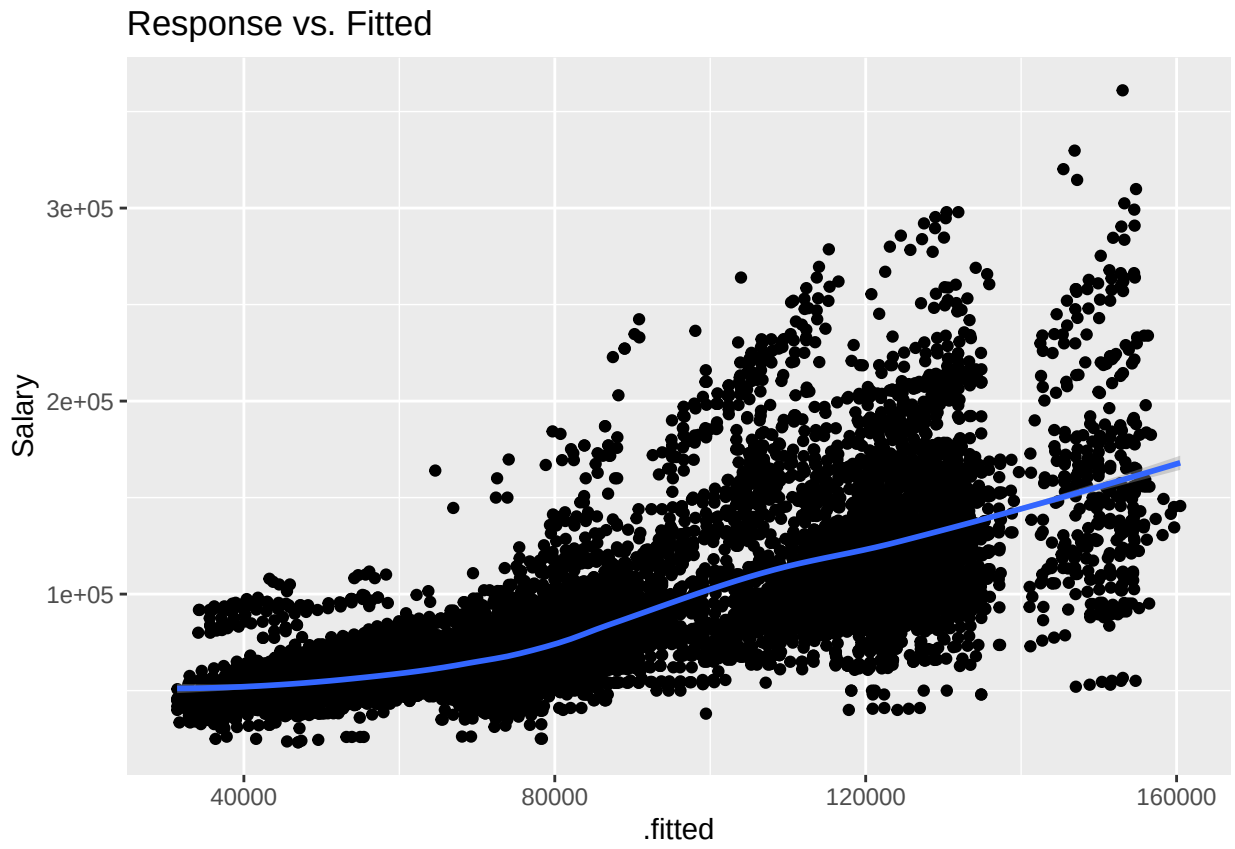



```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

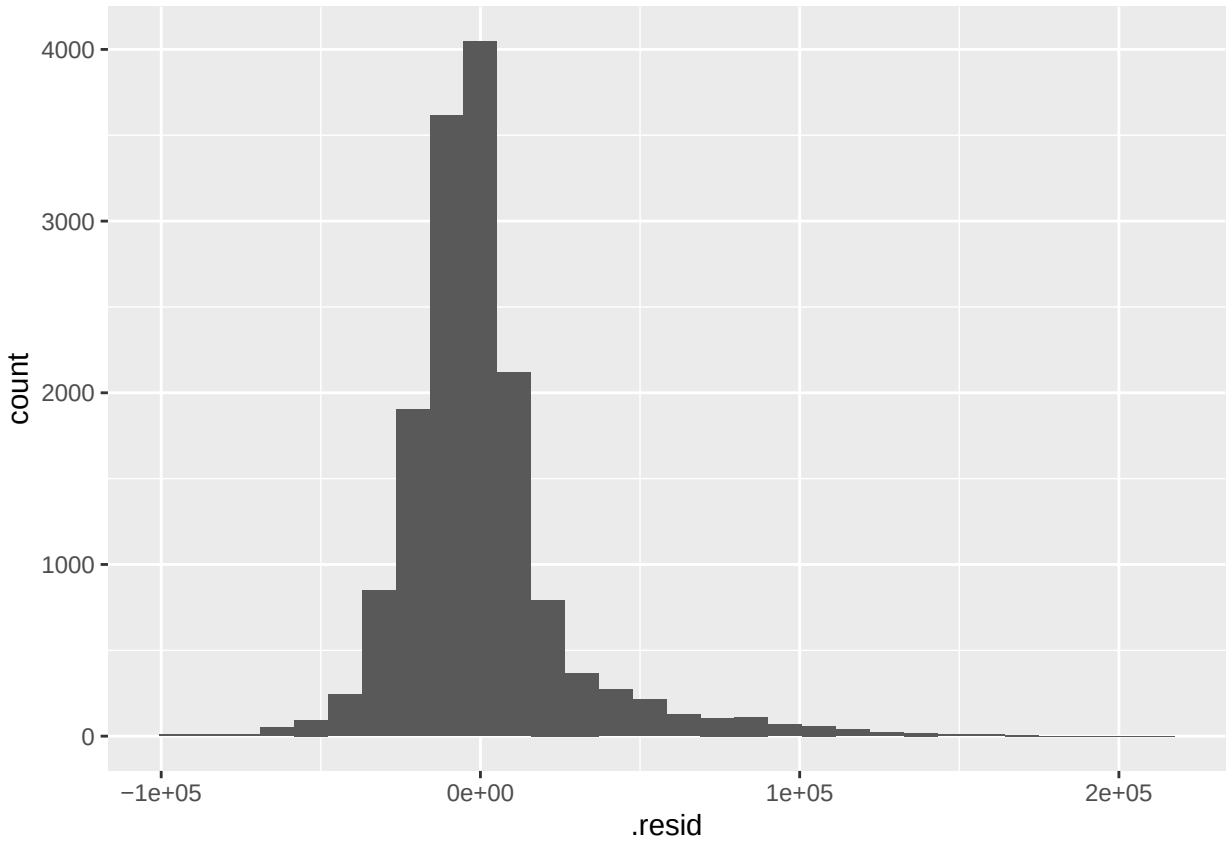


- Does GPA impact an instructor's salary?
- Do GPA and salary relationships vary across course levels?
- Do GPA and salary relationships vary across years?
- Do GPA and salary relationships vary across different departments?

To model the relationship between salary and our feature variables of interest, a generalized additive model (GAM) was fitted on the data. As we can see from the plot of Salary vs. the GAM's fitted values, the model does an adequate job of prediction on the data. The trend of this graph seems to be nearly linear. In addition, the distribution of the model's residuals is relatively close to normal. Based on our model, it does not seem that GPA has any sort of influential impact on Salary.




```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



After removing GPA from our model, the trend of the model's predictions vs. the actual response variable is still linear, and the residuals are still largely normally distributed.

Part 3: Analysis + Conclusion