

MATH 130 Exploratory Data Analysis - GPA

Michael Pratt

2024-02-27

SPECIAL NOTE ON THE DATA FOR THIS PROJECT

Data for this project is from the Office of Graduate Studies at Chico State provided through the Data Acquisition and Technical Analysis (DATA) Workgroup of the Grad Tech Committee. It is composed of data collected for Fall Admits for data that is at least five years old. Due to FERPA Privacy Laws, the exact dates of admits cannot be disclosed.

```
library(readxl)
gpa <- read_excel("C:/Users/mpratt4/Box/GRAD/SPECIAL PROJECTS/GRAD Tech Cmte/Data Acquisition and Techn
head(gpa)
```

```
## # A tibble: 6 x 6
##   empl_ID designation          ba_school ugrd_gpa comp_dgr grad_gpa
##   <dbl> <chr>                   <chr>     <dbl>   <dbl>   <dbl>
## 1     1 Psychology - Marriage & Family T~ CALIFORN~ 3.68     1     3.98
## 2     2 English                   CALIFORN~ 3.4      1     3.97
## 3     3 Psychology - Psychological Scien~ CALIFORN~ 2.95     0     3.64
## 4     4 Geosciences - Hydrology         CALIFORN~ 3.5      0     3.66
## 5     5 Kinesiology                 CALIFORN~ 2.69     1     3.75
## 6     6 Public Administration          CALIFORN~ 4        1     3.9
```

Introduction

This project aims to explore the relationship between undergraduate cumulative Grade Point Average (GPA) at the time of admission to graduate school and its impact on two crucial outcomes: the final graduate GPA and the successful completion of the graduate program, culminating in the earning of a Master's degree. The primary variables of interest in this study are: 1. Undergraduate GPA - The cumulative GPA at the time of admission to graduate school. 2. Final Graduate GPA - The cumulative GPA upon completion of the graduate program. 3. Program Completion - A binary variable indicating whether the student successfully completed the graduate program and earned a Master's degree.

Research Question

The research question is: "How does the undergraduate GPA at the time of admission to graduate school influence the final graduate GPA and the likelihood of successful program completion resulting in earning a Master's degree?" The main interest is in understanding whether a higher undergraduate GPA is a strong predictor of success in graduate studies, both in terms of academic achievement (as measured by GPA) and program completion.

Univariate Exploration

Undergraduate GPA

Summary Statistics

The minimum value indicated in the summary is significantly lower than the 2.5 GPA required for entry to graduate school under Title 5. This is because they are considered Special Admits. The applicant has to essentially make a compelling case to the Program Coordinator to be admitted. These are meant to be exceptionally rare and have mitigating circumstances. Even though these might be considered outliers, I am keeping them in to determine if they have any interesting interactions in the analysis.

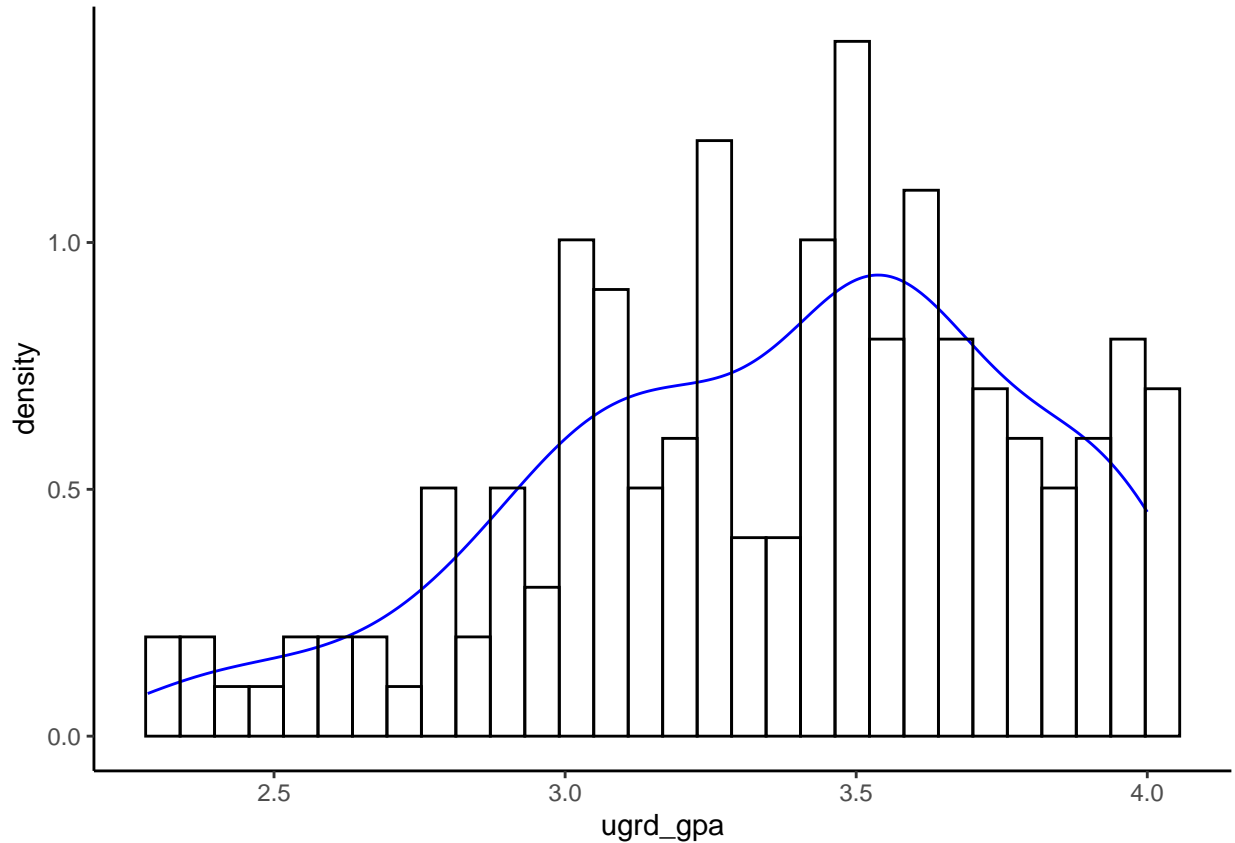
```
summary(gpa$ugrd_gpa)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  2.283   3.086   3.447   3.372   3.663   4.000
```

Graphic

To get a better sense of the distribution in the histogram, I utilized the density overlay to visualize the data more clearly. What is demonstrated is a relatively normal bell-curve distribution with a slight left-skew. This may indicate that we have slightly selective admit processes for some of our programs, but overall we are appropriately selective given the mission of the university as a public institution.

```
ggplot(gpa, aes(x=ugrd_gpa)) + geom_density(col="blue") +
  geom_histogram(aes(y=..density..), color="black", fill=NA) + theme_classic()
```



Final Graduate GPA

Summary Statistics

The minimum is significantly below the 3.0 cumulative GPA needed to stay in our graduate programs. That is appropriate because several admits are academically disqualified each year, so the dataset includes them.

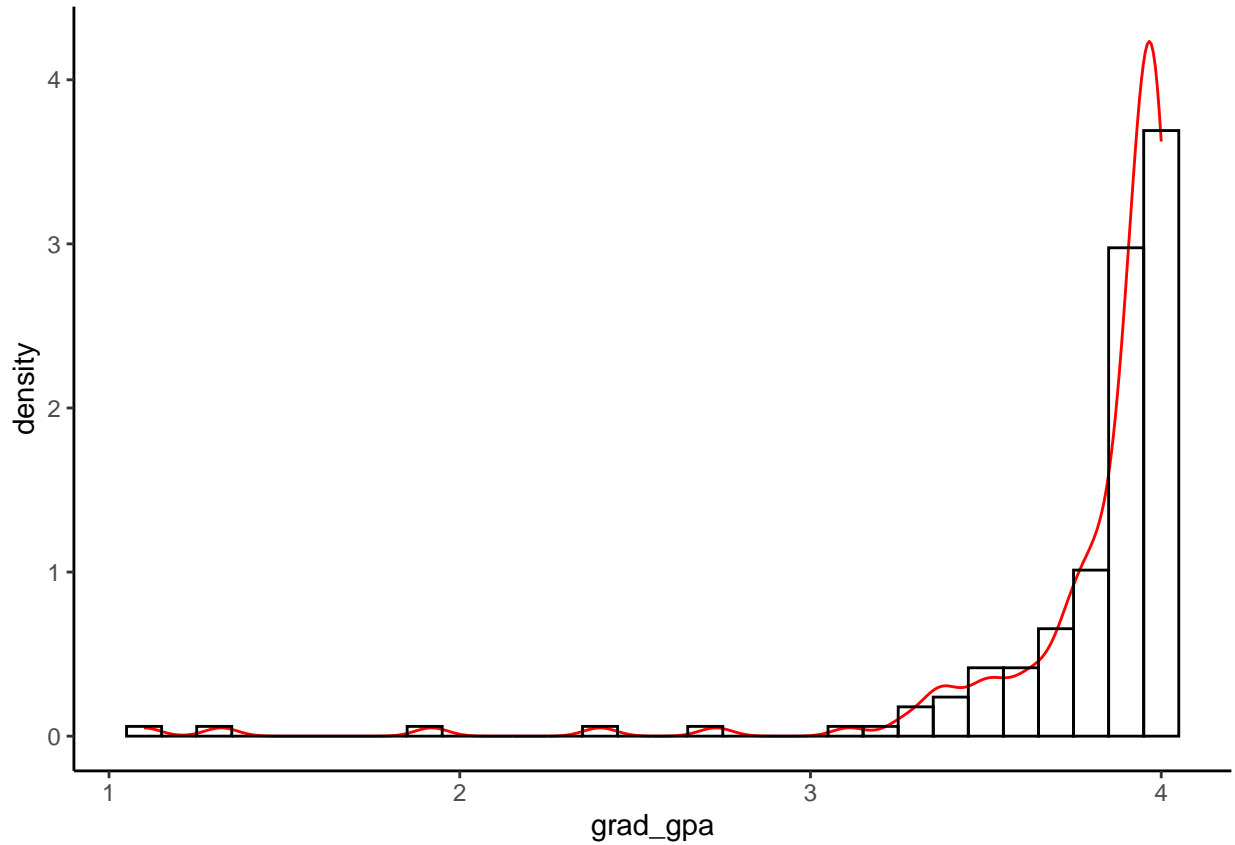
```
summary(gpa$grad_gpa)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.100   3.785   3.930   3.797   3.980   4.000
```

Graphic

The subpar GPAs are noted as below 3.0 and would typically be seen as outliers in the graph. Of greater interest is the intense left-skew of the graph instead of a normal-ish bell-curve distribution between 3.0 and 4.0 on the GPA scale. This might indicate grade inflation of some programs- or it could be a normal distribution for grad school.

```
ggplot(gpa, aes(x=grad_gpa)) + geom_density(col="red") +
  geom_histogram(aes(y=..density..), color="black", fill=NA) + theme_classic()
```



Program Completion

Summary Statistics

A value of '1' indicates that the student earned a Master's degree. A value of '0' indicates that the student did not earn a Master's degree. 140/168 or 83.33% of students in the dataset earned a Master's degree. 28/168 or 16.67% of students in the dataset did not earn a Master's degree.

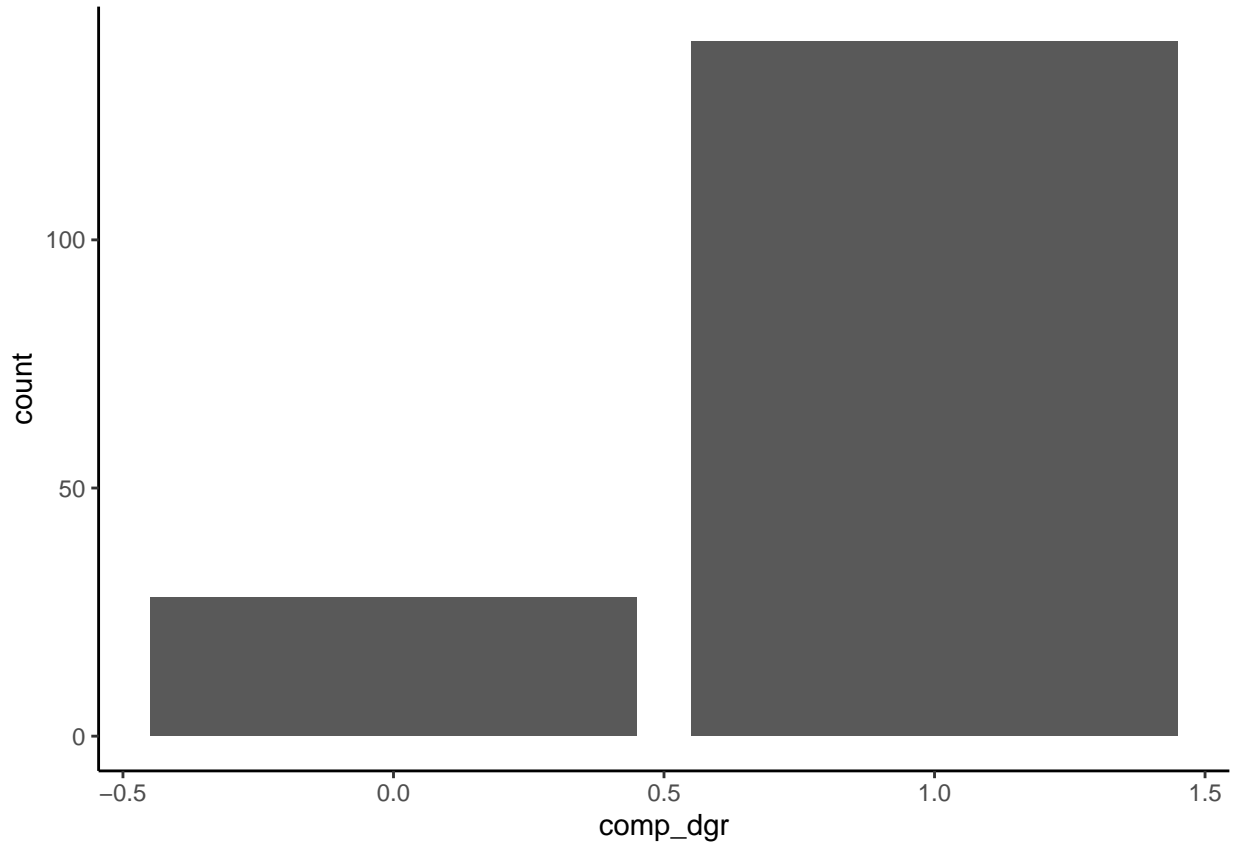
```
table(gpa$comp_dgr)
```

```
##
##  0  1
## 28 140
```

Graphic

The graphical representation highlights that the vast majority of admitted students do go on to earn their degrees.

```
ggplot(gpa, aes(x=comp_dgr)) + geom_bar() + theme_classic()
```



Bivariate Exploration

Undergraduate GPA and Graduate GPA

Grouped Summary Statistics

The slight increase between each of the ranges demonstrates that there is a slight positive correlation between undergraduate cumulative GPA and final graduate GPA.

```
#group undergrad GPA into more manageable categories or ranges
gpa$ugrd_gpa_ranges <- cut(gpa$ugrd_gpa, breaks = c(0, 2.5, 3.0, 3.5, 4.0),
                           include.lowest = TRUE, labels = c("0-2.5", "2.5-3.0",
                                                             "3.0-3.5", "3.5-4.0"))

#calculate graduate GPA within each undergraduate GPA range
avg_grad_ugrd_gpa <- aggregate(gpa$grad_gpa ~ ugrd_gpa_ranges, data = gpa,
                               FUN = mean)

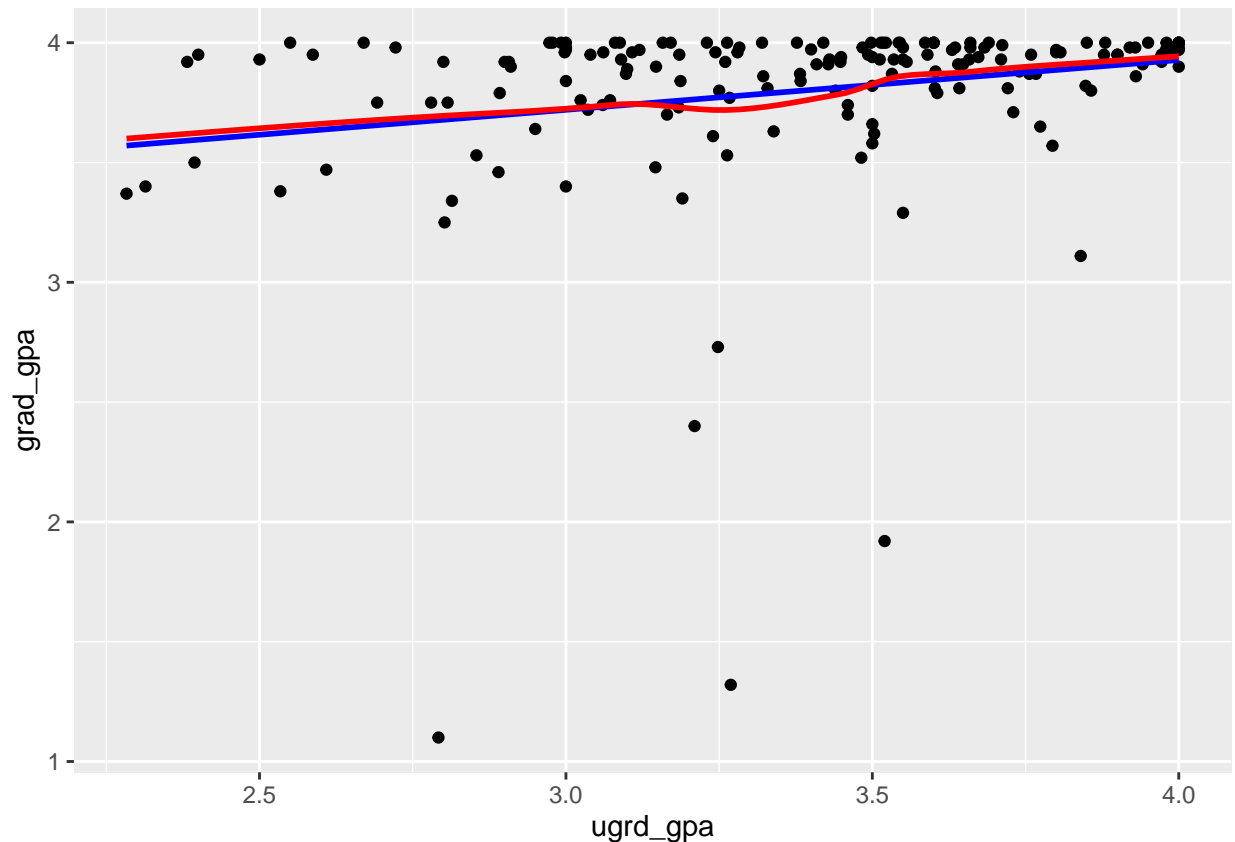
print(avg_grad_ugrd_gpa)
```

```
##   ugrd_gpa_ranges gpa$grad_gpa
## 1      0-2.5      3.678333
## 2     2.5-3.0      3.687931
## 3     3.0-3.5      3.764787
## 4     3.5-4.0      3.878750
```

Graphic

The graph below appears to show that while nearly everyone admitted can do well on their GPA in grad school, there is a slight positive correlation between UGPA and GGPA.

```
ggplot(gpa, aes(x=ugrd_gpa, y=grad_gpa)) + geom_point() +  
  geom_smooth(se=FALSE, method="lm", color="blue") +  
  geom_smooth(se=FALSE, color="red")
```



Undergraduate GPA and Program Completion

Grouped Summary Statistics

The values for completion rate should go up with GPA, but they seem to defy that expectation. I'm not sure what might cause this phenomenon. My best guess is that people with above a 3.5 GPA in undergrad are probably already well-equipped to handle graduate school and are well-informed when it comes to how to access resources and assistance. The sub-2.5 GPAs are probably accomplishing a higher completion rate because they know they have to push themselves to succeed and they might be more willing to seek out assistance and resources. The group in the middle might have typically just gotten by, and as a result fail to pursue assistance even when they might benefit from it.

```
avg_comp_ugrd_gpa <- aggregate(gpa$comp_dgr ~ gpa$ugrd_gpa_ranges, data = gpa,  
  FUN = mean)  
print(avg_comp_ugrd_gpa)
```

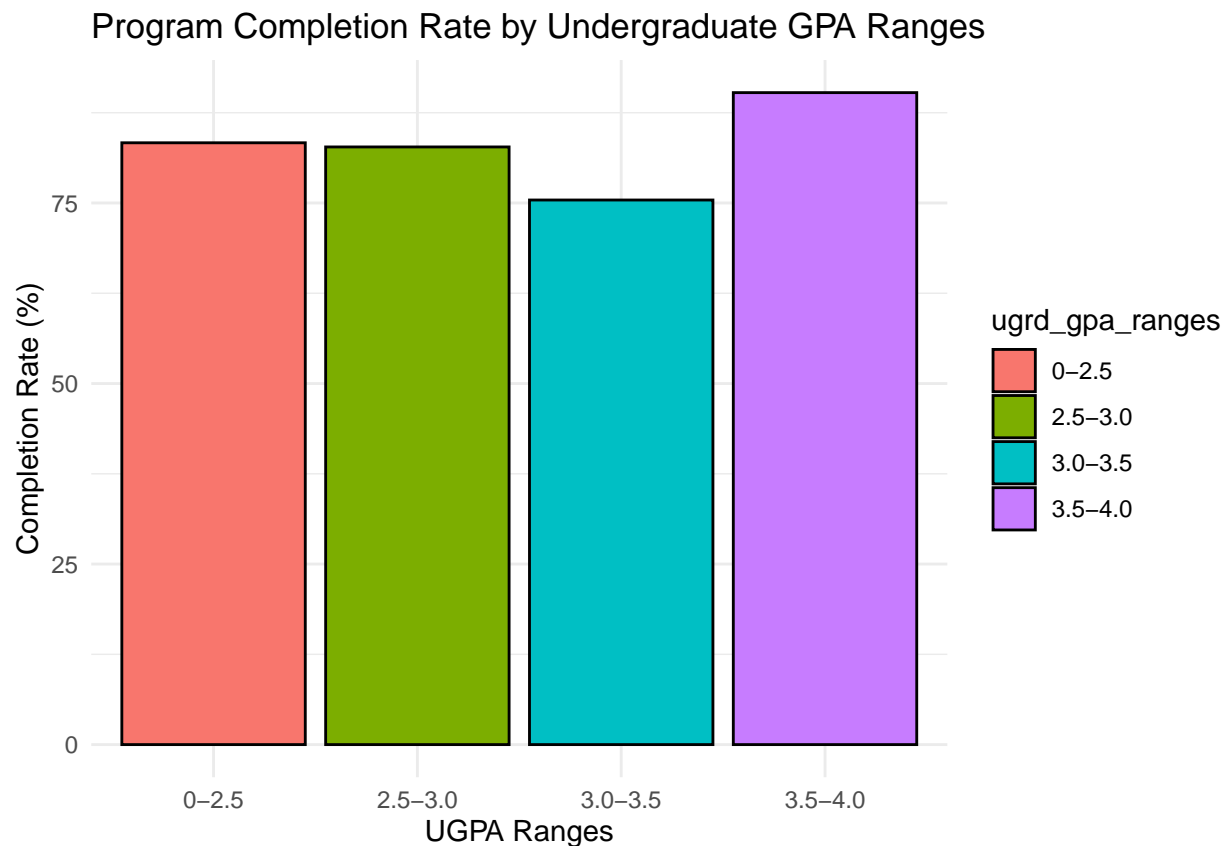
```
##   gpa$ugrd_gpa_ranges gpa$comp_dgr
## 1          0-2.5      0.8333333
## 2          2.5-3.0      0.8275862
## 3          3.0-3.5      0.7540984
## 4          3.5-4.0      0.9027778
```

Graphic

The graph certainly paints an interesting picture that is completely out of line with what I would have believed going into this exploratory data analysis project.

```
#Calculating the percentage of program completion for each UGRD GPA range
completion_rates <- gpa %>%
  group_by(ugrd_gpa_ranges) %>%
  summarize(Completion_Rate = mean(comp_dgr) * 100)

ggplot(completion_rates, aes(x = ugrd_gpa_ranges, y= Completion_Rate, fill =
                             ugrd_gpa_ranges)) +
  geom_bar(stat = "identity", color = "black") +
  theme_minimal() +
  labs(title = "Program Completion Rate by Undergraduate GPA Ranges",
       x = "UGPA Ranges",
       y = "Completion Rate (%)",
       scale_fill_brewer(palette="Pastell1"))
```



Conclusion

The exploration into the impact of undergraduate GPA (UGPA) on graduate school outcomes has yielded some insightful considerations. By categorizing UGPA into ranges and analyzing the corresponding final graduate GPA (GGPA) and program completion rates, a notable trend has been observed that suggests a positive correlation between UGPA and academic success in graduate school with regard to GGPA, but perhaps not quite the same is true for UGPA's impact on graduate school completion rates. This is certainly not what I expected to find to be the case because all of the literature seemed to suggest the rate of graduate school completion tied very strongly to UGPA.

Key Findings

UGPA and GGPA Relationship

Students with higher UGPAs tend to achieve higher GGPAs, indicating that undergraduate academic performance is a strong predictor of graduate school success. This correlation underscores the importance of foundational undergraduate education in preparing students for the rigor of graduate study.

UGPA and Program Completion

The analysis also revealed that UGPA may not be the best indicator of graduate school completion. This suggests that some other variable may be at work when it comes to the successful completion of graduate school.

Implications

These findings emphasize the potential predictive value of UGPA in graduate admissions processes. However, it's also critical to maintain a holistic view of each applicant, considering the diverse factors that contribute to academic and professional success.

Limitations and Future Research

While this study provides valuable insights, it is important to acknowledge its limitations. The analysis does not account for external factors such as financial support, personal circumstances, or the specific nature of the graduate programs. Future research could explore these variables in depth, examining how they interact with UGPA to influence graduate school outcomes.

Additionally, longitudinal studies could offer a more dynamic view of how undergraduate performance impacts long-term career success and satisfaction among graduate school alumni.