Data Science Honors Program

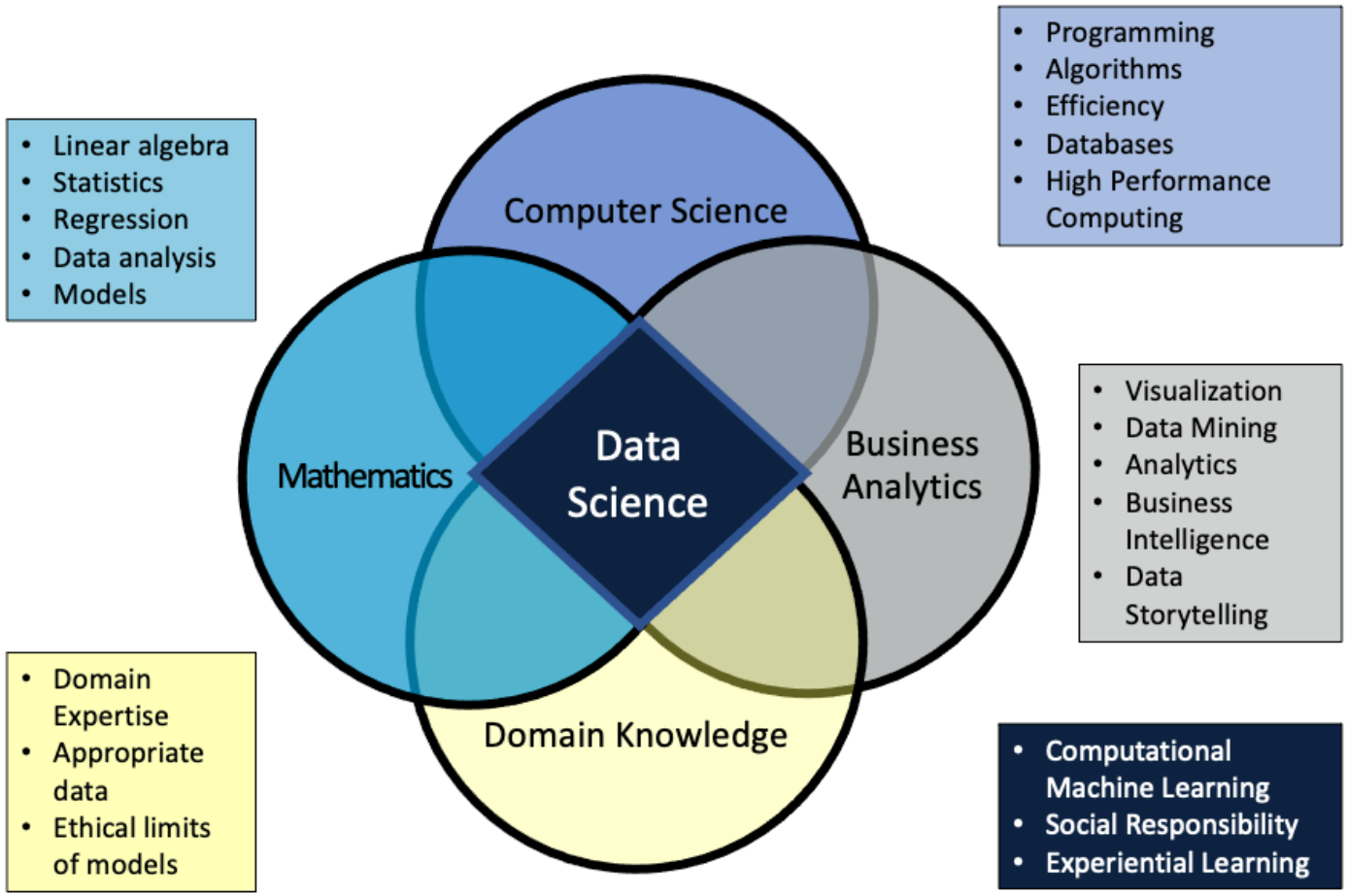
First-year Data Science Summer Internship, 2023

Analysis of Persistence and Retention Data in Xavier's 2020-23 Cohorts of Incoming Students

Eden Abernethy, Matthew Conover, Olivia Delffs, Aiden Leamer,
Doyinsola Okwesa, Jack Swaisgood, Daniel Welch, David Gerberry

Data Science Honors Program at Xavier

- Cohort-based honors program
  - 2021 Cohort: 4 students
  - 2022 Cohort: 8 students
  - 2023 Cohort: 11 students



Visiting Data Science Educator program at Microsoft

- Embedded in COSINE Data Science Team for month in Summer 2020
- Goal of giving educators first-hand experience with DS careers
- Activities fell into the following categories:
  - In-depth, small group discussions with COSINE data scientists and key leaders
  - Attendance at recurring meetings across the larger Data Team
  - Independent learning of DS topics using internal resources
  - Mock interviews.

Importance of “soft skills”

Communication ability.

Finding DS with excellent communication skills who can “tell a story with data,” is more difficult than finding the DS with necessary technical skills. One of the primary roles that DS play is that of a consultant. Even at an organization like Microsoft, DS often involves being able to successfully collaborate with individuals that lack necessary quantitative skills for a project. A successful data scientist must be well-rounded and communicate well enough to understand the question at hand in the client’s domain, work to find appropriate answers for the question and communicate the results so that appropriate actions can be taken. This is one aspect in which the liberal arts tradition of Xavier would be uniquely positioned to produce successful data scientists.

“Early and Often” project experience

- American Statistical Association (ASA) DataFest is a celebration of data in which teams of undergraduates work around the clock to find and share meaning in a large, rich, and complex data set.
  - Weekend competition @ Miami University in Spring
- First-year Data Science Summer Internship
- DSCI 210: Data and American Democracy

  Introduce 2nd-year Data Science students to ways data is collected, analyzed, and used in political campaigns and decision-making, as well as the potential consequences and ethical implications of these practices. It will achieve these goals via
  - Hands-on data analysis work done in collaboration with sophomore PPP students;
  - Discussion of ways data intersects with American democracy both past and present through readings, discussions, reflections, and essays.

First-year Summer Internship, 2023

- Organized primarily by Craig This and team in Office of Institutional Research

Question #1: For students who did not return (persist) from Fall term to Spring term, what variables associate with them?

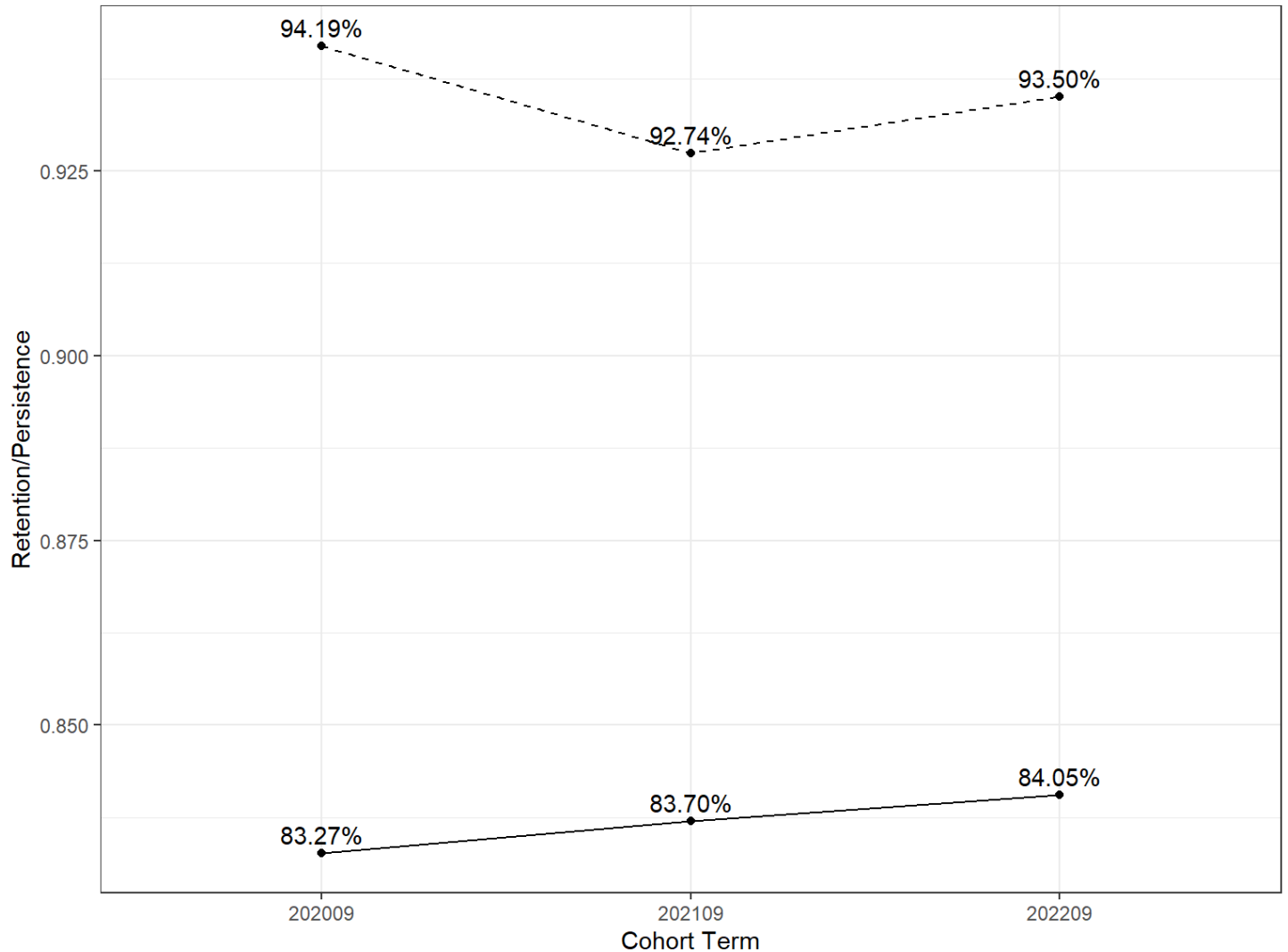
Question #2: For students who did not return (retain) from Fall term to Fall term, what variables associate with them?

- Profile of students who did not persist
- Profile of students who do not retain
- An intervention that Xavier University could employ to retain these students (we do want them to return the following Fall)

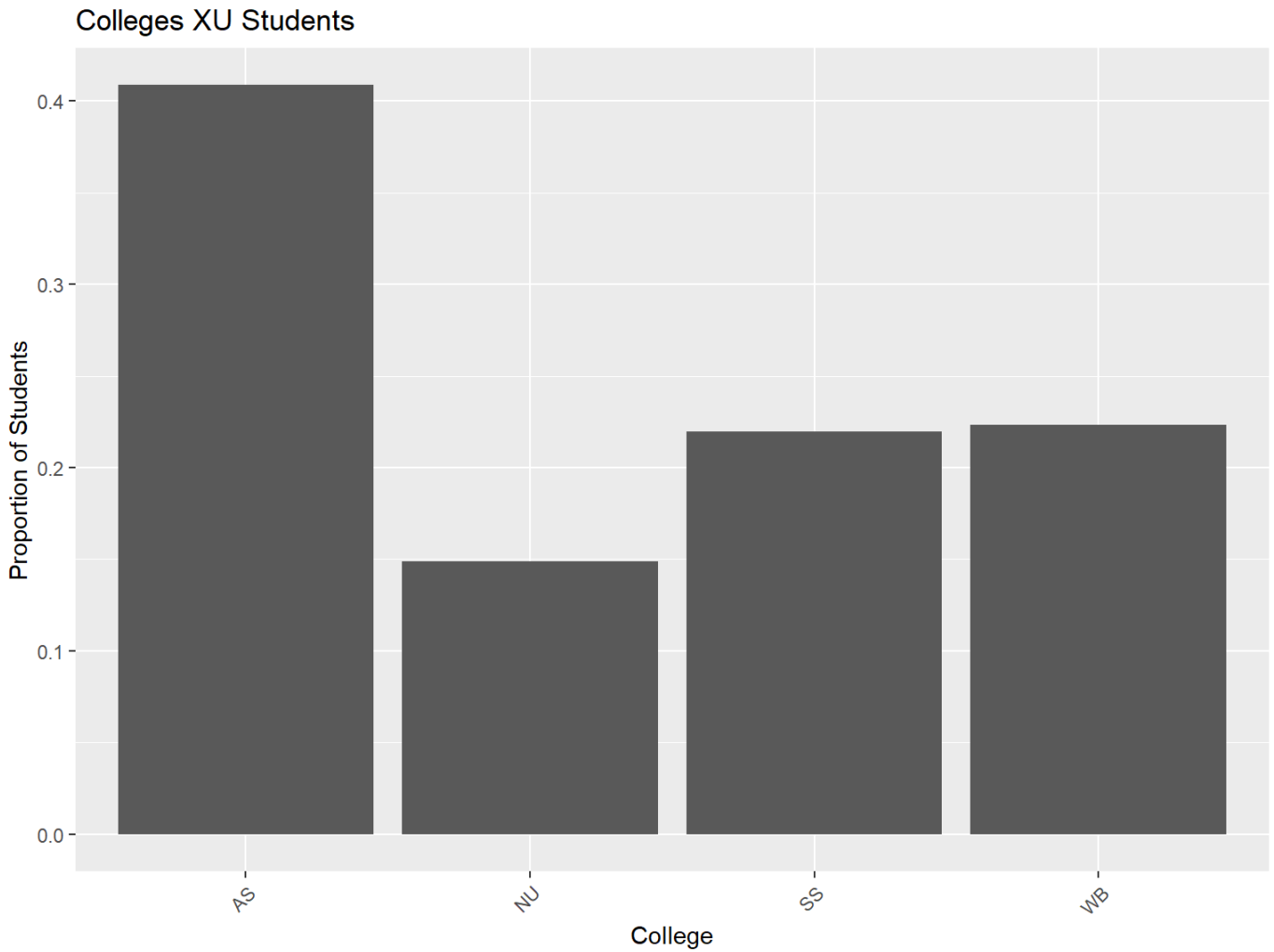
Context to remember…

- Students worked on this July 10 – August 4
- Financials
  - Students hired for 40 hours per week for 5 weeks, room and board paid for by the Data Science Fund (gift from Xavier alum Kevin Novak)
- Student background
  - First experience programming in R for all but one student
  - Very limited Statistics background

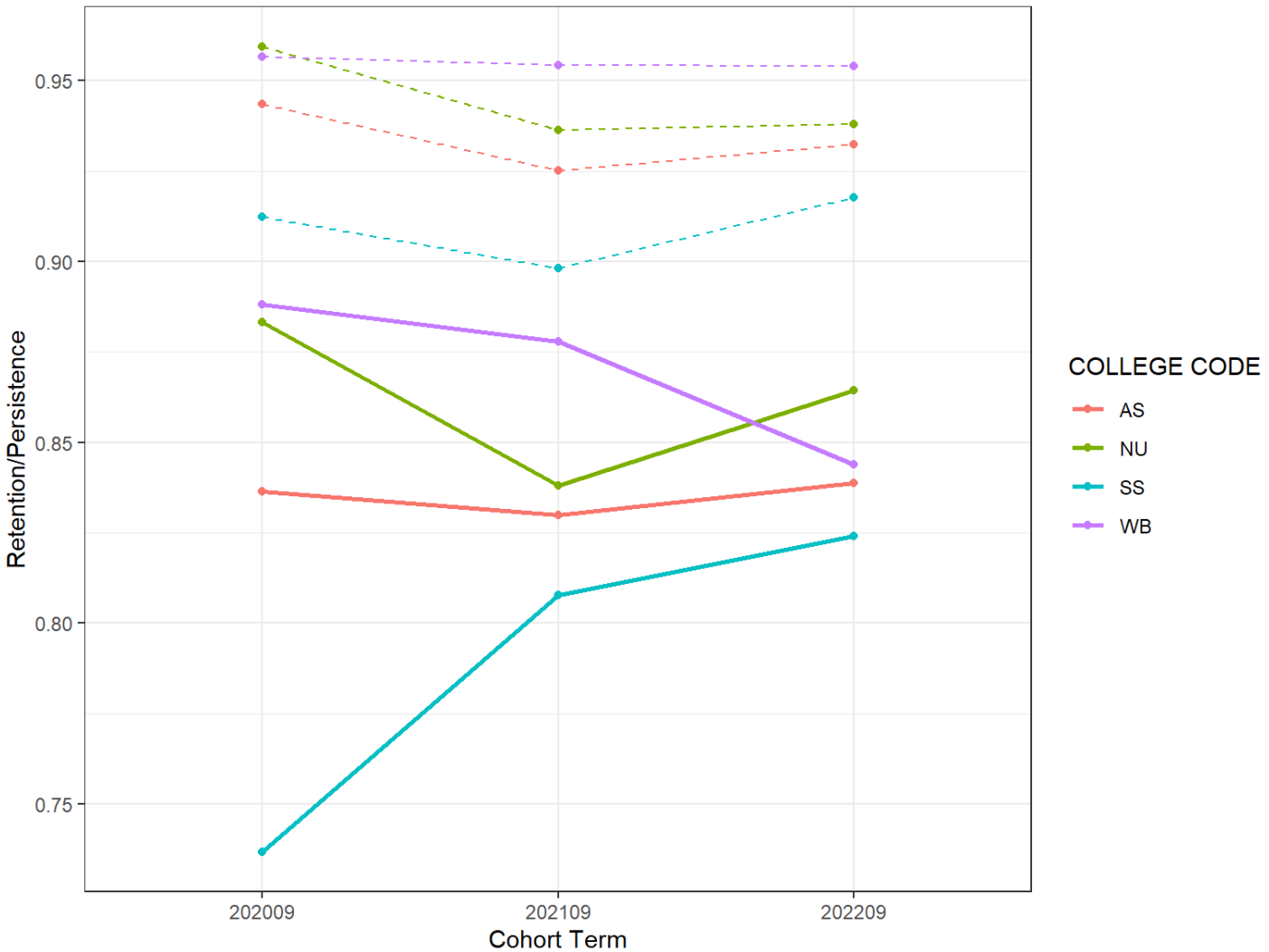
Overall Retention and Persistence Rates



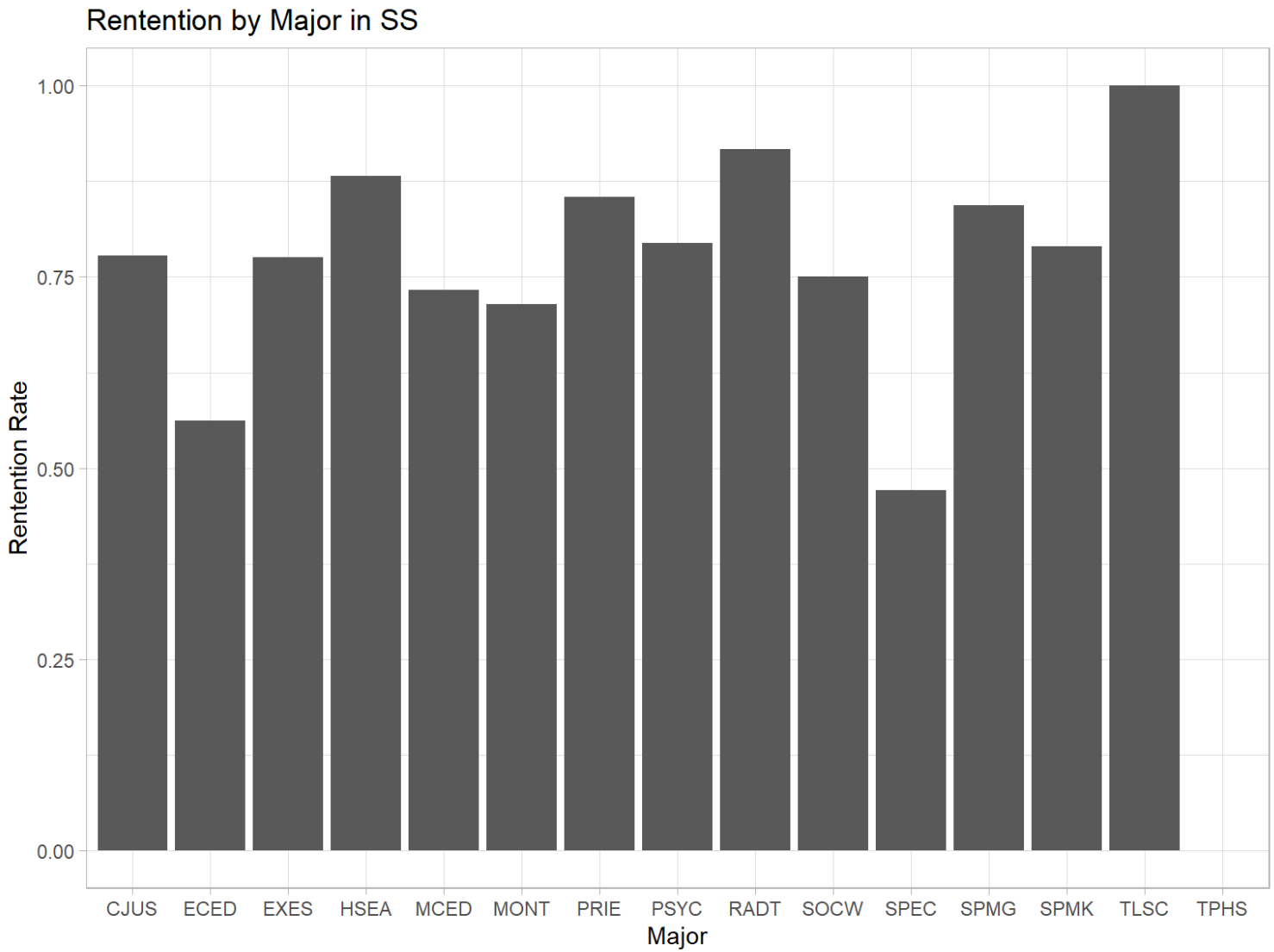
Distribution by College



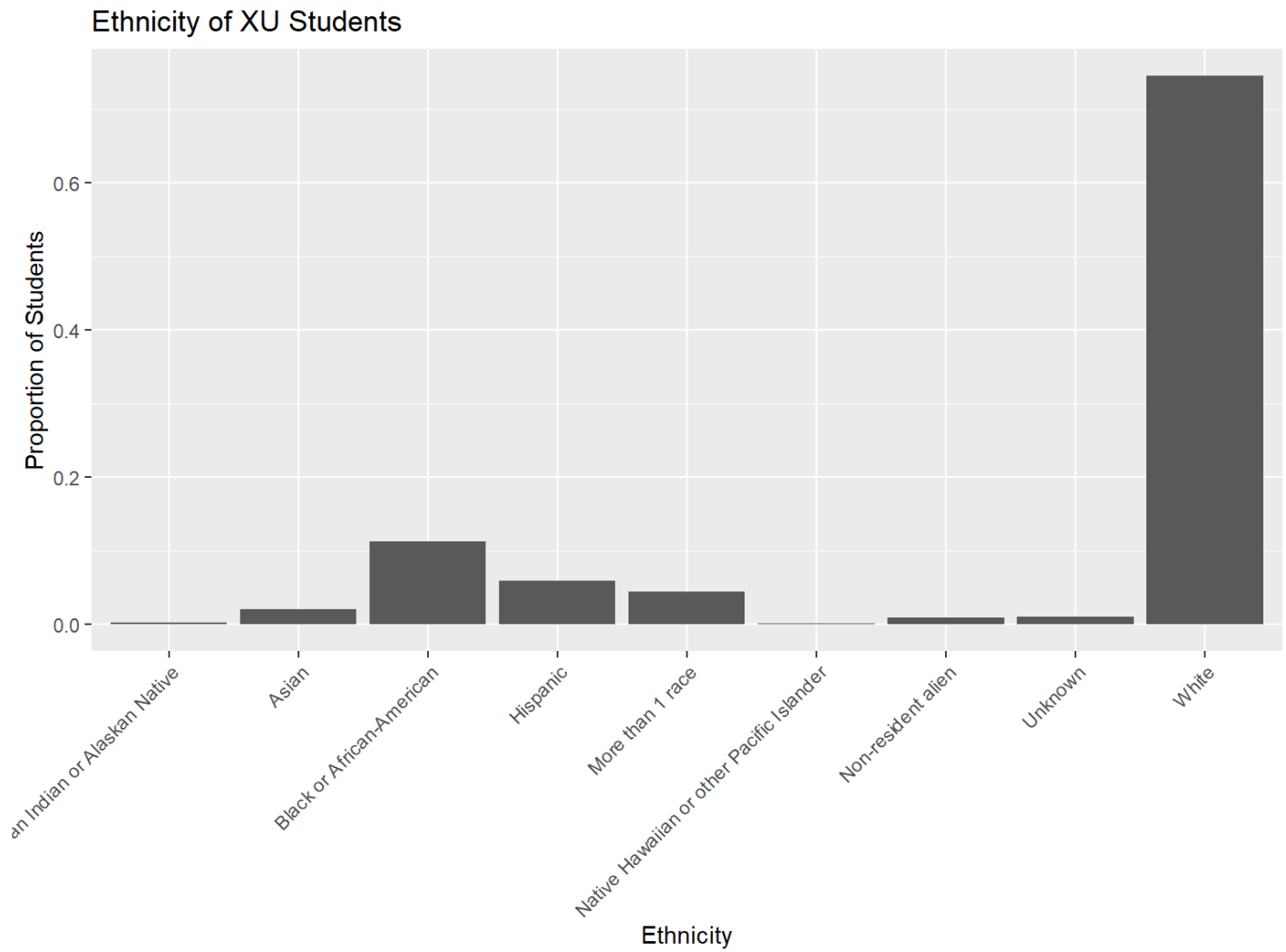
Retention and Persistence Rates by College



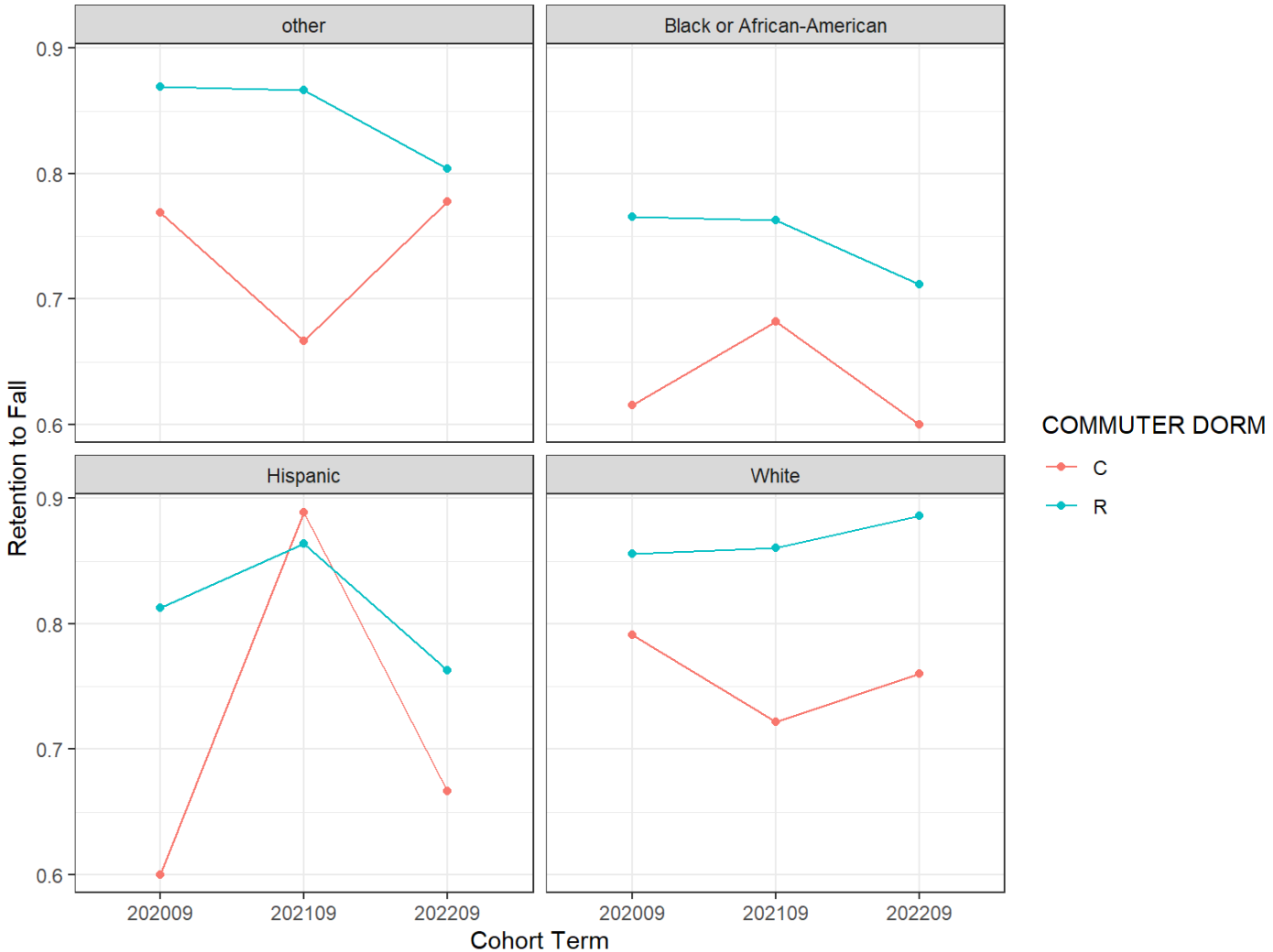
Retention by Major in Professional Sciences



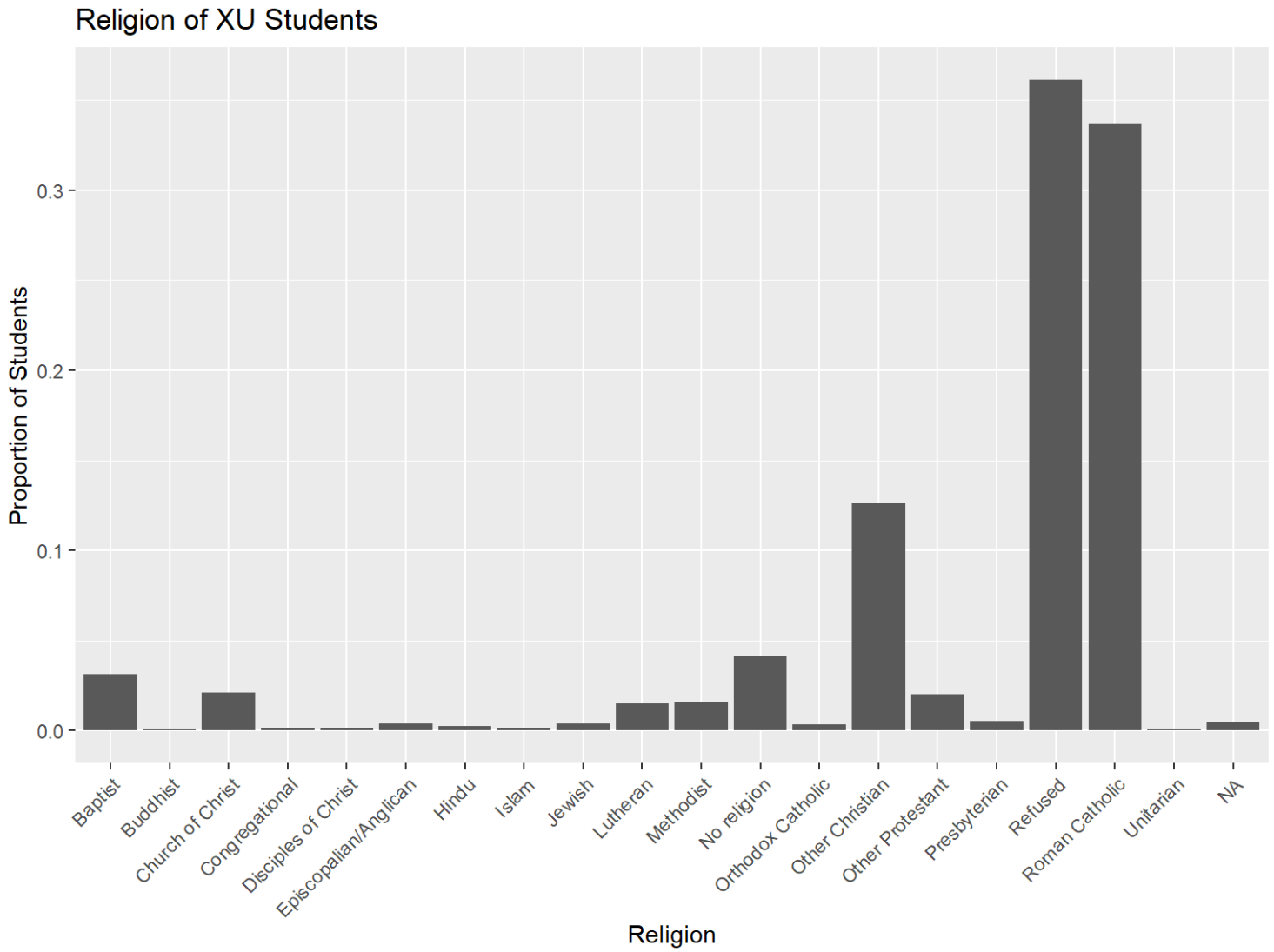
Distribution of Reported Ethnicities



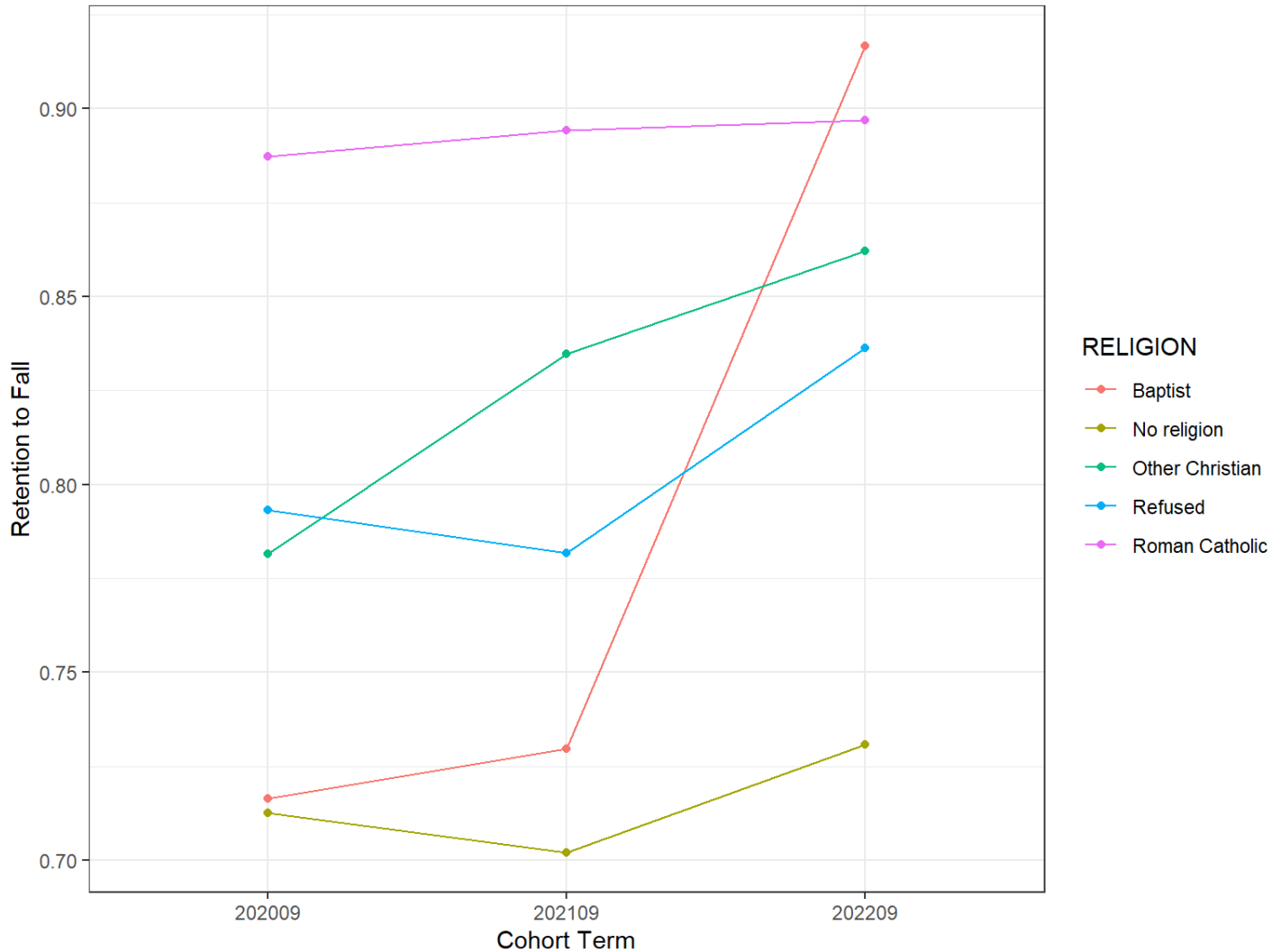
Retention by Race and Residential Status



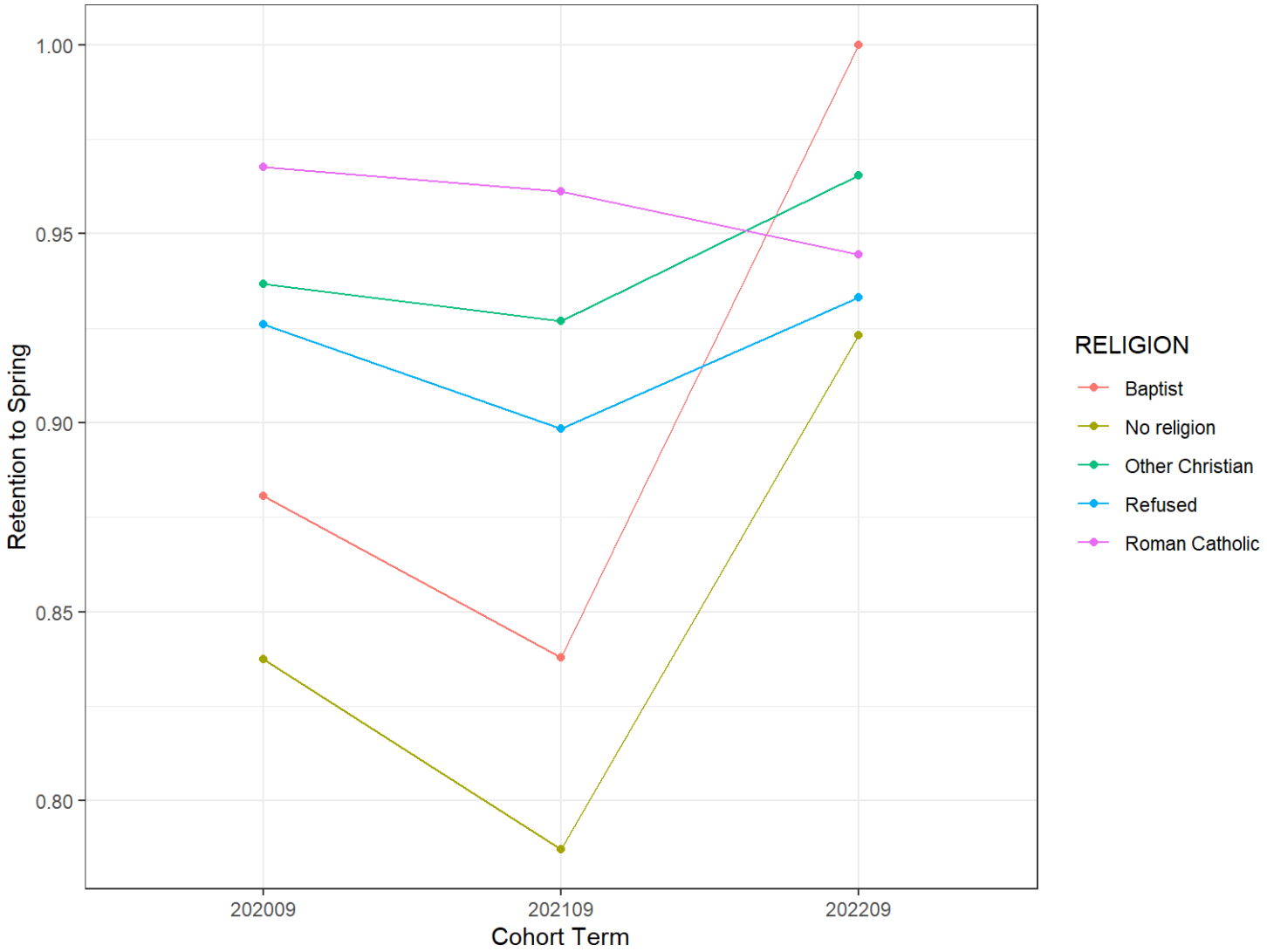
Distribution of Religions



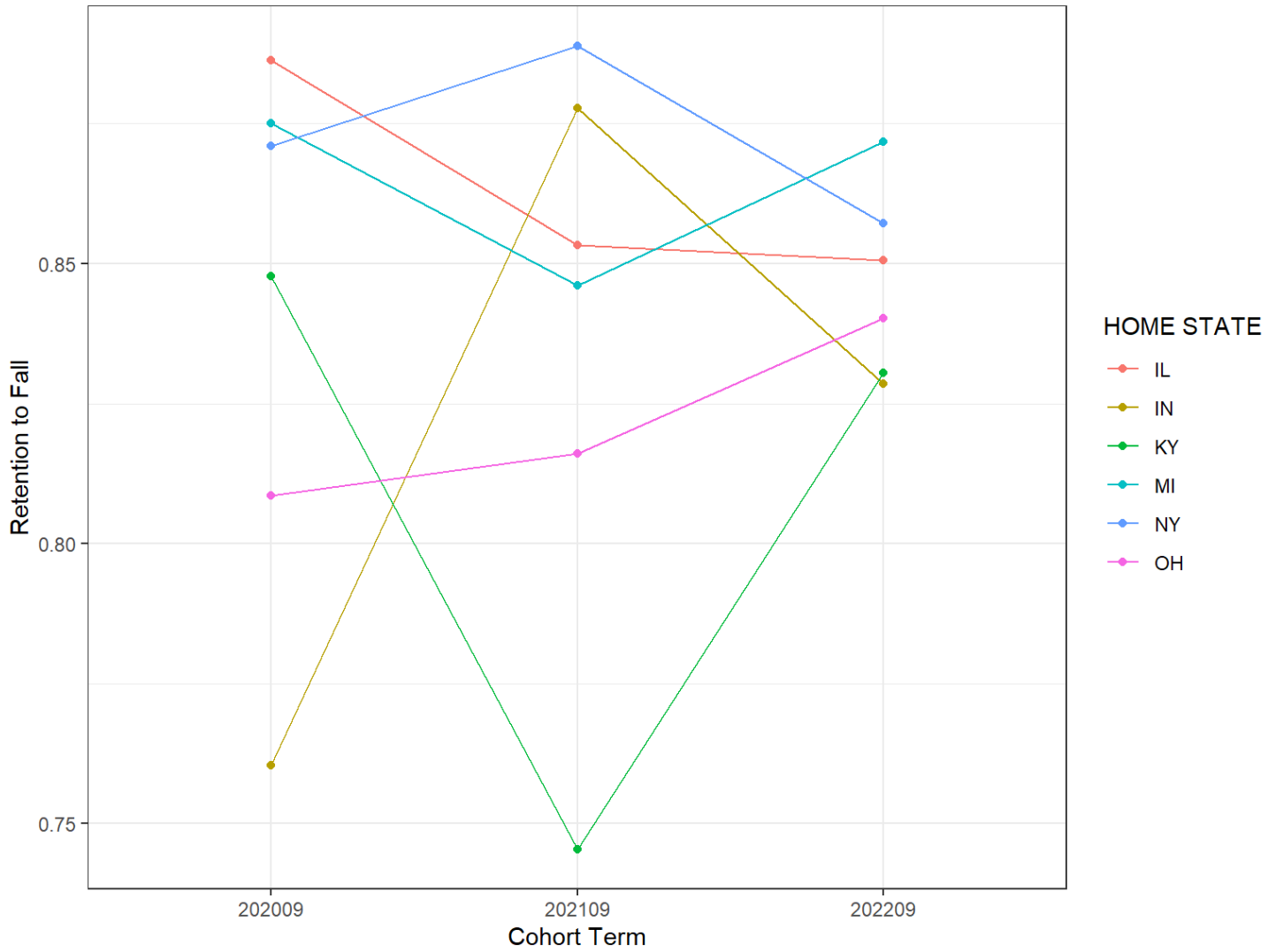
Retention Based on Religion



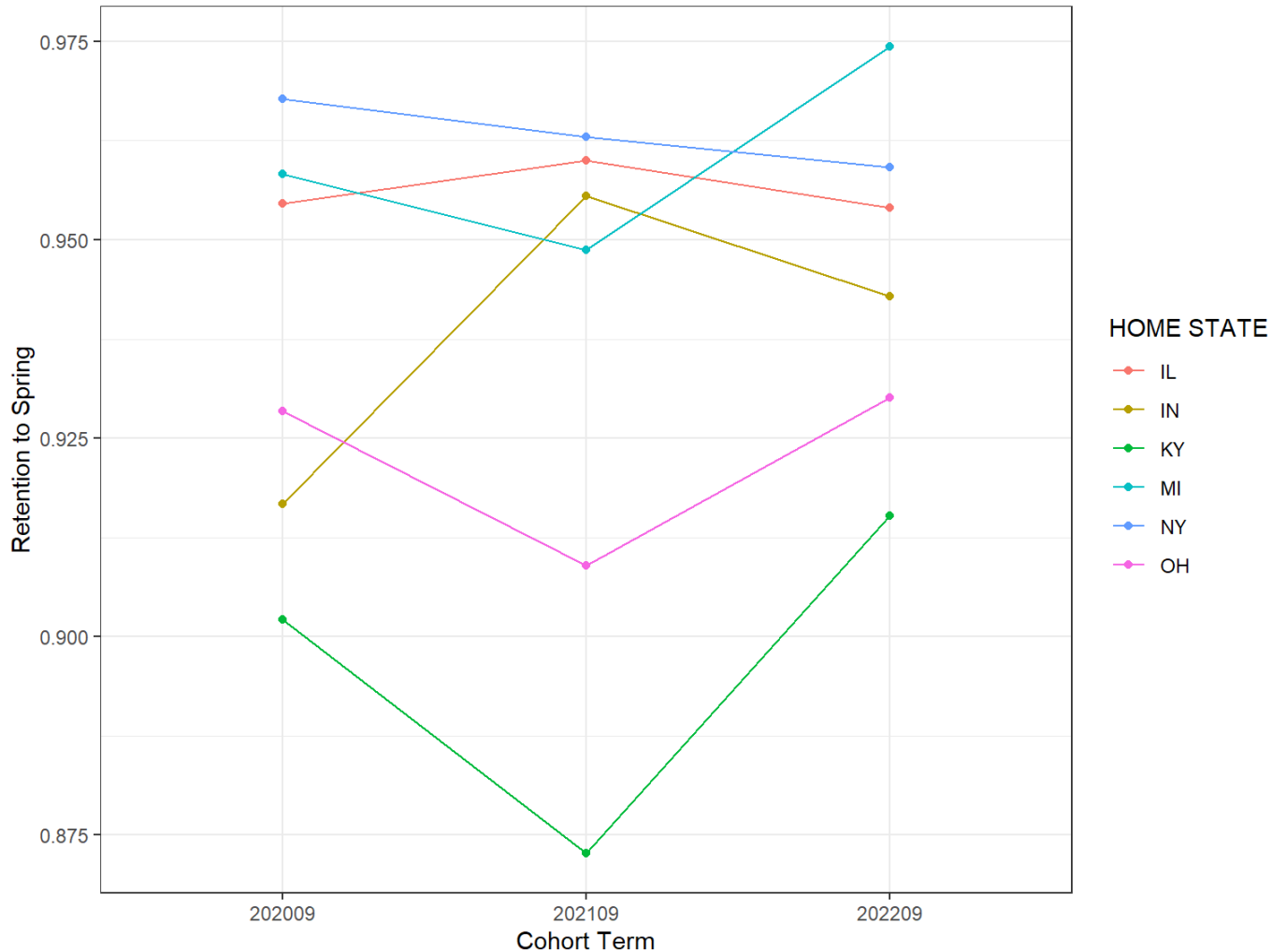
Persistence Based on Religion



Retention by Home State

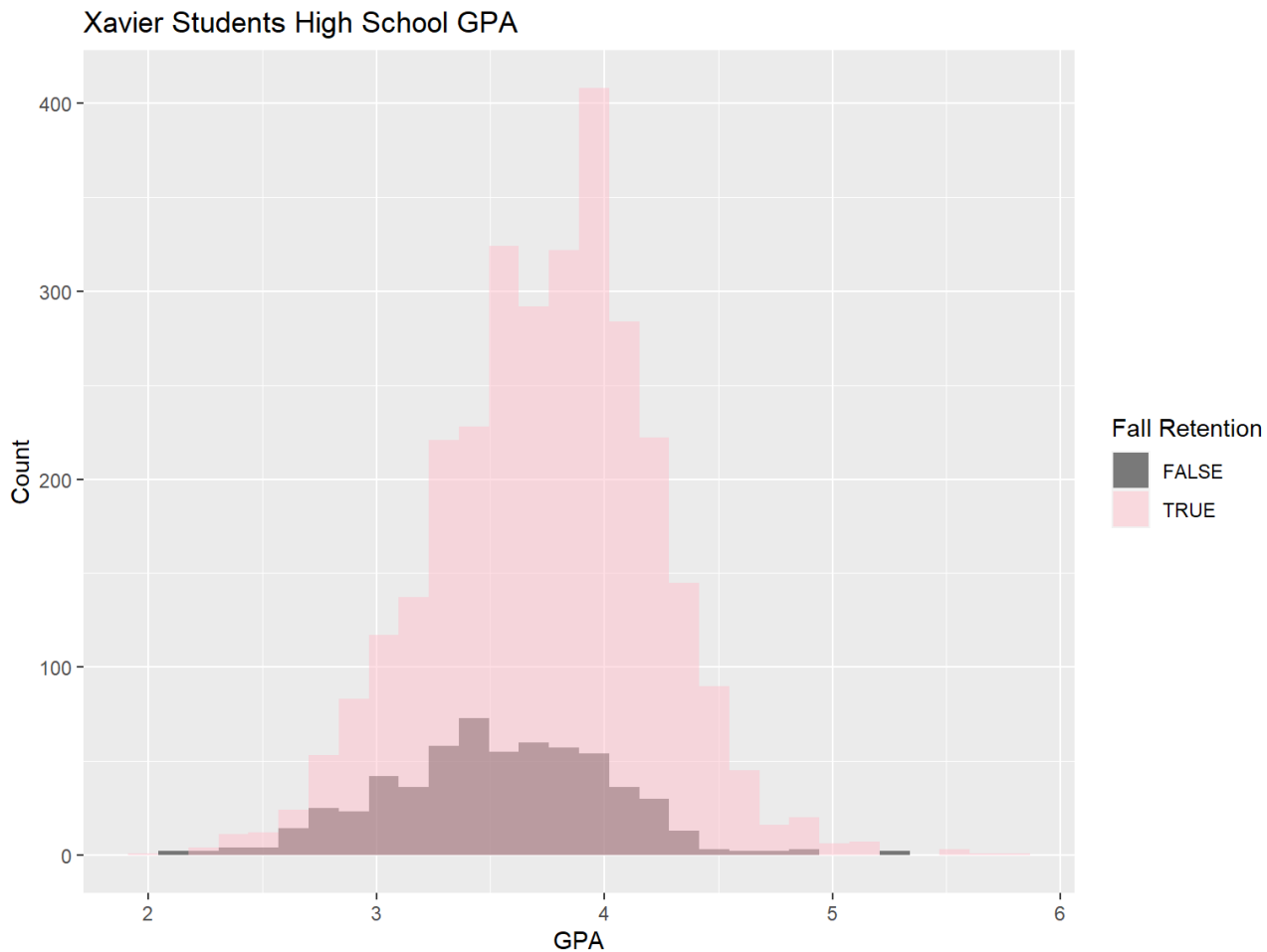


Persistence by Home State

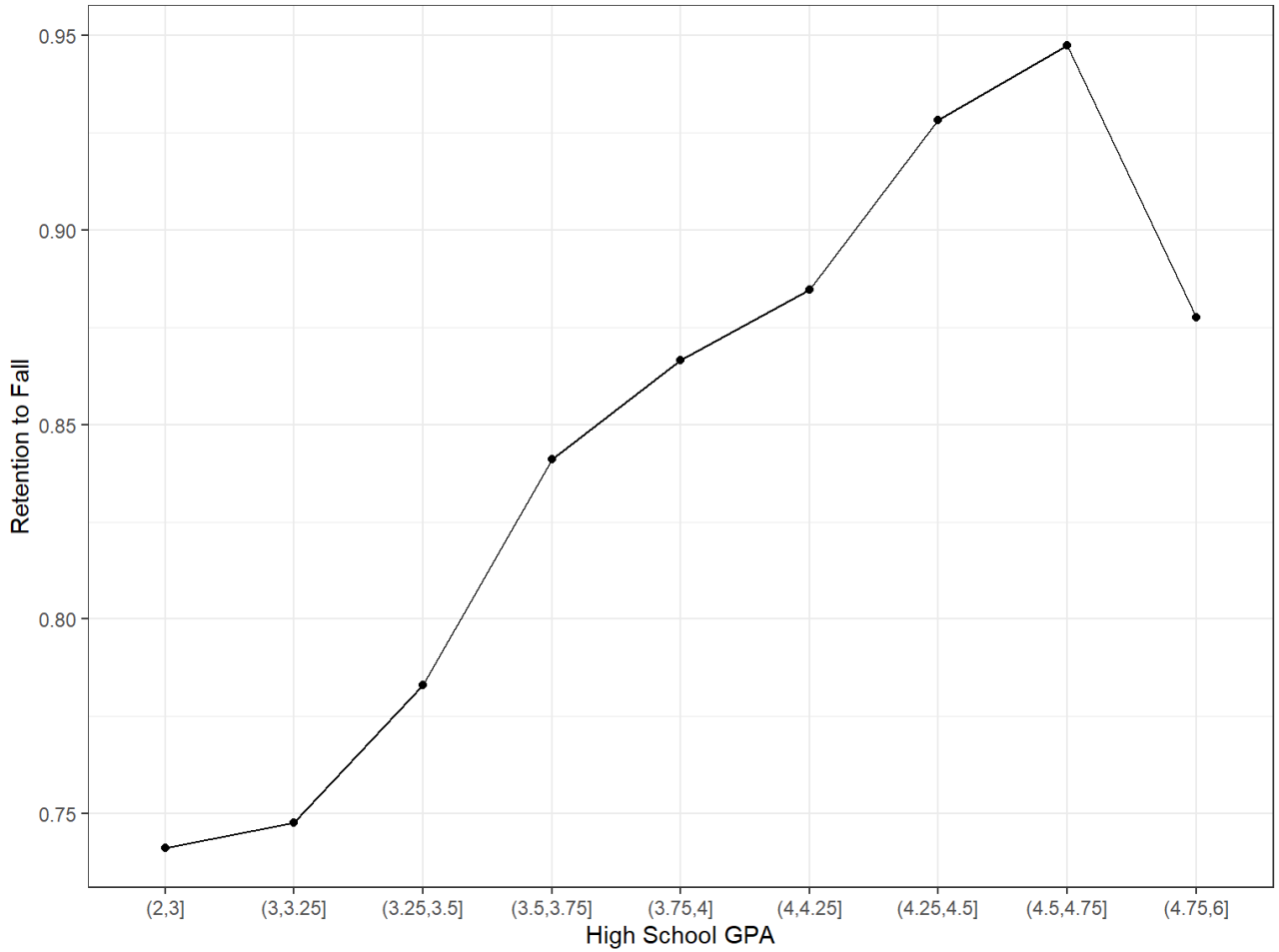


Retention by HS GPA

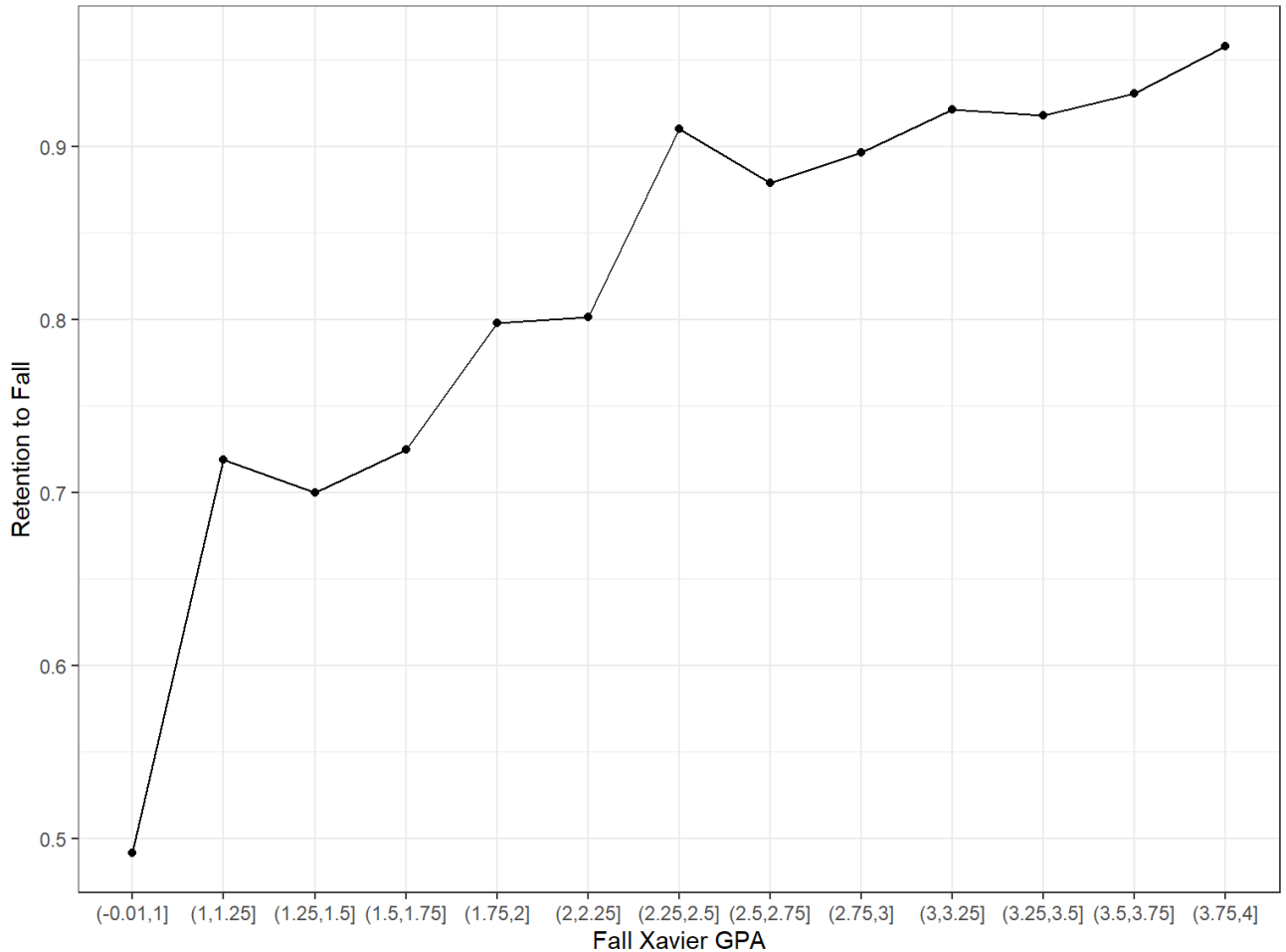
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



High School GPA

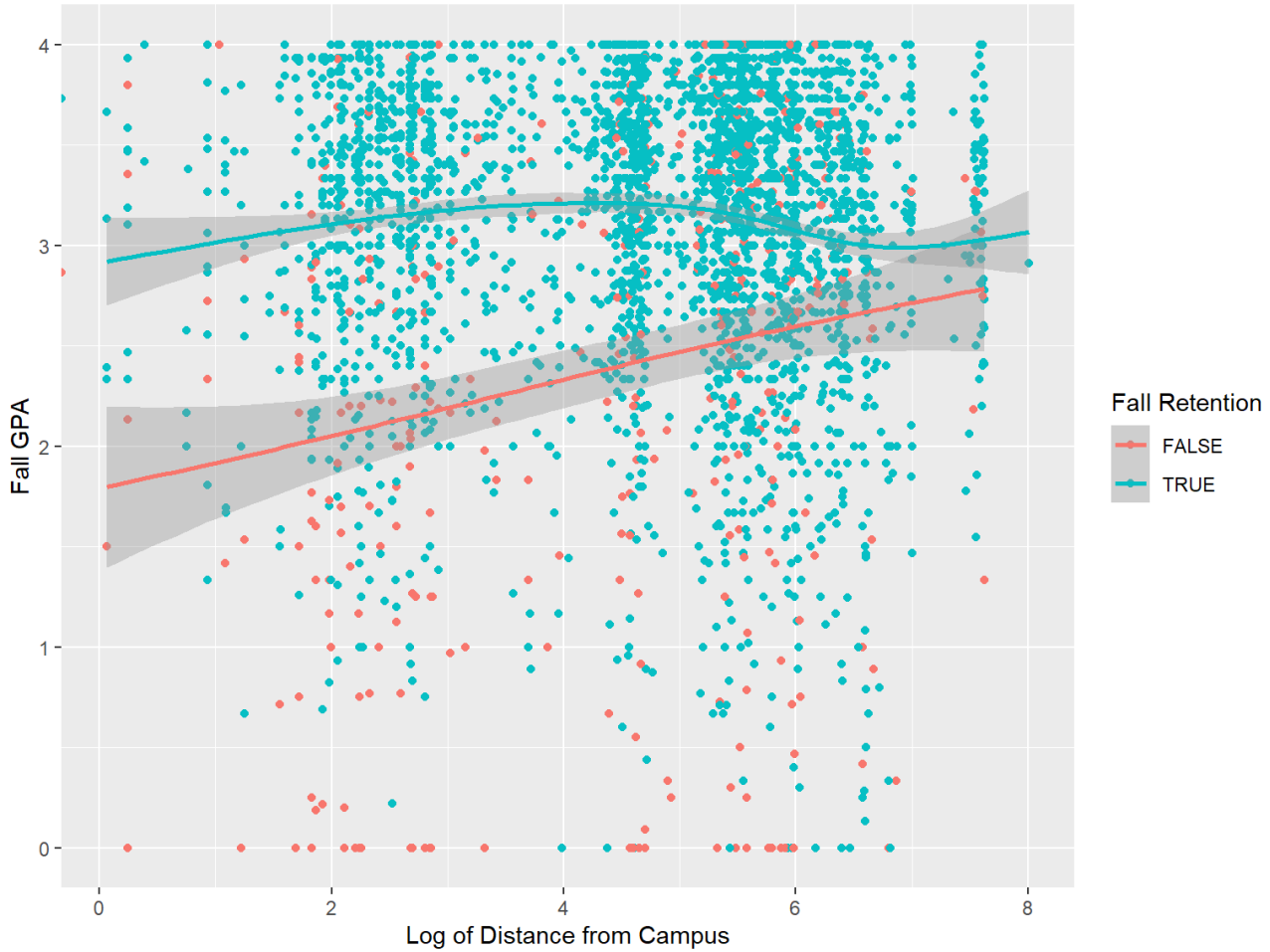


GPA


Retention to Fall
0.9
0.8
0.7
0.6
0.5
(-0.01,1]
(1,1.25]
(1.25,1.5]
(1.5,1.75]
(1.75,2]
(2,2.25]
(2.25,2.5]
(2.5,2.75]
(2.75,3]
(3,3.25]
(3.25,3.5]
(3.5,3.75]
(3.75,4]
Fall Xavier GPA

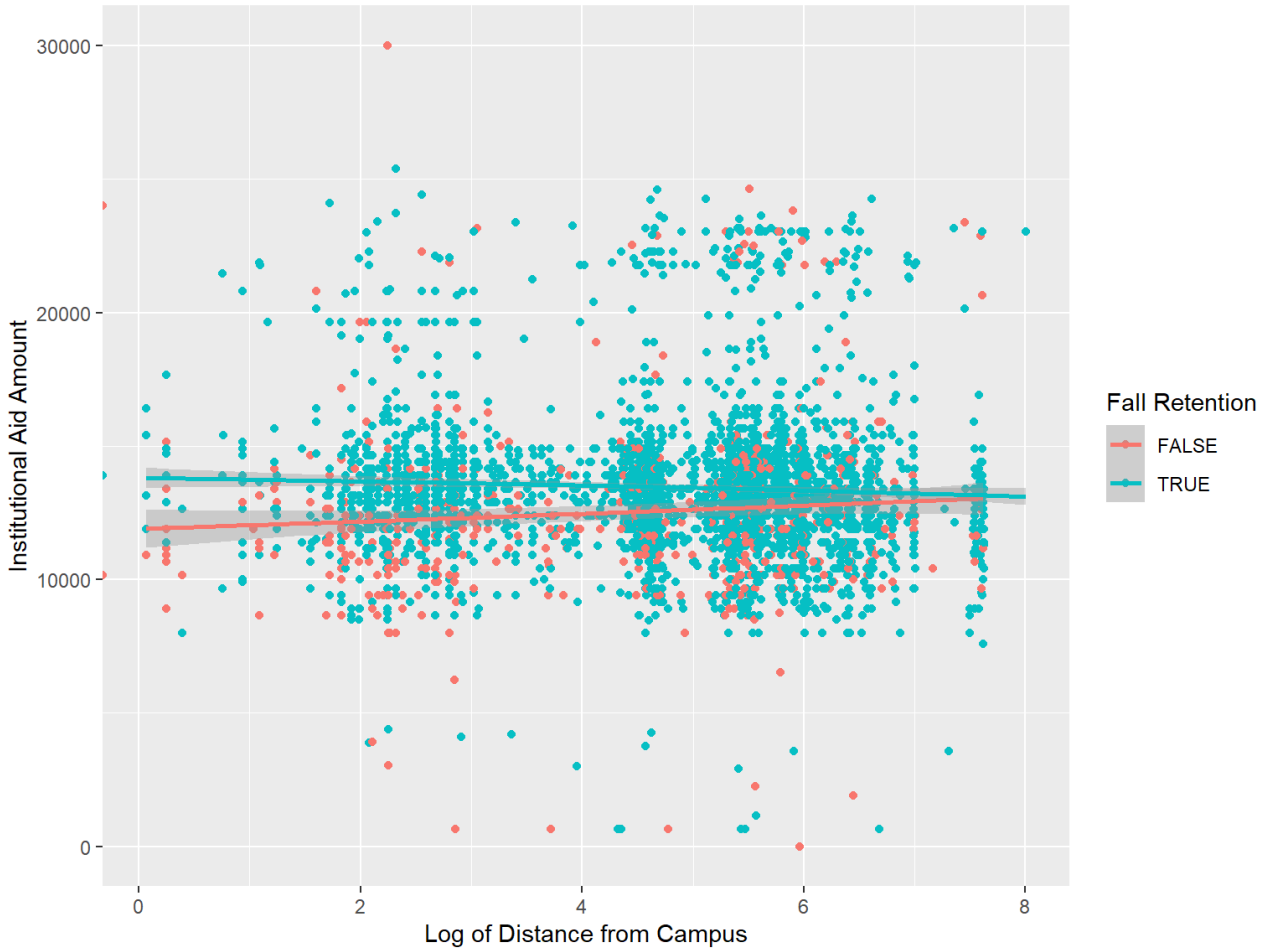
Retention by Home Distance from Xavier and Fall GPA

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

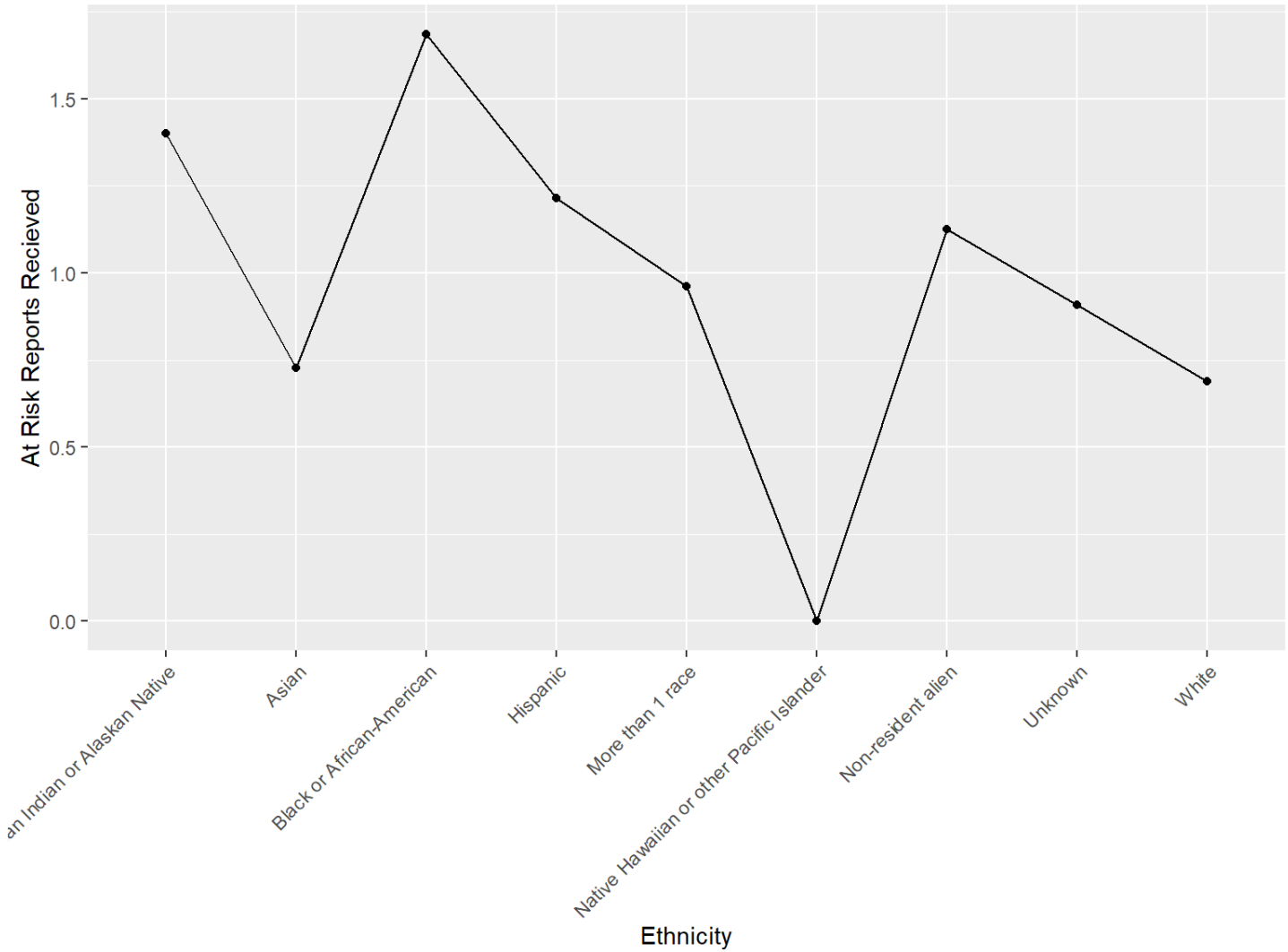


Retention by Home Distance from Xavier and Instant Aid Amt

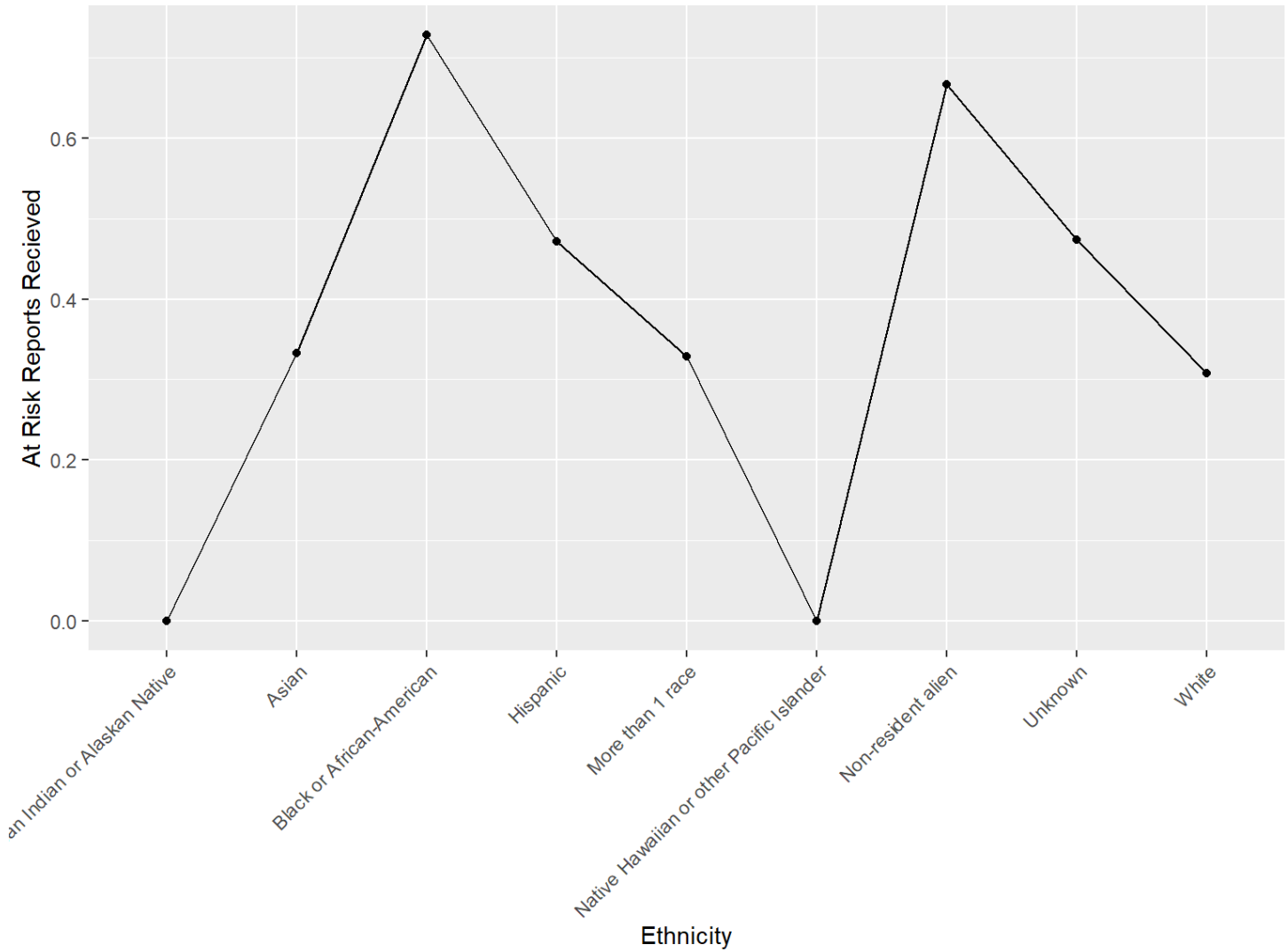
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs =
"cs")'
```



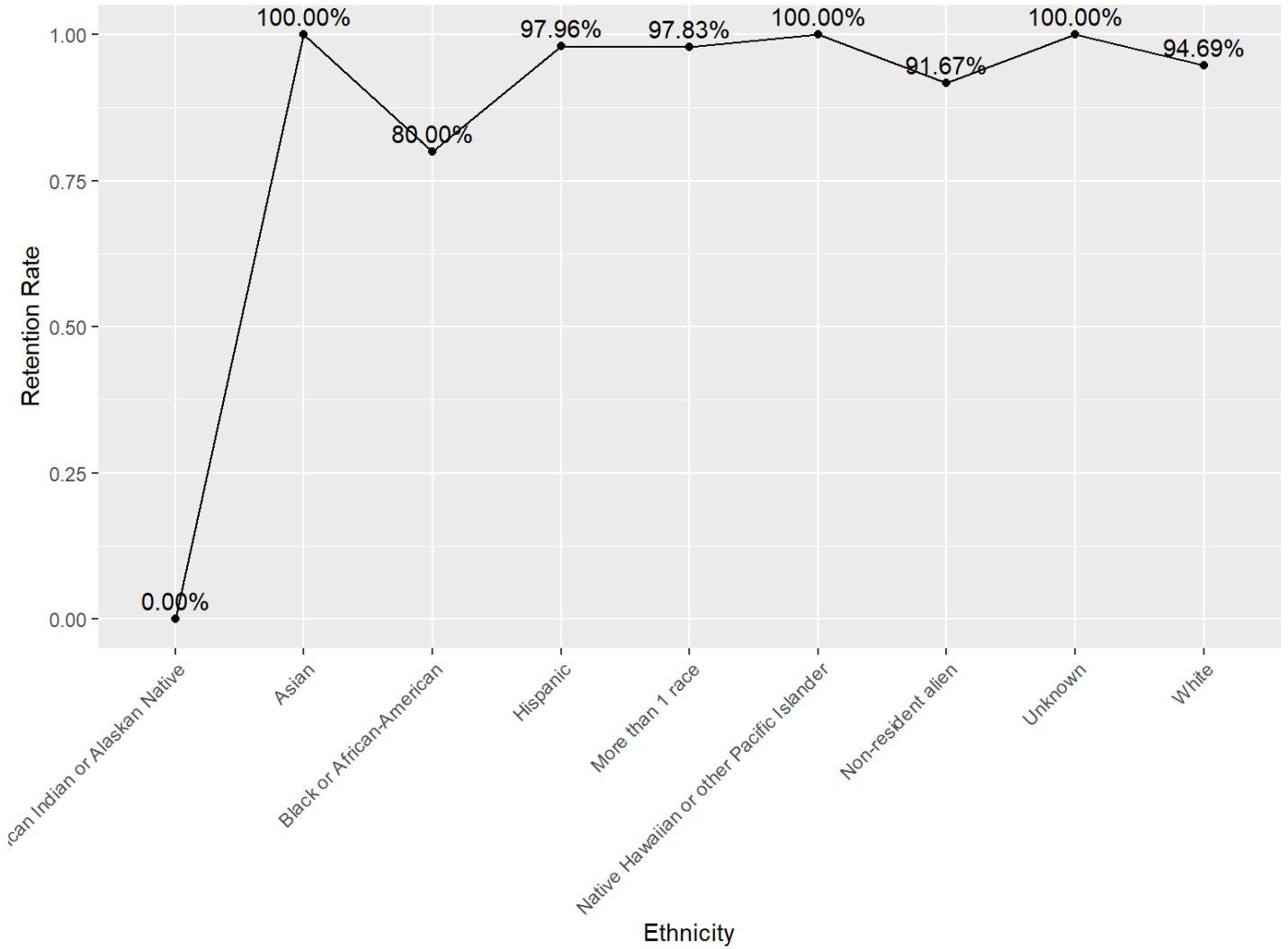
At Risk Reports by Race



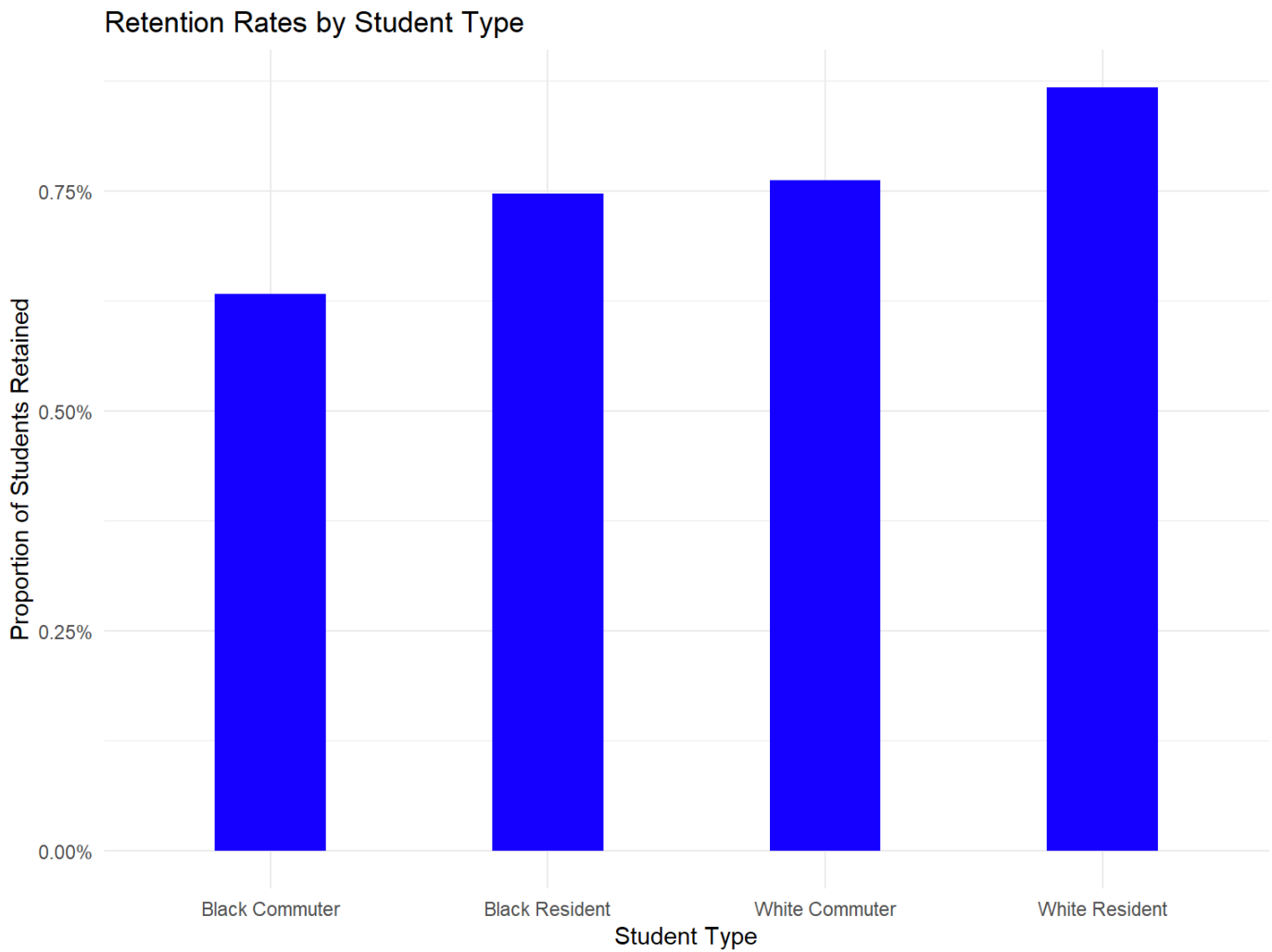
At Risk Reports by Race Above 2.2 GPA



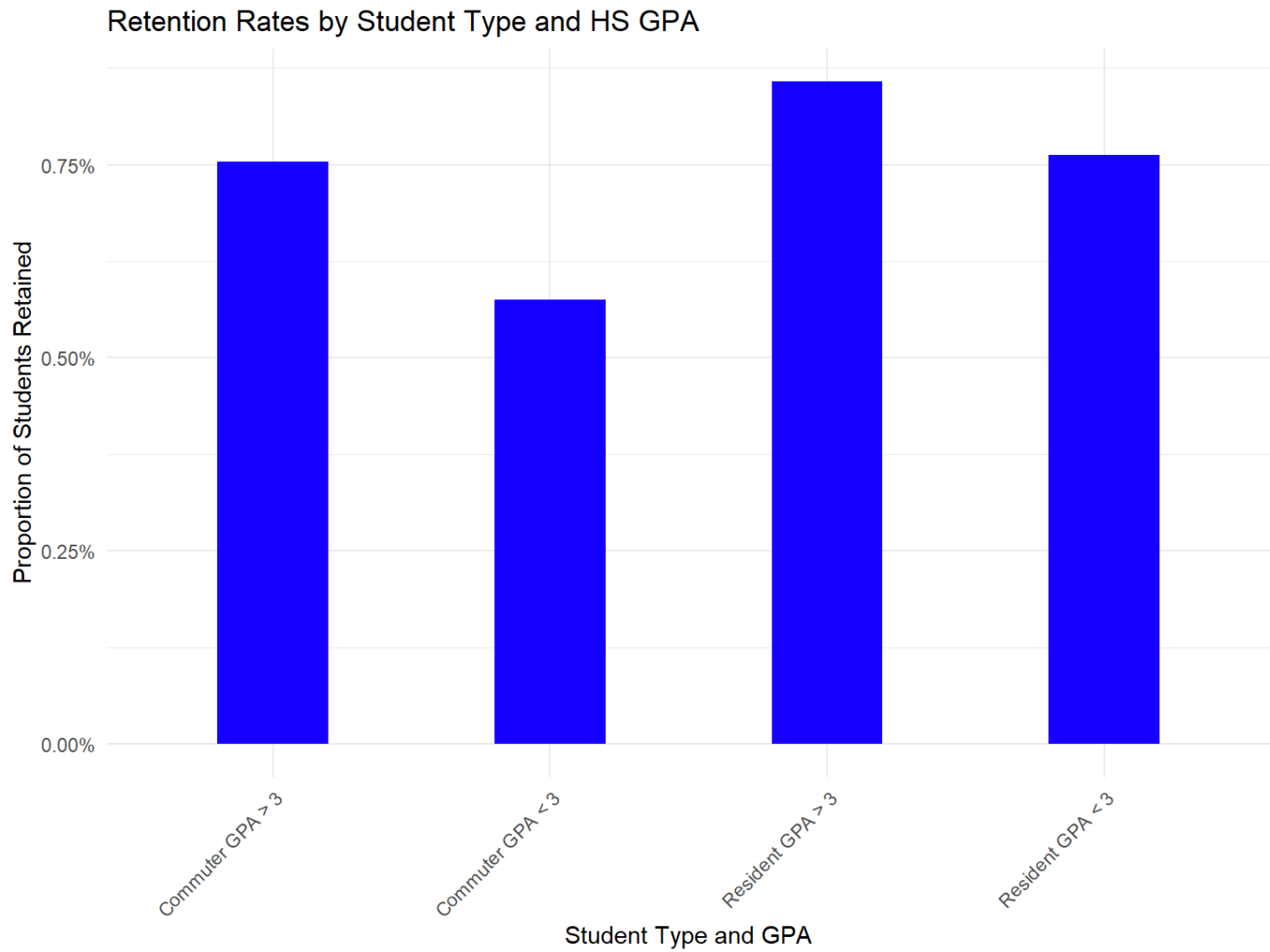
Retention Rates by Students Above a 3.5 GPA



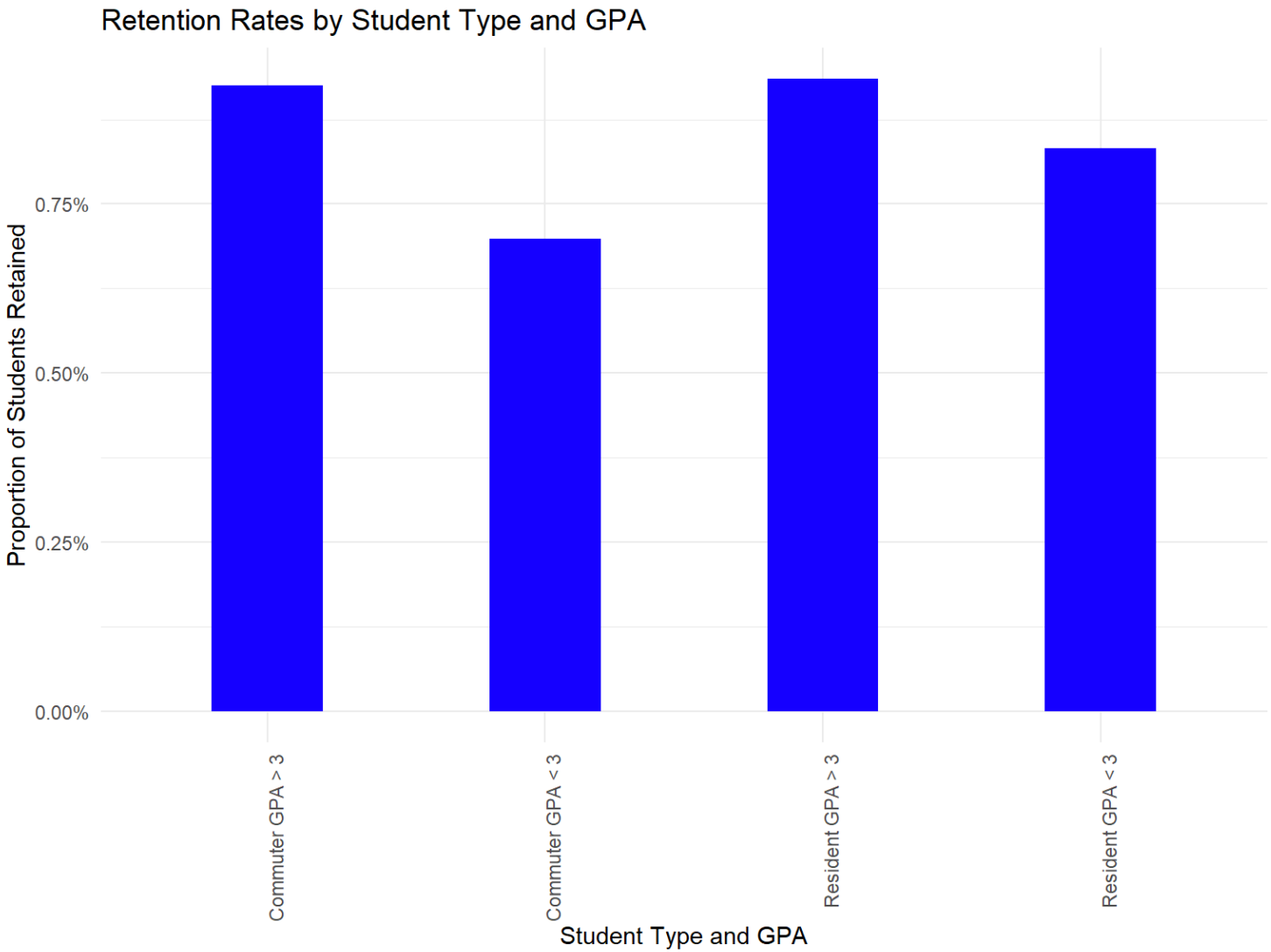
Black/White/Commuter/Resident Retention



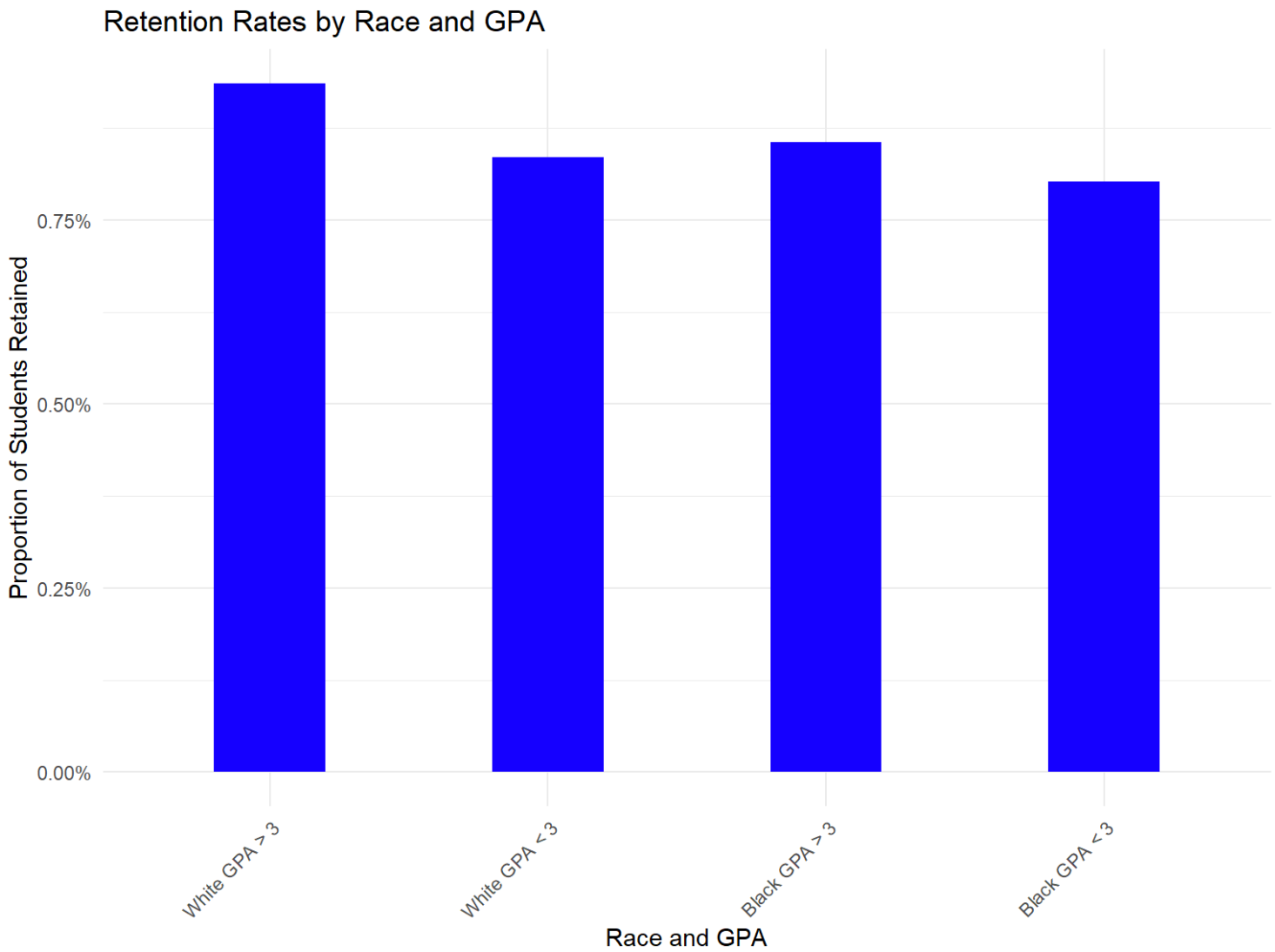
Commuter/Resident/HSGPA<3/HSGPA>3



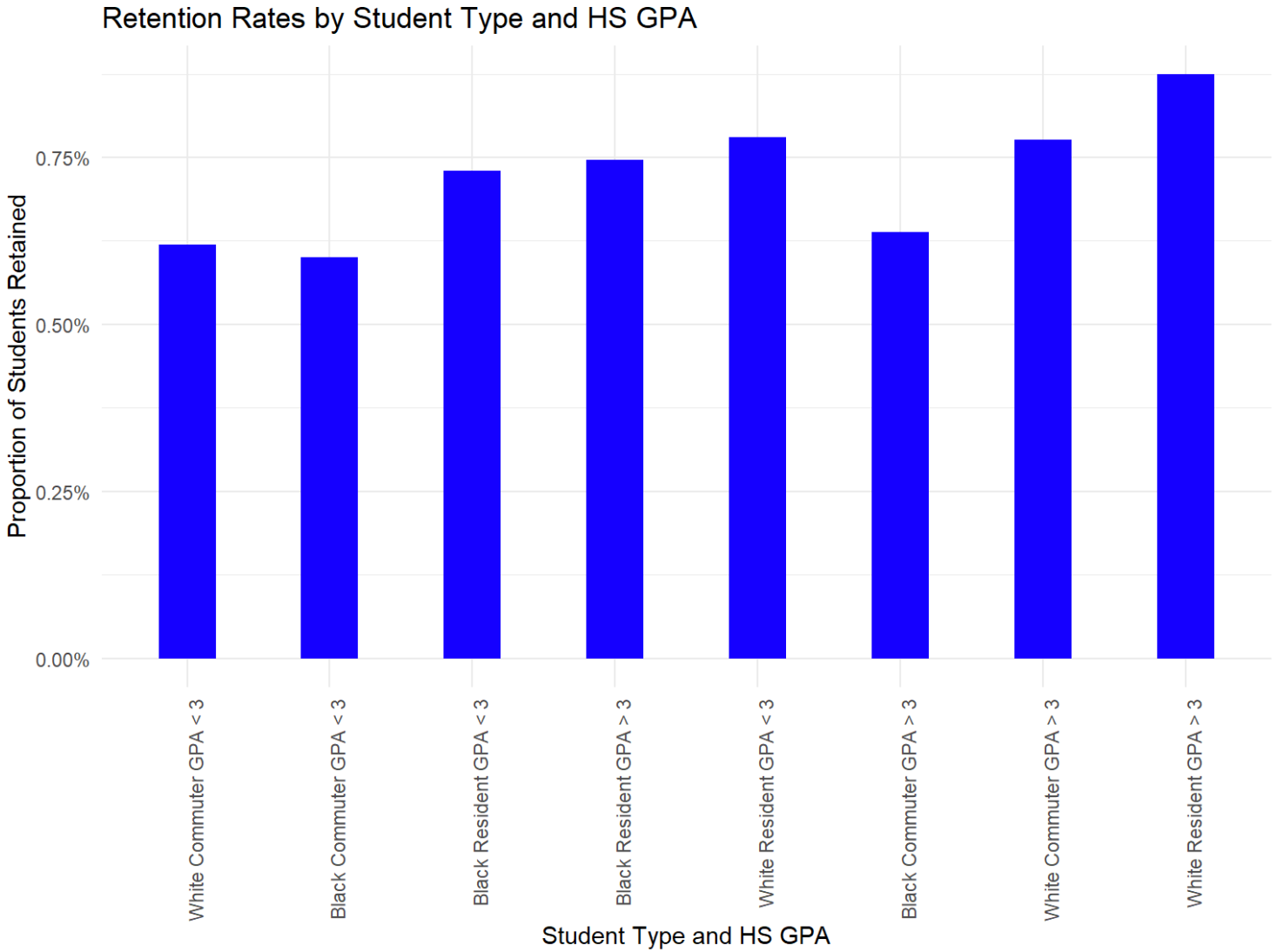
Com/Res/Fall GPA



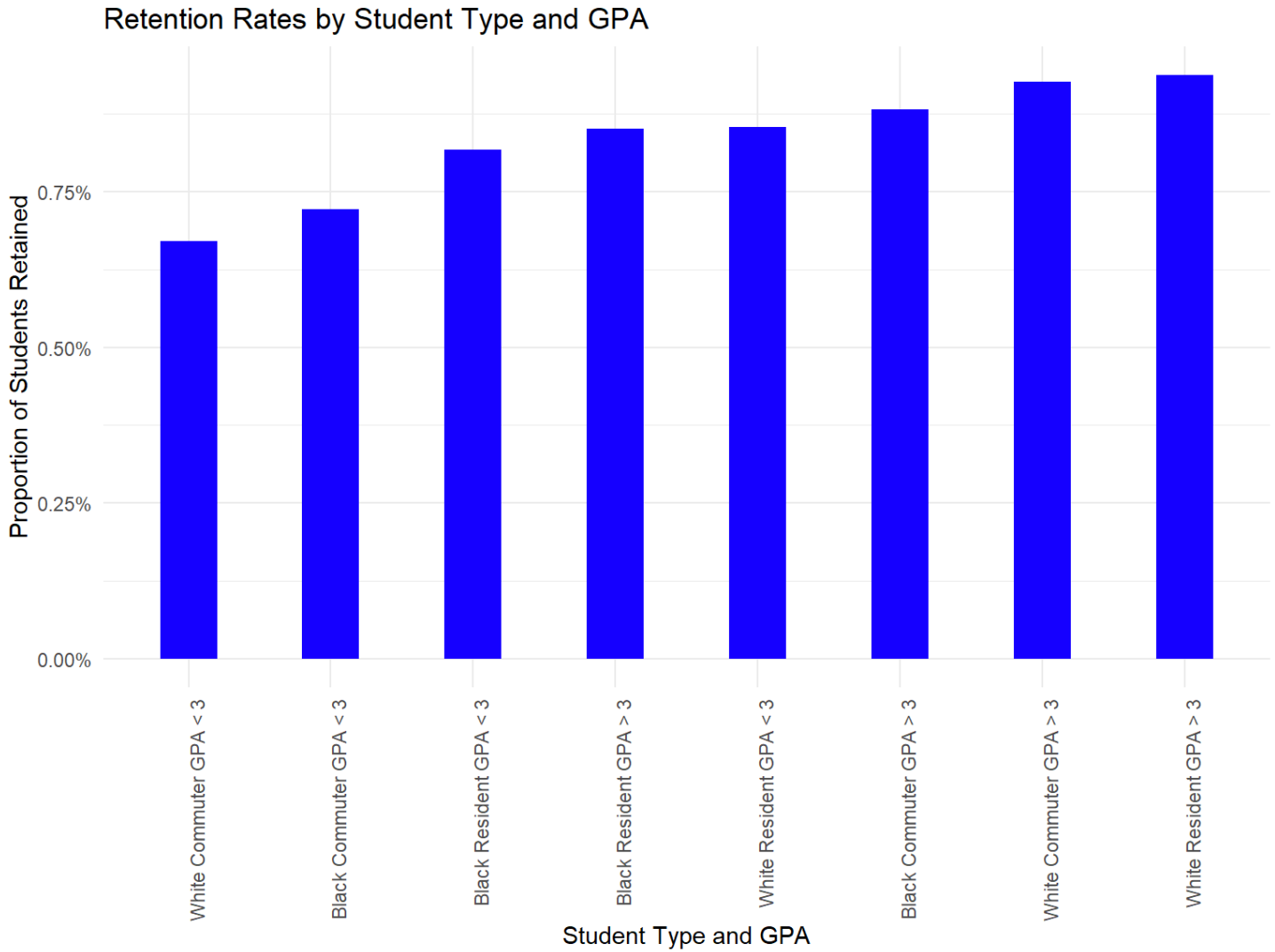
Black/White/Fall GPA>3/Fall GPA<3



Black/White/Commuter/Resident Retention and HS GPA



Black/White/Com/Res/Fall GPA>3/Fall GPA<3



Conclusions

- Black students at higher risk than non-black
- Commuter students at higher risk than resident students
- Black commuters stand out as most at-risk
- Non-religious students are retained at lower rates
- GPA both Fall and HS are indicative of retention

- Fall at-risk reports successful at identifying at-risk students