Your project title

Team Name: Team member 1, Team member 2, Team member 3, Team member 4

Invalid Date

```
Rows: 634
Columns: 15
$ subject_id            <dbl> 185, 158, 209, 102, 174, 184, 255, 265, 343, 137~
$ study                 <dbl> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, ~
$ cohort                <chr> "lac1", "lac1", "lac1", "lac1", "lac1", "lac1", ~
$ demo_race             <dbl> 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ demo_gender           <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, ~
$ demo_firstgen         <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, ~
$ bedtime_mssd          <dbl> 0.11672695, 0.14168084, 1.52928949, 0.13014845, ~
$ TotalSleepTime        <dbl> 432.2000, 391.9310, 344.3043, 392.6207, 423.4211~
$ midpoint_sleep        <dbl> 458.6600, 364.4655, 560.8913, 416.4828, 368.7632~
$ frac_nights_with_data <dbl> 0.8620690, 1.0000000, 0.7931034, 1.0000000, 0.65~
$ daytime_sleep         <dbl> 24.160000, 13.137931, 14.956522, 54.551724, 10.5~
$ cum_gpa               <dbl> 3.00, 3.66, 3.57, 3.61, 3.21, 3.20, 3.40, 3.86, ~
$ term_gpa              <dbl> 3.38, 2.60, 3.07, 3.56, 4.00, 3.36, 3.19, 3.28, ~
$ term_units            <dbl> 73, 64, 63, 61, 61, 60, 60, 60, 60, 59, 59, 58, ~
$ Zterm_units_ZofZ      <dbl> 4.0552949, 2.4825341, 2.3077829, 1.9582805, 1.95~
```

Your written report goes here!

> ❗ Important
>
> Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.

Exploratory Data Analysis

Description of the data set and key variables.

The data was originally collected in 2019, with the participants being first-year students at the following three universities: Carnegie Mellon University (CMU), a STEM-focused private

university, The University of Washington (UW), a large public university, and Notre Dame University (ND), a private Catholic university. To collect data on sleep, each participating student was given a Fitbit device to track their sleep and physical activity for a month in the spring term, and grade and demographic data was provided by university registrars.

There are 634 observations, representing the 634 participants in this study. Race is a binary variable separated into underrepresented students and non-underrepresented students with 0 being underrepresented and 1 being non-underrepresented. Students are considered underrepresented if either parent is Black, Hispanic or Latino, Native American, or Pacific, and students are deemed non-underrepresented if both parents have White or Asian ancestry. The gender of the subject is also binary with 0 being male and 1 being female. First-generation status is binary with 0 being non-first gen and 1 being first-gen. The mean successive squared difference of bedtime measures the bedtime variability, specifically the average of the squared difference of bedtime on consecutive nights.

Clean Data

```
Rows: 634
Columns: 16
$ subject_id          <dbl> 185, 158, 209, 102, 174, 184, 255, 265, 343, 137~
$ study               <dbl> 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, ~
$ cohort              <chr> "lac1", "lac1", "lac1", "lac1", "lac1", "lac1", ~
$ demo_race           <dbl> 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
$ demo_gender         <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, ~
$ demo_firstgen       <dbl> 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, ~
$ bedtime_mssd        <dbl> 0.11672695, 0.14168084, 1.52928949, 0.13014845, ~
$ TotalSleepTime      <dbl> 432.2000, 391.9310, 344.3043, 392.6207, 423.4211~
$ midpoint_sleep      <dbl> 458.6600, 364.4655, 560.8913, 416.4828, 368.7632~
$ frac_nights_with_data <dbl> 0.8620690, 1.0000000, 0.7931034, 1.0000000, 0.65~
$ daytime_sleep       <dbl> 24.160000, 13.137931, 14.956522, 54.551724, 10.5~
$ cum_gpa             <dbl> 3.00, 3.66, 3.57, 3.61, 3.21, 3.20, 3.40, 3.86, ~
$ term_gpa            <dbl> 3.38, 2.60, 3.07, 3.56, 4.00, 3.36, 3.19, 3.28, ~
$ term_units          <dbl> 73, 64, 63, 61, 61, 60, 60, 60, 60, 59, 59, 58, ~
$ Zterm_units_ZofZ    <dbl> 4.0552949, 2.4825341, 2.3077829, 1.9582805, 1.95~
$ university          <chr> "stem_priv", "stem_priv", "stem_priv", "stem_pri~
```
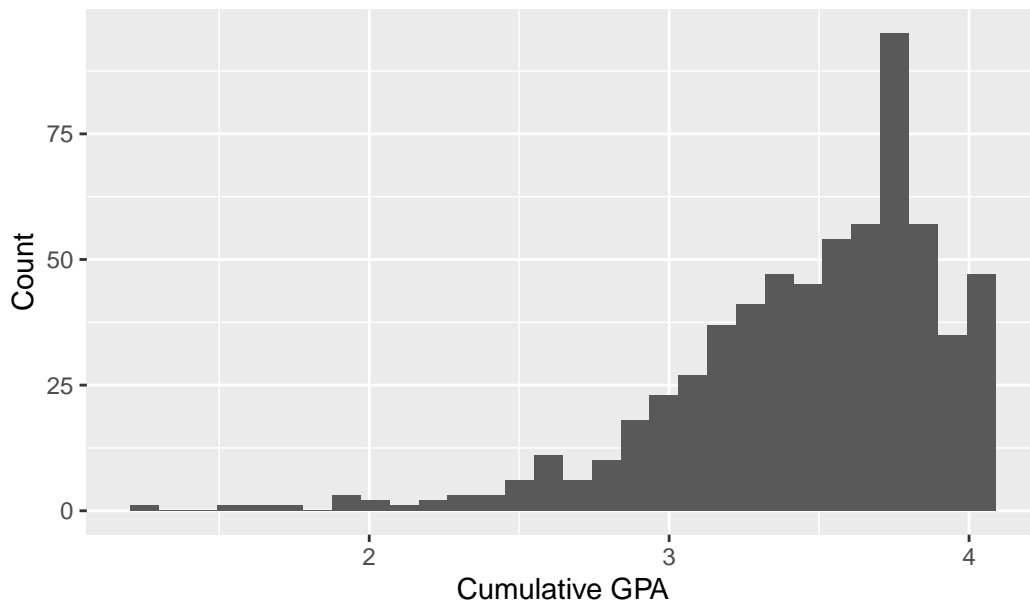
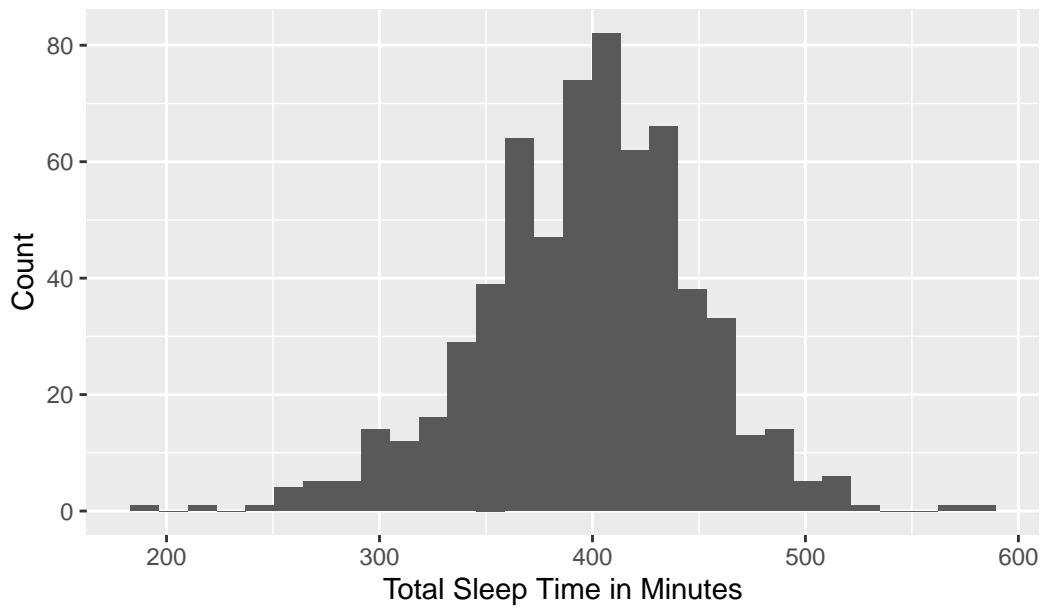Univariate EDA of The Response & Key Predictor Variables
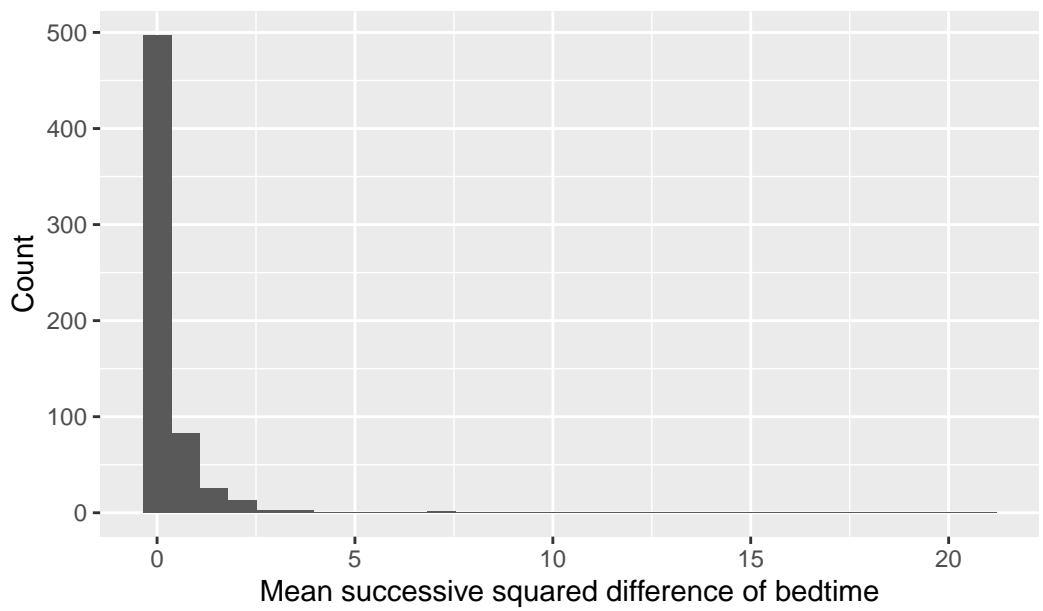
Distribution of the Term GPA
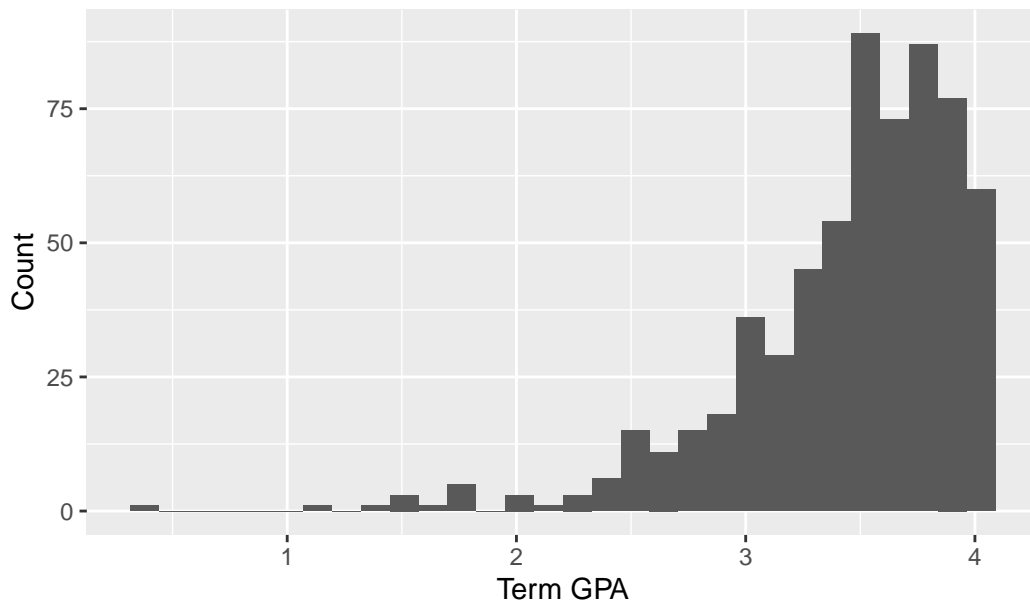
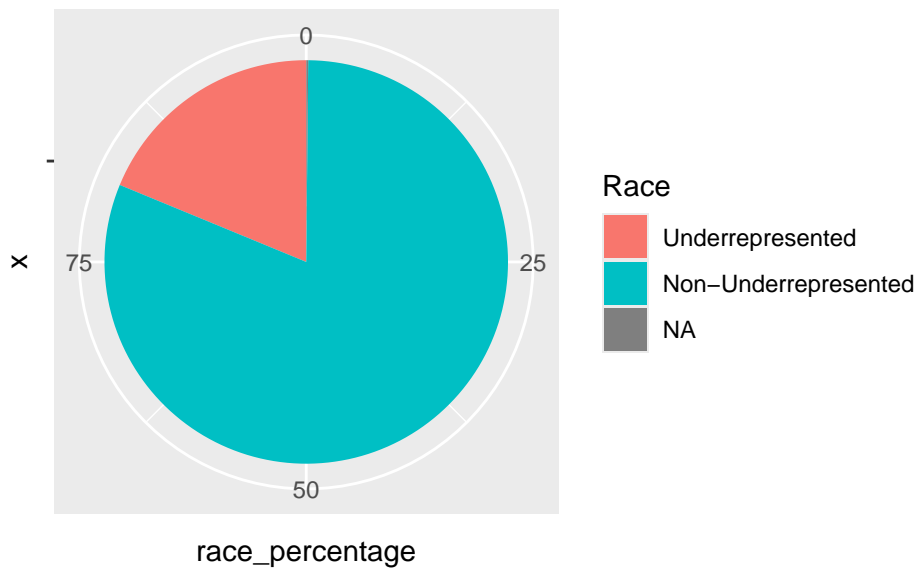## Distribution of the Cumulative GPA

Distribution of the Total Sleep Time
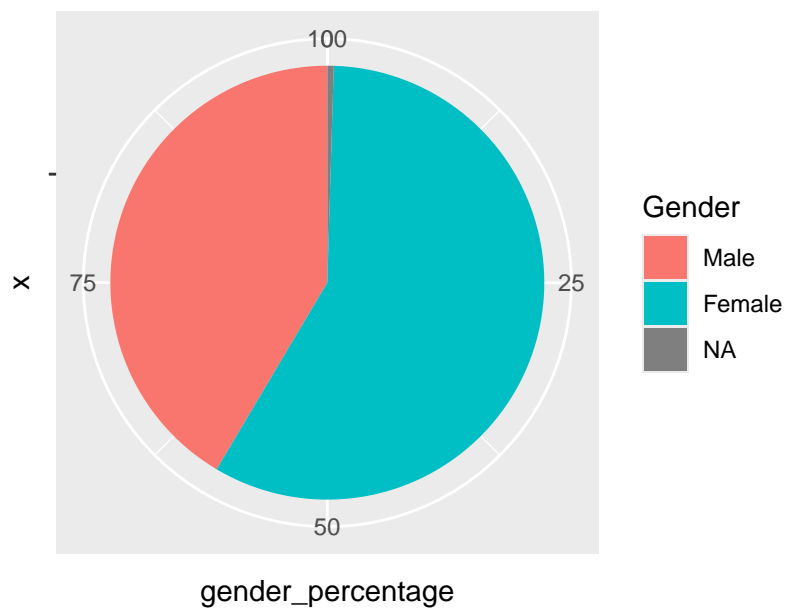

Distribution of the Bedtime Variability

Distribution of Underrepresented Vs. Non−Underrepresented Students

race_percentage

Distribution of Gender

gender_percentage

Bivariate EDA of The Response & Key Predictor Variables

Distribution of Cumulative GPA by University