# Activity One Biometry

EMH

2025-09-27

## Data Checking

I decided to look at the dataset row by row using filtering functions in R. I compared each column against the acceptable ranges and values defined in the data dictionary. By doing that, I found the following errors:

- **Education**: One entry was coded as `"N/A"` instead of `"> high school"` or "`<= high school"`.
- **Race (Other)**: One subject had a value of `"p"` in the `race_other` column, which is not an acceptable value.
- **Age**: One subject had an age of `555`, which is far outside the valid adult range of `18-100`
- **Sex**: One subject had a value of `2` for sex, which is invalid according to the dictionary (`0 = Male`, `1 = Female`).
- **ESS (Epworth Sleepiness Scale)**: One subject had a value of `30`, which exceeds the maximum allowable value of `24`.
- **Subject ID**: One Subject-ID number was listed twice `15-00609` which is an error.
- **Avg__DAily__CAP** : The data values in this column are 8 decimal points long. While this is not an error this is uncharacteristically long and I would reccomend they be rounded to 2 decimal points.

*Step 9. 4 charts and their description.*

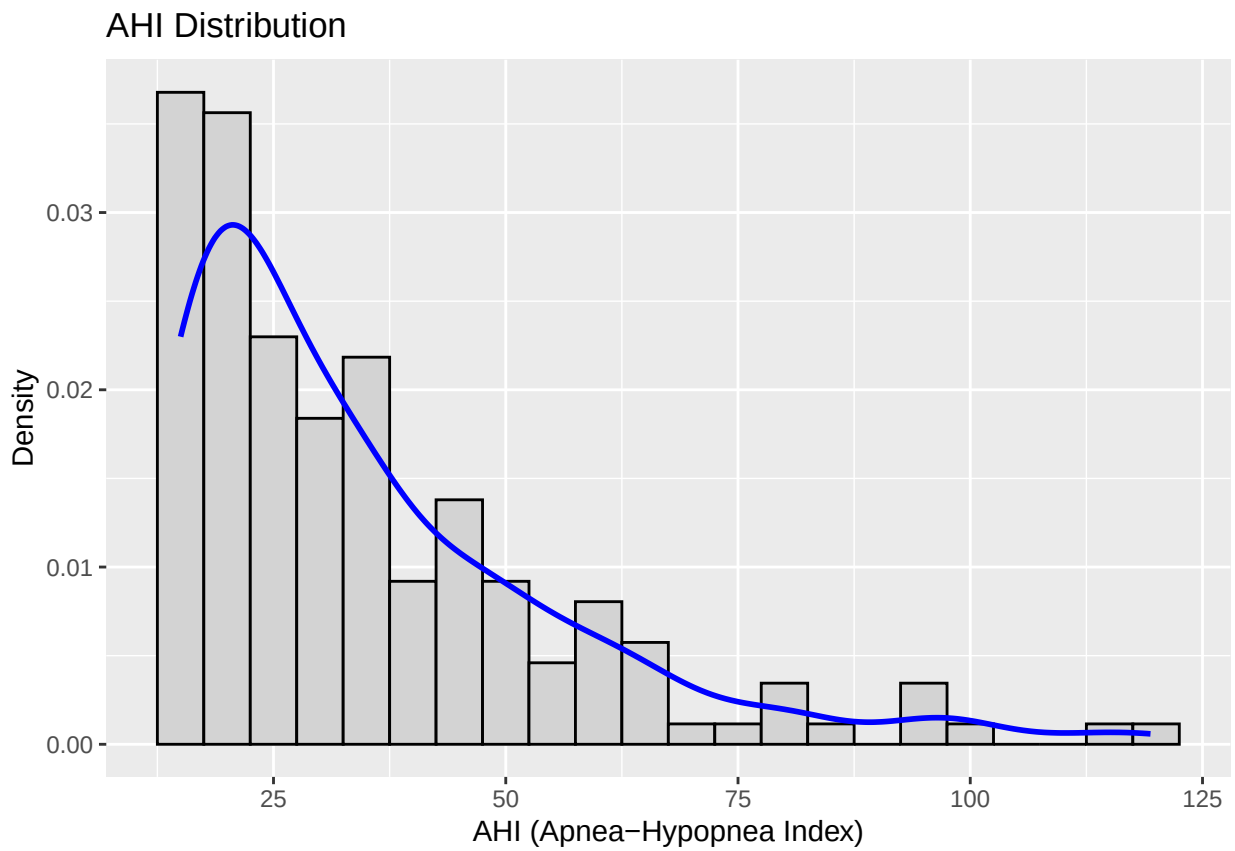


**Figure 1** Histogram showing the distribution of AHI

The distribution of the AHI values are right skewed indicating that most of the participants have a lower AHI value. This can suggest that even though many participants experience sleep apnea severe cases of sleep apnea are not common. The distribution display is univariate since it is only visualizing one variable which is AHI.

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (‘stat_boxplot()‘).
```

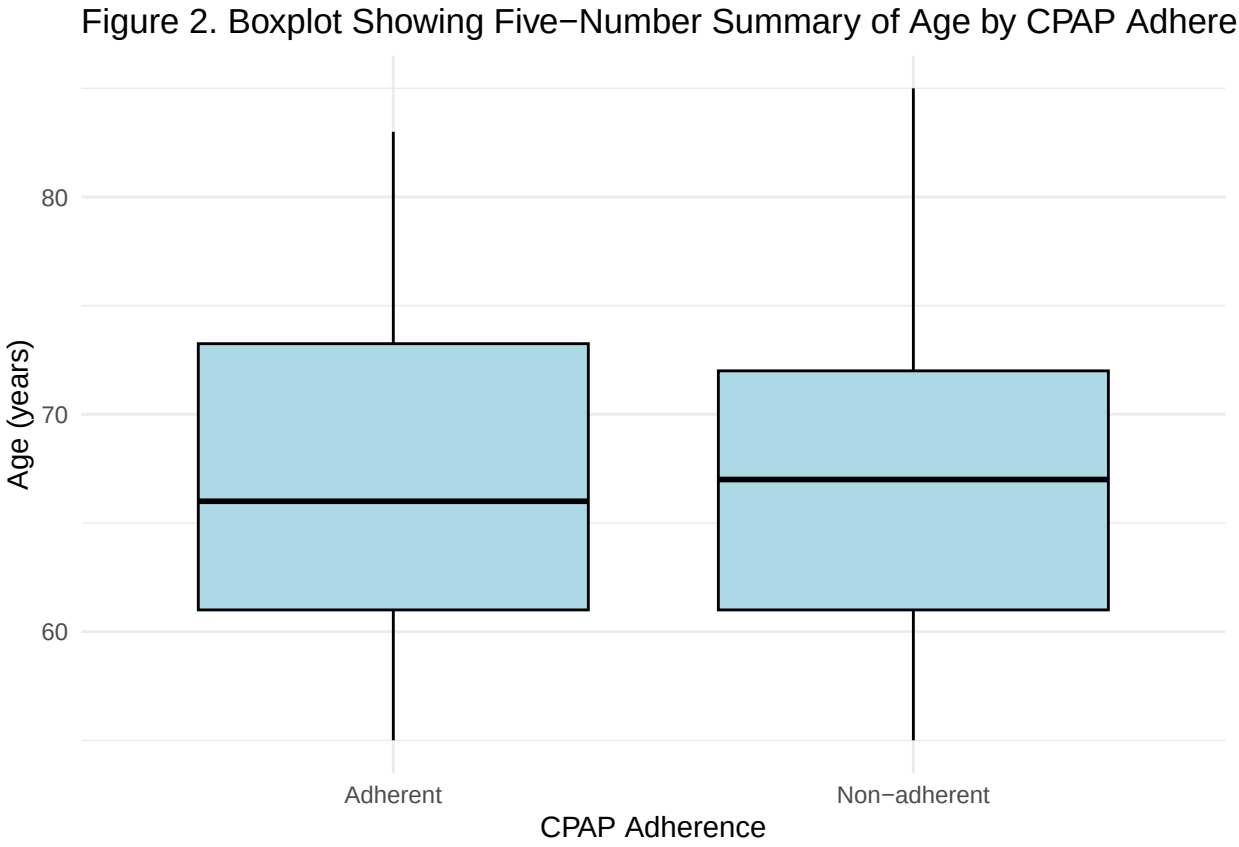


**Figure 2** Boxplot displaying the distribution of participant ages

In this boxplot, the adherent group shows a right-skewed distribution, indicated by the median line being slightly below the center of the box and a wider interquartile range (IQR). This suggests that the ages in the adherent group are just slightly more spread out, with some older individuals pulling the distribution to the right. The non-adherent group has a more symmetrical distribution, with the median line near the center of a smaller IQR, indicating that ages are more tightly clustered. This is a bivariate display.

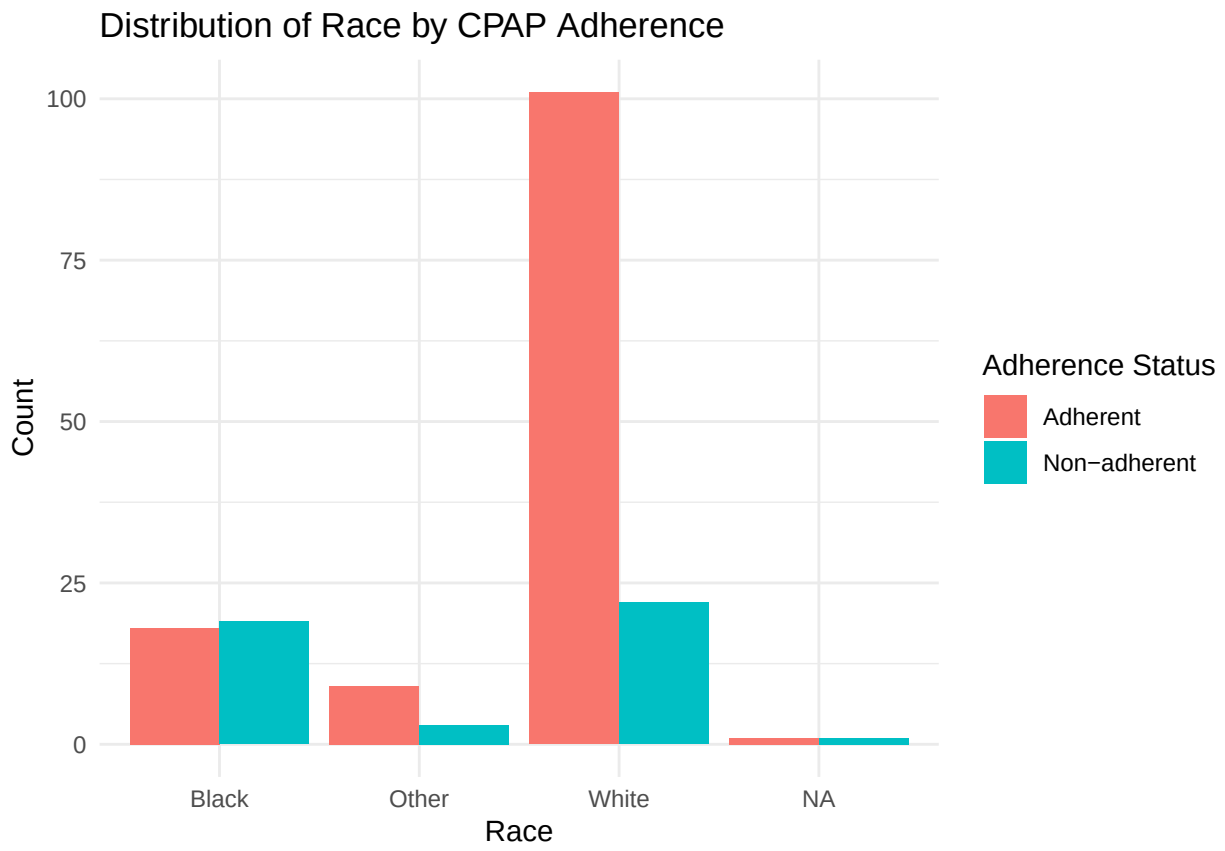


**Figure 3.** Bar graph showing the count of participants by race and their CPAP adherence status.

This bar plot displays the distribution of race alongside CPAP adherence compliance. It is clear that the number of White participants who are adherent is much higher compared to non-adherent White participants. In contrast, the difference between adherent and non-adherent Black participants is less pronounced, with slightly more individuals being non-adherent. This is a bivariate display.

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

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range
## (`geom_point()`).
```

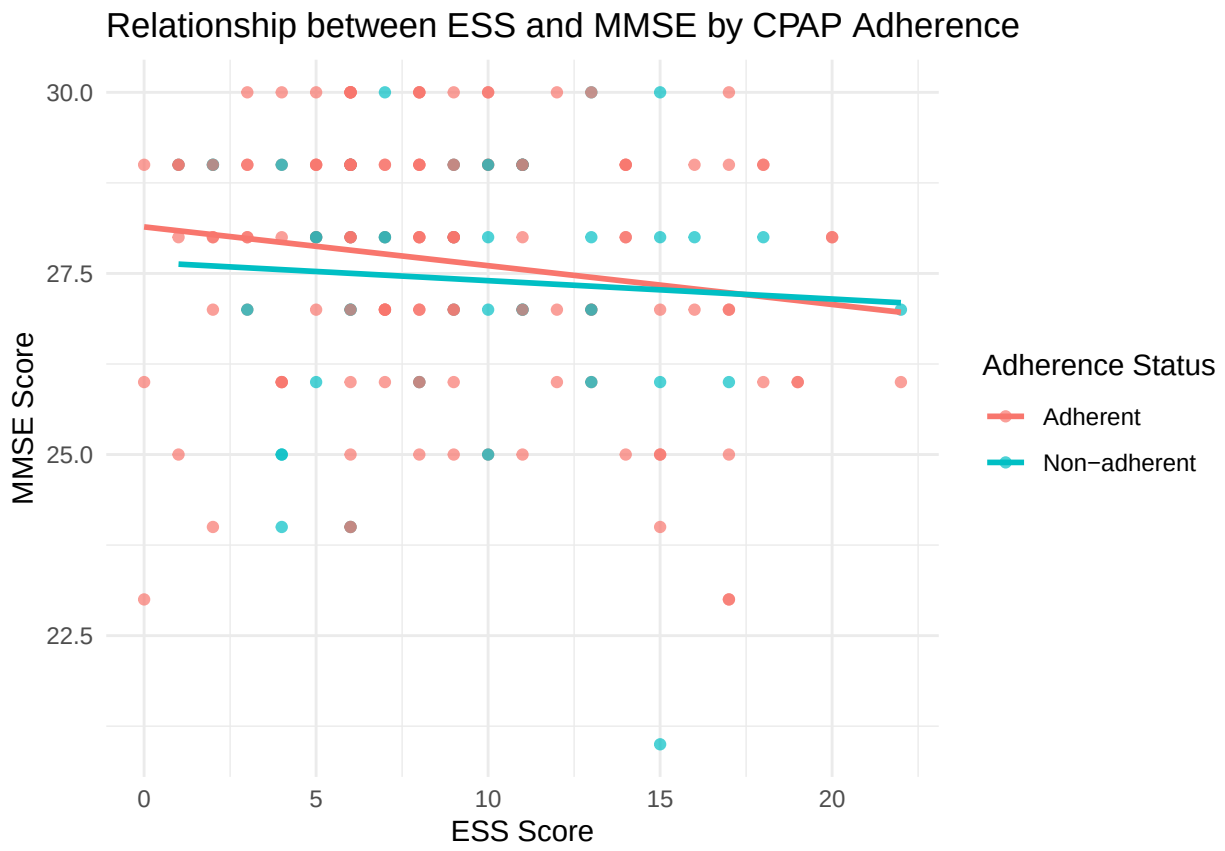


**Figure 4** Scatter plot displaying ess to mmse scores

This scatter plot shows the relationship between daytime sleepiness (ess) and cognitive function (mmse), with points and trend lines colored by CPAP adherence status. It highlights how sleepiness and cognitive scores relate across individuals who are adherent versus non-adherent to CPAP therapy.This is a bivariate display.

```
##                   Stratified by adherence_label
##                    level                    Adherent      Non-adherent  p
##   n                                           129            45
##   age (mean (SD))                          66.98 (7.51)  66.76 (7.50)   0.861
##   sex (%)          0                          70 (54.7)     24 (53.3)   1.000
##                    1                          58 (45.3)     21 (46.7)
##   ethnicity (%)    Hispanic or Latino         10 ( 7.8)      3 ( 6.7)   1.000
##                    Not Hispanic or Latino    119 (92.2)     42 (93.3)
##   education (%)    <= high school             24 (18.8)     12 (26.7)   0.362
##                    > high school             104 (81.2)     33 (73.3)
##   race_black (%)   0                         111 (86.0)     26 (57.8)  <0.001
##                    1                          18 (14.0)     19 (42.2)
##   race_white (%)   0                          27 (21.1)     23 (51.1)  <0.001
##                    1                         101 (78.9)     22 (48.9)
##   race_other (%)   0                         120 (93.0)     41 (93.2)   1.000
##                    1                           9 ( 7.0)      3 ( 6.8)
##   ahi (mean (SD))                          34.34 (21.18) 36.03 (19.91)  0.642
##   ess (mean (SD))                           8.79 (5.06)   9.13 (4.78)   0.691
##   mmse (mean (SD))                         27.66 (1.77)  27.42 (1.82)   0.444
##                   Stratified by adherence_label
```

```
##                     test
##   n
##   age (mean (SD))
##   sex (%)
##
##   ethnicity (%)
##
##   education (%)
##
##   race_black (%)
##
##   race_white (%)
##
##   race_other (%)
##
##   ahi (mean (SD))
##   ess (mean (SD))
##   mmse (mean (SD))
```

**Table 1**

This study examines older adults with sleep apnea, focusing on how consistent use of CPAP therapy (defined as greater than or equal to 4 hours/night) relates to sleepiness and cognitive function. The study sample included 174 participants, with 129 classified as adherent and 45 as non-adherent. The average age was similar between groups (approximately 67 years), and sex distribution was balanced, with just over half identifying as female in both groups. Most participants were not Hispanic or Latino and had more than a high school education. There were significant racial differences between adherence groups: a higher proportion of non-adherent individuals identified as Black (42.2%) compared to the adherent group (14.0%), and fewer were White (48.9% vs. 78.9%, respectively). Overall, adherence to CPAP was associated with notable racial disparities but not with differences in age, sex, education, or clinical characteristics.