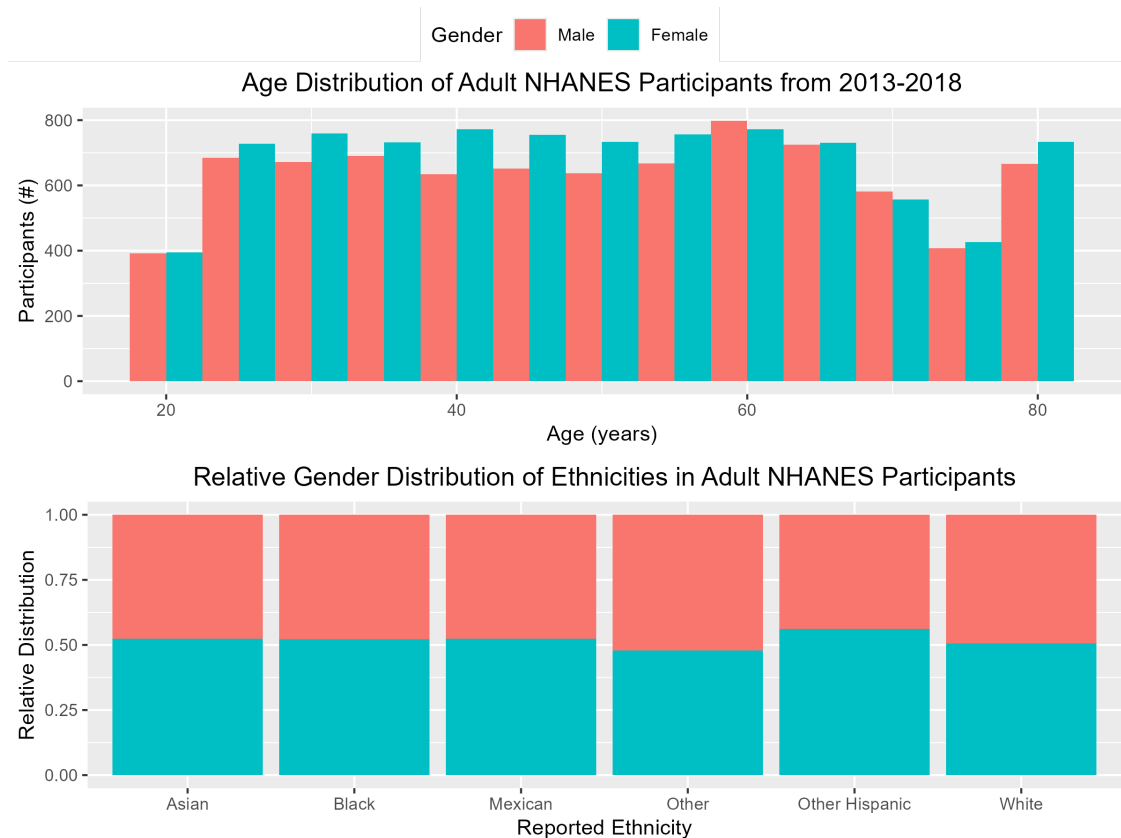# BIOS640 - Week 3 - Assessment

Harrison Saulnier

2025-09-21

Exercise 1 - Reproducing and arranging ggplot2 figures

Presented are the re-worked plots for the assessment, with their associated code.

The initial combination noted in the snippet of code, has the issue of a common legend occupying space, and making the combination look rather messy. Using ggarrange(), I was able to consolidate the legends, however the side-by-side arrangement smooshed the titles together. Thus, I opted to re-arrange the plots, to be on top of each other, with a consolidated legend, for a cleaner finish (presented above).

Exercise 2 - Visualizing key characteristics
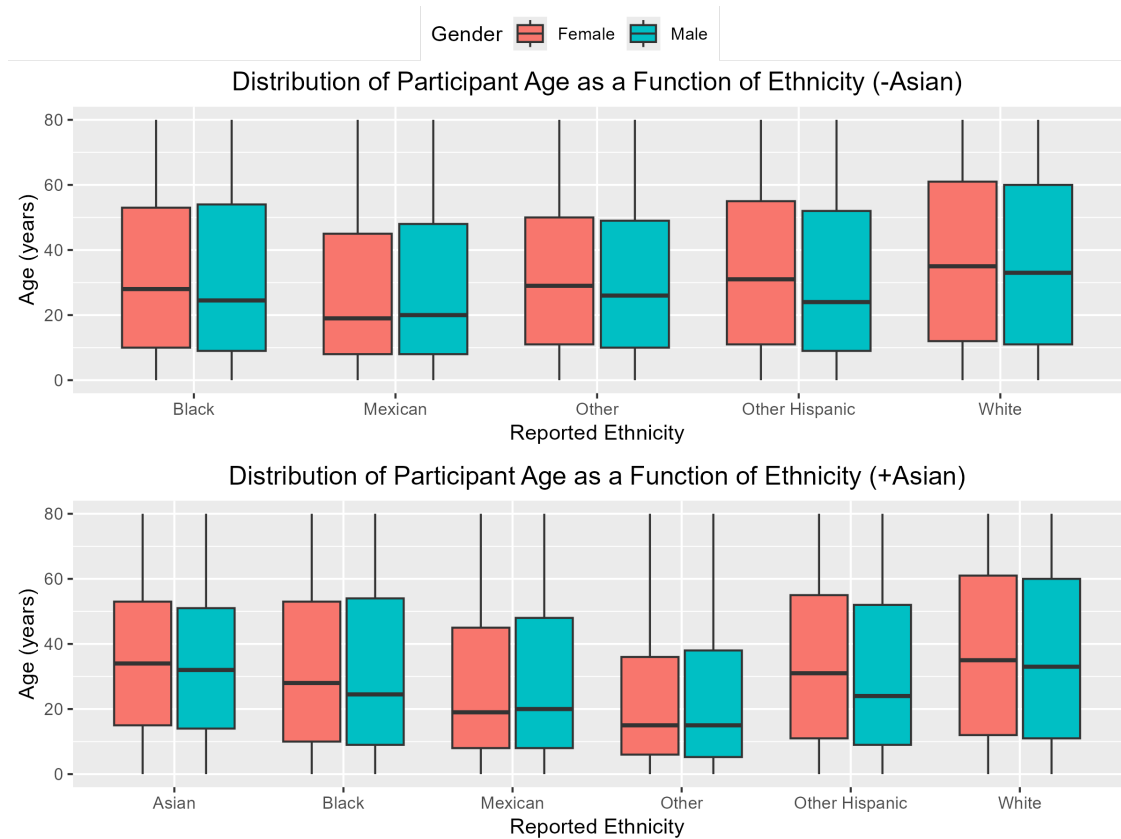
The second set of demographics, which identifies "Asian" as a specific demographic, is the superior set of a data to analyse. As can be seen on the lower bar graph, those who identify as "Asian" account for many older adults, whereas those who identify as other are significantly younger that other ethnicity groups. When the data was consolidated in "other", the data was not as descriptive of the true distribution of demographics.

Exercise #3 - Improving ggplot figure

The issues with this graph I noted include: *The plot is cluttered with several lines, making it difficult to approach at first glance.* The number of different participant plots makes it difficult on the eyes to trace the individual's trajectory without putting your finger on the screen. *The number of individuals has exhausted the color palette for discrete differentiation, thus a subtle change in blue means the several participants cannot be differentiated. Even if individual symbols were added to each participant, this would add to the problem of crowding.* There is no title to the plot, no units for the weight axis (kg vs. lbs), and the x-axis is a variable of time measured in weeks. The legend has no title either. *The baseline of each individual cannot be represented in a useful way, on a single consolidated chart.* The general trend of weight progress for the cohort is not represented, which is a useful piece of information to have.

Change in Weight for Individual Dieting Participants
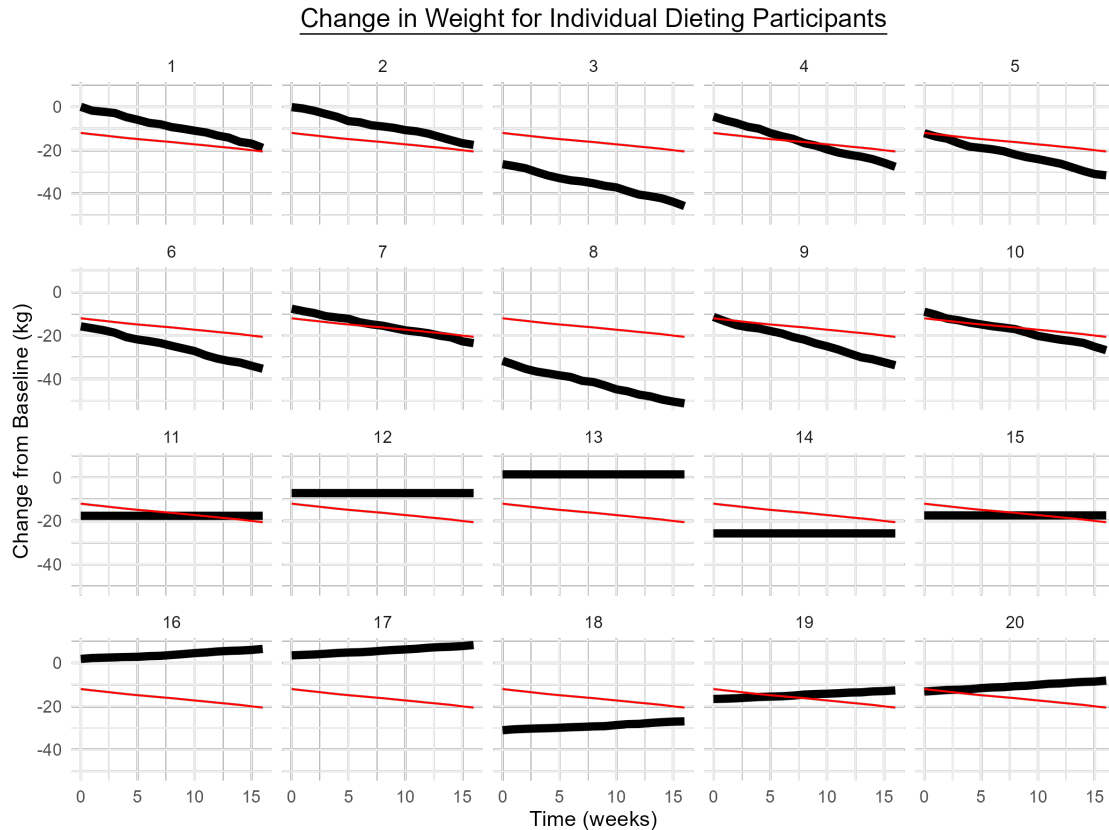
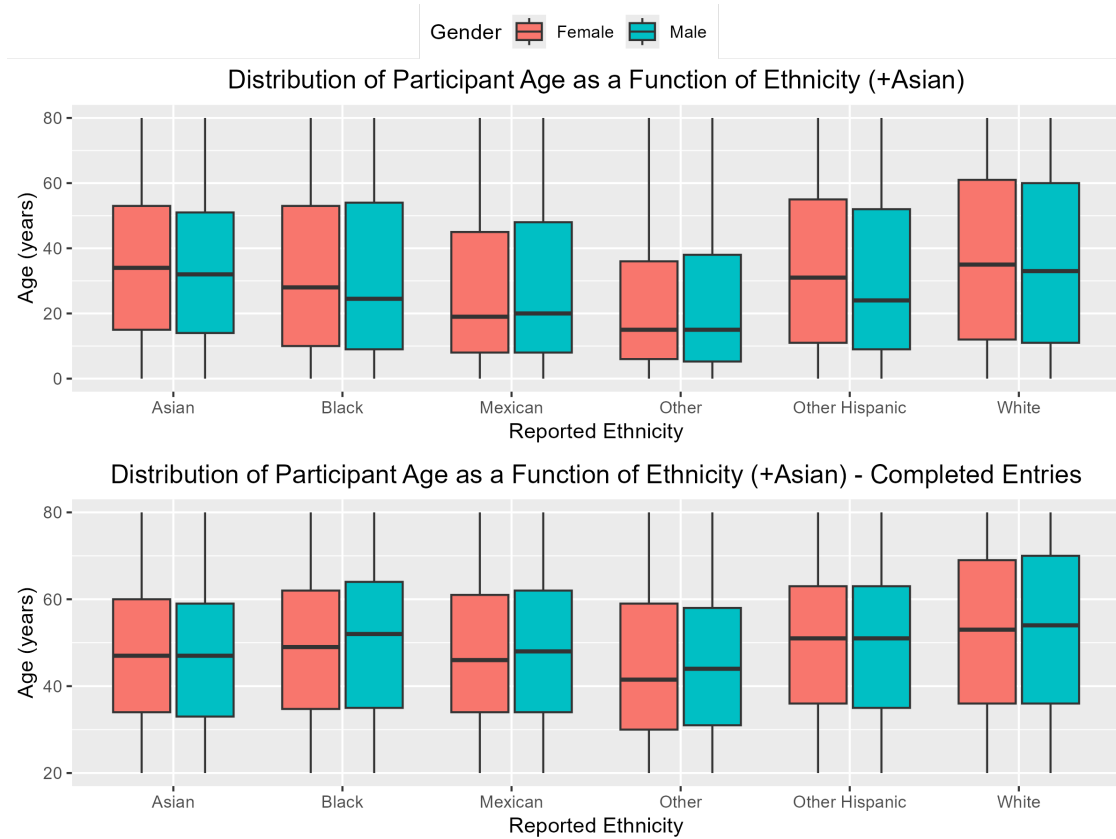**Exercise #4 - Exploring relationships through visualizations**

Appraisal of NHANES Adult Participant Blood Pressure Data from 2013-2018

Introduction

The National Health and Nutrition Examination Survey is a CDC managed project that collects data from participants across several demographics. This report addresses gaps in information collection, and the general trends in blood pressure as functions of age and gender, for adult patients. The term adult is attributed to participants within the ages of 20 and 80 years old; those below the age of 20 are considered pediatric, and those above 80 geriatric.

Age distribution of participants accross ethnicities

The NHANES study had two stratification for ethnicity; one that included diversity in the title of "Asian", while another placing these participants in the "other" category. Given the importance in genetic differences between the diverse nations of Asian, the more descriptive data set was used for this report. When reviewing the data for incomplete entries, interestingly the average age of participants shifted upward, as seen in the figure below. This prompts one to question why participants of a younger age, may not be completing their questionnaires and evaluations, and proposes to review the method of the study to increase younger participation.

Distribution of Participant Age as a Function of Ethnicity (+Asian)

Distribution of Participant Age as a Function of Ethnicity (+Asian) - Completed Entries

# Blood Pressure Categorizations across Age brackets and Gender

This data review notes the differences in prevalence of different stages in hypertension, of non-specified cause, among adult participants in the NHANES. The first half of the chart below notes a significantly large population of normotensive women when compared to men across all age groups, with more Elevated and Stage 1 Hypertensive men than women. The prevalence of Stage 2 HTN participants appears to be approximately the same for both genders.

The following chart notes an expected trend of worsening blood pressure, as correlated with age. Once again, we see a reduced overall participant population in the youth, as mentioned in the previous section, but a relative trend of increasing blood pressure correlated with age. Participants over the age of 60, have slightless less Elevated and Stage 1 HTN individuals as compared to normotensive ones, with the majority suffering from Stage 2 HTN.

Relative Blood Pressure Characterization between Genders in Adult NHANES Participants
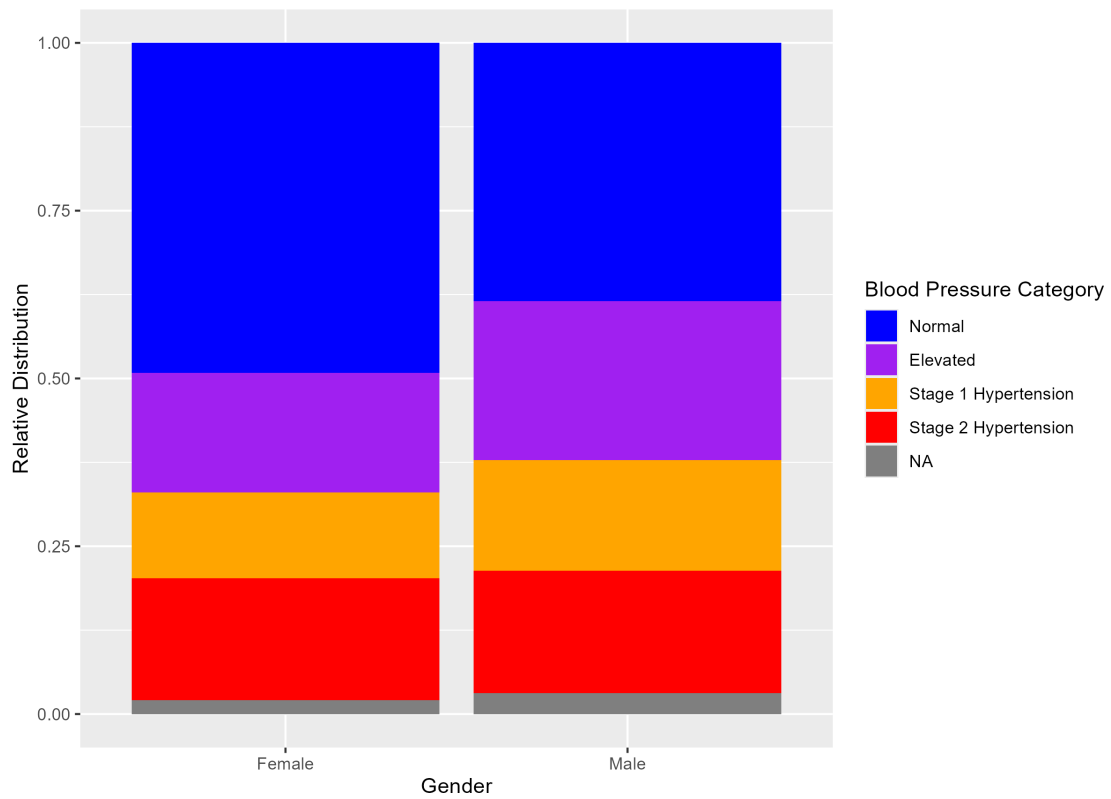
NHANES Blood Pressure Characteristics of Age Categories
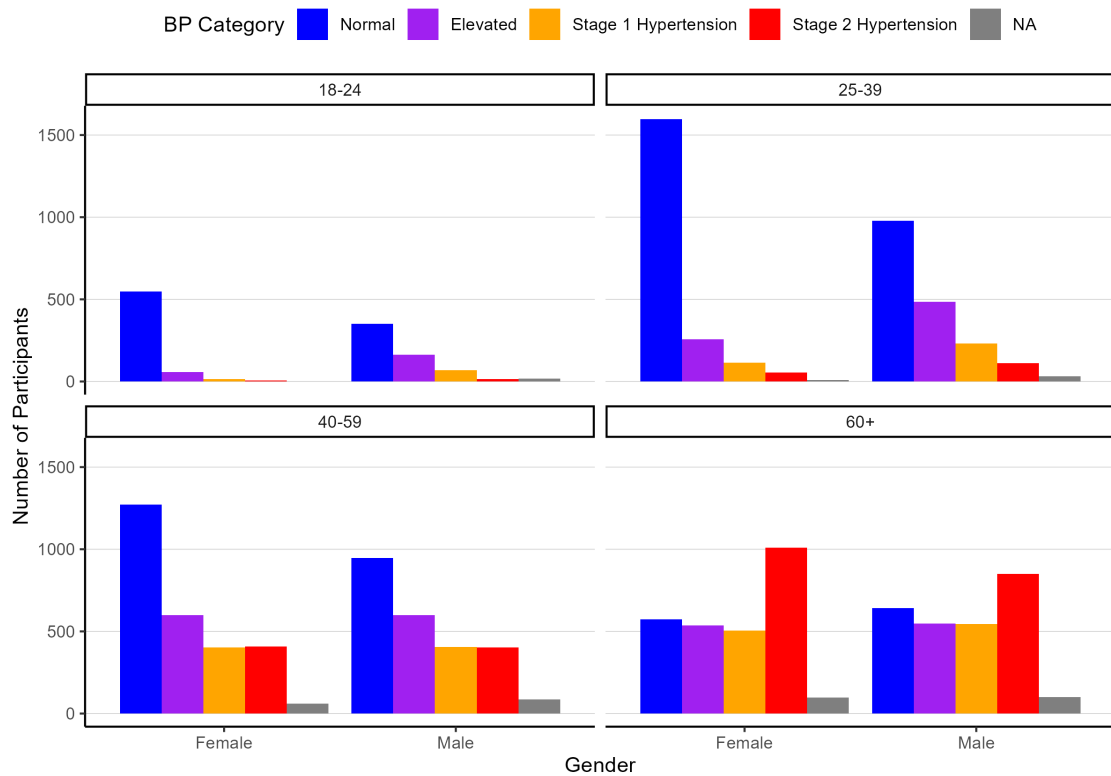
Conclusions and Recommendations

Overall, removing participants with incompleted entries for gender, age and average SBP resulted in an increase in the overall age of the sample size. Across all ethnicities, we note the minimum age rising to above 30 years old, whereas before, every ethnicity category had entries below 20 years old. There is also a narrowing of the age gap between ethnicity groups, after this filtering of the data. One issue which you can see in the second plot, regarding blood pressure, is that there are still participants labelled as "NA", despite my manipulation of the data. I'm not sure where I went wrong, and look forward to feedback from the grader to prevent this issue in the future.

Based on these findings, it would behoove the NHANES team to identify individuals with incomplete NHANES data sets, and release surveys with following assessment to identify possible barriers to participation, issues with the assessment itself and possible solutions to increase participation for a more accurate data set in following years.

Writer's Issues with Assessment and Coding

As stated in the code, I had issues with a few points that I would love to gain feedback on. Despite trialing several NA-dropping codes, I continued to have an NA column in my final chart (Exercise #4). Due to time constraints I was not able to find a solution prior to submission, and I do apologize for that burden on the corrector. Furthermore, in Exercise #3 I could not find a way to have a label for the red line on the graph itself, identifying it as the study's general trend of weight loss. Finally, on two occasions I wanted to have my code inserted into the final report, but despite enabling the echo = TRUE, this did not occur. After troubleshooting, I could not find the code that was preventing the output. Assistance with that would be must appreciated as well.

Appendix A - Raw code from R-Studio for Exercise 4