M5_Lab4

Emily Shepherd

2023-02-04

The Data

The data I chose was used to determine stroke risk based on eleven indicators. It is a large data set, so I chose to focus the visualizations on single women who lived in urban areas. I chose this data set because I am interested in health-related data.

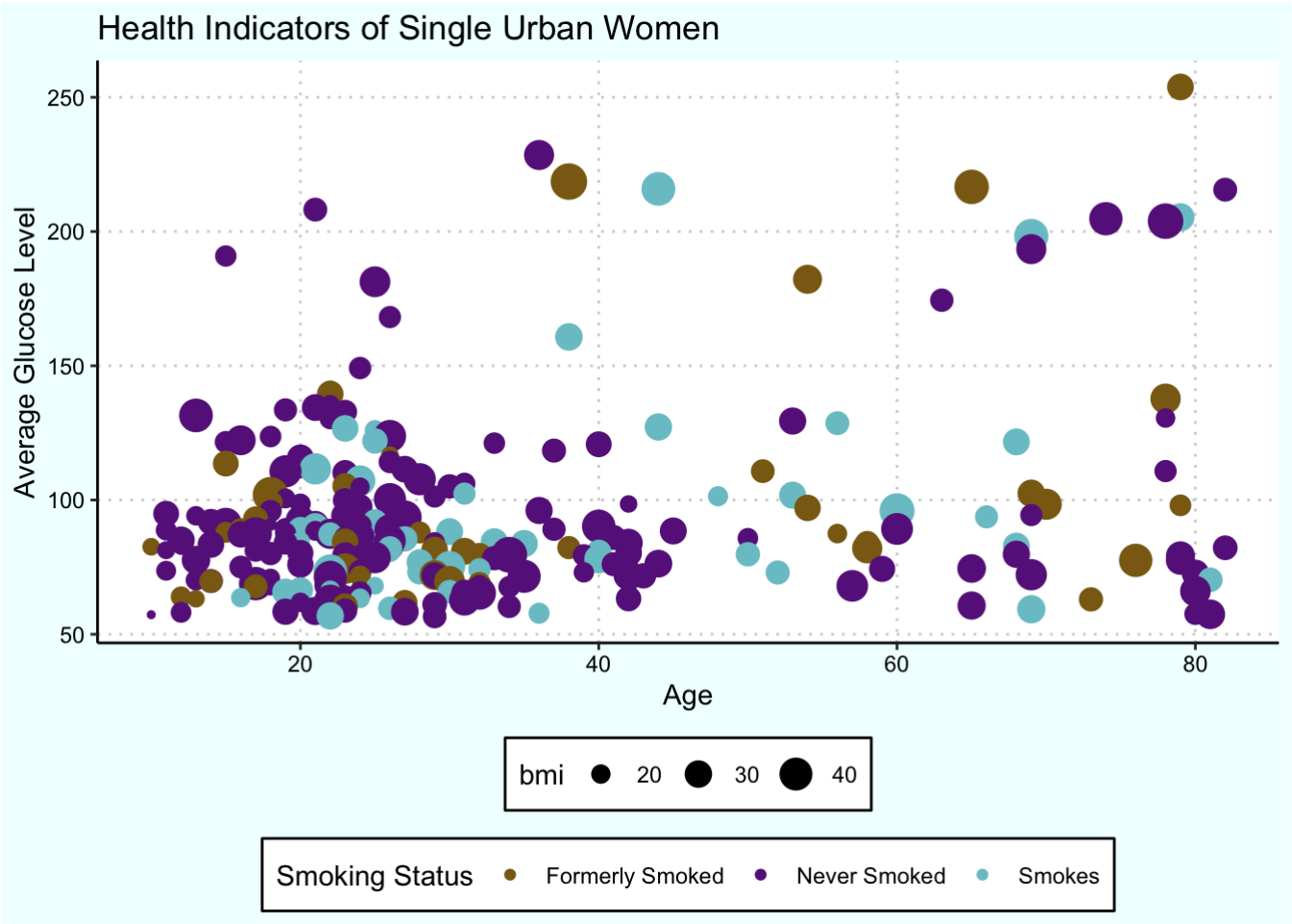
The Bubble Plot

My visualization has age plotted along the x-axis, average glucose level plotted on the y-axis. The size of the points are determined by BMI. Additionally, the color of each bubble was determined whether the woman never smoked, formerly smoked, or is currently a smoker.

I chose to choose the colors manually. I avoided reds and greens because they are hard to discern by many people. The colors I chose were goldenrod4 (Formerly smoked), darkorchid4 (Never Smoked), and cadetblue (Smokes.) I chose the background panel to be azure, because I thought it went with the general palette.

There were three main differences between my bubbleplot and the one constructed in the demo. First, I manually chose the colors. Second, I formatted the text in the legend, so it was not the variable names, but what the variables represented. Lastly, I manually determined the theme to be used by choosing the background and grid colors.

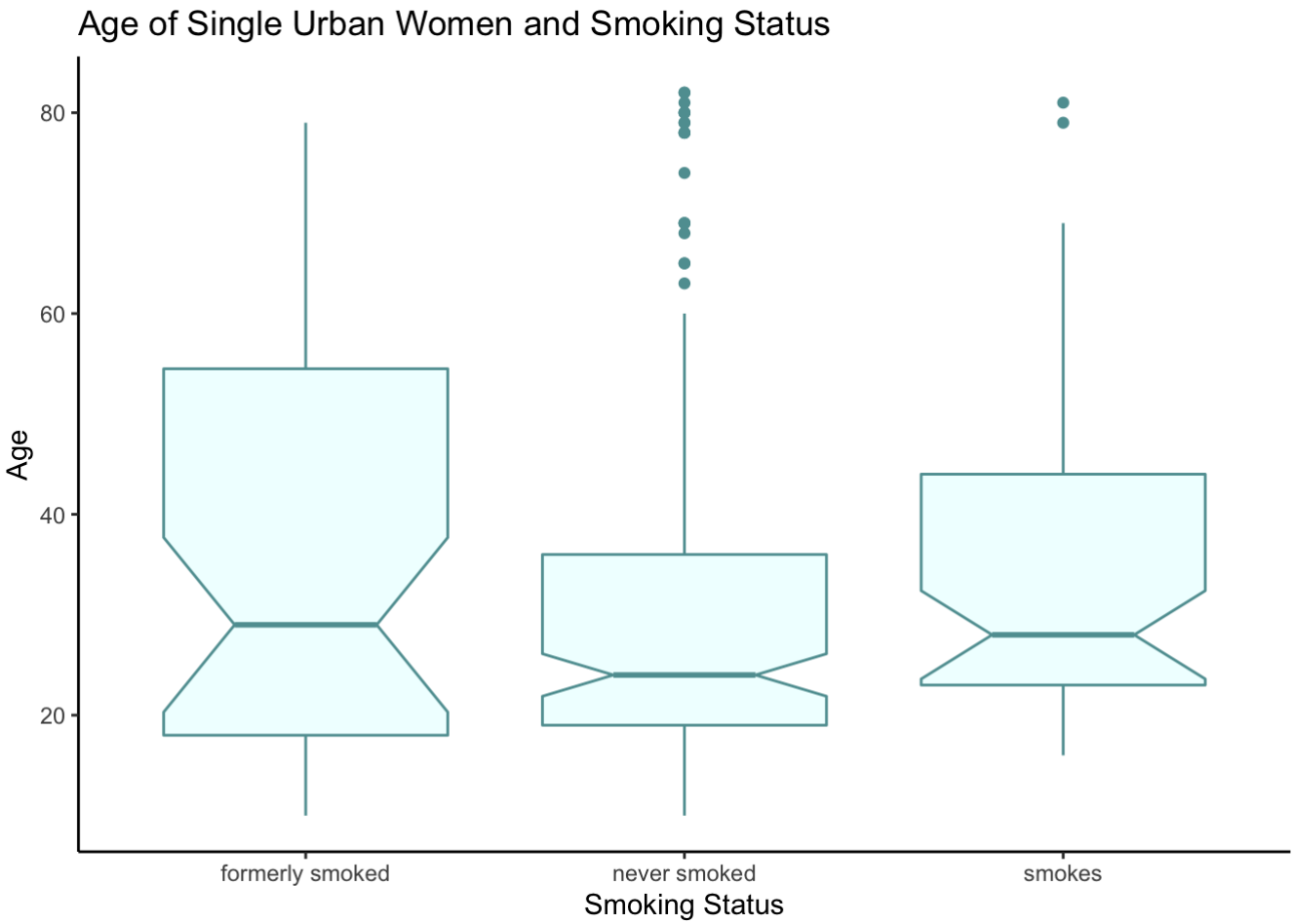
From the plot, there did not appear to be a strong relationship among age, average glucose level, and bmi. However, I did notice that there were very few older smokers in the plot. I decided to explore this relationship with a boxplot.



The Boxplot

One way in which my boxplot differed from the one in the demonstration is that I chose to use ggplot to create it. I plotted smoking status along the x-axis and age along the y-axis. I chose to create the outline cadetblue and the fill azure. Part of the reasoning that I did this was so it would coordinate with the bubbleplot. Instead of manually choosing my theme, I picked theme_classic(), because it is clean and easy to read. I changed the notch width to 0.5.

I noticed several features in the graph. First of all, the median age of formerly smoked is higher than both never smoked and smokers. This makes sense because the women in this category had to start smoking and then decide to quit. I also noticed that the observation from the bubbleplot that there were very few older women who smoked was verified in the boxplot. The maximum appeared to be about 70 years old with two outliers above 80. This observation makes sense because smoking has been scientifically linked to premature death.


Age of Single Urban Women and Smoking Status
80
60
40
20
Age
formerly smoked
never smoked
smokes
Smoking Status