

SEIS631 Final Project

Nikolai

5/4/2022

Introduction

Personality has always been a fascination of mine. While many personality tests have little validity or predictive value, The Big Five personality tests have a reputation of being a good measure with predictive power. I have found a publicly available dataset of demographic information and personality data that I wish to analyze using techniques taught in class.

What

Why

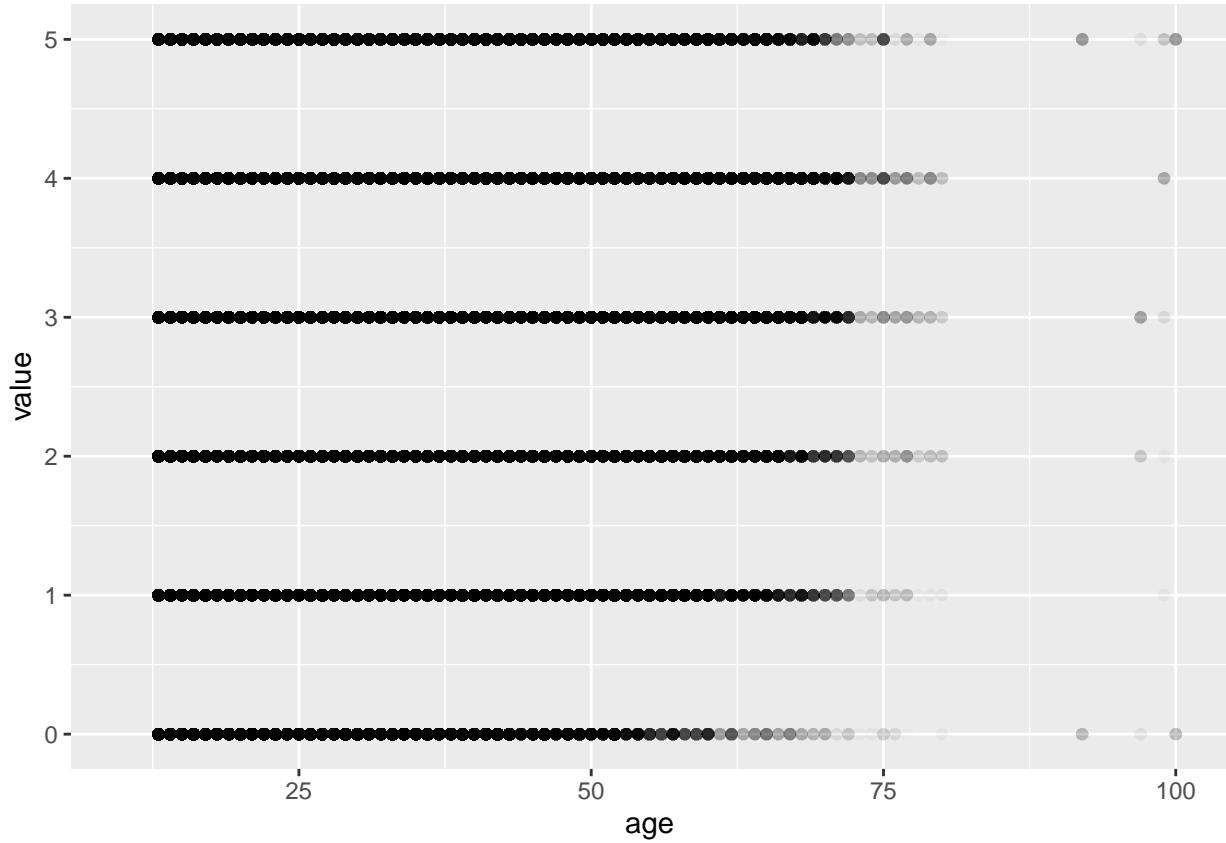
How

Body

Cleaning the data.

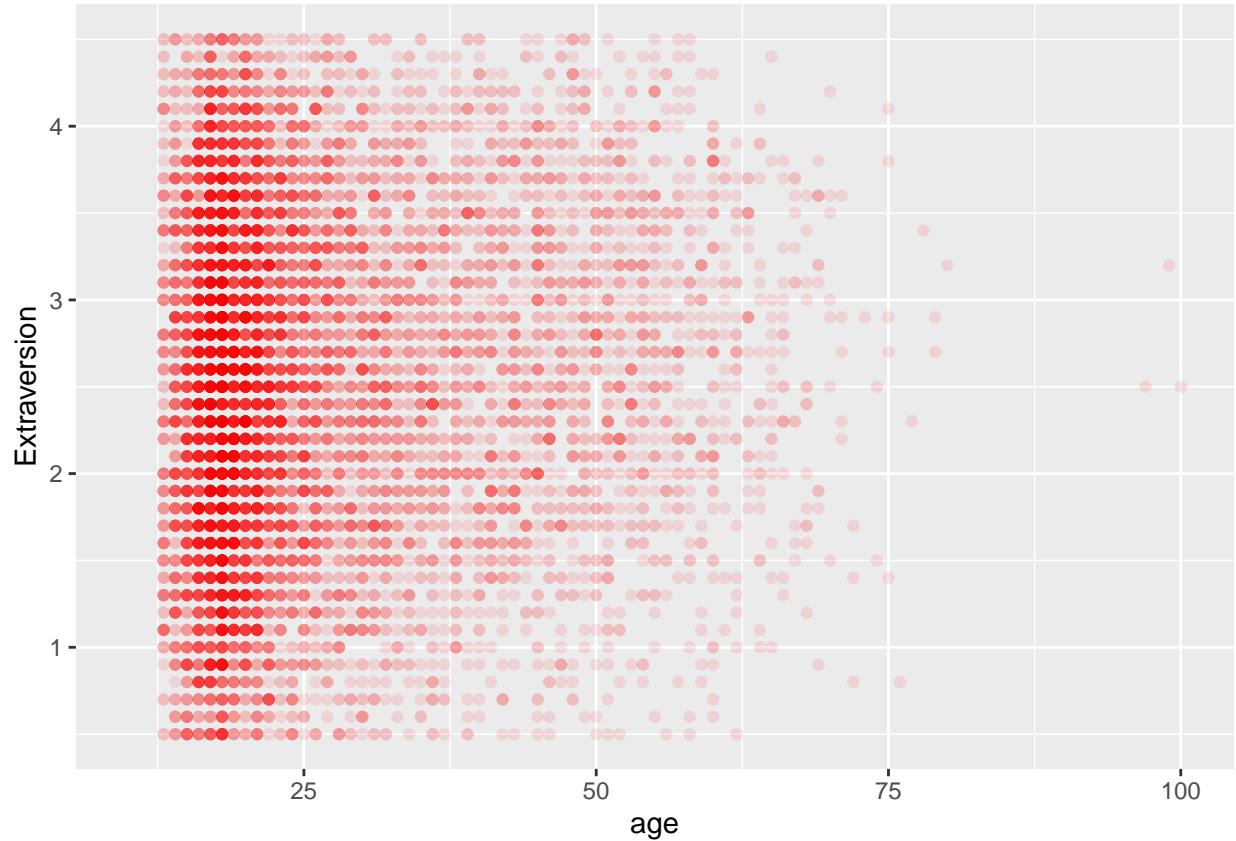
In order to account for the preference of a person to respond positively or negatively to questions, OCEAN tests usually have answers that mean similar things but phrased in a positive or negative manner. In order to compare these types of data, first I will have to convert these negative values to positive ones. I have created a function to do this, and using the codebook that came with the data I can determine which questions were negative or positive. I'm going to first look at extroversion versus introversion, clean the data, and plot it for US populations to see if there are any trends.

```
## Warning: Removed 4150 rows containing missing values (geom_point).
```

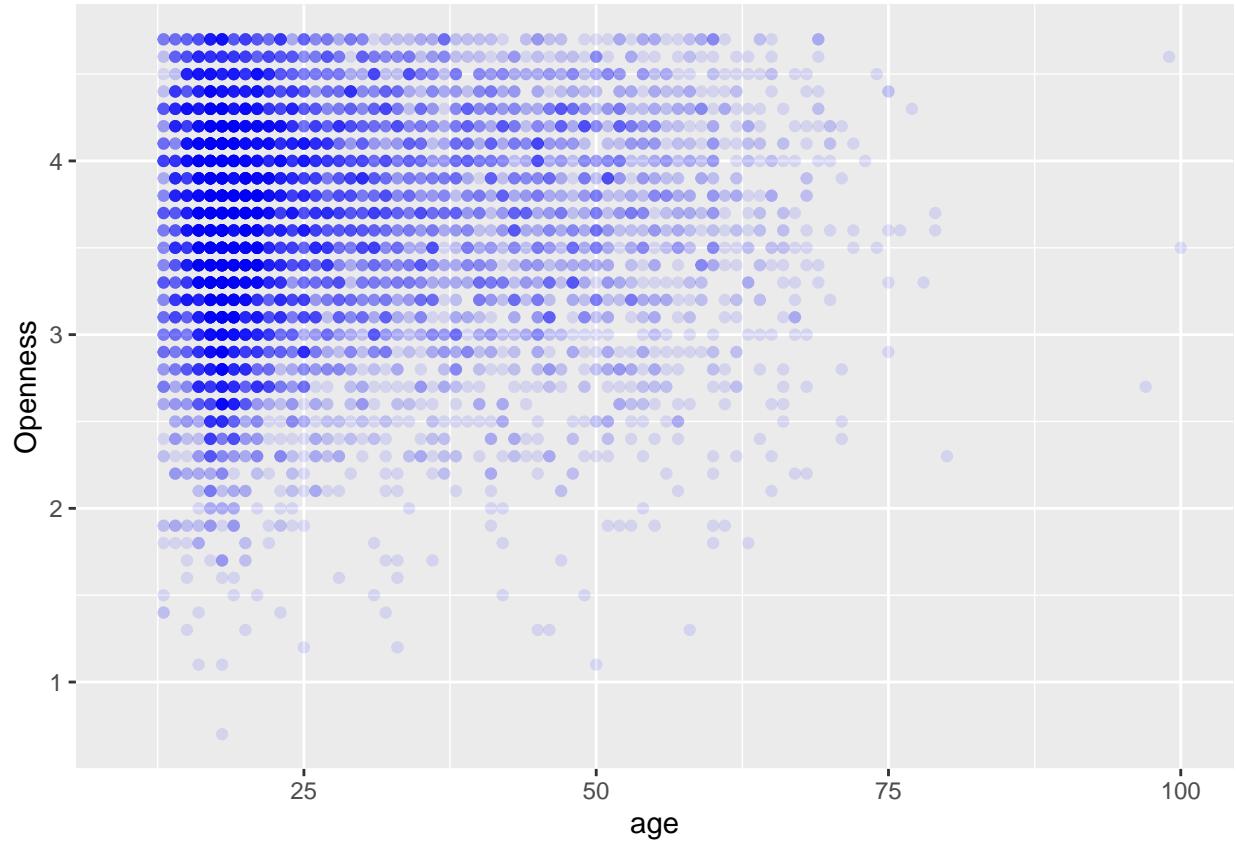


looking at the above graph of age versus extroversion, it seems there is little relation between age and being introverted or extroverted with the exception of after age 60 it appears like less people are strongly introverted.

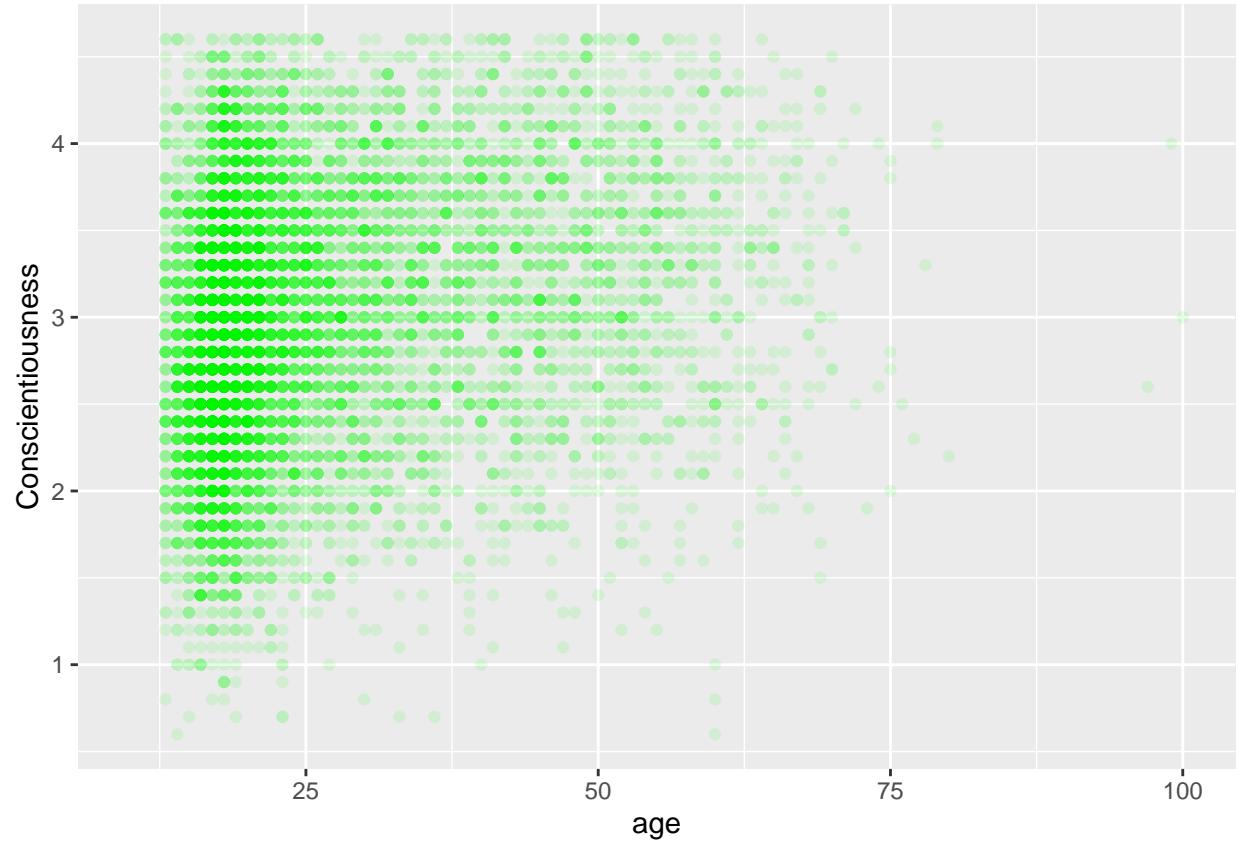
```
## Warning: Removed 21 rows containing missing values (geom_point).
```



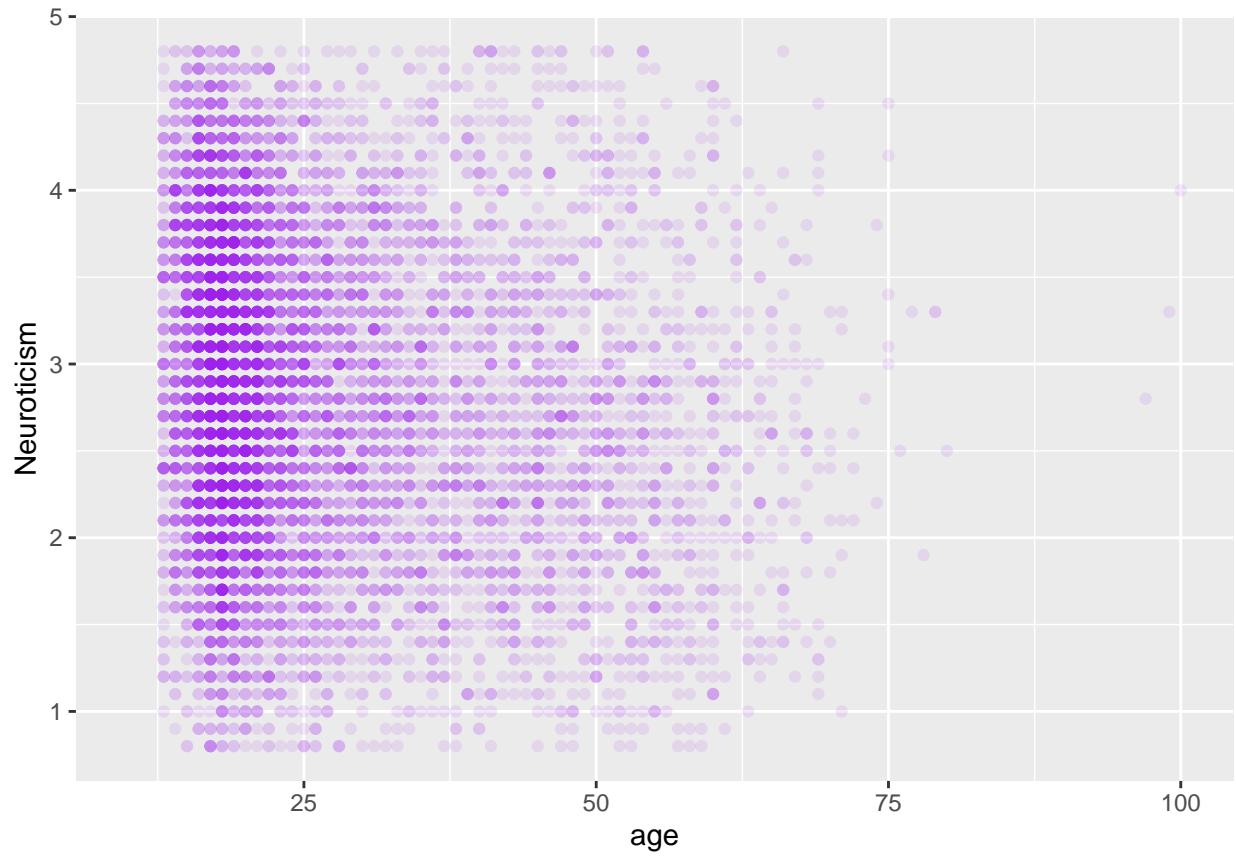
```
## Warning: Removed 21 rows containing missing values (geom_point).
```



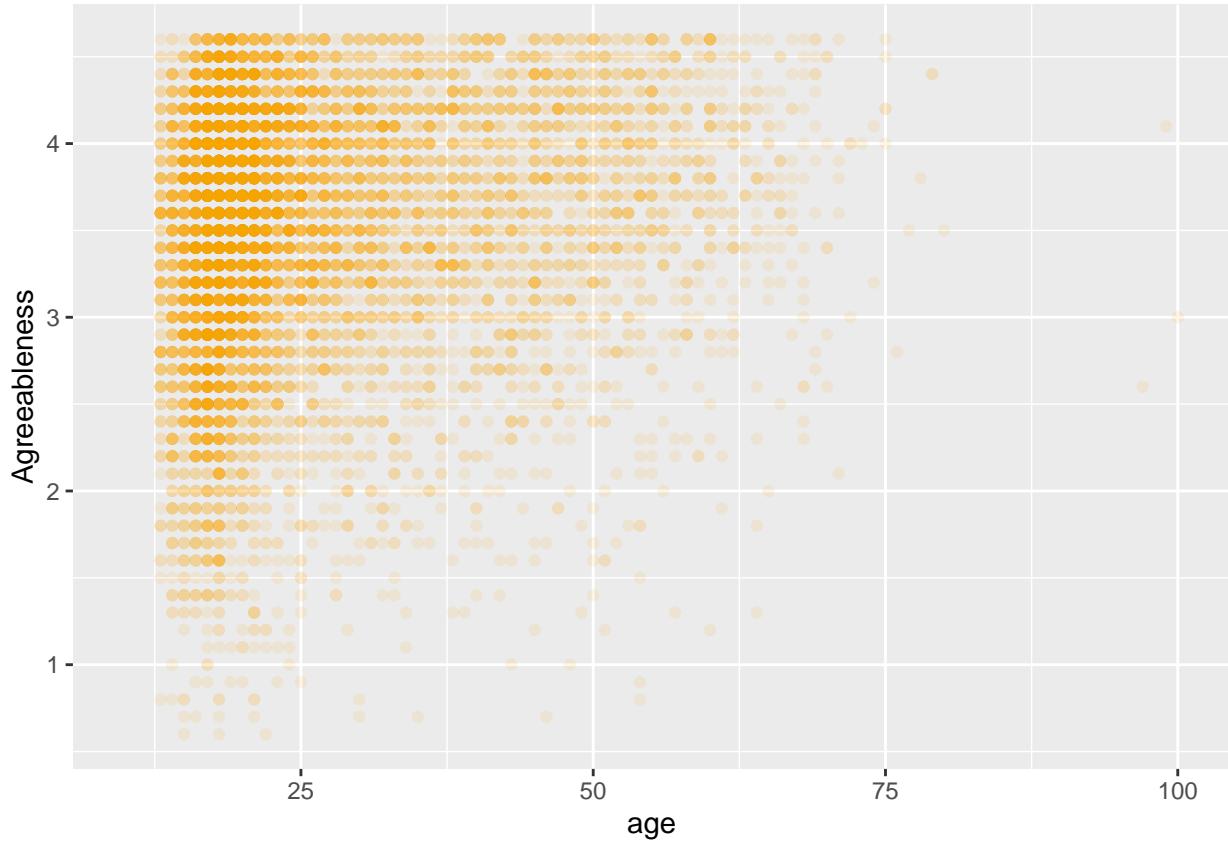
```
## Warning: Removed 21 rows containing missing values (geom_point).
```



```
## Warning: Removed 21 rows containing missing values (geom_point).
```

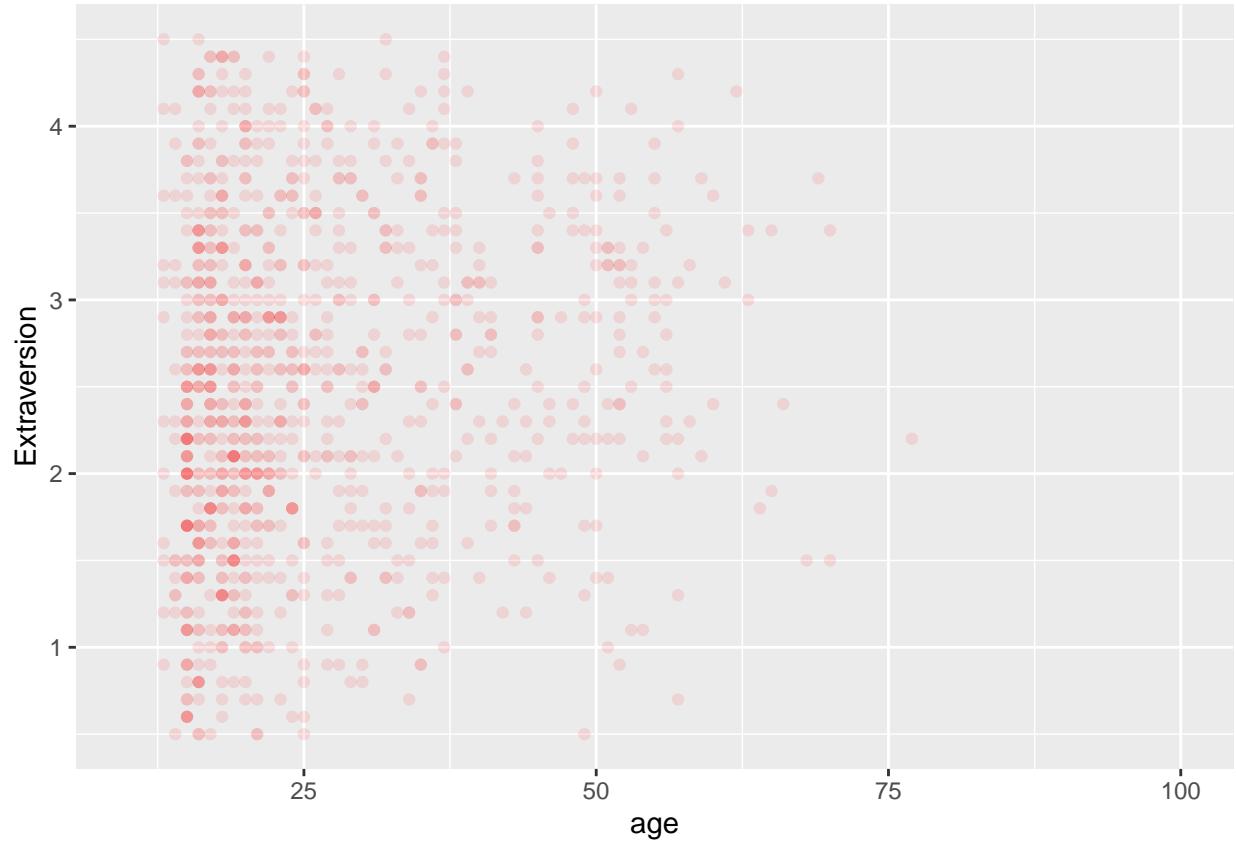


```
## Warning: Removed 21 rows containing missing values (geom_point).
```

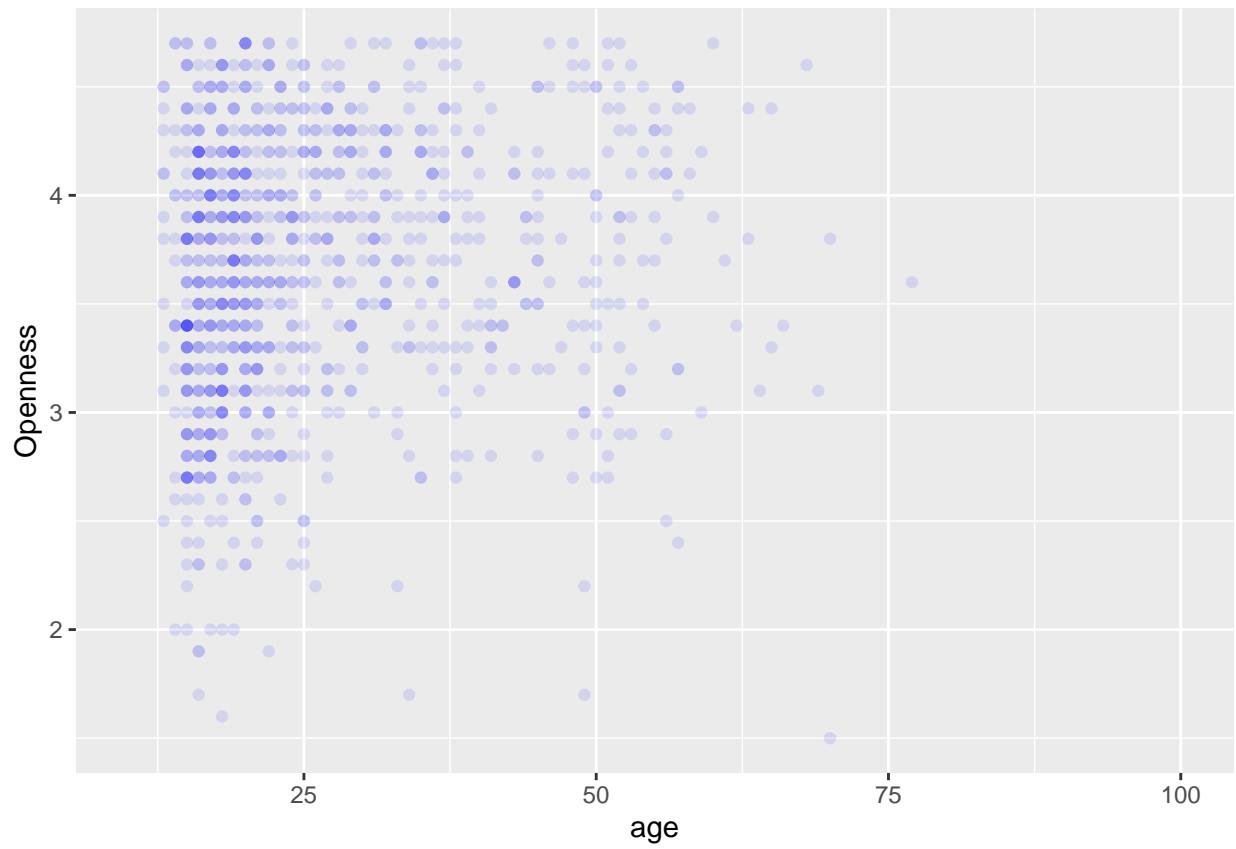


If we take the averages of all of the question's scores for each trait for respondents in the US, we see that most respondents were age 25 or younger. We also see trends, such as Agreeableness and Openness tend to be higher than 2. Conscientiousness and Neuroticism seem to avoid the valences of 1 and 5. Extraversion, on the other hand, appears to be across the board.

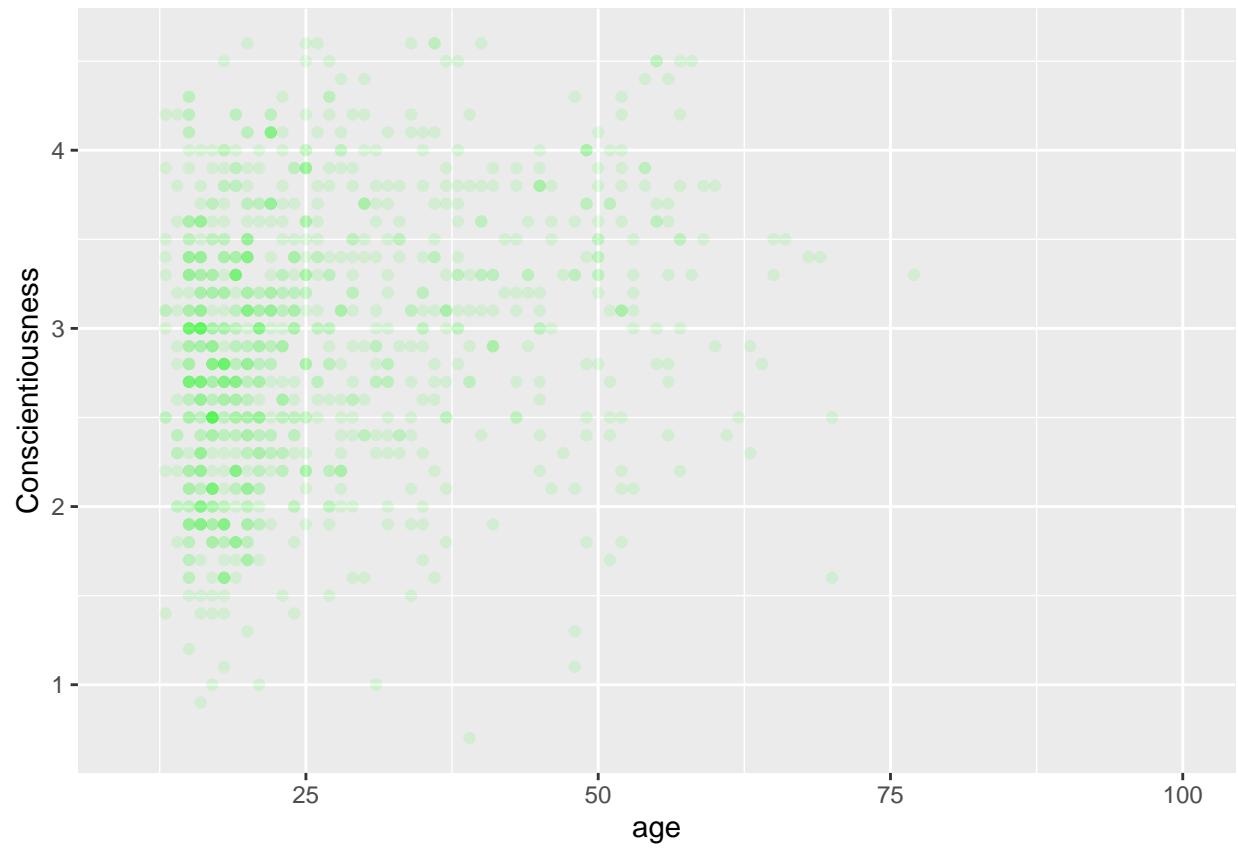
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



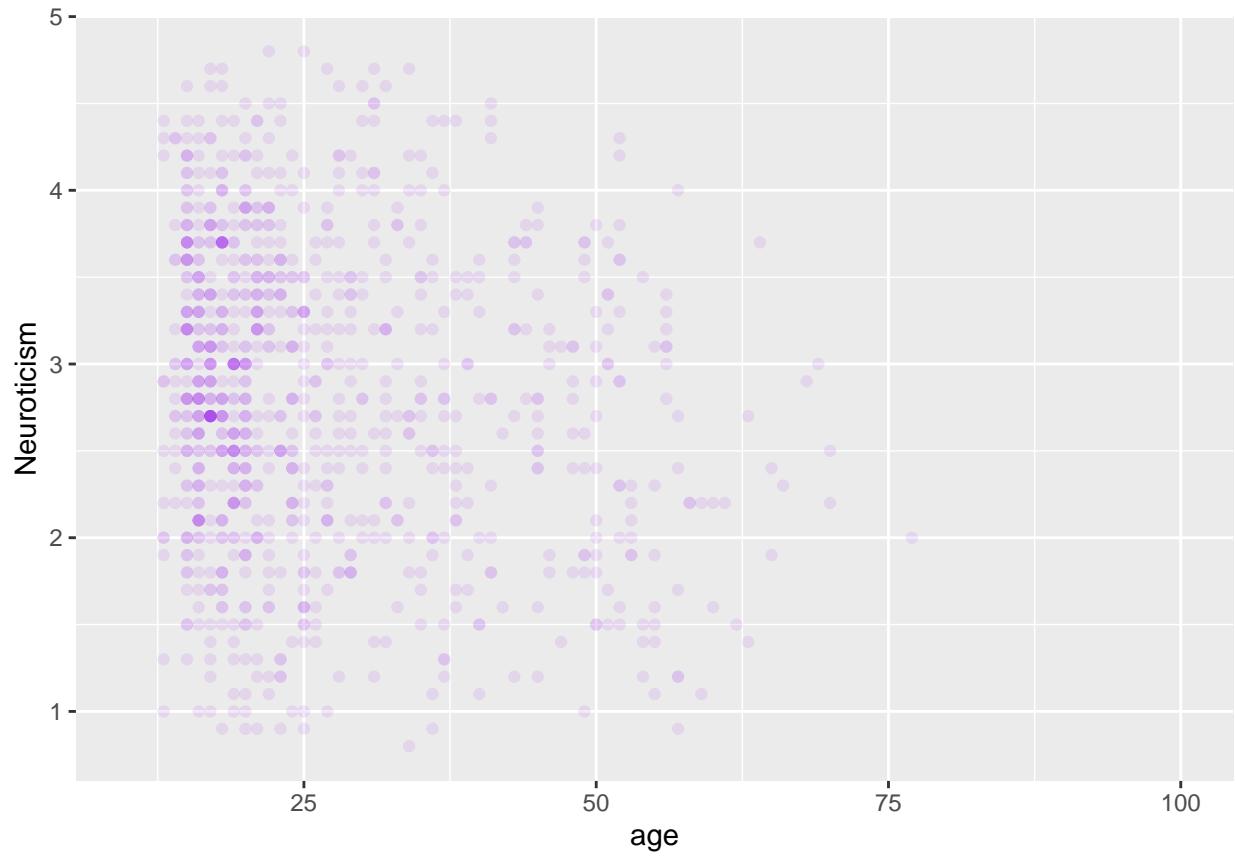
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



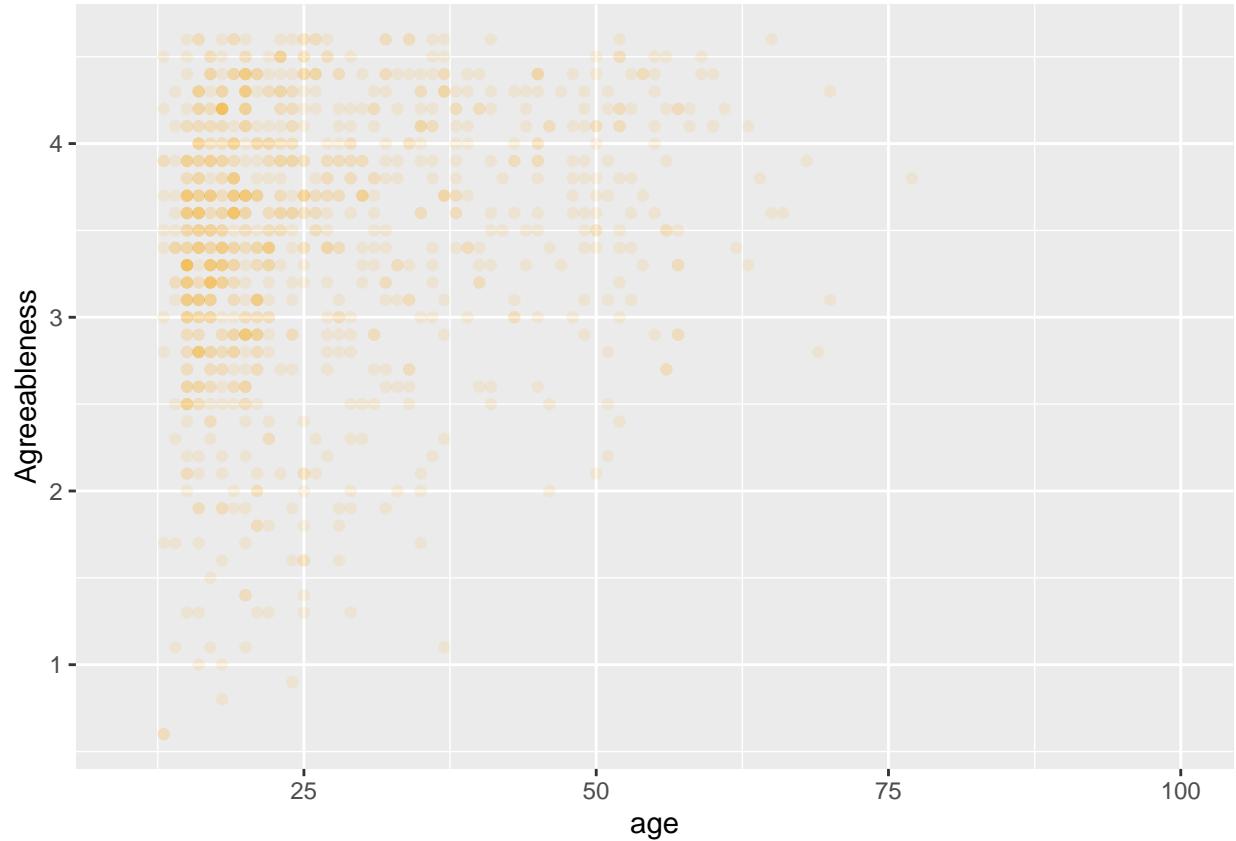
```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```
## Warning: Removed 2 rows containing missing values (geom_point).
```



```
## Warning: Removed 2 rows containing missing values (geom_point).
```



Topics From Class

- (a) Git
- (b) RMarkdown
- (c) Statistical concepts such as normal distributions, mean, standard deviations, percentiles, and areas under the curve
- (d) Geometric distributions
- (e) T tests to determine if two populations are significantly different.

Conclusion