

# Sta 440 Case Study 1

*Amanda Levenberg, Jeffrey Ho, Lucy Lu*

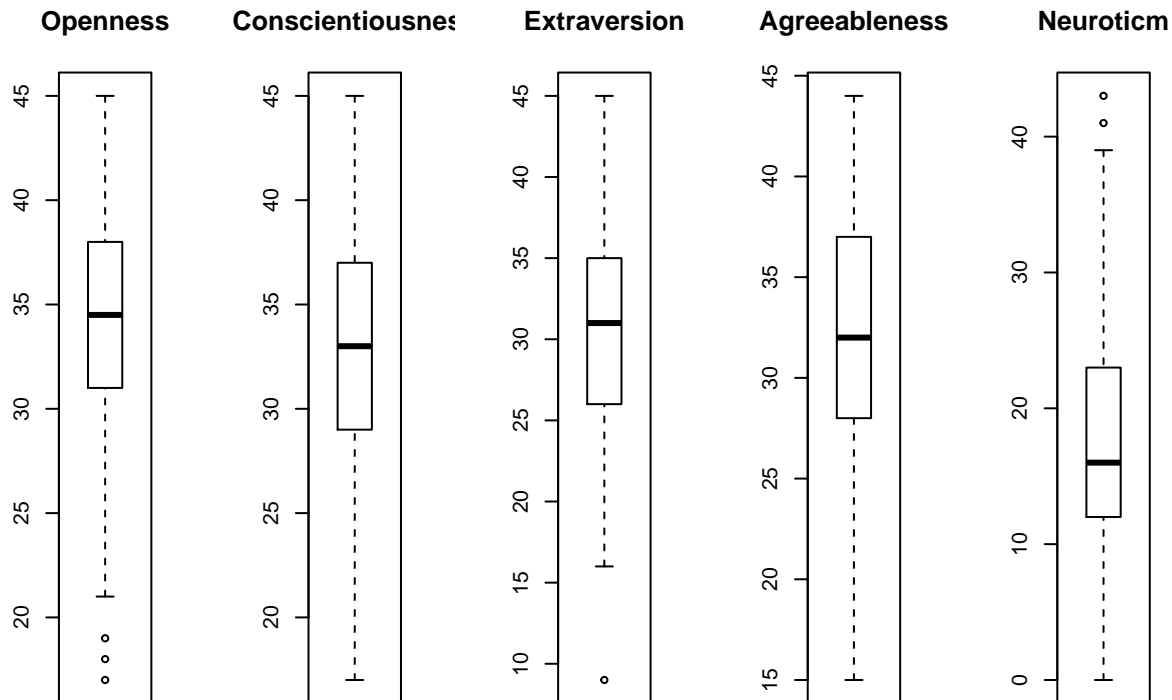
*September 7, 2016*

## Background

The study of neuroscience aims to examine the human nervous system, specifically as it relates to the rest of the human body. In particular interest is the relationship between brain structure and human behavior, such as the link between specific brain connections and cognitive traits. This analysis aims to study the relationship between brain fiber connections in regions of the brain and the commonly studied Big Five personality traits: openness, conscientiousness, extraversion, agreeableness, neuroticism. Each personality trait is measured on a spectrum between two extremes. For instance, the openness of an individual is qualified as a value between inventive/curious and consistent/cautious. Through network modeling and analysis, this study will aim to elucidate whether connections between certain regions of the brain, in particular those across hemispheres, are associated with personality tendencies in the Big Five personality traits when controlling for confounding variables such as the age and sex of the subject.

## Exploratory Data Analysis

Based on an exploratory data analysis, the variables for the Big Five personality traits appear fairly normally distributed with only a slight skew.



Openness: mean = 33.93, median = 34.50, variance = 37.64

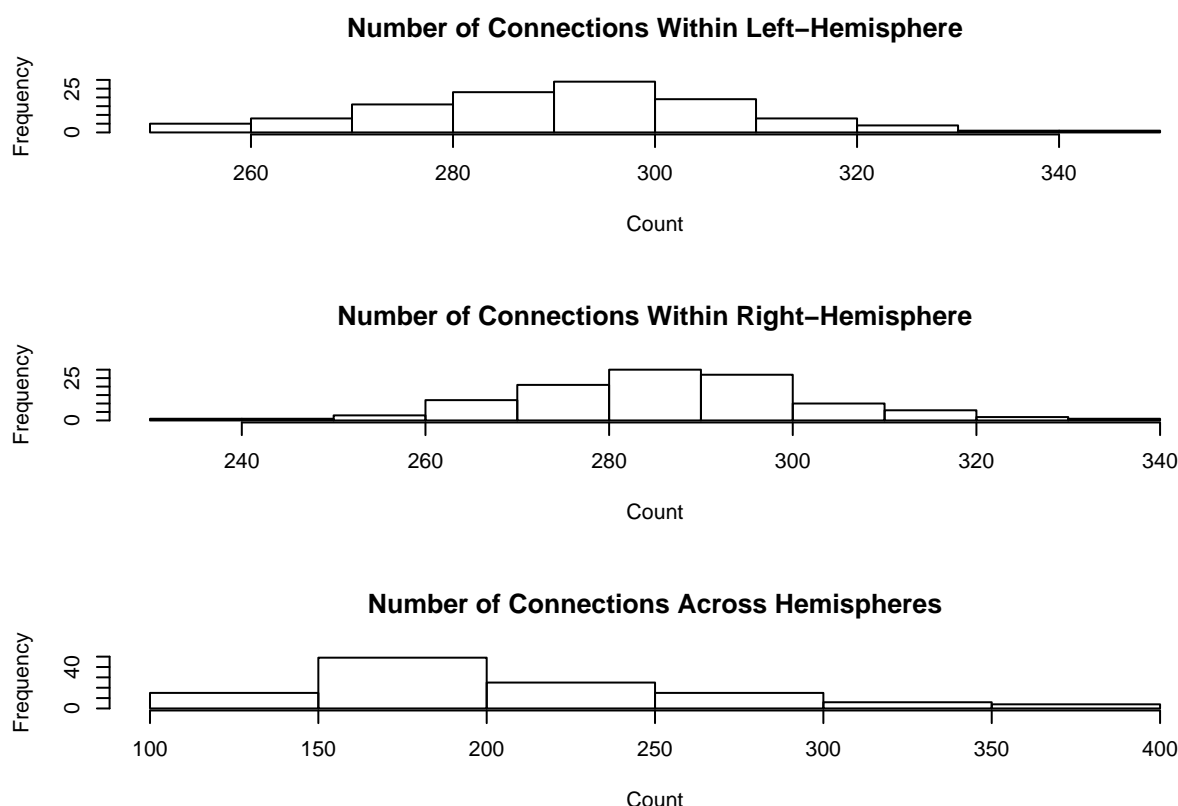
Conscientiousness: mean = 32.76, median = 33.00, variance = 37.33

Extraversion: mean = 30.25, median = 31.00, variance = 39.37

Agreeableness: mean = 31.92, median = 32.00, variance = 44.18

Neuroticism: mean = 17.61, median 16.00, variance = 74.83

Each variable appears similar in center and spread, with the exception of neuroticism, which contains a substantially lower center and a larger variance. Some of the variables (openness, extraversion, and neuroticism) contain outliers, though none of the outliers appear extreme enough to warrant concern for the analysis. Similarly, an analysis of the FSIQ and CAQ variables do not provide significant concern. FSIQ is normally distributed, with a median of 120, a mean of 119.4, and a variance of 188.5. The variable for CAQ is significantly right skewed, with a mean of 17.81 and a median of 13. Consequently, it may be necessary to transform the variable in future analysis.



Each subject's brain connectivity is represented by a 68 by 68 matrix, of which every element corresponds to the number of fiber connections between two brain regions. A first approach to study the connectivities is to transform the matrices to be binary. Since only about 20% of all entries are non-zero, 0 is set as the threshold above which an entry is counted as 1. After the transformation, 4 counts are collected for each subject: the number of connections within the left hemisphere, within the right hemisphere, between hemispheres and overall. Examining the distributions of these values, it can be seen that within-hemisphere connections appear normally distributed, whereas the distribution of across-hemisphere connections is right-skewed and has a greater variance.

Additionally, the exploratory analysis aims to look at covariance between the covariates. Below, two correlation matrices demonstrate the correlations between various covariates. The

correlations across the two plots are the same (represented by the numbers), but the boxes that are colored vary across the two plots. The shaded boxes indicate a significant correlation between the two covariates. The first plot has boxes shaded with a 95% confidence level, and the second plot has boxes shaded with a 99% confidence level. As seen in the matrices, gender has an association with some of the OCEAN covariates, including conscientiousness and agreeableness. Additionally, some of the other metrics exhibit a significant positive correlation to each other, such as CCI and FSIQ as well as CCI and openness.

With regards to the data itself, the indicator and URSI covariates were removed for the analysis. Additionally, there was some missing data. In order to continue with the analysis, if a value was missing, it was replaced with the column mean as an estimate. Future analysis may potentially involve an alternative method to estimating these missing values. For instance, one can insert the mean of a subgroup (i.e. gender) that the subject belongs to as opposed to the mean of the entire group.

