Political Orientation & Brain Region Development:

Visualization Using Density Estimation

Margaret Catherman

October 1, 2023

Abstract:
Is there a relationship between the size of two regions of the brain, amygdala and acc, and an individual's political orientation to the left or right? And are these two regions independent of each other? Using various statistical approaches and resulting visualizations, this analysis will show first, that amygdala and acc have a low interdependence; and second, how one may infer there is a difference between brain structure composition and political views. Throughout the analysis, when examining the conditional mean for these features, we noticed that acc is lower for liberal political orientation, and increases as political orientation moves to the conservative end of the spectrum. The inverse is true for amygdala; with its mean highest among those claiming to be most liberal, or '2", and decreasing to its lowest among those at the conservative end of the spectrum, '5'. (It should be noted that this is the *opposite* of findings in all other studies on this topic.) There have been numerous similar analysis over the past 30 or so years; this analysis is based on the study by Kanai, R., Feilden, T., Firth, C. and Rees, G., 2011, " Political orientations are correlated with brain structure in young adults", Current biology, 21(8), pp.677-680., data was collected on 90 university students to study whether or not the two brain regions (amygdala, acc) are likely to be independent of each other, and if the size of these regions correlate to different political views in individuals.

Data; The data set n90pol.csv contains information on 90 university students who participated in Kanai and Feilden's psychological experiment. This is a small excerpt of the original data set, which may account for the following analysis reaching a different conclusion than that of the original experiment.

Methods:
Throughout this analysis, we will use various statistical approaches to visualize the relationship between the composition of of amygdala and acc levels in participants' brain and their political orientation: one and two dimensional histograms and kernel density estimator (KDE); scatter plots and line charts of the mean, variance, and skew based on variable orientation and conditional distributions. Furthermore, we will establish that the two variables (amygdala, acc) have a low interdependent using probability and statistics theory, with visualization of their respective mutual information and permutation scores.

Before we review the findings of this analysis, let us evaluate the strengths and weaknesses of two of the methods we will use: KDE and the histogram.

" A kernel density estimator is essentially as smooth version of a histogram" (An Introduction to Statistical learning with Applications in R, Second Edition James, G., et all)

One advantage of employing histograms in statistical analysis is their capacity to aggregate data without necessitating the retention of the entire dataset in memory. This attribute renders them memory-efficient, as they do not require the storage of all individual data points. However, when dealing with high-dimensional datasets, histograms exhibit certain drawbacks. Specifically, they become less sample-efficient due to the numerous bins required in high-dimensional spaces. The section of bin numbers significantly influences the resulting histogram, making it susceptible to pronounced fluctuations in appearance, even with slight alterations in bin size. Too many bins introducing 'noise'. Furthermore, histograms may pose challenges when dealing with datasets comprising only a few sample points.

Comparatively, Kernel Density Estimation (KDE) offers better statistical performance guarantees than histograms. Another advantage of KDE lies in its flexibility, allowing for accurate representation even with limited sample points. The outcome of KDE is a smooth, continuous representation of data density, which contrasts with the discrete nature of histograms. Additionally, KDE is less sensitive to the selection of bins, reducing the impact of bin placement on the analysis results. However, there are limitations associated with KDE. To utilize KDE effectively, it is imperative to retain the entire dataset in memory, which can be a significant drawback when dealing with large datasets, as it necessitates substantial memory resources. Consequently, KDE may not be a practical choice for datasets with a considerable number of samples due to its computational expense. (Xie, Y., Density Estimation PDF.)

Findings:

(a) First, we will form the 1-dimensional histogram and KDE to estimate the distributions of amygdala and acc, respectively. ignore the variable orientation. Decide on a suitable number of bins so you can see the shape of the distribution clearly. Set an appropriate kernel bandwidth h > 0.

After trying numerous options, it was decided the suitable number of bins that best captured the shape of the distribution was 12 for the histogram, & a KDE bandwidth of .2 for both features.

Note: (amygdala, acc) have been standardized throughout this analysis.

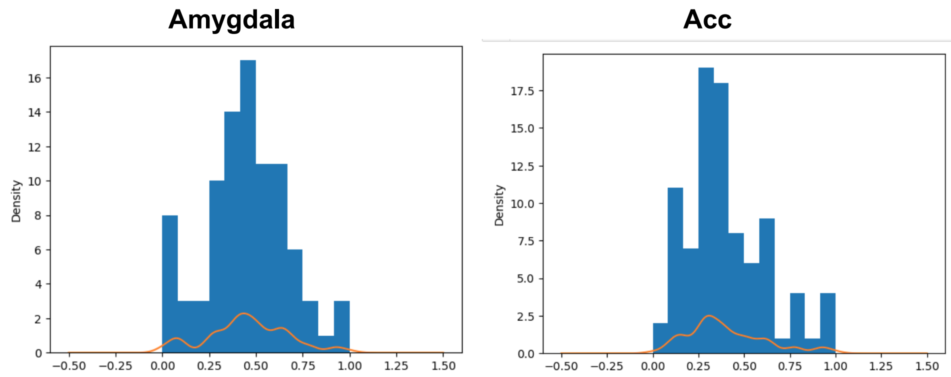
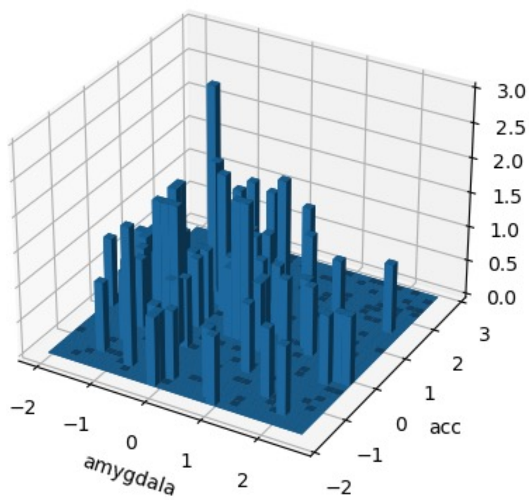


Figure 1: Histogram bin count 12; KDE bandwidth .2 for both features.

(b) Form 2-dimensional histogram for the pairs of variables (amygdala, acc). Decide on a suitable number of bins so you can see the shape of the distribution clearly.

For the 2-dimensional histogram, 30 bins produced the best clarity for (amygdala, acc). Outliers are noted along the edges.



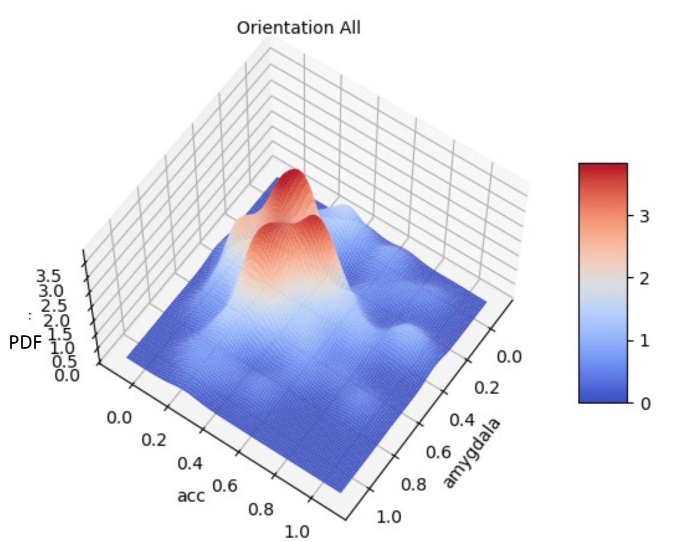
(c) Use kernel-density-estimation (KDE) to estimate the 2-dimensional density function of (amygdala, acc) (this means for this question, you can ignore the variable orientation). Set an appropriate kernel bandwidth h > 0. Please show the two-dimensional KDE (e.g., two-dimensional heat-map, two-dimensional contour plot, etc.) Please explain what you have observed: is the distribution unimodal or bi-modal? Are there any outliers? Are the two variables (amygdala, acc) likely to be independent or not? Please support your argument with reasonable investigations.

In the (KDE) 2-dimensional density function of (amygdala, acc), we see the distribution is bi-modal, *right.*

The concentration of density is noted by the color bar chart, with areas in dark red indicating the highest density, down to areas in dark blue exhibiting the lowest.

Outliers are denoted by the small ‘hills’ near the edges of the space.

While a strong interdependence is not evident in this figure, there does appear to be a joint drifting of both values lower, illustrated in the second “peak.’



2c. Continued, next page

2 c., (cont.) Are the two variables (amygdala, acc) likely to be independent or not? (NOTE: It actually involves prerequisite knowledge. From traditional probability and statistics, how do you show that two random variables are independent? Once you can answer that, then you need to visually represent that rule somehow and test for independence.) Please support your argument with reasonable investigations.

Independence of Variables: From traditional probability and statistics, we show two random variables are independent as per the definitions, below:

"Two random variables are independent if knowing the value of one does not change the probability of the other. This means that if X and Y are independent, we can write:

$$P(Y=y \mid X=x)=P(Y=y)$$

for all x,y.

"The probability that a discrete random variable X takes on a particular value x i.e. P(X=x) is denoted by f(x) and is called the probability mass function (p.m.f.).discrete: probability mass function (p.m.f.). It is referred to as the probability density function (p.d.f.) for continuous random variables." https://towardsdatascience.com/independence-covariance-and-correlation-between-two-random-variables-197022116f93

There are several **Investigations & tests for independence:**

i. Mutual information: returns zero for independent variables

The computation of the mutual information function is based on the application of nonparametric techniques, based on entropy estimation from k-nearest neighbors distances. This methodology is drawn from the foundational concepts initially introduced by L. F. Kozachenko, & N. N. Leonenko in 1987 [1]. Mutual information (MI) serves as a mathematical measure quantifying the dependency between two variables. This measure assumes non-negative values. MI attains a value of zero if the two variables are independent, with larger MI values indicative of greater interdependency.

Sklearn.feature_selection "mutual_info_regression " estimates mutual information for a continuous target variable. See "mutual_info_classif" for a discrete target variable.[2]

1. L. F. Kozachenko, N. N. Leonenko, “Sample Estimate of the Entropy of a Random Vector:, Probl. Peredachi Inf., 23:2 (1987), 9-16
2. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.mutual_info_regression.html

While built-in functions from the libraries noted above will be used for calculating MI in this analysis, the mathematics for calculating Mi is noted, below. Note the similarity to the equation from traditional probability and statistics, above.

"Let (X, Y) be a pair of random variables with values over the space $\mathcal{X} \times \mathcal{Y}$. If their joint distribution is $P_{(X,Y)}$ and the marginal distributions are P_X and P_Y, the mutual information is defined as

$$I(X;Y) = D_{KL}(P_{(X,Y)} || P_X \otimes P_Y)$$

where D_{KL} is the Kullback–Leibler divergence, and $P_X \otimes P_Y$ is the outer product distribution which assigns probability $P_X(x) \cdot P_Y(y)$ to each (x, y). Notice, as per property of the Kullback–Leibler divergence, that $I(X;Y)$ is equal to zero precisely when the joint distribution coincides with the product of the marginals, i.e. when $X and Y$ are independent (and hence observing Y tells you nothing about X). $I(X;Y)$ is non-negative, it is a measure of the price for encoding (X, Y) as a pair of independent random variables when in reality they are not." https://en.wikipedia.org/wiki/Mutual_information

2 C. (cont.)
Are the two variables (amygdala, acc) likely to be independent or not? (NOTE: It actually involves prerequisite knowledge. From traditional probability and statistics, how do you show that two random variables are independent? Once you can answer that, then you need to visually represent that rule somehow and test for independence.) Please support your argument with reasonable investigations.

investigations and visualizations to test independence of the variables (amygdala, acc :

i. Mutual information (cont.)

Mutual information scores with scatter plots for (amygdala, acc)

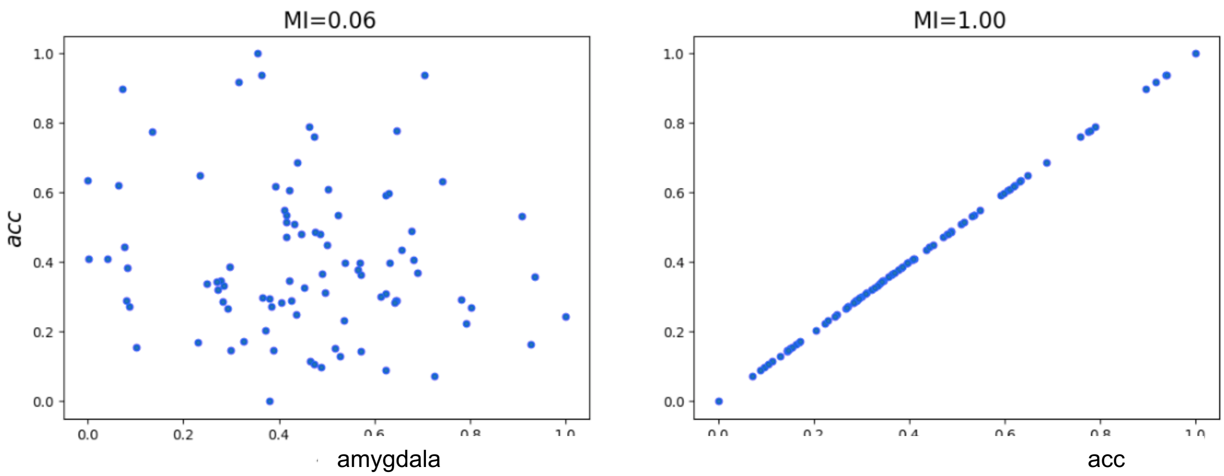


Figure 2: The scatter plot on the left (amygdala, acc) visualizes the low interdependence between these two variables.

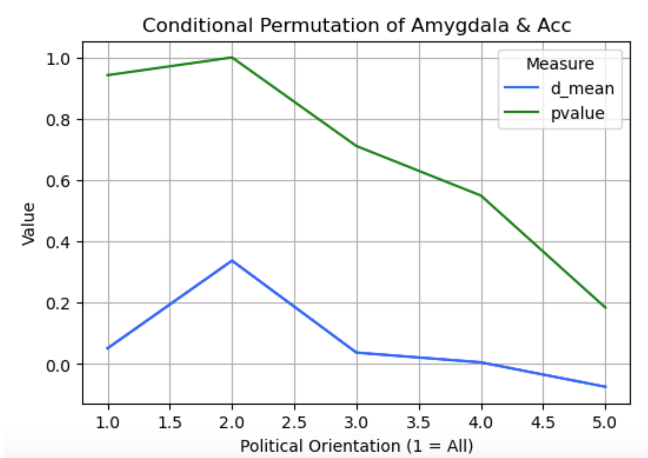
Full independence of variables is indicated by an MI score of 0. The MI score for the two variables (amygdala, acc) is .06. This score indicates **these variables do have a low interdependence and are *not* independent**. This low Mi score is visualized in the scatter plot for (amygdala, acc) in Figure 2, *left*. We see a drifting of acc values downward as amygdala values increase. Also note the concentration of both features in the middle range of each of their values. In contrast, the scatter plot, *right*, of (acc , acc), with a higher MI of 1, results in a linear scatter plot . This is because (of course!) a variable shares a high dependance with itself.

ii. Permutation: A p-value greater than 0.05 means that deviation from the null hypothesis is not statistically significant. P-value for a hypothesis test whose null hypothesis is that two samples have no ordinal correlation.

We see in the chart, right, that's although the p-value for the variables in this analysis decreases as the political orientation moves from 2 to 5, nonetheless it remains well above **0.05,** indicating the interdependence between (amygdala, acc) is low & not statistically significant.

This is in agreement with the Mi findings*: the interdependence between (amygdala, acc) is low.*

iii. The low interdependence between these two variables will continue to be observed and noted throughout the following sections



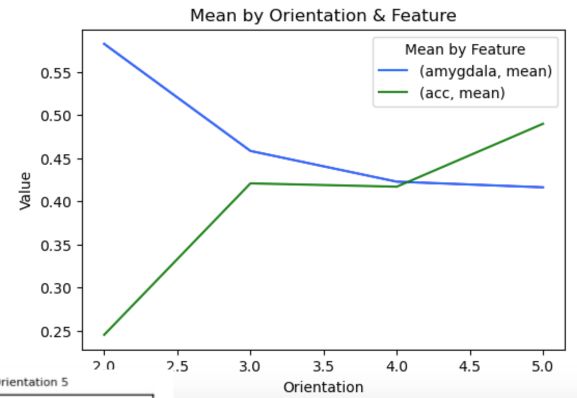
	1	2	3	4	5
d_mean	0.051590	0.337416	0.037621	0.005813	-0.073711
pvalue	0.942011	0.999760	0.710923	0.549585	0.184478

Figure 3: Here, 'd_mean' indicates difference between the means of the two variables. Orientation '1' indicates the full data set.

2. Density estimation: Psychological experiments (cont.)

(d) We will consider the variable orientation and consider conditional distributions. Here we plot the estimated conditional distribution of amygdala & acc, conditioning on political orientation: p(amygdala | orientation = c), c = 2; : : : ; 5, using KDE. We set an appropriate kernel bandwidth h > 0. We also fill out the conditional sample mean for the two variables:

	orientation	2	3	4	5
amygdala	mean	0.582403	0.458249	0.422584	0.416051
acc	mean	0.244987	0.420629	0.416771	0.489761



Here we explain, based on the results, we can infer that the conditional distribution of amygdala and acc, respectively, are different from c = 2; : : : ; 5. This is a type of scientific question one could infer from the data: whether or not there is a difference between brain structure and political view.

Figure 4: Conditional mean for (amygdala, acc), above.

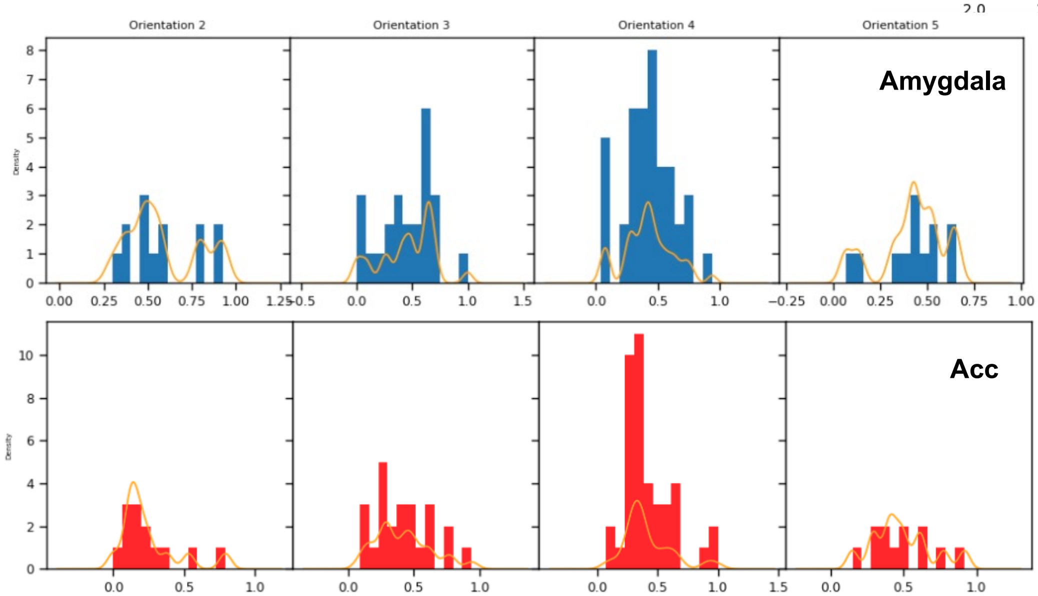


Figure 5: KDE & histogram. The Y axis represents both Density (KDE) & Frequency (histogram).

Conclusion: Yes, one can infer that the conditional distribution of amygdala and acc, respectively, are different from c = 2; : : : ; 5, for each political orientation. There is a difference between brain structure and political views. As we observe (amygdala, acc) conditionally, we notice increased interdependence. The two variable have an inverse movement of their means as we progress from c = 2; : : : ; 5. When examining the conditional mean for these features, we noticed that acc is lower for liberal political orientation, and increases as political orientation moves to the conservative end of the spectrum. The inverse is true for amygdala, with its mean highest among those claiming to be most liberal, or ‘2”, and decreasing to its lowest among those at the conservative end of the spectrum, ‘5’. (Oddly, this is the opposite of findings in all other studies on this topic.)

This similarity of movement is noticed, further, when comparing the skewedness of (amygdala, acc), as observed in both the KDE & histogram in Figure 5, & as we as in the Skew plots in Figure 6. The two variables follow parallel patterns of increasing/decreasing variance among Political Orientations 2, 3, and 4; however for Orientation 5, amygdala's variance is at its lowest point, while acc is at its second highest.

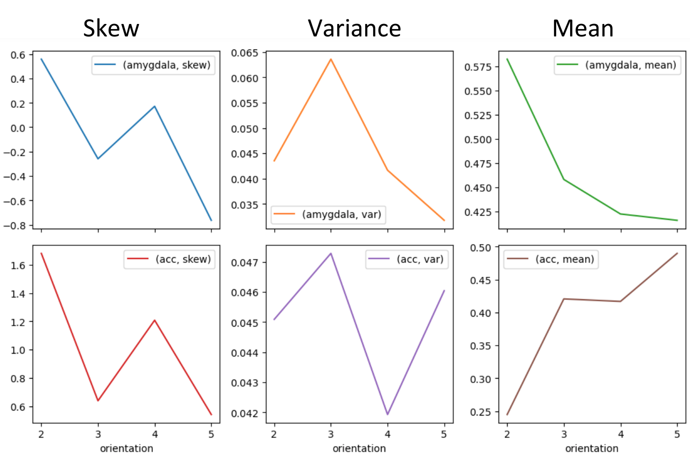


Figure 6: Conditional Skew, *Variance & Mean for (amygdala, acc), left & below.*

	orientation	2	3	4	5
amygdala	skew	0.558530	-0.259143	0.171706	-0.763869
	var	0.043525	0.063584	0.041663	0.031774
	mean	0.582403	0.458249	0.422584	0.416051
acc	skew	1.681008	0.638290	1.208162	0.540886
	var	0.045091	0.047283	0.041928	0.046042
	mean	0.244987	0.420629	0.416771	0.489761

2. Density estimation: Psychological experiments (cont.)

(e) Again we will consider the variable orientation. We will estimate the conditional joint distribution of the volume of the amygdala and acc, conditioning on a function of political orientation: p(amygdala,acc I orientation = c), c = 2; : : : ; 5. We will use two-dimensional KDE to achieve the goal; et an appropriate kernel bandwidth h > 0. Please show the two-dimensional KDE (e.g., two-dimensional heat-map, two-dimensional contour plot, etc.).

Please explain based on the results, can you infer that the conditional distribution of two variables (amygdala, acc) are different from c = 2; : : : ; 5? This is a type of scientific question one could infer from the data: Whether or not there is a difference between brain structure and political view.

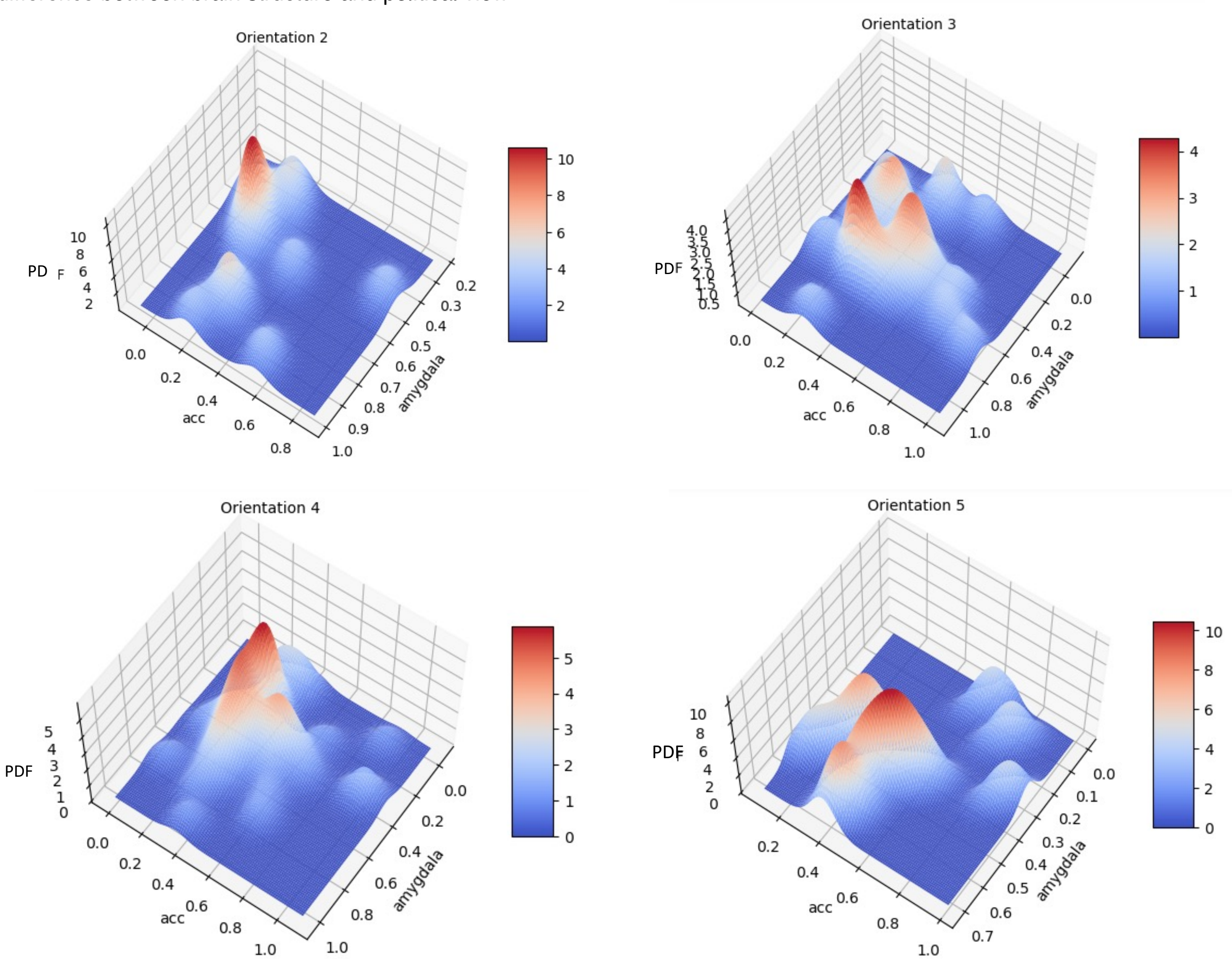
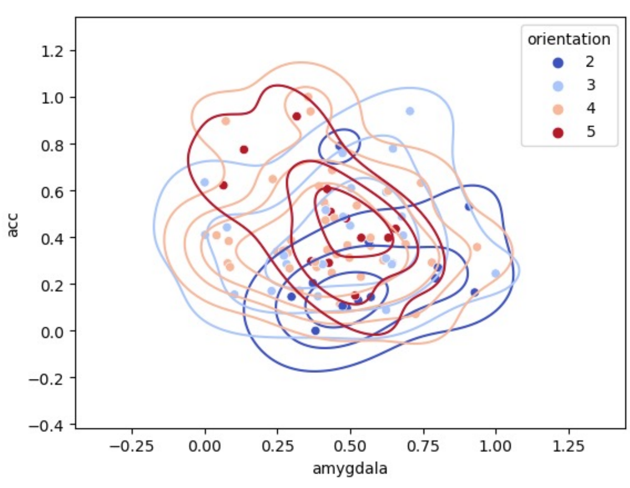


Figure 7: a) Two-dimensional KDEs of the conditional distribution of amygdala and acc, above. b) Scatterplot & kde, combined, of the conditional distribution, below.



Yes, one can see in these two-dimensional KDEs that the conditional distribution of amygdala and acc, respectively, are different from c = 2; : : : ; 5, for each political orientation. There is a difference between brain structure and political views.

Notice the movement of density at its highest level is very subtle. In Figure 7a, Orientations 2 and 5 have the two highest PDF scales, at 30 and 10, respectively. In contrast, Orientations 3 and 4's PDF scales reach 4 and 5.

Orientations 2 & 5 in Figure 7b exhibit some similarities in appearance to the orthogonal eigenvectors of PCA, in tat the overall distribution moves in almost perpendicular directions.

2. Density estimation: Psychological experiments (cont.)

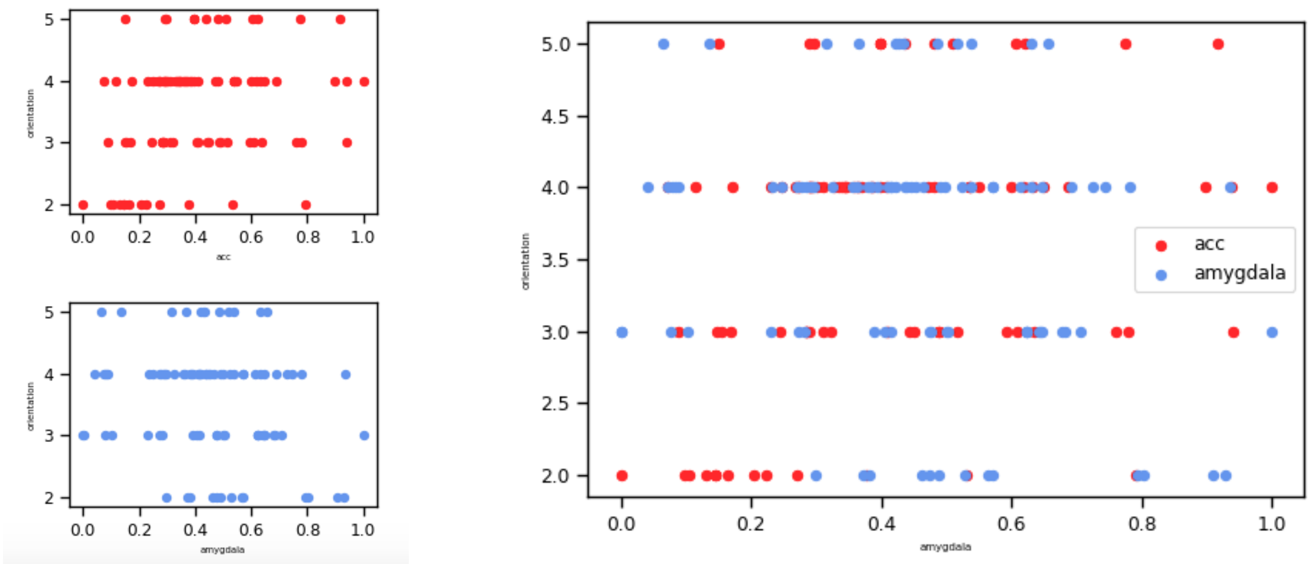
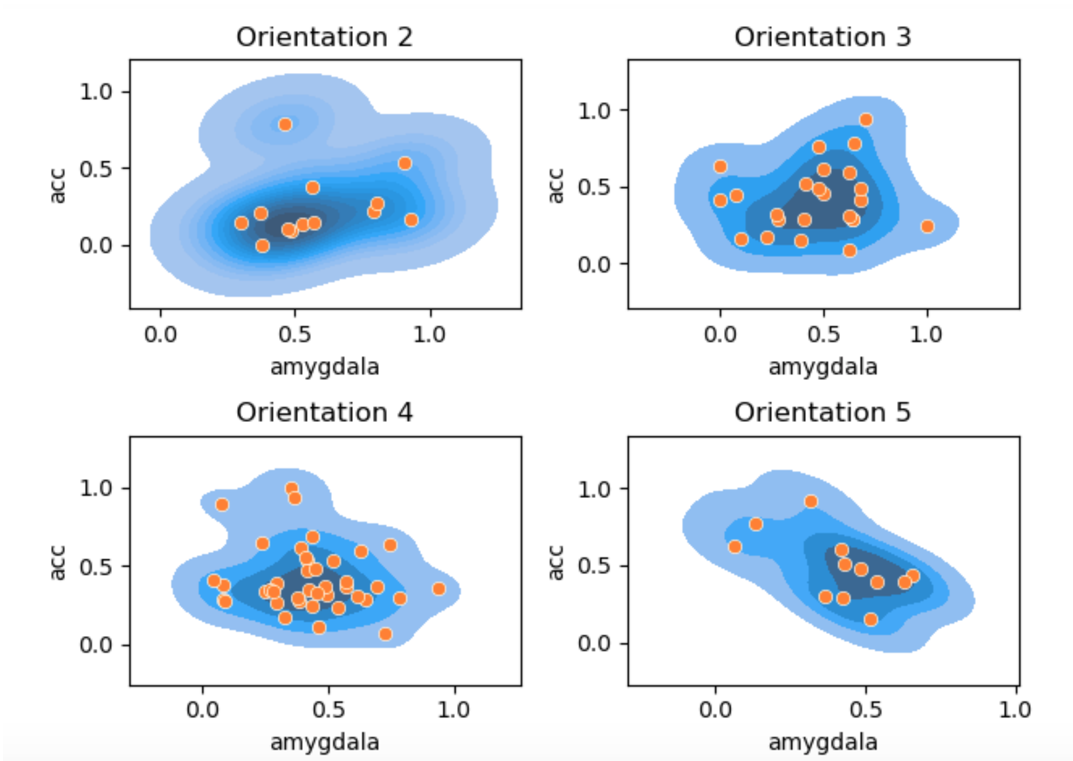


Figure 8: a) Additional Scatter Plots illustrating the distribution of acc and amygdala, *above*. Here again we see how acc is in lower volumes at Orientation 2, and moves to higher volumes at Orientation 5. We see the reverse is true for amygdala. b) KDE & scatterplot combined, *below*. Orientation 2 & 5 remind me of a perpendicular distribution, as we see in PCA.



KDE: "The units on the density axis are a common source of confusion. While kernel density estimation produces a probability distribution, the height of the curve at each point gives a density, not a probability. A probability can be obtained only by integrating the density across a range. The curve is normalized so that the integral over all possible values is 1, meaning that the scale of the density axis depends on the data values."
https://seaborn.pydata.org/generated/seaborn.kdeplot.html