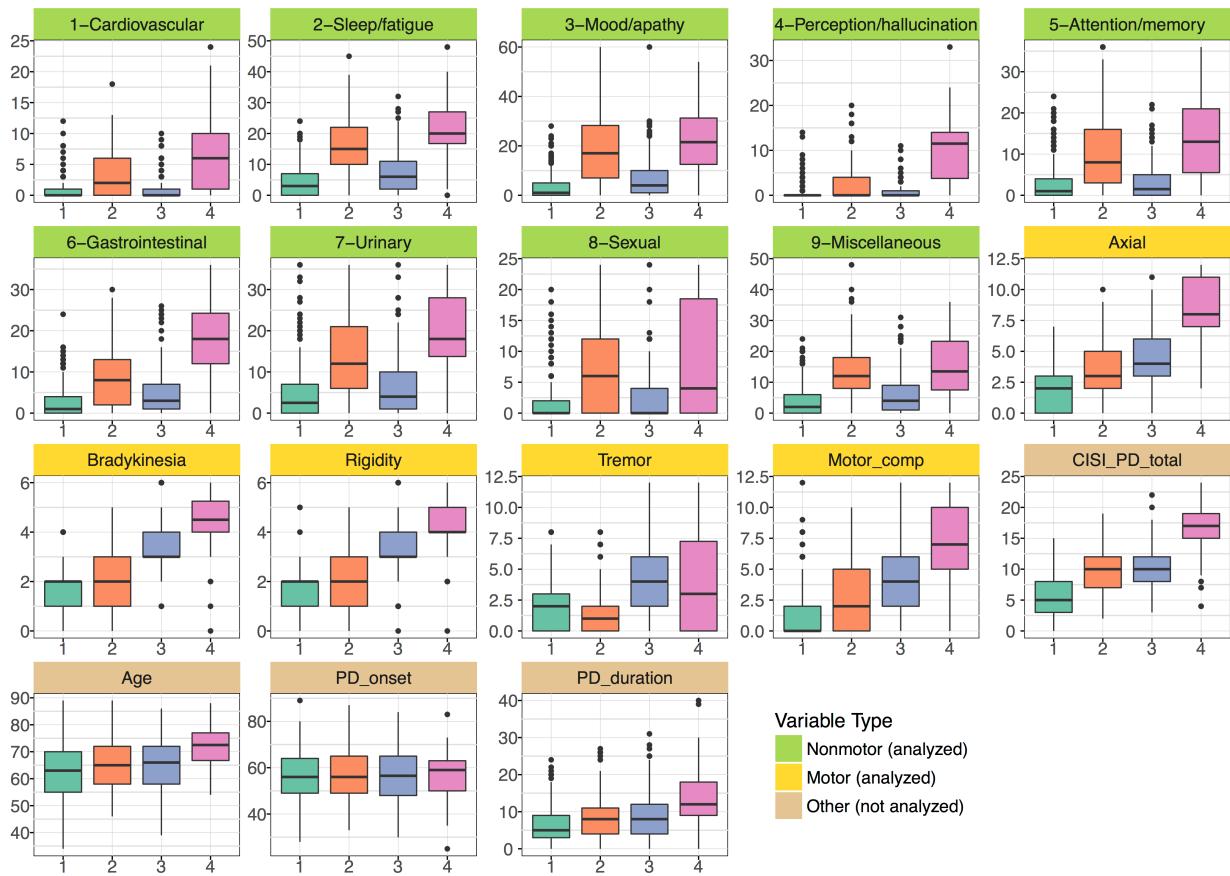
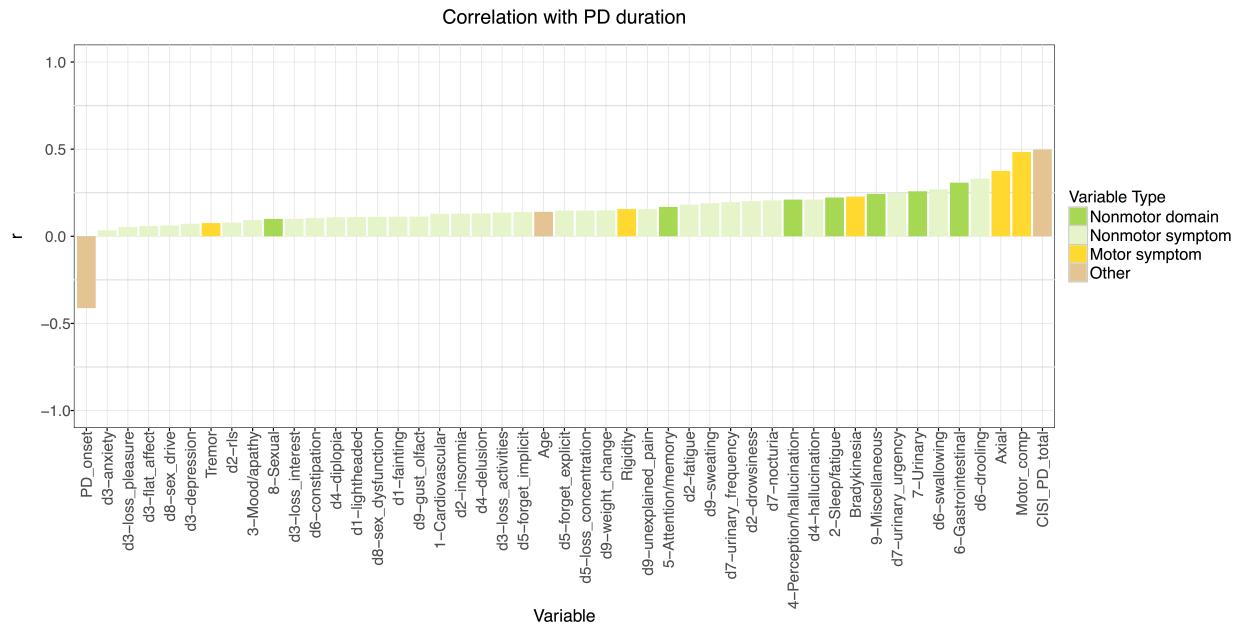


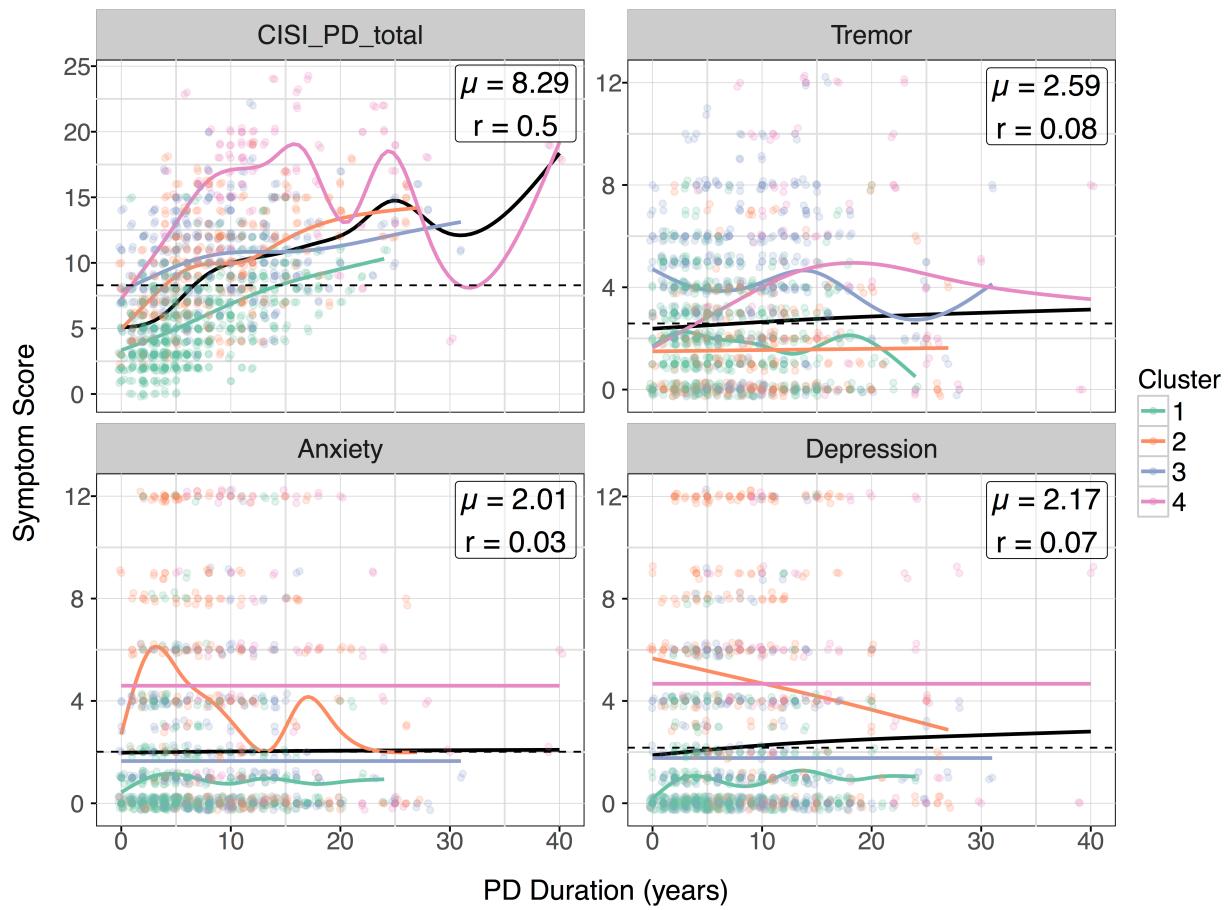
Supplementary Figure 1. Gap statistics. Plot of the gap statistic $\text{Gap}(k)$ versus number of clusters with k-means on 500 bootstrapped samples of a) the domains clustering, and b) the symptoms clustering. Error bars represent ± 1 standard error (se). Per the method described in Tibshirani *et al.* (2001), the optimal number of clusters is the smallest k such that $\text{Gap}(k) \geq \text{Gap}(k + 1) - \text{se}_{k+1}$. For the domains clustering, $k = 4$; for the symptoms clustering, $k = 6$. The gap statistic for the optimal k and the comparison to $k + 1$ are marked with dotted lines.



Supplementary Figure 2. Domains clustering boxplots. Boxplots for domains clustering for each symptom and cluster.



Supplementary Figure 3. Correlation with disease duration. Correlation of applicable variables with disease duration.



Supplementary Figure 4. Symptoms against disease duration. For clarity, scatterplot points are colored according to cluster and jittered. Smoothed loess curves for each cluster are drawn in their respective cluster colors. The black curve is the curve for the entire population, and the global mean score is marked with a dotted line.